INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

Identification of Electrical Devices Applying Big Data and Machine Learning Techniques to Power Consumption Data

[1]Professor Kannadasan R ,
Department of Software systems, SCOPE ,VIT University,Vellore-632014

[2]P V V Subhash ,
UG Scholar, School of Computer Science and Engineering, VIT Univeristy, Vellore

Abstract

The government agencies and the large multinational companies across the world focus 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: Load profiling, big data, electricity consumption, behavior dynamics, demand response

1. INTRODUCTION

Countries around the world have set aggressive goals for the restructuring of monopolistic power system towards liberalized markets especially on the demand side. In a competitive retail market, load serving entities (LSEs) will be developed in great numbers [1]. Having a better understanding of electricity consumption patterns and realizing personalized power managements are effective ways to enhance the competitiveness of LSEs [2]. Meanwhile, smart grids have been revolutionizing the electrical generation and consumption through a two-way flow of power and information. As an important information source from the demand side, advanced metering infrastructure (AMI), has gained increasing popularity worldwide; AMI allows LSEs to obtain electricity consumption data at high frequency, e.g., minutes to hours [3]. Large volumes of electricity consumption data

reveal information of customers that can potentially be used by LSEs to manage their generation and demand resources efficiently and provide personalized service. 90% of the world's data was generated in the last few years. Due to the advent of new technologies, devices, and communication means like social networking sites, the amount of data produced by mankind is growing rapidly every year. The amount of data produced by us from the beginning of time till 2003 was 5 billion gigabytes. If you pile up the data in the form of disks it may fill an entire football field. The same amount was created in every two days in 2011, and in every ten minutes in 2013. This rate is still growing enormously. Though all this information produced is meaningful and can be useful when processed, it is being neglected. Big Data is a collection of large datasets that cannot be processed using traditional computing techniques. It is not a single technique or a tool rather it involves many areas of business and technology. Load profiling, which refers to electricity consumption behaviors of customers over a specific period, e.g., one day, can help LSEs understand how electricity is actually used for different customers and obtain the customers' load profiles or load patterns. Load profiling plays a vital role in the Time of Use (ToU) tariff design [4], nodal or customer scale load forecasting [5], demand response and energy efficiency targeting [6], and non-technical loss (NTL) detection [7].

The core of load profiling is clustering which can be classified into two categories: direct clustering and indirect clustering [8]. Direct clustering means that clustering methods are applied directly to load data. Heretofore, there are a large number of clustering techniques that are widely studied, including k-means [9], fuzzy k-means [10], hierarchical clustering [11], self-organizing maps (SOM) [12], support vector clustering [13], subspace clustering [14], ant colony clustering [15] and etc. The performance of each clustering technique could be evaluated and quantified using various criteria, including the clustering dispersion indicator (CDI), the scatter index (SI), the Davies-Bouldin index (DBI), and the mean index adequacy (MIA). [16].

The deluge of electricity consumption data with the widespread and high-frequency collection of smart meters introduces great challenges for data storage, communication and analysis. In this context, dimension reduction methods can be effectively applied to reduce the size of the load data before clustering, which is defined as indirect clustering. Such clustering can be categorized into two sub-categories, feature extraction-based clustering and time series-based clustering. Feature extraction which transforms the data in the high-dimensional space into a space of fewer dimensions [17], is often used to reduce the scale of the input data. Principal component analysis (PCA) [18] [19] is a frequently used linear dimension reduction method. It tries to retain most of the covariance of the data features with the fewest artificial variables.

The existing studies on load profiling mainly focus on individual large industrial/commercial customer, medium or low voltage feeder, or a combination of small customers, load profiles of which shows much more regularity [25]. It should be noted that although these dynamic characteristics are always "deluged" in a combination of customers, they could be described by several typical load patterns. However, with regard to residential customers, at least two new challenges will be faced. One challenge is the high variety and variability of the load patterns.

In this paper we are analyzing Electricity data by using hadoop tool along with some hadoop ecosystems like hdfs, mapreduce, sqoop, hive and pig. By using these tools we can process no limitation of data, no data lost problem, we can get high throughput, maintence cost also very less and it is a opensoure software,it is compatible on all the platforms since it is Java based.

2. LITERATURE REVIEW

In [3]There is growing interest in discerning behaviors of electricity users in both the residential and commercial sectors. With the advent of high-resolution time-series power demand data through advance metering

mining this data could be costly from the computational viewpoint. One of the popular techniques is clustering, but depending on the algorithm there solution of the data can have an important influence on the resulting clusters. This paper shows how temporal resolution of power demand profile saffects the quality of the clustering process, the consistency of cluster membership (profiles exhibiting similar behavior), and the efficiency of the clustering process. This work uses both raw data from household consumption data and synthetic profiles. The motivation for this work is to improve the clustering of electricity load profiles to help distinguish user types for tariff design and switching, fault and fraud detection, demand-side management, and energy efficiency measures. The key criterion for mining very large data sets is how little information needs to be used to get a reliable result, while maintaining privacy and security.

