

Application of big data for distribution and consumption of power

Olagunju Mukaila¹, Adeniyi Abidemi Emmanuel², Ogundokun Roseline Oluwaseun³,
Ojo Olufemi Samuel⁴, Kolawole Paul Oluwatoba⁵

¹Department of Computer Science, Faculty of Science, Federal University, Oye-Ekiti, Ekiti State, Nigeria

^{2,4}Department of Computer Science, College of Pure and Applied Sciences, Landmark University,
Omu-Aran, Kwara State, Nigeria

⁵Department of Computer Science, University of Ilorin, Ilorin, Kwara State, Nigeria

Article Info

Article history:

Received Apr 8, 2020

Revised Apr 26, 2021

Accepted May 5, 2021

Keywords:

Analytics

Big data

Electricity

Energy

Power consumption

ABSTRACT

The exponentially growing and tremendous collection of data stored in the power sector, combined with the need for data analysis, has produced an urgent need for powerful tools to extract hidden data as to effectively distribute the power for proper consumption for the household. This research work was embarked on to show the business value of big data analytics in Energy and utilities with a focus on how analytics can help solve problems of inefficiency and wastages in electricity generation, production and distribution and how raw energy datasets can be converted into insights that can be used by energy policy makers to make major business decisions. To explicitly show how raw data can be turned into insights, the study deploys the use of the Hadoop on Hortonworks' open-source apache-Hive licensed data warehousing framework run on a windows operating system to turn raw datasets (in excel formats converted to .csv format) gotten from the prepaid meters of 196,000 consumers (households and businesses) in 11 business units of Ikeja Electricity Distribution Company (IKEDC, Nigeria) to analyze the distribution and consumption of power.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Adeniyi Abidemi Emmanuel

Department of Computer Science

Landmark University

KM 4 Omu-Aran Road, Ipetu, Kwara State, Nigeria

Email: adeniyi.emmanuel@lmu.edu.ng

1. INTRODUCTION

About 2.5 quintillion bytes of data are produced on daily basis. Almost 90 percent of data in the world today have been created in the last two years alone. The produced data comes from various outlets: climate information, social media, from digital pictures and videos posted on the internet, cell phone GPS signal, and records from the purchased transactions [1]. Such colossal amount of data that is being produced continuously is what can be coined as big data. With the growth of technologies and services comes an increase in the amount of data. These types of data can be structured and unstructured from different sources that could contain billions of records of people information that can be gotten from social media, web sales, audios, images [2]. There has been numerous demand in the storage and processing of data in this twenties century. In order to support the process of this larger amount of requested data, cloud computing was developed and the implementation was used successfully in data storage and processing [3].

Big data is a concept which could be used in any industry and a vast amount of data could be used to ones personal profit, but the focus here is the Nigerian Electricity Distribution Companies (Discos), which

provides supply value chain services in electricity distribution. Discos are the known as the middle man between the customers and the electricity grid. Navigant Research reported that the smart meters installed worldwide will surpass 1.1 billion estimated by the year 2022 [4].

The analytics of big data is the process of extracting and analyze useful information in order to uncover the hidden patterns of data through the use of advanced techniques mostly data mining and statistical method to fetch hidden patterns [5]. Big data has grown too large and this makes it difficult to work with using traditional database management system [6], [7]. Business analytics is one of the new technologies that is characterized with low risks and quick paybacks which can help the organizations to improves and understand their business and leverage opportunities presented by abundant data domain specific analytics [8]. When big data are empty, they lacking value or they are being useless. The prospective quality of big data is to act effectively for decision making, in order to activate this decision making process, the organizations will need an efficient method to turn high volumes of diverse data into meaningful perceptivity. There was widespread awareness of the value of enhancing building energy efficiency to save energy and improve building sustainability. One successful way to achieve this aim is to discover and derive useful information from the construction of operational data.

As the technologies keep increasing, smart grids distribution also increases which serve as the moving forces for the approval of big data analytics. These new technologies undergo development which made better quality to capture electricity used by the customers at any time of the day from the meter [9]. There are still campaigns in many countries about the adoption of smart grids technologies. The European Union member states have fully initiated the installation of the intelligent meters, as a way to improve the efficiency of the energy system and their target is to reach 80 percent roll out by the year 2020. The development of the smart grid and meter infrastructures intelligent system gave birth to new level of data for electric power information technology and business leaders. This makes big data analytics extend to many areas of technology and business intelligence [8].

