CSE488: Big Data Analytics [SPRING 2023]

Assignment 5

Submitted by:

Student ID: 2020-1-60-215
Student Name: Md. Abdul Ahad Rifat

Browse and run installed applications

Java - WordCount/src/WordCount.java - Eclipse

File Edit Source Refactor Navigate Search Project Run Window Help

Quick Access   Java

Package   WordCount.java   MapForWordCount.java   ReduceForWordCount.java

```java
15      }Configuration config = new Configuration();
16      Path input = new Path(args[0]);
17      Path output = new Path(args[1]);
18
19      //Instantiate a Job object for your job's configuration.
20      @SuppressWarnings("deprecation")
21      Job job = new Job(config, "WordCount");
22      /*
23       * Specify the jar file that contains your driver, mapper, and reducer.
24       * Hadoop will transfer this jar file to nodes in your cluster running
25       * mapper and reducer tasks.
26      */job.setJarByClass(WordCount.class);
27      job.setMapperClass(MapForWordCount.class);
28      job.setReducerClass(ReduceForWordCount.class);
29      job.setOutputKeyClass(Text.class);
30      job.setOutputValueClass(IntWritable.class);
31      FileInputFormat.addInputPath(job, input);
32      FileOutputFormat.setOutputPath(job, output);
33      /*
34       * Start the MapReduce job and wait for it to finish.
35       * If it finishes successfully, return 0. If not, return 1.
36      */ boolean success = job.waitForCompletion(true);
37      System.exit(success ? 0 : 1);
38      }
39      }
```

- SalesEmployeeFinder
- training
- WordCount
  - src
    - (default packag
      - MapForWord
      - ReduceForWo
      - WordCount.ja
  - JRE System Library
  - Referenced Librari
  - conf

Problems @ Javadoc Declaration Console

No consoles to display at this time.

Writable   Smart Insert   39 : 4

Hue - File Browser - p...   Java - WordCount/src/...   cloudera@quickstart:~

Browse and run installed applications

Java - WordCount/src/MapForWordCount.java - Eclipse

File Edit Source Refactor Navigate Search Project Run Window Help

Quick Access   Java

Package   WordCount.java   MapForWordCount.java   ReduceForWordCount.java

```java
4   import org.apache.hadoop.io.LongWritable;
5   import org.apache.hadoop.io.Text;
6   import org.apache.hadoop.mapreduce.Mapper;
7
8   public class MapForWordCount extends Mapper<LongWritable, Text, Text, IntWritable> { @Override
9       public void map(LongWritable key, Text value, Context context)
10              throws IOException, InterruptedException {
11
12          String line = value.toString();
13          String[] words = line.split(",");
14          for(int i=0;i<words.length-1;i++){
15              for(int j=i+1;j<words.length;j++){
16                  if((words[i].charAt(0)<(words[j].charAt(0))){
17                      context.write(new Text(words[i]+", "+words[j]),new IntWritable(1));
18                  }
19                  else{
20                      context.write(new Text(words[j]+", "+words[i]),new IntWritable(1));
21                  }
22
23              }
24
25          }
26
27      }}
28
```

- SalesEmployeeFinder
- training
- WordCount
  - src
    - (default packag
      - MapForWord
      - ReduceForWo
      - WordCount.ja
  - JRE System Library
  - Referenced Librari
  - conf

Problems @ Javadoc Declaration Console

No consoles to display at this time.

Writable   Smart Insert   28 : 1

Hue - File Browser - p...   Java - WordCount/src/...   cloudera@quickstart:~

Access documents, folders and network places

Java - WordCount/src/ReduceForWordCount.java - Eclipse

File   Edit   Source   Refactor   Navigate   Search   Project   Run   Window   Help

Quick Access   Java

Package

- SalesEmployeeFinder
- training
- WordCount
  - src
    - (default package)
      - MapForWord
      - ReduceForWo
      - WordCount.ja
  - JRE System Library
  - Referenced Librari
  - conf

