**Distributed Databases**

* A database that is **physically distributed across multiple locations** but still functions as a **single system**.
* **Key Characteristics**:
  + Data is distributed to improve **load balancing** and **performance**.
  + **No redundant data management**, meaning data is stored in fragments.
  + Typically optimized for **transactional processing** rather than analytical queries.
* **Purpose**: Used for **highly scalable** and **available** transactional applications.
* **Example**: A multinational company with databases distributed across different countries

**Data Warehouse**

* A **subject-oriented, integrated, non-volatile, and time-variant** collection of data supporting management decisions (**W.H. Inmon, 1996**).
* **Key Characteristics**:
  + **Subject-Oriented**: Organizes data around business topics.
  + **Integrated**: Combines data from multiple heterogeneous sources.
  + **Non-Volatile**: Data is stable and not frequently updated or deleted.
  + **Time-Variant**: Allows historical analysis and time-series evaluation.
* **Purpose**: Used for analytical processing (OLAP), reporting, and business intelligence.
* **Example**: A retail company using a data warehouse to analyze sales trends across years.

**Federated Databases**

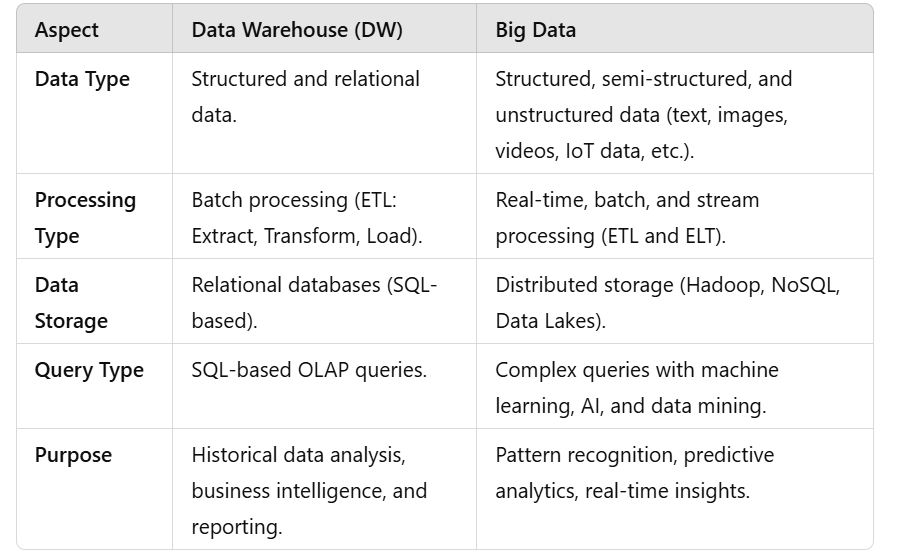
* A system that integrates multiple **autonomous** and **heterogeneous** databases without centralizing them.
* **Key Characteristics**:
  + Ensures **greater autonomy** among databases.
  + No **specific analytical purpose** like data warehouses.
  + **No read-access optimization**, meaning queries can be slower.
* **Purpose**: Used when organizations need to **combine multiple databases without modifying them**.
* **Example**: A hospital system integrating databases from different medical departments while keeping them independent

