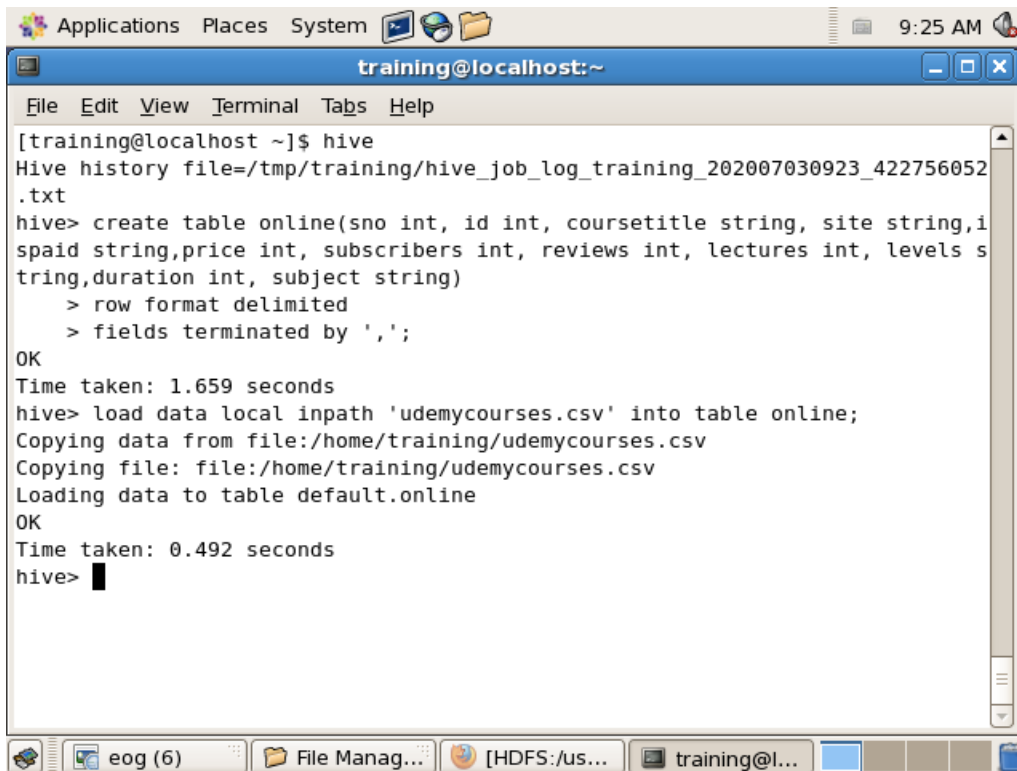


## HIVE

### Load Data into hive environments.

1. Open the terminal.
2. Write the Hive command on the terminal.
3. This will open the hive shell.
4. Now you need to create the table in hive and load data into it from HDFS.

