

A Lesson Learned of Cloud Computing for Fleet Management in Taiwan

Wei-Yen Lin¹

*Geographic Information System Research Center, Feng Chia University
100, Wenhwa Rd., Taichung, Taiwan
E-mail: william@gis.tw*

Wen-Yuan Ku

*Geographic Information System Research Center, Feng Chia University
100, Wenhwa Rd., Taichung, Taiwan
E-mail: cool@gis.tw*

Chien-Fu Chen

*Geographic Information System Research Center, Feng Chia University
100, Wenhwa Rd., Taichung, Taiwan
E-mail: jeff@gis.tw*

Jackie C. Y. Mu

*Geographic Information System Research Center, Feng Chia University
100, Wenhwa Rd., Taichung, Taiwan
E-mail: jackie@gis.tw*

In recent years, GPS-based fleet management systems (FMS) have become a powerful tool for fleet companies to better manage their transportation resources in a more efficient and economic manner. Vehicles equipped with a GPS-enabled car kit periodically report their positions to the control center via wireless communication every 30 seconds or so, resulting in rapid accumulation of records in databases and enormous demand on computing resources.

SkyEyes, one of leading FMS service providers in Taiwan, has currently established an online commercial fleet with up to 1,500 vehicles. To fulfill the potential market expansion in the future, SkyEyes began a pilot study of introducing Apache Hadoop as the cloud computing platform to enhance the performance and scalability of the current architecture. This paper re-examines the track of study and evaluate the system performance, providing a lesson learned in fleet management industry for future research.

*The International Symposium on Grids and Clouds and the Open Grid Forum
Academia Sinica, Taipei, Taiwan
March 19 - 25, 2011*

¹ Speaker

1. Introduction

Over the past few years when GPS and digital maps are much available and affordable to the public, fleet management systems (FMS) have become a powerful tool for both commercial fleet companies and government sectors with public service vehicles to better manage their transportation resources in a more efficient and economic manner [1, 2, 5]. In recent years, more and more fleet companies have further adopted a web-based FMS, rather than premise-based solutions as before, to dispense the investment on internal IT staff and disregard annoyance induced by frequent updates and upgrades of the software [8].

Most fleet management systems collect real-time position data from their vehicles equipped with GPS-enabled car kits every 30 seconds or so, which results in the rapid accumulation of records in databases and enormous demand on computing resources to process those spatial data while the number of vehicles grows. Hence, how to fulfill the massive computing demand and meanwhile remain scalable has turned into an important issue for SaaS fleet management [1, 3].

SkyEyes GPS Technology Co., one of leading FMS service providers in Taiwan, has currently established an online commercial fleet with up to 1,500 vehicles [9]. For a typical work day with 12 service hours, the database has to accommodate around 2 million records each day in order to render the daily vehicle trail on the map upon the request by fleet managers, or 0.18 billion records every 3 months for fleet company executive officers to produce managerial reports for business analysis. For near real-time monitoring, it requires a great amount of CPU time for spatial and analytical computation; for example, mapping geographic coordinates into map features, such as roads, factories, or administrative zones, with surrounding buffers for Georeferencing or Geofencing [8].

To fulfill the potential market expansion in the future, SkyEyes began a pilot study of introducing cloud computing to enhance the performance and scalability of the current architecture, with an ultimate goal of shifting the entire system onto the 'Cloud'. In its early stage of testing, we imported the 8-month historic vehicle trails and driving status data into HBase [6], a non-relational, distributed database modeled in Java for Hadoop [7], a software framework that supports data-intensive distributed applications. Under Hadoop architecture, we implemented a socket server to take in real-time vehicle position data as well as a web service for clients to request for their historic vehicle records. This research reexamines the track of study and addresses the evaluation on performance, providing a lesson learned in fleet management industry for future research work.