Data Analysing using smart grid makes it possible to identify clusters that have excessive electrical load, clusters that have high power outage frequencies lines with high failure probability. As a result, it is possible for example to identify grid upgrades, transformations and maintenance and to effectively forecast energy management [10], [11]. An ideal power grid continually balances power generation and consumption for grid stability. The traditional power grid is based on one-directional approach from concentrated electrical power generation to the grid distribution by the transmission line, which does not allow the reverse flow. Currently, the boom of renewable energy, local energy production (from the original consumer called a “pro-consumer”), electrical mobility with rechargeable battery systems, energy storage and many other applications have forced the decentralization of the power system. In detail, the power grid system is evolving from a concentrated power plant to a micro one, where each local consumer may become a producer by, for example, photovoltaic panels, installed at home, storage, and fuel cell application. In this new scenario, the electronic meter is the gateway that connects the consumer/pro-consumer to the power grid. It is a key device for big data: data are frequently and rapidly acquired and subsequently analyzed by the power utilities [11], [12].

Since 2000, the Enel Company has replaced 38 million analogic meters with electronic meters, becoming the first power utility in the world ready for the smart grid application. The increasing number of distributed power stations, such as wind farms, concentrated heat power units, mini hydro plants and photovoltaic systems creates a virtual power plant (VPP). This new system can replace a conventional power plant, while providing more flexibility and higher efficiency. However, VPP is a complex system that requires difficult control optimization and secure communication. Big data analysis may help resolve problems and increase the reliability of the system.

Currently, the wind and solar energy resources can be connected to the power grids, since there is closely relationship between the capacities of power generation of new energy resources and the feature randomness intermittency of climate conditions. The intermittent renewable new energy sources can be efficiently managed if only the big data of power grids in effectively analysed. This will help the new energy resources generated to be allocated to the region with shortage of electricity [13]-[15].

Several authors have worked on both daily and seasonal trends in energy consumption, but due to the geographical region, most of their studies focused on consumption patterns within a specific country. These authors [16], worked on electricity demand in India which based on aggregate macro data at both national and state level. Their work uses econometric analysis to determine the income in consumption of electricity and the relationship between the consumption and gross domestic product per capital and price of electricity over a given period of time.

Yigzaw and Yohanis [17] presented on residential houses billing system in Hong Kong and the United Kingdom respectively. Consumption of energy patterns were considered as basic units of analysis. In [18] proposed MapReduces and apache Hadoop big data techniques was used by this author to analyze and

generate insights for data of energy collected in order to improve the energy efficiency. He also used the energy data collected to evaluate different prediction models and forecast future consumption on the basis of previous energy consumptions.

Whaley and Saman [19] uses smart meters as means to gather various household daily energy consumptions. The household energy consumption was done through segmentation of electric appliances and new technologies for heat generation. This makes it easy to identify consumers with similar needs and behaviors. The major key role in the evolution of smart grids is the application of big data analysis in order to assist in extraction as well as to analyze only the efficient energy consumption from the smart grids so there will be efficient management and discovering of hidden energy consumptions [20]-[23].

2. RESEARCH METHOD

The material used in this research is the data of about 196,000 homes and businesses that are customers of Ikeja Electric, one prepaid meter per entity. The smart meter generated information every day for a month, and there is one year of data available. All data is available in Excel and was later converted to CSV format. This primary data was obtained via the monitoring department of IKEDC's database and also secondary data via the information gathered from recent whitepapers, research materials, and prepared texts about big data analytics on the cloud and Nigerian electric power distribution. Hive as a data warehousing tool and Hadoop file system (HDFS) can ingest various structured, unstructured and semi-structured datasets, but for the purpose of this study and due to its peculiarities viz-a-via data collection device (smart meter).

Table 1. Data fields description

S/N	Attribute-Name	Description
1	Feeder	a feeder line is part of an electric distribution network, usually a radial circuit of intermediate voltage.
2	Undertaking Office (UT)	A field office of Ikeja Electric for distribution of electric power within a Business unit.
3	Business Unit (BU)	The various business divisions of the distribution company.
4	Tariff Plan	The tariff plan a customer belong to.
5	Type	The type metering device, ie AMR or AML.
6	Customer numbers	The unique identifier for each consumer.
7	Monthly Consumption (October 2015 through October 2016)	Electric power consumption for each month in kilowatts per hour.
8	Account Numbers	Customers account number for recharging their prepaid meters
9	Address	Different households or offices with the installed prepaid meter.
10	Average	Average power consumption for each month.

The apache Hive information warehouse software encourages questioning and overseeing substantial datasets link in the case of distributed computing Hive is a valuable asset for ELT, Hadoop knowledge inventory control, and Hadoop repository. Conversely, as opposed to conventional repositories, it is comparatively sluggish. It does not have any of the structured query language (SQL) functionality or even any of the database functions that standard repositories do. However, it embraces SQL, functions as a repository, and provides more users with access to Hadoop technologies (even those who are not programmers). It provides a method for converting unorganized and semi-organized data into functional template data. If you want to create a master data processing system? Hive allows you to do this. Do you want to develop a data storage facility? You can do the same thing with Hive, but you'll need to learn the techniques to make Hive an effective weapon for ELT tool [24-27]. As shown in Figure 1, the transformation process basically involves putting the datasets in the most appropriate format suitable for analysis. This entails the creation of a database, temporary table, with which the data is transferred from HDFS to Hive metastore, Thereafter, creating an appropriate schema to represent the needed fields and records most suitable for querying.

