LTI
Let's Solve

Data warehouse- An Overview

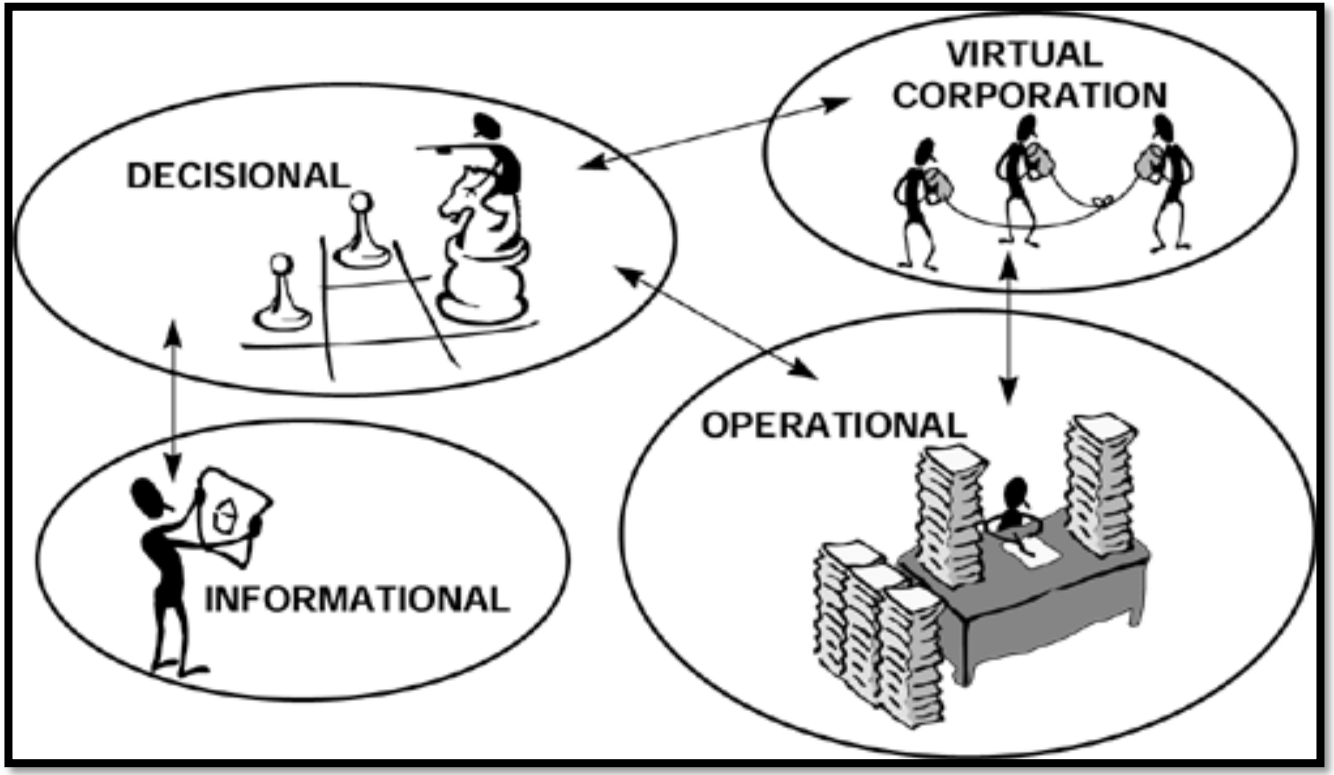
Introduction

BUSINESS PERSPECTIVE

- From the business perspective, the requirements of the enterprise fall into categories illustrated and described below:

  - OPERATIONAL
  - INFORMATIONAL
  - DECISIONAL ‘&’
  - VIRTUAL CORPORATION

View in Business Perspective



Operational

- Technology supports the smooth execution and continuous improvement of day-to-day operations.

- The identification and correction of errors through exception reporting and workflow management, and the overall monitoring of operations.

- Information retrieved about the business from an operational viewpoint is used to either complete or optimize the execution of a business process.

