SADAF SULTAN

Bytewise fellowship

Task #4

What is ETL ? in detail.

ETL stand for "**Extract, Transform and load"**
In data warehousing to bring the data from different data sources into one centralized database we need to **extract** data from original source,**transform** that data by deduplication, combination and ensuring the quality of data and then **load** the data into target database.
EXTRACT(extract the data from multiple sources)
TRANSFORM(in this ETL process, rules and regulations can be applied that ensure data quality and accessibility)
LOAD(in the ETL process is to load the newly transformed data into a new destination (data lake or data warehouse.) Data can be loaded all at once (full load) or at scheduled intervals (incremental load)**)**

For Example:
In power-bi 1st of all we download the dataset from internet which we need for our project this step is known as extraction of data, secondly we transform that dataset according to our demand and requirement we filter it,sorting,verification ,make some measurements its know as transformation of data and at the end we load that data into our tool and start make visualization.

A typical ETL process collects and refines different types of data, then delivers the data to a data lake or data warehouse such as Redshift, Azure, or BigQuery. Most companies today rely on an ETL tool as part of their data integration process. ETL tools are known for their speed, reliability, and cost-effectiveness, as well as their compatibility with broader data management strategies. ETL tools also incorporate a broad range of data quality and data governance features.

What is ELT ? in detail.

ELT is the next generation of ETL.ELT is a modern take on the older process of extract, transform, and load in which transformations take place before the data is loaded. While the purpose of ETL is the same as ELT, the method is evolved for better processing.Traditional ETL software extracts and transforms data from different sources before loading it into a data warehouse or data lake. With the introduction of the cloud data warehouse, there was no longer the need for data cleanup on dedicated ETL hardware before loading into your data warehouse or data lake. The cloud enables a push-down ELT architecture with two steps changed from the ETL pipeline.

EXTRACT(extract the data from multiple sources)

LOAD(load it into cloud data warehouse)

TRANSFORM(transfrom it using the power and scalability of the target cloud platform)

ETL Tools (any 3).

- Integrate.io

  Cloud type

  Integrate.io is a leading low-code data integration platform with a robust offering (ETL, ELT, API Generation, Observability, Data Warehouse Insights) and hundreds of connectors to build and manage automated, secure pipelines in minutes. This platform is highly scalable with any data volume or use case, while enabling you to easily aggregate data to warehouse, databases, data stores, and operational systems.

- SAS Data Management

  Enterprise type

  SAS Data Management is a data integration platform built to connect with data wherever it exists, including cloud, legacy systems, and data lakes.

  Its provide holistic view of organization's business processes. It's a flexible tool and work in variety of computing environments and database as well as integration with third party data modeling tools to produce compelling visualizations.

- Dataddo

  Cloud type

  Dataddo is a no-code, cloud-based ETL platform that enables technical and non-technical users to flexibly integrate data.t offers a wide range of connectors, fully customizable metrics, a central system for simultaneous management of all data pipelines, and can be seamlessly incorporated into existing technology architecture.

  Users can deploy pipelines within minutes of account creation and all API changes are managed by the Dataddo team, so pipelines require no maintenance

Task # 5

What is Historical Load?

Historical load is the one-time initial load of data that the Source already had before the creation of the Pipeline.**Data collected about past events and circumstances pertaining to a particular subject**. By definition, historical data includes most data generated either manually or automatically within an enterprise.

What is Full Load?

In a Full Data Load, the **complete dataset is emptied or loaded** and then entirely overwritten (i.e. deleted and replaced) with the newly updated dataset in the next data loading run. While comparing the Incremental Data Load vs Full Load, you also don't need to maintain extra information such as timestamps to carry out a Full Data Load.

What is Incremental Load?

Incremental load is the periodic load to keep the data warehouse updated with the most recent transactional data. This is an on going process that continues till the life of the warehouse/mart. The periodicity of the incremental loads is dependent on the availability time of the source data.

A less comprehensive but more manageable approach is incremental loading. Incremental loading compares incoming data with what's already on hand, and only produces additional records if new and unique information is found. This architecture allows smaller, less expensive data warehouses to maintain and manage business intelligence.