NOVA IMS Information Management School

MGI

Master's Degree Program in

Information Management

Data Migration in the Era of Digital Transformation

Migrating to Cloud-Based Solutions

Mariana Domingues Alves

Internship Report

presented as partial requirement for obtaining the Master's Degree Program in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

DATA MIGRATION IN THE ERA OF DIGITAL TRANSFORMATION

Migrating to Cloud-Based Solutions

By

Mariana Domingues Alves

Internship Report presented as a partial requirement for obtaining the Master's Degree in Information Management, with a specialisation in Knowledge Management and Business Intelligence.

Supervisor: Mijail Naranjo-Zolotov

July 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, 15th of July 2023

ABSTRACT

This internship report provides an overview of data migration, a crucial process in modern organisations' digital transformation efforts. It explores the importance of data migration for optimising data management strategies and gaining a competitive edge. The report discusses the steps of data migration, including planning, data extraction, transformation, loading, validation and reconciliation. It addresses common challenges such as data mapping complexities and system compatibility issues, and suggests strategies to mitigate risks. Additionally, the report highlights the importance of data security and privacy considerations during data migration, including compliance with regulations and the adoption of encryption and access controls. It contains best practices and recommendations for successful data migration projects, stressing collaboration, monitoring, and evaluation for accurate and stable data migration outcomes.

KEYWORDS

Data Migration; Digital Transformation; Cloud Software; On-Premises Software;

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ETL – Extract, Transform, Load

DBMS – Database Management Systems

SAP – Systems Analysis and Program Development

GCP – Google Cloud Platform

AWS – Amazon Web Services

VPN – Virtual Private Network

AI – Artificial Intelligence

ADF – Azure Data Factory

HTML – Hypertext Markup Language

PDF – Portable Document Format

SQL – Structured Query Language

1. INTRODUCTION

1.1. ACADEMIC CONTEXT

This internship report has been done in the context of the second-year Master's program in Information Management at the Information Management School of the University Nova of Lisbon, with the goal of summarising my work as a Tech Analyst as part of a consulting team for Deloitte Portugal's industries of Products, Services, Utilities, and Resources. Beginning in September 2022 and ending in February 2023, my internship lasted 6 months.

1.2. COMPANY OVERVIEW

Deloitte is one of the largest professional services companies in the world. It is part of the “Big Four” accounting firms, alongside PwC, KPMG, and EY. The company provides a large diversity of services including audit, consulting, tax, risk management and financial advisory. Aside from having a strong reputation in the market, Deloitte is well-known for its expertise in industries like healthcare, technology, and financial services, and is praised for its dedication to sustainability and corporate responsibility.

Deloitte Portugal has been operating in the market for over 60 years and is a leading professional services firm, being considered one of the most important consulting and audit companies in Portugal, with a team of over 1000 professionals. Its wide range of clients includes large multinational companies, government entities, as well as local businesses. The company is actively involved in promoting business ethics, corporate governance and environmental sustainability, and supports various community initiatives and charitable causes.

1.3. BACKGROUND AND PROBLEM IDENTIFICATION

Data migration can be a time-consuming and complex process. Several challenges and considerations must be taken into account by organisations in order to plan and execute a data migration project. The migrated data should be accurate, complete and consistent. The storage of the data and the capacity for handling it must be ensured by the organisation.

With the required timeframe, the migration process should be completed. According to Azeroual & Jha (2021), data migration holds a significant importance in the process of digitalisation within companies. With the implementation of new software systems, the transfer of pre-existing content and information from diverse data sources becomes a necessity. Consequently, the objective of quality assurance is to identify and rectify errors within the data, data migration programs, and the underlying infrastructure.

According to Sarmah (2018), data loss is a common risk encountered during the data migration process, and its frequency is a significant concern. To address this issue, migration strategies should be devised to minimise the challenges associated with data loss. The most effective mitigation policy involves employing a comprehensive data backup of the source system before initiating the migration to the target system.

Hussein (2021) states that prior to the migration, it is crucial to perform a thorough backup of all data, specifically focusing on the files intended for migration. This precautionary measure allows for the

ability to address any potential problems that may occur during the migration process, such as encountering corrupt, incomplete, or lost files. With a comprehensive backup in place, the data can be restored to its original state, effectively rectifying any errors that may arise.

1.4. INTERNSHIP GOALS

Companies frequently begin data migrations as they transition from locally installed systems to cloud-based hosts, in order to improve or restructure their business and enhance their performance and competitiveness. When a new system or location for data is introduced, it is important for them to seek out the most suitable methods to maintain their data, which leads to the process of data migration. Older systems are constantly being upgraded or replaced by new applications that will use the same dataset.

In this internship, the project I was a part of had the goal of migrating an existing dataset from SAS Enterprise Guide to Microsoft Azure Data Factory, as the client concluded that it was advantageous to choose cloud over on-premise software, and therefore, Microsoft Azure over SAS, after comparing what each product has to offer.

For this project, several questions were taken into consideration, such as the tools to be used, the frequency of the status reports to keep the client informed and updated throughout the progress and defining who has the authority to edit and remove data from the source system. During the process, each function is recreated, and the expected output will be tested and compared against the old system to ensure that the results are the same and therefore the migration plan has been successfully performed.

2. LITERATURE REVIEW

2.1. DATA MIGRATION

Data migration is the process of moving data from one location or system to another. It is a typical activity in lots of businesses and frequently required when upgrading or combining systems or migrating to the cloud. Additionally, data migration can enhance data security, compliance, and quality. The body of knowledge on data migration has expanded recently, with a focus on the methods, constraints, and optimum practices for completing it efficiently.

Companies may need to improve their current systems or integrate various systems in order to increase efficiency and reduce spending. Data migration is frequently necessary in these circumstances to transfer data from the old systems to the new ones. Organisations may decide to move their data to the cloud to benefit from the benefits of cloud software, such as improved performance, scalability, and cost savings.

Azeroual & Jha (2021) state that data migration entails more than simply transferring data from an outdated data structure or database to a new one. It also encompasses the crucial tasks of error correction and enhancing the overall data quality and functionality.

According to Mathur vd. (2016), data migration is a difficult, multi-step process that needs careful planning and execution. Data extraction, data transformation, data loading, and data validation are frequently included in the process. Data transformation requires changing the data to conform to the structure of the target system, while data extraction requires locating and retrieving the data that must be migrated. Data validation confirms that the data has been successfully migrated and is accurate, while data loading entails moving the data to the target system.

2.2. STEPS IN DATA MIGRATION

1. Planning: In this step, the target system, the data that needs to be migrated, and the migration schedule are all identified. It also involves deciding any system dependencies that must be considered and assessing the impact of migration on the organisation's operations.

2. Data Extraction: The data that requires to be migrated from the source system must be identified and extracted at this stage. Various methods, including export/import tools, data replication, and data warehousing, can be used to extract the data.

3. Data Transformation: This step includes transforming the data to conform to the target system's organisational structure. Data reformatting, data cleansing, or data mapping may be required.

4. Data Loading: This stage entails migrating the data to the correct system. Data can be imported in several ways, including incremental loading, bulk loading, and real-time data replication.[5]

5. Data Validation: This process makes sure the data was accurately moved and hasn't been interfered with. Data comparison between the source and target systems, data quality checks, or testing the target system's data may all be necessary in this situation.

6. Data reconciliation: Any discrepancies between the source and target systems are identified and resolved.[6]

In figure 1, a flow chart of steps in data migration is represented.

