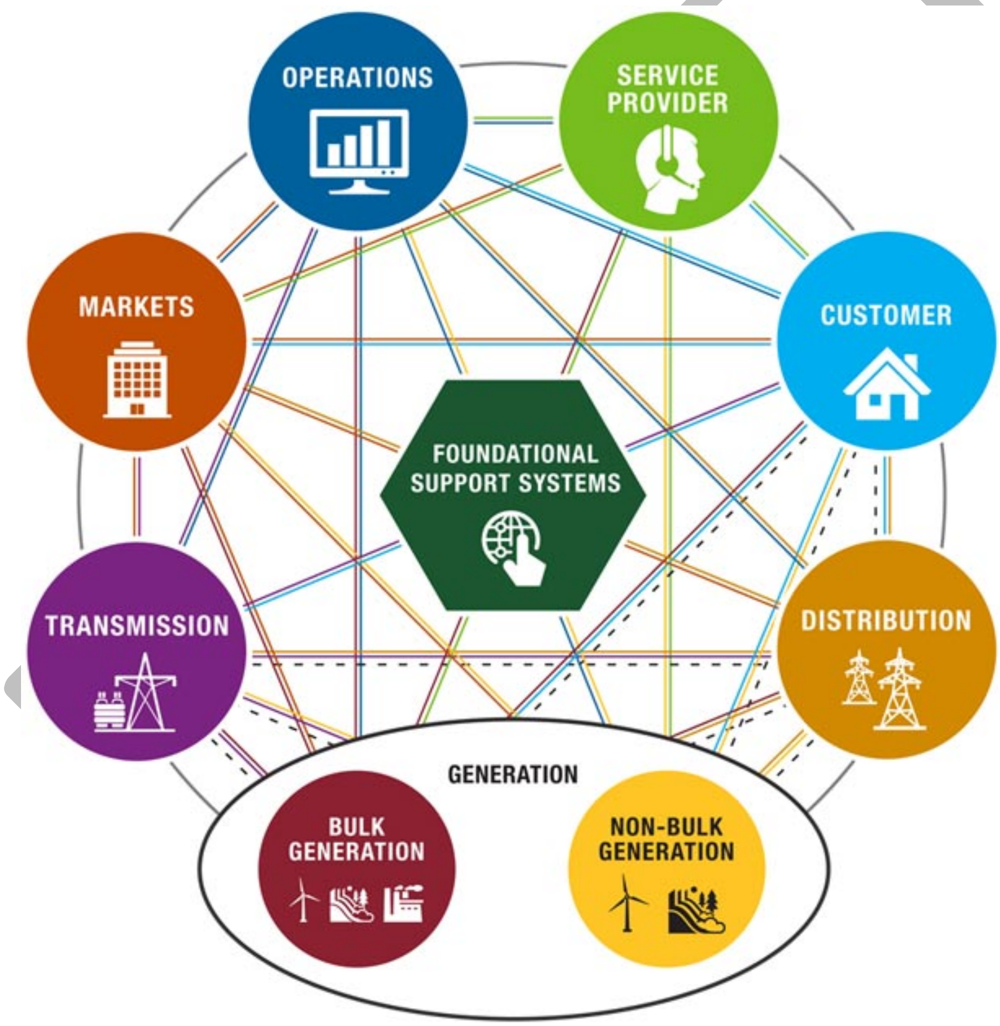
Big Data Analytics in the Smart Grid

White Paper #1 – Draft

Topic: Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid: Introduction, Benefits, Challenges and Issues

Authored by: IEEE Smart Grid Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid Working Group



CONTRIBUTORS

Working Group on Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid

Chair

Laura L. Pullum	IEEE Computer Society

Contributors

Anish Jindal	IEEE Communications Society
Mehdi Roopaei	IEEE
Abhishek Diggewadi	IEEE Power and Energy Society
Merlinda Andoni	IEEE (Student)
Ahmed Zobaa	IEEE Power and Energy Society
Aftab Alam	IEEE Power and Energy Society
Abedalsalam Bani-Ahmed	IEEE Power and Energy Society
Young Ngo	IEEE Power and Energy Society
Shashank Vyas	IEEE Power and Energy Society
Rajesh Kumar	IEEE Industry Applications Society
Valentin Robu	
David Flynn	

Staff

Phyllis Caputo IEEE Smart Grid

Angelique Rajski Parashis IEEE Smart Grid

ACKNOWLEDGEMENT

IEEE Smart Grid Community brings together IEEE's broad array of technical societies and organizations through collaboration to encourage the successful rollout of technologically advanced, environment-friendly and secure smart-grid networks around the world. As the professional community and leading provider of globally recognized Smart Grid information, IEEE Smart Grid Community is intended to organize, coordinate, leverage and build upon the strength of various entities within IEEE with Smart Grid expertise and interest. Additional information on IEEE Smart Grid can be found at http://smartgrid.ieee.org.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABSTRACT

The concept of smart grid incorporates a network of generation, transmission and distribution components that undertake power delivery from bulk generation power plants and distributed generation to various types of loads. The components are governed and managed by intelligent devices from generation to consumption, and can be optimized based on environmental and economic constraints. A smart grid allows utilities to engage consumers in power generation at the residential and industrial level, and may implement a bidirectional power exchange. To enable being “smart”, a huge amount of data is exchanged between grid components and the enterprise systems that manage these components. Based on the application, information exchanged enables economically optimized bidirectional power flow between a utility and its customers. Data exchange is essential for controlling, monitoring and coordination between smart equipment in a smart grid subsystem. For optimal performance, big data analytics are a necessity, and local autonomous control is achieved when artificial intelligence is applied using machine learning techniques. This paper reviews the applications of big data analytics, machine learning and artificial intelligence in the smart grid. Benefits, challenges, impacts and problems of employing these techniques are presented. Some big data analytics approaches for computing and transmitting data are detailed.

Keywords: Smart Grid, Big Data Analytics, Machine Learning, Artificial intelligence, Cloud Computing, Edge Computing, Internet of Things, Data Acquisition Framework, Cyber-Security

1. IEEE SMART GRID BIG DATA ANALYTICS, MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE WORKING GROUP WHITE PAPER SERIES

This white paper is the first in a series of white papers developed by the IEEE Smart Grid Big Data Analytics, Machine Learning and Artificial Intelligence (BDA/ML/AI) working group. The intent of the series is to provide a concise view into the current status of, benefits of, challenges to, best practices in and standards for BDA/ML/AI in the smart grid.

The IEEE Smart Grid BDA/ML/AI White Paper Series will comprise the following white papers:

1. Introduction to BDA/ML/AI, Benefits, Challenges and Issues
2. Best Practices in Big Data Analytics for the Smart Grid
3. Big Data Analytics in the Smart Grid: Recommended Standards, Existing Frameworks and Future Needs
4. Potential Applications and Improvements / Solutions to Issues: A sub-series of application- and solution-specific white papers organized by IEEE Smart Grid domain and sub-domain categorization. The intent is to have this subseries of smart grid analytics white papers cover the scope of these domains and subdomains.

2. INTRODUCTION

The smart grid refers to an advanced communication and information infrastructure that enables optimization in energy production, transmission, distribution and storage. Other benefits involve system management automation, educated planning, lower costs and effort, and electricity system reliability improvement [1]. The characteristics of smart grids involve the whole spectrum of the power system, from generators and energy suppliers to end-consumers [2]. Smart grids include the ability to enable active customer participation, and facilitate accommodation of power generation and storage options. A perspective view of the smart grid shows one entity consisting of multiple domains. These domains can be viewed as a chain of domains for power service, starting from the generation and ending with the customer. However, these domains are coupled with the help of functional support systems that involve many aspects of data management and communications, insuring system resiliency and efficiency and subsequently economic and environmental projections. The domain definitions were adapted by IEEE Smart Grid based on National Institute of Standards and Technology (NIST) definitions [4]. A conceptual model of the smart grid domains and their interactions is shown in Fig. 1.

In addition to integration of distributed energy resources (DER), the key drivers for the development of the smart grid are recent technology breakthroughs in energy storage, electric vehicles (EV) and operation and efficiency improvements required to ensure network resilience and security of supply. Future energy systems shall include the legacy power equipment within the grid infrastructure, with estimates that the US electricity grid will require $2 trillion in network upgrades by 2030 [5]. According to the European Commission, the transition towards a more sustainable and secure energy system would require an investment of €200 billion per year in the EU for generation, networks and energy efficiency developments [3].

