

Research and Realization for Information System of Shenhua Ningxia Coal Integration Based on XML

Xiaoyan Zhang¹, Huan Li¹ and Shiheng Liu²

¹Xi'an University of Science and Technology, China

²Shenhua Ningxia coal industry group company, China

Abstract—Shenhua Ningxia Coal Group as a modern large-scale coal enterprises, attaches great importance to information management, and promote the development of the coal industry. There are many information management systems within the group, playing an important role, but the various systems are independent, the system can not be shared between the data, which led to the emergence of "information silos". In response to this problem, The data relation between several systems of the group adopts the way of XML technology and middleware to establish the heterogeneous data fusion model, and uses the similarity algorithm to match the attribute columns to break the heterogeneity among the databases and realize the data integration among the systems.

Keywords—information silos; XML technology; middleware; data integration

I. INTRODUCTION

In recent years, the coal industry information construction has been rapidly developed, and the information management system has been established in various departments of the enterprise. With the development of information and network era, more and more enterprises need to realize the integrated management and realize the data sharing among various information systems within the enterprise. However, since each independent system uses different databases and data models, it is difficult to synchronize data and thus form an information island. Therefore, it is an urgent issue to realize data synchronization and data sharing in heterogeneous databases.

Shenhua Ningxia Coal Group Company has taken the forefront of the coal industry in information construction. At present, 23 sets of management information systems of Shenhua Ning Coal Group Company are put into operation. However, due to the fact that the management of various system data of the Group is not truly unified, The data of each system can not be shared. Aiming at this problem, Ningxia coal Group proposed to achieve group information integration by the end of 2016. In this paper, the unified integration of coal quality information of Ningxia coal group is analyzed and studied, and the coal information sharing between five coal-related systems is realized.

II. ANALYSIS OF THE CURRENT SITUATION OF SHENHUA NINGXIA COAL INFORMATION SYSTEM.

Shenhua Ningxia Coal Group attaches great importance to information construction and information management. Currently 23 sets of management information systems within the group to facilitate the user's operation, and play a good role in

their respective application scenarios. However, due to the independence of each system, different system developers use different databases and system environments, resulting in the system can not be shared between the data, so that the system can not be unified management of information, the lack of connectivity between systems, making the system Between the "information island" phenomenon. A large number of data and information are scattered to different systems(1), data sharing and data integration can not be carried out, and there is data interaction between independent information systems. Therefore, data integration is needed to achieve data sharing between systems.

The systems involved in the management of coal quality information include the coal quality management information system, the Shenhua Ningxia ERP management system, the transportation and marketing management information system, the coal metering information system and the coal washing plant information system. Each system is independent of others and each has different emphasis. Therefore, we can not see all the data inside the group from a macro perspective. And there is data exchange among all systems. Users are required to repeat input for the amount of the system, which increases the work load. In addition, multiple management systems are run in the group at the same time. For information department managers, they need to learn a different system of operating and maintenance methods, the task is arduous, so the system need to be solved between heterogeneous, as this can lighten the user's workload and can realize the group information integrated management.

A. Analysis of Heterogeneous Data of Shenhua Ningxia Coal

Shenhua Ningxia Coal Group involved in coal quality information system has five independent subsystems: coal quality management information system, Shenhua Ningxia ERP management system, transportation management information system, coal measurement information system and coal washing plant information system, the use of subsystems Database and system architecture vary, and each system's database table naming format, constraints and table structure is inconsistent, so to achieve heterogeneous database system information integration needs to break the heterogeneity between the database to achieve each Data conversion between databases and transparent access between heterogeneous databases. There are data exchange among these systems in the group, and each system is independent and connected with each other. In order to ensure the consistency of data among the systems and prevent users from doing duplicate entry work, the quality management system of the coal quality system is overlapped with several other systems to realize data sharing.

