International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 26 Issue 5 – AUGUST 2019.

# DISTRICT POWER ANALYZER USING DATA TECHNOLOGIES

M. RASYA BANU[#1] and Dr. A.V. SEETHALAKSHMI [*2]

[#] *M. Phil. Scholar, Dept. of Computer Science, PRIST University, Thanjavur Campus, India*

[*] *Computer Science & Computer Application, PRIST University, Thanjavur Campus, India*

***Abstract***— **T A large number of cloud services require users to share data like electric meter records for data analysis or mining, bringing important concerns. Anonymzing data sets via generalization to satisfy certain requirements such as k-anonymity is a widely used category of some techniques. At present, the scale of data in many cloud applications enlarge considerably in following with the Big Data tendency, thereby making it a provocation for commonly used software tools to capture, manage, and process such large-scale data within a acceptable elapsed time. As a result, it is a provocation for existing undetermined approaches to achieve privacy preservation on privacy-sensitive large-scale data sets due to their inadequacy of scalability. In this paper, we propose a scalable two-phase top-down specialization (TDS) approach to anonymize large-scale data sets using the MapReduce substructure on cloud. In both phases of our approach, we consciously design a group of original MapReduce jobs to concretely realize the specialization assessment in a highly scalable way. Experimental evaluation results indicate that with our approach, the scalability and efficiency of TDS can be appreciably improved over existing approaches.**

***Index Terms***— **Energy Efficient Buildings, Smart Metering, Big Data**

## I. INTRODUCTION

Smart Energy has been a dominant conceptual prototype for future energy use. Because of limited nonrenewable energy measure available on Earth and also high costs of receiving renewable energies (REs), how to make energy use more well organized and constructive is reproving for future social and economic developments. Smart grids (SGs) have been a key enabler for smart energy, which refers to power networks that can intelligently integrate the behaviors and actions of all stakeholders connected to it, e.g., generators, customers, and those that do both—in order to efficiently deliver sustainable, economic, and secure electricity supplies. While there are many definitions for SGs, one commonly used conceptual framework is that of the National Institute of Standards and Technology (NIST) which defines seven important domains: bulk generation, communication, distribution, customers, service providers, performance, and markets.

Key technological challenges facing SGs include intermittency of RE generation that affects electricity quality; large scale networks of small distributed generation mechanisms, e.g., photovoltaic (PV) panels, batteries, wind and solar, plug-in hybrid electric vehicles (PHEVs), that result in high complexities. Another significant issue is how to use information and communication technologies (ICTs), advanced electronic and analytic technologies to enhance efficiency and cost effectiveness of energy use. Managing SGs to deliver smart energy require advanced data analytics for acquiring accurate information and automated decision support and handling events in a timely fashion. Significant progresses have been made for using field data obtained from intelligent devices installed in substations, feeders, and various databases and models across the utility enterprises. Some of the examples can be found in and references therein. Typical information sources include market data, lighting data, power system data, geographical data, weather data which can be processed and converted into information and knowledge that can be used for state estimation, situational awareness, fault detection and forewarning, stability assessment, wind or solar forecasting. Information acquisition is a key for timely data sensing, processing, and knowledge extraction. So far, the most talked-about information about power network operations is from data collected from intelligent electronic devices installed in substations and various parts of the transmission and distribution networks

A simple classification is consumers, electricity companies (utilities), and environment. In some literature, a further level of granularity has been added to the electricity company class by expanding to metering company (distributor), utility company, and supplier (retailer). However, a more comprehensive classification would be those by NIST, where the functionality of the whole energy usage cycle is defined to include bulk generation, transmission, distribution, customers, service providers, operations, and markets. In terms of bulk generation, transmission, and distribution, which involve meter company and grid company, smart meters will complement well with existing infrastructure to provide a more accurate and timely view of the energy consumption by regions. Events such as suspicious usage areas and potential faults will be noticed more easily and on-time actions taken subsequently when necessary. It may also enable more accurate prediction of electricity flows enabling better network maintenance planning. Smart meters can enable consumers to directly review their electricity usage, even down to the level of separate appliances, and thus adjust their behaviors to reduce energy cost. Customized rate plans are another key benefit to consumers.

