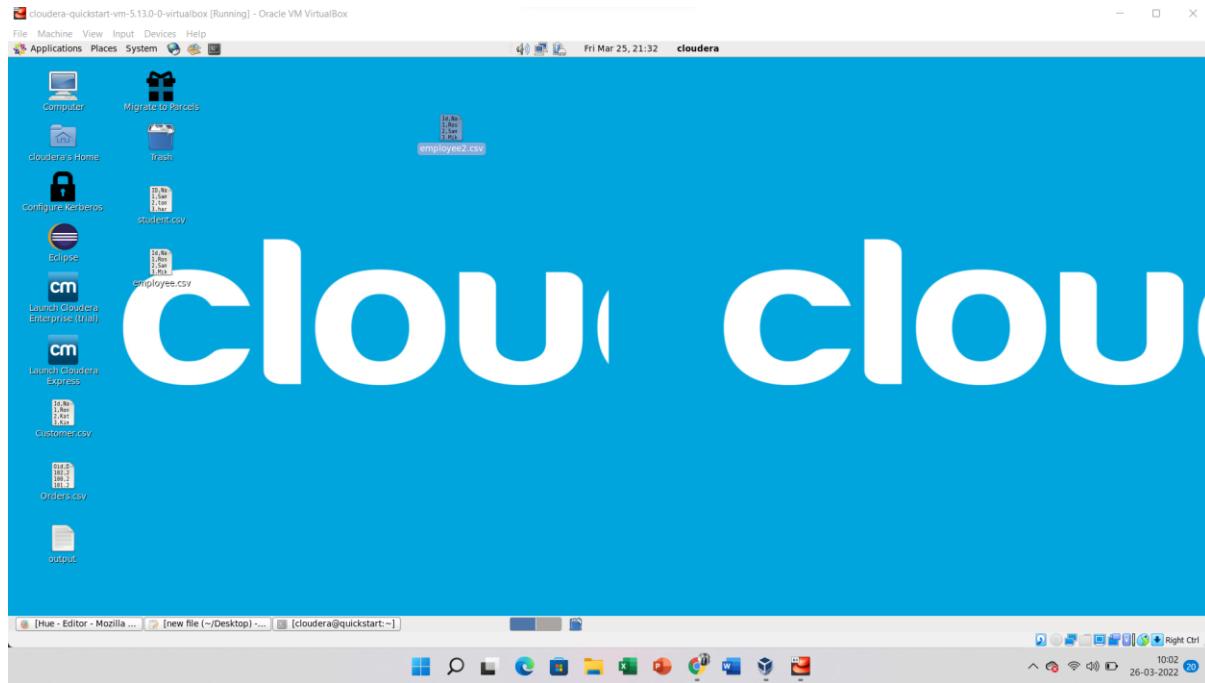
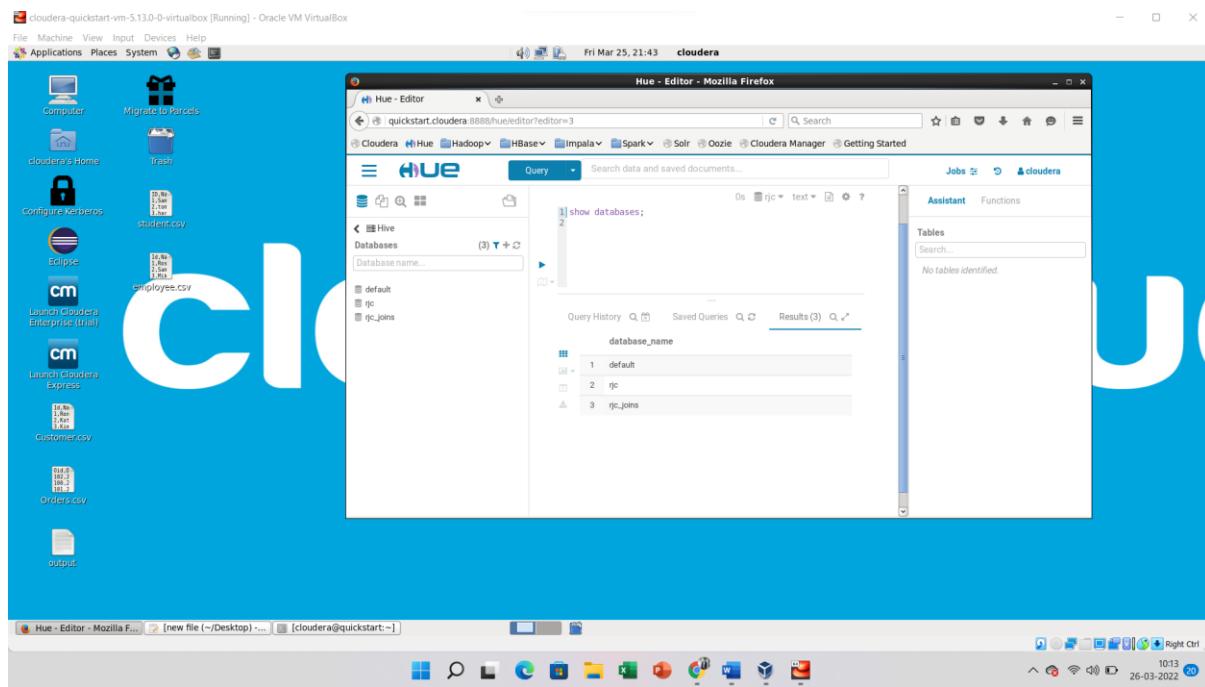


