VISVESVARAYA TECHNOLOGICAL UNIVERSITY

JNANA SANGAMA, BELGAVI-590018, KARNATAKA

DATA WAREHOUSING

(AS PER CBCS SCHEME 2022)

SUB CODE: BAD515B

PREPARED BY:

INDHUMATHI R (ASST.PROF DEPT OF DS (CSE), KNSIT)

DEPARTMENT OF COMPUTER SCIENCE (DATA SCIENCE) AND ENGINEERING

K.N.S INSTITUTE OF TECHNOLOGY

HEGDE-NAGAR, KOGILU ROAD,
THIRUMENAHALLI, YELAHANKA,
BANGALORE-5600

MODULE 1: INTRODUCTION TO DATA WAREHOUSING

CHAPTER 1: THE COMPELLING NEED FOR DATA WAREHOUSE

ESCALATING NEED FOR STRATEGIC INFORMATION

- **Strategic Information Need**:

Explanation: Companies require strategic information to make informed decisions that enhance competitiveness. This information helps in identifying market trends, customer preferences, and operational efficiencies.

Real-world Application: A retail chain analyzes customer purchase data to determine which products to promote during seasonal sales, thereby increasing revenue

- **Operational versus Strategic Information Systems**:

- Operational Systems:

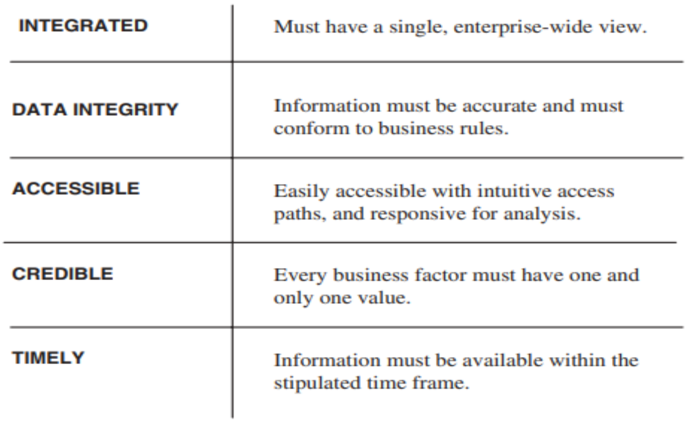
  Explanation: These systems manage day-to-day transactions and are designed for efficiency in processing current data. They focus on individual transactions and are often referred to as Online Transaction Processing (OLTP) systems.

  Real-world Application: An order processing system that tracks sales in real-time, ensuring that inventory levels are updated immediately after a sale.

- Strategic/Informational Systems:

Explanation: These systems are designed to support decision-making processes by providing historical and aggregated data. They focus on analysis rather than transaction processing.

Real-world Application: A business intelligence tool that analyzes sales data over the past year to identify trends and forecast future sales



INTEGRATED	Must have a single, enterprise-wide view.
DATA INTEGRITY	Information must be accurate and must conform to business rules.
ACCESSIBLE	Easily accessible with intuitive access paths, and responsive for analysis.
CREDIBLE	Every business factor must have one and only one value.
TIMELY	Information must be available within the stipulated time frame.

Figure 1-2 Characteristics of strategic information.

- **Information Crisis**:
- Definition: Organizations have accumulated vast amounts of data over the years, but they struggle to convert this data into actionable strategic information.
- Issue: The data is often stored in disparate systems that are incompatible, making it difficult to access and analyse for decision-making.
- Operational Data: While operational systems provide data for daily operations, they do not offer the comprehensive insights needed for strategic planning.
- **Technology Trends (1960-2010):**
- Computing Technology: The evolution from mainframes to mini-computers, PCs, and networking systems illustrates how technology has become more accessible and powerful.
- Human/Machine Interface: The transition from punch cards to graphical user interfaces (GUIs) and voice recognition shows the increasing user-friendliness of technology.
- Processing Options: The shift from batch processing to online and networked systems highlights the need for real-time data processing capabilities.
- **Opportunities and Risks**:
- Opportunities Examples:
    a) Telecom: Sales personnel can make informed decisions that enhance competitiveness.
    b) Banking: Quick access to strategic information helps retain valued customers.
    c) Healthcare: Improved programs lead to significant reductions in emergency visits and hospital admissions.
    d) Retail: Enhanced inventory management ensures products are available when needed.
    e) Pharmacy: Understanding customer purchasing patterns improves marketing effectiveness.
- Risks Examples:
    a) Car Rental: Poor fleet management can lead to financial losses.
    b) Manufacturing: Inconsistent data can hinder benchmarking and quality control.
    c) Utilities: Without strategic information, companies may fail to compete effectively in deregulated markets.

FAILURES OF PAST DECISION-SUPPORT SYSTEMS

- **History of Decision-Support Systems**:

1. *Ad hoc Reports*: Users request specific reports, leading to one-off solutions.
2. *Special Extract Programs*: IT anticipates requests and prepares extract programs to generate reports.
3. *Small Applications*: Simple applications are created for generating reports based on extracted data.
4. *Information Centers*: Centralized locations where users can request reports or view screens.
5. *Decision-Support Systems*: More sophisticated systems aimed at providing strategic insights.
6. *Executive Information Systems*: Designed for executives, but often limited in flexibility and usability.

Inability to Provide Information: Every one of the past attempts at providing strategic information to decision makers was unsatisfactory. Figure 1-4 depicts the inadequate attempts by IT to provide strategic information. As IT professionals, we are all familiar with the situation

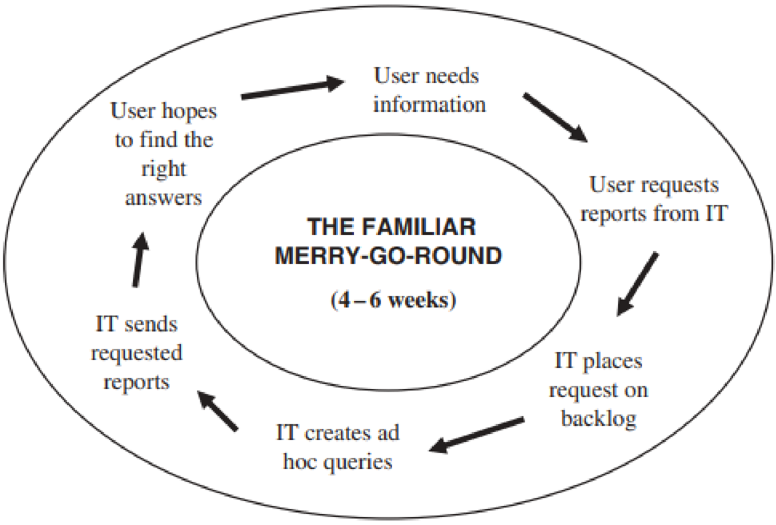

User needs information
User requests reports from IT
IT places request on backlog
IT creates ad hoc queries
IT sends requested reports
User hopes to find the right answers
THE FAMILIAR MERRY-GO-ROUND
(4 – 6 weeks)

Figure 1-4 Inadequate attempts by IT to provide strategic information.

