Big Data Analytics for Distribution System Monitoring in Smart Grid

Shibily Joseph* and Jasmin E A

Government Engineering College, Thrissur, Kerala, India

Abstract

Advanced Metering Infrastructure (AMI) is an integral part of Smart Grid. Utilities made huge investments in AMI to fully utilize the benefits of smart grid. Efforts must be made to utilize this infrastructure beyond billing purpose. Smart meters can act as a good sensor for distribution system monitoring. But system monitoring based on AMI requires real time computing of smart meter data. Considering the size of smart meters deployed, real time big data analytical tools and techniques are required for system monitoring based on AMI data. Stream Computing is a real time data analytical technique used in big data. In our paper system monitoring applications for outage detection and peak load tracking from AMI data is implemented using stream computing technique. This work revealed that stream computing is a promising big data analytics that can be used in real time applications of Smart Grid.

Keywords: Advanced metering infrastructure (AMI), Big data analytics, Real time Monitoring, Outage detection

1. Introduction

In the end of 20th century and beginning of 21st century witnessed a paradigm shift in manufacturing, education, service and management sector. This is mainly due to the impact of information and communication technology (ICT). Utility industries alone cannot resist the change. Utilities incorporated ICT in their generation, transmission and distribution section of their business. This gave birth to a new kind of power grid called smart grid. Actually power grid is transformed to smart grid by using ICT to gather information to improve the efficiency, reliability, economics and sustainability of the production, transmission and distribution of electricity [1]. The major components of smart grid are power stations, power transmission lines, pluggable electric vehicles (PEV) charging stations, residential sub division installed with solar panels, residential complex with advanced metering infrastructures (AMI), and energy smart houses with electric appliances connected to the smart grid. One of the major characteristics of smart grid is the seamless interaction of these components using networks. All these components generate data with high velocity, variety and finally lead to huge volume of data [2].

According to U.S Department of Energy(DoE), future smart grid utilize data from advanced measuring instruments[3]. Now utilities are massively installing communicating sensors and smart meters for the real time monitoring, controlling, management and planning of the system. The data from the smart grid are used by demand response(DR) programs, distribution grid monitoring system, outage management system, renewable energy planning systems, conservation voltage reduction (CVR) programs, strategic asset management systems, predictive analysis, billing system, load management system [4, 5]. Now utilities are facing big challenge from data generated by smart grid as this data is classified as big data. The characteristics of big data are volume, velocity, variety and value [6]. Big data analytics enable utilities to derive great value from data to support real time operations, strategic decisions and system planning.

Utilities are now tapping mater data to monitor their distribution system to improve performance indices. IEEE P1366-2003 standard defined four key reliability indices, which are System Average Interruption Duration Index (SAIDI), Consumer Average Interruption Duration Index (CAIDI), System Average Interruption Frequency Index (SAIFI) and Momentary Average Interruption Frequency Index (MAIFI)[7]. BRIDGE survey shows that utilities who makes outage and restoration management (ORM) a top priority have higher customer satisfaction index[8]. So companies are now giving more importance to OMS activities. Outages affect the normal load prediction and to minimize the effect, utilities are now giving more priority to outage management system (OMS)[9].

Loads and their characteristics are important to distribution companies. Customer loads have many common characteristics, like load levels vary through the day, peaking in the afternoon or early evening. Estimation of load characteristics is an important function of distribution system monitoring. Demand, load factor(LF), coincident factor (CF), diversity factor (DF) and responsibility factor(RF) are the important load characteristics. To calculate these values, peak load of the distribution system is required. For example load factor (LF) is the ratio of the average load over the peak load[10]. So utilities must use advanced computing techniques to track the peak load of the distribution system at certain time intervals.

In this paper we discussed the function of an outage management system and its advantages. The importance of peak load tracing and outage detection in distribution system monitoring is the theme of this work. Methods to detect outages and their deficiencies are also mentioned in the following sections. Introspection of a smart grid distribution system with quantities of data expected, big data applications and big data technologies are presented in Section 2. Design of applications on stream computing framework for outage detection and peak load tracking using AMI data for distribution system monitoring is given in Section 3. Distribution monitoring system designed upon a open source stream computing platform for faster processing of data is the novelty of this paper.

