VIVA Institute of Technology
10th National Conference on Role of Engineers in Nation Building – 2022 (NCRENB-2022)

VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

Study of Latest Trends in Big Data Management of Smart Grids

Ritesh Chavan
Assistant Professor, Department of Electrical Engineering, VIVA Institute of Technology

Abstract: *Electrical power grid plays a crucial role for sustainable development of each domain of universal resources. Modernisation of the grid enables communication with help of latest technologies and delivers electricity with high resiliency, reliability and efficiency. The smart grid is modernisation technique to provide smart features such as monitor and control of grid-connected systems activities, consumer centric electricity delivery, and real-time data collection. Big data analytics is latest technique endorsed by researchers for establishment of strong smart grid using analysis of association between big data, cloud services and smart grid. The big data analytics includes major features of data management, data analysis, data processing and data realisation. This article put forward latest trends in domain of big data analytics and their applications in smart grid with help of relevant examples.*

Keywords – *Big data, Data management, Grid modernization, Security, Smart grid*

I. INTRODUCTION

Smart grids connect a large number of customers, and an even larger number of devices. As a larger percentage of devices become "smart", the amount of data to be collected by the utility increases dramatically. Evolving technologies in the energy and utilities industry can provide companies with unprecedented capabilities. At the same time, these advances also generate unprecedented data volume, speed and complexity. The earliest reference of Big Data can be traced back to Nutch which is an open source project launched by Apache Software Foundation. According to reference, big data has characteristics, which are volume, Variety and Velocity because of which the traditional data processing techniques are not always applicable for big data application. Electric power big data can be found in all sections of electric power generation and management. The key technology allows electric power to overcome the challenges of limited resources and harsh environment. Moreover, having identified those benefits, it should minimize the costs of infrastructure needed to obtain and process the data necessary to deliver these benefits. In this paper, author have enlightened about latest trends in data management domain and its application in smart grid sector.

II. IMPORTANT FEATURES OF DATA ANALYTICS IN SMART GRID

2.1 Data Analysis in Smart Grid

The smart grid involves all the operations of a system such as generation, transmission and distribution. Millions of sensors include phasor measurement units (PMUs) and smart meters. These sensors can generate a huge amount of data in real time. The smart grid combines all the power generation resources such as solar energy, wind energy and nuclear power. The electric network control in turn involves the sensors, smart meters and control strategies. The data in smart grid could be classified into three categories, smart grid operating and monitoring data, business marketing data and enterprise management data. All these data are structured and unstructured. Unstructured data including video monitoring and graph/image processing are not convenient to store using two-dimensional logical table. For example, power grid dispatching and controlling need real time data. A smart grid is a modernised electrical grid that combines digital information technology, communication technology, computer technology, and power infrastructure. Smart grids provide a number of benefits, including increased energy economy, as well as increased power supply safety and dependability.

Utilities are putting more emphasis on business intelligence and advanced analytics to assist data-driven decision making and planning in light of the changing data environment. These analytics necessitate a unified picture of company data as well as data alignment across various operational units and lines of business. Utilities that integrate and analyse this data can obtain insight into their operations and assets, allowing them to take proactive rather than reactive action. Higher profitability, decreased carbon footprint, increased safety, improved regulatory engagement, and improved customer satisfaction are all possible outcomes.

2.2 Cloud Computing

Cloud Computing, which is now one of the most in-demand technologies is giving every business a new form by delivering on-demand virtualized services/resources. Every firm, from small to medium to large, uses cloud-computing services to store information and access it from anywhere and at any time using just the internet. We will learn more about cloud computing's underlying architecture in this essay.. The cloud-computing is partitioned into two viz. frontend and backend as shown in fig. 1 below.

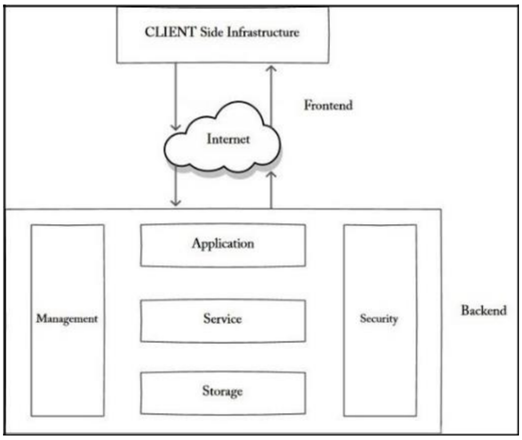

Fig. 1: Cloud Computing Architecture

2.2.1 Frontend

The client side of a cloud computing system is referred to as the frontend of the cloud architecture. It contains all of the client's user interfaces and apps for accessing cloud computing services and resources. For instance, to access the cloud platform, you may utilise a web browser. The client interacts with the front end. It includes client-side interfaces and applications for interacting with cloud computing services. The internet web browsers such as Google Chrome, Firefox, Microsoft Edge and Internet Explorer, every type of clients and users, gadgets like laptops, tablets, and mobile devices all make up the front end.

