

Assignment_46.11272

Student Name : Silvi Dheer

Course : Core Spark

Dataset :

```
1 2 3 4 5 6 7
```

```
name,subject,grade,marks,age
Mathew,science,grade-4,5,12
Mathew,history,grade-2,55,13
Mark,maths,grade-2,23,13
Mark,science,grade-1,76,13
John,history,grade-1,14,12
John,maths,grade-2,74,13
Lisa,science,grade-1,24,12
Lisa,history,grade-3,86,13
Andrew,maths,grade-1,34,13
Andrew,science,grade-3,26,14
Andrew,history,grade-1,74,12
Mathew,science,grade-2,55,12
Mathew,history,grade-2,87,12
Mark,maths,grade-1,92,13
Mark,science,grade-2,12,12
John,history,grade-1,67,13
John,maths,grade-1,35,11|
Lisa,science,grade-2,24,13
Lisa,history,grade-2,98,16
Andrew,maths,grade-1,23,16
Andrew,science,grade-3,44,14
Andrew,history,grade-2,77,11
```

Problem Statement

Problem Statement 1:

1. Read the text file, and create a tupled rdd.
2. Find the count of total number of rows present.
3. What is the distinct number of subjects present in the entire school
4. What is the count of the number of students in the school, whose name is Mathew and Marks is 55

Problem Statement 2:

1. What is the count of students per grade in the school?
2. Find the average of each student (Note - Mathew is grade-1, is different from Mathew in some other grade!)
3. What is the average score of students in each subject across all grades?
4. What is the average score of students in each subject per grade?
5. For all students in grade-2, how many have average score greater than 50?

Problem Statement 3:

Are there any students in the college that satisfy the below criteria:

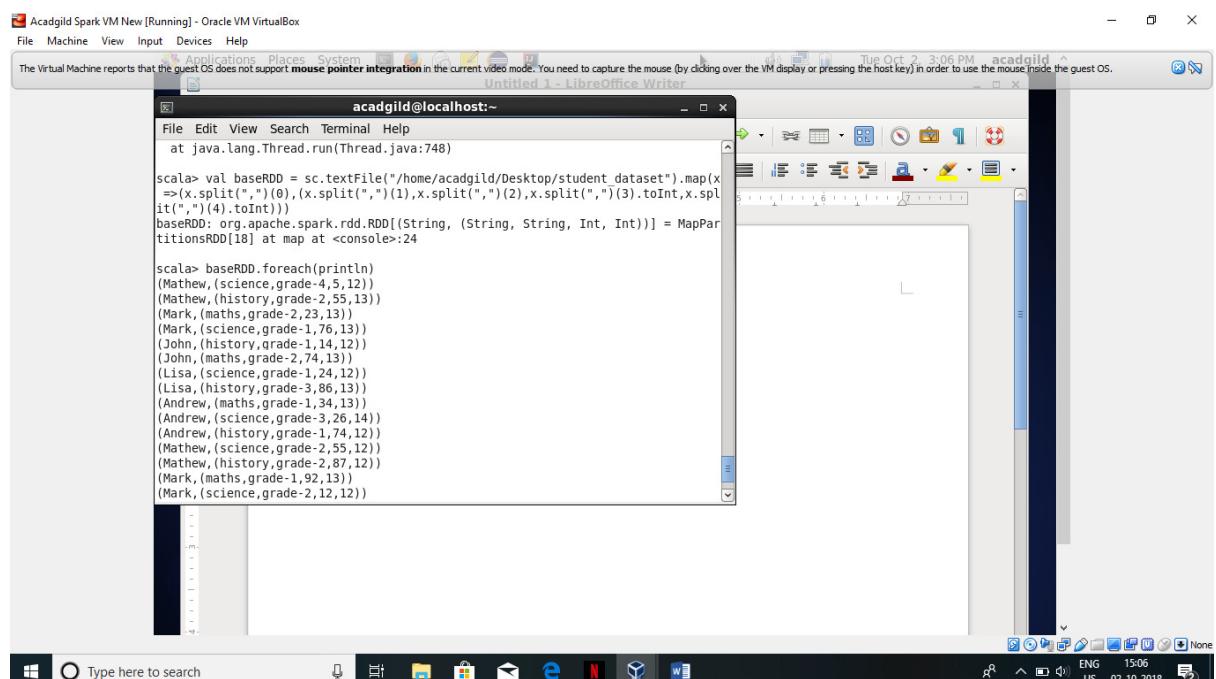
1. Average score per student_name across all grades is same as average score per student_name per grade
2. Hint - Use Intersection Property.

Spark Operations – Problem Statement 1

Task 1.1 Read the text file, and create a tupled rdd.

```
val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x  
=>(x.split(",")(0),(x.split(",")(1),x.split(",")(2),x.split(",")(3).toInt,x.split(",")(4).toInt)))  
baseRDD.foreach(println)
```

We have created a tuple RDD with name as Key and the subject, grades and the marks as values.



