Development of A Real-Time Business Intelligence (BI) Framework based on Hex-Elementization of Data Points for Accurate Business Decision-Making.

Girish Nair

Doctor of Philosophy, 2019

PRIMARY SUPERVISOR: Professor Yi-Chen Lan, Western Sydney University, Sydney, Australia

SECONDARY SUPERVISOR: Professor Bhuvan Unhelkar, University of South Florida, Sarasota-Manatee, Florida, USA.

Western Sydney University

School of Business

NSW, Australia

Dedication

I dedicate this thesis to my father, who passed away 25 years ago, while doing his Ph.D.

This is for you papa.

Acknowledgment

I would like to first acknowledge my teachers (supervisors). Thank you Prof Yi-Chen your guidance, support and encouragement during my candidature. Thank you in believing in me and taking me under your umbrella. A special thanks to Prof Bhuvan. He said this is going to be a character-building exercise, and what an exercise this has been. Thank you for providing the appropriate nudge at appropriate times. Thank you for all your guidance to ease this journey.

Secondly, I thank my mentor and manager Nigel. Thank you for being patient with me and letting me off work on those innumerable Fridays. Thank you for believing in me. I thank him and my colleagues at work for not asking any tough questions regarding my thesis.

I personally thank James Harris and Colleen Berish for kindly reviewing and painfully reading this thesis. Thank you for making this research read well!

I am grateful to my mother for providing me with the upbringing, which was the source of inspiration, resilience and hard-work needed to complete this journey. Finally, I thank my loving wife, Anju. Her strong belief in me, unconditional love and support has sustained me throughout this journey. I am sorry to my son Ralph and daughter Norah for staying away from them to work on this thesis.

A special tribute to “Patience”, without which it was not possible to complete this journey.

Statement of Authenticity

I, Girish Nair, declare that this thesis is submitted in accordance with the requirements of the award of Doctor of Philosophy, in the School of Business, Western Sydney University, and the work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text.

I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

Girish Nair

Candidate

[redacted]

Signature

15-June-2019

Date

Abstract

The desire to use business intelligence (BI) to enhance efficiency and effectiveness of business decisions is neither new nor revolutionary. The promise of BI is to provide the ability to capture interrelationship from data and information to guide action towards a business goal. Although BI has been around since the 1960s, businesses still cannot get competitive information in the form they want, when they want and how they want. Business decisions are already full of challenges. The challenges in business decision-making include the use of a vast amount of data, adopting new technologies, and making decisions on a real-time basis. To address these challenges, businesses spend valuable time and resources on data, technologies and business processes. Integration of data in decision-making is crucial for modern businesses. This research aims to propose and validate a framework for organic integration of data into business decision-making. This proposed framework enables efficient business decisions in real-time. The core of this research is to understand and modularise the pre-established set of data points into intelligent and granular "hex-elements" (stated simply, hex-element is a data point with six properties). These intelligent hex-elements build semi-automatic relationships using their six properties between the large volume and high-velocity data points in a dynamic, automated and integrated manner. The proposed business intelligence framework is called "Hex-Elementization" (or "Hex-E" for short).

Evolution of technology presents ongoing challenges to BI. These challenges emanate from the challenging nature of the underlying new-age data characterised by large volume, high velocity and wide variety. Efficient and effective analysis of such data depends on the business context and the corresponding technical capabilities of the organisation. Technologies like Big

Data, Internet of Things (IoT), Artificial Intelligence (AI) and Machine Learning (ML), play a key role in capitalising on the variety, volume and veracity of data. Extricating the "value" from data in its various forms, depth and scale require synchronizing technologies with analytics and business processes.

Transforming data into useful and actionable intelligence is the discipline of data scientists. Data scientists and data analysts use sophisticated tools to crunch data into information which, in turn, are converted into intelligence. The transformation of data into information and its final consumption as actionable business intelligence is an end-to-end journey. This end-to-end transformation of data to intelligence is complex, time-consuming and resource-intensive.

This research explores approaches to ease the challenges the of end-to-end transformation of data into intelligence. This research presents Hex-E as a simplified and semi-automated framework to integrate, unify, correlate and coalesce data (from diverse sources and disparate formats) into intelligence. Furthermore, this framework aims to unify data from diverse sources and disparate formats to help businesses make accurate and timely decisions.

Keywords: Business Intelligence, Data Integration, Hex-Elementization, Hex-E, Organic Business Intelligence, Automated Business Intelligence, Business Intelligence Framework, Big Data, Artificial Intelligence, Business Decision-making

Table of Contents

Table of Figures

Table of Tables

Abbreviations

No.	Abbreviation	Definition
1	AI	Artificial Intelligence
2	BA	Business Analysis
3	BI	Business Intelligence
4	CAQDAS	Computer-Assisted Qualitative Data Analysis Software
5	CEO	Chief Executive Officer
6	CIO	Chief Investment Officer
7	COO	Chief Operating Officer
8	CRM	Customer Relationship Management
9	DA	Decomposition Analysis
10	DSS	Decision Support System
11	EA	Enterprise Architecture
12	EAP	Enterprise Architecture Planning
13	EIS	Executive Information System
14	ERP	Enterprise Resource Planning
15	EU	European Union
16	GDPR	General Data Protection Regulation
17	HaaS	Hex-E as a Service
18	Hex-E	Hex-Elementization
19	HPC	High Performance Computing
20	HR	Human Resources
21	IDC	International Data Corporations
22	IoT	Internet of Things
23	IPA	interpretative phenomenological analysis (IPA)
24	KPI	Key Performance Indicators
25	MIS	Management Information System
26	ML	Machine Learning
27	NLP	Natural Language Processing
28	OLAP	Online Analytical Processing
29	PCA	Principal Component Analysis
30	ROE	Return on Effort
31	ROI	Return on Investment
32	RTBI	Real-Time Business Intelligence
33	SMM	Social Media Marketing
34	TCP/IP	Transmission Control Protocol / Internet Protocol

Chapter 1: INTRODUCTION

Research Case

Business decision-making is full of challenges. These challenges emanate from the nature of the underlying data - primarily its sheer volume, velocity, and variety. Analysis of such data depends on the outcomes desired by the business and the technical capabilities of the organisation to process that data. The challenge is further complicated with the rapid evolution of technologies such as Big Data, Internet of Things (IoT) and Artificial Intelligence (AI). For example, when business organisations examined the social media giant Facebook, it was not immediately clear how the platform would help generate sales and revenue (Vieira, Costa, de Almeida, Costa, & Coelho, 2018). The value of Facebook, as later realised, is in the value of the information gathered about its users. In other words, the business value is a lot more than the overt data collection by an organisation. Extricating value from the data in its various forms, depth, and scale requires multiple disciplines of technologies, analytics and business processes to work together (Unhelkar, 2017). Transforming this data into useful and actionable intelligence is the capability of data scientists and data analysts. These data professionals make use of sophisticated analytical tools to transform data into information and intelligence. This end-to-end transformation of data to intelligence is complex, time-consuming and resource-intensive.

This research explores approaches to ease this end-to-end data to intelligence transformation. The research suggests a simplified and semi-automated framework, called Hex-Elementization, to help integrate, correlate and coalesce data into intelligence in real-time to help accurate decision-making. The research proposes a framework which helps integrate data into intelligence. The vision of this framework enables efficient business decisions in real-time, and with limited resources and time. Furthermore, the framework aims to unify data from diverse sources and disparate formats.

The core of this research is to modularise the pre-established set of data points into intelligent and granular "hex-elements". These hex-elements enable integrated and semi-automatic relationship-building between the large volume and high-velocity data points. The purpose of this relationship-building in data is to help businesses extract context-based insights which directly helps in decision-making. The proposed business intelligence framework is called "Hex-Elementization" also abbreviated as "Hex-E".

Integrated relationship-building requires that each data point is well-defined, modularised and open to connecting with other data points. Finding, correlating, connecting, and coalescing relationships between disparate data and information sets lead to real-time business intelligence. To achieve this objective, this research illustrates the inherent limitations of the existing way of connecting data points from diverse forms (e.g., text, multimedia and mixed format) in a business environment. The data challenges for businesses today include large volumes (Petabytes), high velocity (giga-bytes per second) and wide varieties (text, machine sensors, audio, graphics, videos and mixed). Data, disparate in content and medium, needs to be correlated in a meaningful manner to enhance decision-making. This data integration not only facilitates accurate decision-making but also enables agility in decision-making (Unhelkar, 2013). Enabling these data points to coalesce semantically is not possible with human capacity due to the sheer size and quantity of data.

Machine-to-machine interaction (Semantic Web) and automated-relationship forming are essential in creating actionable insights within a short time.

Figure 1 shows how datasets from disparate sources of data pass through the Hex-Elementization process to create generic hex-elements. These hex-elements are further used to connect, coalesce and correlate data to form meaningful information. The Context Engine translates the goals from the different entities the organisation deals with to the data available post the Hex-Elementization process.

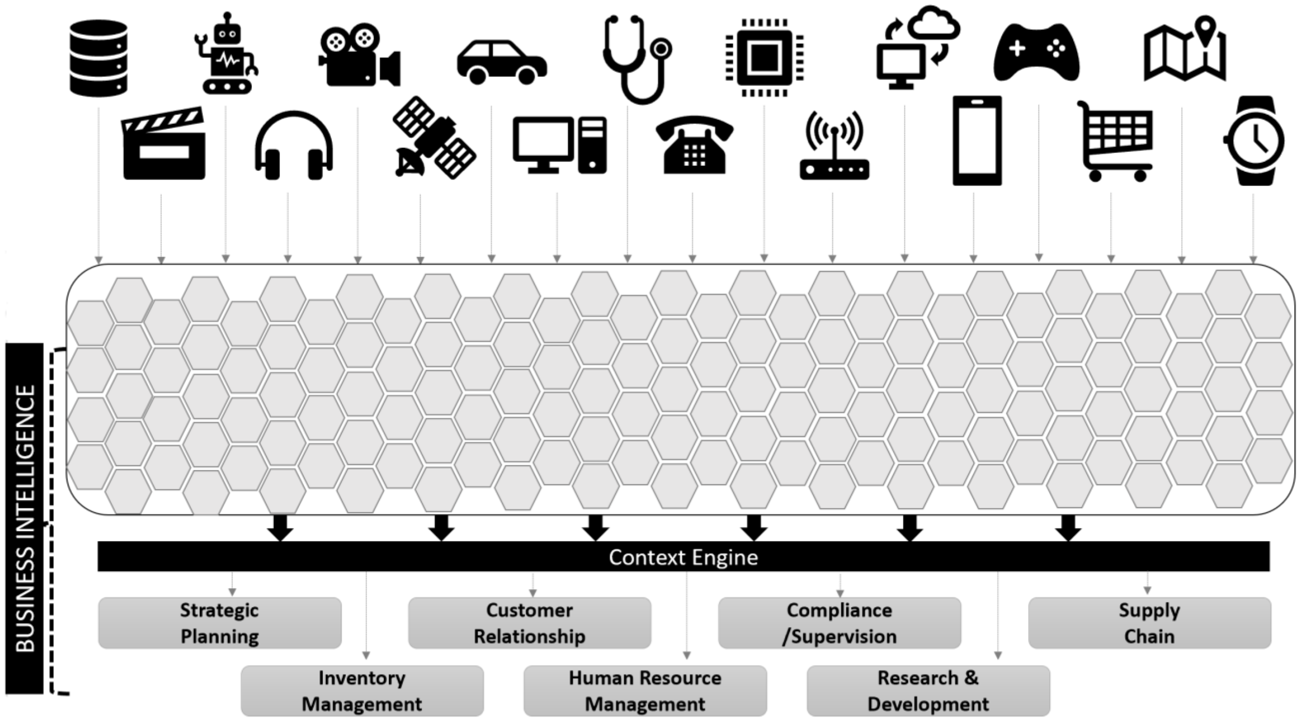

Figure 1: The concept of Hex-Elementization

Thus, the research addresses challenges that include breaking down data complexity from various sources to form atomic level basic data points, automatically connecting data based on context and enabling real-time decision-making by businesses. The research objectives are explained in the next section.

Research Problem Statement

The problem statement in the context of the proposed research is briefly explained below. The problem covers on what do we already know, what do we need to know, why do we know it, and what will the research do to find out.

Data is taking a centre stage in decision-making. Organisations are turning towards a more data-driven approach when it comes to decision making. However, data is constantly evolving as the variety, velocity, volume and veracity changes. Most of the organisations are not aware of the fact that they already capture a lot of data. Extracting intelligence from this data will help them make timely and accurate decisions.

This research addresses the current difficulty in extracting intelligence in real-time due to the complexity inherent in data. This research proposes a methodology called Hex-Elementization which aims to break down data into the most granular form with six unique characteristics. The process of Hex-Elementization aims to break down the data to a common denominator to help correlate, coalesce and connect disparate data sets to enable extracting business intelligence. This extraction of business intelligence is proposed to be contextual driven in this business-intelligence framework.

Organisations make decisions all the time. These organisations need a mechanism to extract business intelligence on a real-time basis. Currently, the business intelligence frameworks are pull-based in which organisational stakeholders define their business goals and aims. Then an army of business analysts, programmers, database administrators along with sophisticated analytical tools, work together to translate business goals into identifiable technical actions to be applied to the data they have. This entire end-to-end journey, from defining business goals to getting insights, takes time and resources. This research proposed a methodology which aims to ease this journey of data into intelligence.

The research will first aim to understand the challenges business stakeholders in varied industries face in trying to extract intelligence to make optimal business decisions. Then the research will aim to propose a model which can be applied in various organisations along with other technologies to help extract business intelligence on a real-time basis.

Research Objectives

Based on the previous discussion, the main objectives of this research are summarised as follows.

1. **Examine the relationship between data and the challenges of real-time business decision-making.**

   This objective examines various types of data used in a corporate environment and how the data is used in decision-making. The aim is to understand if enough granularity is embedded in the data for business to make optimal business decisions. The challenges in decision-making include the need for accuracy, time available to make the decisions and the level of granularity required. The impact of delay in making decisions due to the lack of data or uncorrelated information is studied in this research. Also, the benefit of a real-time solution for business insights is investigated in detail in this research.

2. **Develop a framework that enables data sets to connect, correlate and coalesce without intermediary processes or human intervention.**

   This objective is explored by understanding how the data, and the information emanating from it, flows through the organisation, processes and people. The aim of this research objective is to understand, through existing frameworks (i.e., in Literature Review), how disparate information connects to help generate insights. The impact of data on processes and people is also explored here.

3. **Develop a conceptual Hex-Elementization platform in which data correlation occurs symbiotically in processes based on the inherent nature of things.**

   This objective explores a set of protocols, in Enterprise Architecture platforms, to find an optimal way of organically relating data sets. The conceptual Hex-E platform is based on the premise that each data point is defined in the most optimal way through its six parameters.

4. **Explore the impact of Hex-Elementization on various business competencies, and contextual-based decision.**

   This objective is examined by understanding how businesses pull data and information. When businesses explicitly demand (pull) data and information, they first need to know what they are looking for. While a push-based system (such as Hex-Elementization) provides data, information and intelligence to the business based on pre-defined rules or context. Whether contextual-based information, pushed through a framework, is viable through Hex-Elementization framework, is the main explorative aim of this research. This research also investigates the practicality and impact of information flow on business structure and enterprise architecture.

5. **Develop Hex-Elementization based on business context and desired outcomes rather than entirely on data-analytics.**

   This objective is examined by understanding how businesses can benefit from a decision-making approach based on context, as compared to decisions based entirely on pre-defined business rules.

Philosophical Background of Hex-Elementization

Business intelligence is not a new concept. The desire to use BI to enhance efficiency, effectiveness and tactical business decisions has been around since the advent of computers. With the introduction of Big Data and IoT (Nair & Lan, 2016), there is a need to re-think the way BI is applied. The organisational challenges in real-time decision-making include resources needed to analyse the data in ever-shortening time periods.

Hex-Elementization proposes a platform enabling disparate datasets to communicate with each other in real-time based on the context in which businesses make decisions. The Hex-Elementization model aims to help businesses configure the platform to suit the context in which decisions are made. The Hex-Elementization model aims to find a way to make data, and information from this data, connect organically relative to the business context. Businesses only need to provide the context required for decision-making, and Hex-E responds with the information derived from various datasets seamlessly to businesses, relative to this context. Hex-Elementization begins with the premise that the optimal way to describe each data point is through six attributes or characteristics. Therefore, an essential question in developing this platform is "Why Hex or six sides? ". According to Schneider, every natural pattern of growth or movement inevitably conforms to one or more simple geometric types (Schneider, 2003). Schneider further articulates numbers, shapes and their patterns symbolise omnipresent principles, including wholeness, polarity, structure, balance, cycles, rhythm and harmony, from the most basic building blocks and growth patterns. Furthermore, in surprising ways, multiples of six, particularly twelve, thirty-six, and sixty emerge as a natural framework in mathematics, nature, the symbolic arts and everyday affairs (Schneider, 2003).

The hexagon, as a shape, is found in many basic structures. For example, honeycombs and ice crystals have six sides; and even the recently discovered strong element, graphene, (Ball, 2016) has six sides. The hexagonal shape is trusted by nature. Nature finds this structure helpful, effective and efficient. For example, hexagonal cells of beehives package the maximum space using the least materials, energy and time (Schneider, 2003).

One reason for this harmony is that hex structures lend themselves to the tight knitting of a variety of configurations with the most distinct attributes (Ball, 2016). Other shapes (e.g., octagonal) also lend themselves to configurations, but they quickly start becoming less efficient. This lack of efficiency could be the result of the effort required to select and connect the right properties, or attributes, from the shapes or data points. Conversely, less complicated shapes like squares or triangles limit the number of attributes that can be used to describe a data point resulting in limited effectiveness in connecting. Thus, the hexagon seems to be ideal for integration. Laplace, the famous French astronomer and mathematician said, “All the effects of nature are only the mathematical consequences of a small number of immutable laws” (Laplace, 1921).

The hexagon is a geometrical shape that promises to integrate easily with other hexagonal shapes. By defining data points with six properties, the Hex-Elementization platform can help configure or unify multiple elements more easily than other ways of defining data. The first step is to break the complex and disparate set of data available to businesses into an atomic piece of data. Each data is then described with six distinct attributes or properties. These attributes are used to connect, coalesce and correlate a data point with another data point in real-time under the Hex-E umbrella.

Hex-element is the most basic part of the overall process of Hex-Elementization. Conceptually, a hex-element is a data element defined by six unique properties or attributes. A hex-element is not a byte or bit, data record or an address in the computer’s registry. Hex-element is an outcome of a group of characteristics, in this case, six. Hex-element is an embodiment of six attributes which uniquely identify an intelligible part of data which enables coalescence when combined with other hex-elements. Figure 2 shows how a hex-element renders itself with six properties or characteristics.

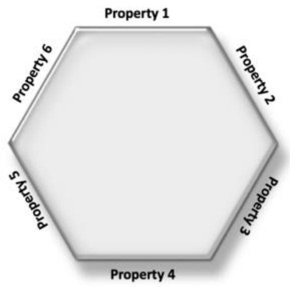

Figure 2: Hex-element with six attributes and six sides

Two hex-elements can connect based on matching properties (see Figure 3). This matching is the essence of the Hex-Elementization process in which disparate atomic data, as denoted by hex-elements, form relationships based on common and shared properties.

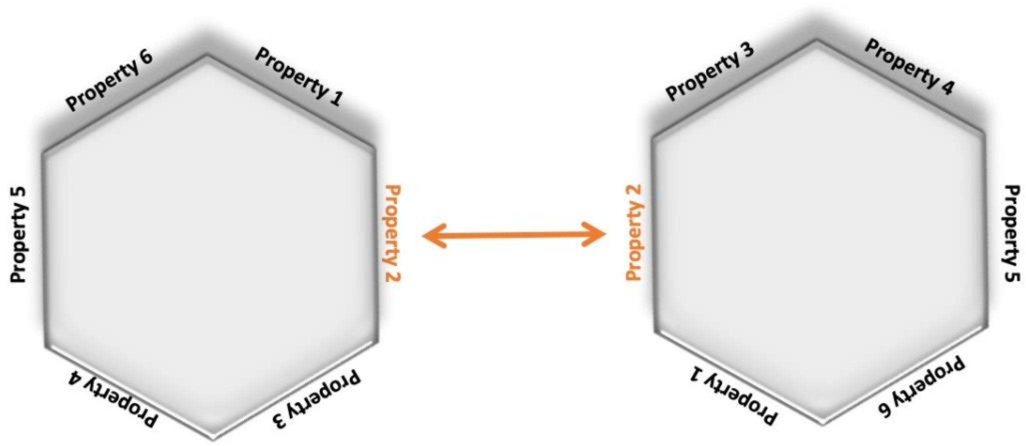

Figure 3: Two hex-elements connecting using common property

The discussion until now presents the concept of Hex-Elementization. The next section presents the research questions.

Research Questions

1. Businesses have access to the vast amount of data that is continuously growing. Businesses extract some information snippets from this data based on available tools and processes. There is a need for businesses to transform this data into information and subsequently into intelligence. How do businesses prepare themselves to utilise data? This will help in answering the first Research Objective 1 – as highlighted below.

   Examine the relationship between data and the challenges of real-time business decision-making.

2. What are the challenges in creating and utilising business intelligence in understanding the current state of the business and how can this business intelligence be used to undertake business transformation? How do businesses enable intelligent decision-making in real-time by streamlining their data ecosystem and creating rapid correlations between otherwise unrelated data sets? This research question is aimed to help achieve the research outcome described below.

   Develop a conceptual Hex-Elementization platform in which data correlation occurs symbiotically in processes based on the inherent nature of things.

3. How does the increased use of unstructured data impact decision-making? How will developments in data science augment decision-making?

4. Is the battle of Quality vs Quantity of data more prominent in an environment in which data science is being incorporated in every known industry? How does the Quality vs. Context continuum survive in such a changing environment?

   Develop Hex-Elementization based on business context and desired outcomes rather than entirely on data-analytics.

5. What are the various business risks associated with the effort to connect, coalesce and correlate datasets that produce intelligence for business to help with real-time decisions? This research question is aimed to help answer the research objectives below.

   - ***Explore the impact of Hex-Elementization on various business competencies, and contextual-based decision.***
   - ***Develop Hex-Elementization based on business context and desired outcomes rather than entirely on data-analytics***

Research Outline

Improving decision-making within a corporate environment is a result of considering as many factors as possible (Kakkar & Gordon, 2018). A key factor in decision-making is the value provided by data. Literature indicates that defining an entity with six elements provides maximum value in integration (Schneider, 2003). This research explores a hexagonal mapping for data and its impact on decision-making.

The six sides of a hexagon make it possible to create rich extendible structures. The richness of hexagonal structures is also extremely malleable (Ball, 2016), resulting in the potential to apply these hexagonal structures to a wide-ranging set of problems and challenges. For example, hexagonal structures can be valuable in robotics, supply chains, financial markets and health care. When nano-robots are tasked to merge and form bigger, unified structures, then the most granular part of this structure must be in the shape of hexagons

The concepts of economics are applicable to data utilisation as well. Economics within technology deals with demand and supply of data and intelligence. For example, there is a continuous wave of data (i.e., supply), however the consumption of data (i.e., demand) is lagging. This lag is attributed to the limited ability of businesses to extract information and insights as fast as the data arrives. Business need to consume data surpasses the need to consume everyday goods and services. This need continues to rise as technology advances. The only difference is there will not be any supply constraints when it comes to data and information (Lan & Unhelkar, 2005). When it comes to data, the traditional concepts of economics will eventually fail because supply exceeds demand. Despite this, data is becoming increasingly expensive (Brynjolfsson & McAfee, 2014). Hex-Elementization is envisaged to fill in this gap between the supply and demand of the data. Hex-Elementization lays the foundation for a framework which can help escalate the connection between different data sets on a real-time basis. This concept reduces the time it takes to relate data and extract information.

The problem of integration becomes evident, even if the data consumption is accelerated. This is because integration is not just confined to one kind of data. Instead, there is a need to integrate data in various forms, including music, e-books, and business strategy (Lan & Unhelkar, 2005). Integration is important in the digital age. Economist Shapiro states that in the next phase of the human-technical evolution, long-held economic principles and economic law will be challenged. Abundance will become the norm rather than scarcity (Guerriero, 2019). Every industry is experiencing this exponential increase of data in different formats. This exponential growth in size and type of data makes it hard to analyse it mainly because the new data does not fit in relational databases. Silo-based integration is the assimilation of data within the same silo – or type of medium (for example, integrating Microsoft Office files or merging two MP3 files). Businesses need a generic framework which can read from different formats of data and work around their volume.

One of the reasons for the Global Financial Crisis (2007-2008) was the mountain of complex financial instruments, so disparate in structure and format that no central system could simplify them to an understandable format. Lack of understanding, and the associated lack of transparency can create complex systems which are hard to manage without adequate controls. Hex-Elementization aims to create a platform which is data-format agnostic. Consider for example, an investment banking organisation that needs to understand the information flow and knowledge control (Lan & Unhelkar, 2005). There are risks associated with this investment banking business, including credit risk, compliance oversight, and liquidity risks. A platform like Hex-Elementization can help this investment bank correlate non-linear data trends in various formats, frequency and volume, and make it easy to understand and incorporate data in various forms. As James Rickards succinctly puts forth; “the next financial collapse will resemble nothing in history” (Rickards, 2014).

Processes capture data and information about the persons, places, things and events discovered while conducting business (Unhelkar, 2012). This needs to be current and of value. For example, one can tap into resources (mostly from the Internet) on what kind of food is healthy for humans relative to age, gender, and any medical condition. However, what is needed are more current properties (information) that make this data and information relevant to the person rather than another quality rating (of good or bad). “We do not need data on what is good, but information on what is better and relevant for us” (Nair & Lan, 2016). For that to happen, it is crucial to combine the existing information – which already exists in a different form – with new and fresh information providing the context. The array of permutations and combinations for customers using services takes the process of globalisation and data integration to an altogether different level (Lan & Unhelkar, 2005). To enable access to this rich dimension of data and relate it to the context of the user, Hex-Elementization proposes the use of Artificial Intelligence (AI), Machine Learning (ML) and Neural Networks (NN). The use of ML and NN can help find automated relationships and predict various dimensions of information from the same set of data for businesses to gain a 360-degree view of their operating landscape.

The volume, variety, and velocity of the data currently managed are unprecedented. By some estimates, the amount of data stored globally, since the year 2000, is ten times larger than any in human history (Brynjolfsson & McAfee, 2014). Every two minutes, the number of photographs uploaded to Instagram, Facebook, Flickr, and Google is more than the combined photographs taken in the entire 19th century (Benzell & Brynjolfsson, 2019). According to economist Brian Arthur, innovation is just combining earlier inventions to put them to practical use (Arthur, 2010).

The nature of information flow has changed. In the past, people received information using visual and tactile cues in brick and mortar shops while buying something. However, technology and the amount of information at hand has changed that behaviour pattern. Majority of consumers, especially Gen-Y type and beyond are tech savvy and subscribe to one or more social media platforms. These consumers start their research on products or services well before they visit a shop. As described by Unhelkar, in a report on the user experience, the understanding of human-business relationships is vital in a densely connected world where the consumers are increasingly Gen-Y type and beyond (Unhelkar, 2016). Unhelkar correctly identifies the fact that the process of decision-making starts way before consumers face the actual system or interface presented by the business (Unhelkar, 2013). In other words, investigating and assimilating data (e.g., prices and model) and information (e.g., quality, shop deals/specials) occurs even before a consumer visits the shops. How do these bits of pre-purchase information, which are not organised or systematically captured, reach the vendors (the businesses, the shop-owners)? How can the vendors gain insights into consumer thinking and shopping patterns while shopping decisions are made? This vendor need is best served through an on-going learning platform. Learning and managing that learning, is an ongoing activity that should find a place even in the routine, business-as-usual operations of a business (Unhelkar, 2012). Every additional or incremental piece of data and information increases knowledge about customers of any business. More time is increasingly spent finding ways to link and relate data than in making decisions. The proposed model of "Hex-Elementization" creates a framework to help real integration, linking data and insights, and saves valuable time during the pre-integration phase of decision-making.

The Journey from Data to Knowledge to Intelligence

In practice, the journey from data to intelligence in Hex-Elementization entails creating knowledge-sets through the interaction between various internal business processes such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), or Supply Chain Management (SCM), and when these processes interact with external data (e.g., Macro or Third-Party data).

Information within a business organisation can flow from three different sources –

1. An Internal Data Source (ERP, SCM. CRM).
2. Business Inherent Knowledge (Insight, Experience) and
3. External Data Sources (Macro Data, Competition, Third-party Data Sources).

Businesses have previously extracted the information from internal sources (i.e., business processes and departments) to help make tactical decisions. It is only a recent phenomenon that businesses (TDWI, 2018) are getting data and information through third-parties, especially in the form of Big Data. Even if the data always existed, the process of integrating internal data to external data is a resource-intensive process, requiring extra people, time and money.

As shown in Figure 4, the flow of internal data through various business processes provides intelligence which could be used to make tactical decisions. The figure shows how the data from internal sources and external sources along with the business rules are introduced to the Hex-E platform to help process it further for decision-making.

Big Data presents businesses with an enormous trove of data previously untapped. Hex-Elementization eases the extraction and integration of internal or external data, unstructured or structured data, and small or big data, to help organisations make timely and accurate business decisions.

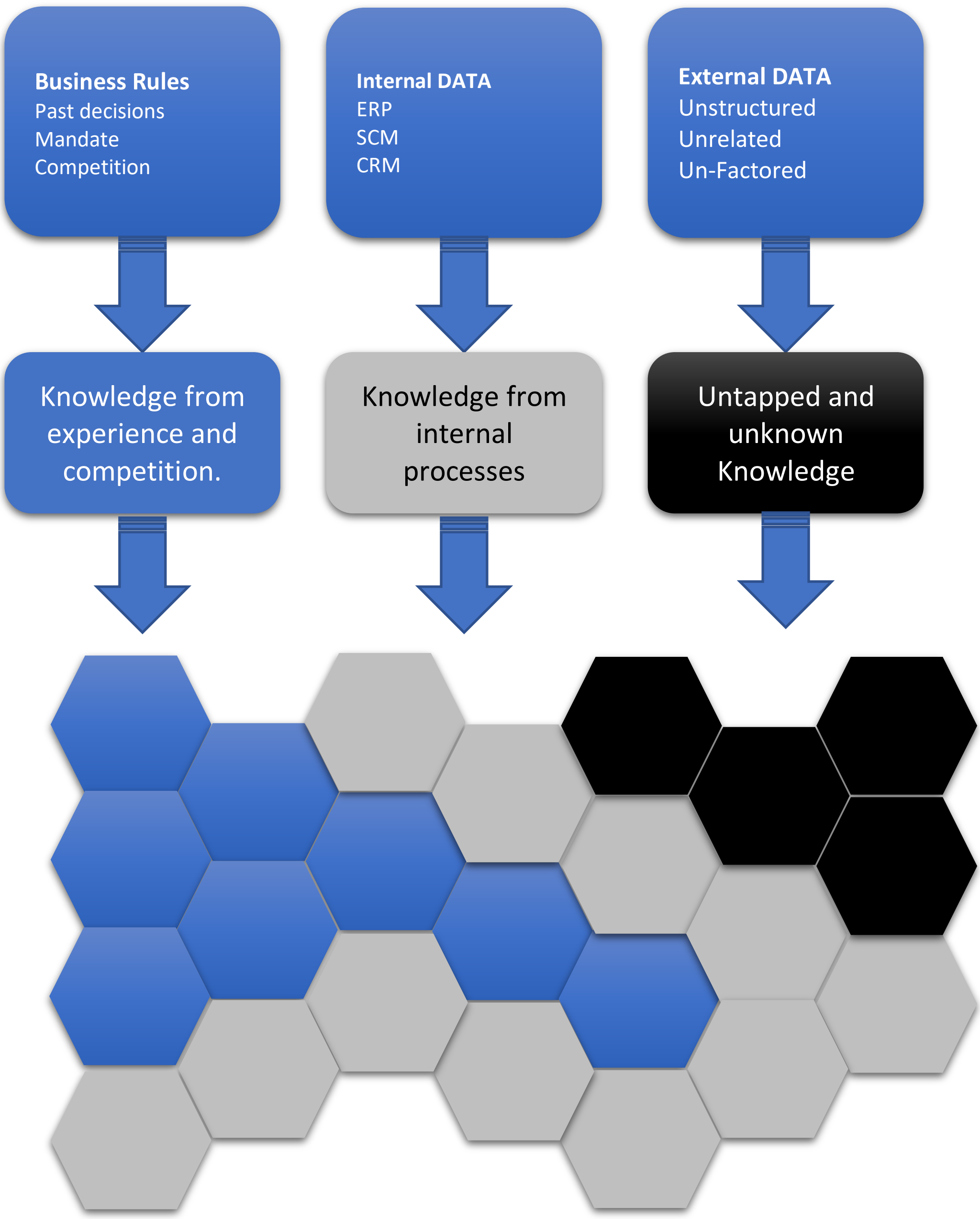

Figure 4: Knowledge flow in Hex-Elementization

The 7th Dimension and Big Data

Hex-Elementization also results in an extra property of data as data is connected. The number of hex-element properties exponentially increases upon integration. A hex-element with six properties, when combined with another hex-element with the same number of properties, does not end up having 12 properties, as would be expected by adding two, six-sided figures, but has 13 properties – as shown in Figure 5. So as more hex-elements are combined, the total properties grow exponentially. Similarly, a group of 7 hex-elements, as shown in Figure 6, has a combined property repository of 56 properties rather than the expected value of 42 – a 33% increase in the number of properties. In other words, more information is created when a piece of data integrates with another piece.

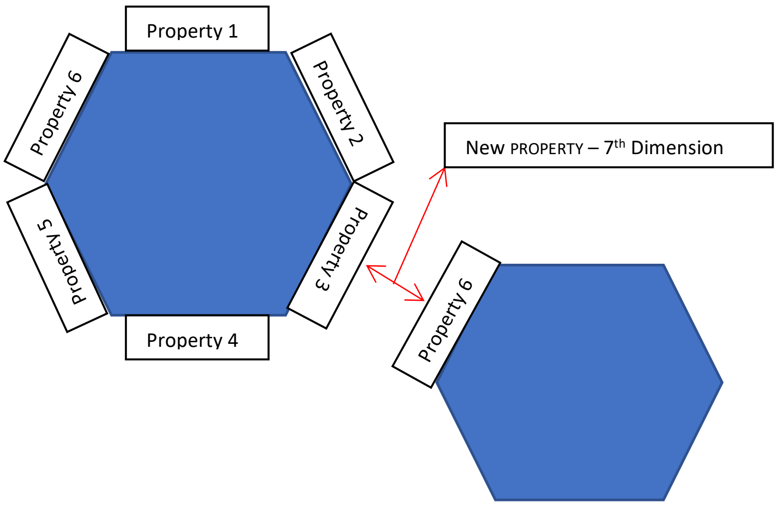

Figure 5: Correlation process between hex-elements and the 7th property out of commonality

The new dimension is the result of the fusion of one of the six hex-dimensions where the properties have something in common. The common thread of these two properties creates a new dimension or a new property. This is referred to as the 7th dimension. The exponential growth in the number of properties when new hex-elements encounter pre-built hex-structures is the main benefit of this structure. The structure also provides tight-knit data elements, as shown in Figure 6.

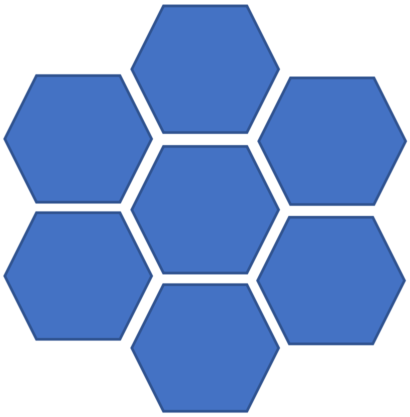

Figure 6: Hex-elements when connected forms a tight and efficient structure

Property (#)	Smartphone	Smartwatch	Heart Sensor
Property 1	GPS Co-ordinate	GPS Co-ordinate	Pulse
Property 2	Heart rate	Heart rate	Heart rate
Property 3	Location	Time-zone	Rhythm
Property 4	Time and Date	Routine	Variability
Property 5	Health	Glucose level	Respiration rate
Property 6	Tracking	Tracking	Temperature

Table 1: Example properties of hex-elements from three devices

Table 1 provides an example set of properties from a hex-element derived from three devices, namely a smartphone, smartwatch and a heart sensor. Each of these properties is one of the six sides of a hex-element. When trying to connect, the common properties across two hex-elements are matched to form a union. For example, the GPS coordinate from both Smartphone and Smartwatch (property 1) is used to create a connection. Property 2, heart rate, is a common property across all the hex-elements. Data from smartphone, smartwatch and heart rate monitor help create a push-based intelligence. For example, when the heart rate of a person rises dramatically, the heart sensor can send a signal to the smartwatch (through Property 2), which in turn can trigger the smartphone to call the emergency services based on its connection with the smartphone using property 2. Thus, hex-elements emerging from various devices coalesce and integrate using common properties between the elements.

Big Data brings in the ability to decode untapped information in a variety of medium of data, which is both a structured and unstructured format. Within these two formats of data, there are innovations made regularly. In the era of the ever-evolving world of Big Data, the greater the quantity of data (Mayer-Schönberger & Cukier, 2014), the bigger the opportunity to gain meaningful information. As data becomes the new factor of production apart from labour, land, and capital, its use in diverse faculties and situations are boundless (Brynjolfsson, 1994). Traditional database concepts and data storing concepts will become obsolete as the volume of the stored data increases.

The web is the default source of information for a lot of basic consumer needs. From school children trying to get help with science projects, to people who are at the highest echelon of their professional career, everyone uses the web in a small or big scale. Content on the web is 95% unstructured (Tanwar, Duggal, & Khatri, 2015). The flow of unstructured data will multiply as innovations and new devices, and the way they communicate, keeps changing. Keeping pace with the ever-growing types of files, music, photos and many other digital formats, is a question that must be answered. Moreover, how can current research make it easier for future generations to integrate and find insights among this ever-increasing disparity between data and its formats?

Hex-Elementization is a potential solution to help form the building blocks of a framework through which disparate, structured and unstructured data, in any volume, can be passed to help the data integrate, communicate, and extract meaningful information beyond the realms of the original use (Nair & Lan, 2016). For example, when data from a GPS, traffic information, and the meteorological department are easily combined, a car driver not only receives information on the optimal path to reach a destination but also whether any current weather pattern that might make it uncomfortable to drive. The system helps the driver find the best path in terms of proximity to the destination and with less traffic, but also chooses an economical path. It also formulates a route which is best for the car given the wear and tear forecasted by the weather data. Thus, the car's Electronic Control Unit (ECU) data integrates with the GPS coordinates, traffic information, transport department, and the meteorological department.

Big data is not only about the volume of data, but it is also about the complexity of data. In other words, the delay in extracting information from data is in part, because of significant volumes, and the physical or digital incompatibility which exists in current systems. Computers do not need to be bigger or smarter, with powerful chips and larger hard-drives; instead require a framework to better break and integrate the data.

Theoretical Contribution and Practical Contribution

This research contributes to both the academic and business worlds, creating a considerable overlap between the two. The theoretical contribution will help further the research by adding to the knowledge base this model creates at the end of the study. The practical objective deals with the applicability of this work in practice by businesses. The contribution of this research to both academic (theoretical) and practical world are listed below:

Theoretical Contribution

1. This research provides a business intelligence framework to enable real-time, resource-saving, accurate decision-making in workplaces. The underpinning theory will add to the decision-making theories explained in the Literature Review section.
2. The contextual underpinning of large undertakings in new-age technologies, like Big Data, is just the beginning. This research provides the basis of a contextual driven business intelligence framework in a push environment. Decision-making in the era of big-data is detailed in the Literature Review. The theories in this subject will be further augmented by the introduction of Hex-Elementization.
3. This research adds a new concept of integration to the area of data and information gathering, especially in the world of Artificial Intelligence, Machine Learning, Internet of Things and Big Data. The current theories around integration of data are not designed with AIML in mind. The integration need to be more automated in the era of big-data + AIML. The underpinning concept of Hex-E with respect to integration of data is further discussed from the lenses of theoretical literatures.
4. The research provides a theory at the atomic level of data to enable automated integration. Apart from the information theory which was conceived in late 19th century, no major attempt in literature has been made to granulise data, which is discussed in detail in the Literature Review. This research will help create a new branch in the body of knowledge which deals with breaking data further into smaller self-contained unit of measurement.

Practical Contribution

1. This research helps save resources, in terms of people, time and systems, by providing a generic framework in which a diverse set of data correlates and connects. This automated connection provides meaningful and timely insight for real-time and accurate decisions.
2. The framework provides the organic growth of intelligence encompassing both internal business data (i.e., from HR, Finance, Marketing, Supply Chain) and external data (i.e., Third-party data vendors, high-frequency consumer data).
3. Resource consumption in enterprise architecture planning is extensive and can be optimised with the flow of information and the ensuing business intelligence. The model proposed by this research provides a solution that works in various situations and architecture.
4. Big Data, and its variety, volume, and velocity, is expected to increase exponentially in various forms. The framework proposed by this research enables easier integration of any new type of data entering the organisation.
5. The proposed Hex-Elementization framework, powered by Artificial Intelligence, including Machine Learning and Neutral Networks, automates the integration of data from disparate sources. Hex-E has the potential to be implemented as-a-service or as-a-protocol within any data-driven organisation.
6. The framework proposed by Hex-E automates triggers and red flags to help detect regulatory violations in real-time. The framework supports the need for information to be collated and correlated from diverse market players to comply with regulatory requirements. Compliance departments and regulators alike, need to process this information on a real-time basis to maintain a credible oversight on business activities.

Chapter 2: LITERATURE REVIEW

Introduction to the Literature Review

The literature reviewed in this chapter covers data, information, AI, ML, IoT and analytics. Furthermore, decision-making across numerous disciplines is also discussed. The purpose of this focused literature review is to lay a solid foundation for the research in Hex-E and decision-making. This literature study includes the capability of humans in making decisions (Zeleny & Cochrane, 1973) and in challenging circumstances, as highlighted in the field of psychology and economics. Inherent bias such as framing, anchoring and representativeness (Kahneman & Tversky, 1996) can unduly influence decisions. This influence steers people away from making objective, data-based decisions and leads to more subjective decisions based on intuition (Morewedge & Kahneman, 2010). The field of economics, on the other hand, highlights the issues in making objective decisions when understanding the economic activity of humans are unsettled (Fine & Milonakis, 2009). Many economic studies have demonstrated how conventional wisdom is often wrong (Levitt & Dubner, 2005), especially when it comes to decision-making. Technology advances challenge the art of decision-making, and the theories of decision-making are faced with reality, which tends to be involved as humans find boundaries in cognition and judgement (Bandura, 1986). Decision-making at times entails balancing the costs and benefits of all necessary or available alternatives (Saaty, 2005). For typical decision-makers in an organisation, experience and knowledge are challenged by the changing environment (Prelec & Loewenstein, 1991). The influx of new tools, new data and the complexity of using these new tools to crunch unconventional datasets are all challenges in extracting value from data. The emerging data and technologies limit the rationality of decision-making. Hence, in this bounded rationality (Gigerenzer & Selten, 2002), decision-makers are faced with the prospect of using optimal resources, while making optimal choices to make optimal decisions. This process then becomes complex, especially when faced with a deluge of data being presented (Unhelkar, 2016).

The process to reach a decision should be an evolutionary one, rather than a revolutionary one (Lindblom, 1959). In the era of big data, the “incrementalist view” of decision-making (Carayannis & Stokes, 2000) is optimal for businesses that are incrementally facing new datasets. Learning from each dataset and the impact of decisions on organisational efficiency should be at the centre of business goals. To be tactical and nimble is tough in the age of Big Data, when billions if not trillions of “things” are connected to the Internet (Nokia, 2019), that are streaming data on every minuscule movement, made anywhere in the organisation. The “organisational, procedural view” of decision-making concentrates on the steps and rules of getting the insights which are necessary to make optimal decisions (March, 1994).

The entire suite of literature surrounding decision-making revolves around organisational structures, examined in detail in this section, organisational procedures and organisational hierarchy. The subsequent literature review tries to synthesise the existing body of knowledge (Mays, Roberts, & Popay, 2001) and literature in the field of business intelligence, decision-making in the context of organisational structures, processes, and implementations. The literature review is a complex process of reviewing existing literature in the context of the current state of business intelligence and decision-making and, at the same time, examines how this pre-established set of knowledge embedded in existing literature leads to problems or further inquiry (Ford & Richardson, 1994).

There are many ways to achieve a well-formed literature review. In order to address concerns, conflicts, issues, challenges or beliefs associated with the existing body of knowledge, a variety of literature and theories are critiqued (Machi & McEvoy, 2016) in developing this literature review.

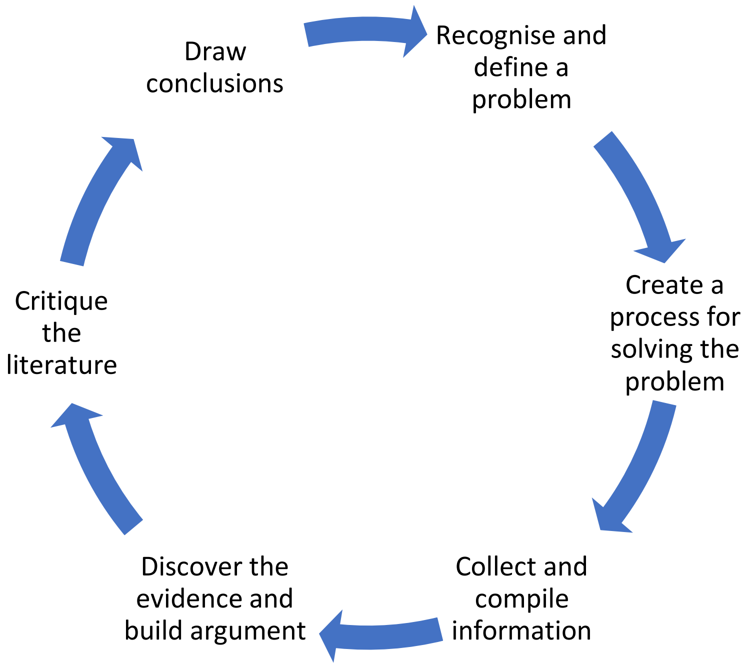

Figure 7: Critical thinking process of Literature Review *(Machi & McEvoy, 2016)*

As shown in Figure 7, the literature review conducted in this thesis follows a logical process of identifying and searching for existing literature on the topics of decision-making and business intelligence, followed by an assessment of the suitability of the literature to the identified research questions. The review process includes analysing existing literature using the inductive argument in order to help explain and discuss the research questions in light of the literature review (Webster & Watson, 2002). This process is also known as the argument of discovery (Mayer, 2004; Machi & McEvoy, 2016). Using constructive critique, each piece of literature is reviewed and analysed to help define the contribution or potential gaps of this research to the existing body of knowledge. The review process is also known as the argument of advocacy as it analyses and critiques the knowledge gained from the argument of the discovery (previous phase) to help support the questions raised in this thesis.

The concept of a framework designed to generate business intelligence organically is not discussed in contemporary literature or theories. The concept of such a framework transcends a single field, like information technology, and encompasses multiple fields including business processes, data analytics and enterprise architecture. In the absence of an optimal enterprise architecture, corporations are not able to create a holistic environment where data and intelligence are shared in a coordinated way. Various enterprise architecture theories, including that of Zachman and Spewak, are discussed in detail in this literature review.

The literature review also discusses the theories and literature surrounding the fourth factor of production, that of data, (land, labour and capital being the other three) and especially Big Data. Businesses are now capturing more data than ever. According to one research, 80% of this newly captured, high-frequency data lies dormant (Paul, Manguerra, & Slawecki, 2018). Businesses either do not have the resources or an environment to transform the data to insights. Hex-E proposes an automated method of crunching and analysing the data in order to generate business intelligence despite the different formats and quality of the data.

The literature review also highlights the concepts of Information Theory, which makes it possible to view data at its most atomic level. Drawing inference from the differences in

Quantum Theory versus General Theory of Relativity, atomic-level behaviour is often "spooky" (Einstein, 1971) and rarely makes sense in real-life situations; the same concept applies to bits and bytes. It is the transformation of data into information and, eventually intelligence, that plays a crucial and useful role in business decision-making and provides relevant insights. Integration, coalescence and correlation are concepts that are considered very early in an organisation's effort at setting up an information technology framework. Without such upfront planning, it is difficult for an operating business to bring about technical agility unless the business was to start from scratch. Hex-E proposes to address the issue of the integration of data and information and its evolution into intelligence in a unified manner. This integration and evolution help businesses achieve agility in decision-making based on a holistic business intelligence framework.

Decision-Making in the Big Data Era

Making accurate and timely decisions is a continuous challenge for most businesses. This challenge has become complicated with the advent of Big Data. While traditional decision-making also uses data, there is a significant reliance on intuition and instinct in those decisions. The high volume, velocity, and variety of Big Data make it almost impossible to undertake decisions based on intuition. Businesses can handle the challenge mentioned above by understanding and applying the power of data and the ensuing information within. Data-based decisions drive the daily activities of an organisation. Data is now considered the fourth factor of production after land, people and capital (Brynjolfsson & McAfee, 2014). Similar to how the deployment of capital, land or people affects overall productivity, and the bottom-line, the importance of the intangible asset of data on productivity and profitability is immutable.

Information technology has been highlighted in the past as an enabler of productivity in the workplace (Brynjolfsson, 1994). Although giant leaps in information technology have been made since the 90s, the use of resources including time, money and people has increased. Despite the giant strides in the fourth industrial revolution, it is not easy for organisations to extract meaning from data (Schwab, 2017). This is due to the changing nature of data in terms of volume and variety. Decision-making is further complicated as the internal data generated by CRM, ERP and SCM systems is difficult to be effectively consumed. This is because these systems have proprietary formats which take time to coalesce and produce meaningful information. Time lags with internal data make it difficult to incorporate that data into the decision-making process. Different departments in a typical organisation often use siloed applications based only on the specific department needs. The needs of these departments are also silo-based as the application is used for a narrowly defined purpose. Data sharing and integration is often an afterthought. When management sees the need to generate insights internally, it needs to create a new process to integrate data. For good decision-making, it is crucial that businesses integrate across their boundaries. The external and internal data needs to be properly integrated to be of value in decisions.

Integrating data in decision-making has inherent issues because data is now multi-structured, multi-formed and non-relational (Kościelniak & Puto, 2015). Businesses can tap into unique datasets (e.g., satellite imagery) to gauge the impact of extraneous factors on their operations. Such data, however, is not at rest. This means it cannot be approached in a store and analyse method. The analysis of data that is in motion and complex in a real-time manner is crucial.

Hex-Elementization proposes a model which tries to extract information from this multi-structural, disparate data set by finding context-based business intelligence. Hex-elements (as explained in Chapter 1) from disparate data sets organically try to find relationships based on the common properties. Hex-E explores relationships and intelligence in multiple data sets. Business needs help in simplifying this data-to-decision process, especially when data is growing exponentially. Doing so helps organisations harness the enormous potential of exploiting data to provide a competitive advantage (Unhelkar, 2017).

Data-driven Decision-Making

Corporations are currently facing a data-rich environment, in which data is streamed from every conceivable source, including internal departments, multiple devices and external parties. The world wide web (Internet) has brought the world closer. However, the collaborative intelligence (Luo, Lan, & Tang, 2012) it promised remains to be realised. The difficulty for decision-makers resides not just in differentiating between good and bad data, but also in extracting information from this data. It is only natural that technologies delivering faster, cheaper, and more accurate information create opportunities to re-invent the decision-making process.

As McElheren and Brynolfsson aptly said, the inertia in adopting new tools means that large corporations are not always nimble creatures (Mcelheran & Brynjolfsson, 2016). Instead of jumping on the latest technological bandwagon, managers need to re-think how to leverage decision-making practices using data objectively. For example, many businesses are keen to apply machine learning (ML) to their data without having a plan to integrate insights with existing business processes. Generating business intelligence, which can provide a competitive edge, becomes increasingly challenging when new age analytical tools and Big Data are used in a general way (or become General Purpose Technology (GPT)). Converting data into information and then into business intelligence provides the edge to businesses to remain competitive and make decisions at a faster rate. Intuition and the experience of the managers are no longer enough to make decisions in fast-changing, high stakes, business-critical, shifting conditions, and time-pressured circumstances (Klein, 2004).

In the new economy, company intuition and experience have taken a back-stage (Provost & Fawcett, 2013), in favour of evidence-based or data-driven decision-making. This is based on objective analysis and interpretation of data. Before the advent of Big Data, IoT, and Machine Learning concepts, organisations used data to make decisions primarily in the area of product pricing, marketing, and research and development (R&D) (Garvin, 2013). In a paper dedicated to the marketing field, Kumar, et al found that almost 30% of the marketing managers did not have enough customer data to make evidence-driven decisions and almost 40% of the organisations did not have the tool or infrastructure to transform the data into actionable business intelligence or insights (Kumar, et al., 2013). Some research has been published to highlight the lack of an eco-system which enables generating business intelligence or insights to help managers make data-driven decisions. Upper echelons of management need to recognise this gap proactively and dedicate resources to establish skill-sets, infrastructure and systems to help them make data-driven decisions (Aksoy, 2013). Another research in 2011 highlighted the fact that in the US along there will be a shortfall of 1.5 million managers who have enough skills and systems to enable data-driven decisions (Manyika, et al., 2011).

Existing works of literature on data-driven decision-making either concentrate on the benefits and alternatives of a data-driven approach to decision-making or propose a way to implement a data-driven approach in decision-making. Some pieces of literature touch on how to obtain, classify and disseminate evidence or insight to support decision-making (Baba & HakemZadeh, 2012). Continually evolving technology makes it increasingly difficult for organisations to narrow their focus on data to get insights before they even proceed to make decisions.

A data-driven process model (Jia, Hall, & Song, 2015) addresses this data-to-knowledge journey as a continuous process which encompasses collecting data, transforming it into knowledge, and monitoring decision implementation. Such a model also includes a feedback loop to continuously review the quality of the decisions made in light of the inputs from external stakeholders. The model provides a structured approach to utilising data in decision-making. The model, as shown in Figure 8, adapts similar, earlier work which posits creating a data-driven framework generic enough to embody creating and capturing knowledge, and application of the same, in various organisational pursuits such as R&D or new product development (Gold, Malhotra, & Segars, 2001). The model also builds upon the concept of organisational data absorption process through acquisition, assimilation, and transformation of data and exploitation of the same from a decision-making perspective (Malhotra, Gosain, & El Sawy, 2005).

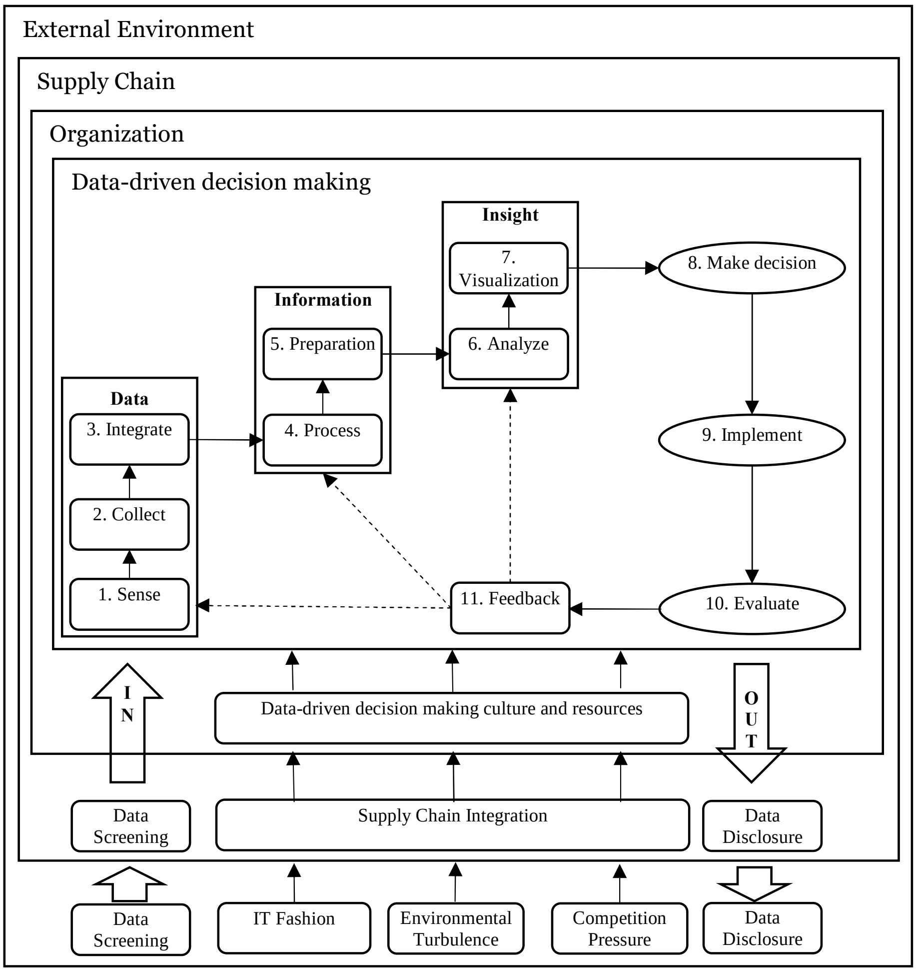

Figure 8: Data-Driven Decision-Making framework *(Jia, Song, Hall, 2015)*

These developments in the space of data-driven decision-making posit a case for an architecture in which disparate data can organically coalesce to provide insights. Every activity, ranging from air temperature to the number of people visiting a shop, human or machine-driven, is now captured in the form of structured or unstructured data. This deluge of data will continue as time progresses. Hex-E provides an avenue to merge and find meaning in this era of a higher volume of data, wide variety of data, and improved veracity of the data to provide seamless business insights which can be more context driven rather than static (as compared to dynamic) rules.

Six Degrees of Separation (Information Theory)

In the quest of finding theoretical precedence for the "hex" (6 sides) in the Hex-Elementization, the Six-Degrees of Separation theory stands more starkly than others (Lohr, 2012). The theory postulates that any person (which can be extended to anything) is connected to any other person within 6 or fewer connections. To illustrate the theory, it can be concluded that only six persons are standing between any given person and the President of the United States. The math supporting this theory is vigorous (see Figure 9). If a person knows six people with 50 unique acquaintances, then the total size of the network is 50^6 (15.6 billion), which is twice the total population on earth today.

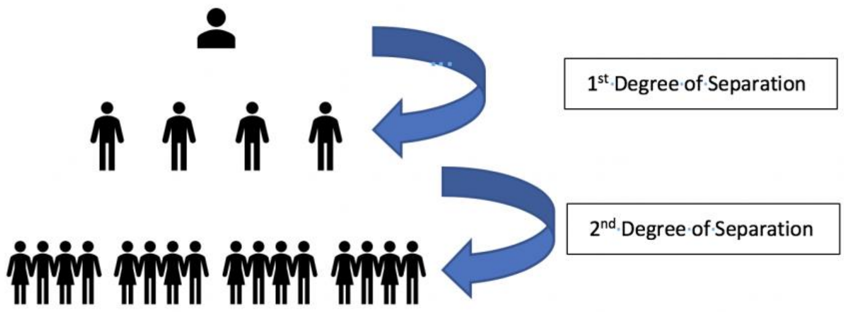

Figure 9: Illustration of 2 out of the 6 degrees of information-flow

The evolution of this theory can be traced back to 1929, in which a Hungarian writer, in a series of short stories, suggested that in the increasingly connected worlds, individuals should be connected with no more than five people (Milgram, 1967). Although mathematical attempts to prove this theory did not succeed in the 1950s by MIT and IBM, a social experiment by a stockbroker in the 1960s shed more light on this postulate. In the experiment known as "the small world problem" (Milgram, 1967), the researcher sent a letter to a group of people asking them if they knew the "target person" whom he intends to meet. The recipients were further instructed to forward the letter to another person they knew, asking if that person knew the “target person”. The experiment would stop once the “target person” was reached, who in turn would send a mail with the names of all the people who kept forwarding the letter before he was finally reached. Sociologist Milgram published the findings in “Psychology Today” and, in turn, popularised the term “six-degrees of separation” (Milgram, 1967). In 2008, Microsoft attempted to validate the findings through its Windows Messenger system, involving 180 million users. They found that on average there were 6.6 connections between each pair of users on its network (Smith, 2008). In 2011, researchers from Cornell and Studi di Milano conducted a study on connectivity among approximately 700 million Facebook users and found the average degree of separation was 3.74 peoples (Backstrom, Boldi, Rosa, Ugander, & Vigna, 2011). Facebook, as recent as 2016, conducted a similar exercise in understanding how connected its users are. Their researchers found that it has reduced the chain-length of its users to 3.5 hops (in other words, only 3.5 people between each pair of 1.5 billion Facebook users existed).

The experiment initiated by Milgram (Milgram, 1967), has given rise to systems which can predict how many days it takes a disease can spread across continents. It can also help in estimating the effect of a (big) bank default on the financial sector or predicting the effect of the sentiment of fake news on the general population in turning the tides in politics. Regarding business intelligence, this theory is particularly interesting when viewed with the lens of Hex-Elementization. As more devices get connected to the Internet (Informatics India, 2014), the more difficult it becomes for businesses to collate information across these devices in a meaningful manner. It is not just the collation of the data which is at stake here, but the more significant this data-set becomes, the more difficult and time-consuming it gets to extract information from it (McAfee & Brynjolfsson, 2012). It seems that many of these technologies were invented with HOW in mind rather than WHAT. IoT sensors can read and collect data from thousands of sensors connected to machinery, devices, vehicles, and collection-posts. Once the data is stored on a higher frequency level, the further use and interpretation of the data to gain insights is an exploratory field and currently being discussed and investigated in practice.

In the context of Hex-Elementization, the six-degrees of separation theory postulates that the maximum hops the environment needs to traverse across various streams of information are limited to six, and possibly less. The connection between different streams of information, which is facilitated by the six common attributes, exponentially grows as it scales up across various paths. Often in an organisation, information is collated, collected and stored in different departments. There are no over-arching processes, which try to correlate the data across different departments. This lack of process is often the case when the information is stored in a completely different format from one department to another. The resources needed to help with this integration to gain insights are enormous. Thus, information often remains dormant and eventually decays. Hex-Elementization proposes facilitating information-exchange among disparate datasets, differentiated not only in form but also in velocity and volume. This facilitation needs to be executed on the least number of hops or paths to generate insights that can be effectively used by organisations in making decisions.