Operational

- Example:
  - Day-to-day transaction includes - ATM Withdrawals which includes Checking Balance, Mini Statement, Withdrawal, etc…

Informational

- Technology makes current, relatively static information widely and readily available to as many people as need access to it.

- Examples include company policies, product and service information, organizational setup, office location, corporate forms, training materials, company profiles.

Decisional

- Technology supports managerial decision-making and long-term planning.

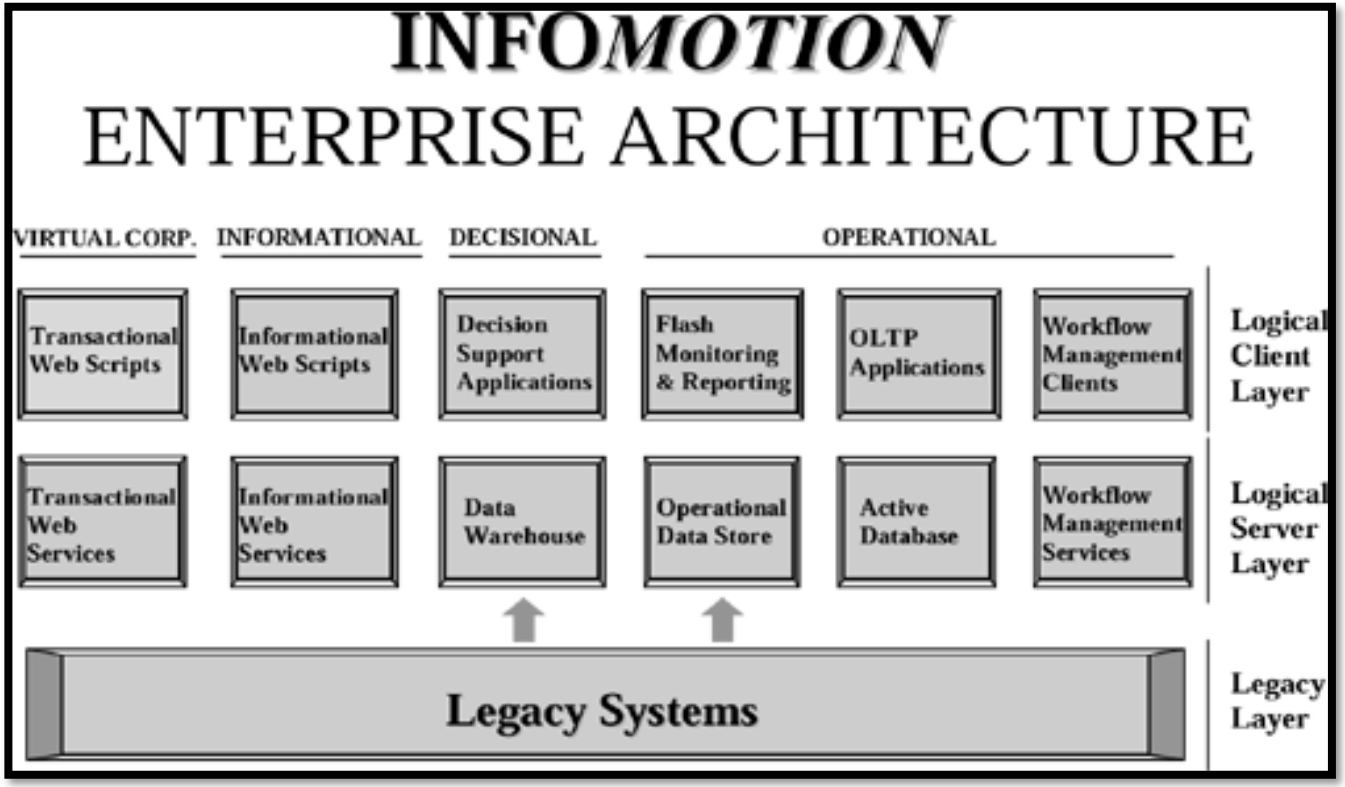
- Decision-makers are provided with views of enterprise data from multiple dimensions and in varying levels of detail. Historical patterns in sales and other customer behaviour are analysed.

