Chapter XI
Ontology-Based Data Warehousing and Mining Approaches in Petroleum Industries

Shastri L. Nimmagadda
Kuwait Gulf Oil Company, Kuwait

Heinz Dreher
Curtin University of Technology, Australia

ABSTRACT

Several issues of database organization of petroleum industries have been highlighted. Complex geo-spatial heterogeneous data structures complicate the accessibility and presentation of data in petroleum industries. Objectives of the current research are to integrate the data from different sources and connect them intelligently. Data warehousing approach supported by ontology, has been described for effective data mining of petroleum data sources. Petroleum ontology framework, narrating the conceptualization of petroleum ontology and methodological architectural views, has been described. Ontology-based data warehousing with fine-grained multidimensional data structures, facilitate to mining and visualization of data patterns, trends, and correlations, hidden under massive volumes of data. Data structural designs and implementations deduced, through ontology supportive data warehousing approaches, will enable the researchers in commercial organizations, such as, the one of Western Australian petroleum industries, for knowledge mapping and thus interpret knowledge models for making million dollar financial decisions.

OVERVIEW

Data in major commercial petroleum industries are complex in nature and often poorly organized and duplicated, and exist in different formats. Business, in these companies, is operated both in space and time. Due to the diverse nature of business products and operations in different

geographic locations, these industries demand more accurate and precise information and data. Businesses operating in multiclient or multiuser environments with redundant data are prone to carry information with several ambiguities and anomalies.

With the widespread use of databases and explosive growth in their sizes, petroleum businesses face a problem of information overload. Effectively utilizing these massive volumes of data is becoming a major challenge for this type of industry. Data searching becomes tedious when specific queries are made, due to the piling up of volumes of data and information accumulated in several places, such as Websites and Web servers. In order to compete and increase profitability in world markets, it is vital for fast growing businesses to carry out mapping and integration of multioperational data structures. This can deliver accurate and precise information, which is crucial for elegant and economic decision support. Information in the form of knowledge or intelligence extracted from business data always adds value to the quality of decision-making. For the purpose of building knowledge from petroleum business data, an ontology approach that supports data warehousing, combined with data mining and visualization techniques, is a significant breakthrough. This chapter addresses issues of importing/exporting data from Web resources or data from offline sources, their logical and physical storage, and accessing, interpreting, and presenting the explored information.

Sedimentary basins, which are known to bear oil and natural gas deposits, may consist of several petroleum systems. We examine these systems as synonymous with other information systems. Ontology is proposed for simplifying the complexity of petroleum exploration and production data of different petroleum systems. This has prompted the development of various conceptual models and translating them into logical data models by a multidimensional data mapping approach. These logical data models will be converted into implementation models, using a contemporary DBMS (for example, Oracle), such as warehouse approach, proposed by us. Specific requests will be made with queries for locating a specific piece of data or information from these warehouses. Simple mining algorithms will be developed and used for extracting patterns, correlations and trends from petroleum data. These patterns and trends are interpreted for a meaningful geological knowledge.

INTRODUCTION

Large amounts of petroleum operational data are routinely collected and stored in the archives of many organizations. Much of the data archived for informational, as well as audit purposes, are still under-utilized or many personnel do not know what to do with them. However, by analyzing the petroleum data of one basin, it would be possible to discover exploration, drilling, and production patterns of other basins and use these patterns for future planning of various classes of drillable exploratory or development wells. Such an approach was not feasible until recently due to limitations in both hardware and software. In the recent past (Pujari, 2002) there has been a tremendous improvement in hardware—several (gigabytes) GB of main memory, multi terabytes of disk space with multi GHz of processing speed on a PC. Thus, computer programs, which sift massive amounts of operational data recognize data patterns and provide hints to formulate hypotheses for tactical and strategic decision-making, can now be executed in a reasonable time. This has opened up a productive area of research to formulate appropriate algorithms for mining archival data to devise and test hypotheses. In this project, an array of ideas from computer science, information technology, statistics and management science are being applied for organizing