INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
*(Volume3, Issue1)*
Available online at: www.ijariit.com

# A Survey on Big Data Energy Based on Smart Grid

**Diksha M**
*M.Tech, Student Dept. Of CSE*
*Cambridge Institute of Technology*
*Bangalore.*
diksha.m364@gmail.com

**Prof. Raghavendra T. S**
*Associate Prof. Dept. Of CSE*
*Cambridge Institute of Technology*
*Bangalore.*
raghavendra.cse@citech.edu.in

***Abstract—: Energy is the most important part of human life. As a significant approach energy in smart grid is interconnected with power grid that involves sensors, deployment strategies, smart meters, and real-time data processing. It generates the data with high velocity, large volume, and diverse variety. In this paper we gave brief introduction on big data, big data architecture, smart grid, big data architecture for smart grid and its advantages and big data applications in smart grid environment and future challenges in energy domain and smart grid communication.***

***Keywords— Big Data, Energy, Smart Grids, Smart Grid Communication.***

## I. INTRODUCTION

BIG DATA technology is a new scientific trend. Driven by data analysis in high-dimensions, big data technology works out data correlations to gain insight to the inherent mechanisms. Data driven results only rely on an unrestrained selection of the system s raw data and a general statistical procedure. On the other side, procedures for traditional model-based analysis, especially decoupling practical interconnected systems are based on assumptions and simplifications. Generally, data driven analysis tools, rather than model based ones, are suitable to complex large- scale interconnected systems with readily accessible data.

Data in the power system have increased dramatically, leaving gaps and challenges; data processing is a major concern and increases with data growth. The 4Vs data i.e. volume, variety , velocity , and veracity in smart grids, which can be handled within a tolerable elapsed time or hardware resources by traditional model-based tools, have encouraged the development of an emerging paradigm—big data technology for power systems . Big data technology has been successfully applied as a powerful data-driven tool for numerous phenomena, such as quantum systems, financial systems, biological systems, as well as wireless communication networks. Major tasks of the architecture for these applications seem similar:1) big data modeling; and 2) big data analysis. The big data technology will also have a wide applied scope in power systems, and the results will be fruitful.

### *A. ROLE OF BIG DATA IN THE SMART GRID*

The data obtained from PMUs, smart meters, sensors and other IEDs have opened up a plethora of chances, such as predictive analytics, real-time vulnerability assessment, theft detection, demand-side-management ,economic dispatch ,energy trading, etc .Big data can help improve the smart grid management to a higher level. For example, by enhancing the accessibility of a customer's electricity consumption data, The demand response will be expanded and energy efficiency will be improved. Similarly, analysing data obtained prevent outages, maximize safety and ensure service reliability. Furthermore, electric utilities are using prediction data analytics for estimating several parameters which are helpful to operate the smart grid in an efficient, economical and reliable way. At the same time, calculating equipment downtime, accessing power system failures, and managing unit commitment can also be done effectively in high correlation with integrating distributed generation. Thus, the grid management state will be enhanced, the efficiency will be improved and the robustness of generation and deployment will be ensured.

 

*B. A THREE-TIER ENERGY BIG DATA ANALYTICS ARCHITECTURE*

A widely accepted analytical framework of big data is shown, which has three layers including data access and computing, data privacy, domain knowledge and big data mining algorithm. The inner core data mining platform is mainly responsible to say, data analysis and task processing are divided into multiple sub tasks and executed on a large number of computing nodes through a parallel program.

The middle layer of the structure plays an important role to connect inner layer and outer one. The data mining technology in the inner layer provides a platform for data-related work in middle layer such as the information sharing, privacy, protection, and knowledge acquisition from areas and applications, etc. In the whole process, the information sharing is not only the guarantee of each phase, but also the goal of processing and analysing with the big data in smart grid. In the outer layer of the architecture, pre-processing is necessary for the heterogeneous, uncertain, incomplete, and multi-source data through data fusion technology. After pre-processing, complex and dynamic data will be excavated, and then, pervasive smart grid global knowledge can be obtained through local learning and model fusion. Finally, model and its parameters need adjustment according to the feedback.