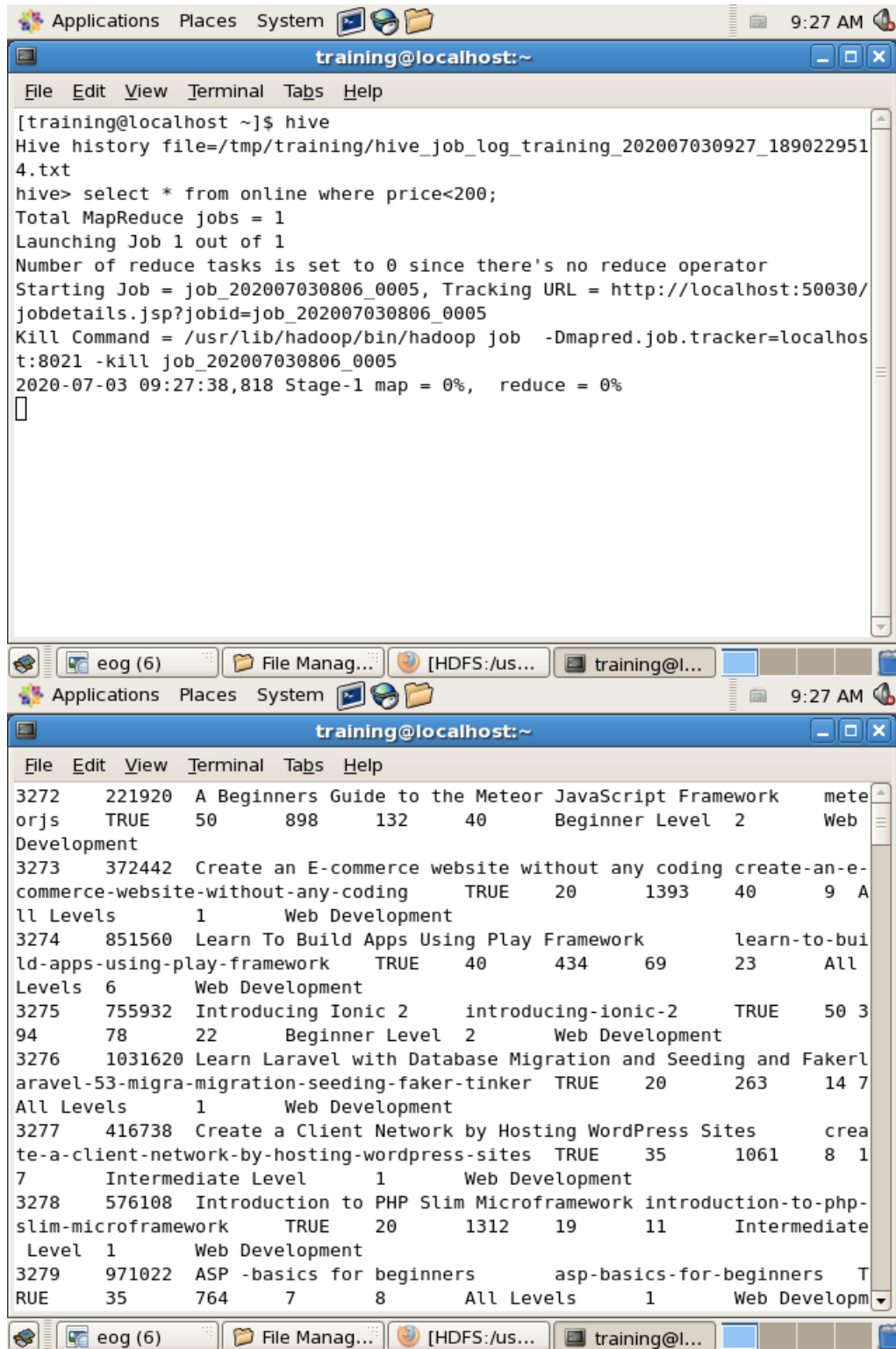


The screenshot shows a terminal window titled 'training@localhost:~'. The window contains the following text:

```
[training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_202007030923_422756052
.txt
hive> create table online(sno int, id int, coursetitle string, site string, i
spaid string, price int, subscribers int, reviews int, lectures int, levels s
tring, duration int, subject string)
    > row format delimited
    > fields terminated by ',';
OK
Time taken: 1.659 seconds
hive> load data local inpath 'udemycourses.csv' into table online;
Copying data from file:/home/training/udemycourses.csv
Copying file: file:/home/training/udemycourses.csv
Loading data to table default.online
OK
Time taken: 0.492 seconds
hive> █
```

The terminal window has a menu bar with 'File', 'Edit', 'View', 'Terminal', 'Tabs', and 'Help'. The top status bar shows 'Applications', 'Places', 'System', and the time '9:25 AM'. The bottom taskbar shows icons for 'eog (6)', 'File Manag...', '[HDFS:/us...', and 'training@l...'.

## Query 1: Which courses are available at affordable prices(price<200)?



The image shows two screenshots of a Linux terminal window. The top screenshot shows the execution of a Hive query. The bottom screenshot shows the results of the query, which is a table of courses.

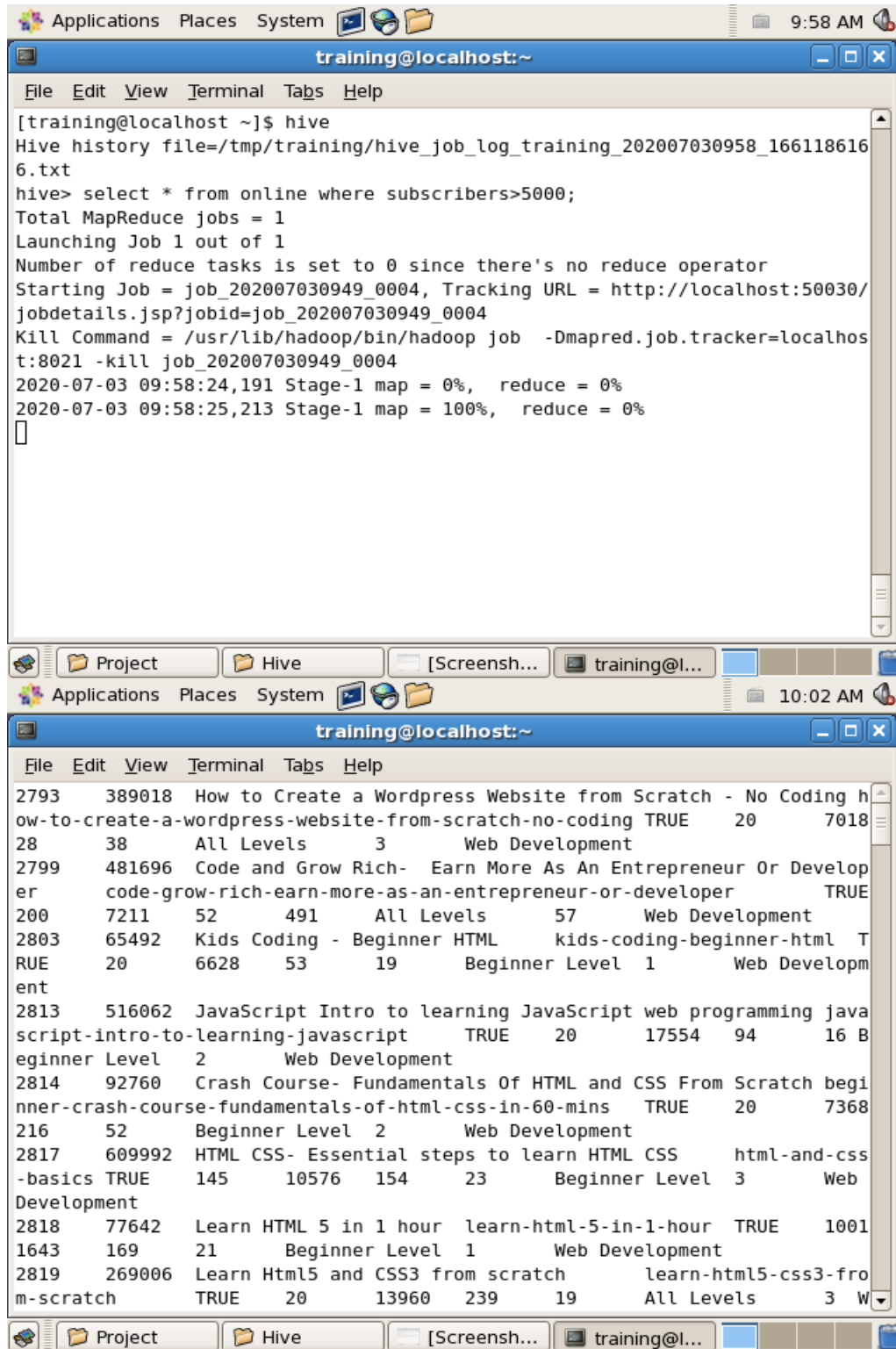
**Top Screenshot: Hive Query Execution**

```
[training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_202007030927_1890229514.txt
hive> select * from online where price<200;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_202007030806_0005, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_202007030806_0005
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007030806_0005
2020-07-03 09:27:38,818 Stage-1 map = 0%, reduce = 0%
[]
```

