Dimensional Modeling Method Discussion for the Profits from Mineral Rights Transfer Management

Jun Tang
Kunming Geofly Technology Co., LTD., Kunming 650041, Yunnan, China
DOI: 10.32629/memf.v3i2.771

Abstract: In the informatization process of the profits from mineral rights transfer management, it is of great practical significance to establish a data warehouse with easy to understand, efficient query performance and strong scalability, so as to provide good decision support services for managers. As an important method of data warehouse modeling, dimensional modeling method has significant advantages over others in data understandability, scalability and data access performance compared with other data warehouse modeling methods. Combined with the business needs of mineral rights transfer income (mineral rights price) assessment and collection decision analysis, the key design ideas of dimensional modeling are expounded from the perspectives of dimension content change processing design, dimension role-playing, shared dimension design and audit dimension design. Taking the actual engineering project as an example, dimension modeling results can be queried easily in order to help domestic peers in the construction of DW/BI projects.
***Keywords*:** the profits from mineral rights transfer, data warehouse, dimensional modeling, dimension role-playing, shared dimension

As an important business of management department in natural resource, there have been a large amount of data accumulated in mineral rights transfer in its information construction, and there will be changes for these data in the future in a continuous manner. Data on mineral rights transfer revenue based management is endowed with the characteristics of multiple management levels, cross-department management, rich and heterogeneous data types. How to effectively realize data mining and knowledge discovery of mineral rights transfer revenue management process based on these data, so as to better support natural resource management departments to make better management decisions, is an urgent problem to be studied and solved.

In the practical application of data mining and knowledge discovery, it is feasible to construct high quality data warehouse to support the upper data mining application. For the purpose of a data warehouse is built for the analysis of the integrated data environment[1], and fundamentally different from the transaction processing system for database, the construction of data warehouse in the technical level to solve the problem is to focus on increasing bulk data analysis of the performance, and improve the business analyst on the data using the convenience of support. Data warehouse modeling methods mainly include dimension modeling and paradigm modeling[2]. Compared with dimension modeling, the entity relationship model established by the paradigm modeling method is too standardized. Although data redundancy is eliminated and data consistency control and other problems are reduced, it cannot improve the performance of data analysis and mining and the convenience of data use like dimension modeling. Therefore, the current mainstream data warehouse modeling method is dimension modeling method, partly mixed with paradigm modeling method[3].

The theory of dimensional modeling was first proposed by Ralph Kimball, and the dimensional modeling technology under the guidance of this theoretical system is the best practice technology of data warehouse multidimensional data model design[4]. In foreign countries, Kimball et al.[4] have provided analysis and design cases of retail, financial services, insurance, telecommunications, education, health care and other industries. In China, there are many relevant literatures describing the design results formed by the application of this technology. For example, Liu Jun et al. [5], Xiang Shiyao et al. [6], Wu Qunyong et al. [7], and Chen Jun et al. [8] applied this technology to conduct research in combination with a specific field of business and expressed the design results of the fact table and dimension table, but there are few descriptions of the ideas on how to form the design results

In terms of studying how to form design results, Liu Yanjun et al. [9] conducted design analysis on consistency dimension and measurement combining with large-scale equipment state analysis. He Chengang[10] studied how to update late data in mechanism, and Gao Liang[11] explained the design of slow change dimension in combination with university management business. It can be said that this technology has been applied to information construction in many fields of global social economy, but there are relatively few reports on the analysis and design process of how to form modeling results, especially

in the management business of mineral rights transfer income. On the other hand, existing reports mainly give focus on the choice of data mode, some of which involve the design of dimension slowly changing, while other key analysis and design ideas of dimension modeling are rarely mentioned in combination with business.

Based on this, from the business requirements on decision analysis on management of mining rights transfer revenue, this paper focuses on dimension modeling in dimension content change processing, design dimension based role playing, shared dimension design, audit dimension design and other aspects. The decision support requirement of revenue management business of a provincial mineral rights transfer is verified as an example, which is expected to be helpful for reference application in DW/BI projects of domestic counterparts.

