Activity 22.1: Using DASK to Create Multiple Files in Parallel (40:00)

Start Assignment

Due Nov 13 by 11:29am **Points** 20 **Submitting** a file upload
Available after Nov 6 at 8am

(https://classroom.emeritus.org/courses/9296/modules/items/1556599)

(https://classroom.emeritus.org/courses/9296/modules/items/1556601)

(https://classroom.emeritus.org/courses/9296/modules/items/1556602)

(https://classroom.emeritus.org/courses/9296/modules/items/1556603)

(https://classroom.emeritus.org/courses/9296/modules/items/1556604)

(https://classroom.emeritus.org/courses/9296/modules/items/1556606)

(https://classroom.emeritus.org/courses/9296/modules/items/1556607)

(https://classroom.emeritus.org/courses/9296/modules/items/1556608)

(https://classroom.emeritus.org/courses/9296/modules/items/1556609)

(https://classroom.emeritus.org/courses/9296/modules/items/1556610)

(https://classroom.emeritus.org/courses/9296/modules/items/1556611)

(https://classroom.emeritus.org/courses/9296/modules/items/1556612)

(https://classroom.emeritus.org/courses/9296/modules/items/1556613)

(https://classroom.emeritus.org/courses/9296/modules/items/1556614)

Live Support s://classroom.emeritus.org/courses/9296/modules/items/1556615)

🎯 Learning Outcome Addressed:

  2. Run parallel operations in DASK.

In this activity, you will practice defining a pandas *dataframe* and then writing the contents to files in parallel by using the DASK *library*.

To complete this activity, follow these steps:

  1. In a Terminal window, run the command below to install the DASK *library* locally:

```
pip install dask
```

Provide a screenshot to show that you successfully installed DASK.

2. Download the **Activity 22_1.py (https://classroom.emeritus.org/courses/9296/files/2494004/download)** file and open it using VS Code. Provide a screenshot to show that you opened the file correctly.

3. In the Activity 22_1.py file, use the pandas `DataFrame` *function* to create a *dataframe* with two columns. The first column, `odd_num`, will contain, as entries, all the odd numbers between 1 and 10. The second column, `even_num`, will contain, as entries, all the even numbers between 1 and 10. Assign this *dataframe* to the `pandas_df` variable. Provide a screenshot to show that you created the `pandas_df` *dataframe* correctly.

4. In the Activity 22_1.py file, set the `npartitions` *argument* inside the `from_pandas` *function* equal to 2. Provide a screenshot to show that you set the `npartitions` *argument* equal to 2.

5. Run and debug your code in VS Code. Provide a screenshot to show that you ran the code without any errors.

6. In a Terminal window, navigate to the `activity22.1` folder. List the files inside of the `activity22.1` folder. Provide a screenshot to show that the `0.part` and `1.part` files are present inside of the `activity22.1` folder.

7. In a Terminal window, use the `cat` command to visualize the contents of the `0.part` and `1.part` files. Provide two screenshots. The first screenshot should show the contents of the `0.part` file. The second screenshot should show the contents of the `1.part` file.

You have now completed this activity and practiced defining a pandas *dataframe* and writing the contents to files in parallel by using the DASK *library*.

Submission Instructions:

Your submission for this activity should be a Word document that includes the following screenshots, each labeled for the step that the screenshot represents:

1. Provide a screenshot to show that you successfully installed DASK.
2. Provide a screenshot to show that you opened the file correctly.
3. Provide a screenshot to show that you created the `pandas_df` *dataframe* correctly.
4. Provide a screenshot to show that you set the `npartitions` *argument* equal to 2.
5. Provide a screenshot to show that you ran the code without any errors.
6. Provide a screenshot to show that the `0.part` and `1.part` files are present inside of the `activity22.1` folder.
7. Provide two screenshots. The first screenshot should show the contents of the `0.part` file. The second screenshot should show the contents of the `1.part` file.

Additional Details:

- **Estimated time:** 40 minutes
- *This is a required activity and counts toward course completion.*

Activity 22.1

Criteria	Ratings		Pts
1. Provide a screenshot to show that you successfully installed DASK.	**3 pts** **Complete** The correct screenshot has been included in the submission.	**0 pts** **Incomplete** The screenshot has not been included in the submission or is the incorrect screenshot.	3 pts
2. Provide a screenshot to show that you opened the file correctly.	**2 pts** **Complete** The correct screenshot has been included in the submission.	**0 pts** **Incomplete** The screenshot has not been included in the submission or is the incorrect screenshot.	2 pts
3. Provide a screenshot to show that you created the pandas_df dataframe correctly.	**3 pts** **Complete** The correct screenshot has been included in the submission.	**0 pts** **Incomplete** The screenshot has not been included in the submission or is the incorrect screenshot.	3 pts
4. Provide a screenshot to show that you set the npartitions argument equal to 2.	**3 pts** **Complete** The correct screenshot has been included in the submission.	**0 pts** **Incomplete** The screenshot has not been included in the submission or is the incorrect screenshot.	3 pts
5. Provide a screenshot to show that you ran the code without any errors.	**3 pts** **Complete** The correct screenshot has been included in the submission.	**0 pts** **Incomplete** The screenshot has not been included in the submission or is the incorrect screenshot.	3 pts
6. Provide a screenshot to show that the 0.part and 1.part files are present inside of the activity22.1 folder.	**3 pts** **Complete** The correct screenshot has been included in the submission.	**0 pts** **Incomplete** The screenshot has not been included in the submission or is the incorrect screenshot.	3 pts

Criteria	Ratings		Pts
7. Provide two screenshots. The first screenshot should show the contents of the 0.part file. The second screenshot should show the contents of the 1.part file.	**3 pts** **Complete** Both correct screenshots have been included in the submission.	**0 pts** **Incomplete** Both screenshots have not been included in the submission or are the incorrect screenshots.	3 pts
		Total Points: 20	

Top Questions

😌 😳 😌

It's all empty here!

If you have any questions ask one