TÉCNICO LISBOA

Big Data Applications in the Energy Sector

A Review of the Current Status

Oğuz Kıranlar

Thesis to obtain the Master of Science Degree in

Energy Engineering and Management

Supervisors: Prof. Carlos Augusto Santos Silva
Eng. Pilar González Fernández

Examination Committee

Chairperson: Prof. José Manuel Costa Dias de Figueiredo
Supervisor: Prof. Carlos Augusto Santos Silva
Member of the Committee: Prof. António Miguel Areias Dias Amaral

December 2017

Acknowledgments

The first and the most important thanks goes to my supervisors in Iberdrola, Pilar Gonzalez Fernandez and Sergio Platón Martínez, for supporting me in this process and steering me in the right direction by providing valuable practical insights.

I would like to acknowledge my thesis supervisor, Professor Carlos Augusto Santos Silva, and thank him for maintaining a fine balance between guiding me and granting space for my ideas and reflections in the process of writing.

Out of many people who supported me during the process of writing this thesis, I would like to single out three who devoted their time and effort in providing valuable feedback to me. Thank you, Claire Harvey, for reading my whole thesis and carefully noting all language nuances that were unclear. Thank you, Kvitka Krotiuk, for helping me with the wording of some phrases and sentences. Thank you, Tiago Mendes, for translating the abstract of this thesis to Portuguese.

Studying in Entech and being a part of InnoEnergy have been a truly rewarding experience for me. Thanks to Entech, the last two years of my life were filled with unprecedented discoveries about the field of energy technologies, environment/world, and my own self. I would like to thank Entech for its unparalleled approach to studying by providing us not only with the best-quality education, but also with opportunities to reflect on our place in the world as professionals and, most of all, as human beings.

Finally, I am grateful for all the people I met during this journey, some of who, hopefully, will remain my lifelong friends.

Resumo

Esta tese de mestrado oferece uma vista geral compreensiva de aplicações analiticas de Big Data no sector de energia, com um foco especial nas oportunidades para a Iberdrola.

Nos recentes anos, os volumes de dados têm vindo a crescer de forma exponensial. A penetração de novas tecnologias contribui para a geração de dados em várias estruturas. Uma vez analisados, esta enorme e complexa amostra de dados pode criar valor em negócio para organizações. O ramo que se dedica a lidar com o tamanho e complexidade de dados é referido como Big Data.

O sector de energy detém uma quatidade vasta de dados recebidos de cada parte da sua cadeia de valor. Para além disso, está sobe uma tranformação em que os sistemas estão a ficar mais descentralizados, as exigências dos consumidores estão a crescer, e a pressão regulatória em mitigar as alterações climáticas é maior que nunca. Estas tendências fazem da analitica de Big Data uma ferramenta promissora para a adaptação a estas mudanças, e ao mesmo tempo, criando novas oportunidades de negócio.

De forma a determinar o potêncial da analítica de Big Data, a cadeia de valor do sector de energia for dividido em diferentes partes de acordo com a sua relevância em relação à fonte de dados: geração, redes de transmissão & distribuição, gestão de flexibilidade e relações com o consumidor. Foram analisadas as actividades das companhias electricas, as propostas de empresas que usam Big Data para criar produtos e serviços, juntamente com outros sectores que são semelhantes ao de energia em termos de tamanho e estrutura de dados. Foram identificados, se aplicáveis, modelos de negócio que derivam destas aplicações. Finalmente, foi elaborado um caso de negócio apropriado à Iberdrola, usando business model canvas e métodos de análise SWOT.

Os resultados mostram que visibilidade em tempo-real dos ativos, juntamente com a análise preditiva, habilitam operações óptimas para ambas as redes de geração e transmissão & distribuição. Os custos operacionais diminuem com um melhor desempenho de activos e melhoram a produtividade dos trabalhadores de campo. Para a geração de energias renováveis, as incertezas relacionadas com o condições climatéricas são reduzidas, o que leva ao aumento da adopção destas tecnologias.

Descobriu-se que as aplicações em gestão de flexibilidade criam várias oportunidades de negócio. A analítica Big Data é usada para agregar e operar recursos destribuidos de energia (RDEs) em quasi tempo-real. Entretanto, é usada para fornecer serviços a consumidores residenciais, comerciais e industriais, o que leva ao envolvimento de mais activos RDEs em programas de gestão de flexibilidade. Estas operações são geridas ou por serviços de utilidade pública ou por agregadores. Do ponto de vista de relações com o cliente, os serviços de utilidade pública convertem os dados relacionados com o cliente em descernimentos energéticos para oferecer aos seus clientes serviços que levam a um maior envolvimento.

Com base em diveresas aplicações identificadas, foi elaborada a *Iberdrola Efficiency Shop*, uma idea de negócio que converte daods de medidores inteligentes em conselhos energéticos para conduzir os clientes a adquirir aplicações domésticas mais eficázes.

Palavras-chave: Analiticas de Big Data, companhias electricas

Abstract

This master thesis provides a comprehensive overview of Big Data analytics applications in the energy sector, with a special focus on the opportunities for Iberdrola.

Data volumes have been growing exponentially in recent years. Penetration of new technologies contributes to the generation of data in various structures. Once analyzed, this large and complex data can create business value for organizations. The field that is dedicated to tackling the size and complexity of data is referred to as Big Data.

The energy sector holds a vast amount of data received from each part of its value chain. Further, it is undergoing a transformation where the systems are getting more decentralized, customer demands are growing and regulatory pressure on mitigating climate change is larger than ever. These trends make Big Data analytics a promising tool to adapt to these changes while creating new business opportunities.

To determine the potential of Big Data analytics, the value chain of the energy sector was divided into different parts according to their relevance to data sources: generation, transmission & distribution network, flexibility management and customer relations. The activities of electricity utilities and the value propositions of companies that use Big Data to create products & services were analyzed. Further, other sectors that are similar to energy in terms of data size and structure were investigated. If applicable, business models that are derived from these applications were identified. Lastly, a business case suitable for Iberdrola was elaborated by using business model canvas and SWOT analysis methods.

The results show that real-time visibility of the assets together with predictive analysis enables optimal operations both for generation and transmission & distribution network. The operational costs are driven down by better asset performance management and enhanced field worker productivity. For renewable energy generation, weather-related uncertainties are reduced which leads to the increased adoption of these technologies.

It was found that applications in flexibility management create various business opportunities. Big Data analytics is used to aggregate and operate distributed energy resources in near-real time. Meanwhile, it is used to provide services for residential, commercial and industrial end-users which leads to involvement of more DER assets in flexibility management programs. These operations are either run by utilities or aggregators. From customer relations standpoint, utilities convert their customer related data into energy insights to provide services for their customers which leads to a better engagement.

Based on various applications identified, *Iberdrola Efficiency Shop*, a business idea that converts smart meter data into energy advice to drive customers to purchase more efficient home appliances, was elaborated. The case study showed the potential of Big Data analytics to create a new line of business for utilities.

Keywords: Big Data analytics, electric utilities

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter aims to give the reader the scope of Big Data by introducing its relation with advanced analytics and business intelligence Later, it emphasizes the importance of Big Data for transformation in the energy sector. Lastly, the research question, objectives and the methodology of the study are presented.

IBM's computer Deep Blue defeated world chess champion Garry Gasparov in 1997. It was a historic moment that showed the capabilities of artificial intelligence (AI) and machine learning. Its algorithm was mapping every possible move against his opponent, collecting this data and analysing it to determine the best option. Deep Blue's success was based on its high computing performance and data storage capacity. AI evolved rapidly since this breakthrough. From self-driving cars to autonomous weapons, it is possible to see it behind every leading-edge technology. A hypothesis called *The Technological Singularity* claims that intelligence explosion will result in an era of machines which are beyond human capacity and cognition [1]. Whether this hypothesis will ever take place or not, machine learning and AI, serving as data analytics tools, are giving decision makers a fact-based understanding that goes beyond intuition for better a judgement.

Business intelligence is a set of tools that turn raw data into meaningful insights, and it has been actively used by organisations for decades. To perform this, enterprises have been recording, aggregating and analysing data about their business operations such as production processes, sales and customer interactions [2]. Traditionally, these data analysis models were designed in a way that every flow of data is planned by the developer. However, powerful machine learning algorithms have brought brought about a paradigm shift in the way business intelligence is handled. Today, simply putting data into a model is enough to receive meaningful information since the modern algorithms define which questions to ask on the path without human interaction. This ability introduces a concept called Analytics 3.0. It refers to a type of data analytics in which huge and complex data streams are analysed to reveal patterns of past events (descriptive), make forecasts about the future (predictive) and determine the most optimum behaviour (prescriptive) [2]. If we consider data analytics as an engine, powered by machine learning and artificial intelligence, data is the fuel. Therefore, there is a direct link between features of the data and the quality of the analysis results. Data, the fuel of business intelligence, does not come in a refined homogeneous form. It is generated in many different forms and at different rates. The field dedicated to tackle the complexity and size of the data is referred as Big Data[1]

1.1 Background

Penetration of technologies such as sensors, smart phones, online shopping and social media resulted in an exponential growth in data generation. A simple daily life activity such as a post on social media or credit card use is enough to produce data about location, finance, sales and so on. It is estimated that 2.5 quintillion[2] bytes of data is being generated daily and the growth pace is so rapid that 90% of the data in the world was generated only in the last 2 years [3]. International Data Corporation (IDC) estimates that global data volume will reach 163 zettabytes (ZB)[3] by 2025, nearly 10 times the 16.1 ZB of data generated in 2016 [4]. This vast amount of data, through analysis, becomes valuable information, which can be converted into knowledge to gain new insights. This value has been acknowledged by businesses which led to a high interest towards Big Data. As seen in the figure 1.1, there is drastic increase in the frequency of searching the keyword *Big Data* on Google starting from 2012. Piecing together all these trends, it is clear that the era of Big Data has arrived.

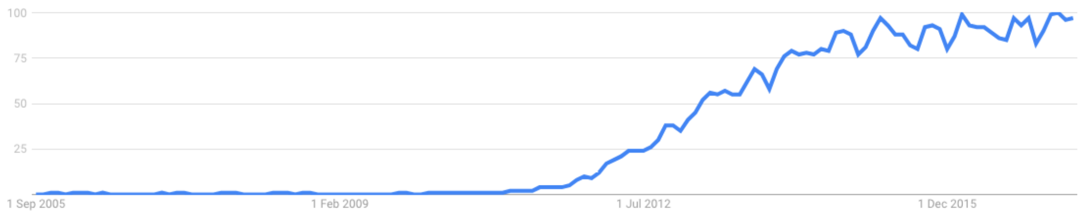

Figure 1.1: Google search frequency for the keyword ”Big Data”

The effective use of Big Data promotes data-driven decision making and thus provides numerous applications for firms in creating new business models, developing new products & services and improving business operations [5]. It is projected that the firms which are able to exploit all of their data and deliver actionable information will benefit an extra 430 billion dollars of income compared with their less analytically driven competitors by 2020. [6]. There is already a great number of use cases available. For instance, The supermarket chain Tesco monitored its refrigerators with sensors which measured the temperature every 3 seconds. Analyzing huge amount of data, the company was able to optimize the refrigerators and reduce costs by 25 million dollars annually [5]. Frito-Lay, a snack producer introduced a biodegradable, environmentally friendly food packaging to the market. However, the new chip bag was extremely noisy and it became a hot topic in social media, endangering the brand name of the company. Analyzing the comments, likes and shares about the product on social media, Frito-Lay

[1] Big Data is written with capital letters along the paper to emphasize the phenomenon, not only the large amount of data.

[2] 10^{18}

[3] Equivalant to 10^{21} bytes

proactively solved the problem before it damaged the brand name [3]. Klaus Schwab, the founder of World Economic Forum, described Big Data as the driving force of fourth industrial revolution, making its value comparable to coal for the first industrial revolution.

The energy sector is undergoing a tranformation. The declining cost of ownership for distributed energy resources (DER)[4] with technological progress, is enabling a higher market penetration. Incentive-based policies contribute to the adoption of DER's as well. An increasing number of DER's is changing customers into prosumers[5]. This situation challenges utilities to meet a growing number of customer choices and demands while continuing their businesses. Higher number of DER's, together with promoted energy efficiency causes a market erosion for traditional utilities. Moreover, having a large amount of distributed power producing and consuming assets makes the grid unstable, complicating the asset management. From a regulatory standpoint, some innovative regulators see the opportunity to allow new players in the sector, such as aggregators[6], bringing more competition in the sector. To tackle climate change, policy makers put a pressure on the sector to cut the green house gases (e.g. EU Emissions Trading Scheme, COP21, Energiewende) and set a target to increase the share of renewables in the generation mix. Growing environmental concerns and competition make utilities more vulnerable to public opinion. Lastly, emerging technologies such as IoT and smart grids, smart meters and devices enable two way communication, increasing connectivity, controllability and automation of the entire value chain of the sector. Based on the trends above, the traditional utilities who are loyal to business models from the 20th century, will be disrupted during this transformation. Similar disruptions are happening in other sectors as well. NetFlix has recently changed the model of media consumption, whereas AirBnb did it with accommodation and Uber with individual transportation. A survey conducted by PwC with 70 utilities shows that 97% of the participants are expecting a disruption in their business by 2020 and 73% anticipate a major business model change by 2030 [7]. Big Data will be a major decisive factor to define the winners of this rapid change in the sector. Analytics 3.0, enabled with Big Data, creates a tremendous value by offering new, customer centric business models, a better understanding of market needs and customer behaviour. Meantime in house use of Big Data optimizes the asset management which reduces operational costs. However, due to its infancy stage, Big Data is not truly embraced by the sector. According to Accenture Technology Vision, only 28% of the businesses think that they are creating value out of their data while 40% accept that they need a plan to implement it [8]. Meanwhile, there is a large number of companies, ranging from multinational technology providers to startups, who are dedicated to leverage data in the energy sector. Some innovative utilities, concentrated in Northern America and Europe, have already started to make use of the phenomenon. In conclusion, Big Data will play a key role in the following years, and therefore having a comprehensive overview of the Big Data applications in the energy sector related to business intelligence will be very helpful for utilities.

From an academic perspective, Big Data applications in electricity sector are still relatively unexplored. There are several overviews, featuring the business opportunities of Big Data in the sector.

[4] It is a comprehensive definition for entire resources connected to distribution grid (e.g. solar PV, small, batteries, CHP and so on.

[5] A consumer who can also produce.

[6] An actor in the sector who aggregates a load capacity from DER's and offers it in the energy market.

However, their scope do not go beyond certain sections throughout the value chain (e.g. smart grid, distribution grid) Moreover, they do not present a holistic approach, including the companies, business models and current activities of utilities.

1.2 Research Question

Based on what has been mentioned earlier, the research question of this study is:

How could Big Data create value for utilities in the rapidly transforming energy sector?

1.3 Purpose of the Study

Iberdrola is a Spanish based utility, serving over 32 million customers in Spain, the UK, the USA and Brazil. The company was ranked the world's fourth largest utility in terms of R&D&I[7] investments [9]. Iberdrola Ventures-PERSEO is the venture capital program of the utility. They aim at fostering innovation in the energy sector by giving investor support to entrepreneurs. Therefore, the team is always up to date about new technologies. The opportunities of Big Data analytics for utilities drew attention by Iberdrola Ventures-PERSEO and they wanted to have a comprehensive overview of the technology, specifically its operational and business opportunities for utilities.

Based on the research question, the purpose of the study is to identify the current applications of Big Data analytics for each part of the energy sector value chain. To conceive an extensive overview, several side objectives must be reached. These are mainly:

- Analysis of companies dedicated to leverage data to create value for utilities.
- Analysis of use cases in other utilities.
- Determination of new business models that are enabled by Big Data analytics applications.
- Analysis of other sectors that are similar to utilities.
- Proposal of a business case for Iberdrola.

1.4 Research Methodology

The number of Big Data analytics applications are so large and the field is so dynamic that identifying them requires a careful research and analysis from a very large variety of resources. Therefore, a research methodology was structured as shown in figure 1.2. It is a general to specific approach, starting with reaching out to as many resources as possible and is followed by determination of the applications. Later, the energy sector value chain is broken down to different parts, and applications were sectioned accordingly. At this point, related business models were also investigated. Lastly, a business case for Iberdrola which uses Big Data analytics was elaborated. These sections are explained in detail below.

[7] R&D&I stands for research, development and innovation.

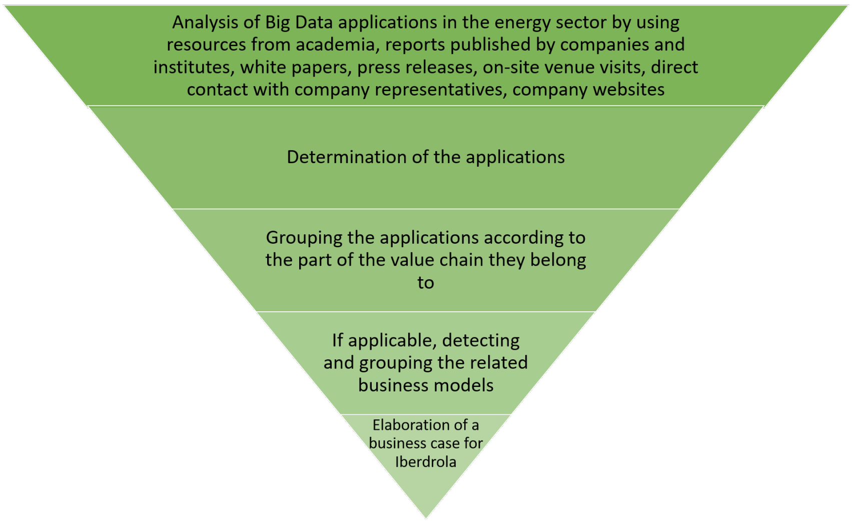

Figure 1.2: Research methodology: From general to specific approach

1.4.1 Research Scope and Resources

The research started with an academic overview. Early results showed that the scope of the articles that review Big Data applications in the electricity sector tend to be limited. They focus on the applications either enabled by smart grids/IoT or a limited part of the electricity value chain [10, 11, 12, 13, 14, 15]. To create a holistic overview, the entire value chain was divided into groups with the objective to address them all. Further, company activities were examined on top of the information from academia. This approach is the essence of research methodology. It reveals the real business value of Big Data analytics since it contains information about latest updates in the sector. Analysis of company activities also leads to identification of innovative business models together with the applications. As an example, when the academic papers emphasize on the theory of an application, analysis of company activities show that the same technology can be used to create different business models. The strategy of company research followed the the steps below:

- Activities of large utilities were investigated. This step gave insights about the current trend of the sector.
- An investigation of big players in each part of the value chain was conducted. For instance, in the case of identifying applications about wind energy generation, largest producers of wind turbines were reviewed.
- The companies who are dedicated to leverage data to create value for utilities were examined. The research started with multinational companies and scaled down to start-ups.
- Activities of other sectors that have millions of customers such as banking and telecom were investigated. These sectors have a degree of similarity with the energy sector in terms of customer relations. The objective is to identify the use cases which could be applied to energy sector as well.

The resources of the study are mainly academia, reports published by companies and institutes, white papers, press releases, on-site venue visits, direct contact with company representatives, company websites, etc. The scope of the study is limited to North America and Europe since the majority of the countries that Iberdrola operates are located in these regions[8]. Moreover, these regions are the pioneers of IT technologies and most of the activities occur here.

1.4.2 Determination of the Applications

At this stage, company activities and value propositions are converted into applications. Since Big Data is not a product but a phenomena which deals with making use of complex datasets, a filtering according to the relevance to Big Data definition was required. Some applications use data and analytics but the data does not match with the Big Data definition, and thus remains out of the scope. The elaboration of the applications start with description of the application. Later, if there is a cluster of business models related to the application, they are explained. Lastly, these are supported by a selection of best practices from related companies and case studies.

1.4.3 Grouping the Applications

The energy sector value chain was broken down to four parts according to its relevance to Big Data analytics. This was necessary since the value creation varies according to the part it belongs to. For instance, generation related applications mostly focus on cost reduction and asset management which are more operational benefits. There are not any business models that can be derived from these. In the meantime, other parts create value for both the utilities and their customers which have potential for business creation. The division is done accordingly:

- **Generation:** The scope of electricity generation was limited to wind and solar energy. This is due to the decentralized nature of these energy resources. The increasing number of assets leads to a more complex dataset, and thus a match with the Big Data definition.
- **Transmission & Distribution Network:** The operational benefits of transmission & distribution network is covered.
- **Flexibility Management:** The applications that aim at managing flexibility of the grid was covered under this part of the value chain. Since the applications are mostly customer-centric services that involve them in the management, there is a difference between the distribution network and flexibility management applications.
- **Customer Relations:** This part refers to interaction between a utility and its customers.

1.4.4 Elaboration of a Business Case

After completion of the research, a case study for a new line of business that uses Big Data analytics was proposed to Iberdrola. The business idea was elaborated by using business model canvas and SWOT analysis techniques.

[8] Iberdrola operates in Spain, the US, the UK and Brazil.

Chapter 2

Big Data

The purpose of this chapter is to give a better understanding of Big Data to the reader by introducing its characteristics and value chain. Further, it introduces the Energy Big Data concept, by explaining the data sources and enabling technologies of Big Data in the energy sector.

Big data is still in its infancy stage of development, and thus there is not a consensus about the definition of the term. In general, it can be defined as an extremely large amount of structured or unstructured data that is not possible to manage with conventional data management techniques. However, this definition does not capture all the aspects of Big Data. There are quite a few definitions with different perspectives available in literature, namely the product oriented perspective, the process-oriented perspective, the cognition-oriented perspective and the social movement perspective [16].

- The product oriented perspective relates Big Data with its size, usually petabytes to zettabytes [16]. It also emphasizes the rate of generation. Facebook's data warehouse which holds 300 petabytes of data with a daily incoming rate of 600 terabytes, fits the product oriented perspective of Big Data [17].
- The process-oriented perspective refers to the authenticity of the operations that are required to capture, store, manage and analyze the data rather than its size.
- The cognition-oriented perspective focuses on the fact that it is not possible for an individual to comprehend and work with Big Data, and therefore it is required to handle it with technological infrastructures, interdiciplinary work and statistical analyses [16].
- The social movement perspective relates the term to socioeconomic, cultural, and political shifts that are caused by it such as scientific discovery, environmental and biomedical research, education and national security [16]. Several definitions from different organizations are shown in table 2.1.

Table 2.1: Big Data definitions from different perspectives

Perspective	Definition	Institution	Source
Product oriented	”Large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, internet transactions, emails, videos, click streams, and/or all other digital sources available today and in the future.”	NSF	[18]
	”Growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data”	White House	[19]
Process oriented	”Big Data is a term encompassing the use of techniques to capture, process, analyse and visualize potentially large datasets in a rea- sonable timeframe not accessible to standard IT technologies.” By extension, the platform, tools and software used for this purpose are collectively called Big Data technologies.”	NESSI	[20]
	”Big Data technologies are a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis.”	IDC	[21]
	”Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization”	Gartner	[22]
Cognition oriented	”Big Data exceed the reach of commonly used hardware environments and software tools to capture, manage, and process it within a tolerable elapsed time for its user population.”	TERADATA	[23]
	”Data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze.”	McKinsey Global Institute	[23]
Social movement	”The potential to quantify and change various aspects of contemporary lift, to revolutionize the art of management”	-	[23]

2.1 Dimensions of Big Data

Big Data was first characterized in a report by Meta Group (currently Gartner) in 2001 with the concept of 3Vs, namely volume, velocity and variety [22]. Later, IBM introduced the veracity dimension and finally in 2013, Demchenko proposed the value dimension which brought about the 5Vs concept.

- Volume is associated with the cumulative sum of the data stream which can reach up to zettabytes.
- Velocity corresponds to the rate of data used to support the interactions and generated by interactions. As the applications which use data become more complex, the velocity of data interaction increases. Consequently, data cannot be stored without processing. For instance, IBM’s hospital monitoring application collects and analyzes more than 100.000 data point per patient per second [24].
- Variety refers to differences and incompatibility between the data types, sources and structures. The data format can be in the forms of documents, emails, text messages, audio, images, video, graphics data, and others [25].
- Veracity describes the degree of reliability of the data. Provenance, quality and noisiness of the data determines the veracity [26]. Provenance is the tracking and recording of the original data. Ideally, using the original data is preferable as it improves the accuracy of processing. Data quality is about objectiveness of data. Many resources such as social media contain subjective data and processing subjective data might result in meaningless conclusions. Lastly, noisiness of data corresponds to uncertainties in the datasets. For instance, the labels of datasets might be inaccurate,

and thus degrades the veracity [26].

- Value is the monetary worth added to an organization by implementation of Big Data [23].

2.2 A Brief History of Big Data

The size of data has been challenging the IT world for the last four decades. As the conception of ”Big” evolved from megabytes in 1970s to exabytes in 2010s, the methods to handle the data has changed. The first data related challenge arose in enterprises as they wanted to store and analyse their historical business data [27]. The solution was ”database machines” that consisted of an integrated hardware and software system. The idea behind integrating software and hardware was to keep the costs low. In 1980s, it became evident that single database machines weren't sufficient since the data amount was growing. This realization led to software-based parallel databases where data is kept in separate hardware. ”Shared-nothing” parallel database systems became the most popular solution where a networked cluster of individual machines are used [28]. These systems pioneered the divide and conquer parallelism, wichh partitions data for storage [12]. The introduction of Web 1.0 in 1990s enhanced data size, and therefore the necessity of search engines. Web 1.0 also enabled unstructured data generation, which the parallel database systems weren't capable of managing. Google introduced the Google File System (GFS) and MapReduce programming model which allowed automatic data parallelization and distribution of computation applications to large clusters of commodity servers [12]. Many other web companies such as Facebook and Yahoo created open source versions of Google's Big Data stack which later became the popular Big Data platform ”Hadoop” [28]. During the mid-2000s, the exponential rise in data generation brought about NoSQL databases, which are scheme-free, fast, highly scalable and reliable [12].

2.3 The Big Data Value Chain

The value chain of Big Data can be divided into four parts: data generation, data acquisition, data storage and data analysis [12]. Data generation and acquisition cover the exploitation of raw material while data storage is needed to keep the data until it is used in the analysis part, where data is converted into value. Figure 2.1 visualizes entire value chain with respect to time, including the technological developments involved.

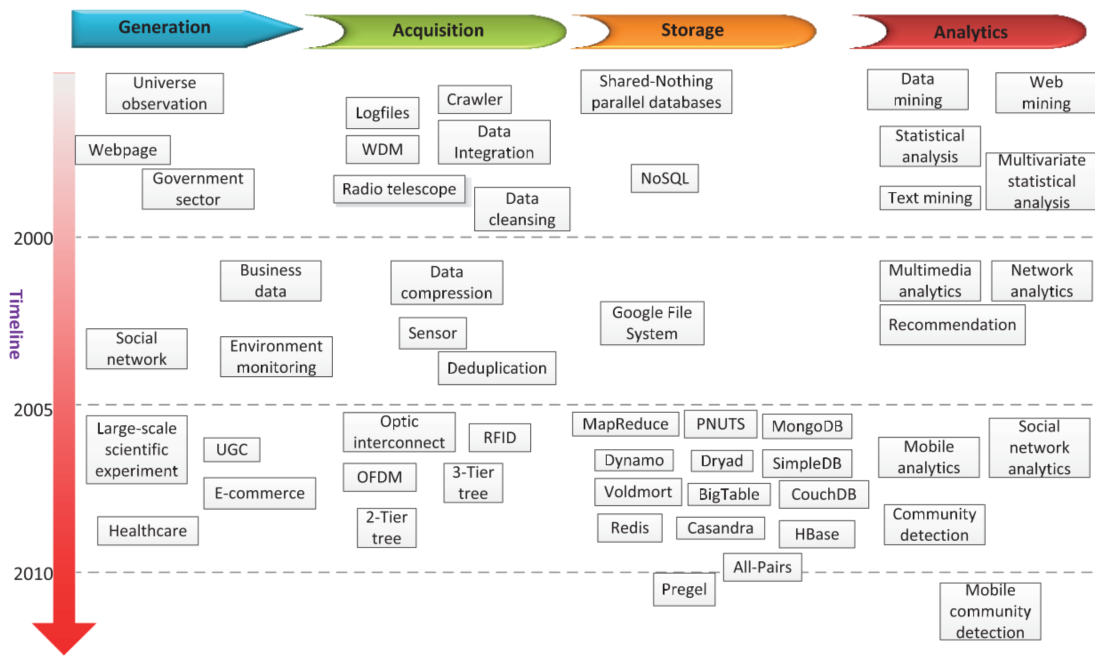

Figure 2.1: Value chain of Big Data with respect to time, including the technological developments [29]

2.3.1 Data Generation

Considering data as a raw material, it was the acceleration of data generation which enabled Big Data applications. Hence, the progress of data generation should be separated in three stages to better understand its evolution. The first stage started with adoption of digital technologies and database systems by organizations in 1990s. They began to store large amount data such as bank and trading transactions, government archives. These datasets were structured and small enough to be managed in organizations' storages systems. The second stage was triggered with the introduction of Web 1.0 systems in late 1990s. Web 1.0 refers to search engines and e-commerce businesses which generate a substantial amount of semi-structured and unstructured data (webpages, transaction logs). In early 2000s, Web 2.0 allowed user generated data to be included. The sources were forums, blogs and social networking and social media websites. The third stage was initiated by mobile devices such as sensors, smart phones, and sensor-based internet-enabled devices [29].