In[5] With the construction of smart grid, lots of renew- able energy resources such as wind and solar are deployed in power system. It might make the power system load varied complex than before which will bring difficulties in short-term load forecasting area. To overcome this issue, this paper proposes a new short-term load forecasting framework based on big data technologies. First, cluster an aliases per formed to classify daily load patterns for individual loads using smart meter data. Next, an association analysis is used to determine critical influential factors. This is followed by the application of a decision tree to establish classification rules. Then, appropriate forecasting models are chosen for different load patterns. Finally, the forecasted total system load is obtained through an aggregation of an individual load's forecasting results. Case studies using real load data show that the proposed new framework can guarantee the accuracy of short-term load forecasting within required limits. In [6]There are several pattern-based clustering methods which are used for different applications such as pattern recognition, data mining, etc. In recent years, some of these methods are implemented in power system studies, especially for clustering load curves for designing suitable tariffs, demand response programs selection, etc. Choice of the best clustering method for certain application is one of the most important issues which is case dependent and should be considered in using of clustering load curves. Demand response programs are widely used in power system for different applications such as peak clipping, demand reduction, etc. since demand response programs are featured with different characteristics. Therefore, selection of suitable programs for different customer classes is of great importance. In this paper, an improved weighted fuzzy average (WFA) K-means for the purpose of demand response programs applications is developed. This method is implemented on 316 load curves of Tehran distribution network and the results are investigated. In[7] This paper proposes a comprehensive framework to detect non-technical losses (NTLs) and recover electrical energy (lost by abnormalities or fraud) by means of a data mining analysis, in the Spanish Power Electric Industry. It is divided into four sections: data selection, data preprocessing, descriptive, and predictive data mining. The authors insist on the importance of the knowledge of the particular characteristics of the Power Company customer: the main features available in databases are described. The paper presents two innovative statistical estimators to attach importance to variability and trend analysis of electric consumption and offers a predictive model, based on the Generalized Rule Induction (GRI) model. This predictive analysis discovers association rules in the data and it is supplemented by a binary Quest tree classification method. The quality of this framework is illustrated by a case study considering a real database, supplied by Endues a Company.

In[26]The increasing US deployment of residential advanced metering infrastructure (AMI) has made hourly energy consumption data widely available. Using CA smart meter data, we investigate a household electricity segmentation methodology that uses an encoding system with a pre-processed load shape dictionary. Structured approaches using features derived from the encoded data drive five sample program and policy relevant energy lifestyle segmentation strategies. We also ensure that the methodologies developed scale to large data sets.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

In[27] Clustering methods are increasingly being applied to residential smart meter data, which provides a number of important opportunities for distribution network operators (DNOs) to manage and plan low-voltage networks. Clustering has a number of potential advantages for DNOs, including the identification of suitable candidates for demand response and the improvement of energy profile modeling. However, due to the high stochastic city and irregularity of household-level demand, detailed analytics are required to define appropriate attributes to cluster. In this paper, we present in-depth analysis of customer smart meter data to better understand the peak demand and major sources of variability in their behavior. We find four key time periods, in which the data should be analyzed, and use this to form relevant attributes for our clustering. We present a finite mixture model- based clustering, where we discover ten distinct behavior groups describing customers based on their demand and their variability. Finally, using an existing bootstrap technique, we show that the clustering is reliable. To the authors' knowledge, this is the first time in the power systems literature that the sample robustness of the clustering has been tested.

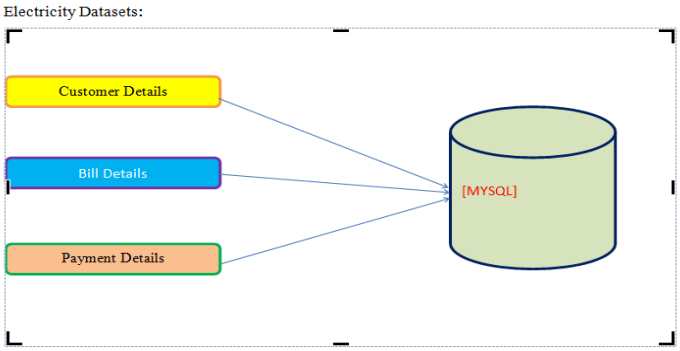
3. SYSTEM MODEL

Propsed concept deals with providing database by using hadoop tool we can analyze no limitation of data and simple add number of machines to the cluster andwe get results with less time, high throughput and maintence cost is very less and we are using joins,partations and bucketing techniques in hadoop. In this paper we are analyzing turism data by using hadoop framework along with some hadoop ecosystems like hdfs, mapreduce, sqoop, hive and pig. By using these tools we can process no limitation of data, no data lost problem, we can get high throughput, maintence cost also very less and it is a opensouresoftware,it is compatible on all the platforms since it is Java based

MODULES

- Data Preprocessing Module
- Data Migration Module With Sqoop
- Data Analytic Module With Hive
- Data Analytic Module With Pig
- Data Analytic Module With MapReduce

