



## Novel Architecture with Dimensional Approach of Data Warehouse

<sup>1</sup>Keshav Dev Gupta, <sup>2</sup>Jyoti Gupta, <sup>3</sup>Prakati Prasoon

<sup>1</sup>Research Scholar

<sup>2,3</sup>B.Tech Student

Department of Engineering & Technology

Jayoti Vidyapeeth Women's University

Jaipur, Rajasthan

India

---

**Abstract:** *Data warehousing is a phenomenon that grew from the huge amount of electronic data stored in recent years and from the urgent need to use that data to accomplish goals that go beyond the routine tasks linked to daily processing. Modeling data warehouses is a complex task focusing, very often, into internal structures and implementation issues. In this paper we discuss that, in order to accurately reflect the user's requirements into an out of error, easy to understand, and easily extendable data warehouse schema, special attention should be paid at the dimensional modeling phase. Based on a real mortgage business warehouse environment, we present a set of user modeling requirements and we discuss the involved concepts. Understanding the semantics of these concepts, allow us to build a conceptual model- namely, the starER model- for their efficient handling. According to William H. Inmon data warehouse is a subject-oriented, integrated, time- variant, nonvolatile collection of data in support of management decisions.*

*Data warehouses store current as well as historical data and are used for creating trending reports for senior management reporting such as annual and quarterly comparisons. Data Warehousing is a recent technology that allows information to be easily and efficiently accessed for decision-making activities by collecting data from many operational, legacy and possibly heterogeneous data sources.*

---

**Keywords:** *Data warehouse, dimension modeling, OLAP, ETL, decision-making*

---

### I. Introduction

Nowadays, data warehousing became an important strategy to integrate heterogeneous information sources in organizations, and to enable On-Line Analytic Processing (OLAP). Unfortunately, neither the accumulation, nor the storage process, seems to be completely credible. The concept of data warehousing originated in 1988 with the work of IBM researchers Barry Devlin and Paul Murphy. The need to warehouse data evolved as computer systems became more complex and handled increasing amounts of data. Warehoused data must be stored in a manner that is secure, reliable, easy to retrieve and easy to manage. On-Line Analytical Processing (OLAP) tools are well-suited for complex data analysis, such as multi-dimensional data analysis, and to assist in decision support activities while data mining tools take the process one step further and actively search the data for patterns and hidden knowledge in the data held in the warehouse [1]. A data warehouse is a relational or multidimensional database that is designed for query and analysis.

Data warehouse design has hitherto focused on the physical data organization (i.e., the "internal" structure) and quite understandable so, because of the volume and the complexity of data. Following the logical structure of data, as described in a data warehouse, several schemas have been developed emphasizing on the star-oriented approach; data unfolds around facts occurring in businesses [2]. The star, the star flake, and the snowflake schema are used widely for this purpose. Although all of these schemas provide some level of modeling abstraction that is understandable to the user, they are not built having his/her needs in mind. Data warehouses to bring multiple databases together and make them available for data mining and other forms of analysis. Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge Worker (executive, manager, and analyst) to make better and faster decisions.

A data warehouse is a database designed to support decision making in an organization. Data from the production databases are copied to the data warehouse so that queries can be performed without disturbing the performance or the stability of the production systems. Our position is that data warehouse modeling- as exactly databases do, many years now- should be exposed, to a higher level of design that is understandable to the user, independent of implementation issues, and that does not use any computer metaphors, such as "table" or "field". The result of this process will be a schema that is formal and complete, so that it can be transformed into the next logical schema without ambiguities Data warehouses offer organizations

the ability to gather and store enterprise information in a single conceptual enterprise repository. Basic data modeling techniques are applied to create relationship associations between individual data elements or data element groups. These associations, or “models,” often take the form of entity relationship diagrams (ERDs). More advanced techniques include the star schema and snowflake data model concepts. Regardless of the technique chosen, the goal is to build a metadata model that conceptually represents the information usage and relationships within the organization. They are not optimized for transaction processing, which is the domain of OLTP systems [13]. Data warehouses usually consolidate historical and transactional data derived from multiple sources. Data warehouses separate analysis workload from transaction workload and enable an organization to consolidate data from several sources. The data in a data warehouse is typically loaded through an extraction, transformation, and loading (ETL) process from one or more data sources such as OLTP applications, mainframe applications, or external data providers. The architecture of the enterprise data warehouse is designed to deliver the analysis capabilities defined in the business requirements document just referenced and likewise to provide the critical success elements defined there. In addition, the warehouse is expected to deliver any additional analysis capabilities delivered by existing campus decision-support systems which were not explicitly documented in the business requirements document.

