Research on coal mine data management and service platform

Yunqi Chen*

China Coal Technology Engineering Group Chongqing Research Institute, No. 6 Kecheng Road, Jiulongpo District, Chongqing, China

Abstract. In order to solve the problems of massive data of multi-source heterogeneous systems in coal mines, such as complex types, difficult aggregation, poor quality, and lack of a unified service platform, this paper analyzes the key technologies and theoretical methods in data collection, transmission, storage, calculation and service, and designs a coal mine data and management service platform based on a multi-layer architecture. It clearly defines that the platform consists of main functional modules such as collection and access management, quality management, storage management, metadata and master data management, operation and maintenance management and visualization management, which provides unified data services for all businesses and systems in the mine, and further improves the level of coal mine intelligence.

1 Introduction

Coal is China's main energy source and one of the most important guarantees for national energy security. With the continuous construction of various intelligent equipment and systems, the effective integration and efficient use of coal mine data has become an industry problem. This is particularly evident in the following aspects:

1) More and more intelligent IoT sensors are installed and used in coal mines, and the total amount of data generated has shown explosive growth, requiring a large amount of computing and storage resources, which has brought a lot of pressure to the existing information infrastructure of coal mine enterprises.

2) There are many types of coal mine information and automation systems. Some large and medium-sized coal mines have nearly 100 heterogeneous systems in place or under construction, which poses a huge challenge to system data integration and management applications.

3) The level of coal mine informationization is relatively backward. Massive monitoring data generally stays in monitoring alarms and data statistics. In terms of data mining analysis, machine learning and prediction and early warning, there is a lack of typical application cases, and the depth of data application is not enough.

4) Coal mine enterprises upload disaster monitoring data, equipment operating data and safety management data to superior group companies and government supervision and

* Corresponding author: chenyunqee@foxmail.com

regulatory departments in real time, which places increasingly high demands on the service capabilities and stability of coal mine enterprise data centers.

In response to the above problems, we should make full use of technologies such as the Internet of Things, cloud computing, and big data, improve data collection standards and specifications, design a reasonable and standardized data middle-office architecture, enhance data collection, analysis, and service capabilities, and promote efficient development of the enterprise.

2 Key technology analysis

2.1 Standardized acquisition technology for massive heterogeneous data

In order to achieve the effective collection and integration of data from nearly 100 subsystems in coal mines, a unified data collection standard specification must be established. The specification needs to be compatible with the existing mainstream technical protocols in the industry, such as PLC, OPC, MQTT, FTP, Web Socket, Web API, relational database, etc., to achieve transparency in access to similar subsystems and ensure the stability and maintainability of the data center[3].

In order to improve the quality of data access and ensure the authenticity, availability and integrity of data, the data collection process also involves data governance-related technologies, including data cleaning and data standardization. Data cleaning is to establish data quality standards and delete or correct the non-standard, incomplete and inaccurate data that is collected to ensure the accuracy of the data. In the process of data standardization, the data needs to be pre-processed to standardize its name, unit of measurement, representation format, etc., so as to achieve unambiguous representation of the same data.

2.2 Multivariate Data Distributed Storage Technology

Coal mine data is mainly divided into structured data, semi-structured data and unstructured data, and the corresponding storage methods include relational database, spatial data, memory database or big data storage[4]. Relational databases include Oracle, MySQL or PostgreSQL, etc. These databases with extension components can also store spatial data, such as PostSQL or Oracle Spatial.

In-memory databases are data directly operated in memory. Compared with disks, the data reading and writing speed of memory is several orders of magnitude higher, which greatly improves data access performance and is particularly suitable for the exchange and processing of massive real-time data. HBase is a high-performance, high-reliability and scalable distributed database for column-oriented storage developed based on Hadoop distributed file system HDFS. It is suitable for accessing massive historical data of coal mines in the form of key-value pairs.

2.3 General Algorithm Model and Service Interface

At present, some coal mines have completed the construction of coal mine data centers and data middle platforms, but lack the corresponding data analysis methods and have not formed an effective decision-making model. Therefore, it is necessary to introduce the current more mature machine learning and deep learning algorithms to provide algorithm model support for mine safety risk prediction and early warning.

