**Student: Abdullah Alakbarli**

**Teacher: Şükür İbadov**

**LESSON 6 - TASK**

**1)Union və union all nədir?**

UNION and UNION ALL in SQL are set operators that combine the result of two SELECT queries. Both share standard features, with one significant difference: UNION only returns a unique record, while UNION ALL returns all the records (including duplicates)

UNION ALL is generally faster than UNION because it does not perform the additional step of removing duplicates.

For example:

**SELECT employee\_id, employee\_name**

**FROM employees1**

**UNION**

**SELECT employee\_id, employee\_name**

**FROM employees2;**

And union all:

**SELECT employee\_id, employee\_name**

**FROM employees1**

**UNION ALL**

**SELECT employee\_id, employee\_name**

**FROM employees2;**

**2)What is the OLAP, MOLAP, ROLAP, HOLAP?**

**OLAP-** OLAP stands for Online Analytical Processing, and it refers to a category of software tools and technologies that enable users to interactively analyze and explore multidimensional data from different perspectives. OLAP systems organize data into multidimensional structures, typically represented as data cubes. These cubes consist of dimensions (categories or attributes) and measures (numeric data). OLAP supports operations such as drilling down (going from a higher level of abstraction to a lower level) and rolling up (going from a lower level to a higher level). These operations help users navigate through the data cube to gain insights.

OLAP cubes are central to OLAP systems. These cubes provide a way to organize and structure data for efficient analysis. Users can navigate through the dimensions and measures to explore data interactively. OLAP systems often pre-calculate and store aggregated data to enhance query performance. This allows for quick responses to user queries, even when dealing with large datasets.

**MOLAP-** MOLAP databases often pre-aggregate and store summarized data at various levels of granularity. This pre-aggregation enhances query performance by allowing the system to quickly retrieve summarized results without having to perform extensive calculations on the raw data. MOLAP systems are known for their fast query response times. This is achieved through the pre-aggregation of data and the efficient indexing of multidimensional structures. MOLAP databases typically involve an offline processing step where data is loaded into the multidimensional structure. This allows for the creation of aggregated data sets and the optimization of cube structures before users interact with the system.

**ROLAP-** ROLAP systems store data in traditional relational database tables. Unlike MOLAP, which uses multidimensional structures, ROLAP leverages the relational model to organize and manage data. ROLAP systems are designed to take advantage of the capabilities of relational database management systems (RDBMS). They use SQL (Structured Query Language) for querying and leverage the indexing and optimization features provided by the underlying database. ROLAP can handle more flexible and complex data models, including normalized data structures. This flexibility allows for a wide range of data representations, making it suitable for scenarios where data models are not easily captured in a predefined cube. ROLAP systems can provide real-time access to data stored in relational databases. This is beneficial for scenarios where up-to-the-minute information is critical. ROLAP is often chosen in situations where the underlying data is already stored in relational databases, and there is a need to perform analytical processing without the need for a separate multidimensional storage structure.

**Holap -** HOLAP stands for Hybrid Online Analytical Processing. HOLAP is an approach that combines elements of both MOLAP (Multidimensional Online Analytical Processing) and ROLAP (Relational Online Analytical Processing) systems. The goal of HOLAP is to leverage the strengths of both MOLAP and ROLAP to provide a balanced solution for efficient and flexible analytical processing. HOLAP systems provide flexibility in data modeling, allowing users to work with both multidimensional structures and relational tables. This flexibility accommodates different data models and complexities.

**3)What is the mapReduce and Hadoop?**

**MapReduce** - MapReduce provides a programming model for processing and generating large datasets that can be parallelized across a distributed computing cluster. MapReduce provides a programming model for processing and generating large datasets that can be parallelized across a distributed computing cluster.

**Hadoop** - Hadoop is an open-source framework that implements the MapReduce programming model. It also includes a distributed file system called Hadoop Distributed File System (HDFS) for storing large datasets.

