ATTE NOPANEN
UTILIZING THE INDUSTRIAL INTERNET FOR POWER PLANTS

Master of Science Thesis

ABSTRACT

The goals of this thesis were to explore the how Industrial Internet solutions could be utilized in power plants and their automation. The contents include a market analysis of the current Industrial Internet market and an analysis on the situation of competing power plant and automation system suppliers. Additionally, this thesis explains the different components of an Industrial Internet platform and studies how such a platform can be utilized to create a real-time process monitoring application for a power plant.

The Industrial Internet market situation is currently in a very fast moving state, however there is no clear market shareholder that has risen above the rest. Analyzing the Industrial Internet offering of companies such as GE, Siemens and ABB lead to the conclusion that many companies are developing their own visions of Industrial Internet platforms, but none have yet to offer anything very advanced or even complete.

These Industrial Internet platforms that are being developed can be summarized to be collections of components and tools that allow the utilization of advanced data collection and analytics capabilities. The key components include data sources, data pipelines, central storage, analytics, visualizations and security. Utilizing these components in the Valmet Industrial platform, this thesis presents a practical implementation of dashboards that can be used to monitor the performance of a power plant. The finished application allows a customer's production manager to monitor process performance from any device over the internet and provides easy access to data in support of business decisions and to analyze any problematic parts of key processes.

Finally, methods on how these newly developed capabilities can be utilized and integrated into business opportunities are studied to explain how the Industrial Internet platform can be utilized to improve current offerings and provide new products and services in the future. For example, predictive maintenance models can provide the capability to optimize maintenance routines and prevent process downtime. This is a logical next step to research when developing new features into the created platform. Moreover, the combination of Industrial Internet capabilities with new technologies such as virtual reality, drones and blockchains can bring completely new business opportunities in the future.

TIIVISTELMÄ

ATTE NOPANEN: Teollisen internetin hyödyntäminen voimalaitoksissa
Tampereen teknillinen yliopisto
Diplomityö, 62 sivua, 1 liitesivua
Lokakuu 2017
Konetekniikan diplomi-insinöörin tutkinto-ohjelma
Pääaine: Voimalaitos- ja polttotekniikka
Tarkastaja: professori Jukka Konttinen

Avainsanat: Teollinen internet, esineiden internet, voimalaitosautomaatio, KPI

Tämän työn tarkoituksena on tutkia teollisen internetin hyödyntämisen mahdollisuuksia voimalaitoksissa. Työ sisältää markkinatutkimuksen, jossa käydään läpi teollisen internetin nykymarkkinat ja selvitetään muun muassa voimalaitos- ja automaatiotoimittajien teollisen internetin tarjonta. Tämän lisäksi työssä tarkastellaan teollisen internetin alustan sisältämiä komponentteja ja niiden käytännön hyödyntämistä voimalaitoksen prosessin seurantasovelluksen kehittämisessä.

Teollisen internetin kehityksessä edetään tällä hetkellä hyvin nopeasti markkinoiden eri osa-alueilla. Monet yritykset, kuten GE, Siemens ja ABB kehittävät omia teollisen internetin alustoja hyödyntääkseen sen tuomia mahdollisuuksia, mutta selvityksen perusteella, kaikilla nämä alustat ovat vielä hyvin keskeneräisiä eivätkä tarjoa yhtään kehittyneempiä ratkaisuja kuin muut.

Nämä kehitteillä olevat teollisen internetin alustat ovat joukkoja työkaluja ja valmiita komponentteja, jotka mahdollistavat tiedon keräämisen melkein mistä vain ja datan kehittyneen analyysin. Alustat koostuvat moninaisista tietolähteistä, tiedonsiirtokomponenteista, keskitetystä tallennustilasta, analytiikasta, visualisoinnista ja tietoturvasta. Tässä työssä toteutetaan visuaaliset näytöt voimalaitoksen suorituskyvyn seurantaan hyödyntäen yllämainittuja komponentteja Valmet Industrial Internet alustassa. Työn aikana valmistunut sovellus antaa asiakkaan tuotantojohtajille mahdollisuuden käyttää oman prosessin tietoa päätöksenteossa ja ongelmatilanteiden analysoinnissa.

Lopuksi työssä pohditaan tapoja hyödyntää teollisen internetin tuomia mahdollisuuksia nykyliiketoiminnassa. Esimerkiksi ennakoivan huollon mallin avulla laitteiden elinikää voidaan arvioida reaaliajassa, minkä avulla voidaan vähentää ennakoimattomia tuotantokatkoksia. Tallaisten kehittyneiden analyysimahdollisuuksien sekä uusien teknologioiden kuten lisätyn todellisuuden, lennokkien ja lohkoketjujen hyödyntäminen teollisen internetin kanssa ovat loogisia seuraavia askelia tämän kehitysalueen tutkimukselle.

PREFACE

CONTENTS

APPENDIX 1: EXAMPLE DATA FILE

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AI	Artificial Intelligence
AR	Augmented Reality
BI	Business Intelligence
CSV	Comma Separated Values
DBMS	Database Management System
DCS	Distributed Control Systems
DEV	Development environment
DV2	Data Vault 2.0
EC2	Amazon Elastic Compute Cloud
EDW	Enterprise Data Warehouse
EKPI	Energy KPI Monitoring
ETLA	The Research Institute of the Finnish Economy
HDD	Hard Disk Drive
IIoT	Industrial Internet of Things
IoT	Internet of Things
IoTSP	Internet of Things, Services and People
IT	Information Technology
JDBC	Java Database Connectivity
KKS	Identification System for Power Stations
KPI	Key Performance Indicator
MDS	Master Data Services
ML	Machine Learning
ODBC	Open Database Connectivity
OT	Operational Technology
PROD	Production environment
S3	Amazon Simple Storage Service
SFTP	SSH File Transfer Protocol
SQL	Structured Query Language
SSD	Solid-Stade Drive
SSH	Secure Shell
TEST	Test environment
VII	Valmet Industrial Internet
VPC	Valmet Performance Center
VR	Virtual Reality

1. INTRODUCTION

In a nutshell, the Industrial Internet is a tool or set of tools that makes it possible to utilize data from networked devices in a way that creates additional insight and value for industrial businesses. In broader terms the Industrial Internet encompasses megatrends such as cloud computing, where data processing and storage is moved from local computers to servers accessed through the internet, big data analysis, which includes the analysis of large amounts of unstructured and currently unutilized data, and interconnected network devices, which allows the comparison of geographically separated devices and machines that can make decisions based on the information they share with each other. [33]

The purpose of this thesis is to study the background and requirements of adopting IoT solutions on an industrial point of view, assess the current market situation and to test the implementation of a current power plant monitoring solution utilizing Industrial Internet components.

1.1 Motivation

The reason this subject was chosen for this thesis is in the potential gains and possibilities the Industrial Internet offers in comparison to what is currently utilized. Currently industrial devices are creating vast amounts of data, which is primarily stored on local servers and where the data can only be accessed through local Intranet. Some of the data is used for real time monitoring where the values are displayed in the control room, after which the data is never looked at again. A very small part of the data has been used for actual analysis and historical comparison and each report has required a large amount of work to produce. Still, most of the data is never even seen by human eyes. [33]

The Industrial Internet enables companies to utilize and show the data that is being collected without having to spend money on large workloads. First, by collecting data in the cloud, the need to maintain physical servers on site for storage is reduced and storage quantity is easily and cheaply scalable. Second, data can be viewed anywhere, not only onsite using a physical connection. This means easy real time data analysis access for users who have not been able to access data before without waiting for an analyst to compile reports. Third, going through and analyzing vast data collections requires computing power. By moving the analysis to the cloud, computing time and cost can be minimized by reducing the need for owned computer clusters and providing easy scalability for when business grows. [37]

Utilizing the benefits of cloud computing and networked devices enables businesses to monitor, control and optimize processes in ways that have been expensive or difficult before. As the Harvard Business Review details in Figure 1 below, smart connected devices make it possible to monitor, control, optimize and even autonomously run devices and processes, with each step building on the last. [39]
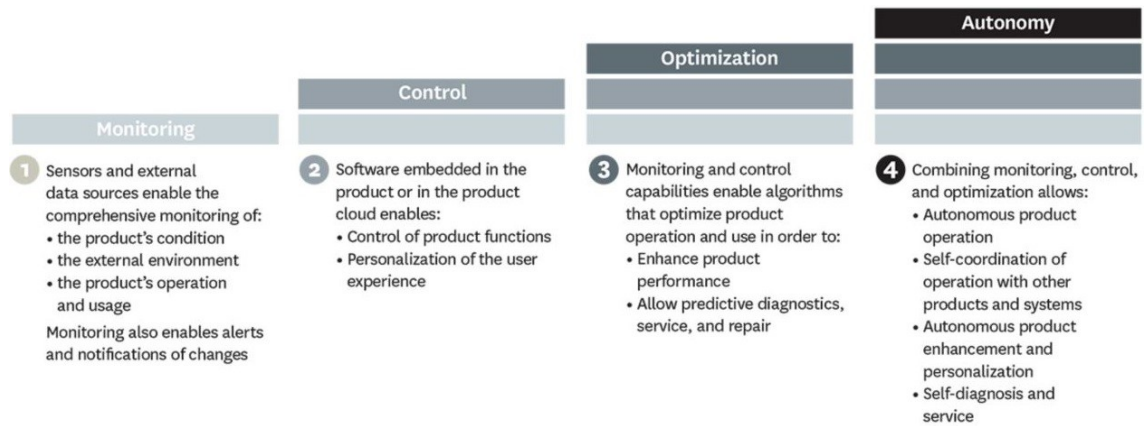
Figure 1. The capabilities of smart connected devices. [39]

Receiving real-time data from devices allows comprehensive monitoring of the devices condition, operation and environment. Combining this monitoring data with device embedded control software allows the remote operation of the device with control decisions based on the monitoring data. After enabling the monitoring and control capabilities, the possibility to optimize product performance arises. This can include analyzing data for the optimal running parameters or predictive diagnostics for maintenance so that downtime can be minimized. Finally, building upon all the earlier steps, autonomously running products can be designed which coordinate their operation with other devices and self-diagnose problems and schedule maintenance without external input. [39] These possibilities are part of the goals that are hoped to be achievable building on the groundwork provided by this thesis.

Not only do these capabilities enable the efficient access of data and process optimization for one site, it also allows the easy connection of one sites data to other similar sites or data external to the site. This makes it possible to efficiently analyze whole chains of connected sites, benchmark performance between power plants and utilize predictions in the process control of multiple sites in connection.

1.2   Thesis Structure

Beginning in chapter 2, this thesis will explore the current Industrial internet market situation with the key drivers, challenges and current market competitors' offerings. After understanding the current market situation, chapter 3 will investigate what kinds of components an Industrial Internet platform contains and how they can be utilized to store data, calculate key values and report data through visualization.

Once the terms and theories related to the Industrial Internet have become familiar, chapter 4 will explore the possible applications of the Industrial Internet and how they can be used to improve future energy segment performance. In this section a use case for a boiler monitoring application is presented with details of each step of development.

Chapter 5 will explore the business opportunities of the developed application and how it can be used to improve current and future business. Finally, the findings and achievements of this thesis will be summarized in chapter 6.

2. IOT MARKET ANALYSIS

Digitalization and IoT development is expected to be a rising trend in the coming years. Companies are investing large sums of money into research and development of IoT solutions to secure a market share for themselves. For example, SAP has stated they will invest 2 billion euros into IoT by 2020 [44] and IBM has pledged to create an IoT Unit and invest 3 billion over 4 years [27]. By 2020, annual revenues for IoT vendors could exceed USD 470B [2] selling hardware, software and comprehensive solutions. The competitor analysis in this thesis will be conducted based on the public materials available from each vendor during the research process in 2017. Focus will be on the main competitors in the energy and power plant automation.

While analyzing the various marketing materials and three layers of the Industrial Internet market were identified:

1. The Cloud Service Providers (Amazon, Microsoft Azure)
2. The Industrial Application Platforms as a Service (PaaS), (Honeywell INspire, Siemens MindSphere)
3. The Applications that utilize the platform or are standalone products (Siemens Plant Data Solutions, Honeywell Uniformance Suite)

Cloud service providers provide the physical servers and computers that the cloud based platforms and applications run on. These server and computing sites require a lot of capital and so it can be cheaper and more flexible to partner up with the cloud service providers to cut the need to open their own server sites. The storage price in a cloud storage can be cheaper as servers can be concentrated in large scale server facilities so that costs can be reduced in comparison to maintaining multiple smaller data servers. In addition, the actual physical storage in the cloud is not limited by the users' physical devices and storage capacity can be flexibly scaled as data increases by simply paying for the storage that is being used at any given point. This is opposed to the need to add actual physical storage devices to a locally managed storage server. [33]

The Industrial Internet platforms are services which contain the necessary layers to gather, store, analyze and present data. The platforms contain the development and customer interfaces to the applications. There are mainly two different types of platforms. First, closed platforms are places where only the platform owner develops first party applications to specific needs. These are usually vendor specific platforms intended for use with their own products and to provide additional services to their customers. Second, open platforms also allow 3rd party developers to register and develop applications by providing the environment to build and sell them.

The third layer includes the applications, which are solutions developed to solve specific problems or to provide additional value to processes. These applications include for example Key Performance Indicator (KPI) monitoring to visualize key measurements or calculate values that indicate how well a process is performing or component lifetime analysis, where the wear of different devices can be predicted.

2.1 IoT Definitions

Before delving deeper into the current situation and to clarify some of the terms that appear in references and are used in this thesis, this chapter will briefly define the most important terms and their relations. The many names of the term Industrial Internet include the Internet of Things, Industry 4.0, M2M or Industrial Internet of things. New names and terms appear as the technology develops and some even fall out of popular use. Even though these terms are considered synonyms by some, there are differences that can be found when a closer look is taken.

On a grass roots level, the Internet of Things is seen to refer to a series of connected devices which transmit data to a central server or cloud, which is a server based storage accessible over the internet as opposed to having data and processing done on the users own computer. On this central platform, data-analysis and presentation can be done by anyone with the right access and regardless of their desktops performance. IoT is the universal term that covers everything related to the phenomenon being studied, but is often attributed more towards consumer products and services. Adding Industrial to IoT results in the Industrial Internet of Things, or Industrial Internet. By using this combined term, we refer to the IoT phenomenon as seen from the industrial perspective and how it can be utilized in the different industrial applications, which narrows down IoT to not cover consumer products.[33]

Industry 4.0 was coined as a name for the German strategic initiative to establish Germany as a market leader in the utilization of IoT in manufacturing solutions. The term is originally restricted to only the manufacturing industry, but it is sometimes used in other contexts as well. [30]

M2M refers to Machine-to-Machine, which means the communication or exchange of information between devices. In many regards, it is a synonym to the Industrial Internet, however, M2M only covers a small portion of what the Industrial Internet contains. [33]

Big Data is defined by Gartner as high-volume, high-velocity and high-variety information assets that require new forms of data processing to utilize for value [20]. To understand the scale of big data, volume can refer to all the data on the Internet or high-velocity and variety can refer to constant data feeds from multiple sources using very different formats. Inmon and Lindstedt (2015) identify Big Data to contain characteristics of very large volume, in inexpensive storage, managed with parallel processing and in an

unstructured format [52]. The composition of corporate data and the part that is big data is characterized in Figure 2.
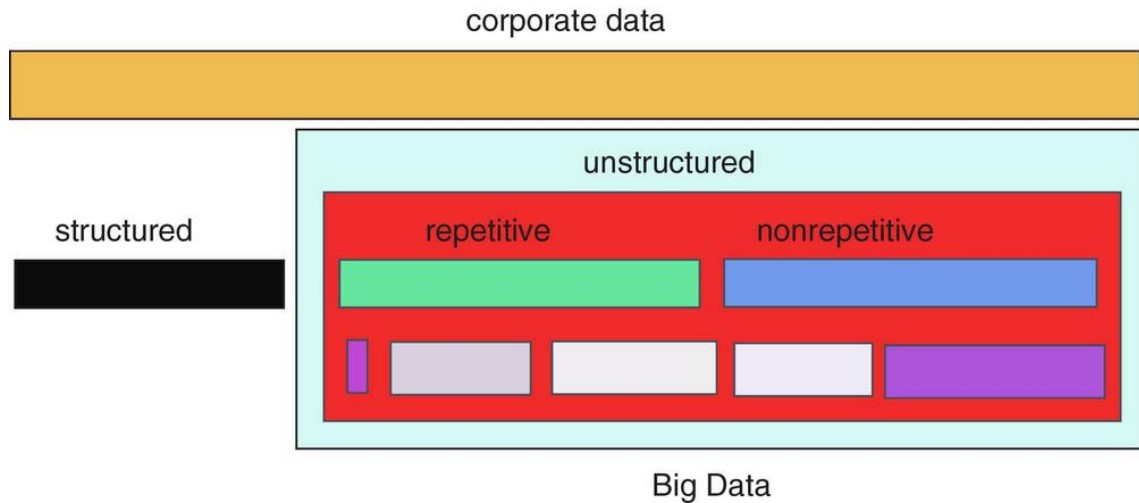
Figure 2. Big Data in corporate data. [52]

Large amounts of data require storage, and for feasibility costs cannot be high, that is why Big Data is characterized as being stored on cheap storage for example cloud storage, which can store data on a petabyte scale. Moreover, obtaining data from many sources at very fast rates makes the structuring of data into manageable databased require a lot of work defining attributes, keys and indexes before utilization. In the case of big data value comes from being able to process unstructured data masses, without the need for this work. Finally, to have a chance the process data at the required scale in a reasonable timeframe needs coordinated parallel processing power to be able to utilize it in a meaningful way.

Since this thesis is focusing on the industrial side of IoT and more specifically the applications for Power Plants, the primary term used in this thesis will be the Industrial Internet.