Data Sources

Currently, main sources of data come from business operations of enterprises, internet of things (IoT), human interaction and position information provided by smart phones and social media and scientific research [12].

- Enterprise data consists of the data generated during the business operations of enterprises such as online trading and analysis, production, inventory, sales, finance, etc [12]. Currently the biggest share of data sources belongs to enterprise data and it is estimated to double in every 1.2 years [30]. For instance, Walmart processes one million trade transactions hourly, which are stored in databases with 2.5 petabyte capacity. [30].
- IoT data is estimated to be the biggest source of Big Data in future. McKinsey Global Institute

predicts that the number of IoT devices will range from 25 billion to 50 billion by 2025 [31]. The IoT data is large-scale and is obtained from distributed devices. The data format might vary depending on the application. For instance, location information is structured, when video streaming is unstructured and thus, heterogeneous. IoT devices are distributed in specific geographical locations and they keep track of time. The information about time and space involved in IoT data facilitates data analysis in further stages. Lastly, a small portion of IoT data is effective. For example, a 24/7 functioning security camera provides valuable information only in case of emergencies [12].

- Networking data involves the internet and mobile networks. The data in this field is generated extremely fast and requires advanced processing ability.
- Scientific data is generated in advanced scientific experiments or observations such as the Large Hadron Collider in CERN, Sloane Digital Sky Survey (SDSS) which is an astronomical catalogue and in the field of computational biology. [29].

2.3.2 Data Acquisition

The second part of the Big Data value chain is data acquisition and it comprises three processes: data collection, data transmission and data pre-processing. As discussed earlier, data is generated in various structures and it needs to be acquired from a data production environment. Later, the acquired data is transformed to storage. Lastly, pre-processing needs to be applied in order to manage noise and redundancy of data. Figure 2.2 summarizes the data acquisition process and the methods applied.

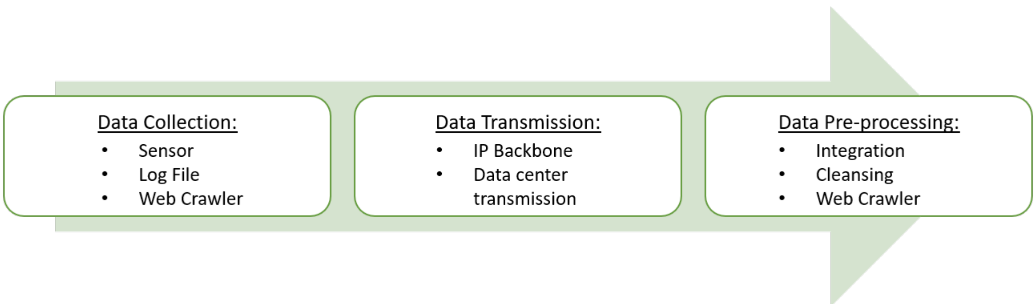

Figure 2.2: Data acquisition process

Data Collection

Data collection methods vary depending on the features of the data sources and the purpose of analyses. Three data collection methods are widely used in Big Data applications: log files, sensors and web crawler.

- Log files are one of the most common data collection methods. They are automatically generated record files of data source systems. They are used in nearly all the applications running on digital devices [12]. For instance, user activities such as clicks and visits of web servers are recorded as log files [32]. There are mainly three types of log files in use: public log file format (NCSA), expanded log format (W3C), and IIS log format (Microsoft) [29].

- Sensors are devices which measure physical quantities and convert them to digital signals. Sensor data includes acoustics, sound, vibration, chemical, electric current, weather, pressure, thermal, etc. [29]. The sensed information may be transformed in a wired or wireless system depending on the application. For instance, a security camera is a wired network when applications like environmental research require sensory data to be collected in a large area.

- Web crawler is program that is used to download and store web pages by search engines. Table 2.2 presents a comparison of different data collection methods and their features.

Table 2.2: Comparison of different data collection methods and their features

Method	Structure	Scale	Complexity	Applications
log file	structured or semi-structured	small	easy	web log, click stream
sensor	structured or unstructured	medium	sophisticated	video surveillance, inventory management
web crawler	mixed	large	medium	web search, SNS analysis

Data Transmission

Data transmission is divided into 2 phases: IP backbone and data center transmission. The IP backbone phase comprises of the transmission of data from the distributed sources to a data center through a trunk line. The rate of data transfer is dependent on physical media and link management methods [29]. Data center transmission refers to the transportation of data inside of a data center. It is necessary for place adjustment and processing. Such transmission is carried out depending on the architecture and the protocols of data centers.

Data Pre-processing

Data pre-processing is required since the data comes with noise, redundancy and inconsistency due its characteristics mentioned before. This action not only decreases the storage costs but also increases the analysing efficiency. The most common techniques are integration, cleansing and redundancy elimination [12].

The function of the integration is to combine data sets from different sources and present users a unified view. There are mainly two methods: data warehouse and data federation. Both methods recollect data sets, combine and serve.

Data cleansing is the process applied to identify inaccurate, incomplete or unreasonable data. Once identified, these parts can be deleted. The data cleansing technique refers to a process which determines inaccurate, incomplete, or unreasonable data and then to amend or remove data to improve quality. In general data cleansing follows four steps: determination of error types and instances; correction of errors; documentation of error types and instances, and modification of data entry procedures to improve future applications [12]. Even though data cleansing promotes a better data analysis, it requires extra computation. Hence, a balance must be established between the cleansing and analysis steps.

Redundancy, data repetitions or surplus data, occurs in many types of data. Elimination of redundancy is crucial to decrease the transmission and storage costs. Several methods were suggested to eliminate redundancy such as redundancy detection, data filtering and compression [29]. Once again, a balance of computation resources must be considered since these operations require extra computation.

2.3.3 Data Storage

Data storage is required to organise data in a suitable format for analysis and value extraction. Two features are desired in a data storage system. Firstly, the storage infrastructure must accommodate the information persistently and reliably. Secondly, it must provide a scalable access interface to query and analyse the data [29]. Data storage subsystems can be viewed in two components, namely storage infrastructure and data management.

Storage Infrastructure

The infrastructure of storage systems can be investigated according to the specific technology that they are manufactured from or the networking infrastructure they use. The most common storage technologies are random access memory (RAM), magnetic disks such as hard disk drive (HDD) and storage class memories such as solid-state storage (SSD). RAM is a volatile storage mechanism where data is lost in the case of a power cut. A HDD includes a number of rapidly rotating disks with magnetic heads which are positioned on top of a actuator arm which reads and writes data on surfaces. The written data on a HDD system remains when the power is off and the capacity costs are low. However, writing and reading velocity is low in a HDD system since the technology is based on mechanical movements. The SSDs do not have mechanical components, and thus work faster and quieter. The drawback of such technology is its high cost. Current data storage devices in the market combine features of HDD and SSD systems which results in an optimization of cost and performance.

There are several types of networking architectures, namely direct attached storage (DAS), network attached storage (NAS) and storage area network (SAN) [12]. DAS is the simplest storage system where several storage devices are directly connected to a computer. It is a simple extension to existing infrastructure and its scalability is limited. Therefore, it is commonly used in personal computers and small sized servers. NAS systems use a network to provide users data access and sharing in form of files. This network is separated from the main server and consists of its own software. Thus, it has a higher scalability. Lastly, SAN is a specialised network with high speed provided by fiber optic cables. It is designed to have better scalability which is the most important feature for Big Data applications. Figure 2.3 indicates the three types of networking architecture.

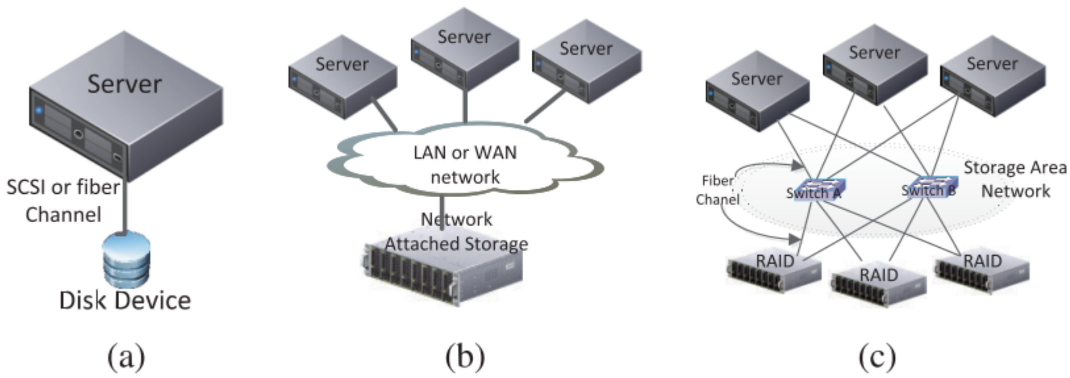

Figure 2.3: Three types of networking architecture [29]: a) DAS b) NAS c) SAN

Data Management Framework

Data management framework aims at providing the most convenient method of data organization for processing and analysing. Currently, there are three methods used in Big Data applications: file systems, data bases and programming models [12].

- File systems which are required in SAN and NAS systems are the basis of data storage technologies. There are several file systems proposed by different companies. For instance, Google designed its own file system, GFS, which uses cheap commodity servers to provide fault tolerance and high performance.
- The data base is an organised collection of data, and is an essential component of data storage systems. Since Big Data applications use various types of data with many different sources, traditional relational data bases cannot address the challenges. Hence, non-SQL (NoSQL) has been the most convenient database for Big Data storage. Multiple specifics such as supporting easy replication, possesing a simple API, eventual consistency and supporting a huge amount of data make NoSQL very popular for Big Data applications [29]. There are three primary types of NoSQL databases available: key-value databases, column-oriented databases and document databases. Key-value databases are composed of a data model and the data is stored according its key value. They came into use several years ago by the Dynamo system proposed by Amazon. Later, many other versions based on Dynamo appeared in the market (e.g, Voldemort, Redis, Tokyo Canbinet, Tokya Tyrant, Memcache, Riak and Scalaris). Column-oriented databases store the data according to the column of the dataframe unlike the tradiitonal databases where the variable is the row. The first column-oriented database is BigTable by Google. Other versions such as Cassandra, HBase and Hypertable were proposed by different companies [33]. Lastly, document databases are commonly used in Big Data applications. They are capable of managing more complex data structures compared to key-value stores due to their data model. The three data bases available are SimpleDB, MongoDB and CouchDB [33]. Table 2.3 indicates the main databases available in the market.

Table 2.3: The most commonly used NoSQL databases used in Big Data applications

Data Model	Name	Producer
Key-Value	Dynamo Voldemort Redis	Amazon LinkedIn Salvatore Sanfilippo
Column-oriented	BigTable Cassandra Hbase Hypertable	Google Facebook Apache Hypertable
Document	SimpleDB MongoDB CouchDB	Amazon 10gen Couchbase

- Programming models are the last element of a data management framework. The drawback of NoSQL databases is the fact that a declarative expression of join operations and querying are not fully supported. This drawback is overcome by programming models which implement the application logics to NoSQL databases. There are three major models, namely generic, graph and stream processing models [29]. Generic processing models facilitate accessibility of NoSQL databases by MapReduce model. MapReduce is a programming model which automatizes paralleling and managing computation based on a large number of commodity computers. It achieves this task by two functions. The map function processes a key-value pair according to the input data and generates intermediate key-value pairs. Later, the reduce function merges all the intermediates related to the intermediate key, and thus reduces the data size [25]. The graph processing model expresses the computational task as a directed graph with vertices and directed edges. The vertices hold a modifiable and user-defined value. Programs are run as a sequence of iterations called supersteps until the algorithm terminates [34]. The main commercial models are Pregel and Graphlab. The last model is the streaming processing model.

2.3.4 Data Analysis

The last phase of the Big Data value chain is data analysis. It is also the most important phase since a tremendous number of applications offered by Big Data comes from data analysis. In general, the purpose of data analysis is the inspection of data reliability, guidance for the decision-making process, future prediction and fault detection [29]. From the objective perspective, entire data analytics can be categorized as descriptive, predictive and prescriptive analysis [35]. Descriptive analysis aims at determining the reasons of incidents by exploiting historical data. It is commonly used in business intelligence. Predictive analysis refers to capturing future trends and possibilities by predictive modelling. Lastly, prescriptive analysis focuses on decision making by observing a system.

Data Analysis Techniques

Data analysis techniques contain many different disciplines such as machine learning, data mining, statistical analysis and data visualization.

- Machine learning is a technique which is capable of revealing valuable knowledge by learning the behaviour from empirical data. It has been one of the most popular research areas of Big

Data technology. It typically follows the data pre-processing, learning and evaluation phases [36]. Data pre-processing is applied to prepare the data for machine learning algorithms. Such algorithms generally performs a selection, performance measuring, error estimation, and statistical tests. Once the output is received, a new algorithm might be applied if the performance of the previous one is not satisfactory. Several dimensions are taken into account to characterise machine learning. These are nature of learning feedback, target of learning tasks, and timing of data availability [36]. Figure 2.4 indicates a multi-dimensional classification of machine learning.

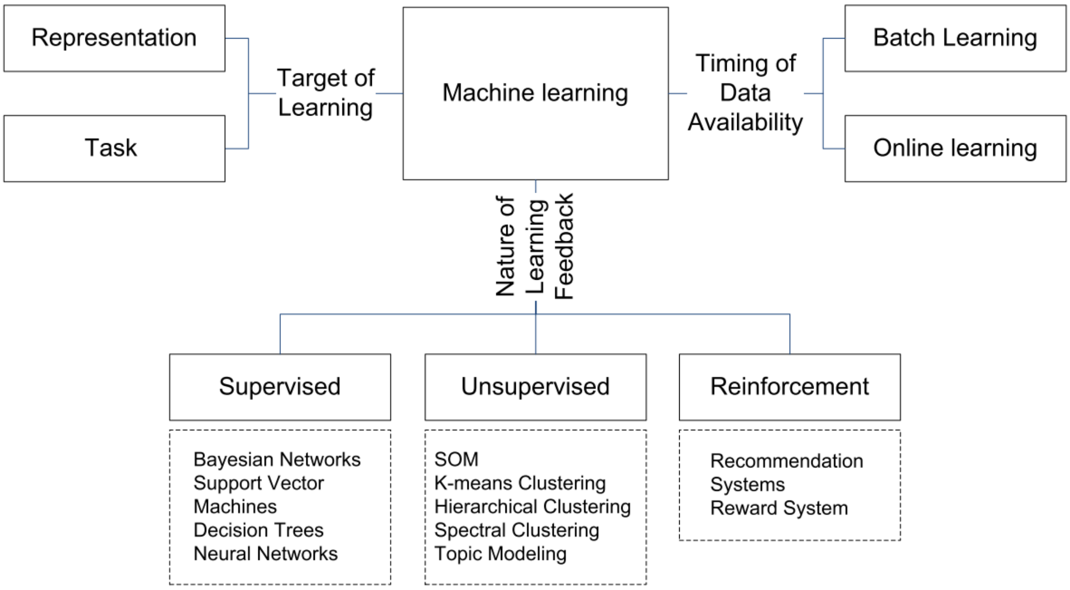

Figure 2.4: Multi-dimensional classification of machine learning [36]

From the nature of learning perspective, machine learning is divided into three types: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the input-output pairs are submitted to a learning system in order for the system to identify the pattern between the pairs. Typical algorithms used are bayesian networks, support vector machines, decision trees and neural networks [36]. Unsupervised learning does not take into account the relation between input and output. It aims at finding the pattern between the input data. The most common algorithms are self organising maps, K-means, clustering, hierarchical clustering, spectral clustering and topic modeling. The reinforcement learning holds features from both supervised and unsupervised learning. It feeds from the feedback of its previous experiences similar to supervised learning. However, it does not try to find the pattern between the input and output. Principal algorithms are recommendation systems and reward systems.

From the target of the learning tasks perspective, machine learning can be described as representation or task learning. The objective of representation is to learn new and more convenient ways of data representation, and thus facilitate information extraction in further steps. Unlike representation, task learning has a concrete desired output which the learning system tries to approach. Clustering, classification and regression are the typical examples of task learning [36].

From the timing of data availability perspective, machine learning is classified as batch learning or online learning. Batch learning uses the entire data set to identify the pattern, whereas online learning updates the model for every input. Online learning is used when there is no computational ability to cover the entire dataset or when the data is generated continuously.

- Data mining is a computing process which is used to define the pattern of a data set. The process starts with domain understanding and continues with data selection, data pre-processing, cleaning and transformation, discovering paterns, interpretation and reporting [37]. Domain understanding is applied to determine for what purpose the results of the process will be used. This enables better gathering of relevant information beforehand. For instance, if the intention is to conduct a video mining of a tennis game, it is important to know the rules of the game in order to capture the score. Data selection, pre-processing and cleaning is applied to prepare the data set for applying the algorithms. The pattern discovery stage, as the name suggests, is the core of the process. Several machine learning, statistical data analysis, neural networks, fuzzy logic and pattern recognition algorithms, including association, classification, clustering, regression, time-series analysis, and visualization are used in this stage [37]. Interpretation is carried out to validate the discovered pattern. Using the example of the tennis game, this stage compares the scoring pattern that the process discovered with the real world scores. Once the comparison is made, data mining ends with reporting.

- Statistical analysis uses applied mathematics to analyse the data sets. It is useful in the context of data analysis in two ways: description and interference. Descriptive statistical analysis is useful for description of the data set, whereas inferential statistical analysis determines the interferences [38].

- Data visualization is an effective tool to facilitate information communication by use of graphics. Charts and maps are commonly used to express the data. It is one of the hot topics of Big Data technology since the traditional methods for data visualization (e.g. spreadsheets) are not feasible to use considering the volume of the data sets.

2.4 Algorithms Used in Big Data Applications

This section explains the commonly used algorithms for Big Data applications.

- Regression algorithms are used in predictive analysis which relate a dependent variable to other independent variables. It is used for forecasting, revealing the cause-effect relations and time series modelling. Some typical questions that can be answered with regression analysis are ”How is sales volume affected by the weather?” or ”How does the title of a book affect its sales?”. The role of the regression is to set up a model out of the data [39]. The main regression algorithms are ordinary least squares regression (OLSR), linear regression, logistic regression, stepwise regression, multivariate adaptive regression splines (MARS), and locally estimated scatterplot smoothing (LOESS) [40].

- Instance-based algorithms use training data which are considered to be significant for the learning model. They compare training data with the test data in order to reveal similarities, and thus make a prediction. Instance-based algorithms are also called lazy learners since their working principle consists of storing the training data and waiting until the test data is introduced. Some common algorithms are: k-Nearest Neighbour (kNN), learning vector quantization (LVQ), self-organizing map (SOM), locally weighted learning (LWL) [40].
- Decision tree algorithms are one of the most commonly used algorithms for Big Data applications. They map observations about an item to conclusions about the target value of the item [40]. The decision tree is called a classification tree when the dataset is finite and is called a regression tree in the case of introducing continuous values. Their popularity is due to many advantages they offer, particularly for machine learning such as easy understanding, interpretation and visualization [41]. Main decisions tree algorithms are the classification and regression tree (CART), iterative Dichotomiser 3 (ID3), C4.5 and C5.0 (different versions of a powerful approach), chi-squared automatic interaction detection (CHAID), decision stump, M5, and conditional decision trees [40].
- Bayesian Algorithms use probability theorem to identify the uncertainties of the suggested model. The most popular algorithms are naive Bayes, Gaussian naive Bayes, multinomial naive Bayes, averaged one-dependence estimators (AODE), Bayesian belief network (BBN) and Bayesian network (BN) [40].
- Clustering is an unsupervised learning algorithm. It is based on the classification of objects into different groups according to their similarities. The commonly used clustering algorithms are k-means, k-medians, expectations maximisation (EM) and hierarchical clustering [40].
- Association rule learning algorithms find out the rules behind the observations. Apriori and Eclat algorithms are the commonly used methods [40].
- Artificial neural networks algorithms are based on the human brain's working principle in terms of processing and learning ability, but in a very limited scale. These methods allow dynamic and non-linear relations to be examined [42]. The typical algorithms are perceptron, back-propagation, hopfield network and radial basis function network (RBFN) [40].
- Deep learning algorithms are an enhanced version of artificial neural network and one of the most promising methods in Big Data applications. Deep learning algorithms can manage the volume, variety and velocity of the data due to distributed representations of the data [25]. The main algorithms are deep Boltzmann machine (DBM), deep belief networks (DBN) , convolutional neural network (CNN) and stacked auto-encoders [40].
- Ensemble algorithms combine several different weak learners such as averaging and voting. These weak learners become the sub model and are used to make an overall prediction. Some popular algorithms are boosting, bagging, bootstrapped aggregation, adaBoost, stacked generalization (blending), gradient boosting machines (GBM), gradient boosted regression trees (GBRT) and random forest [40].

2.5 Energy Big Data

Digitalisation systems in the energy sector boosts the rate of data generation. However, the characteristics of data flowing throughout the system reveals the challenges mentioned in the earlier sections (*i.e.* volume, variety, velocity of the data). Therefore, Big Data is the essential technology to realize the paradigm shift in the sector and Energy Big Data is the term that refers to data related to it.

The sources of Energy Big Data can be investigated in two groups, namely electric utility data and supplementary data [11]. Electric utility data includes all of the enterprise data and the information that a utility can reveal from a smart grid. Electric utility data consists of supervisory control and data acquisition (SCADA) data, phasor measurement units (PMU) data, smart meter data, intelligent electronic devices (IEDs) data, asset management data, digital protective relay (DPR) data, digital fault recorder (DFR) data, sequence of event recorder (SER) data, advanced metering infrastructure (AMI) data, control and maintenance data for equipment, and automated metering reading (AMR) data [43]. Supplementary data refers to all other data sources which are beneficial for Big Data applications such as geographical information system (GIS) data, global positioning system (GPS) data, time-reference data, weather and lightning data, seismic reflection data, animal migration data, financial market data, social media data and regulatory reporting data [11]. Figure 2.5 summarizes entire data sources used for Big Data applications.

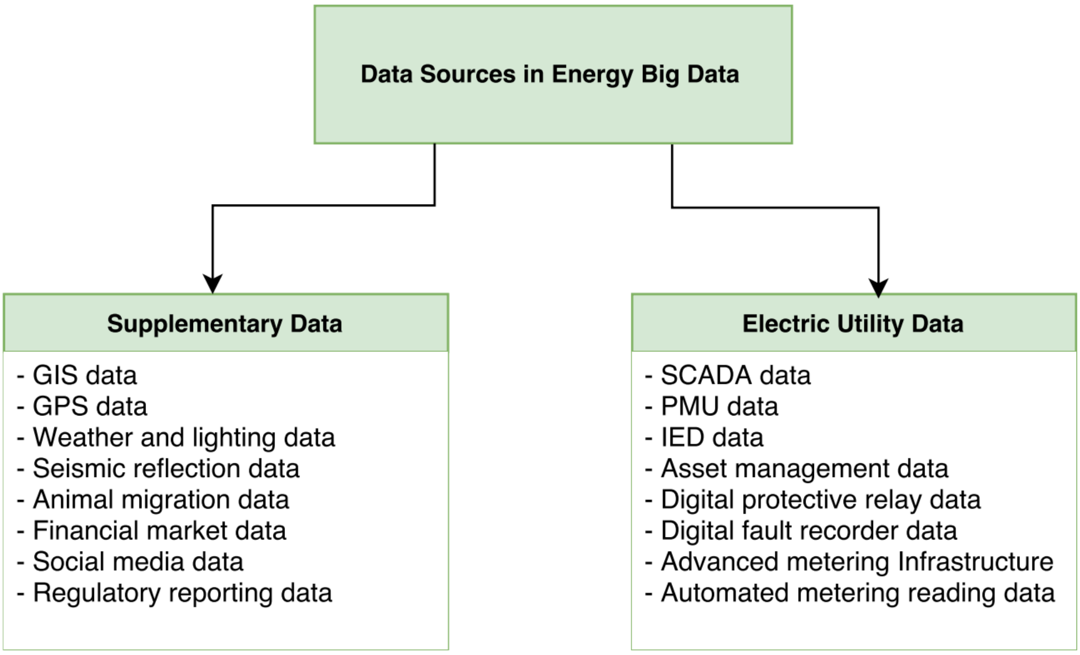

Figure 2.5: Data sources of energy Big Data

2.5.1 Enabling Technologies

There are three essential technologies which accommodate the penetration of Big Data applications in the energy sector. The value created by Big Data is directly linked to the improvement of these accompanying technologies. These are smart grid, internet of things (IoT) and cloud computing.

Smart Grid is a new generation power network which has a bidirectional power and data flow, distributed and automated components and ability to perform real-time balancing of demand-supply by computing and communication features [43]. Table 2.4 summarizes the differences between a traditional and smart grid with respect to some significant features.

Table 2.4: Comparison of a traditional and smart grid

Feature	Traditional Grid	Smart Grid
Power flow property	Undirectional	Bidirectional
Generation profile	Centralized	Distributed
Grid configuration	Radial	Network
Integrating DERs	Very rare	Frequently
Sensor devices	Few	Plenty
Monitoring	Restricted view	Self-monitoring
Control	Limited and passive	Pervasive and active
Outage recovery	Manual restoration	Self reconfiguration

Internet of Things (IoT) is another technology strongly linked to Big Data . It is the communication of physical assets by a wireless infrastructure (i.e. internet). For instance, Amazon recently launched a service in combination with washing machine producer Whirlpool. The machine communicates with Amazon to re-order the washing powder before it is over [44]. The technology is very valuable for Big Data applications as well, since sufficiently equipped products can apply basic commands without the need of processing.

Cloud Computing is the delivery of computing services over the internet. It enables the entire value chain of Big Data to be centralised in the expert company while the value is provided by a software or a platform. Even though the technology is relatively new, the applications are already in the daily lives of people. Social media platforms are good examples of cloud computing. A user can keep and sort the tweets in a dashboard dedicated to him on the internet while all the required processing and storage actually happens in the platform providers assets. The individual's device doesn't use any processing power particularly for these applications. There are two services related to cloud computing, namely software-as-a-service (SaaS) and platform-as-a-service (PaaS) and both are commonly used for Big Data applications in the energy sector [45].

Chapter 3

Applications of Big Data Analytics in the Energy Field

This chapter covers the current Big Data analytics applications in the generation, distribution & transmission network and customer relations parts of the value chain.

3.1 Wind Energy Generation

Wind energy is one of the most favourable renewable energy resources in the world. The cumulative installed capacity has reached to 430 GW worldwide in 2015, with an average growth rate of 20% for the last 10 years [46]. Currently, vast majority of the wind farm installations are on-shore. However, the off-shore wind farm deployment is growing drastically. The share of off-shore installations have reached 3% in 2015, from 1 % in 2010 and this trend is expected to continue [46]. Wind energy generation varies dramatically in time and space, injecting costly uncertainties into the finance and operation of the wind farms. Big Data analytics holds many opportunities to overcome these uncertainties. The big wind turbine producers have already started to offer products and services that leverage Big Data.