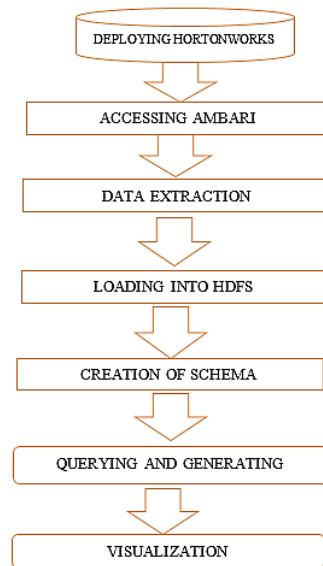


Figure 1. Phases involved in generating results of analysis

3. RESULTS AND ANALYSIS

The Hive tool used in carrying out this research is open-source apache-Hive licensed data warehousing frameworks which can either be configured or pre-configured. A pre-configured machine was used in this study, owned by Hortonworks. It's the vendor's Hadoop distribution. A Hortonworks Data Platform hosted on Microsoft azure platform was used for this analysis due to its low system requirements, cost, and ease of usage. The language used in this analysis is the HiveQL which is a querying language used in the Hive data warehousing environment, it's similar to the SQL and MySQL languages. Figure 2 displays a table for running queries on data.

```

1 CREATE TABLE IF NOT EXISTS smart_meter (
2
3 Customer_no STRING COMMENT 'Customers meter number',
4
5 Tariff_Plan STRING COMMENT 'Individual tariff plans ',
6
7 Address STRING COMMENT 'Home address',
8
9 UT STRING COMMENT 'Nearest Undertaken office',
10
11 DT STRING COMMENT 'Location of distribution Transformer',
12
13 Feeder STRING COMMENT 'Electricity Lines',
14
15 BU STRING COMMENT 'Business Units',
16
17 Type STRING COMMENT 'AMR OR AMI meter',
18
19 Oct_2015 INT COMMENT 'Total power consumption for the month',
20 Nov_2015 INT,
21 Dec_2015 INT,
22 Jan_2016 INT,
23 Jan_2016 INT,
24 Feb_2016 INT,
25 Mar_2016 INT,
26 Apr_2016 INT,
27 May_2016 INT,
28 Jun_2016 INT,
29 Jul_2016 INT,
30 Aug_2016 INT,
31 Sept_2016 INT,
32 Oct_2016 INT,
33 Average INT
34 )
  
```

Figure 2. Creating a table for running queries on data

This is the view of the pre-process panel after the dataset has been imported. Once the data is loaded, Hive recognizes attributes that are shown in the 'attributes box' at the left corner of pre-process panel which shows the list of recognized attributes. The 'Table Name' window above 'attribute box' displays the fields name and type. By clicking on the rightmost icon on the table name, the data present on the table are displayed in the result section. Also, it displays the minimum, maximum, mean and standard deviation of the selected attribute. The most significant power distribution equipment is the distribution transformers and feeders. This analysis seeks to explore the busyness or otherwise of this equipment for effective decision

making. Figure 3 shows the view in which the datasets are been loaded to their permanent location by matching column values of the datasets loaded into the temporary table.

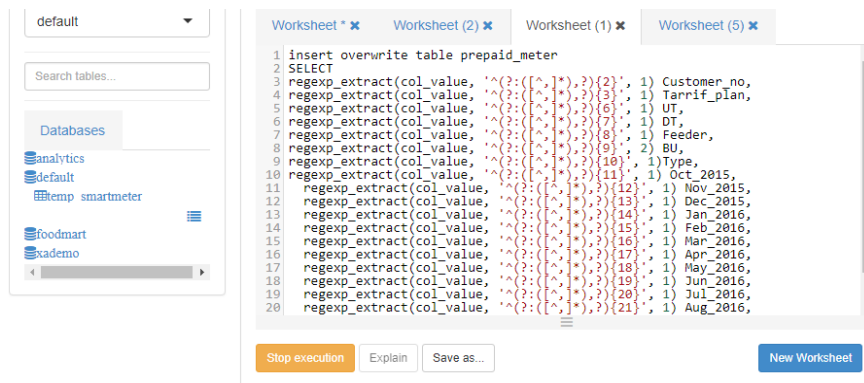


Figure 3. Putting extracted data to the table

- Distribution transformers and feeders

The analysis depicted in the graphs below, clearly shows the transformers that have load more and less than the average (84KWH) for a month, this clearly shows the distribution transformer that needs load reduction in the different localities. The Figure 4 shows the average monthly load of exceeding transformer while Figure 5 shows the below transformer average monthly load. The Figure 6 shows the feeders or electric lines that carry optimum loads, overloaded and perhaps needs prompt replacement or more lines to be added to the locality.

