Enterprise Technologies and Big Data Business Intelligence

Chapter: 04
Instructor: Houssein Dhayne
houssein.dhayne@net.usj.edu.lb

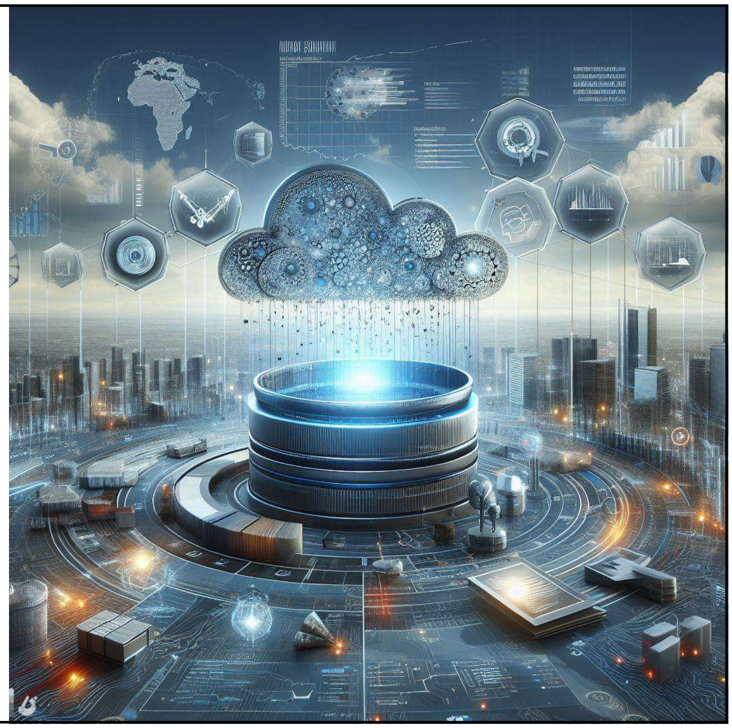


Table of content

Introduction to Enterprise Technologies and BI

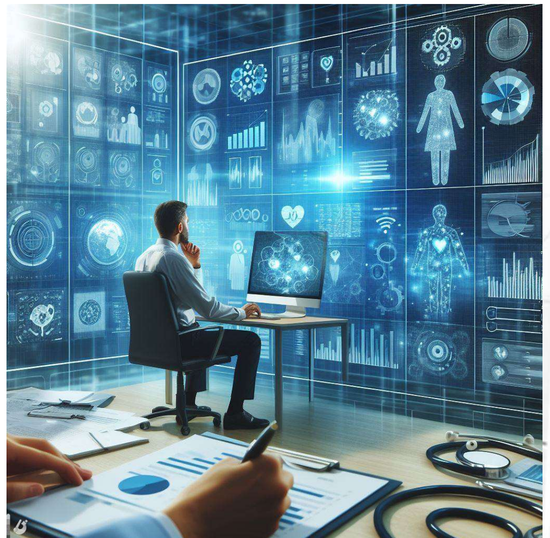
- In a business structure, there are different layers of management:
  - The strategic layer sets the overall direction,
  - the tactical layer implements the strategy,
  - and the operational layer carries out day-to-day tasks.
- Metrics and performance indicators are used to align these layers, providing insights into how processes are running.

Layered System Dynamics

- This chapter explores technologies that facilitate the transformation of data into meaningful insights.
  - Data is initially stored in the operational-level information systems of an organization.
  - Moving up the analytical hierarchy, analytical processing systems come into play. These systems employ multidimensional structures to handle more complex queries, offering deeper insights into business operations.
  - At a broader scale, data is gathered from various parts of the enterprise and stored in a centralized data warehouse.
- Management derives insights into overall corporate performance and Key Performance Indicators (KPIs) from these data repositories.

Scenario

- Imagine you're overseeing the implementation of Big Data Analytics in a large healthcare organization. The primary goal is to:
- enhance patient care,
- optimize operations,
- contribute to medical research.
- Let's explore how the concepts from Chapter 4 are applied in this transformative journey.

