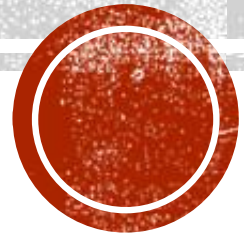
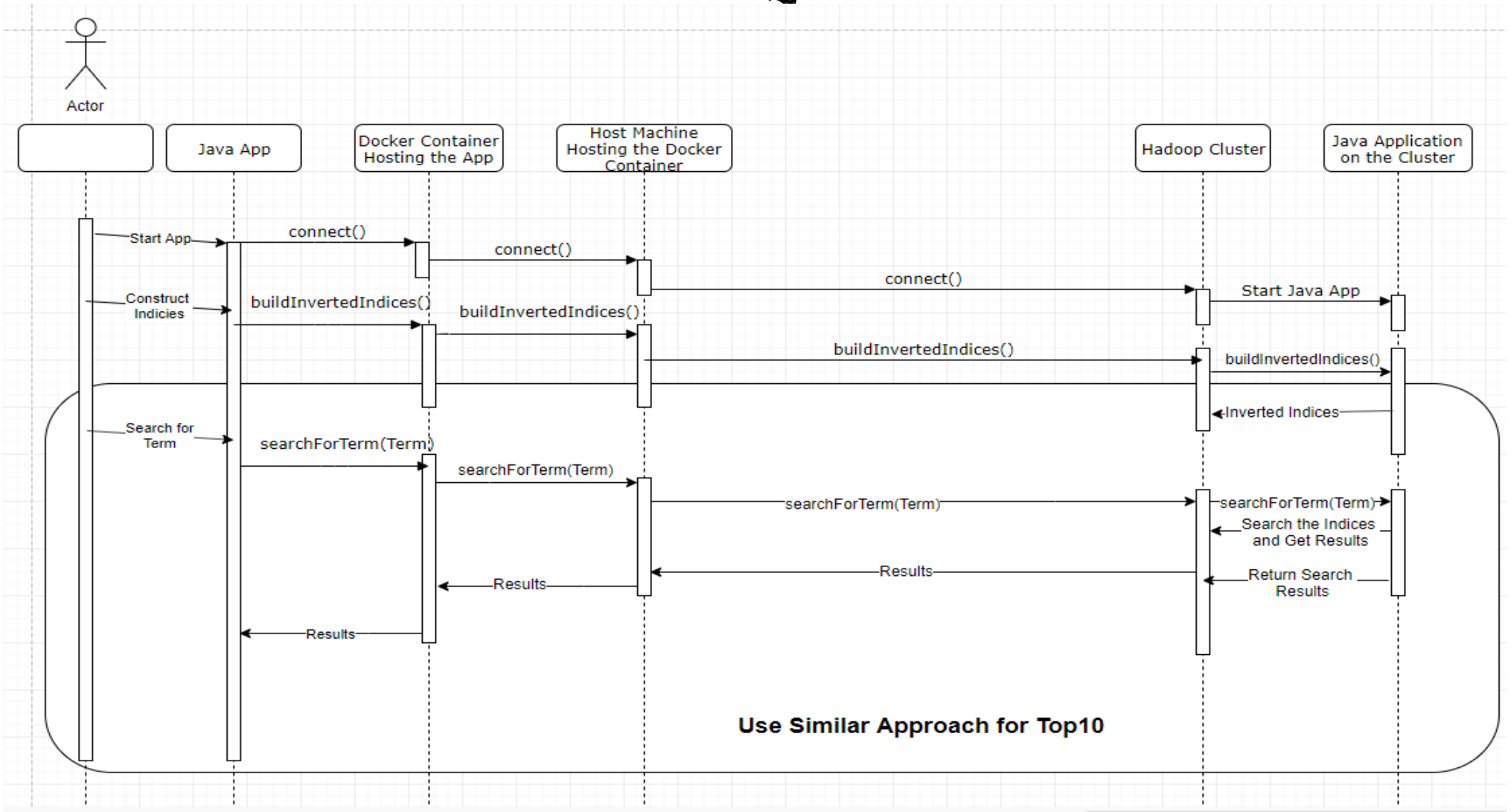
PROJECT-2 DESCRIPTION



