Chethan U Mahindrakar – SEC01 (NUID 002646783)

Big Data System Engineering with Scala
Spring 2023
Assignment No. Spark 2(Working with API'S)

Output:

The top 10 longest songs in the playlist

```
The top 10 longest songs in the playlist are :
The End - New Stereo Mix Advanced Resolution ----- 699533 ms
Desolation Row ----- 681400 ms
Free Bird ----- 550066 ms
Purple Rain ----- 520786 ms
Kashmir - 1990 Remaster ----- 508200 ms
Iron Man - Live ----- 500693 ms
Good Times ----- 495400 ms
Stairway to Heaven - 1990 Remaster ----- 478173 ms
L.A. Woman ----- 471160 ms
A Whiter Shade Of Pale - Live ----- 466506 ms

Command took 0.49 seconds -- by mahindrakar.c@northeastern.edu at 3/14/2023, 2:54:45 AM on Spark 2
```

The artists from the above songs sorted by the most number of followers

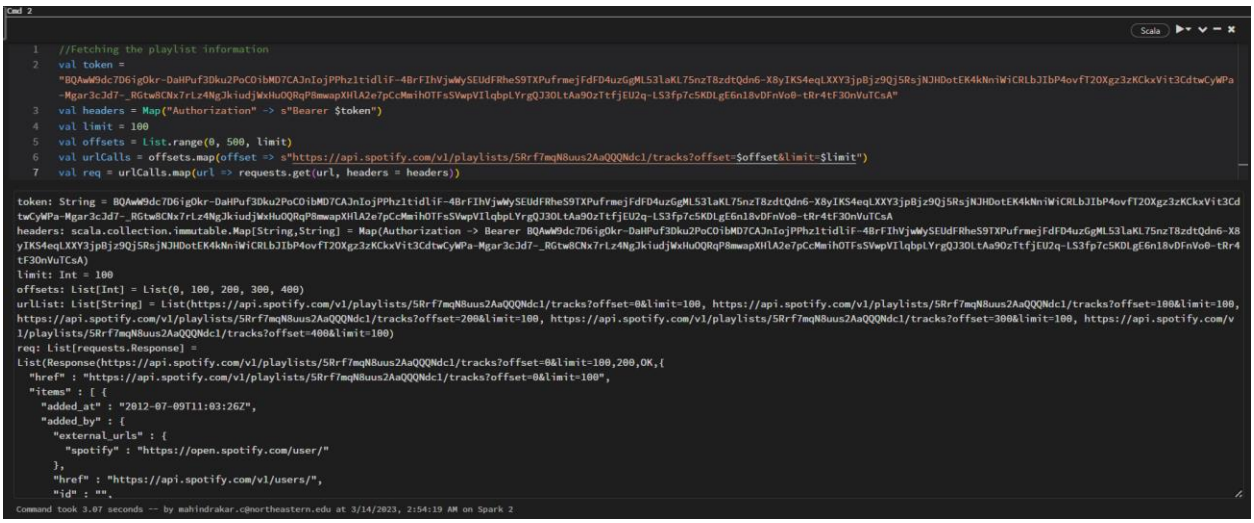
```
Artist details of artists in top 10 longest songs:
Led Zeppelin ----- 13151926 followers
The Doors ----- 7106674 followers
Black Sabbath ----- 6808824 followers
Prince ----- 6646834 followers
Bob Dylan ----- 6020952 followers
Lynyrd Skynyrd ----- 4519016 followers
CHIC ----- 646066 followers
Procol Harum ----- 400525 followers
The Danish National Concert Orchestra and Choir ----- 75 followers

Command took 0.62 seconds -- by mahindrakar.c@northeastern.edu at 3/14/2023, 2:54:49 AM on Spark 2
```

Code:

```
Cmd 1

1  import org.apache.spark.sql.SparkSession
2  import org.apache.spark.sql.functions._
3  import requests._
4  import scala.concurrent.ExecutionContext.Implicits.global
5  import scala.concurrent.Future
6  import ujson._
7
8  val spark = SparkSession
9    .builder()
10   .appName("Spark Assignment 2")
11   .getOrCreate()

import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions._
import requests._
import scala.concurrent.ExecutionContext.Implicits.global
import scala.concurrent.Future
import ujson._
spark: org.apache.spark.sql.SparkSession = org.apache.spark.sql.SparkSession@3f5f5433

Command took 0.43 seconds -- by mahindrakar.c@northeastern.edu at 3/14/2023, 2:51:58 AM on Spark 2
```

