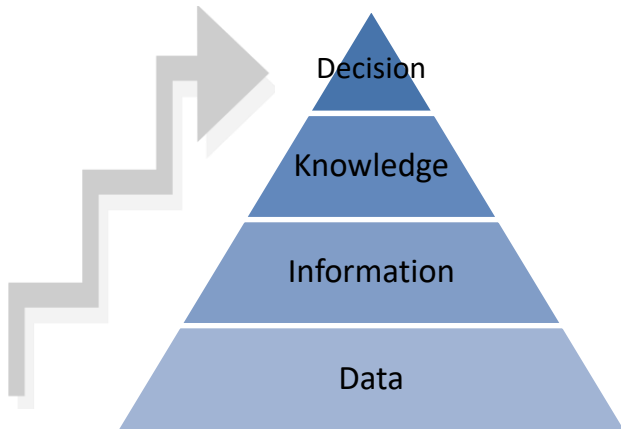


Data Warehouse

By: Abdelwahed Ashraf

@linkedin @gmail

Data → Information → Knowledge → Decision



Introduction to DWH

Motivation to the Data Warehouse (DWH)

- Data could be a product for some companies.
- It could be decision support for other products or businesses.
- It could be reporting the results after passing the data life-cycle from storage (Database). Some challenges are facing the people who work on data management backend:
 - Performance,
 - Integration,
 - and Applying analytical functions.
- Vendors who are working on solving the above challenges are creating their product of DWH. Their ultimate goal is to optimize the above points.

Definition of a Data Warehouse (DWH)

A DWH is a technique for collecting and managing data from varied sources to **provide meaningful business insights.**

The information is subject orientated, recorded over time and may be stored at various degrees of summarization

Motivation to the Data Warehouse (DWH)

- The DWH is not a product but an environment.
- It is a process of transforming data into information and make it available to users in a **timely manner** to make a difference.
- It is an architectural construct of an information system that provides users with current and historical decision support information which is difficult to access or present in the traditional operational data store.
- The DWH is the core of the BI system built for data analysis and reporting.

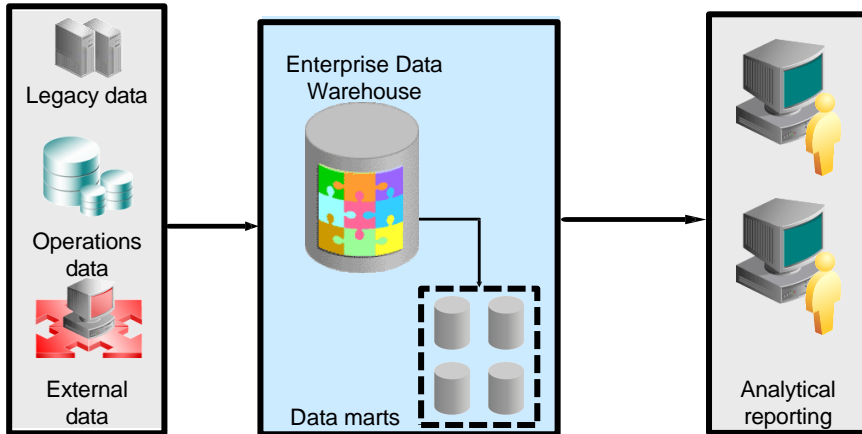
Other names for the Data warehouse system:

- Decision Support System (DSS).
- Business Intelligence Solution.
- Executive Information System.
- Management Information System.
- Analytic Application.
- Data Warehouse.

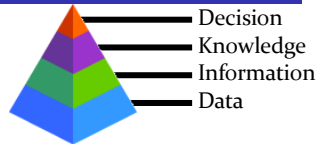
Differences Between DWH and Operational DB

Transactions DB (OLTP)	DWH
Works with small Pieces of Information	Works with Enterprise wide Information
Support Insert, Update, Delete or Select	Read Only
Normalized	Not required (De-normalized in many use cases)
Small To Large Database	Large to Very Large Database
Volatile Data	Non Volatile
Applications that Run the business	Applications that analyze the business