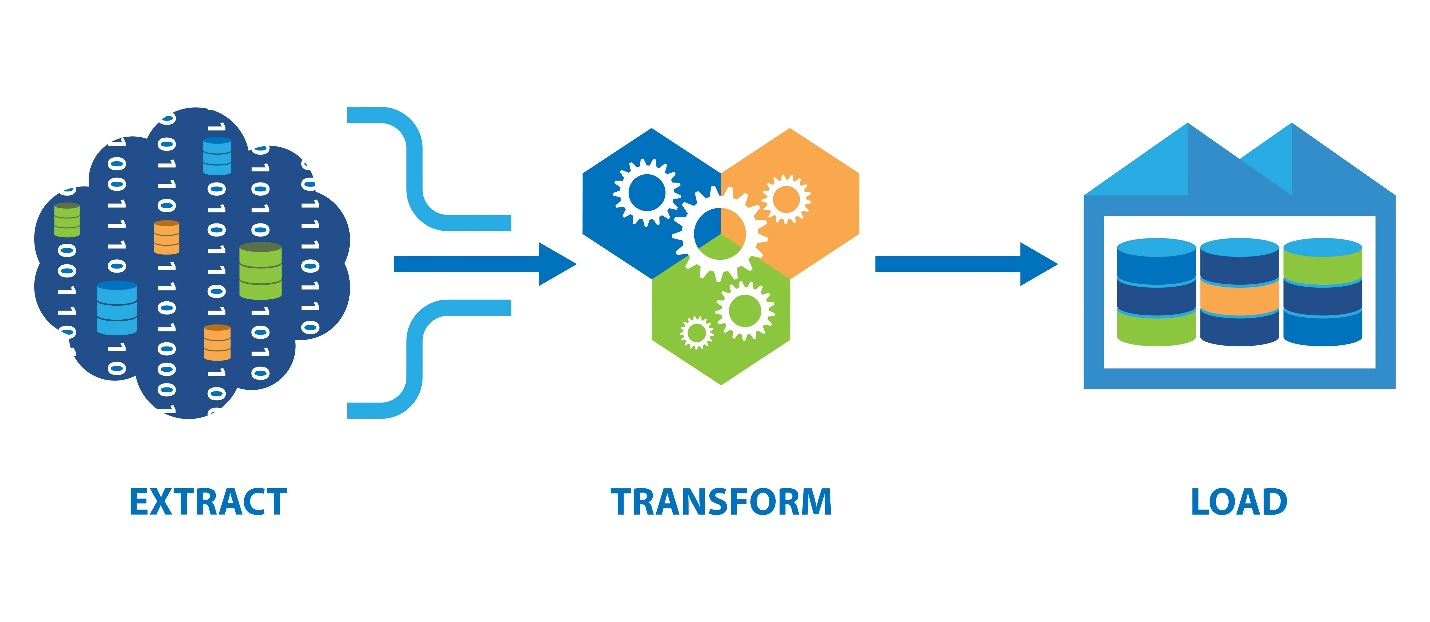
**4)What is ETL?**

ETL is a process that divides 3 part. **Extract, Transform, Load**.

Extract process- Data gathering from various source systems.

Transform- The "Transform" phase involves cleaning, validating, and transforming the raw data into a format that is suitable for analysis and reporting. Transformation activities may include data cleansing (removing or correcting errors), data validation, standardization of formats, and the application of business rules. Aggregation, normalization, filtering, and enrichment are common transformations performed during this phase

Load - Aggregation, normalization, filtering, and enrichment are common transformations performed during this phase. Loading can involve appending new data, updating existing records, or completely overwriting the target system, depending on the data integration strategy. Loading can involve appending new data, updating existing records, or completely overwriting the target system, depending on the data integration strategy.



**5) What is the ERP?**

ERP – ENTERPRISE RESOURCE PLANING - ERP systems typically include a suite of integrated applications to support various business functions. ERP systems use a centralized database that stores data from different business processes. This centralized data repository ensures consistency and a single source of truth for the organization. ERP systems provide real-time or near-real-time access to data, allowing organizations to make informed decisions based on the most up-to-date information. ERP systems enable cross-functional workflows by automating and streamlining business processes across departments. For example, an ERP system can link the order processing in sales with inventory management and production planning. Many ERP systems allow for customization to meet the specific needs of an organization. Additionally, ERP solutions are often scalable, allowing companies to expand their usage as they grow.