

# Wiki - Correlated Pages

**Course:** 大数据软件工程

**Exercise type:** PySpark

**Soft Due:** 2015-05-27 18:00

**Hard Due:** 2015-05-27 18:00

**Note:** hard due has passed. You can no longer submit answer.

**Author:** 王欣明(TEACHER)

**Email:** [wangxm35@mail.sysu.edu.cn](mailto:wangxm35@mail.sysu.edu.cn)

你已经在这道题上花费大量时间，是否题目有让你困惑的地方？如果愿意的话，请在下面**Feedback**那里评价一下这道题。你的反馈将同时用邮件发送给作者。

## Description:

0 Questions and 0 Answers

Ask New Question

In this assignment, we will use Wikipedia traffic statistics data obtained from <http://aws.amazon.com/datasets/4182> . To make the analysis feasible (within the short timeframe of the exercise), we only took a small sample.

The data files contain traffic statistics for all pages in a specific hour. In total there are nine input files, each with around 200MB data:

- smart\_programmer/wiki/part-00097
- smart\_programmer/wiki/part-00098
- smart\_programmer/wiki/part-00099
- smart\_programmer/wiki/part-00100
- smart\_programmer/wiki/part-00101
- smart\_programmer/wiki/part-00102
- smart\_programmer/wiki/part-00103
- smart\_programmer/wiki/part-00104
- smart\_programmer/wiki/part-00105

The content of each file is like the follows:

20090507-040000 aa ?page=http://www.stockphotosharing.com/Themes/Images/users\_raw/id.txt 3 39267

20090507-040000 aa Main\_Page 7 51309

20090507-040000 aa Special:Boardvote 1 11631

20090507-040000 aa Special:Imagelist 1 931

Each line, delimited by a space, contains stats for one page. The schema is:

The <date\_time> field specifies a date in the YYYYMMDD format (year, month, day) followed by a hyphen and then the hour in the HHmmSS format (hour, minute, second). There is no information in mmSS. The <project\_code> field contains information about the language of the pages. For example, project code “en” indicates an English page. The <page\_title> field gives the title of the Wikipedia page. The <num\_hits> field gives the number of page views in the hour-long time slot starting at <data\_time>. The <page\_size> field gives the size in bytes of the Wikipedia page.

Write a spark program to find the 20 pairs of pages that have the closest visit count (i.e. the absolute values of the difference between their visit count are the 20 smallest).

Note: ignoring pages with visit count lower than 10000

The output shall be like:

```
aa XXX 100000 aa AAA 100000
aa YYY 1000 aa BBB 23123
en ZZZ 1234 aa CCC 53112
```

....

Each line contains a pair of page. For each pair of page, the page with fewer visit count is listed first. If the visit count of the two pages are the same, then the page with larger title (compared by string) is shown first.

Note: items in each line is separated by "\t".

The spark program shall consist of two python code file. One is "driver.py", which is given to you by the smart\_programmer system. You can read its code in the assignment page. "driver.py" will create the spark context and load the input data files as initial RDDs, then pass these input RDDs to the function "main" that shall be written by you in "main.py"

To output the result, simply use "print" to print the result to stdout in the main function.

## Hint:

Hint is not available for this exercise.

## Your answer:

Already pass (hard) due date: May 27, 2015, 6 p.m. You can no longer submit your code to smart\_programmer by yourself.

We don't accept late submission in principle. But if you really have a convincing reason, send the answer code to your TA by email and ask her/him to submit it for you. Don't copy the standard answer. It will be considered as plagiarism.

If you want to study the standard answer, you can add it as an optional assignment.

Add this exercise as an optional assignment

## Hard due has passed. The standard answer is unlocked:

driver.py

[\*]main.py

```
1. import
2. from pyspark import SparkContext
3. from      import
4.
5.      = [
6.      "smart_programmer/wiki/part-00097",
7.      "smart_programmer/wiki/part-00098",
8.      "smart_programmer/wiki/part-00099",
9.      "smart_programmer/wiki/part-00100",
10.     "smart_programmer/wiki/part-00101",
11.     "smart_programmer/wiki/part-00102",
12.     "smart_programmer/wiki/part-00103",
13.     "smart_programmer/wiki/part-00104",
14.     "smart_programmer/wiki/part-00105"
15. ]
16.
17. if      == "__main__":
18.     sc = SparkContext(appName="wiki-top10")
19.     RDDs = [ ]
20.     for input file name in input file names:
21.         = .      ("hdfs://eden/user/driver/"+      )
22.         RDDs.append(rdd)
23.         (RDDs)
24.
```

## Discussion:

Discussion is not available for this exercise.

