LTI

Celebrating 20 Years

Data warehouse Concepts

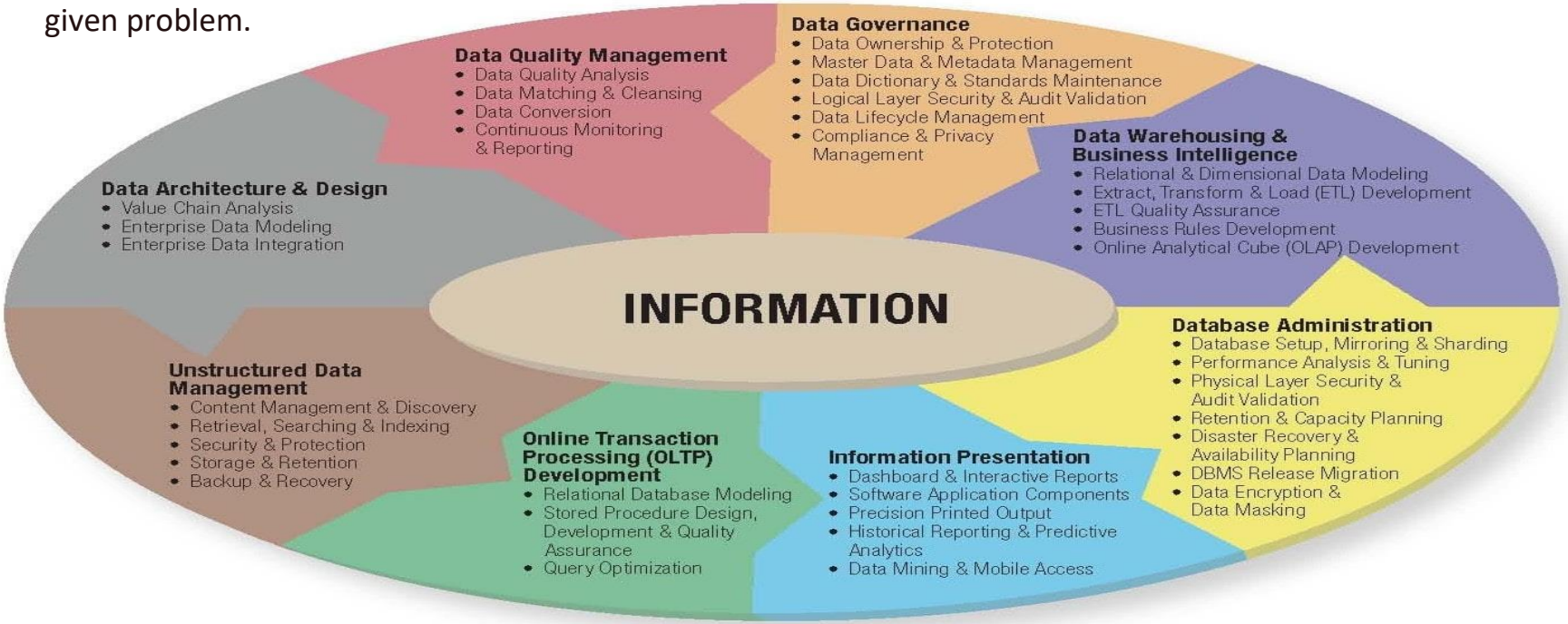
A Larsen & Toubro Group Company

Agenda

- Data as Information/Knowledge
- Why Business Intelligence
- Why Data warehouse
- Data Mart
- Meta Data
- ETL
- Data warehouse Architecture

Data as Information/Knowledge

- The goal of Informational Processing is to turn data into information!
- Why?
- Because business questions are answered using *information* and the *knowledge* of how to apply that information to a given problem.



Data as Information/Knowledge – College Case Study

- Registration
- Admissions
- Billing
- Financial Aid
- Student Services
- Course Management
- Alumni/Development
- Fiscal Management

Data as Information/Knowledge – What Case Study is this?

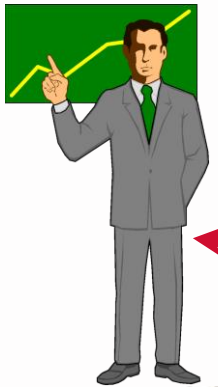
- Client
  - Plan
    - Plan & Design
    - Business Process & Technology Consulting
    - Service Modelling & Product Design
  - Build
    - Network Installation
    - Service Assurance Solutions
    - Product/Technology development
  - Operate
    - Survey
    - Optimization
    - Helpdesk

Need for a Separate informational system

- *Data*: Informational data is distinctly different from operational data in its structure and content

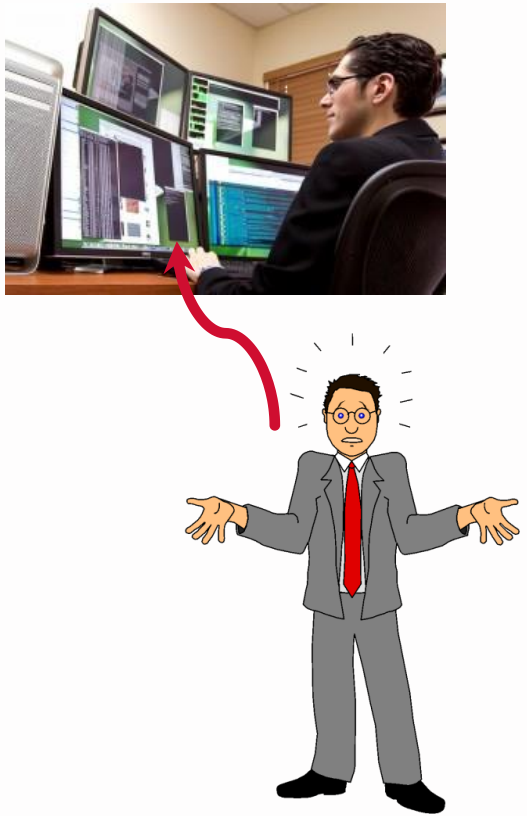
- *Processing:* Informational processing is distinctly different from operational processing in its characteristics and use of data

