SMART METER AND SURVEY OF CHALLENGES AND SOLUTIONS FOR BIG DATA MINING

Prof. Priyanka R. Tarmale
Lecturer, Electrical Power System Department, V.P.M‟s Polytechnic, Thane ,Maharashtra, India

Abstract: Smart grid is bidirectional flow of electricity as well as information which automatically deliver the information with its power grid network. Smart grid uses advanced sensors, computing and communication hardware and software for better delivery which also improve performance of grid. It is combination of electrical infrastructure and information technology. For such as smart grid system, smart meters are required which provide information to the service provider as well as customers, so that it can help them for controlling power consumption which indirectly balances the grid system. To capture and transfer data related to consumption, is the key function of smart meter, which generates large amount of data. Various issues and challenges rises regarding to security and privacy concerns and about speed for data collection and complexity, which result into Big data challenge. The paper presents the literature review of smart grid and smart infrastructure and also security approach in Big data managing is discuss in this paper.

I. INTRODUCTION

In contrast to Traditional Grid which is consisting of mainly four operations i.e electricity generation, transmission distribution and electricity control, Smart Grid uses modern information technologies and is more capable of delivering power in effective manner. Table 1.1 shown below can give comparison between Traditional Grid and Smart Grid. [1]Use of renewable energy resources is increase with the increase of distributed generation and distributed storage, due to change in climate, awareness of energy efficiency, new need of electricity markets.

EXISTING GRID	SMART GRID
Electromechanical	Digital
One-way communication	Two-way communication
Centralized generation	Distributed generation
Few sensors	Sensors throughout
Manual monitoring	Self-monitoring
Manual restoration	Self-healing
Limited control	Pervasive control
Few customer choices	Many customer choices

Table 1.1

The most desirable objective by smart power system is to control the load through efficient networking techniques and also to balance the power system in all situations. With the progressing technologies a new system is introduced i.e Smart Meter which is based on different communication methods. Smart meter capture‟s the data about power consumption and large data regarding power quality and status of meters. [5]The main challenges of smart grids are how to manage different types of front-end intelligent devices such as power assets and smart meters efficiently and how to process a huge amount of data received from these devices. In recent years the sizes of databases are rapidly increasing which made a need to develop tools capable in the automatic extraction of knowledge from data. Database is an organized collection of data which is uploaded as per needed and data mining is extraction requires data stored in databases. Big data is defined as datasets that are of large size and have greater complexity which consist of both structured and unstructured data in it.

II. PROBLEM STATEMENT

Some developers still make use of conventional data storage and processing techniques (such as relational databases) to store and harness information from large data sets which proves to be inefficient these days. Hence, there is a need for them to adopt a new method of developing applications that can process information in a very efficient and fault-tolerant way.

III. EVOLUTION OF ELECTRICITY METERING

A Smart Metering system is based on four main pillars:

- A Smart Metering device, Smart Meter (SM);
- A data gathering device, Data Concentrator (DC);
- A communication system used for data flow;
- A centralized management and control system, Control Center (CC).

Electricity Meters are used to measure the quality of electricity supplied to customer as well as to calculate energy and transportation charges for electricity retailers and network operators. Smart meters provide real-time display of energy use and pricing information, different tariffs. [3]As large amount of data and information will be generated from metering, sensing, and monitoring in smart grid, Task is to manage data modelling, information analysis, integration and optimization in data concentrator. The main function of the Data Concentrator is to gather metering data from the Smart Meters. Control Centre or Data Management System (DMS) is in charge of receiving and storing the metering data for processing purposes. The CC can be seen as a modular system formed by the Meter Data Management System (MDMS), which manages the metering data, and additional secondary modules in charge of end-users applications. Communication system play important role as they transmit data in terms of quality, time, and security. These system should provide good, coverage must be cost effective, proper security features.