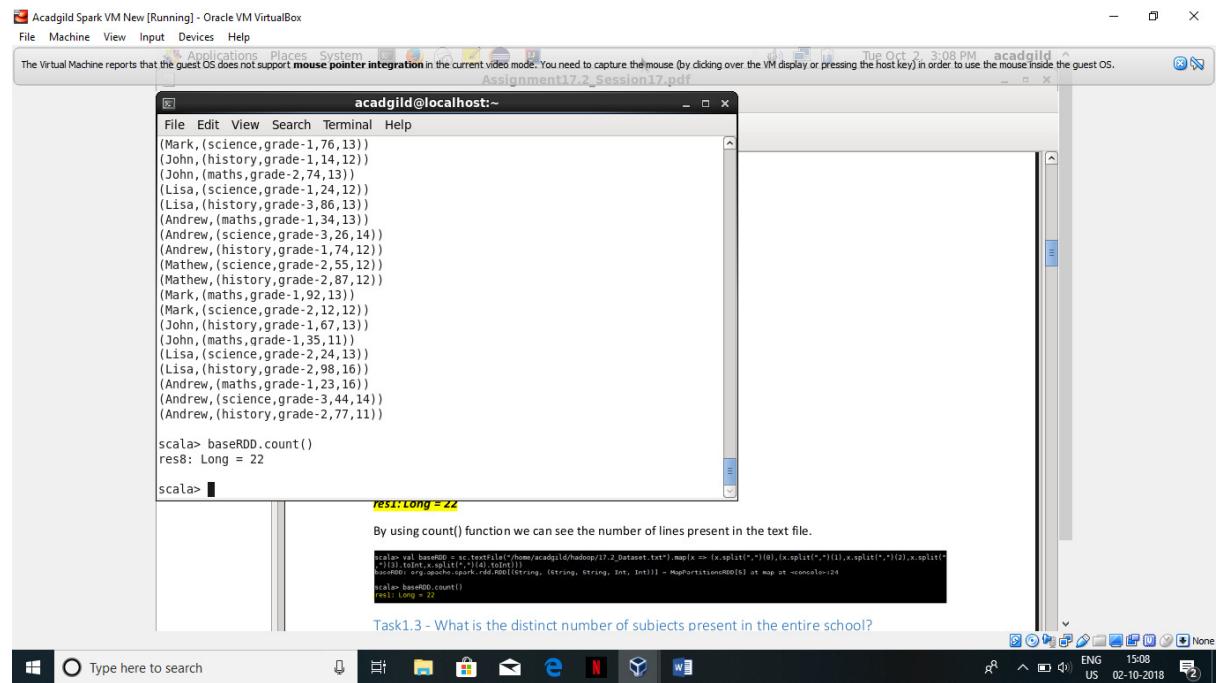
The screenshot shows a Windows desktop environment. In the foreground, a terminal window titled "acadgild@localhost:~" is open, displaying Scala code and its execution results. The code reads a text file and creates a tuple RDD where the first element is the student's name and the subsequent elements are tuples of subjects and their corresponding grades and marks. The output shows multiple rows of data for students like Mathew, Mark, and John across various subjects like Science, History, and Maths. In the background, a LibreOffice Writer document titled "Untitled 1" is visible, showing a blank page with a toolbar at the top.

```
acadgild@localhost:~$  
File Edit View Search Terminal Help  
at java.lang.Thread.run(Thread.java:748)  
scala> val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x  
=>(x.split(",")(0),(x.split(",")(1),x.split(",")(2),x.split(",")(3).toInt,x.split(",")(4).toInt)))  
baseRDD: org.apache.spark.rdd.RDD[(String, (String, String, Int, Int))] = MapPartitionsRDD[18] at map at <console>:24  
scala> baseRDD.foreach(println)  
(Mathew, (science,grade-4,5,12))  
(Mathew, (history,grade-2,55,13))  
(Mark, (maths,grade-2,23,13))  
(Mark, (science,grade-1,76,13))  
(John, (history,grade-1,14,12))  
(John, (maths,grade-2,74,13))  
(Lisa, (science,grade-1,24,12))  
(Lisa, (history,grade-3,86,13))  
(Andrew, (maths,grade-1,34,13))  
(Andrew, (science,grade-3,26,14))  
(Andrew, (history,grade-1,74,12))  
(Mathew, (science,grade-2,55,12))  
(Mathew, (history,grade-2,87,12))  
(Mark, (maths,grade-1,92,13))  
(Mark, (science,grade-2,12,12))
```

Task1.2 - Find the count of total number of rows present

```
baseRDD.count()
```

By using count() function we can see the number of lines present in the text file.



```
File Edit View Search Terminal Help
(acadgild Spark VM New [Running] - Oracle VM VirtualBox)
File Machine View Input Devices Help
Tue Oct 2 3:08 PM acadgild ^ Assignment17.2_Session17.pdf
acadgild@localhost:~ - □ ×
File Edit View Search Terminal Help
(acadgild Spark VM New [Running] - Oracle VM VirtualBox)
File Machine View Input Devices Help
Tue Oct 2 3:08 PM acadgild ^ Assignment17.2_Session17.pdf
(acadgild@localhost:~) 
(Mark, (science,grade-1,76,13))
(John, (history,grade-1,14,12))
(John, (maths,grade-2,74,13))
(Lisa, (science,grade-1,24,12))
(Lisa, (history,grade-3,86,13))
(Andrew, (maths,grade-1,34,13))
(Andrew, (science,grade-3,26,14))
(Andrew, (history,grade-1,74,12))
(Mathew, (science,grade-2,55,12))
(Mathew, (history,grade-2,87,12))
(Mathew, (maths,grade-1,92,13))
(Math, (science,grade-2,12,12))
(John, (history,grade-1,67,13))
(John, (maths,grade-1,35,11))
(Lisa, (science,grade-2,24,13))
(Lisa, (history,grade-2,98,16))
(Andrew, (maths,grade-1,23,16))
(Andrew, (science,grade-3,44,14))
(Andrew, (history,grade-2,77,11))

scala> baseRDD.count()
res8: Long = 22

scala> res8: Long = 22
By using count() function we can see the number of lines present in the text file.

scala> val baseRDD = sc.textFile("/home/acadgild/hadoop/17.2_Dataset.txt").map(x => (x.split(",")(0),(x.split(",")(1),x.split(",")(2),x.split(","))
(baseRDD: org.apache.spark.rdd.RDD[(String, String, Int, Int)]) = MapPartitionsRDD[5] at map at <console>:24
scala> baseRDD.count()
res8: Long = 22
Task1.3 - What is the distinct number of subjects present in the entire school?
Windows Type here to search ENG 15:08
US 02-10-2018 None
```

Task1.3 - What is the distinct number of subjects present in the entire school?

```
val distinctRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(",")(1),1))
```

```
val RDDreduce = distinctRDD.reduceByKey((x,y)=>(x+y))
```

```
RDDreduce.foreach(println)
```

First we are creating a RDD to read the file and selecting only subject name and mapping them with value 1 and counting the values of occurrences using **reduceByKey** to get distinct number of subjects

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Tue Oct 2 3:13 PM acadgild
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
Change desktop appearance and behavior, get help, or log out 1.odt - LibreOffice Writer
acadgild@localhost:~ 
File Edit View Search Terminal Help
(Lisa, (science,grade-2,24,13))
(Lisa, (history,grade-2,98,16))
(Andrew, (maths,grade-1,23,16))
(Andrew, (science,grade-3,44,14))
(Andrew, (history,grade-2,77,11))

scala> baseRDD.count()
res8: Long = 22

scala> val distinctRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(",")(0),1))
distinctRDD: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[21] at map at <console>:24
ap at <console>:24

scala> val RDDreduce = distinctRDD.reduceByKey((x,y)=>(x+y))
RDDreduce: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[22] at reduceByKey at <console>:25
Key at <console>:25

scala> RDDreduce.foreach(println)
(maths,6)
(history,8)
(science,8)

scala> 

```

Task1.4 - What is the count of the number of students in the school, whose name is Mathew and marks is 55

```
val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>((x.split(",")(0),x.split(",")(3).toInt),1))
```

In the first line code, we are reading the text file and creating a tuple RDD as “baseRDD” with name & marks as key and mapping numerical 1 as value.

```
val RDDfilter = baseRDD.filter(x=>x._1._1 == "Mathew" && x._1._2 == 55)
```

Filter the tuple RDD by providing the condition Mather and mark as 55,

```
val RDDreduce = RDDfilter.reduceByKey((x,y)=>x+y).foreach(println)
```

Now we are counting each occurrences using the reduceByKey and the required output is below,

The screenshot shows a Windows desktop environment with several open windows. In the foreground, there is a terminal window titled "acadgild@localhost:~" running Scala code. The code filters an RDD to find pairs where the first element is "Mathew" and the second element is 55, then reduces them by key to get a single count of 2. To the right of the terminal is a LibreOffice Writer document titled "Untitled 1.odt". The Writer document contains some text and a redacted section. The desktop taskbar at the bottom shows icons for File Explorer, Edge browser, File Explorer, and a LibreOffice icon. The system tray indicates the date as 02-10-2018 and the time as 15:19.

