

Output:

1. User Count program

```
package user_frequency_count;

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 user_frequency_count {

    public static class TokenizerMapper
        extends Mapper<Object, Text, Text, IntWritable>{

        private final static IntWritable one = new IntWritable(1);
        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()) {
                word.set(itr.nextToken());
            }
        }
    }
}
```

```

String validation_string = word.toString();
StringTokenizer validation = new StringTokenizer(validation_string,
".");

int count = 0;
while(validation.hasMoreTokens())
{
    count += 1;
    validation.nextToken();
}
if (count == 4)
{
    context.write(word, one);
}
itr.nextToken();
itr.nextToken();
itr.nextToken();
}
}
}

```

```

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(user_frequency_count.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);
}
}

```

2. Extract maximum frequency user program

```

package maximum_user_frequency;

import java.io.DataInput;
import java.io.DataOutput;
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.io.Writable;
import org.apache.hadoop.io.WritableComparable;
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 MaximumUserFrequency{

    static class User implements Writable, WritableComparable<User> {

        private String name = "";
        private Integer freq = new Integer(0);

        public void setName(String n)
        {
            name = n;
        }

        public void setFreq(Integer f)
        {
            freq = f;
        }

        public String getName()
        {
            return name;
        }

        public Integer getFreq()
```

```

    {
        return freq;
    }

    @Override
    public void readFields(DataInput in) throws IOException {
        freq = in.readInt();
        name = in.readLine();
    }

    @Override
    public void write(DataOutput out) throws IOException {
        out.writeInt(freq);
        out.writeBytes(name);
    }

    @Override
    public int compareTo(User o) {
        // TODO Auto-generated method stub
        int result = this.freq.compareTo(o.freq);
        return result;
    }
}

private static User u = new User();

static class UserMaxCountMapper extends Mapper<Object, Text, Text, User> {
    private Text user = new Text();

    @Override

```

```

        public void map(Object key, Text value, Context context) throws IOException,
InterruptedException {

            user.set("user");

            u.setName(value.toString().split(",")[0]);

            u.setFreq(Integer.parseInt(value.toString().split(",")[1]));

            context.write(user, u);

        }

    }

```

```

static class UserMaxCountReducer extends Reducer<Text, User, Text, IntWritable> {

    private User result = new User();

    public void reduce(Text key, Iterable<User> values, Context context)

        throws IOException, InterruptedException {

            result.setFreq(null);

            result.setName("");

            for (User value : values)
            {

                if (result.getFreq() == null || (value.getFreq() > result.getFreq())) {

                    result.setFreq(value.getFreq());

                    result.setName(value.getName());

                }

            }

            key.set(result.getName());

            context.write(key, new IntWritable(result.getFreq()));

        }

    }

```

```

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

    Configuration conf = new Configuration();

    Job job = Job.getInstance(conf);

    job.setJarByClass(MaximumUserFrequency.class);

    job.setJobName("find_max_user_count");

    FileInputFormat.addInputPath(job, new Path(args[0]));

    FileOutputFormat.setOutputPath(job, new Path(args[1]));

    job.setMapperClass(UserMaxCountMapper.class);

    job.setReducerClass(UserMaxCountReducer.class);

    job.setMapOutputKeyClass(Text.class);

    job.setMapOutputValueClass(User.class);

    job.setOutputKeyClass(Text.class);

    job.setOutputValueClass(IntWritable.class);

    System.exit(job.waitForCompletion(true) ? 0 : 1);

}
}

```

3. User Count program output

```

cloudera-quickstart-vm-542-0-virtualbox (basic_commands) [Running] - Oracle VM VirtualBox
Sun Apr 24, 1:59 AM cloudera
cloudera@quickstart:~/Desktop/Assignment1

File Edit View Search Terminal Help
[cloudera@quickstart Assignment1]$ hadoop jar ~/Desktop/Assignment1/user_frequency_count.jar user_frequency_count user_frequency_count_output
22/04/24 01:55:02 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/04/24 01:55:03 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/04/24 01:55:03 INFO mapreduce.JobSubmitter: Total input paths to process : 1
22/04/24 01:55:03 INFO mapreduce.JobSubmitter: number of splits:1
22/04/24 01:55:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1650790069213_0001
22/04/24 01:55:03 INFO Impl.YarnClientImpl: Submitted application application_1650790069213_0001
22/04/24 01:55:04 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8080/proxy/application_1650790069213_0001/
22/04/24 01:55:04 INFO mapreduce.Job: Running job: job_1650790069213_0001
22/04/24 01:55:14 INFO mapreduce.Job: Job job_1650790069213_0001 running in uber mode : false
22/04/24 01:55:14 INFO mapreduce.Job: map 0% reduce 0%
22/04/24 01:55:23 INFO mapreduce.Job: map 100% reduce 0%
22/04/24 01:55:31 INFO mapreduce.Job: map 100% reduce 100%
22/04/24 01:55:31 INFO mapreduce.Job: Job job_1650790069213_0001 completed successfully
22/04/24 01:55:31 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=91
  FILE: Number of bytes written=221073
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1115485
  HDFS: Number of bytes written=80
  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)=6311
  Total time spent by all reduces in occupied slots (ms)=5375
  Total time spent by all map tasks (ms)=6311
  Total time spent by all reduce tasks (ms)=5375
  Total vcore-seconds taken by all map tasks=6311
  Total vcore-seconds taken by all reduce tasks=5375
  Total megabyte-seconds taken by all map tasks=6462464
  Total megabyte-seconds taken by all reduce tasks=3504000

Map-Reduce Framework
  Map input records=10008
  Map output records=15789
  Map output bytes=236835
  Map output materialized bytes=91
  Input split bytes=142
  Combine input records=15789
  Combine output records=5
  Reduce input groups=5
  Reduce shuffle bytes=91
  Reduce input records=5

