Data Warehouse Design and Optimization for Drilling Engineering

Ning Jing*,1, Honghai Fan[1], Yinghu Zhai[1] and Tianyu Liu[2]

[1]China University of Petroleum, Fuxue Road 18, Changping District, Beijing, China; [2]Research Institute of Petroleum Exploration and Development, CNPC, Beijing, China

Abstract: With the development of petroleum informatization and increase of drilling data, data storage, analysis and integration is attached to the key planning process. Therefore, a solution is needed which combines the data integration, management, analysis and decision support. Data warehouse is one of the hot topics in computer technology application, which has solved the problem of data using after the application of information system. This paper puts forward the data warehouse design proposal in drilling engineering. The data warehouse project with drilling engineering is narrated, such as system structure, the design realization, the data demonstration and security policy. The authors also present a method of drilling data integration based on ontology. The data warehouse system which has the well drilling project specialized domain characteristics is developed by using data warehouse and communication technology. This system can provide effective decision support analysis for decision-makers in different levels and departments.

Keywords: Drilling Engineering, Data Warehouse, Ontology, Data Integration, Petroleum Informatization.

1. INTRODUCTION

With the continuous development of oil drilling technology and the continuous expansion of drilling scale, the amount of drilling information is also increasing progressively. Studies on how to effectively store, manage, analyze, and use the drilling data are of great significance, especially for deep exploratory well, which have double meaning of reality and urgency. So the comprehensive information analysis and management system is necessary for drilling data.

Two new decision-supported technologies - Data Warehouse [1] and Data Mining -rose in the mid-1990s are able to play an important role in this research. 2008, Michel Schneider [2] showed a general model for the design of data warehouses. Their proposition leans on a graphic representation which offers a visual help to the user. 2007, Shastri Lakshman and Heinz Dreher [3] designed ontology-based multidimensional modeling warehouse to offer an improved solution in time-depth conversion for seismic interpretation in onshore producing basions. 2007, Mike Dampier [4] described how data warehouse fitted into a service-oriented architecture. He also showed the observations and conclusions successful business process management of a common business scenario utilizing data warehouse within a service-oriented architecture. 2004, Robello Samuel G. [5] presented that through proper preparation and use of technology, something like eKnowledge factory can allow organizations to overcome the demo graphic's battle that the petroleum industry is facing in coming years. 1999, Raghubir Singh [6] described the process of establishing a well engineering data warehouse.

Applications had been modified to interface directly with this warehouse. Data had been separated from applications enabling easy access to one common source of reliable information. Randy E. Raley [7] described the development of a data warehouse that will be used to facilitate inspections through a graphical interface, store all the data in flexible format and allow the mining of the data for new information on the structure for corrosion control. The Data Warehouse would do analysis and predication by using historical data, which can grasp the drilling status quickly, accurately, comprehensively and timely of the whole oil field. And it will improve the level of drilling management and enhance oil drilling efficiency, so as to achieve the purpose of boosting economic benefits.

However, every operating department has different data requirements and databases, which have diverse data type, form and data code. So the data integration is primary task of drilling data warehouse. 2011, N. Prat, and I. Comyn-Wattiau [8] proposed to represent aggregation knowledge with objects (UML class diagrams) and rules in the Production Rule Representation language (PRR). 2006, Nicolas Prat and Jacky Akoka [9] presented a UML-based data warehouse design method that spanned the three design phases of conceptual, logical and physical. This method was proposed to data, which comprised a set of meta models used at each phase. In this paper, we develop a drilling data warehouse based on ontology, which is able to provide the semantic explanation of systematic data for data integration.

