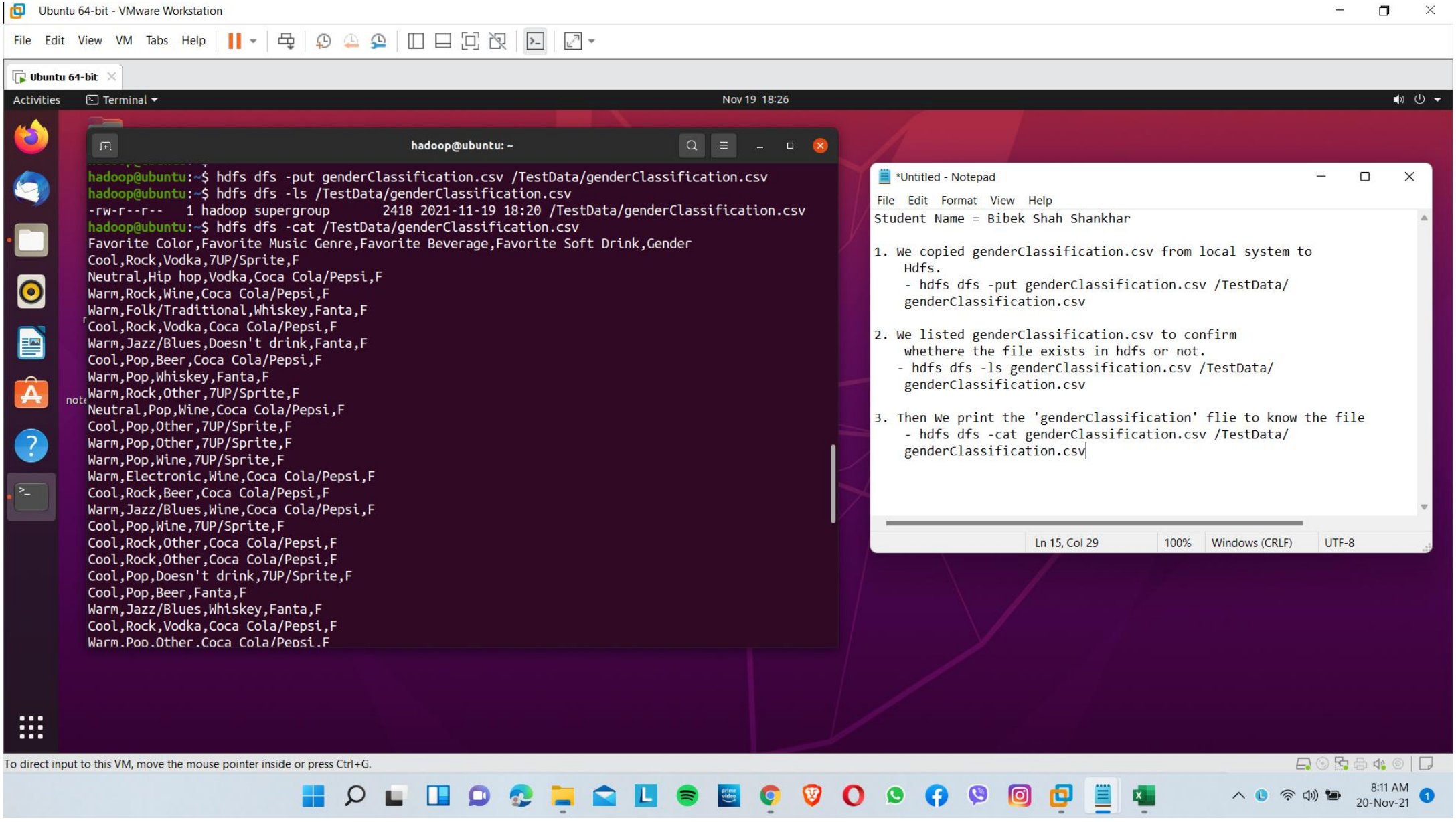
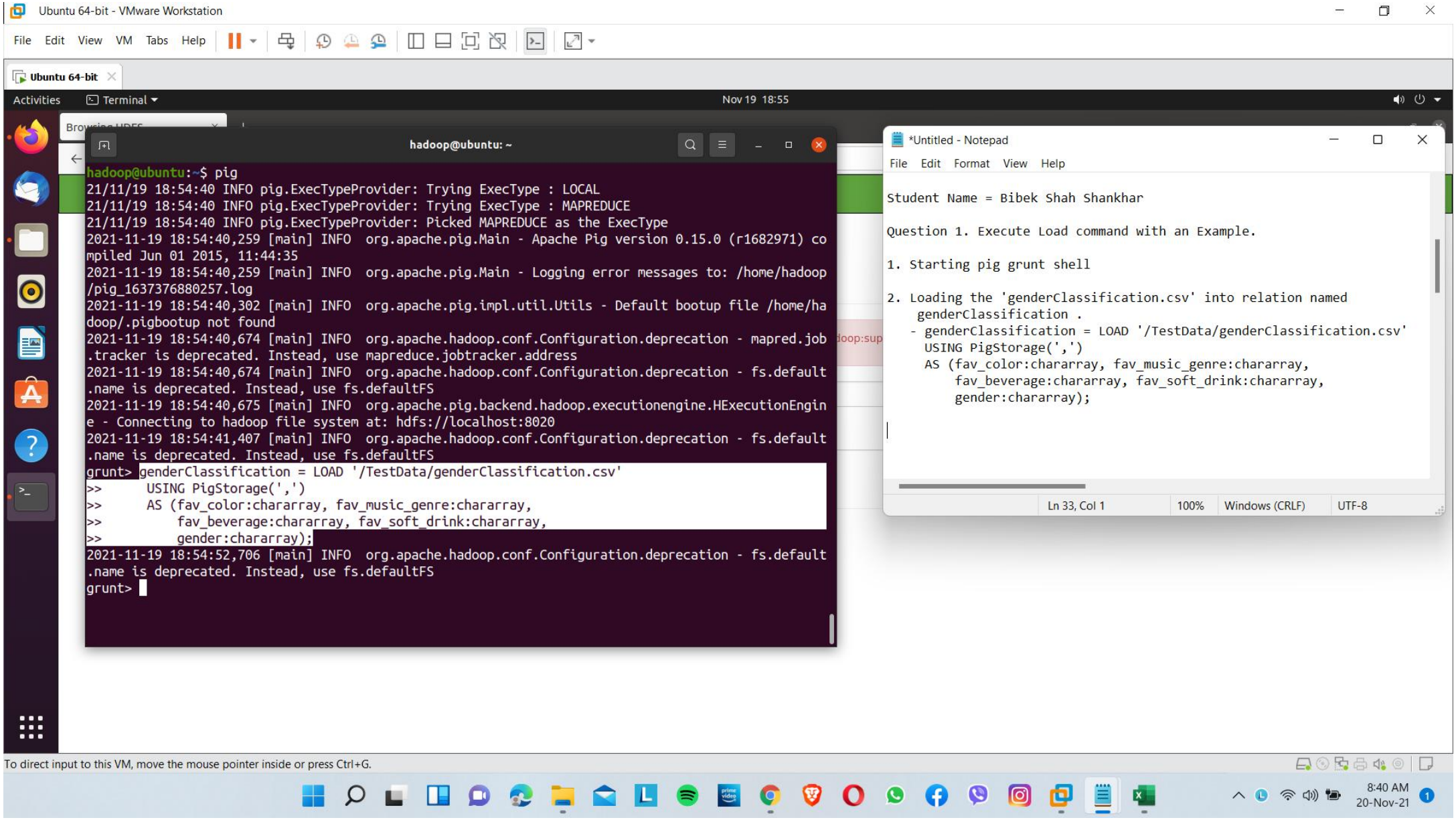
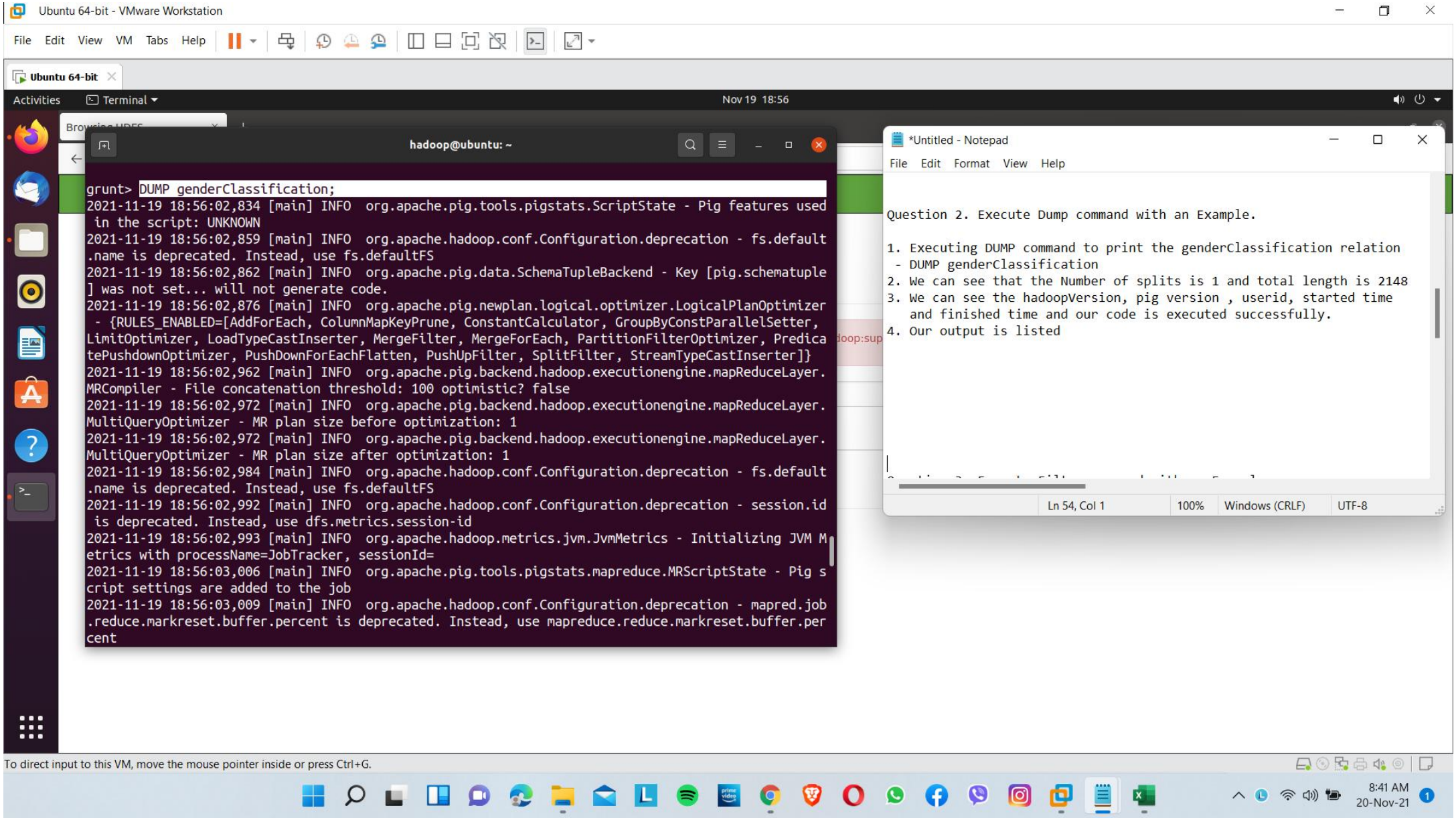
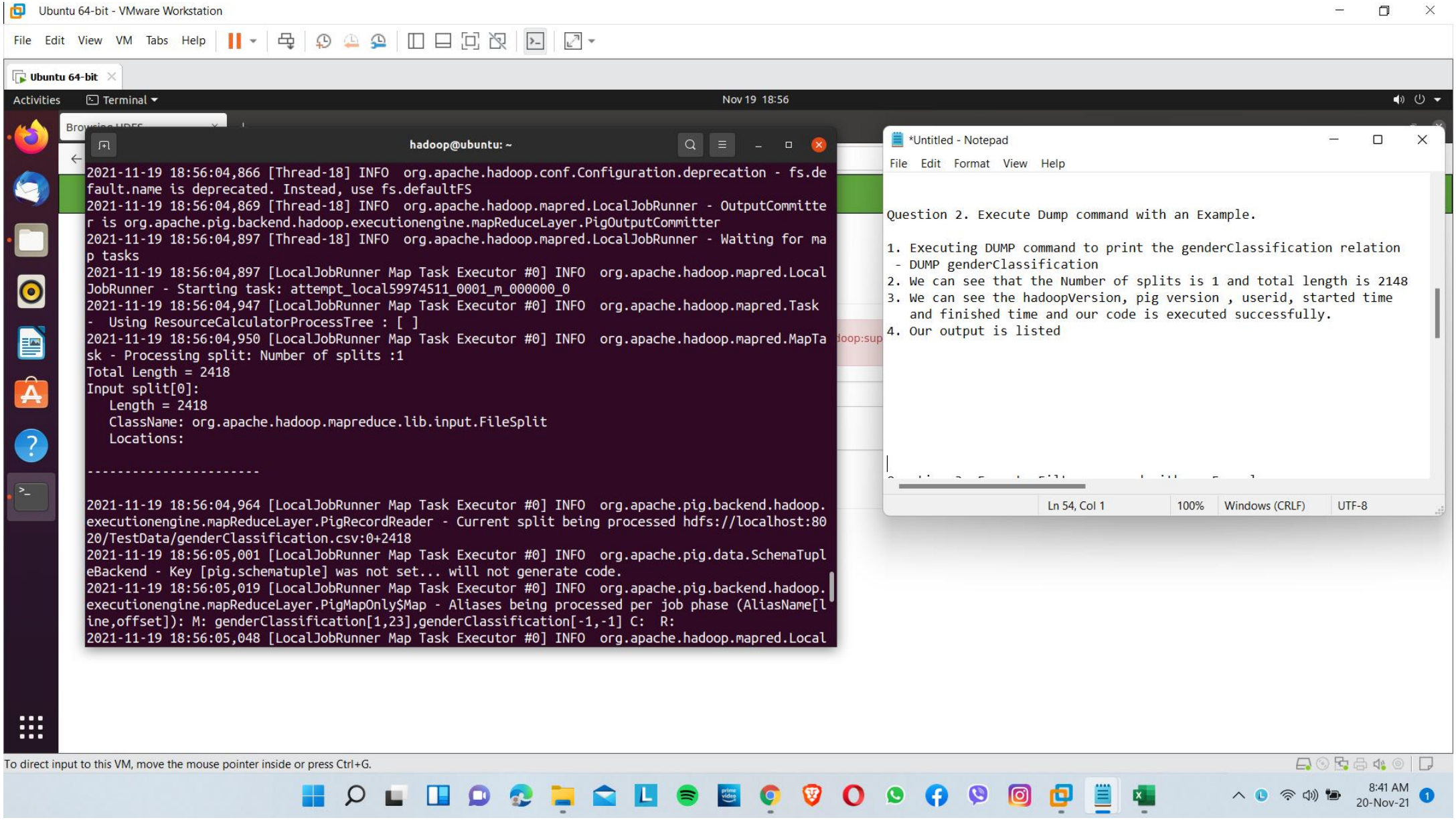
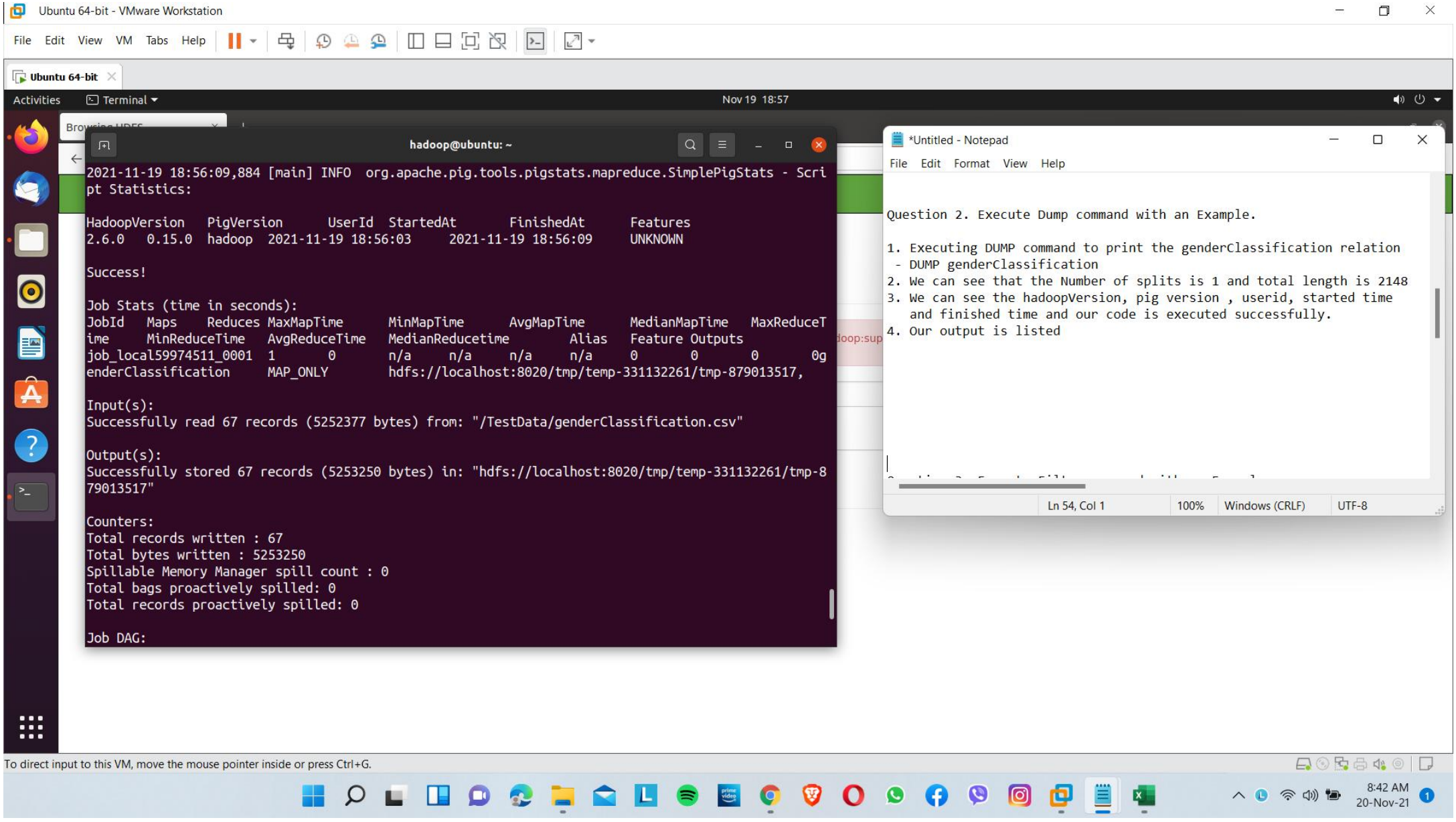
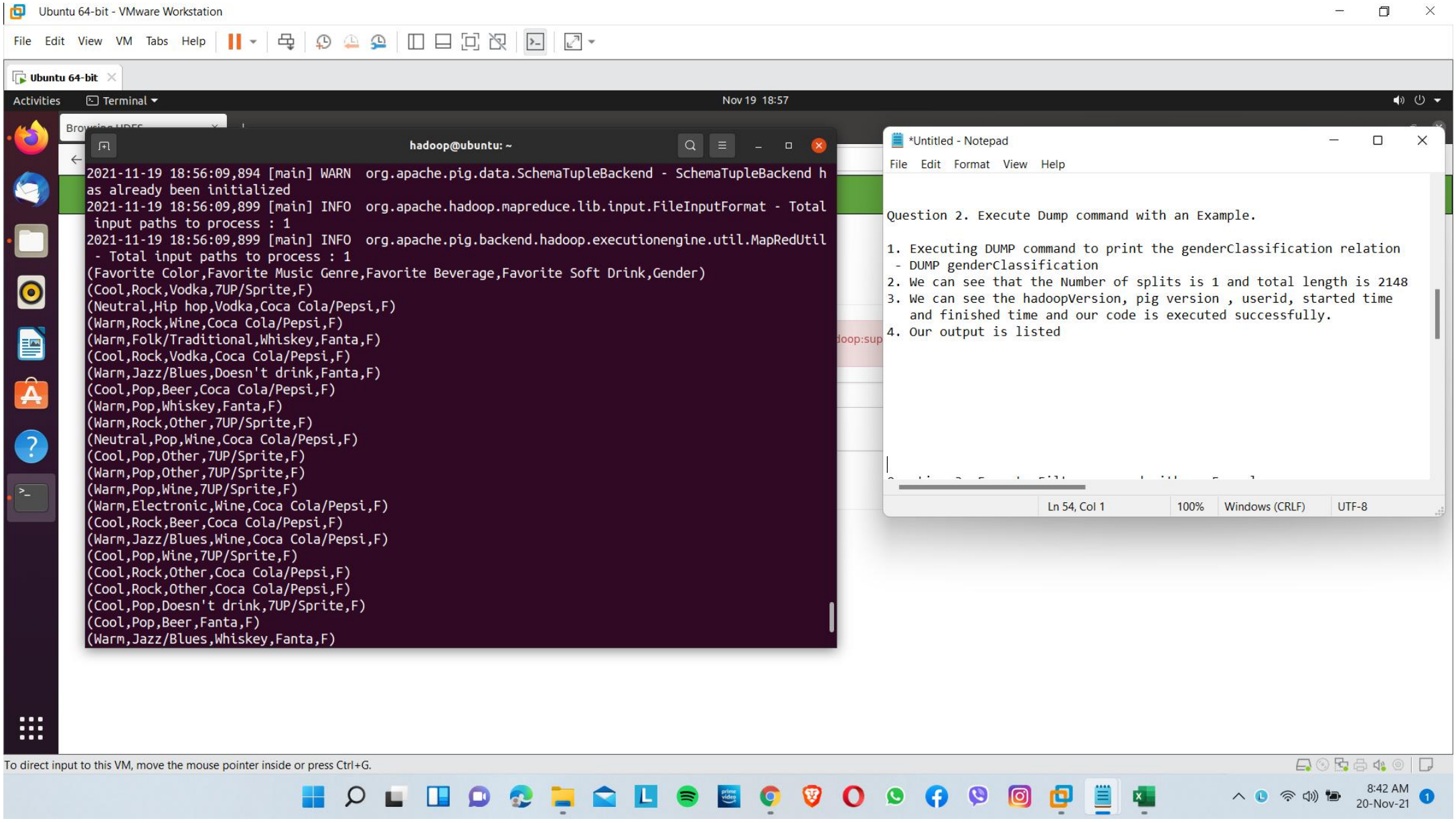

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 18:26
hadoop@ubuntu: ~
hadoop@ubuntu:~$ hdfs dfs -put genderClassification.csv /TestData/genderClassification.csv
hadoop@ubuntu:~$ hdfs dfs -ls /TestData/genderClassification.csv
-rw-r--r-- 1 hadoop supergroup 2418 2021-11-19 18:20 /TestData/genderClassification.csv
hadoop@ubuntu:~$ hdfs dfs -cat /TestData/genderClassification.csv
Favorite Color,Favorite Music Genre,Favorite Beverage,Favorite Soft Drink,Gender
Cool,Rock,Vodka,7UP/Sprite,F
Neutral,Hip hop,Vodka,Coca Cola/Pepsi,F
Warm,Rock,Wine,Coca Cola/Pepsi,F
Warm,Folk/Traditional,Whiskey,Fanta,F
Cool,Rock,Vodka,Coca Cola/Pepsi,F
Warm,Jazz/Blues,Doesn't drink,Fanta,F
Cool,Pop,Beer,Coca Cola/Pepsi,F
Warm,Pop,Whiskey,Fanta,F
Warm,Rock,Other,7UP/Sprite,F
Neutral,Pop,Wine,Coca Cola/Pepsi,F
Cool,Pop,Other,7UP/Sprite,F
Warm,Pop,Other,7UP/Sprite,F
Warm,Pop,Wine,7UP/Sprite,F
Warm,Electronic,Wine,Coca Cola/Pepsi,F
Cool,Rock,Beer,Coca Cola/Pepsi,F
Warm,Jazz/Blues,Wine,Coca Cola/Pepsi,F
Cool,Pop,Wine,7UP/Sprite,F
Cool,Rock,Other,Coca Cola/Pepsi,F
Cool,Rock,Other,Coca Cola/Pepsi,F
Cool,Pop,Doesn't drink,7UP/Sprite,F
Cool,Pop,Beer,Fanta,F
Warm,Jazz/Blues,Whiskey,Fanta,F
Cool,Rock,Vodka,Coca Cola/Pepsi,F
Warm,Pop,Other,Coca Cola/Pepsi,F
*Untitled - Notepad
File Edit Format View Help
Student Name = Bibek Shah Shankhar
1. We copied genderClassification.csv from local system to Hdfs.
- hdfs dfs -put genderClassification.csv /TestData/genderClassification.csv
2. We listed genderClassification.csv to confirm whethere the file exists in hdfs or not.
- hdfs dfs -ls genderClassification.csv /TestData/genderClassification.csv
3. Then We print the 'genderClassification' flie to know the file
- hdfs dfs -cat genderClassification.csv /TestData/genderClassification.csv
Ln 15, Col 29
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:11 AM
20-Nov-21

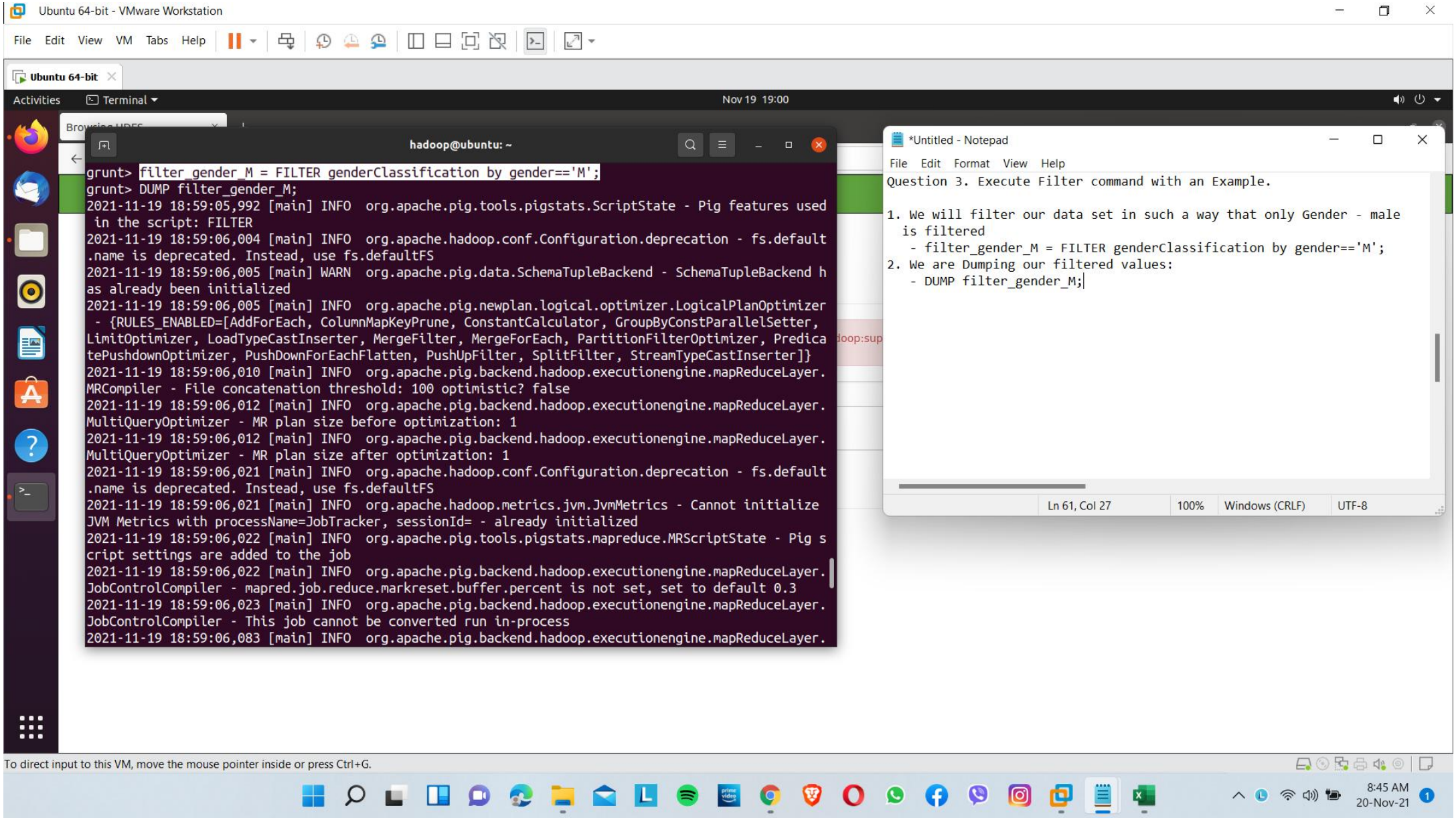

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 18:55
hadoop@ubuntu: ~
hadoop@ubuntu:~$ pig
21/11/19 18:54:40 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/11/19 18:54:40 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
21/11/19 18:54:40 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2021-11-19 18:54:40,259 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2015, 11:44:35
2021-11-19 18:54:40,259 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1637376880257.log
2021-11-19 18:54:40,302 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup not found
2021-11-19 18:54:40,674 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2021-11-19 18:54:40,674 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 18:54:40,675 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:8020
2021-11-19 18:54:41,407 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> genderClassification = LOAD '/TestData/genderClassification.csv'
>> USING PigStorage(',')
>> AS (fav_color:chararray, fav_music_genre:chararray,
>> fav_beverage:chararray, fav_soft_drink:chararray,
>> gender:chararray);
2021-11-19 18:54:52,706 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt>
*Untitled - Notepad
File Edit Format View Help
Student Name = Bibek Shah Shankhar
Question 1. Execute Load command with an Example.
1. Starting pig grunt shell
2. Loading the 'genderClassification.csv' into relation named genderClassification .
- genderClassification = LOAD '/TestData/genderClassification.csv'
USING PigStorage(',')
AS (fav_color:chararray, fav_music_genre:chararray,
fav_beverage:chararray, fav_soft_drink:chararray,
gender:chararray);
Ln 33, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:40 AM
20-Nov-21

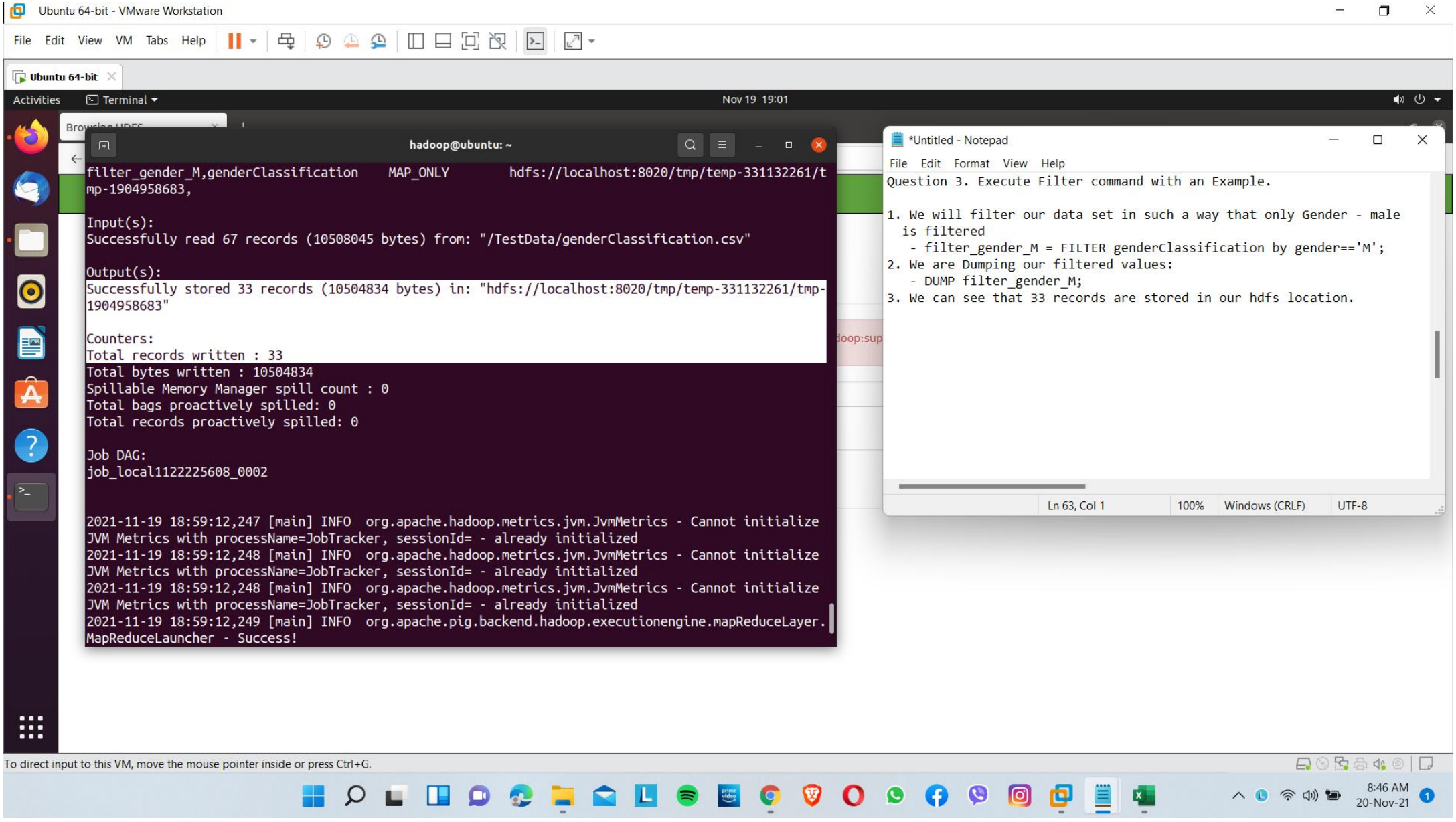

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 18:56
hadoop@ubuntu: ~
grunt> DUMP genderClassification;
2021-11-19 18:56:02,834 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2021-11-19 18:56:02,859 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 18:56:02,862 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 18:56:02,876 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 18:56:02,962 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 18:56:02,972 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 18:56:02,972 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 18:56:02,984 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 18:56:02,992 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is deprecated. Instead, use dfs.metrics.session-id
2021-11-19 18:56:02,993 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metrics with processName=JobTracker, sessionId=
2021-11-19 18:56:03,006 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 18:56:03,009 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
*Untitled - Notepad
File Edit Format View Help
Question 2. Execute Dump command with an Example.
1. Executing DUMP command to print the genderClassification relation - DUMP genderClassification
2. We can see that the Number of splits is 1 and total length is 2148
3. We can see the hadoopVersion, pig version , userid, started time and finished time and our code is executed successfully.
4. Our output is listed
Ln 54, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:41 AM
20-Nov-21

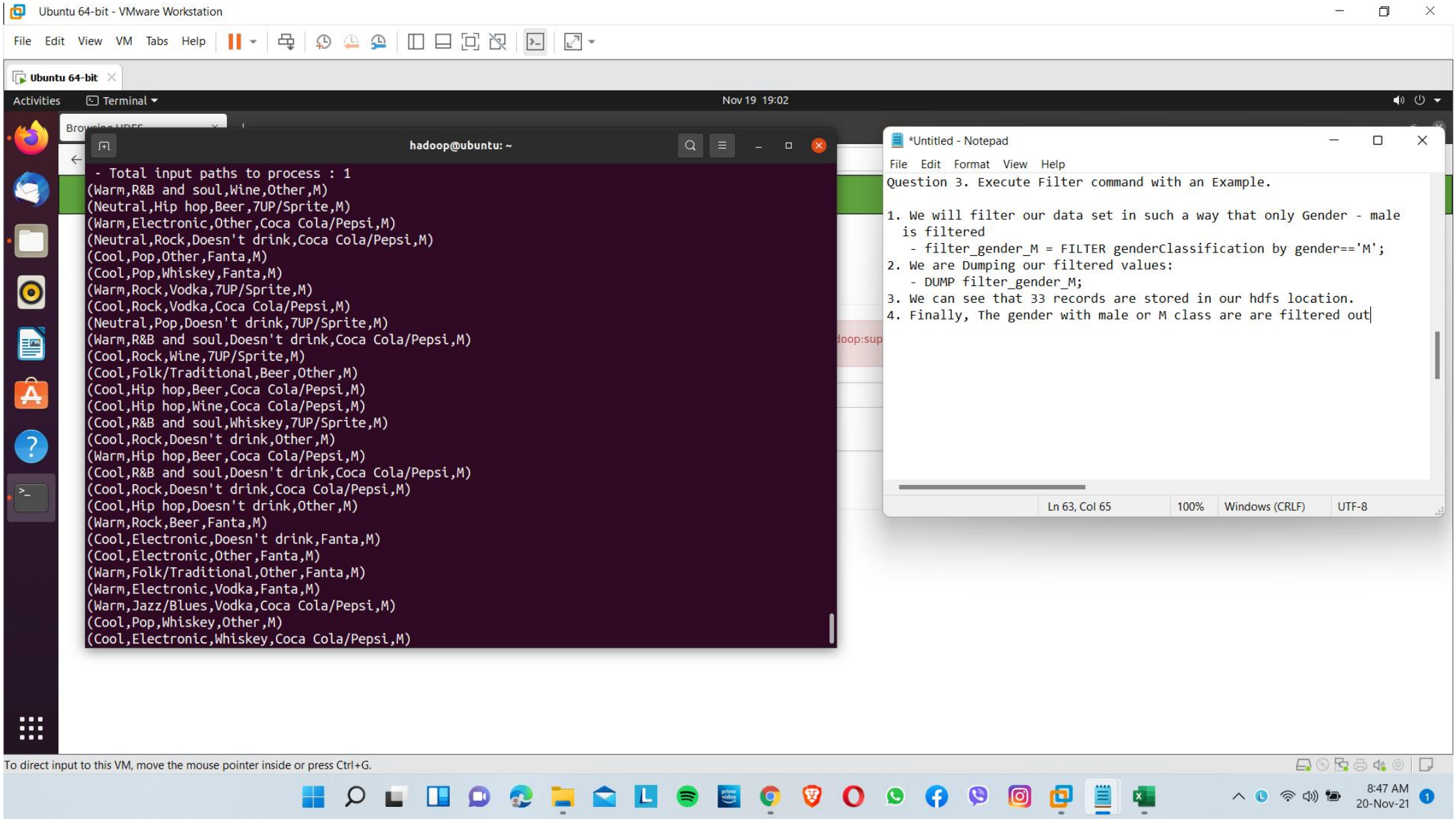

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 18:56
hadoop@ubuntu: ~
2021-11-19 18:56:04,866 [Thread-18] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 18:56:04,869 [Thread-18] INFO org.apache.hadoop.mapred.LocalJobRunner - OutputCommitter is org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigOutputCommitter
2021-11-19 18:56:04,897 [Thread-18] INFO org.apache.hadoop.mapred.LocalJobRunner - Waiting for map tasks
2021-11-19 18:56:04,897 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.LocalJobRunner - Starting task: attempt_local59974511_0001_m_000000_0
2021-11-19 18:56:04,947 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Using ResourceCalculatorProcessTree : [ ]
2021-11-19 18:56:04,950 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.MapTask - Processing split: Number of splits :1
Total Length = 2418
Input split[0]:
Length = 2418
ClassName: org.apache.hadoop.mapreduce.lib.input.FileSplit
Locations:
2021-11-19 18:56:04,964 [LocalJobRunner Map Task Executor #0] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigRecordReader - Current split being processed hdfs://localhost:8020/TestData/genderClassification.csv:0+2418
2021-11-19 18:56:05,001 [LocalJobRunner Map Task Executor #0] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 18:56:05,019 [LocalJobRunner Map Task Executor #0] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: genderClassification[1,23],genderClassification[-1,-1] C: R:
2021-11-19 18:56:05,048 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Local
*Untitled - Notepad
File Edit Format View Help
Question 2. Execute Dump command with an Example.
1. Executing DUMP command to print the genderClassification relation - DUMP genderClassification
2. We can see that the Number of splits is 1 and total length is 2148
3. We can see the hadoopVersion, pig version , userid, started time and finished time and our code is executed successfully.
4. Our output is listed
Ln 54, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:41 AM
20-Nov-21

