**EXPERIMENT 9:** *To use Apache Pig and implement group by, order by, filter and join operations on databases inside HDFS.*

**Name: Adwait Purao**

**UID : 2021300101**

**Division: COMPS B**

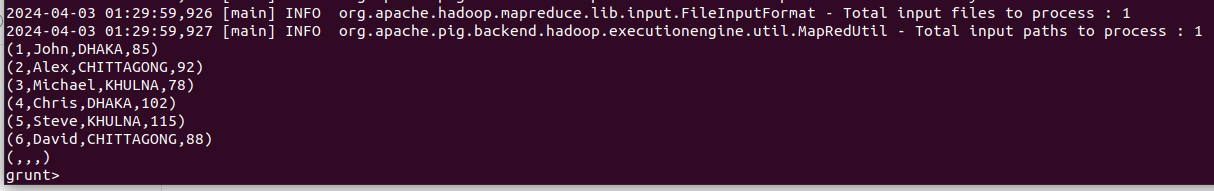
1. **Start Pig in Local Mode:**
   1. **Command:**

| pig -x local |
| --- |



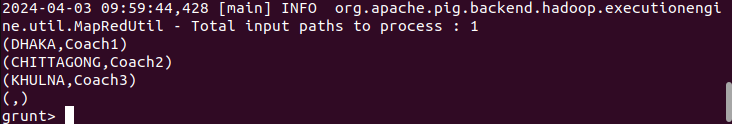
1. **Load data into Apache Pig:**
   1. **Command:**

| cricketers\_data = LOAD 'file:///home/hadoop/Desktop/Pig/cricketers\_data.txt' USING PigStorage(',') AS (player\_id:int, name:chararray, team:chararray, score:int);  dump cricketers\_data; |
| --- |

****

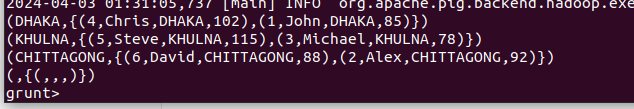
1. **Load data into Apache Pig:**
   1. **Command:**

| team\_data = LOAD 'file:///home/hadoop/Desktop/Pig/team\_data.txt' USING PigStorage(',') AS (team:chararray, coach:chararray);  dump team\_data; |
| --- |



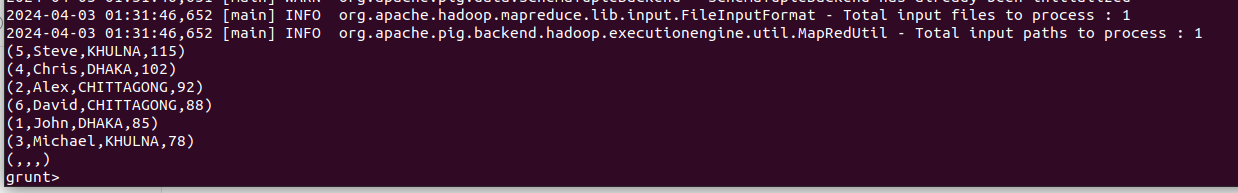
1. **Group By: To group the data by team:**
   1. **Command:**

| grouped\_data = GROUP cricketers\_data BY team;  dump grouped\_data; |
| --- |

****

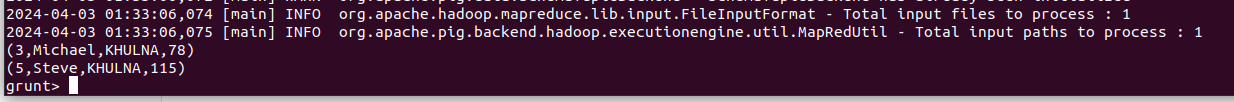
1. **Order By: To order the data by score:**
   1. **Command:**

| ordered\_data = ORDER cricketers\_data BY score DESC;  dump ordered\_data; |
| --- |

****

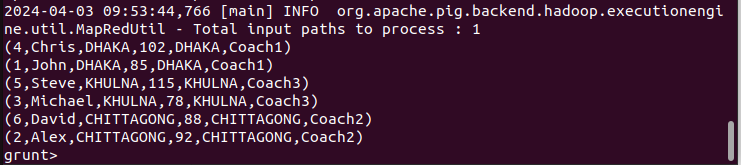
1. **Filter: To filter the data for a specific team, say KHULNA**
   1. **Command:**

| filtered\_data = FILTER cricketers\_data BY team == 'KHULNA';  dump filtered\_data; |
| --- |

****

1. **Join: To join two datasets,on the basis of a common attribute.** 
   1. **Command:**

| joined\_data = JOIN cricketers\_data BY team, team\_data BY team;  dump joined\_data; |
| --- |

****