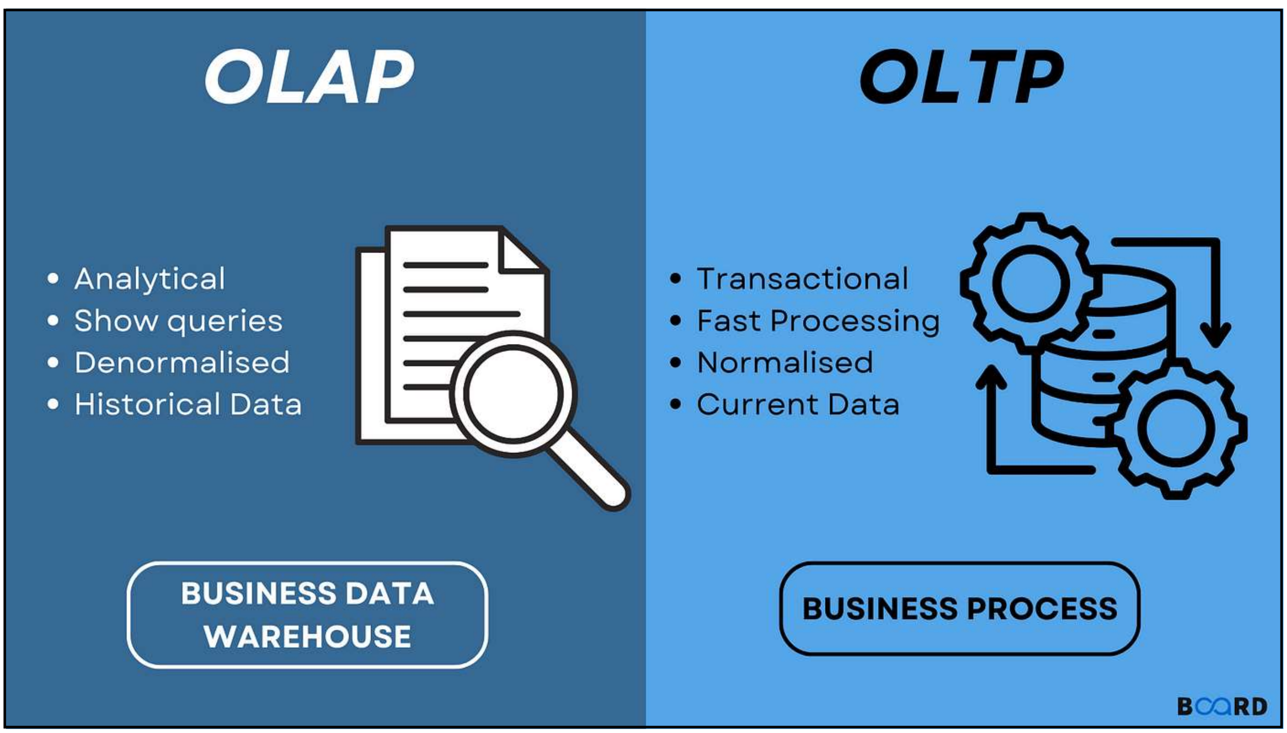


OLTP (Online Transaction Processing)

- OLTP, or Online Transaction Processing, is a critical software system.
- It excels in real-time processing of transaction-oriented data.
- Operational data is normalized and serves as a common source for structured data, feeding analytic processes.
  - The healthcare system uses OLTP to manage real-time patient interactions.
  - For instance, when a patient is admitted, the system records vital signs, treatment details, and billing information instantly, ensuring swift and accurate data processing.

OLAP (Online Analytical Processing)

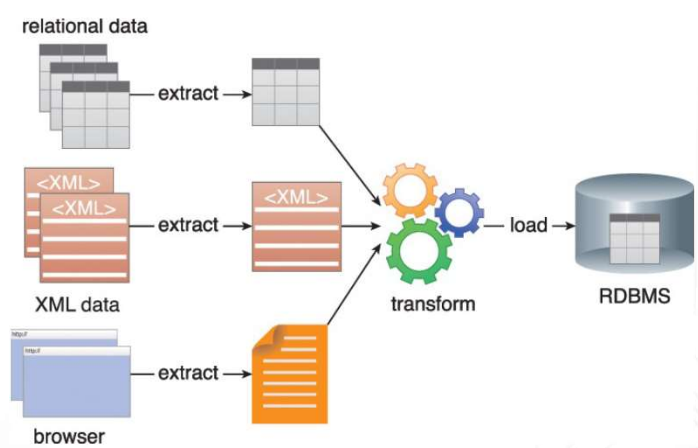
- Online Analytical Processing (OLAP) systems are pivotal in BI, data mining, and machine learning.
- Unlike OLTP, OLAP handles complex data analysis queries, offering deeper insights into business operations.
- Utilizes multidimensional databases, enabling the system to answer intricate queries and provide comprehensive analytics.
- Historical data is aggregated and denormalized to support fast reporting capabilities, forming a cornerstone of BI activities.
  - To analyze historical patient data for trends and patterns, the organization employs OLAP. This includes identifying correlations between specific treatments and outcomes over time, providing valuable insights into patient care effectiveness.