2.2 Current Market Situation

In a survey conducted by IDC in 2015, it can be seen that the majority of respondents from each industry segment are either actively pursuing or considering adopting IoT solutions as displayed in Figure 3. [36]
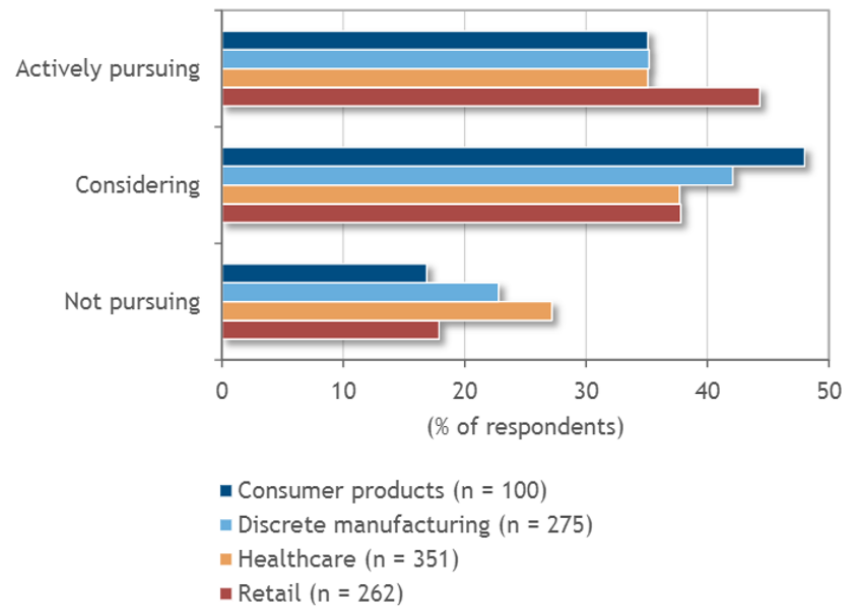
Figure 3. IoT adoption intentions by industry segment. [36]

Only a minority of actors in for example discrete manufacturing are not even considering the possibilities of IoT. Over half of all respondents in all segments are either actively pursuing or considering incorporating IoT solutions in their businesses. [36] As the results show, IoT is not just focused on a small part of the market. To understand why this is, the Bain IoT customer survey studied the motivations behind this pursuit of adopting IoT solutions. The results can be found in Figure 4, where the reasons for implementing IoT solutions for either new revenue possibilities or cost reductions can be found.

Percentage of respondents who say they will implement IoT solutions for the following reasons

Reason	%
Improving quality of service or product, leading to price premium, lower support costs	47%
Improving productivity of workforce	45%
Increasing reliability of operations	44%
Increasing asset productivity	37%
Reducing cost of materials and waste	35%
Accessing new customers	34%
Increasing customer satisfaction to reduce churn and support costs, gain revenue	33%
Insights for new products and services or cost efficiencies	32%
Faster time to market, lower development costs	26%
Reducing risk of theft and other loss	22%
Transitioning to services offerings	18%
Linking price to business outcomes	16%

■ Cost more important ■ Revenue more important □ Both important

Source: Bain IoT customer survey, 2016 (n=533)

Figure 4. Reasons for implementing IoT solutions according to the Bain IoT customer survey. [2]

The top motivation to implement IoT according to the respondents is to improve the quality of a product or service to increase revenue. Improving workforce productivity and reliability of operations are also high on the list. Following these are multiple goals to increase efficiency and allow for the increase of revenue.

As reported by the Research Institute of the Finnish Economy (ETLA), the three megatrends driving the industrial internet include: 1. Globalization, 2. Digitalization and 3. Urbanization. But why is it happening right now? First, it has been enabled by the maturation of intelligent technologies and their lowered prices. Second, a large part of the industrialized world is connected by a high performance and fast networks. Third, the internet makes it possible to monitor and operate processes and businesses from just about anywhere in the world. In addition, the IT skills of the modern workforce have improved and Big Data analysis platforms have developed to the point where large amounts of data can be processed. [34]

Despite being on its way to be the next megatrend, the widespread adoption of Industrial Internet solutions still has challenges to overcome that need to be solved as it is matured

for business. According to the Bain IoT customer survey in Figure 5, the main concerns of IoT buyers are in security, return on investment and integration difficulties. [2][1]

Percentage of IoT buyers who selected this as one of their top three barriers to implementation of solutions

Source: Bain IoT customer survey 2016 (n=533)
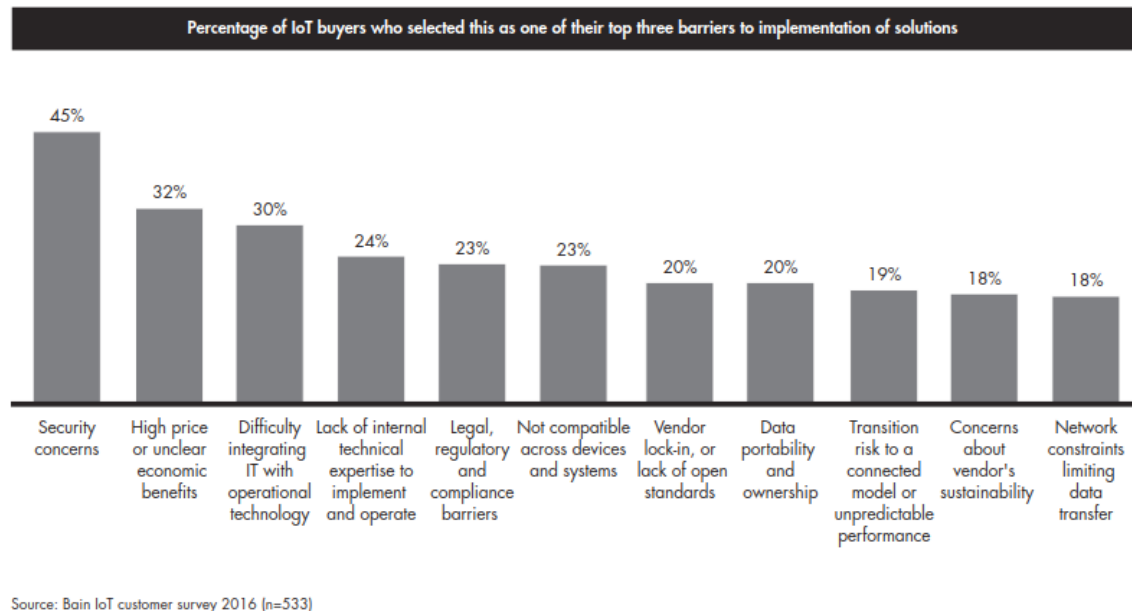
Figure 5. *Bain IoT customer survey results. [2]*

Security and data ownership, the largest issue to overcome, often rise as companies start utilizing the Industrial Internet. When data is stored outside the customer's servers in an external cloud and can be accessed over the internet, access rights and connection security needs to be well implemented to prevent unauthorized access and make sure sensitive data stays confidential. In addition, when contracts on data transfer and storage are made, data ownership needs to be clarified. On the other hand, by having data stored on an external cloud the need for direct access connections to the facilities systems can be reduced in some support or analysis cases, which reduces the risk of unauthorized access.

Industrial Internet integration into existing processes and systems also include their own challenges. Usually there are multiple databases where data needs to be gathered from and each one needs to be configured separately. Due to the lack of standards for IoT, using competing IoT services with a third-party automation system can require additional work inflicting higher costs. Some vendors are developing the system on an open standard basis which will allow the connection of all devices, but may not always be as cost effective and optimal as a tailored solution for a product. In the future, there may also be demand that applications can utilize and integrate with existing cloud services used by a site.

When large amounts of data are transferred, and stored for analysis, it is important that data integrity is considered. A miscalibrated sensor could be sending the wrong values, a data transfer process might change a value because of an error or units used in different

parts of the world may cause normalized values to be wrong. To overcome this challenge, some form of data integrity checking will be required as data amounts rise.

2.3   Competition

As so many companies are currently developing Industrial Internet solutions there is of course overlap between the offerings of competitors. At this stage, it is somewhat difficult to determine the stage at which a company that is developing these solutions is at. Assessing each company based on their marketing material is only accurate to the extent that they give a guideline as to what are the goals of their products. A common note from studying these different companies has been that most appear to be in development stages or adopted in minor cases.

As stated earlier and since the IoT is being developed from many different angles and affects multiple different industries, this thesis focuses on identifying major players in the energy and power plant automation areas. The following companies, assembled in Table 1, were identified to have IoT solutions in development during the course of this thesis.

Company	IoT platform	Power Plant Solutions
General Electric	Predix	Not specified
Siemens	MindSphere	Plant Data Services
Honeywell	Inspire	Uniformance suite
ABB	IoTSPA	ABB Ability
Schneider Electric	EcoStruxure	Plant Struxure
Valmet	Valmet Industrial Internet	Valmet Performance Center

Table 1. *Energy and power plant automation companies developing IoT solutions. [49] [48][45][25][26][23][22][4][3]*

Based in the marketing material available by these companies on their websites, they are all either providing a universal platform to that provides analysis and other solutions to multiple industries or focus on one segment of industry. In this case the energy industry is a common segment they are all focused in to some degree. All the players' base offerings are similar with regards to what the IoT can provide, with some players differentiating themselves in certain areas. The similarities in each companies' platform offerings are as follows:

1. Data gathering to a cloud service.
2. Data availability on multiple devices over the internet
3. Storing history data over time periods for analysis
4. Some degree of IT security

To note, these are the promises that each companies service will provide and currently most features might still be in development. Data gathering to a cloud service includes the storing of both internal (e.g. process data) and external data (e.g. weather, Nordic pool) to an accessible database. Being stored on the cloud, the data can be made available through web browser access, which in turn enables the usage of just about any device capable of browsing the internet including mobile phones, tablets and laptops. The storage of history data is for analysis purposes and reporting. The security offerings aim to keep access of data safe, where access is controlled and data can only be seen and accessed by those with the correct access rights. This can, for example, be a multiple level hierarchy, where a plant manager can access all the data of a plant, while an operator can only see the data related to their work area or equipment.

In addition to these universal platform offerings, each vendor usually provides their own software solutions targeted at certain industries. These solutions or applications utilize the platform and the data stored on the platform to give insight from the data that has been gathered. Examples of such applications include:

- Analyzing the received data and compiling reports
- Real time process surveillance and alarms for unusual process parameters
- Optimization services based on the data analysis
- Preventative maintenance procedures, which estimate the lifetime of components and allow for reduction of process downtime.

These possibilities are similar to what was defined by the Harvard Business Review earlier in Figure 1. Building on the platform, which enables the gathering and storage of data, applications to monitor, optimize, predict and autonomously run are possible to develop.

2.3.1 General Electric

GE is entering the IoT market with its open, multi-industry Predix platform. In this case Predix is a Platform as a Service (PaaS), where developers can create applications and solutions themselves for improving data use in their business. It is envisioned to function similar to an application store for mobile devices, which is a digital distribution platform where anyone can create applications to provide value and sell these applications to paying customers. Keeping the platform open can be a way to bring new ideas into a large company such as GE, where certain possibilities the end user can invent may not normally be considered [15]. According to the GE presenter at a Power Plant Digitalization seminar held in Tampere the applications featured on Predix can even be in competition with one another or a Siemens solution and currently GE has about 35 energy industry related apps

running on the Predix platform [15]. The business model is based on a subscription charge for the user. [23][22][15]

Currently, it seems anyone can register as a developer on the Predix website and start developing apps. GE even offers training programs for developers to learn how to start their own projects. Based on the prerequisites for the training courses, developers would need quite a bit of programming knowledge to get started. There is also course on application design, where the participant would learn how to plan and design an application, after the IT side is implemented, but ultimately a process engineer for example would probably have a hard time completing a whole application without the required IT knowledge. [42]

As a feature launched with Predix, GE is bringing its take on virtual power plant modeling with the GE Digital Twin. The GE Digital Twin works by gathering the data of a product, e.g. a turbine. The design process is digitalized from the start, which allows for a model of the device to be stored in the cloud. By gathering the operating data of the entire installed base of the product, for example 100 turbines, the virtual Digital twin is formed by creating a universal simulation of the device using the expected running parameters of all the connected device. By comparing the expected simulation to the installed products and adding additional external data sources, discrepancies can be noticed, changes to running parameters can be simulated and lifetime estimates can be calculated. These enable taking preventative maintenance measures, optimization and investment calculations. [21]

In a demo video of the Digital Twin, Predix analyzes a running turbine, compares it to the Digital Twin, locates the problems and offers solutions. These solutions each display what the base variables are, what calculations are use used and how confident Predix is with the predicted outcome. The solutions can be either manual adjustments of the running parameters or software based and automated, which utilizes a GE product to do the adjustments for the customer. The presentation also demonstrates the use of VR (Virtual Reality) and AR (Augmented Reality). VR glasses allow a user to see and look around a completely virtual environment. AR, on the other hand, augments what the user sees in a real environment by overlaying virtual elements to what a person is looking at. Utilizing these technologies, the presenter demonstrates how to view the virtual Digital Twin with VR glasses, locate damages and adjust parameters without any physical user interaction with the device. [21]

To gauge the readiness of GE's offering, we can look at the for example the Predix developer network, where services, tools and applications are listed [22]. The Predix platform has a long list of possible tools and services tied to it and in the apps section there are currently 10 applications listed. These include the HyperCube analytics platform, a process outcome optimization demo, control systems status monitoring and a weather application to make weather data available to real time control systems. The rest of the

applications are related to connecting to machines or creating an application. However, these applications are all designated as coming soon, so from this we can assume, that there is still some work to progress through before Predix is fully market ready. As an additional insight stated by an independent Lux Research Analyst Isaac Brown in press release, GE's Predix platform is not as ready as the marketing implies: there is very little market penetration and no large-scale tests have been conducted in practice. [24]

2.3.2 Siemens

The Siemens MindSphere is Siemens' PaaS offering to the Industrial Internet market. Similar to GE's Predix platform, MindSphere is also an open application store type environment. In addition to the basic features of tools for data collection and analysis, it features it advertises simplicity to in connecting devices, pay-per-use billing for flexible business models and an open standard for compatibility with third party equipment, not just Siemens' own products. [48] During the Power Plant Digitalization seminar held in Tampere, Finland, a presenter from Siemens stated that Siemens has over 100 customers using the MindSphere platform, with about 50 energy related applications. [15]

While the MindSphere platform is open to all industries, Siemens does offer a ready Industrial Internet application for the energy industry called Plant Data Services (PDS). The PDS offers solutions to key areas such as performance optimization, asset monitoring and availability and uptime improving. This application includes the customer interface, personalized web portal for accessing reports through your internet browser. Through the interface a customer can access plant wide process details, including KPI and other performance drivers. The solution includes Energy savings and asset analysis to sustainably lower plant operating costs by identifying and recommending upgrades, modernization or retrofitting with investment analysis to calculate the savings potentials of the choices. PDS is not however listed in MindSphere's applications list. [49]

MindSphere's current listed applications include Fleet and Machine Manager for monitoring devices, KeepSecure for security monitoring and Product Insight to analyze insights from product performance. In addition, there is a possibility to create your own apps and dashboards. Based on these listings, Siemens' offering is still tapping into a very small portion of the potential of the Industrial Internet and development of applications is needed before larger market impact is possible. [47]

2.3.3 Honeywell

Honeywell is set to gather partners to develop IIoT solutions to the market with INspire, a program to bring together customers, vendors, licensors, consultants and Honeywell's experts to develop IoT solutions for a multitude of operational challenges. The INspire Ecosystem will feature data management, onsite control, analytics and security solutions. [25]

Honeywell's application and analytics engine targeted at the energy industry is called the Uniformance suite. The Uniformance suite contains the Uniformance PHD historian for the capture and storage of real time process data. The Uniformance Asset Sentinel is the application for managing assets in combination with risk detection and uptime predictions. In addition, KPI analysis is provided with Uniformance KPI. Through these components the Honeywell suite promises easy scalability, reduced data collection efforts and process alarm investigation packaged in to a web browser based service where data and reports can be accessed with configurable dashboards. [26]

Honeywell classes its strengths to be a lean track record in both IT and operations in addition to exclusive IIoT- ready services. This suggests that Honeywell is first aiming to convert its present offering to be classified as IIoT ready. Currently there is little indication of any ready to market applications by Honeywell. [26]

2.3.4 ABB

ABB is making its move towards the Industrial Internet with its Internet of Things, Services and People (IoTSP), which features data analysis, web portal access and 3rd party service integration. Focusing on industrial sector, ABB Ability is a cross-industry digital capability to connect their customers to the Industrial Internet of Things. Cross-industry in this case covers mining, utilities, manufacturing, infrastructure and transport. It is stated to combine current solutions and devices with cloud services. Moreover, recently ABB unveiled it will be collaborating with IBM to utilize the IBM Watson AI to unlock new machine learning solutions, for example defect detection and optimization. [4]

For the power plants ABB Ability is marketed to offer solutions for KPI monitoring, performance monitoring, preventative maintenance and power plant virtualization. In addition, there are more specific targeted solutions like boiler startup optimization, emission control and process optimization. [3]

The situation with ABB seems to be similar to the other companies analyzed. There are lots of areas in development, however the current offering seems to only cover a small portion of the potential. ABB does list a few cases where there have been significant improvements utilizing the digital capabilities of ABB Ability, cases include for example boiler startup reduction by 15% and NOx emission reduction by 5-15% through remote services in addition to reducing 40% of abnormal events in automation. [3]

2.3.5 Valmet

Valmet has also entered the industrial internet market with its own offering. Having a long history in automation and advanced process control, Valmet aims to strengthen the services provided to customers with new and game changing tools. Their ecosystem is

called Valmet Industrial Internet and it's offering includes performance optimization, reliability services and on-demand remote expert support. It is also mentioned that Valmet has been developing VR and AR solutions for data visualization and other applications. [50]

Valmet's four industry specific performance centers, including one for energy, aim to act as a customer's channel to all of Valmet's Industrial Internet services and applications. These physical performance centers have remote monitoring capabilities of customer sites and perform data analysis to reach the performance targets of customers. In addition to this physical center run by Valmet experts, the customer has access to a one stop customer portal through which the customer can view the applications and analysis related to the customer's sites. [51]

Outlined in this the VII marketing material is their long history in automation and process control in addition to highlighting significant customer successes that have been possible with Valmet's process knowhow. In the paper and pulp industry, Valmet has helped customers achieve chemical savings of 8.7%, mill energy consumption decreases of 8 % and production increases of 2.8% with performance agreements and utilizing Industrial Internet components. [51]

Based on the public material Valmet has been testing and developing Industrial Internet capabilities for some years and is preparing to scale up by utilizing the potential of the cloud. However, readymade applications still appear to be in development and not completely market ready.

