

Assignment 2: MapReduce Program Application

COS20028 – Big data Architecture and Applications

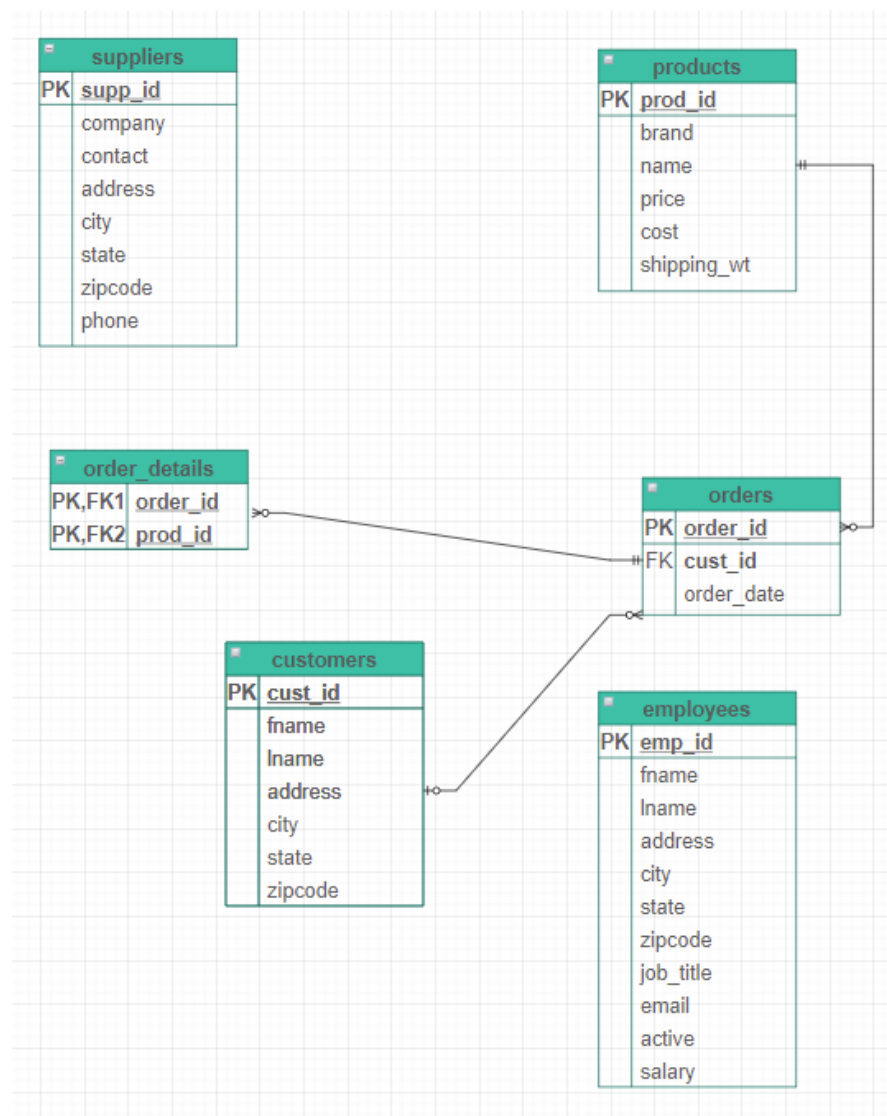
Semester 2, 2021

Name: SYED OMAIR MAQDOOM MOHIUDDIN

Student ID: 102863768

Date: 30/10/2021

Question 1 – ERD



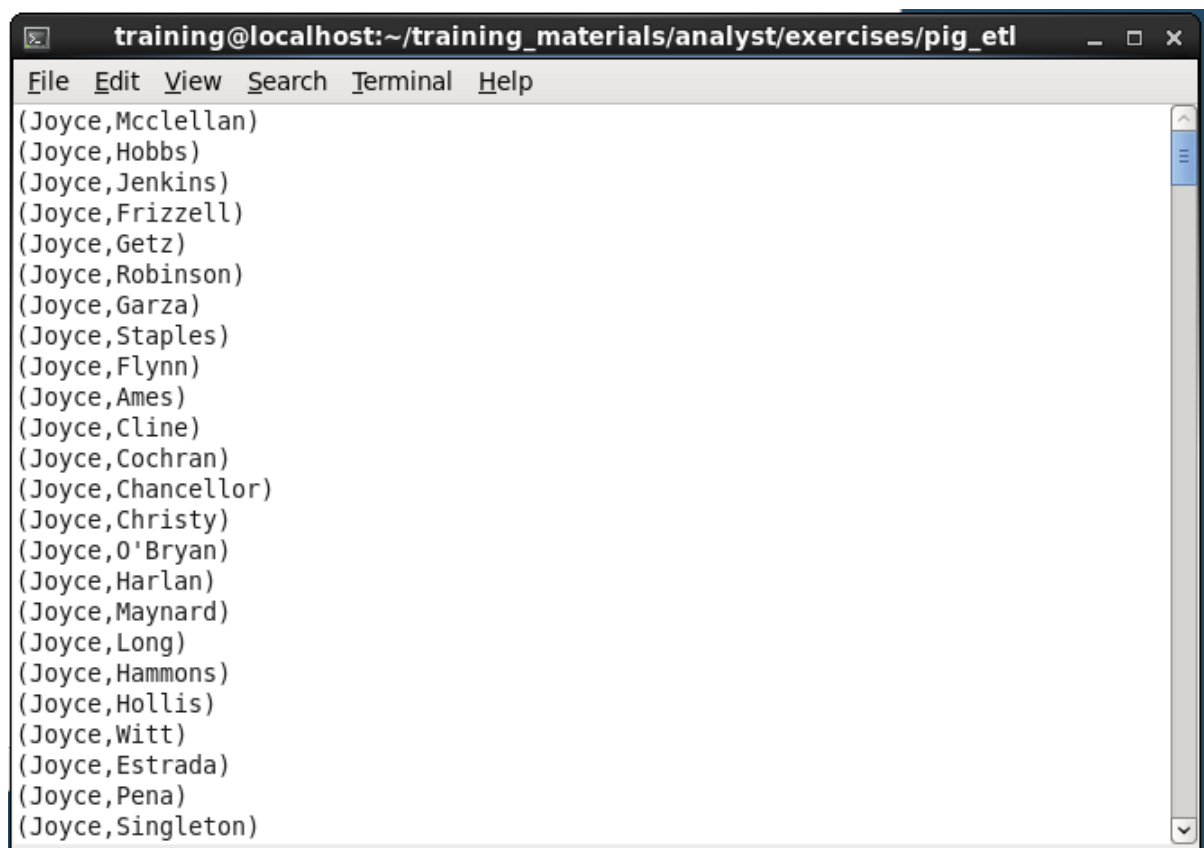
Question 2

a) Code:

```
2a.pig ✕
data = LOAD '/dualcore/customers' AS (cust_id:int, fname:chararray, lname:chararray, address:chararray, city:chararray, state:chararray, zipcode:int);
data2 = FOREACH data GENERATE fname AS fname:chararray, lname AS lname:chararray;
data3 = ORDER data2 BY fname ASC;
DUMP data3;
```