Data Preprocessing Module

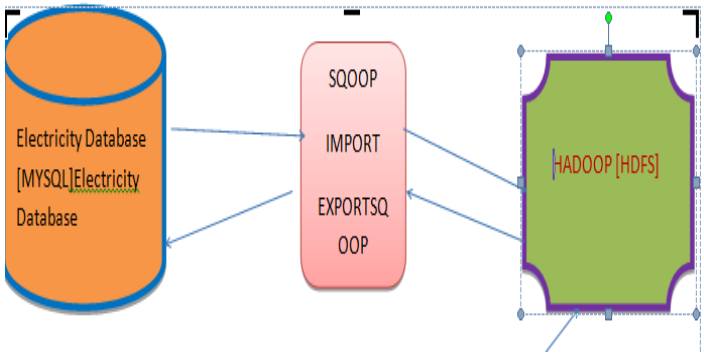


INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

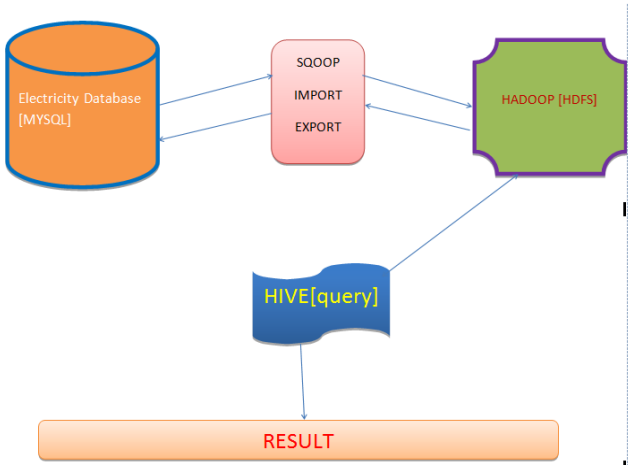
In this module we have to create Data set for Electricity Consumption it contain set of table such that customer details, billing details and payment details for last four years .and this data first provide in MySQL database with help of this dataset we analysis this project.

Data Migration Module with SQOOP

Now we are ready with dataset. So now our aim is transfer the dataset into hadoop(HDFS), that will be happen in this module. Sqoop is a command-line interface application for transferring data between relational databases and Hadoop In this module we fetch the dataset into hadoop (HDFS) using sqoop Tool. Using sqoop we have to perform lot of the function, such that if we want to fetch the particular column or if we want to fetch the dataset with specific condition that will be support by Sqoop Tool and data will be stored in hadoop (HDFS).



Data Analytic Module with HIVE



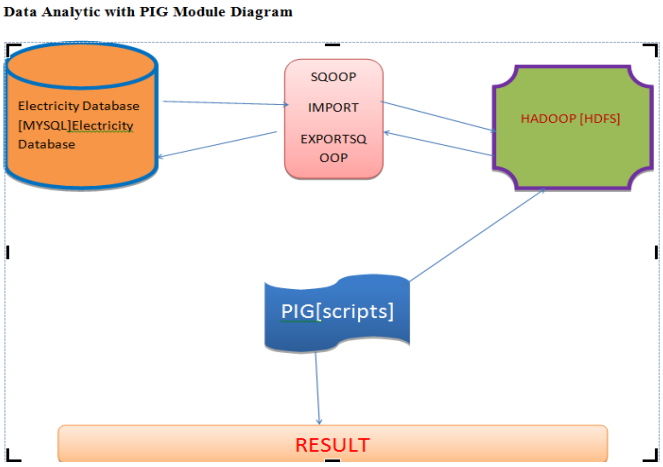
Hive is a data ware house system for Hadoop. It runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. Hive was developed by Facebook. Hive supports Data definition Language (DDL), Data Manipulation Language (DML) and user defined functions. In this module we have to analysis the dataset using HIVE tool which will be stored in hadoop (HDFS).For analysis dataset HIVE

using HQL Language. Using hive we perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language.

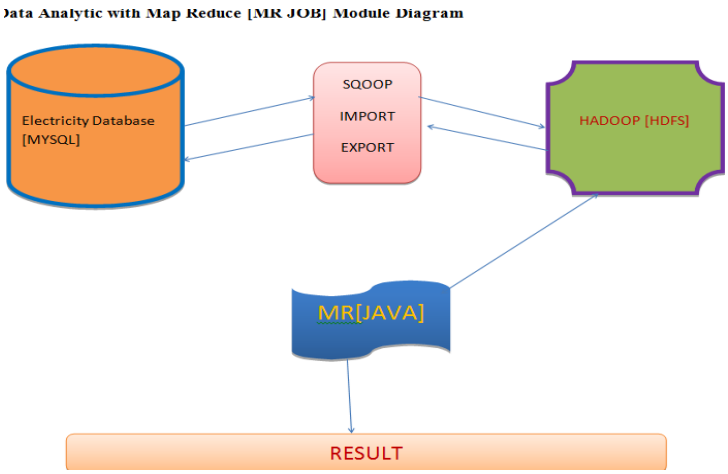
Data Analytic Module with PIG

Apache Pig is a high level data flow platform for execution Map Reduce programs of Hadoop. The language for Pig is pig Latin. Pig handles both structure and unstructured language. It is also top of the map reduce process running background.

In this module also used for analyzing the Data set through Pig using Latin Script data flow language.in this also we are doing all operators, functions and joins applying on the data see the result.



Data Analytic Module with MAPREDUCE



INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017 Nov 19, 2017

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. In this module also used for analyzing the data set using MAP REDUCE. Map Reduce Run by Java Program

4. SYSTEM TECHNIQUES

Map Reduce is a processing technique and a program model for distributed computing based on java. The Map Reduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name Map Reduce implies, the reduce task is always performed after the map job. The major advantage of Map Reduce is that it is easy to scale data processing over multiple computing nodes. Under the Map Reduce model, the data processing primitives are called mappers and reducers. Decomposing a data processing application into mappers and reducers is sometimes nontrivial. But, once we write an application in the Map Reduce form, scaling the application to run over hundreds, thousands, or even tens of thousands of machines in a cluster is merely a configuration change. This simple scalability is what has attracted many programmers to use the Map Reduce model.