Here are some of the factors relating to the inability to provide strategic information:

1) Ad hoc Requests: The volume of requests overwhelms IT departments, resulting in delays.
2) Changing Requirements: Users continually change their requests, complicating the reporting process.
3) Dependence on IT: Users lack the ability to access data independently, leading to bottlenecks.
4) Lack of Flexibility: The systems in place are not designed for the interactive analysis required for strategic decision-making.

OPERATIONAL VS DECISION-SUPPORT SYSTEMS

- Making the Wheels of Business Turn (Operational):

Definition: Operational systems (OLTP) are designed to manage day-to-day transactions and core business processes.

Examples: Processing orders, managing inventory, and handling customer transactions.

Focus: These systems prioritize data input and transaction processing, ensuring smooth operations.

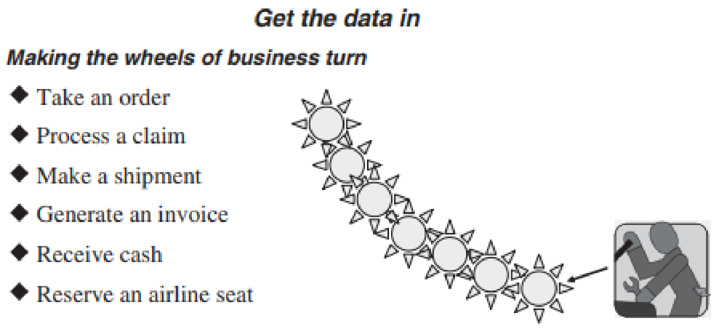


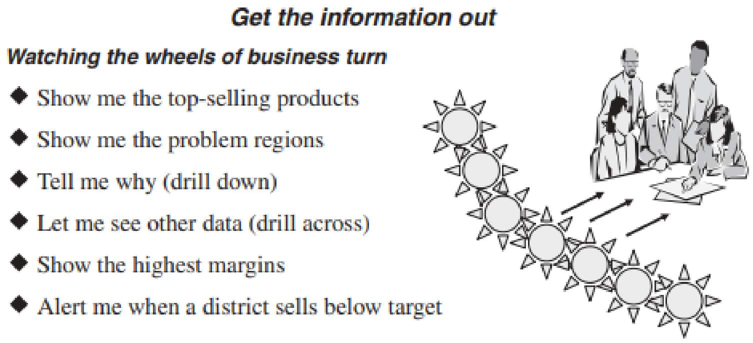
Figure 1-5 Operational systems.

- Watching the Wheels of Business Turn (Decision-Support):

Definition: Decision-support systems are designed to analyze business performance and provide insights for strategic decision-making.

Functions: Enable users to identify trends, analyze performance, and drill down into data for detailed insights.

Focus: These systems prioritize data output and analysis, allowing users to explore data interactively.

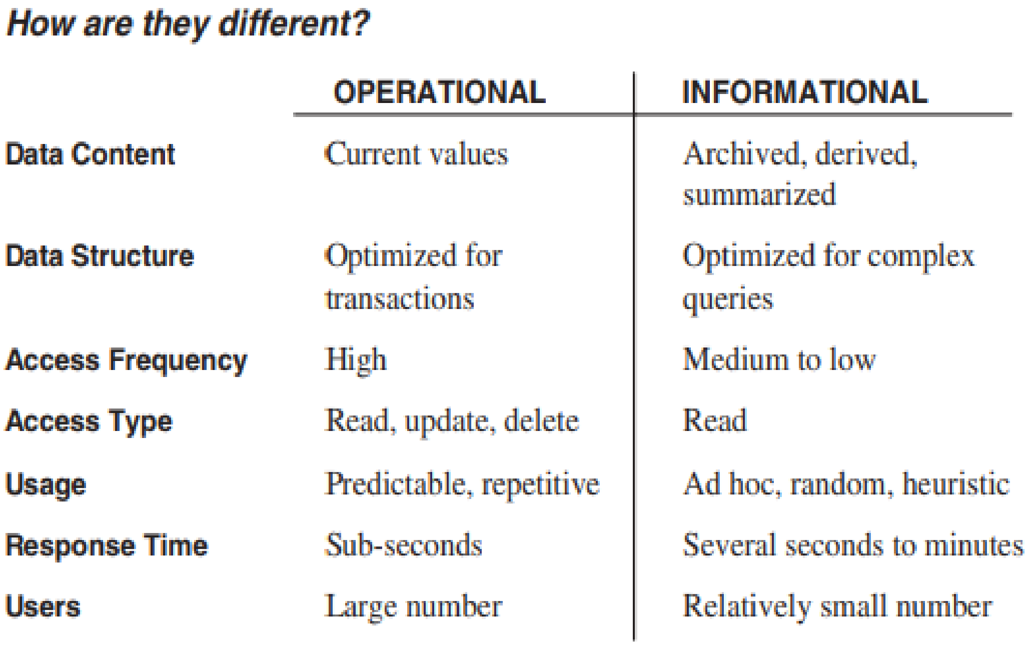


Get the information out

Watching the wheels of business turn

- Show me the top-selling products
- Show me the problem regions
- Tell me why (drill down)
- Let me see other data (drill across)
- Show the highest margins
- Alert me when a district sells below target

Figure 1-6 Decision-support systems.



How are they different?

	OPERATIONAL	INFORMATIONAL
Data Content	Current values	Archived, derived, summarized
Data Structure	Optimized for transactions	Optimized for complex queries
Access Frequency	High	Medium to low
Access Type	Read, update, delete	Read
Usage	Predictable, repetitive	Ad hoc, random, heuristic
Response Time	Sub-seconds	Several seconds to minutes
Users	Large number	Relatively small number

Figure 1-7 Operational and informational systems.

