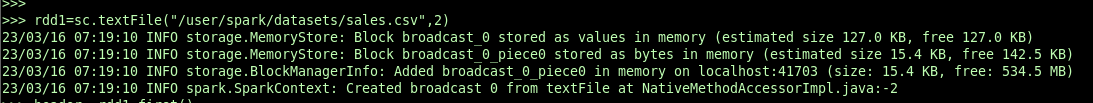
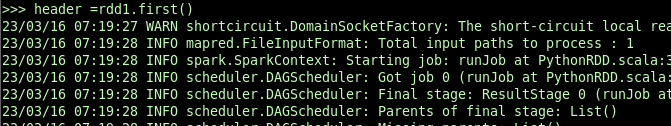
-------------------------------------------Pyspark Project-----------------------

Loading data from hdfs to spark rdd

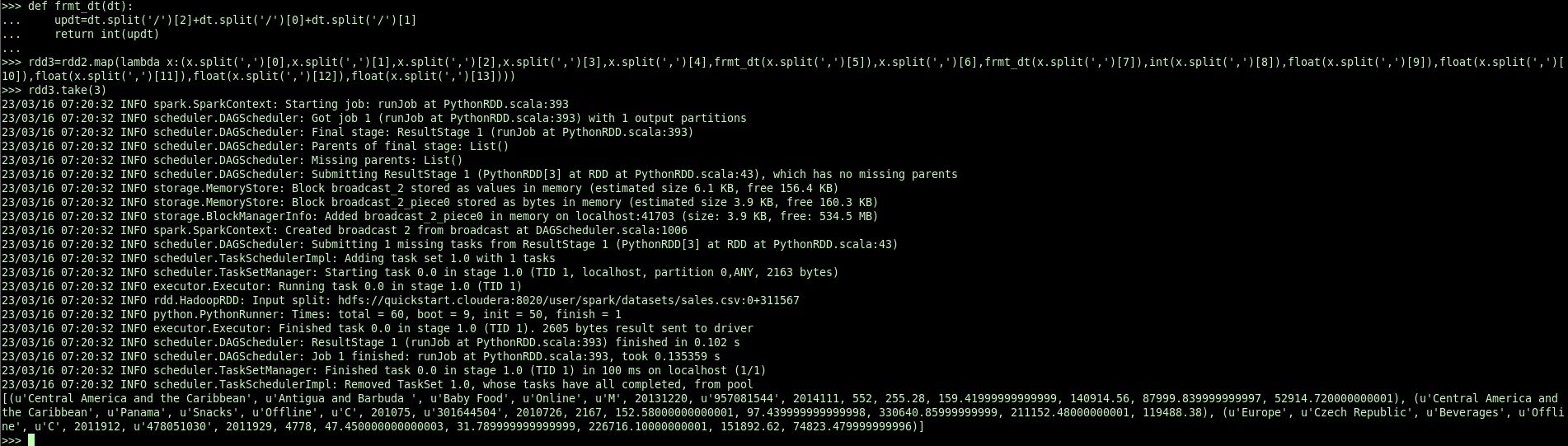


Removing header

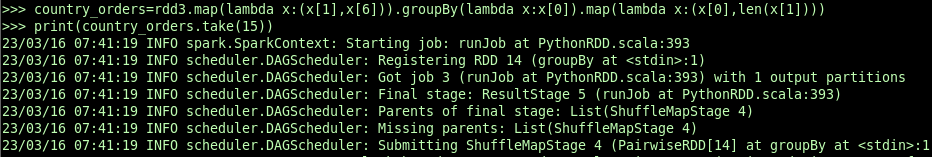


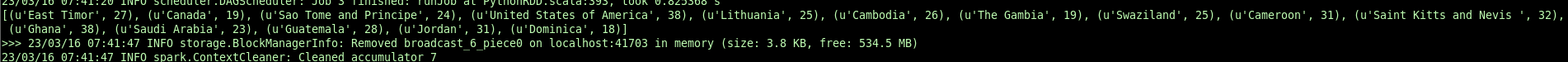