Data Warehousing and Business Intelligence

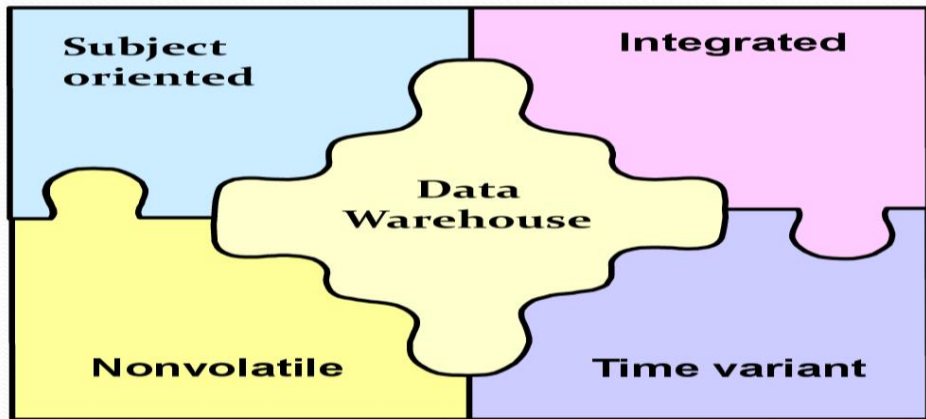


Definition of BI



- BI is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies.
- BI a content-free expression, so it means different things to different people.
- BI helps transform data, to information (and knowledge), to decisions and finally to action.

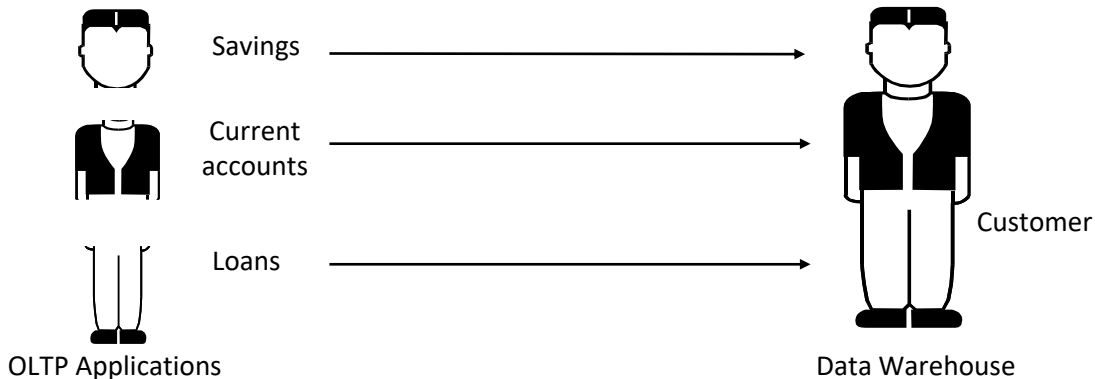
DWH Characteristics



The characteristics of DWH:

● Integrated:

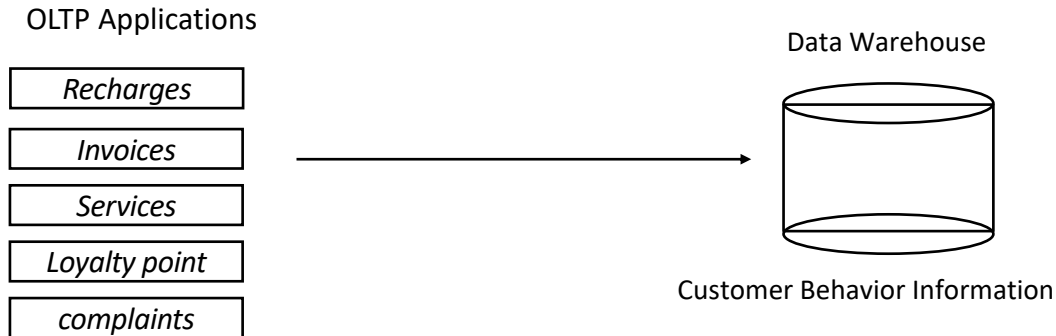
- *DWH is an integrated environment which allows us to integrate different source systems. Data are modeled (organized) in a unified manner.*



The characteristics of DWH:

● Subject-oriented:

- Data is categorized and stored by business subject rather than by application.



The characteristics of DWH:

● Time-Variant:

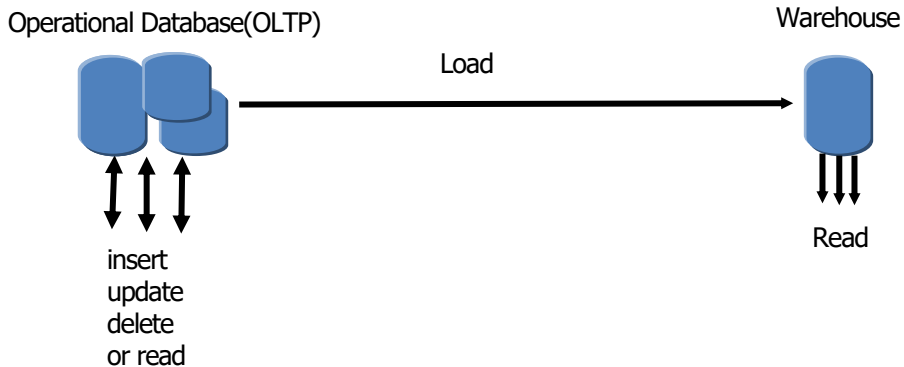
- Data modeled (organized) based on periods (hourly, daily, weekly, monthly, quarterly, yearly).