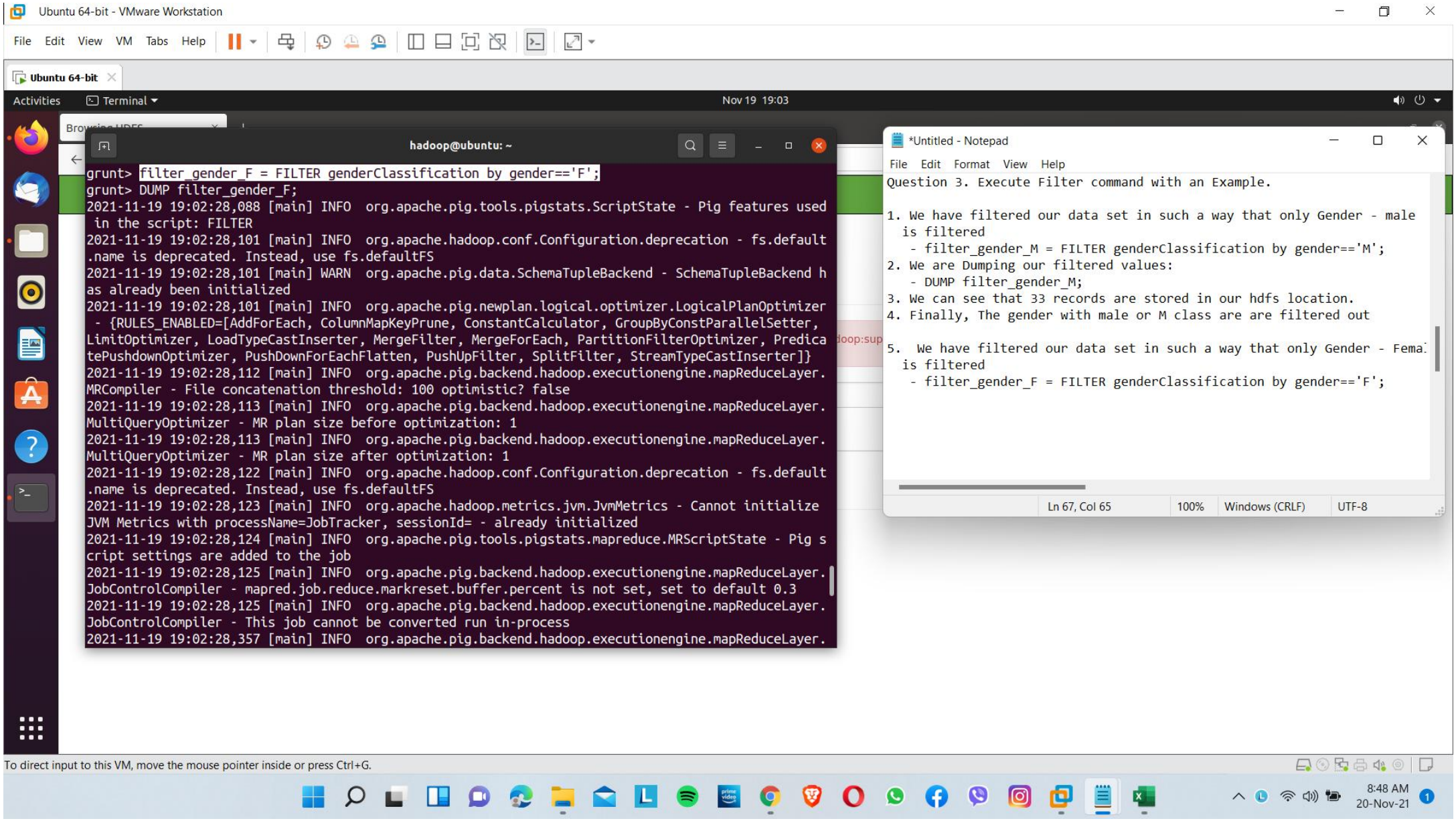

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 18:57
hadoop@ubuntu: ~
2021-11-19 18:56:09,884 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0 0.15.0 hadoop 2021-11-19 18:56:03 2021-11-19 18:56:09 UNKNOWN
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs
job_local59974511_0001 1 0 n/a n/a n/a n/a 0 0 0 0 genderClassification MAP_ONLY hdfs://localhost:8020/tmp/temp-331132261/tmp-879013517,
Input(s):
Successfully read 67 records (5252377 bytes) from: "/TestData/genderClassification.csv"
Output(s):
Successfully stored 67 records (5253250 bytes) in: "hdfs://localhost:8020/tmp/temp-331132261/tmp-879013517"
Counters:
Total records written : 67
Total bytes written : 5253250
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
*Untitled - Notepad
File Edit Format View Help
Question 2. Execute Dump command with an Example.
1. Executing DUMP command to print the genderClassification relation
- DUMP genderClassification
2. We can see that the Number of splits is 1 and total length is 2148
3. We can see the hadoopVersion, pig version , userid, started time and finished time and our code is executed successfully.
4. Our output is listed
Ln 54, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:42 AM
20-Nov-21

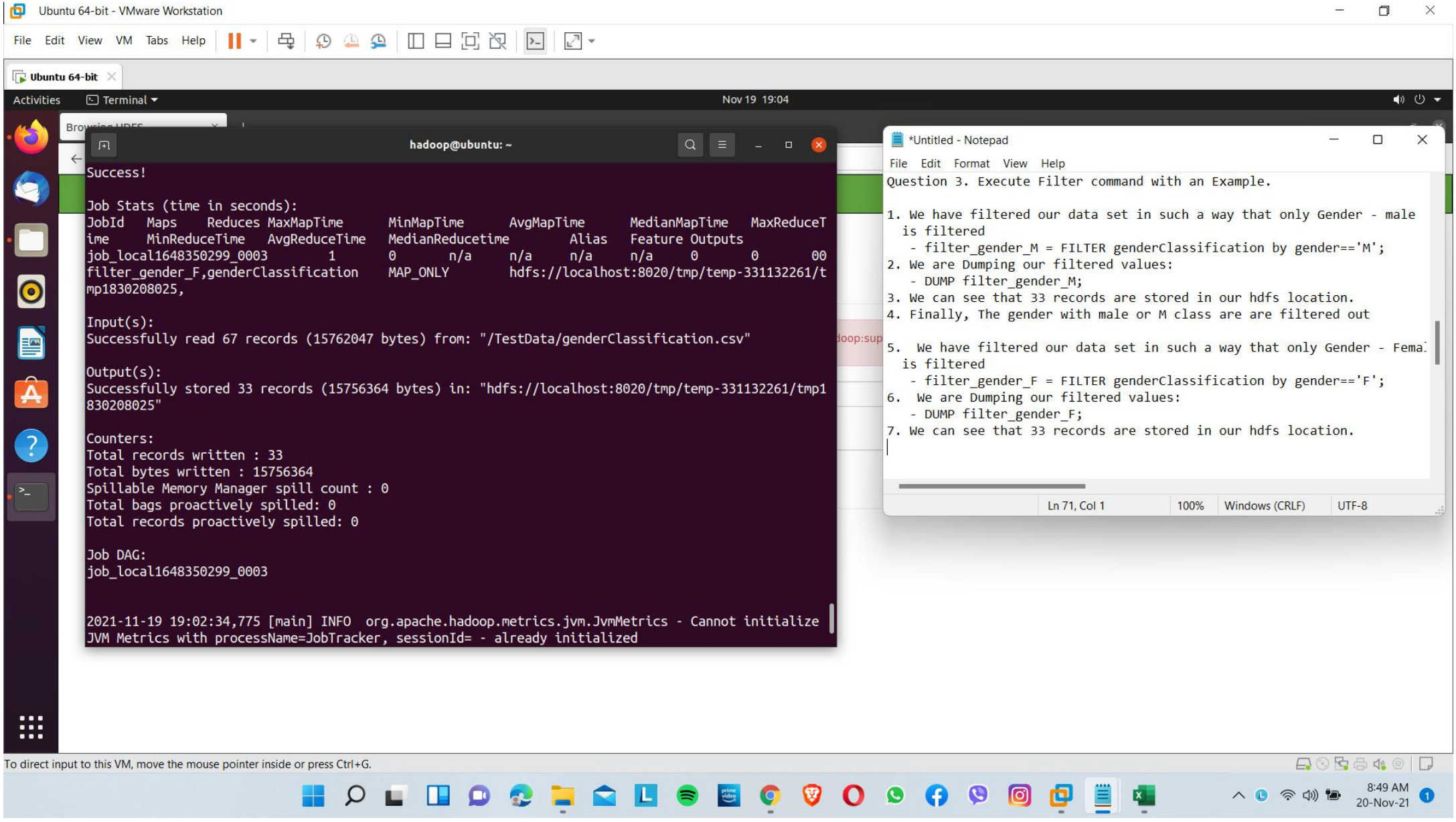

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 18:57
hadoop@ubuntu: ~
2021-11-19 18:56:09,894 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 18:56:09,899 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 18:56:09,899 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Favorite Color,Favorite Music Genre,Favorite Beverage,Favorite Soft Drink,Gender)
(Cool,Rock,Vodka,7UP/Sprite,F)
(Neutral,Hip hop,Vodka,Coca Cola/Pepsi,F)
(Warm,Rock,Wine,Coca Cola/Pepsi,F)
(Warm,Folk/Traditional,Whiskey,Fanta,F)
(Cool,Rock,Vodka,Coca Cola/Pepsi,F)
(Warm,Jazz/Blues,Doesn't drink,Fanta,F)
(Cool,Pop,Beer,Coca Cola/Pepsi,F)
(Warm,Pop,Whiskey,Fanta,F)
(Warm,Rock,Other,7UP/Sprite,F)
(Neutral,Pop,Wine,Coca Cola/Pepsi,F)
(Cool,Pop,Other,7UP/Sprite,F)
(Warm,Pop,Other,7UP/Sprite,F)
(Warm,Pop,Wine,7UP/Sprite,F)
(Warm,Electronic,Wine,Coca Cola/Pepsi,F)
(Cool,Rock,Beer,Coca Cola/Pepsi,F)
(Warm,Jazz/Blues,Wine,Coca Cola/Pepsi,F)
(Cool,Pop,Wine,7UP/Sprite,F)
(Cool,Rock,Other,Coca Cola/Pepsi,F)
(Cool,Rock,Other,Coca Cola/Pepsi,F)
(Cool,Pop,Doesn't drink,7UP/Sprite,F)
(Cool,Pop,Beer,Fanta,F)
(Warm,Jazz/Blues,Whiskey,Fanta,F)
*Untitled - Notepad
File Edit Format View Help
Question 2. Execute Dump command with an Example.
1. Executing DUMP command to print the genderClassification relation - DUMP genderClassification
2. We can see that the Number of splits is 1 and total length is 2148
3. We can see the hadoopVersion, pig version , userid, started time and finished time and our code is executed successfully.
4. Our output is listed
Ln 54, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:42 AM
20-Nov-21

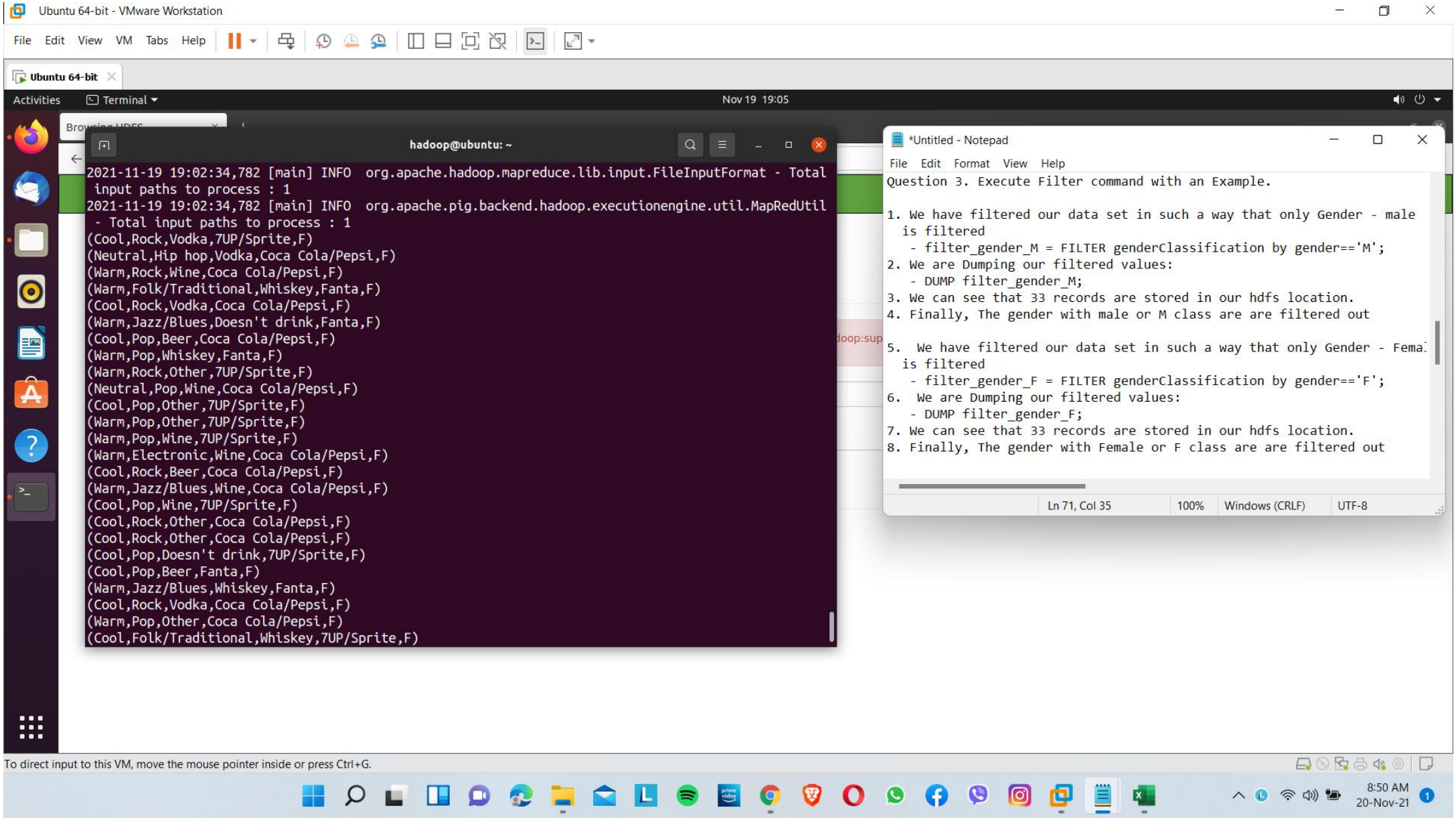

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:00
hadoop@ubuntu: ~
grunt> filter_gender_M = FILTER genderClassification by gender=='M';
grunt> DUMP filter_gender_M;
2021-11-19 18:59:05,992 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2021-11-19 18:59:06,004 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 18:59:06,005 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 18:59:06,005 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 18:59:06,010 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 18:59:06,012 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 18:59:06,012 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 18:59:06,021 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 18:59:06,021 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 18:59:06,022 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 18:59:06,022 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 18:59:06,023 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
2021-11-19 18:59:06,083 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 3. Execute Filter command with an Example.
1. We will filter our data set in such a way that only Gender - male is filtered
- filter_gender_M = FILTER genderClassification by gender=='M';
2. We are Dumping our filtered values:
- DUMP filter_gender_M;
Ln 61, Col 27
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:45 AM
20-Nov-21

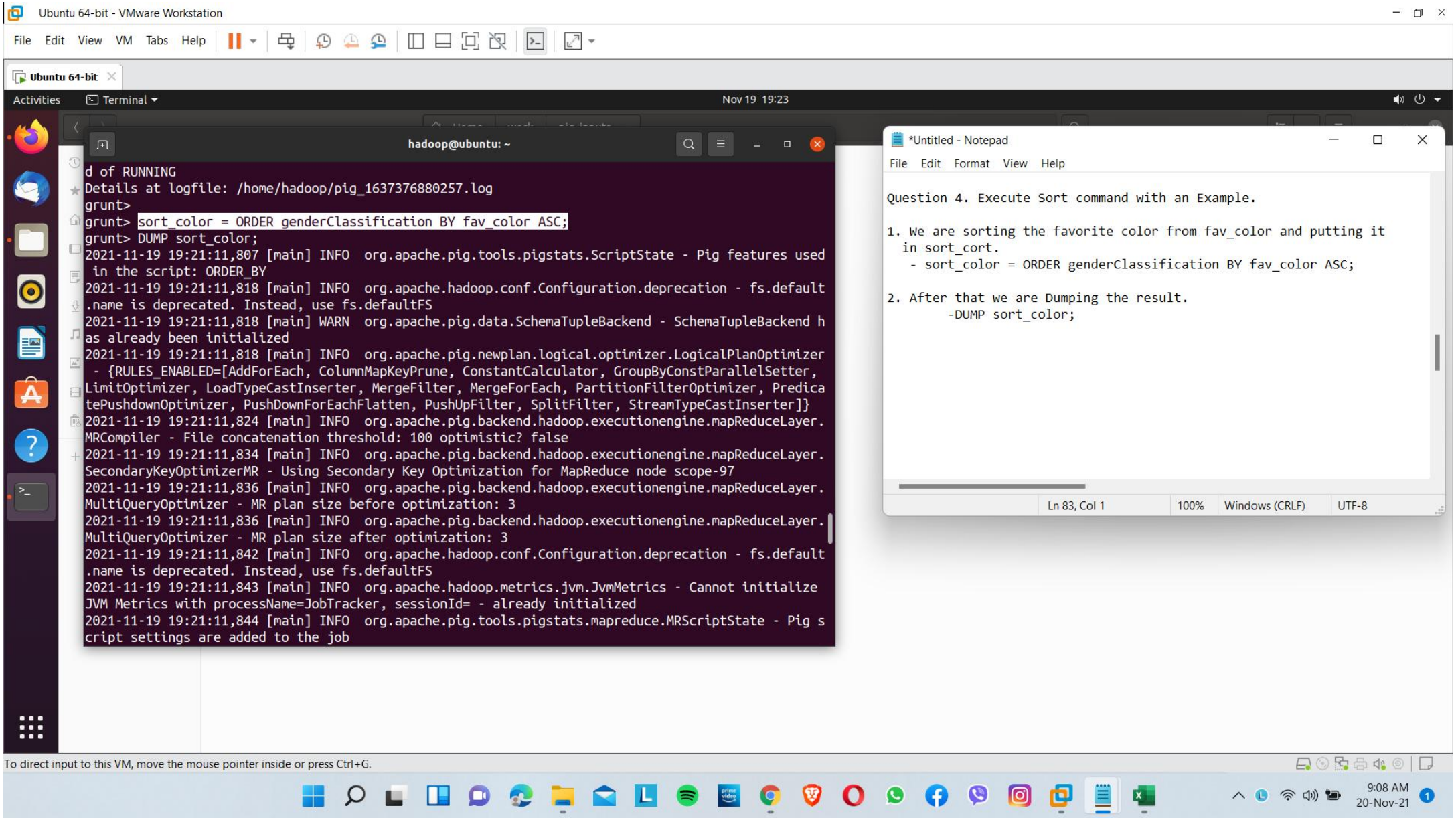

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:01
hadoop@ubuntu: ~
filter_gender_M,genderClassification MAP_ONLY hdfs://localhost:8020/tmp/temp-331132261/tmp-1904958683,
Input(s):
Successfully read 67 records (10508045 bytes) from: "/TestData/genderClassification.csv"
Output(s):
Successfully stored 33 records (10504834 bytes) in: "hdfs://localhost:8020/tmp/temp-331132261/tmp-1904958683"
Counters:
Total records written : 33
Total bytes written : 10504834
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1122225608_0002
2021-11-19 18:59:12,247 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 18:59:12,248 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 18:59:12,248 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 18:59:12,249 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
*Untitled - Notepad
File Edit Format View Help
Question 3. Execute Filter command with an Example.
1. We will filter our data set in such a way that only Gender - male is filtered
- filter_gender_M = FILTER genderClassification by gender=='M';
2. We are Dumping our filtered values:
- DUMP filter_gender_M;
3. We can see that 33 records are stored in our hdfs location.
Ln 63, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:46 AM
20-Nov-21

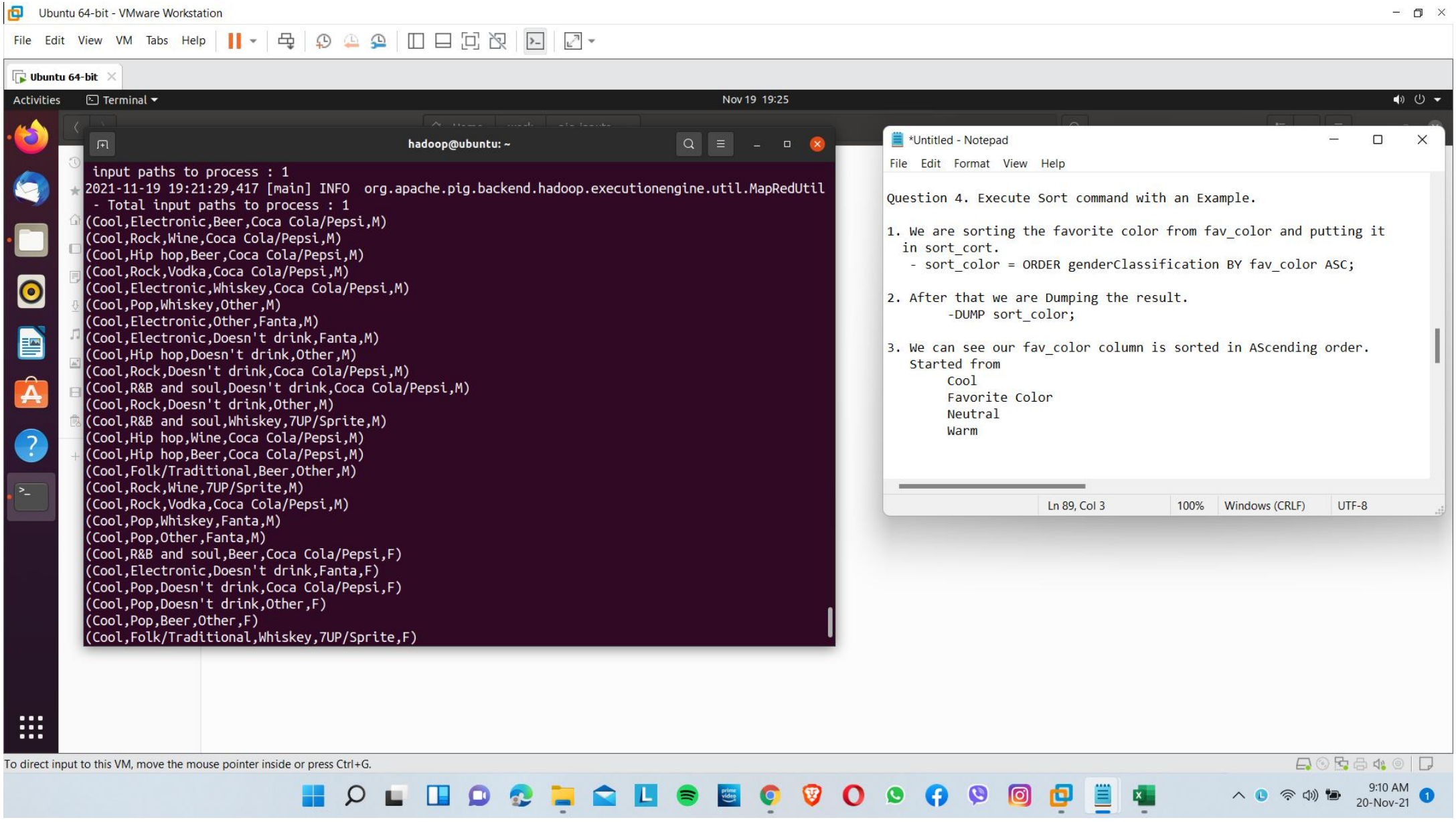

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:02
hadoop@ubuntu: ~
- Total input paths to process : 1
(Warm,R&B and soul,Wine,Other,M)
(Neutral,Hip hop,Beer,7UP/Sprite,M)
(Warm,Electronic,Other,Coca Cola/Pepsi,M)
(Neutral,Rock,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Pop,Other,Fanta,M)
(Cool,Pop,Whiskey,Fanta,M)
(Warm,Rock,Vodka,7UP/Sprite,M)
(Cool,Rock,Vodka,Coca Cola/Pepsi,M)
(Neutral,Pop,Doesn't drink,7UP/Sprite,M)
(Warm,R&B and soul,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Rock,Wine,7UP/Sprite,M)
(Cool,Folk/Traditional,Beer,Other,M)
(Cool,Hip hop,Beer,Coca Cola/Pepsi,M)
(Cool,Hip hop,Wine,Coca Cola/Pepsi,M)
(Cool,R&B and soul,Whiskey,7UP/Sprite,M)
(Cool,Rock,Doesn't drink,Other,M)
(Warm,Hip hop,Beer,Coca Cola/Pepsi,M)
(Cool,R&B and soul,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Rock,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Hip hop,Doesn't drink,Other,M)
(Warm,Rock,Beer,Fanta,M)
(Cool,Electronic,Doesn't drink,Fanta,M)
(Cool,Electronic,Other,Fanta,M)
(Warm,Folk/Traditional,Other,Fanta,M)
(Warm,Electronic,Vodka,Fanta,M)
(Warm,Jazz/Blues,Vodka,Coca Cola/Pepsi,M)
(Cool,Pop,Whiskey,Other,M)
(Cool,Electronic,Whiskey,Coca Cola/Pepsi,M)
*Untitled - Notepad
File Edit Format View Help
Question 3. Execute Filter command with an Example.
1. We will filter our data set in such a way that only Gender - male is filtered
- filter_gender_M = FILTER genderClassification by gender=='M';
2. We are Dumping our filtered values:
- DUMP filter_gender_M;
3. We can see that 33 records are stored in our hdfs location.
4. Finally, The gender with male or M class are are filtered out
Ln 63, Col 65
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:47 AM
20-Nov-21

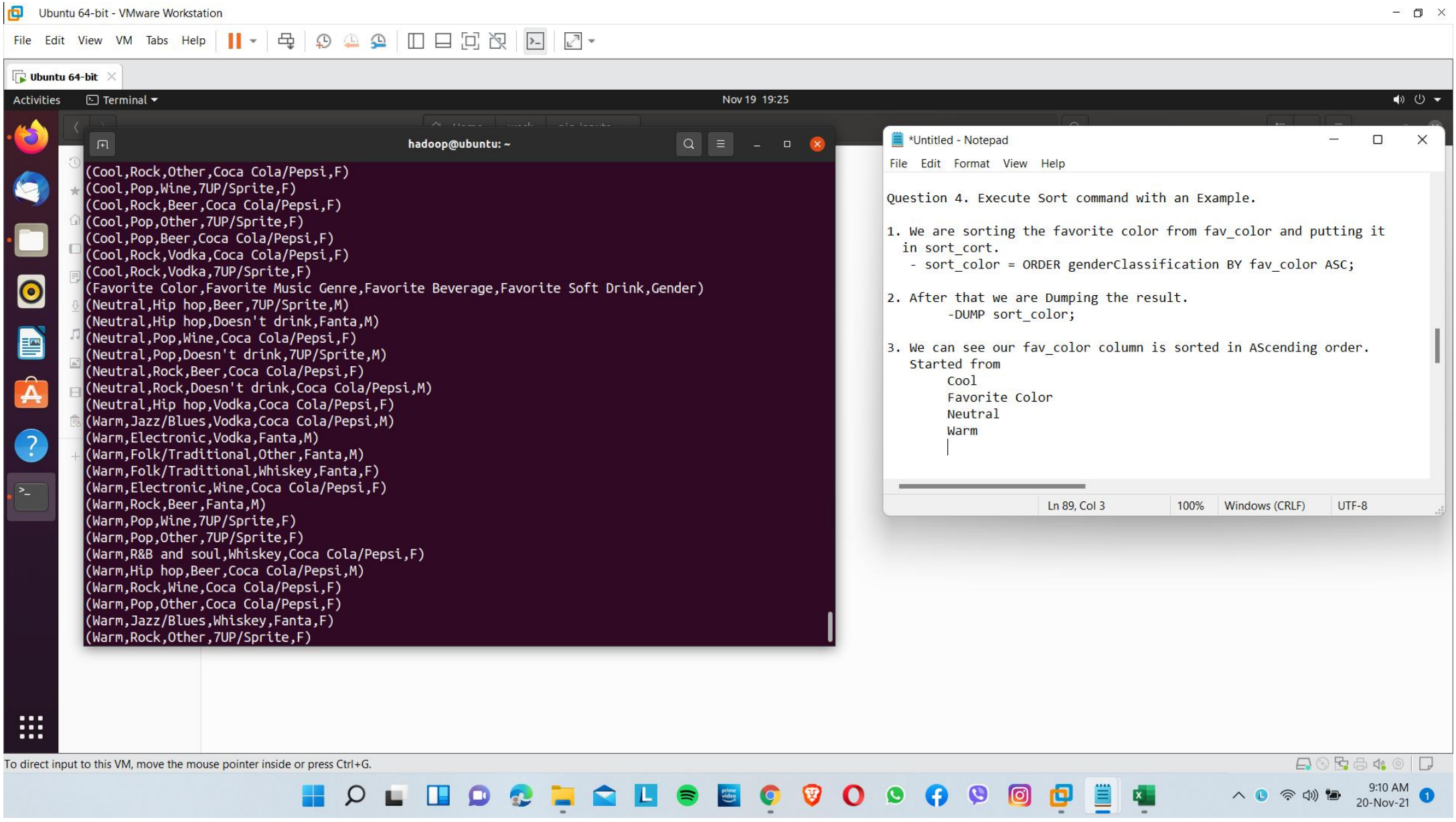

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:03
hadoop@ubuntu: ~
grunt> filter_gender_F = FILTER genderClassification by gender=='F';
grunt> DUMP filter_gender_F;
2021-11-19 19:02:28,088 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2021-11-19 19:02:28,101 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:02:28,101 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 19:02:28,101 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 19:02:28,112 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 19:02:28,113 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 19:02:28,113 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 19:02:28,122 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:02:28,123 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 19:02:28,124 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 19:02:28,125 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 19:02:28,125 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
2021-11-19 19:02:28,357 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 3. Execute Filter command with an Example.
1. We have filtered our data set in such a way that only Gender - male is filtered
- filter_gender_M = FILTER genderClassification by gender=='M';
2. We are Dumping our filtered values:
- DUMP filter_gender_M;
3. We can see that 33 records are stored in our hdfs location.
4. Finally, The gender with male or M class are are filtered out
5. We have filtered our data set in such a way that only Gender - Fema is filtered
- filter_gender_F = FILTER genderClassification by gender=='F';
Ln 67, Col 65
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:48 AM
20-Nov-21

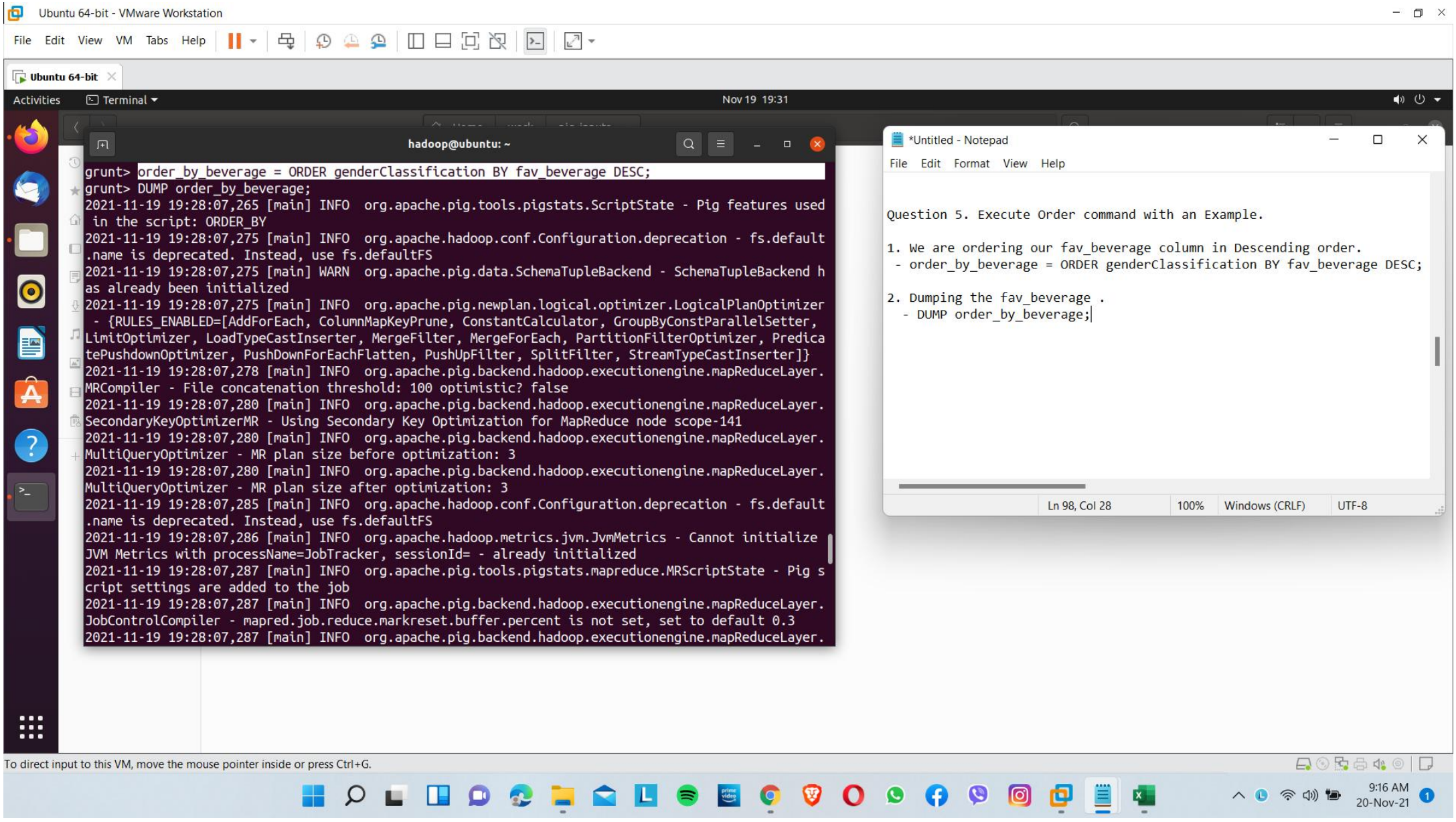

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:04
hadoop@ubuntu: ~
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs
job_local1648350299_0003 1 0 n/a n/a n/a n/a 0 0 00
filter_gender_F,genderClassification MAP_ONLY hdfs://localhost:8020/tmp/temp-331132261/tmp1830208025,
Input(s):
Successfully read 67 records (15762047 bytes) from: "/TestData/genderClassification.csv"
Output(s):
Successfully stored 33 records (15756364 bytes) in: "hdfs://localhost:8020/tmp/temp-331132261/tmp1830208025"
Counters:
Total records written : 33
Total bytes written : 15756364
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1648350299_0003
2021-11-19 19:02:34,775 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
*Untitled - Notepad
File Edit Format View Help
Question 3. Execute Filter command with an Example.
1. We have filtered our data set in such a way that only Gender - male is filtered
- filter_gender_M = FILTER genderClassification by gender=='M';
2. We are Dumping our filtered values:
- DUMP filter_gender_M;
3. We can see that 33 records are stored in our hdfs location.
4. Finally, The gender with male or M class are are filtered out
5. We have filtered our data set in such a way that only Gender - Fema is filtered
- filter_gender_F = FILTER genderClassification by gender=='F';
6. We are Dumping our filtered values:
- DUMP filter_gender_F;
7. We can see that 33 records are stored in our hdfs location.
Ln 71, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:49 AM
20-Nov-21

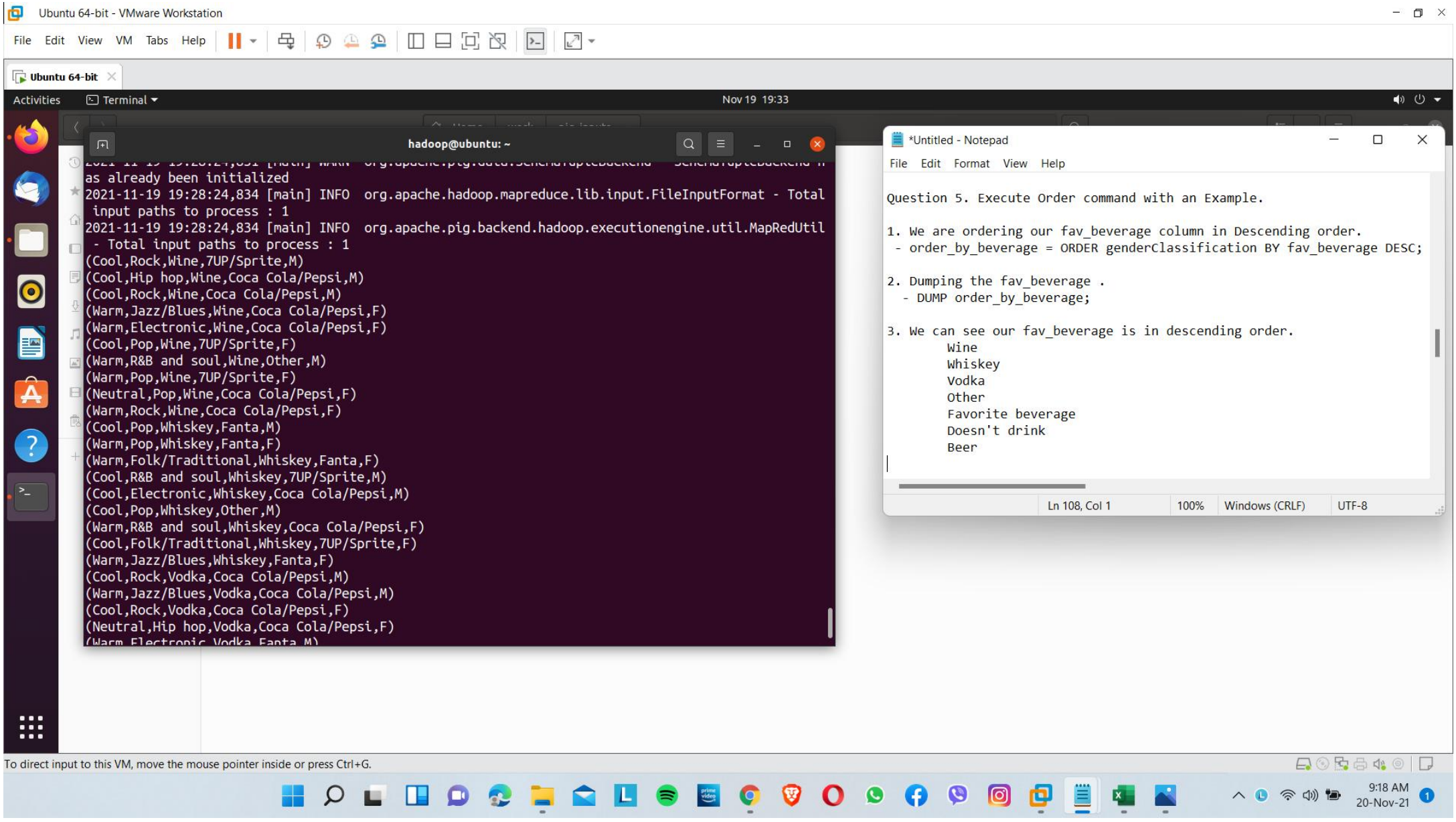

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:05
hadoop@ubuntu: ~
2021-11-19 19:02:34,782 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 19:02:34,782 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Cool,Rock,Vodka,7UP/Sprite,F)
(Neutral,Hip hop,Vodka,Coca Cola/Pepsi,F)
(Warm,Rock,Wine,Coca Cola/Pepsi,F)
(Warm,Folk/Traditional,Whiskey,Fanta,F)
(Cool,Rock,Vodka,Coca Cola/Pepsi,F)
(Warm,Jazz/Blues,Doesn't drink,Fanta,F)
(Cool,Pop,Beer,Coca Cola/Pepsi,F)
(Warm,Pop,Whiskey,Fanta,F)
(Warm,Rock,Other,7UP/Sprite,F)
(Neutral,Pop,Wine,Coca Cola/Pepsi,F)
(Cool,Pop,Other,7UP/Sprite,F)
(Warm,Pop,Other,7UP/Sprite,F)
(Warm,Pop,Wine,7UP/Sprite,F)
(Warm,Electronic,Wine,Coca Cola/Pepsi,F)
(Cool,Rock,Beer,Coca Cola/Pepsi,F)
(Warm,Jazz/Blues,Wine,Coca Cola/Pepsi,F)
(Cool,Pop,Wine,7UP/Sprite,F)
(Cool,Rock,Other,Coca Cola/Pepsi,F)
(Cool,Rock,Other,Coca Cola/Pepsi,F)
(Cool,Pop,Doesn't drink,7UP/Sprite,F)
(Cool,Pop,Beer,Fanta,F)
(Warm,Jazz/Blues,Whiskey,Fanta,F)
(Cool,Rock,Vodka,Coca Cola/Pepsi,F)
(Warm,Pop,Other,Coca Cola/Pepsi,F)
(Cool,Folk/Traditional,Whiskey,7UP/Sprite,F)
*Untitled - Notepad
File Edit Format View Help
Question 3. Execute Filter command with an Example.
1. We have filtered our data set in such a way that only Gender - male is filtered
- filter_gender_M = FILTER genderClassification by gender=='M';
2. We are Dumping our filtered values:
- DUMP filter_gender_M;
3. We can see that 33 records are stored in our hdfs location.
4. Finally, The gender with male or M class are are filtered out
5. We have filtered our data set in such a way that only Gender - Fema is filtered
- filter_gender_F = FILTER genderClassification by gender=='F';
6. We are Dumping our filtered values:
- DUMP filter_gender_F;
7. We can see that 33 records are stored in our hdfs location.
8. Finally, The gender with Female or F class are are filtered out
Ln 71, Col 35
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
8:50 AM
20-Nov-21