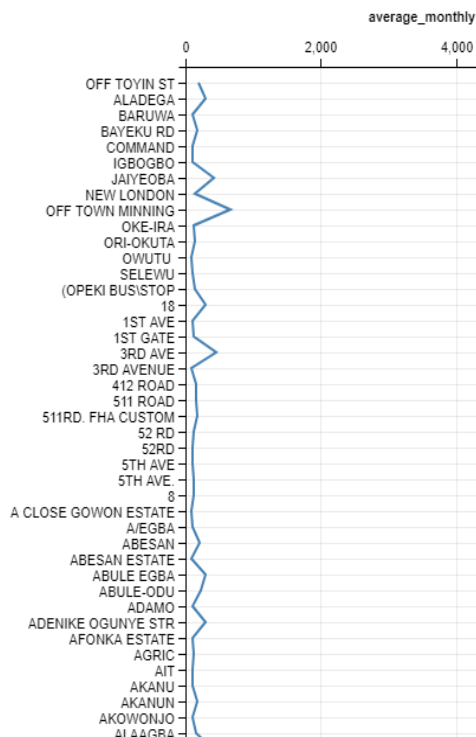


Figure 4. Transformers exceeding the average monthly load

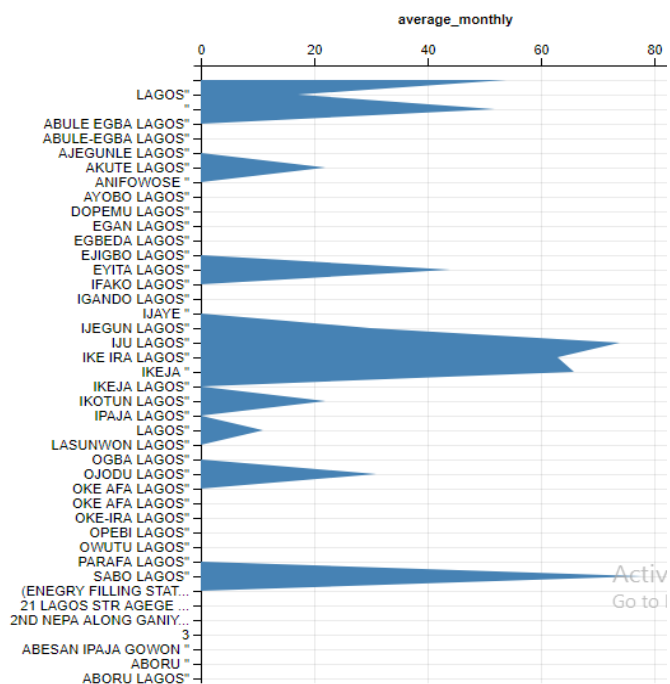


Figure 5. Transformers below the average monthly load

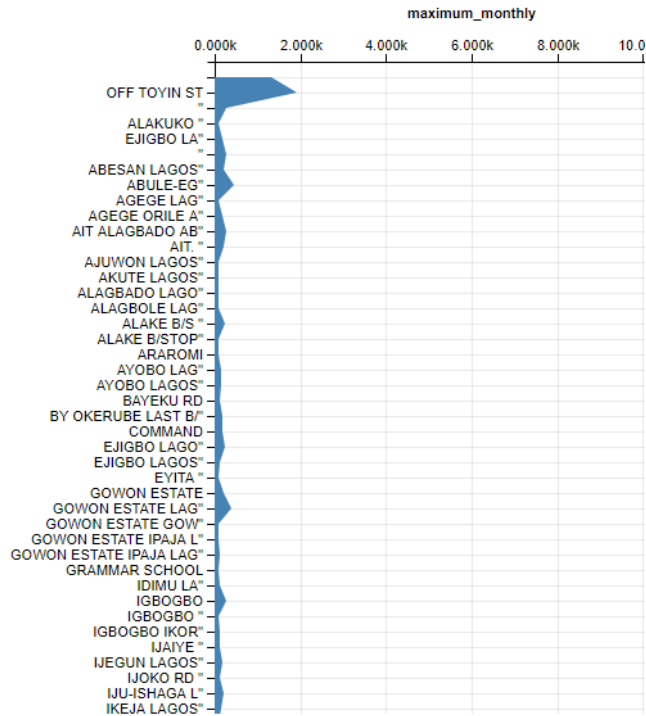


Figure 6. Feeders exceeding the average monthly load

- Effects of climatic conditions on consumption

Figure 7 shows rainfall statistics for the year 2016 in Lagos, it is used to highlight the months with the most and least rainfall. Thereafter a query was run for the average rainfall for those months and the results shown in Figure 8. Figure 9 shows the distribution of months and the corresponding sun intensity, while Figure 10 shows the average power consumption in these months. Figure 11 displayed the measure of the level of non-consumption of power up to the average quota and also the computed standard deviation of the consumption for the year.