```
acadgild@localhost:~$ val RDDfilter = baseRDD.filter(x=> x._1 == "Mathew" && x._2 == 55)
RDDfilter: org.apache.spark.rdd.RDD[((String, Int), Int)] = MapPartitionsRDD[27]
at filter at <console>:25

scala> RDDfilter
res10: org.apache.spark.rdd.RDD[((String, Int), Int)] = MapPartitionsRDD[27] at
filter at <console>:25

scala> RDDfilter.connect
<console>:26: error: value connect is not a member of org.apache.spark.rdd.RDD[((String, Int), Int)]
      RDDfilter.connect
                           ^
name is

scala> RDDfilter.collect
res12: Array[((String, Int), Int)] = Array(((Mathew,55),1), ((Mathew,55),1))

scala> val RDDreduce = RDDfilter.reduceByKey((x,y)=> x+y).foreach(println)
((Mathew,55),2)
RDDreduce: Unit = ()

scala>
```

Spark Operations – Problem Statement 2

Task 2.1 - What is the count of students per grade in the school?

```
val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x => (x.split(",")(2),1)).reduceByKey((x,y)=>x+y).foreach(println)
```

we are reading the text file by creating a tuple RDD with grade as key and mapping numerical 1 as values and reducing the number occurrences using reduceByKey, please see the required output below,

Acadgild Spark VM New [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

Assignment17_2_Session17.pdf

acadgild@localhost:~

```
<console>:24: error: not found: value Val
  Val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x
=> (x.split(",")(2),1)).reduceByKey((x,y)=>x+y).foreach(println)
  ^
<console>:26: error: not found: value Val
val $ires7 = Val.baseRDD
  ^
<console>:24: error: not found: value Val
  Val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x
=> (x.split(",")(2),1)).reduceByKey((x,y)=>x+y).foreach(println)
  ^
<console>:26: error: not found: value Val
val $ires7 = Val.baseRDD
  ^
<console>:24: error: not found: value Val
  Val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x
=> (x.split(",")(2),1)).reduceByKey((x,y)=>x+y).foreach(println)
  ^
(grade-3,3)
(grade-1,9)
(grade-4,1)
(grade-2,9)
baseRDD: Unit = ()
```

ACADGILD

school?

op/17_2_Dataset.txt".map(x =>

key and mapping numerical 1 as values
see the required output below,

(-,*)(2,1)).reduceByKey((x,y)=>x+y).foreach(println)

new is grade-1, is different from

```
val baseRDD =
sc.textFile("/home/acadgild/hadoop/17_2_Dataset.txt").map(x=>(x.split(",")|0),x.split(",")|2),x.split(",")|3.toInt)

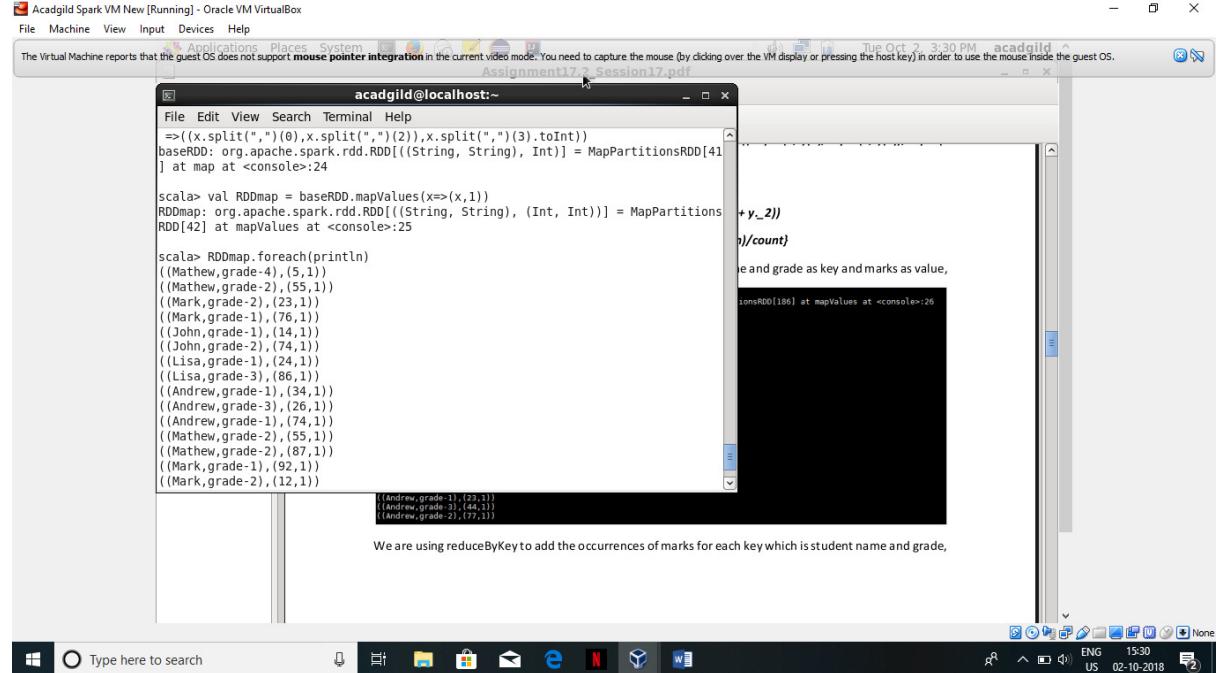
val RDDmap = baseRDD.mapValues(x=>x,1)

val RDDreduce = RDDmap.reduceByKey((x,y) => (x._1 + y._1, x._2 + y._2))
```

Type here to search

Task 2.2 - Find the average of each student (Note - Mathew is grade-1, is different from Mathew in some other grade!)

```
val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x  
=>((x.split(",")(0),x.split(",")(2)),x.split(",")(3).toInt))  
val RDDmap = baseRDD.mapValues(x=>(x,1))  
val RDDreduce = RDDmap.reduceByKey((x,y) => (x._1 + y._1, x._2 + y._2))  
val StudAvg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}  
First we are creating the baseRDD to read the file and selecting name and grade as key and marks as value,
```



```
Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Tue Oct 2 3:30 PM acadgild Assignment17 Session17.pdf
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
File Edit View Search Terminal Help
=>((x.split(",")(0),x.split(",")(2)),x.split(",")(3).toInt))
baseRDD: org.apache.spark.rdd.RDD[((String, String), Int)] = MapPartitionsRDD[41]
] at map at <console>:24

scala> val RDDmap = baseRDD.mapValues(x=>(x,1))
RDDmap: org.apache.spark.rdd.RDD[((String, String), (Int, Int))] = MapPartitions
RDD[42] at mapValues at <console>:25

scala> RDDmap.foreach(println)
((Mathew,grade-4),(5,1))
((Mathew,grade-2),(55,1))
((Mark,grade-2),(23,1))
((Mark,grade-1),(76,1))
((John,grade-1),(14,1))
((John,grade-2),(74,1))
((Lisa,grade-1),(24,1))
((Lisa,grade-3),(66,1))
((Andrew,grade-1),(34,1))
((Andrew,grade-3),(26,1))
((Andrew,grade-1),(74,1))
((Mathew,grade-2),(55,1))
((Mathew,grade-2),(87,1))
((Mark,grade-1),(92,1))
((Mark,grade-2),(12,1))