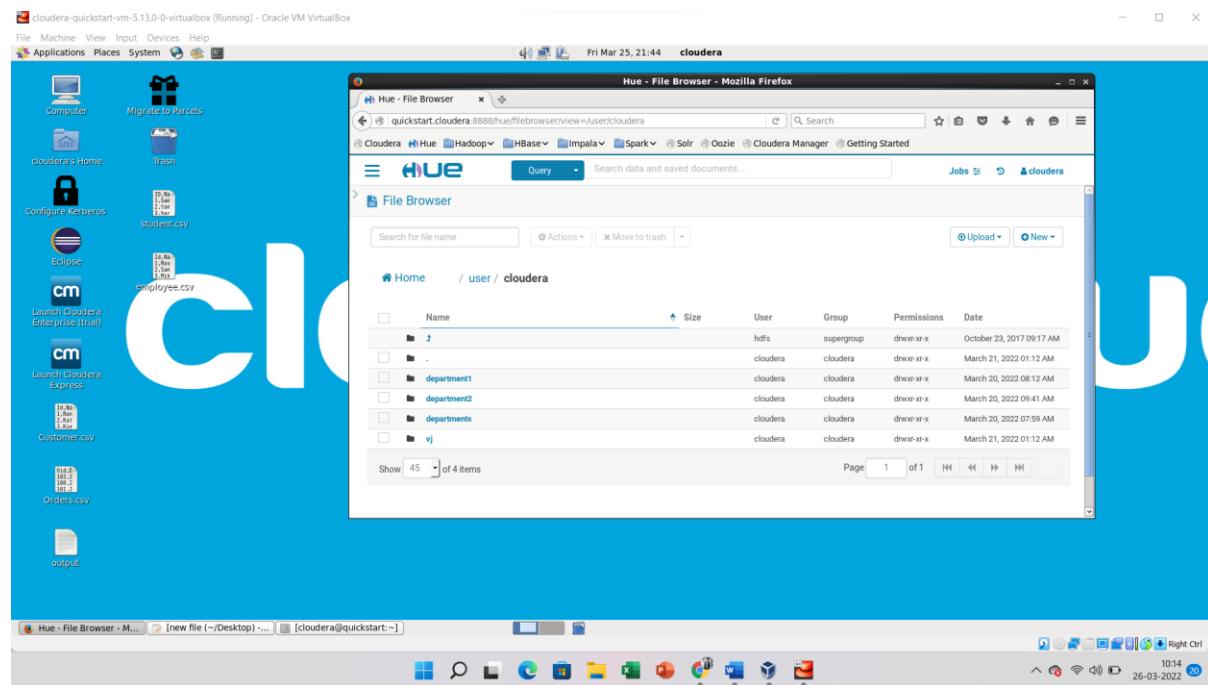
pig



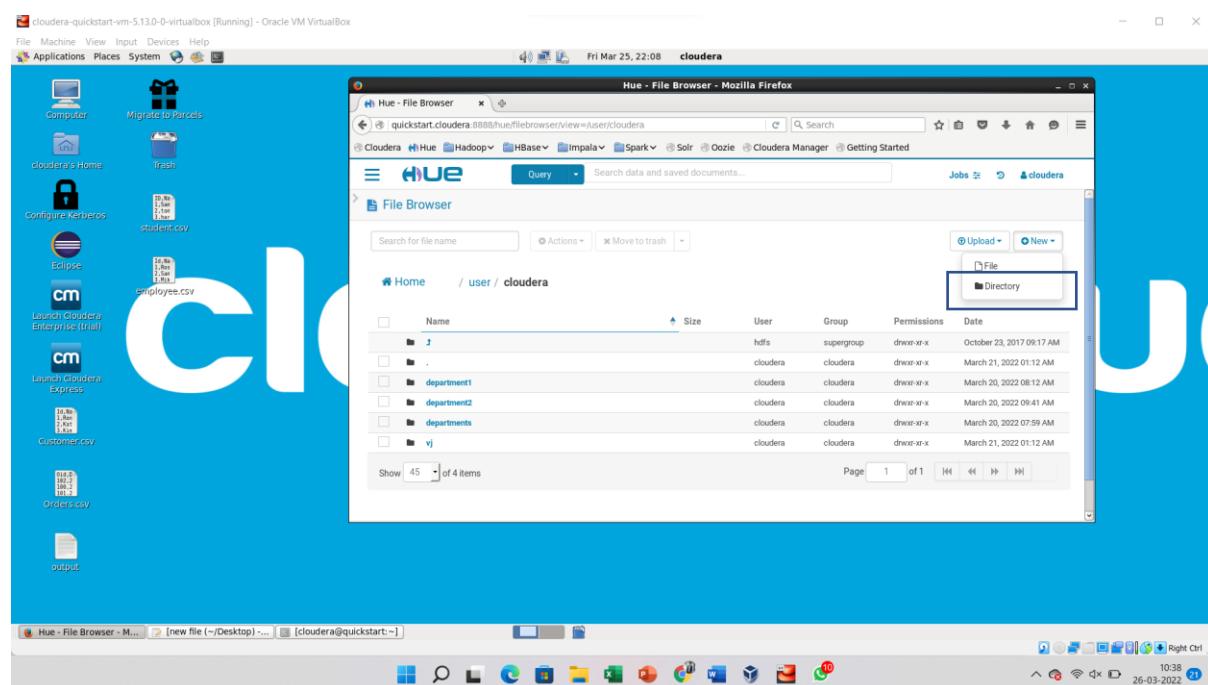
Open the browser And then open Hue and login.



