

Task1.1) Find the number of unique listeners in the data set.

```
Ans)import java.io.IOException
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class UniqueListner_Mapper extends Mapper<LongWritable, Text, Text, IntWritable>{

    private static IntWritable one = new IntWritable(1);
    @Override
    protected void map(LongWritable key, Text value, Mapper<LongWritable, Text,
    Text,IntWritable>.Context context) throws IOException, InterruptedException {

        String[] lineArray = value.toString().split("\\\\|");
        String listner = lineArray[0];
        context.write(new Text(listner), one);

    }

}

// Reducer Class:

public class UniqueListner_Reducer extends Reducer<Text, IntWritable, Text, IntWritable> {

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

        int sum = 0;

        for (IntWritable val:values) {

            sum= sum + val.get();

        }

        context.write(key, new IntWritable(sum));

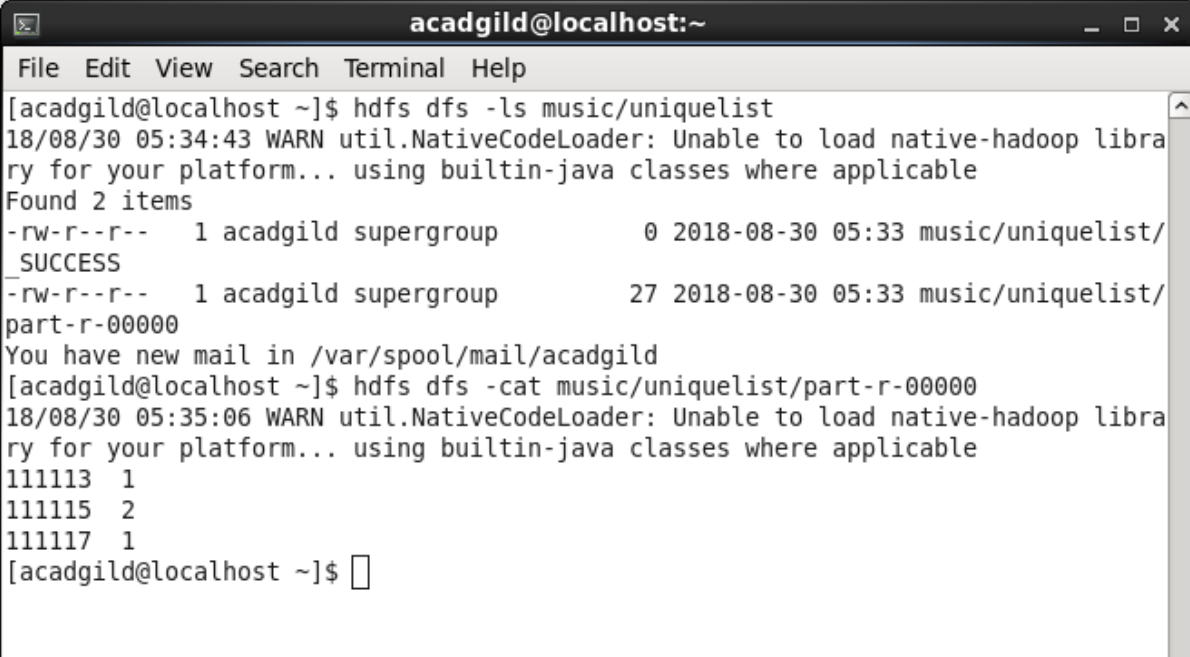
    }

}
```

```

[acadgild@localhost ~]$ hadoop jar Music.jar uniquelister hadoop/music.txt music/unique
18/08/30 05:32:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/30 05:33:00 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/30 05:33:02 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/30 05:33:02 INFO input.FileInputFormat: Total input paths to process : 1
18/08/30 05:33:02 INFO mapreduce.JobSubmitter: number of splits:1
18/08/30 05:33:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_153546909691_0018
18/08/30 05:33:03 INFO impl.YarnClientImpl: Submitted application application_153546909691_0018
18/08/30 05:33:03 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_153546909691_0018/
18/08/30 05:33:03 INFO mapreduce.Job: Running job: job_153546909691_0018
18/08/30 05:33:15 INFO mapreduce.Job: Job job_153546909691_0018 running in uber mode : false
18/08/30 05:33:15 INFO mapreduce.Job:  map 0% reduce 0%
18/08/30 05:33:24 INFO mapreduce.Job:  map 100% reduce 0%
18/08/30 05:33:34 INFO mapreduce.Job:  map 100% reduce 100%
18/08/30 05:33:34 INFO mapreduce.Job: Job job_153546909691_0018 completed successfully

```



```

acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hdfs dfs -ls music/unique
18/08/30 05:34:43 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 acadgild supergroup          0 2018-08-30 05:33 music/unique/_SUCCESS
-rw-r--r--  1 acadgild supergroup       27 2018-08-30 05:33 music/unique/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat music/unique/part-r-00000
18/08/30 05:35:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111113  1
111115  2
111117  1
[acadgild@localhost ~]$