**Bottom Screenshot: Query Results**

id	course_id	course_name	price	author	levels	category	status	created_at	updated_at
3272	221920	A Beginners Guide to the Meteor JavaScript Framework	50	mete	Beginner Level 2	Web Development	TRUE	898	132
3273	372442	Create an E-commerce website without any coding	20	create-an-e-commerce-website-without-any-coding	Web Development	TRUE	1393	40	9
3274	851560	Learn To Build Apps Using Play Framework	40	learn-to-build-apps-using-play-framework	Web Development	TRUE	434	69	23
3275	755932	Introducing Ionic 2	50	introducing-ionic-2	Beginner Level 2	Web Development	TRUE	78	22
3276	1031620	Learn Laravel with Database Migration and Seeding and Faker	20	aravel-53-migra-migration-seeding-faker-tinker	Web Development	TRUE	263	14	7
3277	416738	Create a Client Network by Hosting WordPress Sites	35	create-a-client-network-by-hosting-wordpress-sites	Intermediate Level 1	Web Development	TRUE	1061	8
3278	576108	Introduction to PHP Slim Microframework	20	introduction-to-php-slim-microframework	Web Development	TRUE	1312	19	11
3279	971022	ASP -basics for beginners	35	asp-basics-for-beginners	All Levels 1	Web Development	TRUE	764	7

## Query 2: Which courses are the best rated courses(subscribers>5000)?

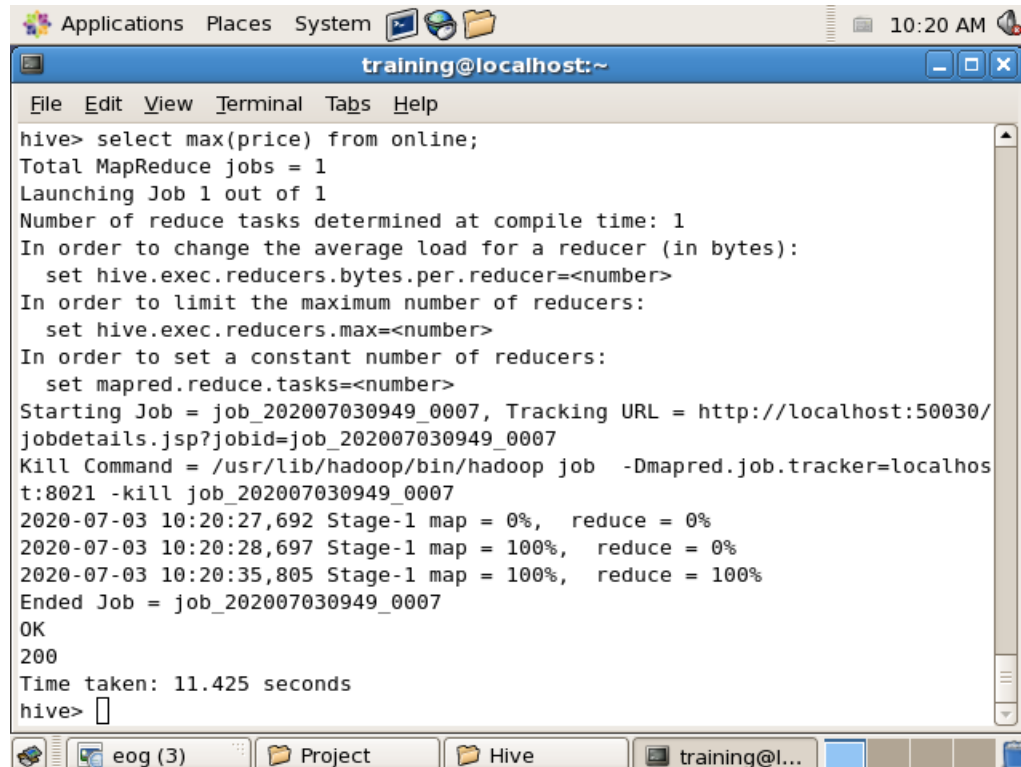


The image shows two screenshots of a terminal window titled "training@localhost:~". The top screenshot shows the execution of a Hive query to select courses with more than 5000 subscribers. The bottom screenshot shows the results of the query, which are displayed as a table with 10 columns: course\_id, title, rating, subscribers, duration, level, category, and several other metrics.

```
[training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_202007030958_1661186166.txt
hive> select * from online where subscribers>5000;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_202007030949_0004, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_202007030949_0004
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007030949_0004
2020-07-03 09:58:24,191 Stage-1 map = 0%, reduce = 0%
2020-07-03 09:58:25,213 Stage-1 map = 100%, reduce = 0%
[]
```

```
2793 389018 How to Create a Wordpress Website from Scratch - No Coding h
ow-to-create-a-wordpress-website-from-scratch-no-coding TRUE 20 7018
28 38 All Levels 3 Web Development
2799 481696 Code and Grow Rich- Earn More As An Entrepreneur Or Develop
er code-grow-rich-earn-more-as-an-entrepreneur-or-developer TRUE
200 7211 52 491 All Levels 57 Web Development
2803 65492 Kids Coding - Beginner HTML kids-coding-beginner-html T
RUE 20 6628 53 19 Beginner Level 1 Web Developm
ent
2813 516062 JavaScript Intro to learning JavaScript web programming java
script-intro-to-learning-javascript TRUE 20 17554 94 16 B
eginner Level 2 Web Development
2814 92760 Crash Course- Fundamentals Of HTML and CSS From Scratch begi
nner-crash-course-fundamentals-of-html-css-in-60-mins TRUE 20 7368
216 52 Beginner Level 2 Web Development
2817 609992 HTML CSS- Essential steps to learn HTML CSS html-and-css
-basics TRUE 145 10576 154 23 Beginner Level 3 Web
Development
2818 77642 Learn HTML 5 in 1 hour learn-html-5-in-1-hour TRUE 1001
1643 169 21 Beginner Level 1 Web Development
2819 269006 Learn Html5 and CSS3 from scratch learn-html5-css3-fro
m-scratch TRUE 20 13960 239 19 All Levels 3 W
```