2 ANALYSIS AND DESIGN OF DRILLING ENGINEERING DATA WAREHOUSE

Definition of Data Warehouse was proposed by William H. Inmon [10] who is called the father of Data Warehouse in 1991 — a subject-oriented, integrated, relatively stable data collection reflecting the historical changes, used to support decision-making of management. The Data Warehouse is a

* Address corresspondence to this author at the China University of Petroleum, Fuxue Road 18, Changping District, Beijing, China;
Tel: (86)13488704284; Fax: (86)01089733221;
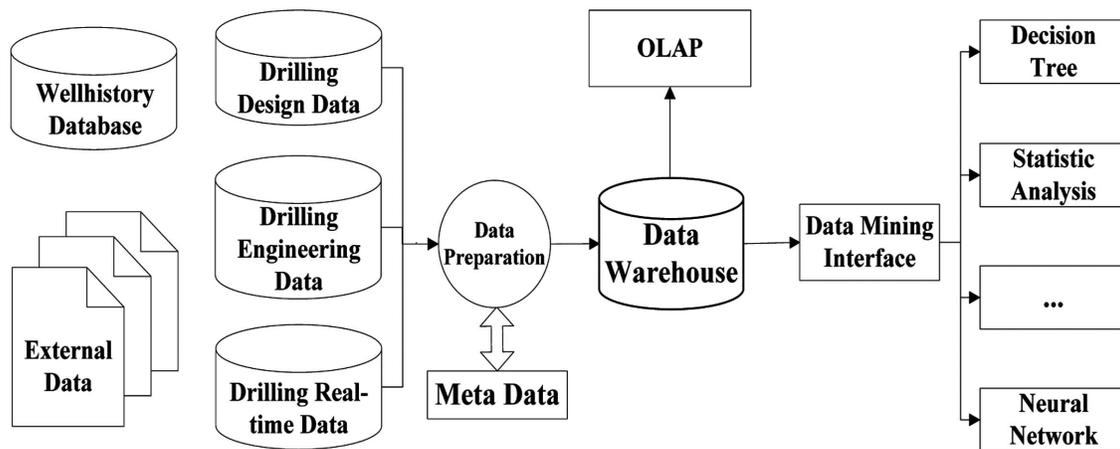E-mails: jingning1222@hotmail.com; jingning1017@126.com

Fig. (1). System Architecture of Drilling Engineering Data Warehouse Solution.

decision support system (DSS) and it is also the structured data environment for online analytical application, which is used for research and solving the problem of accessing to the information from the database.

For Special areas of drilling engineering, including pre-drilling operation, well design, drilling operation, well cementation and well completion, many categories of information and a large amount of data will emerge. New data management system has to not only meet the user's needs of data storage, query and statistical, but also help to obtain effective decision-making basis timely and accurately from the huge amount of data, which proposed to build the data warehouse system which has the well drilling project specialized domain characteristics [11].

2.1. System Architecture

The drilling Data Warehouse based on database is a platform for data re-organization which offers data analysis and data mining to the drilling design and construction decision-making. The data warehouse is a process of problem-solving, rather than a product. Although it needs a certain support of software product, the system must be built based on the characteristics of the industry. Fig. (**1**) illustrates the system architecture of data warehouse solution. The workflow of system as follow:

(1) A variety of raw drilling engineering data were collected to the data preparation area. Then it will be loaded to the data warehouse under the control of the common data model (metadata) after extraction, cleansing and conversion;

(2) Summarize the data in the data warehouse according to the division level of data granularity;

(3) Analyze the data using Data analysis tools (online analytical processing, data mining), so that the report or chart will be presented to final users in the way of multi-dimensional view.

2.2. Functional Requirement

The main function of the data warehouse for drilling engineering, its main is extracting, cleaning, transforming and loading the raw data. At the same time the data would be loaded into the drilling data warehouse accurately and timely. According to the specificity in the areas of drilling engineering, it should achieve the following functions:

(1) Extract, transform and load from the source data regularly and automatically;

(2) Clean data and separate dirty data according to the requirements of user;

(3) Automatically assign the missing data a value, in accordance with user requirements which can be average or high-frequency values;

(4) Convert a typical data to standard data to match the meta data automatically;

(5) Has a logging feature which can capture system abnormalities, and improve the robustness of the system;

(6) Meet the security requirements of the database system, including the integrity of the database, the integrity of the database elements, auditable, access control, user authentication and availability.