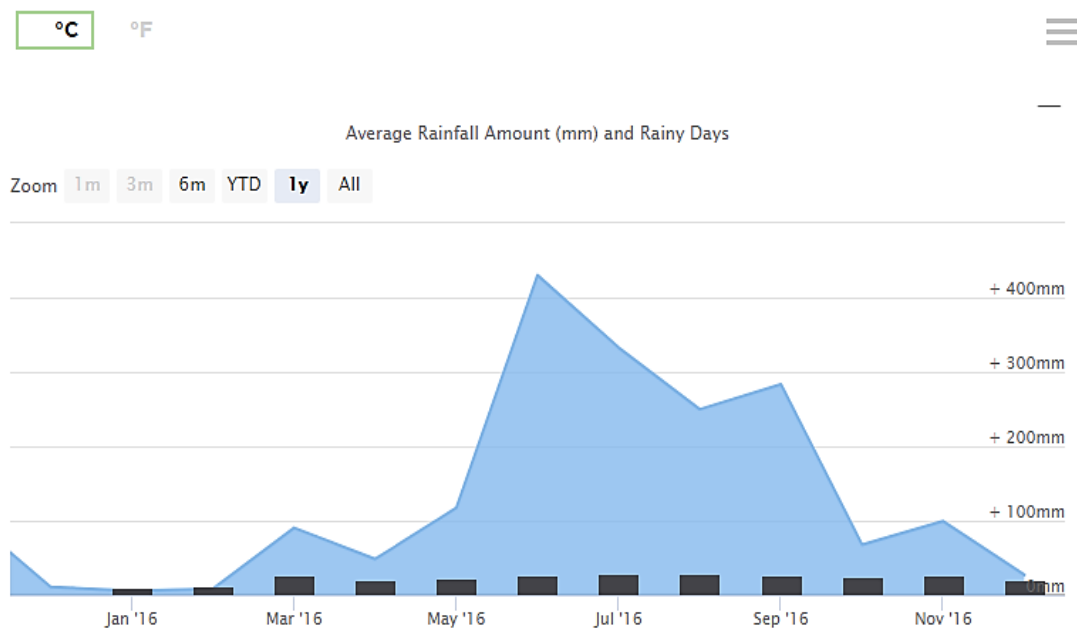


Figure 7. Lagos rainfall statistics for 2016 [28]