Hex-Elementization proposes a unique way to generalise data at the source (department specific datasets) which then have attributes or properties which can be used for coalescence. The proposed environment can help gain business insights from disparate data sets within an organisational setup, which otherwise would have remained dormant.

Quantum Information Theory – Greg Jaeger

The most formal and prevalent unit of information – bits and bytes – was first theorised by Ralph Hartley in 1928, and this formed the basis of data chunks as currently known (Hartley, 1928). The size and the format of this basic unit of information were dependent on the hardware. In other words, the number of bits in a word is usually relative to the size of the registers in a CPU. With the advent of multi-core CPUs coupled with Moore's theory (Coffrin, 2019) of doubling the processing power every 18 months, there is an infinite set of bits and bytes which can now reside even on a home PC. Although this formed the very basis of information systems, there has not been much effort or investigation done in making datasets (as defined by bits/bytes) intelligible. There is a void between the basic computer information processing and information theory.

The birth of Information Theory in the 1900s, although neither formal not modern, first started highlighting the parameters affecting the transmission and processing of information (Jaeger, 2009). Earlier research and investigation focused on the physics of hardware and its limitations. A formal and detailed study on information theory was theorised in a paper titled "Certain Factors Affecting Telegraph Speed" (Nyquist, 1924), in which the physical limitation of channels concerning data transmission was quantified. The modern theory of information was proposed in a publication titled "A Mathematical Theory of Communication" (Shannon, 1949), although most of it highlighted the inherent drawback of communication channels. Shannon aimed to maximise the yield of data in the best possible/optimal way. The information theory hence revolved around the channel through which data is transmitted rather than the data itself (Jaeger, 2009). Encoding the data beforehand with properties of the data has the potential to add value to data. Hex-Elementization of data can help provide these properties for each data point, which can help to form the basis of information processing.

The use of parallel data processing for Big Data application deals with the optimisation (Yin & Wang, 2015) and theories of data integration to find the best solution and meaningful patterns in the “big” data set. The challenges include characteristics of disparity, volume, variety, veracity to the velocity of the underlying data. Most of the literature indicates a level of effort needed to handle each of these characteristics of Big Data to weave a common thread across untapped information. However, the literature review highlights a gap in the integration and coalescence of information. This gap indicates the need for an architecture in which data is set so that meaningful information can be filtered to enable integration. Hex-Elementization aims to (a) minimise the steps needed for data to form information (b) enable a proxy artificial intelligence into each data item/element which transforms it into an intelligible piece of information in a bigger informational context, and (c) help in non-linear integration of data by helping information theory algorithms.

The Zachman ISA Framework (1987 and 1992)

Zachman introduced a schema, structure, and ontology, which starts with structuring enterprise solutions and answers to basic primitive interrogatives (Zachman, 1992). Zachman was instrumental in developing one of the first architecture which documented Information Systems Architecture (ISA) in the context of several hierarchical perspectives' characteristics (Bernard, 2012). He provides a structure which defines the information system and information flow through the way of definitions, as shown in Figure 10.

	Why	How	What	Who	Where	When
Contextual	Goal List	Process List	Material List	Organizational Unit & Role List	Geographical Locations List	Event List
Conceptual	Goal Relationship	Process Model	Entity Relationship Model	Organizational Unit & Role Rel. Model	Locations Model	Event Model
Logical	Rules Diagram	Process Diagram	Data Model Diagram	Role relationship Diagram	Locations Diagram	Event Diagram
Physical	Rules Specification	Process Function Specification	Data Entity Specification	Role Specification	Location Specification	Event Specification
Detailed	Rules Details	Process Details	Data Details	Role Details	Location details	Event Details

Figure 10: Zachman's Framework of Enterprise Architecture *(Zachman, 1992)*

Enterprise Architecture (EA), at its core, helps replace the age-old systems-level thinking with a view that IT is not a support system but should be woven through every department in an organisation. EA facilitates optimal resource utilisation when it comes to information systems. However, substantial time is required to map out the distinct informational needs of various departments. It takes significant time for data architects and data analysts to map out the needs of one department with another and the integration between the two. It takes time to find the shared properties between systems and departments which can correlate and combine to become informative. The proposed Hex-Elementization framework solution supports integration by removing the need for intermediaries to map out and find common properties or attributes among different data sets. Hex-E exhibits a protocol of information interchange from a top-down level. If each distinct and atomic piece of information has enough uniquely identified properties, then the properties seek out the most optimal way to link to another piece of information. In this case, it is not necessary to establish definitions, architectural boundaries or define links between departments. This is so because defining business rules takes time and effort. Differences in systems and frequent upgrade to applications with departments may require rewriting integration rules. The architecture can become complex as the disparity in applications and systems grows.

"The EA improves decision-making by providing comprehensive views of current capabilities and resources, as well as a set of plausible future operating scenarios that reveal changes in processes and resources" (Bernard, 2012). The proposed Hex-E solution improves decision-making by being impervious to the differences within the organisation in any form (departments, structure, architecture, type of resources or systems). By breaking down the data into its simplest form with six properties, Hex-E can help find commonalities across various departments despite the differences in systems and applications. The automated connection seeking hex-elements tries to incorporate data across departments by finding data points with similar characteristics. Many times, decision makers do not get a holistic view of the decisions they make due to the lack of integration of data across departments. Hex-E can help relate seemingly unrelated variables across different departments to provide a holistic view to the decision makers.

The Spewak EA Planning Method

Steven Spewak took the basics of the Zachman's framework and developed it further into the concept of "Enterprise Architecture Planning (EAP)" (Zachman, Spewak, & Hill, 1992). The basis of the framework is that IT and information flow support business functions. Spewak was one of the first proponents of the term enterprise when it comes to information system flows extending beyond individual systems in a hierarchical structure, as shown in Figure 11.

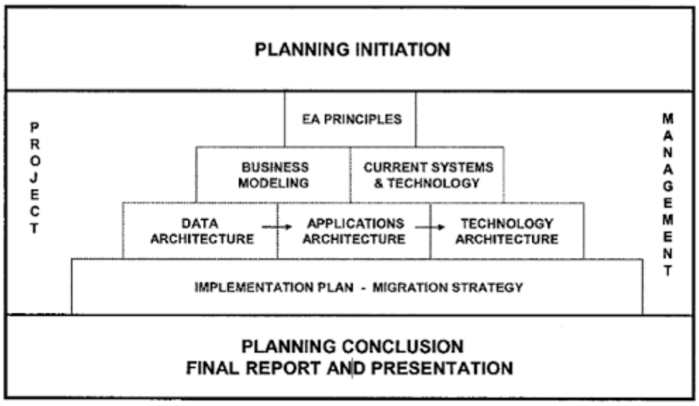

Figure 11: Spewak's Architecture of Enterprise Architecture *(Zachman, Spewak, & Hill, 1992)*

Spewak's EA model delves deeper into the integration and planning phase of enterprise architecture from a system point of view. Although it provides a robust framework of enterprise resource planning, it does not help in demonstrating how the information flows through various divisions to help make better decisions. In the fast-paced business world (Brynjolfsson & McAfee, 2014), real-time decision-making is vital. To enable real-time systems, information-flow needs to be seamless throughout the enterprise.

Hex-Elementization is theorised to create a seamless flow of information dissemination, requiring less time on structuring data warehouses and flow, and more time on providing context. Hex-E is also theorised to work underneath the departmental siloes in businesses. Enterprise Architecture provides a structure which is optimal from a management point of view. However, EA fails to incorporate the informational flow required for business to gain intelligence across the company. For rigid enterprise structures, Hex-E is theorised to provide a solution to integrate with existing legacy systems and is examined later in this study. This integration with legacy systems is essential to extract valuable information beneficial to rest of the company. For new businesses or businesses with malleable enterprise architecture, Hex-E provides a protocol-based solution which can be implemented easily. In such architectures Hex-E provides an easy way to break data and information into granular forms, enabling the integration for real-time business decisions. Hex-E aims to provide an answer to break the barriers set by EA strategies.

EA3 Cube Framework - Scott Bernard

Another Enterprise Architecture model which modularises the way enterprises manage data and information flow is Scott Bernard's EA3 Cube Framework (Bernard, 2012). Bernard treated Enterprise Architecture (EA) as holistic management, planning and documentation activity in his EA3 Cube Framework as described in Figure 12. Bernard's approach distinctly addresses the different lines of business displayed as five individual sub-architecture levels and three common thread areas. It is a component-based architecture in which the building blocks of each sub-architecture can be configured as "plug-and-play". The components are interoperable and integrated due to the standard thread that promotes solutions outside the traditional domain (Bernard, 2012).

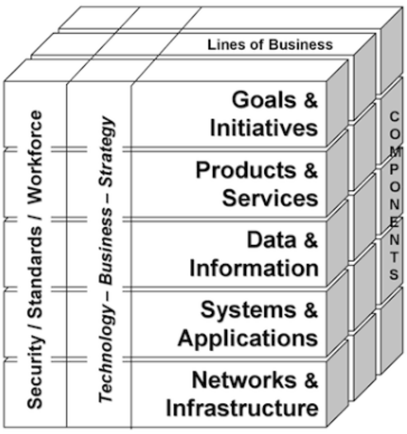

Figure 12: EA3 Model *(Bernard, 2012)*

The Enterprise Architecture (EA) framework Bernard proposed is based on six fundamental principles (Bernard, 2012):

1. In an EA governance process, each EA management function links to other enterprise-level management functions.

2. EA process is a consistently repeatable methodology for a given management function.
3. EA is a framework representing the core elements and layers of the management function.
4. EA process is a union of a set of architectural descriptions.
5. EA process defines a set of tools which act as a repository to support architectural descriptions.
6. EA encompasses a group of associated best practices acting as a guideline of management functions.

Bernard also emphasised the need to understand the organisational context in order to help design an appropriate EA management function. He detailed examples of the impact of the organisational context on the design of an EA management function (Bernard, 2012). Through a novel approach named “segment approach”, Bernard details how EA basics can be established and contextually driven. The contextual underpinning of the EA architecture relates to one of the research goals of Hex-E, which is to enable a push-based intelligence system. The loss of context when designing enterprise architecture is not experienced in organisation functions. This loss of context is experienced by individual functions when trying to share data and provide holistic insights to the management. This is where Hex-E can supplement the architectural designs in order to enable smooth inter-department communication and sharing of intelligence.

Traditional business functions are overwhelmed with information in the era of Big Data. IT departments, which traditionally service other business functions (evident from the six principles discussed above), is permeating to other departments. In other words, each department is employing programmers, analysts and analytical tools to deal with this wave of information. Bernard's EA principles are pivotal to the effective management of enterprise functions, but not ideal for information exchange among these functions. Tech personnel plan the network structure and wiring when setting up an office. Similarly, Hex-E is theorised to be woven across business functions which can help in connecting, integrating and coalescing data across various functions. These connections help management extract intelligence to make effective business decisions.

Consistent with the previous Enterprise Architecture theories, the model provides a top-down view of the processes and data flow within an enterprise. The model maps out the possibilities to address data, applications, and networks to assist in strategic planning, streamlining supply chains, and delivering value to clients. Similar to the framework embodied in Zachman's work, this model supports the information flow structure across various divisions within an enterprise. It portrays the benefit of having distinct divisions which can be uniformly and efficiently serviced by limited resources. Integration and enterprise structure do not individually define how complex divisional information can seamlessly integrate, correlate and assist in effective decision-making. In the era of Big Data and IoT, divisions and departments (silos) are losing relevance, especially given the required synergies between disparate enterprise systems. To effectively and efficiently extract information from data, data at the atomic level needs to be reconsidered.

Hex-E aims to form the basis of this architecture which provides a common platform through which disparate data with unique and sought-after attributes can link to each other automatically and provide contextual-based information to support decision-making, irrespective of the enterprise architecture.

Business Intelligence

Businesses not only face a massive amount of data but also face data in unconventional varieties. Data from sources like mobile phones, social media, satellites, computer vision technology, augmented reality applications, and IoT (Internet of Things) appliances have exploded in the last 5 years. Making sense of unstructured data requires a shift in both business and technology thought-processes. Businesses accustomed to making decisions by using conventional data (structured data-sets) now need to use customer's faces, likes on posts, star-ratings on product and services, pictures, audio, video, blogs, vlogs (video logs) and tweets, among other unpredictable types of data, in making decisions (Kaisler, Armour, & Espinosa, 2013).

Business intelligence (BI) is a broad concept which encompasses various processes including analysis, collection, presentation, and dissemination of business information (Lönnqvist & Pirttimäki, 2006). Businesses have tried and tested systems for BI including MIS (Management Information Systems), DSS (Decision Support Systems), ERP (Enterprise Resource Planning), BIS (Business Information Systems), EIS (Executive Information System), OLAP (Online Analytical Processing) and BA (Business Analytics). These systems are further built upon the core repository of data from customers (CRM), sales data (Finance), suppliers and distribution (Supply Chain) and marketing (e.g., Social Media Marketing - SMM). The common foundations between these systems and application layers rests on data in a structured format (Chen, Chiang, & Storey, 2012). Structured data has identifiable dimensions which include numeric and textual data. Data is maintained in data warehouses, databases, record management in which it is processed and analysed using analytical tools. Analysts identify patterns in these well-formed data sets to help interpret business trends and provide management with information to make better decisions. The benefit of structured data is that the process of comparison and predictions, the two most essential elements of adding value to data, is easier (Bakir, et al., 2007).

Businesses spend time-intensive resources, building ontologies for information extracted from structured data into basic semantics and common representations (Nédellec & Nazarenko, 2006). These ontologies act as a common language between technology enabling business intelligence and the people who make decisions. Ontologies applied to structured data analysis create a series of processes that can be easily consumed to help highlight business insights. However, this trend is changing.

According to one research study, 80% of the data generated or collected within a firm is unstructured in format (Jun, Sung, & Siau, 2001). Moreover, this collection of unstructured data is doubling every three months. This growing amount of data requires reconsideration of the entire value chain of how businesses deploy resources in setting up a BI framework which includes and recognises the different types of data being created. Structured and unstructured data needs to coexist in a holistic and integrated environment. In addition, the environment should support analysis, crunching and consumption to generate new business insights. Otherwise, the different types of data can become overwhelming in the decision-making process.

Research by Blumberg and Atre, summarised that 60% of Chief Technology Officers (CTO) and Chief Information Officers (CIO) surveyed considered unstructured and semi-structured data critical to improve operational processes and create new business opportunities (Blumberg &

Atre, 2003). Merrill Lynch estimates that more than 85% of data accumulated by businesses are in the form of unstructured and semi-structured format (Shilakes & Tylman, 2008). Gartner group estimates that roughly 40% of the lower management tech workforce is used to managing and maintaining semi-structured and unstructured data (Gartner, 2018).

Unstructured Data

Unstructured data is neither in a conventional structured format nor can it be easily analysed using conventional analytical tools (Feldman & Sanger, 2007). It includes, but is not limited to, digital assets such as videos, movies, chats, blogs, graphical images, charts, thumbs-up and star-ratings on the web. A video is encoded at the granular level in bits and bytes, but businesses need to ask some additional questions. Can the video be quickly consumed by businesses to make real-time decisions? Can the data from the video be merged with other structured datasets to know more about the sales pattern of customers in a particular geographical location? Businesses like Amazon aspire to make this a reality (Moniruzzaman & Hossain, 2013). However, there is no existing framework to enable the use of unstructured data for the broader corporate population outside of Amazon. The benefits of such a system are explained here.

1. **Augmented Sales.** Businesses are using a mix of advanced technologies to increase sales. Amazon is using customer iris scans, web surfing patterns, and other personal information to capture consumption trends (Brynjolfsson & Hitt, 2000). Personal information, including facial recognition and iris scans are unstructured datasets, which can be used to help augment sales or services. Unstructured datasets like facial expressions showing happiness, sadness, or a frown can be captured and (machine) learned by businesses to understand the effectiveness of a sales pitch. Satellite images

of geographies showing population dispersion help businesses fine-tune their expansion plans to set up stores or sell specific goods and services to cater to the marginal customer (Florida, Mellander, & Gulden, 2012).

2. **Targeted marketing.** Targeted marketing is a branch of marketing, which identifies a specific group of customers based on certain characteristics and then, customises marketing to individual customer needs and interests (Pitta, 1998). Structured public data can be limited when trying to target customers based on personal details alone. This limitation is circumvented using unstructured data because unstructured data can provide more intrinsic data on consumers than structured data. A few businesses have already employed this marketing strategy by capturing the click-behaviour of customers (Browne, Durrett, & Wetherbe, 2004). Facebook pitches targeted product marketing based on what the user has browsed in recent history. Businesses can pitch cars based on a customer driving style. Eyeglasses can be marketed based on the retinal scans received from the optometrists. Businesses can pitch for a particular type of television based on customer trips to a local movie rental shop. Grocery stores can pass the data on which isle customers spent the most time in (through video) without making a purchase, suggesting they are planning for some event in the future (e.g., child's birth). There are innumerable ways businesses can use unstructured datasets by triangulating it with structured information they already have.

3. **Improving process efficiency**. Industrial businesses using unstructured data have the potential to revolutionise the process efficiency drive. Consider the example of the farmer who used thousands of photos of cucumbers at his sorting facility to train a

sorting machine powered by machine learning algorithms to improve accuracy (in sorting) and speed by a combined 80% (McAfee & Brynjolfsson, 2017). Many businesses are investigating innovative ways of using unstructured data to help improve processes. Incorporating unstructured data can help improve the serviceability of the organisation in terms of products or services. Process improvement using unstructured data can also help increase the accuracy of the process when compared to the same process handled by humans (Katal, Wazid, & Goudar, 2013).

4. **Cost minimisation benefits.** The process improvement explained earlier can not only help in achieving efficiency but can eventually lead to cost reduction. One of the many ways to improve the bottom line (earnings) of a company is by either increasing the sales or reducing the cost per unit of sales (Husted & de Jesus Salazar, 2006). The benefits of an efficient BI platform can sometimes seem intangible. However, there are many identifiable benefits (e.g., inventory management) to the entire organisation in terms of hard dollars saved (Bucher, Gericke, & Sigg, 2009). Incorporating unstructured data into the decision-making process is turning an intangible digital asset into tangible possibilities. Businesses can now try and tap into the unexplored territory of using visual and aural streams of intelligence which were untapped for a significant part of the previous industrial revolution (Schwab, 2017). The benefit of this new kind of intelligence is far-reaching. However, it can also help in minimising the overall cost of making decisions in the new-age world.

5. **Improved supply-chain relationship.** Unstructured data can help improve the company's supply-chain relationship, including that with the suppliers and customers. Many businesses are now tracking shipments via satellite images to get an indication of the earliest possible delivery times for their products (Kitchin & McArdle, 2016). Businesses are tracking seasonal patterns (unstructured Big Data) to time agricultural supplies. The star-ratings for app purchases on the Apple and Android platform are now being incorporated into the initial stages of software development by technology businesses (Kozinets, Hemetsberger, & Schau, 2008). There are increasing numbers of brick-and-mortar consumer businesses who track the foot traffic into their department stores using ML driven CCTV cameras (Bodor, Jackson, & Papanikolopoulos, 2003). These innovative ways of capturing data helps merchants showcase certain products at certain times of the week to attract more customers. Thus, unstructured data sets provide an in-depth insight into supplier and customer behaviours, providing businesses with valuable insights to help sustain and grow their supply-chain relationship.

6. **Accurate long-term decisions.** Accountants believe that in the long-term, everything averages out. However, this theory often fails when applied to the decisions made by businesses on a longer timeline. Businesses often have a vision for the long-term and start deploying capital to accomplish the vision year over year. They often do not see results until the project, or their vision is close to completion. Many times businesses pull out of the project either mid-way or close to the completion period due to unforeseen changes in the competitive landscape or technological changes rendering their plan less commercially viable than conceptualised (Marchewka, 2014).

Unstructured data can help make long-term planning more viable by simulating results when combined with powerful algorithms and high-performance computing (HPC). NASA, for example, has been using unstructured data sets, including videos, pictures, and atmospheric pressure to plan for the upgrade of the Hubble Telescope (Abel, Bryan, & Norman, 2002). NASA reduced the deployment time for the upgrade phase by simulating the deployment and operation of the James Webb Telescope, which is going to replace Hubble. Not only did this help anticipate any improbable scenarios in the future, but it also helped fast-track the project saving millions in long-term actual and perceived costs.

7. **Real-time and tactical short-term decisions.** The benefits of unstructured data are in managing the long-term goals of a company and in making short-term decisions. These can be either real-time decisions to react to changes or tactical to adjust to short-term unexpected incidents. The duration of these short-term decisions can be short-lived; however, certain short-term decisions can form the backbone of how the firm adjusts long-term decisions. For instance, businesses involved in packaging agriculture-based products could monitor the satellite images of weather patterns evolving in a country from which it sources its agricultural produce; this enables the company to place orders with an alternative provider not affected from specific weather issues. Similarly, highway traffic operators could watch videos of stalled cars or vehicles involved in accidents on certain parts of the highway and start diverting traffic to an alternative route to avoid congestion. Unstructured data can, at times, be more straightforward to digest from a cognitive perspective, helping to make decisions at a faster rate.

8. **Enterprise-wide employee satisfaction**. Many businesses conduct firm-wide surveys to identify issues employees are facing, in order to keep or increase employee satisfaction levels. These surveys are branded as employee satisfaction surveys and list many close-ended questions on how satisfied employees are on various issues, ranging from their role, manager behaviour, workplace environment, and the direction of the business. The survey is then concluded by requesting open-ended feedback from its employees. This entire process of conducting employee satisfaction is not only inefficient and laborious, but it might take a long time before management can summarise the result of such an exercise. The more time it takes to analyse and summarise issues highlighted by the employees, the more time it takes to implement solutions addressing those issues. Unstructured data can help businesses capture employee satisfaction on a more granular level and on a real-time basis as employees interact with the organisational elements. Unstructured data not only can speed up the process of gathering employee feedback but can help businesses attain a deeper level of understanding about their employees.

9. **New product/service development.** Unstructured data can play an essential role in developing new products or services. When coupled with other emerging technologies like Machine Learning and Deep Learning, unstructured data provides enormous potential to speed up the rate at which new products and services can be introduced or explored in a market. Consider the Hack Rod, the brainchild of the Primordial Research Project. Hack Rod is a car entirely built with a combination of unstructured and lots of structured (Big Data) data set. The car was designed entirely by an Artificial Intelligence platform which combined machine learning techniques

with advanced computer vision and photogrammetry to create 3D models of car designs. These designs were further tested using simulations to fine-tune various aspects before the final product (car) was ready to be manufactured. This example of the combined power of machine learning and unstructured data provides a glimpse of the products and services that can be designed from scratch. Advance machine learning algorithms with unstructured data have also improved the accuracy of detecting brain cells (vis-a-vis health cells), which has the potential to revolutionise health care service levels. There are innumerable use cases of how unstructured data can be used to help to create new product and services.

Decision-making in the age of Artificial Intelligence

Artificial intelligence has the potential to change how businesses operate and make decisions. "It is change, continuing change, inevitable change, that is the dominant factor in society today. No sensible decision can be made any longer without taking into account not only the world as it is but the world as it will be." (Asimov, Silverberg, & Timmerman, 1978). Human beings are continuously exploring the building blocks of existence. Through continuous scientific discovery, humans have continually increased understanding of the physics of origins and existence. In an attempt to transcend capabilities to machines, specific fields have gained new heights. Certain human skills have been surpassed by machines, including mathematical calculations, board games such as chess, backgammon, Go, competitions like Jeopardy and entrance exams to universities (Brynjolfsson & McAfee, 2014). These developments have been very prevalent in media and research, and have helped to determine if AI can or will replace humans. Without even realizing it, people have benefitted from AI in areas such as avionics, search engines, recommender systems, fraud detection systems, medical diagnoses, and image processing. While not visible to most people, these AI advances have enriched lives (Elliott, 2019).

Not only has the quantity of data exploded in recent years, but the modes of delivery and access have changed dramatically. For example, today, an African citizen with a smartphone has access to more information than was available to all American administrations combined (including the FBI and the Pentagon) 40 years ago (Banerjee & Levy, 2008). However, having access to data does not equate to meaningful insights. The need for easily accessible insights, not data, provides the impetus for knowledge-engineering. Artificial intelligence (AI) aims to slot-in the journey by automating the insight-generation from data. However, AI is the natural progression of how software has been helping humans in decision-making. Figure 13 displays the evolution of software in terms of decision-making and its progress through time, with the latest incarnation being AI (Fulcher, 2008).

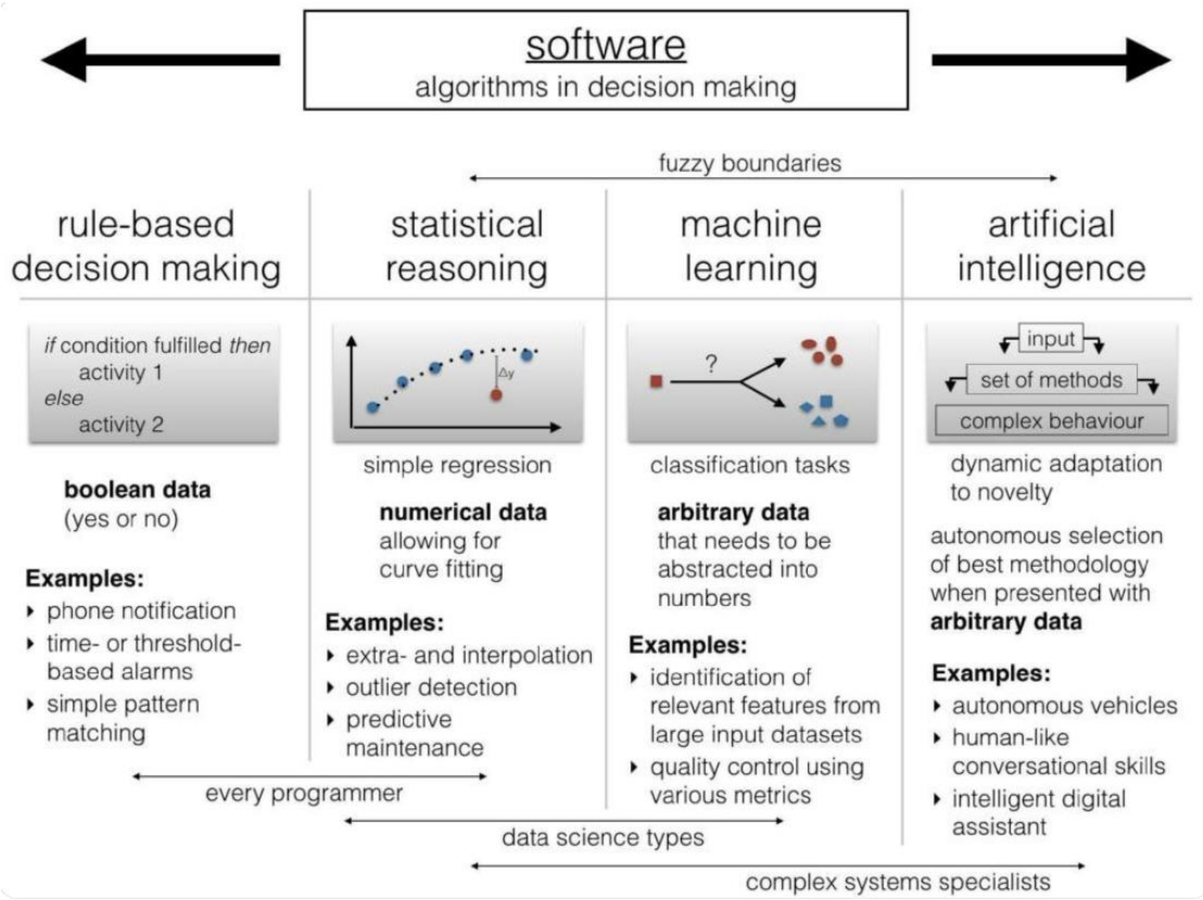

Figure 13: Evolution of software and its applications (Fulcher, 2008)

The issue surrounding AI and its ability is often surrounded by fear (Minsky, 2007). The fear of the unknown is the main criticism of AI. One example of decision-making involving AI is the case of automating the job of a cheese controller in a dairy company. The company wanted to automate the job of the cheese controller, which is to taste the maturity of the cheese by inserting his/her finger into the centre of the cheese before deciding whether it is ready to be sold. The company developed a solution powered by AI and sophisticated sensors to feel the touch of the cheese and automate the decision-making process. The AI system failed to judge the maturity of the cheese in multiple tests, despite being fitted with the most sophisticated sensors. Later it was discovered that it was not touching the cheese which helped the cheese controller judge the maturity of the cheese. It was the cheese odour emitting from the cheese that helped the controller judge the maturity of the cheese (Brynjolfsson & McAfee, 2014). Hence it appears that even people are unable to accurately express how they fully make decisions. As human decisions are fallible due to standard errors, AI (or Machine Learned models) can make errors too, if driven by incorrect data. However, it is essential to distinguish between the two types of errors which can occur. Consider the simple personal decision on whether to eat a mushroom based on whether it is poisonous or not. Table 2 shows the decision matrix of such a choice.

	Edible	**Not-edible**
Poisonous	No	Yes
Non-Poisonous	Yes	No

Table 2: Decision matrix - an example

In the event the model predicts the target mushroom as being edible when it is poisonous, it generates a false positive error. The model could also create a false negative error in which the model predicts the target mushroom being poisonous when it is, in fact, safe to eat. The cost of these errors is not symmetrical - poisoning might kill someone, while not eating a safe mushroom creates feelings of hunger. Hence the fallibility of artificial intelligence is similar to the fallibility of humans. Similarly, as humans learn more by practising, a system based on artificial intelligence can get better with more training (Morton & Aleksander, 1990).

Another example is a Formula One (F1) car. Every part of a F1 car, including tyres, is embedded with microprocessors and sensors. It is estimated that every lap completed by a F1 car creates a gigabyte of data. One would imagine that the F1 teams would have standardised the broadcast of data from these sensors to enable analysis at a faster rate to create real-time decisions. One could extrapolate this example and apply it to an environment which consists of non-standard sensors, devices, and gadgets, which broadcasts data in a non-standard format every microsecond – which needs analysis to make decisions. This technological adaptation is similar to the problem of the mid-1990s users of the Internet who were forced to convert their texts, pictures, or content to hypertext format to enable others to read or digest it in a more natural way. It was easier to ask users of the Internet to follow a standard, however, with current technology new formats are invented frequently indicating the need for an environment which is format and standard independent.

AI has great potential to supplement humans when it comes to decision-making (Makridakis, 2017). AI not only can crunch an enormous amount of data but also draws patterns and understands complex relationships better than traditional business intelligence applications. The ability to crunch data is not solely the domain of AI as high-performance computing has gained traction in many fields, including weather prediction, stock market operations, and supply chain management. High-performance computing assists in augmenting the development of AI. Another element which helps augment AI is the research to understand how humans perform the most complex actions. These actions not only include physical actions like walking and running but includes the mental action of thinking (Bhattacharya, 2018).

Researchers have already quantified some of the standard human skills like vision, language/speech and other cognitive abilities. Advanced algorithms are available in the market place which can emulate human skills, including computer vision, natural language processing, cognitive analytics, and heuristics-based analytics (McAfee & Brynjolfsson, 2017). The success in building these algorithms was achieved with progress sub-technologies which power Al, including Machine Learning, Neural Networks, Deep Learning, Virtual and Augmented Reality. However, before AI can be used in general to represent knowledge in uncertain domains and new domains, it needs to have specific, critical common characteristics (Norvig & Russell, 2016). These essential elements or characteristics of AI dictate the level of involvement in a typical organisation. These common elements set the precedence of how AI is utilised and by which functions of the firm, including decision-making, can be handled by AI (Mladenic, Lavrač, Bohanec, & Moyle, 2003). Steve Moyle et al., list the elements needed in an AI which are discussed below.

1. **Robustness.** AI needs to know when it is at the limits of its capability. Humans can articulate that they are not very good at some task and the associated risks. This acceptance adds robustness. Brittle systems tend to break in novel situations. Many professional codes of practice prohibit members to go beyond their level of expertise. Robustness is essential in decision-making. Without a robust set of rules, processes, or checks, the decision-making process might be either incomplete or can be fallible. Robustness needs to be at the centre of Hex-E, especially when decisions start incorporating alternative (non-traditional, non-fundamental) unstructured data sets. Hex-E is envisaged to stand apart from BI systems which are rule-based. However, automated information seeking and connecting behaviour of Hex-E can be surrounded by overarching big picture rules provided by the businesses. Although

Hex-E is not based on rules, however, it is designed to have boundaries set by the primary goals of the businesses.

2. **Inspect and Correct.** All AI models need to be "debuggable". Many are of the opinion that AI is in essence black boxes, the working of which cannot be understood in detail. AI, as a black-box need to be simplified. Whatever basis AI is using to operate, it must be able to be systematically analysed and corrected, without a complete wiping and starting again. Even with this, finding root causes and identifying the consequences of the repairs are likely to be extremely complex. However, identifying consequences is essential, given that decision-making itself is a complex process involving digesting information from various sources known to a firm. These can be internal or external sources. The process through which information is ingested to make decisions needs to be verifiable in every manner. This verifiability also needs to be granular in order to point to the root cause of any issue adversely impacting the decision-making process. Hex-E provides this granularity and verifiability of information necessary for decision-making. The Hex-E model is a model which breaks data into smaller and atomic level hex-elements. These granular hex-elements with their associated characteristics seek to establish relationship with other hex-elements. Hex-elements emanate from various sources. Hence, from an outsider's perspective, Hex-E potentially is seen a black-box. However, similar to how varied data sources can contribute to the Hex-Elementization process, a verification protocol is envisaged to be part of the Hex-E platform, which can help in the inspection and control function. The verification protocol is designed to not only check the validity of the informational flow but also provides an opportunity to sense-check the results.

3. **Context Sensitive.** AI is aware of whom it is serving. Data and its context are further studied in depth in the next sub-section of this chapter. Here AI can communicate what, when, how, and for whom it is operating. These skills are similar to AI being able to share its models from its user's viewpoint (Michie, 1990). Hex-E is designed to capture the context of businesses. The business context is what Hex-E will use when trying to find the answer through automated connection and coalescence of data from various parts of the business.
4. **Life Preserving.** All models make errors, and AI is no exception. There are processes to deal with human errors. Like the airline industry, which records errors and mistakes in order to learn, the feedback loop surrounding AI should have systems and processes to record all mistakes, learn from the mistakes and the near misses, to keep improving. This mode of learning from mistakes and actions is particularly important in businesses which operate in diverse conditions, amidst the changing competitive landscape. Competitive advantages are short-lived in the era of fast-changing and a fast data-consuming economy. Due to these challenges, businesses need have to heightened awareness to evolve and adapt quickly. The impact of mistakes might not take long to adversely impact the business's viability. Hence, the feedback loop involving AI in decision-making needs to learn, adapt and apply as dynamically and quickly as possible.

Artificial Intelligence as a disruption or enabler for decision-making

Disruption is the driving force for change in organisations (Des Horts, 1991). AI offers various potential benefits to managers, including handling jobs that can be managed by technology. The legal profession, for example, is impacted by AI-based decision-making. The professions in which rules are black and white (e.g. legal professions) may be impacted because decision-making can be automated using those rules. Professionals need to use their experience together with analytics to arrive at the best solution for the specific situation.

AI can disrupt many areas of technical management. Various functions of a typical manager could be facilitated by the development of AI, including managing personnel, supported by the HR department. AI supports personnel decision-making by linking skills to the most relevant jobs with greater accuracy and reliability than traditional methods. This process of linking is demonstrated by affinity matching tools that consider candidates' personalities as well as company culture (Hwang & Lee, 2013).

Ultimately, a significant part of the manager's job is at risk of being "Uberised" (i.e. at rent, on demand) by AI tools and robots (Dial & Storkey, 2017). The question is whether this means the disappearance of managers, as some who advocate the development of liberation management would like to see? Machines are unlikely to replace the human relationships that are essential in the real world. However, machines could facilitate the transformation of the manager of tomorrow into the "Augmented Manager" (Nee & Ong, 2013). It is reasonable to imagine that managers of the future will use AI intelligently within their main sphere of responsibility. However, AI tools and robots are unlikely to fully replace human decision-making. The contemporary manager demonstrates two essential leadership qualities: intuition and emotion. AI and robots can help the manager in his or her duty to manage human relationships by carrying out the automatable part of that activity with greater efficiency, and by helping the manager in their decision-making, thus transforming them into an "augmented manager". This augmentation might continue until AI becomes sophisticated enough to take away the most important functions of a typical manager in existing organisational settings. The manager will then be able to assume the roles of “leader of people and “guarantor of values”, the two roles that no AI tool or robot can fulfil in the short term. The manager, as a “leader of people”, will need to demonstrate even greater human and relational skills than those demanded today. Colleagues will undoubtedly have recourse to AI tools and robots to answers to the majority of common questions, but the manager’s presence will be essential when a colleague has a sensitive personal issue that needs to be handled delicately and confidentially. Furthermore, the manager will still have the final say on decisions concerning colleagues. The type of decisions which a typical manager makes today will be very different from what he makes in an AI-enabled workplace.

The manager, as a “guarantor of values”, is even more essential in businesses that are increasingly turning to AI and robots, particularly ensuring these technologies continue to be used ethically (Anderson & Anderson, 2011). The human factor may not be taken into account when rolling out these technologies, as has already been demonstrated, by the spread of new technologies such as mobile devices in the 2000s (Halverson & Smith, 2009). Moreover, this role of the guarantor is even more important, given that the values of trust, accountability, and cooperation are often negatively associated with the digital revolution.

Humans use their brains to find the optimal solution for a problem in a closed environment. A Natural Artificial Neural Network replicates this human brain process by using algorithms that take specific actions in various environments (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Also, as with humans, it is based on rewards. By exploring the environment, this technology finds the optimal solution that maximises the reward. This technology is in the nascent stages, and the human brain is still superior in many areas such as speed and processing capacity. Today, artificial neural networks are mainly used for image recognition with promising results, even though the networks still need weeks to recognise an image. A futuristic scenario might include how an artificial neural network from one company works with a counterpart from another company to make a new deal, improve efficiency, improve supply chain, or share resources. Such a system could also help customers by highlighting complementary products, among other things. Also, these three stages of AI development are not sequential as depicted in Figure 14 but evolve in parallel and at a very different pace. These technologies, unlike past industrial revolutions, are not constrained to larger organisations. These technologies are available to both large and small firms equally. Because of this equity in availability, businesses that do not embrace technology innovations leave themselves open to disruption from other businesses. Disruptive technology provides other businesses with a competitive advantage in many ways.

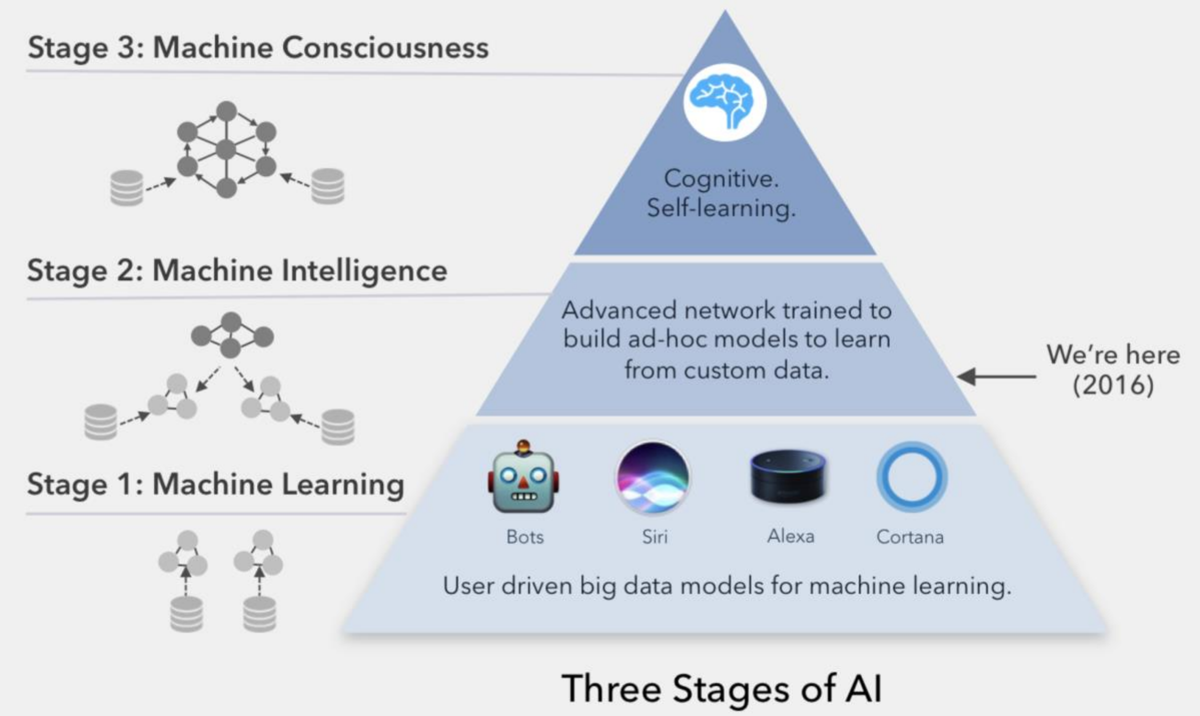

Figure 14: Three stages of AI *(Paka, 2016)*

In practice, AI is a precise tool that should be used in a planned way to achieve specific goals. Implementing it takes prudence and vision, along with a strong understanding of the technical challenges involved (Hryniewicz, 2018). The initial adoption of AI in the field of business intelligence is promising, given the number of activities businesses are trying to quantify. AI can help augment this era of quantification, which has the potential to help improve business decisions. This improvement is not just in the form of better-informed decisions but also factors in more robust variables which were difficult to incorporate in traditional businesses. Until Chief Executive Officers (CEOs) understand the implications and requirements surrounding the use of artificial intelligence in decision-making, enterprises are not prepared to systematically enter the arena of AI (HortonWorks, 2019). Here are three things that every CEO should understand before tackling AI at the strategic level (Carley, 2002).

1. AI Must Be Aligned with Goals

In a McKinsey survey of 3,000 executives, 41 percent said they are uncertain about the benefits of AI (Bughin, et al., 2017). AI can help businesses by automating relatively simple tasks that would take humans 30 seconds or less to complete (HortonWorks, 2018). AI can look more deeply at data to find patterns that humans may never catch that helps identify variables essential to strategic decisions. These are two very different capabilities, and learning how these capabilities impact business outcomes are critical for success. Businesses must decide if automating simple tasks, such as sorting cucumbers on a production line, reading basic loan agreements or even improve medical outcomes reduces bottom line expenses. AI can be used by hospitals to monitor patient movements using wearable devices to predict the immediate risk of a fall or assist doctors in spotting and diagnosing potential health problems on MRI images. Regardless of the organisational goal, it is essential to use AI to support the user experience. A doctor could benefit from diagnostic recommendations, but time constraints in a busy practice may only make this useful if delivered directly into a medical record during a patient consultation (HortonWorks, 2018). A retail customer using natural language AI to help pick clothing has a different set of requirements. The use of AI in decision-making must fit workflows and formats that make sense for users. Hence user interface, business analysis, and workflow are essential to factor and design. The survey conducted by McKinsey highlights how businesses are adopting AI, summarised in Figure 15. Most of the existing adoption is centred around industries which have appropriate use and structure to embrace AI. They are data-centric and could benefit from a strategic partnership with multiple stakeholders in the business, discussed in the next two points.

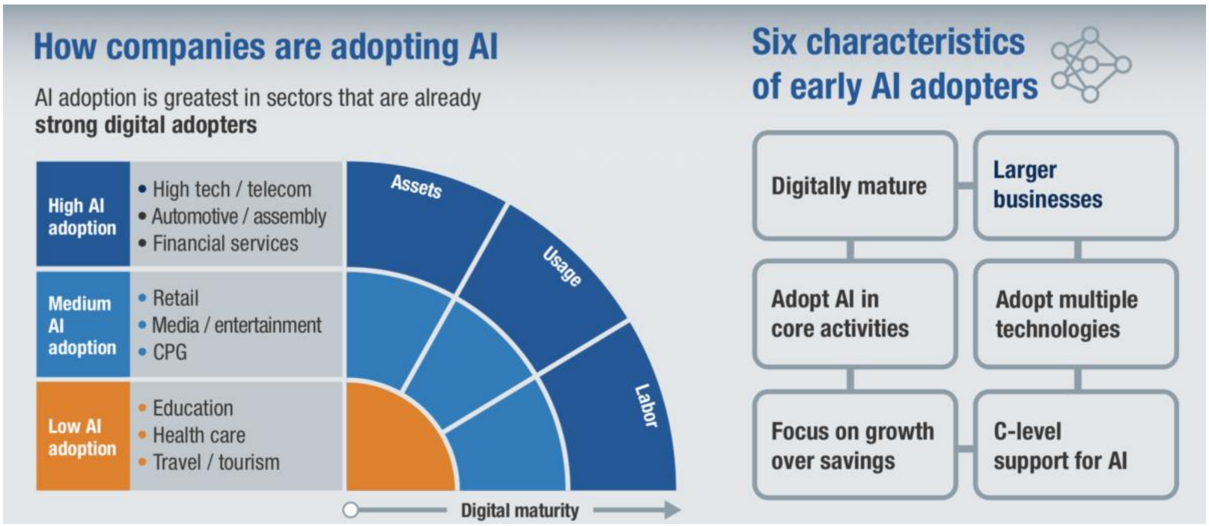

Figure 15: How businesses are adopting AI - McKinsey Survey *(Bughin, et al., 2017)*

2. Organisations Must Be Data-Centric

AI thrives and prospers on data. The neural networks that are typical of AI systems learn to make better decisions using vast amounts of historical data (Kaastra & Boyd, 1996). The networks then produce models that organisations must continuously refine as new data comes into the organisation. The question is whether the organisation continues to provide this stream of data. Most organisations have existing IT infrastructures built up by multiple teams over many years. The result is a fragmented information landscape. Data resides in different systems that do not easily integrate with each other. Management power structures can worsen this problem, creating tensions that cause people to restrict access to their data. Breaking down these human and technological barriers takes a mixture of leadership and investment in technology (Hryniewicz, Horton Networks, 2018). A CEO committed to strategic AI unifies the data architecture. Smooth exchange of data throughout the organisation can help with many projects other than AI It is a foundational practice for the modern, tech-savvy organisation.

3. Organisations Should Pursue Strategic Partnerships

AI may consume a lot of data, but it must be the right data. Understanding which information is needed in order to train AI models requires multiple kinds of expertise. The first is domain knowledge. Organisations need people with an intimate understanding of the organisation's operations and how it uses different kinds of information to achieve specific results. The second type of expertise involves data science. Data scientists work with database administrators to extract, manipulate, and prepare data for AI workflows. If a data science platform is working from this perspective, a developer will struggle to retrain, automate, and deploy models for consumption and "productising" for consumption by a range of applications (Hryniewicz, 2018). This group of skills, from data science to software development to business process design, is a challenge for many businesses as they attempt to transform their businesses. Combine this with the need to engage Cloud services for compute-intensive neural network training and many CEOs may be overwhelmed (Hryniewicz, 2018). Partnering with product and service providers that have a track record of navigating through the AI design, development, and deployment process is a proven way to help overcome these AI hurdles and drive the business to success. Technically adept partners have the technical and business understanding to assist with the use of artificial intelligence in decision-making. These partners also advise how to manage the barriers to data aggregation and management. In such a fast-moving field, it pays to have specialist expertise to help guide businesses in the journey. The results could advance the business quickly, while others spend resources to solve the problem in-house. Businesses and CEOs who are planning to adopt AI should define the workflow so AI could pursue a strategic partnership, internal first and then external, to create an eco-system which can be utilised to help make tactical decisions.

AI technology is a means towards effectiveness and efficiency, improved innovative capabilities, and better opportunities. Several industries have begun to adopt AI into operations to achieve these benefits. According to a survey by Tech Pro Research, up to 24% of businesses currently use or plan on using artificial intelligence. Standout businesses in this effort are in the health, financial services, and automotive sectors (Sincavage, 2018).

Before the resurgence of AI and its commercial application, business stakeholders have relied on inconsistent and incomplete data. With AI, they have data-driven models and simulations to turn towards for decision making. According to PwC's Rao, limitless outcome modelling is one of the breakthroughs in today's AI systems. He adds, "There is an immense opportunity to use AI in all kinds of decision-making" (Rao, 2018). AI systems in the current state start from zero amount of data and feed on a regular flow of big data. This is augmented intelligence in action, which eventually provides executives with sophisticated models as the basis for decision-making (Sincavage, 2018).

The Hex-E platform seeks to leverage the benefits of AI. Business executives face many challenges in decision-making. One of these challenges is the limited time to address issues from various parts of the business. Not only do they need to make tactical decisions to address certain issues in a timely fashion, but they also need to consider the long-term goals of the business. Hex-E is designed to be an able partner in decision-making for business management. Driven by AI, Hex-E will connect data which is relevant to the business and find connections to provide insights to business managers. Hex-E not only aims to provide shorter-term insights but also provide insights on how it might impact various parts of the business from a longer-term perspective.

The Role of Data and Context in Decision-making

Understanding a data point encompasses its storage, processing, and security needs (Unhelkar, 2017). However, the use of a data point depends on an important dimension which is its context. Context drives business decisions. Context not only provides the building blocks of the decision-making process but measures the success of those decisions against the context. Hence, the discussion surrounding context indicates the complexities of definition surrounding the data point, which is needed as an input for decision-making (Watts, Shankaranarayanan, & Even, 2009).

Contextual awareness can be defined in many ways. Context can be understood as one or more reference points surrounding a data point that enable interpretation of the primary data point (Unhelkar, 2017). Data on its own, is factual and does not have meaning. Reference points around the data point provide that meaning or context—and thereby change how data is interpreted and analysed (Nair & Lan, 2016). Context provides essential and adequate information to undertake a reasonable analysis of data and enhances the ability of a business to achieve the desired outcome. Contextual awareness includes understanding the business situation where the data analytics from the data are used, setting and describing that situation, spotting any changes, incorporating the feedback from the results of analytics based on the new context, and reinterpreting data based on the changing context (Unhelkar, 2017).

Context has an iterative relationship with business. Context helps to focus on business outcomes—and understanding the outcomes helps create context (Unhelkar, 2013). Thus, context narrows down the use of data relevant to a specific outcome. Data analytic frameworks need to be context aware (Chen, Chiang, & Storey, 2012) in order to provide outcome-based and agile analytical solutions.

Figure 16 shows the likely reference points and characteristics pivotal in establishing a data point (Agarwal, Govindu, Lodwig, & Ngo, 2016). The context is divided into two layers; a) the immediate or direct layer and b) the secondary level. Location, time, feature, and person are the characteristics providing the context in the immediate layer. These references are in turn made up of additional sub-reference points for the secondary level. For example, as shown in Figure 16, the location is made up of latitude, longitude, and altitude. The people reference point is made up of solution provider, business user, and end user. Additional layers can be added depending on the availability of data and business outcomes defined.

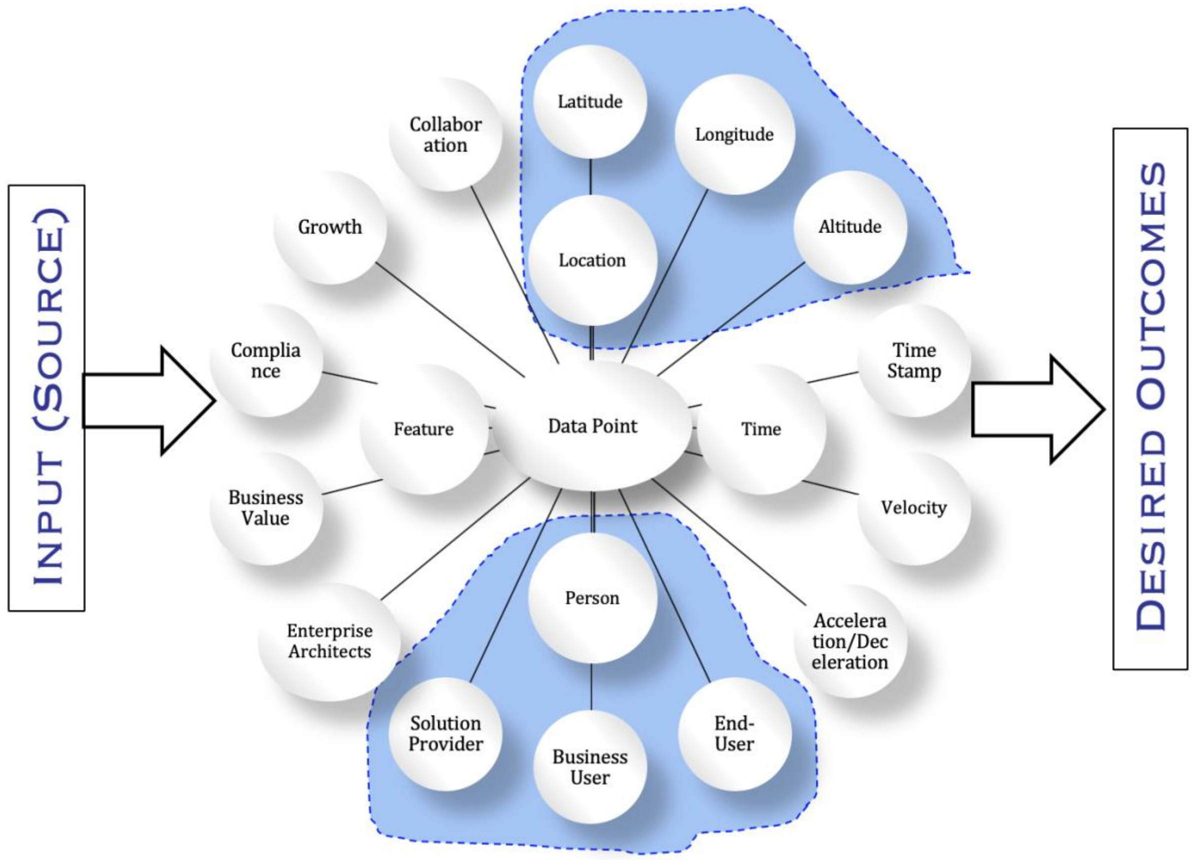

Figure 16: Data point and the context *(Unhelkar, 2017)*

Developing the context and using contextual awareness is of immense value to Big Data analytics (Unhelkar, 2017). At the most granular level, each IoT (e.g., temperature gauges, smartwatches, smoke detectors, smart shoes, and home appliances) is a data point with many additional data points embedded within it (Nair & Lan, 2016). These devices can send, receive, and process data in collaboration with other devices, and the back-end Cloud, in a dynamic fashion and in real time. Strategic use of Big Data goes beyond embedding a sensor, sending signals and receiving data points over the Internet, and processing them (Schmidt, Möhring, Maier, Pietsch, & Härting, 2014). Instead, context awareness is used to analyse data from multiple sensors coming from varied sources using advanced algorithms, in real time, to develop a 360°, holistic view of the data point for enhanced and Agile decision-making (Unhelkar, 2012).

Usually, in most of the circumstances, the context can be ascertained through some or all the five Ws— Where, When, Who, Why and What. For example, a simple IoT device may only need to answer one or two Ws, while a more complex IoT device may need answers to all four Ws and, perhaps, even additional questions, such as How, Why, Which, and How much. Table 3 shows an example of a contextual reference for a cash amount data point when used in ascertaining business outcome (Unhelkar, 2017).

Context (Reference Point)	Description of the Reference Point	Example of the Contextual Reference Point for "Cash Amount"
Who	Stakeholder	Bank customer
Why	Goal	Withdraw cash
What	Technology	ATM
When	Timing	Late night
Where	Location	Bus stop
How	Process	Debit card

Table 3: Context parameters a data-point (cash amount) vis-a-vie Business Outcomes

The contextual references for each data point are further augmented by the advances made in the field of Machine Learning. ML is considered at the cutting edge of data science (Mjolsness & DeCoste, 2001). ML plays a pivotal role in achieving more granularity, such as identifying the precise product for a customer at a particular point in time, narrowing down potential areas of fraud and money laundering, and enabling emergency services (e.g., ambulance and fire) to position themselves for rapid responses on certain days or at certain events. Although there are diverse definitions of ML, in short, it can be understood as machines (e.g., computers) learning iteratively, incrementally and interactively from the data they receive and the feedback they receive from the end-user (Unhelkar, 2017). The feedback from the end-user is pivotal in making the machine understand whether the analysis it performed was correct or not. This feedback loop is essential for machines to learn over time. It is no different from humans who learn from every wrong decision they have made in the past and try to correct their mistakes in the subsequent decisions. The novelty of ML in the age of Big Data is its applicability to a huge amount (volume) of high-velocity data (Tonidandel, King, & Cortina, 2018). The vast amount of data sets, affordable data storage, distributed processing, high-performance computing, and the analytical opportunities available have dramatically increased interest in ML systems (Mitchell, 1999). Branches of ML, including Deep Learning and Neural Networks, focus on creating correlations between data points without explicit instructions. If data points can relate with each other automatically, opportunities open to create previously unknown analytical insights (Unhelkar, 2017). ML is designed to learn iteratively (an Agile characteristic), provided its explorations and learnings are embedded within a set of acceptable business rules.

While Big Data analytics create insights, those insights are still limited to the effort and imagination of the individuals undertaking the analytics (Unhelkar, 2017). One of the issues with businesses not embracing the ML-based models is due to the black-box nature of what happens in the background of an ML algorithm. For example, given specific inputs, the ML comes up with outputs which match the needs of the business. When the ML is trained on many inputs which derive certain outputs, the algorithm creates patterns (i.e., business rules) which it learned iteratively and incrementally. This series of patterns which the machine has created through its development is a black box which is feared by many in the sphere of business. This fear is due to the fear of the unknown. The unknown is the inability to know the exact working of the algorithm which created those patterns. However, if a human is asked the colour of an apple is, the answer comes quickly “apple is red”. If the same individual is asked how he derived his answer and what happened in his brain before he was able to answer that question, he may not have a response. The reason is that humans still do not know which neurons fire in the brain, explicitly or implicitly, to understand how the response was generated. The brain is a black box. However, businesses tend to be positively biased towards organic black boxes (humans) than digital black boxes (ML driven decision-making process).

Another challenge which businesses face is the evolution of automated ML (Auto-ML). Models can be built through evolving, iterative and interactive learning that can be described as metaprogramming. The concept of automated ML-based model building leads to questions (Unhelkar, 2017) such as:

- How should data points be enabled to correlate and integrate in order to form a sensible mosaic that will be of interest to decision makers?
- What are the defining characteristics surrounding a data point to enable it to seek and connect with another data point?
- How can a connection between two data points provide feedback or “learning” mechanism for the background algorithm ultimately helping decisions better after every iteration?
- How many properties of a data point will be the most optimal to enable it to create new (and sensible) links?
- What is the business value of such automated ML-based interconnections? (Unhelkar, 2017)

The existing business intelligence frameworks available in the marketplace fail to answer all the questions highlighted by Unhelkar while discussing Big Data strategies for agile business. The research in this thesis highlights Hex-Elementization as a concept conceived to provide answers to some of these questions. For example, there is a need to arrive at a basis of providing optimum levels of context reference points for an analysable data point. Hex-Elementization involves describing a data point, irrespective of its form, as “hex-elements” with six pieces of reference data points and then integrating those reference points. Figure 16 shows two such data points, each with six reference points that are trying to connect. These data points with six elements are amenable to much easier connections with each other than fewer or more than six elements (Nair & Lan, 2016).

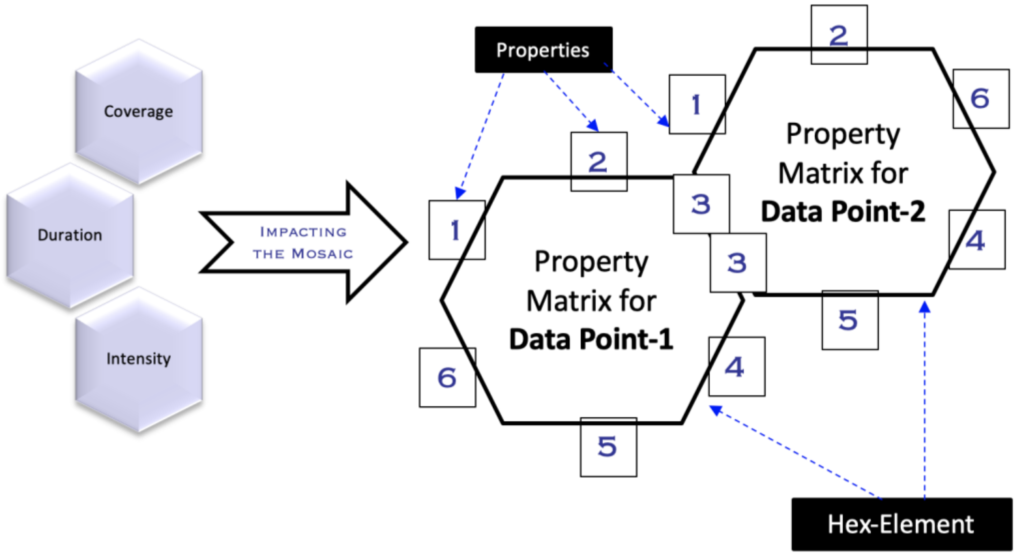

Figure 17: Contextual wrapping around Hex-Elementization

The two data points, shown in Figure 17, can be part of a set of unstructured data or structured data, including microwave or digital signals, sensor data, machine language, a lingua used by robots to communicate with each other, or any other conceivable way of communication. Hex-Elementization treats the data point in its simplest form while embedding enough attributes to help it automatically seek and match with other data points. Figure 17 further shows how the desired coverage, duration, and intensity of analytics can be used to help channel the connection between unrelated pieces of hex-elements.

In theory, this continuous connection between Hex-Elementized data points can go on ad infinitum, resulting in a data mosaic limited only by the physical capacities of the devices.

Consider, for example, IoT devices. The data flow from a given IoT device is broken down into hex-elements with a set of six properties (Figure 47). There is no limit to how many IoT devices can be integrated via Hex-Elementization. Each set of hex-elements from each IoT device tries to interconnect by seeking common properties. This connection enriches the flow of information as the analytical process gathers more data from each new stream of hex-elements emanating from each IoT device. Each flow gathers the related information it seeks to create a new informational flow.

Consider another example of thousands of x-ray photos, computed tomography scans, ultrasound scans, blood reports, and other pieces of data sitting in the digital storage of a hospital. Hex-Elementization provides the basis for potential ML algorithms that can take these data points and produce interesting and even unexpected relationships and information snippets for the medical staff. The information stream or “hex stream” is made up of a chain of hex-elements from disparate sources that are grouped and unified by common properties to create a chain of rich and interconnected information.

