International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 1 Issue 4, June - 2012

Elevate and Forward Motion Approach-Path in Construction of Data Warehouse

Param Deep Singh
M.Tech.
S.A.T.I. Vidisha(m.p.)

Jitendra Raghuvanshi
M.Tech
B.U.I.T. , Bhopal

Divakar singh
H.O.D
B.U.I.T.Bhopal

Abstract:
Data warehouse is a process for building decision support systems and knowledge management environment that supports both day-to-day tactical decision making and long-term business strategies. It is not about the tools; rather, it is about creating a strategy to plan, design, and construct a data store capable of answering business questions. Good strategy is a process that is never really finished; a defined data warehouse development process provides a foundation for reliability and reduction of risk. This process is defined through methodology. The data warehouse development enjoys high visibility; many firms have concentrated on reducing these costs. Standardization and reuse of the development artifacts and the deliverables of the process can reduce the time and cost of the data warehouse's creation. To understand how a warehouse can benefit you and what is required to manage a warehouse, you must first understand how a data warehouse is constructed and established.

Data warehouse 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.

Keywords:
Data Warehouse (DW), Data Warehouse Architecture (DWA), Data Marts, and ETL.

1. Introduction:
Data Warehouse (DW) is a database used for reporting and analysis. The data stored in the warehouse is uploaded from the operational systems. The data may pass through an operational data store for additional operations before it is used in the DW for reporting.
A data warehouse is a collection of consistent, subject-oriented, integrated, time-variant, non-volatile data and processes on them, which are based on available information and enable people to make decisions and predictions about the future. So, defined a data warehouse as follows:

- **Subject-oriented**, meaning that the data in the database is organized so that all the data elements relating to the same real-world event or object are linked together;
- **Time-variant**, meaning that the changes to the data in the database are tracked and recorded so that reports can be produced showing changes over time;
- **Non-volatile**, meaning that data in the database is never over-written or deleted, but retained for future reporting; and
- **Integrated**, meaning that the database contains data from most or all of an organization's operational

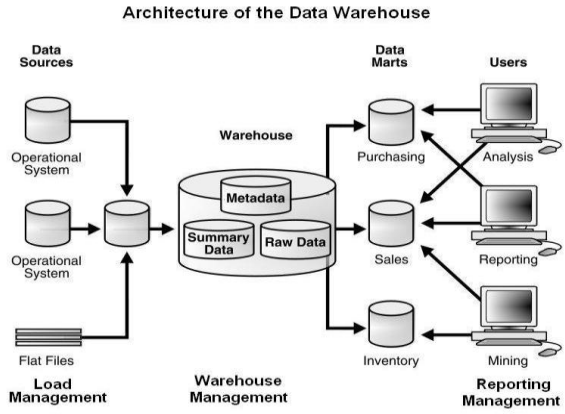
applications, and that this data is made consistent;

"A data warehouse is a repository of integrated information, available for queries and analysis. Data and information are extracted from heterogeneous sources as they are generated; this makes it much easier and more efficient to run queries over data that originally came from different sources." This definition of the data warehouse focuses on data storage. The main source of the data is cleaned, transformed, catalogued and made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. A data warehouse is the main repository of the organization's historical data, its corporate memory. In other words, the data warehouse contains the raw material for management's decision support system. Frequently data in Data Warehouses is heavily de-normalized, summarized and/or stored in a dimension-based model but this is not always required to achieve acceptable query response times. Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata.

2. Data Warehouse and Its Architecture:

Data warehouses are computer based information systems that are home for "secondhand" data that originated from either another application or from an external system or source. Warehouses optimize database query and reporting tools because of their ability to analyze data, often from disparate databases and in interesting ways. In other words, data warehouses are read-only, integrated databases designed to answer comparative and "what if" questions.

Think of a data warehouse as a central storage facility which collects information from many sources, manages it for efficient storage and retrieval, and delivers it to many audiences, usually to meet decision support and business intelligence requirements. The data warehouse enables users to access vast stores of integrated, operational data to track business trends, facilitate forecasting and planning efforts, and make strategic decisions. It is not a single product, but rather a flexible environment comprised of multiple technologies.



