Application of Big Data Technology in Urban Informatization

Junhu Li[1], Yufeng Liu[2], Yi Liu[1], Yu Yao[1], Longling Zhang[1]

[1]School of Date science and technology, Heilongjiang University, Harbin, China, 150080, China

[2]School of Entrepreneurship Education, Heilongjiang University, Harbin, China, 150080, China

Abstract

As the concept of smart cities in China has been proposed, combining urban data and digitization has become a major trend in urban development. However, at present, there are many and complicated sensing devices in cities, and the device models provided by various companies are also different. Therefore, city data has multiple sources and complex characteristics. Therefore, the use of traditional single-node servers to build city management systems has gradually failed to meet demand in terms of performance. With the advent of the concept of big data and related technologies, and the gradual implementation of 5G technology, these technologies have gradually been applied in urban informatization and combined with the current popular visualization technology, provide convenience for city managers, both It improves the efficiency of city analysis and management and improves the performance of the city management system. Based on big data technology, this paper proposes a high-efficiency, high-fault-tolerance urban information system architecture.

Keywords

urban informatization, distributed system.

1. Introduction

China is a country with a large population. In the period when China's Internet technology has not yet developed rapidly, most of the business needs to be handled manually. But after the 20th century, China ushered in an economy and technological take-off, and China's population grew on an unprecedented scale. Today, there are a large number of residents in major cities in China. This caused a variety of social problems: urban congestion, rising land prices, and low per capita distribution of social resources. In addition, the large population also makes the advancement of China's urban informatization much more difficult and complicated than in Western countries. Therefore, this paper discusses the data processing technology based on big data to solve the difficult and complex problems of China's urban informatization work.

1.1. Significance and Background of Urban Informatization

Urban informatization includes many fields, such as electronic office, e-commerce, digital city management, smart city planning, and other fields.

The informatization of cities is mainly to accurately grasp the development status of each city, and can effectively assist government personnel in making decisions such as city planning and management. China 's Internet giant Alibaba has gradually placed its strategic center in the informatization of smart cities. Now Hangzhou has a "city brain" that can collect different city information from multiple aspects and analyze effective results. Not only that, but other companies have also started to implement urban information management systems based on big data platforms.

At the same time, the city is in the "large-scale" direction, many serious social problems continue to produce: traffic safety has become difficult to control the flow of people getting bigger, accident management and control are not timely, use Big data, artificial intelligence,

cloud computing, and other emerging technologies promote the development of urban informatization and improve the efficiency of urban governance. It is a road to building an intelligent city and has a very broad prospect.

2. Urban Information Management Center based on Big Data Technology

In order to deal with the massive data of cities, it is necessary to construct a distributed city information computing system. This paper proposes a system architecture based on the Hadoop framework, which can sense data from the city in real time and send it to the computing system for storage and calculation.
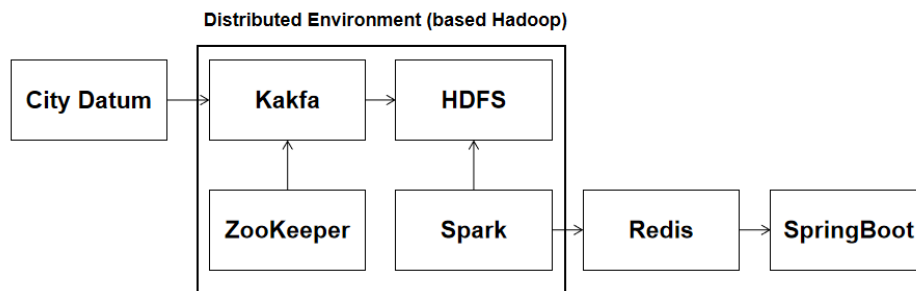

Fig 1. Distributed environment(based Hadoop)

Hadoop is a project of the Apache Foundation and is currently the most popular big data technology framework. It has the characteristics of high fault tolerance, and can run on processors with average performance and low prices. Therefore, developers can use a lower cost to build a stable information system.

3. Based on Kafka's City Information Collection, Cache Warehouse

As a city's central management system, data sources will come from various sensors, including mobile phone signaling information, subway travel records and so on. In order to collect these complex data in an orderly manner, a stable message pair needs to be constructed. The message queue serves as a buffer to prevent external information from directly pouring into the information system, causing blockages and even downtime. Therefore, a message queue with Robust nature is indispensable to the stability of the system.

Kafka can build an excellent distributed message queue, effectively deal with multi-source heterogeneous data from the city, and provide it to the data processing end for analysis and calculation based on the producer - consumer model.
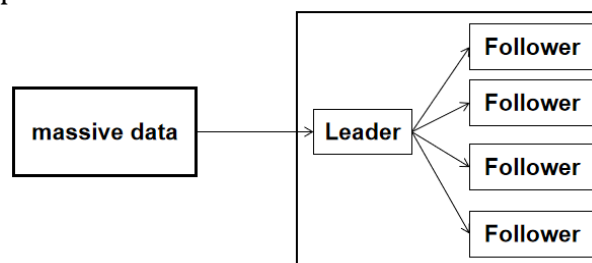

Fig 2. The overview of Kafka

Kafka will distribute it. It can store different types of city data in different nodes according to different 'Topics' , so it can selectively 'resident mobile phone signaling' , 'city PoI (Point of interest) data', or It is other data for classification.

