Big Data Platforms homework guide

This document provides a step-by-step guide for getting started with the homework assignments of the Big Data Platforms course. The homework assignments are based on Jupyter Notebooks. You will need to install Docker and run a Jupyter Notebook container to complete the assignments. The following steps will guide you through the process. If you have any questions, please contact the course staff on telegram or via email (big-data-platforms-f25@helsinki.fi).

Step 1: Download and Install Docker

1. **Windows**:
     - Visit the Docker for Windows page.
     - Follow the instructions to download and install Docker Desktop.
     - After installation, launch Docker from the Start menu.
2. **macOS**:
     - Visit the Docker for macOS page.
     - Follow the instructions to download and install Docker Desktop.
     - After installation, launch Docker from the Applications folder.
3. **Linux**:
     - For Linux, the installation steps depend on your distribution. Follow the instructions provided on the Docker for Linux page.
4. **Check your Docker version**
   Whether you just installed Docker or already have it on your machine, verify the version by running: `docker --version` in your terminal. Make sure the version is 20.10 or newer, which is required to run the Jupyter Docker environment for this course.

Step 2: Get released homework files

You could download the homework files from this Google Drive. We will release the homework files based on the schedule of the course.

Step 3: Start Jupyter Notebook Container

1. Open a terminal or command prompt and navigate to the folder containing the downloaded homework files.

2. Pull the docker image with the following command: (You only need to do this once)

   ```
   docker pull hainingt/big_data_platforms:2025
   ```

3. Start a Jupyter Notebook container with the following command: *(Copy and paste the following command as one line — no line breaks!)*

From **Windows Command Prompt (cmd.exe)**:

```
docker run --name bdp_homework --rm -p 8888:8888 -v
%cd%:/home/jovyan/work hainingt/big_data_platforms:2025
```

From **Windows PowerShell**:

```
docker run --name bdp_homework --rm -p 8888:8888 -v
${PWD}:/home/jovyan/work hainingt/big_data_platforms:2025
```

From **macOS / Linux (bash, zsh, etc.)**

```
docker run --name bdp_homework --rm -p 8888:8888 -v
$(pwd):/home/jovyan/work hainingt/big_data_platforms:2025
```

Some notes about the above command:

- In case of you cannot find the local folder in the Jupyter Notebook, please check the permission of the folder. You can change the permission by running the following command before starting the container:

```
chmod 777 -R .
```

- If you're using Docker Desktop on Windows or macOS, check your Docker Desktop settings. In some cases, you may need to explicitly grant Docker Desktop access to certain directories.
- Make sure there is no other Jupyter Notebook container running on port 8888.
- Space or illegal characters in the path of the folder may cause problems. If you encounter any problems, please try to move the folder to another path or use the absolute path of the folder.

4. Open the jupyter notebook in your browser by copying the URL with the token from the terminal or command prompt. The URL should look like this:

```
http://127.0.0.1:8888/?token=<TOKEN_VALUE>
```

Step 4: Access and Save Jupyter Notebook

1. Open a web browser and paste the copied URL with the token.
2. You will be directed to the Jupyter Notebook interface in your browser.
3. Create or open a notebook, work on it, and save the results within the Jupyter environment.
   - Your submissions will be auto-graded, so please make sure your code can be executed correctly. Before submitting your notebook, you can click on "Kernel" in the Jupyter interface

and select "Restart & Run All" to make sure your code can be executed correctly.
- Please don't change the structure of the notebook files. You should only change the code in the code cells with the comment "`YOUR CODE HERE`" or "`YOUR ANSWER HERE`", and you should not add or remove any cells.
- The container should contain all the necessary packages for the homework assignments. Your code should not require any additional packages.
- There are hidden test cases on the server. It is possible that your code can pass all the test cases on your local machine but fail on the server. Please make sure your code is correct to the best of your knowledge.
- Please contact the course staff if you have any questions about the homework assignments or if you encounter any problems with the auto-grading system.

4. The notebook file will be saved in the folder where you started the container. You can also download the notebook file:
- Click on "File" in the Jupyter interface.
- Select "Download as" and choose the desired format (e.g., Notebook (.ipynb)).

Step 5: Stop and Remove Jupyter Notebook Container

1. To stop the container, press `Ctrl+C` in the terminal or command prompt where you started the container.
2. If the container is running in the background, you can stop it with the following command:

```
docker stop bdp_homework
```

The container will be automatically removed when it is stopped.

Step 6: Upload Jupyter Notebook to Mooc Page

1. Go to the big-data-platforms-25 Mooc Page and sign in with your Mooc account.
2. The submission box is in the bottom of the page. Choose the assignment title and upload your notebook file.

Step 7: Receive Feedback

1. Every one hour the server will auto-grade your submission.
2. Your overall score will be shown on the Mooc page.
3. The detailed feedback result will be sent to your email address (the one you used to register the Mooc account) from bigdata.platforms.feedback@gmail.com. Please check your spam folder if you cannot find the email.
4. You can submit multiple times and the system will only record the best score.