1.1 Problem Statements

Driving records, digitalized and transmitted by the car kits in vehicles, are essential to every fleet management system. They provide fundamental information for vehicle tracking and other managerial measurements for fleet companies. Table 1 shows a simplified example of driving records, including vehicle identification, sending time, driving speed, mileages, position

in a form of coordinates, driving status, and so on. Every vehicle monitored by SkyEyes periodically sends one record to the control center every 30 seconds since the engine is started. As mentioned earlier, SkyEyes currently accommodates about 1,500 vehicles in total from over 200 customers around Taiwan, producing around 2 million driving records a day on a daily basis of 12 service hours for typical commercial vehicles, such as tanker trucks, freight trucks, or waste trucks. The annual driving records can reach up the number of half billion, which poses an enormous burden both on storage and on system performance.

Table 1: A simplified example of driving records

Car	Time	Speed	Mileage	X	Y	Status
ABC-01	2009-12-01 01:23:45	91	132001	167029.5951	2630698.927	Speeding
EFG-99	2009-12-01 01:23:48	25	4500	166781.4617	2630828.813	Running

Daily driving records are used for rendering vehicle location and driving trail on the digital map by fleet company staff in a real-time manner while historic driving records, say 2 months before the query, are often retrieved to analyze the efficiency of vehicle dispatching or serve as the evidence to claim compensation from insurance companies when accidents occur. Due to limitations of relational database (RDB) on which SkyEyes kernel is established, the system only provides 3-month historic driving records for real-time queries. This constraint causes many complaints from our customers when some of them started to adopt data mining techniques for better transportation plans in recent years. As the number of total vehicles in the system keeps growing, SkyEyes has been looking for an effective and financially feasible solution to solve this problem such as to sustain competitive in the market.

1.2 Research Objectives

Featured as high accessibility, availability, and scalability, cloud computing has been brought to enterprises with a more affordable solution to conquer the dilemma of increasing service demands with limited budgets. According to the definition by National Institute of Standards and Technology (NIST),

“cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction” [10].

This paper is mainly to introduce one of cloud computing frameworks, Apache Hadoop, as the solution to those problems addressed in previous section. To sum up, this research will:

1. Establish a ‘private cloud’ using Hadoop framework to provide a distributed computing environment for SkyEyes to accommodate increasing demands on system in near future.
2. Develop a Map/Reduce algorithm to facilitate the data conversion from RDB-based driving records into look-up table with row keys.
3. Summarize and analyze the result of experiments with suggestions for the future research.

2. Fleet Management System (FMS)

2.1 System Architecture

With the capability of global positioning systems (GPS), real-time monitoring comes to realization in modern fleet management. Vehicles to be monitored are equipped with a GPS-enabled car kit, a small box keeping tracks of driving information, such as speed, tire pressure, bearing, etc. [1, 3] When starting the engine, the vehicles regularly transmit the driving data through car kits via wireless communication, either 3G or GPRS, to the centralized control center. Web-based Fleet Management System is mounted on the control center with a collection of map services and web services associated with spatial analysis. Control staff can then real-time monitor all vehicles on the digital map with visual aids in web browsers. Figure 1 illustrates the system architecture of a typical FMS [4].