2.2.2 Backend

The cloud employed by the service provider is referred to as the backend. It maintains the resources and offers security procedures in addition to containing them. It also incorporates large amounts of storage, virtual applications, virtual machines, traffic management systems, and deployment methodologies, among other things. The service provider uses the back end. It oversees all of the resources needed to deliver cloud-computing services. It comprises a massive quantity of data storage, as well as security measures, virtual machines, deployment models, servers, and traffic management systems, among other things.

In the backend, an application is a piece of software or a platform that a client may use. That is, it provides the service in the backend in accordance with the client's needs. Backend services relate to the three main categories of cloud-based services: *Software as a service, Platform as a service*, and *Infrastructure as a service*. It also controls which services the user can have access to. In the backend, runtime cloud refers to the virtual machine's execution and runtime platform/environment.

2.3 Data integration and management techniques

The electric power big data integration and management technologies, which have the potential to provide new application functions for businesses, integrate data from two or more information systems. To put it another way, the technologies combine data from several sources, in various forms, and with various characteristics, into logical storage media. Data fusion and data integration, database management technology, and Extract-Transform-Load (ETL) technology are among the electric power big data integration and management technologies as shown in fig. 2. Data integration problems have always existed, which is why scientists at the University of Minnesota built the first data integration system in 1991. The ETL strategy was employed in this data integration technology, which extracts, processes, and loads data from several sources into uniform perspective.

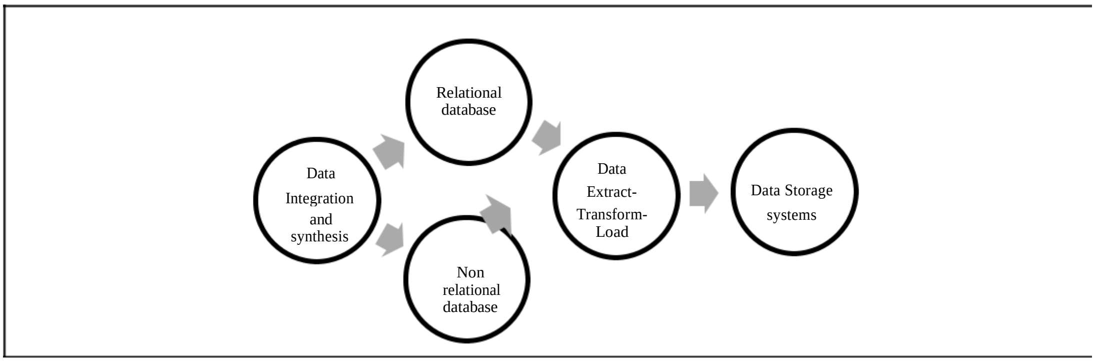

Fig. 2 Data integration and management techniques

Data integration would be useful in this situation to help aggregate data from several departmental databases and use it for reporting and analysis. Building an enterprise data warehouse, transferring data from one or several databases to another, and syncing data between applications are all examples of mission-critical data management initiatives that include data integration. As a result, organisations utilise a wide range of data integration tools, technologies, and methodologies to bring data from diverse sources together and generate a single version of the truth.

2.4 Data Analysis Techniques

The volume of data, the speed with which it is handled, and the diversity of data are all characteristics of big data. Data analytics has moved into the technological domains of artificial intelligence and machine learning as a result of the second characteristic, velocity. Conventional statistical procedures are used in addition to the ever-evolving computer-based analytical capabilities that data harnesses. In the end, how data analysis techniques work inside an organisation is twofold: big data analysis is handled through the streaming of data as it arises, followed by batch analysis of data as it accumulates – to look for behavioural patterns and trends. As the amount of data generated grows, so will the number of data management approaches. Data promotes innovation, as it gets more insightful in terms of speed, size, and depth.

