

Tanggal Praktikum : Tuesday, May 30, 2023

The screenshot shows a Windows desktop environment. In the foreground, a terminal window titled "root@127.0.0.1 (root)" is active. It displays the command prompt after a successful SSH connection to root@127.0.0.1. The terminal output includes:

```

root@127.0.0.1: password:
• Mohabtern Personal Edition v23.1 •
  (SSH client, X server and network tools)

• SSH session to root@127.0.0.1
  • Direct SSH : ✓
  • SSH compression : ✓
  • SSH forwarder : ✓
  • X11-forwarding : x (disabled or not supported by server)

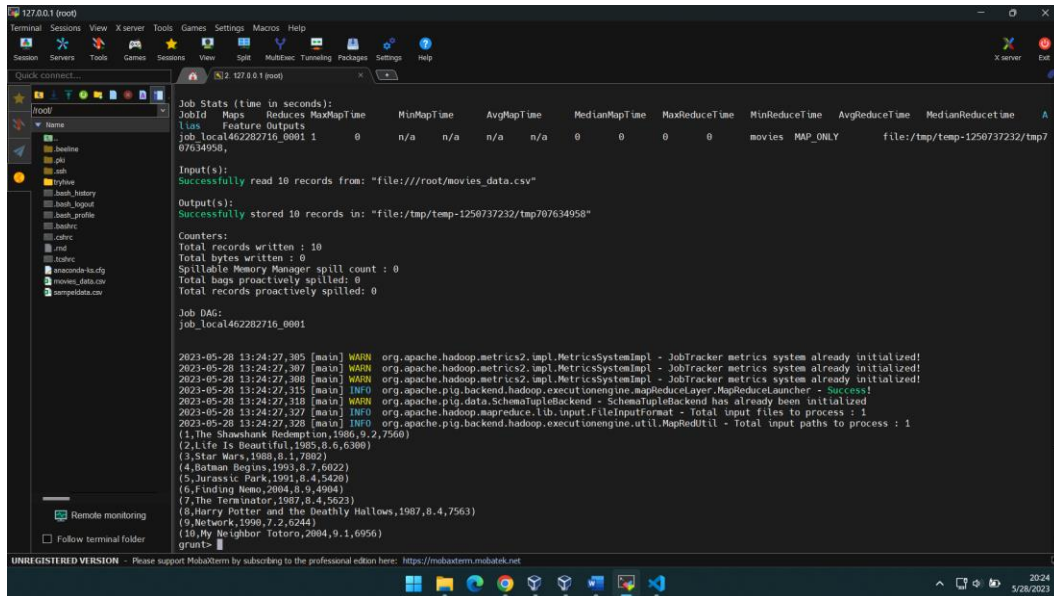
• For more info, ctrl+c/link on help or visit our website.

Last login: Sun May 28 13:00:17 2023
[root@andbox-hdp ~]# ls
anaconda-ks.cfg  sampledata.csv  tryhive
[root@andbox-hdp ~]# cd /
[root@andbox-hdp ~]# ls
anaconda-ks.cfg  movies.data.csv  sampledata.csv  tryhive
[root@andbox-hdp ~]# pig -l local
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
23/05/28 13:21:18 INFO org.apache.pig.Main: Apache Pig version 0.16.0-3.0.0-1.0-187 (rhiversioned directory) compiled Sep 19 2018, 10:13:33
2023-05-28 13:21:18 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecEngine: Connecting to hadoop file system at: file:///
2023-05-28 13:21:18 [main] INFO org.apache.pig.impl.util.Utils: Default bootstrap file /root/.pigbootpath not found
2023-05-28 13:21:18 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecEngine: Connecting to hadoop file system at: file:///
2023-05-28 13:21:18 [main] WARN org.apache.pig.PigServer: Pig Server for the session: PIG-default(044b8dc-1bd3-d452-9518-24f94331fafe)
2023-05-28 13:21:18 [main] WARN org.apache.pig.PigServer: AJS is disabled since yarn.timeline-service.enabled set to false
grunt> movies = load '/usr/local/hadoop/share/pig/lib/*' using PigStorage('-') as (id,name,year,rating,duration);
grunt> dump movies;
2023-05-28 13:24:25 [main] INFO org.apache.pig.tools.pigstats.ScriptStats: Pig features used in the script: BROADCAST
org.apache.pig.mapjoin.logical.optimizer.LogicalPlanOptimizer -- (HILLS_ENABLED=AddressEach, ColumnMapKeyPrune, ConstantCa
culator, GroupByConstParallelSettler, LocalIterator, LocalProjectorInsertion, MergeFilter, MergeGroup, PartitionIterOptimizer, P
ushdownForEachJoin, PushFilter, SplitFilter, StreamTypeCastInserter))
2023-05-28 13:24:25 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager: Selected heap (PS Old Gen) of size 694940192 to monitor, collectionmg
threshold = 48958128, usageThreshold = 48958128
2023-05-28 13:24:25 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCCompiler: File concatenation threshold: 100 optimistic? f
all
2023-05-28 13:24:25 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer: MR plan size before optimization: 1
2023-05-28 13:24:25 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer: MR plan size after optimization: 1
org.apache.hadoop.metrics2.impl.MetricConfigImpl - Loaded properties from hadoop-metrics2.properties

```

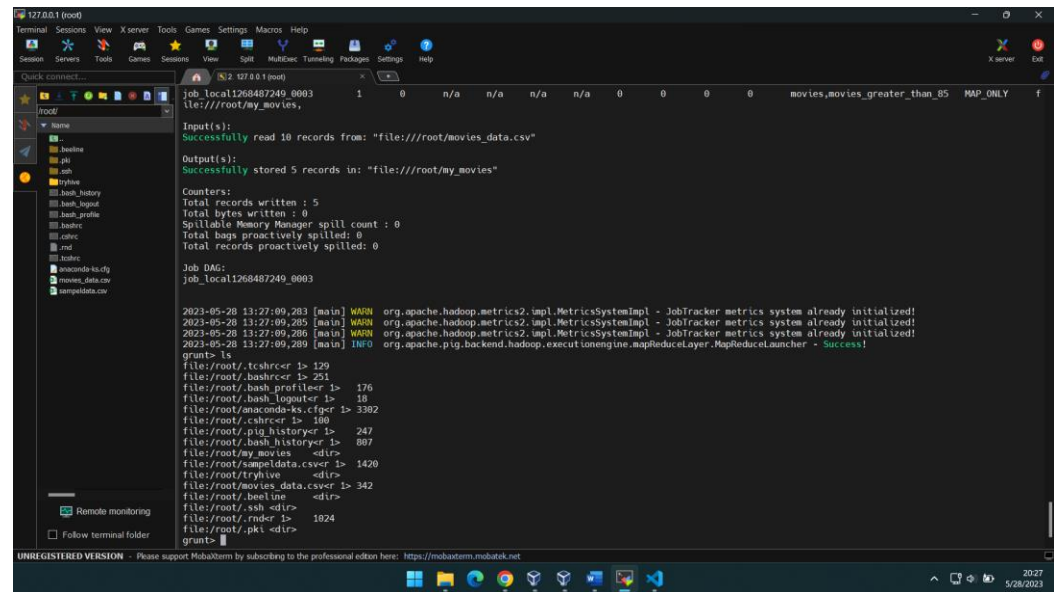
In the background, a file explorer window is open, showing the contents of the "/home" directory. It lists various files and folders including .bash_history, .bash_logout, .bash_profile, .bashrc, .cdirc, .git, .local, .mavenrc, .mavenrc-tcp, .mavenrc-tcp, .sampledata.csv, and .sampledata.csv.