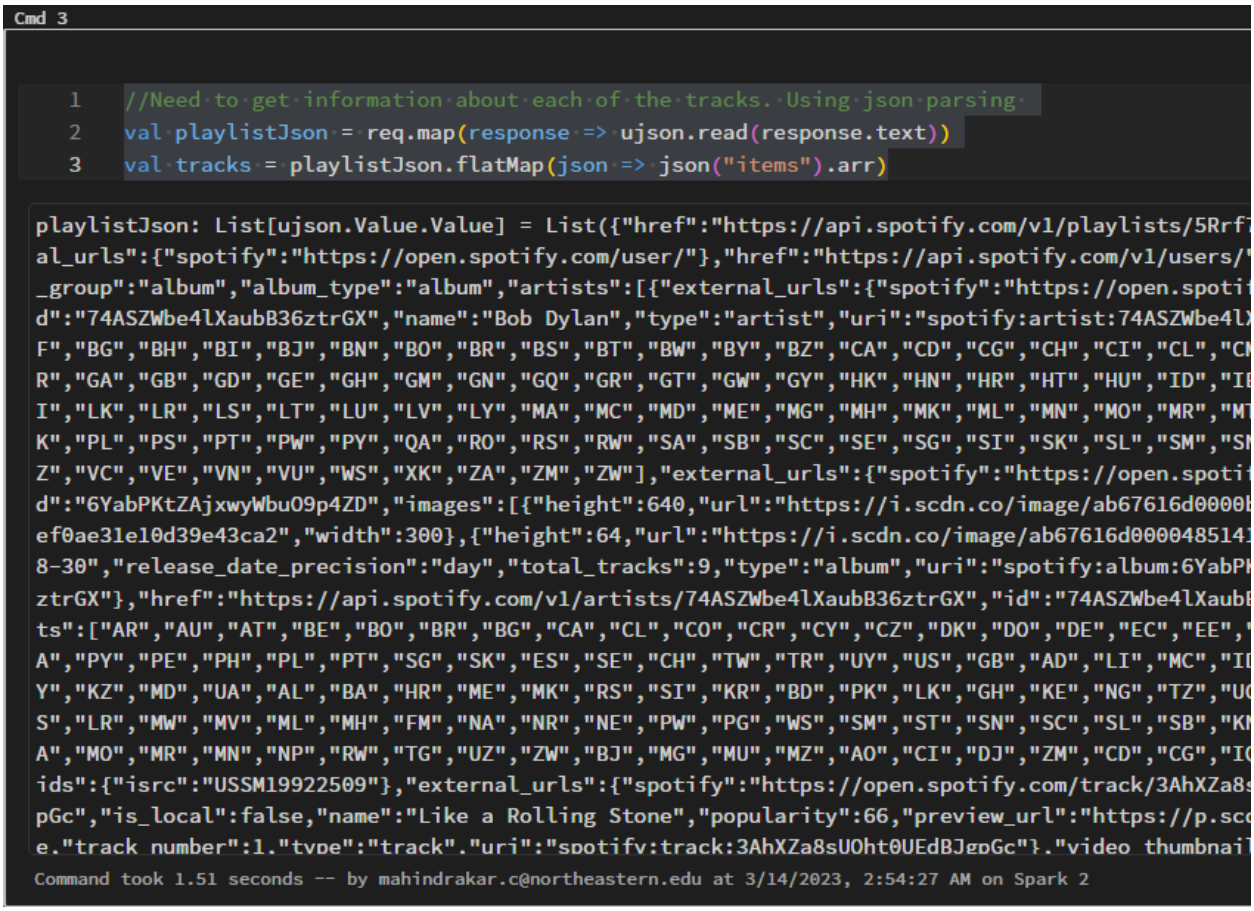
Importing the required libraries and building the spark session(Although it is not required in this particular assignment)

The scala-requests and scala-ujson libraries from https://github.com/com-lihaoyi were needed for this assignment and were imported into the cluster configuration through maven.

