BDA Lab (Hadoop 1) – 1BM17CS086

Login as `hduser`

Password: `bmsce`

```
su hduser
```

```
ubuntu@ip-172-31-90-179:~/lab1$ su hduser
Password:
hduser@ip-172-31-90-179:/home/ubuntu/lab1$ cd /usr/local/hadoop/sbin/
```

change directory to `/usr/local/hadoop/sbin/`

Run these scripts:

`start-dfs.sh`

`start-yarn.sh`

```
cd /usr/local/hadoop/sbin/
./start-dfs.sh
./start-yarn.sh
jps
```

```
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ ./start-dfs.sh
20/11/26 18:14:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-ip-17
2-31-90-179.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-ip-17
2-31-90-179.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-seconda
rynamenode-ip-172-31-90-179.out
20/11/26 18:15:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-ip-1
72-31-90-179.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-i
p-172-31-90-179.out
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ jps
4370 Jps
3464 NameNode
4298 NodeManager
3643 DataNode
3995 ResourceManager
3852 SecondaryNameNode
```

create a new file `file1.txt` in the
home folder:

```
cd
nano file1.txt
cat file1.txt
```

```
hi how are you
how is your job
how is your family
how is your brother
how is your sister
```

Execute the following commands:

```
hadoop fs -ls /

hadoop fs -mkdir /rgs1

hadoop fs -copyFromLocal /home/hduser/file1.txt /rgs1/test.txt

hadoop jar /home/ubuntu/lab1/wordcount.jar WordCount /rgs1/test.txt /rgs1/output

hadoop fs -ls /rgs1/output

hadoop fs -cat /rgs1/output/part-r-00000
```

```
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ cd
hduser@ip-172-31-90-179:~$ nano file1.txt
hduser@ip-172-31-90-179:~$ cat file1.txt
hi how are you
how is your job
how is your family
how is your brother
how is your sister
hduser@ip-172-31-90-179:~$ hadoop fs -ls /
20/11/26 18:16:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
hduser@ip-172-31-90-179:~$ hadoop fs -mkdir /rgs1
20/11/26 18:16:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
hduser@ip-172-31-90-179:~$ hadoop fs -copyFromLocal /home/hduser/file1.txt /rgs1/test.txt
20/11/26 18:17:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
```

```
hduser@ip-172-31-90-179:~$ hadoop jar /home/ubuntu/lab1/wordcount.jar WordCount /rgs1/test.t
xt /rgs1/output
20/11/26 18:18:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
20/11/26 18:18:35 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs
.metrics.session-id
20/11/26 18:18:35 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker,
 sessionId=
20/11/26 18:18:36 INFO input.FileInputFormat: Total input paths to process : 1
20/11/26 18:18:36 INFO mapreduce.JobSubmitter: number of splits:1
20/11/26 18:18:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local887129590
_0001
20/11/26 18:18:36 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
20/11/26 18:18:36 INFO mapreduce.Job: Running job: job_local887129590_0001
20/11/26 18:18:36 INFO mapred.LocalJobRunner: OutputCommitter set in config null
20/11/26 18:18:36 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce
.lib.output.FileOutputCommitter
20/11/26 18:18:36 INFO mapred.LocalJobRunner: Waiting for map tasks
20/11/26 18:18:36 INFO mapred.LocalJobRunner: Starting task: attempt_local887129590_0001_m_0
00000_0
20/11/26 18:18:36 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
20/11/26 18:18:36 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/rgs1/test.tx
t:0+89
20/11/26 18:18:37 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
20/11/26 18:18:37 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
20/11/26 18:18:37 INFO mapred.MapTask: soft limit at 83886080
20/11/26 18:18:37 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
20/11/26 18:18:37 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
20/11/26 18:18:37 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred
.MapTask$MapOutputBuffer
20/11/26 18:18:37 INFO mapred.LocalJobRunner:
20/11/26 18:18:37 INFO mapred.MapTask: Starting flush of map output
20/11/26 18:18:37 INFO mapred.MapTask: Spilling map output
20/11/26 18:18:37 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
20/11/26 18:18:37 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(10485
7280); length = 77/6553600
20/11/26 18:18:37 INFO mapred.MapTask: Finished spill 0
```

```
20/11/26 18:18:37 INFO mapred.Task: Task:attempt_local887129590_0001_m_000000_0 is done. And
 is in the process of committing
20/11/26 18:18:37 INFO mapred.LocalJobRunner: map
20/11/26 18:18:37 INFO mapred.Task: Task 'attempt_local887129590_0001_m_000000_0' done.
20/11/26 18:18:37 INFO mapred.LocalJobRunner: Finishing task: attempt_local887129590_0001_m_
000000_0
20/11/26 18:18:37 INFO mapred.LocalJobRunner: map task executor complete.
20/11/26 18:18:37 INFO mapred.LocalJobRunner: Waiting for reduce tasks