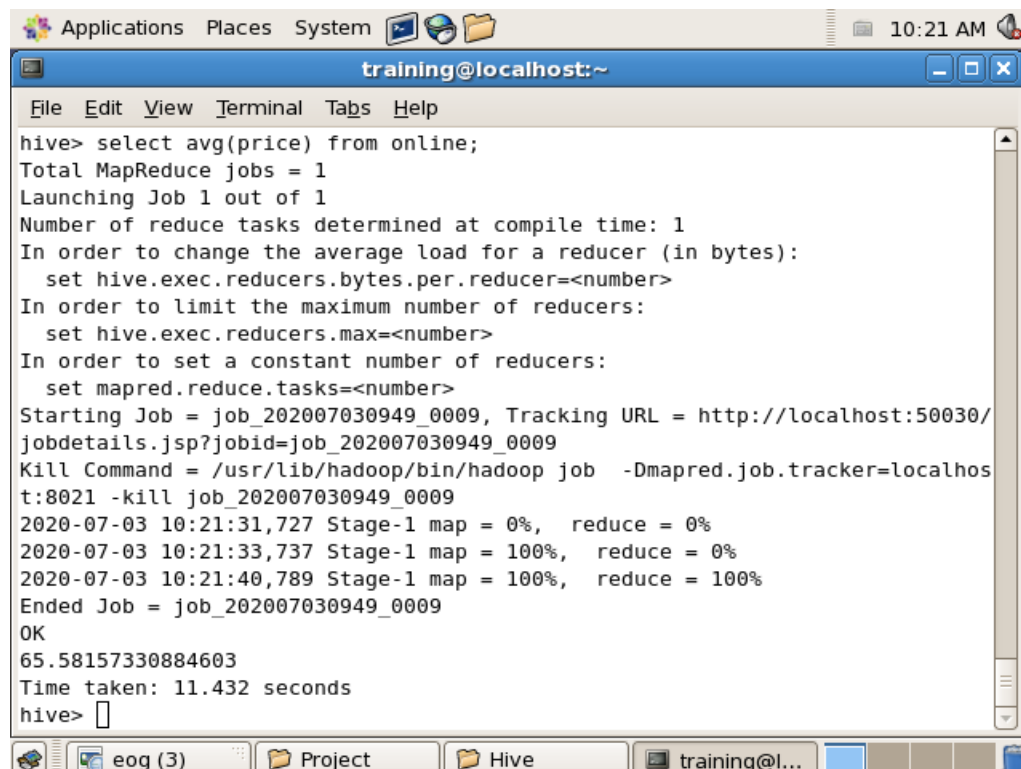
### Query 3: What is the highest price for a course(max(price))?



A terminal window titled 'training@localhost:~' showing the execution of a Hive query. The window has a menu bar with 'File', 'Edit', 'View', 'Terminal', 'Tabs', and 'Help'. The terminal output shows the query 'select max(price) from online;' being executed. It displays job details, including the number of MapReduce jobs (1), the number of reduce tasks (1), and the time taken (11.425 seconds). The output also shows the progress of the job, with Stage-1 map and reduce percentages reaching 100%.

```
hive> select max(price) from online;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202007030949_0007, Tracking URL = http://localhost:50030/
jobdetails.jsp?jobid=job_202007030949_0007
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007030949_0007
2020-07-03 10:20:27,692 Stage-1 map = 0%, reduce = 0%
2020-07-03 10:20:28,697 Stage-1 map = 100%, reduce = 0%
2020-07-03 10:20:35,805 Stage-1 map = 100%, reduce = 100%
Ended Job = job_202007030949_0007
OK
200
Time taken: 11.425 seconds
hive>
```

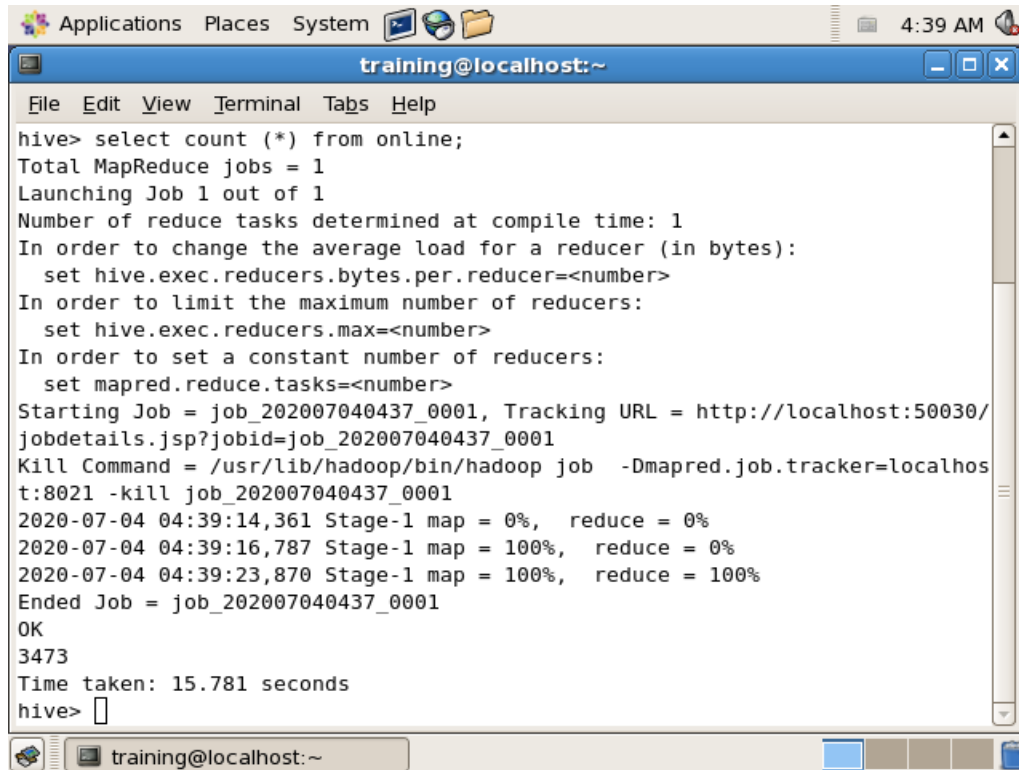
### Query 4: What is the average price from all paid courses?



