PIG

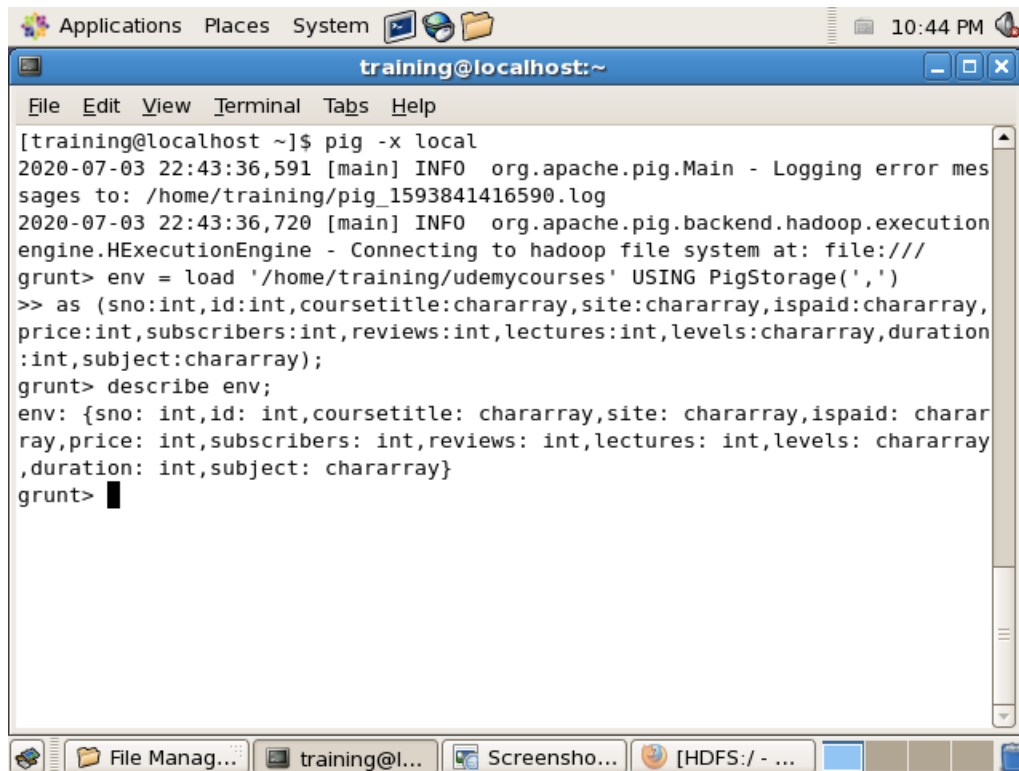**Load Data into Pig environments.**
Local Mode:
1. Open the terminal.