20/11/26 18:18:37 INFO mapred.LocalJobRunner: Starting task: attempt_local887129590_0001_r_0
00000_0
20/11/26 18:18:37 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
20/11/26 18:18:37 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.map
reduce.task.reduce.Shuffle@7ce5cbfd
20/11/26 18:18:37 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=363285696, maxSin
gleShuffleLimit=90821424, mergeThreshold=239768576, ioSortFactor=10, memToMemMergeOutputsThr
eshold=10
20/11/26 18:18:37 INFO reduce.EventFetcher: attempt_local887129590_0001_r_000000_0 Thread st
arted: EventFetcher for fetching Map Completion Events
20/11/26 18:18:37 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map at
tempt_local887129590_0001_m_000000_0 decomp: 111 len: 115 to MEMORY
20/11/26 18:18:37 INFO reduce.InMemoryMapOutput: Read 111 bytes from map-output for attempt_
local887129590_0001_m_000000_0
20/11/26 18:18:37 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 111
, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->111
20/11/26 18:18:37 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
20/11/26 18:18:37 INFO mapred.LocalJobRunner: 1 / 1 copied.
20/11/26 18:18:37 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outpu
ts and 0 on-disk map-outputs
20/11/26 18:18:37 INFO mapred.Merger: Merging 1 sorted segments
20/11/26 18:18:37 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of t
otal size: 105 bytes
20/11/26 18:18:37 INFO reduce.MergeManagerImpl: Merged 1 segments, 111 bytes to disk to sati
sfy reduce memory limit
20/11/26 18:18:37 INFO reduce.MergeManagerImpl: Merging 1 files, 115 bytes from disk
20/11/26 18:18:37 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into
 reduce
```

```
20/11/26 18:18:37 INFO mapred.Merger: Merging 1 sorted segments
20/11/26 18:18:37 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of t
otal size: 105 bytes
20/11/26 18:18:37 INFO mapred.LocalJobRunner: 1 / 1 copied.
20/11/26 18:18:37 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use
 mapreduce.job.skiprecords
20/11/26 18:18:37 INFO mapreduce.Job: Job job_local887129590_0001 running in uber mode : fal
se
20/11/26 18:18:37 INFO mapreduce.Job:  map 100% reduce 0%
20/11/26 18:18:37 INFO mapred.Task: Task:attempt_local887129590_0001_r_000000_0 is done. And
 is in the process of committing
20/11/26 18:18:37 INFO mapred.LocalJobRunner: 1 / 1 copied.
20/11/26 18:18:37 INFO mapred.Task: Task attempt_local887129590_0001_r_000000_0 is allowed t
o commit now
20/11/26 18:18:37 INFO output.FileOutputCommitter: Saved output of task 'attempt_local887129
590_0001_r_000000_0' to hdfs://localhost:54310/rgs1/output/_temporary/0/task_local887129590_
0001_r_000000
20/11/26 18:18:37 INFO mapred.LocalJobRunner: reduce > reduce
20/11/26 18:18:37 INFO mapred.Task: Task 'attempt_local887129590_0001_r_000000_0' done.
20/11/26 18:18:37 INFO mapred.LocalJobRunner: Finishing task: attempt_local887129590_0001_r_
000000_0
20/11/26 18:18:37 INFO mapred.LocalJobRunner: reduce task executor complete.
20/11/26 18:18:38 INFO mapreduce.Job:  map 100% reduce 100%
20/11/26 18:18:38 INFO mapreduce.Job: Job job_local887129590_0001 completed successfully
20/11/26 18:18:38 INFO mapreduce.Job: Counters: 38
        File System Counters
                FILE: Number of bytes read=7008
                FILE: Number of bytes written=506089
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=178
                HDFS: Number of bytes written=69
                HDFS: Number of read operations=13
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
```

```
        Map-Reduce Framework
                Map input records=5
                Map output records=20
                Map output bytes=169
                Map output materialized bytes=115
                Input split bytes=101
                Combine input records=20
                Combine output records=10
                Reduce input groups=10
                Reduce shuffle bytes=115
                Reduce input records=10
                Reduce output records=10
                Spilled Records=20
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=48
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=242360320
        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=89
        File Output Format Counters
                Bytes Written=69
```

```
hduser@ip-172-31-90-179:~$ hadoop fs -ls /rgs1/output
20/11/26 18:18:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 hduser supergroup          0 2020-11-26 18:18 /rgs1/output/_SUCCESS
-rw-r--r--   1 hduser supergroup         69 2020-11-26 18:18 /rgs1/output/part-r-00000
hduser@ip-172-31-90-179:~$ hadoop fs -cat /rgs1/output/part-r-00000
20/11/26 18:19:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
are     1
brother 1
family  1
hi      1
how     5
is      4
job     1
sister  1
you     1
your    4
```