Time	Data
Jan-97	January
Feb-97	February
Mar-97	March

The characteristics of DWH:

● Non-Volatile:

- Typically, data in the data warehouse is not updated or deleted.



The characteristics of DWH:

● Non-Volatile:

● Changing Data.

Operational Database



insert
update
delete
or read

First time load



Warehouse



Refresh



Refresh



Types of DWH

Motivation to Data Warehouse

Types of Data Warehouse

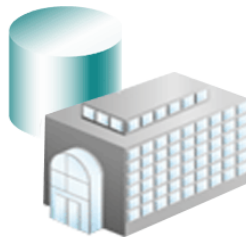
Enterprise Data Warehouse (E-DWH): It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data (DWH Model). It offers data classifications according to the subject with privileges policy.

Operational Data Store (ODS): is a central database that provides an up-to-date (real-time) data from multiple transnational systems for operational reporting into a single DWH.

Data Mart: A departmental data warehouse that stores only relevant data, It specially designed for a particular line of business, such as sales or finance.

Enterprise Data Warehouse (E-DWH)

- Supports large-scale implementation
- Scopes the entire business
- Contains data from all subject areas
- Is developed incrementally
- Is a single source of enterprise-wide data
- Is the single distribution point to dependent data marts



Data Mart: A departmental data warehouse that stores only relevant data

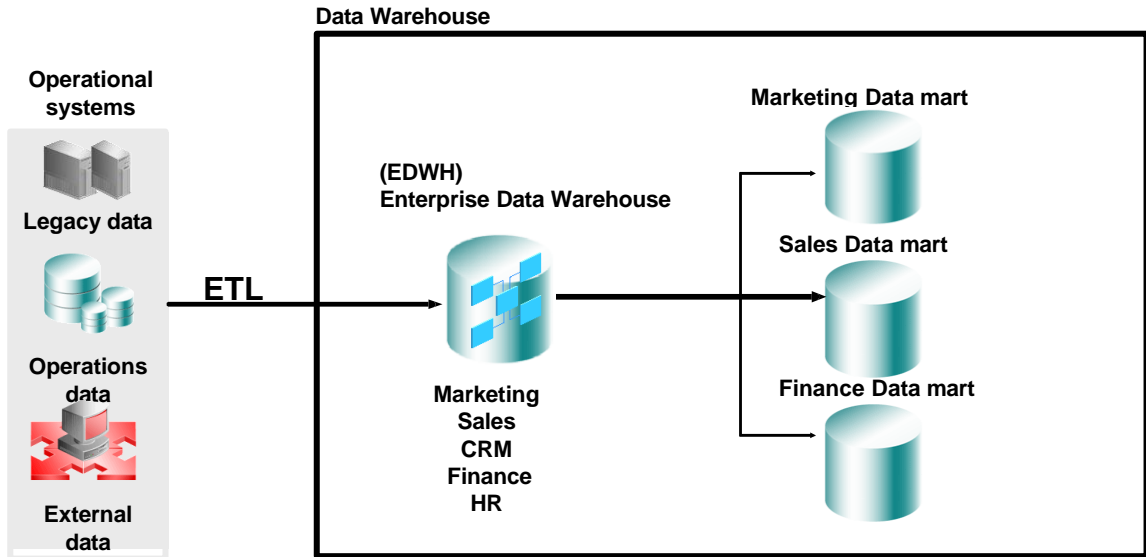
- **Dependent data mart**

A subset that is created directly from a E-data warehouse

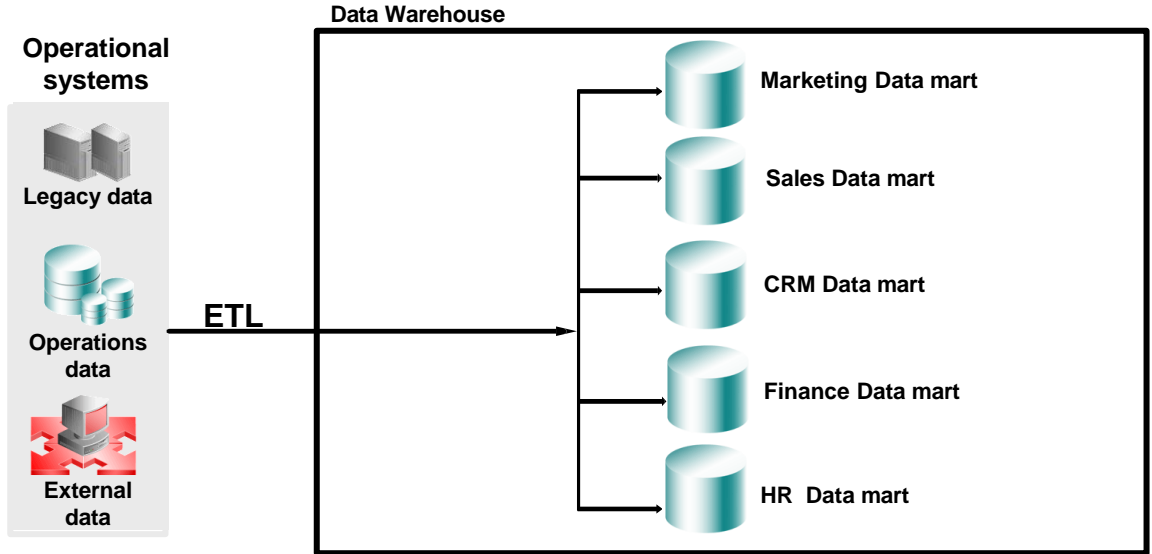
- **Independent data mart**

A small data warehouse designed for a strategic business unit or a department

Dependent Data Mart



Independent Data Mart



DWH vs ODS vs Data Mart

Metric	E-DWH	ODS	Data Mart
Latency	Day -1 (Batch)	Real-time(Stream)	Day -1
Data level	Transnational	Transnational	Summary
Historical	Long-term	Snapshot	Aggregated Long-Term
Size	TB/PB	GB	GB/TB