WordCount.java   MapForWordCount.java   ReduceForWordCount.java
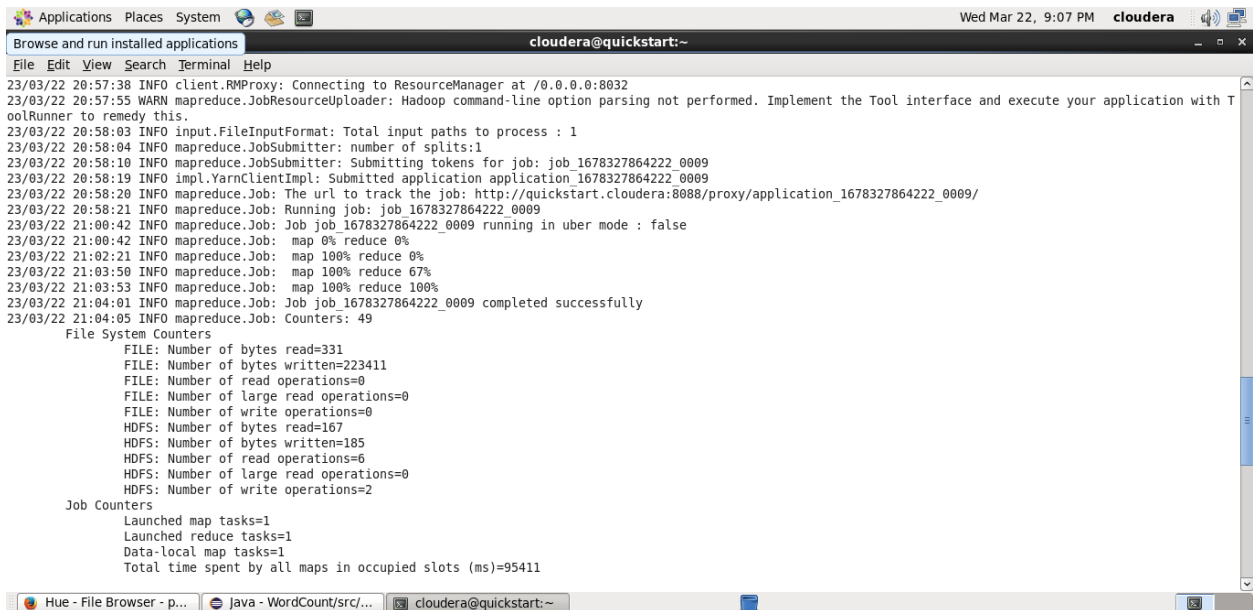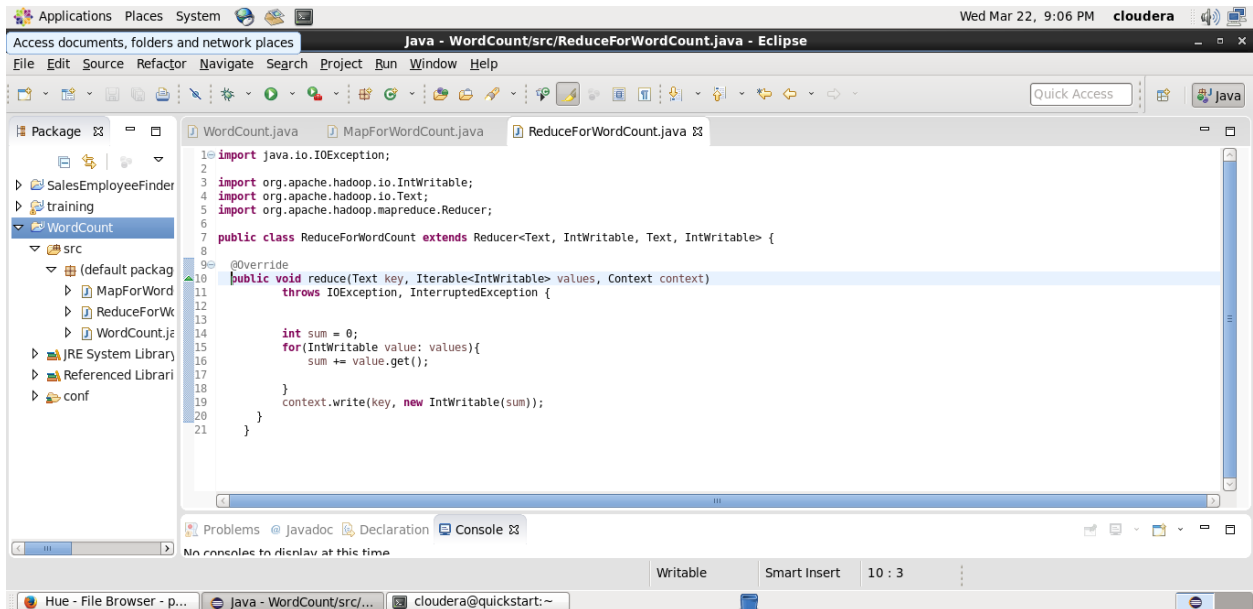
```java
1 import java.io.IOException;
2
3 import org.apache.hadoop.io.IntWritable;
4 import org.apache.hadoop.io.Text;
5 import org.apache.hadoop.mapreduce.Reducer;
6
7 public class ReduceForWordCount extends Reducer<Text, IntWritable, Text, IntWritable> {
8
9     @Override
10     public void reduce(Text key, Iterable<IntWritable> values, Context context)
11             throws IOException, InterruptedException {
12
13
14         int sum = 0;
15         for(IntWritable value: values){
16             sum += value.get();
17
18         }
19         context.write(key, new IntWritable(sum));
20     }
21 }
```