Implemention of Big Data analytics is possible by generating and collecting the right volume of data with the optimum velocity that is compatible with computational skills of the enterprise. General Electric (GE), one of the largest wind turbine producers, develops modern wind farms with 50 sensors in each turbine. Data is generated in different rates according to purpose of the analysis in these sensors. The first level of analysis starts in the turbine every 40 milliseconds to optimize the pitch of the blades and decide if the electricity generated will be stored or fed into grid depending on available storage systems. The second level of analysis occurs at the farm. The farm controller analyses 30 signals from each turbine with 160-millisecond frequency in order to deliver the predictable power. Besides, the farm controller uses 200 data clusters with a 1 second interval from each turbine to evaluate turbine health and performance. Later, pre-processed data is transferred to a remote monitoring center in 1 minute intervals. Data scientists and engineers placed in remote monitoring centers conduct analysis to increase the overall performance of the farms and prepare recommendations. Moreover, historical operating data is utilized in supercomputers in order to reveal correlations and critical issues among all the turbines. These insights are significant for asset performance management including predictive maintenance, spare parts inventory management etc. Lastly, operational data from the farm is combined with other enterprise data (e.g. financial data, electricity market data) to report on the power production and forecast [47].

3.1.1 Pinpointing the Optimal Location for Wind Turbines

Precise placement of a wind turbine has a big impact on its performance. The windiest location doesn't necessarily mean the best choice to erect a turbine. Turbulence is a decisive factor for the service life of the turbine since it strains the components. Besides, having the precise information of a location allows virtual layout of the wind farm before the investment is made. This proactive approach allows the right configuration of the turbines to be obtained such as tower height and rotor diameter. Enabling an actionable knowledge requires Big Data analytics.

Vestas-IBM

Vestas, a global wind turbine producer, combines meteorological data from 35.000 weather stations with the company's historical operational data from 50.000 wind turbines worldwide to obtain a more precise windmap. The size of the data is approximately 2.8 petabytes and includes 160 factors such as temperature, barometric pressure, humidity, precipitation, wind direction and velocity elevation, topography, and satellite images [48, 49]. The company was able to reduce its wind data grids from 27x27km to 10x10m by using the IBM InfoSphere BigInsights Big Data analytics tool [48]. Vestas is now able to predict the amount of the energy that will be generated in a certain location during the project planning with several hours of computation time. A conventional prediction with less accuracy would take up to weeks due to point measurements, estimation of wind and turbulences.

3.1.2 Operation Optimization

The predictive analysis of vast amounts of data makes the operations more visible and predictable. Once the operations are more visible in near-real time scale, daily activities in a wind farm, and the work plans can be automatized, which leads to a decrease in operating costs. Real-time monitoring of the actual and potential power output is compared, including the possible reasons of lacking the full potential. Furthermore, recommendations to fix the issue is provided via historical operation data which suits the particular case [50]. The positioning of the blades is very crucial for the most efficient energy conversion. Ideally, this positioning must be achieved in real time since the wind speed is changing instantly. Real-time performance monitoring, with the necessary control hardware can also increase the power output which leads to higher annual energy production in the wind farm. It is achieved by adjusting parameters such as speed, torque and pitch.

General Electric

General Electric is one of the pioneers of Big Data implementation in the wind energy sector. They provide smart services in many applications related to wind energy generation. As seen in figure 3.1, an overviewing software is capable of giving a vast amount of information about the operational status of the wind farm.

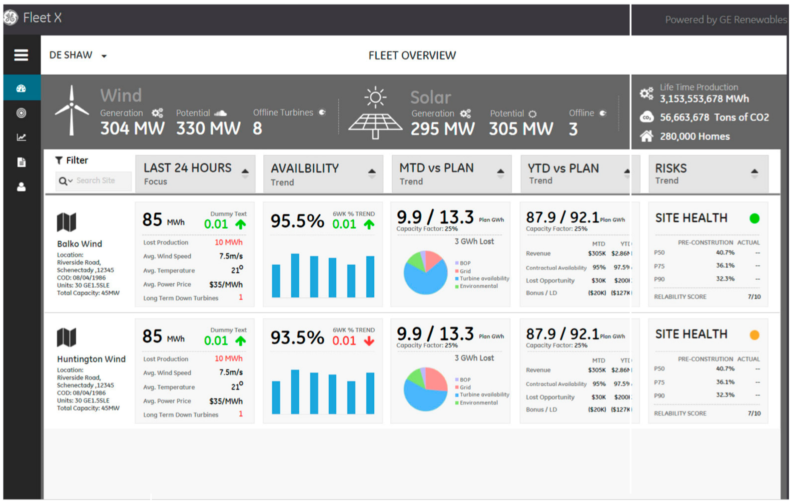

Figure 3.1: Snap from FleetX software [50]

A real-time comparison of the potential and realised generation is made with suggestions to bridge the gap. Digital-plan-of-the-day, a software to optimize daily activities in a wind farm, visualizes the data related to failures in the turbines, the weather and the available staff in the field. It optimizes the best operation decision to increase productivitiy. PowerUp, another service by General Electric, tunes the wind turbine operation by adjusting the parameters such as drive train speed, torque, pitch and yaw to increase annual energy production (AEP) [50]. The service increased the AEP by up to 10% in some of the deployed wind farms [50].

3.1.3 Asset Performance Management

It is essential for wind energy developers to have their turbines ready to operate at any time the wind is available. Maximizing availability of the turbines increases the profitability. Wind energy, being an asset-centric industry, is able to operate safer and more reliably with a data-driven asset performance management (APM). The benefits of a data-driven APM are listed below:

- Reduction in unplanned downtime of the turbines, and thus an increase in the availability.
- Prevention of costly emergency repairs by detecting problems early.
- Reduction in unnecessary routine maintenance which introduces risks and decreases availability.
- Reduction in inventory costs since decision-making on spare parts storage and logistics is based on failure rates.
- Increased health and safety performance due to lower asset-related incidents.

A holistic APM approach consists of machine & equipment (M&E) health monitoring, reliability management and asset strategy optimization. M&E health monitoring provides a comprehensive view of asset performance by exploiting real-time data, alarms, events, and other operational data, and thus enables reliability management. It is the type of management which would prevent unplanned downtime. It starts with recording actions taken by the experts during the mitigation of an error (i.e. converting unstructured inspection notes to structured data sets) and combines with M&E data to reveal predictive analysis. Lastly, asset strategy optimization is an improvement process which updates the asset management procedures depending on the feedback provided by reliability management reports [51].

General Electric

General Electric offers comprehensive asset management with its Digital Wind Farm concept. The monitoring of the assets is realised with sensors inside of each turbine located in the yaw of the nacelle, to measure the torque of the generator and the speed of the blade tips [52]. Later, a predictive analysis is held to determine failures before they happen with the potential risk analysis.

Siemens Wind Power

Siemens is a large, multinational technology corporation founded in 1847 and based in Munich, Germany with a reported 2015 revenue of 90.8 billion euros. Siemens is producing its wind turbines with sensors that measures the vibrations in the transmission case, the generator, and the main shaft bearing at the rotor blades. The sensor data is transmitted to a data center together with weather data. The center uses this data to identify the anomalies and predict the failures. In some cases, a bearing failure can be detected one year prior to the deadline of replacing it [53]. Remote monitoring and diagnosis services give flexibility in scheduling the maintenance and play an important role in cost reduction in asset management.

3.1.4 Forecasting

Forecasting the power output with the shortest time scale and the best accuracy serves well for the utilities since they need this information for managing intermittent generation grid/dispatch, energy trading and portfolio management, financial forecasting and regulatory requirements [54]. Hence, efforts to predict the power output have been made since 1980's, before the era of Big Data had arrived. First attempts were using only wind speed data and statistical algorithms. Therefore, shorter intervals were resulting in inaccurate predictions. Starting from 2000s, more accurate predictions were achieved with the developments in machine learning algorithms and Big Data tools [55].

HyREF by IBM

Hybrid Renewable Energy Forecasting (HyREF), developed by IBM, is a software which analyses large volumes of sensor-collected weather data and instrument-collected power data in order to forecast power output of a wind farm with high accuracy. It is able to estimate power generation every 15 minutes in an extra short term (0-4 hours), short term (4-72 hours) and near term (up to 1 month). Moreover, it can detect extreme conditions and alerts the operator [56]. Figure 3.2 presents a screen shot of the software. The yellow line represents the predicted energy generation while the blue line stands for actual energy generation. The close relation between the lines proves the accuracy (over 90%) of HyREF.

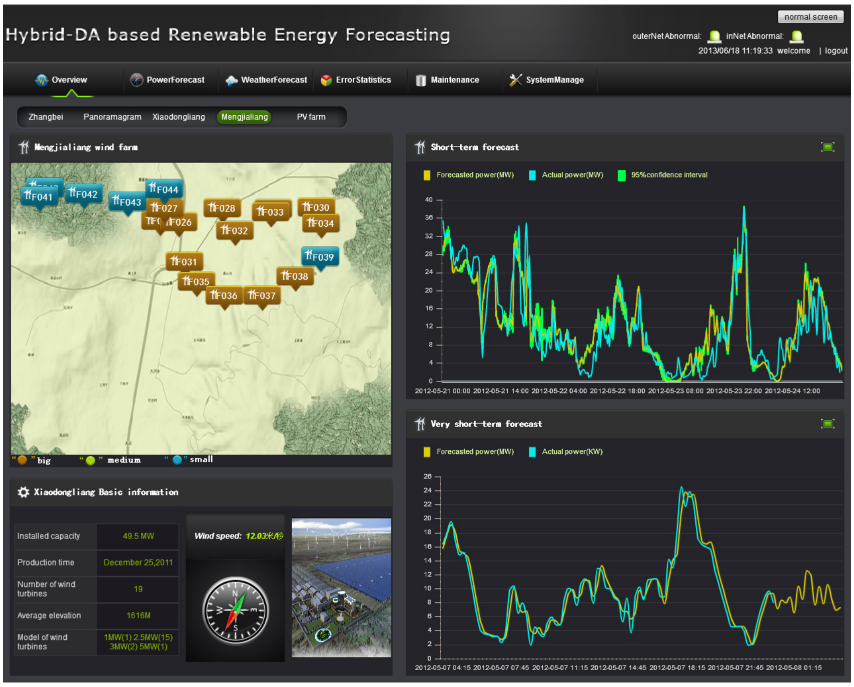

Figure 3.2: Snap from the HyREF software [56]

WeatherSentry by Vaisala and Schneider Electric

WeatherSentry is a software developed by Vaisala, in collaboration with Schneider Electric. Unlike HyREF, WeatherSentry uses only wind speed measurements from the turbines and the historical weather data of the wind farm locations. It is compatible with existing monitoring systems of wind farms. It stands out as a cost effective solution and it is utilized by more than 130 GW of installed wind generation capacity worldwide [57].

3.2 Solar Energy Generation

Solar has been the fastest growing renewable energy source in terms of installed capacity for the last 6 years. The cumulative global capacity has reached to 235 GW in 2015 and it is expected to be doubled by 2018 [58]. There are two methods for extracting electric energy from the sun, namely concentrated solar power (CSP) and photovoltaics (PV). The working principle of CSP is similar to a conventional power plant, in which steam is produced in a boiler and fed into turbines to obtain mechanical energy. The heat source is concentrated solar rays instead of fossil fuels, and thus it is a renewable energy production. In the meantime, PV technology is based on the fact that the sunlight hitting a semiconductor results in voltage and current difference in a cell which is the source of electricity generation. In 2016, the share of CSP deployment was 3.5 GW, corresponds to less than 1% of total capacity [59]. This is mainly due to lower costs of PV comparing to CSP. Moreover, PV systems can be installed in different scales. This feauture facilitates their adoption by smaller parties. These parties can be categorised as residential (0-20 kW), commercial (20-100 kW), industrial (100-1000 kW) and utility-scale (5 MW+) according to their installment capacity [59]. The segmented ownership of PV technology facilitates numerous Big Data analytics applications in the sector.

3.2.1 Reducing the Soft Costs

The increasing maturity level of solar PV technology resulted in a drastic decline in the price of residential, grid connected PV systems during the last decade. In 2016, the cost of a PV system in Europe was 1350 euro/kW without VAT [58]. The hardware of the system accounts for 67% of the total costs. The non-hardware related costs resulting from customer acquisition, installer overhead, financing, contracts, inspection, permitting, interconnection, and installation labor are called soft costs [59]. Big Data analytics is able to reduce the soft costs by offering smarter marketing strategies, more precise preliminary investigation of the projects and facilitating the financing for solar developers.

kWh Analytics

One of the major expenses in a solar project is the high financial costs. Being a weather dependent energy source, solar projects hold uncertainty. The risky nature of the investment results in higher interest rates. Moreover, unlike wind, solar projects are mostly developed by smaller size companies which don't have performance guarantees. kWh Analytics, a US based start-up, aims at being the trustful 3rd party to investigate project performance for the investors. This will decrease the interests on solar project loans by increasing the credibility of the investment. The company achieves this by gathering data from more than 70.000 solar projects on its platform together with its risk management software.

Kevala

In the US, the ultimate electricity price depends not only on wholesale pricing but also the local distribution system which is called Locational Marginal Price (LMP). It is locational due to the additional costs resulting from transmission of electricity from generation source (node) to end user. Hence, it is very valuable to know if a potential solar project site is close to a node or not. Kevala maps the node and substation locations and ranks them according to their profitability by taking into account the energy value derived from annual hourly solar production through weather data and historical wholesale prices. This helps investors to narrow their site search down to more profitable locations. Further, the characteristics of the grid components also influence the costs. Kevala determines the relationship between the feeders and the parcels in order to estimate the capacity of a feeder for the designed electricity carriage. By this information, the investor can know if the planned site is overreaching the grid capacity which would require an extra infrastructure investment. Another difficult-to-estimate cost is interconnection of the site to the grid. The parameters such as proximity of the selected site to feeder lines and substations as well as their characteristics influence the costs. Kevala maps all the parcels served by a particular feeder and visualizes them according to their cost effectiveness. The company provides these services by gathering various publicly available data such as utility filings, outage reports, permit applications, visual inspection and historical wholesale pricing data [60].

CertainSolar

CertainSolar is a Washington, US based start-up. The company aims at reducing the uncertainty on potential savings of a solar project. One of the risks related to solar projects is the uncertainty of utility rates in the market. Estimated utility rates are taken into consideration while calculating the savings of solar installations. A poor estimation might lead to unprofitable investments. This uncertainty leads to higher loaning expenses that hinders solar projects. CertainSolar, using Big Data analytics, makes accurate predictions on utility rates, and thus reduces the costs [61].

3.2.2 Facilitating Residential Solar PV Adoption

Marketing and installation costs still have a significant share in the total cost due to the conventional marketing techniques usually applied. Big Data analytics can facilitate an increase in the adoption rate of solar PV systems by offering solutions to households, the contractors and the developers.

PowerScout

PowerScout is an US based startup, founded in 2015. The company uses light detection and ranging (LIDAR) data with detailed consumer data and advanced image recognition technology to determine households that would adopt solar PV systems in the close future. Each home is tagged with 1.200 data points, including income and education levels, political affiliation, the type of the resident and the car [62]. This insight would cut off the marketing expenses sharply. They also provide a platform for households to facilitate decision-making for solar PV installation. The platform is able to predict the potential electricity output of the house by leveraging the same database. They create a digital surface model of the roofspace, determining the roof facet, the orientation and the slope. Then, a calculation takes place to identify the optimum panel installation and power output. Current practise to inquire such information is to send an expert who would perform a survey to decide whether the house is suitable or not. It also contains close-by contractors with the best installation prices and gives the household an opportunity to choose from them. The contractors register to the platform and increase their business volume while a better structured work plan increases their productivity.

Geostellar

Geostellar, a US based company found in 2010, also provides a marketplace in which home owners can determine how much electricity they can generate without a site investigation, resulting in lower acquisition costs for the installers and smarter decision-making for the home owners. The company automates the solar electricity generation model by taking into account weather data, the size, orientation and slope of a roof; and shadows from trees, structures and terrain. Later, a financial calculation is conducted by using utility rates, energy usage profiles and energy incentives [63]. The accuracy of the method was verified by National Renewable Energy Laboratory (NREL). 76 % of the calculations were matching with the on-site performance tests results [64].

PickMySolar and PVimpact

PickMySolar is a web based market place for solar panel installers. The company allows the home owners to compare and choose the best contractors only by managing the platform. This allows the company to access age, gender, household income, financial status, credit scores, building details data.

Later, the application PVimpact leverages data obtained from PickMySolar and combines it with smart meter data from the households. The application enables home owners to have an easy access to their savings, receive alerts and recommendations about their system with Big Data analytics. It also gives an option to share their data for benchmarking with their neighbours on a dynamic map. PVimpact is still under development. However, once it is widely spread, it has a potential to give insights to utilities about adoption rate of solar PV in certain areas and generation/consumption data for demand response programs.

3.2.3 Performance Improvement

Solar PV sites are monitored to keep the performance at the highest possible level at all times. For a utility scale site, the monitored data is gathered from inverters, combiners, trackers, module optimizers, meters, transformers, switchgear and substations. Moreover, weather sensors measure irradiance, temperature and wind speed. The monitoring data combined with third party data sources can improve the annual energy production and develop better asset management methods.

Alectris

Alectrics is a PV operations, maintenance and asset management company which was established with the merging of Advartia and Origis Energy in 2012 [65]. They combine the monitoring data with the financial and business operation data of the operator to deliver actionable information. For instance, the software can observe a low internal rate of return and connect this observation to a lack of performance due to initial bad design of an asset [66].

NexTracker

NexTracker is a subsidiary of Flex, a multinational electronics company. The company is focused on producing tracking systems for solar panels. Solar power plants lose electricity generation efficiency due to construction variability, terrain undulation and changing weather conditions. NexTracker proposes a tracking system, leveraging cloud positioning, fog or haze data to correct the panel direction by using machine learning algorithms. This quick change in the panel position increases the annual energy output up to 6 % by preventing the unnecessary losses [67].

3.3 Transmission & Distribution Network

3.3.1 Asset Performance Management

The biggest responsibility of a utility is to ensure that the grid infrastructure is functional at all times. However, this is a major operational challenge since the grid consists of numerous distributed components such as transformers, cables, towers, circuit breakers, etc. For developed countries, the current grid infrastructure was built in the 20th century. The average equipment age is 30-35 years [68]. As a result, asset performance management faces challenges. Much of the equipment will soon be operating beyond its designed life time and needs to be replaced. The equipment tends to break down. This leads to costly urgent repairs to keep the grid running. Taking into account the size of the grids, a very large organization of field work is required. The traditional approach follows the points below:

- Prediction of the failures from historical data based on intuition and experience.
- Creating management strategies which targets the entire equipment rather than considering the condition of each component in it. Lack of a risk/criticality assessment approach.
- A limited view on equipment health conditions.

The predictive analysis capabilities of Big Data analytics transforms the traditional approach by giving pro-active insights of current status of the grid. These insights are then converted to optimized work plans. The new data-driven approach consists of predictive maintenance, a digital worker concept and asset planning.

Predictive Maintenance

The data collected from sensors, site inspections, GIS sources and historical maintenance records give insights about the condition of the equipment. Later, analytics converts these near-real time insights into actionable information. A possible failure in the equipment can be detected with a lead time of days, weeks or even months. A maintenance strategy can be formed, supported with automatic recommendations. This type of maintenance approach is called predictive maintenance. Decision makers can base their maintenance strategy on risk analysis and prioritize the components which require urgent care.

C3 IoT

C3 IoT is a SaaS provider for predictive business intelligence. Having its headquarters in the US, the company operates both in North America and Europe. The products of the company mostly aim at utilities. They use IoT, enterprise and third party data to create services such as predictive maintenance, fraud detection, customer segmentation, etc. Figure 3.3 shows a snap from the user interface of their predictive maintenance platform. Once a service territory is chosen on the map, the platform shows a risk score of the assets.

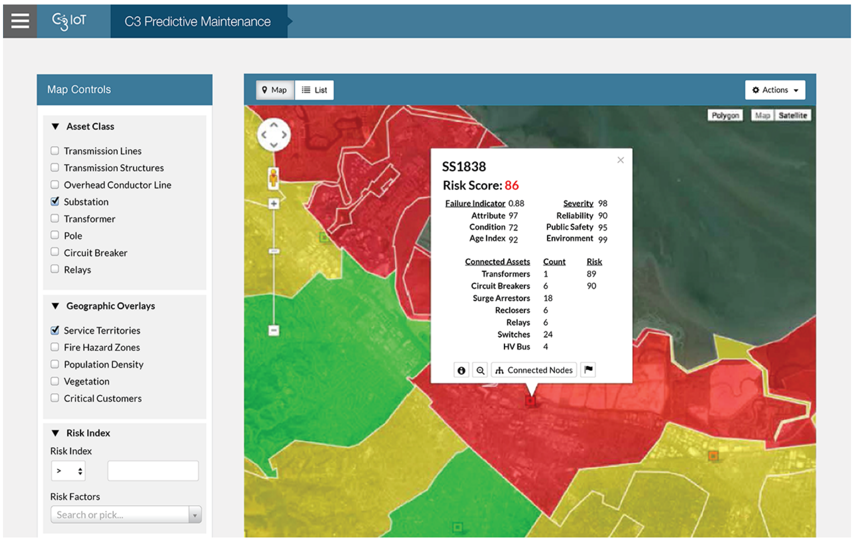

Figure 3.3: User interface of C3 IoT predictive maintenance platform [69]

Enel, based in Italy, is one of the largest utilities in Europe operating more than 10 countries globally and serving over 61 million customers [70]. They implemented C3 Predictive Maintenance across 16.000 substations in Italy which corresponds to 1 million customers. The platform predicts the failures in feeders by collecting data from SCADA, maintenance work orders, fault protection, asset management, historical equipment failures, known network issues, power quality, lightning, terrain and vegetation, and weather [71].

Digital Worker

The predictive maintenance decisions can be used to empower field workers with mobile devices in which optimized work plans are present. This increases the productivity. Further, a better data collection is possible since the paperwork is removed.

AusGrid

AusGrid is one of Australia's largest utilities providing electricity for 1.6 million customers. Owning 500.000 power poles, 30.000 substations, and over 50.000 km of cables in a enormously large area requires a intense work load. The company distributed mobile devices to its field workers. The devices allow inspection data to be collected instantly and fed into analytics tools. The company announced a 73% increase in productivity with this approach due to less time spent on paperwork, and more optimized repair site visits. Moreover, data collection was handled with less errors due to the simplified method [72].

Arria

Arria is a British based natural language generation company. They convert large datasets into simple-to-read texts. An example of the conversion is presented in figure 3.4. Such simplification can be used by field workers. An advanced text language generation is able act as a senior engineer that gives simultaneous work orders.

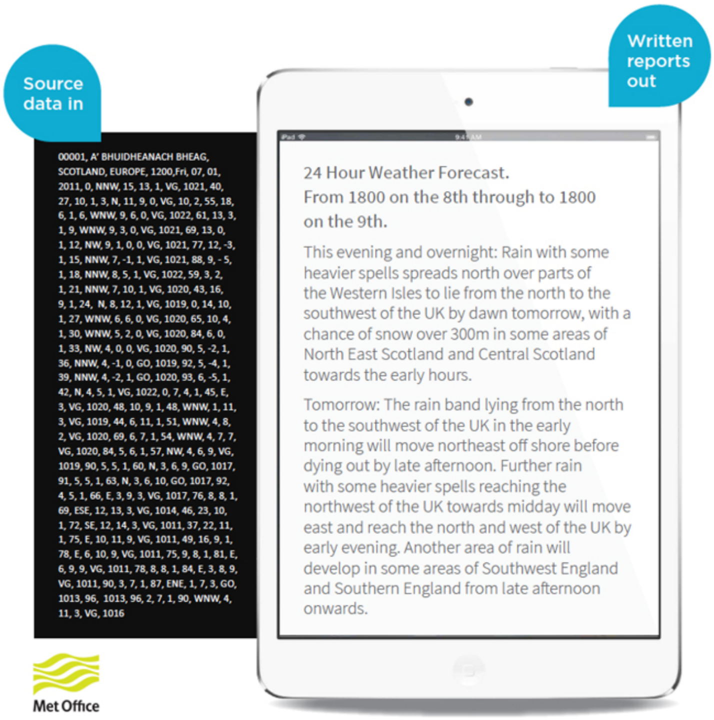

Figure 3.4: Conversion of raw dataset into easy-to-read text [73]

Arria implemented the technology in a large utility in the USA[1] [74]. The response time to manage an outage is very critical for utilities. An engineering expertise is required to understand the reasons behind the outage. However, analysis of the problem takes time. Utility needed a system which acts like a virtual senior engineer to decrease this response time. The system should instantly analyse the data and communicate with the workers in a plain English. Arria interviewed senior engineers and analysed historical outage management reports to learn from historical data to achieve the goal. The implementation of the system enabled the know-how of the company to be documented, the press releases related to outages to be generated much faster and more efficient decision making [74].

Asset Planning

Big Data analytics can also support decision makers for their investments. The analysis of the grid can optimize the appropriate size of a transformer for a certain area or can warn when it is necessary to add new transformers.

3.3.2 Proactive Outage Management

Having a holistic asset performance management approach isn't enough to eliminate electricity outages. 70% of the outages result from weather conditions in the US [75]. According to the US Department of Energy 679 outages occurred in the country due to weather conditions with an economic cost of 18-33 billion dollars annually between 2003 and 2012 [75]. These weather dependent outages create direct

[1] The company doesn't reveal its name.

costs to utilities since they mobilize crews and replace/repair the assets without any lead time. Moreover, a long time without electricity leads to a bad image of utility. Big Data analytics can help utilities to predict outages by use of historical weather and outage data.

Weather Company Outage Management

Weather Company is a weather forecast company acquired by IBM in 2016 [76]. They convert weather data into insights for various sectors. Their software product outage management combines weather data with outage data and learn from them to predict the number of outages and infrastructure damage within a utility service territory with a 72 hours lead time. The model is based on predictive analysis. Figure 3.5 presents a snap from the user interface of the product. Different colors represent potential number of outages. The area can be adjusted as an entire territory, a region or a substation. The novelty of the product is its ability to combine 16 different weather forecast datasets to obtain the most accurate results [75]. The model uses parameters such as wind speed (sustained and gust), precipitation, temperature, atmospheric pressure, humidity, soil moisture, and foliage from the weather data and vegetation, asset condition (location and/or health), population from other sources [77]. Having been released in June, 2017 the product is one of the most cutting-edge Big Data applications in the sector.

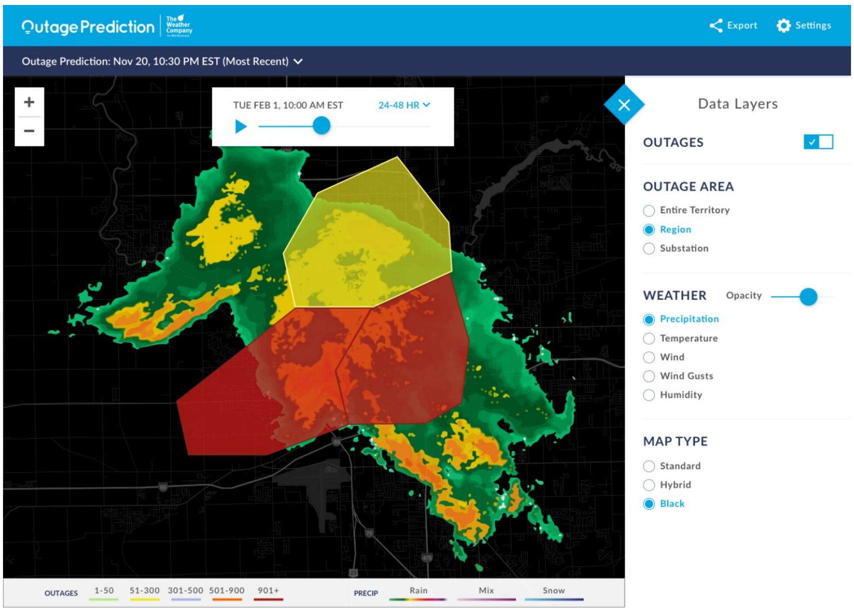

Figure 3.5: User interface of Weather Company Outage Management [78]

3.3.3 Theft Detection

Electricity theft is the major reason of non-technical revenue losses for utilities. The thefts commonly occur at the distribution grid where millions of customers are connected. The large and complex structure of the distribution grid makes it difficult to detect anomalies. The problem is greater for developing countries. The losses related to theft reaches up to 30% of the total electricity generation in India, while the world average is 8% and the average of OECD countries is 6% [79]. For instance, Brazil[2] suffers

[2] Iberdrola serves 13.4 million customers in Brazil.

a theft of 27.000 GWh annually [80]. The reasons of theft aren't only infrastructure related but also socio-economic features play an important role. For instance, fevalas[3] in Brazil are the main source of theft, while in the South East of Turkey, where there is an active terrorist group, there is the record rates of theft, up to 73% [81]. Big Data analytics is cabaple of pro-actively detecting the theft and mitigating the problem.