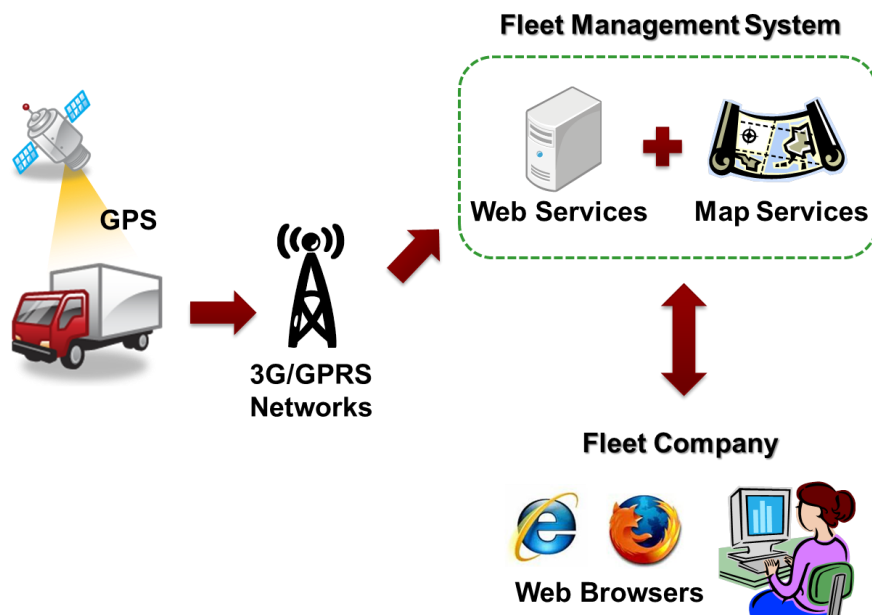


Figure 1: Architecture of Fleet Management Systems

2.2 Functionalities

There are various functionalities provided by FMS to assist fleet control managers in monitoring and managing vehicles. Those functionalities can be categorized as operation, safety, and analysis [2]. For operations, FMS provides managers with an easy way to map vehicles (as shown in Figure 2), find the closest capable drivers for specific mission, specify geofencing to confine vehicles within a designated area or on paths preventing them from unauthorized use of vehicles, and most importantly, communicate with drivers with a 2-way messaging channel [5, 8]. For safety concerns, FMS can automatically monitor the vehicle speed, and gives speeding alerts both to managers and to drivers when any vehicle drives over the posted speed limit. For analysis, maintenance reporting tracks service history for each vehicle notifying fleet managers to schedule preventive maintenance. Figure 3 shows the vehicle historic trails implemented by SkyEyes Corporation that can help managers with claims of insurance compensations when

damage occurs. Other reports such as fuel efficiency, man-hours, billing, and routing schedule provide useful information for managers with Key Performance Indicators (KPI) for hundreds of vehicles in order to improve their business [9].

To provide fleet managers with these functionalities, driving records sending from vehicles are essential to a FMS. Speeding alerts and travel time prediction module require vehicle speeds to process while geofencing and real-time monitoring require vehicle positions to render vehicles on the map. The data received and processed by a FMS can reach a million on a monthly basis for a small fleet company with 25 vehicles. SkyEyes is a web-based platform serving hundreds of fleet companies with more than 1,500 vehicles in total, producing about 2 million driving records in a single day. Storage and processing performance are two major concerns for its system scalability in the future.