DATA WAREHOUSING—THE ONLY VIABLE SOLUTION

- **New System Environment Features**:

- Analytical Database: Designed specifically for analytical tasks rather than transaction processing.
- Multi-Application Integration: Combines data from various operational systems for a comprehensive view.
- User -Friendly Interface: Facilitates easy access and interaction by end-users.
- Read-Intensive Usage: Optimized for querying and analysis rather than frequent updates.
- User Interaction: Allows users to directly query the system without IT intervention.
- Periodic Updates: Data is updated at scheduled intervals rather than in real-time.
- Historical Data Inclusion: Maintains historical data for trend analysis and decision-making.

- **Processing Requirements in the New Environment:**

- Running of Simple Queries and Reports: Users should be able to execute straightforward queries and generate reports against both current and historical data. This facilitates quick access to information for decision-making.
- Ability to perform “What If” Analysis: Users should be able to conduct scenario analysis to explore various potential outcomes based on different data inputs. This capability allows decision-makers to evaluate the impact of different strategies.
- Ability to Query, Step Back, Analyze, and Continue: Users should have the ability to dig deeper into the data by querying, analyzing results, and adjusting their queries based on findings. This iterative process encourages thorough exploration of data.
- Ability to Spot Historical Trends: The system should enable users to identify trends over time and apply these insights to future decisions. This historical perspective is crucial for strategic planning and forecasting.

- **Strategic Information from the Data Warehouse**:
- The data warehouse serves as the source of strategic information for the enterprise, enabling informed decision-making.
- It is designed to support analytical tasks, providing a comprehensive view of the organization's data.

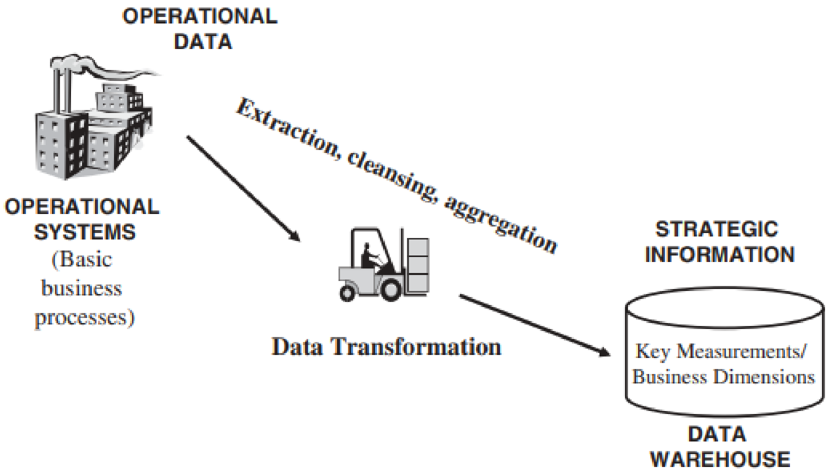

OPERATIONAL DATA
Extraction, cleansing, aggregation
OPERATIONAL SYSTEMS (Basic business processes)
Data Transformation
STRATEGIC INFORMATION
Key Measurements/ Business Dimensions
DATA WAREHOUSE

Figure 1-8 General overview of the data warehouse.

The process starts by extracting data from various operational systems. This data is then cleansed and transformed to address inconsistencies and to be formatted suitably for strategic analysis. For example, data from different systems might use different formats for dates or have duplicative customer records that need to be merged. Once cleansed and transformed, this data is loaded into the data warehouse.

>>***Key Components***:

- **Extraction, Transformation, Loading (ETL):** The process by which data is extracted from source systems, transformed into a consistent format, and then loaded into the data warehouse.
- **Data Modeling:** The design of the data warehouse structure that defines how data is interconnected.
- **Analysis and Reporting Tools:** These are used to query the data warehouse and generate reports for decision support.

>>**Benefits:**

- **Strategic Decision Making:** Enables detailed and strategic analysis that isn't feasible with operational systems.
- **Historical Analysis:** Provides capabilities to analyze data over different periods which help in identifying trends and patterns.
- **User Autonomy:** Reduces reliance on IT departments by empowering end-users to generate their own reports and analyze data directly.

➢ **Data Warehouse Defined**:

Functional Definition: The data warehouse is an informational environment that:

- Provides an integrated and total view of the enterprise.
- Makes the enterprise's current and historical information easily available for strategic decision-making.
- Allows decision-support transactions without hindering operational systems.
- Ensures consistency in the organization's information.
- Offers a flexible and interactive source of strategic information.

➢ **A Simple Concept for Information Delivery**:

- Data warehousing is fundamentally about transforming existing data into useful strategic information.
- It is not about generating new data but rather about leveraging the vast amounts of data already available within the organization

➢ **An Environment, Not a Product**:

  - A data warehouse is not a single software or hardware product but a comprehensive computing environment designed for data analysis and decision support.
  - It is user-centric, allowing users to find and analyze the strategic information they need.

A BLEND OF MANY TECHNOLOGIES:

The term "Blend of Technologies" in the context of data warehousing refers to the combination of various technological tools and methodologies that work together to handle the complex requirements of creating, managing, and utilizing a data warehouse. This blend involves several specific areas of technology, each contributing to the overall functionality and effectiveness of the data warehousing environment. Here's a breakdown of these technologies:

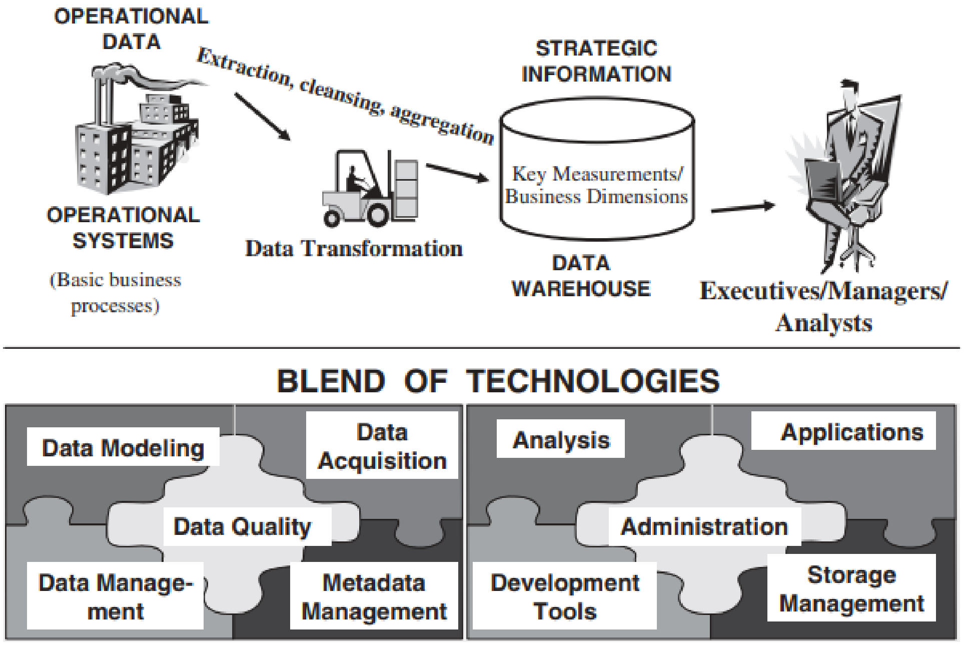


Figure 1-9 The data warehouse: a blend of technologies.

1. Data Modeling

Data modeling involves designing the data structures that will store the information in the data warehouse. This includes defining how tables are structured, how data elements interrelate, and ensuring that the design supports the analytical needs of the organization effectively.

2. Data Acquisition

This involves the processes and technologies used to extract data from various source systems (which can be both internal and external), and to load it into the data warehouse. This often includes integration tools that can handle different data formats and sources.

3. Data Quality

Maintaining data quality is critical in a data warehouse because it ensures the reliability and accuracy of reports and analyses derived from the warehouse. Technologies and processes here include validation, cleansing (removing or correcting data errors), and enrichment (enhancing data with additional sources).

4. Metadata Management

Metadata management involves handling data about the data in the warehouse, such as its source, how it's been transformed, and how it's structured. This makes it easier to manage the data warehouse and ensures that users can understand and trust the data they are working with.

5. Administration and Management

This includes the tools and processes used for the ongoing management of the data warehouse environment. Responsibilities include managing user access, monitoring performance, configuring hardware and software resources, and ensuring data security.

6. Storage Management

Efficient storage management is vital to handle the large volumes of data typically found in a data warehouse. This includes not only physical data storage but also the organization and indexing of data to optimize fast retrieval and efficient use of storage resources.

7. Development Tools

These are the tools used to build, maintain, and modify the data warehouse, including the interfaces for managing ETL processes, performing data modeling, and creating queries and reports.

8. Analysis Applications

After the data is stored and managed within the data warehouse, various analytical tools and applications are used to examine, manipulate, and report on the data. These tools support business intelligence activities like generating reports, conducting complex analyses, and providing data visualization.

THE DATA WAREHOUSING MOVEMENT

- **Adoption and Growth**:
  - As organizations began to recognize the effectiveness of data warehousing, the movement gained momentum.
  - Initially, large companies adopted data warehousing due to their resources, followed by medium-sized businesses.
- **Data Warehousing Milestones**:

Key historical milestones that marked the growth of data warehousing include:

- 1983: Teradata introduces a database management system for decision-support systems.
- 1991: Bill Inmon publishes "Building the Data Warehouse," establishing foundational concepts.
- 1995: The Data Warehousing Institute is founded, promoting education and research in the field.
- **Initial Challenges**:

Despite early successes, organizations faced challenges such as:

- Increased customer expectations for service and quality.
- The need for leaner operations due to deregulation and competition.
- Fragmented data views and inconsistent data across systems.

EVOLUTION OF BUSINESS INTELLIGENCE:

- The challenges faced in early data warehousing led to the evolution of business intelligence (BI).
- BI encompasses both the transformation of data into information and the derivation of knowledge from that information.
- BI is viewed as a combination of applications and technologies that improve business decision-making.

- **BI: Two Environments**:
- Data to Information: Focuses on gathering, cleansing, and storing corporate data.
- Information to Knowledge: Involves analyzing stored data to derive insights and knowledge.

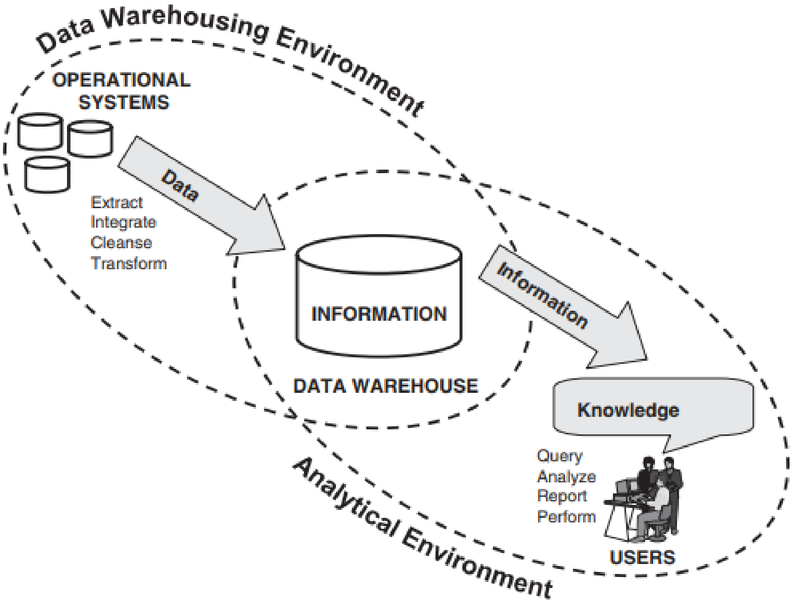


Figure 1-10 BI: data warehousing and analytical environments.

1. Data to Information Environment:

- **Purpose:** This environment focuses on taking raw data from multiple sources within a company, such as sales records, customer interactions, and financial transactions.
- **Process:**
  - **Extraction:** Data is pulled from various operational systems where it is generated.
  - **Integration:** Data from different sources is combined to provide a unified view.
  - **Cleansing:** The integrated data is cleaned to remove errors, duplicates, and inconsistencies.
  - **Transformation:** This cleaned data is then transformed into a format that’s easier to analyze.
  - **Storage:** Finally, this processed data is stored in specially designed data repositories, known as data warehouses.

Essentially, this environment turns raw operational data into organized and cleaned information that's ready for analysis.