((Andrew,grade-1),(23,1))
((Andrew,grade-3),(44,1))
((Andrew,grade-2),(77,1))

We are using reduceByKey to add the occurrences of marks for each key which is student name and grade,
```

We are using reduceByKey to add the occurrences of marks for each key which is student name and grade,

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ File Edit View Search Terminal Help
((Lisa,grade-2),(98,1))
((Andrew,grade-1),(23,1))
((Andrew,grade-3),(44,1))
((Andrew,grade-2),(77,1))

scala> val RDDreduce = RDDmap.reduceByKey((x,y) => (x._1 + y._1, x._2 + y._2))
RDDreduce: org.apache.spark.rdd.RDD[((String, String), (Int, Int))] = ShuffledRDD
D[43] at reduceByKey at <console>:25

scala> RDDreduce.foreach(println)
((Lisa,grade-1),(24,1))
((Mark,grade-2),(35,2))
((Lisa,grade-2),(122,2))
((Andrew,grade-2),(77,1))
((Andrew,grade-1),(131,3))
((Lisa,grade-3),(86,1))
((John,grade-1),(116,3))
((Mathew,grade-4),(5,1))
((John,grade-2),(74,1))
((Mark,grade-1),(168,2))
((Andrew,grade-3),(70,2))
((Mathew,grade-2),(197,3))

scala> 
[[Andrew,grade-3),(70,2)]
[(Mathew,grade-2),(197,3)]

Now we are calculating average by summing the marks and dividing by its count for each key. Below screenshot shows the final result,
scala> val StudAvg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}
StudAvg: org.apache.spark.rdd.RDD[((String, String), Double)] = MapPartitionsRDD[188] at mapValues at <console>:30
scala> StudAvg.foreach(println)
((Lisa,grade-1),24.0)

```

Now we are calculating average by summing the marks and dividing by its count for each key. Below screenshot shows the final result,

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ File Edit View Search Terminal Help
((John,grade-2),(74,1))
((Mark,grade-1),(168,2))
((Andrew,grade-3),(76,2))
((Mathew,grade-2),(197,3))

scala> val StudAvg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}
StudAvg: org.apache.spark.rdd.RDD[((String, String), Double)] = MapPartitionsRDD
D[44] at mapValues at <console>:25

scala> StudAvg.foreach(println)
((Lisa,grade-1),24.0)
((Mark,grade-2),17.5)
((Lisa,grade-2),61.0)
((Andrew,grade-2),77.0)
((Andrew,grade-1),43.666666666666664)
((Lisa,grade-3),86.0)
((John,grade-1),38.666666666666664)
((Mathew,grade-4),5.0)
((John,grade-2),74.0)
((Mark,grade-1),84.0)
((Andrew,grade-3),35.0)
((Mathew,grade-2),65.66666666666667)

scala> 

```

Task 2.3 - What is the average score of students in each subject across all grades?

```

val baseRDD =
sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>((x.split(",")(0),x.split(",")(1)
),x.split(",")(3).toInt))
val RDDmap = baseRDD.mapValues(x=>(x,1))

```

```
val RDDreduce = RDDmap.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
val SubAvg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}
```

We are first creating baseRDD to read the text file and we are extracting name and subject as key and marks as value,

```
scala> baseRDD.foreach(println)
[Stage 0:>
((Mathew,science),5)
((Mathew,history),55)
((Mark,maths),23)
((Mark,science),76)
((John,history),14)
((John,maths),74)
((Lisa,science),24)
((Lisa,history),86)
((Andrew,maths),34)
((Andrew,science),26)
((Andrew,history),74)
((Mathew,science),55)
((Mathew,history),87)
((Mark,maths),92)
((Mark,science),12)
((John,history),67)
((John,maths),35)
((Lisa,science),24)
((Lisa,history),98)
((Andrew,maths),23)
((Andrew,science),44)

Now using mapValues we are mapping each value with 1-
scala> val RDDmap = baseRDD.mapValues(x=>(x,1))
RDDmap: org.apache.spark.rdd.RDD[((String, String), (Int, Int))] = MapPartitionsRDD[3] at mapValues at <console>:26
```

Now using mapValues we are mapping each value with 1-

```
scala> val RDDmap = baseRDD.mapValues(x=>(x,1))
RDDmap: org.apache.spark.rdd.RDD[((String, String), (Int, Int))] = MapPartitionsRDD[3] at mapValues at <console>:25

scala> RDDmap.foreach(println)
((Mathew,science),(5,1))
((Mathew,history),(55,1))
((Mark,maths),(23,1))
((Mark,science),(76,1))
((John,history),(14,1))
((John,maths),(74,1))
((Lisa,science),(24,1))
((Lisa,history),(86,1))
((Andrew,maths),(34,1))
((Andrew,science),(26,1))
((Andrew,history),(74,1))
((Mathew,science),(55,1))
((Mathew,history),(87,1))
((Mark,maths),(92,1))
((Mark,science),(12,1))
((John,history),(67,1))
((John,maths),(35,1))
((Lisa,science),(24,1))

Task 2.3 - What is the average score of students in each subject across all grades?
Codes,
val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>((x.split(","))
```

Now we are adding the marks and number of occurrences for each key using reduceByKey.

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ 
File Edit View Search Terminal Help
((Lisa,science),(24,1))
((Lisa,history),(98,1))
((Andrew,maths),(23,1))
((Andrew,science),(44,1))
((Andrew,history),(77,1))

scala> val RDDreduce = RDDmap.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
RDDreduce: org.apache.spark.rdd.RDD[(String, String), (Int, Int)] = ShuffledRDD[4] at reduceByKey at <console>:25

scala> RDDreduce.foreach(println)
((Lisa,history),(184,2))
((Mark,maths),(115,2))
((Andrew,science),(70,2))
((Mark,science),(88,2))
((Mathew,science),(60,2))
((Andrew,maths),(57,2))
((Mathew,history),(142,2))
((John,maths),(109,2))
((John,history),(81,2))
((Lisa,science),(48,2))
((Andrew,history),(151,2))

scala> 

```

(0).x.split("")(1).x.split("")(3.toInt)
val RDDmap = baseRDD.mapValues(x=>(x,1))
val RDDreduce = RDDmap.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
val SubAvg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}

calculating average by dividing the sum of marks and count of occurrences for each key.