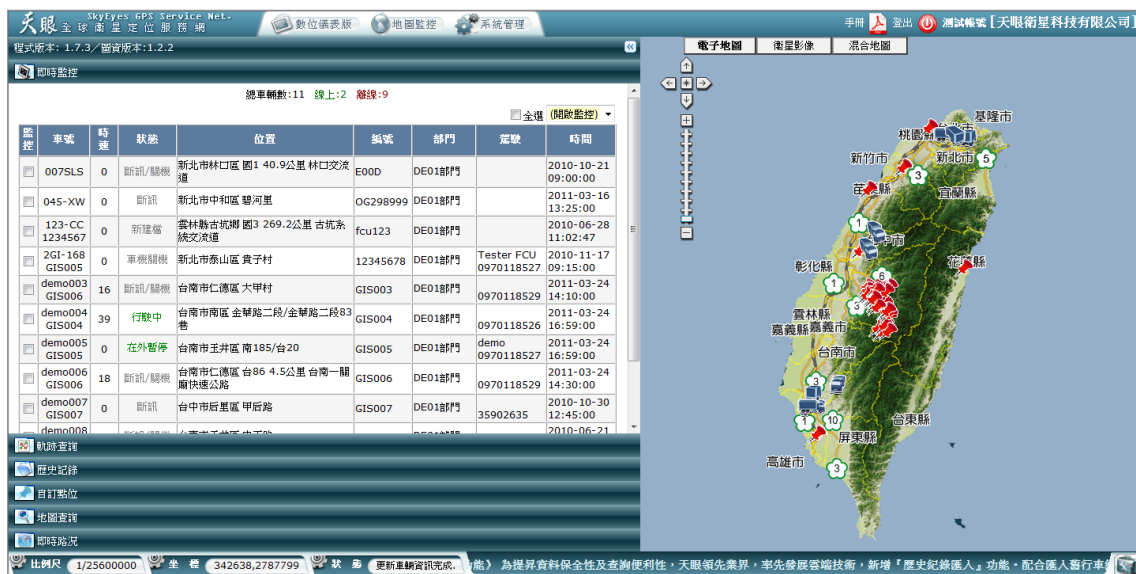


Figure 2: Real-time monitoring of vehicles on FMS



Figure 3: Animated vehicle historic trails

3. Framework of Cloud Computing

In order to enhance scalability and performance of SkyEyes, Cloud Computing was introduced and a framework based on open source projects, Apache Hadoop, was adopted. The core of Hadoop framework is Hadoop Common, which provides access to the filesystem supported by Hadoop, HDFS. The HDFS filesystem uses name of the network switch as its location awareness when replicating data, keeping different copies of the data on different racks. HDFS and HBase provide a high throughput access to application data, and support large structured data storage. The framework also consists of Map/Reduce, a software framework to support distributed computing on large data sets on clusters of computers.

A typical Hadoop cluster includes a single master and multiple slave nodes. The master node consists of a ‘jobtracker’, ‘tasktracker’, ‘namenode’, and ‘datanode’. A slave node consists of a ‘datanode’ and ‘tasktracker’. Figure 4 illustrates the framework of ‘private cloud’ adopted by SkyEyes for the experimental study. This framework consists of a master and two slaves, each of which is built on a PC with two-core 2.3 GHz CPU and 4 GBytes memory. The master provides the HBase APIs for SkyEyes Socket Server to transmit historic driving records into HBase as well as the ‘HBase Thrift’ for clients to place queries on the database.

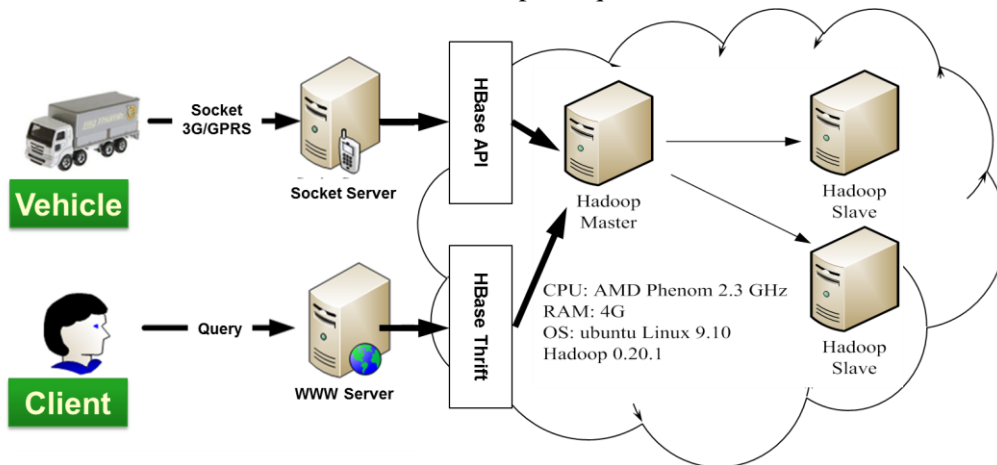


Figure 4: Framework of SkyEyes on the ‘Cloud’

The Map/Reduce algorithm developed in our framework splits the writing process into multiple threads on different slaves with distributed computing resource, and thus greatly decreases the processing time while writing driving records into HBase (as shown in Figure 5).