Result:

```
[training@localhost Assignment2]$ pig 2a.pig
2021-09-05 15:01:25,518 INFO org.apache.pig.Main: Apache Pig version 0.10.0-cdh4
.2.1 (rexpoted) compiled Apr 22 2013, 12:04:54
2021-09-05 15:01:25,519 INFO org.apache.pig.Main: Logging error messages to: /ho
me/training/training_materials/analyst/exercises/pig_etl/Assignment2/pig_1630868
485512.log
(Aaron,Ortega)
(Aaron,Parr)
(Aaron,Daniel)
(Aaron,Crawford)
(Aaron,Rios)
(Aaron,Bellows)
(Aaron,Darnell)
(Aaron,Hall)
(Aaron,Fairley)
(Aaron,Breckenridge)
(Aaron,Norman)
(Aaron,Matlock)
(Aaron,Lott)
(Aaron,Leyva)
(Aaron,Mercurio)
(Aaron,Gibbs)
```



A terminal window titled "training@localhost:~/training_materials/analyst/exercises/pig_etl" displays a list of 25 names, each in the format "(Joyce, Lastname)". The names are listed vertically, starting with "(Joyce, McClellan)" at the top and ending with "(Joyce, Singleton)" at the bottom. The terminal has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". A scrollbar is visible on the right side of the terminal window.

```
(Joyce,McClellan)
(Joyce,Hobbs)
(Joyce,Jenkins)
(Joyce,Frizzell)
(Joyce,Getz)
(Joyce,Robinson)
(Joyce,Garza)
(Joyce,Staples)
(Joyce,Flynn)
(Joyce,Ames)
(Joyce,Cline)
(Joyce,Cochran)
(Joyce,Chancellor)
(Joyce,Christy)
(Joyce,O'Bryan)
(Joyce,Harlan)
(Joyce,Maynard)
(Joyce,Long)
(Joyce,Hammons)
(Joyce,Hollis)
(Joyce,Witt)
(Joyce,Estrada)
(Joyce,Pena)
(Joyce,Singleton)
```

b) Code:

```
Question2b.pig
data = LOAD '/dualcore/customers' AS (cust_id:int, fname:chararray, lname:chararray, address:chararray, city:chararray, state:chararray, zipcode:int);
data1 = FILTER data BY city == 'Louisville';

data2 = FOREACH data1 GENERATE fname AS fname:chararray, lname AS lname:chararray;
STORE data2 INTO 'dualcore/Customers_Louisville';
```

Result:

```
[training@localhost Assignment2]$ pig Question2b.pig
2021-09-05 15:10:11,053 INFO org.apache.pig.Main: Apache Pig version 0.10.0-cdh4
.2.1 (rexported) compiled Apr 22 2013, 12:04:54
2021-09-05 15:10:11,054 INFO org.apache.pig.Main: Logging error messages to: /ho
me/training/training_materials/analyst/exercises/pig_etl/Assignment2/pig_1630869
011048.log
[training@localhost Assignment2]$
```

part-m-00000 (~/.training_materials/analyst/exercises/pig_etl/Assignment2/Customers_Louisville) - gedit

File Edit View Search Tools Documents Help

Open Save Undo Cut Copy Paste Print Run

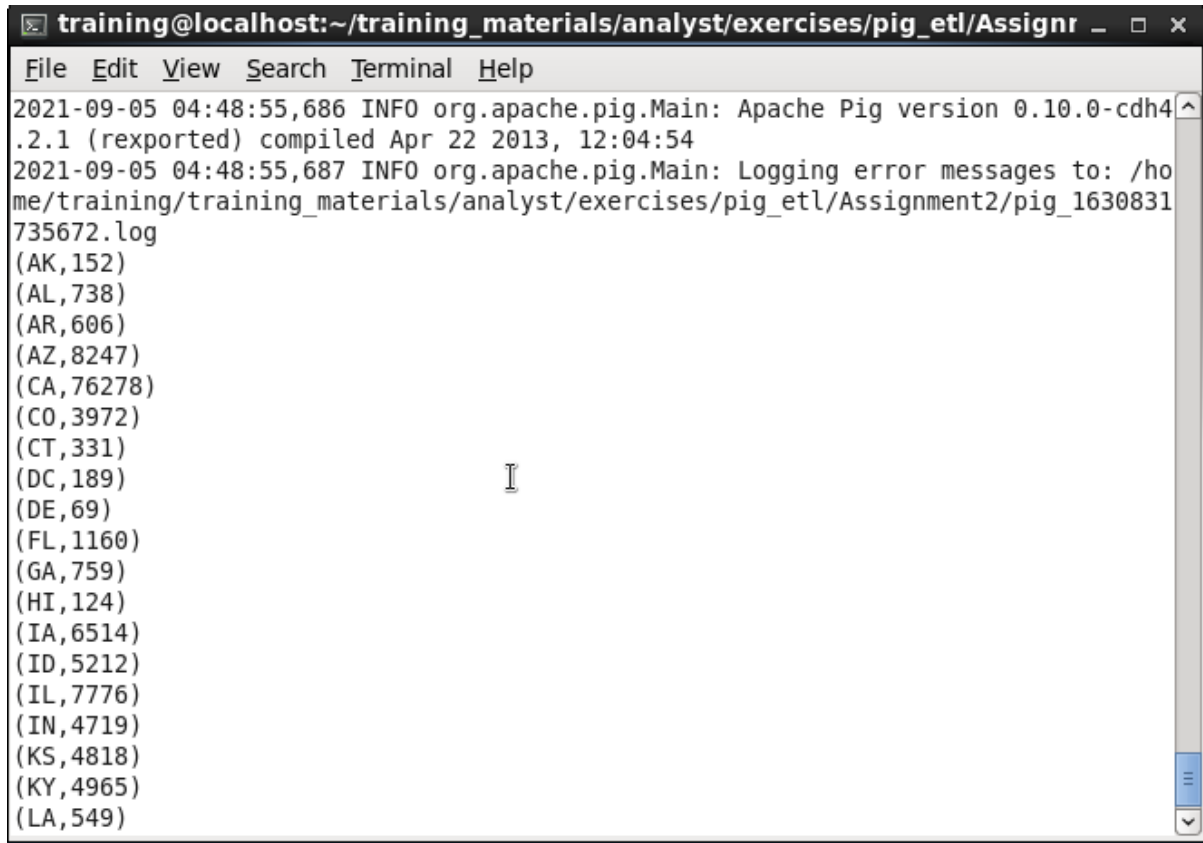
part-m-00000

Christie	Boudreaux
Federico	Wells
Jennifer	Cantu
Elizabeth	Roby
Darin	Dewitt
Mary	Brockway
Blaine	Javier
Shirley	Dagostino
Traci	Hurd
Suzanne	Simon
Leah	Younger
Joel	Hunt
Williams	Metz
Raymond	Ingram
Nancy	Pena
Debra	Miller
Elizabeth	O'Neal
Marsha	Bouldin
Angelo	Gilliland
Mark	Nolan
.....

