

Data Hadoop & Spark Training - ACADGILD Assignment 4.2

27-Nov-17

BIG DATA – PIG ASSIGNMENT

BY – SAHIL KHURANA

Problem Statement

Create a sample dataset and implement the below Pig commands on the same dataset.

- 1) Concat**
- 2) Tokenize**
- 3) Sum**
- 4) Min**
- 5) Max**
- 6) Limit**
- 7) Store**
- 8) Distinct**
- 9) Flatten**
- 10) IsEmpty**

As not dataset is given by Acadgild. So, I will be using the following dataset.

Dataset -

File Name “customer_data.csv” & “order_data.csv”
customer_data.csv

1	Tracy	Stokley	990000112345	1/1/2001	Tracy Stokley1@gmail.com
2	Ken	Grigg	990000112346	1/1/2001	Ken Grigg2@gmail.com
3	Jacques	Sandoz	990000112347	1/1/2001	Jacques Sandoz3@gmail.com
4	Rickie	Ciampa	990000112348	1/1/2001	Rickie Ciampa4@gmail.com
5	Andreas	Vicini	990000112349	1/1/2001	Andreas Vicini5@gmail.com
6	Kurt	Gariepy	990000112350	1/1/2001	Kurt Gariepy6@gmail.com
7	Hoyt	Ruddick		1/1/2001	Hoyt Ruddick7@gmail.com
8	Moshe	Bloomberg	990000112352	1/1/2001	Moshe Bloomberg8@gmail.com
9	Steve	Lindamood	990000112353	1/1/2001	Steve Lindamood9@gmail.com
10	Sandy	Yearta	990000112354	1/1/2001	Sandy Yearta10@gmail.com
11	Tyler	Alves	990000112355	1/1/2001	
12	Kerry	Prindle	990000112356	1/1/2001	Kerry Prindle12@gmail.com
13	Lucio	Spiro	990000112357	1/1/2001	Lucio Spiro13@gmail.com
14	Shannon	Dufour	990000112358	1/1/2001	Shannon Dufour14@gmail.com
15	Leonardo	Negrete	990000112359	1/1/2001	
16	Kim	Mccurley	990000112360	1/1/2001	Kim Mccurley16@gmail.com
17	Wilson	Demelo	990000112361	1/1/2001	Wilson Demelo17@gmail.com
18	Chuck	Deleo	990000112362	1/1/2001	Chuck Deleo18@gmail.com
19	Seth	Rayford	990000112363	1/1/2001	Seth Rayford19@gmail.com
20	Conrad	Melville	990000112364	1/1/2001	Conrad Melville20@gmail.com
21	Sebastian	Dorton	990000112365	1/1/2001	Sebastian Dorton21@gmail.com
22	Dana	Toms	990000112366	1/1/2001	
23	Wyatt	Strope	990000112367	1/1/2001	Wyatt Strope23@gmail.com
24	Cletus	Exum	990000112368	1/1/2001	Cletus Exum24@gmail.com
25	Josh	Heffley	990000112369	1/1/2001	Josh Heffley25@gmail.com
26	Lesley	Mccreary	990000112370	1/1/2001	Lesley Mccreary26@gmail.com
27	Chad	Lepley	990000112371	1/1/2001	Chad Lepley27@gmail.com
28	Freddy	Como	990000112372	1/1/2001	Freddy Como28@gmail.com
29	Hugh	Chalk	990000112373	1/1/2001	Hugh Chalk29@gmail.com
30	Marshall	Skipworth	990000112374	1/1/2001	Marshall Skipworth30@gmail.com
31	Jim	Jonason	990000112375	1/1/2001	Jim Jonason31@gmail.com
32	Ed	Gambrel	990000112376	1/1/2001	Ed Gambrel32@gmail.com
33	Stacey	Tostado	990000112377	1/1/2001	Stacey Tostado33@gmail.com
34	Everett	Mooneyham	990000112378	1/1/2001	Everett Mooneyham34@gmail.com

