**Parameter substitution in Pig**

Assume ,we have a file called pig11 with below data.

1 sasi 5000 12

2 siva 3000 23

3 amar 4000 34

5 divya 7990 12

6 rakesh 9876 56

7 samba 6539 34

8 nandu 7896 57

9 jana 9843 12

If we want to list departments equal to 12 ,then we write pig latin code like below.

empdata = load ‘pig11’ using PigStorage(‘ ‘) as (id:int, name:chararray, salary:int, deptid:int);

res\_data = filter empdata by deptid==12;

Dump res\_data;

Usually we write above code in a file .let us assume we have written it in a file called pig1prog.pig

And we write the above code in file. Then will execute by using below command

pig –f /home/xxx/pig1prog.pig  
  
If you want to run same program through grunt shell  
  
run /home/xxx/pig1prog.pig

Later if we want to see only numbers equals to 34, then we change second line to

res\_data = filter empdata by deptid==34;

and we re-run the code using same command.

But Its not a good practice to touch the code in production ,so we can make this script dynamic by using –param option of Piglatin.

Whatever values we want to decide at the time of running we make them dynamic .now we want to decide number to be filtered at the time running job,we can write second line like below.

res\_data = filter empdata by deptid==$departmentid

and we run code like below.

Pig –param departmentid=12  –f /home/xxx/pig1prog.pig

Assume we even want to take path at the time of running script, now we write code like below

empdata = load ‘$ipfile’ using PigStorage(‘ ‘) as (id:int, name:chararray, salary:int, deptid:int);

res\_data = filter empdata by deptid==$departmentid;

Dump res\_data;

And run like below

Pig –param ipfile=pig11 –param departmentid =34 –f pig1program.pig

If you feel this code is missing readability, we can specify all these dynamic values in a file like below

##Dyna.params (file name)

ipfile = pig11

departmentid = 34

Then you can run script with param-file option like below.

Pig –param\_file dyna.params –f pig1program.pig

This type of feature is not available in apache hive.

So what are the benefits we gain using this feature.

**1.       We can avoid hard coding in pig scripts**

**2.       Of course, we make scripts more reusable and dynamic.**

**3.       We can have better productivity using reusable scripts.**