A. The Algorithm

- Generally MapReduce paradigm is based on sending the computer to where the data resides!
- MapReduce program executes in three stages, namely map stage, shuffle stage, and reduce stage.

Map stage : The map or mapper's job is to process the input data. Generally the input data is in the form of file or directory and is stored in the Hadoop file system (HDFS). The input file is passed to the mapper function line by line. The mapper processes the data and creates several small chunks of data.

Reduce stage : This stage is the combination of the **Shuffle** stage and the **Reduce** stage. The Reducer's job is to process the data that comes from the mapper. After processing, it produces a new set of output, which will be stored in the HDFS.

5. REQUIREMENTS

GENERAL

These are the requirements for doing the project. Without using these tools and software's we can't do the project. So we have two requirements to do the project.
They are

1. Hardware Requirements.
2. Software Requirements.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

B. HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the system does and not how it should be implemented.

PROCESSOR : PENTIUM IV 2.6 GHz, Intel
RAM : 4GB DD RAM
MONITOR : 15" COLOR
HARD DISK : 40 GB

C. SOFTWARE REQUIREMENTS

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the team's and tracking the team's progress throughout the development activity.

Framework : Hadoop
Database : MY SQL 5.5
Lanuage : Pig, Hive, CoreJava
DataAccessTool : Sqoop
Operating System : cent os
IDE : Eclipse

6. CASE STUDY

A. *Description of the Data Set*

The data set used in this paper was provided by Research Perspective, Ltd. and contains the electricity consumption of 6,445 customers (4,511 residents, 391 industries, and 1533 unknown) over one and a half years (537 days) at a granularity of 30 minutes [36]. The whole data set consists of total 3.46 million ( ) daily load profiles. The bad load profiles are roughly identified by detecting the load profiles with missing values or all zeroes. Among these massive load data, we eliminated 6187 bad load profiles, which is a very small sample (approximately 0.18%) of the whole data set.

B. *Modeling Dynamics of Electricity Consumption for Each Customer*

According to the regular routine of electrical customers, we reasonably divide a day into four periods: Period 1 (00:00- 06:30, 22:00-24:00, overnight period), Period 2 (06:30-11:30, morning period), Period 3 (11:30-17:00, daytime period), and Period 4 (17:00-22:00, night period). On this basis, the load data are transformed into PAA representations which also vary from 0 to 1.

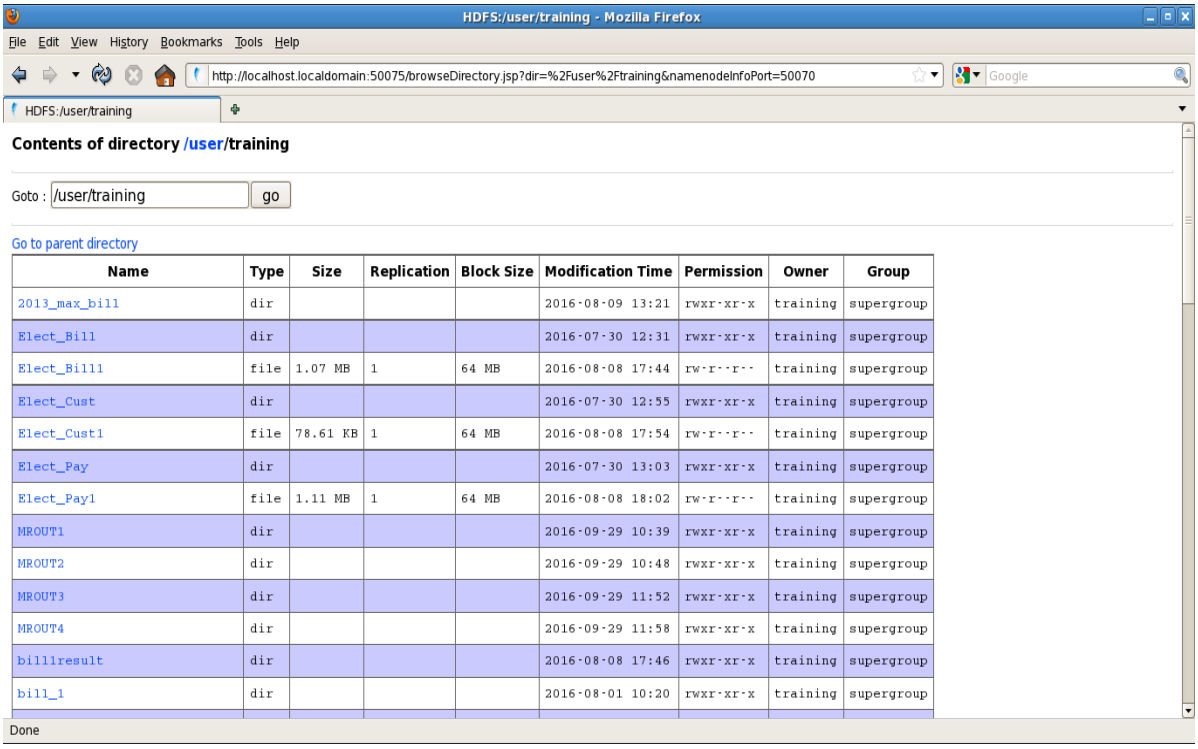
INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

