Ans 1) ETL stands for Extract, Transform and Load. It is the process to extract raw data from multiple sources, transform it into clean and organised data and load it into data warehouse.

Ans 2) For example:
The company possesses a large volume of data but faces challenges in utilising it due to its complexity and difficulty in analysis. The data is dispersed across various systems, making access difficult. The company aims to identify the relevant data and ensure quick and easy access, enabling fact-based decision-making that will drive growth and maximise profits.

Ans 3) Challenges that a data analyst might face during the transformation phase of ETL are:
(i) Incomplete Data
(ii) Duplicate Data
(iii)Summarising data from multiple sources

Problems can be address using ETL processes by eliminating duplicate data, by sorting and applying different functions.

Ans 4) A data warehouse, or enterprise data warehouse (EDW), is a system that aggregates data from different sources into a single, central, consistent data store to support data analysis, data mining, artificial intelligence (AI) and machine learning.

Relationship with ETL process: ETL organises data like arranging puzzle pieces, preparing them for storage in the data warehouse, where they form a complete picture for analysis.

Ans 5) **DATABASE:** A database is a collection of data that is stored and accessed electronically. Databases can store and manage large amounts of data, and can be used for a variety of activities, including data storage, analysis, and management.

DATA WAREHOUSE: A Data Warehouse (DW) is a relational database that is designed for query and analysis rather than transaction processing.

Ans 6) In a business environment, a database is primarily used for transactional processing, storing current and operational data, while a data warehouse is utilised for analytical purposes, storing historical and aggregated data for reporting and analysis.

Ans 7) Imagine a university managing student records. For daily tasks like course registration and updating grades, it uses a database like MySQL, which handles real-time updates efficiently. For analysing long-term trends in academic performance and enrollment, the university uses a data warehouse like Amazon Redshift. Data warehouses store extensive historical data and support complex queries, helping with strategic planning. So, the database is for everyday operations, while the data warehouse is for in-depth analysis and future planning.

Ans 8) **Data Warehouse Tools:**
(iv)Amazon REDSHIFT
(v) Google Big Query
(vi)Snowflake
(vii)PostgreSQL
(viii)IBM DB2

ETL Tools:
(i) Cloud DataFlow
(ii) Oracle
(iii)Azure Data Factory
(iv)Pentaho
(v) Informatica

Database Tools:
(i) Oracle
(ii) MySQL
(iii)Microsoft SQL Server
(iv)SQLite
(v) Microsoft Azure SQL DataBase

Ans 9) **Data Analyst**: A Data Analyst is someone who gathers, analyzes, and interprets data to assist organizations in making informed decisions. They use tools like SQL, Excel, and Python to process large datasets and reveal valuable insights. By identifying patterns

and trends within the data, they create reports and visualizations to communicate their findings effectively. Data Analysts play a vital role across different industries, helping businesses utilize data for strategic planning and decision-making.

Business Analyst: A Business Analyst serves as a link between different departments within an organization, analyzing processes, gathering requirements, and proposing solutions to enhance efficiency. They bridge the gap between business needs and technical solutions, translating stakeholder requirements into actionable plans using tools like flowcharts and documentation. Business Analysts play a crucial role in project management, ensuring projects align with organizational goals and deliver value to the business.

Data Scientist: A Data Scientist is a professional skilled in advanced statistical methods, machine learning, and programming. They analyze large datasets using tools like Python or R, uncovering insights to make data-driven decisions. Data Scientists play a crucial role across industries, developing predictive models and communicating findings through reports and visualizations to drive innovation and optimize processes within organizations.

Ans 10) Let's consider a small online bookstore trying to decide which genres of books to feature prominently on their homepage.

Without data visualization: The bookstore owner would have to manually analyze sales data, which might be overwhelming and time-consuming. They'd have to go through spreadsheets or databases listing book titles, genres, and sales figures.

With data visualization: The owner can create a simple pie chart or bar graph illustrating the distribution of sales across different genres. For instance, they might see that mystery novels account for 40% of sales, followed by romance at 30% and fantasy at 20%.

Seeing this visual representation, the owner can quickly identify which genres are the most popular among customers. Based on this insight, they can adjust their homepage layout to feature more mystery novels prominently, perhaps with a "Top Mystery Picks" section, while still showcasing other genres but with less prominence.

By using data visualization, the bookstore owner can make informed decisions about how to allocate homepage space effectively, maximizing customer engagement and potentially increasing sales.