Name: Khoa Cu Dang (Cody) Cao
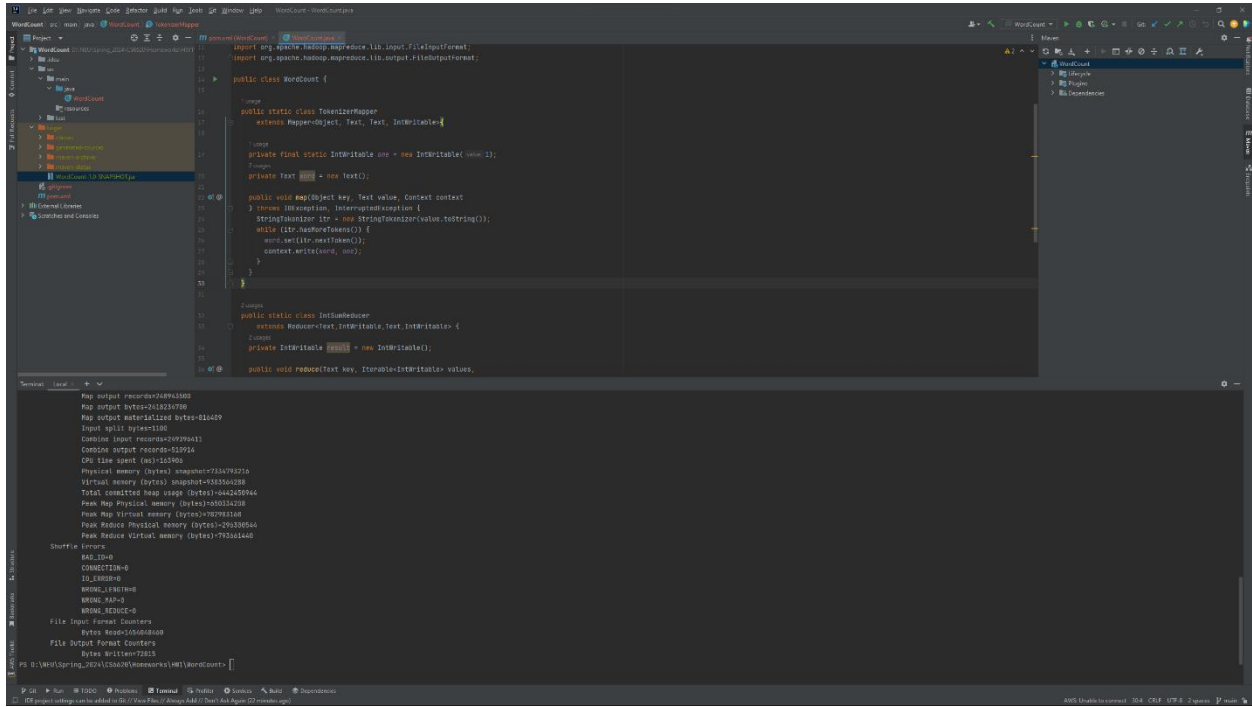
Course: CS6620 Spring 2024

Homework 1 Report

Local IDE Results

AWS Results

EMR Cluster Configuration

Step Execution Result

		Step ID		Status		Name		Log files ↗	Creation time (UTC-08:00) ▼	Start time (UTC-08:00) ▼	Elapsed time
☐	⊟	s-0653764279X26LYLZ23Z		⊘ Completed		WordCount JAR		controller syslog stderr stdout ↻	February 5, 2024 at 14:09	February 5, 2024 at 14:09	5 minutes, 46 seconds

Jar location	Permissions	Main class
s3://cs6240-hw1-cody/WordCount-1.0-SNAPSHOT.jar	-	-

Action on failure	Argument
Continue	🗐 WordCount s3://cs6240-hw1-cody/input/hw1.txt s3://cs6240-hw1-cody/output/

Input file on S3

Amazon S3 > Buckets > cs6240-hw1-cody > input/

input/   🗐 Copy S3 URI

Objects Properties

Objects (1) Info   ↻ 🗐 Copy S3 URI 🗐 Copy URL ⬇ Download Open ↗ Delete Actions ▼ Create folder 🗐 Upload

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 Inventory ↗ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more ↗

🔍 Find objects by prefix   < 1 > ⚙

☐	Name ▲	Type	Last modified ▼	Size ▼	Storage class ▼
☐	🗋 hw1.txt	txt	January 29, 2024, 14:03:22 (UTC-08:00)	1.4 GB	Standard

Output Directory on S3

Amazon S3 > Buckets > cs6240-hw1-cody > output/

output/   🗐 Copy S3 URI

Objects Properties

Objects (4) Info   ↻ 🗐 Copy S3 URI 🗐 Copy URL ⬇ Download Open ↗ Delete Actions ▼ Create folder 🗐 Upload