The journey of a Context-Based Data Point

The journey of a data point is relative to its representation with six reference points or six characteristics, and the input provided by the context-modelling and analytics engine. Figure 18 shows such a journey of the data point (defined by Hex-Elementization) via a context and an analytics engine (Unhelkar, 2017). The data is based on available heterogeneous sources, and its representation is based on the context analytics engine. The context analytics engine also enriches itself by learning from the previously related data-context relationships based on past decisions. This enrichment is driven by technology such as Machine Learning. Data is then combined from these heterogeneous sources and varying formats depending on the context. The output of the context engine is the context, which is used by the analytical engine in order to manipulate the data points. Pre-trained ML models supplemented with visuals and other forms of presentations can be used to provide feedback into the context engine. This feedback then allows dynamic updating of the context and further refinement of the analytics in real time. The automation of this journey of a data point and its replication to all other data points provides the business objective and determines the optimum level to which analytics can be performed.

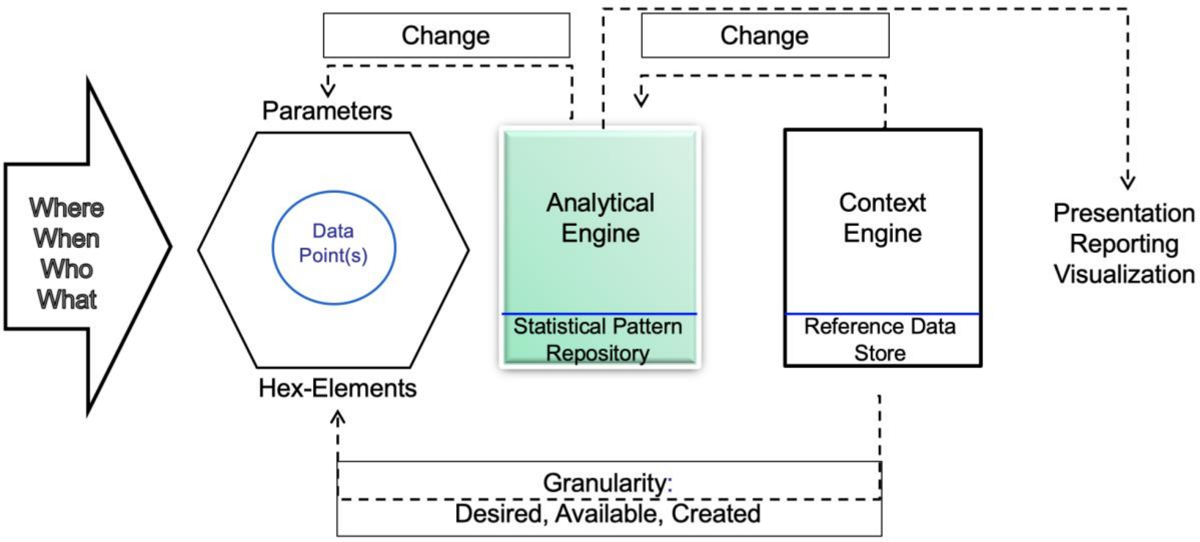

Figure 18: Journey of data-point via Context Engine & Analytic Engine the Feedback Loop (Unhelkar, 2017)

The Granularity of Data, Analytics, and Processes

The concept of treating the most atomic level of data can be explained further through granularity, generalised in many aspects. Granularity is an essential differentiator of data from Big Data (Unhelkar, 2017). The granularity of processes is the level of activities within the process where decision-making is undertaken. While data is fed from the top of the analytics funnel (which can undertake predictive, prescriptive, and exploratory analytics), it becomes increasingly granular as it moves down the funnel. The more granular the analytics is, the greater the opportunity for business agility and, the greater opportunity to make optimal decisions. Optimal decisions can be made because granularity enables precision and speed in decision-making within a business process.

Granularity can be applied to any aspect of the context of a data point (e.g., time and location) (Unhelkar, 2017). Granularity can be applied to business functions and processes (e.g., replenishing an inventory item or directing a cooling tower). Granularity also applies to the input and storage, and at the process level.

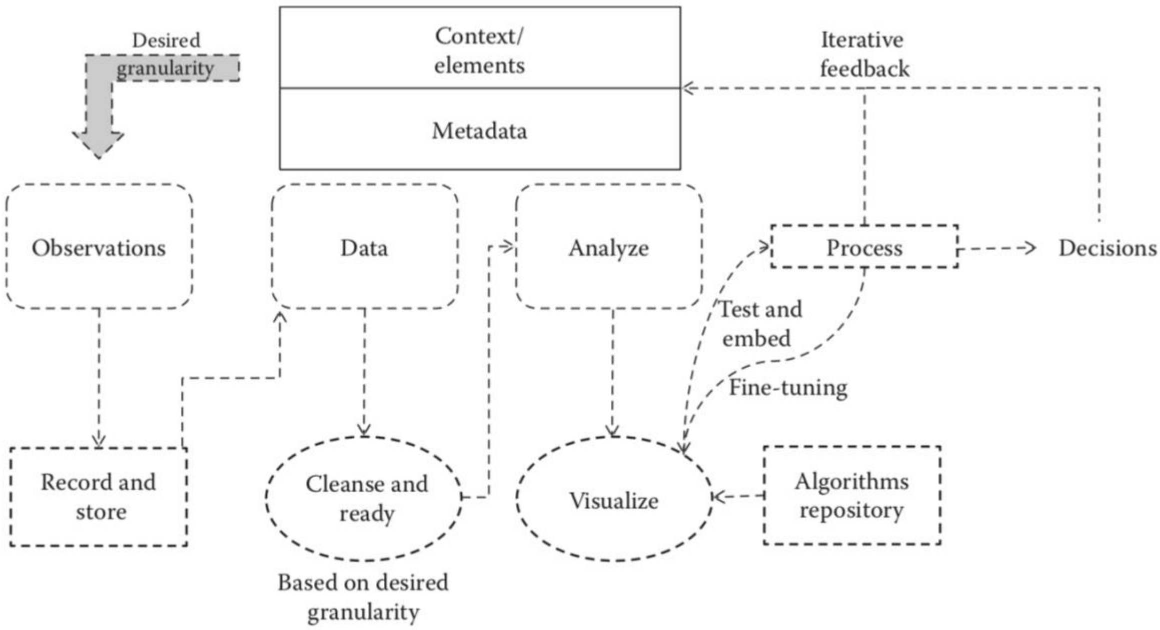

Figure 19: Data analytic processes based on granularity (*adapted from Unhelkar, 2017)*

Granularity can enrich the customer experience (Unhelkar, 2017). This enrichment exists because the context and analytics engine provide metadata and characteristics that can help narrow down customer choices, create personalised messages, provide instantaneous offers, and handle any issues or complaints. Figure 19 shows the details of the data analytics process that caters to fine granular requirements. Fine granular analytics can help understand the customer decision-making process. Organisations can use these insights to develop more effective services and products, more dynamic pricing, and appropriate post-sales strategies. Probabilistic models for decision-making are more finely granular than deterministic. The models apply more readily to individual departments in the business rather than the entire organisation. Granularity can help business in;

- Understanding the narrow window of opportunity for decision-making and make it convenient for a business stakeholder to make rapid decisions within the limited time frame.
- Improving the quality of the product or service based on precise, personalised insights that include factors specific to an individual business stakeholder.
- Enhancing the predictability of outcomes in terms of product sales and service quality due to the finer granular analytics to consider multiple factors outside of the standard analytical factors.
- Improving agility and sensitivity to the stimuli to the organisation from internal and external sources because of the speed with which those stimuli are received and incorporated in the finer granular analytics.

Factors that impact the levels of granularity of analytics:

- **Desired business outcomes from the analytics**: The greater the need for precision and speed of response, the finer the level of granularity needed.

- **The maturity of the context engine and the ease of plugging them into analytics**: The more mature contextual patterns are, the easier it is to undertake finer levels of granularity for business analysis.
- **Level of Hex-Elementization within the data population**: This is the clarity of definition of properties around each data point. The greater the clarity, the finer the level analytics.
- **Level of integration**: Required between various collaborating systems in order to produce insights. The more complex the need for integration, the costlier it gets to set finer granularity levels.
- **Structure of data**: Structured and unstructured data is proposed to be handled through a framework like Hex-Elementization, considering the properties of the data need for analysis holistically.
- **The volume of data on which the analytics are to be carried out**: The greater the volume, the better the opportunity to identify a reliable pattern, but the costs in terms of resources (computing needs, storage, and analysts) of finer granularity also go up.
- **The velocity of data in relation to the required speed of analytics**: Faster incoming data can lose its value quickly if the analytical response in the form of tactical decisions, is not produced within an equally fast timeframe.
- **Performance**: Speed of the desired response based on the "currency" of data to create an analytical result. The higher the need for performance, the greater the costs and effort associated with making the granularity levels finer. This process of "granularising" has an overarching impact on when and how the decisions can be made.

- **Communications**: Networks and security layers within the solutions environment. Enhanced communications enable faster transmission of responses, thereby enabling finer granularity analytics to provide value to the user.
- **Breadth of focus**: Looking at the number of users, the extent of their decision-making, and the volume and variety of data. While broader focus means richer data sources, it also means more effort in making fine granular decisions.

Summary

The literature review conducted in this chapter is sourced from various fields, including Business Intelligence, Decision-making, Information Technology, Psychology, and Organisational Management. The topics covered here including viewing data from the most atomic level and building context around different data sets, and the subsequent impact of granularity, in terms of data and processes, are important considerations for business intelligence. The impact of each of these features dictates the mode of extracting information efficiently to enable decision-making. Decision-making, to be a successful process, needs to have all these components working efficiently in tandem. Data, context, and granularity is something Hex-Elementization targets to unify to help business managers extract the business intelligence they need to make optimal business decisions in the least amount of time.

Some of the theories explored include Zachman's and Spewak's enterprise architecture and its impact on decision-making, Big Data theories and its impact on real-time decision-making, and the online presence of organisations and the necessity to make decisions tactical and timely. These fields were selected for the literature review to determine if there was any precedence to the concept of Hex-Elementization in any form or manner, which might make this research redundant. In literature, the concept of Hex-Elementization stands out as unique and has not been addressed in any manner. The application of an atomic level of data in business management with a view of integration, in current or future circumstances, is absent. The framework of Hex-Elementization was conceptualised with a view of enabling decision-making faster with impending challenges with data one faces in this new age. The concept of Hex-Elementization can be applied to various activities of business management and not just decision-making. The concept of context-driven decision-making on a real-time basis is difficult to achieve due to the gap which exists between data collection, data analysis, and subsequent decision-making.

Psychological evaluation, database structures, Big Data frameworks, and the statistical probability of social network connections are some of the things which were discussed in the literature review. Although each of these concepts works in isolation and each can help make decisions at a faster rate or more accurately, none of these examples of literature is scalable. Big Data is nothing but a repository of a collection of little data in either higher frequency, or velocity, or quantity, or variety. Although data existed in large quantities before, the advent of affordable HPC (high-performance computing) and distributed computing has changed the responsibility of analysis from the originator of data to these new technologies. A department no longer needs to expend an enormous amount of resources (mostly in capital) to acquire servers, computers, and software to crunch data to help decision-making in a typical business. Businesses can easily leverage these new technologies at a margin of what it used to cost ten years ago. Technology has made it possible to extract information from data dramatically more cost-efficiently over the years.

The data analysts often overlook the incremental use of an extra piece of data. This oversight is because Big Data is often confused with just a large quantity of data. With the advent of IoT, the previously narrow-minded view of Big Data has started to change. IoT entailed in hundreds and thousands of chips, devices, gadgets, machinery – equipped with the latest microprocessors which now costs pennies compared to 10-years ago. The literature review highlights gaps in the existing body of knowledge amidst the issues discussed surrounding decision-making.

To conclude, Hex-Elementization stands apart from the existing works of literature and adds to the body of knowledge relating to business intelligence by looking at the data at an atomic level in order to help generate insights in an organic way.

# No	Theory	Contribution to the Research
1.	**Decision Making in the Big Data Era**	**Discusses various theories which deal with making decisions with the explosion of data. The theories do no provide a framework to make effective decisions.**
2.	**Data-Driven Decision-Making**	**Discusses theories which highlight the challenges when it comes to incorporating more data into decision-making versus traditional means (intuition)**
3.	**Six-Degrees of Separation (Information Theory)**	**The six-degrees of separation in the context of information theory which tries to highlight how connected is the world and how data and decisions are so closely linked even if its not explicitly obvious.**
4.	**Quantum Information Theory**	**Quantum information theory highlights the gap in the theoretical world in which data has not be granularised further to ease data integration.**
5.	**The Zachman ISA Framework**	**Discusses Zachman's framework in the context of decision-making and the gap it has when it comes to making decisions in the era of Big-data and AIML.**

6.	**The Spewak EA Planning Method**	**Similar to Zachman's model Spewak's planning model for decision-making highlights how organisational structure needs to be rethought in the era of fast-paced decision-making.**
7.	**EA3 Cube Framework**	**EA3 cube framework, which is one of the best theories followed fpr decision making structures, lead to inefficient frameworks which is not optimal for modern-era decision-making.**
8.	**Business Intelligence Theories**	**These theories highlight the fact that business intelligence is mostly pull-based in which managers need to know when they want from these systems. The gap which Hex-E tries to plug here is to propose a push-based BI system.**
9.	**Decision making in the age of Artificial intelligence**	**AI will change the way decisions are made. The theories around AI based decision making is discussed in the context of what Hex-E can offer.**
10.	**The Role of Data and Context in Decision-making**	**Contextual decision making is time-consuming and resource intensive. The theories are discussed in context with the benefits provided by Hex-E.**

Chapter 3: RESEARCH METHODOLOGY

The relevance of Research Method to this research

This research into Hex-Elementization is undertaken formally in order to understand and resolve the research questions highlighted in Chapter 1. The need to identify and follow a formal research method for this research could not have been greater. This need for a robust research method is explained in detail further in this chapter. In addition to the research method itself, there is also a need to utilise a tool to undertake analytics. This chapter also explains the tool, together with the research approach and the research methodology utilised in this research.

This research explores a framework (Hex-E) to facilitate generation of business intelligence in the corporate world. The unique contribution of this research is the concept of atomising diverse types of data to enable easy integration with Hex-Elementization. Use of data in the business world is currently limited by the silo-based structure of a database or a specific framework. Existing systems are based on a common shared architecture or environment, instead of incorporating the differences or diversity in the design and architecture of hardware or software. As the use of IoT and Artificial Intelligence broadens the difference between hardware and software, dependencies on architectures are reduced. The literature review discussed the challenges rigid enterprise architecture brings to decision-making. In rigid enterprise architecture, the generation of intelligence is dependent on architecture design. This issue of addressing the dependability is at the core of Hex-E. Hex-E aims to reduce or eradicate this dependence on architecture designs to enable the free flow of information.

Information permeating across departments to form intelligence is valuable to the business decision makers.

A Mixed-method Approach

Research methods are traditionally classified into qualitative or quantitative methods. The mixed method research approach is more than just the combination of "qualitative and quantitative approaches in the methodology of a study" (Tashakkori & Teddlie, 2010). Mixed method research helps achieve a complete and corroborated result, as it counterbalances the weakness of one primary method with a second method. This approach is more powerful than exclusively using either quantitative or qualitative in conducting the research. The mixed method approach enables investigation on multiple inquiry lines, which can be missed by deploying just a single line of inquiry using a single method (Shannon-Baker, 2015). Mixed method research, in essence, helps the researcher to experiment with certain segments of the research before the actual research begins.

Hex-Elementization is a new concept in theory which helps integrate the disparate set of data from different sources, systems and environments. A study like this, which is innovative in practice, needs a holistic understanding of the work and progress towards creating a model for change (Walton, 2014). Exploratory concepts, like the one at the centre of this research, pose a unique problem to solve. A mixed method is ideal for this scenario as both words (qualitative) and numbers (quantitative) help solve tangible problems, by combining inductive and deductive logic through abductive thinking (Morgan, 2007).

Creswell aptly defines the core characteristics of mixed method research (Creswell, 2014) by providing four distinct actions when conducting mixed method research. Those actions are:

1. Collects and analyses both the quantitative and qualitative aspects to address the questions posed by the research.
2. Combines and contrast the two forms of analysis and the ensuing results.
3. Organises the procedures to adhere to the research design to provide logic flow and process for the study.
4. Frames the procedures within theory and philosophy to support the hypothesis.

These core concepts of mixed research aptly apply to the framework of Hex-E. The concept of Hex-E is exploratory in nature and encompasses fields from business management to information technology. To address the broad-ranging concerns of these fields in the development of the framework, it is important to employ a robust research methodology. The Hex-E platform will be able to leverage the core concepts of a mixed method, in which the triangulation of both quantitative and qualitative analysis provides logical flow to the study. A mixed method appropriately supports the bases which Hex-E is designed to cover in this research.

Exploratory Sequential Design – the Mixed Method Approach

There are many classification and methods developed by experts over many years. Creswell and team have developed various typologies for mixed methods throughout the last 15 years. Their present typology of the core mixed-method designs (Creswell & Plano-Clark, 2017) simplifies the method into three core designs a) Convergent design b) Explanatory sequential design, and c) Exploratory sequential design. As Creswell and Plano-Clark explain, the differences in these designs lie with the researcher's intent when conducting the research. In convergent mixed method design, data and analysis is conducted through two different methods (qualitative and quantitative) and then results are compared, contrasted and combined to answer research questions. The explanatory sequential design is used to explain the result of the first phase of research, that of quantitative data, by conducting qualitative research as the second phase.

This research design explains the statistically significant or insignificant results, outliers, or positive-performing exemplars of the research. In this design, the qualitative data helps identify why and where the quantitative results are pointing (Creswell, 2014). The third design, the exploratory sequential design, is used to explore a phenomenon or concept which cannot be analysed using off-the-shelf tools (Tashakkori & Teddlie, 1998). The exploration is needed because there are either no tools of measurement available, or the variables driving the need for research are unknown. At times there are no guiding or existing theories or frameworks to follow, and a need exists for a measure to fill in the gaps highlighted by the participants in the research.

Given Hex-Elementization proposes a model to create a business intelligence framework encompassing disparate sources and mediums of data in a unified and organic manner, the research needed is supported by the defining aspects of exploratory sequential design. The research design in Figure 20 shows the flow of research through various phases and the intended output at the end of each phase.

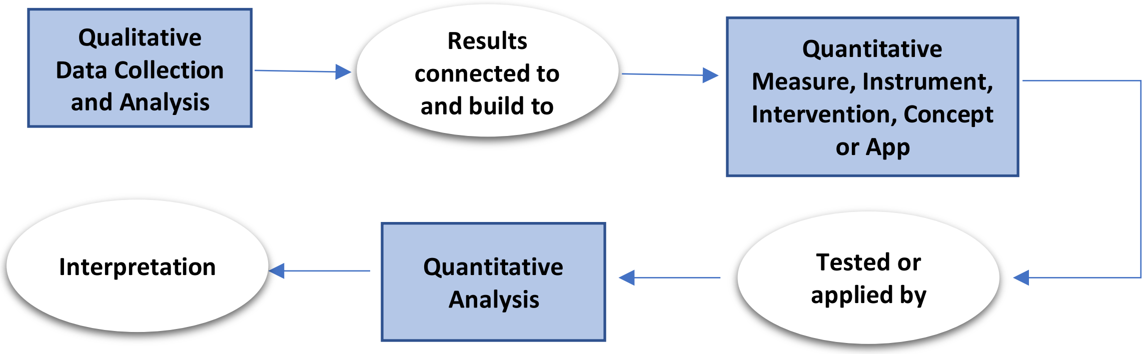

Figure 20: The Exploratory Sequential Design

This research is conducted using selective qualitative methods post gathering the data and analysis stage of quantitative information. Planned interviews are considered suitable for verification using the qualitative method as compared to quantitative. Further rationale for selecting the exploratory sequential design can be attributed to five major reasons as highlighted by Creswell (Creswell, 2012).

1. **Exploratory**. This research is exploratory by trying to understand how businesses undertake decisions in the era of Big Data.
2. **Complexity**. Given there are few theories or practical examples of a data integration model for business intelligence to enable real-time decision-making, the research is highly complex in both concept and goals.
3. **Context**. The decision-making process revolves around context. And context is subjective, and it depends on the organisational circumstances. Since the context itself continues to evolve in the era of exponentially growing data, this research is suited to be conducted using a mixed method.
4. **Explanation**. The research provides links between data, the creation of information, business intelligence and decision-making. These links are better explained by a theoretical hypothesis which is qualitative, rather than quantitative.

5. **Difficulty in measurability.** Although it is understood that the use of data in business decision-making is important, there are limited ways to accurately quantify this phenomenon. The research targets creating business intelligence in a real-time and accurate manner for improved decision-making. Again, measuring the impact of this research on business decision-making cannot be accurately quantified, and needs to be validated through qualitative judgement.

The dominance of the qualitative method in this research paper is attributed to the proposed Hex-Elementization model to be verified for practical business application. The research conducts investigations and refines the pillars of the research objectives when qualitative feedback is received throughout the research process. The research is conducted in phases (Figure 21) with considerable overlap between each phase.

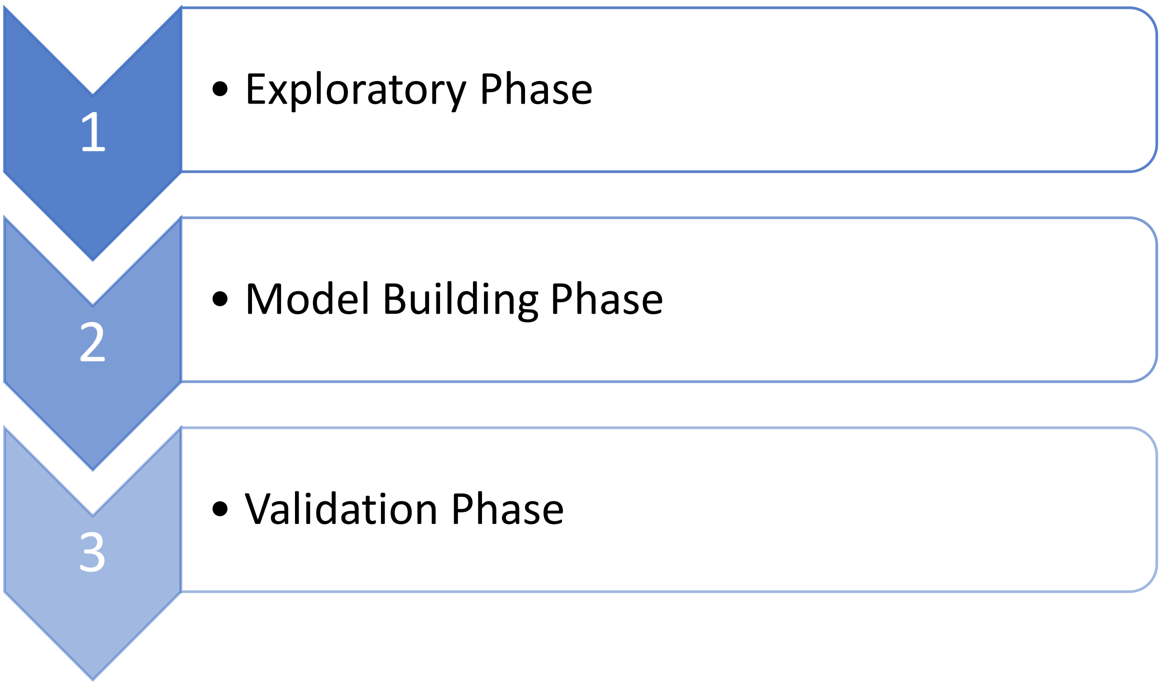

Figure 21: Research process flow

Phase 1: Exploratory

The initial phase of this research is exploratory, as the goal is to understand the existing methods of decision-making. This phase explores the ways businesses in the current environment incorporate data, and the intelligence created from the data, to make accurate and timely decisions. Collecting and analysing data both quantitatively and qualitatively promotes understanding current practices and collecting information (data) to help validate and verify the applicability of the proposed framework to business. For the quantitative analysis, a web-based survey was sent to many individuals to gather their opinion on the state of decision-making in the current environment. For the qualitative phase, a small representative population of the individuals were selected to conduct interviews, to help observe and build learning models through at least ten organisations.

The interviews targeted senior stakeholders, including Chief Information Officers, Chief Executive Officers, and middle-level managers. These stakeholders take charge of the businesses they operate and manage, and consume business intelligence to set their Business Strategies (Kvale & Brinkmann, 2004). This group of stakeholders provide the best judgment of how efficiently data and information are turned into business intelligence and also highlight how real-time business intelligence provides a more tactical approach to business decisions. The research involved interviewing 10 senior- to mid-level managers from various organisations. The mid-level manager acts as Point-Of-Obligatory-Passage (Raufflet & Mills, 2009) between data or information and the business intelligence that is ultimately passed on to the senior management. The proposed Hex-E model challenges businesses to rethink the use of data for business intelligence (Chevalier & Buckles, 2013). In some cases, the discussion encompasses various departments of the business to understand the flow and use of business intelligence, which had to break through existing corporate thinking within these firms. Thus, the validations were reinforced by qualitative methods.

To explore current decision-making processes, the interviews were conducted by carefully selecting a few organisations which fulfil the following criteria:

1. A data-driven company in which data is captured and utilised in various forms.
2. A company in which data and the information derived from this data is needed to conduct business strategy (both long-term and short-term).
3. Innovative in finding new ways of utilising data for a wide variety of purposes with a mild interest in emerging technologies such as Big Data and IoT.
4. Consumes data and information for various parts of its business and are looking for real-time solutions for faster and more accurate decision-making.
5. A large size company, which is information-oriented and faces difference competencies via different departments to help generate business intelligence, and where there is a potential for wider-reaching benefits of adopting Big Data strategies (Hex-Elementization).

Interviews help accumulate primary data and information from participants through sharing practices, beliefs and opinions. The interviews were conducted through a semi-structured approach, which facilitated open discussion on what the business stakeholders think holistically about data and information in overall ability to make timely and accurate decisions. The discretionary element of semi-structured interviews help probe the issue at hand at greater depth. Decision-making is not always objective but involves a subjective element. The information regarding this subjectivity in decision-making by business stakeholders is best ascertained through a conversational approach – which is precisely what semi-structured interviews offer (Chevalier & Buckles, 2013). This approach helped the researcher delve deeper into the use of data and the flow of information from start to end. This examination helped fine-tune the foundation of Hex-Elementization.

The relationship of the research to the interviews serves two purposes. The first is understanding how the data and information flows to the stakeholders in the form of business intelligence. Second is examining the lag between data accumulation, segregation, integration, aggregation and correlation of the information before it becomes useful to the business stakeholders.

These interviews were initially conducted face-to-face to enable a proper understanding of the organisation and the importance of business intelligence to the firm in general. Interviews in this manner are comprehensive and tend to be dominated by open-ended questions. The questions were a combination of descriptive questions (highlighting areas not considered explicitly), structural questions (understanding relationship between things or groups similar context) and contrast questions (exploring terms discussed during interviews) (Knowles, 2007).

Additional granular information was collected via web-survey to be more efficient and less intrusive than mail surveys or phone calls. Anonymity was upheld in these web surveys so as not to create any ethical issues. Once the interview materials are collated, the next step is to transcribe the material for further investigation and analysis. Verification of the information collected through interviews becomes essential to ascertain the validity of interview findings.

Prior to verification, a mathematical framework is established to test the sample information collected through interviews and how it can help validate the concept of Hex-Elementization.

Phase 2: Model building

The next stage would encompass formalising the Hex-E model theoretically. The basis of the theoretical framework factors in the lessons learned through the interviews (qualitative phase) and literature review. Although the concept of Hex-E, as conceived, serves as a larger perspective to this research, the additional insights gained through interviews helped refine the model. This phase involved going to the root of the concept of Hex-E to help understand if any missing factors of integration should be incorporated to further improve the model.

The insights gained from interviews provided a practical viewpoint for the concept of business intelligence created in real-time to support accurate and real-time decisions. The insights helped determine if the optimal solution space is steering away from the conceptual base of the model. Another consideration is extending the single dimension of the hex-element to a three-dimensional (3D) hex-element to support more detailed purpose and effect of decision-making to the entire organisation. For example, a decision to lock-in long term raw materials supply does not just affect production but also affects future profitability (finances) of the company if raw material prices fall. These additional details gained from interviews (phase 1) helped create and formalise the model of Hex-Elementization, in Phase 2.

Phase 3: Validation

The next phase of the research involved validating the model based on the results of the qualitative and quantitative analysis. The quantitative analysis is performed on the data collected through the survey, while the qualitative analysis is based on the semi-structured interviews. The focus of this research is not the development of a new business intelligence theory, but whether the flow of business intelligence to the ultimate business user is optimal. Hex-Elementization concept is explored via methodologies explained earlier on whether the new protocol it suggests will be more useful than the existing process flow from data to business intelligence. Given the need and use of real-time business intelligence is not commonplace and cannot be validated just through quantitative data analysis, qualitative analysis can provide more in-depth and experimental insights (Mustafa, 2008) into the application of Hex-E in a business environment. The aim of combining the finding of qualitative analysis with the insights gained from the quantitative analysis is to show the applicability of the Hex-E concept in business uses.

A Method to Convert Subjective Survey Responses to Quantitative Measure

In this study, the quantitative investigation of issues is accomplished through survey responses. The data analysis or quantitative assessment of the survey responses is divided into two parts. The first part requires a high level of knowledge of the participants. This initial part helped summarise the attributes of the participants. It is essential to understand the attributes of the participants because it sets the tone of the responses relative to the issues encountered in day-to-day life.

The second part of the data analysis entails quantifying the qualitative responses provided by the participants. This quantification of the qualitative responses enables statistical analysis and pattern recognition. The quantification also helps standardise responses across similar areas of interest within the survey (Standardisation). For example, a group of questions which highlight similar kind of issues from different perspectives can be grouped to understand if there was a consensus among the common issues. This process is called normalisation. The normalisation and standardisation of the survey responses and the subsequent quantification helps in performing trend analysis. The trend analysis was conducted through regression studies. Complex questions, with qualitative responses, were joined together in a single linear dependent dataset. The independent time series of data was based on the quantification of the seniority of the participants. Both the dependent and independent variables of the regression analysis are used to plot the data series to calculate the trend line. The coefficient of regression is then used to ascertain whether there is a good fit. In other words, this measure of regression attempts to highlight whether issues discussed in the survey and the severity increases with seniority. The trend line and the regression coefficient help quantify the relationship.

Standardisation of data

Data standardisation is important in conducting advanced statistical analysis. It is a critical step in analysing data sets which are not in the numerical format or not quantised. The aim of standardisation is manifold. Standardisation not only helps convert qualitative data into a quantitative measure but also helps compare data across multiple questions and responses.

During the data collection phase, participants based on the criteria listed in Chapter 4 responded to approximately 25 questions. Some questions demand a Boolean answer (Yes or No), while others are based on a 5-point scale or 7-point scale, ranging from the most positive to the most negative response. The challenge is to compare the responses across a few questions which strike a common theme. Furthermore, standardisation also helps aggregate data further in the analysis phase. The method of aggregation is explained in detail under the next topic of Normalisation.

There is no standard method to standardise. Standardisation method is based on the values of the data, the goal of the standardisation and what needs to be achieved once the data is standardised. To enable comparison across a few questions in the research, a scale of values from 1 to X, in increments of 1, is used. This scaling is based on the premise that more negative responses get a higher value. For example, survey response scale from **Strongly Agree, Agree, Not Sure, Do not Agree, Strongly Disagree** are assigned numbers from 1 to 5, with one being **Strongly Agree**. All the responses are quantified in numerical values, in which negative responses across questions are standardised with a higher value and positive responses have a lower value. This normalisation eases the transition from qualitative input to quantitative insights.

Normalisation of data

Data normalisation is discussed in detail in the various annals of data management. However, the following discussion differs from the normalisation process in databases. A standard database normalisation entails re-structuring and re-organising data to make the database relational, adding in flexibility and eliminating redundancy and inconsistent dependency (Parent & Spaccapietra, 2000). This process of database normalisation is then defined by systematic decomposition of tables into granular datasets known as normal forms. The term normalisation refers to the normalisation applied in statistical studies, in which data in different formats, measures, and units are brought together for comparison and more importantly, aggregation. So, in essence, this section explains data normalisation from a statistical point of view rather than a database point of view.

The data which needs to be analysed in the later part of this research are the survey responses, which vary in format and scale. In the section preceding this, the discussion about standardising the data (from survey responses) was discussed. While standardisation quantifies data from a standard scale of good to bad, normalisation shifts and scales variables to compare corresponding normalised values across different datasets. These normalised values help eliminate the effects of gross influences (De Oliveira & Levkowitz, 2003), similar to an anomalous time series (Rebbapragada, & Protopapas, 2009). The process of normalisation is also known as Feature Scaling. In the area of machine learning, objective functions often do not work correctly without appropriate normalisation. This stands true, particularly for the majority of classification algorithms where it is pivotal to calculate the distance of each data point by the Euclidean distance (Sathianwiriyakhun, Janyalikit, & Ratanamahatana, 2016). In stochastic gradient descent approaches, which is an iterative method for optimising an objective function, normalisation of data helps improve the convergence speed of the machine learning algorithm. In this research, normalisation helped scale the responses in order to assist in aggregation. Few of the issues during the survey generated extreme responses, either positive or negative. The goal of the data analysis later in the research is to merge the negative responses collected from the participants in order to understand the effect of the issues discussed during the survey on their decision-making.

Each issue affecting participant decision-making adversely is normalised in the first stage on a scale between 1 to 100, in which 100 would mean the most severe or negative response.

Once each issue (responses from questions) is normalised to this scale (1-100), a few questions can be aggregated together by summing the response values.

Original Response	Standardise	Normalise
Strongly Agree	1	1
Agree	2	25
Not Sure	3	50
Do Not Agree	4	75
Strongly Disagree	5	100

Table 4: Sample Response for a given Question A from Respondent AA

Original Response	**Standardise**	**Normalise**
Yes	1	1
Maybe	3	50
No	5	100

Table 5: Sample Response for a given Question B from Respondent AA

Question A: Normalised Response	**Question B: Normalised Response**	**Aggregated Response: (Sum of A + B)**
1	**1**	**2**
25	**0**	**25**
50	**50**	**100**
75	**-**	**75**
100	**100**	**200**

Table 6: Sample Aggregation of the two questions from Respondent AA

As illustrated in Tables 4, 5 and 6, the first two responses from a sample survey question are standardised and then normalised. In the last step, a sample aggregation is shown by summing up the normalised scores from the two sample questions. The aggregation is summed from the normalised values from the two questions. There are no rules governing the aggregation method. The normalised responses can be based on average, median or any other statistical measure without any bias to outliers at the same time applying the same morphing techniques consistently to all the responses.

A method of finding objective patterns using Qualitative Analysis

To supplement the quantitative analysis of the survey data, described in the previous section, a qualitative study was conducted. This qualitative study helped establish the concept, views and issues surrounding the topic of the research. The survey helped in gathering closed-ended responses, which could be quantitatively analysed. However, qualitative analysis of open-ended interviews was conducted to get deeper into issues of incorporating data into decisions. The interviews also helped understand if the proposed concept of Hex-Elementization provides a more profound understanding of how business intelligence is being utilised by practitioners. Qualitative research is suitable when the research entails an investigation into a new field of study or intends to establish and theorise prominent issues (Corbin, Strauss, & Straus, 2014). Qualitative methods, in essence, enables the researcher to have an in-depth and extensive understanding of the issues by employing textual interpretation; the most common types of qualitative methods are interviewing and observation (Creswell, 2012). Hence, in order to meet the objectives of this study, semi-structured interviews were conducted.

Semi-structure Interviews

Semi-structured interviews form a branch of qualitative research conducted through intense and prolonged contact with participants in a naturalistic setting to investigate the everyday lives of individuals, groups, societies, and organisations (Agosto & Hughes-Hassell, 2005). The researcher's role is to gain a holistic (systemic, encompassing, and integrated) overview of the context under study (Mertens, 2014). Interviews are usually conducted with relatively little standardised instrumentation as the researcher is essentially the main instrument in the study (Kvale, 2008).

The major benefit of interviews, as a method of research, is compatibility with several methods of data analysis, including discourse analysis, grounded theory and interpretive phenomenology. Interviews are easy to arrange and conduct, and include a variety of methods. Unstructured interviews have more elements of a conversation than structured interviews and are most often than not skewed towards the interest of the interviewer (Rubin & Rubin, 2011). Non-directive interviews, which is a form of unstructured interview, do not have a pre-prepared set of questions. Focus interviews are conducted by interviewers who know the interviewee very well (Merton & Kendall, 1946). At times when the focus of the respondent strays away from the topic at hand, the interviewer brings the focus back to the core of the issue. Informal interviews are also unstructured interviews where questions usually emerge during the conversation (Legard, Keegan, & Ward, 2003).

In comparison, semi-structured interviews are in-depth interviews in which the respondents answer pre-formulated open-ended questions. Semi-structured interviews are widely employed by different researchers who tend to gain a more open response to their research.

These interviews provide freedom of thought and opinion and do not restrict thought-process within any frames (Corbin, Strauss, & Straus, 2014). Semi-structured, in-depth interviews are utilised extensively as an interviewing format with an individual and sometimes a group (Corbin, Strauss, & Straus, 2014). The benefit of this type of interview is that it only needs to be conducted once, with an individual or with a group, and generally lasts 30 minutes to an hour (DiCicco-Bloom & Crabtree, 2006). For optimum use of interview time, interview guides serve the purpose of exploring many respondents more systematically and comprehensively as well as keeping the interview focused on the desired line of action (DiCicco-Bloom & Crabtree, 2006). The interview process comprises core questions and many associated subordinate questions. In order to have the interview data captured more effectively, recording the interviews is considered an appropriate choice. Handwritten notes during the interview help form central themes for each question being asked to the respondents. These themes can then help form the coding (keywords) which are further used to link responses across all the respondents on the same theme. Recording the interview makes it easier for the researcher to focus on the interview content and verbal prompts. An automated-transcription software can further supplement the recorded responses to generate a verbatim transcript of the interview. The automatic transcription through an app on the smartphone helps save monetary resources spent on human-driven transcription. The automated transcription is an efficient process which helps the researcher analyse the data encoded in the transcription of the interviews to start drawing insights.

Rigorous data analysis using software (NVivo)

Mixed methods research design typically is comparative, convergent or sequential, with the priority given to one-type (descriptive) of data (Clark & Creswell, 2008). There have been no explicit methods available to integrate data from mixed method research, and it mainly left at the discretion of the researcher's creativity to extract patterns from descriptive data (Andrew, Salamonson, & Halcomb, 2008). Developments made in computer science have enabled the fusion of qualitative and quantitative analysis into a single dataset (Bazeley, 2013). Software, such as NVivo, termed as Computer Assisted Qualitative Data Analysis System (CAQDAS), enables researchers to integrate and consolidate data from various quantitative methods to help create a rich repository of data which can be analysed to find patterns and validate the research (Reis, Costa, & de Souza, 2016). This ability of data triangulation across various descriptive qualitative methods, including interviews, case studies, journal exhibits, and other unstructured data, is the most significant benefit of a software (NVivo). Moreover, using a systematic way of identifying explicit and implicit themes through analytical software like NVivo brings objectivity to the analysis. The data analysed in this manner is dependable, as the subjectivity of the researcher is systematically removed by connecting responses from individuals (in the form of an interview), and other unstructured opinions gathered as a part of the research, to draw findings (Hancock & Algozzine, 2016). Although there are many software packages available to analyse qualitative datasets, the primary reason of choice should be the research design, research goals of the researcher and the topic (Cameron, et al., 2007). The software chosen for this project was NVivo version 12 because of its ability to import qualitative data captured during the interviews and integrate it tightly with the quantitative data captured through the survey.

Inductive coding interview responses and structuring themes

The primary analytical challenge for researchers conducting qualitative analysis is finding coherent descriptions and explanations that present gaps, inconsistencies, and contradictions inherent in personal and social life (Sayer, 1992). The researcher can jeopardise the outcomes of qualitative analysis by forcing logic, structure or order, and the plausibility that leads to a theory which could be grounded on the uneven, sometimes random, nature of social life.

In this research, after the one-on-one in-depth interviews were conducted and transcribed, the raw text was brought into Microsoft Word for further processing, including correcting spelling mistakes and grammatically incorrect sentences. Once the text was processed, the text was sense-checked again matching the audio tapes to the text. The responses were aligned with the speakers and appropriately differentiated among the interviewer (researcher) and the interviewee (respondent). The text was then imported into NVivo software and grouped by each respondent. Next, a series of steps were undertaken in NVivo to derive common themes across all participants. Data analysis began with open coding, where categories were formed about the phenomena by segmenting information (Creswell & Tashakkori, 2007). This process included assigning codes or themes to a set of field notes and interview transcripts. The stage of coding, in which themes were identified and related to interview responses, allowed comparison of data and identification of common problems and issues (Charmaz, 2006). Coding itself went through multiple phases to help sift through themes which might have been missed through the random coding exercise. After the initial coding of broad view (big picture) themes, the second stage deployed Focused coding, whereby the codes were more directed, selective and conceptual than in the first stage (Charmaz, 2006). Focused coding used the most significant and frequently occurring codes to sift through the vast amounts of data to determine each code's suitability for use (Charmaz, 2006). The third stage of coding involved conducting Axial coding, whereby the data was assembled in new ways (Corbin & Strauss, 1990). Axial coding identified central phenomenon, explored interrelationships, identified strategies, classified context, and defined the most analytical way to categorise the phenomenon (Creswell, 2012). The fourth and final stage of coding involved in theoretical coding. Theoretical coding enabled building final relationships between all categories identified by the earlier coding to help supplement the hypothesis in the centre of the thesis (Charmaz, 2006).

The next step involved sorting through the coded transcripts in order to identify common phrases, relationships between variables, themes, categories, distinct differences between subgroups, and common sequences, in order to find patterns. Furthermore, this pre-processing phase involved isolating these patterns and processes, and commonalities and differences, and perform data analysis by tying the isolated elements back to the quantitative data analysis. The reflections or remarks noted down during the interview process were referenced to leverage the analytic memo into overarching themes. The next step involved gradually elaborating a small set of assertions, propositions and generalisations that cover the consistencies discerned in the database. Furthermore, these generalisations were compared with a formalised body of knowledge in the form of constructs or theories, specifically in the findings highlighted by the quantitative analysis.

The quantity of data collected during the qualitative analysis process can be overwhelming. The entire range of activities of selecting, focusing, abstracting, simplifying and transforming data from the qualitative data collected, including the field notes, interview transcripts, documents and other unstructured materials, is known as data condensation (Malterud, 2012). Data becomes stronger and robust as it goes through the condensing process.

Moreover, the use of a computer-aided method to analyse the qualitative data makes the data robust for further analysis and drawing conclusions.

Objectivity, accuracy, transparency, dependability

Qualitative research is often criticised because it does not serve evidence-based practice well. There are some who argue that much of the qualitative research is of a poor standard (Hammersley, 2007). Typically, the criticism is around the fact that qualitative studies are often subjectively designed and lack a clearly defined set of quality criteria to judge the study and hence is of uncertain quality. There are two assumptions driving these criticisms that need to be addressed. First, it is assumed that clearly defined criteria of quality are already available for quantitative research. The second assumption is that explicit assessment criteria are needed. There are also beliefs that unless researchers operate by such criteria, their research is of poor quality, and that users of research require some reliable means of judging its quality. A set of criteria would meet this need. Steiner Kvale aptly highlights the most common ten criticisms (Kvale, 1994) as listed below:

1. Is not scientific, but only common sense.
2. Is not objective, but subjective.
3. Is not trustworthy, but biased.
4. Is not reliable, but rests upon leading.
5. Is not intersubjective; different interpreters find different meanings.
6. Is not a formalised method; it is too person-dependent.
7. Is not scientific hypothesis-testing; it is only explorative.
8. Is not quantitative, only qualitative.
9. Is not yielding generalizable results; there are too few subjects.

10. Is not valid, but rests on subjective impressions.

To address the limitations of the qualitative analysis highlighted earlier, computer software was used to assist in data triangulation. The data triangulation brought objectivity, accuracy, transferability and dependability to this research project. Data triangulation refers to the use of several methods and different data sources in qualitative research to formulate a comprehensive understanding of the hypothesis at the centre of the research (Patton, 1999). By converging information from different sources, data triangulation has also been viewed as a qualitative research strategy to test validity. As Norman Denzin said, “the greater the triangulation, the greater the confidence in the observed findings” (Denzin, 1970). There are four main methods of triangulation as postulated by (Denzin, 1970) and (Patton, 1999): a) method triangulation, b) investigator triangulation, c) theory triangulation, and d) data source triangulation.

This research utilises the method of data triangulation to conduct a qualitative analysis. Data triangulation uses a variety of data sources, including time, space and persons, in a study (Hales, 2010). Data triangulation is a robust method because findings can be corroborated, and any weaknesses in the data can be compensated for by the strengths of other data. Data triangulation helps increase the validity and reliability of results. This approach of data triangulation as a qualitative research methodology has been used in many sectors to strengthen conclusions about findings and to reduce the risk of false interpretations. The use of computer software, NVivo, in performing this data triangulation brought objectivity to this research method. The subjective elements of a researcher’s line of inquiry and guiding the response from the participants has also been the centre of the criticism of qualitative methods. However, the use of computer software to analyse all the text (transcripts) across all the responses to find common threads (nodes, themes) provides a level of transparency to the entire process, reducing subjectivity and providing more objectivity.

Ethical Considerations In This Research

Without the insights from the participants of the interview process, it is hard to conduct a proper qualitative analysis. The ability to influence others by providing the needed data is the basis of the success of a research initiative (Kahaner, 1996). Unlike research related to sociology, medicine or political science, in which ethical standards are required to make interviews easy and less stressful for participants, this research did not cause duress of any kind to its participants. This lack of duress is because the research proposes a new concept that is exploratory, and the participant views of the underlying issues were pivotal for this research. However, high ethical standards were observed and implemented when conducting this research. Ethical clearance (Appendix 2) was requested and obtained from the University of Western Sydney (UWS) Human Research Ethics Committee well before the actual interviews were conducted. Every participant in the interview was made fully aware of the study and the benefits of participating in it. Three documents were sent in advance to each participant including Participant Information Sheet (Appendix 3), Consent Form (Appendix 4) and the interview questions (Appendix 5). A brief research proposal outlining the project details was included in the Participant Information Sheet. The Consent Form included the confidentiality agreement and signed by each participant. This provided a guarantee that no raw data would be made public, and the participant remains anonymous. The goal was to derive common themes across all the participant responses, and once this was accomplished, all links to the participant, directly or indirectly, were stripped from the research and associated data.

Chapter 4: DATA COLLECTION AND ANALYSIS3

Introduction

Data collection in the research project was conducted via two separate processes. The first set of data was collected from individuals, from various industries and business segments, who use data in a variety of formats to assist with making business decisions. A web-based survey was used in this stage to quickly collate the thoughts of data-savvy individuals, which would eventually help fine-tune the Hex-E framework at a later stage. This first stage also set a precedent for the second stage of detailed one-to-one interviews to help understand the detailed use of data from a selected group of individuals from the stage 1 survey. Both these stages included the collection of valuable data and helped gain insight into how professionals use and perceive data in their day-to-day decision-making and fine-tuned the model to factor in variables which were not considered before the data collection.

Data Collection Process

Quantitative Method

Quantitative: Collection Method (Survey)

The first stage of this research involved assessing the current state of business intelligence in the context of decision-making. An online survey for the first stage of exploration was less time-consuming than traditional means of collecting data. The benefit of the digital data collection is that it provides access to individuals and corporations across geographical boundaries and diverse businesses which would have otherwise impossible to gain access using traditional method (Garton, Haythornthwaite, & Wellman, 1999). As postulated by Wellman, almost three decades before social networks became prominent, the Internet connected like-minded people, but with different ideas and interests, to share and educate others (Wellman, 1997). Hence, the decision was made to approach individuals from different organisations, involved in a variety of fields, and at diverse organisation levels, to share their view on how important data is in making business decisions in this fast-changing environment. Surveys conducted online also saves time for the researcher to reach to a diverse and geographically dispersed set of individuals while being able to concentrate on other research activities (Llieva, Baron, & Healey, 2002). Online surveys are also cost effective as it saves the researcher the cost of paper, printing, data-entry, cost of equipment, and travel (Watt, 1999).

One of the disadvantages often associated with online surveys is the issue with sampling (Howard, Rainie, & Jones, 2001). It is postulated in many research papers that this method of collecting data limits the researcher's understanding and the characteristics of the participants beyond necessary demographic information (Stanton, 1998). Another issue with conducting online surveys is the presence of lurkers in the massive open online communities which can skew the results of the survey by drifting away from the domain knowledge which is required to provide an informed opinion in the surveys (Couper, 2000). The third, and most important issue arising from conducting online surveys is inherited from behavioural economics and is called framing. Framing is a systematic bias in which an individual frame of references can influence the participants in a biased manner (Thaler, 2017). Each one of these concerns was carefully taken into consideration when selecting the participants in the first stage of the data collection (survey). How this research addressed each of these concerns is enumerated in the next section, Selection of Participants.

Qualtrics was chosen as the survey provider, given its breadth and depth of reach in the Internet survey space. Qualtrics has been adopted as the default Internet survey provider across more than 100 universities and many international corporations. Qualtrics provides a rich set of question types, and an intuitive interface to easily create an Internet-based survey. Another benefit of Qualtrics is advanced data crunching methods built-into the same infrastructure, making it easy for researchers to perform a large amount of data analysis. There is less need to export the data out of the system unless one needs to perform statistical analysis outside the scope of the study.

Quantitative: Selection of Participants

The research tries to explore the use of data in various forms, with a focus on the decision-making endeavours of business organisations. Hence, the research needed insights from individuals who are tech-savvy and data-savvy, and also from participants who envision increasing the use of data in future. Understanding how decision-makers at differing levels of the organisation perceive data being consumed during the BI process is another goal of the research. The research was not industry specific but was aimed for organisations and individuals who value data in its current form or future value. Labour-intensive organisations or individuals working in such an organisation were excluded based on an informed decision, the basis of which is explained earlier.

To address the shortfalls of conducting web-based online surveys highlighted in the previous section, a concise and precise list of individuals was created, mapping the firms, current responsibilities, type of work they perform, and the industry. Many of these individuals were selected due to their in-depth knowledge of the business and industry. Creating a concentrated list of individuals ensured that the survey had a good chance of capturing informed responses from individuals. The survey was never posted on the public domain (on social networks which are designed for personal use rather than professional use). Not publishing via social media avoided the issue of lurkers highlighted in the previous sections which would skew the results of the survey in a direction away from the desired goals (Preece, Nonnecke, & Andrews, 2004). The researcher also avoided priming the individuals by providing too much information on what this completed research aims to accomplish. Framing the opinions of the individuals could have skewed the results of the survey overwhelmingly in favour of what this research aims to demystify. Avoiding this bias from responders was essential to capture the true nature of how data is used by these individuals in day-to-day decision-making. The survey aimed to gauge if there is a gap between data collection and generating business intelligence in the current roles of these individuals. Identifying this gap was one of the primary goals of this exercise.

Quantitative: Response collection methodology

Conducting the survey online was optimal as participants were geographically dispersed and worked in various industries. Participants from four different continents and 10 different countries were selected for this targeted survey. More than 20 different web survey portals are available to use, many of which are entirely free to set up and gather survey response. These web-based survey platforms are easy to use, with an attractive and intuitive user interface. Qualtrics was selected for this research due to its link to hundreds of academic institutions, including Western Sydney University. Qualtrics supported survey responses from multiple devices across numerous operating system. Qualtrics supported advanced response types in which the researcher could gather responses in a variety of question formats. Options that include drop-down selections, sliders, and free-floating text with decision trees made it stand apart from other survey platforms. Apart from emulating the ease of use and intuitiveness of other platforms, Qualtrics supported advanced analytics. The ability to cut and splice data at the host platform (survey platform) saved valuable time.

Respondent's personal information was not recorded in the survey platform to safeguard their privacy. An anonymous link was generated for the survey, and the researcher sent the personalised, individual emails to each respondent. Survey links were emailed to 120 respondents along with the summary of the research (Project Proposal) and a consent form, with a request complete the survey within a month.

Qualitative Method of data collection

Data analysis is not only an interactive model but also an iterative model (Miles, Huberman, & Saldana, 2014). Apart from data collection, the data display, data condensation and conclusion derived from the data are cyclical, as shown in Figure 22, in which the three nodes are in constant use to derive the final output (final findings) (Miles, Huberman, & Saldana, 2014).

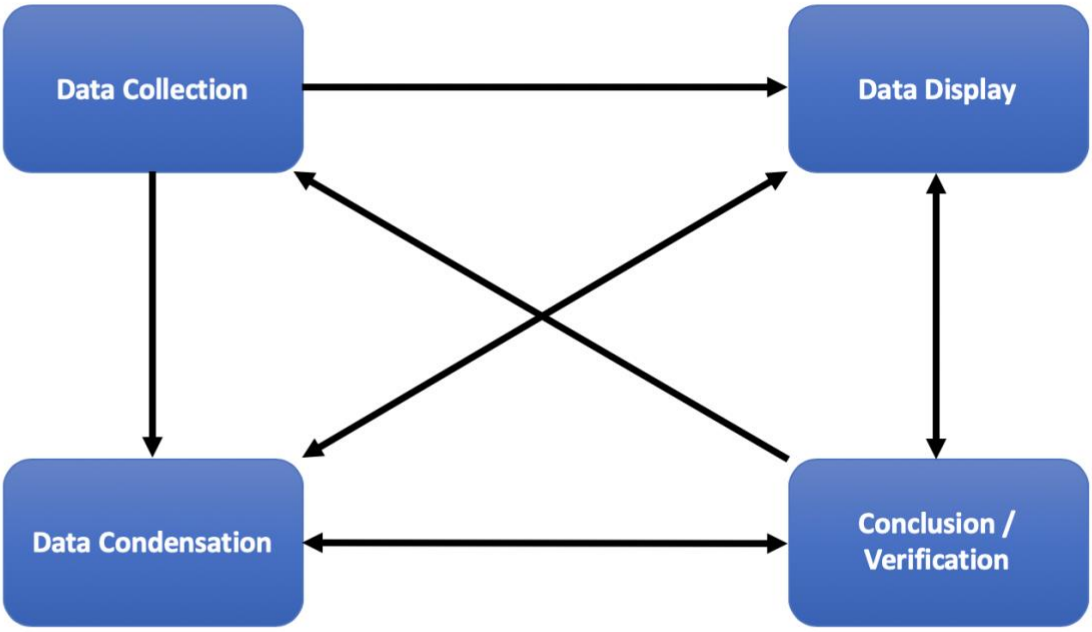

Figure 22: Components of Data Analysis: An Interactive Model *(Miles, Huberman, & Saldana, 2014)*

For instance, the process of coding data, using NVivo, is known as Data Condensation. This process leads to new ideas based on what can be visually transformed in terms of data display. Triangulating data from field notes and interview transcripts led to further data condensation, which was essential in drawing the preliminary conclusion. The conclusions can then be tested against new data sets to find gaps in the analysis. Thus, the journey of qualitative analysis followed in this research has been iterative, in which the researcher constantly tried to identify the most important themes based on the responses.

Qualitative: Interview Participants

Ten interviews were conducted based on the criteria of the individuals explained in Chapter 3 for this stage of the analysis. A conscious effort was made to achieve demographic diversification among the chosen participants. The profile of the interview participants is summarised in Table 7.

#	Type of Organisation	Business Size	No of employees	For-Profit or NGO	Role	Years of Experience	Interview Duration
1.	Private	Small	22	Profit	CEO	42	24 min
2.	Private	Small	10	Profit	CEO	40	54 min
3.	Government	Medium	600	Non-Profit	Middle-Management	15	40 min
4.	Private	Small	10	Profit	CEO	42	40 min
5.	Public	Large	85	Profit	Middle-Management	20	30 min
6.	Private	Small	2	Profit	CEO	35	28 min
7.	Private	Large	19,000	Profit	Head of Department	15	23 min
8.	Public	Large	300,000	Profit	Head of Department	30	22 min
9.	Government	Medium	110	Non-Profit	COO	10	36 min
10.	Private	Medium	1500	Profit	Head of Department	15	40 min

Table 7: Table showing participant information – including the duration of each interview

Participants were advised that the interview would take between 30 to 60 minutes. The shortest interview took 23 minutes to complete and the longest took 54 minutes. The average duration of the interviews was 33.7 minutes. Participants were selected to equally represent the private and public sector, for-profit and not-for-profit and small-to-large enterprises. Most interviewees were from middle to top level management. Having an experienced group of individuals was a conscious decision made by the researcher to gather insights from senior-level industry participants who are at the centre of decision-making. Decision-making is the most important aspect of this research, and how participants use data or a business intelligence environment is what the research is trying to validate through the qualitative analysis. The participants were diverse in terms of experiences and industries. Bias was minimised because participants were diverse in terms of experience and industry, nor did they have a high level of technical knowledge or expertise. The participants were well represented from an age perspective so that the younger, tech-hungry participants balanced the vision and wisdom of the older participants, who frequently have less technical acumen. The fact that the participants represented different industries highlighted diversity in the principles and processes which drove their decision-making, discussed in detail later in this chapter.

Qualitative: Semi-Structure Interview - Questions and Process

Semi-structured interviews are well suited in a mixed-method approach for many tasks, particularly when some of the close-ended questions (survey) require follow-up queries, which was the case in this research. This type of interviews is particularly suitable when the researcher needs to ask probing, open-ended questions in order to gain additional information from each respondent independently of the others. Semi-structured interviews are also ideally suited for conditions when there is a need to ask probing, open-ended questions on topics where sitting with peers in a focus group prohibit candid responses (Newcomer, Hatry, & Wholey, 2015). This approach is also valuable when conducting a formative program evaluation, and one-on-one interviews are needed with key program managers, staff, and front-line service providers. This is especially important when the researcher is examining uncharted territory with unknown but potential momentous issues and the interviewers or researcher needs maximum latitude to spot and pursue useful leads (Newcomer, Hatry, & Wholey, 2015). With this in mind, the interview participants were approached well in advance via email asking for a time and date for the interview. Each participant was sent an email invite with Participant Information Sheet, Consent Form and Interview Questions. On the day of each interview, following the code of conduct described in the ethics approval process and the participant information sheet, participants were asked if they granted explicit permission to record the interview. Each interview audio was recorded on an iPad and on an iPhone, which was running an automated transcription software. This transcription (speech recognition) software converted audio into text with 85% accuracy. The transcription in text format was encrypted to secure the content. Each participant was asked the seven questions listed below.

1. What is the importance of data in your decision-making as compared with your (and your employees') experience and intuition?
2. Is your organisation actively increasing its data capture? Please provide examples of new types – e.g. unstructured and various sources of data.
3. What are the key challenges you face in deriving business insights from data transformation? E.g., volume, the variety of data.
4. What are the key challenges in undertaking Data-driven decisions? Are business decisions based on lagged or outdated data? Does context get lost?
5. What percentage of your organisation's time is spent regularly in cleaning (pre-processing) and linking data from disparate sources before it can be consumed for decision-making?
6. Do you think the cost of the journey from data-to-intelligence has increased overtime? Cost in terms of time (to clean, understand, contextual data), the cost in terms of capital (capital to acquire new data science tools, R&D, high-performance computing, specialised labour) and labour (i.e. the need for specific skill sets like data scientists, new data science specialists, analysts?
7. This research is proposing a Hex-E business intelligence framework that will enable linking of data with inbuilt intelligence; What do you think would be the benefits and risks of such a framework to a business? (especially in the light of the issues you highlighted in previous questions concerning challenges of extracting insights from data).

During the interview session, the researcher took field notes including important and thematic keywords used by the participants during each question. These notes and keywords served as the stepping-stone to either drill down into sub-questions or summarise and identify trends which can be later used to get additional detail from the interviewees. The notes and keywords also served the purpose of quasi-bookmarks in the audio transcripts. The same process was used to conduct all 10 interviews which provided a large set of qualitative data for the researcher to analyse across seven distinct questions.

Qualitative: Data Preparation

The data from this semi-structured interview process came in the form of the transcription, which was generated by Otter speech recognition software (Otter, 2018). The transcription was then swept to remove any spelling or grammatical mistakes, either due to noise interference during the recording or the lag in the auto-transcription service. The researcher meticulously went through the entire transcription text in tandem with the audio recording to find any errors. The text was appropriately aligned to each speaker (i.e., interviewer vs. interviewee) to differentiate the participants' responses. The text was then imported into a Microsoft Word document to identify spelling and grammatical errors. Names and personal details were removed from the transcript, including the organisation name to avoid revealing the identity of the participants either directly or indirectly.

Qualitative: Coding and Data Condensation

The method of indexing and categorising text gathered during the semi-structured interview process in order to derive themes, patterns and common-thread, is known as coding (Gibbs, 2018). Coding is conducted with the most commonly occurring keywords which are then grouped into themes. For this research, the process of coding was carried out using the NVivo 12, which is computer-aided qualitative research software. At the end of the interview process, the entire text generated was close to 80 pages. The analysis of large amounts of such unstructured or qualitative data can be painstaking and time-consuming. Software packages such as NVivo helped handle this task efficiently. The software helped to "organise, analyse, and find insights in unstructured, or qualitative data" (QSR, 2019). The use of computer-assisted qualitative data analysis software (CAQDAS) has become widespread in recent years and is valued by many qualitative researchers for its ability to organise and categorise sources based on relevant and essential characteristics. The benefits include categorising data by themes, ease of searching, retrieving relevant and associated material, and developing visual representations of the data (Talanquer, 2014). Using software also helps improve the accuracy of results due to the systematic coding and analysis process (Lee & Esterhuizen, 2000). Additionally, its ability to create an audit trail of the data analysis process can be used by others in gauging the methodological standards of the study (Welsh, 2002).

The process of coding used in this research was based on a combination of selective coding and thematic analysis. In broad terms, these elements also relate to inductive and deductive approaches to coding. Deductive coding involves assigning research material to predefined categories related to a theoretical framework or specific research questions and is generally used for testing existing theories or examining applicability to a particular context (Zhang & Wildemuth, 2009). The use of established themes or codes assists in comparing similar studies conducted in different settings. Deductive coding helps categorise interview responses by classifying the perceived common issues when using data to make business decisions. Deductive coding also helped highlight the most common issues faced by respondents in countering the use of data in decision-making. Inductive coding is most often used in research conducted within the interpretivist/constructivist theoretical paradigm, which focuses on understanding phenomena from the perspective of individual realities. This approach played an essential part in addressing a few of the research questions which are discussed later in the chapter.