7. SIMULATION RESULLT

We implement the proposed distributed clustering algorithm by HADOOP on a standard PC, with an PENTIUM IV 2.6GHz INTEL and 4.0 GB RAM. Distance calculation consumes most of the time at the global modelling stage. The overall computation time reduced greatly.

```
[training@localhost ~]$ mysql -u training -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 22
Server version: 5.0.77 Source distribution

Type 'help;' or '\h' for help. Type '\c' to clear the buffer.

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| movielens          |
| training           |
+--------------------+
3 rows in set (0.25 sec)

mysql> use movielens;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+----------------------+
| Tables_in_movielens  |
+----------------------+
| bill                 |
| cust                 |
| payment              |
+----------------------+
3 rows in set (0.00 sec)

mysql>
```

Fig : 1 Dataset Providing In The MYSQL Database



Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
2013_max_bill	dir				2016-08-09 13:21	rwxr-xr-x	training	supergroup
Elect_Bill	dir				2016-07-30 12:31	rwxr-xr-x	training	supergroup
Elect_Bill1	file	1.07 MB	1	64 MB	2016-08-08 17:44	rw-r--r--	training	supergroup
Elect_Cust	dir				2016-07-30 12:55	rwxr-xr-x	training	supergroup
Elect_Cust1	file	78.61 KB	1	64 MB	2016-08-08 17:54	rw-r--r--	training	supergroup
Elect_Pay	dir				2016-07-30 13:03	rwxr-xr-x	training	supergroup
Elect_Pay1	file	1.11 MB	1	64 MB	2016-08-08 18:02	rw-r--r--	training	supergroup
MROUT1	dir				2016-09-29 10:39	rwxr-xr-x	training	supergroup
MROUT2	dir				2016-09-29 10:48	rwxr-xr-x	training	supergroup
MROUT3	dir				2016-09-29 11:52	rwxr-xr-x	training	supergroup
MROUT4	dir				2016-09-29 11:58	rwxr-xr-x	training	supergroup
bill1result	dir				2016-08-08 17:46	rwxr-xr-x	training	supergroup
bill_1	dir				2016-08-01 10:20	rwxr-xr-x	training	supergroup