1. Dimensional modeling design

1.1 Brief introduction to decision support requirements of mineral rights transfer income management

The management business of mineral rights transfer income varies according to the types and reasons of mineral rights transfer, which is jointly managed by the national, provincial, city-county natural resource management departments at four levels. Externally, it is also associated with the non-tax revenue collection and payment management departments, mining rights holders, evaluation agencies, etc. In the same level of natural management department, it also spans mining rights management, reserves management, mining rights transaction management, financial management and other relevant functional departments, resulting in a relatively complex management process. In the construction of its data warehouse, first of all, we need to consider how the business data with complex structure and uneven quality generated in the history can well support the decision analysis needs of diverse management groups. Second, you need to consider to provide good data understanding support for non-IT (Information Technology) professionals on the front end. Finally, consider high-performance support for BI (Business Intelligence) analysis tools.

1.2 Key design

1.2.1 Design of dimension content change processing

In the process of revenue management of mineral rights transfer, there are changes in some dimension content records. For example, the most commonly used administrative information will be affected by the adjustment of administrative information published by the Ministry of Civil Affairs every year, which need to be updated accordingly. The following takes the historical changes of administrative information of Jinning District, Kunming City, Yunnan Province in the administrative dimension table as an example to illustrate the design.

If the construction for data based warehouse started in 2013, the information of Jinning District in the data warehouse at that time will be shown in Table 1.

Since 2016, the administrative code of Jinning District has been modified to 530115, so after the dimension table in the data warehouse is updated at that time, the information recorded in the dimension table should be shown in Table 2.

If the Ministry of Civil Affairs needs to adjust the administrative code of Jinning District to 530177 by the end of 2028, the dimension table records in the data warehouse should be modified again, and the adjusted dimension table information should be shown in Table 3.

Table 1. The oldest record content of Jinning district in administrative region dimension table

Administrative dimension	Code for Administrative Division (History)	Code for Administrative Division (Updated)	……	Effective date	Expiration date	Validation or not
13	530122	530122	……	1982-01-01	9999-12-31	Validation

Table 2. The record content of Jinning district in administrative region dimension table after one change

Administrative dimension	Code for Administrative Division (History)	Code for Administrative Division (Updated)	……	Effective date	Expiration date	Validation or not
13	530122	530115	……	1982-01-01	2015-12-31	Validation
87	530115	530115	……	2016-01-01	9999-12-31	Invalidation

Table 3. The record content of Jinning district in administrative region dimension table after two change

Administrative dimension	Code for Administrative Division (History)	Code for Administrative Division (Updated))	……	Effective date	Expiration date	Validation or not
13	530122	530177	……	1982-01-01	2015-12-31	Invalidation
87	530115	530177	……	2016-01-01	2028-12-31	Invalidation
234	530177	530177	……	2029-01-01	9999-12-31	Validation

Change and treatment of main contents of the above dimensional content:

(1) Whenever the information of an administrative region changes, a new row will be added to record the complete information without affecting the original information that has changed. The dimension key column value of the previously existing fact record at rows in the fact table is not required to be changed.

(2) While new rows are added, three auxiliary management columns need to be added, namely, row validation date, row expiration date, current effective description respectively: The first two columns indicate how long the information at rows have been in effect over the course of history, while the current validation description indicates which row is the last valid record for that dimension object. When a new row is added, the value of the "row expiration date" column of the new row should be set to 9999-12-31, and the column in the previous row should be changed to the confirmed expiration date. In the new row, to set the current validity description column to Effective, and the column in the previous row to invalidation.

(3) In the process of content change of a dimension record, columns that need to be used as grouping conditions are usually added to the dimension table as "tracking columns" according to business requirements. Tracking columns assume the responsibility of tracking the historical changes of attribute information described by them. The contents of this column must remain unchanged during historical changes. The "administrative division code (latest)" column in the example table above assumes this responsibility.

After the above techniques are used, the facts can be grouped according to the historical administrative division code or the latest administrative division code, which can better cope with the uncertain business requirements of the front end, but at the same time, it also brings a certain complexity of dimension table structure design and maintenance.

1.2.2 Dimensional role-playing

In mineral rights transfer based earnings management business, it is common to have multiple date dimensions with different meanings in a single fact table because relational database systems do not support multiple foreign keys in the same table associated with another table. Therefore, to implement the above business requirements, you are needed to create multiple physical date dimension tables, or create multiple views based on the physical date dimension tables, and then associate the physical tables or views with multiple date-like foreign keys in the fact table.The view approach is a better choice because the subsequent work of ETL processing is easier because there is no need to maintain the contents of two physical tables. This approach is called dimensional role-playing, as illustrated in Figure 1.

