Data Marts

A data mart is a subset of a data warehouse that is designed to serve a specific business unit, department, or function within an organization. Unlike a data warehouse, a data mart contains a smaller subset of data that is tailored to the needs of specific users. Data marts can be created by extracting data from a data warehouse or by integrating data from multiple sources.

Data Lakehouse

A data lakehouse is a hybrid architecture that combines the flexibility and scalability of a data lake with the performance and reliability of a data warehouse. A data lakehouse architecture allows organizations to store raw data in a centralized location and use it for a variety of analytics use cases. The data is structured using schema-on-read, which means that the schema is applied to the data only when it is read or queried, allowing for more flexible and agile data analysis.

Data Mesh

Data mesh is a new approach to organizing and managing data within an organization. It emphasizes the creation of decentralized, domain-specific data teams that are responsible for managing the data for their specific domains. Data mesh advocates for a self-serve data platform that provides data teams with the tools and services they need to manage their data independently. This approach aims

to address the challenges of traditional centralized data management, such as data silos and bottlenecks, and to promote a more agile and collaborative approach to data management.

Data Warehouse vs Data Lake

Data Warehouse (DWH)	Data Lake
Schema-on-write	Schema-on-read
Structured data	Structured, semi-structured and unstructured data
Batch processing	Supports batch and real-time processing
Data is optimized for analysis	Data is stored in its raw format
Designed for specific use cases	Designed to support a variety of use cases
Data is stored in predefined schemas	Data is stored in a centralized location without predefined schemas
Data is cleaned and transformed before loading	Data is loaded in its raw form and can be transformed as needed
Queries are optimized for fast, consistent results	Queries are more flexible but may be slower than in a data warehouse

OLTP vs OLAP

OLTP (Online Transaction Processing) is a type of data processing that is used to manage day-to-day business operations. OLTP systems are typically transaction-oriented and handle small, frequent transactions in real time. These systems are designed to support high levels of concurrency and low-latency processing. Examples of OLTP systems include point-of-sale systems, online banking systems, and airline reservation systems.

OLAP (Online Analytical Processing) is a type of data processing that is used to support business intelligence and decision-making. OLAP systems are typically analysis-oriented and handle large, infrequent transactions in batch-processing mode. These systems are designed to support complex queries and provide fast, interactive access to large volumes of data. OLAP systems often use denormalized or star schema data models to optimize for reads and complex queries. Examples of OLAP systems include data warehouses, data marts, and business intelligence tools.

OLTP (Online Transaction Processing)	OLAP (Online Analytical Processing)
Transactional system	Analytical system
Supports day-to-day business operations	Supports business intelligence and decision making
Real-time processing	Batch processing
Transaction-oriented	Analysis-oriented
Relatively simple queries	Complex queries
Frequent, small transactions	Infrequent, large transactions
Low latency	High latency
Normalized data model	Denormalized or star schema data model
Optimized for updates	Optimized for reads and complex queries
Data volume is smaller	Data volume is larger