2. Write the Pig Command (pig –x local) on the terminal.

3. This will open the grunt shell in the local mode.

4. Now you need to create a table and load data into it from local file system.
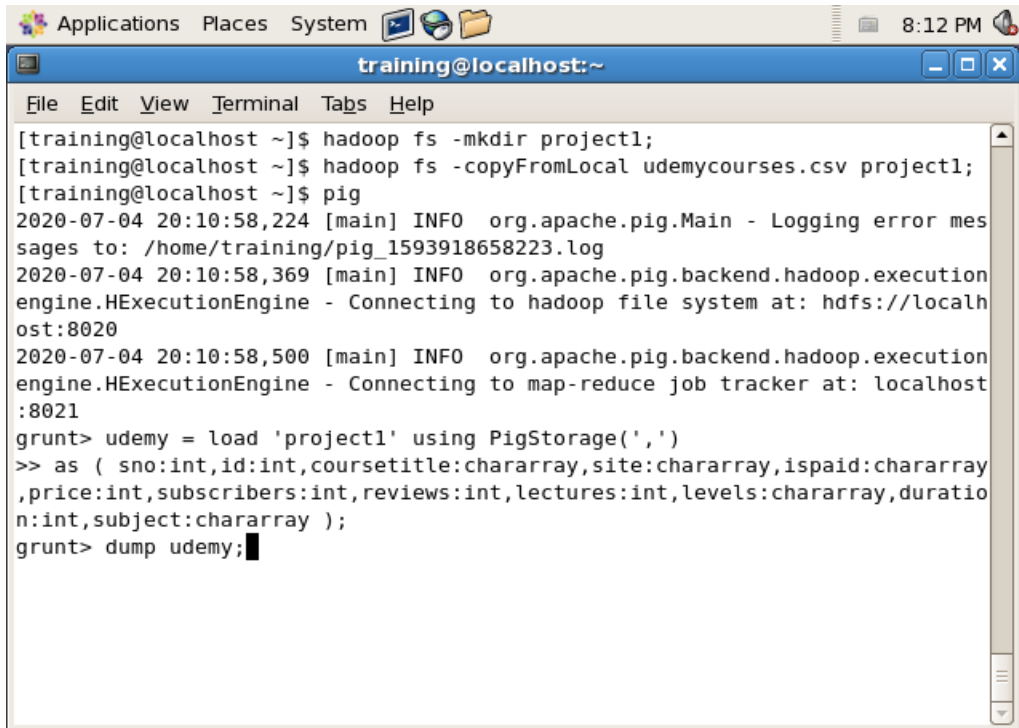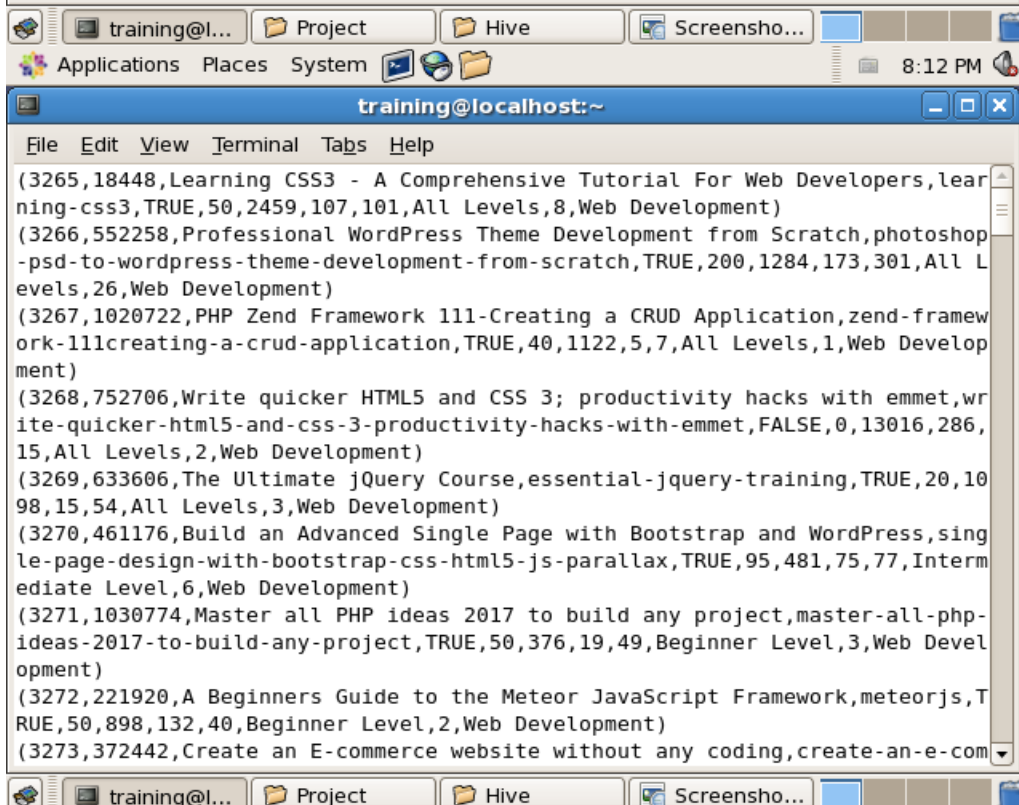

MapReduce Mode:
1. Open the terminal.