- Examples of Decisional systems support includes decision-making and planning through scenario-based modelling, what-if analysis, trend analysis, and rule discovery.

Virtual Corporation

- Technology enables the creation of strategic links with key suppliers and customers to better meet customer needs.
- In the past, such links were feasible only for large companies because of economies of scale.
- Now, the affordability of Internet technology provides any enterprise with this same capability.

HIGH LEVEL ARCHITECTURE



Business System

- We can divide IT systems into transactional (OLTP) and analytical (OLAP).

- In general we can assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

OLTP / Transactional Systems

- *OLTP (On-line Transaction Processing)* is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE).

- The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second.

- In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).

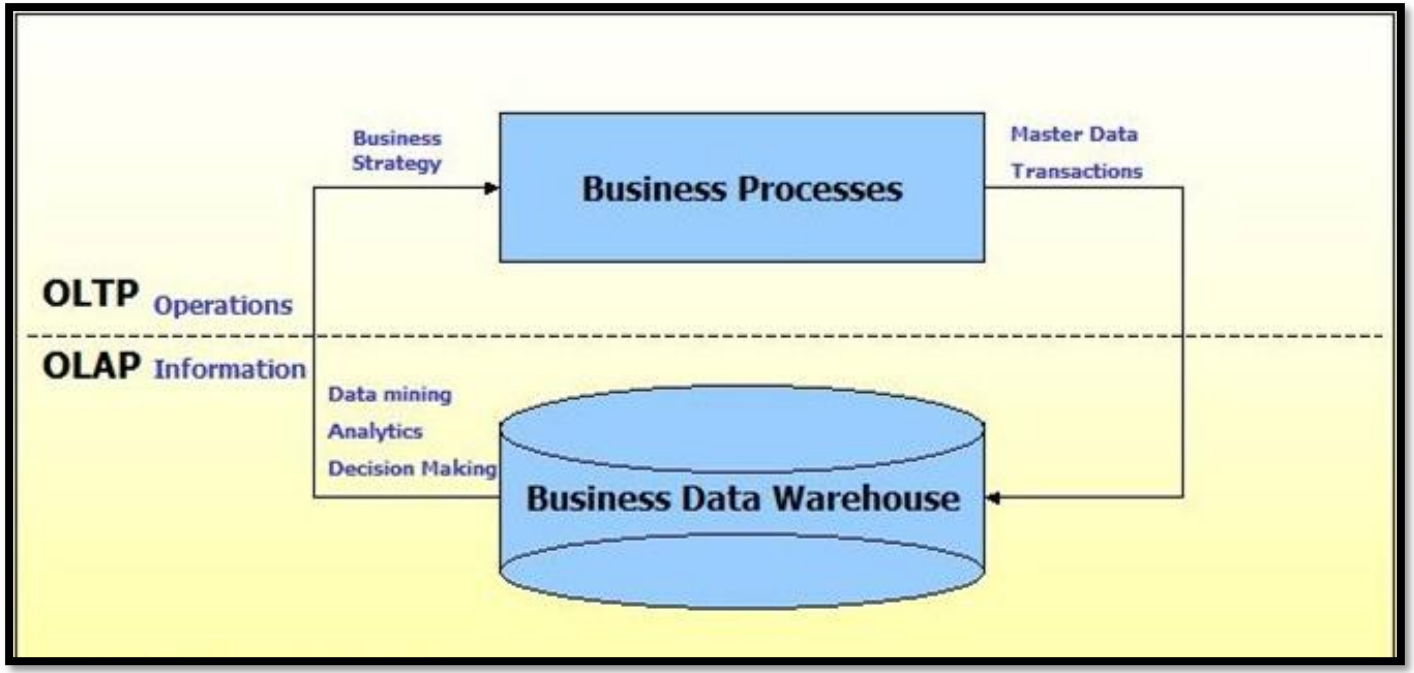
OLAP / Analytical Systems

- *OLAP (On-line Analytical Processing)* is characterized by relatively low volume of transactions.

- Queries are often very complex and involve aggregations. For OLAP systems a response time is an effectiveness measure. OLAP applications are widely used by Data Mining techniques.