A Data Warehouse Architecture (DWA) is a way of representing the overall structure of data, communication, processing and presentation that exists for end-user computing within the enterprise. Data warehouse architecture involves the following components;

2.1 Load management includes all of the software and utilities required to:

- Extract source system data and move it to the warehouse environment;
- Complete basic transformation to ensure that non-essential data is eliminated and other data is converted to appropriate data types;

- Fast load data into a staging area where it can be subsequently manipulated;
- Extract data from staging area and load clearing house tables;
- Extract data from clearing house tables and load the data warehouse;
- Extract data from the data warehouse and load data marts.

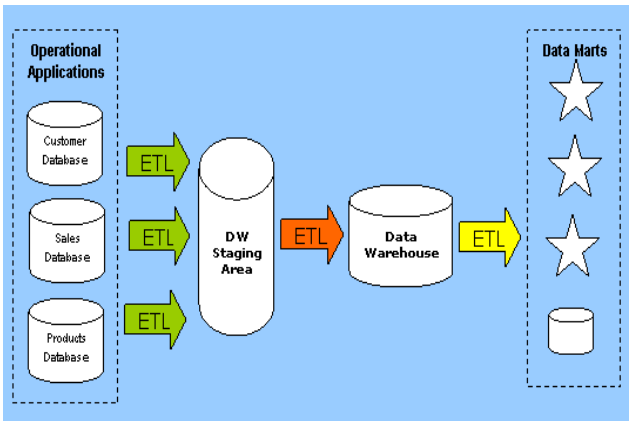
2.2 Warehouse management involves all of the software and system management utilities required to:

- Clean and transform data;
- Create temporary holding tables to accommodate merging data for analysis or cleansing purposes;
- Create and/or maintain indexes, views, and table partitions;
- Aggregate data as necessary;
- De-normalize data if needed for performance purposes;
- Archive data in each of the data warehouse environments; and
- Complete incremental or full back-ups as needed.

2.3 Reporting management involves the software required to ensure that business intelligence reporting tools direct queries to the data that will provide the quickest query response. Reporting business intelligence software such as Business Objects or Micro strategy generally handles reporting management with support from custom routines developed using relational database management software.

3. Construction of Data Warehouse:

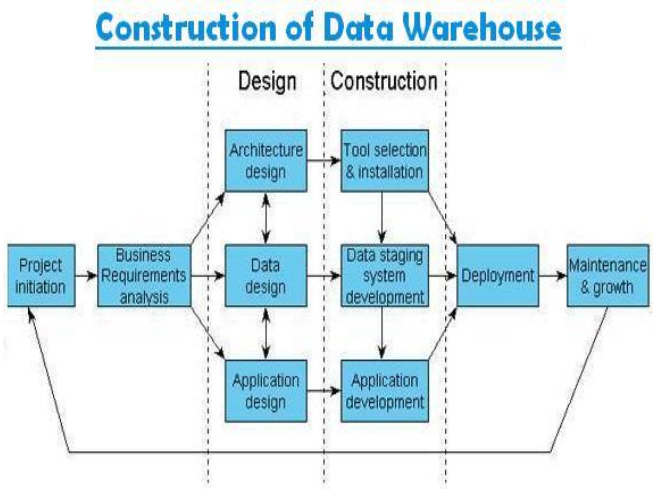
The Data warehouse into a set of conformed Data Marts that, are accessible by decision makers and the typical data warehousing environment shown as:



The construction of data warehouse can be summarized and provides an opportunity to:

- understand new concepts and processes, and identify potential problems;
- make more realistic plans and manage expectations;
- evaluate alternative tools;
- demonstrate benefits and gain management commitment.

So, the construction of data warehouse shown as:



3.1 Project initiation

No data warehousing project should commence without:

- a clear statement of business objectives and scope;
- a sound business case, including measurable benefits;

- an outline project plan, including estimated costs, timescales and resource requirements;
- high level executive backing, including a commitment to provide the necessary resources;

A small team is usually set up to prepare and present a suitable project initiation document. This is normally a joint effort between business and IT managers.