The Information Center



- **Management requires business information**
- **A request for a report is made to the Information Center**
- **Information Center works on developing the report**
- **Requirements for the report must be clarified**

The Information Center

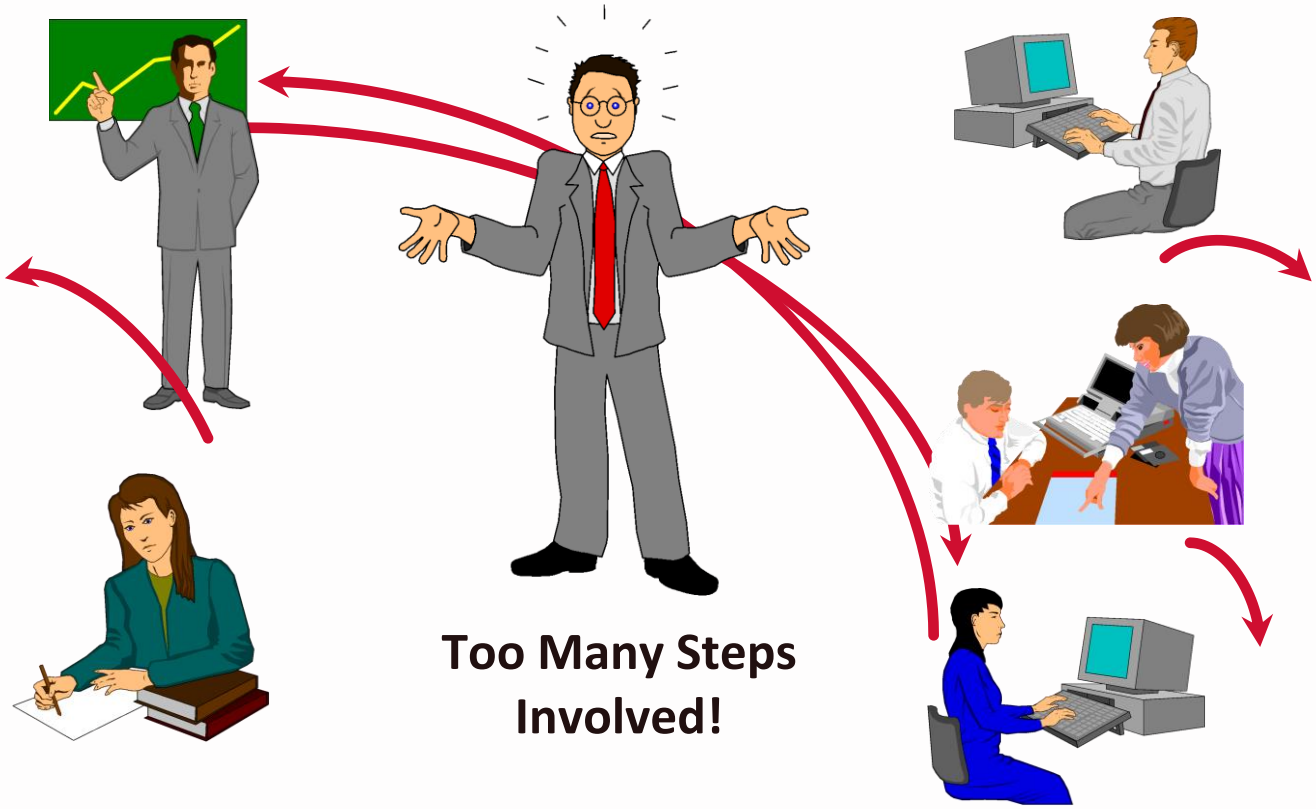


- **Report provided to analyst**
- **Analyst manipulates data for decision making**
- **Management receives information, but...**

What took so long? and

How do I know it's right?