2. Information to Knowledge Environment:

- **Purpose:** This environment takes the cleaned and organized information and uses analytical tools to turn it into actionable insights or knowledge.
- **Process:**
  - **Access:** Users can access the stored information using various BI tools.
  - **Analysis:** Through techniques like data visualization, statistical analysis, and predictive modeling, users analyze the information.
  - **Insight Generation:** The result of this analysis is knowledge that can inform decision-making, helping leaders and managers make informed strategic choices based on facts and trends observed in the data.

How They Work Together:

While these two environments function separately with specific tasks, they are complementary. The first environment ensures data is accurate and accessible, while the second empowers users to derive meaningful insights from this data. Together, they form a complete BI system that supports fact-based decision-making, improving the effectiveness and efficiency of business strategies.

Real-world Example:

Imagine a large online retailer:

- **Data to Information:** The retailer gathers data from their website traffic, customer purchase histories, and inventory levels, cleans and integrates this data, and stores it in a data warehouse.
- **Information to Knowledge:** Marketing analysts then use BI tools to study this information, identifying shopping trends, forecasting demand for products, and crafting targeted marketing campaigns.

CHAPTER 2: DATA WAREHOUSE THE BUILDING BLOCK

DEFINING THE DATA WAREHOUSE

Formal Definitions:

- Bill Inmon's Definition: "A Data Warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management's decisions."
  Subject-Oriented: Data is organized around key subjects relevant to the business rather than specific applications.

1. Integrated: Data is collected from various sources and made consistent.
2. Nonvolatile: Data is stable and not frequently changed; updates occur at scheduled intervals.
3. Time-Variant: Data is stored over time, allowing for historical analysis.

- Sean Kelly's Definition: The data in the data warehouse is:
  Separate: Distinct from operational systems.

1. Available: Accessible for analysis.
2. Integrated: Combined from various sources.
3. Time Stamped: Contains time-related data for historical analysis.
4. Subject-Oriented: Focuses on business subjects.
5. Nonvolatile: Data does not change frequently.
6. Accessible: Easy for users to retrieve and analyze.

DEFINING FEATURES OF DATA WAREHOUSES:

1. **Subject-Oriented Data:**
   - **Operational Systems:** In these systems, data is organized around specific applications like order processing, customer billing, or stock checks. Each application has its own data sets designed to support its specific functions efficiently.
   - **Data Warehouse:** Contrary to operational systems, data in a data warehouse is organized by real-world business subjects or events, not by the applications. For example, all data related to "claims" in an insurance company would be linked

and stored together, regardless of whether it comes from auto insurance or workers' compensation insurance.

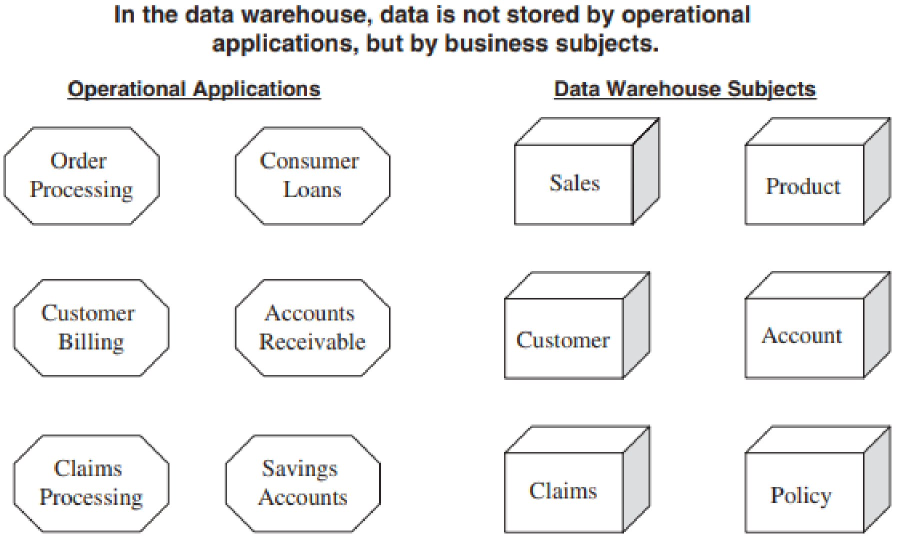


Figure 2-1 The data warehouse is subject oriented.

INTEGRATED DATA:

- **Integration Process:** Since operational data comes from various sources with different formats, data integration is a crucial process in data warehousing. This process involves:
  - **Standardization:** Aligning data formats, naming conventions, and measurement units.
  - **Transformation:** Converting data into a format suitable for analysis and storage in the data warehouse.
  - **Consolidation:** Merging data from different sources to create a unified set of data

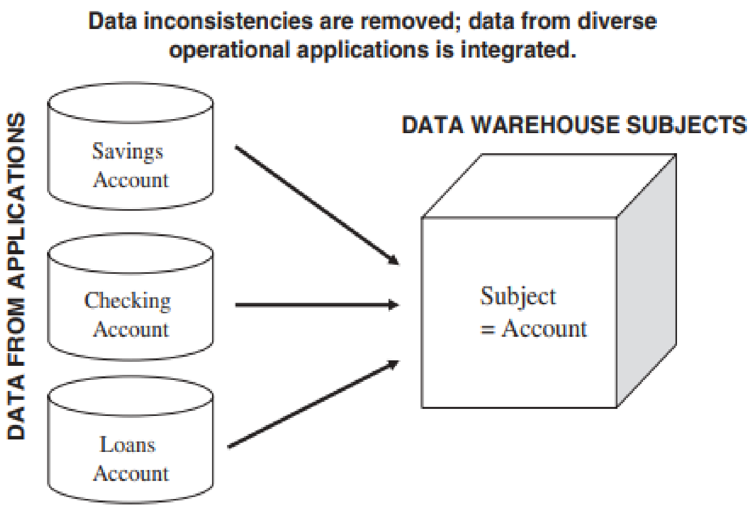


Figure 2-2 The data warehouse is integrated.

TIME-VARIANT DATA

- **Definition:** In a data warehouse, data is stored to reflect changes over time, not just the current snapshot. This means the data warehouse contains historical data alongside the most recent data, allowing for trend analysis and forecasting.
- Operational Systems: Primarily store current values (e.g., current account balance).
- Data Warehouse: Stores historical data, enabling trend analysis and forecasting.
- Importance: Historical snapshots allow users to analyze past performance and make informed predictions about the future.