A terminal window titled 'training@localhost:~' showing the execution of a Hive query. The window has a menu bar with 'File', 'Edit', 'View', 'Terminal', 'Tabs', and 'Help'. The terminal output shows the query 'select avg(price) from online;' being executed. It displays job details, including the number of MapReduce jobs (1), the number of reduce tasks (1), and the time taken (11.432 seconds). The output also shows the progress of the job, with Stage-1 map and reduce percentages reaching 100%. The final result of the query is 65.58157330884603.

```
hive> select avg(price) from online;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202007030949_0009, Tracking URL = http://localhost:50030/
jobdetails.jsp?jobid=job_202007030949_0009
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007030949_0009
2020-07-03 10:21:31,727 Stage-1 map = 0%, reduce = 0%
2020-07-03 10:21:33,737 Stage-1 map = 100%, reduce = 0%
2020-07-03 10:21:40,789 Stage-1 map = 100%, reduce = 100%
Ended Job = job_202007030949_0009
OK
65.58157330884603
Time taken: 11.432 seconds
hive>
```

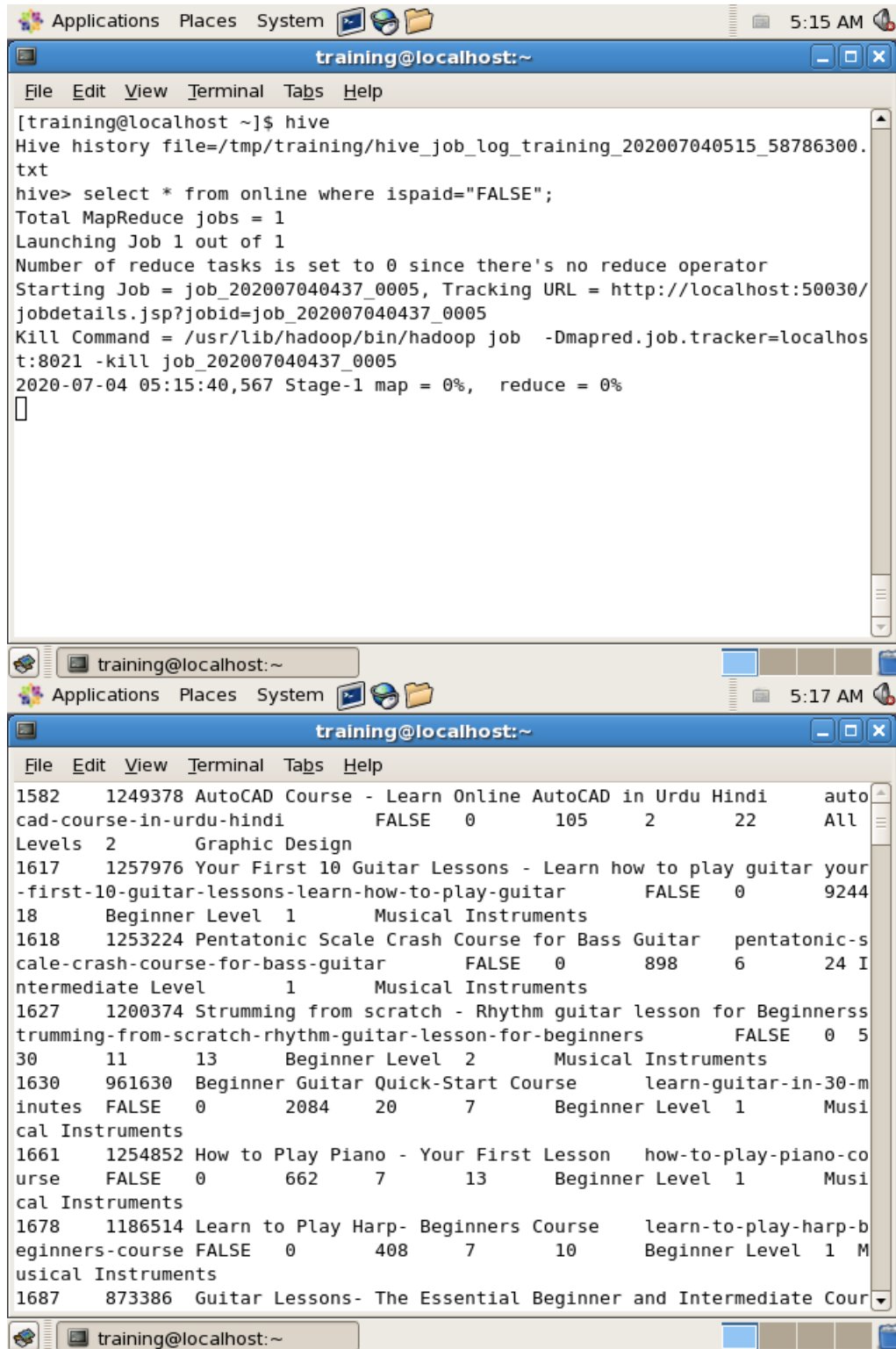
### Query 5: How many course entries are there in present data?



The screenshot shows a terminal window titled 'training@localhost:~' with a menu bar (File, Edit, View, Terminal, Tabs, Help) and a system bar at the top (Applications, Places, System, 4:39 AM). The terminal displays the output of a Hive query: 'select count (\*) from online;'. The output includes job details, progress updates, and completion status.

```
hive> select count (*) from online;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202007040437_0001, Tracking URL = http://localhost:50030/
jobdetails.jsp?jobid=job_202007040437_0001
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007040437_0001
2020-07-04 04:39:14,361 Stage-1 map = 0%,  reduce = 0%
2020-07-04 04:39:16,787 Stage-1 map = 100%,  reduce = 0%
2020-07-04 04:39:23,870 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_202007040437_0001
OK
3473
Time taken: 15.781 seconds
hive> 
```