c) Code:

```
data = LOAD '/dualcore/customers' AS (cust_id:int, fname:chararray, lname:chararray, address:chararray, city:chararray, state:chararray, zipcode:int);
data1 = GROUP data BY state;
data2 = FOREACH data1 GENERATE group AS state, COUNT(data) as CustomerCount;
DUMP data2;
```

Result:

A screenshot of a terminal window titled 'training@localhost:~/training_materials/analyst/exercises/pig_etl/Assignr'. The terminal shows the output of a Pig script. It starts with two INFO messages from org.apache.pig.Main: 'Apache Pig version 0.10.0-cdh4.2.1 (rexpoted) compiled Apr 22 2013, 12:04:54' and 'Logging error messages to: /home/training/training_materials/analyst/exercises/pig_etl/Assignment2/pig_1630831735672.log'. Following these, the output lists 15 states with their corresponding customer counts in parentheses: (AK,152), (AL,738), (AR,606), (AZ,8247), (CA,76278), (CO,3972), (CT,331), (DC,189), (DE,69), (FL,1160), (GA,759), (HI,124), (IA,6514), (ID,5212), (IL,7776), (IN,4719), (KS,4818), (KY,4965), and (LA,549).

```
training@localhost:~/training_materials/analyst/exercises/pig_etl/Assignr
File Edit View Search Terminal Help
2021-09-05 04:48:55,686 INFO org.apache.pig.Main: Apache Pig version 0.10.0-cdh4.2.1 (rexpoted) compiled Apr 22 2013, 12:04:54
2021-09-05 04:48:55,687 INFO org.apache.pig.Main: Logging error messages to: /home/training/training_materials/analyst/exercises/pig_etl/Assignment2/pig_1630831735672.log
(AK,152)
(AL,738)
(AR,606)
(AZ,8247)
(CA,76278)
(CO,3972)
(CT,331)
(DC,189)
(DE,69)
(FL,1160)
(GA,759)
(HI,124)
(IA,6514)
(ID,5212)
(IL,7776)
(IN,4719)
(KS,4818)
(KY,4965)
(LA,549)
```

d) Code:

```
2d.pig
data = LOAD '/dualcore/products' AS (prod_id:int, brand:chararray, name:chararray, price:int, cost:int, shipping_wt:chararray);

data1 = DISTINCT data;

data2 = GROUP data1 BY brand;

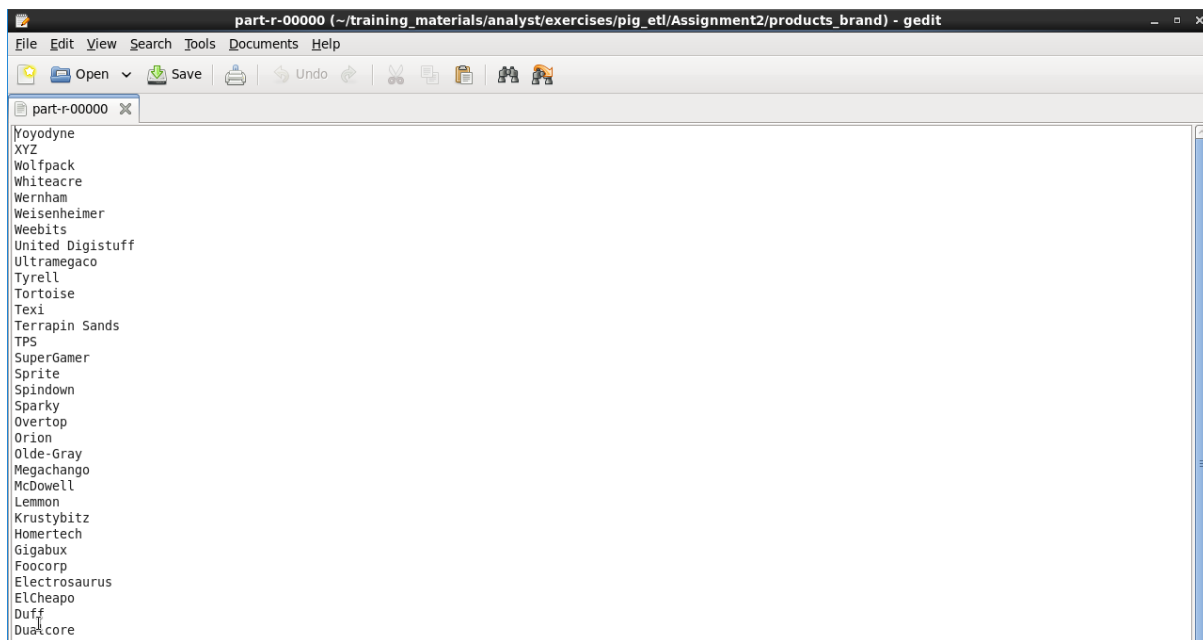
data3 = FOREACH data2 GENERATE group AS brand;

data4 = ORDER data3 BY brand DESC;

STORE data4 INTO '/dualcore/products_brand';
```

Result:

```
[training@localhost Assignment2]$ pig 2d.pig
2021-09-05 15:23:21,479 INFO org.apache.pig.Main: Apache Pig version 0.10.0-cdh4
.2.1 (rexported) compiled Apr 22 2013, 12:04:54
2021-09-05 15:23:21,480 INFO org.apache.pig.Main: Logging error messages to: /home/training/training_materials/analyst/exercises/pig_etl/Assignment2/pig_1630869
801474.log
[training@localhost Assignment2]$
```



The screenshot shows a gedit window titled "part-r-00000 (~/.training_materials/analyst/exercises/pig_etl/Assignment2/products_brand) - gedit". The window contains a list of product brands, each on a new line. The brands are: Toyodyne, XYZ, Wolfpack, Whiteacre, Wernham, Weisenheimer, Weebits, United Digistuff, Ultramegaco, Tyrell, Tortoise, Texi, Terrapin Sands, TPS, SuperGamer, Sprite, Spindown, Sparky, Overtop, Orion, Olde-Gray, Megachango, McDowell, Lemmon, Krustybitz, Homertech, Gigabux, Foocorp, Electrosaurus, ElCheapo, Duff, and Dualcore. The list is sorted in descending order of brand name.

