1. Output the full names of the Top 5 Salespeople

sales\_raw = LOAD 's3://data-eng-resources/big-data/adventureworks/salespeople.csv' USING CSVLoader AS (BusinessEntityID:int,TerritoryID:int,SalesQuota:int,Bonus:int,CommissionPct:float,SalesYTD:float,SalesLastYear:float,rowguid:chararray,ModifiedDate:datetime);

people\_raw = LOAD 's3://data-eng-resources/big-data/adventureworks/people.csv' USING CSVLoader AS (BusinessEntityID:int,PersonType:chararray,NameStyle:int,Title:chararray,FirstName:chararray,MiddleName:chararray,LastName:chararray,Suffix:chararray,EmailPromotion:int,AdditionalContactInfo:chararray,Demographics:chararray,rowguid:chararray,ModifiedDate:datetime);

sales\_be = FOREACH sales\_raw GENERATE BusinessEntityID, SalesYTD;

people\_be = FOREACH people\_raw GENERATE BusinessEntityID, FirstName, MiddleName, LastName;

sp\_join = JOIN sales\_be BY BusinessEntityID, people\_be BY BusinessEntityID;

data1 = FOREACH sp\_join GENERATE CONCAT( people\_be::FirstName, ' ', people\_be::MiddleName, ' ', people\_be::LastName) AS name, sales\_be::SalesYTD;

order\_data1 = ORDER data1 BY sales\_be::SalesYTD DESC;

limit\_data1 = LIMIT order\_data1 5;

OUTPUT:

(Linda C Mitchell,4251368.5)

(Jae B Pak,4116871.2)

(Michael G Blythe,3763178.2)

(Jillian NULL Carson,3189418.2)

(Ranjit R Varkey Chudukatil,3121616.2)

2. Find the average salesperson's sales (year to date) per territory name. Also provide the count of salespeople per territory.

terr\_raw = LOAD 's3://data-eng-resources/big-data/adventureworks/territories.csv' USING CSVLoader AS (TerritoryID:int,Name:chararray,CountryRegionCode:chararray,Group:chararray,SalesYTD:int,SalesLastYear::int,CostYTD:int,CostLastYear:int,rowguid:chararray,ModifiedDate:datetime);

st\_join = JOIN sales\_raw BY TerritoryID, terr\_raw BY TerritoryID;

st\_group = GROUP st\_join BY terr\_raw::Name;

avg\_std = FOREACH st\_group GENERATE st\_join.terr\_raw::Name, AVG(st\_join.sales\_raw::SalesYTD), COUNT(st\_join.sales\_raw::SalesYTD);

OUTPUT:

({(Canada),(Canada)},2029130.125,2)

({(France)},3121616.25,1)

({(Central)},3189418.25,1)

({(Germany)},1827066.75,1)

({(Australia)},1421810.875,1)

({(Northeast)},3763178.25,1)

({(Northwest),(Northwest),(Northwest)},1500717.4583333333,3)

({(Southeast)},2315185.5,1)

({(Southwest),(Southwest)},3354952.0,2)

({(United Kingdom)},4116871.25,1)

3. Output a file containing a list of all names of employees as initials and surname (e.g. "D. R. Harvey")

names = FOREACH people\_raw GENERATE CONCAT(SUBSTRING(FirstName, 0, 1), '. ', SUBSTRING(MiddleName, 0, 1), '. ', LastName) AS Initals, PersonType;

names\_em = FILTER names BY PersonType=='EM';

names\_final = FOREACH names\_em GENERATE Initals;

STORE names\_final INTO 'csvoutput' USING PigStorage('\t', '-schema');