2. Write the Pig Command (pig –x mapreduce OR pig) on the terminal.

3. This will open the grunt shell in the mapreduce mode.

4. The Pig scripts are converted into MapReduce jobs and then executed on Hadoop cluster (hdfs).

training@localhost:~
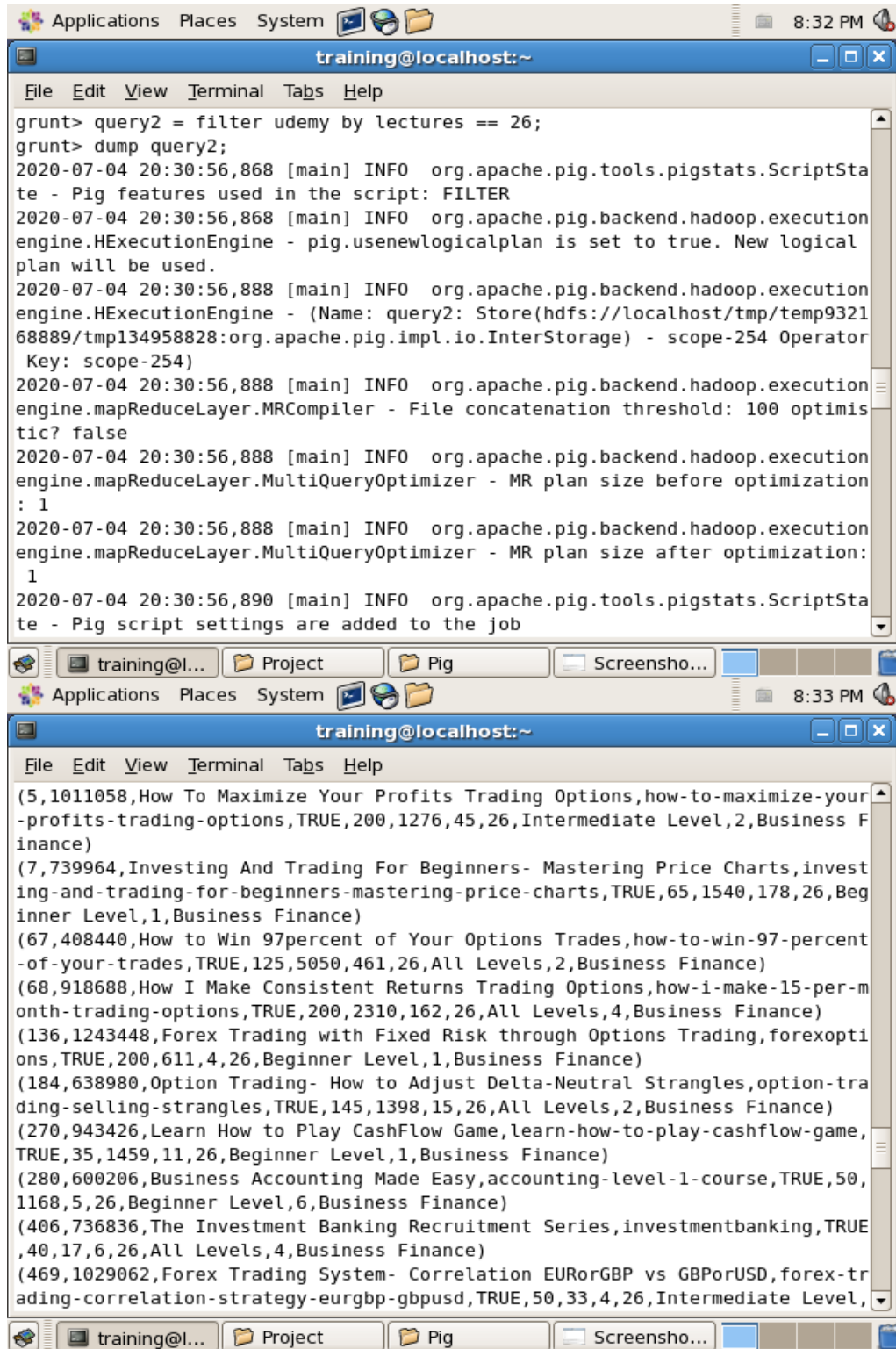
File   Edit   View   Terminal   Tabs   Help

```
[training@localhost ~]$ hadoop fs -mkdir project1;
[training@localhost ~]$ hadoop fs -copyFromLocal udemycourses.csv project1;
[training@localhost ~]$ pig
2020-07-04 20:10:58,224 [main] INFO  org.apache.pig.Main - Logging error mes
sages to: /home/training/pig_1593918658223.log
2020-07-04 20:10:58,369 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - Connecting to hadoop file system at: hdfs://localh
ost:8020
2020-07-04 20:10:58,500 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost
:8021
grunt> udemy = load 'project1' using PigStorage(',')
>> as ( sno:int,id:int,coursetitle:chararray,site:chararray,ispaid:chararray
,price:int,subscribers:int,reviews:int,lectures:int,levels:chararray,duratio
n:int,subject:chararray );
grunt> dump udemy;█
```