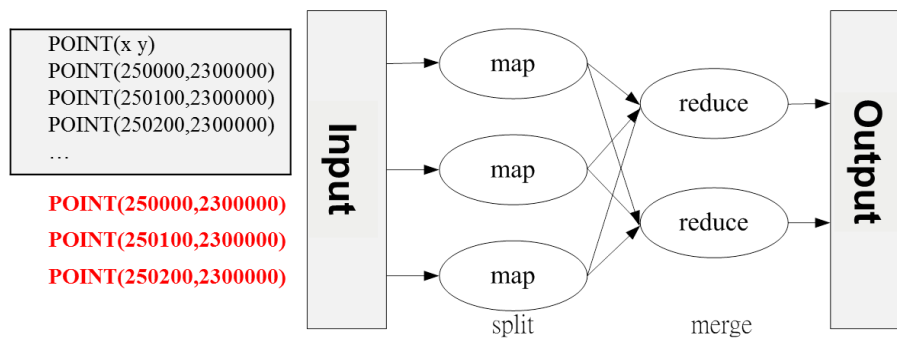


Figure 5: Framework of SkyEyes on the ‘Cloud’

4. Experiment and Results

The experiment of cloud computing consists of two parts: data feeding and querying.

In the phase of data feeding, we built a cluster of 3 nodes, which totally provide a computing capacity of 30 maps at a time to process the data transmission between HBase and the relational database where original driving records are stored. In this experiment, we fed the HBase with a real driving record set from SkyEyes between August 6, 2010 and August 28, 2010, with total records of 27.24 million. The Map process time is 2 hours and 11 minutes. Table 2 details the statistics of experiment by using Hadoop Map/Reduce Administration Tool.

Table 2: Summary of experiment: data feeding

Cluster Summary (Heap Size is 8.94 MB/963 MB)

Maps	Reduces	Total Submissions	Nodes	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node
0	0	64	3	30	6	12.00

Completed Jobs

Jobid	Priority	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete
job_201103020920_0071	NORMAL	SkyEyes	100.00%	29	29	100.00%

Hadoop job_201103020920_0071

Job Name: SkyEyes

Job File:

hdfs://cloud-a:9000/opt/hadoop-data/mapred/system/job_201103020920_0071/job.xml

Job Setup: Successful

Status: Succeeded

Started at: Mon Apr 18 11:40:39 CST 2011

Finished at: Mon Apr 18 13:51:42 CST 2011

Finished in: 2hrs, 11mins, 3sec

Job Cleanup: Successful

	Counter	Map	Reduce	Total
Job Counters	Launched map tasks	0	0	29
	Data-local map tasks	0	0	29
FileSystemCounters	HDFS_BYTES_READ	1,692,848,859	0	1,692,848,859
Map-Reduce Framework	Map input records	11,882,564	0	11,882,564
	Spilled Records	0	0	0
	Map input bytes	1,692,775,113	0	1,692,775,113
	Map output records	0	0	0

In the phase of querying, we developed a web service for clients to place queries on the historic driving records. The average response time of querying upon this data collection is less than 10 ms, which is over one hundred times than previous system without cloud computing. Figure 6 is the user interface of a client to place a query with conditions of vehicle plate number and time interval of query.

