Big Data Assignment 1

**1. HDFS 15 points**

Running any version of Hadoop (HPC Dataproc, docker, AWS, or any), submit screen grabs (a picture in jpg or other suitable format) of the following:

a)  create a directory **in HDFS** with this format: **netid-bd23** (e.g. mine will be ‘jcr365-bd23’). Submit a screen grab of the output of a *Hadoop file listing* showing your home directory and your new directory in it.

Text

Description automatically generated

b)  Create a directory for the homework problem 1.2 (trigram count), and extract all input files into it. Call this directory as follows: hw1.2, e.g. mine will be hw1.2. Submit a picture of directory listings or otherwise show the input files in it.

Sol: Created a directory hw1.2 in the directory sk10689-bd23

Text

Description automatically generated

Put all the input files in the hw1.2 directory that is created.

Text

Description automatically generated

**2. Language Models with MapReduce 85 Points**

Code is done in Java.

Gave the solutions for the HW files and Cat example.

Same for solution 3