**Experiment 10: Week 11:**

1. Installation of PIG and Write Pig Latin scripts sort, group, join, project, and filter your data.

**8.2 RESOURCES:**

VMWare, Web browser, 4 GB RAM, Hard Disk 80 GB, VMWare, Web browser, 4 GB RAM, Hard Disk 80 GB.

**8.3 PROGRAM LOGIC:**

**STEPS FOR INSTALLING APACHE PIG**

1) Extract the pig-0.15.0.tar.gz and move to home directory

2) Set the environment of PIG in bashrc file.

3) Pig can run in two modes

Local Mode and Hadoop Mode

Pig –x local and pig

4) Grunt Shell

Grunt >

5) LOADING Data into Grunt Shell

DATA = LOAD <CLASSPATH> USING PigStorage(DELIMITER) as (ATTRIBUTE :

DataType1, ATTRIBUTE : DataType2…..)

6) Describe Data

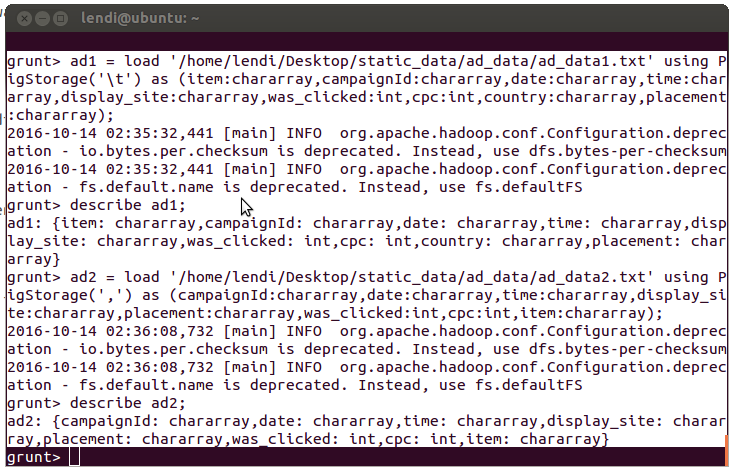
Describe DATA;

7) DUMP Data

Dump DATA;

**8.4 INPUT/OUTPUT:**

Input as Website Click Count Data



**PROGRAM LOGIC:**

**FILTER Data**

FDATA = FILTER DATA by ATTRIBUTE = VALUE;

**GROUP Data**

GDATA = GROUP DATA by ATTRIBUTE;

**Iterating Data**

FOR\_DATA = FOREACH DATA GENERATE GROUP AS GROUP\_FUN,

ATTRIBUTE = <VALUE>

**Sorting Data**

SORT\_DATA = ORDER DATA BY ATTRIBUTE WITH CONDITION;

**LIMIT Data**

LIMIT\_DATA = LIMIT DATA COUNT;

**JOIN Data**

JOIN DATA1 BY (ATTRIBUTE1,ATTRIBUTE2….) , DATA2 BY

(ATTRIBUTE3,ATTRIBUTE….N)

**INPUT / OUTPUT :** 