**4. Top / Bottom 5: catagery-1/ catagery-2 / page /users (Exclude status code other than 200, also exclude record related to css/js/image)**

**Top 5 and bottom 5 records**

**Category, total\_number\_views**

**page, total\_number\_views**

**IP, total\_number\_of\_views**

**TOP 5 records by Category1:-**

**Filter the records in File1 having status code as 200:-**

**Command:-**

File1\_status\_code\_200 = FILTER File1 by code == 200;

**Group the filtered records by category1**

**Command:-**

GRPD\_Category1 = group File1\_status\_code\_200 by Category1;

**Generate count for each category1:-**

**Command:-**

CNT\_FOR\_Category1 = foreach GRPD\_Category1 generate group,COUNT(File1\_status\_code\_200.Category1) as counting;

**Sort the results from above step in descending order .**

**Command:-**

SORTED\_CATEGORY1\_DEC = ORDER CNT\_FOR\_Category1 by counting DESC;

**Limit the sorted output to 5 in order to obtain top 5 category1:-**

**Command:-**

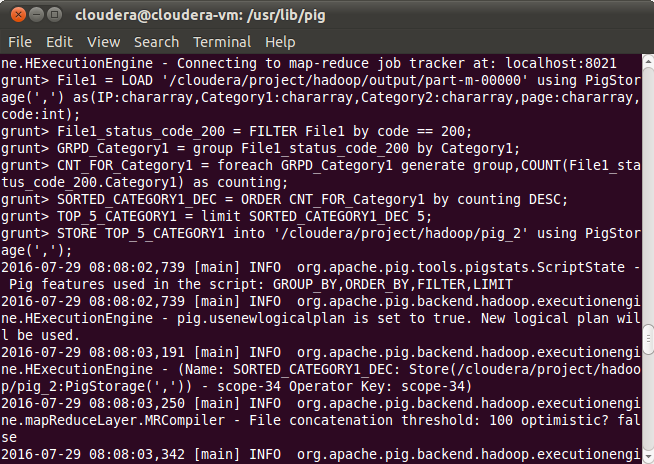
TOP\_5\_CATEGORY1 = limit SORTED\_CATEGORY1\_DEC 5;

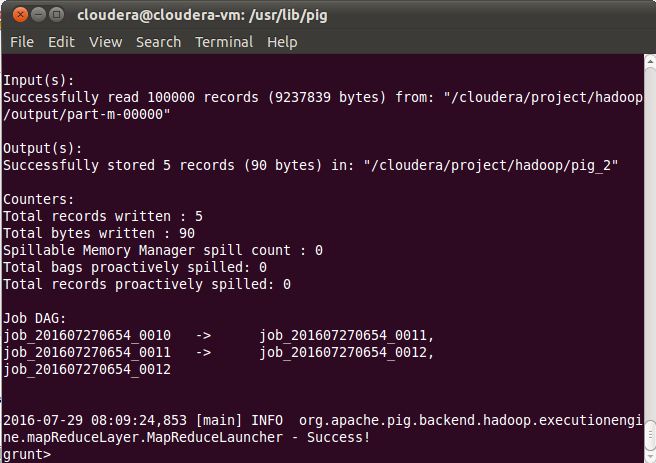
**Store the output to HDFS:-**

**Command:-**

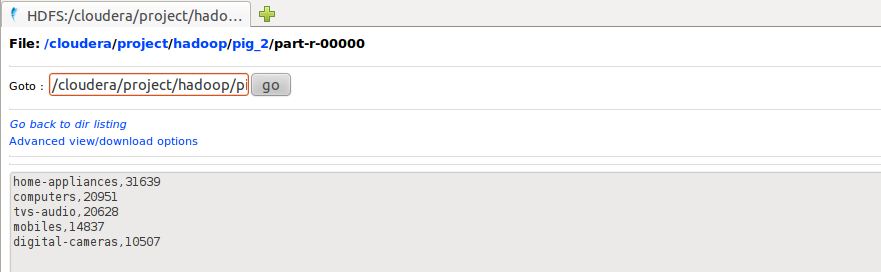
STORE TOP\_5\_CATEGORY1 into ‘/cloudera/project/hadoop/pig\_2’ using PigStorage(‘,’);