The data received from utility such as SCADA, AMR systems, AMI, MDA systems, GIS, CIS and AMS is combined with external data on the weather, business registries, demographic data, and so on. The analysis in real time is then fed to a dashboard using data visualization techniques. For instance, if a line is located in an agricultural area where there isn't a significant urbanization and load on the line is expected to be low. A unexpected load in this line is most probably a theft due to illegal use electricity for farming. The level of data input has a significant influence on the accuracy of the results. However, the thefts are higher in the areas where metering is not widespread. Therefore, use of external sources and customized algorithms and methods are applied in different areas. A trend of having local partners to develop customized theft detection systems are observed in the market [82]. Once the theft is detected, a machine learning algorithm is applied to proactively identify a potential theft and risk analysis.

Choice

Choice is a Luxemburg based company, focused on detecting electricity theft by using Big Data analytics. Their product Revenue Intelligence prioritizes and optimises the potential targets of theft for investigators [83].

Light S.A. is a Brazilian utility providing electricity to more than 10 million customers in Rio de Janeiro. The company used Choice's Revenue Intelligence service to tackle their theft problem. The first results after 4 months showed 21.9 GWh of energy theft prevention. As the program went on, the annual amount reached up to 160 GWh which corresponds to 40 million dollars of savings [84].

3.4 Customer Relations

Historically the utility-customer relationship has been relatively simple. A continuous interaction between the parties didn't overreach monthly bill payments. Today, however, customers are empowered with more choices in how they access energy. Moreover, they are more climate-conscious, efficiency-minded and cost-savvy. This change in customer profile is driving utilities to run their businesses more customer-centric, which means they need to offer more than electricity as a service. According to Market Strategies International report, if given the chance, 39% of utility customers would switch to another energy provider and another 66% believe that utility communications are not sufficient in the US [115]. Big Data analytics, in combination with the enabling technologies mentioned earlier, converts electricity into a bundled service which consolidates the bond between the utilities and customer through personalization. Further, the analysis of the data brings more customer insights to utilities and makes customer

[3]Brazilian shanty town near by big cities.

operations easier.

Advanced customer relations through Big Data applications is particularly suitable for residential end-users since individually generated data[4] can be related to every single resident of households to bring insights. The relationship of a utility and its commercial & industrial customers are improved through energy management systems and tailored DR programs as elaborated in the 7th chapter. Therefore, this chapter focuses on the application related to residential customers.

3.4.1 Leveraging Data to Map the Buildings

The starting point of reaching customers through bundled services is accurately estimating the characteristics of end-users' buildings. The accuracy depends on the variety of data and the performance of the analysis algorithms. The aggregated data mainly consists of:

- Historical energy consumption and weather data.
- Publically available demographics and building related data, comprising of the year the building was built, the materials used for the construction, location, presence of energy efficiency upgrades and so on. The demographic data is about how many residents are occupying the building, the residents' prior purchasing experiences and their eagerness to purchase certain products.
- Once the relationship is established, the end-user can provide very specific data through a platform. For instance, if an analysis shows that the air conditioning system of a household is consuming more than the others, it can ask the customer if there is an issue with the equipment. The data brought by the answer increases the accuracy for further analysis.

The algorithms relate the energy consumption of buildings to the interactions between weather, resident behavior and schedules, building characteristics, and individual devices present in the building [116]. Detecting the individual devices in the building through consumption data is provided by disaggregation[5]. This interconnected calculation of energy consumption captures the physical processes that drive the energy use and enables the actual suggestions that would have an influence on consumption. For instance, these algorithms can associate the energy saving potential of a building by insulation, or changing the windows, changing the thermostat setting and so on. Moreover, as one of the data inputs is generated by end-users, the more they are used the more accurate they get with machine learning algorithms. The algorithms are able to accomplish a task such as *"Identify residential consumers who spend more than 40 percent of their energy on HVAC on days when the temperature is above 30 degrees; who own their homes; have incomes below 30.000; and live in the North Eastern United States."*[116]. Having such a specific insight about the customers, their buildings and energy use enables opportunities for utilities and there is a growing number of companies dedicated to create these services for utilities.

[4] The data generated through e-commerce, mobile phones, social media and so on.

[5] It is a set of algorithms which separate the use of energy into appliances (e.g. refrigerator , television, air conditioning and so on.

Customer Segmentation and Targeting

Segmentation is a marketing term which refers to division of the customers into groups according to their similarities. It is commonly used to reduce marketing costs since the offers can be delivered more specifically. Utilities offer various programs and initiatives to their customers such as rate plans, appliance rebates, home audits and so on. The described *map the buildings* technique involves many insights that allow an effective customer segmentation according to their demographic background and consumption behaviour. Through the segmentation a targeted marketing can take place through the following methods:

- **Automated tip targeting:** Big Data analytics enables recognition of customers who are most receptive to participating in utility programs or reacting to efficiency tips when they are offered. These "better recognised" customers are then proposed the most relevant automated tips through recommendation engine which can take place visually either on a web portal, mobile app or the bills.

- **Utility program recommendation:** The same technique can be applied to promote the best utility program to the customers as well. The programs range from demand response participation to electricity tariff plans. For instance, Iberdrola offers its customers a variety of time of use tariffs, offering different tariffs for different times of the year, week, or the day [117]. If a building is used only during the summer period, the recommendation engine would capture it and propose *Summer Plan* to the customer.

Home Energy Reports

Home energy reports are another value that can be added to individual building insights. They are detailed energy consumption reports which empowers customers to make decisions on their energy use. There are two Big Data analytics enabled features of an advanced energy report:

- **Benchmarking:** Providing customers the insights about their energy usage performance in comparison to their neighbours or buildings with similar characteristics in a significant motivation for to engage them with the provided services.

- **Disaggregation:** The ability to breakdown the total energy usage data into appliances by advanced algorithms. A more accurate division has been possible in recent years which allows customers to benefit through direct feedback and personalised tips.

Pro-active High Bill Notifications

The billing model of electric utilities is based on the payment after the service is provided. The bill, coming with obligation of a payment and information about electricity use makes it the most fundamental contact between the customer and the utility. According to a survey held by OPower with 175.000 utility customers in the US, customers spend four to six minutes to assess their bill immediately after the arrival

[119]. In some of the cases, the bill is assessed as unexpectedly high. The same survey indicates that 72% of customers experienced at least one particularly high electric bill, while 39% adds that it happened in the last 12 months [119]. The reaction to an unexpectedly high bill is either a frustration which could trigger a churn for the utility or an immediate call to their customer service. The volume of the calls for the utility customer services peaks during the week of bill distribution, while 41% of the calls are related to billing [119]. Utilities which are capable of making real-time predictive analysis as described above could increase the customer satisfaction by providing notifications about unexpected high bills before the issuing date. A notification to a customer who is on track to a high bill can also be supported by personalized tips to decrease it and suggestions for the most suitable program participation. Further, reduction in unexpected bills would decrease the volume of customer service calls, resulting in cost savings.

Best Practises: Related Companies and Case studies

OPower (Oracle)

OPower is a cloud-based software solution companies providing services for utilities. The company was founded 2007 in the US and was acquired by Oracle in 2016. The company's Big Data platform stores and analyses over 600 billion meter reads from 60 million end-users, enabling customer satisfaction with mentioned tools above.

E.ON UK is the third largest utility of United Kingdom, serving about 5 million customers. The UK electricity market is highly competitive and the customer profile is in correlation with the empowered customer definition mentioned earlier. According to a survey held by Ofgem in 2013, 43% of the customers did not trust their energy providers to be transparent in their deals with them [120]. This situation encouraged E.ON UK to change their customer relations strategy, aiming at improved trust and satisfaction, less churn and ultimately enhanced company performance.

The new strategy was based on delivering energy saving services with Opower's *Energy Saving Toolkit*. The solution is a SaaS product which is reachable through a web portal. It provides personalized advice by:

- Energy consumption tracking
- Social benchmarking
- What appliance consumes the most
- Energy saving tips
- Energy saving plan [120].

Through these features, the product allows customers to monitor their energy usage and compare with similar sized buildings. The comparison feature promotes a behavioural change and encourage customers to follow the personal tips provided in the application. It also provides easy to read charts with an energy disaggregation. Lastly, the results of energy saving improvements can be shared in social media like Facebook and Twitter.

The web portal was launched in 2013 with a strong communications campaign through social media, TV and cinema advertisement, e-mail and so on. By 2015, a million customer visited the portal. The results exceeds the expectations [120].

- In 2014, E.ON UK website attracted 30 million unique visitors, double the amount compared to the previous year.
- 30.000 energy saving results were posted on social media.
- A significant rise in online product switches was observed.
- The number of customers who manage their account online was doubled.
- Over 1 million customers signed up for rewards programs which promote energy consumption reduction in the peak times.

In conclusion, being the energy advisor of the customers through Big Data analytics, created a big value for E.ON UK.

3.4.2 Utility Market Places

An online marketplace is a website which enables visitors to find products and services by aggregating them from different vendors. In the past few years, online marketplaces such as Amazon, Uber and Airbnb attained a big success as they turn purchasing into a personalized experience for customers. For instance, Amazon recommends the most interesting products in an individual level by tracking the previous purchases of its customers. It also provides *"customers who bought this item also bought"* option which makes other people's choices visible to the user. The concept is favorable not only for B2C, but also B2B customers. According to Forrester Research, 93% of B2B decision-makers prefer to procure products and services through online market places [121]. For an online marketplace operator, the revenue streams are mainly marketing fees, through to advertising and lead generation, and data capture and monetization.

In the meantime, utilities are able to reach out to their customers in a very personal level through the Big Data analytics as mentioned above with products and services for energy efficiency, demand response, distributed energy resources, and more. The combination of these two creates an opportunity for utilities to position themselves as data-driven marketplaces for their customers who would like to be energy efficient. There are already a variety of companies which work with utilities to provide them a white label marketplace. There are two types of marketplaces which is suitable for a utility:

- **Project marketplaces:** They connect customers who are seeking an energy efficiency solution or a technology to contractors. For instance, Serviz is a US based online marketplace that focuses on building maintenance applications. Some of their services include applications which can increase the energy efficiency of the buildings such as insulation.

- **Product marketplaces:** They enable customers to purchase a product or a technology directly for self-installation or they connect them with the contractors who would sell and install the product. From an energy standpoint, these products can be DERs such as solar panels or HVAC units but also home appliances such as televisions, refrigerators and freezers, projectors, washing machines and dryers, dishwashers, air conditioners, lightbulbs, monitors, tablets, and video games consoles.

With an online marketplace, utilities can initiate value-added conversations and interactions with customers on top the existing communication which facilitates the adoption of utility offers and creates new revenue streams. Further, the utility can have more insights about the preferences of their customers. Each online session for a user, in terms of how they search for products can be analysed. Which products they look at, how they sort it, energy calculator preferences they pick and more can be collected in an individual level to create an index to identify the customer's possible interest in energy related utility programs. Lastly, a marketplace is a powerful tool to attract customers to utilities' website. A customer who is digitally engaged to the utility is less costly comparing to a traditional engagement. Table 3.1 indicates the largest utilities of North America and Europe according to their revenues and their current status related to marketplace operations.

Table 3.1: The largest utilities of North America and Europe and their current marketplace operating status

Region	Utility	Marketplace	Marketplace Provider
North America	Exelon	✓	Simple Energy
	Duke Energy	✓*	-
	Southern Company	✓	Simple Energy
	PG&E	✓	Enervee
	American Electric Power	✓*	efi
	NextEra Energy	-	-
	AES	-	-
	FirstEnergy	✓	efi
	Consolidated Edison	✓	Enervee
	Edison International	-	-
Europe	E.ON	✓	Enervee
	Enel	-	-
	EDF	✓	Enervee
	Engie	-	-
	RWE	-	-
	SSE	-	-
	Centrica	-	-
	Iberdrola	-	-
	Gas Natural Fenosa	-	-
	EnBW	-	-

* A limited marketplace only for rebated products, not a full scale service

Best Practises: Related Companies and Case studies

Enervee

Enervee is a California, US based SaaS company that combines data science, behavioral science and digital marketing to drive consumer energy savings. The company provides white label marketplaces for utilities which are open to their customers who are interested in making energy-efficient home appliances choices. In this sense, Enervee is a product marketplace provider for utilities. The company's core technology is the *Enervee Score*. It is an index on a 0-100 scale which compares energy efficiency of a certain product model with all other products available in the market that have similar features. Unlike, EU energy efficiency labels, Enervee Score is updated daily. For instance, if a new television is launched with an outstanding energy efficiency potential, the algorithm behind the index will compare it to the others and the older models' scores will decrease with the introduction of the new product in the market. The company combines data from product offerings, energy consumption profiles, utility rates and rebates. Figure 3.6 shows a screenshot from a utility marketplace powered by Enervee. Once a product is chosen, a "clear cost" is determined by taking into account the energy use and life time of the product, in combination with utility electricity tariffs. Further, motivating insights about the use energy efficient product is provided such as the direct impact to environment with numbers [122].

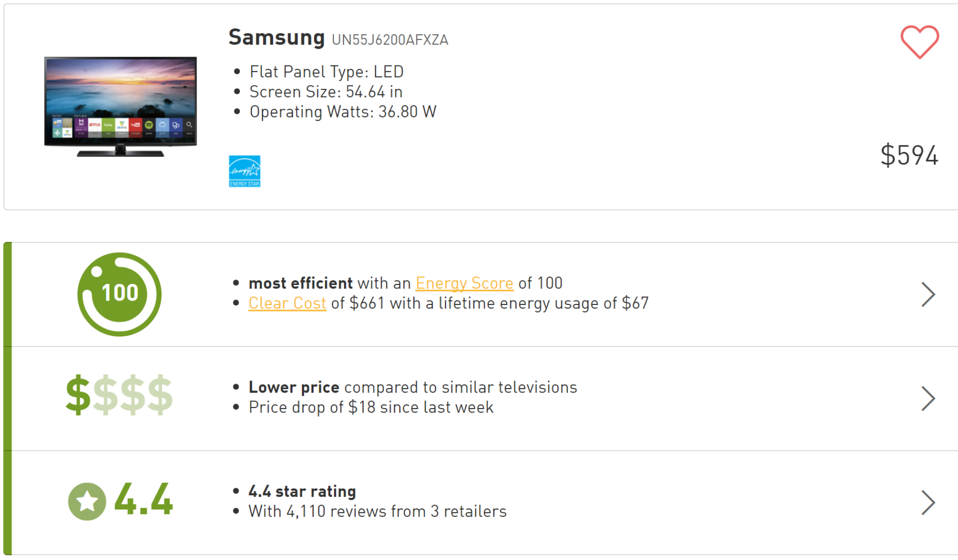

Figure 3.6: Screenshot of Enervee Marketplace [123]

The second feature of the marketplace is the *Customer Engagement Score*. It is an index which determines the level of interest towards energy related products and services among the customers of a utility. Enervee achieves it by collecting 200 event data on the marketplace from each visit. The score is not a standard measure, but a customised one [122]. For instance, the utility can prioritize the interest towards certain products such as solar panels. The results will rank the customers according to their interest accordingly. Another priority could be the score related to energy efficiency programs. In this case, events such as sorting by Enervee Score and overriding the defaults in the energy saving calculator could be used to analyse the customers.

The marketplace is attaining attention from the utilities in the US and Europe. Avista, Con Edison, Duquesne Light, Eversource, Pacific Gas & Electric, San Diego Gas & Electric, Snohomish PUD and United Illuminating from the US; E.ON and EDF from Europe have already adopted the marketplace [122]. By the end of 2016 the marketplace achieved:

- More than 22 million households served.

- More than $22 million's worth of products sold.
- Average appliance purchase value of $1.078.

Simple Energy

Simple Energy is another SaaS company based in Colorado US. They combine behavioural science, Big Data analytics and digital marketing techniques to create a platform for utilities to engage with their customers through energy efficiency. Recently, the company added a utility marketplace solution to their product portfolio. Their platform consists of three main features. The first one is *Energy Insights*. It is a combination of personalized energy saving tips and targeted utility program recommendation. Second feature is *Energy Community*, a dashboard dedicated to benchmarking of household energy efficiency performance. It allows customers to compare themselves with their peers which is a motivation for further engagement. The last feature is *Energy Rewards*. It is a marketplace with local products and services. Customers, using the platform collect energy saving points when they participate in utility programs. These points can be used in this marketplace to redeem services. Some examples of the offered rewards are tickets to a local cultural event or a dinner in a restaurant [124]. Their latest product utility market place allows engaged customers to the purchase efficiency products such as smart thermostats, advanced power strips and so on. There are two main differences between Enervee and Simple Energy. The most fundamental difference is that Simple Energy's online marketplace functions as an e-commerce website which positions the utility as a vendor while Enervee connects utility customers to multiple vendors. This allows Enervee offer a larger variety of products. Secondly, Enervee's pulling power is the intuitive energy efficiency performance index for a variety of home appliances and electronics while Simple Energy uses its know-how on targeted segmentation and marketing, resulting from the engagement platform, to drive the customers to use the marketplace.

Xcel Energy is the 10th largest utility of the US in terms of the number of customers served. In 2015, the utility launched its marketplace with the objective of selling smart thermostat's which would be a part of their demand response program. Simple Energy's engagement platform was used to offer the demand response program to the targeted customers. The platform then directed interested customers to the marketplace. A rebate was applied to the customers who are willing to participate in the demand response program offered by Xcel Energy. The pilot stage took 18 months. By the end of 2016, the program exceeded the expectations:

- Over 8.100 smart thermostats were sold via the marketplace (with limited marketing) which accounts for 75% of the total thermostat sales of the utiltiy.
- Over 2.000 customers participated to the demand response program which is higher than the objectives of Xcel Energy.
- 90% of the customers stated that they were satisfied with their purchasing experience [125].

Commonwealth Edison (ComEd) is one of the largest utilities in the US, serving over 3.8 million customers. They launched Simple Energy's marketplace with a market campaign during the *Black Friday*

Week in 2016. The sales volume during on week period overreached $1.3 million with 12.000 total products such as smart thermostats, advanced power strips, lighting, and connected home products. The campaign showed the success of targeted marketing, an easy way of shopping and online rebates to boost sales [126].

3.4.3 Analysis of Other Sectors

Utilities provide services to millions of customers through a continuous relationship. As a result, a vast amount of customer data is accumulated in their systems. From this perspective, there are several more sectors which are similar to utilities. Banking and telecommunication stand out in terms of their similarity to the energy sector. Traditionally, they are both focused on providing a commodity service and are both currently being disrupted by digital transformation. Therefore, an analysis of how they are using Big Data analytics gives insights to the energy sector. However, the scope of the analysis must be limited to the applications which bring value to their customers since the customer related part of the value chain has similarities with the energy sector. Therefore, it would be irrelevant to analyse other parts. For instance, fraud detection is used commonly in both sectors, although it doesn't give any insight to the energy sector. Further, another limitation is if the application brings value to the customer or not. A common application in both sectors is to identify potential customers according to their demographic data. This is clearly beneficial for the company but it doesn't have a remarkable value for the customer. Therefore, the scope of the analysis is identifying the applications which bring value to the customers by leveraging their data. Lastly, an analysis of other sectors which hold a large amount of data from their customers is conducted. These are mainly social media networks, online retailers, internet television networks. The big players of these sectors are very good examples of how Big Data analytics can empower customers and bring engagement. Therefore, the pioneer companies of these sectors are analysed under the "Examples from Other Sectors" section.

Telecommunication Sector

Telecommunication is one of the most suitable sector for Big Data applications due to the amount of data they have. The operators collect data from many different sources including their customers demographics, the devices used, location, service usage and the data generated by customer services [127]. Big Data analytics is used to create various applications throughout the value chain of telecommunication sector such as improving network operations, fraud detection, call drop analysis. However, customer related applications are elaborated below:

- **Churn Prediction:** Customer churn rates are considerably high in the sector, ranging between 20-40% [127]. Losing a customer has an obvious negative impact for the operator (less revenues and market share) and capability of preventing it is very valuable. However, from a customer experience standpoint, the period of time which leads to the churn point is unpleasant, and thus a prediction this dissatisfaction is even bigger value for providing better services. To achieve this, data collected on customer usage, complaints, transactions, social media is analysed and the customers who are on track to leave the company are determined. As the analysis gives insights about the reasons of dissatisfaction, personalized offers can be made to the customers. For instance, a customer who considers leaving the company due to his old and expensive contract can be offered a new contract with more advantageous tariffs. T-Mobile USA, one of the leading companies in the country, analysed the frequency of calls to contacts who are using a new operator [128]. An increase in the frequency was a sign that the customer's circle is switching to other operators and soon the customer might too. The company offered the same price for all calls to these customers in order to prevent the churn. This also prevented the customer from the hassle of changing the operator. The company succeeded a decrease in the churn rate by 50% in one quarter .

- **Customer Segmentation:** An individual level of segmentation is made in telecommunication. The segmentation is used for targeted marketing and recommendations, similar to the energy sector as mentioned above. The recommendations are mainly a personalized response to analysed usage data. The *Rocket Recharge* service mentioned in the Telenor India case study represents a good example of this application. There are several different applications, following the segmentation. Telecommunication customers receive a vast amount of information from the operators through channels such as e-mail, SMS, phone calls and so on. Receiving unrelated information is highly disturbing for the customers. Using analytics, telecommunication companies reach out to their customers with relevant information through preferred channels. For instance, if a customer sends request by SMS to the operator to find out his balance frequently, the algorithms can capture this and start to send him daily balance sheet via SMS. Further, the capability of keeping the track of the usage data in real-time allows self service platforms for customers.

- **Location Based Services:** A real-time geographic location information provides many opportunities for operators as it is the direct link to customers' daily life habits.

Case Study: Telenor India

India is one of the most competitive telecommunication markets in the world. There are more than 10 operators that serve 1.21 billion customers. Moreover, the market is extremely price sensitive which leads to ownership of more than one SIM-card per customer. Consequently, identification of high value customers and mitigation of customer churn are essential tasks for telecom operators. Telenor India is a big player in the market that focuses on pre-paid segment. The company serves over 53 million customers [129]. However, Telenor figured out that a big portion of their customers prefer them as the "second SIM-card". To be more preferable, the company decided to change their relation with the customers by implementing data-driven applications. Taking into account the data characteristics needed to develop these applications, they partnered with TERADATA for implementation of Big Data solutions. The applications were mainly creating a self service customer management platform and providing value added services.

- *Self Service Manager* is a platform which enables customers to reach out the relevant data about their use of service and run some operations with near-real-time data. For instance, customers can check their balances, receive data settings, find stores, check complainant status and so on. Having a personalized contact with the customer allows Telenor India to offer their customers the most optimum recharge option. The service is called *Rocket Recharge*. Once the customers dial $*$234#, the system analyses individual usage patterns and recommends the optimal recharge amount. Further, the system proposes additional bonuses such as extra talking time or SMS depending on the customers historical usage data [129]. A combination of recommended recharge amount and additional bonuses slowly transforms the company into a trusted telecommunication partner. Since Telenor's target is not the most possible recharge amount but it varies in every dial according to their usage, the customer builds a trust bond. As a result, the revenue received from recharge via *Self Service Manager* increased by 25% with the introduction of *Rocket Recharge*.

- Another value was created for the retailers. Telenor India has around 2.000 franchisee retail stores where customers can recharge. Using the location data, the company was able to determine the customers who live $<$ 5km to these stores. The data includes information about who has low balances, who are most likely to visit the store on a specific day, or who would be interested in current promotions, and was shared with the retailers. This allowed retailers to make calls to near-by customers to offer recharge. The retailers were incentivized with a 7% commission for each recharge. Four months after the start of the application 14% rise in customer visit was observed, resulting in a revenue increase both for the company and its retailer partners, while the customers were individually informed in a local level. A decrease in the blind calls also contributed to the reduction in the negative "spammer" perception of Telenor among the customers [129].

Banking Sector

Banking is another sector which has a vast amount of data related to their customers, and thus Big Data analytics applications in the sector gives insights to energy sector. The data used in banking mainly comes from ATMs, call centers, branches, credit cards, debts and financial business forecasts from various sources such as news, industry data, trading data regulatory data and analyst reports. The customer related applications of this data is similar to telecommunication. The banking sector also takes personalization to the center, in order to bring better services. This is due to the similarity of their business. They both are highly interconnected with the customers and use similar channels. The list below provides several use cases related to Big Data applications in banking.

- **Personalized Loyalty Programs:** Loyalty programs have been used in the sector as tool to retain customers. The method is based on bringing more benefits to the customer as long as they use the banking services. Analysing the transactional data, several banks determine the spending habits of their customers and create personalized loyalty programs. For instance, The Royal Bank of Canada offers promotions in pet shops to their customers who has previous pet related purchases. Singapore Citibank offers customer discounts at retailers and restaurants where the customers are frequently visiting [128].

- **Easier Use of Services:** Similar to telecommunication, customers of the banking sector use services very often. Therefore, there are various banks, intending to make these services more easy-going for customers. For instance, HDFC bank, from India, records the individual choice of language in the ATM use for each customer. In the next visit of the customer, the ATM operates in the previously used language. The bank shortened the use of time by 40% [130].

Open Banking

Banking is one of the sectors that was significantly disrupted by digital technologies, mainly by fintechs[6]. These are software companies, ranging from multinational scale to start-ups, which use analytics to create financial services for people. Paypal is one of the well-known fintechs, facilitating financial transactions for individual and SME[7] users. Another example is Robinhood, which is a mobile app, enabling people to invest in publicly traded companies. Citigroup estimates that the revenues of fintechs will exceed $200 billion in North America alone by 2023 [131]. These non-bank players who are based on creating services to customer in an individual level is pushing traditional banks to a new concept: Open Banking, which is based on banks sharing their customers to third parties in order for them to create new services. Such a partnership enables development of Big Data applications by not only the banks themselves but any developer who has a good idea, creating even bigger value for customers. The technology behind is called application program interface (API). It is a secured data sharing connection between companies. For instance, Uber uses Google Maps API to operate [132]. This emerging concept is adopted in different ways. For instance, BBVA, one of the largest banks of Spain recently launched *BBVA API Market*, a platform which is open to any developer from Spain to use their data [133]. Nordea, a bank from Norway is planning to create an ”App-store” dedicated to financial services. The last approach is more conservative, keeping the data between trusted partners rather than putting it to an open platform. The list below provides highlighted use cases of open banking.

- **Personal Finance Management Platforms:** These platforms help users to track their income and expenditure. The data obtained from different banks can be combined to give insights about customer’s spendings. Further automated advice that rely on Big Data analytics can guide the users about investment products available in the market.

  Moneyhub is a British based personal financial management tool company that aggregates data related to customer’s bank accounts, credit cards, investments, savings and borrowing, including property and pension in one dashboard [134]. The predictive text analysis of the platform categorises the expenditures to give a better understanding about user’s spendings. Later, the spendings are compared to other months and categories to give insights to user. For instance, if the user puts in a saving target for a certain date, the application can notify the user by stating: ”There are 10 days left this month. If you spend less than 12 euros per day, you will be able to reach your saving target”.