Latest Submission Grade: 10 (submitted on 2015-05-25 17:04)

=====[Phase1: check\_plagiarism\_cloudera]=====  
Pass.

=====[Phase2: cloudera\_prepare]=====  
Pass syntax checking. You got 10 points.

=====[Phase3: cloudera\_pyspark\_cluster]=====  
The spark execution fails. You get 0 out of totally 90 points.  
Failure: timeout (exceeds hard time limit 150 seconds)

Output content:

Detailed log:

```
15/05/25 17:16:35 INFO SparkContext: Running Spark version 1.3.0
15/05/25 17:16:36 INFO SecurityManager: Changing view acls to: driver
15/05/25 17:16:36 INFO SecurityManager: Changing modify acls to: driver
15/05/25 17:16:36 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users
15/05/25 17:16:36 INFO Slf4jLogger: Slf4jLogger started
15/05/25 17:16:36 INFO Remoting: Starting remoting
15/05/25 17:16:36 INFO Remoting: Remoting started; listening on addresses :[akka.tcp://sparkDriver@eden:42
15/05/25 17:16:36 INFO Remoting: Remoting now listens on addresses: [akka.tcp://sparkDriver@eden:42719]
15/05/25 17:16:36 INFO Utils: Successfully started service 'sparkDriver' on port 42719.
15/05/25 17:16:36 INFO SparkEnv: Registering MapOutputTracker
15/05/25 17:16:36 INFO SparkEnv: Registering BlockManagerMaster
15/05/25 17:16:36 INFO DiskBlockManager: Created local directory at /tmp/spark-0331b338-550d-491c-bd73-35e
15/05/25 17:16:36 INFO MemoryStore: MemoryStore started with capacity 265.1 MB
15/05/25 17:16:36 INFO HttpFileServer: HTTP File server directory is /tmp/spark-2d6aedd1-a32e-4c6e-91c3-5f
15/05/25 17:16:36 INFO HttpServer: Starting HTTP Server
15/05/25 17:16:36 INFO Server: jetty-8.y.z-SNAPSHOT
15/05/25 17:16:36 INFO AbstractConnector: Started SocketConnector@0.0.0.0:58368
15/05/25 17:16:36 INFO Utils: Successfully started service 'HTTP file server' on port 58368.
15/05/25 17:16:36 INFO SparkEnv: Registering OutputCommitCoordinator
15/05/25 17:16:37 INFO Server: jetty-8.y.z-SNAPSHOT
15/05/25 17:16:37 INFO AbstractConnector: Started SelectChannelConnector@0.0.0.0:4040
15/05/25 17:16:37 INFO Utils: Successfully started service 'SparkUI' on port 4040.
15/05/25 17:16:37 INFO SparkUI: Started SparkUI at http://eden:4040
15/05/25 17:16:37 INFO Utils: Copying /tmp/smart_programmer/assignment-4159/submission-1170884/./driver.py
15/05/25 17:16:37 INFO SparkContext: Added file file:/tmp/smart_programmer/assignment-4159/submission-1170
15/05/25 17:16:37 INFO Executor: Starting executor ID <driver> on host localhost
15/05/25 17:16:37 INFO AkkaUtils: Connecting to HeartbeatReceiver: akka.tcp://sparkDriver@eden:42719/user/
15/05/25 17:16:37 INFO NettyBlockTransferService: Server created on 55059
15/05/25 17:16:37 INFO BlockManagerMaster: Trying to register BlockManager
15/05/25 17:16:37 INFO BlockManagerMasterActor: Registering block manager localhost:55059 with 265.1 MB RA
15/05/25 17:16:37 INFO BlockManagerMaster: Registered BlockManager
15/05/25 17:16:38 INFO EventLoggingListener: Logging events to hdfs://eden:8020/user/spark/applicationHist
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(270507) called with curMem=0, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_0 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=270507, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 21.
15/05/25 17:16:38 INFO BlockManagerInfo: Added broadcast_0_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:38 INFO BlockManagerMaster: Updated info of block broadcast_0_piece0
15/05/25 17:16:38 INFO SparkContext: Created broadcast 0 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=292593, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_1 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=563148, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 21.
15/05/25 17:16:38 INFO BlockManagerInfo: Added broadcast_1_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:38 INFO BlockManagerMaster: Updated info of block broadcast_1_piece0
15/05/25 17:16:38 INFO SparkContext: Created broadcast 1 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=585234, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_2 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=855789, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 21.
15/05/25 17:16:38 INFO BlockManagerInfo: Added broadcast_2_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:38 INFO BlockManagerMaster: Updated info of block broadcast_2_piece0