Fig : 2 All data default stored location:/in this user/training

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

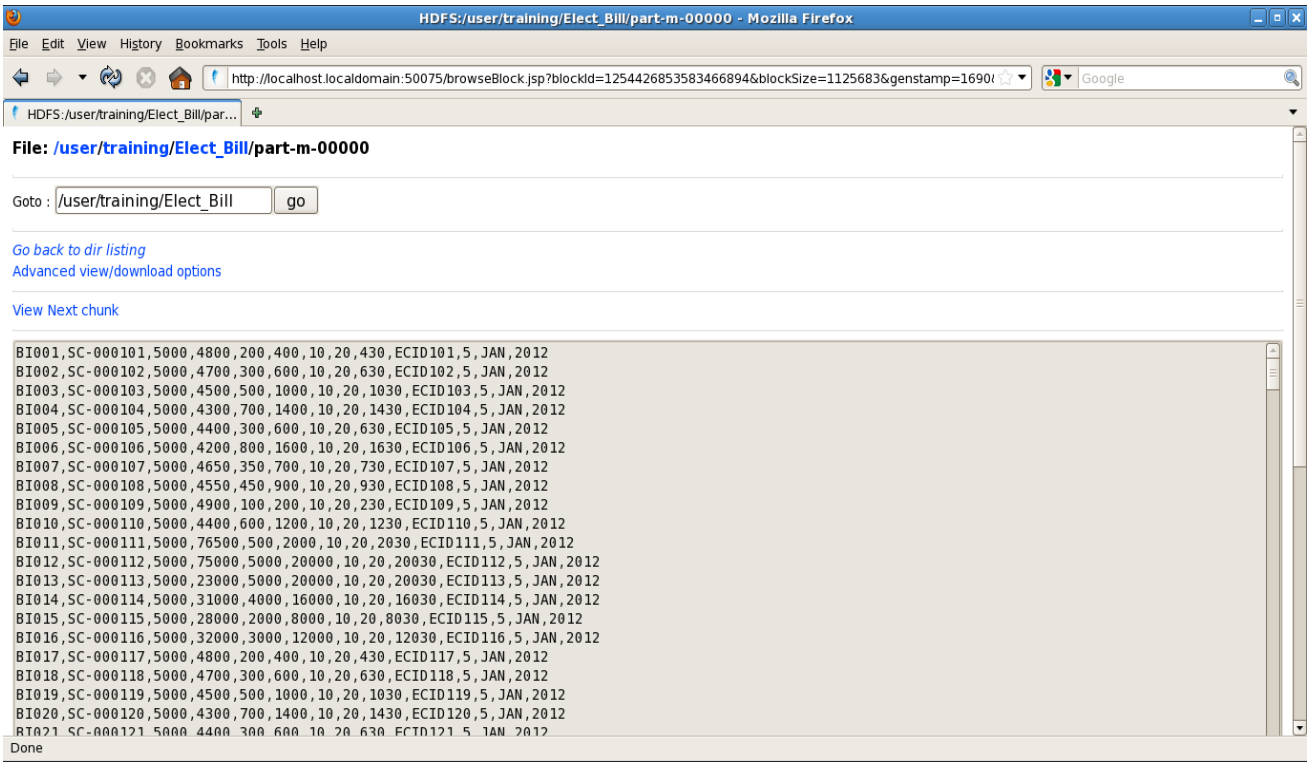


Fig : 3 SQOOP

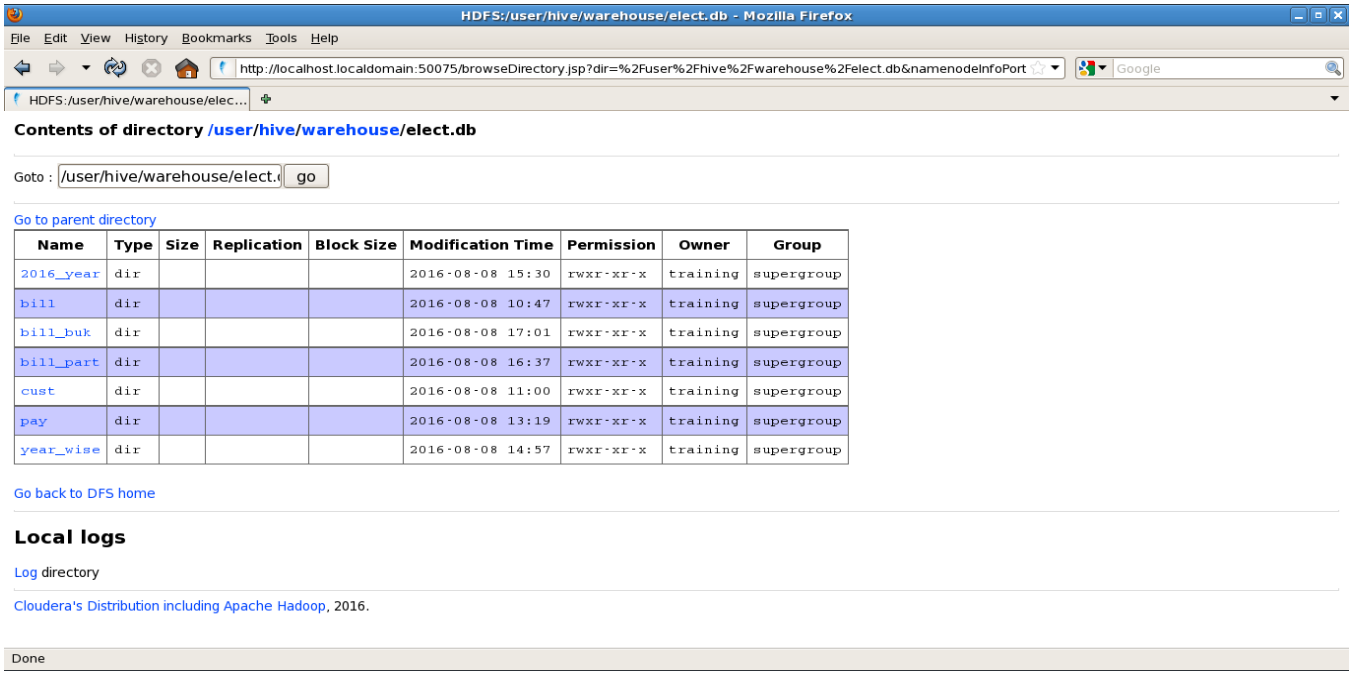


Fig : 4 HIVE

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

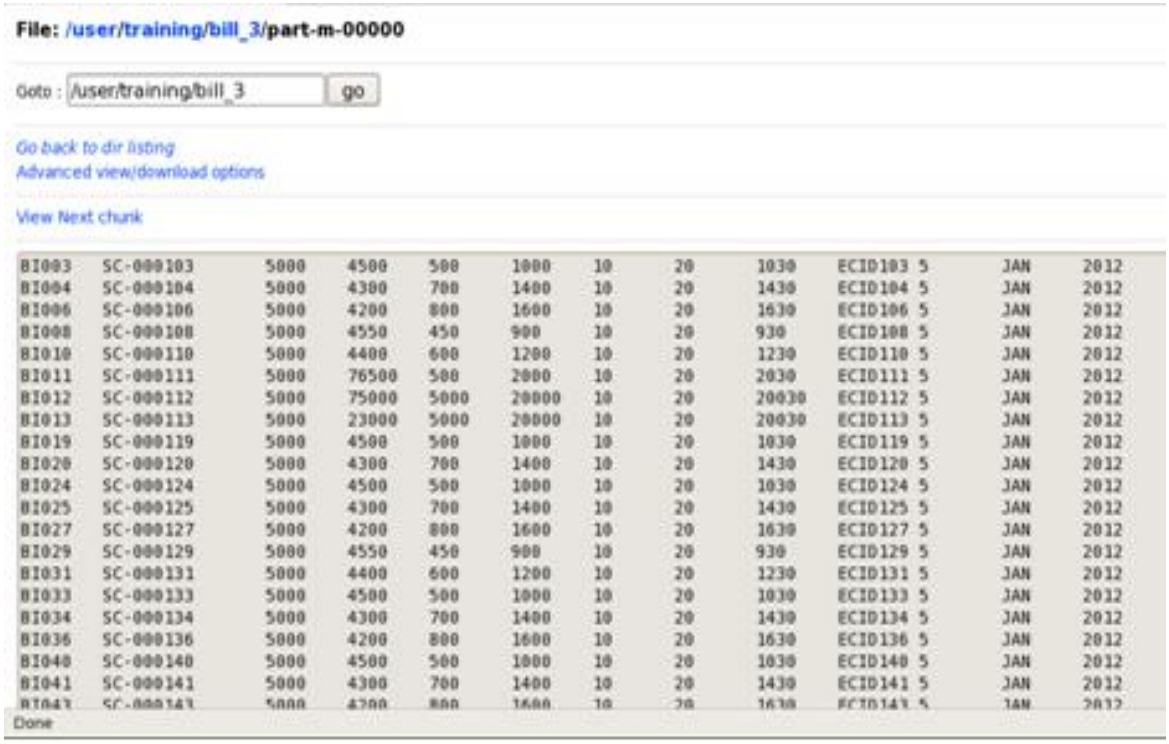


Fig : 5 PIG

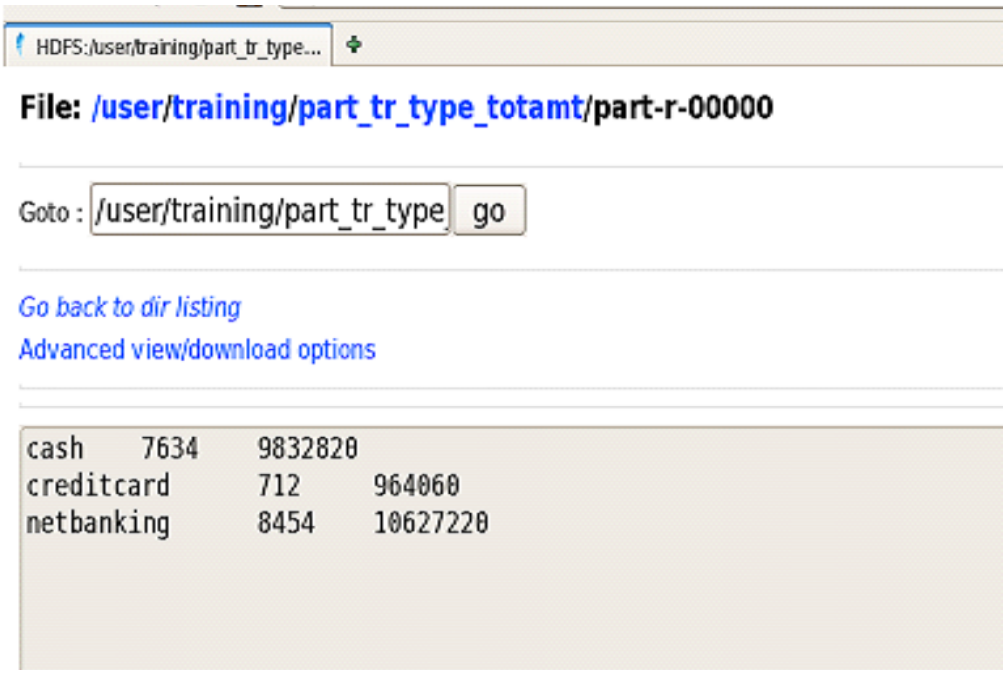


Fig : 6 Map Reduce

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 6 (2017) PAGES 2473 – 2486
RECEIVED : 25.10.2017 PUBLISHED : 19.11.2017
Nov 19, 2017

8. APPLICATION OF THE SYSTEM

Facebook using Hadoop:

At Facebook, Hadoop has traditionally been used in conjunction with Hive for storage and analysis of large data sets. Most of this analysis occurs in offline batch jobs and the emphasis has been on maximizing throughput and efficiency. These workloads typically read and write large amounts of data from disk sequentially. As such, there has been less emphasis on making Hadoop performant for random access workloads by providing low latency access to HDFS. Instead, we have used a combination of large clusters of MySQL databases and caching tiers built using memcached. In many cases, results from Hadoop are uploaded into MySQL or memcached for consumption by the web tier.

Twitter using Hadoop:

Twitter has large data storage and processing requirements, and thus we have worked to implement a set of optimized data storage and workflow solutions within Hadoop. In particular, we store all of our data LZO compressed, because the LZO compression turns out to strike a very good balance between compression ratio and speed for use in Hadoop. Hadoop jobs are generally IO-bound, and typical compression algorithms like gzip or bzip2 are so computationally intensive that jobs quickly become CPU-bound. LZO in contrast was built for speed, so you get 4-5x compression ratio while leaving the CPU available to do real work. For more discussion of LZO, complete with performance comparisons we did a while back

FUTURE ENHANCEMENTS

We are using spark we can get result hundred times faster than Hadoop. The secret is that it runs in-memory on the cluster, and that it isn't tied to Hardtop's MapReduce two-stage paradigm. This makes repeated access to the same data much **faster**. **Spark** can run as a standalone or on top of Hadoop YARN, where it can read data directly from HDFS.

CONCLUSION

To reach the 2050 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.