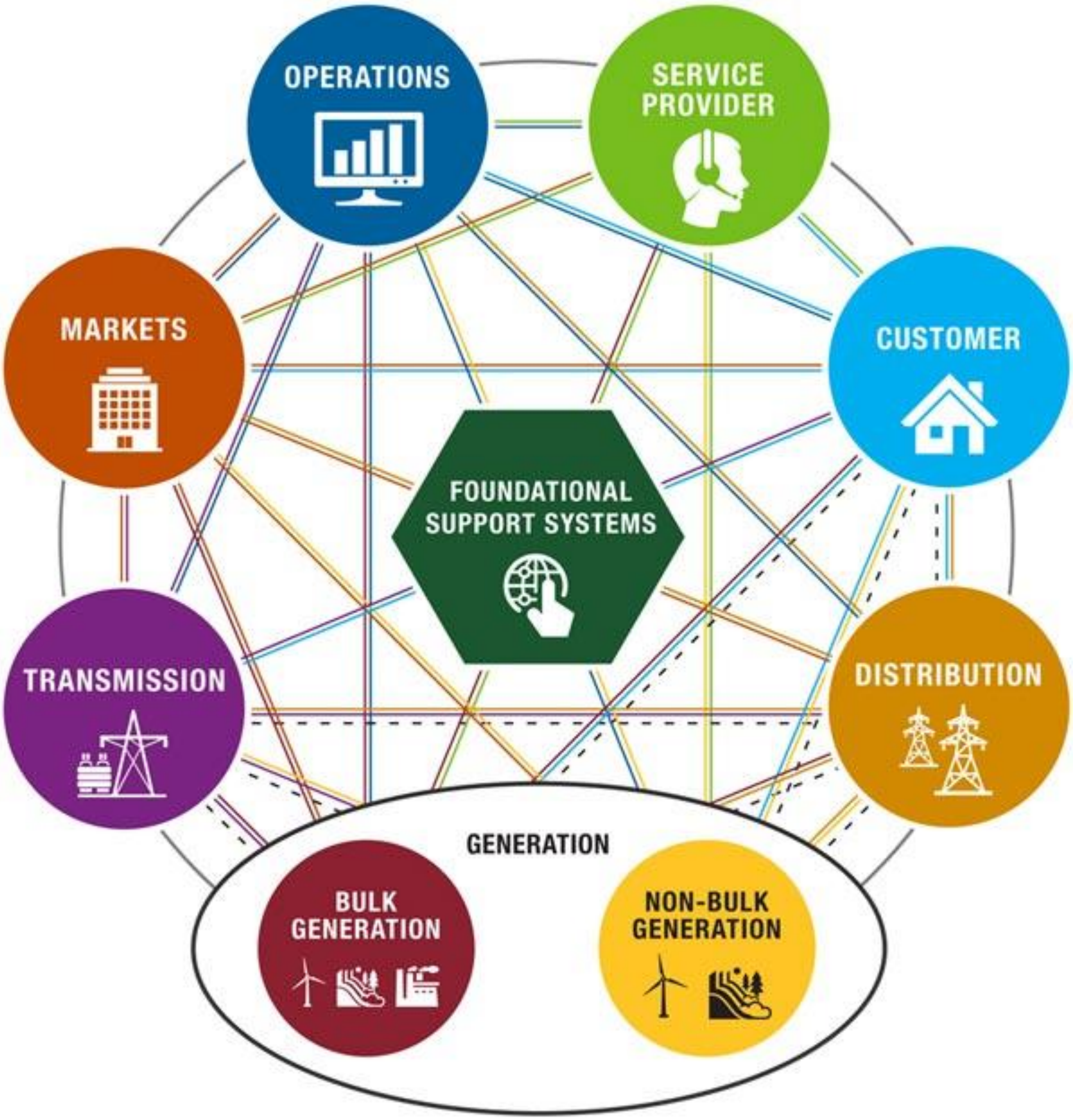


Figure 1: Interaction of roles in smart grid domains [12]

The bidirectional flow of electricity and information is an essential field in the smart grid. With the growing electricity supply from smaller-scale, decentralized generators, i.e., wind farms and residential rooftop photovoltaics (PV) panels, and microgrids, advanced sensor and metering technologies with integrated security protocols allow for envisioned advanced features of the smart grid, such as demand response, autonomous control, self-healing and self-configuration

[6]. Essentially, the smart grid provides significant improvements to traditional power systems that include six essential building blocks, namely, network, user, hardware, software, servers and data [7].

These characteristics make it challenging for traditional analysis, but ideal for the application of artificial intelligence, machine learning techniques and big data analytics. In this paper, we will use the term Big Data Analytics (BDA) to refer to the collective data analytics, machine learning and AI. The objective of BDA is to investigate the very large volumes of data produced by various components in the smart grid, and transform the data into meaningful inputs such as patterns of operation, alarm trends, fault detection, and control commands. For example, advanced machine learning applications for distribution transformers analyze the data aggregated in real-time for each transformer. The outcome of these learning applications may identify some operating trends leading to failure patterns of these devices and help anticipate future failures, and consequently, provide timely and accurate insights for predictive maintenance.

Research efforts in smart grid deployments have focused on advanced metering infrastructure (AMI), such as smart meters, communication, information, control and energy management systems for utilities and consumer-based equipment (e.g., smart home energy controllers and building monitoring systems). Moreover, other application areas include the integration of automation, control and real-time monitoring of advanced sensors and monitoring equipment. This can be accomplished with field devices, such as phasor measurement units and intelligent electronic devices (IED), at the transmission level and automated feeder switches, and network protection relays, voltage regulators, and capacitor controllers at the distribution level. These actions aim to enhance power system performance and diagnostics that will lead to cost reduction.

A significant portion of the smart devices being deployed is related to the massive rollout of smart meters currently taking place in many countries. The number of smart meter readings for a large utility company is expected to rise from 24 million a year to 220 million per day [7]. Approximately 22GB of smart meter data is being generated by 2 million customers per day [8]. Assuming that an application requires data collection in 15 minute intervals, 1 million devices would result in 35.04 billion data entries with a total volume of 2920 Tb per year [7]. The other portion of smart grid devices relates to cutting-edge network devices, such as IEDs being installed in power system networks to monitor key network parameters and generation and consumption in real time, control power flows, exchange information with each other and have local decision-making capability.

Traditional approaches of data analysis are inadequate to cope with the high volume and frequency of data generated within the smart grid paradigm by various distributed sources. This makes optimization and smart management challenging and computationally intensive. Data are generated by multiple heterogeneous sources including sensors, IEDs, smart meters, smart appliances, distribution automation data, third-party data, asset management data and weather station data playing an increasingly important role for managing intermittent DERs [7]. Data need to be transformed into actionable insights by applying high volume data management and advanced analytics (i.e., BDA) [9]. Essentially, BDA represents advanced analytics, such as predictive analytics, data mining, statistical analysis, machine learning and AI techniques, which operate on large data sets having one or more features of big data [10].

Local and distributed control architectures can provide solutions that can reduce the data transmission load and computational resources required, as opposed to centrally controlled decision-making. BDA techniques can provide solutions as the complexity of the power system continues to grow [11].

3. CONCEPTUAL DEFINITIONS of BIG DATA ANALYTICS, MACHINE LEARNING, AND ARTIFICIAL INTELLIGENCE

Before moving forward, we need to understand several terms which are interrelated (as illustrated in Fig. 2) and commonly used in different contexts while performing analytics in the smart grid. These terms are:

Statistics: It is the study of the collection, analysis, interpretation, presentation, and organization of data. Further, it can also be defined as the mathematics of estimating parameters of populations based on data from different representative samples of those populations. In statistics, the standard procedure for statisticians is to start with a null hypothesis (a default position that there is no relationship between two quantities) which is compared with an alternate hypothesis (a position that states there is a relationship between two quantities). The decision to reject a hypothesis is taken on the basis of various statistical tests which are performed on different population samples.

Data Analytics: It is the discovery and communication of meaningful patterns in data [13]. Data analytics is a (sometimes automated) process used to discover novel, valid, useful and potentially interesting knowledge from large data sources which is otherwise difficult to uncover. If statistics is to be considered a branch of mathematics, data analytics is inclined towards performing the same functionality for computer science. Visual tools and techniques are the preferred means of communicating the results of performing data analytics.

Machine Learning: It is the ability of machines (associated with computers) to learn automatically without being explicitly programmed. It deals with representation and generalization of data and creates a representation of instances and functions which are evaluated on these data. Generalization is the unique property that the machine learning systems will try to yield, that is, the ability of the systems to perform well even on unseen data instances.

Artificial Intelligence: It is the intelligence exhibited by machines, as opposed to natural intelligence exhibited by humans or animals. AI encompasses techniques which can endow an object or a program with human-like intelligence. AI also includes intelligent agents, entities that perceive their environment and take actions based on that perception.

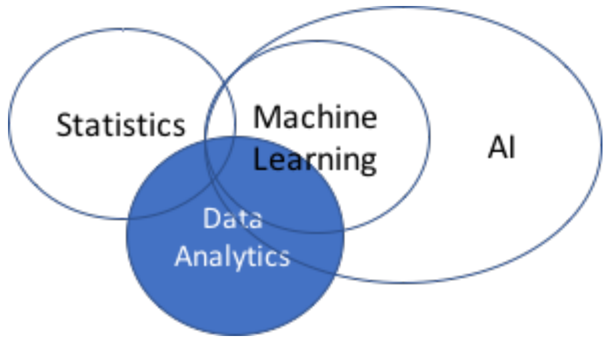


Figure 2. Venn diagram illustrating the interconnectedness of statistics, data analytics, machine learning and artificial intelligence.

The purpose of different types of analytics change as we move along the continuum of value (Fig. 3) as follows:

- *Descriptive analytics* aim to provide information about what happened and it comprises the first step that tries to identify useful information/data for further processing. It might include data visualization, data mining or aggregation of reports.
- *Diagnostic analytics* aim to understand the cause of events and system behavior and tries to identify challenges and opportunities.
- *Predictive analytics* are used to make probabilistic predictions to identify trends with the aim to determine what might happen in the future.
- *Prescriptive analytics* are applied to identify the best outcome to events, given the system's parameters, and draw strategies to deal with similar events in the future. It uses tools such as simulation techniques and decision support to explore optimal strategies to best take advantage of a future opportunity or to mitigate a future risk [14].

