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

• When you load data from an external
source, you load it into a suspense table.
You can then review the data in the
suspense table and modify it.
To load data into the suspense table, position the source file or tape, specify the location of the
source, and run the appropriate load external data process.

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

• 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.

• 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, butoptimized for
interactive computational
workloads.

4.) Describe the functions of DASK.

• 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.

• Apache Cassandra is an open source, user-available, distributed, NoSQL DBMS which is
designed to handle large amounts of data across many servers. It provides zero point of failure.
Cassandra offers massive support
for clusters spanning multiple datacentres.