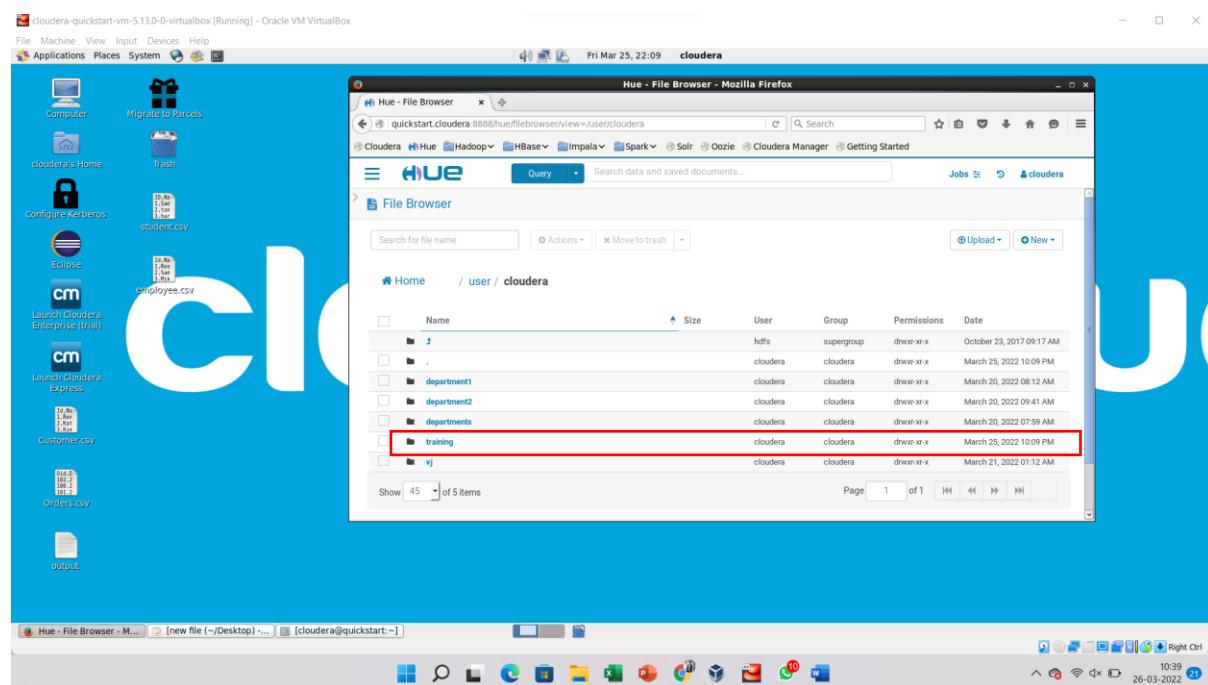
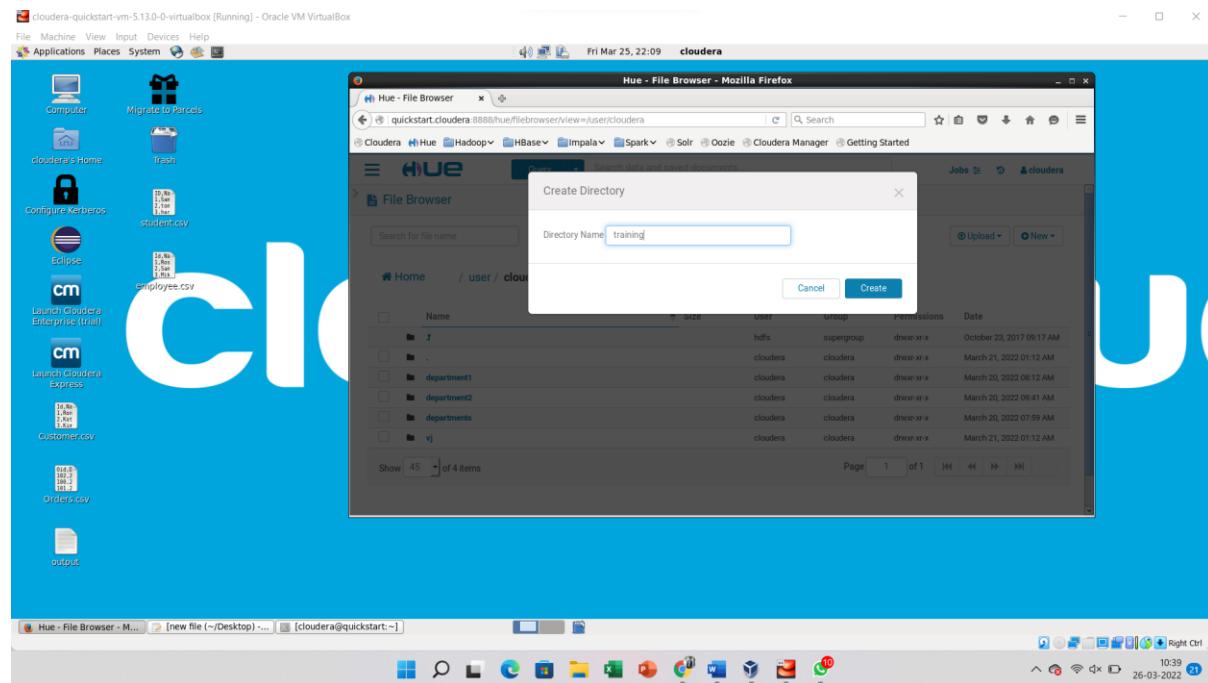
File Browser.



