

## Data Warehousing Methods and its Applications

<sup>1</sup>Dr. C. Suba

<sup>1</sup>(Department of Computer Applications, Sri Lakshmi College of Arts & Science, Bangaram, India)

---

**Abstract:** A Data Warehouse is a process of Extracting data from source system or mining knowledge from large amount of data. Data warehouse commonly called ETL, which abbreviation for Extraction, Transformation and Loading. Commonly ETL refers to process. This paper discuss about Overview about Data Warehouse, Different Design approach of Data Warehouse and its advantages &disadvantages,Differetnt types of Extraction methods in Data Warehouse, Characteristics of Data Warehouse, Different Benefits of Data Warehousing, Different Component used in Data Warehousing and its applications of Data Warehousing.

**Keywords:** Characteristics and Benefits of Data Warehouse, Data Warehousing, Design Approach of Data Warehousing, Different Extraction Methods of Data Warehousing, Main Components of Data Warehouse and Data Warehousing Applications.

---

### I. Introduction

Data warehouse is a useful tool, gives benefit from the ability to store and analyze data, and this can allow in making sound business decisions. It is also important to make sure that the correct information is published, and it should be easy to access by the people who are responsible for making decisions[5].

A Data warehousing is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process. Data Warehouse is new technologies include: Agile BI development products, Data Warehouse appliances, Big data analytics, In-memory data, I workspaces and dashboards, Collaborative sharing of BI content. Mash-ups, Complex event processing and Mobile BI and data federation.

Hardware and software that support the efficient consolidation of data from multiple sources in a Data Warehouse for Reporting and Analytics include

- ETL (Extract, Transform, Load)
- EAI (Enterprise Application Integration)
- CDC (Change Data Capture)
- Data Replication
- Data Reduplication
- Compression
- Big Data technologies such as Hardtop and Map Reduce, and Data Warehouse Appliances



Figure 1: The New Kind of Warehousing

### II. Data Warehouse Design Approach

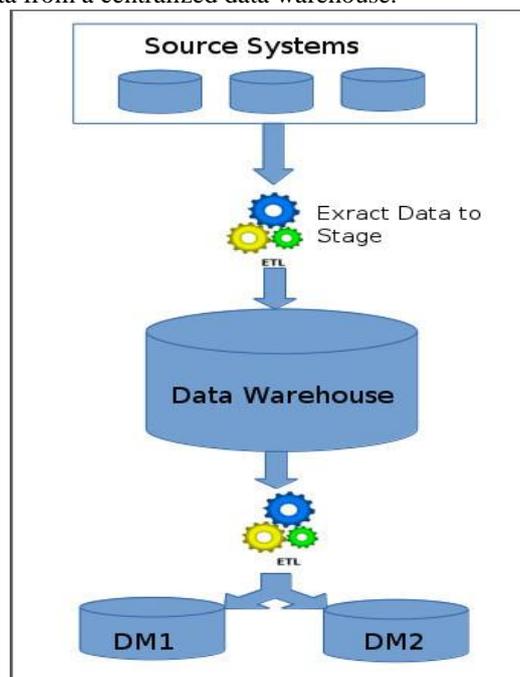
Data Warehouse design approaches are very important aspect of building data warehouse. Selection of right data warehouse design could save lot of time and project cost. There are two different Data Warehouse Design Approaches normally followed when designing a Data Warehouse and based on the requirements. Four major approaches to building a data warehousing environment exist. These architectures are generally referred to as

- a. Top-Down Approach
- b. Bottom-up Approach
- c. Hybrid Approach, and

## d. Federated Approach

**a. Top-Down Approach**

“Bill Inmon” is sometimes also referred to as the “father of data warehousing”; his design methodology is based on a top-down approach. It is one part of the overall business intelligence system. If you use a top-down approach, you will have to analyze global business needs, plan how to develop a data warehouse, design it, and implement it as a whole. The data warehouse holds atomic or transaction data that is extracted from one or more source systems and integrated within a normalized, enterprise data model. From there, the data is summarized, dimensionalized, and distributed to one or more “dependent” data marts. These data marts are “dependent” because they derive all their data from a centralized data warehouse.



