Q.1. What is ETL?

i. ETL (Extract, Transform, Load) is a process used in data warehousing & data integration.

ii. Extract - retrieving raw data from multiple sources like databases, API's or flat files.

iii. Transform - cleaning, filtering, aggregating & converting data into a suitable format.

iv. Load - storing the processed data in a target systems like a data consistency, accuracy & reliability, making it essential for business intelligence & reporting.

v. Modern ETL tools, such as Informatica, Talend, & Apache Nifi, help automate this process.

Q.2. What are the different components of BI?

i. Data sources - structured & unstructured data from databases, API's, & external sources.

ii. ETL Tools - extract, transform & load data into a centralized repository.

iii. Data Warehouse - a centralized storage systems for historical & current data.

iv. OLAP - Online Analytical Processing allows multidimensional analysis of data.

v. Reporting & visualization - Dashboards, reports, & graphs using tools like Tableau & PowerBI.

Q.3. What is Data Extraction?

i. Data extraction is the process of retrieving raw data from various sources, such as relational databases, cloud storage, APIs, spreadsheets, web scraping & unstructured sources like emails or PDFs.
ii. The extracted data can be structured, semi-structured or unstructured depending on its origin.
iii. This process is the first step in ETL & is essential for data warehousing, business intelligence & analytics.
iv. Data extraction can be done in batch mode or real-time mode.
v. Tools like Apache Nifi, Talend & Python scripts are commonly used for extraction.

Q.4. What do you mean by data Transformation?

i. Data Transformation is the process of converting raw data into structured, meaningful format to support analysis & decision-making.
ii. It includes data cleaning, data

aggregation, data enrichment & data mapping.

iii. Transformation is crucial in ETL as it ensures consistency & accuracy before loading data into a warehouse.

iv. Tools like Apache Spark, Talend & SQL queries facilitate this process.

v. Effective transformation improves data usability for business intelligence & analytics, making it easier to generate reports & identify patterns.

Q5. What is Data Warehouse?

i. A data warehouse is a centralized repository where structured & historical data is stored for analysis & reporting.

ii. It integrates data from multiple sources, including transactional databases, CRM systems & external API's ensuring consistency & Reliability.

iii. Unlike operational databases, which handle real-time transactions, data warehouses are optimized for analytical queries & support-large-scale decision making.

iv. Key characteristics include subject-oriented, time-variant, integrated & non-volatile.

v. Technologies like Amazon Redshift, Google BigQuery & snowflake are popular

for cloud based data warehousing.

Q. 6. What data mart?

i. A data mart is subset of a data warehouse designed for specific business functions like sales, marketing, finance or HR.
ii. It contains focused, subject-specific data, making it easier for teams to access relevant insights without querying the entire warehouse.
iii. Data marts can be dependent or independent.
iv. They improve performance by reducing query complexity & enabling faster reporting.
v. Common technologies for data marts include Amazon Redshift, Snowflake & Microsoft Azure Synapse.