

# Assignment 9.1

```
petrol_data = LOAD 'petrol_dataset' USING PigStorage(',') AS  
(dist_id:chararray,dist_name:chararray,buy_rate:chararray, sell_rate:chararray, volume_in:int,  
volume_out:int,year:int);
```

```
dump petrol_data;
```

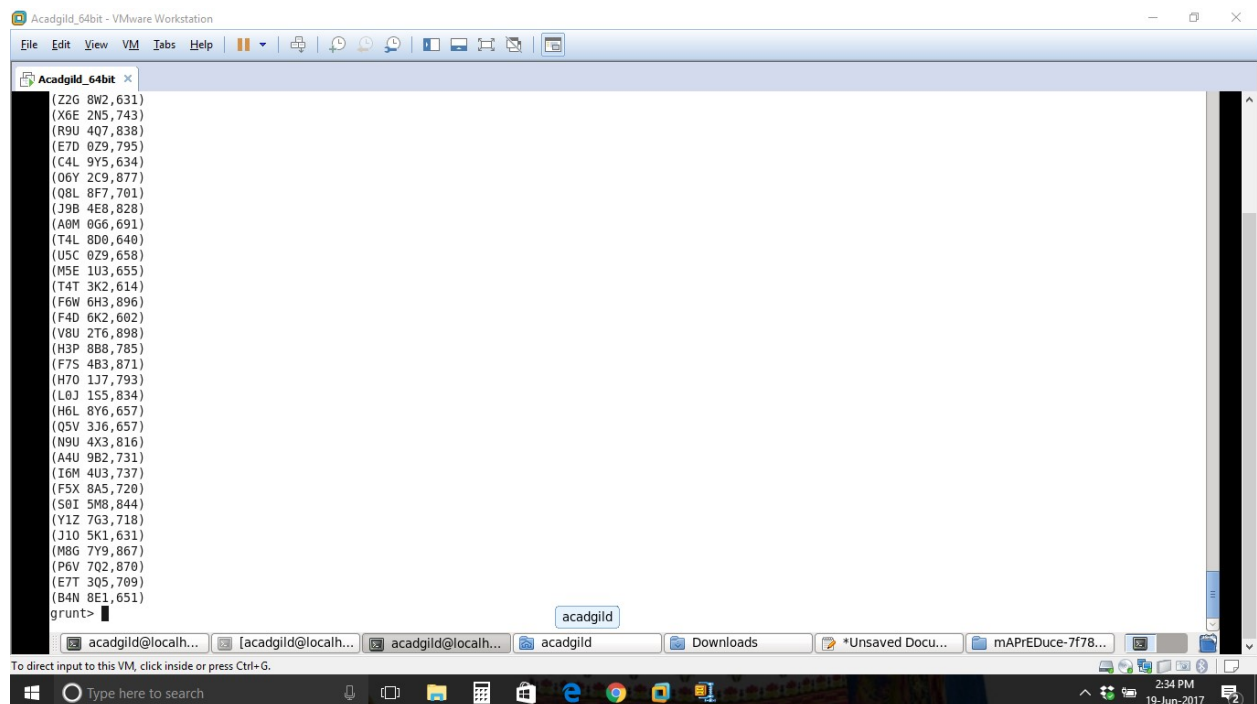
## 1.Problem Statement

**What is the total amount of petrol in volume sold by every distributor?**

```
petrol_volume_sold = foreach petrol_data generate dist_id, volume_out;
```

```
dump petrol_volume_sold;
```