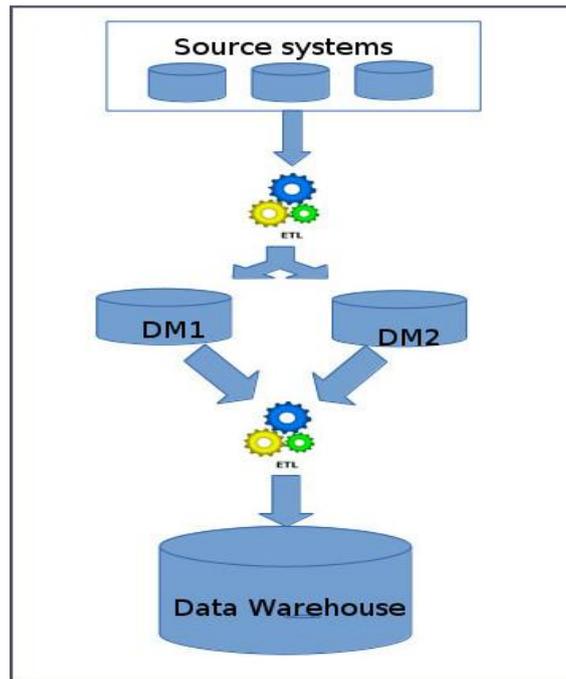
**Figure 2:** Top-down Data Warehouse Design Approach

**Pros and Cons of top-down Approach**

- It provides an integrated, flexible architecture to support downstream analytic data structures. First, this means the data warehouse provides a departure point for all data marts, enforcing consistency and standardization so that organizations can achieve a single version of the truth.
- Second, the atomic data in the warehouse lets organizations re-purpose that data in any number of ways to meet new and unexpected business needs. For example, a data warehouse can be used to create rich data sets for statisticians deliver operational reports, or support operational data stores (ODS) and analytic applications.
- Major problem of top-down approach may take longer and cost more to deploy than other approaches, especially in the initial increments.

**b. Bottom-Up Approach**

Ralph Kimball is a renowned author on the subject of data warehousing. His data warehouse design approach is called dimensional modeling or the Kimball methodology. This methodology follows the bottom-up approach. In a bottom-up approach, the goal is to deliver business value by deploying dimensional data marts as quickly as possible. Unlike the top-down approach, these data marts contain all the data both atomic and summary. The integration of data marts is implemented using data warehouse bus architecture. In the bus architecture, a dimension is shared between facts in two or more data marts. These dimensions are called conformed dimensions. These conformed dimensions are integrated from data marts and then data warehouse is built.



**Figure 3:** Bottom-up Data Warehouse Design Approach

#### **Pros and Cons of Bottom-up Approach**

- Major benefit of a bottom-up approach is that it focuses on creating user-friendly, flexible data structures using dimensional, star schema models. It also delivers value up front, rapidly because it doesn't lay down a heavy infrastructure.
- The data marts contain both summary and atomic data, users do not have to "drill through" from a data mart to another structure to obtain detailed or transaction data.
- This model contains consistent data marts and these data marts can be delivered quickly.
- One problem with a bottom-up approach is that it requires organizations to enforce the use of standard dimensions and facts to ensure integration and deliver a single version of the truth.

#### **c. Hybrid Approach**

- The hybrid approach tries to blend the best of both "top-down" and "bottom-up" approaches. It attempts to capitalize on the speed and user-orientation of the "bottom-up" approach without sacrificing the integration enforced by a data warehouse in a "top down" approach. The hybrid approach relies on an extraction, transformation, and load (ETL) tool to store and manage the enterprise and local models in the data marts as well as synchronize the differences between them.

#### **Pros and Cons of Hybrid Approach**

- It combines rapid development techniques within an enterprise architecture framework.
- It develops an enterprise data model iteratively and only develops a heavyweight infrastructure once it's really.
- Backfilling a data warehouse can be a highly disruptive process that delivers no ostensible value and therefore may never be funded.
- Heavily on an ETL tool to synchronize meta data between enterprise and local versions, develop aggregates, load detail data, and orchestrate the transition to a data warehousing infrastructure.
- The hybrid approach may make it too easy for local groups to stray irrevocably from the enterprise data model.

#### **d. Federated Approach**

The federated approach is sometimes confused with the hybrid approach above or "hub-and-spoke" data warehousing architectures that are a reflection of a top-down approach. However, the federated approach-as defined by its most vocal proponent, Doug Hackney. Hackney says the federated approach is architecture of architectures". Hackney concedes that a federated architecture will never win awards for elegance or be drawn up on clean white boards as an "optimal solution".

### Pros and Cons of Hybrid Approach

- The approach merely encourages organizations to share the “highest value” metrics, dimensions, and measures wherever possible, however possible. Multiple data marts, data warehouses, or analytic applications.
- The major problem with the federated approach is that it is not well documented.
- Another potential problem is that without a specific architecture in mind, a federated approach can perpetuate the continued decentralization and fragmentation of analytical resources, making it harder to deliver an enterprise view in the end.

### III. Different Extraction Methods In Data Warehouse

The Extraction is the first step of the ETL (Extract, Transform and Load Process). Extraction is the process of extracting data from the source system for further use in the data warehouse environment [10].