35	Greg	Mcvey	990000112379	1/1/2001	Greg Mcvey35@gmail.com
36	Roberto	Collier	990000112380	1/1/2001	Roberto Collier36@gmail.com
37	Desmond	Carasco	990000112381	1/1/2001	Desmond Carasco37@gmail.com
38	Dewey	Marler	990000112382	1/1/2001	Dewey Marler38@gmail.com
39	Ivan	Buss	990000112383	1/1/2001	Ivan Buss39@gmail.com
40	Sonny	Cutter	990000112384	1/1/2001	Sonny Cutter40@gmail.com
41	Eldridge	Wilcox	990000112385	1/1/2001	Eldridge Wilcox41@gmail.com
42	Christian	Anders	990000112386	1/1/2001	Christian Anders42@gmail.com
43	Leif	Mcauley	990000112387	1/1/2001	Leif Mcauley43@gmail.com
44	Ernesto	Onorato	990000112388	1/1/2001	Ernesto Onorato44@gmail.com
45	Renato	Staats	990000112389	1/1/2001	Renato Staats45@gmail.com
46	Tommy	Wattles	990000112390	1/1/2001	Tommy Wattles46@gmail.com
47	Antwan	Coffman	990000112391	1/1/2001	Antwan Coffman47@gmail.com
48	Rolando	Reagle	990000112392	1/1/2001	Rolando Reagle48@gmail.com
49	Lindsay	Kreutzer	990000112393	1/1/2001	Lindsay Kreutzer49@gmail.com
50	Antonio	Mckeller	990000112394	1/1/2001	Antonio Mckeller50@gmail.com
51	Rosalee	Estell	990000112395	1/1/2001	Rosalee Estell51@gmail.com
52	Olimpia	Smithers	990000112396	1/1/2001	Olimpia Smithers52@gmail.com
53	Shandra	Polzin	990000112397	1/1/2001	Shandra Polzin53@gmail.com
54	Cristine	Faller	990000112398	1/1/2001	Cristine Faller54@gmail.com
55	Waneta	Miro	990000112399	1/1/2001	Waneta Miro55@gmail.com
56	Joanna	Ebner	990000112400	1/1/2001	Joanna Ebner56@gmail.com
57	Cristie	Klem	990000112401	1/1/2001	Cristie Klem57@gmail.com
58	Dorotha	Shack	990000112402	1/1/2001	Dorotha Shack58@gmail.com
59	Alejandrina	Felt	990000112403	1/1/2001	Alejandrina Felt59@gmail.com
60	Suzette	Harriston	990000112404	1/1/2001	Suzette Harriston60@gmail.com
61	Jillian	Janecek	990000112405	1/1/2001	Jillian Janecek61@gmail.com
62	Euna	Grays	990000112406	1/1/2001	Euna Grays62@gmail.com
63	Rosy	Nuckols	990000112407	1/1/2001	Rosy Nuckols63@gmail.com
64	An	Manna	990000112408	1/1/2001	An Manna64@gmail.com
65	Beatris	Woodward	990000112409	1/1/2001	Beatris Woodward65@gmail.com
66	Sha	Rochin	990000112410	1/1/2001	Sha Rochin66@gmail.com
67	Rossie	Hamer	990000112411	1/1/2001	Rossie Hamer67@gmail.com
68	Danille	Sharples	990000112412	1/1/2001	Danille Sharples68@gmail.com
69	Elouise	Monteleone	990000112413	1/1/2001	Elouise Monteleone69@gmail.com
70	Brianne	Cazares	990000112414	1/1/2001	Brianne Cazares70@gmail.com
71	Rhea	Mccollom	990000112415	1/1/2001	Rhea Mccollom71@gmail.com
72	Ashlee	Guthridge	990000112416	1/1/2001	Ashlee Guthridge72@gmail.com
73	Jule	Pratts	990000112417	1/1/2001	Jule Pratts73@gmail.com
74	Shelley	Remy	990000112418	1/1/2001	Shelley Remy74@gmail.com
75	Wendie	Rodriguez	990000112419	1/1/2001	Wendie Rodriguez75@gmail.com
76	Christena	Ruhland	990000112420	1/1/2001	Christena Ruhland76@gmail.com
77	Irene	Bozek	990000112421	1/1/2001	Irene Bozek77@gmail.com
78	Meaghan	Geddie	990000112422	1/1/2001	Meaghan Geddie78@gmail.com
79	Rosita	Gebhardt	990000112423	1/1/2001	Rosita Gebhardt79@gmail.com

80	Liza	Fell	990000112424	1/1/2001	Liza Fell80@gmail.com
81	Marjory	Slate	990000112425	1/1/2001	Marjory Slate81@gmail.com
82	Concetta	Hornberger	990000112426	1/1/2001	Concetta Hornberger82@gmail.com
83	Rebecca	Harcrow	990000112427	1/1/2001	Rebecca Harcrow83@gmail.com
84	Eryn	Tineo	990000112428	1/1/2001	Eryn Tineo84@gmail.com
85	Maisie	Domenico	990000112429	1/1/2001	Maisie Domenico85@gmail.com
86	Danika	Winer	990000112430	1/1/2001	Danika Winer86@gmail.com
87	Fredericka	Rux	990000112431	1/1/2001	Fredericka Rux87@gmail.com
88	Kiyoko	Hamid	990000112432	1/1/2001	Kiyoko Hamid88@gmail.com
89	Armida	Kitchell	990000112433	1/1/2001	Armida Kitchell89@gmail.com
90	Melita	Boroughs	990000112434	1/1/2001	Melita Boroughs90@gmail.com
91	Rubie	Riche	990000112435	1/1/2001	Rubie Riche91@gmail.com
92	Hilma	Wasielewski	990000112436	1/1/2001	Hilma Wasielewski92@gmail.com
93	Sonja	Zalenski	990000112437	1/1/2001	Sonja Zalenski93@gmail.com
94	Amada	Deweese	990000112438	1/1/2001	Amada Deweese94@gmail.com
95	Jetta	Garica	990000112439	1/1/2001	Jetta Garica95@gmail.com
96	Trina	Dilullo	990000112440	1/1/2001	Trina Dilullo96@gmail.com
97	Betsy	Bevington	990000112441	1/1/2001	Betsy Bevington97@gmail.com
98	Masako	Wallick	990000112442	1/1/2001	Masako Wallick98@gmail.com
99	Asha	Mcsween	990000112443	1/1/2001	Asha Mcsween99@gmail.com
100	Camila	Setzer	990000112444	1/1/2001	Camila Setzer100@gmail.com

