A3Test

Cmd 1

```
1  val businessfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/business.csv")
2
3  val business = businessfile.map(a=>a.split('^')).map(a=>(a(0),
   (a(0),a(1),a(2))))
4
5  val reviewfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/review.csv")
6  val review = reviewfile.map(a=>a.split('^')).map(a=>(a(2),a(3).toDouble))
7  val reviewMedium = review.mapValues((_, 1))
8  val average1 = reviewMedium.reduceByKey((x,y)=>(x._1 + y._1, x._2 + y._2))
9  val average2 = average1.mapValues{case (sum, count) => (1.0 * sum) / count
   }
10
11 val afterjoin = average2.join(business)
12 val newmap = afterjoin.map(x=>(x._2._2, x._2._1))
13 val top10 = newmap.takeOrdered(10)
   (Ordering[Double].reverse.on(x=>x._2)).foreach(println)
```

▸ (1) Spark Jobs

```
((2LzZkqb5PLf0zzUVa4O9Gg,1821 N Fordham BlvdSte 1Chapel Hill, NC 27514,List(Hair
Salons, Beauty and Spas)),5.0)
((P9O4_6XLsft3-aBQtRoUrg,2422 Robinhood StWest UniversityHouston, TX 77005,List
(Veterinarians, Pets)),5.0)
((2LzZkqb5PLf0zzUVa4O9Gg,1821 N Fordham BlvdSte 1Chapel Hill, NC 27514,List(Hair
Salons, Beauty and Spas)),5.0)
((pSNEDaUGljLSHBZz5JNpuA,8935 Towne Centre DrSuite 105San Diego, CA 92122,List(A
ctive Life, Golf)),5.0)
((aVbVhFMQOyoIXsvEiOBghQ,4507 Brooklyn Ave NEUniversity DistrictSeattle, WA 9810
5,List(Event Planning & Services, Party & Event Planning, Caterers)),5.0)
((P9O4_6XLsft3-aBQtRoUrg,2422 Robinhood StWest UniversityHouston, TX 77005,List
(Veterinarians, Pets)),5.0)
((aVbVhFMQOyoIXsvEiOBghQ,4507 Brooklyn Ave NEUniversity DistrictSeattle, WA 9810
5,List(Event Planning & Services, Party & Event Planning, Caterers)),5.0)
((gqlCplBbmlMsGFZLNWd5wg,183 Angell StCollege HillProvidence, RI 02906,List(Fash
ion, Shopping)),5.0)
((iIw-ahkNV8c_xPCGesMfoA,304 W Weaver StSte 203Carrboro, NC 27510,List(Active Li
fe, Pilates, Fitness & Instruction)),5.0)
((pSNEDaUGljLSHBZz5JNpuA,8935 Towne Centre DrSuite 105San Diego, CA 92122,List(A
ctive Life, Golf)),5.0)
businessfile: org.apache.spark.rdd.RDD[String] = /FileStore/tables/eej38o6b15011
11865624/business.csv MapPartitionsRDD[33] at textFile at <console>:53
business: org.apache.spark.rdd.RDD[(String, (String, String, String))] = MapPart
```

```
itionsRDD[35] at map at <console>:55
reviewfile: org.apache.spark.rdd.RDD[String] = /FileStore/tables/eej38o6b1501111
865624/review.csv MapPartitionsRDD[37] at textFile at <console>:57
review: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[39] at map
at <console>:58
reviewMedium: org.apache.spark.rdd.RDD[(String, (Double, Int))] = MapPartitionsR
DD[40] at mapValues at <console>:59
average1: org.apache.spark.rdd.RDD[(String, (Double, Int))] = ShuffledRDD[41] at
reduceByKey at <console>:60
average2: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[42] at m
apValues at <console>:61
afterjoin: org.apache.spark.rdd.RDD[(String, (Double, (String, String, Strin
g)))] = MapPartitionsRDD[45] at join at <console>:63
newmap: org.apache.spark.rdd.RDD[((String, String, String), Double)] = MapPartit
ionsRDD[46] at map at <console>:64
top10: Unit = ()

Command took 1.56 seconds -- by louhy1128@gmail.com at 2017/7/26 下午6:36:12 on A3
```

Cmd 2

```
1  val businessfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/business.csv")
2  val business =
   businessfile.map(a=>a.split('^')).filter(a=>a(1).contains("Stanford")).map(
   a=>(a(0),""))
3
4  val reviewfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/review.csv")
5  val review = reviewfile.map(a=>a.split('^')).map(a=>(a(2),(a(1),a(3))))
6
7  val afterjoin = business.join(review)
8  val result = afterjoin.map(x=>(x._2))
9
10 result.collect().foreach(println)
```