15/05/25 17:16:38 INFO SparkContext: Created broadcast 2 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=877875, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_3 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=1148430, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 21.
15/05/25 17:16:38 INFO BlockManagerInfo: Added broadcast_3_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:38 INFO BlockManagerMaster: Updated info of block broadcast_3_piece0
15/05/25 17:16:38 INFO SparkContext: Created broadcast 3 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=1170516, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_4 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=1441071, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_4_piece0 stored as bytes in memory (estimated size 21.
15/05/25 17:16:38 INFO BlockManagerInfo: Added broadcast_4_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:38 INFO BlockManagerMaster: Updated info of block broadcast_4_piece0
15/05/25 17:16:38 INFO SparkContext: Created broadcast 4 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=1463157, maxMem=278019440
```



```
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_5 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:38 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=1733712, maxMem=278019440
15/05/25 17:16:38 INFO MemoryStore: Block broadcast_5_piece0 stored as bytes in memory (estimated size 21.6
15/05/25 17:16:38 INFO BlockManagerInfo: Added broadcast_5_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:38 INFO BlockManagerMaster: Updated info of block broadcast_5_piece0
15/05/25 17:16:38 INFO SparkContext: Created broadcast 5 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=1755798, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_6 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=2026353, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_6_piece0 stored as bytes in memory (estimated size 21.6
15/05/25 17:16:39 INFO BlockManagerInfo: Added broadcast_6_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:39 INFO BlockManagerMaster: Updated info of block broadcast_6_piece0
15/05/25 17:16:39 INFO SparkContext: Created broadcast 6 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=2048439, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_7 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=2318994, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 21.6
15/05/25 17:16:39 INFO BlockManagerInfo: Added broadcast_7_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:39 INFO BlockManagerMaster: Updated info of block broadcast_7_piece0
15/05/25 17:16:39 INFO SparkContext: Created broadcast 7 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(270555) called with curMem=2341080, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_8 stored as values in memory (estimated size 264.2 KB,
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(22086) called with curMem=2611635, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_8_piece0 stored as bytes in memory (estimated size 21.6
15/05/25 17:16:39 INFO BlockManagerInfo: Added broadcast_8_piece0 in memory on localhost:55059 (size: 21.6
15/05/25 17:16:39 INFO BlockManagerMaster: Updated info of block broadcast_8_piece0
15/05/25 17:16:39 INFO SparkContext: Created broadcast 8 from textFile at NativeMethodAccessorImpl.java:-2
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.232:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.236:50010
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.232:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.232:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.232:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.236:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.232:50010
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.232:50010
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.232:50010
15/05/25 17:16:39 INFO FileInputFormat: Total input paths to process : 1
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.236:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.234:50010
15/05/25 17:16:39 INFO NetworkTopology: Adding a new node: /default/172.16.21.233:50010
15/05/25 17:16:39 INFO SparkContext: Starting job: sortBy at /tmp/smart_programmer/assignment-4159/submiss
15/05/25 17:16:39 INFO DAGScheduler: Registering RDD 28 (reduceByKey at /tmp/smart_programmer/assignment-4
15/05/25 17:16:39 INFO DAGScheduler: Got job 0 (sortBy at /tmp/smart_programmer/assignment-4159/submissior
15/05/25 17:16:39 INFO DAGScheduler: Final stage: Stage 1(sortBy at /tmp/smart_programmer/assignment-4159/