training@localhost:~

File   Edit   View   Terminal   Tabs   Help

```
(3265,18448,Learning CSS3 - A Comprehensive Tutorial For Web Developers,lear
ning-css3,TRUE,50,2459,107,101,All Levels,8,Web Development)
(3266,552258,Professional WordPress Theme Development from Scratch,photoshop
-psd-to-wordpress-theme-development-from-scratch,TRUE,200,1284,173,301,All L
evels,26,Web Development)
(3267,1020722,PHP Zend Framework 111-Creating a CRUD Application,zend-framew
ork-111creating-a-crud-application,TRUE,40,1122,5,7,All Levels,1,Web Develop
ment)
(3268,752706,Write quicker HTML5 and CSS 3; productivity hacks with emmet,wr
ite-quicker-html5-and-css-3-productivity-hacks-with-emmet,FALSE,0,13016,286,
15,All Levels,2,Web Development)
(3269,633606,The Ultimate jQuery Course,essential-jquery-training,TRUE,20,10
98,15,54,All Levels,3,Web Development)
(3270,461176,Build an Advanced Single Page with Bootstrap and WordPress,sing
le-page-design-with-bootstrap-css-html5-js-parallax,TRUE,95,481,75,77,Interm
ediate Level,6,Web Development)
(3271,1030774,Master all PHP ideas 2017 to build any project,master-all-php-
ideas-2017-to-build-any-project,TRUE,50,376,19,49,Beginner Level,3,Web Devel
opment)
(3272,221920,A Beginners Guide to the Meteor JavaScript Framework,meteorjs,T
RUE,50,898,132,40,Beginner Level,2,Web Development)
(3273,372442,Create an E-commerce website without any coding,create-an-e-com
```

