

Finding Second Highest Salary Using Apache Pig

Use Case

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Finding Second Highest Salary Using Apache Pig

Table of Contents

Problem Statements:.....	2
Important Links:	2
Dataset Description:	2
Technology/Software Used:	2
Work Flow Diagram:	3
Implementation:.....	4

edureka!

Finding Second Highest Salary Using Apache Pig

Problem Statements:

- How to find out second highest salary.

Important Links:

- **Pig Installation:**

<https://edureka.wistia.com/medias/lpb6yiupps>

Dataset Description:

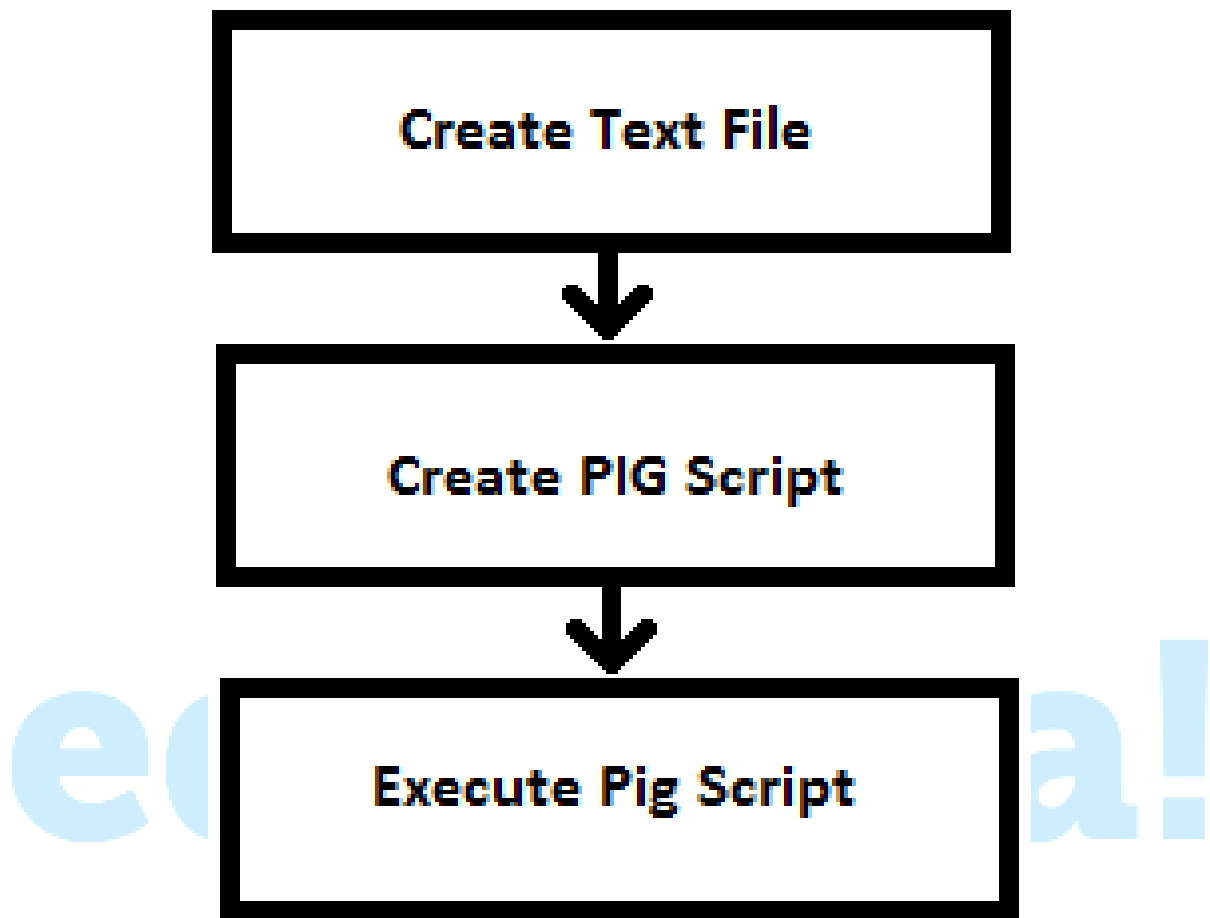
Here, we have taken a sample dataset which is shown in below image. It contains 6 rows and 2 columns. Columns are user and salary. To make it more explanatory, we have taken repeated values in salary column.

