**MapReduce job run in Google Dataproc**

1. Prepared PySpark script WordCount.py

from pyspark import SparkContext

sc =SparkContext()

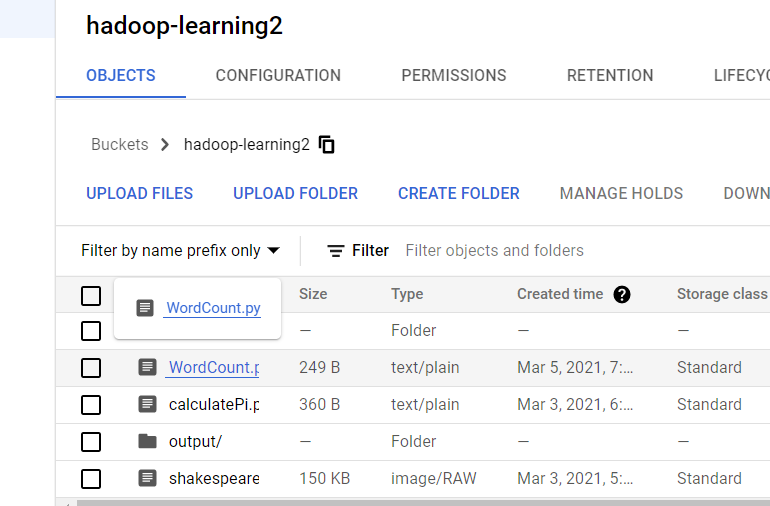
file = "gs://hadoop-learning2/shakespeare.raw"

words = sc.textFile(file).flatMap(lambda line: line.split(" "))

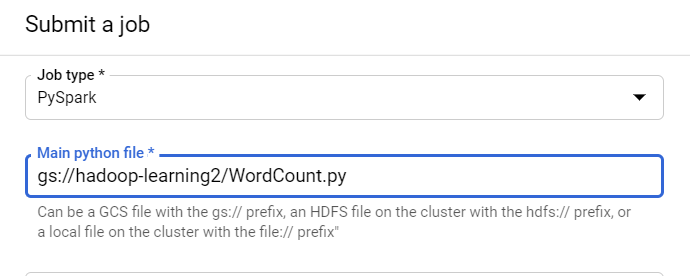
wordCounts = words.map(lambda word: (word, 1)).reduceByKey(lambda a,b:a +b)

wordCounts.saveAsTextFile("gs://hadoop-learning2/output2/")

1. Uploaded the script file and data file to Google Cloud bucket



1. Submit a job with the files in Google Dataproc



1. Job complete
2. Results generated successfully.