## Query 6: Which courses are available at free of cost(ispaid=FALSE)?

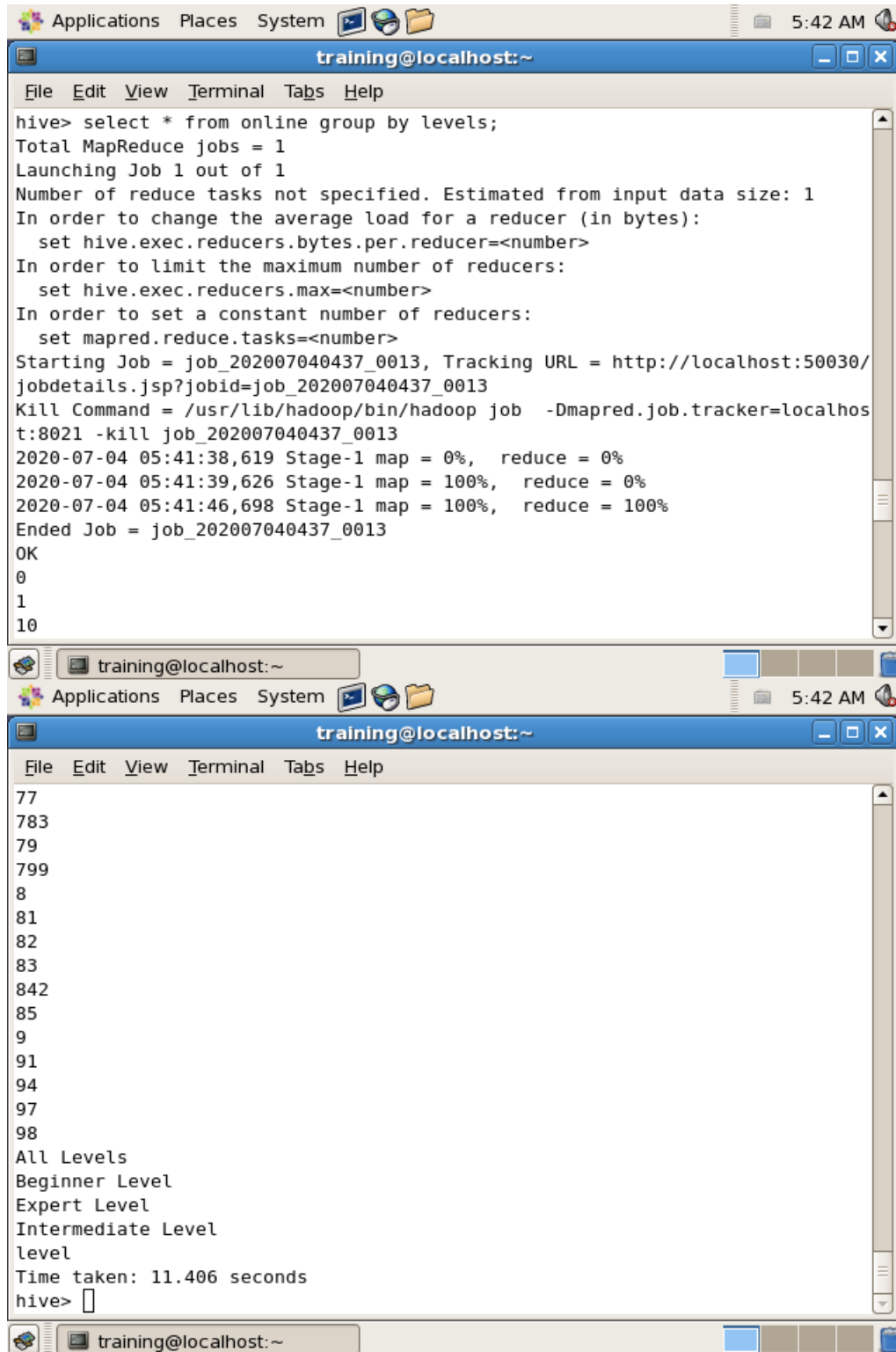


The image shows two screenshots of a Linux terminal window. The top screenshot shows the execution of a Hive query to select courses where 'ispaid' is 'FALSE'. The bottom screenshot shows the output of the query, which is a table of course details.

```
[training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_202007040515_58786300.txt
hive> select * from online where ispaid="FALSE";
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_202007040437_0005, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_202007040437_0005
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007040437_0005
2020-07-04 05:15:40,567 Stage-1 map = 0%, reduce = 0%
```

id	course_id	course_name	ispaid	duration	levels	category
1582	1249378	AutoCAD Course - Learn Online AutoCAD in Urdu Hindi	FALSE	105	2	auto cad-course-in-urdu-hindi
1617	1257976	Your First 10 Guitar Lessons - Learn how to play guitar your first-10-guitar-lessons-learn-how-to-play-guitar	FALSE	0	9244	Beginner Level 1 Musical Instruments
1618	1253224	Pentatonic Scale Crash Course for Bass Guitar	FALSE	898	6	pentatonic-scale-crash-course-for-bass-guitar
1627	1200374	Strumming from scratch - Rhythm guitar lesson for Beginners	FALSE	0	5	strumming-from-scratch-rhythm-guitar-lesson-for-beginners
1630	961630	Beginner Guitar Quick-Start Course	FALSE	2084	20	7
1661	1254852	How to Play Piano - Your First Lesson	FALSE	662	7	13
1678	1186514	Learn to Play Harp- Beginners Course	FALSE	408	7	10
1687	873386	Guitar Lessons- The Essential Beginner and Intermediate Cour				