3. Jalankan Script untuk menmbuat variable dan menampilkan data

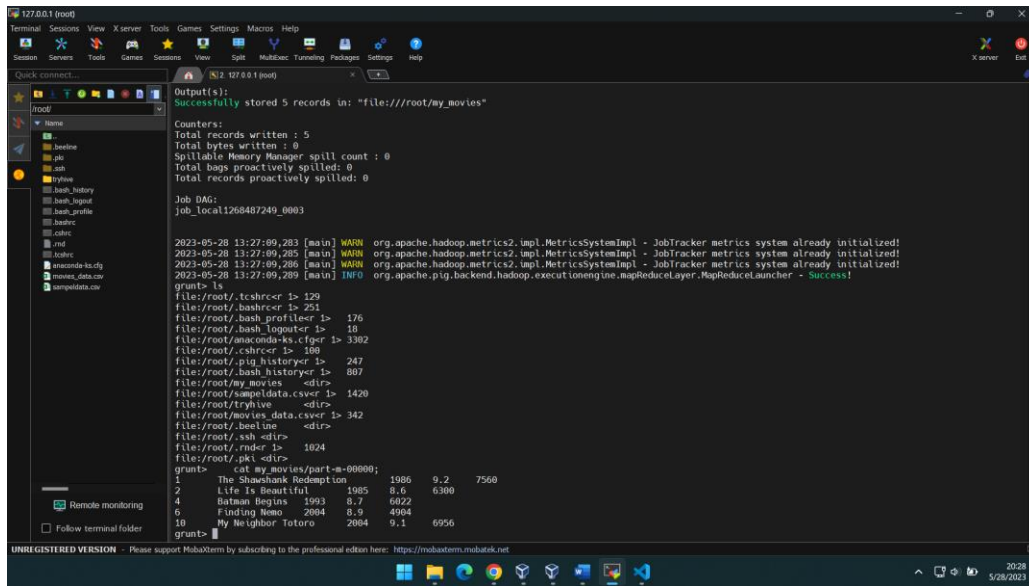


```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
root@127.0.0.1:~$
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime
127.0.0.1:~$
Job local462282716_0001 1 0 n/a n/a n/a n/a 0 0 0 0 movies MAP_ONLY file:/tmp/temp-1250737232/tmp707634958,
07634958,
Input(s):
Successfully read 10 records from: "file:///root/movies_data.csv"
Output(s):
Successfully stored 10 records in: "file:/tmp/temp-1250737232/tmp707634958"
Counters:
Total records written: 10
Total bytes written: 0
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local462282716_0001
2023-05-28 13:24:27,305 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:24:27,307 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:24:27,308 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:24:27,315 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2023-05-28 13:24:27,318 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-05-28 13:24:27,327 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:24:27,328 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,The Shawshank Redemption,1986,9.2,7560)
(2,Life Is Beautiful,1985,8.6,6389)
(3,Star Wars,1980,8.1,7092)
(4,Batman Begins,1993,8.7,6022)
(5,Jurassic Park,1991,8.4,5428)
(6,Finding Nemo,2004,8.9,4904)
(7,The Terminator,1987,8.4,5623)
(8,Harry Potter and the Deathly Hallows,1987,8.4,7563)
(9,Network,1999,7.2,6244)
(10,My Neighbor Totoro,2004,9.1,6956)
grunt>
```

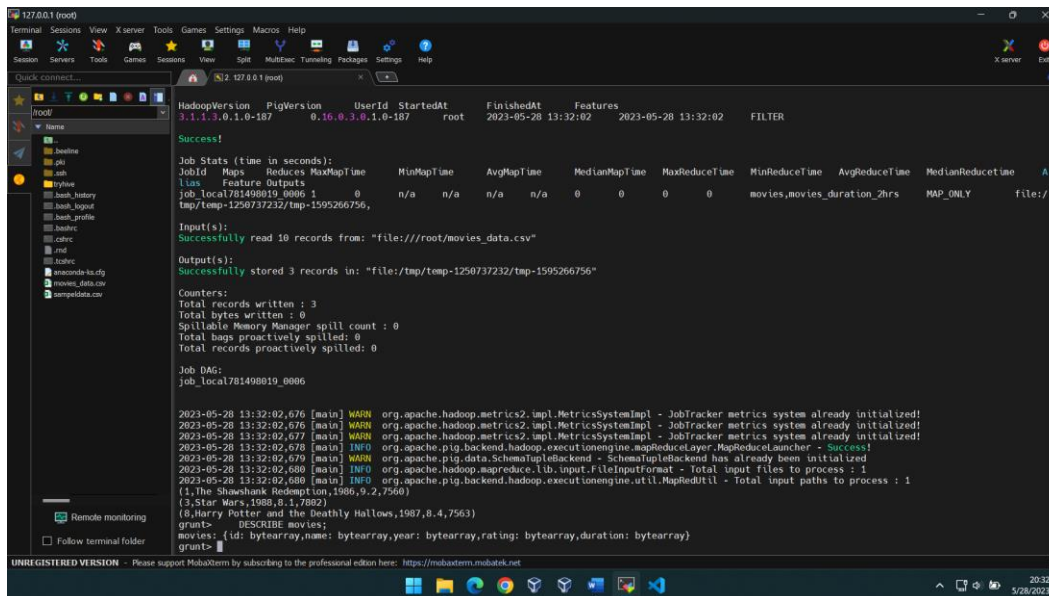
4. Menyimpan Variable



```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
root@127.0.0.1:~$
Job local1268487249_0003 1 0 n/a n/a n/a n/a 0 0 0 0 movies,movies_greater_than_85 MAP_ONLY f
ile:///root/my_movies,
Input(s):
Successfully read 10 records from: "file:///root/movies_data.csv"
Output(s):
Successfully stored 5 records in: "file:///root/my_movies"
Counters:
Total records written: 5
Total bytes written: 0
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1268487249_0003
2023-05-28 13:27:09,283 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:27:09,285 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:27:09,286 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:27:09,289 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
grunt> ls
file:/root/.tcshrcer 129
file:/root/.bashrcer 1251
file:/root/.bash.profiler 176
file:/root/.bash.logouter 18
file:/root/.anaconda-ks.cfg 3382
file:/root/.cshrcer 186
file:/root/.pig.history 247
file:/root/.bash.history 807
file:/root/my_movies <dir>
file:/root/sampledata.csv 1420
file:/root/tryhive <dir>
file:/root/movies_data.csv 342
file:/root/beeLine <dir>
file:/root/.ssh <dir>
file:/root/.rmdir 1024
file:/root/.pk1 <dir>
grunt>
```