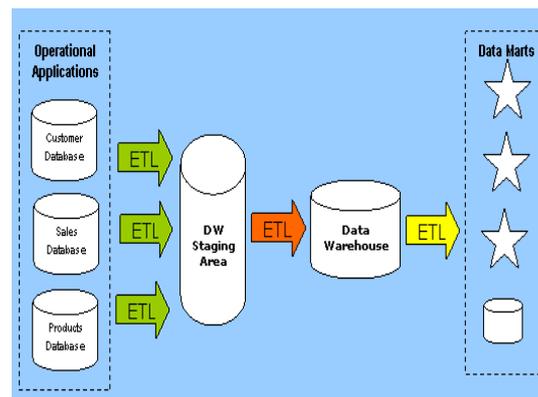


Figure 4: Extraction Methods

#### Types of Data Warehouse Extraction Methods

There are two types of data warehouse extraction methods [8]:

1. Logical Extraction Methods
2. Physical Extraction Methods

#### 1. Logical Extraction Methods

Logical Extraction it has two types:

- a. Full Extraction
- b. Incremental Extraction

##### a. Full Extraction

- In this method, data is completely extracted from the source system. The source data will be provided and no additional logical information is necessary on the source system.
- For example, exporting complete table in the form of flat file.

##### b. Incremental Extraction

- In incremental extraction, the changes in source data need to be tracked since the last successful extraction. Only these changes in data will be extracted and then loaded.
- Many Data warehouse system do not use change capture technique. Instead they extract the entire table from the source system into stage area and compare the data with previous version table and identify the data which has changed.

#### 2. Physical Extraction Methods

Physical Extraction has two methods:

- a. Online Extraction
- b. Offline Extraction

##### a. Online Extraction

In this process, extraction processes directly connect to the source system and extract the source data.

**b.Offline Extraction**

- The data is not extracted directly from the source system but is staged explicitly outside the original source system.

- The following common structure in offline extraction:

(i).Flat File: Generic format

(ii).Dump File: Database Specific file

**Example for Data Warehouse Extraction Methods**

- ✓ Extract into Flat File using External Tables
- ✓ Dump File
- ✓ Flat File using nzsqli

**IV. Characteristics Of Datawarehouse And Benefits**

**a. CHARACTERISTICS OF DATA WAREHOUSE**

There are three prominent data warehouse characteristics [9]:

1. **Integrated:** The way data is extracted and transformed is uniform, regardless of the original source.
2. **Time-variant:** Data is organized via time-periods (Weekly, Monthly, annually etc).
3. **Non-volatile:** A data warehouse is not updated in real-time. It is periodically updated via the uploading of data, protecting it from the influence of momentary change.

**b.DATABASES DATA WAREHOUSE**

**Database:** Databases are real-time repositories of information, which are usually tied to specific applications.

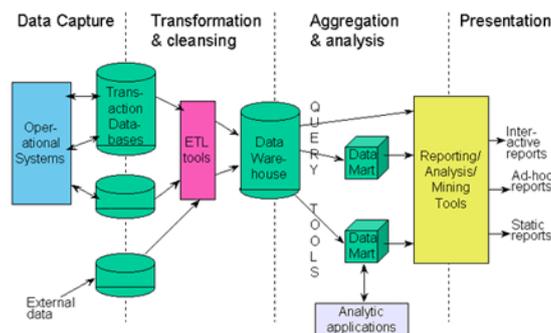
**Data Warehouse:** Data warehouses pull information from various sources (including databases), with a focus on the storage, filtering, retrieval and, specifically, analysis of huge volumes of structured data.

**c.DATAWAREHOUSE BENEFITS**

There are three main benefits to utilizing data warehouse:

1. Enablement of better decision making
  2. Quick and easy data access
  3. Consistent quality data
1. **Enablement of better decision making:** The corporate decision making no longer have to hedge their bets or make important business decisions based on partial or limited data.
  2. **Quick and Easy Data Access:** Users can access an array of information, stored across multiple sources, almost instantly. It means you won't be wasting time attempting to manually pull information from various sources.
  3. **Consistent quality data:** Data warehouse gather information from countless sources, but they convert it into a unified format to be used throughout your organization.

**V. Main Components Of Data Warehouse**  
Warehouse Components



**Figure 5:** Components Of Data Warehouse

The main components of data warehouse are [10]:

1. Overall Architecture
2. Data Warehouse Database
3. Sourcing, Acquisition, cleanup and Transformation Tools

4. Meta data
5. Access Tools
6. Data Marts
7. Data warehouse Administration and Management
8. Information Delivery System

### **1. Overall Architecture**

The data warehouse architecture is based on a relational database management system. The central information repository is surrounded by a number of key components designed to make the entire environment functional manageable and accessible by the both the operational systems.

The transformation process may involve conversion, summarization, filtering and condensation of data.