2.3.6 Other IIoT players

In addition to the earlier players identified in the competitor analysis process, s few other companies of note related to the IoT market are gathered in this chapter. These companies are KONE, Wärtsilä, Fortum and Maintpartner.

KONE has turned to IBM Watson for its IoT venture in the elevator and lift business. In early 2017, KONE launched its KONE Care customizable maintenance service and 24/7 Connected Services utilizing the IBM Watson IoT platform. According to the press release, these new service offerings have been piloted and will be commercially available during 2017. In the 24/7 Connected Services, Watson is utilized for gathering operating data of elevators, analyzing it and using it to optimize and prioritize maintenance stops and lower downtime. Henrik Ehrnrooth, President and CEO of KONE predicts that this is just the beginning, and KONE will be incorporating AI into its services to help predict and suggest resolutions to potential problems.[35]

Wärtsilä, focusing on the maritime industry, has launched the Wärtsilä Genius Services to provide onshore or offshore monitoring optimization of engine efficiency and propulsion. It also includes benchmarking across the monitored and installed base, health indicators and security solutions.[53]

Fortum has started utilizing VR and AR in the running of power plants with the Daqri Smart Helmet. This VR helmet supports overlaying data on the operators view, thermal vision, work instruction displays and remote expert communication giving an operator a greater situational understanding and providing quickly accessible help. Fortum utilizes the VR helmet for 360 video tours of sites and immersive training solutions. For example, a nuclear power plant safety training has been virtualized, in which a control room fire can be or other control situations simulated. In addition, it is planned to be utilized for remote maintenance, guidance and visualizing key points in meeting. In a nuclear power plants, the VR helmet makes it possible to visit and see areas that cannot be access during normal operation. [14]

On the predictive maintenance front, Maintpartner has developed a predictive maintenance system, MP Intelligence, to maximize the reliability of the customer's production facility. This self-learning system either monitors data flows of a site or is fed history data, from which the system learns the normal operating parameters of the site. After this process, the software then compares the new data to the universal model, to detect discrepancies in any measurement. Alarms can be configured to go off at certain deviations from the norm, which provide insight into wearing components, operator errors and possible process malfunctions. Maintpartner claims the MP Intelligence prevents from 30-80% of all production disruptions. The business model is based on a monthly subscription, that can be terminated with a month's notice. [40]

3. CLOUD BASED INDUSTRIAL INTERNET TOOLS AND STRUCTURE

The following chapters will explore the terms and components related to a cloud based Industrial Internet platform to provide a basic understanding of the functions before moving on to one practical application of this information in chapter 4. An Industrial Internet platform contains the following components:

1. Data gathering and transfer
2. Cloud storage and computing
3. Data warehouse and data modeling
4. BI tools and analytics

From a power plant or process engineer end user point of view, this type of expertise is needed in the first step, where the process data is gathered and the last step, where the engineer's knowhow allows the relevant analysis and visualization of the data. However, to get a complete understanding of how data is used and transferred, this thesis will also explore the Information Technology (IT) side of the Industrial Internet platform, which will explain the workings of the cloud storage and data modeling done in between the two steps visible to the end user.

3.1 Sensors and Data

Almost every new power plant has an automation system or distributed control system (DCS), which gathers continuous data from different parts of the process to control and monitor. This data is captured by sensors, which are electrical devices that measure physical, chemical phenomena or actions and convert the gained information into an electrical output. Common sensor types include: acceleration, speed, temperature, humidity, pressure, flow, chemical content, vibration, electrical voltage or current, radiation and sound. Sensors make up the bulk and base layer of the Industrial Internet. Large amounts of data, but one source is not much unless it has the context of many. Analytics can then find the correlations in unexpected places. [15]

Sensors measurements can be divided into two, analog and discrete. Analog measurements output a value of the measured variable over a period of time. Discrete measurements have certain values, they are usually used for component status indication by have the value 0 for when the device is not running and 1 when it is running, no other possible values.

These DCS sensors are what provide the bulk of the process information from power plants. Utilizing IoT devices however, additional wireless or wired sensors could be

added to supplement process data with this information. In addition to be able to utilize external sources like weather forecasts and electricity pricing it can be possible to find new connections in the stored data to improve processes.

To get the gathered sensor data from the DCS into an accessible central cloud server, the data needs to be transferred. There are multiple ways to transfer data over the internet, some examples include email, file transfer, external database connections or even direct IoT sensor feeds over the internet.

Modern DCS have a local historian computer to store all the gathered data from sensors. From this historian, data can be uploaded to a cloud storage by gathering data into files and sending them over the internet by for example an email attachment or utilizing a file transfer protocol, to securely transfer files from a computer to another. It can also be possible to read data directly from a historian database, however, this way the benefits of cloud analysis speed and reliability are lost since queries go outside the central storage and connections must remain stable for the data to be available.

3.2 Data Modeling: Enterprise Data Warehouse

This chapter will familiarize the reader with the terms and components required to the software side of an enterprise data warehouse and introduce the reader to data modeling using one technique. First a quick introduction into the terminology of database storage.

An Enterprise Data Warehouse (EDW) is an enterprise wide database that stores unified business information across a company and makes it accessible to all users in the company.

A Database Management System (DBMS) is a software system to control and manage a database. It controls the placement, access, update and deletion of data on the database's storage, so that users can do create and read data without the need for the repetitive and complex tasks managed by the DBMS. [52]

Structured Query Language (SQL) was developed by IBM in the 1970s to allow for database users to access data without the need to develop programs to make the queries. It is a language that expresses the users request for certain data in the desired order. [32]

Open Database Connectivity (ODBC) is an Application Program Interface (API) for database access utilizing SQL as its language. It is designed to allow a single application to access different DBMSs using the same code. [32]

To determine how an EDW is populated, a data model is needed. The work in this thesis utilizes Data Vault 2.0 (DV2), which is a model that includes: modeling, methodology,

architecture and implementation best practices. It is a common foundational data warehouse architecture, that has been utilized in developing the Valmet Redshift Data Warehouse. Compared to Data Vault 1.0, which focused primarily on data modeling, DV2 has included components to relate directly to business intelligence. It is designed to support Big Data with NoSQL databases and to provide guidance and best practice. [52]

As introduced by W.H. Inmon and Daniel Linstedt, "A Data Vault model is a detail-oriented, historical tracking, and uniquely linked set of normalized tables that support one or more functional areas of business." [52] Basically it is a collection of tables, a database, where the table are unified in structure. Historical tracking mean all changes and additions can be tracked by time and source. Uniquely linked means data connections between separate tables can be made using keys, a sensor measurement table can be linked to a site table, through which one can find additional details of the customer. The Data Vault model concept is presented in Figure 6.
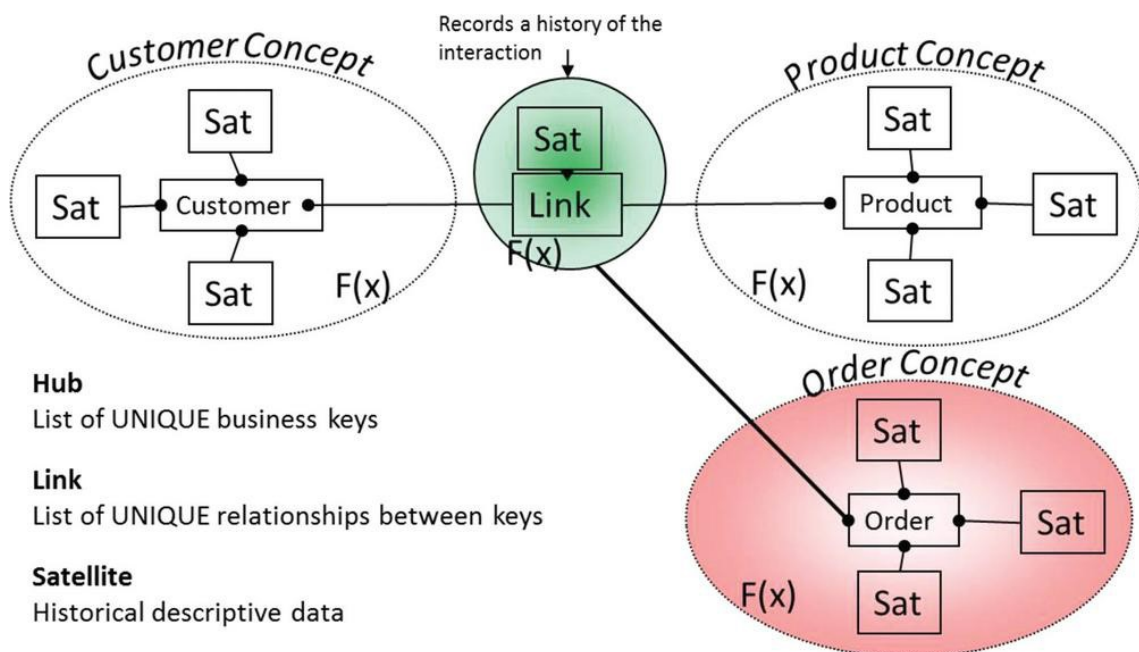
Figure 6. Data Vault model concept. [52]

The components of the data vault model include hubs, links and satellites. Hubs are collections of unique business keys that exist across an enterprise. Links describe the relationships between different business keys. And finally, satellites are the actual data tables were unchanging and time series data is stored. To support big data DV2 uses business key related hash keys for uniquely identifying data. This enables parallel processing by removing the dependency of DV1 to process hubs first, then links, and satellites last. Utilizing these components, a data vault model is made and new data processing can be

automated to place incoming data into the right satellites, with all the connections available. The design parameters allow the data vault model to scale as needed with no inherent limit to model size or amount of data. The only limitations that come into play are imposed by the infrastructure. Moving on to the Data Vault architecture, meaning the organization of data found in relation to sources and end user data, a visualization of this architecture can be found in Figure 7. [52]
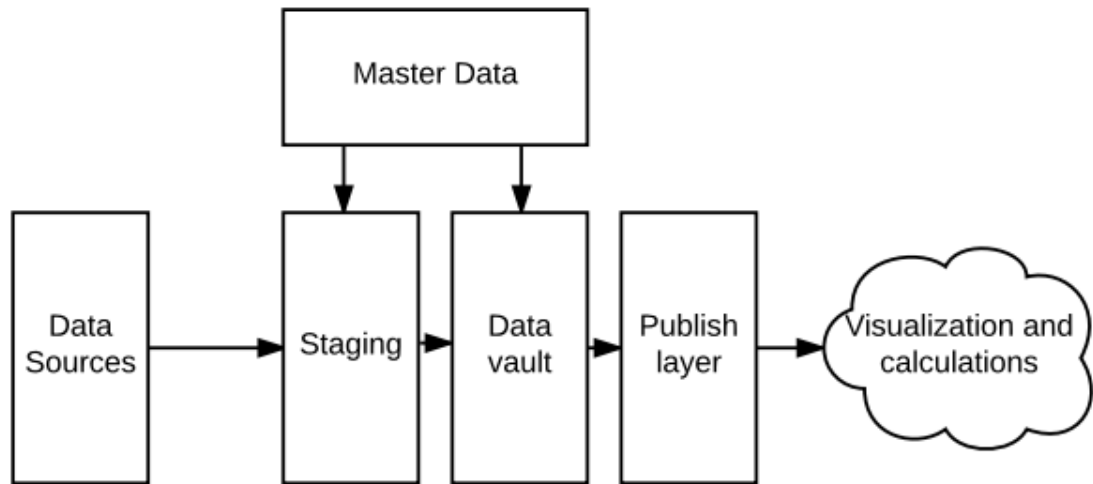
Figure 7. *Data Vault architecture.*

Data Vault architecture has three tiers, staging, data warehouse and the information delivery layer. The first tier, staging, is where data arrives in batches or as real time data flow. It is temporary storage from where the data is moved into the correct data warehouse tables, satellites, by a set of hard rules. In the beginning all data is transferred through staging, but as data connections are developed, for real time data it is ideal to skip the staging phase and have real time data written directly to the data warehouse data model to reduce data transfer times. The second tier, data vault, which is the actual enterprise data warehouse (EDW), is where the data is structured into the data models satellites, which are connected by the data vault models links and hubs. This is where data is stored in a structured model. To access certain data from the EDW, the third publish layer containing information marts or information delivery views are snapshots or views of the data in the EDW model, that are accessed for visualization, calculations or other appliances. In addition to these three layers, there is separate application for master data that can be used to for example fill in missing data based on average values or correlate sensor tag to additional information like site and device. [52]

In addition to the Data Vault data model, DV 2.0 contains the methodology and best practices to develop DV projects and applications utilizing for example Capability Ma-

turity Model Integration (CMMI), Total Quality Management (TQM), Project Management Professional (PMP) and Six Sigma techniques for process improvement. However, these are not the focus of this thesis. [52]

3.3   AWS Cloud Storage and Computing

All cloud applications ultimately require a physical server for data storage and computing power for analysis. To understand the environment the VII data warehouse runs on, an introduction the Amazon Web Services (AWS) cloud offering is presented in this chapter. The AWS storage classes can be divided into 3 categories: Slow storage, raw data storage and data warehouse storage for modeled data, which is represented by Redshift. A price comparison between the different storage components of the AWS offering can be found in Table 2. This table contains four different storage types and for each type the storage cost, query prices, data transfer prices, structure and speed are specified. Data prices contains is a monthly fee that contains the storage of data on the cloud server. Request prices are incurred when the data is accessed from the cloud storage with for example queries. Data transfer prices include the cost of moving data between storages, for example over the internet. Availability is the speed at which the data is accessible and structure defines how the data can be stored, for example it can be modeled or without structure at all.

Type	Data price per month	Data price per year	Request prices	Data Transfer	Data availability	Structure
AWS Glacier	$0.004 per GB	$48.00 per TB	$0.055 per 1,000 requests	$0.01 per GB	Average 3-5 hours	No structure
AWS S3 Standard	$0.023 per GB	$276.00 per TB	$0.005 per 1,000 requests	$0.004 per 10,000 requests	Fast	Raw Data Buckets
Redshift DS	$0.230 per GB	$2760.00 per TB	$1.742 per hour of EC2 host use	Free between S3 and redshift	Fast (HDD) up to 3.30GB/s	Data Model
Redshift DC	$0.998 per GB	$11976.00 per TB	$1.742 per hour of EC2 host use	Free between S3 and redshift	Very Fast (SSD) up to 3.70GB/s	Data Model

Table 2. *Example of storage prices on AWS. [10][9][7][6]*

The prices selected for comparison are based on a 1 year contract term with an upfront payment of 50% at the start. Lower prices are achieved with a longer contract term and higher upfront payments. However, all AWS services can also be utilized on demand and without any yearly term or upfront payments. This does result in higher costs on all features, but allows for flexible cloud usage to, for example, make a proof of concept. [10][9][7][6]

The AWS Glacier is classified as a slow storage space, where large amounts of data can be stored for the cheapest price. It is ideal for example long term backups or data logs and

other information that is rarely needed for access. The minimum storage time purchased is 90 days and data can then be accessed based on three levels of speed which increase in cost: expedited, standard and bulk. The fastest and most expensive access speed level is expedited, in which data can be accessed within 1-5 minutes and costs $0.03 per GB retrieved. In the standard level, data can be accessed in 3-5 hours' time with a price of $0.01 per GB retrieved and an additional $0.055 per 1,000 requests. Bulk access allows for retrieving very large datasets with a cost of $0.0025 per GB retrieved and $0.0275 per 1,000 requests. [7]

The S3 standard storage is a place where large amounts of data can be added cheaply with the possibility to access the data at any time. Ideally raw data feeds can be uploaded unstructured right into the S3 into buckets which contain the raw data in different form. From the S3 the data can then be moved to the redshift data vault, where it is modeled to include relations to other tables and data sources. Data transfer between the S3 and redshift is free and does not incur additional costs, whereas transfer over the internet does incur a minor fee based on the amount of data transferred. [11]

The Redshift data warehouse designed for cost-effective structured data storage on a petabyte scale to allow for efficient data analysis. For storage, there are currently two possibilities based on the speed of the storage. On the one hand, Dense Storage (DS) is based on HDD storage for a large amount of data and a low-price point. Dense Compute (DC) on the other hand utilizes SSD storage smaller datasets but with higher performance for calculations. As mentioned above, data transfer between redshift and the S3 is free, but otherwise data can be accessed based on the computing hours used when running an EC2 instance. [8]

For computing needs, the Amazon Elastic Compute Cloud (EC2) provides scalable computing capacity in the AWS. Using EC2, virtual servers can be deployed and configured as needed to provide the computing power required for development and the deployment of applications without the need for you own computing hardware. EC2 has a price structure of on demand pricing for use at any time, where charges are hour based for what is used. In addition, there is a possibility to make reserved instances or dedicated hosts for reducing prices in long term commitments. Dedicated host servers allow enterprises to utilize existing server-bound software licenses on the cloud. [6][5]

3.4   BI Tool

Once the data is available on the cloud, and modeled in the EDW, it can then be utilized for value. Business Intelligence (BI) tools are utilized to access the data on the EDW for visualization and reporting purposes.

In the modern view, BI tools are meant to provide visual data discovery as opposed to an IT-centric approach, where database skills and knowledge are required to access data.

These new discovery methods allow the user's selections and even written text to be translated into database queries, without the need for the user to understand the query language behind the data access. Combining this data discovery aspect with easy to use visual chart making and dashboards customization allows users with process knowledge do work that previously required data scientist qualifications. [43]

Birst is a BI and analytics tool, that was founded in 2004. It is server oriented to serve as a platform for entire businesses as opposed to a local desktop oriented tool. It was aimed to decrease deployment time enterprise scale analytical applications by reducing the complexity of configuration and by creating automated analytics techniques. [12]

On the VII platform Birst works as the developer and customer interface for viewing dashboards. Birst can connect to numerous different types of databases and sources either with local data or live database connections. As Birst runs on a cloud, no local computer installations are needed to access it. A web browser is used instead and the user logs on with the right credentials and then has access to the spaces and data that is defined for the user's credentials.