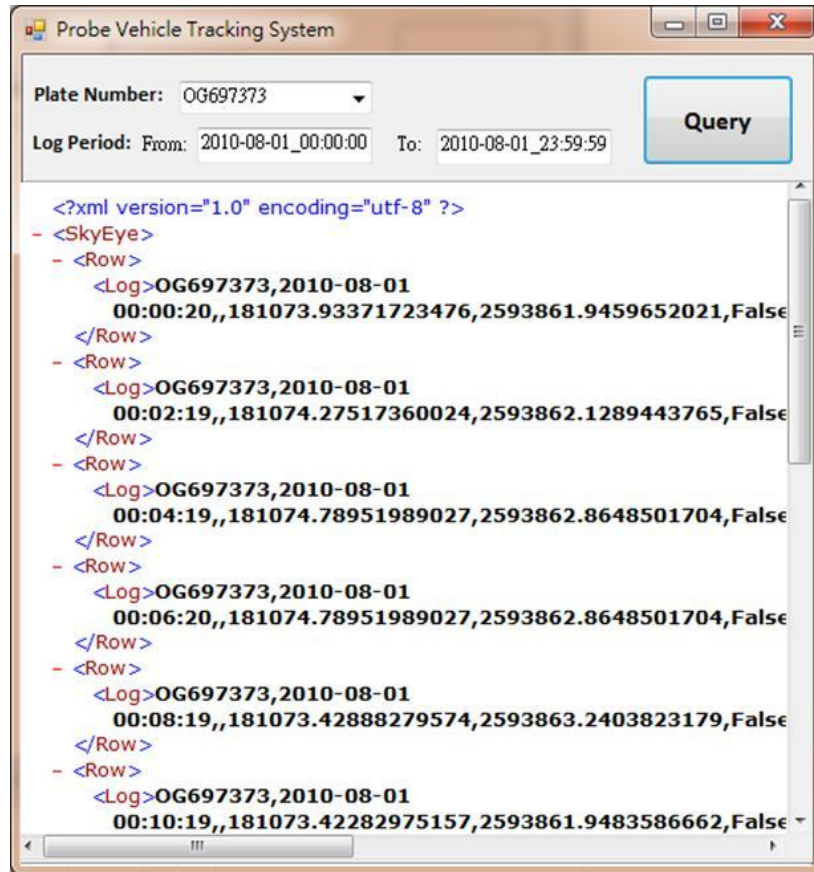


Figure 6: User interface of querying driving records in HBase

5. Conclusion and Future Work

From our experiment conducted by this study, we conclude that cloud computing can provide an effective distributed infrastructure for large-size data storage by clustering several separated hosts, but processing the jobs as a whole. Moreover, it can also provide a parallel computing capability to deal with queries on huge data storages, greatly reducing the cost of hardware and process resources, and meanwhile retain the satisfactory performance of querying conducted by web clients.

In the future, we plan to divide a digital map of Taiwan area into 1 meter by 1 meter grids and prepare a look-up table for feature identification. It is estimated less than 10 ms to identify any {x, y} coordinate as a map feature from over 80 billion grids.

References

- [1] Stojanovic, D.; Predic, B.; Antolovic, I.; Dordevic-Kajan, S., "Web information system for transport telematics and fleet management," *Telecommunication in Modern Satellite, Cable, and Broadcasting Services*, 2009. TELSIS '09. 9th International Conference, pp.314-317, 2009.
- [2] Thong, S.T.S.; Chua Tien Han; Rahman, T.A., "Intelligent Fleet Management System with Concurrent GPS & GSM Real-Time Positioning Technology," *Telecommunications*, 2007. ITST '07. 7th International Conference on ITS , pp.1-6, 2007.
- [3] Zeimpekis, V.; Giaglis, G.M.; Minis, I., "Development and Evaluation of an Intelligent Fleet Management System for City Logistics," *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual* , pp.72, 2008.
- [4] Giacobbe, M.; Puliafito, A.; Villari, M., "A service oriented system for fleet management and traffic monitoring," *Computers and Communications (ISCC)*, 2010 IEEE Symposium, pp.784-786, 2010.
- [5] Zantout, R.; Jrab, M.; Hamandi, L.; Sibai, F.N., "Fleet management automation using the global positioning system," *Innovations in Information Technology*, 2009. IIT '09. International Conference, pp.30-34, 2009.
- [6] HBase, Web page, <http://hadoop.apache.org/hbase>.
- [7] Hadoop, Web page, <http://hadoop.apache.org/hadoop>.
- [8] Telogis, Web page, <http://www.telogis.com/>.
- [9] SkyEyes, Web page, <http://skyeyes.com.tw/>.
- [10] The NIST Definition of Cloud Computing, http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf.