

GUIDANCE FOR ELETRICITY CONSUMPTION DATASETS

Ms S.DEEPIKA Ms.S.ELAKKIYA Ms.B.MONISHA Ms. R.MONISHA

Mr .S.BASKER M.E.,(ph.d)

¹(B.E (CSE), Kongunadu College of Engg & Tech, Trichy, India, kkmilit@gmail.com)

²(Principal, Kongunadu College of Engg & Tech, Trichy, India)

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. 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 .

KEYWORDS: *Smart Grid Meters , Energy Consumption Data , Apache Hadoop, Decision Making*

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. Having a better understanding of electricity consumption patterns and realizing personalized power managements are effective ways to enhance the competitiveness of LSEs.

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. 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.

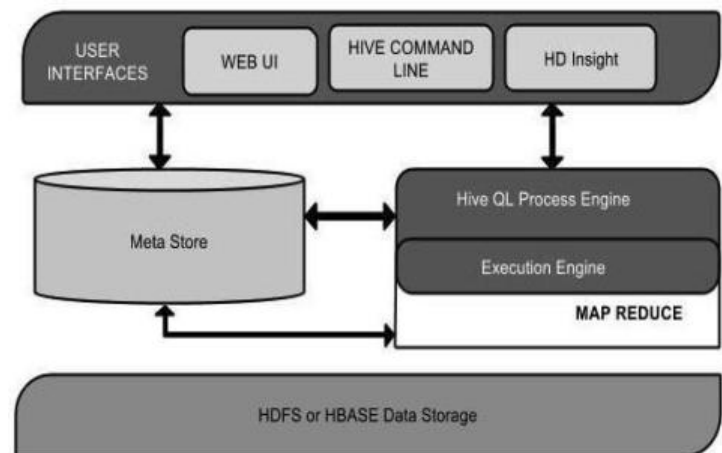


Fig 1.1: Electricity dataset analysis

2. RELATED WORK

MySQL was used as the backend in the existing system that has various drawbacks which is limitation of data where the processing large set of datas. If once the data is lost they can't be recovered .so the hadoop tool is used for

the proposed system. In the existing system, it takes more time and maintenances cost is very high.

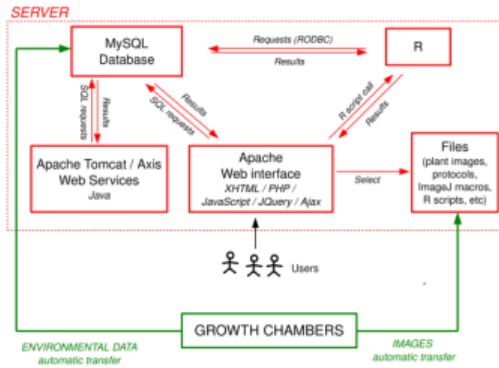


Fig 2.1: Mysql database analysis

3. PROPOSED SYSTEM

Proposed 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 and we get results with less time, high throughput and maintenance cost is very less and we are using joins, partitions and bucketing techniques in hadoop.

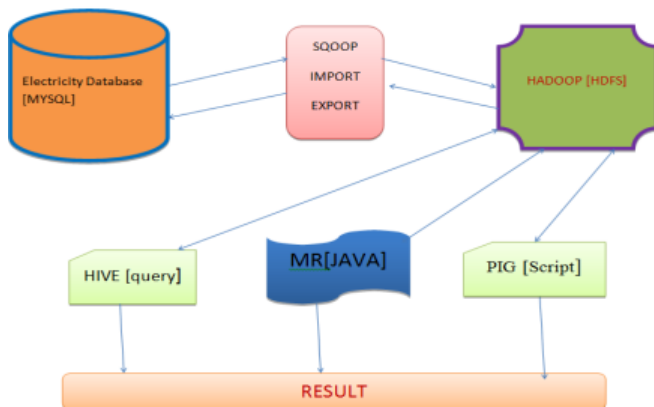


Fig3.1: Hadoop database model

3.1 MODULES

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

- Data Analytic Module With Map Reduce

3.1.1 DATA PREPROCESSING MODULE

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.

