Hadoop Famework for Monitoring the Smart Home Devices

Mr. Ravindran U[1], Ms. Sandhiya A[2]

[1]Assistant Professor, Dept. of CSE, Kingston Engineering College, Tamil nadu, India-632059

[2]Student, Dept. of CSE, Kingston Engineering College, Tamil nadu, India-632059

---***---

ABSTRACT-Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater operational efficiencies, reduced risk and cost reduction. Smart home has more electronic devices when the human makes use of these electronic devices regularly there are lots of chances that his/her health may get disturbed and may face lots of inconvenience due to poor health conditions that caused because of those electronic devices. This project focuses on the utilization of energy i.e., electrical energy by means of voltages from various home for each electronic devices and can identify which electronic devices is widely used in most of the home and can analysis the cause i.e., disease occurred due to the usage of these gadgets. Huge volume of data is required to perform analysis and yield better results. Relational Data Base Management System is not sufficient to store these large amount of data to analyze which device consumes large amount of electricity, cause that is the disease. In order to perform such analysis we make use of powerful Hadoop framework. Hadoop framework application provides distributed storage and distributed processing of large datasets.

Keywords: Smart Home, Electronic Devices, Energy Consumption, Hadoop, Diseases.

INTROUDCTION

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 has become a complete subject, which involves various tools, techniques and frameworks. Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater operational efficiencies, cost reductions, and reduced risks for the business. To harness the power of big data, you would require an infrastructure that can manage and process huge volu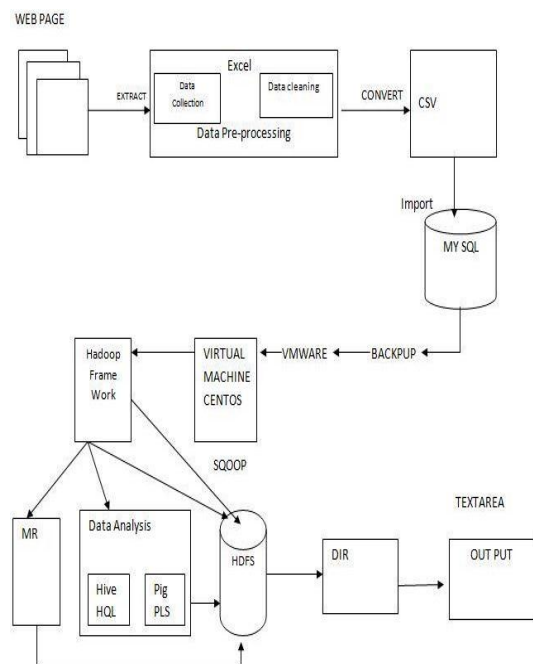mes of structured and unstructured data in realtime and can protect data privacy and security. A smart home is a residence that uses internet-connected devices to enable the remote monitoring and management of appliances and systems such as lighting and heating etc., Smart home provides home owners security, comfort, convenience and energy. By consumption of over energy at smart homes it will raise electrical bill and may cause many health issues like eye problems, brain tumors, and breathing problems etc., We use big data hadoop framework for smart homes to identify which electronic device is most widely used by many homes and detect the illness that caused by which electronic device.

SYSTEM DESIGN

DATA PREPROCESSING MODULE

Smart Home Dataset such as Appliances, Diseases, Energy Consumption are collected from multiple websites and feed into MS Excel for data cleaning in order to remove noisy and missing data. The Excel sheet is now converted into CSV file and imported into MySQL Database for easy access. The CSV file format is not fully standardized.

The basic idea of separating fields with a comma is clear, but that idea gets complicated when the field data may also contain commas or even embedded line breaks.

CSV implementations may not handle such field data, or they may use quotation marks to surround the field. Quotation does not solve everything: some fields may need embedded quotation marks, so a CSV implementation may include escape characters or escape sequences.