5. Menampilkan skema



6. Transformasi Data

The screenshot shows a terminal window with a dark theme. At the top, there's a title bar for a file named '17/08/21 (pool)'. Below it, a menu bar includes 'Terminal', 'Sessions', 'View', 'X Server', 'Tools', 'Games', 'Settings', 'Macros', and 'Help'. A toolbar with various icons is visible. The main terminal area displays the output of a Hadoop job. It starts with 'Quick connect...' and a file explorer sidebar on the left showing a directory structure under 'root/'. The job output includes statistics for 'Job Stats (Time in seconds):', a table of metrics (MinMapTime, AvgMapTime, MedianMapTime, MaxReduceTime, MinReduceTime, AvgReduceTime, MedianReduceTime), and a list of counters. It then shows the DAG (Directed Acyclic Graph) for the job. Finally, it displays a series of log messages from the Hadoop metrics2 system, indicating that the metrics system is already initialized and listing various input files being processed.

```

Job Stats (Time in seconds):
JobId      Maps      Reduces  MapSpTime
  lias      Feature outputs
job_local-1250376710_0007 1    0    n/a    n/a    n/a    n/a    0    0    0    0    movie_duration_mmt,movies    MAP_ONLY    file:/
tmp/tmp-12503737232/tmp-1624846836,

Input(s):
Successfully read 10 records from: "file:///root/movies_data.csv"

Output(s):
Successfully stored 10 records in: "file:/tmp/tmp-12503737232/tmp-1624846836"

Counters:
Total records written : 10
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1250376710_0007

2023-05-28 13:33:39.090 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:33:39.092 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:33:39.092 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:33:39.094 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2023-05-28 13:33:39.095 [main] WARN org.apache.pig.data.SchemaUpBackend - SchemaUpBackend has already been initialized
2023-05-28 13:33:39.096 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:33:39.097 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapReduceUtil - Total input paths to process : 1
(The Shawshank Redemption,126.0)
(Life Is Beautiful,105.0)
(Star Wars,130.0)
(Batman Begins,100.0)
(Jurassic Park,99.0)
(Finding Nemo,81.0)
(The Terminator,93.0)
(Harry Potter and the Deathly Hallows,126.0)
(Network,104.0)
(My Neighbor Totoro,115.0)
grants

UNBREGISTERED VERSION - Please support MobazTerm by subscribing to the professional edition here: https://mobazterm.mobaztek.net
  
```

7. Group By Data

The image shows a Windows desktop environment. In the foreground, a terminal window titled "cmd (17280 x 1000)" is open, displaying the execution of Hadoop commands. The first command is `hadoop fs -ls /root/movies_data.csv`, which successfully reads 10 records from the file. The second command is `hadoop fs -ls /tmp/1250737232/tmp-1624846836`, which successfully stores 10 records. The third command is `hadoop fs -ls /tmp/1250737232/tmp-1624846836`, which shows the contents of the file, including a list of movies and their counts. The output of the third command is as follows:

```
Output(s):
Successfully stored 10 records in: "file:/tmp/1250737232/tmp-1624846836"

Counters:
Total records written: 10
Total bytes written: 0
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local288376710_0007

2023-05-28 13:33:39,090 [main] WARN org.apache.hadoop.metrics2Impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:33:39,092 [main] WARN org.apache.hadoop.metrics2Impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:33:39,094 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MapReduceLauncher - Success!
2023-05-28 13:33:39,095 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process: 1
2023-05-28 13:33:39,096 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process: 1
(Life Is Beautiful,105.0)
(Star Wars,130.0)
(Batman Begins,100.0)
(Jurassic Park,90.0)
(Finding Nemo,81.0)
(The Terminator,83.0)
(Harry Potter and the Deathly Hallows,126.0)
(Network,184.0)
(My Neighbor Totoro,115.0)
grunt> group by year = group movies by year;
2023-05-28 13:34:27,966 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2023-05-28 13:34:27,967 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-05-28 13:34:27,967 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_INT 4 time(s).
grunt> count by year = FOREACH group by year GENERATE group, COUNT(movies);
2023-05-28 13:34:32,168 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2023-05-28 13:34:32,168 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-05-28 13:34:32,168 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_INT 4 time(s).
grunt>
```


In the background, a file explorer window titled "N2: 127.0.0.1 (port)" is open, showing the contents of the root directory. The file explorer shows a tree view on the left with folders like "root/" and "root/127.0.0.1 (port)". The main pane shows the contents of the root directory, which includes a file named "movies_data.csv". The file explorer also shows the file's properties, including its size (10 bytes) and its location (file:/tmp/1250737232/tmp-1624846836).


```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split Multitac Tunneling Packages Settings Help
Quick connect...
root
Name
beetle
phi
csh
python
hadoop_history
hadoop_logout
hadoop_profile
hadoop
calvc
cmd
hadoop
anacorde-ku.dfg
movies_data.csv
tempdata.csv
Remote monitoring
Follow terminal folder

Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime
1 1 n/a n/a n/a n/a n/a n/a n/a n/a
BY_COMBINE file:/tmp/tmp-1250737232/tmp-408242062,

Input(s):
Successfully read 10 records from: "file:///root/movies_data.csv"

Output(s):
Successfully stored 8 records in: "file:/tmp/tmp-1250737232/tmp-408242062"

Counters:
Total records written : 8
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1328132894_0009

2023-05-28 13:35:31.281 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:31.286 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:31.292 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:31.301 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2023-05-28 13:35:31.311 [main] WARN org.apache.pig.data.SchemaUpBackend - SchemaUpBackend has already been initialized
2023-05-28 13:35:31.550 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:35:31.550 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1985,1)
(1986,1)
(1987,2)
(1988,1)
(1990,1)
(1991,1)
(1993,1)
(2004,2)
grunt>
```

```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split Multitac Tunneling Packages Settings Help
Quick connect...
root
Name
beetle
phi
csh
python
hadoop_history
hadoop_logout
hadoop_profile
hadoop
calvc
cmd
hadoop
anacorde-ku.dfg
movies_data.csv
tempdata.csv
Remote monitoring
Follow terminal folder

l.us Feature Outputs
job_local404706609_0008 1 n/a n/a n/a n/a n/a n/a n/a grouped_by_year,movies GROUP_BY file:/tmp/temp-1250737232/tmp-2015053306,

Input(s):
Successfully read 10 records from: "file:///root/movies_data.csv"

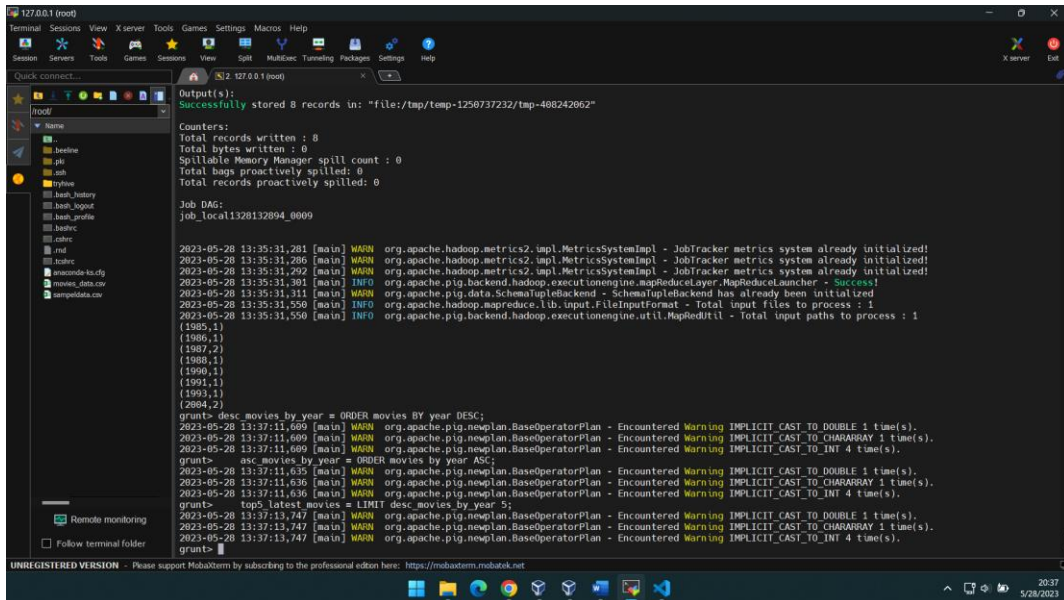
Output(s):
Successfully stored 8 records in: "file:/tmp/tmp-1250737232/tmp-2015053306"

Counters:
Total records written : 8
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local404706609_0008

2023-05-28 13:35:21.136 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:21.137 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:21.138 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:21.140 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2023-05-28 13:35:21.140 [main] WARN org.apache.pig.data.SchemaUpBackend - SchemaUpBackend has already been initialized
2023-05-28 13:35:21.142 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:35:21.142 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1985,((2,Life Is Beautiful,1985,8,6,6380)))
(1986,((1,The Shawshank Redemption,1986,9,2,7568)))
(1987,((0,Harry Potter and the Deathly Hallows,1987,8,4,7563),(7,The Terminator,1987,8,4,5623)))
(1988,((3,Star Wars,1988,8,1,7892)))
(1990,((9,Network,1990,7,2,6244)))
(1991,((5,Jurassic Park,1991,8,4,5428)))
(1993,((4,Batman Begins,1993,8,7,6022)))
(2004,((10,My Neighbor Totoro,2004,9,1,6956),(6,Finding Nemo,2004,8,9,4904)))
grunt> dump count by year;
2023-05-28 13:35:25.140 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY
2023-05-28 13:35:25.152 [main] INFO org.apache.pig.data.SchemaUpBackend - Key [pig.schemaup] was not set... will not generate code.
2023-05-28 13:35:25.153 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCo
llocator, GroupByKeyParallelSorter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, P
ushdownForEachFlatten, PushFilter, SplitFilter, StreamTypeCastInserter]}
```