3. INTEGRATION OF HETEROGENEOUS DATA

Drilling engineering is huge and systematic, which has various kinds of business and a large amount of information. The business systems established by each unit lack general plan and design coordination in business and data association. They use different development forms and databases. Also they have different data types, storage methods and explaining standards, so that heterogeneous information islands would be formed. It can neither apply data comprehensively, nor meet the needs of integrated business management and data support on department decision [12]. This paper presents a method of heterogeneous data integration based on ontology, which is able to integrate the data of drilling engineering and improve the efficiency of drilling data warehouse.

3.1. Ontology

Ontology, which was originally a branch of philosophy, is used to represent the essence and organization of things. Philosophers use it to answer the basic questions of approaching things. In 1993, Gruber proposed the definition of ontology — Ontology is an explicit specification of conceptual model. More generally, ontology is used to describe the concept and the relationship between the

concepts of a field or a more extensive range. It would offer vocabulary that represents and communicates the knowledge in a special field, as well as the relationship collection that contains the vocabulary term at the conceptual level, so that these concepts and relationships have explicit and unique definition which can be recognized easily by people within a sharable scope. In this way, the communication between machines or human-machine would come true. An ontology is a normative description of a specific field, which includes concepts, attributes and constraints. Study on data integration based on ontology is very active, being widely used in information retrieval, information integration and machine translation etc.

Semantic heterogeneity of data sources in data integration has become increasingly prominent. Ontology is a conceptualized description of the basic properties of things, so we can use ontology through a computer-readable way to describe the data source information and global data model, and use the global body to establish a shared vocabulary and domain knowledge of a to-be integrated field. All the distributed data sources take advantage of the shared vocabulary and shared knowledge in the global ontology to decrease the semantic heterogeneity problems of data from each data resources to the greatest degree.

3.2. Drilling Data Integration

For drilling engineering data warehouse system, we can adopt hybrid ontology method. The hybrid ontology method is that each information source has its own ontology to describe its semantics. On the uppermost level, we establish a drilling shared vocabulary set, including basic terms in this field. The advantage of this method is able to support access and evolution of ontology, making it scalable, and its structure is shown in Fig. (**2**). Heterogeneous data integration system, with the establishment of mapping from data source to local ontology and local ontology to the global ontology, establishes a unified semantic of data source to complete the logical focus of heterogeneous data sources.

3.2.1. Construction of Local Ontology

Local ontology corresponds to the bottom basic database such as drilling design database, drilling wells history database etc. And data dictionary extracted from the basic
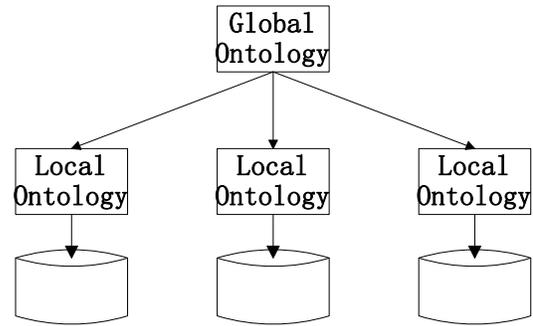

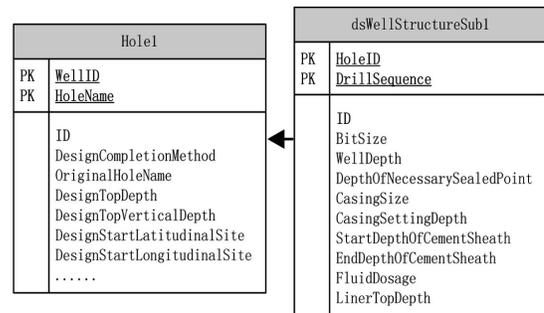**Fig. (2).** Chart of Hybrid Ontology Integration.

Fig. (3). Relation Diagram of Well Structure Design S1

database is used to build the local ontology. Take the well structure design database S1 for example, this database mainly records borehole data and well structure details information. The relationship between tables is shown in Fig. (**3**).