NONVOLATILE DATA

- **Definition:** Once data is stored in a data warehouse, it doesn’t change as a result of operational system updates. The data warehouse is updated in batches at scheduled intervals, not continuously. It is primarily used for analysis and querying rather than transaction processing.
- Operational Systems: Data is frequently updated in real-time based on transactions.
- Data Warehouse: Data is updated at specific intervals (e.g., daily, weekly) and is not altered with each transaction.
- Purpose: The nonvolatile nature allows for stable data analysis and reporting without interference from daily operations.

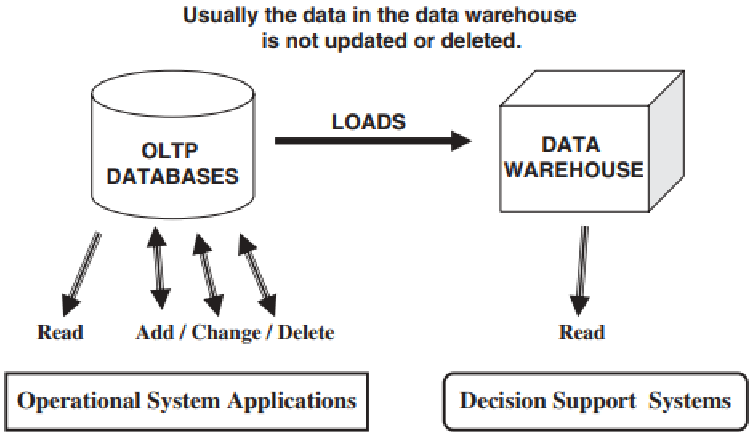
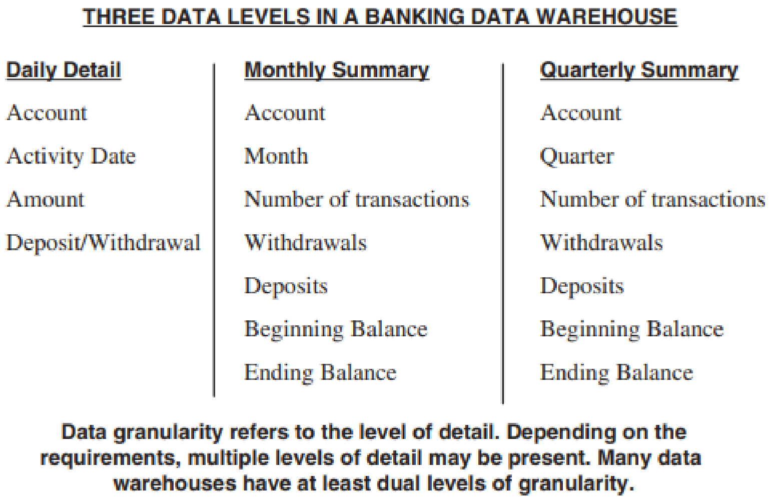

Usually the data in the data warehouse is not updated or deleted.
OLTP DATABASES
LOADS
DATA WAREHOUSE
Read
Add / Change / Delete
Read
Operational System Applications
Decision Support Systems

Figure 2-3 The data warehouse is nonvolatile.

DATA GRANULARITY

- **Definition:** Refers to the level of detail or fineness of data stored in the data warehouse. Granularity can vary from highly detailed data (e.g., transactions per second) to summarized data (e.g., monthly sales totals).
- Operational Systems: Typically store data at the lowest level of detail (e.g., individual transactions).
- Data Warehouse: Maintains data at multiple levels of granularity, including summary data for efficient querying.



THREE DATA LEVELS IN A BANKING DATA WAREHOUSE

Daily Detail	Monthly Summary	Quarterly Summary
Account	Account	Account
Activity Date	Month	Quarter
Amount	Number of transactions	Number of transactions
Deposit/Withdrawal	Withdrawals	Withdrawals
	Deposits	Deposits
	Beginning Balance	Beginning Balance
	Ending Balance	Ending Balance

Data granularity refers to the level of detail. Depending on the requirements, multiple levels of detail may be present. Many data warehouses have at least dual levels of granularity.

Figure 2-4 Data granularity.

DATA WAREHOUSES AND DATA MARTS:

- **Definitions**:
  - *Data Warehouse*: A centralized repository that stores data from across the organization, providing a comprehensive view for analysis.
  - *Data Mart*: A smaller, departmental subset of a data warehouse, focused on specific business areas or subjects.

- **Differences**:
- Scope: Data warehouses serve the entire organization, while data marts cater to specific departments or functions.
- Data Integration: Data warehouses integrate data from various sources, while data marts may pull data from the warehouse or specific operational systems.

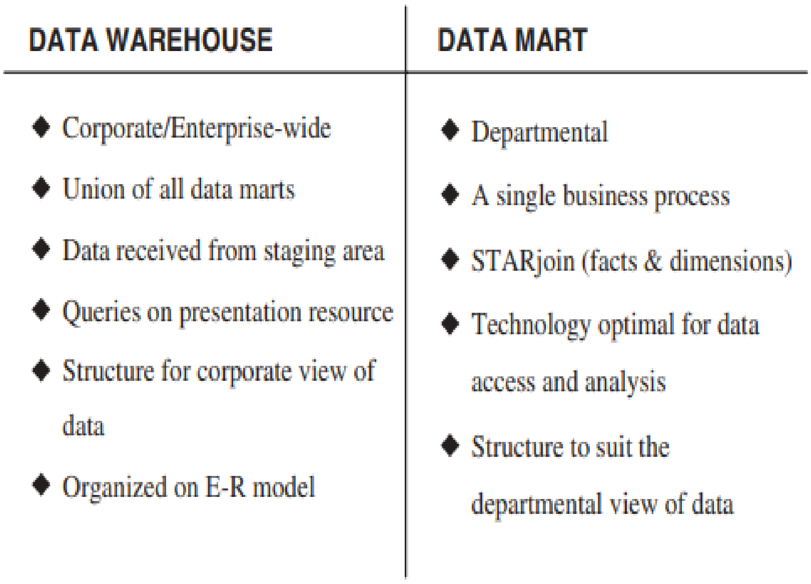
- **Approaches**:

1. *Top-Down Approach (Data Warehouse First)*

- **Scenario**: The platform decides to build a comprehensive data warehouse that integrates data from all user interactions, posts, and transactions across the platform.
- **Purpose**: To have a unified view of data that can aid in strategic decision-making, such as new feature development based on trending content types or user feedback.
- **Real-Time Use**: The platform uses the data warehouse to analyze trends, such as which types of content are most engaging or what time users are most active, to tailor the user experience and improve engagement.

