

Paper 170-27

IDW -- The Next Generation Data Warehouse

Larry Bramblett, Data Warehouse Solutions, LLC, San Ramon, CA

ABSTRACT

Data Warehouse systems collect, clean and manage mission critical information. Using statistical and targeted intelligence, they provide the business community with needed information and filter out unwanted data. SAS provides the software to make Data Warehouses come to life. Seems like everything is known about Data Warehouse development so much so that Data Warehouse systems are no longer considered an 'Emerging Technology'. Does that mean that everything about Data Warehousing has been done? This paper will take another look at Data Warehouse system design and what is needed to support the next generation of information platforms.

INTRODUCTION

The first generation Data Warehouse supports the cleaning, transformation and loading of corporate data into libraries. The warehouse corporate data is often the most primitive or lowest level of knowledge of the company. To make this knowledge available, Metadata is defined and processed using software such as SAS/Warehouse Administrator®. Metadata describes data attributes, transformations, aggregation levels, and job process information. Metadata helps socialize the Data Warehouse to the knowledge workers so that they can discover information contained and often hidden within the data.

The next generation Data Warehouse systems will be constructed as Intelligent Data Warehouses (IDW). IDW's are Data Warehouses that are managed as active data sources by a Knowledge Warehouse. The Knowledge Warehouse applies knowledge objects that are created and controlled by software engines using expert system models. Knowledge objects supply the rules, methods and procedures in a reusable way and are activated to not only link Data Warehouse data with a specific business process, but to re-package data so that the IDW can manage the business.

The Knowledge Warehouse manages business objects, rules and procedures to form business knowledge. Business knowledge forms the foundation and infrastructure to manage the business process component of the Intelligent Data Warehouse.

The tools are now available and the SAS *system* provides the toolset to link it all together. All that is needed is an understanding of what is missing from the first generation Data Warehouse model.

Current Business Intelligence Strategies

Today there are many authorities who point to or talk about Intelligent Data Warehousing. They advocate various methods to enhance the business intelligence in the first generation Data Warehouse. Many take the view that Data Mining provides intelligence to the warehouse. Some take the view that content management provides intelligence, while others talk about portal views into intelligence.

The Intelligent Data Warehouse requires the use of data mining to leverage knowledge discovery. Data mining extends intelligence to uncover relationships between warehouse data and business process. Without data mining strategies the IDW could not apply lessons learned from a historical perspective nor could the IDW bring those strategies into the present. By itself, data mining does not imply an Intelligent Data Warehouse but without data mining strategies that support the business process objectives the IDW is incomplete.

Software packages exist that support knowledge management; their main focus being knowledge content management. Knowledge content management provides a sophisticated means to manage documents and interaction between groups. E-mail, work group collaborations and document searches are the mainstay of this type of software. They support the IDW by providing information about the company. However, the knowledge contained within the Data Warehouse Metadata and the knowledge contained in content management is not connected.

One solution employed to solve this disconnect between data and knowledge is data portals. Data portals bring together data sources and content information. So one would say that portals provide a way to bring intelligence into the warehouse. This is true, but what about reusability of information? Often knowledge contained within content is not reusable and has a short life span that requires timely updates to be of value. Data portals support an important component of the IDW but are still only one piece of the IDW infrastructure.

These methods are all first generation business intelligence models; a second generation Intelligent Data Warehouse has all of these components and more.

The IDW Design Requirements

The Intelligent Data Warehouse includes a Knowledge Warehouse that supports linking reusable knowledge objects with flexible methods and processes. The structure of the Knowledge Warehouse and the Data Warehouse must be able to change with the company in a way that leverages information to meet company objectives.

To elevate a first generation Data Warehouse to an Intelligent Data Warehouse requires a four-part design. The design must support Knowledge Warehouse management of knowledge objects, and near to real time execution of rules and procedures on the company's business process. All of this must conform to a standard of software interoperability that allows these very different software objectives to work together.

IDW design and SAS

First, the warehouse must be managed by software that retains the Metadata. SAS/Warehouse Administrator® manages Metadata and supports the Object Management Group's (OMG) Common Warehouse Metamodel (CWM).

The CWM provides the framework needed to describe the data resources, data analysis and warehouse management information. The CWM defines a number of sub-Metamodels for Metadata interchange among Data Warehousing, business intelligence, knowledge management and portal technologies. By supporting the CWM the

SAS/Warehouse Administrator® software establishes the interoperability for the IDW between the Data Warehouse and Knowledge Warehouse.

Next, the warehouse must communicate across the business network to manage data in near real time; SAS® Integration Technologies supports this requirement.

Third, there must be a Knowledge Warehouse that contains knowledge base libraries. Within these knowledge bases there are objects with specific business rules, methods, and procedures focused on specific tasks. The knowledge objects may perform several functions but are always predictable. Knowledge bases work with other knowledge bases to enhance reusability. SAS procedures, methods and catalogs fall into this category. The Knowledge Warehouse is not engine specific but must conform to the CWM for interoperability.

Finally, the IDW must contain a software engine that employs an expert system to apply data from the Data Warehouse using the knowledge objects contained in the Knowledge Warehouse for specific business goals. This is an emerging area and there are number of third parties that support this requirement, including one that uses the SAS language.

The IDW Mission

The primary mission of the IDW is to link the Knowledge Warehouse and the Data Warehouse processes with operational systems to bring about quality, functionality and information sharing. The IDW Knowledge Warehouse like the Data Warehouse contains information about the business it supports. Information in the Knowledge Warehouse is at a level that supports process control, reporting and analysis, from both OLTP and Data Warehouse data. Each knowledge base found within the Knowledge Warehouse is dedicated to a specific subject of the business. The Data Warehouse will contain data in support of Knowledge Warehouse specific business goals by presenting data from a historical data source, data validation function point or for near real time analytics. The Data Warehouse could maintain data not related to the Knowledge Warehouse without impact to the IDW but the question should be why maintain data that is not defined by a knowledge process? What does it support in the business? The IDW reveals data that may not

have enough business value to be maintained in the Data Warehouse.