- In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema).

OLTP (vs.) OLAP



Functions of Data Warehouse Tools and Utilities

- The following are the functions of data warehouse tools and utilities:

  - Data Extraction - Involves gathering data from multiple heterogeneous sources.
  - Data Cleaning - Involves finding and correcting the errors in data.
  - Data Transformation - Involves converting the data from legacy format to warehouse format.
  - Data Loading - Involves sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
  - Refreshing - Involves updating from data sources to warehouse.

Differences Between OLTP and OLAP

	OLTP	OLAP
users	clerk, IT professional	knowledge worker
function	day to day operations	decision support
DB design	application-oriented	subject-oriented
data	current, up-to-date detailed, flat relational isolated	historical, summarized, multidimensional integrated, consolidated
usage	repetitive	ad-hoc
access	read/write index/hash on prim. key	lots of scans
unit of work	short, simple transaction	complex query
# records accessed	tens	millions
#users	thousands	hundreds
DB size	100MB-GB	100GB-TB
metric	transaction throughput	query throughput, response

DWH Properties

“ A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process.”

- The properties of DWH are as Follows:
  - Subject-Oriented
  - Integrated
  - Non-Volatile
  - Time Variant

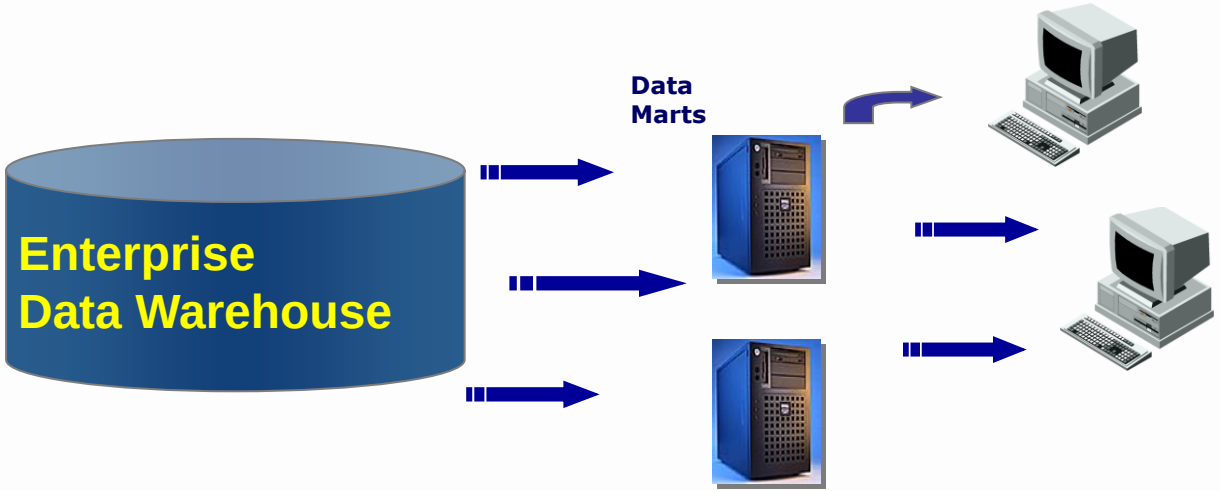
DWH Layers

- In general, all data warehouse systems have the following layers:
  - Data Source Layer
  - Data Extraction Layer
  - Staging Area
  - ETL Layer
  - Data Storage Layer
  - Data Logic Layer
  - Data Presentation Layer
  - Metadata Layer
  - System Operations Layer

What is a Data mart?

- Data mart is a decentralized subset of data found either in a data warehouse or as a standalone subset designed to support the unique business unit requirements of a specific decision-support system.

- Data marts have specific business-related purposes such as measuring the impact of marketing promotions, or measuring and forecasting sales performance etc,.