In role-playing, column names in different views should have their names and meanings redefined so that similar fields from different views can be expressed differently and accurately in BI reports. "Payment Information Push Date" and "actual payment date" in Figure 2.

1.2.3 Shared dimension design

In the process of mineral rights transfer based earnings management, reserve management, mining right trade based management departments, financial department has its own demands on business analysis.Based on this, a corresponding fact sheet can be established for the business of assessing, contract signing and fee collection to meet the analysis needs of the above three types of users respectively. However, another problem is usually faced at this time, as the superior managers of these three departments, they are often required to understand multiple measurement information in each sale through a comprehensive report. For example, the evaluation price in a mineral rights transfer process, the transfer price in the contract, collection fees levied. In this case, the best choice is to use the shared dimension (or consistency dimension) to model the evaluation business contract signing business fee collection business, and then at the BI layer you can easily aggregate an example of multiple fact tables using consistency dimension as shown in Figure 2.

In the figure above, the consistency dimension -- mining rights assignment dimension needs to be established in the ETL process first, and then shared by different business facts. Finally, the mining rights assignment dimension can be easily implemented through SQL to achieve cross-table drilling. In the process of cross-table drilling, attention can be given to the use of full external connection, so as not to exclude the assignment records that have been evaluated but have not yet been signed or Transfer record of contract signed but not yet paid from the final report record. Table 4 shows an example of the content of a statistical report formed by cross-table drill-down

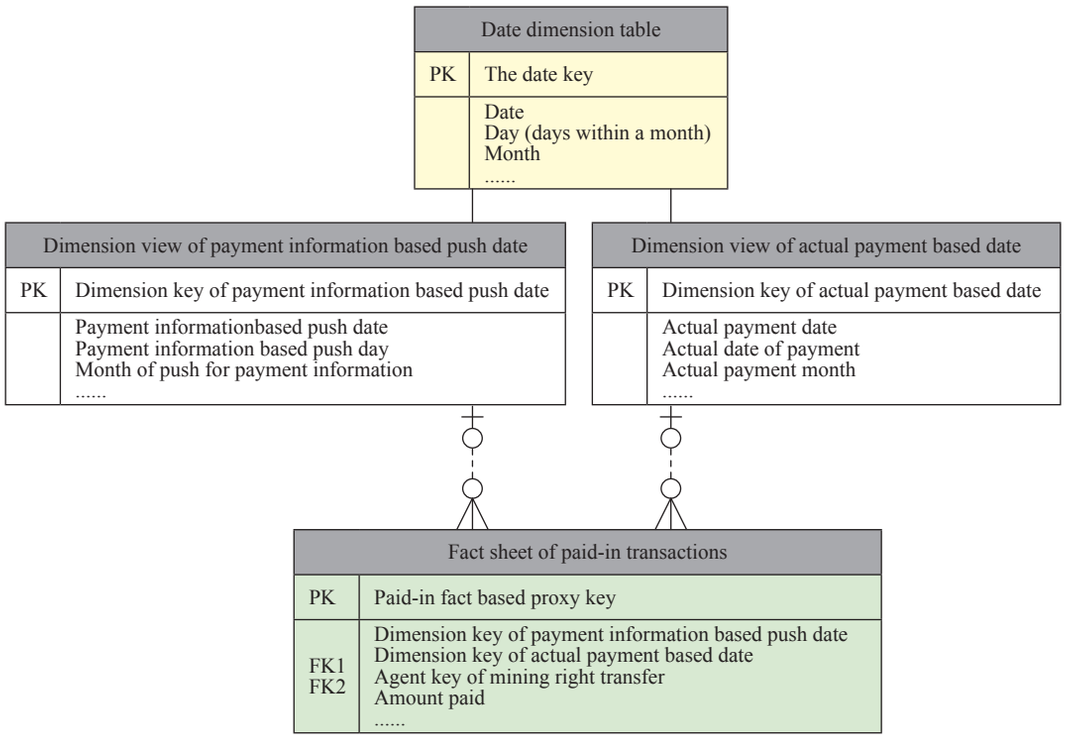

Figure 1. The example of implementing role playing of date dimension by view in relational database

