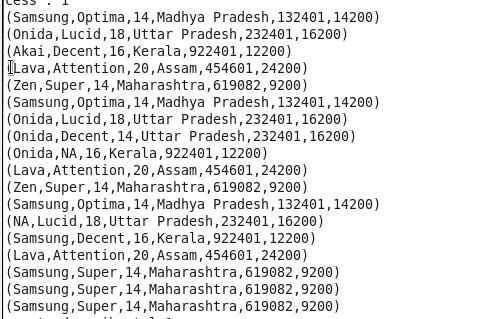
**Create a sample dataset and implement the below Pig commands on the same dataset.**

1. **Concat:-- Concatenates two expressions of identical type**

**CONCAT (expression, expression)**

**CONCAT**

****

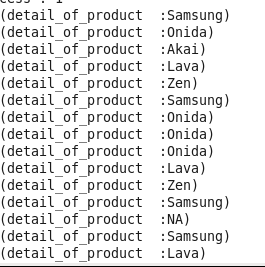
**After desc**

****

**After we concatenate**

****

**After which result will be**

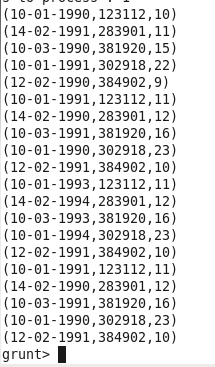
****

**2)Tokenize: Here we are having a data set .**

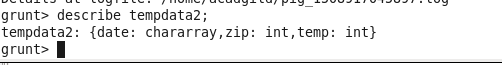
**TOKENIZE**

**Splits a string and outputs a bag of words**

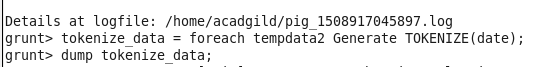
**TOKENIZE(expression [, ‘field\_delimiter’])**

****

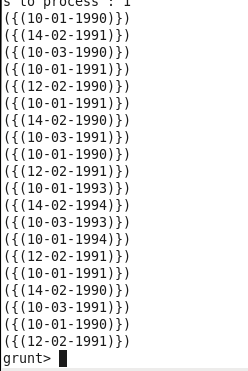
**After we describe the dataset we come to now about the schema of dataset.**