This is the conceptual or semantic modeling phase, and the benefits of its use have been praised a lot: communication between the designer and the user, early detection of modeling errors, and easily extendable schemas are among them. The conceptual modeling phase is part of a design methodology- which is classical in the database area, and has been already proposed for the data warehouse area- following the user requirements analysis and specifications phase and, is followed by the logical design focusing on workload refinement and schema validation.

The ex ante treatment of the metadata repository is enabled by a full set of steps, i.e., quality question, which constitute our methodology for data warehouse quality management and the quality-oriented evolution of a data warehouse based on the architecture, process and quality meta models. Our approach extends GQM, based on the idea that a goal is operationally defined over a set of questions. Thus, we provide specific “questions” for the full lifecycle of a goal: this way the data warehouse metadata repository is not simply define statically, but it can be actually exploited in a systematic manner.

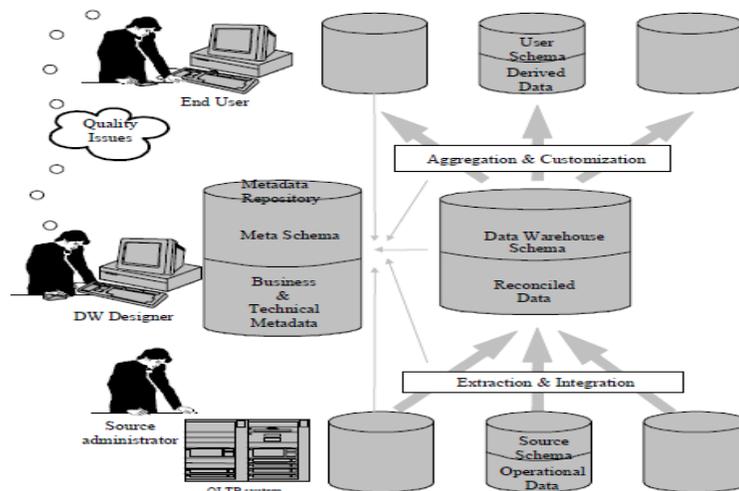
## II. Novel Architecture of Data warehouse:

The data warehouse architecture is primarily based on the business processes of a business enterprise taking into consideration the data consolidation across the business enterprise with adequate security, data modeling and organization, extent of query requirements, meta data management and application, warehouse staging area planning for optimum bandwidth utilization and full technology implementation [2].

We made an interesting point above, by stating that the efficient management of the lifecycle of a data warehouse is enabled by the inspection of three viewpoints concerning the data warehouse: its architecture, processes and substances. The architecture constitutes of the static components comprising the data warehouse; the processes capture the dynamic part of the data warehouse environment. Finally, quality is a measure of the fulfillment of the expectations of the involved stakeholders in such an environment[3].

For each viewpoint, a respective meta model can be derived. It is important, of course, that all three meta models respect a coherent framework, and fit gracefully with each other. We will immediately proceed to present the architecture meta model for a data warehouse, as well as the general framework for metadata management. In the next section we will make a proposal for a quality meta model and in the fourth section we will present templates for quality management in a data warehouse environment.

The data warehouse processes -which we believe to be the more interesting part of the whole environment



**Figure 1.1. A generic architecture for a data warehouse**

Although many data warehouses have already been built, there is no common methodology, which supports database system administrators in designing and evolving a data warehouse. The problem with architecture models for data warehouses is that practice has preceded research in this area and continues to do so. Consequently, the task of providing an abstract model of the architecture becomes more difficult. Formally, an architecture model corresponds to the schema structure of the meta-database that controls the usually distributed and heterogeneous set of data warehouse components and therefore is the essential starting point for design and operational optimization. The purpose of architecture models is to provide an expressive, semantically defined and computationally understood meta modeling language, based on observing existing approaches in practice and research. Expressiveness and services of the metadata schema are crucial for data warehouse quality [9].