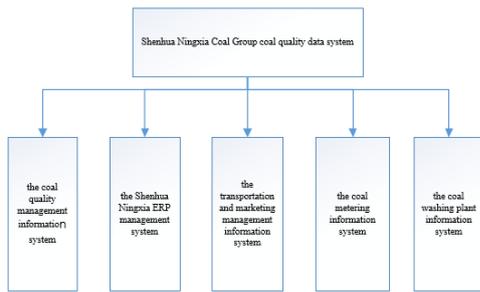


FIGURE 1. SHENHUA NINGXIA COAL QUALITY DATA SYSTEM

As shown in Figure 1, this paper mainly focuses on the data integration of the above five systems. The five systems have different emphases. Many of the information between the systems is duplicated. Because of the independent existence of each other, the staff need to repeatedly input data. Shenhua Ningxia ERP Management System in order to settle the coal quality data in the coal quality management information system and the sales volume of the railway and highway in the transportation system, the coal quality management system needs to obtain the "quantity" of coal from the measurement system, And some data are obtained from transportation management information system. There is data exchange between the five systems. In order to facilitate the operation and use of the users, it is necessary to realize the integration among several systems.

B. The Key Technology of Coal Quality Data Integration

Data integration is to integrate data from different systems and different storage media into the same data environment and present it as a unified data view(2). Data integration of multiple data sources that is, to break the heterogeneity between the database, make full use of the scattered data resources in various systems, and truly achieve data sharing.

At present, the way to solve heterogeneous databases is federation, data warehousing and middleware(3). The federated database system consists of a set of local databases that work in coordination with each other but maintain their autonomy. There is no unified mode between systems, which is suitable for data volume Relatively small heterogeneous database; data warehouse is the heterogeneous database data extracted, filtered and stored in a data warehouse, so as to facilitate data query, only for small changes in the data and a small amount of data in the database; and The Mediator method is by far the most popular way to create an integrated database. The middleware gets the user's request and feeds it back to the data layer, which parses the data layer's query results into an XML format and returns it to the client.

Database intermediate technology to provide users with access to heterogeneous data sources unified interface, so users do not need to know how to communicate between the various databases, how to connect, data extraction, data synthesis and data query transparency, the user only need to operate the query . The use of middleware to eliminate the heterogeneity between the database, the general relational database consists of many sheets of data tables, and each table contains a different record, each record will be composed of many data columns, so the smallest of the relational database Data that is a column, and the data ta-

ble is the database entity, the table by the primary key and foreign key association, so the system integration in order to make heterogeneous database data unification, the need to transform the table in the database into xml Document form, convenient data reading and transmission between tables.

XML language is a cross-platform language, has a good data storage format, and flexibility is good, scalability, ease of network transmission. Data can be exchanged between different databases through xml language, and the middleware method based on XML language can be used to integrate the data of heterogeneous databases, which can solve the problem of "information isolated islands" between independent heterogeneous databases.

III. IMPLEMENTATION OF COAL DATA INTEGRATION

A. Mapping between XML Documents and Databases

Xml document data model can be seen as a "tree" shape, by the parent node and child nodes, for each database information table corresponds to an XML document, the data table name as the XML document's parent node, each attribute column Is a child node, the information in the table traverses until the end of the last line, you can map all the information in the data table into xml form.As follow.

TABLE I. PERSON INFO

Name	Age	Position	Sex	Worktime	Firdegree
wang	26	operator	female	2011-09	undergraduate
li	25	technican	male	2010-10	undergraduate

```

<?xml version="1.0" standalone="yes"?>
  < PersonInfo >
    <Name> wang</Name>
    <Age>26</Age>
    <Position> operator </Position>
    <Sex>female</Sex>
    <Worktime>2011-09 </Worktime>
    <Firdegree> undergraduate</Firdegree>
  </ PersonInfo >
  < PersonInfo >
    <Name>li </Name>
    <Age>25</Age>
    <Position>technican </Position>
    <Sex>male</Sex>
    <Worktime>2010-10 </Worktime>
    <Firdegree>undergraduate</Firdegree>
  </ PersonInfo >
  
```

B. Heterogeneous Database Data Integration Process

Middleware technology is used to realize the information sharing of heterogeneous databases. By adding middleware layer to the database layer and presentation layer, the data between each database is processed. First, the user requests data from the presentation layer, and the request is received by the middleware. The middleware analyzes and decomposes the query request, and finally publishes this function to each database. After the database receives the query data and returns the query result to the corresponding wrapper, the wrapper converts the result of the query Into xml documents, and data mapping,

synonymous data and synonymous data matching mapping, and finally these results are integrated into a global view file, returned to the user.

