Leveraging Big Data and IoT technology into Smart Homes

Sana Ramzan*

* Instructor, Accounting Department. Algonquin College, Kuwait

DOI: 10.29322/IJSRP.10.09.2020.p10545
http://dx.doi.org/10.29322/IJSRP.10.09.2020.p10545

Abstract- The energy sector generates huge amounts of data that can be used by utility providers in energy conservation. As data is increasing drastically, the conventional energy meter is not sufficient to keep track of this data that incorporates demand and supply patterns of energy. The conventional energy meter manually retrieved information from a residential customer or an industrial customer and manually billed the consumption. This process was tedious, time consuming and costly for the utility providers. Subsequently, this led to the adoption of smart grid systems for energy management and efficient data processing. Another important facet of this research is the incorporation of residential sector, which is the highest consumer of energy and generates humongous amounts of data. Hence, to understand the data transmission process in a residential sector, it is necessary to understand the data transmission process in one household automation system that consists of a smart meter system and does not contain the conventional energy meter. The large volume of data, popularly known as energy big data, generated from smart meters holds valuable insights that can be utilized by the utility companies and home owners to make strategic control decisions in terms of power generation, consumption, and monitoring. In this research paper, a smart home lifecycle is demonstrated with a clear description regarding the progression of storing, processing, analyzing and forming of visualized data from smart homes. This can be used by consumers and utility providers to gain insights and track consumption patterns of a single smart home unit. This paper will thus provide a summarized synopsis to those who want to gain information regarding available techniques and methodology for smart system automation process for any sector such as healthcare, education, information technology, oil and gas etc.

Index Terms- Energy Management, Hadoop, Internet of Things, OLAP, Smart Homes

I. Introduction

Massive amount of data is generated every day by business sectors, residential sectors, industrial sectors, transportation facilities, telecommunication domains etc. Today, data is increasing exponentially and reaching levels of Terabytes (10^{12}), Petabytes (10^{15}), Exabytes (10^{18}), Zettabytes (10^{21}) etc. When compared to traditional business data, human data is increasing at an overall 10x faster growth rate, whereas machine data is increasing at a 50x growth rate [16]. This massive amount of data that involves structured and unstructured forms of data is known as big data, and is creating new opportunities for businesses to grow into a customer oriented and information driven environment. Hence, this enormous amount of data can be stored, analyzed, processed and developed into meaningful insights that can help utility providers to formulate ways to improve energy consumption, generation, operation and management. Since the cost of fossil fuels are increasing, carbon emissions are increasing, environment impacts are increasing, non renewable resources are depleting, equipment are leading towards obsolescence, security levels of energy resources are increasing and data generated by energy sectors are substantial, it is necessary for utility providers to store, analyze, process and develop meaningful results through new technological advancements while replacing traditional systems. Meaningful results from the data processed can be formulated into visualizations such as graphs and reports which can be used by utility providers, consumers and other stakeholders such as government, businesses etc.

Residential and commercial sectors are responsible for the highest consumption of energy in any economy. The big data generated by the residential sector, especially the smart homes, periodically can be leveraged into insightful results that can aid in predicting customer demands, observing customer consumption patterns and providing a complete load profile analysis of energy consumption. A smart home concept incorporates energy consumption units that are transferred to intelligent network nodes where energy is locally produced and intelligently managed [14]. On the other hand, like a smart grid lifecycle involves data generation, data acquisition, data storing, data processing, data querying and analytics and monitoring of the data, the same is true with respect to a smart home facility. Moreover, the smart home concept follows "ubiquitous sensing" which means that network of sensors are integrated to a network of processing devices which eventually structure and build multivariate information [15]. Hence, this paper describes this ubiquitous sensing concept by demonstrating the four data transmission stages in smart homes; collection, storage, processing, and analysis and visualization.

II. Brief Overview on Smart Homes

A smart grid is an advanced version of an electric grid where grid performance is utilized in satisfying the demand and supply of users efficiently through information and communication technologies (ICTs) and material engineering that focuses on maintenance of a sustainable system [18]. This has become a fundamental research domain for the energy sector to optimize energy consumption. Smart Homes and Smart meters are the two elements of the smart grid application of Internet of Things [14]. Smart homes are based on the concept of Internet of things, which is a primary source of sensory data [4] whereas smart meter is an interface unit between the smart grid of utility provider and the end user in a smart home [14].

