

Large Scale Spatial Data Management on Mobile Phone data set Using Exploratory Data Analysis

K.S.Pandiyar, Dr.S.Sivakumar

Research scholar, Professor, Department of Geology,
Anna University, Chennai

Abstract — Now a days the mobile phone available in the market has enabled with the GPS. The mobile phones enhanced with the GPS will generate large amount of GPS trajectories. There is chance to derive a lot of information from that spatial data such as position of the user, transportation planning etc. There are lots of data analysis techniques, methods and data collection techniques are available. The challenges among them are which method is choosing for your application. This paper proposes the exploratory data analysis method combined with hadoop cloud computing software framework. It able to handle billion of data records which are collects from the spatial.

Keywords— GPS, Mobile Phone, Spatial Analysis, Cloud Computing, Hadoop.

I.INTRODUCTION

With the improvement of technology in communication and information technology such as GPS and GPRS. The GPS system provides service to the people to identify their location. The location information are not only gathered by the geo-location also from the base stations and cell of mobile. With the help of GPS at the end of call the call details of the customer are sending to the customers. Data from the mobile phones are larger and continuously increased. So that the data management of mobile phone records need a reliable and scalable storage and management system. For providing data in timely and scalable manner to the millions of people we need a cloud system. The clients of the facebook and google are managed and served by cloud system. The cloud computing technology plays a important role in data mining area. The cloud system provides several services such as software service, platform service, data storage, computing service and elasticity of data to the users. The remaining portion of the paper describes the concept of GPS system, available data mining techniques, Cloud computing technology with hadoop software framework and system overview.

II .OPERATION OF GLOBAL POSITIONING SYSTEM

The global positioning system is generally used for navigation with help of satellites which are in orbits. The GPS satellite rotates the earth twice a day and cover the whole earth. The GPS satellite sends the information to the earth station and this information is used to identify the location of the users. The GPS receiver is used to receive the signal; the receiver must at least receive two or three satellite signals. The GPS satellites will send the message to the GPS receiver continuously; the message contains signal transmitted time and the position at the time of where the message is from and GPS receiver computes the distance of the satellite position using this message.

Fig.1 Operation of GPS



The current GPS system includes three segments user segment, control segment and space segment. Space segment consist of satellites which around 30 satellites on medium earth satellite orbit.

Some of the important application of GPS is as follows

- Aviation
- Road transport
- Science
- Shipping and rail transport
- Heavy vehicle guidance
- Security
- Survey, mapping and geo physics
- Telecommunications and social activities

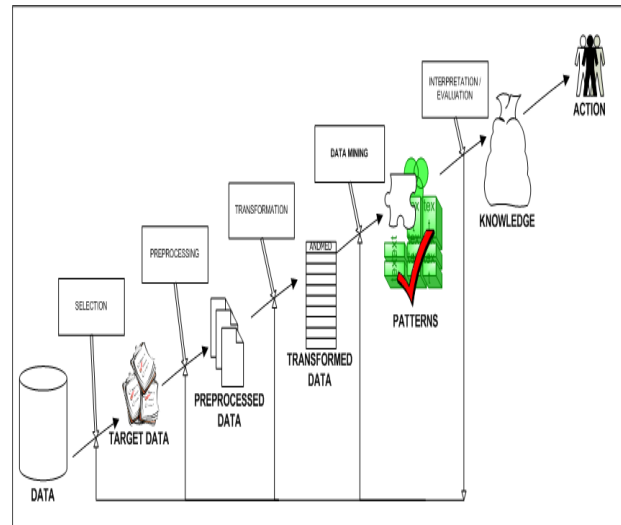
III. DATA MINING TECHNIQUE

The data mining is the process of gathering data from the different sources and analyzing the data and providing the useful information to the clients. For preceding the data mining methods we must know how to relate, map, associate and cluster the data from different sources. Some of the important data mining techniques are association, classification, prediction, clustering, decision tree, sequential pattern and combination. Association is most commonly used data mining technique to classify the two data items. The classification method provides a better data classification than association; which provides the classification based on what kind of data it is and some other attributes of gathered data. Clustering method is used to grouping the similar data by examining one or more attributes of data. The clustering technique is commonly used for big data analysis and management. Clustering can be used to derive different information by identifying the correlation between them, so it is well suitable for the large scale spatial data management on mobile phones.

The data analysis is commonly based on prediction and classification, for example the expectation of the marketing profit can be predicted by analyzing the collected data. Prediction is used with the combination of other data mining technique classification, relation and pattern matching. These methods predict the future by analyzing the past scenarios of the data.

Sequential pattern which is used to identify the frequently occurred or used data. For example the customer purchase a similar type of bath soap in a shopping at different months, so that it can be predicted and placed in a customer shopping cart.

