

Assignment 10 Solution

STUDENT NAME: Chin-Ting Ko

1) Scala Program source code paste here.

```
/**
 * Created by Ko, Chin-Ting on 12/3/16.
 * This program is to calculate movie average ratings.
 */

import org.apache.spark.SparkContext
import org.apache.spark.SparkContext._
import org.apache.spark.SparkConf

object movie_rating{
  def main(args: Array[String]) {

    val conf = new SparkConf().setAppName("Movie Rating")
    val sc = new SparkContext(conf)
    sc.setLogLevel("ERROR");

    val movieRatingRDD = sc.textFile("hdfs://localhost:9000/u.data")
    val movieRatingTupleRDD = movieRatingRDD.map (line => (line.split("\t")(1), line.split("\t")
(2).toFloat ))
    val groupedMovieRatingTupleRDD =
movieRatingTupleRDD.groupByKey().mapValues(_._2.toList)

    def average[T] (ts: Iterable[T]) (implicit num: Numeric[T]) = {
      num.toDouble( ts.sum ) / ts.size
    }

    val averageGroupedMovieRatingTupleRDD = groupedMovieRatingTupleRDD.map(rec =>
(rec._1, average(rec._2)))

    val movieItemRDD = sc.textFile("hdfs://localhost:9000/u.item")

    val movieItemTupleRDD= movieItemRDD.map(
      line => (line.split("\\|")(0), (line.split("\\|")(0), line.split("\\|")(1), line.split("\\|")(2),
line.split("\\|")(4) )))

    val joinedRDD = movieItemTupleRDD.join(averageGroupedMovieRatingTupleRDD)

    val finalResultRDD = joinedRDD.map (
      rec => rec._2._1._1 + ", " + rec._2._1._2 + ", " + rec._2._1._3 + ", " + rec._2._1._4 + ", " +
rec._2._2)

    finalResultRDD.saveAsTextFile ("hdfs://localhost:9000/RatingResult")
    sc.stop()
  }
}
```

2) Scala build file:

name := "Movie Rating"

version := "1.0"

scalaVersion := "2.10.6"

**libraryDependencies += "org.apache.spark" %%
"spark-core" % "1.6.3"**

3) Sample run should be demonstrated at command line. Rather than taking screen shots, you can use script command line unix tool. Here is a demo:

\$ script myrun.out

Script started, file is myrun.out

{This starts a new shell and everything you type goes to your to the file called myrun.out, then run commands, then exit}

\$ spark-submit

\$ exit

Script done, file is myrun.out

You can simply paste your output file here.

```
Script started on Sun 04 Dec 2016 12:37:33 AM PST
j0;hdadmin@hdserver:~/MovieRating[?1034h[hdadmin@hdserver MovieRating]$ spark-submit --master
[Cyarn --deploy-mode cluster
-- [Knum-executors 2 --driver-memory [K 600m --executor-memory 600m target/scala2[K-2.10
/movie-rating_2.10-1.0.jar
16/12/04 00:42:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
16/12/04 00:42:33 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
16/12/04 00:42:34 INFO yarn.Client: Requesting a new application from cluster with 1 NodeManagers
16/12/04 00:42:34 INFO yarn.Client: Verifying our application has not requested more than the maximum
memory capability of the cluster (8192 MB per container)
16/12/04 00:42:34 INFO yarn.Client: Will allocate AM container, with 984 MB memory including 384 MB
overhead
16/12/04 00:42:34 INFO yarn.Client: Setting up container launch context for our AM
16/12/04 00:42:34 INFO yarn.Client: Setting up the launch environment for our AM container
16/12/04 00:42:34 INFO yarn.Client: Preparing resources for our AM container
Java HotSpot(TM) Server VM warning: You have loaded library /tmp/libnetty-transport-native-
epoll11219871908556726122.so which might have disabled stack guard. The VM will try to fix the stack
guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z
noexecstack'.
16/12/04 00:42:35 INFO yarn.Client: Uploading resource file:/usr/local/hadoop/spark-1.6.3-bin-without-
hadoop/lib/spark-assembly-1.6.3-hadoop2.2.0.jar -> hdfs://localhost:9000/user/hdadmin/.sparkStaging/
application_1480831749536_0001/spark-assembly-1.6.3-hadoop2.2.0.jar
16/12/04 00:42:38 INFO yarn.Client: Uploading resource file:/home/hdadmin/MovieRating/target/
scala-2.10/movie-rating_2.10-1.0.jar -> hdfs://localhost:9000/user/hdadmin/.sparkStaging/
application_1480831749536_0001/movie-rating_2.10-1.0.jar
16/12/04 00:42:39 INFO yarn.Client: Uploading resource file:/tmp/spark-
d59d21db-9f3e-47ed-87c6-928ef484c950/___spark_conf___7921801327606231057.zip -> hdfs://localhost:
9000/user/hdadmin/.sparkStaging/application_1480831749536_0001/
___spark_conf___7921801327606231057.zip
16/12/04 00:42:39 INFO spark.SecurityManager: Changing view acls to: hdadmin
16/12/04 00:42:39 INFO spark.SecurityManager: Changing modify acls to: hdadmin
16/12/04 00:42:39 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls
disabled; users with view permissions: Set(hdadmin); users with modify permissions: Set(hdadmin)
16/12/04 00:42:39 INFO yarn.Client: Submitting application 1 to ResourceManager
16/12/04 00:42:40 INFO impl.YarnClientImpl: Submitted application application_1480831749536_0001
16/12/04 00:42:41 INFO yarn.Client: Application report for application_1480831749536_0001 (state:
ACCEPTED)
16/12/04 00:42:41 INFO yarn.Client:
    client token: N/A
    diagnostics: N/A
    ApplicationMaster host: N/A
    ApplicationMaster RPC port: -1
```