Although not a common practice at present, smart meters enable demand–response for consumers where limiting or even cutting off the supply depending on market situations is possible. When all consumers being aware of both consumption and production of energy, adapt their energy usage during a period of high demand, high pricing or lower supply, more reliable and stable supply, better energy awareness, savings and efficiency will be achieved. In combination, these activities have been called demand side management (DSM) which is essential to really benefit consumers. Consumer awareness of the benefits from smart meters as well as their functionality will be a key factor in the successful adoption of this technology. As discussed in, another key factor could be the use of disaggregation techniques to extract underlying end use and appliance-level information from an aggregated energy signal. For retailers, the availability of vast volumes of data which could be used to profile and understand customers, their needs and behaviors enable better service provision and build stronger loyalty. Better consumer awareness is expected to result in reduced energy consumption thus reducing the need for additional power plants which generate greenhouse gases. Restricting and reducing electricity usage during peak periods can result in cutting down on the need of using peeker plants which generally make higher carbon emissions. Load control feature in smart meters enables switching individual appliances ON and OFF as required. Retailers could offer this feature to customers when the cost of power is very high, while distributors could use it when a section of the network is close to capacity.

The “DISTRICT POWER ANALYZER” is the main type of data that utilities are collecting is customer energy usage. The inability of customers and third parties to access this data—be it daily, hourly, or in near real-time—significantly limits the benefits that smart meters can provide. Having access to their own usage data allows consumers to track and manage their energy use, decrease their costs. That said, most consumers are not in a good position to turn independent data points into useful instruction. The real value will come from making the data accessible to third-party dispensed energy measure and energy service companies that can process the data, providing construct and actionable insights to consumers. Finally, we analyze the power energy in each district and display which district need high power energy and which district need old power energy. This will helpful for reduce the power energy in each sector. Chapter-II describes the methodology.

## II. RELATED WORK

### *A. Big Data Platforms For The Internet Of Things, Radu Ioan Ciobanu, Valentin Cristea, Ciprian Dobre And Florin Pop.2014,Springer*

This paper focuses on how Big Data could change the research direction in the business model by providing services along with products. Technology shift generate more data through various applications like wireless sensors, smart devices, social media etc., this paper focuses on the improvement the performance of the old services and offer new services in an open and dynamic environment. Also discusses the expected challenges and upcoming trends in the context aware environments for the Internet of Things. IoT aims to integrate and collect the information from smart objects of various domains. IoT infrastructure is best suited for integration, collection, processing, transmission and delivery of context information. It combines context model with event based organisation of services. This paper insisting on IoT is the backbone for the development of many applications which includes people, things, mobility and governance. Opportunistic networks facilitate the mobile communication when things are unable to establish the communication or it is offloaded to handle with large throughputs. The exchange of data is for the users in a closed place and the mobility happens through the short range transmission protocols. To handle the dynamic environment, the mobile devices are working in store-carry-and forward paradigm. The contact acts as the opportunities for the data to move to the destination. In this network data distribution happens through publish/subscribe model. This paper discussed from IoT point of view and various opportunities analyzed using new categories.

### *B. Fog Computing: A Platform For Internet Of Things And Analytics, Flavio Bonomi, Rodolfo Milito, Preethi Natarajan And Jiang Zhu,2014,Springer*

This paper proposed a hierarchical distributed architecture for IoT. Fog computing proposes a new breed of applications and services to have a productive interaction between existing cloud and Fog. Special focus given to Analytics and challenges of Big Data. Fog computing is the next level of the cloud computing and it uses common resources. A Smart Traffic Light System (STLS) and Wind form systems were taken as the use cases in Fog computing. The outcome of the use case studies discussed various attributes between Fog and Cloud. Requirement and necessity of Fog deeply analysed and discussed in the use cases. Relevance of Fog emphasized for IoT and Big Data. Technical requirement of Fog also discussed under the Fog architecture. Primary aim of this paper is how cloud computing can be extended into Fog.

### *C. Using Materialized View As A Service Of Scallop4sc For Smart City Application Services Shintaro Yamamoto, Shinsuke Matsumoto,Sachio Saiki, And Masahide Nakamura Kobe University, 1-1 Rokkodai-Cho, Nada-Ku, Kobe, Hyogo 657-8501, Japan*

This paper focuses on developing materialized view architecture as a smart service for Smart city. In a Smart city environment, houses and infrastructure which are connected to smart city produces huge amount of data and it is a big challenge to process the data and to get the required information from available data. Big Data is required to meet the challenge. Processing everything with raw data will take enormous time. This paper proposed architecture named Materialized View as a Service. Within the abstract cloud service it encapsulates all the transactions related to materialized views. The architecture designed in data platform Scallop4SC. The architecture designed in MapReduce on Hadoop and HBase KVS. It takes care of the design and implementation part of house logs. So finally the MVaaS converts the house logs into application specific materialised view. Also the effectiveness of the system demonstrated in three case studies.