**Distributor with petrol volume sold**



## 2.Problem Statement

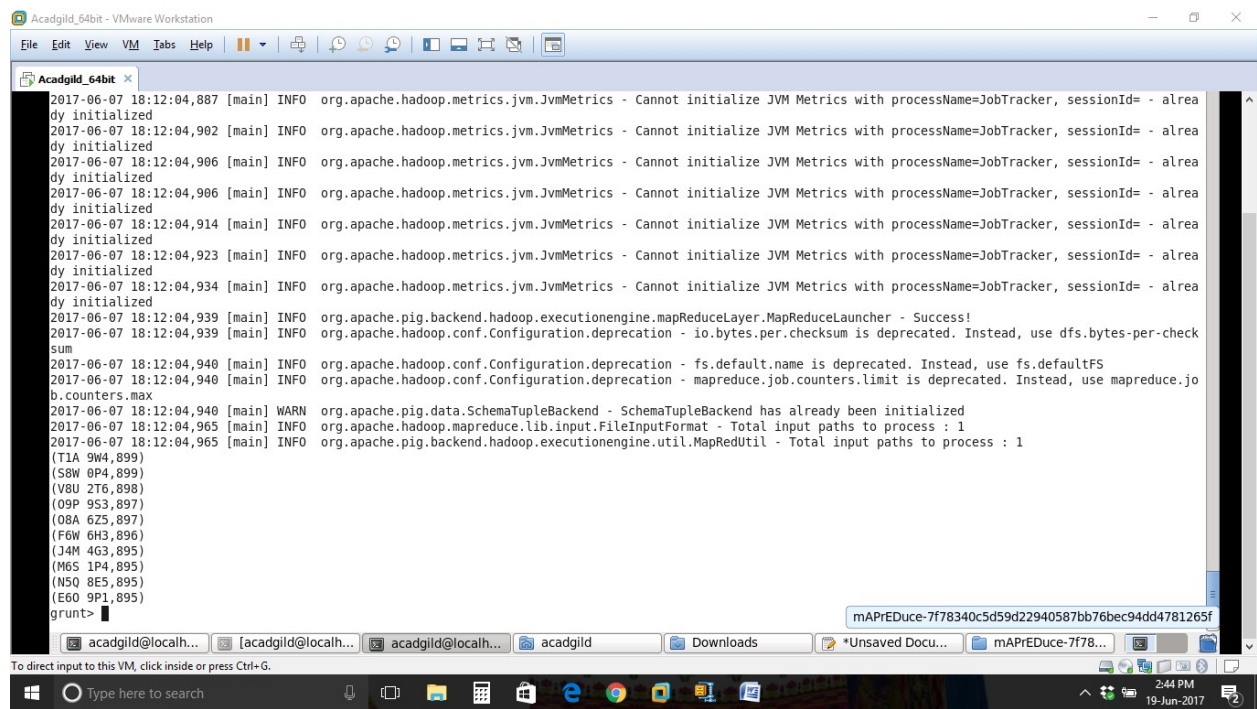
Which are the top 10 distributors ID's for selling petrol? Also display the amount of petrol sold in volume.

```
petrol_volume_top10 = ORDER petrol_volume_sold by volume_out desc;
```

```
top_10_distributor = limit petrol_volume_top10 10;
```

```
dump top_10_distributor;
```

**Top 10 distributor with volume sold details**



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
2017-06-07 18:12:04,887 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-07 18:12:04,902 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-07 18:12:04,906 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-07 18:12:04,906 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-07 18:12:04,914 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-07 18:12:04,923 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-07 18:12:04,934 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-07 18:12:04,939 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-07 18:12:04,939 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-07 18:12:04,940 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-07 18:12:04,940 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-07 18:12:04,940 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-07 18:12:04,965 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-07 18:12:04,965 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(TIA 9W4,899)
(SBW 0P4,899)
(VBU 2T6,898)
(O9P 9S3,897)
(O8A 6Z5,897)
(F6W 6H3,896)
(J4M 4G3,895)
(M6S 1P4,895)
(NSQ 8E5,895)
(E6Q 9P1,895)
grunt>
mAPrEDuce-7f78340c5d59d22940587bb76bec94dd4781265f
acacgild@localh... [acacgild@localh... acacgild@localh... acacgild Downloads *Unsaved Docu... mAPrEDuce-7f78...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 2:44 PM 19-Jun-2017
```

### 3.Problem Statement

List 10 years where consumption of petrol is more with the distributor id who sold it.

