The application of data warehouse and data mining in fracturing engineering system

Shuangshuang Ren 1,∗, Fei Shen 1, Hao Xu 2, Meng Li 3 and Jun Wu 1
1Research Institute of Drilling and Production of Jidong Oilfield, China
2China National Oil & Gas Exploration and Development Co., Ltd., China
3CNOOC Research Institute Co., Ltd., China

∗Corresponding author e-mail: rsshuang@126.com

Abstract. With the development of low permeability oil reservoirs in oil fields, the continuous development of fracturing technology and the optimization of fracturing design, the analysis of effect needs to be fast and accurate. The fracturing project is a systematic project that requires full-process, comprehensive data support and analysis. To achieve this, it is necessary to analyse a large amount of data and share resources. The development of data warehouse system and information network technology exactly meets the above requirements. At the same time, data warehouses, with the aid of online analytical processing (OLAP) and data mining tools (DM), can help engineering technologies manage, synthesize and analyse vast amounts of data in a better way to support management decisions. This paper takes fracturing engineering as an example, illustrate the contents of the data warehouse system's structure, design and implementation, and data presentation, proposes a data warehouse solution for fracturing engineering. Through the use of data warehouse technology and data mining technology, this paper has constructed a data warehouse with specialties in the field of fracturing engineering, providing technical support for engineering management and technical personnel at different levels and departments to provide effective decision analysis.

1. The purpose of developing a data warehouse in a fracturing project

The traditional databases have achieved great success in daily transaction processing. However, the demand for decision analysis of managers cannot be met because managers often want to be able to understand the development trend of the business by analyzing a large amount of data in the organization. The traditional database only retains current business processing information and lacks much historical information needed for decision analysis. In order to meet the management personnel's decision analysis needs, it is necessary to generate a data environment that adapts to decision analysis on the basis of the database - data warehouse [1,2]. The data warehouse is a topic-oriented, integrated, non-volatile, time-varying set of data used to support executive decision-making. Its main role is to conduct long-term trend analysis and provide technical support to decision-makers.

The oil and gas fracturing project have the characteristics of large scale, large investment, and high risk. The technical methods, decision analysis and target schemes of fracturing projects all have a direct bearing on the benefits and successes of construction. In addition, the oil and gas fracturing
project include many aspects such as reservoir conditions, drilling operations, cementing operations, completion operations, fracturing design and fracturing operations, etc., involving many functional departments distributed in different regions [3,4]. Because of the variety of information and the huge amount of data, most of the accumulated data has been stored in traditional ways, basically only satisfying users' needs for data storage, query, and statistics. However, for fracturing decision-makers, how to obtain timely, accurate, scientific and effective decision-making basis from these large amounts of data is very important. Based on the above requirements, a data warehouse solution with special characteristics in petroleum is proposed.

2. System structure
The development of a fracturing engineering data warehouse is the same as the development of other data warehouses, using the same engineering methodology, or the same software support tools, but the fracturing project has its own strong professional characteristics and cannot be copied from other data warehouses directly. Because the data warehouse is not a product, although it needs a certain software product to support it, the data warehouse itself must be constructed according to the characteristics of the enterprise's own application. Figure 1 shows the system structure of the data warehouse solution in the fracturing project.

![Figure 1. System structure of the data warehouse solution in the fracturing project](image)

The workflow of the system is as follows:

The data acquisition system collects all kinds of engineering data under the control of the common data model and stores the data in the data warehouse after being reorganized and classified. According to the data granularity classification level, the data entering the data warehouse is aggregated in different degrees. It provides end users with multidimensional views, reports, or charts through data presentation tools.
3. Design implementation

3.1. Data Structure Design
Multidimensional database organization is a form of data organization that directly addresses OLAP analysis operations. The data in the data warehouse is organized thematically. Fact themes are the criteria for categorizing data at a higher level and often reflect the issues that policy makers are most concerned about. Through the investigation of the decision-makers, the subject of the fracturing engineering data warehouse can be determined based on the operational requirements of the fracturing project. The following uses the example of “fracturing technique” as an example to illustrate the data warehouse data structure definition process. Select five dimensions related to fracturing technique analysis which is time, stratum, tools, hydraulics, fracturing fluid, and get a multi-dimensional architecture as shown in Table 1.

| Table 1. Multi-dimensional architecture of the data warehouse of fracturing technology |
|-----------------------------------------------|-----------------|-----------------|-----------------|-----------------|
| **Dimensions**                              | **Operation time** | **Formation parameters** | **Tool parameters** | **Hydraulic parameter** | **Fracturing fluid parameters** |
| **Category**                                |                  |                      |                  |                        |                                 |
| **Year**                                    |                  |                      |                  |                        |                                 |
| **Formation interval**                      |                  |                      |                  |                        |                                 |
| **Tool types**                              |                  |                      |                  |                        |                                 |
| **Structural features**                     |                  |                      |                  |                        |                                 |
| **Temperature**                             |                  |                      |                  |                        |                                 |
| **Pump pressure**                           |                  |                      |                  |                        |                                 |
| **Differential pressure**                   |                  |                      |                  |                        |                                 |
| **Pump displacement**                       |                  |                      |                  |                        |                                 |
| **Fracturing fluid parameters**             |                  |                      |                  |                        |                                 |
| **Type of fracturing fluid**                |                  |                      |                  |                        |                                 |
| **viscosity**                               |                  |                      |                  |                        |                                 |
| **Fluid loss coefficient**                  |                  |                      |                  |                        |                                 |
| **Consistency coefficient**                 |                  |                      |                  |                        |                                 |
| **Flow index**                              |                  |                      |                  |                        |                                 |

3.2. Design a multidimensional database
The logical structure design of each topic generally adopts a star model or a snowflake model, which is implemented in a data warehouse through a series of data tables. According to the multi-dimensional structure of the data warehouse, the definitions are developed from the topics. According to the star or snow model modeling method, a fact table (containing all dimension table detailed information entities code and all metric index values) is established for each topic. Every dimension in this fact table creates a dimension table (containing the dimension's detail entity code, category entity code, hierarchy code) and a corresponding detailed level or category table (code and name for each level or category). These tables all contain common fields as part of the subject code, which are linked together and merged into one topic.

