Creating a Corporate Data Warehouse — 1

• Steps to Build a Data Warehouse:
1. Business Requirement Analysis: Identify what kind of data is necessary for decision-making.
2. Data Collection: Source data from multiple systems such as operational databases, third-party systems, and external data sources.
3. Data Cleansing & Integration: Data is cleansed to remove inconsistencies and duplicates, then integrated to form a unified dataset.
4. Data Modeling: The data is organized into schemas (e.g., star schema or snowflake schema) for easy retrieval.
5. ETL Process: (Extract, Transform, Load) enables the transfer of data into the data warehouse.
6. User Access & BI Tools: BI tools (like Tableau, Power BI) are connected for reporting, analysis, and visualization.

Tools for Data Warehousing — 2

• Popular Tools:
◦ Microsoft SQL Server: Provides integrated tools for DW and BI.
◦ Oracle: Offers comprehensive features like data integration, analytics, and in-memory processing.
◦ Amazon Redshift: A cloud-based data warehouse solution offering scalability and quick access to data.
◦ Google BigQuery: A fully managed, serverless data warehouse that allows large-scale data analysis.

Data Mining — 3

• Definition: The process of extracting meaningful patterns, trends, and insights from large datasets using statistical and computational methods.
• Use: Data mining supports decision-making processes by identifying customer behaviors, predicting trends, and detecting anomalies.

KDD Process (Knowledge Discovery in Databases)
1. Data Selection: Identifying and retrieving relevant data from large datasets.
2. Data Preprocessing: Cleaning, transforming, and preparing the data for mining (handling missing values, noise).
3. Data Transformation: Converting data into a form suitable for mining (e.g., normalization, aggregation).
4. Data Mining: Applying algorithms to extract patterns and knowledge (e.g., classification, clustering).
5. Evaluation: Interpreting and validating the discovered patterns to assess their usefulness.
6. Knowledge Presentation: Presenting the insights to decision-makers through reports or dashboards.

Enterprise Information Management (EIM) — 4

• Definition: EIM is a set of practices and tools used to manage enterprise data in a secure, consistent, and scalable manner. It ensures that data across the organization is accurate, accessible, and aligned with business objectives.
• Key Components:
◦ Data Governance: Policies and standards for managing data assets.
◦ Master Data Management (MDM): Ensuring consistency and accuracy of key data entities across the enterprise.
◦ Data Quality Management: Ongoing processes to improve data accuracy and integrity.