Question 3

a) Code:

```
3a.pig X
data = LOAD '/dualcore/orders' AS (order_id:int, cust_id:int, order_date:chararray);
data1 = LOAD '/dualcore/order_details' AS (order_id:int, prod_id:int);
data3 = FILTER data BY order_date MATCHES '^2013-05-.*$';
data4 = FILTER data1 BY prod_id == 1274348;
data5 = JOIN data1 BY order_id LEFT OUTER, data4 BY order_id;
data6 = FOREACH data5 GENERATE data1::order_id AS order_id:int, data1::prod_id AS prod_id;
data7 = JOIN data3 BY order_id, data6 BY order_id;
data8 = FOREACH data7 GENERATE data3::order_id AS order_id:int, data6::prod_id AS prod_id:int;
data9 = DISTINCT data8;
data10 = GROUP data9 BY order_id;
data11 = FOREACH data10 GENERATE group, COUNT(data9.prod_id) AS Total:long;
data12 = GROUP data11 ALL;
data13 = FOREACH data12 GENERATE MAX(data11.Total);
DUMP data13;
```

Result:

```
[training@localhost bonus_02]$ cd /home/training/training_materials/analyst/exercises/pig_etl/Assignment2/
[training@localhost Assignment2]$ pig 3a.pig
2021-09-05 09:46:11,555 INFO org.apache.pig.Main: Apache Pig version 0.10.0-cdh4.2.1 (rexpoted) compiled Apr 22 2013, 12:04:54
2021-09-05 09:46:11,556 INFO org.apache.pig.Main: Logging error messages to: /home/training/training_materials/analyst/exercises/pig_etl/Assignment2/pig_1630849571549.log
(8)
[training@localhost Assignment2]$
```

b) Code:

```
loyalty_program.pig
-- load the data sets
orders = LOAD '/dualcore/orders' AS (order_id:int,
    cust_id:int,
    order_dtm:chararray);

details = LOAD '/dualcore/order_details' AS (order_id:int,
    prod_id:int);

products = LOAD '/dualcore/products' AS (prod_id:int,
    brand:chararray,
    name:chararray,
    price:int,
    cost:int,
    shipping_wt:int);

orders_2012 = FILTER orders BY order_dtm MATCHES '^2012.*$';

grouping = GROUP orders BY cust_id;
orders_count = FOREACH grouping GENERATE group, COUNT(orders) AS num_orders;
orders_greater_6 = FILTER orders_count BY num_orders > 6;

orders_customer = JOIN orders_greater_6 BY group, orders_2012 BY cust_id;

order_details_customer = JOIN orders_customer BY orders_2012::order_id, details BY order_id;
order_product_customer = JOIN order_details_customer BY details::prod_id, products BY prod_id;

price_customer = FOREACH order_product_customer GENERATE orders_greater_6::group AS cust_id, products::price AS price;

customer_total_group = GROUP price_customer BY cust_id;

customer_total = FOREACH customer_total_group GENERATE group, SUM(price_customer.price)/100 AS Total;

SPLIT customer_total INTO
Platinum IF Total > 10000,
Gold IF Total >= 7000 AND Total < 10000,
Silver IF Total >= 5000 AND Total < 7000,
Unknow IF Total < 5000;

STORE Platinum INTO 'dualcore/loyalty/Platinum_Records';
STORE Gold INTO 'dualcore/loyalty/Gold_Records';
STORE Silver INTO 'dualcore/loyalty/Silver_Records';
STORE Unknow INTO 'dualcore/loyalty/Unknow_Records';
```

```
3b.pig
platinum = LOAD 'dualcore/loyalty/Platinum_Records/part-r-00000' AS (cust_id:int, price:long);
gold = LOAD 'dualcore/loyalty/Gold_Records/part-r-00000' AS (cust_id:int, price:long);
silver = LOAD 'dualcore/loyalty/Silver_Records/part-r-00000' AS (cust_id:int, price:long);
unknow = LOAD 'dualcore/loyalty/Unknow_Records/part-r-00000' AS (cust_id:int, price:long);

data = GROUP platinum ALL;
data1 = FOREACH data GENERATE 'Platinum_Count' AS groups:chararray, COUNT(platinum.cust_id) AS Total_Platinum:long;
DUMP data1;

data2 = GROUP gold ALL;
data3 = FOREACH data2 GENERATE 'Gold_Count' AS groups:chararray, COUNT(gold.cust_id) AS Total_Gold:long;
DUMP data3;

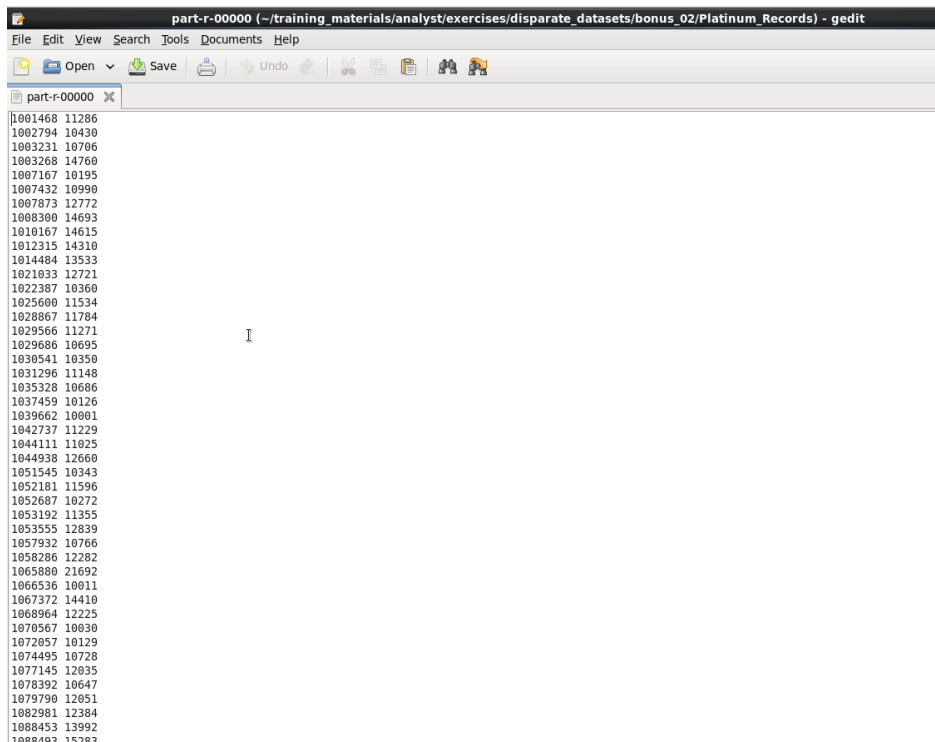
data4 = GROUP silver ALL;
data5 = FOREACH data4 GENERATE 'Silver_Count' AS groups:chararray, COUNT(silver.cust_id) AS Total_Silver:long;
DUMP data5;

data6 = GROUP unknow ALL;
data7 = FOREACH data6 GENERATE 'Unknown' AS groups:chararray, COUNT(unknow.cust_id) AS Total_Unknow:long;
DUMP data7;
```