- **More Transparent Credit Scoring:** Traditional credit scoring is limited in terms of capturing the

[6] Fintech is an abridgement for financial technology.
[7] SME is an abridgement for small and medium enterprises.

real performance of a customer due to the limited data. A better analysis of individuals is possible by combining bank-owned data with third party data. A more transparent credit scoring enables a better match between the customers and the available products for them.

Credit Kudos is an innovative credit reference agency, focused on people who do not have enough financial history either due to their age or if they are new to the country, etc. Such a customer usually either gets rejected or is offered an expensive product. Credit Kudos collects the data from the customer by a audit. On top of the customer provided data, it connects to an online banking account. A combination of these two turns into a credibility score which allows the customer to reach out to banking products [135].

SafetyNet is an online lending service for overdrafts. Normally, if an account reaches a negative point, the bank issues an overdraft loan which has high interest. SafetyNet, using the customer data, issues a loan with a lower interest before the automatic loan occurs. The company automatically recollects the loans once the account has sufficient funds [136].

- **Integrated Banking Services:** In the last 1-2 years, more applications were created by using banking APIs. These allow customers to access banking services from different platforms.

  In the USA, Facebook users are able to transfer money to their friends on messenger [137].

  Amazon Echo is a hands-free speaker which connects to internet to access Amazon enabled services and more. The partnership between Capital One, a bank operating in the US, and Amazon allows customers to access banking services from their device with their own voice [138].

  O2 Deutschland is a German based telecommunication company. They initiated a mobile only banking service which allows customers to access to small amount of loans, send or receive money by only their mobile phone number. The company is using the API from Fidor Bank, an online banking company operating in the country [139].

Even though the Open Banking concept is at its early stage, the recent innovative services prove its potential. By creating a platform, banks are able put their label on the customer centric services which leads to a better customer engagement. Moreover, they create revenue streams either by direct monetization of data or indirect revenues such as a commission on every transaction occurs in the bank owned platform [132].

Examples from Other Sectors

Amazon

Amazon started operating as a online book store in 1994. Today, it is the largest online retailer of the world whose service vary from online media consumption to supermarket shopping. Moreover, they brought hardware products to the market which allows a ”physical engagement” with the customers. Undoubtedly, one of the main factors of the company’s success is the use of Big Data to create customer-centric services.

Amazon gathers the biggest possible variety of products under the same roof. However, accessing this large amount of products at once is overwhelming for customers. Psychologically, customers might give up with the purchase, feeling unsure if they are choosing the best product. The solution Amazon created to tackle this problem is the recommendation engine. It turns the customers data into insights to recommend the best option. For instance, data is converted into ”Employed males between 18 and 45, living in a rented house with an income of over $30.000 who enjoy foreign films, are also tend to purchase projectors.” insight. Amazon uses these insights to deliver innovative services to its customers listed below [140].

- Amazon Echo is a hands-free speaker which also features as a digital assistant. It connects the customers to their Amazon account for easier purchases. Moreover, it can distinguish individual voices which leads to multiple use by different accounts at the same time. On top of being the access point to the account, it creates value added services. For instance, *Ask My Buddy* is a feature which automatically calls the trusted contacts of the user in case of an emergency. The platform is open for anyone to access the data, allowing them to develop customised services.
- Amazon Echo Look is another device recently launched. It is a camera addition to the previous model which gives customers insights about their outlook. The user asks the device to take a photo of himself. The photo is then used to compare with the latest fashion trends or the conventional style of the individual. Device is then able to offer new products to the user. For instance, it is capable of stating ”The colors you chose are quite out-fashioned. Considering your style, would you like me to purchase this pink colored skirt for you?”.

Amazon is maybe the best example of how a company can diversify its services by making use of their customers data.

Chapter 4

Flexibility Management

This chapter aims at revealing the applications related to flexibility management. It starts with flexibility management methods and their current status. Later, the relation of Big Data analytics with these methods are analysed and related business models are explained.

The transformation in the energy sector is driving a change in the composition and the operation of the power grid. Flexibility of the grid, the degree of balancing the generation and consumption ability at all times, is also being affected by this change. The traditional flexibility management includes the monitoring of the supply side and adjusting the generation accordingly. However, this method is not sufficient with the growing decentralised structure.

The disturbance that is brought to the grid by this transformation can be investigated in two levels: system level and local level. In the system level, the electricity generated from renewable sources has priority to be fed into the grid.

The local level of disturbance occurs in the distribution grid due to the fact that all of the distributed generation (DG)[1] units and some of the renewable energy sources are connected to it [85]. These are mainly voltage problems, reverse power flows, congestions and grid losses.

- Voltage problems happen due to the fact that DG units connected to the medium and low voltage networks change the profile in their surroundings by the injection of active power. If there is not enough demand, the voltage would locally increase, resulting in a risk in customer assets.
- Reverse power flows occur when the local power generation overcomes the local load. However, the existing infrastructure is not designed to handle this phenomena. This brings a stress to the grid, and hence decreases the reliability.
- Congestions occur when the flow exceeds the design capacity of a grid component. This is observed when a sudden load is introduced to a certain area, e.g. implementation of electric vehicles

[1] The term refers to power generating units connected to a distribution grid, such as combined heat and power (CHP), small size solar, wind and hydro plants.

[85]. This problem is mainly solved by system upgrades at a local level. However, the pace of improvements can not catch up with rapid adoption of local power generation. Moreover, it is not always cost effective to make an upgrade to a component which is rarely experiencing a congestion.

- Grid losses refers to the DG impact on all the flows in the grid and they have a dynamic behaviour, depending on the power produced. If the local production is close to the demand, it decreases the grid losses. However, if the DG location is remote, it results in a increase in the grid losses [85].

Taking into account the system and local level problems related to grid operations, the sector urges flexibility for a low cost operation and increased reliability. The mitigation of these disturbances are achieved by providing firm capacity and an operating reserve. Firm capacity refers to the precise available load for a certain time, while operating reserve is the short-term available capacity which becomes crucial due to fluctuated profile of generation. There are two approaches to increase the flexibility capacity. The first approach is demand response (DR) and it has been applied for decades. However, Big Data analytics takes it to the next level by enhancing predictability and controllability. The second approach comprises of aggregating DER's in order to create a load capacity and offer it to energy markets. DR is in fact considered a DER asset. However, differentiating it while defining Big Data applications gives a better understanding since it is based on demand reduction while other DER assets are related to power generation. Big Data analytics makes aggregation of DER's possible while increasing their participation in flexibility management with more advanced models.

4.1 Demand Response

Demand response (DR) is a set of programs that motivate residential, commercial and industrial (C&I) consumers to reduce or shift their electricity consumption during peak periods [86]. The use of DR started in 1970s as a collaboration between large (C&I) parties and utilities. The utility would warn a large industrial customer by phone to reduce its electricity consumption without any feedback from the customer about their operational performance. It was a tool to mitigate demand peaks or recuperate system emergencies. The liberalization of energy markets changed the role of DR, making it an important tool to reduce wholesale electricity prices, and therefore more parties were involved. There has been many DR programs suggested in literature which can be reduced to two main groups, namely price based DR programs and incentive based DR programs [87].

- Price-based DR programs involve variation of the price over time in liberalized markets. The customers are motivated to adjust their consumption to reduce the peaks. In other words, the price of electricity gets higher during the peaks [87].

- Incentive-based DR programs are rewarding systems in which the customers benefit from reducing their electricity consumptions upon request. Depending on the agreement, the utility might have a level of control over electricity use of the customer. There are mainly six types of incentive-based

DR programs, namely direct load control, interruptible/curtailable demand bidding, emergency DR, capacity market programs and ancillary services market programs [88]. Table 4.1 summarizes all available DR programs with their explanations.

Table 4.1: A summary of demand response programs

Price-based	**Incentive-based**
Time of use: An average cost of electricity generation and tranmission & distribution is reflected in the prices depending on the time of the day **Real time pricing**: The price of electricity changes hourly during the day. The customers are notified a day or an hour before the changes. **Critical peak pricing**: It normally works with time of use principle. The only difference is that when some specific conditions appear such as system reliability, much higher prices are valid for peak times.	**Direct load control**: The utility is able to shut down or shifts the demand of the customer. It is more convenient for residential of small commercial customers. **Interruptible/curtailable**: In this system, the customers guarantee to curtail their consumption upon request. They receive a discounted rate or direct payments in return. The typical customers are mainly large industrial partners since their capacity to influence the grid is higher. **Demand bidding**: Customers propose a demand reduction capacity for certain times and the price to utilities. **Emergency DR programs**: Direct incentive payments are offered to customers for their load reduction when the shortfall arise. **Capacity market programs**: A load curtailment from the customer offerings is used to support system capacity, instead of excess generation. The incentives are up-front reservation payments and the agreed customers face penalties in case of not obeying the agreement

DR programs are beneficial for all the parties involved. The customers benefit either by increasing their revenues with incentives payments or by reducing their bills. The most expensive electricity generation means are in use during the peak times. DR lowers wholesale electricity prices by shedding the peaks since the costly generation means are utilized less than expected. Furthermore, a lasting DR program enables utilities to invest less in aggregate system capacity requirements, allowing further savings on bills over the long term. DR programs also enhance the reliability of the grid. Lastly, they relieve congestion, manage contingencies and avoiding outages in the grid [88]. In spite of their benefits, the adoption of DR programs is hindered by some barriers:

- Aggregation of load is one of the most important concepts for penetration of DR programs. An aggregator is an actor who is able to create a pool of load out of smaller resources. It allows smaller players such as small commercial and residential customers to be engaged in the programs. A survey conducted in UK with 191 commercial and industrial customers indicates that 83 % of them are using an aggregator to participate in DR programs [89]. However, aggregation services are not legally available in all member countries of European Union. The countries are divided into three group according to their legal status. The first group of countries do not allow an aggregator at all. This group involves Portugal, Spain, Italy, Croatia, the Czech Republic, Bulgaria, Slovakia, Hungary, the Baltics, Cyprus and Malta. The second group allows the retailers (utilities) to be aggregators, including the countries Germany, the Nordics, the Netherlands and Austria. Lastly, the group of countries who allow any actor to be an aggregator are Belgium, France, Ireland and the UK [90]. Even though the concept is not widespread throughout Europe, the legislation is expected to change in country level in the close future.

- The real value of DR can only be uncovered with the participation from the end-users. The barriers concerning rate of adoption towards DR programs are mainly financial turnover insecurity and the behavioural change reluctance [91]. It is important to make the financial benefits visible for the end-users to convince them to participate. Besides, an optimized load reduction with the minimum influence on daily activities of the end-users needs to be obtained.

4.1.1 Utility Focused Big Data Applications: DRMS

Big Data analytics, in combination with IoT, enables aggregation and visibility of data. Data gathered a from utility's advanced metering infrastructure (AMI), meter data management system (MDMS), and legacy systems such as SCADA, customer information systems (CIS), customer billing data, weather data, and geographic information systems (GIS) can be used to optimize and manage DR programs from one center which can be either be implemented on premise or as software as a service (SaaS) [92]. Such systems are called demand response management systems (DRMS). On premise implementation is a model in which the data storage and service hardware is owned by the customer while in the SaaS model, the data is stored in a cloud which is provided by Big Data services. The SaaS model is widely preferable comparing to on premise configuration due to its faster implementation time and lower costs. A SaaS can be implemented within two months and less than $100.000 while an on premise implementation might take up to a year with over $1.000.000 investment [92]. In 2016, 46.1 million dollars were spent on DRMS and it is predicted to reach 232 million $ by 2025 [92]. The benefits of DRMS, enabled by Big Data analytics, will be discussed in the following section with the barriers, hindering its adoption.

Benefits of DRMS

Increased DR Load Impact

The improvements of DR load impact can be divided into 3 sections, namely increased reliability, augmented load shed per participant and higher participant response to the programs [93].

Increased Reliability: The most important attributes of a load shed is dispatchability, predictability and repeatability. Once a load shed satisfies these, it can replace an equivalent amount of generation. This type of load shed is called dispatch grade. It is possible to upgrade the DR portfolio to dispatch grade by analysis of data.

Increased Load-shed per Participant: The best way of increasing the load shed per participant is to offer them tailored programs. Some customers may prefer peak time shedding programs while others participate in direct load programs. Having a better understanding of the customer portfolio increases their satisfaction and thus their loyalty to the program. The customers can be segmented using GIS, building size, usage profile and demographic data [93].

Higher Number of Participants: An efficient way of increasing the number of participants is to offer them some services beyond low prices or incentive payments. This can be achieved by advanced analytics. The first opportunity is to show the participants their benefits in real-time. This could overcome the hindering factor related to financial benefit insecurity. Besides, a management system which gives actionable insights to the customers about their systems (i.e. malfunctioning warning) increases the participation.

Reduced Program Costs

The programs costs are related to customer engagement, marketing, technology infrastructure and integration and Big Data analytics offer solutions to each of these cost items [93]. The increased load-shed per participant, in combination with tailored marketing techniques lowers the engagement costs. Measurement & verification is a set of statistical research tools to validate the load shedding impact of a DR program. Traditionally it is realised by using the data received from spreadsheets. Such methods bring administrative costs, high error rates and long waiting times. One software platform to manage DR programs mitigates these inconveniences, and therefore reduces the costs. Lastly, accumulation of different infrastructure for each program results in complexity and redundancy. Gathering entire data in one software by using Big Data tools can standardize programs and facilitate the integration.

4.1.2 End-user Focused Big Data Applications: EMS

Energy Management Systems (EMS) focuses on cost effective energy efficiency solutions for residential and non-residential end-users. Intelligent efficiency, a term highly linked to EMS, is defined as the ability to save energy by gathering large volumes of data [94]. It consists of all products and services that monitor, control or analyze energy consumpion. These solutions leverage data from traditional control and automation systems, smart meter interval electricity consumption data, supplemental submeters and advanced sensors or other business intelligence offerings. The energy efficiency applications are provided by controlling end-user's applications such as heating, ventilation and air conditioning (HVAC), lighting and plug loads citation as needed. Big Data analytics, using predictive and prescriptive analysis, doesn't only increase the effectiveness of the service but also offers new solutions. Intelligent software solutions improve energy efficiency by;

- **Visualization and monitoring:** An interactive dashboard provides periodic summaries of energy use which can be visualized as a heat map.
- **Fault detection and diagnosis:** The ability to simultaneously monitor and analyse real-time data from multiple assets allows benchmarking. The comparison of the end-users and pattern learning algorithms can detect a malfunction and alert the end-user.
- **Predictive Maintenance & Continuous Improvement:** The predictive analysis determine the potential future malfunctions, allowing better capital planning.
- **Optimization:** The historical data related to energy performance can be prescriptively analysed to optimize the end-users activities.

Based on the solutions above, two opportunities for energy efficiency become prominent. Firstly, end-users receive greater information about their energy consumption, as well as tips to reduce their energy use. This is a motivation for behavioural change. Secondly, the actionable knowledge received by analytics can be used by automated systems to optimize energy use. Needless to say, having an intelligence control over the assets makes for smooth integration into DRMS systems, and allows them to be valuable DR assets for advanced DR programs. Therefore, many companies in this field offer energy management solutions to residential and non-residential end-users to engage them into a DR program.

4.1.3 Findings from Related Companies and Case Studies

Having analysed 23 companies that provide DR services for utilities by using Big Data, several trends and business models were identified. The companies range from multinational technology providers to startups. The sector is fairly dynamic, facing acquisitions, merges and partnerships. These are happening either between the DR providers, a DR provider and a hardware manufacturer (mostly HVAC) or a DR company and a utility. The latest example is the acquisition of EnerNOC[2] by Enel[3] earlier this year. Such dynamism, particularly the interest of utilities towards intelligent DR, proves its value.

Figure 4.1 presents a comprehensive overview of the business relationship between the DR providers, end-users and utilities. The term DR provider refers to companies who use Big Data analytics in order to provide DR services to utilities either by turnkey sales of firm capacity and ancillary services, implementation of DRMS or an energy efficiency SaaS which connects end-users to utilities. To better understand the subject, the providers are grouped into two categories depending on their customer segment (Industrial and commercial focused and residential/behavioural focused DR providers). This categorisation is necessary since not all the DR providers serve entire electricity end users.

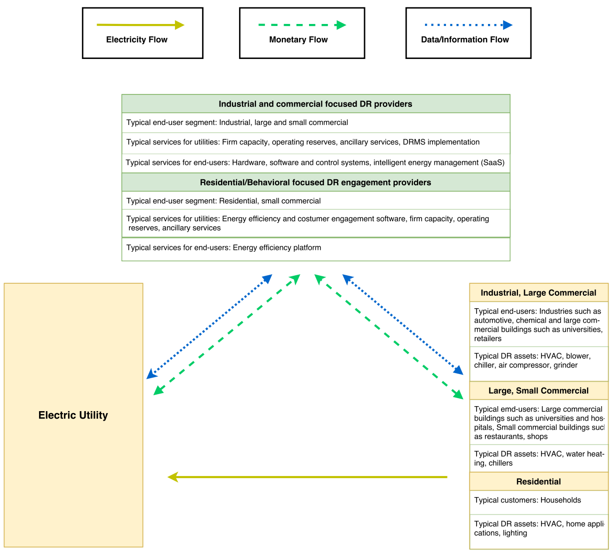

Figure 4.1: Business relation between the DR providers, electric utilities and end-users

The business models of all DR providers have similarities to a certain degree. They aim at the highest possible end-user acquisition rate which would turn into load reduction for utilities. They are positioned

[2] A US based DR provider.
[3] A major Italian electricity utility.

as a mediator between utilities and the end-users, controlling the data/information and monetary flow. However, there are fundamental differences in the way they achieve it, particularly in the methods of customer engagement. Moreover, the improving Big Data analytics abilities bring new players to the market with different business models and value proposals. The first group, industrial and commercial focused DR providers are the largest and the most mature group, consisting of some companies traced back to early 2000s. There are mainly three different business models inside this group:

- The first business model provides intelligent energy management systems for industrial and large commercial end-users by software that optimizes their businesses energy consumption. This software gives the businesses the option to participate in DR programs and the provider sells the firm capacity and ancillary services to utilities as a turnkey solution without any further interference. Their revenue streams are subscription fees for their intelligent EMS and the brokerage fee that they agree with the utility for the load reduction.

- The second business model is provided by DRMS companies. As mentioned earlier, the software enables a dispatch grade load reduction capability by centralization of whole DR portfolio, real time monitoring, tailored DR programs for end-users and faster M&V reporting. The integration of the DRMS happens in different levels; in some cases it is just the software implementation, either as a SaaS or on premise, in others it is implementation as a turnkey service including entire life cycle of a DR. For instance, Comverge took responsibility of introducing the advanced DR concept to Eskom, the largest utility of South Africa, providing 95% of the country. They did not only implement the DRMS but also defined its borders with the know-how they have [95]. These companies develop patented algorithms which uses the DR assets as much as possible without discomforting the end-user. Their revenues are directly linked to utilities since their service does not involve end-users like the previous group. The revenue streams are either software sales or subscription fees.

- The companies who produce hardware such as HVAC systems are also active in the field. DR is a very advantageous business opportunity to them as DR participation of a HVAC system increases its value proposal, and this works both ways. Therefore, many companies offer an intelligent commercial/residential EMS system which can be embedded in utilities DRMS to increase the DR participation and customer engagement rates. However, the model is not limited to hardware producers. For instance, a business model called *Bring Your Own Thermostat* allows HVAC owners to connect their asset with the utility through a thermostat and many DRMS software are designed, compatible to this trend. These companies make revenue with software sales of either energy efficiency SaaS, or entire DRMS on top the hardware sales.

The second group, residential/behavioural DR providers, are the new players in the market. The vast majority of them are start-ups which were founded in the last 5 years. However, they are growing fast, receiving a large amount of investment. Their value proposal is similar. They provide an energy efficiency platform for residential end-users which gives actionable information about their energy consumption and tips to reduce it. It encourages end-users to participate in DR programs through behavioural change. The reliability of DR is provided with learning algorithms which conduct an analysis to predict the participation of every house. They receive revenues through software sales to utilities. It is important to mention that the real value of these companies are the customer insights that a utility can receive. The subject is elaborated under the customer relations chapter.

Best Practices: Related Companies and Case Studies

EnerNOC

EnerNOC is a Boston, US based energy intelligence software company, founded in 2003. It was acquired by Enel in 2017. It is one of the most important actors in the market and a good representative of the first group C&I focused DR provider. The company is using Big Data analytics for energy management in commercial and industrial customers. Their software Energy Intelligence for businesses won the ”Most Insightful: Big Data and Analytics Innovation” award for having the most accurate energy cost accruals that enable proactive decision making and a predictive algorithm, enabling users to reduce demand charges [96]. The expertise the company has in the energy management field allows a diverse portfolio of DR program participants. Combining this ability and their energy market knowledge makes them a valuable company for DR program implementation.

Southern California Edison Company (SCE) is the largest utility operating in California, serving more than 14 million people over a 130.000 km^2. They collabrate with EnerNOC to receive a 110 MW curtailment. All of the participants are from the C&I category since EnerNOC energy intelligence software is suitable for this type of customer. A demand reduction of 100 KW per participant is achieved and the participants are receiving up to $25.000 incentive payments annually [97].

PNM is the largest utility of New Mexico with 499.000 customers. Their summer peak reaches up to 1.900 MW in hot summer days. The utility uses EnerNOC’s turnkey load curtailment service. As a part of the agreement, EnerNOC implements its energy intelligence software which allows the company to analyse and remotely control C&I customers’ energy consuming units such as lighting and HVAC which results in a 30 MW of load curtailment [98].

Midwest Energy is a utility which operates in Kansas. They have 48.000 electricity and 42.000 gas customers. The utility partnered with EnerNOC to start an automated agricultural DR program. The utility serves approximately 2.000 agricultural customers whose peak consumption is over 30 KW and it accounts to 45 MW peak irrigation load. The program achieves a 20 MW load curtailment by shutting down the irrigation pumps automatically. Moreover, the Big Data analytics enables program participants to monitor their energy consumption and savings in real-time [99].

Comverge

Comverge is founded in the US in 1997 as a energy management company. It provides hardware and software for energy efficiency and management. IntelliSOURCE, the company’s DR management system, starts with customer acquisition in which registered customers are provided with online appointment scheduling for device installation. Later, the software automates all processes related to device deployment. The core of the software is the control function, which enables a dispatch-grade load shedding through analytics of real-time operating status of all demand side resources [100].

Pepco Holdings Inc is a holding consisting of several utilities such as Pepco Maryland, Dalmarva Power Maryland, Atlantic City Electric, Pepco District of Columbia and Dalmarva Power Delaware. The total number of customers reaches over 2 million. Pepco Holding partnered with Comverge to expand its DR programs. The pulling power of Comverge is its expertise in hardware, related to DR such as smart thermostats, HVACs and control switches. They handled the marketing of the programs, installation of the devices and integration of DRMS in the utilities. The program achieved 390.000 participants from residential and C&I customer segments with a load curtailment capacity of 360 MW [100].

Eskom is the largest utility in South Africa, generating 95% of the electricity in the country and 45% of the entire continent. The total customers served reaches over 5.5 million. There has not been any DR activities in South Africa until the program was launched with the help of Comverge. The need arose due to the old infrastructure of the grid, and therefore a very tight reserve margin. Since it was the first open market DR attempt, all the rules had to be decided and managed at once. Hence, IntelliSOURCE, the DRMS of Comverge was used to implement the program rapidly. The software helped Eskom to qualify and register the C&I customers, including their load profiles, to schedule and dispatch DR with real-time performance monitoring and M&V of curtailment data. The result was 300 MW load curtailment with a rate of 3000MWh per month [95].

4.2 Distributed Energy Resources

Distributed energy resources (DER) is a comprehensive definition for entire resources connected to distribution grid. It includes various assets which can be categorised into four groups: distributed generation, distributed energy storage, demand response and microgrids.

Distributed generation (DG): This group consists of all the power generating assets connected to distribution grid:

- *Combined heat and power (CHP):* The main purpose of a CHP unit is delivering the required heat for industrial, commercial of residential processes and utilizing the waste heat in order to generate power.
- *PV solar and wind power:* The smaller size PV solar and wind power plants are included in DERs portfolios. They also provide a rapid load curtailment opportunity, although it is not yet accessible since renewable production is prioritised, which results in maximum capacity operation of these sources. Forecasting, another Big Data analytics application, which was discussed in the earlier sections, increases the reliability of these sources in flexibility management.

Distributed energy storage: There are various energy storage technologies in different scales such as pumped hydro storage, compressed-air energy storage, flywheels and power to gas storage [85]. However, residential, C&I and electric vehicle (EV) batteries are examined, taking into account the fact that Big Data analytics offer solutions precisely for these storage types.

- *Residential and C&I owned batteries:* The penetration of batteries among residential and C&I end-users is increasing with decreasing hardware costs. Moreover, companies such as Stem and

Sonnen are pioneering the smart battery concept, which is connected to a solar PV system and capable of optimizing its utilization accordingly.

- *Batteries of EVs:* The batteries of electric cars can be used in 2 different operation modes: Grid-to-vehicle, which operates as a DR program. The charging time of the EV can be optimised according to load curtailment requirement of the grid. The second operation mode is vehicle-to-grid, in which EV batteries are used as DERs [85].

Demand response load sources: As mentioned in the earlier section, load curtailment resulting from DR programs can function as a DER asset if they are dispatch grade. The residential and commercial loads are mainly heating, ventilation and air conditioning (HVAC) systems. Other potential loads are smart home applications such as washing machines, tumble dryers, dishwashers, refrigerators, etc. The industrial load is aggregated from energy intensive industries such as automotive, mining, waste water treatments etc. Their energy demanding components such as pumps, chillers and blowers constitute their portfolio. Table 4.2 summarizes available DER assets, according to their category and segment.

Table 4.2: DERs which can be involved in a DERMS according to their segment and category

Customer Type	**Segment**	HVAC	Pump	Blower	Chiller	Cooling Tower	Comp. Air	PV	CHP	Battery	EV
Industrial	Automotive	✓			✓	✓		✓		✓	
	Mining	✓	✓	✓			✓				
	Waste Water		✓	✓					✓		
	Chemical	✓	✓	✓	✓	✓	✓		✓		
Commercial	Hospital	✓			✓			✓	✓	✓	
	University	✓						✓	✓	✓	✓
	Refrigerated Warehouse		✓		✓	✓		✓		✓	
	Commercial Building	✓	✓		✓	✓		✓	✓	✓	
Residential	Households	✓						✓		✓	✓
Agricultural	Farm		✓					✓			
	Greenhouse			✓					✓		

Microgrids: A microgrid is a local energy grid with a control feature. It is capable of disconnecting from the traditional grid and still operating due to the DER assets included [101]. Ideally, a microgrid can be used as a DER asset.

The DER deployment is increasing rapidly. The newly added capacity of DGs will exceed the newly added centralized generation capacity by 2018 with a projection of over 300 GW by 2026 [102]. The battery costs are decreasing gradually, resulting in further market penetration. Moreover, EV adoption shows an ascending trend, expected to reach between 100 to 140 million cars globally by 2030 according to different scenarios [103]. In conclusion, DER-heavy energy systems are inevitable and all the players of the sector, particularly the utilities need to adopt to this change.

From a Big Data standpoint, increasing number DERs means more data points with great complexity which brings about value creation. There are mainly two trends related to Big Data applications in a DER environment.

- The first group of applications focuses on end-users. The companies working in this direction develop new technologies that optimize DER assets for end-users. They aim at reducing the energy related costs of residential or non-residential end-users by the intelligent use of DER assets. Typical examples are smart batteries and solar-plus-storage concept.

- The second group of applications are utility focused. A comprehensive overview of a DER-heavy environment, in which every single DER asset can be monitored and controlled, creates a spectacular value, particularly for utilities.