**Screenshot:-**





**O\P:- Top 5 records by category1:-**



**BOTTOM 5 records by Category1:-**

**Sort the CNT\_FOR\_Category1 in ascending order:-**

**Command:-**

SORTED\_CATEGORY1\_ASC = ORDER CNT\_FOR\_Category1 by counting ASC;

**Limit the sorted output to 5 in order to obtain top 5 category1:-**

**Command:-**

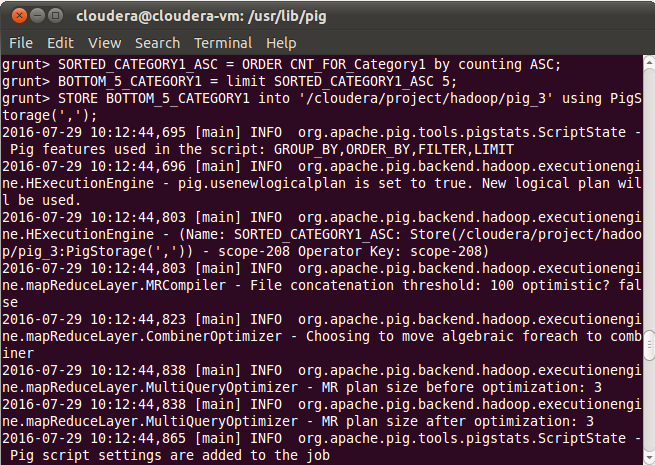
BOTTOM\_5\_CATEGORY1 = limit SORTED\_CATEGORY1\_ASC 5;

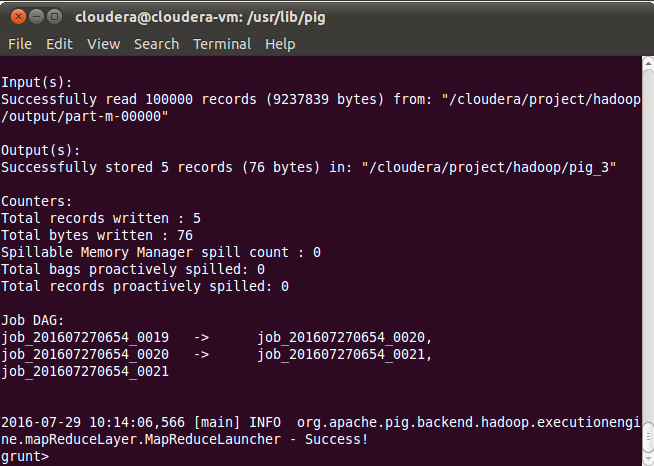
**Store the output to HDFS:-**

**Command:-**

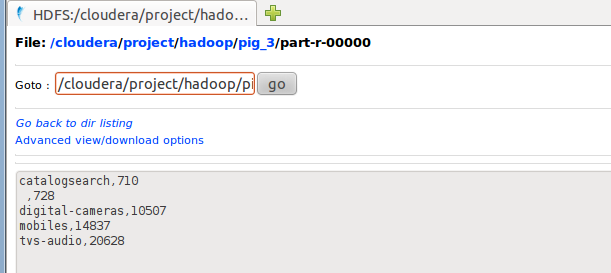
STORE BOTTOM\_5\_CATEGORY1 into ‘/cloudera/project/hadoop/pig\_3’ using PigStorage(‘,’);

**Screenshot:-**





**O\P:- Bottom 5 records by category1:-**



**Category2**

**Top 5 records by Category2**

**Filter the records in File1 having status code as 200:-**

**Command:-**

File1\_status\_code\_200 = FILTER File1 by code == 200;

**Group the filtered records by category2**

**Command:-**

GRPD\_Category2 = group File1\_status\_code\_200 by Category2;