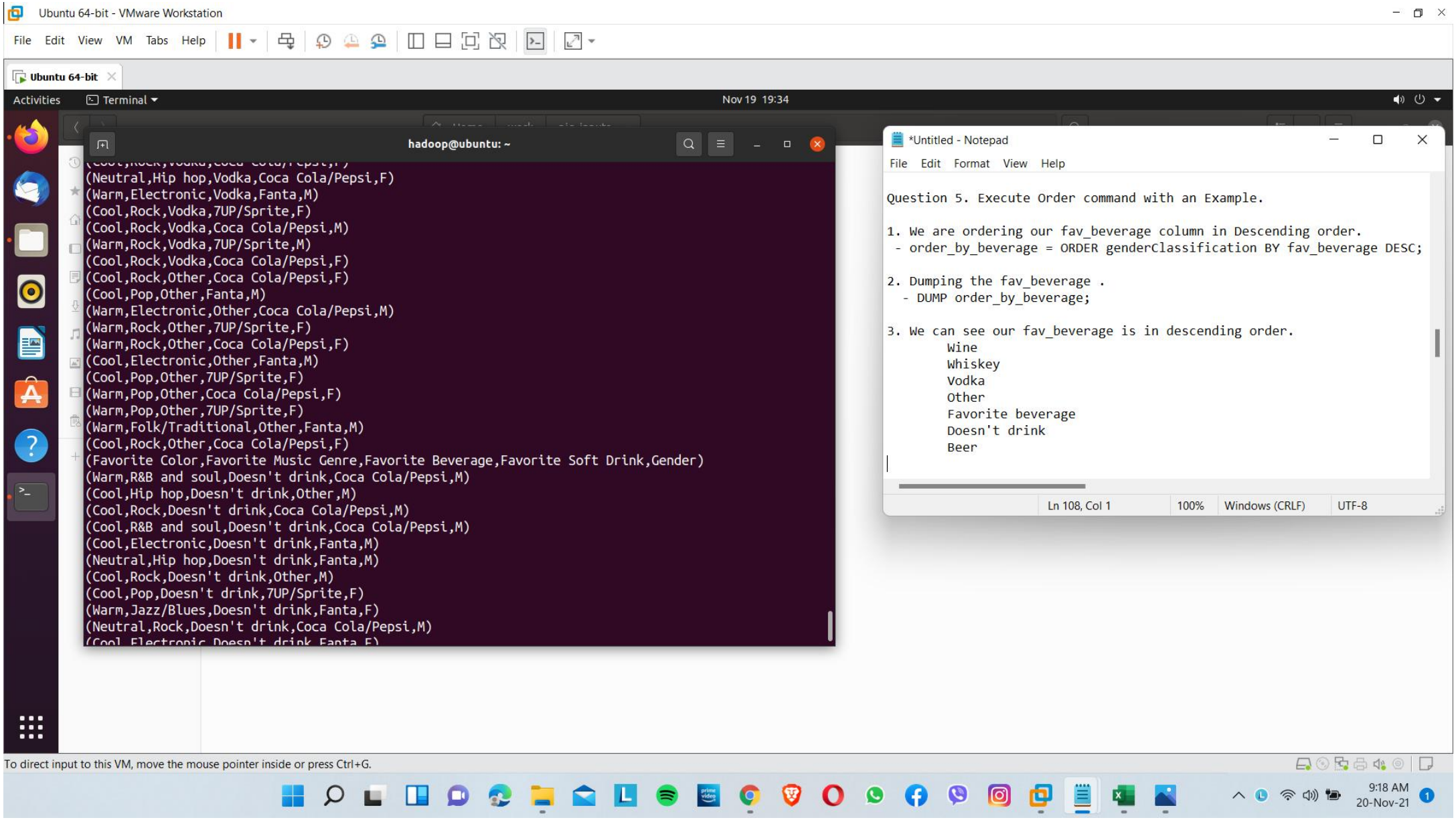

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:23
hadoop@ubuntu: ~
d of RUNNING
Details at logfile: /home/hadoop/pig_1637376880257.log
grunt>
grunt> sort_color = ORDER genderClassification BY fav_color ASC;
grunt> DUMP sort_color;
2021-11-19 19:21:11,807 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: ORDER_BY
2021-11-19 19:21:11,818 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:21:11,818 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 19:21:11,818 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 19:21:11,824 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 19:21:11,834 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope-97
2021-11-19 19:21:11,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 3
2021-11-19 19:21:11,836 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 3
2021-11-19 19:21:11,842 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:21:11,843 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 19:21:11,844 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
*Untitled - Notepad
File Edit Format View Help
Question 4. Execute Sort command with an Example.
1. We are sorting the favorite color from fav_color and putting it in sort_cort.
- sort_color = ORDER genderClassification BY fav_color ASC;
2. After that we are Dumping the result.
-DUMP sort_color;
Ln 83, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:08 AM
20-Nov-21

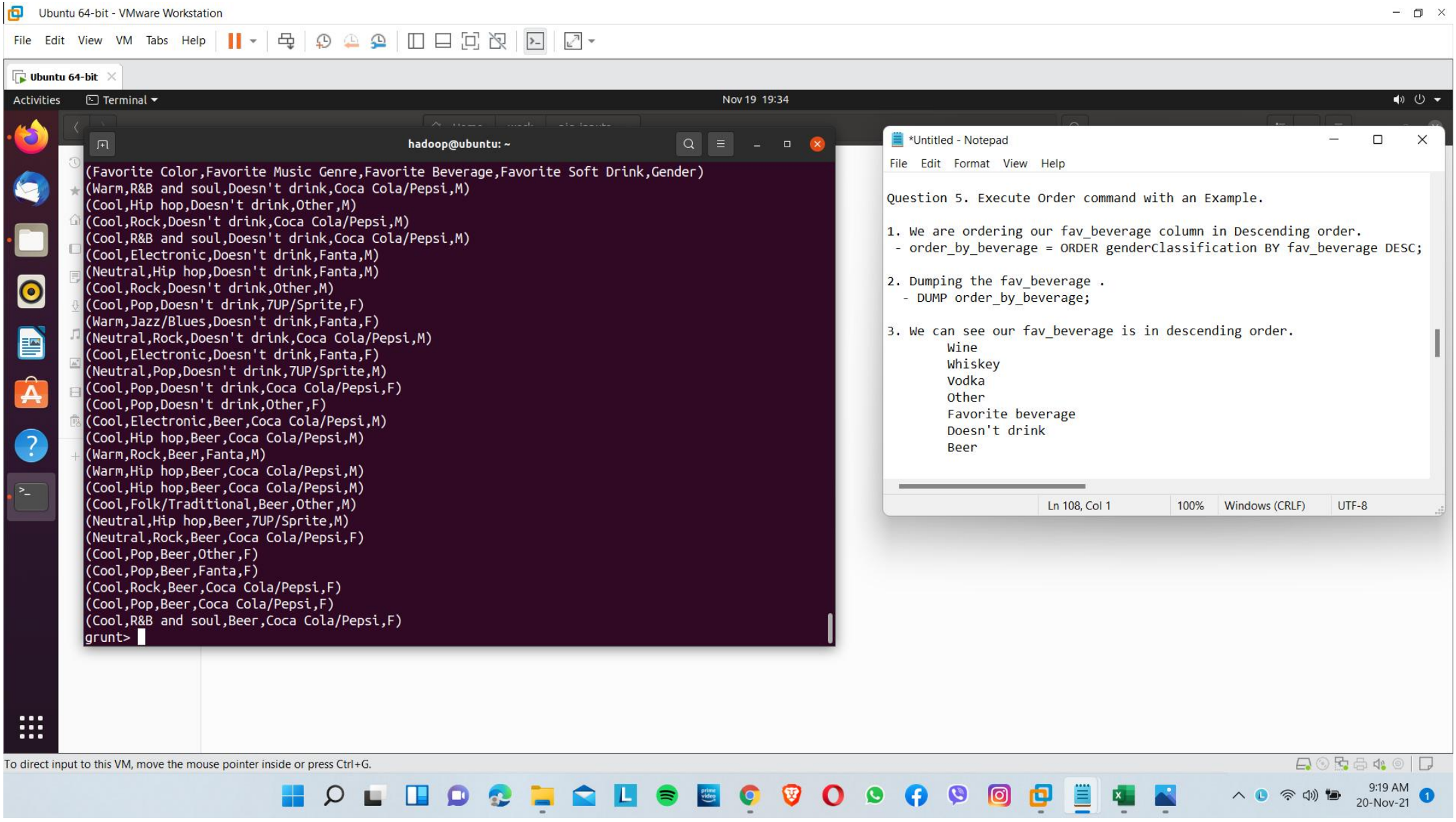

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:25
hadoop@ubuntu: ~
input paths to process : 1
2021-11-19 19:21:29,417 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil
- Total input paths to process : 1
(Cool,Electronic,Beer,Coca Cola/Pepsi,M)
(Cool,Rock,Wine,Coca Cola/Pepsi,M)
(Cool,Hip hop,Beer,Coca Cola/Pepsi,M)
(Cool,Rock,Vodka,Coca Cola/Pepsi,M)
(Cool,Electronic,Whiskey,Coca Cola/Pepsi,M)
(Cool,Pop,Whiskey,Other,M)
(Cool,Electronic,Other,Fanta,M)
(Cool,Electronic,Doesn't drink,Fanta,M)
(Cool,Hip hop,Doesn't drink,Other,M)
(Cool,Rock,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,R&B and soul,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Rock,Doesn't drink,Other,M)
(Cool,R&B and soul,Whiskey,7UP/Sprite,M)
(Cool,Hip hop,Wine,Coca Cola/Pepsi,M)
(Cool,Hip hop,Beer,Coca Cola/Pepsi,M)
(Cool,Folk/Traditional,Beer,Other,M)
(Cool,Rock,Wine,7UP/Sprite,M)
(Cool,Rock,Vodka,Coca Cola/Pepsi,M)
(Cool,Pop,Whiskey,Fanta,M)
(Cool,Pop,Other,Fanta,M)
(Cool,R&B and soul,Beer,Coca Cola/Pepsi,F)
(Cool,Electronic,Doesn't drink,Fanta,F)
(Cool,Pop,Doesn't drink,Coca Cola/Pepsi,F)
(Cool,Pop,Doesn't drink,Other,F)
(Cool,Pop,Beer,Other,F)
(Cool,Folk/Traditional,Whiskey,7UP/Sprite,F)
*Untitled - Notepad
File Edit Format View Help
Question 4. Execute Sort command with an Example.
1. We are sorting the favorite color from fav_color and putting it in sort_cort.
- sort_color = ORDER genderClassification BY fav_color ASC;
2. After that we are Dumping the result.
-DUMP sort_color;
3. We can see our fav_color column is sorted in AScending order.
Started from
Cool
Favorite Color
Neutral
Warm
Ln 89, Col 3
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:10 AM
20-Nov-21

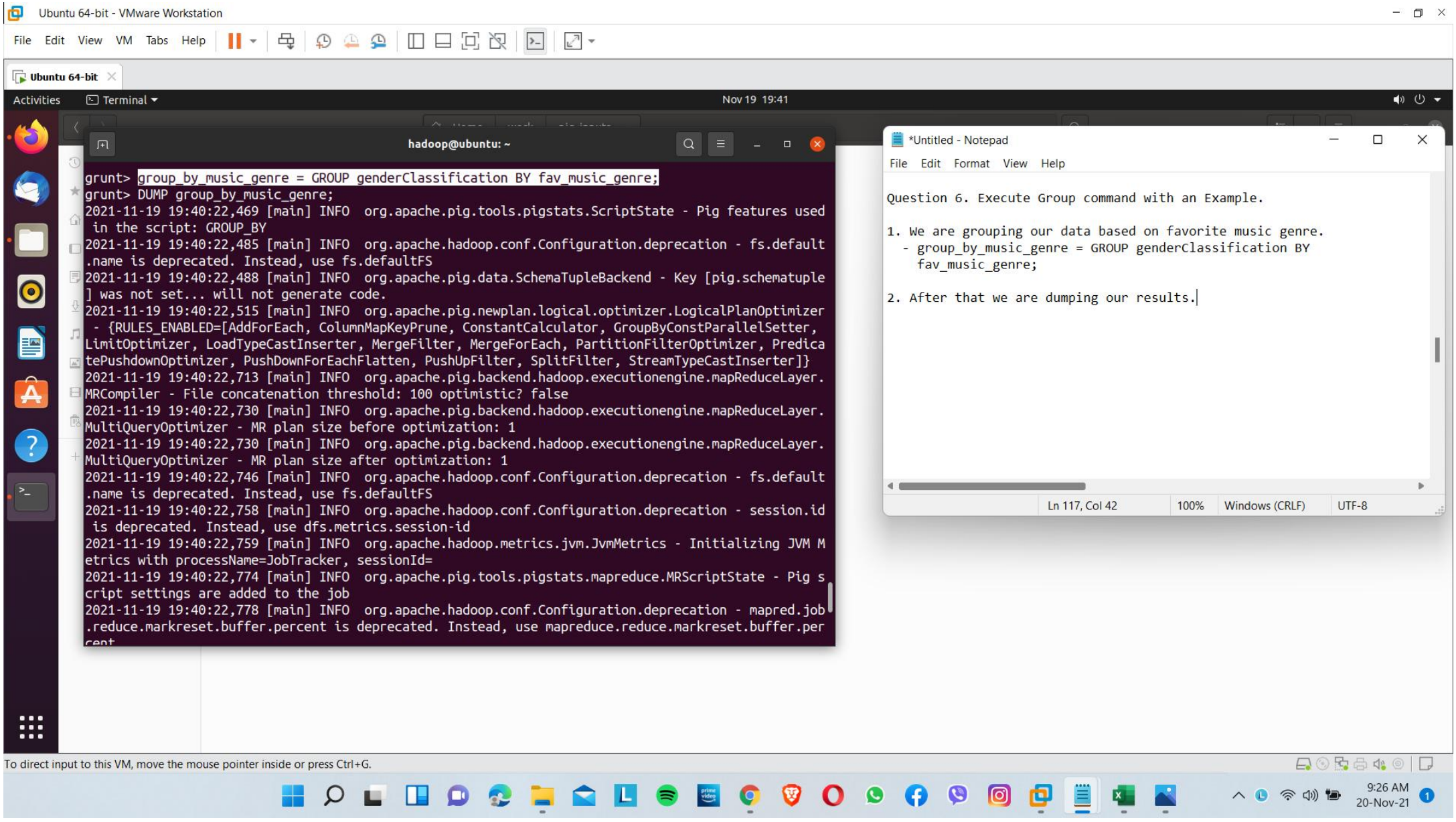

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:25
hadoop@ubuntu: ~
(Cool,Rock,Other,Coca Cola/Pepsi,F)
(Cool,Pop,Wine,7UP/Sprite,F)
(Cool,Rock,Beer,Coca Cola/Pepsi,F)
(Cool,Pop,Other,7UP/Sprite,F)
(Cool,Pop,Beer,Coca Cola/Pepsi,F)
(Cool,Rock,Vodka,Coca Cola/Pepsi,F)
(Cool,Rock,Vodka,7UP/Sprite,F)
(Favorite Color,Favorite Music Genre,Favorite Beverage,Favorite Soft Drink,Gender)
(Neutral,Hip hop,Beer,7UP/Sprite,M)
(Neutral,Hip hop,Doesn't drink,Fanta,M)
(Neutral,Pop,Wine,Coca Cola/Pepsi,F)
(Neutral,Pop,Doesn't drink,7UP/Sprite,M)
(Neutral,Rock,Beer,Coca Cola/Pepsi,F)
(Neutral,Rock,Doesn't drink,Coca Cola/Pepsi,M)
(Neutral,Hip hop,Vodka,Coca Cola/Pepsi,F)
(Warm,Jazz/Blues,Vodka,Coca Cola/Pepsi,M)
(Warm,Electronic,Vodka,Fanta,M)
(Warm,Folk/Traditional,Other,Fanta,M)
(Warm,Folk/Traditional,Whiskey,Fanta,F)
(Warm,Electronic,Wine,Coca Cola/Pepsi,F)
(Warm,Rock,Beer,Fanta,M)
(Warm,Pop,Wine,7UP/Sprite,F)
(Warm,Pop,Other,7UP/Sprite,F)
(Warm,R&B and soul,Whiskey,Coca Cola/Pepsi,F)
(Warm,Hip hop,Beer,Coca Cola/Pepsi,M)
(Warm,Rock,Wine,Coca Cola/Pepsi,F)
(Warm,Pop,Other,Coca Cola/Pepsi,F)
(Warm,Jazz/Blues,Whiskey,Fanta,F)
(Warm,Rock,Other,7UP/Sprite,F)
*Untitled - Notepad
File Edit Format View Help
Question 4. Execute Sort command with an Example.
1. We are sorting the favorite color from fav_color and putting it in sort_cort.
- sort_color = ORDER genderClassification BY fav_color ASC;
2. After that we are Dumping the result.
-DUMP sort_color;
3. We can see our fav_color column is sorted in AScending order.
Started from
Cool
Favorite Color
Neutral
Warm
Ln 89, Col 3
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:10 AM
20-Nov-21

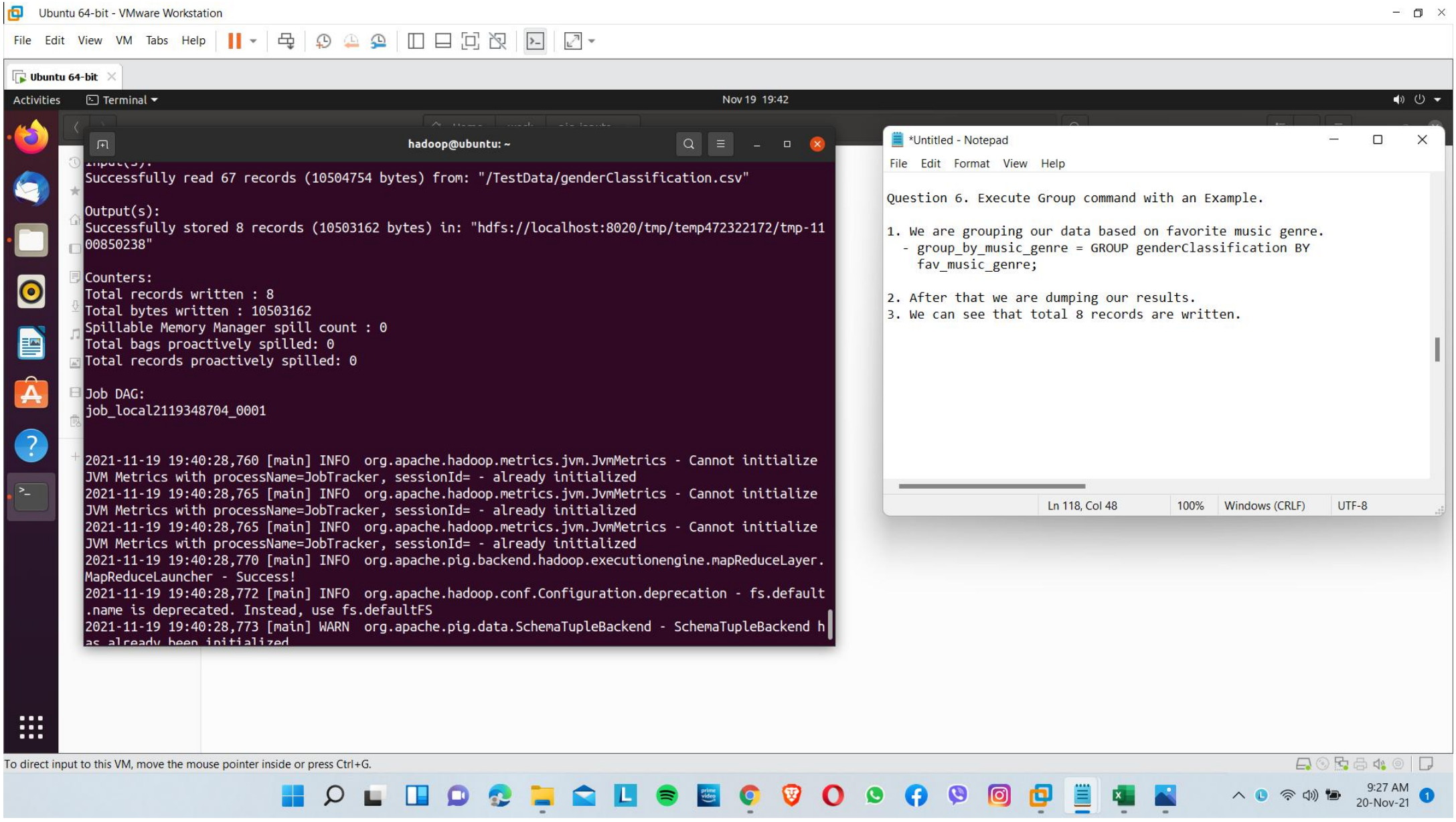

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:31
hadoop@ubuntu: ~
grunt> order_by_beverage = ORDER genderClassification BY fav_beverage DESC;
grunt> DUMP order_by_beverage;
2021-11-19 19:28:07,265 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: ORDER_BY
2021-11-19 19:28:07,275 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:28:07,275 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 19:28:07,275 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 19:28:07,278 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 19:28:07,280 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.SecondaryKeyOptimizerMR - Using Secondary Key Optimization for MapReduce node scope-141
2021-11-19 19:28:07,280 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 3
2021-11-19 19:28:07,280 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 3
2021-11-19 19:28:07,285 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:28:07,286 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 19:28:07,287 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 19:28:07,287 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 19:28:07,287 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 5. Execute Order command with an Example.
1. We are ordering our fav_beverage column in Descending order.
- order_by_beverage = ORDER genderClassification BY fav_beverage DESC;
2. Dumping the fav_beverage .
- DUMP order_by_beverage;
Ln 98, Col 28
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:16 AM
20-Nov-21

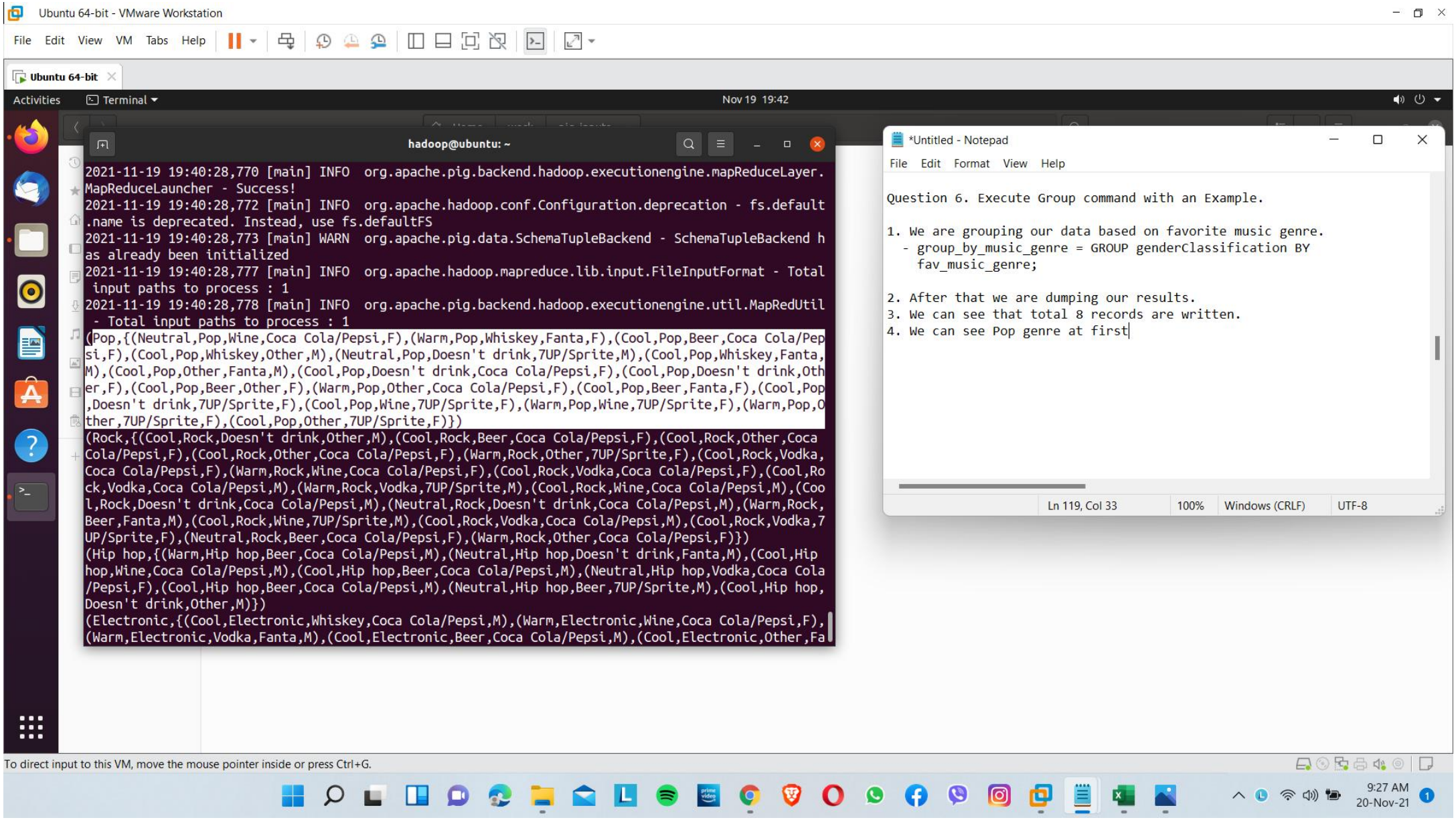

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:33
hadoop@ubuntu: ~
as already been initialized
2021-11-19 19:28:24,834 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 19:28:24,834 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Cool,Rock,Wine,7UP/Sprite,M)
(Cool,Hip hop,Wine,Coca Cola/Pepsi,M)
(Cool,Rock,Wine,Coca Cola/Pepsi,M)
(Warm,Jazz/Blues,Wine,Coca Cola/Pepsi,F)
(Warm,Electronic,Wine,Coca Cola/Pepsi,F)
(Cool,Pop,Wine,7UP/Sprite,F)
(Warm,R&B and soul,Wine,Other,M)
(Warm,Pop,Wine,7UP/Sprite,F)
(Neutral,Pop,Wine,Coca Cola/Pepsi,F)
(Warm,Rock,Wine,Coca Cola/Pepsi,F)
(Cool,Pop,Whiskey,Fanta,M)
(Warm,Pop,Whiskey,Fanta,F)
(Warm,Folk/Traditional,Whiskey,Fanta,F)
(Cool,R&B and soul,Whiskey,7UP/Sprite,M)
(Cool,Electronic,Whiskey,Coca Cola/Pepsi,M)
(Cool,Pop,Whiskey,Other,M)
(Warm,R&B and soul,Whiskey,Coca Cola/Pepsi,F)
(Cool,Folk/Traditional,Whiskey,7UP/Sprite,F)
(Warm,Jazz/Blues,Whiskey,Fanta,F)
(Cool,Rock,Vodka,Coca Cola/Pepsi,M)
(Warm,Jazz/Blues,Vodka,Coca Cola/Pepsi,M)
(Cool,Rock,Vodka,Coca Cola/Pepsi,F)
(Neutral,Hip hop,Vodka,Coca Cola/Pepsi,F)
*Untitled - Notepad
File Edit Format View Help
Question 5. Execute Order command with an Example.
1. We are ordering our fav_beverage column in Descending order.
- order_by_beverage = ORDER genderClassification BY fav_beverage DESC;
2. Dumping the fav_beverage .
- DUMP order_by_beverage;
3. We can see our fav_beverage is in descending order.
Wine
Whiskey
Vodka
Other
Favorite beverage
Doesn't drink
Beer
Ln 108, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:18 AM
20-Nov-21

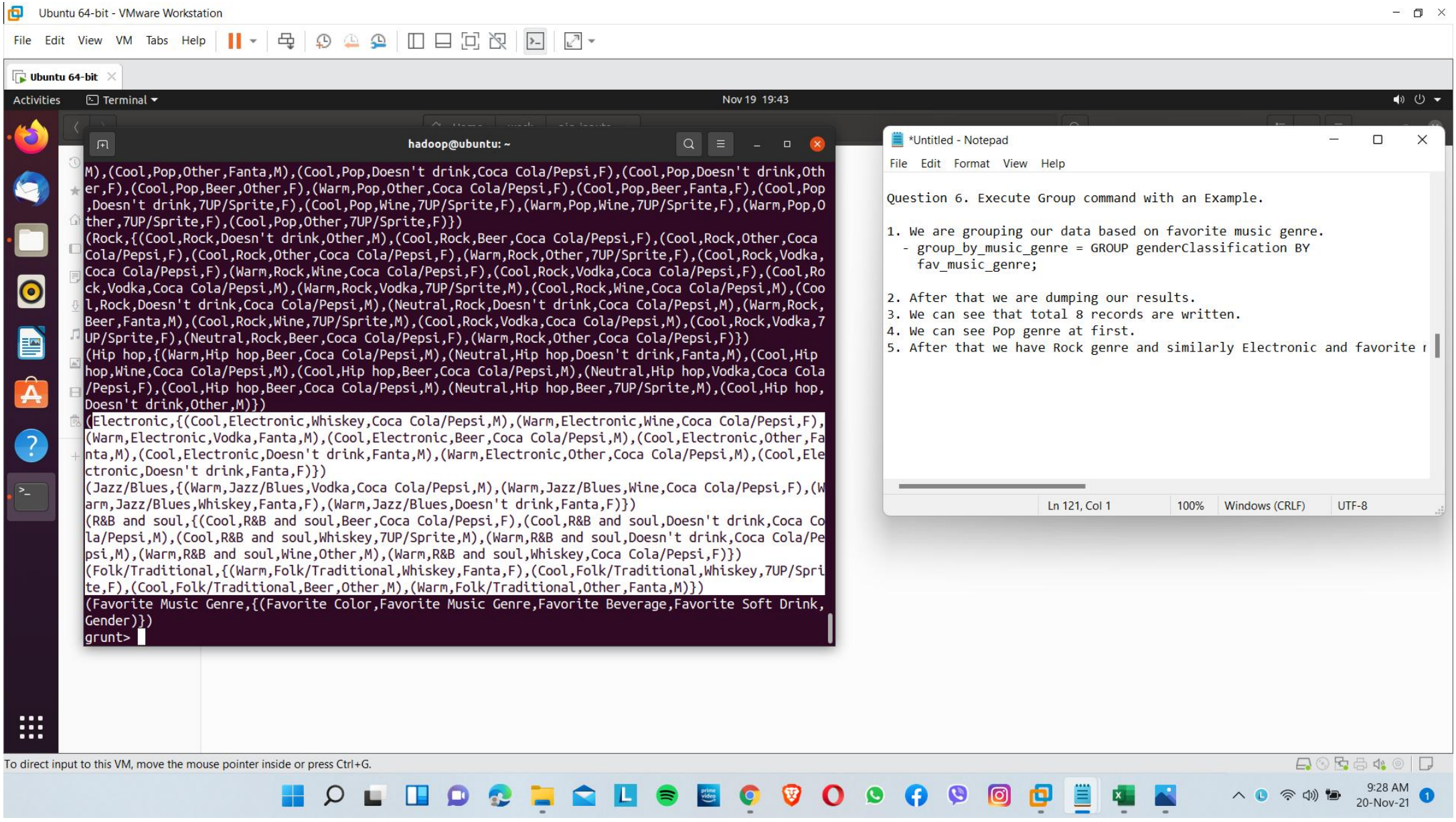

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:34
hadoop@ubuntu: ~
(Neutral,Hip hop,Vodka,Coca Cola/Pepsi,F)
(Warm,Electronic,Vodka,Fanta,M)
(Cool,Rock,Vodka,7UP/Sprite,F)
(Cool,Rock,Vodka,Coca Cola/Pepsi,M)
(Warm,Rock,Vodka,7UP/Sprite,M)
(Cool,Rock,Vodka,Coca Cola/Pepsi,F)
(Cool,Rock,Other,Coca Cola/Pepsi,F)
(Cool,Pop,Other,Fanta,M)
(Warm,Electronic,Other,Coca Cola/Pepsi,M)
(Warm,Rock,Other,7UP/Sprite,F)
(Warm,Rock,Other,Coca Cola/Pepsi,F)
(Cool,Electronic,Other,Fanta,M)
(Cool,Pop,Other,7UP/Sprite,F)
(Warm,Pop,Other,Coca Cola/Pepsi,F)
(Warm,Pop,Other,7UP/Sprite,F)
(Warm,Folk/Traditional,Other,Fanta,M)
(Cool,Rock,Other,Coca Cola/Pepsi,F)
(Favorite Color,Favorite Music Genre,Favorite Beverage,Favorite Soft Drink,Gender)
(Warm,R&B and soul,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Hip hop,Doesn't drink,Other,M)
(Cool,Rock,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,R&B and soul,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Electronic,Doesn't drink,Fanta,M)
(Neutral,Hip hop,Doesn't drink,Fanta,M)
(Cool,Rock,Doesn't drink,Other,M)
(Cool,Pop,Doesn't drink,7UP/Sprite,F)
(Warm,Jazz/Blues,Doesn't drink,Fanta,F)
(Neutral,Rock,Doesn't drink,Coca Cola/Pepsi,M)
*Untitled - Notepad
File Edit Format View Help
Question 5. Execute Order command with an Example.
1. We are ordering our fav_beverage column in Descending order.
- order_by_beverage = ORDER genderClassification BY fav_beverage DESC;
2. Dumping the fav_beverage .
- DUMP order_by_beverage;
3. We can see our fav_beverage is in descending order.
Wine
Whiskey
Vodka
Other
Favorite beverage
Doesn't drink
Beer
Ln 108, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:18 AM
20-Nov-21

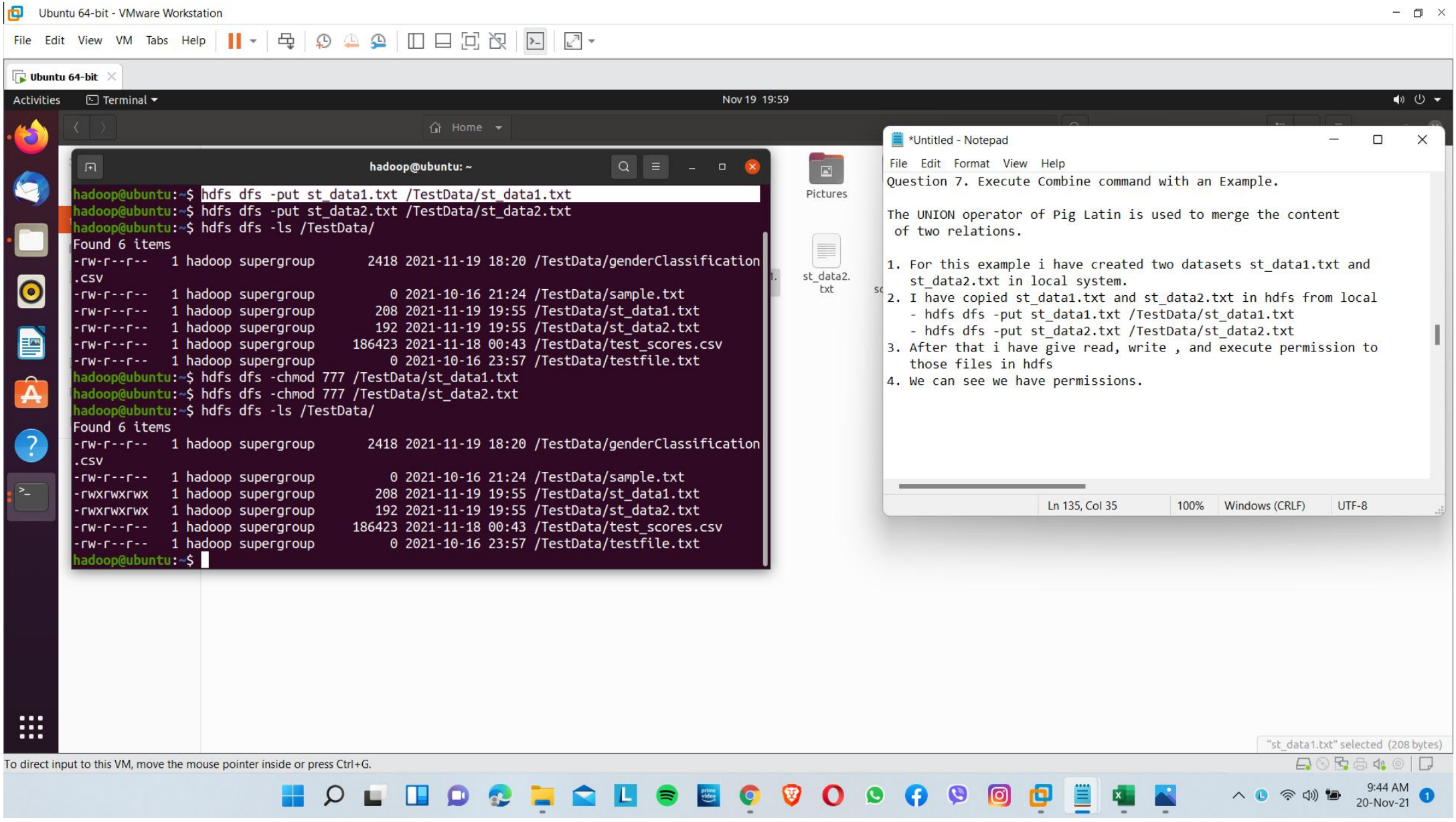

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:34
hadoop@ubuntu: ~
(Favorite Color,Favorite Music Genre,Favorite Beverage,Favorite Soft Drink,Gender)
(Warm,R&B and soul,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Hip hop,Doesn't drink,Other,M)
(Cool,Rock,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,R&B and soul,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Electronic,Doesn't drink,Fanta,M)
(Neutral,Hip hop,Doesn't drink,Fanta,M)
(Cool,Rock,Doesn't drink,Other,M)
(Cool,Pop,Doesn't drink,7UP/Sprite,F)
(Warm,Jazz/Blues,Doesn't drink,Fanta,F)
(Neutral,Rock,Doesn't drink,Coca Cola/Pepsi,M)
(Cool,Electronic,Doesn't drink,Fanta,F)
(Neutral,Pop,Doesn't drink,7UP/Sprite,M)
(Cool,Pop,Doesn't drink,Coca Cola/Pepsi,F)
(Cool,Pop,Doesn't drink,Other,F)
(Cool,Electronic,Beer,Coca Cola/Pepsi,M)
(Cool,Hip hop,Beer,Coca Cola/Pepsi,M)
(Warm,Rock,Beer,Fanta,M)
(Warm,Hip hop,Beer,Coca Cola/Pepsi,M)
(Cool,Hip hop,Beer,Coca Cola/Pepsi,M)
(Cool,Folk/Traditional,Beer,Other,M)
(Neutral,Hip hop,Beer,7UP/Sprite,M)
(Neutral,Rock,Beer,Coca Cola/Pepsi,F)
(Cool,Pop,Beer,Other,F)
(Cool,Pop,Beer,Fanta,F)
(Cool,Rock,Beer,Coca Cola/Pepsi,F)
(Cool,Pop,Beer,Coca Cola/Pepsi,F)
(Cool,R&B and soul,Beer,Coca Cola/Pepsi,F)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 5. Execute Order command with an Example.
1. We are ordering our fav_beverage column in Descending order.
- order_by_beverage = ORDER genderClassification BY fav_beverage DESC;
2. Dumping the fav_beverage .
- DUMP order_by_beverage;
3. We can see our fav_beverage is in descending order.
Wine
Whiskey
Vodka
Other
Favorite beverage
Doesn't drink
Beer
Ln 108, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:19 AM
20-Nov-21

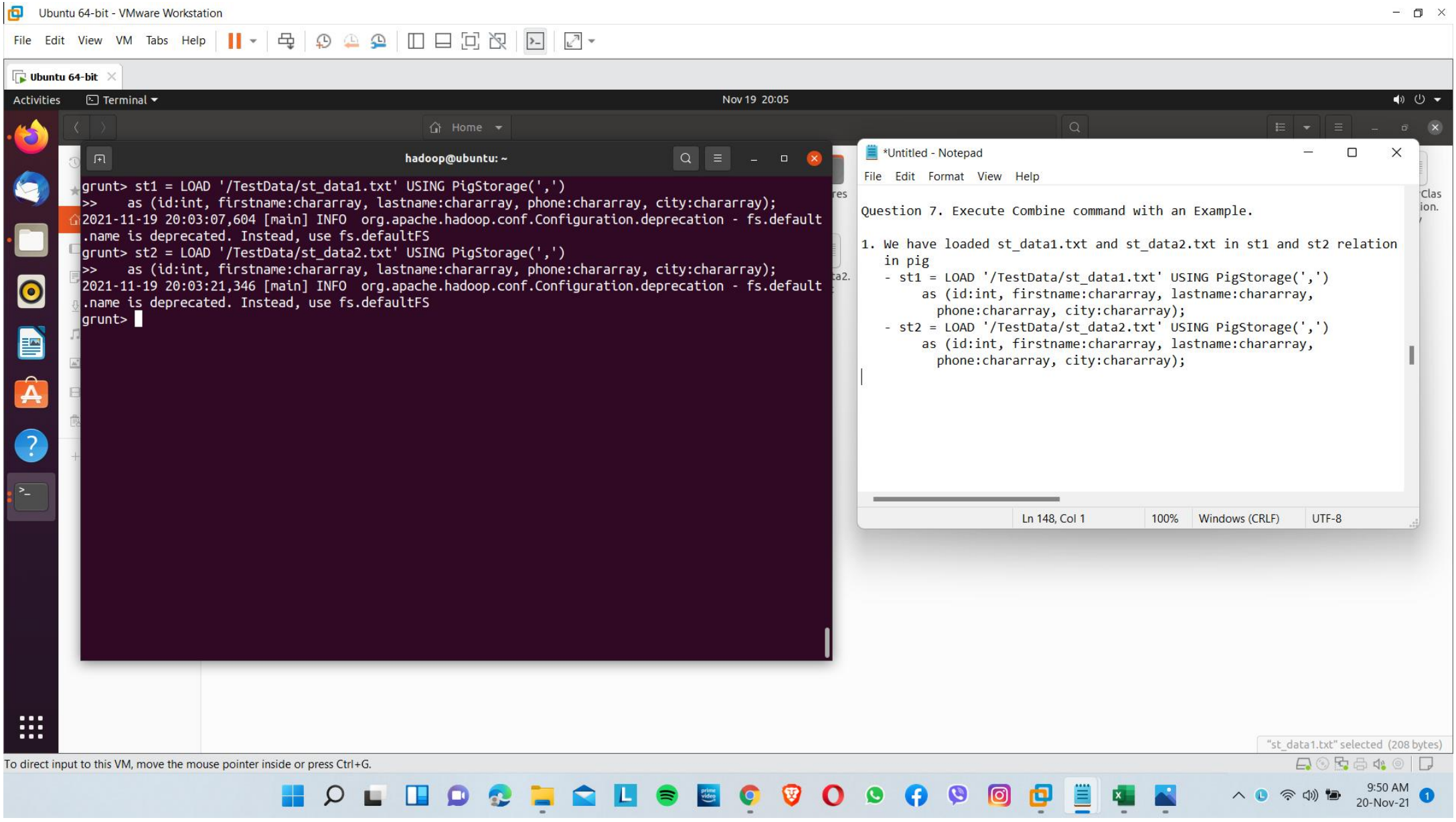

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:41
hadoop@ubuntu: ~
grunt> group_by_music_genre = GROUP genderClassification BY fav_music_genre;
grunt> DUMP group_by_music_genre;
2021-11-19 19:40:22,469 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY
2021-11-19 19:40:22,485 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:40:22,488 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 19:40:22,515 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 19:40:22,713 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 19:40:22,730 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 19:40:22,730 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 19:40:22,746 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:40:22,758 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is deprecated. Instead, use dfs.metrics.session-id
2021-11-19 19:40:22,759 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metrics with processName=JobTracker, sessionId=
2021-11-19 19:40:22,774 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 19:40:22,778 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
*Untitled - Notepad
File Edit Format View Help
Question 6. Execute Group command with an Example.
1. We are grouping our data based on favorite music genre.
- group_by_music_genre = GROUP genderClassification BY fav_music_genre;
2. After that we are dumping our results.
Ln 117, Col 42
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:26 AM
20-Nov-21

