Design and implementation of ITS Architecture based on Big Data

Qiong Long¹, Xian-Li Duan²,*
¹Hunan City University, China, 413000
²Yiyang Communications Planning Survey and Design Institute, China, 413000
*Corresponding author e-mail: pangdui@outlook.com

Abstract. With the progress of social economy and the rapid development of transportation industry, in the face of the rapid growth of urban road traffic data, this paper puts forward the design scheme of intelligent traffic analysis system based on Hadoop. In this paper, HBase distributed database is used to store urban road static RDF data, hive data warehouse is used to store urban road traffic data, and MapReduce programming model is used to store massive and heterogeneous urban road traffic data. The urban road traffic data are analyzed, and the whole scheme is verified by the prototype system.

Keywords: ITS, Architecture, Big Data

1. Introduction
With the development of economy and the improvement of people's living standard, the number of urban cars is increasing rapidly. Traffic accidents, traffic jams and so on all cause great burden on urban road traffic. In the era of big data, urban road traffic data is also growing exponentially. In the face of its multi-source, heterogeneous, large number and other characteristics, how to analyze data quickly and efficiently, and make full use of the analysis results, so as to improve the efficiency of urban traffic operation is a major problem to be solved urgently[1]. Based on the analysis of the characteristics of traffic big data and the current big data technology, this paper designs the overall architecture and application response process based on Hadoop related components and computing models, and finally realizes the rapid and efficient processing of massive multi-source heterogeneous traffic big data.

2. System Design

2.1. Characteristics and Problems of Traffic Big Data
Traffic data mainly includes road information and vehicle information in traffic field, which are usually divided into static data and dynamic data. Static data includes road information data, road
facilities data, parking lot data, etc.; dynamic data includes traffic information data collected from coil equipment, video equipment, etc. Through the summary, it can be concluded that traffic big data has the following characteristics: multi-source, multi-dimensional, massive, dynamic and heterogeneous ". For static data, it is mainly structured data, which is smaller than dynamic data. And relatively fixed, but due to the differences of data collection institutions and equipment, it has a greater semantic heterogeneity, so the system introduces ontology model to process static data, to solve the problem of semantic heterogeneity; for dynamic data, because of the rapid development of transportation and the increasing demand for data in the era of big data, it has a stronger multi-source heterogeneity, and dynamic The volume of state data is huge [2]. How to store dynamic data and meet the needs of rapid query and calculation is the main problem faced by big traffic data. Static data ontology needs to be represented by RDF triplets. In order to meet the characteristics of stable storage, fast query and convenient expansion, HBase component of Hadoop framework is introduced for storage; for dynamic data, because of its huge volume and multi-source heterogeneous characteristics, combined with semantic network technology, hive component of Hadoop framework is introduced for storage to meet the computing needs. In the aspect of traffic big data calculation, in addition to the traditional data mining method, we also need to consider the parallel implementation of traffic analysis related algorithms during data mining [3]. MapReduce computing model can be used for the calculation of large-scale data sets, and it is convenient and efficient to run programs on distributed systems.

2.2. System Architecture Design
The design and implementation of this system is based on the Hadoop distributed system architecture. The core part of Hadoop framework includes HDFS and MapReduce [4]. HDFS provides support for the underlying data storage. MapReduce provides computing support for massive data. The main advantages of Hadoop lie in its scalability, efficiency and reliability. According to the characteristics of transportation big data, the overall architecture of intelligent transportation analysis platform is designed based on Hadoop framework. As shown in Figure 1. The architecture consists of data acquisition layer, data storage layer, data analysis layer and application layer.

3. Hierarchy of the System

![Figure 1. Platform Structure](image)

3.1. Data acquisition layer
The data acquisition layer is divided into dynamic data acquisition and static data acquisition [5]. The dynamic data includes the flow data, vehicle model data and vehicle speed data collected by the coil; the bayonet data includes license plate data, acquisition time and other data. Static data includes road
information, equipment information and other data that will not change due to different acquisition time. The dynamic data is integrated and stored in the distributed database, and the static data is mapped into RDF data according to the later ontology model for storage.

