Task No: 9 Date:	Write Pig Latin scripts sort, group, join, project, and filter the data **Tools: Pig, LINUX/WINDOWS**	CO4 K3

AIM:

To Implement the Pig Latin scripts sort, group, join, project, and filter the data using Apache Pig in Windows Operating System.

PROCEDURE:

1) Install the prerequisite Stable software's Java Development Kit and Java Runtime Environment, Apache Hadoop.
2) Visit the Apache Pig download page: https://pig.apache.org/downloads.html
3) Download the latest stable release of Pig.
4) Extract the Pig Archive and Set Environment Variables for java, Apache Hadoop and Apache Pig

   For Java:

   JAVA_HOME=C:\Program Files\Java\jdk-1.8

   Path = C:\Program Files\Java\jdk-1.8\bin

   For Hadoop:

   HADOOP_HOME=C:\ApacheHadoop2.9.2

   Path = C:\ApacheHadoop2.9.2\bin

   For Pig:

   PIG_HOME = C:\Apachepig

   Path = C:\Apachepig\bin

   Path = C:\Apachepig\conf

5) Verify the Paths, Run following commands in a NEW Command Window

   echo %PIG_HOME%

6) Open the pig. cmd file in edit mode, and change the value of the HADOOP_BIN_PATH

   Old value:- %HADOOP_HOME%\bin

   New Value:- %HADOOP_HOME%\libexec

7) Edit Pig Configuration, go to the conf directory within your Pig installation directory, rename the pig.properties.template file to pig.properties. and set the exectype property to "local"

   exectype=local

8) Start Apache Pig, run the following command in a new Command Prompt as administrator

   C:\Users\Lenovo>echo %PIG_HOME%

   O/P: C:\ApachePig

C:\Users\Lenovo>pig -version

Apache Pig version 0.17.0 (r1797386)

compiled Jun 02 2017, 15:41:58

C:\Users\Lenovo>pig

Grunt Shell started:

grunt>

Output:

Pig Latin scripts:

Input.txt

Rajiv,42
siddarth,45
Rajesh,40
Preethi,23
Trupthi,34
Archana,21
Robin,22
BOB,23
Maya,23
Sara,25
David,23
Maggy,22

Addressfile.txt

Rajiv,Chennai
Rajesh,Delhi
Trupthi,Hyderabad
Robin,Pune
Maya,Hyderabad
Anderson,Chennai
Antolina,Chennai

Load Data:
data = LOAD './input.txt' USING PigStorage(',') AS (name:chararray, age:int);

Sort Operator:
sortbyage = ORDER data BY age ASC|DESC;
dump sortbyage;

Group Operator:
grouped_data = GROUP data BY age;
dump grouped_data;

Filter Operator:

```
filterbyage = FILTER data BY (age>40);
dump filterbyage;
```

Inner Join Operator:

```
table1 = LOAD './input.txt' USING PigStorage(',') AS (name:chararray, age:int);
table2 = LOAD './addressfile.txt' USING PigStorage(',') AS (name:chararray, address:chararray);

joinbyname = JOIN table1 BY name, table2 BY name;
```

OuterJoin Operator;

```
LO = JOIN table1 BY name LEFT OUTER, table2 BY name;
dump LO;

RO = JOIN table1 BY name RIGHT, table2 BY name;
dump RO;

FO = JOIN table1 BY name FULL OUTER, table2 BY name;
dump FO;
```

Store Operator:

```
grouped_data = GROUP data BY age;
result = FOREACH grouped_data GENERATE group AS age, COUNT(data) AS count;
STORE result INTO 'output';
```

Result:

Thus the Apache Pig Latin scripts sort, group, join, project, and filter the data are executed Successfully.