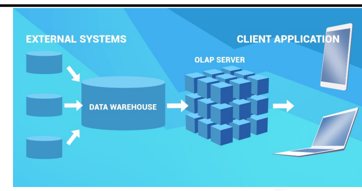
ETL Process and Data Warehouses

- Extract, Transform, Load (ETL) is a pivotal process in data management.
- ETL operations facilitate the movement of data from source to target systems, a core function in data warehousing.
- The process involves extraction from various sources, application of transformation rules, and loading into the target system.
- ETL is fundamental for feeding data warehouses, enabling a structured and consolidated repository of historical and current data.
  - Data from various sources, such as Electronic Health Records (EHR), medical imaging systems, and billing databases, undergoes ETL processes daily. This ensures that diverse data types are integrated into a cohesive format, ready for advanced analytics.

An ETL process can extract data from multiple sources and transform it for loading into a single target system.

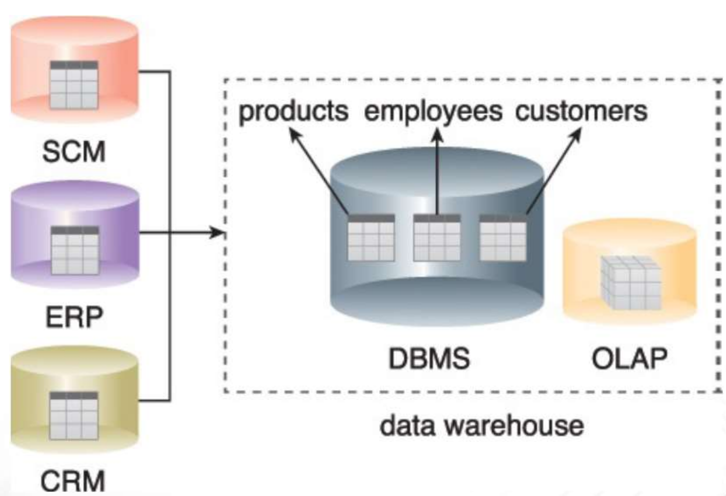


Data Warehouses and Analytical Databases

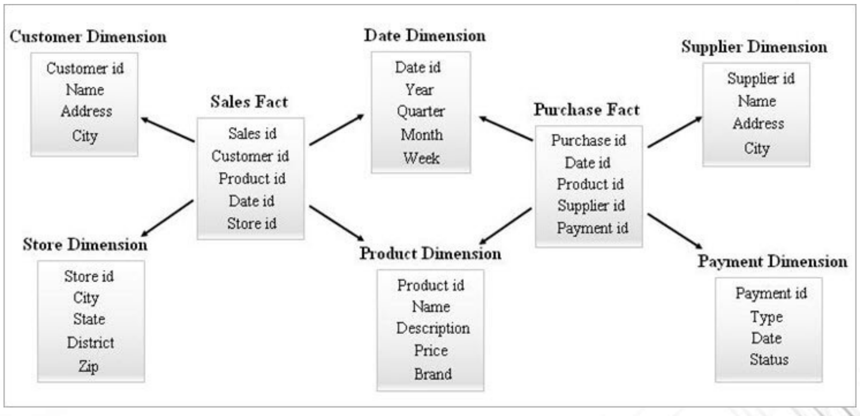


- A data warehouse is a central, enterprise-wide repository consisting of **historical** and current data. **Data related to multiple business entities from different operational systems** is periodically extracted, validated, transformed and consolidated into a single denormalized database.
- With periodic data imports from across the enterprise, the amount of data contained in a given data warehouse will continue to increase.
  - Over time this leads to slower query response times for data analysis tasks.
  - To resolve this shortcoming, data warehouses usually contain optimized databases, called analytical databases, to handle reporting and data analysis tasks. An analytical database can exist as a separate DBMS, as in the case of an OLAP database.
  - Example: A centralized data warehouse stores aggregated and validated information from across the healthcare enterprise. This includes data from operational systems like ERP, CRM, and SCM. Periodic batch jobs load data into the warehouse, forming the foundation for in-depth analysis.

