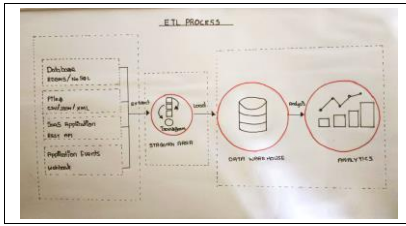
TASK 1	ANSWERS
What is a ETL ?	ETL, which stands for **Extract**, **Transform** and **load** is a data integration process. That combines data from multiple data sources into a single, consistent data store that is loaded into a **data warehouse** or other target system.
What is a ELT?	ELT, which stands for "**Extract, Load, Transform,**" is another type of data integration process, similar to its **counterpart ETL**, "Extract, Transform, Load". This process moves *raw data from a source system to a destination resource*, such as a data warehouse. While similar to ETL, ELT is a fundamentally different approach to data pre-processing which has only *more recently gained adoption with the transition to cloud environments.*
List out the various tools used in ETL	1.Hevo Data 2.Pentaho 3.Talend 4.AWS Glue 5.Informatica PowerCenter 6.Azure Data Factory 7.IBM Infosphere DataStage 8.Blendo 9.StreamSets 10.Integrate.io 11.Xtract.io 12.Jaspersoft 13.Sybase ETL 14.SAP BusinessObjects Data Integrator
Draw the mechanism of ETL	
What is a datalake	.A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale. You can store your data as-is, without having to first structure the data, and run different types of analytics—from dashboards and visualizations to big data processing, real-time analytics, and machine learning to guide better decisions.
What is a data warehouse	A data warehouse is a type *of data management system that is designed to enable and support business intelligence (BI) activities*, *especially analytics*. Data warehouses are solely intended to perform queries and analysis and often contain large amounts of historical data.A data warehouse is a *centralized storage system that allows for the storing, analyzing, and interpreting of data in order to facilitate better decision-making*. Transactional systems, relational databases, and other sources provide data into data warehouses on a regular basis
OLTP VS OLAP	**OLAP (Online Analytical Processing) OLAP (Online Analytical Processing)** ~It is well-known as an online database query management system. ~Consists of historical data from various Databases. ~It makes use of a data warehouse. ~t is subject-oriented. Used for Data Mining, Analytics, Decisions making, etc. ~In an OLAP database, tables are not normalized. **OLTP (Online Transaction Processing)** ~It is well-known as an online database modifying system. ~Consists of only operational current data. ~It makes use of a standard database management system (DBMS). ~It is application-oriented. Used for business tasks. ~In an OLTP database, tables are normalized (3NF).

What are all the various analytical tools that can be connected with Datawarehouse?

Amazon Redshift
Microsoft Azure
Google BigQuery
1.Snowflake
2.Micro Focus Vertica
3.Teradata
4.Amazon DynamoDB
5.PostgreSQL
6.Amazon RDS
7.Amazon S3
8.SAP HANA
9.MarkLogic
10.MariaDB
11.IBM Db2 Warehouse
12.Exadata

Explain the different stages used in ETL

Extract
During data extraction, raw data is copied or exported from source locations to a staging area. Data management teams can extract data from a variety of data sources, which can be structured or unstructured. Those sources include but are not limited to:

~SQL or NoSQL servers
~CRM and ERP systems
~Flat files
~Email
~Web pages

Transform
In the staging area, the raw data undergoes data processing. Here, the data is transformed and consolidated for its intended analytical use case. This phase can involve the following tasks:

~Filtering, cleansing, deduplicating, validating, and authenticating the data.

Explain 3 use cases for NOSQL

Personalization.

A personalized experience requires data, and lots of it – demographic, contextual, behavioral and more. The more data available, the more personalized the experience. However, relational databases are overwhelmed by the volume of data required for personalization. In contrast, a distributed NoSQL database can scale elastically to meet the most demanding workloads and build and update visitor profiles on the fly, delivering the low latency required for real-time engagement with your customers.

Profile Management.

User profile management is core to Web and mobile applications to enable online transactions, user preferences, user authentication and more. Today, Web and mobile applications support millions – or even hundreds of millions – of users. While relational databases can struggle to serve this amount of user profile data as they are limited to a single server, distributed databases can scale out across multiple servers. With NoSQL, capacity is increased simply by adding commodity servers, making it far easier and less expensive to scale.

Content Management.

What are all the characteristics of a nosql

It's more than rows in tables—NoSQL systems store and retrieve data from many formats: key-value stores, graph databases, column-family (Bigtable) stores, document stores, and even rows in tables.

It's free of joins—NoSQL systems allow you to extract your data using simple interfaces without joins.

It's schema-free—NoSQL systems allow you to drag-and-drop your data into a folder and then query it without creating an entity-relational model.

It works on many processors—NoSQL systems allow you to store your database on multiple processors and maintain high-speed performance.

It uses shared-nothing commodity computers—Most (but not all) NoSQL systems leverage low-cost commodity processors that have separate RAM and disk.

It supports linear scalability—When you add more processors, you get a consistent increase in performance.

It's innovative—NoSQL offers options to a single way of storing, retrieving, and manipulating data. NoSQL supporters (also known as NoSQLers) have an inclusive

What are all the tools used to analyse nosql

1. MongoDB Compass
2. Panoply (SQL/NoSQL)
3. Knowi (NoSQL)
4. Izenda (SQL/NoSQL)
5. FineReport (SQL/NoSQL)
6. SolarWinds Database Performance Monitor
7. ClusterControl
8. MongoJS Query Analyzer
9. FineReport
10. Tableau