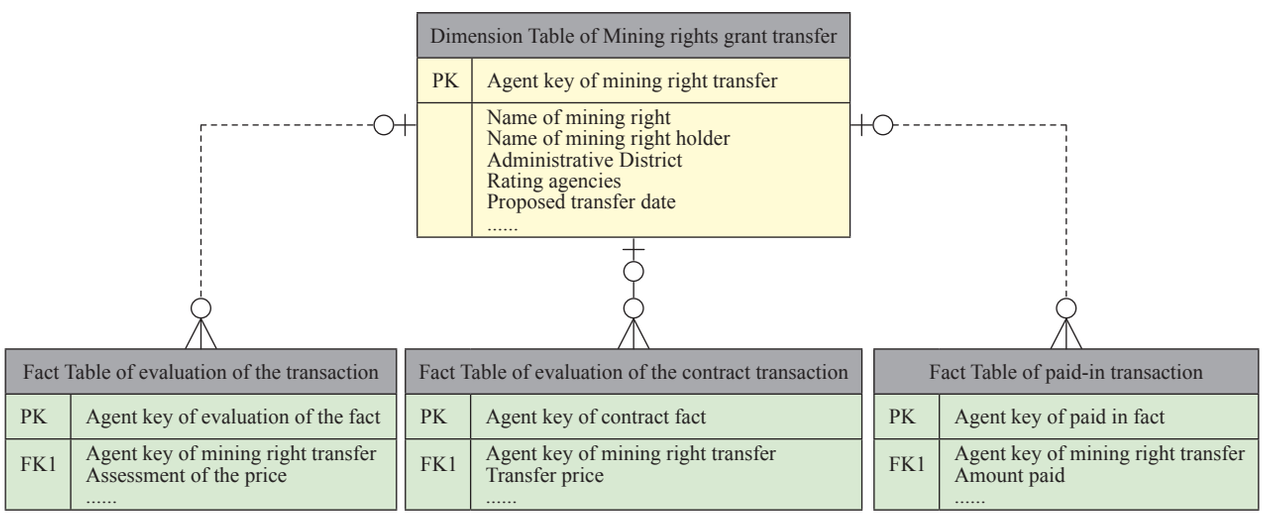

Figure 2. The example of common dimension-multiple fact tables co-referenced the same dimension table

Table 4. The example of one statistical report's records by drilling across different tables

Transfer number	Transfer date	Name of mining right	……	Assessed Price (ten thousand yuan)	Transfer Price (Ten thousand yuan)	Amount paid (ten thousand yuan)
111	20170819	XXX coal mine	……	180.53	208.00	208.00
112	20170819	XXX tin ore	……	19650.00	19650.00	1189.00
113	20170821	XXX lead zinc ore	……	3809.00	3809.00	2076.00

1.2.4 Design from audit dimension

In the process of establishing the business fact sheet for income management of mineral rights transfer, There are many data quality problems in historical fact records, so in the process of ETL, a lot of descriptive information about data quality will be generated, such as outliers, missing field content and so on, etc., as well as the ETL processing information such as batch number, batch processing time, processing personnel, etc. All these can be classified as ETL audit information. It is a good design idea to set up the corresponding dimension table of audit information and refer to it by the fact table. On the one hand, it helps to track the impact of ETL efforts on the quality of fact tables, and on the other hand, it helps to answer the questions of senior business managers about the reliability and authenticity of data content. An example of the audit dimension is shown in Figure 3.

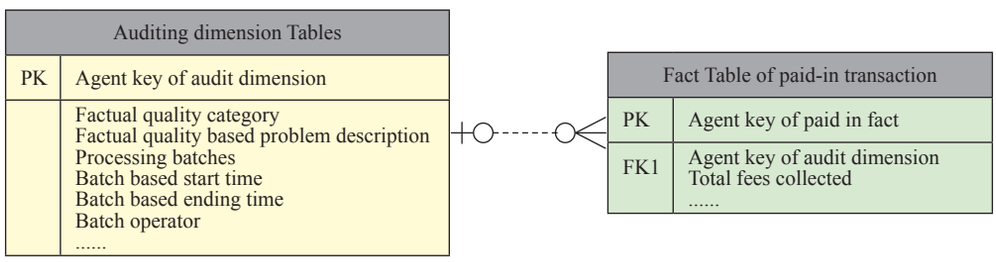

Figure 3. The example of fact table referenced audit dimension table

When the audit dimension is used, it is convenient to answer questions such as "how many contract records fail to be traced in their corresponding original evaluation report records" or "What is the total amount of the paid payment of the non-tax certificate".