Birst solutions are called spaces, which are collections of dashboards, data, reports and user groups. Each space can be considered a solution or product geared towards different needs. During the course of this Thesis I worked on a solution that will be called Energy KPI Monitoring, which is displayed in Figure 8 below.
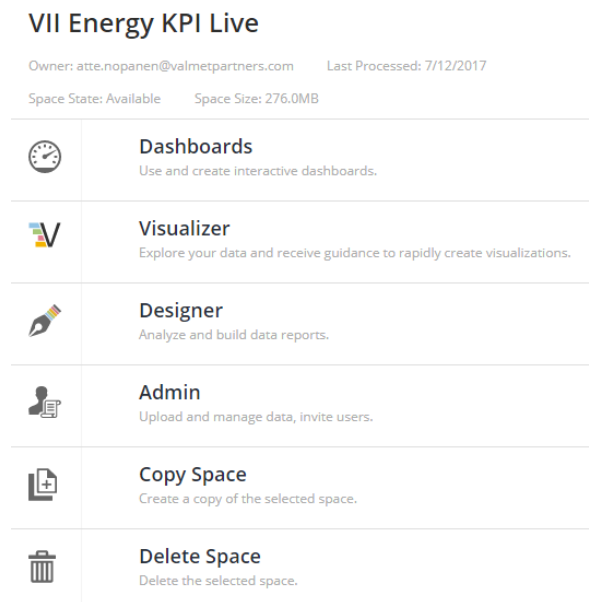
Figure 8. *Birst space for Energy KPI and its contents.*

Each Birst space contains components that are visible based on the access rights a person has to the space. For Energy KPI, I was the creator and owner, so all the different components were visible to me. These include: Dashboards, Visualizer, Designer, Admin and the space actions copy and delete.

The Dashboards component is where reports are displayed in interactive visual tables, charts and KPI indicators. These dashboards compromise the information displayed in the product, they are the views the customer sees when he accesses the solution. Visualizer, shown in Figure 9, is the main chart and table creation tool. It has a graphic interface that supports drag and drop and, in my experience, was very fast to learn and quite easy to use when the EDW data model is in place.
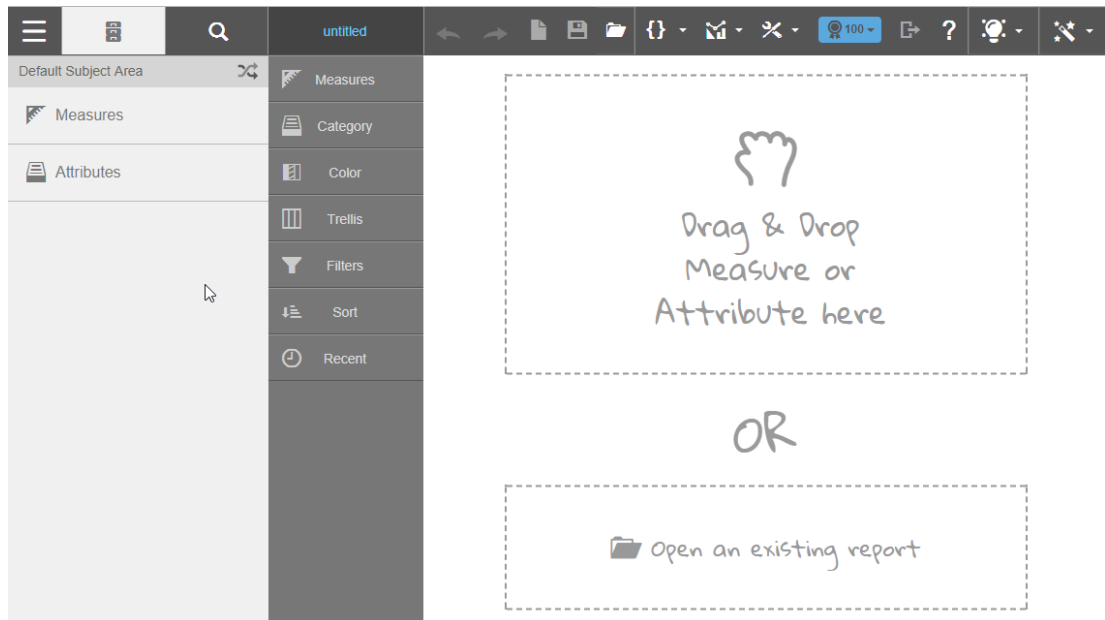
Figure 9. *Birst Visualizer.*

Creating charts is as fast as selecting the measure from the measures list and adding an attribute by which it is visualized. As an example, to make the chart visualized in Figure 10, we could choose the fuel power measurement a measure and tag timestamp as the category to create a chart that shows fuel power values at each measurement time. To limit the number of measurements, a filter for day is added, so that only values for the filtered day are shown. Selecting these values from the measures and categories list draws the fuel power trend shown in Figure 10. The chart is drawn in a matter of seconds, and when it is done, it is interactive, e.g. the user can mouse over a point in the trendline to view additional information or a certain set of points can be selected by clicking and drawing over the chart to filter data only to the selected points.
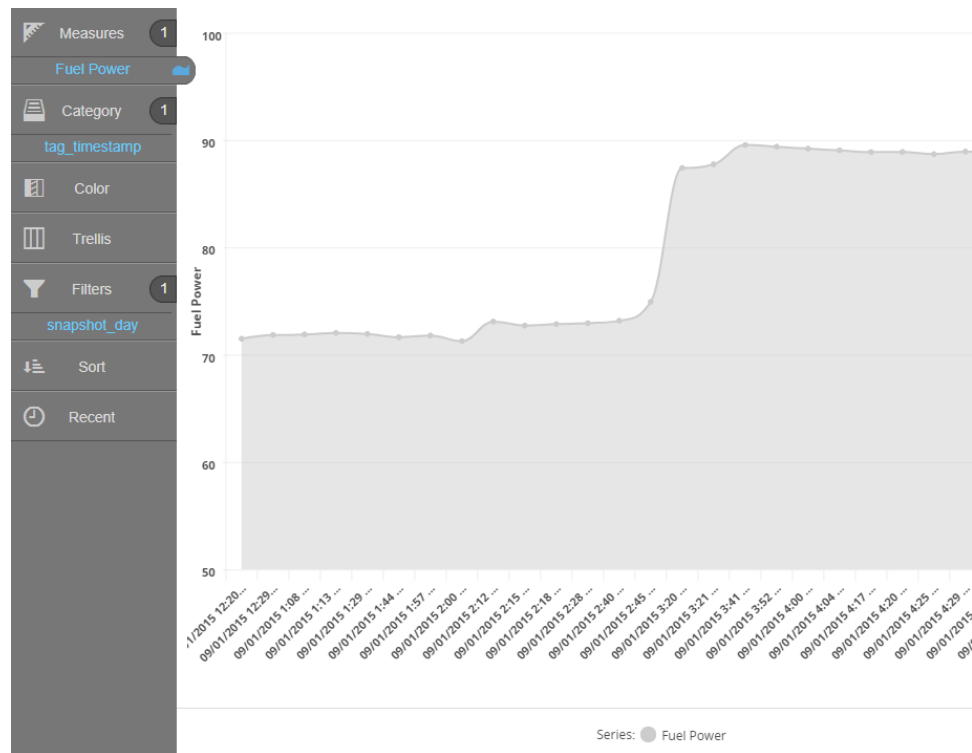
Figure 10. Simple fuel power trend, drawn with Birst visualizer.

In addition to these automatic features, there are many additional formatting settings available for the developer. Chart colors can be manually determined or simply chosen from a list of themes that unify the look of the chart with others. Conditional formatting parameters can be added to for example color a certain part of the graph in red when a value is below a specified limit. All the fonts, labels and axes sizes are customizable by the user, most simply by clicking the area that needs to be edited and typing in a new name. Created charts can be then saved and even used as a template for another trend, just by changing the measure in the case of the example trend. Simply selecting another measure in place of fuel power returns the same type of graph, but with new trend data.

Continuing with the space components, designer is another older version of a chart and report building tool. Currently it has some features that the visualizer does not and it supports making very detailed reports where you can control almost every aspect of appearance. Most of the features of designer are being added to the visualizer little by little.

The Birst Admin component, shown in Figure 11, lets you control the space settings, add data connections, manage your data model, process data and manage access. Space settings include naming, time zone and query settings. Through the data sources tab connections to databases such as redshift can be made, or data can be loaded to Birst directly from external files such as excel. If using a data warehouse the data modeling part does not need user input, and if using an external file such as excel, the user can quickly set up a data model with a little bit of prerequisite knowledge from a Birst training exercise for example.
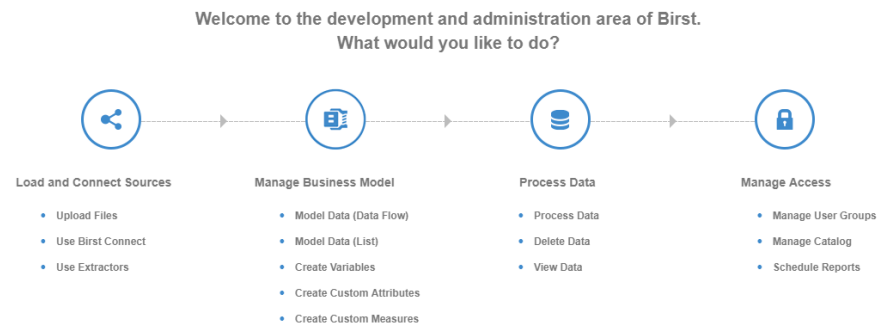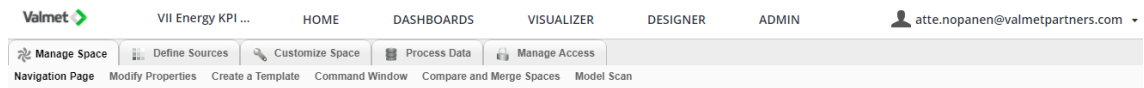
Figure 11. Birst Admin module.

To help new developers starting with BI tools for the first time, Birst has an active community, where Birst users can ask questions, share solutions and offer ideas for further improving the Birst tool. From this Birst Community you can search for any problem you are having and might find a ready solution for it if someone has brought it up before. If nothing is found, the user can also create a new discussion for the problem. As Birst is being constantly developed and new features are added, the Birst community also offers the possibility for users to vote for ideas to be implemented. That way ideas that get a lot of support are usually added to Birst during version updates.

The remaining space actions copy and delete do what their names imply. Copy allows a person with space admin rights to make a backup copy or a test copy that can be edited without affecting the original. Delete completely removes the selected space. In addition to these actions, there is a possibility to swap spaces in the admin component, where you can swap an updated space to replace one that is in use with minimal effect on users.

While Birst supports uploading data in multiple ways, making a database connection is done by selecting Birst connect in the Admin module. Birst Connect is Birst's own tool for connecting to different data sources. It supports multiple database types, as seen in Figure 12.
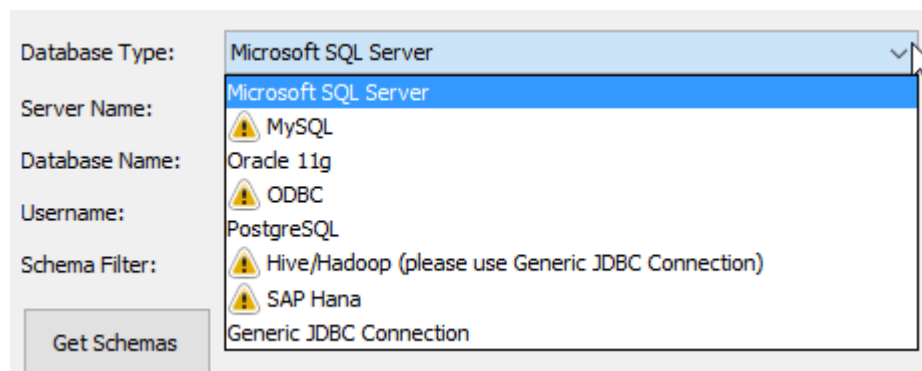
There are two basic ways to make the database connection. The first is to connect to a database and read the data into Birst for use. This method however requires scheduling and new data is retrieved only every time the task is scheduled to run. It can be useful for data that does not change often. For continuous time series data, a live connection to the database can be made, where Birst queries the database every time data needs to be retrieved. This way the newest data is always available for dashboards, when a dashboard is looked at.

For both these connections the steps include adding a source for Birst Connect to query. To define the source the user must fill in the necessary database information like server name, username and password for data access as is shown in Figure 13. For a live connection, this is all that is needed, when Birst connect can run on a server continuously. The database can then be accessed in the Birst Admin module and the data is ready for utilization. For a database connection to read a table into Birst, a task is made where a known table is defined and selected to connect to. This task can then be scheduled to run, and Birst will gather the new data according to this schedule.

Figure 13. Birst Connect.

As soon as data connections are defined, the user can start visualizing charts. And once the user has created charts, they can be assembled in the dashboards component for view-

ing. With the dashboard component, the user can insert saved charts and resize and arrange them as needed. In addition, KPI indicators displaying a certain measurement value can be created. User selectable filters are also added at this stage, there are multiple options a developer can use. Filter types include but are not limited to checklist, calendar and slider. These display filters can apply on either all the charts on the dashboard or only the ones selected and can allow the user to see insightful views with last month or year to date values.

Once the dashboards are designed, they function as the user interface to explore data. The same Birst space for a boiler for example can be used for all boiler customers, the end user will see the same charts but with different data, only the data that the customer has access to is shown. In addition, different Birst spaces can be created for different applications and customers can be allowed to access only the applications they have access rights to.

3.5 Analytics

In the beginning the primary focus of a platform is to gather data in a central location. The strengths of utilizing a cloud location comes into play, when data from multiple sites and many different sources can be utilized in analytics.

In principle, there is nothing new about gathering process data, utilizing it for calculations and visualizing the results. These calculations have previously been restricted to a local site area with only the capabilities available locally or in some cases have connected data from multiple sites in limited capabilities. Even remote access has been in available to give access to support technicians from service companies to get some view of the situation onsite from a remote location. However, with the utilization of the cloud in an Industrial Internet platform these capabilities can be done on a much larger scale and combine data from sources that has been difficult before. This is where the Industrial Internet can bring value into analytics. [33]

As stated above, calculations done on site usually only have limited access to external data sources much less a competing power plants process data. By gathering this data into a central cloud storage, where the storage capacity easily scales with the data being added, a vast database of data from different sources are available for comparison.

As a security notice the disclosing of a customer's data to other customers is not acceptable, but a comparing analysis calculated based on each of these customers can provide valuable insights into their own process while keeping the source data anonymous. This way for example similar processes can be benchmarked and compared to other sites to notice potential improvement areas. Anonymity can be enforced by not having any defining information in the data that is being compared, for example data can be numbered in

a way that the customer only sees their site highlighted but has no knowledge whose the other sites are.

Another important capability provided by analytics and when a gathered dataset reaches a suitable timeframe, is prediction and other predictive analytics. With enough process data that contains different operation cycles a baseline behavior pattern or model can be calculated for the process, allowing the search for discrepancies in process operation to reduce breaks and downtimes and even predicting how a process should behave with certain values. These analytics function as a base for predictive maintenance, where process components lifetimes can be estimated by comparing their data to previous components datasets and lifetimes. These models become more accurate as more data from different situations is accumulated.

Analytics is one of the cornerstones of the Industrial Internet and finding ways to implement these more advanced analysis properties is an important step in the future.

3.6 Security

Industrial Internet security concerns can be divided into two categories, information technology (IT) security and operational technology (OT) security. According to Gartner, IT encompasses all the information processing technologies related to software, hardware, communications and services, while OT, a relatively new term, is the technology through which actual physical change in process and physical devices are detected and caused. [19][18]

A basic principle before moving an Industrial Internet solution into production is to have IT security in place. There are multiple standards to help achieve a good level of security and they encompass for example firewalls, multilevel security, update control and security policies. Firewalls control connections between computers and multilevel security allows for different types of access control for users and prevent access to unauthorized information. Through update control the latest security updates can be managed in devices, so that vulnerabilities are not long lived. Security policies help organize and document the rules to keep a company's network secure. In addition, an important part of security is user authentication and authorization. Through authentication, a user is verified to be the person that they claimed to be and through authorization the rules of what a user is allowed to do is determined, e.g. read only access or deletion rights. Risks of the lack of IT security include data loss, data theft and service disruptions, which can each cause significant economic consequences. [33]

As the Industrial Internet encompasses more and more Industrial systems, OT security is becoming more important. OT is needed to prevent unauthorized access to control the physical devices and processes that an industrial site contains. Risks of lacking OT security include physical damage to devices, process disruptions and danger to employees.

Some key considerations to take into account when planning OT security for Industrial Internet environments include: planning, mapping and defining the networks, limiting connections and only using two way connections if needed. It is important to know the topology of the network, e.g. the arrangement of devices connected to the network. Attaching new devices needs to be planned and executed so that the network stays secure and for example older devices that were never meant to be connected to public networks are not accessible by internet. Each new connection needs to be documented and only access from authorized addresses should be allowed between devices. Moreover, if the connection is only gathering data from a device, there is no need for a two way connection, which closes another possible entry point into the system. [33]

As a glimpse into the future, new types of security solutions can arise from a technology called blockchains. Blockchains are currently being utilized with virtual currencies, which is virtual currency that is not regulated by banks, as a method to confirm and trace transactions. The blockchain security is based on transactions that are authenticated in a built-in mechanism where every record is traceable back to its origin. Currently, however, it is unclear how the blockchain could be utilized outside of the world of virtual currency. [33]

4. POWER PLANT APPLICATION INTEGRATION INTO AWS

One of the objectives of this thesis was to test the power plant monitoring solution Valmet Energy Performance Center's integration with AWS and Birst. To test this capability the following design parameters were issued:

1. Data connections and data availability in the cloud needed to be solved
2. A BI tool to design dashboards and make the specified visualizations
3. The possibility to control data viewing and allow customer access

After learning the basics and understanding what the capabilities of Birst and the AWS cloud were, an Energy KPI monitoring solution demo was selected as the target of this test. This solution was to contain the necessary dashboards to monitor a power plants performance, superheater corrosion and measurement discrepancy highlighting. To be able to understand the capabilities and requirements of this, this chapter will go over what was done to get an Energy KPI Monitoring demo running on Birst utilizing the AWS cloud. The steps needed to achieve the design parameters specified earlier were defined as follows:

- Data transfer from the pilot site to the Valmet EDW
- Birst Space creation and configuration for the data
- Data visualization and charts
- Dashboard assembly and style consistency
- Development to production and allowing customer access.