C. AN INTEGRATED BIG DATA ARCHITECTURE IN THE SMART GRID

Based on big data analytics and cloud technology, an integrated architecture can be used in the smart grid in many ways, such as optimizing power transmission, controlling power consumption, keeping the balance between power demand and supply, etc This architecture takes advantages of the three technologies including big data analytics, smart grid and cloud computing, which composes an enhanced version of smart grid. The improved smart grid has the following functionalities:

- _Analyse the historic data and estimate the energy production.
- _ Analyse the consumer behaviour patterns about electricity consumption to estimate the demand in advance.
- _ Keep record of the energy production from different sources and make decision to switch demands between the high/low priorities.
- _ Balance the load of the demand/supply chain.
- _ Do efficiently on the storage/transfer of the generated power.

## II. SMART GRID

**S**MART grid is a term referring to the next generation power grid in which the electricity distribution and management is upgraded by incorporating advanced two-way communications and pervasive computing capabilities for improved control, efficiency, reliability and safety. A smart grid delivers electricity between suppliers and consumers using two-way digital technologies. It controls intelligent appliances at consumers' home or building to save energy, reduces cost and increase reliability, efficiency and transparency. A smart grid is expected to be a modernization of the legacy electricity network. It provides monitoring, protecting and optimizing automatically to operation of the interconnected elements. It covers from traditional central generator and/or emerging renewal distributed generator through transmission network and distribution system to industrial consumer and/or home users with their thermostats, electric vehicles, and intelligent appliances. A smart grid is characterized by the bidirectional connection of electricity and information flows to create an automated, widely distributed delivery network. It incorporates the legacy electricity grid the benefits of modern communications to deliver real-time information and enable the near-instantaneous balance of supply and demand management.

Many technologies to be adopted by smart grid have already been used in other industrial applications, such as sensor networks in manufacturing and wireless networks in the below telecommunications, and are being adapted for use in new intelligent and interconnected paradigm. In general, smart grid communication technologies can be grouped into five key areas: advanced components, sensing and measurement, improved interfaces and decision support, standards and groups, and integrated communications.

## IV. BIG DATA ARCHITECTURE FOR SMART GRIDS AND ITS ADVANTAGES

A.BIG DATA ARCHITECTURE FOR SMART GRIDS

It consists of two independent procedures to connect smart grids and big data—big data modelling as an engineering procedure, following by the big data analysis as a mathematical procedure. During the engineering procedure, the raw data source **is** acquired during the mathematical procedure, the following steps are conducted.

1) real-time analysis: focusing on the real time data window whose last edge of the sampling time area is current time [i.e., **A** Time(end) = *t*0]; and 2) block-calculation for decoupling interconnected system: focusing on a smaller window consisting of data only in designated dimensions, but not in all dimensions.

*B. Advantages in Data Processing*

*1) Algorithm:* The procedure of traditional data processing algorithms, in most cases, relies highly on specific simplifications and assumptions to build models. Taking genetic algorithm for an example, two steps are essential to achieve the result. One is to trans code the engineering variables to gene as the input of the gene model. It is a subjective selection procedure for the specific roles in engineering system, and only a few variables can be taken into account in final model. The other one is to perform the genetic algorithm through operations of selection, crossover, and mutation. Many problems, such as improper settings of the population size, of the crossover probabilities, or of the mutation probabilities, will inevitably make the result worse. Compared to traditional algorithms, big data analysis, driven by data, enables us to analyse the interrelation and interaction among all the elements in system, seen as correlations indicated by high-dimensional parameters. Using a pure mathematical procedure without physical models and hypotheses, big data analysis is easier in logic and faster in speed.

*2) Distributed Calculation for Interconnected Grids:* Smart grids operation is featured with autonomous sources and decentralized controls. Due to the potential transmission cost and privacy concerns, aggregating distributed data sources to a centralized site for mining is systematically prohibitive. On the other hand, although we are able to carry out model based mining at each distributed site, the decoupling procedure for connected sources is highly related to simplifications and assumptions. Accordingly the result is often barely satisfied and leads to biased views and decisions.

