


Commands :

- For starting the cluster there is no need to do anything special , A docker compose up would be sufficient.
- Also there is a image called data-loader which loads files that should be in the data directory and loads them into hdfs filesystem.
- Also we wrote a shell script to automatically jar the files and run the spark submit. So There is nothing to be done
- So the only thing that one should do is to load the parquet and the csv file into the data directory.

Github-link : <https://github.com/ct-team-spring-2024/project-5>

For videos Go to :  Project-5