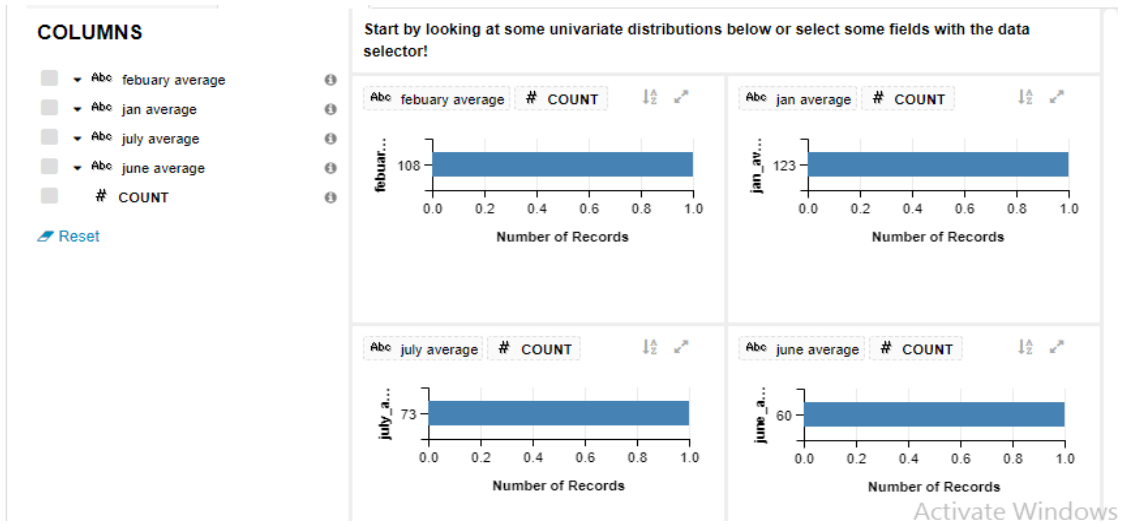


Figure 8. Power consumption in months with high and low rainfall

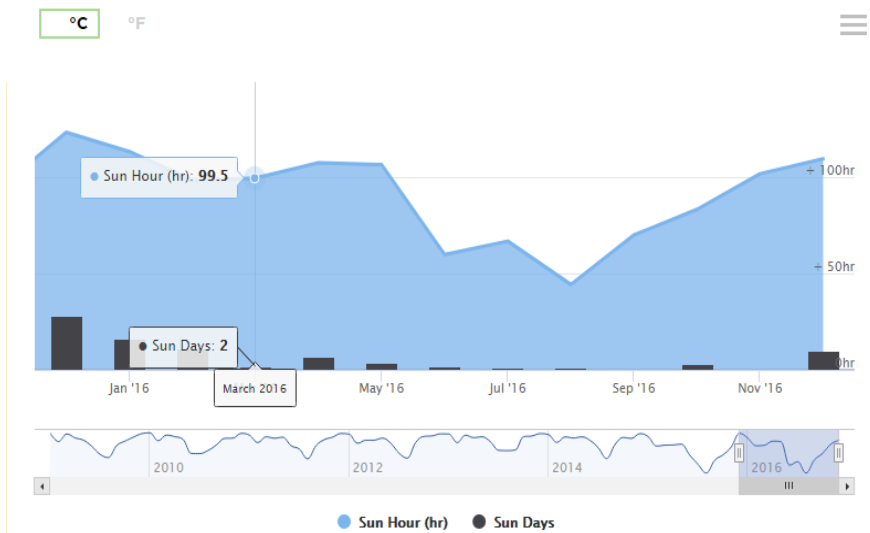


Figure 9. Sunshine stats for Lagos in 2015/2016 [28]

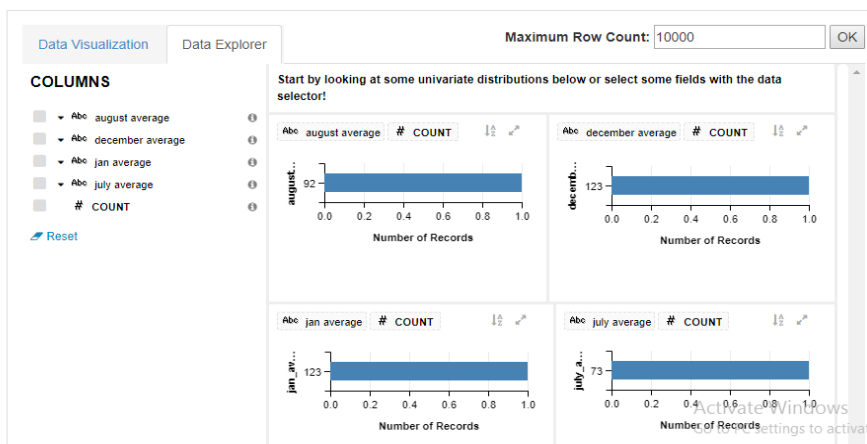


Figure 10. Sunshine stats for Lagos in 2015/2016

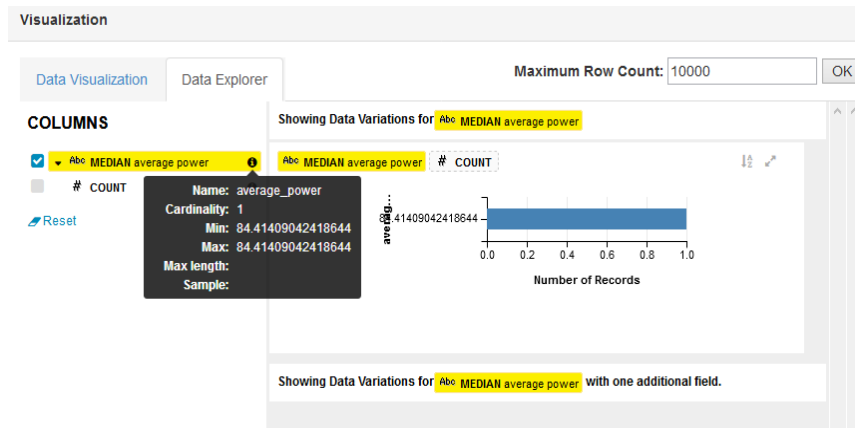


Figure 11. Standard deviation for consumption in a year

- Degree of deviation from the average
- Interpretation of results
 - a. During the course of analysis, it was discovered that the total power consumed for the location under consideration is 7.6 GW (i.e. 7.6 billion watts), by about 196,000 consumers, this insight could be useful for load request and estimation from the transmission companies. The average consumption for a household in a month is 84KW/H, this value can be used to support cost analysis and projected revenue figures for a year.
 - b. The equipment analysis provides knowledge in form of load analysis for the distribution equipment; it clearly shows there're lots of electric lines with more than optimum electric traffic or throughput. Also, the graphs show the transformers that are typically overloaded (about 30% of them) clearly need their loads reduced.
 - c. Analysis resulting from the effects of climatic conditions on power availability opposes the norm that there's more power available or consumed during periods of high rainfall. The results show that there's twice as much power consumed in January (period of low rainfall), compared to June (a typical month with high rainfall) sunshine.
 - d. The standard deviation value (0.64) shows that 60% of households did not consume up to the average 84kw/h which shows the level of non-availability of power across households in the 11 business divisions.