User	Salary
Jay	1000
Aroob	800
Max	600
Cris	800
Justin	400
Rock	1000

Technology/Software Used:

- Hadoop environment
- Apache Pig

Work Flow Diagram:

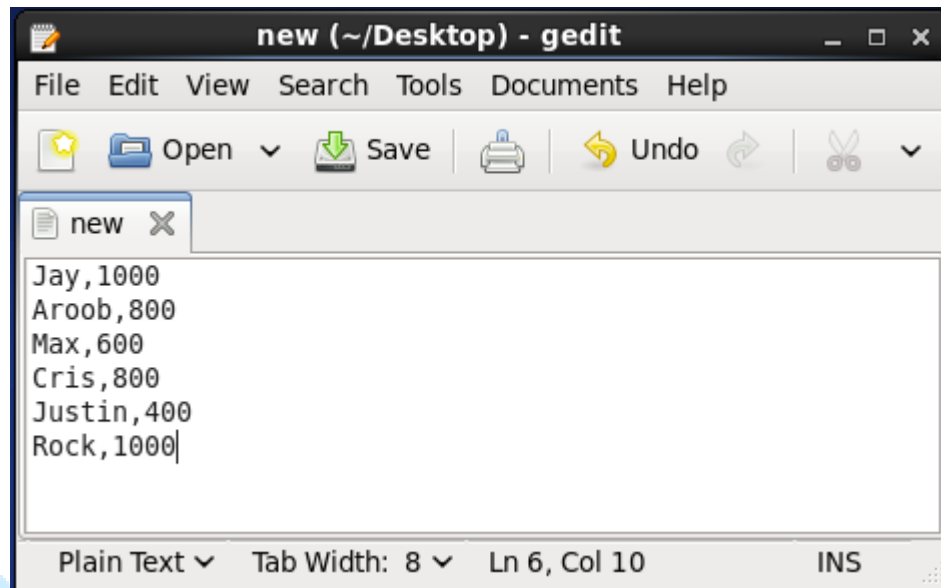


Implementation:

Please follow the below procedure to get the desired output:

Data File Creation:

Here, we have put comma (,) as a separator. After creating the file, we moved the file to HDFS. Please check the below image:



Pig Script Creation:

Here, we have put all the commands in one file and saved it as a **salary.pig**. Check the below text box (you can copy the below commands):

```
A = load 'hdfs:/new' using PigStorage(',') as (name:chararray, salary:int);  
B = group A by salary;  
C = foreach B generate A.name, group as salary1;  
D = ORDER C by salary1 DESC;  
E = limit D 2;  
F = ORDER E by salary1;  
G = limit F 1;  
H = foreach G generate flatten($0),$1;  
dump H;
```

Execution of Pig Script in Terminal:

By using following command, you can run the pig script.

```
[edureka@localhost ~]$ pig /home/edureka/Desktop/salary.pig
2014-12-26 18:05:07,077 [main] INFO org.apache.pig.Main - Apache Pig version 0.
12.1-SNAPSHOT (rexported) compiled Mar 22 2014, 07:41:29
2014-12-26 18:05:07,077 [main] INFO org.apache.pig.Main - Logging error message
s to: /home/edureka/pig_1419597
2014-12-26 18:05:07,339 [main] INFO org.apache.hadoop.util.NativeCodeLoader - U
nable to load native-hadoop lib... using builtin-java clas
ses where applicable
2014-12-26 18:05:07,535 [main] INFO org.apache.hadoop.mapreduce.util.Util - Default bo
otup file /home/edureka/.pigboot
2014-12-26 18:05:07,671 [main] INFO org.apache.hadoop.conf.Configuration.deprec
ation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.addr
ess
```

Give the complete path of salary.pig file where you have written pig script, after pig keyword. Script will run in MR mode.

Output:

Check the output:

```
2014-12-26 18:08:17,884 [main] INFO org.apache.hadoop.mapreduce.inputFormat - Total input paths to process : 1
2014-12-26 18:08:17,884 [main] INFO org.apache.hadoop.mapreduce.util.MapReduceUtil - Total input paths to process : 1
(Cris,800)
(Aroob,800)
2014-12-26 18:08:17,925 [main] INFO org.apache.hadoop.mapreduce.configuration - mapreduce.job.counters.limit is deprecated
counters.max
2014-12-26 18:08:17,925 [main] INFO org.apache.hadoop.mapreduce.configuration - dfs.permissions is deprecated. Instead, use
2014-12-26 18:08:17,925 [main] INFO org.apache.hadoop.mapreduce.configuration - dfs.permissions is deprecated. Instead, use
```