In Smart Homes, air conditioning, lighting, heating and other systems are activated only when a consumer needs to use a particular system. This leads to optimization of energy consumption. The big data provided by smart homes detects patterns, nuances and anomalies, which could not be found otherwise. Consequently, this helps utility providers to analyze the consumption pattern and behavior of clients and convert this streaming data to visualized reports. Accordingly, providers can use these reports to monitor and ensure the anticipated changes in supply and demand with regard to client's electricity consumption behavior. This shows that data from smart homes helps utility providers in forecasting energy consumption, satisfying energy demand, improving the smart grid performance and lowering energy cost by supplying energy when needed.

III. Data sources in smart homes

The physical architecture of a smart home or any smart system integrates or bridges the gap between the physical world and computer-based systems. The physical architecture of a smart home involves wireless sensors, actuators, microcontrollers, processors, control center etc. [3][11]. The purpose of these devices is to communicate with each other in a network and provide meaningful insights. The technology used in an IoT system is driven by new innovations and has solved problems with regard to accuracy, cost and the ability to detect things that couldn't be detected previously. Consequently, IoT systems incorporate wireless sensors and actuators in its physical architecture where sensors convert a physical energy to an electrical impulse whereas, on the other hand, actuators convert an electrical input to a physical action [11]. Hence, sensors are devices that react to a stimulus and send a command to the control center, which activates an actuator to manipulate the physical environment. For example, a temperature sensor detects heat from a flame, which is the stimulus in this case, and sends a command to the control center. The control center sends the command to a sprinkler actuator, which eventually turns on and puts out the flame.

Sensors track and supervise the dweller of a house. There are different types of sensors used in a smart home system such as motion sensors, water sensors, doorbell sensors, temperature sensors, light sensors, weather sensors, synthetic sensors etc. [13], and there are different types of actuators as well in a smart system such as electric motors, hydraulic systems etc. These sensors and actuators are connected by wireless connections, mobile applications, wireless sensor networks and mobile sensor networks [3] [4]. WSNs are used to transmit data among sensor nodes through large geographic areas to increase communication capabilities regarding issues such as environmental research, water quality monitoring, civil engineering and wildlife habit monitoring [4]. Therefore, through the network structure, these technologies and the process of "Sensor to Actuator Flow" generate tremendous data, which is processed and analyzed in later stages to drive decision-making. The data collected is processed for immediate decision making with regard to threat detection, immediate crash statistics, abrupt shutdowns, etc. [11]. The data that is collected from these devices are entry and exit times, level and hours of usage, temperature levels etc., which is then sent to a database storage system for further processing [7].

IV. Data Storage

There are two types of databases for storage. This includes databases that follow the structured query language systems which are considered as traditional RDMS systems such as Oracle databases and open source systems that follow no SQL systems. There are various reasons why the development of big data has challenged the relational database management systems (RDMS) and has replaced these traditional systems with open source systems such as Hadoop and OLAP. RDMS are legacy systems that are not compatible with other technologies and are not set up to be simultaneously accessed for reporting and analysis. It analyzes structured data where it stores and analyzes data that are queried using a conventional SQL language. RDMS inculcates traditional form of row and column-oriented database and inundates parallel processing which is present in Hadoop and OLAP. The storage capacity is also limited when compared to the storage systems of Hadoop. RDMS includes the SQL query language where schemas are defined but, on the other hand, Hadoop follows a NOSQL query language where schemas are not required. RDMS does not possess the capability of processing data in real time like Hadoop due to lack of query performance and low latency caused by data storage issues. Subsequently, smart grid technology precisely smart homes, generate petabytes of data that create issues with respect to storage, processing and analyzing and this cannot be performed by traditional RDMS. Hadoop is thus good at indexing rich data, as it is flexible, scalable and cost effective. Therefore, a smart home setting should incorporate big tools like Hadoop and OLAP in analyzing the data achieved by these settings which can be used by consumers and utility providers to gain insights and facilitate demand and supply based on consumption needs. Therefore, through a Hadoop and OLAP system, the smart grid concept, which incorporates smart meter and smart homes, becomes more efficient as large data can be processed with faster data access as compared to traditional RDMS.

Hadoop incorporates a storage facility known as Hadoop Distributed File System (HDFS) that satisfies the issues regarding scalability, reliability, and fault tolerance. Through the data from advanced metering infrastructure (AMI) installed in smart grids and sensors, large bulk of data files are formed into blocks and stored in allocated HDFS cluster nodes which are termed as data nodes [5]. These data blocks are replicated into multiple data nodes where these data nodes are derived from name nodes, which store information regarding files and blocks [5]. These data blocks are further moved for processing to the Map Reduce processing stage in Hadoop. This multi-node structure of HDFS makes its fault tolerant [2].

