



Safety Platform of Traffic in Cloud Computing Environment

Chetan.R¹, Ranjith .J², Umesh.M³, Usha N⁴

Assistant Professor, Dept. of ISE, SJBIT, Bangalore, Karnataka, India^{1,2,3,4}

ABSTRACT: The ever-increasing road traffic collisions, improved gathering, analyzing and tackling constructive information in view of traffic safety has become one of most vital concerns for local traffic authorities. The ultimate goal is to guarantee sound traffic management and rapid access to updated road traffic information for city travelers. Therefore, an effort to devise a traffic safety information platform based on cloud computing technology, which serves as an outstanding channel for message between traffic authority, staff and the public. This platform relies on data gathering & storing using Big data methodologies, analyzing the data using Data warehousing, and mining model, eagerly personalized to present traffic environment and trustworthy data from traffic accidents. The platform is completely integrated with matching variables such as people, vehicles, and environment, weather and road conditions.

KEYWORDS: data mining, cloud computing, data warehousing, clustering

I. INTRODUCTION

A. Overview of traffic safety

According to the statistics published by World Health Organization (WHO), annual road traffic deaths and injuries around the world has reached 1.2 million and 50 million respectively, which regularly resulted in everlasting damages. With the modern development of information technology and the eternally growing threat to life from road traffic collisions, it has become one of most vital concerns for local traffic authorities to better gather, analyze and tackle helpful information in view of traffic safety with a vision to ensure sound traffic management and for city travelers to get speedy access to updated road traffic information in the expect of guiding their travel behaviors as well [2]. Therefore, an endeavor to design a traffic safety information management platform based on cloud computing, data warehousing and data mining, which serves as an outstanding channel for communication between traffic authority, staff and the public. It aims to obtain full benefit of road safety information database in effort to simplify, judge and predict different safety information such as traffic accident black spots, vehicle types legally responsible to serious accidents and weather based on already occurred accidents on one hand and share detailed traffic conditions without hindrance such as road condition, potential risks, disasters and weather [21].

B. Data Warehouse

Data warehouse is a database used for reporting and data analysis. Integrating data from one or more contrasting sources creates a middle storehouse of data. It store existing and historical data and are used for creating, trending reports for senior management reporting such as yearly and quarterly comparisons. The data stored in the warehouse is uploaded from the operational systems. The typical extract transform load (ETL) based data warehouse uses staging, data integration, and access layers to house its key functions. The staging layer or staging database stores raw data extracted from each of the dissimilar source data systems. The integration layer integrates the dissimilar data sets by transforming the data from the staging layer often storing this transformed data in an operational data store (ODS) database. The integrated data are then moved to yet another database, often called the data warehouse database, where the data is arranged into hierarchical groups often called dimensions and into facts and aggregate facts. The access layer helps users retrieve data. A data warehouse constructed from an integrated data source system does not require ETL, staging databases, or operational data store databases. This integrated data warehouse architecture supports the drill down from the aggregate data of the data warehouse to the transactional data of the integrated source data systems. The data warehouse focuses on data storage. The main source of the data is cleaned, transformed, catalogued and made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support. However, the means to retrieve and analyze data, to extract, transform and load data, and



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

to manage the data dictionary, business intelligence tools, tools to extract data into the repository, and tools to manage and retrieve metadata are also considered important components of a data warehousing system.

C. Data Mining

Data mining is the computational procedure of discovering patterns in huge data sets involving methods at the intersection of artificial, machine learning, statistics, and database systems. The general objective of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. apart from the raw analysis stage, it involves database & data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. The real data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unfamiliar remarkable patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). These patterns can then be seen as a kind of summary of the input data and may be used in further analysis.

Data mining involves six common classes of tasks

- Anomaly detection (Outlier/change/deviation detection) – The detection of unusual data records, that might be interesting or data errors that require additional investigation.
- Association rule learning (Dependency modeling) – Searches for relationships between variables.
- Clustering – is the job of discovering groups and structures in the data that are in some way or another "similar", without using recognized structures in the data.
- Classification – is the task of generalizing known structure to apply to new data.
- Regression – attempts to find a function which models the data with the least error.
- Summarization – providing a more dense representation of the data set, including visualization and report generation.

II. LITERATURE SURVEY

Hadoop [22] was created by Doug Cutting, the creator of Apache Lucene, the widely used text search library. Hadoop has its origins in Apache Nutch, an open source web search engine, itself a part of the Lucene project. Mike Cafarella and Doug Cutting estimated a system supporting a billion page index would cost around half a million dollars in hardware, with a monthly running cost of \$30,000.

