Distributed Storage and Computing Technology in New Energy Real-time on-Board Data

Yingzi Wang*, Jue Hou and Sisi Chu

China Automotive Technology & Research Center Co., Ltd. Hebei, 300300

*Corresponding author e-mail: yingzimizhan@sina.com

Abstract. The development and application of new energy vehicles has always been the focus of the government. As far as the users of new energy vehicles are concerned, the safety and energy-saving effects of operating conditions are of vital importance. Therefore, to fully promote the development of new energy vehicles, we must do a good job in the statistics and processing of new energy real-time vehicle data. This article focuses on the analysis of the effects of distributed storage and computing technology on real-time on-board data processing for new energy vehicles to provide solutions to the difficulties of countries and enterprises in the research and development of new energy vehicles and promote the energy saving of the state and enterprises.

Keywords: Distributed storage, Computing technology, New energy vehicles, Real-time on-board data, Research and application

Whether or not new energy vehicles can run safely and stably is a matter of close concern to the public. In the past, wireless remote monitoring methods were often used to analyze on-board data whose functions are relatively simple. Most of the collected data must be manually processed, which is time-consuming and error prone [1, 2]. Based on distributed storage and computing technology, a real-time on-board data statistical analysis platform for new energy vehicles is established, which can dynamically display real-time data such as energy consumption, mileage, and faults of new energy vehicles. It can also realize the statistical analysis of the vehicle's power consumption and fuel consumption in a certain period of time, which is an important basic guarantee for the safe and stable operation of new energy vehicles. Doing a good job in this area can promote the development of new
energy vehicles.

1. The basic concepts of distributed storage and computing technology
The distributed storage is mainly realized by the HDFS distributed file system, HBase distributed database, and Hive large database, which are interdependent. Distributed computing is based on systems such as MapReduce, Spark, Apache Hama, Apache Giraph, Open MPI, and Apache HBase. The collection and monitoring of new energy real-time on-board data is realized by using tools such as Ganglia and Sqoop.

2. The establishment of distributed storage and computing platform
The new energy real-time on-board data statistics and analysis platform based on distributed storage and computing technology mainly consists of six parts, that is, intelligent vehicle information unit, regional server, real-time data input server, database management system, web server and statistical analysis query browser client. The main function of intelligent vehicle unit is to collect and transfer the real-time vehicle data of new energy such as operation status, energy consumption parameters, GPS [3]. The main function of the regional server is to accept the processed real-time vehicle data of new energy, verify and analyze it, and transmit it to the server through the Internet in the form of information. The server contains the distributed storage and calculation module. The related content of the distributed storage and calculation module has been described above. The main function of the server is to make statistical analysis of the collected data and generate dynamic modules. The web server supports the data of the whole system and returns the final information to the client browser, so that the user can better view the information [4, 5].

In the analysis process of new energy real-time on-board data, "real time" contains two meanings, one of which is the original data collected by intelligent vehicle unit, which is processed to generate kinematic segments for better data processing. The other meaning is to use the existing technology to process the collected real-time data and display the processing in a dynamic form, and users can analyze the real-time operation state and energy consumption of new energy vehicles according to the dynamic graph. The traditional new energy real-time on-board data analysis and statistics work is often completed by manual off-line processing. Once one of the important data changes, the calculation results will be biased, which needs to be fully calculated to ensure the accuracy and effectiveness of the data. Therefore, this new energy real-time on-board data processing method is difficult to meet the basic condition of "real-time on-board data" [6, 7].

3. Distributed storage and computing technology for real-time energy consumption statistical analysis of new energy vehicles
Distributed storage and computing technology is used for the analysis of real-time on-board data of new energy, which can divide the energy consumption of new energy vehicles into several aspects for statistical analysis. With this technology, we can effectively grasp the energy consumption of each vehicle in the running state, and can also screen out some new energy vehicles for analysis. The conditions that can be used for screening are running area, vehicle type, power type and whether there
is external power supply. In the actual operation of new energy vehicles, the energy consumption is not a constant value, but a constantly changing form. Distributed storage and computing technology counts the real-time energy consumption of new energy vehicles in two forms: (1) statistics of energy consumption at every moment; (2) statistics of total energy consumption in a period of time. The statistics of energy consumption at every moment can make the user master the change rule of the energy consumption of the car, and guide the driver to effectively control the car. The statistics of total energy consumption in a period of time can reveal the total energy consumption of the car, which is conducive to the evaluation of the energy saving situation of the car by the enterprise personnel and users[8].

The application of distributed storage and distributed computing has greatly improved the calculation efficiency of real-time energy consumption of new energy vehicles. The distributed computing process is calculated as follows: (1) Basic integral calculation on the smart vehicle information terminal. The main calculation content is the fuel consumption and power consumption per 2 seconds of the vehicle running state; (2) Information processing on the database management system. The main calculation content is the fuel consumption and power consumption of 100 kilometers, and the battery recycling situation of 100 kilometers [9]; (3) Combined calculation of external data. The external area contains the operating area, vehicle type, and operating time of the new energy vehicle, which are effectively collected and displayed to the user in the form of graphics.