Qualitative: Visualisation to gain insights

Finally, appropriate visualisation of the large unstructured data set was important to derive insights from the qualitative data. Data visualisation is viewed in the context of quantitative data represented through a variety of visualisation methods such as graphs, bar charts, scatterplots, tables, pie charts, and smart visualisations. Visualisation helps the researcher synthesise and coalesce findings in an easily interpretable manner. Data visualisation of qualitative research has received minimal attention to date (Mikalef, Pappas, Krogstie, & Giannakos, 2018). One of the first teams to describe the importance of data visualisation methods was Miles and Huberman (1994) by demonstrating the effectiveness of using visual tools such as matrices and network displays in qualitative research dissemination. CAQDAS (NVivo) and other qualitative research tools have also played an essential role in making it possible to visualise qualitative data. For example, tables and matrices, network mapping, word clouds, and word trees have become more common in qualitative dissemination efforts (Henderson & Segal, 2013). These analytic display methods are useful to summarise emerging theories from the data and primary thematic results (Verdinelli & Scagnoli, 2013).

Visualisation was used to summarise the analysis using qualitative data collected post semi-structured interviews. The visualisation used in this research supplemented the themes that emerged from the interview analysis. Moreover, this research, which is based on mixed methods, is rich with visualisation in the quantitative analysis section (detailed in the next section). Hence, selective visualisation options were used in the qualitative section, helping the researcher objectively identify the most overarching themes.

Data Analysis

Part 1: Quantitative Analysis of survey data

Introduction

The survey conducted in this research provides an opportunity to understand the research hypothesis from a group of participants with varied backgrounds and experience. In this section of the research, quantitative analysis is conducted on the survey responses. This quantitative data analysis is divided into three parts. The first part contains a section in which survey participant characteristics are validated and illustrate the diversity within the group of survey participants. This validation also helps understand the context in which participants might have provided survey responses. Characteristics are quantified from geographical diversity, represented sectors or industries, experience level and management experience. All these features are single dimensions which help validate the relevance to this research. The analysis shows the breakdown of the respondent by each of these characteristics.

The second part of the data analysis involves analysing more than one dimension of the data. In this section, more than one feature or variable is combined with another to provide additional information and a higher level of understanding. For example, participant experience level or management level is taken into consideration to provide context around the lag in receiving data to make business decisions. This step of the data analysis is essential as it highlights how individuals with different characteristics and backgrounds have different perceptions of the same issue. Urgency and relevancy of issues may change based on role in the organisation or experience level. This entire process of overlaying issues (features) from the survey is called multi-dimensional analysis and is used to perform cross-sectional data analysis. The multi-dimensional analysis helped understand the survey results respective to the attributes of the survey participants.

The third part entails grouping issues together based on the commonality using Principal Component Analysis (PCA), which is followed by regression analysis. The PCA analysis helps in understanding which factors within the data are more important to the analysis. The regression analysis then takes the factors highlighted by PCA to conduct a statistical analysis of the relationship of the factors to the responses provided by the participants in order to find patterns in data.

Geographical diversity

Former US President Barack Obama said “The study of geography is about more than just memorising places on a map. It is about understanding the complexity of our world, appreciating the diversity of cultures that exists across continents. So, in the end, it is about using all that knowledge to help bridge divides and bring people together.” (Obama, 2018). From a business perspective, geographical diversity is essential to understand and incorporate into the decision-making process. Businesses around the world evolve at varying speeds, creating potentially different growth phases for each. Each growth phase brings challenges and issues which help organisations adapt and innovate. This process of adaptation and innovation helps design and formulate future decisions.

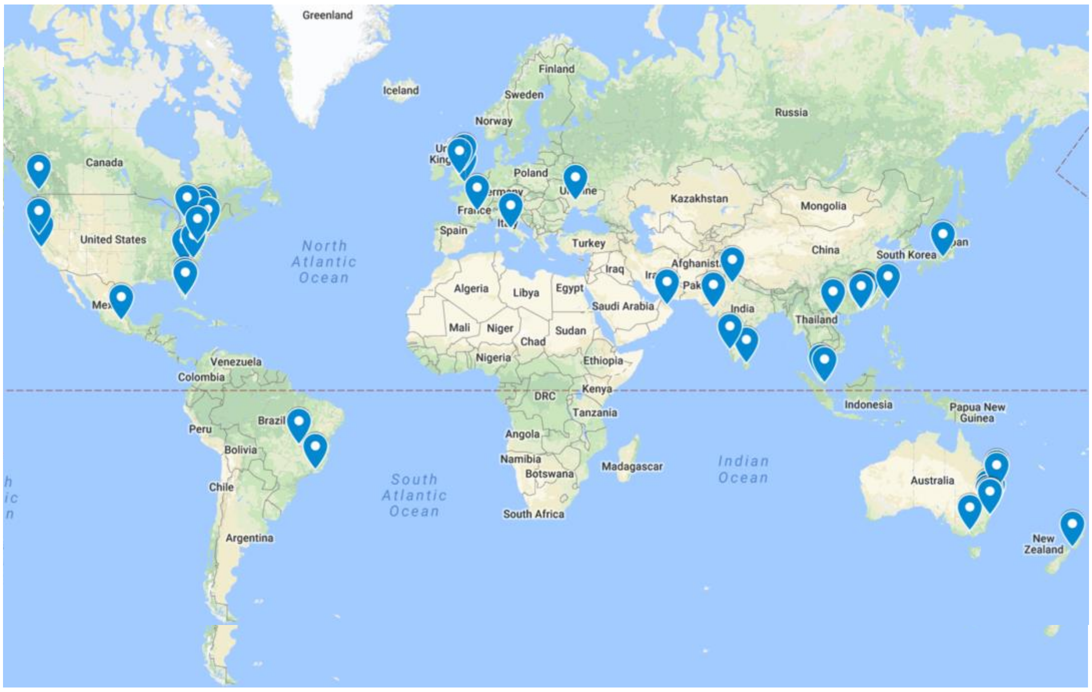

Figure 23: Geographical diversity of the survey respondents (courtesy Google Maps)

Keeping this important but often forgotten aspect of decision-making in mind, the survey was sent to a group of individuals who are geographically dispersed. As shown in Figure 23, individuals selected hailed from many countries across a few continents. This diversity brought not only varied experience, but also different insights on how data use in decision-making has evolved over the years, especially in the context of globalisation.

Information diversity is vital to decision-making. Geographical diversity brings diverse experiences to the table. These diverse experiences bring new information on decision-making, especially when considering globalisation and how information technology can blur the line between geographical divisions. This diversity brings a unique perspective from participants who have seen the effect of the globalisation and the adoption of information technology in unique ways within a particular region. Although the underlying assumptions and technology base is uniformly shared among all, adoption, implementation, use and application of information technology differ by country. Consider how e-commerce, which initially started in the United States spread to emerging markets, had to evolve to provide a bigger scale of operation (population) and address the demographic make-up of the country (younger generation). Implementing these technologies is challenging in the countries where the population is equally dispersed among the urban and rural population. These demographic challenges introduce complexity in making decisions which are not only more significant in scale but also applicable to the general population.

Hence, getting opinions of individuals who are geographically dispersed was necessary to capture perceptions of how data is used in decision-making and the form in which data is easily consumed for various decision support systems within an organisation.

Sector/Industry Diversity

The survey conducted for this research received responses from 50 participants worldwide. The participants were given an option of 22 sectors recognised as standard sectors by the Australian Securities & Investments Commission (ASIC). The survey attracted responses from individuals representing 19 of these 22 sectors, which shows broad sectoral diversity among the participants, as highlighted in Table 8.

Industry	Upper Management (e.g. CEO, CIO)	Lower Management (e.g. Support, Tech)	Middle Management (e.g. Department Manager)	Total
Agriculture, Forestry and Fishing	0	0	0	**0**
Accommodation and Food Services	0	0	0	**0**
Retail Trade	1	0	0	**1**
Professional, Scientific and Technical Services	1	0	0	**1**
Manufacturing	0	0	1	**1**
Electricity, Gas, Water and Waste Services	0	1	0	**1**
Rental, Hiring and Real Estate Services	2	2	0	**4**
Financial and Insurance Services	2	2	6	**10**
Construction	0	0	1	**1**
Information Media and Telecommunications	1	3	2	**6**
Wholesale Trade	0	0	1	**1**
Transport, Postal and Warehousing	0	1	1	**2**
Other	0	1	0	**1**
Mining	0	0	0	**0**
Administrative and Support Services	1	1	1	**3**
Public Administration and Safety	0	1	1	**2**
Education and Training	1	2	1	**4**
Health Care and Social Assistance	3	2	1	**6**
Arts and Recreation Services	1	0	1	**2**
Other Services (e.g. Consulting)	2	2	0	**4**
Total	**15**	**18**	**17**	**50**

Table 8: Breakdown of survey participants by industry and experience level

Broad sectoral representation brings diverse experience from the participants in applying the concepts under discussion in this research. Each industry faces challenges when it comes to products manufactured, services produced, or customers served. The operational processes within each of these industries vary in nature as do the decisions surrounding the variables of input, output and process. Within the same sector, businesses have varied operating structures, which define and drive the fundamentals of decisions. For example, within the technology sector, there are many products and services, and each tech business operates in unique ways. Each sector has its own KPIs to monitor and report, which helps in making decisions. Business intelligence differs by sector, and by companies within each sector, depending on what is important to monitor, report and act upon. Table 8 highlights a heavy tilt toward financial and insurance services. This is due to the diversity within this broad sector, which includes sub-sectors such as consumer banking, commercial insurance, investment banking, commercial banking and accounting services. Each of these sub-sectors is large and have illustrated different needs related to data for decision-making.

Cross-sectional statistical analysis of this survey, from a sector difference perspective, provided intriguing insights. The intrigue was in how unique sector issues were in data usage for decision-making. Despite the difference in the way each sector demands resources, the participants unanimously agreed on issues in the way data is consumed. Sectoral differences highlighted the perception that data is often difficult to consume in an optimal manner to make tactical or real-time decisions. Although many participants highlighted sector-specific issues when dealing with data, the common denominator was the multiple ways data is presented before it turns into an actionable insight.

The survey was effective in capturing diverging views of participants from various sectors and how data-to-insights takes a different course but presents common challenges in each sector.

Experience breakdown

The 50 respondents to the survey had varying degrees of experience within their profession, with over 300 years of experience among the group. The average years of experience was 12.7 with a median of 12. The average number of years of experience is greater than the median, highlighting that the survey attracted more senior respondents. As shown in the breakdown in Figure 24, 41% of respondents had more than 15 years of experience, while 26% of the respondents had work experience between 10-15 years. It can be surmised that 67% or two-thirds of the respondents had more than 10 years of experience, suggesting the survey attracted responses from more senior participants. 33% of the respondents had less than ten years of experience. Only 3% of the respondents had less than three years of experience.

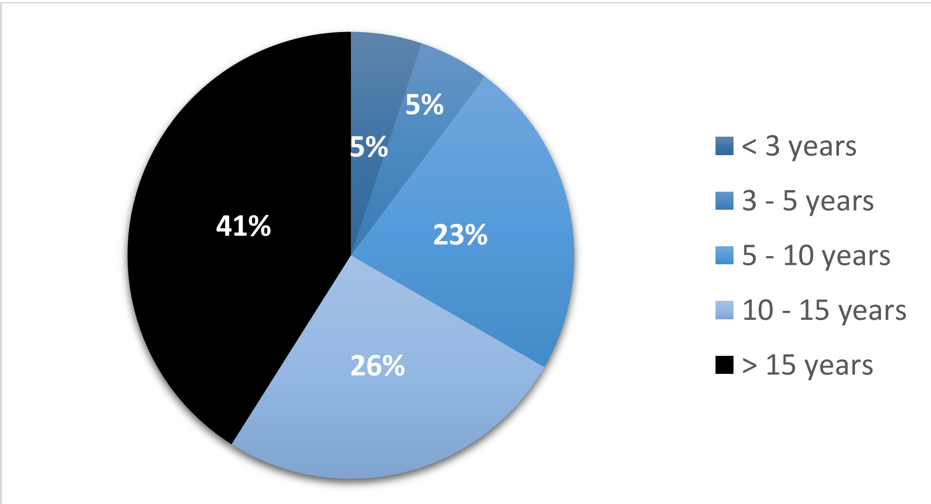

Figure 24: Breakdown of survey participants by experience

Although tenure or number of years of experience in a field is pivotal to understanding the long-standing role of data and its importance in decision-making, the survey attracted plenty of "junior" respondents (i.e. less experienced) who are at the forefront of new technology. This group may have a heightened understanding of the importance of the technological shift currently underway in Big Data, artificial intelligence, and computer vision.

The experienced respondents (i.e. senior) brought attention to many issues from past experience in turning data into insights or converting data into intelligence. The younger respondents were more familiar with these concepts, having been born in an era dominated by technological pioneers in the form of Google, Facebook and Amazon, who have made advanced technological concepts mainstream.

Psychological research relates the impact of experience on decision-making. There are a few essential factors that impact the fundamentals of leadership and decision-making. Critical elements of decision-making incorporate past experiences, an assortment of psychological predispositions, past sunk results, and personal differences, including age and financial status. A combination of these personal predispositions influences the decisions made by individuals. Past experiences influence the course of future decisions (Juliusson, Karlsson, & Gärling, 2005). Individuals who have gained positive results from past decisions tend to make decisions using similar techniques, and conversely, individuals gaining insight from past mistakes avoid repeating them (Sagi & Friedland, 2007). Decision-making using historical precedence tends to include various biases. These biases include cognitive, hindsight, belief, omission, and confirmation (Marsh & Hanlon, 2007; Blank & Nestler, 2008; Stanovich & West, 2008).

Individuals making decisions using past experiences tend to use data without over thinking the insights gained, even though experienced-based decision-making is laden with issues. In other words, intuition and heuristics can be captured more naturally from more experienced individuals in an organic way rather than in a manufactured way.

Management level breakdown

The survey also attracted a mix of individuals varying in experience as shown in Figure 25. Many respondents (33%) were from middle management. Respondents from upper management, in the post of CEO, CIO, Chief Operating Officer (COO) or Chief Financial Officer (CFO) made up 31% of the respondents. Lower management respondents, including people who work in the back-office, business support and technology departments made up 21% of the total respondents. 15% of the respondents preferred not to divulge the management level.

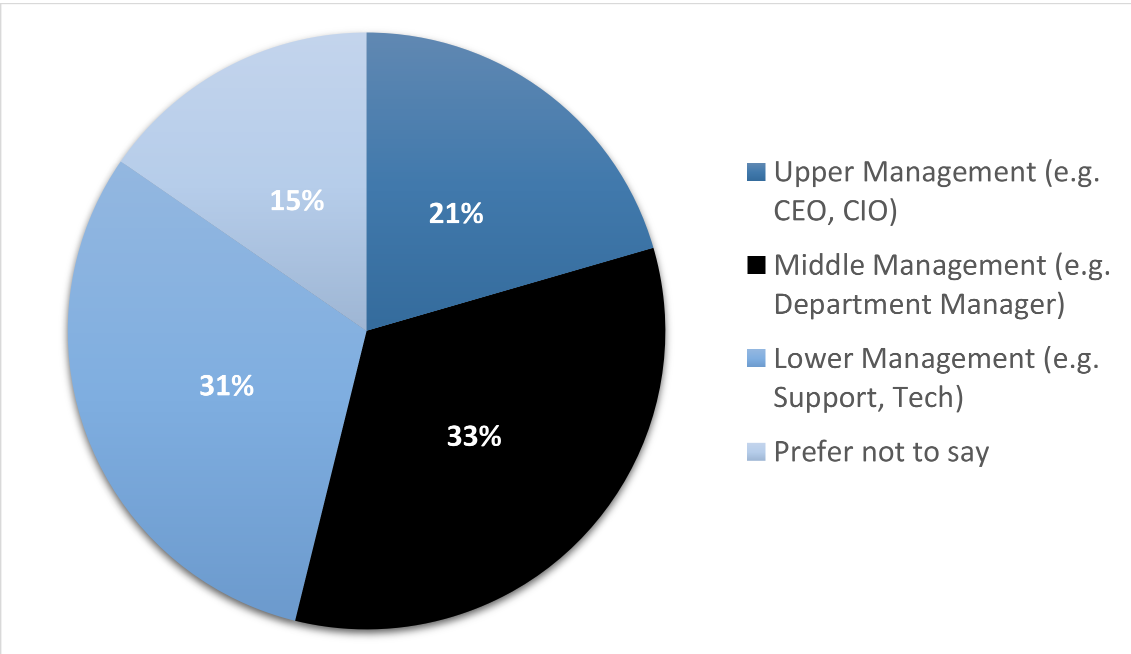

Figure 25: Breakdown of participants by management level

To understand the impact of data on business decisions, it was essential to gain insights from people at various levels of management. In the context of a business-intelligence framework, it is important for mixed levels of experienced professionals to make decisions in a collective manner. Individuals at various management levels bring a varied perspective to business decisions. Upper management is usually the senior stakeholders of the organisation and tend to make decisions which support long-term vision for the organisation's growth. Senior levels of management tend to make long-term decisions based on where they see the organisation in future or the larger mission, creating long term goals and defining the vision of the firm. These big-picture goals are usually set in place to help current and future shareholders and creditors of the business understand objectives, without whom growth cannot be accomplished. Upper management sets a series of Key Performance Indicators (KPIs) over the long-term which it strives to accomplish by defining standards and setting broader goals at an organisation level. The decision-making process to formulate and implement strategic plans at the top management level is time-consuming but at the same time considers adverse variables which could contradict their mission statements. This, in turn, provides a unique perspective from upper management on how they envision data playing a role in decision-making over the long term. Data usage in making longer-term decisions has been well documented (Boone, Ganeshan, Jain, & Sanders, 2018) in the time-series analysis using historical data, especially in making financial decisions. However, strategic goals can also benefit from historical data analysis irrespective of operational sector, market capitalisation and competitive landscape. Data-based decisions are agnostic of business size, sector or standing in society.

Middle management differs from senior management in that it finds itself dealing with day-to-day activities of the firm. Middle managers handle functions including human resources, materials management, staffing, business control, marketing, manufacturing and resourcing. While senior management makes long-term decisions, middle management makes the shorter-term, tactical decisions. Controlling and managing shorter-term goals embodied in granular mission statements, which when aggregated upwards, lend to the greater goal of the organisation set by upper management. The shorter-term tactics set by middle management on a daily, weekly or monthly basis adhere to the principles set by the upper management. These smaller goals are often set at the department level and then further drilled down to individual teams. Middle management performance is often measured in short-term goals achieved within the business segment for which they are responsible. The respondents from middle management brought valuable insights in terms of how data might help make tactical decisions daily. Since this research project proposes a business-intelligence framework to help make organic real-time decisions, middle management would benefit from implementing a business intelligence framework such as Hex-E because of the fast-paced nature of decisions required in this role.

Lower management individuals at the operational line of supervision, give insights to data provided to them, which assists middle-level management. This assistance to the middle management comes in the form of enriched data from an operational perspective. Lower management identified the challenges in normalising, standardising and tokenising the data received in a consistent format to provide relevancy. This group was at the forefront of facing data in various formats, size, velocity and variety. These respondents dealt with structured and unstructured data. A combination of individuals from these three levels of management bought a balance on the views towards data and its journey from insights to decision-making.

Data usage breakdown

The importance of the data-driven decision model was examined earlier in the literature review (Chapter 2). However, getting a practitioner's view was an insightful by-product of this survey. Participants were requested to share views on how important data is during decision-making and whether its importance and usage have increased over time. Additional insights were provided by the cross-sectional analysis overlaying the experience of the participants, highlighting the results in a multi-dimensional way. Figure 26 shows how important data is in decision-making and further break downs the responses by the participant's experience.

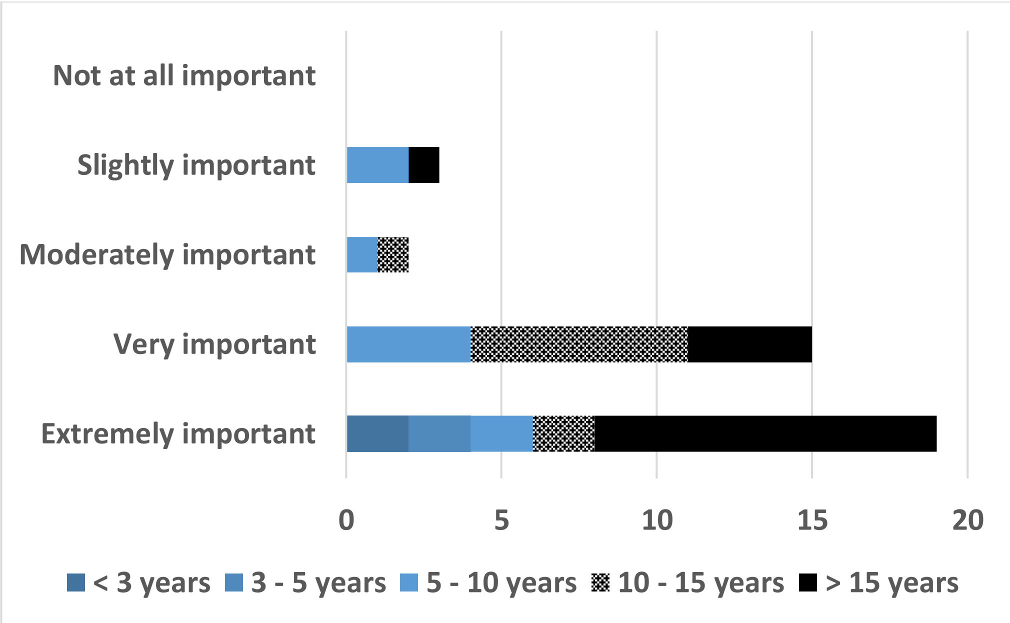

Figure 26: Participant responses on "how important is data in decision-making."

67% of the respondents (33 out of the 50 participants) believed data was either very important or extremely important in decision-making. The remaining participants (33%) either thought data was moderately important or slightly important to decision-making. None believed data had no importance. Overlaying experience, as an additional dimension to this response, reveals that more experienced participants thought data was vital to decision-making. An overwhelming majority (94%) of the more experienced participants in the survey thought data was essential, in extreme degrees to decision-making.

The survey results from this question were unexpected, given the literature reviews conducted in Chapter 2 highlighted the fact that certain managers, especially more experienced ones, rely more on intuition and experience rather than an objective approach where they start with evidence-based insights. As businesses face the mounting challenge from the enormous amount of information at their disposal, it would be difficult to ignore hard data with soft (untested) intuition. The challenge does not just lie in business intelligence or technology aspects of the current environment, but also in the psychological aspects of decision-making by managers. Psychological research has aptly identified the inherent biases people have based on past experiences. The only way to manage biases and prevent impact on the business environment is to make the decision-making process more objective and data-driven.

The survey results align with the literature review on data-driven decision-making and how managers are slowly deploying data-driven approaches to supplement their decision-making process, rather than relying solely on experience and intuition. Data-driven processes provide affirmation for managers on intuitive actions in decision-making.

Business Intelligence from existing data

Following up from the previous survey question in which the importance of data was gauged from a decision-making perspective, the following response analysis gauges participant perspective on data effectiveness for gaining business intelligence for decision-making.

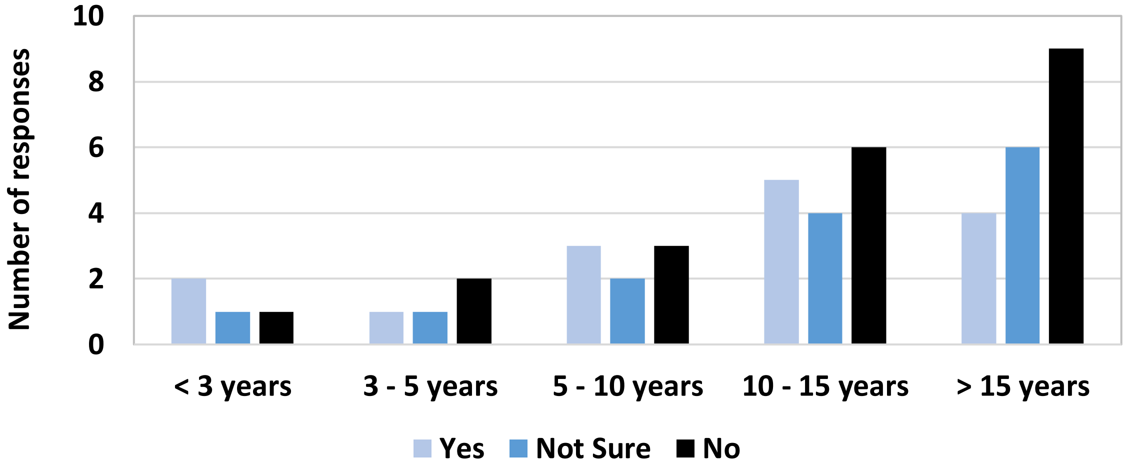

Figure 27: Survey response on optimal business intelligence from existing data

Figure 27 depicts the results from the question on participant perceptions that internal data is optimal for gaining business intelligence. 30% of the respondents responded positively that internal data does provide optimal business intelligence. The remaining 70% of the respondents either were not sure or did not think internal data provided optimal business intelligence. Consistent with the previous analysis, an overlay of a second dimension (experience) on this survey response provided additional insights on how people with varying degrees of experience thought about their internal data's ability to provide optimal business intelligence.

The responses from participants with less than five years of experience were mixed, with an equal number of participants responding positively and negatively to this question. The impact of less than optimal internal data for gaining insights, perhaps is less known to employees who are at the operation level of management. Low-level managers are also at the point of capturing data from various activities undertaken directly or managed at a lower scope. There can be a disconnect between the data captured at this level and how senior management understands and consume the data. This lack of understanding is one of the current overarching problems, in which businesses are capturing an increasing amount of data daily but fail to understand how to use it in a meaningful way to support future decision-making. A Forrester research paper recently highlighted between 60% and 73% of data collected by businesses goes unused (Forrester, 2018). There are multiple reasons why this issue arises from an organisation's point-of-view. Initial design analysis for a typical business intelligence framework considers the quantity, format, velocity and quality (veracity) of the data in question. As time progresses, new systems or hardware (Internet of Things) are put in place to capture additional data-sets not encapsulated in the initial design. Businesses fail not only to envision future technology trends but also potential paths of future technology. The framework itself should be dynamic and flexible enough to consume new datasets in varying formats, structure, and velocity. This ability to incorporate the unknown is an area of focus for the Hex-Elementization framework.

Lag in data for decision-making

The earlier questions and the responses discussed in this chapter dealt with the top-level view of data and the practitioners using it to make business decisions. The following series of recorded responses discuss the form, quality, and timing of data received in day-to-day duties, and how it impacts decision-making.

Borrowed from the field of economics, data can be viewed as a leading or lagging (indicator) of an independent activity (e.g. stock market). Leading indicators, as the name suggests, leads a particular economic activity. The following are a few examples of data-patterns which can be considered leading indicators. Analyst earnings forecast of businesses in the next 12-months is a good predictor of equity market returns. Attendance records in the school lead to better overall quality of education in schools and foot traffic in a mall is a good predictor of consumer spending. Leading indicators are predictive and can help organisations fine-tune strategies in advance by pre-empting the impact of threats on business operations and overall profitability, including technological disruption, competitive landscape, and natural disasters. Businesses need to understand the vast amount of data at hand, including internal data, which can help pre-emptive decision-making in any domain of the operation.

Conversely, lagging indicators, are data series which are historical in nature. The impact on specific activity is gauged retrospectively once the trend is known. Although lagging indicators have a place in various fields, such as economics, these indicators are not helpful in formulating dynamic decisions. In the field of decision theory, decision-lag is a concept which is discussed in detail. Decision-lag is the time it takes for decision-making (by upper management) in order to address an organisational or economic issue at hand. Decision-lag occurs due to the options upper management has on hand, each with a cascading impact on its future operations in varying degree. These lags can be detrimental to the operations of a firm because decisions become reactive rather than proactive and are based on old data.

With that in mind, participants were asked the degree of lag experienced in day-to-day activities and whether it adversely impacts decision-making. As summarised in Figure 28, 66% of the participants agreed that the lag in the data adversely impacts their ability to make decisions.

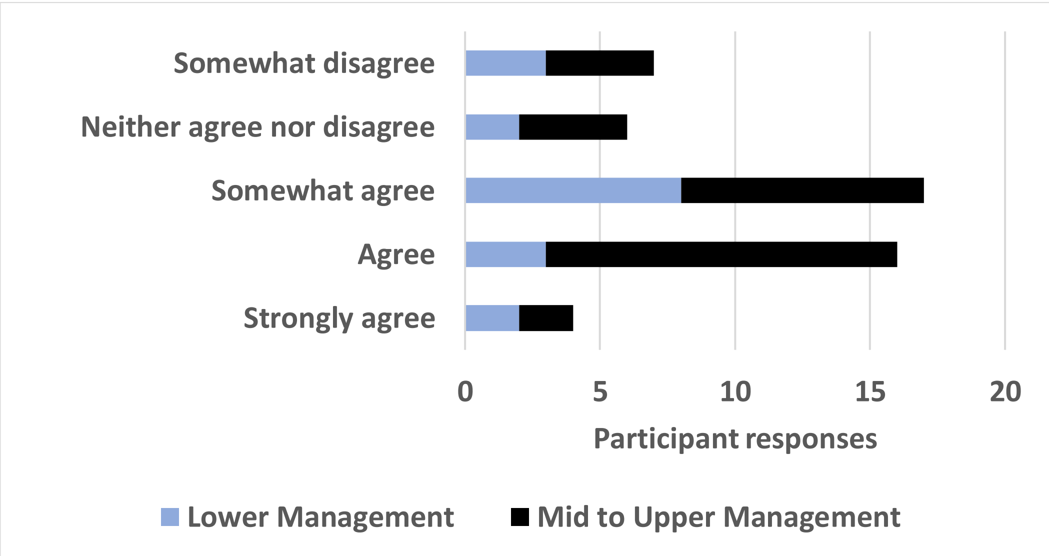

Figure 28: Participants response to the lag in the data and its impact on decision-making

33% of the respondents were either neutral or disagreed that delayed data adversely impacts decision-making. The analysis was further enhanced by overlaying seniority of the participants and classifying into two categories – lower management and mid-to-upper management. This simple bifurcation helped differentiate the strategic decision-makers from the individuals who are involved in the daily operations of the business. 69% of the participants in the mid-to-upper management thought that a lag in data adversely impacts decision-making. A side note was that the timing of the data lag was important to eventual use and impact. Lag erodes the impact data can have on the decision-making. The longer it takes to convert data into information, the harder it becomes for businesses to take timely decision. This lag then adversely impacts the ability of the businesses to be nimble in changing business landscape. Hence, the lag in data makes it less useful for making timely and accurate business decisions.

Quantity vs. Quality of Data

Following up from the question related to timing and frequency of data, the next question asked the participants about the relevance of the volume of data in decision-making. The tremendous volume of data presented to organisations was discussed in great length earlier in the literature covering data-driven decision-making. Evidence-based decision-making has gained popularity because of the ability to be consistent and because every action is measurable against a benchmark. Quantification of various business activities has increased over the decade as new and inexpensive technology has come to the forefront. In 2017 the cost of storage fell from $2 per GB on a Cloud platform to $0.02 (Wang & McElheran, 2017). This quantification shows the proliferation of evidence-based activities conducted by businesses on a broader scale, including the decision-making process. As more devices get connected to the Internet (IoT), and more activities are quantified and monitored, the deluge of data is hard to contain. This deluge of data was evident in the response from the participants who overwhelmingly (82%) agreed quantity of data was at an all-time high.

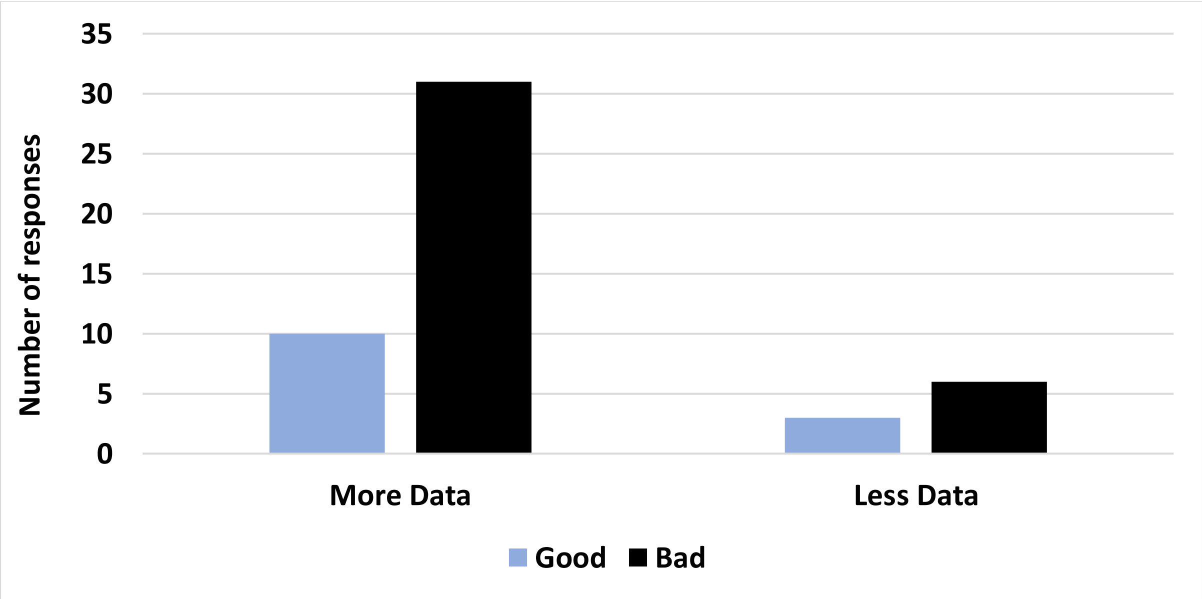

Figure 29: Survey responses on the quality of data they receive

As shown in Figure 29, only 18% of respondents highlighted that they use less data in their decision-making process. Additionally, the quality of data was also added as a dimension to this data analysis. 76% of the respondents who agreed that they are facing more data than ever, also agreed that the quality of data is poor or bad.

The Quantity vs. Quality discussion has been going on for many years when it comes to the topic of Big Data (Gawande, 2009). A lot of data (Big Data) is not necessarily synonymous with good data. However, if one is capturing millions of clicks on a website, millions of conversion rates on e-commerce operations, or temperature changes on a millisecond basis, then some outliers in the underlying data is expected. The ultimate derivation of analytics from the data is what dictates the quality of the data. If the purpose is to create a long-term trend, then a higher margin of error is acceptable, provided its inconsequential to the overall analysis. However, there are many cases in which the quality of the data dictates the outcome which forms the basis of decision-making. In cases like these, the volume of data is addressed along with improving the quality of data for optimal results. Optimal suggests that a few issues with quality are acceptable provided the impact is minimal on the overall observations.

It is interesting to note that participants highlighted that most of the data they receive is “bad” in quality. This issue of bad quality suggests that time and valuable resources in the form of labour and technology are required in pre-processing the data before it can be converted into an easier form to correlate, coalesce and extract business intelligence.

Quantity vs. Context

As Artificial Intelligence becomes more prominent in extracting meaning from otherwise unfathomable data, context plays a significant role. Advanced machine learning concepts like Natural Language Processing, Neural Networks, Deep Learning, and Text Analytics are used to emulate the way humans make decisions. In most of these initiatives, AI is involved in making sense of data. Heuristics drives human decisions, and when a machine can be taught the concept of heuristics and complemented with sufficient data to test the fundamentals of the decisions, it creates a complex system which amplifies (in quality, quality, speed and value) what human beings can achieve when it comes to making decisions.

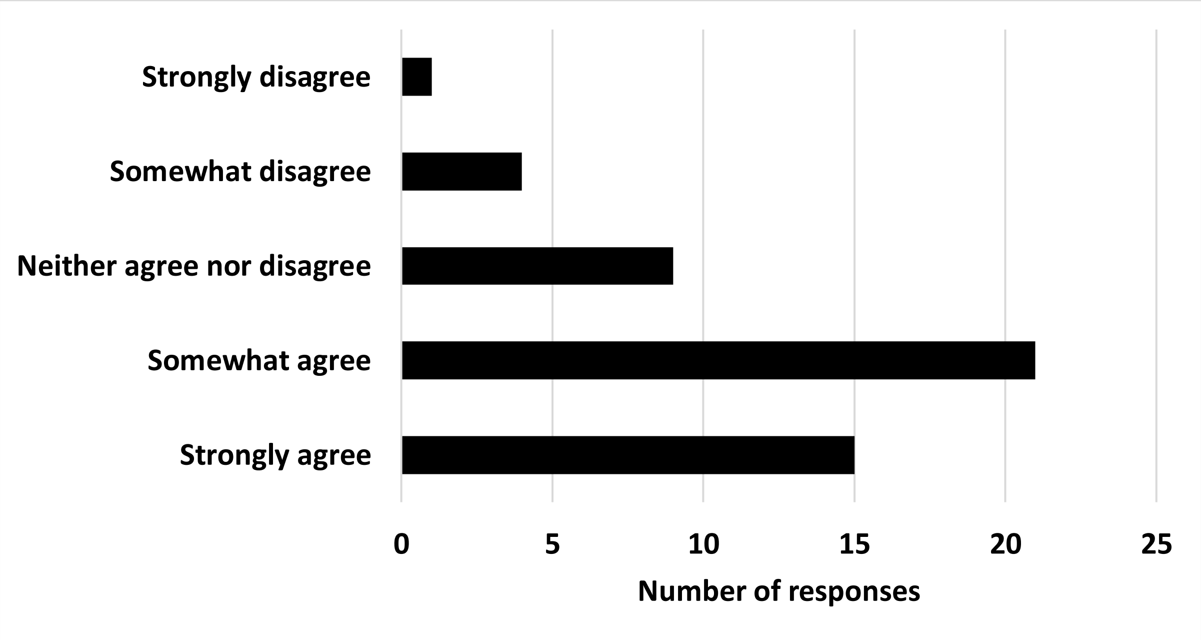

Figure 30: Responses on whether context gets lost amidst a huge quantity of data

The participants in the survey were asked if they felt that the business context is lost amidst the vast quantity of data. The results, which are plotted in Figure 30, show that 72% of respondents were positive that context gets in the volume of the data. Looking at the same data set from an additional dimension highlighted a more pronounced aspect of the quantity vs context conundrum. This additional dimension was analysed through the cross-sectional analysis of the same question by experience level. This analysis revealed most of the participants with greater than ten years of experience, agreed that context got lost in the enormous amount of data consumed. As shown in Figure 31, the lower the experience level of the respondents, the less prominent their view on lost context.

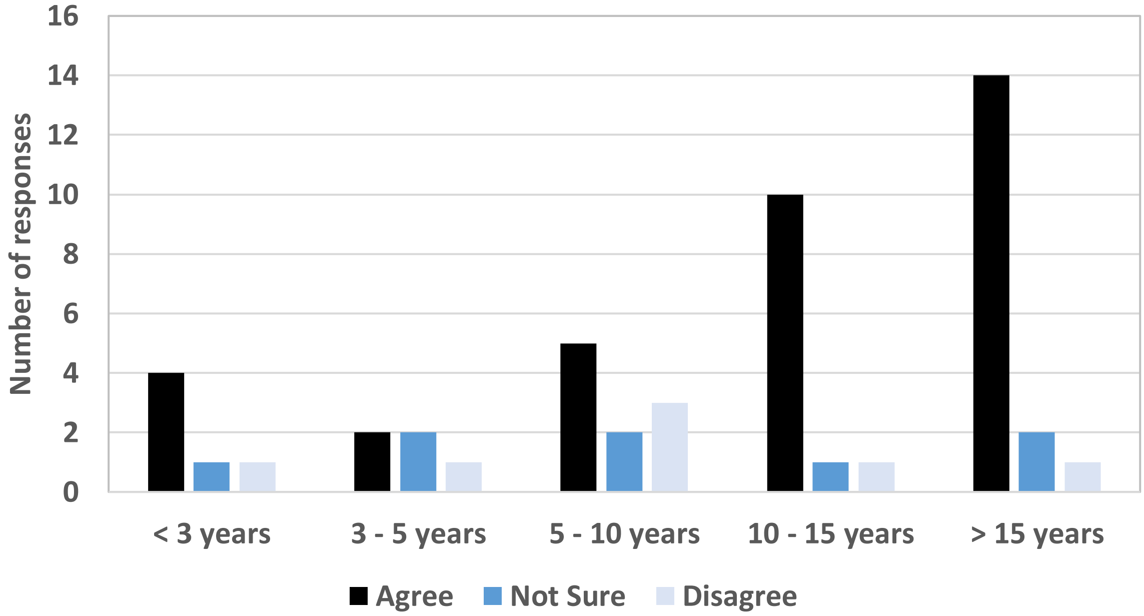

Figure 31: Survey response on the loss of context by management level

This loss of context is due to the functional difference between the decision makers versus the personnel who provides the insights in a typical firm. As upper management sets big picture goals, they are in a better position to judge the context of the insights provided to make decisions. Middle and lower management are often responsible for extracting insights from the data deluge. This process can be cumbersome and cause the context to become less important than the issues associated with format, quantity or quality. Experience of those consuming the data helps tie the context to the goals.

Major challenges with the existing data

Data, as highlighted in the literature review, is now considered the fourth factor of production for an organisation (Brynjolfsson & McAfee, 2014). The other three being land, labour and capital. Many new-age enterprises, which are platform based, including Airbnb, Coursera, and Alibaba, are entirely built on the Cloud with entirely data-driven decision-making ability. As evidence-based and data-driven decision models gain popularity, it will become more difficult for subjective decision makers to ignore data in the changing and highly dynamic competitive environment. However, data is not without its limitations. Businesses currently spend a lot of resources, including software, hardware and people, in transforming data into information and insights.

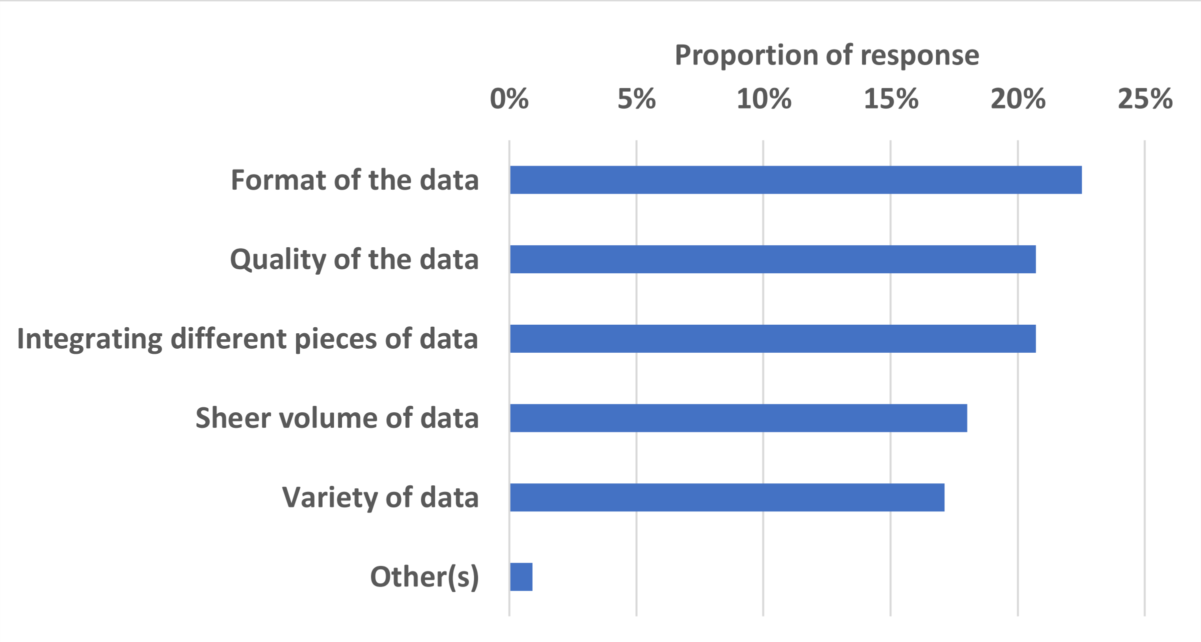

Figure 32: Various challenges faced by survey participants with existing data

The survey asked the respondents to highlight major challenges with the existing data. This question gathers information about activities or functions within the data management domain that takes most of their time and resources. The results are highlighted in Figure 32. “Format of the data” is highlighted as the main issue faced by most participants in the survey.

This issue is not surprising given there are more than 1500 file formats (Kaplan & Haenlein, 2010) currently used across different operating systems, applications, software and platforms. These formats are summarised in a table in Appendix 1. The conflicts across data formats are not going to disappear amidst an increasing number of platforms. This complex array of formats makes it difficult for businesses to adopt and implement a business intelligence system.

Hex-E is envisaged to behave like a protocol which provides an environment for data to enrich in its granular form. Data in this most granular format, as envisaged by the proposed by Hex-Elementization, is platform agnostic. An example of this concept is the way the Internet is accessed through multiple operating systems using any of the browsers available. No matter which operating system is used, and which browser is utilised, the content of a website is consistently displayed. This agnostic behaviour of the Internet is the basis for system agnostic Hex-Elementization. It will not matter which system or machine feeds Hex-E with data; the output is a consistent form which is contextually driven.

The second most challenging aspect of data is “Quality”. As explained in the previous question, the quality of the data is a challenge, especially when it is associated with the quantity of data. There is no correct level of quality when it comes to extracting trends from Big Data. Quality is subjective, and there is an optimal level of quantity vs quality profile, which is acceptable when it comes to handling Big Data.

The third most challenging aspect of data highlighted by the participants was “integrating different pieces of data”. The issues with data integration from disparate sources include technology platforms, finances and human skill set. As highlighted in the response from the most challenging aspect when dealing with data, the number of formats (variety of Big Data), the volume, velocity and veracity of Big Data is essential to be tackled before value can be extracted (Unhelkar, 2017). Data combination is time-consuming and technologically challenging. The capital and skill set essential for such a task are usually beyond the capabilities of a medium-to-small enterprise.

"Sheer volume of data" and "Variety of data" were the other two challenges the participants of the survey highlighted. Increasing volume of data poses a conundrum to businesses on what data they should concentrate on. With businesses keen to capture as much as data as possible from various parts of their operations, it becomes harder to sift through them on a timely fashion to generate insights. The data is not just increasing in sheer size, but also expanding at a faster pace (high-frequency). For examples, IoT devices can capture and stream data on a microsecond. The challenge is how to consume this high-frequency Big Data to generate insights which can be impactful and accurate on a timely basis. The challenge of "Variety of data", was discussed earlier in this section, which is challenging to incorporate into a business intelligence framework with the ever-increasing "formats" of data in use.

The Benefit of Unstructured Data in Decision-Making

Unstructured data analysis and optimization is a new domain. Videos, audio files, images, and star-ratings on social media websites are some of the examples of unstructured data. Unstructured data provides a method to gain insights from a dataset which could not have been analysed until high-performance computing, distributed storage and distributed processing and Software as a Service (SaaS) became more readily available. To remain competitive, it is important for businesses to gather as much information as possible in form and frequency. In many instances, unstructured data is used to super-charge the insights already generated by the structured data set.

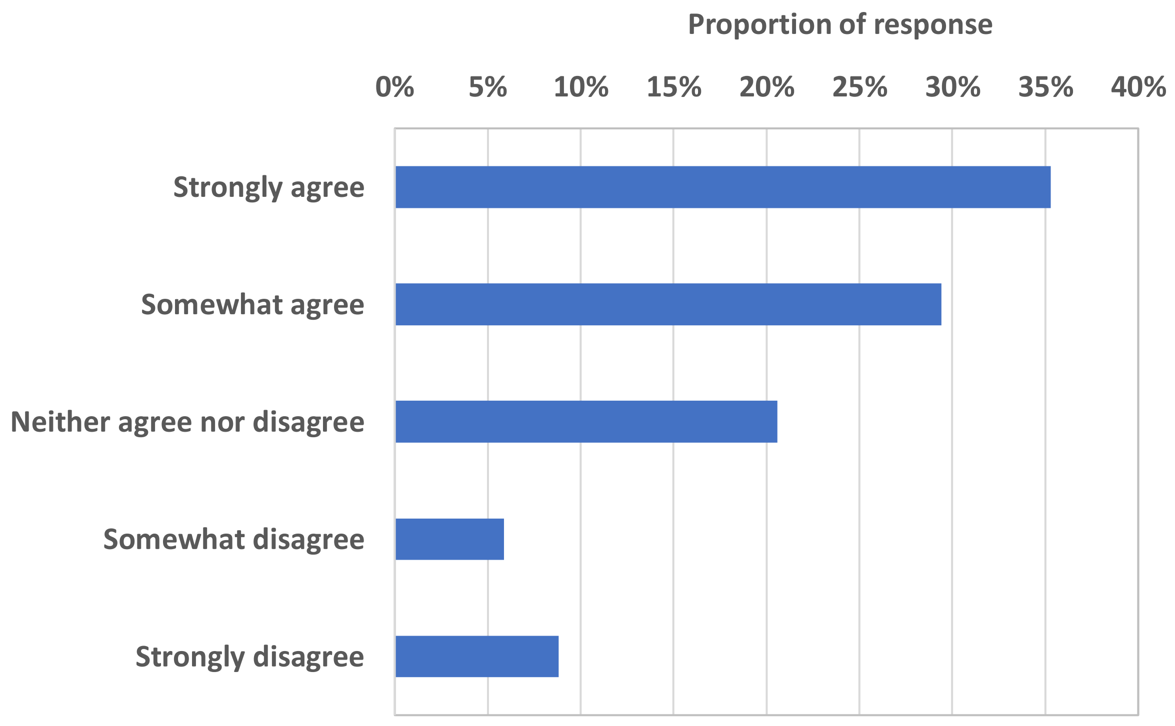

Figure 33: Survey response on whether unstructured data can assist in decision-making

The participants in the survey were asked if they think unstructured data can help in making better decisions. Almost 65% of the respondent agreed or somewhat agreed (Figure 33) that unstructured data could add value to decisions. 22% stayed neutral, by neither agreeing nor disagreeing. Approximately 13% of the respondents did not agree that unstructured data can help. This response can be the result of the series of challenges associated with unstructured data. Firstly, these challenges include the sheer volume of the dataset. For example, pictures and videos are some of the large-sized digital assets seen in the market today. Secondly, algorithm sifting through unstructured dataset remains a black-box. Machine Learning techniques like Neural Networks and Natural Language Processing are used to classify unstructured datasets efficiently and effectively. Although these techniques do a good job extracting information at a faster pace into a structured format which can be quickly consumed, these results are not necessarily clear. Businesses, especially in the age of increased regulation, are obsessed with understanding how things work. This desire to learn how things work is not just for regulators, customers or shareholders, but important for businesses to understand how it impacts future sustainability. Thirdly, the domain of extracting information from unstructured datasets is believed to be the domain of specialist individuals with unique skillsets (e.g. data scientists). Combining click stream, weblogs, social media, or unstructured data, with structured data sets such as transactional systems (e.g., the supply chain system) provides an opportunity for businesses to get a holistic view of clients. Such a task requires a high level of skill which resides with a limited number of individuals (i.e. data scientists).

Finally, traditional data mining is thought to be impossible using unstructured data. It is easier to find a relationship with a high volume of structured data than doing the same with unstructured data. Advanced search engines powered by computer vision, texts analytics, and visualisation tools ease the mining of value from the unstructured data set. Businesses can then leverage the combination of unstructured and structured data sets to generate insights to support long-term strategic decisions.

The use of unstructured data has enormous potential in highlighting new products and services, improving profitability, reducing costs and providing the ability to react to dynamic markets.

Perceived benefits of the Hex-Elementization business intelligence framework

To conclude the survey, the participants were asked about their view on the proposed framework – Hex-Elementization. The participants were given a brief research proposal, in which the research questions, research outcomes and the goals of the model were explained in detail to help them form a view on the framework. The participants were provided a list of perceived benefits a framework like Hex-Elementization would bring if it was implemented at their workplace.

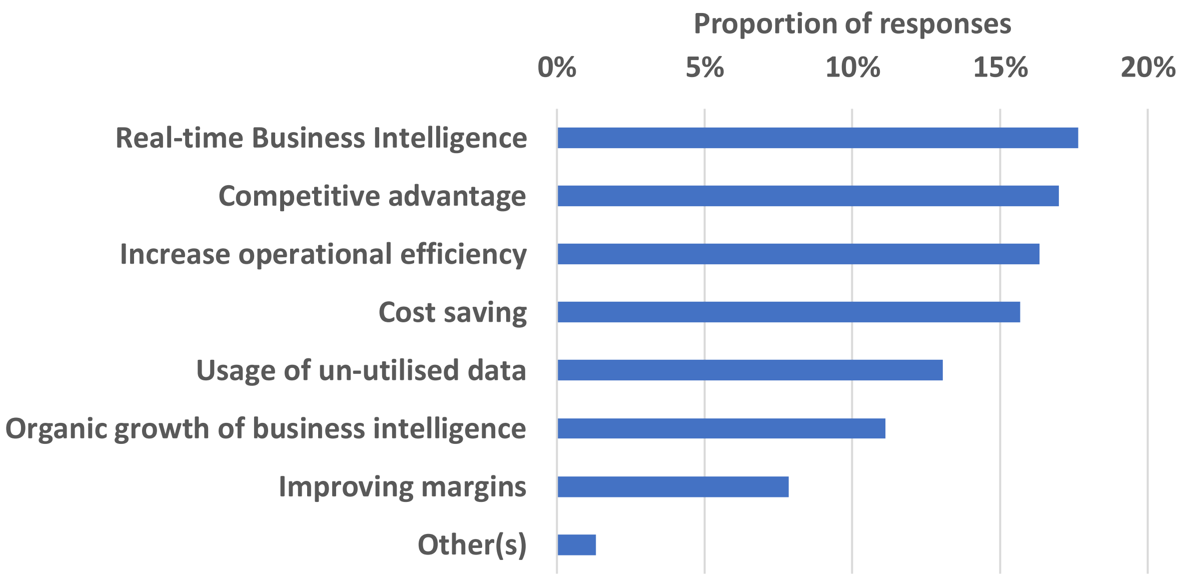

Figure 34: Survey response on the perceived benefit of Hex-E framework

As shown in Figure 34, 18% of the participants of the survey picked "Real-time Business Intelligence" as the number one perceived benefit from the Hex-Elementization framework. Real-time Business Intelligence (RTBI) facilitates up-to-date insights and intelligence in an amalgamated form, across various parts of the business. Such a system taps data from operational, distribution, manufacturing, marketing, and sales departments of the business and, extracts intelligence to make faster and more tactical decisions. In order to make decisions in an era when technology and consumer tastes change quickly, business needs to collect data from various points and needs to accumulate the information in a data warehouse. The issues with the format, quantity and quality, as highlighted in this section earlier, makes it difficult for any existing system or framework to extract the necessary insights in a timely fashion. The data-to-intelligence journey has many obstacles, as discussed at length in the early part of this chapter. Hence, the top pick by the participants to highlight Hex-E as being beneficial in generating Real-time Business intelligence is understandable, given the existing issues linked to generating insights on a real-time basis.

At 17%, “Competitive Advantage” was the second most popular choice by the participants. Operating a business entails more than just achieving profitability and sales targets and high customer satisfaction (Chen & Popovich, 2003). It requires a vision of the future in which market share increases, employee satisfaction rises, products and services use cutting edge technology, and business keeps pace with changing times. These wins are only possible with careful planning through business decisions which gives the business a competitive advantage (Oliver, 2000). It is usually easy to have a competitive advantage in the growth stage. However, this advantage can quickly erode if the business does not incorporate new data sets regarding any aspect of its day to day activities, and uses it to generate insight to help constantly monitor the competitive landscape (Sarvary, 1999). Business-intelligence frameworks not only provide intelligence that is intrinsic to the business, but also extrinsic intelligence, which keeps an organisation competitive with others in the market.

“Increase in operational efficiency” came third with 16% of participants choosing this as a perceived benefit of a framework like Hex-Elementization. According to a survey conducted by Gartner in 2010, 71% of responses gathered from businesses of various sizes (market capitalisation) voted “operational efficiency” as the top reason to implement business intelligence. Business intelligence provides the right information to the right people in order to make the right decisions. BI is also used to measure how efficiently the information is used when making optimal internal decisions. Data can become a liability if the flow from various departments is not unified to help generate business insights. BI is not only a means to gain competitive advantage, but it also helps businesses manage data efficiently. Compliance can also benefit from business intelligence. An appropriate BI framework can help spot trends or inappropriate corporate behaviour that could result in regulatory action. Operational efficiency is measured by profitmaking activities and by monitoring internal activities to help improve processes or adhere to compliance regulations. The risk of being non-compliant can far outweigh the profits a firm generates from its core activities.

At 15.6%, “Cost saving” was the fourth popular choice of the perceived benefits of the Hex-Elementization framework. As business intelligence taps into the data throughout the business, encompassing every department, it helps business stakeholders examine different cost saving avenues. BI can also help identify inefficient parts of the businesses as well as the effort needed to reduce the costs of the inefficiency. For example, business intelligence can cross-triangulate information between distribution and manufacturing departments and highlight excess inventory. Excess inventory not only involves unwanted storage costs but also impacts the timely distribution of the end-product. An optimal business-intelligence framework can help reduce marketing and advertising costs by pointing to a strategy which maximises return on investment (ROI). Business intelligence also provides helpful insights into human resource (HR) departments in gauging employee contribution to every output of the business. Thus, BI helps make a quantified decision on reducing headcount when manual processes become automated. Business intelligence can help trigger a self-fulfilling cycle where profitability is improved as a result of higher ROI through improved processes and employee productivity; a cycle that is repeatable through BI. Consistent service, high-quality delivery, and stringent internal controls help control the cost and buys precious consumer loyalty. A real-time business intelligence framework monitors various aspects of the firm in real-time, which supports tactical decisions to help reduce overhead. For example, a real-time business intelligence framework might highlight issues of transporting goods through an area which suffers from seasonal weather conditions (e.g. typhoon). A suggestion to alter the transportation route could be slightly most expensive than the normal route but saves a considerable amount of time and money in the long-term.

Principle Components Analysis (PCA)

There are several survey results highlighted in Part 1 and Part 2 of the Quantitative Data analysis reviewed earlier in this chapter, including common elements and issues. To help classify common questions together, a Principal Component Analysis (PCA) was conducted. PCA helps identify and group underlying components that are the target measures of the survey questions. PCA identifies and quantifies a linear relationship between variables (survey response in this instance). PCA is a statistical method which helps in dimensionality-reduction (Mersereau & Farrell, 2005). PCA helps reduce the variables considered in the survey to help group highly correlated variables. By reducing the number of variables (dimensions) in a dataset, the research effort is concentrated on a fewer number of variables which might have the most positive impact in terms of extracting the most valuable insights.

This exercise aims to determine factors which represent common themes in the questions. Identifying factors helps combine questions that load into the same factors, which is then compared in the next stage of data exploration using regression analysis. In this post survey data collection phase, PCA helps validate what the survey is measuring by slightly sacrificing accuracy, which is a natural outcome of dimensionality-reduction. Figure 35 visually demonstrates the entire process of PCA. The process shows the flow from the very first step of using the raw data (survey responses) to the outcome of gaining the PCA results.

As the first step, 16 questions were classified as eligible for PCA. These 16 questions were selected because the responses could be standardised and normalised (as explained in the methodology section). The categorical questions, which attracted subjective and textual responses, were excluded from the PCA process. Questions addressed directly to the respondents in order to help answer research questions were classified as explicit measures. The implicit measures are common components which were not openly asked in the survey. They were addressed during the Qualitative Data Analysis which is discussed in detail in the next section.