Orientation	Multi sources	Multi sources	Product
Business Units	Multi organizational units	Product team	Business team

Use Cases of Operational DB vs DWH

A Retail Sales company named XSales.

- They have lots of systems.
- One of this systems is a CRM system as example of operational DB.
- The CRM system handles the customer activities with the company including (sales, Return, inquiries and other activities).
- This system has a backend database (OLTP).
- CRM team can report their sales and customer activities from their database.
- Product owner can take a decision based on their system backend reports.

What is the need for DWH?

- This company has other systems Marketing, Stock, Call center
- They need to report information related to the CRM, Stock , Call center source systems in one report.
 - So, they need to ingest (transfer) the data from the source systems to one single database.
 - The decision from the DHW is a **global and strategical decision**.
 - If the company needs to build a machine learning model which needs data from different sources. They need to load the data from a centralized database rather than read each source alone.

Use case (ODS)

- Why do we need the ODS?
- How does it fit in our system?

XSales has a call center system which handles the customer inquiries.

This system requires the some data related to order status, Return, customer information, billing details, shipping to be calculated and accumulated in **real-time** to be able to give the customer the right answer for his inquires.

Use case (ODS)

So, What is the challenge for this system?

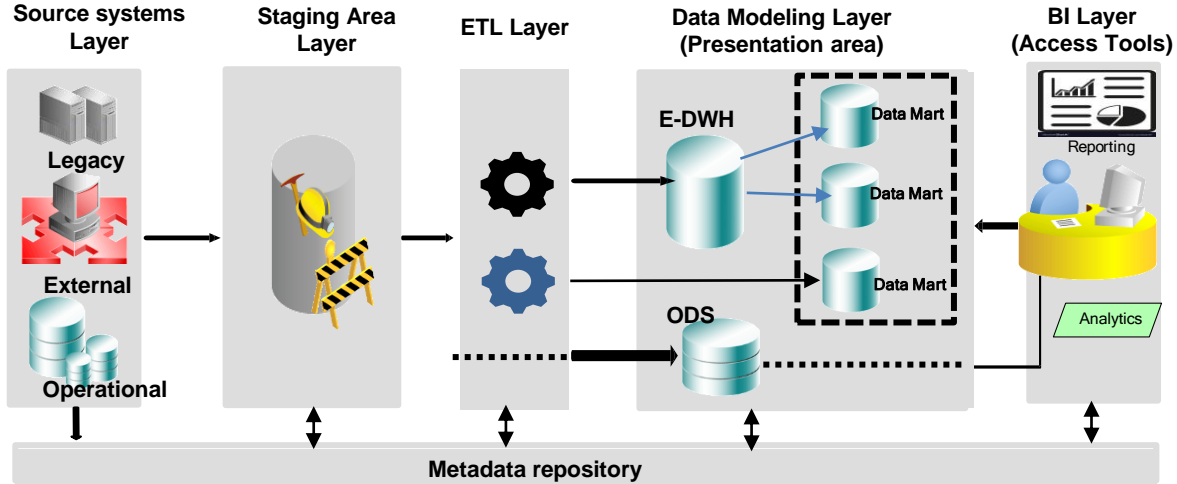
- It needs specific information from different source systems.
- It requires to track the source system database changes or update in **real-time**.
- It's functionality is based on the aggregate data not the transactions