4. CONCLUSION

There are several other big data analytics methods of deriving insights from large data pools; Hive clearly provides an easy and convenient way of doing this. This study shows successful deployment of big data technologies in the cloud, which depicts the interoperation of two disruptive technologies (cloud meets big data). The Hive query language provide a convenient way of querying both structured and unstructured data, while offering programmers or analyst with prior SQL knowledge an easy way to process MapReduce jobs. Finally, this study has shown that there are enormous useful insights or business intelligence that can be gotten from inundated large datasets using Hadoop tools and most conveniently hosted in the cloud. This is so because of the open source nature of this technology and hence it is inexpensive to deploy. The only bottleneck is the expertise required to carry out such important task. A further research can be done in the area of AMI data analytics to get power consumption insights on an hourly basis, how big data analytics can be used to curb energy/electricity data theft, analytics for energy and Utilities management.

ACKNOWLEDGEMENTS

The authors' appreciation goes to Landmark University Centre for Research and Development (LUCRED), Landmark University, Omu-Aran, Kwara State, Nigeria for funding this reseach and for PDF sponsorship.

REFERENCES

- [1] J. Camacho-Rodríguez, *et al.*, "Apache hive: From mapreduce to enterprise-grade big data warehousing," *In Proceedings of the 2019 International Conference on Management of Data*, pp. 1773-1786, 2019, doi: 10.1145/3299869.3314045
- [2] S. Sagioglu and D. Sinanc, "Big data: A review," *In 2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, USA, 2013, pp. 42-47, doi: 10.1109/CTS.2013.6567202.
- [3] I. A. T. Hasem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Ghani, and S. U. Khan, "The rise of "big data" on cloud computing, Review and Open research work. *Information Systems*," 2014. Review and open research issues," *Information systems*, vol. 47, pp. 98-115, 2015, doi: 10.1016/j.is.2014.07.006.
- [4] N. Research, "Smart gas meter penetration," 2016. [Online]. Available at: www.pikersearch.com/newsroom/smart-gas-meter-penetration-to-reach-11-by-2016 2016
- [5] R. Russom, "Big Data Analytics," *TDWI best practices report*, fourth quarter, pp. 1-34, 2011.
- [6] A. Elgend and P. Elragal, "ERP and Big Data: The Inept Technology," *Procedia Technology*, vol. 16, pp. 242-249, 2014, doi: 10.1016/j.protcy.2014.10.089.
- [7] M. Abiodun *et al.*, "Cloud and Big Data: A Mutual Benefit for Organization Development," *Journal of Physics: Conference Series*, vol. 1767, no. 1, pp. 1-11, 2021, doi: 10.1088/1742-6596/1767/1/012020.
- [8] C. Chiang and C. Storey, "Business Intelligence and Analytics: from Big Data to Big Impact," *MIS Quarterly*, vol. 36, pp. 1165-1188, 2012, doi: 10.2307/41703503.
- [9] G. Mom, "The Electric Vehicle: Technology and Expectations in the Automobile Age," Baltimore: *Johns Hopkins University Press*, pp. 548-551, 2004.
- [10] V. Simonyan, and R. Mazumder, "High-performance integrated virtual environment (Hive) tools and applications for big data analysis," *Genes*, vol. 5, no. 4, pp. 957-981, 2014, doi: 10.3390/genes5040957.
- [11] A. Kazeem *et al.*, "Applicability of MMRR load balancing algorithm in cloud computing," *International Journal of Computer Mathematics: Computer Systems Theory*, vol. 6, no. 1, pp. 7-20, 2021, doi: 10.1080/23799927.2020.1854864.
- [12] V. Hayn and W. Bertsch, "Electricity load Profiles in Europe: The importance of household segmentation," *Energy Research and Social Science*, vol. 3, pp. 30-45, 2014, doi: 10.1016/j.erss.2014.07.002.
- [13] G. Amir and H. Murtaza, "Beyond the types, Big Data Concepts methods and analytics," *International journal of information management*, vol. 35, no 2, pp. 137-144, 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.
- [14] IBM Developers World "Retrieved from Hive as an alternative ELT tool," 2017. [Online]. Available at: <https://www.ibm.com/developerworks/library/bd-hivetool/index.html>
- [15] A. Ahmed, "Applying Big Data Analytics for Energy Efficiency," *International Design Business Management*, pp 17-35, 2014.
