

Research on Tax Big Data Supervision System and Method

Zihang Fei, Chuansheng Wu

Liaoning University of Science and Technology, China

Abstract: The tax big data supervision system and method, including the distributed real-time management and control system deployed in the superior unit and each subordinate unit, the superior distributed real-time management and control system is used to receive the management and control rules, and issue the management and control rules to the subordinate distributed real-time management and control system: The distributed real-time management and control system writes the massive data in the data source at the same level into the distributed message queue Kafka in real time, and uses the distributed stream computing engine Flink to compare and calculate based on the management and control rules, and write the data that conforms to the management and control rules. Enter the alarm information into Kafka; the lower-level distributed real-time management and control system pushes the alarm information in the Kafka subscription at this level to Kafka in the upper-level distributed real-time management and control system; the upper-level distributed real-time management and control system pushes the alarm information in the Kafka subscription to users, at the same time push the alarm information to the distributed search engine ElasticSearch to save, and provide the historical alarm information query function.

Keywords: Tax, Big data, Supervision.

1. Introduction

A large amount of real-time data generated by various taxation systems starts from the front-end sensing equipment, and is aggregated to the big data platform through multi-link circulation, and then from the lower-level unit to the upper-level unit. On the one hand, the data delay gradually increases, and the massive data is managed and controlled in real time. The timeliness of the data needs to be further improved. On the other hand, after a large amount of data is aggregated to the superior unit, it takes up a lot of physical resources and computing resources, and the existing resources of the subordinate cannot be used, resulting in the failure to use the resources reasonably and effectively.

2. Project Content

The embodiments of this project provide a tax big data supervision system and method that overcomes the above problems or at least partially solves the above problems. The specific solutions are as follows:

2.1. Provide tax big data supervision system

Including the distributed search engine ElasticSearch and the distributed real-time management and control system deployed at the superior and each subordinate, the superior distributed real-time management and control system is used to receive the management and control rules, and issue the management and control rules to each subordinate distributed real-time management and control system;

The distributed real-time management and control system uses the massive data in the data source at the same level to be written into the distributed message queue Kafka in real time to complete the real-time access of massive data, and uses the distributed streaming computing engine Flink to subscribe the real-time data in Kafka. The data is compared with the management and control rules, and the alarm information is written to Kafka for the data that conforms to

the management and control rules;

The lower-level distributed real-time management and control system is also used to push the alarm information in the Kafka subscription at this level to the Kafka of the upper-level distributed real-time management and control system; the upper-level distributed real-time management and control system is also used to push the alarm information in the Kafka subscription to the user interface Display and push alarm information to the distributed search engine ElasticSearch;

The distributed search engine ElasticSearch is used to save the alarm information pushed by the upper-level distributed real-time management and control system, and provide the function of querying historical alarm information.

The distributed real-time management and control system includes a management and control management module, a data access module, a management and control comparison module and a Mysql database; the management and control management module of the upper-level distributed real-time management and control system is used to receive management and control rules, and issue the management and control rules to each lower-level distributed management and control system. The management and control management module of the real-time management and control system; the management and control management module is used to write the received management and control rules into the Mysql database in the system at this level; the data access module is used to write the massive data in the data source at this level to the distribution system in real time. In the message queue Kafka, the real-time access of massive data is completed; the management and control comparison module is used to load the management and control rules from the Mysql database, and uses the distributed streaming computing engine Flink to compare the real-time data in the Kafka subscription with the management and control rules. For computing, write alarm information to Kafka for data that conforms to the management and control rules.

The upper-level distributed real-time management and control system is deployed in the upper-level unit, and the

distributed real-time management and control systems of each lower-level unit are respectively deployed in each lower-level unit.

The system also includes a distributed management and control platform. The distributed management and control platform manages the management and control authority of system users. According to the authority of system users, the upper-level distributed real-time management and control system is used to issue management and control rules to authorized subordinate units, and receive distributed real-time control rules from upper and lower levels. The alarm information pushed by the management and control system is pushed to the user interface for display.

The distributed management and control platform includes a unified user portal, through which the user's query request is received, and historical alarm information is queried from the distributed search engine ElasticSearch based on the query request, and fed back to the user.