*D. Mukherjee, A.; Datta, J.; Jorapur, R.; Singhvi, R.; Haloi, S.; Akram, W. (18-22 Dec. 2012) "Shared Disk Big Data Analytics With Apache Hadoop"*

This paper discusses the necessity of Big Data and Big Data techniques which is required to process huge amount of data and to discover insights. Hadoop is a open source platform used for implementing Mapreducer Model. The performance of VERITAS Storage Foundation Cluster File System ( SF CFS) is compared with Hadoop distributed file system (HDFS) for shared data Big Data analytics. Analytics with clustered file system is best suited for this proposed model.

*E. "Towards MapReduce Performance Optimization: A Look Into The Optimization Techniques In Apache Hadoop For Big Data Analytics" Kudakwashe Zvarevashe1, Dr. A Vinaya Babu*

Traditional database management system can't handle huge distributed, structured and unstructed data. Big Data plays a role in solving the issues of handling huge, complicated and dynamic data. Hadoop and NoSQL databases supported to eradicate these problems. Various technologies associated to MapReduce discussed in this paper. Difference research problems related to the improvement of the MapReduce problem discussed.

## III. METHODOLOGY

### A. *APACHE HADOOP ECOSYSTEM*

There are several techniques and tools for solving many IoT data management challenges like Big data, cloud computing, semantic sensor web, data fusion techniques, and middleware. • Big Data Analytics and Tools many techniques or methodologies that can solve IoT data processing and analytics issues in many concepts, fig.1 showed the Apache Hadoop ecosystem.

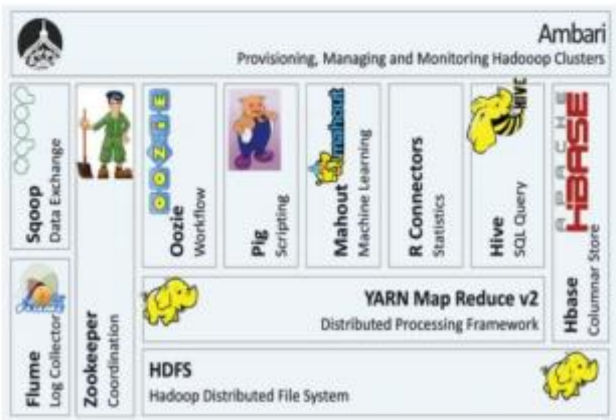


Figure 1: Apache Hadoop ecosystem

### B. *Hadoop:*

Hadoop is an open source assignment that managed by the Apache Software Foundation. Big data can be collected and handled by Hadoop. Hadoop is proposed to parallelize data processing through computing nodes to hurry computations and hide latency. There are two main components for Hadoop: Hadoop Distributed File System (HDFS) and Map Reduce engine. HDFS maintain enormous data constantly set and reproduce it to the user application at high bandwidth. MapReduce is a framework that is used for processing massive data sets in a distributed fashion through numerous machines

### C. *Map Reduce:*

MapReduce was constructed as a broad programming paradigm. Some of the original employments offered all the key needs of parallel execution, fault tolerance, load balancing, and data manipulation. The Map Reduce named with this name because it includes two abilities from existing functional computer languages: map and reduce. The MapReduce framework gets all sets with the common key from all records and joins them together. Therefore, it requires forming one group for each one of the different produced keys. MapReduce is one of the most new technologies, but it is just an algorithm, a technique for how to fit all the data. To acquire the best from MapReduce, we require more than just an algorithm. We require a collection of products and technologies created to manage the challenges of Big data.

### D. *HBase:*

HBase: It is a database model inside the Hadoop framework that looks like the original system of Big Table. The HBase has a column that operates as the key and is the only index that can be used to get back the rows. The data in HBase is also saved as (key, value) sets, where the subject in the non-key columns can be represented by the values.

### E. *Hive:*

The already deployed tools for data warehousing are not able to be suitable especially in the situation wherever, data is accessible everywhere; they are costly and often privately-operated. Such as the notion like MapReduce is there, it requests for the ability to write job procedures. Map Reduce jobs are difficult to track the characteristics of reusable code as some jobs are business particular some of the time. Hive may require be thought as the necessary portion of Hadoop system and views at the top that principally is the organization for the data warehouse. Hive cannot treat with applications and transactions of the real time those are achieved online. The motivation behind it is a complicated technique.