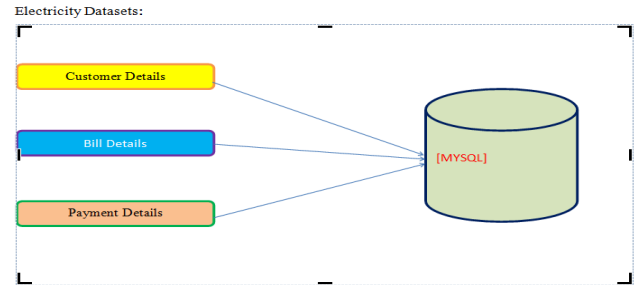


Fig 3.1.1: Preprocessing Weather dataset

3.1.2 DATA MIGRATION MODULE WITH SQOOP

This Data migration module is ready with dataset. So that the 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 the fetch dataset into hadoop (HDFS) using sqoop Tool. Using sqoop to perform lot of the function, such that if the module want to fetch the particular column or if the module want to fetch the dataset with specific condition that will be support by Sqoop Tool and data will be stored in hadoop (HDFS).

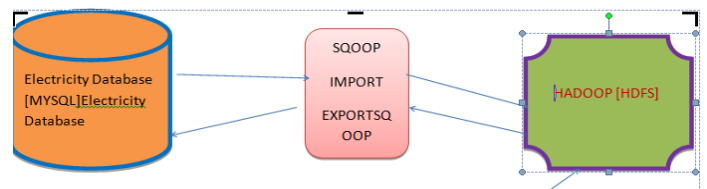


Fig3.1.2: processing dataset with sqoop

3.1.3 DATA ANALYTIC MODULE WITH HIVE

Hive is a data ware house system for Hadoop. It runs SQL like queries called HQL (Hive query language) that 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 the analysis of

dataset using HIVE tool which will be stored in hadoop (HDFS).For analysis dataset HIVE using HQL Language. Using hive to perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language.

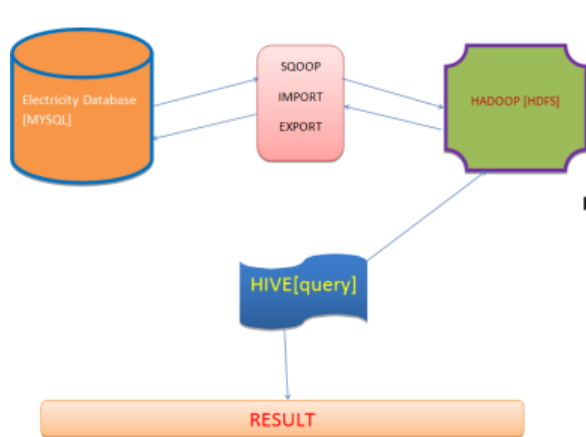


Fig3.1.3 :processing dataset with hive

3.1.4 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 all operators, functions and joins applying on the data.