Type casting the required columns and getting date to int format of yyyymmdd

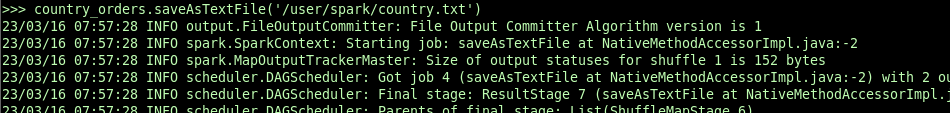


1.country wise number of orders



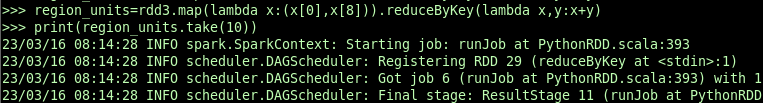


Saving the result rdd to hdfs locations



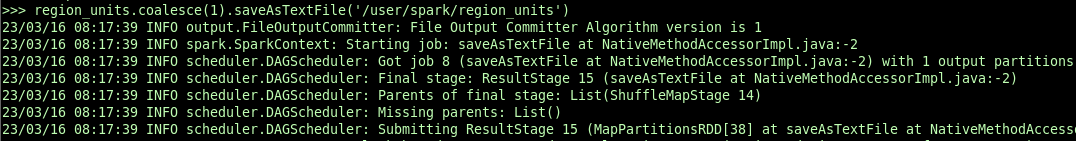


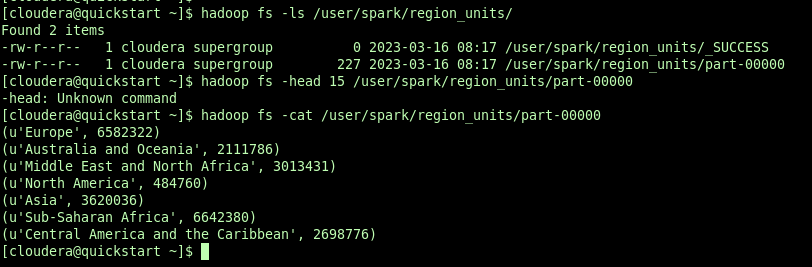
2.Display the number of units sold in each region.