The Information Center



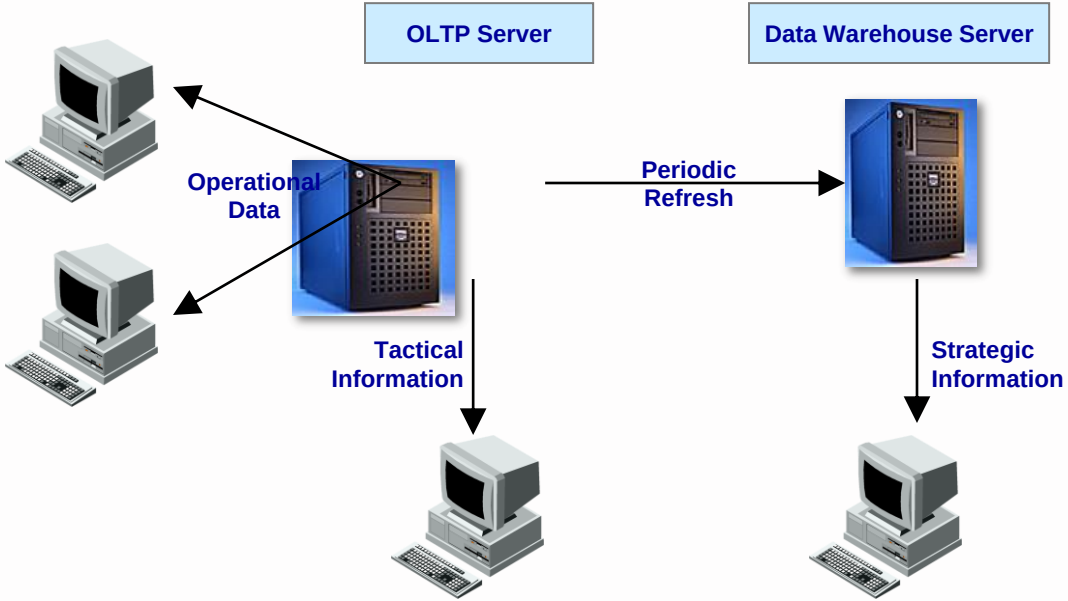
Tactical Information

- Supports day to day control operations
- Transaction Processing
- High Performance Operational Systems
- Fast Response Time
- Initiates immediate action

Strategic Information

- Understand Business Issues
- Analyze Trends and Relationships
- Analyze Problems
- Discover Business Opportunities
- Plan for the Future

Why we need Tactical and Strategic Information



- Operational data helps the organization meet operational and tactical requirements for *data*.
- While the Data Warehouse data helps the organization meet *strategic* requirements for *information*

Information System: OLTP(DB) Vs OLAP(DWH)

Operational Systems	Data warehouse Systems
Operational systems are generally designed to support high-volume transaction processing (i.e. OLTP) with minimal back-end reporting.	Data warehousing systems are generally designed to support high-volume analytical processing (i.e. OLAP) and subsequent, often elaborate report generation.
Operational systems are generally process-oriented or process-driven, meaning that they are focused on specific business processes or tasks. Example tasks include billing, registration, etc	Data warehousing systems are generally subject-oriented, organized around business areas that the organization needs information about. Such subject areas are usually populated with data from one or more operational systems. As an example, revenue may be a subject area of a data warehouse that incorporates data from operational systems that contain student tuition data, alumni gift data, financial aid data, etc.
Operational systems are generally concerned with current data.	Data warehousing systems are generally concerned with historical data.

Information System: OLTP(DB) Vs OLAP(DWH)

Operational Systems	Data warehouse Systems
Data within operational systems are generally updated regularly according to need.	Data within a data warehouse is generally non-volatile, meaning that new data may be added regularly, but once loaded, the data is rarely changed, thus preserving an ever-growing history of information. In short, data within a data warehouse is generally read-only.
Operational systems are generally optimized to perform fast inserts and updates of relatively small volumes of data ~5GB.	Data warehousing systems are generally optimized to perform fast retrievals of relatively large volumes of data ~100TB.
Operational systems are generally application-specific, resulting in a multitude of partially or non-integrated systems and redundant data(e.g. billing data is not integrated with payroll data).	Data warehousing systems are generally integrated at a layer above the application layer, avoiding data redundancy problems.
Operational systems generally require a non-trivial level of computing skills amongst the end-user community.	Data warehousing systems generally appeal to an end-user community with a wide range of computing skills, from novice to expert users.

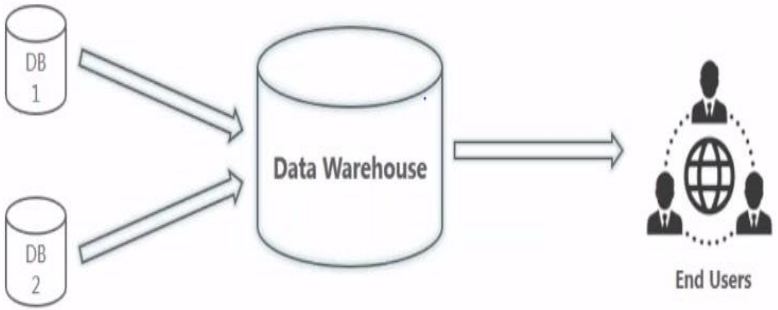
Information System: OLTP(DB) Vs OLAP(DWH) - Examples

- OLTP
  - A Super market server captures every single record purchased at the market
  - A bank server which records every transaction is made for a particular account
  - A railway reservation server which records the transactions of a Passenger
- OLAP
  - Bank manager wants to know how many customers are utilizing the ATM at his/her branch. On this he may take a business decision whether to use the ATM at his/her branch or to relocate.
  - An insurance company wants to know the number of polices for each agent has sold. This will give an insight on the agent's performance

