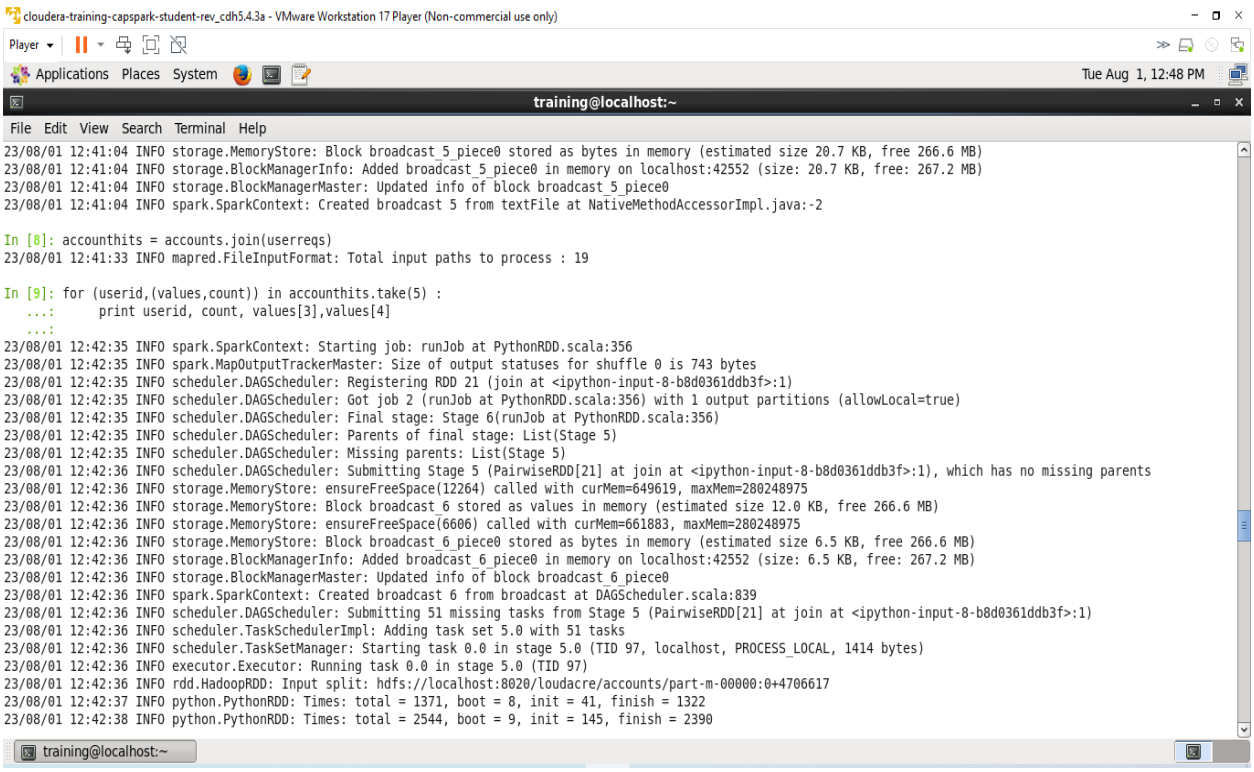
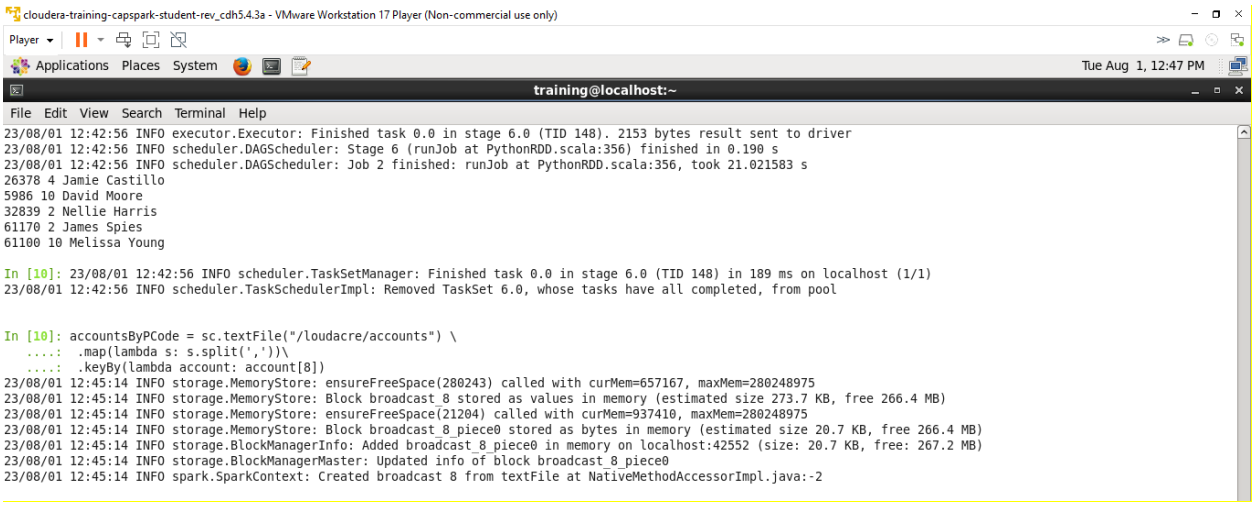
# LAB 12 BIG DATA

**Submitted by:**
**ABHISHEK ARYA**
**NET ID: axa220149**




cloudera-training-capspark-student-rev_cdh5.4.3a - VMware Workstation 17 Player (Non-commercial use only)
Player
Applications Places System
Tue Aug 1, 12:47 PM
training@localhost:~
File Edit View Search Terminal Help
23/08/01 12:42:56 INFO executor.Executor: Finished task 0.0 in stage 6.0 (TID 148). 2153 bytes result sent to driver
23/08/01 12:42:56 INFO scheduler.DAGScheduler: Stage 6 (runJob at PythonRDD.scala:356) finished in 0.190 s
23/08/01 12:42:56 INFO scheduler.DAGScheduler: Job 2 finished: runJob at PythonRDD.scala:356, took 21.021583 s
26378 4 Jamie Castillo
5986 10 David Moore
32839 2 Nellie Harris
61170 2 James Spies
61100 10 Melissa Young
In [10]: 23/08/01 12:42:56 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 6.0 (TID 148) in 189 ms on localhost (1/1)
23/08/01 12:42:56 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool
In [10]: accountsByPCode = sc.textFile("/loudacre/accounts") \
....: .map(lambda s: s.split(','))\
....: .keyBy(lambda account: account[8])
23/08/01 12:45:14 INFO storage.MemoryStore: ensureFreeSpace(280243) called with curMem=657167, maxMem=280248975
23/08/01 12:45:14 INFO storage.MemoryStore: Block broadcast_8 stored as values in memory (estimated size 273.7 KB, free 266.4 MB)
23/08/01 12:45:14 INFO storage.MemoryStore: ensureFreeSpace(21204) called with curMem=937410, maxMem=280248975
23/08/01 12:45:14 INFO storage.MemoryStore: Block broadcast_8_piece0 stored as bytes in memory (estimated size 20.7 KB, free 266.4 MB)
23/08/01 12:45:14 INFO storage.BlockManagerInfo: Added broadcast_8_piece0 in memory on localhost:42552 (size: 20.7 KB, free: 267.2 MB)
23/08/01 12:45:14 INFO storage.BlockManagerMaster: Updated info of block broadcast_8_piece0
23/08/01 12:45:14 INFO spark.SparkContext: Created broadcast 8 from textFile at NativeMethodAccessorImpl.java:-2