**Using DASK to Read and Analyze Multiple Files in Parallel**

A folder titled **Activity\_22.2** was first created. The **GenerateFilesWithDask.py** file was downloaded to the **Activity\_22.2** folder.

A screenshot of a computer

Description automatically generated

The **GenerateFilesWithDask.py** file was run, leading to the creation of a **/data** folder containing some large files.

A screenshot of a computer program

Description automatically generated

One navigated out of the **/data** folder, but within the **Activity\_22.2** folder, and a new Python file named **Activity22-2.py** was created.

A screenshot of a computer

Description automatically generated

The necessary DASK libraries were imported into the **Activity22-2.py** file using the specified command.

A screenshot of a computer

Description automatically generated

All of the files generated in the **/data** folder were read using a wildcard. The command to read the CSV files was added, and the data read into the DASK dataframe was displayed.

The **GenerateFilesWithDask.py** Python file was run and a screenshot was provided, showing the head of the DASK dataframe and confirming that the first five rows were displayed correctly.

A screenshot of a computer

Description automatically generated

The data was processed by calculating and displaying the mean of the **x** column using the provided code.

After running the file, a screenshot of the Terminal window was provided, showcasing the output after the computed mean of the dataframe was printed.

A screenshot of a computer

Description automatically generated

The number of columns in the dataframe was computed using the given code.

After entering the code and running the **GenerateFilesWithDask.py** Python file again, a screenshot of the Terminal window was provided to display the number of columns in the dataframe.

A screenshot of a computer

Description automatically generated

The number of rows in the dataframe was computed with the specified code.

Once the code was added and the **GenerateFilesWithDask.py** Python file was run again, a screenshot of the Terminal window was provided, displaying the number of rows in the dataframe.

A screen shot of a computer

Description automatically generated