3.2 Requirements analysis

Establishing a broad view of the business' requirements should always be the first step. The understanding gained will guide everything that follows, and the details can be filled in for each phase in turn. Collecting requirements typically involves 4 principal activities:

- Interviewing a number of potential users to find out what they do, the information they need and how they analyze it in order to make decisions.
- Interviewing information systems specialists to find out what data are available in potential source systems, and how they are organized.
- Analyzing the requirements to establish those that are feasible given available data.
- Running facilitated workshops that bring representative users and IT staff together to build consensus about what is needed, what is feasible and where to start.

3.3 Design

The goal of the design process is to define the warehouse components that will need to be built. The architecture, data and application designs are all inter-related, and are normally produced in parallel.

3.3.1 Architecture design

The warehouse architecture describes all the hardware and software components that form the data warehousing environment and explains:

- how components will work together;
- where they are located;
- who uses them;
- who will build and maintain them.

The architecture provides a framework for the selection of tools and the detailed design of individual components during the first and subsequent phases of development.

3.3.2 Data design

This step determines the structure of the primary data stores used in the warehouse environment, based on the outcome of the requirements analysis. It is best to produce a broad outline quickly, and then break the detailed design into phases, each of which usually progresses from logical to physical; Once the logical design is established, the next step is to define the *physical* characteristics of individual data stores and any associated indexes required to optimize performance.

3.3.3 Application design

The application design describes the reports and an analysis required by a particular group of users, and usually specifies:

- a number of template report layouts;
- how and when these reports will be delivered to users;
- functional requirements for the user interface.

3.4 Construction

Warehouse components are usually developed iteratively and in parallel. That said, the most efficient sequence to begin construction is probably as follows:

3.4.1 Tool selection & installation

Selecting tools is best carried out as part of a pilot exercise, using a sample of real data. This allows the development team to assess how well competing tools handle problems specific to their organization and to test system performance before committing to purchase. The most important choices are:

- ETL tool
- Database for the warehouse usually relational and marts

- Reporting and analysis tools

It pays to define standards and configure the development, testing and production environments as soon as tools are installed, rather than waiting until development is well underway.

3.4.2 Data staging system

This comprises the physical warehouse database, data feeds and any associated data marts and aggregates. The following steps are typical:

- Create target tables in the central warehouse database;
- Request initial and regular extracts from source systems;
- Write procedures to transform extract data ready for loading;
- Create and populate any data marts;
- Write procedure to load regular updates into the warehouse;
- Write validation/exception handling procedures;
- Write archiving/backup procedures;
- Document the whole process.

3.4.3 Application development

This step can begin once a sample or initial extract has been loaded, but it is usually best to leave the bulk of application development until the underlying data mart and associated meta-data are stable. It is a good idea to involve users in the development of reports and analytic applications, preferably through prototyping, but at least by asking them to carry out acceptance testing.

3.5 Deployment

It is too often assumed that the first version of a data warehouse can be rolled out in a matter of weeks, simply by showing all the users how to use the new reporting tools. As well as training, planning for deployment needs to cover:

- Installing and configuring desktop PCs - any hardware upgrades or amendments to the 'standard build' need to be organized well in advance;
- Implementing security measures - to control access to applications/data;
- Providing more advanced tool training later, when users are ready, and assisting potential power users to develop their first few reports.

If the first users find errors and inconsistencies in the data, don't feel comfortable with the tool; then time spent building the warehouse may ultimately be wasted. The following guidelines will help to reduce these risks:

- Do not start deployment until the data are ready;
- Use a small, representative group to try out the finished system before rolling out;
- Do not grant system access to users until they have been trained.