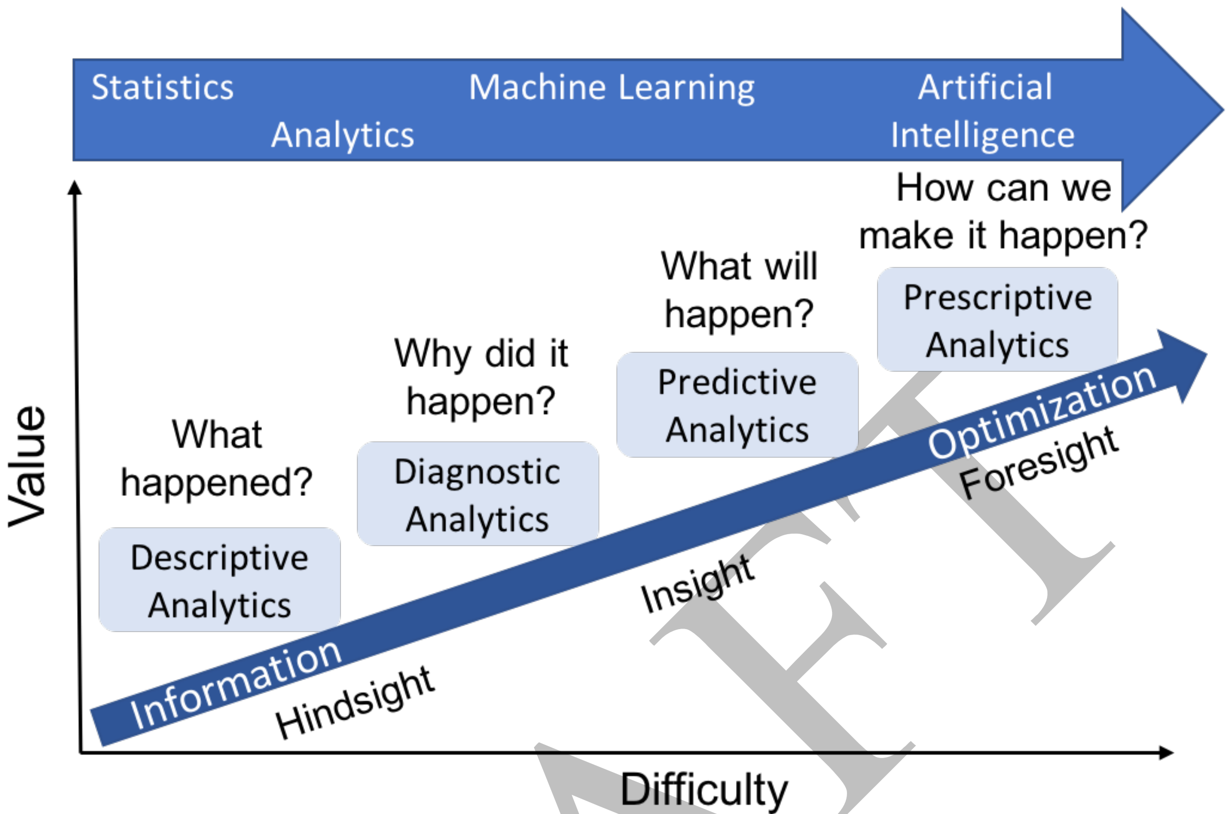


Figure 3. Statistics, Analytics, Machine Learning and Artificial Intelligence in context of type of analysis conducted. (Adapted from [15])

3.1 Typical Applications of Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid

To illustrate how big data analytics, machine learning and artificial intelligence are used in the smart grid, this section provides examples of typical applications relevant to the smart grid. These examples are provided as illustration and are, in no way, comprehensive. Section 4 lists additional expected smart grid-relevant applications. At present, there is a lack of typical applications of artificial intelligence in the smart grid beyond the application of ML. However, there are a growing number of new models, e.g., deep learning and reinforcement learning, that show promise towards enabling AI use in the smart grid.

3.1.1 Big Data Analytics Applications

Data generated in the smart grid are difficult to handle with traditional analysis techniques to produce actionable information within useful timeframes, as required by the nature of smart grid operations. Smart grid data can be classified as big data according to the 5Vs (Volume, Velocity, Variety, Veracity and Value) model shown in Fig. 4. Smart grid data exhibits each feature of the model as described in Table 1.

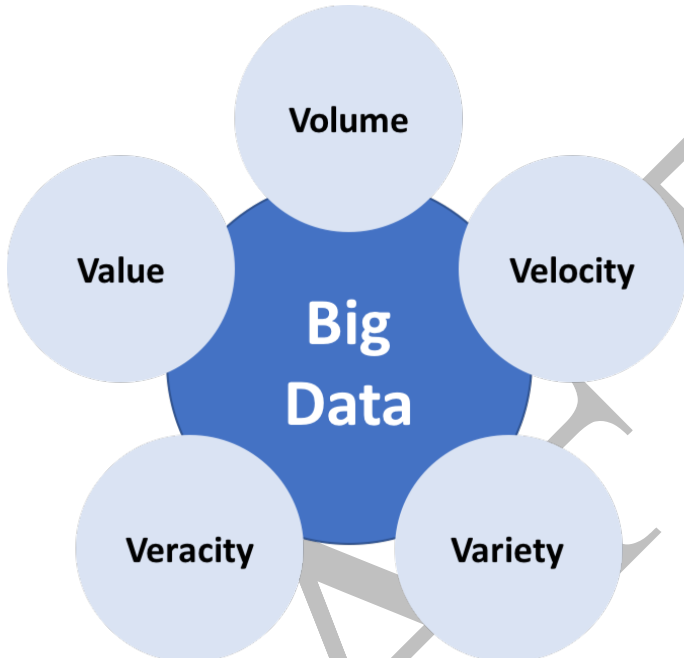


Figure 4: The features of 5Vs Big Data model [7]

Table 1 Smart grid data compliance with the 5Vs Big Data model [7]

Feature	5Vs Model	Smart Grid
Volume	Number of records and required storage	High volumes of data from smart meters and advanced sensor technology
Velocity	Frequency of data generation, transfer or collection	If smart meter data are collected every 15 minutes, 1 million devices result in 35.04 billion data entries or 2920 Tb per year [7]. The frequency data are collected is crucial for real-time monitoring and analysis.
Variety	Diversity of sources, formats, multidimensional fields	Existence of structured (e.g., relational data), semi-structured (e.g., web service data) and unstructured data (e.g., video data)
Veracity	Reliability and quality of data	Reliable data are crucial to ensuring safe system operation and stability.
Value	Extracting useful benefits and insights	Applications derive value from smart grid data, e.g., predicting future generation and demand.

The results of big data analytics can be used to predict and understand end-consumer behavior, to improve network resilience and faults, to enhance security and monitoring, to enhance performance and to optimize available resources and future planning.

3.1.2 Machine Learning Applications

Machine learning algorithms are particularly used for clustering data gathered from the smart grid domain [16]. Data are clustered to form groups with similar characteristics (natural classification), e.g., grouping together data points with similar active/reactive power profiles for transmission system operator (TSO) studies. Other examples include identifying low voltage (LV) feeders with similar load patterns for distribution system operator (DSO) studies to be compressed or summarized into cluster prototypes (e.g., generating representative days for wind production and their inclusion into network simulations). Further, ML can use smart grid data to understand the underlying structure, to gain useful insights, detect anomalies and generate hypotheses, etc. (e.g., detect theft and understand user consumption behavior at a particular feeder). Predictions play a significant role in power systems as they are typically used to plan future aggregated electricity demand, future system supply, estimating flexibility and reserve services requirements or operational management of distribution networks.

3.1.3 Multi-Agent System Applications

An application of AI in the smart grid context is multi-agent system (MAS) modelling. As the power system becomes more decentralized, market-oriented, multi-variable and complex, control and decision-making through a centralized approach becomes challenging as it requires significant computational power to determine optimal decisions for the entire system without significant delays. An improved approach is to divide the power system into more autonomous units that can make some of the decisions locally, following decentralized or distributed control. MAS approaches can be used to solve complex problems in an efficient, scalable and distributed way [17]. Potential applications of MAS in the context of the smart grid include the control of microgrids, fault management and disturbance diagnosis, self-healing and power restoration, voltage control, frequency control, demand side management based on agent architecture, energy consumption optimization and scheduling and coordination of storage devices.

3.2 The Need for Data Analytics in the Smart Grid

The smart grid gathers data from diverse sources and stores it to be consumable by analytics. Managing smart grids to provide smart energy requires advanced machine learning techniques to collect accurate information in an automated fashion, automate decision-making and control events in a timely manner at both the local and system-wide level. Important progress has been made for using field data acquired from smart devices mounted in substations, feeders, and numerous databases and models across the utility enterprise. There are several sources of data in smart grids on markets, equipment, geography and power system data which can be used to predict states, provide situational awareness, analyze stability, detect faults and provide advance warning. Therefore, analytics (comprising BDA, ML and AI) have a significant role to make the grid more intelligent, efficient and productive. Analytics can be applied to signal, event, state, engineering operations, and customer analytics, in sum enabling high-level and detailed insights into grid situational awareness. There are several types of analytics models, namely descriptive, diagnostic, predictive, and prescriptive models (recall Fig. 3). These can be applied for the smart grid, for instance, descriptive models describe customer behaviors for demand response programs. Diagnostic models are used to understand specific customer behaviors and analyze their power-related decisions. Each type of model can provide valuable input to create models that predict future customer decisions and hence, power needs. Finally, prescriptive models can provide high level analytics to influence smart grid marketing, engagement strategies and decision making.

Power systems are required to evolve for dynamic and flexible interaction with consumers participating in the electricity markets, LV control automation, distribution management system (DMS) integration, microgrid control and balancing, proactive fault identification, self-healing and resource optimization. Smart grid systems are becoming increasingly complex and interconnected, exhibiting characteristics of a “system of systems”.

As explained above, energy systems need to evolve to account for distributed power generation and the dynamic processes of demand management, load control and energy storage management. The energy system is currently experiencing significant changes, due to changes in regulatory frameworks and policies that relate to sustainability. This has resulted in significant growth in the volume, variety and velocity of data, a significant increase in stakeholder number and diversity, but also providing new business opportunities for improved economics and reliability. The need for data analytics and novel technologies is relevant to every stakeholder of the energy system, including the system operator, market operator, regulator, service provider, consumers, transmission and distribution system operators and service providers and generators.

3.2.1 Cloud Computing

Cloud computing provides the ability to connect to software and data on the cloud (the Internet) instead of a local computing network or a local hard drive. It is the most recent successor to virtualization, cluster computing, utility computing, and grid computing. Cloud computing centers largely on the outsourcing of computing needs and storage to cloud services. It is a system where users can connect to a vast network of computing resources, data and servers that are available usually on the Internet.