Query 1: Filter the Udemy Courses acc. to lectures (lectures = 26)?
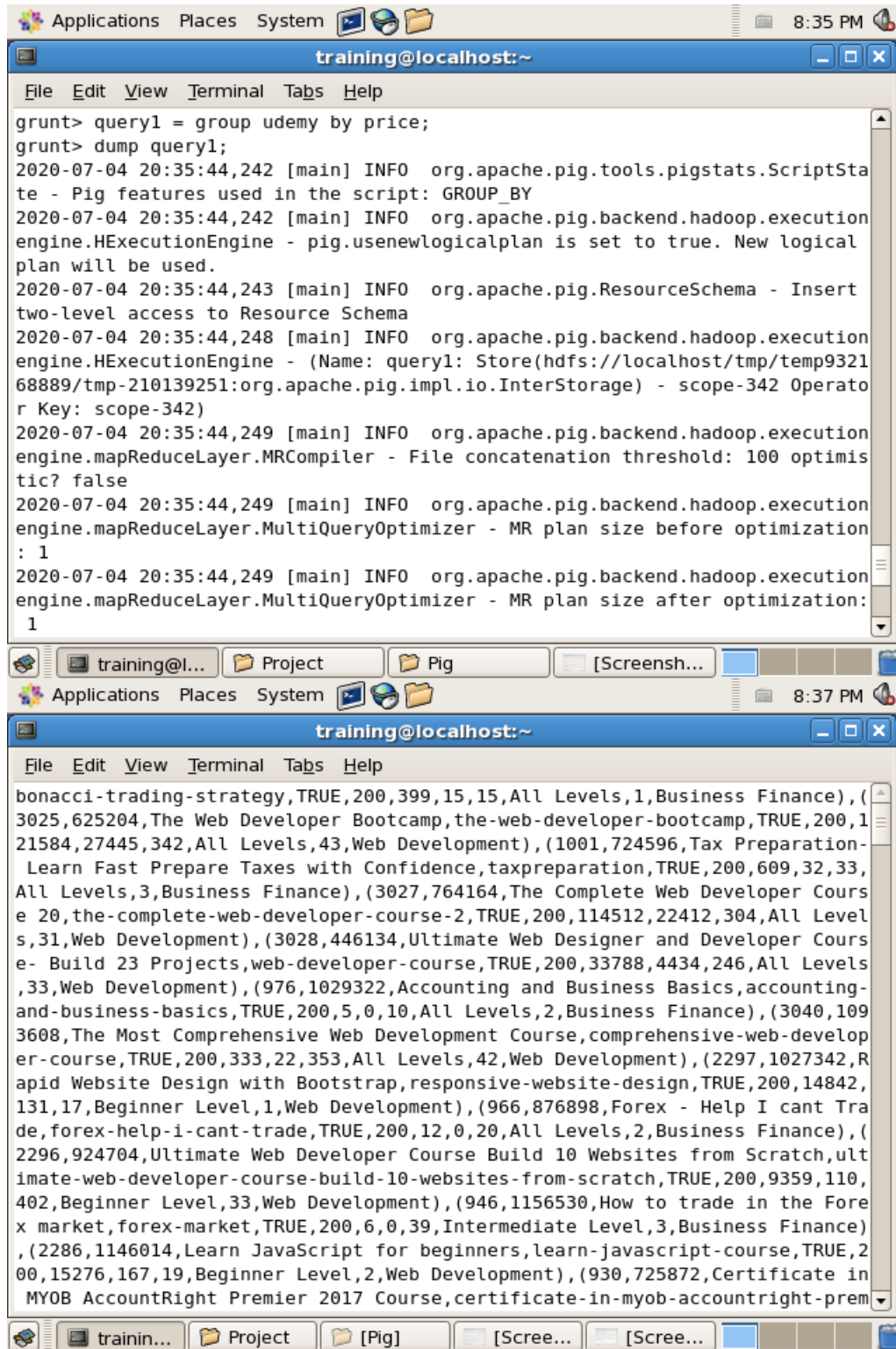
```
grunt> query2 = filter udemy by lectures == 26;
grunt> dump query2;
2020-07-04 20:30:56,868 [main] INFO  org.apache.pig.tools.pigstats.ScriptSta
te - Pig features used in the script: FILTER
2020-07-04 20:30:56,868 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - pig.usenewlogicalplan is set to true. New logical
plan will be used.
2020-07-04 20:30:56,888 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - (Name: query2: Store(hdfs://localhost/tmp/temp9321
68889/tmp134958828:org.apache.pig.impl.io.InterStorage) - scope-254 Operator
 Key: scope-254)
2020-07-04 20:30:56,888 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimis
tic? false
2020-07-04 20:30:56,888 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization
: 1
2020-07-04 20:30:56,888 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization:
 1
2020-07-04 20:30:56,890 [main] INFO  org.apache.pig.tools.pigstats.ScriptSta
te - Pig script settings are added to the job
```

```
(5,1011058,How To Maximize Your Profits Trading Options,how-to-maximize-your
-profits-trading-options,TRUE,200,1276,45,26,Intermediate Level,2,Business F
inance)
(7,739964,Investing And Trading For Beginners- Mastering Price Charts,invest
ing-and-trading-for-beginners-mastering-price-charts,TRUE,65,1540,178,26,Beg
inner Level,1,Business Finance)
(67,408440,How to Win 97percent of Your Options Trades,how-to-win-97-percent
-of-your-trades,TRUE,125,5050,461,26,All Levels,2,Business Finance)
(68,918688,How I Make Consistent Returns Trading Options,how-i-make-15-per-m
onth-trading-options,TRUE,200,2310,162,26,All Levels,4,Business Finance)
(136,1243448,Forex Trading with Fixed Risk through Options Trading,forexopti
ons,TRUE,200,611,4,26,Beginner Level,1,Business Finance)
(184,638980,Option Trading- How to Adjust Delta-Neutral Strangles,option-tra
ding-selling-strangles,TRUE,145,1398,15,26,All Levels,2,Business Finance)
(270,943426,Learn How to Play CashFlow Game,learn-how-to-play-cashflow-game,
TRUE,35,1459,11,26,Beginner Level,1,Business Finance)
(280,600206,Business Accounting Made Easy,accounting-level-1-course,TRUE,50,
1168,5,26,Beginner Level,6,Business Finance)
(406,736836,The Investment Banking Recruitment Series,investmentbanking,TRUE
,40,17,6,26,All Levels,4,Business Finance)
(469,1029062,Forex Trading System- Correlation EURorGBP vs GBPorUSD,forex-tr
ading-correlation-strategy-eurgbp-gbpusd,TRUE,50,33,4,26,Intermediate Level,
```