2. *Bottom-Up Approach (Data Marts First)*

- **Scenario**: The platform begins by developing specific data marts for departments like advertising and user engagement.
- **Purpose**: Each department gets fast, actionable insights relevant to its specific needs. For example, the advertising department analyzes the performance of different ad formats and campaigns.
- **Real-Time Use**: The advertising team quickly adjusts campaigns based on real-time data on ad performance, optimizing for better user engagement and ROI.



DATA WAREHOUSE	DATA MART
◆ Corporate/Enterprise-wide	◆ Departmental
◆ Union of all data marts	◆ A single business process
◆ Data received from staging area	◆ STARjoin (facts & dimensions)
◆ Queries on presentation resource	◆ Technology optimal for data access and analysis
◆ Structure for corporate view of data	◆ Structure to suit the departmental view of data
◆ Organized on E-R model	

Figure 2-5 Data warehouse versus data mart.

- **Practical Approach**:

Combination of Both: A balanced strategy that considers the enterprise-wide view while implementing data marts based on priority needs.

Steps:

- Plan and define requirements at the corporate level.
- Establish an overall architecture for the data warehouse.
- Standardize and conform data across the organization.
- Implement the data warehouse incrementally through supermarts.

ARCHITECTURAL TYPES

- Overview of Architectural Types:
  1) *Centralized Data Warehouse*: A single, comprehensive repository that serves the entire organization. All user data is stored centrally. Queries for strategic analysis, such as determining the effectiveness of new features across all users, are run here.
  2) *Independent Data Marts*: Separate data marts that operate independently of the centralized warehouse. Each department, say, content moderation and

marketing, operates its own data mart with data relevant only to its functions. Might lead to inconsistencies but allows quick, department-specific analytics

3) *Federated Architecture*: A combination of independent data marts and a centralized warehouse that allows for data sharing and integration.
4) *Hub-and-Spoke Architecture*: A central data warehouse (hub) that feeds data to dependent data marts (spokes).A central data warehouse acts as a 'hub' with 'spokes' (dependent data marts) for different departments. Ensures consistency and supports both detailed departmental analysis and broad strategic analysis.
5) *Data-Mart Bus Architecture*: A centralized data warehouse that serves as a bus, allowing data to be shared among various data marts. Starts with specific business needs (e.g., analyzing user engagement on posts) and builds out data marts that conform in dimensions like user demographics and engagement metrics. Over time, these conformed data marts integrate to provide a comprehensive view of all aspects of the platform's operations.

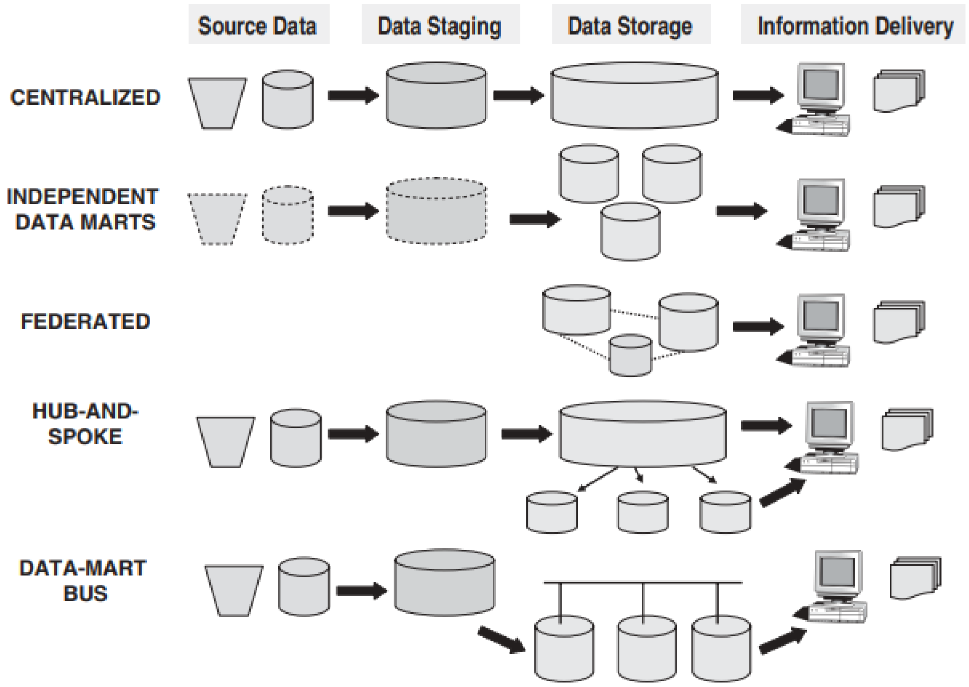


Figure 2-6 Data warehouse architectural types.

COMPONENTS OF A DATA WAREHOUSE

- **Source Data Component**: The operational systems and external data sources that provide data to the data warehouse.

  This refers to the data coming from different sources into the warehouse. It can be internal (within the organization) or external (from third-party sources).