Typical machine learning algorithms include decision trees, support vector machines, K-means, etc., which have been used in business scenarios such as coal mine gas concentration prediction and early warning, and personnel trajectory similarity analysis[5]. Deep learning is a subfield of machine learning. It uses a deep neural network model to automatically extract features from data. In coal mining enterprises, it is mainly used in fault identification such as face recognition, violation detection, and belt coal pile monitoring[6].

2.4 Multi-service and multi-terminal data synchronization technology

At present, the data synchronization technologies widely used in the coal mining industry include database automatic synchronization technology, three-party ETL tools, message queues, and FTP file upload technology under standard protocols. Mainstream relational databases such as MySQL, Oracle, and SQL Server all have full and incremental data replication functions, but they are mainly used for data copying and data migration between data models of the same type and architecture. ETL tools[7] include Kettle, Flume, and Informatica, which support data extraction from structured data sources and synchronization to data warehouses, but these do not support complex logical data requirements well, and most data sources are relational databases[7].

Coal mining companies use more FTP and kafka message docking technology protocols, follow the perception data specifications issued by regulatory authorities, and achieve data synchronization and upload with company platforms and regulatory platforms at all levels.

3 Platform architecture and functional design

3.1 Platform technical architecture

The data management service platform consists of the data generation layer, subsystem layer and data middle platform service, which in turn consists of modules such as data acquisition, data governance analysis, data storage, computing services and data sharing services. Its architecture is shown in the figure below.

The data generation layer is mainly composed of time series data continuously generated by various disaster monitoring sensors such as gas, water, and fire, as well as mechanical and electrical equipment operating data and management, safety, finance, personnel and other data related to the coal mine production process.

The subsystem layer is based on the data generation layer, and according to management and production needs, the data is classified and collected through multiple different subsystems for online monitoring and management. Typical examples include relational data, spatial geographic databases, etc. to store structured data, various Excel and CAD files to store semi-structured data, and text documents, audio and video files to store unformatted data.

The data acquisition module uses interface protocols such as FTP, OPC, Kafka, and Web API to achieve fast, complete, and effective collection of massive heterogeneous system data, providing a data transmission channel for data fusion upwards and providing an interface for issuing collaborative linkage control instructions downwards. The data governance module includes data analysis and data quality management. The platform submits the standardized data to the data quality management module after analysis, and classifies and stores it in the warehouse after format verification and consistency check.

Depending on the data category and purpose, data with high real-time requirements is stored in the memory database. The result data after statistics and analysis is stored in the

relational database. For long-term historical data information, data migration is performed regularly and stored in a large database. The computing service module is an important part of data service management, which provides computing services for upper-level businesses. These computing services include real-time data computing, historical data mining and analysis, and AI model services.