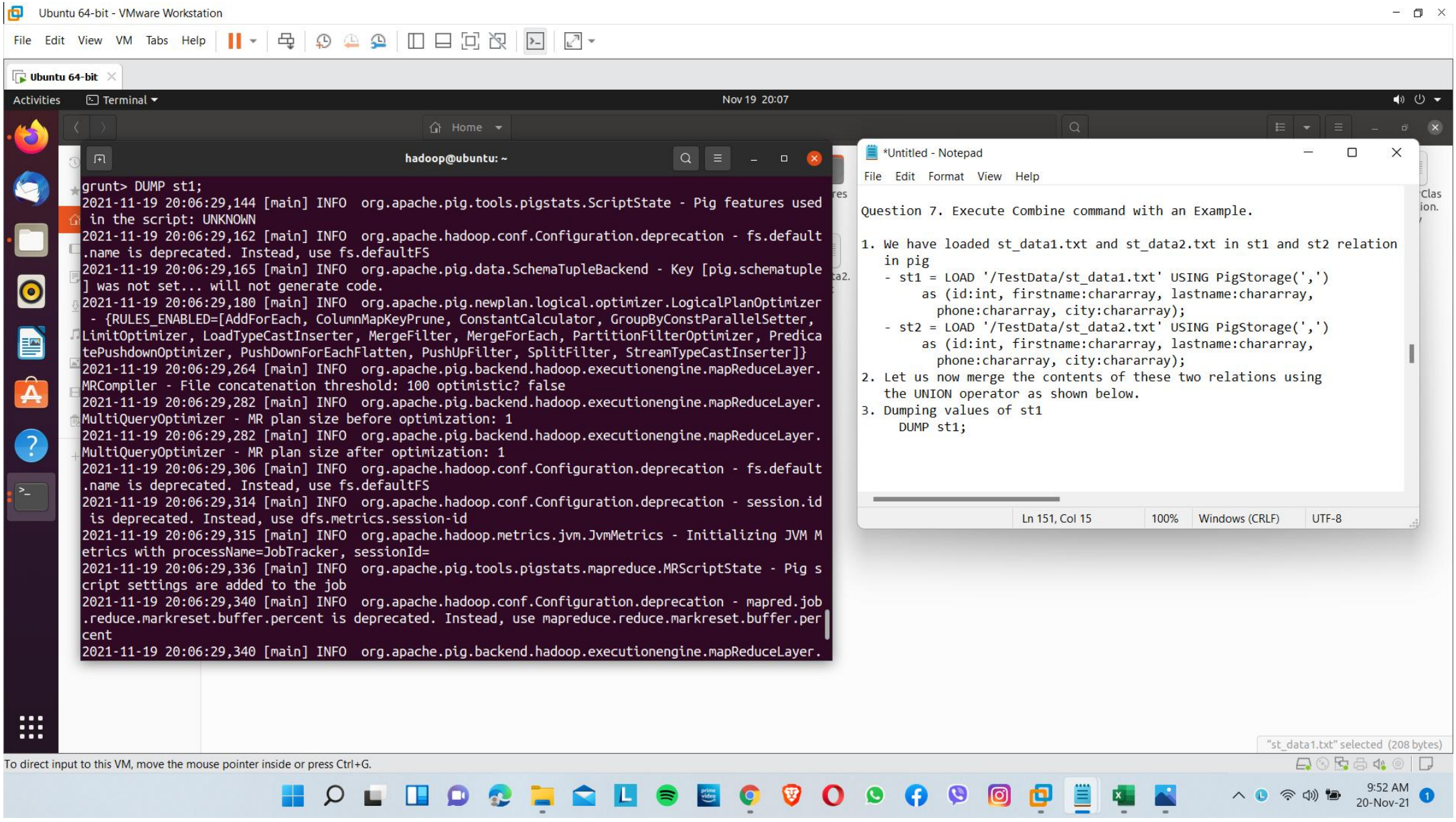

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:42
hadoop@ubuntu: ~
Successfully read 67 records (10504754 bytes) from: "/TestData/genderClassification.csv"
Output(s):
Successfully stored 8 records (10503162 bytes) in: "hdfs://localhost:8020/tmp/temp472322172/tmp-1100850238"
Counters:
Total records written : 8
Total bytes written : 10503162
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local2119348704_0001
2021-11-19 19:40:28,760 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 19:40:28,765 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 19:40:28,765 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 19:40:28,770 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 19:40:28,772 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:40:28,773 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
*Untitled - Notepad
File Edit Format View Help
Question 6. Execute Group command with an Example.
1. We are grouping our data based on favorite music genre.
- group_by_music_genre = GROUP genderClassification BY fav_music_genre;
2. After that we are dumping our results.
3. We can see that total 8 records are written.
Ln 118, Col 48
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:27 AM
20-Nov-21

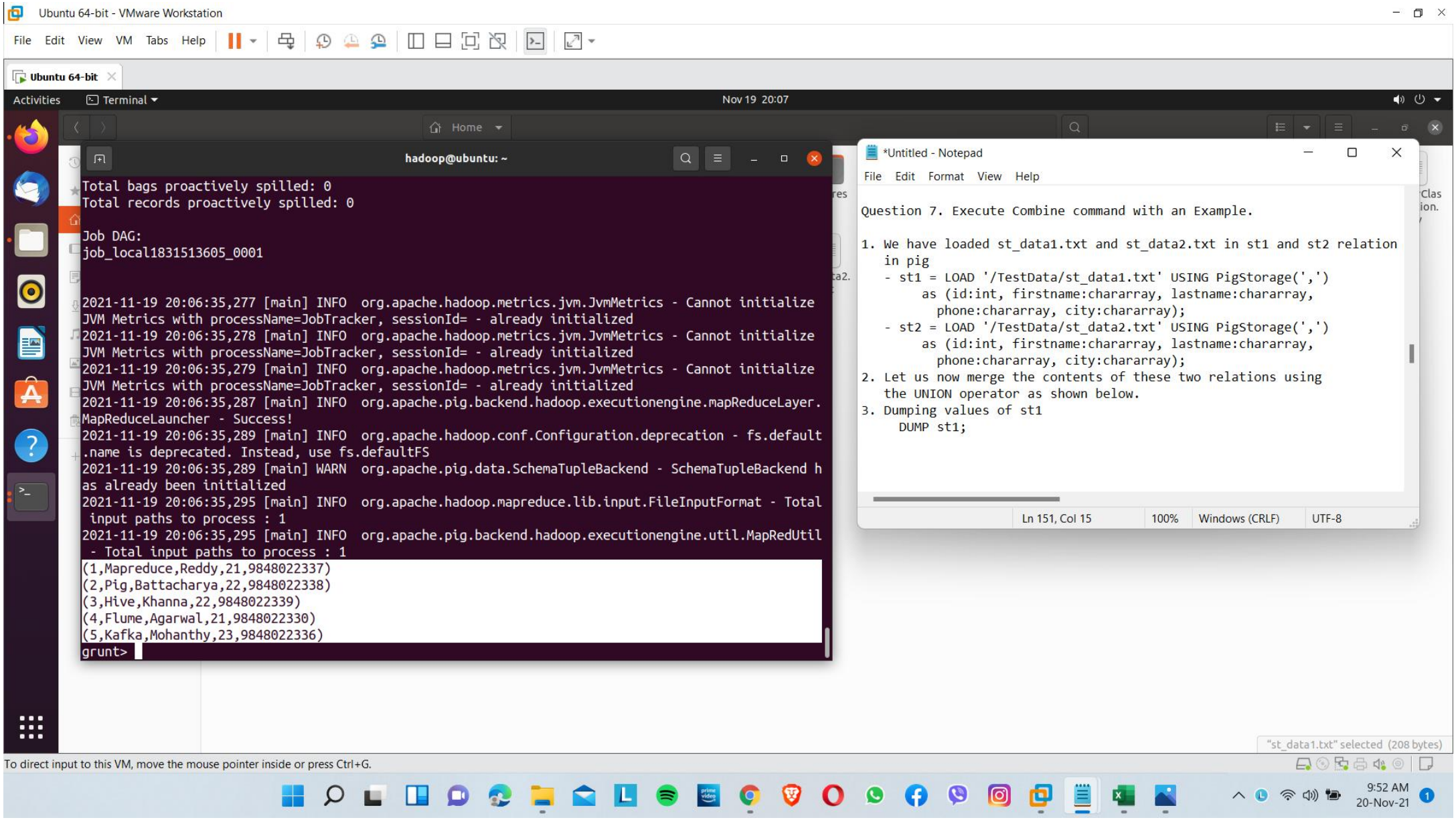

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:42
hadoop@ubuntu: ~
2021-11-19 19:40:28,770 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 19:40:28,772 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 19:40:28,773 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 19:40:28,777 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 19:40:28,778 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Pop,{(Neutral,Pop,Wine,Coca Cola/Pepsi,F),(Warm,Pop,Whiskey,Fanta,F),(Cool,Pop,Beer,Coca Cola/Pepsi,F),(Cool,Pop,Whiskey,Other,M),(Neutral,Pop,Doesn't drink,7UP/Sprite,M),(Cool,Pop,Whiskey,Fanta,M),(Cool,Pop,Other,Fanta,M),(Cool,Pop,Doesn't drink,Coca Cola/Pepsi,F),(Cool,Pop,Doesn't drink,Other,F),(Cool,Pop,Beer,Other,F),(Warm,Pop,Other,Coca Cola/Pepsi,F),(Cool,Pop,Beer,Fanta,F),(Cool,Pop,Doesn't drink,7UP/Sprite,F),(Cool,Pop,Wine,7UP/Sprite,F),(Warm,Pop,Wine,7UP/Sprite,F),(Warm,Pop,Other,7UP/Sprite,F),(Cool,Pop,Other,7UP/Sprite,F)})
(Rock,{(Cool,Rock,Doesn't drink,Other,M),(Cool,Rock,Beer,Coca Cola/Pepsi,F),(Cool,Rock,Other,Coca Cola/Pepsi,F),(Cool,Rock,Other,Coca Cola/Pepsi,F),(Warm,Rock,Other,7UP/Sprite,F),(Cool,Rock,Vodka,Coca Cola/Pepsi,F),(Warm,Rock,Wine,Coca Cola/Pepsi,F),(Cool,Rock,Vodka,Coca Cola/Pepsi,F),(Cool,Rock,Vodka,Coca Cola/Pepsi,M),(Warm,Rock,Vodka,7UP/Sprite,M),(Cool,Rock,Wine,Coca Cola/Pepsi,M),(Cool,Rock,Doesn't drink,Coca Cola/Pepsi,M),(Neutral,Rock,Doesn't drink,Coca Cola/Pepsi,M),(Warm,Rock,Beer,Fanta,M),(Cool,Rock,Wine,7UP/Sprite,M),(Cool,Rock,Vodka,Coca Cola/Pepsi,M),(Cool,Rock,Vodka,7UP/Sprite,F),(Neutral,Rock,Beer,Coca Cola/Pepsi,F),(Warm,Rock,Other,Coca Cola/Pepsi,F)})
(Hip hop,{(Warm,Hip hop,Beer,Coca Cola/Pepsi,M),(Neutral,Hip hop,Doesn't drink,Fanta,M),(Cool,Hip hop,Wine,Coca Cola/Pepsi,M),(Cool,Hip hop,Beer,Coca Cola/Pepsi,M),(Neutral,Hip hop,Vodka,Coca Cola/Pepsi,F),(Cool,Hip hop,Beer,Coca Cola/Pepsi,M),(Neutral,Hip hop,Beer,7UP/Sprite,M),(Cool,Hip hop,Doesn't drink,Other,M)})
(Electronic,{(Cool,Electronic,Whiskey,Coca Cola/Pepsi,M),(Warm,Electronic,Wine,Coca Cola/Pepsi,F),(Warm,Electronic,Vodka,Fanta,M),(Cool,Electronic,Beer,Coca Cola/Pepsi,M),(Cool,Electronic,Other,Fa
*Untitled - Notepad
File Edit Format View Help
Question 6. Execute Group command with an Example.
1. We are grouping our data based on favorite music genre.
- group_by_music_genre = GROUP genderClassification BY fav_music_genre;
2. After that we are dumping our results.
3. We can see that total 8 records are written.
4. We can see Pop genre at first
Ln 119, Col 33
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:27 AM
20-Nov-21

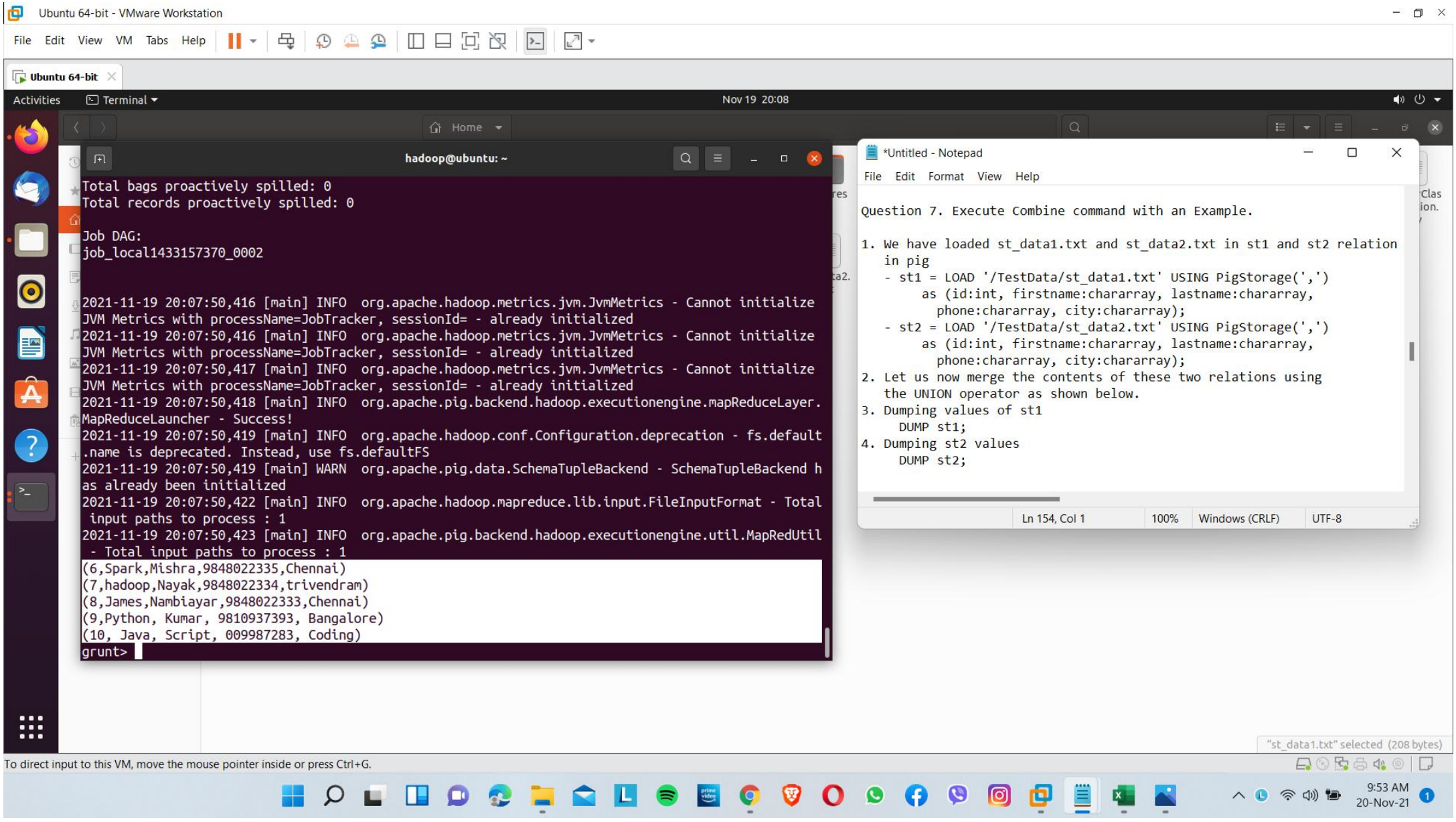

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:43
hadoop@ubuntu: ~
M),(Cool,Pop,Other,Fanta,M),(Cool,Pop,Doesn't drink,Coca Cola/Pepsi,F),(Cool,Pop,Doesn't drink,Oth
er,F),(Cool,Pop,Beer,Other,F),(Warm,Pop,Other,Coca Cola/Pepsi,F),(Cool,Pop,Beer,Fanta,F),(Cool,Pop
,Doesn't drink,7UP/Sprite,F),(Cool,Pop,Wine,7UP/Sprite,F),(Warm,Pop,Wine,7UP/Sprite,F),(Warm,Pop,O
ther,7UP/Sprite,F),(Cool,Pop,Other,7UP/Sprite,F)})
(Rock,{(Cool,Rock,Doesn't drink,Other,M),(Cool,Rock,Beer,Coca Cola/Pepsi,F),(Cool,Rock,Other,Coca
Cola/Pepsi,F),(Cool,Rock,Other,Coca Cola/Pepsi,F),(Warm,Rock,Other,7UP/Sprite,F),(Cool,Rock,Vodka,
Coca Cola/Pepsi,F),(Warm,Rock,Wine,Coca Cola/Pepsi,F),(Cool,Rock,Vodka,Coca Cola/Pepsi,F),(Cool,Ro
ck,Vodka,Coca Cola/Pepsi,M),(Warm,Rock,Vodka,7UP/Sprite,M),(Cool,Rock,Wine,Coca Cola/Pepsi,M),(Coo
l,Rock,Doesn't drink,Coca Cola/Pepsi,M),(Neutral,Rock,Doesn't drink,Coca Cola/Pepsi,M),(Warm,Rock,
Beer,Fanta,M),(Cool,Rock,Wine,7UP/Sprite,M),(Cool,Rock,Vodka,Coca Cola/Pepsi,M),(Cool,Rock,Vodka,7
UP/Sprite,F),(Neutral,Rock,Beer,Coca Cola/Pepsi,F),(Warm,Rock,Other,Coca Cola/Pepsi,F)})
(Hip hop,{(Warm,Hip hop,Beer,Coca Cola/Pepsi,M),(Neutral,Hip hop,Doesn't drink,Fanta,M),(Cool,Hip
hop,Wine,Coca Cola/Pepsi,M),(Cool,Hip hop,Beer,Coca Cola/Pepsi,M),(Neutral,Hip hop,Vodka,Coca Cola
/Pepsi,F),(Cool,Hip hop,Beer,Coca Cola/Pepsi,M),(Neutral,Hip hop,Beer,7UP/Sprite,M),(Cool,Hip hop,
Doesn't drink,Other,M)})
(Electronic,{(Cool,Electronic,Whiskey,Coca Cola/Pepsi,M),(Warm,Electronic,Wine,Coca Cola/Pepsi,F),
(Warm,Electronic,Vodka,Fanta,M),(Cool,Electronic,Beer,Coca Cola/Pepsi,M),(Cool,Electronic,Other,Fa
nta,M),(Cool,Electronic,Doesn't drink,Fanta,M),(Warm,Electronic,Other,Coca Cola/Pepsi,M),(Cool,Ele
ctronic,Doesn't drink,Fanta,F)})
(Jazz/Blues,{(Warm,Jazz/Blues,Vodka,Coca Cola/Pepsi,M),(Warm,Jazz/Blues,Wine,Coca Cola/Pepsi,F),(W
arm,Jazz/Blues,Whiskey,Fanta,F),(Warm,Jazz/Blues,Doesn't drink,Fanta,F)})
(R&B and soul,{(Cool,R&B and soul,Beer,Coca Cola/Pepsi,F),(Cool,R&B and soul,Doesn't drink,Coca Co
la/Pepsi,M),(Cool,R&B and soul,Whiskey,7UP/Sprite,M),(Warm,R&B and soul,Doesn't drink,Coca Cola/Pe
psi,M),(Warm,R&B and soul,Wine,Other,M),(Warm,R&B and soul,Whiskey,Coca Cola/Pepsi,F)})
(Folk/Traditional,{(Warm,Folk/Traditional,Whiskey,Fanta,F),(Cool,Folk/Traditional,Whiskey,7UP/Spri
te,F),(Cool,Folk/Traditional,Beer,Other,M),(Warm,Folk/Traditional,Other,Fanta,M)})
(Favorite Music Genre,{(Favorite Color,Favorite Music Genre,Favorite Beverage,Favorite Soft Drink,
Gender)})
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 6. Execute Group command with an Example.
1. We are grouping our data based on favorite music genre.
- group_by_music_genre = GROUP genderClassification BY
fav_music_genre;
2. After that we are dumping our results.
3. We can see that total 8 records are written.
4. We can see Pop genre at first.
5. After that we have Rock genre and similarly Electronic and favorite r
Ln 121, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:28 AM
20-Nov-21

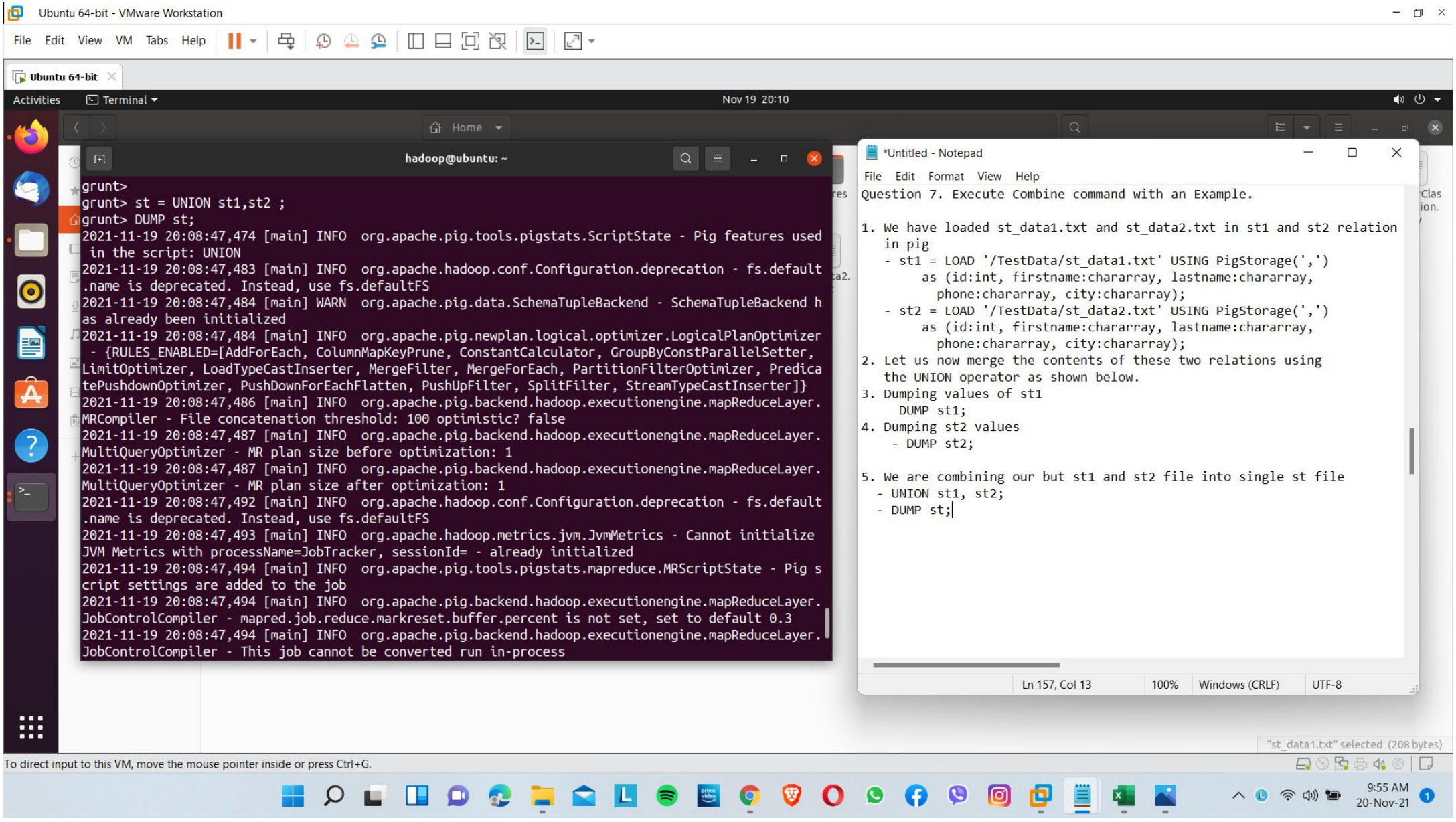

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 19:59
Home
hadoop@ubuntu: ~
hadoop@ubuntu:~$ hdfs dfs -put st_data1.txt /TestData/st_data1.txt
hadoop@ubuntu:~$ hdfs dfs -put st_data2.txt /TestData/st_data2.txt
hadoop@ubuntu:~$ hdfs dfs -ls /TestData/
Found 6 items
-rw-r--r-- 1 hadoop supergroup 2418 2021-11-19 18:20 /TestData/genderClassification.csv
-rw-r--r-- 1 hadoop supergroup 0 2021-10-16 21:24 /TestData/sample.txt
-rw-r--r-- 1 hadoop supergroup 208 2021-11-19 19:55 /TestData/st_data1.txt
-rw-r--r-- 1 hadoop supergroup 192 2021-11-19 19:55 /TestData/st_data2.txt
-rw-r--r-- 1 hadoop supergroup 186423 2021-11-18 00:43 /TestData/test_scores.csv
-rw-r--r-- 1 hadoop supergroup 0 2021-10-16 23:57 /TestData/testfile.txt
hadoop@ubuntu:~$ hdfs dfs -chmod 777 /TestData/st_data1.txt
hadoop@ubuntu:~$ hdfs dfs -chmod 777 /TestData/st_data2.txt
hadoop@ubuntu:~$ hdfs dfs -ls /TestData/
Found 6 items
-rw-r--r-- 1 hadoop supergroup 2418 2021-11-19 18:20 /TestData/genderClassification.csv
-rw-r--r-- 1 hadoop supergroup 0 2021-10-16 21:24 /TestData/sample.txt
-rwxrwxrwx 1 hadoop supergroup 208 2021-11-19 19:55 /TestData/st_data1.txt
-rwxrwxrwx 1 hadoop supergroup 192 2021-11-19 19:55 /TestData/st_data2.txt
-rw-r--r-- 1 hadoop supergroup 186423 2021-11-18 00:43 /TestData/test_scores.csv
-rw-r--r-- 1 hadoop supergroup 0 2021-10-16 23:57 /TestData/testfile.txt
hadoop@ubuntu:~$
Pictures
st_data2.txt
*Untitled - Notepad
File Edit Format View Help
Question 7. Execute Combine command with an Example.
The UNION operator of Pig Latin is used to merge the content of two relations.
1. For this example i have created two datasets st_data1.txt and st_data2.txt in local system.
2. I have copied st_data1.txt and st_data2.txt in hdfs from local
- hdfs dfs -put st_data1.txt /TestData/st_data1.txt
- hdfs dfs -put st_data2.txt /TestData/st_data2.txt
3. After that i have give read, write , and execute permission to those files in hdfs
4. We can see we have permissions.
Ln 135, Col 35
100%
Windows (CRLF)
UTF-8
"st_data1.txt" selected (208 bytes)
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:44 AM
20-Nov-21

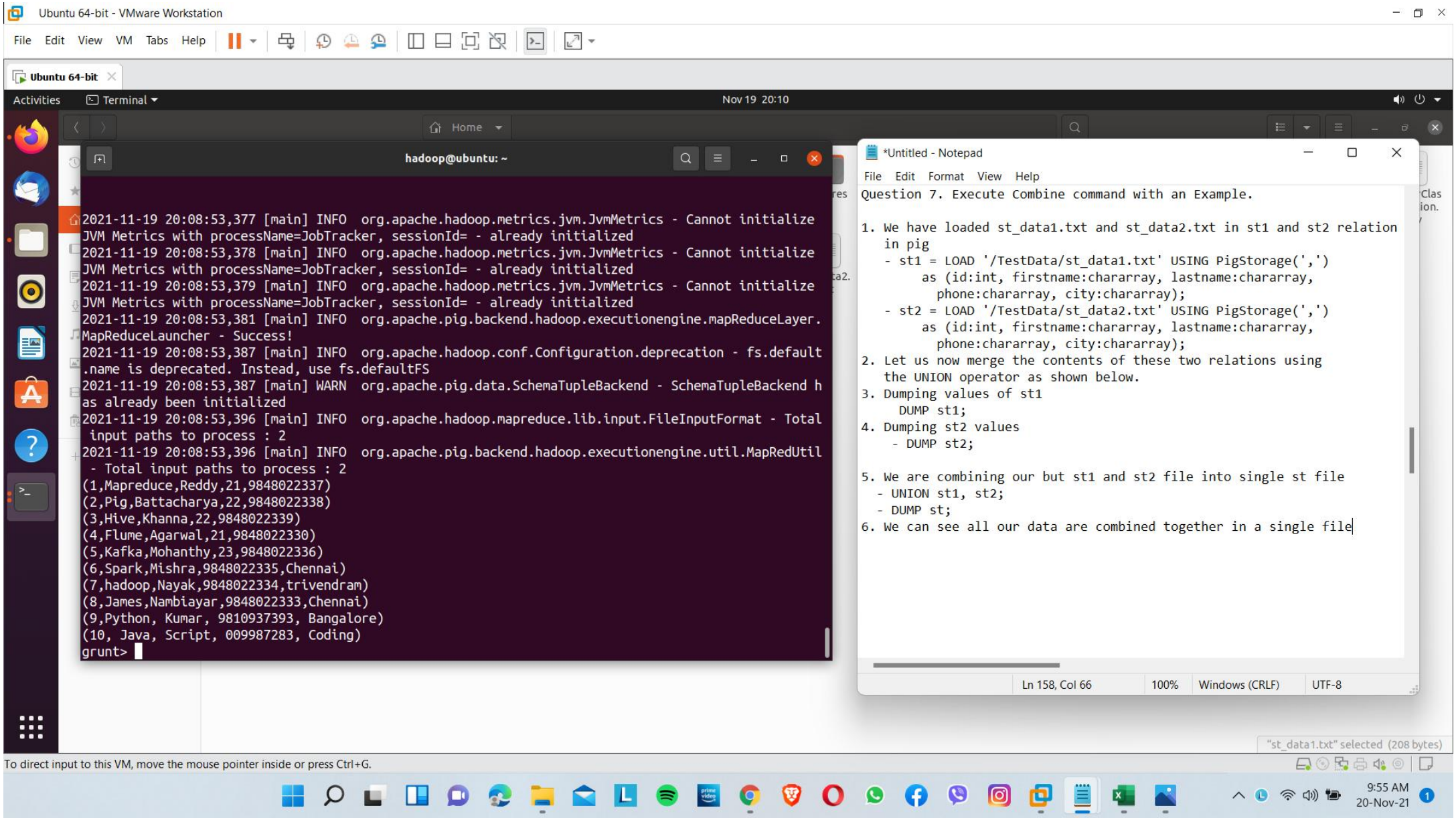

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 20:05
Home
hadoop@ubuntu: ~
grunt> st1 = LOAD '/TestData/st_data1.txt' USING PigStorage(',')
>> as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
2021-11-19 20:03:07,604 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
.name is deprecated. Instead, use fs.defaultFS
grunt> st2 = LOAD '/TestData/st_data2.txt' USING PigStorage(',')
>> as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
2021-11-19 20:03:21,346 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default
.name is deprecated. Instead, use fs.defaultFS
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 7. Execute Combine command with an Example.
1. We have loaded st_data1.txt and st_data2.txt in st1 and st2 relation
in pig
- st1 = LOAD '/TestData/st_data1.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray,
phone:chararray, city:chararray);
- st2 = LOAD '/TestData/st_data2.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray,
phone:chararray, city:chararray);
Ln 148, Col 1
100%
Windows (CRLF)
UTF-8
"st_data1.txt" selected (208 bytes)
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:50 AM
20-Nov-21

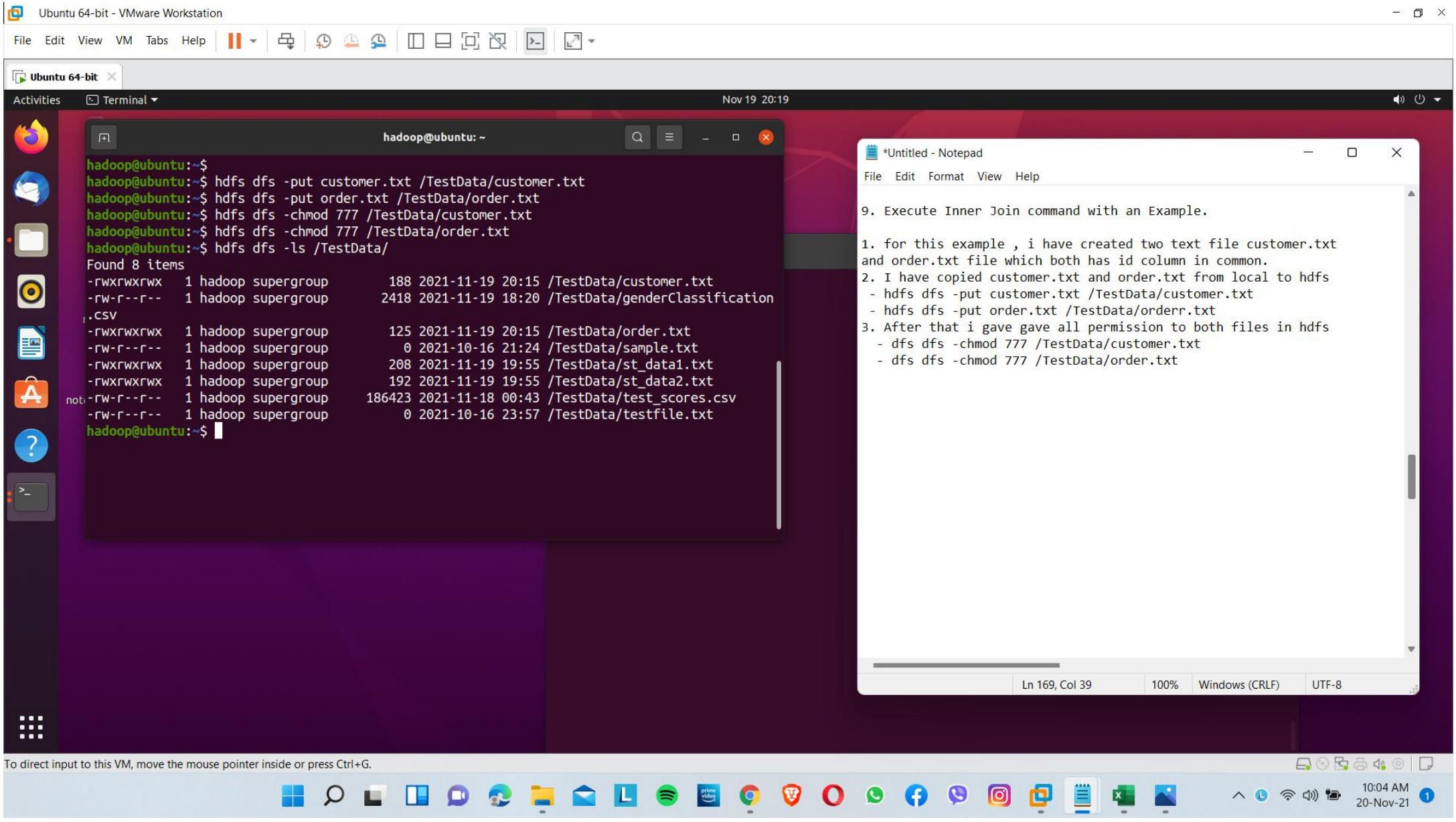

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 20:07
Home
hadoop@ubuntu: ~
grunt> DUMP st1;
2021-11-19 20:06:29,144 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2021-11-19 20:06:29,162 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 20:06:29,165 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 20:06:29,180 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 20:06:29,264 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 20:06:29,282 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 20:06:29,282 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 20:06:29,306 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 20:06:29,314 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is deprecated. Instead, use dfs.metrics.session-id
2021-11-19 20:06:29,315 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metrics with processName=JobTracker, sessionId=
2021-11-19 20:06:29,336 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 20:06:29,340 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
2021-11-19 20:06:29,340 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 7. Execute Combine command with an Example.
1. We have loaded st_data1.txt and st_data2.txt in st1 and st2 relation in pig
- st1 = LOAD '/TestData/st_data1.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
- st2 = LOAD '/TestData/st_data2.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
2. Let us now merge the contents of these two relations using the UNION operator as shown below.
3. Dumping values of st1
DUMP st1;
Ln 151, Col 15
100%
Windows (CRLF)
UTF-8
"st_data1.txt" selected (208 bytes)
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:52 AM
20-Nov-21

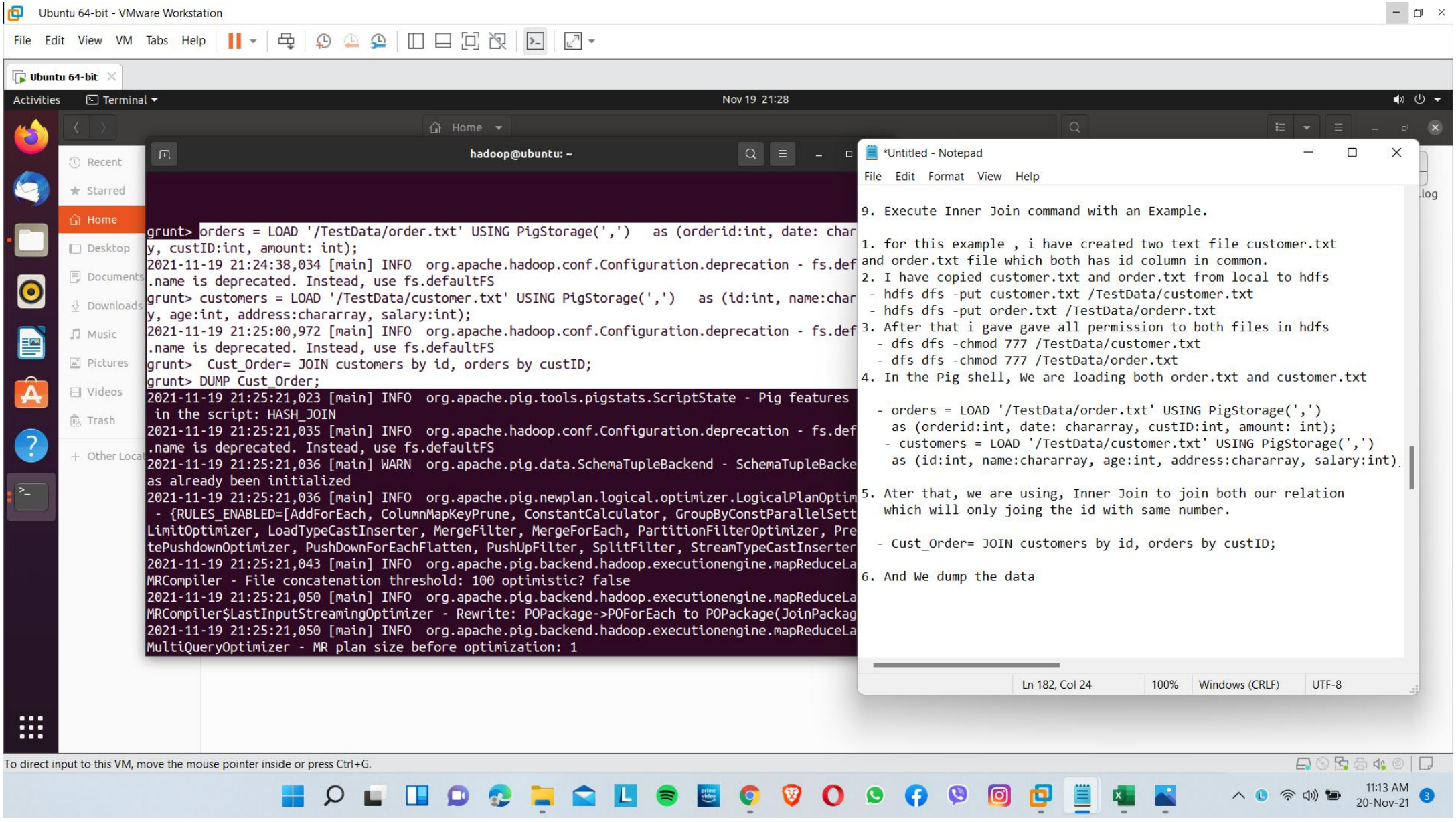

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 20:07
Home
hadoop@ubuntu: ~
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1831513605_0001
2021-11-19 20:06:35,277 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:06:35,278 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:06:35,279 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:06:35,287 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 20:06:35,289 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 20:06:35,289 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 20:06:35,295 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 20:06:35,295 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Mapreduce,Reddy,21,9848022337)
(2,Pig,Battacharya,22,9848022338)
(3,Hive,Khanna,22,9848022339)
(4,Flume,Agarwal,21,9848022330)
(5,Kafka,Mohanthy,23,9848022336)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 7. Execute Combine command with an Example.
1. We have loaded st_data1.txt and st_data2.txt in st1 and st2 relation in pig
- st1 = LOAD '/TestData/st_data1.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
- st2 = LOAD '/TestData/st_data2.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
2. Let us now merge the contents of these two relations using the UNION operator as shown below.
3. Dumping values of st1
DUMP st1;
Ln 151, Col 15
100%
Windows (CRLF)
UTF-8
"st_data1.txt" selected (208 bytes)
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:52 AM
20-Nov-21