The A/B data analysis approach compares a control group against a range of test groups to see which treatments or adjustments will enhance a particular objective variable. Analysis of what content, text, graphics, or style can boost conversion rates on an e-commerce site is an example given by McKinsey. Big data fits into this paradigm once again since it can test large numbers; however, this can only be done if the groups are large enough to acquire noticeable differences. The insights are more efficient and perhaps more accurate than if they were created from a single source of data by utilising a collection of approaches that analyse and integrate data from numerous sources and solutions. Within database administration, data mining is a typical approach used in big data analytics to uncover patterns from massive data sets by integrating methods from statistics and machine learning.

2.5 Data Processing Techniques

Data processing is a technique for manipulating information. It refers to the transformation of unstructured data into material that is both meaningful and machine-readable. It is essentially a method of transforming unstructured data into useful information as shown in fig.3 . It can relate to the processing of business data using automated methods. To analyse massive numbers of comparable data, this usually involves very basic, repetitive processes. Raw data is the source of information that is processed to provide useful results. Distributed computing, memory computing, and stream processing are some of the data processing techniques used in electric power big data.

Commercial data processing is a way of using relational databases in a commercial setting, which involves batch processing. It entails feeding the system a vast amount of data and producing a significant volume of output with fewer processing processes. It essentially integrates commerce and computers in order to make it helpful for a company. Because the data handled by this system is typically standardised, there is a significantly smaller risk of mistakes. Unlike corporate data processing, scientific data processing employs a large number of computing operations while needing minimal deliverables. Computing procedures include arithmetic and comparison activities. In this type of processing, any chance of errors is unacceptable even though it would contribute to false choices. As a result, the data is verified, categorised, and standardised with extra attention, and a multitude of scientific techniques are used to ensure that no inaccurate connections or conclusions are reached.

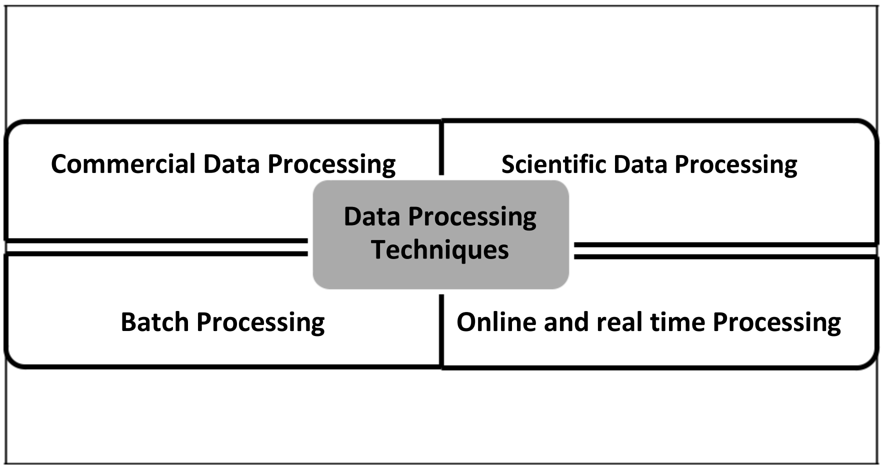

Fig. 3 Types of data processing techniques

Data processing in a corporate context takes much longer. Scientific data processing includes things like processing, organising, and disseminating scientific datasets, as well as allowing scientific study of algorithm and calibration data.

2.6 Data Visualisation Techniques

Data visualisation is among the most significant abilities in modern machine learning and data mining. Data visualisation is one of the most significant instruments for determining a qualitative understanding. This might be useful when trying to examine a dataset and extract information about it, as well as spotting trends, destruction of data, outliers, and other things. Data visualisations may be used to convey and highlight serious correlations in graphs and plots that are more useful to yourself and clients than measurements of association or importance if we have some domain expertise. A graphical depiction of information and data is referred to as data visualisation as shown in fig. 4. Data conceptual models make it easy to identify and comprehend patterns, anomalies, and characteristics of the data by employing visual components like infographics, dashboards, and maps.

