**Instructions**

I performed the assignment based on the question to develop MapReduce based word count program in **2 Ways** as below**:**

**Firstly, create the Microsoft Azure account from** [**https://azure.microsoft.com/en-in/free/cloud-services/search/**](https://azure.microsoft.com/en-in/free/cloud-services/search/) **.**

1. **By creating the Microsoft Azure HDInsight Hadoop Cluster and then using the YARN to run Program in Python**

I first created the mapper and reducer program using python Programming Language in software PyCharm and then created the Azure HDInsight Hadoop Cluster to run this program for the local input file using the Hadoop-straming jar file

**Now, to execute in this way first create the Azure HDInsight cluster form**

[**https://portal.azure.com/#create/Microsoft.Template**](https://portal.azure.com/#create/Microsoft.Template)

**Instruction:**

1. Install PyCharm if not installed.
2. Develop MapReduce Word Count Program
3. Create Microsoft Azure HDInsight Cluster
4. Install OpenSSH client and server if not installed using Windows Powershell
5. Run command

yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar \

-files mapper.py,reducer.py \

-mapper mapper.py \

-reducer reducer.py \

-input /example/data/wordcountinput.txt \

-output /example/wordcountoutput

1. **By installing Hadoop & Java on Windows 10 and then using in-build hadoop-mapreduce-examples.jar**

I installed Hadoop & Java on Windows 10 first and then used the in-build hadoop-mapreduce-examples.jar to run the wordcount program

**Instructions:**

1. Install 7 Zip
2. Install Java 8
3. Install Hadoop 3.2.2
4. Start all hadoop daemons
5. Create input file
6. Put input file into hdfs
7. Run the program using below command:

“hadoop jar hadoop-mapreduce-examples-3.2.2.jar wordcount /Users/Komal/WordCountInput.txt /Users/Komal/WordCountOutput.txt”

1. View output using below command

“hadoop dfs -cat /Users/Komal/WordCountOutput.txt/part-r-00000”