****

**Here we are using tokenize command for display of only date.**

****

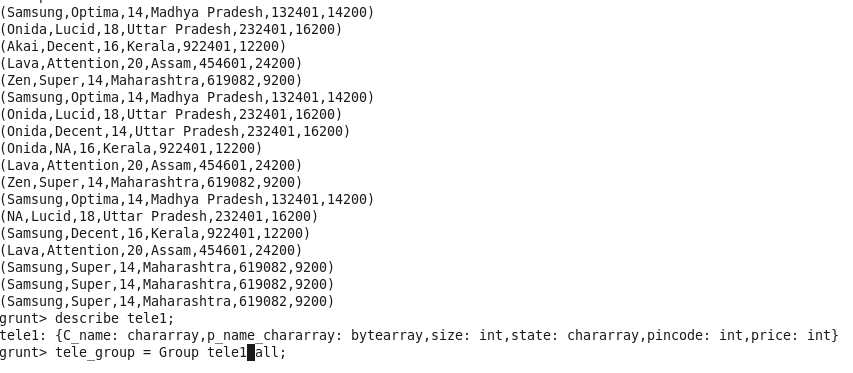
**Output will be**

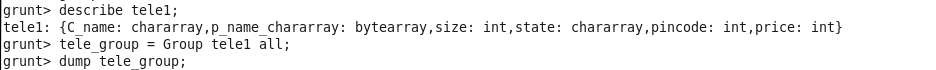
****

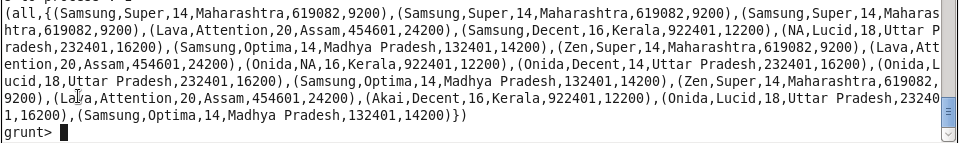
**3)Sum :- SUM**

**SUM(expression)**

**Computes the sum of the numeric values in a single-column bag**

****

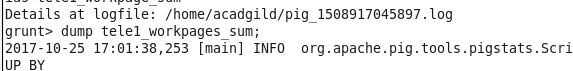
****

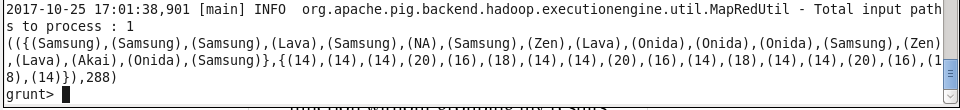
****

**We are using the sum command here**

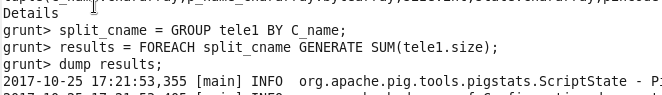
****

**Here after that will get the output**

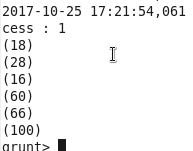
****

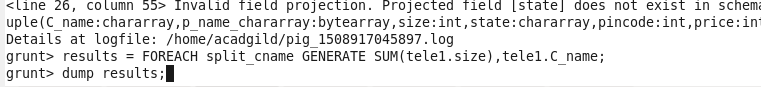
****

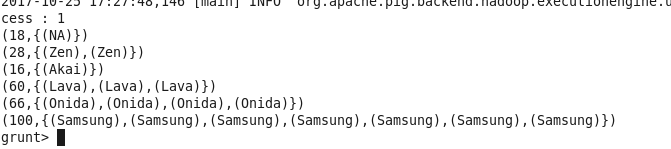
**Sum is processed using GROUP BY**

****

**And the output will belike.**

****

****

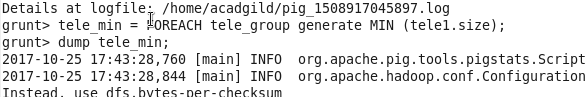
****

**4)MIN: MIN**

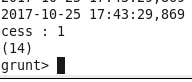
**MIN(expression)**

**Computes the minimum of the numeric values or chararrays in a single-column bag.**

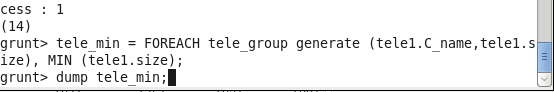
**Here on above data set for each tele\_group we have created we will find the minimum size of television using the below syntax.**

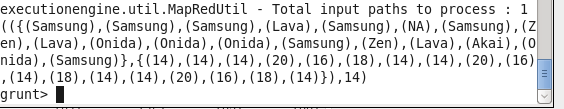
****

**Output will be as given below**

****

**Can use it with groupby also as we have used in below examole**

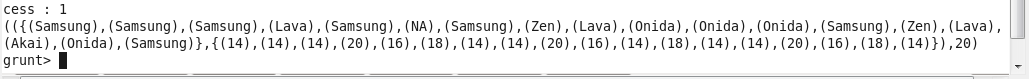
****

****

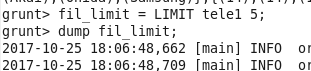
**5)MAX -- MAX(expression)**

**Computes the maximum of the numeric values or chararrays in a single-column bag**

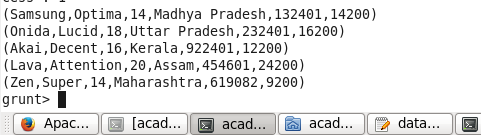
****

****

**6)Limit:-- limit is used to limit the output of some query**

****

**Here we are limiting the output of tele1 to only 5 so only 5 output values will be shown.**

****

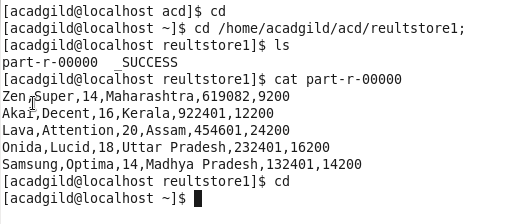
**7)Store: Store is used to store the output of the query we are running into some location, so as to used afterwards ,because if we are not storing the data results then once we come out of hdfs shell data will get removed.**

**Here in below example we are storing the result of above query into the locatin /home/acadgild/acd/resultstore1**

****

****

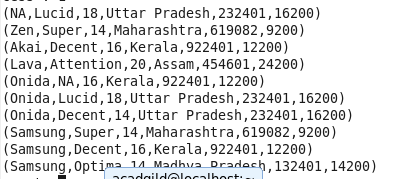
**Can see the result using the location of result followed by –cat command.**

****

**8)Distinct:--Distinct is used to get the unique value ,i.e if some values are oming again anad again then it will show that only once.**

****

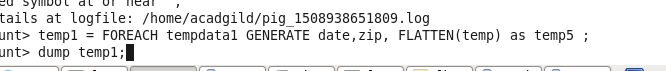
**Here we are showing that we want only unique or distinct value from tele1 so only once occuranc will be shown here.**

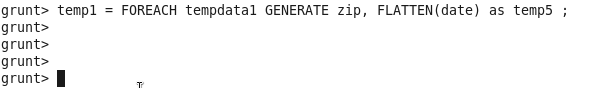
****

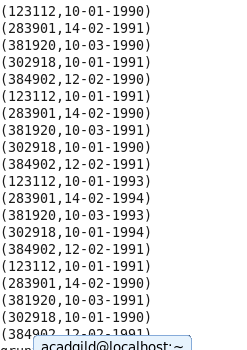
**9)Flatten: Flatten Operator**

**Flatten un-nests tuples as well as bags**

**consider a relation that has a tuple of the form (a, (b, c)). The expression GENERATE $0, flatten($1), will cause that tuple to become (a, b, c).**

****

****

****

**10IsEmpty: IsEmpty(expression)**

**Checks if a bag or map is empty**