Data marts - Main Features

Main Features:

- Low cost
- Controlled locally rather than centrally, conferring power on the user group.
- Contain less information than the warehouse
- Rapid response
- Easily understood and navigated than an enterprise data warehouse.
- Within the range of divisional or departmental budgets

Advantages of Data mart over Data warehouse

- Typically single subject area and fewer dimensions
- Limited feeds
- Very quick time to market (30-120 days to pilot)
- Quick impact on bottom line problems
- Focused user needs
- Limited scope
- Optimum model for DW construction
- Demonstrates ROI
- Allows prototyping

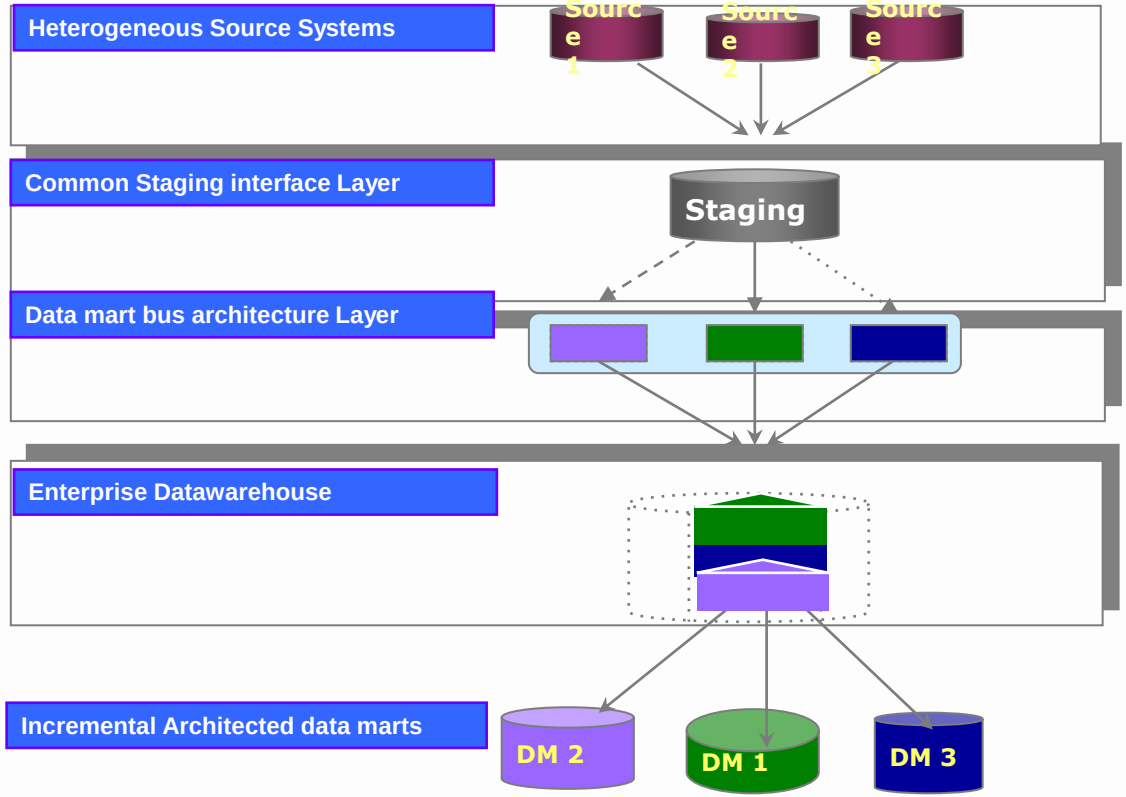
DWH DESIGN APPROACHES

- Data warehouse design is one of the key technique in building the data warehouse.
- Choosing a right data warehouse design can save the project time and cost.
- Basically there are two data warehouse design approaches are popular.

Top-Down Design:

- In the top-down design approach the, data warehouse is built first.
- The data marts are then created from the data warehouse.

