Python Advance Assignment - 3

Name:K.Durga Surya Prasad

Roll No: 19B21A0549

**1.What is the process for loading a dataset from an external source?**

Ans: You can load data using the IMPORT and EXPORT statements and combine that with further post-processing through subselects. You can even specify a statement instead of a table name for the external data source which is executed on that system, for example, to load a certain part of a table into Exasol.

**2. How can we use pandas to read JSON files?**

Ans: To read the files, we use read\_json() function and through it, we pass the path to the JSON file we want to read. Once we do that, it returns a “DataFrame”( A table of rows and columns) that stores data.

**3. Describe the significance of DASK**.

Ans: Dask is a flexible library for parallel computing in Python. Dask is composed of two parts: Dynamic task scheduling optimized for computation. This is similar to Airflow, Luigi, Celery, or Make, but optimized for interactive computational workloads.

**4. Describe the functions of DASK.**

Ans: Dask is a free and open-source library for parallel computing in Python. Dask helps you scale your data science and machine learning workflows. Dask makes it easy to work with Numpy, pandas, and Scikit-Learn, but that's just the beginning.

**5. Describe Cassandra's features.**

Ans: The features of cassandra’s are

* High Scalability. Cassandra is highly scalable which facilitates you to add more hardware to attach more customers and more data as per requirement.
* Rigid Architecture. ...
* Fast Linear-scale Performance. ...
* Fault tolerant. ...
* Flexible Data Storage.