Query 2: Group the Udemy Courses having same price?

```
training@localhost:~                                    _ □ ✕

File  Edit  View  Terminal  Tabs  Help

grunt> query1 = group udemy by price;
grunt> dump query1;
2020-07-04 20:35:44,242 [main] INFO  org.apache.pig.tools.pigstats.ScriptSta
te - Pig features used in the script: GROUP_BY
2020-07-04 20:35:44,242 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - pig.usenewlogicalplan is set to true. New logical
plan will be used.
2020-07-04 20:35:44,243 [main] INFO  org.apache.pig.ResourceSchema - Insert
two-level access to Resource Schema
2020-07-04 20:35:44,248 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - (Name: query1: Store(hdfs://localhost/tmp/temp9321
68889/tmp-210139251:org.apache.pig.impl.io.InterStorage) - scope-342 Operato
r Key: scope-342)
2020-07-04 20:35:44,249 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimis
tic? false
2020-07-04 20:35:44,249 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization
: 1
2020-07-04 20:35:44,249 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization:
 1
```
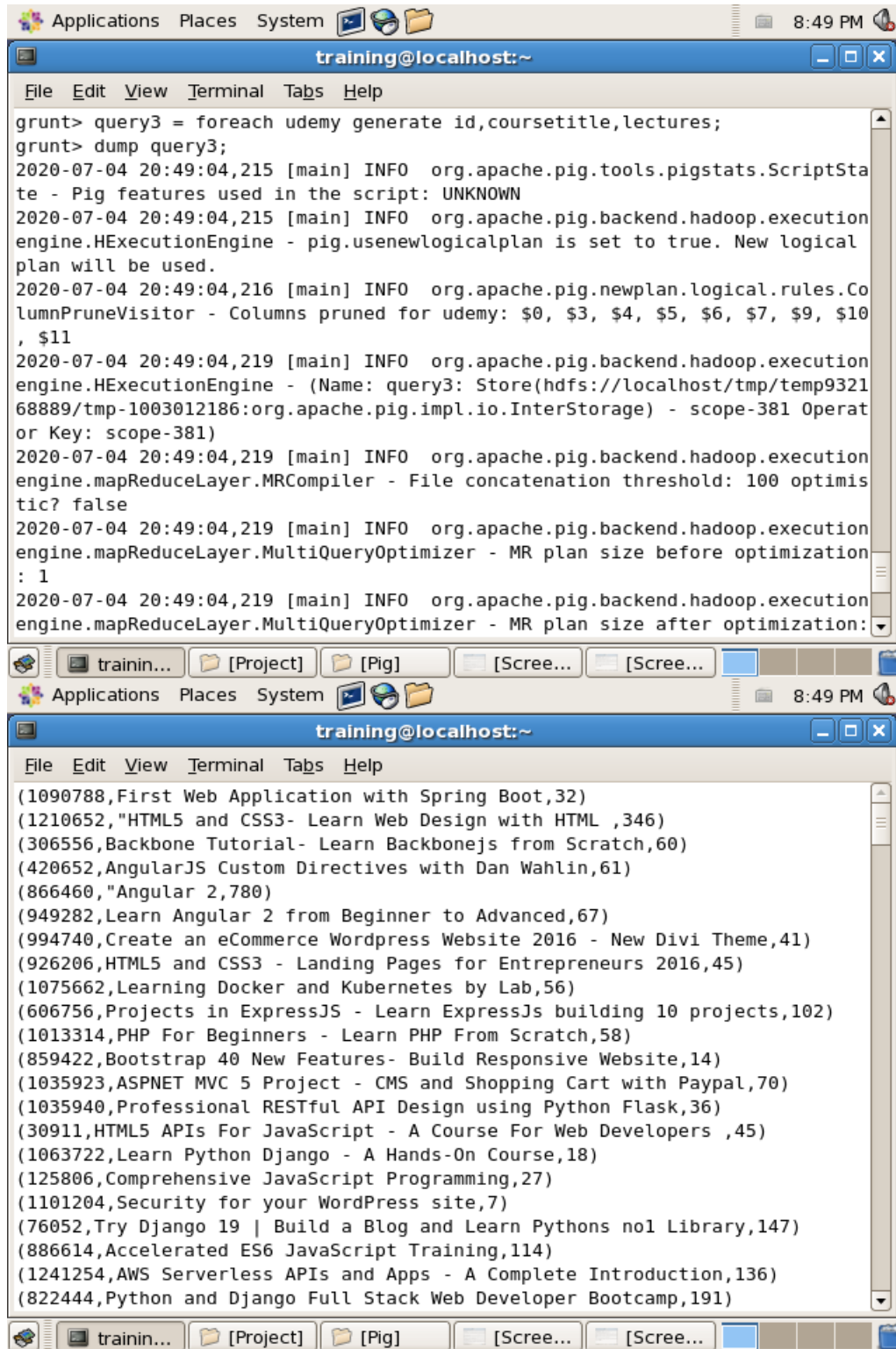
```
training@localhost:~                                    _ □ ✕

File  Edit  View  Terminal  Tabs  Help

bonacci-trading-strategy,TRUE,200,399,15,15,All Levels,1,Business Finance),(
3025,625204,The Web Developer Bootcamp,the-web-developer-bootcamp,TRUE,200,1
21584,27445,342,All Levels,43,Web Development),(1001,724596,Tax Preparation-