2. Process of design and practice

2.1 Identification of the business process to be modeled

Business processes in dimensional modeling are generally micro activities of organizations[4]. The business process selected in this paper is the tasks allocated in the paid information of mineral rights transfer income from the non-tax collection and payment management department. The business process of receiving paid-in information is supported by a developed business software system, which records a series of business facts, such as payment of principal, late fees, capital occupancy expense and other numerical measures.

2.2 Determination of the granularity of the facts

The process of determination of fact granularity precisely defines the business content expressed by a single row in a fact table [4]. The granularity suitable for modeling in this paper is defined as: record one line of paid-in information fed back to the business software system by the non-tax collection and payment management department.

2.3 Determination of dimensions and facts

The dimensions are typically designed to form a collection of dimension tables that contain consistent dimensions viewed from the entire data warehouse level, as well as proprietary dimensions that are limited to a particular business item. At the granularity identified above, the relevant set of dimensions is shown in Table 5.

The related dimension and fact table structures are shown in Figure 4.

Illustration on key issues related to the above design:

(1) In addition to the primary key and dimension based foreign key of the fact table, the columns in the fact table usually only contain the fact column. Efforts should be made to try to avoid placing any descriptive information based columns into the fact table column. The descriptive information should generally be in the dimension table.

(2) The "ID of the associated contract of this payment" in the above fact table is a degraded dimension key, which is a dimension key without corresponding dimension table. The fact table retains that the function of the associated contract ID of this payment, as the key of degradation dimension, is to group and summarize the installment payment based information according to the contract ID. It's still essentially a dimensional foreign key.

(3) The fact table contains four facts: To withdraw the principal, confiscate the fine for delaying payment, the capital occupancy expense,and confiscate the total cost. The total cost collected is the sum of the preceding three items.The addition

