# What is the process of loading dataset from external souce

## Set the defaults for network resources in the network administration dialog. You must specify the name of a data set that you have already allocated in the Output data set name field. The dialog requires that this value be present so you can update the database. Because the dialog does not permit you to enter only the value for that field, you must enter values for all fields, even though the dialog does not use the information.

## Load the external data into the dialog work table.

## Verify the resource information.

## Process the database update statements to load the data in the work table into the IBM Z Performance and Capacity Analytics database.

# 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. If we want to read a file that is located on remote servers then we pass the link to its location instead of a local path.

Example 1: Reading JSON file

import pandas as pd

df = pd. read\_ json ("FILE\_ JSON. Json ")

df . head()

# Describe the significance 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.

# Describe the functions of DASK?

## Technical: Harness the power of all of the cores on the laptop/workstation in parallel;

## Technical: Support larger-than-memory computation, allowing datasets that fit on disk, but not in RAM

## Social: Invent nothing. We wanted to be as familiar as possible to what users already knew in the Py Data stack.

# Describe Cassandra's features?

## Distributed: Each node in the cluster has has same role.

## Supports replication & Multi data center replication: Replication factor comes with best configurations in cassandra.

## Scalability

## Fault-tolerance

## MapReduce Support

## Query Language