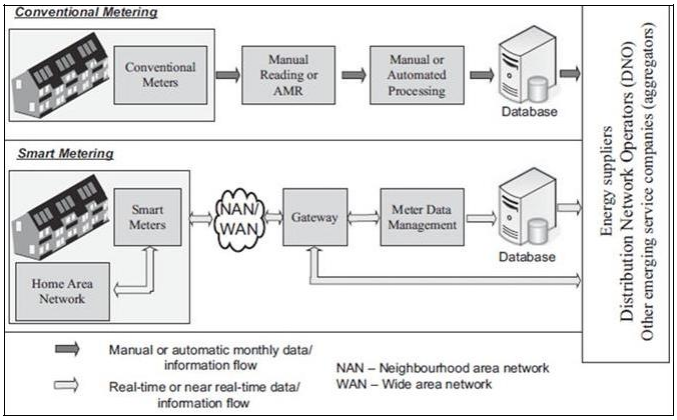

Fig 1.1 Conventional and smart metering comparison

IV. SMART METER SYSTEM ISSUES

A. Issues in Smart Meter Data Analytics

To achieve metering intelligence, number of technical issues must be solved properly. [6] The main requirement is ability to work with very large volumes of data, and also with varieties of data content which can be done with different techniques for efficient data fusion and integration in order to achieve consumer‟s acceptance and support for smart meters.

B. Smart meters and big data

Big data in smart grid is generated from various sources like power utilization by users, energy consumption data measured by smart meter, management, control and maintenance data for devices and equipment's in power generation, transmission and distribution network.

V. BIG DATA IN SMART GRID

In smart meter, big data analytics application is to convert tens of billions of data points coming from millions of smart meters which are located in different locations and turning that data into actionable information for the grid operations. It can also generate data of customer‟s consumption records. It is also possible to calculate data about unaccountable consumption when energy is being diverted or stolen. The aim is to improve grid reliability, outrage response and reducing cost of distributions operations.

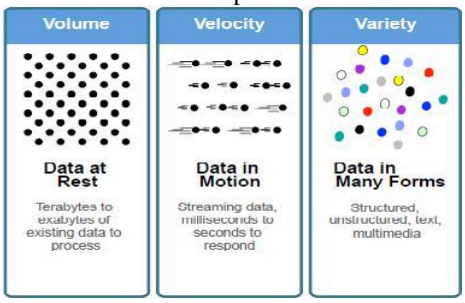

Fig 1.2 Big data characteristics

Big data refers to rapidly growing datasets with sizes beyond the capability of traditional data base tools to store, manage and analyses them. Big data is a heterogeneous collection of both structured and unstructured data. Increase of storage capacities, Increase of processing power and availability of data are the main reason for the appearance and growth of big data. Big data refers to the use of large data sets to handle the collection or reporting of data that serves businesses or other recipients in decision making. The data may be enterprise specific or general and private or public

[2]. Big data are characterized by 3 V‟s: Volume, Velocity, and Variety.

A] Security and Privacy Challenges for Big Data

Extraordinary benefits of Big Data, such as ability to collect larger datasets which are constantly increasing gives growth concerns over privacy and data protection. Information security is problem where massive amount of data will be correlated, analyzed and mined for meaningful patterns [8].

Any security control used for big data must meet the following requirements:

- It must not compromise the basic functionality of the cluster.
- It should scale in the same manner as the cluster.
- It should not compromise essential big data characteristics.
- It should address a security threat to big data environments or data stored within the cluster

Security of big data can be enhanced by using the techniques of authentication, authorization, encryption and audit trails. The following are some of the methods used for protecting big data:

- Using authentication methods
- Use file encryption
- Implementing access controls
- Use key management
- Use secure communication

VI. TECHNIQUES FOR BIG DATA

Big data has great potential to produce useful information for companies which can benefit the way they manage their problems. Due to frequently occurring different patterns of data and hidden rules, it is a big task for Big Data for automatic discovering of intelligence of data and this cannot be possible without the help of computational tools. New emerging technologies like Hadoop framework and MapReduce, offers new ways to process and transform big data which is unstructured and complex into meaning knowledge.

A. HADOOP

Hadoop is a scalable, open source, fault tolerant Virtual Grid operating system architecture for data storage and processing. It uses HDFS (Hadoop Distributed File System) which is fault-tolerant high bandwidth clustered storage architecture.[8] It runs MapReduce for distributed data processing and is works with structured and unstructured data. For handling the velocity and heterogeneity of data, tools like Hive, Pig and Mahout are used which are parts of Hadoop and HDFS framework. Hadoop and HDFS by Apache is widely used for storing and managing big data. Hadoop consists of distributed file system, data storage and analytics platforms and a layer that handles parallel computation, rate of flow (workflow) and configuration administration.