Then we can get the ontology description O1. The mapping information from O1 to S1 is shown in Table **1** and Table **2**.

Table 1.   Relational Mapping from O1 to S1.

ID	Class	Table
1	Hole	Hole
2	WellStructure	WellStructure

Table 2. Mapping from O1 the data type to S1 the field.

ID	Property	Type	Oid	ID	Property	Type	Oid
1	ID	int	1	11	Hole	int	2
2	Well	int	1	12	DrillSequence	nvarchar	2
3	HoleName	nvarchar	1	13	BitSize	real	2
4	DesignCompletionMethod	nvarchar	1	14	WellDepth	real	2
5	OriginalHole	nvarchar	1	15	SealedPointDepth	real	2
6	DesignTopDepth	real	1	16	CasingSize	real	2
7	DesignTopVDepth	real	1	17	CasingDepth	real	2
8	DesignLatitudinalSite	real	1	18	StartDepth	real	2
9	DesignLongitudinalSite	real	1	19	EndDepth	real	2
10	…		1	20	FluidDosage	real	2
				21	LinerDepth	real	2

Similarly, drilling assembly design database S2, stores information of drill name, steel grade, tensile safety coefficient, torsional strength. The relationship between tables is shown in Fig. (**4**). The construction method of local ontology O2 is same as O1.

3.2.2. Construction of Global Ontology

Global ontology, integrated from the local ontology, corresponds to the logical structure of the system database and forms the related mapping information. Fig. (**5**) shows the global ontology built by local ontology O1 and O2. The mapping information from global ontology to the local ontology is shown in Table **3**.

Table 3.   Class Mapping from Global Ontology to Local Ontology.

ID	GClass	LClass	Sourse
1	Hole	Hole	S1
2	WellStructure	WellStructure	S1
3	BHAUse	BHAUse	S2
4	BHA	BHA	S2

3.2.3. Integrated Ontology Mapping

After constructing the global ontology and local ontology, we need to establish the mapping of global ontology and local ontology, namely ontology integration mapping, to achieve the integration of heterogeneous system, which will relate two levels technology that is the concept merging and ontology mapping relation table [13, 14]. The concept merging refers to complete the calculation by the similarity of concepts and their attributes for semantic interpretation; ontology mapping relation table is created by merging mapping calculation.

4. FUNCTION MODULE DESIGN

4.1. Data Preparation

Data preparation is the key to the whole system as a link to connect the bottom original database and data warehouse, including metadata management module, ETL management module which will finish data extraction, calibration, cleaning, and conversion here.

4.1.1. Metadata Management Module

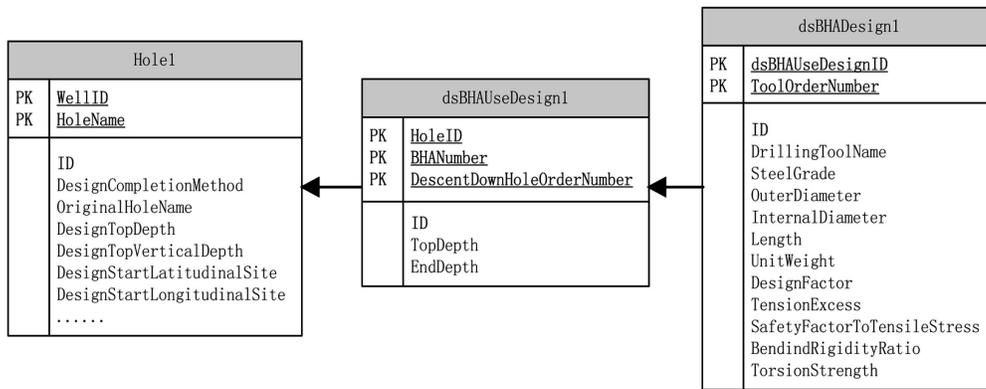Meta data drives ETL process of the entire system and metadata management module will mainly accomplish the