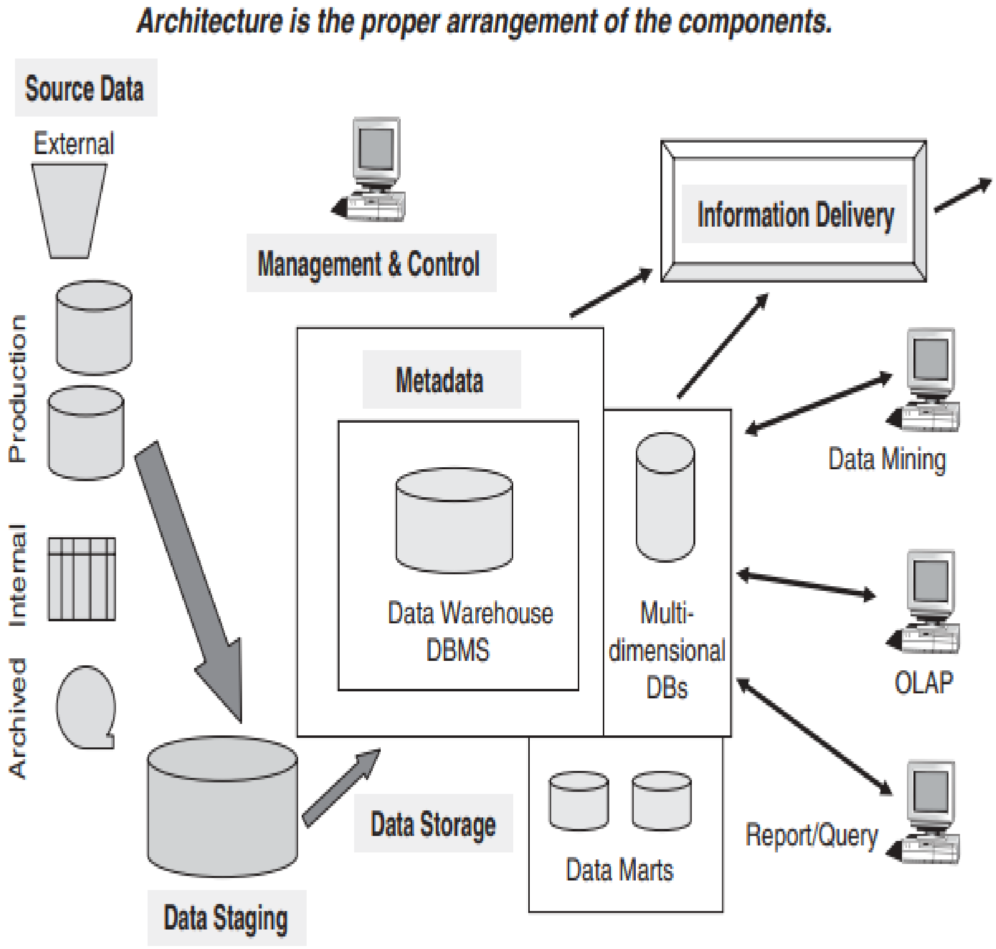


Figure 2-7 Data warehouse: building blocks or components.

- Production Data: This category of data comes from the various operational systems of the enterprise. These normally include financial systems, manufacturing systems, systems along the supply chain, and customer relationship management systems. Based on the infor mation requirements in the data warehouse, you choose segments of data from the different operational systems
- Internal Data: In every organization, users keep their "private" spreadsheets, documents, customer profiles, and sometimes even departmental databases. This is the internal data, parts of which could be useful in a data warehouse
- Archived Data: Operational systems are primarily intended to run the current business. In every operational system, you periodically take the old data and store it in archived files
- External Data: Most executives depend on data from external sources for a high percentage of the information they use. They use statistics relating to their industry produced by external agencies and national statistical offices. They use market share data of competitors. They use standard values of financial indicators for their business to check on their performance.

❖ **Data Staging Component**: The area where data is cleansed, transformed, and prepared for loading into the data warehouse. Data staging is a step in the data warehousing process where data from various sources is prepared before it's loaded into the warehouse. It involves cleaning, transforming, and integrating the data to make it ready for analysis.

- Data Extraction: This is the process of pulling data from various sources, which may be in different formats. It is a necessary first step to prepare data for transformation and loading into the warehouse.
- Data Transformation: Once data is extracted, it needs to be transformed into a standard format to ensure consistency. This step involves cleaning, organizing, and formatting the data.

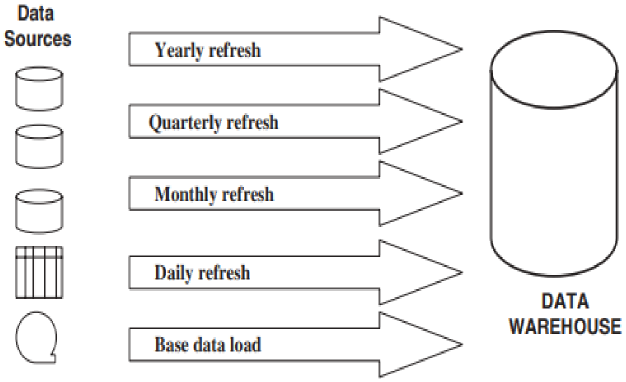


Figure 2-8 Data movements to the data warehouse.

- Data Loading: After the data is transformed, it's loaded into the data warehouse where it will be stored and made available for querying and analysis.

❖ **Data Storage Component**: The repository that stores the integrated and transformed data.

❖ **Information Delivery Component**: The tools and interfaces that enable users to access and analyze the data. The information delivery component includes reports, OLAP (Online Analytical Processing) tools, and dashboards that let users query and visualize the data

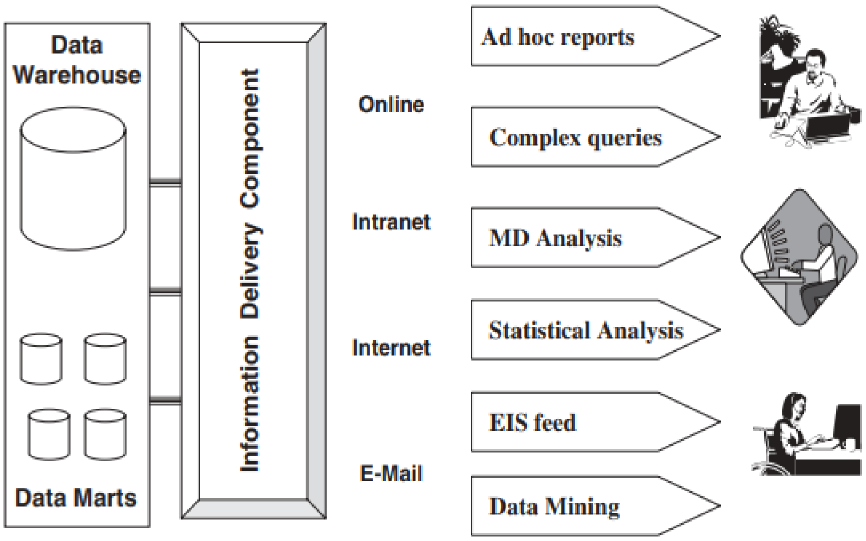


Figure 2-9 Information delivery component.

- **Metadata Component**: The data that describes the structure, content, and context of the data warehouse.
- **Management and Control Component**: The processes and tools that manage the data warehouse, ensuring data quality, security, and performance.

METADATA IN THE DATA WAREHOUSE

- Importance of Metadata:
- Definition: Metadata is data that describes the structure, content, and context of the data warehouse.
- Types of Metadata:
    - Technical metadata (e.g., data formats, storage locations) and
    - Business metadata (e.g., data definitions, business rules).
- Significance: Metadata is crucial for data discovery, data quality, and data governance, enabling users to understand and trust the data.