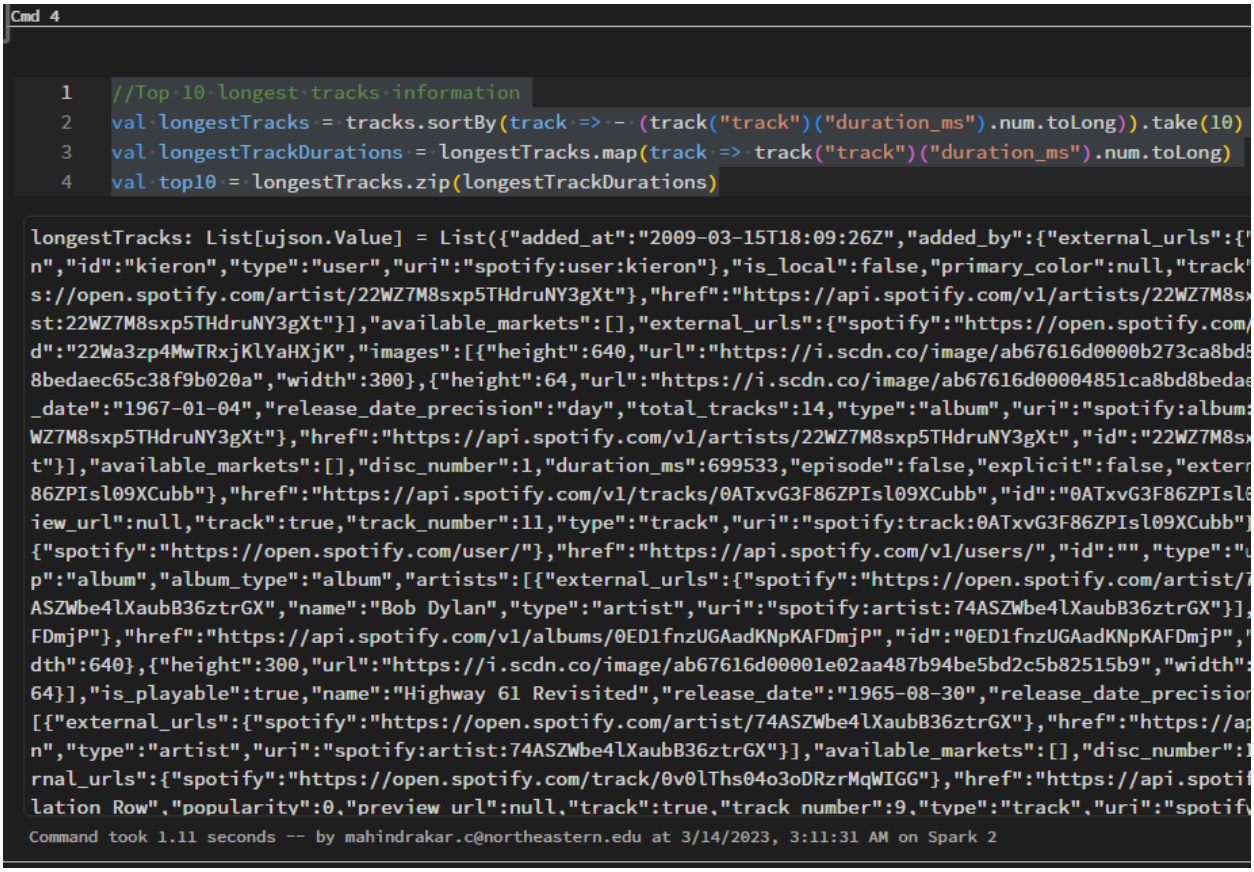


This code I basically used for generating the API request. For this, we need to use a token that is provided by Spotify for developers. The token needs to be changed on the next use as it is most likely to have expired. The token is passed in the get request as a

part of the header. Since there is a limit on the number of songs from a playlist that can be obtained in one go, we need to call the api multiple times. This is handled by the limit and offsets. We make 5 calls for 500 songs extracting 100 songs with each request. All of the tracks are stored in a list which will be decomposed in the next step.



This code uses the ujson package to parse the responses from the previous API queries (stored in the req variable) into JSON. Then, a new list of track information is created after information about each track has been extracted from the JSON data. The "items" array, which holds details about each track, is then extracted from each JSON object in the playlistJson list using the flatMap function, and the list of arrays that results is flattened into a single list of track details.