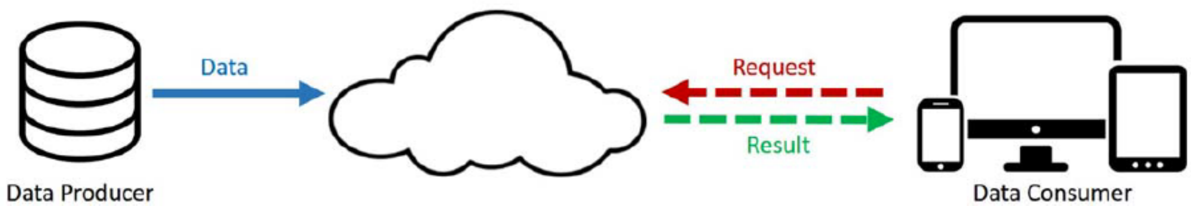
Virtualization is the foundation of cloud computing. Cloud computing, as defined by Forrester [18] is given as a pool of abstracted, highly scalable, managed compute infrastructure capable of hosting end-customer applications and billed by consumption. The key feature of cloud computing is that both the software and the information are stored on the massive network of cloud servers rather than on an end-user's computer. Cloud computing is a reliable choice for performing analytics as it has abundant resources accessible anywhere and at any time. There are many cloud computing platforms, like those offered by Amazon Web Services, AT&T's Synaptic Hosting, and to an extent, the HP/Yahoo/Intel Cloud Computing Testbed, and the IBM/Google Cloud. The grid can be made to run more efficiently by using cloud platforms v. massive local networks.

The opportunities and challenges of emerging and future smart grids can also be assisted by cloud computing. The advantages of using cloud computing are:

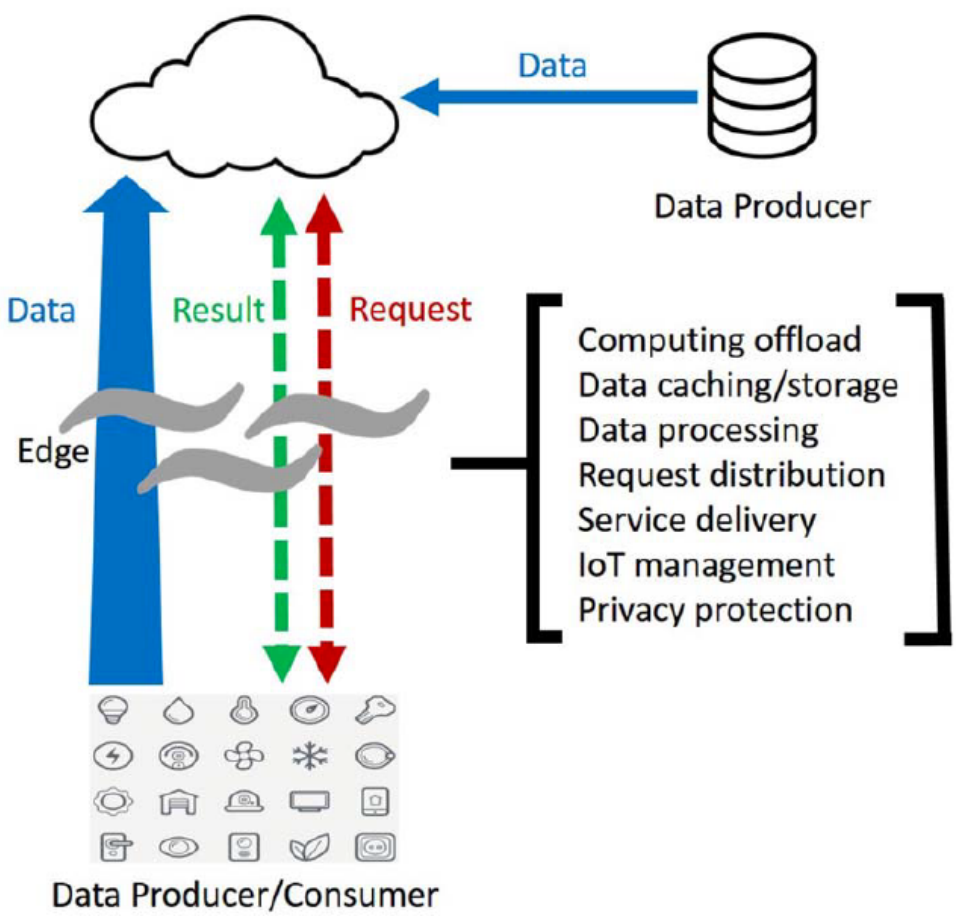
- **Self-service on-demand:** The user can individually provision computing capabilities as needed. Human interaction with each service provider is not required, as the service is provided automatically.
- **Broad network access:** Capabilities are available over the network. It can be accessed through standard internet access mechanisms.
- **Swift elasticity:** Cloud computing also supports the elastic nature of memory devices and storage. Depending on user demand, it can expand and contract.
- **Measured service:** Cloud computing also offers metering infrastructure to users. Users are thus able to provision and pay just for their consumed resources.

3.2.2 Edge Computing

In relevance to the smart grid, the Internet of Things (IoT) is a concept that brings large amounts of data generated by an embedded component of a variety of devices, sensors and networked entities. Forming a subsystem of the smart grid, these devices may be bundled in a cyber manner in order to aggregate predefined data at different data rates, e.g., customer power usage. Many applications are developed to utilize the data sourced at the edge (v. center) of the network, and they process the aggregated data locally, mitigating unnecessary data transmission to the cloud, and leading to an evolutionary transition to the concept of edge computing. It is estimated that by 2019; 45% of IoT-created data will be stored, processed, analyzed at the edge of the network [36]. Figure 5 illustrates the concepts of cloud computing and edge computing. The last resort of the data in cloud computing is a data consuming application, while in edge computing, the data are produced and consumed locally. Putting the computing close to the data sources reduces response time and energy consumption compared to a cloud computing solution. The smart grid benefits of both cloud and edge computing depend on the specific application of same.



a) cloud computing platform



b) Edge computing platform

Figure 5. Computing platforms a) Cloud and b) Edge [37]

3.3 Potential Impact of using Data Analytics in the Smart Grid

According to a NIST report, the benefits of modernization of power grids are as high as five times the one-time development cost [19]. Initial assessments by the American Council for an Energy-Efficient Economy predict the use of information and communication technology (ICT) and smart appliances would save about $80 billion in America's annual electricity bill [20]. This would only be possible if analytics (BDA, ML and AI) are performed on data gathered in the smart grid. The major benefits of performing analytics include increased customer satisfaction, better resource utilization and improved quality of service. However, to conduct analytics, a proper data acquisition framework is required to collect, process and analyze the data.

3.3.1 Data Acquisition Framework

The general framework for collection and analysis of data in the smart grid is depicted in Fig. 6. The entities present in a smart grid environment include power generation unit(s), transmission lines, distribution stations and end-users. The end-users may be industrial, commercial or residential users. Apart from these, end-users also include EVs and plug-in hybrid EVs (PHEV), which have made their way into the electricity market due to their growing popularity in the transportation sector.

Data are gathered in the cloud infrastructure due to its various advantages as discussed above. As ICT is an integral part of the smart grid, data can be easily gathered from its associated entities as shown in Fig. 5. The data of smart homes can be gathered by placing smart meters and other sensors in smart homes. EV/PHEV data can also be collected whenever they communicate (with the smart grid or cloud) to get services or to exchange information. The access points (AP) are placed at appropriate places which help these entities exchange information directly or indirectly via the cloud. The data from EV/PHEV and the smart grid is exchanged directly or through the APs connected to the cloud via the internet. The data collected from smart meters in smart homes and offices are sent to the cloud using APs. In Fig. 5, distribution stations consist of various charging stations to provide charging to, or discharging of, EVs. These charging stations also send their consumption data to the cloud with the help of APs. The communication technologies that are generally used for transmitting such a large amount of data to the cloud are mentioned in Table 2. In this table, an object can be a smart home, PHEV or any other end-user entity. The terms Object-to-access_point and Access_point-to-cloud indicate medium- and long-range communication, respectively.

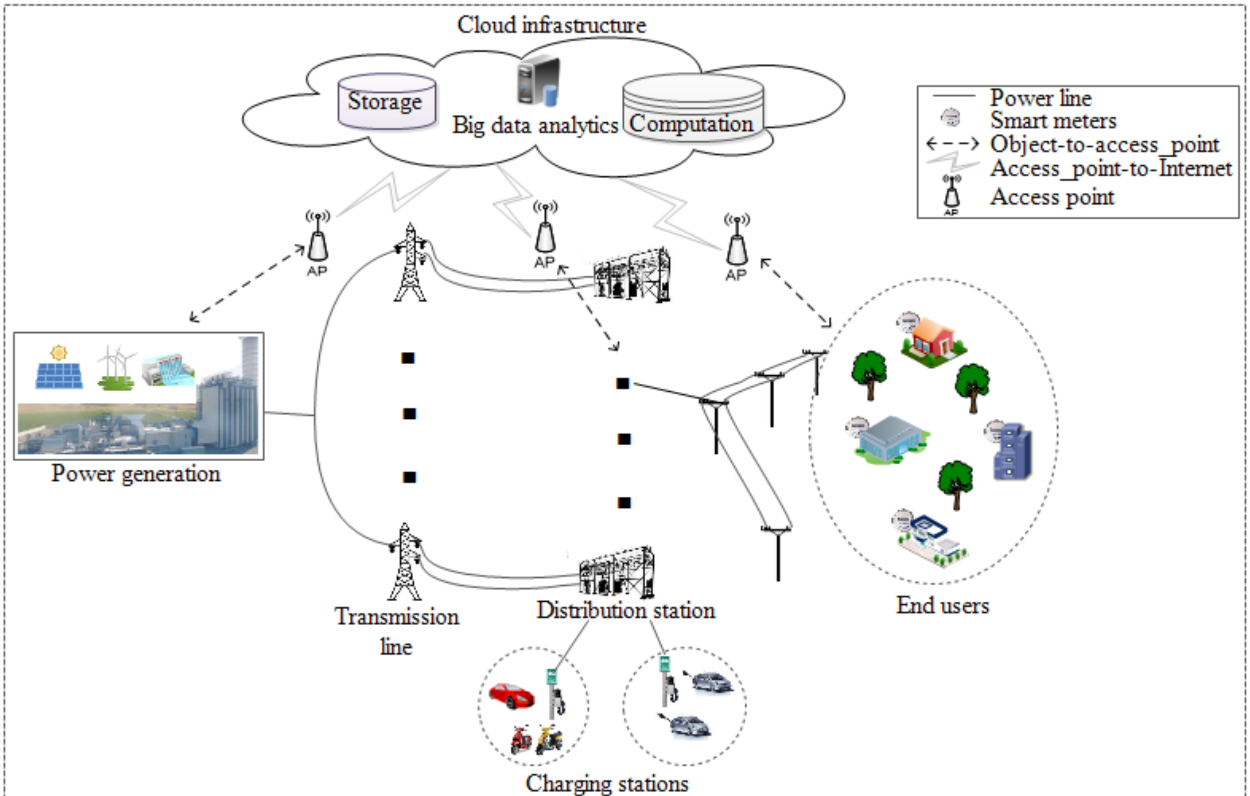


Figure 6: Framework for data acquisition and analysis in the Smart Grid

Table 2: Communication technologies used for data transmission.