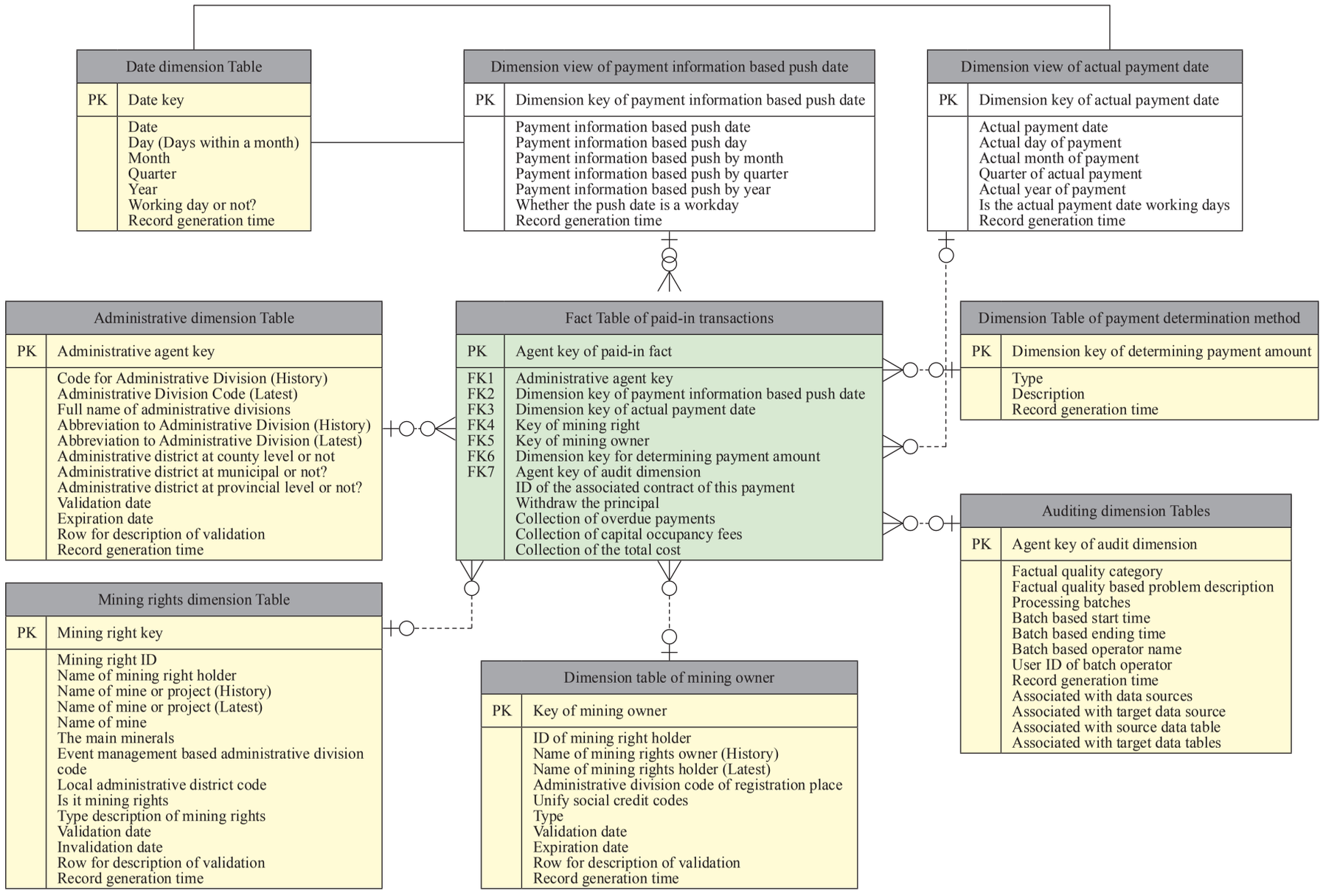

Figure 4. The E-R diagram of dimension tables and fact table

Table 5. The example of dimension set

Dimension names	Types	5W2H related items	Factors for consideration
Payment information push date	Conformed Dimension	when	Date dimension role-play can be used to filter or group statistics by year, month and day, scroll up and drill down between different levels
Actual payment date	Conformed Dimension	when	Date dimension role-play can be used to filter or group statistics by year, month and day, scroll up and drill down between different levels
administrative region	Conformed Dimension	where	It can be used in accordance with the administrative division screening or group statistics, scroll up and drill down between different levels
mining right	Conformed Dimension	what	It can be used for screening or grouping statistics according to the attributes related to mining rights (such as type of mining right type)
Person with mining right	Conformed Dimension	who	It can be used for screening or grouping statistics according to the related attributes of mining rights owners (such as type of economy and type of mining rights owners)
Method for determining the amount of payment	Special business dimensions	what	It can support screening or grouping statistics according to the classification of competitive transfer by agreement
Contract ID	Special business dimensions/ Degenerate Dimension	what	The contract details in transactional systems can be viewed in BI system and can be used as a condition for grouping statistics
Audit	Special business dimensions	why	It Supports group statistics and filtering based on data quality types

of this redundant column to the fact table reduces computation in subsequent summaries and improves performance, but consumes more storage space, which is also consistent with the design philosophy of analytical data structures

3. Application effect

The results of the dimensional modeling design described in this paper are carried by the Greenplum data warehouse. After setting up the dimension and fact tables described in the previous section in Greenplum, you can perform the query shown in Figure 5 to summarize the fact tables:

```
Select
        DWS_COMMON_D_REGION.JC_LATEST as "administrative region",
        DWS_COMMON_D_KYQ.KYQMCMS as "Type of mining rights",
        sum(DWS_CRSY_F_SJ.BJ) as "Principal (ten thousand yuan)",
        sum(DWS_CRSY_F_SJ.ZJZYF) as "Capital occupancy expense (ten thousand yuan)",
        sum(DWS_CRSY_F_SJ.ZNJ) as "overdue fine(ten thousand yuan)",
        sum(DWS_CRSY_F_SJ.ZFY) as "Total fees collected (ten thousand yuan)"
From
        DWS_COMMON_D_KYQ, DWS_COMMON_D_REGION,
        DWS_COMMON_D_DATE, DWS_CRSY_F_SJ
Where
        DWS_COMMON_D_DATE.YEAR=2017 and
        DWS_COMMON_D_DATE.MONTH=10 and
        DWS_COMMON_D_REGION.ZSJXZQBJ='yes' and
        DWS_COMMON_D_DATE.KEY_DATE=DWS_CRSY_F_SJ.KEY_DATE and
        DWS_COMMON_D_REGION.KEY_REGION=DWS_CRSY_F_SJ.KEY_REGION and
        DWS_COMMON_D_KYQ.KEY_KYQ=DWS_CRSY_F_SJ.KEY_KYQ
Grouping by
        DWS_COMMON_D_REGION.JC_LATEST,, DWS_COMMON_D_KYQ.KYQMCMS
```

