**222BDA41**

**PIG CODES**

Copy files to read

* hadoop fs -copyFromLocal emp.csv pig/emp.csv

Load Data

* A = load '/user/root/pig/emp.csv' using PigStorage(',') as (eid:int,ename:chararray,epos:chararray,esal:int,ecom:int,edpno:int);
* Dump A;
* A2 = load '/user/root/pig/emp.csv’;
* describe A2;

Aggregate (by row)

* B = filter A by edpno==20;
* B2 = filter A by edpno==20 and epos=='MANAGER’,
* C = limit B 3;
* D = order C by esal desc;

Store Data

* store D into '/pig/pigout1’ using PigStorage(',’);

Transform (by column)

Select existing column

* E = foreach A generate eid;

Create new column

* F = foreach A generate \*, ecom\*2 as Bonus,esal\*5 as Incentive;

Transform columns

* G = foreach A generate SUBSTRING(ename,0,4);

Advanced codes

* H = foreach A generate $0,$1;
* I = group A by edpno;
* J = foreach I generate group as edpno, COUNT($1) as count;
* K = foreach A generate MAX(A.esal) as maxsal,MIN(A.esal) as minsal, SUM(A.esal) as sumsal, COUNT($1) as count;
* L = group A by (edpno, epos)

* SPLIT A into B if edpno==10, C if edpno==20, D if epos=='MANAGER';

Joins

* A = load '/emp.csv' using PigStorage(',') as (eid:int,ename:chararray,epos:chararray,esal:int,ecom:int,edpno:int);
* B = load '/dept.csv' using PigStorage(',') as (edpno:int,epos:chararray,ecity:chararray);
* C = JOIN A by edpno,B by edpno;
* D = foreach C generate A::eid,B::epos;
* E = JOIN A by edpno RIGHT OUTER, B by edpno;

Word Count

* lines = load '/plaintext.txt' as (line:chararray);
* token = foreach lines generate TOKENIZE(line);
* flats = foreach token generate FLATTEN($0);
* group\_words = group flats by $0;
* count\_word = foreach group\_words generate group as word, COUNT($1) as word\_occurence;