

Data Source Layer

This shows different data sources that feed data into the data warehouse. The data source can be any format plain text file. Relational database other type of databases, Excel file, SQL data base, access MySQL, PostgreSQL data base. All can act as a data source. Many different types of data can be data source

- Operations: such as HR data, Sales data production data, Inventory data, Marketing data, System data
- Web server logs with user browsing data
- Internal marketing research data
- Third party data are such as demographics data or survey data
- All these data sources together form the data source layer.

Data Extraction Layer

Data is pulled from data source into D.W. system. There is likely some minimal data cleansing, but it is may a case for much more transformation on the data.

Staging Area

This is where data can sit prior to being scrubbed & transformed into Data Warehouse /Data Mart rather than all data source as a fetching area or source of data having common data area makes it easier for subsequent data pressing/ integration.

ETL Layer

This stage, where data gets its “intelligence” where there is process of transformation is applied so, transactional data is transferred to analytical data to make decision for getting exact application as a repair to create different decision making structures. This layer is also where data cleansing happens.

Data Storage Layer

This is where the transformed & cleansed data sits. Based on scope of functionality, 3 types or entity can be found here. (1) Data Warehouse (2) Data Mart (3) Operational data store (ODS)
In any given system, actually system contain any of three two of three or three of three as an entity for further functionality.

Data Logic Layer

There is the area of D.W. where business logic is stored. These business rules does not affect any transformation process of the data but provide a structure for getting a report as a decision making process.

Data Presentation Layer

This restores to the information that reaches to the users. This can be a form of tabular, graphical report in a browser an emailed report that gets automatically generated and sent everywhere as per the given domain of users or an alert that warns users of exceptions among others.

Metadata Layers

This is where information about the data stored in the Data Warehouse system is stored. A logical data model would be an example of something that is in metadata layer.

System Operations Layers

This layer includes information on how the data warehouse system operates, such as ETL job status, system performance and user access history.

How OLTP and Data Warehousing Is Contrasting

OLTP	FEATURE	DATA WAREHOUSE
Complex Data Structure (3NF Databases)		Multidimensional Data Structure
Few	Indexes	Many
Many	Joins	Some
Normalized DBMS	Duplicated Data	Demoralized DBMS
Complex long running queries by system	To update the data	Normalized small database. Speed will be not an issue due to smaller database and normalization will not degrade performance.
Snapshot of business processes	User perspective View	Multi-dimensional views of business activities of planning and decision making Queries and Process Scripts OLTP
Supports thousands of concurrent users.	Supporting Number of Users	Supports few concurrent users relative to OLTP
Optimized for a common set of transactions, usually adding or retrieving a single row at a time per table.	Transaction process	Optimized for bulk loads and large, complex, unpredictable queries that access many rows per table.
Optimized for validation of incoming data during transactions; uses validation data tables.	Data updating	Loaded with consistent, valid data; requires no real time validation

Table.2 Key difference between an OLTP sys & data warehouse

One major difference between the types of system is that data warehouse are not usually in third normal form (3 NF) a type of data normalization common in OLTP Environments. Data warehouses and OLTP system have very different requirements.

Some Examples of The Difference Between Typical Data Warehouses And OLTP System Workload

Data Warehouse is designed to accommodate ad hoc queries. The workload of Data Warehouse is in to advance, so a Data Warehouse should be optimized to perform well for a wide variety of possible query operations.

OTPL system support only predefined operations. Applications might be specifically tuned or design to support only those Operations.

Data Modification

A data warehouse is updated on a regular basis by the ETL process using bulk data modification techniques. The end user of data warehouse does not directly update the Data Warehouse. While in OLTP systems, end users

routinely issue individual data modification statements to Data base. The OLTP database is always up to date and reflects current state of each business transactions.

The Factors That Are Affect the Selection of Architecture

No two organizations are the same and consequently companies may differ on their architecture selection decision procedure, there is no possibility to provide single architecture for all of the companies which are best in any situation. From experts, some potential factors are identified to select particular architecture. Some of the factor related to the relational theory, such as the information processing theory of the firm, while others are retained on social, political theories such as power and politics.