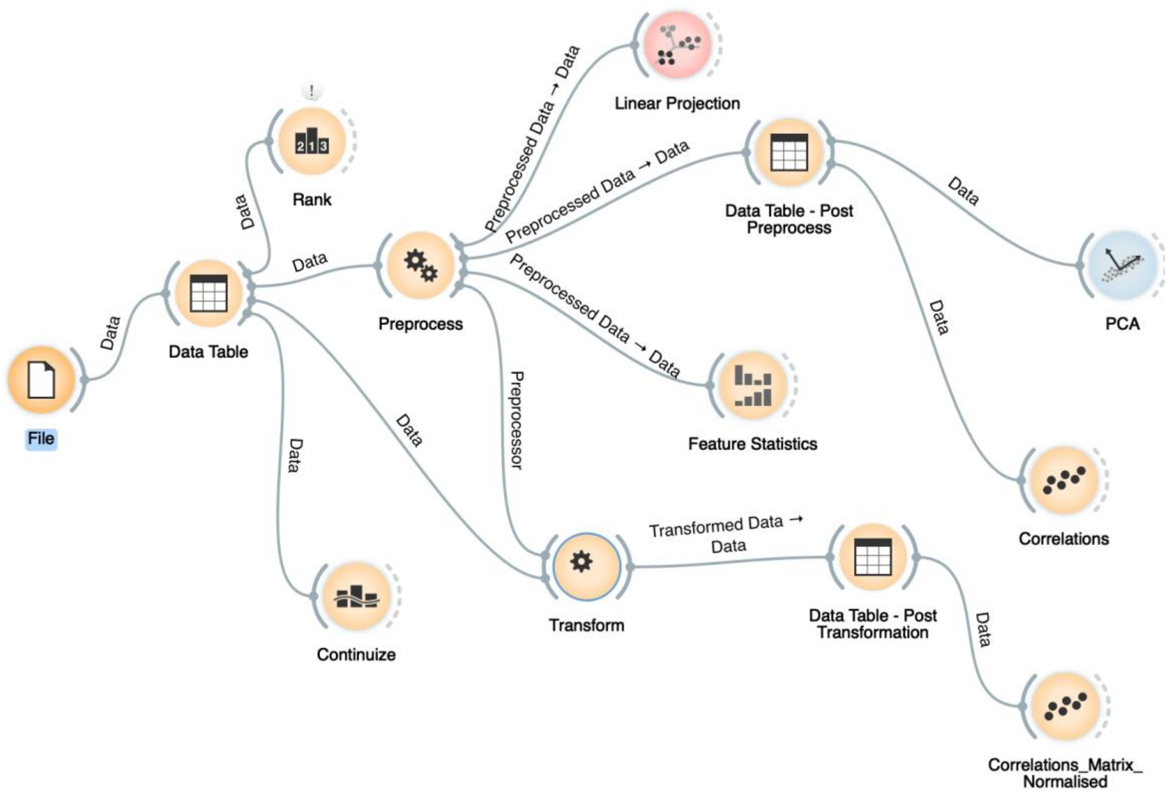

Figure 35: Principal Component Analysis process

The pre-processing completed on the data resulted in a standardised dataset for the 16 questions. A covariance matrix was then created for every combination of two variables (i.e. a pair of two questions) to help understand the relationship between the two. This relationship was represented using the formula:

$$Cor\ (x, y) = \frac{cov\ (x, y)}{\sigma x \times \sigma y}$$

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Q1	1.00															
Q2	-0.17	1.00														
Q3	0.01	**0.42**	1.00													
Q4	0.06	0.11	-0.03	1.00												
Q5	-0.15	-0.03	-0.07	**0.51**	1.00											
Q6	0.00	-0.15	0.08	0.10	0.01	1.00										
Q7	0.07	-0.02	-0.13	-0.34	-0.14	-0.26	1.00									
Q8	0.11	0.18	-0.27	0.24	0.09	-0.52	0.16	1.00								
Q9	-0.08	-0.22	0.08	-0.15	0.22	0.10	0.33	-0.23	1.00							
Q10	-0.01	-0.22	0.13	-0.12	0.23	0.07	0.33	-0.24	**0.98**	1.00						
Q11	0.01	-0.33	-0.12	-0.18	0.08	**0.69**	0.15	-0.21	0.09	0.06	1.00					
Q12	-0.20	0.03	-0.03	-0.12	-0.13	-0.31	0.01	0.36	-0.38	-0.37	-0.21	1.00				
Q13	-0.02	-0.04	-0.42	0.23	0.43	-0.27	0.32	**0.55**	0.06	0.04	-0.12	0.10	1.00			
Q14	0.05	0.23	-0.05	0.19	0.15	-0.30	-0.14	0.38	-0.53	-0.48	-0.30	**0.61**	0.17	1.00		
Q15	0.03	-0.05	-0.15	0.37	0.32	0.30	-0.03	-0.01	0.21	0.16	0.00	-0.17	0.39	0.06	1.00	
Q16	0.00	0.08	**0.61**	-0.22	-0.09	0.14	0.05	-0.47	0.07	0.13	-0.02	0.20	-0.49	0.23	-0.20	1.00

Table 9: Correlation matrix of standardised responses

The covariance matrix for the 16 responses was then plotted, as shown in Table 9. This numerical matrix helped visualise the similarities between issues highlighted by the survey questions. The most correlated pairs of questions are highlighted in bold in Table 9 to illustrate which questions had the strongest correlation. In the next step, the 10 most positively correlated pairs of questions were pre-processed to find which combinations of questions (pair of two) had the strongest explanatory power for similar issues. The next step in this process is to calculate the eigenvectors and eigenvalues of the linear transformation to simplify the characteristics of the issues posed in the survey. Eigenvectors help quantify the core characteristics of a given vector of data which does not change by any linear transformation. Eigenvalues are just the characteristics values associated with the eigenvectors. The eigenvectors corresponding to the principal components, and the eigenvalues to the variance, are explained by the principal components. The principal component analysis of the correlation matrix in Table 9 provided an orthonormal eigenbasis for the space of the plotted data. The largest eigenvalues relate to the principal components that are linked with most of the co-variability among several observed data. Data-mining software, Orange, was utilised to help conduct this analysis, in which eigenvectors and eigenvalues were multiplied by the correlation matrix to fit the transformed data into a new series. Orange enables interactive data exploration for rapid qualitative analysis with clean visualizations. Easy to use graphic user interface enables to focus on exploratory data analysis instead of coding, while clever defaults make fast prototyping of a data analysis workflow extremely easy (Orange, 2019). Although the software does the computation, the math behind the transformation can be simplified using the below equation.

$$[correlation\,matrix \times eigenvector] = [eigenvalues] \times [eigenvector]$$

Applying the above equation yields the series of eigenvalues and eigenvectors. The next step re-orients the data by multiplying the original data by the eigenvectors to fit into the new axes. This re-oriented data can be depicted in a bi-plot (XY scatterplot), as shown in Figure 36. These combinations of principal components were plotted on the scatter plot to visualise and quantify their relationship through linear regression, as shown in Figure 36. This analysis also helped group the questions which were not highlighted as having high correlation using the covariance matrix.

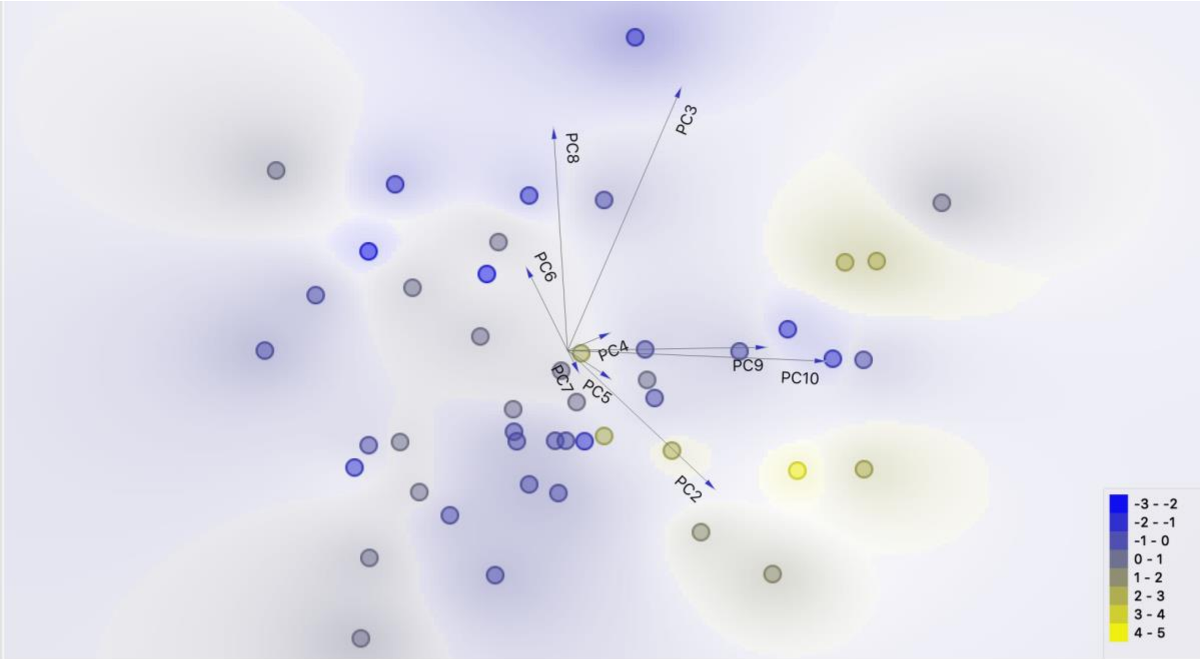

Figure 36: Linear projection of principal components

In the scatter plot (Figure 36), the axes are standardised, and the arrows help identify the original variable values. The X and Y axes are the principal components and the data points, or dots, are the standardised cross-scores of the survey answers. The arrows point in the direction of increasing values of the original answers provided. The closeness of the arrows means the two variables are highly correlated. In other words, the closer arrows indicate a higher level of connection between the responses.

Regression analysis on composite issues discussed in the survey – Study 1

The previous section of data analysis explored how participants viewed the importance of data and the challenges with the use of data in the decision-making process. Use of different types (unstructured vs structured) of data in use was also discussed by the participants in the pursuit of business intelligence. Questions around quality and lag were highlighted as significant issues by participants. Medium to lower level management participants highlighted issues with the operational data in day-to-day workflow.

It was essential to analyse the level of severity of the issues highlighted by the participants and then weight the issues by seniority. This weighted-severity helps consolidate the issues highlighted by different participants and shows which stakeholders are highly affected most by them. Senior management typically drives major initiatives, like adoption or implementation of a business-intelligence framework. Hence, weighting by seniority was important when visualizing the issues. This measure highlights that the degree of severity is higher when the participant is more experienced. This does not mean that the issues are irrelevant to the middle-to-lower management stakeholders. Middle to lower level managers are mostly responsible for making short-term operational decisions. However, upper management is responsible for making long-term, long-lasting, highly impactful, and sometimes revolutionary changes to the vision of the organisations.

Several important issues in this research needed to be combined to understand common themes and issues. The Principal Component Analysis helped highlight questions which trended and were correlated. The top seven principal components highlighted seven questions which were highly correlated. These seven questions were then compared against the seniority of the participants with the help of a linear-regression to help quantify the relationship.

Linear regression studies help quantify the relationship between multiple variables (Dupont & Plummer, 1998). In most cases, linear regression is used in the predictive analysis in which an independent variable (or multiple independent variables) is analysed against a dependent variable to understand if the former can be used to predict the latter. For example, the use of gas in the U.S. increases during the winter season due to higher demand, which in turn affects gas prices just before the winter season begins. In this case, independent variables like temperature, season, and consumer demand, can be regressed against the gas prices to understand and quantify the relationship. A typical regression analysis helps establish causality between the independent variable and the dependent variables. The more closely a group of variables impact an independent variable, the stronger the causality or relationship. This strength of the relationship is usually denoted in R2 (the coefficient of regression which ranges between -1.0 to +1.0). The higher the value, the stronger the relationship between dependent and independent variables. A coefficient closer to 0 indicates either a coincidental relationship or no relationship at all. Another important metric of a typical linear regression analysis is the value of the beta. The beta is the slope of the regression line and is plotted with the known values of dependent variables and estimates of independent variable. A rising regression line has a positively sloping trajectory which indicates the beta is positive. A beta of 1 suggests a perfect and symmetrical relationship between the movement of the dependent variable(s) on the independent variable. A beta above 1 suggests a cyclical relationship, where a given change in the dependent variables causes a significantly positive impact on the independent variable. A beta below 1 but above 0 suggests less elastic sensitivity. A beta below 0 (i.e. negative) suggests an inverse relationship between the dependent variable and independent variable where a small positive change in the dependent variable causes an opposite (i.e. negative) impact on the independent variable.

To prepare the data, questions which had the most significant impact on the research were amalgamated for standardisation. Seven questions were selected which had highlighted various issues ranging from the quality of data to the lost opportunity of not using unstructured data sets for decision-making.

The regression study was divided into two parts. The first study was conducted grouping four questions together. These four questions are listed below.

1. In your opinion, is there a lag between when the data is received and when the decisions are made? In other words, are business decisions based on lagged or outdated data?
2. How would you rate the quality of the data used for day-to-day decision-making? e.g. If the data needs no further processing/cleaning, then the quality of the data is considered Extremely good.
3. Do you think you get optimal business intelligence from the information gathered within the firm?
4. What proportion of your time is spent on a regular basis in cleaning and linking data from disparate sources before it can be consumed as information to make business decisions? - % of the total time.

The kind of response to each question listed above was different. For example, the first question was based on a five-scale response, in which the participant had options ranging from STRONGLY AGREE to DO NOT AGREE. Similarly, the second question also gathered responses from a five-scale option. Responses to these questions needed to be standardised in order to be compared with each other for further analysis. As a first step, the responses to each question were quantified by providing a score to each answer. On a five-scale response type, the most negative responses (DO NOT AGREE, or EXTREMELY BAD) were given a score of 5, and the most positive responses (STRONGLY AGREE, EXTREMELY GOOD) were given a score of 1. The rest of the responses were scaled between 5-1 (from most negative to positive). Once each response was scored across the same question, they were turned into a percentile rank using the below map (Table 10) – to enable standardising the response with a value between 0 and 100. A small example showing the transformation and standardisation of the responses is demonstrated in Table 10.

No.	Original Response	First Transformation	Second Transformation
1	STRONG DISAGREE	5	100
2	DO NOT AGREE	4	75
3	NOT SURE	3	50
4	AGREE	2	25
5	STRONG AGREE	1	1

Table 10: Sample transformation and standardisation of responses

After the second transformation was completed for the four questions of the first study, an aggregate score was created. As highlighted in Table 10, the most negative responses were given higher values across all the four questions to capture the aggregate severity of issues at the core of the four questions. This aggregated and standardised score formed the first variable (dependent) set for the regression analysis.

With the second (independent) variable needed for the regression analysis, the seniority of the participants was taken into consideration. The survey included three questions which were designed to gauge the seniority of the participants. These 5th, 6th and 7th questions are listed below.

1. How long have you been involved in the industry you currently work in?
2. At which management level are you currently working?
3. Are you responsible for making firm-wide or department-wide decisions in your firm?

Similar to the dependent variables, independent variable responses were first transformed by scoring and then standardising from 1 to 100. To manage the scale of the analysis, responses were aggregated by taking a mean of the three standardised scores from the responses to the three questions listed below. A standardised score of 100 suggested the participant is more senior than the average of the rest of the participants. A standardised score of 1 suggested the participant has less seniority compared to the average seniority of the participants. At the end of this exercise, the independent standard score of seniority was calculated across all the participants and cross-triangulated against each of their responses.

Furthermore, to understand the trend between the severity of the issues highlighted by the participants and their seniority, one of the six possible regression trend line estimation could be used to quantify the relationship. The six different ways to fit the data includes linear, logarithmic, polynomial, power, exponential, and moving-average. As the data for the analysis does not include a historical time-series, only a few of the above listed six approaches to regression applies to this study. These six methods of regression are briefly explained below (Anderson & Anderson, 2011; Microsoft, 2019).

1. **Linear** - A linear trend line is a best-fit straight line that is used with simple linear data sets. The data is linear if the pattern in its data points resembles a line. A linear trend line usually shows that something is increasing or decreasing at a steady rate.

2. **Logarithmic** - A logarithmic trend line is a best-fit curved line that is most useful when the rate of change in the data increases or decreases quickly and then levels out. A logarithmic trend line can use negative and positive values.

3. **Polynomial** - A polynomial trend line is a curved line used when data fluctuates. It is useful for analysing gains and losses over a large data set. The order of the polynomial can be determined by the number of fluctuations in the data or by how many bends (hills and valleys) appear in the curve. An Order 2 polynomial trend line generally has only one hill or valley. Order 3 generally has one or two hills or valleys. Order 4 generally has up to three.

4. **Power** - A power trend line is a curved line that is best used with data sets that compare measurements that increase at a specific rate — for example, the acceleration of a race car at one-second intervals.

5. **Exponential** - An exponential trend line is a curved line that is most useful when data values rise or fall at increasingly higher rates. An exponential trend line cannot be created if the data contains zero or negative values.

6. **Moving Average** - A moving average trend line smooths out fluctuations in data to show a pattern or trend more clearly. A moving average trend line uses a specific number of data points, averages them, and uses the average value as a point in the trend line. If Period is set as 2, for example, then the average of the first two data points is used as the first point in the moving average trend line. The average of the second and third data points is used as the second point in the trend line, and so on.

A polynomial trendline was selected in analysing responses to show and quantify the relationship. This was because the peaks and troughs, indicating volatility in the independent and dependent variables are limited and the series does not relate to time. The result of the analysis is shown in Figure 37.

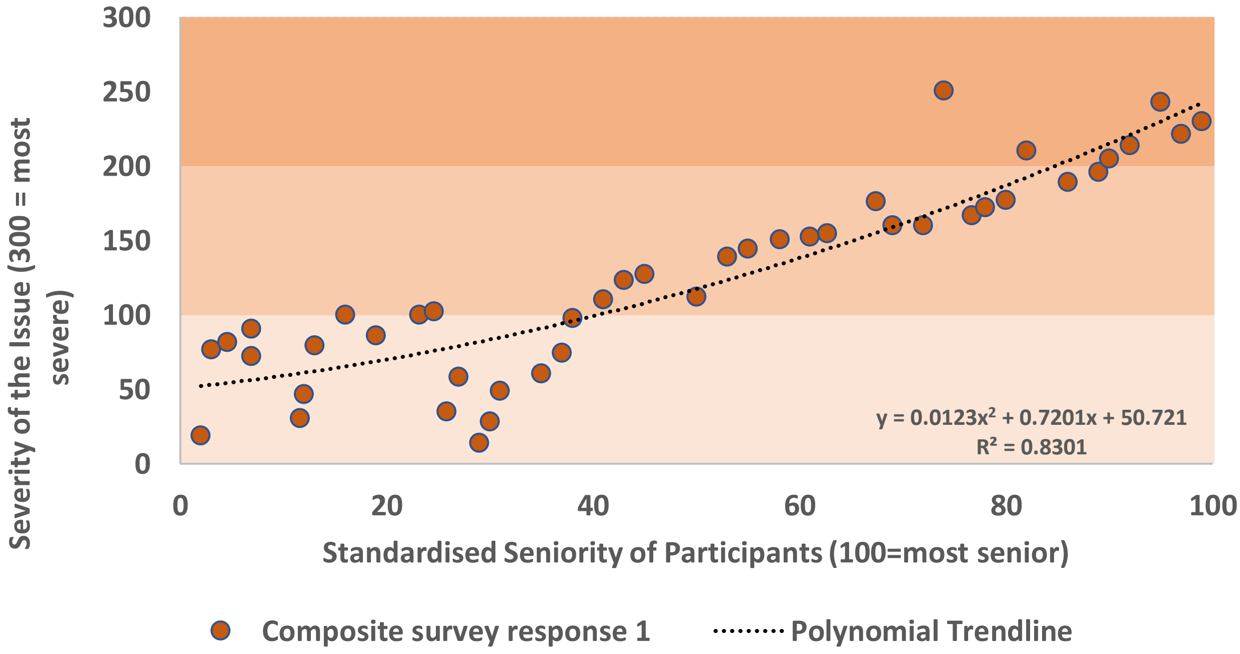

Figure 37: Regression analysis of Composite Survey Response set 1

The Y-axis in Figure 37 shows the aggregated severity) response from positive to negative from the four questions highlighted earlier. The higher the value of the Y-axis the more severe, or negative, the issue. The increasing severity is also shown visually on the chart by three, coloured bands which become darker as the severity or negativity rises. The X-axis plots the seniority of the participants from left to right. The higher the value, the more senior the participant. The rounded markers plotted on the chart are the aggregated response from the four questions from each participant. The chart also shows the equation of polynomial regression. The polynomial line of regression is shown on the chart in black-dotted lines. The chart also shows the value of R2 or R-squared. R-squared is the measure of how close the data is to the fitted regression line. It is the statistical measure of goodness of fit of the trend line to the data. An R-squared (R2) value of 1 suggests a perfect fit.

How to interpret the chart: The plotted dots (i.e. severity vs seniority) seems to have a good positive relationship. The dots on the chart have a slightly rising trend and trajectory. This increasing severity is also visually confirmed by the slightly rising regression line (i.e. the black dotted line). The relationship also shows robust causality of the dependent variable (i.e. responses) to the seniority. As seniority rises, so does the severity of the issues. The R-squared (R2) value of 0.83 also suggests a strong relationship between these variables. The polynomial equation suggests the trend in the data could be explained by order of two as the data is more linearly dispersed and has fewer troughs and peaks.

The regression study quantified specific issues related to data lag, data quality, internal data, and time spent on pre-processing data. These are the issues of immediate concern to senior management. Senior management is aware of the delay in consumption of data after it is collected. They would necessarily not know how much delay is built in the data after it is collected but before it can be consumed. However, from this study it's evident that the senior level participants know the overall impact of data-lag on their decision-making ability. Senior management continually meet external stakeholders and businesses who demonstrate products or services. This provides an idea of the kind of information that is being consumed by these external parties. Senior management can sense that internal data is not enough to make optimal decisions. Overlaying external datasets to supplement internal data provides a more holistic view. Senior management also understand the issues with data quality and the amount of time spent in cleaning bad data causing delays in generating timely insights.

Regression analysis on composite issues discussed in the survey – Study 2

Similar to the regression study 1, a second study was conducted aggregating three more questions which signified a different set of issues faced by many participants in the survey. This second study was conducted in the same manner as the previous study in which the answers were first normalised into scores and then standardised to allow aggregation across multiple responses.

The three questions which were amalgamated in this second study are listed below.

1. Do you think business-context gets lost in the enormous amount of information?
2. In your opinion, can your decision-making benefit from unstructured data?
3. What are the challenges in fitting the data to the business context before it gets used for making business decisions?

Similar to the previous study the aggregated quantified value of the responses to the three questions listed above were then regressed against the seniority level of the participants to understand how the level of severity changes by seniority.

The other two questions pertain to “Context” in business decisions. Businesses are inundated with so much data that it is easy to get overwhelmed with datasets distracting the organisation from its main goals. There are other factors other than the amount of data, which impacts the organisation from the “contextual-loss”. High frequency (velocity of data) and complex (mixed format data), datasets require special tools and skills to combine, comb, and analyse data for the goals (context) the organisation wants to achieve. However, the complexity around handling the data, and then extracting insights, might take so long the context is forgotten or becomes obsolete.

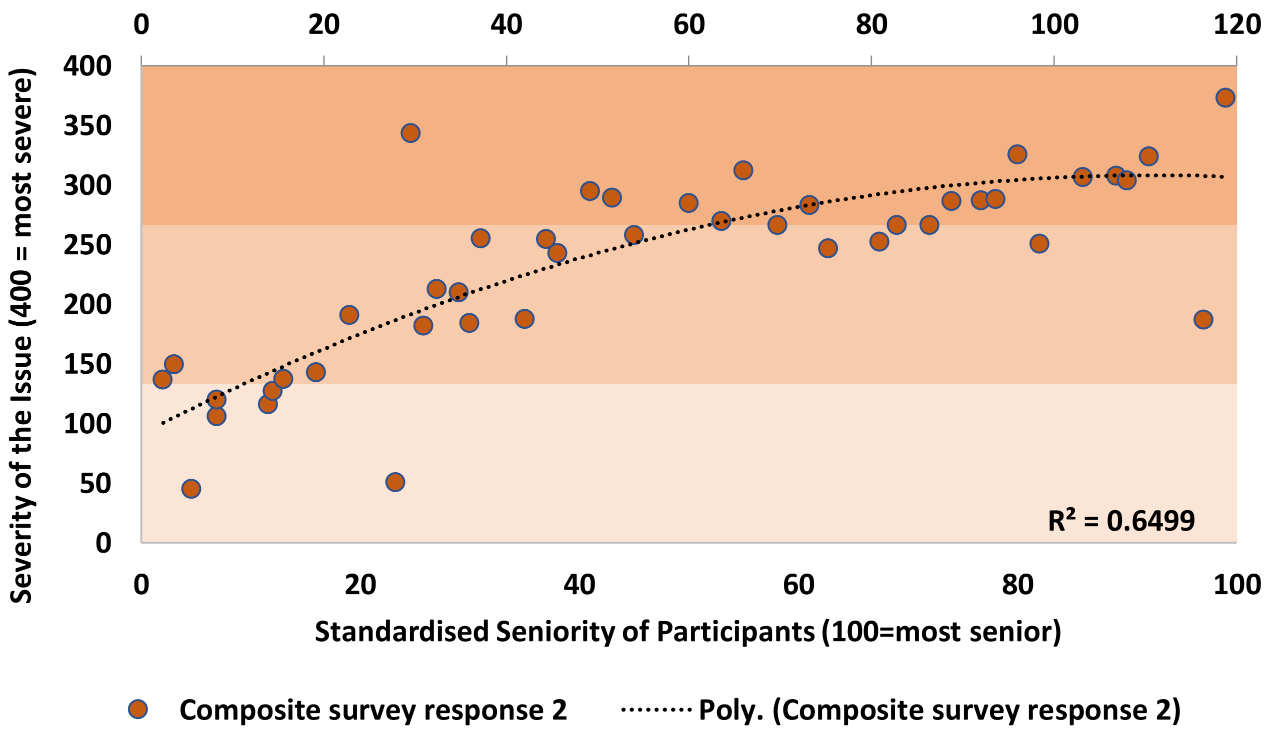

Figure 38: Regression analysis of Composite Survey Response set 2

This loss of context was quantified by the second aggregated response from the three questions discussed earlier. Again, as applied in the first regression study, the severity of the aggregated response from the survey was overlaid with the experience of the participants. Like the previous study, the aggregated severity of the scores from the three questions was treated as a dependent variable and regressed against the normalised score representing and quantifying the seniority of the respondents.

The Y-axis on Figure 38 shows the aggregate severity of the issues highlighted by the three questions highlighted earlier in this section. The X-axis quantifies the seniority of the participants facing the issues based on three questions; as listed below.

1. How long have you been involved in the industry you currently work in?
2. At which management level are you currently working?
3. Are you responsible for making firm-wide or department-wide decisions in your firm?

The aggregate seniority based on these three questions were scaled from 1 to 100, where 100 is the person(s) with the most experience. Both the independent variable, the aggregate severity, and the dependent variable, the seniority of the participants, was then plotted on the scatter plot. A polynomial regression trend line was then generated to fit the dependent and independent data. The coefficient of the regression line (R2) at 0.65 suggested a strong relationship between the independent and the dependent variables. In other words, the severity of the issues discussed increased with the seniority of the participants.

Regression analysis, in general, highlights the direction of the relationship and the strength of that relationship (Bruce & Bruce, 2017). In the earlier two regression studies, a multiple regression analysis was used to find the impact of some issues (independent variables) on the dependent variable (seniority of the respondents). Although the regression analysis explicitly highlights the relationship between a few variables, the impact of these variables on business performance can be implicit and often profound. For example, in the second regression analysis, a positive relationship between loss of context and seniority was established. In other words, many senior stakeholders experience the inability to extract timely insights from the vast amount of data, given all the challenges in breaking down the complex datasets into easily ingestible sub-sets.

Part 2: Qualitative Data Analysis of the semi-structured interview

Preparation for analysing qualitative data

Post completion of the semi-structured interviews, the transcripts were cleaned, sifted through and analysed using the CASDAQ (NVivo) to find overarching themes. Before examining the themes identified by this research, this section describes why thematic analysis was the chosen type of qualitative analysis and the process of a thematic analysis of the data collected through the interview process.

Discovering themes is the basis of most social science research (Ryan & Bernard, 2003). Qualitative approaches are unbelievably diverse, complex and nuanced (Holloway & Todres, 2003), and thematic analysis is often seen as a foundational method for qualitative analysis. Many experts believe it is the first qualitative method of analysis that researchers should learn (Morse, 1994), as it provides core skills useful for conducting many other forms of qualitative analysis. Holloway and Todres identified "thematising meaning" as a shared generic skill across qualitative analysis (Holloway & Todres, 2003). Following this, Boyatzis characterises it not as a specific method but as a tool to use across different methods (Boyatzis, 1998). Along the same lines, Ryan and Bernard locate thematic coding as a process performed within major analytic traditions (such as grounded theory), rather than a specific approach in its own right (Ryan & Bernard, 2003).

Thematic analysis is a method for identifying, analysing, and reporting patterns and themes within data (Bazeley, 2013). It helps in organising and describing the data set in rich detail. However, more often than not, it helps interpret various aspects of the research topic (Boyatzis, 1998).

While, thematic analysis is widely used, there is no concrete definition of thematic analysis and the methodology behind it (Attride-Stirling, 2001). In many annals of research, thematic analysis is often not explicitly claimed as the method of analysis, when, in reality, considerable analysis is essentially thematic. Thematic analysis is claimed as something else, such as discourse analysis, or even content analysis (Meehan, Vermeer, & Windsor, 2000) or not identified as any particular method at all – for example, data were "subjected to qualitative analysis for commonly recurring themes" (Braun & Wilkinson, 2003). The problem with the analysis in terms of evaluating research or comparing and synthesising a typical study with other studies in the same topic is hard to evaluate how researchers went about analysing their data, or what assumptions informed their analysis (Attride-Stirling, 2001).

The thematic analysis differs from other analytic methods that seek to describe patterns across qualitative data – such as thematic discourse analysis, thematic decomposition analysis, interpretative phenomenological analysis (IPA) and grounded theory. Although IPA and grounded theory are theoretically bounded, these tend to seek patterns in the data. IPA is closely linked to a phenomenological epistemology (Eatough & Smith, 2008), which gives experience primacy (Holloway & Todres, 2003), and is about understanding people's everyday experience of reality in great detail to gain an understanding of the phenomenon in question (McLeod & Balamoutsou, 2001). It is interesting to note that grounded theory comes in different forms (Charmaz, 2002). A grounded theory analysis aims to generate a plausible and valuable theory of the phenomena that is grounded in the data (McLeod & Balamoutsou, 2001). However, the grounded theory seems to be used increasingly in a way that is essentially grounded theory "lite/light" - as a set of procedures for coding data very similar to thematic analysis. Such analysis does not appear to have the theoretical commitments of a "full-fat" grounded theory, which requires analysis to be directed towards theory development (Holloway & Todres, 2003).

On the other hand, thematic discourse analysis is used to refer to a wide range of pattern-type analysis of data, ranging from thematic analysis within a social constructionist epistemology which is where patterns are identified as socially produced, but no discursive analysis is conducted. This entails forming analysis very similar to the interpretative collection form of decomposition analysis (DA) (Clarke & Baniassad, 2005). Thematic decomposition analysis (Stenner, 1993) is a specifically named form of 'thematic' discourse analysis which tries to identify patterns within data and theorises language as components of meaning.

These methods find recurring themes or patterns across the entire dataset rather than within a data item, such as an individual interview, as in the case of biographical or case-study forms of analysis such as narrative analysis (Murray, Camic, Rhodes, & Yardley, 2003; Riessman, 1993). The methods more or less overlap with thematic analysis. As thematic analysis does not require the detailed theoretical and technological knowledge of approaches such as grounded theory and DA, it can offer a more accessible form of analysis, especially in early qualitative research study.

In contrast to IPA or grounded theory, including sub-methods like narrative, discourse or DA, thematic analysis is not linked to any pre-existing theoretical framework, and hence it can be used within different theoretical frameworks. Thematic analysis can be an essentialist or realist method, which reports experiences, meanings and the reality of the participant (Braun & Clarke, 2006). Thematic analysis can also be a constructionist method, which examines how events, realities, meanings, and experiences are the effects of a range of discourses operating within society. Thematic analysis can also be a contextualist method, slotting in between the two poles of essentialism and constructionism, and characterised by theories such as critical realism (Willing, 1999). This approach acknowledges the way individuals make meaning of experiences, and, in turn, the way the broader social context impacts those meanings, while retaining focus on the material and other limits of reality (Corker, 1999). Therefore, thematic analysis can be a method which works to reflect reality and to unpick or unravel the surface of reality. A good thematic analysis helps make this investigation transparent and let it speak for itself (Fereday & Muir-Cochrane, 2006).

Thematic analysis is conducted in stages as described in Table 11. This thematic analysis process is further detailed below.

Phases	Stages
Initialisation	Reading transcriptions and highlighting meaningful units; Coding and looking for abstraction; Writing reflective notes
Construction	Classifying; Comparing; Labelling; Translating & Transliterating; Defining & Describing
Rectification	Immersion and distancing: Relating themes to established knowledge; Stabilising
Finalisation	Developing the storyline

Table 11: Phases and associated stages of themes from qualitative content and thematic analysis

Phase 1: Becoming Familiar with the Data

During a typical qualitative analysis, the researcher may either collect the data or source it from a third-party. If the data is collected through interactive means, the researcher comes to the analysis with some prior knowledge of the data and possibly some initial analytic interests or thoughts. Regardless, it is crucial for the researcher to be immersed in the data to the extent that he is familiar with the depth and breadth of the content. Similar to the banking industry in which KYC (know-your-customer/client) is an important aspect of conducting business, in the business of qualitative analysis, it is essential to know the data.

Phase 2: Generating Initial Codes

<table>
<tr><th>Types of Codes</th><th>Example</th><th>Extracted code</th><th>Principles of coding</th></tr>
<tr><td>Conceptual code</td><td>I know the concept of Big Data and what the benefits are, but I know nothing about how to use it in practical sense</td><td>Lack of knowledge of Big Data</td><td rowspan="5">In line with the reductionist nature of qualitative data management, the researcher converts large set of data into smaller and more manageable codes;
Coding leads to breaking data into common incidents and patterns by examining similarities and differences;
The coding process is a cyclic process and researcher's effort dictate the level of abstraction:
Investigator triangulation as independent coding and analysis of some of the data by two or more researchers might be ideal to enhance the data</td></tr>
<tr><td>Relationship code</td><td>Instructor's presence with students in clinical placement is necessary to make the collaboration of students in medication administration in clinical practice possible</td><td>Necessity of instructor's supervision in medication education</td></tr>
<tr><td>Participant perspective code</td><td>I believe that managers are fully able to check the accuracy of the analyst’s interpretation of the data to make decisions</td><td>Positive attitude towards data accuracy</td></tr>
<tr><td>Participant characteristic code</td><td>Although I am not a technical person, I would love to understand the business intelligence which lays latent in my business to make better decisions</td><td>A non-technical manager's eagerness to understand and assimilate</td></tr>
<tr><td>Setting code</td><td>In critical care settings, I have been provided with more opportunity to administer medication</td><td>Critical care settings</td></tr>
</table>

Table 12: Examples of coding *(Polit & Beck 2010)*

Coding, as the process of data reduction, is an element of data organisation in most qualitative approaches (Burnard, Gill, & Stewart, 2008). To simplify coding, different types of codes are recognised in qualitative content analysis and thematic analysis (Table 12)(Polit & Beck, 2010). Conceptual code identifies key elements, domains and dimensions of the study phenomenon; relationship code identifies links between elements, domains and dimensions;

participant perspective code identifies the participant's positive, negative, or indifference comments about a particular experience; participant characteristic code and setting code show the general characteristics of participants and the place in which the phenomenon has happened, respectively (Vaismoradi, Jones, Turunen, & Snelgrove, 2016). Such a classification not only helps researchers organise codes but also enables detailed comparison and sub-classification before the subsequent analytical steps.

Phase 2, which involves coding, begins when the researcher reads and becomes familiar with the data, and generates an initial list of ideas about what in the content of the data and why it is interesting. Then, phase 2 involves producing initial codes from the data. Codes identify a feature of the data including semantic or latent content that appears interesting to the analyst, and refers to "the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon" (Boyatzis, 1998).

Phase 3: Hunting and Mining for Themes

Phase 3 begins when all data is initially coded & collated, and the researcher has a long list of the different codes which have been identified across the data set. This phase, which re-focuses the analysis at the broader level of themes, rather than codes, involves classifying different codes into potential themes and collating all the relevant coded data extracts within the identified themes. The researcher analyses the codes and considers how different codes may combine to form an overarching theme. It may be helpful in this phase to use visual representations to help sort the different codes into themes. This analysis uses a combination of tables, mind-maps, word clouds, word trees and cross-matrix tables.

Phase 4: Revising Themes

Phase 4 begins when the researcher has devised a set of candidate themes, and begins refining those themes. During this phase, it became evident that some candidate themes are not really themes while others might collapse into each other (Braun, Clarke, Hayfield, & Terry, 2019). Other themes might need to be broken down into separate themes (Patton, 1990). Data within themes should meaningfully bind together, while there should be clear and identifiable distinctions between themes. This phase involves two levels of reviewing and refining the themes. Level one involves reviewing at the level of the coded data extracts. Level two involves refining the themes when new insights are found in the coded data. This phase was an iterative process in which both codes and themes were iteratively identified and enriched within the qualitative analysis process.

Phase 5: Defining and Naming Themes

Phase 5 begins when the researcher has a satisfactory thematic map of the data (Braun & Clarke, 2006). At this point, the researcher then further refines the themes presented for the analysis. Define and refine means identifying the 'essence' of each theme (Bloomberg & Volpe, 2018) and determining which aspect of the data each theme captures. When refining themes, it is important that themes do not have a high level of diversity, complexity or diversity. The researcher does this by going back to collated data extracts for each theme and organising them into a coherent and internally consistent account, with accompanying narrative. It is vital that researchers identify interesting aspects of data as well as why it is interesting rather than just paraphrase the content of the data extracts. This way the interesting aspects of the data will be objectively analysed rather than treated subjectively to fit the need of the research.

Phase 6: Producing the Report

Phase 6 begins when a full set of themes has been defined; this phase starts the final analysis and write-up of the report. The task of writing-up thematic analysis, whether it is for publication, research assignment or dissertation, is to tell the complicated story of the data in a way which convinces the reader of the merit and validity of the analysis. It is essential that the analysis provides a concise, coherent, logical, non-repetitive, and interesting account of the story the data tell – within and across themes. The write-up must provide sufficient evidence of the themes within the data.

Identified Themes

In this study, five major themes were identified through the semi-structured interviews and many of these themes emanated from analysing the qualitative data. For example, Table 13 shows the most frequently used words across all the interviews.

Word	Length	Count	Weighted Percentage
data	4	655	8.25%
making	6	134	1.69%
times	5	132	1.66%
decision	8	120	1.51%
information	11	88	1.11%
company	8	75	0.94%
people	6	69	0.87%
challenges	10	64	0.81%
process	7	64	0.81%
organisations	13	50	0.63%
system	6	50	0.63%
different	11	48	0.60%
capturing	9	46	0.58%
business	7	42	0.53%

taking	6	41	0.52%
customers	9	40	0.50%
costs	5	39	0.49%
points	6	38	0.48%
experiences	11	36	0.45%
example	7	35	0.44%
managers	8	35	0.44%
issues	6	35	0.44%
create	6	34	0.43%
sources	7	32	0.40%
quality	7	31	0.39%
important	9	30	0.38%
better	6	29	0.37%
getting	7	28	0.35%
terms	5	28	0.35%
working	7	27	0.34%
industry	8	27	0.34%
able	4	26	0.33%
years	5	26	0.33%
understand	10	26	0.33%
market	6	25	0.31%
problem	7	24	0.30%
increasing	10	23	0.29%
change	6	22	0.28%
investments	11	22	0.28%
sort	4	22	0.28%
trying	6	22	0.28%
intelligence	12	21	0.26%
structured	10	21	0.26%
collect	7	21	0.26%
Cloud	5	21	0.26%
applications	12	21	0.26%

Table 13: Table of words most frequently used in the interviews

Each theme is discussed in detail below accompanied by the visualisations, where appropriate. These five themes were based on the recurrence of the topics which were most commonly highlighted by the interviewees. There were many other themes which emerged in the post-interview analysis. However, these five were overarching and materialised with almost every respondent. For instance, every respondent, without being led or framed, highlighted the fact that “data” is indispensable for decision-making and survivability and hence, data becoming the fourth factor of production was chosen as a theme.

Theme 1: Data is the fourth factor of production

Classical economics, made popular by Adam Smith and David Ricardo (Stirati, 1994), focused on physical resources in defining its factors of production and discuss the distribution of cost and value among these factors. The main factors of production to produce something of economic value (or output) include land, labour, and capital. Almost every organisation that wants to produce or create service of economic value is built upon a combination of these three factors of production. However, as globalisation extends to the corporate world, an additional factor of production, that of data, is quickly emerging (Hughes, 2012). A plethora of theory and literature has chronicled the onslaught of data and its growth, especially during the past decade. The use of data, and its impact on corporate decision-making, is overwhelming (Senge, 2006). A recent study by International Data Corporation (IDC) sponsored by EMC (Tucci, 2010) predicts the amount of digital information created annually will grow by a factor of 44 by 2020, as all major forms of media – voice, TV, radio, print – complete the journey from analogue to digital (Farmer, 2010). Most organisations, irrespective of size, are thrust into an era in which data and information are pivotal to survive into the future.

Harnessing information as a primary factor of production means organisations need to recognise and effectively plan for the four “V’s” of data: volume, velocity, variety, and veracity (Unhelkar, 2017). Volume pertains to the sheer amount of data being digitised, maintained, secured, and then utilised. Understanding the organisation’s current needs and having a plan for utilisation is fundamental to the future growth of the business. Velocity refers to the speed at which data is ingested, moved, stored, transformed, managed, analysed or reported on in order to maintain a competitive advantage (Firica, 2017). This velocity of data varies by organisation and industry or sector. It also varies by applications used. Within the same organisation, different departments may use data with a different velocity. Variety points to different types of data – audio, video, graphics, and sensor. This difference must be well understood because competitiveness requires access to the right types of data more than ever. Date types can range from age-old flat files with structured data to spatial and unstructured data, which makes it difficult to factor these disparate formats into the decision-making process. Veracity is the truthfulness or quality of data. Lack of veracity in data leads to poor understanding and decision-making in business. Lower veracity of data can lead to incorrect analytics and, eventually, incorrect decisions. Higher veracity in data, however, enables deep insights which fuels business agility and helps generate new ideas. Data quality is the most critical frontier in business for successful and holistic decision-making. Mastering the four Vs puts the business in a much better position to use information broadly for a competitive advantage (Unhelkar, 2017).

Broad usage of information is key to making timely and optimal business decisions. Information needs to be ingested at a faster rate due to the changing competitive landscape. The higher the latency to insight and decisions, the more likely a competitor can outmaneuver, using the advantages of time and information to their benefit. This issue of latency and its adverse impact on decision-making was evident from the responses of the interviewees from the first two questions posed during the semi-structured interviews.

1. What is the importance of data in your decision-making as compared with your (and your employees') experience and intuition?
2. Is your organisation actively increasing its Data capture? Please provide examples of new types – e.g., unstructured and various sources of data?

Responses to these two questions from the respondents were wide ranging and not narrowed to the hypothesis of this research. Respondents provided great discussions on the pros and cons of using data in decision-making as well as the merits of continued human intervention when data by itself could not be used to make long-term strategic decisions. Respondents specifically discussed the flaws of decision-making based entirely on intuition and experience. 93% of the respondents thought data is increasingly more important in decision-making. Many of the respondents mentioned that their organisations already incorporate data-based decision-making on a daily basis; others used the power of data in decision-making on an ad-hoc basis. However, all the respondents unanimously agreed that decision-making based on gut feeling is dissipating faster as Big Data-driven analytics comes to the fore. The sentiment is captured in these responses from the respondents;

"Data is the key and the biggest asset that gives me the best return on investment."

> *"In my business, I deal with B2B customers. I would do anything to have any kind of data to quantify their behaviour to create patterns. This way I will be able to make sound and informed decisions. Data is not available readily and if it is then it is not in a format, I can use readily to make business decisions."*

The respondents were a group of very experienced people with an average experience of 26 years which ranged between 10 to 42 years. The respondents were also aware of developments in the field of data science. These respondents were moving away from traditional thinking when it came to making decisions. They cited a high awareness of the growth and use of data in the current business environment. This up-to-date knowledge seems to be driving their openness to embrace the change when it came to decision-making. This openness was evident across almost all respondents. The more experienced the respondents, the higher the understanding of the latent power of data. These respondents were also aware that they are not utilising the data within their organisation to its full potential for decision-making. While responding to the second question (which asked if they were capturing more data), the respondents also emphasised the fact that the world is getting quantified in various ways. Every instrument in day to day business is getting a sensor embedded (IoT), or trends are monitored which were previously overlooked. Furthermore, these customer patterns, based on the data, are getting quantified in an unprecedented fine-granular level (Agarwal, Govindu, Ngo, & Lodwig, 2016).

As illustrated in Table 8, interview respondents hailed from different industries. Despite the diversity in experience, industry and domain expertise, data was explicitly and implicitly embedded in their vernacular, especially concerning decision-making in a business environment. A word frequency query was run on the entire set of responses, which included more than 300 hours of audio and transcription. The result of this query is graphically shown as a word-cloud (Figure 39) and the word data dominated the conversation the most. Although the hypothesis of this research is structuring data for integration, at no point in the interview process were the respondents influenced to talk or respond in favour of data. Terms and codes associated closely around the data shows the implicit link of data and its associated terms – including data in decision-making, data in a timely manner, the cost associated with data, different data sets, and data transformation into information before becoming intelligence.

Figure 39: Word frequency chart using all responses

Thus, it was apparent even in the first two questions in the interview process that data plays an essential role in respondent day-to-day decision-making. The respondents who were already using data extensively were trying to seek out more data and those who were using an enormous amount of data wanted to make sense of it. Ease of integration across multiple knowledge domains was also important for these respondents.

Theme 2: The path from data to business intelligence is not straight-forward

Business intelligence is pivotal for success for any corporation in this era. Data-driven business intelligence empowers organisations to improve productivity, enhance the decision-making process and regularly make strategic adjustments to ensure business goals are being met (Davis, Edgar, Porter, Bernaden, & Sarli, 2012). BI and data-analytics allow key stakeholders to learn more about critical areas of their business, competitors' businesses, industry, and audience. BI helps provide better visibility into performance effectiveness of the entire organisation's operational processes (Ranjan, 2008). Theme 2 emanated from the responses to questions three and four during the semi-structured interview process. The questions posed to the respondents were:

1. What are the key challenges you face in deriving business insights from data transformation?
2. What are the key challenges in undertaking Data-driven decisions?

The mind-map (Figure 40) visually represents the core of this theme and maps the keywords originating from the discussion around decision-making with the interview respondents. The first question brought up the types of challenges the respondents have experienced when transforming data into insights. Also, the second question highlighted what respondents thought the main challenges in undertaking data-driven decisions were. Both these questions were somewhat interrelated, which was evident from the responses of the interviewee/respondents.

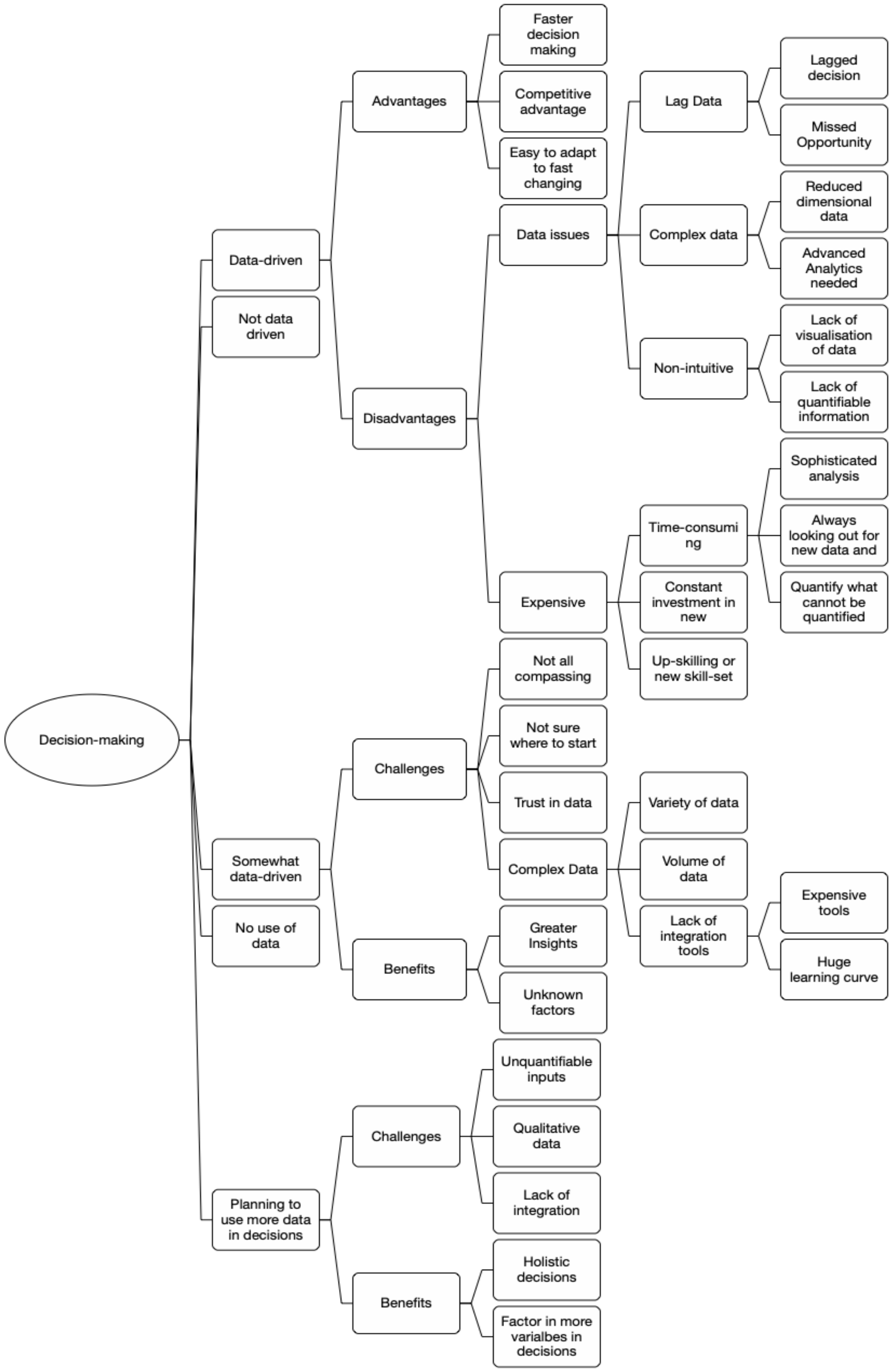

Figure 40: Mind Map of decision-making as a node

As shown in Figure 40, every sub-node (i.e. related keywords) emanating from the key node, decision-making, presented challenges at some point. These challenges were complex and included objective and subjective challenges discussed in detail next.

1. **The complexity of more data sources:** Typically, more data equals more sources (Klein, 2001). It can be challenging to sort out the information and decipher what will be most helpful for a particular action. An average BI user regularly works with no fewer than 30 distinct data sources, according to a recent study from Aberdeen (Techwire, 2016). Some 40% of BI-sourcing businesses need platforms to analyse unstructured data. An example of such multiple sources is social media. Social media can provide entirely different insights about customers than data from web analytics. People on social media do not always play the role of a typical customer as compared to people on an e-commerce website. People leave breadcrumbs of their personality or lifestyle on social media when they like a video or provide a thumbs-up to a friend's new pair of shoes. Therefore, social media provides a different set of insights versus plain vanilla web analytics. There are lots of social media platforms (e.g. Facebook, Twitter, Instagram, Etsy), which mean multiple sources – the same goes for web analytics. The more data sources one has, the more insights can be extracted (McAfee, Brynjolfsson, Davenport, & Barton, 2012). However, adding more data also makes it harder to manage, unless the organisation has a platform that can handle the additional data. When complex data is handled the right way, it delivers significant value. Developing a way to make use of the data and providing multiple departments easy access is pivotal for data-driven decision-making. This explosion of data from

various sources was one of the underpinning issues highlighted by most of the respondents during the interview process.

2. **Data Quality:** Every respondent raised the issue of how data quality is the most crucial aspect to consider when deriving business intelligence. Most highlighted the point that if data quality is actively managed then decisions made on the insights generated could provide optimal results. On the flip side, most respondents also pointed out that bad quality data results in wrong insights and eventually leads to bad decisions.

Data quality issues are not just central to the topics emphasised by the group of respondents in the research. There is a growing trend in which businesses are increasingly aware of the potential of their data. An increasing number of businesses see data as an asset which can be used to harness business intelligence, and in turn get closer to customers and provide better products, services, experiences and value. In a study conducted by Experian in 2015, CIOs stated they could increase their organisations' profit by an average of 15% if their data were of the highest quality (Experian, 2015). Organisations who proactively manage data as an asset together with data quality are the ones that will reap its full strategic value (Experian, 2015). The same research also highlighted that more than 92% of the organisations find data quality challenging in some way.

Most of the respondents believed that with the explosion of the volume of data (Big Data), quality control needs to evolve to keep up the same level of assurance. However, few highlighted that with Big Data it is vital to understand the context of the data being analysed. For example, in a controlled medical trial involving 100 volunteers, the effect of the common cold can be analysed and treated appropriately. The effect of the common cold can be carefully studied, and the spread of the virus can be understood in greater detail. However, the same kind of information can be obtained by analysing Google searches during the winter season. Millions of Internet users suffering from the common cold tend to search for remedies and medication online. However, not everyone who is searching for that medication might be affected by the common cold. Even if 50% of those Internet users are affected by cold, the sample size is far greater in number than the 100-person medical trial group. Hence, the challenge is to understand that although Big Data might not be accurate enough, its use needs to be put in the context relative to the insights which are being derived. This dilemma of greater dataset vs. accuracy was highlighted by respondents during the interview process.

One of the respondents highlighted different types of quality problems. The first type is human typographical errors. This first kind can be classified as a mistake which drove the bad quality of the data. The second type is when a person completely forgets to enter it in a database/spreadsheet, or the data is added or entered in the wrong place. This second kind of quality issue is when somebody or some system adds/puts the wrong value. For example, one of the respondents gave this instance of the bad quality of data which is unintentional.

> *"If a certain system is looking at different types of dog, and they call it a poodle, it may be because they do not know any different. They think it is a poodle it may be that the picklist provided does not support that type of dog. So, they*

have just thrown in the one that they think is closest to it. Sometimes they genuinely do not know. Other times, what one calls a poodle in Australia is called something else in Malaysia. And so how are you going to match that across?"

This semantic and contextual problem is a classic example of a data issue an organisation encounters when interacting with external systems. Other issues with data quality stem when many applications are developed for different departments, but with no interaction or integration between each application. The issue is to understand how these disparate applications can talk to each other and match common elements to help gather deep insights. Businesses sometimes face smaller, data-related problems which are hard to tackle, let alone complex systems which factor in numerous variables. For example, a business which needs to uniquely identify an employee by matching the first name and family name, finds multiple "John Smiths". Even if the business just has structured data, like date of birth, in one application it might be "date_birth", and in another application, it might be in an entirely different format (e.g., birth_date, date_of_birth, DOB). These were some of the challenges highlighted by the respondents during the interview process. Almost two-thirds of the respondents (63%) highlighted the lack of a coordinated, centralised approach to data quality within their organisation, with others taking a siloed approach to data quality management. Thus, data quality issues were at the top of the challenges highlighted by the respondents in this survey.

3. **Data is different to different organisations:** This sub-theme was highlighted by a few respondents and came out as an outlier to the other themes or sub-themes highlighted in this section. The core of this sub-theme can be summarised through a comment made by one of the respondents as:

   > *"To stay ahead of the game, we are always trying to find different ways to analyse the data to provide newer insights to our clients."*

   Respondents highlighted that the same dataset is used differently by different people in their organisation. An example is how customer sales data is analysed by the finance department to forecast sales and earnings estimates for a given fiscal period, and also used by the R&D department to help design the next product or service. On a broader scale, businesses in the same industry with access to the same data, end up using it uniquely and analysing it differently. The one that finds the most value or derives a unique insight advantage most likely to benefit from the data.

   Another way to look at the same issue is to examine how businesses develop different metadata for the same dataset. Metadata is data about data. It is the abstracted information about data that resides in information systems (Even, Shankaranarayanan, & Watts, 2006). Numerous studies have theorised that providing metadata, especially quality and process metadata, to end-users helps them better assess the data quality in the context of the decision task. Quality metadata consists of measurements such as processes and task accuracy, and completeness. Quality metadata also is referred to as data tags (Wang, 1993), or data quality information

(Fisher, 2003). Providing users with such metadata during the decision process can improve decision outcomes (Chengalur-Smith, 1999).

Process metadata (Marco, 2000) captures information about data creation and delivery, such as sources, processing methods, storage, and end-usage targets (Shankaranarayanan, 2004). Providing such metadata to business users was argued to improve their ability to assess data quality and thus enhance decision-making (Shankaranarayanan & Watson, 2003). However, the context of the business-goals drives the depth to which this metadata needs to be defined. As an example, the metadata for a video file created by a musician might have data about the instruments, tone, lyrics, and the genre of music. For the musician, this metadata might be sufficient for selling the album to prospective record businesses or promoters. However, the same video might have different metadata if it was going to be used in an online music course. The context of the business goals defines the metadata which needs to be linked to the data asset. Also, businesses competing in the same domain or industry often have access to the same marketplace data. Access to the aggregate sales, buying patterns, demographic differences and consumer affluence drives the supply and demand of their goods and services. These businesses also have access to the same talent pool of analysts who can thus analyse data. Often anomalies are seen in decision-making from these businesses leading to entirely different business results. A good example of this anomaly can be found in the consumer goods industry in Australia, which is dominated by two businesses, Coles and Woolworths. The differences in their business performance can be attributed to the way they uniquely value their process, and the data which drives it.

Today's marketplace is not a single market economy (Brønn & Vrioni, 2001). Economies are connected directly or indirectly. Data in isolation does not provide any insights. One respondent said:

> *"I get paid by reaching the set of data I can use and connect to create a story. It is all about bringing perspective to the decision makers, and it is all about providing insights from data and provide the trend for the future."*

Another respondent explained in detail:

> *"How am I going to use the same set of data or similar kinds of data to show and predict the future and see from the perspective of decision makers especially when they do not have the visibility of data. The story must be visible, as what the decision-makers want to drive is just difficult to understand from an individual perspective. The questions I have is who is managing and providing the data, which of the key information they need to pick, how are they sequencing and connecting these raw data sets. I have to create a story,*

create an information-set so that the decision maker or the stakeholders can see the relevance and see it from their perspective".

Thus, the use of data defines the metadata which is needed to learn more about that data in an iterative way, as shown in Figure 41.

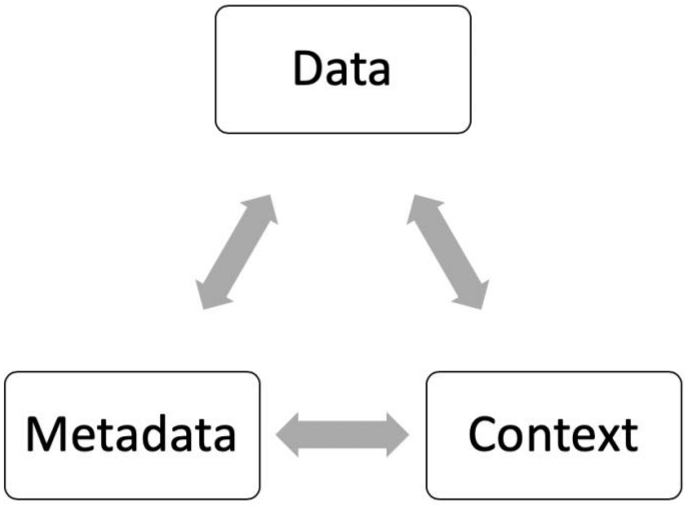

Figure 41: The iterative process of creating metadata based on context

4. **Volume, Variety, Veracity and Value of Data:** The V's of data were discussed in this thesis earlier as being volume, variety, veracity and value. During the interview process, it was found that each respondent highlighted each of these Vs in context relating to their work. For example, one of the respondents highlighted the enormous volume of data relevant to investment banking. The respondent specifically highlighted the activity of trading in the stock market and the fact that market data now encompasses more than fifty different stock exchanges. Over the years, the volume of market data has gone up, the number of exchanges and the number of

exchange products have greatly increased. On top of this, there are more businesses in the market. This increased competition has created enormous pressure on existing market participants to use data to differentiate themselves from the norm. One respondent said:

> *"We are really only limited by our imagination as to what we can do with it (data)."*

With the storage of data becoming inexpensive, the traditional way of analysing data internally in an organisation is challenged. While it is sometimes easy to access a huge volume of data, it may not be useful. Hence the challenge is to identify what is quality (veracity) data amidst a massive amount of data. When it came to veracity, one respondent added:

> *"The challenge is not just that the data is of low quality, the challenge is actually identifying what is high quality. How do you identify what high-quality data is? Simply identifying where the data is, what are the possible sources of data is not sufficient. This is because we find that even if we go looking for new types of data, it is not necessarily immediately obvious where it is or what types of data we should be using."*

Another challenge was the variety of data and respondents highlighted that a lot of the new type of data is unstructured. In the past, structured data has been straightforward to use and easy to manipulate on a single computer. However, as the search for new and differentiated data increases, corporations are trying unique ways of quantifying data which is or was difficult to quantify. More often than not, the search has yielded data in an unstructured format. The respondents highlighted the challenge of turning the unstructured data into more structured data to allow further analysis.

Businesses at the forefront of business intelligence have started using advance analytical algorithms, including Natural Language Processing (NLP) in conjunction with Machine Learning or other AI-based tools. This usage also closely aligns with the last V on the list – which is Value. Respondents highlighted that although volume can be handled given the cost of storage has gone down in the last decade, investing in appropriate tools and skill set is challenging. This challenge emanates from the fact that the Value extracted out of this enormous Volume and Variety of data might not yield Valuable insights. Having access to all new tools and new or differentiated datasets does not guarantee value. According to the respondents in this interview, successful businesses need to know what data is meaningful and align business goals to what they seek from data. With capital resources, businesses can efficiently manage the tools and data needed to create insights; however, business needs to decide which insights are necessary for success. This point was also summarised by one of the respondents as:

> *"There is so much out there that it seems like we are always in the unknown. However, if I don't make up my mind on what I need, then I will always be in that abyss."*

5. **Business Intelligence is a costly exercise:** The practice of converting raw data into actionable business intelligence, which is optimal for business decisions, involves significant investment throughout the BI process. The BI process, from development through implementation, deployment and maintenance incurs a significant cost at an escalating rate. According to the research conducted by IDC, most of a typical BI solution spanning at least three years is staffing related (Analytics, 2007). IDC grouped the Business Intelligence costs into five major categories, as shown in Table 14. These research papers by independent research agencies suggest that BI is highly resource-intensive.

#	BI Process	Proportion of Costs
1	Annual Staffing	37%
2	Software	25%
3	Hardware	15%
4	Initial build	13%
5	Downtime	10%

Table 14: Distribution of cost in traditional BI *(Analytics, 2007)*

The cost of Cloud computing and storage has gone down over the years. However, the cost of understanding the data and deriving insights has gone up. One respondent aptly put this as:

> *"Data-driven business intelligence environment has increased my cost four times more than what I used to do before using traditional means."*

According to the respondents, the cost of data, the analyst, and the tools have increased over time. To keep pace with changing technology, respondents emphasised that changing toolsets and skillsets to match what is required is a challenge. Respondents highlighted that business stakeholders who understand the value of data also require increased resources, including time and capital. These stakeholders are ready to deploy these resources if the potential benefit of these insights can be matched to the future growth of the business. According to the respondents, data and the tools needed to crunch that data is continually evolving. One respondent appropriately described the challenge his business is currently facing:

> *"So, one of the challenges is to work out which tools to use, whether you should use natural language processing, whether you should use machine learning. And there were so many different types. Which one should you use? Whether you should use A.I.? There are just so many tools available, and there are so many different ways of doing the analysis, which is often very standalone and doesn't talk to each other. This makes it very difficult even to get started."*

Respondents highlighted tools varied from various fifth-generation programming languages to stand-alone PaaS (Platform as a service), SaaS (Software as a service) and IaaS (Infrastructure as a service). These "as-a-service" offerings are often more economical than a stand-alone application; however, the implementation is challenging due to the fact they need to fit in with the existing systems architecture. The "as-a-service" business model has helped businesses in the B2B space generate constant revenue from their products. These service models are being eyed by consumer businesses hungry for income that lasts beyond the initial product purchase. Through "servitisation" (Vandermerwe, 1988), i.e. combining products with services, businesses can innovate faster and deepen relationships with customers by providing more value. Value includes data insights derived from internal and external structured data sources, IoT-powered devices including thermostats to wind turbines, and real-time supply chain. AI and other emerging technologies are driving the growth of “as-a-Service” models because it helps businesses quantify the value of services. In a survey conducted by DJS Research, three-fourths of senior managers and C-level executives surveyed believed “as-a-Service” has positively influenced their business model, products and service (DJS Research, 2018). Most respondents in this research conveyed the same message by emphasising the impact of these offerings. Respondents also highlighted how these services have helped them push beyond traditional markets, expand internationally, and increase revenue. Many respondents have seen customer retention and personalisation of services rise incrementally. However, the respondents emphasised that this path to business intelligence was a costly one.