Result:

```
[training@localhost bonus_02]$ pig 3b.pig
2021-09-06 16:44:03,474 INFO org.apache.pig.Main: Apache Pig version 0.10.0-cdh4
.2.1 (rexpoted) compiled Apr 22 2013, 12:04:54
2021-09-06 16:44:03,475 INFO org.apache.pig.Main: Logging error messages to: /home/training/training_materials/analyst/exercises/disparate_datasets/bonus_02/pig
_1630961043469.log
(Platinum_Count,109)
(Gold_Count,156)
(Silver_Count,281)
(Unknown,110497)
[training@localhost bonus_02]$
```

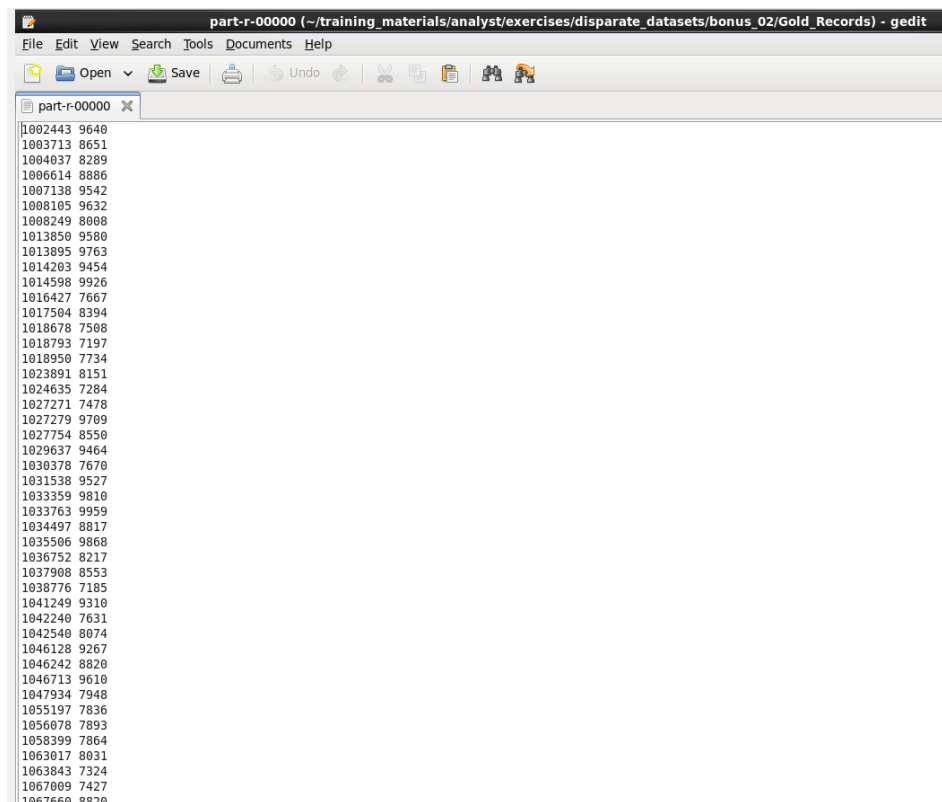

Platinum:



The screenshot shows a gedit window titled "part-r-00000 (~/.training_materials/analyst/exercises/disparate_datasets/bonus_02/Platinum_Records) - gedit". The window contains a list of numbers, each with a tab number in brackets at the start. The numbers are arranged in a single column and appear to be a sequence of values, possibly representing a dataset or a list of records. The window has a standard menu bar (File, Edit, View, Search, Tools, Documents, Help) and a toolbar with icons for Open, Save, Undo, and other editing functions.

```
[1001468 11286
1002794 10430
1003231 10706
1003268 14760
1007167 10195
1007432 10990
1007873 12772
1008300 14693
1010167 14615
1012315 14310
1014484 13533
1021033 12721
1022387 10360
1025600 11534
1028867 11784
1029566 11271
1029686 10695
1030541 10350
1031296 11148
1035328 10686
1037459 10126
1039662 10001
1042737 11229
1044111 11025
1044938 12660
1051545 10343
1052181 11596
1052687 10272
1053192 11355
1053555 12839
1057932 10766
1058286 12282
1065880 21692
1066536 10011
1067372 14410
1068964 12225
1070567 10030
1072057 10129
1074495 10728
1077145 12035
1078392 10647
1079790 12051
1082981 12384
1088453 13992
1088493 15283
```

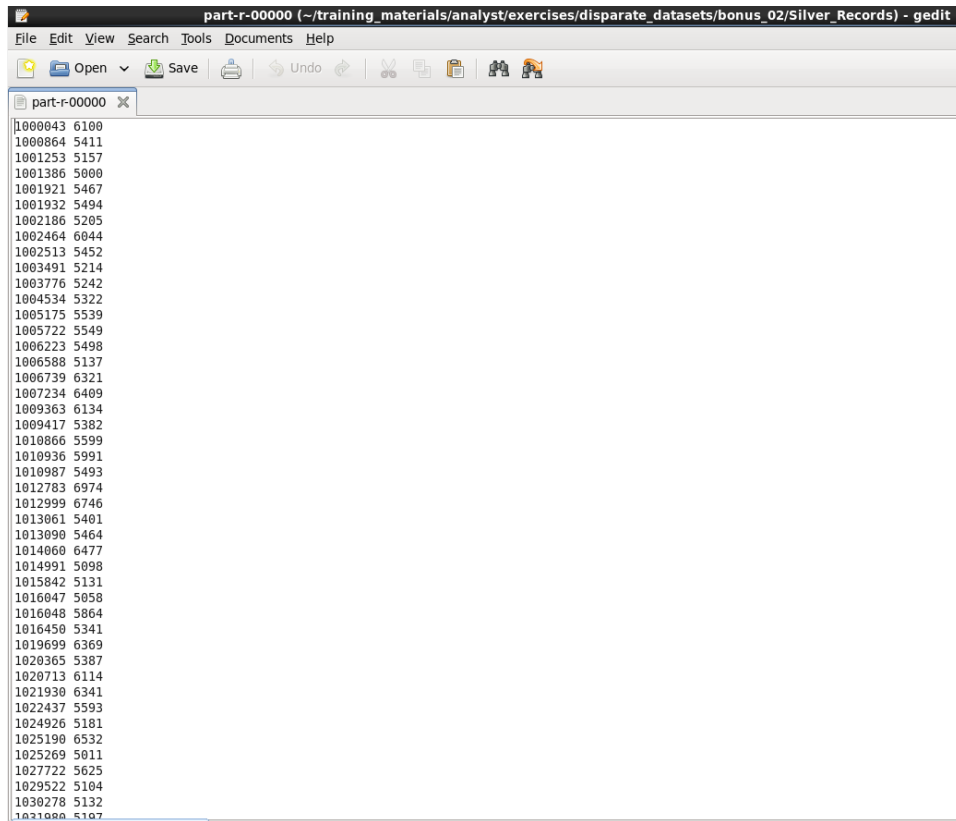
Gold:



The screenshot shows a gedit window titled "part-r-00000 (~/.training_materials/analyst/exercises/disparate_datasets/bonus_02/Gold_Records) - gedit". The window contains a list of numbers, each with a tab number in brackets at the start. The numbers are arranged in a single column and appear to be a sequence of values, possibly representing a dataset or a list of records. The window has a standard menu bar (File, Edit, View, Search, Tools, Documents, Help) and a toolbar with icons for Open, Save, Undo, and other editing functions.

