



9/9/2018

Pig & Spark Outputs

SPARK Project-1

Name: RUKSANA BHANU SHEIK

PIG OUTPUTS

PIG OUTPUT₁: LOOKUP



```
ec2-user@ip-172-31-84-37:~$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig1_single_row_lookup/part*
2,2017-10-01 00:15:30,2017-10-01 00:25:11,1,2.17,1,N,141,142,1,9.0,0.5,0.5,2.06,0.0,0.3,12.36
ec2-user@ip-172-31-84-37:~$
```

```
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig1_single_row_lookup/part*
```

```
2,2017-10-01 00:15:30,2017-10-01 00:25:11,1,2.17,1,N,141,142,1,9.0,0.5,0.5,2.06,0.0,0.3,12.36
```

```
[ec2-user@ip-172-31-84-37 ~]$
```

PIG OUTPUT₂: FILTER



```
ec2-user@ip-172-31-84-37:~$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig2_filter/part* | wc -l
31029
ec2-user@ip-172-31-84-37:~$
```

```
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig2_filter/part* | wc -l
```

```
31029
```

```
[ec2-user@ip-172-31-84-37 ~]$
```

PIG OUTPUT₃: GROUPBY



```
ec2-user@ip-172-31-84-37:~$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig3_groupby_orderby/part*
4,85287
3,292090
2,17648209
1,36492406
[ec2-user@ip-172-31-84-37 ~]$
```

```
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/pig3_groupby_orderby/part*
```

4,85287

3,292090

2,17648209

1,36492406

```
[ec2-user@ip-172-31-84-37 ~]$
```

SPARK OUTPUTS

SPARK OUTPUT₁: LOOKUP



```
login as: ec2-user
Authenticating with public key "imported-openssh-key"
Last login: Sat Sep  6 15:42:13 2014 from 122.167.89.100
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark1_single_row_lookup/part*
2,2017-10-01 00:15:30,2017-10-01 00:25:11,1,2.17,1,N,141,142,1,9,0.5,0.5,2.06,0,0.3,12.36
[ec2-user@ip-172-31-84-37 ~]$
```

```
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark1_single_row_lookup/part*
```

2,2017-10-01 00:15:30,2017-10-01 00:25:11,1,2.17,1,N,141,142,1,9,0.5,0.5,2.06,0,0.3,12.36

```
[ec2-user@ip-172-31-84-37 ~]$
```

SPARK OUTPUT₂: FILTER



```
ec2-user@ip-172-31-84-37:~$  
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark2_filter/part* | wc -l  
31029  
[ec2-user@ip-172-31-84-37 ~]$
```

user/spark_assignment/output/spark2_filter/part* | wc -l

31029

[ec2-user@ip-172-31-84-37 ~]\$

SPARK OUTPUT₃: GROUPBY



```
ec2-user@ip-172-31-84-37:~$  
[ec2-user@ip-172-31-84-37 ~]$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark3_groupby_orderby/part*  
(4,85287)  
(3,292090)  
(2,17648209)  
(1,36492406)  
[ec2-user@ip-172-31-84-37 ~]$
```

[ec2-user@ip-172-31-84-37 ~]\$ hadoop fs -cat /user/ec2-user/spark_assignment/output/spark3_groupby_orderby/part*

(4,85287)

(3,292090)

(2,17648209)

(1,36492406)

[ec2-user@ip-172-31-84-37 ~]\$