**ACD\_BDD\_Session\_8\_Assignment\_2\_Main**

**Problem Statement**

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

1) Concat

2) Tokenize

3) Sum

4) Min

5) Max

6) Limit

7) Store

8) Distinct

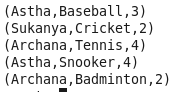
9) Flatten

10) IsEmpty

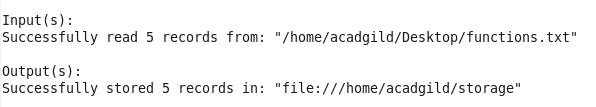
**Solution**

Data stored on local: /usr/local/hadoop-2.6.0/sbin/functions.txt'

Astha Baseball 5  
Sukanya Cricket 3  
Astha Snooker 4  
Sukanya Snooker 2  
Archana Badminton 3  
 Cricket 6

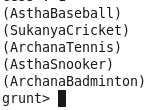
grunt> data = load '/home/acadgild/Desktop/functions.txt' AS (f1:chararray, f2:chararray, f3:int);  
  
grunt> DUMP data;  
  


grunt> **STORE** data INTO 'storedata';



grunt> next = FOREACH data GENERATE **CONCAT**(f1,f2);

grunt> DUMP next;



grunt> token = FOREACH data GENERATE **flatten**(**TOKENIZE**(f1)) AS word;  
  
grunt> DUMP token;

(Astha)

(Sukanya)

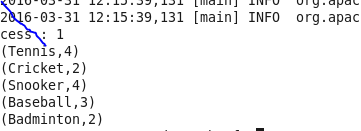
(Archana)  
(Astha)

(Archana)

grunt>now = GROUP data BY f2;  
  
grunt> DUMP now;

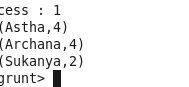
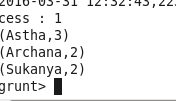
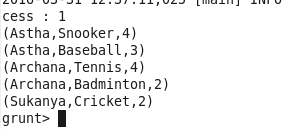


grunt>now2 = GROUP data1 by f2;  
grunt> try = FOREACH now2 GENERATE group, **SUM**(data.f3);

  
  
grunt> DESCRIBE try;

try: {group: chararray,long}

test = GROUP data1 by f1;

grunt> max1 = FOREACH test GENERATE group, **MAX**(data.f3);  
  
grunt> DUMP max1;  
  
(  
  
grunt> Min = FOREACH test GENERATE group, **MIN**(data.f3);  
  
grunt> DUMP Min;  
  
  
  
grunt> dist = **DISTINCT** data;  
  
  
  
grunt> lim = **LIMIT** data 2;  
  
