**Assignment 1 – How to run my code in HDFS**

**1. Task 1: common word count**

* File structure

|  |  |
| --- | --- |
| Folder | Files |
| assignment1\_task1 | stopwords.txt task1-input1.txt task1-input2.txt |
|  | WordCount.jar |

docker cp WordCount.jar d46320c58d59:WordCount.jar

docker cp assignment1\_task1/stopwords.txt d46320c58d59: assignment1\_task1/stopwords.txt

docker cp assignment1\_task1/task1-input1.txt d46320c58d59:assignment1\_task1/task1-input1.txt

docker cp assignment1\_task1/task1-input2.txt d46320c58d59:assignment1\_task1/task1-input2.txt

* Set up directory and files in HDFS

docker exec -it master bash

hadoop fs -mkdir -p assignment1\_task1

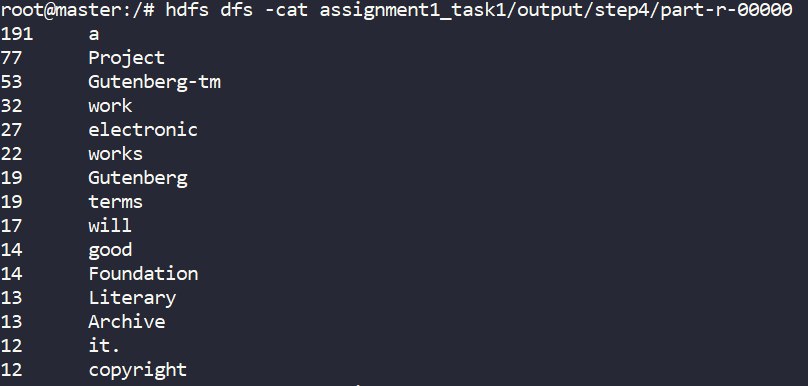
hdfs dfs -put ./assignment1\_task1/\* assignment1\_task1

* Run jar file in Hadoop

hadoop jar WordCount.jar CommonWords assignment1\_task1/task1-input1.txt assignment1\_task1/output/step1 assignment1\_task1/task1-input2.txt assignment1\_task1/output/step2 assignment1\_task1/output/step3 assignment1\_task1/output/step4

* Check output

Output directory is assignment1\_task1/output/step4



**2. Task 2: recommendation system**

* File structure

|  |  |
| --- | --- |
| Folder | Files |
| assignment1\_task2 | data.csv |
|  | RecommendationSystem.jar |

docker cp RecommendationSystem.jar d46320c58d59: RecommendationSystem.jar

docker cp assignment1\_task2/data.csv d46320c58d59: assignment1\_task2/data.csv

* Set up directory and files in HDFS

hadoop fs -mkdir -p assignment1\_task2

hdfs dfs -put ./assignment1\_task2/\* assignment1\_task2

docker cp RecommendationSystem.jar d46320c58d59:RecommendationSystem.jar

* Run jar file in Hadoop

hadoop jar RecommendationSystem.jar Recommend

* Check output

hdfs dfs -cat assignment1\_task2/recommend/step5/part-r-00000