Big Data – Road to Smart Cities

R. S. Balaji

M. Tech Cloud Computing, Department of Information Technology, SRM University, SRM Nagar, Kattankulathur, Tamil Nadu, India, 603203

Abstract: ***The recent trends in social development show that in the next decades an increasing larger share of the population worldwide will live in cities. Decarbonizing the energy footprint of urban areas becomes therefore a critical point in the outlook of meeting challenging environmental targets set by Governments in many Countries and eventually fighting climate change. Optimal utilization of local resources represents a strategic area to improve the environmental efficiency of cities, and integrated operation and planning of the overall urban system is needed to optimize such efficiency. Hence, while various energy vectors such as gas, electricity, heat, and so on, have been traditionally decoupled in terms of operation and planning, the need to reduce emissions and the availability of smart grid technologies represent a significant opportunity to rethink urban energy systems. The objective of this paper is to discuss the main issues relevant to wastage of energy in the buildings and method to reduce it through the smart control devices.***

Keywords: Big Data, Hadoop distribution File System (HDFS), Android mobiles, smart switch and n/w extender (THA-101).

1. Introduction

The topic of energy efficiency has come to dominate the energy policy world widely. Day by day the energy demand is increasing. This is required to meet the advancement in all areas of any nation and set to continue into the future as a result of political requirements and concern about the environment and climate. The opportunity to gain cost and competitive advantages through the more efficient use of energy will also ensure that the topic does not lose its importance. Although efforts have been made in the area of energy efficiency in the past, there is still significant room for improvement.

In a commercial building, the productivity and quality of life of the people depends on the comfort level within it. Visual and thermal comforts are the main comfort factors. These are provided by lighting and air conditioning systems. The more the comfort level provided, the more is the energy consumption. Energy scarcity, these days, necessitates a balance between the energy consumption and the customers‟ comfort.

Commercial and residential buildings use almost 40% of primary energy and approximately 77% of the electricity. Commercial and residential users are the fastest growing consumers of energy use. The residential sector has experienced changes, accounting for 14.6% of total energy consumption in 1990 rising to 21% in 2005.

Energy consumption in the commercial building sector will continue to increase until buildings can be designed to produce enough energy to offset the growing energy demand.

Energy use of commercial and residential buildings can be reduced through upgrading thermal insulation of walls and roofs, and through measures to reduce uncontrolled energy consumption. But, any examination of energy efficiency problem cannot be restricted to the more efficient use of energy or "passive techniques". Particularly, a more integrated approach that incorporates the use of state-of-the-art technology is considered as necessary. Energy efficiency is a part of energy intelligence. The intelligent use of energy means getting the most out of the primary energy used. It combines clever energy savings with improved ease-of-use and sustained cost savings.

Energy can be saved to a great extent by properly managing the energy while ensuring reliability of the critical loads. This also minimizes the total energy consumption. The saved energy can be utilized later by the critical loads. The intelligent energy management system used in commercial buildings can serve multiple purposes, like reducing the total energy consumption, ensuring effective and efficient use of energy, ensuring supply to the critical loads, and minimizing the energy wastage.

So in this paper I am going to explain how the android device temperature data and number of people as data from the building based on timing help to reduce the energy consumption.

2. Data Collection Methods

2.1 Andriod Mobile

°C or the pressure in hPa. Also, unlike motion sensors and position sensors, which often require high-pass or low-pass filtering, environment sensors do not typically require any data filtering or data processing.

The application sense the current environment temperature (or) room temperature, time, date and current location, stores the data in mobile by creating the file on SD card. Then text file which is create is feed into Hadoop distribution file system.

2.2 Data From the Building's

Data is collected from the organization, school, college based on the timing how many a people"s is there while they are entering into particular building as an employee"s or students. This raw data is continuously collected from some organization or school and the data is stored in the system for further process.

This data is finally feed into Hadoop Distribution file system to compare with the already feuded temperature data with time and current location to make a decision for the electricity usage pattern.