▸ (1) Spark Jobs

```
(,(ln-8dqz9uu5TwjVg1vYESg,3.0))
(,(GEu83t4SjJ5S7SdoGkZIDg,3.0))
(,(S91kz40EtAo3AKUrDyEeDg,1.0))
(,(qy1HyF1nCKaMJCg-vvB1Xw,2.0))
(,(hqojrGeufv6qYpN00R-OsA,3.0))
(,(ln-8dqz9uu5TwjVg1vYESg,3.0))
(,(GEu83t4SjJ5S7SdoGkZIDg,3.0))
(,(S91kz40EtAo3AKUrDyEeDg,1.0))
(,(qy1HyF1nCKaMJCg-vvB1Xw,2.0))
(,(hqojrGeufv6qYpN00R-OsA,3.0))
(,(Kz1YTeNziSfVJBMKuA_aTA,5.0))
(,(_G96tUVqQrEw_g3SQUmKdw,5.0))
```

```
(,(Kz1YTeNziSfVJBMKuA_aTA,5.0))
(,(_G96tUVqQrEw_g3SQUmKdw,5.0))
(,(-Y71iV2dg5SdNoBf-hHHXA,4.0))
(,(8Y9qdkwqQcdc8EhTmT8Drg,5.0))
(,(68NVxMw8wwaqV8B__xIzeg,5.0))
(,(DZuxbyCn4LLtuOVUbGnJCg,5.0))
(,(ZKvsRhd91j-Hthga4NGLTQ,1.0))
(,(p4-tNsuQB6O11lp77KZKJg,2.0))
```

Command took 1.94 seconds -- by louhy1128@gmail.com at 2017/7/26 下午6:37:04 on A3

Cmd 3

```
1  val reviewfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/review.csv")
2
3  val review = reviewfile.map(a=>a.split('^')).map(a=>(a(1),a(1)))
4  val userfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/user.csv")
5  val user = userfile.map(a=>a.split('^')).map(a=>(a(0),a(1)))
6
7  val afterjoin = review.join(user)
8
9  val result = afterjoin.map(x=>(x._2, 1))
10 val topTen = result.reduceByKey(_+_).takeOrdered(10)
   (Ordering[Int].reverse.on(x=>x._2)).foreach(println)
11
```

▸ (1) Spark Jobs

```
((-iLH3Q2Wg4AMrNUXcgvliA,A T.),258)
((HUmClCl luKP5Ur6X7e306Q,John L.),240)
((3x8lZ-EoBhg-mw21BRITuQ,William J.),198)
((itXMelaTleEjLIFWCJtnwg,Christina G.),165)
((CQUDh80m48xnzUkx-X5NAw,David N.),165)
((U4KYlRjP3KmavdPbtfOWJQ,Lisa W.),162)
((wZPizeBxMAyOSl0M0zuCjg,Jess L.),142)
((tCqYnhAdQhPO3JAAnc09ig,Melissa M.),139)
((1kpMAKRZuAz3OzxBav3XTg,Ligaya T.),137)
((WKfoKNPk_-KbzVpCCY261g,Mae S.),136)
reviewfile: org.apache.spark.rdd.RDD[String] = /FileStore/tables/eej38o6b1501111
865624/review.csv MapPartitionsRDD[62] at textFile at <console>:47
review: org.apache.spark.rdd.RDD[(String, String)] = MapPartitionsRDD[64] at map
at <console>:49
userfile: org.apache.spark.rdd.RDD[String] = /FileStore/tables/eej38o6b150111186
5624/user.csv MapPartitionsRDD[66] at textFile at <console>:50
user: org.apache.spark.rdd.RDD[(String, String)] = MapPartitionsRDD[68] at map a
t <console>:51
afterjoin: org.apache.spark.rdd.RDD[(String, (String, String))] = MapPartitionsR
DD[71] at join at <console>:53
```

```
result: org.apache.spark.rdd.RDD[((String, String), Int)] = MapPartitionsRDD[72]
at map at <console>:55
topTen: Unit = ()
```

Command took 2.31 seconds -- by louhy1128@gmail.com at 2017/7/26 下午6:37:44 on A3

Cmd 4

```
1  val businessfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/business.csv")
2
3  val business =
   businessfile.map(a=>a.split('^')).filter(a=>a(1).contains("TX")).map(a=>
   (a(0),a(0)))
4
5  val reviewfile =
   sc.textFile("/FileStore/tables/eej38o6b1501111865624/review.csv")
6
7  val review = reviewfile.map(a=>a.split('^')).map(a=>(a(2),a(2)))
8
9  val afterjoin = review.join(business)
10 val result = afterjoin.map(x=>(x._1, 1))
11
12 val count = result.reduceByKey(_+_)
13
14 count.collect().foreach(println)
```

