SSH-in-browser

```
vp2359_nyu_edu@nyu-dataproc-m:~$ ls
 Lab5           TotalOrderSort.txt   __MACOSX      activations_vp.zip  'app_(2).jar'     part-r-00000   par
 PageRank.txt   _SUCCESS             activations   'app_(1).jar'       pagerank_bdad.txt part-r-00001   par
vp2359_nyu_edu@nyu-dataproc-m:~$ cd Lab5
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5$ cd ..
vp2359_nyu_edu@nyu-dataproc-m:~$ hadoop fs -rm -r loudacre/accounts-models
Deleted loudacre/accounts-models
```

```
[success] Total time: 0 s, completed Mar 8, 2024 8:07:12 PM
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5$ sbt package
[info] welcome to sbt 1.9.9 (Temurin Java 1.8.0_392)
[info] loading project definition from /home/vp2359_nyu_edu/Lab5/project
[info] loading settings for project lab5 from build.sbt ...
[info] set current project to XML Reader (in build file:/home/vp2359_nyu_edu/Lab5/)
[info] compiling 1 Scala source to /home/vp2359_nyu_edu/Lab5/target/scala-2.12/classes ...
[success] Total time: 10 s, completed Mar 8, 2024 8:07:31 PM
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5$ ls
build.sbt  project  src  target
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5$ cd project
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5/project$ ls
build.properties  target
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5/project$ cd ..
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5$ cd target
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5/target$ ls
global-logging  scala-2.12  streams  task-temp-directory
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5/target$ cd scala-2.12/
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5/target/scala-2.12$ ls
classes  sync  update  xml-reader_2.12-1.0.jar  zinc
```

```
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5/target/scala-2.12$ spark-submit --deploy-mode client --class XMLReader xml-reader_2.12-1.0.jar
24/03/08 20:08:22 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
24/03/08 20:08:22 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
24/03/08 20:08:22 INFO org.apache.spark.SparkEnv: Registering BlockManagerMasterHeartbeat
24/03/08 20:08:22 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
24/03/08 20:08:23 INFO org.sparkproject.jetty.util.log: Logging initialized @4055ms to org.sparkproject.jetty.util.log.Slf4jLog
24/03/08 20:08:23 INFO org.sparkproject.jetty.server.Server: jetty-9.4.40.v20210413; built: 2021-04-13T20:42:42.668Z; git: b881a572662e1
24/03/08 20:08:23 INFO org.sparkproject.jetty.server.Server: Started @4167ms
24/03/08 20:08:23 INFO org.sparkproject.jetty.server.AbstractConnector: Started ServerConnector@42e627a8{HTTP/1.1, (http/1.1)}{0.0.0.0:4
24/03/08 20:08:23 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at nyu-dataproc-m/192.168.1.77:8032
24/03/08 20:08:24 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at nyu-dataproc-m/192.168.1.77:1
24/03/08 20:08:24 INFO org.apache.hadoop.conf.Configuration: resource-types.xml not found
24/03/08 20:08:24 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Unable to find 'resource-types.xml'.
24/03/08 20:08:25 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl: Submitted application application_1704906963891_22668
24/03/08 20:08:26 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at nyu-dataproc-m/192.168.1.77:8030
24/03/08 20:08:28 INFO com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics: Detected potential high latency for operation op_get_file_s
ontext=gs://nyu-dataproc-temp/spark-job-history
24/03/08 20:08:28 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageImpl: Ignoring exception o
s with desired state.
24/03/08 20:08:28 INFO com.google.cloud.hadoop.fs.gcs.GhfsStorageStatistics: Detected potential high latency for operation op_mkdirs. la
://nyu-dataproc-temp/spark-job-history
24/03/08 20:08:30 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat: Total input files to process : 66
24/03/08 20:08:30 INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat: Total input files to process : 66
9763:MeeToo 1.0
426:Titanic 1000
383:Sorrento F00L
404:MeeToo 1.0
393:iFruit 1
53:MeeToo 1.0
96:iFruit 1
283:Sorrento F00L
464:MeeToo 1.0
475:Titanic 1000
479:iFruit 1
306:Titanic 1000
14045:MeeToo 1.0
```

```
349:MeeToo 1.0
124:MeeToo 1.0
247:MeeToo 1.0
88:Sorrento F00L
2939:Titanic 1000
314:MeeToo 1.0
24/03/08 20:08:35 INFO org.sparkproject.jetty.server.AbstractConnector: Stopped Spark@42e627a8{HTTP/1.1, (http/1.1)}{0.0.0.0:0}
^Cvp2359_nyu_edu@nyu-dataproc-m:~/Lab5/target/scala-2.12$ hadoop fs -ls loudacre/accounts-models
Found 3 items
-rw-r--r--   1 vp2359_nyu_edu vp2359_nyu_edu        0 2024-03-08 20:08 loudacre/accounts-models/_SUCCESS
-rw-r--r--   1 vp2359_nyu_edu vp2359_nyu_edu  2188453 2024-03-08 20:08 loudacre/accounts-models/part-00000
-rw-r--r--   1 vp2359_nyu_edu vp2359_nyu_edu  1555759 2024-03-08 20:08 loudacre/accounts-models/part-00001
vp2359_nyu_edu@nyu-dataproc-m:~/Lab5/target/scala-2.12$
```