15/05/25 17:16:39 INFO DAGScheduler: Parents of final stage: List(Stage 0)
15/05/25 17:16:39 INFO DAGScheduler: Missing parents: List(Stage 0)
15/05/25 17:16:39 INFO DAGScheduler: Submitting Stage 0 (PairwiseRDD[28] at reduceByKey at /tmp/smart_prog
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(11832) called with curMem=2633721, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_9 stored as values in memory (estimated size 11.6 KB,
15/05/25 17:16:39 INFO MemoryStore: ensureFreeSpace(6123) called with curMem=2645553, maxMem=278019440
15/05/25 17:16:39 INFO MemoryStore: Block broadcast_9_piece0 stored as bytes in memory (estimated size 6.0
15/05/25 17:16:39 INFO BlockManagerInfo: Added broadcast_9_piece0 in memory on localhost:55059 (size: 6.0
15/05/25 17:16:39 INFO BlockManagerMaster: Updated info of block broadcast_9_piece0
15/05/25 17:16:39 INFO SparkContext: Created broadcast 9 from broadcast at DAGScheduler.scala:839
15/05/25 17:16:39 INFO DAGScheduler: Submitting 18 missing tasks from Stage 0 (PairwiseRDD[28] at reduceBy
15/05/25 17:16:39 INFO TaskSchedulerImpl: Adding task set 0.0 with 18 tasks
15/05/25 17:16:39 INFO TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, ANY, 1606 bytes)
15/05/25 17:16:39 INFO TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, localhost, ANY, 1606 bytes)
```



```
15/05/25 17:16:39 INFO TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, localhost, ANY, 1606 bytes)
15/05/25 17:16:39 INFO TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, localhost, ANY, 1606 bytes)
15/05/25 17:16:39 INFO TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, localhost, ANY, 1591 bytes)
15/05/25 17:16:39 INFO TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, localhost, ANY, 1591 bytes)
15/05/25 17:16:39 INFO TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, localhost, ANY, 1576 bytes)
15/05/25 17:16:39 INFO TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, localhost, ANY, 1576 bytes)
15/05/25 17:16:39 INFO Executor: Running task 0.0 in stage 0.0 (TID 0)
15/05/25 17:16:39 INFO Executor: Running task 1.0 in stage 0.0 (TID 1)
15/05/25 17:16:39 INFO Executor: Running task 2.0 in stage 0.0 (TID 2)
15/05/25 17:16:39 INFO Executor: Running task 4.0 in stage 0.0 (TID 4)
15/05/25 17:16:39 INFO Executor: Running task 5.0 in stage 0.0 (TID 5)
15/05/25 17:16:39 INFO Executor: Running task 3.0 in stage 0.0 (TID 3)
15/05/25 17:16:39 INFO Executor: Running task 6.0 in stage 0.0 (TID 6)
15/05/25 17:16:39 INFO Executor: Running task 7.0 in stage 0.0 (TID 7)
15/05/25 17:16:39 INFO Executor: Fetching file:/tmp/smart_programmer/assignment-4159/submission-1170884/./
15/05/25 17:16:39 INFO Utils: /tmp/smart_programmer/assignment-4159/submission-1170884/./driver.py has bee
15/05/25 17:16:39 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00099:0+
15/05/25 17:16:39 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00100:0+
15/05/25 17:16:39 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00099:11
15/05/25 17:16:39 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00097:0+
15/05/25 17:16:39 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00097:11
15/05/25 17:16:39 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00098:0+
15/05/25 17:16:40 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00098:11
15/05/25 17:16:40 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00100:11
15/05/25 17:16:40 INFO deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id
15/05/25 17:16:40 INFO deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id
15/05/25 17:16:40 INFO deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap
15/05/25 17:16:40 INFO deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.parti
15/05/25 17:16:40 INFO deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id
15/05/25 17:17:14 INFO PythonRDD: Times: total = 35187, boot = 336, init = 158, finish = 34693
15/05/25 17:17:14 INFO Executor: Finished task 1.0 in stage 0.0 (TID 1). 2082 bytes result sent to driver
15/05/25 17:17:14 INFO TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, localhost, ANY, 1561 bytes)
15/05/25 17:17:14 INFO Executor: Running task 8.0 in stage 0.0 (TID 8)
15/05/25 17:17:14 INFO TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 35422 ms on localhost (1/
15/05/25 17:17:15 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00101:0+
15/05/25 17:17:42 INFO PythonRDD: Times: total = 27796, boot = -122, init = 155, finish = 27763