4.2.1 Utility Focused Big Data Applications: DERMS and VPP

A distributed energy resource management system (DERMS) is a software-based solution that enhances utility's real-time visibility into the DER assets. It is highly linked to DRMS in terms of functioning. The main difference is that DERMS has a higher scale since it deals with more assets connected. Therefore, many DERMS vendors include the DRMS function embedded in their software solutions.

A typical life cycle of DER management follows eight steps. It starts with registering the assets through a portal. Modelling and interconnecting follows the registration. A simulation of each DER asset's effect on the distribution grid is modelled and physical interconnection is provided in these steps. Later, the system optimizes the resource dispatch while calculating the end-user credits and debts. Lastly, the system calculates the billing for end-users and provides asset performance analytics as mentioned in the previous chapter. Such an automatized and optimized management system creates an opportunity: virtual power plants [104].

A virtual power plant (VPP) is a system which includes the smart aggregation and optimization of DERs in order to create a dispatch capacity that is comparable to a conventional centralised power plant. VPPs, leveraging Big Data analytics and smart grid, can accommodate the complexity of DERs, providing regulation service, voltage management, fast DR, contingency reserve, peak demand management, and renewable integration [102].

Benefits of VPPs

The concept is considerably new in the market. However, various benefits of VPPs have been proven in the investigated case studies. These are mainly:

Portfolio Diversity: The high numbers of accessible DERs from different categories and segments provides a large asset portfolio. Moreover, these products provide gradual adjustments via advanced analytics (e.g. CPower-Stem partnership) which enables real-time tracking of the market condition and instant response.

Instant Ramping Capability: Ramping capability, described as how rapidly a generation source can react to demand or supply of a power system, is essential for balancing the grid. Current fossil fuel solutions such as combined cycle power plants can ramp in approximately 15 minutes while a VPP can respond within seconds [102]. Besides, ancillary services can be provided in 4 seconds, bringing enormous flexibility for the grid.

100% Green Electricity Option: The VPP enables operators to decide the composition of their DER portfolio. A portfolio consisting of renewable resources can be certified as 100% green electricity, resulting in a value creation such as elimination of environmental impact and customer satisfaction. Moreover, such a portfolio is not bound to marginal cost fluctuations related to fuel costs.

Relief in Local Community Concerns: One of the major issues related to renewable energy site development is the resistance from local communities (i.e. not in my backyard problem). The portfolio of VPP is mostly composed of already existing, privately owned sources, and thus they do not encounter public resistance.

Lower Cost Compared to Conventional Power Plants: The SaaS structure of a VPP service results in a much lower cost compared to conventional power plants. A VPP implementation is 80$/MW while a coal fired power plant is 3.000.000$/MW in US [102].

4.2.2 End-user Focused Big Data Applications: Intelligent Batteries

End-user focused Big Data applications include activities which monitor, optimize and control the DER assets in order to reduce energy related costs. The DER assets suitable for these applications are intelligent batteries and solar-plus-storage co-optimization solutions. There is a strong similarity between the energy management systems and these applications. They both aim at providing solutions to end-users while they are used, as a tool to increase the customer engagement for utilities. The main difference is that when EMS optimize the assets with an switch on/off feature, end-user focused applications optimize batteries by charging/discharging. This allows end-users to avoid electricity consumption during the high electricity prices, and therefore brings about savings. Moreover, a platform dedicated to end-users helps them monitor the savings they receive in real-time. These battery systems are already being used by pioneering companies. 10% of the Fortune 500 companies have already implemented storage systems while another 30% is actively considering them [105]. The intelligent battery concept also encourages solar PV adoption since it mitigates the uncertainty of solar systems. In conclusion, Big Data analytics, when used for batteries, can decrease the energy costs of end-users and increase the DER adoption rate.

4.2.3 Findings from Related Companies and Case Studies

The analysis of 30 companies and their case studies show that utilization of DER assets for flexibility management is the most active field of Big Data applications. A large number of trends and business models already exist and new ones are arising either in pilot scale or in research stage. Broadly speaking, the value proposal of DER related business models are similar to DR. They both help utilities manage the flexibility of the grid to avoid grid related problems, reduce the generation costs, avoid unnecessary investments and engage with prosumers. This is expected due to the fact that DR assets are actually a part of DERs. However, as mentioned earlier, DR assets only reduce the load simply by switching off when necessary while other DER assets actually provide power to the grid. This fundamental difference results in different business models to evolve in time. There are mainly three groups of models who provide DER management services:

- The first group includes the companies who provide DERMS and VPP for utilities by a SaaS or a on-premises solution. Almost all of the DRMS providing software companies also have DERMS solutions. It is expected since the required skills and know-how for the services are highly linked. The revenues of these companies are either direct sales of software or subscription fees. The degree of implementation of DER management is flexible. The providers either offer entire life cycle of implementation as a turnkey method or they only implement the software and engaging end-users becomes the responsibility of the utilities.

- The second group of companies dealing with DER management by using Big Data analytics are firm capacity providers for wholesale electricity markets. Instead of offering their digital solutions to utilities as a key partner, they aggregate a firm capacity (i.e. VPP) and participate in power markets. Their asset portfolio is mainly distributed generation due to its relatively high firm capacity/asset feature.

- The third group consists of companies which provide end-user focused Big Data applications through intelligent batteries. They use the relation that they have with the end-users to create a portfolio of assets which can be sold to utilities as firm capacity, operating reserves and ancillary services. Figure 4.2 indicates the business model which these companies apply. The companies, being the mediator between the utility and the end-users receive revenues from both parties. They don't necessarily sell the products to end-users but they may also finance them through leasing and loaning. This decreases the initial investment and makes the assets more affordable. Depending on the agreement, they also receive revenues through the bill reduction their system manages. The revenues related to utilities come from direct sales of firm capacity, operating reserves and ancillary services. The general structure of the business model is depicted in figure 4.2.

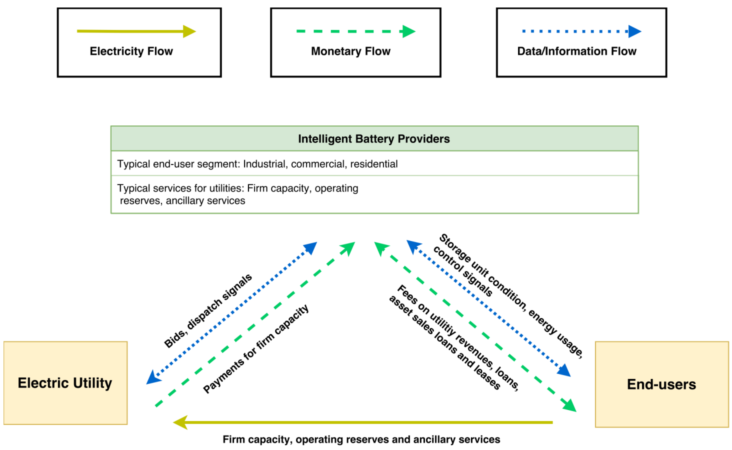

Figure 4.2: Business relation between the intelligent battery providers, electric utilities and end-users

Best Practices: Related Companies and Case Studies

AutoGrid

AutoGrid is a California, US based energy intelligence company which offers solutions for energy management and analysis. Since its founding in 2010, the company received 42 million $ of funding, including the German utility E.ON. They offer both DRMS and DERMS as SaaS which makes its implementation quick and low cost. Being a software provider, the company is a good example of the first group of companies described above. Autogrid has many utility clients concentrated in North America and Europe. Some clients are E.ON, Bonneville Power Administration, Florida Power & Light, Southern California Edison, Eneco, Portland General Electric, City of Palo Alto [106].

Eneco is a Rotterdam, Netherlands based energy provider, reaching over 2 million customers. They worked with Autogrid to create their own VPP, Software Defined Power Plant. The software forecasts, optimizes and controls a capacity of 100 MW DER assets such as CHPs from greenhouses and C&I DR assets [107, 108].

Enbala (in partnership with ABB)

Enbala, a Canadian-based company with a focus of creating reliable, efficient and predictable power grids. ABB is a Swiss-based multinational technology company, operating in electrification products, robotics and motion, industrial automation and power grids. They have operations in over 100 countries with 132.000 employees [109]. Enbala partnered with ABB to develop a DERMS and VPP software called Symphony in early 2017 with a showcase of 100.000 DER asset optimization. The software's driving force is Big Data analytics and machine learning which enables a self-learning mechanism for better predictions. Symphony leverages sensor data received from the smart distribution grid in order to monitor the asset performances in real-time. The recurrent neural network algorithms can reveal patterns which are difficult to detect, and therefore unexpected events in the grids can be determined. Moreover, agent-based control and optimization can assess the assets in sub-seconds. The software has a 99.9% accuracy in DER performance monitoring and can manage up to 1 million assets with a five-second response time [110].

Siemens

Siemens, the German based multinational technology company whose activities are related to almost the entire value chain of the energy sector also provides DERMS embedded in their utility operating systems. Siemens has a partnership with Accenture, called Omnetric, to facilitate implementation of its utility operating systems [92]. The software combines utility informational and operational systems, including customer information, billing and meter data, SCADA, and real-time electricity market information on price and supply to provide an actionable dispatch decisions. The software also allows the users to define aggregated load by substations which improves the reliability. The know-how on sales and implementation obtained from Accenture and developed features of the software brings many customers such as IOUs, municipal utilities, electric cooperatives, and retail electric suppliers [92].

Arizona Public Services (APS) is the largest utility operating in Arizona, US serving over 1.2 million customers [111]. In 2014, they initiated a program called *Solar Partner Program (SPP)*. The program is a trial for a new type of business model in which the utility leases the rooftops of end-users to install solar panels, taking care of entire life cycle of the them. This program made APS the first utility in the US to remotely control solar panels installed in rooftops. The program is actually a pilot to determine the viability of this business model. It reached to 1.670 volunteer participants who agreed to lease their rooftop to the utility with an agreement of 30$ monthly bill reduction for 20 years. The total capacity of the project reached to 10 MW. The DERMS software to manage the project was provided by Siemens.

The final report of the project with the findings will be published in October 2017. However, the interim report which was published in 2016 showed that utility could decrease the feeder load of the grid by 21 percent. Even though the program is in pilot scale, it is significant to point out the potential business models that DERMS can enable[112].

NextKraftwerke

Next Kraftwerke, describing itself as a digital utility, was found in Cologne, Germany in 2009. They are the largest VPP in Europe, operating not only in Germany but also in Austria, Belgium, France, the Netherlands, Poland, Switzerland and Italy. The total capacity reaches 3.200 MW with 4.500 DG units such as CHPs, biogas plants, solar PV sites, small hydro power plants and wind farms. The company also offers ancillary services for several German transmission service operators. (e.g. mprion, EnBW Transportnetze, 50Hertz Transmission and TenneT) [113]. NextKraftwerke is a good representative of the second group of companies who are disruptive for utilities. Eneco has acquired a minority interest of 34% of the company in May, 2017 which shows the interest of utilities for VPP concept.

Sunverge Energy

Sunverge Energy is a California, USA based company which provides solar PV and intelligent battery storage systems to home owners since its establishment in 2010. The company is a representative of the third group mentioned above since they operate a VPP with their accessible assets [114]. Sunverge partnered with several utilities to try different business models in the last years. They aimed at giving the control of their assets to utilities instead of selling them firm capacity.

PowerStream is a municipally owned Canadian utility serving over 360.000 customers in the province of Ontario. The utility initiated a pilot program, similar to APS, mentioned earlier. The program aims at exploring the full ownership of a solar-intelligent battery system in partnership with Sunverge Energy. The program includes 20 residential end-users, equipped with 5 kW panels and 6.8 kW/12 kWh batteries. The business model they test follow several steps: design, build, own, operate, maintain and energize. Basically, implementation and operation of the system is realised by the utility with an up-front 3.500 $ of payment received from the end-users. The payback period of the investment is expected to be 5 years via the savings on top the benefits such as outage management and resilience.

Chapter 5

Elaboration of a Business Case: Iberdrola Efficiency Shop

The objective of this chapter is to elaborate a business case in which Big Data analytics is used. It starts by describing the business idea, and continues with business planning by using the business canvas model. Further, the market opportunities and risks are determined with a SWOT analysis. Lastly, a operational and financial planning is presented.

5.1 The Idea

The analysis of Big Data applications in the electricity sector shows that customer oriented applications are the most suitable for creating a new line of business. Iberdrola, having 30.22 million electricity customers worldwide, has the chance to create revenue streams by offering new services to their customers. *Iberdrola Efficiency Shop* is an idea that positions Iberdrola as a trusted energy advisor for customers' energy consumption behaviour and efficient home appliances choices.

5.1.1 Iberdrola Energy Advisor

The idea provides two main services to the customers. The first service is called *Iberdrola Energy Advisor*. It makes the energy consumption of the customers more transparent to them. It is based on the applications mentioned in the 6^{th} chapter. A disaggregation of energy use data is conducted to identify the electricity consumption at an appliance level. This information in real-time is then utilized to create the platform which is available as a web portal and a mobile app. The platform provides many features that increase the engagement. These are mainly:

- Energy home reports

- Pro-active high bill notifications
- Personalized tips on energy reduction
- Utility offers according to customers consumption habits
- A benchmarking with the similar buildings' energy consumption
- Connected to social media accounts to share the energy related achievements

There are various platforms that use some of these features for utility customer engagement. The difference of the idea is to connect this platform to an online marketplace, *Iberdrola Efficiency Shop*, to motivate its customers to buy energy efficient products. To achieve this connection successfully, the platform has the following characteristics:

- It reaches out to the customer at the right moment.
- It is easy-to-use.
- It incentivizes the customers to use it more.

Figure **??** represents three snaps from the prototype of the idea. They aim at giving an understanding of the customer journey.

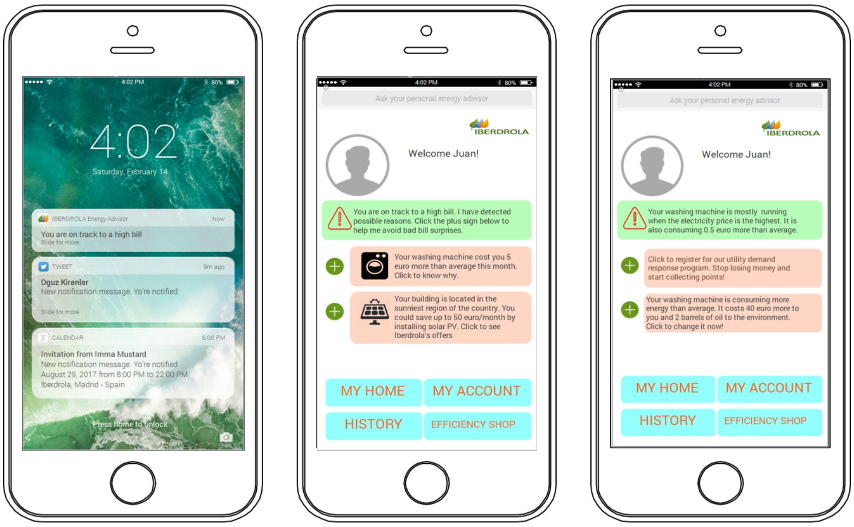

Figure 5.1: The prototype of the idea

The first snap indicates the initiation of the customer interaction. It is important to get involved with the customer in the moments that customer would benefit the most. In this example, the platform detects a customer on track to a high bill and creates a pop-up warning notification. Once the customer responds to the notification, the landing page appears as shown in the second snap. The center of the page is dedicated to potential reasons of the high bill notification. The issues related to excess energy use is explained to the customer in a simple way, using icons and the direct effect of the problem (usually the monetary correspondence of the problem). In this example, the cause of the high bill is determined as the washing machine. As a secondary information, a targeted advertisement is present, stating that the building location is very suitable for solar PV installation and Iberdrola has related offers. It is important to keep the statements simple, clear and personalized to keep the attention of the customer at all times. As the customer clicks the washing machine tab, the solutions appear as shown in the last snap. In this case, the contributor to the high bill is the fact that washing machine is commonly operating in the times that the electricity price is higher and it is not an energy efficient model. The solution to the first cause is a utility program recommendation and a direct link to register. Yet the uniqueness of the idea appears with the second solution: a direct link to the *Iberdrola Efficiency Shop*.

There are four tabs located in the landing page of the platform: My Home, My Account, History and Efficiency Shop. My Home connects the users to a page where they can access energy home reports. A dynamic energy consumption graph is present here. It is adjustable to read it in various modes. For instance, the user can see the energy consumption break down in an appliance level or pick on the appliance and see its performance comparing to others. The history tab allows customers to see their previous energy consumption. They can compare with previous months and receive motivational feedback in case of an energy efficiency which is then available to be shared on social media. My Account is one of the most important tabs of the platform since it holds a loyalty program feature in it. *Iberdrola Energy Advisor* rewards the customers who interact with the platform. This feature resembles a gamification process but it has a tangible benefit, and thus a more effective engagement tool. Basically, the customer collects *Efficiency Points* by participating in utility programs, posting his energy efficiency achievements in social media, by inviting his peers to be a Iberdrola customer, etc. The *Efficiency Point* offers are dynamic and they change according to marketing needs of Iberdrola. The customer can use these points as a discount on *Iberdrola Efficiency Shop*. Further, the points can have more weight in certain products depending on Iberdrola's efficiency goals. For instance, if Iberdrola plans to launch a utility thermostat program, a special offer through the online marketplace can be arranged. In short, connecting a loyalty program to the online marketplace empowers Iberdrola to achieve its strategic goals while satisfying and creating value for its customers. Lastly, the Efficiency Shop tab is a constant link to the online marketplace.

Another tool in the app which aims at simplifying energy consumption insights and recommendations to the customer is the *Ask your personal energy advisor* tab. It is a search engine located in the upside of the screen at all times. It simplifies the app which is heavily filled with information. For instance the customer can type "How is my TV doing?" and a energy consumption comparison of the TV with the best available models in the market and the average models will pop-up. Another example is "How much savings could I have if I had a solar PV in my roof?" As a result, a screen with the results shows up. The capabilities of this tab are open to development. Further, it can be connected to a smart speaker such as Amazon Echo to answer energy related questions. Similar examples were previously discussed in the banking sector Big Data applications section.

5.1.2 Iberdrola Efficiency Shop

Iberdrola Efficiency Shop's objective is to help customers choose the most energy efficient products available in the market. It is based on Enervee online marketplace. As mentioned in the 6^{th} chapter, it is a white label online marketplace provider. Once, the customer is directed to the *Iberdrola Efficiency Shop*, he can compare the the products according to their energy efficiency, price and the customer reviews collected from different retailers. The marketplace also offers two features: YouSave and ClearCost. YouSave calculates monetary savings of the customer in case of purchasing the selected product. It takes into account the average use time, electricity price and life-time of the product. ClearCost makes the same calculation to determine the price of the product with its energy consumption. For instance, a more energy efficient product is expected to be cheaper in the long run. The marketplace also tracks the changes in the price of the selected product and can create alerts when there is a drop in the price. Finally,*Iberdrola Efficiency Shop* connects the customer to online or onsite retailers. This is very important for the feasibility of the idea. If Iberdrola would position as an online retailer, it would be a totally new business area and would require an expertise. Being the online retailer brings about different customer needs such as product returns, insurance, maintenance and it is not suitable for a utility and it is risky for public opinion. After all, the main motivation of the idea is to improve customer engagement while creating new services and revenues. Hence, it is crucial to keep *Iberdrola Efficiency Shop* as a marketplace that connects the customers to retailers rather than selling to them directly.

In conclusion, utility data can be used to create an energy awareness and efficiency platform which drives its customers to a online marketplace where they can buy energy efficient products. To achieve this, contacting the customers in the right moment with a easy-to-use interface is necessary. Finally, a loyalty program is crucial for supporting Iberdrola to achieve its own marketing or energy efficiency regulatory objectives.

5.2 Business Model Canvas of Iberdrola Efficiency Shop

The business model canvas suggested by Alexander Osterwalder was used to describe the business plan of the idea [141]. It consists of eight sections. These describe the value proposition of the idea, the required activities to realise it, cost and income sources, necessary competences and customer segments. Each section is explained below. Besides, a full view of the canvas can be found in appendix B.

5.2.1 Value Proposition

The features of *Iberdrola Efficiency Shop* and its innovative method of reaching out to customers creates value for the Iberdrola, its customers and environment.

- **Customers:** The idea empowers the customers by making their energy related activities more transparent. It helps them lower their monetary losses related to inefficient use of energy without any discomfort. It acts as an advisor for environmentally concerned customers who need guidance to lower their carbon footprint. Moreover, it creates an alternative to the channel of purchasing new home appliances. Lastly, the loyalty program creates tangible benefits such as discounts and special offers as long as they are interacting with the platform and the marketplace.

- **Iberdrola:** The most tangible outcome of the idea for Iberdrola is the new revenue stream created by *Iberdrola Efficiency Shop*. Secondly, being the energy advisor boosts the customer engagement and makes Iberdrola the favourable energy provider in an open market. The loyalty program, *Efficiency Points*, acts as a tool for Iberdrola to drive people to achieve company objectives such as efficiency targets, promoting new utility programs, creating company awareness through social media posts, reducing customer churn and peer-to-peer engagement. Lastly, being the operator of such a platform, Iberdrola pioneers a pro-environment movement which brings about a positive public opinion about the company.

- **Environment:** Since the core of the idea is based on energy efficiency, it creates an awareness towards efficient home appliances and behavioral change.

5.2.2 Key partners

Being a Big Data analytics application, the idea requires skills and experience in information technology, and thus it is not feasible to build in-house. As discussed in previous chapters, there are various software companies providing these skills to utilities. The realization of *Iberdrola Energy Advisor* platform and *Iberdrola Efficiency Shop* marketplace depends on a partnership with technology providers.

- **White label marketplace providers:** As explained in the 6^{th} chapter, there are software companies which provide white label marketplaces. Simply Energy and Enervee draw attention as they were recently preferred by other large utilities in North America and Europe. Between these, Enervee is more favourable since their marketplace promotes energy efficiency by using behaviour science. Moreover, it positions *Iberdrola Efficiency Shop* as a connection between the retailers and the customers, avoiding being the online retailer responsibility.

- **Customer energy insights platform providers:** The term refers to software companies which are capable of creating *Iberdrola Energy Advisor* platform. This platform requires use of Big Data storage and analysis capabilities. The most important parameters to consider while choosing the right partner is the scalability and its ease-of-customization. As stated before, Iberdrola serves over 30.22 million customers. Therefore, the platform must be capable of scaling up to millions of customers when necessary. Lastly, the proposed idea requires some innovative features that are not proposed by any providers yet, such as loyalty programs and the *Ask your personal search tab*. Therefore, a good level of customization is needed for realization of the idea.

5.2.3 Key Activities and Resources

The key activities to realize the idea starts with the implementation of the platform and the marketplace. As mentioned earlier, development of the proposed features which aim at more effective customer engagement is needed. Once the technology is ready, a marketing campaign to create awareness must be run. Lastly, a continuous improvement of the platform and the marketplace according to objectives of Iberdrola must take place. The key resources related to these activities are mainly human, financial and intellectual.

- **Human:** A in-house team dedicated to operate the business must be constituted. The team should have competences in IT technologies, business development and marketing.
- **Financial:** A budget for implementing the technology and marketing is necessary.
- **Intellectual:** This resource is related to initiation of the project. The integration if IT system might require an external consultancy service.

5.2.4 Customer Relations

The customer relations are intended to be kept as automated as possible. Being built on Big Data analytics, the platform and marketplace are capable of managing customer needs and inquiries without human capital. In contrast, *Iberdrola Energy Advisor* could decrease the call center load. An automated customer relations can be maintained by using mobile app, web portal, e-mail, electronic and paper bills. Lasly, a personal assistant chatbot can be integrated to the marketplace to create automated answers.

5.2.5 Channels

The channels to reach out to the customer in order to create awareness about the business can be divided into two groups: Iberdrola customers and others. There is already an existing channel between Iberdrola and its customers. E-mails, electronic and paper bills and company website can be used. Further, motivating customers to post their energy related achievements on social media as a part of loyalty program leads to spreading of the business. Therefore, social media is an important channel to engage new customers. The potential customers who are not yet an Iberdrola customer can be reached through a conventional marketing campaign that uses television and social media advertisements.

5.2.6 Customer Segments

Any Iberdrola customer who purchases a home appliance is the target of this idea. Therefore, the main segment is residential Iberdrola customers since households are the main consumers of home appliances. Secondly, small and medium size commercial customers such as restaurants, bars and retailers are suitable.

5.2.7 Cost Structure

The realization of the business requires several fixed and variable costs. The initial fixed costs are salaries of the team members and marketing. Depending on the key partner's business model, a subscription or a commission fee might be paid proportional to the volume of sales on the marketplace.

5.2.8 Revenue Streams

The direct revenue stream is the commission fees for the products sold in the marketplace. An indirect revenue is created by increased utility program participation rates.

5.3 SWOT Analysis

A SWOT analysis was conducted to evaluate the opportunities and the risks related to the *Iberdrola Efficiency Shop*. Table 5.1 shows the analysis. At the top of the table, the strengths and the weaknesses of the idea are represented. The left hand side indicates the opportunities and the threats in the market. The intersections refer to the resulting effects. For instance, the intersection of strengths and threats describes how the strengths of the idea can mitigate the market related threats. Further, applicable action plans to overcome the threats and weaknesses are are pointed out.

Table 5.1: SWOT analysis for Iberdrola Efficiency Shop

Efficiency Shop	**Strengths** • Integration with the *Iberdrola Energy Advisor* platform to attract customers • Interaction with the customer in the right moment when the they need advice the most • Easy-to-use features such as "Ask you enery advisor" search tab • Dynamic energy efficiency score of the products • Embedded loyalty program in the platform	**Weaknesses** • Involvement of different parties to realize the idea • Possible high initial cost due to licensing and consultancy fees
Opportunities • Increasing awareness towards energy efficiency • Energy efficiency targets • Increasing adoption rate of digital solutions • Paradigm shift in people's purchasing habits towards online retailers • Customer-centric service creation trend among utilities	• The idea complements market trends and potential future government regulations • The strong need of a customer oriented service is a good motivation to realize the idea	• The opportunities and the market needs overcomes the effort resulting from complexity and potential high cost of the idea
Threats • Uncertainty of the adoption of the idea since it is a new approach	• The easy-to-use features and the tangible benefits of the loyalty program involves customers with different interests. Thus, a high rate of adoption is possible	• Special attention must be given to integration consultancy to overcome the complexity smoothly • A strong marketing campaign must be established to reach out to the people effectively

5.4 Operational & Financial Planning

Operational and financial planning is made for 2 years and divided into 3 stages. These stages differ from each other according to the targeted market and actions to take. Therefore, revenues and costs change in every stage.

5.4.1 1st Stage: Spanish Market

The first stage aims at implementing the idea in Spain, and lasts for 3 months. The total market size of home appliances in Spain was 3.26 billion euro in 2014 [142]. Iberdrola has 11 million customers in the country, corresponding to 37.9% of total electricity customers [143, 144]. Hence, it can be assumed that 37,9% of total home appliance market consists of Iberdrola customers. This amount corresponds to 103.2 million euro per month. At this stage,the target is to reach 3% of the total market, which will be the main revenue stream.