8. Sorting dengan order



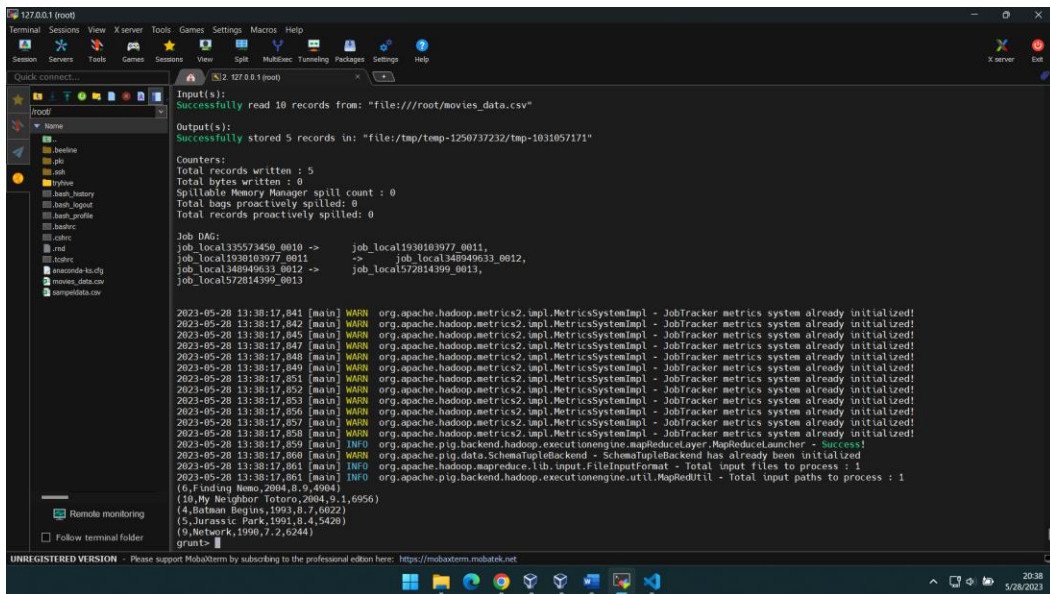
The screenshot shows a terminal window with the following output:

```
Output(s):
Successfully stored 8 records in: "file:/tmp/temp-1250737232/tmp-408242662"

Counters:
Total records written: 0
Total bytes written: 0
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1320132894_0009

2023-05-28 13:35:31,281 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:31,286 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:35:31,292 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Success!
2023-05-28 13:35:31,311 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-05-28 13:35:31,550 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:35:31,550 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1980,1)
(1986,1)
(1987,2)
(1988,1)
(1990,1)
(1991,1)
(1993,1)
(2004,2)
grunt> desc_movies_by_year = ORDER movies BY year DESC;
2023-05-28 13:37:11,669 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2023-05-28 13:37:11,669 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-05-28 13:37:11,669 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_INT 4 time(s).
grunt> asc_movies_by_year = ORDER movies BY year ASC;
2023-05-28 13:37:11,635 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2023-05-28 13:37:11,636 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-05-28 13:37:11,636 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_INT 4 time(s).
grunt> top5_latest_movies = LIMIT desc_movies_by_year 5;
2023-05-28 13:37:13,747 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2023-05-28 13:37:13,747 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
2023-05-28 13:37:13,747 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_INT 4 time(s).
grunt>
```