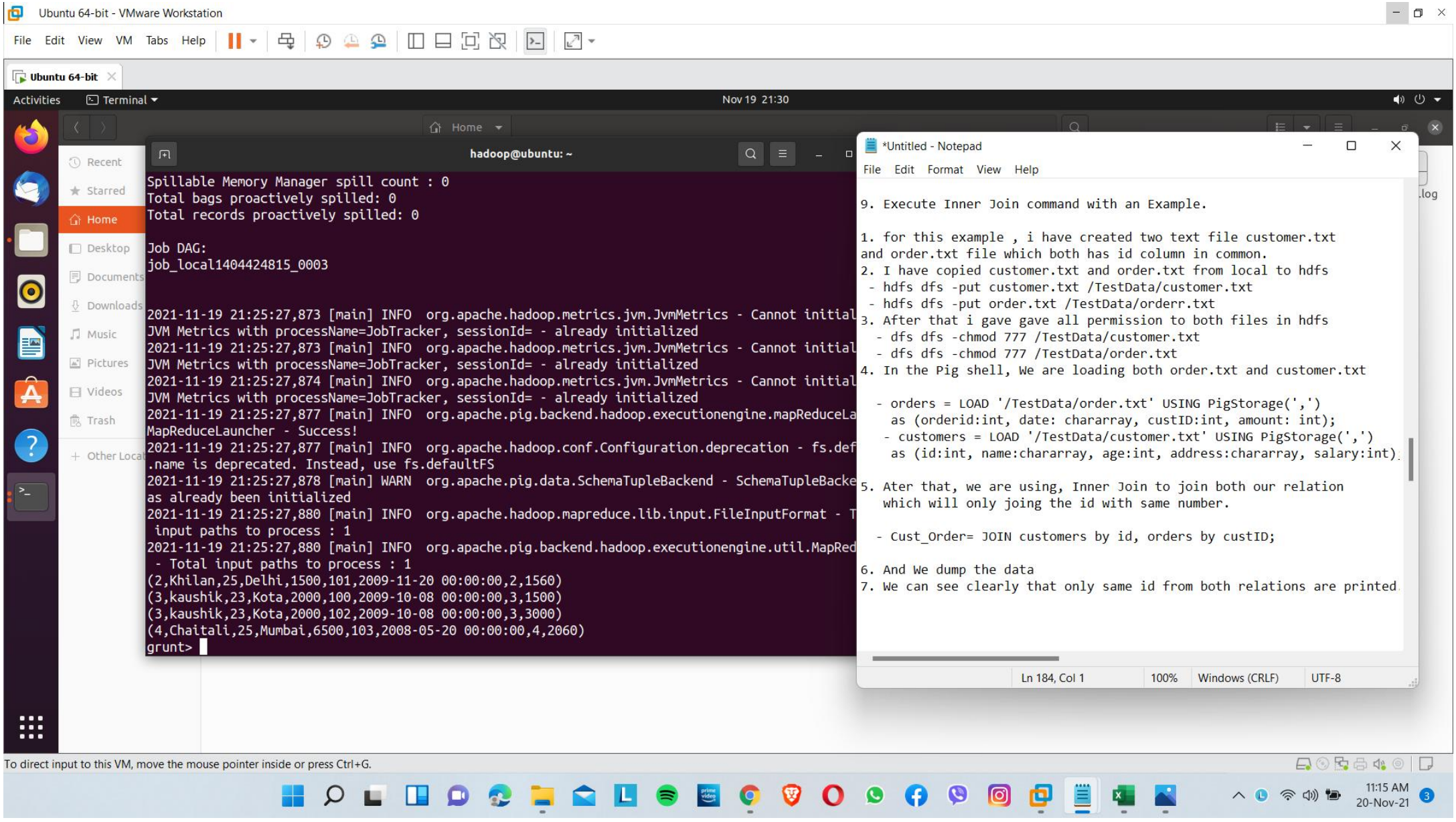

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 20:08
Home
hadoop@ubuntu: ~
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1433157370_0002
2021-11-19 20:07:50,416 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:07:50,416 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:07:50,417 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:07:50,418 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 20:07:50,419 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 20:07:50,419 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 20:07:50,422 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 20:07:50,423 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(6,Spark,Mishra,9848022335,Chennai)
(7,hadoop,Nayak,9848022334,trivendram)
(8,James,Nambiayar,9848022333,Chennai)
(9,Python, Kumar, 9810937393, Bangalore)
(10, Java, Script, 009987283, Coding)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 7. Execute Combine command with an Example.
1. We have loaded st_data1.txt and st_data2.txt in st1 and st2 relation in pig
- st1 = LOAD '/TestData/st_data1.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
- st2 = LOAD '/TestData/st_data2.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
2. Let us now merge the contents of these two relations using the UNION operator as shown below.
3. Dumping values of st1
DUMP st1;
4. Dumping st2 values
DUMP st2;
Ln 154, Col 1
100%
Windows (CRLF)
UTF-8
"st_data1.txt" selected (208 bytes)
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:53 AM
20-Nov-21

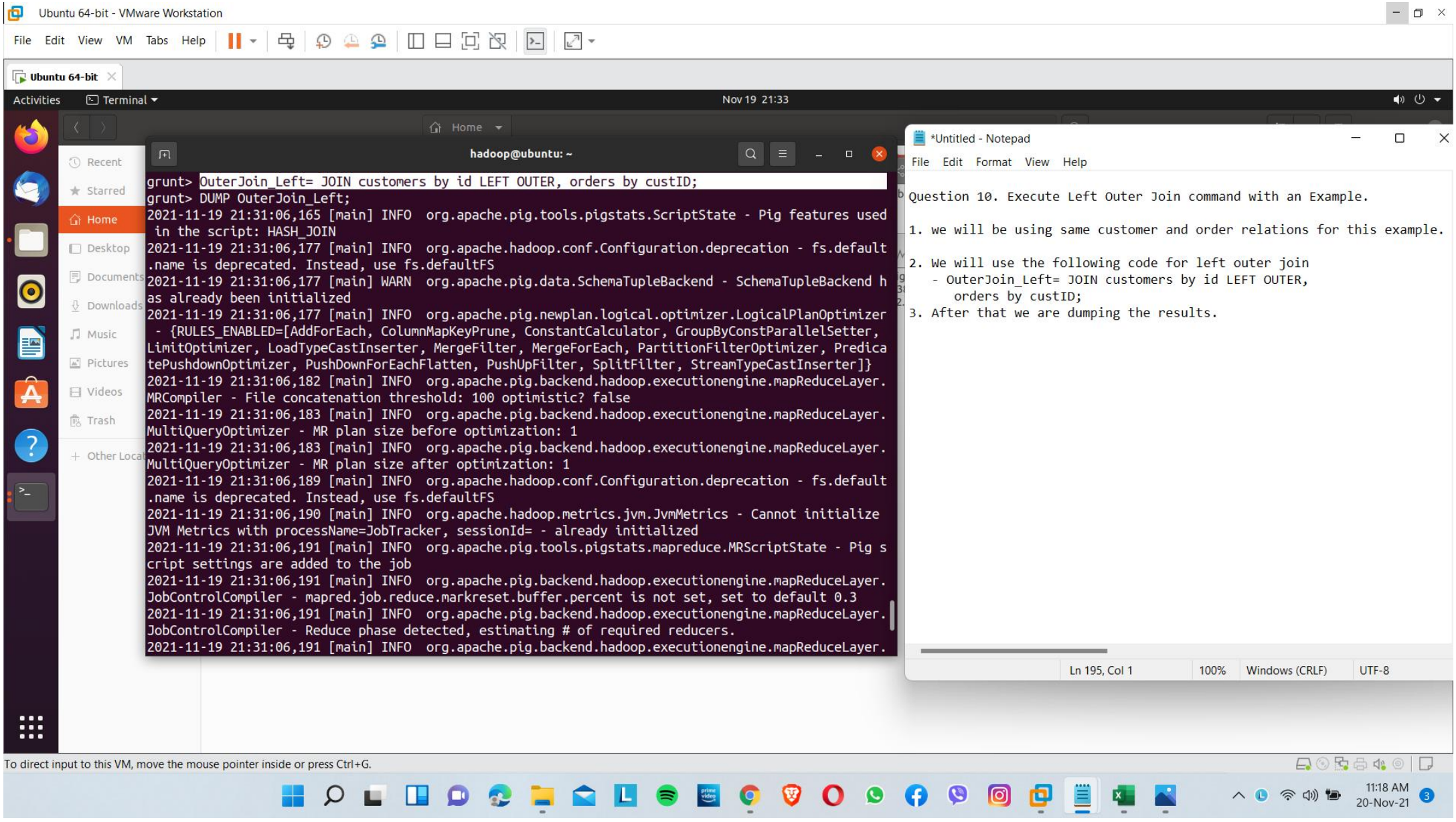

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 20:10
Home
hadoop@ubuntu: ~
grunt>
grunt> st = UNION st1,st2 ;
grunt> DUMP st;
2021-11-19 20:08:47,474 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNION
2021-11-19 20:08:47,483 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 20:08:47,484 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 20:08:47,484 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 20:08:47,486 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 20:08:47,487 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 20:08:47,487 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 20:08:47,492 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 20:08:47,493 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:08:47,494 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 20:08:47,494 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 20:08:47,494 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
*Untitled - Notepad
File Edit Format View Help
Question 7. Execute Combine command with an Example.
1. We have loaded st_data1.txt and st_data2.txt in st1 and st2 relation in pig
- st1 = LOAD '/TestData/st_data1.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
- st2 = LOAD '/TestData/st_data2.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
2. Let us now merge the contents of these two relations using the UNION operator as shown below.
3. Dumping values of st1
DUMP st1;
4. Dumping st2 values
- DUMP st2;
5. We are combining our but st1 and st2 file into single st file
- UNION st1, st2;
- DUMP st;
Ln 157, Col 13
100%
Windows (CRLF)
UTF-8
"st_data1.txt" selected (208 bytes)
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:55 AM
20-Nov-21

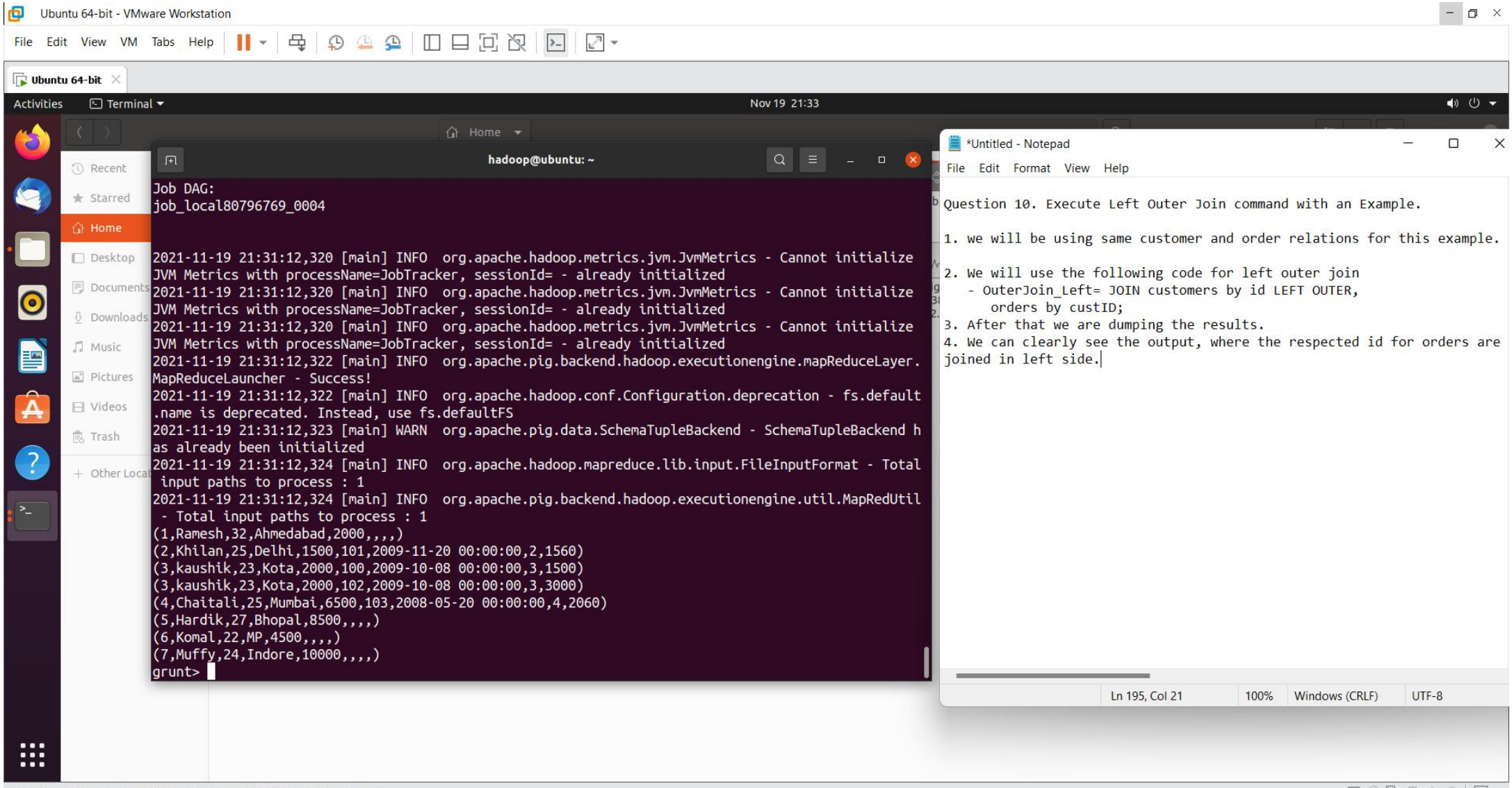

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 20:10
Home
hadoop@ubuntu: ~
2021-11-19 20:08:53,377 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:08:53,378 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:08:53,379 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 20:08:53,381 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 20:08:53,387 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 20:08:53,387 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 20:08:53,396 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 2
2021-11-19 20:08:53,396 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 2
(1,Mapreduce,Reddy,21,9848022337)
(2,Pig,Battacharya,22,9848022338)
(3,Hive,Khanna,22,9848022339)
(4,Flume,Agarwal,21,9848022330)
(5,Kafka,Mohanthy,23,9848022336)
(6,Spark,Mishra,9848022335,Chennai)
(7,hadoop,Nayak,9848022334,trivendram)
(8,James,Nambiayar,9848022333,Chennai)
(9,Python, Kumar, 9810937393, Bangalore)
(10, Java, Script, 009987283, Coding)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 7. Execute Combine command with an Example.
1. We have loaded st_data1.txt and st_data2.txt in st1 and st2 relation in pig
- st1 = LOAD '/TestData/st_data1.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
- st2 = LOAD '/TestData/st_data2.txt' USING PigStorage(',')
as (id:int, firstname:chararray, lastname:chararray, phone:chararray, city:chararray);
2. Let us now merge the contents of these two relations using the UNION operator as shown below.
3. Dumping values of st1
DUMP st1;
4. Dumping st2 values
- DUMP st2;
5. We are combining our but st1 and st2 file into single st file
- UNION st1, st2;
- DUMP st;
6. We can see all our data are combined together in a single file
Ln 158, Col 66
100%
Windows (CRLF)
UTF-8
"st_data1.txt" selected (208 bytes)
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
9:55 AM
20-Nov-21

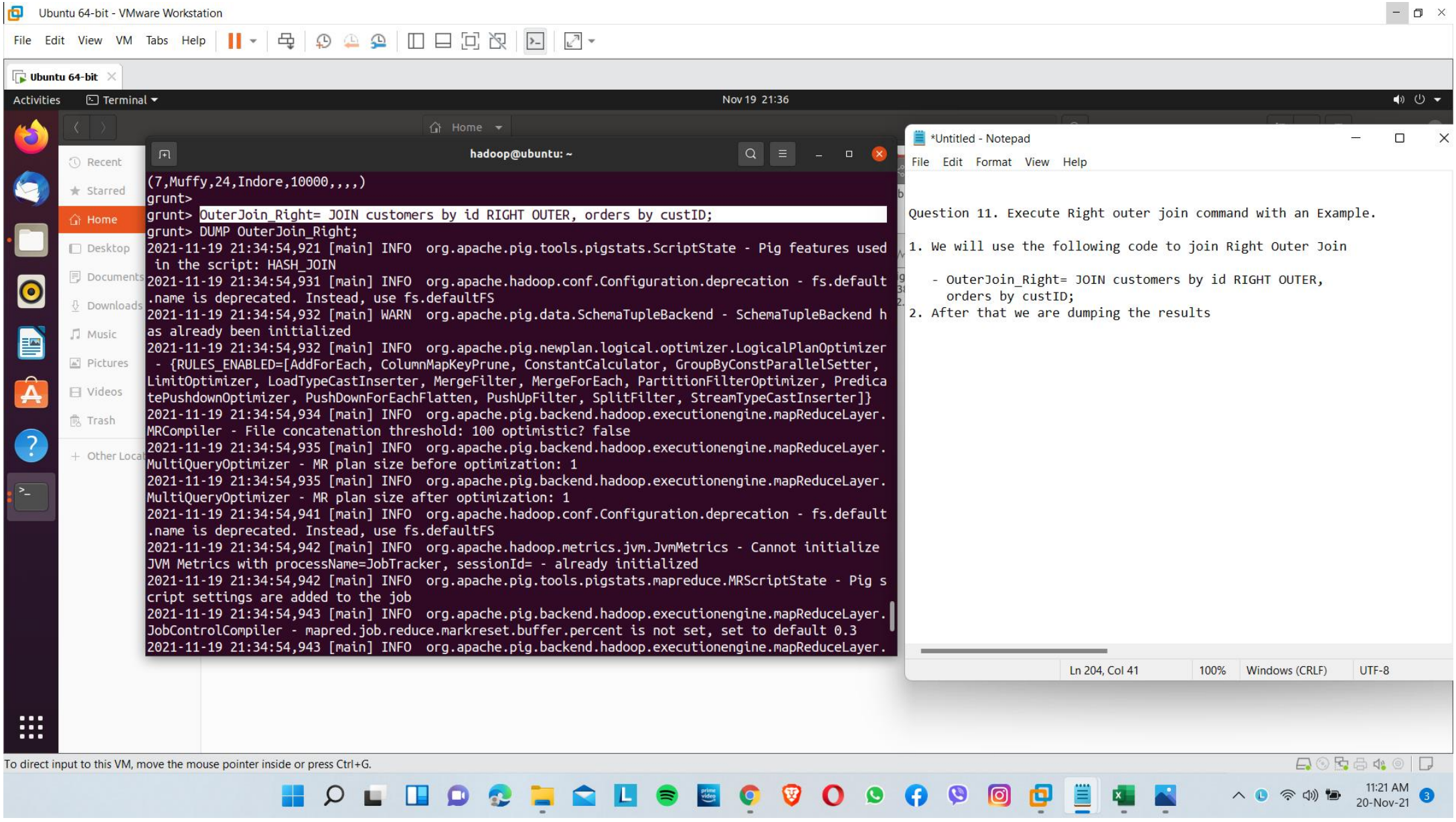

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 20:19
hadoop@ubuntu: ~
hadoop@ubuntu:~$
hadoop@ubuntu:~$ hdfs dfs -put customer.txt /TestData/customer.txt
hadoop@ubuntu:~$ hdfs dfs -put order.txt /TestData/order.txt
hadoop@ubuntu:~$ hdfs dfs -chmod 777 /TestData/customer.txt
hadoop@ubuntu:~$ hdfs dfs -chmod 777 /TestData/order.txt
hadoop@ubuntu:~$ hdfs dfs -ls /TestData/
Found 8 items
-rwxrwxrwx 1 hadoop supergroup 188 2021-11-19 20:15 /TestData/customer.txt
-rw-r--r-- 1 hadoop supergroup 2418 2021-11-19 18:20 /TestData/genderClassification.csv
-rwxrwxrwx 1 hadoop supergroup 125 2021-11-19 20:15 /TestData/order.txt
-rw-r--r-- 1 hadoop supergroup 0 2021-10-16 21:24 /TestData/sample.txt
-rwxrwxrwx 1 hadoop supergroup 208 2021-11-19 19:55 /TestData/st_data1.txt
-rwxrwxrwx 1 hadoop supergroup 192 2021-11-19 19:55 /TestData/st_data2.txt
-rw-r--r-- 1 hadoop supergroup 186423 2021-11-18 00:43 /TestData/test_scores.csv
-rw-r--r-- 1 hadoop supergroup 0 2021-10-16 23:57 /TestData/testfile.txt
hadoop@ubuntu:~$
*Untitled - Notepad
File Edit Format View Help
9. Execute Inner Join command with an Example.
1. for this example , i have created two text file customer.txt and order.txt file which both has id column in common.
2. I have copied customer.txt and order.txt from local to hdfs
- hdfs dfs -put customer.txt /TestData/customer.txt
- hdfs dfs -put order.txt /TestData/orderr.txt
3. After that i gave gave all permission to both files in hdfs
- dfs dfs -chmod 777 /TestData/customer.txt
- dfs dfs -chmod 777 /TestData/order.txt
Ln 169, Col 39
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
10:04 AM
20-Nov-21

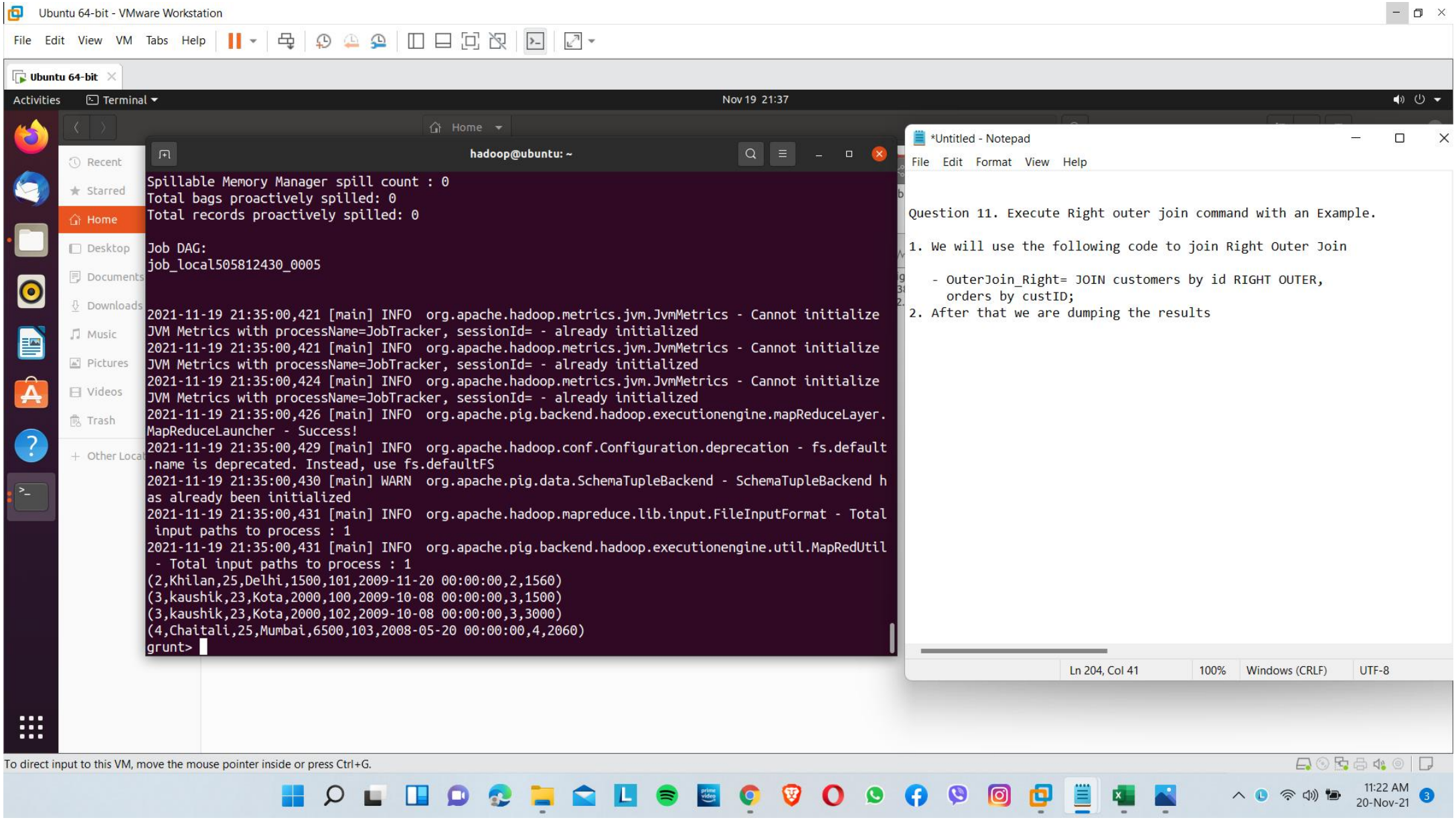

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:28
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
grunt> orders = LOAD '/TestData/order.txt' USING PigStorage(',') as (orderid:int, date: char
y, custID:int, amount: int);
2021-11-19 21:24:38,034 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.def
.name is deprecated. Instead, use fs.defaultFS
grunt> customers = LOAD '/TestData/customer.txt' USING PigStorage(',') as (id:int, name:char
y, age:int, address:chararray, salary:int);
2021-11-19 21:25:00,972 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.def
.name is deprecated. Instead, use fs.defaultFS
grunt> Cust_Order= JOIN customers by id, orders by custID;
grunt> DUMP Cust_Order;
2021-11-19 21:25:21,023 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features
in the script: HASH_JOIN
2021-11-19 21:25:21,035 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.def
.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:25:21,036 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBacke
as already been initialized
2021-11-19 21:25:21,036 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptim
- {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSett
LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, Pre
tePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter
2021-11-19 21:25:21,043 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLa
MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 21:25:21,050 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLa
MRCompiler$LastInputStreamingOptimizer - Rewrite: POPackage->POForEach to POPackage(JoinPackag
2021-11-19 21:25:21,050 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLa
MultiQueryOptimizer - MR plan size before optimization: 1
*Untitled - Notepad
File Edit Format View Help
9. Execute Inner Join command with an Example.
1. for this example , i have created two text file customer.txt
and order.txt file which both has id column in common.
2. I have copied customer.txt and order.txt from local to hdfs
- hdfs dfs -put customer.txt /TestData/customer.txt
- hdfs dfs -put order.txt /TestData/orderr.txt
3. After that i gave gave all permission to both files in hdfs
- dfs dfs -chmod 777 /TestData/customer.txt
- dfs dfs -chmod 777 /TestData/order.txt
4. In the Pig shell, We are loading both order.txt and customer.txt
- orders = LOAD '/TestData/order.txt' USING PigStorage(',')
as (orderid:int, date: chararray, custID:int, amount: int);
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
as (id:int, name:chararray, age:int, address:chararray, salary:int);
5. Ater that, we are using, Inner Join to join both our relation
which will only joing the id with same number.
- Cust_Order= JOIN customers by id, orders by custID;
6. And We dump the data
Ln 182, Col 24
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:13 AM
20-Nov-21

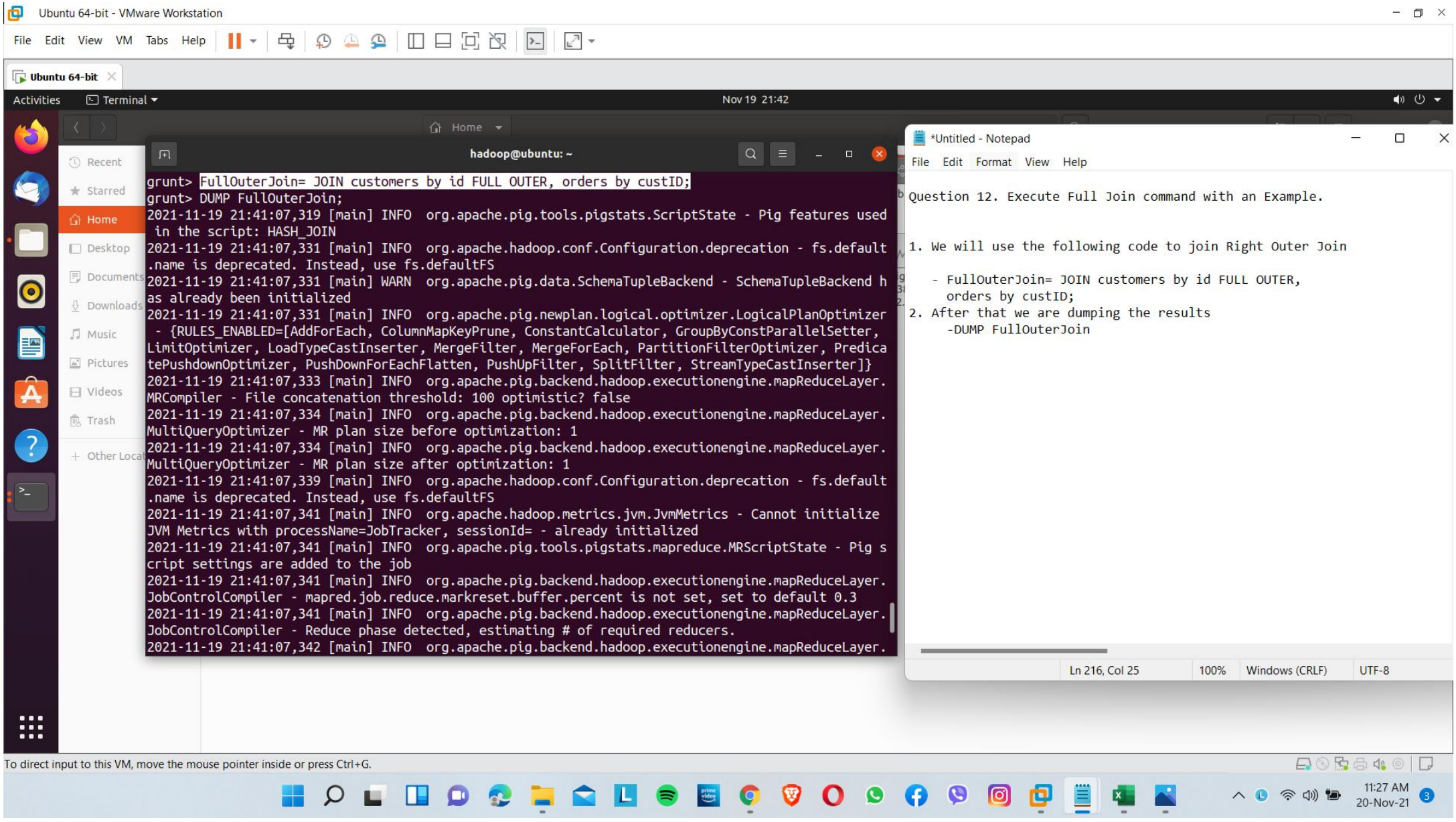

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:30
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1404424815_0003
2021-11-19 21:25:27,873 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initial
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:25:27,873 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initial
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:25:27,874 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initial
JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:25:27,877 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLa
MapReduceLauncher - Success!
2021-11-19 21:25:27,877 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.def
.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:25:27,878 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBacke
as already been initialized
2021-11-19 21:25:27,880 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - T
input paths to process : 1
2021-11-19 21:25:27,880 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRed
- Total input paths to process : 1
(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)
grunt>
*Untitled - Notepad
File Edit Format View Help
9. Execute Inner Join command with an Example.
1. for this example , i have created two text file customer.txt
and order.txt file which both has id column in common.
2. I have copied customer.txt and order.txt from local to hdfs
- hdfs dfs -put customer.txt /TestData/customer.txt
- hdfs dfs -put order.txt /TestData/orderr.txt
3. After that i gave gave all permission to both files in hdfs
- dfs dfs -chmod 777 /TestData/customer.txt
- dfs dfs -chmod 777 /TestData/order.txt
4. In the Pig shell, We are loading both order.txt and customer.txt
- orders = LOAD '/TestData/order.txt' USING PigStorage(',')
as (orderid:int, date: chararray, custID:int, amount: int);
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
as (id:int, name:chararray, age:int, address:chararray, salary:int);
5. Ater that, we are using, Inner Join to join both our relation
which will only joing the id with same number.
- Cust_Order= JOIN customers by id, orders by custID;
6. And We dump the data
7. We can see clearly that only same id from both relations are printed
Ln 184, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:15 AM
20-Nov-21

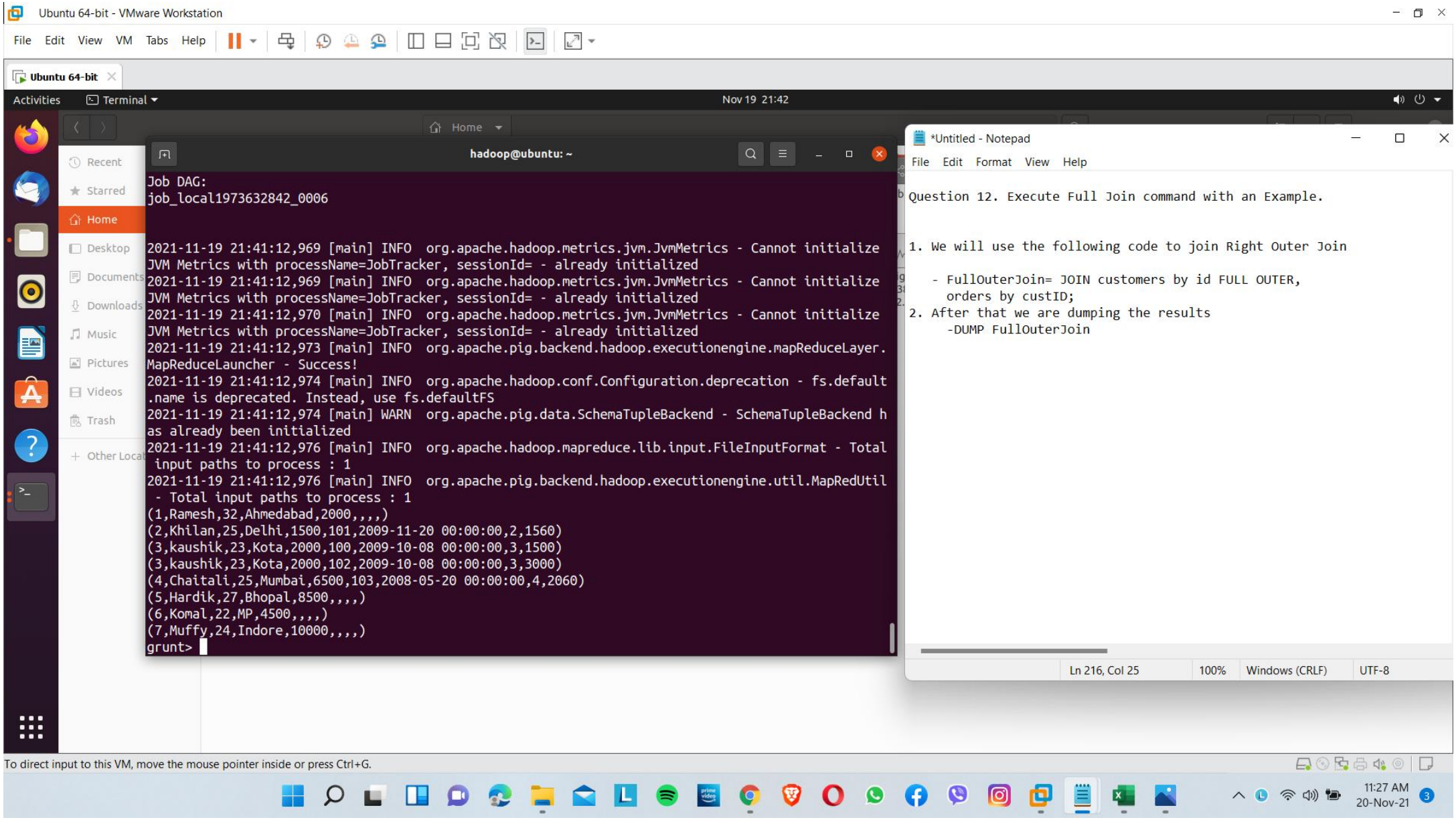

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:33
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
grunt> OuterJoin_Left= JOIN customers by id LEFT OUTER, orders by custID;
grunt> DUMP OuterJoin_Left;
2021-11-19 21:31:06,165 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN
2021-11-19 21:31:06,177 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:31:06,177 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 21:31:06,177 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 21:31:06,182 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 21:31:06,183 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 21:31:06,183 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 21:31:06,189 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:31:06,190 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:31:06,191 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 21:31:06,191 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 21:31:06,191 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2021-11-19 21:31:06,191 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 10. Execute Left Outer Join command with an Example.
1. we will be using same customer and order relations for this example.
2. We will use the following code for left outer join
- OuterJoin_Left= JOIN customers by id LEFT OUTER,
orders by custID;
3. After that we are dumping the results.
Ln 195, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:18 AM
20-Nov-21

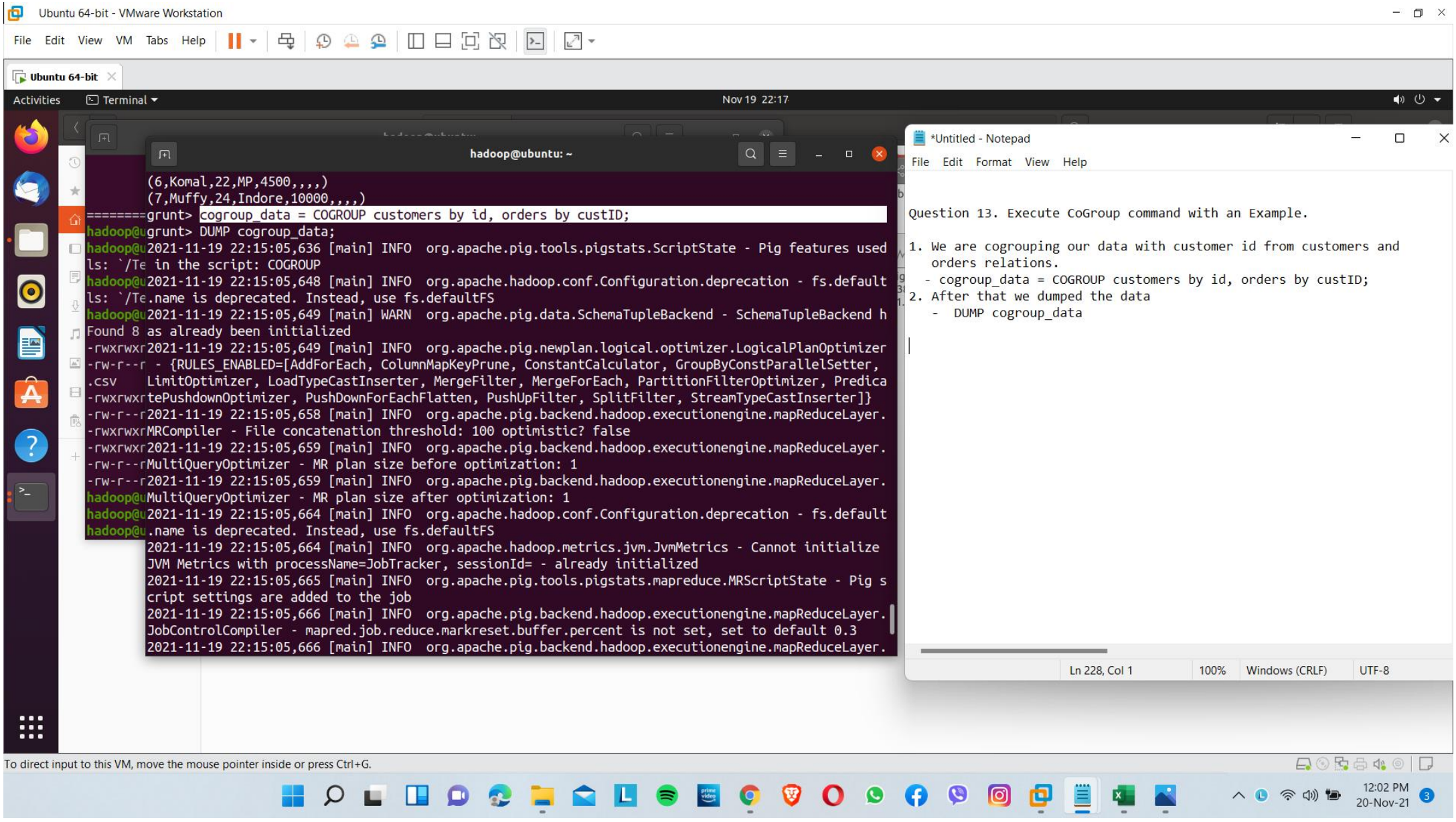

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:33
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
Job DAG:
job_local80796769_0004
2021-11-19 21:31:12,320 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:31:12,320 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:31:12,320 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:31:12,322 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 21:31:12,322 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:31:12,323 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 21:31:12,324 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 21:31:12,324 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Ramesh,32,Ahmedabad,2000,,,,)
(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)
(5,Hardik,27,Bhopal,8500,,,,)
(6,Komal,22,MP,4500,,,,)
(7,Muffy,24,Indore,10000,,,,)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 10. Execute Left Outer Join command with an Example.
1. we will be using same customer and order relations for this example.
2. We will use the following code for left outer join
- OuterJoin_Left= JOIN customers by id LEFT OUTER,
orders by custID;
3. After that we are dumping the results.
4. We can clearly see the output, where the respected id for orders are joined in left side.
Ln 195, Col 21
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:18 AM
20-Nov-21

