Hadeel Abdeljalil

HDFS commands:

hadoop fs -mkdir /user/cloudera/sportshop

hadoop fs -mkdir /user/cloudera/sportshop/input

I copied the data provided and added it to a file called salesData.txt to /home/cloudera =>file path

hadoop fs -put salesData.txt /user/cloudera/sportshop/input/

hadoop fs -cat /user/cloudera/sportRetailer/input/salesData.txt

Using HDFS commands, show to do the following:

• run the job

hadoop jar test.jar test. SalesAnalysis /user/cloudera/sportshop/input /user/cloudera/sportshop/output

Note: I added SalesAnalyis at test packege

 • browse the content of 'output' directory

hadoop fs -ls /user/cloudera/sportshop

• show the content of the output files

Hadoop fs -cat /user/cloudera/sportshop/output

the full code of the MapReduce job:

import org.apache.hadoop.conf.Configuration;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.io.LongWritable;

import org.apache.hadoop.mapreduce.Job;

import org.apache.hadoop.mapreduce.Mapper;

import org.apache.hadoop.mapreduce.Reducer;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;

import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

```java
import java.io.IOException;

public class SalesAnalysis {

    // First Job Mapper
    public static class SalesMapper extends Mapper<Object, Text, Text, LongWritable> {
        private final static LongWritable totalSalesAmount = new LongWritable();
        private Text retailerCity = new Text();

        @Override
        public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
            if (value.toString().startsWith("Retailer")) {
                return;
            }

            String[] fields = value.toString().split(",");
            if (fields.length == 6) {
                try {
                    String retailer = fields[0].trim();
                    String city = fields[2].trim();
                    int pricePerUnit = Integer.parseInt(fields[4].replace("$", "").replace(",", "").trim());
                    int unitsSold = Integer.parseInt(fields[5].replace(",", "").trim());
                    long salesAmount = (long) pricePerUnit * unitsSold;

                    retailerCity.set(retailer + ", " + city);
                    totalSalesAmount.set(salesAmount);

                    context.write(retailerCity, totalSalesAmount);
                } catch (NumberFormatException e) {
                    System.err.println("Error parsing record: " + value.toString() + " " + e.getMessage());
```

```java
                }
            }
        }
    }


    // First Job Reducer
    public static class SalesReducer extends Reducer<Text, LongWritable, Text, LongWritable> {

        private LongWritable result = new LongWritable();


        @Override
        public void reduce(Text key, Iterable<LongWritable> values, Context context) throws IOException,
InterruptedException {
            long sum = 0;
            for (LongWritable val : values) {
                sum += val.get();
            }
            result.set(sum);
            context.write(key, result);
        }
    }


    // Second Job Mapper
    public static class SortMapper extends Mapper<Object, Text, LongWritable, Text> {
        private Text retailerCity = new Text();
        private LongWritable salesAmount = new LongWritable();


        @Override
        public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
            String[] fields = value.toString().split("\t");
            if (fields.length == 2) {
                retailerCity.set(fields[0]);
```

```java
            salesAmount.set(Long.parseLong(fields[1]));


            // The negative sign makes the LongWritable value descending

            context.write(new LongWritable(-salesAmount.get()), retailerCity);

        }

    }

}


    // Second Job Reducer

    public static class SortReducer extends Reducer<LongWritable, Text, Text, LongWritable> {

        @Override

        public void reduce(LongWritable key, Iterable<Text> values, Context context) throws IOException,
InterruptedException {

            for (Text val : values) {

                // Negate the key back to its original positive value

                context.write(val, new LongWritable(-key.get()));

            }

        }

    }


    public static void main(String[] args) throws Exception {

        // Configuration and First Job

        Configuration conf = new Configuration();

        Job job1 = Job.getInstance(conf, "sales sum");


        job1.setJarByClass(SalesAnalysis.class);

        job1.setMapperClass(SalesMapper.class);

        job1.setCombinerClass(SalesReducer.class);

        job1.setReducerClass(SalesReducer.class);


        job1.setOutputKeyClass(Text.class);
```

```java
        job1.setOutputValueClass(LongWritable.class);

        FileInputFormat.addInputPath(job1, new Path(args[0]));
        FileOutputFormat.setOutputPath(job1, new Path(args[1] + "/temp"));

        boolean success = job1.waitForCompletion(true);
        if (!success) {
            System.exit(1);
        }

        // Configuration and Second Job
        Job job2 = Job.getInstance(conf, "sort by sales amount");

        job2.setJarByClass(SalesAnalysis.class);
        job2.setMapperClass(SortMapper.class);
        job2.setReducerClass(SortReducer.class);

        job2.setOutputKeyClass(LongWritable.class);
        job2.setOutputValueClass(Text.class);

        FileInputFormat.addInputPath(job2, new Path(args[1] + "/temp"));
        FileOutputFormat.setOutputPath(job2, new Path(args[1] + "/final"));

        System.exit(job2.waitForCompletion(true) ? 0 : 1);
    }
}
```
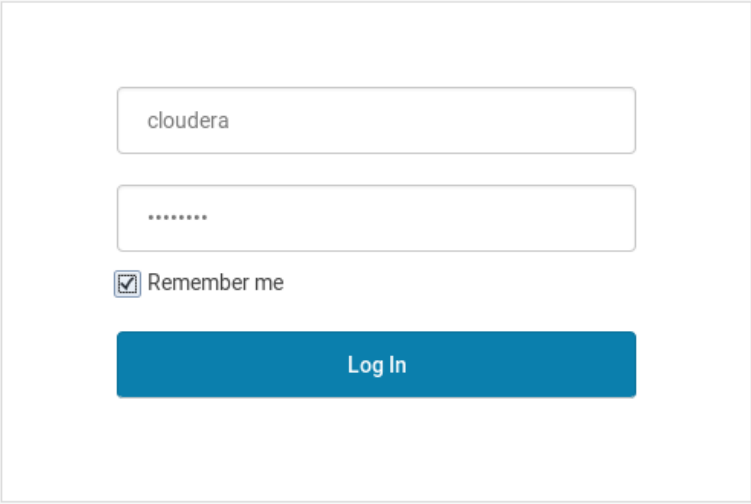
```
                          cloudera@quickstart:~                      _  □  ×
File  Edit  View  Search  Terminal  Help
[QuickStart] Shutting down CDH services via init scripts...
kafka-server: unrecognized service
[QuickStart] Disabling CDH services on boot...
error reading information on service kafka-server: No such file or directory
[QuickStart] Starting Cloudera Manager server...
[QuickStart] Waiting for Cloudera Manager API...
[QuickStart] Starting Cloudera Manager agent...
[QuickStart] Configuring deployment...
Submitted jobs: 81
[QuickStart] Deploying client configuration...
Submitted jobs: 82
[QuickStart] Starting Cloudera Management Service...
Submitted jobs: 90
[QuickStart] Enabling Cloudera Manager daemons on boot...


Success! You can now log into Cloudera Manager from the QuickStart VM's browser:

    http://quickstart.cloudera:7180

    Username: cloudera
    Password: cloudera

[cloudera@quickstart ~]$
```


