**Pig Programming: Create Your First Apache Pig Script**

In our [***Hadoop Tutorial Series***](http://www.edureka.co/blog/hadoop-tutorial?utm_source=blog&utm_medium=content-link&utm_campaign=pig-script), we will now learn how to create an Apache Pig script. Apache Pig scripts are used to execute a set of Apache Pig commands collectively. This helps in reducing the time and effort invested in writing and executing each command manually while doing this in Pig programming. This blog is a step by step guide to help you create your first Apache Pig script.

**Apache Pig script Execution Modes**

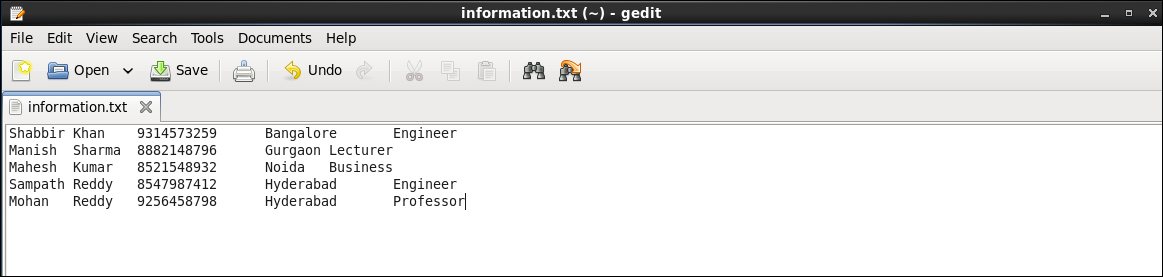
**Local Mode**: In ‘local mode’, you can execute the pig script in local file system. In this case, you don’t need to store the data in Hadoop HDFS file system, instead you can work with the data stored in local file system itself.

**MapReduce Mode**: In ‘MapReduce mode’, the data needs to be stored in HDFS file system and you can process the data with the help of pig script.

**Apache Pig Script in MapReduce Mode**

Let us say our task is to read data from a data file and to display the required contents on the terminal as output.

**The sample data file contains following data:**

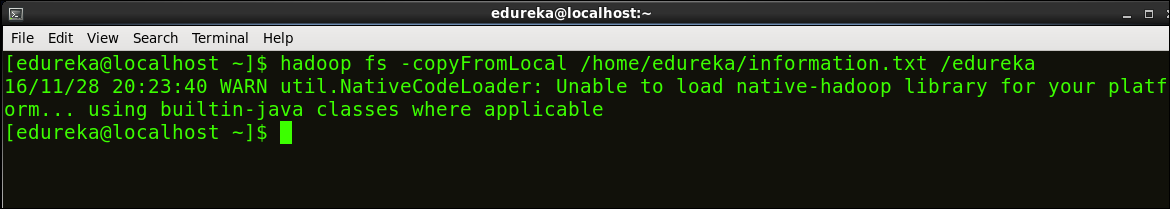


Save the text file with the name ‘information.txt’

The sample data file contains five columns *FirstName*, *LastName*, *MobileNo*, *City*, and *Profession* separated by **tab key**. Our task is to read the content of this file from the HDFS and display all the columns of these records.

To process this data using Pig, this file should be present in Apache Hadoop HDFS.

***Command****: hadoop fs –copyFromLocal /home/edureka/information.txt /edureka*

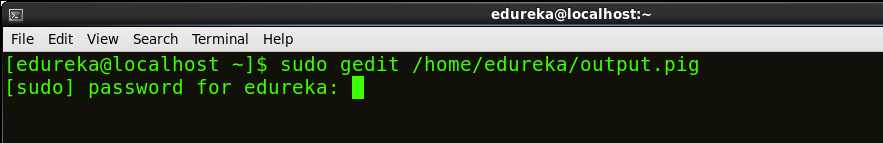


***Step 1:*** **Writing a Pig script**

Create and open an Apache Pig script file in an editor (e.g. gedit).

***Command****: sudo gedit /home/edureka/output.pig*

This command will create a ‘output.pig’ file inside the home directory of edureka user.



Let’s write few PIG commands in output.pig file.

|  |  |
| --- | --- |
| 1  2  3  4  5 | A = LOAD ‘/edureka/information.txt’ using PigStorage (‘\t’) as (FName: chararray, LName: chararray, MobileNo: chararray, City: chararray, Profession: chararray);    B = FOREACH A generate FName, MobileNo, Profession;    DUMP B; |

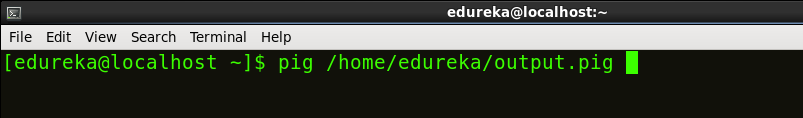
Save and close the file.