**Generate count for each category1:-**

**Command:-**

CNT\_FOR\_Category2 = foreach GRPD\_Category2 generate group,COUNT(File1\_status\_code\_200.Category2) as counting;

**Sort the results from above step in descending order.**

**Command:-**

SORTED\_CATEGORY2\_DEC = ORDER CNT\_FOR\_Category2 by counting DESC;

**Limit the sorted output to 5 in order to obtain top 5 category1:-**

**Command:-**

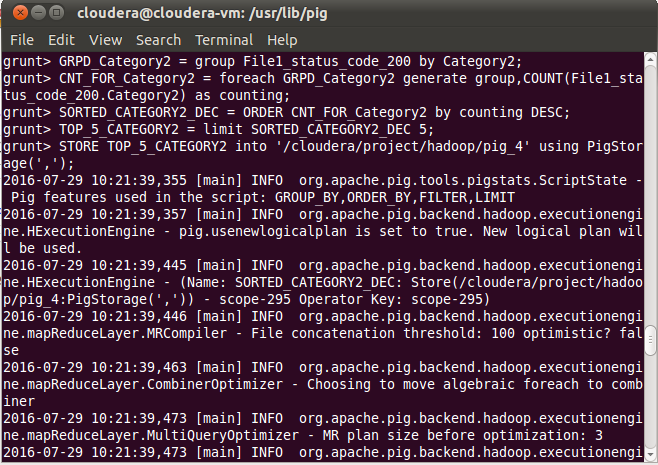
TOP\_5\_CATEGORY2 = limit SORTED\_CATEGORY2\_DEC 5;

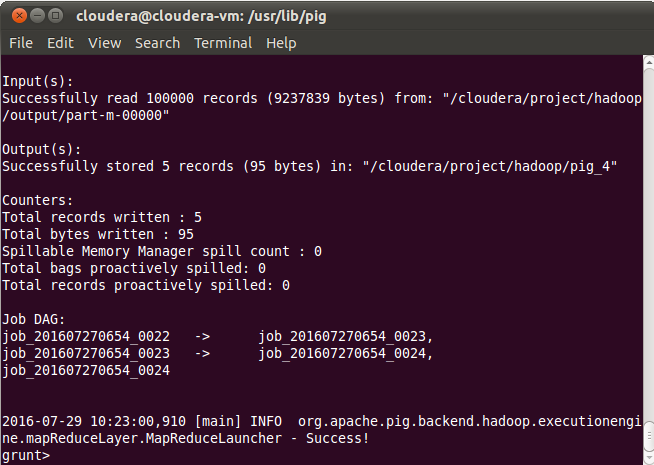
**Store the output to HDFS:-**

**Command:-**

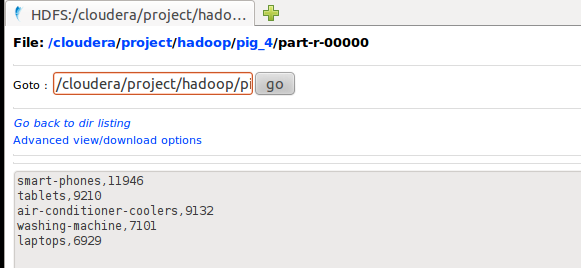
STORE TOP\_5\_CATEGORY2 into ‘/cloudera/project/hadoop/pig\_4’ using PigStorage(‘,’);

**Screenshot:-**





**O\P:-Top 5 records by category2:-**



**BOTTOM 5 records by Category2:-**

**Sort the CNT\_FOR\_Category2 in ascending order:-**

**Command:-**

SORTED\_CATEGORY2\_ASC = ORDER CNT\_FOR\_Category2 by counting ASC;