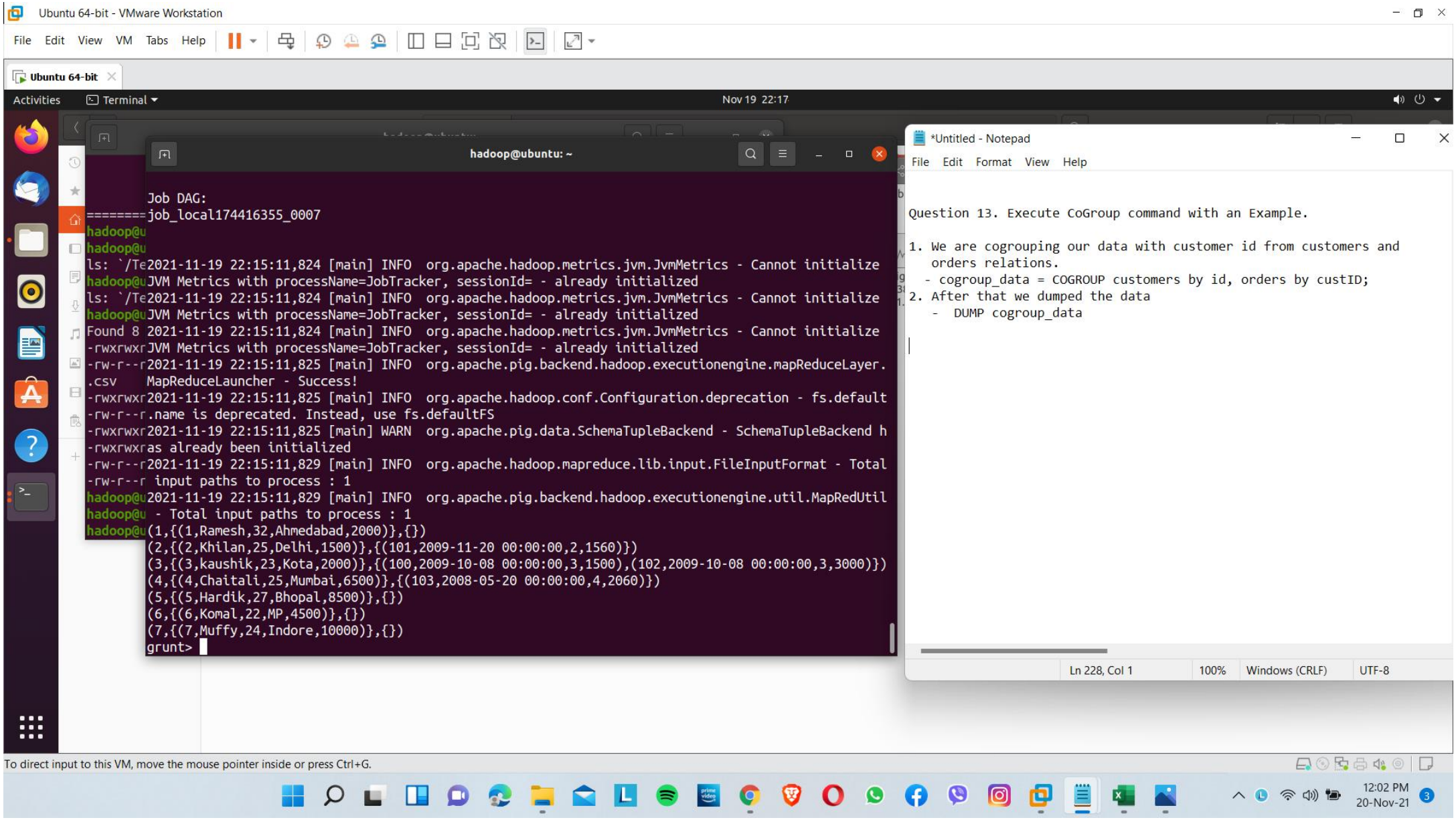

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:36
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
(7,Muffy,24,Indore,10000,,,,)
grunt>
grunt> OuterJoin_Right= JOIN customers by id RIGHT OUTER, orders by custID;
grunt> DUMP OuterJoin_Right;
2021-11-19 21:34:54,921 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN
2021-11-19 21:34:54,931 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:34:54,932 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 21:34:54,932 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 21:34:54,934 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 21:34:54,935 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 21:34:54,935 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 21:34:54,941 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:34:54,942 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:34:54,942 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 21:34:54,943 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 21:34:54,943 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 11. Execute Right outer join command with an Example.
1. We will use the following code to join Right Outer Join
- OuterJoin_Right= JOIN customers by id RIGHT OUTER,
orders by custID;
2. After that we are dumping the results
Ln 204, Col 41
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:21 AM
20-Nov-21

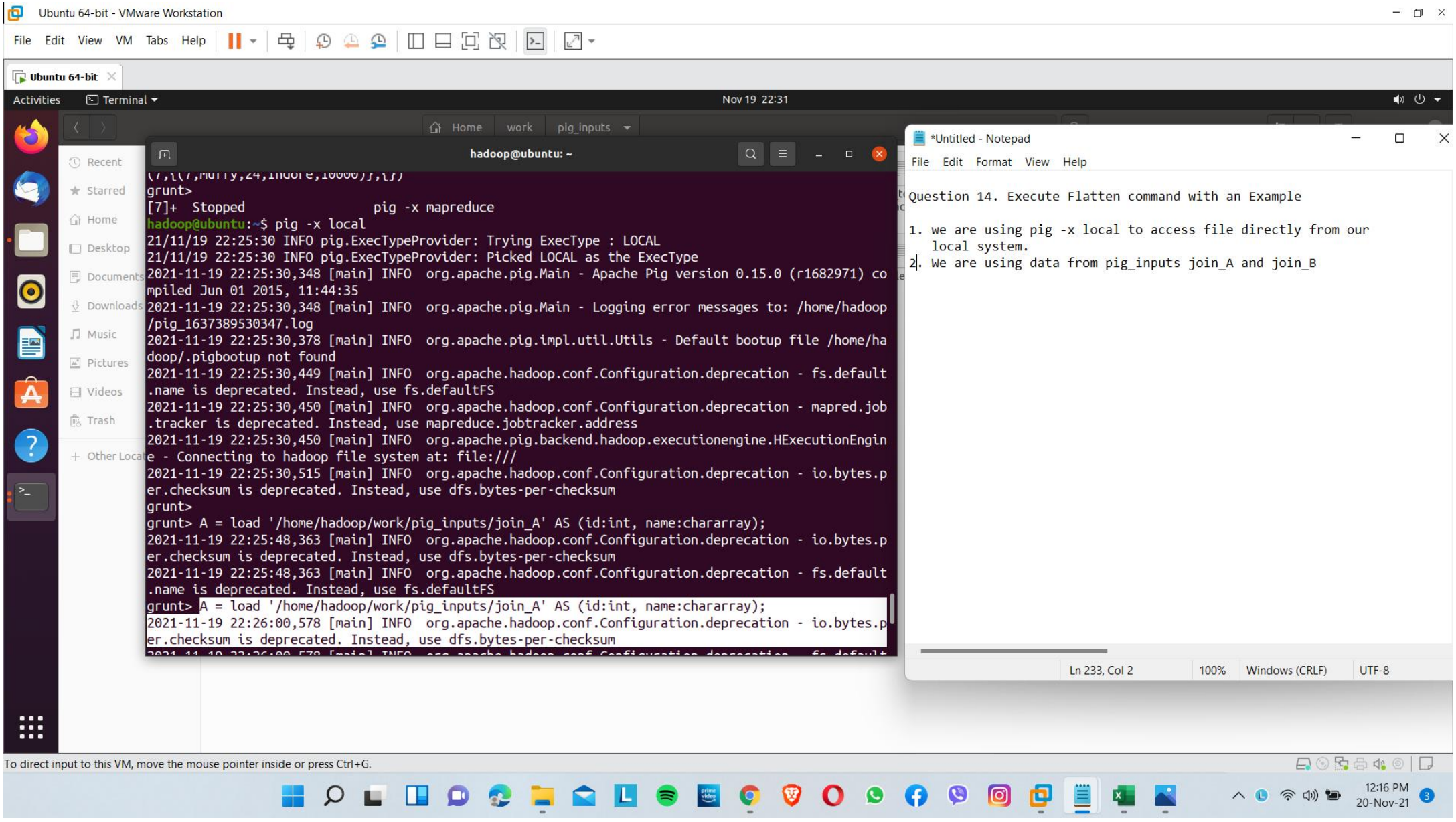

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:37
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
hadoop@ubuntu: ~
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local505812430_0005
2021-11-19 21:35:00,421 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:35:00,421 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:35:00,424 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:35:00,426 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 21:35:00,429 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:35:00,430 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 21:35:00,431 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 21:35:00,431 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 11. Execute Right outer join command with an Example.
1. We will use the following code to join Right Outer Join
- OuterJoin_Right= JOIN customers by id RIGHT OUTER, orders by custID;
2. After that we are dumping the results
Ln 204, Col 41
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:22 AM
20-Nov-21

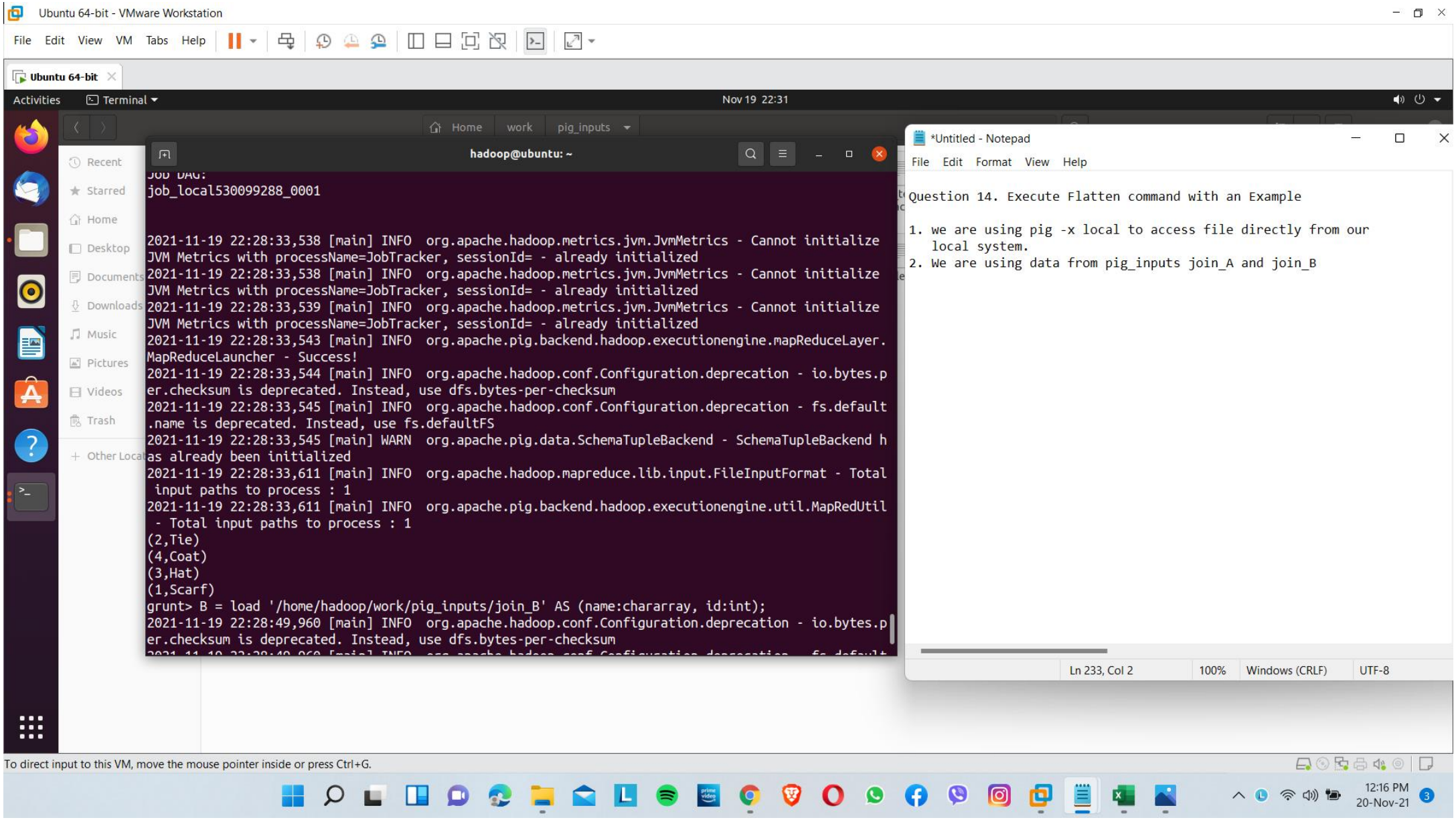

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:42
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
grunt> FullOuterJoin= JOIN customers by id FULL OUTER, orders by custID;
grunt> DUMP FullOuterJoin;
2021-11-19 21:41:07,319 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN
2021-11-19 21:41:07,331 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:41:07,331 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 21:41:07,331 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 21:41:07,333 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 21:41:07,334 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 21:41:07,334 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 21:41:07,339 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:41:07,341 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:41:07,341 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 21:41:07,341 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 21:41:07,341 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Reduce phase detected, estimating # of required reducers.
2021-11-19 21:41:07,342 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 12. Execute Full Join command with an Example.
1. We will use the following code to join Right Outer Join
- FullOuterJoin= JOIN customers by id FULL OUTER, orders by custID;
2. After that we are dumping the results
-DUMP FullOuterJoin
Ln 216, Col 25
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:27 AM
20-Nov-21

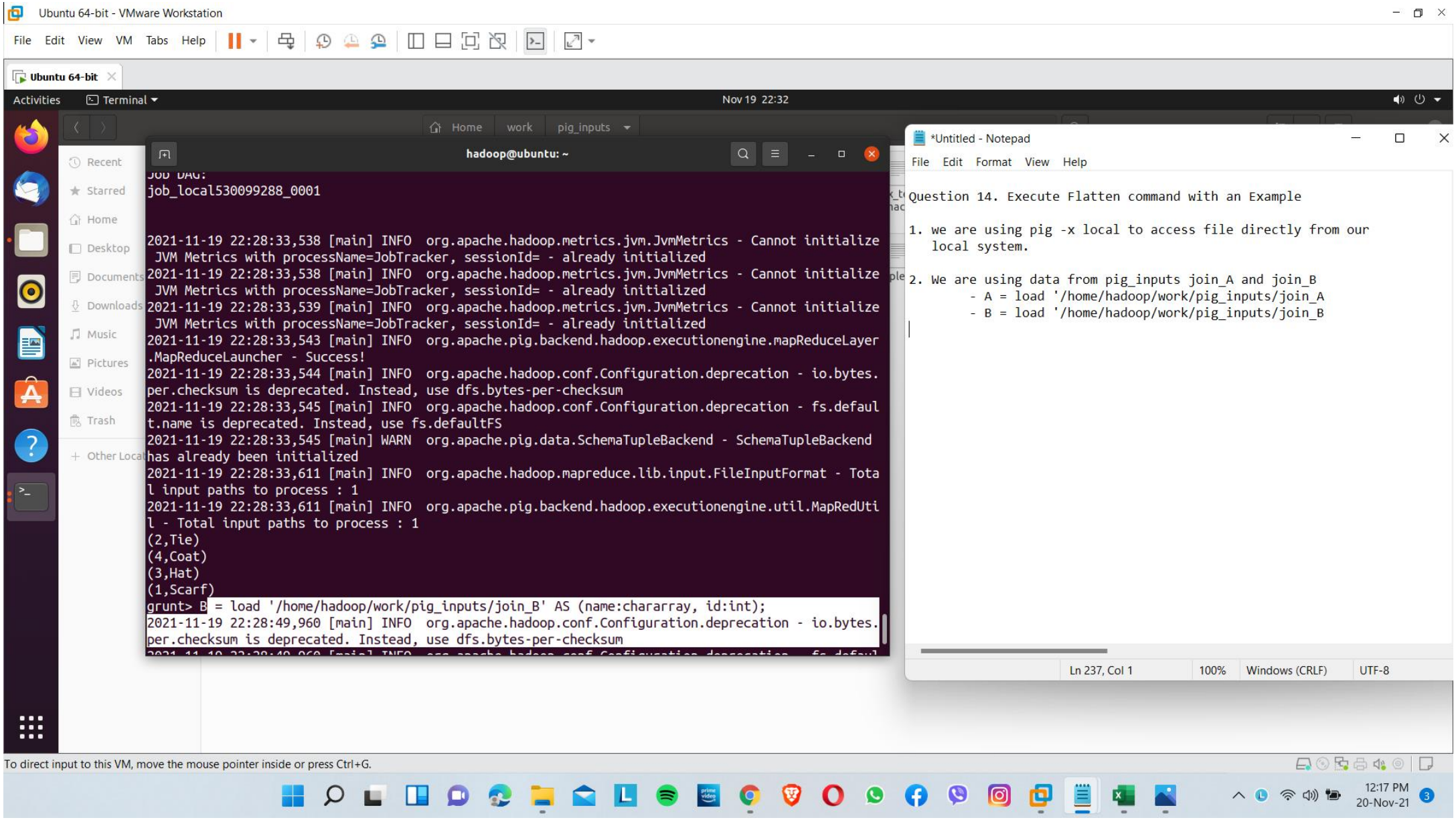

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 21:42
Home
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
hadoop@ubuntu: ~
Job DAG:
job_local1973632842_0006
2021-11-19 21:41:12,969 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:41:12,969 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:41:12,970 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 21:41:12,973 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 21:41:12,974 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 21:41:12,974 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 21:41:12,976 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 21:41:12,976 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Ramesh,32,Ahmedabad,2000,,,,)
(2,Khilan,25,Delhi,1500,101,2009-11-20 00:00:00,2,1560)
(3,kaushik,23,Kota,2000,100,2009-10-08 00:00:00,3,1500)
(3,kaushik,23,Kota,2000,102,2009-10-08 00:00:00,3,3000)
(4,Chaitali,25,Mumbai,6500,103,2008-05-20 00:00:00,4,2060)
(5,Hardik,27,Bhopal,8500,,,,)
(6,Komal,22,MP,4500,,,,)
(7,Muffy,24,Indore,10000,,,,)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 12. Execute Full Join command with an Example.
1. We will use the following code to join Right Outer Join
- FullOuterJoin= JOIN customers by id FULL OUTER, orders by custID;
2. After that we are dumping the results
-DUMP FullOuterJoin
Ln 216, Col 25
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
11:27 AM
20-Nov-21

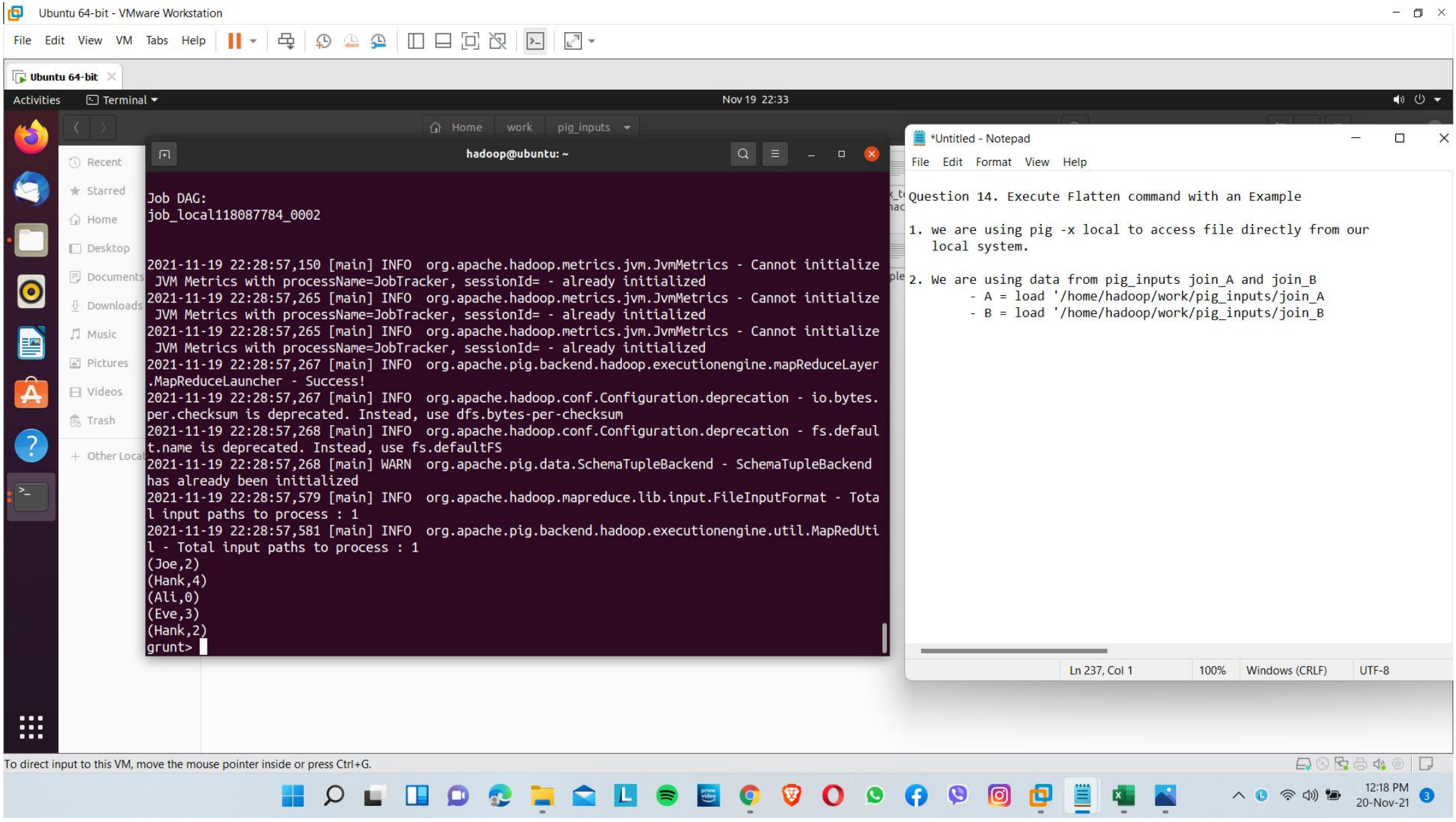

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:17
hadoop@ubuntu: ~
(6,Komal,22,MP,4500,,,,)
(7,Muffy,24,Indore,10000,,,,)
grunt> cogroup_data = COGROUP customers by id, orders by custID;
grunt> DUMP cogroup_data;
2021-11-19 22:15:05,636 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: COGROUP
2021-11-19 22:15:05,648 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:15:05,649 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 22:15:05,649 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 22:15:05,658 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 22:15:05,659 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 22:15:05,659 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 22:15:05,664 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:15:05,664 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:15:05,665 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 22:15:05,666 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 22:15:05,666 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.
*Untitled - Notepad
File Edit Format View Help
Question 13. Execute CoGroup command with an Example.
1. We are cogrouping our data with customer id from customers and orders relations.
- cogroup_data = COGROUP customers by id, orders by custID;
2. After that we dumped the data
- DUMP cogroup_data
Ln 228, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:02 PM
20-Nov-21

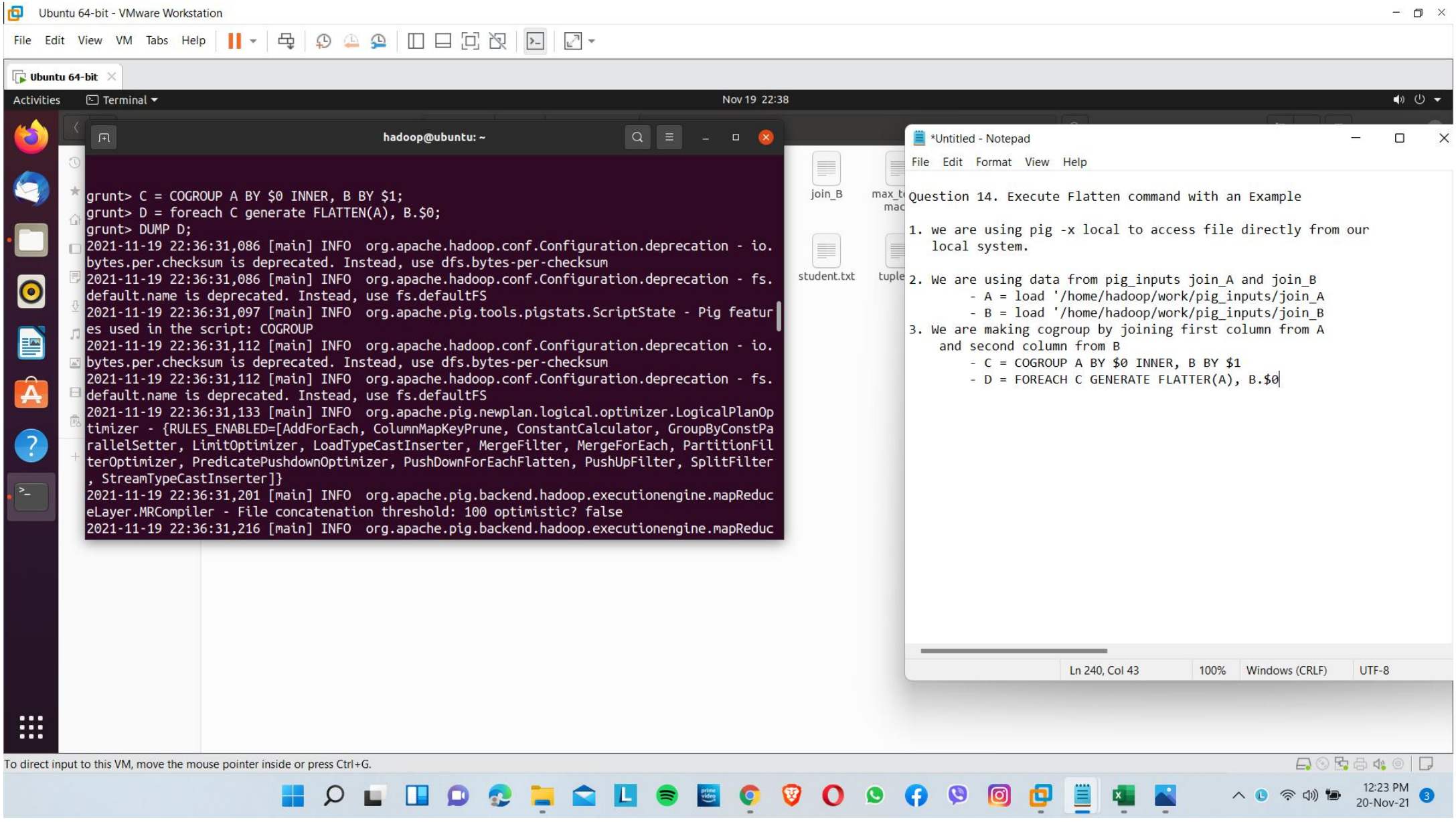

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:17
hadoop@ubuntu: ~
Job DAG:
job_local174416355_0007
2021-11-19 22:15:11,824 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:15:11,825 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 22:15:11,825 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:15:11,825 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 22:15:11,829 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 22:15:11,829 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,{(1,Ramesh,32,Ahmedabad,2000)},{})
(2,{(2,Khilan,25,Delhi,1500)},{(101,2009-11-20 00:00:00,2,1560)})
(3,{(3,kaushik,23,Kota,2000)},{(100,2009-10-08 00:00:00,3,1500),(102,2009-10-08 00:00:00,3,3000)})
(4,{(4,Chaitali,25,Mumbai,6500)},{(103,2008-05-20 00:00:00,4,2060)})
(5,{(5,Hardik,27,Bhopal,8500)},{})
(6,{(6,Komal,22,MP,4500)},{})
(7,{(7,Muffy,24,Indore,10000)},{})
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 13. Execute CoGroup command with an Example.
1. We are cogrouping our data with customer id from customers and orders relations.
- cogroup_data = COGROUP customers by id, orders by custID;
2. After that we dumped the data
- DUMP cogroup_data
Ln 228, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:02 PM
20-Nov-21

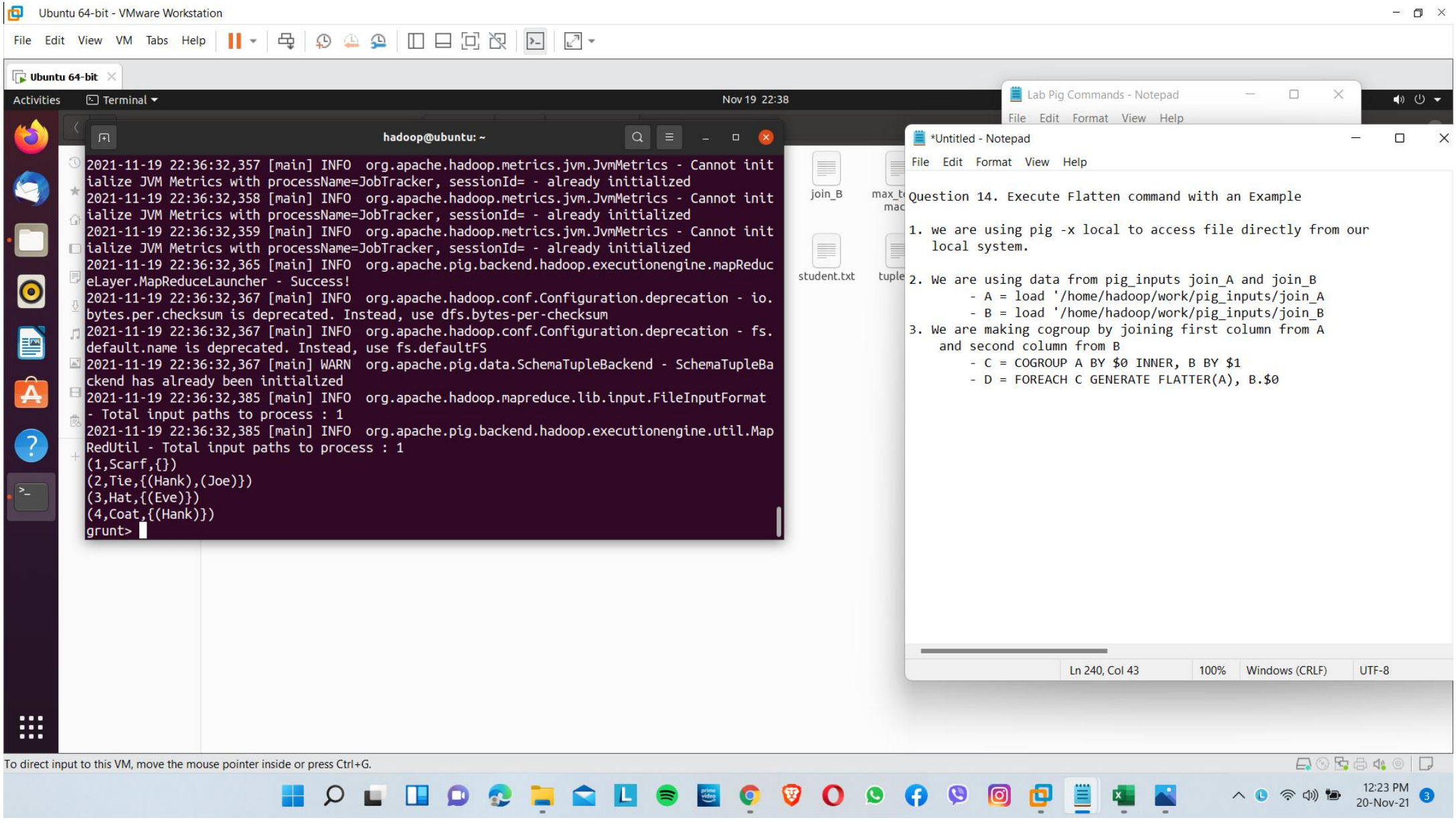

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:31
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
hadoop@ubuntu: ~
grunt>
[7]+ Stopped pig -x mapreduce
hadoop@ubuntu:~$ pig -x local
21/11/19 22:25:30 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/11/19 22:25:30 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2021-11-19 22:25:30,348 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2015, 11:44:35
2021-11-19 22:25:30,348 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1637389530347.log
2021-11-19 22:25:30,378 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup not found
2021-11-19 22:25:30,449 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:25:30,450 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2021-11-19 22:25:30,450 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2021-11-19 22:25:30,515 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
grunt>
grunt> A = load '/home/hadoop/work/pig_inputs/join_A' AS (id:int, name:chararray);
2021-11-19 22:25:48,363 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-19 22:25:48,363 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A = load '/home/hadoop/work/pig_inputs/join_A' AS (id:int, name:chararray);
2021-11-19 22:26:00,578 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
*Untitled - Notepad
File Edit Format View Help
Question 14. Execute Flatten command with an Example
1. we are using pig -x local to access file directly from our local system.
2. We are using data from pig_inputs join_A and join_B
Ln 233, Col 2
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:16 PM
20-Nov-21

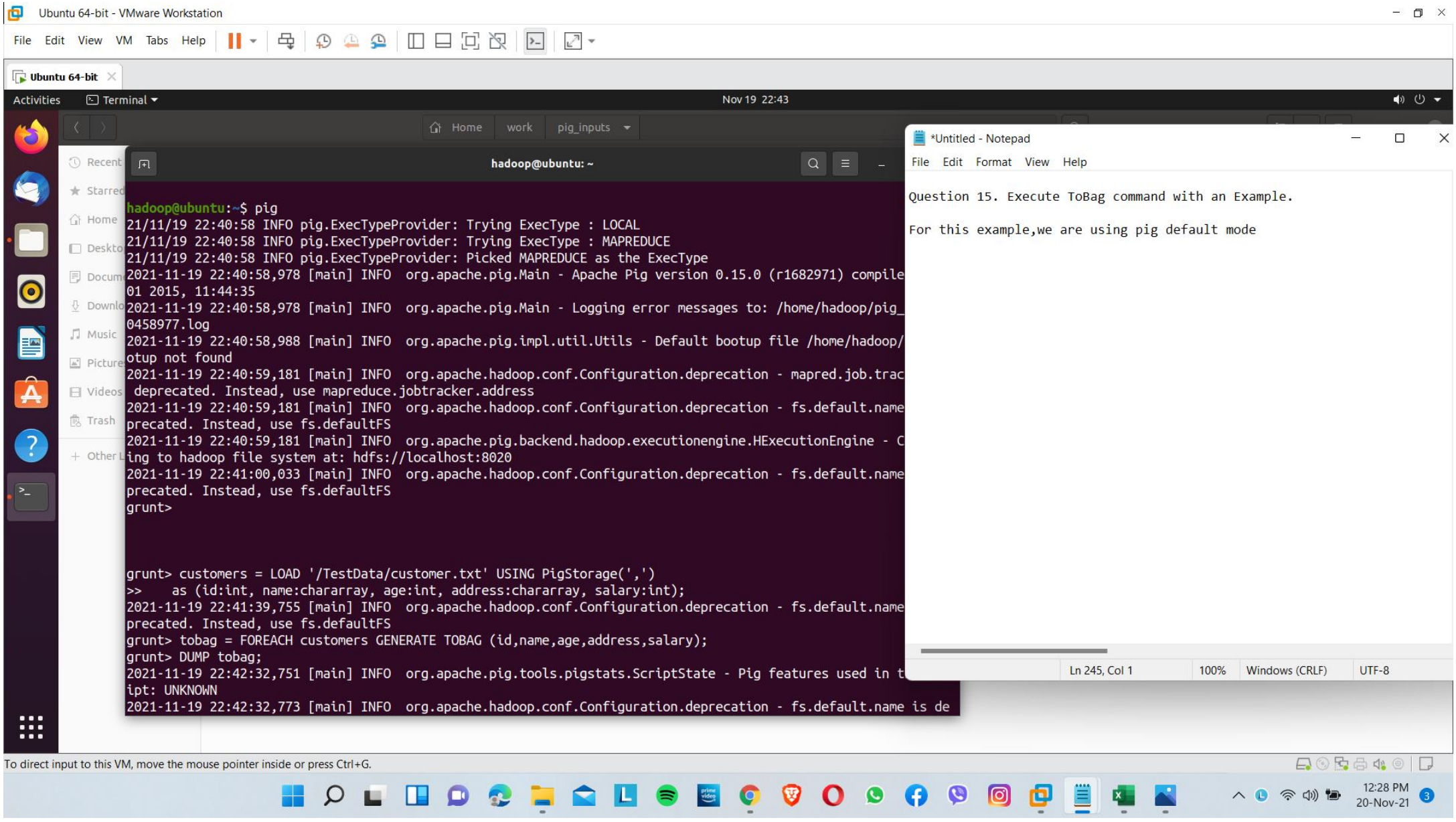

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:31
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
job_local530099288_0001
2021-11-19 22:28:33,538 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:33,538 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:33,539 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:33,543 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 22:28:33,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-19 22:28:33,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:28:33,545 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 22:28:33,611 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 22:28:33,611 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(2,Tie)
(4,Coat)
(3,Hat)
(1,Scarf)
grunt> B = load '/home/hadoop/work/pig_inputs/join_B' AS (name:chararray, id:int);
2021-11-19 22:28:49,960 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
*Untitled - Notepad
File Edit Format View Help
Question 14. Execute Flatten command with an Example
1. we are using pig -x local to access file directly from our local system.
2. We are using data from pig_inputs join_A and join_B
Ln 233, Col 2
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:16 PM
20-Nov-21

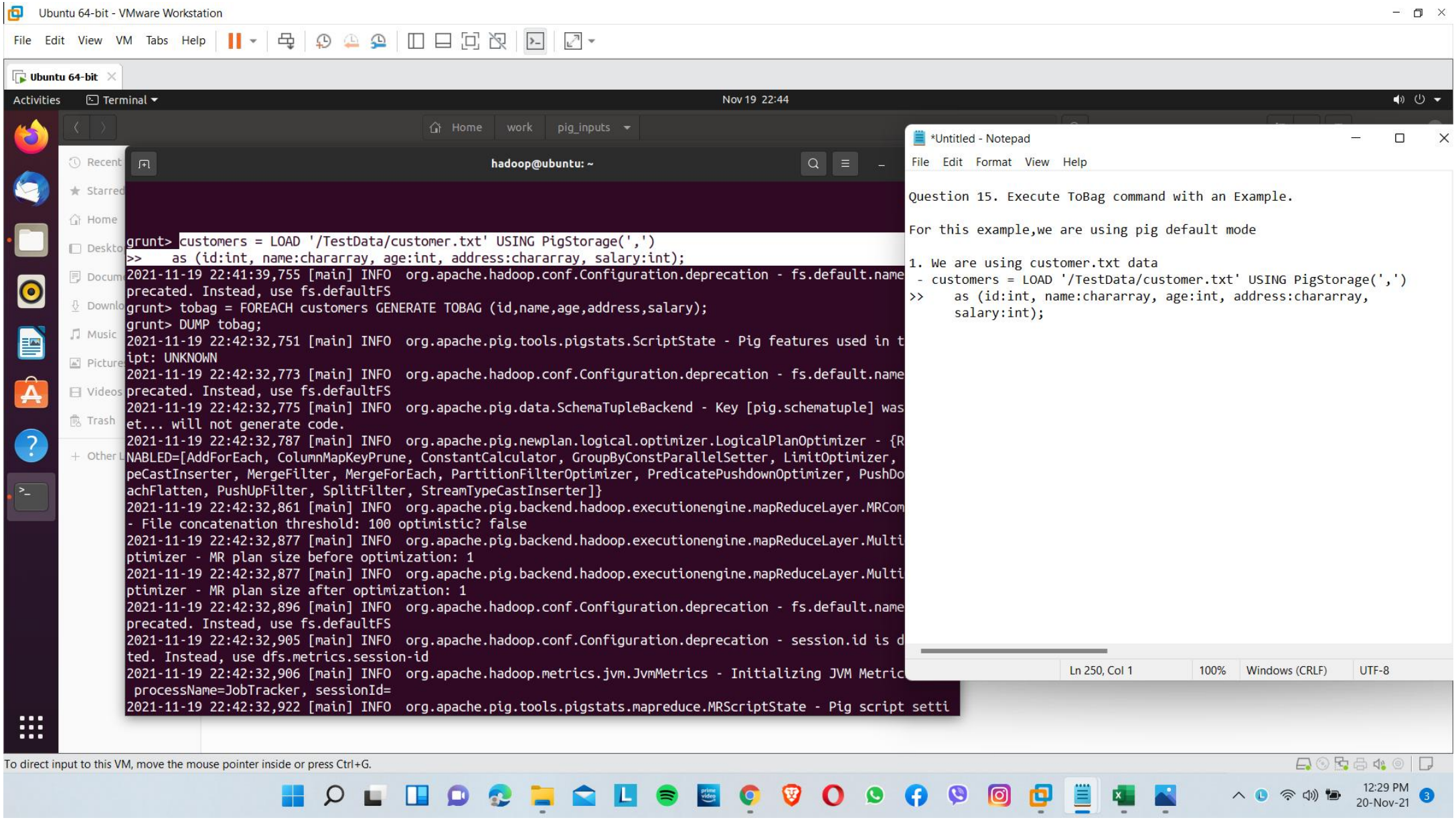

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:32
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
job_local530099288_0001
2021-11-19 22:28:33,538 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:33,538 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:33,539 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:33,543 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 22:28:33,544 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-19 22:28:33,545 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:28:33,545 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 22:28:33,611 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 22:28:33,611 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(2,Tie)
(4,Coat)
(3,Hat)
(1,Scarf)
grunt> B = load '/home/hadoop/work/pig_inputs/join_B' AS (name:chararray, id:int);
2021-11-19 22:28:49,960 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
*Untitled - Notepad
File Edit Format View Help
Question 14. Execute Flatten command with an Example
1. we are using pig -x local to access file directly from our local system.
2. We are using data from pig_inputs join_A and join_B
- A = load '/home/hadoop/work/pig_inputs/join_A
- B = load '/home/hadoop/work/pig_inputs/join_B
Ln 237, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:17 PM
20-Nov-21