15/05/25 17:17:42 INFO Executor: Finished task 8.0 in stage 0.0 (TID 8). 2137 bytes result sent to driver
15/05/25 17:17:42 INFO TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, localhost, ANY, 1561 bytes)
15/05/25 17:17:42 INFO Executor: Running task 9.0 in stage 0.0 (TID 9)
15/05/25 17:17:42 INFO TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 27926 ms on localhost (2/
15/05/25 17:17:42 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00101:10
15/05/25 17:17:47 INFO PythonRDD: Times: total = 67580, boot = 325, init = 118, finish = 67137
15/05/25 17:17:47 INFO Executor: Finished task 4.0 in stage 0.0 (TID 4). 2137 bytes result sent to driver
15/05/25 17:17:47 INFO TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, localhost, ANY, 1546 bytes)
15/05/25 17:17:47 INFO Executor: Running task 10.0 in stage 0.0 (TID 10)
15/05/25 17:17:47 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00102:0+
15/05/25 17:17:47 INFO TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 67811 ms on localhost (3/
15/05/25 17:17:48 INFO PythonRDD: Times: total = 69295, boot = 332, init = 127, finish = 68836
15/05/25 17:17:49 INFO Executor: Finished task 2.0 in stage 0.0 (TID 2). 2137 bytes result sent to driver
15/05/25 17:17:49 INFO TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, localhost, ANY, 1546 bytes)
15/05/25 17:17:49 INFO Executor: Running task 11.0 in stage 0.0 (TID 11)
15/05/25 17:17:49 INFO TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 69511 ms on localhost (4/
15/05/25 17:17:49 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00102:11
15/05/25 17:17:52 INFO PythonRDD: Times: total = 73333, boot = 330, init = 166, finish = 72837
15/05/25 17:17:53 INFO Executor: Finished task 6.0 in stage 0.0 (TID 6). 2137 bytes result sent to driver
15/05/25 17:17:53 INFO TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, localhost, ANY, 1531 bytes)
15/05/25 17:17:53 INFO TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 73540 ms on localhost (5/
15/05/25 17:17:53 INFO Executor: Running task 12.0 in stage 0.0 (TID 12)
15/05/25 17:17:53 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00103:0+
15/05/25 17:17:54 INFO PythonRDD: Times: total = 74412, boot = 335, init = 150, finish = 73927
15/05/25 17:17:54 INFO Executor: Finished task 0.0 in stage 0.0 (TID 0). 2137 bytes result sent to driver
15/05/25 17:17:54 INFO TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, localhost, ANY, 1531 bytes)
15/05/25 17:17:54 INFO Executor: Running task 13.0 in stage 0.0 (TID 13)
15/05/25 17:17:54 INFO TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 74672 ms on localhost (6/
15/05/25 17:17:54 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00103:11
15/05/25 17:17:54 INFO PythonRDD: Times: total = 74930, boot = 346, init = 153, finish = 74431
15/05/25 17:17:54 INFO Executor: Finished task 7.0 in stage 0.0 (TID 7). 2137 bytes result sent to driver
15/05/25 17:17:54 INFO TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, localhost, ANY, 1516 bytes)
15/05/25 17:17:54 INFO Executor: Running task 14.0 in stage 0.0 (TID 14)
15/05/25 17:17:54 INFO TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 75163 ms on localhost (7/
15/05/25 17:17:54 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00104:0+
15/05/25 17:17:57 INFO PythonRDD: Times: total = 78165, boot = 321, init = 398, finish = 77446
15/05/25 17:17:57 INFO Executor: Finished task 5.0 in stage 0.0 (TID 5). 2137 bytes result sent to driver
15/05/25 17:17:57 INFO TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, localhost, ANY, 1516 bytes)
15/05/25 17:17:57 INFO TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 78350 ms on localhost (8/
15/05/25 17:17:57 INFO Executor: Running task 15.0 in stage 0.0 (TID 15)
15/05/25 17:17:57 INFO PythonRDD: Times: total = 78269, boot = 339, init = 130, finish = 77800
15/05/25 17:17:57 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00104:12
15/05/25 17:17:58 INFO Executor: Finished task 3.0 in stage 0.0 (TID 3). 2137 bytes result sent to driver
15/05/25 17:17:58 INFO TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, localhost, ANY, 1501 bytes)
15/05/25 17:17:58 INFO Executor: Running task 16.0 in stage 0.0 (TID 16)
15/05/25 17:17:58 INFO TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 78459 ms on localhost (9/
15/05/25 17:17:58 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00105:0+
15/05/25 17:18:23 INFO PythonRDD: Times: total = 29113, boot = -87, init = 162, finish = 29038
```



```
