

Roll No: 20BCE204

Course Name and Course Code: 2CS702 Big Data Analytics

Practical No: 5

Aim: Design and implement the MapReduce program to find the sum of given values.

Code:

IntSum.java

```
import java.io.IOException;
import java.util.StringTokenizer;

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.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;

public class IntSum {

    public static class TokenizerMapper
        extends Mapper<Object, Text, Text, IntWritable>{
        private Text word = new Text();

        public void map(Object key, Text value, Context context
            ) throws IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {
                String s = itr.nextToken();
                word.set("SUM");
                IntWritable one = new IntWritable(Integer.parseInt(s));
                context.write(word, one);
            }
        }
    }

    public static class IntSumReducer
```

```

    extends Reducer<Text,IntWritable,Text,IntWritable> {
private IntWritable result = new IntWritable();

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

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "word count");
    job.setJarByClass(IntSum.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(IntSumReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

Steps in Terminal:

hadoop fs -mkdir /Prac5/Input

touch input5.txt

hadoop fs -put /Users/dhyan/Desktop/Practical5/input5.txt /Prac5/Input

code IntSum.java

export HADOOP_CLASSPATH=\$(hadoop classpath)

```
echo $HADOOP_CLASSPATH
```

```
javac -classpath ${HADOOP_CLASSPATH} -d /Users/dhyan/Desktop/Practical5/classes  
/Users/dhyan/Desktop/Practical5/IntSum.java
```

```
jar -cvf IntSum.jar -C /Users/dhyan/Desktop/Practical5/classes .
```

```
hadoop jar /Users/dhyan/Desktop/Practical5/IntSum.jar IntSum /Prac5/Input /Prac5/Output
```

```
hadoop fs -cat /Prac5/Output/part-r-00000
```

Screenshots:

```

Practical5 -- dhyam@Dhyans-MacBook-Pro -- -zsh -- 136x41
(base) dhyam@Dhyans-MacBook-Pro ~ % hadoop fs -mkdir /Prac5/Input
2023-09-22 09:18:27,975 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
(base) dhyam@Dhyans-MacBook-Pro ~ % cd desktop
(base) dhyam@Dhyans-MacBook-Pro desktop % touch input5.txt
(base) dhyam@Dhyans-MacBook-Pro desktop % hadoop fs -put /Users/dhyam/Desktop/Practical5/input5.txt /Prac5/Input
2023-09-22 09:21:41,617 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
(base) dhyam@Dhyans-MacBook-Pro desktop % cd Practical5
(base) dhyam@Dhyans-MacBook-Pro Practical5 % ls
input5.txt
(base) dhyam@Dhyans-MacBook-Pro Practical5 % code IntSum.java
(base) dhyam@Dhyans-MacBook-Pro Practical5 % mkdir classes
(base) dhyam@Dhyans-MacBook-Pro Practical5 % ls
IntSum.java      classes          input5.txt
(base) dhyam@Dhyans-MacBook-Pro Practical5 % export HADOOP_CLASSPATH=$(hadoop classpath)
(base) dhyam@Dhyans-MacBook-Pro Practical5 % export JAVA_HOME=/usr/libexec/java_home -v 1.8.0_333
(base) dhyam@Dhyans-MacBook-Pro Practical5 % javac -classpath $(HADOOP_CLASSPATH) -d /Users/dhyam/Desktop/Practical5/classes /Users/dhyam/Desktop/Practical5/IntSum.java
(base) dhyam@Dhyans-MacBook-Pro Practical5 % jar -cvf IntSum.jar -C /Users/dhyam/Desktop/Practical5/classes .
added manifest
adding: IntSum.class(in = 1479) (out= 812)(deflated 45%)
adding: IntSum$IntSumReducer.class(in = 1730) (out= 732)(deflated 57%)
adding: IntSum$TokenizerMapper.class(in = 1711) (out= 760)(deflated 55%)
(base) dhyam@Dhyans-MacBook-Pro Practical5 % hadoop jar /Users/dhyam/Desktop/Practical5/IntSum.jar IntSum /Prac5/Input /Prac5/Output
2023-09-22 09:26:43,750 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2023-09-22 09:26:44,481 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2023-09-22 09:26:46,025 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-09-22 09:26:46,107 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/dhyam/.staging/job_1695353368388_0001
2023-09-22 09:26:46,977 INFO input.FileInputFormat: Total input files to process : 1
2023-09-22 09:26:48,018 INFO mapreduce.JobSubmitter: number of splits:1
2023-09-22 09:26:48,713 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1695353368388_0001
2023-09-22 09:26:48,714 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-09-22 09:26:49,017 INFO conf.Configuration: resource-types.xml not found
2023-09-22 09:26:49,017 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-09-22 09:26:49,364 INFO impl.YarnClientImpl: Submitted application application_1695353368388_0001
2023-09-22 09:26:49,444 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1695353368388_0001/
2023-09-22 09:26:49,444 INFO mapreduce.Job: Running job: job_1695353368388_0001

```