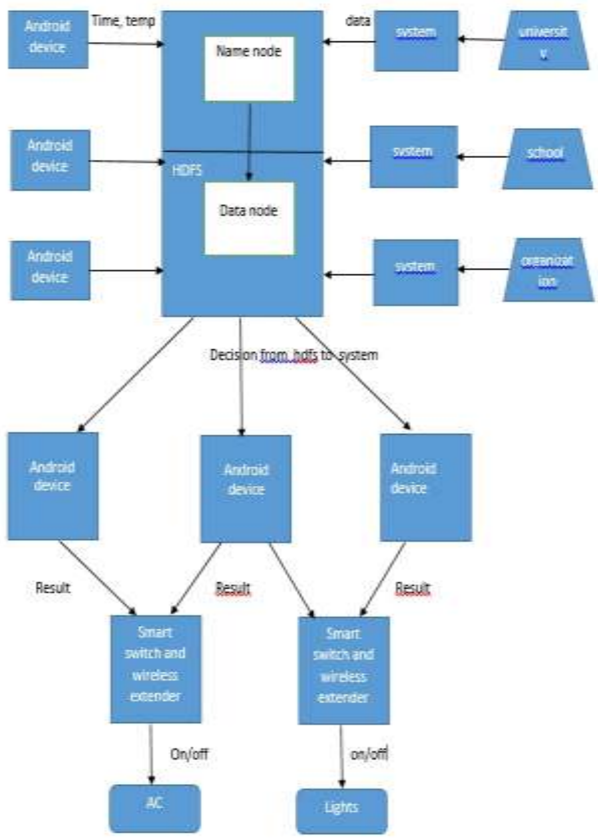

Figure 1: Architecture of the proposed system

Hadoop Distributed File System (HDFS)
The **Hadoop distributed file system** (**HDFS**) is a distributed, scalable, and portable file-system written in Java for the Hadoop framework. A Hadoop cluster has nominally a single namenode plus a cluster of datanodes, although redundancy options are available for the namenode due to its criticality. Each datanode serves up blocks of data over the network using a block protocol specific to HDFS. The file system uses TCP/IP sockets for communication. Clients use remote procedure call (RPC) to communicate between each other.

HDFS stores large files (typically in the range of gigabytes to terabytes) across multiple machines. It achieves reliability by replicating the data across multiple hosts, and hence theoretically does not require RAID storage on hosts (but to increase I/O performance some RAID configurations are still useful). With the default replication value, 3, data is stored on three nodes: two on the same rack, and one on a different rack. Data nodes can talk to each other to rebalance data, to move copies around, and to keep the replication of data high. HDFS is not fully POSIX-compliant, because the requirements for a POSIX file-system differ from the target goals for a Hadoop application. The tradeoff of not having a fully POSIX-compliant file-system is increased performance for data throughput and support for non-POSIX operations such as Append.
HDFS added the high-availability capabilities, as announced for release 2.0 in May 2012,letting the main metadata server (the NameNode) fail over manually to a backup. The project has also started developing automatic fail-over.

The HDFS file system includes a so-called *secondary namenode*, a misleading name that some might incorrectly interpret as a backup namenode for when the primary namenode goes offline. In fact, the secondary namenode regularly connects with the primary namenode and builds snapshots of the primary namenode's directory information, which the system then saves to local or remote directories. These check pointed images can be used to restart a failed primary namenode without having to replay the entire journal of file-system actions, then to edit the log to create an up-to-date directory structure. Because the namenode is the single point for storage and management of metadata, it can become a bottleneck for supporting a huge number of files, especially a large number of small files. HDFS Federation, a new addition, aims to tackle this problem to a certain extent by allowing multiple namespaces served by separate namenodes.

3. Hadoop Architecture

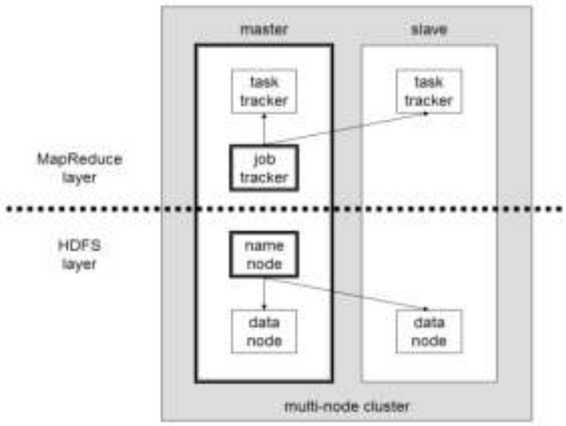

Figure 2: Hadoop Architecture (multi node cluster)

An advantage of using HDFS is data awareness between the job tracker and task tracker. The job tracker schedules map or reduce jobs to task trackers with an awareness of the data location. For example: if node A contains data (x,y,z) and node B contains data (a,b,c), the job tracker schedules node

B to perform map or reduce tasks on (a,b,c) and node A would be scheduled to perform map or reduce tasks on (x,y,z). This reduces the amount of traffic that goes over the network and prevents unnecessary data transfer. When Hadoop is used with other file systems, this advantage is not always available. This can have a significant impact on job-completion times, which has been demonstrated when running data-intensive jobs. HDFS was designed for mostly immutable files and may not be suitable for systems requiring concurrent write-operations. HDFS can be mounted directly with a Filesystem in Userspace (FUSE) virtual file systemon Linux and some other Unix systems.

File access can be achieved through the native Java API, the Thrift API to generate a client in the language of the users' choosing (C++, Java, Python, PHP, Ruby, Erlang, Perl, Haskell, C#, Cocoa, Smalltalk, and OCaml), the command-line interface, browsed through the HDFS-UI webapp over HTTP, or via 3rd-party network client libraries.