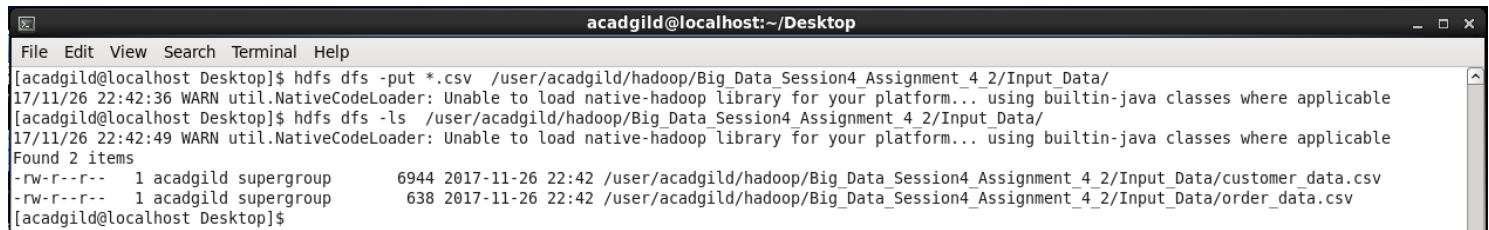
order_data.csv

OD01	1	1	450	1/24/2016	1	Apple iPhone 4S
OD01	1	1	450	1/24/2016	2	Apple iPhone 5S
OD01	1	1	450	1/24/2016	3	Apple iPhone 6
OD02	2	2	400	1/24/2016	1	Apple iPhone 4S
OD02	2	2	400	1/24/2016	2	Apple iPhone 5S
OD02	2	2	400	1/24/2016	3	Apple iPhone 6
OD03	2	1	350	1/25/2016	1	Apple iPhone 4S
OD03	2	1	350	1/25/2016	2	Apple iPhone 5S
OD03	2	1	350	1/25/2016	3	Apple iPhone 6
OD04	3	1	500	1/25/2016	1	Apple iPhone 4S
OD04	3	1	500	1/25/2016	2	Apple iPhone 5S
OD04	3	1	500	1/27/2016	3	Apple iPhone 6
OD05	4	1	650	1/26/2016	1	Apple iPhone 4S
OD05	4	1	650	1/27/2016	2	Apple iPhone 5S
OD05	4	1	650	1/27/2016	3	Apple iPhone 6

Put the dataset in HDFS

Hdfs location :-

/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/



```

acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
[acadgild@localhost Desktop]$ hdfs dfs -put *.csv /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/
17/11/26 22:42:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost Desktop]$ hdfs dfs -ls /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/
17/11/26 22:42:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 6944 2017-11-26 22:42 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/customer_data.csv
-rw-r--r-- 1 acadgild supergroup 638 2017-11-26 22:42 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv
[acadgild@localhost Desktop]$
  
```

MapReduce Mode - To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation. You can specify mapreduce mode using the -x flag

pig -x mapreduce

Command 1:- Concat

```

grunt>customers_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/customer_data.csv' using PigStorage(',') as
(Customer_ID:int,
First_Name:chararray,
Second_Name:chararray,
Contact_No:Long,
Created_Date:chararray,
Email_ID:chararray);

grunt>trim_customers_details = FOREACH customers_details GENERATE Customer_ID, TRIM(First_Name) as
First_name,Second_Name,Contact_No,Created_Date,Email_ID;

grunt>customers_details_concat = foreach trim_customers_details Generate CONCAT (First_name,'_',Second_Name);

grunt>STORE customers_details_concat INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_01_concat';
  
```

Output

```
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
ggregation.
2017-11-27 01:57:24,382 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
grunt>
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_01_concat/part-m-000000
Tracy_Stokley
Ken_Grigg
Jacques_Sandoz
Rickie_Ciampa
Andreas_Vicini
Kurt_Gariepy
Hoyt_Ruddick
Moshe_Bloomberg
Steve_Lindamood
Sandy_Yearta
Tyler_Alves
Kerry_Prindle
Lucio_Spiro
Shannon_Dufour
Leonardo_Negrete
Kim_Mccurley
Wilson_Demelo
Chuck_Deleo
Seth_Rayford
Conrad_Melville
Sebastian_Dorton
Dana_Toms
Wyatt_Strope
Cletus_Exum
Josh_Heffley
Lesley_Mccreary
Chad_Lepley
Freddy_Como
Huah_Chalk
[*new file (~/.Desktop)...] [acadgild@localhost:~] [Browsing HDFS - Moz...] acadgild@localhost:~/...
```

