**Student Names who passed exam**

-- load student dataset

stu = LOAD 'pig/student' USING PigStorage('\t') AS (name: chararray, rollno: int);

-- load results dataset

res = LOAD 'pig/results' USING PigStorage('\t') AS (rollno: int, result: chararray);

-- get passed results

pass\_res = FILTER res BY (result == 'pass');

-- join stu and pass\_res

pass\_stu = JOIN stu by rollno, pass\_res by rollno;

-- get student names

stu\_nms = FOREACH pass\_stu generate name;

-- dump stu\_nms

dump stu\_nms;

**Result:**

(raj)

(ajeet)

(ramesh)

(priyanka)

(suresh)

(ritesh)

(hitesh)

(raman)

(aman)

(ravi)

(shravi)

(chavvi)

**Weather data**

----loading and parsing data-----

A = load 'pig/weatherPIG.txt' using TextLoader as (data:chararray);

AF = foreach A generate TRIM(SUBSTRING(data, 6, 14)), TRIM(SUBSTRING(data, 46, 53)), TRIM(SUBSTRING(data, 38, 45));

store AF into 'pig/data9' using PigStorage(',');

S = load 'pig/data9/part-m-00000' using PigStorage(',') as (date:chararray, min:double, max:double);

-------Hot Days------

X = filter S by max > 25;

-------Cold Days------

X = filter S by min < 0;

-------Hottest Day-----

H1 = group S all; /\* puts S's data in H1's Tuple \*/

I = foreach H1 generate MAX(S.max) as maximum;

X = filter S by max == I.maximum;

-------Coldest Day------

H2 = group S all;

J = foreach H2 generate MIN(S.min) as minimum;

X = filter S by min == J.minimum;

-----UDF-----

register udf\_corrupt.jar;

A = load 'pig/weatherPIG.txt' using TextLoader as (data:chararray);

AF = foreach A generate TRIM(SUBSTRING(data, 6, 14)), IfCorrupted(TRIM(SUBSTRING(data, 46, 53))), IfCorrupted(TRIM(SUBSTRING(data, 38, 45)));

store AF into 'pig/data2' using PigStorage(',');

S = load 'pig/data2/part-m-00000' using PigStorage(',') as (date:chararray, min:double, max:double);

dump S;

**Health Care Use Case**

REGISTER deidentify.jar;

A = LOAD 'pig/healthcare\_Sample\_dataset2.csv' using PigStorage(',') AS (PatientID: int, Name: chararray, DOB: chararray, PhoneNumber: chararray, EmailAddress: chararray, SSN: chararray, Gender: chararray, Disease: chararray, weight: float);

B = LOAD 'pig/healthcare\_Sample\_dataset1.csv' using PigStorage(',') AS (PatientID: int, Name: chararray, DOB: chararray, PhoneNumber: chararray, EmailAddress: chararray, SSN: chararray, Gender: chararray, Disease: chararray, weight: float);

C = UNION A, B;

D = FOREACH C GENERATE PatientID, DeIdentifyUDF(Name,'12345678abcdefgh'), DeIdentifyUDF(DOB,'12345678abcdefgh'), DeIdentifyUDF(PhoneNumber,'12345678abcdefgh'), DeIdentifyUDF(EmailAddress,'12345678abcdefgh'),DeIdentifyUDF(SSN,'12345678abcdefgh'), DeIdentifyUDF(Disease,'12345678abcdefgh'),weight;

STORE D into 'pig/deidentifiedDir';