DATA PREPROCESSING MODULE DIAGRAM
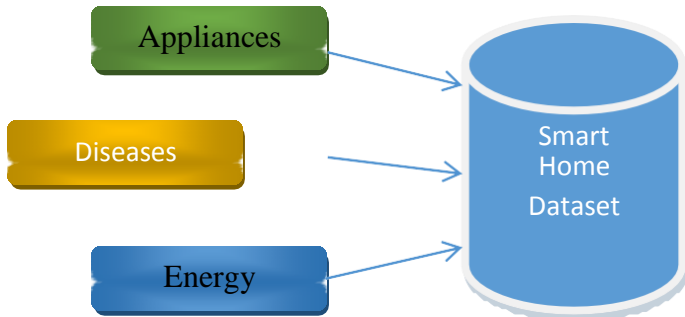
DATA MIGRATION USING SQOOP

Now we are ready with smart home 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 ANALYSIS USING HIVE

Hive is a Data warehouse 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. The smart home datasets in particular area like such that Appliances, Diseases, Energy Consumption are analysed. The analysed that dataset using HIVE tool will be stored in hadoop (HDFS).For

analyzing the dataset HIVE using HQL Language. Using hive we perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language. Hive is an ETL and Data warehousing tool developed on top of Hadoop Distributed File System (HDFS). Hive makes job easy for performing operations like

- Data encapsulation

- Ad-hoc queries

- Analysis of huge datasets
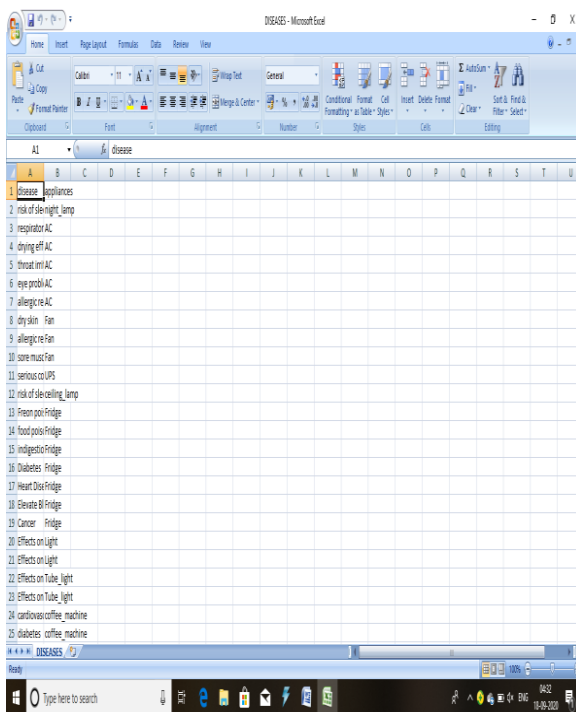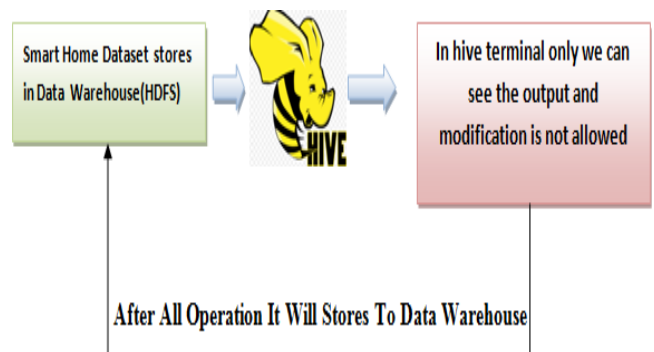

