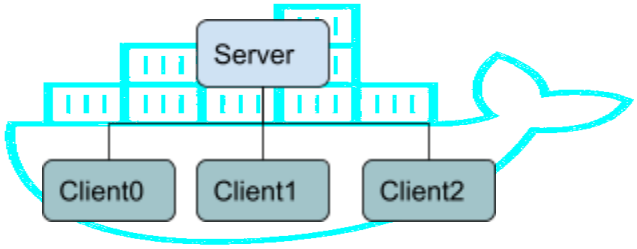
Department of Computer Science & Software Engineering

COMP6231 Fall 2022
Assignment 3:
Docker Containers And MPI

Problem Description

Use docker containers to build a cluster of 4 containers (1 server and 3 clients) to run the following tasks in a distributed manner.



Tasks

Run COMP6231 assignments (1,2) inside a cluster of docker containers.

- **Task1 (T1):** Run Assignment 1 **file server.**
  - Run the file server inside the first container and the client inside the 3 containers then creates the tree structure for each client on the server side.
  - List the IP/Ports used in each container and the shared storage paths if any. **(5 P)**
  - List the commands used to build the docker files and cluster in the proper order. **(5 P)**
  - Explain each used command. **(10 P)**
  - Log the output from the server and clients **(5 P)**.
- **Task2 (T2):** Run assignment 2 **parallel pandas using MPI only**.
  - Run the master node and the 3 workers inside a cluster of 4 containers.
  - The Master node distributes the dataset rows across 3 workers and aggregates the final outputs.
  - List the IP/Ports used in each container and the shared storage paths if any. **(5 P)**
  - List the commands used to build the docker files and cluster in the proper order and report the MPI execution results. **(20 P)**
  - Explain each used command. **(20 P)**
  - Log the output from the master and workers. **(5 P)**

MPI Execution Analysis

Prepare a report that shows the impact of varying the number of containers on **T2** and plot a line graph (number of containers on the x-axis and time taken on the y-axis).

Docker base images:

- Use docker pull python command to pull the python docker base image for task 1.
- Use docker pull husseinabdallah2/mpi4py-cluster:master command to pull the mpi4py image for task 2.

Grading Distribution

- Tasks are worth 75 points (25p for T1 and 50p for T2).
- The analysis report is worth 25 points.

Submission Instructions

- The assignment is due at 11:59 PM on Friday, November 18, 2022.
- Your submission should be a zip file structured as follows:

```
.zip
|___Tasks
        |__T1_file_server
        |   |___file_server_commands_explanation.pdf
        |   |___file_server_ DockerFile.txt
        |__T2_parallel_pandas
            |____MPI_commands_explanation.pdf
            |____MPI_DockerFile.txt
            |____MPI_analysis.pdf
```

- The zip file name should have the format: <first_name>_<last_name>_<ID>_A3.zip (e.g. john_doe_11111111_A3.zip).
- If you need clarification about an unclear part of the assignment, send an email to hussein.abdallah@mail.concordia.ca
- If you require help in programming, please schedule a POD session with your respective tutor and prepare your questions. The tutors may assist you with the programming and APIs but are not able to provide solutions to the assignment.
- This is an **individual** assignment. You are not allowed to copy/share your solutions with your colleagues. Doing so is considered cheating that disqualifies both submissions (0%) and may be reported to the department.

Late Policy

- 0-24 hours late = 25% penalty.
- 24-48 hours late = 50% penalty.
- More than 48 hours late = you lose all the points for this assignment.
- Submissions of corrupted files and blank files will be considered late submissions.