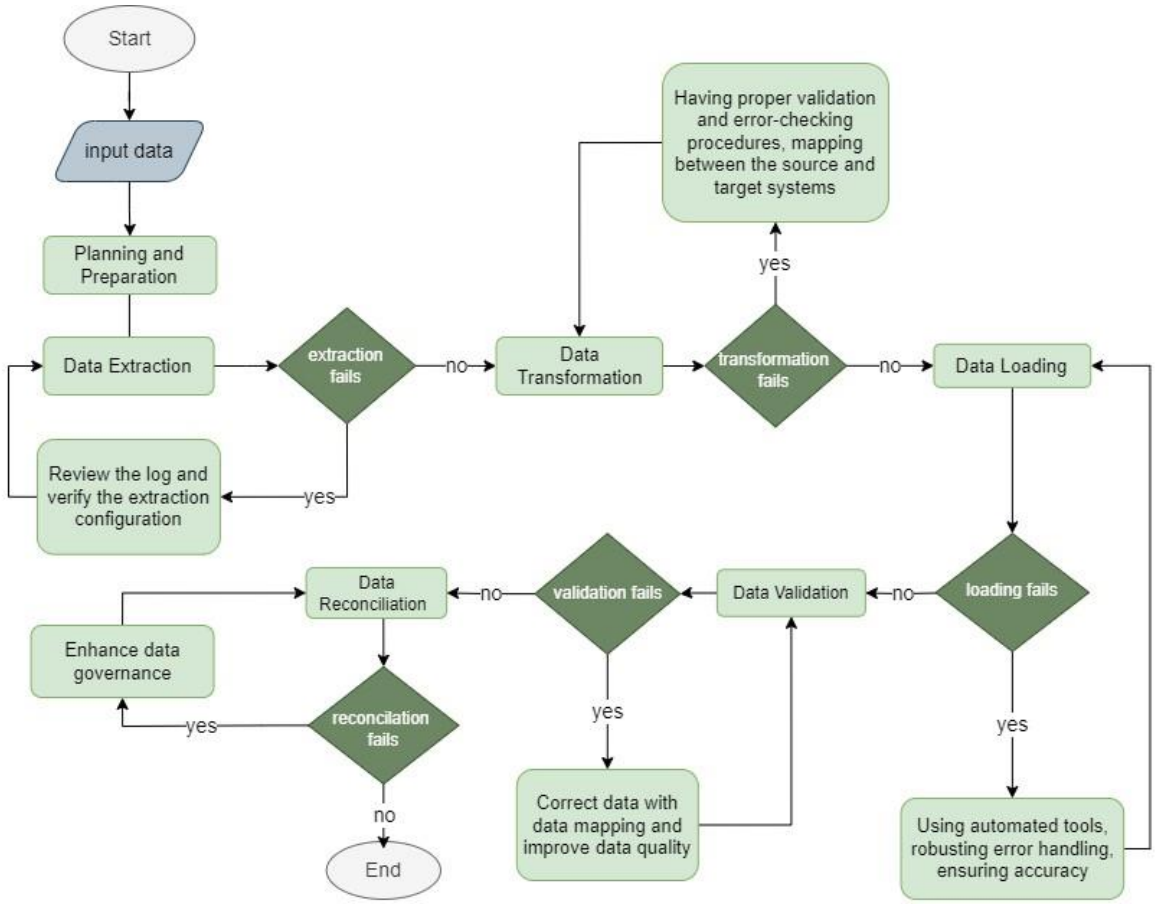

Figure 1 – Illustrative Steps in Data Migration

2.3. METHODS OF DATA MIGRATION

ETL: Data is extracted from the source system, changed to match the structure of the destination system, and then loaded into the target system using the ETL technique of data migration. ETL is a popular technique for data migration because it enables complex data transformation and validation before loading the data into the target system.[7]

Data replication: Data replication is a method of migrating data that includes transferring data in real-time or almost real-time from the source system to the target system. Without any downtime, this method can be used to migrate data in a live environment, but it necessitates a lot of resources and is challenging to implement. [8]

Data Integration: Data can be transferred to the cloud using cloud-based migration methods, including employing cloud-based data migration tools or cloud-based data integration services like Azure Data Factory.[9]

Database Links: Database links are a method for data migration that involves creating a link between the source and target systems, allowing the data to be transferred directly between the two systems.[10]

Data Synchronization: The process of ensuring that data between two or more systems is consistent. This method involves copying changes made in one system to other systems, resulting in all systems having the same data.[11]

“Data migration is a regular activity of the IT department in many organisations. It often causes major issues due to various reasons like the staff, downtime of the environment, the poor performance of the application and these will, in turn, affect the budgets. To prevent these types of issues, organisations need a reliable and consistent methodology that allows the organisations to plan, design, migrate and validate the migration. Further, they need to evaluate the need for any migration software/tool that will support their specific migration requirements, including operating systems, storage platforms, and performance. To keep in check all the points, an organisation needs robust Planning, Designing, Assessment and proper execution of the Project and its variables.” (Hussein, 2021)

The chosen method of data migration can impact the success of a data migration project. It should be selected based on the organisation’s specific needs, purposes and constraints. By carefully selecting the appropriate method and executing it properly, organisations can ensure that their data is moved with minimal disruption and that the new system is populated with accurate, complete, and usable data.

2.4. Cloud Solutions vs On-Premises

Cloud software is hosted and managed by a third-party provider on remote servers located in secure data centers and accessed through a web browser or app. These servers are specifically designed to provide high levels of reliability, scalability, and security, ensuring optimal performance for users accessing the software through a web browser or dedicated application. On-premise software, on the other hand, is installed and runs locally on a user's own physical server or computer.[12]

The decision of whether to use cloud or on-premises software is a critical one that organisations must make. While both approaches have their advantages and disadvantages, the choice ultimately depends on an organisation's specific needs and circumstances. From a managerial perspective, it's important to carefully evaluate the potential pros and cons of each approach and determine which one aligns best with the organisation's goals and objectives. In this answer, we'll explore some of the key advantages and disadvantages of cloud and on-premises software, providing a framework for managers to make an informed decision.

2.4.1. Cloud software from a managerial perspective

Cloud software provides businesses with a flexible, scalable, and cost-effective solution that is accessible from anywhere and requires minimal maintenance. In recent years, cloud software has become a popular choice for businesses due to its numerous advantages. However, there are some

potential disadvantages to choosing cloud software. In table 1, the advantages and disadvantages of choosing cloud software are illustrated.[13]

Table 1 – Advantages and Disadvantages of choosing Cloud Software

Advantages	Disadvantages
Flexibility and scalability: Cloud software allows for easy adjustment of resources based on demand, ensuring optimal performance and cost efficiency.	Security and privacy concerns: Organisations need to trust cloud providers to handle sensitive data securely. Thorough evaluation, implementing security safeguards, and selecting certified providers can help mitigate risks.
Accessibility: Cloud software enables access from anywhere with an internet connection, facilitating collaboration among remote teams.	Integration challenges: Integrating cloud services with existing on-premises infrastructure may require careful planning and potential investments in middleware or tools.
Cost-effectiveness: The subscription-based pricing model of cloud software provides cost-effectiveness, as users only pay for the resources they use without a large upfront investment.	Dependence on service providers: Organisations need to consider their reliance on cloud providers for critical services and negotiate appropriate service level agreements.
Maintenance and upgrades: Cloud providers handle maintenance, upgrades, and security, saving users time and resources.	Limited control: Organisations may have limited control over the cloud infrastructure and services. Evaluating customization options and considering hybrid cloud solutions can address this concern.
Disaster recovery: Cloud software offers robust disaster recovery capabilities by storing data in multiple locations, minimizing downtime and enhancing business continuity.	Compliance: Organisations need to evaluate regulatory requirements and confirm the compliance of their cloud providers to ensure adherence to applicable legislation.

2.4.2. On-Premises Solutions from a managerial perspective

Despite the growing popularity of cloud software, on-premise software still has its own set of advantages that make it a good choice for businesses that prioritise data privacy and control, and have specific needs that may not be met by cloud software. However, there may be a few disadvantages of choosing on-premises solutions. Table 2 outlines the advantages and disadvantages of choosing on-premises solutions.

Table 2 – Advantages and Disadvantages of choosing On-Premises solutions

Advantages	Disadvantages
Data control: Users have full control over their data, ensuring compliance with data privacy regulations and keeping sensitive information within their infrastructure.	High capital costs: Requires a substantial upfront investment in hardware, software, and infrastructure. Organisations may need to explore alternative financing options and carefully evaluate long-term costs.
Performance: On-premises software typically offers better performance due to proximity, reducing latency and improving response times.	Maintenance costs: Organisations are responsible for maintaining and upgrading hardware and software, which can be costly and time-consuming. Careful evaluation and outsourcing options can help mitigate these costs.
Customisation: Users can tailor the software to meet specific needs and requirements, providing greater flexibility and adaptability.	Limited scalability: On-premises infrastructure may pose challenges when business needs change, requiring additional investment in infrastructure or consideration of hybrid cloud solutions.
Offline access: On-premises software allows for uninterrupted productivity in areas with limited or unreliable internet connectivity.	Accessibility: On-premises infrastructure may be less accessible than cloud services, impacting remote work and collaboration. Remote access technologies like VPNs and collaboration tools can help address this.

2.5. Data Migration Strategies

1. Big Bang Data Migration

The big bang data migration is in a single action, therefore it is fast and simple. During the migration, all the systems will be unavailable to users due to its deployment. This strategy is better suited for small businesses with lesser amounts of data. According to Nyeint & Soe (2019), the big bang migration strategy involves immediately shutting down all applications and databases, dedicating full efforts to data migration. This approach may appear favourable as it ensures the shortest possible migration duration and minimises the risk of unforeseen issues. However, it can potentially disrupt the organisation's workflow, particularly for companies reliant on real-time data. Being disconnected from data can present significant challenges. To mitigate the negative impact of big bang migrations, many companies choose to initiate the process after work hours or during holidays. This scheduling approach aims to minimise problems caused by temporarily cutting off data access, thereby reducing the potential disruptions.