The screenshot shows a terminal window with the following output:

```
Input(s):
Successfully read 10 records from: "file:///root/movies_data.csv"

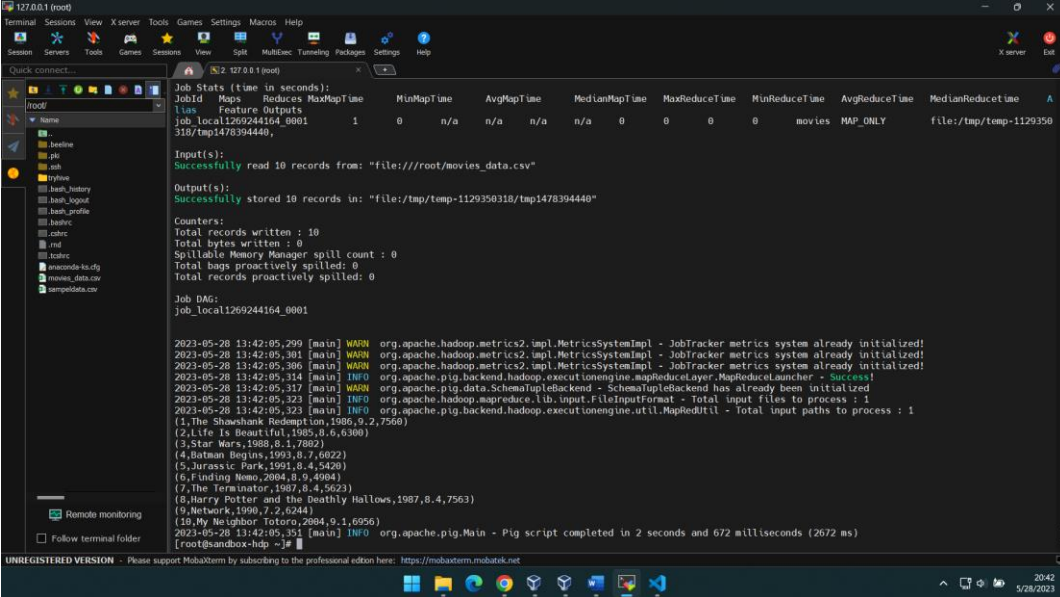
Output(s):
Successfully stored 5 records in: "file:/tmp/temp-1250737232/tmp-1031057171"

Counters:
Total records written: 5
Total bytes written: 0
Spillable Memory Manager spill count: 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local135573450_0010 -> job_local1930103977_0011,
job_local1930103977_0011 -> job_local348949633_0012,
job_local348949633_0012 -> job_local572014399_0013,
job_local572014399_0013

2023-05-28 13:38:17,841 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,842 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,845 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,847 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,848 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,849 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,851 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,852 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,853 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,856 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,857 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,858 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:38:17,859 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce.Launcher - Success!
2023-05-28 13:38:17,860 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-05-28 13:38:17,861 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:38:17,861 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(16, Finding Nemo, 2004, 8.9, 4964)
(18, My Neighbor Totoro, 2004, 9.1, 6956)
(4, Batman Begins, 1993, 8.7, 6022)
(5, Jurassic Park, 1993, 8.4, 5426)
(9, Network, 1996, 7.2, 6244)
grunt>
```

9. Pig Batch Mode



```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split Multitab Tunneling Packages Settings Help

Quick connect...
root@
Job Stats (time in seconds);
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReduceTime
job_local1269244164_0001 1 0 n/a n/a n/a n/a 0 0 0 0
310/tmp1478394440, movies MAP_ONLY file:/tmp/temp-1129350

Inputs:
Successfully read 10 records from: "file:///root/movies_data.csv"

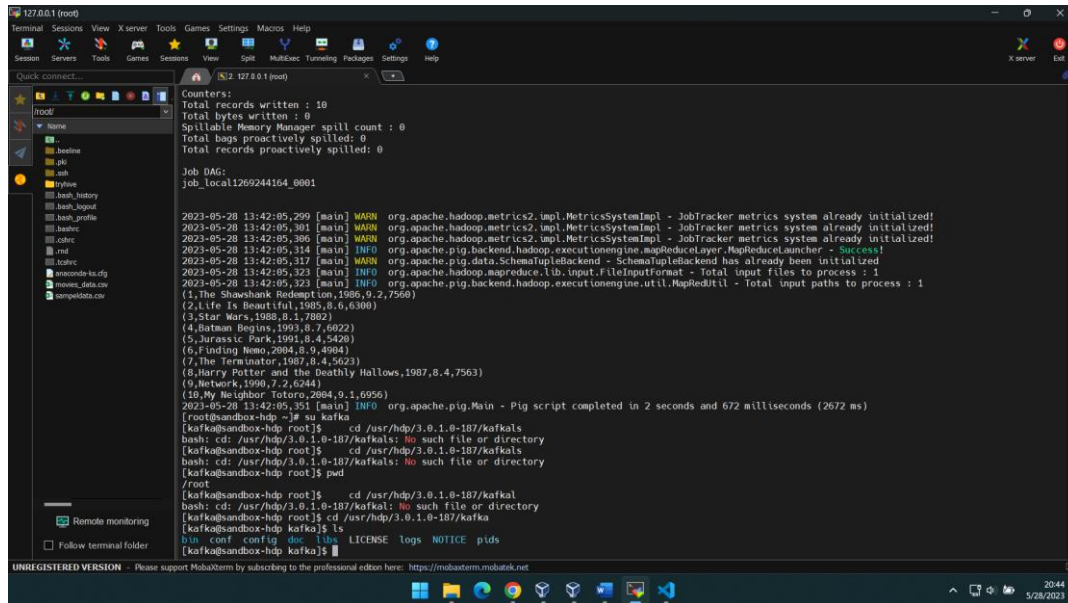
Outputs:
Successfully stored 10 records in: "file:/tmp/temp-1129350310/tmp1478394440"

Counters:
Total records written : 10
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1269244164_0001

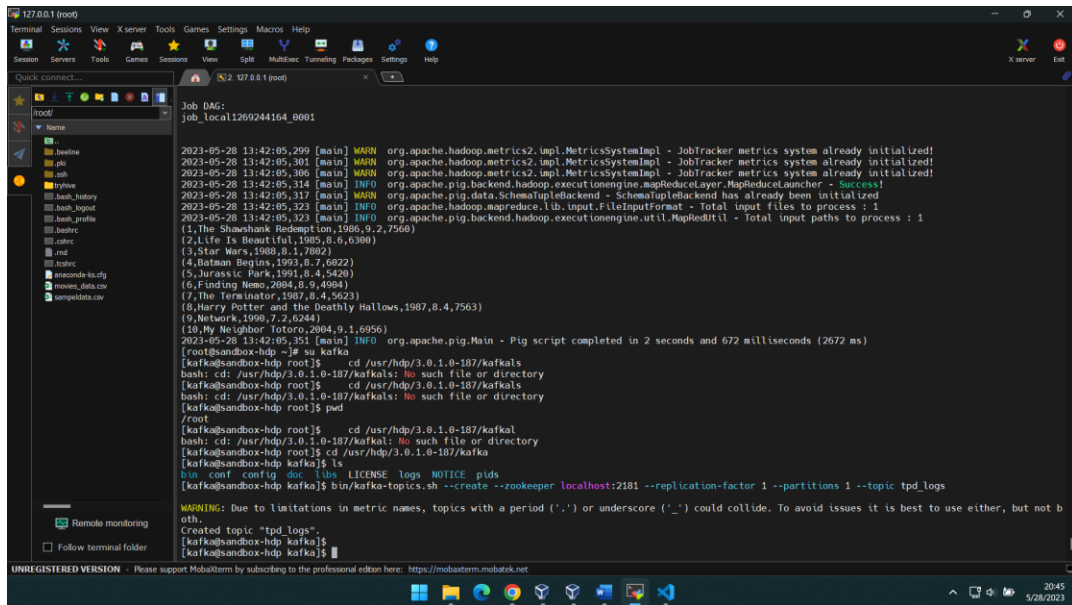
2023-05-28 13:42:05.299 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05.301 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05.306 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05.314 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduce_layer.MapReduceLauncher - Success!
2023-05-28 13:42:05.317 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-05-28 13:42:05.323 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:42:05.323 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,The Shawshank Redemption,1986,8.2,7568)
(2,Life Is Beautiful,1995,8.6,6380)
(3,Star Wars,1988,8.1,7802)
(4,Batman Begins,1993,8.7,6022)
(5,Jurassic Park,1991,8.4,5420)
(6,Finding Nemo,2004,8.9,4904)
(7,The Terminator,1987,8.4,5623)
(8,Herry Potter and the Beethly Hallows,1987,8.4,7563)
(9,Network,1990,7.2,6244)
(10,My Neighbor Totoro,2004,9.1,6956)
2023-05-28 13:42:05.351 [main] INFO org.apache.pig.Main - Pig script completed in 2 seconds and 672 millseconds (2672 ms)
[root@sandbox-hdp ~]#
```