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ 
File Edit View Search Terminal Help
((Mathew,history),(142,2))
((John,maths),(109,2))
((John,history),(81,2))
((Lisa,science),(48,2))
((Andrew,history),(151,2))

scala> val SubAvg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}
SubAvg: org.apache.spark.rdd.RDD[(String, String), Double] = MapPartitionsRDD[5] at mapValues at <console>:25

scala> SubAvg.foreach(println)
((Lisa,history),92.0)
((Mark,maths),57.5)
((Andrew,science),35.0)
((Mark,science),44.0)
((Mathew,science),30.0)
((Andrew,maths),28.5)
((Mathew,history),71.0)
((John,maths),54.5)
((John,history),40.5)
((Lisa,science),24.0)
((Andrew,history),75.5)

scala> 

```

(0).x.split("")(1).x.split("")(3.toInt)
val RDDmap = baseRDD.mapValues(x=>(x,1))
val RDDreduce = RDDmap.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
val SubAvg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}

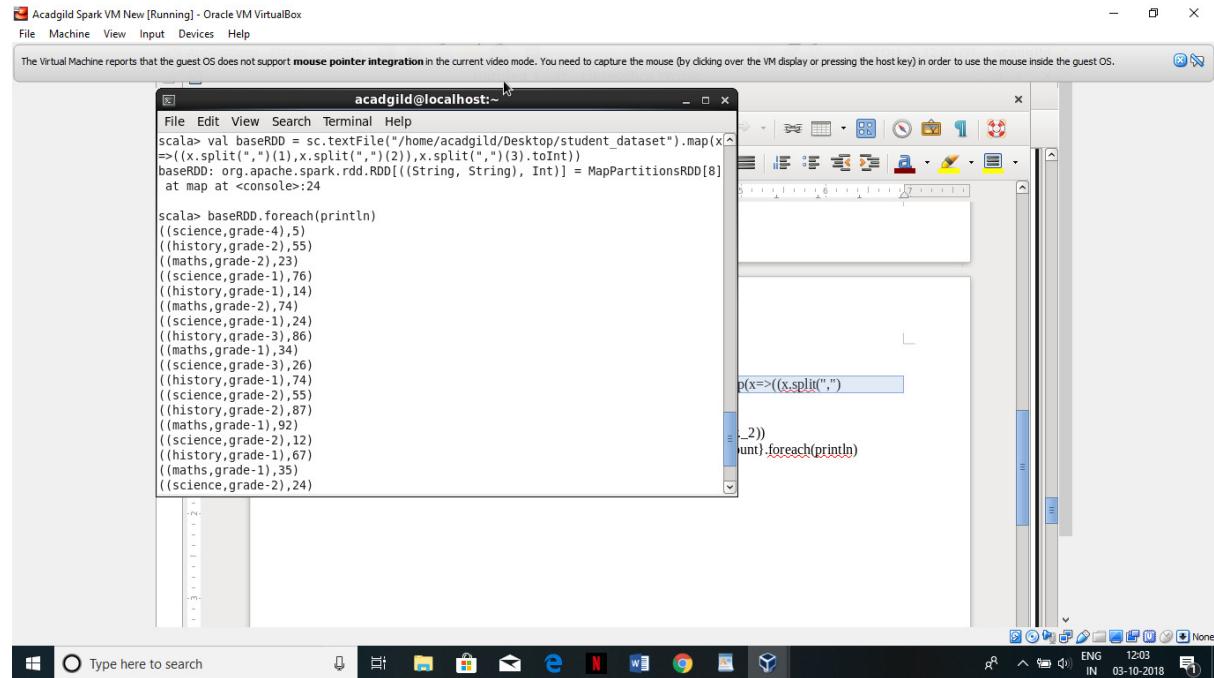
Task 2.4 - What is the average score of students in each subject per grade?

```

val baseRDD =
sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>((x.split(",")(1).x.split(",")(2),x.split(",")(3.toInt)))
val RDDmapvalue = baseRDD.mapValues(x=>(x,1))
val RDDreduce = RDDmapvalue.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
val Avg_Grade = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}.foreach(println)

```

In first step we are creating paired RDD named as baseRDD to read the file and extracting subject and grade as key and marks as value



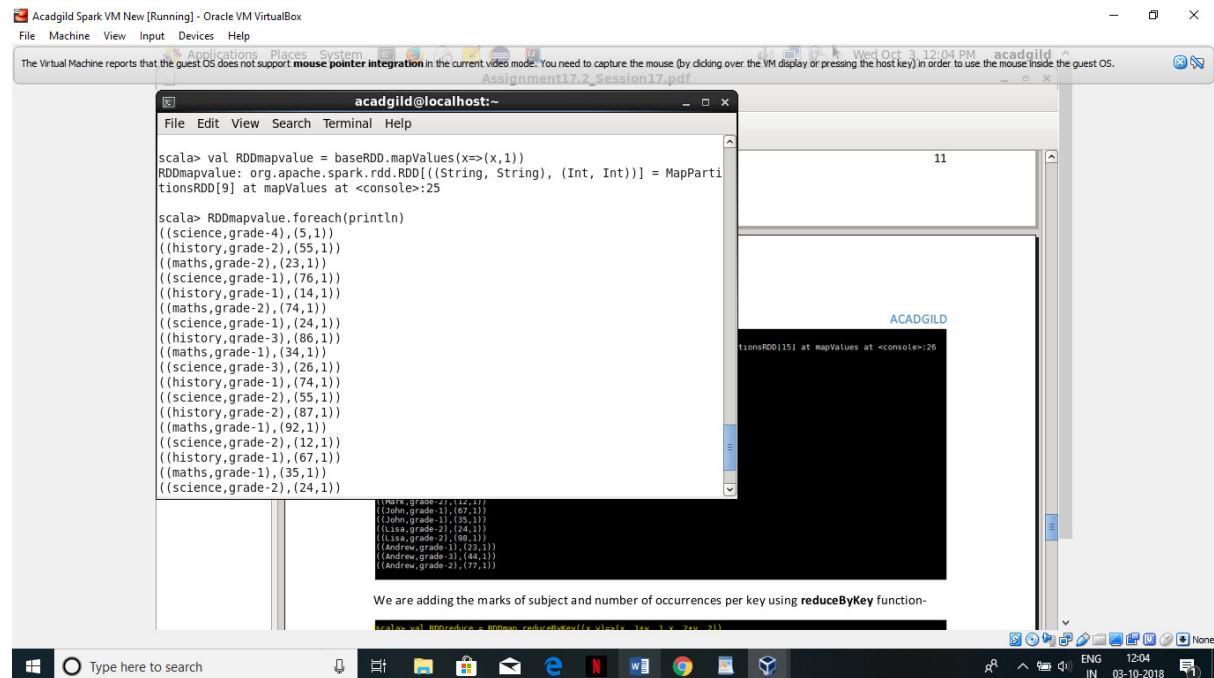
```
Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

acadgild@localhost:~$ File Edit View Search Terminal Help
scala> val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(",")(1),x.split(",")(2)),x.split(",")(3).toInt)
baseRDD: org.apache.spark.rdd.RDD[(String, String), Int] = MapPartitionsRDD[8]
at map at <console>:24

scala> baseRDD.foreach(println)
((science,grade-4),5)
((history,grade-2),55)
((maths,grade-2),23)
((science,grade-1),76)
((history,grade-1),14)
((maths,grade-2),74)
((science,grade-1),24)
((history,grade-3),86)
((maths,grade-1),34)
((science,grade-3),26)
((history,grade-1),74)
((science,grade-2),55)
((history,grade-2),87)
((maths,grade-1),92)
((science,grade-2),12)
((history,grade-1),67)
((maths,grade-1),35)
((science,grade-2),24)