2. Trickle Data Migration

The trickle data migration strategy involves gradually migrating data to the new system over a period of time, rather than all at once. This strategy is often used when there is a large amount of data to be migrated, or when it is important to minimise the risk of data loss or downtime during the migration process.

Trickle also allows for a more gradual transition, which can minimise disruptions to business operations and minimise the risks of data loss or corruption. Trickle migration is performed during the normal work of all involved systems. The idea behind trickle data migration is not to shut the whole system at once, but operate only on its chosen areas so that all other could be accessible at the moment of migration. In this approach, employees keep the access to data even if the migration is ongoing continuously without interruptions (Nyeint & Soe, 2019).

A comparison of two data migration strategies can be seen in table 3.

Table 3 – Comparison of the two data migration strategies

	Big Bang	Trickle
Downtime	high	low
Cost	low	high
Complexity	low	high
Accessibility	low	high

Both strategies have their own strengths and weaknesses. In Big Bang, the amount of time to migrate all data to a new system at once is short; however, the risk is higher because the process could result in significant downtime or data loss. The Big Bang method is often used for smaller-scale migrations, it has a clean cutover from the old system to the new system, but it can result in downtime. The Big Bang Method is unsuitable for big data migration. On the other hand, the Trickle method takes longer to migrate data because the data is moved in smaller, more manageable chunks. It has no system downtime, and it can handle large amounts of data. But the time required to migrate is longer, and the process is more complicated. Constant synchronization between systems is required. The choice between two methods will depend on the specific requirements of the project and the size of the data.

2.6. Types of Data Migration

There are eight types of data migration.

1. **Storage Migration:** Data transfer from one storage system to another. Storage migration is caused by the replacement of old storage equipment with newer, more advanced storage technology equipment.[15]
2. **Application Migration:** The process of moving an application and its associated data from one environment to another. It happens whenever a company updates its application software or switches vendors for its application providers.[16]
3. **Database Migration:** The transfer of data from one database to another, typically from an older database to a newer one. It is managed by Database Management Systems(DBMS). [17]

4. **Cloud migration:** The movement of data, applications, and workloads from on-premises infrastructure to the cloud.[18]
5. **Platform migration:** The process of moving an application or workload from one operating system or platform to another.
6. **Server consolidation migration:** The process of moving multiple workloads or applications to a smaller number of servers.
7. **Virtualization migration:** The process of moving an application or workload from a physical environment to a virtual environment.
8. **Disaster recovery migration:** The movement of data and applications to a secondary location in order to ensure business continuity in the event of a disaster.[19]

2.6.1. Differences between Data Migration Types

The main difference between platform migration and server consolidation migration is the scope of the migration. Platform migration involves moving an application or workload from one operating system or platform to another, whereas server consolidation migration involves moving multiple workloads or applications to a smaller number of servers. This can involve reducing the number of physical servers or virtual servers or consolidating workloads onto a cloud-based platform.

The main difference between storage migration and cloud migration is the type of infrastructure involved. Storage migration involves moving data from one storage system to another, typically within a company's own on-premises infrastructure. On the other hand, cloud migration involves moving data, applications, and workloads from on-premise infrastructure to a cloud-based infrastructure, such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform (GCP). Cloud migration can offer benefits such as increased scalability, availability, and reduced costs compared to traditional on-premises framework. However, it can also present challenges such as security concerns, data privacy regulations, and potential compatibility issues with existing systems and applications. In contrast, storage migration can be less complex, but may not provide the same level of benefits as cloud migration, especially in terms of scalability and availability. [20]

Data migration encompasses various types of migrations, including storage migration and database migration. While storage migration involves transferring data from one storage system to another, database migration focuses on moving data from one database to another, often with a different schema. Cloud migration specifically targets the movement of data, applications, and workloads to the cloud, while platform migration centers around transferring an application or workload from one operating system or platform to another. Application migration focuses on moving an application and its associated data from one environment to another, while platform migration is concerned with moving the application or workload to a different operating system or platform. Server consolidation migration involves reducing the number of servers by relocating multiple workloads to fewer servers, whereas virtualization migration entails moving a workload from a physical to a virtual environment. Disaster recovery migration aims to ensure business continuity by moving data and applications to a backup location in the event of a disaster, while cloud migration focuses on the movement of data, applications, and workloads to the cloud.

2.7. RISK ASSESSMENT

Risks of data migration are illustrated in table 4 below.

Table 4 – Risks of Data Migration

Failure event	Probability	Plan of action
Data Loss	Depends on the size and complexity of the data being migrated	Regular back-up of data, testing of migration methods and procedures.
Data Corruption	Depends on hardware failure, software bugs, or network issues.	Implementing best practices such as error handling procedures, verifying the accuracy of data during and after the migration.
Data Integrity	Depends on the size and complexity of the data, method of the migration and the quality of data management processes.	Developing a comprehensive data migration plan, implementing appropriate data validation and testing procedures and documenting the migration process.

As shown in table 4, data migration can present a variety of risks. There is a risk of losing data during the migration process if the data is not properly transferred from the source to the target system. The probability of data loss during migration is determined by the size and complexity of the data being migrated. To prevent data loss, organisations should take steps such as regularly checking and backing up the data so that it can be restored if necessary, as well as testing the migration methods, approaches, and procedures.

Various factors, such as software glitches or hardware failures, can lead to data corruption. To minimise data corruption, actions such as validating the data being migrated to guarantee that it is consistent and free of errors before and after migration and implementing the best practices should be considered by the organisations.

Data integrity is concerned with the accuracy and consistency of the data. The issues that can arise from data integrity are due to a variety of reasons including system failures, a lack of data backups, data mapping errors, data conversion errors, and the size and complexity of the data being migrated. To mitigate these issues, organisations should implement appropriate data validation and testing procedures and document the migration process.[21]

These risks can have serious consequences for an organisation's operations, such as downtime, revenue loss, and reputational damage. The likelihood of data migration issues can be reduced and effectively managed by implementing the actions.

2.8. Functional and Performance Requirements

The migrated data must be accurate and complete, with no missing or corrupted data, to meet functional requirements. It must be consistent across all systems, with no duplicate or contradictory information. Unauthorised access and data breaches must be avoided. The migrated data must be saved and retained in accordance with the data retention policies of the organisation. Authorised users must be able to access the migrated data in a timely and efficient manner. In terms of performance, the migration process must be completed on time and with as little downtime as possible. The migration process must be scalable in order to handle large amounts of data. The migration process must be consistent and reliable, with few mistakes or failures. Data that has been migrated must be protected from loss, corruption, or degradation. The migration process must be optimised to reduce resource utilization and costs.

2.9. Recent Studies about Data Migration Techniques and Technology

The study by Ansar, Ashraf & Fatima (2018) titled "Data Migration in Cloud: A Systematic Review" investigates the advantages and limitations of cloud migration. Cloud migration offers numerous benefits such as cost savings in data storage, enhanced collaboration, flexible scalability, reliable backup facilities, pay-per-use billing, and superior disaster recovery. However, there are also limitations associated with slow internet connections during data migration, which can result in business downtime. Additionally, issues may arise due to regular downloading, potentially impacting the smooth execution of the migration process. Data security can be compromised, as certain cloud applications may not support data conversion or transfer into the system. Consequently, organizations may be compelled to transition from one cloud accounting market to another, a process that can be time-consuming and expensive. Platform dependency may also introduce security concerns when transferring critical data, potentially exposing privacy and security vulnerabilities.

The study by Rai, Mehfuz & Sahoo (2013) titled "Efficient Migration of Application to Clouds: Analysis and Comparison" explores the process of migrating applications to the cloud, highlighting the benefits and considerations involved. Migrating to the cloud offers numerous advantages, but it necessitates thoughtful planning and assessment. Organizations must carefully select suitable applications and infrastructure for migration, taking into account both business and technical factors. Not all applications are suitable for the cloud, so a strategic phased approach is recommended, allowing organizations to navigate potential challenges while moving applications to the cloud. This ensures that the migration costs do not outweigh the benefits. It is crucial for different types of organizations, including large corporations and governments, to consider factors specific to their individual needs before embarking on the cloud migration journey. Despite the need for customization, the cloud holds promise for all organizations. The study further concludes that application performance does not decline when migrating to the cloud, underscoring the potential advantages of leveraging this technology.