Large random matrices provide a natural and universal data driven solution. For a specific zone-dividing interconnected system, each site is able to form a small matrix only with its own data. In this way, the integrated matrix can be divided into blocks for distributed calculation and vice versa. For the overall system, we can conduct high-dimensional analysis by integrating the regional matrices, or even by processing a few regional high-dimensional parameters

*C. Advantages in Management Mode*

Some brief but novel introductions and analyses about the development of power grids, mainly under the perspective of data managing mode and information communication technology (ICT), are given as the related background for applying big data into smart grids. Meanwhile, new situations and challenges, and advanced management mode for future grids are further discussed from the perspective. Especially, it is group-work mode, in our opinion, that breaks through the regional limitation for energy flows and data flows. As a result, some data-driven functions, e.g., comparative analysis, and distributed calculation, are able to be carried out. Generally, the power grids evolutions are summarized as three generations—G1–G3 .Meanwhile, their data flows and energy flows, as well as corresponding data management systems and work modes, are quite different, respectively.

*1) G1 (Small-Scale Isolated Grids):* G1 was developed from around 1900 to 1950, featured by small-scale isolated grids. For G1, components interchange energy and data within the isolated grid to keep stable. The components are fully controlled by decentralized control system and operating under individual-work mode. It means that each apparatus collects designated data, and makes corresponding decisions only with its own application. The individual-work mode works with an easy logic and little information communication. Whereas, it means few advanced functions and inefficient utilization for resources. It is only suitable for small grids.

*2) G2 (Large-Scale Interconnected Grids):* G2 was developed from about 1960 to 2000, featured by zone-dividing large-scale interconnected grids. For G2, utilities interchange energy and data within the adjacent ones. The components are dispatched by control center and operating under team-work mode. The regional team leaders, likes local dispatching centres, substations, or micro grid control centres, aggregate their own team-members (i.e., components in the region) into a standard black-box model. These standard models will be further aggregated by the global control center for the control or prediction purposes. The two aggregations above are achieved by four steps, which are data monitoring, data pre-processing, data storage, and data processing, respectively. However, lots of engineering technologies or sciences are essential as the foundations—cognitive radio wireless network , specific communication service mapping as ICT ,cloud storage, parallel computing as computer science, and modelling building, parameter identification as mathematical modeling .

*3) G3 (Smart Grids):* The development of G3 was launched at the beginning of the 21st century; and for China, it expected to be completed around 2050 [27].clear-cut partitioning is no longer suitable for G3, as well as the team-work mode which is based on the regional leader. For G3, the control force of the regional center is greatly released by individual units. The high-performance and self-control individuals results in much more flexible flows, for both energy exchange and data communication, to improve utilization by sharing resources among the whole grid Accordingly, the group work mode is proposed. Under this mode, the individuals play a dominant part in the system under the authority of the global control centers Virtual power plants and multi-micro grids for instance, are typically G3 utilities. These group-work mode utilities provide a relaxed environment to benefit both the individuals and the grids: the former, driven by their own interests and characteristics, are able to create or join a relatively free group to benefit mutually from sharing their respective superior resources; meanwhile, these utilities are generally big and controllable enough to be good customers or managers to the latter.

## V.APPLICATIONS

We take a brief overview on three application directions of big data: renewable energy, demand response and EVs. Each direction contains numerous technical fields there are no clear boundaries between them. Demand response should be considered in both renewable energy and EVs while EVs is also a concrete implementation of renewable energy.

