

# LAB - 10

Name: Praful Patil

NetID: pvp220001

## Screenshot - 10.2

```
prafulpatil — training@localhost:~ — ssh -oHostKeyAlgorithms+=ssh-rsa training@10.182.131.54 — 153x49
23/07/24 10:28:50 INFO DAGScheduler: Final stage: Stage 1(collect at <ipython-input-8-0a9f21ece71b>:1)
23/07/24 10:28:50 INFO DAGScheduler: Parents of final stage: List()
23/07/24 10:28:50 INFO DAGScheduler: Missing parents: List()
23/07/24 10:28:50 INFO DAGScheduler: Submitting Stage 1 (file:/home/training/training_materials/data/frostroad.txt MapPartitionsRDD[1] at textFile at NativeMethodAccessorImpl.java:-2), which has no missing parents
23/07/24 10:28:50 INFO MemoryStore: ensureFreeSpace(2696) called with curMem=310311, maxMem=280248975
23/07/24 10:28:50 INFO MemoryStore: Block broadcast_2 stored as values in memory (estimated size 2.6 KB, free 267.0 MB)
23/07/24 10:28:50 INFO MemoryStore: ensureFreeSpace(1659) called with curMem=313007, maxMem=280248975
23/07/24 10:28:50 INFO MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 1659.0 B, free 267.0 MB)
23/07/24 10:28:50 INFO BlockManagerInfo: Added broadcast_2_piece0 in memory on localhost:45356 (size: 1659.0 B, free: 267.2 MB)
23/07/24 10:28:50 INFO BlockManagerMaster: Updated info of block broadcast_2_piece0
23/07/24 10:28:50 INFO SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:839
23/07/24 10:28:50 INFO DAGScheduler: Submitting 1 missing tasks from Stage 1 (file:/home/training/training_materials/data/frostroad.txt MapPartitionsRDD[1] at textFile at NativeMethodAccessorImpl.java:-2)
23/07/24 10:28:50 INFO TaskSchedulerImpl: Adding task set 1.0 with 1 tasks
23/07/24 10:28:50 INFO TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, localhost, PROCESS_LOCAL, 1321 bytes)
23/07/24 10:28:50 INFO Executor: Running task 0.0 in stage 1.0 (TID 1)
23/07/24 10:28:50 INFO HadoopRDD: Input split: file:/home/training/training_materials/data/frostroad.txt:0+730
23/07/24 10:28:50 INFO Executor: Finished task 0.0 in stage 1.0 (TID 1). 2540 bytes result sent to driver
23/07/24 10:28:50 INFO DAGScheduler: Stage 1 (collect at <ipython-input-8-0a9f21ece71b>:1) finished in 0.019 s
23/07/24 10:28:50 INFO DAGScheduler: Job 1 finished: collect at <ipython-input-8-0a9f21ece71b>:1, took 0.062238 s
Out[8]:
[u'Two roads diverged in a yellow wood,',
 u'And sorry I could not travel both',
 u'And be one traveler, long I stood',
 u'And looked down one as far as I could',
 u'To where it bent in the undergrowth;',
 u'',
 u'Then took the other, as just as fair,',
 u'And having perhaps the better claim,',
 u'Because it was grassy and wanted wear;',
 u'Though as for that the passing there',
 u'Had worn them really about the same,',
 u'',
 u'And both that morning equally lay',
 u'In leaves no step had trodden black.',
 u'Oh, I kept the first for another day!',
 u'Yet knowing how way leads on to way,',
 u'I doubted if I should ever come back.',
 u'',
 u'I shall be telling this with a sigh',
 u'Somewhere ages and ages hence:',
 u'Two roads diverged in a wood, and I--',
 u'I took the one less traveled by,',
 u'And that has made all the difference.']

In [9]: 23/07/24 10:28:50 INFO TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 18 ms on localhost (1/1)
23/07/24 10:28:50 INFO TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
In [9]:
```

## Screenshot 10.3.5

```
In [8]: jpglogs.take(10)
Out[8]:
[u'217.150.149.167 - 4712 [15/Sep/2013:23:56:06 +0100] "GET /ronin_s4.jpg HTTP/1.0" 200 5552 "http://www.loudacre.com" "Loudacre Mobile Browser MeeToo 1.0"',
 u'104.184.210.93 - 28402 [15/Sep/2013:23:42:53 +0100] "GET /titanic_2200.jpg HTTP/1.0" 200 19466 "http://www.loudacre.com" "Loudacre Mobile Browser MeeToo 2.0"',
 u'37.91.137.134 - 36171 [15/Sep/2013:23:39:33 +0100] "GET /ronin_novelty_note_3.jpg HTTP/1.0" 200 7432 "http://www.loudacre.com" "Loudacre Mobile Browser iFruit 3"',
 u'177.43.223.203 - 90653 [15/Sep/2013:23:31:17 +0100] "GET /ifruit_3.jpg HTTP/1.0" 200 19578 "http://www.loudacre.com" "Loudacre Mobile Browser Sorrento F31L"',
 u'19.250.65.76 - 44388 [15/Sep/2013:23:31:10 +0100] "GET /sorrento_f24l.jpg HTTP/1.0" 200 5730 "http://www.loudacre.com" "Loudacre Mobile Browser iFruit 3A"',
 u'134.72.143.150 - 24554 [15/Sep/2013:23:13:42 +0100] "GET /sorrento_f24l.jpg HTTP/1.0" 200 703 "http://www.loudacre.com" "Loudacre Mobile Browser iFruit 1"',
 u'48.202.252.134 - 24990 [15/Sep/2013:23:12:00 +0100] "GET /ifruit_3a.jpg HTTP/1.0" 200 1730 "http://www.loudacre.com" "Loudacre Mobile Browser Titanic 2000"',
 u'100.30.199.161 - 9834 [15/Sep/2013:23:06:14 +0100] "GET /sorrento_f40l.jpg HTTP/1.0" 200 15995 "http://www.loudacre.com" "Loudacre Mobile Browser iFruit 1"',
 u'58.46.139.19 - 10399 [15/Sep/2013:23:03:16 +0100] "GET /ronin_novelty_note_3.jpg HTTP/1.0" 200 17725 "http://www.loudacre.com" "Loudacre Mobile Browser Titanic 1100"',
 u'135.41.174.97 - 58228 [15/Sep/2013:22:29:56 +0100] "GET /sorrento_f31l.jpg HTTP/1.0" 200 14615 "http://www.loudacre.com" "Loudacre Mobile Browser iFruit 3"']
```

## Screenshot 10.3.6

```
In [9]: sc.textFile(logfile).filter(lambda line: \
...: ".jpg" in line).count()
[Stage 3:> (0 + 0) / 311]
Out[9]: 64978
```

# LAB - 10

Name: Praful Patil  
NetID: pvp220001

## Screenshot 10.3.11

