ReadMe

Initial Configuration:

- Programming Language : Python
- Execution Environment : Google Cloud
- Programming Framework : Apache Spark

File Attached:

1. **WordCountAssignment :** Contains code for all the three parts of the assignment.
2. **Part1Output.txt** : Sample output file for part 1 of the assignment
3. **Part2Output.txt** : Sample Output file for part 2 of the assignment
4. **Part3Output.txt** : Output file for part 3 of the assignment
5. **Results and Snapshots**: Contains results and associated snapshots. The results were received as expected.

Please note :

The no of output files for Part 1 of the assignment is 14. Each file is 60KB each except last one
The no of output files for Part 2 of the assignment is 48. Each file is 60KB each except last one.

Technical Specifications:

I used pyspark library of python for writing the code. For the purpose of solving above problems i made use of functions like map(), collect(), join(), reduceByKey() etc. The detailed explanation is documented in the WordCountAssignment and results and snapshot file.