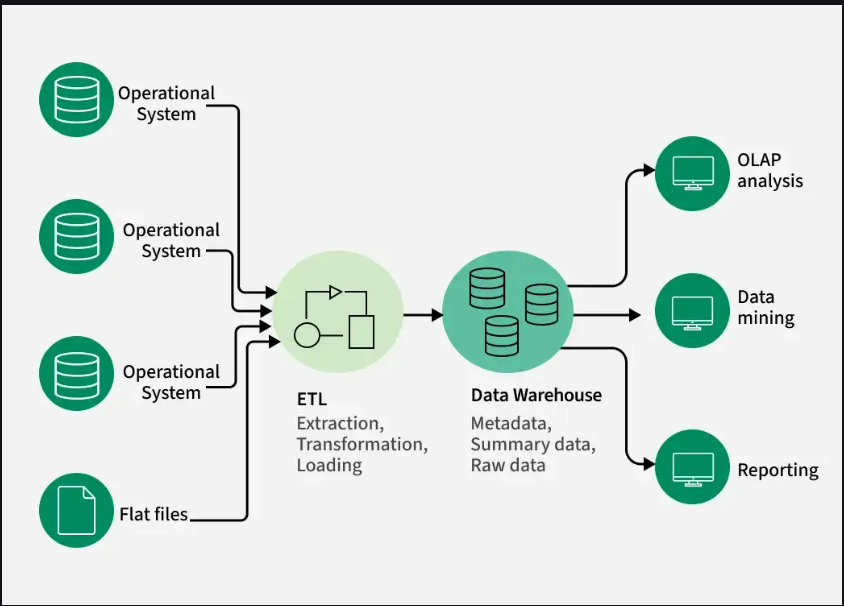
Data Warehousing

Introduction to Data Warehousing

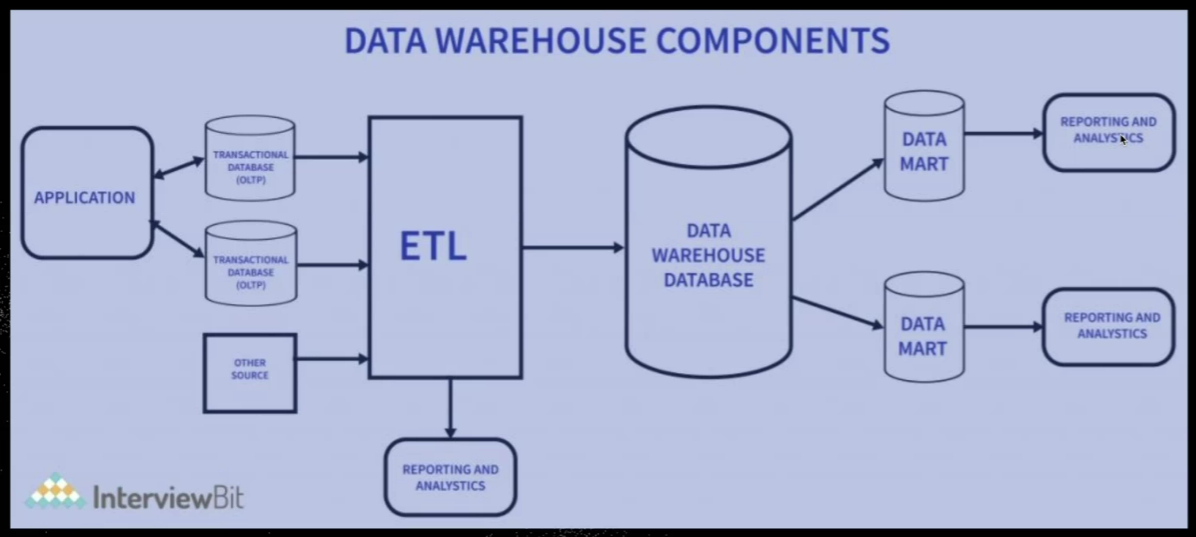
A data warehouse is a system designed for reporting and data analysis, typically storing large volumes of data from multiple sources. It supports decision-making activities in organizations by providing a comprehensive view of their data.

Key Characteristics of Data Warehousing



Need for Data Warehousing

* Handling Large Volumes of Data: Unlike traditional databases (MBs to GBs), data warehouses manage vast datasets (TBs), storing massive historical data.
* Enhanced Analytics: Data warehouses are optimized for complex queries, enabling deeper insights from historical data.
* Centralized Data Storage: They integrate data from multiple sources, offering a unified view of business operations for better decision-making.
* Trend Analysis: With historical data, data warehouses help analyse trends and predict future outcomes.
* Business Intelligence Support: Data warehouses power BI tools, giving decision-makers easy access to critical information for improved efficiency and data-driven strategies.



Components of a Data Warehouse

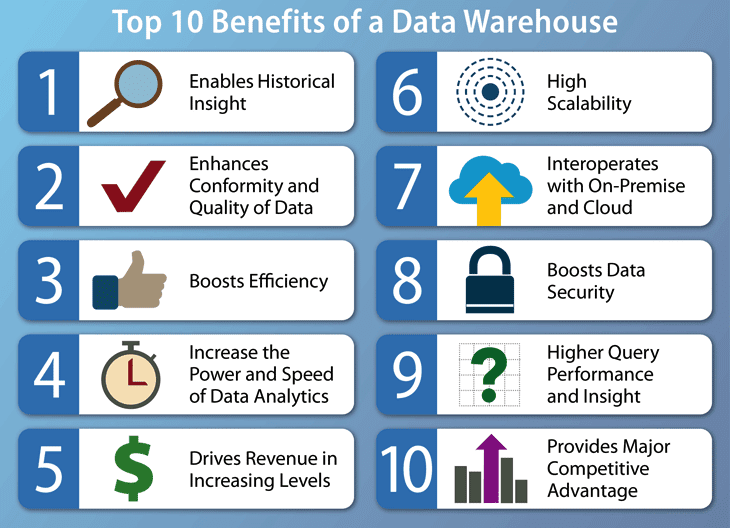
* Data Sources: Operational systems, databases, and external feeds that provide raw data.
* ETL Process: Extracts data, transforms it into a usable format, and loads it into the warehouse.
* Data Warehouse Database: The central repository where cleaned data is stored, organized for efficient querying.
* Data Marts: Smaller, focused data subsets tailored to specific business needs or departments.
* OLAP Tools: Tools that allow multi-dimensional data analysis, supporting complex queries.
* End-User Access Tools: Business logic tools or dashboards that allow business users to query data and generate reports.

Key Characteristics:

* Subject Oriented: Data is organized around key subjects (e.g., sales, finance) rather than transactions.
* Integrated: Data from different sources is standardized into a common format.
* Time Variant: Historical data is stored for analysis over different time periods.
* Non-Volatile: Once data is entered into the warehouse, it is not changed or updated.
* Optimized for analysis: Tailored for quicker access and reporting related to specific area.
* Easier to Manage: Smaller size makes it easier to maintain compared to a full data warehouse.

Benefits of Data Warehousing:

* Improved Business Decision-Making: By centralizing data, businesses can get a more holistic view, aiding in better decision-making.
* Time Efficiency: Helps in faster query processing and reduces the time to generate reports.
* Data Quality and Consistency: Standardized and cleaned data ensures consistency and reliability in reports.



Challenges:

* High Initial Setup Costs: Building a data warehouse can be expensive due to hardware, software, and setup costs.
* Data Quality: If the incoming data is not clean, it can affect the quality of the warehouse.
* Maintenance: Regular updates, backups, and management are required to ensure optimal performance.