```
petrol_volume_sold_year = foreach petrol_data generate dist_id, volume_out,year;
```

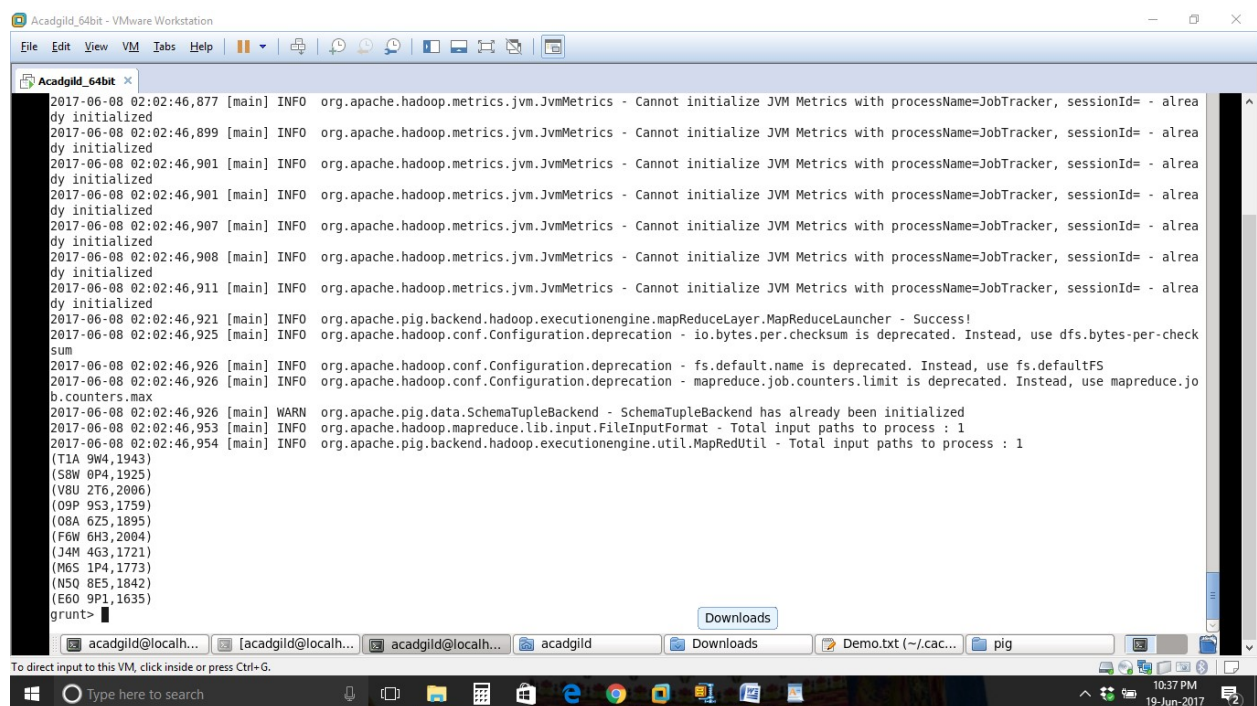
```
petrol_volume_top10_year = ORDER petrol_volume_sold_year by volume_out desc;
```

```
top_10_distributor_year = limit petrol_volume_top10_year 10;
```

```
top_consumption_year = foreach top_10_distributor_year generate dist_id,year;
```

```
dump top_consumption_year;
```

Top 10 years where petrol consumption was more



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
2017-06-08 02:02:46,877 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-08 02:02:46,899 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-08 02:02:46,901 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-08 02:02:46,901 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-08 02:02:46,907 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-08 02:02:46,908 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-08 02:02:46,911 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - alrea
dy initialized
2017-06-08 02:02:46,921 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-08 02:02:46,925 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-08 02:02:46,926 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-08 02:02:46,926 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-08 02:02:46,926 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-08 02:02:46,953 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-08 02:02:46,954 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(T1A 9W4,1943)
(58W 0P4,1925)
(V8U 2T6,2006)
(09P 9S3,1759)
(08A 6Z5,1895)
(F6W 6H3,2004)
(J4M 4G3,1721)
(M6S 1P4,1773)
(N5Q 8E5,1842)
(E60 9P1,1635)
grunt>
```

#### 4.Problem Statement

**Find the distributor name who sold petrol in least amount.**

```
petrol_volume_sold = foreach petrol_data generate dist_id,dist_name ,volume_out;
```