* The first command loads the file ‘information.txt’ into variable A with indirect schema (FName, LName, MobileNo, City, Profession).
* The second command loads the required data from variable A to variable B.
* The third line displays the content of variable B on the terminal/console.

***Step 2:* Execute the Apache Pig Script**

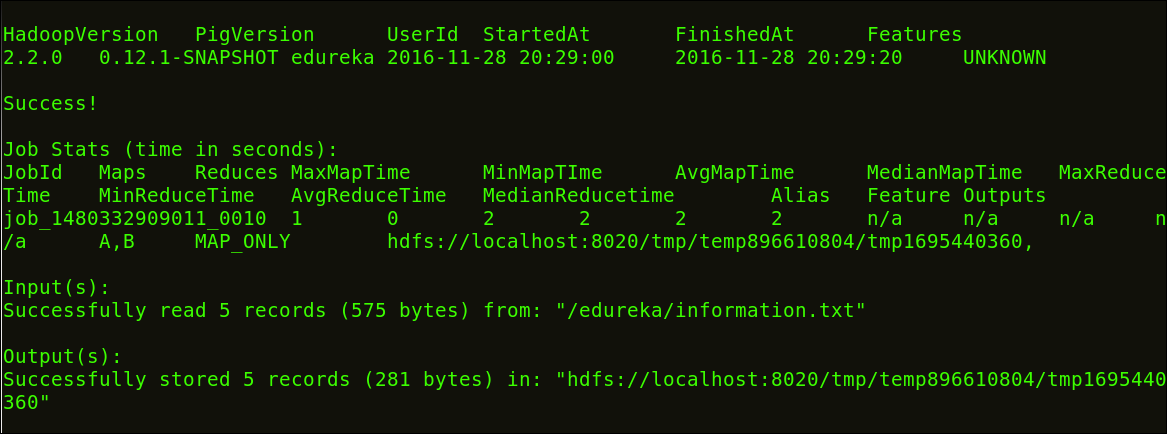
To execute the pig script in HDFS mode, run the following command:

***Command****: pig /home/edureka/output.pig*

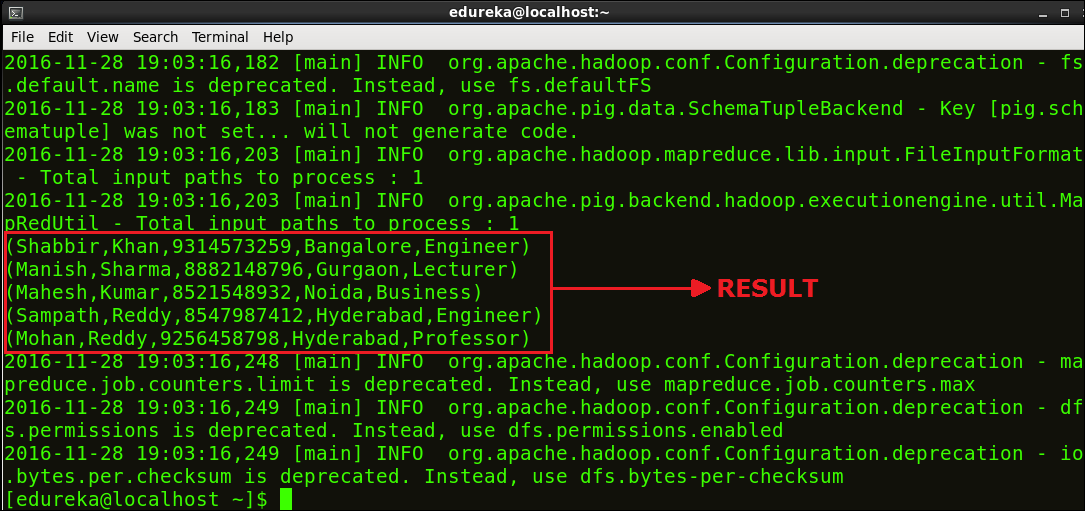


After the execution finishes, review the result. These below images show the results and their intermediate map and reduce functions.

Below image shows that the Script executed successfully.



Below image shows the result of our script.



*Congratulations on executing your first Apache Pig script successfully!*

Now you know, how to create and execute Apache Pig script. Hence, our next blog in [***Hadoop Tutorial Series***](http://www.edureka.co/blog/hadoop-tutorial?utm_source=blog&utm_medium=content-link&utm_campaign=pig-script) will be covering how to [***create UDF (User Defined Functions) in Apache Pig***](http://www.edureka.co/blog/pig-programming-apache-pig-script-with-udf-in-hdfs-mode?utm_source=blog&utm_medium=content-link&utm_campaign=pig-script) and execute it in MapReduce/HDFS mode.