A team of 6 white collar employees need to be hired for implementing and operating the business. Further, launching of business must start with a heavy marketing campaign, including both television[1] and internet advertisements. Television advertisements are planned to run two times per day in four different channels with a total duration of one month. In the meantime, internet advertisements are kept active at all times. Service fee of white label platform and marketplace providers are subject to bilateral business agreements and is not publicly open. Thus, the cost structure of these services need to be assumed. For the platform provider a monthly subscription fee of 0.1% of total number of customers is assumed. For marketplace provider 20% of total sales volume is estimated. Based on these expenditures, **the initial investment** is predicted to be 1.425.463 €. Table 5.2 shows details of initial investment and monthly distribution of costs while table 5.3 indicates monthly net balance starting from the second month. The revenue stream is the commission fee of 7.5% received for the items sold on *Iberdrola Efficiency Shop.*

Table 5.2: Monthly expenses of the 1st stage

	Explanation	**Unit**	**Cost/ Unit (€)**	**Cost (€)**	**Source**
Labor Costs	White Collar Employees	6	3.000	18.000	[145]
Media & Marketing	TV ads (off-peak)	120	1.000	120.000	[146]
	TV ads (prime-time)	120	10.000	1.200.000	[146]
	Social Media Ads	100.000	0.3	30.000	[147]
Key Partners	Platform provider	11.000.000	0.001	11.000	Assumption
	Marketplace provider	Sales	20 (%)	46.462 (€)	Assumption

Table 5.3: Monthly net balance of 1st stage

Total Revenue (€)	232.314
Total Cost (€)	105.462
Net Balance (€)	126.851

[1] Both on prime-time and off-peak viewing time

5.4.2 2nd Stage: British Market

The objective of this stage is to expand *Iberdrola Efficiency Shop* to the UK, which is the second country Iberdrola is operating in Europe. This stage is planned as 9 months. The UK's home appliances market size is 7.13 billion euro and Iberdrola has a market share of 12.5% in the country [143, 148, 149]. This leads to a monthly home appliances market size of 74.3 million euro for Iberdrola. The targeted share of this amount is 3%.

The market entry strategy and cost structure is very similar to the 1st stage. The labor and marketing costs vary according to local conditions while costs related to key partners don't change. The marketing campaign is the largest expenditure of the stage, and it is run for the first month. Thus, an **additional investment** of 2.709.943 euro is needed at this stage. A breakdown of the costs are shown in table 5.4. Table 5.5 indicates the net balance of the company after the investment is completed. The revenue stream is identical to 1st stage.

Table 5.4: Monthly expenses of the 2nd stage

	Explanation	**Unit**	**Cost/ Unit (€)**	**Cost (€)**	**Source**
Labor Costs	White Collar Employees	4	4.500	18.000	[150]
Media & Marketing	TV ads (off-peak)	120	2.000	240.000	[151]
	TV ads (prime-time)	120	20.000	2.400.000	[151]
	Social Media Ads	50.000	0.3	15.000	[147]
Key Partners	Platform provider	11.000.000	0.001	3.500	Assumption
	Marketplace provider	Sales	20 (%)	33.443 (€)	Assumption

Table 5.5: Monthly net balance of 2nd stage

Total Revenue (€)	167.219
Total Cost (€)	69.943
Net Balance (€)	97.275

5.4.3 3rd Stage: US Market

An expansion to the US market is planned during the 3rd stage, which takes 12 months. The US home appliances market is relatively large, reaching 36.6 billion euro per year [152]. Iberdrola has 2.5 million customers, which leads to 1.6% of total electricity consumers in the country [143]. This results in a 48.556.224 euro of total market size. Even though the market entry strategy is similar, a milder marketing plan is considered at this stage. This is due to extremely high cost of marketing in the US. From a financial standpoint, the expected market share by a similar marketing campaign is not viable. Therefore, one month long TV advertising step is skipped. However, social media advertisements at all times continue. As a result, 1.5% of the total market is expected to achieved. Table 5.6 and table 5.7 represent the breakdown of the costs and the net balance respectively.

Table 5.6: Monthly expenses of the 3rd stage

	Explanation	Unit	Cost/ Unit (€)	Cost (€)	Source
Labor Costs	White Collar Employees	2	7.100	14.200	[153]
Media & Marketing	Social Media Ads	80.000	0.3	24.000	[147]
Key Partners	Platform provider	2.500.000	0.001	2.500	Assumption
	Marketplace provider	Sales	20 (%)	10.925 (€)	Assumption

Table 5.7: Monthly net balance of 3rd stage

Total Revenue (€)	54.625
Total Cost (€)	51.625
Net Balance (€)	3.000

5.4.4 Cumulated Cash Flow Diagram

Cumulated cash flow diagram shown in figure 5.2 is a summary of financial projections. The anticipated break-even point is around 20 month, which proves that the idea is viable. It must be noted that the most striking value proposal of the idea is qualitative benefits for both customers and Iberdrola as mentioned above. Thus, a larger marketing campaign which would increase the customer engagement could be planned even though this would shift the break-even point to later months.

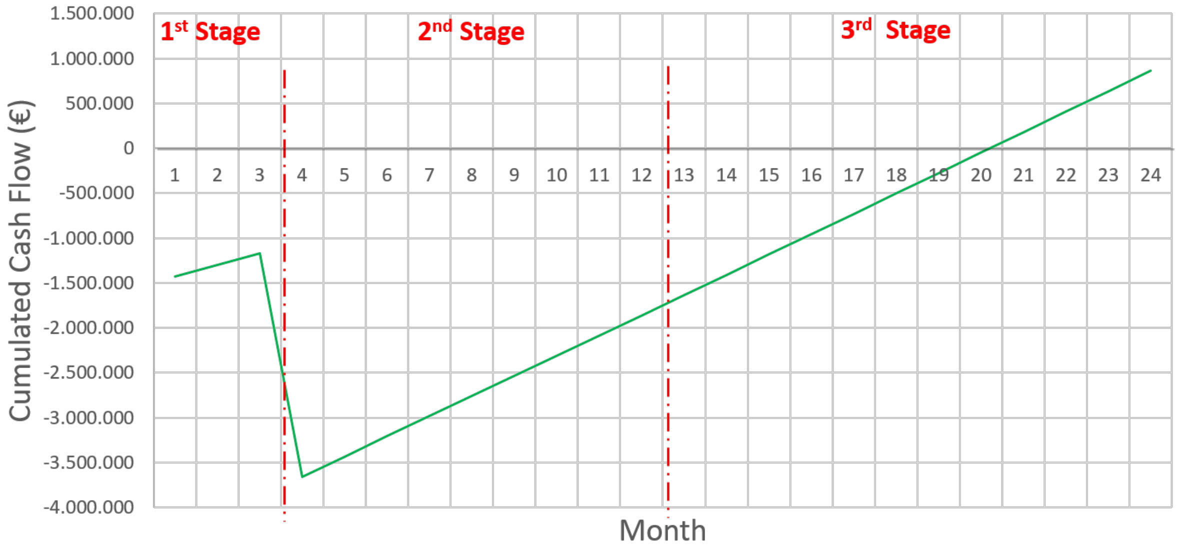

Figure 5.2: Cumulated cash flow diagram

Chapter 6

Conclusions and Recommendations

This chapter concludes the study by summarizing the previous sections. Later, recommendations are made for Iberdrola.

This master thesis provided a comprehensive overview of Big Data analytics applications in the energy sector. Prioritising the interests of a utility company, entire value chain of the electricity sector was divided into four parts and the applications were determined for each of them. Activities of academia, utilities, and other sectors together with value propositions of companies[1] were used for determination of applications. When applicable, new business models enabled by Big Data were identified. Lastly, a business case suitable for Iberdrola was elaborated to emphasize the business opportunities created by Big Data analytics.

The field is highly dynamic. Especially, in North America and Europe many utilities adopt Big Data analytics applications either to improve their operations or to create a new line of business. Their corporate venture capital divisions invest in related companies. Partnerships or acquisitions of service provider companies are also common. Most of the applications were created in the recent years and the dynamism appears to continue.

From the generation perspective, the applications enhance the adoption and performance of renewables. The applications related to wind energy generation are focused on reducing the costs by providing better forecast and operational excellence. Being an asset-centric generation, actionable information about the health status of the turbines boosts the efficiency of field operations. Big Data analytics is either implemented by turbine manufacturers as a turn-key service directly, or utilities develop it in-house or with an external partner. In the meantime, solar energy, being a DER asset, is highly linked to customer collaboration and adoption. Therefore, many applications are derived to promote its adoption by decreasing financing and installation costs. This creates new business model opportunities for utilities.

There is a strong link between smart grid and transmission & distribution network applications. In this section, all of the applications aim at reducing the costs related to operation of these assets. Predictive

[1] The companies that leverage data to create products and services.

analysis enables pro-active insights about the asset status. These insights are utilized to empower field workers. At this point, natural language generation technique simplifies the complex data sets and creates easy-to-understand work orders. Acting as a senior engineer, the technology amplifies the effectiveness in the field. Better monitoring and control over the grid also allows a more optimized asset planning. Further, outages related to severe weather conditions can be foreseen to plan a proactive management. Lastly, a more precise electricity theft investigation to protect revenues is possible for the regions of the world where theft is a major issue.

Flexibility management is one of the most important challenges of modern distribution systems. Big Data analytics is transforming this challenge into business opportunity by various applications. Having the highest number of related business models detected, this section involves many actors such as energy intelligence companies, DR providers, VPP operators, energy management companies, smart home appliances manufacturers, etc. The field is very active in North America and some countries in Europe (France, the UK, Germany, the Netherlands, Belgium).

Customer relations applications offer one of the highest values for utilities. This is mainly due to change in the customer expectations and customer-centric response to this change by utilities. Having an individual and personalized contact with the customers by using their data provides benefits to both sides. Customers can be empowered with more knowledge about their energy use, and start saving without an effort. Moreover, they can receive recommendations not only about their energy use but also their home assets. The marketplace link of this personalized knowledge empowers them to optimize their home related purchases. Such service scales up to projects as well. They can decide on an upgrade for their buildings (windows, roof, solar adoption, etc.), according to its energy related benefits. From a utility standpoint, reaching to the customers at the right time with the relevant content positions them as an energy advisor and bonds the customer engagement. Findings from other similar sectors show that creating bundled services with data increases the business volume and ensures the prevention of customer churn.

6.1 Recommendations

Based on the conclusions above, the list below provides recommendations for Iberdrola.

- As elaborated in the 7th chapter, Iberdrola could use its customer data to become an energy advisor and a marketplace operator. The personalized level of communication is a strong motivation for customer engagement and creating business out of it. In parallel, a project marketplace can be launched to provide building upgrades to customers.

- The open banking concept described in the 6th chapter could be adopted from banking sector. Positioning Iberdrola as an *App-Store* operator would involve independent developers to create services for customers where every interaction in the store returns to Iberdrola as revenue and brand awareness.

- Iberdrola Smart Solar initiative can be supported by solar energy generation applications. Firstly, tailored marketing services that use demographic and LIDAR data can decrease customer acquisition costs. Further, installation design can be managed off-site, without the need of field inspection. The programs can also involve a contractor marketplace for installation and maintenance to empower the customers. Lastly, partnering with a third party financial risk analyser company (see kWh Analytics) can decrease the financing costs of customers.

- Flexibility management is essential in the new era of energy. As mentioned in the 5^{th} chapter, utilities in the US and Europe position themselves as energy management partners of C&I and residential customers. In return they involve their customers in demand side management programs. Iberdrola already provides various products and services at energy management. Implementation of a DERMS would position Iberdrola as a VPP provider and help the company balance the grid while creating a new business line. However, this opportunity is highly linked to legislative decisions.
- Natural language generation is a very promising application. It could cut the costs of grid operations by instantly converting data into simple texts and work orders.
- Brazil is a problematic field of operation of Iberdrola in terms of electricity theft. A collaboration with a theft analytics company (see Choice) could mitigate the problem.
- Centralization in asset performance management of wind farms is an essential Big Data application. Yet, it is not elaborated since the application is already adopted by Iberdrola.

This study shows that Big Data analytics has already proved itself to be a trustful tool to manage the paradigm shift in the energy sector. It is used by various actors ranging from multinational companies to garage start-ups. However, it can still be considered at its infantry stage. There is still a lot of space for new applications and business models. Therefore, innovative utilities should examine the phenomena carefully and concentrate their R&D&I activities on the topic.

Bibliography

[1] *Introduction*. MIT Press, 2015, pp. xv–xxiv. [Online]. Available: http://www.jstor.org/stable/j.ctt17kmw13.5

[2] T. H. Davenport, “Analytics 3.0,” *Harvard Business Review*, vol. 91, no. 12, p. 64, 2013. [Online]. Available: https://hbr.org/2013/12/analytics-30

[3] P. Zikopoulos, D. deRoos, C. Bienko, R. Buglio, and M. Andrews, *Big Data Beyond the Hype*. McGraw Hill, 2015.

[4] D. Reinsel, J. Gantz, and J. Rydning, “Data Age 2025 - The Evolution of Data to Life-Critical: Don ’ t Focus on Big Data ; Focus on the Data That ’ s Big,” pp. 1–25, 2017.

[5] I. Lee, “Big data: Dimensions, evolution, impacts and challenges,” *Business Horizons*, vol. 60, pp. 293–303, 2017.

[6] D. Vesset, A. Nadkarni, B. McDonough, S. Bond, Q. Li, C. W. Olofson, A. Zaidi, D. Schubmehl, S. Kusachi, and P. Carnelley, “Idc futurescape: Worldwide big data and analytics 2016 predictions,” International Data Corporation, Tech. Rep., 2016.

[7] PwC, “14th PwC Global Power & Utilities Survey,” Tech. Rep. [Online]. Available: http://www.pwc.com/cl/es/publicaciones/assets/14th-pwc-global-p-u-survey-web-12-05-15.pdf

[8] Accenture, “Accenture Technology Vision 2015,” pp. 1–24, 2015.

[9] Iberdrola, “Iberdrola, the world’s fourth largest utility.” [Online]. Available: https://www.iberdrola.com/top-stories/iberdrola-shares-with-you/iberdrola-supports-innovation

[10] K. Zhou, C. Fu, and S. Yang, “Big data driven smart energy management: From big data to big insights,” *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 215–225, 2016.

[11] S. S. Refaat, H. Abu-Rub, and A. Mohamed, “Big data, better energy management and control decisions for distribution systems in smart grid,” *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, pp. 3115–3120, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7840966/{%}5Cnhttps://www.scopus.com/inward/record.uri?eid=2-s2.0-85015160024{&}doi=10.1109{%}2FBigData.2016.7840966{&}partnerID=40{&}md5=f7871e5f36ad22990814a2d9fba5bd4d

[12] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” *Mobile Networks and Applications*, no. 19, pp. 171–209, 2014.

[13] N. Koseleva and G. Ropaite, “Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges,” 2017.

[14] P. D. Diamantoulakis, V. M. Kapinas, and G. K. Karagiannidis, “Big Data Analytics for Dynamic Energy Management in Smart Grids,” *Big Data Research*, 2015.

[15] D. Tayal, “Leveraging innovation for electricity utilities,” *Electricity Journal*, 2017.

[16] H. Ekbia, M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, V. R. Suri, A. Tsou, S. Weingart, and C. R. Sugimoto, “Big data, bigger dilemmas: A critical review,” *Journal of the Association for Information Science and Technology*, vol. 66, no. 8, pp. 1523–1545, 2015.

[17] Vagata Pamela and Wilfong Kevin, “Scaling the Facebook data warehouse to 300 PB — Engineering Blog — Facebook Code.” [Online]. Available: https://code.facebook.com/posts/229861827208629/scaling-the-facebook-data-warehouse-to-300-pb/

[18] National Science Foundation, “Core Techniques and Technologies for Advancing Big Data Science & Engineering,” pp. 1–17, 2011.

[19] J. Podesta, P. Pritzker, E. J. Moniz, J. Holdren, and J. Zients, “Big Data: Seizing Opportunities,” Tech. Rep. May, 2014. [Online]. Available: https://www.whitehouse.gov/sites/default/files/docs/big{_}data{_}privacy{_}report{_}may{_}1{_}2014.pdf

[20] Networked European Software and Services Initiative (NESSI), “Big Data: A New World of Opportunities,” no. December, p. 25, 2012. [Online]. Available: http://www.nessi-europe.com/Files/Private/NESSI{_}WhitePaper{_}BigData.pdf

[21] T. Becker, E. Curry, A. Jentzsch, and W. Palmetshofer, "The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches," pp. 277–291, 2016. [Online]. Available: http://link.springer.com

[22] D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety," *Application Delivery Strategies*, p. 4, 2001.

[23] H. Wang, Z. Xu, H. Fujita, and S. Liu, "Towards felicitous decision making: An overview on challenges and trends of big data," *Information Sciences*, vol. 367-368, pp. 747–765, 2016.

[24] P. Zikopoulos, D. DeRoos, C. Bienko, R. Buglio, and M. Andrews, *Big Data Beyond the Hype*, New York, 2015.

[25] B. Furht and F. Villanustre, *Big Data Technologies and Applications*. Springer International Publishing, 2016. [Online]. Available: https://books.google.com.tr/books?id=5n8XDQAAQBAJ

[26] A. L'Heureux, K. Grolinger, H. F. ElYamany, and M. Capretz, "Machine Learning with Big Data: Challenges and Approaches," *IEEE Access*, pp. 1–22, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7906512/

[27] V. R. Borkar, M. J. Carey, and C. Li, "Big Data Platforms: What's Next?" *XRDS: Crossroads, The ACM Magazine for Students*, vol. 19, no. 1, p. 44, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2331042.2331057

[28] V. Borkar, M. J. Carey, and C. Li, "Inside " Big Data Management ": Ogres , Onions , or Parfaits ?" *EDBT '12 Proceedings of the 15th International Conference on Extending Database Technology*, pp. 3–14, 2012.

[29] H. Hu, Y. Wen, T. S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, 2014.

[30] C. E. A. Hagan, "Big Data and the Creative Destruction of Today ' s Business Models," pp. 1–18, 2013. [Online]. Available: https://www.atkearney.com/documents/10192/698536/Big+Data+and+the+Creative+Destruction+of+Todays+Business+Models.pdf/f05aed38-6c26-431d-8500-d75a2c384919

[31] J. Manyika, M. Chui, P. Bisson, J. Woetzel, R. Dobbs, J. Bughin, and D. Aharon, "The Internet of Things: Mapping the value beyond the hype," p. 144, 2015.

[32] M. Helmy, A. Wahab, M. Norzali, H. Mohd, H. F. Hanafi, M. Farhan, and A. Background, "Data Pre-processing on Web Server Logs for Generalized Association Rules Mining Algorithm," *World Academy of Science, Engineering and Technology*, vol. 48, no. December, pp. 190–197, 2008.

[33] J. Akoka, I. Comyn-Wattiau, and N. Laoufi, "Research on Big Data – A systematic mapping study," *Computer Standards & Interfaces*, vol. 54, pp. 105–115, 2017.

[34] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, "Pregel: a system for large-scale graph processing," *Proceedings of the 2010 international conference on Management of data - SIGMOD '10*, pp. 135–146, 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1807167.1807184

[35] Gavin Blackett, "The OR Society: Special Interest Groups - Analytics Network." [Online]. Available: https://www.theorsociety.com/Pages/SpecialInterest/AnalyticsNetwork_analytics.aspx

[36] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, 2017.

[37] C. Amitkumar and B. Mohan, "Multimedia data mining : state of the art and challenges," no. November 2010, pp. 35–76, 2011.

[38] C. L. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data - Data-intensive applications, challenges, techniques and technologies A survey on Big Data.pdf," *Information Sciences*, vol. 275, pp. 314–347, 2014. [Online]. Available: http://www.crisismanagement.com.cn/templates/blue/down_list/llzt_dsj/Data-intensiveapplications,challenges,techniquesandtechnologiesAsurveyonBigData.pdf

[39] L. Igual and S. Seguí, *Introduction to Data Science*. Switzerland: Springer International Publishing, 2017.

[40] Y. Singh, P. K. Bhatia, and O. Sangwan, "A review of studies on machine learning Techniques," *International Journal of Computer Science and Security*, vol. 1, no. 1, pp. 395–399, 2007. [Online]. Available: http://books.google.com/books?hl=en{%}5C{&}lr={%}5C{&}id=7xRB3hRaR2sC{%}5C{&}oi=fnd{%}5C{&}pg=PA70{%}5C{&}dq=A+review+of+studies+on+machine+learning+techniques{%}5C{&}ots=J{%}5C{_}r2yiD-fJ{%}5C{&}sig=4mw7{%}5C{_}{%}5C{_}NxU5CdNTZBs3dkQWuQuvY

[41] "1.10. Decision Trees — scikit-learn 0.18.2 documentation." [Online]. Available: http://scikit-learn.org/stable/modules/tree.html

[42] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, pp. 1310–1315, 2016. [Online]. Available: https://www.engineeringvillage.com/share/document.url?mid=cpx{_}4dbc01df158d4d5e77dM6b8b10178163171{&}database=cpx

[43] N. S. Nafi, K. Ahmed, M. A. Gregory, and M. Datta, "A survey of smart grid architectures, applications, benefits and standardization," *Journal of Network and Computer Applications*, vol. 76, no. June, pp. 23–36, 2016.

[44] Amazon, "What is Dash Replenishment." [Online]. Available: https://developer.amazon.com/de/dash-replenishment-service

[45] Microsoft, "What is cloud computing? A beginner's guide — Microsoft Azure." [Online]. Available: https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/

[46] T. Telsnig, C. Hernandez Vazquez, and A. Villalba Pradas, "JRC Wind Energy Status Report 2016 Edition wind energy," Tech. Rep., 2017.

[47] General Electric Company, "The Case for an Industrial Big Data Platform," 2013. [Online]. Available: http://www.ge.com/digital/sites/default/files/Industrial{_}Big{_}Data{_}Platform.pdf

[48] V. Wind, "Vestas," *2011*.

[49] Pittman DAvid, "Lords of the Data Storm: Vestas and IBM Win Big Data Award — IBM Big Data & Analytics Hub," 2012. [Online]. Available: http://www.ibmbigdatahub.com/blog/lords-data-storm-vestas-and-ibm-win-big-data-award

[50] G E Renewable and G. E. Digital, "Digital Wind Operations Optimization from GE Renewable Energy."

[51] G E Renewable, "Digital Wind Asset Performance Management from GE Renewable Energy," 2017.

[52] T. Kellner, "Wind in the Cloud? How the Digital Wind Farm Will Make Wind Power 20 Percent More Efficient - GE Reports," 2015. [Online]. Available: http://www.ge.com/reports/post/119300678660/wind-in-the-cloud-how-the-digital-wind-farm-will/

[53] U. Kreutzer, "From Big Data to Smart Data: Remote Maintenance - Digitalization & Software - Pictures of the Future - Innovation - Home - Siemens Global Website," 2014. [Online]. Available: https://www.siemens.com/innovation/en/home/pictures-of-the-future/digitalization-and-software/from-big-data-to-smart-data-remote-service-wind-turbines.html

[54] N. Haouas and P. R. Bertrand, "Wind Farm Power Forecasting," vol. 2013, 2013.

[55] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable & Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.

[56] "IBM Research - China: Smarter Energy Solution Suite." [Online]. Available: http://research.ibm.com/labs/china/smarter_energy.html

[57] P. Pinson, G. Kariniotakis, and D. Mayer, "Wind Power Forecasts," 2010.

[58] A. Jäger-waldau, "PV Status Report 2016 October 2016," Tech. Rep. October, 2016.

[59] K. Ardani, D. Seif, C. Davidson, J. Morris, S. Truitt, R. Torbert, and R. Margolis, "Preliminary non-hardware ('soft') cost-reduction Roadmap for residential and small commercial solar photovoltaics, 2013-2020," pp. 3463–3468, 2013.

[60] B. E. Sprague, "Three Ways Data Will Accelerate Solar Electricity Development in the Next Five Years," pp. 1–10, 2015.

[61] CertainSolar, "Solution." [Online]. Available: http://www.certain.solar/

[62] "PowerScout Secures $5.2 Million In Funding For Machine-Learning Clean Energy — CleanTechnica," 2016. [Online]. Available: https://cleantechnica.com/2016/09/27/powerscout-secures-5-2-million-funding-machine-learning-clean-energy/

[63] "Geostellar's Big DataSolar Power Geomatics," 2017. [Online]. Available: https://de.scribd.com/document/53298510/Geostellar-Big-Data-Solar-Geomatics-1-0-041611

[64] National Renewable Energy Laboratory, "Geostellar : Remote Solar Energy Assessments Personalized," Tech. Rep., 2013. [Online]. Available: https://www.nrel.gov/docs/fy16osti/63615.pdf

[65] "Alectris Company History — Alectris." [Online]. Available: http://alectris.com/alectris-company-history/

[66] Alectris Solar Asset Innovation, "Puglia , Italy Deeply Distressed PV Plants Revived by Alectris Resulting in Dramatic Increase of the Solar Portfolio's Profitability," 2010.

[67] "NEXTracker Announces TrueCapture^{TM}, a Revolutionary Intelligent Control System for Solar Power Plants — NEXTracker," 2017. [Online]. Available: https://www.nextracker.com/2017/07/nextracker-announces-truecapture-a-revolutionary-intelligent-control-system-for-solar-power-plants/

[68] International Electrotechnical Commission, "Strategic asset management of power networks," 2015.

[69] C3 IoT, "Machine Learning Software Solutions." [Online]. Available: https://c3iot.com/products/c3-iot-applications/

[70] Enel, "Over 61 Million Customers Choose Enel." [Online]. Available: https://www.enel.com/en/media/news/d201405-over-61-million-customers-choose-enel.html

[71] C3 IoT, "Enel and C3 Energy's Machine Learning-Based Predictive Maintenance — C3 IoT." [Online]. Available: https://c3iot.com/enel-and-c3-energys-machine-learning-based-predictive-maintenance/

[72] Avoka, “Data collection in the field,” 2016.

[73] Arria, “NLG Services.” [Online]. Available: https://www.arria.com/nlg-services/

[74] “Case Study: Automate Expert Reports Anytime Anywhere,” pp. 1–4.

[75] The Weather Company, “Utility Outage Prediction 2.0.”

[76] T. W. Company, “Timeline.” [Online]. Available: http://www.theweathercompany.com/timeline

[77] The Weather Company, “Proactive preparation and response to weather-related power outages.” [Online]. Available: https://business.weather.com/products/outage-prediction

[78] T. W. Company, “Outage Prediction.” [Online]. Available: https://business.weather.com/products/outage-prediction

[79] W. B. Data, “Electric power transmission and distribution losses (percent of output).” [Online]. Available: https://data.worldbank.org/indicator/EG.ELC.LOSS.ZS?locations=ES

[80] D. Da Silva Costa Lima, “Solar Power Versus Electricity Theft in Brazilian Favelas,” 2015. [Online]. Available: https://energypedia.info/images/d/d9/Diego_Costa_-_Solar_power_versus_electricity_theft_in_Brazilian_favelas.pdf

[81] Ç. Yurtseven, “The causes of electricity theft: An econometric analysis of the case of Turkey,” *Utilities Policy*, vol. 37, pp. 70–78, 2015.

[82] Awesense, “About Awesense.” [Online]. Available: https://www.awesense.com/about/{#}our-partners

[83] Bloomberg, “Choice Technologies Holding SARL: Private Company Information.” [Online]. Available: https://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=340807271

[84] Choice, “Success Story: Light S.A.” [Online]. Available: http://choiceholding.com/light-s-a/

[85] Sweco, Ecofys, Tractbel Engineering, and PWC, “Study on the effective integration of demand energy recourses for providing flexibility to the electricity system,” Tech. Rep. April, 2015. [Online]. Available: https://ec.europa.eu/energy/sites/ener/files/documents/5469759000EffectiveintegrationofDERFinalver2{_}6April2015.pdf

[86] “Demand Response — Department of Energy.” [Online]. Available: https://energy.gov/oe/activities/technology-development/grid-modernization-and-smart-grid/demand-response

[87] P. Siano, “Demand response and smart grids—A survey,” *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[88] US Department of Energy, “Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them,” Tech. Rep. February, 2006. [Online]. Available: https://emp.lbl.gov/sites/all/files/report-lbnl-1252d.pdf

[89] The Energyst, “Demand Side Response: Turning Interia into Inertia,” Tech. Rep., 2016. [Online]. Available: http://theenergyst.com/wp-content/uploads/2016/09/DSR-2016.pdf

[90] P. Bertoldi, P. Zancanella, and B. Boza-Kiss, “Demand Response status in EU Member States,” Tech. Rep., 2016. [Online]. Available: https://iet.jrc.ec.europa.eu/energyefficiency/sites/energyefficiency/files/publications/demand{_}response{_}status{_}in{_}eu28{_}member{_}states-online.pdf

[91] S. Nolan and M. O’Malley, “Challenges and barriers to demand response deployment and evaluation,” *Applied Energy*, vol. 152, pp. 1–10, 2015.

[92] B. Feldman and L. Callaway, “Navigant Research Leaderboard Report : Demand Response Management Systems,” Navigant Research, Tech. Rep., 2016.