 Learn Fast Prepare Taxes with Confidence,taxpreparation,TRUE,200,609,32,33,
All Levels,3,Business Finance),(3027,764164,The Complete Web Developer Cours
e 20,the-complete-web-developer-course-2,TRUE,200,114512,22412,304,All Level
s,31,Web Development),(3028,446134,Ultimate Web Designer and Developer Cours
e- Build 23 Projects,web-developer-course,TRUE,200,33788,4434,246,All Levels
,33,Web Development),(976,1029322,Accounting and Business Basics,accounting-
and-business-basics,TRUE,200,5,0,10,All Levels,2,Business Finance),(3040,109
3608,The Most Comprehensive Web Development Course,comprehensive-web-develop
er-course,TRUE,200,333,22,353,All Levels,42,Web Development),(2297,1027342,R
apid Website Design with Bootstrap,responsive-website-design,TRUE,200,14842,
131,17,Beginner Level,1,Web Development),(966,876898,Forex - Help I cant Tra
de,forex-help-i-cant-trade,TRUE,200,12,0,20,All Levels,2,Business Finance),(
2296,924704,Ultimate Web Developer Course Build 10 Websites from Scratch,ult
imate-web-developer-course-build-10-websites-from-scratch,TRUE,200,9359,110,
402,Beginner Level,33,Web Development),(946,1156530,How to trade in the Fore
x market,forex-market,TRUE,200,6,0,39,Intermediate Level,3,Business Finance)
,(2286,1146014,Learn JavaScript for beginners,learn-javascript-course,TRUE,2
00,15276,167,19,Beginner Level,2,Web Development),(930,725872,Certificate in
 MYOB AccountRight Premier 2017 Course,certificate-in-myob-accountright-prem
```