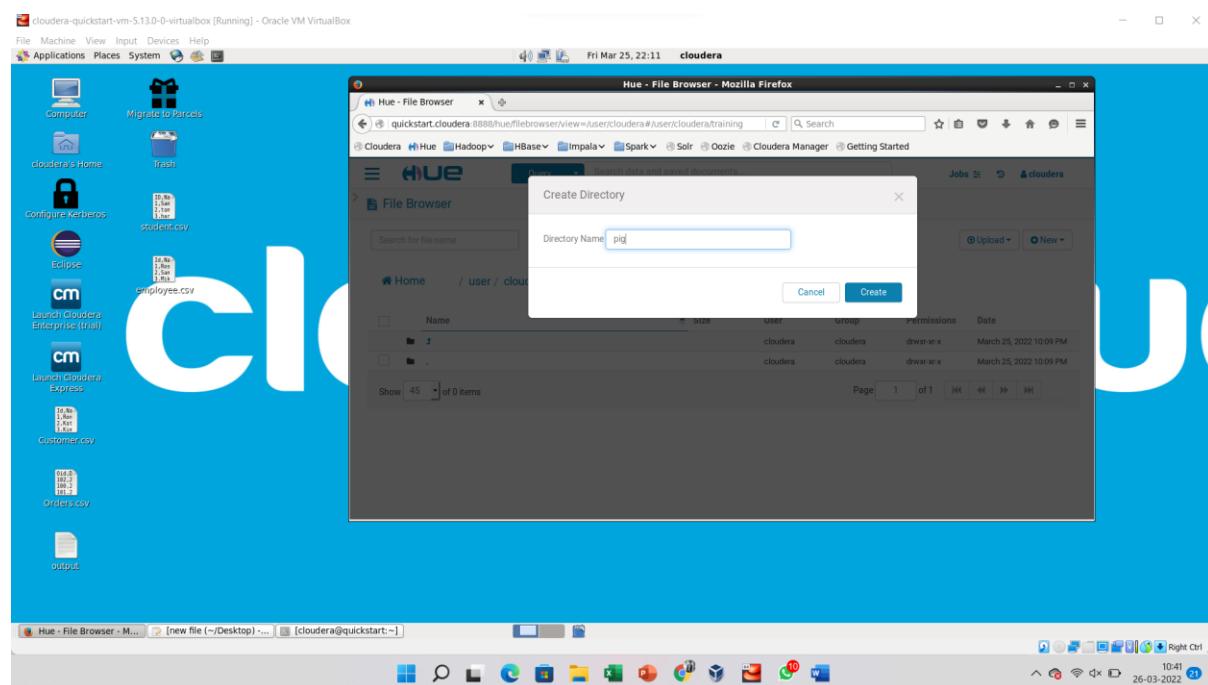
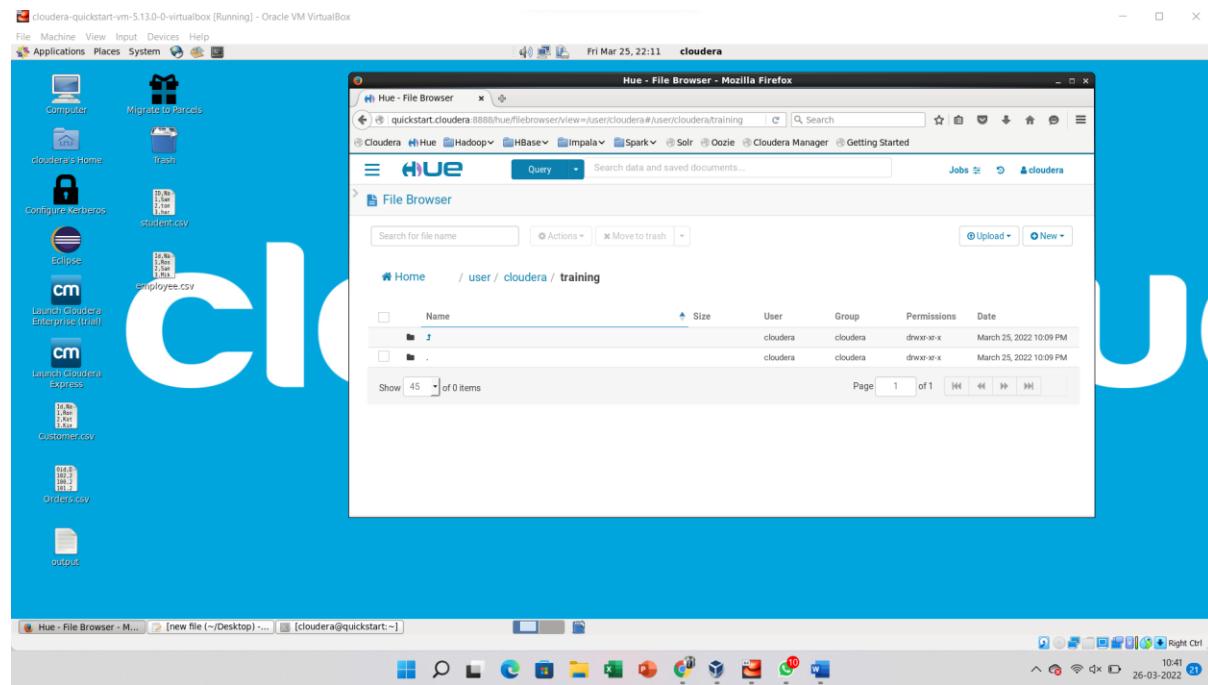
### Create Directory training