Batch jobs periodically load data into a data warehouse from operational systems like ERP, CRM and SCM.

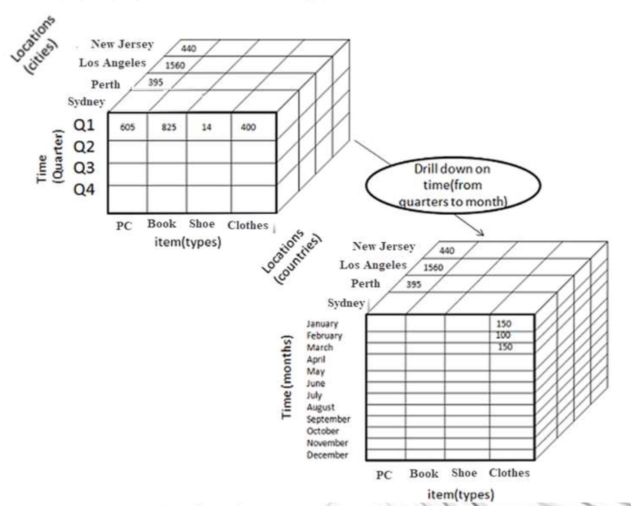


Data Warehouses - example



Data Warehouses - OLAP cube

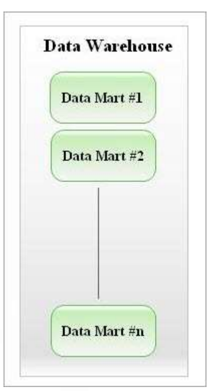
- Analysts frequently need to group, aggregate and join data. These OLAP operations in data mining are resource intensive. With OLAP data can be pre-calculated and pre-aggregated, making analysis faster.
- OLAP databases are divided into one or more cubes. The cubes are designed in such a way that creating and viewing reports become easy.

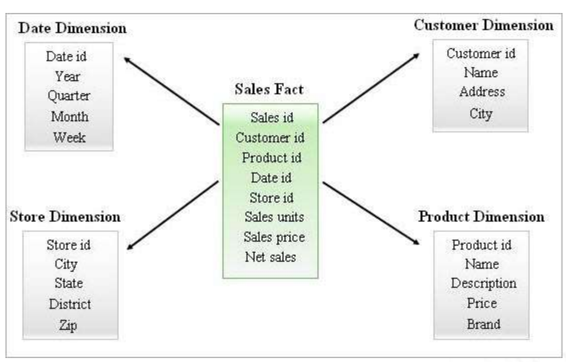


Data Marts - Specialized Data Views

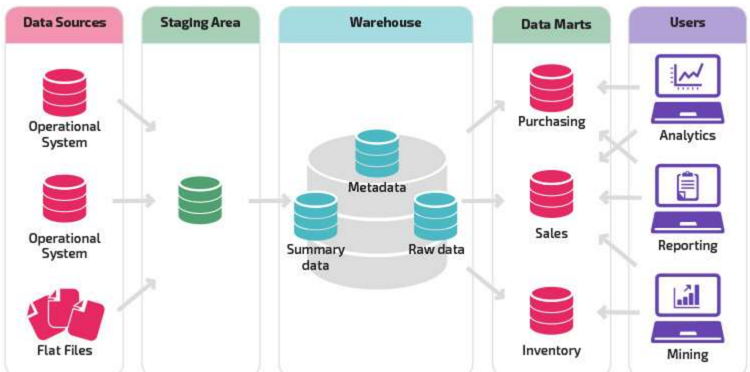
- Data marts, subsets of data warehouses, cater to specific departments, divisions, or lines of business.
- Extraction, transformation, and loading (ETL) processes persist domain-specific entities into data marts.
- Offers specialized views, enhancing the effectiveness of BI by providing tailored insights.
  - Specialized data marts are created to focus on specific medical domains. For instance, a cardiology data mart might contain detailed information on cardiac patients, allowing for targeted analysis and research within that department.