Tracy_Stokley

Ken_Grigg

Jacques_Sandoz

Rickie_Ciampa

Andreas_Vicini

Kurt_Gariepy

Hoyt_Ruddick

Moshe_Bloomberg

Steve_Lindamood

Sandy_Yearta

Tyler_Alves

Kerry_Prindle

Lucio_Spiro

Shannon_Dufour

Leonardo_Negrete

Kim_Mccurley

Wilson_Demelo

Chuck_Deleo

Seth_Rayford

Conrad_Melville

Sebastian_Dorton

Dana_Toms

Wyatt_Strope

Cletus_Exum

Josh_Heffley

Lesley_Mccreary

Chad_Lepley

Freddy_Como

Hugh_Chalk

Marshall_Skipworth

Jim_Jonason

Ed_Gambrel

Stacey_Tostado

Everett_Mooneyham

Greg_Mcvey

Roberto_Collier

Desmond_Carasco

Dewey_Marler

Ivan_Buss

Sonny_Cutter

Eldridge_Wilcox

Christian_Anders

Leif_Mcauley

Ernesto_Onorato

Renato_Staats

Tommy_Wattles

Antwan_Coffman

Rolando_Reagle

Lindsay_Kreutzer

Antonio_Mckeller

Rosalee_Estell

Olimpia_Smithers

Shandra_Polzin

Cristine_Faller

Waneta_Miro

Joanna_Ebner

Cristie_Klem

Dorotha_Shack

Alejandrina_Felt

Suzette_Harriston

Jillian_Janecek

Euna_Grays

Rosy_Nuckols

An_Manna

Beatris_Woodward

Sha_Rochin

Rossie_Hamer

Danille_Sharples

Elouise_Monteleone

Brianne_Cazares

Rhea_Mccollom

Ashlee_Guthridge

Jule_Pratts

Shelley_Remy

Wendie_Rodriguez

Christena_Ruhland

Irene_Bozek

Meaghan_Geddie

Rosita_Gebhardt

Liza_Fell

Marjory_Slate

Concetta_Hornberger

Rebecca_Harcrow

Eryn_Tineo

Maisie_Domenico

Danika_Winer

Fredericka_Rux

Kiyoko_Hamid

Armida_Kitchell

Melita_Boroughs

Rubie_Riche

Hilma_Wasielewski

Sonja_Zalenski

Amada_Deweese

Jetta_Garica

Trina_Dilullo

Betsy_Bevington

Masako_Wallick

Asha_Mcsween

Camila_Setzer

Command 2:- Tokenize

```
grunt>customers_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/customer_data.csv' using
PigStorage(',') as
```

```
(Customer_ID:int,
```

```
First_Name:chararray,
```

```
Second_Name:chararray,
```

```
Contact_No:Long,
```

```
Created_Date:chararray,
```

```
Email_ID:chararray);
```

```
grunt>customers_details_tokenize = foreach customers_details Generate TOKENIZE(Email_ID);
```

```
grunt>STORE customers_details_tokenize INTO
```

```
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_02_tokenize';
```

Output

```
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_02_tokenize/part-m-00000
{(Tracy), (Stokley1@gmail.com)}
{(Ken), (Grigg2@gmail.com)}
{(Jacques), (Sandoz3@gmail.com)}
{(Rickie), (Ciampa4@gmail.com)}
{(Andreas), (Vicini5@gmail.com)}
{(Kurt), (Gariepy6@gmail.com)}
{(Hoyt), (Ruddick7@gmail.com)}
{(Moshe), (Bloomberg8@gmail.com)}
{(Steve), (Lindamood9@gmail.com)}
{(Sandy), (Yearta10@gmail.com)}

{(Kerry), (Prindle12@gmail.com)}
{(Lucio), (Spiro13@gmail.com)}
{(Shannon), (Dufour14@gmail.com)}

{(Kim), (Mccurley16@gmail.com)}
{(Wilson), (Demelo17@gmail.com)}
{(Chuck), (Deleo18@gmail.com)}
{(Seth), (Rayford19@gmail.com)}
{(Conrad), (Melville20@gmail.com)}
{(Sebastian), (Dorton21@gmail.com)}

{(Wyatt), (Strope23@gmail.com)}
{(Cletus), (Exum24@gmail.com)}
{(Josh), (Heffley25@gmail.com)}
{(Lesley), (Mccreary26@gmail.com)}
{(Chad), (Lepley27@gmail.com)}
{(Freddy), (Como28@gmail.com)}
{(Hugh), (Chalk29@gmail.com)}
{(Marshall), (Skipworth30@gmail.com)}
{(Jim), (Jonason31@gmail.com)}
{( ), ( )}
new file (~/Desktop) - gedit
```