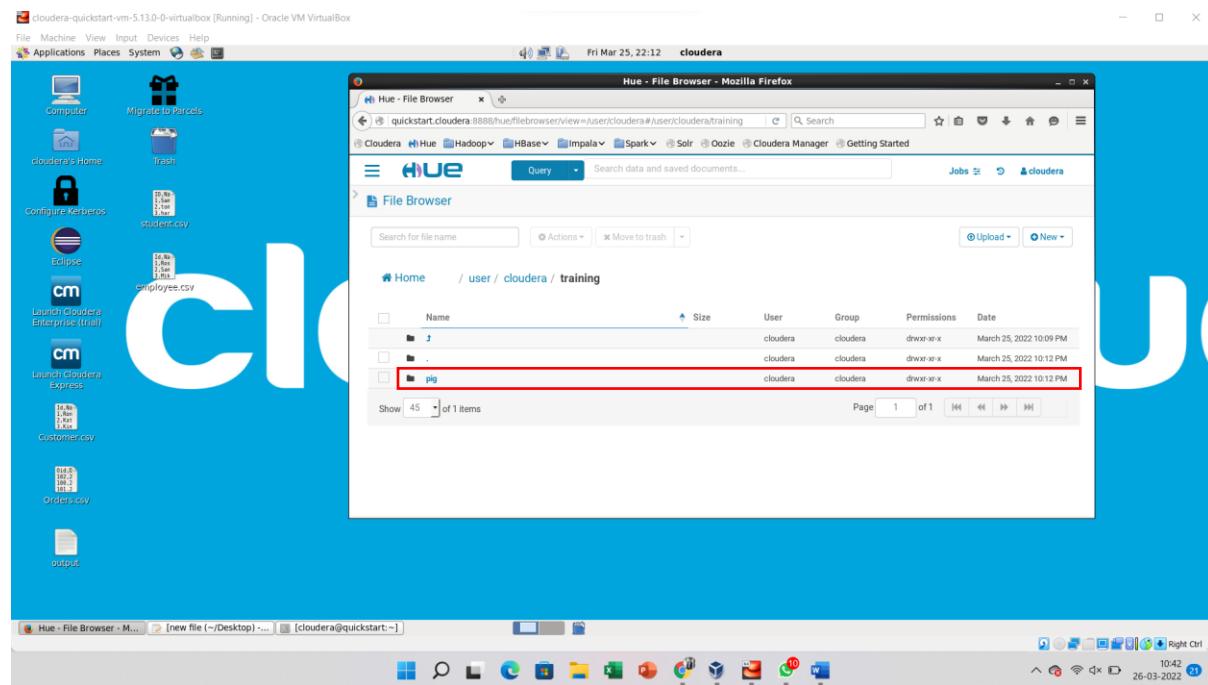


### Create training directory

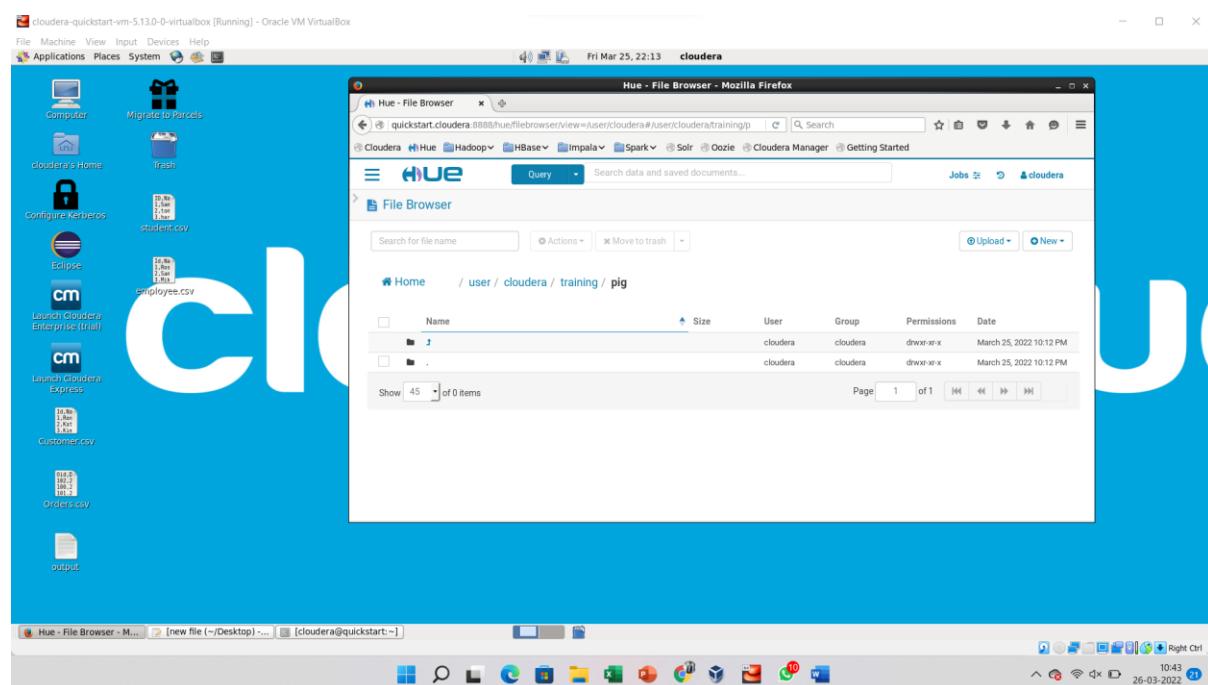


## Open the training directory

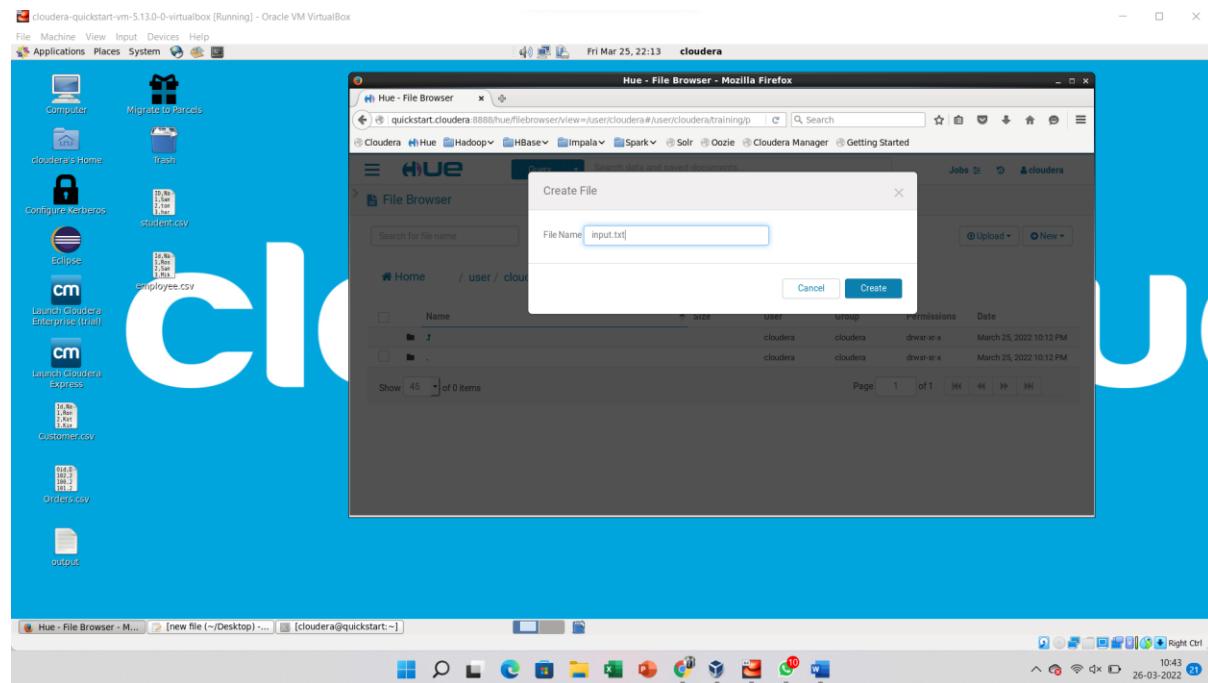




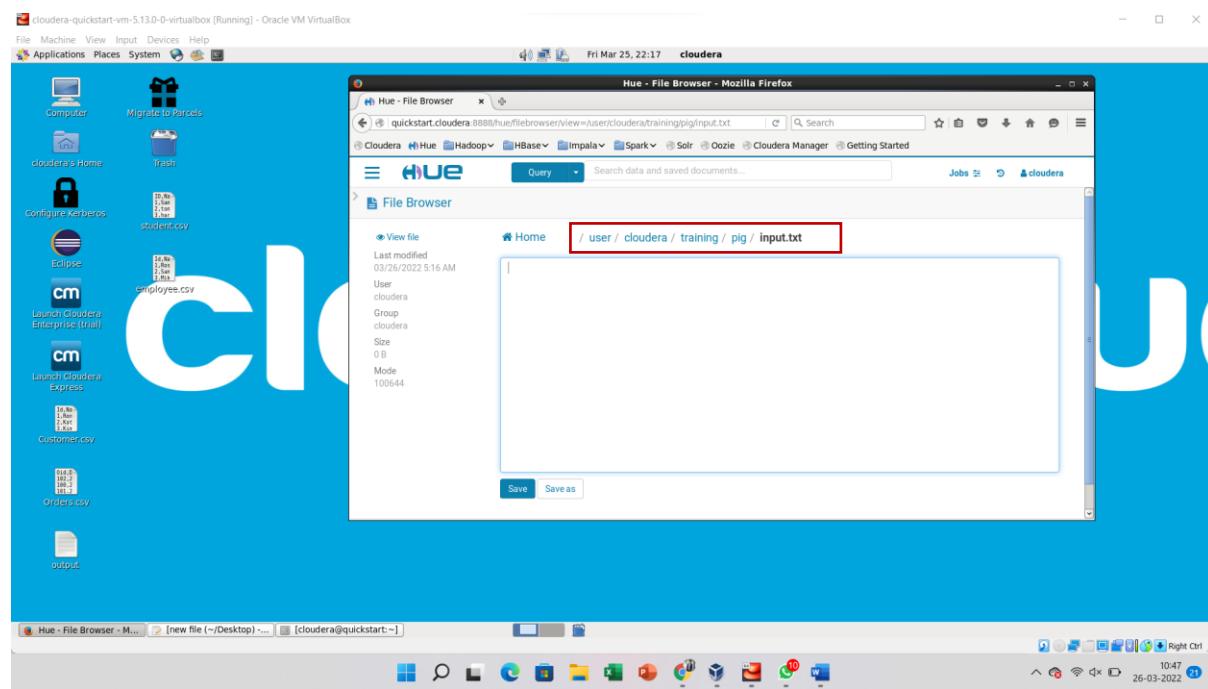
Open the pig directory

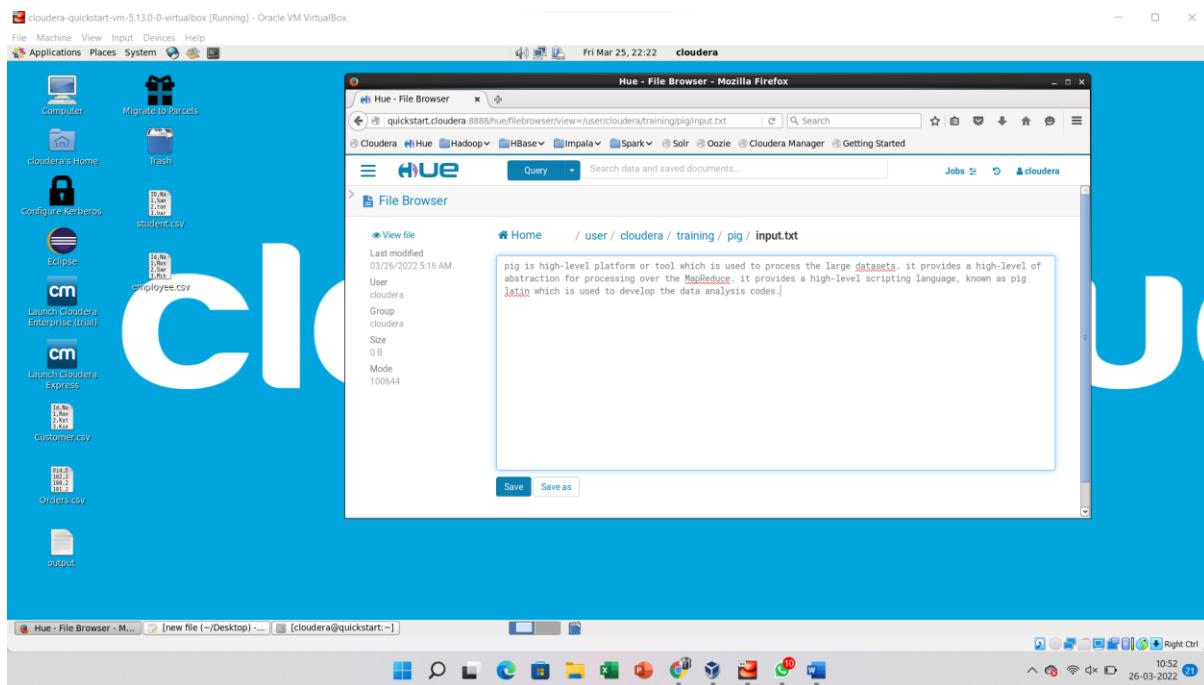


Create new file inside the pig directory and give the name as input.txt

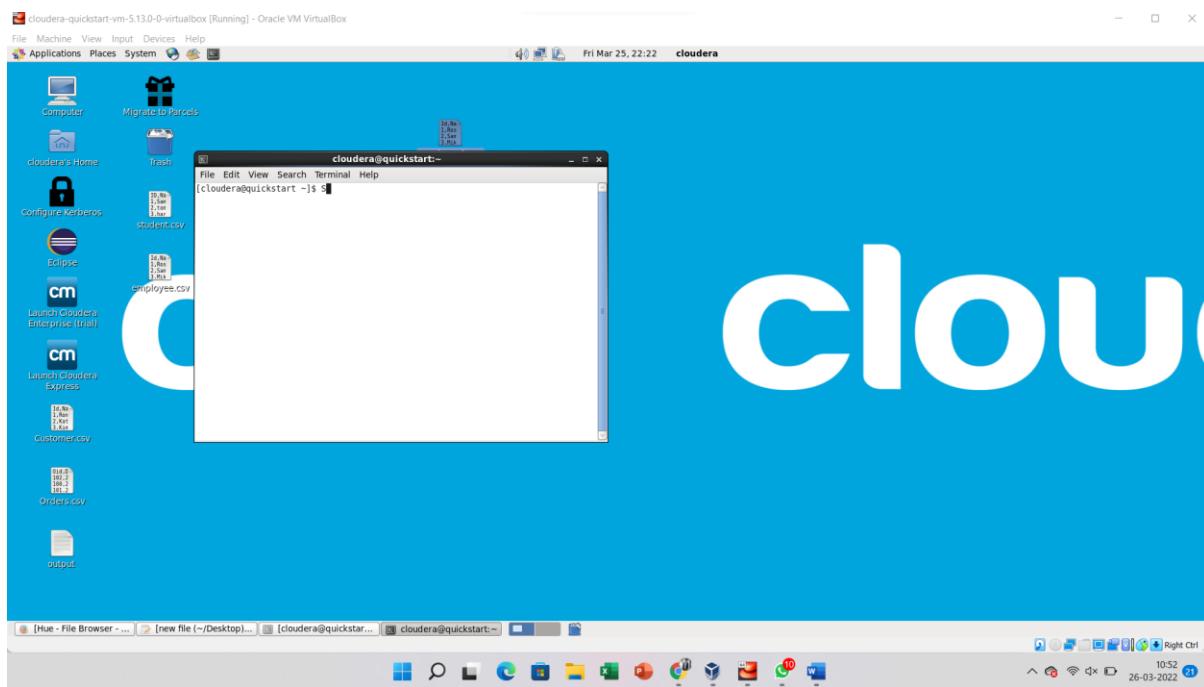


Now add some text in input.txt





Now open the terminal for pig command

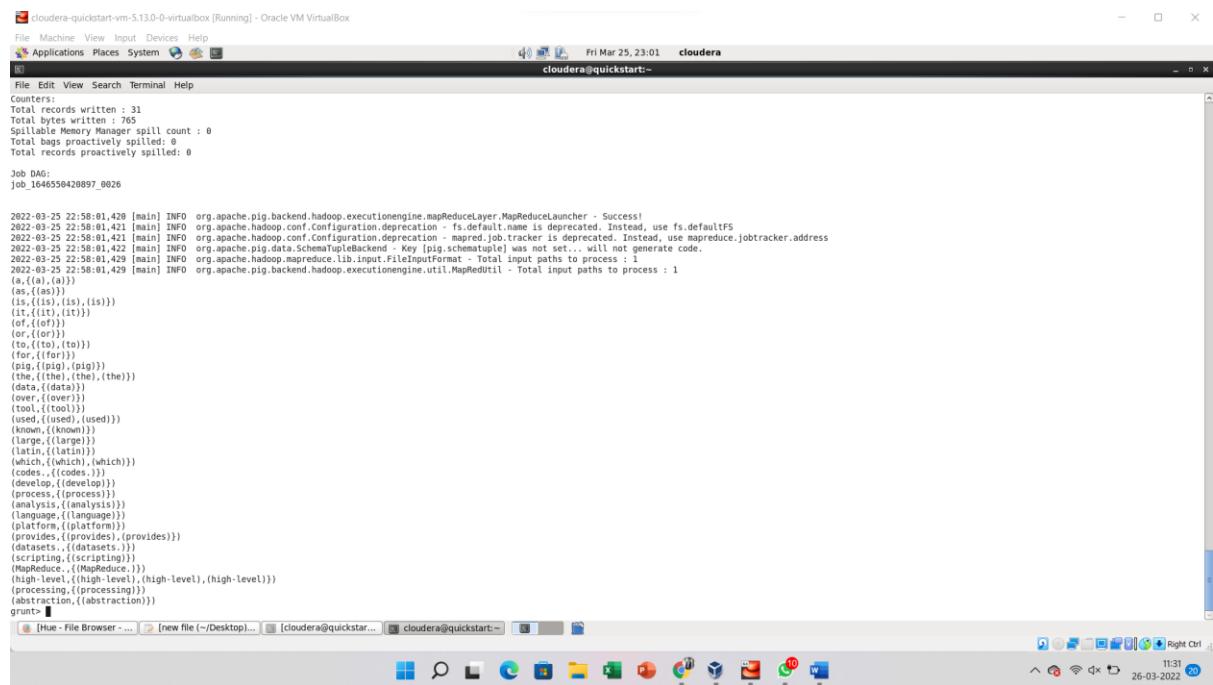












```

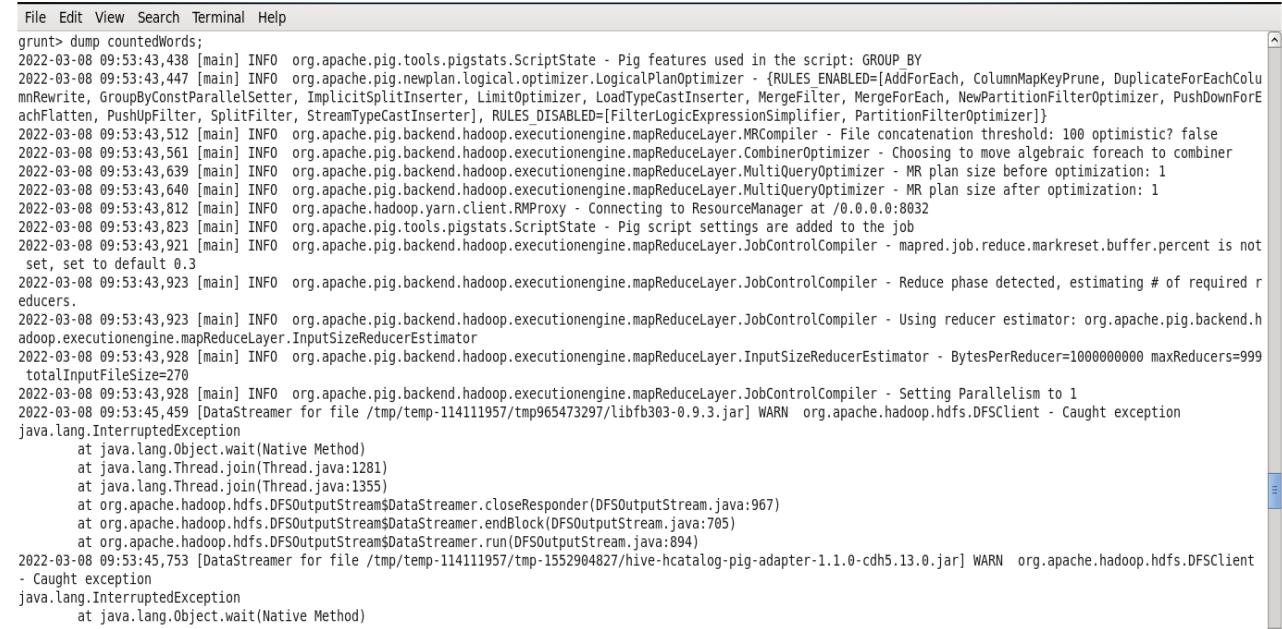
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Fri Mar 25, 23:01 cloudera
cloudera@quickstart:-
File Edit View Search Terminal Help
Counters:
Total records written : 31
Total bytes written : 765
Spilled records : 0 Spill count : 0
Total records proactively spilled: 0