- [16] R. Bose and M. Shukla, "Elasticity of electricity demand in India," *Energy Policy*, vol. 27, no. 3, pp. 137-146, 1999, doi: 10.1016/S0301-4215(99)00011-7.
- [17] Y. Yizaw and F. Yohanis, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data," *Mcgraw-hill Osborne Media*, 2011.
- [18] H. Ahmed, "Applying Big Data Analytics for Energy Efficiency" 2014. *Master's Degree Programme in International Design Business Management (IDBM)*. vol. 08, no. 29, pp. 2550, 2014. [Online]. Available at: https://aaltodoc.aalto.fi/doc_public/printed_theses/?la=en.
- [19] L. Whaley and K. Saman, "Electricity Demand Profile of Australia Low Energy Houses," *Australia: Energy Procedia*, vol. 62, pp. 91-100, 2014, doi: 10.1016/j.egypro.2014.12.370.
- [20] A. Gandomi and M. Haider, "Beyond the types, Big Data Concepts methods and analytics," *International journal of information management*, vol. 35, no. 2, pp. 137-144, 2014, doi: 10.1016/j.ijinfomgt.2014.10.007.
- [21] K. Bakshi, "Considerations for big data: Architecture and approach," *2012 IEEE Aerospace Conference*, 2012, pp. 1-7.
- [22] E. Battre *et al.*, "A programming model and execution framework for web-scale analytical processing," *Proceedings of the 1st Symposium on Cloud Computing.*, NewYork. 2010, pp. 119-30, doi: 10.1145/1807128.1807148.
- [23] Big commerce, "Retrieved from Big commerce support," 2017. [Online]. Available at: <https://support.bigcommerce.com/articles/Public/What-is-a-CSV-file-and-how-do-I-save-my-spreadsheet-as-one>
- [24] F. Shariatzadeh *et al.*, "Demand response for sustainable energy systems: A review, application and implementation strategy," *Renewable and Sustainable Energy Reviews*, vol. 45, pp. 343-350, 2015, doi: 10.1016/j.rser.2015.01.062.
- [25] L. Dias *et al.*, "City energy modelling-Optimising local low carbon transitions with household budget constraints," *Energy Strategy Reviews*, vol. 26, pp. 100-387, 2019, doi: 10.1016/j.esr.2019.100387.
- [26] S. Hashem, I. Yaoob, N. B. Anuar, and S. Mokhtar, "The rise of "big data" on cloud computing, Review and Open research work". *Information Systems*, vol. 47, pp. 98-115, 2014, doi:10.1016/j.is.2014.07.006.
- [27] M. Institute, "Curbing Global energy demand growth: The energy productivity Oppurtunity," *Mckinsey Global institute*, pp. 7-9, 2007.
- [28] Climate FieldView. [Online]. Available: [weatherstats.com \(https://climate.com/\)](https://weatherstats.com).

BIOGRAPHIES OF AUTHORS

Olagunju Mukaila is a Senior Lecturer in the Department of Computer Science, Faculty of Science, Federal University, Oye-Ekiti, Ekiti State. His area of research interest is software visualization and modelling.



Adeniyi Abidemi Emmanuel is a Staff in the Department of Computer Science, College of Pure and Applied Sciences, Landmark University, Omu-Aran, Nigeria. He received Bachelor of Science in Computer Science from Ladoke Akintola University of Technology, Ogbomoso. His area of research includes computer security and data mining.



Ogundokun Roseline Oluwaseun is a Lecturer at the Department of Computer Science, College of Pure and Applied Sciences, Landmark University, Omu Aran, Kwara State, Nigeria, she holds Bachelor of Science in Management Information System from Covenant University, Ota; Master of Science in Computer Science from the University of Ilorin, Ilorin; Post Graduate Diploma in Education (PGDE) from the National Teachers' Institute (NTI), Kaduna and; currently a PhD student in the Department of Computer Science, University of Ilorin, Ilorin. Her research interests include machine learning, deep learning, steganography and cryptography, information security, artificial intelligence, data mining, information science and human computer interaction.



Ojo Olufemi Samuel is a former E-Learning Administrator at Landmark University Nigeria. He is currently a lecturer at Ajayi Crowther University, Oyo, Nigeria.



Kolawole Paul Oluwatoba pursued Bachelor of technology in computer Engineering from Ladoke Akintola University of technology, Ogbomoso, oyo state in 2013 and Masters of Science in university of Ilorin. Currently he looks forward to apply and pursue PhD in Department of computer Science University of Ilorin. He is currently a researcher.