{(Tracy),(Stokley1@gmail.com)}

{(Ken),(Grigg2@gmail.com)}

{(Jacques),(Sandoz3@gmail.com)}

{(Rickie),(Ciampa4@gmail.com)}

{(Andreas),(Vicini5@gmail.com)}

{(Kurt),(Gariepy6@gmail.com)}

{(Hoyt),(Ruddick7@gmail.com)}

{(Moshe),(Bloomberg8@gmail.com)}

{(Steve),(Lindamood9@gmail.com)}

{(Sandy),(Yearta10@gmail.com)}

{(Kerry),(Prindle12@gmail.com)}

{(Lucio),(Spiro13@gmail.com)}

{(Shannon),(Dufour14@gmail.com)}

{(Kim),(Mccurley16@gmail.com)}

{(Wilson),(Demelo17@gmail.com)}

{(Chuck),(Deleo18@gmail.com)}

{(Seth),(Rayford19@gmail.com)}

{(Conrad),(Melville20@gmail.com)}

{(Sebastian),(Dorton21@gmail.com)}

{(Wyatt),(Strope23@gmail.com)}

{(Cletus),(Exum24@gmail.com)}

{(Josh),(Heffley25@gmail.com)}

{(Lesley),(Mccreary26@gmail.com)}

{(Chad),(Lepley27@gmail.com)}

{(Freddy),(Como28@gmail.com)}

{(Hugh),(Chalk29@gmail.com)}

{(Marshall),(Skipworth30@gmail.com)}

{(Jim),(Jonason31@gmail.com)}

{(Ed),(Gambrel32@gmail.com)}

{(Stacey),(Tostado33@gmail.com)}

{(Everett),(Mooneyham34@gmail.com)}

{(Greg),(Mcvey35@gmail.com)}

{(Roberto),(Collier36@gmail.com)}

{(Desmond),(Carasco37@gmail.com)}

{(Dewey),(Marler38@gmail.com)}

{(Ivan),(Buss39@gmail.com)}

{(Sonny),(Cutter40@gmail.com)}

{(Eldridge),(Wilcox41@gmail.com)}

{(Christian),(Anders42@gmail.com)}

{(Leif),(Mcauley43@gmail.com)}

{(Ernesto),(Onorato44@gmail.com)}

{(Renato),(Staats45@gmail.com)}

{(Tommy),(Wattles46@gmail.com)}

{(Antwan),(Coffman47@gmail.com)}

{(Rolando),(Reagle48@gmail.com)}

{(Lindsay),(Kreutzer49@gmail.com)}

{(Antonio),(Mckeller50@gmail.com)}

{(Rosalee),(Estell51@gmail.com)}

{(Olimpia),(Smithers52@gmail.com)}

{(Shandra),(Polzin53@gmail.com)}

{(Cristine),(Faller54@gmail.com)}

{(Waneta),(Miro55@gmail.com)}

{(Joanna),(Ebner56@gmail.com)}

{(Cristie),(Klem57@gmail.com)}

{(Dorotha),(Shack58@gmail.com)}

{(Alejandrina),(Felt59@gmail.com)}

{(Suzette),(Harriston60@gmail.com)}

{(Jillian),(Janecek61@gmail.com)}

{(Euna),(Grays62@gmail.com)}

{(Rosy),(Nuckols63@gmail.com)}

{(An),(Manna64@gmail.com)}

{(Beatris),(Woodward65@gmail.com)}

{(Sha),(Rochin66@gmail.com)}

{(Rossie),(Hamer67@gmail.com)}

{(Danille),(Sharples68@gmail.com)}

{(Elouise),(Monteleone69@gmail.com)}

{(Brianne),(Cazares70@gmail.com)}

{(Rhea),(Mccollom71@gmail.com)}

{(Ashlee),(Guthridge72@gmail.com)}

{(Jule),(Pratts73@gmail.com)}

{(Shelley),(Remy74@gmail.com)}

{(Wendie),(Rodriguez75@gmail.com)}

{(Christena),(Ruhland76@gmail.com)}

{(Irene),(Bozek77@gmail.com)}

{(Meaghan),(Geddie78@gmail.com)}

{(Rosita),(Gebhardt79@gmail.com)}

{(Liza),(Fell80@gmail.com)}

{(Marjory),(Slate81@gmail.com)}

{(Concetta),(Hornberger82@gmail.com)}

{(Rebecca),(Harcrow83@gmail.com)}

{(Eryn),(Tineo84@gmail.com)}

{(Maisie),(Domenico85@gmail.com)}

{(Danika),(Winer86@gmail.com)}

{(Fredericka),(Rux87@gmail.com)}

{(Kiyoko),(Hamid88@gmail.com)}

{(Armida),(Kitchell89@gmail.com)}

{(Melita),(Boroughs90@gmail.com)}

{(Rubie),(Riche91@gmail.com)}

{(Hilma),(Wasielewski92@gmail.com)}

{(Sonja),(Zalenski93@gmail.com)}

```
{(Amada),(Deweese94@gmail.com)}
{(Jetta),(Garica95@gmail.com)}
{(Trina),(Dilullo96@gmail.com)}
{(Betsy),(Bevington97@gmail.com)}
{(Masako),(Wallick98@gmail.com)}
{(Asha),(Mcsween99@gmail.com)}
{(Camila),(Setzer100@gmail.com)}
```

Command 3:- Sum

```
grunt> orders_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv' using
PigStorage(',') as
```

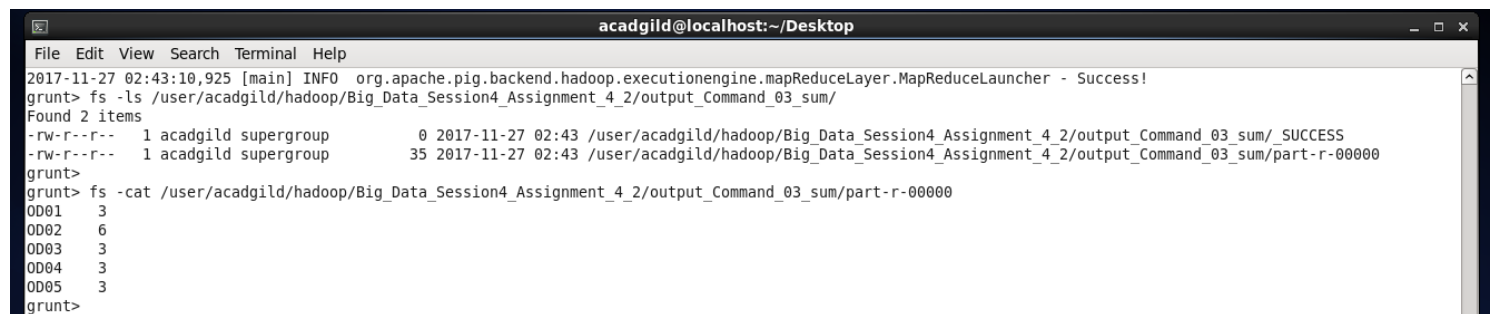
```
(Order_ID:chararray, Customer_ID:int, Order_Quantity:int, Order_Value:
float, Order_Date:chararray, Product_ID:int,
Product_Name:chararray);
```

```
grunt> group_by = group orders_details BY Order_ID;
```

```
grunt> aggr_by_order = foreach group_by generate group, SUM(orders_details.Order_Quantity);
```

```
grunt> STORE aggr_by_order INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_03_sum';
```

Output



```
acadgild@localhost: ~/Desktop
File Edit View Search Terminal Help
2017-11-27 02:43:10,925 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> fs -ls /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_03_sum/
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-11-27 02:43 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_03_sum/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 35 2017-11-27 02:43 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_03_sum/part-r-000000
grunt>
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_03_sum/part-r-000000
OD01 3
OD02 6
OD03 3
OD04 3
OD05 3
grunt>
```

OD01 3

OD02 6

OD03 3


OD04 3

OD05 3

Command 4:- Min

```
grunt>orders_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv' using
PigStorage(',') as
(Order_ID:chararray,Customer_ID:int,Order_Quantity:int,Order_Value:
float,Order_Date:chararray, Product_ID:int,
Product_Name:chararray);
grunt>group_by = group orders_details BY Order_ID;
grunt>min_by_order = foreach group_by generate group,MIN(orders_details.Order_Quantity);
grunt>STORE min_by_order INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_04_min';
```