```

```
cloudera-quickstart-vm-5.4.2-0-virtualbox (basic_commands) [Running] - Oracle VM VirtualBox
Sun Apr 24, 2:00 AM cloudera
cloudera@quickstart:~/Desktop/Assignment1

Total time spent by all map tasks (ms)=6311
Total time spent by all reduce tasks (ms)=5375
Total vcore-seconds taken by all map tasks=6311
Total vcore-seconds taken by all reduce tasks=5375
Total megabyte-seconds taken by all map tasks=6462464
Total megabyte-seconds taken by all reduce tasks=5504800

Map-Reduce Framework
  Map input records=16008
  Map output records=15789
  Map output bytes=236835
  Map output materialized bytes=91
  Input split bytes=142
  Combine input records=15789
  Combine output records=5
  Reduce input groups=5
  Reduce shuffle bytes=91
  Reduce input records=5
  Reduce output records=5
  Spilled Records=10
  Shuffled Maps => 1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=137
  CPU time spent (ms)=2100
  Physical memory (bytes) snapshot=353996800
  Virtual memory (bytes) snapshot=3807397888
  Total committed heap usage (bytes)=226365440

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=111543
File Output Format Counters
  Bytes Written=80

[cloudera@quickstart Assignment1]$ hadoop fs -ls user_frequency_count_output
Found 2 items
-rw-r--r-- 1 cloudera cloudera 0 2022-04-24 01:55 user_frequency_count_output/SUCCESS
-rw-r--r-- 1 cloudera cloudera 80 2022-04-24 01:55 user_frequency_count_output/part-r-00000
[cloudera@quickstart Assignment1]$ hadoop fs -cat user_frequency_count_output/part-r-00000
10.128.2.1 4257
10.129.2.1 1652
10.130.2.1 4056
10.131.0.1 4198
10.131.2.1 1626
[cloudera@quickstart Assignment1]$
```

4. Maximum User frequency program output

```
cloudera-quickstart-vm-5.4.2-0-virtualbox (basic_commands) [Running] - Oracle VM VirtualBox
Sun Apr 24, 2:06 AM cloudera
cloudera@quickstart:~/Desktop/Assignment1

[cloudera@quickstart Assignment1]$ hadoop jar ~/Desktop/Assignment1/user_maximum_frequency.jar user_frequency_count_output webMax
22/04/24 02:05:47 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/04/24 02:05:48 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/04/24 02:05:48 INFO input.FileInputFormat: Total input paths to process : 1
22/04/24 02:05:49 INFO mapreduce.JobSubmitter: number of splits:1
22/04/24 02:05:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1650790069213_0002
22/04/24 02:05:49 INFO impl.YarnClientImpl: Submitted application application_1650790069213_0002
22/04/24 02:05:50 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8080/proxy/application_1650790069213_0002/
22/04/24 02:05:50 INFO mapreduce.Job: Running job: job_1650790069213_0002
22/04/24 02:06:03 INFO mapreduce.Job: Job job_1650790069213_0002 running in uber mode : false
22/04/24 02:06:03 INFO mapreduce.Job: map 0% reduce 0%
22/04/24 02:06:12 INFO mapreduce.Job: map 100% reduce 0%
22/04/24 02:06:22 INFO mapreduce.Job: map 100% reduce 100%
22/04/24 02:06:23 INFO mapreduce.Job: Job job_1650790069213_0002 completed successfully
22/04/24 02:06:24 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=111
  FILE: Number of bytes written=221451
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=231
  HDFS: Number of bytes written=16
  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)=7479
  Total time spent by all reduces in occupied slots (ms)=8029
  Total time spent by all map tasks (ms)=7479
  Total time spent by all reduce tasks (ms)=8029
  Total vcore-seconds taken by all map tasks=7479
  Total vcore-seconds taken by all reduce tasks=8029
  Total megabyte-seconds taken by all map tasks=7658496
  Total megabyte-seconds taken by all reduce tasks=8221696

Map-Reduce Framework
  Map input records=5
  Map output records=5
  Map output bytes=95
  Map output materialized bytes=111
  Input split bytes=151
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=111
  Reduce input records=5
```



```
cloudera-quickstart-vm-542-0-virtualbox (basic_commands) [Running] - Oracle VM VirtualBox
Sun Apr 24, 2:07 AM cloudera
cloudera@quickstart:~/Desktop/Assignment1

File Edit View Search Terminal Help
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)=7479
  Total time spent by all reduces in occupied slots (ms)=8029
  Total time spent by all map tasks (ms)=7479
  Total time spent by all reduce tasks (ms)=8029
  Total vcore-seconds taken by all map tasks=7479
  Total vcore-seconds taken by all reduce tasks=8029
  Total megabyte-seconds taken by all map tasks=7658496
  Total megabyte-seconds taken by all reduce tasks=8221696
Map-Reduce Framework
  Map input records=5
  Map output records=5
  Map output bytes=95
  Map output materialized bytes=111
  Input split bytes=151
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=111
  Reduce input records=5
  Reduce output records=1
  Spilled Records=10
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=207
  CPU time spent (ms)=1170
  Physical memory (bytes) snapshot=346046464
  Virtual memory (bytes) snapshot=3007389696
  Total committed heap usage (bytes)=226365440
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=0
File Output Format Counters
  Bytes Written=10
[cloudera@quickstart Assignment1]$ hadoop fs -cat webMax/part-r-00000
10.128.2.1 4257
[cloudera@quickstart Assignment1]$
```