Fig.2 Data Mining Process Overview



IV. CLOUD COMPUTING

The word cloud computing refers internet or network which is commonly represented as a cloud symbol. It provides storage and data access over the internet by remotely login with the cloud; you need only an internet connection. The cloud can be designed as four models and provides three services. The four model of cloud configuration are as follows

- Public Cloud
- Private cloud
- Community cloud
- Hybrid cloud

With the private cloud the file access and sharing are done only within the given organization, which is monitored and controlled by the cloud administrator. But the public cloud used as an open source, which has contradiction with private cloud which is accessible publicly any one can use the cloud services, here is somewhat security issues than private cloud. Community cloud is somewhat similar to the private cloud; the data access and sharing are doing among the specific groups. Hybrid cloud is a combination private, public and community cloud. The services that are provided by the cloud are

- Software as a service
- Platform as a service
- Infrastructure as a service

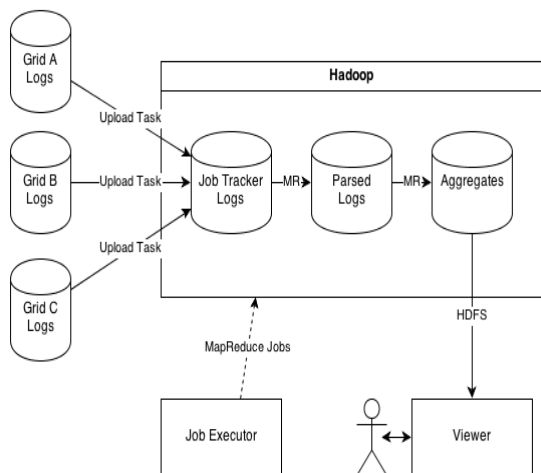
V. SYSTEM OVERVIEW

This paper proposes the new design for managing large scale spatial data of mobile phone data, which is going to be designed by using cloud environment. For processing and managing large

amount of data and for clients to accessing their data and navigation information timely with more scalable the cloud computing is used here. The reason for choosing cloud for building this architecture is; it globalize your access securely, more virtualization, elasticity, extend the scalability, streamlined process, no need to install or purchase the apps.

The spatial data management need more scalability and constrained access of information so that the cloud is going to combined with the data mining model. This research suggested to used the Hadoop cloud computing platform to manage huge amount of data. The components of the system are data receiver, data processing, data sources, permanent storage, and spatial information access interface. The data are directly collected by the mobile devices through air and directly received by the data receiver, the data receiver receives the data and forward the data in to appropriate manner. The received data are furtherly processed and the information is accessed by the users via spatial data interface applications. The main core of the system is Hadoop cloud computing environment which is used by google and facebook to maintain the million of clients at a same time.

Fig.3 System Architecture



VI. CONCLUSION

The GPS play an important role in much application to identify the location and navigation of path. This kind of data are collected as a data base and provide a prediction of navigation and deriving some other knowledge based information to the users is the main aim of this paper.

This can be achieved and implemented using the Hadoop cloud environment, which provides more scalable and secure information storage and retrieval. In future there may be some other application or data storage method is used to reduce the computational cost, cost of maintenance and security issues. The future work is to keep searching and improve this technology better and better.

REFERENCES

1. IBM, "Data mining techniques" 11 dec 2012.
2. Leon S., Ouri W., Philip S. Y., Bo Xu., 2011. Transportation Mode Detection using Mobile Phones and GIS Information. ACM SIGSPATIAL, Chicago, USA.
3. Tutorialspoint.com, "Cloud computing tutorial"
4. Jansen, Wayne & Grance, Timothy. Guidelines on Security and Privacy in Public Cloud Computing. National Institute of Standards and Technology, 2011.
5. Strowd, Harrison & Lewis, Grace. T-Check in System-of-Systems Technologies: Cloud Computing (CMU/SEI-2010-TN-009). Software Engineering Institute, Carnegie Mellon University, 2010. <http://www.sei.cmu.edu/library/abstracts/reports/10tm009.cfm>
6. "Cloud Computing: Clash of the clouds". *The Economist*. 2009-10-15. Retrieved 2009-11-03.
7. "Gartner Says Cloud Computing Will Be As Influential As E-business". Gartner. Retrieved 2010-08-22.
8. Gruman, Galen (2008-04-07). "What cloud computing really means". *InfoWorld*. Retrieved 2009-06-02.
9. "The economy is flat so why are financials Cloud vendors growing at more than 90 percent per annum?". FSN. March 5, 2013.
10. Figure 8, "A network 70 is shown schematically as a cloud", US Patent 5,485,455, column 17, line 22, filed Jan 28, 1994.
11. Yang, J., Wu, S., 2010. Studies on Application of Cloud Computing Techniques in GIS. In: proceeding of The International Conference on Geoscience and Remote Sensing, China, Vol.2, pp. 492-495.