Assignment one: Hadoop Examples
Due Date: Oct 13th, 2017
Bring a hardcopy of your solutions to class on Oct 13th!

Download the files in the below link:

https://drive.google.com/open?id=0B1an055wg0hqM3VVWjVGdzBxTjg

Task1: Hadoop-Wordcount Example in Java and Python:

a) Compile and execute the wordcount code in the wordcount/java folder using the Makefile. Show the screenshot of the output.
b) Go to the wordcount/streaming directory. Write the wordcount map and reduce codes in python (may.py and reduce.py). Also edit the Makefile in the same directory to compile your code. Hand-in the code, makefile, and a screenshot of the output.

Task2: Hadoop-Sparse Matrix-Matrix Multiply in Python: Open the matmult directory and

a) Write the hadoop map and reduce files in python to read the input files for matrices A and B from the input folder and print their multiplication. The rank of each matrix is 100 and they are both sparse square matrices.
b) Create a makefile similar to the one in the wordcount example to run your code.
c) Hand-in the code, makefile, and a screenshot of the output.

Task3: Hadoop-Dense Matrix-Matrix Multiply in Python:

d) Write a python code that generates two dense N by N matrices.
e) Write the hadoop map and reduce codes that multiply the two matrices and Hand-in the code, makefile, and the screenshot of the output for an example 10 by 10 matrix.
f) Describe your algorithm. Can you do better? If so, describe your idea/s and implement it. (hint: In a table or graph demonstrate the performance--execution time, number of mappers and reducers, etc.--of your code while increasing N.)
g) **Extra points:** Develop a model for the communication cost of your algorithms. (hint: Model the maximum number of reducers and the network cost of your algorithm based on parameters such as tile-size/block-size.)