## 2. Big Data Analytics in Smart Grid

Big data is the term applied to data sets whose size, type and speed is beyond the ability of the traditional databases to capture, manage, and process. Big data is characterized by one or more of the following features - high volume, high velocity or high variety. Big data comes from sensors, devices, video/audio, networks, transactional applications, web, and social media - much of it generated in real time and in a very large scale. Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include different types such as structured/unstructured and streaming/ batch, and different sizes from terabytes to zeta bytes [11].

Analyzing big data allows analysts, researchers, and business users to make better and faster decisions using data that was previously inaccessible or unusable. Using advanced analytical techniques such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, businesses can analyze previously untapped data sources independent or together with their existing enterprise data to gain new insights resulting in significantly better and faster decisions. This section is discussing the size of data in smart grid, big data applications in smart grid and technologies for big data analysis.

2.1. Big Data in Smart Grid

Power grid is the largest network exists in the world which is unidirectional. Smart grid well connects these people using information communication technique (ICT). Before attempting to solve any problem in smart grid it is required to estimate the size of the data to be handled by the technology. Literature review on the earlier attempts to quantify the data requirement of smart grid is given in this section.

In [12] author made an estimate of number of messages per second in a typical distribution system. It is 1 million (10,00,000) messages per second in a distribution system with one lakh users connected using 10000 feeders and considering the messages exchanged between the feeders[13]. As per DLMS/COSEM standard allows speed negotiation up to 9600 baud and supports segmentation and reassembly[14]. At this rate the total data reached from 1000000 messages will be more than 1GB data per second at distribution control center[9]. Schematic diagram of this distribution substation is shown in Figure 1[35].
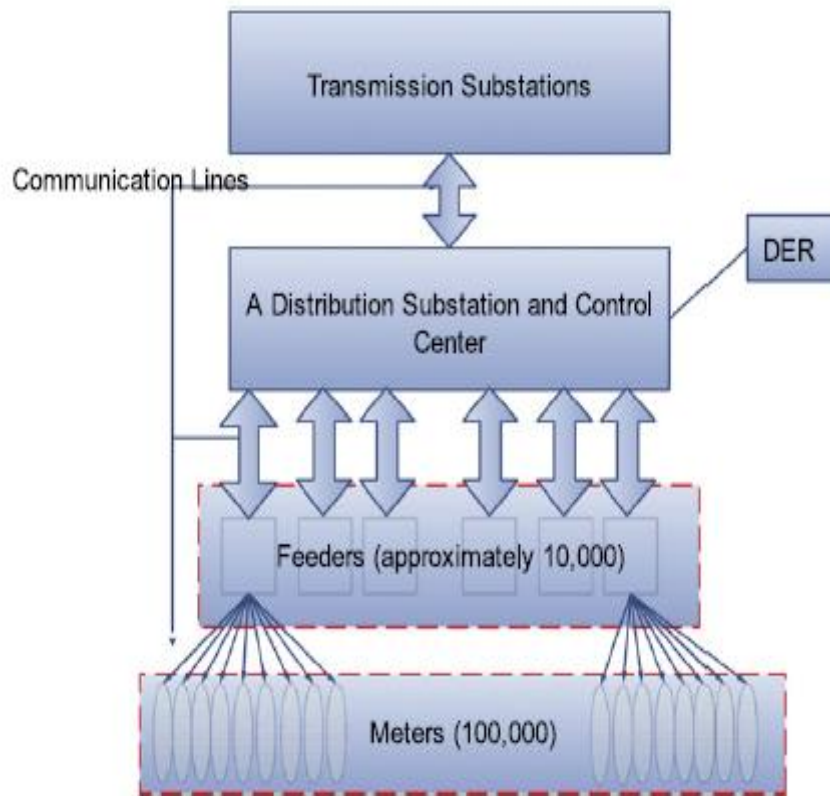
Figure 1. Distribution Substation Architecture

In[15] authors made an attempt to quantify the data generated on Dutch Grid. For a sampling period of 15 minutes data from Dutch Grid touches 1.5 Peta bytes per year by 8 million people. In [16] authors mentioned that data processing technology used in smart grid should be highly scalable because the increase in smart meter penetration will increase the data generated in smart grid. A technique to process high velocity big data without storage is stream computing. Two most popular stream processing systems are Apache S4 and Storm.