**Limit the sorted output to 5 in order to obtain top 5 category1:-**

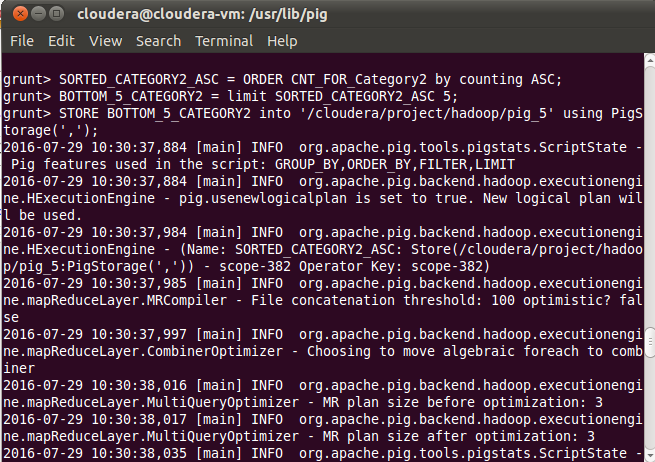
**Command:-**

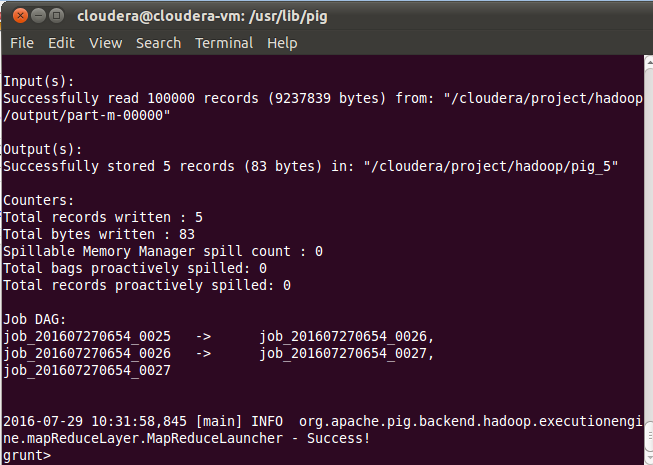
BOTTOM\_5\_CATEGORY2 = limit SORTED\_CATEGORY2\_ASC 5;

**Store the output to HDFS:-**

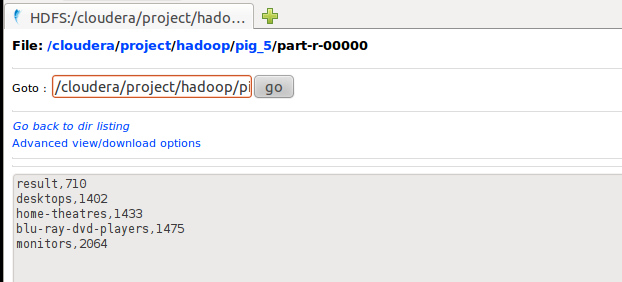
**Command:-**

STORE BOTTOM\_5\_CATEGORY2 into ‘/cloudera/project/hadoop/pig\_5’ using PigStorage(‘,’);





**O\P:-Bottom 5 records by category2**



**TOP 5 records by Page:-**

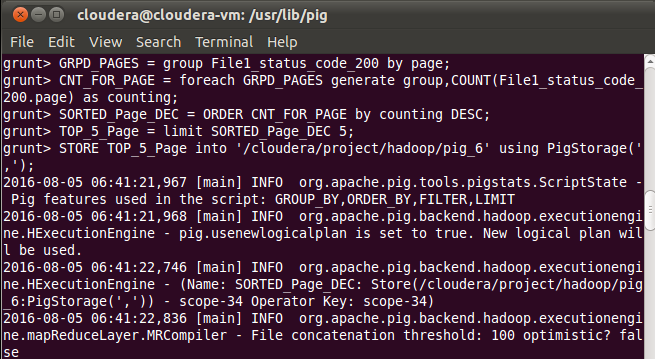
GRPD\_PAGES = group File1\_Sstatus\_code\_200 by page;

