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

1. Concat

sample=LOAD '/home/acadgild/hadoop/pig.txt' USING PigStorage(',')

as (id:int,first\_name:chararray,last\_name:chararray,grade:int);

concate\_rows = FOREACH sample GENERATE CONCAT(first\_name , '' last\_name)

1. Tokenize

Tokenize\_example = FOREACH sample GENERATE TOKENIZE (first\_name);

1. Sum

emp\_detail = LOAD '/home/acadgild/hadoop/employee.txt' USING Pigstorage(',') as (id :int, name:chararray, wdate:chararray,type1:int);

emp\_gp =Group emp\_detail All;

emp\_sum=FOREACH emp\_gp GENERATE (emp\_ detail.name,emp\_ detail.type1),sum(emp\_ detail.type1)

DUMP emp\_sum;

1. Min

emp\_detail = LOAD '/home/acadgild/hadoop/employee.txt' USING Pigstorage(',') as (id :int, name:chararray, wdate:chararray,type1:int);

emp\_gp =Group emp\_detail All;

emp\_Min= FOREACH emp\_gp GENERATE (emp\_ detail.name,emp\_ detail.type1),MIN(emp\_ detail.type1)

DUMP emp\_Min;

1. Max

emp\_detail = LOAD '/home/acadgild/hadoop/employee.txt' USING Pigstorage(',') as (id :int, name:chararray, wdate:chararray,type1:int);

emp\_gp =Group emp\_detail All;

emp\_Max= FOREACH emp\_gp GENERATE (emp\_ detail.name,emp\_ detail.type1),MAX(emp\_ detail.type1)

DUMP emp\_Max;

1. Limit

emp = LOAD '/home/acadgild/hadoop/employee.txt' USING Pigstorage(',') as (id :int, name:chararray, wdate:chararray,type1:int);

emp\_limit = limit emp 2;

Dump emp\_limit;

1. Store

emp = LOAD '/home/acadgild/hadoop/employee.txt' USING Pigstorage(',') as (id :int, name:chararray, wdate:chararray,type1:int);

STORE emp INTO ‘hdfs://localhost:9000/pig\_Output/ ' USING PigStorage (',');

hdfs dfs -ls 'hdfs://localhost:9000/pig\_Output/'

hdfs dfs -cat 'hdfs://localhost:9000/pig\_Output/part-m-00000'

1. Distinct

emp = LOAD '/home/acadgild/hadoop/employee.txt' USING Pigstorage(',') as (id :int, name:chararray, wdate:chararray,type1:int);

Emp\_dist = distinct emp;

Dump emp\_dist;

1. Flatten

players = LOAD '/home/acadgild/hadoop/baseball.txtl' as (name:chararray, team:chararray, position:bag{t:(p:chararray)}, bat:map[]);

pos = foreach players generate name, flatten(position) as position;

bypos = group pos by position;

1. IsEmpty

emp\_sales = LOAD /home/acadgild/hadoop/emp\_sales.txt' USING PigStorage(',')

as (sno:int, name:chararray, age:int, salary:int, dept:chararray);

emp\_bonus = LOAD /home/acadgild/hadoop/emp\_bonus.txt' USING PigStorage(',')

as (sno:int, name:chararray, age:int, salary:int, dept:chararray);

cogroup\_data = COGROUP emp\_sales by age, emp\_bonus by age;

isempty\_data = filter cogroup\_data by IsEmpty(emp\_sales);

Dump isempty\_data;