ANALYSIS OF SMART METER DATA USING HADOOP

[1]Balaji K. Bodkhe, [2]Dr. Sanjay P. Sood
MESCOE Pune, CDAC Mohali
Email:[1]balajibodkheptu@gmail.com,[2]spsood@gmail.com

Abstract— The government agencies and the large multinational companies across the world focuses on energy conservation and efficient usage of energy. The need of using energy in a efficient way is the need of developing countries like India and China .The emergence of smart grid meters gave us access to huge amount of energy consumption data. This data provided by smart meters can be used efficiently to provide insights into energy conservation measures and initiatives. Various energy distribution companies harness this data and get unpredictable results about customer's usage pattern; they then after performing analysis predict the demand and consumption of users. This analysis helps them to decide the tariff at different point of time. The companies are trying to overcome the bottleneck in capital investment cost of data .Further, processing Big Data for chart generation and analytics is a slow process and is not fast enough to support real-time decision making. Our paper showcases a Business Intelligence tool which uses Apache Hadoop to efficiently handle the existing problems. Taking the advantage of this tool, energy distribution companies can reduce the investment by using community hardware that runs Hadoop. The usage of distributed computing tools also reduces the processing time significantly to enable real-time monitoring and decision making .This tool will also reduce carbon footprint and other related problems in energy distribution including loses and theft .In future this same analysis can be done on other utility resources such as gas and water.

Index Terms—About four key words or phrases in alphabetical order, separated by commas.

I. INTRODUCTION

Analytics of energy consumption data to gain insights into customer usage patterns is what energy distributors are trying to achieve for several target applications such as time-of use tariff, demand response management and billing accuracy. This smart meter collects data every minute which results in generating large amount of data while the old mechanical meter collects data by hourly or monthly. This huge data storage capabilities and the complexity of data processing intelligence varies significantly with different applications. Traditional RDBMS of utility companies is a bottleneck in executing this approach. As a result, for the industry to truly benefit from the smart grid investment, it is critical that the massive amount of data made available by smart meters be handled efficiently in an organized manner that helps grid operators make timely decisions to operate grid safely, economically and reliably. Apache Hadoop is the solution available to tackle above problems which runs on commodity machines only. It is distributed computing tool which have large storage as well as processing capability.

We are using Apache Hadoop framework that allows for the distributed processing of large data sets across clusters of computers. Hadoop MapReduce is a system for parallel processing of large data sets. Traditional RDBMS or other applications are much slower and inefficient in handling big data generated by smart meters as compared to Hadoop framework. So for the

industry and customers to gain benefits like billing accuracy, energy theft detection, analyzing customer usage patterns and demand response management etc. It is always profitable to use Hadoop which runs on cheap commodity hardware.

With the evolution of smart meters for smart distribution and efficient use of energy, electricity, the generated power should be utilized properly with fair economy gains to distributors and the consumers. Thus with this focus of energy distribution in the domain of energy consumption, which will result in reduction of carbon prints, the analytics for the data received from the smart meters should be done. This massive size of analytics will need large computation which can be done with the help of distributed processing framework, Hadoop. The framework's use will provide multipurpose beneficial outputs which include: billing accuracy, time-of-use tariff plans etc. Thus this concept, smart meter data analytics, is implemented with a view of future use.

II. HADOOP

Smart meters send energy consumption data to the server at regular intervals of time which results in generating big amount of data and existing tools were not capable of handling such large amounts of data. Apache Hadoop is an open source framework for developing distributed applications that can process very large amounts of data. It is a platform that provides both distributed storage and computational capabilities.

Hadoop has two main layers:

1. Computation layer: The computation tier uses a framework called MapReduce.
2. Distributed storage layer: A distributed file system called HDFS provides storage.

Hadoop Advantages:

- Hadoop is an open source, versatile tool that provides the power of distributed computing.
- By using distributed storage & transferring code instead of data, Hadoop reduces the costly transmission step when working with large data sets to a great extent.
- Redundancy, hadoop can recover from a situation when a single node fails.
- Ease to create programs with Hadoop As it uses the MapReduce framework.