Jana and Bandyopadhyay [3] analyzed several security threats and its measures, and suggested additional recommendations on top of best practices adopted for Identity and Access Management(IAM) for mobile cloud users. Pirker et al [4] presented privacy preserving cloud resource payment to enable mobile clients to anonymously consume resources of a cloud service provider such that the provider is not able to track users' activity patterns.

Satyanarayanan [5] discussed a wide range of issues in areas such as privacy, software licensing, and business models with the emergence of cloudlet based hardware/software ecosystem to support for example, cognitive assistance for attention challenged mobile users, scalable crowd sourcing of first person video, and ubiquitous mobile access to one's legacy world.

Abuelela and Olariu[6] envision VANETs as cloud computing networks as well. Vehicles will share the capability of computing power, Internet access and storage to form conventional clouds

Santos et al. [7] proposed a new platform to achieve trust in conventional clouds. Krautheim [8] also proposed a third party to share the responsibility of security in cloud computing between the service provider and client, decreasing the risk exposure to both. The bilinear aggregate signature has been extended to simultaneously audit multiple users.

Observing the fact that many devices (computing, sensing and storing devices) on vehicles are idle for a long while, Olariu and his colleagues [9], [10], [11] proposed to share these devices as the computational engine of the cloud. Ristenpart et al. [12] presented experiments of locating co-residence of other users in cloud virtual machines.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

The mobile cloud application models are based on the standard cloud service model that includes Infrastructure as a Service (IaaS) [14], Platform as a Service (PaaS) [15], and Software as a Service (SaaS) [16], [17]. Therefore, based on the working of the application models, any of these service layers can be utilized. Some of the well known services for mobile cloud computing include Amazon Elastic Compute Cloud (EC2) [18], GoogleApp Engine [15], and Microsoft Azure [19].

III. THE TECHNOLOGIES AND CONCEPTS

A. Cloud Computing

Cloud Computing can be used as a platform and one type of application program abreast of the latest computing development such as distributed computing, parallel computing and grid computing. It uses a computer network to offer computing resources such as data or software as a service to users who pay for it on demand. Thus, it breaks free of the computing power and storage space margins of the traditional local computing model. User’s computers, mobile phones, and other private devices might only contain an operation system and an Internet explorer, and they do not have to identify where the data is stored or who offer the software. Users present their computing tasks that cannot be skilled by their local devices to the clouds. The clouds offer computing services and return the computing results to them.

Cloud computing benefits from several key characteristics

- *Reliability* The data is stored and the applications are running on the servers in the clouds. Users do not have to worry about lost or corrupt data.
- *Agility* The clouds can distribute computing resources according to the user’s needs or preferences to provide flexible management.
- *Utility*. Users do not have to buy expensive computing devices. They only need to pay for the computing services provided by the clouds.
- *Application programming interface (API)* accessibility to software that enables machines to interact with cloud software in the same way that a traditional user interface facilitates interaction between humans and computers. Cloud computing systems typically use Representational State Transfer (REST) based APIs.
- *Device and location independence* facilitate users to access systems using a web browser despite of their location or what device they use. As infrastructure is off site and accessed via the Internet, users can connect from anywhere.
- *Multitenancy* enables sharing of resources and costs across a large pool of users thus allowing for Centralization of infrastructure in locations with lower costs, Peak load capacity increases & Utilisation and efficiency improvements for systems that are often only 10–20% utilised.
- *Performance* is monitored and consistent and loosely coupled architectures are constructed using web services as the system interface.
- *Security* can get better due to centralization of data, improved security focused resources, etc., but concerns can stick with about loss of control over definite perceptive data, and the lack of security for stored kernels.

Cloud Computing Drivers

Table 1. Attributes which draws organization towards Cloud Computing

Attribute	Why It Can Draw an Organization Toward Cloud Computing
Availability	Users have the ability to access their resources at any time through a standard internet connection.
Collaboration	Users begin to see the cloud as a way to work simultaneously on common data and information.
Elasticity	The provider transparently manages a user’s resource utilization based on dynamically changing needs.
Lower Infrastructure Costs	The pay-per-usage model allows an organization to only pay for the resources they need with basically no investment in the physical resources available in the cloud. There are no infrastructure maintenance or upgrade costs.
Mobility	Users have the ability to access data and applications from around the globe.
Risk Reduction	Organizations can use the cloud to test ideas and concepts before making major investments in technology.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

Scalability	Users have access to a large amount of resources that scale based on their demand.
Virtualization	Each user has a single view of the available resources, independently of how they are orderly in terms of physical devices.