Communication scenario	**Alternatives**	**Protocols used**	**Frequency bands**	**Data rate**
Object-to-access_point	DSA	IEEE 802.11af [21]	470-790 MHz	1 Mbps
	DSRC/WAVE	IEEE 802.11p [22]	5.850-5.925 GHz	3-27 Mbps
Access_point-to-cloud	Wi-Fi	IEEE 802.11 a/b/g [23]	2.4 - 5 GHz	1-54 Mbps
	WiMAX	IEEE 802.16 [24]	1.25 - 20 MHz	30 Mbps – 1 Gbps
	LTE/LTE-A	-	20 MHz – 100 MHz	300 Mbps – 3 Gbps

DSA - Dynamic Spectrum Access
DSRC - Dedicated short-range communication
WAVE - Wireless Access in Vehicular Environment
WiMAX - Worldwide Interoperability for Microwave Access
LTE - Long Term Evolution

LTE-A – Long Term Evolution Advanced

Apart from the framework depicted in Fig. 5, data can also be collected from other sources involving different technologies as given in Table 3.

Table 3: Other sources of data.

Data type sources	**Technology involved**	**Remarks**
Advanced metering infrastructure	Smart meters	Due to increase in adoption of smart meters in homes, data generated by these meters has increased significantly.
Distribution automation	Grid equipment	For real-time monitoring and control of the grid, the sensors are deployed in distribution systems which take multiple samples of data per unit of time.
Off-grid data	Third party datasets	To study the effect of utility policies on consumer behavior, utilities are integrating data from third party sources.

Once the data are gathered, data analytics techniques can be applied to these data and the results can be communicated back to the smart grid or utilities (and other entities, if necessary) for decision-making.

3.3.2 Extracting Value by using Data Analytics

The main aim of using BDA is to extract useful information (value) from the data. This value can be extracted from the gathered data after performing analytics on the data as illustrated in Fig. 7. Utilities and consumers may make informed decisions based on the resulting value.

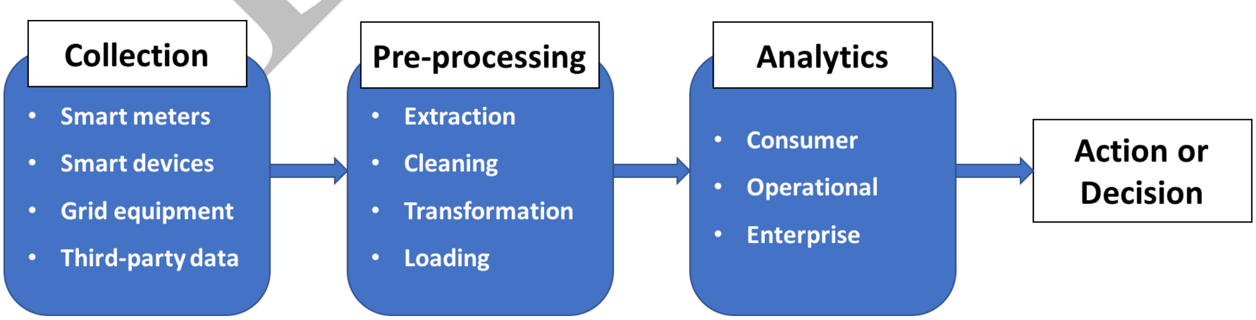


Figure 7: Example of extracting value using BDA [25].

As shown in Fig. 7, the first step is to collect data on which analytics will be performed. The data can be collected from many sources, e.g., smart meters, smart devices, and third-party datasets. Once the data are collected, the next step is data pre-processing. In this step, the data from various sources (in a variety of formats and possibly containing missing or erroneous values) are extracted. These data values are then cleaned to remove erroneous values. The data are then transformed to the target repository's format, after which data are loaded into a repository. Now data analytics techniques are applied on the pre-processed data to extract value (i.e., information) based upon which some informed actions or decisions can be made. The analytics performed to extract value can be divided into three broad categories viz. consumer analytics, operational analytics and enterprise analytics [25]. Consumer analytics include energy forecasting, consumption analysis and theft detection. Operational analytics include asset maintenance, outage management, and distribution optimization. Enterprise analytics include real-time grid awareness and visualization of data.

3.4 Security in the Smart Grid

In the smart grid environment, millions of devices and infrastructures are connected and interrelated using communication links and this exposes the grid to possible security vulnerabilities. Furthermore, cloud computing enables applications to be virtualized; however, sharing the platform with millions of users creates some security concerns. Besides, the smart grid must be highly scalable and accessible in real time application where low latency might be a huge challenge. Therefore, there are many cyber-security dimensions such as security availability, integrity, confidentiality and accountability that increase the risk of compromising the smart grid.

3.4.1 Cyber-Security Detection

The electricity industry has experienced the first known successful cyber-attack on a power grid suffered by the Ukrainian distribution company in December of 2015, where 30 substations were switch off, and over 200 thousand people were left without electricity for a duration from 1 to 6 hours [45]. Following the incident, the investigation uncovered a complex cyberattack consisting of spear-phishing emails with malware compromising the corporate networks, seizure of the supervisory control and data acquisition (SCADA) system and remotely switching off the substations, destruction of server files, and disabling of information technology (IT) infrastructure components. While the threat landscape continues to evolve, organizations are most concerned about cybercriminals using phishing and unknown malware methodologies. If malware evolves and re-engineers itself constantly as it spreads, the traditional malware detection methods of searching for specific code signatures will not be effective. Given the enormous amount of historical and ongoing cyber-security data collection available for analysis, AI and machine learning algorithms are powerful tools available to incorporate into the cyber-security strategy. The opportunities for significant impact in cyber security defense using AI may include the following: earlier detection of new or unknown malware based on historical baselines, improving the efficiency of updating the malware signature database, and increasing IT staff productivity through reduction of number of incidents that must be investigated and remediated.

3.4.2 Cyber-Physical Theft

Electricity theft in the power grid has been a major concern for utilities and causes huge economic losses. According to the Northeast Group, LLC, utilities lose up to $90 billion worldwide as a result of electricity theft [26]. Generally, electricity theft can be categorized into physical and cyber-attacks. Physical attacks are carried out by manually tapping the electrical supply from a neighborhood or directly from the feeder. However, with advancements in communication technology, one can also initialize remote theft on the grid using the Internet, known as a cyber-attack. The most commonly occurring cyber-attack on the grid is cyber-tampering, through which the attacker maliciously alters meter consumption data, ultimately leading to a reduced electricity bill for the building associated with the meter.

In part to improve the grid's resilience against such attacks, utility companies are investing billions of dollars on smart grid infrastructure. However, infrastructure changes alone will not prevent cyber-attacks on the grid. To address this issue, advanced ML and AI techniques can be leveraged by the utilities. These techniques are capable of learning user behavior from historical data and can easily identify anomalies in user behavior [27]. For example, a sudden reduction in the consumer's use can serve as an indicator for theft. Machine learning and artificial intelligence techniques need data to formulate precise models for detecting cyber-attacks. This data can be gathered from various sensors deployed across the electricity network and can be processed in real-time using cloud services and edge computing. So, deployment of machine learning models to analyze this data can help to generate useful information which can be used to identify and thwart cyber-attacks.

For example, the authors in [27] used the data gathered from sensors deployed on transmission lines, the distribution station and transformer levels to check power lines that are prone to physical tapping. The authors leveraged consumers' smart meter data and applied ML techniques on those data to identify the consumers who tampered with the smart meters.

3.4.3 Internet of Things and the Smart Grid

In 2016, the number of connected devices used worldwide was reported as 6.4 billion by Gartner [28] and it is expected to reach 20.8 billion by 2020. This report also notes that about 5.5 million new devices were connected each day in 2016 alone. The smart grid, as a big consumer of autonomous connected devices, not only utilizes millions of IoT devices, but also analyzes very large volumes of created data to make better decisions about smart grid networks. Ninety-one million smart meters are expected by 2020, along with 36 million smart interacting thermostats, and 183 million IoT residential devices. The predicted number of annual data points for 15-minute intervals will approach 11 trillion among these categories alone. This stream of data will empower deeper analysis of the grid at an increasingly granular level to act and react in real-time, thereby improving the efficiency of energy use [29]. Big data analytics and IoT are integral parts of the smart grid. Therefore, there is a critical need for integration of machine learning with IoT sensors and devices at different smart grid levels to analyze the whole ecosystem to optimize the cost, balance the energy resources and consequently form an intelligent smart grid as shown in Figure 8.