The study by Hussein (2021) titled "Data Migration Need, Strategy, Challenges, Methodology, Categories, Risks, Uses with Cloud Computing, and Improvements in Its Using with Cloud Using Suggested Proposed Model (DMig 1)" examines the importance of a well-defined data migration strategy. This strategy should encompass various aspects, including dealing with legacy data, mapping data from the old system, addressing the challenges of identifying source data, adapting to

continuously changing targets, meeting data quality requirements, employing suitable project methodologies, and developing expertise in migration processes.

According to a study by Nyeint & Soe (2019), data migration holds significant importance for data-dependent companies, enabling them to update their systems and adapt to evolving business conditions. While data migration may initially seem challenging, it is an essential responsibility for effective management in today's world. Ensuring high data quality during the migration process is paramount. However, the complexity of data migration becomes apparent as one delves deeper into the process, particularly in terms of minimizing its impact on the company. The choice of migration strategy plays a crucial role in mitigating the effects of data migration. Two primary strategies, namely Big Bang Migration and Trickle Migration, offer distinct approaches to the migration process, each employing different methods and approaches to achieve successful data migration.

3. METHODOLOGY

3.1. APPROACHES AND METHODOLOGIES

There are several approaches and methodologies for data migration, including manually transferring data from one system to another. This approach is typically used when the volume of data is small or when the data needs to be manipulated or transformed before being transferred. In cases where data migration involves complex transformations or requires extensive data cleaning, manual transfer allows for greater control and customization during the migration process.[24]

Another approach that can be used is to use automated tools, which are software programs designed to automate the data migration process. These tools can handle large volumes of data large volumes of data, which ensure data accuracy and consistency while reducing the risk of data loss or corruption. Examples of common automated data migration technologies include data migration software, ETL (extract, transform, load) tools, and database migration tools.

In this project, since the data required to be transformed before being migrated, it was decided that we would have to manually transfer it, with the guidance of the Agile methodology framework. Agile methodology is an iterative approach to project management that emphasises flexibility, collaboration, and customer satisfaction. It is a mindset and a set of principles that prioritise delivering value to the customer through continuous improvement and feedback.

3.1.1. SCRUM Methodology

The agile project management framework Scrum is popular in the software industry as well as other sectors. It is a cooperative method of project management that prioritises cooperation, openness, and incremental development. Since Scrum promotes iterative and adaptive planning, it allows teams to quickly respond to changes and deliver high-quality products in shorter timeframes.[25]

3.1.1.1. SCRUM Roles

Scrum is based on a set of roles, events, and artifacts that help teams to work together effectively. The roles in Scrum include the Product Owner, Scrum Master, and Development Team. In this project, the Product Owner was responsible for defining the product backlog, which is a prioritised list of features or tasks that need to be completed. The Scrum Master was responsible for facilitating the Scrum process and ensuring that the team was following the framework's practices. The Development Team, of which I was a member, was responsible for delivering the product increment during each sprint.

3.1.1.2. SCRUM Events

Scrum events include Sprint Planning, Daily Scrum, Sprint Review, and Sprint Retrospective. These events are designed to provide structure and transparency to the project, help the team stay on track, and identify areas for improvement. By regularly engaging in these Scrum events, teams can foster collaboration, maintain a shared understanding of project goals, and continuously enhance their work processes.[26]

Sprint Planning: The sprint planning meeting is a collaborative meeting between the Product Owner and the Development Team. During this meeting, the team reviews the product backlog and selects

the items they will work on during the sprint. The team then creates a sprint goal and a sprint backlog, which is a list of tasks that the team plans to complete during the sprint.

Daily Scrum: The daily Scrum is a 15-minute stand-up meeting held every day during the sprint. During the daily Scrum, each team member reports on their progress since the last meeting, discusses any impediments or roadblocks they are facing, and plans their work for the day.

Sprint Review: The sprint review is a meeting held at the end of the sprint to review the work completed during the sprint. During the sprint review, the development team presents the product increment they have completed during the sprint to the Product Owner and other stakeholders. The team and stakeholders then collaborate to review the product increment and provide feedback.

Sprint Retrospective: The sprint retrospective is a meeting held at the end of the sprint to reflect on the sprint and identify areas for improvement. During the sprint retrospective, the team discusses what went well during the sprint, what didn't go well, and identifies actionable items to improve their process for the next sprint.

3.1.1.3. SCRUM Artifacts

In Scrum, artifacts are tangible items created during the project that provide transparency and information about the project's progress. These artifacts provide transparency, alignment, and focus for the Scrum team and stakeholders. They are used throughout the project to track progress, make informed decisions, and adapt to changes in requirements or priorities. In figure 2, the SCRUM process is shown.[27]

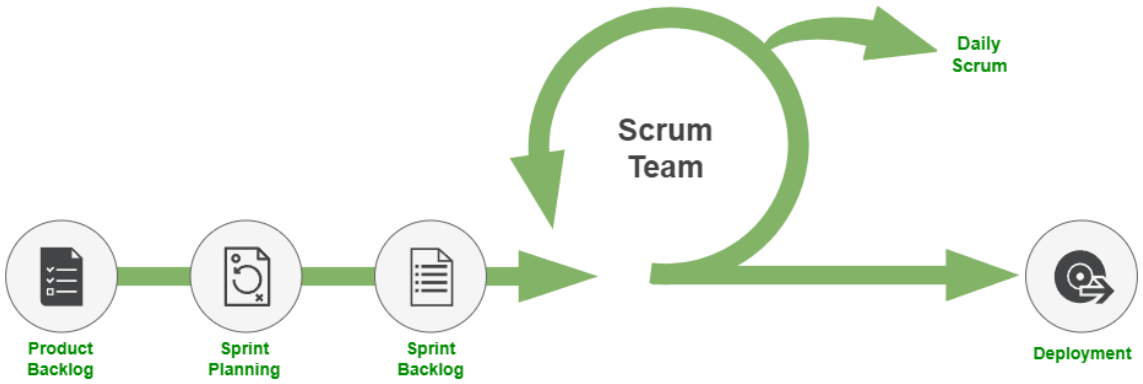

Figure 2 – SCRUM Artifacts

Product Backlog: A prioritised list of features, requirements, enhancements, and bug fixes that need to be completed in the product. The Product Owner is responsible for maintaining the product backlog and updating it regularly based on stakeholder feedback and changing priorities.

Sprint Backlog: A set of items from the product backlog that the development team has committed to completing during a specific sprint. The sprint backlog is created during the sprint planning meeting and is used by the development team to guide their work during the sprint.

Increment: The sum of all the completed product backlog items at the end of a sprint. The increment represents the progress made by the development team during the sprint and should be potentially releasable. Each sprint builds on the previous sprint, and the increment should be delivered incrementally and iteratively until the product is complete.

3.2. Data Migration Motivation and Plan

The objective of data migration is to move data from one system or location to another, while minimizing disruption to the business and maintaining data accuracy and integrity. Before agreeing to begin the data transfer process in this project, the client conducted an evaluation of the company and considered the following motivations:

Consolidation: if the company decides to merge or acquire other organisations, they may need to consolidate their data into a single system or location to facilitate collaboration and streamline operations. This task will be simpler if the data is stored and maintained in cloud software.

Upgrading or replacing systems: as technology advances, the company strives to keep pace by upgrading or replacing outdated systems.

Moving to the cloud and performance optimization: the company was interested in migrating their data to the cloud in order to benefit from features such as scalability, cost-effectiveness, and remote access, and to ensure that the sensitive data is stored securely. By migrating the data to the cloud, the company was also seeking to optimise performance and reduce access times.

Before beginning the migration process, the team developed a detailed plan for the migration process, including timelines, resources, and potential risks. This plan was reviewed and approved by all the relevant stakeholders.

Figure 3 - Data Migration Plan Step by Step

To ensure a successful migration with minimal risk, the team followed best practices such as:

- Defining a clear migration plan and timeline
- Conducting a thorough data assessment and analysis
- Cleaning and transforming data before migration
- Testing the migration process before it is run on live data
- Conducting regular data backups and recovery testing
- Monitoring the migration process and fixing any issues as they arise
- Verifying the accuracy and completeness of the migrated data.

Since there was a large amount of data to be migrated, a decision was made to adopt the Trickle data migration strategy and progressively migrate the data to the new system, which allows the old and new systems to be used concurrently and minimises disruptions to business operations.

3.3. TOOLS AND TECHNOLOGIES

Migrating data from SAS Enterprise Guide to Databricks was accomplished through the use of the Azure Data Factory. The Azure Data Factory is a cloud-based data integration service that enables the creation, scheduling, and management of data pipelines. In this project, the team used a few tools to assure data quality and integrity, efficiently replace the old software with the new, and complete the migration. Jira software, SAS Enterprise Guide, Databricks, and Azure Data Factory were the tools used.