The Data Warehouse front-end systems use the IDW Knowledge Warehouse to control end-user requests, call standardized expert system routines and control information sharing. Business rules, procedures and process workflow are accessible to the end-users. This supports a stable, consistent method of retaining, transferring and adding knowledge to the business community.

The cost to retain knowledge, transfer knowledge or add knowledge is a significant expense to all businesses. The Knowledge Warehouse reduces the cost of training, information sharing and time to market by providing accessible business information. Raising this type of information to the sharable level enhances not only the Data Warehouse but also the very fabric of the company.

First Generation Business Intelligence and Data Warehousing

First generation business intelligence objectives focused on opening information within an organization to end-users. Data Warehouses, Data Mining, content management, and data retrieval all evolved to solve the problem of poor knowledge sharing. While this is better than ad-hoc reporting, first generation business intelligence fails to capture the knowledge life cycle of the business. It's not the data that has intelligence it's the process.

First generation business intelligence requires the use of technology to gain control of knowledge. Strategies for coding, transfer and transformation of data form the foundation to bring about sense out of the chaos of data that exists in every business. Frequently, there is very little understanding of the human process that produced the data in the first place. This knowledge is found in the business process and the life cycle of that process.

Knowledge Life Cycle

In the human act of conducting business, knowledge is socialized within subjects of the business. Claims of knowledge, rules and actions are maintained in both verbal and written form. There is often no attempt to validate knowledge as true or incorrect because there is no easy way to relate data to a process.

The knowledge life cycle recognizes the natural taxonomy of knowledge. Existing knowledge is maintained in the Knowledge Warehouse in a way that allows new knowledge claims to ultimately become integrated into the company's knowledge base. To bring this knowledge forward a view of the subject life cycle is created, documented for rules and claims and then validated with data.

First generation Data Warehousing and business intelligence does not address this vast amount of knowledge and does not have a formal process to manage the information.

The Impact of the Knowledge Process on Data Warehousing

Data that is collected, transformed, mined and reported in a first generation Data Warehouse is, by definition, historical, static and in today's market old news. The primary function of this data is to manage the business with after the fact information. One reason that Data Warehouse projects fail to live up to management expectations is the failure to understand the relationship of the data to the business process.

Data Warehouse projects often try to mix operational data in the form of an operational data store (ODS) with the Data Warehouse to gain access to real time information. This is a mixing of time variant data and creates volumes of discussions on how best to handle the situation. A different view might be that this is really a mixing of the knowledge process.

In the Intelligent Data Warehouse, the knowledge process is managed from within the Knowledge Warehouse using both historical and active data in near to real time. There may still be a need for an ODS depending on the process that is being managed. The IDW becomes a means of sending high performance information to the operational systems and keeping Data Warehouse data tightly coupled to the business process life cycle. While the IDW still maintains the traditional definition of the warehouse within its boundaries, current operational data becomes integrated and useable.

To expand the Data Warehouse into an Intelligent Data Warehouse the knowledge process components for process definitions, business rules, knowledge structures, and knowledge bases are added and managed by the Knowledge Warehouse software. Merging these components advances the

Data Warehouse from just data management to knowledge process management.

Second Generation Business Intelligence and Data Warehousing

Second generation business intelligence is the linkage of knowledge production and knowledge integration with the business process environment. With this linkage comes an understanding of the data and the procedures that are used to create business objectives. Higher levels of information can be generated to produce better methods, shorter time to market, higher product quality and a reduction in overhead costs.

Second generation business intelligence combines what the business knows about itself with what the business knows how to do and conforms that knowledge into the socialized structure thus creating an Intelligent Data Warehouse environment.

CONCLUSION

Developing an Intelligent Data Warehouse will require the effort to understand the business process as well as the business data. Linking historical data and employing data mining within operational systems can realize quantum gains in productivity and reduction of cost. The effort to map data with process is in no way small but as with all Data Warehousing efforts the projects can be targeted and expanded.

The IDW Knowledge Warehouse knowledge bases, methods and objects in themselves can generate revenue by being licensed as knowledge products within the same vertical market or even across markets. Effort given to create repeatable and reusable knowledge bases will produce company assets that may be more valuable than the products produced by those processes.

SUGI 27 Presentation

Presentation will provide an understanding of an IDW model, differences between rule based and object based approaches and how SAS can be used to bring Data Warehouses to the level of an Intelligent Data Warehouse. The presentation is intended for architects, business managers, and IT managers.

REFERENCES

Dymond. A.M. (2001), *Automated Consultant™ Reference and Design Manuals*, Dymond and Associates, Concord CA.

McElroy, M.W. (2001), "Second-Generation KM A White Paper," *Proceedings The 5th Annual KM World 2001 Conference and Exposition, Santa Clara, California*

Object Management Group, *Common Warehouse Metamodel (CWM) Specifications*, 250 First Avenue Needham, MA 02494 USA, <http://www.omg.org>

ACKNOWLEDGEMENTS

SAS and all other SAS Institute Inc. Product or service name are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® Indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

CONTACT INFORMATION

Larry D. Bramblett
Data Warehouse Solutions, LLC
2582 Shadow Mountain Drive
San Ramon, California 94583

(925) 831-1373
(925) 831-8890 (FAX)
dwsinfo@dnai.com

A SAS Institute Inc. Alliance Partner

Intelligent Data Warehouse