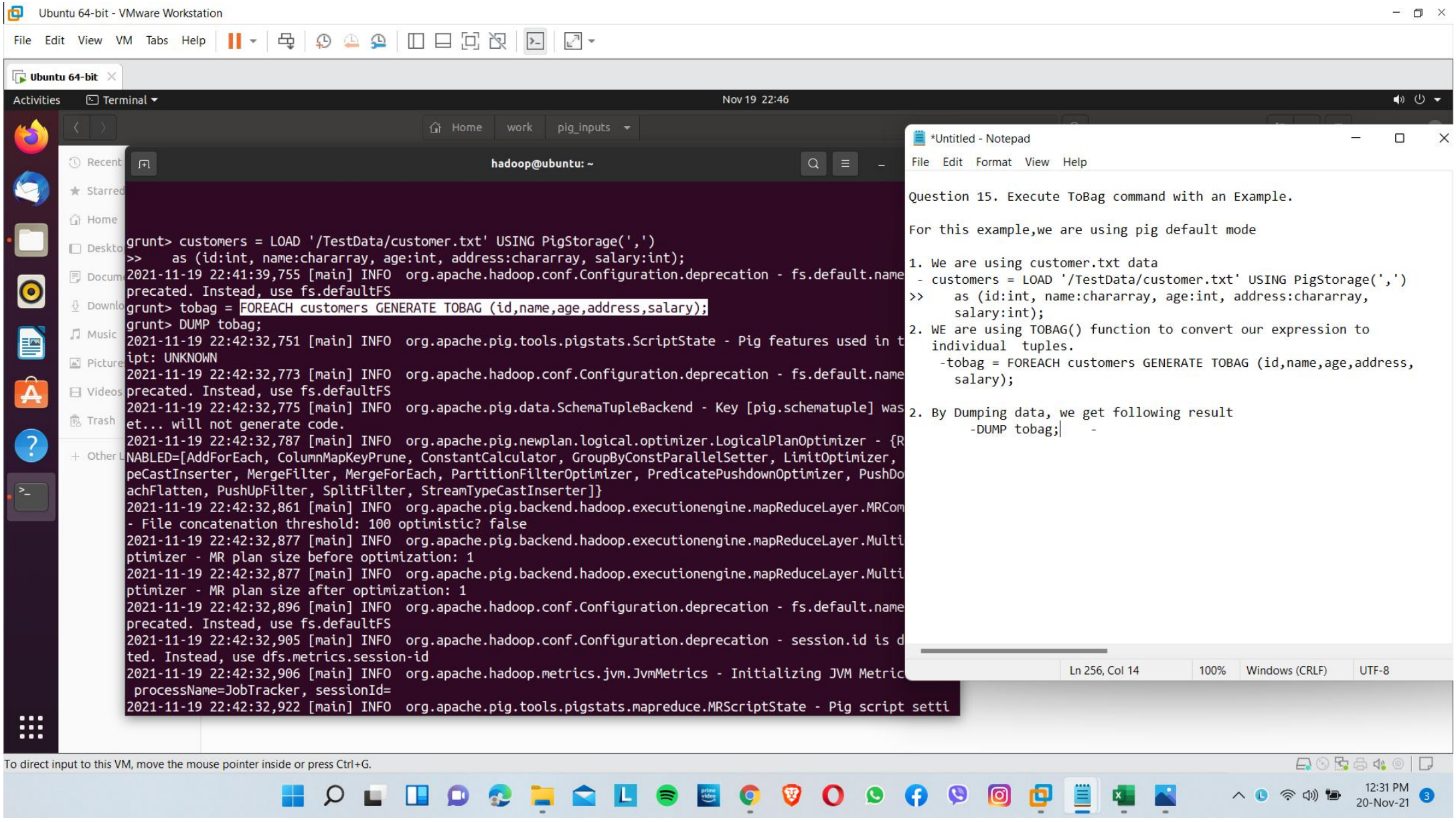

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:33
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locat
hadoop@ubuntu: ~
Job DAG:
job_local118087784_0002
2021-11-19 22:28:57,150 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:57,265 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:57,265 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:28:57,267 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 22:28:57,267 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-19 22:28:57,268 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:28:57,268 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 22:28:57,579 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 22:28:57,581 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Joe,2)
(Hank,4)
(Ali,0)
(Eve,3)
(Hank,2)
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 14. Execute Flatten command with an Example
1. we are using pig -x local to access file directly from our local system.
2. We are using data from pig_inputs join_A and join_B
- A = load '/home/hadoop/work/pig_inputs/join_A
- B = load '/home/hadoop/work/pig_inputs/join_B
Ln 237, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:18 PM
20-Nov-21

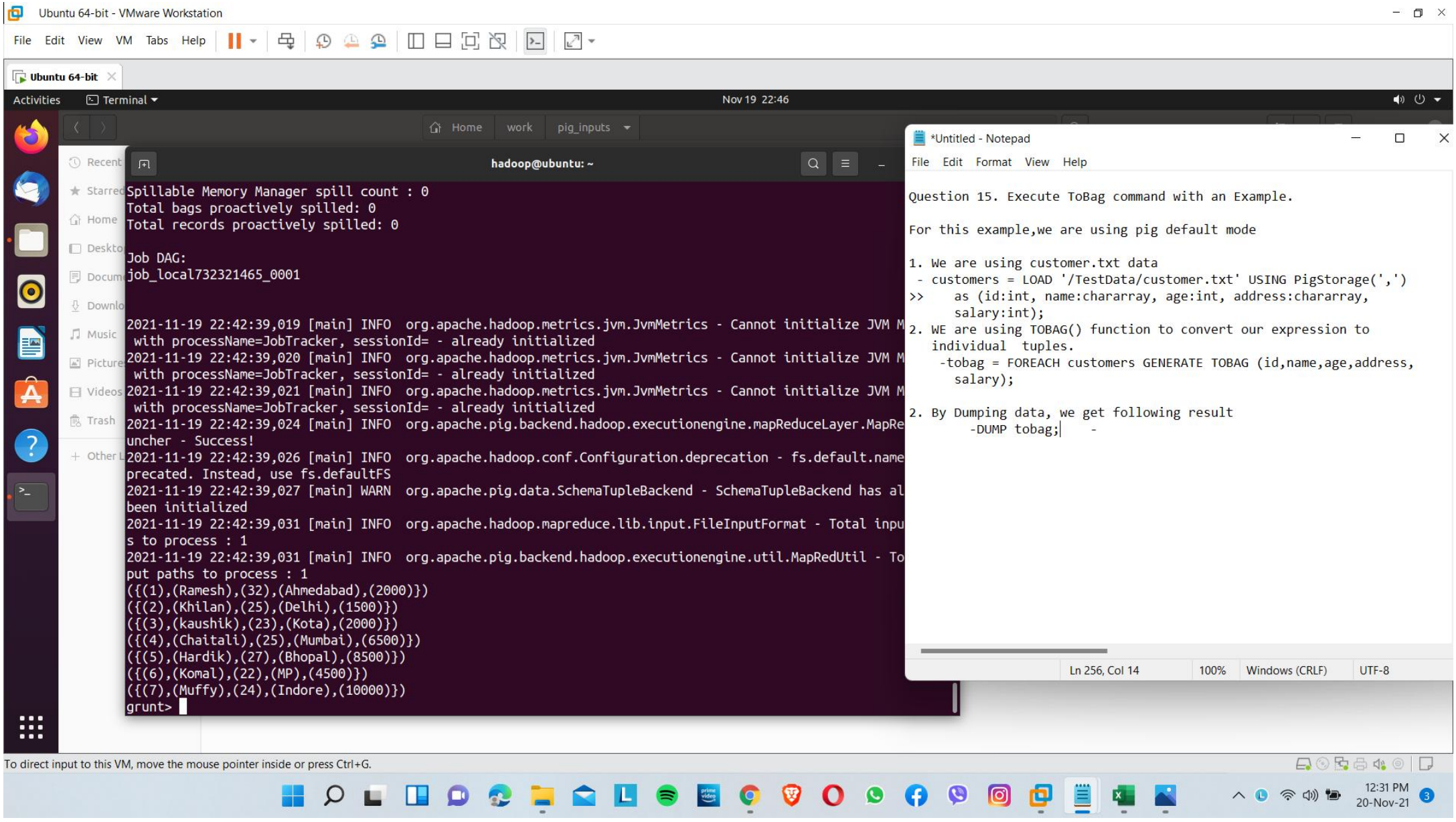

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:38
hadoop@ubuntu: ~
grunt> C = COGROUP A BY $0 INNER, B BY $1;
grunt> D = foreach C generate FLATTEN(A), B.$0;
grunt> DUMP D;
2021-11-19 22:36:31,086 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-19 22:36:31,086 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:36:31,097 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: COGROUP
2021-11-19 22:36:31,112 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-19 22:36:31,112 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:36:31,133 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 22:36:31,201 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 22:36:31,216 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduc
join_B
student.txt
*Untitled - Notepad
File Edit Format View Help
Question 14. Execute Flatten command with an Example
1. we are using pig -x local to access file directly from our local system.
2. We are using data from pig_inputs join_A and join_B
- A = load '/home/hadoop/work/pig_inputs/join_A
- B = load '/home/hadoop/work/pig_inputs/join_B
3. We are making cogroup by joining first column from A and second column from B
- C = COGROUP A BY $0 INNER, B BY $1
- D = FOREACH C GENERATE FLATTER(A), B.$0
Ln 240, Col 43
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:23 PM
20-Nov-21

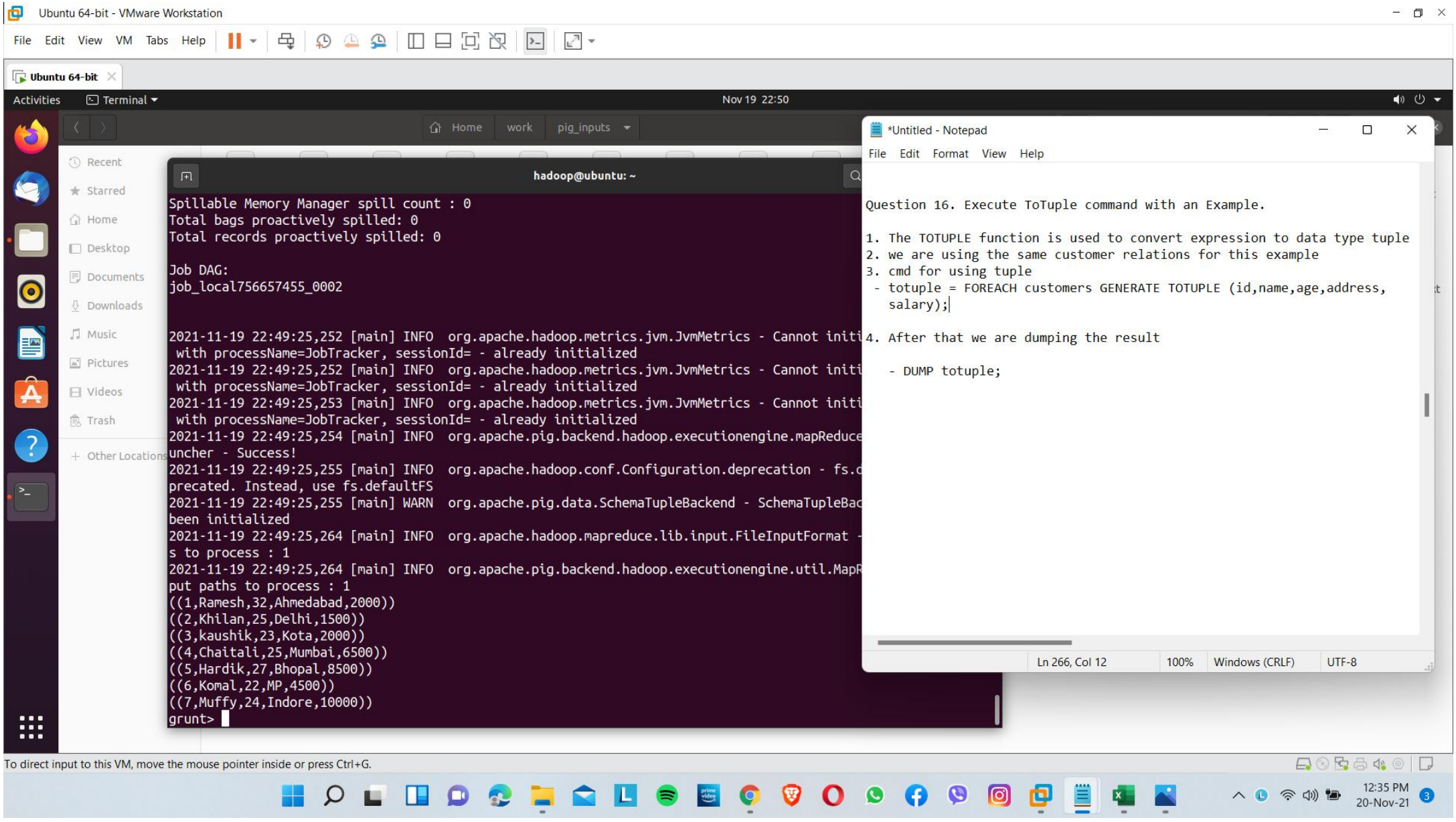

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:38
hadoop@ubuntu: ~
2021-11-19 22:36:32,357 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:36:32,358 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:36:32,359 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:36:32,365 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 22:36:32,367 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-11-19 22:36:32,367 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 22:36:32,367 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 22:36:32,385 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 22:36:32,385 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,Scarf,{})
(2,Tie,{(Hank),(Joe)})
(3,Hat,{(Eve)})
(4,Coat,{(Hank)})
grunt>
join_B
student.txt
Lab Pig Commands - Notepad
File Edit Format View Help
*Untitled - Notepad
File Edit Format View Help
Question 14. Execute Flatten command with an Example
1. we are using pig -x local to access file directly from our local system.
2. We are using data from pig_inputs join_A and join_B
- A = load '/home/hadoop/work/pig_inputs/join_A
- B = load '/home/hadoop/work/pig_inputs/join_B
3. We are making cogroup by joining first column from A and second column from B
- C = COGROUP A BY $0 INNER, B BY $1
- D = FOREACH C GENERATE FLATTER(A), B.$0
Ln 240, Col 43
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:23 PM
20-Nov-21

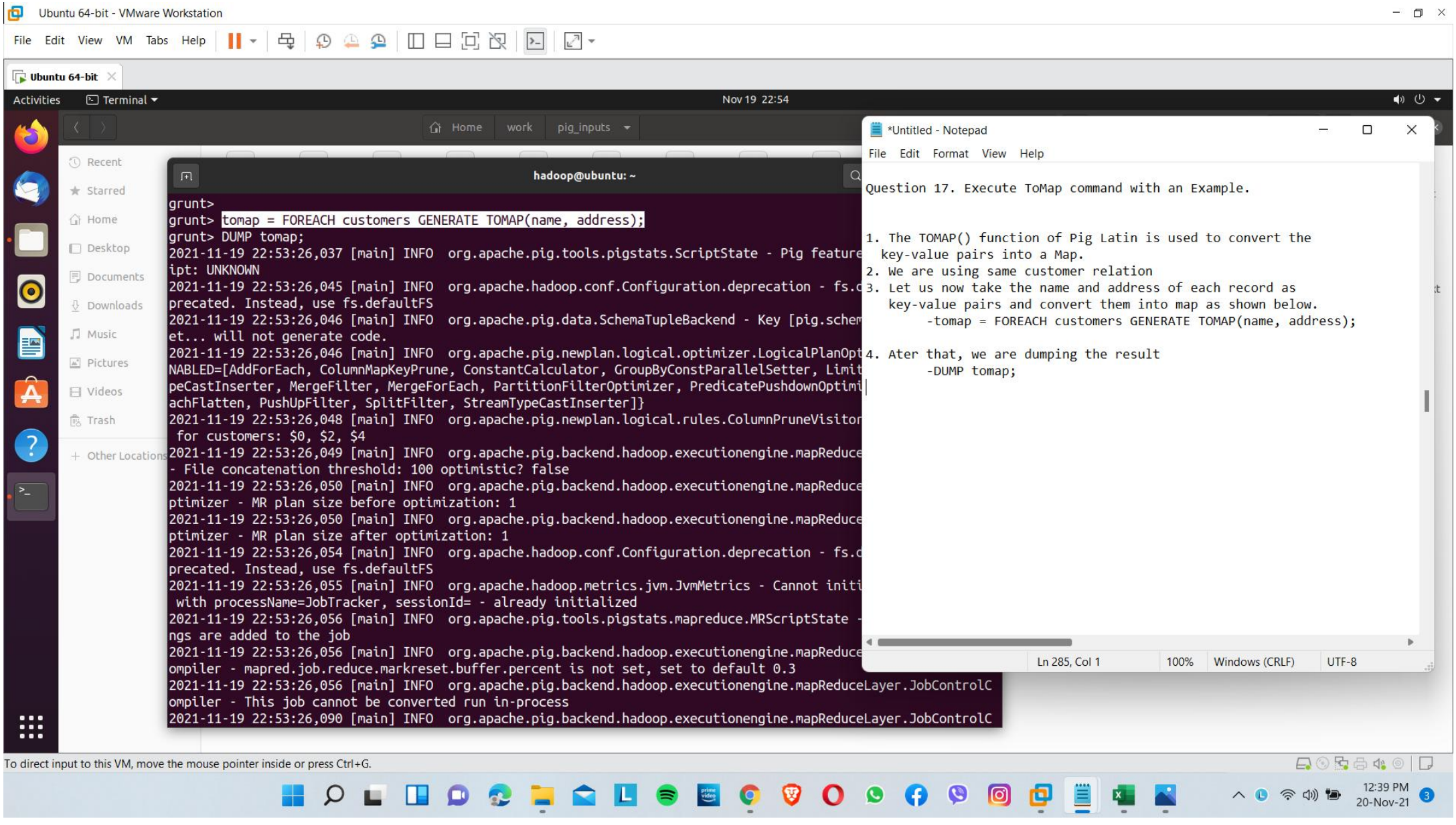

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:43
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
hadoop@ubuntu: ~
hadoop@ubuntu:~$ pig
21/11/19 22:40:58 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
21/11/19 22:40:58 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
21/11/19 22:40:58 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2021-11-19 22:40:58,978 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compile
01 2015, 11:44:35
2021-11-19 22:40:58,978 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_
0458977.log
2021-11-19 22:40:58,988 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/
otup not found
2021-11-19 22:40:59,181 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.trac
deprecated. Instead, use mapreduce.jobtracker.address
2021-11-19 22:40:59,181 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
2021-11-19 22:40:59,181 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - C
ing to hadoop file system at: hdfs://localhost:8020
2021-11-19 22:41:00,033 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
grunt>
grunt> customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2021-11-19 22:41:39,755 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
grunt> tobag = FOREACH customers GENERATE TOBAG (id,name,age,address,salary);
grunt> DUMP tobag;
2021-11-19 22:42:32,751 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in t
ipt: UNKNOWN
2021-11-19 22:42:32,773 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is de
*Untitled - Notepad
File Edit Format View Help
Question 15. Execute ToBag command with an Example.
For this example,we are using pig default mode
Ln 245, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:28 PM
20-Nov-21

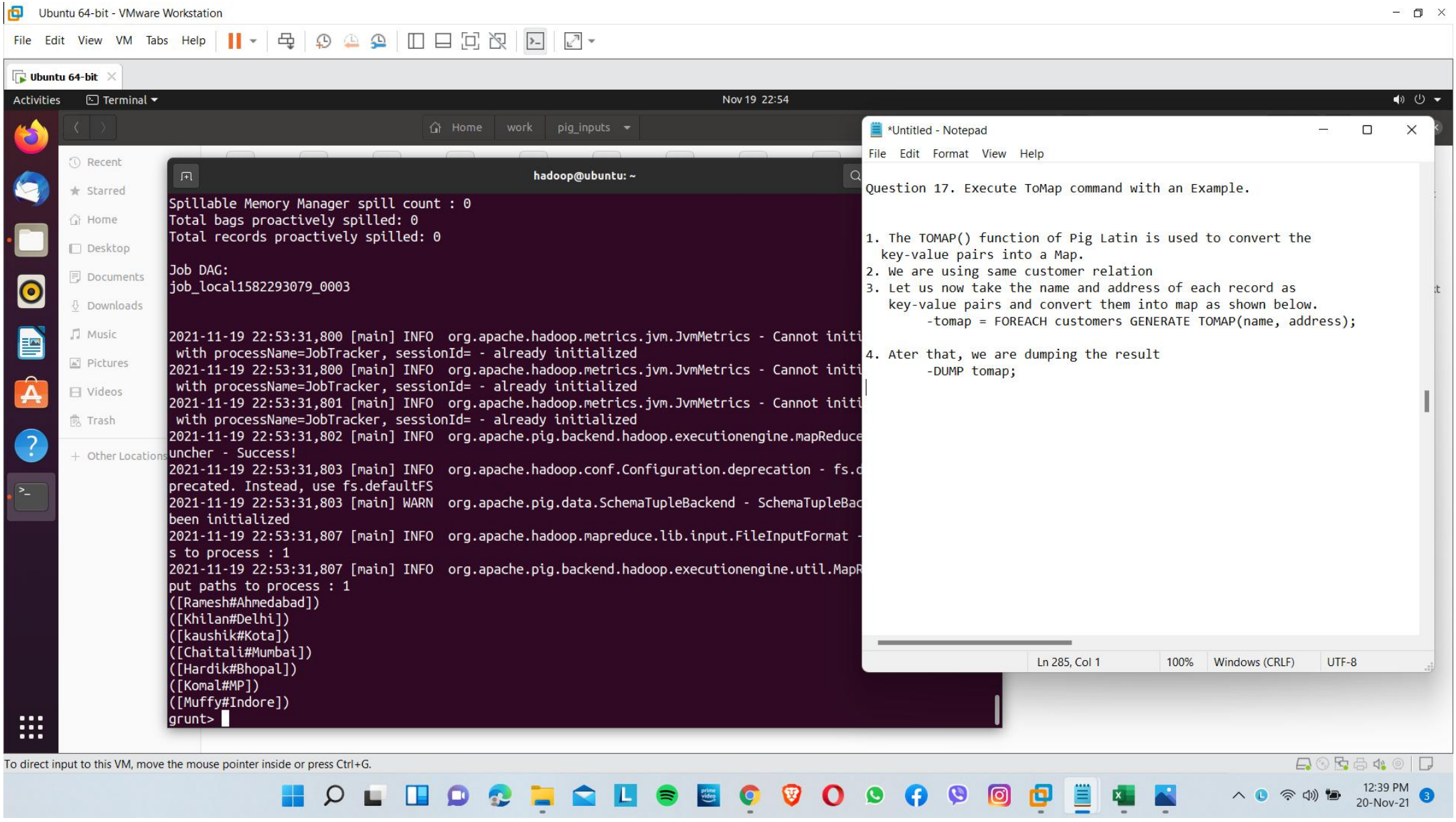

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:44
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
hadoop@ubuntu: ~
grunt> customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2021-11-19 22:41:39,755 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
grunt> tobag = FOREACH customers GENERATE TOBAG (id,name,age,address,salary);
grunt> DUMP tobag;
2021-11-19 22:42:32,751 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in t
ipt: UNKNOWN
2021-11-19 22:42:32,773 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
2021-11-19 22:42:32,775 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was
et... will not generate code.
2021-11-19 22:42:32,787 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {R
NABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer,
peCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDo
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 22:42:32,861 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCom
- File concatenation threshold: 100 optimistic? false
2021-11-19 22:42:32,877 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.Multi
ptimizer - MR plan size before optimization: 1
2021-11-19 22:42:32,877 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.Multi
ptimizer - MR plan size after optimization: 1
2021-11-19 22:42:32,896 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
2021-11-19 22:42:32,905 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is d
ted. Instead, use dfs.metrics.session-id
2021-11-19 22:42:32,906 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metric
processName=JobTracker, sessionId=
2021-11-19 22:42:32,922 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script setti
*Untitled - Notepad
File Edit Format View Help
Question 15. Execute ToBag command with an Example.
For this example,we are using pig default mode
1. We are using customer.txt data
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray,
salary:int);
Ln 250, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:29 PM
20-Nov-21

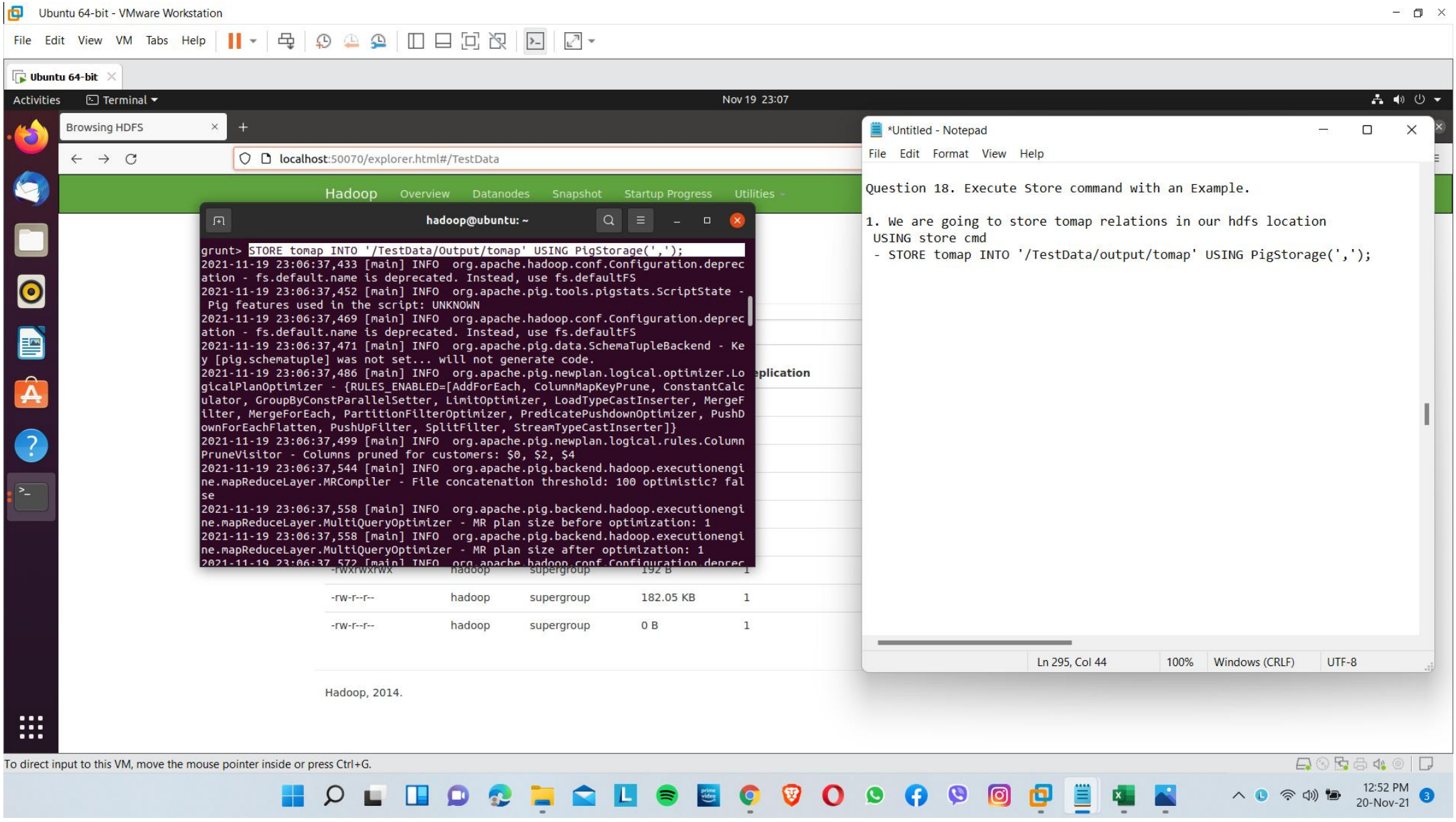

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:46
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
hadoop@ubuntu: ~
grunt> customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2021-11-19 22:41:39,755 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
grunt> tobag = FOREACH customers GENERATE TOBAG (id,name,age,address,salary);
grunt> DUMP tobag;
2021-11-19 22:42:32,751 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in t
ipt: UNKNOWN
2021-11-19 22:42:32,773 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
2021-11-19 22:42:32,775 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was
et... will not generate code.
2021-11-19 22:42:32,787 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {R
NABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer,
peCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDo
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 22:42:32,861 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCom
- File concatenation threshold: 100 optimistic? false
2021-11-19 22:42:32,877 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.Multi
ptimizer - MR plan size before optimization: 1
2021-11-19 22:42:32,877 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.Multi
ptimizer - MR plan size after optimization: 1
2021-11-19 22:42:32,896 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
2021-11-19 22:42:32,905 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is d
ted. Instead, use dfs.metrics.session-id
2021-11-19 22:42:32,906 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metric
processName=JobTracker, sessionId=
2021-11-19 22:42:32,922 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script setti
*Untitled - Notepad
File Edit Format View Help
Question 15. Execute ToBag command with an Example.
For this example,we are using pig default mode
1. We are using customer.txt data
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray,
salary:int);
2. WE are using TOBAG() function to convert our expression to
individual tuples.
-tobag = FOREACH customers GENERATE TOBAG (id,name,age,address,
salary);
2. By Dumping data, we get following result
-DUMP tobag; -
Ln 256, Col 14
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:31 PM
20-Nov-21

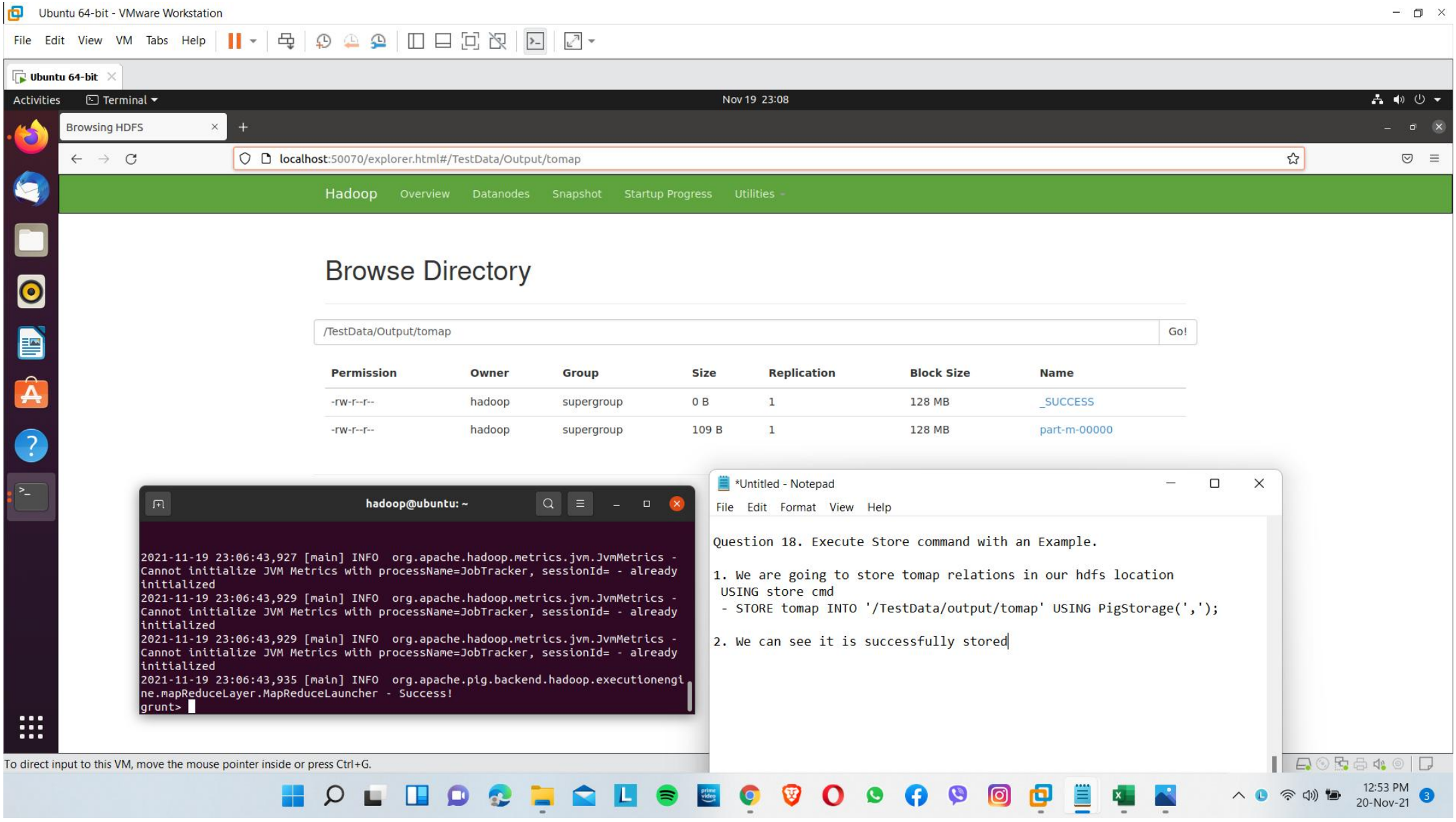

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:46
Home work pig_inputs
hadoop@ubuntu: ~
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local732321465_0001
2021-11-19 22:42:39,019 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM M
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:42:39,020 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM M
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:42:39,021 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM M
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:42:39,024 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapRe
uncher - Success!
2021-11-19 22:42:39,026 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name
precated. Instead, use fs.defaultFS
2021-11-19 22:42:39,027 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has al
been initialized
2021-11-19 22:42:39,031 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total inpu
s to process : 1
2021-11-19 22:42:39,031 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - To
put paths to process : 1
({(1),(Ramesh),(32),(Ahmedabad),(2000)})
({(2),(Khilan),(25),(Delhi),(1500)})
({(3),(kaushik),(23),(Kota),(2000)})
({(4),(Chaitali),(25),(Mumbai),(6500)})
({(5),(Hardik),(27),(Bhopal),(8500)})
({(6),(Komal),(22),(MP),(4500)})
({(7),(Muffy),(24),(Indore),(10000)})
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 15. Execute ToBag command with an Example.
For this example,we are using pig default mode
1. We are using customer.txt data
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray,
salary:int);
2. WE are using TOBAG() function to convert our expression to
individual tuples.
-tobag = FOREACH customers GENERATE TOBAG (id,name,age,address,
salary);
2. By Dumping data, we get following result
-DUMP tobag;
Ln 256, Col 14
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:31 PM
20-Nov-21

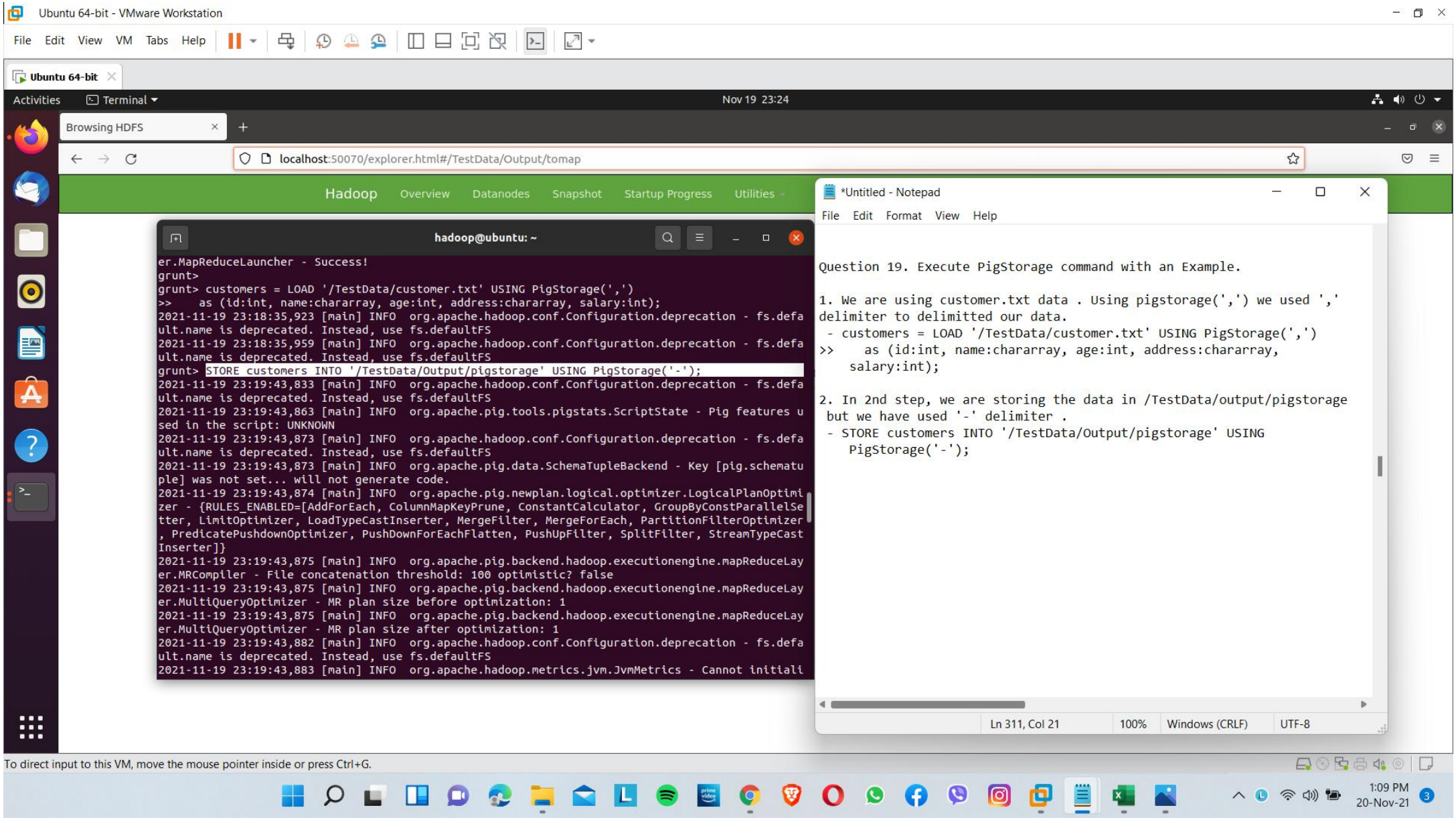

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:50
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
hadoop@ubuntu: ~
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local756657455_0002
2021-11-19 22:49:25,252 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initi
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:49:25,252 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initi
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:49:25,253 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initi
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:49:25,254 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduce
uncher - Success!
2021-11-19 22:49:25,255 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.d
precated. Instead, use fs.defaultFS
2021-11-19 22:49:25,255 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBac
been initialized
2021-11-19 22:49:25,264 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat -
s to process : 1
2021-11-19 22:49:25,264 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapR
put paths to process : 1
((1,Ramesh,32,Ahmedabad,2000))
((2,Khilan,25,Delhi,1500))
((3,kaushik,23,Kota,2000))
((4,Chaitali,25,Mumbai,6500))
((5,Hardik,27,Bhopal,8500))
((6,Komal,22,MP,4500))
((7,Muffy,24,Indore,10000))
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 16. Execute ToTuple command with an Example.
1. The TOTUPLE function is used to convert expression to data type tuple
2. we are using the same customer relations for this example
3. cmd for using tuple
- totuple = FOREACH customers GENERATE TOTUPLE (id,name,age,address,
salary);
4. After that we are dumping the result
- DUMP totuple;
Ln 266, Col 12
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:35 PM
20-Nov-21