Why Business Intelligence

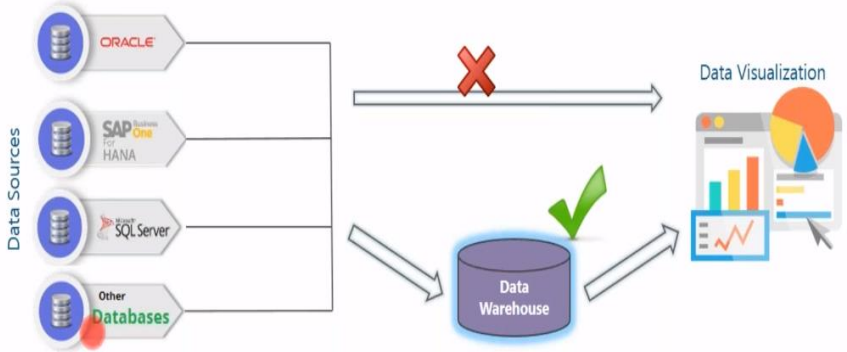
- Why Business Intelligence
  - Data has to be secure and distributed efficiently for important up-to-date **business** decisions. ... Hence data can help maximize revenues and reduce costs.
  - A **Business Intelligence** (BI) solution helps in producing accurate reports by extracting data directly from your data source. Which in turn help in contributes to the growth of the company
    - Planning Data Gathering Data Analysis Business Action
      - It is act of transforming raw/operational data into useful information for **business** analysis

- How it works?
  - BI is based on Data warehouse technology **extracts** information from a company's operational systems
  - The data is **transformed** (Cleaned and integrated), and **loaded** into Data warehouse
  - Since this data is credible, it is used for business insights.

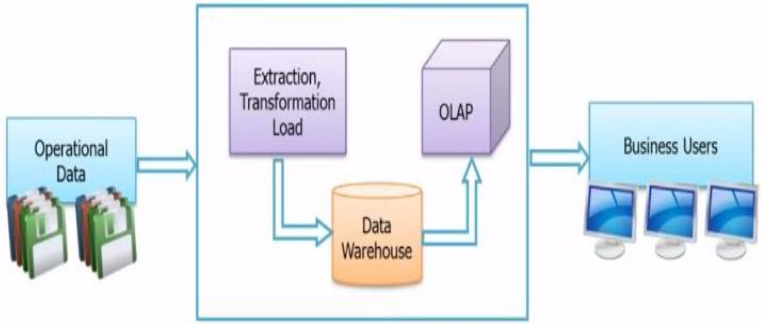


But Why Data Warehousing

- Let's understand the challenges in achieving Business Intelligence
- Data Collected from various sources & stored in various databases cannot be directly visualized
- The data first needs to be **integrated** and then **processed** before visualization takes place

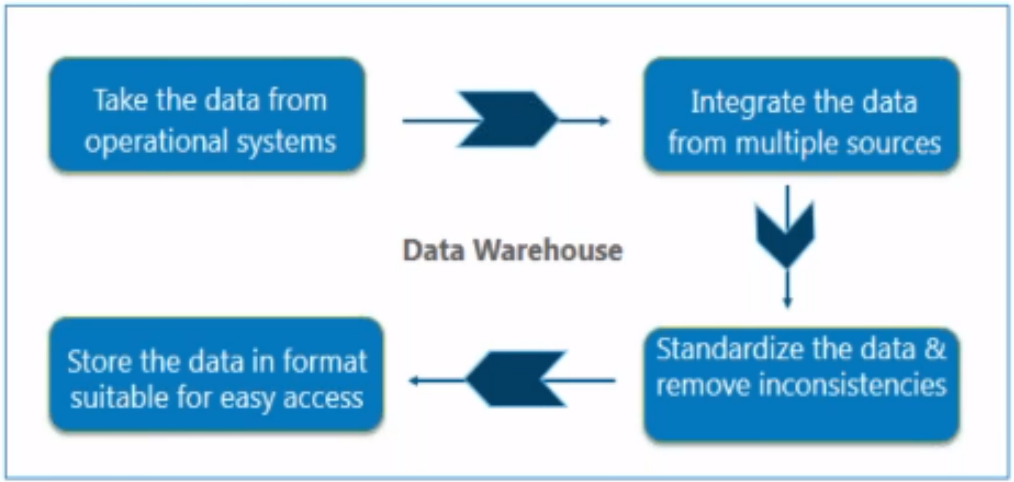


- Let's understand what is a Data warehouse
  - A Central location where consolidated data from multiple location (databases) are stored
  - DWH is maintained separately from an organization's operational database
  - End users access it whenever any information is needed
  - DWH is not loaded every time new data is added to



Advantages of Data Warehousing

- Strategic questions can be answered by studying trends
- Data warehousing is more faster and accurate
- DWH is not a product that a company can go and purchase, it needs to be designed & depends on entirely on company's requirements



Properties of Data Warehousing

- "A data warehouse is a *subject-oriented, integrated, time-variant and non-volatile* collection of data in support of management's decision making process" - Bill Inmon (Father of Data Warehousing)
- "A data warehouse is a copy of transaction data specifically structured for query and analysis" - Ralph Kimball
- Subject-Oriented
  - Data is categorized and stored by business subject rather than by application
- Integrated
  - Data on given subject is collected from disparate sources and stored in a single place
- Time-variant
  - Data is stored as series of snapshots, each representing a period of time
- Non-volatile
  - Typically data in the data warehouse is not updated or deleted

Evolution of Data Warehousing

Traditional approaches to computer system design during 1980's

- Not optimized for analysis and reporting
- Company wide reporting couldn't be supported from a single system
- For developing reports often required writing specific computer programs which was slow and expensive

- 60's: Batch reports
  - hard to find and analyze information
  - inflexible and expensive, reprogram every new request
- 70's: Terminal-based DSS(Decision support System) and EIS (executive information systems)
  - still inflexible, not integrated with desktop tools
- 80's: Desktop data access and analysis tools
  - query tools, spreadsheets, GUIs
  - easier to use, but only access operational databases
- 90's: Data warehousing with integrated OLAP engines and tools