Problems   Javadoc   Declaration   Console

No consoles to display at this time

Writable   Smart Insert   10 : 3

Hue - File Browser - p...   Java - WordCount/src/...   cloudera@quickstart:~

Browse and run installed applications

cloudera@quickstart:~

File   Edit   View   Search   Terminal   Help

```
23/03/22 20:57:38 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/03/22 20:57:55 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with T
oolRunner to remedy this.
23/03/22 20:58:03 INFO input.FileInputFormat: Total input paths to process : 1
23/03/22 20:58:04 INFO mapreduce.JobSubmitter: number of splits:1
23/03/22 20:58:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1678327864222_0009
23/03/22 20:58:19 INFO impl.YarnClientImpl: Submitted application application_1678327864222_0009
23/03/22 20:58:20 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1678327864222_0009/
23/03/22 20:58:21 INFO mapreduce.Job: Running job: job_1678327864222_0009
23/03/22 21:00:42 INFO mapreduce.Job: Job job_1678327864222_0009 running in uber mode : false
23/03/22 21:00:42 INFO mapreduce.Job:  map 0% reduce 0%
23/03/22 21:02:21 INFO mapreduce.Job:  map 100% reduce 0%
23/03/22 21:03:50 INFO mapreduce.Job:  map 100% reduce 67%
23/03/22 21:03:53 INFO mapreduce.Job:  map 100% reduce 100%
23/03/22 21:04:01 INFO mapreduce.Job: Job job_1678327864222_0009 completed successfully
23/03/22 21:04:05 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=331
                FILE: Number of bytes written=223411
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=167
                HDFS: Number of bytes written=185
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=95411
```

Hue - File Browser - p...   Java - WordCount/src/...   cloudera@quickstart:~

Hue - File Browser - part-r-00000 - File Viewer - Mozilla Firefox

Cloudera Live : Welcom...   ×   Google Account   ×   Hue - File Browser - ...   ×

quickstart.cloudera:8888/filebrowser/view=/user/cloudera/hjf/part-r-00000#p1   Search

Cloudera   Hue   Hadoop   HBase   Impala   Spark   Solr   Oozie   Cloudera Manager   Getting Started

HUE   Query Editors ⌄   Data Browsers ⌄   Workflows ⌄   Search   Security ⌄   File Browser   Job Browser   cloudera ⌄

📄 File Browser

ACTIONS

▏▎ View as binary

✏ Edit file

⬇ Download

📄 View file location

↻ Refresh

INFO

Last modified
March 22, 2023 9:03
p.m.
User
cloudera

🏠 Home   /   user /   cloudera /   hjf /   **part-r-00000**   Page   1   of 1

```
B, C   1
B, D   1
B, K   1
C,  A   1
C,  D   1
D,  A   2
D,  B   1
D,  K   1
E,   A   2
E,   B   1
E,   C   1
E,  C   1
E,  D   1
```

Hue - File Browser - p...   🔵 Java - WordCount/src/...   cloudera@quickstart:~