The present Hadoop ecosystem, as shown in Figure 8, consists of the Hadoop kernel, MapReduce, the Hadoop distributed file system (HDFS) and a number of related components such as Apache Hive, HBase, Oozie, Pig and Zookeeper.

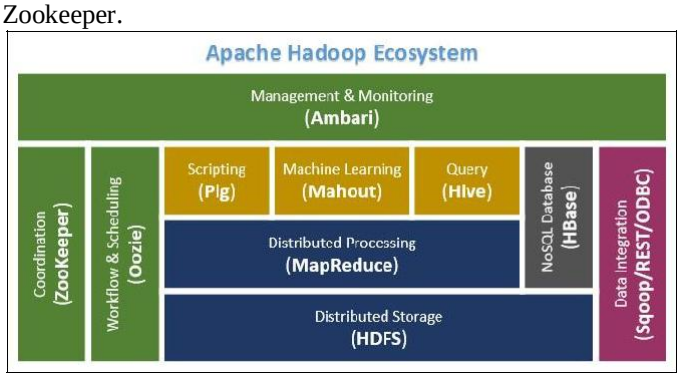

Fig1.3 Hadoop Architecture Tools

For handling the velocity and heterogeneity of data, tools like Hive, Pig and Mahout are used which are parts of Hadoop and HDFS framework. [2] It is interesting to note that for all the tools used, Hadoop over HDFS is the underlying architecture. Oozie and EMR with Flume and Zookeeper are used for handling the volume and veracity of data, which are standard Big Data management tools

i] Mode of Operation

A MapReduce job usually splits the input data set into independent blocks of data which are assigned to Map tasks (functions) in a completely parallel manner. The output of the map is then sorted and given as input to the Reduce tasks (functions) to produce the final result. Hadoop MapReduce works exclusively on <key, value> pairs. It views the input to the job as <key, value> pairs and produces its output from the job as a set of <key, value> pairs. It makes use of an interface known as the Writable interface to serialize the Key and Value classes. These classes have to implement the interface. Also, the key class has to implement the Writable Comparable interface in order to facilitate sorting by the framework. The input and output of the MapReduce jobs can be depicted as:

(Input)<k1,v1> à map à <k2,v2> à combine à <k2,v2> à reduce à <k3, v3> (output)

Applications which implement MapReduce usually implement the Mapper and Reducer Interfaces which provide several methods for different tasks.

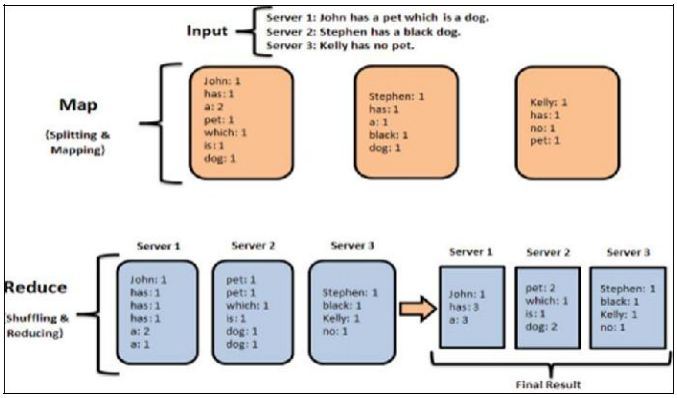

Fig 1.4 Example of MapReduce Word Count

ii] Existing Users

Hadoop MapReduce is applicable in all areas of life where data is being generated. This is because the data being generated will certainly need to be stored and processed. Due to its awesome benefits, a number of "Big" companies have already started using MapReduce. They include:

- Facebook
- Yahoo
- Amazon
- eBay
- Google
- IBM
- The New York Times
- Walmart

Amazon implements MapReduce in their Amazon Elastic MapReduce (Amazon EMR). It is a web service that makes it easy to process vast amounts of data quickly and cost-effectively. Every day, a number of users adopt Hadoop MapReduce for processing their data and join in enjoying its benefits.

iii] Benefits of Hadoop MapReduce

MapReduce provides a lot of benefits. They include:

1. Cost-efficiency
2. Simplicity
3. Scalability
4. Speed
5. Recovery
6. Minimal Data Motion.

FUTURE SCOPE

As far as the future of big data is concerned it is for certain that data volumes will continue to grow and the prime reason for that would be the drastic increment in the number of hand held devices and internet connected devices, which is expected to grow in an exponential order. SQL will remain as the standard for data analysis and Spark, which is emerging, will emerge as the complimentary tool for data analysis. Although there has been much opposition to smart meters due to privacy and health concerns, it is obvious that smart meters are here to stay and that the smart grid and smart metering will be a „way of life" in the future. A number of different dimensions to smart meters have been highlighted including the smart meter technology and the process, the various stakeholders, existing analytics technologies and tools, the current technological revolutions such as big data. The smart grid and smart meters will be part of a much wider internet of things in the future integrating multiple aspects of human needs and services to satisfy such needs, and the analytics requirements discussed such as big data, real time analytics, stream analytics will need to be built in to the processes and workflows for diagnostics in real time.

V. CONCLUSION

Due to the potential importance of SG, this paper comprehensively explores the complete survey of SG. In this it is surveyed that the major SG projects/programs and three

major technical systems in SG: the smart infrastructure system, the smart management system, and the smart protection system. We also get about challenges and future research directions worth exploring for each of these three systems. A number of different dimensions to smart meters have been highlighted including the smart meter technology and the process, the various stakeholders, existing analytics technologies and tools, the current technological revolutions such as big data, cloud computing and the internet of things. Managing that mass of data is extremely challenging- finding actionable market intelligence and using that data to rapidly respond to market developments is even more so. Given the impending exponential growth of data coming from the "Smart Grid," increased regulatory oversight of energy markets, and the lost commercial opportunities buried in all that data, traditional capture, storage, retrieval and analysis techniques will prove ineffectual for many companies in this market.

In conclusion, big data will have a large impact on the management of power utilities due to the fast deployment of ICT and intelligent sensing within the power and energy network. Smart grid operation and future energy management will be hugely data-intensive. There are many challenges which affect the success of big data applications in smart grid. Presently, experience in integrating big data with smart grid is limited. Great efforts have to be spent to develop more advanced and efficient algorithms for data analysis. More demonstration projects are needed to gain practical experience and evaluate benefits, standardize the design and deployment procedures, to drive company performance and minimize risks from new regulations and political interference under a low carbon economy and green environment.

REFERENCE

[1] Smart Grid – The New and Improved Power Grid: A Survey Xi Fang, Student member, IEEE, Satyajayant Misra, Member, IEEE, Guoliang Xue, Fellow, IEEE, and Dejun Yang, Student Member, IEEE.

[2] ISSUES, CHALLENGES, AND SOLUTIONS:BIG DATA MINING Jaseena K.U.1 and Julie M. David2 1,2 department of Computer Applications, M.E.S College, Marampally, Aluva, Cochin, India 1jaseena.mes@gmail.com,2julieeldhosem@yahoo.com.

[3] State of the Art and Trends Review of Smart Metering Electricity Grids Noelia Uribe-Pérez 1,*, Luis Hernández 2, David de la Vega 3 and Itziar Angulo 3.

[4] Model of Smart System Based On Smart Grid, Smart Meter and Wireless Based Smart Appliances Rao M.Navid-Ur-Rehman1, Zubair Ahmed1, Adnan Faroqi1 1University College Of Engineering And Technology, The Islamia University Of Bahawalpur, Punjab,Pakistan.

[5] Issues in Smart Metering Systems Obaid Ur-Rehman, Natasa Zivic, Christoph Ruland Chair for Data Communications Systems University of Siegen Hoelderlinstrasse 3, 57076 Siegen, Germany email: {obaid.ur-rehman, natasa.zivic, christoph.ruland}@uni-siege n.de.

[6] Smart Electricity Meter Data Intelligence for Future Energy Systems: A SurveyDamminda Alahakoon, Member, IEEE and Xinghuo Yu, Fellow, IEEE

[7] Security Issues in Smart Metering Systems Obaid Ur-Rehman, NatasaZivic, ChristophRuland Chair for Data Communications Systems University of Siegen Hoelderlinstrasse 3, 57076 Siegen, Germany email: {obaid.ur-rehman,natasa.zivic, christoph.ruland}@uni-siegen.de

[8] Advanced Analytics for Harnessing the Power of Smart Meter Big Data DammindaAlahakoon School of Information and Business Analytics Deakin University Australia d.alahakoon@deakin.edu.auXinghuo Yu Platform Technologies Research Institute RMIT University Australia x.yu@rmit.edu.au