Data Warehousing - Snapshot

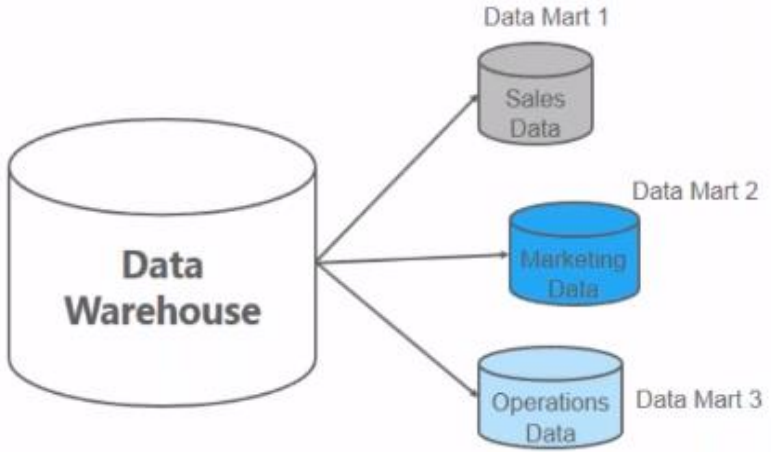
- Features
  - Strategic enterprise level decision support
  - Multi-dimensional view on the enterprise data
  - Caters to the entire spectrum of management
  - Descriptive, standard business terms
  - High degree of scalability
  - High analytical capability
  - Historical data only
- Business Benefits
  - Understand Business Trends
  - Better forecasting decisions
  - Better products to markets in timely manner
  - Analyze daily sales information and make quick decisions
  - Solution for maintaining your company's competitive edge

- Some Key Application Areas
  - Risk management
  - Financial analysis
  - Marketing programs
  - Profit trends
  - Procurement analysis
  - Inventory analysis
  - Statistical analysis
  - Claims analysis
  - Manufacturing optimization
  - Customer relationship management

Data Mart

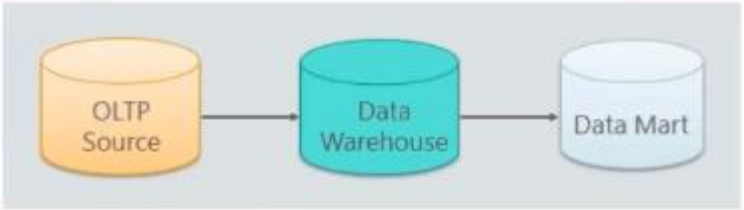
- Data Mart is smaller version of DWH which deals with a single subject
- Data Mart are focused on one area. Hence they draw the data from a limited number of sources
- Time taken to build Data Mart is very less compared to the time taken to build Data warehouse

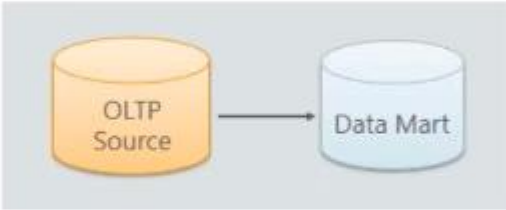
Data Warehouse	Data Marts
Enterprise wide data	Department wide data
Multiple subject areas	Single subject area
Multiple data sources	Limited data sources
Occupies large memory	Occupies limited memory
Longer time to implement	Shorter time to implement

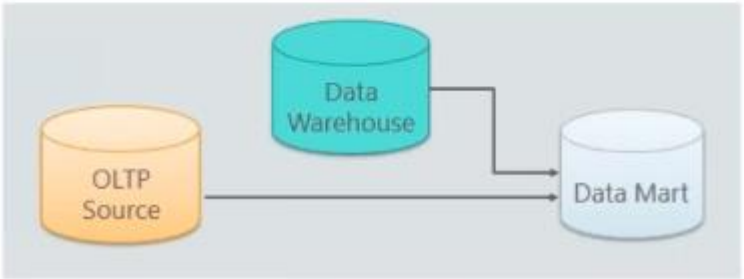


Data Mart - Types

- Dependent Data Mart
  - The data is first extracted from OLTP system and then to populated in Central DWH
  - From the DWH, the data travels to Data Mart
- Independent Data Mart
  - The data directly received from the source system
  - This is suitable for small organizations or smaller groups within an organization
- Hybrid Data Mart
  - The data is fed from both OLTP as well as from Data warehouse