In this Section we focus on the special modeling needs of a data warehouse at the conceptual phase, as they are drawn from theoretical and practical experience for a mortgage company. The listed user requirements reveal a set of new modeling concepts that need to be handled. Based on these, later on, well-known models and schema are to be combined and extended with new constructs, improving their ability to conveniently design data warehouse environments.

In this paper we firstly address the modeling requirements of a data warehouse, from the user point of view. For this purpose, we use a real mortgage business environment. The understanding of the requirements reveals a set of concepts that need to be accommodated at the conceptual modeling phase. We propose the use of a new model, namely the starER model, for the conceptual modeling of data warehouses. The starER model combines the semantically rich constructs of the well-known Entity- Relationship (ER) model with the star structure that rules the data in data warehouses. Our starting point is that, the ER model has been tested for years, and proved powerful enough to model complex applications such as spatiotemporal, and multimedia. In all cases, when new modeling techniques are needed to capture the new demands, new constructs are added to ER; but the core of the model is the same. So, there is no reason to change such a model, which designers are familiar and happy with. On the other hand, as mentioned before, data warehouses impose a new modeling structure [5].

The architecture of the enterprise data warehouse is designed to deliver the analysis capabilities defined in the business requirements document just referenced and likewise to provide the critical success elements defined there. In addition, the warehouse is expected to deliver any additional analysis capabilities delivered by existing campus decision- support systems which were not explicitly documented in the business requirements document. These systems include the following:

- BAIRS
- BIS
- Cal Profiles
- The pilot student data warehouse
- FASDI
- The Office of Student Research database/ reporting file.

Describes the number of stages and how data is processed to convert raw / transactional data into information for end user usage.

The data staging process includes three main areas of concerns or sub- processes for planning data warehouse architecture namely “Extract”, “Transform” and “Load” These interrelated sub-processes are sometimes referred to as an “ETL” process.

- 1) Extract- Since data for the data warehouse can come from different sources and may be of different types, the plan to extract the data along with appropriate compression and encryption techniques is an important requirement for consideration.
- 2) Transform- Transformation of data with appropriate conversion, aggregation and cleaning besides de-normalization and surrogate key management is also an important process to be planned for building a data warehouse. [12]
- 3) Load- Steps to be considered to load data with optimization by considering the multiple areas where the data is targeted to be loaded and retrieved is also an important part of the data warehouse architecture plan.

### **III. Dimensional Model of Data warehouse**

A Dimensional Model is a database structure that is optimized for online queries and Data Warehousing tools. It is comprised of "fact" and "dimension" tables. Dimensional Models are designed for reading, summarizing and analyzing numeric information, whereas Relational Models are optimized for adding and maintaining data using real-time operational systems [24]. Dimensional models use facts and dimensions to describe data for the business. In dimensional modeling, the best unit of analysis is the business process in which the organization has the most interest. Dimensional modeling is very flexible for the user perspective. Dimensional data model is mapped for creating schemas. Whereas ER Model is not mapped for creating schemas and does not use in conversion of normalization of data into demoralized form [6].

Dimensional Models A demoralized relational model Made up of tables with attributes Relationships defined by keys and foreign keys Organized for understandability and ease of reporting rather than update Queried and maintained by SQL or special purpose management tools [4]. A dimensional model includes fact tables and lookup tables. While it is

universally recognized that a DW leans on a multidimensional model, there is no agreement on the approach to conceptual modeling. On the other hand, an accurate conceptual design is the necessary foundation for building a “good” information system [10].

\_ The Entity/Relationship model is widespread in the enterprises, but...

"Entity relation data models [...] cannot be understood by users and they cannot be navigated usefully by DBMS software. Entity relation models cannot be used as the basis for enterprise data warehouses." (Kimball, 96)[5, 6].

