

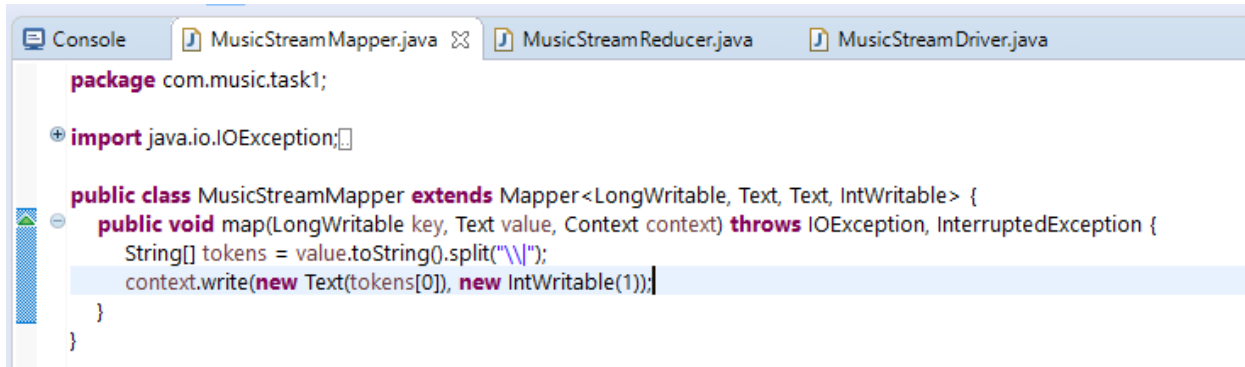
Session 5:
ADVANCE MAP REDUCE AND
INTRODUCTION TO UNIX
CONCEPTS
Assignment 1

Task 1:

Find the number of unique listeners in the data set.

Programs: (Package: com.music.task1)

MusicStreamMapper.java

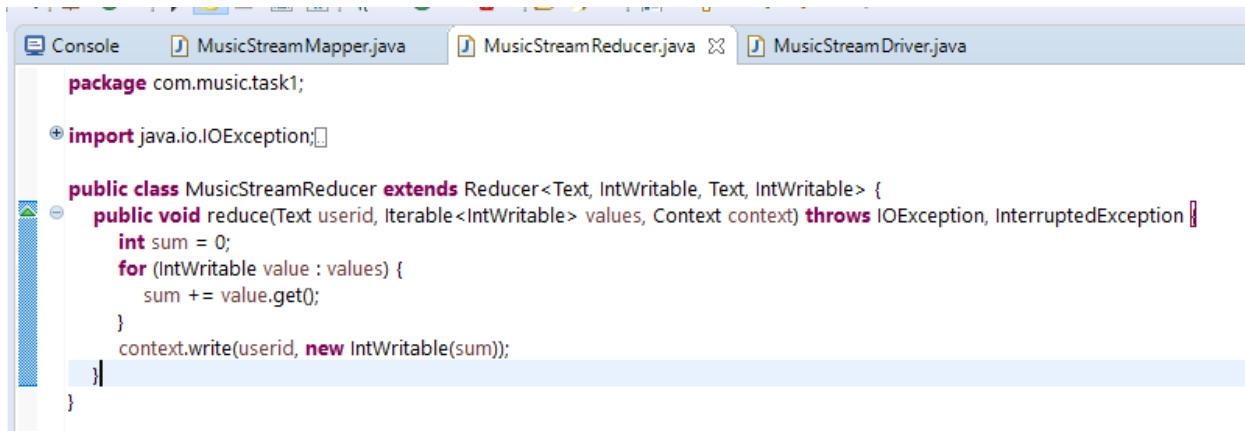


```
package com.music.task1;

import java.io.IOException;

public class MusicStreamMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
    public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
        String[] tokens = value.toString().split("\\\\");
        context.write(new Text(tokens[0]), new IntWritable(1));
    }
}
```

MusicStreamReducer.java



```
package com.music.task1;

import java.io.IOException;

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

MusicStreamDriver.java

```
package com.music.task1;

import org.apache.hadoop.conf.Configuration;

public class MusicStreamDriver {
    @SuppressWarnings("deprecation")
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = new Job(conf, "Music Stream Analyzer");

        // Configure Mapper, Reducer and Driver classes
        job.setJarByClass(MusicStreamDriver.class);
        job.setMapperClass(MusicStreamMapper.class);
        job.setReducerClass(MusicStreamReducer.class);

        // Setup mapper output key and value types
        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(IntWritable.class);