3.3.1. JIRA Software

Jira Software is a project management tool used by software development teams to plan, track, and manage their work efficiently and collaboratively. It allowed our team to create and track issues, bugs, and tasks in a collaborative environment. It provided a range of attributes and functionalities, including project and issue tracking, agile boards, customizable workflows, and real-time reporting. One of the key features of Jira Software was its flexibility. Jira's interface is visible in figure 3 below.

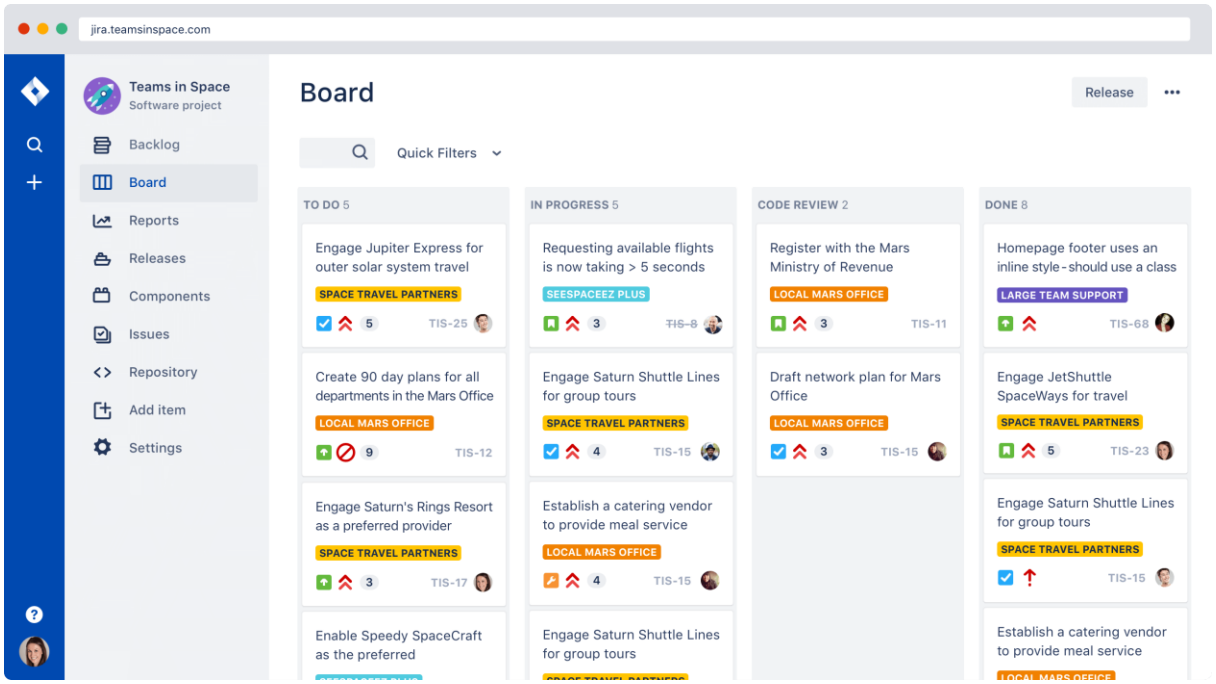

Figure 4 - Jira Software. Source: (*Jira Software - STAGIL*, 2022)

3.3.2. SAS Enterprise Guide

SAS Enterprise Guide is a powerful and user-friendly graphical user interface (GUI) which allows the user to achieve a wide area of data analysis, data mining, and data visualization tasks. SAS Enterprise Guide provide a range of tools for managing data, including importing and exporting data, data filtering, data sorting, and data transformation.

It also offers a range of statistical analysis and reporting tools, including summary statistics, regression analysis, time series analysis, and forecasting. It supported our team to report generation in a variety

of formats, including PDF, HTML, and Excel. Moreover, it provides a range of data visualization tools, including charts, graphs, and tables. Trends and patterns are quickly identified due to the visual descriptions of data made by the tools. Because it supports collaboration among team members, it allows multiple users to work on the same project simultaneously. It also provides version control and audit trail functionality, allowing to track changes and sustain a record of project history.

3.3.3. Databricks

Databricks is a cloud-based data processing platform that provides an optimised Apache Spark-based unified analytics engine for big data processing and machine learning tasks. It offered a scalable and secure platform for data processing, analytics, and AI, allowing the development team to write and run Spark code, as well as SQL and Python, in a collaborative and interactive workspace, making it easy to quickly and efficiently analyse and visualise large datasets.

This platform brought benefits like data standardization, data cleaning, and data conversion. It was used during the migration procedure to validate data. Due to Spark's data processing abilities to run complex data quality checks and identify any data issues that needed to be addressed before migration, as well as handling errors that occurred during the migration process, completeness, accuracy, and consistency of the data in the source and target systems were guaranteed throughout the process. The following figure 4 displays the databricks architecture.

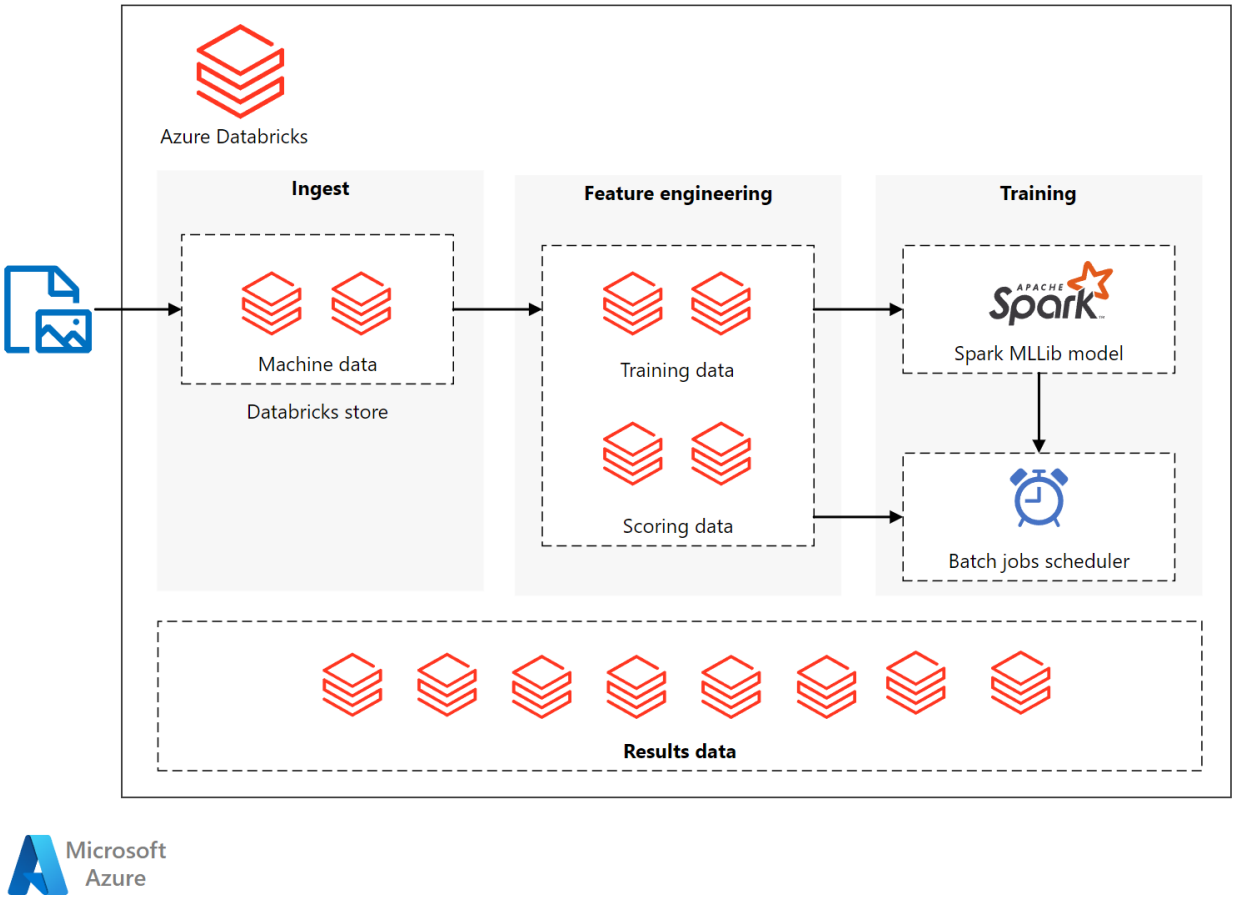

Figure 5 – Databricks Architecture. Source: (Ehrlinger, 2023)

3.3.4. Azure Data Factory

Azure Data Factory (ADF) is a cloud-based data integration service provided by Microsoft as part of the Azure cloud platform. It allowed our development team to create, schedule, and manage data pipelines that can move data from various sources to various destinations. ADF provides a graphical interface for designing data pipelines and supports a range of data sources and destinations, including SQL Server, Oracle, and more.

