Assignment 8.2

Problem Statement

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

1. **Concat**

student\_details = LOAD '/home/ec2-user/pigdata/student\_details.txt' USING PigStorage(',') as (id:int, firstname:chararray, lastname:chararray, age:int, phone:chararray, city:chararray, gpa:int);

**student\_name\_concat = foreach student\_details generate id, CONCAT(firstname,lastname) as name, phone,city,gpa;**

1. **Tokenize**

news = LOAD '/home/ec2-user/pigdata/news.txt' USING PigStorage() as (headline:chararray);

**words\_tkn = foreach news generate TOKENIZE(headline) as words;**

1. **Sum**

studentgrp = GROUP student\_details all;

**sum\_of\_students\_age = foreach studentgrp generate SUM(student\_details.age) as age\_sum;**

1. **Min**

studentgrp = GROUP student\_details all;

**min\_gpa = foreach studentgrp generate MIN(student\_details.gpa);**

1. **Max**

studentgrp = GROUP student\_details all;

**max\_gpa = foreach studentgrp generate MAX(student\_details.gpa);**

1. **Limit**

student\_ordered\_age = ORDER student\_details By age;

**youngest\_student = LIMIT student\_ordered\_age 1;**

1. **Store**

**STORE student\_name\_concat into '/home/ec2-user/pigdata/student\_name\_concat' using PigStorage(',');**

1. **Flatten**

**words\_flattened = foreach words\_tkn generate FLATTEN(words) as word;**

1. **IsEmpty**

sales = LOAD '/home/ec2-user/pigdata/emp\_sales.txt' USING PigStorage(',') as (sno:int, name:chararray, age:int, salary:int, dept:chararray);

bonus = LOAD '/home/ec2-user/pigdata/emp\_bonus.txt' USING PigStorage(',') as (sno:int, name:chararray, age:int, salary:int, dept:chararray);

co\_grp\_data = COGROUP sales by age, bonus by age;

**isempty\_data = filter co\_grp\_data by IsEmpty(sales);**

1. **Distinct**

**words\_distinct = DISTINCT words\_flattened;**