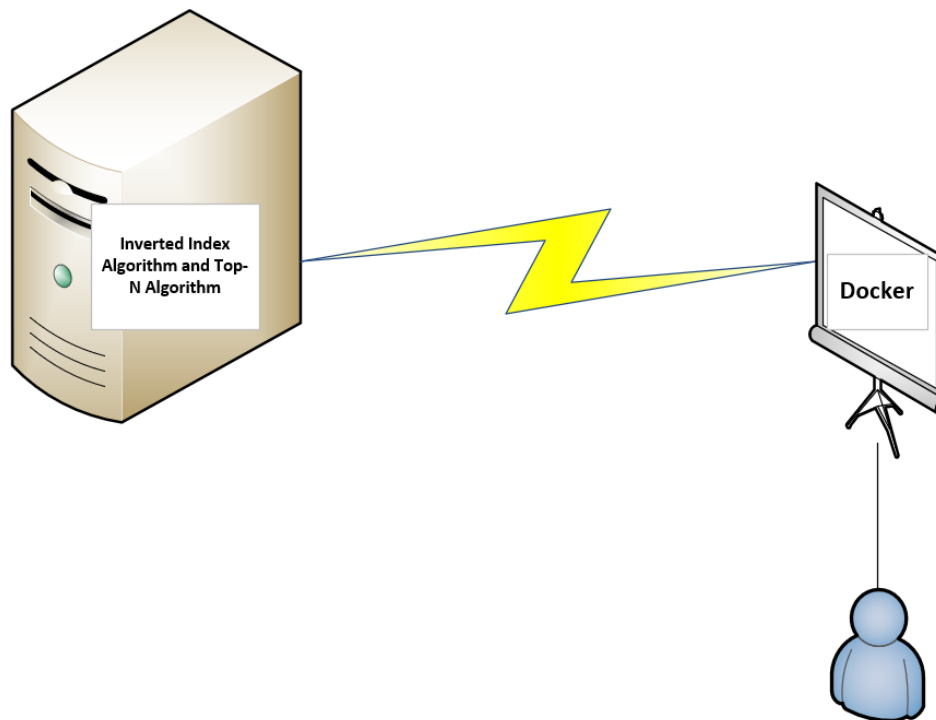


[Fall-2021]

Course Project – Option 2

Deadline: November 21st 11:59PM EST (8:59PM PST)

Mini Search Engine



General Description:

You are asked to build a system that would allow the users to search in large data files with efficiency.

In developing this system, you will build two applications:

1. The first application communicates with the User (accepts user input and displays the output). This application is deployed on a Docker container. This application shouldn't conduct any heavy-processing. Rather, it should contact the second application to conduct any required heavy-processing.
2. The second application processes the requests received from the first application. These requests include constructing Inverting indices (to build indices for large data files) and searching these indices along with finding

the Top-N frequent terms in these indices. This application is deployed on the Google Cloud Platform cluster and it doesn't communicate with users directly.

Example Workflow:

Please check the Mockup PDF that is posted on Canvas. Although the Mockup shows a Graphical User Interface, you may use the terminal for your applications. Building the Graphical User Interface provides extra-credit.

Important Guidelines:

- You may use the data files located in the Data folder on Canvas in testing and validating your applications.
- Use ReadMe.md file on your repository to list any assumptions, steps and any important information to share.
- Keep any private keys for your GCP account outside of the code on GitHub. Instead, submit them on Canvas along with your repository URL.

Submission Guidelines:

- Post URL for your GitHub repository to Canvas. Make sure to keep your GitHub repository public.
- You should complete this project individually. No group-work is offered for this project. However, you are welcome to share ideas. If you are using external references, refer to them.
- Your GitHub repository should have a ReadMe.md file that lists the "exact" steps on how to get this application to work. I will follow the steps in your ReadMe file and if I can't get it running on my machine, I will deduct considerable number of points from your project grade.
- You should record a video demonstrating two elements:
 1. Code Walkthrough while you are explaining your code changes.
 2. Demoing the running application while you are navigating through EVERY functionality that is working in your application. I will use this

video to help assessing your grade. You may lose points for the functionalities that are not demonstrated in the demo.

- Your video size may be large to be uploaded to GitHub. You may use Box to upload the video and add the URL to your ReadMe.md file in your GitHub repository.
 1. Make sure that your video is publicly shared. Private videos won't be visible to the instructor and TAs and therefore, your project grade will be impacted.

Grading Criteria:

- First Java Application Implementation and Execution on Docker: 40% of the total project grade.
- Docker to Local (or GCP) Cluster Communication: 20% of the total project grade
- Inverted Indexing MapReduce Implementation and Execution on the Cluster (GCP) (with stop-list): 20% of the total project grade
- Term Search and Top-N Search (including execution time): 20% of the total project grade
- Extra-credit: building Graphical User Interface for this application: +20% of the total project grade.

Suggested Project Task Schedule (You may run ahead of schedule):

Week	Task
End of Week-4	Build the first application to run on the terminal (or GUI-based if you are seeking the extra-credit).
End of Week-5	Create the docker container and add the application to it.
End of Week-8	Manage the docker-cluster authentication.
End of Week-10	Create the Inverted Indexing algorithm
End of Week-11	1. Complete the 2 nd application 2. Start working on your video recording for the demo and the code walkthrough
End of Week-12	1. Try to finish the GUI for the Application. Otherwise, leave it with the terminal interface. If you got it to work, re-update your video recording with the code walkthrough.

	2. Finish your ReadMe.md file to list all your steps, assumptions, and any information you find important to share.
--	---

Common Penalties:

- Your GitHub repository is not public: 100% reduction (won't be graded)
- Late submissions on Canvas or GitHub: 100% reduction (won't be graded)
- Not submitting the GitHub video (for both code walkthrough and functionality demo) or having the video not publicly shared: 20% penalty (calculated from maximum project grade).
- Not providing clear details in the ReadMe file on how to run the application (or any variables that need to be updated/replaced): 10% penalty (calculated from maximum project grade)