**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

**Dataset.txt**

**MEDAL NAME GAMES SPORT EVENT**

Silver Vijay Kumar 2012 London Shooting Men’s 25 Rapid Fire pistol

Silver Sushil Kumar 2012 London Wrestling Men’s 66 kg Freestyle

Bronze Saina Nehwal 2012 London Badminton Women’s singles

Bronze Mary Kom 2012 London Boxing Women’s flyweight

Bronze Gagan Narang 2012 London Shooting Men’s 10 m Air Rifle

Bronze Yogeshwar Dutt 2012 London Wrestling Men’s 60 kg Freestyle

olympic = load '/home/acadgild/pig/dataset.txt' using PigStorage('\t') as (medal:chararray,name:chararray,games:chararray,sport:chararray,event:bytearray,income:bytearray);

**1) CONCAT**

A= foreach olympic generate CONCAT(medal,',',name);

OUTPUT:

(Silver,Vijay Kumar)

(Silver,Sushil Kumar)

(Bronze,Saina Nehwal)

(Bronze,Mary Kom)

(Bronze,Gagan Narang)

(Bronze,Yogeshwar Dutt)

**2)TOKENIZE**

A = foreach olympic generate TOKENIZE(games);

({(2012),(London)})

({(2012),(London)})

({(2012),(London)})

({(2012),(London)})

({(2012),(London)})

({(2012),(London)})

**3) SUM**

A = group olympic by medal;

B= foreach A generate group,SUM(olympic.income);

(Bronze,145.0)

(Silver,56.0)

**4)MIN**

A = group olympic by medal;

B= foreach A generate group,MIN(olympic.income) as min\_income;

(Bronze,23.0)

(Silver,22.0)

**5)MAX**

A = group olympic by medal;

B= foreach A generate group,MAX(olympic.income) as max\_income;

(Bronze,45.0)

(Silver,34.0)

**6)LIMIT**

A = LIMIT olympic 4;

(Silver,Vijay Kumar,2012 London,Shooting,Men�s 25 Rapid Fire pistol,22)

(Silver,Sushil Kumar,2012 London,Wrestling,Men�s 66 kg Freestyle,34)

(Bronze,Saina Nehwal,2012 London,Badminton,Women�s singles,45)

(Bronze,Mary Kom,2012 London,Boxing,Women�s flyweight,23)

**7)STORE**

store A into 'session8' using PigStorage('|');

OUTPUT:

[acadgild@localhost pig]$ cd session8/

[acadgild@localhost session8]$ ll

total 4

-rw-r--r--. 1 acadgild acadgild 254 Mar 29 23:14 part-r-00000

-rw-r--r--. 1 acadgild acadgild 0 Mar 29 23:14 \_SUCCESS

[acadgild@localhost session8]$ cat part-r-00000

Bronze|Mary Kom|2012 London|Boxing|Women�s flyweight|23

Bronze|Saina Nehwal|2012 London|Badminton|Women�s singles|45

Silver|Vijay Kumar|2012 London|Shooting|Men�s 25 Rapid Fire pistol|22

Silver|Sushil Kumar|2012 London|Wrestling|Men�s 66 kg Freestyle|34

**8)DISTINCT**

B= DISTINCT (foreach olympic generate sport);

DUMP B;

(Boxing)

(Shooting)

(Badminton)

(Wrestling)

**9)FLATTEN**

B= group olympic by sport;

C = foreach B generate FLATTEN(group);

(Boxing)

(Shooting)

(Badminton)

(Wrestling)

**10)ISEMPTY**

sports = load '/home/acadgild/pig/dataset\_join.txt' as (sport:chararray);

(Shooting)

(Wrestling)

(Badminton)

A = cogroup olympic by sport,sports by sport;

B = filter A by IsEmpty(sports);

(Boxing,{(Bronze,Mary Kom,2012 London,Boxing,Women�s flyweight,23)},{})