D. Alahakoon *et. al.*, stated that smart meters are the building block of smart grid[17]. Smart meters are mainly used to capture and transfer consumption data and events related to power quality and meter status (Connected/disconnected). The data has high volume, velocity and variety. To discover the value from this data, advanced analytics using appropriate tools and technology is very important. In essence it is big data.

A. Martin *et. al.*, pointed out that, the increased penetration of smart meters resulted in huge volume of data and hence, the data processing technology used in smart grid should be highly scalable[35]. Store and process approach is not suitable in this big data environment. Stream computing is the suitable technology for this situation.

All around the world, major utilities are now concentrating on the rollout of smart meters. Due to this rollout, by 2020, the major utilities of the world has to handle an amout of several Terra bytes(TB) per day[38]. This is contributed by different sources like smart meters, energy market pricing and bidding data, management data, control and maintenance data, financial data of operations, whether data and GIS data.

2.2. Big Data Applications in Power Distribution Systems

Before the wide spread deployment of AMI meters, power utilities have no visibility of the system beyond distribution substation. Only data available from distribution system was monthly billing data. AMI meters are equipped with sophisticated sensors to measure a variety of parameters like quality of power, consumption pattern, apart from billing data. Meters are reporting value at 15 min or 30 minutes interval. Utilities are now using this data to make decisions for grid operations, electric system planning, whole sale market participation and policy discussions.

Big data applications in distribution system planning can be divided into two categories i) Short term operations and ii) Long term planning studies[18]. Short term applications are detection of energy theft, outage detection, peak load monitoring, customer consumption behavior modeling, special load and renewable forecast, distribution system visualization, state estimation and distribution system planning, in which the first three applications are qualified to be very short term applications. Applications in Long term system planning studies include modeling customer consumption behavior under various incentives and pricing structures, transformation of distribution system planning process.

In [19] authors pointed out the potential applications of big data analytics in smart grid. Uncertain data mining, Non intrusive load monitoring, real time system peak load tracking, demand response, direct load control, electric vehicle charging system planning and electric vehicle real time dispatch are the potential smart grid application scenarios related to the big data analysis.

In[37] authors pointed out that making the power grid smart, is a big data challenge that requires advances in information processing technique and cyber infrastructure. Energy data, streaming from millions of communicating digital meters, which are sampled in every 15/30 minute interval, must be received and correlated with historical data of consumers. Implementation of this application requires a cloud based software platform for data driven analysis.

Authors in[38] presented a survey on big data application for smart grid. The survey pointed that, demand response for load management, common behavior analysis, state estimation using AMI & intelligent electronics devises(IEDs), pricing analysis for incentive implementation, optimization of grid infrastructure, demand forecast, asset management, service quality analysis, analysis of economic situation and prediction of social impact are the major big data applications for smart grid.

2.3. Big Data Analytical Technologies

Analyzing big data allows analysts, researchers and business users to make better and faster decision using data that was previously inaccessible or unknown. Scientists have developed a wide variety of techniques and technologies to capture, curate, analyze and visualize big data. The techniques involve a number of disciplines, including statistics, clustering algorithms, data mining, machine learning, signal processing, pattern recognition, optimization and visualization method[20]. Big data analysis can be done in two ways depending on the application needs, batch wise or real time processing[2]. Batch processing analytics commonly based on Map Reduce system where as Real Time systems analyze the data on the fly. The three major type of big data processing technologies are Data Warehouses(DWH), Batch processing using Hadoop and stream computing for real time analytics.

2.3.1. Data Warehouses (DWH)

DWHs are central repositories of integrated data from one or more disparate sources [21]. They store current and historical data and are used for creating analytical reports for knowledge workers throughout the enterprise. Analytical tools like clustering, machine learning and statistical tools can be applied over the stored data to generate actionable insights or information. Data warehouse is the historic source of truth stored using advanced tools and technologies.

2.3.2. Hadoop Based Batch Processing

At present, one of the most famous and powerful batch process-based big data tools is Apache Hadoop. It provides infrastructures and platforms for specific big data applications in business and commerce. Hadoop is a platform which implemented the map-reduce programming paradigm. There are three steps in the MapReduce model: Map, Shuffle and Reduce. Hadoop parallel process parallelizable problems across huge data sets using a large number of computing nodes[19].

