* Getting started with dask
* Overview of dask features
* Data processing using dask dataframes
* Create dask dataframe using csv files
* Get the row and column count
* Overview of data processing APIs of dask dataframes
* Write data in dask dataframe to csv files
* Real world example of data processing using dask
* Exercise and Solution

1. What is the main benefit of using dask?

a. It is faster than NumPy

b. It allows you to work with datasets that are too large to fit into memory

c. It has better visualization capabilities than pandas

d. It is easier to use than PySpark

Answer: b. It allows you to work with datasets that are too large to fit into memory

1. What is the main advantage of dask arrays over NumPy arrays?

a. Dask arrays can handle larger datasets

b. Dask arrays have better performance

c. Dask arrays have more advanced indexing capabilities

d. Dask arrays have better visualization capabilities

Answer: a. Dask arrays can handle larger datasets

1. How can you create a dask dataframe from a CSV file?

a. Using pandas: pd.read\_csv()

b. Using dask: dd.read\_csv()

c. Using NumPy: np.read\_csv()

d. None of the above

Answer: b. Using dask: dd.read\_csv()

1. How can you specify the number of partitions when reading a CSV file into a dask dataframe?

a. Using the nrows parameter

b. Using the blocksize parameter

c. Using the partitions parameter

d. None of the above

Answer: c. Using the partitions parameter

1. What is the difference between using dask.dataframe.read\_csv() and dask.dataframe.from\_pandas() to create a dask dataframe?

a. read\_csv() can handle larger datasets than from\_pandas()

b. from\_pandas() can handle larger datasets than read\_csv()

c. read\_csv() reads data directly from a CSV file, while from\_pandas() reads data from an existing pandas dataframe

d. from\_pandas() reads data directly from a CSV file, while read\_csv() reads data from an existing pandas dataframe

Answer: c. read\_csv() reads data directly from a CSV file, while from\_pandas() reads data from an existing pandas dataframe

1. How can you get the number of columns in a dask dataframe?

a. Using the len() function

b. Using the .count() method

c. Using the .shape[1] attribute

d. None of the above

Answer: c. Using the .shape[1] attribute

1. What is the purpose of the .map() method in dask dataframes?

a. To apply a function to each element of a dask dataframe

b. To apply a function to each row of a dask dataframe

c. To apply a function to each column of a dask dataframe

d. None of the above

Answer: c. To apply a function to each column of a dask dataframe

1. What is the default delimiter used when writing a dask dataframe to a CSV file using the .to\_csv() method?

a. Comma (,)

b. Tab (\t)

c. Semi-colon (;)

d. None of the above

Answer: a. Comma (,)

1. What is a common use case for dask in the field of data science?

a. Processing and analyzing large datasets that do not fit in memory

b. Building machine learning models

c. Creating interactive data visualizations

d. None of the above

Answer: a. Processing and analyzing large datasets that do not fit in memory

1. How can you write a dask dataframe to a CSV file named "output.csv" using the default delimiter?

a. df.write\_csv("output.csv")

b. df.to\_csv("output.csv")

c. df.save\_csv("output.csv")

d. None of the above

Answer: b. df.to\_csv("output.csv")