1. Masuk ke kafka



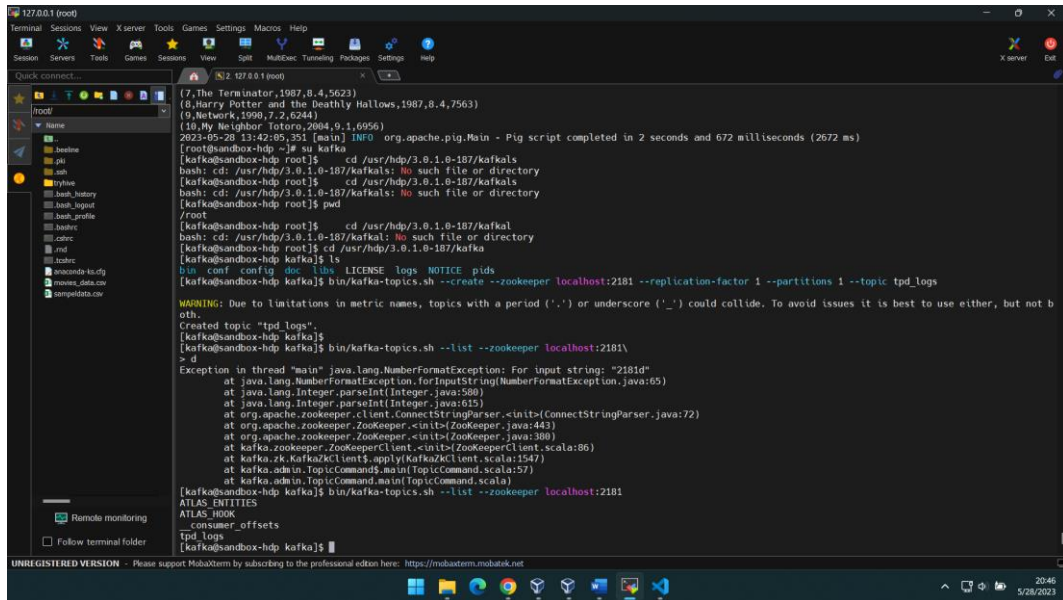
```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
root@127.0.0.1: ~
Counters:
Total records written : 10
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_local1269244164_0001
2023-05-28 13:42:05,299 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05,301 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05,306 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05,314 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.MapReduceLauncher - Success!
2023-05-28 13:42:05,317 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-05-28 13:42:05,323 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:42:05,323 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,The Shawshank Redemption,1986,9.2,7566)
(2,Life Is Beautiful,1985,8.6,6300)
(3,Star Wars,1988,8.1,7802)
(4,Batman Begins,1993,8.7,6022)
(5,Jurassic Park,1993,8.4,5420)
(6,Finding Nemo,2004,8.9,4904)
(7,The Terminator,1987,8.4,5623)
(8,Harry Potter and the Deathly Hallows,1987,8.4,7563)
(9,Network,1990,7.2,6244)
(10,My Neighbor Totoro,2004,9.1,6956)
2023-05-28 13:42:05,351 [main] INFO org.apache.pig.Main - Pig script completed in 2 seconds and 672 milliseconds (2672 ms)
[root@sandbox-hdp ~]# su kafka
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ pwd
/root
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
[kafka@sandbox-hdp kafka]$ ls
bin conf config doc libs LICENSE logs NOTICE pids
[kafka@sandbox-hdp kafka]$
```

2. Buat topic



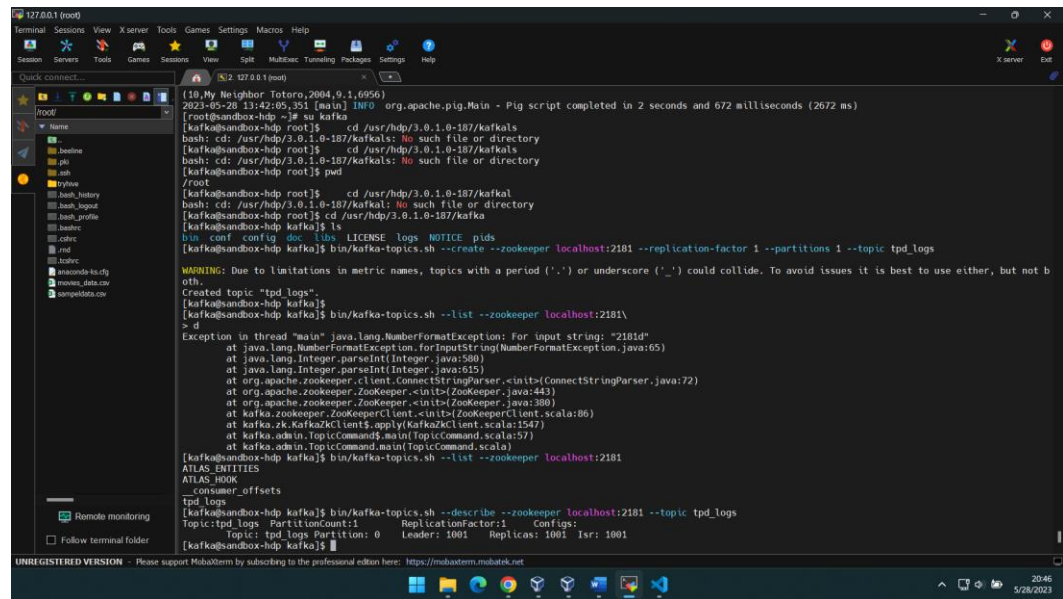
```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Quick connect...
root@127.0.0.1: ~
Job DAG:
job_local1269244164_0001
2023-05-28 13:42:05,299 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05,301 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05,306 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-05-28 13:42:05,314 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreducelayer.MapReduceLauncher - Success!
2023-05-28 13:42:05,317 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-05-28 13:42:05,323 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-05-28 13:42:05,323 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,The Shawshank Redemption,1986,9.2,7566)
(2,Life Is Beautiful,1985,8.6,6300)
(3,Star Wars,1988,8.1,7802)
(4,Batman Begins,1993,8.7,6022)
(5,Jurassic Park,1993,8.4,5420)
(6,Finding Nemo,2004,8.9,4904)
(7,The Terminator,1987,8.4,5623)
(8,Harry Potter and the Deathly Hallows,1987,8.4,7563)
(9,Network,1990,7.2,6244)
(10,My Neighbor Totoro,2004,9.1,6956)
2023-05-28 13:42:05,351 [main] INFO org.apache.pig.Main - Pig script completed in 2 seconds and 672 milliseconds (2672 ms)
[root@sandbox-hdp ~]# su kafka
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ pwd
/root
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
[kafka@sandbox-hdp kafka]$ ls
bin conf config doc libs LICENSE logs NOTICE pids
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic tpd_logs
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not b
oth.
Created topic 'tpd_logs'.
[kafka@sandbox-hdp kafka]$
[kafka@sandbox-hdp kafka]$
```