Output



```
File Edit View Search Terminal Help
ng to job history server
2017-11-27 03:00:16,197 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retr
ggregation.
2017-11-27 03:00:16,197 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_04_min/part-r-00000
OD01      1
OD02      2
OD03      1
OD04      1
OD05      1
grunt>
```

OD01 1

OD02 2

OD03 1

OD04 1

OD05 1

Command 5:- Max

```
grunt>orders_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv' using
PigStorage(',') as
```

```
(Order_ID:chararray,Customer_ID:int,Order_Quantity:int,Order_Value:
float,Order_Date:chararray, Product_ID:int,
Product_Name:chararray);
```

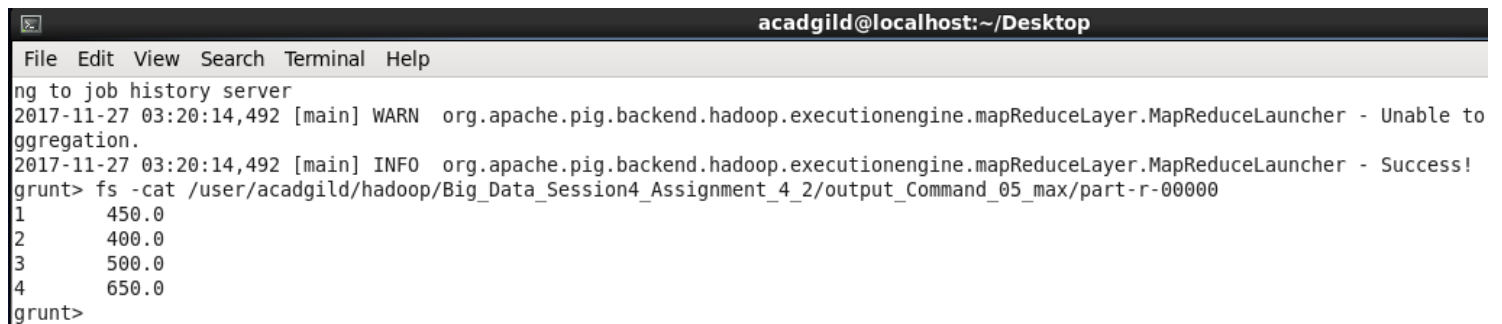
```
grunt>group_by = group orders_details BY Customer_ID;
```

```
grunt>max_by_order = foreach group_by generate group,MAX(orders_details.Order_Value);
```

```
grunt>STORE max_by_order INTO
```

```
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_05_max';
```

Output



```
File Edit View Search Terminal Help
ng to job history server
2017-11-27 03:20:14,492 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to
ggregation.
2017-11-27 03:20:14,492 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_05_max/part-r-00000
1      450.0
2      400.0
3      500.0
4      650.0
grunt>
```

1 450.0

2 400.0

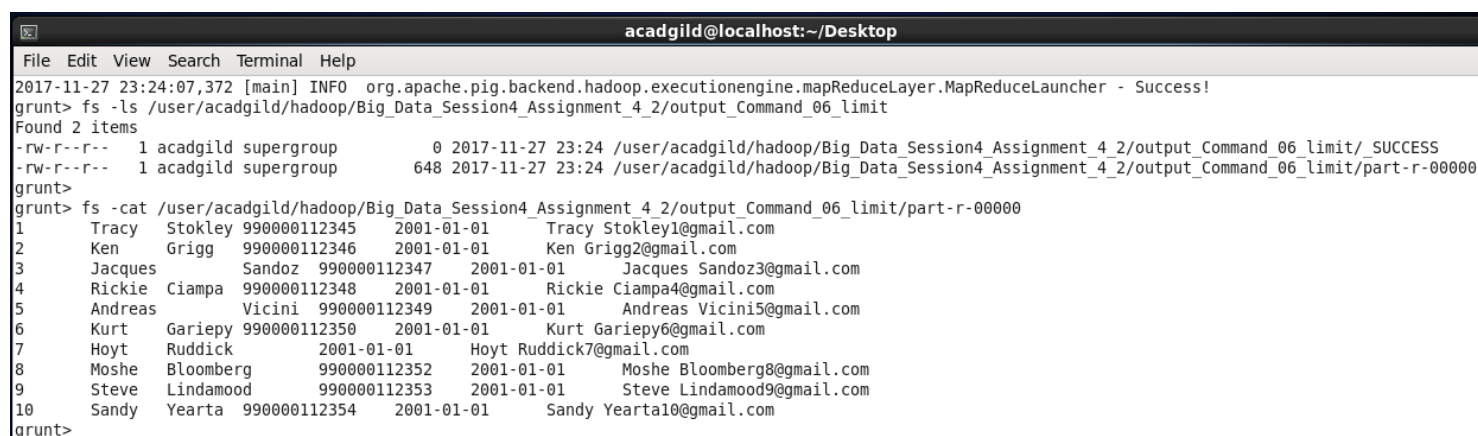
3 500.0

4 650.0

Command 6:- Limit

```
grunt>customers_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/customer_data.csv' using
PigStorage(',') as (Customer_ID:int, First_Name:chararray,
Second_Name:chararray,
Contact_No:Long,
Created_Date:chararray,
Email_ID:chararray);
grunt>limit_order = Limit customers_details 10;
grunt>STORE limit_order INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_06_limit';
```