2.3.3. Stream Computing Framework

Stream computing provides a computing framework to effectively capture and process the data streams. Stream processing solutions are designed to handle high volume in real time with a scalable, highly available and fault tolerant architecture. Stream computing enables analysis of data in motion. Common operations on streams are filter, aggregate, rule, correlation, time windows, transformation, automate, predict, act, monitor and alert. The data flow graph of a stream processor consists of stream source, filter, operators and stream sink as shown in Figure 2 [33]. A stream processing application is a collection of operators connected by streams[34]. Essential components of stream computing frame work are schematically shown in Figure 3.
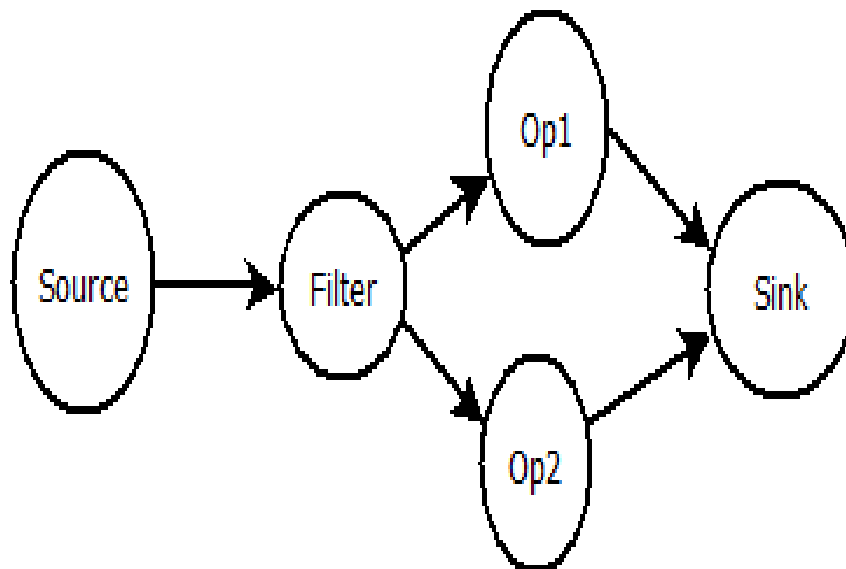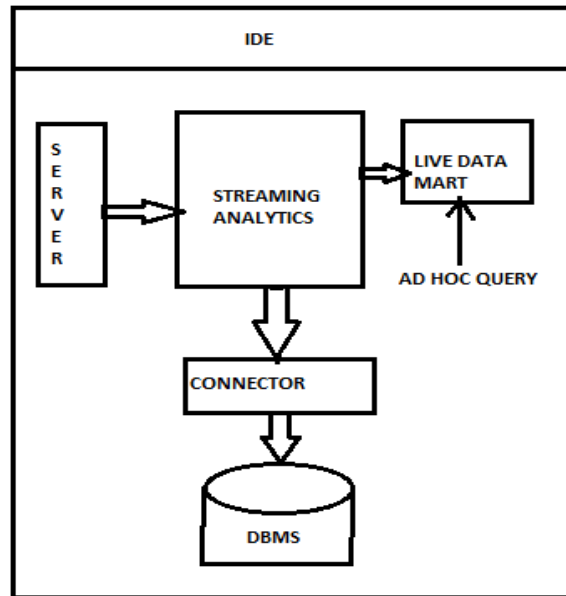

Figure 2. Data Flow Diagram of Stream Computing

Figure 3. Functional Units of Stream Computing Framework

For stream data applications, such as electric power system operation, would require real-time response for data processing platforms for example, Apache Storm which is designed for real-time stream data analytics. Widely used stream processing platforms are S4[22], Storm[23], StreamBase[24] *etc*. Features of above three big data processing technologies are compared in Table 1.