3.2. Data storage layer
The data storage layer stores the collected dynamic data and static data under the HA loop computer cluster [6]. The computer cluster adopts the master / slave architecture, the master node is the management node, and there is one, which is responsible for recording the data storage location and other information; the slave node is the data node, and there are many, which is the real physical storage location of data. In this framework, dynamic data is integrated into a large data table and stored in hive data warehouse; static data is mapped according to ontology model to become RDF data and then written in HBase distributed database ".

3.3. Data analysis layer
The data analysis layer uses MapReduce programming model to process and calculate the data according to different traffic big data analysis needs. In this system, data cleaning, association analysis, clustering analysis. 0d algorithm and other modules are mainly introduced. Due to the interference of external factors and the defects of data acquisition equipment itself, first of all, the data need to be cleaned to remove unreasonable and abnormal data, such as the license plate data is all "0" data.

3.4. Application layer
By using the results of data analysis layer, it can provide services to all users in the transportation industry in the form of interface. Provide different calculation results according to different needs of users. At the same time, new computing modules can be added in the data analysis layer to meet the new needs. For example, OD calculation results can be used to analyze individual travel behavior and group travel behavior, and clustering analysis results can be used to estimate travel time.

4. Application Response Process Design
According to the overall architecture of the system, the application response process is designed, as shown in Figure 2. The application process mainly includes user request, request verification, application matching and data calculation module. Users make application requests according to their personal needs, and the system automatically matches the algorithm modules and data needed by their application. If there is no such application, incomplete data or insufficient input conditions, the error information will be returned to the user. If the input conditions meet the calculation requirements, the application matching stage will be entered through request verification.
Application matching includes data matching and algorithm matching. According to the user's request, the data type needed by the application is found by ontology reasoning. For example, for 0d algorithm, the user needs to input the road range, time and license plate set, and then infer the coil code, license plate number, coil coordinates and acquisition time and other information to be found according to these ranges and the data required by od algorithm itself, and then search the order in semantic network Paragraph. According to the data type, the semantic web search part uses SPARQL to query HBase database to find out the actual address of the required data, transmits the address results to the big data search module, extracts the big table data stored in hive, and obtains the final data to be calculated. On the other hand, the algorithm module of complex application is obtained by encapsulating and recombining the algorithm of minimum force. Since the algorithms corresponding to each application have been encapsulated and the service is provided in the form of interface, the algorithm matching part only needs to search in the algorithm library to obtain the algorithm required for this request. Here, take the bicycle 0d algorithm as an example. Finally, the data and algorithm are combined to get the calculation results and return them to the user.

5. Summary
Based on the current situation of urban traffic and the development of big data, this paper puts forward the construction method of intelligent traffic analysis platform, completes the overall architecture design of the system, the application response process design and the implementation of the prototype system, and initially realizes the rapid calculation of traffic big data. The scalability, efficiency and reliability of Hadoop also provide guarantee for rapid and accurate traffic big data analysis. The analysis results based on this platform can also be reused to provide support for traffic research, traffic management and traffic planning, which can help to alleviate urban traffic congestion and other problems.

Acknowledgement
Key projects of Hunan Education Department(16A037).
Reference

[1] Engineering - Automobile Engineering; New Findings from University of Defence Update Understanding of Automobile Engineering (Application of fuzzy inference system for analysis of oil field data to optimize combustion engine maintenance)[J]. Energy Weekly News, 2019.

[2] Yifan Huang, Shengwang Meng. Automobile insurance classification ratemaking based on telematics driving data [J]. Decision Support Systems, 2019, 127.

[3] Drisya Manghat, Karuppanagounder Krishnamurthy. Performance evaluation of urban corridors using the speed profile data of automobiles [J]. Travel Behaviour and Society, 2020, 18.

[4] Wu Libing, Qiu Xin, Ye Luyao, et al. Research on the performance of SQL query engine based on Hadoop [J]. Journal of central China Normal University (NATURAL SCIENCE EDITION), 2016, 50 (2): 174-182.

[5] Zhao Yi. Research and practice of building data warehouse based on big data platform [J]. China financial computer, 2017 (05): 37-42.

[6] Lin Ziyu. Principle and application of big data technology [M]. Beijing: People's post and Telecommunications Press, 2017: 184-187.