Open access • Proceedings Article • DOI:10.1109/CCECE.2004.1345217

An intelligent project lifecycle data mart-based decision support system

N. Houari, Behrouz H. Far

Institutions: University of Calgary

Published on: 02 May 2004 - Canadian Conference on Electrical and Computer Engineering

Topics: Data mart, Intelligent decision support system, Decision support system, Data warehouse and Knowledge extraction

An Intelligent Project lifecycle Data Mart-based Decision Support System

Nora Houari and Behrouz Homayoun Far
Department of Electrical and Computer Engineering
University of Calgary
2500 University Drive, NW, Calgary, Alberta, Canada T2N 1N4
Email: {nhouari, far@ucalgary.ca}

Abstract

Mission critical decision making in enterprises depends heavily on intelligent systems for extracting, analyzing and interpreting information from multiple heterogeneous, distributed data and knowledge sources. It is assumed that data warehouses (DW), data marts (DM) are required for optimized data accessibility and use. This paper discusses issues with the current DW/DM systems and propose a novel architecture based on multi-agents technology to support information and knowledge extraction over distributed data sources in order to use them in the decision making process. The proposed framework is applied to a real-world project lifecycle case that is EPC (Engineering Procurement and Construction) project.

Keywords: *Intelligent agents, Data warehouse, Data mart, Decision Support Systems, Multi-agent system.*

1. INTRODUCTION

Information outsourcing is one of the major concerns of large companies. In recent years advances in technologies such as networks and databases have enabled businesses to store enormous internal and external data to support decision making. Today, any enterprise to be competitive and able to react in an appropriated manner to market requisites needs to have an effective support to decision-making, which means effective and flexible Data Warehousing (DW) systems [11, 4, 6].

However, it is quite inefficient to dig in all raw data collected in the data warehouse of a project due to the fact that decisions are made to target specific enterprise's activities and area. Data Mart (DM) follows the same lead as DW in a small scale. It provides access to critical data faster than the enterprise data warehouse, allowing business managers to initiate business intelligence strategies and reap the rewards sooner.

Intelligent software agent technology can play an important role in the design and development of a DM for decision support. Intelligent agents can help discover, locate, and report the information [7]. When applied to a DW, agent technology allows improving its performance, making possible to transform conventional search and integration mechanisms with self-adaptation abilities to day-by-day needs of enterprise's decision-makers. By allocating regular data warehousing systems activities to an agent community, it is possible to: reduce management, monitoring and maintenance costs of a data warehouse system. It helps optimize selection, extraction, and integration processes; which will improve the quality of information through the application of intelligent techniques on consistency and redundancy control activities; and ensure higher levels of confidence on decision-making processes.

The main purpose of this work was to develop a decision support system (DSS) capable of dealing with the huge amounts of data available cross a distributed project lifecycle that is carried by an enterprise and its partner alliance and derive the metamorphoses of these data towards valuable and accurate knowledge retrieval for quicker, better quality, and confident decisions. The system provides several types of data access capabilities to retrieve and analyze the data contained in a data warehouse, on-line transaction processing operational (OLTP), and legacy systems for providing the critical information needed by business decision makers. It is applied to a real world project lifecycle case EPC (Engineering Procurement and Construction).

The structure of this paper is as follows: in Section 2 we briefly give an overview of data warehousing types and their drawbacks. In Section 3 we present the proposed framework, and finally conclusion and future work are given in Section 4.

2. BACKGROUND

A data warehouse is a "subject-oriented, integrated, time-varying, non-volatile collection of corporate data" [11]. So far several architectures of DW have been applied to multiple research applications and commercial

CCECE 2004- CCGEI 2004, Niagara Falls, May/mai 2004
0-7803-8253-6/04/$17.00 ©2004 IEEE

products. They all fit into one of the three types: *centralized DW*, *data mart*, and *distributed DW* [2, 1]. Figure 1 shows the architectures of these three types.

