Fariz khanzada

Task 4

Q.1

The ETL process is a critical component of data integration and management, as it enables organizations to consolidate data from multiple sources and transform it into a format that is consistent and usable for analysis and reporting. ETL tools and platforms, such as Apache NiFi, Talend, or Microsoft SQL Server Integration Services (SSIS), provide a visual interface for designing and managing ETL workflows, as well as scheduling and monitoring ETL jobs.

Q.2

ELT stands for Extract, Load, Transform. It is a data integration process that involves loading raw data into a target system and transforming it into a usable format for analysis and reporting.

Q.3

The three-tier architecture in data engineering typically refers to a design pattern that separates the data processing system into three distinct layers, as follows:

1. Data Storage Layer: This layer involves the storage of raw data, typically in a data lake or a data warehouse. The data storage layer provides a scalable and flexible storage environment for the raw data, which can be processed and transformed later.
2. Data Processing Layer: This layer involves the processing and transformation of the raw data into a usable format for analysis and reporting. The data processing layer typically uses technologies such as Apache Spark, Apache Flink, or Apache Beam to perform data processing tasks such as cleaning, filtering, aggregating, and enriching the data.
3. Data Analytics Layer: This layer involves the analysis and reporting of the processed data, typically using business intelligence (BI) tools such as Tableau, Power BI, or QlikView. The data analytics layer provides insights and visualizations to help stakeholders make informed decisions based on the data.

The three-tier architecture in data engineering is a widely used design pattern that provides scalability, flexibility, and agility for data processing and analytics.

Q.4

Apache NiFi – An open-source data integration platform that provides a web-based interface for designing and managing ETL workflows.

Talend – A popular data integration and ETL tool that supports a wide range of data sources and destinations, as well as a drag-and-drop interface for designing ETL workflows.

Microsoft SQL Server Integration Services (SSIS) – A data integration and ETL tool that is part of the Microsoft SQL Server suite of tools, which provides a visual interface for designing and managing ETL workflows, as well as scheduling and monitoring ETL jobs.