Big Data Processing Using Virtual Machines in Smart Cities Environment

Sushma. C[1] Satish B Basapur[2]
[1]PG Scholar [2]Assistant Professor
[1,2]Department of Information Science and Engineering
[1,2]Dr.AIT, Bangalore, Karnataka, India

Abstract— Smart city is the means of urban development vision to integrate multiple information and communication technology in a secure fashion to manage information about schools, colleges libraries, transportation systems, hospitals, power plants, water supplies, waste management, law enforcement, and other community services. The goal of smart city is to improve quality of life using technology to improve the efficiency of services and meet resident's needs. Processing and analyzing these data had become a major challenge in the recent emerging technologies. In this paper we explore the big data processing that is data related to smart cities by using the virtual machines. Here we store the data in the virtual machines which is geographically located at different distributed data centers. In order to process the big data in this paper we use machine learning techniques.

Key words: Smart Cities Environment, Big Data Processing

I. INTRODUCTION

In the recent years several governments are initializing to use the big data facilitating the development of the smart cities which had led to the cities to know the principles and requirements for the smart city development by considering some features. These features includes sustainability, resilience, governance, enhancement in the quality of life and management of the natural resources and facilities of the smart city[1]. To use the services provided by the smart cities the organizations requires a very large computations and the storage capabilities. One way to use such facilities is to relay on the big data and its applications currently researches are going on this fields which had led to know the significance of the big data in the applications and services of the smart city.

In the urbanized environment there is a large volumes of different sources of data. Such as enormous mobile sources of data like the smart phones, smart cards sensors etc. Sensors gives information which may lead to know the urban dynamic patterns in the smart city environment. Enormous sensors are installed around the cities most of them are incorporated in the indoor spaces, which may generate the large volume of the data. This has led to find the correct tools and operations to process the machines by using the new methods. The information which is been obtained from the systems which is connected to the internet is not yet been explored widely. The volume of the retrieved data is very large to be significantly used by most current management systems of the cities. Due to this "Large amount of the data with very less information" is a very common problem. The Smart City makes Convenience for the users by providing efficient services. It also contains issues related to the privacy and then security. Two different techniques are utilized to overcome these issues. One method is to prevent the heavy collection of the private data to be secured by using some algorithms like encryption algorithms. The infrastructures of the data plays a very significant part in computing the smart city. Networks such as the wireless sensor networks can be used for retrieving the data and spatial temporal database for the purpose of storing the data

Two important applications have been considered to initiate the effectiveness of computing the smart city which includes Smartgrid and then smart railway transport. By considering the domain of the smart grid, Huang et al, had proposed the framework for integrating the operations related to smart buildings by scheduling the energy for the huge power systems by proactively building the participations of the demand. This method provides buildings to actively explore and communicate their energy consumption by using the smart grid operators instead of retrieving and reacting for the marketing signals and the instructions such as time varied electrical prices. By using the innovative methods and further higher performance algorithms are useful for studying the smart city from the view of computing perspective.

Smart phones plays a major role in the smart city, not only it could be used as the communication device but also can be used for the health assistants, work secretaries and electronic ID. It is very useful to store the data in the smart phones and it can be used anywhere and at anytime, sensitive data can also be placed in the smart phones like account number, passwords, confidential and personal information. Placing the data in the smart phones provides the efficient security for the users. By the increased growth in the Internet of Things and the Internet[2][3] technologies in the context of the smart cities[4][5], a large volume of the data is generated which requires to manage properly for different applications using the standard and the integrated method.

The structured Method contains the general process for finding the steps which should be following by using various approaches, tools and services. These steps may include collecting the data, storing it, processing, analyzing and then generating the applications for the smart city.

Smart city applications produces large volume of the data were big data systems use this data to produce information which increases the efficiency of the smart cities applications. The big data systems stores, processes and then maintains the information about the smart cities in the significant manner to generate information which increases the various services of the Smart City. In this paper we are using machine learning techniques such as the kmeans algorithm in order to process the big data.

All rights reserved by *www.ijsrd.com*

II. RELATED WORK

In the Smart City Environment Security and Privacy contains various aspects, which includes services related to the Security, organization of the sensitive data, Protocol communication, Management of the key and authorization. At the present most commonly used methods for Privacy and Security in the Smart City are obtained from the distributed Systems [6]. These methods produces safe secure and reliable integration of the resources which are distributed. The main example which uses the distributed resources is the Smart grid[7] . These techniques are not sufficient for the smart city, specifically for the smart devices. Since due to the different formats of the data and the communication protocols, Processing of the data had became very complex, which had made the process of encryption very difficult[8].

