

Apache Pig - Running Scripts

Word count Program:

To create a data file

```
[cloudera@localhost ~]$ gedit datafile7
```

```
hai hello how are you  
are you fine  
the world is very beautiful  
my name is raj  
i live in bangalore  
i am fine
```

Save & Close

To put this file into hadoop

```
[cloudera@localhost ~]$ hadoop fs -put datafile7 /user/cloudera/
```

To create a pig latin script

```
[cloudera@localhost ~]$ gedit wcount7.pig
```

```
file = load '/user/cloudera/datafile7' as (line:chararray);  
t = foreach file generate flatten(TOKENIZE(line)) as word;  
gr = group t by word;  
wc = foreach gr generate group,COUNT(t.word);  
dump wc;  
store wc into '/user/cloudera/script_o7' using PigStorage();
```

Save & Close

To run the pig latin script

```
[cloudera@localhost ~]$ pig wcount7.pig
```

```
2018-03-28 00:24:21,302 [main] INFO
```

```
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

To check the output

```
[cloudera@localhost ~]$ hadoop fs -ls /user/cloudera/script_o7
```

```
Found 3 items
```

```
-rw-r--r--  3 cloudera cloudera      0 2018-03-28 00:24 /user/cloudera/script_o7/_SUCCESS
drwxr-xr-x  - cloudera cloudera      0 2018-03-28 00:23 /user/cloudera/script_o7/_logs
-rw-r--r--  3 cloudera cloudera    128 2018-03-28 00:24 /user/cloudera/script_o7/part-r-00000
```

```
[cloudera@localhost ~]$ hadoop fs -cat /user/cloudera/script_o7/part*
```

```
i      2
am     1
in     1
is     2
my     1
are    2
hai    1
how    1
raj    1
the    1
you    2
fine   2
live   1
name   1
very   1
hello  1
world  1
bangalore  1
beautiful  1
```

Grep Program:

To create a pig latin script

```
[cloudera@localhost ~]$ gedit grep.pig
```

```
file = load '/user/cloudera/datafile1' as (line:chararray);
token = foreach file generate flatten(TOKENIZE(line)) as word;
t = filter token by word matches 'you';
gr = group t by word;
wc = foreach gr generate group,COUNT(t.word);
dump wc;
store wc into '/user/cloudera/script_o8' using PigStorage();
```

Save & Close

To run the pig latin script

```
[cloudera@localhost ~]$ pig grep.pig
```

```
2018-03-28 00:43:05,939 [main] INFO
```

```
org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
```

To check the output

```
[cloudera@localhost ~]$ hadoop fs -ls /user/cloudera/script_o2
```

```
Found 3 items
```

```
-rw-r--r--  3 cloudera cloudera      0 2018-03-28 00:43 /user/cloudera/script_o8/_SUCCESS
drwxr-xr-x  - cloudera cloudera      0 2018-03-28 00:42 /user/cloudera/script_o8/_logs
-rw-r--r--  3 cloudera cloudera      6 2018-03-28 00:42 /user/cloudera/script_o8/part-r-00000
```

```
[cloudera@localhost ~]$ hadoop fs -cat /user/cloudera/script_o8/part*
```