-ri-
-r-
-r-
-r-
-r-
-r-
-r-
-r-
-m-
```

Then we are mapping the values of baseRDD with numerical value 1 using function mapValues



```
Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

acadgild@localhost:~$ Applications Places System
assignment17_2_Session17.pdf Wed Oct 3 12:04 PM acadgild ^

File Edit View Search Terminal Help
scala> val RDDmapvalue = baseRDD.mapValues(x=>(x,1))
RDDmapvalue: org.apache.spark.rdd.RDD[(String, String), (Int, Int)] = MapPartitionsRDD[9] at mapValues at <console>:25

scala> RDDmapvalue.foreach(println)
((science,grade-4),(5,1))
((history,grade-2),(55,1))
((maths,grade-2),(23,1))
((science,grade-1),(76,1))
((history,grade-1),(14,1))
((maths,grade-2),(74,1))
((science,grade-1),(24,1))
((history,grade-3),(86,1))
((maths,grade-1),(34,1))
((science,grade-3),(26,1))
((history,grade-1),(74,1))
((science,grade-2),(55,1))
((history,grade-2),(87,1))
((maths,grade-1),(92,1))
((science,grade-2),(12,1))
((history,grade-1),(67,1))
((maths,grade-1),(35,1))
((science,grade-2),(24,1))

11
ACADGILD
tionsRDD[15] at mapValues at <console>:26

We are adding the marks of subject and number of occurrences per key using reduceByKey function-
scala> val reduced = baseRDD.reduceByKey(f=identity, s=Sum(0))
```

We are adding marks and number of occurrences for each key using reduceByKey

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ 
File Edit View Search Terminal Help
((history,grade-1), (67,1))
((maths,grade-1), (35,1))
((science,grade-2), (24,1))
((history,grade-2), (98,1))
((maths,grade-1), (23,1))
((science,grade-3), (44,1))
((history,grade-3), (77,1))

scala> val RDDreduce = RDDmapvalue.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
RDDreduce: org.apache.spark.rdd.RDD[((String, String), (Int, Int))] = ShuffledRDD
D[10] at reduceByKey at <console>:25

scala> RDDreduce.foreach(println)
((history,grade-2), (317,4))
((history,grade-3), (86,1))
((science,grade-4), (5,1))
((maths,grade-1), (184,4))
((science,grade-3), (70,2))
((science,grade-1), (100,2))
((science,grade-2), (91,3))
((history,grade-1), (155,3))
((maths,grade-2), (97,2))

scala> 

```

We are calculating average by dividing the sum of marks with number of occurrences and the requested output is shown below,

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ 
File Edit View Search Terminal Help
((history,grade-2), (317,4))
((history,grade-3), (86,1))
((science,grade-4), (5,1))
((maths,grade-1), (184,4))
((science,grade-3), (70,2))
((science,grade-1), (100,2))
((science,grade-2), (91,3))
((history,grade-1), (155,3))
((maths,grade-2), (97,2))

scala> val Avg_Grade = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}.foreach(println)
((history,grade-2),79.25)
((history,grade-3),86.0)
((science,grade-4),5.0)
((maths,grade-1),46.0)
((science,grade-3),35.0)
((science,grade-1),50.0)
((science,grade-2),30.33333333333332)
((history,grade-1),51.66666666666664)
((maths,grade-2),48.5)
Avg_Grade: Unit = ()

scala> 

```

Task 2.5 - for all students in grade-2, how many have average score greater than 50?

```

val baseRDD =
sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>((x.split(",")(0),x.split(",")(2)),x.split(",")(3).toInt))
val RDDmap = baseRDD.mapValues(x=>(x,1))

```

```

val RDDreduce = RDDmap.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
val RDDavg = RDDreduce.mapValues{case(sum,count)=>(1.0*sum)/count}
val RDDfiltermap = RDDavg.filter(x=>x._1._2 == "grade-2" && x._2>50).count()
val RDDfiltermap = RDDavg.filter(x=>x._1._2 == "grade-2" && x._2>50).foreach(println)

```

Creating a paired RDD named as baseRDD to read the file and extracting name and grade as key and marks as value

```

File Edit View Search Terminal Help
Type :help for more information.

scala> val baseRDD = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(",")(0),x.split(",")(2)).x.split(",")(3).toInt)
baseRDD: org.apache.spark.rdd.RDD[(String, String), Int] = MapPartitionsRDD[2] at map at <console>:24

scala> baseRDD.foreach(println)
[Stage 0:>
((Mathew,grade-4),5)
((Mathew,grade-2),55)
((Mark,grade-2),23)
((Mark,grade-1),76)
((John,grade-1),14)
((John,grade-2),74)
((Lisa,grade-1),24)
((Lisa,grade-3),86)
((Andrew,grade-1),34)
((Andrew,grade-3),26)
((Andrew,grade-1),74)
((Mathew,grade-2),55)
((Mathew,grade-2),87)
((Mark,grade-1),92)
((Mark,grade-2),12)

```

Now we are mapping each value of baseRDD with 1 as shown below,

```

File Edit View Search Terminal Help
Type :help for more information.

scala> val RDDmap = baseRDD.mapValues(x=>(x,1))
RDDmap: org.apache.spark.rdd.RDD[(String, String), (Int, Int)] = MapPartitionsRDD[3] at mapValues at <console>:25

scala> RDDmap.foreach(println)
((Mathew,grade-4),(5,1))
((Mathew,grade-2),(55,1))
((Mark,grade-2),(23,1))
((Mark,grade-1),(76,1))
((John,grade-1),(14,1))
((John,grade-2),(74,1))
((Lisa,grade-1),(24,1))
((Lisa,grade-3),(86,1))
((Andrew,grade-1),(34,1))
((Andrew,grade-3),(26,1))
((Andrew,grade-1),(74,1))
((Mathew,grade-2),(87,1))
((Mathew,grade-2),(55,1))
((Mark,grade-1),(92,1))
((Mark,grade-2),(12,1))
((John,grade-1),(67,1))
((John,grade-1),(35,1))
((Lisa,grade-2),(24,1))

```

We are adding the marks of subject and number of occurrences per key using reduceByKey function-

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

acadgild@localhost:~ % File Edit View Search Terminal Help
((Lisa,grade-2),(98,1))
((Andrew,grade-1),(23,1))
((Andrew,grade-3),(44,1))
((Andrew,grade-2),(77,1))

scala> val RDDreduce = RDDmap.reduceByKey((x,y)>(x, 1+y, 1,x, 2+y, 2))
RDDreduce: org.apache.spark.rdd.RDD[(String, String), (Int, Int)] = ShuffledRDD
D[4] at reduceByKey at <console>:25

scala> RDDreduce.foreach(println)
((Lisa,grade-1),(24,1))
((Mark,grade-2),(35,2))
((Lisa,grade-2),(122,2))
((Andrew,grade-2),(77,1))
((Andrew,grade-1),(131,3))
((Lisa,grade-3),(66,1))
((John,grade-1),(116,3))
((Mathew,grade-4),(5,1))
((John,grade-2),(74,1))
((Mark,grade-1),(168,2))
((Andrew,grade-3),(78,2))
((Mathew,grade-2),(197,3))

scala>