Another challenge that respondents highlighted was related to hardware. One respondent said,

> *“As I talked about, the challenge of the quantity and volume of data introduces the need for new tools and hardware. If you can't do it on a standalone PC, you need to have access to cluster computing and Cloud computing. And that introduces all sorts of challenges.”*

The above statement touches on a few concepts, each of which has its challenges. For example, businesses dealing with Big Data (i.e. huge quantity/volume of data) cannot perform analytics using traditional standalone desktops. The need for processing power is higher to crunch Big Data. This is where Cloud computing enters. Many businesses which provide data warehousing capabilities offer different flavours of Cloud computing. An organisation can either choose to have a public Cloud or private Cloud. The choice depends on the appetite of the business to deploy capital and commitment. However, the biggest challenge, as highlighted by the respondents, is the lack of understanding around the Cloud. One respondent provided a well-rounded comment on this issue as below:

> *"I proposed the use of Cloud computing to help us generate insight from this big high-frequency data we acquired. However, my upper management guys have no clue about the Cloud. They hear the news on hacking and phishing and have formed an opinion that the Cloud is not safe. However, I wonder that my business's cybersecurity is as robust as the security measures deployed by the likes of Microsoft, Google, IBM and Amazon, on their Cloud services? The biggest cost I see is the cost of ignorance."*

6. **Context gets lost in many ways.** There are many instances in which it is difficult to convert data into insights. The difficulty arises because businesses do not plan the context which drives decision-making. When organisations are presented with an opportunity to use Big Data, they tend first to tackle the issue of storing this data in vast data lakes and data warehouses. These Big Data sets have the potential to provide a 360-degree data of the customer, which can help design better products, improve serviceability and increase market share. However, organisations usually do not lay down the contextual foundation for the existence of the data and the planned usage.

Businesses underestimate the challenge of deriving insights from data due to large volume and variety of the data as well as the lack of necessary resources such as skills, time and capital. The respondents during the interview process often highlighted the need to see “meaningful results” when presented with analysis post data crunching. When requested to elaborate, one respondent said:

> *“Because you can come up with lots of different relationships using purely a data-driven analysis it does not mean any of those conclusions are reasonable or meaningful. So, the challenge is always to ensure meaningful results.”*

Another respondent highlighted the context is typically embedded in the mind of the business analyst or data scientist who is trying to analyse the data for deriving insights. The main business goals of a business, as communicated by the senior most stakeholders, get translated into smaller and more achievable goals acted on by middle-level managers. Business analysts who create the analysis have a good understanding of the big picture as to what needs to be achieved. Equipped with a hypothesis, they try to prove or disapprove issues surrounding specific business functions or provide insights which might be impactful for the business in the short term. What they create is meaningful and contextual. In the same vein, another respondent emphasised that businesses sometimes confuse causality with correlation. The respondent provided an excellent example:

> *“If event A preceded event B, it doesn’t always mean they are related events.*

And it doesn't mean that there's a causal relationship. So just because you match the volume to the number of golf memberships, it doesn't mean that you're going to increase your sales volume if you lobby more golfing people. So, there's the context to it."

Another respondent provided a psychological aspect of context concerning the decision-making exercise. The respondent mentioned that context in decision-making should incorporate the perspective of the stakeholders at all times, adding that it is not only essential to capture which decisions were made at what time, but it is also critical to capture why the decision was made relative to the perspective of the stakeholders. The process of decision-making needs to be captured from the perspective of the context. The respondents added there is a long gap between when decisions are made and when the business gets results. The success or failure of those results is then associated with the decisions made in the past. Due to the delay in seeing results, the basis on which the decisions were made or the context on which the decisions were formulated is lost. A respondent further added:

"Because if you look at the decision, ten years in future, and try to analyse the same set of information, without the context, I would have made a different choice or decision. What happens in commercial business is that you are making significant choices, and the context behind those choices or decisions is always left to the person and when the person moves on from the position that innate knowledge of the context gets lost."

Another issue emphasised by the respondents is the context of the dependencies which affect a particular decision. Multi-faceted decisions have dependencies which are not captured or are considered when the data is chosen. Decisions are made taking into account what can be factored in at any point in time. However, the process of decision-making loses information on the dependency factors. The context behind the decision gets lost. The success or failure of an individual decision is attributed to the decision-makers, and even the factors used to make those decisions. However, the context behind why a certain factor was included in the decision-making remains elusive given the context was not captured appropriately as a part of the knowledge base.

Another respondent highlighted the loss of context even in the data that businesses have been using for a long time to make decisions. For example, regulators tend to use forward-looking (i.e. forecasts) indicators to help make decisions. However, as regulators face an increasing number of issues to monitor and regulate, forecasts are no longer serving the purpose they once used to. The primary purpose of forward-looking indicators was to help regulators foresee signs of deteriorating conditions of an economic variable enabling appropriate action. Moreover, as the current economic system becomes more complex due to its interconnectedness, the correlation between forward-looking indicators to the main economic variables keep changing. To understand these changing dynamics, it is essential to put the changing behaviour in context for business decisions.

7. Consistency in decision-making and to be human

When discussing the efficacy of decision-making supplemented by data, the respondents highlighted the pros and cons of a decision-making process fully and solely driven by humans. Most of the respondents highlighted that humans are irrational when making decisions. Irrationality in decision-making has been thoroughly studied in psychology. People go wrong at every turn (Stanovich, 2009). People have difficulty figuring out what they want (Gladwell, 2005). In general, preferences are not very stable or coherent, nor is there a high degree of certainty in assessing risks and reward. One respondent said:

> *"Human decision-makers should accept that experience, emotion, and human touch are just not reliable every time, and they are not consistent."*

Taking emotions out of the decision-making process is a difficult challenge, particularly among key stakeholders during processing and analysing stages. Emotions tend to make the decision-making process less consistent or repeatable. Human traits that may lead to erroneous decisions and highlighted by the respondents, include overconfidence, base-rate neglect, sunk costs and escalation effects, representation effects, hindsight bias, and confirmation bias. Although explaining these concepts does not fall in the domain of this research, each one of these human-led concepts which are pivotal in making decisions a success or failure, are explained in brief here.

An organisation or individual is said to be overconfident in some judgment, estimate or decision if expressed confidence systematically exceeds the accuracy achieved

(Lichtenstein, 1977). This mostly occurs when the organisation achieves success in the first few decisions. This initial success manifests itself as overconfidence in the subsequent decision-making process. Overconfidence at an individual and organisation level prevents incorporating all available information, which is essential in making an overall successful decision. The second folly of humans, which at times can lead to an erroneous decision, is base-rate neglect (Tversky & Kahneman, 1980). Typically, an individual or organisation must combine two kinds of information into an overall judgment: a relatively stable long-run average for some class of events (e.g., the frequency of a particular disease in a certain target group) and some specific information about a member of that class (e.g., a diagnostic test on a particular individual). As an example, in a military setting, a commander might have a strategy for an attack on the enemy-base as a result of his past experience (National Research Council, 1998). However, information from forward-observers, based on a quick analysis of sightings can, at times, force the commander to override his initial plans. Bias can be compensated at the command level by training, or aid can be provided to the commander to overcome base-rate neglect and remain a subject of discussion in the psychology of decision-making (National Research Council, 1998).

The next phenomenon, sunk cost, is something which most decision makers might have experienced, either at an individual level or an organisational level. Some studies have demonstrated tendencies to persist in failing courses of action to which one has initially become committed and to treat nonrecoverable costs as appropriate considerations in choosing future actions (Arkes & Blumer, 1985). Most of the decision makers are so emotionally committed to the initial decisions that they tend not to evaluate their decisions based on their emotions and accept something such as the sunk cost. The inability to accept that a decision is incorrect and taking alternative action is one of the biases which adversely impact human decision-making (National Research Council, 1998).

The next concept which can adversely affect decisions made by humans is called representation effect, made popular by Prospect Theory (Tversky & Kahneman, 1992). The main point of this concept is that outcomes are evaluated in comparison with some subjectively set reference point and that preferences differ above and below this point. Since a given outcome can often be framed either as a gain compared with one reference level or as a loss compared with another, outcome preferences are vulnerable to what appears to be purely verbal effects. For example, (McNeil, Pauker, Sox, & Tversky, 1982) found that both physicians' and patients' preferences between alternative therapies were influenced by whether the therapies were described in terms of mortality or survival rates. Hence, the wording or framing of outcomes can influence how businesses make decisions (National Research Council, 1998).

The next two topics related to the learning of judgment and decision-making skills have been of interest to decision behaviour researchers and are briefly addressed here: hindsight bias and confirmatory search. Hindsight bias is the phenomenon (Fischhoff & Beyth, 1975) in which decision-makers tend to recall greater confidence in an outcome's occurrence or non-occurrence than before the fact. In retrospect, they feel that ''we knew it all along”. For decision-makers who are exposed to this bias, the primary effect impedes their learning over a series of decisions by making the outcome of each less surprising than it should be. If, in retrospect, the realised outcome seems to have been highly predictable from the beginning, one has little to learn. Confirmation bias is a second potential obstacle to learning. It is the tendency to shape one's search for information toward sources that can only confirm current beliefs, but not test the belief (National Research Council, 1998). A confirmatory search of this sort might, for example, direct a physician to examine a patient for symptoms commonly found in the presence of the disease the physician suspects. A recruiter might similarly examine the performance of the personnel he or she had recruited. Such a search often has a rather modest information yield (Klayman & Ha, 1987).

These human traits not only make humans humane but also introduce biases which at times yield sub-optimal results from business decisions. One of the respondents described this across in detail as below:

> *"I think the thesis needs to be explicit in saying that human factor and psychological factors have to layer upon the foundational data framework. Otherwise, they (humans) will miss the point again, which I have been facing in the many years of experience. I don't think it is the data, which is providing an inconsistent message, I think it is the human factor, and the environment which keeps changing. The business environment is not static, and neither are dynamics between departments within the same organisation. So, you can consider how different it would be in different countries, different regions, different political environment. I think the thesis has to explicitly factor in that*

> *the data can provide a boost in confidence and the new framework you are proposing would provide a consistent way of providing a story to the decision makers in a human way."*

Theme 3: A large proportion of time is spent on making data useful

The modern proponents of economics have highlighted time and information as the other factors of production quickly gaining popularity, in addition to land, labour and capital. Speed and time provide a crucial competitive advantage. Incorporating fast-changing data in making tactical decisions on a timely manner can help corporations stay ahead of their competitors. Timely actions also win appreciation from customers on being relevant and current. Quality of data is important; however, understanding the difference between stationary data versus data-in-motion is vital for optimal decision-making. Swiftly understanding and translating customer needs from concept to practice, in many ways, determine the success of an enterprise. Rapid technology innovation has forced the business world to keep pace. In turn, technology innovation enables all organisations to compete on speed and time. This was evident from the responses to question five which inquired:

> "What approximate percentage of your organisation's time is spent on a regular basis in cleaning (pre-processing) and linking data from disparate sources before it can be consumed for decision-making?"

The answer to this question sometimes was a precise numerical estimate, while many responses were more of a discussion on how time plays a significant role in decision-making. According to one respondent:

“Smaller businesses cannot afford the time or resources to cleanse the data. 80% of times, on average, is spent in cleansing the data.”

Another respondent highlighted the fact that business decisions become easier if the data is presented in a way which is easier to consume and process. The respondent said:

“Decision-making process only takes 5%. The rest of the 95% is spent in data capture, data collection, data sanitising, data cleansing, data linking, and data modelling. The 5% of the time for the decision-making process is looking at the data and making the right choice.”

The traditional factors of production, land, labour and capital are flexible and added or subtracted based on the requirement of the business. However, time is a limited resource. Time, respondents highlighted, was the single most critical resource, which they do not have enough. One of the respondents highlighted:

“Decision-making, if considered as a dependent variable, then time is the most important independent variable.”

Timely business decisions are pivotal for success (Yeoh, Koronios, & Gao, 2008). However, the elements to make an optimal decision depend on differentiated business insights. Businesses have access to similar set of data to make decisions. However, how this data can be turned into insights and how fast can it be done, defines the differentiated insights they might have from their competition. Many respondents highlighted the inability to make optimal business decisions on time. In their opinion, it is time-consuming to collect the data from different sources within the firm. Further, there is a need to consolidate and correlate internal data with trends found in external data to gain understanding on which business variables need to be adjusted to achieve optimal results. These results or goals might include increasing sales growth, raising market share or lowering operational costs. The opinion is that integration of data is time-consuming which at times can delay the decision-making process to achieve desired results. Respondents also highlighted that as Big Data is more prevalent, it is harder to understand emerging formats and protocols, which is time-consuming to incorporate in business strategy.

Furthermore, new evolving formats and concepts in data science demand appropriate and new skill sets, which are time-consuming to find and source. Up-skilling of internal employees to keep pace with the changing technology landscape is also a time-consuming activity, as emphasised by the respondents. Pre-processing data was also highlighted as an activity by the respondents, which is time-consuming. Pre-processing data, or cleaning data, is an essential activity in order to make sense of and derive insights from the data. Another insight provided by the respondents, which is closely aligned with time, is related to lagged data. Lagged data points are historical and helpful in some manner to understand relationships between factors affecting business. However, lagged data cannot be incorporated directly into a real-time decision-making process. If it is incorporated, appropriate adjustments should be made to adjust to the lagged information. However, the lag in the data can adversely impact the decision-making process. One respondent said;

> *"And by the time we see the data and use it to make decisions, it feels like the world had already moved on."*

Sometimes the lag is not in the data, but the time it takes to perform analysis. An example of this was highlighted by one of the respondents during the interview process. This respondent detailed his practice of hiring new talents at his firm. He emphasised how difficult it was to hire somebody with the right talent because of the lengthy process of analysing the data (e.g. resume, social media, online profiles, third-party reference collectors). A resume is a dataset that lacks quantification which could cause a most significant delay in hiring someone with the right skills. In this example, the Human Resource (HR) department collected the resumes. The HR department then scanned through the information to make sure each candidate met all the criteria the business had advertised. After this initial scanning, the HR department then sends matching resumes to a middle-level manager, who again goes through the resumes to screen and select a few individuals for the first round of interviews. This process goes on for multiple weeks before the final group of individuals are selected. Before this final screening occurs, many prospective interviewees drop out because they got an offer from another business. Organisations not only need real-time data to make real-time decisions, but they also need real-time analysis of the data to make optimal decisions. In the above example, the hiring exercise would have been robust with the help of additional analytics and data about interviewees. Also, Big Data analytics would provide more holistic and well-rounded information on the interviewees, making the hiring process much faster to conduct and conclude. During the interview, this was described by one of the respondents as:

> *"I need data in time; I need predictive analysis that is based on this data. I need the most current data to make any difference using those predictions."*

Coding for the node “time” using NVivo created the most intrinsic and connected word tree (Figure 42). This figure shows the connection between time and decision-making among the respondents. Time was the concept which was discussed across many questions during the interview process, as the respondents felt strongly about the lag time it takes for business intelligence to reach the respondents. Across different topics of discussion, time was discussed repeatedly within various contexts. From the time it takes to capture data, to the time it takes to process different kinds (i.e. variety, volume, veracity) of data, to the complexity introduced through the fast-paced competitive landscape, time was in the centre of many discussions explicitly or implicitly.

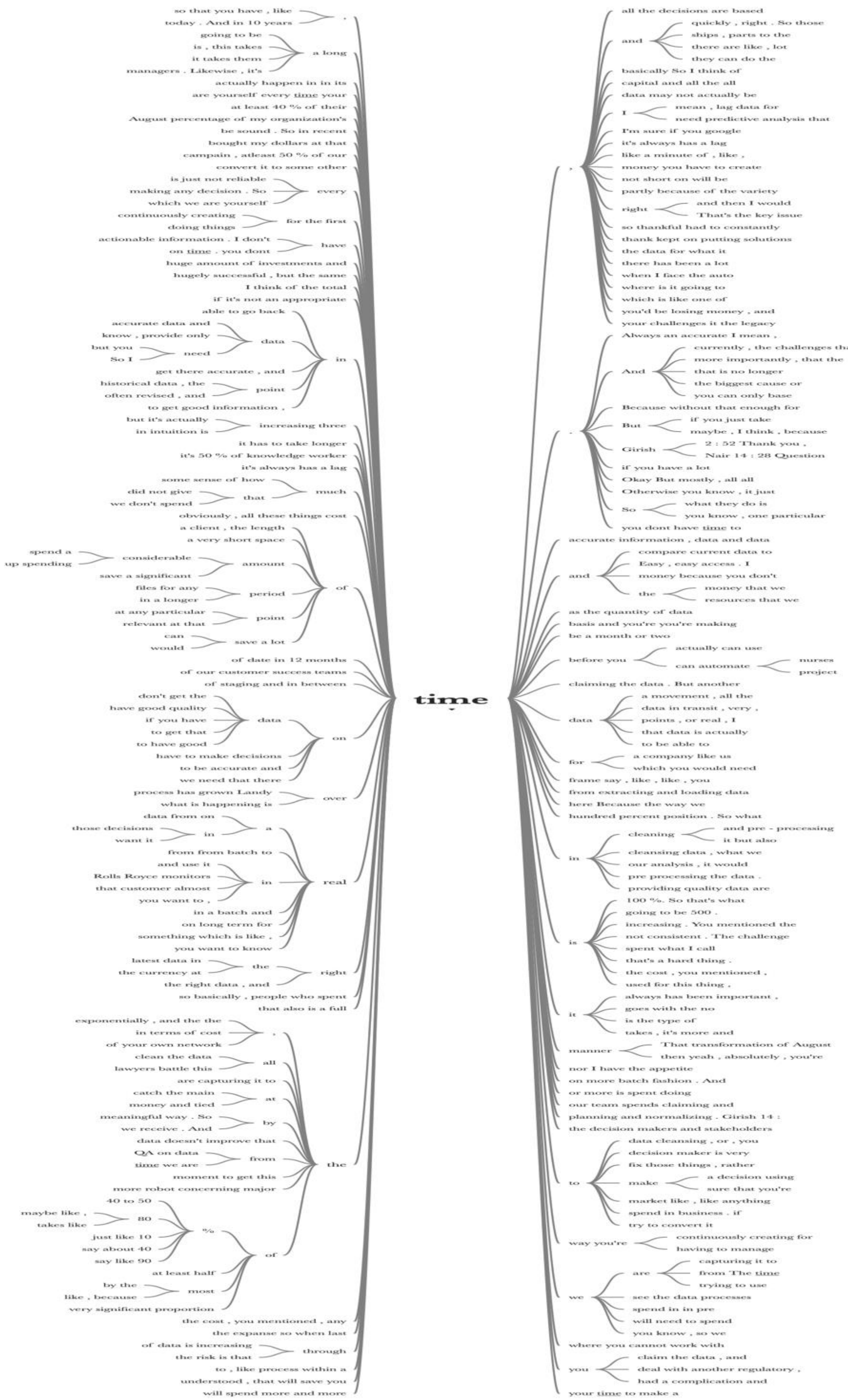

Figure 42: Word tree for the node "time"

Theme 4: Data and business intelligence is pivotal for competitive advantage

As data analysis and technology evolves, BI provides businesses with the ability to see their operations in greater detail. Business intelligence is more than a tool to understand how the business operates. It is a tool to gain insights which can provide actionable information to make optimal and smart decisions. With the trend to quantify every variable essential for success, businesses must be alerted to understand how they compare to the competition. BI can help the organisation optimise operations of the business by reducing the cost of labour, manufacturing and advertising, inventory management and process efficiencies. BI helps the business align itself with the business objectives and streamline operations. It can be used to see where the business has been, where it is now and where it is going, especially in the context of competition. During the interview process, the respondents highlighted five main benefits of how data can translate into business intelligence into a competitive advantage.

1. **Discovering true costs:** For businesses, a significant way to reduce wasteful spending is to streamline the inventory management system. The goal for businesses is to have enough stock on hand to fulfil customer orders, without overstocking. Overstocking ties up capital that is best used elsewhere. Business intelligence applications monitor inventory turnover by each product or service. Over time, it is easy to see how much supply is on hand. Another danger of overstocking is the risk of deadstock. Deadstock is stock that remains on the shelf for so long that it cannot be sold or returned. Deadstock can be caused by a product losing popularity, or by the development of newer versions of the product making the existing stock outdated. Deadstock can also be stock that has perished. With deadstock, not only has the investment capital been lost, but can also take up warehouse space for profitable products. Operational expenses also take a significant toll on the bottom-line if not managed effectively.

Business intelligence tools can quickly help identify which processes are contributing to business efficiency. KPIs extracted from business intelligence monitor and measure performance. Ultimately BI tools have the potential to reduce costs and increase the bottom line. The advantage of business intelligence helps businesses identify the current costs of processes. In the age of digital assets and virtual organisations, keeping a close eye on costs is pivotal for increasing efficiency and maintaining profitability. Respondents highlighted the positive impact of business intelligence from various facets of running or managing a business. In addition to the back-office expenses such as inventory management, payroll, and customer service, BI can also highlight which processes are the costliest and which have less impact on resources. Business intelligence can help highlight costs that are directly tied to the overall profits. One of the most significant benefits of business intelligence, as highlighted by several respondents, is the ability to monitor progress and adjust as necessary. Effectively managing costs is one of the ways to stay ahead of the competition.

2. Discovering new opportunities

Respondents during the interview process also emphasised the benefits of business intelligence to discover a new product or service opportunity. For example, business intelligence helps sales managers to quickly identify which customers are buying and which products are in decline. With BI, businesses can quickly visualise customer spending trends by monitoring purchases on a daily, weekly, monthly, or annual basis, effectively showing the ebb and flow of product desirability. Sophisticated business intelligence systems which incorporate secondary market research such as similar sector reports, industry trends, product preferences, or consumer survey results, tend to highlight new ideas for products and services which might be popular in the future. Holistic business intelligence which incorporates primary and secondary market research can help predict future consumer trends.

One of the respondents highlighted that the power of BI lies in identifying consumer trends which, in turn, emphasises new sales opportunities. An outstanding opportunity lies with cross-selling complementary products. For example, a store may sell washing machines and dryers. A customer may need to replace only their washing machine, but by bundling the two appliances together at a very attractive price point, there is an incentive to purchase the pair. If a distributor finds that they are only selling washing machines to a customer, it may be that the customer is getting a better deal on dryers from another supplier. Similarly, the washing machine manufacturer might think about selling laundry accessories like detergents and fabric softener as complimentary soft consumer products to their hardware (e.g. Miele).

Another respondent highlighted that:

> *"Changes in the market and regulatory environment creates new needs, thereby, posing new opportunities for product development."*

Business intelligence systems at times incorporate market research in the form of surveys, interviews, competitive analysis, and analysing secondary sources of information. Secondary information sources such as social media, patents, news blogs, industry analysis and reports can reveal emerging trends, highlight whitespaces and opportunities for new products or services. BI also helps businesses identify strengths and weaknesses that eventually helps form a market strategy for the product. Competitive analysis should be a regular process conducted by the business in order to know and respond effectively to competitor moves. Another respondent emphasized the importance of social media. With the rise of social media, analysis becomes more important to analyse conversations and identify patterns across social channels. As consumers become increasingly vocal about their opinions on social media, businesses are analysing this data in real-time to make decisions about product development and customer services. Most importantly, market research through social media provides a fair view of a product's performance by tracking the impact of the marketing campaigns run by businesses. This medium simultaneously helps to uncover the opportunities to improve and stay competitive in the market.

3. **Discovering the nature of customers:** The respondents during the interview process emphasised that their businesses are collecting a staggering amount of data about their customers. These datasets are no longer just limited to basic demographic information such as age, gender, and location. Businesses can now easily track and monitor past purchases, spending patterns, changes in demand, the influence of marketing initiatives, and more. Although collecting and housing this data is a good start, enterprises do not get any value from the information unless it is effectively reviewed, organised, and insights pulled to get a better understanding of their customer base.

Respondents provided some examples of use-cases of BI at their firms. BI software is being used to rank existing customers based on the recentness, frequency and value of their purchases. This feature of BI has provided enterprises with the ability to optimally organise customers into more targeted opportunity groups for up-selling and cross-selling. With business intelligence, businesses ensure that sales and marketing efforts are aimed at retaining and optimising the right customers and attracting the right prospects with the right offers, be it for an up-sell to an existing customer or an incentive to pull in new customers.

Respondents also highlighted that business intelligence could help business planners improve forecast accuracy and the overall sales and operational planning process. For example, BI software can track orders and fill-rates, so the business is ready to handle seasonal spikes and drops in orders. A high-end food and beverage producer can use this data to stock online, and in-store physical shelves with the right items when demand is the highest. When a drop-in demand is expected, the producer knows to pull back on production and save inventory space for other products. Business intelligence software also allows businesses to set the threshold for outliers to eliminate past events that could impact future customer demand forecasts, so the data is more accurate.

A high number of respondents highlighted that BI software helps efficiently plan, monitor and assess the success of promotional activities in relation to customer reactions to marketing campaigns or promotions. In this way, marketing budgets can be adjusted and allocated to more successful campaigns that produce the best possible ROI (return-on-investment). For instance, the same high-end food and beverage producer, discussed in the earlier example, might want to track the effectiveness of free shipping for the holidays offer versus the effectiveness of a 20% off coupon. The business can determine which offer was used more frequently or which offer prompted customers to add more items to their online shopping cart. By comparing actual to expected results by campaign, businesses can market smarter and better identify opportunities to increase sales and growth.

Business intelligence leveraging Big Data, can help businesses better understand, forecast and influence the behaviour of customers by providing rich insights into how existing customers behave, think, act, and spend. With a better understanding of customer needs, wants, and behaviours, enterprises are in a better position to serve and increase profits from existing customers, as well as pull in more targeted new customers (Hennel, 2018).

4. Discovering true profitability

True profitability is what the business produces in economic value factoring in all the indirect and direct cost associated in generating that economic value (Weissenrieder, 2997). This is the measure which factors in the real costs of running the business, time value of money and real costs of having or losing customers, each one of which was discussed earlier in this section. The culmination of improved efficiency of resources, better implementation of strategic plans, and the ability to tweak strategies faster makes business intelligence a critical aspect of corporate profitability. Not only does BI help highlight factors which directly impact profitability, including costs of producing products, or capital expenditures of machinery, it also helps in highlighting indirect factors (customer satisfaction, buying behaviour and patterns, market intelligence) which impact profitability. Respondents highlighted that if BI tools were more real-time, it would significantly provide a boost to overall profitability.

5. **Discovering hidden insights and trends**

Most executives and business managers have a sound idea of the overall shape of their business and what is needed to stay ahead of the competition. However, with the advent of Big Data and increased used of social media to gauge and record consumer preferences, it is hard to keep on top of fast-changing factors one needs to make optimal real-time decisions. However, many respondents mentioned that after a BI solution was implemented, they discovered new information or hidden insights which was were confusing factors in isolation. Having all the information allows businesses to take advantage of unidentified opportunities and to address unrecognised problems before they can have a severe impact.

Few respondents highlighted these hidden trends have arisen out of the data-driven business-intelligence tools. One respondent said:

> *"An organisation, which is not data-driven, is not going to sustain or survive. People who spent time in providing quality data are reaping the benefits."*

Another respondent emphasised:

> *"Many businesses, which embraced business intelligence and developed a data-driven culture, reinvented their businesses and they now understand what their core business is. This includes businesses like Computer Associates and Texas Instruments."*

Through a mix of market research and business intelligence, businesses have helped themselves derive insights in order to understand their business areas that are profitable. Businesses gauge the ROI (return-on-investment) based on the predictions generated using data. Details not previously obvious using traditional intelligence quickly become visible. Current business intelligence provides hidden insights on the target audience, market size, and competitor landscape. However, the landscape of data analytics which forms the pillars of a business intelligence system is changing rapidly with new technologies. Businesses need to incorporate these fast-paced changes in technology and datasets which can help them factor in the next wave of changes related to their products or services and consumers, to be able to make real-time decisions. One respondent summarised this aptly as:

> *"I have read research from Gartner, Forrester, and all these research groups who are data specialists, and they have said that the market of data management is $100 billion today. And in 10 years it is going to be $500 billion. Technology lifespan is shortening, and something which is new today will go out of date in 12-months' time. Hence businesses need to constantly keep on investing in new tools, new intelligent system so that they can continue to utilise the data and continue to maintain a competitive advantage through insights which are not available elsewhere."*

Theme 5: Various risks to a business intelligence framework

Most of the discussion with the interview respondents revolved around the use of data, and the strength and benefits of adopting business intelligence. The part of the discussion which added a lot of value was how these respondents highlighted various risks which might impact or drive business intelligence framework, especially in the light of the research proposal. Most of the respondents discussed these risks when asked the last question in the interview process, which is stated below.

> This research is proposing a business intelligence framework called Hex-E that will enable linking of data with inbuilt intelligence; What do you think would be the benefits and risks of such a framework to a business? (especially in the light of the issues you highlighted in previous questions with respect to challenges of extracting insights from data).

This fifth theme involved discussing risks highlighted by the respondents which not only helped to enhance the proposed Hex-Elementization model but could also help anyone who is embracing business intelligence for the first time. Among others, these risks ranged from regulatory issues, to cyber-security, fraud detection and legacy applications.

Privacy and data security

Data security and privacy issues, both personal and corporate, were one of the most widely discussed types of risk among the respondents. Along with quality assurance, the respondents highlighted the importance of incorporating data governance in any business intelligence system, including in the proposed Hex-E model. The respondents were interested to learn the proposed Hex-E model in envisaging a business intelligence platform which eases the integration of data from various applications, appliances, and disparate systems. However, they highlight the risk of the data getting compromised in terms of security or privacy. One of the respondents said that:

> *"Most importantly, for all businesses, it is crucial to maintain data integrity and data privacy. For example, a system incorporating personal details in a business intelligence platform built in the US will not want the social security number to be compromised. Even businesses like Facebook are trying their best not to let the data get compromised. However, as technology is changing so fast, businesses such as Facebook who are the forefront of technological innovation can have holes which can be capitalised by personal or state-sponsored agents to perform malicious activities. An example is the alleged Russian meddling in the US election of 2016. So, the people using Facebook do not know what to believe and what not do, and hence privacy and security are very important to incorporate in any business-intelligence system."*

Globalisation has not only brought the world together with the help of technology, but it has also increased societal complexity due to interconnectedness. This complex interconnectedness becomes difficult to interpret for any specific context because of the presence of so many players. These players include businesses who are trying to understand and serve their customers in a better way. These players include state players like governments, which tap into the social network to understand better ways of governing. These players include independent research firms who scour through the network for patterns and insights which in turn they sell it for commercial purposes. These players also include people or teams with malicious intent, with the primary aim of spreading chaos to either manipulate results or gain from people falling prey to their tactics. Hence data security becomes a major factor which businesses must consider managing the risk of privacy and security of data being compromised in any manner.

Risk of non-integration with legacy systems

A few respondents raised the important issue of legacy systems, highlighting legacy systems as a challenge and a big hurdle in climbing the technology ladder. Legacy systems are still running critical business operations like back-office processing, payroll and inventory management systems. In many cases, it is difficult to scale these systems and support the complexity and variety of current products and services.

Moreover, legacy functionality has not kept pace with current technology. Some large banks, for example, run systems so old they have had to “wrap” these applications in expensive data management systems to give customers online access. Constraints with the legacy systems may prevent businesses from rising ahead in the marketplace, introducing new products, moving into new geographies, or expanding services to customers on new platforms.

Further complicating matters, legacy systems are not easy to replace. Legacy hardware and software platforms often form the foundation of operations (Schneider A. , 2019). Moreover, businesses have incrementally enhanced these systems over many years, to the point where the systems are exceedingly complex and almost impossible to duplicate with a replacement. These businesses have become prisoners of legacy systems (Ross & Vitale, 2000). Legacy systems have security issues and challenges where certain vulnerabilities may not be as easy to fix due to the large, inflexible nature of the system (Rassa, 2019). Even if there is a fix, the patch is typically considerably delayed. Legacy systems data structure and infrastructure cannot keep up with all the new entrants into space, including Big Data and AI-ML. Replacing legacy systems is an investment to support data intelligence and can be costly to an organisation.

However, these legacy systems cannot be ignored entirely, nor can new business intelligence (or any other) system work in isolation (Alvarez Technology Group, 2019). Legacy systems form a big part of the business ecosystem. New systems, built on new technology, with nimble designs need to have backward compatibility. New systems must integrate with older systems to have a holistic approach to business intelligence until the technology supports all functionality of the legacy system. The new will have to integrate with the old. This integration is not evident in the existing implementations of business intelligence. When discussing the possible merits of Hex-E, one respondent said,

> *"You need to consider Hex-E being able to integrate with legacy systems, because if you don't consider that, then what that means is, you're not using the new framework to its full potential. If Hex-E has plans to break data into its most atomic level to help with integration, then legacy systems are best suited because that is where much latent intelligence resides. If Hex-E fails to incorporate that, then we are not using the complete set of information. If it is to be ready for future standards and protocols, then its test is to work with the old standards."*

The respondent made a valid point, which was not part of the initial hypothesis and model design of Hex-Elementization. From a model building perspective, this point was one of the most significant contributions to the design. Incorporating legacy systems became part of the model when this issue was uncovered during the interview process.

Risk of challenging the norms and countering change management process

The risk of challenging existing norms and hurdles brought in by the change management process is closely related to legacy systems discussed earlier. A strategy that includes legacy systems is a stop and start strategy. It deals with change in big chunks, followed by long periods of static, unchanging business. That was the mainstream way to run a business throughout the industrial age, and many businesses, especially brick-and-mortar businesses, have incorporated it into their DNA: “Have a short period of adaptation to make any necessary changes, then stop until another wave of necessary changes come along” (Freeman & Perez, 1988). However, business, in general, no longer works this way. Intermittent changes leave legacy systems that last until the next phase of change (Alvarez Technology Group, 2019), weighing the business down in an impractical model. Data needs to be processed within some defined timeframe in order to benefit from the results. The internal approval processes for any big change is lengthy. The biggest hurdle is to navigate through the entire change management process, which is difficult if the business has a traditional pedigree, in which change is slow to accept. These change-management processes involve everything from strategy and planning, risk management, executive buy-in budgets, resources, people, processes, information, systems, skills and training. A respondent who emphasised that the biggest hurdle to a proposed system like Hex-Elementization is the Change Management process itself. The respondent stated:

> *“Change management is a conversation in itself. However, even if you have the data which you fix with Hex-E, even if you have the systems in place, even if the systems integrate through Hex-E , you still have the entire business transformation change to address, and that is both within the organisation and with channel partners, with stakeholders and everybody else.”*

Businesses need to have technical acumen to embrace changes promised by a business intelligence implementation. Departments within a business should be ready to share data and knowledge to help a platform like Hex-E to be successful. Unfortunately, technical acumen does not guarantee success. The real test is whether business users embrace a new solution. Many times, businesses do not understand the technology being adopted and rush to procure or implement a new system. However, the challenges which arise out of asking the right questions often brings resistance. This resistance to change often arises out of the challenges business faces when planning to implement a new BI system. For example, a BI system built on a Cloud infrastructure promises to be accessible and faster than a system which is built on in-house infrastructure. However, the lack of knowledge about Cloud technology coupled with the fear of losing control – security – can force the hand of the business to abandon the plan altogether. The lack of knowledge or technical acumen creates a hurdle in achieving user adoption. As one of the respondents highlighted;

> *"The biggest hurdle is human inertia— people in organisations don't like to change. It takes significant physical and mental energy for people to change the way they do things in an organisation. Even when people know their habits are counterproductive or even self-destructive, they find it difficult to change."*

This inertia is magnified when groups of people work together in an organisation. The group quickly establishes unwritten rules to achieve objectives. Changing those rules—mostly the corporate culture—takes an extraordinary amount of conscious effort (March, Schulz, & Zhou, 2000). BI leaders need to understand the principles of change management to help their organisations turn data into actions. Change managers mainly look at which stakeholders are affected by the changes. Project managers focus on aspects like timelines, tasks and technology. With any change, the businesses cannot focus only on the business goals and profitability; they also must think about the workers. Technology thus cannot be left to the IT department. In business intelligence, this is especially challenging with the added complexity of increased enterprise involvement, unlike other business IT applications, where the enterprise and IT partner up. In BI the enterprise owns many components and work streams. Basically, in the world of business intelligence – technology is everyone’s job (Watson & Wixom, 2007). An important aspect of implementation is that BI projects and initiatives are iterative, but BI change management is ongoing.

Regulatory risks

There has been an extraordinary increase of regulations in the last 8 to 10 years, especially after the Global Financial Crisis of 2008. Increased regulation is felt globally. Respondents emphasised the beginning of increased regulations with the US Dodd-Frank Act during the Great Recession, which permeated into Europe with a series of regulations, most recently around data privacy and protection concerns in the form of General Data Protection Regulation (GDPR). However, respondents also noted that the rapid adoption of new data technologies, including artificial intelligence and Cloud-based storage, has made data risk a more pervasive concern for organisations in nearly every industry. While addressing the concerns around privacy and ethical use have become very important, businesses will continue on the path of digitisation and become more data-driven. Respondents highlighted that increased regulation and rules on sharing, moving, securing and classifying the data might turn out to be the stumbling block of the BI architecture proposed in this research (Hex-E). One respondent from investment banking highlighted the difficulties in adopting the BI framework proposed by Hex-E, by saying:

> *"Regulations in the financial sector are especially numerous and complicated. These regulations frequently overlap and sometimes contradict each other. The schedules are complex, and deadlines change frequently, but institutions are expected to be ready regardless. Ten years since the financial crisis, we see institutions still struggle to meet the complexity of these regulations. Implementation of regulatory requirements is one of the longest-running processes of any industry, and the Hex-E framework you propose might struggle with coalescing data amidst a difficult regulatory regime."*

Accurately identifying, analysing and acting to mitigate potential operational or strategic business risks is vital to enable continuous growth, as well as the ongoing client, partner and employee safety and satisfaction. One of the respondents highlighted that aspects of many organisations' operations – from across most major industries and in both B2B (business-to-business) and B2C (business-to-consumers) environments – are also required to comply with external regulatory and legal guidelines. Complicated governance requirements demand more than meeting basic external legal and regularity obligations. Market-leading organisations are increasingly using GRC (Governance, Regulatory, Compliance) to generate added business value by improving operational decision-making and strategic planning. By identifying, measuring, tracking and analysing progression towards, and adherence to, established benchmark metrics via reporting and analytics capabilities, these businesses show this value added to their potential investors or existing stakeholders.

Another respondent, who works for a local regulator highlighted there has been a lot of press coverage on GDPR, urging organisations to audit data proactively. However, there has been far less coverage and media attention on how to responsibly manage analytics, business-intelligence processes, and report on sensitive data. That is also a problem, especially for users who use BI tools to access sensitive data regularly. The respondent asked a rhetorical question:

> *"How do you ensure there's no maverick behaviour – even when it is done without malice, such as importing a database into Excel for a "quick project" – and ensure that everyone looking at data provided by accessing many troves of disparate data sets across various departments, connected automatically through Hex-E understands what can and cannot happen and why? Who is to monitor what Hex-E highlights as insights?"*

There are high chances that regulators are looking to make early examples of businesses that breach GDPR requirements, which puts BI teams and analytics in the hot seat. How data records are stored and deleted, and how they are encrypted and transferred internationally, will all have to be reassessed. This regulatory overhead poses an acute challenge for proposed frameworks like Hex-E which must define levels of sensitive data before generating insight and engineer private and sensitive information from the results.

Institutions that invest in data infrastructure and data management have a stronger chance of obtaining a comprehensive view of assets and risk exposures. Moreover, that can lead to more accurate reporting for regulatory purposes, as well as efficiencies and cost savings. Effective in 2018, all organisations in any industry that process the personal data of European Union (EU) residents are accountable for protecting sensitive data, regardless of whether those organisations have a physical presence in the EU. The potential penalties for noncompliance can reach as high as 4% of global revenue or 20 million euros (Bandyopadhyay & Bandyopadhyay, 2018), whichever is greater. In such an era of regulations, a business intelligence system will have to fulfil two needs a) the need of the business users to gain hidden insights to maintain their competitive edge, without comprising the second need – b) the need of the regulators to keep the data safe.

Validation of research questions

After completing the data analysis of the survey conducted in this research, validation of the results and the findings relative to the research hypothesis and research questions, is an essential part of this thesis. Validation of the data analysis and the research questions not only helps in assessing the findings of the research but is also helpful in solidifying the issues underpinning the research questions (Elo, et al., 2014). The dependability on the questions to provide validating responses during surveys can be influenced by multiple, hard-to-control factors. These unknown set of factors can make the validity neither quick nor easy to perform (Stajkovic, Locke, & Blair, 2006). However, unlike the challenges in research conducted in social sciences in which individuals are interviewed on a one-to-one basis, or in the field of medical research in which the effects of certain compounds of chemicals are hard to measure and quantify, this study involved researching in two phases. Phase one was a survey which formed the basis of the quantitative analysis, and phase two semi-structured interviews were conducted which formed the basis of the qualitative analysis. In this research, the questions posed during the survey and the interviews were linked to how business decisions are made. The researcher tried to understand the factors directly attributable to the use of data and business intelligence in a typical business setting. The survey and interviews were not designed to understand how decisions are made in general, but how a business intelligence framework would assist in making decisions. The survey and the interview questions were specific in asking respondents about the use and flow of data in their decision-making process, and what might improve it in the context of the proposed Hex-Elementization model.

The results of the data analysis from the survey cannot be extrapolated over a large population as it was not large enough to be statistically significant. However, the results were in line with the foundations of the research hypothesis set out in the introduction of this thesis. The results from the quantitative data analysis of the survey combined with the qualitative analysis of the semi-structured interviews provided a meaningful validation of the research questions posed earlier in this research. The key findings relative to each research question is discussed below.

Validation of Participants

Before the research delves deeper into validating research questions, a brief comment on validating participants involved in the data analysis stage is vital to understand the basis of the investigation in this research. Validating the data analysis of the survey is incumbent on the fact of justifying the suitability of the participants to the research (Kelliher, 2011). This suitability was addressed in the previous section (Research Methodology) which detailed reasons why the participants were chosen for this survey. In the initial part of the data analysis, details about the diversity in the participant's geographical location, experience, and field of expertise were classified and highlighted. The simple analysis of who the participants are and what they do, demonstrates the survey intended to attract responses from a wide array of people without any predisposition or biases associated with the aim of this research. Data is important in every facet, irrespective of the type of the industry or the dynamics of that industry. The research would have gained a skewed response if it only included participants from a specific industry (e.g. people working in Information Technology (IT)), or participants from the same department (e.g. people working in the IT department within a business). Hence the neutrality of the participants from a data consumption point of view was validated at the onset of the research.

Validation of Research Questions and Findings

This section is dedicated to summarising the results of the qualitative and quantitative data analysis discussed earlier in this chapter, concerning the research questions discussed in Chapter 1. Validity refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores (Messick, 1989). Each research question posed in Chapter 1, is validated by collating the knowledge gained through the three stages of the research, including literature review, qualitative analysis of the survey and the quantitative analysis conducted post the semi-structured interviews. The validation entails highlighting whether each research question was answered using the three stages of the research and if not, what were the learning outcomes for the researcher to modify and improve the Hex-Elementization model.

Findings related to Research Question 1

Research question 1 seeks out the ways organisations prepare to transform data into actionable insights in the age of Big Data.

The question:

Businesses have access to the vast amount of data that is continuously growing. Businesses extract some information snippets from this data based on available tools and processes. There is a need for businesses to transform this data into information and subsequently into intelligence. How do businesses prepare themselves to utilise data?

Organisations have access to the vast amount of data that is continuously growing (Big Data). Organisations extract some information snippets from this data based on processes and available tools. There is an acute need for organisations to transform this data into information regularly and, subsequently, transform that suite of information into intelligence. How do organisations prepare themselves to utilise data?

The question had few sub-points to be addressed, starting with how organisations are utilising Big Data in decision-making and the challenges in doing so. Another sub-question was to understand what organisations use in terms of processes and tools to achieve the transformation of data into insights.

This question was handled through all three phases of the research investigation process, including literature review, quantitative analysis and qualitative analysis. In the literature review (Chapter 2), the specific investigation was done to understand the existing body of knowledge relating to the use of Big Data in decision-making (Chapter 2). This literature review highlighted the benefits Big Data brings to decision-making by providing information from every aspect of the business, which the stakeholders need to make a practical decision. The literature review aptly recognised the fact that data is explicitly mentioned as a possible fourth factor of production after land, capital and labour. Data is an intangible asset which can help organisations gain competitive advantage. The literature also delved into facets of data-driven decision-making in Chapter 2. The enormous amount of data available to decision makers provides an opportunity to leverage information to make real-time tactical decisions. The same issue was highlighted from the quantitative analysis of survey responses. Survey participants were asked about the importance of data in their decision-making, and 67% of the respondents agreed that data was significant in decision-making. Nobody (0%) thought data had no impact in decision-making, despite the fact a few of the respondents were from industries which are not data-driven nor data-savvy. Furthermore, the importance of data was validated through the qualitative analysis post the semi-structured interviews. Respondents pointed out that data has become the fourth factor of production in any business. Traditional factors of land and labour are becoming obsolete given the new age mega-corporations like Uber, Airbnb, and Facebook, that make more money from online and digital assets than brick-and-mortar. During the interview, it was amply evident that data is pivotal for corporate survivability. Broad usage of information is key to making timely and optimal business decisions.

However, the transformation of data into information and insights was presented with a multitude of challenges, as theorised by the Hex-Elementization. Hex-Elementization is modelled on the premises that this path of transformation (i.e. from data to information to insights) is not accessible in the existing environment with the current set of enterprise tools. The literature review into these challenges started with highlighting different enterprise architectural frameworks. Zachman, Spewak and Bernard's architectural models were discussed in detail to provide a background on how these structures entail into structures which do not have enough flexibility to incorporate sweeping changes to how decisions are made. Enterprise structures and architectures are essential. However, the rigid structures created around functions (departments), makes architectures less nimble when decisions must be made on shared data and knowledge. This challenge of the transformation of data into insights was further validated during the survey process in which respondents were asked how easy it is to get insights from their existing data sets. An overwhelming 70% of respondents highlighted that they do not get optimal data from their existing business intelligence systems. Theme 2 discussed in Chapter 4 of the quantitative analysis of semi-structured interviews also highlighted the path from data to intelligence is full of hurdles. These hurdles included the complexity of data, quality of data, the different way corporations use or perceive the data, the cost of maintaining a business intelligence framework, the loss of context, and the lack of consistency when humans make decisions. The literature review and qualitative and quantitative analysis also highlighted the challenges of high volume, increasing velocity and changing variety of Big Data which businesses need to handle in order to extract insights on time. Also discussed in the literature review and subsequent data analysis were the various challenges faced by the organisations. These challenges include technical challenges when it comes to integrating disparate datasets and organisation inertia in accepting the challenges of incorporating a business intelligence culture. This issue was discussed in detail as one of the Risks (management change) in Theme 5 (Chapter 4) in the qualitative analysis section of this thesis.

In essence, the investigation to validate research Question 1 highlighted that data and its transformation are critical for organisations to not only gain competitive advantage but also to survive successfully into the future. Organisations are finding ways to handle the complex array of applications and systems to handle the influx of data (in various forms and quantity) to extract insights. As highlighted by Gartner, less than 1% of the global data repository is analysed for any purpose (Gartner, 2018), and the reason is that organisations many times do not know what they hope to achieve. There is a lack of understanding around what data needs to be utilised to make decisions, and what decisions need to be made to stay relevant. With changing technology, organisations do not have a more straightforward way to connect and transform data, and find relationships in an easier manner. Hex-Elementization is the framework, envisaged to break data into its most atomic level to enable integration.

Irrespective of the type of the data, size of the data, and the format of the data, Hex-E aims to present the data in the way websites are presented to our devices. Similar to how TCP/IP protocols translate the language of the Internet to the language understood by the devices, Hex-E aims to translate data in a way that can be related and coalesced seamlessly. Validating research Question 1 highlighted the fact that the goals of the Hex-E model continue to remain acceptable.

Findings related to Research Question 2

Research Question 2 stemmed from the issue that many businesses struggle to utilise the resource of data generated internally or externally. Business intelligence in its current state faces challenges from various angles, including technological challenge, integration challenges, the challenge with legacy system and more significant challenge in transforming data into insights. Research Question 2 set out to understand the challenges faced by different organisations and whether one of the goals of the Hex-E framework (i.e. ease the integration of data to enable real-time insights and decisions) continues to hold true. Like research Question 2, the three stages of the research, namely literature review, qualitative data analysis and quantitative analysis helped answer this question.

The question:

What are the challenges in creating and utilising business intelligence in understanding the current state of the organisation and how can this business intelligence be used to undertake business transformation? How do organisations enable intelligent decision-making in real-time by streamlining their data ecosystem and creating rapid correlations between otherwise unrelated data sets?

The world is inter-connected and co-mingled as per the literature review in Chapter 2 on six degrees of separation. The theory postulated that over time, with increased globalisation fuelled by technology, the world has become smaller and closer. The theory postulates that not more than six connections connect two random people in the world. However, recent research has shown that it takes less than 3.7 connections to connect any two random people. The rationale behind showing the interconnectedness of the world was to show that businesses are more connected than ever. For example, a slowdown in China will affect a grocery business in Australia based on the inability to map out the supply chain information in detail. Business does not operate in isolation which makes it harder to factor in so many variables impacting a given business. Businesses can never list all the issues that may have an impact, because of the complex array of relationships created by a connected world.

The other issue with the transformation of data is the data itself. To provide a bit of background, Quantum Information Theory was discussed in the literature review (Chapter 2). Going through this literature also helped set the stage for the way information is viewed from the most microscopic level. Hex-E proposes breaking the data element to its more fine-granular level with properties of six sides called hex-element. To find the granular data structures, few works of literature around the basic definitions of data from bits and bytes to other basic structures of data were analysed. The literature review highlighted the gap in knowledge base around granular structures as postulated by Hex-E. The gap originated from the search for the ability to join data together organically to provide automated business intelligence. However, the literature review highlighted the inability of data to integrate into an automated fashion and that this stems from the way information theory defines data. Data is changing in volume, variety and velocity as businesses try to quantify and factor in as many variables as possible for making faster and accurate decisions. Therefore, the transformation of data into insights must be re-thought from a grass-roots level, rather than an afterthought. These challenges were well documented in the literature review conducted to answer this research question.

The challenges with the transformation of data to insights were also summarised during the survey in which participants were asked to list the significant challenges they face in the existing suite of data they use for making business decisions. The challenges highlighted during the survey, detailed in Chapter 4 sheds a practical light on what was discovered in the literature review. The survey results highlighted the biggest challenge in transforming data into insights in the current environment was the “format of data”, followed by “quality of data”, “integrating different pieces of data” and lastly “sheer volume of data”. These challenges are the areas in which the proposed Hex-E architecture is theorised to add the most value. By breaking the data (i.e. in any format, frequency, or quality) into the most granular element, the architecture would become agnostic to the nature of the source of data. These basic data-elements in Hex-E are then left to coalesce with each other based on the common characteristics they share.

Furthermore, participants in the survey were asked about the perceived benefits of a business-intelligence framework like Hex-E provided on top of the existing systems. These benefits were discussed in detail in Chapter 4. Participants were provided with a brief research proposal in which the primary goals of the proposed architecture were described to help them answer the survey question. Participants of the survey were given an array of benefits they could choose from if the concept of Hex-Elementization is implemented as a business intelligence framework in their organisation. The benefits highlighted by the survey participants provides a good definition of the proposed system and any business intelligence framework which should be ideal for achieving business goals. Among the benefits participants highlighted, “real-time business intelligence” was at the top. Although the sample size is small, this shows what the existing business intelligence systems lack in terms of the functionality. It also shows time is the element business stakeholders value the most when it comes to the ability to make decisions. So, both the literature review and the ensuing survey results highlighted that real-time intelligence is challenging to achieve in the perceptions of the business stakeholders and theory.

To further analyse the challenges highlighted by the survey participants, regression analysis was conducted in the quantitative analysis stage (Chapter 4). This study helped aggregate and collate broader issues across the entire spectrum of data transformation and its ensuing use in business intelligence. The results were discussed in Chapter 4 and highlighted significant issues by the survey participants. More importantly, the regression study showed a strong relationship between the issues (with data transformation and issues related to business intelligence) were higher with the participants with more professional experience. Another challenge which was highlighted as a part of the survey was the lag in data impacting business decisions. The results of this survey response were discussed in Chapter 4. Not factoring data on time into a business intelligence framework can provide skewed insights and sub-optimal business decisions.

These challenges were further connected during the qualitative analysis of the interview responses. Theme 2 and Theme 3 (in Chapter 4) highlighted various challenges in the data transformation process for business intelligence. The issue of not having real-time analytics was elaborated exclusively in a separate Theme 3 (Chapter 4) which delved deep into the dimension of time and the issue of not receiving data or information in time to convert to insights, which is useful in making tactical and real-time decisions. These themes touched on issues and challenges as highlighted by the conversationalist type of semi-structured interviews included in the research. The challenges encompassed technical, structural, corporate outlook and culture, among other things.

The challenges of data transformation highlighted in the process of answering the second research question helped in forming the features of the Hex-E business-intelligence framework. The issues of integration of data, finding correlations within disparate data sets, the cost of running a business intelligence application with specialised software and personnel, the speed at which data gets transformed into insights formed the main features of the Hex-E as postulated earlier in the research design.

Findings related to Research Question 3

Research Question 3 entails gauging the need for a new type of data, mainly unstructured, to be incorporated in decision-making systems. The investigation of existing literature sought to identify the discussion of this new type of data in the decision-making process and if strategies exist to handle unstructured data. On similar lines of investigation, survey participants were asked close-ended questions on whether they are using or plan to use unstructured data in their decision-making. During the open-ended interview process, respondents were also asked if they have started incorporating unstructured data in their decision-making process. All these lines of inquiries were designed to gauge the current state of unstructured data and its ability to add value to the decision-making process, for what the Hex-E architecture is proposing.

The question:

How does the increased use of unstructured data impact decision-making? How will developments in data science augment decision-making?

The use of the term unstructured data was used to cover separate sets or kinds of data. Firstly, it was used in the context of a dataset which is not structured in form. A dataset which is not easily consumable with traditional data analytics, since the beginning of digitisation. The second meaning of unstructured is to represent not only new age data formats such as likes in social media, videos, audio files, images, signals from computer vision signals, satellite signals, and IoT data, but also any new kind of mobile phone signals (6G) (Pradeep & Kumar, 2017) or any new kind of communication protocols. Currently, the number of devices connected to the Internet is four times the number of people who have done so (Lueth, 2018). Many budding enterprise applications are storing more data than in the past. However, less than 1% of the data stored globally is currently being analysed in any way (Kumar, 2017). This statistic indicates an untapped opportunity for enhancing decision making process. However, through the investigation process for Question 3, it becomes clearer that the existing systems cannot consume unstructured data. With new protocols and communication channels being introduced to the world, there is a need to embrace a platform which is agnostic to the format or frequency of data. This concept is at the heart of the proposed Hex-E architecture.

While analysing the literature, it was found that 80% of incremental data captured by businesses is unstructured in nature and is doubling in size every three months. Existing literature fails to grasp the benefits of unstructured data in making real-time tactical decisions. Various journal articles hinted at upsides to the firms that embrace unstructured data sets, including augmented sales, targeted marketing, improving process efficiency, minimising costs, improved supply-chain relationships, accurate long-term decisions, the ability to make real-time tactical decisions, enhanced employee satisfaction and new product/services idea generation. Most of the discussion in the annals of literature was surrounded by the prospect of incorporating unstructured data set into decision-making as a stand-alone process, rather than highlighting the power of merging unstructured data into structured dataset to augment the insights provided to decision-makers. How this augmentation will occur was revealed in the quantitative analysis of survey data and the qualitative data from the interview process.

Participants in the survey were asked (Chapter 4) if unstructured data provides an avenue to gain hidden insights which were not so obvious using just structured data. High-performance computing, distributed databases, and Cloud storage has made it easier to incorporate unstructured data sets into their decision-making process. A majority 65% of the participants agreed that the unstructured data sets have the potential to enhance decision-making. Although no direct questions related to unstructured data were asked of the respondents during the semi-structured interview process, many highlighted the opportunities and challenges in incorporating it in Theme 1, Theme 2 and Theme 4 (Chapter 4). Few respondents highlighted the opportunity of using spatial data in quantifying data which was previously difficult to factor in making decisions. For example, the geospatial photographs of active industries in China provide a way to quantify the industrial production in China. This data is not collated or distributed by any statistical agency in the world in order to benefit global businesses. Many respondents emphasised the importance of capturing star ratings or likes on social media regarding products or competitors' products to help manage supply chains and profitable products or services. This unstructured data set helped fine-tune strategies with the help of real-time analysis. Another respondent highlighted the advances made in agriculture in which drones are deployed to measure temperature, humidity, wind and nutrients in the atmosphere to help automate irrigation systems and spraying of fertilisers by automated machinery. However, respondents highlighted the challenges in incorporating unstructured data because of difficulties in incorporating the data in closed systems. Businesses, which continue to have legacy systems run core applications, are facing difficulties in incorporating the new age data sets (unstructured). Other respondents highlighted the reticence of the businesses to incorporate new age data sets thinking the hype will soon die, giving them the license not to make changes. Thus, from technological issues to management culture, respondents highlighted issues pertaining to the increased usage of unstructured data in their existing decision management systems.

Although it is challenging to incorporate cultural issues and change management issues in any new proposed system, Hex-E proposes a novel blueprint for a system which can ease the technological challenges of incorporating unstructured data sets, either existing or future variants. By breaking down the data stream of any unstructured data set to the most granular format, Hex-E provides a platform which is agnostic to any specific form of data. Going back to the example of the Internet, in which devices are agnostic to how an Internet page is displayed, a business-intelligence system needs to be not only ready for the existing array of unstructured data sets but also future variants of data sets. Thus, validating this question on unstructured data set, helped make the designs of the proposed model of Hex-E more robust than initially envisaged.

Findings related to Research Question 4

Research Question 4 involved concepts which permeated through this entire research. The three main concepts related to the research question are quantity, quality and context of data. The literature review, the survey responses and the discussion during the interview process, made it clear that these three concepts were at the heart of any discussion surrounding data and its use in the decision-making process.

The question:

Is the battle of Quality vs Quantity of data more prominent in an environment in which data science is being incorporated in every known industry? How does the Quality vs Context continuum survive in such a changing environment?

The issue of quantity of data and quality of data were highlighted in the literature review in Chapter 2 which was specifically investigated to understand the use of Big Data in decision-making and data-driven decision-making. These two concepts quality and quantity of data were also indirectly covered when analysing literature for unstructured data in Chapter 2 within the literature review. In the various works of literature reviewed during this research, academics have struggled to understand the balance between quality and quantity of data. When the quantity of data is smaller, it is easier to spend time cleaning the data to extract meaningful data which can then be used in a decision-making process. With the explosion of velocity and volume of data, the question is if the increased quantity of data still warrants the quality to be perfect, or does the quantity of data be used to generalise findings which might exceed the limits of what a quality data can provide? (Unhelkar, 2017) This was the point of contention in many academic works of literature. The example quoted in Chapter 4, in which a medical trial involving 100 people was conducted to understand the common cold, is outpaced by a simple search of people scouring the Internet for remedies during the typical cold season in the US. In this example, a 70% hit rate among the volunteers in the medical trial would highlight 70 people out of 100 who showed positive symptoms of common cold. However, even a 50% success rate using the sample of Internet users and searches on search engines regarding the medication which they use relative to their common cold symptoms might run into millions of people. The question is whether a 70% hit rate encompassing 70 people is more valid than the 50% hit rate involving a few million people in determining the main symptoms of common cold. Academic literature discussed this issue in multiple facets. However, this quality vs quantity issue became more apparent with the quantitative and qualitative data analysis which ensued. On the contextual issue with data, not many pieces of literature were found tying the context of data when it came to business intelligence applications using Big Data. The context was a concept discussed in isolation when business goals were discussed in the literature. However, the continuum of context concerning changes in the supply of enormous trove of data was not individually researched.

The survey included three questions related to quantity, quality and context of data. The results from the survey regarding these issues were discussed in detail in Chapter 4, and the regression analysis conducted in Chapter 4. These survey responses highlighted a prevalent feeling that a lot of data did not necessarily mean good data. Most believed that the derivation of the data used in decision-making dictates how much accuracy one needs in the data. If the aim is to create long-term goals using long-term trends, then a higher margin of error is acceptable when consuming Big Data. However, if the scope of the decisions is short-term, then the accuracy of the data needs to be higher to support accurate decisions. Most of the participants in the survey highlighted that a significant proportion of the recently acquired data is bad in quality. Concerning the context in data, 72% of the participants in the survey agreed that context gets lost amidst a massive quantity of data. This relates to another question in the survey which asked the participants what proportion of time is spent cleaning and pre-processing the data. Participants highlighted that on average they spent 80% of time cleaning and pre-processing the data before it can be used in further analysis. The more time it takes to process and clean the underlying the data, the longer the delay in using it to extract insights. These insights are then weighed on costs and benefits spectrum before they can be used to make decisions. Context is lost in this delay in transforming data into insights. Context is also lost when analysts find it difficult to understand which dimension of the Big Data set would yield the biggest return-on-effort (ROE). ROE is low on Big Data sets which are hard to understand or crunch for analytic purposes. The context of the business goals gets lost at the lowermost management level which deals directly with the data. This loss of context was evident from the results of the survey.

Finally, two themes in the qualitative analysis of interview responses were dedicated to understanding the quality vs quantity vs context continuum regarding data. These were Theme 2, Theme 3 and Theme 4 (Chapter 4). Theme 2 highlighted the challenges faced by the respondents in the journey from data to business intelligence. Most of the respondents believed that with the explosion of data volume and variety, businesses need to increase quality assurance. This explosion of data was explicitly highlighted by respondents who work in data-centric industries including banking and consumer services. Many agreed that organisations which proactively manage data as an asset and strive to improve the quality will reap the full strategic value.

Quality has also become the issue of the regulatory framework in many countries, which is now expanding globally. Quality of data management is essential to the survivability of organisations not only from a competitive advantage but also from a governance perspective. Many businesses are now highlighting higher quality of data as a Unique Selling Proposition (USP) for goods and services. For example, recently Apple advertised how they value individual privacy when it comes to dealing with the data on their smartphones and tablets, especially when the likes of Facebook and Google are in constant trouble with regulators. Respondents in the interview process also agreed that context gets lost in many ways during the data transformation journey.