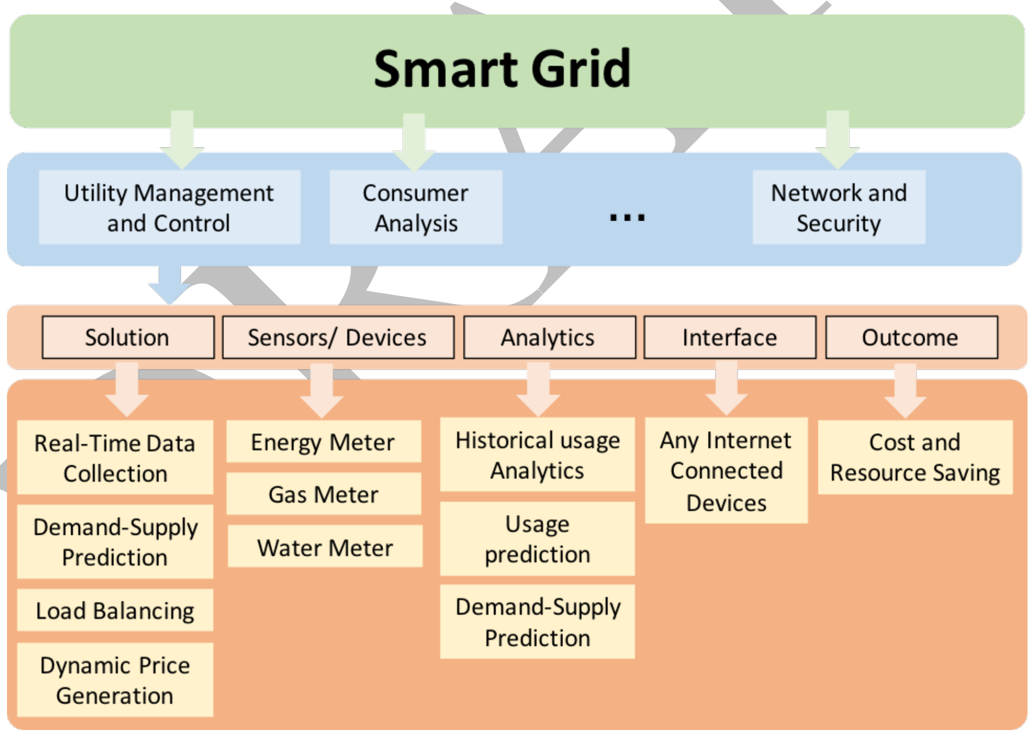


Figure 8. Machine Learning and IoT Integration: Utility Management and Control in Smart Grid

The smart grid is considered one of the largest beneficiaries of the IoT. IoT technology can support smart grids by providing high penetration of information sources such as power production, storage, transmission, distribution and consumption, and enhancing connectivity, automation and monitoring of each device. The modern and intelligent smart grid will not be cost effective without IoT technology. Smart grids have already attained extensive adoption in sensing, transmission and processing of information, and now IoT technology plays an important role in grid configuration and connecting legacy equipment. The application of the IoT in smart grids may fall into one or more of the following categories: (i) IoT is applied for employing different IoT smart devices for monitoring equipment status, (ii) IoT is applied for collecting information from equipment with the support of its linked IoT smart sensors and devices through diverse communication tools and (iii) IoT is applied for supervising the smart grid across application interfaces. In each of these categories, machine learning is used to provide real-time data analytics and generate informed decisions.

4. Overview of Benefits, Challenges and Issues of using Data Analytics in the Smart Grid

4.1 Current and Expected Use

Several US states are actively exploring how to use tools and technologies to realize a smarter grid [30]. Example US projects include:

- DMS Platform by the University of Hawaii [31]: The project showed both battery energy storage system and demand response technologies can be effective in reducing peak loads on the Maui Electric Company (MECO) system and of individual substations. The experience gained in this project will help MECO integrate distributed and renewable energy resources (PV, wind) with the operation of its central generators and transmission system.
- Perfect Power by Illinois Institute of Technology (IIT) [46]: The IIT and partners proposed to develop and demonstrate a system and supporting technologies to achieve "Perfect Power" at the main campus of IIT. Plans are for a self-healing, learning and self-aware smart grid that identifies and isolates faults, reroutes power to accommodate load changes and generation, dispatches generation and reduces demand based on price signals, weather forecasts, and loss of grid power.
- Some utilities are using predictive tools utilizing phasor measurement data for forecasting possible geomagnetic disturbances and protecting transformers from damage [38].
- A popular application is asset management where utilities are utilizing predictive analysis to determine when their assets require maintenance [39].
- Oklahoma Gas and Electric is coupling AMI with time-based rates and in-home displays to reduce peak power usage. This may enable the utility to defer the construction of a 170 MW peaking power plant [32].
- On July 5, 2012 a severe windstorm in Chattanooga, TN resulted in 80,000 customers losing power. The city was able to restore power to half of the affected residents within 2 seconds using automated feeder switching [32].
- Using synchrophasor data for real-time control, the Western Electricity Coordinating Council determined that it can increase the energy flow along the California-Oregon Intertie by 100 MW or more, with an estimated reduction in energy costs of $35-75 million over 40 years without additional high-voltage capital investments [32].
- Data mining and machine learning-based analytics on historical events data has been provided in an integrated hardware-software platform for real-time event detection by

Tollgrade communications. This has demonstrated the use of predictive grid-analytics on data collected by the company's proprietary smart-grid sensors for preventing blackouts to occur [43]. The platform identifies anomalies on a distribution feeder before they magnify and thus enables a proactive, predictive strategy for managing real-time events, similar to the concept envisaged in [44] for predictive detection of unintentional islanding in feeders. Utilities like Western Power Distribution have used this platform and have experienced improved reliability indices.

Some example smart grid projects from the UK include:

- Network Constraints Early Warning Systems by Scottish Power Energy Networks [40]: The projects aim to utilize large-scale, distributed smart meter data to monitor and detect the risk of power surges and voltage excursions outside technical operational limits, in different parts of the distribution network (or subnetworks). With millions of consumers, one cannot collect data in real time from every smart meter/consumer in the network, therefore an important question to answer is what data granularity is needed for different parts of the network to detect these excursions and provide early warnings of potential vulnerabilities.
- Thames Valley Vision by Scottish and Southern Electricity Networks [41]: The project was the first scaled deployment of LV substation monitoring of real time electricity data and their integration into a distribution management system. Data were used for substation categorizations, aggregation and forecasting of future network loading.
- Home-Offshore [42]: The project aims to use advanced robotic monitoring and sensing techniques for the remote inspection and asset management of offshore wind farms and their connections. Data collected will generate insights for diagnostic and prognostic schemes which will allow improved information to be streamed into multi-physics operational models for offshore wind farms.

4.2 Benefits and Impacts

The use of BDA can provide numerous benefits to all stakeholders within the power industry. For instance, DSOs can achieve improved coordination of the supply and the demand in distribution networks, consumers can improve their energy efficiency and achieve significant savings in their energy bills by closely monitoring their energy demand, energy suppliers or retailers can gain insights into consumer behavior and improve operational efficiency and energy generators can increase profits by optimizing their generation assets and production. TSOs can enhance their long-term planning by studying the behavior of different types of consumers in groups such as residential, commercial and industrial and by applying ML algorithms that effectively capture the impact of dispersed generation, which is often invisible at a transmission level.

Deregulated energy systems have increasingly relied on service providers such as demand aggregators, storage providers and virtual power plants. Service providers' optimal operation and profit maximization strategies are subject to uncertainty and need to account for the strategic behavior of multiple players participating in the energy market. Not only is the decision-making process multi-variable, but it needs to account for interdependencies of the decision variables at different time frames. For example, long term decisions might introduce constraints to medium term goals and operation. Useful techniques to cope with these issues can be derived from the field of game theory, which is suitable to study strategic interactions between multiple actors and to identify market equilibria. These combined with probabilistic modelling and stochastic optimization tools, allow service providers to reduce costs and maximize revenues. Agents can identify and learn bidding strategies in the energy markets, optimal dispatching, balancing and operation of their assets. In addition, service providers need to adopt predictive maintenance, prognostics and health management to adopt market and operation strategies that prolong the useful lifetime of their assets. The same tools and principles can be used for microgrid control in a decentralized or distributed fashion and coordination with the centralized grid.

Utility companies, especially at a distribution level, need to adopt big data techniques as the energy system continues to evolve. Most utilities are currently restricted to using descriptive and diagnostic analytics that aim to analyze historical data or events to understand the reasons behind their outcomes, e.g., system fault management or outage management. However, as energy systems grow to become more decentralized and reliant on intermittent DER, utilities need to deploy predictive analytics to evaluate potential future scenarios, such as evaluating future grid investment requirements and forecasting load impact and asset monitoring. Further, prescriptive analytics can lead to actionable insights for planning of generation and transmission/distribution capacity or optimization of renewable integration. For a more detailed look at challenges utilities currently face in network management, see Table 4.

Table 4: Network management issues

Grid Management Issue	Example	Consequence
Incomplete real-time monitoring	Lack of ● Equipment loading information ● Status of switches, transformer tap changers	● Inefficient equipment utilization ● Difficult to enable customers to connect distributed generators to

	● System momentary fault location ● Status of distributed resources ● Customer demand/load	grid and maintain system reliability ● No understanding of automated operations on feeder
Lack of system interoperability	Non-integrated systems for ● Customer information system ● Geographic information system ● Crew management ● Switch order management ● AMI ● SCADA	● Inefficient work processes ● Redundant/inaccurate data ● Longer outage duration ● Possible non-compliance of work processes with possible safety issues
Lack of Diagnostics	Lack of applications for ● Fault location ● Restoration switching analysis ● Voltage/Reactive power control ● Distribution state estimation	● Longer outage durations ● Inefficient use of crew hours ● No chance to reduce customer demand through voltage control at peak times ● Higher system losses ● Increased customer complaints for voltage out of range
Lack of Prognostics	● Reactive-based maintenance of transmission and distribution network assets ● SCADA needs improved scalability to deal with high volume of data monitored ● Lack of environmental monitoring	● Expensive maintenance ● Increased customer interruption ● Assumption that condition monitoring data are not dependent on circumstance data, such as the environmental conditions where a sensor is deployed, leading to potentially erroneous information

Table 5 describes some of the issues that can be improved significantly by the use of BDA in smart grids.