```

Here we are calculating the average of each student,

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

acadgild@localhost:~ % File Edit View Search Terminal Help
((John,grade-2),(74,1))
((Mark,grade-1),(168,2))
((Andrew,grade-3),(70,2))
((Mathew,grade-2),(197,3))

scala> val RDDavg = RDDreduce.mapValues{case(sum,count)>=(1.0*sum)/count}
RDDavg: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[5] at mapValues at <console>:25

scala> RDDavg.foreach(println)
((Lisa,grade-1),24.0)
((Mark,grade-2),17.5)
((Lisa,grade-2),61.0)
((Andrew,grade-2),77.0)
((Andrew,grade-1),43.666666666666664)
((Lisa,grade-3),86.0)
((John,grade-1),38.666666666666664)
((Mathew,grade-4),5.0)
((John,grade-2),74.0)
((Mark,grade-1),84.0)
((Andrew,grade-3),35.0)
((Mathew,grade-2),65.66666666666667)

scala>

```

Now in below step we are filtering the above result with student belonging to grade-2 and having marks greater than 50, the number of the counts and the data is shown in the below.

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ 
File Edit View Search Terminal Help
((Lisa,grade-2),61.0)
((Andrew,grade-2),77.0)
((Andrew,grade-1),43.666666666666664)
((Lisa,grade-3),86.0)
((John,grade-1),38.666666666666664)
((Mathew,grade-4),5.0)
((John,grade-2),74.0)
((Mark,grade-1),84.0)
((Andrew,grade-3),35.0)
((Mathew,grade-2),65.6666666666667)

scala> val RDDfiltermap = RDDavg.filter(x=>x._1._2 == "grade-2" && x._2>50).count()
RDDfiltermap: Long = 4

scala> val RDDfiltermap = RDDavg.filter(x=>x._1._2 == "grade-2" && x._2>50).foreach(println)
ach println
((Lisa,grade-2),61.0)
((Andrew,grade-2),77.0)
((John,grade-2),74.0)
((Mathew,grade-2),65.6666666666667)
RDDfiltermap: Unit = ()

scala>

```

Spark Operations – Problem Statement 3

Are there any students in the college that satisfy the below criteria:

1. Average score per student_name across all grades is same as average score per student_name per grade

Solution- To find the solution of above problem we will first calculate average of each student across all grades i.e. irrespective of grade. Below is the code used to find the same

```

val baseRDD1 =
sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(",")(0),x.split(",")(3).toInt))
val studAvg = baseRDD1.mapValues(x=>(x,1)).foreach(println)
val studReduce = studAvg.reduceByKey((x,y)=> (x._1+y._1,x._2+y._2))
val Avg_Stud = studReduce.mapValues{case (sum,count) => (1.0 * sum)/count}

```

We created a paired RDD named as baseRDD1 by extracting only name and marks,

Acadgild Spark VM New [Running] - Oracle VM VirtualBox

The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

```
acadgild@localhost:~
```

```

File Edit View Terminal Help
scala> val baseRDD1 = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(",")(0),x.split(",")(3).toInt))
baseRDD1: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[13] at map at <console>:24
scala> baseRDD1.foreach(println)
(Mathew,5)
(Mathew,55)
(Mark,23)
(Mark,76)
(John,14)
(John,74)
(Lisa,24)
(Lisa,86)
(Andrew,34)
(Andrew,26)
(Andrew,74)
(Mathew,55)
(Mathew,87)
(Mark,92)
(Mark,12)
(John,67)
(John,35)
(Lisa,24)

per grade
val baseRDD1 = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(","))
(0),x.split(",")(3).toInt)
val studAvg = baseRDD1.mapValues(x=>(x,1).foreach(println))
scala> val studReduce = studAvg.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
scala> val Avg_Stud = studReduce.mapValues(case (sum,count) => (1.0 * sum)/count)

```

Then we are mapping each value of above RDD with 1,

Acadgild Spark VM New [Running] - Oracle VM VirtualBox

The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

```
acadgild@localhost:~
```

```

File Edit View Terminal Help
scala> val studAvg = baseRDD1.mapValues(x=>(x,1)).foreach(println)
(Mathew,(5,1))
(Mathew,(55,1))
(Mark,(23,1))
(Mark,(76,1))
(John,(14,1))
(John,(74,1))
(Lisa,(24,1))
(Lisa,(86,1))
(Andrew,(34,1))
(Andrew,(26,1))
(Andrew,(74,1))
(Mathew,(55,1))
(Mathew,(87,1))
(Mark,(92,1))
(Mark,(12,1))
(John,(67,1))
(John,(35,1))
(Lisa,(24,1))
(Lisa,(98,1))
(Andrew,(23,1))
(Andrew,(44,1))
(Andrew,(77,1))

per grade
val baseRDD1 = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>(x.split(","))
(0),x.split(",")(3).toInt)
val studAvg = baseRDD1.mapValues(x=>(x,1))
val studReduce = studAvg.reduceByKey((x,y)=>(x._1+y._1,x._2+y._2))
val Avg_Stud = studReduce.mapValues(case (sum,count) => (1.0 * sum)/count)

```

Then we are adding the marks and number of occurrences for each student using `reduceByKey` as shown below, and calculating the average of each student,

```

File Edit View Terminal Help
scala> val studReduce = studAvg.reduceByKey((x,y)=> (x._1+y._1,x._2+y._2))
studReduce: org.apache.spark.rdd.RDD[(String, (Int, Int))] = ShuffledRDD[16] at
reduceByKey at <console>:25

scala> studReduce.foreach(println)
(Mark, (203,4))
(Andrew, (278,6))
(Mathew, (202,4))
(John, (198,4))
(Lisa, (232,4))

scala> val Avg_Stud = studReduce.mapValues{case (sum,count) => (1.0 * sum)/count}
Avg_Stud: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[17] at m
apValues at <console>:25

scala> Avg_Stud.foreach(println)
(Mark,50.75)
(Andrew,46.33333333333336)
(Mathew,50.5)
(John,47.5)
(Lisa,58.0)

scala>

```

Now the second step of this problem is to find the average of each student per grade. We have used below code to find,

```

val baseRDD2 =
sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>((x.split(",")(0),x.split(",")(2),
),x.split(",")(3).toInt)).foreach(println)
val grade = baseRDD2.mapValues(x=>(x,1))
val gradeReduce = grade.reduceByKey((x,y)=> (x._1+y._1,x._2+y._2))
val gradeAvg = gradeReduce.mapValues{case(sum,count) => (1.0*sum)/count}

```

So first we are creating another paired RDD named as baseRDD2 by extracting name and grade as key and marks as value from the input file,

```

File Edit View Search Terminal Help
scala> val baseRDD2 = sc.textFile("/home/acadgild/Desktop/student_dataset").map(x=>((x.split(",")(0),x.split(",")(2),x.split(",")(3).toInt)).foreach(println)
((Mathew,grade-4),5)
((Mathew,grade-2),55)
((Mark,grade-2),23)
((Mark,grade-1),76)
((John,grade-1),14)
((John,grade-2),74)
((Lisa,grade-1),24)
((Lisa,grade-3),86)
((Andrew,grade-1),34)
((Andrew,grade-3),26)
((Andrew,grade-1),74)
((Mathew,grade-2),55)
((Mathew,grade-2),87)
((Mark,grade-1),92)
((Mark,grade-2),12)
((John,grade-1),67)
((John,grade-1),35)
((Lisa,grade-2),24)
((Lisa,grade-2),98)
((Andrew,grade-1),23)
((Andrew,grade-3),44)
((Andrew,grade-2),77)