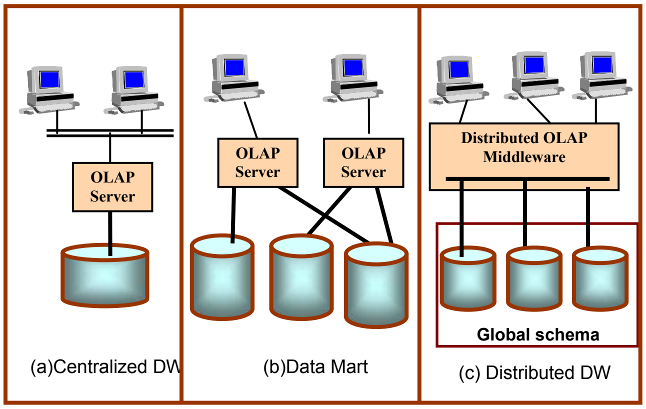

Figure1. The three types of DW

In a centralized DW all business data from different operational and functional departments of an enterprise is integrated and stored in a single database with a single enterprise model. Centralized DW are created to provide knowledge workers with consistent and integrated data from major business area. They offer the possibilities to correlate information across the enterprise. But due to the huge subject area and the vast amount of different operational sources that need to be integrated, centralized data warehouses are very difficult and time-consuming to implement.

Data Mart stands for highly-focused version of a DW. Because the data is highly focused, setup costs and time can be significantly reduced. The value of a data mart is measured by the efficiency of delivering a relevant and complete subset of data residing in data source to perform decision support. Data in data mart may be captured from legacy and operational systems and/or enterprise data warehouse.

Although the DM idea is quite attractive and offers the opportunity to build a corporate wide data warehouse from the bottom up, the benefits of data marts can easily be outweighed if there is no corporate wide data warehouse strategy. If data mart design is not done carefully, independent islands of information are created and cross-functional analysis, as in the enterprise data warehouse, is not possible.

The pragmatic approach is to integrate the DMs into a common schema. The result is a distributed enterprise data warehouse, also called federated data warehouse [10] or enterprise DM [3]. Under the distributed DW architecture, individual data mart can be built, managed and maintained flexibly. The complexity of the global schema remains a big challenge facing distributed DW. To solve this complexity issue we introduce a novel approach based on agent concept, in the next paragraph we look at what is an agent and what is the motivation for the choice of multi-agents as a solution technology.

3. PROPOSED SOLUTION

The aim of the proposed intelligent project lifecycle data mart based decision support system (DSS) is to improve the way project's lifecycle are handled in today's cross-organizational project. The system will not make decisions for the managers; it rather assists the manager to tailor the development process in a reliability driven manner. The DSS framework is geared towards improving the way information is gathered, managed, distributed and utilized by key business users, in order to assess managing complex projects lifecycle that are highly decentralized, dynamically changing and often in unpredictable environments.

3.1 Software Agents

A software agent is an autonomous computational entity with communication capability (social), able to react to environmental changes (react) and act in a goal-oriented way (proactive) [12]. Due to their characteristics intelligent software agents are strong candidates for acting as mediators between source and application of knowledge. Software agents possess several common characteristics, such as their ability to communicate, cooperate, and coordinate with other agent in a multiple agent system. Each agent is capable of acting autonomously, cooperatively, and collectively to achieve the collective goal of a system. The coordination capability helps manage problem solving so that cooperation agents work together as a coherent team. The coordination is achieved, for example by exchanging data, providing partial solution plans and constraints among agents [5]. Intelligent agents play the role of assistants by allowing managers to delegate work that they could have done to the agent software.

The choice of agents as a solution technology for our project lifecycle DSS was motivated by the following:

- The domain involves distribution of data and source of knowledge both physically and logically.
- The integrity of existing organizational structure and the autonomy of its subparts must be maintained.
- Interactions are fairly sophisticated, including information sharing, and coordination.

3.2 Framework

A snapshot of the framework is build up on the model of the decision support given by [8, 9] and is shown in Figure 2. The framework is composed of:

- *Multiagents* (MA): As we want to find something from the distributed databases, we need a multi-agents based system. At the same time, in any cases, the decision should afford on-line and dynamic information. Various agent types are used to realize the system goals. The agents' roles, capabilities, intelligence, autonomy, cooperation, communication language and protocol as well as shared ontology are considered.

- *Local Data Base Model*: this module is responsible for gathering and storing the relational database of the knowledge acquired.
- *Model Base System*: this module assists with generating alternatives (combinations of practices), and storing decision making model.
- *Knowledge based module:* for this module an expert system approach is chosen to implement the mapping of the practices. This module stores knowledge and incorporates mathematical solvers.