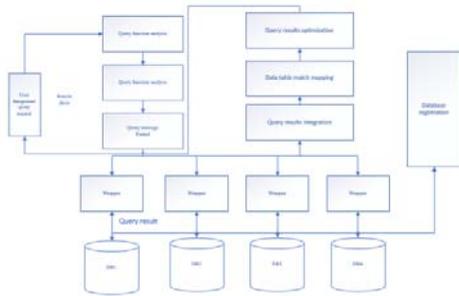


FIGURE II. COAL QUALITY DATA INTEGRATION ARCHITECTURE DIAGRAM

C. System Integration and Results

System integration using middleware way to achieve heterogeneous database data communication, xml technology to achieve heterogeneous database table data conversion, matching the mapping of data results using common matching algorithm for attribute matching and will be similar to the meaning of The words are unified named to facilitate the user's query, and finally all the results will be displayed as a virtual global view, the final results will be displayed in the client interface. In the process of data integration for the same meaning but naming a few different forms of matching mapping mapping algorithm used to achieve. Data integration between two systems, need to match the attribute column between database tables, can we truly achieve the data sharing between two systems, property is obtained by using the edit distance algorithm into the column of similarity, the higher the similarity value of property, the closer the relationship between can match(4). For example, the two systems of coal quality management information system and coal industry transportation and marketing group enterprise management system have data interaction. The coal quality management system needs to obtain the information such as the highway, railway, coal mass and pricing standard of commercial coal, System of mine name, mine point, brand and customer name and marketing system to establish the corresponding relationship in order to pull the data in the distribution of coal quality information system.

TABLE II. TWO TABLE NAMES AND ATTRIBUTES

num	The transportation mine.	The name of the coal mine.	NULL	Default
1	Jinfeng Coal Mine	Jinfeng mine	Y	NULL
2	Baijigou Coal Mine	Baijigou mine	Y	NULL
3	Hongliu Coal Mine	Hongliu mine	Y	NULL
4	Golden company		Y	NULL

TABLE III. SIMILARITY CALCULATION

num	The transportation mine.	The name of the coal mine.	Similarity value
1	Jinfeng Coal Mine	Jinfeng mine	0.75
2	Baijigou Coal Mine	Baijigou mine	0.75
3	Hongliu Coal Mine	Hongliu mine	0.75
4	Golden company		0

IV. CONCLUSION

With the development of informatization, the integrated management of information in coal enterprises is also gradually taking shape. In order to meet the actual demand, coal enterprises may use different systems, but these systems exist independently and cause some data to be unavailable and effective Use, you need to integrate the data between the various systems, this paper uses middleware to achieve information fusion between heterogeneous databases, using xml document data mapping between database tables to achieve data sharing within the group system, user-friendly The use of truly integrated information management group.

ACKNOWLEDGMENT

Supported by Industrial Project of Science and Technology Department of Shaanxi Province (2014K05-37)

Supported by the Serving Local Project of Education Department of Shaanxi Province (14JF016)

REFERENCES

- [1] Liu xingzhi. The Present Situation and Promotion Measures of Shandong Information Construction. information Technology.2015
- [2] Jianjiang Zhang.Research on Heterogeneous Database Integration Middleware Based on xml. [D].2010
- [3] Yidan Zhou.Research on query rewrite in heterogeneous data integration system.[D] .2017
- [4] Ke wei.Design and implementation of similarity query system in heterogeneous database.[D].2016