In this project, in a scenario where the data to be migrated was located on-premises, Azure Data Factory (ADF) was a powerful tool used in this project to help move and transform data from various sources to various destinations. It was used to ingest data from various sources, including on-premises and cloud-based data sources. It uses connectors to connect to various data sources and pull data into Azure. Figure 5 below illustrates the Azure Data Factory Architecture.

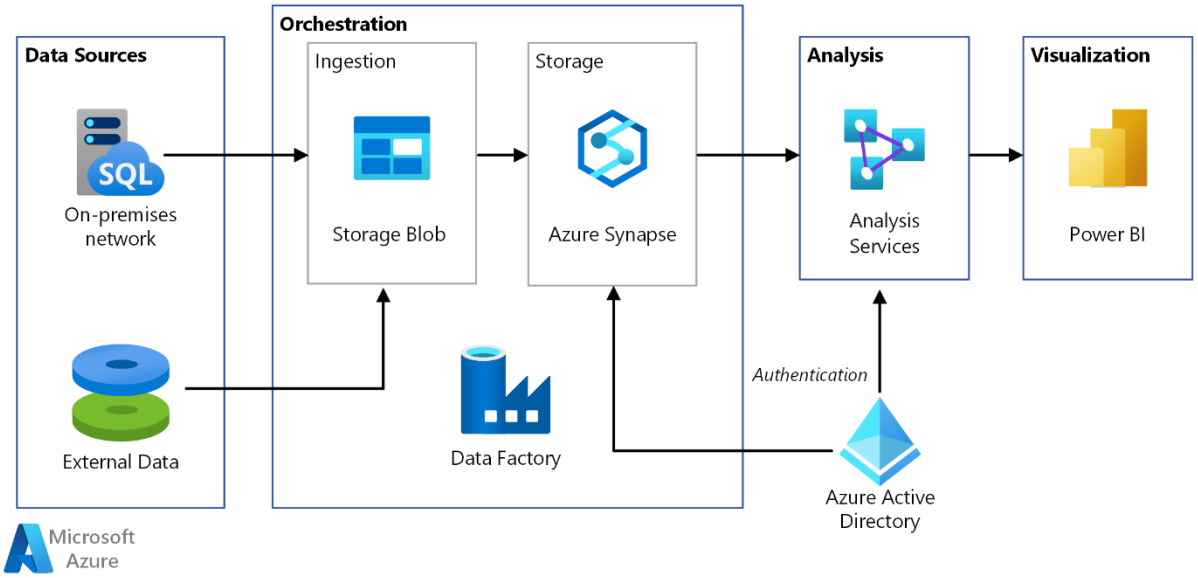

Figure 6 - Azure Data Factory Architecture. Source: (Martinekuan, 2023)

4. RESULTS AND DISCUSSION

4.1. SUCCESSFUL MIGRATION

The data migration process from SAS Enterprise Guide to Microsoft Azure Data Factory is expected to be executed with remarkable success, ensuring the seamless transfer of all the necessary data. The migration project, meticulously planned and meticulously executed, aims to meet and surpass expectations, as to be accomplished within the designated timelines and budget. Throughout the migration journey, every critical step is carefully orchestrated to ensure the preservation of data integrity and consistency. The meticulous planning and diligent execution of the migration process will result in a smooth transition, minimising any potential disruptions and maximizing efficiency. By adhering to the established timelines and budget, the migration team demonstrated exceptional project management skills and a keen understanding of the complexities involved. The successful completion of the data migration from SAS Enterprise Guide to Microsoft Azure Data Factory will provide the organisation with a powerful platform for advanced analytics and machine learning, empowering data-driven decision-making processes and opening new avenues for growth and innovation.

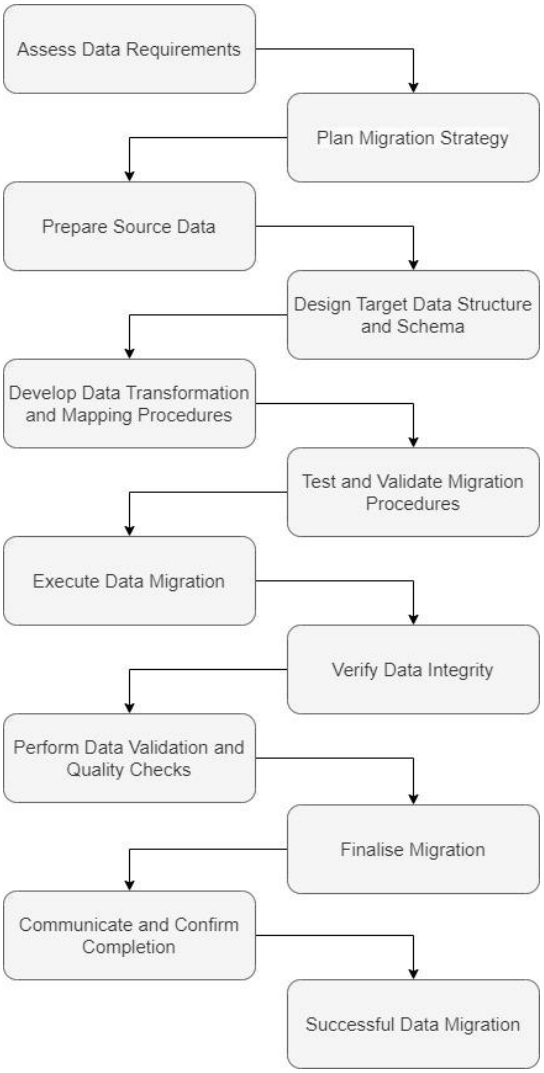

Figure 7 – Representation of the Steps to a Successful Migration

4.2. Data Accuracy and Integrity

Maintaining data accuracy and integrity was of paramount importance throughout the entire data migration process. Rigorous measures were implemented to ensure that the transferred data remained consistent and error-free, as well as to mitigate any potential discrepancies or data inconsistencies.

To begin with, a comprehensive data mapping exercise was conducted, where each element of the source data from SAS Enterprise Guide was meticulously matched with the corresponding target structure in Microsoft Azure Data Factory. This mapping process enabled a clear understanding of the data attributes and facilitated accurate transformation and migration.

Furthermore, before initiating the migration, a series of quality checks and validation methods were employed to identify and rectify any inconsistencies or anomalies in the source data. Additionally, data cleansing techniques were employed to eliminate duplicate records, resolve missing values, and standardise data formats.

Additionally, a sample-based testing approach is employed to validate the migrated data. Random samples are extracted and analysed to verify the correctness of the transformation and ensure that the migrated data retained its original meaning and context. To further bolster data integrity, comprehensive backup and recovery strategies are in place throughout the migration process, ensuring that data can be restored to its original state in case of any unexpected issues or failures.

By employing these meticulous quality checks, validation methods, and backup strategies, the migration team ensures that the transferred data remains consistent, accurate, and error-free. These measures provide stakeholders with confidence in the migrated data, enabling them to leverage the full potential of Microsoft Data Factory for advanced analytics and informed decision-making.

4.3. Minimised Disruption

The data migration process aims to be executed with utmost care to minimise disruption to the business operations. The migration plan was designed strategically, taking into consideration potential risks and challenges that could have impacted the company's day-to-day activities. To ensure a smooth transition, a meticulous assessment of the existing infrastructure, systems, and dependencies was conducted prior to the migration. This allowed the migration team to identify and address any compatibility issues proactively.

A comprehensive risk mitigation strategy was implemented to minimise the impact on business operations. Additionally, communication and coordination with the relevant stakeholders were prioritised, ensuring that everyone was aware of the migration plan and its potential impact on their workflows.

To further mitigate risks, a phased approach was adopted for the migration. Instead of attempting to transfer all the data at once, smaller subsets or incremental portions were migrated in a controlled manner. This approach allowed for thorough testing, validation, and troubleshooting at each stage, ensuring that any issues could be identified and resolved promptly before moving on to the next phase.

