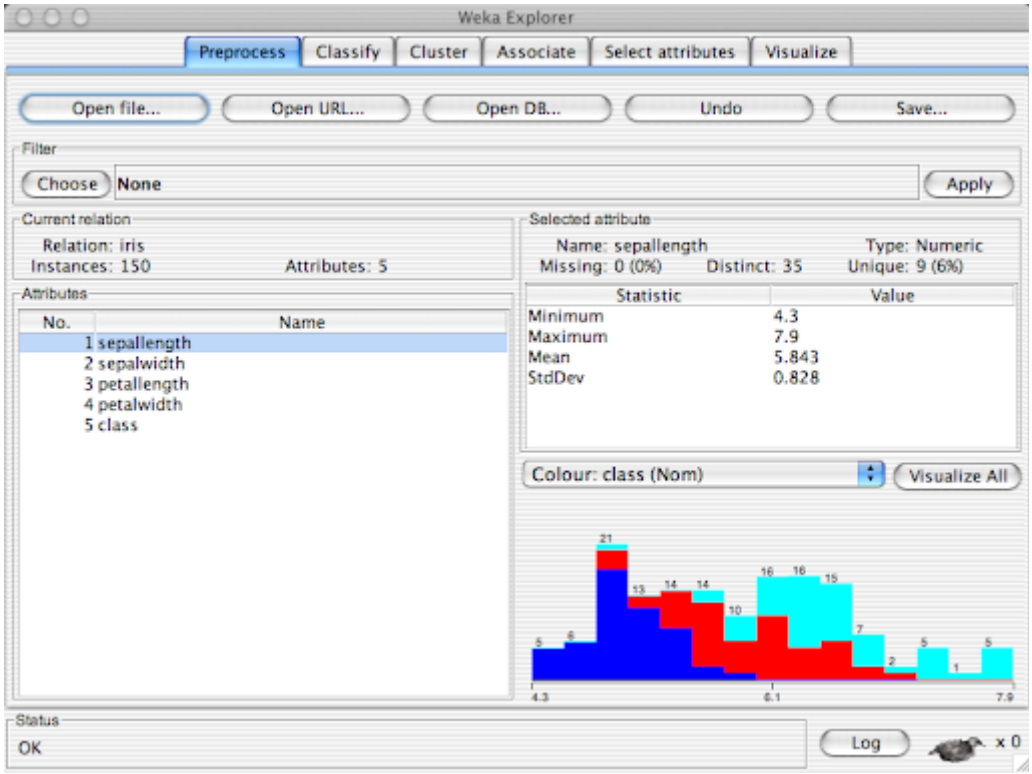
Name-TUSHAR RAJ VERMA

1.Setup ETL Datastudio and copy files from one location to other location.Identify components that could be used to transform data in between.



Q2.) What are the 2 types of systems present? Take one industry and explain how the 2 systems can be applied to that industry.

OLTP

An OLTP system captures and maintains transaction data in a database. Each transaction involves individual database records made up of multiple fields or columns. Examples include banking and credit card activity or retail checkout scanning.
In OLTP, the emphasis is on fast processing, because OLTP databases are read, written, and updated frequently. If a transaction fails, built-in system logic ensures data integrity.

Example of OLTP system

An example of OLTP system is ATM center. Assume that a couple has a joint account with a bank. One day both simultaneously reach different ATM centers at precisely the same time and want to withdraw total amount present in their bank account.
However, the person that completes authentication process first will be able to get money. In this case, OLTP system makes sure that withdrawn amount will be never more than the amount present in the bank. The key to note here is that OLTP systems are optimized for transactional superiority instead data analysis.

OLAP

Online Analytical Processing, a category of software tools which provide analysis of data for business decisions. OLAP systems allow users to analyze database information from multiple database systems at one time.
The primary objective is data analysis and not data processing.

Example of OLAP

Any Datawarehouse system is an OLAP system. Uses of OLAP are as follows :

- The OLAP will be used to analyse historic data and extract information accordingly.These two systems are required for banking softwares and customer will have a relevent result.

Q3.) i) Explain the need of DWH in airline industry.

Solution:-

Airline industry uses basics of OLTP for tickets confirmation and seat allotment of customers as being a transactions to commit.As in OLAP the dynamic time events such as occupancy and vacancy are done by analysis.The routes of flight are also being analysed to have a optimul path.So a datawarehouse will be mantain these types of transactional and analytical formations.

Q4.) What are the 4 terms that are used to describe a data warehouse? Take an IT company as an example and come with the possible subjects around which a data warehouse can be created.

Solution:-

1. **Subject Oriented** − A data warehouse is subject oriented because it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue, etc. A data warehouse does not focus on the ongoing operations, rather it focuses on modelling and analysis of data for decision making.

2. **Integrated** − A data warehouse is constructed by integrating data from heterogeneous sources such as relational databases, flat files, etc. This integration enhances the effective analysis of data.

3. **Time Variant** − The data collected in a data warehouse is identified with a particular time period. The data in a data warehouse provides information from the historical point of view.

4. **Non-volatile** − Non-volatile means the previous data is not erased when new data is added to it. A data warehouse is kept separate from the operational database and therefore frequent changes in operational database is not reflected in the data warehouse.

Example:-
Amazon Web Services (AWS)

The whole shift in data storage and warehousing to the cloud over the last several years has been momentous and Amazon has been a market leader in that whole paradigm. Amazon offers a whole ecosystem of data storage tools and resources that complement its cloud services platform. For example, there is Amazon Redshift, a fast, fully managed, petabyte-scale data warehouse cloud solution;AWS Data Pipeline, a web service designed for transporting data between existing AWS data services; and Elastic MapReduce, which provides an easily managed Hadoop solution on top of the AWS services platform.According to Gartner, Amazon was the overall leader in data warehousing customer satisfaction and experience in last year’s survey.

Q5.) Explain the difference between a data warehouse, data mart, ODS.

Data Store Object	Data Warehouse	Data Mart
1) Act as source to populate DW and marts.	Provides management reporting	Specific application or workgroup.
2) Often used for operational reporting.	Summarized Data.	Narrow Scope.
3) Detailed,Atomic Data.	Tuned to optimize query performance.	Customized or stand alone analysis.
4) Huge Data Volumes	Multiple Departments or processes.	Interactive query.
5) Integrated,clean data.	May act as staging area for data marts.	Highly Summarized.
6) Cross-Functional and Cross-Departmental.	Uses dimensional data modeling	Single subject and department oriented.
7) Easy,fast access to relevant data for specific applications.	Historical data form many sourcesstored in one place	Fast queries on smaller amounts of real-time or near-real time data for operations.

Q6.) I am a call centre executive dealing with customers on call. A customer calls me and asks me current information related to his account. I use a tool which references a data warehouse and I try to get the information. Is this the right approach?

Solution:-

Yes, it is a right approach as to extract useful information from a transactional database in an optimal time.The information is accurate and less time complex as its fetching speed is very less.

Q7.) Your team lead wants you to implement an ETL process and as a first step wants you to come with a blue print on how you will go about performing the task.

Solution:-

Focusing on gathering as an object resource as it needs to be added as a format of data the first thing to look into as structured,semi structured and unstructured data then to apply various use cases and test cases.As a blue print the terms like data marts,data cubes etc. Needs to be cleared and as a combination of OLAP and OLTP.

The concept of Busniess Intelligence (Business intelligence comprises the strategies and technologies used by enterprises for the data analysis of business information.)needs to be clear for logic implementation and fetching of relavent results in an optimal time.

Q8.) Detail some challenges faced while performing ETL.

Solution:-