**Solution 1 -> Daily Show Use Case**

**A** = **load** '/dialy\_shows' **using** PigStorage(',') **AS** (**year**:chararray,occupation:chararray,**date**:chararray,**group**:chararray,gusetlist: chararray);

B = foreach **A** generate occupation,**date**; **C** = foreach B generate occupation,ToDate(**date**,'MM/dd/yy') **as date**;

D = **filter C by** ((**date**>ToDate('1/11/99','MM/dd/yy')) **AND** (**date**<ToDate('6/11/99','MM/dd/yy'))); #**Date range** can be modified **by the user**

E = **group** D **by** occupation; F = foreach E generate **group**, **COUNT**(D) **as** cnt;

G = **order** F **by** cntdesc;

H = **limit** G 5;

*• In relation* ***A****, we are loading the dataset using PigStorage along with the schema of the file.*

*• In relation* ***B****, we are extracting the required columns i.e.,* ***occupation*** *and* ***date****.*

*• In relation* ***C****, we are converting the* ***date*** *in string format to date using* ***ToDate****function in Pig.*

*• In relation* ***D****, we are filtering the dates in a specific range. - Here, we have given the date range from 1/11/99 to 6/11/99 i.e., we are taking out the data for 6*

*months.*

*• In relation* ***E****, we are grouping relation* ***D*** *by occupation. - If you describe relation E,then you can see the schema of the relation as shown below:*

**describe** E;

E: {**group**: chararray,D: {(occupation: chararray,**date**:datetime)}}

*• In relation* ***F****, we are generating the group and the Count of values. Here, we will get the occupation of the guest and the number of times that occupation guest came to the show within this span of 6 months.*

*• In relation* ***G****, we are ordering the relation* ***F*** *by descending order.*

*• In relation* ***H****, we are limiting the records of relation* ***G*** *to 5.*

*• With this, we will get the top five GoogleKnowlege\_Occupation guests in the show in a particular period.*

**Solution 2 -> Daily Show Use Case**

**A** = **load** '/dialy\_shows' **using** PigStorage(',') **AS** (**year**:chararray,occupation:chararray,**date**:chararray,**group**:chararray,gusetlist: chararray);

**B** = foreach **A** generate **year**,**group**;

**C** = **filter** B **by group** == 'Politician';

**D** = **group C by year**;

**E** = foreach D generate **group**, **COUNT**(**C**) **as** cnt;

**F** = **order** E **by** cntdesc;

*• In relation* ***A****, we are loading the dataset using PigStorage along with the schema of the file.*

*• In relation* ***B****, we are extracting the required columns i.e., year and the group.*

*• In relation* ***C****, we are filtering the group by Politician.*

*• In relation* ***D****, we are grouping the relation* ***C*** *by year.*

*- If you describe relation then you can see the schema of the relation as shown below:*

**describe** D;

D: {**group**: chararray,C: {(**year**: chararray,**group**: chararray)}}

• *In relation* ***E****, we are generating the group and the Count of values in the relation* ***C****.*

*• In relation* ***F****, we are ordering the values in the relation* ***F*** *by descending order.*