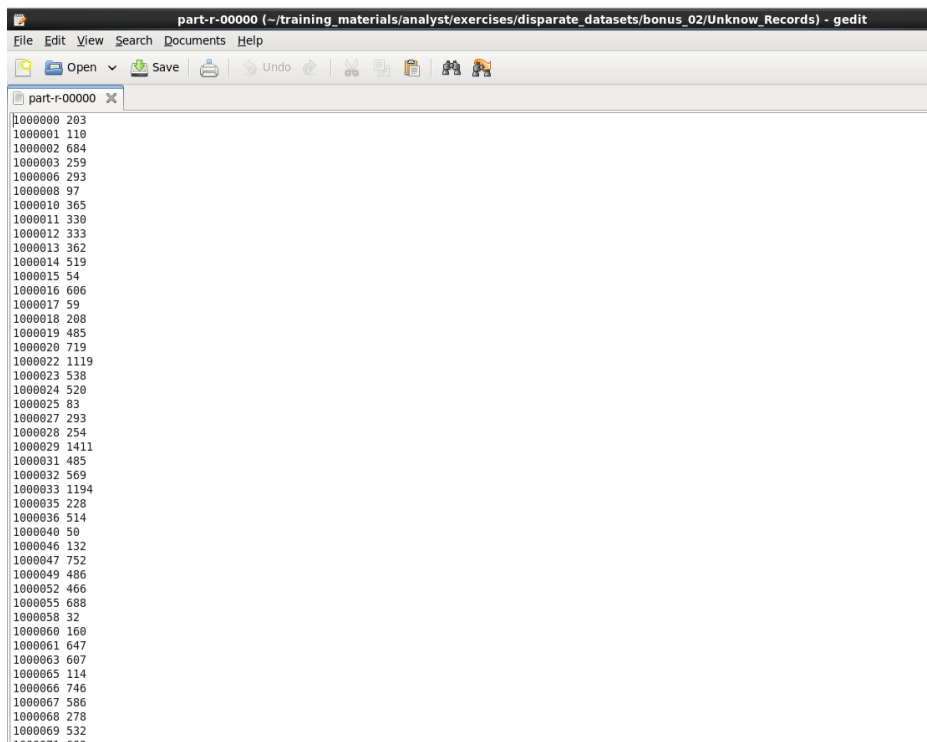
```
[1002443 9640
1003713 8651
1004037 8289
1006614 8886
1007138 9542
1008105 9632
1008249 8008
1013850 9580
1013895 9763
1014203 9454
1014598 9926
1016427 7667
1017504 8394
1018678 7508
1018793 7197
1018950 7734
1023891 8151
1024635 7284
1027271 7478
1027279 9709
1027754 8550
1029637 9464
1030378 7670
1031538 9527
1033359 9810
1033763 9959
1034497 8817
1035506 9868
1036752 8217
1037908 8553
1038776 7185
1041249 9310
1042240 7631
1042540 8074
1046128 9267
1046242 8820
1046713 9610
1047934 7948
1055197 7836
1056078 7893
1058399 7864
1063017 8031
1063843 7324
1067009 7427
1067660 8820
```

Silver:



The screenshot shows a gedit window titled "part-r-00000 (~/.training_materials/analyst/exercises/disparate_datasets/bonus_02/Silver_Records) - gedit". The window contains a list of numbers, each preceded by a tab character. The numbers are: 6100, 5411, 5157, 5000, 5467, 5494, 5205, 6044, 5452, 5214, 5242, 5322, 5539, 5549, 5498, 5137, 6321, 6409, 6134, 5382, 5599, 5991, 5493, 6974, 6746, 5401, 5464, 6477, 5098, 5131, 5058, 5864, 5341, 6369, 5387, 6114, 6341, 5593, 5181, 6532, 5011, 5625, 5104, 5132, 5197.

Unknown:

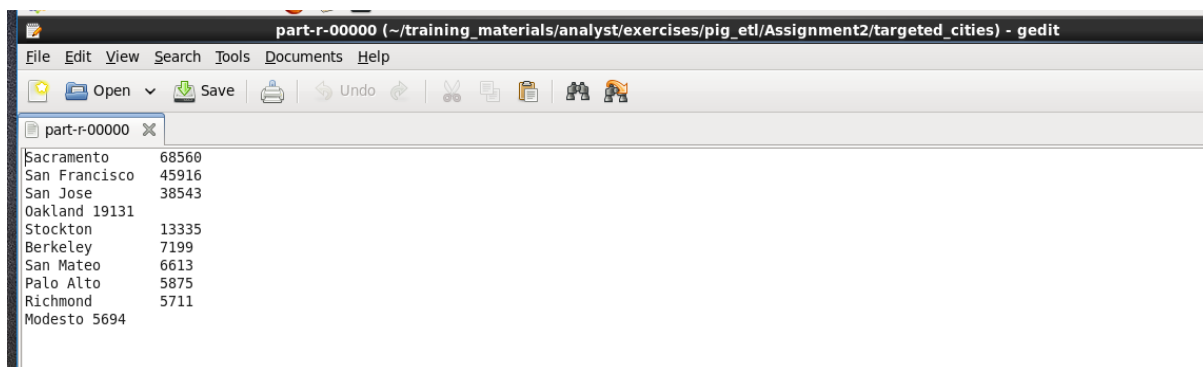


The screenshot shows a gedit window titled "part-r-00000 (~/.training_materials/analyst/exercises/disparate_datasets/bonus_02/Unknow_Records) - gedit". The window contains a list of numbers, each preceded by a tab character. The numbers are: 203, 110, 684, 259, 293, 97, 365, 330, 333, 362, 519, 54, 606, 59, 208, 485, 719, 1119, 538, 520, 83, 293, 254, 1411, 485, 569, 1194, 228, 514, 50, 132, 752, 486, 466, 688, 32, 160, 647, 607, 114, 746, 586, 278, 532, 505.

c) Code:

```
top_ten_cities_customers.pig X
data = LOAD '/dualcore/customers' AS (cust_id:int, fname:chararray, lname:chararray, address:chararray, city:chararray, state:chararray, zipcode:int);
data1 = LOAD '/dualcore/orders' AS (order_id:int, cust_id:int, order_date:chararray);
data2 = DISTINCT data1;
data3 = JOIN data BY cust_id, data2 BY cust_id;
data4 = GROUP data3 BY data::city;
data5 = FOREACH data4 GENERATE group, COUNT(data3) AS Total:long;
data6 = ORDER data5 BY Total DESC;
data7 = LIMIT data6 10;
STORE data7 INTO 'dualcore/targeted_cities';
```

Result:



```
part-r-00000 X
Sacramento      68560
San Francisco   45916
San Jose        38543
Oakland 19131
Stockton        13335
Berkeley        7199
San Mateo       6613
Palo Alto       5875
Richmond        5711
Modesto 5694
```

Question 4

a) Code:

Query Results: Unsaved Query

DOWNLOADS
[Download as CSV](#)
[Download as XLS](#)
[Save](#)
MR JOB (1)
[job_202109020415_0361](#)

Did you know? You can click on a row to select a column you want to jump to.