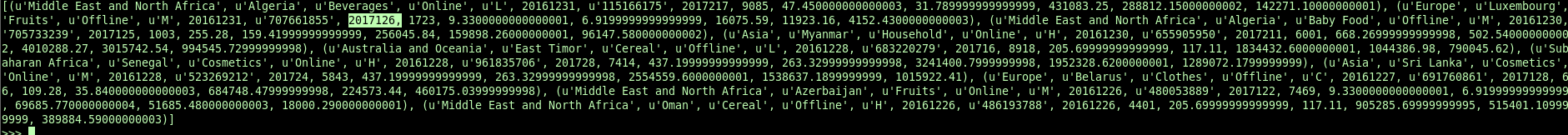


Saving result to hdfs location as single partitition

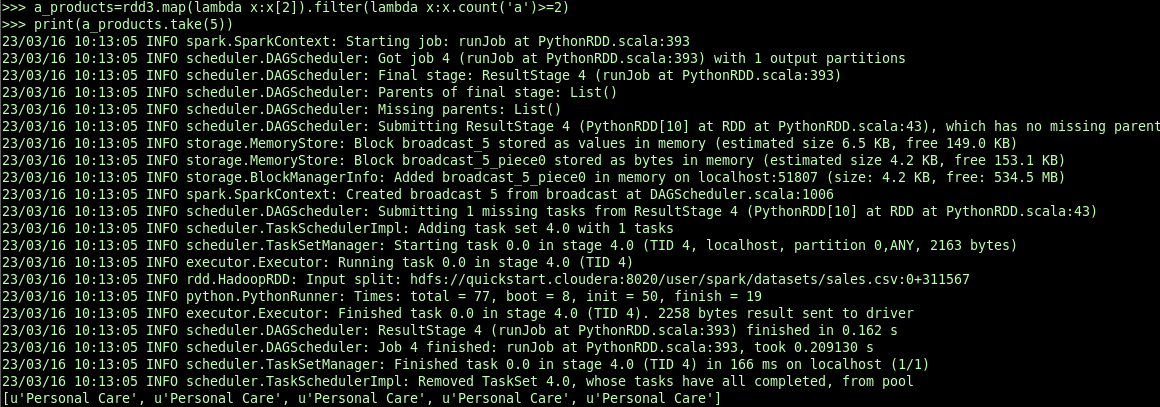




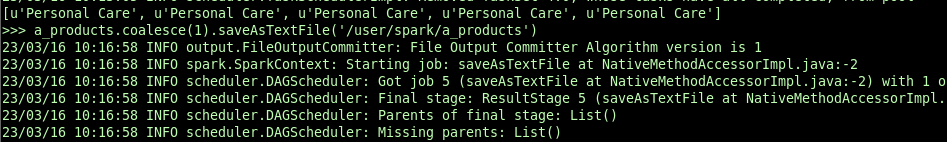
3.Display the 10 most recent sales.

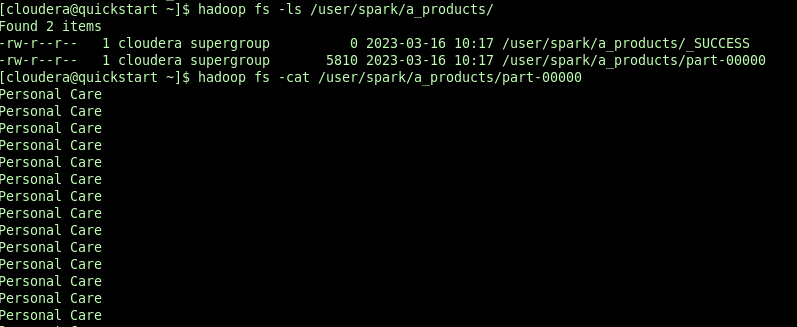


4.Display the products with atleast 2 occurences of 'a'



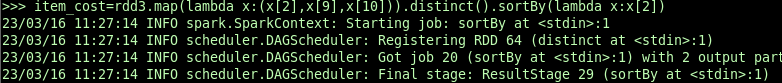
Saving results as 1 partition textfile in hdfs





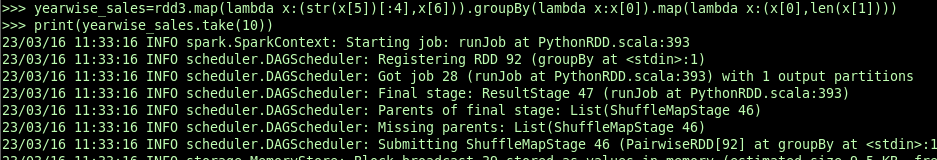
5.Display country in each region with highest units sold.

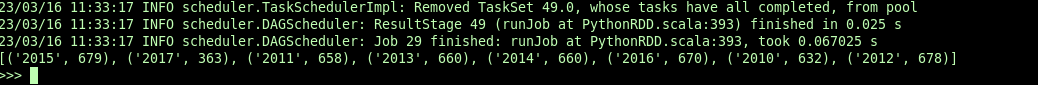
6.Display the unit price and unit cost of each item in ascending order.



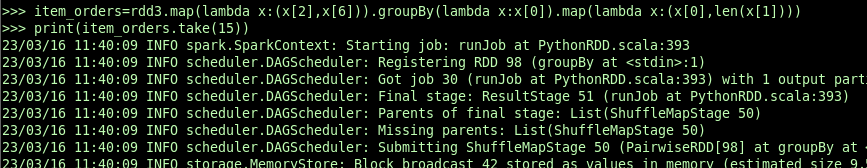


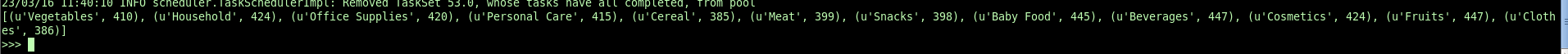
7.Display the number of sales yearwise.





8.Display the number of orders for each item.





Saved results in hdfs in text format