Table 1. Features of DWH, Hadoop and Stream Computing

Features	DWH	Hadoop	Stream Computing
Type of Data Stored	Structured	Structured & Unstructred	No storage
Storage purpose	Reporting & dash board	Long running computations	Real time analytics
Age of data	old	past	Current/new data
Size of data	Terra/Peta bytes	Giga Bytes	Kilo Bytes
Speed of processing	Peta bytes /day	Kbps	Mbps
Implementation cost	High	Medium	Low
Volume	High	High	Low
Volume	High	High	Low
Velocity	Nil	Nil	High
Variety	Nil	High	High

3. Stream Computing Based Distribution System Monitoring

"Distribution intelligence refers to the part of the Smart Grid that applies to the utility distribution System, that is, the wires, switches, and transformers that connect the utility substation to you, the customers" [25]. A key component of distribution intelligence is outage detection and response. Today many utilities rely on customer phone calls to know which area of their distribution system are being affected by power outage. Along with

smart meters distribution intelligence will help to quickly pin point the source of power outage so that repair crew can be immediately dispatched to the problem area. This can be materialized only through real time analysis of smart meter data. Outage intelligence helps distribution managers, outage managers, community liaisons and regulatory affair managers apply analytics for predicting, preventing, detecting, assessing and responding to outages[26]. Use analytics to pin point outages before customers have time to call. Faster notification leads to faster restoration. According to Mark Donsky, AMI meter play a crucial role in speedy service restoration[27]. They give immediate last gasp alarm whenever power is lost and service restoration notification when power is restored. Social media networks like twitter can be used as an alternative source to sense power outages, but response time is counted in fraction of hours[31]. Beyond smart meters, automatic switches or reclosers and faulted circuit indicators (FCI) are two technologies which helps to speed up power restoration[32]. Outage management system perform a crucial role in quality utility operations. In this paper two typical components of distribution system monitoring application are presented. Outage detection and peak load tracking applications are implemented on Apache Storm stream computing framework using AMI data.

3.1. Outage Detection System

Outage Management is the second most important beneficiary of AMI data[28]. The first step in outage management is the detection of outages. Last gasp messages along with location information of the meters can be used to pin point the outages and the severity.

3.2. Peak Load Tracking system

The electrical load of a feeder is the sum of loads of all individual customers[10]. In radial distribution systems, peak load is calculated as maximum of the sum of the load of the individual consumers of the system at the fixed interval. Load values are sent by customers on 15/30 minutes interval. The data reported by the utility contain meter_ID, timestamp, kWh, kW, voltage, current, frequency, PF_LAG_OR_LEAD, kVA, kVAh.

Apache Storm is a widely used platform to design stream computing applications. Spouts and Bolts are the building blocks of the program. A Spout is a generator of data which can filter and transform the tuples received from other process or files and send to appropriate bolts. Bolts implement the business logic of the application and can pass the tuples to next level bolts. The interconnection of Spouts and Bolts form an acyclic graph in the program and is termed as Topology [29].

The number of Spouts and Bolts of different kind determines the speed up of the program and can be configured to utilize maximum parallelism available in the machine or cluster. Smart meters will report load data in 30 minutes interval and last gasp messages when power outages or restoration happens. Sketch of the topology used is shown in the Figure 4.

4. Implementation and Results

Topology is implemented using Apache Storm. Tuple structure used is (meterID, time, outage, value) for last gasp messages and (meterID, timestamp, kWh, kW, voltage, current, frequency, PF_LAG_OR_LEAD, kVA, kVAh) for load values. Topology tested on a local cluster with Ubuntu 14.04 OS loaded Intel Xeon 3.70 GHz processor machine with 8 cores. System has 16 GB RAM and 300 GB hard disk.

Storm UI is a user interface which can be displayed on a web browser to get the status of the topology running on a Storm cluster. Main sections displayed on a Storm UI are cluster summary, topology summary and supervisor summary. Topology summary gives details about the names of topologies running on a cluster. Under each topology Spouts

and Bolts on the topology with a detailed list of statistics are displayed. Storm UI can also give a visual representation of the topology [30]. Figure 5 shows the screen shot of the topology implemented for distribution system monitoring on Apache Storm.