Data Mart – Main Features/Advantages/Disadvantages

Data Mart 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

Data Mart Advantages over DWH

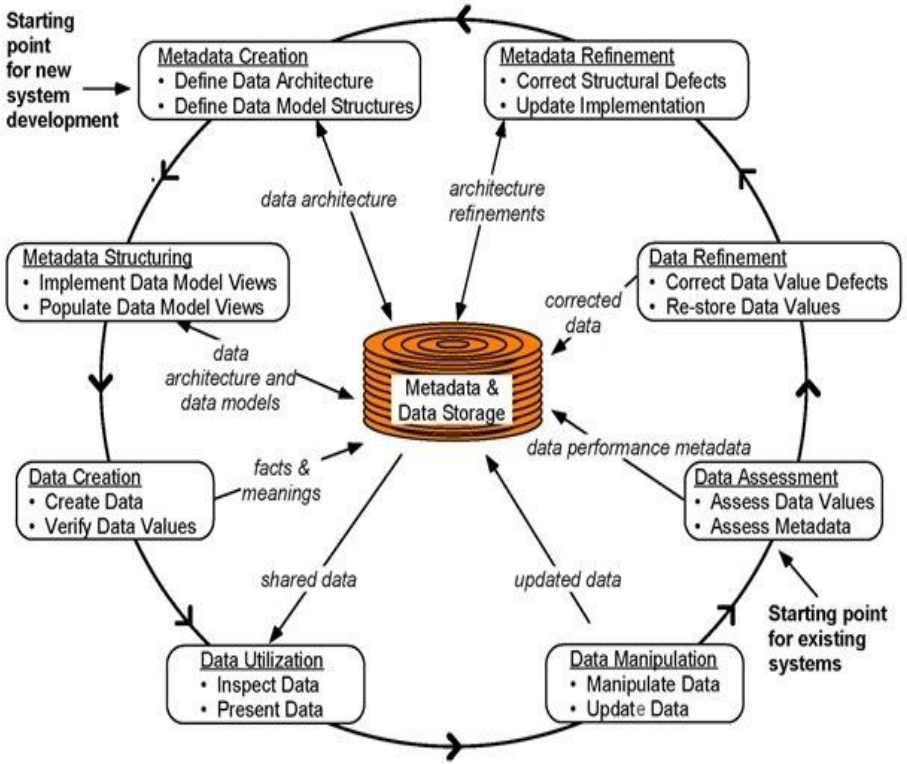
- Typically single subject area and fewer dimensions
- Limited feeds
- Very quick time to market (30-120 days to pilot)
- Quick impact on bottom line problems
- Focused user needs
- Limited scope
- Optimum model for DW construction
- Demonstrates ROI
- Allows prototyping

Data Mart Dis-advantages

- Does not provide integrated view of business information.
- Uncontrolled proliferation of data marts results in redundancy
- More number of data marts complex to maintain
- Scalability issues for large number of users and increased data volume

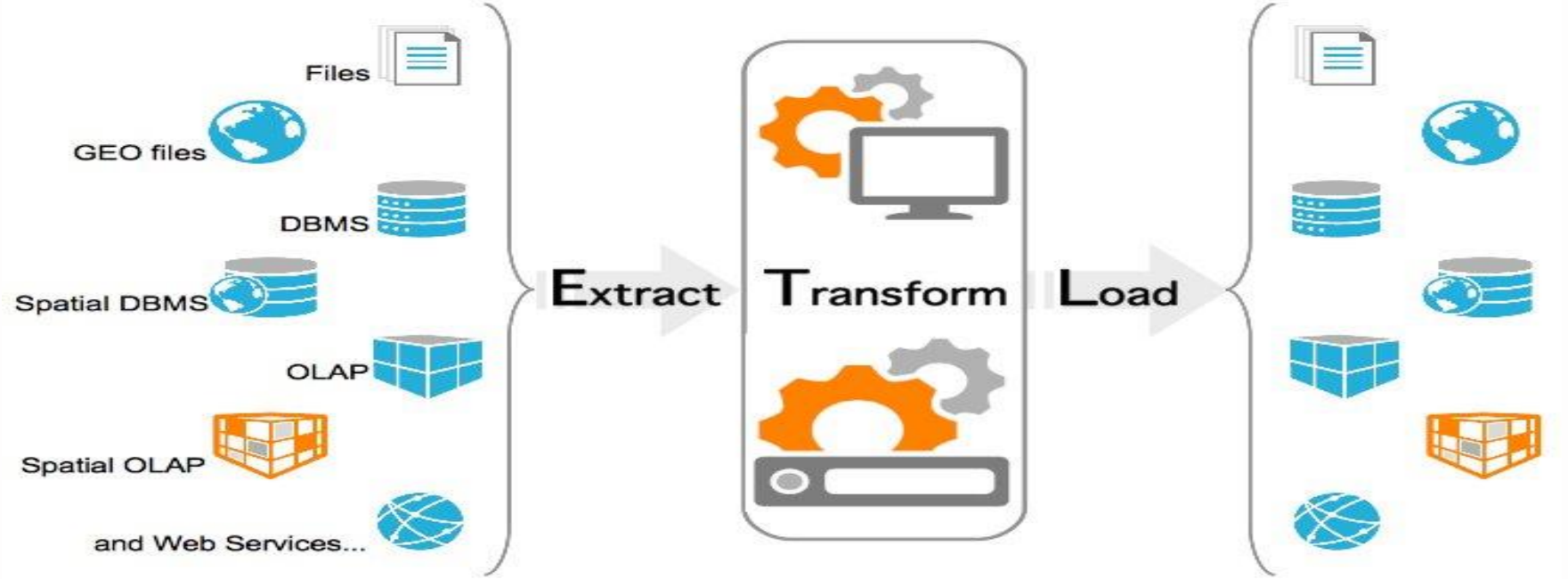
Meta Data

- Metadata means "data about data“
- Metadata in as DWH defines the Source data i.e. Flat files, Regional Data bases and other objects
- Metadata is defined as the data providing information about or or more aspects of the data
- it is used to summarize basic information about data which can make tracking and working with specific data easier
  - Some examples include:
    - Means of creation of the data
    - Purpose of the data
    - Time and date of creation
    - Creator or author of the data
    - Location on a computer network where the data was created
    - Standards used
    - File size