```

Task 1.2) What are the number of times a song was fully heard?

```
Ans) import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
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 SongfullyHeard {
    public static void main(String[] args) throws IOException,
ClassNotFoundException, InterruptedException {

        Configuration conf= new Configuration();
        Job job = new Job(conf,"SongFullyHeard") ;
        job.setJarByClass (SongfullyHeard.class);
        job.setMapperClass (Map.class);
        job.setReducerClass (Reduce.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);

    }

    public static class Map extends Mapper<LongWritable, Text,
Text,IntWritable>{
protected void map(LongWritable key,  Text value, Context context)
throws IOException ,InterruptedException {
    String [] lineArray = value.toString().split("[]");
    String trackID = lineArray[1];
    int SongHeard = Integer.parseInt(lineArray[4]);
    if(SongHeard == 1) {
        context.write(new Text(trackID), new IntWritable(1));
    }
}

    }

    public static class Reduce extends
Reducer<Text,IntWritable,Text,IntWritable>{

protected void reduce(Text key, Iterable<IntWritable> value, Context
context) throws IOException, InterruptedException{
    int sum=0;
    for(IntWritable val : value) {
        sum += val.get();
    }
    context.write(key,new IntWritable(sum));
} } }
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Music.jar SongfullyHeard /user/  
/acadgild/hadoop/music.txt /user/acadgild/music/songheard  
18/08/29 19:30:35 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
18/08/29 19:30:39 INFO client.RMProxy: Connecting to ResourceManager at localhos  
t/127.0.0.1:8032  
18/08/29 19:30:40 WARN mapreduce.JobResourceUploader: Hadoop command-line option  
parsing not performed. Implement the Tool interface and execute your applicatio  
n with ToolRunner to remedy this.  
18/08/29 19:30:41 INFO input.FileInputFormat: Total input paths to process : 1  
18/08/29 19:30:41 INFO mapreduce.JobSubmitter: number of splits:1  
18/08/29 19:30:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15  
35546909691_0003  
18/08/29 19:30:42 INFO impl.YarnClientImpl: Submitted application application_15  
35546909691_0003  
18/08/29 19:30:42 INFO mapreduce.Job: The url to track the job: http://localhost  
:8088/proxy/application_1535546909691_0003/  
18/08/29 19:30:42 INFO mapreduce.Job: Running job: job_1535546909691_0003  
18/08/29 19:30:57 INFO mapreduce.Job: Job job_1535546909691_0003 running in uber  
mode : false  
18/08/29 19:30:57 INFO mapreduce.Job: map 0% reduce 0%  
18/08/29 19:31:10 INFO mapreduce.Job: map 100% reduce 0%
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hdfs dfs -ls music/songheard  
18/08/29 19:32:18 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
Found 2 items  
-rw-r--r-- 1 acadgild supergroup 0 2018-08-29 19:31 music/songheard/_  
SUCCESS  
-rw-r--r-- 1 acadgild supergroup 6 2018-08-29 19:31 music/songheard/p  
art-r-00000  
[acadgild@localhost ~]$ hdfs dfs -cat music/songheard/part-r-00000  
18/08/29 19:32:43 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
223 1  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$
```

Task3) What are the number of times a song was shared.

Ans)

```
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.LongWritable;
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 songShared {

public static void main(String[] args) throws IOException,
ClassNotFoundException, InterruptedException {

        Configuration conf= new Configuration();
        Job job = new Job(conf,"Songshared") ;
        job.setJarByClass (songShared.class);
        job.setMapperClass (Map.class);
        job.setReducerClass (Reduce.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);

    }

public static class Map extends Mapper<LongWritable, Text,
Text,IntWritable>{

    protected void map(LongWritable key,  Text value, Context context)
    throws IOException ,InterruptedException {

        String [] lineArray = value.toString().split("[|]");
        String trackID = lineArray[1];
        int SongShared = Integer.parseInt(lineArray[2]);
        if(SongShared == 1) {
            context.write(new Text(trackID),new IntWritable(1));
        }
    }
}

public static class Reduce extends
Reducer<Text,IntWritable,Text,IntWritable>{
    protected void reduce(Text key, Iterable<IntWritable> value, Context
context)throws IOException, InterruptedException{
        int sum=0;
        for(IntWritable val : value) {
            sum += val.get();
        }
        context.write(key,new IntWritable(sum));
    }
}
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hadoop jar /home/acadgild/Music.jar songShared /user/acadgild/hadoop/music.txt /user/acadgild/music/songshared  
18/08/29 19:36:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
18/08/29 19:36:52 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032  
18/08/29 19:36:53 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
18/08/29 19:36:54 INFO input.FileInputFormat: Total input paths to process : 1  
18/08/29 19:36:54 INFO mapreduce.JobSubmitter: number of splits:1  
18/08/29 19:36:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1535546909691_0004  
18/08/29 19:36:55 INFO impl.YarnClientImpl: Submitted application application_1535546909691_0004  
18/08/29 19:36:55 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1535546909691_0004/  
18/08/29 19:36:55 INFO mapreduce.Job: Running job: job_1535546909691_0004  
18/08/29 19:37:05 INFO mapreduce.Job: Job job_1535546909691_0004 running in uber mode : false  
18/08/29 19:37:05 INFO mapreduce.Job: map 0% reduce 0%  
18/08/29 19:37:13 INFO mapreduce.Job: map 100% reduce 0%  
18/08/29 19:37:22 INFO mapreduce.Job: map 100% reduce 100%  
18/08/29 19:37:22 INFO mapreduce.Job: Job job_1535546909691_0004 completed successfully
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hdfs dfs -ls music/songshared  
18/08/29 19:39:25 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 acadgild supergroup 0 2018-08-29 19:37 music/songshared/_SUCCESS  
-rw-r--r-- 1 acadgild supergroup 6 2018-08-29 19:37 music/songshared/part-r-000000  
[acadgild@localhost ~]$ hdfs dfs -cat music/songshared/part-r-000000  
18/08/29 19:39:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
225 2  
[acadgild@localhost ~]$
```