Total bytes proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_104550420897_0026

2022-03-25 22:58:01,426 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-25 22:58:01,421 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-25 22:58:01,421 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-25 22:58:01,422 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-25 22:58:01,429 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-25 22:58:01,429 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MappedUtil - Total input paths to process : 1
(a,((a),(a)))
(as,{(as)})
(is,{(is),(is),(is)})
(it,{(it),(it),(it)})
(or,{(or),(or),(or)})
(or,{(or),(or),(or)})
(to,{(to),(to),(to)})
(for,{(for),(for),(for)})
(pig,{(pig),(pig),(pig)})
(the,{(the),(the),(the),(the)})
(data,{(data)})
(over,{(over),(over),(over),(over)})
(used,{(used),(used),(used),(used)})
(known,{(known),(known),(known),(known)})
(large,{(large),(large),(large),(large)})
(lal,{(lal),(lal),(lal),(lal)})
(which,{(which),(which),(which),(which)})
(codes,{(codes),(codes),(codes),(codes)})
(develop,{(develop),(develop),(develop),(develop)})
(prime,{(prime),(prime),(prime),(prime)})
(analyse,{(analyse),(analyse),(analyse),(analyse)})
(language,{(language),(language),(language),(language)})
(platform,{(platform),(platform),(platform),(platform)})
(provide,{(provide),(provide),(provide),(provide)})
(datasets,{(datasets),(datasets),(datasets),(datasets)})
(scripting,{(scripting),(scripting),(scripting),(scripting)})
(MapReduce,{(MapReduce),(MapReduce),(MapReduce),(MapReduce)})
(hadoop,{(hadoop),(hadoop),(hadoop),(hadoop)})
(High-level,{(High-level),(High-level),(High-level),(High-level)})
(processing,{(processing),(processing),(processing),(processing)})
(abstraction,{(abstraction),(abstraction),(abstraction),(abstraction)})
(grunt>)