        // Setup input file format
        job.setInputFormatClass(TextInputFormat.class);

        // Configure input and output file paths
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

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

Execution:

```
hadoop jar /home/acadgild/workspace/Mapreduce/MusicStreamAnalysis_Task1.jar
/mapreduce/musicdata.txt /mapreduce/music-task1-out
```

Output:

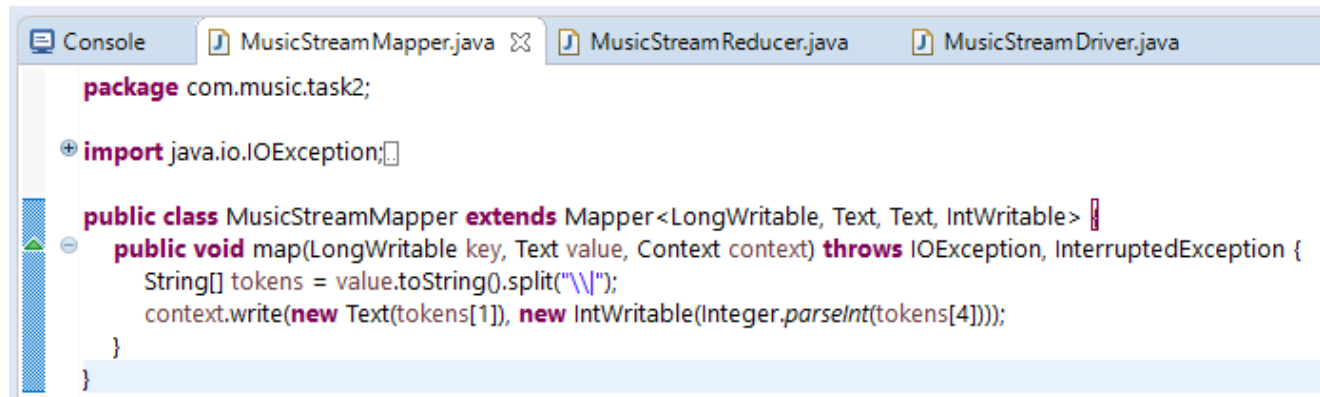
```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/bin
File Edit View Search Terminal Help
[acadgild@localhost bin]$ hadoop fs -copyFromLocal /home/acadgild/workspace/Mapreduce/musicdata.txt /mapreduce
19/02/19 22:51:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost bin]$ hadoop fs -ls /mapreduce
19/02/19 22:51:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 13 items
drwxr-xr-x - acadgild supergroup          0 2019-02-19 00:09 /mapreduce/companysales-out
drwxr-xr-x - acadgild supergroup          0 2019-02-19 21:33 /mapreduce/countrysales-out
drwxr-xr-x - acadgild supergroup          0 2019-01-13 07:02 /mapreduce/input
-rw-r--r-- 1 acadgild supergroup        201 2019-01-13 06:54 /mapreduce/input.txt
drwxr-xr-x - acadgild supergroup          0 2019-02-19 22:29 /mapreduce/movies
-rw-r--r-- 1 acadgild supergroup          72 2019-02-19 22:51 /mapreduce/musicdata.txt
drwxr-xr-x - acadgild supergroup          0 2019-01-13 07:05 /mapreduce/output
drwxr-xr-x - acadgild supergroup          0 2019-02-18 23:00 /mapreduce/productsales-out
drwxr-xr-x - acadgild supergroup          0 2019-02-18 22:29 /mapreduce/sales-out
drwxr-xr-x - acadgild supergroup          0 2019-02-18 23:26 /mapreduce/sales-output
drwxr-xr-x - acadgild supergroup          0 2019-01-13 07:30 /mapreduce/standarddeviation
-rw-r--r-- 1 acadgild supergroup        733 2019-02-17 17:52 /mapreduce/television.txt
drwxr-xr-x - acadgild supergroup          0 2019-01-13 07:25 /mapreduce/wordmean
[acadgild@localhost bin]$ hadoop fs -cat /mapreduce/music-task1-out/part-r-00000
19/02/19 22:55:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111113 1
111115 2
111117 1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$
```

Task 2:

What are the number of times a song was heard fully.