Java file

```java
import java.util.Iterator;
import org.apache.hadoop.mapreduce.Reducer;
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.fs.Path;
```

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.util.GenericOptionsParser;
import org.apache.hadoop.conf.Configuration;


public class WordCount
{
    public static void main(final String[] array) throws Exception {
        final Configuration configuration = new Configuration();
        final String[] remainingArgs = new GenericOptionsParser(configuration, ar
ray).getRemainingArgs();
        if (remainingArgs.length != 2) {
            System.err.println("Usage: WordCount <in> <out>");
            System.exit(2);
        }
        final Job job = new Job(configuration, "wordcount");
        job.setJarByClass((Class)WordCount.class);
        job.setMapperClass((Class)Map.class);
        job.setReducerClass((Class)Reduce.class);
        job.setCombinerClass((Class)Reduce.class);
        job.setOutputKeyClass((Class)Text.class);
        job.setOutputValueClass((Class)IntWritable.class);
        FileInputFormat.addInputPath(job, new Path(remainingArgs[0]));
        FileOutputFormat.setOutputPath(job, new Path(remainingArgs[1]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }

    public static class Map extends Mapper<LongWritable, Text, Text, IntWritable>
    {
        private static final IntWritable one;
        private Text word;

        public Map() {
            this.word = new Text();
        }

        public void map(final LongWritable longWritable, final Text text, final M
apper.Context context) throws IOException, InterruptedException {
            final StringTokenizer stringTokenizer = new StringTokenizer(text.toSt
ring());
            while (stringTokenizer.hasMoreTokens()) {
                this.word.set(stringTokenizer.nextToken());
                context.write((Object)this.word, (Object)Map.one);
            }
        }

        static {
            one = new IntWritable(1);
        }
    }

    public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritab
le>
    {
        private IntWritable result;
```

```java
        public Reduce() {
            this.result = new IntWritable();
        }

        public void reduce(final Text text, final Iterable<IntWritable> iterable,
final Reducer.Context context) throws IOException, InterruptedException {
            int n = 0;
            final Iterator<IntWritable> iterator = iterable.iterator();
            while (iterator.hasNext()) {
                n += iterator.next().get();
            }
            this.result.set(n);
            context.write((Object)text, (Object)this.result);
        }
    }
}
```