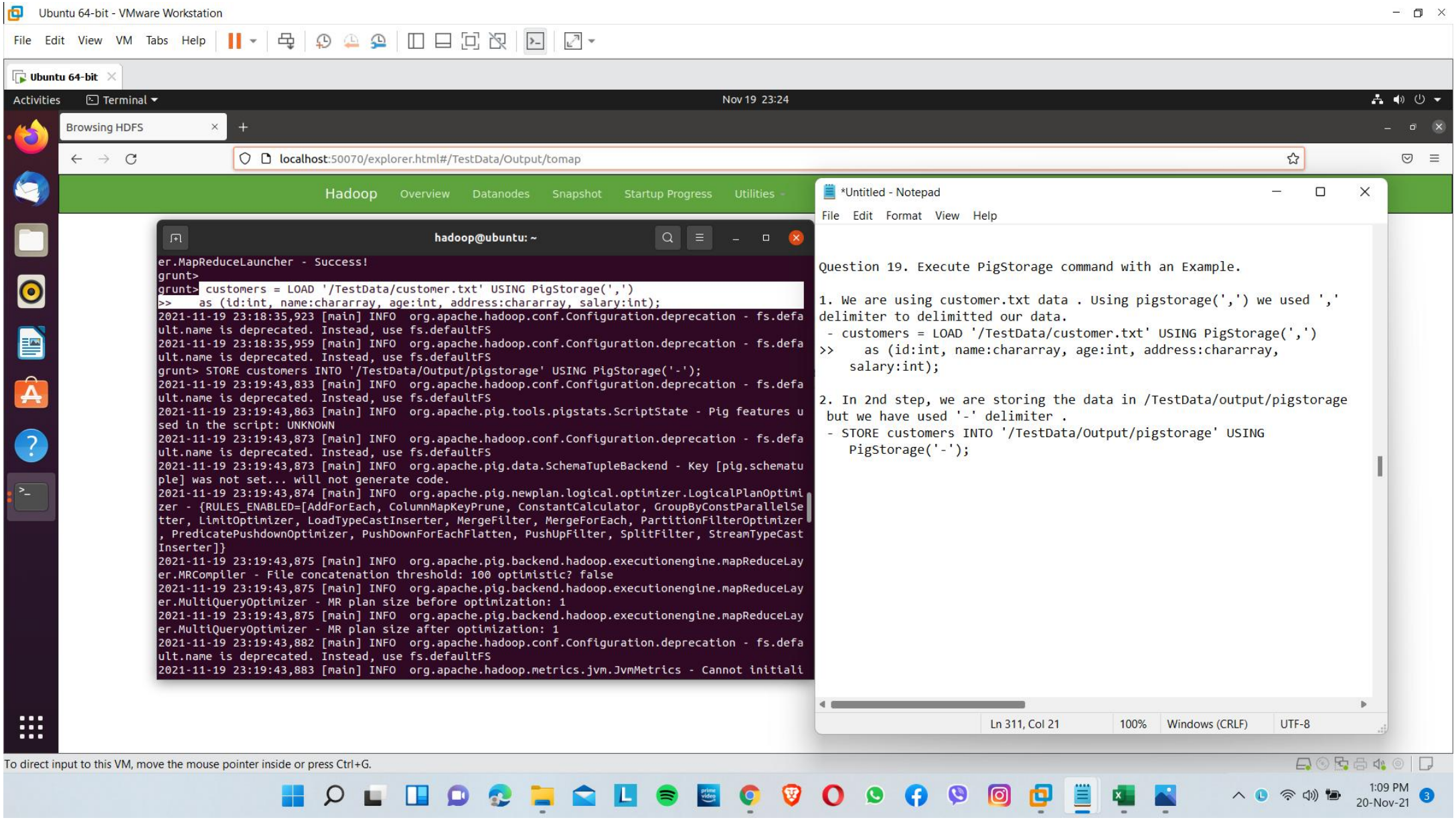

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:54
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
hadoop@ubuntu: ~
grunt>
grunt> tomap = FOREACH customers GENERATE TOMAP(name, address);
grunt> DUMP tomap;
2021-11-19 22:53:26,037 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig feature
ipt: UNKNOWN
2021-11-19 22:53:26,045 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.d
precated. Instead, use fs.defaultFS
2021-11-19 22:53:26,046 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schem
et... will not generate code.
2021-11-19 22:53:26,046 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOpt
NABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, Limit
peCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimi
achFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 22:53:26,048 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor
for customers: $0, $2, $4
2021-11-19 22:53:26,049 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduce
- File concatenation threshold: 100 optimistic? false
2021-11-19 22:53:26,050 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduce
ptimizer - MR plan size before optimization: 1
2021-11-19 22:53:26,050 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduce
ptimizer - MR plan size after optimization: 1
2021-11-19 22:53:26,054 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.d
precated. Instead, use fs.defaultFS
2021-11-19 22:53:26,055 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initi
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:53:26,056 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState -
ngs are added to the job
2021-11-19 22:53:26,056 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduce
ompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 22:53:26,056 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlC
ompiler - This job cannot be converted run in-process
2021-11-19 22:53:26,090 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlC
*Untitled - Notepad
File Edit Format View Help
Question 17. Execute ToMap command with an Example.
1. The TOMAP() function of Pig Latin is used to convert the
key-value pairs into a Map.
2. We are using same customer relation
3. Let us now take the name and address of each record as
key-value pairs and convert them into map as shown below.
-tomap = FOREACH customers GENERATE TOMAP(name, address);
4. Ater that, we are dumping the result
-DUMP tomap;
Ln 285, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:39 PM
20-Nov-21

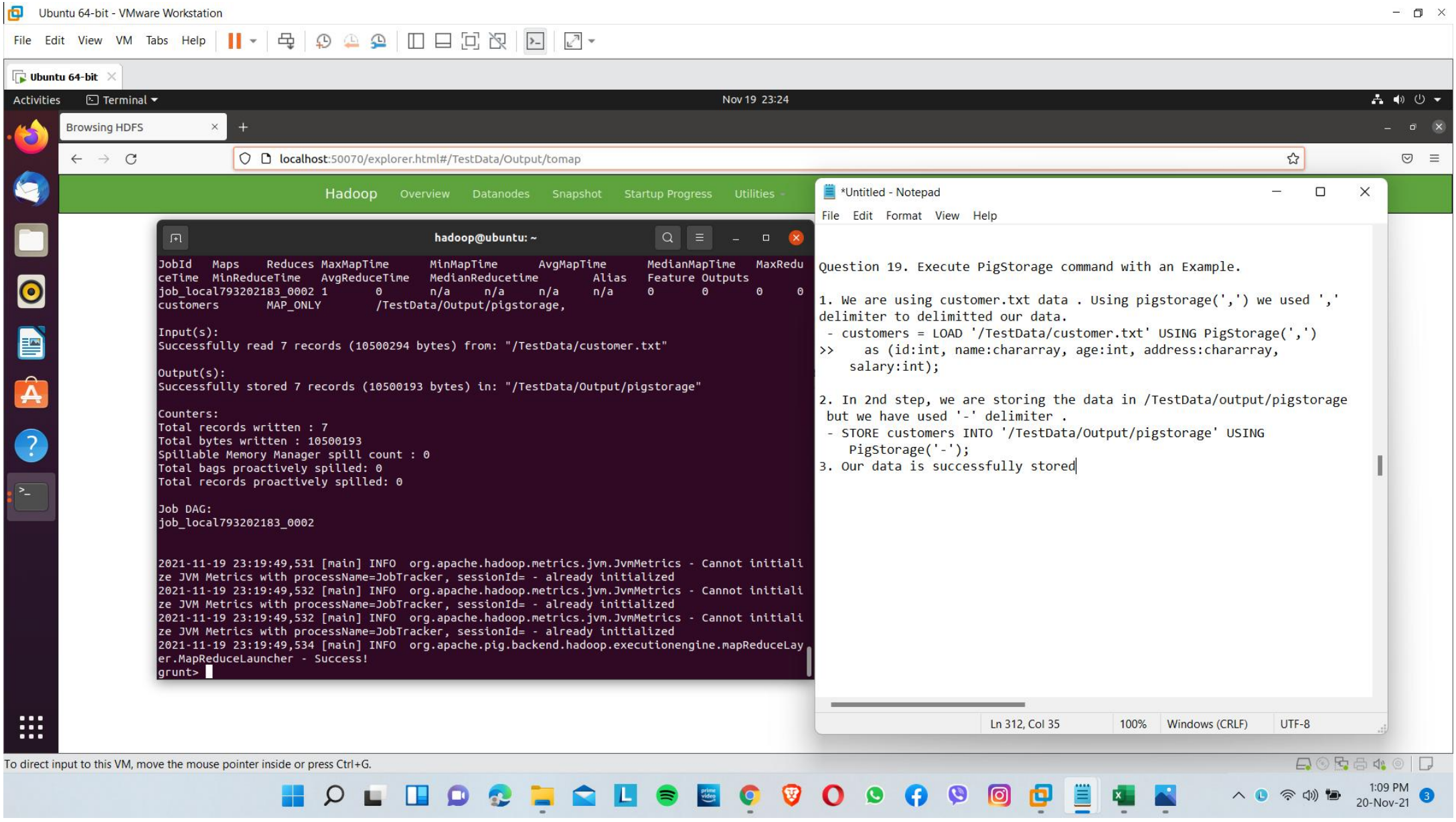

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 22:54
Home
work
pig_inputs
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
hadoop@ubuntu: ~
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1582293079_0003
2021-11-19 22:53:31,800 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initi
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:53:31,800 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initi
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:53:31,801 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initi
with processName=JobTracker, sessionId= - already initialized
2021-11-19 22:53:31,802 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduce
uncher - Success!
2021-11-19 22:53:31,803 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.d
precated. Instead, use fs.defaultFS
2021-11-19 22:53:31,803 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBac
been initialized
2021-11-19 22:53:31,807 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat -
s to process : 1
2021-11-19 22:53:31,807 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapR
put paths to process : 1
([Ramesh#Ahmedabad])
([Khilan#Delhi])
([kaushik#Kota])
([Chaitali#Mumbai])
([Hardik#Bhopal])
([Komal#MP])
([Muffy#Indore])
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 17. Execute ToMap command with an Example.
1. The TOMAP() function of Pig Latin is used to convert the
key-value pairs into a Map.
2. We are using same customer relation
3. Let us now take the name and address of each record as
key-value pairs and convert them into map as shown below.
-tomap = FOREACH customers GENERATE TOMAP(name, address);
4. Ater that, we are dumping the result
-DUMP tomap;
Ln 285, Col 1
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:39 PM
20-Nov-21

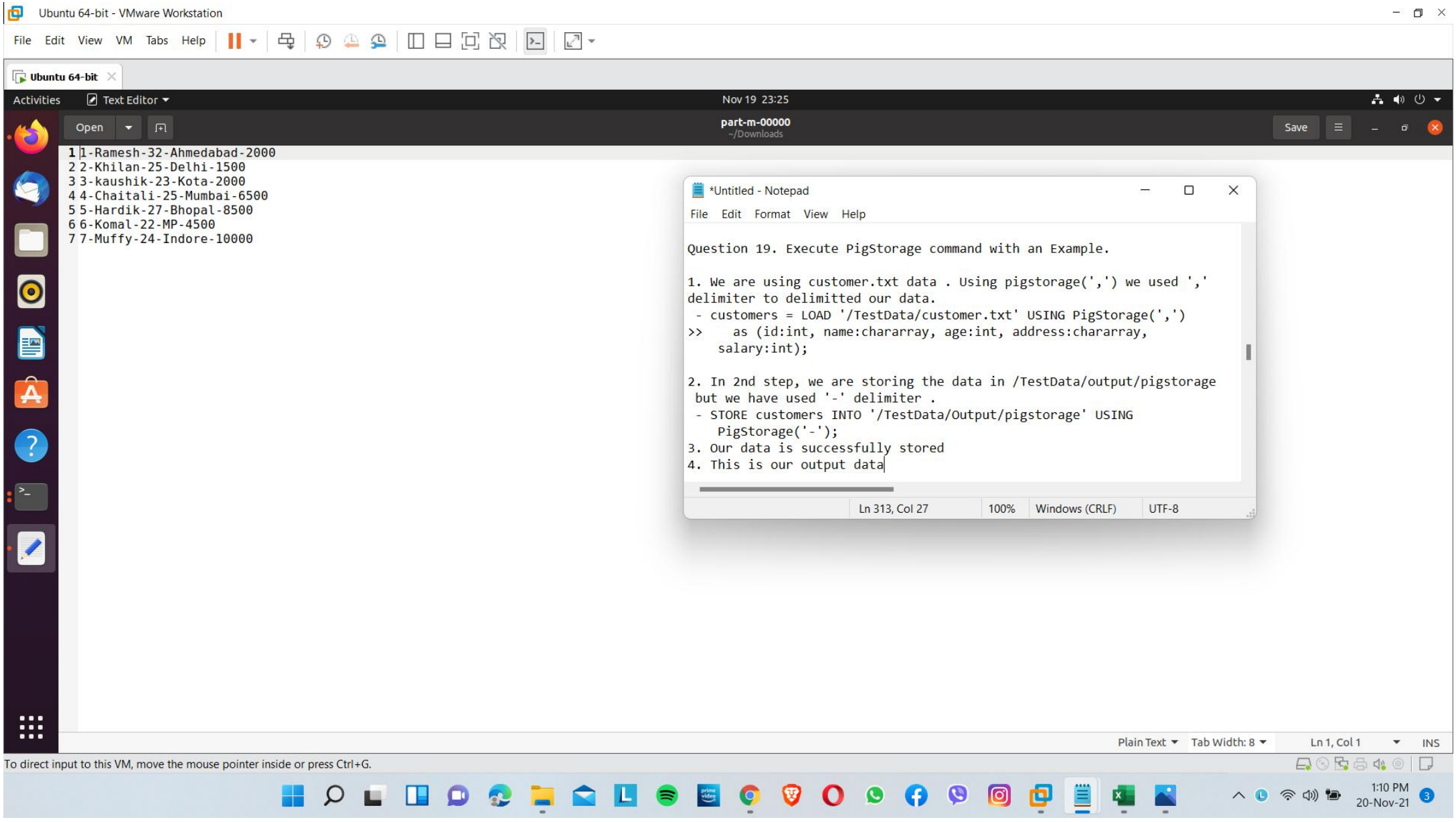

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 23:07
Browsing HDFS
localhost:50070/explorer.html#/TestData
Hadoop
Overview
Datanodes
Snapshot
Startup Progress
Utilities
hadoop@ubuntu: ~
grunt> STORE tomap INTO '/TestData/Output/tomap' USING PigStorage(',');
2021-11-19 23:06:37,433 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:06:37,452 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2021-11-19 23:06:37,469 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:06:37,471 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 23:06:37,486 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 23:06:37,499 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for customers: $0, $2, $4
2021-11-19 23:06:37,544 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 23:06:37,558 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 23:06:37,558 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
-rwxrwxrwx hadoop supergroup 192 B 1
-rw-r--r-- hadoop supergroup 182.05 KB 1
-rw-r--r-- hadoop supergroup 0 B 1
Hadoop, 2014.
*Untitled - Notepad
File Edit Format View Help
Question 18. Execute Store command with an Example.
1. We are going to store tomap relations in our hdfs location
USING store cmd
- STORE tomap INTO '/TestData/output/tomap' USING PigStorage(',');
Ln 295, Col 44
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:52 PM
20-Nov-21

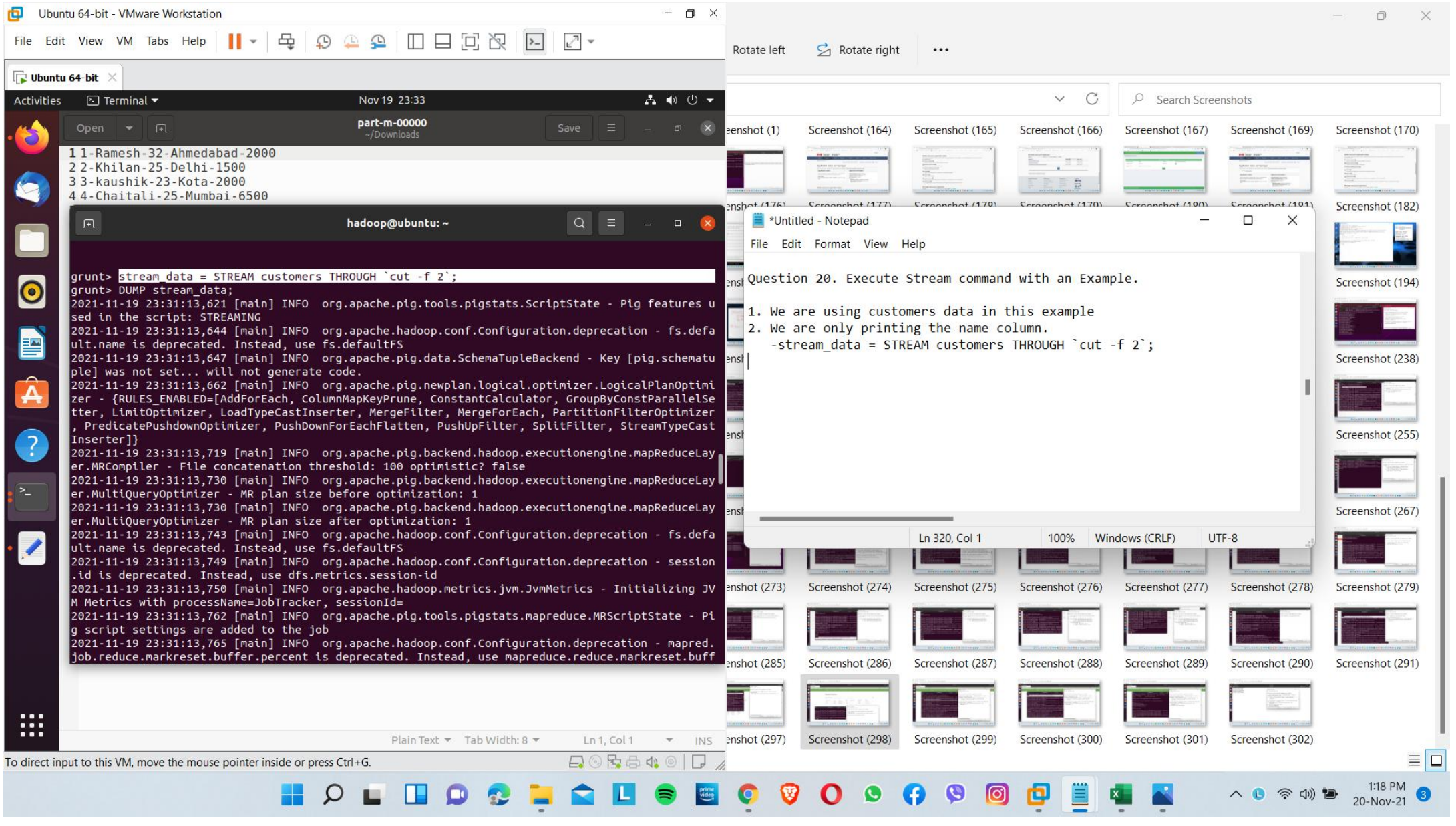

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 23:08
Browsing HDFS
localhost:50070/explorer.html#/TestData/Output/tomap
Hadoop
Overview
Datanodes
Snapshot
Startup Progress
Utilities
Browse Directory
/TestData/Output/tomap
Go!
Permission
Owner
Group
Size
Replication
Block Size
Name
-rw-r--r--
hadoop
supergroup
0 B
1
128 MB
_SUCCESS
-rw-r--r--
hadoop
supergroup
109 B
1
128 MB
part-m-00000
hadoop@ubuntu: ~
2021-11-19 23:06:43,927 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:06:43,929 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:06:43,929 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:06:43,935 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 18. Execute Store command with an Example.
1. We are going to store tomap relations in our hdfs location
USING store cmd
- STORE tomap INTO '/TestData/output/tomap' USING PigStorage(',');
2. We can see it is successfully stored
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
12:53 PM
20-Nov-21

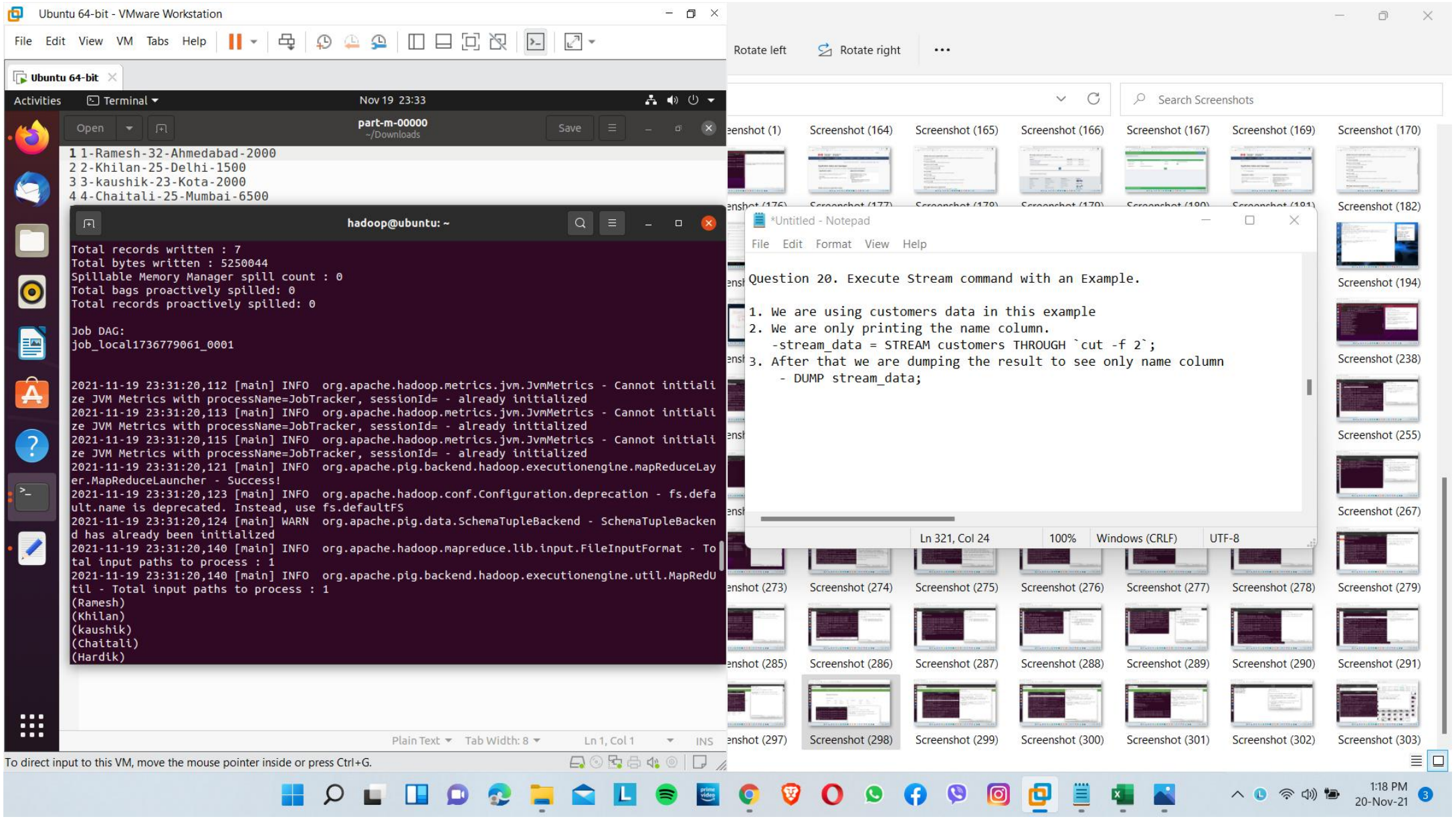

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 23:24
Browsing HDFS
localhost:50070/explorer.html#/TestData/Output/tomap
Hadoop
Overview
Datanodes
Snapshot
Startup Progress
Utilities
hadoop@ubuntu: ~
er.MapReduceLauncher - Success!
grunt>
grunt> customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2021-11-19 23:18:35,923 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:18:35,959 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> STORE customers INTO '/TestData/Output/pigstorage' USING PigStorage('-');
2021-11-19 23:19:43,833 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:19:43,863 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2021-11-19 23:19:43,873 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:19:43,873 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 23:19:43,874 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 23:19:43,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 23:19:43,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 23:19:43,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 23:19:43,882 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:19:43,883 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initiali
*Untitled - Notepad
File Edit Format View Help
Question 19. Execute PigStorage command with an Example.
1. We are using customer.txt data . Using pigstorage(',') we used ',' delimiter to delimitted our data.
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2. In 2nd step, we are storing the data in /TestData/output/pigstorage but we have used '-' delimiter .
- STORE customers INTO '/TestData/Output/pigstorage' USING PigStorage('-');
Ln 311, Col 21
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
1:09 PM
20-Nov-21

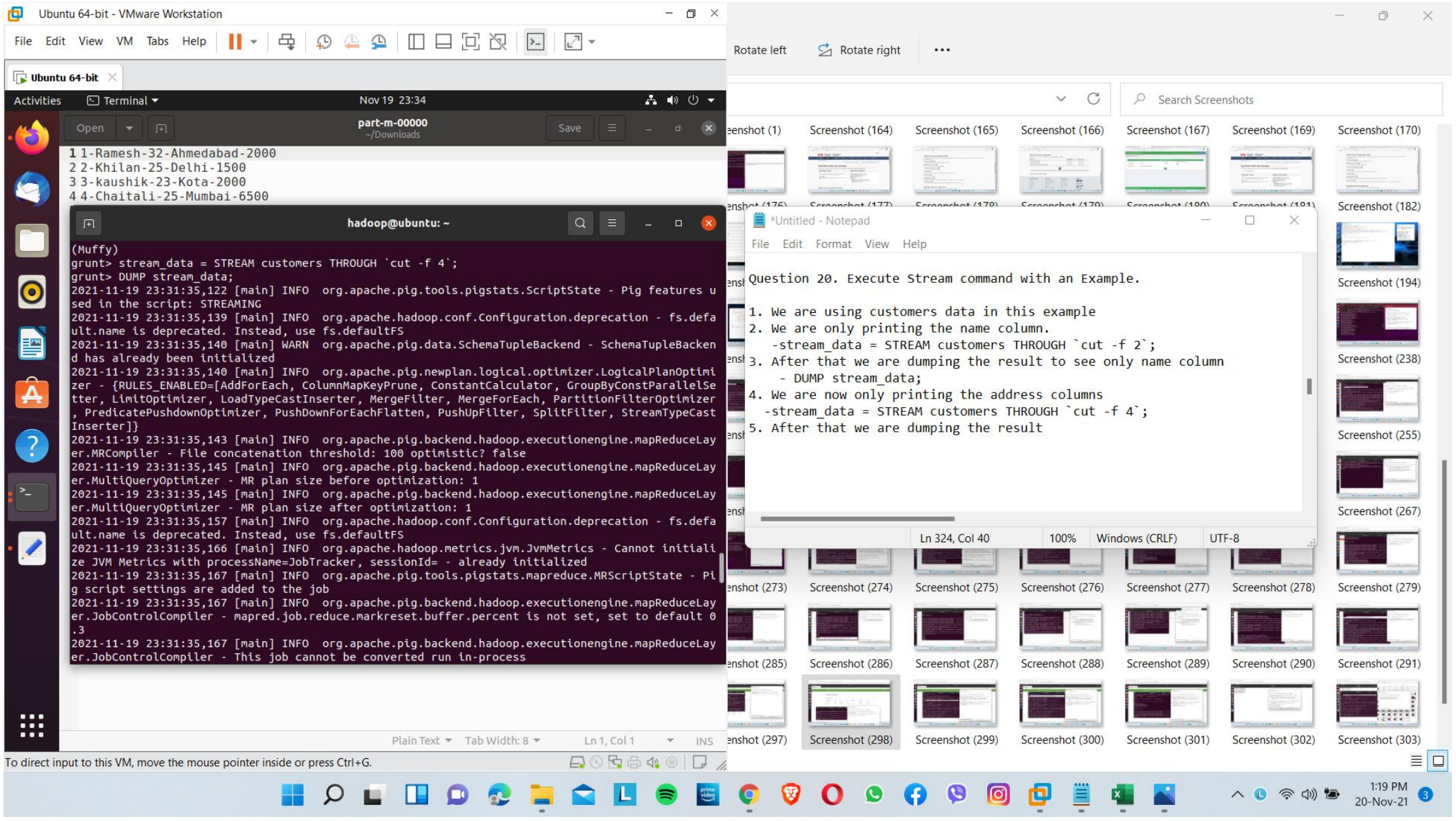

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 23:24
Browsing HDFS
localhost:50070/explorer.html#/TestData/Output/tomap
Hadoop
Overview
Datanodes
Snapshot
Startup Progress
Utilities
hadoop@ubuntu: ~
er.MapReduceLauncher - Success!
grunt>
grunt> customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2021-11-19 23:18:35,923 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:18:35,959 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> STORE customers INTO '/TestData/Output/pigstorage' USING PigStorage('-');
2021-11-19 23:19:43,833 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:19:43,863 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2021-11-19 23:19:43,873 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:19:43,873 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 23:19:43,874 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 23:19:43,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 23:19:43,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 23:19:43,875 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 23:19:43,882 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:19:43,883 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initiali
*Untitled - Notepad
File Edit Format View Help
Question 19. Execute PigStorage command with an Example.
1. We are using customer.txt data . Using pigstorage(',') we used ',' delimiter to delimitted our data.
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2. In 2nd step, we are storing the data in /TestData/output/pigstorage but we have used '-' delimiter .
- STORE customers INTO '/TestData/Output/pigstorage' USING PigStorage('-');
Ln 311, Col 21
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
1:09 PM
20-Nov-21

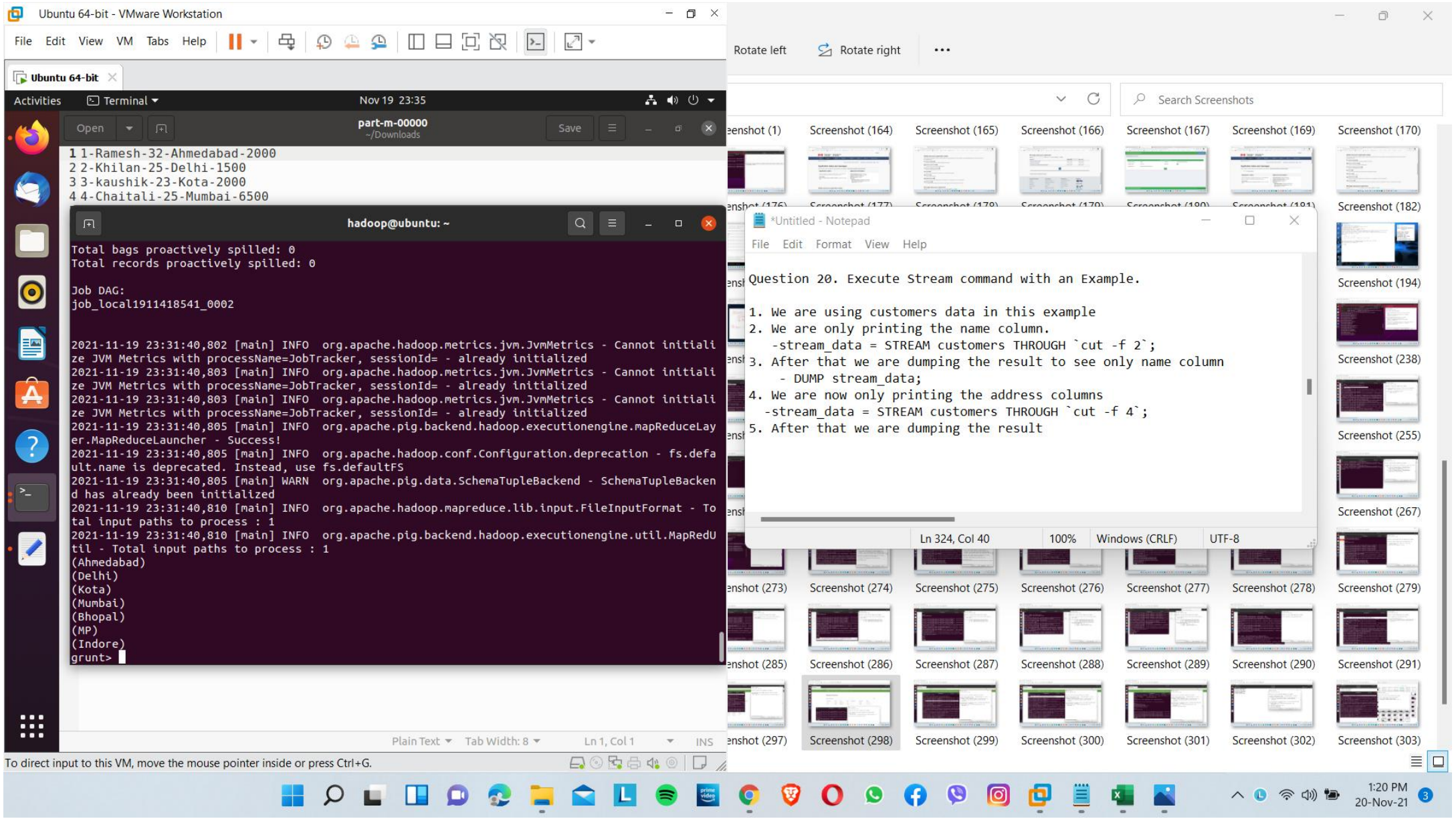

Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 23:24
Browsing HDFS
localhost:50070/explorer.html#/TestData/Output/tomap
Hadoop
Overview
Datanodes
Snapshot
Startup Progress
Utilities
hadoop@ubuntu: ~
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs
job_local793202183_0002 1 0 n/a n/a n/a n/a 0 0 0 0
customers MAP_ONLY /TestData/Output/pigstorage,
Input(s):
Successfully read 7 records (10500294 bytes) from: "/TestData/customer.txt"
Output(s):
Successfully stored 7 records (10500193 bytes) in: "/TestData/Output/pigstorage"
Counters:
Total records written : 7
Total bytes written : 10500193
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local793202183_0002
2021-11-19 23:19:49,531 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:19:49,532 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:19:49,532 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:19:49,534 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
*Untitled - Notepad
File Edit Format View Help
Question 19. Execute PigStorage command with an Example.
1. We are using customer.txt data . Using pigstorage(',') we used ',' delimiter to delimitted our data.
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray, salary:int);
2. In 2nd step, we are storing the data in /TestData/output/pigstorage but we have used '-' delimiter .
- STORE customers INTO '/TestData/Output/pigstorage' USING PigStorage('-');
3. Our data is successfully stored
Ln 312, Col 35
100%
Windows (CRLF)
UTF-8
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
1:09 PM
20-Nov-21


Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Text Editor
Nov 19 23:25
Open
part-m-00000
~/Downloads
Save
1 1-Ramesh-32-Ahmedabad-2000
2 2-Khilan-25-Delhi-1500
3 3-kaushik-23-Kota-2000
4 4-Chaitali-25-Mumbai-6500
5 5-Hardik-27-Bhopal-8500
6 6-Komal-22-MP-4500
7 7-Muffy-24-Indore-10000
*Untitled - Notepad
File Edit Format View Help
Question 19. Execute PigStorage command with an Example.
1. We are using customer.txt data . Using pigstorage(',') we used ','
delimiter to delimitted our data.
- customers = LOAD '/TestData/customer.txt' USING PigStorage(',')
>> as (id:int, name:chararray, age:int, address:chararray,
salary:int);
2. In 2nd step, we are storing the data in /TestData/output/pigstorage
but we have used '-' delimiter .
- STORE customers INTO '/TestData/Output/pigstorage' USING
PigStorage('-');
3. Our data is successfully stored
4. This is our output data
Ln 313, Col 27
100%
Windows (CRLF)
UTF-8
Plain Text
Tab Width: 8
Ln 1, Col 1
INS
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
1:10 PM
20-Nov-21


Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 23:33
Open
part-m-00000
~/Downloads
Save
1 1-Ramesh-32-Ahmedabad-2000
2 2-Khilan-25-Delhi-1500
3 3-kaushik-23-Kota-2000
4 4-Chaitali-25-Mumbai-6500
hadoop@ubuntu: ~
grunt> stream_data = STREAM customers THROUGH `cut -f 2`;
grunt> DUMP stream_data;
2021-11-19 23:31:13,621 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: STREAMING
2021-11-19 23:31:13,644 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:31:13,647 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2021-11-19 23:31:13,662 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 23:31:13,719 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 23:31:13,730 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 23:31:13,730 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 23:31:13,743 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:31:13,749 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - session.id is deprecated. Instead, use dfs.metrics.session-id
2021-11-19 23:31:13,750 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Initializing JVM Metrics with processName=JobTracker, sessionId=
2021-11-19 23:31:13,762 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 23:31:13,765 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buff
Plain Text
Tab Width: 8
Ln 1, Col 1
INS
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
Rotate left
Rotate right
Search Screenshots
Screenshot (1)
Screenshot (164)
Screenshot (165)
Screenshot (166)
Screenshot (167)
Screenshot (169)
Screenshot (170)
Screenshot (182)
Screenshot (194)
Screenshot (238)
Screenshot (255)
Screenshot (267)
Screenshot (273)
Screenshot (274)
Screenshot (275)
Screenshot (276)
Screenshot (277)
Screenshot (278)
Screenshot (279)
Screenshot (285)
Screenshot (286)
Screenshot (287)
Screenshot (288)
Screenshot (289)
Screenshot (290)
Screenshot (291)
Screenshot (297)
Screenshot (298)
Screenshot (299)
Screenshot (300)
Screenshot (301)
Screenshot (302)
*Untitled - Notepad
File Edit Format View Help
Question 20. Execute Stream command with an Example.
1. We are using customers data in this example
2. We are only printing the name column.
-stream_data = STREAM customers THROUGH `cut -f 2`;
Ln 320, Col 1
100%
Windows (CRLF)
UTF-8
1:18 PM
20-Nov-21


Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities Terminal Nov 19 23:33
Open part-m-00000 ~/Downloads Save
1 1-Ramesh-32-Ahmedabad-2000
2 2-Khilan-25-Delhi-1500
3 3-kaushik-23-Kota-2000
4 4-Chaitali-25-Mumbai-6500
hadoop@ubuntu: ~
Total records written : 7
Total bytes written : 5250044
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1736779061_0001
2021-11-19 23:31:20,112 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:31:20,113 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:31:20,115 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:31:20,121 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 23:31:20,123 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:31:20,124 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 23:31:20,140 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 23:31:20,140 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Ramesh)
(Khilan)
(kaushik)
(Chaitali)
(Hardik)
Plain Text Tab Width: 8 Ln 1, Col 1 INS
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
Rotate left
Rotate right
Search Screenshots
Screenshot (1)
Screenshot (164)
Screenshot (165)
Screenshot (166)
Screenshot (167)
Screenshot (169)
Screenshot (170)
Screenshot (182)
Screenshot (194)
Screenshot (238)
Screenshot (255)
Screenshot (267)
Screenshot (273)
Screenshot (274)
Screenshot (275)
Screenshot (276)
Screenshot (277)
Screenshot (278)
Screenshot (279)
Screenshot (285)
Screenshot (286)
Screenshot (287)
Screenshot (288)
Screenshot (289)
Screenshot (290)
Screenshot (291)
Screenshot (297)
Screenshot (298)
Screenshot (299)
Screenshot (300)
Screenshot (301)
Screenshot (302)
Screenshot (303)
*Untitled - Notepad
File Edit Format View Help
Question 20. Execute Stream command with an Example.
1. We are using customers data in this example
2. We are only printing the name column.
-stream_data = STREAM customers THROUGH `cut -f 2`;
3. After that we are dumping the result to see only name column
- DUMP stream_data;
Ln 321, Col 24 100% Windows (CRLF) UTF-8
1:18 PM
20-Nov-21


Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities
Terminal
Nov 19 23:34
Open
part-m-00000
~/Downloads
Save
1 1-Ramesh-32-Ahmedabad-2000
2 2-Khilan-25-Delhi-1500
3 3-kaushik-23-Kota-2000
4 4-Chaitali-25-Mumbai-6500
hadoop@ubuntu: ~
(Muffy)
grunt> stream_data = STREAM customers THROUGH `cut -f 4`;
grunt> DUMP stream_data;
2021-11-19 23:31:35,122 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: STREAMING
2021-11-19 23:31:35,139 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:31:35,140 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 23:31:35,140 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2021-11-19 23:31:35,143 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2021-11-19 23:31:35,145 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2021-11-19 23:31:35,145 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2021-11-19 23:31:35,157 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:31:35,166 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:31:35,167 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2021-11-19 23:31:35,167 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2021-11-19 23:31:35,167 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
Plain Text
Tab Width: 8
Ln 1, Col 1
INS
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
Rotate left
Rotate right
Search Screenshots
Screenshot (164)
Screenshot (165)
Screenshot (166)
Screenshot (167)
Screenshot (169)
Screenshot (170)
Screenshot (182)
Screenshot (194)
Screenshot (238)
Screenshot (255)
Screenshot (267)
Screenshot (273)
Screenshot (274)
Screenshot (275)
Screenshot (276)
Screenshot (277)
Screenshot (278)
Screenshot (279)
Screenshot (285)
Screenshot (286)
Screenshot (287)
Screenshot (288)
Screenshot (289)
Screenshot (290)
Screenshot (291)
Screenshot (297)
Screenshot (298)
Screenshot (299)
Screenshot (300)
Screenshot (301)
Screenshot (302)
Screenshot (303)
*Untitled - Notepad
File Edit Format View Help
Question 20. Execute Stream command with an Example.
1. We are using customers data in this example
2. We are only printing the name column.
-stream_data = STREAM customers THROUGH `cut -f 2`;
3. After that we are dumping the result to see only name column
- DUMP stream_data;
4. We are now only printing the address columns
-stream_data = STREAM customers THROUGH `cut -f 4`;
5. After that we are dumping the result
Ln 324, Col 40
100%
Windows (CRLF)
UTF-8
1:19 PM
20-Nov-21


Ubuntu 64-bit - VMware Workstation
File Edit View VM Tabs Help
Ubuntu 64-bit
Activities Terminal Nov 19 23:35
Open part-m-00000 ~/Downloads Save
1 1-Ramesh-32-Ahmedabad-2000
2 2-Khilan-25-Delhi-1500
3 3-kaushik-23-Kota-2000
4 4-Chaitali-25-Mumbai-6500
hadoop@ubuntu: ~
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1911418541_0002
2021-11-19 23:31:40,802 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:31:40,803 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:31:40,803 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
2021-11-19 23:31:40,805 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2021-11-19 23:31:40,805 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2021-11-19 23:31:40,805 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2021-11-19 23:31:40,810 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2021-11-19 23:31:40,810 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Ahmedabad)
(Delhi)
(Kota)
(Mumbai)
(Bhopal)
(MP)
(Indore)
grunt>
Plain Text Tab Width: 8 Ln 1, Col 1 INS
To direct input to this VM, move the mouse pointer inside or press Ctrl+G.
Rotate left
Rotate right
Search Screenshots
Screenshot (1)
Screenshot (164)
Screenshot (165)
Screenshot (166)
Screenshot (167)
Screenshot (169)
Screenshot (170)
Screenshot (182)
Screenshot (194)
Screenshot (238)
Screenshot (255)
Screenshot (267)
Screenshot (273)
Screenshot (274)
Screenshot (275)
Screenshot (276)
Screenshot (277)
Screenshot (278)
Screenshot (279)
Screenshot (285)
Screenshot (286)
Screenshot (287)
Screenshot (288)
Screenshot (289)
Screenshot (290)
Screenshot (291)
Screenshot (297)
Screenshot (298)
Screenshot (299)
Screenshot (300)
Screenshot (301)
Screenshot (302)
Screenshot (303)
*Untitled - Notepad
File Edit Format View Help
Question 20. Execute Stream command with an Example.
1. We are using customers data in this example
2. We are only printing the name column.
-stream_data = STREAM customers THROUGH `cut -f 2`;
3. After that we are dumping the result to see only name column
- DUMP stream_data;
4. We are now only printing the address columns
-stream_data = STREAM customers THROUGH `cut -f 4`;
5. After that we are dumping the result
Ln 324, Col 40 100% Windows (CRLF) UTF-8
1:20 PM
20-Nov-21