3. List all topic



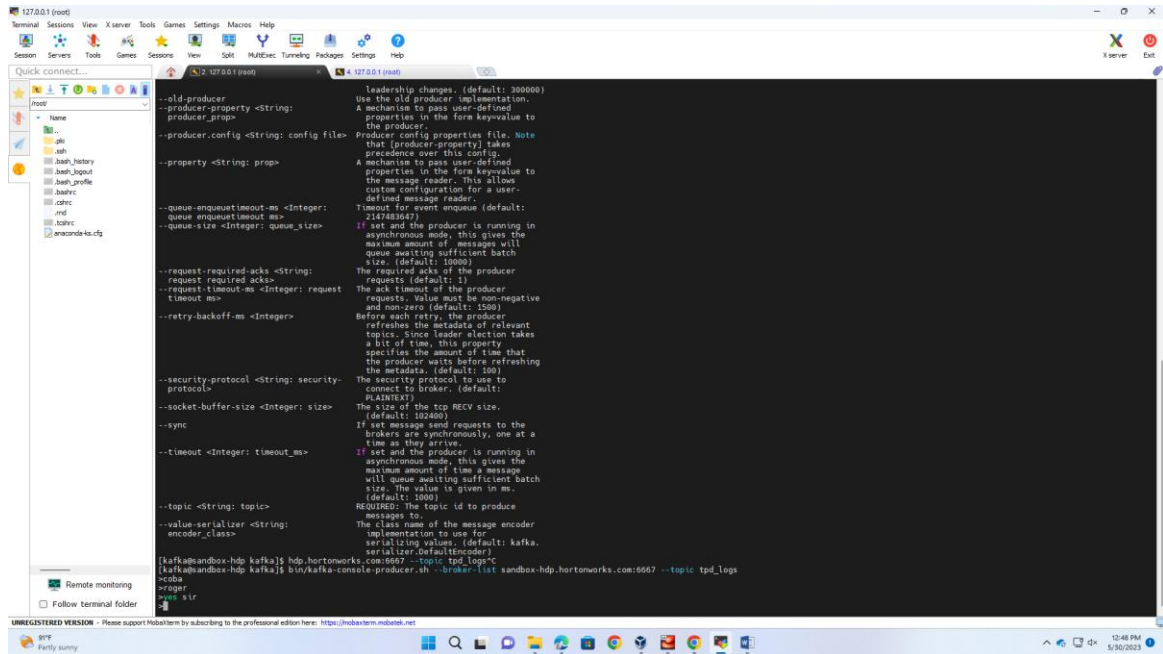
```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
(7,The Terminator,1987,8.4,5623)
(8,Harry Potter and the Deathly Hallows,1987,8.4,7563)
(9,Network,1990,7.2,6244)
(10,My Neighbor Totoro,2004,9.1,6956)
2023-05-28 11:42:05,351 (main) INFO org.apache.pig.Main - Pig script completed in 2 seconds and 672 milliseconds (2672 ms)
[root@sandbox-hdp ~]# su kafka
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ pwd
/root
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
[kafka@sandbox-hdp kafka]$ ls
bin conf config doc libs LICENSE logs NOTICE pids
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic tpd_logs
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not b
oth.
Created topic "tpd_logs".
[kafka@sandbox-hdp kafka]$
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --list --zookeeper localhost:2181
> d
Exception in thread "main" java.lang.NumberFormatException: For input string: "2181d"
    at java.lang.NumberFormatException.forInputString(NumberFormatException.java:65)
    at java.lang.Integer.parseInt(Integer.java:580)
    at java.lang.Integer.parseInt(Integer.java:615)
    at org.apache.zookeeper.client.ConnectStringParser.<init>(ConnectionStringParser.java:72)
    at org.apache.zookeeper.ZooKeeper.<init>(ZooKeeper.java:443)
    at org.apache.zookeeper.ZooKeeper.<init>(ZooKeeper.java:380)
    at kafka.zookeeper.ZooKeeperClient.<init>(ZooKeeperClient.scala:86)
    at kafka.zk.KafkaZkClient$.apply(KafkaZkClient.scala:1547)
    at kafka.admin.TopicCommand$.main(TopicCommand.scala:57)
    at kafka.admin.TopicCommand.main(TopicCommand.scala)
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --list --zookeeper localhost:2181
ATLAS ENTITIES
ATLAS_HOOK
consumer_offsets
tpd_logs
[kafka@sandbox-hdp kafka]$
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

4. Detail topic



```
127.0.0.1 (root)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help
Quick connect...
(10,My Neighbor Totoro,2004,9.1,6956)
2023-05-28 11:42:05,351 (main) INFO org.apache.pig.Main - Pig script completed in 2 seconds and 672 milliseconds (2672 ms)
[root@sandbox-hdp ~]# su kafka
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ pwd
/root
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
bash: cd: /usr/hdp/3.0.1-0-187/kafka: No such file or directory
[kafka@sandbox-hdp root]$ cd /usr/hdp/3.0.1-0-187/kafka
[kafka@sandbox-hdp kafka]$ ls
bin conf config doc libs LICENSE logs NOTICE pids
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic tpd_logs
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not b
oth.
Created topic "tpd_logs".
[kafka@sandbox-hdp kafka]$
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --list --zookeeper localhost:2181
> d
Exception in thread "main" java.lang.NumberFormatException: For input string: "2181d"
    at java.lang.NumberFormatException.forInputString(NumberFormatException.java:65)
    at java.lang.Integer.parseInt(Integer.java:580)
    at java.lang.Integer.parseInt(Integer.java:615)
    at org.apache.zookeeper.client.ConnectStringParser.<init>(ConnectionStringParser.java:72)
    at org.apache.zookeeper.ZooKeeper.<init>(ZooKeeper.java:443)
    at org.apache.zookeeper.ZooKeeper.<init>(ZooKeeper.java:380)
    at kafka.zookeeper.ZooKeeperClient.<init>(ZooKeeperClient.scala:86)
    at kafka.zk.KafkaZkClient$.apply(KafkaZkClient.scala:1547)
    at kafka.admin.TopicCommand$.main(TopicCommand.scala:57)
    at kafka.admin.TopicCommand.main(TopicCommand.scala)
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --list --zookeeper localhost:2181
ATLAS ENTITIES
ATLAS_HOOK
consumer_offsets
tpd_logs
[kafka@sandbox-hdp kafka]$ bin/kafka-topics.sh --describe --zookeeper localhost:2181 --topic tpd_logs
Topic:tpd_logs PartitionCount:1 ReplicationFactor:1 Configs:
Topic:tpd_logs Partition: 0 Leader: 1001 Replicas: 1001 Isr: 1001
[kafka@sandbox-hdp kafka]$
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