Table 5: Network management issues and benefits from BDA

Category	**Applications**
Connectivity model improvement	● Detect incorrect distribution transformer connectivity ● Correct meter phasing ● Auto-generate secondary circuit models
Asset management	● Identification of overloaded or high utilization assets (transformer replacement, cable monitoring) ● Identification of under-loaded transformers or stranded assets ● Identification of transformer voltage issues
Theft and consumer behavior	● Improve understanding of near real-time load profile of distribution feeder ● Identification of theft or tampered meters ● Using AMI or smart meters data ● Non-intrusive load monitoring
Fault location and diagnosis	● Identification of fault occurrence with the use of data from sensors and monitoring equipment, e.g., from substations, feeders or relay data ● Study fault patterns and types to help plan mitigating action such as tree trimming, equipment maintenance, etc.
Reliability analysis	● Health state estimation by data-driven methods and ML

4.3 Expected Barriers/Challenges/Issues

Significant challenges arise when applying BDA in the smart grid. These are summarized below:

- Data originate from numerous sources and come in different formats [33], such as smart meters, consumer data, web data, weather data, population data, geographic information system data, DMS data, energy management system, SCADA, etc. Data need to be integrated and interoperability between different devices and control levels must be ensured. New regulation and standardized processes are required for data collection and governance [7]. There is a lack of standards for data description and communication, essential for interoperability. Moreover, data integration and data sharing practices across institutions need to be defined for the benefit of all stakeholders [33].

- A significant issue is related to incomplete or missing data. Sensing equipment and smart meters are not installed everywhere in the grid and might be subject to failures or malfunctions. In this case, missing data may need to be extrapolated from available data. The issue becomes extremely important when it comes to ML techniques that rely on good quality and availability of data for algorithm training. Although there have been significant advances in ML approaches that handle the issue of missing data well, more needs to be done. Data integrity needs to be ensured by pre-processing techniques and error detection systems that ensure data quality.
- Utility companies need to install new IT equipment, such as sensor equipment, hardware and software tools and data storage devices [7], presumably at significant installation and operation costs. Moreover, utilities are uncertain regarding the transition from a centralized to a decentralized power network and the role of BDA in future energy systems [34].
- Research challenges remain in the big data analytics space. Cloud computing, distributed file management and new databases are developed, such as NoSQL databases for parallel processing. Several ML libraries have been developed for batch processing and novel techniques are being developed for stream processing. However, further improvement is required to adhere to the time response requirements of the smart grid [34].
- Data visualization can be challenging due to the size of the datasets and their high dimensionality. Current data analytics tools tend to have poor performance for visualization of large data sets, have high response times and are not easily scalable [34]. The scale of data can be handled somewhat by aggregation, however, in some cases aggregation may hide the specificity needed for insight.
- A continuously growing amount of data needs to be stored, typically in the cloud. This can lead to data explosion and the scalability issue [34]. Technical challenges such as the network bandwidth capacity and data security issues remain to be solved. Fig. 8 shows the increase in data volume as analytics evolve towards the smart grid model.
- Another issue for ML techniques includes the potential for inadequate training data, which may decrease confidence in the results of supervised ML models for previously unwitnessed situations. Advances in semi-supervised and unsupervised ML need to continue to address this issue.

- Special care is required for privacy, confidentiality and personal data protection. Significant problems to be resolved include data security protection, intellectual property protection, personal privacy protection, commercial security, network security and financial information [34].
- Security of data and smart grid infrastructure is critical. Smart grids can be vulnerable to cyber-attacks or cyber-physical attacks. Smart grid system functionality is vital for society overall and contains sensitive information that needs to be protected from malicious attacks and vulnerabilities. This includes all access points and protection of equipment which are inherently distributed, but also data storage security that can be obtained by adopting advanced cryptographic techniques and verification mechanisms [7].
- Energy consumption required for potentially intensive demand on computational power is another important issue that requires a cost-efficient and sustainable solution [35].
- Finally, as with every significant transition, there is a requirement for a well-trained workforce, both current and future human resources [7].
- Additional issues include analytics verification and validation, certification and regulatory compliance; and interoperability of all elements of the smart grid.

5. Conclusion

This white paper introduced the concepts and possible use cases of big data analytics, machine learning and artificial intelligence in the smart grid. The smart grid includes generation units, transmission lines and distribution systems. Distributed energy resources are integrated either separately or at the end-customer and then the power flows in different directions needing more modern devices for monitoring, transmitting and metering all power components. Monitoring, collecting and transmitting all the grid parameters involves a large amount of data. Meanwhile, it is infeasible to swiftly and correctly analyze these enormous data by traditional methods, which brings challenges to effectively enabling the smart grid. Therefore, powerful data analytics will be adopted to manage the massive number of data. Big data analytics, machine learning and artificial intelligence are approaches employed in the smart grid to manage the data collected from all power meters, sensors and other appliances. By the combination of these techniques, it is possible to know and expect load demand, generation volume and system disturbance, and then regulate the control automatically and quickly so the grid can be improved and avoid instabilities. Applying analytics effectively in the smart grid still faces numerous difficulties. Most power utilities are still uncertain of big data analytics, machine learning and artificial intelligence and thus this article defined terms related to advanced analytics approaches used in smart grid and explained the advantages, challenges and problems of utilizing these approaches. A brief review was also given on security matters, different impact and communication technologies in smart grid.

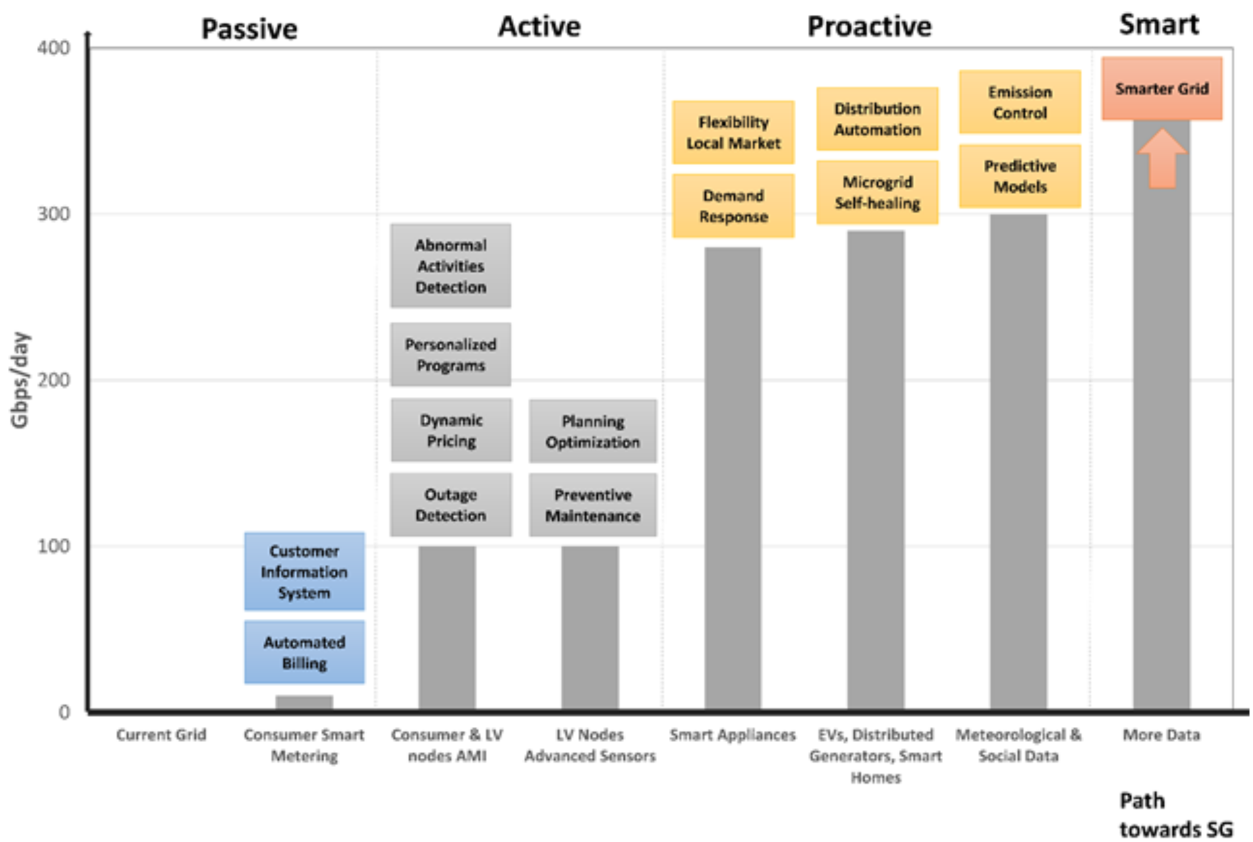


Figure 9: The evolution of grid analytics and future smart grid in big data applications: diverse data sources create high volumes of data and create more value. Adapted from [14]

References

[1] U.S. Department of Energy, "Smart Grid system report," 2009. [Online] Available: https://www.smartgrid.gov/files/systems_report.pdf.

[2] European Commission, "European SmartGrids Technology Platform: Vision and Strategy for Europe's Electricity Networks of the Future," 2006. [Online] Available: http://cordis.europa.eu/pub/fp7/energy/docs/smartgrids_en.pdf.

[3] European Commission, "A framework strategy for a resilient energy union with a forward-looking climate change policy," 2015. [Online] Available: https://www.eea.europa.eu/policy-documents/com-2015-80-final

[4] National Institute of Standards and Technology, "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0," 2014, [Online] Available: http://dx.doi.org/10.6028/NIST.SP.1108r3.

[5] P. Bronski, J. Creyts, M. Crowdis, S. Doig, L. Guccione, J. Mandel, B. Rader, H. Touati and D. Seif, "The Economics of Load Defection: How Grid-Connected Solar-Plus-Battery Systems Will Compete with Traditional Electric Service–Why it Matters, and Possible Paths Forward," Rocky Mountain Institute, 2015. [Online] Available: https://rmi.org/insights/reports/economics-load-defection/.