**Relation of Data Marts to These Concepts:**

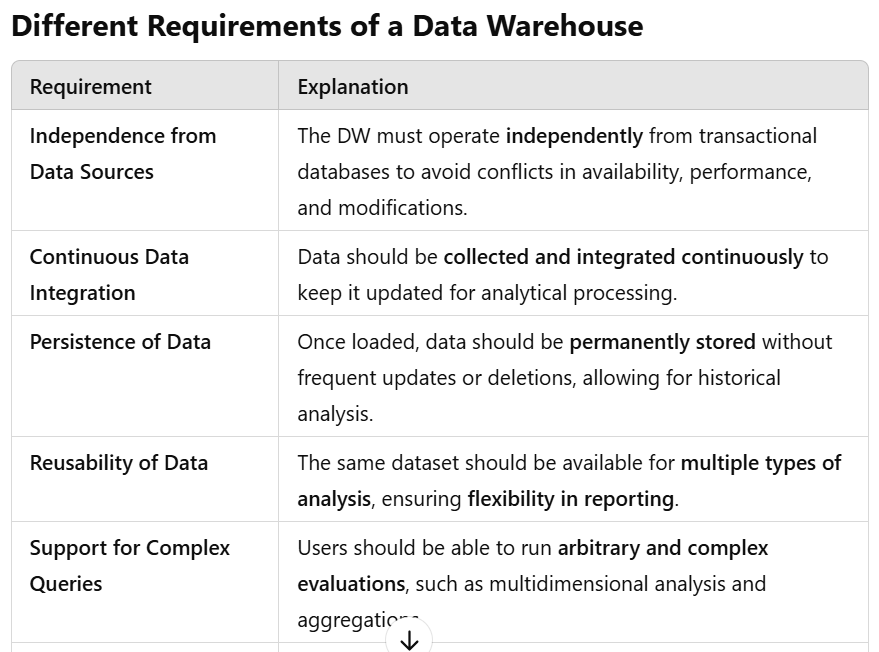
* **Data Marts** are **subsets** of a Data Warehouse, designed for **specific departments or functions** like sales or finance.
* Data marts extract and store relevant data from the data warehouse, making it easier for users to perform specialized analyses. They serve as a bridge between the vast data repository of a data warehouse and the specific needs of individual business units.
* They can be:
  + **Dependent** (created from an existing Data Warehouse).
  + **Independent** (built separately from other data systems).
* **Purpose**: Improve **query performance** and provide **focused analysis**.
* **Example**: A marketing department's Data Mart containing only customer-related data.

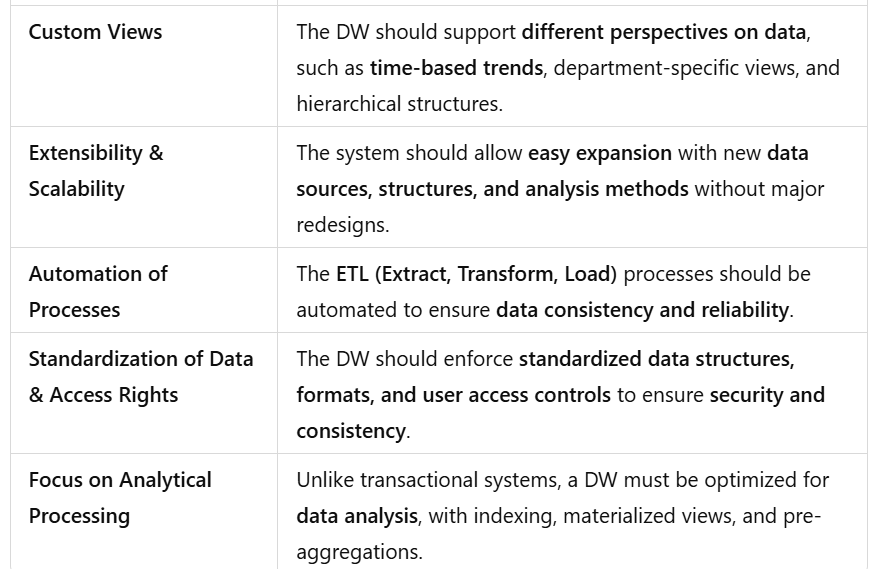
**Classification of Data Warehouse in the Context of Big Data**

A **Data Warehouse (DW)** plays a crucial role in the **Big Data ecosystem** by acting as a structured repository that supports **analytical processing and business intelligence**. However, with the rise of Big Data, traditional data warehouses are being integrated with modern technologies to handle **large-scale, high-velocity, and unstructured data**.



**Different Requirements of a Data Warehouse**





**Common Aspects Across All Requirements**

All the above requirements aim to ensure that the Data Warehouse is:

1. **Integrated** → Combining data from multiple sources into a unified format.
2. **Efficient & Scalable** → Handling large data volumes with fast query response times.
3. **Reliable & Consistent** → Ensuring data accuracy, security, and accessibility.
4. **Optimized for Decision-Making** → Supporting complex queries for business intelligence.
5. **Automated & Maintainable** → Reducing manual intervention in data processing.