This implementation chapter will go over each step of this process in more detail to explain how the Energy KPI demo dashboards were completed and what the capabilities of the VII platform are.

4.1 Data Transfer from Site to EDW

The first step to starting a new solution is to gather data. In this case, data transfer from the site to the cloud was needed. As a simplification to the data pipeline, the connection from site to data warehouse is as follows: Site server -> transfer server (SFTP) -> raw data storage (S3) -> data warehouse (Redshift). All the gathered sensor data was found on site from the automation system's historian.

To be able to differentiate different sensors form each other, a naming convention or tagging system was in place to facilitate the gathering of relevant data. In Valmet power plants, each sensors measurement is tagged with a tag name that derives itself by the Identification System for Power Stations (KKS) [28] system which gives each tag a

unique name which distinguishes it from other measurements. A single power plant contains tens of thousands of tags depending on what sensors are needed and what phenomena need to be measured. A list of a few hundred tags had been specified beforehand, which was used to gather the needed information to be sent to the cloud. This data contained measurements for temperatures, pressures and some calculated process values as well among other data.

At the beginning of the pipeline, on the site's historian database for this example case, the collected data was gathered into Comma Separated Values (CSV) files that contain three information columns: scan date and time, tag name and measurement value. The CSV files were named according to the naming convention:

```
Naming convetion:
siteid_businessline_lineid_source_type_datetime.csv
Example:
123456_energy_123_source1_datatype1_20170701130000.csv
```

It was decided to gather the data as one minute averages. The collecting was done utilizing an R-code script that gathers the necessary information from the local historian database. The tags that needed to be gathered were listed into a separate file and the timestamp from which gathering was to start was also defined. The script then reads the definitions, gathers the tag data and writes the information into a .csv that has the correct format that was specified above and named with the necessary naming convention. In addition, the script creates the control file (.ini) with the same name which contains information on the number of rows outputted and start time and end time. The script also compresses the data into a zip file and transfers completed files into the designated transfer folder, from where they are transferred off the site. The resulting data files were named and contained rows of data in the following format for each minute and measurement:

```
Names:
123456_energy_123_source1_datatype1_20170701130000.zip
123456_energy_123_source1_datatype1_20170701130000.ini

Data:
scan_dttm,tag_name,value
2017-07-01 13:00:00,12345678,25.12342367
2017-07-01 13:01:00,12345678,22.14512364
2017-07-01 13:02:00,12345678,26.14973245
```

An example data file containing 30 minutes of random data can be found in Appendix 1 to give a sense of the structure of the raw data. Before the script ends, the date and time in the definition file is updated to match the time up to which data was collected. This enables the scheduling of a continuous data collection, where new data is collected starting from the previous task's end.

After the data has been gathered into suitable files for transfer a method for secure transfer of the files from the customer site to the cloud is utilized. For this purpose, SSH File

Transfer Protocol (SFTP) was chosen. As defined by company behind Secure Shell (SSH): "SSH is a software package that enables secure system administration and file transfers over insecure networks, such as the Internet." [46] and SFTP is a file transfer protocol that makes it possible to transfer files securely over an SSH connection. [46]

With this method, a site specific SFTP account was created that allows the user to add files into the SFTP server running in the AWS cloud environment. Continuous data transfer was then configured to copy the collected data files from the designated transfer directory on the customer site an send those files to the target directory on the SFTP server using the SFTP credentials provided. The SFTP server acts as a gateway, to which a secure connection can be made for file transfer, without having access to the rest of the cloud platform.

On the SFTP server, the ini control file triggers the next step. In this part, it is important that the zip file with all the data is uploaded completely, before the ini file, so that the next step is not triggered before all the data is ready to be sent.

The next step for the data is the Amazon Simple Storage Service (S3), which is an internet storage, which can store and retrieve large amounts of data from anywhere on the web. The service is aimed to have the benefit of scaling to very large amounts of data. [11] For these reasons the S3 is used as the target for all the raw data that needs to be stored from the customer sites and other sources. There are multiple ways to add data to the S3 in addition to SFTP, but those are not utilized in this case example.

In the S3 the .zip data file arrives in the target bucket, which is the term used for S3 data directories. From there, the files are unzipped into csv files and in the process, some additions to the data are made. To each row, the source's site id and line id found in the filename are added for later identification and filtering of the data. In addition, the csv is renamed to include the target table name, where it will be transferred into the staging area of the EDW. After these modifications, the data filename and content is as follows:

```
Name:
123456_energy_123_source1_datatype1_20170701130000.process_data.csv

Data:
scan_dttm,tag_name,value,milled,lineid
2017-07-01 13:00:00,12345678,25.12342367,123456,123
2017-07-01 13:01:00,12345678,22.14512364,123456,123
2017-07-01 13:02:00,12345678,26.14973245,123456,123
```

One more step is required, before the data is ready to be moved into the data vault. The uncompressed and finalized data is then converted into JavaScript Object Notation (JSON) files. JSON is described as a lightweight text-based data interchange format that is on the one hand easy for humans to read and write and second easy for machines to parse and generate [31]. It is based on the ECMA-404 standard by ECMA International

[16]. This JSON file, contains the data in a format that can easily be written into the staging area by an automated process.

This automated process reads the JSON file contents and writes it's contents into the staging area and finally inserts data from the staging area into the data vault as a structured database for efficient storage and analysis. Redshift contains several layers including the temporary staging area, data vault and publish layer. When moving process data into redshift, it is first transferred into the staging area. In staging the data is organized into the correct column/row format with the correct column names for the data and the right values under each column. From staging the data is then moved into the data vault into satellites, links and hubs based on the data model that is configured. This enables the data to be linked to other data tables containing data from other sources. In the case of process data of which there is a continuous flow, the staging area is then emptied after each data dump.

Finally, the publish layer houses the tables and views that are available for access by Birst developers. A connection using Birst Connect is made to the table that contains the process data, after which Birst can be utilized to create visualizations and dashboards based on the data. These publish tables can be utilized to select certain snapshots of data from the data vault, combine data with master data or even make aggregations utilizing SQL scripts.

Once these steps were complete, a continuous data feed from the customer site was in place. Data was gathered every 10 minutes and sent forth into the cloud storage and ready to be accessed for visualization.

4.2   Birst Space Creation and Configuration

To begin developing the dashboards for the Energy KPI Monitoring solution a Birst space called VII Energy KPI was created as seen in Figure 8. This space functioned as the base for collecting data, permissions and visualizations to make a demo of the cloud platforms capabilities.

Because of the continuous nature of process data, a live connection to the data warehouse was essential in order to allow for near real time monitoring of selected process KPIs. Birst Connect is used to allow Birst to access databases and files in this fashion, so Birst Connect was set up to run on a separate Birst server on the cloud. There Birst Connect functions as a live connector to allow Birst to run queries into the Redshift data warehouse. Setting up Birst connect requires the spaces name, id and access credentials and the task can then be scheduled to run at certain intervals, which is the case for user credentials and access control.

Access control in Birst is implemented by a Role Based Access Control (RBAC), which is a collection of user and customer information tables running on a Master Data Service

(MDS) instance, which is a Microsoft SQL server add-on. The MDS is used as a master data repository, which stores primarily unchanging data. In RBAC user identifying information is connected to the users site, application access rights and data viewing rights. In the RBAC MDS a user's rights can be defined to only have access to certain applications, what data can be seen (for example certain sites or certain components) and the type of access (admin, developer or viewer). These role-based rules are kept up to date on a table in redshift, from where Birst spaces can check and control user security and access. Passwords are not stored in the MDS, only user information that enforce what a username can access and see.

Birst connect was configured to allow the live access of redshift and access the RBAC user roles table. After this connection was made, the changes in the Birst admin module were made. First the space settings were configured to utilize dynamic groups for access. Then the RBAC source that was connected to with Birst Connect was enabled and its primary keys were set in addition to defining is hierarchy. Moreover, with a short script a script based source was made to allow for the deletion on user rights where needed. Next these user groups were defined and their rights configured. For energy KPI the user groups presented in Figure 14 were made.

*Figure 14. Energy KPI user groups for access control.*

The two user groups defined were Developer and Viewer, where the developer has full access to all the spaces actions, but the viewer only has bookmarking and download options. That mean the viewer can only view dashboards, download data views and create bookmarks for personal use. The developer on the other hand can edit dashboards and

settings in addition to utilizing the visualizer and designer to create reports. When the RBAC user groups were in place, it was made possible to allow other users to view the space and work in progress and even allow other developers to help creating reports.

The second part of defining sources involved configuring the right data warehouse tables for live process data access. In the information delivery layer of redshift, publish, there was a view of the process data in the satellites. In addition, access to the site information was also needed to later provide data level filtering based on what a user can see. The two tables accessed are called:

```
src_fuel_diet_energy_process_data
d_site_compass
```

The first table contains tag name, date and measurement value in addition to site id and other discerning information. d_site_compass contains all site information, so that it can be connected to the src_ table through site id through it other information such as company name can also be accessed. These data sources and their connections can be seen in Figure 15.
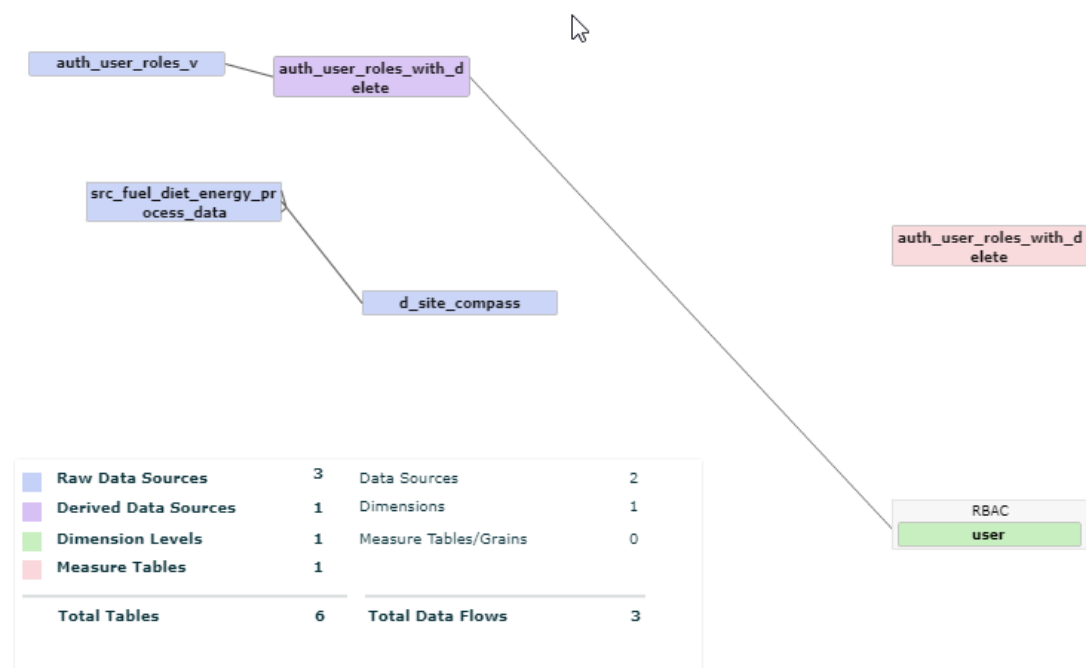
Figure 15. *Birst data sources connections.*

Once these connections were made the Birst space was ready to start utilizing the stored data for visualization. The next chapter explains the types of visualizations needed and how they were implemented.

4.3   Data Visualization and Dashboards Introduction

The primary tool for quickly creating trends, bar charts and even table views is the visualizer. Most of the needed figures were done using this tool. As a specification to what the Energy KPI demo needed to contain, the following dashboards related to power plant performance were specified:

- Boiler performance
- Superheater corrosion
- Operation point deviation

These three dashboards functioned as the base from which reports were designed. Each dashboard is explained in the following chapters as the visualizations are completed.

4.3.1   Boiler Performance

Starting with boiler performance, key process measurements to display were superheater temperature, boiler fuel input, steam output and a combustion diagram. Superheater temperatures were chosen to be able to follow the conditions at the superheaters and make sure that steam is being reheated properly. Fuel input and steam output are direct performance and the boilers efficiency can be calculated from their relation. The combustions diagrams purpose is to show how the boiler is operating compared to its specifications.

The combustion diagram shows the relation between the fuel mix and steam output of the boiler. In the diagram, the operating window is determined by the boiler operating specifications with different fuel types. With high energy content fuel, less fuel flow is needed to reach full operating power. Using lower energy content fuel, for example wet biomass, a larger fuel flow is needed to reach full boiler load in terms of power output. With the right process parameters, it is also possible to exceed the full boiler load specification, and achieve a higher than normal power output. The chart that was used as a base for this can be found in Figure 16.
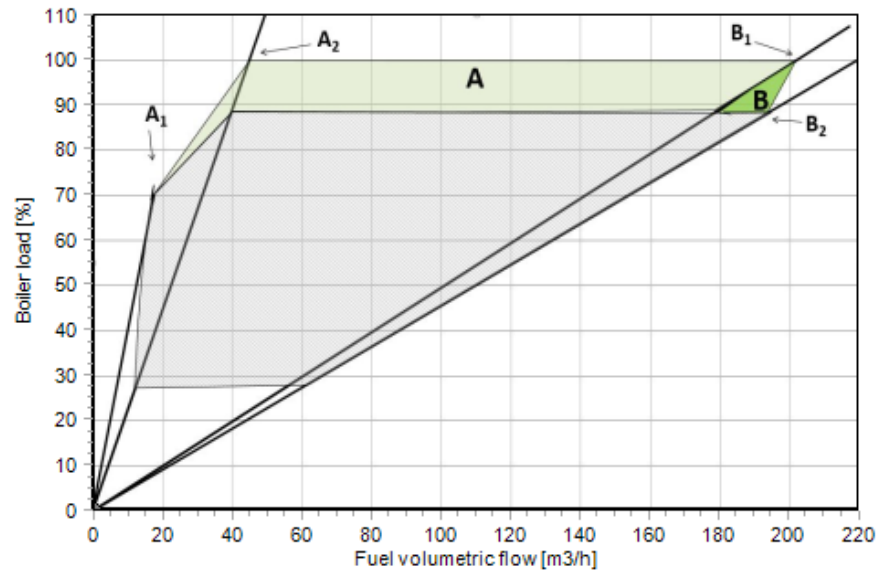
Figure 16. *Combustion diagram operating window.*

The operating window consists of the lines A1, A2, B1 and B2. These lines represent a certain fuel types fuel flow to boiler load ratio. Combined with the boiler's operating specs: full load 100%, 89% load and minimum load 28% areas A and B are created, which represent the specified boiler operating area at normal load.

To create this diagram with Birst, it was decided to draw the outer lines only to encompass the area of operation. Since this chart was complex, the visualizer did not support all the features needed. For that reason, the chart was drawn utilizing the older designer, where all the right customization features were available. To create a drawing of the operating window, it was first needed to define the values for each border line. To represent boiler load, boiler steam output was used on the y-axis and fuel flow on the x-axis was represented by the fuel mix input flow. The boiler's technical specifications reported its steam output capacity as 150MW, so that a logical choice for 100% load. From this figure the values for 89% load and 28% load could be calculated. The maximum fuel flow value was set as 200 m³/h as in the original chart. Utilizing these values, drawing points for each line of the operating were generated, based on which the visualization was done, as shown in Figure 17. This was simply done by creating a separate table where all the different point's coordinates were gathered and generated by the designers TRANSFORM function. On top of this operating window, the data value points were set to appear, by showing the average value correlation of fuel input and steam output for every hour and filtered to show one day at a time.
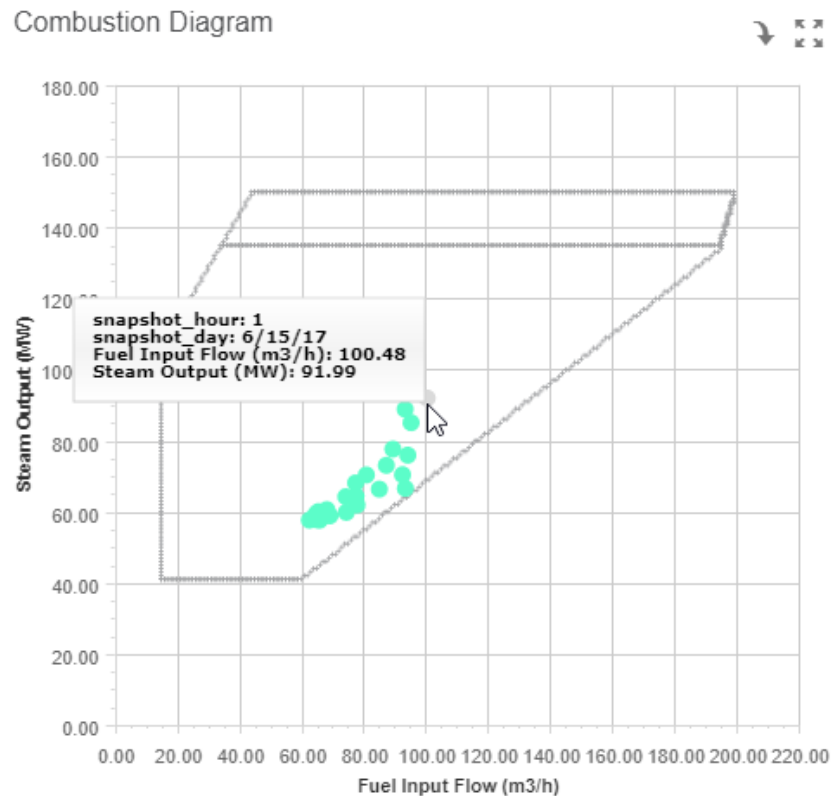
Figure 17. Combustion diagram in Birst.