15/05/25 17:18:23 INFO TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13). 2137 bytes result sent to driver
15/05/25 17:18:23 INFO TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, localhost, ANY, 1501 bytes)
15/05/25 17:18:23 INFO Executor: Running task 17.0 in stage 0.0 (TID 17)
15/05/25 17:18:23 INFO TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 29219 ms on localhost (17/05/25 17:18:23 INFO HadoopRDD: Input split: hdfs://eden/user/driver/smart_programmer/wiki/part-00105:13
15/05/25 17:18:31 INFO PythonRDD: Times: total = 48333, boot = -122, init = 158, finish = 48297
15/05/25 17:18:31 INFO Executor: Finished task 9.0 in stage 0.0 (TID 9). 2137 bytes result sent to driver
15/05/25 17:18:31 INFO TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 48440 ms on localhost (11/05/25 17:18:33 INFO PythonRDD: Times: total = 35220, boot = -75, init = 83, finish = 35212
15/05/25 17:18:33 INFO Executor: Finished task 16.0 in stage 0.0 (TID 16). 2137 bytes result sent to driver
15/05/25 17:18:33 INFO TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 35369 ms on localhost (15/05/25 17:18:39 INFO PythonRDD: Times: total = 50611, boot = -98, init = 211, finish = 50498
15/05/25 17:18:39 INFO Executor: Finished task 11.0 in stage 0.0 (TID 11). 2137 bytes result sent to driver
15/05/25 17:18:39 INFO TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 50721 ms on localhost (15/05/25 17:18:49 INFO PythonRDD: Times: total = 61726, boot = -112, init = 202, finish = 61636
15/05/25 17:18:49 INFO Executor: Finished task 10.0 in stage 0.0 (TID 10). 2137 bytes result sent to driver
15/05/25 17:18:49 INFO TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 61838 ms on localhost (15/05/25 17:18:49 INFO PythonRDD: Times: total = 26350, boot = -91, init = 112, finish = 26329
15/05/25 17:18:49 INFO Executor: Finished task 17.0 in stage 0.0 (TID 17). 2137 bytes result sent to driver
15/05/25 17:18:49 INFO TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 26475 ms on localhost (15/05/25 17:18:51 INFO PythonRDD: Times: total = 58605, boot = -109, init = 203, finish = 58511
15/05/25 17:18:51 INFO Executor: Finished task 12.0 in stage 0.0 (TID 12). 2137 bytes result sent to driver
15/05/25 17:18:51 INFO TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 58725 ms on localhost (15/05/25 17:18:53 INFO PythonRDD: Times: total = 58637, boot = -75, init = 143, finish = 58569
15/05/25 17:18:53 INFO Executor: Finished task 14.0 in stage 0.0 (TID 14). 2137 bytes result sent to driver
15/05/25 17:18:53 INFO TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 58757 ms on localhost (15/05/25 17:18:55 INFO PythonRDD: Times: total = 57934, boot = -73, init = 158, finish = 57849
15/05/25 17:18:55 INFO Executor: Finished task 15.0 in stage 0.0 (TID 15). 2137 bytes result sent to driver
15/05/25 17:18:55 INFO TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 58046 ms on localhost (15/05/25 17:18:55 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
15/05/25 17:18:55 INFO DAGScheduler: Stage 0 (reduceByKey at /tmp/smart_programmer/assignment-4159/submiss
15/05/25 17:18:55 INFO DAGScheduler: looking for newly runnable stages
15/05/25 17:18:55 INFO DAGScheduler: running: Set()
15/05/25 17:18:55 INFO DAGScheduler: waiting: Set(Stage 1)
15/05/25 17:18:55 INFO DAGScheduler: failed: Set()
15/05/25 17:18:55 INFO DAGScheduler: Missing parents for Stage 1: List()
15/05/25 17:18:55 INFO DAGScheduler: Submitting Stage 1 (PythonRDD[32] at sortBy at /tmp/smart_programmer/
15/05/25 17:18:55 INFO MemoryStore: ensureFreeSpace(7288) called with curMem=2651676, maxMem=278019440
15/05/25 17:18:55 INFO MemoryStore: Block broadcast_10 stored as values in memory (estimated size 7.1 KB,
15/05/25 17:18:55 INFO MemoryStore: ensureFreeSpace(4365) called with curMem=2658964, maxMem=278019440
15/05/25 17:18:55 INFO MemoryStore: Block broadcast_10_piece0 stored as bytes in memory (estimated size 4.
15/05/25 17:18:55 INFO BlockManagerInfo: Added broadcast_10_piece0 in memory on localhost:55059 (size: 4.3
15/05/25 17:18:55 INFO BlockManagerMaster: Updated info of block broadcast_10_piece0
15/05/25 17:18:55 INFO SparkContext: Created broadcast 10 from broadcast at DAGScheduler.scala:839
15/05/25 17:18:55 INFO DAGScheduler: Submitting 18 missing tasks from Stage 1 (PythonRDD[32] at sortBy at
15/05/25 17:18:55 INFO TaskSchedulerImpl: Adding task set 1.0 with 18 tasks