Organizations Units' Information Integration

The high level of information interdependence is, when the work of One Organizational unit depends on different organizational unit. In this situation the ability to share consistent, integrated information is important. It is understood that firms with high information. Interdependences select an enterprise wise architecture.

Upper Management Is Information Needs

For carrying out job responsibilities, higher management often requires information from lower organizational levels. It may need to monitor progress on meeting company goals, drill down into area of interest, aggregate layer/level data and be confident that company is a compliance with regulation.

Urgency of Need for A Data Warehouse

An organization can be Data Warehouse or Data Mart and so business rules needed to implement Data Warehouse fast. Some architecture is more quickly implemented than others. So as per the requirement Data Warehouse architecture is selected.

Nature of End User Task

Some User perform non-routine task queries with specific structures reports are not sufficient in manner to provide their needs. They have to analyse their data as per the end users requirement. These users require an architecture that provides enterprise wide data that can be analysed “on the fly” in creative way.

Constraints on Resources

Some Data Warehouse require more resources to implement it than others, resources as it personal, business unit personal and monetary resource can impact the selection of the architecture.

View of Data Warehouse Prior To Implementation

Organization differ in their view or plans for Data Warehouse or Data marts. Some may require building Data Warehouse as a part of their strategic plans while other organizations may not. As a result it may be developed to provide Point solution” to a particular business units need. It may be project which is supporting decision. Support infrastructure to support a range of applications, it may be critical enabler to support a company’s strategy business objective. So as per the required view, Data Warehouse implementation is done.

Expert Influence

When building a data warehouses. There are many places turn for help – consulting, the literature, conferences and seminars internal experts and end users. These varying degrees can influences the

architecture that is selected for example an implemented last successful architecture, pointed out by consultant, which is made successful completion and satisfactory reports from prior Data Warehouse.

Compatibility with Existing System

Many benefits are included to the existing system it is work as a foundation of building of Data Warehouse & new other steps or new other changes are done in already established foundation. This may include compatibility with source systems, metadata integration, data access tools and technology vendors.

The Perceived Ability of the In House IT Staff

The building of Data Warehouse can also affect the factors like. IT staff's technical skill, successful experience with similar projects, and level of confidence. All these factor make system of making Data Warehouse is improvement.

Source on Sponsorship

The source of sponsorship for a Data Warehouse may vary from a single department or business unit to the top management within an organization. Influence from the sponsors may control many aspect of Data Warehouse act initiative such as monetary resources and the architecture selected.

Technical Issues

A variety of technical consideration can affect the choice of architecture scalability in terms of no of users' volume of data, query performance, scalability in terms of the number technical changes etc.

Conclusion

The research model that relates the factors to the architectures is shown as below:

Architecture Selection Factors

- Information Interdependence
- Upper management's information needs
- Urgency of need
- Nature of end user task
- Resource availability
- View of the data warehouse
- Export influence
- Compatibility with exciting system
- Received ability of the in house IT staff
- Source of sponsorship
- Technical issue

Data Warehouse Architectures

- Independent data marts
- Data mart bus architecture
- Hub
- Centralized data Warehouse
- Federated

References

- [1] Nong Ye, The Handbook of Data Mining (Lawrence Erlbaum Associates, Mahwah, NJ. Publication, 2003).
- [2] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques (Morgan Kaufmann Publishers, University of Illinois at Urbana-Champaign).
- [3] Bharat Bhushan Agarwal and Sumit Prakash Taval, Data Mining and Data Warehousing (Laxmi Publications, New Delhi - 110002, India).
- [4] Ralph Kimball, Joe Caserta, Data Warehouse. ETL Toolkit. Practical Techniques for. Extracting, Cleaning, Conforming, and. Delivering Data (Wiley Publication, Inc).

Author Details:

Sweety Patel
Department of Computer Science, Fairleigh Dickinson University, USA