You did not have to do worry about partitioning the data, determining which nodes will perform which tasks, or handling communication between nodes as it is all done by Hadoop for you.

- Hadoop leaving you free to focus on what is most important to you and your data and what you want to do with it.

Hadoop Key Features:

Distributed computing is the very vast field but following key features has made Hadoop very distinctive and attractive.

A. Accessible:

Hadoop runs on large clusters of commodity machines or on cloud computing services such as Amazon's Elastic Compute Cloud (EC2).

B. Robust:

As Hadoop is intended to run on commodity hardware, it is architected with the assumption of frequent hardware malfunctions. It can gracefully handle most such failures.

C. Scalable:

Hadoop scales linearly to handle larger data by adding more nodes to the cluster.

III. SYSTEM ARCHITECHTURE

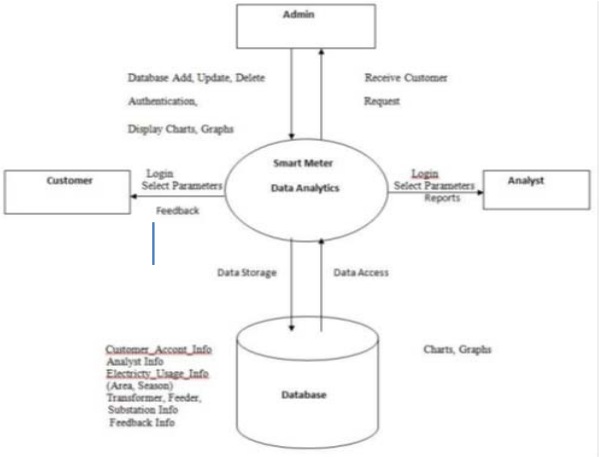

Fig. 1 System Architechture

IV. SMART METER DATA ANALYSIS

Electricity generation in India majorly depends upon the non-renewable sources [1]. Though India is rich in these resources but these resources are depleting at an alarming rate such that they will be exhausted very soon whereas the renewable resources are not utilized to their capacity. Thus India needs to concentrate more on the renewable sources of energy such as wind

energy harvesting in southern and western India where wind velocity is high.

Poor metering, power theft, lack of proper planning, overload on the resources are the few reasons to the present poor grid conditions in India. The transmission and distribution costs are so high that the government loses a lot of money on every unit of electricity sold.

Keeping all the above factors in mind the government of India has taken many steps towards the betterment and improvement of the electricity grid. Smart metering, Variable tariffs, the Electricity act 2003, etc. are some of the initiatives taken.

The Smart grid concept presented in this paper is a step by step process specially tailored to Indian conditions which when followed will lead to a very effective and well managed smart electricity grid by 2024. This will include well established production and transmission devices, smart metering, transparency in the working of the management, visual analytics for both the provider and the consumer.

Various hurdles will have to be jumped over to reach this stage the biggest one will be that of corruption, once that is overcome then rest of the problems can be collectively solved by heuristic methods. Secondly government should try and involve the IT companies into this so that experts from these companies can make necessary amendments. The top ranked colleges can be involved in this process where the students might provide some valuable insights.

To save energy, reduce cost, and increase reliability, billions of dollars are being invested by the U.S. government and private industries to build the smart grid infrastructure [2]. Higher resolution measurements are made available to more equipment at wider areas by the wide deployment of modern information technology into power grid control and communication networks.

Consider Fig.2 for example. According to this figure, data is collected by a smart meter by the minute while the data is collected hourly or monthly by an old mechanical meter; 30-60 data points are collected by a phasor measurement unit (PMU) per second which is much faster than the sampling rate of the traditional mechanical system named supervisory control and data acquisition (SCADA) system, which is 1 data point per 1-2 second.