Task2.1) Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table.

Ans) The commands used are as below.

- i. Initially created a table in mysql with the same schema as that of hdfs.

```
mysql> create table employee(emp_id INT NOT NULL AUTO_INCREMENT, emp_name VARCHAR(100), emp_sal FLOAT, PRIMARY KEY(emp_id));
Query OK, 0 rows affected (0.65 sec)
```

- ii. Used below sqoop export command to export the data which is initially present in sqoopout folder.

```
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --password Root@123 --table employee --export-dir /sqoopout --input-fields-terminated-by ',' --input-lines-terminated-by '\n' -m1 --batch
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
18/08/30 09:03:46 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/08/30 09:03:46 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/08/30 09:03:46 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/08/30 09:03:46 INFO tool.CodeGenTool: Beginning code generation
Thu Aug 30 09:03:47 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/08/30 09:03:48 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
```

- iii. Checked the output in employee table in database db1.

```
mysql> show tables;
+-----+
| Tables_in_db1 |
+-----+
| emp            |
| employee       |
+-----+
2 rows in set (0.00 sec)

mysql> select * from emp;
+-----+-----+-----+
| emp_id | emp_name | emp_sal |
+-----+-----+-----+
| 1      | ashu     | 434000  |
| 2      | ashwini  | 650000  |
| 3      | atmakuri | 100000  |
| 4      | xzy      | 7        |
| 6      | xzy      | 7        |
| 7      | xyz      | 8        |
| 10     | xyz      | 8        |
+-----+-----+-----+
7 rows in set (0.03 sec)
```


Task2.2) Use Sqoop tool to import data present in SQOOPOUT folder made while demo of Import table with parameter person_id =3.

Ans) The command used to import data with person id=3.

```
sqoop import --connect jdbc:mysql://localhost/simplidb --table
Person --username root --password Root@123 --target-dir /sqoopout
-m 1 --where person_id=3
```

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplidb -
-table Person --username root --password Root@123 --target-dir /Sqoopout -m1 --w
here person_id=3
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/./hca
talog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/./acc
umulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
18/08/30 07:38:42 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/08/30 07:38:42 WARN tool.BaseSqoopTool: Setting your password on the command-
line is insecure. Consider using -P instead.
18/08/30 07:38:42 INFO manager.MySQLManager: Preparing to use a MySQL streaming
resultset.
18/08/30 07:38:42 INFO tool.CodeGenTool: Beginning code generation
Thu Aug 30 07:38:44 IST 2018 WARN: Establishing SSL connection without server's
identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ an
d 5.7.6+ requirements SSL connection must be established by default if explicit
option isn't set. For compliance with existing applications not using SSL the ve
rifyServerCertificate property is set to 'false'. You need either to explicitly
disable SSL by setting useSSL=false, or set useSSL=true and provide truststore f
or server certificate verification.
18/08/30 07:38:47 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `Person` AS t LIMIT 1
```

```
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -ls /Sqoopout
18/08/30 07:43:28 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-08-30 07:40 /Sqoopout/_SUCCESS
S
-rw-r--r--  1 acadgild supergroup        37 2018-08-30 07:40 /Sqoopout/part-m-
000000
[acadgild@localhost ~]$ hdfs dfs -ls /Sqoopout/part-m-000000
18/08/30 07:44:24 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
-rw-r--r--  1 acadgild supergroup        37 2018-08-30 07:40 /Sqoopout/part-m-
000000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /Sqoopout/part-m-000000
18/08/30 07:44:45 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
New York,United States,Brown,James,3
[acadgild@localhost ~]$ █
```