A.RENEWABLE ENERGY

The new and renewable energy are integrated in the electric power generation, which is different from the traditional power generation mode. This difference causes the measurement and management of generated data from electric power become increasingly difficult and complex. Since the big data technology helps to make better prediction, management and processing complex big data in the energy domain, it has become increasingly popular and widely applied in renewable energy companies once emerging. On marine energy. The marine energy has attracted investments which made marine energy technologies become cost-competitive. Wool *et.al.* looked at the tribological design of three green marine energy systems: tidal, offshore wind, and wave machines. They pointed out those most marine renewable energy conversion systems need tribological components to promote rotational motion of wind or tidal streams to generate electricity. Therefore, they studied and highlighted current research thrusts to address present issues related to the tribology of marine energy conversion technologies. As for another important renewable energy sources, wind energy has also attracted wide attention. For instance, in order to improve the efficiency of electric power generation, the Vestas Wind Technology Group of Denmark optimized the wind turbine geographical placement by using big data technologies to analyse the data information including weather reports, tidal conditions, satellite images and geographical information . Besides, they also updated the underlying architecture for better data collection and used big data processing technologies to meet their high-performance computing needs. Furthermore, Kaldellis proposed an optimum autonomous wind power system sizing for remote consumers by using long-term wind speed data. Aiming to improve the life quality level of remote consumers, they used the proposed autonomous wind power systems to address the electricity demand requirements, especially in high wind potential locations. Since the historical electricity data is recorded in continuous time series, Continuous Markov Chain can be applied to analyse energy big data in order to help power enterprises make optimal investment of renewable energy and evaluate optimal energy configuration.

### B. ENERGY DEMAND RESPONSE

The demand of electric power consumers has been expanding along with the increasing use of machines and emerging types of appliances such as EVs and smart home furniture. Thus, the concern on the stability and reliability of electricity supply has also been aroused. However, since the traditional power grid lacks real-time response between demand and supply, it cannot meet these demands due to its inflexible distribution. Therefore, demand response is expected to be an aspect for the future smart grid.

### C. ELECTRIC VEHICLES

Electric Vehicles becomes a promising alternative transportation method which is related to the smart grid domain. With the increasing number of EVs, many problems in various areas of EVs such as performance evaluation, driving range and battery capacity become great concerns by researchers. Besides, EV is one of the typical application examples of big data. Therefore, most of the issues above can be handled with big data technologies. For instance, Wu *et al.* launched the EV grid integration study based on driving on a patterns analysis. Recently, most researches are focused on analyzing the energy efficiency, addressing routing problems by recording driving data, estimating the driving range. Su and Chow proposed an Estimation of Distribution Algorithm (EDA) based on big data theories to allocate electric energy intelligently to EVs. The simultaneous connection of a large number of EVs to the power grid may have a influence on quality and stability of the overall power system.

## VI. CHALLENGES

Coming from wide various sources, the volume of energy big data is increasing at an exponential speed. At the same time, Difficulties also arise up in data storage, mining, querying, processing, etc. Therefore, cryptography technologies, fuzzy data computing, qualified data processing are all essential for big data applied better in smart grid.

### A. DATA UNCERTAINTY

Data uncertainty comes from data complexity. Because of the complicated and distributed smart grid, the sources and forms of energy big data become diverse. There is no need or unrealistic to obtain accurate data samples or give a precise response. Therefore, uncertain data mining and imprecise data querying are presented for data uncertainty.

#### 1) UNCERTAIN DATA MINING

Previous applications in smart grid using traditional data mining techniques require the data samples be accurate
or with definite values. However, in the latest decades, there have been some researchers conducted to capture the uncertain factors of the data, which are common in the big data applications. These techniques often assume a distribution of the data values based on a probabilistic method, and then generate the probabilistic results of data mining. For example, the K-means clustering algorithm was optimized to the UK-means clustering algorithm through specifying a data object from an uncertain region with an uncertain probabilistic distribution method. In the UK-means clustering algorithm, the precise error and the cluster number of a data object are less important. Tsang *et al.* developed a decision tree classifier used for uncertain data and proposed the related improved algorithms. Extension of classical decision tree in building algorithms was proposed in to handle data tuples without certain values.

#### 2) IMPRECISE DATA QUERYING

In the traditional data querying techniques used in smart grid, the precise results are required in the Data Base Management System (DBMS) according to the query conditions. In recent years, a lot of querying and indexing techniques began to consider about the uncertainty of the data. Instead of giving the precise response, these techniques estimate data uncertainty and provide probabilistic guarantee. For instance, the uncertainty of the data has been modelled as the stochastic value within certain limits, and the data queries have been classified into different classes. Different algorithms are then developed to compute the probabilistic answers. Chen and Cheng studied location data with uncertainty and classified the query issues into two types, according to the degree of data uncertainty.