Citizens in the Smart City places their data in the Smart Devices, Which may lead to the security and privacy issues, Example viruses in the mobile devices and applications, intercepting the packet in the middle of the communication, wrong permission and hacking on the resources like the servers. Due to the presence of the large volume of the data various machine learning techniques are used which aims on scaling the centralized learning algorithm. One method is to decompose the data in to smaller parts and then parallelizing the process of learning by using the processors for each and every parts. This method explains the scenario where the deep learning is used [9],[10],[11]. This method does not concerns on the security and the privacy issues. Other approaches proposes the distributed and the decentralized algorithms.

The authors[12] had presented the distributed method of Support Vector Machine(SVM). In the view to learn the model from the physically distributed data, they uses the iterative method which trains the SVM on each local data set and then they exchanges the learned support vectors with other locations. The support vectors are added to the datasets which are local and then the local SVM are trained again.

The researchers[13] had proposed the distributed learning algorithms for the machines which are arbitrarily connected. In this approach the first phase consists of learning phase which is based on the local data and then the iterative method by which machines are completed with the end model. Another identical solution had been proposed[14] in which two distributed learning algorithms for the links in the networks. One method depends on decentralized consensus algorithm. The other method depends on Alternative Direction Methods of Multiplier both uses the iterative solutions.

III. CHALLENGES AND APPLICATIONS

In the view to process the data which is been gathered by the large number of static and mobile sensors incorporated in the cities, It is important to use various approaches related to the data mining. Data Mining approach has raised as the efficient solution for exploring the large amount of data by finding the patterns and the methods which is helpful for analysing the process by the user. The use of this kind of approaches had been led it to be adaptable for processing the large amount of data, without giving up the efficiency of the computational techniques by preserving the robustness of the results which is been obtained. The use of these techniques for the large amount of the data is known as the Big data.

The incorporation of the Big Data approaches is useful in few areas of the smart cities. For example Managing the public transportation gives the present and the previous description of the human flow at various time periods like the working days and the weekends, In the view to process the data by the increased number of static and the mobile sensors incorporated in the cities

There are many applications of big data serving smart cities:

A. Smart Education

ICT had provided the solution to increase the efficiency of the education processes, effectiveness and productivity by using the smart services for the education which will be flexible and intelligent and gives the better usage of the information, increased control and assessments, greater support for learning throughout the life for all people. The application of the smart education will allow the people to actively engage in learning activities, it also allows to adapt quickly for the increased changes in society and the environment. Further by relaying on the big data gathered in this field and processing it accurately to provide the requested information, this may impact the positive levels in gaining the knowledge, teaching and learning tools to deliver or to acquire the knowledge. Big data in the field of education can be used to investigate the shortages and to increase the study curriculum.

B. Smart Traffic Lights

One of the important aspect of the smart city is to effectively control the traffic flow in the city which will increase the efficiency of maintaining the transport systems and improves the traffic patterns in the cities. When there is the growth in the population then the problems related to the related to the traffic, pollution and problems related to the economic arises. Because of this the use of smart traffic lights and signals plays a important role in the smart cities. Smart traffic lights and signals should be interconnected across the traffic grids to provide more information regarding the traffic patterns. Every sensors find the various parameter for the traffic flow. The system takes decisions by using the valves obtained from these parameters and provides the appropriate instructions for the lights and signals.

C. Smart Grid

Smart Grid is one of the important component of the smart city. It is the electrical grid system that uses the information and communication technology to gather and process the available data, like the data regarding the characteristics of the suppliers and consumers, in the automated manner to add some values. It increases the efficiency, reliability economics and sustainability of the production and distribution of the electric power. Smart grid uses the remote controls by using the computer with double sided communication technology between the power producers and consumers to increase the efficiency and reliability of the grid by using the system.