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 Inventory ↗ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more ↗

🔍 Find objects by prefix   < 1 > ⚙

☐	Name ▲	Type	Last modified ▼	Size ▼	Storage class ▼
☐	🗋 _SUCCESS	-	February 5, 2024, 14:15:30 (UTC-08:00)	0 B	Standard
☐	🗋 part-r-00000	-	February 5, 2024, 14:15:25 (UTC-08:00)	23.3 KB	Standard
☐	🗋 part-r-00001	-	February 5, 2024, 14:15:25 (UTC-08:00)	23.8 KB	Standard
☐	🗋 part-r-00002	-	February 5, 2024, 14:15:28 (UTC-08:00)	24.0 KB	Standard

JAR file location on S3

Amazon S3 > Buckets > cs6240-hw1-cody

cs6240-hw1-cody Info

Objects Properties Permissions Metrics Management Access Points

Objects (5) Info   ↻ 🗐 Copy S3 URI 🗐 Copy URL ⬇ Download Open ↗ Delete Actions ▼ Create folder 🗐 Upload

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 Inventory ↗ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more ↗

🔍 Find objects by prefix   < 1 > ⚙

☐	Name ▲	Type	Last modified ▼	Size ▼	Storage class ▼
☐	🗋 bootstrap_debug.sh	sh	February 4, 2024, 23:35:10 (UTC-08:00)	22.0 B	Standard
☐	🗀 input/	Folder	-	-	-
☐	🗀 logs/	Folder	-	-	-
☐	🗀 output/	Folder	-	-	-
☐	🗋 WordCount-1.0-SNAPSHOT.jar	jar	February 5, 2024, 13:58:56 (UTC-08:00)	4.6 KB	Standard

Last lines of syslog.txt

```
syslog.txt                                    ×        +

File    Edit    View

2024-02-05 22:15:21,950 INFO org.apache.hadoop.mapreduce.Job (main):  map 98% reduce 20%
2024-02-05 22:15:22,953 INFO org.apache.hadoop.mapreduce.Job (main):  map 98% reduce 21%
2024-02-05 22:15:23,956 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 21%
2024-02-05 22:15:25,961 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 67%
2024-02-05 22:15:28,973 INFO org.apache.hadoop.mapreduce.Job (main):  map 100% reduce 100%
2024-02-05 22:15:30,985 INFO org.apache.hadoop.mapreduce.Job (main): Job job_1707169248635_0002 completed successfully
2024-02-05 22:15:31,228 INFO org.apache.hadoop.mapreduce.Job (main): Counters: 61
        File System Counters
                FILE: Number of bytes read=6025225
                FILE: Number of bytes written=12766216
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2178
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=22
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
                HDFS: Number of bytes read erasure-coded=0
                S3: Number of bytes read=1454182711
                S3: Number of bytes written=72815
                S3: Number of read operations=0
                S3: Number of large read operations=0
                S3: Number of write operations=0
        Job Counters
                Killed map tasks=1
                Killed reduce tasks=1
                Launched map tasks=22
                Launched reduce tasks=3
                Data-local map tasks=22
                Total time spent by all maps in occupied slots (ms)=20207016
                Total time spent by all reduces in occupied slots (ms)=6692032
                Total time spent by all map tasks (ms)=841959
                Total time spent by all reduce tasks (ms)=209126
                Total vcore-milliseconds taken by all map tasks=841959
                Total vcore-milliseconds taken by all reduce tasks=209126
                Total megabyte-milliseconds taken by all map tasks=646624512
                Total megabyte-milliseconds taken by all reduce tasks=214145024
        Map-Reduce Framework
                Map input records=21907700
                Map output records=248943500
                Map output bytes=2418234700
                Map output materialized bytes=1916143
                Input split bytes=2178
                Combine input records=248943500
                Combine output records=232012
                Reduce input groups=5273
                Reduce shuffle bytes=1916143
                Reduce input records=232012
                Reduce output records=5273
                Spilled Records=696036
                Shuffled Maps =66
                Failed Shuffles=0
                Merged Map outputs=66
                GC time elapsed (ms)=14964
                CPU time spent (ms)=410400
                Physical memory (bytes) snapshot=11395403776
                Virtual memory (bytes) snapshot=58716188672
                Total committed heap usage (bytes)=6135898112
                Peak Map Physical memory (bytes)=490098688
                Peak Map Virtual memory (bytes)=2315685888
                Peak Reduce Physical memory (bytes)=279863296
                Peak Reduce Virtual memory (bytes)=2599497728
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1454182711
        File Output Format Counters
                Bytes Written=72815
```