“Top Down "Approach – Innmon’s



Advantages of top-down design

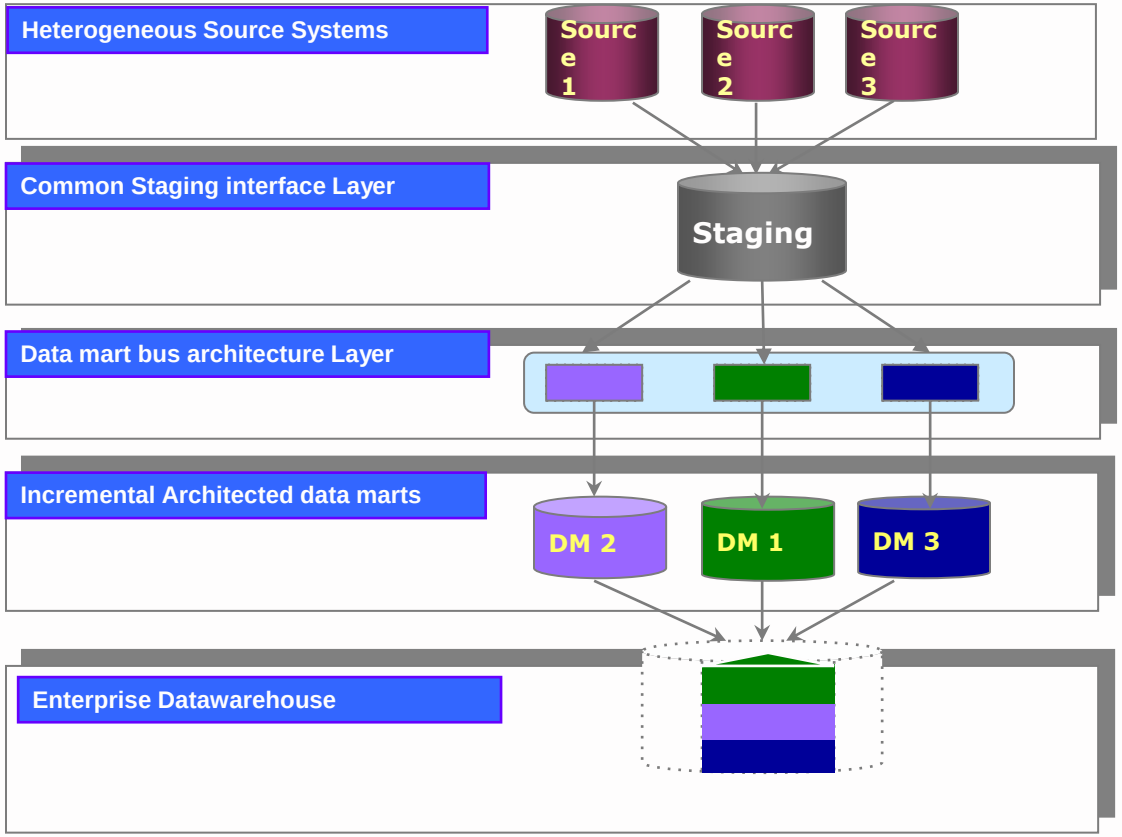
- Provides consistent dimensional views of data across data marts, as all data marts are loaded from the data warehouse.

- This approach is robust against business changes. Creating a new data mart from the data warehouse is very easy.

Bottom-Up Design:

- In the bottom-up design approach, the data marts are created first to provide reporting capability.
- A data mart addresses a single business area such as sales, Finance etc.
- These data marts are then integrated to build a complete data warehouse.

“Bottom up "Approach – Kimball’s



Advantages of bottom-up design are:

- This model contains consistent data marts and these data marts can be delivered quickly.
- As the data marts are created first, reports can be generated quickly.
- The data warehouse can be extended easily to accommodate new business units. It is just creating new data marts and then integrating with other data marts.

OLTP Vs. ODS Vs. DWH

Characteristic	OLTP	ODS	Data Warehouse
Data redundancy	Non-redundant within system; Unmanaged redundancy among systems	Somewhat redundant with operational databases	Managed redundancy
Data stability	Dynamic	Somewhat dynamic	Static
Data update	Field by field	Field by field	Controlled batch
Data usage	Highly structured, repetitive	Somewhat structured, some analytical	Highly unstructured, heuristic or analytical
Database size	Moderate	Moderate	Large to very large
Database structure stability	Stable	Somewhat stable	Dynamic

Let's Solve