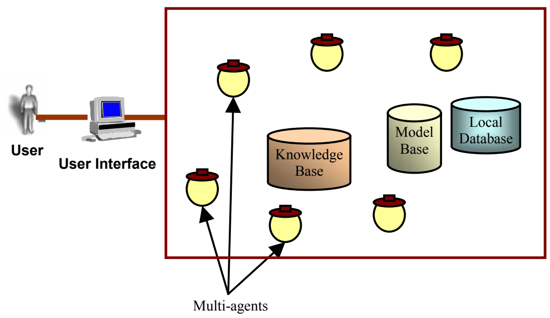

Figure 2. The DSS Framework

The system is organized into four layers (Figure 3).

- *The Source layer*- This bottom layer embodies all the data sources located in an enterprise project lifecycle environment, and this includes OLTP, legacy systems and DW.
- *The Acquisition and Integration layer*- While the previous layer guarantees the storage of the proper data, it is up to this layer to dig inside to extract valuable information from it. Here data is extracted, transformed, and normalized towards its analytical uses. There are also all the skills concerning data navigation and their queering manipulation.
- *The Decision layer*- Decision layer of the overall system is the DSS itself. In this layer the decision making task is accepted from the user and the relative decision model is selected and the decision process is carry out, and generate the result to the user in a more comprehensive way.
- *The Access layer*- This layer receives the user description of the decision to be performed, and deliver the results back of the decision process performed.

3.3 System Structure

The DSS is designed using (a) *user interaction* agent, (b) *decision making* agent, (c) *report/visualization* agent, (d) *knowledge acquisition* agent, (e) *ETL* agent, and (f) *Data-Guide* agent.

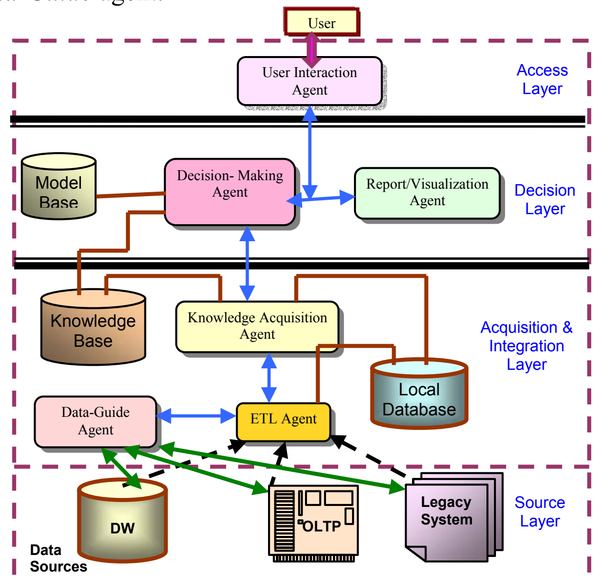

Figure 3. Overview of the DSS architecture

User Interaction Agent interacts with the user in assisting him/her performs the decision support activities. The user can provide a general description of the problem at hand in terms of high level goals and objectives, and decision to be performed. The user interaction agent is responsible for receiving user specifications and delivering results back. It also keeps track of user preferences and profile.

Decision-Making Agent is responsible for coordinating the various tasks that need to be performed. The structure of decision-making agent is shown in Figure 4. From the user's specified high level objectives, this agent converts them into specific lower level tasks and generates a plan of action. The decision making agent has knowledge about the task domain, as well as the capabilities of other agents; it may seek the services of a group of agents that work cooperatively and synthesize the final result. Thus, the decision agent is responsible for performing the actual decision activity and providing the results.

Report/Visualization Agent incorporates the result of the decision making process and generates the final results in a more comprehensible and natural way, making it more appealing to users. This agent contains pre-formatted report templates as well as a suite of visual representations for expressing the results of the decision support operation. It implements a variety of visualization techniques and generates the decision report.

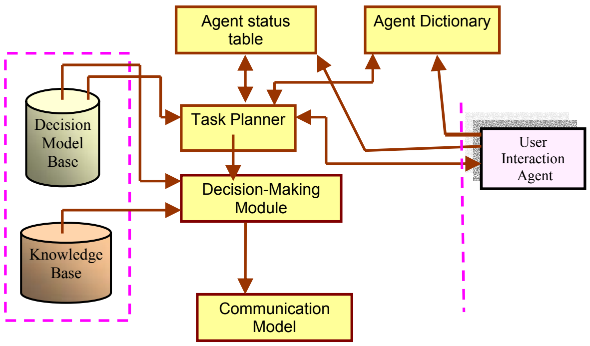

Figure 4. Decision making Agent Structure

The knowledge acquisition process is carried by the *Knowledge Acquisition Agent*. The acquired knowledge should be changed or modified corresponding with variety of data provided by ETL agents. Different agents should have different knowledge acquisition algorithm and feedback mechanism because they deal with different user requests. Agent should be able to valuate knowledge acquired by other agent before using it in decision making. Furthermore, they should be able to verify or reinforce the knowledge.