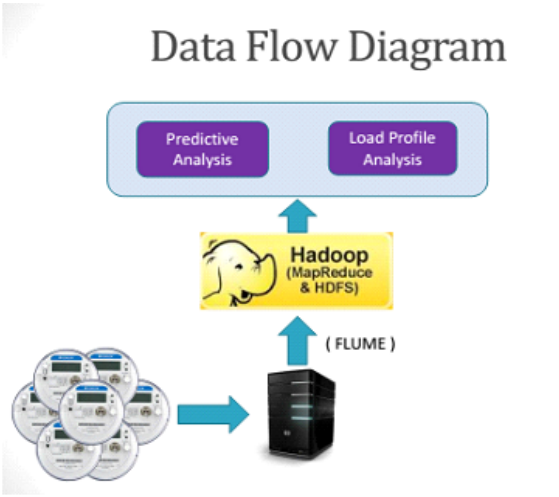

Fig.2: Data Flow Diagram

V. IMPLEMENTATION AND RESULTS

The system that we implemented is useful for both electricity providers and customers in terms of cost efficiency and analysis. Smart meters collect and send electricity consumption data at particular intervals which results in generating large amount of data. The best software that we can use for handling such big data is Hadoop. Hadoop is infrastructure developed by Apache and leading distributed computing platform. Hadoop was derived from Google's MapReduce and Google File System (GFS). Hadoop is open-source platform and works on master-slave configuration. Hadoop runs on commodity hardware.

In our implementation HDFS (Hadoop Distributed File System) acts as our database where we are going to store all the data which includes electricity consumption data, user account information.

GUI (Graphical User Interface):

Our main module is the GUI which tells entirely about the project that we have implemented. It describes different functionalities provided to user of the system. Here two types of user can gain access to the system, one is the analyst and another is consumer. Only authenticated users can be allowed to use the system. Both users can view electricity consumption data of their choice based on different given parameters (week, month etc.) available in the system. Then depending upon their selected choice, graph is generated so that they can analyze the consumption of electricity and further achieve various objectives mentioned. Main aim of GUI

is to provide flexible and efficient system to users.

ANALYSIS MODEL 1:

Module Name	**Energy Consumption Prediction**
Functionality	Following functionality is provided in this module 1. Can access different parameters (week, season etc.) 2. Can access graph of choice
General Workflow	Following steps are included in general workflow of module 1. Login 2. Select Parameters 3. Select Graph
Dependency	Following module on which this module is dependent Login for username and password

Table I Analysis Model 1

ANALYSIS MODEL 2:

Module Name	**Drill Down**
Functionality	Following functionality is provided in this module 1. Can access different parameters (week, season etc.) 2. Can access graph of choice
General Workflow	Following steps are included in general workflow of module 1. Login 2. Select Parameters 3. Select Graph
Dependency	Following module on which this module is dependent Login for username and password

Table II Analysis Model 2

CONSUMER MODEL:

Module Name	**Customer**
Functionality	Following functionality is provided in this module 1. Can access different parameters (week, season etc.) 2. Can access graph of choice
General Workflow	Following steps are included in general workflow of module 3. Login 4. Select Parameters 5. Select Graph
Dependency	Following module on which this module is dependent Login for username and password

Table III Customer Model

The fig 4 shows the comparison of energy consumption in the month of February and fig 5 shows the running hadoop job on slave 3