B. Analyzing the Data with Hadoop

Hadoop

As infrastructure of distributed system, Users can explore the distributed programs and make full use of cluster of high-speed process and storage without shrewd details about the distributed architecture substrates [22]. Hadoop is a group of associated subprojects that descend beneath the umbrella of infrastructure for distributed computing. These projects are hosted by the Apache Software Foundation, which provides support for a community of open source software projects. To take advantage of the parallel processing that Hadoop provides, it is essential to express query as a Map Reduce job. After some local, small-scale testing will be able to run it on a cluster of machines.

Map and Reduce

Map Reduce is a programming model for data processing. The model is simple, yet not too simple to express useful programs in. Map Reduce works by breaking the processing into two phases: the map phase and the reduce phase. Each phase has key-value pairs as input and output, the types of which may be chosen by the programmer. The programmer also specifies two functions: the map function and the reduce function. The input to our map phase is the raw data, which choose a text input format that gives each line in the dataset as a text value. The key is the offset of the start of the line from the beginning of the file. The map function is just a data preparation phase, setting up the data in such a mode that the reducer function can do its work on it. The map function is also a good place to drop bad records. As can be seen from the Fig. 1., it contains a lot of components.



Fig.1. Hadoop component

Avro

A data serialization system for efficient, cross language RPC, and relentless data storage. (At the time of this writing, Avro had been created only as a new subproject, and no other Hadoop subprojects were using it yet.)

HDFS

Hadoop consists of distributed filesystem called HDFS, which stands for Hadoop Distributed Filesystem. A distributed filesystem that runs on huge clusters of product machines. HDFS is a filesystem intended for storing very large files with streaming data access patterns, running on clusters on commodity hardware.

Pig

A data flow language and execution environment for exploring very big datasets. Pig runs on HDFS and Map/Reduce clusters.

HBase

HBase is an open source, non-relational, distributed data base modeled after Google's BigTable and is written in Java. HBase uses HDFS for its fundamental storage, and supports both batch style computations using Map/Reduce and point queries (random reads). It is developed as part of Apache Software Foundation's Apache Hadoop project and runs on apex of HDFS, providing BigTable like capabilities for Hadoop.

ZooKeeper

A distributed, highly available coordination service. ZooKeeper provides primitive such as distributed locks that can be used for building distributed applications.

Hive and Mahout

Hive is a data warehouse platform based on Hadoop. By utilizing Hive, data extraction, transformation and loading can easily realize. Furthermore, it's potential to convert the program from QL to corresponding Map/reduce and base them on Hadoop by HQL (*Hive Query Language*). Mahout is a distributed framework of machine learning and data mining provide some traditional algorithm in the extensible ground of machine learning.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

IV. PROBLEMS AND CONSTRUCTION MODE

A. Problems

Public traffic information service has been in the interest worldwide. But, there is still a batch of weaknesses in this area:

- Limited information service ability.
At present the majority information services rely on government and information gather, tackling & releasing are self-governing to several extent. What's added information content seems to be unsophisticated and unusable.
- Drawback of service content, approach, eminence and range.
The present service content chiefly touches upon travel information and service approach is considered to be obsolete and thus put down much room for development.
- Lack of frequently updated traffic safety information.
At present time, the significant traffic safety information is partial for information service provider & it's really sturdy congregation, storing and processing group information.

For the given problems above, an attempt to generate a traffic safety information platform based on cloud computing for local traffic authorities and the public through establishing traffic safety information system. With the platform any users can attach cloud service platform with their own terminal apparatus and keep them well educated of any traffic safety information they are concerned about.

B. The main research content

An effort to design a traffic safety information management platform based on cloud computing, data warehousing and data mining, which serves as an outstanding feed for communication between traffic authority, workers and the public [23]. It aims to take complete benefit of road safety information database in effort to simplify, critic and envisage different safety information such as traffic accident black-spots, vehicle types liable to severe accidents and climate based on previously occurred accidents on one hand and collect, deal with and divide complete traffic circumstances without delay such as road condition, potential risks, accidents, disasters and weather on the other hand. The key problem requests tackling.

There are three key technologies:

- Cloud storage of huge traffic data
- precise analysis and proposal of massive traffic information
- Cloud terminal technology

Support different mobile terminal including mobile phones, ipad and personalized cloud terminal in various interactive forms so that users can use cloud services anytime and anywhere and distribute traffic safety information.