Organisations currently storing every interaction or engagement with clients, tapping into primary and secondary market research to understand customers, using social media to capture shopping behaviour and preferences of their and competitors' customers are rarely utilising the data to its full potential. The main issue, as highlighted by the respondents, is the loss of context from initial planning to subsequent implementation and time lag between the two activities. For example, a business might want to understand the buying patterns of customers on third-party marketplaces, like Amazon, to help build new products. Some actions include conducting primary research, purchasing consumer buying patterns on these third-party marketplaces or capturing likes and dislikes of customers on various products. By the time all that data is cleaned, processed, and analysed, the insights obtained might be irrelevant or obsolete to react to competitors or other market changes such as introducing a new product. In the time it takes to process and analyse data, a competitor enters the market with the proposed product the business might be planning to develop. Real-time insights are needed to stay nimble in this highly competitive market. To gain this competitive advantage, businesses need to define context which is achievable relative to their ability to convert data into insights.

The discussion around quantity, quality and context were important to understand as the Hex-Elementization model develops. Although the Hex-E based business intelligence system aims to cater to Big Data (quantity), quality and context remain a more significant factor to incorporate in the model. Contextual-based business intelligence systems are hard to configure or develop given it is relative to each business with the same industry or different across the various industry in the same economy. Context is relative, and the ability to customise a business intelligence architecture based on context is what drives the connections and correlations between elements in a Hex-E architecture. Context remains the underpinning pillar over which Hex-E will prosper as a business intelligence framework.

Findings related to Research Question 5

The question:

What are the various business risks associated with the effort to connect, coalesce and correlate datasets that produce intelligence for business to help with real-time decisions?

Research Question 5 was the most complicated to answer among all the research questions. The complication arose from the need to discuss various kinds of business risks associated with a typical business intelligence system. No credible literature review was found covering the business risks from a business intelligence system which aims to connect, coalesce and correlate data sets to produce insights to enable accurate real-time decisions. The topic of discussion on business risk is lengthy and verbose and could not be covered in the survey sent to participants. However, the semi-structured interview provided a platform to discuss these business risks in detail.

The open-ended questions gave enough freedom to the respondents on various topics of risk in Question 5. When asked about the key risks associated with the proposed business intelligence platform (Hex-E), the respondents emphasised a plethora of risks. One risk is cybersecurity, especially when businesses want to open the Hex-E framework to correlate data from external sources to provide insights to the management. Some respondents highlighted the risks which are inherent in the organisations when one department is forced to share the data with another department. The value of the proposed business-architecture (Hex-E) is to help find a relationship between datasets, internal or external, and find patterns and trends which were not explicitly sought by decision makers. The ability to surprise management with insights not factored in the decision-making process provides the highest value. Hence, privacy and data security issues were highlighted by the respondents as an essential cog in the business intelligence framework proposed by this research.

Another risk, which was not initially considered in this research, was the ability to integrate with legacy systems. The risk of losing out on an existing, large amount of information, which although might be dormant, holds enormous value when it comes to making long-term decisions. The risk of leaving out the legacy systems helped design the Hex-E framework to have backward compatibility along with future potential. The third kind of risk respondents highlighted was regulatory risks. Introduction of new regulations, like the recent GDPR, can change the way businesses store, access, and share data within the organisation. Securing the data at multiple levels for access controls to avoid malicious activities needs to be factored in a business intelligence framework like Hex-E. The ability to find a hidden relationship when correlating external data with internal data risks exposing corporate networks to the outside world. To keep data secure is crucial for the business intelligence framework at the same time not compromising on speed or accuracy remains a challenge under such settings. The proposed framework of Hex-E needs to consider these risks to gain confidence in the process of generating insights. The competitive advantage gained from unique insights should not come at the risk of eroding data privacy or security of data.

Many of these business risks discussed above were not factored in a while designing the initial model of Hex-Elementization. Factoring in these risks into the Hex-E framework was the most significant benefit of investigating the research Question 5. The insights gained during the Question 5 added value in terms of enriching Hex-E from perspectives not considered when constructing this model. The model will now be adjusted to cater to some of these business risks highlighted by the respondents. A revised plan for the model will be highlighted in future outcomes of this research in the next chapter.

Chapter 5: Conclusions and Future Direction

Theoretical Contribution and Practical Contribution

This research contributes to both the academic and business worlds, creating a considerable overlap between the two. The theoretical contribution will help further the research by adding to the knowledge base this model creates at the end of the study. The practical objective deals with the applicability of this work in practice by businesses. The contribution of this research to both academic (theoretical) and practical world are listed below:

Theoretical Contribution

5. This research provides a business intelligence framework to enable real-time, resource-saving, accurate decision-making in workplaces. The underpinning theory will add to the decision-making theories explained in the Literature Review section.
6. The contextual underpinning of large undertakings in new-age technologies, like Big Data, is just the beginning. This research provides the basis of a contextual driven business intelligence framework in a push environment. Decision-making in the era of big-data is detailed in the Literature Review. The theories in this subject will be further augmented by the introduction of Hex-Elementization.
7. This research adds a new concept of integration to the area of data and information gathering, especially in the world of Artificial Intelligence, Machine Learning, Internet of Things and Big Data. The current theories around integration of data are not designed with AIML in mind. The integration need to be more automated in the era of big-data + AIML. The underpinning concept of Hex-E with

respect to integration of data is further discussed from the lenses of theoretical literatures.

8. The research provides a theory at the atomic level of data to enable automated integration. Apart from the information theory which was conceived in late 19th century, no major attempt in literature has been made to granulise data, which is discussed in detail in the Literature Review. This research will help create a new branch in the body of knowledge which deals with breaking data further into smaller self-contained unit of measurement.

Practical Contribution

7. This research helps save resources, in terms of people, time and systems, by providing a generic framework in which a diverse set of data correlates and connects. This automated connection provides meaningful and timely insight for real-time and accurate decisions.
8. The framework provides the organic growth of intelligence encompassing both internal business data (i.e., from HR, Finance, Marketing, Supply Chain) and external data (i.e., Third-party data vendors, high-frequency consumer data).
9. Resource consumption in enterprise architecture planning is extensive and can be optimised with the flow of information and the ensuing business intelligence. The model proposed by this research provides a solution that works in various situations and architecture.
10. Big Data, and its variety, volume, and velocity, is expected to increase exponentially in various forms. The framework proposed by this research enables easier integration of any new type of data entering the organisation.

11. The proposed Hex-Elementization framework, powered by Artificial Intelligence, including Machine Learning and Neutral Networks, automates the integration of data from disparate sources. Hex-E has the potential to be implemented as-a-service or as-a-protocol within any data-driven organisation.
12. The framework proposed by Hex-E automates triggers and red flags to help detect regulatory violations in real-time. The framework supports the need for information to be collated and correlated from diverse market players to comply with regulatory requirements. Compliance departments and regulators alike, need to process this information on a real-time basis to maintain a credible oversight on business activities.

Conclusions

This research investigates and explores approaches to enable the transformation of data to intelligence to help businesses make timely, accurate and wide-ranging decisions. The research highlights the concept of Hex-Elementization which is theorised as a semi-automated and seamless framework enabling integration, correlation and coalescence of data into intelligence. This intelligence provides value in real-time to help make accurate decision-making. The traditional business intelligence approach is time consuming, modular and linear (Moss & Atre, 2003). The move to the world of Big Data, however, is not merely a quantitative increase in volume and processing complexity. There is a qualitative jump in the challenge of handling data that is high-volume, high-velocity and of wide-variety (IBM, 2011). The 4th V – Veracity (or quality) is added to create the contemporary definition of Big Data (Sathi, 2012). There is an acute need to automate the creation of sensible higher units of data from the vast and complex Big Data eco-system that can produce information and knowledge for business. This research presents the Hex-Elementization (Hex-E) framework, that aims to contribute to further automation in business intelligence and business decision-making.

The research started by discussing the philosophical background behind the concept of Hex-Elementization (Hex-E). This first chapter delved deep into how Hex-E is formulated at the most granular level. The starting point for Hex-E is the definition of a data point through six parameters, hence the term Hex. Optimising a hex-element to parameters of six, as argued in this chapter, is based on the benefits of a hexagon, which are structure, efficiency, and symmetry. Further to defining data through the six parameters, an essential contribution of Hex-E is enabling data points to connect automatically. Given the massive and ever-growing size of Big Data – within and outside of business organisations – it is evident that unless the connections between data points are intelligently automated, it becomes increasingly difficult to make sense of this data, especially for business decision-making. Chapter 1 further outlines the research outcomes followed by discussing the main research questions. This first chapter also summarises the theoretical and practical contribution to the concept of Hex-Elementization.

Chapter 2 involved investigating existing literature to seek concepts and discussions similar to the concept of Hex-Elementization. The literature review included the investigation into existing relevant theories around Big Data, the interconnectedness of people and economies, information theory, enterprise architectures, unstructured data sets and AI, and context in data management. Much of the literature studied showed research in the aforementioned areas but in isolation. A knowledge gap between holistically combining these concepts versus isolated usage was identified. The literature review provided both confidence and corroboration the Hex-E model has all the elements of a new body of knowledge.

The next chapter (Chapter 3) involved outlining the research methodology to answer the research questions highlighted in Chapter 1 and to help fill the literature gap highlighted in Chapter 2. In Chapter 3, the research methodology was discussed in detail, highlighting why a mixed-method approach was adopted for this research. After carefully going through various mixed-methods, the exploratory-sequential design approach was considered a best fit the goals of the research investigation. This exploratory-sequential design method involved using both quantitative methods and qualitative methods to help validate and answer the research questions. Both, quantitative and qualitative aspects of the research methodology was discussed in detail in Chapter 3.

Chapter 4 details the data collection methods for both quantitative and qualitative research methods. This chapter then details the methods of each instrument used for the two research methods. For the quantitative analysis, a survey was sent to 120 people from various industries. Fifty responses were received and analysed using quantitative techniques to understand various aspects of business intelligence, including challenges and hurdles in the existing environment. For the qualitative analysis, ten participants were picked from various industries and with varied experience. Semi-structured interviews were conducted with each participant. The textual analysis using NVivo was conducted on the transcripts of these interviews. Five major themes emerged from this process and the themes were discussed in detail in Chapter 4. After finishing both quantitative and qualitative analysis, this chapter detailed the validation of each of the research questions discussed in Chapter 1.

Research investigation using a mixed-method approach based on the gaps identified in the literature review, provided clarity on the aims of the proposed model of Hex-Elementization. The validation of research questions in Chapter 4 highlighted that the Hex-E model includes most of the gaps in practice and theory. A few improvements to the Hex-E model were also realised during this phase, which were subsequently incorporated in the model.

Limitations & Challenges of Hex-Elementization

New Form of Unstructured data

Big Data and the IoT ecosystem can collate, store, sift, aggregate, visualise, and analyse complex datasets in relational, structured, unstructured, sequential, and geospatial form. Hex-Elementization as an architectural framework would need to create a platform to understand, decode, disintegrate, relate, encode, process, compute, and create directional decisional paths before reaching the end user. Text analytics (Dey, 2003), which is how consumer interactions are captured (via text, phone calls, emails, call centre logs, globs, surveys, and social networks), is difficult to perform due to the element of language. The same challenges apply to continuously evolving unstructured datasets (e.g., Vines or Live Photos).

Hex-Elementization will need to create and simplify the complex structure of this new kind of data to harness its untapped power. The process of Hex-Elementization starts with creating meaningful and identifiable hex-elements, each of which holds six distinct properties. The challenge is to break unstructured data (e.g., voice) into identifiable and self-sustained granular elements. The processing of unstructured data and the data exchanges, when compared to structured datasets, are less precise or complete (Feldman & Sanger, 2007). The Hex-E process will be augmented once AI enters the realms of text and voice expression to help understand the complex set of data in simple terms.

IoT Node Interface, Communication Protocols and Channels, and Hex-E

The uptake of IoT, which is evolving all the time, is going to explode in the coming years across every part of society (Kumar & Mallick, 2018). Although Hex-Elementization can provide a platform to ease the communication, processing, and data interchange between IoT devices, the question is whether each IoT data pipe will need an individual gateway that has to be intelligent enough to break the outgoing data into hex-elements. Given that current IoT devices have limited processing capabilities for the Hex-E requirements, this requirement will demand that IoT hosts become smarter and faster. As the need for data increases, communication channels and protocols match pace and become better and faster. The capabilities of 5G networks compared to 4G networks is exponential. This exponential speed in networks means exponential flow of data to businesses, which in turn means increased need to incorporate this data into decision-making systems.

Similarly, future communication channels could boast superior speed and robustness compared to older protocols. The capabilities of Hex-E could grow to deal with the changing landscape brought in by future communication channels. For example, microwave-based mobile networks (e.g. 5G) can be decoded as data to help integrate with trillions of IoT sensors in the future. Hex-E can help enable this communication integration by breaking down data into distinguishable and identifiable hex-elements. Once this breakdown is successfully completed, Hex-E can leverage its integration capabilities to coalesce data from various sources and apply to business insights.

Heterogeneity of IoT

Heterogeneity of existing IoT systems complicates the Hex-Elementization framework. Ideally, each IoT stream is a process that decrypts its data set before these elements enter the Hex-Elementization phase (see Figure 6), where correlations and decision trees start formulating. Each stream will have to adhere to specific common goals and protocols for Hex-E to work seamlessly.

Pull or Push Architecture?

Another protocol that needs to be discussed at length and decided upon is the question of whether IoT will push or pull the data for intelligent processing. Going back to the example of autonomous cars, the smart meters operated by city governments need to push data out of their domain so that IoT devices (and hex-elements) can access this data to find and build relationships. The autonomous car is only beneficial if it makes decisions autonomously across various functions (parking, distance to the destination, traffic diversion), not just when it comes to driving. The most desirable aspect of the Hex-Elementization framework is having real-time intelligence based on readily accessible and related datasets. The protocols built on pull data over a security layer would push the need for smarter gateways for each IoT usage.

Future Direction

3-Dimensional Hex-E

The future direction of the Hex-Elementization extends and explores the concept in several other relevant fields of application. The idea envisaged in a two-dimensional aspect in Chapter 1 can be further developed on a 3-dimensional (3D) framework. The three-dimensional rendering of a single Hex-Element (explained in Chapter 1) helps explore possibilities beyond the current six-side/properties of a two-dimensional hexagon. A 3D metamorphosis of the 2D Hex-Element can be rendered in either a hexagonal prism or in a hexahedron or a truncated octahedral. The hexagonal prism has eight faces, eighteen edges and twelve vertices compared with just six faces six vertices in a simple 2D Hex-Element.

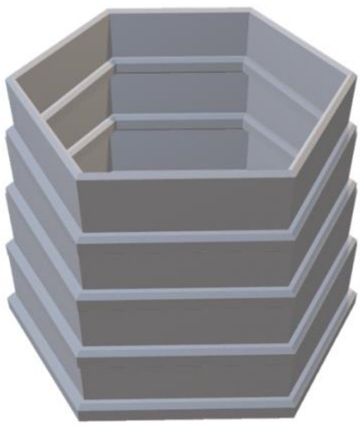

Figure 43: 3D rendering of "hex-elements" in the form of a hexagonal prism

While the 2D hex-element discussed in this thesis serves the purpose of explaining the power of granularising data into its most atomic level and ease of integration, the 3D manifestation of the same can entail in additional dimensionality of the data. For example, the 3D prism shown in Figure 43 can be manifested as a knowledge bank of a multitude of 2D hex-elements stacked on top of each other. This 3D rendering of hex-element can also serve as a history of the decisions it helped create in the past. Previous decisions and the combination of characteristics which were pivotal in making those decisions can be compared against the efficacy of the decision itself. Based on the success or failure of the decision, the 3D hexagonal prism rendering of data can be useful in correcting itself for future decisions. The future direction can be theorised as the concept of Hex-Elementization as a learning model based on the 3D rendering of the concept.

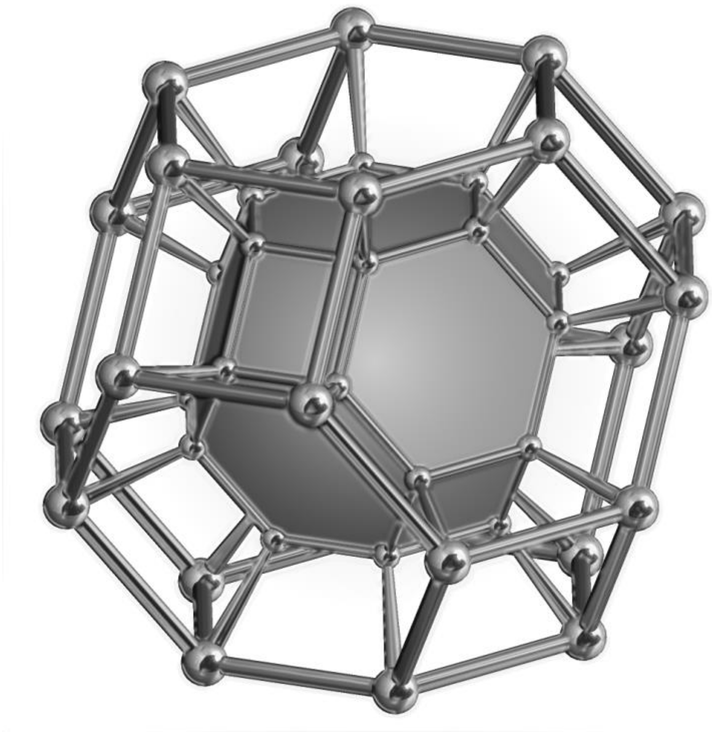

Figure 44: Truncated octahedron

In the case of the second 3D rendering of hex-elements, in the form of a truncated octahedron (Figure 44), the structure is a mix of hexagons and squares (for proper adhesion). This structure has 14 faces, 36 edges, and 24 vertices. A data point broken in such a metaphysical form not only contain more features, properties, and characteristics but also helps connect with other data points more easily. This ease of interconnectivity is because of organically generated form of hexagon structures, which symbolises efficiency and structure. As explained in Chapter 1, the interconnectivity in beehives is what makes them stronger and efficient. For business intelligence to be organically generated, data in disparate form, quantity and structure needs to be in a coalesced state. Extracting of business insights from a coalesced data lake becomes easier because the connection between data items at the most granular level is already made. Business can not only extract insights faster but also make real-time decisions.

IoT and Hex-E

The characterisation of Big Data as high Volume, Velocity, and Variety also points to the most typical source data with these characteristics – the Internet of Things (IoT). The IoT (IoT Paradigm) is a generic term used to describe any sensor-based gadget or device that collects and sends data to a repository – usually residing on a Cloud. The challenge of automation is data connectivity, which is what Hex-E is designed to handle. This is because the speed with which a group of IoT devices collect and send data (stream data) results in ever-increasing volumes and velocity of data. This data also has variety, including audio, video, graphics and sensor data, that presents challenges of defining the data points correctly. Furthermore, such data is time sensitive. Once the currency of such data is over, its corresponding analytics start losing their meaning for business decision-making. The Hex-E platform directly taps into the IoT devices in order to define data points dynamically through their six parameters.

The Internet of Things forms the basis for many devices and corresponding applications. IoT devices are being extended to machine-to-machine (M2M), transportation, healthcare, nanorobotics, and utility domains. IoT devices are becoming integral to both the vertical markets as well as cross-market (horizontal) amongst business processes. IoT is a disruptive technology as it dismantles and reorganises business processes across the Internet and in turn impacts the very business models.

IoT, at its core, is sensing and ingesting data from its surrounding – including other objects, machines, systems, and human beings. The challenge is to utilise IoT for business decision-making – by navigating through the myriad of data types, information exchange systems, applications, and hardware and communication systems. Organisations are eager to quantify consumer behaviour, individual tastes and preferences, and behavioural patterns through IoT. Such quantification through analytics enables organisations to customise and personalise product and services on an individual basis.

Analytics of Big Data generated by IoT devices is not just a technical or statistical challenge. The need to devise an IoT-based business strategy has never been higher because, as a disruptive technology, IoT challenges business models, economic principles and user expectations. As IoT becomes ubiquitous, it leads to a lot of un-interpreted data. Such data that cannot be interpreted for decision-making is not only a waste, but it puts an additional burden on the organisation owning it – such as its storage, backup and security. The challenge in handling IoT is to ensure that the data emanating from the IoT is stored and connected in an increasingly meaningful way (Ali, Ajmi, & Ali, 2018). However, with many large organisations using millions of IoT devices which communicate with each other in many ways and using a variety of protocol, it becomes very challenging to find a common thread to connect the data in a meaningful way.

Hex-Elementization aims to create a platform (i.e., an architecture, an environment) encompassing end-to-end integration between business architecture and business strategy utilising IoT. The traditional way of thinking about a data point as atomic is insufficient in the IoT-driven Big Data world. The larger and more diverse the data is, the more it needs to be broken down into smaller chunks that enable and ease interaction. These smaller chunks (i.e., hex-elements) can be combined and related to meaningful information as shown in Figure 45.

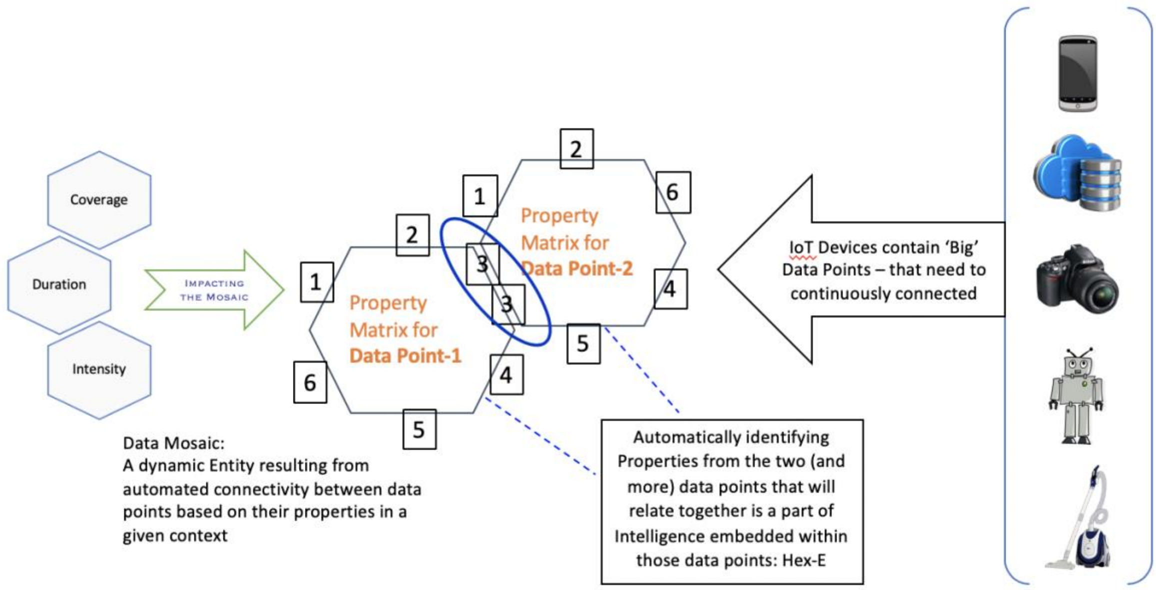

Figure 45: Hex-Elementization as a mechanism for Context of a data point in an IoT environment

IoT, Big Data and Hex-E

The propensity to gather data will increase exponentially with the adoption of IoT. However, for IoT to become synonymous with pervasive or ubiquitous computing, its real potential lies in real-time data analytics. Business intelligence, business planning, product planning, and product design are some of the activities that benefit from real-time data analytics. For example, real-time smartwatch usage information, including data collected from internal accelerometers and GPS, can help both product designers in gaining valuable insights on various usages types. This insight into versatile use of smartwatch can help product designers improve the design of their product. Also, data collected from the smartwatch and way the operating system collates and stores data in the smartwatch can help software developers fine-tune user interaction. IoT, with its ability to drag a large amount of data (Big Data) with it, has the potential to improve the productivity of various industries, including healthcare, transportation, energy, and the IT industry.

There are three specific aspects of IoT and Big Data that can be handled through Hex-E:

1. Big Data precedes IoT in form and adoption, but with IoT becoming more mainstream, the adoption of Big Data is accelerated in the form of Cloud-based and real-time Big Data processing solutions. Hex-E can help make big-data be more meaningful and easier to understand. The biggest issue with Big Data collected through IoT is integration with other datasets to create value. Integration is the key in Big Data which can be simplified with the adoption of Hex-E.
2. Hardware is becoming more intelligent as IoT devices are becoming smarter with each iteration of a device as a result of higher levels of processing onto smaller chip sizes. Increased processing power could cause businesses to sacrifice ergonomics and aesthetics. Finding a balance between these two factors of ergonomics and aesthetics is challenging. Hex-E aims to break the silo-based hardware structure set up in existing enterprise designs. Hardware, whether it is a server, desktop, tablet, smartphone or sensor, has a protocol to deal with data. Hex-E aims to break this silo-based approach of hardware by providing a common platform in which to interact. Hex-E proposes a platform which can translate the language spoken by different hardware to provide understanding of the data from each.

3. IoT will help accelerate the adoption of business intelligence and analytics on a real-time basis to make real-time decisions, but the question is how? The traditional business intelligence approach is time-consuming, modular and linear. This linear design of business intelligence systems makes it less nimble when it comes to incorporating new and changing data outside the domain of expertise. Traditional BI platforms are designed in a rigid structure, which is hard to customise. This rigid structure seeks insights to attain specific business goals, rather than working towards the big-picture context of the stakeholders. Thus, it makes traditional BI powerful from a management and oversight perspective, but not from an Agile approach. Hex-E proposes a business intelligence platform that is context driven. Businesses have goals to achieve and these goals can become complicated to achieve given the contribution needed from various departments within a typical organisation. These goals can be achieved once the basic building blocks are generalised with the help of a platform. it is a challenge for businesses to have this collaboration between departments in order to share and leverage data which can be used to achieve the business goals. Hex-E aims to be that platform, providing a common-ground on which departments can organically and easily correlate their insights.

Currently, IoT process streams have a linear flow of information (after processing Big Data) from an IoT device to the user and vice versa. However, IoT should encompass the availability of intelligence and data analytics across its entire value chain. An example is the autonomous car, which can send real-time information to tire manufacturers on what is causing damage to tires, to engine control unit (ECU) producers on optimal processing levels, to traffic departments on which roads numerous cars are avoiding, to municipalities on how certain streets are becoming congested, to smartphone manufacturers on which apps the user had problems with when inside the car, and many more. IoT and Big Data can share intelligence not only between the provider and the receiver but across the entire IoT value chain.

Hex-Elementization breaks down the barriers between the three stages of data analytics: data collection (Big Data), data analysis, and response. Once data from disparate IoT devices is broken into hex-elements, the Hex-Elementization process will find information paths as shown in Figure 4. These information paths can provide snippets of associated information to the IoT user or real-time insights to a business that has incorporated IoT at various stages of its workflow. By automatically finding relationships based on the common properties between hex-elements, Hex-Elementization aims to turn Big Data into quick and intelligent data. Marketing and sales campaigns, product designs, and customer response are more effective when they are real-time. In the fast-moving consumer goods (FMCG) sector, tracking fleet, transportation, distribution and inventory levels on a real-time basis will help FMCG businesses respond to adverse macroeconomic conditions without any lags.

Vertical and Horizontal IoT Solutions that can be optimised with Hex-E

The proliferation of IoT products has forced current IoT producers to adopt either a vertical or horizontal business model. Vertical business models are constrained within the same product line and can operate vertically within the product’s value chain (Tadejko, 2015). The issue is that within this widely adopted model, users can demand different systems, which are designed to achieve pre-defined tasks with their array of support services. Horizontal models, on the other hand, work across different value chains and different product lines (Severi, et al., 2014). The goal is to nurture innovation in the IoT industry by allowing multiple players to share, contribute, and adopt open functionality. The objective should be to make IoT devices and data exchange more readily available and distributed across vertical and horizontal business models. A hybrid business model is ideal where a complex information system facilitates free and easy sharing of data between IoT devices.

Hex-Elementization framework enables creating hybrid IoT solutions by breaking down the barriers inherent in vertical and horizontal IoT solutions. A technology business can utilise a Hex-Elementization framework to enhance its software services and products; it can also learn about the user interactions with the software in order to give input to its hardware division so it can improve its product design. For example, Microsoft could use its knowledge of how its users interact with MS Office applications to better design Microsoft hardware (e.g., Surface, keyboards, mice, gaming controllers), or it could pass on the learnings from its hardware users to make its software more intuitive and easier to use — all on a real-time basis.

Consumer Services Improvements through Hex-E

The impact of IoT will be most significant on the consumer. From health-tracking bands to tablets, smart glasses and self-driving cars, it is hard to imagine any aspect of a daily life that will not be impacted by advances made by IoT. As IoT devices multiply, a vast array of usage data is broadcast. This data usage might emanate from a single device, a multitude of devices, or in a multitude of ways of communicating with the same device, forcing action on individual chains of information from each IoT channel. For example, people interact with phones through motion, voice, touch, and position to applications or other people. An IoT ecosystem in which all the disparately built, differently programmed, and varied interactions across the IoT devices can communicate on the same level playing field is the optimal desired end result.

Hex-Elementization aims to facilitate this communication by laying down a series of ubiquitous communication channels. An ecosystem built on Hex-E not only lets different devices from different producers communicate with each other but helps in ubiquitous and real-time processing of data to make day-to-day life and corresponding decision-making easier. The need is for IoT to expose the data and information through the IoT devices and provide real-time help in everyday activities through a combined and concentrated effort across all IoT devices.

Hex-Elementization will find the best-suited paths and decision trees while processing through the relationship set between hex-elements. Figure 4 shows how the information processing and information flow will occur after getting the inputs from various devices linked to the Hex-E framework. Based on the unique but related characteristics, Hex-E will provide a platform for self-sustained and directional decision or information branches to provide real-time analytics to the end user. Having ubiquitous processing and computations is pivotal for the success of IoT. Finding correlations in real time will enable IoT to take centre stage in the next data revolution.

Artificial Intelligence and Hex-E

Artificial Intelligence (AI), is progressively evolving into a key driver that shapes day-to-day business processes and business intelligence decision-making. AI in this context is the broad umbrella of technologies including machine learning, neural networks, and deep learning applications. Given what AI has achieved in numerous fields, including supply chain, health care, weather forecasting, product designing, it is clear that it will play a big part in the future of business intelligence. Several recently developed AI capabilities include voice recognition and digital assistants, task automation, medical diagnoses, and facial recognition systems. Because of advances in cognitive computing, businesses can now use sophisticated algorithms to gain insights into consumer behaviour, use the real-time insights to identify trends and make informed decisions thus, providing a competitive advantage.

Business intelligence needs to evolve from being reactive to providing proactive analytics. The proliferation of new Big Data sources, including smartphones, communication signals, tablets and IoT devices, means businesses no longer wish to be weighed down by the huge repository of latent and static reports generated by BI systems. Businesses need the ability to provide actionable insights. BI that moves from reactive analytics to proactive analytics can offer alerts and real-time insights. These analytics allow businesses to make better use of operational data while it is fresh and actionable. Although the initial designs of BI applications were to provide descriptive analytics, some advances have been made in recent years to provide predictive analytics. Predictive analytics enable businesses to have future insights. Although no statistical algorithm can provide accurate predictions 100% of the time, organisations are using analytics to forecast future events with a higher probability than before. BI needs to move into prescriptive analytics that enables users to prescribe various possible alternatives and recommend actions accordingly towards viable solutions. Prescriptive analytics is all about providing advice and suggesting solutions. These AI-powered analytics not only predict what will happen but also explain why it will happen.

Embedding of intelligence within data is becoming increasingly difficult, if not impossible, because of high volume, high-velocity data input. Hex-E concept, elaborated in this research, aims to utilise AI to create automated correlations between data points. Should data points be able to relate with each other automatically, opportunities are opened to create previously unknown analytical insights. Hex-E implemented alongside AI can democratise data and improve analytics adoption. As displayed in Figure 46, AI can work alongside Hex-E, providing a layer of data assurance, assisted search, automated process flow and machine-learning driven metadata creation. This hybrid BI framework involving Hex-E and AI will need a management layer responsible for security, access, messaging, monitoring and interfaces.

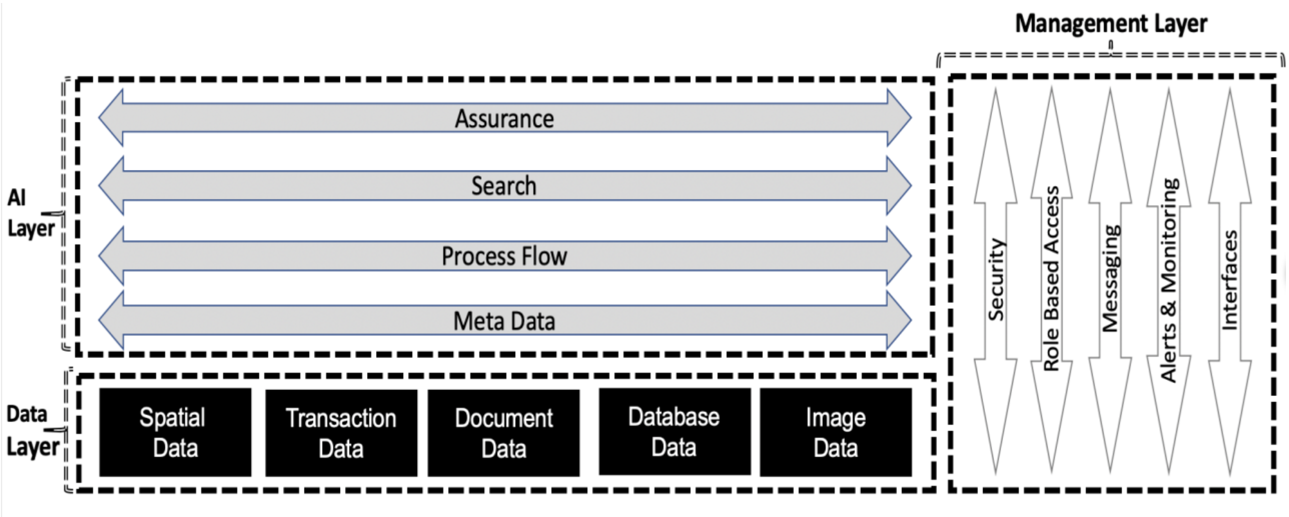

Figure 46: Conceptual framework of Hex-E + A.I. Layer

Radical advances in computing power, predictive analytics, machine learning, neural networks and other AI algorithms have opened the door to a new generation of BI built upon the foundation of Hex-E. If appropriately implemented, Hex-E plus AI can help pull actionable insights from complex data sets and automatically prescribe actions. Hex-E combined with AI can provide data points exactly where and when they are needed. No additional work is envisaged by the business user. By analysing all the data that has been collected, Hex+AI enables the use of data for competitive advantage.

One of the ways to visualise Hex-E architecture with Neural Networks (one of the sub-technologies under AI) is demonstrated in Figure 47.

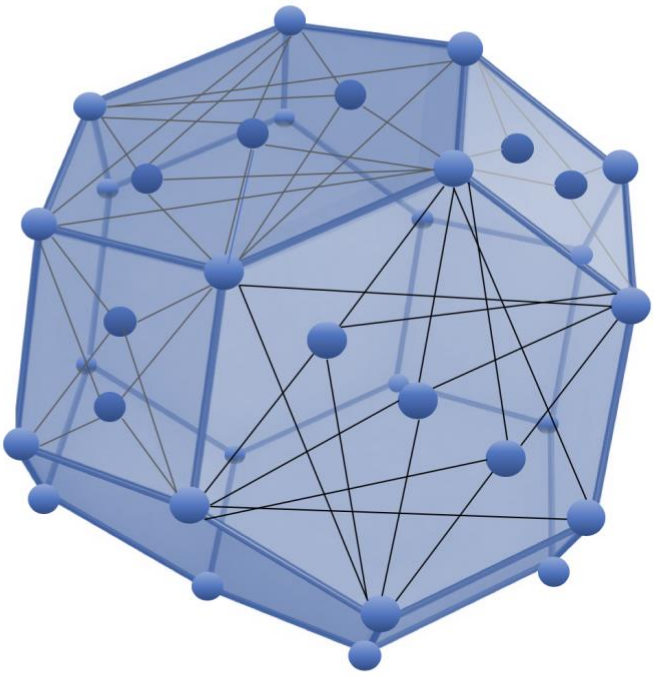

Figure 47: Rendering of Hex-E with Neural Networks

The theoretical implementation of Hex-E shows how an example neural network can run with Hex-E. The two-dimensional Hex-E framework, conceptualised in Chapter 1, is rendered here as a three-dimensional hex-element. The sides and edges on each dimension have embedded the unique characteristics for that hex-element. The characteristics are connected through input, output and hidden layers of a neural network to form the connection between the characteristics which are used when multiple hex-elements from various sources need to be connected. Thus, neural networks can be the way to characterise the unique features of each hex-element in two-dimensional or three-dimensional implementation of Hex-E.

Finally, one of the significant benefits of Hex-E + AI is to bridge the gap from data to intelligence for business users who do not have the in-depth technical knowledge, which is currently one the biggest hurdles in adopting BI (discussed in Chapter 4). Hex-E + AI will have the ability to analyse massive quantities of data and provide recommendations. It makes analytics and Big Data insights accessible and understandable to the average user, not just data scientists. Hex-E + AI will be based on the context of a typical business user. This context is driven by business goals such as maximising profits, increasing market share or streamlining supply chain issues. Where a sharp divide once existed between two groups of users, those experienced enough with data science to gather value from the data themselves, and those who are not experienced, Hex-E + AI can bridge the gap and deliver insights in a digestible format.

AI, like any technology, has limitations especially in areas of cybersecurity or inherent bias. However, the potential benefits are enormous. Despite the potential benefits, many businesses are lagging in recognising AI as a trend in the future of business analytics and intelligence. Businesses that begin to implement AI will soon be outperforming competitors when it comes to improved overall business performance and higher revenue.

Robotics and Automation with Hex-E

According to a study done by Oxford University researchers, 45% of American manufacturing will be done by robots in the next two decades (Rutkin, 2013). Robots will take IoT from its current role of simple monitoring to a higher level. This higher level will entail a role change of IoT sensors from mere data capturing devices to becoming an integral part of data integration. Robots will need to have a common ground for communication as the adoption of IoT and Big Data becomes mainstream. This common ground will help businesses in making both strategic decisions and tactical decisions (Unhelkar, 2017). As IoT augments the adoption of Big Data, AI will increase the use of robots and show how robots impact automation and manage daily tasks.

In a scene from the movie *Transformers*, robots are shown emanating from small metal balls. The scene shows how, in the future, smaller, AI-driven, self-sustaining granular and intelligent robots (nanorobots) will come together to create a larger and unified intelligent robot (or an object). The fictional work in the movie can be made more realistic through the application of Hex-E. Hex-Elementization can lend its basic principles to this physical transformation function, as well as laying down the principles of tight-fit integration in the future world of nanorobots. In reality, it is difficult for small metal balls (circles in 3D) to combine to create a larger structure but is easier for six-sided hex-elements to do so. Hexagons shapes not only can facilitate tight integration, but as discussed in Chapter 1, are symbolic when it comes to structural strength and efficient integration, which circular (balls) cannot offer.

From assisting this physical transformation in robotics to driving the development of AI, Hex-E framework can be extended in a multitude of directions. The basics of AI are about machine learning which is the ability to find and learn patterns for decision-making. Hex-Elementization can help establish integration and data interchange protocols between robots where the AI and the integration of IoT data and devices assist with IoT analytics. For example, if a smart dirt sensor learns from the smartwatch that the user is allergic to the dust it has detected, it can instruct the robotic vacuum cleaner to clean the floors by the time the user gets home — which it determines after ascertaining the estimated arrival time of the user from the user’s smart car.

Managing Risks in Decision-Making

Decision-making in an organisation often occurs in the face of uncertainty about whether a choice will lead to benefit or disaster. The risk is the potential that a decision will lead to a loss or an undesirable outcome (Renn, 1998). Typically, any human decision carries some risk, but some decisions are riskier than others. In organisational management, decision-making is always closely associated with risk management. Decisions are full of risks and uncertainties, and the effective management of the latter creates successful outcomes for an organisation.

Risk can be summed up in two parts: the probability of something going wrong, and the negative consequences if it does. Risk can be hard to spot and factor in during the decision-making process. However, preparing for and managing risks is pivotal for organisations to survive. Not planning for risk increases the chances of financial losses, reputational losses and intangible resource losses, such as time. Risk analysis and management are essential tools of any organisational decision system. Risk management can help identify and understand the risks that one could face, and in turn, helps manage risks, and minimise impact on organisational plans. Before highlighting how Hex-E can help manage risk, the next section lists the different type of risks an organisation can face under different circumstances.

Types of Business Risk and Hex-E

This section briefly lists the types of risk which an organisation can come across during the usual course of business. Understanding types of risk can help businesses factor undesirable outcomes into the decision-making process. Some of these risks can be part of a business's business intelligence framework. The application of Hex-E could help manage some of these unforeseen risks. How Hex-E can add value is briefly highlighted next in each risk.

1. **Human** – Illness, death, injury, or other loss of a key individual. This risk is less relevant from a business intelligence perspective; however, appropriate succession planning is important to avoid such risks. Hex-E aims to remove the subjectivity brought by humans by objectively analysing the factors impacting decisions from various sources, based on the context provided by humans.
2. **Operational** – This risk includes disruption to supplies and operations, loss of access to essential assets, or failures in distribution. For example, sanctions against a country can disrupt businesses which depend on that country for certain resources (e.g. oil from Iran). Business intelligence is pivotal in understanding the full extent of the operational risk. Pivotal parts of operations which are susceptible to risk, include supply-chain and distribution channels. The benefit of a business intelligence system like Hex-E is that it can be driven contextually. When an information path breaks down in Hex-E (Figure 47), the system will find the next alternative path, which is optimal to fulfil the context of finding an optimal supply chain. The automatic connection of common attributes helps Hex-E to seek out alternatives without explicitly instructing it to do so.
3. **Reputational** – This risk includes loss of customer or employee confidence, or damage to market reputation. There are some parts of this risk which can be managed by an organisation, while others are harder to manage and can be unforeseen. For example, customers need to be satisfied with good quality products or services at a good price and at the right time. Quality, price and time are the three most important aspects to keep the customers happy. These are part of the operations which can be quantified and controlled. Business intelligence systems like Hex-E should be able to manage the variables which can be quantified and controlled easily. Quality control can be built

into Hex-E through machine learning, where the end products are effectively passed through quality assurance checks. Price of products or services is partly driven by how much it costs to produce or manufacture. Again, the variables governing price is, to certain extent, a controllable factor. Hex-E could manage and track effective supplier channels which ensures ingredients or parts used to manufacture a good, is sourced at the best available price. Keeping the cost of manufacturing down can help manage the end-prices offered to the customers. The variables of a reputation risk, which are hard to quantify and track, are hard to manage and control. Finding processes within various departments, which are hard to quantify or monitor from a risk perspective, help businesses design mitigation measures. Risk control measures can help track activities which could potentially adversely impact reputation. Implementing risk controls throughout the organisation can help a business intelligence system like Hex-E monitor and track risks which can adversely impact reputation.

4. **Procedural** – This risk includes failures of accountability, internal systems or controls, or from fraud. Accountability and internal systems controls are the facets of an organisation which have, in the past, been managed by business intelligence applications. Again, similar to the risk discussed earlier, the activities which can be easily quantified and monitored can be managed and handled by a BI system like Hex-E. However, fraud detection is the hardest kind of risk. Fraud mainly involves people who act (intentionally) for their own benefit, secretly depriving something valuable to the business. There are many ways one may conduct fraud and cheat on others. The traditional methodology of data analysis has been used for the detection of fraud by many organisations. However, this data analysis requires complex techniques and is a very time-consuming practice, requiring wide knowledge based on the kind of fraud

conducted. Therefore, there are various training provided to security agencies in order to handle the situation that is unforeseen and implement the corrective measures. One way to systematically detect fraud in an organisation would be to deploy Hex-E BI framework with AI technologies. Statistical intelligence can be embedded into a Hex-E driven business-intelligence system, which can be powered by machine learning algorithms. Machine learning algorithms including analysing internal user profiles, creating learning models on user engagement internally and externally, and creating probability weighted user interactions, are a few techniques which could detect symptoms of fraud. Embedded machine learning and neural network algorithms within Hex-E architecture could perform automated time series analysis, anomalies detection, clustering and classification of behaviours which can potentially work as early warning system within the business intelligence application.

5. **Project** – This risk includes issues such as going over budget, taking too long on key tasks, or experiencing issues with product or service quality. Project management is again a quantifiable activity in which the progress of a project can be measured at every stage of completion. Business intelligence systems like Hex-E could handle verification of activities, such as cost overruns and deadlines, which could provide an early signal on whether a venture is running on time and optimal costs trajectory. Hex-E could help monitor planned cost outlays or deadline. Hex-E could also provide alternatives based on the current progress of the project, suggesting that the system can not only highlight preventive measures but also helps in prescriptive advice.
6. **Financial** – This umbrella of risk includes business failure, stock market fluctuations, interest rate changes, or non-availability of funding/credit risk. While economic slowdown and frequent interest rate changes are factors which can be quantified and

measured accurately, BI systems built on Hex-E architecture should be able to factor in and provide alternative courses to help navigate through difficult market conditions. Stock market fluctuations are hard to predict; however, easier to quantify and relate. There are some factors like economic conditions and political actions, such as fiscal stimulus, which are easy to monitor. Often, stock market cycles follow economic cycles. By quantifying macro-economic data, a Hex-E based BI system can factor in the impact of various economic scenarios on the business. Given that Hex-E can coalesce data across different datasets, it should be able to provide optimal insights to the management based on various scenarios, which can help businesses make appropriate and timely decisions.

7. **Technical and Cyber risk** – Of the all the risks highlighted in this section, the technical risks are broader and the most difficult to address. The difficulty in addressing these risks is due to the constant advances in technology. As a result, it is difficult to comprehend and pinpoint a technical failure. Cybersecurity is also one of the biggest sub-threat within technical risks and has the potential to damage not just the reputation of the business but has the potential to shut down the business completely.

Business data and information are among the most valuable assets of any business. Businesses are increasingly conscious of the importance of the data for success in the current market economy and the competitive advantage it provides. Hence, businesses need to understand how to protect this information. Big Data analytics professionals are making use of preventative technologies, as well as managed detection and response services. Businesses use these preventive technologies to deal with the constantly evolving, sophisticated cyber threats caused by the increased amounts of data being generated daily. Cybersecurity is a field which is at the top of the priority when it comes to managing a different kind of risks. Businesses are increasingly dealing with a) data residing within the vast number of applications in their network, b) data acquired and utilised from external third parties, and c) increasing number of accessibilities through mobile devices, tablets and wearable technology. Businesses are collecting vast amounts of data about their users, and hence, privacy and security have undoubtedly become primary concerns.

There are four significant aspects which businesses need to monitor against cyber threats, and these are network, applications, data and users. The way businesses can handle the cyber threats to these four things systematically is by assessing threats and risk, analysing previous attacks, and monitoring in real-time. The security-related information available from Big Data reduces the time required to detect and resolve an issue, allowing cyber analysts to predict and avoid the possibilities of intrusion and invasion. According to a research, 84% of businesses use Big Data to help block these attacks (Morgan, 2015). The report also described a decline in security breaches after introducing Big Data analytics into their operations. Insights from Big Data analytics tools can be used to detect cybersecurity threats, including malware and ransomware attacks, compromised and weak devices, and malicious insider programs. This is where a combination of Hex-E powered BI, and Big Data analytics looks most promising in improving cybersecurity. BI analytics built on the Hex-E framework can leverage machine learning techniques to enable a business to perform an in-depth analysis of collected information. Ultimately, a system built on Hex-E plus Big Data plus AI can provide heightened awareness of a potential threat to the integrity of the business. A business can create baselines based on statistical data that highlight what is normal and what is not. With such insights, organisations and business managers can know when there is a deviation from the norm using the data collected. New statistical and predictive models and possibilities can also be created using this historical data by the use an architecture which includes Hex-E and AI.

Another dimension of cyber security from within the organisation is monitoring internal data. It is valuable for any business to use an employee system monitoring program that relies on a Hex-E based BI system along with Big Data analytics. This hybrid system can provide insights which the business's compliance officer or the Human Resource manager can use to replay the behavioural characteristics of an insider. Businesses of any size, small or big, can prevent the compromise of the integrity of their systems by establishing a hybrid monitoring system with Hex-E + BI + AI. A hybrid system like this could limit the access to sensitive information only to staff that are authorised to access it. Staffers may use their logins and other system applications to change data and view files that they are permitted to touch. The hybrid system will continuously monitor this accessibility of sensitive data. The data collected through the monitoring system can help the hybrid system to create learning models to differentiate between genuine vs malicious use of the data.

Such a hybrid system can also help develop an intrusion detection system. The hybrid system can monitor traffic flowing through firewalls, establish and monitor multi-factor authentication and effectively manage data encryption. Thus, such a hybrid system could be built as an intrusion detection system to provide early warning for activities it classifies as having malicious intent. A hybrid system propelled by Hex-E would keep the business context in perspective when managing cyber threats. It is also essential to comprehend business requirements in order to make an informed decision about deploying such a system. Sometimes, the need to incorporate external data might be pivotal to gain specific industry insights. There might be a genuine reason to create a protocol to be able to do this without creating a hole in the existing firewall or evading the firewall. Such activities could be added as an exception to the hybrid system, which will keep the business goals (context) in mind when monitoring and allowing access across the internal systems.

Another benefit of a hybrid system based on the Hex-E, BI, AI combination provides a method to combat cyber threats of the future. Technology is evolving at a fast rate, and hence the tools used by malicious users and cyber thieves are becoming increasingly complex and complicated. Businesses must build defences that can withstand increasingly sophisticated cyber-attacks. The benefit of the proposed hybrid system is that the AI side of it will use machine learning over time to learn the increasing level of sophisticated attacks over time. Such a system will automatically learn and evolve with each incremental attack. Also, given the way Hex-E has been modelled, it is easier for an architecture built on Hex-E to be introduced to a new system or technology. Hex-E, using the hex-elements route, can seamlessly incorporate the technical rules of the new system or algorithm in the BI system. Then AI part of the hybrid system can easily incorporate the new rules of engagement for any future threats.

Big Data means big responsibility (Wiewiórowski, 2015). There are many layers of IT security with an on-premises data environment. Firewalls, intrusion prevention systems, vulnerability management scanning apps, and private and public Cloud hardware and software are only the beginning of safeguarding corporate or organisational data. A BI system involving Hex-E could provide a single and consolidated solution to effectively handle the cyber threats reducing the chance of adversely impacting business integrity.

8. **Natural** – Natural risks are risk related with weather, natural disasters, or disease adversely impacting business operations. Unforeseen natural disasters, adverse weather conditions and epidemics have the potential to impact the profits of businesses adversely. There are a few industries, like health care and consumer staples, which might be more impacted than others; however, incorporating weather data has been an emerging trend in recent years. Incorporating weather data into the business intelligence domain has the potential to safeguard the business operations to some extent. When it comes to weather, it is not about risk avoidance, as natural disasters cannot be avoided. In the domain of risks related to nature, the business strategy is more about risk minimisation. The past few years have seen many advances and increased amounts of weather sensor data. Most of this data is flowing from IoT (Internet of Things) and smartphones. This data deluge, coupled with the rise of Cloud computing and data science technologies, has brought unprecedented advances in forecasting accuracy. Some businesses in particular industries are starting to capitalise on it, while others are pondering alternatives to incorporate this data into their business intelligence platforms. Highly precise and near-real-time forecasts help insurance carriers reduce impairment claims, retailers optimise staffing, packaged

goods businesses sell more shampoo, and consumer staples businesses manage perishable goods inventory.

As discussed earlier, a hybrid BI implementation using Hex-E and AI could augment the ability of businesses to factor in weather data in operations. Hex-E, as postulated in Chapter 1, will have the ability to incorporate data from different sources by breaking it into the most granular set. Weather data comes in the form of a structured big-data set, which can be crunched using big-data technology within the Hex-E framework. The ability of Hex-E to then link weather data with other data sets, including supply-chain management, distribution management and inventory management can help businesses manage the impact of adverse weather conditions.

Moreover, such a hybrid system has the potential to open the use of weather data in businesses which have not found a way to incorporate this data into day-to-day activities. Increasing or decreasing demand for certain goods and services, in specific industries, coincides with changes in weather. Restaurant chain owners could use the hybrid BI system to analyse the order patterns and customer flow with weather patterns to make few decisions. These decisions may include adjusting menus, or how much meat or vegetables to buy, and store, based on results from the weather-based analysis. Consumer staples businesses can figure out what times of the year they should increase advertising, showcase specialty soups, or offer a discount on cold desserts. Operations managers can use weather and climate data to control the electricity use and tune the temperature control systems of buildings (Evans & Gao, 2016). By embedding smart systems that report on usage and combine it with weather data, operations managers can better forecast and control operational costs. These are just a few examples of adding value by incorporating weather into other functions of the business using a hybrid BI system powered by Hex-E.

9. **Political** – Political risks include the risk of changes in tax, public opinion, government policy, or foreign influence. Political risk results from numerous extraneous factors that can negatively impact a business's revenue or complicates it's business strategy. Apart from the factors listed above, other factors including macroeconomic issues like monetary policy, civil unrest, fiscal policy (tariff rises), and authoritarian government actions, can adversely impact the ability of businesses to operate normally. A change in the ruling political party and the ensuing changes to fiscal policy may mean a business will be unable to convert foreign currency, export or import goods and supplies, or protect in-country assets.

To manage political risk through a business intelligence framework, business stakeholders first need to identify and classify business risk. Next, managers must quantify the impact of specific political risks on the business's performance using financial modelling. Business stakeholders then need to associate the financial impact on the business's risk tolerance. Moreover, the business stakeholders need to implement a standard risk response to manage the risk. This four-stage process is an essential component of political risk management within a BI framework.

Along the lines of what has been discussed in earlier risk management topics, a hybrid BI system built on Hex-E and AI can help in varying degrees in these four stages. The hybrid system can identify political risk in an automatic and real-time basis by monitory high-frequency news sentiment data. There are third party news aggregators like RavenPack, DJ News, Alexandria Technologies, Bloomberg Event-Based news which can be tapped into by the Hex-E based hybrid BI system. Once the business stakeholders classify and contextualise the political risks which might adversely impact their business, Hex-E can constantly monitor the news to highlight any growing political risk on a real-time basis to the business owners. This hybrid BI system should be able to handle the second stage of political risk management – which is to quantify the impact of specific political risk on the business's revenue using financial modelling. Advances in machine learning and neural networks can effectively handle linear and non-linear variables in a financial model and can easily forecast the impact on the revenue given a certain political change. The third step is where business stakeholders need to work closely with the hybrid system to specify the risk tolerance based on the results from the financial modelling. Risk tolerance is a measure of how much a business is ready to forego (in losses) in the event of a political issue. Once business stakeholders define their risk tolerance, it can be quantified by the hybrid BI system. This quantification of risk tolerance can be a learning model based on machine learning or deep learning which continually adjusts the risk tolerance based on past occurrences of events and the probability of an event happening in the future. The more the algorithm understands and quantifies a political risk, the better its predictions become in future. The fourth and final part of political risk management is to provide advice to the business stakeholders on possible risk response options. Given Hex-E is envisaged to tap into every internal data system and useful external data systems to help contextualise business goals, adding risk management parameters will increase the number of criteria or constraints it needs to solve to provide optimal and automated results. Such a hybrid BI system with cheap computing and storage resources can easily incorporate a multitude of such risk parameters to provide a holistic prescriptive risk response to business owners.

10. **Regulatory** – Regulatory risk is the impact of a change in regulations or legislation on businesses' operation. Businesses must accept the regulations set by governing bodies that oversee their industry. There are sometimes first-order impacts of regulation in which the business's industry or sector gets directly affected by new regulation, and other times there are second and third order impacts in which a business's partners or customers get impacted by an impending regulatory change. Hence, any change in regulations can cause a rippling effect across an industry or sector.

    Regulations can increase costs of operations, introduce legal and administrative hurdles, and sometimes even restrict a business from doing business. Regulatory requirements continue to increase in scope and complexity, profoundly impacting the operating costs associated with compliance. Impending regulatory changes by government agencies, or even the frequent discussion of regulatory changes, also pose a risk to businesses. The regulatory environment is now one of the top issues with potentially high impact on a business according to 400 U.S. CEOs across all major industries surveyed by KPMG (KPMG, 2014). The same report suggested that 34% of CEOs are spending more time with regulators or government officials or considering doing so. New regulations are expensive in terms of compliance, as businesses need to transform data tracking and gathering systems, reporting functions and, in some cases, change their organisational structures. These regulations can also limit revenue growth and profitability by limiting certain products or activities. For example, the first

224 rules of the Dodd-Frank Act impacting financial institutions, is 7,365 pages (Forbes, 2019). The effort for financial institutions to comply with these regulations will be staggering and they need to dedicate an enormous amount of time and resources in order to ensure compliance.

Businesses have limited choice as the regulatory burdens continue to increase over time. Complying with regulations generally creates a drag on businesses apart from financial losses. Regulatory compliance can add costs, slow down processes and restrict expansion. The management of regulatory risk is not as straight forward as in the previous types of risk. One of the ways to monitor and manage regulatory risk through a BI system is to take data and insights that in the past have been used almost exclusively for compliance purposes and use them to drive additional value. The positive aspect of regulatory risks is that there are plenty of examples and plenty of data available for past regulatory actions. For example, sales data required solely for indirect tax compliance purposes may be analysed to provide new or better insights on product and customer profitability that otherwise was not known.

A hybrid BI system structured on Hex-e and AI can help generate these insights helping in better decision-making. Although “better” is a relative word used here, given regulatory risks can neither be minimised nor can be avoided in any circumstances. The regulatory environment is so impactful that businesses are changing their business models as a result of regulations (Forbes, 2019). Businesses are therefore faced with a risk balancing exercise. However, a hybrid BI system can systematise the regulatory requirements and help monitor whether adherence to those requirements is being fulfilled. It can help in assessing the risks of not complying on a timely manner and help notify the appropriate responsible officer. The risks associated with not complying must be measured against the risks associated with breaking the blocking statute or regulations or both, and the possibility of a criminal conviction and fines against the business or its officers or both. Monitoring risks can be effectively carried out by the proposed hybrid BI system given the regulatory rules and conditions are adequately defined. The risk of non-compliance is quantified through fines and punishments, and the responsibilities of various business functions are clearly defined. The hybrid BI system can also help the compliance department of a business in communicating with the authorities and legal contacts in the host jurisdiction to speed up the treaty route for information transfer, or by finding the information in another jurisdiction that does not have such treaty. All these aspects of regulatory risks are measurable and quantified information which the hybrid BI system can utilise to provide both descriptive and prescriptive advice.

Blockchain and Hex-E

Blockchain is a technology that has recently received much publicity owing to its association with crypto-currencies (especially Bitcoin). While the hype surrounding Bitcoin may or may not be a fad, the fact remains that blockchain as a technology has much broader applications. Blockchain, as a technology, has the potential to become one of the greatest disruptors since the advent of the Internet (Scott, Loonam, & Kumar, 2017). Blockchain can fundamentally alter electronic communication with the proven potential to affect all sorts of transaction processing systems. However, blockchain has the potential to show added value in many ways beyond the transactional processing systems, one of the first enterprise systems to embrace blockchain in practice. Blockchain has already passed the early adoption stage and is now on track of becoming a disruptive force in the market. Other possible applications of blockchain discussed in the literature and practice alike include autonomous marketplaces, invoicing, secured business transactions, smart legal contracts, robo-financial advisors and enhancing product quality and recall effectiveness. Such applications could be seen leveraging the advantages of the distributed ledger platform which forms the basis of the blockchain (Bahga & Madisetti, 2016).

Blockchain can be an effective partner of BI applications, especially Hex-E. For example, blockchain can be used in managing risks, which were discussed earlier in this chapter. Blockchain does not make fraud impossible, but it makes it harder by making fraud detection easier. Ease is relative though, and the speed of fraud detection becomes a crucial factor in defining success for both platforms as well as applications. Thus, detecting fraud might become the first major application for Hex-E based BI. Another example of blockchain being an effective partner of Hex-E based BI is in the area of integration (Tauro, 2019). Hash is a concept in the blockchain which is a result of an algorithm instrumental in encrypting a block of information or transaction. Processing hashes might need to become more common especially in the Extract, Transform, Load (ETL) function. Encrypted hashing suffers from the same problem as plain-vanilla encryption in general. The longer it takes to encrypt, the longer it takes to break (decrypt) via brute force, thus making the process more secure. However, security in blockchain comes at the cost of processing power. Blockchain's hashing hence provides a mechanism through which hex-elements can be secured before they connect and coalesce to extract insights. This mechanism provides two-pronged benefits – one is it allows effect risk management from internal fraud perspective by limiting the ability to assess information on a need-to-basis (Tauro, 2019). Any attempt to work backwards through related hex-elements can trigger an internal warning system to catch the fraud. The second benefit is to attach these hashes on hex-elements which are designed to deal with outside/external data. This attachment provides an additional layer of security from hackers or parties trying to pry on a business's business intelligence or strategies.

Benefits from the marriage between blockchain and Hex-E based BI system includes the elimination of concurrency issues when committing transactions, prevention of historical transaction re-writing, an effective and secured platform for sharing internal sensitive data, and sophisticated way to store granular business intelligence actions based on point-in-time insights and information. To elaborate on the last point here, it is necessary to understand the "block" in the blockchain. A block in a blockchain is immutable (indestructible) (Vigil, Pathak, Upadhyay, Garg, & Singh, 2018). It is not uncommon to deal with immutable data today. However, in a Hex-E based BI systems, there is a real-time search for insights, by constantly connecting common characteristics of data across many domains and functions of an organisation. At any point in time, the Hex-E system might provide a piece of advice, based on the integrated data, risk management variables, specific AI-powered algorithm, business context, among many other impactful analyses. Once the prescriptive advice is provided, it becomes action in the past, which may or may not be remembered in the future. Blockchain provides a method of storing the entire series of events, data, logic, algorithm, business variables, and the context which were woven together by the Hex-E to provide advice, in a block. This block, by definition, remains immutable, providing a valid repository of past decisions and variables. The series of blocks in this decision path can be either tapped by Hex-

E to help make future decisions or to provide support to other functions within the business which can benefit from it.

The ultimate goal in BI is to collect and produce just-in-time data, integrate it with existing data assets and channel it through on-demand modelling to help identify significant operational and strategic actions. BI also helps in disseminating valuable information across the business on time. Blockchain can become a backbone of valuable, just-in-time data that can have a transformative impact on the business. The combination of blockchain with the concept of Hex-E could potentially improve business intelligence agility by providing an event-centred methodology to collect, produce and distribute intelligence across business units, services and applications.