3.1JobTracker and TaskTracker: the MapReduce engine

Above the file systems comes the MapReduce engine, which consists of one *JobTracker*, to which client applications submit MapReduce jobs. The JobTracker pushes work out to available *TaskTracker* nodes in the cluster, striving to keep the work as close to the data as possible. With a rack-aware file system, the JobTracker knows which node contains the data, and which other machines are nearby. If the work cannot be hosted on the actual node where the data resides, priority is given to nodes in the same rack. This reduces network traffic on the main backbone network. If a TaskTracker fails or times out, that part of the job is rescheduled. The TaskTracker on each node spawns off a separate Java Virtual Machine process to prevent the TaskTracker itself from failing if the running job crashes the JVM. A heartbeat is sent from the TaskTracker to the JobTracker every few minutes to check its status. The Job Tracker and TaskTracker status and information is exposed by Jetty and can be viewed from a web browser.

Known limitations of this approach are:

The allocation of work to TaskTrackers is very simple. Every TaskTracker has a number of available *slots* (such as "4 slots"). Every active map or reduce task takes up one slot. The Job Tracker allocates work to the tracker nearest to the data with an available slot. There is no consideration of the current system load of the allocated machine, and hence its actual availability.

If one TaskTracker is very slow, it can delay the entire MapReduce job – especially towards the end of a job, where everything can end up waiting for the slowest task. With speculative execution enabled, however, a single task can be executed on multiple slave nodes.

Data from the different source such as application installed android device may be in travel around the cities like low temperature area, high temperature area, data from these device will be as the date, location, time, and environment temperature is collected and stored in the Hadoop distribution file system continuously for the decision process and data from the organization, school, university etcwill be compared based on time, temperature, people using electricity. The result is passed to the android device with application from any internet connection operates the smart switch and network extender(THA-101) to turn on or off a light, fans, machine etc.Place the THA-101 in an area of your home with low wireless coverage the powerful built-in N300 wireless extender connects to your existing wireless router and broadcasts a strong wireless network.

4. Conclusion

As the decision made by Hadoop distribution file system using naïve bayesain algorithm and it is passed to android device as the result. Use the free Apple or Android™ mobile app with the result from any internet connection to turn on and off a light, fan, electronic device, or appliance. Create a daily schedule to turn on/off the outlet to match your needs or to give the illusion that you're home.It is consider as an suitable energy management plan for any organization or schools based on the climate, time and number of people in the particular building. To reduces the wastage of energy used unnecessary.

5. Future Work

In this paper i have discussed reduction of energy usage using of a predictive model for organization, school and university based on the time, temperature, and number of people inside the particular buildings. We have proposedto reduce the unnecessary used energy in particular place when no one over there. So the future work will be carried out of any like urban transport, urban waste management, public safety, water management.

References

[1] Ying-Wen Bai and Chi-Huang Hung, "Remote power On/Off control and current measurement for home electric outlets based on a low-power embedded board and ZigBee communication", IEEE International Symposium on Consumer Electronics, 2008. 14-16 April 2008, pp. 1-4.

[2] L.D.D. Harvey, "Energy savings by treating buildings as systems", Physics of Sustainable Energy, Using Energy Efficiently and Producing it Renewably, American Institute of Physics Conference Series, pp. 67–87, 2008.

[3] D. Hawkes and W. Forster, "Energy Efficient Buildings, Architecture,Engineering, and Environment", 2002, New York, Norton

[4] A. Wagner, S. Herkel, G. Löhnert and K. Voss, „Energy efficiency incommercial buildings", EuroSolar 2004, Freiburg, Available :www.solarbau.de

[5] European Standard EN 15232. "Energy performance of buildings –Impact of Building Automation, Controls and Building Management").

[6] L. D. D. Harvey, "Reducing energy use in the buildings sector:measures, costs, and examples", Energy Efficiency, v.2, 2009, pp. 139–163, DOI 10.1007/s12053-009-9041-2.

[7] Xiaohong Guan, Fellow, IEEE, Zhanbo Xu, and Qing-Shan Jia, Member,IEEE, "Energy-Efficient Buildings

Facilitated by Microgrid", *IEEE Transactions on Smart Grid*, Vol. 1, No. 3, December 2010, pp. 243-252.
[8] Dr. Gary L. Johnson, "Wind Energy Systems", Chapter-4-Wind Turbine Power, pp.1-54.
[9] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domesticelectricity use: A high-resolution energy demand model", *Energy andBuildings*, 42 (10), pp 1878-1887, October 2010.
[10] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions,"*FutureGener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.[Online]. Available: http://dx.doi.org/10.1016/j.future.2013.01.010