Program: (Package: com.music.task2)

MusicStreamMapper.java

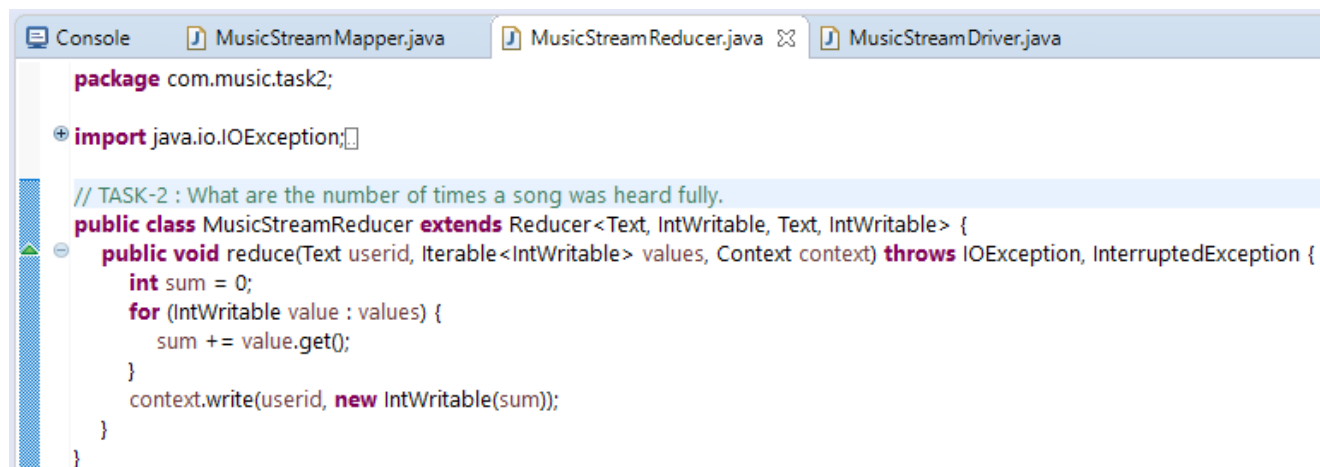


```
package com.music.task2;

import java.io.IOException;

public class MusicStreamMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
    public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
        String[] tokens = value.toString().split("\\\\");
        context.write(new Text(tokens[1]), new IntWritable(Integer.parseInt(tokens[4])));
    }
}
```

MusicStreamReducer.java



```
package com.music.task2;

import java.io.IOException;

// TASK-2 : What are the number of times a song was heard fully.
public class MusicStreamReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
    public void reduce(Text userid, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable value : values) {
            sum += value.get();
        }
        context.write(userid, new IntWritable(sum));
    }
}
```

MusicStreamDriver.java

```

package com.music.task2;

import org.apache.hadoop.conf.Configuration;

public class MusicStreamDriver {
    @SuppressWarnings("deprecation")
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = new Job(conf, "Music Stream Analyzer");

        // Configure Mapper, Reducer and Driver classes
        job.setJarByClass(MusicStreamDriver.class);
        job.setMapperClass(MusicStreamMapper.class);
        job.setReducerClass(MusicStreamReducer.class);

        // Setup mapper output key and value types
        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(IntWritable.class);

        // Setup input file format
        job.setInputFormatClass(TextInputFormat.class);

        // Configure input and output file paths
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

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

```

Execution:

```

hadoop jar /home/acadgild/workspace/Mapreduce/MusicStreamAnalysis_Task2.jar
/mapreduce/musicdata.txt /mapreduce/music-task2-out

```

Output:

```

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$ hadoop fs -cat /mapreduce/music-task2-out/part-r-00000
19/02/19 23:15:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
222 0
223 1
225 0
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$

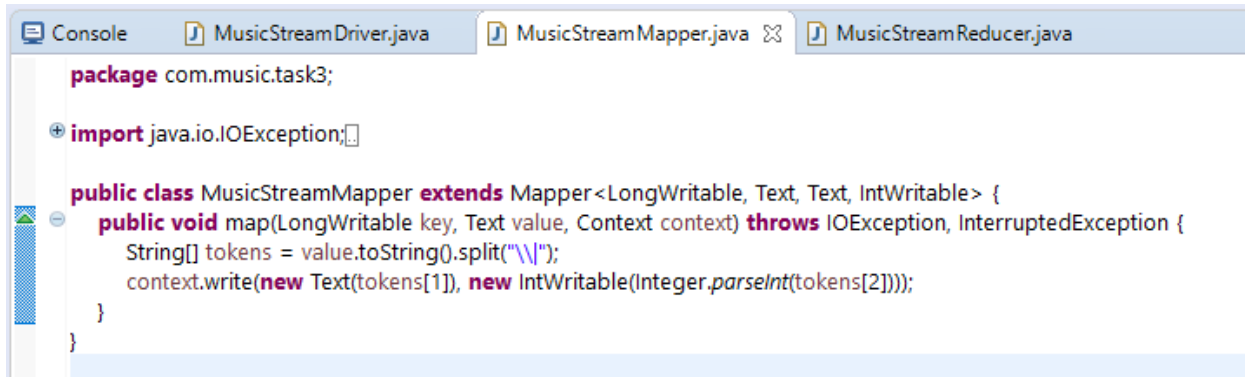
```

Task 3:

What are the number of times a song was shared.