C. Architecture design of intelligent traffic safety information platform based on cloud computing.

Based on the responsiveness and understanding of above problem, attempt to design the following system structure. Figure 2. Briefly explained as follows. Those bunch data gathered by each branch system grant support for isolating users in the field of traffic safety. In most cases driver's casual behaviors often result in uneven pattern, which is further suitable for cluster analysis.

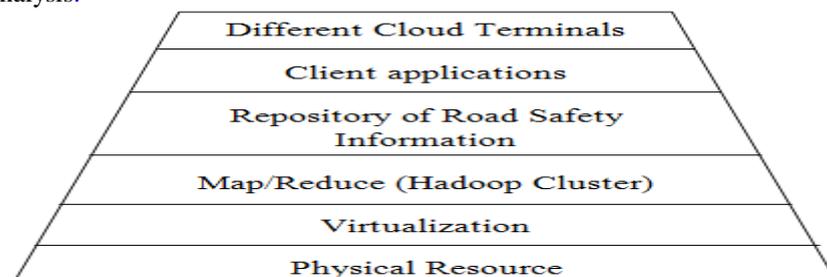


Fig.2. Architecture design of intelligent traffic safety information platform based on cloud computing

Due to the large scale of user data, we choose to take up Hadoop for parallel computing and distributed storage. Moreover, to advance the consumption rate of resources we have realized the virtualization of server layer, built traffic



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

information security awareness based on all kinds of data mining algorithm with Mahout and blueprint mobile terminal access for solving all kinds of cloud terminal access in the application layer.

D. Algorithm usage case

There is no extensively accepted effective algorithm on report of incomplete statistical data so far. It is an attempt to confine influencing factors of road safety in a grey realm by adopting clustering algorithm [24] and estimate overall level of road traffic safety based on information screening, processing, annex and extension.

The procedures will be as follows:

- Format conversion: Data format have to be transformed into the input format that can be handled by clustering algorithm. What clustering algorithm can deal directly with the format is sequence file in Mahout. So we should write a Class which is used to convert the format into sequence file file:

InputMapper and a Map function to implement InputMapper.

Map function is defined as follows:

```
“Public void Map (LongWritable key,  
Text values,  
OutputCollector<Text, VectorWritable>Output,  
Reporter reporter)”
```

- Invoke Apache Mahouts
- Invoke parallel clustering algorithm of Mahout, which contains a couple of clustering algorithm. K - means clustering algorithm is a case in point.
- Obtain the clustering results of HDFS
- Analyze the clustering results, namely obtain the results directly in HDFS by Hive and extract them to local client for analysis.

V. CONCLUSION

It is not so simple to begin from graze and thus call for additional analysis and exploration. Its purpose in the traffic field also depends on the expansion of related areas such as IT and logistics. Due to its potential theoretical and practical values in the field of traffic management, cloud computing has aroused wide concern among researchers. The expansion and application of cloud computing in different vital industries can also make possible its progress in traffic field.

REFERENCES

- [1] Shi J., Li, X. (2011). Research on Traffic Information Cloud Computing and Its Application. *Journal of Transportation Systems Engineering and Information Technology*,(01):179-184.
- [2] Zeng, K., Yan, J. (2011). Cloud Computing and Its Application in Intelligent Transportation. *Modern Science &Technology of Telecommunication*,(05):45-51.
- [3] Debasish Jana and Debasis Bandyopadhyay, "Management of Identity and Credentials in Mobile Cloud Environment", *Proceedings of the 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS 2013)*, September 28-29, 2013, Bali, Indonesia
- [4] Martin Pirker, Daniel Slamanig, and Johannes Winter. 2012. Practical privacy preserving cloud resource-payment for constrained clients. In *Proceedings of the 12th international conference on Privacy Enhancing Technologies (PETS'12)*, Simone Fischer-Hübner and Matthew Wright (Eds.). Springer-Verlag, Berlin, Heidelberg, 201-220.
- [5] Mahadev Satyanarayanan. 2013. Cloudlets: at the leading edge of cloudmobile convergence. In *Proceedings of the 9th international ACM Sigsoft conference on Quality of software architectures (QoSA '13)*. ACM, New York, NY, USA, 1-2.
- [6] M. Abuelela and S. Olariu, "Taking vanet to the clouds," *Proceedings of The 8th International Conference on Advances in Mobile Computing and Multimedia MoMM 2010*, pp. 8–10, 2010.
- [7] N. Santos, K. P. Gummadi, and R. Rodrigues, "Towards trusted cloud computing," in *Proceedings of HotCloud*, June 2009.
- [8] F. J. Krauthem, "Private virtual infrastructure for cloud computing," 2009 conference on Hot topics in cloud computing, pp. 1–5, 2009.
- [9] M. Abuelela and S. Olariu, "Taking vanet to the clouds," *Proceedings of The 8th International Conference on Advances in Mobile Computing and Multimedia MoMM 2010*, pp. 8–10, 2010.
- [10] M. Eltoweissy, S. Olariu, and M. Younis, "Towards autonomous vehicular clouds," in *Proceedings of AdHocNets'2010*, Victoria, BC, Canada, August 2010.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