```

Practical5 -- dhyam@Dhyans-MacBook-Pro -- -zsh -- 136x41
2023-09-22 09:26:49,364 INFO impl.YarnClientImpl: Submitted application application_1695353368388_0001
2023-09-22 09:26:49,444 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1695353368388_0001/
2023-09-22 09:26:49,444 INFO mapreduce.Job: Running job: job_1695353368388_0001
2023-09-22 09:27:00,948 INFO mapreduce.Job: Job job_1695353368388_0001 running in uber mode : false
2023-09-22 09:27:00,955 INFO mapreduce.Job: map 0% reduce 0%
2023-09-22 09:27:06,140 INFO mapreduce.Job: map 100% reduce 0%
2023-09-22 09:27:12,293 INFO mapreduce.Job: map 100% reduce 100%
2023-09-22 09:27:14,411 INFO mapreduce.Job: Job job_1695353368388_0001 completed successfully
2023-09-22 09:27:14,613 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=16
    FILE: Number of bytes written=551401
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=123
    HDFS: Number of bytes written=8
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  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)=3389
    Total time spent by all reduces in occupied slots (ms)=3497
    Total time spent by all map tasks (ms)=3389
    Total time spent by all reduce tasks (ms)=3497
    Total vcore-milliseconds taken by all map tasks=3389
    Total vcore-milliseconds taken by all reduce tasks=3497
    Total megabyte-milliseconds taken by all map tasks=3470336
    Total megabyte-milliseconds taken by all reduce tasks=3580928
  Map-Reduce Framework
    Map input records=5
    Map output records=5
    Map output bytes=40
    Map output materialized bytes=16
    Input split bytes=109
    Combine input records=5
    Combine output records=1

```

```
Practical5 -- dhyan@Dhyan-MacBook-Pro -- -zsh -- 136x41

Total vcore-milliseconds taken by all map tasks=3389
Total vcore-milliseconds taken by all reduce tasks=3497
Total megabyte-milliseconds taken by all map tasks=3470336
Total megabyte-milliseconds taken by all reduce tasks=3580928

Map-Reduce Framework
  Map input records=5
  Map output records=5
  Map output bytes=40
  Map output materialized bytes=16
  Input split bytes=109
  Combine input records=5
  Combine output records=1
  Reduce input groups=1
  Reduce shuffle bytes=16
  Reduce input records=1
  Reduce output records=1
  Spilled Records=2
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=126
  CPU time spent (ms)=0
  Physical memory (bytes) snapshot=0
  Virtual memory (bytes) snapshot=0
  Total committed heap usage (bytes)=494927872

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=14
File Output Format Counters
  Bytes Written=8

(base) dhyan@Dhyan-MacBook-Pro Practical5 % hadoop fs -cat /Prac5/Output/part-r-00000
2023-09-22 09:27:58,582 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
SUM      150
(base) dhyan@Dhyan-MacBook-Pro Practical5 % echo $HADOOP_CLASSPATH
```

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Browse Directory

/Prac5/Input

Show 25 entries

Permission

Owner

-rw-r--r--

dhyan

Showing 1 to 1 of 1 entries

Hadoop, 2023.

File information - input5.txt

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information -- Block 0

Block ID: 1073741906

Block Pool ID: BP-1319230486-127.0.0.1-1692333972593

Generation Stamp: 1082

Size: 14

Availability:

- 10.2.77.153

File contents

10

20

30

40

50

Close

Block Size

Name

8 MBinput5.txt

Previous

1

Next