Figure 5. Schematic diagram of SQL statement summarizing dimensional modeling results

The SQL statement above has the following connotations:Grouping can be made according to the latest abbreviation to administrative region and type of mining rights.

Statistics on the payment of administrative departments in Yunnan province(municipality) in October 2017 from the four measurement of principal, capital occupancy expense, overdue fines and total amount collected can be made.You can load the query results to the BI system or service software for display.

4. Conclusion

This paper describes a design practice process of dimension modeling in the construction of data warehouse in the field of mineral rights transfer based revenue management. The application effect of design results in actual projects is shown as follows:

(1) Dimension modeling is applied to design data warehouse model of mineral rights transfer revenue management,which can effectively solve some problems in the construction of data warehouse for mining rights transfer based revenue management in approach, model expansion and other aspects.

(2) It is possible to design a data warehouse that is easy to be understood with high query performance for management personnel for mining rights revenue, which provides a practical case for reference for the establishment of high-quality industrial data warehouse in the future.

Virtually, the theoretical system of dimensional modeling technology is well developed with complexion.Part of theoretical system was elaborated in this paper. If the needs for industry data analysis proves to be complex and changeable, other aspects of dimensional modeling techniques need to be further studied to meet the needs of actual projects in different industries.

References

[1] Bianchengxiaochao. Basic characteristics of data warehous[EB/OL]. (2019-12-04)[2022-01-28]. https://blog.csdn.net/u012955829/article/details/103396155.

[2] 1024 Xiaoshen. Easy to understand warehouse modeling: Paradigm modeling and dimensional modeling[EB/OL]. (2021-12-13)[2022-01-28].https://blog.csdn.net/weixin_44786530/article/details/121910025.

[3] Beilamei. Data warehouse modeling method and modeling case[EB/OL]. (2021-12-13)[2022-01-28]. https://blog.csdn.net/ytp552200ytp/article/details/107629384.

[4] Ralph Kimball, Margy Ross. *The Data Warehouse Toolkit: The Definitive Guide to Demensional Modeling, Thrid Edition*[M]. Beijing: Tsinghua University Press, 2015.

[5] Liu Jun, Yang Hongquan. Research on data warehouse scheme for dispatching and maintenance analysis of EMU Depot[J]. *Railway Transport and Economy*, 2018, 40(3):6.

[6] Xiang Shiyao. Research on the construction, analysis and application of multi-dimensional attribute space data cube of moving objects[D]. China University of mining and Technology (Beijing).

[7] Wu Qunyong, Liu Mengxin, Li Xijie, et al. Spatial OLAP system integration for environmental data set[J]. *Journal of Earth Information Science*. 2013, 15(01):90-96.

[8] Chen Jun, Zhuang Yifei, Cun Meili, et al. Multidimensional analysis model and method of urban public transport travel based on APTS big data[J]. *Transportation system engineering and information*, 2019, 19(1):7.

[9] Liu Yanjun, Feng Yu, Wu Qianhui, et al. Construction technology of distributed real-time data warehouse for state analysis of large equipment[J]. *Computer integrated manufacturing system*, 2017, 23(10):10.

[10] He Chengang. Research on late data in data warehouse[J]. *Computer applications and software*, 2013, 30(12):4.

[11] Gao Liang. Research on the construction of data warehouse in Colleges and Universities[J]. *Journal of Wuhan University: Science Edition*, 2012(S1):6.