#### IV. Conclusion

Data warehouses separate analysis workload from transaction workload and enable an organization to consolidate data from several sources. The data in a data warehouse is typically loaded through an extraction, transformation, and loading (ETL) process from one or more data sources such as OLTP applications, mainframe applications, or external data provider. Dimensional modeling is a logical design technique for structuring data so that it's intuitive to business users and delivers fast query performance. Data presented to the business intelligence tools must be grounded in simplicity to stand any chance of success. Simplicity is a fundamental requirement because it ensures that users can easily understand databases, as well as allows software to efficiently navigate databases.

#### References :

- [1]. Anahory, S., and Murray, D., 1997. Data Warehousing in the Real World. Addison-Wesley.
- [2]. Busborg, F., Christiansen, J.B., Jensen, K.M., and Jensen, L., 1998a. A Method fo Data Warehouse Development. Dat5
- [3]. Report, part 2. CS Department. Aalborg University.
- [4]. Busborg, F., Christiansen, J.B., Jensen, K.M., and Jensen, L., 1998b. Data Warehouse Modeling: The Nykredit Case Study. Dat5 Report/Part I. Computer Science Department. Aalborg University.
- [5]. Chen, P.S., 1976. The Entity-Relationship Model: Toward a unified view of Data. ACM TODS, 1(1):9-36.
- [6]. Goffarelli, M., Maio, D., and Rizzi, S., 1998. Conceptual Design of Data Warehouses from E/R Schemas. Proceedings of the 13th Hawaii International Conference on SystemSciences. Kona, Hawaii.
- [7]. AdCom Services. Data Warehouse - Project Proposal.
- [8]. <http://pulsar.adcom.uci.edu:80/dwh/dwpro4/prohome.html>.
- [9]. N. Agmon and N.Ahituv. Assesing data reliability in an information system. J. Management Information Systems, vol. 4, no. 2, 1987.
- [10]. R. Agrawal, A. Gupta, and S. Sarawagi. Modelling multidimensional databases.
- [11]. Technical report, IBM Almaden Research Center, San Jose, California,
- [12]. [Akao, Y., ed. Quality Function Deployment. Productivity Press, Cambridge MA. .]
- [13]. [P. Andritsos, A. Papagianni. Development of an OLAP tool. Technical Report KDBSLDIPL-
- [14]. Knowledge and Database Systems Laboratory, NTUA.
- [15]. Arbor Software Corporation. Arbor Essbase.<http://www.arborsoft.com/essbase.html>
- [16]. American Supplier Institute, Inc.. Taguchi Methods: Introduction to Quality Engineering. Allen Park, Mich.
- [17]. V. R. Basili, G.Caldiera, H. D. Rombach. The Goal Question Metric Approach.
- [18]. Inc. Available at <http://www.cs.umd.edu/users/basili/papers.html>
- [19]. [BaPa82] D.P. Ballou and H.L. Pazer. The impact of inspector failability on the inspection policy
- [20]. Serial production system. Management Science, vol. 28, no.4..
- [21]. [D.P. Ballou and H.L. Pazer. Modelling data and process quality multi-input, multi output
- [22]. Information systems. Management Science, vol. 31, no.
- [23]. D.P. Ballou and H.L. Pazer. Cost/quality tradeoffs for control procedures information systems. OMEGA: Internatinal J. Management Science,
- [24]. E. Baralis, S. Paraboschi and E. Teniente. Materialized View Selection in a Multidimensional Database. Proceedings of the 23rd International Conference on Very Large Databases (VLDB)

#### AUTHOR BIOGRAPHY



**First Author:** Keshav Dev Gupta, M.Sc., M.C.A., M. Tech., Ph .D (pursuing), in the department of Engineering and Technology, Assistant Professor of computer science and engineering, Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India. He has over Six years of experience in teaching computer science and engineering to graduate and post graduate students. He is member of IAE, IAE-Societies, IAS, CSI, and IEEE. He has attended four national and international conferences of computer science. He has published five international papers in journal. He has published two books viz. OOPS with Java, and Java Programming with HTML.



**Second Author:** Jyoti Gupta, B.Tech (pursuing), in the department of Engineering and Technology, Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India. She has attended four national and international conferences of computer science. She has published many international papers in journal.



**Third Author:** Prakarti Prason, B.Tech (pursuing), in the department of Engineering and Technology, Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India. She has attended four national and international conferences of computer science. She has published many international papers in journal.