ResultsQueryLogColumns

```
SELECT brand,MIN(name) AS Name, COUNT(name) AS Product_Count FROM products WHERE name LIKE '%USB%' GROUP BY brand
```

Query Results: Unsaved Query

DOWNLOADS
[Download as CSV](#)
[Download as XLS](#)
[Save](#)
MR JOB (1)
[job_202109020415_0495](#)

Did you know? You can click on a row to select a column you want to jump to.

ResultsQueryLogColumns

```
SELECT COUNT(DISTINCT brand) AS Brand_Count FROM products WHERE name LIKE '%USB%'
```

Result :

Query Results: Unsaved Query

DOWNLOADS
[Download as CSV](#)
[Download as XLS](#)
[Save](#)
MR JOB (1)
[job_202109020415_0361](#)

Did you know? You can click on a row to select a column you want to jump to.

ResultsQueryLogColumns

	brand	name	product_count
0	ACME	16 GB USB Flash Drive	5
1	BDT	USB 3.0 Front Panel - 2 Port	2
2	Bigdeal	Quad USB Power Adapter	2
3	Bitbucket	16 GB USB Flash Drive	5
4	Byteweasel	16 GB USB Flash Drive	3
5	Chestnut	USB 3.0 4-Port Ultra Mini Hub	3
6	DevNull	500 GB Portable External USB 2.0 Disk	4
7	Dorx	16 GB USB Flash Drive	6
8	Dualcore	750 GB Portable External USB 3.0 Disk	10
9	Duff	Wireless N USB Adapter	1
10	ElCheapo	16 GB USB Flash Drive	9
11	Electrosaurus	2.5" External USB 3.0 HDD Case	1
12	Homertech	16 GB USB Flash Drive	5
13	Krustybitz	32 GB USB Flash Drive	3
14	Lemmon	32 GB USB Flash Drive	2
15	McDowell	USB 2.0 4-Port Mini Hub	3
16	Megachango	32 GB USB Flash Drive	6
17	Orion	16 GB USB Flash Drive (Blue)	8
18	Overtop	USB 3.0 Front Panel - 4 Port	2
19	Spindown	16 GB USB Flash Drive	3

Query Results: Unsaved Query

<div>DOWNLOADS</div> <div>Download as CSV</div> <div>Download as XLS</div> <div>Save</div> <div>MR JOB (1)</div> <div>job_202109020415_0495</div>	Results	Query	Log	Columns
	brand_count			
	0	29		

b) Code:

Query Results: Unsaved Query

<div>DOWNLOADS</div> <div>Download as CSV</div> <div>Download as XLS</div> <div>Save</div> <div>MR JOBS (2)</div> <div>job_202109020415_0366</div> <div>job_202109020415_0367</div>	Results	Query	Log	Columns
	<pre>SELECT COUNT(c.cust_id) AS Count_of_Orders,MIN(fname) AS Fname,MIN(lname) AS Lname FROM customers c JOIN orders o ON (c.cust_id = o.cust_id) WHERE fname LIKE 'Diana' AND lname LIKE 'Lamb'</pre>			
	Did you know? You can click on a row to select a column you want to jump to.			

Result:

Query Results: Unsaved Query

<div>DOWNLOADS</div> <div>Download as CSV</div> <div>Download as XLS</div> <div>Save</div> <div>MR JOBS (2)</div> <div>job_202109020415_0366</div> <div>job_202109020415_0367</div>	Results	Query	Log	Columns
	count_of_orders	fname	lname	
	0	10	Diana	Lamb

c) Code:

Query Results: [Unsaved Query](#)

DOWNLOADS
[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOBS (2)
[job_202109020415_0478](#)
[job_202109020415_0479](#)

ResultsQueryLogColumns

SELECT brand, COUNT(prod_id) AS Product_Count FROM products
GROUP BY brand ORDER BY brand DESC

Result:

Query Results: [Unsaved Query](#)

DOWNLOADS
[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOBS (2)
[job_202109020415_0478](#)
[job_202109020415_0479](#)

ResultsQueryLogColumns

	brand	product_count
0	Yoyodyne	4
1	XYZ	46
2	Wolfpack	20
3	Whiteacre	6
4	Wernham	20
5	Weisenheimer	12
6	Weebits	1
7	United Digistuff	27
8	Ultramegaco	39
9	Tyrell	51
10	Tortoise	8
11	Texi	7
12	Terrapin Sands	12
13	TPS	46
14	SuperGee	7

d) Code:

Query Results: [Unsaved Query](#)

DOWNLOADS

[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOB (1)
[job_202109020415_0358](#)

ResultsQueryLogColumns

SELECT brand, MIN(name) AS name, MIN(price) AS price, MIN(cost) AS Lowest_Cost FROM products
GROUP BY brand

Did you know? You can click on a row to select a column you want to jump to.

Result:

Query Results: [Unsaved Query](#)

DOWNLOADS

[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOB (1)
[job_202109020415_0358](#)

ResultsQueryLogColumns

	brand	name	price	lowest_cost
0	ACME	1.0 TB SATA3 Disk (10K RPM)	279	153
1	ARCAM	42" Wide Screen TV	30999	25802
2	Argo	22" Wide Screen LCD Monitor	17719	17482
3	Artie	27" Wide Screen TV	13269	9564
4	BDT	1GB DDR2 800 Desktop RAM	389	356
5	Bigdeal	1/4 in. Standard Phone Female to Female Adapter	59	28
6	Bitbucket	1.5 TB SATA3 Disk	2029	1436
7	Bitmonkey	All-in-one Desktop (24 in. display)	43459	39281
8	BuckLogix	1GB DDR 400 Desktop RAM	499	472
9	Bytefortress	Office Suite (Home Edition)	23259	14413
10	Byteweasel	16 GB USB Flash Drive	1279	1064
11	Chatter Audio	MP3 Player (16 GB internal memory)	4099	3520
12	Chestnut	1/4 in. Standard Phone Male to 3.5mm Mini Female Adapter	49	38
13	DevNull	1/4 in. Standard Phone Female to Female Adapter	39	17
14	Dorx	1 TB NAS Server	1929	1790
15	Dualcore	1.5 TB SATA3 Disk	119	119
16	Duff	1GB DDR2-667 (PC2-5300) Desktop Memory Kit (2x512MB)	19	7
17	ElCheapo	1/4 in. Standard Phone Male to 3-Pin XLR Female Adapter	279	124
18	Electrosaurus	1024MB PCI-X Graphics Card with HDMI output (OEM)	409	384
19	Foocorp	DSL Modem	2989	2754
20	Gigabux	1/4 in. Standard Phone Female to Female Adapter	399	198