The next step was to create a fuel power and steam power comparison trend with the newer visualizer tool. The y-axis was specified by simply selecting and dragging two measures to the measure field of the visualizer, where the measure in both cases was tag_value, which contains all the data for every measurement. To change this to only show one certain tag, '3LBA10DU902:av' for example, which is the tag specified for steam power, a little modification to the measure's query was added, as seen in the code below.

```
[Avg: tag_value] where [src_fuel_diet_energy_process_data.tag_cus-
tomer_name] = '3LBA10DU902:av'
```

Changing the aggregation from sum to average changed the outputted value into an average across all tag values in the source table. The where clause in the code above makes the query only select values that have the tag name column with the specified name. This modified query now returns an average value for all measurements with the tag name '3LBA10DU902:av'. The same modification was done to the fuel power measure, but with a different tag name. At this point, there were two average values for two different tags.

To specify the x-axis, the category hour was selected from the attributes list and inserted into the visualizer's category field. This reorganized the two average values over all time into average values for each hour. And once a filter for date was added, the chart displays

the hourly average value for both steam power and fuel power over the course of the selected day filter as shown in Figure 18.
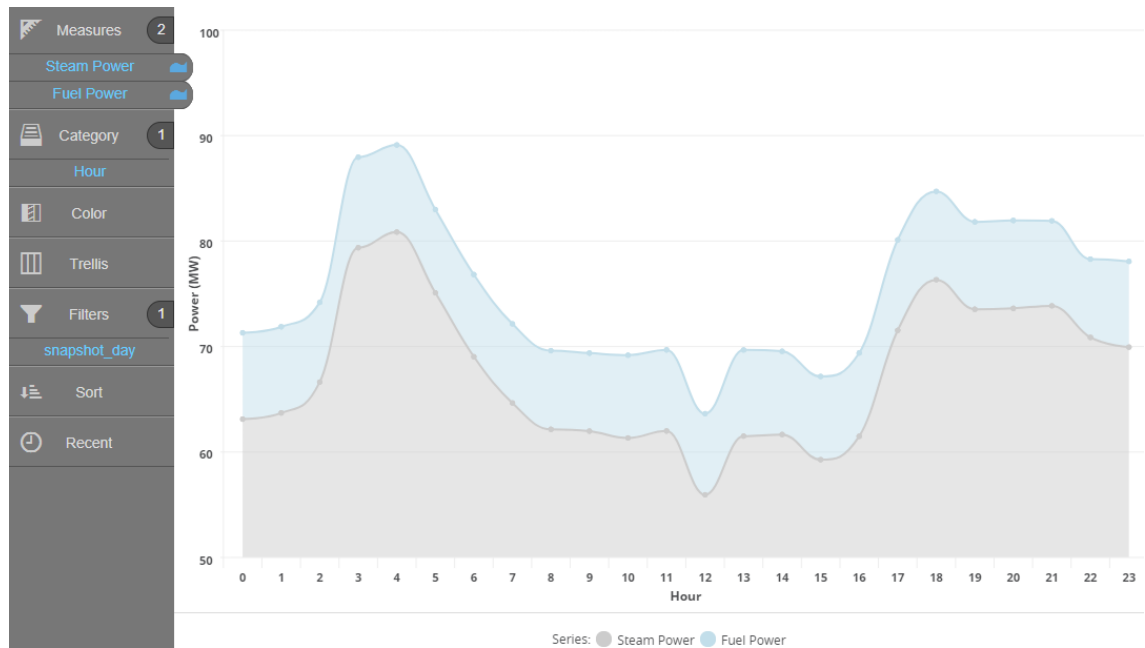
***Figure 18.** Visualizer view for boiler performance chart creation.*

This chart was then saved as a report for later use in the dashboards. To create another similar chart for steam temperature at the first and second superheater, the newly created chart was used as a template. Simply modifying the correct tag name for each tag_value measure draws the same chart with the data for the defined tags to be saved as a finished report for later utilization.

4.3.2 Superheater Corrosion

The superheater corrosion index is a calculated percentage value, that is used to indicate the level of corrosion risk in the power plants superheaters. Especially when using renewable bio-fuels, conditions at the a superheater's pipes can cause corrosion that can shorten the lifespan significantly. The main causes of corrosion lie in the fuel's Sulphur and Chlorine content.

To visualize the corrosion risk development over time a trend line was the logical selection. In addition, a color indicator was added to the chart to classify when the risk is high, medium and low. According to the specs, an acceptable and normal risk is under 60%. A percentage of 60%-70% indicates an elevated risk and over 70% indicates a high risk of corrosion. A natural addition to explain the effect of risk, would have been the addition of superheater pipe diameter measurements and lifetime analysis, which are calculated on site, however, that data was not accessible at the time.

The chart was created in a similar fashion as the steam and fuel power chart in chapter 4.3.1, but as there only was one tag that appears in this graph. Instead of modifying each measure query, a tag_customer_name filter was added to only select values with the correct tag name for the two measures needed for the chart. To have two types of visualization on this one chart, one of the measures was selected to be a trendline and the other a bar chart. This was done by simply selecting the measure and setting the chart type for each measure from an selection of different chart types. This way the aggregation could also be changed to average, maximum or minimum etc., but the average aggreation was decided to provide the necessary information. Now Birst draws two chart components in the same chart as can be seen in Figure 19.
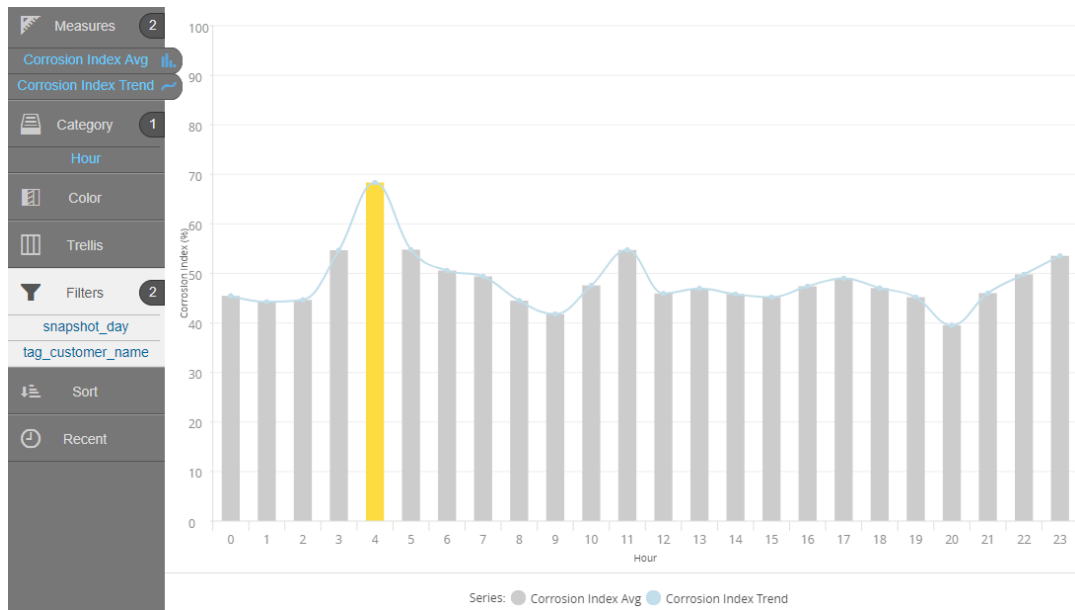
Figure 19. Corrosion index trend and bar indicator.

The final addition to this chart was the addition of a conditional formatting rule from chart options. With this setting an IF conditional expression was defined as:

```
([Avg:  tag_value]  where  [src_fuel_diet_energy_process_data.tag_cus-
tomer_name] = '3COR20CX805') >= 70; color: #E70033
```

This colors the all the bars that have a measurement value over 70 with the color code #E70033, which is red for high risk. Another conditional is set for values above 60 to be colored yellow as a warning for medium risk. Otherwise the chart colors are left to the unified color palette to determine. After these additions, the chart was ready for saving.

4.3.3 Operation Point Deviation

The final dashboard type was for exploring the possibility of highlighting deviations in the monitored data. For this, four measurements were selected to be used in the testing: steam temperature, steam temperature after superheater, flue gas temperature and flue gas

O^2. First, a limit value for each of these measurements was selected to highlight the possibility. In practice, this limit value can be found in the Boiler's technical specifications and would be utilized, but for this demonstration a placeholder value was used. These limit values were defined in the Birst admin component by creating a custom measure that had a constant value. From here the measure can be edited in the future to change the limits. As a note for future improvement, these limits could also be dynamic and change according to the stage of the process, for example low load, high load or startup. All it would need is a new table defined into the EDW and something to identify the process phase.

This limit and two measures were added to be drawn in visualizer. The category was kept as hourly, so that a period of one day could easily be monitored. Again, a universal tag filter was added so that the correct values were displayed and one measure was selected to be a trendline average, while the other was a bar chart. As there are many measurements available during the course of one hour with minute averaged data, the bar chart was selected to show the maximum value of the measurement for each hour. This way one can see if the value has been above the limit for even one minute during an hour.

Now there were three chart components drawn in one chart: the measurement limit constant, average value trend and maximum value bar chart. The same way as before, a conditional formatting expression was added to color values above the limit of 470 with the color red. The result is seen in Figure 20.
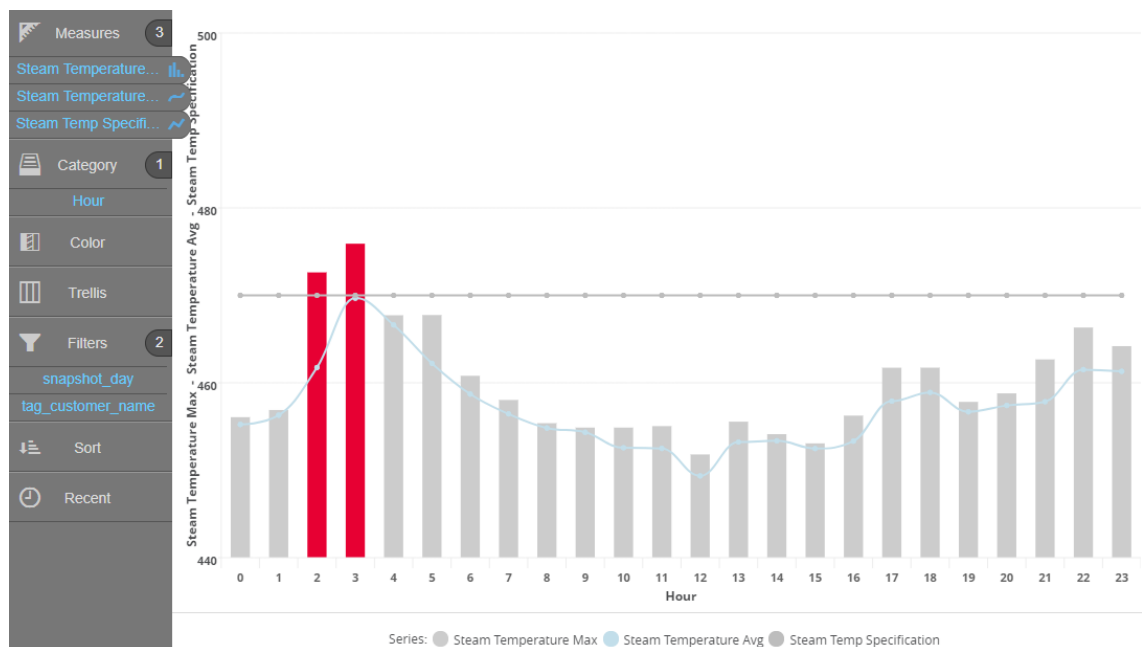
Figure 20. *Steam temperature after secondary superheater chart in visualizer.*

Now that this one chart was ready, it was saved and it was time to create charts for the three other selections. The first chart was again used as a template, in which the tag_customer_name filter was changed to correspond with the right data. In addition, the

tag_name and limit value in the conditional filter needed to be changed to correspond with the right limit. This made the process of creating additional charts fast and simple.

In the future, adding warnings for upper and lower limits including alarm limits is possible using the same method as above. There is also the potential to have alarms for unusual process behavior to send notifications by email or other form of messaging to notify operators or managers of dangerous process parameters.

4.4   Dashboards Assembly and Style

Now that the basic visualization reports have been made the dashboards can be assembled. A few common notices in dashboard design include how to display data in a way that is clear and quick to decipher. Some basic pointers in dashboard design include the need to be for the displayed data to be timely, accurate, valuable and actionable in order to be effective. This means the end-user should not have to wait for too long before the information is visible, the data does not have errors or missing pieces, it is meaningful in some way that brings value and it is thorough enough to be able to base decisions on. These pointers guide the type of data that is displayed on the dashboards. [13]

For the visual side of the dashboard, the overall display of the data needs to be simple to understand and decipher very quickly, clear as not to cause confusion and concise so as not to contain too much information on one view. So, once the data has been selected, it is divided into suitable overview and drill down information in a way that it can be displayed in a visually appealing way and does not provide too much in a single glance. Another important factor for the visual display is the actual layout of the dashboard and how to size, place and order the different chart elements, KPIs and filters on the dashboard. Two factors that can be used as a base for deciding dashboard layout is the Golden Ratio [38] and the F-shaped pattern in which a human user reads web content [17]. The Golden Ratio is a mathematical term which has been used as a most visually appealing ratio in artwork throughout history. Consider a line AB, with the segments AC and CB. If the ratio of the length of AC to that of CB is the same as the ratio of AB to AC, then the line AB is cut in a golden ratio. This ratio can be utilized in deciding the layout of a dashboard. [38]

In an eye tracking study conducted by the Nielsen Norman Group, it was discovered that the users read web content in an F-shaped pattern, which most commonly contained two horizontal movements and one vertical movement concentrated on the left side of the content. Reading progressed so that the eye tracking patterns created an approximate F-shaped area where content is read and other areas are left with less attention as is displayed in Figure 21. In this figure, red represents the areas of most eye activity and the color scales from yellow to blue for areas of less activity. Areas that the eye did not focus on have no color indicator. Translating these findings into dashboard design, the most important display of information should be on the top left side. [17]

Figure 21. *Human eye tracking study indicating an F-Shaped pattern when a user is reading web content. [17]*

The dashboards in this case were assembled utilizing these parameters for making visually effective dashboards. The actual assembly of the dashboards in Birst is simple. Selecting the dashboards module allows the user to create new empty dashboards at the click of a button. These empty canvases can then be renamed and the charts created in chapter 4.3 can simply be inserted into the dashboard, resized and dragged to the position that is wanted. In addition, on this dashboard creation screen other components like single KPI indicators, text, images filters can be added into the dashboards as shown in Figure 22.
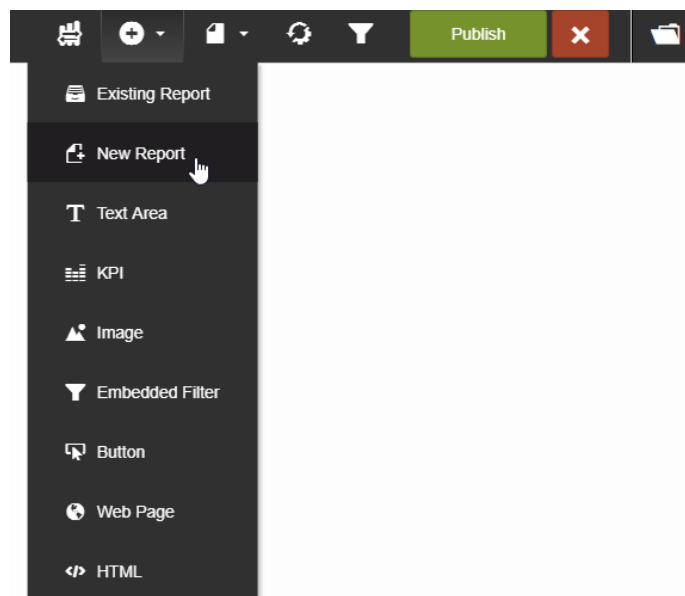
Figure 22. *Adding reports to a dashboard.*

The first dashboard, performance overview, was defined to display key performance information, such as a combustion diagram and steam and fuel power consumption to allow a user to see how performance has been during the course of a day. The charts created in

chapter 4.3.1 are inserted into the dashboard and resized as seen in Figure 23. To get some understanding of the rate of production, a local temperature indicator is added as output usually varies by temperature. For data filtering, a date selection is created, as all the dashboards will be filtered based on the selected date. This provides the user with a day view of any date that is selected.
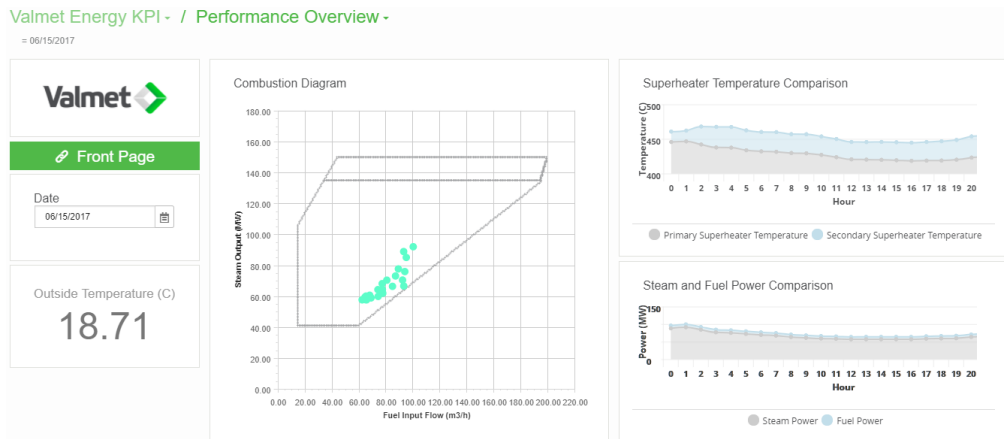
Figure 23. Performance Overview.