Furthermore, the migration plan included a robust rollback strategy in case unforeseen complications arose. Regular backups and checkpoints were maintained to facilitate a smooth rollback process if needed, reducing the potential impact on business operations and data integrity. By proactively addressing potential risks and challenges, adopting a phased approach, and maintaining effective communication with stakeholders, the data migration process was carried out with minimal disruption to the company's day-to-day activities. This ensured that critical business operations could continue seamlessly, empowering the organisation to leverage the benefits of the new Microsoft Azure Data Factory environment without compromising productivity or efficiency.

4.4. Improved Collaboration and Streamlined Operations

Migrating the dataset to a single system or location in the cloud will bring numerous benefits to the company, fostering collaboration and streamlining operations. By consolidating the data, the organisation will gain significant efficiencies and improve its overall data management processes, promoting data-driven decision-making and facilitating a more cohesive and unified approach to data management within the organisation.

First and foremost, having a centralised and unified data repository in the cloud will facilitate seamless collaboration among teams and departments within the company. Previously, data might have been scattered across multiple systems or silos, making it difficult to access, share, and integrate. With the migration, stakeholders across the organisation can now easily access the consolidated dataset, eliminating the need for time-consuming data transfers or manual integration. This streamlined access to data promotes cross-functional collaboration, allowing teams to work together more efficiently, share insights, and make informed decisions based on a comprehensive and up-to-date understanding of the data.

Furthermore, the consolidation of data allows for better scalability and flexibility. Cloud-based platforms, such as Microsoft Azure Data Factory, offer scalable infrastructure and computing resources. As the company's data needs grow, it can easily expand its storage and processing capabilities, accommodating larger datasets and more advanced analytics. This scalability also facilitates agility in adapting to changing business requirements, ensuring that the company remains competitive and responsive to market dynamics.

In summary, migrating the dataset to a single system in the cloud will bring significant benefits to the company. Collaboration among teams will improve, as data is easily accessible and shareable. Data governance and control will be strengthened, promoting compliance and security. Operational efficiencies will gain through streamlined data management processes. And the scalability and flexibility offered by the cloud will position the company for future growth and innovation. Overall, consolidating the data in the cloud will empower the organisation to harness the full potential of its data, driving better insights, informed decision-making, and ultimately, business success.

4.5. Technology Advancement and Optimisation

The migration to Microsoft Azure Data Factory will allow the business to take advantage of the unique capabilities and benefits of the cloud platform while also keeping up with technical improvements. Scalability is a remarkable advantage, as Azure Data Factory offers the flexibility to quickly increase or decrease computer resources in response to demand. With this flexibility, the business can manage

increasing data volumes and challenging analytics workloads without having to shell out a lot of money for infrastructure. Pay-as-you-go pricing allows the business to only pay for the resources and services they really utilise, which is another significant benefit. This eliminates the need for initial purchases of hardware or software licenses, saving a lot of money.

Additionally, the migration has led to improved performance, which translates into faster data processing, more accurate predictions, and improved overall efficiency. Metrics such as reduced processing times, increased model accuracy, and improved resource utilisation can serve as tangible examples of the optimisation achieved through the migration. Overall, the adoption of Microsoft Azure Data Factory will empower the company to embrace cutting-edge technologies, achieve cost-efficiency and deliver enhanced performance for data-driven decision-making processes.

4.6. STAKEHOLDER SATISFACTION

The future satisfaction of the client and other stakeholders with the outcomes of the data migration project is anticipated to be exceedingly high. The successful completion of the migration, within the planned timelines and budget, will ensure that the client can fully leverage the power of Microsoft Azure Data Factory for their data-driven initiatives. The consolidation of data, improved collaboration, and streamlined operations resulting from the migration will enable the client to make more informed decisions, drive innovation, and gain a competitive edge in their industry. With enhanced data accuracy, integrity, and accessibility, stakeholders across the organisation will experience increased efficiency and productivity in their daily workflows. The optimised data management processes, cost savings, and improved performance achieved through the migration will undoubtedly contribute to the overall satisfaction of the client and other relevant stakeholders, as they witness the tangible benefits and transformative potential of the data migration project.

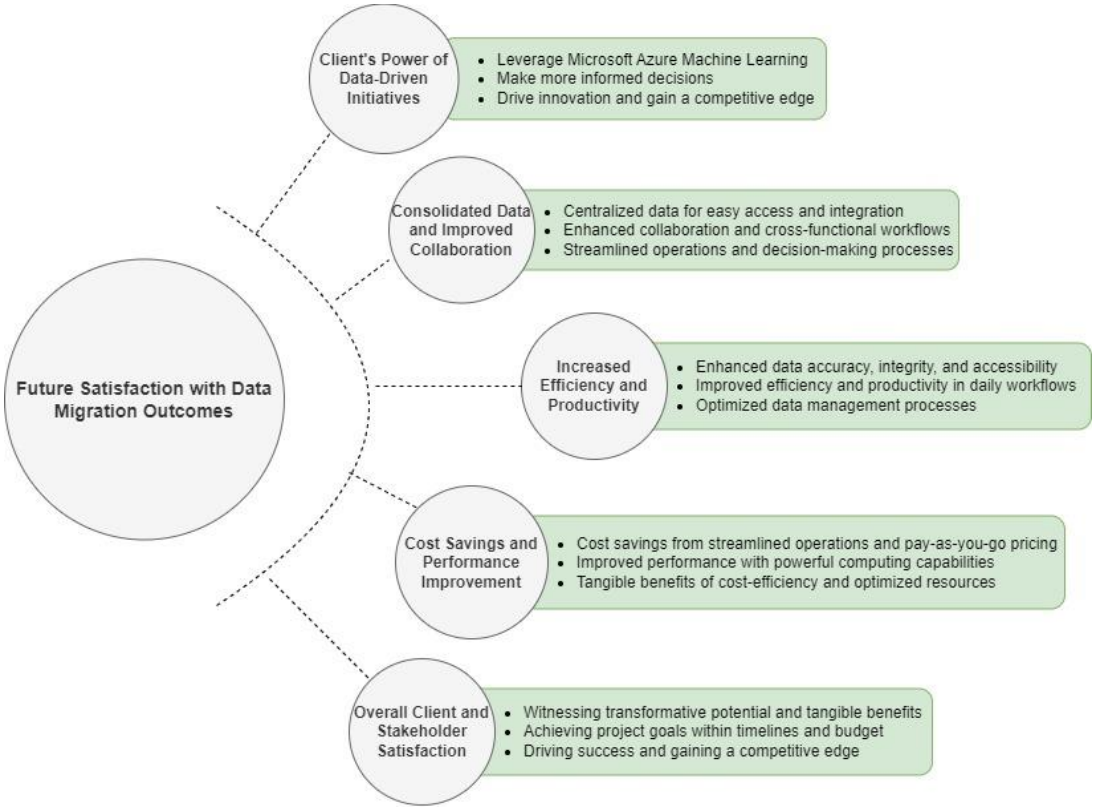

Figure 8 – Future Satisfaction with Data Migration Outcomes

4.7. LESSONS LEARNED AND RECOMMENDATIONS

This data migration project provided valuable lessons that can guide future endeavors in similar contexts. One key lesson learned was the importance of meticulous planning and thorough assessment of the existing infrastructure and data dependencies before initiating the migration. This helped in identifying potential challenges and compatibility issues upfront. The project also encountered some unexpected complexities during the mapping and transformation of data, which required agile problem-solving and close collaboration between the migration team and stakeholders. These challenges were overcome by maintaining open lines of communication, conducting regular testing and validation, and having a robust rollback strategy in place.

For future data migration projects, several best practices, tools, and methodologies can further enhance the process. Firstly, conducting a comprehensive data profiling exercise and data cleansing before migration ensures data accuracy and integrity. Implementing automated data validation checks, including data type validation, range validations, and referential integrity checks, can help identify and resolve issues more efficiently. Additionally, adopting a phased approach for migration, where data subsets are migrated and validated incrementally, can reduce the risk of large-scale failures and enable early detection and resolution of problems. Leveraging cloud-based platforms, such as Microsoft Azure, provides scalability, cost-effectiveness, and improved performance. Employing backup and recovery strategies ensures data resilience and minimises disruptions. Lastly, thorough documentation of the migration process, including lessons learned and best practices, will serve as a valuable resource for future projects.

By incorporating these recommendations and leveraging best practices and tools, organisations can enhance the efficiency, accuracy, and success of their data migration projects, ultimately ensuring a smooth transition and maximising the benefits of the new data environment.

5. CONCLUSIONS

In conclusion, the successful completion of the data migration process from SAS Enterprise Guide to Microsoft Azure Data Factory showcases the critical importance of meticulous planning, thorough assessment, and effective collaboration throughout the project. One of the key accomplishments of this data migration endeavor is the consolidation of data.