Did you know? You can click on a row to select a column you want to jump to.

e) Code:

DOWNLOADS

[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOBS (3)
[job_202109020415_0328](#)
[job_202109020415_0329](#)
[job_202109020415_0330](#)

ResultsQueryLogColumns

SELECT products.prod_id, COUNT(products.prod_id) AS TotalSale, name, brand
FROM products JOIN order_details ON
(products.prod_id = order_details.prod_id) GROUP BY
products.prod_id,name,brand ORDER BY TotalSale DESC LIMIT 5

Result:

Query Results: **Unsaved Query**

DOWNLOADS

[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOBS (3)
[job_202109020415_0328](#)
[job_202109020415_0329](#)
[job_202109020415_0330](#)

ResultsQueryLogColumns

	prod_id	totalsale	name	brand
0	1274348	119801	Tablet PC (10 in. display, 64 GB)	Byteweasel
1	1274021	29053	8GB DDR3-1600 (PC3-12800) Dual Channel Desktop Memory Kit (2x4GB)	Dualcore
2	1273970	14566	F Jack Male-to-Male Cable (12 in.)	Megachango
3	1273853	14525	Extension cord (18 in.)	Dualcore
4	1273662	14282	USB Power Adapter (single)	Tyrell

f) Code:

Query Results: **Unsaved Query**

DOWNLOADS

[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOBS (3)
[job_202109020415_0386](#)
[job_202109020415_0387](#)
[job_202109020415_0388](#)

ResultsQueryLogColumns

SELECT PRINTF("%.2f",SUM(price)/100) AS Total_Revenue FROM products p JOIN order_details od ON (p.prod_id = od.prod_id) JOIN orders o ON (od.order_id = o.order_id) WHERE order_date LIKE '2013-05-%'

Result:

Query Results: **Unsaved Query**

DOWNLOADS

[Download as CSV](#)
[Download as XLS](#)
[Save](#)

MR JOBS (3)
[job_202109020415_0386](#)
[job_202109020415_0387](#)
[job_202109020415_0388](#)

ResultsQueryLogColumns

	total_revenue
0	\$32068207.24

g) Code:

Query Results: [Unsaved Query](#)

DOWNLOADS

Download as CSV

Download as XLS

Save

MR JOBS (3)

job_202109020415_0389

job_202109020415_0390

job_202109020415_0391

Did you know? You can click on a row to select a column you want to jump to.

ResultsQueryLogColumns

```
SELECT DAY(order_date) AS Order_Day, PRINTF("$.2f",SUM(price)/100) AS Total_Revenue FROM products p JOIN order_details od ON (p.prod_id = od.prod_id) JOIN orders o ON (od.order_id = o.order_id) WHERE order_date LIKE '2013-05-%' GROUP BY DAY(order_date)
```

Result:

Query Results: [Unsaved Query](#)

DOWNLOADS

Download as CSV

Download as XLS

Save

MR JOBS (3)

job_202109020415_0389

job_202109020415_0390

job_202109020415_0391

Did you know? You can click on a row to select a column you want to jump to.

ResultsQueryLogColumns

	order_day	total_revenue
0	1	\$383618.19
1	2	\$1064939.68
2	3	\$1100699.17
3	4	\$1028571.67
4	5	\$1066038.15
5	6	\$1075537.38
6	7	\$1033746.80
7	8	\$1073944.64
8	9	\$1100788.81
9	10	\$1062743.60
10	11	\$1059573.17
11	12	\$1065223.57
12	13	\$1072153.66
13	14	\$1080752.28
14	15	\$1081561.73
15	16	\$1033992.27
16	17	\$1087067.38
17	18	\$1045830.21
18	19	\$1058442.66
19	20	\$1035860.45
20	21	\$1050509.84

Question 5

(codes and table contents)

Query Editor

```
1 CREATE TABLE elcheapo_products AS SELECT * FROM products
2 WHERE brand LIKE 'elCheapo';
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
```

Query Results: Unsaved Query

Query Results: Unsaved Query

DOWNLOADS

- Download as CSV
- Download as XLS
- Save

MR JOBS

No Hadoop jobs were launched in running this query.

Did you know? You can click on a row to select a column you want to jump to.

Results Query Log Columns

SELECT * FROM elcheapo_products

Query Results: Unsaved Query

Query Results: Unsaved Query

Save

MR JOBS

No Hadoop jobs were launched in running this query.

Did you know? You can click on a row to select a column you want to jump to.

Results Query Log Columns

DROP TABLE elcheapo_products

Query Results: Unsaved Query

DOWNLOADS

Download as CSV

Download as XLS

Save

MR JOBS

No Hadoop jobs were launched in running this query.

Results

Query

Log

Columns

↑

tab_name

0

customers

1

elcheapo_products

2

order_details

3

orders

4

products

Hue

Query Editor

My Queries

Saved Queries

History

Tables

Settings

training

Query Results: Unsaved Query

DOWNLOADS

Download as CSV

Download as XLS

Save

MR JOBS

No Hadoop jobs were launched in running this query.

Results

Query

Log

Columns

↑

prod_id

brand

name

price

cost

shipping_wt

0

1273653

EiCheapo

Composite AV Cable (24 in.)

2489

982

1

1

1273654

EiCheapo

Composite AV Cable (24 in.)

2189

1669

1

2

1273656

EiCheapo

Composite AV Cable (36 in.)

2489

1605

1

3

1273665

EiCheapo

Dual USB Power Adapter

2109

898

1

4

1273667

EiCheapo

Dual USB Power Adapter (auto)

2969

2521

1

5

1273671

EiCheapo

High Speed HDMI Cable (24 in.)

2009

1184

1

6

1273677

EiCheapo

High Speed HDMI Cable (60 in.)

4749

4053

1

7

1273686

EiCheapo

Headphone Adapter (1/4 in. to 1/8 in.)

779

736

1

8

1273688

EiCheapo

Mobile Bluetooth Keyboard and Stand for Tablets

7009

3975

3

9

1273734

EiCheapo

Auto Charger

3989

2164

2

10

1273749

EiCheapo

Batteries (AAA, 4 pack)

749

664

2

11

1273783

EiCheapo

Keyboard (basic PC101)

3079

2372

3

12

1273801

EiCheapo

Wireless laser mouse (three button, with wheel)

3949

3348

1

13

1273839

EiCheapo

Surge Protector (8-Outlet)

2499

1152

2

14

1273874

EiCheapo

Extension cord (10 ft., outdoor)

2229

2707

2

Did you know? You can click on a row to select a column you want to jump to.

Query Results: Unsaved Query

DOWNLOADS

Download as CSV

Download as XLS

Save

MR JOBS

No Hadoop jobs were launched in running this query.

Results

Query

Log

Columns

↑

tab_name

0

customers

1

order_details

2

orders

3

products

Did you know? You can click on a row to select a column you want to jump to.