2.2. Provide tax big data supervision methods

Methods include:

The upper-level distributed real-time management and control system receives the management and control rules, and issues the management and control rules to each lower-level distributed real-time management and control system; the distributed real-time management and control system writes the massive data in the data source at this level into the distributed message queue Kafka in real time, Complete the real-time access of massive data, and use the distributed streaming computing engine Flink to compare and calculate the real-time data in the Kafka subscription with the management and control rules, and write the alarm information to Kafka for the data that meets the management and control rules; lower-level distributed The real-time management and control system pushes the alarm information in the Kafka subscription at this level to Kafka in the upper-level distributed real-time management and control system, and the upper-level distributed real-time management and control system pushes the alarm information in the Kafka subscription to the user interface for display, and pushes the alarm information to the distributed real-time management and control system. Search engine ElasticSearch.

The distributed real-time management and control system includes a management and control management module, a data access module, a management and control comparison module and a Mysql database; the management and control rules are received through the management and control management module of the upper-level distributed real-time management and control system, and the management and control rules are issued to each subordinate distributed real-time management and control system. The management and control management module of the management and control system; the management and control management module of the distributed real-time management and control system at all levels writes the received management and control rules into the Mysql database in the system at this level; The massive data in the data source is written into the distributed message queue Kafka in real time to complete the real-time access to the massive data; the management and control comparison modules of the distributed real-time management and control systems at all levels load the management and control rules from the Mysql database, and use the distributed flow The Flink computing engine compares the real-time data in the Kafka subscription with the management and control

rules, and writes the alarm information to Kafka for the data that conforms to the management and control rules.

The upper-level distributed real-time management and control system is deployed in the upper-level unit, and the distributed real-time management and control systems of each lower-level unit are respectively deployed in each lower-level unit.

The method further includes: managing the management and control authority of the system user through the distributed management and control platform, using the upper-level distributed real-time management and control system to issue management and control rules to the authorized subordinate units according to the system user's authority, and receiving the information pushed by the upper-level distributed real-time management and control system. Alarm information, push the alarm information to the user interface for display.

The method further includes: setting a unified user portal on the distributed management and control platform, receiving query requests from users through the unified user portal, querying historical alarm information from the distributed search engine ElasticSearch based on the query requests, and feeding back to the users.

Description of drawings

Figure 1 is a structural diagram of the tax big data supervision system provided by this project embodiment.

3. Specific Implementations

The technical solutions in the embodiments of this project will be clearly and completely described below with reference to the accompanying drawings in the embodiments of this project. Obviously, the described embodiments are only a part of this project, not all of the embodiments. Based on the embodiments in this project, all other embodiments obtained by persons of ordinary skill in the art without creative work fall within the protection scope of this project.

3.1. As the first embodiment of this project,

Provide tax big data supervision system, including distributed search engine ElasticSearch and distributed real-time management and control system deployed at superiors and subordinates. Distributed real-time management and control system; also supports users through the unified upper-level distributed real-time management and control system, according to the authority of the system user, select the scope of management and control, according to the global management and control or designate lower-level management and control, and issue management and control rules to all lower-level distributed real-time management and control systems , or issued to the designated lower-level distributed real-time management and control system;

The distributed real-time management and control system uses the massive data in the data source at this level to be converted and processed into the distributed message queue Kafka of the distributed real-time management and control system in real time, so as to complete the real-time access to the massive data, and use the distributed stream Formula computing engine Flink, compares the real-time data in the Kafka subscription with the management and control rules, determines whether there is data that meets the management and control rules, and writes the corresponding alarm information to Kafka if there is data that meets the management and control rules;

The lower-level distributed real-time management and

control system is also used to push the alarm information in the Kafka subscription at this level to the Kafka of the upper-level distributed real-time management and control system; the upper-level distributed real-time management and control system is also used to push the alarm information in the Kafka subscription to the user interface Display and push alarm information to the distributed search engine Elasticsearch;

The distributed search engine Elasticsearch is used to save the alarm information pushed by the upper-level distributed real-time management and control system, and provide the function of querying historical alarm information.

Among them, the upper-level distributed real-time management and control system is deployed in the upper-level unit, and each lower-level distributed real-time management and control system is deployed in each lower-level unit. The distributed real-time management and control system includes a management and control management module, a data access module, a management and control comparison module and a Mysql database. ,As shown in Figure 1.