Data Marts - example





From Data Warehouse to Data Mart



Traditional BI - Descriptive Analytics

- Traditional BI primarily employs descriptive and diagnostic analytics.
- It offers insights into historical and current events, answering well-formulated questions.
- Utilizes ad-hoc reports and dashboards to communicate information.
- Ad-hoc reporting involves manual processing for custom reports, while dashboards provide graphical, periodic views of key business areas.
  - Dashboards provide management with key metrics, including top diagnoses, procedures, and readmission rates over time. This traditional BI approach aids in monitoring historical and current events in healthcare operations.

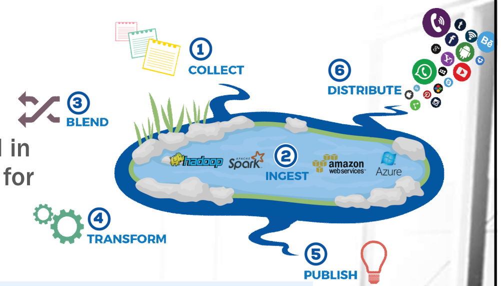
Big Data BI - Beyond Traditional Boundaries

- Big Data BI builds upon traditional BI by incorporating cleansed, consolidated enterprise-wide data.
- Combines structured, semi-structured, and unstructured data for predictive and prescriptive analytics.
- Analyzes multiple business processes simultaneously, uncovering patterns and anomalies.
- Requires a next-generation data warehouse, acting as a unified repository for various data formats, eliminating the need for multiple data source connections.
  - Building upon traditional BI, the organization incorporates Big Data BI to integrate machine data from IoT devices and genomic information. This enables predictive and prescriptive analytics to identify patients who may benefit from personalized treatments.

Next-Generation Data Warehouses

- Next-generation data warehouses are essential for Big Data BI, acting as a standardized data access layer.
- Incorporate new features and technologies to store cleansed data from various sources uniformly.
- Create a hybrid data warehouse, centralizing structured, semi-structured, and unstructured data.
- Eliminate the need for BI tools to connect to multiple sources, streamlining data access.

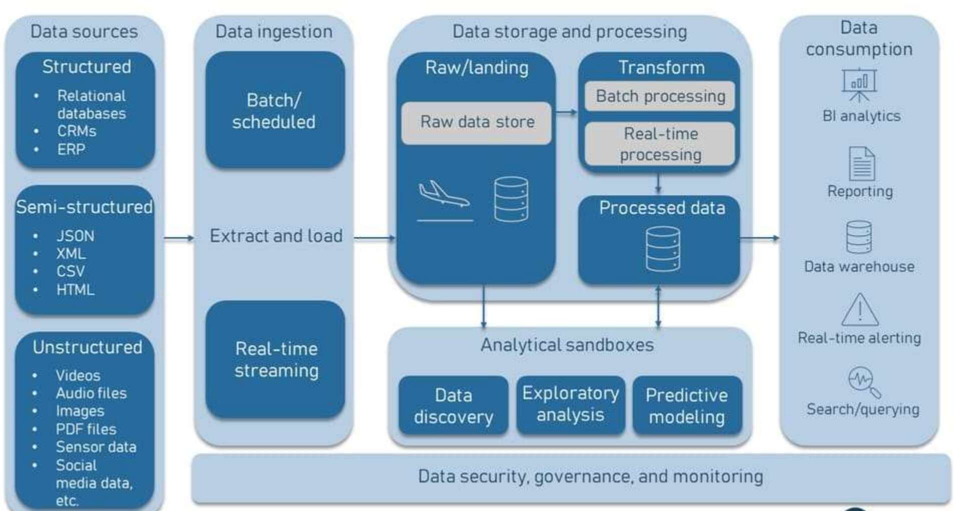
Data lake



- A data lake is a vast pool of raw data that can be stored in various formats, while a data warehouse is a repository for structured, filtered data that has been processed for a specific purpose. Data lakes are more flexible and can handle both structured and unstructured data.