[6] F. Khan, A. ur Rehman, M. Arif, M. Aftab and B. Jadoon, "A survey of communication technologies for smart grid connectivity," *IEEE International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, Quetta, 2016, pp. 256-261.

[7] S. Sagiroglu, R. Terzi, Y. Canbay and I. Colak, "Big data issues in smart grid systems," *IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, Birmingham, 2016, pp. 1007-1012.

[8] A. Mahmood, M. Aamir and M. I. Anis, "Design and implementation of AMR Smart Grid System," *IEEE Canada Electric Power Conference*, Vancouver, BC, 2008, pp. 1-6.

[9] P. Diamantoulakis, V. Kapinas and G. Karagiannidis, "Big data analytics for dynamic energy management in smart grids," *Big Data Research,* vol. 2, no. 3, pp. 94-101, 2015.

[10] P. Russom, "Big data analytics," *TDWI Best Practices Report, TDWI Research*, pp. 1-6, 2011. [Online] Available: http://www.teradata.com/Resources/Analyst-Reports/TDWI-Best-Practices-Report-Big-Data-Analytic-(1)

[11] U. Ahsan and A. Bais, "Distributed big data management in smart grid," *26th Wireless and Optical Communication Conference (WOCC)*, Newark, NJ, 2017, pp. 1-6.

[12] IEEE Smart Grid R&D Committee, "IEEE Smart Grid Survey - Structure of Emerging Technologies," October 1, 2017. [Online] Available:

https://smartgrid.ieee.org/images/files/white-papers/white-paper-ieee-smart-grid-survey-structure-of-emerging-technologies.pdf.

[13] J. Strickland, “Predictive Modeling and Analytics,” Lulu, Inc., Colorado Springs, CO, 2014.

[14] T. H. Dang-Ha, R. Olsson and H. Wang, “The Role of Big Data on Smart Grid Transition,” IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), Chengdu, 2015, pp. 33-39.

[15] M. Davis, “Top 10 Moments from Gartner’s Supply Chain Executive Conference,” May 28, 2013. [Online] Available: http://blogs.gartner.com/matthew-davis/top-10-moments-from-gartners-supply-chain-executive-conference/.

[16] A. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651-666, 2010.

[17] A. L. Dimeas and N. D. Hatziargyriou, “Agent based control of Virtual Power Plants,” IEEE International Conference on Intelligent Systems Applications to Power Systems, Toki Messe, Niigata, 2007, pp. 1-6.

[18] S. Tilley, “Software Testing in the Cloud: Perspectives on an Emerging Discipline,” *IGI Global*, PA, 2012.

[19] National Institute of Standards and Technology, “Strategic R&D Opportunities for the Smart Grid: Advancing Measurement Science and Standards for Smart Grid Technologies,” pp. 1-36, 2013, [Online] Available: http://www.nist.gov/smartgrid/upload/Final-Version-22-Mar-2013-Strategic-R-D-Opportunities-for-the-Smart-Grid.pdf.

[20] J.A. Laitner, M.T. McDonnell and K.E. Martinez, “The Energy Efficiency and Productivity Benefits of Smart Appliances and ICT Enabled Networks: An Initial Assessment,” American Council for an Energy-Efficiency Economy (ACEEE), 2014. [Online] Available: http://aceee.org/research-report/f1402.

[21] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine and S. Pandey, “IEEE 802.11af: A Standard for TV White Space Spectrum Sharing,” *IEEE Communications Magazine*, vol. 51, no. 10, pp. 92-100, 2013.

[22] IEEE Std 802.11p, “IEEE Standard for Information Technology Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 6: Wireless Access in Vehicular Environments,” 2010. [Online] Available: https://www.ietf.org/mail-archive/web/its/current/pdfqf992dHy9x.pdf.

[23] IEEE 802.11: Wireless LANs. [Online] Available: http://standards.ieee.org/about/get/802/802.11.html.

[24] IEEE 802.16: Broadband Wireless Metropolitan Area Networks (MANs). [Online] Available: http://standards.ieee.org/about/get/802/802.16.html.

[25] C. L. Stimmel, "Big Data Analytics Strategies for the Smart Grid," Boca Raton, FL, USA: CRC Press, 2014.

[26] PR Newswire, "World Loses $89.3 Billion to Electricity Theft Annually, $58.7 Billion in Emerging Markets," 2014. [Online] Available: http://www.prnewswire.com/news-releases/world-loses-893-billion-to-electricity-theft-annually-587-billion-in-emerging-markets-300006515.html.

[27] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar and S. Mishra, "Decision Tree and SVM-based Data Analytics for Theft Detection in Smart Grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005-1016, 2016.

[28] Gartner, Inc., "Gartner Says 6.4 Billion Connected 'Things' Will Be in Use in 2016, Up 30 Percent From 2015," 2015. [Online] Available: http://www.gartner.com/newsroom/id/3165317.

[29] Navigant Consulting, Inc., "Navigating the Energy Transformation: Building a Competitive Advantage for Energy Cloud 2.0," 2016. [Online] Available: https://www.navigant.com/insights/energy/2016/navigating-the-energy-transformation.

[30] Litos Strategic Communication, "Smart Grid: An Introduction," for US Department of Energy, 2008. [Online] Available: https://www.smartgrid.gov/files/The_Smart_Grid_Introduction_200804.pdf.

[31] Hawaii Natural Energy Institute, "Maui Smart Grid Demonstration Project: Managing Distribution System Resources for Improved Service Quality and Reliability, Transmission Congestion Relief, and Grid Support Functions - Final Technical Report," for US Department of Energy, 2014. [Online] Available: https://www.hnei.hawaii.edu/projects/maui-smart-grid-demonstration.

[32] US Department of Energy, "2014: Smart Grid System Report, A Report to Congress", 2014. [Online] Available: https://www.smartgrid.gov/files/2014-Smart-Grid-System-Report.pdf.

[33] F. Luo, Z. Y. Dong, J. Zhao, X. Zhang, W. Kong and Y. Chen, "Enabling the big data analysis in the smart grid," IEEE Power & Energy Society General Meeting, Denver, CO, 2015, pp. 1-5.

[34] C. S. Lai and L. L. Lai, "Application of Big Data in Smart Grid," *IEEE International Conference on Systems, Man, and Cybernetics*, Kowloon, 2015, pp. 665-670.

[35] X. Miao, “Big Data and Smart Grid,” *International Conference on Big Data Science and Computing,* Beijing, China, 2014.

[36] D. Evans, “The Internet of Things: How the next evolution of the Internet is changing everything,” CISCO White Paper, vol. 1, pp. 1-11, 2011.

[37] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, “Edge Computing: Vision and Challenges,” *IEEE Internet of Things Journal*, 3(5): 637-646.

[38] PingThings, “Predictive Grid,” accessed 10/12/2017. [Online] Available: http://www.pingthings.io/.

[39] IBM, “Strategic Asset Management to Improve Electricity Network Planning and Maintenance,” [Online] Available: https://www-935.ibm.com/industries/energy/solutions/asset-workforce-management.html.

[40] SP Energy Networks, “Network Innovation Allowance Progress Report,” July 2017, [Online] Available: http://www.smarternetworks.org/Project.aspx?ProjectID=2039.

[41] Scottish & Southern Electricity Networks, “Thames Valley Vision,” [Online] Available: https://www.ssepd.co.uk/thamesvalleyvision/.

[42] M. Barnes, et al., “HOME-Offshore: Holistic Operation and Maintenance for Energy from Offshore Wind Farms,” University of Manchester Technical Report, [Online] Available: http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/P009743/1.

[43] “Bringing Our Electric Grid into the 21st Century,” 2014, [Online] Available: https://www.engerati.com/resources/bringing-our-electric-grid-21st-century.

[44] S. Vyas, R. Kumar and R. Kavasseri, “Data Analytics and Computational Methods for Anti-islanding of Renewable Energy Based Distributed Generators in Power Grids,” *Renewable and Sustainable Energy Reviews*, vol. 69, March 2017, pp. 493-502.

[45] Kim Zetter, “Inside the Cunning, Unprecedented Hack of Ukraine’s Power Grid,” *Wired*, 03/03/2016, [Online] Available: https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/.

[46] Illinois Institute of Technology, “Public/Private Partnership Creates Opportunity to Fundamentally Address Local Energy Crises,” [Online] Available: https://web.iit.edu/mediaroom/press-releases/2008/nov/20/publicprivate-partnership-creates-opportunity-fundamentally.

Acronyms

5V	Five Vs of Big Data Analytics (Volume, Velocity, Variety, Veracity, Value)
AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
AP	Access Point
BDA	Big Data Analytics
DER	Distributed Energy Resource
DMS	Distribution Management System
DSA	Dynamic Spectrum Access
DSO	Distribution System Operator
DSRC	Dedicated Short-Range Communication
EU	European Union
EV	Electric Vehicle
GB	Gigabyte
Gbps	Gigabits per second
GHz	Gigahertz
ICT	Information and Communication Technologies
IED	Intelligent Electronic Device
IEEE	Institute of Electrical and Electronics Engineers
IIT	Illinois Institute of Technology
IoT	Internet of Things
IT	Information Technologies
LLC	Limited Liability Corporation
LTE	Long-Term Evolution
LTE-A	Long-Term Evolution-Advanced
LV	Low Voltage
MAS	Multi-Agent System
Mbps	Megabits per second
MECO	Maui Electric Company
MHz	Megahertz
ML	Machine Learning
MW	Megawatt
NIST	National Institute of Standards & Technology
PHEV	Plug-in Hybrid Electric Vehicle
PV	Photovoltaics
SCADA	Supervisory Control and Data Acquisition
SQL	Structured Query Language

Tb	Terabyte
TN	Tennessee
TSO	Transmission System Operator
UK	United Kingdom
US	United States
WAVE	Wireless Access in Vehicular Environment
WiMAX	Worldwide Interoperability for Microwave Access