4. Application of distributed storage and computing technology in the real-time mileage and fault statistics analysis of new energy vehicles

The main content of the new energy real-time mileage statistical analysis is to collect statistics on the total mileage of new energy vehicles, and draw a mileage trend chart for presentation on a monthly or daily basis. Its main function is to approximate the mileage of the vehicle in a certain period of time, which has a certain role in the user evaluation of the actual mileage between the two places. The fault analysis summarizes the total number of monthly faults and types of faults in new energy vehicles, evaluates the overall situation of the vehicle, and concludes the high-frequency faults of the vehicle. The combination of new energy vehicle mileage and fault statistical analysis can be used to evaluate the stability and reliability of new energy vehicles.

5. Application of distributed storage and computing technology in the statistics analysis of real-time road traffic

The analysis of real-time road traffic is also an important part of new energy real-time on-board data. Accurate analysis of real-time road condition allows users to grasp the road conditions in a certain area at the first time to fully plan the road according to their own needs, which is a practical function for users. With the help of distributed storage and computing technology, real-time statistical analysis of road conditions is presented and presented in a dynamic form. The most effective method is fast clustering. Under the fast clustering method, kinematic segments are divided into two types, namely unobstructed road sections and congested road sections. Among them, the kinematic sections of the congested road section are shown as short moving distance and long idling time. In contrast, the non
idling time of unobstructed road sections is long, and the running distance of vehicle in unit time is long. In the distributed computing module, in order to reduce the calculation workload and generate dynamic road conditions, the running distance and non-idle time are selected as the analysis indicators of the road conditions\textsuperscript{10}.

In the distributed computing mode, the congestion degree of the road is displayed in the form of data, and users can better understand and make choices. The calculation method of crowding feature value is the ratio of the average value of moving distance and non-idle time in the kinematic segment. Generally, the deviation between the ratio of moving distance and non-idle time in the kinematic segment and crowding feature value is within 10%. In this case, it can be used as crowding feature value. Under normal circumstances, the intelligent on-board information unit will process the vehicle data and working condition data within 0.1s delay of the mobile phone. These data are effectively compared and transmitted to the user, and the accuracy and effectiveness of the road condition data are guaranteed. It is worth noting that the more data collected and recorded by the new energy real-time on-board data platform over time is, the more the road condition data generated will be more in line with the specific conditions of the region, which is also an important performance of data stability.

6. Conclusion
To sum up, the establishment of new energy real-time on-board data processing platform based on distributed storage and computing technology can guarantee the effectiveness and accuracy of vehicle data, and the display in the form of dynamic graph greatly improves the intuitiveness of data, which plays a role in promoting the development of new energy vehicles. Its main application advantages can be summarized as follows: (1) The establishment of new energy real-time on-board data processing platform under distributed storage and computing technology meets the needs of the government and enterprises, and is an important means to objectively and accurately grasp the development of new energy vehicle technology; (2) It can provide data support for the technical research and development of new energy vehicles, including national policies, industry standards and laws and regulations; (3) It directly establishes the relationship between vehicles and road conditions, providing support for road condition analysis; (4) It can serve the state, enterprises and scientific research institutions and promote the implementation of decisions and the research and development of technology.

References
[1] Li Yingping. Research and Application of Real-time Aggregation Technology Based on Distributed New Energy Access [J]. Inner Mongolia Science Technology and Economy, 2016 (2): 89-89.
[2] Wang Dong. Research on Real-time Flow Calculation and Processing of Railway Power Supply Monitoring Information Based on Storm [D].
[3] Wen Jing, Chen Da, Yu Lu. Research and Application of Distributed Storage Technology in Cloud Computing Environment [J]. Telecommunication Engineering Technology and Standardization (Issue 8): 57-61.
[4] Zhang Yanling, Bai Yameng. Research on Application of Hadoop Technology in Distributed Data Storage and Computing Platform of Smart Campus [J]. Journal of Jiaozuo University, 2017, v.31; No.100 (04): 76-79.

[5] Xie Hui, Yan Fangchao, Zhuang Jihui, et al. Development and Application of a Real-time Statistical Analysis Platform for Energy-Saving and New Energy Vehicles [C] // The 5th China Intelligent Transportation Annual Conference and the 6th International Energy-Saving and New Energy Vehicle Innovation Development Forum Excellent Proceedings (Volume 2)-New Energy Vehicles. 2009.

[6] Yang Nan. Research on Prediction of Power Consumption Based on Distributed Storage and Computing Platform [D]. 2015.

[7] Li Jun. Research on Key Technologies of Distributed Storage in Cloud Computing Environment [D].

[8] Gong Li, Shi Yang. Research and Application of Distributed Storage System [J]. Network Security Technology and Application, 2014 (9): 73-73.

[9] Wang Erxi, Zhou Yi. Research on Analysis Method of Distributed New Energy Monitoring Data Based on Cloud Platform [J]. Energy and Environmental Protection, 2018.

[10] Sun Jianzheng. Calculation and Analysis of Wind Turbine Condition Monitoring Data Based on Spark [D].