What is ETL? in detail.

1. ETL stands for Extract, Transform, Load. It is a process used in data integration to extract data from various sources, transform it into a common format, and load it into a target system, such as a data warehouse or data lake. The ETL process involves three main stages:

* Extraction: Data is extracted from various sources, which may include databases, files, or applications.
* Transformation: The extracted data is transformed to a common format, which may involve data cleaning, data enrichment, data validation, and data consolidation.
* Loading: The transformed data is loaded into a target system, such as a data warehouse or data lake, where it can be further processed and analyzed.

What is ELT? in detail.

* ELT stands for Extract, Load, Transform. It is a process used in data integration that is similar to ETL, but with a different sequence of stages. In ELT, data is first extracted from various sources and loaded into a target system, such as a data lake, and then transformed into a common format. The ELT process involves three main stages:
* Extraction: Data is extracted from various sources, which may include databases, files, or applications.
* Loading: The extracted data is loaded into a target system, such as a data lake, where it can be stored and processed.
* Transformation: The data is transformed into a common format, which may involve data cleaning, data enrichment, data validation, and data consolidation.

3 Tier Architecture in DE

1. Three-tier architecture is a data engineering design pattern that separates the data processing and storage into three layers:

* Presentation layer: The top layer is the presentation layer, which is responsible for displaying the data to end-users. This layer includes tools for querying and analyzing data, such as BI reporting tools, dashboards, and data visualization software.
* Application layer: The middle layer is the application layer, which contains the business logic and data processing components. This layer includes ETL tools, data integration software, and data transformation engines.
* Data storage layer: The bottom layer is the data storage layer, which is responsible for storing and managing data. This layer includes data warehouses, data lakes, and databases.

ETL Tools (any 3)

1. There are many ETL tools available in the market, here are three popular ones:

* Informatica PowerCenter: Informatica PowerCenter is a popular ETL tool that supports complex data integration requirements. It provides a visual interface for designing and building ETL workflows, and it supports various data sources and targets, including databases, files, and applications.
* Apache NiFi: Apache NiFi is an open-source ETL tool that is used for data ingestion and data flow management. It provides a web-based interface for designing and building ETL workflows, and it supports various data formats and protocols, including JSON, XML, and REST.
* Microsoft SQL Server Integration Services (SSIS): SSIS is a popular ETL tool that is included in Microsoft SQL Server. It provides a visual interface for designing and building ETL workflows, and it supports various data sources and targets, including databases, files, and applications.

What is Historical Load?

* Historical load is a type of ETL (Extract, Transform, Load) process in which all historical data from source systems is extracted and loaded into a target system, such as a data warehouse or data lake. The data is transformed and loaded in a way that allows historical analysis and reporting of the data.

What is Full Load?

* Full load is a type of ETL process in which all data from source systems is extracted and loaded into a target system, such as a data warehouse or data lake. This process involves extracting all data from source systems, transforming it into a common format, and loading it into a target system. Full load is typically used when the target system is being set up for the first time or when a significant change has occurred in the source system.

What is Incremental Load?

* Incremental load is a type of ETL process in which only new or changed data since the last load is extracted and loaded into a target system, such as a data warehouse or data lake. This process involves comparing the source data to the target data and extracting only the records that have changed or been added since the last load. Incremental load is used to update data in the target system without having to reload all the data each time. This process reduces the time and resources required for data integration and enables near real-time analytics.