Data Analytic with PIG Module Diagram

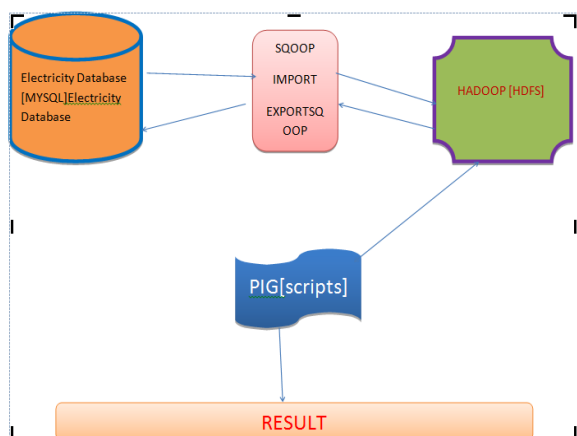


Fig3.1.4 processing dataset with pig

3.1.5 DATA ANALYTIC MODULE WITH MAPREDUCE

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

Data Analytic with Map Reduce [MR JOB] Module Diagram

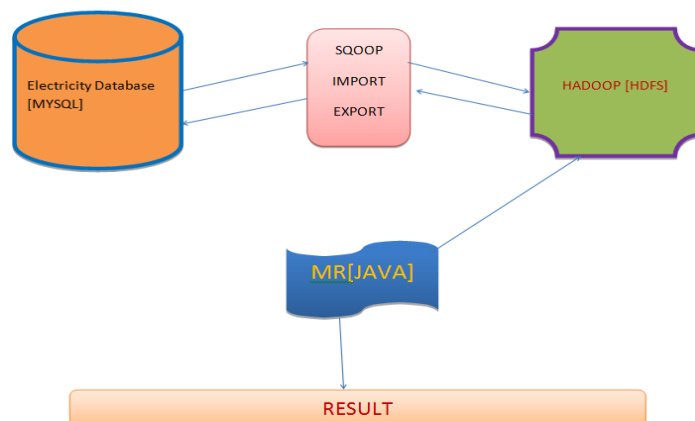


Fig3.1.5 :processing dataset with map reduce

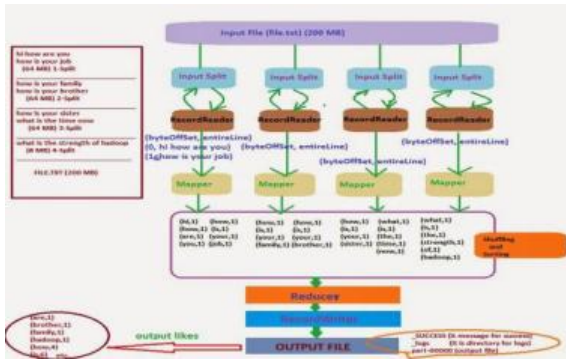
4.ALGORITHM

Commonly Map Reduce paradigm is based on sending the computer to where the data resides.

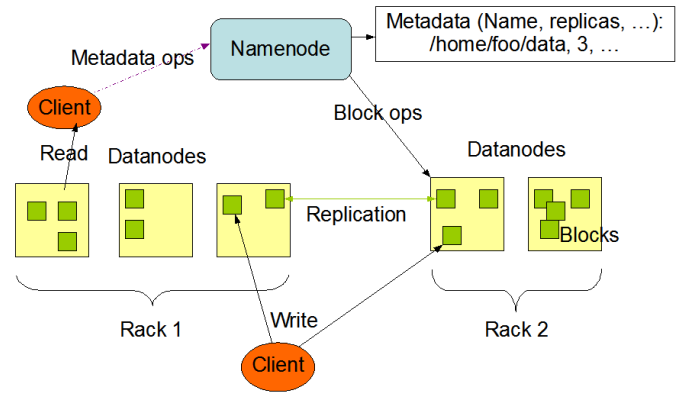
Map Reduce program executes in three stages, namely map stage, shuffle stage, and reduce stage.

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 arrangement of **Shuffle** stage and **Reduce** stage. In that Reducer’s job is to process the data that comes from the mapper. After the handing it produces a new form of output, which will be stored in the HDFS.



HDFS Architecture



5. USE CASE DIAGRAM

Use Case Diagram: System Design

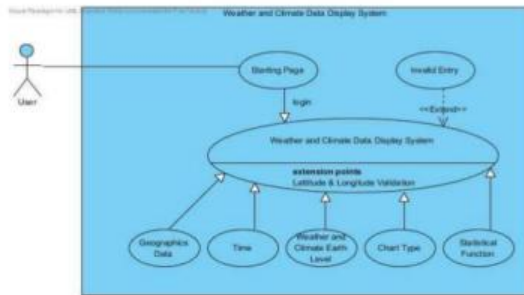


Fig 5.1: usecase model

The user has to login the page initially. After the login user will reach the page which consists of data of electricity consumption and bill. The user can get the details of geographics data, time, weather and climate earth level, chart type and statistical function.

6. HDFS

Hadoop File System was developed using distributed file system design. It runs on commodity hardware. The other distributed systems, HDFS is highly fault tolerant and it was designed using low-cost hardware. HDFS holds very excess of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion to set free the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing.

7. RESULT

7.1 Data preprocessing Module:

The Energy consumption dataset collecting customer details, billing details and payment details are represented in the module data preprocessing. That the collected datasets are accessed in form of Preprocessing with datasets of cities. In the module, Data preprocessing handles in the gathered datasets of the process.

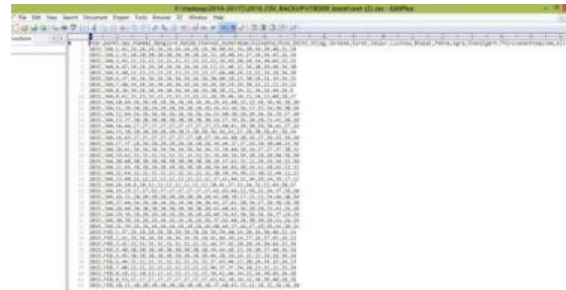


Fig 7.1.1 data processing module

7.2. TO RETRIEVE THE PARTICULAR COLUMNS

The gathered datasets are retrieved from Hadoop distributed file system Database. As per the needs of user the data is retrieved. The large amount of data is accessed and the retrieval of weather data.