3.6 Maintenance

A data warehouse is not like an OLTP system: development is never finished, but follows an iterative cycle (analyze – build – deploy). The most important activities are:

- Monitoring the realization of expected benefits;
- Providing ongoing support to users;
- Assisting with the identification and cleansing of dirty data;
- Maintaining both feeds & meta-data as source systems change over time;
- Tuning the warehouse for maximum performance;
- Recording successes and using these to continuously market the warehouse.

4. Benefits of Data Warehouse:

A data warehouse maintains a copy of information from the source transaction systems. This architectural complexity provides the opportunity to:

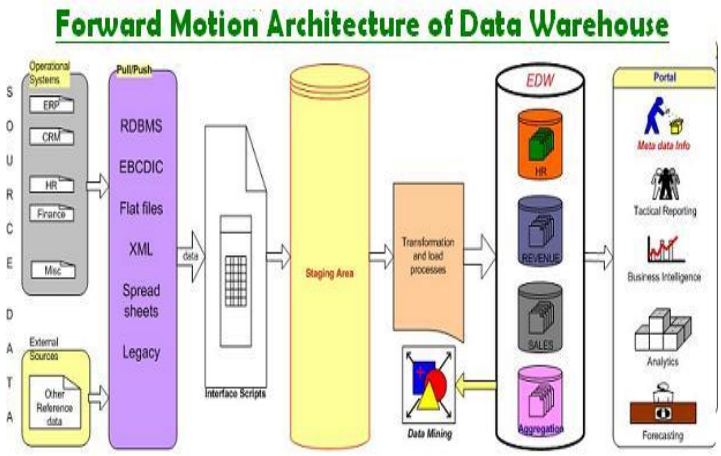
- Maintain data history, even if the source transaction systems do not.

- Integrate data from multiple source systems, enabling a central view across the enterprise.
- Improve data, by providing consistent codes and descriptions, flagging or even fixing bad data.
- Present the organization's information consistently.
- Provide a single common data model for all data of interest regardless of the data's source.
- Restructure the data so that it makes sense to the business users.
- Restructure the data so that it delivers excellent query performance, even for complex analytic queries, without impacting the operational systems.
- Add value to operational business applications, customer relationship management systems.

5. Conclusion:

Over the last few years, data warehouses enjoy a lot of attention both from the industrial and the research community. The reason lies in their great importance: making predictions about the future, has always been desirable for business companies. The growth of data warehousing is going to be enormous with new products and technologies coming out frequently. In order to get the most out of this period, it is going to be important that data warehouse planners and developers have a clear idea of what they are looking for and then choose strategies and methods that will provide them with performance today and flexibility for tomorrow.

In this paper, we have elevate and forward motion approach-path in construction of data warehouse; which incorporates new guidelines to address specific warehouse needs and which can naturally be embedded into the traditional database and data warehouse design process. Finally after the above explanation symbolize elevate and forward motion architecture of data warehouse as:



6. References:

[1]M.Golfarelli,D.Maio,S.Rizzi,"Conceptual design of data warehouses from E/R.

[2]D. Theodoratos, T. Sellis (DWQ project). Designing Data Warehouses. DKE '99.

[3]C. Adamson, M. Venerable. Data Warehouse Design Solutions. J. Wiley & Sons, Inc. 1998.

[4]A. Bauer, H. G¨unzel, Data Warehouse Systeme Architektur, Entwicklung, 2001.

[5]http://www.1keydata.com/datawarehousing/data-warehouse-architecture.html.

[6]J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, SanFrancisco, 2000.

[7]R. Kimball. The Data Warehouse Toolkit. J. Wiley & Sons, Inc. 1996.

[8]A. Marotta. A transformations based approach for designing Data Warehouses Internal Report. InCo. Universidad de la República, Montevideo, Uruguay. 1999.

[9]W. J. Labio, Y. Zhuge, J. N. Wiener, H. Gupta, H. Garcia-Molina, Prototype for Data Warehouse Creation & Maintenance, 1997.

[10] Immon, W. H., 1996. Building the Data Warehouse. Wiley Computer Publishing (2nd Edition).

[11]Anahory, S., and Murray, D., 1997. Data Warehousing in the Real World. Addison-Wesley.