```

`countedWords = foreach groupedWords generate group, COUNT(wordsInEachline);`

`dump countedWords;`

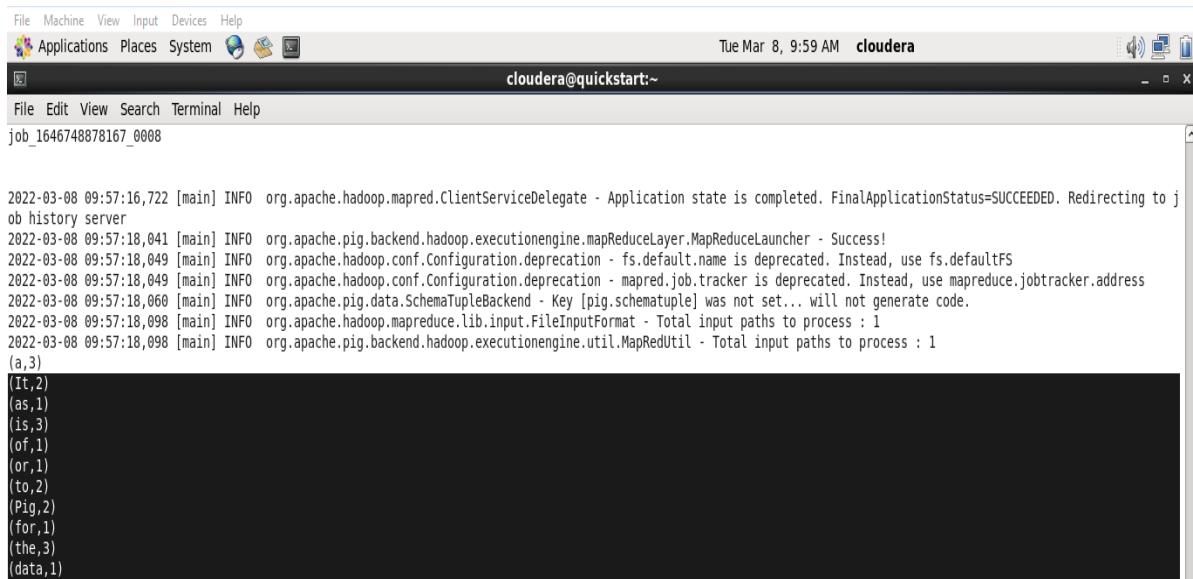


```

File Edit View Search Terminal Help
grunt> dump countedwords;
2022-03-08 09:53:43,438 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP BY
2022-03-08 09:53:43,447 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForAchFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
2022-03-08 09:53:43,512 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-03-08 09:53:43,561 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.CombinerOptimizer - Choosing to move algebraic foreach to combiner
2022-03-08 09:53:43,639 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-03-08 09:53:43,640 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-03-08 09:53:43,812 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at /0.0.0.0:8082
2022-03-08 09:53:43,823 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-03-08 09:53:43,921 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-03-08 09:53:43,923 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2022-03-08 09:53:43,923 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator
2022-03-08 09:53:43,928 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=270
2022-03-08 09:53:43,928 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting Parallelism to 1
2022-03-08 09:53:45,459 [DataStreamer for file /tmp/temp-114111957/tmp-1552904827/hive-hcatalog-pig-adapter-1.1.0-cdh5.13.0.jar] WARN org.apache.hadoop.hdfs.DFSClient - Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
2022-03-08 09:53:45,753 [DataStreamer for file /tmp/temp-114111957/tmp-1552904827/hive-hcatalog-pig-adapter-1.1.0-cdh5.13.0.jar] WARN org.apache.hadoop.hdfs.DFSClient - Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)

```





The screenshot shows a terminal window titled "cloudera@quickstart:~". The window displays log output from a Hadoop job. The logs include informational messages from org.apache.hadoop.mapred.ClientServiceDelegate, org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher, org.apache.hadoop.conf.Configuration.deprecation, org.apache.hadoop.conf.Configuration.deprecation, org.apache.pig.data.SchemaTupleBackend, org.apache.hadoop.mapreduce.lib.input.FileInputFormat, and org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil. The log also lists tuples being processed, such as (It,2), (as,1), (is,3), (of,1), (or,1), (to,2), (Pig,2), (for,1), (the,3), and (data,1). The timestamp for the logs is Tue Mar 8, 9:59 AM.

```
File Machine View Input Devices Help
Applications Places System
Tue Mar 8, 9:59 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
job_1646748878167_0008

2022-03-08 09:57:16,722 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2022-03-08 09:57:18,041 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-03-08 09:57:18,049 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-03-08 09:57:18,049 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-03-08 09:57:18,066 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2022-03-08 09:57:18,098 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-03-08 09:57:18,098 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(a,3)
(It,2)
(as,1)
(is,3)
(of,1)
(or,1)
(to,2)
(Pig,2)
(for,1)
(the,3)
(data,1)
```