MAP REDUCE
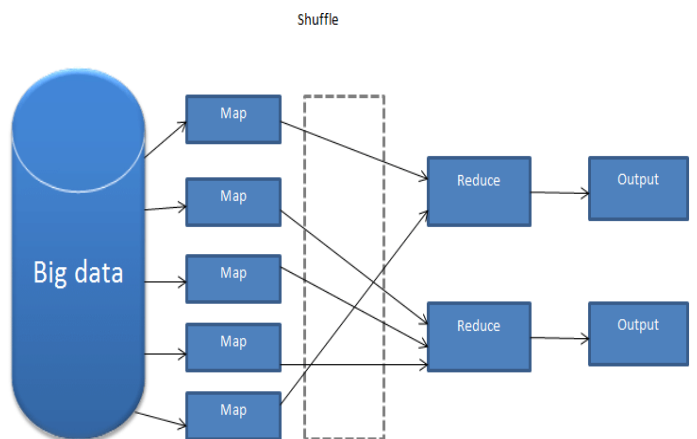
MapReduce is a Distributed Data Processing Algorithm introduced by Google. MapReduce Algorithm is mainly inspired by Functional Programming model MapReduce algorithm is useful to process huge amount of data in parallel, reliable and efficient way in cluster environments. It divides input task into smaller and manageable sub-tasks to execute them in-parallel MapReduce algorithm is based on sending the processing node (local system) to the place where the data exists.

MapReduce Algorithm works by breaking the process into 3 phases.

- Map Phase

- Sort & Shuttle phase

- Reduce phase

In MapReduce, each phase has key-value pairs as input and output. MapReduce will always expect the input in the form of Key & Value pairs from HDFS layers. Once the MapReduce processing completed, it will produce the o/p again on top of HDFS in the form of (Key, Value) pair. It will produce which electronic device consumes large amount of electricity and because of that large consumptions of electricity what diseases are occur.

RESULT AND DISCUSSION

The analysis is completely based on the tools such as Hive and Pig which makes use of HQL(Hive Query Language) and Pig Latin Scripting Language and upon terminal the analysis logic is performed based on use cases. The result is same upon both the tools where as the performance may vary and has both merits and demerits on comparison with those analysis tools. The predicted output can be the home which uses electronic devices during a specified time interval and the disease occurred due to this usage.

CONCLUSION

In this project we focused on the utilization of energy i.e., electrical energy by means of voltages from various home for each electronic devices and identified which electronic devices is widely used in most of the home and the cause i.e., disease occurred due to the usage of these gadgets. To yield better results we used huge volume of data hence Hadoop framework which is used for Big Data was focused.

FUTURE ENHANCEMENT:

Basically Spark is a framework - in the same way that Hadoop is - which provides a number of inter-connected platforms, systems and standards for Big Data projects. Like Hadoop, Spark is open-source and under the wing of the Apache Software Foundation. Essentially, open-source means the code can be freely used by anyone. Beyond that, it can also be altered by anyone to produce custom versions aimed at particular problems, or industries. Volunteer developers, as well as those working at companies which produce custom versions, constantly refine and update the core software adding more features and efficiencies. In fact Spark was the most active project at Apache last year. It was also the most active of all of the open source Big Data applications, with over 500 contributors from more than 200 organizations. Unlike Hadoop, Spark does not come with its own file system - instead it can be integrated with many file systems including Hadoop's HDFS, MongoDB and Amazon's S3 system.

REFERENCES

[1] L. M. Camarinha-Matos, "Collaborative smart grids–A survey on trends," Renewable and Sustainable Energy Reviews, vol. 65, pp. 283-294, 2016.

[2] R. Bayindir, E. Hossain, E. Kabalci, and R. Perez, "A comprehensive study on microgrid technology," International Journal of Renewable Energy Research, vol. 4, pp. 1094-1107, 2014.

[3] E. Hossain, R. Perez, S. Padmanaban, and P. Siano, "Investigation on the Development of a Sliding Mode Controller for Constant Power Loads in Microgrids," Energies, vol. 10, p. 1086, 2017.

[4] E. Hossain, E. Kabalci, R. Bayindir, and R. Perez, "Microgrid testbeds around the world: State of art," Energy Conversion and Management, vol. 86, pp. 132-153, 2014.