queue: default
start time: 1480840960415
final status: UNDEFINED
tracking URL: http://localhost:8088/proxy/application_1480831749536_0001/
user: hdadmin

16/12/04 00:42:42 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:43 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:44 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:46 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:47 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:48 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:49 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:50 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:51 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:52 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:53 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:54 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:55 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:56 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:57 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:58 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:42:59 INFO yarn.Client: Application report for application_1480831749536_0001 (state: ACCEPTED)
16/12/04 00:43:00 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:00 INFO yarn.Client:
client token: N/A
diagnostics: N/A
ApplicationMaster host: 192.168.2.12
ApplicationMaster RPC port: 0
queue: default
start time: 1480840960415
final status: UNDEFINED
tracking URL: http://localhost:8088/proxy/application_1480831749536_0001/
user: hdadmin

16/12/04 00:43:01 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:02 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:03 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)


[illegible]

16/12/04 00:43:33 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:34 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:35 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:36 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:37 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:38 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:39 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:41 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:42 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:43 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:44 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:45 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:46 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:47 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:48 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:49 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:50 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:51 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:52 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:53 INFO yarn.Client: Application report for application_1480831749536_0001 (state: RUNNING)
16/12/04 00:43:54 INFO yarn.Client: Application report for application_1480831749536_0001 (state: FINISHED)
16/12/04 00:43:54 INFO yarn.Client:
 client token: N/A
 diagnostics: N/A
 ApplicationMaster host: 192.168.2.12
 ApplicationMaster RPC port: 0
 queue: default
 start time: 1480840960415
 final status: SUCCEEDED
 tracking URL: http://localhost:8088/proxy/application_1480831749536_0001/
 user: hdadmin
16/12/04 00:43:55 INFO util.ShutdownHookManager: Shutdown hook called
16/12/04 00:43:55 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-d59d21db-9f3e-47ed-87c6-928ef484c950
]0;hdadmin@hdserver:~/MovieRating[hdadmin@hdserver MovieRating]\$ eix[K[Kxit
exit
Script done on Sun 04 Dec 2016 12:45:33 AM PST

4) **Screen shot of home of page ResourceManager demonstrating that a Spark job called “Movie Rating” was run successfully.**

Quick Start - Spark ... x Apache Hadoop 2.7.... x All Applications x

localhost:8088/cluster

 **All Applications**

Cluster

- About
- Nodes
- Node Labels
- Applications
 - NEW
 - NEW SAVING
 - SUBMITTED
 - ACCEPTED
 - RUNNING
 - FINISHED
 - FAILED
 - KILLED
- Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	1	0	0 B	8 GB	0 B	0	8	0	1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCore

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress
application_1480831749536_0001	hdadmin	movie_rating	SPARK	default	Sun Dec 4 00:42:40 -0800 2016	Sun Dec 4 00:43:53 -0800 2016	FINISHED	SUCCEDED	

Showing 1 to 1 of 1 entries

- 5) For the movie id 376 and 495, what is the movie title, release date, IMDB URL, average rating that you got as result using Hive. Please fill the data grid below:

Movie id	Title	Release date	IMDB URL	Average rating
376	Houseguest(1994)	01-Jan-1994	http://us.imdb.com/M/title-exact?Houseguest%20(1994)	2.833333333333333
495	Around the world in 80 Days (1956)	01-Jan-1956	http://us.imdb.com/M/title-exact?Around%20the%20World%20in%2080%20Days%20(1956)	3.6610169491525424