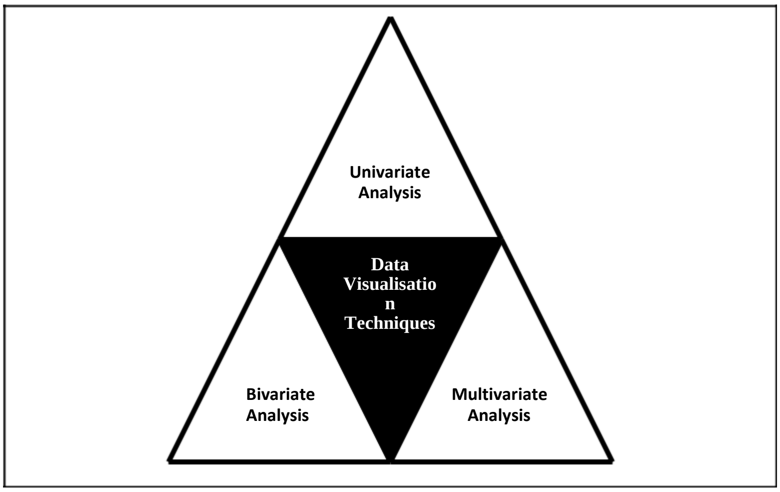

Fig. 4 Types of data visualization techniques

Our society is pictorial, encompassing everything from artwork and advertisements to films and television, even though our eyes can record patterns and textures. As a result, we can immediately distinguish the red area from the blue, the square from the circular. As a result, data visualization is another form of visual communication that entices our curiosity and keeps our attention on the information sent via the eyes. We immediately spot the patterns and outliers in the dataset whenever we show a graphic. It is an effective method for examining data and producing presentable and understandable findings.

III. Data Management Model For Smard Grid Application

3.1 Requirement of data management for Smart Grid applications

Smart grid systems create humongous data; for example, a SCADA dataset obtained every 2–5 seconds, whereas an AMI system gathers data every 1–15 minutes, and so on. As a result, utilities confront a slew of issues in data management, ranging from strategy to execution. The electric power big data integration and management technologies, which have the potential to provide new application functions for businesses, integrate data from multiple data systems. Millions of equipment smart sensors metres must be accessed, with real-time data sampled and transmitted on a regular basis, along with historic data gathered out over many years that had to be in larger quantities.

A smart grid is a diverse and complex ecosystem that includes a variety of devices, networking, services, and data. Networks with lower or higher latency, devices either with power limits, active or passive systems, uninterrupted or non-continuous data, and so on are examples. In terms of transmission limits, faults, limited resources, and high scalability, smart grids encounter a variety of requirements and obstacles when it comes to data integration. If protocols are not supplied, providers will have to deal with a variety of interfaces, each with its own set of definitions and communication procedures, making integration difficult.

3.2 Data management model for Smart Grid

Energy providers can use Big Data technology to develop novel technique, functional module, and solutions, as well as better data handling in smart grids. Big Data may be regarded as a large number of datasets, but it also has additional characteristics. Moreover to volume, Big Data study relied on variety, which presents multiple data modes (organised, moderately organised, or unorganised), velocity, which provides timeliness requirements, value, which provides the ability to extract significance from the data extracted, variability, which provides the data inconsistency concept, and factuality, which works on the system's reliability. The several data classes that were utilised to obtain the trustworthy datasets. Functional data refers to the grid's power data, which includes actual and reactive flow of power, load management capability, voltage, and so on. Non-operational information pertaining to primary data, information on power reliability, and so on. It has nothing to do with grid power. Meter utilization data refers to information about electricity energy needs, such as median, maximum, and schedule of day. The data in event messages originates from smart grid equipment events such as voltage failure, fault detection, and so on. Finally, Metadata is a type of data that is being used to manage and summarize all other types of data. All of this information is gathered from a variety of sources, including meter, detectors, gadgets, and power stations.

3.2.1 Data integration techniques for Smart grid

All corporate systems employ *Service Oriented Architecture* (SOA), which combines a large number of software components, each of which has its own method of delivering value to customers. The challenge is determining how to set up and maintain all of these systems. Like a strategy, SOA allows software to communicate with one another in a unified manner, making data unification simpler and more treatable.

The *Enterprise Service Bus* (ESB) is built on a variety of ways for managing communication across many types of systems, such as Geographical Information System, Outage Management System, Billing and Customer Information System. In terms of administration, monitoring, and integration deviation, the ESB provides several benefits that minimize cost and time.

Common Information Models (CIM) are pivotal to the effectiveness or loss of data management since they are employed for smart grid resilience and incorporated data structure. Unified Modeling Language frameworks for the electricity market are referred to as CIM. In terms of data unification, availability, and expenditure, it is critical in energy administration platforms.