REFERENCES

[1] USA Department of Energy, Smart Grid / Department of Energy, http://energy.gov/oe/technology-development/smart-grid, 2014.

[2] I. P. Panapakidis, M. C. Alexiadis and G. K. Papagiannis, "Load profiling in the deregulated electricity markets: A review of the applications," in *European Energy Market (EEM), 2012 9th International Conference on the*, 2012, pp. 1-8.

[3] R. Granell, C. J. Axon and D. C. H. Wallom, "Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles," *IEEE Trans. Power Systems*, vol. 30, pp. 3217-3224, 2015.

[4] N. Mahmoudi-Kohan, M. P. Moghaddam, M. K. Sheikh-El-Eslami, and E. Shayesteh, "A three-stage strategy for optimal price offering by a retailer based on clustering techniques," *International Journal of Electrical Power & Energy Systems*, vol. 32, pp. 1135-1142, 2010.

[5] P. Zhang, X. Wu, X. Wang and S. Bi, "Short-term load forecasting based on big data technologies," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 3, pp. 59-67, 2015.

[6] N. Mahmoudi-Kohan, M. P. Moghaddam, M. K. Sheikh-El-Eslami, and S. M. Bidaki, "Improving WFA k-means technique for demand response programs applications," in *Power & Energy Society General Meeting, 2009. PES '09*. IEEE, 2009, pp. 1-5.

[7] C. Leon, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, and R. Millan, "Variability and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies," *IEEE Trans. Power Systems*, vol. 26, pp. 1798-1807, 2011.

[8] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology*, vol. 20, pp. 117-129, 2015.

[9] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of Low Voltage Network Templates-Part I: Substation Clustering and Classification," *IEEE Trans. Power Systems*, vol. 30, pp. 3036-3044, 2015.

[10] K. Zhou, S. Yang and C. Shen, "A review of electric load classification in smart grid environment," *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 103-110, 2013.

[11] G. J. Tsekouras, P. B. Kotoulas, C. D. Tsirekis, E. N. Dialynas, and N. D. Hatziargyriou, "A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers," *Electric Power Systems Research*, vol. 78, pp. 1494-1510, 2008.

[12] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, "Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps," *IEEE Trans. Power Systems*, vol. 21, pp. 1672-1682, 2006.

[13] G. Chicco and I. S. Ilie, "Support Vector Clustering of Electrical Load Pattern Data," *IEEE Trans. Power Systems*, vol. 24, pp. 1619-1628, 2009.

[14] M. Piao, H. S. Shon, J. Y. Lee, and K. H. Ryu, "Subspace Projection Method Based Clustering Analysis in Load Profiling," *IEEE Trans. Power Systems*, vol. 29, pp. 2628-2635, 2014.

[15] G. Chicco, O. Ionel and R. Porumb, "Electrical Load Pattern Grouping Based on Centroid Model with Ant Colony Clustering," *IEEE Trans. Power Systems*, vol. 28, pp. 1706-1715, 2013.

[16] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, pp. 68-80, 2012.

[17] I K. Fodor, "A Survey of Dimension Reduction Techniques," *Perpinan*, vol. 205, pp. 351-359, 2003.

[18] M. Abrahams and M. Kattenfeld, "Two-stage fuzzy clustering approach for load profiling,", in *Universities Power Engineering Conference (UPEC), 2009 Proceedings of the 44th International*. pp. 1-5, 2009.

[19] M. Koivisto, P. Heine, I. Mellin, and M. Lehtonen, "Clustering of Connection Points and Load Modeling in Distribution Systems," *IEEE Trans. Power Systems*, vol. 28, pp. 1255-1265, 2013.

[20] G. Chicco, R. Napoli and F. Piglione, "Comparisons Among Clustering Techniques for Electricity Customer Classification," *IEEE Trans. Power Systems*, vol. 21, pp. 933-940, 2006.

[21] E. D. Varga, S. F. Beretka, C. Noce, and G. Sapienza, "Robust Real-Time Load Profile Encoding and Classification Framework for Efficient Power Systems Operation," *IEEE Trans. Power Systems*, vol. 30, pp. 1897-1904, 2015.

[22] S. Zhong and K. Tam, "Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain," *IEEE Trans. Power Systems*, vol. 30, pp. 2434-2441, 2015.

[23] J. Torriti, "A review of time use models of residential electricity demand," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 265-272, 2014.

[24] Y Xiao, J Yang, H Que, "Application of Wavelet-based clustering approach to load profiling on AMI measurements," in *Electricity Distribution (CICED), 2014 China International Conference on. IEEE*, pp. 1537-1540, 2014.

[25] A Notaristefano, G Chicco, F Piglione. "Data size reduction with symbolic aggregate approximation for electrical load pattern grouping," *Generation, Transmission & Distribution, IET*, vol. 7, pp. 108-117, 2013.

[26]J Kwac, J Flora, R Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid*, vol. 5, pp. 420-430, 2014.

[27] S. Haben, C. Singleton and P. Grindrod, "Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data," *IEEE Trans. Smart Grid*, vol. 7, pp. 136-144, 2016.

[28] W. Labeeuw and G. Deconinck, "Residential Electrical Load Model Based on Mixture Model Clustering and Markov Models," *IEEE Trans. Industrial Informatics*, vol. 9, pp. 1561-1569, 2013.