File Name	Type	Size	Last Modified	Permissions	Owner	Group
h000006	dir		2016-07-21 10:24	rwxr-xr-x	training	supergroup
avg_temp_max	dir		2016-07-13 01:58	rwxr-xr-x	training	supergroup
h010006	dir		2016-07-08 23:00	rwxr-xr-x	training	supergroup
h0100062	dir		2016-07-08 23:05	rwxr-xr-x	training	supergroup
maxtemp12	dir		2016-07-11 05:23	rwxr-xr-x	training	supergroup
maxtempmax	dir		2016-07-11 04:40	rwxr-xr-x	training	supergroup
rw1	dir		2016-07-19 18:58	rwxr-xr-x	training	supergroup
subu123	dir		2016-07-08 22:50	rwxr-xr-x	training	supergroup
wecher	dir		2016-10-06 11:02	rwxr-xr-x	training	supergroup
wecher1	file	280.49 KB	1 64 MB	rw-r--r--	training	supergroup
wetresult	dir		2016-07-11 03:24	rwxr-xr-x	training	supergroup

Fig7.2.1 Retrieval of data

These data are framed and gives tabular form of output data. By using their hadoop framework tool the large set of weather data.

Year	Month	Day	Temp1	Temp2	Temp3	Temp4
2015	JAN	1	42	24	38	24
2015	JAN	2	41	28	22	20
2015	JAN	3	41	22	22	22
2015	JAN	4	43	24	12	24
2015	JAN	5	40	23	17	23
2015	JAN	6	37	34	29	34
2015	JAN	7	40	34	34	34
2015	JAN	8	36	34	30	34
2015	JAN	9	42	33	28	33
2015	JAN	10	44	34	26	34
2015	JAN	11	39	28	24	28
2015	JAN	12	44	36	22	36

7.3.TABLES CREATIONS

The Tabular form has created by the input set of data processed from the database. The output has retrieved from Tabular form .In the dataset created in the database system.

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
wet_buck	dir				2016-07-11 01:29	rwxr-xr-x	training	supergroup
wet	dir				2016-07-10 23:05	rwxr-xr-x	training	supergroup
wet_part	dir				2016-07-11 01:13	rwxr-xr-x	training	supergroup

Fig7.3.1:Table creation

2015, JAN, 1, 42, 24, 24, 24, 24, 24, 24, 24, 24, 24, 38, 48, 42, 24, 20, 26, 20, 40, 22, 28
 2015, JAN, 2, 41, 20, 20, 20, 28, 28, 28, 26, 32, 18, 46, 24, 27, 24, 34, 45, 24, 34
 2015, JAN, 3, 41, 22, 22, 22, 22, 22, 22, 22, 22, 19, 45, 20, 28, 24, 36, 44, 23, 32
 2015, JAN, 4, 43, 24, 24, 24, 24, 24, 24, 24, 12, 21, 44, 22, 24, 26, 28, 40, 34, 30
 2015, JAN, 5, 40, 23, 23, 23, 23, 23, 23, 23, 17, 44, 46, 24, 32, 22, 26, 18, 34, 38
 2015, JAN, 6, 37, 34, 34, 34, 34, 34, 34, 34, 20, 40, 18, 23, 30, 24, 24, 19, 34, 32
 2015, JAN, 7, 40, 34, 34, 34, 34, 34, 34, 34, 34, 36, 19, 34, 38, 23, 22, 21, 33, 22
 2015, JAN, 8, 36, 34, 34, 34, 34, 34, 34, 34, 38, 38, 21, 34, 32, 24, 44, 34, 12
 2015, JAN, 9, 42, 33, 33, 33, 33, 33, 33, 33, 33, 28, 39, 44, 34, 22, 34, 23, 40, 28, 17
 2015, JAN, 10, 44, 34, 34, 34, 34, 34, 34, 34, 26, 41, 40, 33, 12, 34, 34, 36, 36, 20
 2015, JAN, 11, 39, 28, 28, 28, 28, 28, 28, 24, 43, 36, 34, 17, 33, 34, 38, 30, 34
 2015, JAN, 12, 44, 36, 36, 36, 36, 36, 36, 36, 22, 40, 38, 28, 28, 34, 34, 39, 27, 30

Fig7.3.2:Retrieved dataset

The table has to be created and then the data is inserted. After the data is inserted they are processed and then they are retrieved.

8.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 Hadoop's Map Reduce 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.

9.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.

11.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.

[6] H. Gao, J. Tang, X. Hu, and H. Liu, "Content-aware point of interest recommendation on location-based social networks," in *Proc. 29th Int. Conf. AAAI, 2015*, pp. 1721-1727.

[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. Hatzigiorgiou, "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 SelfOrganizing Maps," *IEEE Trans. Power Systems*, vol. 21, pp. 1672-1682, 2006.