The tracks list is sorted by the "duration ms" parameter in descending order using the sortBy method. The result is then placed in a new variable named longestTracks after being reduced to the first 10 items with the take function. The "duration ms" attribute is then extracted from each track object in the longestTracks list using the map function and converted to a Long integer type. Finally, a new list of tuples called top10 is produced by using the zip function to combine each track object in the longestTracks list with its matching duration value in the longestTrackDurations list. The top 10 longest tracks in the playlist are represented by the tuples in top10, each of which comprises a track object and its matching duration.

Cmd 5

```
1  //Displaying the top 10 longest songs in the playlist
2  println("The top 10 longest songs in the playlist are :")
3  top10.foreach { case (track, durationMs) =>
4    println(track("track")("name").str + s" ----- $durationMs ms")
5  }
```

```
The top 10 longest songs in the playlist are :
The End - New Stereo Mix Advanced Resolution ----- 699533 ms
Desolation Row ----- 681400 ms
Free Bird ----- 550066 ms
Purple Rain ----- 520786 ms
Kashmir - 1990 Remaster ----- 508200 ms
Iron Man - Live ----- 500693 ms
Good Times ----- 495400 ms
Stairway to Heaven - 1990 Remaster ----- 478173 ms
L.A. Woman ----- 471160 ms
A Whiter Shade Of Pale - Live ----- 466506 ms
```

Command took 0.52 seconds -- by mahindrakar.c@northeastern.edu at 3/14/2023, 3:11:35 AM on Spark 2

This code is used only used for displaying the top 10 songs in terms of duration.