Output



```
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
2017-11-27 23:24:07,372 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> fs -ls /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_06_limit
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-11-27 23:24 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_06_limit/SUCCESS
-rw-r--r-- 1 acadgild supergroup 648 2017-11-27 23:24 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_06_limit/part-r-00000
grunt>
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_06_limit/part-r-00000
1 Tracy Stokley 990000112345 2001-01-01 Tracy Stokley1@gmail.com
2 Ken Grigg 990000112346 2001-01-01 Ken Grigg2@gmail.com
3 Jacques Sandoz 990000112347 2001-01-01 Jacques Sandoz3@gmail.com
4 Rickie Ciampa 990000112348 2001-01-01 Rickie Ciampa4@gmail.com
5 Andreas Vicini 990000112349 2001-01-01 Andreas Vicini5@gmail.com
6 Kurt Gariepy 990000112350 2001-01-01 Kurt Gariepy6@gmail.com
7 Hoyt Ruddick 2001-01-01 Hoyt Ruddick7@gmail.com
8 Moshe Bloomberg 990000112352 2001-01-01 Moshe Bloomberg8@gmail.com
9 Steve Lindamood 990000112353 2001-01-01 Steve Lindamood9@gmail.com
10 Sandy Yearta 990000112354 2001-01-01 Sandy Yearta10@gmail.com
grunt>
```

1	Tracy Stokley	990000112345	2001-01-01	Tracy Stokley1@gmail.com
2	Ken Grigg	990000112346	2001-01-01	Ken Grigg2@gmail.com
3	Jacques Sandoz	990000112347	2001-01-01	Jacques Sandoz3@gmail.com
4	Rickie Ciampa	990000112348	2001-01-01	Rickie Ciampa4@gmail.com
5	Andreas Vicini	990000112349	2001-01-01	Andreas Vicini5@gmail.com
6	Kurt Gariepy	990000112350	2001-01-01	Kurt Gariepy6@gmail.com
7	Hoyt Ruddick		2001-01-01	Hoyt Ruddick7@gmail.com
8	Moshe Bloomberg	990000112352	2001-01-01	Moshe Bloomberg8@gmail.com
9	Steve Lindamood	990000112353	2001-01-01	Steve Lindamood9@gmail.com
10	Sandy Yearta	990000112354	2001-01-01	Sandy Yearta10@gmail.com

Command 7:- Store

From Command 1 to Command 6, I have used the Store to store the results in a folder.

One more example of Store is given below

```
grunt>orders_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv' using
PigStorage(',') as
```

```
(Order_ID:chararray,Customer_ID:int,Order_Quantity:int,Order_Value:
```

```
float,Order_Date:chararray, Product_ID:int,
```

```
Product_Name:chararray);
```

```
grunt>quantity_value_product = foreach orders_details generate (Order_Quantity * Order_Value) as
quantity_value_product,Order_ID,Customer_ID,ROUND(Order_Value) as Trimmed_Value;
```

```
grunt>STORE quantity_value_product INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_07_store';
```

Output

```
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
gggregation.
2017-11-28 00:44:45,996 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> fs -ls /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_07_store/
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-11-28 00:44 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_07_store/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 255 2017-11-28 00:44 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_07_store/part-m-000000
grunt>
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_07_store/part-m-000000
450.0 OD01 1 450
450.0 OD01 1 450
450.0 OD01 1 450
800.0 OD02 2 400
800.0 OD02 2 400
800.0 OD02 2 400
350.0 OD03 2 350
350.0 OD03 2 350
350.0 OD03 2 350
500.0 OD04 3 500
500.0 OD04 3 500
500.0 OD04 3 500
650.0 OD05 4 650
650.0 OD05 4 650
650.0 OD05 4 650
grunt>
```

450.0 OD01 1 450

450.0 OD01 1 450

450.0 OD01 1 450

800.0 OD02 2 400

800.0 OD02 2 400

800.0 OD02 2 400

350.0 OD03 2 350

350.0 OD03 2 350

```

350.0 OD03 2      350
500.0 OD04 3      500
500.0 OD04 3      500
500.0 OD04 3      500
650.0 OD05 4      650
650.0 OD05 4      650
650.0 OD05 4      650

```

Command 8:- Distinct

```

grunt> orders_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv' using
PigStorage(',') as

```

```

(Order_ID:chararray, Customer_ID:int, Order_Quantity:int, Order_Value:
float, Order_Date:chararray, Product_ID:int,
Product_Name:chararray);

```

```

grunt> orders_temp = foreach orders_details generate Order_ID;

```

```

grunt> distinct_orders = distinct orders_temp;

```

```

grunt> STORE distinct_orders INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_08_distinct';

```

Output



```

acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
gggregation.
2017-11-28 00:58:44,303 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> fs -ls /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_08_distinct
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-11-28 00:58 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_08_distinct/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 25 2017-11-28 00:58 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_08_distinct/part-r-00000
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_08_distinct/part-r-00000
OD01
OD02
OD03
OD04
OD05
grunt>
grunt>
grunt>

```

OD01

OD02

OD03

OD04

OD05

Command 9:- Flatten

```
grunt> orders_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv' using
PigStorage(',') as
```

```
(Order_ID:chararray, Customer_ID:int, Order_Quantity:int, Order_Value:
```

```
float, Order_Date:chararray, Product_ID:int,
```

```
Product_Name:chararray);
```

```
grunt> groupd = group orders_details by Customer_ID;
```

```
grunt> customer_product = foreach groupd generate group as
Customer_ID, orders_details.Product_Name as Products_Bought;
```

```
grunt> flattened_customer_product = foreach customer_product generate Customer_ID,
flatten(Products_Bought) as Products;
```

```
grunt> STORE flattened_customer_product INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_09_flatten';
```