All rights reserved by www.ijsrd.com

IV. TECHNIQUES TO PROCESS THE BIG DATA

In order to process the big data in this paper we explore the machine learning techniques such as the Kmeans algorithm. Machine learning is the kind of artificial intelligence that gives the computers with the capability to learn without being explicitly programmed. Machine Learning mainly aims on the development of the computer programs that can learn themselves to grow and change accordingly when it is exposed to the new data. The technique of the machine learning is identical to that of the data mining. Machine learning makes the Usage of the data to find the patterns in the data and to adjust the program actions accordingly.In this paper we use virtual machines in order to process big data these virtual machines are located in data centers which is located at different geographic regions. Here we use Kmeans Clustering algorithm which is based on machine learning technique here virtual machines are clustered and data is processed according to the Kmeans algorithm.

Kmeans is one of the simplest and the best Known learning algorithms, and can be used for a variety of machine learning tasks, such as detecting abnormal data, clustering of text documents, and analysis of a dataset prior to using other classification of regression methods. In the Kmeans algorithm we don't need to know any of the cluster categories in advance this algorithm will find possible categories based solely on the data. In general this clustering uses iterative techniques to group a dataset in to clusters that contain similar characteristics. These groupings are useful for exploring data, identifying anamolies in the data, and eventually for making prediction. Clustering models also can help to identify relationship in a dataset.

V. CONCLUSION

In this paper we explore the information about the smart cities and how smart cities benefits the citizens and how the information is gathered and collected in the form of big data and we explore the previous techniques to process the big data and we had done a literature survey of the challenges of the big data and the applications of the smart cities and how big data is used to process the information of the smart cities. How the big data is used to process the information of the smart cities. In this paper we use the virtual machines in order to process the big data here the virtual machines are clustered and we use machine learning techniques such as Kmeans algorithm in order to process the data among the clusters of the virtual machines.

REFERENCES

[1] Khan Z, Anjum A, Kiani SL. Cloud Based Big Data Analytics for Smart Future Cities. In Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing. IEEE Computer Society; 2013. pp. 381–386.
[2] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," Future Gener. Compute. Syst., vol. 29, no. 7, pp. 1645–1660, Sep. 2013.
[3] D. Bandyopadhyay and J. Sen., "Internet of things: Applications and challenges in technology and standardization," Wirel. Pers. Commun., vol. 58, no. 1, pp. 49–69, May 2011.
[4] D. Ludlow and Z. Khan, "Participatory democracy and the governance of smart cities," in Proceedings of the 26th Annual AESOP Congress, Ankara, Turkey, July 2012.
[5] Z. Khan and S. Kiani, "A cloud-based architecture for citizen services in smart cities," in ITAAC Workshop 2012, Utility and Cloud Computing (UCC), 2012 IEEE Fifth International Conference on, 2012, pp. 315– 320.
[6] J. Blom, D. Viswanathan, M. Spasojevic, J. Go, K. Acharya, and R. Ahonius, "Fear and the city: role of mobile services in harnessing safety and security in urban use contexts," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010, pp. 1841–1850.
[7] A. Paverd, A. Martin, and I. Brown, "Security and privacy in smart grid demand response systems," in Smart Grid Security. Springer, 2014, pp. 1–15.
[8] D. Damopoulos, G. Kambourakis, M. Anagnostopoulos, S. Gritzalis, and J. Park, "User privacy and modern mobile services: are they on the same path" Personal and ubiquitous computing, vol. 17, no. 7, pp. 1437–1448, 2013.
[9] A. Bordes, S. Ertekin, J. Weston, and L. Bottou, "Fast kernel classifiers with online and active learning," J. Mach. Learn. Res., vol. 6, pp. 1579–1619, Dec. 2005.
[10] J. Dean, et al. "Large scale distributed deep networks," in NIPS, 2012, pp. 1223–1231.
[11] A. Coates, et al., "Deep learning with cots hpc systems," JLMR, vol. 28, no. 3, pp. 1337–1345, 2013.
[12] A. Navia-V´azquez and E. Parrado-Hernandez, "Distributed support vector machines," IEEE TNN, vol. 17, no. 4, pp. 1091–1097, 2006.
[13] L. Georgopoulos and M. Hasler, "Distributed machine learning in networks by consensus," Neuro computing, vol. 124, pp. 2 – 12, 2014.
[14] S. Scardapane, et al.,"Distributed learning for random vector functional- link networks," Information Sciences, vol. 301, pp. 271 – 284, 2015.

All rights reserved by *www.ijsrd.com*