Program: (Package – com.music.task3)

MusicStreamMapper.java

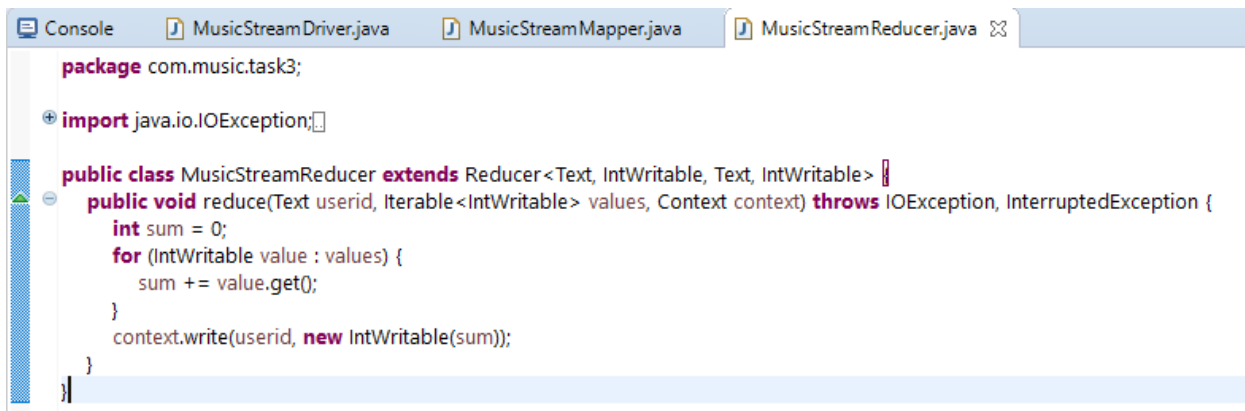


```
package com.music.task3;

import java.io.IOException;

public class MusicStreamMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
    public void map(LongWritable key, Text value, Context context) throws IOException, InterruptedException {
        String[] tokens = value.toString().split("\\\\");
        context.write(new Text(tokens[1]), new IntWritable(Integer.parseInt(tokens[2])));
    }
}
```

MusicStreamReducer.java



```
package com.music.task3;

import java.io.IOException;

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

MusicStreamDriver.java

```
package com.music.task3;

import org.apache.hadoop.conf.Configuration;

//TASK-3 : What are the number of times a song was heard fully.
public class MusicStreamDriver {
    @SuppressWarnings("deprecation")
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = new Job(conf, "Music Stream Analysis - Task 3");

        // Configure Mapper, Reducer and Driver classes
        job.setJarByClass(MusicStreamDriver.class);
        job.setMapperClass(MusicStreamMapper.class);
        job.setReducerClass(MusicStreamReducer.class);

        // Setup mapper output key and value types
        job.setMapOutputKeyClass(Text.class);
        job.setMapOutputValueClass(IntWritable.class);

        // Setup input file format
        job.setInputFormatClass(TextInputFormat.class);

        // Configure input and output file paths
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

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

Execution:

```
hadoop jar /home/acadgild/workspace/Mapreduce/MusicStreamAnalysis_Task3.jar
/mapreduce/musicdata.txt /mapreduce/music-task3-out
```

Output:

```
drwxr-xr-x  - acadgild supergroup          0 2019-01-13 07:25 /mapreduce/wordmean
[acadgild@localhost bin]$ hadoop fs -cat /mapreduce/music-task1-out/part-r-00000
19/02/19 22:55:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
111113 1
111115 2
111117 1
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$ hadoop fs -cat /mapreduce/music-task2-out/part-r-00000
19/02/19 23:15:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
222 0
223 1
225 0
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$ hadoop fs -cat /mapreduce/music-task3-out/part-r-00000
19/02/19 23:25:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
222 0
223 0
225 2
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$
hadoop jar /home/acadgild/workspace/Mapreduce/MusicStreamAnalysis_Task3.jar /mapreduce/musicdata.txt /mapreduce/
```