Collected 2.6 lakhs records reported in 30 minutes interval data for one month from 440 meters of residential consumers installed by a distribution company in India as part of their smart grid pilot project implementation. Total processing time taken is 32 seconds as reported by Storm UI by using 2 cores of the system. This shows that 160 μs was taken to process each event and able to achieve a speed of 6250 messages per second. It takes more than 2.5 minutes to process 1 million messages which is generated in one second. A cluster setup with 10 machines of 1 Gbps Ethernet cards can handle 1.2 GB data per second which will be generated under a feeder. The status of the meter is saved to a file and displayed using web server. Outage detection system relying on social media as a sensor requires 10 minutes to report the event with less than 100% accuracy[31]. Stream computing is 20 times faster in accurately reporting outages compared to social sensors. The screen shot of the OMS is shown in figure 6a. Peak load for different dates is shown in Figure 6b.
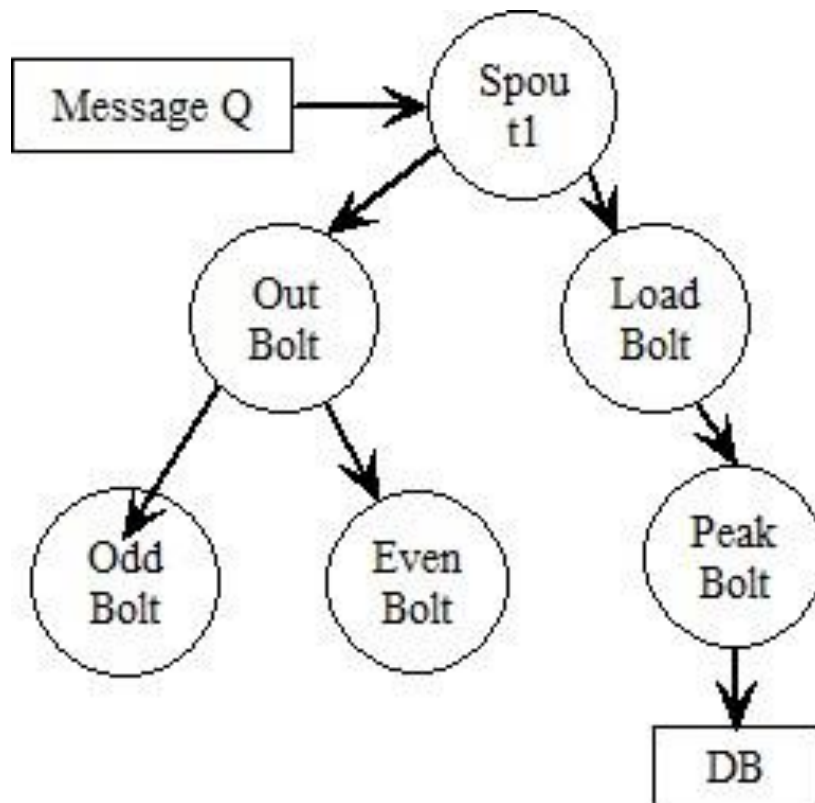

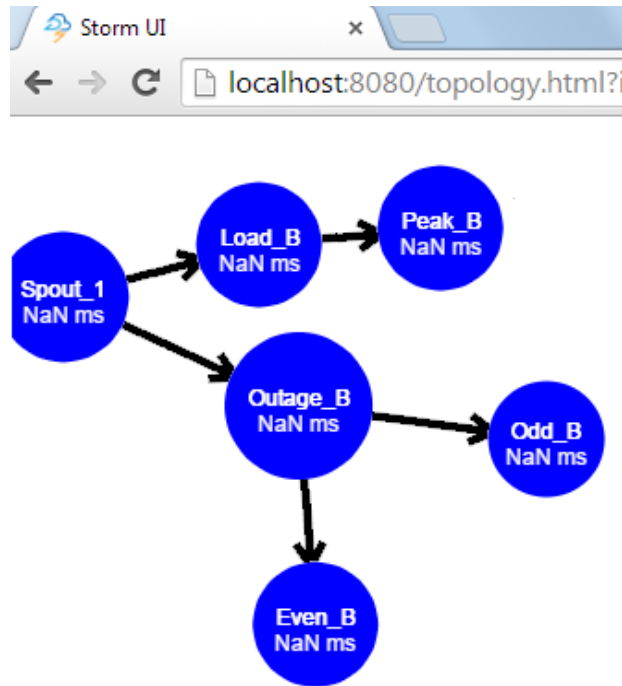**Figure 4. Distribution Monitoring Topology**