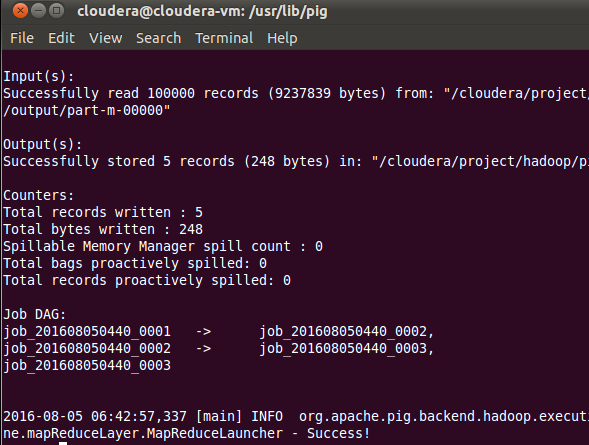
CNT\_FOR\_PAGE = foreach GRPD\_PAGES generate group,COUNT(File1\_status\_code\_200.page) as counting;

SORTED\_PAGE\_DEC = ORDER CNT\_FOR\_PAGE by counting DESC;

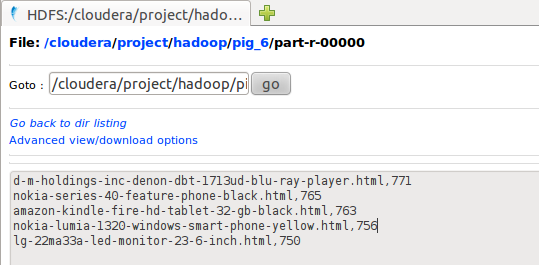
TOP\_5\_Page = limit SORTED\_PAGE\_DEC 5;

STORE TOP\_5\_Page into ‘/cloudera/project/hadoop/pig\_6’ using PigStorage(‘,’);





**Output:-TOP 5 Page count:-**

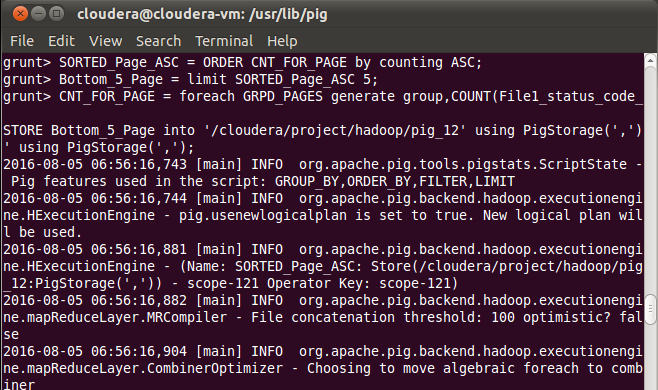


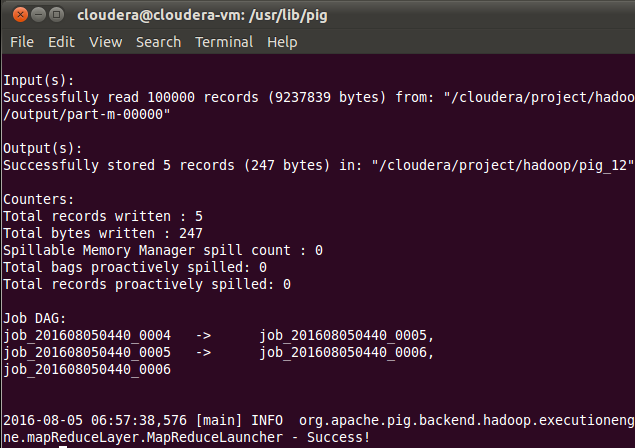
Bottom 5 page:-

SORTED\_PAGE\_ASC = ORDER CNT\_FOR\_PAGE by counting ASC;

Bottom\_5\_Page = limit SORTED\_PAGE\_ASC 5;

STORE Bottom\_5\_Page into ‘/cloudera/project/hadoop/pig\_12’ using PigStorage(‘,’);





**Output:-Bottom 5 pages:-**