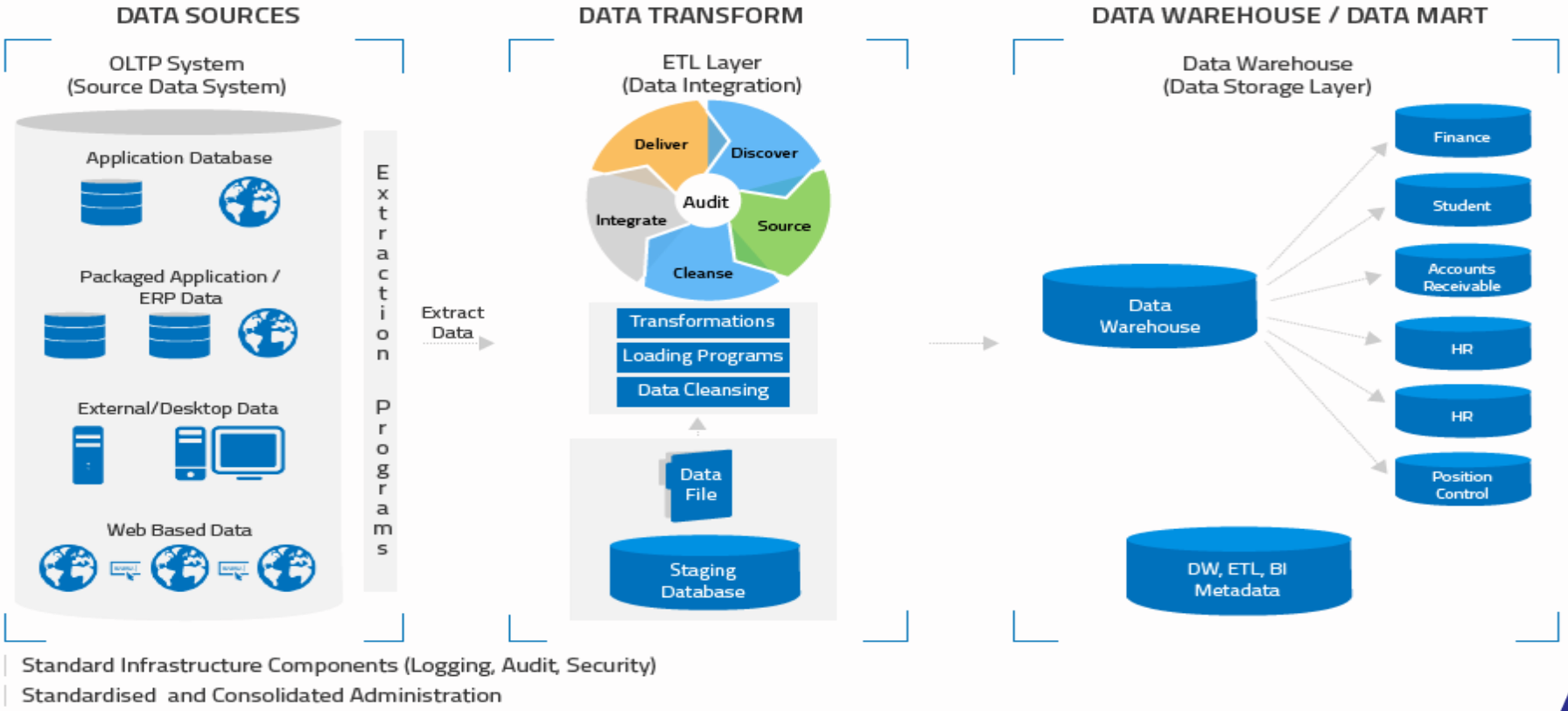


Extract Transform Load

- ETL is the process of Extracting the data from various sources, transforming this data to meet your business requirements and then loading it into a target data warehouse