V. Data Processing

Hadoop MapReduce, a programming model, is known for efficient and cost-effective processing of big data due to its batch processing capability. Each data stored in the HDFS cluster is sent into the mapper function where the Map Reduce algorithm is applied [5]. In the MapReduce processing layer, data is split into small parts, which is then formed into tasks. The process of this layer is reflected in the name of this layer "MapReduce". Map refers to the stage where tasks are mapped according to the query in terms of key and value pair. The Map task transfers the pairs to the reduce stage where these key and value pairs are reduced, sorted and made into a new set of output [10]. The master node chooses the tasks and allocates to the required job tracker [8].

Along with MapReduce, there are other data processing frameworks that support Map Reduce. One of such Hadoop release is YARN (Yet Another Resource Negotiator), which solves issues regarding scalability and resource utilization. To eliminate these issues, YARN divides the Job Tracker into two components; Resource manager, which manages resources in the form of containers available across nodes, and Application manager, which manages and supervises the tasks of the resource manager and the individual nodes [6]. Hive is another tool of MapReduce, which is used for data querying. This particular tool is devised to execute SQL queries to process data through Map Reduce algorithm [6]. Hence, Hive is eventually connecting HDFS or Hadoop data warehouse with MapReduce framework.

OLAP is another tool that can be integrated with Hadoop and can be used for processing of data. It acquires data from a data warehouse, such as HDFS and HBase, in a multi-dimension form. This benefit present in Online Analytic Processing (OLAP) is not modeled in the relational database, Online Transactional Processing (OLTP) as OLTP is a two-dimensional data structure, which includes rows and columns. OLAP data warehouse stores aggregated data, which can be filtered, sliced and sorted into multidimensional schemas in the form of a cube. The data stored in for OLAP processing is used for future decision-making and analytical processing.

Depending on the needs and requirements of the consumer and utility provider, data can be processed in the form of measures and dimensions such as time, location and appliances etc., as shown in figure 1. These dimensions are further divided into hierarchies, levels and attributes for further aggregation [17]. Utility providers can use the levels to determine the consumption trends based on the appliances for each quarter in a particular location. Consequently, with regard to smart home automation, utility providers can use OLAP processing model for consumption analysis, demand patterns, demand forecast and trend analysis to provide real time information feedback to consumers. Utility providers can adjust the demand and supply of energy by analyzing the low power consumption of users of smart homes or redistribute the demand and supply of energy by analyzing the high-power consumption of users.

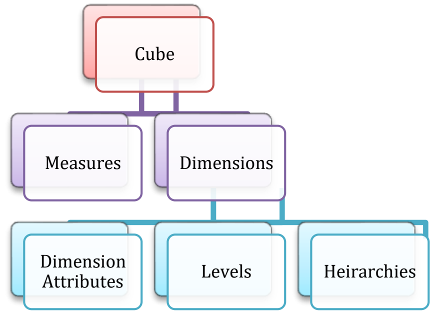

Figure 1: Data Process in OLAP

Complex data structures can be built by creating schemas or data points also known as dimensions, according to the OLAP system. Each dimension is then categorized into different attributes of a dimension. This is therefore shown in figure 2, where data obtained from smart homes can be formed into different types of appliances based on each quarter of the year in each city.

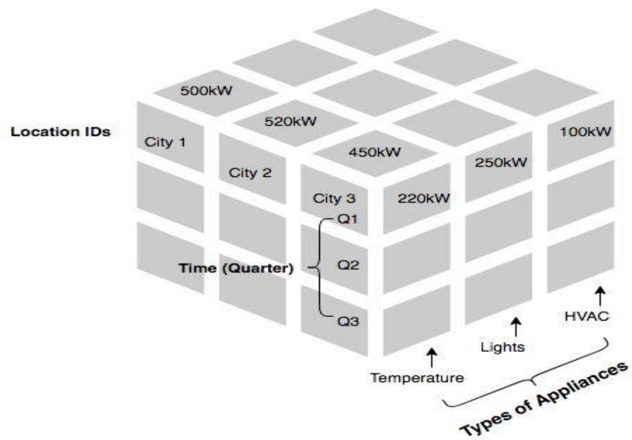

Figure 2: Multidimensional view of data in OLAP

This cube can further become more informative by further magnifying the data points as shown in figure 3:

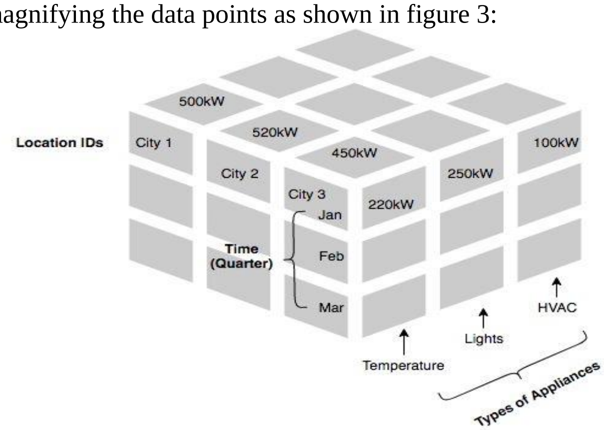

Figure 3: Multidimensional view of data concerning quarter 1

VI. Data Analytics (Visualization)

Map Reduce processing incorporates further analytical processing which is performed with the inculcation of other tools of Hadoop. Through analytic algorithms such as artificial intelligence, machine learning and other algorithms such as PIG of Hadoop, utility providers will be able to provide energy efficiently and will be able to make intelligent decisions on energy management. These predictive analyses incorporate neural networks and genetic algorithms that help in forecasting and generating predictive models that give insight on consumption behavior. An example of a scientific development that incorporates neural networks and genetic algorithms is the building of artificial intelligence. Artificial intelligence in Pig, with the help of machine learning, analyses the "reasoning, behavior, rationality and thought process" [15] of a human being. Neural networks of artificial intelligence and support vector machines are known for flexible monitoring of smart houses [15], and play a vital role in predicting the energy requirements and analyzing the heating, ventilation and air-conditioning (HVAC) units of a smart home facility [3]. Moreover, an empirical research on energy consumption demonstrated that, through predictive analysis and optimization methods, there was 23% of energy savings as compared to energy consumption in the previous month in a smart facility [3].

OLAP is also involved in further analytical processing of data through tools such as Pentaho Business Intelligence. Data stored in Hadoop can be integrated with Pentaho business intelligence OLAP tool as Pentaho has the capability to integrate data, mold data into multidimensional form, report, mine and analyze data, and create a dashboard that showcase Pentaho reports and architectural data after ETL operations [12]. Due to the data integration capability of Pentaho, data is retrieved from the HDFS layer of Hadoop and transported to the Pentaho Stack [9], which is then transformed to a multidimensional model for OLAP operations as described above in figure 2 and 3. These multidimensional models in the form of a cube are then transformed to graphs and reports that can be used by utility providers and other stakeholders for further forecasting and energy management.

VII. Conclusion and Recommendations

In conclusion, after analyzing different studies and frameworks, it is noticed that a theoretical big data solution involves the data sources stage followed by a data transformation stage which involves converting raw data to tables, schemas and clusters during the storage process. This data is transformed using big data platform and tools, which analyze and process this data into the end product through application and analytics. This process is explained in figure 4 and can be considered as the theoretical framework of big data solution.

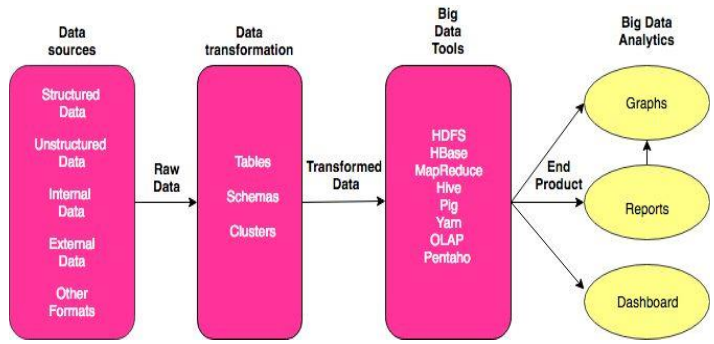

Figure 4: Theoretical framework of Big Data

Keeping in mind the theoretical framework of big data solution and various secondary researches, it can be seen that data transmission process from a residential unit to a utility sector involves four stages. These four stages are essential for an energy management progression. A utility provider needs to keep in mind the four stages facilitated in a smart home to track the demand and supply of energy consumption. Different stakeholders such as the home owner, the community owner, the state owner, the country owner [1] and the utility provider are affected by this humongous data transmission procedure that incorporates four stages; data generation, data storage, data processing, and data analytics. This procedure is briefly mapped in figure 5:

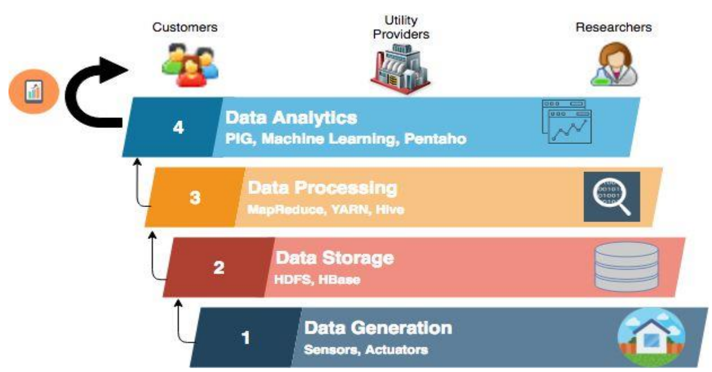

Figure 5: Four stages of data transmission from a smart home facility

Lastly, the world is going towards a technological change where various countries are focusing on energy management and are including smart cities and smart homes in their construction and infrastructural plans. Consequently, it is necessary for such countries that are adopting the concept of smart homes to keep in mind the stages of data transmission from a smart home facility and how this data can be used for forecasting demand and supply of energy, energy management and energy conservatism by different stakeholders in the energy sector.

References

[1] Al-Ali, A. R. et al. "A smart home energy management system using IoT and big data analytics approach." IEEE Transactions on Consumer Electronics 63 (2017): 426-434.

[2] Borthakur, Dhruba. "HDFS Architecture Guide." *Hadoop*, 22 Aug. 2019, hadoop.apache.org/docs/r1.2.1/hdfs_design.html.

[3] Cano, María Victoria Moreno et al. "Applicability of Big Data Techniques to Smart Cities Deployments." IEEE Transactions on Industrial Informatics 13 (2017): 800-809.

[4] Chen, Min, et al. "Big Data Generation and Acquisition." *Big Data*, 2014, pp. 19–32.

[5] Fanibhare, Vaibhav and Vijay R. Dahake. "SmartGrids: MapReduce framework using Hadoop." 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN) (2016): 400-405.

[6] Gavali, Swati M. and Supriya Sarkar. "Survey Paper on Big Data Processing and Hadoop Components." (2016).

[7] Hu, Zhubing. "A Data Acquisition and Control System in Smart Home Based on the Internet of Things." (2016).

[8] Humbetov, S.. "Data-intensive computing with map-reduce and hadoop." 2012 6th International Conference on Application of Information and Communication Technologies (AICT) (2012): 1-5.

[9] Jerez, Alex Rayón. "Kettle: Pentaho Data Integration Tool." *LinkedIn SlideShare*, 9 Jan. 2014, www.slideshare.net/alrayon/kettle-pentaho-data-integration-tool.

[10] Kiran, Ravi. "MapReduce Tutorial: Mapreduce Example in Apache Hadoop." *Edureka*, 11 Feb. 2020, www.edureka.co/blog/mapreduce-tutorial/.

[11] Misra, Joydeep. "IoT System: Sensors and Actuators Overview." *Bridgera*, 26 June 2017, bridgera.com/iot-system-sensors-actuators/.

[12] "Pentaho Data Integration." *Pentaho Documentation*, 20 May 2017, help.pentaho.com/Documentation/7.1/0D0/Pentaho_Data_Integration.

[13] Price, Dan. "6 Sensors Every Smart Home Should Have." *MakeUseOf*, 25 Aug. 2017, www.makeuseof.com/tag/sensors-every-smart-home/.

[14] Roehr, Andrew J. 2010, *Smart Home: The Human Side of the Smart Grid*, www.smartgrids-cre.fr/media/documents/1003_CapG_SmartHome.pdf.

[15] Sánchez, Veralia Gabriela et al. "A Review of Smart House Analysis Methods for Assisting Older People Living Alone." J. Sensor and Actuator Networks 6 (2017): 11.

[16] *The Exponential Growth of Data*. 16 Feb. 2017, insidebigdata.com/2017/02/16/the-exponential-growth-of-data/.

[17] "The Multidimensional Data Model." *Oracle OLAP Application Developer's Guide*, 2002, docs.oracle.com/cd/B13789_01/olap.101/b10333/multimodel.htm.

[18] Yu, Xinghuo et al. "The New Frontier of Smart Grids." IEEE Industrial Electronics Magazine 5 (2011): 49-63.

AUTHORS

First Author – Sana Ramzan, M.Sc., Department of Accounting, Algonquin College, sramzan@ac-kuwait.edu.kw