▸ (1) Spark Jobs

```
(r-MdoGkMYp3Kt7IE9MUf1A,768)
(yja5Oi3Gt6DbcuhZ-zvI8A,18)
(qnJ8FQPgJ1Jq-3G9l1fB3A,84)
(2bBSnkg6KQF4jq90IHWhHw,54)
(3NBPta-oH4MCNbR5k_wQjw,2)
(E36AGF--3QrQZbYtRrW1Mg,10)
(uHnV3vqHGH1H1kqO8jjiyw,10)
(Idg-Xgp_NTDZKzWTgdVDsw,8)
(vrCnZuoWHkJxTeSSFQiEHA,8)
(KK-6rXbIo9B4-b_P5WO5Qg,940)
(5vcgEM1R2qUWfR9VaNLVzA,18)
(dqux3WWVy2Gtyr3UwHL1wQ,142)
(Qt5qXt4coAHSM_EqyDyCSA,16)
(65x8x1k4lGg3WJbOTNiXFQ,4)
(ieytJEosKkDQnKb-KuTTXQ,30)
(hfYwn1tvJNyEH6Bdgc-Z5Q,6)
(j4sLo2IOkFxrB_EzvKIJXg,14)
(sJM4ogveiQAlHQOrzHVpQw,6)
(zq5jw3zVytbtwjgmdv_7-Q,4)
(B_8VlsVTtVvpv2dL0XvbMA,14)
(Cx5KQaTZWk9SUfjO5BjVCg,4)
(YJ8U4WYQZVj8uZ39yPoSlQ,12)
```

```
(IoL6l_mU7JB_uylOhwf1qw,256)
(BFcpJ6F86ygOJNPbW-QYmQ,28)
(1hN5HymVOc5khFJYhCppUA,4)
(juRDs3kMqnUdFXZTmluSCQ,6)
```

Command took 1.60 seconds -- by louhy1128@gmail.com at 2017/7/26 下午6:38:51 on A3

Cmd 5

```
1  val reviewfile = sc.textFile("/FileStore/tables/eej38o6b1501111865624/review.csv")
2  val review = reviewfile.map(a=>a.split('^')).map(a=>(a(1),a(3).toDouble))
3
4  val reviewMedium = review.mapValues((_, 1))
5  val average1 = reviewMedium.reduceByKey((x,y)=>(x._1 + y._1, x._2 + y._2))
6  val average2 = average1.mapValues{case (sum, count) => (1.0 * sum) / count }
7
8  val userfile = sc.textFile("/FileStore/tables/eej38o6b1501111865624/user.csv")
9
10 //val content = Console.readLine()
11
12 val user = userfile.map(a=>a.split('^')).filter(a=>a(1).contains("Matt J.")).map(a=>(a(0),a(1)))
13
14 val afterjoin = user.join(average2)
15 val newMap = afterjoin.map(x=>((x._2._1, x._1),x._2._2)).sortBy(_._2, false)
16
17 newMap.collect().foreach(println)
18
```

▸ (2) Spark Jobs

```
((Matt J.,og3gjB0zMlOLPqbmKmpR5w),5.0)
((Matt J.,Swc47jpKGsDn3lGoxDBwgg),5.0)
((Matt J. H.,IsmyZZtqsprEJqWh6MvoVQ),4.333333333333333)
((Matt J.,oB_SieEDwfbzsk5QaCUmbQ),4.0)
((Matt J.,aJOWk8y8YtBwFtjIYjnppA),4.0)
((Matt J.,BdL9uiLrPPp5KjRFB4T8Rg),4.0)
((Matt J.,5UeW6xMmWClWcOrw6cPXXg),3.4285714285714284)
((Matt J.,U0KwjsrX92pu2V99x05b6A),3.125)
((Matt J.,yzycOzJCh8VUnXSRcECcuA),2.5)
((Matt J.,IGRS8RTjlKtVoXZALyF-tA),1.0)
reviewfile: org.apache.spark.rdd.RDD[String] = /FileStore/tables/eej38o6b1501111865624/review.csv MapPartitionsRDD[90] at textFile at <console>:64
review: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[92] at map at <console>:65
reviewMedium: org.apache.spark.rdd.RDD[(String, (Double, Int))] = MapPartitionsR
```

```
DD[93] at mapValues at <console>:67
average1: org.apache.spark.rdd.RDD[(String, (Double, Int))] = ShuffledRDD[94] at
reduceByKey at <console>:68
average2: org.apache.spark.rdd.RDD[(String, Double)] = MapPartitionsRDD[95] at m
apValues at <console>:69
userfile: org.apache.spark.rdd.RDD[String] = /FileStore/tables/eej38o6b150111186
5624/user.csv MapPartitionsRDD[97] at textFile at <console>:71
user: org.apache.spark.rdd.RDD[(String, String)] = MapPartitionsRDD[100] at map
 at <console>:75
afterjoin: org.apache.spark.rdd.RDD[(String, (String, Double))] = MapPartitionsR
DD[103] at join at <console>:77
newMap: org.apache.spark.rdd.RDD[((String, String), Double)] = MapPartitionsRDD
[109] at sortBy at <console>:78

Command took 2.08 seconds -- by louhy1128@gmail.com at 2017/7/26 下午6:40:38 on A3
Cmd 6
```