15/05/25 17:18:55 INFO TaskSetManager: Starting task 0.0 in stage 1.0 (TID 18, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO TaskSetManager: Starting task 1.0 in stage 1.0 (TID 19, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO TaskSetManager: Starting task 2.0 in stage 1.0 (TID 20, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO TaskSetManager: Starting task 3.0 in stage 1.0 (TID 21, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO TaskSetManager: Starting task 4.0 in stage 1.0 (TID 22, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO TaskSetManager: Starting task 5.0 in stage 1.0 (TID 23, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO TaskSetManager: Starting task 6.0 in stage 1.0 (TID 24, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO TaskSetManager: Starting task 7.0 in stage 1.0 (TID 25, localhost, PROCESS_LOCAL, 1
15/05/25 17:18:55 INFO Executor: Running task 1.0 in stage 1.0 (TID 19)
15/05/25 17:18:55 INFO Executor: Running task 2.0 in stage 1.0 (TID 20)
15/05/25 17:18:55 INFO Executor: Running task 6.0 in stage 1.0 (TID 24)
15/05/25 17:18:55 INFO Executor: Running task 5.0 in stage 1.0 (TID 23)
15/05/25 17:18:55 INFO Executor: Running task 7.0 in stage 1.0 (TID 25)
15/05/25 17:18:55 INFO Executor: Running task 4.0 in stage 1.0 (TID 22)
15/05/25 17:18:55 INFO Executor: Running task 0.0 in stage 1.0 (TID 18)
15/05/25 17:18:55 INFO Executor: Running task 3.0 in stage 1.0 (TID 21)
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:18:56 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms
15/05/25 17:19:01 INFO PythonRDD: Times: total = 5107, boot = -6880, init = 6913, finish = 5074
15/05/25 17:19:01 INFO Executor: Finished task 6.0 in stage 1.0 (TID 24). 964 bytes result sent to driver
15/05/25 17:19:01 INFO TaskSetManager: Starting task 8.0 in stage 1.0 (TID 26, localhost, PROCESS_LOCAL, 1
15/05/25 17:19:01 INFO Executor: Running task 8.0 in stage 1.0 (TID 26)
15/05/25 17:19:01 INFO TaskSetManager: Finished task 6.0 in stage 1.0 (TID 24) in 5114 ms on localhost (17/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
```



```
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
15/05/25 17:19:01 INFO PythonRDD: Times: total = 5421, boot = -24733, init = 24765, finish = 5389
15/05/25 17:19:01 INFO Executor: Finished task 7.0 in stage 1.0 (TID 25). 964 bytes result sent to driver
15/05/25 17:19:01 INFO PythonRDD: Times: total = 5423, boot = -16292, init = 16337, finish = 5378
15/05/25 17:19:01 INFO TaskSetManager: Starting task 9.0 in stage 1.0 (TID 27, localhost, PROCESS_LOCAL, 1
15/05/25 17:19:01 INFO Executor: Finished task 2.0 in stage 1.0 (TID 20). 964 bytes result sent to driver
15/05/25 17:19:01 INFO Executor: Running task 9.0 in stage 1.0 (TID 27)
15/05/25 17:19:01 INFO TaskSetManager: Starting task 10.0 in stage 1.0 (TID 28, localhost, PROCESS_LOCAL,
15/05/25 17:19:01 INFO TaskSetManager: Finished task 2.0 in stage 1.0 (TID 20) in 5429 ms on localhost (2/
15/05/25 17:19:01 INFO TaskSetManager: Finished task 7.0 in stage 1.0 (TID 25) in 5428 ms on localhost (3/
15/05/25 17:19:01 INFO Executor: Running task 10.0 in stage 1.0 (TID 28)
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
15/05/25 17:19:01 INFO PythonRDD: Times: total = 5512, boot = -4255, init = 4291, finish = 5476
15/05/25 17:19:01 INFO Executor: Finished task 4.0 in stage 1.0 (TID 22). 964 bytes result sent to driver
15/05/25 17:19:01 INFO TaskSetManager: Starting task 11.0 in stage 1.0 (TID 29, localhost, PROCESS_LOCAL,
15/05/25 17:19:01 INFO Executor: Running task 11.0 in stage 1.0 (TID 29)
15/05/25 17:19:01 INFO TaskSetManager: Finished task 4.0 in stage 1.0 (TID 22) in 5520 ms on localhost (4/
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
15/05/25 17:19:01 INFO PythonRDD: Times: total = 5594, boot = -127, init = 155, finish = 5566
15/05/25 17:19:01 INFO Executor: Finished task 3.0 in stage 1.0 (TID 21). 964 bytes result sent to driver
15/05/25 17:19:01 INFO TaskSetManager: Starting task 12.0 in stage 1.0 (TID 30, localhost, PROCESS_LOCAL,
15/05/25 17:19:01 INFO Executor: Running task 12.0 in stage 1.0 (TID 30)
15/05/25 17:19:01 INFO TaskSetManager: Finished task 3.0 in stage 1.0 (TID 21) in 5603 ms on localhost (5/
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Getting 18 non-empty blocks out of 18 blocks
15/05/25 17:19:01 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
15/05/25 17:19:02 INFO PythonRDD: Times: total = 6718, boot = -22731, init = 22777, finish = 6672
15/05/25 17:19:02 INFO Executor: Finished task 1.0 in stage 1.0 (TID 19). 964 bytes result sent to driver
```