### B. DATA QUALITY

Due to the multiple origins of big data in the complicated power system, the smart grid database always contains incomplete, inconsistent and uncorrected data. Therefore, a series of data pre processing technologies are necessary to improve the data quality, especially after uncertain data mining and imprecise querying .Data pre processing contains data manipulation, compression and normalization of the older data into improved format. Data manipulation refers to transformation, which is used to reduce the impact from data fault sensing and provide more reliable energy data. The compression operation is used for massive data measurements to be reduced into a reasonable extent. Meanwhile, the normalization technology is used to regulate the collected data by minimizing its noise, inconsistencies and incompleteness of data. These three operations are main components of data pre-processing but the order is not unique. Furthermore, different regulations on big data cause conflicts between collaborators and lead to inconvenience in smart grid. Therefore, a unified and complete system standard needs to be set up in the future.

### C. QUANTUM CRYPTOGRAPHY FOR DATA SECURITY

Since the large amount of data is stored in multiple distributed grids, it is difficult to monitor the security of data storage in real time. However, most data involve personal privacy, commercial secrets, financial information, etc. It causes the problem of data larceny, which has become increasingly serious. Therefore, some kinds of cryptography technologies have been proposed by researches to solute security problems of big data in energy domain. Modern cryptography technique is based on the Rack Scale Architecture (RSA), which is especially used for the hardly factoring large data. In 1994, Peter Shor discovered that quantum computers can handle discrete logarithm operation and this would lead to the loss of the efficiency of the RSA architecture. Since the quantum computers have been put into use recently, a new data cryptograph technology called quantum cryptography is considered as the next generation cryptography solution. The quantum cryptography can absolutely guarantee the security of smart grid data based on the quantum mechanics. In this technology, the data is encoded as qubits instead of transmitting bits and different directions will affect the qubit's polarization. The quantum key distribution begins when a large amount of qubits are sent from the sender to the receiver. If an eavesdropper wants to get the information about any qubit, he will have to measure it. However, having no idea about the qubit's polarization, the eavesdropper can only process it by guess and then transmit another randomly polarized qubit to the receiver. All these operations are under the surveillance. At the end of quantum key distribution, the key will be identified safe or not by a brief testing.

## CONCLUSION

In this paper, we gave an introduction on big data and power grid, presented the background in the aspects of related work, techniques and tools for energy big data. Then, we discussed the important role of big data, which brought efficiency and accuracy to the power system. Furthermore, an integrated architecture was introduced for the big data analysis. We also discussed the applications of energy big data in branches of renewable energy, energy demand response, and electric vehicles. Finally, we listed some challenges need to be addressed about data uncertainty, quality and security.

## REFERENCES

[1] C. W. Gellings, M. Samotyj, and B. Howe, ``The future's smart delivery system [electric power supply]," *IEEE Power Energy Mag.*, vol. 2, no. 5, pp. 40_48, Sep./Oct. 2004.
[2] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, ``A survey on smart grid communication infrastructures: Motivations, requirements and challenges," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 5_20, 1st Quart., 2013.
[3] K.Wang, Y. Shao, L. Shu, Y. Zhang, and C. Zhu, ``Mobile big data faulttolerant processing for eHealth networks," *IEEE Netw.*, vol. 30, no. 1, pp. 1_7, Jan. 2016.
[4] B. Gu, X. Sun, and V. S. Sheng, ``Structural minimax probability machine," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, Apr. 2016, doi: 10.1109/TNNLS.2016.2544779.
[5] F. Rahimi and A. Ipakchi, ``Demand response as a market resource under the smart grid paradigm," *IEEE Trans. Smart Grid*, vol. 1, no. 1,pp. 82_88, Jun. 2010.
[6] S. M. Amin and B. F. Wollenberg, ``toward a smart grid: Power deliveryfor the 21st century," *IEEE Power Energy Mag.*, vol. 3, no. 5, pp. 34_41, Sep. 2005.
[7] T. Niimura, M. Dhaliwal, and K. Ozawa, ``Fuzzy regression models to represent electricity market data in deregulated power industry," in *Proc. IFSA World Congr. NAFIPS Int. Conf.*, vol. 5. Jul. 2001, pp. 2556_2561.

 