Fig. (4). Relation Diagram of Drilling Assembly Design S2.
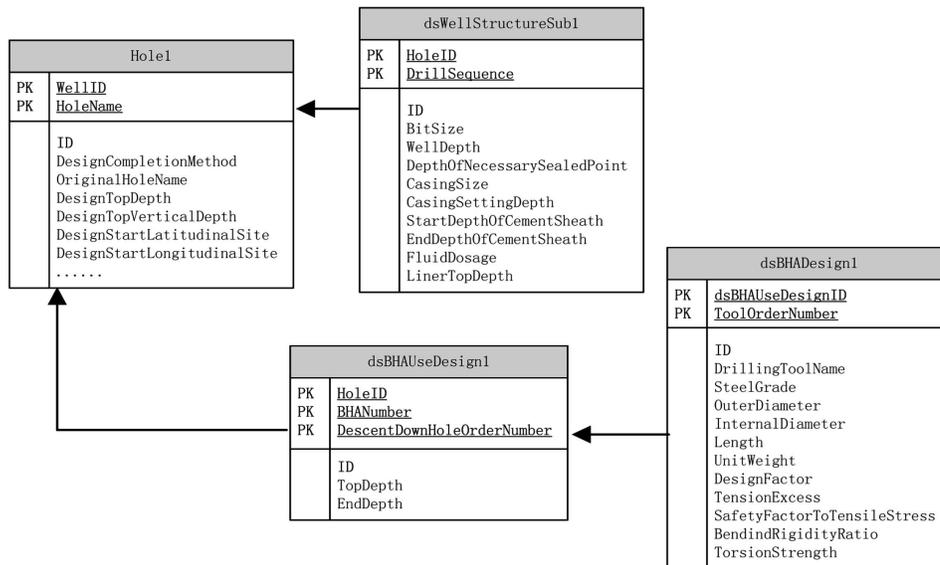

Fig. (5). Global Ontology Diagram after Integrating Data Source S1 and S2

functions of access on source and target database metadata, metadata storage and data query etc. The metadata include the source and target database metadata (database, database tables, and table field properties), the log data (task name, the beginning and end time, conversion bar, successful and failure number items, etc.), task metadata (source field, target field, the conversion rules, cleaning rules, task type, etc.).

4.1.2. ETL Management Module

ETL management module performs data conversion tasks stored in metadata which extracts tasks and resolve them to the rules of data cleansing, transformation and loading, and cleans them according cleaning rules of the relevant data, then carries through data conversion based on conversion rules and at last generates dynamic INSERT statements based on loading rules which will load data into the target table of the drilling data warehouse.

4.2. Tasks Management

The tasks management module, including two functions of task configuration and task scheduling management, is used to record all kinds of configuration information. These functions will run through the various modules of the whole system [15].

4.3. Client Data Show

Client data presentation uses tools of online analytical processing (OLAP), optimization of the query, statistical analysis and data mining to process and display data according the user query and analysis needs of the different levels, which includes a multidimensional view to create data, generation of statistical tables, and generation of a variety of graphics and images. By using the PivotTable provided by SQL Server [16] it can implement drilling data warehouse multidimensional data display in Excel, application or web page in order to meet the needs of decision analysis of users at different levels. Table **4** shows the difference between conventional database and optimized data warehouse.

4.4. Security Policies

Security policy plays a vital role in the process of the development and application of database systems. According to the characteristics of the drilling data warehouse, we consider the security of the system from three aspects, which are availability, integrity and confidentiality.

4.4.1. Usability

The drilling data warehouse system contains a lot of drilling engineering data such as logging, drilling fluid, well history. Various types of data should be named and identified in strict accordance with the international drilling industry standard and stores them in corresponding database to ensure the long-term usability of the data.

4.4.2. Integrity

Drilling Engineering-oriented data warehouse based on SQL Server uses SQL-based security policy mechanism to ensure the physical integrity of the database. Ontology model reorganizes and maps the drilling data in accordance with the actual project which improves the logical integrity of the database.