Query 3: For each courses generate specified data-(id,coursetitle,lectures)?

training@localhost:~

File Edit View Terminal Tabs Help

```
grunt> query3 = foreach udemy generate id,coursetitle,lectures;
grunt> dump query3;
2020-07-04 20:49:04,215 [main] INFO  org.apache.pig.tools.pigstats.ScriptSta
te - Pig features used in the script: UNKNOWN
2020-07-04 20:49:04,215 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - pig.usenewlogicalplan is set to true. New logical
plan will be used.
2020-07-04 20:49:04,216 [main] INFO  org.apache.pig.newplan.logical.rules.Co
lumnPruneVisitor - Columns pruned for udemy: $0, $3, $4, $5, $6, $7, $9, $10
, $11
2020-07-04 20:49:04,219 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - (Name: query3: Store(hdfs://localhost/tmp/temp9321
68889/tmp-1003012186:org.apache.pig.impl.io.InterStorage) - scope-381 Operat
or Key: scope-381)
2020-07-04 20:49:04,219 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimis
tic? false
2020-07-04 20:49:04,219 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization
: 1
2020-07-04 20:49:04,219 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization:
```

trainin...   [Project]   [Pig]   [Scree...   [Scree...]

training@localhost:~

File Edit View Terminal Tabs Help

```
(1090788,First Web Application with Spring Boot,32)
(1210652,"HTML5 and CSS3- Learn Web Design with HTML ,346)
(306556,Backbone Tutorial- Learn Backbonejs from Scratch,60)
(420652,AngularJS Custom Directives with Dan Wahlin,61)
(866460,"Angular 2,780)
(949282,Learn Angular 2 from Beginner to Advanced,67)
(994740,Create an eCommerce Wordpress Website 2016 - New Divi Theme,41)
(926206,HTML5 and CSS3 - Landing Pages for Entrepreneurs 2016,45)
(1075662,Learning Docker and Kubernetes by Lab,56)
(606756,Projects in ExpressJS - Learn ExpressJs building 10 projects,102)
(1013314,PHP For Beginners - Learn PHP From Scratch,58)
(859422,Bootstrap 40 New Features- Build Responsive Website,14)
(1035923,ASPNET MVC 5 Project - CMS and Shopping Cart with Paypal,70)
(1035940,Professional RESTful API Design using Python Flask,36)
(30911,HTML5 APIs For JavaScript - A Course For Web Developers ,45)
(1063722,Learn Python Django - A Hands-On Course,18)
(125806,Comprehensive JavaScript Programming,27)
(1101204,Security for your WordPress site,7)
(76052,Try Django 19 | Build a Blog and Learn Pythons no1 Library,147)
(886614,Accelerated ES6 JavaScript Training,114)
(1241254,AWS Serverless APIs and Apps - A Complete Introduction,136)
(822444,Python and Django Full Stack Web Developer Bootcamp,191)
```

trainin...   [Project]   [Pig]   [Scree...   [Scree...]