Your submission record:

Submission Date	Grade	Diff
<a href="#">2015-05-25 17:04</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-25 16:48</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-21 23:10</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-21 21:51</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-21 00:39</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-20 23:57</a>	44	<a href="#">Diff</a>
<a href="#">2015-05-20 22:16</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-20 21:45</a>	24	<a href="#">Diff</a>
<a href="#">2015-05-20 21:28</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-19 18:13</a>	0	<a href="#">Diff</a>
<a href="#">2015-05-19 16:29</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-19 16:09</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-19 16:05</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-19 15:55</a>	10	<a href="#">Diff</a>
<a href="#">2015-05-19 15:34</a>	10	<a href="#">Diff</a>

Top performers in the class:

Name	Points	Total Trials	Last Submission
郑穗展	100	9	2015-05-21 23:57
贴膜BOY	100	21	2015-05-22 14:22
把我飘准的普通发凡我	100	4	2015-05-22 23:49
无法无天	100	4	2015-05-24 21:47
怎么会这样...	100	11	2015-05-25 00:51
Duang	100	14	2015-05-25 12:37
张炜杰	100	9	2015-05-25 14:02





	<div>Output content:</div> <div>Traceback (most recent call last):</div> <div>File "/tmp/smart_programmer/assignment-4159/submission-164114/./driver.py", line 18, in &lt;module&gt;</div> <div>sc = SparkContext(appName="wiki-top10")</div> <div>File "/opt/cloudera/parcels/CDH-5.4.0-1.cdh5.4.0.p0.27/lib/spark/python/pyspark/context.py", line 111, in __init__</div> <div>conf, jsc, profiler_cls)</div> <div>File "/opt/cloudera/parcels/CDH-5.4.0-1.cdh5.4.0.p0.27/lib/spark/python/pyspark/context.py", line 159, in _do_init</div> <div>self._jsc = jsc or self._initialize_context(self._conf._jconf)</div>
BDSE_2015	<div>才执行了union，map，reduceBeKey三个操作就超时了，后面还能做啥，服务器又崩了？？？？？</div> <div>求老师解救</div>
BDSE_2015	<div>Nearly impossible to finish.</div> <div>老师还是给点思路吧。。。。 ORZ</div>
BDSE_2015	<div>为何只有10分，，，，，，，</div>
BDSE_2015	<div>标准答案也没有按照规则啊</div>

Please contact **Dr. Wang (roadlit at gmail.com)** to report bugs.

Thanks to:

- 吴浩坚同学 and 梁展瑞同学
- [Django](#), [Gunicorn](#), [TinyMCE](#), [Sandbox](#), [Nginx](#)
- [Valgrind](#), [Google Gode Style](#) [SOClone](#)

### Recent messages:

- [06/01 14:50] From 王欣明: 最近一段时间经常发生有些同学的作业卡队列的现象。如果发现请及时通知我重启服务队列。
- [06/01 14:49] From 王欣明: 本周是大数据软件工程的期末课程项目，题目会在周二晚上6点开放，同学们可以自己在宿舍做，一周时间内完成。
- [05/31 00:38] From 谢议尊: test

Send message to an user with real name or nick name:

Name:

Message:

[Send Message](#)