## IV. RESULT & DISCUSSION

### A. *MODULES*

Data Partition
Anonymization
Merging
Specialization
Obs

### B. *DATA PARTITION:*

In this section the data partition is performed on the cloud. Here we gather the large no of data sets.

We are dividing the large into small data sets.

Then we gives the random no for each data sets.

### C. ANONYMIZATION:

After gathering the individual data sets we apply the anonymization.

The anonymization used to hide or remove the sensitive field in data sets.

Then we get the intermediate middle result for the small data sets

The middle results are used for the specialization process.

All middle anonymization levels are merged into one in the second phase. The merging of anonymization levels is completed by merging cuts. To protect that the merged intermediate anonymization level ALI never violates privacy requirements, the more general one is selected as the merged one.

### D. MERGING:

The middle results of the several small data sets are merged here.

The MRTDS driver is utilized to organize the small intermediate result.

For combining, the combined data sets are collected on cloud.

The combining result is again applied in anonymization called specialization.

### E. SPECIALIZATION:

After gathering the intermediate result those results are merged into one.

Then we applies the anonymization on the merged data it called specialization.

Here we are utilize the two kinds of jobs such as IGPL UPDATE AND IGPL INITIALIZATION.

The jobs are created by web using the driver.

### F. OBS:

The OBS called optimized balancing scheduling.

Here we mainly focus on the two kinds of the scheduling called time and size.

Here some data sets are split in to the specified size and applied anonymization on specified time.

The OBS approach is to provide the high ability on handles the large data sets.

## V. CONCLUSION AND FUTURE ENHANCEMENT

We have checked the scalability problem of large-scale data anonymization by TDS, and proposed a highly scalable two-phase TDS approach using MapReduce on cloud. Data sets are divided and anonymized in parallel in the first phase, producing intermediate results. Then, the middle results are merged and further anonymized to produce consistent k-anonymous data sets in the second phase. We have introduced MapReduce on cloud to data anonymization and deliberately designed a group of innovative MapReduce jobs to concretely accomplish the specialization computation in a highly scalable way. Current experimental results on real-world data sets have demonstrated that with our approach, the scalability and efficiency of TDS are improved significantly over existing approaches. In cloud environment concept, the electric metric for data analysis, share and mining is a challenging research issue due to increasingly larger volumes of data sets, thereby requiring intensive investigation. We will checked the adoption of our approach to the bottom-up generalization algorithms for data anonymization. Based on the contributions herein, we plan to further explore the next step on scalable electric aware analysis and scheduling on large-scale data sets. Optimized balanced scheduling strategies are expected to be developed towards overall scalable electric aware data set scheduling.

## REFERENCES

[1] Rathore, M.M.; Ahmad, A.; Paul, A.; Rho, S. Urban planning and building smart cities based on the Internet of Things using Big Data analytics. Comput. Netw. 2016, 101, 63–80.

[2] Zhu, C.; Leung, V.C.M.; Shu, L.; Ngai, E.C.H. Green Internet of Things for Smart World. IEEE Access 2015, 3, 2151–2162.

[3] Botta, A.; de Donato, W.; Persico, V.; Pescapé, A. Integration of Cloud computing and Internet of Things: A survey. Future Gener. Comput. Syst. 2016, 56, 684–700.

[4] Jaradat, M.; Jarrah, M.; Bousselham, A.; Jararweh, Y.; Al-Ayyoub, M. The Internet of Energy: Smart Sensor Networks and Big Data Management for Smart Grid. Procedia Comput. Sci. 2015, 56, 592–597.

[5] Hancke, G.; Silva, B.; Hancke, G., Jr. The Role of Advanced Sensing in Smart Cities. Sensors 2012, 13, 393–425.

[6] Kyriazis, D.; Varvarigou, T.; White, D.; Rossi, A.; Cooper, J. Sustainable smart city IoT applications: Heat and electricity management amp; Eco-conscious cruise control for public transportation. In Proceedings of the 2013 IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks".

[7] Strategic Opportunity Analysis of the Global Smart City Market. Available online: http://www.egr.msu.edu/~aesc310-web/resources/SmartCities/Smart%20City.

[8] Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Gener. Comput. Syst. 2013, 29, 1645–1660.

[9] Atzori, L.; Iera, A.; Morabito, G. The Internet of Things: A survey. Comput. Netw. 2010, 54, 2787–2805.

[10] Internet of Things in 2020: Roadmap for the Future. Available online: http://www.smart-systemsintegration.org/public/documents/publications/Internet-of-Things_in_2020_EC-EPoSS_Workshop_ Report_2008_v3.pdf (accessed on 24 February 2017).