5. Producer



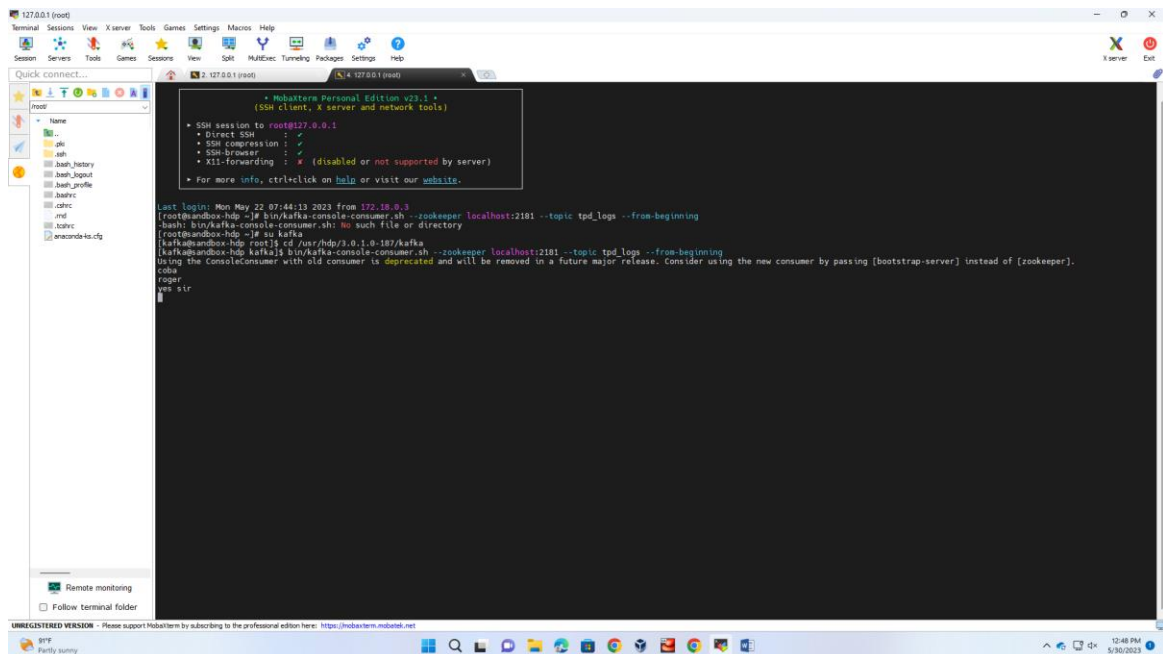
The screenshot shows a terminal window with the following content:

```
--old-producer
--producer-property <String: producer_prop>
--producer.config <String: config file>
--property <String: prop>
--queue-enqueue-timeout-ms <Integer: queue_enqueue_timeout_ms>
--queue-size <Integer: queue_size>
--request-required-acks <String: request_required_acks>
--request-timeout-ms <Integer: request_timeout_ms>
--retry-backoff-ms <Integer>
--security.protocol <String: security_protocol>
--socket-buffer-size <Integer: size>
--sync
--timeout <Integer: timeout_ms>
--topic <String: topic>
--value.serializer <String: encoder_class>

[kafka@andbox-hdp kafka]$ bin/kafka-topics.sh --zookeeper localhost:2181 --topic tpd_logs --create --partitions 1
[kafka@andbox-hdp kafka]$ bin/kafka-console-producer.sh --broker-list sandbox-hdp.hortonworks.com:6667 --topic tpd_logs
> cba
> raga
> yes sir
```

The terminal also displays detailed configuration options for the producer, such as `--old-producer`, `--producer-property`, `--producer.config`, `--property`, `--queue-enqueue-timeout-ms`, `--queue-size`, `--request-required-acks`, `--request-timeout-ms`, `--retry-backoff-ms`, `--security.protocol`, `--socket-buffer-size`, `--sync`, `--timeout`, `--topic`, and `--value.serializer`.

6. Consumer



The screenshot shows a terminal window with the following content:

```

* Mobaxterm Personal Edition v23.1 *
(SSH client, X server and network tools)

+ SSH session to root@127.0.0.1
+ Direct SSH
+ SSH compression : ✓
+ SSH forwarding : ✓
+ X11-forwarding : x (disabled or not supported by server)

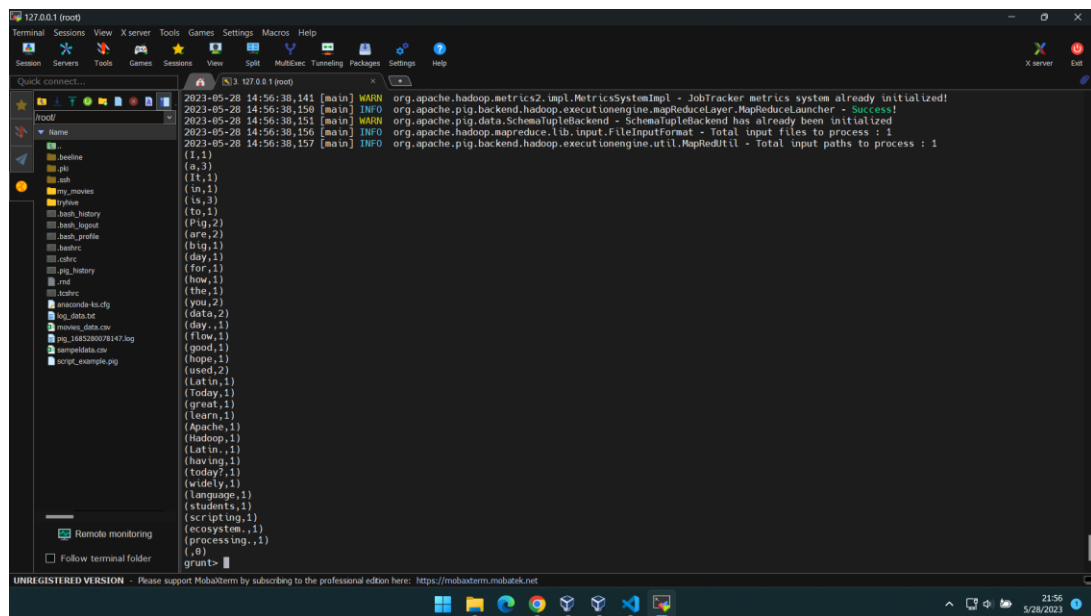
+ For more info, ctrl+click on help or visit our website.

Last login: Mon May 22 07:44:13 2023 from 172.28.0.1
[rosh@andbox-hdp ~]$ bin/kafka-console-consumer.sh --zookeeper localhost:2181 --topic tpd_logs --from-beginning
-bash: bin/kafka-console-consumer.sh: No such file or directory
[rosh@andbox-hdp ~]$ su kafka
[kafka@andbox-hdp root]$ cd /usr/hdp/3.0.1.0-187/kafka
[kafka@andbox-hdp kafka]$ bin/kafka-console-consumer.sh --zookeeper localhost:2181 --topic tpd_logs --from-beginning
Using the ConsoleConsumer with old consumer is deprecated and will be removed in a future major release. Consider using the new consumer by passing [bootstrap-server] instead of [zookeeper].
coba
roger
yes sir
```

The terminal also displays a message about the deprecated `ConsoleConsumer` and suggests using the new consumer by passing `[bootstrap-server]` instead of `[zookeeper]`.

Penugasan

1. Lakukan pemrosesan data menggunakan Pig script (interactive atau batch) untuk menghitung jumlah tiap kata pada tulisan di bawah ini.



```
lines = LOAD 'text_data.txt' AS (line:chararray);

words = FOREACH lines GENERATE FLATTEN(TOKENIZE(line)) as word;

grouped = GROUP words BY word ;

word_count = FOREACH grouped GENERATE group, COUNT(words);

DUMP word_count;
```


2. Screenshot output messages dari Producer ke Consumer menggunakan Apache Kafka pada langkah 7 dan 9