4. Store Huge City Information based on HDFS System

HDFS (Hadoop Distributed File System) file system is a distributed file storage system, which belongs to Hadoop technology. It makes full use of the storage space of each node, and divides the roles of NameNode and DataNode to store the file content and file index. In large cities in China, the city information collected from various sensors is very large every day. Even using a large-capacity data storage center is difficult to deal with this problem. And once this single data center fails, it is necessary to rely on digital information. The automated management of urban systems is a devastating blow.

The number of backups of each node is configured in HDFS. That is, the same data will be copied to multiple redundant backups and stored in each node. In this way, when a node stops working due to a failure, the computing system can still obtain complete data from other nodes for calculation.

5. City Information Real-time Processing System based on Spark Technology

The informatization of cities is not just for the convenience of managing cities. How to display and calculate data in real time is also a problem for urban informatization: for example, cities need to continuously collect data and calculate the users of urban areas based on a short window Flow and movement trajectory. This requires that the information system not only has the ability to receive a large amount of traffic data, but also has enough computing power to calculate these huge and massive data.

Traditional information systems require data to be stored in a persistent data warehouse (such as MySQL, Oracle, and SQL Server, which are currently popular on the market), and then the data can be retrieved and calculated. Such processing methods and processing efficiency can no longer meet the requirements of urban informatization.

Therefore, this article proposes a solution: that is, use the Spark framework to build a streaming data processing framework. The Spark framework provides technologies for different scenarios: such as Spark Mitlib for machine learning, or Spark Streaming for streaming data processing. And choose the latter here. Under normal circumstances, it will work with Kafka 's message queue: according to the agreed time window, a discrete data stream is obtained from the Kafka message queue, and the operation is performed in units of one RDD .

For an efficient urban informatization system, real-time processing capability is a key criterion. The more efficient the calculation speed, the faster city managers can grasp the movement of the city.

6. The City Information System Depends on a More Complete Server System

With the increasing amount of data today, relying solely on a huge SQL database can no longer meet the demand. Traditional databases such as MySql are based on disk read and write operations, and the query speed is relatively low (relatively speaking). If the overloaded city information query service is forwarded to a single database, it is easy to cause the system to collapse.

For urban informatization systems, in order to ensure security, traditional data is often combined with some caching technologies. Redis is a data developed based on C language and running in machine memory. It usually uses message caching, message middleware, etc. It is one of the NoSQL- type databases.

For a system that has a large flow of data at all times, it is necessary to maintain the real-time and smoothness of messages. At present, mainstream server-side technologies still rely on development around the Spring framework (the framework is written in Java and is one of the most popular server application development frameworks on the market). Construction of the mechanism on the server-side listener, and Redis via ' message subscription ' communicate manner.

In addition, in the urban informatization system, in order to enable a robust and effective data interaction between the view page and the data system, technicians have gradually switched from the traditional HTTP protocol to the WebSocket protocol based on long connection.

7. Application of Visualization Technology in Urban Informatization

In recent years, visualization technology is also booming. It allows managers to get rid of the original complicated data analysis and can observe the city data intuitively only by means of charts. There are already many Internet companies in China dedicated to providing visualization services. The Chinese Internet giant Baidu also provides an easy-to-use, open-source framework based on JavaScript technology: Echarts.

Visualization technology can be perfectly applied to urban informatization work: such as urban traffic management. The traffic management department only needs to determine the operation of urban transport based on the chart. With the continuous improvement and development of 5G technology, the future urban information system can be connected to all video surveillance equipment in the city. Through the continuous transmission of images, CNN convolutional neural networks and other technologies are used to dynamically identify and track vehicles and portraits. Therefore, visualization technology has broad development prospects in urban informatization.

8. Conclusion

In today's urban informatization, big data technology is a new force that cannot be ignored. With the continuous development of urbanization in China, it is a general trend to integrate new technologies into the urban system. In the future, with the help of technologies such as the Internet of Things and 5G signal transmission, the city's information management will cover all aspects of people's lives.

In the future, the city's information management will combine 5G transmission, visualization, and other technologies. In order to be able to cope with large amounts of high-concurrency city data, a reliable system is required as a guarantee. The data processing technology based on the distributed systems will provide strong support for urban informatization.

Acknowledgements

References

[1] Pan Y, Tian Y, Liu X, et al. Urban big data and the development of city intelligence[J]. Engineering, 2016, 2(2): 171-178.

[2] Wei L, Yang S Y. Based on Big Data Technology Analysis on the Mode and Countermeasures of Smart City Construction Operation Management[C]//2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE). IEEE, 2018: 184-186.

[3] Bibri S E. The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability[J]. Sustainable cities and society, 2018, 38: 230-253.

[4] Yi Liu, Jiawen Peng, and Zhihao Yu. 2018. Big Data Platform Architecture under The Background of Financial Technology: In The Insurance Industry As An Example. In Proceedings of the 2018 International Conference on Big Data Engineering and Technology (BDET 2018). Association for Computing Machinery, New York, NY, USA, 31–35.

[5] Yanjun Z, Xiaodong Y, Yi L, et al. Research on the Construction of Wisdom Auditing Platform Based on Spatio-temporal Big Data [J][J]. Computer and Digital Engineering, 2019, 47(03): 616-619.

[6] Z. Zhao, J. Wang and Y. Liu, "User Electricity Behavior Analysis Based on K-Means Plus Clustering Algorithm," 2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC), Dalian, China, 2017, pp. 484-487, doi: 10.1109/ICCTEC.2017.00111.