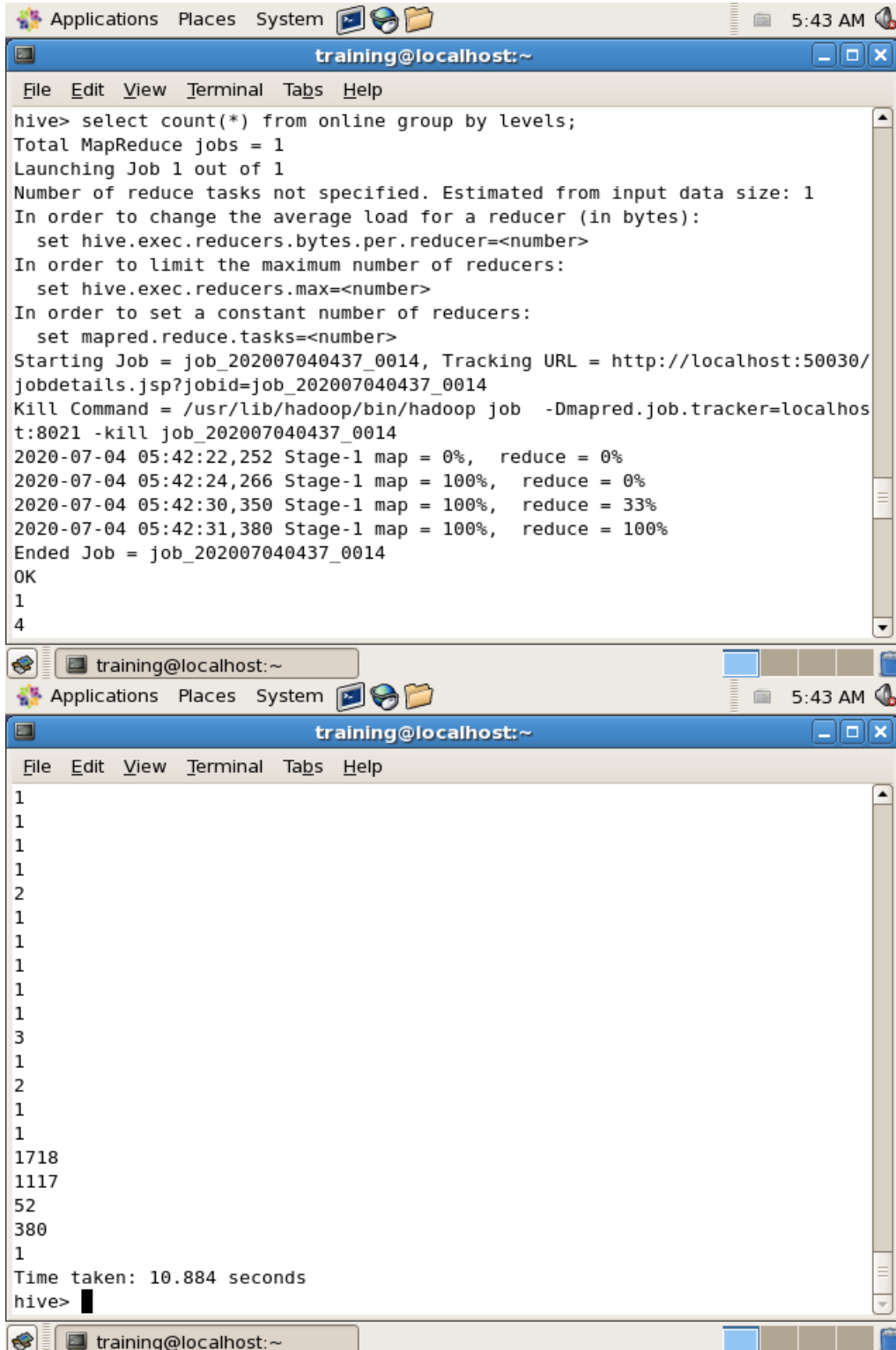
## Query 7: Group the Udemy courses by each levels?



```
Applications Places System 5:42 AM
training@localhost:~
File Edit View Terminal Tabs Help
hive> select * from online group by levels;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202007040437_0013, Tracking URL = http://localhost:50030/
jobdetails.jsp?jobid=job_202007040437_0013
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007040437_0013
2020-07-04 05:41:38,619 Stage-1 map = 0%, reduce = 0%
2020-07-04 05:41:39,626 Stage-1 map = 100%, reduce = 0%
2020-07-04 05:41:46,698 Stage-1 map = 100%, reduce = 100%
Ended Job = job_202007040437_0013
OK
0
1
10

training@localhost:~
Applications Places System 5:42 AM
training@localhost:~
File Edit View Terminal Tabs Help
77
783
79
799
8
81
82
83
842
85
9
91
94
97
98
All Levels
Beginner Level
Expert Level
Intermediate Level
level
Time taken: 11.406 seconds
hive>
```

## Query 8: How many courses are there in each level(All,Beginner,Intermediate,Expert)?



The image shows two screenshots of a terminal window on a Linux system, likely Ubuntu, with the username 'training' and hostname 'localhost'. The top screenshot shows the execution of a Hive query to count courses by level. The bottom screenshot shows the results of the query.