4. Data conversion service
One of the challenges in creating a data warehouse is the data loading work after building a data warehouse [5]. Because the data in the data warehouse usually comes from multiple heterogeneous data sources, the data needs to be extracted, verified, cleaned, converted, and transmitted before entering the data warehouse. Data loading mainly includes the aspects as following:

1. Data extraction. Collect and extract data from external data sources based on data warehouse requirements.
2. Data cleaning. Due to the complexity of data sources, the data quality is uneven. In order to ensure the analysis and decision-making effect, it is necessary to strictly clean the input data warehouse data to ensure the data quality.
(3) Data conversion. Data conversion is to convert data in a data source into the data in a data warehouse according to conversion rules. Data conversion is divided into two steps. The first step is the composition of the conversion rules. The second step is the implementation of conversion rules.

(4) Data loading and data refresh. Data loading is the process of cleaning and converting the data of the data source into the data warehouse. The data loading is the procedure when data are processed and added to the data warehouse from the data source for the first time. The data refresh is date adding procedure that the data needs to be continuously updated in different time periods after data loading.

By using the DTS service provided in SQL Server 2000, the data loading of the fracturing engineering data warehouse can be quickly and easily completed. At the same time, the application for assisting DTS to complete data loading can be wrote.

5. Data service

Based on the idea of componentization, the functional units of the service are componentized and serviced, the degree of coupling between functional modules is reduced, the modules are reused, the requirements for business flexibility are met, and business applications are quickly delivered.

(1) Service decomposition. Based on service-oriented thinking, taking the business segment as the main body and taking the specific business itself as a carrier to sort out the business, then turn the business relationship into a domain model and build an enterprise resource service catalog.

(2) Process Management. Optimize the traditional workflow and sort out the business process.

(3) Functional design. Corresponding process classification and process nodes, combine function requirements and perform functional design.

(4) Function summary. According to the actual business needs, summarizes the business-specific functions, business sharing functions and technical common functions.

(5) Service refinement. Refine the summarized functions, output services and implement code reuse.

(6) Service registration. The service extracted will be registered on the enterprise service bus according to the enterprise resource service catalog.

(7) Service Orchestration. Services on the service bus are flexibly arranged according to business requirements to form business applications.

(8) The data mining-based fracturing process optimization service. First, use gray correlation to analysis the weight of production factors. At the same time, support vector classification machine and support vector regression are used as mining tools to complete output prediction and grade evaluation based on sample information of exploration wells. Finally, by referring to the construction parameters of similar reservoirs, solve the optimization problem of decision function in regression problem, and optimize the arguments, which is the optimization of fracturing construction parameters.

6. Front-end data display

The purpose of building a fracturing data warehouse is to meet the needs of various users for decision support in the fracturing engineering industry. They must query or retrieve data from the data warehouse when making different decisions and analyses. These queries refer not only to the query of record-level data but also to the query of the analysis results (trends or schema summary). Front-end data presentation uses OLAP, optimized query, statistical analysis, reporting and data mining tools to process and display data according to different levels of user query and analysis requirements. It includes creating multidimensional views of data for users to perform multi-dimensional analysis, generate various statistical reports and generate various graphics and images, etc. By using the PivotTable service provided by SQL Server, multidimensional data representation of the data warehouse can be implemented through programming in applications, Excel or Web pages to meet the needs of different levels of users for decision analysis.
7. Conclusions
Data warehouse technology provides an effective solution for the development of decision support system (DSS). The integrated data at different levels in the data warehouse can be used by decision makers at all levels for comprehensive data analysis, especially strategic analysis. The data warehouse solution based on fracturing engineering proposed in this paper can effectively provide DSS analysis for users at all levels of the petroleum industry, thus providing an implementation path for achieving networked, automated and scientific management of fracturing projects.

The application of data warehouse solutions has a broad space for expansion in the field of fracturing engineering. It can be modified on a data warehouse that has been built and successfully run to complete the data warehouse and expand the functionality to upgrade the data warehouse. For example, it can extend knowledge discovery and data mining analysis capabilities of data warehouses, build web warehouses to support web analytics, build distributed data warehouses and build enterprise-class data warehouses based on ERP. With the continuous development and improvement of data warehouses and related technologies, the application of data warehouse technology to large-scale, super large-scale integrated application systems in the oil field will surely become more and more mature.

References
[1] W.H.Inmon 2005 Building the data warehouse pp 21-53
[2] Y. Hongjin, Z. Junping and A. Chi 2001 Oil and gas drilling engineering economy pp 170-198
[3] J. Bischoff 1997 Data warehouse: practical advice from the experts pp 54-77
[4] P. Genmu 2002 Data warehouse technology and implementation pp 247-294
[5] Z. W. Ming 2002 Data Warehouse Principle and Application pp 240-243