Figure 5. UI Storm Topology

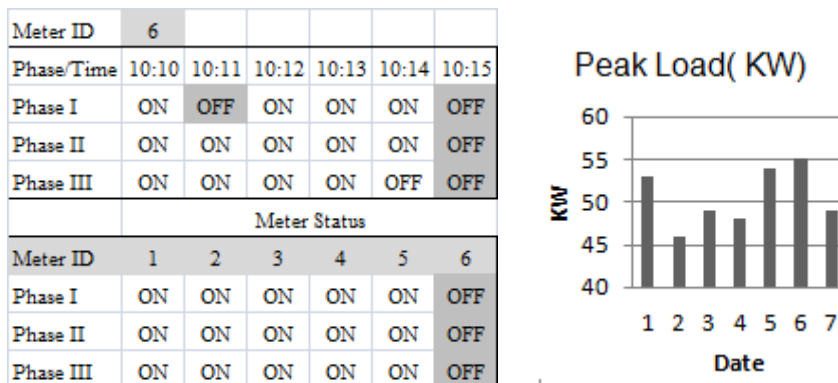| Meter ID	6					
Phase/Time	10:10	10:11	10:12	10:13	10:14	10:15
Phase I	ON	OFF	ON	ON	ON	OFF
Phase II	ON	ON	ON	ON	ON	OFF
Phase III	ON	ON	ON	ON	OFF	OFF
	Meter Status					
Meter ID	1	2	3	4	5	6
Phase I	ON	ON	ON	ON	ON	OFF
Phase II	ON	ON	ON	ON	ON	OFF
Phase III	ON	ON	ON	ON	ON	OFF

Figure 6. a)   Meter Status   b) Peak Load of Distribution Substation

5. Conclusion

Outage Detection and peak load tracking are two important functions of distribution system monitoring. On line analysis of last gasp messages from smart meters make outage restoration activities faster. Time stamps of last gasp messages can be used to update reliability indexes (SAIDI, CAIDI, SAIFI, MAIFI) also. Peak load tracking is required to compute load characteristics values LF, CF, DF and RF. Peak load is calculated from the load values reported in 30 minutes interval.

As the smart grid data reached the volume, velocity and variety of big data, big data analytics have to be used in smart grid computations. Outage detection and peak load tracking requires on line analysis of AMI messages. Here we implemented a distribution system monitoring framework on Apache Storm stream computing platform. Stream Computing system attained 20 times speed up compared to outage detection system relying on social media as a sensor. This paper verified the suitability and applicability of stream computing framework for distribution system monitoring through two critical functions.

References

[1]   U.S. Department of Energy, Smart Grid / Department of Energy, http://www.energy.gov Retrieved 2015-03-10.

[2]   Shibily Joseph, Jasmin E.A, Soumya C, "Stream Computing: Opportunities and Challenges in Smart Grid" In proceedings of the International Conf. on Smart Grid Technologies, Aug 2015,

[3]   United States Department of Energy, "GRID 2030, A National Vision for Electricity's Second 100 Years."

[4]   Jeff St.John, Big Data on the Smart Grid: 2013 in Review and 2014 Outlook. Greentechgrid, article, December 2013

[5]   Dong, Ming, Paulo CM Meira, Wilsun Xu, and Walmir Freitas. "An event window based load monitoring technique for smart meters." Smart Grid, IEEE Transactions on 3, no. 2 (2012): 787-796.

[6]   Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan, The rise of big data on cloud computing: Review and open research issues, Information Systems, Volume 47, January 2015, p. 98-115.

[7]   Measuring Performance of Electric Power Distribution Systems. IEEE Standard 1366, 2003.

[8]   Mandira Srivastava, "Utilities with Outage Management as a Top Priority Excelling in Customer Service Rating", Bridge Energy Group, Jan. 2014

[9]   Shibily Joseph, Jasmin E.A "Stream Computing Framework for Outage Detection in Smart Grid" In proceedings of the IEEE International Conf. on Power Instrumentation Computing and Control, Dec 2015

[10]   Thomas Allen. Electric power distribution handbook. CRC press, 2014.

[11]   What is big data analysis [Online]. Available:http://www-01.ibm.com/software/data/infosphere/hadoop/what-is-big-data-analytics.html.

[12]   A. Aggarwal, S. Kunta, P.K Verma, A proposed communications infrastructure for the Smart Grid, in: 2010 Innovative Smart Grid Technologies ISGT 2010, pp. 1–5.

