**1. Define ETL and explain its importance in data management.**

**Ans.** ETL-Extract, Transform, Load.

**Extract-** The data is taken from various sources, such as data ware-house,

Database,Kaggle

**Transform-** The data which is taken from the extracted data is clean ,formatted

and convert data into particular formats necessary structured way.

**Load-** The transformed data is loaded into the particular system

**Importance** Data Integration, Data Quality ,Efficiency, Decision Making

**2. Describe a scenario where ETL could be beneficial in a business setting.**

BigBasket can extract data from various stores, transform the data accordingly and load it into a data warehouse and perform analysis on it to know about best selling products, etc

**3. What challenges might a data analyst face during the transformation phase of ETL and how can they be addressed?**

**Challenges**: Inconsistent data formats, data quality issues, complex data transformations.

**Answers**:

**Data Cleaning:** Inconsistent formats, missing values, and duplicates require cleansing techniques .

**Data Derivation:** Calculations, aggregations, and manipulations are needed to derive meaningful metrics and insights.

**Data Validation:** Ensuring data accuracy and consistency with business rules is crucial.

**4. Explain the concept of data warehousing and its relationship with ETL processes.**

A Data Warehouse is a relational database that is designed for query and analysis rather than transaction processing. and central repository designed for historical data analysis.

Relationship with ETL: ETL is the foundation for populating the data warehouse with clean, transformed data.

**5. Define a database and a data warehouse.**

**Database** : is a collection of data that is stored and accessed electronically.

**Data Warehouse:**

is a relational database that is designed for query and analysis rather than transaction processing.

**6. How do the purposes of a database and a data warehouse differ in a business environment?**

**Databases:** Support day-to-day business operations, ensuring efficiency and data integrity in core transactions**. and** used for OLTP (Online Transaction Processing)

**Data Warehouses:** Enable data analysis to gain insights into trends, customer behavior, and performance over time. and used for OLAP (Online Analytical Processing).

**7. Can you illustrate with an example when you would use a database versus a data warehouse?**

**Database:** An e-commerce website's database stores customer information, product details, and real-time order processing. and used for day to day customer transactions at e-commerce websites

**Data Warehouse:** used for knowing the best selling product last year.

**8. List 5 Popular Data Warehouse, ETL Tools and Database.**

**Data Warehouses:**Amazon Redshift**,** Microsoft Azure**,** BigQuery , Snowflake

**ETL Tools:** Informatica ,IBM DataStage, Pentaho Data Integration

**Databases:** MySQL, PostgreSQL, Microsoft SQL Server, Oracle Database

**9. Who is Data Analyst, Business Analyst and Data scientist?**

**Data Analyst:** The person who analyzes the data and provides actionable insights.

**Business Analyst:** The person who focuses on improvising the business processes and Understands business processes and uses data to recommend improvements.

**Data Scientist:** : The person who uses advanced data analysis, including machine learning, to extract deeper insights. Develops and applies machine learning models to extract knowledge from data.

**10. Illustrate with an example how data visualization can assist in business decision-making**

A sales manager uses a dashboard to visualize sales across India. The visualization can show the declining sales, growing sales, overall sales, etc

**Heat Maps:** Highlight regions with high and low sales.

**Bar Charts:** Compare monthly sales against targets.

**Line Graphs:** Show sales trends over time.