The management and control management module of the upper-level distributed real-time management and control system is used to receive the management and control rules, and issue the management and control rules to the management and control management modules of each lower-level distributed real-time management and control system;

The control management module is used to write the received control rules into the Mysql database in the system at this level;

The data access module is used to convert the massive data in the data source at the same level into a unified format, and write it into the distributed message queue Kafka in real time to complete the real-time access of the massive data;

The management and control comparison module is used to load the management and control rules from the Mysql database, and use the distributed streaming computing engine Flink to compare and calculate the real-time data in the Kafka subscription with the management and control rules, and write the alarm information to the data that conforms to the management and control rules. in Kafka.

Among them, the system also includes a distributed management and control platform. The distributed management and control platform manages the management and control authority of system users. According to the authority of system users, the upper-level distributed real-time management and control system is used to issue management and control rules to authorized subordinate units, and receive the upper-level distributed management and control rules. The alarm information pushed by the real-time management and control system is pushed to the user interface for display; in addition, the distributed management and control platform includes a unified user portal, which receives user query requests through the user unified portal, and queries the history from the distributed search engine Elasticsearch based on the query request. Alarm information and feedback to users.

3.2. Supervision methods for tax big data

Methods include:

The upper-level distributed real-time management and control system receives the management and control rules, and issues the management and control rules to each lower-level distributed real-time management and control system;

The distributed real-time management and control system writes the massive data in the data source at this level into the distributed message queue Kafka of the distributed real-time management and control system in real time after conversion processing, completes the real-time access of massive data, and uses distributed streaming The calculation engine Flink compares the real-time data in the Kafka subscription with the management and control rules, and determines whether there is data that meets the management and control rules. If there is data that meets the management and control rules, it writes the corresponding alarm information to Kafka;

The lower-level distributed real-time management and control system pushes the alarm information in the Kafka subscription at this level to Kafka in the upper-level distributed real-time management and control system, and the upper-level distributed real-time management and control system pushes the alarm information in the Kafka subscription to the user interface for display, and pushes the alarm information at the same time to the distributed search engine Elasticsearch.

This project transforms and upgrades the traditional centralized management and control system to a distributed management and control system. By deploying a set of distributed real-time management and control systems in the superior unit and each subordinate unit, the computing power is allocated to the subordinate units as needed, and the subordinate units submit the calculation results to the subordinate units. The aggregation of the upper-level system significantly improves the utilization rate of software and hardware resources, enables comparison and calculation in the data perception stage, improves the comparison and early warning timeliness and the actual combat efficiency, and solves the massive data in the tax system. The technical problem of real-time control and comparison timeliness , which can greatly reduce the problem of unreasonable use of resources.

4. Conclusion

The tax big data supervision system and method provided by this project transforms and upgrades the traditional centralized management and control system to a distributed management and control system. Subordinate units, the subordinates submit the calculation results to the superior system for aggregation, which significantly improves the utilization of software and hardware resources, and can perform comparison calculations in the data perception stage, improve the comparison and early warning timeliness and actual combat efficiency, and solve the massive tax system. The technical problem of real-time data management and control compared with timeliness can greatly reduce the problem of unreasonable resource utilization.

5. About the Author

Fei Zihang, male, Han nationality, born on June 19, 2002, from Huludao, Liaoning Province, is a third-year undergraduate student in the Department of Network Engineering, School of Computer and Software Engineering, Liaoning University of Science and Technology. , data structure and C# programming, and have HTML5 and CSS learning experience.

References

- [1] Research on enterprise tax risk management under the background of big data [J]. Pang Kaiwen. Times Economic and Trade. 2021(07).
- [2] Discussing enterprise tax risk management under big data [J]. Sang Zhijie. Small and Medium Enterprise Management and Technology (Early Issue). 2020(11).
- [3] Analysis and strategies of enterprise tax risk management under the background of big data [J]. Ren Changling. Small and Medium Enterprises in China. 2020(11).
- [4] Analysis and strategy of enterprise tax risk management under the background of big data [J]. Hou Wen. Marketing Industry. 2020(42).
- [5] Application of big data technology in tax risk prevention and control [J]. Liang Qian. Science and Technology Information. 2020(26).
- [6] Analysis and strategies of corporate tax risk management under the background of big data [J]. Rong Yan. Chinese and Foreign Entrepreneurs. 2020(19).
- [7] Research on corporate tax risk management under the background of big data [J]. Tang Ziyang. Tax Payment .2020 (17).
- [8] Analysis and strategy of corporate tax risk management under the background of big data [J]. Tan Juan. Taxation .2020 (04).