Use case (ODS)

- ODS is based on change data capture (CDC).
- This approach used to determine the data change and apply action based on this change.
- ODS uses the real-time aggregations to support the online systems from different source systems.

DWH Architecture

Data Warehouse Components



Data Warehousing Architectures

- Issues to consider when deciding which architecture to use:
 - Which database management system (DBMS) should be used?
 - Will parallel processing and/or partitioning be used?
 - Will data migration tools(ETL) be used to load the data warehouse?
 - What tools will be used to support data retrieval and analysis?

Hosted Data Warehouses

- Benefits:

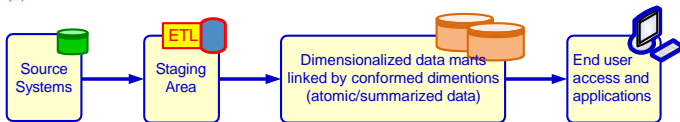
- Requires minimal investment in infrastructure
- Frees up capacity on in-house systems
- Frees up cash flow
- Makes powerful solutions affordable
- Enables powerful solutions that provide for growth
- Offers better quality equipment and software
- Enables users to access data remotely
- Allows a company to focus on core business
- Meets storage needs for large volumes of data

Five Main DW Architectures

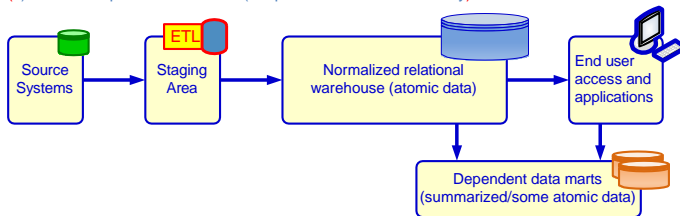
(a) Independent Data Marts Architecture



(b) Data Mart Bus Architecture with Linked Dimensional Datamarts

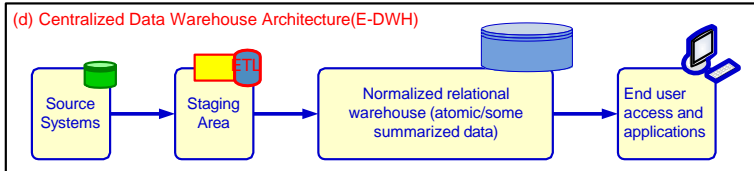


(c) Hub and Spoke Architecture (Corporate Information Factory)

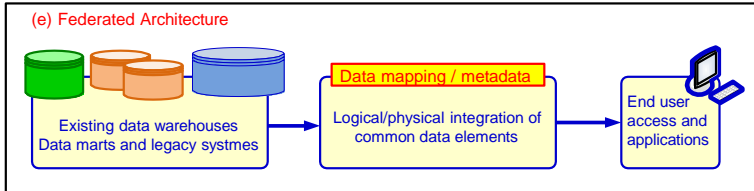


Five Main DW Architectures

(d) Centralized Data Warehouse Architecture(E-DWH)



(e) Federated Architecture



Five Main DW Architectures

1. Independent Data Marts
 2. Data Mart Bus Architecture
 3. Hub-and-Spoke Architecture
 4. Centralized Data Warehouse
 5. Federated Data Warehouse
- Each has pros and cons!

Data Warehouse Development

- Data warehouse development approaches
 - Inmon Model: EDW approach (top-down)
 - Kimball Model: Data mart approach (bottom-up)
 - Which model is best?
 - There is no one-size-fits-all strategy to DW
 - One alternative is the hosted warehouse
- Data warehouse structure:
 - The Star Schema vs. Relational