PROJECT REQUIREMENTS

- You are required to develop two Java Applications.
  - First application communicates with the User (accepts user input and displays the output). This application is deployed on Docker container. Your GUI should give the users the ability to search for a given term or find the top N-terms.
  - Second application processes User requests (such as constructing Inverting indices and processing user search requests). This application is deployed on the GCP Cluster (or AWS/Azure). This second application doesn't communicate with users directly.
    - Top-N algorithm SHOULD be implemented using Map/Reduce pattern and executed on the cluster for the input that the user will input via the GUI.
- Use the **Mockup** as guide for understanding the project requirements.
- Next page has Proposed Sequence diagram.
  - You don't need to follow the sequence. It's clarification for the communication theme required for the project.
- Graphical User Interface is required for this project.
- Use the data files uploaded in the assignment section.
  - For simplicity, you can limit your application to the data files provided in the attachment section (i.e. you don't need to upload the files from the GUI). However, your application should be smart enough to generate word counts for specific file or all files combined.

PROPOSED PSEUDO-SEQUENCE DIAGRAM



PROJECT GRADING CRITERIA

Item	Points
First Java Application Implementation and Execution on Docker	5 Points
Docker to Local (or GCP) Cluster Communication	2 Points
Inverted Indexing MapReduce Implementation and Execution on the Cluster (GCP)	5 Points
Term and Top-N Search	3 Points
Total	15 Points

EXTRA CREDIT OPTION

Option	Extra Points
Display the returned results of term search and Top-N in Jtable.	2 points

- Notes:
  - You need to apply this extra-credit for for both functionality (term search and Top-N). If you applied it for only one of them, no points are granted.
  - If you implement this extra-credit, please help me by noting that you completed it in your ReadMe file. Unobvious/unmentioned notes are at-risk of not being caught by me.

PROJECT SUBMISSION GUIDELINES

- You should complete this project individually. No group-work is offered for this project.
- You should submit URL for your GitHub Repository containing the project code by April 7th, 11:59 PM EST.
- Your GitHub repository should be **public**. Private repositories won't be graded.
- Your GitHub repository should have a ReadMe.md file that lists the "exact" steps on how to get this application working on a new machine. I will follow the steps 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 application that is working in your application. I will use this video to help assessing your grade. You may lose points for the applications that are not demonstrated in the demo.
- Your video size may be large to be uploaded to GitHub. You may use OneDrive to upload the video and add the URL to your ReadMe.md file in your GitHub repository.
  - Make sure that your video is publicly shared. Private videos won't be visible by the instructor and therefore, your project grade will be **impacted**

PROJECT POSSIBLE PENALTIES

- This list contains a group of common penalties that may be applied:
  - Late submissions on Canvas or GitHub: **100% reduction (won't be graded).**
  - GitHub repository is Private and course instructor doesn't have access to it: **100% reduction (won't be graded)**.
  - Not submitting the GitHub video (for **both** code walkthrough and functionality demo): **50% penalty**.
  - Not providing clear details in the ReadMe file on how to run the applications (or any authentication variables that need to be updated/replaced): **30% penalty (5 points)**

USEFUL LINKS/NOTES

- Those who checked-in their credentials are advised to disable the billing on their GCP clusters to avoid misuse of their accounts (DO NOT make changes to your repository after the deadline or your GitHub repository to avoid me marking it as late submission)
- Authenticate to GCP cluster: https://cloud.google.com/docs/authentication/getting-started
- Filmora Wondershare Screen Recorder: https://filmora.wondershare.com/
- How to record video using Filmora: https://www.youtube.com/watch?v=yjlyvvCOaMc