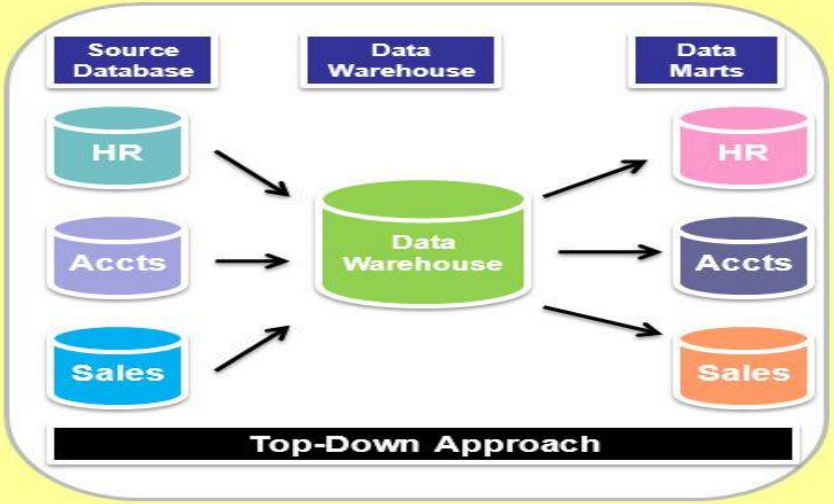
Data Warehouse Architecture

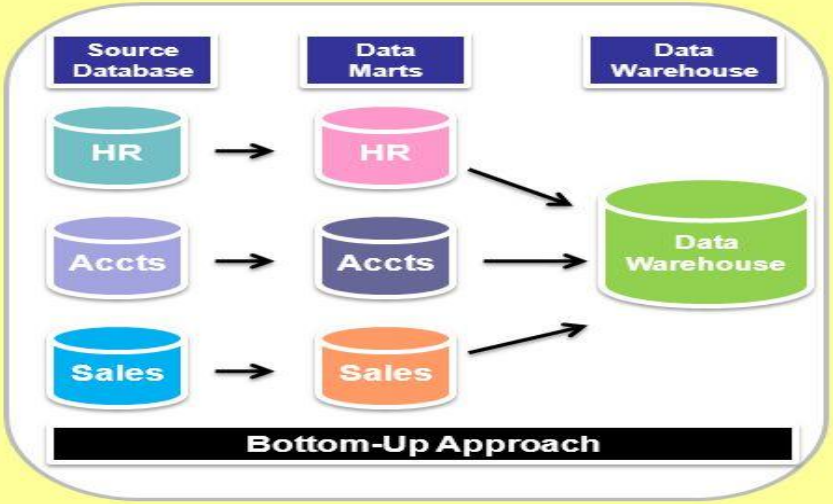


Approaches to Data warehouse designing

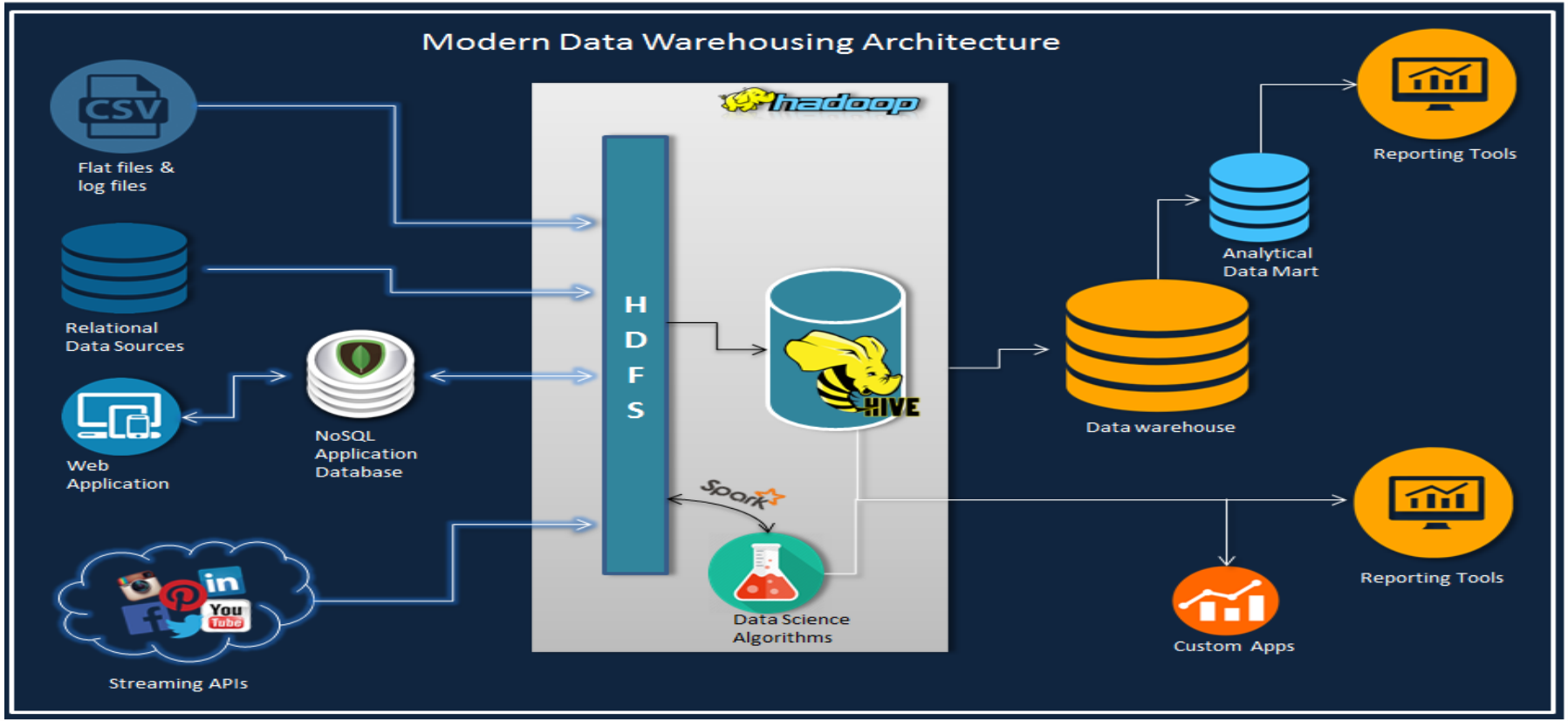
Approaches to Data Warehouse Designing

- **Top – Down Approach:** Build Enterprise-Wide data first and later build Data Marts from Data Warehouse. This type of approach takes long time to build and has a high risk of failure. It also requires experienced professionals to build this type of approach otherwise it could be dangerous
- **Bottom – Up Approach:** The priority is set to each data mart and the departmental marts are built one-by-one. The data is fragmented but it is faster and easier to build and the risk exposure to failure is less.





Modern Data warehouse Architecture



Let's Solve