Implementation of Hex-E

Businesses need to navigate through the noise produced by IoT and its increased use of Big Data. Big Data has enabled businesses to accumulate and accommodate massive troves of data in sophisticated repositories. This data must be converted into information and from information to intelligent contextual information to be of value to the business. The content itself is not sufficient — what is needed is the contextual underpinnings of Big Data associated with IoT. In the example above, the context provided around the information flow is vital to its success. One does not want the dirt sensor to provide the information on the amount of allergen-containing dust on the floor to satisfy one's idle curiosity; one wants it to use that information to solve the problem before they get home. For IoT to be meaningful, the Big Data emanating from these devices need to have contextual boundaries to provide smart info (Roca, Milito, & Nemirovsky, 2018). For example, if the context is "wellness," then content in terms of metadata (which can be "amount of dust [in cc]," "allergen levels per cc of dust," "allergen levels triggering an allergic response in homeowner," etc.) can be derived and passed on to create boundaries for this wellness-related IoT application.

Hex-Elementization can provide this context through semantics embedded in the hex-elements within the same value chain (i.e., hex-elements derived from the same vertical chain of IoT products). This auto-embedding can be achieved using AI overlaid on Hex-E, which can help limit the flow of information and decision-enabling data through contextualisation. For example, a phone manufacturer does not just need to know how much sales fell after a new product launch, but why they fell and what macro or micro-environmental factors affected its customers' buying patterns. With this information in hand, the business might decide to stock its retail shelves not with an equal number of each model, but to tailor the stock according to the age and other demographical attributes of each store's location (e.g., targeting feature-filled phones to younger markets vs. basic, functional phones to an elder community).

The practical implementation of Hex-E (Hex-Elementization) can be undertaken in two ways. First, it can be developed as a protocol. Similar to how TCP/IP translates the World Wide Web into a language understood by any device, operating system, and platform, Hex-E can be built as a set of protocols which can work with TCP/IP to ease the integration of data inherent in various devices, platforms, databases, and networks. This architecture is graphically shown in Figure 48. Hex-E aims to slot in between the outside world (World Wide Web) and internal corporate networks interacting through a secured gateway. Hex-E can be configured by businesses to decide how much external data needs to be incorporated in generating intelligence used for decision-making. Businesses can turn this dial of external vs internal resourcing for business intelligence. Second, Hex-E can be built "as a Service" (HaaS). The HaaS can be rolled out as a layer between the TCP/IP protocols (via a gateway) and existing Enterprise Data Layer. The positioning of this protocol is also shown in Figure 48.

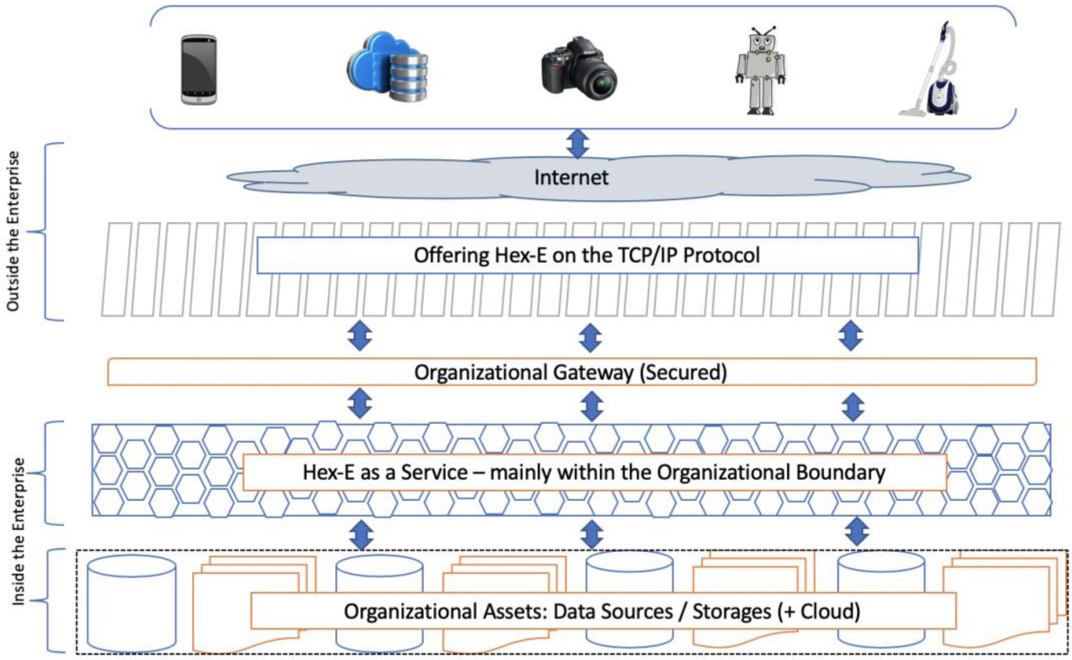

Figure 48: Implementation options of Hex-E both as a set of protocols or "as-a-Service"

Hex-E as a Service

Until a decade ago, advanced technology assets were only accessible for businesses with deep pockets. However, in the last decade, technology has been made accessible to all levels of enterprises, from small to medium size, through "as-a-service" offerings (Dempsey, 2018). Few examples of such offerings are "Software-as-a-service (SaaS)," "Platform-as-a-services (PaaS), "Infrastructure-as-a-service (IaaS)" among others. These modular offerings allow businesses to plug and play modules along with existing enterprise systems, without disrupting the IT structure.

Hex-E can be envisaged to be offered as-a-service to the organisation. As shown in Figure 48, Hex-E is placed in between the outside world and inside of the organisation underneath a security layer. Hex-E can be the point-of-passage between the hoard of existing enterprise applications and the decision-makers. The contextual underpinning of the decision-makers will drive the information exchange between the Hex-E layers and the enterprise applications. The primary goal of any business is to maximise profits, increase sales/revenue, minimise cost, and satisfy customers. Each of these big-picture goals are broken down into granular goals in a typical organisation. Each department then sets its own goals in order to play its part in the bigger picture goal. Each department uses its applications or provides and consumes data through an enterprise software system. However, these smaller silos get convoluted over time created inefficiency in information interchange. This inefficiency leads to lack of coalescence between departments in order to align their smaller goals.

Hex-E can help solve this by creating a context-driven approach. Businesses need to contextualise goals to the Hex-E framework, which in turn will tap into the large amount of application software and systems. Hex-E can then find patterns between different departmental software, systems, and structures to combine data and information to provide intelligence in decision-making.

Hex-E as a Protocol

TCP/IP protocols created big advances in society and how people use the Internet. The Internet helped connect billions of devices and networks around the world, which metaphorically made the world smaller. It takes seconds to get in touch with somebody on the other side of the world, and the credit goes to the Internet. Another uniqueness of the

Internet is that it is agnostic to operating system, hardware or network. One can access the Internet on any device, be it a PC, smartphone or tablet. These devices can run on iOS, Microsoft, Android, Unix or any other operating system. Internet can be accesses via an intranet or complex array of networks. Internet was made possible through TCP/IP protocols (Noura, 2018). The rules and procedures defined in the set of TCP/IP protocols make it possible to access the World Wide Web. These protocols translate the language of the Web into something which can be understood by any hardware. TCP/IP enables the agnostic capabilities of the Internet (Santofimia, et al., 2018).

The disadvantage of the TCP/IP protocols is that they were never designed for integration. Integration in the era of Big Data in which volume, variety, and velocity are continuously changing. A quick Internet search provides a list of known file format across various software service providers. This list currently has more than 1500 formats (see Appendix 1). The TCP/IP suite of rules do not currently read across these different formats to make it easy for businesses to make decisions.

Hex-E can potentially handle data, the way TCP/IP handled the Internet. Hex-E can be developed as a suite of protocols working alongside TCP/IP protocols enabling data integration. This suite of protocols can help break the silos currently built by major technology corporations like Microsoft, Apple, Google, and Amazon. Each of these businesses tries to tie customers (personal or businesses) by offering various applications and services within the same value chain. It is hard for customers to pick and choose applications or systems from various vendors. Even if this happens, it becomes difficult to source intelligence across the multitude of systems to make quick decisions. Hex-E protocols can help break this barrier by connecting to the intranet (which is currently run using TCP/IP) which in turn can tap into the disparate systems. In this manner, Hex-E can be inherent in the existing networks which businesses already have without making any significant investment.

A Use case of Hex-Elementization in Practice

The dataflow from a given IoT device gets broken down into hex-elements with a set of six properties. There is no limit to how many IoT devices can be integrated through Hex-E. Each set of "hex-elements" from each IoT device tries to interconnect by seeking common properties (see Figure 49). This automated interconnectedness will enrich the flow of information as it gathers more data from each new stream of hex-elements emanating from each IoT device. Regardless of the origin of the stream, the related information it seeks to create a new informational flow.

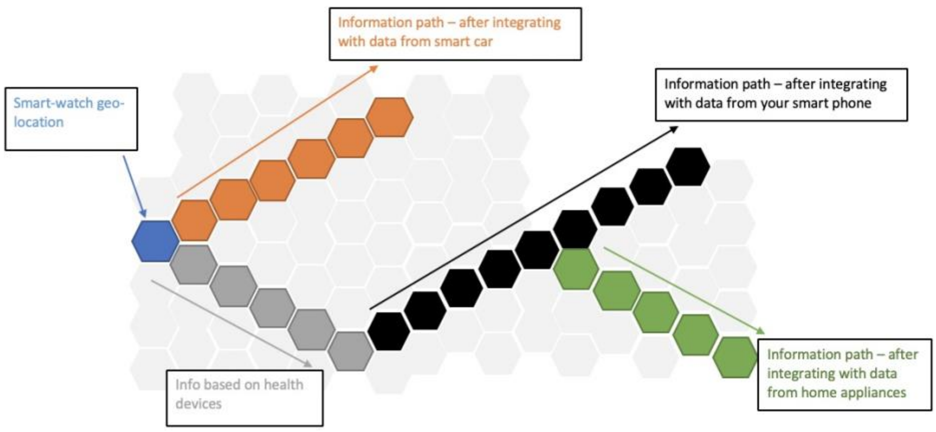

Figure 49: An example hex-element from a smartwatch and how it finds a common thread with a health monitor

As an example, consider the Hex-E pathways shown in Figure 49. In that example, the hex-elements from a smartwatch seek out and connect with the Hex-E from the car and the smart meters on the street where the car is parked. As a result, a unified snippet of information is gleaned which states: “Your restaurant is a five-minute (250m) walk from this parking spot, and you have 45 minutes before the time-based parking restrictions will come into force.” In this example, the information stream or hex stream, is made up of a chain of hex-elements from disparate sources that are grouped and unified by a common property to result in a chain of rich and interconnected information. The process is analogous to a creek getting bigger and turning into a stream and then into a river as it merges with smaller creeks and amasses increasing flow of water.

--------------------------- x-------------------------------

This signifies the end of the thesis but not the end of the thought process behind Hex-Elementization.

Appendices

Appendix 1: Format Categories and number of formats

No.	Format Category	No. of formats
1	Video Game Data	168
2	Scientific Data/Data Exchange formats	112
3	Computer Aided Design	97
4	Raster image formats	92
5	Document file formats	90
6	3D Graphics	88
7	Audio file formats	87
8	Archive/Compression file formats	86
9	Database file formats	65
10	Developer file formats	65
11	Programming Language formats	55
12	Video formats	48
13	Executable/Object Codes file formats	45
14	Security and Certification keys	42
15	System file formats	42
16	Website file formats	32
17	Electronic Design Automation	30
18	Geographical Information Systems	25
19	Virtual Machine/Data	23
20	Font file	22
21	Presentation file formats	22
22	Physical recordable media archiving	21
23	eBooks	20
24	Desktop publishing	18
25	Data file formats	15
26	Settings file formats	15
27	Vector image formats	14
28	Page Description Language formats	12
29	Spreadsheet formats	10
30	Financial records	9
31	Page layout files	8
32	Ceramics glazed recipes	7
33	Links and shortcuts	6
34	Reference Management Software	6
35	Personal Information Manager	5
36	Mathematical	4
37	Others	65
	Total	**1571**

Table 15: Format Categories and number of formats (sorted high to low)

Appendix 2: Ethics Approval

Locked Bag 1797
Penrith NSW 2751 Australia
Research Engagement, Development and Innovation (REDI)

REDI Reference: H12297 Risk Rating: Low 1 - LNR

12 October 2017

Professor Yi-Chen Lan School of Business

Dear Yi-Chen,

HUMAN RESEARCH ETHICS COMMITTEE

I wish to formally advise you that the Human Research Ethics Committee has approved your research proposal H12297 "Development of A Real-Time Business Intelligence Framework for Accurate Business Decision-making based on Hex-Elementization of Data Points ", until 12 October 2020 with the provision of a progress report annually if over 12 months and a final report on completion.

In providing this approval the HREC determined that the proposal meets the requirements of the National Statement on Ethical Conduct in Human Research.

This protocol covers the following researchers:

Yi-Chen Lan, Girish Nair
Conditions of Approval
1. A progress report will be due annually on the anniversary of the approval date. 2. A final report will be due at the expiration of the approval period.

3. Any amendments to the project must be approved by the Human Research Ethics Committee prior to being implemented. Amendments must be requested using the HREC Amendment Request Form: https://www.westernsydney.edu.au/__data/assets/word_doc/0012/1096995/FORM_Amendment_Request.docx

4. Any serious or unexpected adverse events on participants must be reported to the Human Research Ethics Committee via the Human Ethics Officer as a matter of priority.

5. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the Committee as a matter of priority

6. Consent forms are to be retained within the archives of the School or Research Institute and made available to the Committee upon request.

7. Project specific conditions:
There are no specific conditions applicable.

Please quote the registration number and title as indicated above in the subject line on all future correspondence related to this project. All correspondence should be sent to the e-mail address humanethics@westernsydney.edu.au as this e-mail address is closely monitored.

Yours sincerely

Professor Elizabeth Deane
Presiding Member,
Western Sydney University Human Research Ethics Committee

Appendix 3: Participant Information Sheet

Participant Information Sheet – General (Unspecified)

Project Title: Development of A Real-Time Business Intelligence (BI) Framework for Accurate Business Decision-making based on Hex-Elementization of Data Points

Project Summary:

You are invited to participate in a research study being conducted by Girish Nair, the Chief Investigator, student of Western Sydney University in the School of Business, under the Supervision of Professor Yi-Chen Lan. This research aims to explore approaches to ease the end-to-end transformation of data into intelligence for business. This research suggests a simplified and semi-automated framework to integrate, correlate and coalesce data into intelligence in real-time to help accurate decision-making. The research aims to propose a framework which would help in the integration of data organically into intelligence. The envisaged result of this framework is to enable business decisions in real-time in the most efficient manner (i.e., with limited resources and time). Furthermore, this framework aims to facilitate collaboration from diverse sources and disparate formats of data.

How is the study being paid for?

This study is fully funded by Western Sydney University.

What will I be asked to do?

You will be asked to share your experience in the role you play in your firm on how you make decisions. If you use data in making your day-to-day decisions and how it impacts your overall ability to make fast, accurate and timely decisions. Please refer to the Questionnaire to learn more about the information I seek to gather from interviewing you.

How much of my time will I need to give?

The interview (survey) might take up to 60 minutes.

What benefits will I, and/or the broader community, receive for participating?

The benefits can be summarised in two ways – theoretical contribution to the society and practical contribution.

Theoretical Contribution

1. This research will aim to provide a business intelligence framework to enable real-time, resource-saving, accurate decision-making in workplaces

2. Contextual underpinning of large undertakings in the age of new-age technologies, like Big DATA is just starting and this research will provide a basis of a contextual driven business intelligence framework which will be push-based.

3. This research will add a new concept to concept to the area of data and information integration especially in the world of Artificial Intelligence, Machine Learning, Internet of Things and Big Data.

4. The research will provide a theory on the most atomic level of data and information to enable integration for decision-making.

Practical Contribution

1. This research will help save resources, in terms of people, time and systems, by providing a generic framework in which disparate set of data will correlate and connect to provide meaningful and timely insight to the management to help them make real-time and accurate decisions.

2. It provides the organic growth of intelligence encompassing both internal business data (HR, Finance, Marketing, Supply Chain) and external data (Third-party data vendors, high frequency consumer data).

3. The model proposed by this research aims to provide a solution which is not architecture agnostic and will work in various situations and architecture.

4. The framework proposed by this research will create a generic framework and protocol which will entail in easier integration of any new “type” (Big data) of data entering the organisation in future.

Will the study involve any risk or discomfort for me? If so, what will be done to rectify it?

This study and the interview process do not involve/cause any risk or discomfort to the participant. The study aims to capture generic business knowledge you possess. In case the study causes any discomfort to you or if you feel it poses risk to you in any form, please approach the Primary Supervisor or the Ethics Committee (contact details further down this sheet) to escalate your concerns.

How do you intend to publish or disseminate the results?

It is anticipated that the results of this research project will be published and/or presented in a variety of forums. In any publication and/or presentation, information will be provided in such a way that the participant cannot be identified. The individual responses from the participants will be aggregated and generalised, upon completion of which, the individual responses will be deleted (destroyed) safely.

Will the data and information that I have provided be disposed off?

Yes, the individual responses collected from you will be deleted upon aggregation/generalization of data across all the participants. The aggregated data will be used as per Western Sydney University's Open Access Policy. This means that data collected from this study can be made available online and world-wide in perpetuity.

Can I withdraw from the study?

Participation is entirely voluntary, and you are not obliged to be involved. If you do participate you can withdraw at any time without giving reason.

If you do choose to withdraw, any information that you have supplied will be deleted (destroyed) under appropriate supervision. Your contribution to the study will be removed and the data will be re-aggregated excluding your response.

Can I tell other people about the study?

Yes, you can tell other people about the study by providing them with the Chief Investigator's contact details. They can contact the Chief Investigator to discuss their participation in the research project and obtain a copy of the information sheet and sending the survey link to other people.

What if I require further information?

Please contact any of the below person should you wish to discuss the research further before deciding whether to participate

1. Girish Nair, Chief Investigator, Student. Mobile: +61 402 093 626 email: girish124@gmail.com
2. Prof Yi-Chen Lan, Primary Supervisor. Phone +61-2-9683 8107 email: Y.Lan@westernsydney.edu.au

What if I have a complaint?

If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through Research Engagement, Development and Innovation (REDI) on Tel +61 2 4736 0229 or email humanethics@westernsydney.edu.au

Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

If you agree to participate in this study, you may be asked to sign the Participant Consent Form. The information sheet is for you to keep and the consent form is retained by the researcher/s.

This study has been approved by the Western Sydney University Human Research Ethics Committee. The Approval number is H12297.

Appendix 4: Participant Consent form

Consent Form – General (Unspecified)

Project Title: Development of A Real-Time Business Intelligence (BI) Framework for Accurate Business Decision-making based on Hex-Elementization of Data Points

I hereby consent to participate in the above-named research project. Please note that you can either print this form with your signature and send it to the researcher or just click on "Yes, I consent to participate in the survey" when you start your survey.

I acknowledge that:

- I have read the participant information sheet (or where appropriate, have had it read to me) and have been given the opportunity to discuss the information and my involvement in the project with the researcher/s
- The procedures required for the project and the time involved have been explained to me, and any questions I have about the project have been answered to my satisfaction.

I consent to:

☐ *Participating in an interview*

AND

☐ *Having my information audio recorded*

OR

☐ *Send my response via an online survey or via email.*

Data publication, reuse and storage

This project seeks consent for the data provided to be used in any other projects in the future.

To make reuse of the data possible it will be stored under Western Sydney University's Open Access Policy.

I understand that in relation to publication of the data

☐ **My involvement is confidential, and the information gained during the study may be published but no information about me will be used in any way that reveals my identity.**

☐ **The researchers intend to make the non-identified (aggregated) data from this project available for other research projects**

☐ **I can withdraw from the study at any time without affecting my relationship with the researcher/s, and any organisations involved, now or in the future.**

[Researchers must remove the option that is not relevant to the research]

Signed:

Name:

Date:

This study has been approved by the Human Research Ethics Committee at Western Sydney University. The ethics reference number is: H12297.

What if I have a complaint?

If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through Research Engagement, Development and Innovation (REDI) on Tel +61 2 4736 0229 or email humanethics@westernsydney.edu.au.

Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

Appendix 5: Interview Questions

Interview Questions.

1. What is the importance of data in your decision-making as compared with your (and your employees') experience and intuition?

2. Is your organisation actively increasing its Data capture? Please provide examples of new types – e.g. unstructured and various sources of data?

3. What are the key challenges you face in deriving business insights from data transformation? E.g. volume, variety of data.

4. What are the key challenges in undertaking Data-driven decisions? (for example, is there a lag between when the data is received and when the decisions are made? In other words, are business decisions based on lagged or outdated data? Does context get lost?

5. What approximate percentage of your organisation's time is spent on a regular basis in cleaning (pre-processing) and linking data from disparate sources before it can be consumed for decision-making?

6. Do you think the journey from data-to-intelligence has increased overtime? The cost in terms of **time** (to clean, understand, contextual data), cost in terms of **capital** (capital to acquire new data science tools, R&D, high-performance computing, specialised labour) and **labour** (i.e. the need for specific skill sets like data scientists, new data science specialists, analysts)?

7. I am proposing a Hex-E business intelligence framework that will enable linking of data with inbuilt intelligence; What do you think would be the benefits and risks of such a framework to a business? (especially in the light of the issues you highlighted in previous questions with respect to challenges of extracting insights from data).

Bibliography

Abel, T., Bryan, G., & Norman, M. L. (2002). The formation of the first star in the universe. *Science*, 93-98.

Adamson, C. (2010). *Star schema the complete reference.* NY, New York, United States: McGraw-Hill Osborne Media.

Agarwal, A., Govindu, R., Lodwig, S., & Ngo, F. T. (2016). Solving the Jigsaw Puzzle: An Analytics Framework for context awareness in the Internet of Things. *Cutter IT Journal*, 6-11.

Agarwal, A., Govindu, R., Ngo, F., & Lodwig, S. (2016). Solving the Jigsaw Puzzle: An Analytics Framework for Context Awareness in the Internet of Things. *The Cutter Journal*, 6-11.

Agosto, D. E., & Hughes-Hassell, S. (2005). People, places, and questions: An investigation of the everyday life information-seeking behaviors of urban young adults.". *Library & information science research*, 141-163.

Aksoy, L. (2013). How do you Measure what you Can't Define? The Current State of Loyalty Measurement and Management. *Journal of Service Management, 4*(24), 356-381.

Ali, O. A., Ajmi, Q., & Ali, S. H. (2018). Stay Connected–Internet of Things. *International Journal of Applied Engineering Research*, 4599-4605.

Analytics, B. B. (2007). *Comparing the Total Cost of Ownership of Business Intelligence Solutions.* Birst.

Anderson, M., & Anderson, S. L. (2011). *Machine ethics.* London: Cambridge University Press.

Andrew, S., Salamonson, Y., & Halcomb, E. J. (2008). Integrating mixed methods data analysis using NVivo: An example examining attrition and persistence of nursing students. *International Journal of Multiple Research Approaches*, 36-43.

Arkes, H., & Blumer, C. (1985). The psychology of sunk cost. *Organisational behavior and human decision processes*, 124-140.

Arthur, W. B. (2010). *The nature of technology: What it is and how it evolves. .* London, London, United Kingdom: Penguin Books.

Asimov, I., Silverberg, R., & Timmerman, H. (1978). *The bicentennial man.* London: Panther.

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative research*, 385-405.

Baba, V. V., & HakemZadeh, F. (2012). Toward a Theory of Evidence Based Decision-making. *Management Decision, 5*(50), 832-867.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., & Vigna, S. (2011). Four Degrees of Separation. *Proceedings of the 4th Annual ACM Web Science Conference* (pp. 33-42). ACM.

Bakir, G., Hofmann, T., Schölkopf, B., Smola, A., Tasker, B., & Vishwanathan, S. V. (2007). *Predicting structured data.* London: MIT Press.

Ball, P. (2016). *Patterns in nature: Why the natural world looks the way it does.* Chicago, IL, United States: University of Chicago Press.

Bandura, A. (1986). *Social foundations of thought and action. Englewood Cliffs.* NJ: Englewood Cliffs.

Banerjee, I., & Levy, M. R. (2008). Urban entrepreneurs, ICTs, and emerging theories: A new direction for development communication. *Asian Journal of Communication, 18*(4), 304-317.

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies.* California: Sage.

Beiske, B. (2007). *Research methods: Uses and limitations of questionnaires, interviews, and case studies.* Germany: GRIN Verlag GmbH.

Bernard, R. H. (2012). *Social research methods: Qualitative and quantitative approaches.* CA, United States: SAGE Publications.

Bernard, S. A. (2012). *An introduction to enterprise architecture* (3rd Edition ed.). United States: Authorhouse.

Bhattacharya, P. (2018). Artificial Intelligence in the Boardroom: Enabling 'Machines' to 'Learn'to Make Strategic Business Decisions. *Fifth HCT Information Technology Trends (ITT) - IEEE*, 170-174.

Blank, H., & Nestler, S. (2008). How many hindsight biases are there? *Cognition*, 1408-1440.

Bloomberg, L. D., & Volpe, M. (2018). *Completing your qualitative dissertation: A road map from beginning to end.* Thousand Oaks: Sage.

Blumberg, R., & Atre, S. (2003). The problem with unstructured data. *Dm Review*, 42-49.

Bodor, R., Jackson, B., & Papanikolopoulos, N. (2003). Vision-based human tracking and activity recognition. *11th Mediterranean Conference on Control and Automation.*

Boisot, M. H. (1998). *Knowledge assets: Securing competitive advantage in the information economy.* New York: OUP Oxford.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Thousand Oaks: Sage.

Brønn, P. S., & Vrioni, A. B. (2001). Corporate social responsibility and cause-related marketing: an overview. *International journal of Advertising*, 207-222.

Bradbury, H. (2007). *The sage handbook of action research: Participative inquiry and practice.* London, United Kingdom: SAGE Publications.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 77-101.

Braun, V., & Wilkinson, S. (2003). Liability or asset? Women talk about the vagina. *Psychology of women Section Review*, 28-42.

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. *Handbook of Research Methods in Health Social Sciences*, 843-860.

Browne, G. J., Durrett, J. R., & Wetherbe, J. C. (2004). Consumer reactions toward clicks and bricks: investigating buying behaviour on-line and at stores. *Behaviour & Information Technology, 23*(4), 237-245.

Bruce, P., & Bruce, A. (2017). *ractical statistics for data scientists: 50 essential concepts.* New York: O'Reilly Media.

Brynjolfsson, E. (1994). Information assets, technology and organisation. *Management Science*, 1645–1662.

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organisational transformation and business performance. *Journal of Economic perspectives, 14*(4), 23-48.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* New York, NY, United States: W.W. Norton & Business.

Bucher, T., Gericke, A., & Sigg, S. (2009). Process-centric business intelligence. *Business Process Management Journal, 15*(3), 408-429.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., . . . Trench, M. (2017, June). Artificial Intelligence: The Next Digital Frontier? *Mckinsey Global Institute*.

Burnard, P., Gill, P., & Stewart, K. (2008). Analysing and presenting qualitative data. *British dental journal*, 429.

Cameron, J. I., Rappolt, S., Lewis, M., Lyons, R., Warner, G., & Silver, F. (2007). Development and implementation of the Ontario Stroke System: the use of evidence. *International journal of integrated care,* , 7(3).

Carley, K. M. (2002). Computational organisation science: A new frontier. *Proceedings of the National Academy of Sciences* (pp. 7257-7262). The National Academy of Sciences.

Charmaz, K. (2002). Stories and silences: Disclosures and self in chronic illness. *Qualitative inquiry*, 302-328.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* California: Sage.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from Big Data to big impact. *MIS Quarterly*, 1165-1188.

Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM) People, process and technology. *Business process management journal*, 672-688.

Chengalur-Smith, I. B. (1999). The Impact of Data Quality Information on Decision-making: An Exploratory Study. *IEEE Transactions on Knowledge and Data Engineering* , 853-864.

Chevalier, J. M., & Buckles, D. J. (2013). *Participatory action research: Theory and methods for engaged inquiry.* London, United Kingdom: Taylor & Francis.

Clark, V. P., & Creswell, J. W. (2008). *The mixed methods reader.* California: Sage.

Clarke, S., & Baniassad, E. (2005). *Aspect-oriented analysis and design.* Boston: Addison-Wesley Professional.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 3-21.

Corbin, J., Strauss, A., & Straus, A. L. (2014). *Basics of qualitative research.* California: Sage Publications.

Corker, M. (1999). Differences, conflations and foundations: the limits to'accurate'theoretical representation of disabled people's experience? *Disability & society*, 627-642.

Couper, M. P. (2000). Web-based surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*(4), 464–494.

Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches* (3rd Edition ed.). Thousand Oaks, CA, United States: SAGE Publications.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th Edition ed.). Thousand Oaks, CA, United States: SAGE Publications.

Creswell, J. W. (2014). *A concise introduction to mixed methods research.* CA: Sage.

Creswell, J. W., & Tashakkori, A. (2007). The New Era of Mixed Methods. *Journal of Mixed Methods Research*, 303-308.

Creswell, W. J., & Plano Clark, V. L. (2017). *Designing and conducting mixed method research* (Vol. III). Thpusand Oaks, LA: SAGE.

Davis, J., Edgar, T., Porter, J., Bernaden, J., & Sarli, M. (2012). Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Computers & Chemical Engineering*, 145-156.

De Oliveira, M. F., & Levkowitz, H. (2003). From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 378-394.

Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods.* Routledge.

Des Horts, C.-H. B. (1991). The relationship between organisational innovation and technology: an exploratory research.

Dey, I. (2003). *Qualitative data analysis: A user friendly guide for social scientists.* London: Routledge.

Dial, M., & Storkey, C. (2017). *Futureproof: How to Get Your Business Ready for the Next Disruption.* London: Pearson UK.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical education*, 314-321.

DJS Research. (2018, April). Retrieved from https://info.unit4.com/rs/900-SZD-631/images/U4-ALL-GEN-IG-Unit4-Infographic-XaaS-market-research-final-IG180510INT.jpg

Drucker, P. (2012). *Managing in a time of great change.* London: Routledge.

Dupont, W., & Plummer, W. (1998). Power and sample size calculations for studies involving linear regression. *Controlled clinical trials*, 589-601.

Eatough, V., & Smith, J. A. (2008). Interpretative phenomenological analysis. *The Sage handbook of qualitative research in psychology*, 179.

Einstein, B. (1971).

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). *Qualitative content analysis: A focus on trustworthiness.* Thousand Oaks: SAGE open.

Evans, R., & Gao, J. (2016, July 20). *DeepMind*. Retrieved from DeepMind: https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/

Even, A., Shankaranarayanan, G., & Watts, S. (2006). Enhancing Decision-making with Process Metadata: Theoretical Framework, Research Tool, and Exploratory Examination. *Proceedings of the 39th Hawaii International Conference on System Sciences.* Hawaii: IEEE.

Experian. (2015). Retrieved from Experian: https://www.edq.com/globalassets/uk/papers/global-research-2015_20pp-ext-apr15.pdf

Farmer, D. (2010, May 04). *Dell EMC*. Retrieved from Dell EMC: https://www.emc.com/about/news/press/2010/20100504-01.htm

Fayos, J. (1999). Possible 3D carbon structures as progressive intermediates in graphite to diamond phase transition. *Journal of Solid State Chemistry, 148*(2), 278–285.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27-34.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data.* London: Cambridge University Press.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data. .* Cambridge university press.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 80-92.

Fine, B., & Milonakis, D. (2009). *From economics imperialism to freakonomics: The shifting boundaries between economics and other social sciences.* London: Routledge.

Firica, G. (2017, February 8). *TDWI*. Retrieved from TDWI: https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once—future things. *Organisational Behavior and Human Performance*, 1-16.

Fisher, C. C.-S. (2003). The Impact of Experience and Time on the Use of Data Quality Information in Decision-making. *Information Systems Research*, 170-188.

Florida, R., Mellander, C., & Gulden, T. (2012). Global metropolis: assessing economic activity in urban centers based on nighttime satellite images. *The Professional Geographer, 64*(2), 178-187.

Foddy, W. H. (1994). *Constructing questions for interviews and questionnaires: Theory and practice in social research.* Cambridge, United Kingdom: Cambridge University Press.

Ford, R. C., & Richardson, W. D. (1994). Ethical decision-making: A review of the empirical literature. *Journal of business ethics*(13(3)), 205-221.

Forrester. (2018, 12 20). *Forrester Research*. Retrieved from Forrester: https://go.forrester.com/research/

Freeman, C., & Perez, C. (1988). Structural crises of adjustment: business cycles. *Technical change and economic theory*.

Fulcher, J. (2008). Computational intelligence: an introduction. *Computational intelligence: a compendium*, 3-78.

Galbraith, J. R. (1973). *Designing complex organisations.* Reading, MA, United States: Addison-Wesley Pub. Co.

Galletta, A., & Cross, W. E. (2013). *Mastering the Semi-Structured interview and beyond: From research design to analysis and publication.* New York, NY, United States: New York University Press.

Gartner. (2018). *Spending in Key IT Segments Will Shift to the Cloud.* Sydney: Gartner.

Garton, L., Haythornthwaite, C., & Wellman, B. (1999). *Doing Internet Research: Critical Issues and Methods for Examining the Net.* CA: Sage.

Garvin, D. A. (2013). How Google Sold its Engineers on Management. *Harvard Business Review, 12*(91), 74-82.

Gawande, A. (2009). *The cost conundrum.* New York: The New Yorker.

Gibbs, G. R. (2018). *Analyzing qualitative data* (Vol. 6). Thousand Oaks: Sage.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox.* Berlin, Germany: MIT Press.

Giles, & Giles, J. (2012). *Nimble elephant: Agile delivery of data models using a pattern-based approach.* Technics Publications.

Gladwell, M. (2005). *The power of thinking without thinking.* London: Penguin.

Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge Management: An Organisational Capabilities Perspective. *Journal of Management Information Systems, 1*(18), 185-214.

Greenwood, D. J., & Levin, M. (2006). *Introduction to action research: Social research for social change.* Thousand Oaks, CA, United States: Sage Publications.

Hales, D. (2010). An introduction to triangulation. *UNAIDS Monitoring and Education Division*.

Hall, R. (1993). A framework linking intangible resources and capabiliites to sustainable competitive advantage. *Strategic Management Journal, 14*(8), 607–618.

Halverson, R., & Smith, A. (2009). How new technologies have (and have not) changed teaching and learning in schools. *Journal of Computing in Teacher Education, 26*(2), 49-54.

Hammersley, M. (2007). The issue of quality in qualitative research. *International Journal of Research & Method in Education.*, 287-305.

Hancock, D., & Algozzine, B. (2016). *Doing case study research: A practical guide for beginning researchers.* New York: Teachers College Press.

Hartley, R. V. (1928). Transmission of information. *The Bell System Technical Journal*, 535-563.

Hay, D. C. (2011). *Enterprise model patterns: Describing the world.* United States: Technics Publications.

Henderson, S., & Segal, E. H. (2013). Visualizing qualitative data in evaluation research. *New Directions for Evaluation, 2013*(139), 53-71.

Henderson-Sellers, B., & Unhelkar, B. (2000). *Open modeling with UML.* New York, NY, United States: Addison-Wesley Educational Publishers.

Holloway, I., & Todres, L. (2003). The status of method: flexibility, consistency and coherence. *Qualitative research*, 345-357.

Holzner, S., & Levy, J. R. (2010). *Facebook marketing: Leverage social media to grow your business* (4th Edition ed.). Indianapolis, IN, United States: Que Corporation,U.S.

HortonWorks. (2019, April 03). *The Use Of Artificial Intelligence In Decision-making.* Retrieved from HortonWorks: https://hortonworks.com/blog/three-things-ceos-should-know-about-the-use-of-arti.

Howard, P., Rainie, L., & Jones, S. (2001). Days and nights on the Internet: The impact of a diffusing technology. *American Behavioral Scientist*, 383–404.

Hryniewicz, R. (2018, July 11). *Horton Networks*. Retrieved from Horton Networks: https://hortonworks.com/blog/three-things-ceos-should-know-about-the-use-of-artificial-intelligence-in-decision-making/

Hryniewicz, R. (2018, July 16). *The Use of Artificial Intelligence in Decision-Making*. Retrieved from DZone: https://dzone.com/articles/the-use-of-artificial-intelligence-in-decision-mak

Hughes, R. (2012). Turning Big Data into big benefits. *Cutter IT Journal*.

Husted, B. W., & de Jesus Salazar, J. (2006). Taking Friedman seriously: Maximizing profits and social performance. *Journal of Management studies, 43*(1), 75-91.

Hwang, J., & Lee, K. C. (2013). Exploring potentials of personality matching between users and target systems by using fuzzy cognitive map. *System sciences (HICSS). 46th Hawaii international conference* (pp. 417-424). Hawaii: IEEE.

IBM. (2011). *IBM Big Data*. Retrieved from IBM Big Data: http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Informatics India. (2014, Oct 16). *VC FUNDS TO INVEST OVER $1 BILLION IN IOT (Number of Devises Connected through the Virtual Medium Has Reached 13.69 Billion Devices so Far)."* . Retrieved from India Business Insight,: Informatics India Limited

Jaeger, G. (2006). *Quantum information: An overview.* New York, NY, United States: Springer-Verlag New York.

Jaeger, G. (2009). *Entanglement, information, and the interpretation of quantum mechanics.* Berlin, Germany: Springer-Verlag Berlin and Heidelberg GmbH & Co. K.

Jensen, M. C. (2003). *A theory of the firm: Governance, residual claims, and organisational forms.* Cambridge, MA, United States: Harvard University Press.

Jia, L., Hall, D. J., & Song, J. (2015). The Conceptualization of Data-driven Decision-making Capability. *Conference: the Americas Conference on Information Systems.*

Joe Tucci. (2010, May 04). *Security Week*. Retrieved from Security Week: https://www.securityweek.com/content/emc-digital-universe-data-growth-study-projects-nearly-45-fold-annual-data-growth-2020

Juliusson, Á. E., Karlsson, N., & Gärling, T. (2005). Weighing the past and the future in decision-making. *European Journal of Cognitive Psychology*, 561-575.

Jun, L., Sung, & Siau, K. (2001). A review of data mining techniques. *Industrial Management & Data Systems, 101*(1), 41-46.

Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 215-236.

Kahaner, L. (1996). *Competitive Intelligence: From Black Ops to Boardrooms_how Business Gather, Analyze, and Use Information to Succeed in the Global Markeplace.* New York: Simon & Schuster.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions.

Kaisler, S., Armour, F., & Espinosa, A. (2013). Big data: Issues and challenges moving forward. In System sciences (HICSS). *46th Hawaii international conference* (pp. 995-1004). Hawaii: IEEE.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 59-68.

Kappelman, L. A., & Zachman, J. A. (n.d.). The enterprise and its architecture: Ontology & challenges. *Journal of Computer Information Systems, 53*(4), 87–95.

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. In Contemporary Computing (IC3),. *2013 Sixth International Conference* (pp. 404-409). Noida, India: IEEE.

Kelliher, F. (2011). Interpretivism and the pursuit of research legitimisation: an integrated approach to single case design. *Leading issues in business research methods*, 123-131.

King, N., Horrocks, C., & King, P. W. (2009). *Interviews in qualitative research.* London, United Kingdom: Sage Publications.

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society, 3*(1), 2053951716631130.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, 211.

Klein, B. (2001). User perceptions of data quality: Internet and traditional text sources. *Journal of Computer Information Systems*, 9-15.

Klein, G. A. (2004). *The power of intuition: How to use your gut feelings to make better decisions at work.* Crown Business.

Knowles. (2007). *Handbook of the arts in qualitative research: Perspectives, methodologies, examples, and issues.* Los Angeles, LA, United States: Sage Publications.

Kościelniak, H., & Puto, A. (2015). BIG DATA in decision-making processes of enterprises. *Procedia Computer Science, 65*, 1052–1058.

Kozinets, R. V., Hemetsberger, A., & Schau, H. J. (2008). The wisdom of consumer crowds: Collective innovation in the age of networked marketing. *Journal of Macromarketing, 28*(4), 339-354.

KPMG. (2014). *KPMG Insights*. Retrieved from KPMG Insights: https://images.forbes.com/forbesinsights/StudyPDFs/KPMG_CEO_Survey_STUDY.pdf

Kudyba, S., & Hoptroff, R. (2001). *Data mining and business intelligence: A guide to productivity.* United States: dea Group Pub.

Kumar, N. M., & Mallick, P. K. (2018). Blockchain technology for security issues and challenges in IoT. *Procedia Computer Science*, 1815-1823.

Kumar, V. (2017, March 26). *Big Data Facts*. Retrieved from Analytics Week: https://analyticsweek.com/content/big-data-facts/

Kumar, V., Chattaraman, V., Neghina, C., Skiera, B., Aksoy, L., Buoye, A., & Henseler, J. (2013). Data- Driven Services Marketing in a Connected World. *Journal of Service Management, 3*(24), 330- 352.

Kvale, S. (1994). Ten standard objections to qualitative research interviews. *Journal of phenomenological psychology*, 147-173.

Kvale, S. (2008). *Doing interviews.* Thousand Oaks: Sage.

Kvale, S., & Brinkmann, S. (2004). *Interviews: Learning the craft of qualitative research interviewing* (2nd Edition ed.). Thousand Oaks, CA, United States: Sage Publications.

Lönnqvist, A., & Pirttimäki, V. (2006). The measurement of business intelligence. *Information systems management, 23*(1), 32.

Lan, Y.-C., & Unhelkar, B. (2005). *Global enterprise transitions: Managing the process.* United States: Idea Group Publishing.

Laplace, P. S. (1921). Essai Philosophique sur les Probabilités. *Journal of the Röntgen Society, 17*(69), 184–184.

Lee, R. M., & Esterhuizen, L. (2000). Computer software and qualitative analysis: trends, issues and resources. *International journal of social research methodology,*, 231-243.

Legard, R., Keegan, J., & Ward, K. (2003). In-depth interviews. *Qualitative research practice: A guide for social science students and researchers*, 138-169.

Levitt, S., & Dubner, S. J. (2005). *Freakonomics.* New York: William Morrow.

Lichtenstein, S. &. (1977). Do those who know more also know more about how much they know? *Organisational behavior and human performance*, 159-183.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision-making. *Journal of Behavioral Decision-making, 14*(5), 331–352.

Llieva, J., Baron, S., & Healey, N. M. (2002). Online surveys in marketing research: Pros and cons. *International Journal of Market Research, 44*(3), 361–367.

Lohr, S. (2012, 11). The age of Big Data. *New York Times*, p. 11.

Lueth, K. L. (2018, August 8). *IoT Analytics*. Retrieved from IoT Analytics: https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/

Luo, L., Lan, Y.-C., & Tang, Q. (2012). Corporate incentives to disclose carbon information: Evidence from the CDP Global 500 report. *Journal of International Financial Management & Accounting*(23(2)), 93-120.

Machi, L. A., & McEvoy, B. T. (2016). *The Literature Review, Six Steps to Success.* Thousand Oaks: Corwin, Sage Publishing.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 46-60.

Malhotra, A., Gosain, S., & El Sawy, O. A. (2005). Absorptive Capacity Configurations in Supply Chains: Gearing for Partner-Enabled Market Knowledge Creation. *MIS Quarterly, 1*(29), 145-187.

Malterud, K. (2012). Systematic text condensation: a strategy for qualitative analysis. *Scandinavian journal of public health*, 795-805.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011, 12 31). *Big data: the next frontier for innovation, competition, and productivity*. Retrieved from McKinsey Global Institute: http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_ frontier_for_innovation

March, J. G. (1994). *Primer on decision-making: How decisions happen.* Simon and Schuster.

March, J. G., Schulz, M., & Zhou, X. (2000). *The dynamics of rules: Change in written organisational codes.* Palo Alto: Stanford University Press.

Marchewka, J. T. (2014). *Information technology project management.* NJ: John Wiley & Sons.

Marco, D. (2000). *Building and Managing the Meta Data Repository: A Full Lifecycle Guide.* New York: Willey and Sons, Inc.

Marsh, D., & Hanlon, T. (2007). Seeing what we want to see: Confirmation bias in animal behavior research. *Ethology*, 1089-1098.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American psychologist*(59(4)), 14.

Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think.* United States: Eamon Dolan/Houghton Mifflin Harcourt.

Mays, N., Roberts, E., & Popay, J. (2001). Synthesising research evidence. *Studying the organisation and delivery of health services: Research methods*, 200.

McAfee, A., & Brynjolfsson, E. (2012, Oct). Big Data: The Management Revolution. *Harvard Business Review*, p. 9.

McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future.* WW Norton & Business.

McAfee, A., Brynjolfsson, E., Davenport, T., & Barton, D. (2012). Big data: the management revolution. *Harvard business review 90*, pp. 60-68.

Mcelheran, K., & Brynjolfsson, E. (2016). The rise of data-driven decision-making is real but uneven. *EEE Engineering Management Review*(45(4)), 103-105.

McLeod, J., & Balamoutsou, S. (2001). A method for qualitative narrative analysis of psychotherapy transcripts. *Psychological Test and Assessment Modeling*, 128.

McNeil, B. J., Pauker, S. G., Sox, H., & Tversky, A. (1982). 23On the Elicitation of Preferences for Alternative Therapies. *Preference, Belief, and Similarity*, 583.

Meehan, T., Vermeer, C., & Windsor, C. (2000). Patients' perceptions of seclusion: a qualitative investigation. *Journal of advanced nursing*, 370-377.

Mersereau, R. M., & Farrell, M. D. (2005). On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. *IEEE Geoscience and Remote Sensing Letters, 2*(2), 192-195.

Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods.* Thousand Oaks: Sage.

Merton, R. K., & Kendall, P. (1946). The focused interview. *American journal of Sociology*, 541-557.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher*, 5-11.

Michie, D. (1990). Personal models of rationality. *Journal of statistical Planning and Inference*, 381-399.

Microsoft. (2019, April 04). *Choosing The Best Trendline For Your Data - Access.* Retrieved from Microsoft Support: https://support.office.com/en-us/article/choosing-the-best-trendline-for-your-da

Miles, M., Huberman, M. A., & Saldana, J. (2014). *Qualitative data analysis.* New York: Sage.

Milgram, S. (1967). The Small World Problem. *Psychology Today*.

Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind.* Simon and Schuster.

Mitchell, T. (1999). Machine learning and data mining. *Communication of the ACM, 42*(11), 30-36.

Mjolsness, E., & DeCoste, D. (2001). Machine learning for science: state of the art and future prospects. *Science*, 2051-2055.

Mladenic, D., Lavrač, N., Bohanec, M., & Moyle, S. (2003). *Data mining and decision support: Integration and collaboration.* Springer Science & Business Media.

Moniruzzaman, B. A., & Hossain, S. A. (2013). Nosql database: New era of databases for Big Data analytics-classification, characteristics and comparison. *rXiv preprint arXiv:1307.0191*.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in cognitive sciences, 14*(10), 435-440.

Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research. *Qualitative Health Research*(8(3)), 362-376.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*(1(1)), 48-76.

Morgan, D. L. (2014). *Integrating qualitative and quantitative methods : A pragmatic approach.* Thousand Oaks: SAGE.

Morgan, S. (2015, September 02). *CSO Online*. Retrieved from CSO Online: https://www.csoonline.com/article/2978858/is-poor-software-development-the-biggest-cyber-threat.html

Morse, J. M. (1994). Emerging from the data: The cognitive processes of analysis in qualitative inquiry. *Critical issues in qualitative research methods*, 23-43.

Morton, H., & Aleksander, I. (1990). *An introduction to neural computing.* London: Chapman & Hall.

Moss, L. T., & Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications.* Boston: Addison-Wesley Professional.

Murray, M., Camic, P., Rhodes, J., & Yardley, L. (2003). Qualitative research in psychology: Expanding perspectives in methodology and design. *US: American Psychological Association*, 95-112.

Mustafa, A. (2008). *Case study method: Theory and practice (research and management approaches).* New Delhi, Delhi, India: Atlantic Publishers & Distributors Pvt.

Nair, G., & Lan, Y.-C. (2016). A common thread: Applying hex elementalization in IoT data analytics. *Cutter IT Journal, 29*(4), 31.

Nédellec, C., & Nazarenko, A. (2006). Ontologies and information extraction. rXiv preprint cs/0609137.

Nee, A. C., & Ong, S. K. (2013). *Virtual and augmented reality applications in manufacturing.* Springer Science & Business Media.

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). *Handbook of practical program evaluation.* John Wiley & Sons: Hoboken.

Nokia. (2019, April 03). *Our Strategy - Nokia.com.* Retrieved from Nokia: https://www.nokia.com/sites/default/files/nokia_ar17_en_strategy_web.pdf.

Norvig, P., & Russell, S. J. (2016). *Artificial intelligence: a modern approach.* Malaysia: Pearson Education Limited.

Noura, M. A. (2018). Interoperability in Internet of things: Taxonomies and open challenges. *Mobile Networks and Applications*.

Nyquist, H. (1924). Certain factors affecting telegraph speed. *Bell System Technical Journal, 3*(2), 324–346.

Obama, B. (2018, 10 30). *Geography Education*. Retrieved from Geography Education: https://geographyeducation.org/2012/09/06/president-obama-on-geography-education/

Oliver, R. W. (2000). Real Time Strategy: Sustainable Competitive Advantage? *Journal of Business Strategy*, 7-9.

Pachos, J. K. (2012). *Introduction to topological quantum computation.* Cambridge, CA, United States: Cambridge University Press.

Paka, A. (2016, October 1). *Medium.* Retrieved from https://amitpaka.com/three-stages-of-ai-9d2df56dbd08

Parent, C., & Spaccapietra, S. (2000). *Database integration: the key to data interoperability.* Lausanne: MIT Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods.* Thousand Oaks: Sage.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health services research*, 1189.

Paul, S., Manguerra, H. B., & Slawecki, T. (2018). The State of the Art of Big Data Analytics-A Watershed Management Perspective. *Proceedings of the Water Environment Federation*, (pp. 3209-3217).

Pitta, D. (1998). Marketing one-to-one and its dependence on knowledge discovery in databases. *Journal of consumer marketing,, 15*(5), 468-480.

Polit, D., & Beck, C. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International journal of nursing studies*, 1451-1458.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior, 20*(2), 201–223.

Prelec, D., & Loewenstein, G. (1991). Decision-making over time and under uncertainty: A common approach. *Management Science*(37(7)), 770-786.

Pretzlaff, R. K. (2005). Should age be a deciding factor in ethical decision-making? *Health Care Analysis, 13*(2), 119–128.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision-making. *Big Data, 1*(1), 51-59.

QSR. (2019, 04 30). *QSR International*. Retrieved from qsrinternational.com : www.qsrinternational.com

Ranjan, J. (2008). Business justification with business intelligence. *Vine*, 461-475.

Rao, A. (2018). *http://usblogs.pwc.com/emerging-technology/category/artificial-intelligence/*. Retrieved from PwC Blogs: http://usblogs.pwc.com/emerging-technology/category/artificial-intelligence/

Raufflet, E., & Mills, A. J. (2009). The dark side of business. *The critical need for critical cases*, 1-10.

Rebbapragada,, U., & Protopapas, P. (2009). Finding anomalous periodic time series. *Machine learning*, 281-313.

Reis, L. P., Costa, A. P., & de Souza, F. N. (2016). A survey on computer assisted qualitative data analysis software. *11th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-6). Las Palmas: IEEE.

Re-mapping corporate environmental management paradigms. (2006). *International Studies of Management and Organisation, 32*(6), 54–72.

Renn, O. (1998). Three decades of risk research: accomplishments and new challenges. *Journal of risk research*, 49-71.

Rickards, J. (2014). *The death of money: The coming collapse of the international monetary system.* London, United States: Portfolio/Penguin.

Riessman, C. K. (1993). *Narrative analysis.* Thousand Oaks: Sage.

Rigby, C., Day, M., Forrester, P., & Burnett, J. (2000). Agile supply: Rethinking systems thinking, systems practice. *International Journal of Agile Management Systems, 2*(3), 178–186.

Roca, D., Milito, R., & Nemirovsky, M. (2018). Tackling IoT Ultra Large Scale Systems: fog computing in support of hierarchical emergent behaviors. *Fog computing in the Internet of things* , 33-48.

Ross, J., & Vitale, M. R. (2000). The ERP revolution: surviving vs. thriving. *Information systems frontiers*, 233-241.

Rothman, M. J., Witsil, K., & Nanek, D. (2003). *Washing DC, USA Patent No. U.S. Patent No. 6,505,168.*

Rubin, H., & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data.* Thousand Oaks: Sage.

Rutkin, A. H. (2013, September 12). *Technology Review*. Retrieved from MIT Technology Review: https://www.technologyreview.com/s/519241/report-suggests-nearly-half-of-us-jobs-are-vulnerable-to-computerization/

Ryan, G. W., & Bernard, R. H. (2003). Techniques to identify themes. *Field methods, 15*(1), 85-109.

Saaty, T. L. (2005). *Theory and applications of the analytic network process: decision-making with benefits, opportunities, costs, and risks.* Pittsburgh PA: RWS publications.

Sagi, A., & Friedland, N. (2007). The cost of richness: The effect of the size and diversity of decision sets on post-decision regret. *Journal of personality and social psychology*, 515.

Santofimia, M., Villa, D., Aceña, O., del Toro, X., Trapero, C., Villanueva, F. J., & Lopez, J. C. (2018). Enabling smart behavior through automatic service composition for Internet of Things–based Smart Homes. *International Journal of Distributed Sensor Networks*, 14.

Sarvary, M. (1999). Knowledge management and competition in the consulting industry. *California management review*, 95-107.

Sathi, A. (2012). *Big data analytics: Disruptive technologies for changing the game.* Boise: MC Press.

Sathianwiriyakhun, P., Janyalikit, T., & Ratanamahatana, C. A. (2016). Fast and accurate template averaging for time series classification. *In 2016 8th International Conference on Knowledge and Smart Technology (KST)*, 49-54.

Sayer, A. (1992). *Method in social science: A realist approach.* London: Psychology Press.

Schmidt, R., Möhring, M., Maier, S., Pietsch, J., & Härting, R.-C. (2014). Big data as strategic enabler-insights from central european enterprises. *International Conference on Business Information Systems* (pp. 50-60). Cham: Springer .

Schneider, M. S. (2003). *A beginner's guide to constructing the universe: The mathematical archetypes of nature, art, and science.* New York, NY, United States: HarperCollins Publishers.

Schrage, M. D. (1999). *Serious play: How the world's best businesses simulate to innovate.* Boston, MA, United States: Harvard Business School Press.

Schwab, K. (2017). *The Fourth Industrial Revolution.* Crown Business.

Senge, P. M. (2006). *The Fifth Discipline: The Art & Practive of the Learnings Organisation.* New York.

Severi, S., Sottile, F., Abreau, G., Pastrone, C., Spirito, M., & Berens, F. (2014). M2M technologies: Enablers for a pervasive Internet of Things. *In 2014 European Conference on Networks and Communications (EuCNC)* (pp. 1-5). IEEE.

Shankaranarayanan, G. a. (2004). Managing Metadata in Data Warehouses: Pitfalls and Possibilities. *Communications of the AIS* , 247-274.

Shankaranarayanan, G., & Watson, S. (2003). A Relevant Believable Approach for Data Quality Assessment. *International Conference on Information Qualit.* Boston.

Shannon, C. E. (1949). Communication theory of secrecy Systems. *Bell System Technical Journal, 28*(4), 656–715.

Shannon-Baker, P. (2015). "But I wanted to appear happy": How using arts-informed and mixed methods approaches complicate qualitatively driven research on cultural shock. *International Journal of Qualitative Methods*(14(2)), 34-52.

Shapiro, C., & Varian, H. R. (1998). *Information rules: A strategic guide to the network economy.* Boston, MA, United States: Harvard Business School Press.

Shilakes, C., & Tylman, J. (2008). *Enterprise Information Portals.* USA: Merrill Lynch.

Sincavage, D. (2018). *tenfold*. Retrieved from Sales Blog: tenfold: https://www.tenfold.com/business/artificial-intelligence-business-decisions

Sloane, D. J. (2009). Visualizing qualitative information. *The Qualitative Report, 14*(3), 488-497.

Smith, D. (2008, August). *https://www.pleacher.com/mp/mlessons/stat/sixdegre.pdf*. Retrieved from The Observer: https://www.pleacher.com/mp/mlessons/stat/sixdegre.pdf

Stajkovic, A. D., Locke, E. E., & Blair, E. E. (2006). A first examination of the relationships between primed subconscious goals, assigned conscious goals, and task performance. *Journal of Applied Psychology*, 1172.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought.* New Haven: Yale University Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology*, 672.

Stanton, J. M. (1998). An empirical assessment of data collection using the Internet. *Personnel Psychology*, 709–725.

Stenner, P. (1993). Discoursing jealousy. *Discourse analytic research: Repertoires and readings of texts in action*, 94-132.

Stirati, A. (1994). *The theory of wages in classical economics: a study of Adam Smith, David Ricardo, and their contemporaries.* Edward Elgar Publishin.

Tadejko, P. (2015). Application of Internet of Things in logistics–current challenges. *Ekonomia i Zarządzanie*, 7.

Talanquer, V. (2014). Using qualitative analysis software to facilitate qualitative data analysis. *Tools of chemistry education research*, 83-95.

Tan, F. B. (2008). Global information technologies. *Global information technologies*.

Tanwar, M., Duggal, R., & Khatri, S. K. (2015). Unravelling Unstructured Data: A Wealth of Information in Big Data. *IEEE*.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks: SAGE.

Tashakkori, A., & Teddlie, C. (2010). *SAGE handbook of mixed methods in social & behavioral research.* Thousand Oaks, CA: Sage.

Taylor, S. J., Bogdan, R., & DeVault, M. (015). *Introduction to qualitative research methods: A guidebook and resource.* Hoboken, NJ, United States: John Wiley & Sons.

Techwire. (2016, August 11). *Techwire Asia*. Retrieved from Techwire: A recent study from Aberdeen Research found that the average

Thaler, R. H. (2017). *Misbehaving: The making of behavioral economics.*

Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organisational science. *Organisational Research Methods*, 525-547.

Travers, M. (2001). *Qualitative research through case studies.* Thousand Oaks, CA, United States: SAGE Publications.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. *Progress in social psychology*, 49-72.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 297-323.

Unhelkar, B. (2012). *The art of agile practice: A composite approach for projects and organisations.* London, United Kindgom: Auerbach Publishers.

Unhelkar, B. (2013). User Experience Analysis Framework: From Usability to Social Media Networks. *Cutter IT Journal, 13*(3), 13.

Unhelkar, B. (2016). *Green IT strategies and applications: using environmental intelligence.* Boca Raton: CRC Press.

Unhelkar, B. (2017). *Big Data Strategies for Agile Business: Framework Practices, and Transformation Roadmap.* Boca Raton: CRC Press (Auerbach Publications).

Unhelkar, B., Ghanbary, A., & Younessi, H. (2009). *Collaborative business process engineering and global organisations: Frameworks for service integration.* Philadelphia, PA, United States: Business Science Reference.

Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice*, 100-110.

Vandermerwe, S. &. (1988). Servitization of business: adding value by adding services. *European management journal*, 314-324.

Verdinelli, S., & Scagnoli, N. I. (2013). Data display in qualitative research. *International Journal of Qualitative Methods*, 359-381.

Walton, J. B. (2014, Jan 31). *Doctoral dissertation.* Retrieved from University of Cincinnati: http://rave.ohiolink.edu/etdc/view?acc_num=ucin1406810984

Wang, J., & McElheran, K. (2017). Economies Before Scale: Survival and Performance of Young Plants in the Age of Cloud Computing. *Rotman School of Management Working Paper*.

Wang, R. K. (1993). Data Quality Requirements Analysis and Modeling. *In the proceedings of the 9th International Conference on Data Engineering* (pp. 670-677). IEEE Computer Society Press.

Watson, H., & Wixom, B. (2007). The current state of business intelligence. *Computer*, 96-99.

Watt, J. H. (1999). Internet systems for evaluation research. In G. Gay, & T. L. Bennington, *Information Technologies in Evaluation: Social, Moral, Epistemological, and Practical Implications* (pp. 23–44). San Francisco: Jossey-Bass.

Watts, S., Shankaranarayanan, G., & Even, A. (2009). Data quality assessment in context: A cognitive perspective. *Decision Support Systems*, 202-211.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Weissenrieder, F. (2997). Value Based Management: Economic Value Added or Cash Value Added. *FWC AB Study*, 3.

Wellman, B. (1997). An electronic group is virtually a social network. In S. Kiesler, *Culture of the Internet.* Mahwah, NJ: Lawrence Erlbaum.

Welsh, E. (2002). Dealing with data: Using NVivo in the qualitative data analysis process. *In Forum Qualitative Sozialforschung/Forum: Qualitative Social Research.*

Wiewiórowski, W. (2015, November 27). *EDPS*. Retrieved from EUROPEAN DATA PROTECTION SUPERVISOR: https://edps.europa.eu/press-publications/press-news/videos/big-data-means-big-responsibility_en

Willing, C. (1999). *Applied discourse analysis: Social and psychological interventions.* London: Open University Press.

Yeoh, W., Koronios, A., & Gao, J. (2008). Managing the implementation of business intelligence systems: a critical success factors framework. *International Journal of Enterprise Information Systems (IJEIS)*, 79-94.

Yin, J., & Wang, J. (2015). Optimise parallel data access in Big Data processing. *15th IEEE/ACM International Symposium on Cluster*.

Yin, J., Zhang, J., & Feng, W. (2014). A novel scalable data access framework for parallel BLAST. *Parallel Computing, 40*(10), 697–709.

Yin, R. K. (2011). *Applications of case study research.* Thousand Oaks, CA, United States: SAGE Publications.

Yin, R. K. (2013). *Case study research: Design and methods (applied social research methods)* (4th Edition ed.). Thousand Oaksq, CA, United States: Sage Publications.

Zachman, J. A. (1992). *Enterprise architecture planning.* Boston, MA, United States: QED Publishing.

Zachman, J. A., Spewak, S., & Hill, S. (1992). *Enterprise architecture planning.* Boston, MA, United States: QED Publishing Group.

Zeleny, M., & Cochrane, J. L. (1973). *Multiple criteria decision-making.* Columbia: University of South Carolina Press.

Zhang, Y., & Wildemuth, B. M. (2009). *Qualitative analysis of content. Applications of Social Research Questions in Information and Library.*

Zhang, Y., Cao, T., Li, S., Tian, X., Yuan, L., Jia, H., & Vasilakos, A. V. (2016). Parallel processing systems for Big Data: A survey. *Proceedings of the IEEE*.