	Data lake	Data warehouse
Type	Structured, semi-structured, unstructured	Structured
	Relational, non-relational	Relational
Schema	Schema on read	Schema on write
Format	Raw, unfiltered	Processed, checked
Sources	Big data, IoT, social media, streaming data	Application, business, transactional data, batch reporting
Scalability	Easy to scale at a low cost	Difficult and expensive to scale
Users	Data scientists, data engineers	Data warehouse professionals, business analysts
Use cases	Machine learning, predictive analytics, real-time analytics	Core reporting, BI

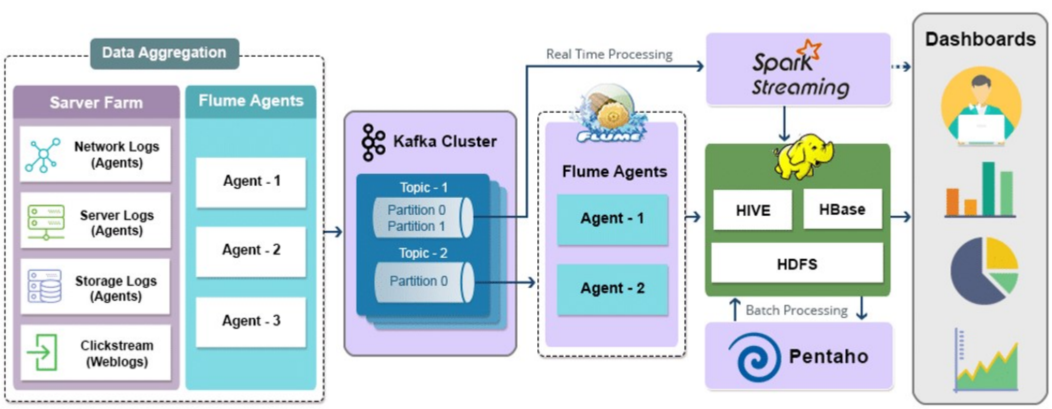
Data lake architecture



Data Visualization in Traditional BI

- Traditional data visualization involves graphical communication of analytical results using charts, maps, data grids, infographics, and alerts.
- Primarily offers static charts and graphs in reports and dashboards.
- Draws data from relational databases, OLAP systems, data warehouses, and spreadsheets.
- Provides descriptive and diagnostic analytics, helping users understand trends and patterns.
  - A dashboard displays real-time information on emergency department wait times, patient satisfaction scores, and the availability of critical resources, helping hospital administrators make informed decisions on resource allocation.

Data Visualization for Big Data Solutions



Data Visualization for Big Data Solutions

- Big Data demands advanced visualization architecture and tools capable of connecting to structured, semi-structured, and unstructured data sources.
- These tools handle vast data records using in-memory analytical technologies, reducing latency.
- Advanced data visualization in Big Data incorporates predictive and prescriptive analytics, eliminating the need for extensive data pre-processing like ETL.
- Allows direct connection to diverse data sources, providing a user-friendly format for viewing data on dashboards.
  - An advanced visualization tool for Big Data is employed to explore patterns in patient genomic data, allowing researchers to visually identify correlations between specific genetic markers and treatment responses

Features of Visualization Tools in Big Data

- Advanced data visualization tools in Big Data share common features:
  - Aggregation: Offers a holistic view of data across multiple contexts.
  - Drill-down: Enables detailed inspection of specific data subsets.
  - Filtering: Focuses on particular data sets by removing irrelevant information.
  - Roll-up: Groups data across categories to display subtotals and totals.
  - What-if analysis: Dynamically changes related factors, visualizing multiple outcomes.

Conclusion and Future Directions

- This chapter delves into essential enterprise technologies and BI concepts.
- Understanding OLTP, OLAP, ETL, data warehouses, and data marts is crucial for effective data management.
- Traditional BI focuses on historical events, while Big Data BI expands boundaries for predictive and prescriptive analytics.
- The new generation data lake helps work through massive volumes of data used for Machine Learning projects which are persistently growing. Data scientists and data engineers have turned to Data Lakes to store very large volumes of data and find meaningful insights.
- Contemporary data visualization tools enhance user understanding, reducing dependency on static reporting methods.