The next was the corrosion overview dashboard, meant to display the current risk of superheater corrosion and link that into the fuel composition. In addition to placing the charts created in chapter 4.3.2, numerical risk indicators to display the corrosion risk were added to the top of the dashboard, where the most important indicators should be. Below them is a trend indicating how the corrosion risk has developed during the day. As additional information on the bottom of the dashboards, the main corrosion components found in the fuel are displayed as numbers. These charts are then filtered by a date selector to display information over the course of a day. The finished dashboard can be seen in Figure 24.
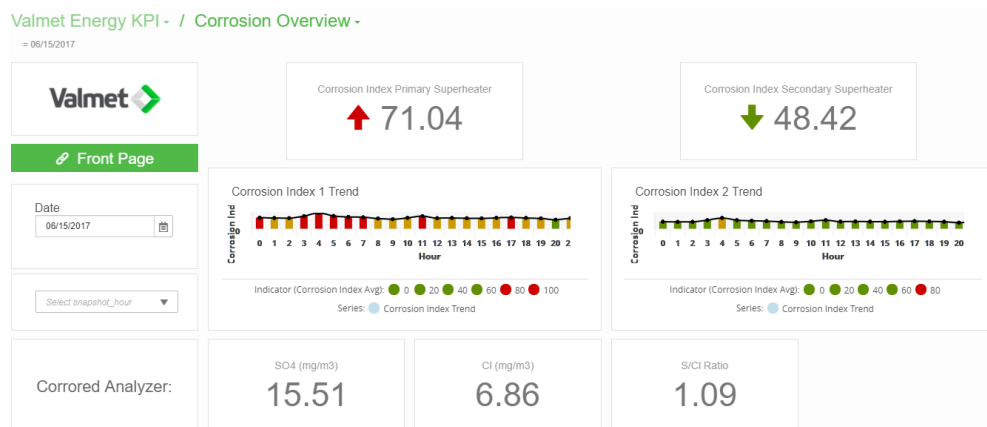
Figure 24. Corrosion Overview.

The third and final dashboard that was specified was an operation point deviation (OPD) dashboard, to demonstrate how discrepancies in incoming data can be seen at a glance. The four main data-areas of interest selected were live steam temperature, reheated steam temperature, flue gas temperature and flue gas O2 content. Steam temperature is followed to make sure it stays within operating limits and to prevent damage suboptimal temperatures might cause. Flue gas temperature should also stay under a certain limit, otherwise heat energy is lost as it leaves the power plant with the flue gas. The flue gas oxygen content on the other hand tells of how combustion has performed, whether there has been too much oxygen or even too little for optimal conditions. These four charts where places in two rows on the OPD dashboard. As with the other dashboards, a date filter to show daily values was added to restrict the data show. Demonstrated in Figure 25, the red indicators for values above a certain limit are quite easily seen at a quick glance.
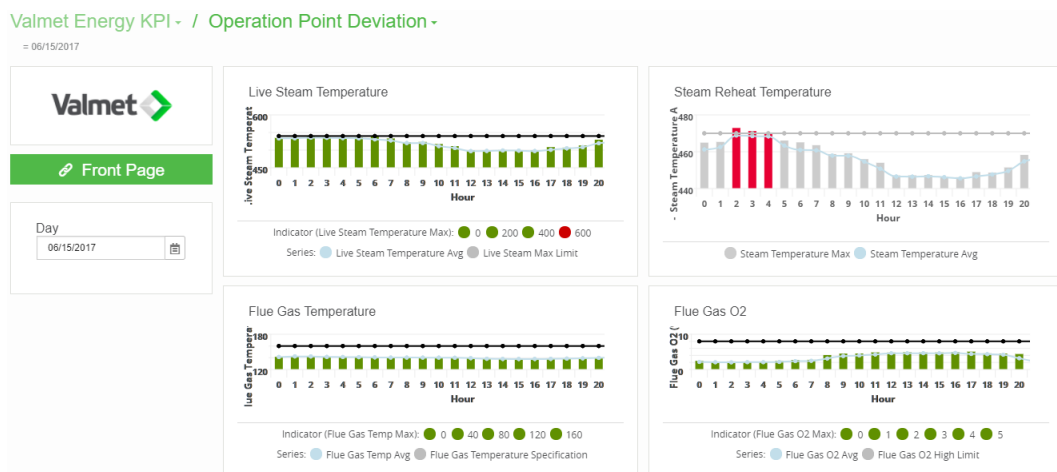
Figure 25. *Operation Point Deviation.*

Another feature of OPD, the ability to drill down to a selected hour was added by creating a supportive dashboard. So, for example, the user can select a value in red from a chart and be directed to a drilled in view of the hour with additional information. This additional view has a chart that shows performance for each minute during that hour and it is easier to discern the problematic moment as shown in Figure 26.
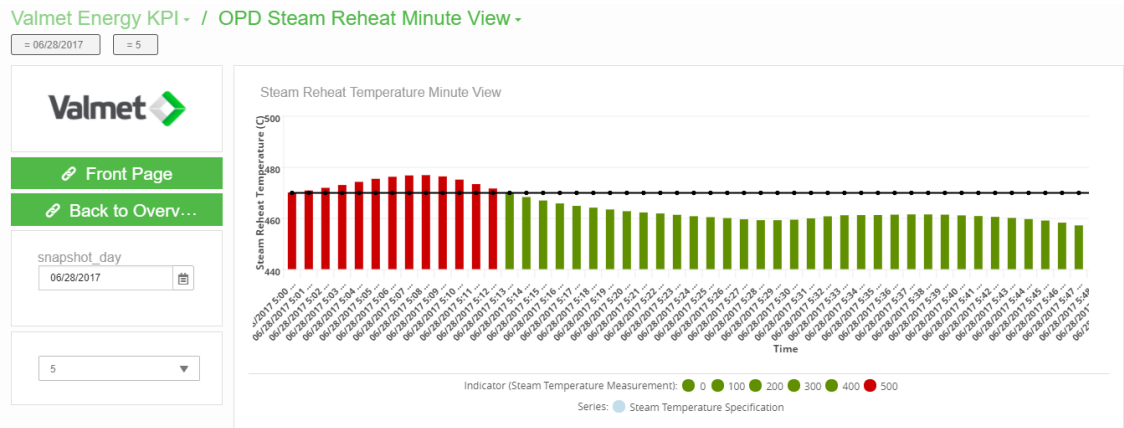
Figure 26. *Operation Point Deviation drill down view.*

After this the dashboards are complete. New data arriving in the cloud is automatically displayed in the dashboards when the appropriate date is selected. In addition, history data can be viewed by selecting an earlier date.

4.5   Production Environment and Customer Access

Once dashboards development is complete and they are ready to be released for the customers, they are moved into the production environment. The VII platform consists of three environments that exist side by side and contain all the same components. The first is the development environment (DEV), where developers from Valmet can create new content, explore data and make changes as necessary to develop new content. In this environment access rights and data control is easy to request when needed. The other two environments are test (TEST) and production (PROD), where user and developer access is much more restricted. TEST is a duplicate of PROD to allow or testing of new developments before being moved into production. Only content that works and has been tested is allowed into production, where customers and end-users access the content. These environments make development easier for developers as they have the required access rights to do what is needed in DEV and safeguards the customer experience by limiting the risk of untested changes being available in production. Another strength of having multiple environments means that data is backed up in multiple separate environments, so that the complete failure of one does not destroy all the data gathered.

Moving a solution into production requires the duplication of the data pipeline utilized in DEV, so that data can be found in the new environment. As of the time of writing the TEST environment is being deployed, so it has not been utilized during the course of this thesis, but in the future, all developments will first be tested before production.

To move the Energy KPI data pipeline into production all the necessary definitions were done in AWS and data from the site was uploaded to the PROD redshift. Once data was

in the right places, the Birst space was deployed into production. This was done by creating a space template that stored all the visualizations, access rights and dashboards from the DEV environment. This template is then manually moved into the production environment Birst server where it was used to create the same space in production. After the template was loaded, the space was ready to be accessed in production. All that was needed was to add the same database connections as used in DEV so that the space accesses data on the PROD redshift. After that the dashboards functioned correctly using the same queries as in DEV.

This production space will be what customers access to see dashboards. In production, the user only has viewer rights to see the spaces dashboards. No editing can be done with visualizer and admin modules as they are not visible. The visual appearance of the dashboards is the same as in DEV, only without edit buttons.

Now that the space was in production and data connections were functioning, Valmet end-users can access the content through Birst if they have applied for the necessary rights. For customers and other outside access, a universal portal based on Salesforce is used. Salesforce acts a one stop web portal, through which a customer can access all the VII content that has been agreed on. For accessing a specific solution's dashboards, the user can see tiles for each application, and by selecting one of the tiles, the user is directed to the spaces main dashboard in Birst production. From here the user can explore the solutions dashboards and see the data that has been defined for him in RBAC.

To create a customer account, it is simply requested to create a username to which solution roles can be added. Once created in Salesforce, the new account is set with roles regarding data access and solutions roles in RBAC to connect the salesforce account into the Birst data authorization. After these steps the new credentials should have access to the solutions, when logging into Salesforce.

4.6   VII Security

An important part of the platform in security. The VII platform security is handled in accordance with Valmet security policy. In the Development environment, access is only possible through the Valmet internal network. Authentication is based on the internal user authentication. Roles or authorizations are assigned to each user as either a viewer with only viewing rights, developer with modifying and creation rights and administer with all rights. Moreover, additional data security and authorization in defined through RBAC, where each user is defined the rights to see certain data and access certain solutions.

On the production side, security covers customers, Valmet developers and the data pipeline. For dashboards, which need to be accessible through the internet by customers, authentication is done through salesforce, with a two-step authentication, where, in addition to logging in with their credentials, the user needs to input a code from their mobile phone

authenticator application. This is used to authenticate that the user is who he says is he is. Authorization is also more restricted in the productions side, as customers by default do not have any other rights except viewing. RBAC roles also define the data and solutions a customer can access. Moreover, Valmet developers also have very restricted rights, to prevent any disruptions to the customer experience. Database modifications are made through a promotions system, where certain approvers check changes before they are deployed. Additionally, developers do not automatically see the data they need, the authorization for data in the production side is done by product owners.

The data pipeline is currently only a one-way connection, where data is sent from the site. This protects the site from risks that its systems are compromised from this data pipeline. The actual data in this practical case is sent securely by SFTP, with authentications in place for each site. From the SFTP server onwards, the data transfer is done through internal cloud connections and AWS security.

In summary, the Valmet Industrial Internet platform includes the necessary security procedures explored in chapter 3.6. Authentications, authorizations and network security is in place to make the VII platform as secure as possible.

4.7 Development Summary

This concludes the integration of a power plant application into the VII platform. The three design parameters defined earlier have all been covered and found that the platform is at a mature enough stage so that all the necessary components to create a functioning product exist i.e. we have data gathering capabilities, a BI tool to design dashboards and the possibility to control data and customer access to the solutions. These first dashboards have been targeted at production managers and other decision makers in the customer organization so that they can more easily access data and use it as a base for business and process decisions.

The performance dashboards allow the customer or Valmet experts to easily access a site's data to monitor performance and new data is populated into the dashboards automatically. The template made here can be used for other boilers as well without the need to build new dashboards for each site, which reduces the workload of scaling this service in the future. The corrosion dashboards allow production managers to follow the corrosion risk to the superheaters and use it to support decisions to be made in, for example, selecting fuel suppliers. Finally, the Operation Point Deviation views let users visually inspect key measurements and analyze the timeframes of any alarming readings. This allows them to take action to figure out what went wrong and why to prevent this from happening in the future or how a process could be more optimal in the future. All these features are securely accessible by the customer over the internet and from any device.

5. BUSINESS OPPORTUNITIES

Now that the practical integration of a solution with the VII platform has been examined, this chapter with go over how these solutions can be tied to business and what types of report opportunities exist for the future. The idea of packaging VII solutions with boiler deliveries is also studied based on the current capabilities of the platform in addition to presenting alternatives to benefit from the VII platform in other business areas.

Some competitors are working on deploying very large platforms to encompass multiple industry segments in addition to energy production. The competition could go in the way of one platform surpassing the others and grabbing a large market share before the others, or another possibility could be that customers choose the platform that is offered by their industrial supplier. If the big open platforms move towards the application store approach that we see today in mobile phones, a big challenge for single applications could be to differentiate themselves from the masses, there might not even be clear winners in terms of features, but just some applications that become popular. As applications update and improve, the stream of different solutions can also become difficult to sift through.

A strong point for Valmet in this situation is to focus on its own products and process knowledge strengths to offer their customers apps designed specifically for either Valmet's own products or even competitors' products in its core business areas. By providing their customer the simple and efficient platform and focusing on what brings the most value to customers it can be possible at this point to penetrate the market where no-one has much foothold. However, as a downside a smaller closed platform might not have the chance to pursue every solution that a large open platform can provide, but it can combat that weakness by offering a focused quality product.

5.1   Opportunities for Report Improvements

While working on the Energy KPI dashboards, opportunities for future improvements and other ideas were identified. They have been gathered in this chapter to have record of them for future reference.

Currently the landing dashboard of the Energy KPI could do with improving. It was not displayed here as it was just a simple portal to access the other dashboards and displayed the basic information of the only boiler currently connected to dashboards. In the future, this need to be reworked when multiple boilers and customers have access. An overview should show KPI indicators for the most important figures in each other dashboard, through which a drill down to the separate dashboard can be made when a user wants more information.

In the performance dashboards, additional information related to electricity and heat production could be added. In addition, information on the running times, startup times could be available here. The combustion diagram still need some visual tweaking to make it more clear and more descriptive. Otherwise connecting performance with saving and costs would be another avenue to explore.

Currently the superheater corrosion dashboards only display the risks of corrosion at any given moment. An improvement to this would be to add actual thickness measurements, lifetime estimates and corrosion rates, as they are already calculated and available from the sites database. In addition, fuel compositions effect on corrosion could be explored more thoroughly through visualizations.

Operation point deviation has many possibilities to implement. A process tree functionality, where from the main page, a user could drill down as deep as needed to find out the roots of any problem. Moreover, displays of the count or time of measurement deviations would provide an additional layer of information to the analysis. Finally, a needed feature is alarms from faulty or measurements over limits. This would allow notifications to be sent when dangerous values arrive to the cloud and would make it possible to notice and react to problems in a more customer oriented way.

A much-needed addition to the Energy KPI dashboard collection is an emissions dashboard. All power plants need to monitor their emissions and keep them under the guidelines provided by authorities. Since all the emission calculations are available on-site, they can also be added to the cloud and visualized in dashboards. This would give customers a new and simple way to follow their own emissions and make the necessary changes to stay in the right values. The basic emission components SO2, NOx, CO and particles could be visualized in trends and displayed with additional analysis like number of high emission records and minutes above the limit per time day for example.

Energy efficiency is another item that would benefit from its own dashboards. There are many parts of energy production, where losses appear and leave room for optimization. For example, the plants own power usage, which is directly removed from sold electricity could be monitored. Leaks and pump power consumption are also possibilities, through which discrepancies can be noticed and fixed if noticed early.

These are all possibilities that should be explored in the future development of the Energy KPI solution and would improve the quality and relevancy of the dashboards.

5.2   VII Valmet Business Integration

The Valmet Industrial Internet platform offers new types of possibilities to benefit from data analysis and visualization. Two basic cases where the VII platform could be utilized

to augment Valmet's current business include improving current physical products and improving service offerings.

On the service side VII opens up possibilities for agreements that compliment Valmet's current offerings. Monthly based agreements can be more economical than delivering physical products, as they provide steady income that is not as heavily tied to the customer's investment capabilities and the current economic trend. An example of such a possibility is a service or performance agreement, where a contract to improve performance is sold to the customer. In this agreement, a certain level of performance can be agreed upon and when it is reached the customer pays for the agreed sum. Monitoring the agreed upon key performance indicators is integral and through them it can be verified if the targets have been met. These types of agreements can be sold as parts of other service contracts.

Examples of possible agreements for power plants includes performance, e.g. electricity or heat production improvements, asset uptime performance for maintenance, remote support for Valmet's professionals to notice discrepancies in the customer's processes before they become a problem. In the future, it could even be possible to sell production instead of plants. For example, instead of buying a power plant, a customer might want to buy 100MW of electricity as a service, which could then be provided to the customer without the customer having to worry about selecting the right products and solutions and managing the facility.

The deliveries of physical products can also be augmented by these Industrial Internet capabilities. Having monitoring capabilities in every delivered device allows for the close monitoring of the product to make sure it works and fulfils the contractual agreements. In addition, the data feeds from the products can be utilized right from the start to improve future product deliveries and speed up deployment. Data gathering also allows Valmet to enhance its reference portfolio used in marketing and the sales of further new products.

To summarize, there are at least two ways the VII platform could be simple to implement in augmenting current offerings. One is by offering new types of agreements that synergize well with the VII platform and provide clear value to the customer. The other is to improve the sales and support of current physical products, so that both Valmet and the customer can benefit from the data by making the fulfilment of contracts easier and making future deliveries more efficient.

5.3   Valmet Boiler Readiness for VII

One of the considerations to keep in mind as the thesis was written was the ability to package VII with each boiler sold by Valmet. The basic requirements for connecting a boiler or power plant to the VII cloud platform could be summarized as:

1. Data sending device and the agreement to send the customer's data over the internet
2. The necessary sensor measurements utilized in the dashboards are present
3. The platforms readiness to accept new plants

The first condition that needs to be met is that the customer must agree to sending their data over the Internet. This step can be hard for some customer's in the very conservative power plant industry. Luckily there are open minded customers, with whom it has been possible to take the first steps of developing an Industrial Internet platform. With these first pilot customers, it is hopefully possible to demonstrate enough value of VII applications so that it becomes easier to persuade new customers of the benefits and security measures in place to keep their data safe. If enough value can be demonstrated it should not be too much of a problem to package with each new boiler, which in turn enhances the analysis possibilities of the platform as more data is accumulated. In any case, the ownership of the data needs to be agreed upon with the customers as well.