[13]   Gao, Jingcheng, et al. "A survey of communication/networking in smart grids."Future Generation Computer Systems 28.2 (2012): 391-404.

[14]   EXCERPT DLMS UA Blue Book: COSEM interface classes and OBIS identification system, Seventh Edn.

[15]   Aiello, Marco, and Giuliano Andrea Pagani. "The smart grid's data generating potentials." In Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on, pp. 9-16. IEEE, 2014.

[16]   Martin, Andre, Rodolfo Silva, Andrey Brito, and Christof Fetzer. "Low Cost Energy Forecasting for Smart Grids Using Stream Mine 3G and Amazon EC2." In Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on, pp. 523-528. IEEE, 2014.

[17]   Alahakoon, Damminda, and Xinghuo Yu. "Advanced analytics for harnessing the power of smart meter big data." In Intelligent Energy Systems (IWIES), 2013 IEEE International Workshop on, pp. 40-45. IEEE, 2013.

[18]   Yu, Nanpeng, et al. "Big data analytics in power distribution systems."Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society. IEEE, 2015.

[19]   Luo, Fengji, et al. "Enabling the big data analysis in the smart grid." Power & Energy Society General Meeting, IEEE, 2015.

[20]   Chen, CL Philip, and Chun-Yang Zhang. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data." Information Sciences 275 (2014): 314-347.

[21]   Data warehouse [online] . Available: https://en.wikipedia.org /wiki/ Data _ warehouse

[22]   S4: Distributed Stream Computing Platform [online]. Available: http://incubator.apache.org/s4/

[23]   Storm: Distributed and Fault-Tolerance Real Time Computation [online]. Available: https://storm.apache.org/

[24]   StreamBase [online], Available: http://www.streambase.com/

[25]   Distribution Intelligence [Online], Available: https://www.smartgrid.gov/the_smart_grid /distribution_intelligence.html

[26]   Outage Intelligence [Online], Available: http://www.spacetimeinsight.com/solutions/utilities/ outage-intelligence/

[27]   Mark Donsky, Accelerating Outage Restoration with Smart Grid Technology [online] , Available: http:// www.silverspringnet.com/ article/accelerating-outage-restoration-with-smart-grid-technology/

[28]   Outage Management: The Second Most Important Beneficiary of AMI Data [Online], Available: http://www.energycentral.net/ article/07/06/ outage-management-second-most-important-beneficiary-ami-data

[29]   Tutorial "Storm" The Apache Software Foundation [Online]: Available : https://storm.apache.org/ documentation/ Tutorial.html

[30]   Romesh Malinga Perera, "Reading and Understanding Storm UI" [Online], Available : http:// www. Malinga .me / reading-and-understanding-the-storm-ui-storm-ui-explained

[31]   Sun, Hiafeng, et al. "Data-Driven Power Outage Detection by Social Sensors.", IEEE Transactions on smart Grid, Vol.7, No.5, September 2016

[32] Krajnak, David J. "Faulted circuit indicators and system reliability." Rural Electric Power Conference, 2000. IEEE, 2000.

[33] Fernandez, Raul Castro, Matthias Weidlich, Peter Pietzuch, and Avigdor Gal. "Scalable stateful stream processing for smart grids." In Proceedings of the 8th ACM International Conference on Distributed Event-Based Systems, pp. 276-281. ACM, 2014.

[34] Simmhan, Yogesh, Baohua Cao, Michail Giakkoupis, and Viktor K. Prasanna. "Adaptive rate stream processing for smart grid applications on clouds." InProceedings of the 2nd international workshop on Scientific cloud computing, pp. 33-38. ACM, 2011.

[35] Gao, Jingcheng, et al. "A survey of communication/networking in smart grids."Future Generation Computer Systems 28.2 (2012): 391-404.

[36] Martin A, Andre, Rodolfo Silva, Andrey Brito, and Christof Fetzer. "Low Cost Energy Forecasting for Smart Grids Using Stream Mine 3G and Amazon EC2." In Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on, pp. 523-528. IEEE, 2014.

[37] Simmhan, Yogesh, et al. "Cloud-based software platform for big data analytics in smart grids." Computing in Science & Engineering 15.4 (2013): 38-47.

[38] Lai, Chun Sing, and Loi Lei Lai. "Application of Big Data in Smart Grid." Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on. IEEE, 2015.