### **2. Data Warehouse Database**

The central data warehouse database is the core store of the data warehouse environment. These approaches include:

- a. Parallel Relation Database
- b. Innovative Approach
- c. Multidimensional database (MDDBs)

### **3. Sourcing, Acquisition, cleanup and Transformation Tools**

The data sourcing, cleanup, transformation and migration tools perform all of the conversions, summarization, key changes, structural changes and condensations needed to transform disparate data into information that can be used by decision support tool.

The functionality includes:

- Removing unwanted data from operational database.
- Converting to common data names and definitions.
- Establishing defaults for missing data.
- Accommodating source data definition changes

The data sourcing, cleanup, transformation and migration tools have to deal with some significant issues include:

- Database heterogeneity DBMSs are very different in data models, data access language, data navigation operations, concurrency integrity, recovery etc.
- Data heterogeneity, this is difference in the way data is defined and used in different models.

### **4. Meta data**

Meta data is data about data that describes the data warehouse it is used for building, maintaining, managing and using the data warehouse.

Meta data can be classified into:

- **Technical Meta Data**, which contains information about warehouse data for use by warehouse designers and administrations.
- **Business Meta Data**, which contains information that gives users an easy to understand perspective of the information stored in the data warehouse.

### **5. Access Tools**

Users interact with the data warehouse using front-end tools. Many of these tools requiring information specialist, many of the end users develop expertise in the tools.

Tools fall into four main categories:

1. Query and Reporting tools
2. Application Development tools
3. Online Analytical processing tools
4. Data Mining Tools

### **6. Data Marts**

The concept of a data mart is causing a lot of excitement and attracts much attention in the data warehouse industry.

These types of data mart, called dependent data marts because their data is sourced from the data warehouse. Each independent data mart makes to own assumptions about how to consolidate the data, and the data across several data marts may not be consistent.

### **7. Data warehouse Administration and Management**

Data warehouse tend to be as much as 4 times as large as related operational databases, reaching terabytes in size depending on how much history needs to be saved.

Managing data warehouses includes security and priority management, monitoring updates from the multiple sources.

### **8. Information Delivery System**

The information delivery component is used to enable the process of subscribing for data warehouse information and having it delivered to one or more destinations.

The web removes a lot of these issues by giving users universal and relatively inexpensive access to data couple this access with the ability.

## **VI. Different Types Of Data Warehouses**

There are three types of data warehouses [11]

1. Enterprise Data Warehouse
2. ODS (Operational Data Store)
3. Data Mart

**1. Enterprise Data Warehouse:** An enterprise data warehouse provides a central database for decision support throughout the enterprise.

**2. ODS (Operational Data Store):** This has a broad enterprise wide scope, but unlike the real enterprise data warehouse, data is refreshed in near real time and used for routine business activity.

**3. Data Mart:** Data Mart is a subset of data warehouse and it supports a particular region, business unit or business function.

## **VII. Data Warehouse Applications**

A data warehouse helps business executives [10] to organize, analyze, and use their data for decision making. A data warehouse serves as a sole part of a plan-execute-assess “closed loop” feedback system for the enterprise management. Data warehouse is widely used in the following fields.

- Financial services
- Banking services
- Consumer goods
- Retail sectors
- Controlled manufacturing
- Information Processing
- Analytical Processing
- Data Mining
- Real Life
- Various Industries
- Statistical Analysis
- Decision Making
- Mailing Box Applications

## **VIII. Conclusion**

The data warehouse is a database, which is kept separate from the organizations operational database. A data warehouse provides us generalized and consolidated data in multidimensional view, Along with generalized and consolidated view of data, a data warehouse also provides us Online Analytical Processing (OLAP) tools. These tools help us to interactive and effective analysis of data in a multidimensional space. The analysis results in data generalization and data mining. Discuss here overview of datawarehouse, Data warehouse design approach, different extraction methods in data warehouse, Characterizes and benefits of data warehouse, Components of data ware, different types of data warehouse and data warehouse applications are discussed.

### **References**

- [1]. PaulrajPonniah,Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals published by John Wiley & Sons, Inc.
- [2]. Data warehousing in the Real World – Sam Anahory& Dennis Murray. Pearson Edn Asia.
- [3]. Data Warehousing Fundamentals – Paulraj PonniahWiley Student Edition.
- [4]. The Data Warehouse Life cycle Tool kit – Ralph KIMBALL Wiley Student Edition.
- [5]. Sandeep Singh and SonaMalhotra (2011),”Data Warehouse and its Methods”, Volume 2, No. 5, May 2011,Journal of Global Research in Computer Science,ISSN-2229-371X.
- [6]. <http://tdan.com/four-ways-to-build-a-data-arehouse/4770>.
- [7]. <http://www.folkstalk.com/2011/04/data-warehouse-design- approaches.html>.
- [8]. <http://www.dwgeek.com.different-extraction-methods>.
- [9]. <https://automic.com>.
- [10]. [www.tutorial.com](http://www.tutorial.com).
- [11]. <https://tekslate.com/types-data-warehouses>