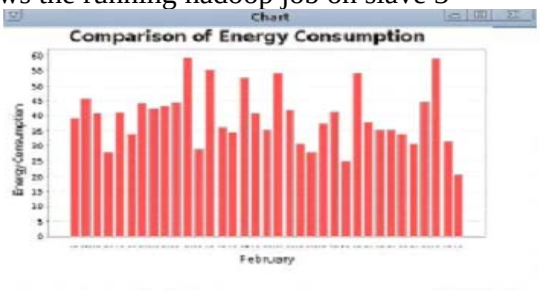

Fig. 4: Comparison of energy consumption

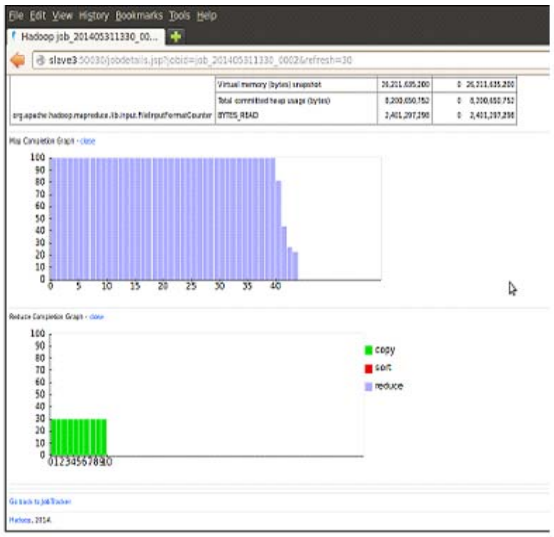

Fig. 5: Ongoing Hadoop Jobs of Slave 3

VI. CONCLUSION

To reach the 2020 energy efficiency as well as renewable energy targets and also for the future smart grids, effective use of smart metering technology is crucial. Rational energy use is a must for a larger group of companies, municipalities and public organizations because of the gain in importance of the energy costs and environmental issues. Hence proper information about their consumption is needed by them along

with and its distribution between different activities. A total picture of their energy use, potential for savings, along with costs can be given to them by smart meter data analytics, enabling effective energy management. Smart meter sends energy consumption data at small intervals resulting in generating big data. Time and storage are two important factors that affect a lot on building any application. The solution for handling such big data is Hadoop.

VII. FUTURE SCOPE

For a good understanding of how customer, environmental and structural features affect the usage, the techniques for the analysis of energy consumption data by using smart meters will be helpful. Analysis of customer behavior can be performed easily by analyzing this data which will be helpful to reduce energy consumption. A utility for better management of power outages and restoration events as well as reduction in outage duration and costs is allowed by Outage Management System (OMS). Reduced CO2 emission is resulted by demand response because of avoided use of polluting power plant. Also, as a result, reduced peak prices are found because of avoided use of expensive peak load production. Because of customers' awareness, reduced consumption is found as a result of feedback as well as load management regarding energy consumption. Analysis of smart meter data in accordance with weather can also be performed. In India, main challenge is to install the smart meter all over the country as internet facility is still not present in most of the regions.

REFERENCES

[1] Amir-Hamed Mohsenian-Rad, Member, IEEE, and Alberto Leon-Garcia, Fellow, IEEE, "Optimal Residential Load Control With Price Prediction in Real-Time Electricity Pricing Environments", IEEE TRANSACTIONS ON SMART GRID, VOL. 1, NO. 2, SEPTEMBER 2010

[2] Ning Lu, Pengwei Du, Xinxin Guo, and Frank L. Greitzer , "Smart Meter Data Analysis.

[3] Dr. Sanjay Gupta, Ashish Mohan Tiwari, Cognizant, "How to Generate Greater Value from Smart Meter Data."

[4] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung Google, "The Google File System."

[5] Amir-Hamed Mohsenian-Rad, Member, IEEE, and Alberto Leon-Garcia, Fellow, IEEE, "Optimal Residential Load Control With Price Prediction in Real-Time Electricity Pricing Environments", IEEE TRANSACTIONS ON SMART GRID, VOL. 1, NO. 2, SEPTEMBER 201

[6] Li Ping Qian, Member, IEEE, Ying Jun (Angela) Zhang, Senior Member, IEEE, Jianwei Huang, Senior Member, IEEE, and Yuan Wu, Member, IEEE, "Demand Response Management via Real-Time Electricity Price Control in Smart Grids", IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 31, NO. 7, JULY 2013 "SMART METER DATA ANALYTICS USING HADOOP (SMDA)" MES COLLEGE OF ENGINEERING, PUNE 87

[7] Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall and Werner Vogels Amazon.com, "Dynamo: Amazon's Highly Available Key-value Store"

[8] Fahad Javed, Naveed Arshad, Fredrik Wallin, Iana Vassileva , Erik Dahlquist, "Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting"

[9] Shannon Warner, Samrat Sen, Cognizant, "How Predictive Analytics Elevate Airlines' Customer Centricity and Competitive Advantage"2, Aug. 1987, pp. 740–741 [Dig. 9th Annu. Conf. Magnetics Japan, 1982, p. 301].