**Terminal Window 1 (Top):**

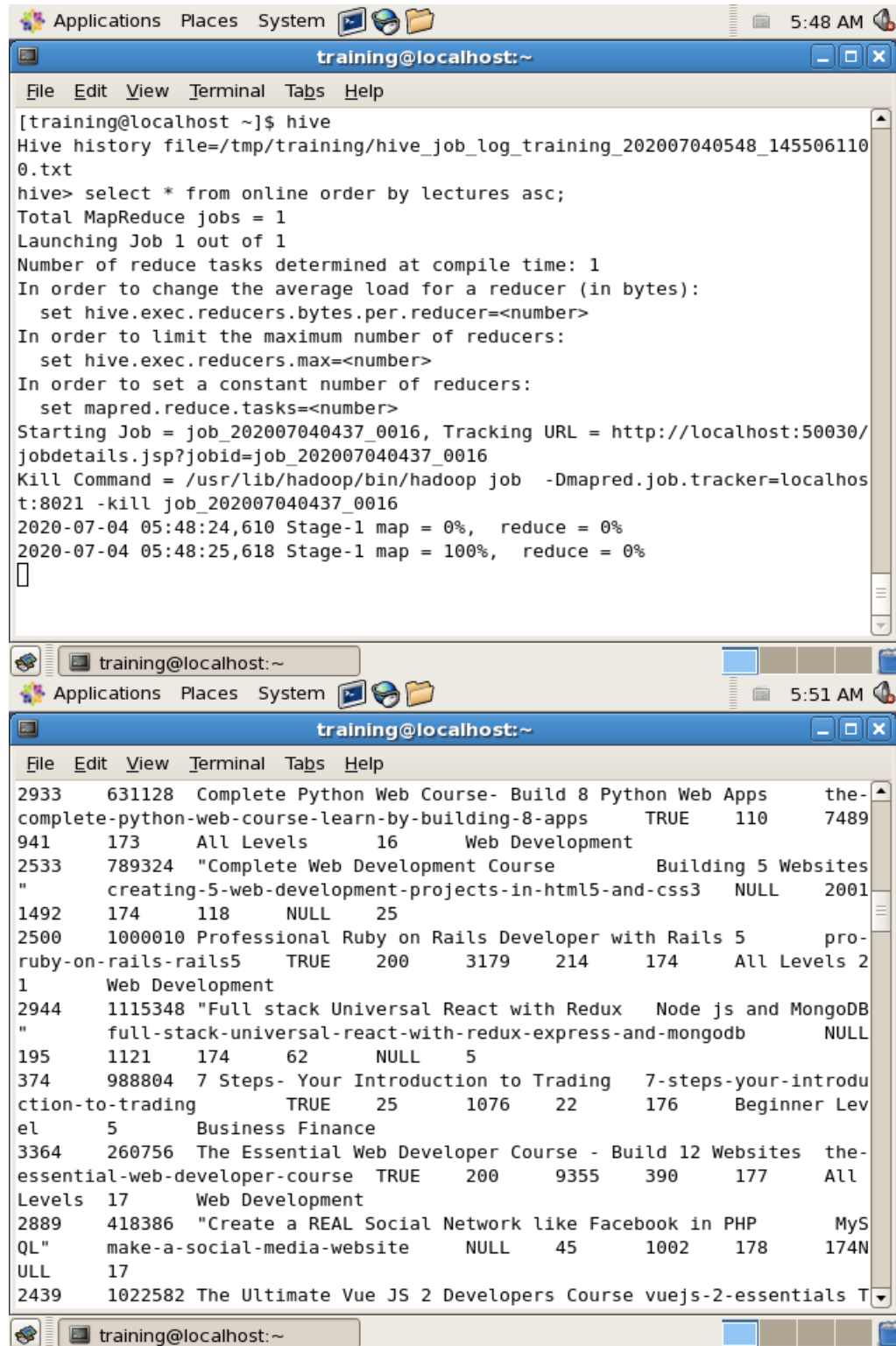
```
File Edit View Terminal Tabs Help
hive> select count(*) from online group by levels;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202007040437_0014, Tracking URL = http://localhost:50030/
jobdetails.jsp?jobid=job_202007040437_0014
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007040437_0014
2020-07-04 05:42:22,252 Stage-1 map = 0%, reduce = 0%
2020-07-04 05:42:24,266 Stage-1 map = 100%, reduce = 0%
2020-07-04 05:42:30,350 Stage-1 map = 100%, reduce = 33%
2020-07-04 05:42:31,380 Stage-1 map = 100%, reduce = 100%
Ended Job = job_202007040437_0014
OK
1
4
```

**Terminal Window 2 (Bottom):**

```
File Edit View Terminal Tabs Help
1
1
1
1
2
1
1
1
1
1
3
1
2
1
1
1718
1117
52
380
1
Time taken: 10.884 seconds
hive>
```



## Query 9: Select courses according to ascending order of lectures?



The image shows two screenshots of a Linux terminal window. The top screenshot shows the execution of a Hive query to select all courses ordered by the number of lectures in ascending order. The bottom screenshot shows the output of the query, which is a table of course details.

```
[training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_202007040548_1455061100.txt
hive> select * from online order by lectures asc;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_202007040437_0016, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_202007040437_0016
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007040437_0016
2020-07-04 05:48:24,610 Stage-1 map = 0%, reduce = 0%
2020-07-04 05:48:25,618 Stage-1 map = 100%, reduce = 0%
[]
```

```
2933 631128 Complete Python Web Course- Build 8 Python Web Apps the-
complete-python-web-course-learn-by-building-8-apps TRUE 110 7489
941 173 All Levels 16 Web Development
2533 789324 "Complete Web Development Course Building 5 Websites
" creating-5-web-development-projects-in-html5-and-css3 NULL 2001
1492 174 118 NULL 25
2500 1000010 Professional Ruby on Rails Developer with Rails 5 pro-
ruby-on-rails-rails5 TRUE 200 3179 214 174 All Levels 2
1 Web Development
2944 1115348 "Full stack Universal React with Redux Node js and MongoDB
" full-stack-universal-react-with-redux-express-and-mongodb NULL
195 1121 174 62 NULL 5
374 988804 7 Steps- Your Introduction to Trading 7-steps-your-introdu-
ction-to-trading TRUE 25 1076 22 176 Beginner Lev
el 5 Business Finance
3364 260756 The Essential Web Developer Course - Build 12 Websites the-
essential-web-developer-course TRUE 200 9355 390 177 All
Levels 17 Web Development
2889 418386 "Create a REAL Social Network like Facebook in PHP MyS
QL" make-a-social-media-website NULL 45 1002 178 174N
ULL 17
2439 1022582 The Ultimate Vue JS 2 Developers Course vuejs-2-essentials T
```

## Query 10: Partition the courses by subject?

Applications Places System 8:32 PM

training@localhost:~

File Edit View Terminal Tabs Help

```
[training@localhost ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_202007032029_956922407.txt
hive> create table subjectwise(sno int,id int,coursetitle string,site string,ispaid string,price int,subscribers int,reviews int,lectures int,levels string,duration int)
> partitioned by (subject string)
> row format delimited
> fields terminated by ',';
OK
Time taken: 2.202 seconds
hive> insert into subjectwise partition(subject) select sno,coursetitle,ispaid,price,subscribers,reviews,lectures,levels,duration from online;
```

[Hive] [Screensh... training@l... [HDFS:/us...]

Applications Places System 4:56 AM

training@localhost:~

File Edit View Terminal Tabs Help

```
hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode=true;
hive> insert overwrite table subjectwise partition(subject) select * from online;
Total MapReduce jobs = 2
Launching Job 1 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_202007040437_0002, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job_202007040437_0002
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job_202007040437_0002
2020-07-04 04:52:51,695 Stage-1 map = 0%, reduce = 0%
2020-07-04 04:52:52,703 Stage-1 map = 100%, reduce = 0%
2020-07-04 04:52:53,723 Stage-1 map = 100%, reduce = 100%
Ended Job = job_202007040437_0002
Ended Job = -794022535, job is filtered out (removed at runtime).
Moving data to: hdfs://localhost/tmp/hive-training/hive_2020-07-04_04-52-46_956_4341013175116580121/-ext-10000
Loading data to table default.subjectwise partition (subject=null)
Loading partition {subject=1}
Loading partition {subject=10}
Loading partition {subject=11}
```

training@localhost:~