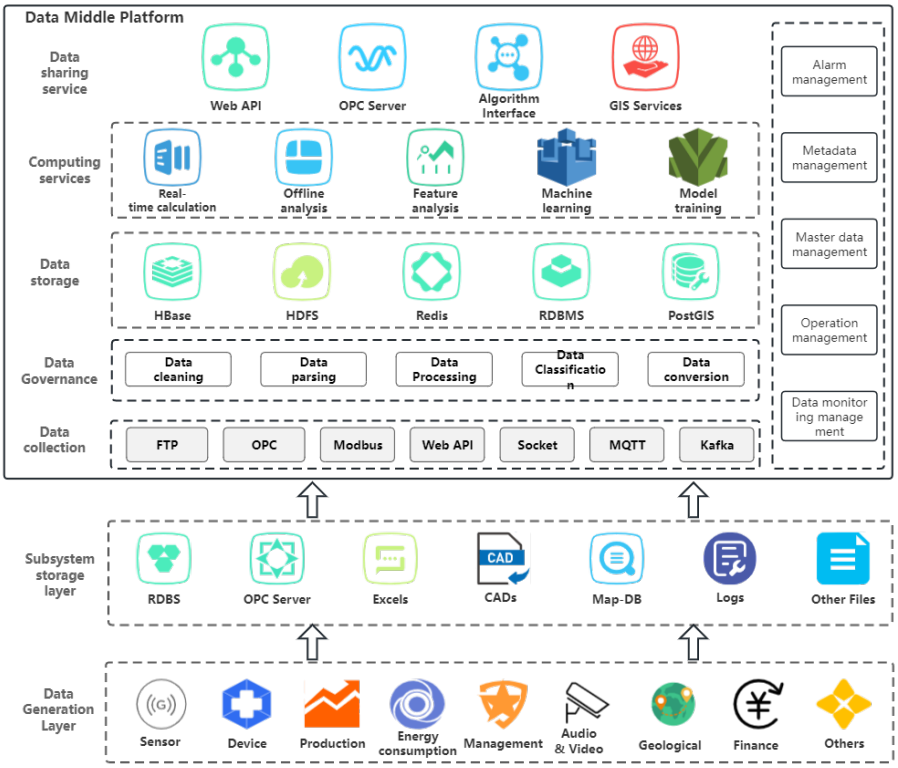

Fig. 1. Architecture of data management service platform.

The data service layer provides various data middle-end services to the management and control platform and third-party systems through the API interface. It mainly includes coal mine data center data sharing services, which provide query, statistics and analysis services for safety monitoring, production management and other data through the standardized unified data model and data service interface of the entire mine.

3.2 Main functions of the platform

The main functions of the platform include collection and access management, data quality management, data storage management, metadata management, master data management, data operation and maintenance management, visualization management.

Data collection and access management is responsible for the access of security, monitoring, production, management, and operation data to each subsystem in accordance with the agreed standard specifications and technical protocols. Data quality management implements the cleaning, deduplication, and standardization of access data to ensure that it meets business needs before storage. Data storage management stores the data in real-time databases, relational databases, large databases, or spatial databases according to the type of data after analysis.

Metadata management is used to develop and manage unified data standards, including naming specifications, data type definitions, business vocabulary, etc., and can realize metadata online configuration management. Master data management is used to clearly define and identify related elements and entities in coal mines such as safety, production, operation, management, equipment, spatial location, etc., and is used for accurate representation and communication of information between multiple businesses, multiple systems, and multiple departments.

Data operation and maintenance management mainly provides data center operation and maintenance and data operation and maintenance functions to coal mine data operation and maintenance and management personnel, conveniently realizes data query and statistics, and has functions such as report query, data export, chart comparison, and data backup. Data visualization is biased towards providing management personnel and senior leaders with comprehensive data after analysis, providing reliable data support for the safety management and business decision-making of the entire mine.

4 Conclusion

Through the research and application of data management and service platforms, standardized specifications for multi-source heterogeneous data in the entire mine have been formed, the data collection interface has been unified, data is classified and stored by subject for different data types and business applications, and a unified data sharing and interaction interface has been built, which has improved the data quality, data service and data management capabilities of the data center and promoted the continuous improvement of the intelligence and informatization level of coal mines.

This work was supported by funds from the Special project of CCTEG Chongqing Research Institute (2022YBXM11).

References

1. WANG Guofa.New technological progress of coal mine intelligence and its problems[J].Coal Science and Technology,2022,50(1) :1-27.
2. FANG Qian, ZHANG Xiaoxia, WANG Lin, et al. Research, practice and application of key technologies of intelligent coal mine bigdata governance[J]. Journal of Mine Automation,2023,49(5):37-45, 73.
3. YAN Jiahe, LI Honghui, MA Ying, et al. Multi-source Heterogeneous Data Fusion Technologies and Government BigData Governance System[J].Computer Science,2024,51(02): 1-14.
4. MA Shuqun.Design of high performance storage architecture for large concurrent big datainformation center of ship[J].Ship Science and Technology, 2020, 42(02): 148-150.
5. Kaur Rupinder, Jain Dr. Anubha, Saini Pushpanjali, et al. A Review Analysis Techniques of Flower Classification Based on Machine Learning Algorithms[J].Electrochemical Society Transactions,2022(1):9609.
6. Rajyalakshmi, Vankadhara, Kuruva Lakshmanna. A review on smart city-IoT and deep learning algorithms, challenges. International journal of engineering systems modelling and simulation 13.1 (2022): 3-26.
7. Cheng Xueqi, Liu Shenghua, Zhang Ruqing. Thinking on new system for big data technology. Bulletin of Chinese Academy of Sciences, 2022, 37(1): 60-67.