The second step states that the necessary measurements for the dashboards need to be in place for the user to be able to see the data on the dashboards. In all new boilers, the required measurements are in place, so no new sensors should be needed to install as everything is usually part of the automation system. However, some of the displayed data is based on calculations that might not currently be done at every customer's site. If this is the case, then those calculations either need to be packaged with new boilers, calculated on the cloud or the customer needs to have a limit on which dashboards are visible based on the available data from the boiler. All are currently possible implement, however the feasibility of each alternative needs to be explored before decision.

The third and final criteria would be the VII platform's readiness to accept every new boiler's data and allow new customers to the interface. Based on the components that are completed for the platform and the first customer accesses ready to be deployed, this should not be a limiting factor. Of course, before scaling to multiple boilers it should be recommended to test all access rights, data pipelines and interfaces with a limited group first, which if proved feasible should then be able to be scaled as needed.

In summary, the VII has reached a stage that testing integration as a package with all future Valmet boiler's is reasonable to start considering. With the first few additions it should become clear if there are still any issues to be ironed out and scaling can then be done.

6. CONCLUSION

This conclusion will go over the findings of the Industrial Internet market analysis and provide a summary of the current situation. Moreover, the development achievements and potential benefits of the Industrial Internet will be gathered and summarized in chapter 6.2.

6.1 Market Competition Conclusion

Currently there are multiple platforms in all kinds of shapes and sizes in development and even deployed to the market. Industrial customers are also starting to take interest in Industrial Internet solutions and though some are ready to take advantage of the new solutions, challenges still need to be overcome before the Industrial Internet's value is clear for new customers.

Based on all the public marketing materials available, Valmet's large competitors are all working on Industrial Internet solutions for power plants and the marketed offerings are mostly similar. However, as the state of readiness was explored for each competitor, it was found that all the industrial internet offerings are still in an incomplete state. There has not been any evidence of large-scale deployments recorded and any applications that are ready to be offered do not have very advanced scalable capabilities.

It is clear that the market is continuing in the direction of digitalization and the utilization of IoT capabilities and it is a subject that is of interest to customers. As development continues it can be feasible to focus on completing such features that bring enough improvement to be considered minimum viable products to ease development and act as stepping stones towards future developments. The first to the finish line with working products and advanced features have the possibility to establish a strong position on the market

6.2 Benefits Conclusion

In this thesis, an application to monitor the real-time process from a power plant was completed. The current dashboards target managerial staff, who can then easily access data to drive their business decisions and have a clearer picture what is happening in their process. The whole pipeline from data transfer to visualization to customer access has been completed and a way towards realizing the potential benefits of the Industrial Internet that were explored in chapter 2.2 is now possible. These benefits can be summarized into the following categories:

- New and Improved products and services

- The availability of data to a wider audience
- Improving deployment costs and scalability.
- Analytics possibilities and benchmarking

In a business point of view a clear benefit of utilizing the Industrial Internet is the possibility to create new products and services in addition to improving a business's current offerings. These new products and improvements lead to new forms of revenue and have the possibility to increase sales if a product offers new types of capabilities a competitor does not have. Examples of such new offerings include General Electric's digital twin technology, that is only made possible by the real-time data gathering of multiple turbines into a centralized database. As explained in 2.3.1, it allows the comparison of performance and optimization based on large sets of data from multiple devices. Additionally, in chapter 2.3.6, the predictive maintenance service from MaintPartner is also made possible by making a model from a large historical dataset of a process to compare to real-time measurement data. This allows the automated analysis of any unexpected variables that might cause service breaks.

Another benefit, that is one of the fundamental goals behind the Industrial Internet, is the availability of data to a much wider audience than before. Modern BI tools offer simple data access and visualization capabilities that allow users with minimal data science experience to explore data. Moreover, experts in many industrial fields should have an easier time to test and follow up on any process changes and products they make. For end customers, data can be delivered in a form that allows them to understand how a power plant is doing even though they might not have a degree in energy engineering. All these points help people in different parts of organizations do their work more efficiently and better understand what they are working with.

Moving into product development and running industrial complexes, Industrial Internet capabilities can offer shorter deployment times of new solutions as more data of previous is available and more easily accessible. Cloud storage offers the flexible scalability needed to be able to quickly react to new needs and expand business as needed without requiring large capital investments to build server capacity.

Finally, the new ways to collect data from many different sources into a central storage space allows the development of analytics and benchmarking that has not been easy or possible before. As computing capacity can flexibly be scaled in a cloud environment, even very large amounts of data can be analyzed in reasonable timeframes that can provide new insight such as predictions and more optimal process parameters. Data streams into a cloud storage from multiple different sources allow the comparison and connection of data that has not been possible on a single customer site scale. This makes it possible to compare the performance of multiple different sites and optimize whole networks of, for example, power plants.

These four categories summarize the benefits of the Industrial Internet and utilizing its different components. Some of these benefits are realizable today while others are still part of the potential that we are heading towards today. Some of the possibilities to realize more of these benefits are presented in the next chapter, where suggestions for future research are presented.

7. SUGGESTIONS FOR FURTHER WORK

There is currently a lot of potential that has yet to be realized in the Industrial Internet space. As different platforms are starting to take shape it is becoming possible to start heading towards the more advanced features and applications that can bring real value.

Now that data gathering is being ramped up, the near future focus in business should be on finalizing minimum viable products to have something concrete to offer. Moreover, moving towards more advanced analytic capabilities such as prediction and statistical analysis should lead to new applicable solutions in optimization and maintenance. These advanced analytics capabilities and how they could be utilized at scale are subjects that would be the logical next step to research after this thesis.

Additionally, there are benefits that could be gained from exploring the possibilities brought by combining Industrial Internet capabilities with other new technologies such as artificial intelligence, VR, blockchains and 3D printing, where three dimensional objects could be printed as spare parts. Some of these technologies have clear synergies with the Industrial Internet, while others have potential that still requires research on their usefulness. Currently, all these suggested topics are at the leading edge of technology and should provide interesting topics to research in the future.

REFERENCES

[1]   A. Bosche, D. Crawford, D. Jackson, M. Schallehn and P. Smith, Defining the Battlegrounds of the Internet of Things, Bain Brief, 2016, 8p. Available: http://www.bain.com/Images/BAIN_BRIEF_Defining_The_Battlegrounds_Of_The_Internet_Of_Things.pdf

[2]   A. Bosche, D. Crawford, D. Jackson, M. Schallehn and P. Smith, How Providers Can Succeed in the Internet of Things, Bain Brief, 2016, 12 p. Available: http://www.bain.com/Images/BAIN_BRIEF_How_Providers_Can_Succeed_In_the_IoT.pdf

[3]   ABB Ability, ABB, Available (accessed on 4.5.2017): http://new.abb.com/abb-ability

[4]   ABB IoTSP, ABB, Available (accessed on 4.5.2017): http://new.abb.com/control-systems/features/industrial-IoT-services-people-use-cases

[5]   Amazon EC2 documentation, Amazon Web Services inc., Available (accessed on 12.7.2017): http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html

[6]   Amazon EC2 pricing, Amazon Web Services inc., Available (accessed on 12.7.2017): https://aws.amazon.com/ec2/dedicated-hosts/pricing/

[7]   Amazon Glacier pricing, Amazon Web Services inc., Available (accessed on 12.7.2017): https://aws.amazon.com/glacier/pricing/

[8]   Amazon Redshift Documentation, Amazon Web Services inc., Available (accessed on 5.7.2017): https://aws.amazon.com/documentation/redshift/

[9]   Amazon Redshift pricing, Amazon Web Services inc., Available (accessed on 12.7.2017): https://aws.amazon.com/redshift/pricing/

[10]   Amazon S3 pricing, Amazon Web Services inc., Available (accessed on 12.7.2017): https://aws.amazon.com/s3/pricing/

[11]   Amazon Simple Storage Service Documentation, Amazon Web Services inc., Available (accessed on 5.7.2017): https://aws.amazon.com/documentation/s3/

[12]   Birst Business Intelligence Tool, Birst Inc., Available (accessed on 14.7.2017): https://www.birst.com/company/

[13]   C. Moynihan, Nobody Puts Data In A Corner: Creating beautiful and compelling dashboards with Birst, Proceedings of the Birst Forward Conference, San Francisco, California, USA, April 24-26, 2017, 72p.

[14]   Daqri Smart Helmet, Daqri, Available (accessed on 14.8.2017): https://daqri.com/products/smart-helmet/

[15]   Digitaalisuuden mahdollisuudet voimalaitoksissa, Suomen Automaatioseura Ry, Seminar, Tampere, Finland, 28.4.2017.

[16]   ECMA-404 The JSON Data Interchange Format Standard, ECMA International, 2013, 14p, Available: http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf

[17]   F-Shaped Pattern For Reading Web Content, Nielsen Norman Group, 2006, Available (accessed on 8.9.2017): https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/

[18]   Gartner IT glossary: Information Technology, Gartner inc., Available (accessed on 18.9.2017): http://www.gartner.com/it-glossary/it-information-technology

[19]   Gartner IT glossary: Operational Technology, Gartner inc., Available (accessed on 18.9.2017): http://www.gartner.com/it-glossary/operational-technology-ot/

[20]   Gartner IT Glossary: What is Big Data?, Gartner Inc. Available (accessed on 13.7.2017): https://research.gartner.com/definition-whatis-big-data?resId=3002918&srcId=1-8163325102

[21]   GE Digital Twin, General Electric, Available (accessed on 4.5.2017): https://www.ge.com/digital/power-digital-twin

[22]   GE Predix Developer Network, General Electric, Available (accessed on 4.5.2017): https://www.predix.io/

[23]   GE Predix, General Electric, Available (accessed on 4.5.2017): https://www.ge.com/digital/predix

[24]   GE's Predix Is Heavy on Marketing, but Light on Capabilities, Lux Research, press release. Available (accessed 14.8.2017): http://www.luxresearchinc.com/news-and-events/press-releases/read/ge%E2%80%99s-predix-heavy-marketing-light-capabilities

[25]   Honeywell Inspire Program, Honeywell, (accessed on 4.5.2017): https://www.honeywellprocess.com/en-US/online_campaigns/iiot-partner-Signup/Pages/index.html

[26]   Honeywell Uniformance Suite, Honeywell, Available (accessed on 4.5.2017):
       https://www.honeywell.com/newsroom/news/2016/07/honeywell-introduces-
       iiot-analytics-platform-uniformance-suite

[27]   IBM to invest $3B in building Internet of Things business, cnet, article. Availa-
       ble (accessed on 3.4.2017): https://www.cnet.com/news/ibm-to-invest-3b-in-
       building-internet-of-things-business/

[28]   Identification System For Power Stations (KKS), Available (accessed on
       3.10.2017): https://www.vgb.org/en/db_kks_eng.html

[29]   Immersive training solutions, Fortum Oy, Available (accessed on 14.8.2017):
       https://www.fortum.com/en/products-and-services/powersolutions/psnu-
       clear/trainingsolutions/pages/default.aspx

[30]   Industrie 4.0, Germany Trade & Invest, Available (accessed on 5.5.2017):
       https://industrie4.0.gtai.de/INDUSTRIE40/Navigation/EN/Topics/industrie-4-
       0.html

[31]   Introducing JSON, Available (accessed on 6.7.2017): http://www.json.org/

[32]   Introduction to ODBC and SQL, Microsoft Corporation, Available (accessed on
       25.7.2017): https://opbuildstorageprod.blob.core.windows.net/output-pdf-
       files/en-us/SQL.sql-content/live/odbc/reference.pdf

[33]   J. Collin and A. Saarelainen, Teollinen Internet, Talentum, Helsinki, Finland,
       2016, 333p.

[34]   J. Juhanko, M. Jurvansuu, T. Ahlqvist, H. Alilisto, P. Alahuhta, J. Collin, M. Ha-
       len, T. Heikkilä, H. Kortelainen, M. Mäntylä, T. Seppälä, M. Sallinen, M. Si-
       mons and A. Tuominen, Suomalainen teollinen internet – haasteesta mahdolli-
       suudeksi, ETLA Raportit No 42., 2015, 66p. Available: https://www.etla.fi/wp-
       content/uploads/ETLA-Raportit-Reports-42.pdf

[35]   KONE revolutionizes elevator maintenance with new customizable KONE Care
       service offering and 24/7 Connected Services, KONE press release, 2017, Avail-
       able: http://www.kone.com/en/media/releases/kone-revolutionizes-elevator-
       maintenance-with-new-customizable-kone-care-tm--service-offering-and-24-7-
       connected-services-2017-02-08.aspx

[36]   L. Dunbrack, S. Ellis, L. Hand, K. Knickle, V. Turner, IoT and Digital Transfor-
       mation: A Tale of Four Industries, IDC White Paper, March 2016, 14p.

[37]   M. Annunziata and G. Bell, Powering the Future – Leading the digital transfor-
       mation of the future, General Electric White Paper, 2015, 24p. Available:

https://powergen.gepower.com/content/dam/gepower-pgdp/global/en_US/docu-ments/product/fuel-capability/digital-energy-transformation-whitepaper.pdf

[38]   M. Livio, The Golden Ratio: The Story Of Phi, The World's Most Astonishing Number, Broadway Books, New York, USA, 2008, 294p.

[39]   M. Porter and J. Heppelmann, How Smart, Connected Products Are Transform-ing Competition, Harvard Business Review, Available: https://hbr.org/2014/11/how-smart-connected-products-are-transforming-compe-tition

[40]   Maintpartner Predictive maintenance, Maintpartner, Available (accessed on 3.4.2017): http://maintpartner.fi/en/services/operation-and-maintenance/mp-in-telligence

[41]   Plant Struxure, Scheider Electric, Available (accessed on 4.5.2017): http://www2.schneider-electric.com/documents/solutions/domains/PlantStrux-ureWhitepaperbyARC.pdf

[42]   Predix Training, Predix Developer Network, Available (accessed on 14.8.2017): https://www.predix.io/resources/training

[43]   R. Sallam, C. Howson, C. Idoine, T. Oestreich, J. Richardson and J, Tapadinhas, Magic Quadrant for Business Intelligence and Analytics Platforms, Gartner, 2017, Available: https://www.gartner.com/doc/reprints?id=1-3TYE0CD&ct=170221&st=sb

[44]   SAP to invest 2 billion euros in Internet of Things by end 2020, Reuters, article. Available (accessed on 3.4.2017): http://in.reuters.com/article/sap-iot-id-INKCN11Y0QU

[45]   Scheider Electric EcoStuxure, Schneider Electric, Available (accessed on 4.5.2017): http://www.schneider-electric.com/b2b/en/campaign/innovation/over-view.jsp

[46]   Secure Shell (SSH), SSH Communications Security, Available (accessed on 5.7.2017): https://www.ssh.com/ssh/

[47]   Siemens MindApps, Siemens, Available (accessed on 14.8.2017): https://www.plm.automation.siemens.com/store/en-ie/mindsphere/mindapps/in-dex.html#

[48]   Siemens MindSphere, Siemens, Available (accessed on 4.5.2017): https://www.siemens.com/global/en/home/products/software/mindsphere.html

[49]   Siemens Plant Data Services, Siemens, Available (accessed on 4.5.2017):
       http://www.industry.siemens.com/services/global/en/portfolio/plant-data-ser-
       vices/Pages/index.aspx

[50]   Valmet Industrial Internet offering launch, Valmet Oy, Available (accessed on
       14.8.2017): http://www.valmet.com/media/news/press-releases/2017/valmet-
       launches-new-industrial-internet-offering-and-starts-partnership-with-tieto/

[51]   Valmet Industrial Internet, Valmet Oy, Available (accessed on 14.8.2017):
       http://www.valmet.com/about-us/industrial-internet/

[52]   W.H. Inmon and D. Linstedt, Data Architecture: A Primer for the Data Scientist:
       Big Data, Data Warehouse and Data Vault. Morgan Kaufmann Publishers, 2015,
       378p.

[53]   Wärtsilä Genius Services, Wärtsilä, Available (accessed on 4.5.2017):
       http://www.wartsila.com/services/areas-of-expertise/wartsila-genius-services

APPENDIX 1: EXAMPLE DATA FILE

```
scan_dttm;tag_name;value
2017-06-02 00:00:00;LBA12DU123;4.17662000656128
2017-06-02 00:01:00;LBA12DU123;5.29157018661499
2017-06-02 00:02:00;LBA12DU123;6.36908388137817
2017-06-02 00:03:00;LBA12DU123;5.40953826904297
2017-06-02 00:04:00;LBA12DU123;4.34052848815918
2017-06-02 00:05:00;LBA12DU123;5.16318559646606
2017-06-02 00:06:00;LBA12DU123;6.01008987426758
2017-06-02 00:07:00;LBA12DU123;4.7961847782135
2017-06-02 00:08:00;LBA12DU123;3.67347502708435
2017-06-02 00:09:00;LBA12DU123;3.73207378387451
2017-06-02 00:10:00;LBA12DU123;3.7661018371582
2017-06-02 00:11:00;LBA12DU123;3.77935361862183
2017-06-02 00:12:00;LBA12DU123;4.13686990737915
2017-06-02 00:13:00;LBA12DU123;1.4420204162598
2017-06-02 00:14:00;LBA12DU123;4.5011596679688
2017-06-02 00:15:00;LBA12DU123;8.48655796051025
2017-06-02 00:16:00;LBA12DU123;9.80646324157715
2017-06-02 00:17:00;LBA12DU123;5.02940225601196
2017-06-02 00:18:00;LBA12DU123;4.87069797515869
2017-06-02 00:19:00;LBA12DU123;2.29962253570557
2017-06-02 00:20:00;LBA12DU123;2.08181047439575
2017-06-02 00:21:00;LBA12DU123;4.30907821655273
2017-06-02 00:22:00;LBA12DU123;4.16013145446777
2017-06-02 00:23:00;LBA12DU123;5.16993808746338
2017-06-02 00:24:00;LBA12DU123;6.12857627868652
2017-06-02 00:25:00;LBA12DU123;5.12389159202576
2017-06-02 00:26:00;LBA12DU123;5.27419018745422
2017-06-02 00:27:00;LBA12DU123;5.3965699672699
2017-06-02 00:28:00;LBA12DU123;5.52737760543823
2017-06-02 00:29:00;LBA12DU123;5.69724702835083
2017-06-02 00:30:00;LBA12DU123;5.66223645210266
```