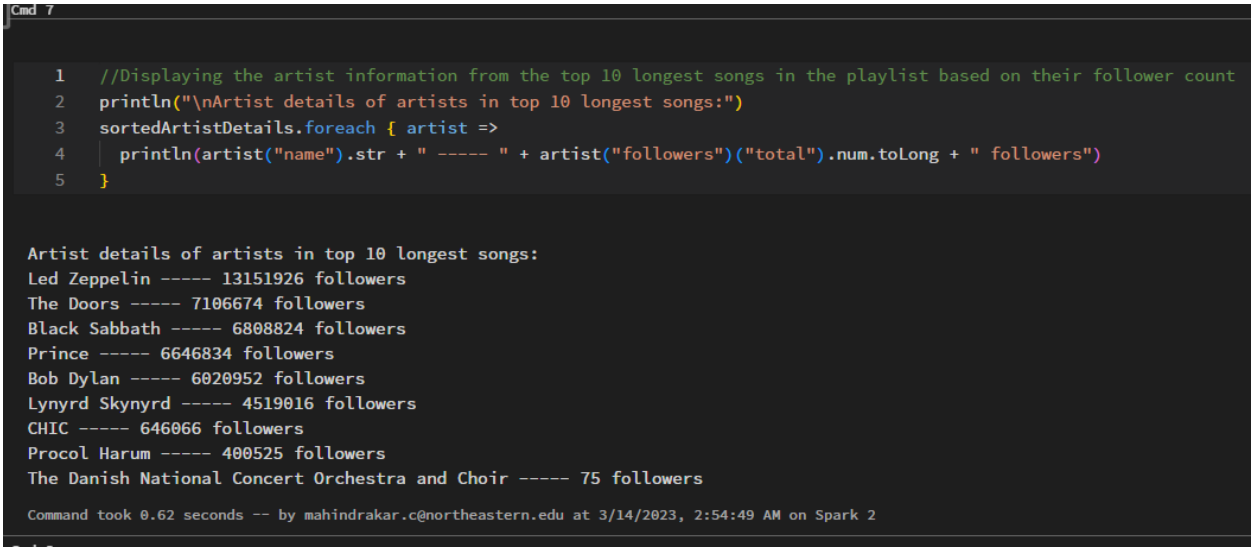
Cmd 6

```
1  //Getting information about the artists using another the API call
2  val artistIds = longestTracks.flatMap(track => track("track")("artists").arr.map(artist => artist("id").str)).distinct
3  val artistDetails = artistIds.map { artistId =>
4    val artistRequest = requests.get(s"https://api.spotify.com/v1/artists/$artistId", headers = headers)
5    ujson.read(artistRequest.text)
6  }
7  val sortedArtistDetails = artistDetails.sortBy(artist => -(artist("followers")("total").num.toLong))
```

```
artistIds: List[String] = List(22WZ7M8sxp5THdruNY3gXt, 74ASZWbe4lXaubB36ztrGX, 4MVyzYMgTwdP7Z49wAZHx0, 5a2EaR3hamoenG9rDuVn8j,
5TJr7n4is453VOY4C, 16VcXOpFIPL82u5CfIIz9R)
artistDetails: List[ujson.Value.Value] = List({"external_urls":{"spotify":"https://open.spotify.com/artist/22WZ7M8sxp5THdruNY3
ock","classic rock","psychedelic rock","rock"],"href":"https://api.spotify.com/v1/artists/22WZ7M8sxp5THdruNY3gXt","id":"22WZ7M
1610000e5eb440959e022afc20e819050bd","width":640},{"height":320,"url":"https://i.scdn.co/image/ab67616100005174440959e022afc20
000f178440959e022afc20e819050bd","width":160}],"name":"The Doors","popularity":76,"type":"artist","uri":"spotify:artist:22WZ7M
st/74ASZWbe4lXaubB36ztrGX"},"followers":{"href":null,"total":6020952},"genres":["classic rock","country rock","folk","folk roc
spotify.com/v1/artists/74ASZWbe4lXaubB36ztrGX","id":"74ASZWbe4lXaubB36ztrGX","images":[{"height":1000,"url":"https://i.scdn.co
0,"url":"https://i.scdn.co/image/ab6772690000dd22f79ca05dcc1581c712bf3658","width":640},{"height":200,"url":"https://i.scdn.co
l":"https://i.scdn.co/image/ab67726900008f74f79ca05dcc1581c712bf3658","width":64}],"name":"Bob Dylan","popularity":74,"type":"
tify":"https://open.spotify.com/artist/4MVyzYMgTwdP7Z49wAZHx0"},"followers":{"href":null,"total":4519016},"genres":["album roc
k","southern rock"],"href":"https://api.spotify.com/v1/artists/4MVyzYMgTwdP7Z49wAZHx0","id":"4MVyzYMgTwdP7Z49wAZHx0","images":
e474c2bf","width":640},{"height":320,"url":"https://i.scdn.co/image/ab67616100005174451eb07ddc6e68d3e474c2bf","width":320},{"h
c2bf","width":160}],"name":"Lynyrd Skynyrd","popularity":75,"type":"artist","uri":"spotify:artist:4MVyzYMgTwdP7Z49wAZHx0"}, {"
n8j"},"followers":{"href":null,"total":6646834},"genres":["funk","funk rock","minneapolis sound","synth funk"],"href":"https:/
8j","images":[{"height":640,"url":"https://i.scdn.co/image/ab6761610000e5ebeaca358712b3fe4ed9814640","width":640},{"height":32
dth":320},{"height":160,"url":"https://i.scdn.co/image/ab6761610000f178eaca358712b3fe4ed9814640","width":160}],"name":"Prince"
j"}, {"external_urls":{"spotify":"https://open.spotify.com/artist/36QJpDe2go2KgaRleHCDTp"},"followers":{"href":null,"total":13
s://api.spotify.com/v1/artists/36QJpDe2go2KgaRleHCDTp","id":"36QJpDe2go2KgaRleHCDTp","images":[{"height":600,"url":"https://i.
t":200,"url":"https://i.scdn.co/image/b0248a44865493e6a03832aa89854ada16ff07a8","width":200},{"height":64,"url":"https://i.scd
Zeppelin","popularity":79,"type":"artist","uri":"spotify:artist:36QJpDe2go2KgaRleHCDTp"}, {"external_urls":{"spotify":"https:/
```

Command took 1.62 seconds -- by mahindrakar.c@northeastern.edu at 3/14/2023, 2:54:47 AM on Spark 2

The artist IDs for each track in the longestTracks list are retrieved using the flatMap method. It takes the "artists" array from each track and maps it to a new array of artist IDs. The list of artist IDs that comes is flattened using flatMap and cleaned up with distinct before being saved in a variable called artistIds. The artist information is then requested from the Spotify API by iterating over each artist ID in the artistIds array using the map function. The artistDetails list is sorted by the total number of followers of each artist in descending order using the sortBy method.



This code displays the artist information from the top 10 songs in descending order of the number of followers.