

EXERCISE-3

```
grunt> --In MapReduce mode with input file on HDFS:  
grunt> a = LOAD '/ngraminput' as (ngram:chararray,year:int,count:int);  
18/11/10 00:15:29 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS  
grunt>  
grunt> b = FILTER a BY year == 2002;  
grunt> c = ORDER b BY count DESC;  
grunt> d = LIMIT c 1;  
grunt>  
grunt> store d into '/home/hadoop/pig_wordcount_ex3';
```

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.7.1-amzn-0	0.14.0-amzn-0	hadoop	2018-11-10 00:15:29	2018-11-10 00:18:04	ORDER_BY,FILTER,LIMIT

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReducetime
job_1541796622508_0007	13	0	60	42	56	57	0	0	0	a,b
job_1541796622508_0008	1	1	5	5	5	5	4	4	4	c
job_1541796622508_0009	1	1	7	7	7	7	4	4	4	c
job_1541796622508_0010	1	1	4	4	4	4	4	4	4	c

MAP_ONLY
SAMPLER
ORDER_BY,COMBINER
/home/hadoop/pig_wordcount_ex3,

Input(s):
Successfully read 66149971 records (1673556390 bytes) from: "/ngraminput"

Output(s):
Successfully stored 1 records (17 bytes) in: "/home/hadoop/pig_wordcount_ex3"

Counters:
Total records written : 1
Total bytes written : 17
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1541796622508_0007 -> job_1541796622508_0008,
job_1541796622508_0008 -> job_1541796622508_0009,
job_1541796622508_0009 -> job_1541796622508_0010,
job_1541796622508_0010

18/11/10 00:18:04 INFO mapreduce.SimplePigStats: Script Statistics:

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.7.1-amzn-0	0.14.0-amzn-0	hadoop	2018-11-10 00:15:29	2018-11-10 00:18:04	ORDER_BY,FILTER,LIMIT

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReducetime
job_1541796622508_0007	13	0	60	42	56	57	0	0	0	a,b
job_1541796622508_0008	1	1	5	5	5	5	4	4	4	c
job_1541796622508_0009	1	1	7	7	7	7	4	4	4	c
job_1541796622508_0010	1	1	4	4	4	4	4	4	4	c

MAP_ONLY
SAMPLER
ORDER_BY,COMBINER
/home/hadoop/pig_wordcount_ex3,

Input(s):
Successfully read 66149971 records (1673556390 bytes) from: "/ngraminput"

Output(s):
Successfully stored 1 records (17 bytes) in: "/home/hadoop/pig_wordcount_ex3"

Output:

1 .	2002	2438961
-----	------	---------

EXERCISE-4

```
grunt> a = LOAD '/ngraminput' as (ngram:chararray,year:int,count:int);
18/11/10 00:30:15 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS
18/11/10 00:30:16 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS
grunt>
grunt> b = FOREACH a GENERATE ngram, year, count;
2639449 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1200: <line 31, column 14> Syntax error, unexpected symbol at or near 'GERERATE'
18/11/10 00:30:16 ERROR grunt.Grunt: ERROR 1200: <line 31, column 14> Syntax error, unexpected symbol at or near 'GERERATE'
Details at logfile: /mnt/var/log/pig/pig_1541807176772.log
grunt> c = GROUP b BY year;
2639453 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1200: <line 31, column 0> Syntax error, unexpected symbol at or near 'c'
18/11/10 00:30:16 ERROR grunt.Grunt: ERROR 1200: <line 31, column 0> Syntax error, unexpected symbol at or near 'c'
Details at logfile: /mnt/var/log/pig/pig_1541807176772.log
grunt> d = FOREACH c {sorted = ORDER b BY count DESC; limited = LIMIT sorted 1; GENERATE FLATTEN(limited)};
2639569 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1200: <line 31, column 30> expression is not a project expression
18/11/10 00:30:16 ERROR grunt.Grunt: ERROR 1200: Pig script has errors
<line 31, column 30> expression is not a project expression
Details at logfile: /mnt/var/log/pig/pig_1541807176772
grunt>
[grunt> store d into '/home/hadoop/pig_wordcount_ex4';
```

```
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs
job_1541796622508_0011 13 0 57 39 53 54 0 0 0 a,b MAP_ONLY
job_1541796622508_0012 1 1 5 5 5 5 4 4 4 4 c SAMPLER
job_1541796622508_0013 1 1 7 7 7 7 4 4 4 4 c ORDER_BY_COMBINER
job_1541796622508_0014 1 1 4 4 4 4 4 4 4 4 c /home/hadoop/pig_wordcount_ex4

Input(s):
Successfully read 66149971 records (1673556390 bytes) from: "/ngraminput"

Output(s):
Successfully stored 1 records (17 bytes) in: "/home/hadoop/pig_wordcount_ex4"

Counters:
Total records written : 1
Total bytes written : 17
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1541796622508_0011 --> job_1541796622508_0012,
job_1541796622508_0012 --> job_1541796622508_0013,
job_1541796622508_0013 --> job_1541796622508_0014,
job_1541796622508_0014

18/11/10 00:32:46 INFO mapreduce.SimplePigStats: Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.7.1-amzn-0 0.14.0-amzn-0 hadoop 2018-11-10 00:30:17 2018-11-10 00:32:46 ORDER_BY,FILTER,LIMIT

Success!

Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs
job_1541796622508_0011 13 0 57 39 53 54 0 0 0 a,b MAP_ONLY
job_1541796622508_0012 1 1 5 5 5 5 4 4 4 4 c SAMPLER
job_1541796622508_0013 1 1 7 7 7 7 4 4 4 4 c ORDER_BY_COMBINER
job_1541796622508_0014 1 1 4 4 4 4 4 4 4 4 c /home/hadoop/pig_wordcount_ex4

Input(s):
Successfully read 66149971 records (1673556390 bytes) from: "/ngraminput"

Output(s):
Successfully stored 1 records (17 bytes) in: "/home/hadoop/pig_wordcount_ex4"

Counters:
Total records written : 1
Total bytes written : 17
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
```