ETL (Extraction, Transformation, and Load) process is carried by *ETL agents* (Figure 5) who are distributed on databases that include DW, legacy and OLTP systems. These agents identify the relevant data sources, extract the data, then perform the necessary data cleansing and transforming. The extracted information then is loaded for use. An ETL agent task is the process where business data became business information.

The *Data-guide agent* is responsible for keeping track of what data is stored where and actively maintain meta-data about each of the necessary data for the project lifecycle. It takes into account the heterogeneity of the data sources as well as resolve conflicts in data definition and representation.

4. CONCLUSION AND FUTURE WORK

Data accessing and DSS system are increasingly becoming critical to organizations that wish to exploit operational and other available data to improve quality of decision making, as well as gain critical competitive advantage. The proposed framework in this paper is the first step towards this direction. In this work we look at both structure's drawbacks and opt for an agent oriented framework. The DSS presented in this paper was conceived upon multi agents and was applied to a real-world project lifecycle of an EPC project. Some challenges need further investigation, such as: how does the performance fall as the number of agents involved increased? For the moment we assume that the number of agents is determined with the initial scope of the project, but questions such as this is crucial if our system is to extend further more.

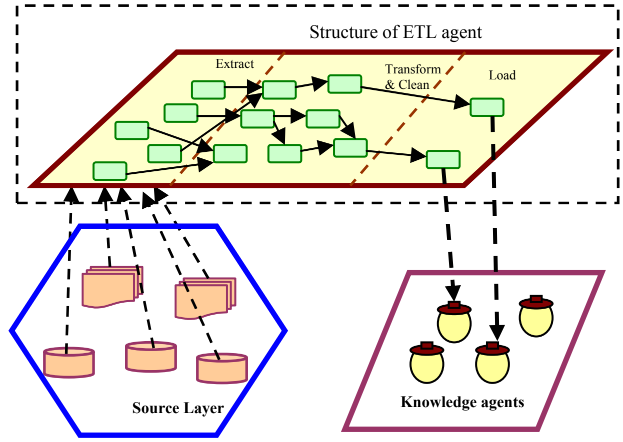

Figure 5. Structure of ETL Agent

References

[1] Albrecht. J, Lehner. W, "ON-Line Analytical Processing In Distributed Data Warehouses," 1998.

[2] D.Wells, P. Carnelley, "Ovum evaluation: the Data Warehouses," Ovum Ltd., London, 1996.

[3] Enterprise-Scalable Data Marts: A New Strategy for building and Deploying Fast, Scalable Data Warehousing Systems, White Paper, Informatica, http://www.informatica.com, 1997.

[4] P. Gray and H. Watson, "Decision Support in the Data Warehouse," The Data Warehousing Institute Series from Prentice Hall, 1998.

[5] P. Maes, "Agents that Reduce the work and Information Overload", Com. of ACM, vol.37, no. 7, pp. 30-40, 1994.

[6] R. Barquin and H. Edelstein, editors. Building, Using, and managing the Data Warehouse. The Data Warehousing institute, Prentice Hall, 1997.

[7] R. Chi, E. Turban, "Distributed Intelligent Executive System," Decision Support Systems, vol. 14, pp. 117-130, 1995.

[8] Sprague R.H. Jr, Watson H.J., "Decision Support Systems -Putting Theory into Practice." 2nd Edition, Prentice Hall, Eaglewood Cliffs, NJ, 1980.

[9] Turban, E., "Decision Support and Expert Systems: Management Support Systems." 4th Edition, Macmillan Publishing Company, NY, 1995.

[10] Wells, D., Carnelley, P., "Ovum evaluates: The Data Warehouse," Ovum Ltd., London, 1996.

[11] W. Inmon, "Building the Data Warehouse," John Wiley & Sons, 1996.

[12] Wooldridge, M. and Jennings, N., "Intelligent agents: Theory and Practice", The Knowledge Engineering Review, vol. 10, no. 2, pp. 115-152, 1995.