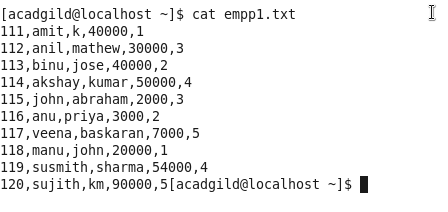
**ASSIGNMENT 2.2**

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

**Dataset:**

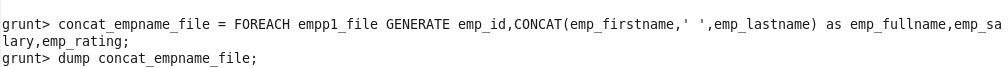
****

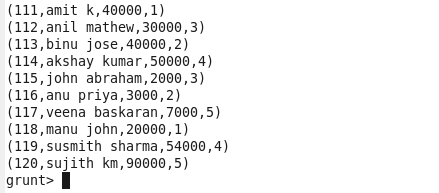
**Loading the file to hdfs:**

**C:\Users\skm\Pictures\snap\A5\Capture1.PNG**

****

**1)Concat**

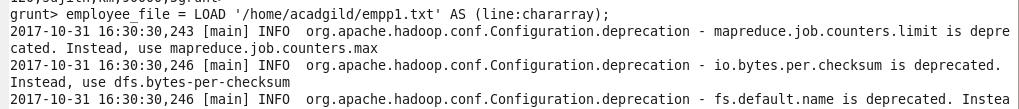
****

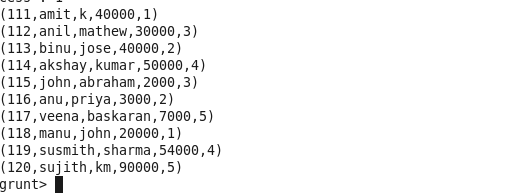
****

**C:\Users\skm\Pictures\snap\A5\Capture5.PNG**

**2) Tokenize**

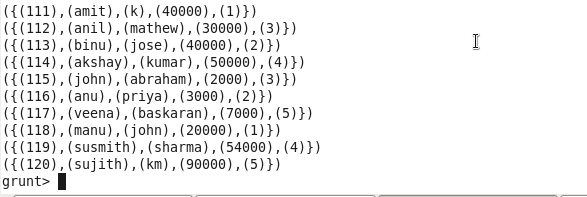
First load the file as single line

****

****

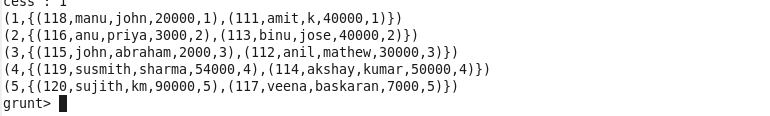
**C:\Users\skm\Pictures\snap\A5\Capture8.PNG**

**Output for tokenized file:**

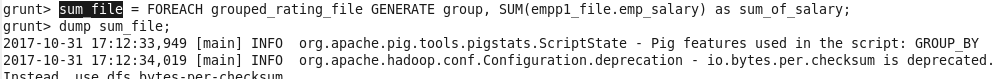
****

Grouping the data based on rating for performing aggregate operation

**C:\Users\skm\Pictures\snap\A5\10.PNG**

****

**3)Sum**

****

**output**

****

**4)Min**

**C:\Users\skm\Pictures\snap\A5\14.PNG**

**5)Max**

**C:\Users\skm\Pictures\snap\A5\16.PNG**

**6)Limit**

**First sort the file:**

**C:\Users\skm\Pictures\snap\A5\18.PNG**

**Apply limit**

**C:\Users\skm\Desktop\new.PNG**

**C:\Users\skm\Desktop\new1.PNG**

**Output**

**C:\Users\skm\Pictures\snap\A5\19.PNG**

**7)Store**

**C:\Users\skm\Pictures\snap\A5\20.PNG**

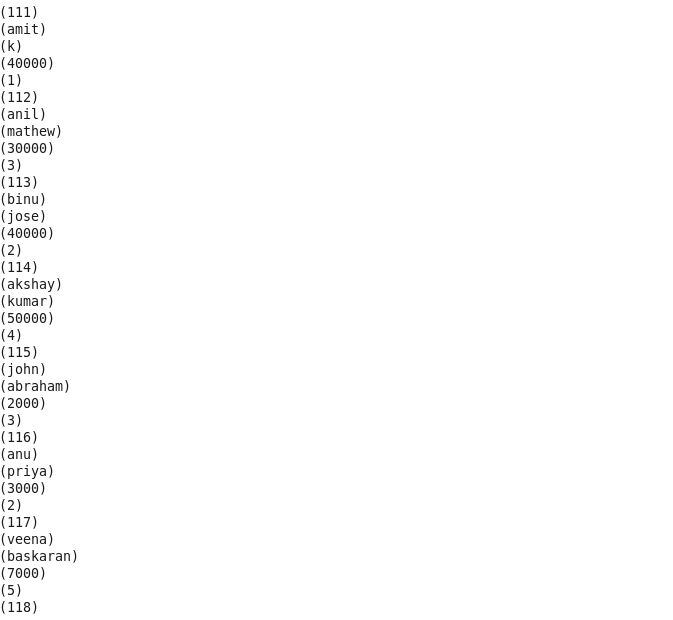
**output**

**8)Distinct**

**C:\Users\skm\Pictures\snap\A5\23.PNG**

**10) Flatten**

**C:\Users\skm\Pictures\snap\A5\25.PNG**

****

****