- [11] S. Olariu, I. Khalil, and M. Abuelela, "Taking vanet to the clouds," *International Journal of Pervasive Computing and Communication*, vol. 7, no. 1, pp. 7–21, 2011.
- [12] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds," in *Proceedings of the 16th ACM conference on Computer and communications security*, ser. CCS '09, 2009, pp. 199–212.
- [13] Thomas Ristenpart, Eran Tromer, Hovav Shacham, and Stefan Savage. 2009. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In *Proceedings of the 16th ACM conference on Computer and communications security (CCS '09)*. ACM, New York, NY, USA, 199-212.
- [14] Rackspacecloud. <http://www.rackspace.com/cloud>, Accessed April 10th, 2012.
- [15] Google app engine. <https://cloud.google.com/products/app-engineappengine-search> Accessed November 15th, 2011.
- [16] Google apps for business. <https://www.google.com/enterprise/apps/business/products.html> Accessed April 10th, 2012.
- [17] Salesforce. <http://www.zdnet.com/why-salesforce-integrator-rework-is-recasting-itself-as-a-cloud-broker-7000027448/> Accessed April 10th, 2012.
- [18] Amazon elastic compute cloud (ec2), <https://www.google.co.in/url>. Accessed December 10th, 2011.
- [19] Microsoft azure. <http://www.localmoxie.com/web.php> Accessed April 10th, 2012. [Online].
- [20] Basics About Cloud Computing, *Grace Lewis*, September 2010.
- [21] Zheng, X. (2012). Application of Cloud Computing in the Intelligent Transportation System in the Future. *Guangxi Journal of Light Industry*,(03):88-89.
- [22] White, T. (2011). Hadoop: The Definitive Guide. : 30-32 About Mahout <https://wiki.apache.org/MAHOUT/quickstart.html>
- [23] Gray Evaluation Method Based on the Relative Merits of the Clustering of Road Traffic Safety. *Road Traffic & Safety*. 2006(11)
- [24] Jaworski, P., Edwards, T., Moore, J., Burnham, K. (2011). Cloud Computing Concept for Intelligent Transportation Systems. 2011 14th International IEEE Conference on Intelligent Transportation Systems. 2011: 931-936

BIOGRAPHY

Mr. Chetan R is a Assistant Professor in the Information Science and Engineering Department, SJB Institute of Technology Bangalore, VTU University. He received Master of Technology (M.Tech) degree in 2010 from PESIT, Bangalore, Karnataka, India. His research interests are Big Data, Data Mining, Cloud Computing, Web Mining etc.

Mr. Ranjith J is a Assistant Professor in the Information Science and Engineering Department, SJB Institute of Technology Bangalore, VTU University. He received Master of Technology (M.Tech) degree in 2008 from NIE, Mysore, Karnataka, India. His research interests are Computer Networks (wireless Networks, Ad-Hoc), Data Mining, Cloud Computing, Storage Area Network etc.

Mr. Umesh M is a Assistant Professor in the Information Science and Engineering Department, SJB Institute of Technology Bangalore, VTU University. He received Master of Technology (M.Tech) degree in 2011 from Dr.AIT, Bangalore, Karnataka, India. His research interests are Software Testing, Software Engineering, Computer Networks (wireless Networks), Data Mining, Cloud Computing etc.

Mr. Usha N is a Lecturer in the Information Science and Engineering Department, SJB Institute of Technology Bangalore, VTU University. She is pursuing her Master of Technology (M.Tech) degree from RVCE, Bangalore, Karnataka, India. Her research interests are Process Migration, Operating System, Computer Networks (wireless Networks), Cloud Computing etc.