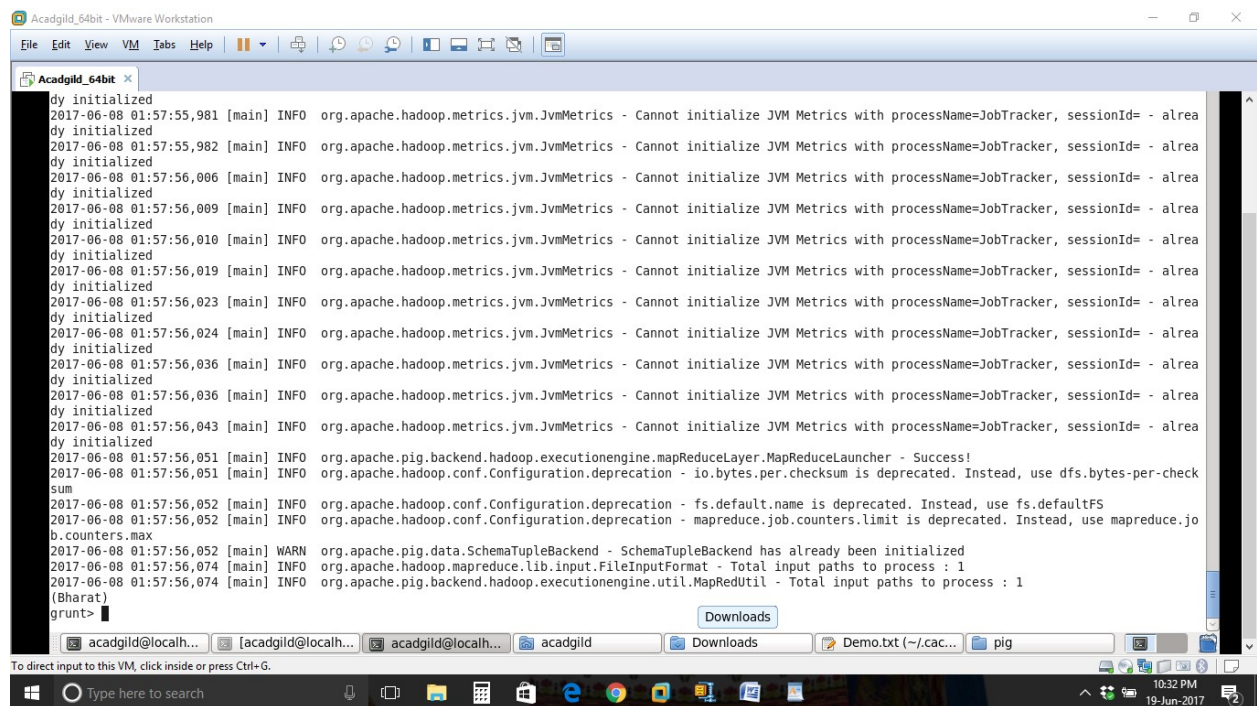
```
petrol_volume_ascending = ORDER petrol_volume_sold by volume_out asc;
```

```
distributor_with_least_sales = limit petrol_volume_ascending 1;
```

```
distributor_least_sale_name = foreach distributor_with_least_sales generate dist_name;
```

```
dump distributor_least_sale_name;
```

**Name of distributor who sold least amount of Petrol**



The screenshot shows a VMware Workstation window titled "Acadgild\_64bit - VMware Workstation". Inside, a terminal window titled "Acadgild\_64bit" displays the output of a Hadoop and Pig job. The logs show multiple "INFO" messages from "org.apache.hadoop.metrics.jvm.JvmMetrics" indicating that JVM Metrics could not be initialized with processName=JobTracker and sessionId=alrea. There are also "INFO" messages from "org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher" indicating success, and "WARN" messages from "org.apache.pig.data.SchemaTupleBackend" indicating it has already been initialized. The terminal window has a "Downloads" button and a list of files: "acadmild", "Downloads", "Demo.txt (~/.cac...", and "pig". The bottom of the window shows the Windows taskbar with the search bar and system clock.

```
dy initialized
2017-06-08 01:57:55,981 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:55,982 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,006 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,009 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,010 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,019 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,023 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,024 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,036 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,036 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,043 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processName=JobTracker, sessionId= alrea
dy initialized
2017-06-08 01:57:56,051 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-06-08 01:57:56,051 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-check
sum
2017-06-08 01:57:56,052 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-08 01:57:56,052 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.jo
b.counters.max
2017-06-08 01:57:56,052 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-06-08 01:57:56,074 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-06-08 01:57:56,074 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Bharat)
grunt>
```