Output



```
acadgild@localhost:~/Desktop
File Edit View Search Terminal Help
2017-11-28 01:51:33,955 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> fs -ls /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_09_flatten
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-11-28 01:51 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_09_flatten/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 265 2017-11-28 01:51 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_09_flatten/part-r-00000
grunt>
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_09_flatten/part-r-00000
1 Apple iPhone 4S
1 Apple iPhone 5S
1 Apple iPhone 6
2 Apple iPhone 5S
2 Apple iPhone 6
2 Apple iPhone 4S
2 Apple iPhone 6
2 Apple iPhone 5S
2 Apple iPhone 4S
3 Apple iPhone 5S
3 Apple iPhone 4S
3 Apple iPhone 6
4 Apple iPhone 4S
4 Apple iPhone 5S
4 Apple iPhone 6
grunt>
grunt>
```

1 Apple iPhone 4S

1 Apple iPhone 5S

1 Apple iPhone 6

2 Apple iPhone 5S

2 Apple iPhone 6

2 Apple iPhone 4S

- 2 Apple iPhone 6
- 2 Apple iPhone 5S
- 2 Apple iPhone 4S
- 3 Apple iPhone 5S
- 3 Apple iPhone 4S
- 3 Apple iPhone 6
- 4 Apple iPhone 4S
- 4 Apple iPhone 5S
- 4 Apple iPhone 6

Command 10:- IsEmpty

```
grunt> orders_details = load
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/Input_Data/order_data.csv' using
PigStorage(',') as
(Order_ID:chararray, Customer_ID:int, Order_Quantity:int, Order_Value:
float, Order_Date:chararray, Product_ID:int,
Product_Name:chararray);
grunt> order_bag = foreach orders_details generate
TOBAG(Order_ID, Customer_ID, Order_Quantity, Order_Value, Order_Date, Product_ID, Product_Name) as
Order_Customer_ID;
grunt> product_order_bag = filter order_bag by not IsEmpty(Order_Customer_ID);
grunt> STORE product_order_bag INTO
'/user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_10_IsEmpty';
```

Output

acadgild@localhost:~/Desktop

File Edit View Search Terminal Help

```

gggregation.
2017-11-28 03:02:56,588 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
grunt>
grunt> fs -ls /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_10_IsEmpty
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2017-11-28 03:02 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_10_IsEmpty/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 895 2017-11-28 03:02 /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_10_IsEmpty/part-m-000000
grunt>
grunt> fs -cat /user/acadgild/hadoop/Big_Data_Session4_Assignment_4_2/output_Command_10_IsEmpty/part-m-000000
{(OD01),(1),(1),(450.0),(2016-01-24),(1),(Apple iPhone 4S)}
{(OD01),(1),(1),(450.0),(2016-01-24),(2),(Apple iPhone 5S)}
{(OD01),(1),(1),(450.0),(2016-01-24),(3),(Apple iPhone 6)}
{(OD02),(2),(2),(400.0),(2016-01-24),(1),(Apple iPhone 4S)}
{(OD02),(2),(2),(400.0),(2016-01-24),(2),(Apple iPhone 5S)}
{(OD02),(2),(2),(400.0),(2016-01-24),(3),(Apple iPhone 6)}
{(OD03),(2),(1),(350.0),(2016-01-25),(1),(Apple iPhone 4S)}
{(OD03),(2),(1),(350.0),(2016-01-25),(2),(Apple iPhone 5S)}
{(OD03),(2),(1),(350.0),(2016-01-25),(3),(Apple iPhone 6)}
{(OD04),(3),(1),(500.0),(2016-01-25),(1),(Apple iPhone 4S)}
{(OD04),(3),(1),(500.0),(2016-01-25),(2),(Apple iPhone 5S)}
{(OD04),(3),(1),(500.0),(2016-01-27),(3),(Apple iPhone 6)}
{(OD05),(4),(1),(650.0),(2016-01-26),(1),(Apple iPhone 4S)}
{(OD05),(4),(1),(650.0),(2016-01-27),(2),(Apple iPhone 5S)}
{(OD05),(4),(1),(650.0),(2016-01-27),(3),(Apple iPhone 6)}
grunt>
grunt>

```

{(OD01),(1),(1),(450.0),(2016-01-24),(1),(Apple iPhone 4S)}

{(OD01),(1),(1),(450.0),(2016-01-24),(2),(Apple iPhone 5S)}

{(OD01),(1),(1),(450.0),(2016-01-24),(3),(Apple iPhone 6)}

{(OD02),(2),(2),(400.0),(2016-01-24),(1),(Apple iPhone 4S)}

{(OD02),(2),(2),(400.0),(2016-01-24),(2),(Apple iPhone 5S)}

{(OD02),(2),(2),(400.0),(2016-01-24),(3),(Apple iPhone 6)}

{(OD03),(2),(1),(350.0),(2016-01-25),(1),(Apple iPhone 4S)}

{(OD03),(2),(1),(350.0),(2016-01-25),(2),(Apple iPhone 5S)}

{(OD03),(2),(1),(350.0),(2016-01-25),(3),(Apple iPhone 6)}

{(OD04),(3),(1),(500.0),(2016-01-25),(1),(Apple iPhone 4S)}

{(OD04),(3),(1),(500.0),(2016-01-25),(2),(Apple iPhone 5S)}

{(OD04),(3),(1),(500.0),(2016-01-27),(3),(Apple iPhone 6)}

{(OD05),(4),(1),(650.0),(2016-01-26),(1),(Apple iPhone 4S)}

{(OD05),(4),(1),(650.0),(2016-01-27),(2),(Apple iPhone 5S)}

{(OD05),(4),(1),(650.0),(2016-01-27),(3),(Apple iPhone 6)}