3.2.2 Data storage techniques for Smart grid

Because the smart grid is centred on gathering data from various sources and providing it to advanced monitoring in quick feedback operations per second, data storage is essential. To address Big Data needs, a sophisticated and adaptable data storage mechanism is required.

The *Distributed File System* (DFS) is a file and storage sharing system that allows numerous users on various computing devices to exchange files and resources. Its storage technique is customer centric and it allows each user to acquire a backup document of the log records. DFS is used by a variety of technologies, including Google Files Systems, QF System, Hadoop Distributed File System and others.

In the event of big data, *not only SQL databases* are a novel database technique that overcomes the limits of standard relational SQL databases. There are three types of architectures for this type of database includes key-value solutions, column oriented services and document storage services.

IV. INTRODUCTION OF BIG DATA ANALYTICS IN SMART GRID

4.1 Integration of renewable power sources in Smart Grid

The incorporation of fresh and sustainable sources into the electric power generating process alters the traditional power supply methodology. Electric power generating data assessment and evaluation has grown increasingly difficult. Big Data technology can help electricity generation and grid businesses can make better estimations. By analyzing big data such as climate information, wave statistics, geospatial data, and radar data, the power players optimizes the topographical structure of wind turbines and increases the productivity of power generation. A well know intelligent system production organization proposed a stutter stepping approach for turbine park position choosing focused on slightly elevated weather forecasts that enhances wind resource efficiency while lowering costs. Though renewable technologies can match market demand and provide unlimited energy, their inconsistency hampers grid management and presents a problem to both energy organizations and customers.

4.2 Conditioning and monitoring of renewables using big data

Wind turbine condition monitoring is an important aspect of operations and maintenance, since operations encompass the administration, surveillance, and elevated onsite controlling of turbine green stage, whereas maintenance contains the treatments essential to maintain the facility in good working order. Preventive servicing replaces components at the succeeding action, surely until a pertaining failure takes place; a planned servicing strategy that focuses on condition monitoring can forewarn maintenance regarding elements those are prone to failure and have them replaced in due time. Corrective maintenance does not use condition monitoring and replaces individual components when issues are found or accumulate.

4.3 Consumer data analysis using big data

For electricity firms, big data studies of consumer information has become a need rather than a privilege. Smart metres provide consumers more command over their own demand, allowing them to participate in microgrid as end customers. Utilities employ Demand Response initiatives to acquire real supply demand data at various locations of expenditure in order to normalize and make a prediction correctly. As a result, the production curve may be more effectively managed in response to demand, reducing overcapacity losses. This will also allow for real-time diagnostics of metres and hardware near to the consumer, as well as the issuing of notifications and the execution of self control systems. One of the goals of demand response is to improve customer involvement by allowing companies to connect with consumers' power requirements even when there is a blackout.

V. CONCLUSION

Smart grid technology manage to accumulate number of information in order to make the grid smarter. At the same time, organizations have issues in dealing with the structure, accessibility, and authentic constraints of the data acquired. It is a requirement to get an assessment of such prospects, principles, and problems of data administration in smart grids throughout this article, as well as a summary of Big Data methodologies that may be utilised to meet smart grid needs such as analysis, retention, and even presentation. In addition to provide an effective and flexible advanced analytics, it is also needed to understand the procedures, resources, and operational specifications for establishing and implementing Big Data analysis for smart grid.

REFERENCES

[1] J. Zhan, J. Huang, L. Niu, X. Peng, D. Deng and S. Cheng, "Study of the key technologies of electric power big data and its a pplication prospects in smart grid," 2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2014, pp. 1-4.

[2]X. Daqing and H. Yinghua, "An Adaptive Data Management Model for Smart Grid," 2015 8th International Conference on Intelligen t Computation Technology and Automation (ICICTA), 2015, pp. 126-129.

[3]J. Lee, J. Han and X. Li, "A Unifying Framework of Mining Trajectory Patterns of Various Temporal Tightness," in IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 6, pp. 1478-1490, 1 June 2015.

[4]A. El Khaouat and L. Benhlima, "Big data based management for smart grids," 2016 International Re newable and Sustainable Energy Conference (IRSEC), 2016, pp. 1044-1047.

[5]Daki, H., El Hannani, A., Aqqal, A. et al. Big Data management in smart grid: concepts, requirements and implementation. J Big Data 4, 13 (2017).