Output:

Book -	1574	139
1 .	1575	153
[In	1576	8
A .	1586	43
"rage	1588	8
' ry	1602	3
1 .	1603	36
And now	1604	3
shall not	1612	22
"Bermudas	1613	6
the primaries	1614	53
"employment	1627	3
1 .	1630	49
City .	1638	42
of national	1639	7
"	1640	14
"true	1641	3
1 .	1642	43
the signatures	1643	3
1 .	1651	124
And now	1653	21
the year	1654	18
"done	1664	39
A .	1665	114
"done	1666	81
"done	1667	78
"done	1668	34
"place	1669	72
cannot be	1677	51
"done	1678	15
"true	1679	21
cannot be	1680	34
1 .	1681	242
1 .	1682	401
"	1690	42
1 .	1692	105
"Philadelphia	1693	411
cannot be	1694	144
cannot be	1695	48
"done	1703	331
cannot be	1704	133
cannot be	1705	280
"true	1706	158
cannot be	1707	167
1 .	1708	349
cannot be	1716	159
' ry	1717	293
cannot be	1718	195
on shore	1719	453
cannot be	1720	225
cannot be	1721	226
cannot be	1729	526
0 .	1730	686
cannot be	1731	595
of Great	1732	1318
cannot be	1733	236
cannot be	1734	502
the air	1742	1440
cannot be	1743	676
cannot be	1744	883
cannot be	1745	741
cannot be	1746	443
1 .	1747	5750
cannot be	1755	1272
"Rome	1756	511
cannot be	1757	1307

. Hilary	1541	2
the year	1581	14
Rom .	1584	120
"	1592	144
"	1593	9
lye .	1596	2
"place	1608	6
"" of"	1618	14
"	1620	58
by her	1621	3
"alfo	1631	134
A .	1632	10
"place	1634	16
"Rome	1635	45
cannot be	1644	166
. Secondly	1647	89
the empire	1648	139
- house	1649	82
cannot be	1658	14
ing .	1659	2
1 .	1660	388
"done	1661	28
"place	1670	124
1 .	1671	209
1 .	1672	24
"place	1673	117
1 .	1674	735
by her	1675	95
1 .	1683	733
the year	1684	70
"alfo	1685	131
"Rome	1686	78
"Council	1687	113
cannot be	1688	157
1 .	1696	23
"place	1697	116
A .	1698	172
1 .	1699	91
cannot be	1700	231
1 .	1701	329
"true	1709	46
cannot be	1710	265
"Rome	1711	252
1 .	1712	72
cannot be	1713	132
"Rome	1714	257
cannot be	1722	246
the Mirza	1723	448
"	1724	375
"true	1725	313
cannot be	1726	493
cannot be	1727	704
A .	1735	651
cannot be	1736	492
cannot be	1737	462
cannot be	1738	451
"Rome	1739	1193
cannot be	1740	626
cannot be	1748	1045
1 .	1749	861
"Rome	1750	1657
cannot be	1751	1084
cannot be	1752	824
cannot be	1753	818
cannot be	1761	955
	1873	76635
"	1874	80076
"	1875	86340
"	1876	82555
"	1877	81858
"	1885	102023
"	1886	100278
"	1887	98829
"	1888	95988
"	1889	92736
"	1890	91038
"	1898	118425
1 .	1899	130921
"	1900	133433
"	1901	136853
1 .	1902	145878
"	1903	149827
1 .	1911	162717
1 .	1912	171872
1 .	1913	179910
1 .	1914	180261
1 .	1915	172815
1 .	1916	171664
1 .	1924	137143
1 .	1925	147549
1 .	1926	154667
1 .	1927	158860
1 .	1928	172232
1 .	1929	156530
1 .	1937	172880
1 .	1938	176652
1 .	1939	185750
1 .	1940	185913
1 .	1941	164449
1 .	1942	182255
1 .	1950	294184
1 .	1951	274962
1 .	1952	296970
1 .	1953	295762
1 .	1954	293544
1 .	1955	322510
1 .	1963	614019
1 .	1964	551242
1 .	1965	624613
1 .	1966	634940
1 .	1967	669708
1 .	1968	696429
1 .	1976	805146
1 .	1977	814441
1 .	1978	829113
1 .	1979	903219
1 .	1980	887364
1 .	1981	845599
1 .	1989	1152428
1 .	1990	1279262
1 .	1991	1211158
1 .	1992	1229918
1 .	1993	1264221
1 .	1994	1303655
1 .	2002	2438961
1 .	2003	2791900
1 .	2004	2828348
1 .	2005	2632338
1 .	2006	2582425
1 .	2007	2473654