

-- for customer.tbl

customer = LOAD '/user/bigdata/customer.tbl' using PigStorage('|')

as

(CUSTOMER_ID:int,CUSTOMER_CODE:chararray,COMPANY_NAME:chararray,CONTACT_NAME:chararray,CONTACT_TITLE:chararray,CITY:chararray,REGION:chararray,POSTAL_CODE:int,COUNTRY:chararray,PHONE:int,FAX:int);

-- for order_detail.tbl

order_detail = LOAD '/user/bigdata/order_detail.tbl' using PigStorage('|')

as

(ORDER_ID:int,PRODUCT_ID:int,CUSTOMER_ID:int,SALESPERSON_ID:int,UNIT_PRICE:float,QUANTITY:int,DISCOUNT:float);

-- for product.tbl

product = LOAD '/user/bigdata/product.tbl' using PigStorage('|')

as

(PRODUCT_ID:int,PRODUCT_NAME:chararray,UNIT_PRICE:float,UNITS_IN_STOCK:int,UNITS_ON_ORDER:int,DISCONTINUED:chararray);

-- for salesperson.tbl

salesperson = LOAD '/user/bigdata/salesperson.tbl' using PigStorage('|')

as

(EMPLOYEE_ID:int,LASTNAME:chararray,FIRSTNAME:chararray,TITLE:chararray,BIRTHDATE:chararray,HIREDATE:chararray,NOTES:chararray);

```

-- for order.tbl

orders = LOAD '/user/bigdata/order.tbl' using PigStorage('|')

as

(ORDER_ID:int,ORDER_DATE:chararray,SHIP_VIA:int,SHIP_CITY:chararray,SHIP_REGION:chararray,SHIP_
POSTAL_CODE:chararray,SHIP_COUNTRY:chararray);

-- (1)

london = FILTER orders BY SHIP_CITY == 'London ';
londonGroup = Group london All;
londonCount = foreach londonGroup Generate COUNT(london);
dump londonCount;

(33)

-- (2)

orderJoin = join orders by ORDER_ID, order_detail by ORDER_ID;
not1996 = filter orderJoin by not(ORDER_DATE matches '.*1996.*');
getProductID = foreach not1996 generate order_detail::PRODUCT_ID;
distinctProduct = distinct getProductID;
productGroup = Group distinctProduct All;
productCount = foreach productGroup Generate COUNT(distinctProduct);
dump productCount;

(77)

-- (3)

```

```

orderFields = foreach order_detail generate ORDER_ID, UNIT_PRICE * QUANTITY as value;
distinctOrder = distinct orderFields;
filterOrderID = FILTER distinctOrder BY (ORDER_ID > 10269 AND ORDER_ID < 10280);
orderGroup = group filterOrderID by (ORDER_ID);
dump orderGroup;

orderSum = foreach orderGroup generate (filterOrderID.ORDER_ID), SUM(filterOrderID.value) as
ordervalue;

dump orderSum;

```

```

({(10270),(10270)},1376.0) ({(10271)},48.0) ({(10272),(10272),(10272)},1455.9999694824219)
({(10273),(10273),(10273),(10273),(10273)},2142.399932861328)
({(10274),(10274)},538.5999908447266) ({(10275),(10275)},307.1999969482422)
({(10276),(10276)},420.0) ({(10277),(10277)},1200.8000183105469)
({(10278),(10278),(10278),(10278)},1488.7999877929688) ({(10279)},468.0)

```

```
-- (4)
```

```

getDetails = foreach order_detail generate SALESPERSON_ID, UNIT_PRICE, QUANTITY, UNIT_PRICE *
QUANTITY as ordervalue;

groupDetails = group getDetails by SALESPERSON_ID;

totalSales = foreach groupDetails generate group, SUM(getDetails.ordervalue) as totalOrderValue;

orderTotalSales = order totalSales by totalOrderValue desc;

top3 = limit orderTotalSales 3;

employeeTop3 = join top3 by group left outer, salesperson by EMPLOYEE_ID;

getTop3 = foreach employeeTop3 generate EMPLOYEE_ID, FIRSTNAME, LASTNAME;

dump getTop3;

```

```
(1, Nancy , Davolio )
```

```
(3, Janet , Leverling )
```

```
(4, Margaret , Peacock )
```

-- (5)

odxProduct = join order_detail by PRODUCT_ID, product by PRODUCT_ID;

filterTofulkura = filter odxProduct BY PRODUCT_NAME == 'Tofu ' OR PRODUCT_NAME == 'Ikura ';

distinctTofulkura = distinct filterTofulkura;

TofulkuraGroup = Group distinctTofulkura All;

totalCount = foreach TofulkuraGroup Generate COUNT(distinctTofulkura);

dump totalCount;

(55)