```

Then we are mapping each value of baseRDD2 with 1 using mapValues function

The screenshot shows a Windows desktop environment. In the foreground, there is a Microsoft Word document titled "Manage your email, contacts and schedule". In the background, a Scala REPL window is open in a terminal emulator. The code being run is:

```
scala> val grade = baseRDD2.mapValues(x=>(x,1))
grade: org.apache.spark.rdd.RDD[(String, String), (Int, Int)] = MapPartitionsRDD[24] at mapValues at <console>:25
scala> grade.foreach(println)
((Mathew,grade-4),(5,1))
((Mathew,grade-2),(55,1))
((Mark,grade-2),(23,1))
((Mark,grade-1),(76,1))
((John,grade-1),(14,1))
((John,grade-2),(74,1))
((Lisa,grade-1),(24,1))
((Lisa,grade-3),(86,1))
((Andrew,grade-1),(34,1))
((Andrew,grade-3),(26,1))
((Andrew,grade-1),(74,1))
((Mathew,grade-2),(55,1))
((Mathew,grade-2),(87,1))
((Mark,grade-1),(92,1))
((Mark,grade-2),(12,1))
((John,grade-1),(67,1))
((John,grade-1),(35,1))
((Lisa,grade-2),(24,1))
```

Then we are adding the marks and number of occurrences of 1 for each key using reduceByKey() function

The screenshot shows a Windows desktop environment. In the foreground, there is a Microsoft Word document titled "Manage your email, contacts and schedule". In the background, a Scala REPL window is open in a terminal emulator. The code being run is:

```
scala> val gradeReduce = grade.reduceByKey((x,y)=> (x._1+y._1,x._2+y._2))
gradeReduce: org.apache.spark.rdd.RDD[(String, String), (Int, Int)] = ShuffledRDD[25] at reduceByKey at <console>:25
scala> gradeReduce.foreach(println)
((Lisa,grade-2),(98,1))
((Andrew,grade-1),(23,1))
((Andrew,grade-3),(44,1))
((Andrew,grade-2),(77,1))
scala> gradeReduce.foreach(println)
((Lisa,grade-1),(24,1))
((Mark,grade-2),(35,2))
((Lisa,grade-2),(122,2))
((Andrew,grade-2),(77,1))
((Andrew,grade-1),(131,3))
((Lisa,grade-3),(86,1))
((John,grade-1),(116,3))
((Mathew,grade-4),(5,1))
((John,grade-2),(74,1))
((Mark,grade-1),(168,2))
((Andrew,grade-3),(78,2))
((Mathew,grade-2),(197,3))
```

In below step we are calculating average of each key by dividing the sum of marks with the count

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ % File Edit View Search Terminal Help
((John,grade-2),(74,1))
((Mark,grade-1),(168,2))
((Andrew,grade-3),(76,2))
((Mathew,grade-2),(197,3))

scala> val gradeAvg = gradeReduce.mapValues{case(sum,count) => (1.0*sum)/count}
gradeAvg: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[0][26] at mapValues at <console>:25

scala> gradeAvg.foreach(println)
((Lisa,grade-1),24.0)
((Mark,grade-2),17.5)
((Lisa,grade-2),61.0)
((Andrew,grade-2),77.0)
((Andrew,grade-1),43.66666666666664)
((Lisa,grade-3),86.0)
((John,grade-1),38.66666666666664)
((Mathew,grade-4),5.0)
((John,grade-2),74.0)
((Mark,grade-1),84.0)
((Andrew,grade-3),35.0)
((Mathew,grade-2),65.66666666666667)

scala>

```

Now to proceed further we are extracting name and marks from above RDD,

In below step we are using intersection function between flatgradeAvg and flatnameAvg rdd's to find whether any common student is there.

```

val flatgradeAvg = gradeAvg.map(x=> x._1._1 + "," + x._2.toDouble)
val flatAvg_Stud = Avg_Stud.map(x => x._1 + "," + x._2)
val commandval = flatgradeAvg.intersection(flatAvg_Stud)

```

```

Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.
acadgild@localhost:~ % File Edit View Search Terminal Help
((John,grade-2),74.0)
((Mark,grade-1),84.0)
((Andrew,grade-3),35.0)
((Mathew,grade-2),65.66666666666667)

scala> val flatgradeAvg = gradeAvg.map(x=> x._1._1 + "," + x._2.toDouble)
flatgradeAvg: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[27] at map at <console>:25

scala> flatgradeAvg.foreach(println)
Lisa,24.0
Mark,17.5
Lisa,61.0
Andrew,77.0
Andrew,43.66666666666664
Lisa,86.0
John,38.66666666666664
Mathew,5.0
John,74.0
Mark,84.0
Andrew,35.0
Mathew,65.66666666666667

scala>

```

```
Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

acadgild@localhost:~ 
File Edit View Search Terminal Help
<console>:25: error: not found: value flatAvg_Stud
      val commanval = flatgradeAvg.intersection(flatAvg_Stud)
                                         ^
scala> com
com  compat
scala> commanval.foreach(println)
<console>:24: error: not found: value commanval
      commanval.foreach(println)
                                         ^
scala> val flatAvg_Stud = Avg_Stud.map(x => x._1 + "," + x._2)
flatAvg_Stud: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[28] at map at
<console>:25
scala> flatAvg_Stud.foreach(println)
Mark,50.75
Andrew,46.33333333333336
Mathew,50.5
John,47.5
Lisa,58.0
scala> 
scala> val commanval = flatgradeAvg.intersection(flatAvg_Stud)
```

So the command `commanval.foreach(println)` shows that no common students are there having average score per `student_name` across all grades is same as average score per `student_name` per grade-

```
Acadgild Spark VM New [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
The Virtual Machine reports that the guest OS does not support mouse pointer integration in the current video mode. You need to capture the mouse (by clicking over the VM display or pressing the host key) in order to use the mouse inside the guest OS.

acadgild@localhost:~ 
File Edit View Search Terminal Help
      commanval.foreach(println)
                                         ^
scala> val flatAvg_Stud = Avg_Stud.map(x => x._1 + "," + x._2)
flatAvg_Stud: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[28] at map at
<console>:25
scala> flatAvg_Stud.foreach(println)
Mark,50.75
Andrew,46.33333333333336
Mathew,50.5
John,47.5
Lisa,58.0
scala> val commanval = flatgradeAvg.intersection(flatAvg_Stud)
commanval: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[34] at intersection at <console>:27
scala> com
com  commanval  compat
scala> commanval.foreach(println)
scala> 
scala> val commanval = flatgradeAvg.intersection(flatAvg_Stud)
```