By migrating data from various sources into a centralised platform like Microsoft Azure Data Factory, the organisation gains a unified view of its data assets. Moreover, the data migration project contributes to improved data accuracy and integrity. Through careful data validation and cleansing processes, any inconsistencies or errors in the source data are identified and rectified. This ensures that the migrated data is reliable and trustworthy, providing a solid foundation for subsequent analysis and decision-making.

Streamlined operations is another noteworthy outcome of the data migration project. By transitioning to Microsoft Azure Data Factory, the organisation benefits from a more efficient and scalable data management infrastructure. The platform's capabilities enable automated workflows, data transformation, and seamless integration with other data tools and services. These streamlined operations lead to enhanced productivity and reduced manual effort, allowing teams to focus on higher-value tasks.

Enhanced collaboration is a crucial aspect that emerges from the successful data migration. Microsoft Azure Data Factory provides a collaborative environment where multiple users can work on data-related tasks simultaneously. This fosters cross-functional collaboration, knowledge sharing, and synergy among team members, leading to more effective data-driven decision-making processes.

Furthermore, the data migration project brings about significant cost savings for the organisation. By leveraging the capabilities of Microsoft Azure Data Factory, organisations can optimize their data infrastructure, reducing the need for expensive on-premises hardware and software licenses. The scalability and pay-as-you-go model offered by cloud-based solutions like Azure further contribute to cost efficiencies.

The successful completion of the data migration project positions the organisation to leverage the power of Microsoft Azure Data Factory for advanced analytics and informed decision-making. With a unified and reliable data foundation, the organisation can harness the platform's advanced analytics tools, machine learning capabilities, and data visualization features to derive valuable insights and drive innovation.

In summary, the data migration project from SAS Enterprise Guide to Microsoft Azure Data Factory has demonstrated the importance of meticulous planning, thorough assessment, and effective collaboration. The achieved outcomes, including data consolidation, improved accuracy and integrity, streamlined operations, enhanced collaboration, and cost savings, empower the organisation to make data-driven decisions and pave the way for future growth and innovation.

6. BIBLIOGRAPHICAL REFERENCES

1. Azeroual, O., & Jha, M. (2021). Without Data Quality, There Is No Data Migration. Big Data and Cognitive Computing, 5(2), 24. https://doi.org/10.3390/bdcc5020024

2. Sarmah, S. (2018). Data Migration. Scifed Journal of Computer Science and Information Technology, 9(1), 1-10.

3. Hussein, A.-E.-E. (2021). Data Migration Need, Strategy, Challenges, Methodology, Categories, Risks, Uses with Cloud Computing, and Improvements in Its Using with Cloud Using Suggested Proposed Model (DMig 1) Open Access. Journal of Information Security, 12(1), 79-103. https://doi.org/10.4236/jis.2021.121004

4. Mathur, V. (2016). Data Migration – A Perspective. Retrieved from Academia.edu https://www.academia.edu/41202536/Data_Migration_A_Perspective

5. Anderson, S. (2022). Documenting the Steps in Your Data Migration Process. StreamSets Blog. Retrieved from https://streamsets.com/blog/documenting-data-migration-process/

6. P.N., N. (2018). How to plan a successful Data Migration project. LinkedIn. Retrieved from https://www.linkedin.com/pulse/how-plan-successful-data-migration-project-nitin-pn/

7. N. Saranya, R. Brindha, N. Aishwariya, R. Kokila, P. Matheswaran and P. Poongavi, "Data Migration using ETL Workflow," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1661-1664, doi: 10.1109/ICACCS51430.2021.9441840.

8. Faraz, M. (2023). What is Data Replication? The Comprehensive 101 Guide. Retrieved from https://hevodata.com/learn/data-replication/

9. J. Sreemathy, K. Naveen Durai, E. Lakshmi Priya, R. Deebika, K. Suganthi and P. Aisshwarya, "Data Integration and ETL: A Theoretical Perspective," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1655-1660, doi: 10.1109/ICACCS51430.2021.9441997.

10. Microsoft. (2023). Import or link to data in another Access database. Retrieved from Microsoft Support https://support.microsoft.com/en-us/office/import-or-link-to-data-in-another-access-database-095ab408-89c7-45b3-aac2-58036e45fcf6

11. Campos, L. (2022). Data Synchronization: What It Is and How to Sync Data for Beginners. HubSpot Blog. Retrieved from https://blog.hubspot.com/website/data-synchronization

12. K. M. Varma and G. B. Se, "Efficient Scalable Migrations in the Cloud," 2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science (BCD), Danang, Vietnam, 2022, pp. 3-6, doi: 10.1109/BCD54882.2022.9900725.

13. Team Cleo. (2023). On Premise vs. Cloud: Key Differences, Benefits and Risks. Cleo Blog. Retrieved from https://www.cleo.com/blog/knowledge-base-on-premise-vs-cloud

14. Nyeint, K. A., & Soe, K. M. (2019). Database Migration based on Trickle Migrations Approach. National Journal of Parallel and Soft Computing, 1(1), March.

15. C. Prasanna, R. Subha, J. Sreemathy, P. Aravindh Ramanathan and M. Jainaveen, "Data Validation and Migration-A Theoretical Perspective," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 1540-1543, doi: 10.1109/ICACCS51430.2021.9441784.

16. Zaw latt, W. (2019). Data Migration Process Strategies. In Proceedings of the 7th International Conference on Computer Applications (ICCA 2019) (pp. 306-311). Retrieved from https://onlineresource.ucsy.edu.mm/bitstream/handle/123456789/1226/ICCA%202019%20Proceedings%20Book-pages-306-311.pdf?sequence=1&isAllowed=y

17. CFI Team. (2021). The process of transferring data from one data storage system to another and also between data formats and applications. Corporate Finance Institute. Retrieved from https://corporatefinanceinstitute.com/resources/data-science/data-migration/

18. Vijayaragavan, V., & Sivasankar, K. (2014). Data integrity in cloud computing - A survey. International Journal of Applied Engineering Research, 9(23), 23285-23297.

19. John Savill, "Backup, High Availability, Disaster Recovery, and Migration," in Microsoft Azure Infrastructure Services for Architects: Designing Cloud Solutions , Wiley, 2020, pp.297-323, doi: 10.1002/9781119596608.ch9.

20. Maniah, B., Soewito, F., Gaol, E., & Abdurachman, E. (2021). A Systematic Literature Review: Risk Analysis in Cloud Migration. Journal of King Saud University - Computer and Information Sciences, 34. https://doi.org/10.1016/j.jksuci.2021.01.008

21. Y. Chen, L. Li and Z. Chen, "An Approach to Verifying Data Integrity for Cloud Storage," 2017 13th International Conference on Computational Intelligence and Security (CIS), Hong Kong, China, 2017, pp. 582-585, doi: 10.1109/CIS.2017.00135.

22. Ansar, M., Ashraf, M. W., & Fatima, M. (2018). Data Migration in Cloud: A Systematic Review. American Scientific Research Journal for Engineering, Technology, and Sciences, 48(1), 73–89. Retrieved from https://asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/4450

23. Rai, R., Mehfuz, S., & Sahoo, G. (2013). Efficient Migration of Application to Clouds: Analysis and Comparison. GSTF Journal on Computing (JoC), 3, 10.7603/s40601-013-0023-z.

24. M. Alkhonaini and H. El-Sayed, "Migrating Data Between Physical Server and Cloud: Improving Accuracy and Data Integrity," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA, 2018, pp. 1570-1574, doi: 10.1109/TrustCom/BigDataSE.2018.00226.

25. Peek, S. (2023). What Is Agile Scrum Methodology? Business News Daily. Retrieved from https://www.businessnewsdaily.com/4987-what-is-agile-scrum-methodology.html

26. Kneafsey, S. (2020). The Five Scrum Events. The Scrum Master. Retrieved from https://www.thescrummaster.co.uk/scrum/the-five-scrum-events/

27. Harris, C. (2023). Agile scrum artifacts. Atlassian. Retrieved from https://www.atlassian.com/agile/scrum/artifacts#:~:text=Summary%3A%20Agile%20scrum%20artifacts%20are,%2C%20sprint%20backlog%2C%20and%20increments.

28. *Jira Software—STAGIL*. (2022). STAGIL GmbH. https://www.stagil.com/jira-software

29. Ehrlinger, J. (2023). *Batch scoring of Spark models on Azure Databricks—Azure Reference Architectures*. https://learn.microsoft.com/en-us/azure/architecture/reference-architectures/ai/batch-scoring-databricks

30. Martinekuan. (2023). *Automated enterprise BI - Azure Architecture Center*. https://learn.microsoft.com/en-us/azure/architecture/reference-architectures/data/enterprise-bi-adf