Start HDFS & yarn(mapReduce), check health*

| Hadoop | Overview | Datanodes | Datanode Volume Failures | Snapshot | Startup Progress | Utilities ▾ |

Overview 'quickstart.cloudera:8020' (active)

Started:	Sat Aug 24 07:37:42 -0700 2024
Version:	2.6.0-cdh5.13.0, r42e8860b182e55321bd5f5605264da4adc8882be
Compiled:	Wed Oct 04 11:08:00 -0700 2017 by jenkins from Unknown
Cluster ID:	CID-a24185f9-a545-40fe-9553-84c3fdca489f
Block Pool ID:	BP-1067413441-127.0.0.1-1508775264580

Summary

Security is off.

Safemode is off.

DFS Remaining:	42.57 GB (78.17 %)
Block Pool Used:	832.38 MB (1.49%)
DataNodes usages% (Min/Median/Max/stdDev):	1.49% / 1.49% / 1.49% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	Sat Aug 24 07:37:42 -0700 2024

---

hadoop — All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	3 GB	0 B	0	3	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	0	0	0

User Metrics for dr.who

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Show 20 entries

Search:

ID ▾	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	Progress	Tracking UI
							No data available in table								

Showing 0 to 0 of 0 entries

First Previous Next Last

Cluster
- About
- Nodes
- Applications
  - NEW
  - NEW_SAVING
  - SUBMITTED
  - ACCEPTED
  - RUNNING
  - FINISHED
  - FAILED
  - KILLED
- Scheduler

Tools

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

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
2	0	0	2	0	0 B	3 GB	0 B	0	3	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	0	0	0

User Metrics for dr.who

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Show 20 ⌄ entries   Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	Progress	Tracking UI
application_1724510262132_0002	cloudera	sort by sales amount	MAPREDUCE	root.users.cloudera	Sat Aug 24 07:52:19 -0700 2024	Sat Aug 24 07:53:22 -0700 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A		History
application_1724510262132_0001	cloudera	sales sum	MAPREDUCE	root.users.cloudera	Sat Aug 24 07:49:55 -0700 2024	Sat Aug 24 07:52:17 -0700 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A		History

Showing 1 to 2 of 2 entries   First Previous 1 Next Last