[93] AutoGrid, “White Paper: The Economic Value of Next-Generatino DR,” Tech. Rep., 2017.

[94] Ethan A. Rogers, R. Neal Elliott, Sameer Kwatra, Dan Trombley, and Vasanth Nadadur, “Intelligent Efficiency: Opportunities, Barriers, and Solutions,” 2013. [Online]. Available: http://www.aceee.org/sites/default/files/publications/researchreports/e13j.pdf

[95] Comverge, “PHI to Reach More Than 370 , 000 Customers for Demand Response.” [Online]. Available: http://www.comverge.com/Comverge/media/pdf/CaseStudies/Eskom-Case-Study.pdf

[96] “EnerNOC Wins MITX Big Data and Analytics Innovation Award (NASDAQ:ENOC).” [Online]. Available: http://investor.enernoc.com/releasedetail.cfm?releaseid=917948

[97] EnerNOC, “EnerNOC Provides Key Demand Response Resources to Southern California Edison,” 2009. [Online]. Available: https://www.enernoc.com/sites/default/files/media/pdf/case-studies/P14141_cs_sce.pdf

[98] E. Company, “Case Study PNM Partners with EnerNOC to Cut Peak Demand and Avoid Expensive Capacity Additions,” 2007. [Online]. Available: https://www.enernoc.com/sites/default/files/media/pdf/case-studies/P14139_cs_pnm.pdf

[99] EnerNOC, “Midwest Energy Grows New Energy Supply with EnerNOC Agricultural Demand Response.” [Online]. Available: https://www.enernoc.com/sites/default/files/media/pdf/case-studies/P14138_cs_midwest-energy.pdf

[100] Comverge, “IntelliSOURCE Enterprise Brochure.” [Online]. Available: http://www.comverge.com/comverge/media/pdf/Brochures/Comverge-IntelliSOURCE-Enterprise-Brochure.pdf

[101] “How Microgrids Work — Department of Energy.” [Online]. Available: https://energy.gov/articles/how-microgrids-work

[102] P. Asmus, “Stacking Value with Virtual Power Plants in Today’s Digital Power Grid Moving Distributed Networked Energy into the Mainstream,” 2017.

[103] International Energy Agency, “Global EV Outlook 2016 Electric Vehicles Initiative,” *Iea*, p. 51, 2016.

[104] B. Kellison, “DERMS : A TERM GONE TOO FAR ? DER Lifecycle Management Is Not Limited to,” 2017.

[105] A. Sandres, “How Big Data Changed the Way Buildings Use Batteries — Realcomm Advisory.” [Online]. Available: https://www.realcomm.com/advisory/745/2/how-big-data-changed-the-way-buildings-use-batteries

[106] “AutoGrid — AutoGrid Wins Greentech Media 2016 Grid Edge Award.” [Online]. Available: http://www.auto-grid.com/news/autogrid-wins-greentech-media-2016-grid-edge-award

[107] “AutoGrid — Eneco Partners with AutoGrid to Deploy Industry’s First Software Defined Power Plant to Integrate Renewable Energy into the Dutch Electric Grid.” [Online]. Available: http://www.auto-grid.com/news/2895-2

[108] “History — Eneco Group.” [Online]. Available: https://www.enecogroup.com/who-we-are/about-eneco-group/history/{#}/2010

[109] “ABB invests in Enbala Power Networks to co-develop cutting-edge grid software.” [Online]. Available: http://cdn.news.e.abb.com/api/public/attachment/2017/2/23/901/ABBinvestsinEnbalaPowerNetworkstoco-developcutting-edgegridsoftware.pdf

[110] “Enbala Announces Symphony by Enbala Energy Balancing Platform Product Enhancements.” [Online]. Available: http://www.prweb.com/releases/2017/02/prweb14025544.htm

[111] “AutoGrid — AutoGrid Wins Greentech Media 2016 Grid Edge Award.” [Online]. Available: http://www.auto-grid.com/news/autogrid-wins-greentech-media-2016-grid-edge-award

[112] J. Paper, “Smart Inverters in Action: Initial Findings From APS’ Utility-Owned Solar Program — Greentech Media.” [Online]. Available: https://www.greentechmedia.com/articles/read/Smart-Inverters-in-Action-Initial-Findings-From-APS-Utility-Owned

[113] “Next Kraftwerke in Facts & Figures.” [Online]. Available: https://www.next-kraftwerke.com/company/facts-figures

[114] P. Asmus and M. Lawrence, “Making Sense of New Public Power DER Business Models The Business Case for Energy Storage,” 2016.

[115] EnergySavvy, “The Journey to Personalization: How Amazon, Netflix, Lyft, and Uber Might Deliver the Utility Customer Experience.”

[116] Tendril, “The Tendril Physics-based Building Model White Paper,” 2014.

[117] Iberdrola, “My home.” [Online]. Available: https://www.iberdrola.es/home

[118] Bidgely, “Customer Engagement: Gain an Edge through Energy Disaggregation,” no. April, 2017.

[119] OPower, “Moments that Matter : A deep dive into the customer experience during utility billing Customers are ready for a new billing experience,” pp. 1–12, 2015.

[120] R. Bigliani, “E . ON UK Rebuilds Trust with Customer Engagement and Digital Transformation,” 2015.

[121] Ecova, “The Digital Marketplace.”

[122] R. Bigliani, “Make Your Customers Feel Good And Become a Truly ”Trusted Energy Advisor” It’s Time for a Change A Way to Get There,” pp. 1–8, 2017.

[123] Enervee, “Samsung 54.64 in LED Television - Energy Score 100/100 - UN55J6200AFXZA.” [Online]. Available: https://marketplace.pge.com/televisions/134398587-Samsung-UN55J6200AFXZA-most-energy-efficient

[124] Simple Energy, “Engagement Platform — Simple Energy.” [Online]. Available: http://simpleenergy.com/engagement-platform/

[125] Company Simple Energy, “Simple Energy’s Marketplace Exceeds Xcel Energy Goals.” [Online]. Available: http://simpleenergy.com/simple-energys-marketplace-exceeds-xcel-energy-goals/

[126] C. S. Energy, “Huge Success for ComEd Marketplace.” [Online]. Available: http://simpleenergy.com/huge-success-for-comed-marketplace/

[127] Deloitte, “Opportunities in Telecom Sector : Arising from Big Data Descriptive,” p. 32, 2015.

[128] Y.-P. Shee, D. Crompton, H. Richter, and S.-P. Maehle, “Big data in banking for marketers - How to derive value from big data,” p. 70, 2014. [Online]. Available: https://www.evry.com/globalassets/insight/bank2020/bank-2020---big-data---whitepaper.pdf

[129] Teradata, “Enhanced User Experience And Cost Savings Are Making Telenor India Even More Competitive,” pp. 1–10.

[130] HDFC Bank, “Case Study HDFC Bank: Transforming Customer Experience at ATMs,” 2010.

[131] Citigroup, “Citi Is Embracing the ‘Fintech’ Revolution.” [Online]. Available: http://fortune.com/citigroup-fintech/

[132] Evry, “Open Banking Transformation,” pp. 1–12, 2016.

[133] K. Cooke, “BBVA launches an open API marketplace for developers.” [Online]. Available: https://www.maparesearch.com/bbva-launches-an-open-api-marketplace-for-developers/

[134] MoneyHub, “Demo Video.” [Online]. Available: https://www.moneyhub.com/demo-video/

[135] Credit Kudos, “Alternative Credit Scoring.” [Online]. Available: https://creditkudos.com/

[136] SafetyNet, “How it works?” [Online]. Available: https://www.safetynetcredit.com/how-it-works

[137] Facebook, “Payments in Messenger — Facebook Help Center.” [Online]. Available: https://www.facebook.com/help/863171203733904/?helpref=hc{_}fnav

[138] Capital One, “Capital One is on Amazon Echo.” [Online]. Available: https://www.capitalone.com/applications/alexa/

[139] Tearsheet, “7 examples showing the power of banking APIs.” [Online]. Available: http://www.tearsheet.co/financial-robots/7-examples-showing-the-power-of-banking-apis

[140] B. Marr, *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. Wiley, 2016. [Online]. Available: https://books.google.ch/books?id=wAOmCgAAQBAJ

[141] A. Osterwalder and Y. Pigneur, *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*, ser. Wiley Desktop Editions. Wiley, 2010. [Online]. Available: https://books.google.de/books?id=fklTInjiPQAC

[142] Marketline, “Marketline Report Store - Household Appliances in Spain.” [Online]. Available: https://store.marketline.com/report/ohec3267--household-appliances-in-spain/

[143] Iberdrola, “Integrated Report,” 2017.

[144] El Confidencial, “Noticias de Iberdrola: Endesa e Iberdrola pierden clientes por tercer año consecutivo en España . Noticias de Empresas.” [Online]. Available: https://www.elconfidencial.com/empresas/2017-06-20/endesa-iberdrola-factura-recibo-electricidad-clientes-comercializadoras{_}1402021/

[145] PayScale, “Software Engineer Salary (Spain).” [Online]. Available: https://www.payscale.com/research/ES/Job=Software{_}Engineer/Salary

[146] ZIGT Media, “Television in Spain: Prices — ZIGT Media.” [Online]. Available: https://zigt.com/television-in-spain-prices/

[147] AdEspresso, “Understanding Facebook Ads Cost,” 2016.

[148] Marketline, “Marketline Report Store - Household Products in the United Kingdom.” [Online]. Available: https://store.marketline.com/report/ohme4359--household-products-in-the-united-kingdom/

[149] Ofgem, “Retail Energy Markets in 2016,” no. August, pp. 1–25, 2016. [Online]. Available: https://www.ofgem.gov.uk/system/files/docs/2016/08/retail{_}energy{_}markets{_}in{_}2016.pdf

[150] REED, “Engineering salaries in the UK.” [Online]. Available: https://www.reed.co.uk/average-salary/engineering/systems

[151] Business Insider, “How much it costs to buy a 30-second ad on UK television - Business Insider.” [Online]. Available: http://www.businessinsider.com/how-much-it-costs-to-buy-a-30-second-ad-on-uk-television-2017-2/{#}itv-1

[152] Marketline, “Marketline Report Store - Household Products in the United Kingdom.” [Online]. Available: https://store.marketline.com/report/ohme4359--household-products-in-the-united-kingdom/

[153] Michigan Tech, “2018 Engineering Salary Statistics — College of Engineering — Michigan Technological University.” [Online]. Available: https://www.mtu.edu/engineering/outreach/welcome/salary/

Appendix A

List of Investigated Companies

Company	Definition	Location	Area of activity
ABB	It is a Swiss-based multinational technology company, operating in electrification products, robotics and motion, industrial automation and power grids. They have partnership with Enbala, a Canadian DERMS provider. This partnership combines software skills of Enbala with grid expertise of ABB.	Switzerland	Flexibility Management, Distribution & Transmission Network
Aclara	Company provides is specialized on smart grid solutions, providing both hardware and software. The grid analytics software of allows utilities to monitor and control their distribution network in near-real-time. They also offer a customer engagement platform which is based on making energy use of utility customers more transparent.	Washington, USA	Flexibility Management, Distribution & Transmission Network, Customer Relations
Alectris	The company combines monitoring data with financial and business operation data of the operator to deliver actionable information. For instance, the software can observe a low internal rate of return and connect this observation to a lacking performance due to initial bad design of an asset.	Greece	Generation
Amazon	It is the largest online retailer of the world whose service vary from online media consumption to supermarket shopping.	Washington, USA	Online retailer
American Electric Power	It is one of the largest electric utilities in the US, serving nearly 5.4 million customers in 11 states.	Ohio, USA	Utility
AppOrchid	It is a natural language processing company which provides custom-made applications for utilities. They aim at simplifying data to knowledge. For instance, the employee can ask a question like ”How do we optimize generation dispatch at 2 PM tomorrow if there is 40% cloud cover?” or “Which transformers are most likely to fail in the first week of August and how long will it take to restore service?” and the results can be seen in the apps. Inspite of the promising value proposal of the company, it does not provide publicly open use cases.	California, USA	Generation, Transmission & Distribution, Flexibility Management
Arria	It is a natural language generation company. They convert large datasets into simple-to-read texts.	United Kingdom	Distribution & Transmission Network
Autogrid	It is a software company providing DR and energy efficiency programs to utilities. Futher, they provide grid analytics services.	California, USA	Flexibility Management
Bidgely	The company provides a customer engagement platform based on data disaggregation for utilities	California, USA	Customer Relations
C3 IoT	The company use IoT, enterprise and third party data to create services such as predictive maintenance, fraud detection, customer segmentation for utilities.	California, USA	Generation, Flexibility Management, Distribution & Transmission Network

Ceiva Energy	Company uses Big Data analytics to control and manage distribution grid of utilities. They also offer white label customer engagement platform.	California, USA	Flexibility Management, Customer Relations
CertainSolar	Company aims at reducing the uncertainty on potential savings of a solar project by accurately predicting future utility rates.	Massachusetts, USA	Generation
Choice	Company is focused on detecting electricity theft by using Big Data analytics. Their product Revenue Intelligence prioritizes and optimises the potential targets of theft for investigators.	Luxembourg	Distribution & Transmission Network
Cpower	Cpower is an energy management company, focusing on C&I customers. They create optimized energy management strategies, composing a portfolio of customers and making them participate in DR programs. Their pulling power is the partnership they have with Stem, a data-driven intelligent storage company. The partnership combines the DR expertise of Cpower with intelligent battery storage of Cpower. The result is a more profitable choice between attending a DR program or store the energy in a TUO DR market.	New York, USA	Flexibility Management
Comverge	The company provides software, hardware, and services to utilities for them to implement demand response programs.	Georgia, USA	Flexibility Management
conEdison	It is a utility, operating in the USA. The total number of customers served is 10 million.	New York, USA	Utility
Deepki	The company leverages data to provide energy management services to businesses. Their product Deepki Collect centralizes and visualizes energy data. Further, the second producs Deepki Ready is capable of calculating energy performance of the assets without necessity of an audit.	France	Flexibility Management
Duke Energy	It is a utility company, operating in North Carolina, USA and serving over 7.5 million customers.	North Carolina, USA	Utility
Ecofactor	Since its foundation in 2006, the company is providing utilities DRMS, focused on HVAC systems. They offer turnkey tailored energy efficiency programs on top the SaaS software they provide.	California, USA	Flexibility Management
Ecova	The company provides utilities customer engagement solutions by providing a platform. The platform is a real-time visualization of energy use tool for C&I and residential customers. They use this product also to offer turnkey energy management solutions for businesses.	Washington, USA	Customer Relations
EDF	It is the largest utility in France in terms of revenues. The company has operations in 5 more countries in the world. The total number of customers served is 37 million.	France	Utility
Enbala	It is a software company providing DERMS and VPP for utilities.	Vancouver, Canada	Flexibility Management
EnBW	Operating in Baden-Wurttemberg, the company serves 5.5 million customers	Germany	Utility
Encycle	Company is a DR provider, specialized on HVAC systems of C&I customers. They don't offer a software for utilties to manage DR programs, instead they offer firm capacity with their assets.	Ontario, Canada	Flexibility Management
Enel	It is the largest utility of Italy. The company has operations in 30 countries, serving nearly 64 million customers worldwide.	Italy	Utility
EnergyHub	It is a cloud-based software solutions company founded in 2007. They provide services for utilities, homeowners and thermostat manufacturers. Their expertise is mostly on DER management. However, their DRMS Mercury, particularly focuses on residential BYOT model of DR programs.	New York, USA	Flexibility Management

EnergySavvy	Company provides utilities customer engagement platform. The platform provides targeted marketing campaigns, customer segmentation, and so on for utilities while empowers customers with energy saving recommendations.	Washington, USA	Customer Relations
EnerNOC	EnerNOC is an energy intelligence software company that was founded in 2003. The company is using Big Data analytics for energy management for commercial and industrial customers. Further, they sell firm capacity to utilities.	Boston, USA	Flexibility Management
Enervee	It is a SaaS company that combines data science, behavioral science and digital marketing to drive consumer energy savings. The company provides white label marketplaces for utilities which are open to their customers who are interested in making energy-efficient home appliances choices.	California, USA	Customer Relations
Engie	It is a utility company, operating in France. It has 2.2 million customers.	France	Utility
E.ON	It is the largest utility of Germany, also operating in 30 countries. The total number of customers served is 33 million.	Germany	Utility
Exelon Corporation	It is a utility services holding company operating in the US and Canada. The subsidiary utilities are Atlantic City Electric, BGE, ComEd, Delmarva Power, PECO and Pepco. The total number of customers served is over 10 million.	Illinois, USA	Utility
First Energy	Having 10 distribution utilities, the company serves 6 million customers.	Ohio, USA	Utility
FirstFuel	It is a customer engagement platform provider for utilities.	Massachusetts, USA	Customer Relations
Gas Natural Fenosa	Company operates in several countries, including Spain, Mexico, Brazil, Colombia, Chile, Argentina and Peru. The total number of electricity customers in Spain is 3.7 million.	Spain	Utility
General Electric	It is one of the largest technology companies in the world whose activities include aviation, health care, energy, transportation and many others.	Boston, USA	Generation, Transmission & Distribution, Flexibility Management
GE Energy Connections	It is a division of General Electric, specialised on hardware, software and control systems for electricity sector. Their DRMS product DRBizNet handles every step of a DR program; including, program creation, marketing, program performance tracking, customer engagement, device installation scheduling, device asset management, work order management of field installation of the demand response devices, capacity forecasting, dispatch optimization, event creation, customer event notification, event dispatching, tracking of customers' participation, real-time performance monitoring and measurement and verification of participation. The software is being used by over 200 organizations which results in management of 10 GW of DR from over 1,000,000 end-users.	Boston, USA	Flexibility Management
Geostellar	Company provides a marketplace in which home owners can determine how much electricity they can generate without a site investigation, resulting in lower acquisition costs for the installers and smarter decision-making for the home owners.	West Virginia, USA	Generation
Gridium	Company uses only smart meter data and zip code to identify energy management recommendations to commercial customers. They also partner with utilities for them to act as an energy manager for their customers.	California, USA	Distribution & Transmission Network

Honeywell	It is an multinational company providing products such as control products for heating, ventilation, humidification, and air conditioning equipment; security and fire alarm systems; home automation systems; lighting controls; building management systems; and home comfort consumer products. Their thermostats are suitable for automated demand response programs.	Minnesota, USA	Flexibility Management
Iberdrola	Iberdrola is a Spanish based utility, serving over 32 million customers in Spain, the UK, the USA and Brazil.	Spain	Utility
IBM	It is a multinational IT company, providing hardware and software for businesses. Their Big Data analytics services are used by various sectors, including energy	New York, USA	Information Technologies
Innovari	Company provides SaaS for utilities to manage DER assets. Their platform allows the utilities to become VPP operators.	Oregon, USA	Flexibility Management
Kevala	Kevala maps the node and substation locations and ranks them according to their profitability by taking into account the energy value derived from annual hourly solar production through weather data and historical wholesale prices. This helps investors to narrow their site search down to more profitable locations.	California, USA	Generation
kWh Analytics	Company aims at being the trustful 3rd party to investigate project performance for the investors. This will decrease the interests on solar project loans by increasing the credibility of the investment. The company achieves this by gathering data from more than 70.000 solar projects on its platform together with its risk management software.	California, USA	Generation
Lockheed Martin	It is an aerospace and defence company, providing services in energy management as well. Their product SEEsuite is a SaaS, managing the life cycle of entire DR program.	Maryland, USA	Flexibility Management
Mapbis	Company combines utility data with third party data such as demographics and parcel data to detect electricity theft.	Turkey	Distribution & Transmission Network
Nexant	It is a software company providing DR and energy efficiency programs to utilities. Futher, they provide grid analytics services.	California, USA	Flexibility Management, Distribution & Transmission Network
NextEra Energy	Having Florida Power & Light Company as its subsidiary, the company serves nearly 5 million customers.	Florida, USA	Utility
NextKraftWerke	It is the the largest VPP in Europe, operating not only in Germany but also in Austria, Belgium, France, the Netherlands, Poland, Switzerland and Italy. The total capacity reaches 3.200 MW with 4.500 DG units such as CHPs, biogas plants, solar PV sites, small hydro power plants and wind farms.	Germany	Flexibility Management
NexTracker	Company is focused on producing tracking systems for solar panels. Solar power plants lose electricity generation efficiency due to construction variability, terrain undulation and changing weather conditions. NexTracker proposes a tracking system, leveraging cloud positioning, fog or haze data to correct the panel direction by using machine learning algorithms.	California, USA	Generation
Nnergix	Company provides a generation forecast of renewable by using weather data	Spain	Generation
OATI	Open Access Technology International (OATI) is an energy software company founded in 1995. Their SaaS product webDistribute combines DR with DER assets. Its ability to model an entire distribution grid and locate the DER asset enhances the performance of DR.	Minnesota, USA	Flexibility Management

Open Energy	Company provides intelligent demand shifting capabilities for the assets in C&I customers for them to avoid peak prices.	United Kingdom	Flexibility Management
Opower & Oracle	OPower is a cloud-based software solution company providing services for utilities. The company was founded 2007 in the US and was acquired by Oracle in 2016.	California, USA	Customer Relations
PickMySolar	PickMySolar is a data-driven marketplace for panel installers.	California, USA	Generation, Customer Relations
PG&E	It is a utility operating in the USA. The total number of customers served is 5.4 million.	California, USA	Utility
PowerScout	Company uses light detection and ranging (LIDAR) data with detailed consumer data and advanced image recognition technology to determine if the households would adopt solar PV systems in a close future.	California, USA	Generation
REstore	It is a Belgian based DR provider. They hold 500 MW of power with a 99.8% reliability from their industrial partners. The company operates these DR assets as a VPP. Further, they provide SaaS DRMS solutions to utilities, for them to create their own DR portfolio.	Belgium	Flexibility Management
RWE	Company supplies 23.4 million customers in Germany, the Netherlands, the United Kingdom and Eastern Europe.	Germany	Utility
SaS	It is a multinational business intelligence company operating in many sector ranging from defence to energy.	North Carolina, USA	Information Technologies
Schneider Electric	It is a multinational company specialized on energy management and automation.	France	Generation, Distribution & Transmission Network
Siemens	It is a large multinational technology company operating in energy, healthcare, transportation, white goods and many other sectors.	Munich, Germany	Generation, Transmission & Distribution, Flexibility
Silver Spring Networks	Company provides smart grid hardware and software solutions. They offer utilities real-time information for grid monitoring & control and demand-side management solutions that	California, USA	Flexibility Management, Distribution & Transmission Network
Simple Energy	They combine behavioural science, Big Data analytics and digital marketing techniques to create a platform for utilities to engage with their customers through energy efficiency. Recently, the company added a utility marketplace solution to their product portfolio.	Colorado, USA	Customer Relations
Sobolt	Company uses data received from satellites to create various services. Their product DustDetect gives insights about the performance decrease of solar panels and recommends a cleaning schedule.	Netherlands	Generation
Southern California Edison	It is one of the largest utilities of the USA, serving over 15 million customers.	California, USA	Utility
Southern Company	It is a utility operating in four states in the USA. The total number of customers served is 9 million.	North Carolina, USA	Utility
Spirea	The company provides products and services for distributed power and control systems. The company maintains and improves grids, particularly microgrids. Their DERMS product Wave fulfils duties such as network awareness, asset monitoring and control, scheduling and dispatch, active and reactive power import and export control, voltage control, constraint management, forecasting, resource valuation, and optimal demand response. It is the only DERMS which takes into account the topology of the grid. It automatically responds to topology changes.	Colorado, USA	Flexibility Management
Sunverge Energy	The company provides solar PV and intelligent battery storage systems to home owners since its establishment in 2010. They connect their assets for operating a VPPfor utilities	California, USA	Flexibility Management

Tendril	Company aims at customer engagement and DR programs for utilities at the same time. Their customer engagement platform promotes behavioural DR and energy efficiency which can be used by utilities.	Colorado, USA	Flexibility Management, Customer Relations
TERADATA	It is a multinational business analytics and data management company.	Ohio, USA	Information Technologies
Tiko	It is a energy management provider for residential customer and a DR provider for the utilities. Further, it offers customer engagement platforms.	Switzerland	Flexibility Management
Vaisala	Company is one of the big players of weather forecasts to predict renewable energy generation.	Finland	Generation
Vestas	It is a large multinational wind turbine manufacturer.	Aarhus, Denmark	Generation
Weather Company	It is a weather forecast company acquired by IBM in 2016. Their software product outage management combines weather data with outage data and learn from them to predict the number of outages and infrastructure damage within a utility service territory with a 72 hours lead time.	Georgia, USA	Distribution & Transmission Network

Appendix B

Business Model Canvas

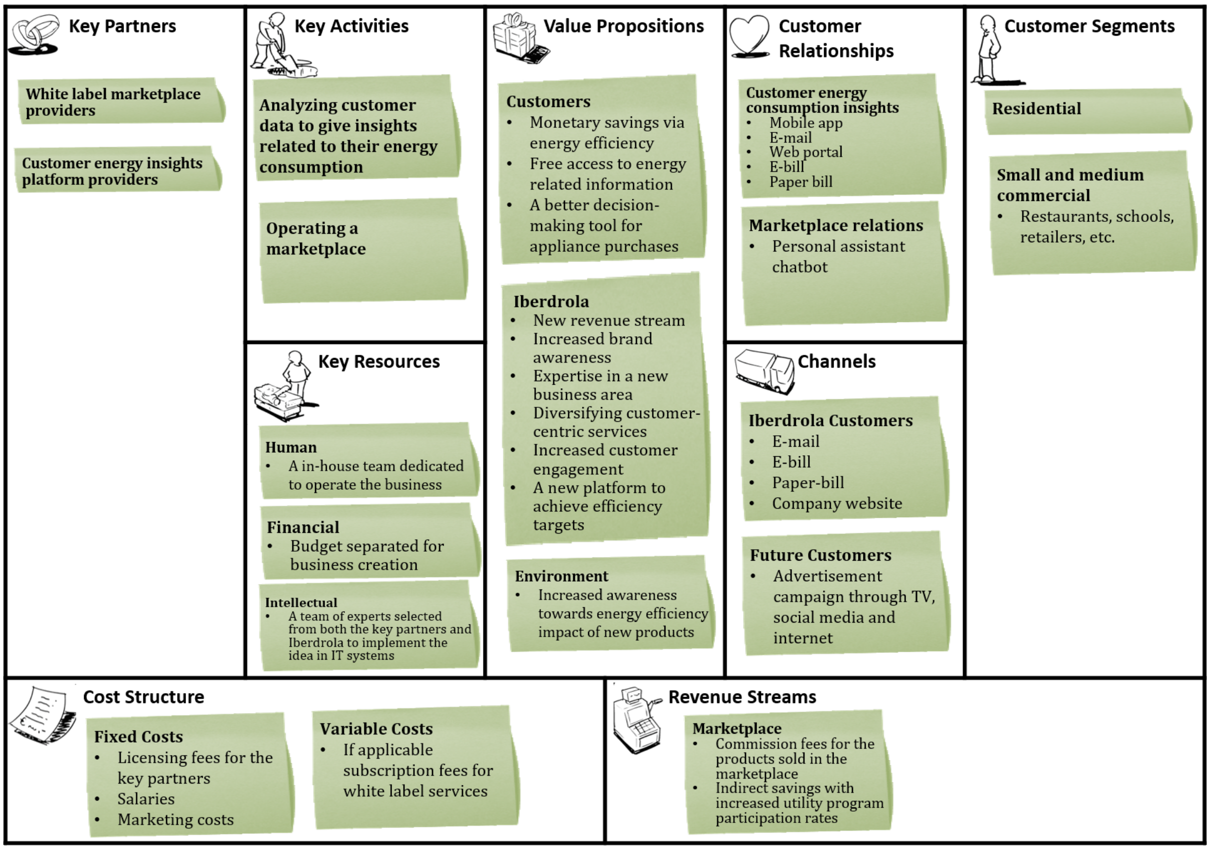

Figure B.1: Business Model Canvas of *Iberdrola Efficiency Shop*