4.4.3. Confidentiality

The authorization mechanism is an important way to achieve security and protection of a relational database. User authentication and access control have been used in the drilling data warehouse, thus ensuring the confidentiality of the system

CONCLUSIONS

(1) Data Warehouse has a wide application in the field of drilling engineering. This paper proposes a new design method of data warehouse for drilling engineering and discusses the metadata management of drilling data warehouse, as well as the process of extraction, transformation and loading to achieve data re-organization.

(2) The integration model of multi-source heterogeneous data based on ontology can solve the problem of multi-source heterogeneous data in drilling engineering and realize the exchange and sharing of the drilling data.

(3) The establishment of this drilling data warehouse can provide effective decision analysis and other technical support for project management and technical personnel of different departments.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENT

REFERENCES

[1]   W. Shan, *Data warehouse and on-line analytical processing*, Electronic Industry Press, 1998.
[2]   M. Schneider, "A general model for the design of data warehouses", *International Journal of Production Economics*, vol. 112, pp. 309-325, 2008.
[3]   S.L. Nimmagadda, H. Dreher, and M.R. Rajab, Ontology-based Warehoused Time-Depth Data Modelling Framework for Improved Seismic Interpretation in Onshore Producing Basins, In: *International Petroleum Technology Conference*, Dubai, U.A.E., 2007.
[4]   M. Dampier, P.M. Haines, S. Tassin, and J.Y. Miller, What Role Does a Data Warehouse Play in a Service-Oriented Architecture?, In: *Digital Energy Conference and Exhibition*, Society of Petroleum Engineers, Houston, Texas, U.S.A., 2007.
[5]   G.R. Samuel, and G. McColpin, Accelerating Drilling Technology with Knowledge Factory: A Changing Paradigm, In: *SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers, Houston, Texas, 2004.
[6]   R. Singh, H. Albuquerque, and A.K.A. Hinai, "Data Warehousing Within Well Engineering," In: *Middle East Oil Show and Conference*, Copyright 1999, Society of Petroleum Engineers Inc., Bahrain, 1999.
[7]   R.E. Raley, and M.C. Wang, "Data Warehouse and Data Mining Techniques Airframe for Corrosion Control," In: *NACE International*, 1999.
[8]   N. Prat, I. Comyn-Wattiau, and J. Akoka, "Combining objects with rules to represent aggregation knowledge in data warehouse and OLAP systems", *Journal data and Knowledge Engineering*, vol. 70, pp. 732-752, 2011.
[9]   N. Prat, J. Akoka, and I. Comyn-Wattiau, "A UML-based data warehouse design method", *Decision Support Systems*, vol. 42, pp. 1449-1473, 2006.

[10]   W.H. Inmon, *Building the Data Warehouse*, Beijing, 2000. Available from: inmoncif.com/inmoncif-old/www/library/white-prsltt build.pdf

[11]   L. Qi, and X. Yingzhuo, "Intelligent Decision Support System Based on Data Warehouse for Drilling Engineering", *ACTA Petrol Ei Sinica*, vol. 24, pp. 77-80, 2003.

[12]   S. Xinghua, "Data Integrating Technology Based on Data Warehouse for Drilling Engineering", *Computer Applications of Petroleum*, vol. 6, pp. 19-23, 2009.

[13]   L. Wentao, C. Zhihong, X. Yan, and L. Xingyi, "Traffic Heterogeneous Data Integrating System Based on Ontology", *Computer Systems & Application*, vol. 19, pp. 7-11, 2010.

[14]   W. Lingli, Y. Jianqiao, and S. Rongrong, "An Ontology - Based Approach to Heterogeneous Database Semantic Integration", *Computer Systems & Application*, vol. 17, pp. 31-33, 2008.

[15]   S. S. Arktos, "Towords modeling, design, control and execution of ETL Processes", *Information System*, vol. 26, pp. 537-561, 2001.

[16]   Y. Pengfei, *SQL Server database application development technology*, *The People's Posts and Telecommunications Press*, Beijing, 1998.