Announcing the Ridiculously Committed Mentor Award   Nominate Now

(https://www.edureka.co/blog/all)	Blogs (https://www.edureka.co/blog/)	Videos (https://www.edureka.co/blog/videos/)	Interview Questions (https://www.edureka.co/blog/interview-questions/)

Pig Programming: Create Your First Apache Pig Script

🔖 Recommended by 117 users
Shubham Sinha (https://www.edureka.co/blog/author/shubham-sinha/) | Jul 16,2018

Pig Programming: Create Your First Apache Pig Script

In our **Hadoop Tutorial Series** (https://www.edureka.co/blog/hadoop-tutorial/), 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.
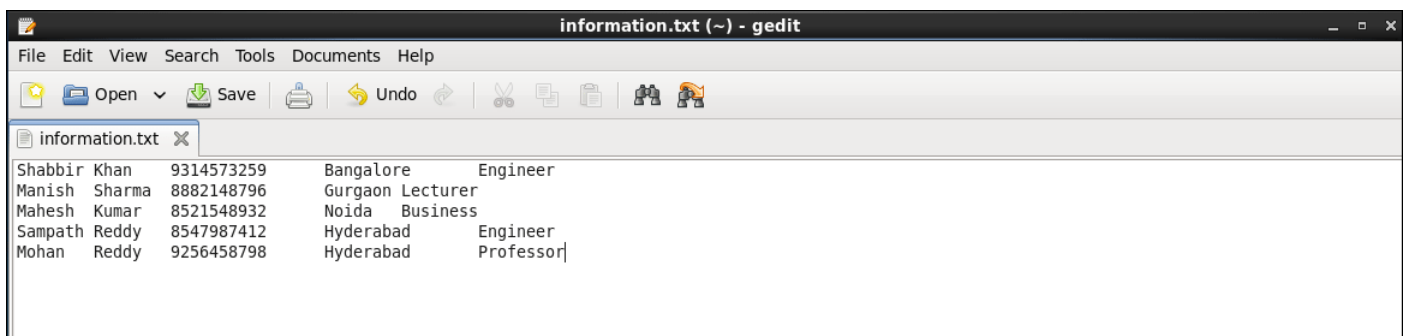
Apache Pig script Execution Modes

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:
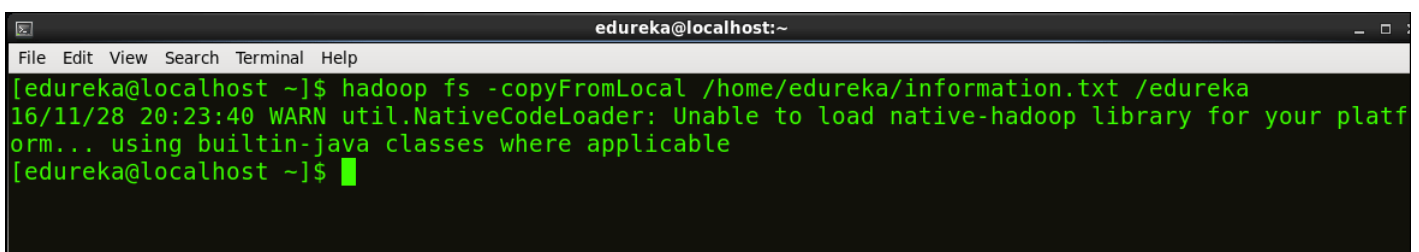
Save the text file with the name 'information.txt'

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**

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

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

```
                                                    edureka@localhost:~
File  Edit  View  Search  Terminal  Help
[edureka@localhost ~]$ sudo gedit /home/edureka/output.pig
[sudo] password for edureka: █
```

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

```
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*

```
                                                    edureka@localhost:~
File  Edit  View  Search  Terminal  Help
[edureka@localhost ~]$ 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.

```
HadoopVersion   PigVersion       UserId   StartedAt        FinishedAt        Features
2.2.0    0.12.1-SNAPSHOT edureka 2016-11-28 20:29:00       2016-11-28 20:29:20      UNKNOWN

Success!

Job Stats (time in seconds):
JobId    Maps    Reduces MaxMapTime      MinMapTIme      AvgMapTime      MedianMapTime   MaxReduce
Time     MinReduceTime   AvgReduceTime   MedianReducetime        Alias   Feature Outputs
job_1480332909011_0010  1       0       2       2       2       2       n/a     n/a     n/a     n
/a       A,B     MAP_ONLY        hdfs://localhost:8020/tmp/temp896610804/tmp1695440360,

Input(s):
Successfully read 5 records (575 bytes) from: "/edureka/information.txt"

Output(s):
Successfully stored 5 records (281 bytes) in: "hdfs://localhost:8020/tmp/temp896610804/tmp1695440
360"
```

Below image shows the result of our script.

```
                                                    edureka@localhost:~                          _ □
File  Edit  View  Search  Terminal  Help
2016-11-28 19:03:16,182 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs
.default.name is deprecated. Instead, use fs.defaultFS
2016-11-28 19:03:16,183 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.sch
ematuple] was not set... will not generate code.
2016-11-28 19:03:16,203 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat
 - Total input paths to process : 1
2016-11-28 19:03:16,203 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.Ma
pRedUtil - Total input paths to process : 1
(Shabbir,Khan,9314573259,Bangalore,Engineer)
(Manish,Sharma,8882148796,Gurgaon,Lecturer)
```
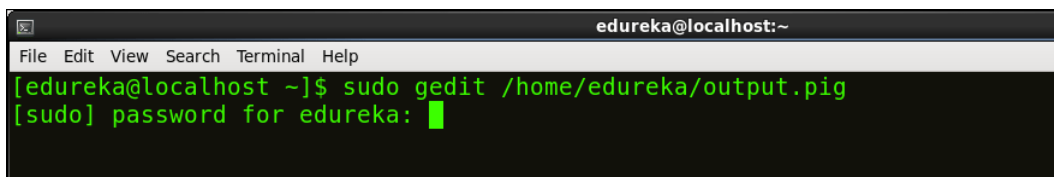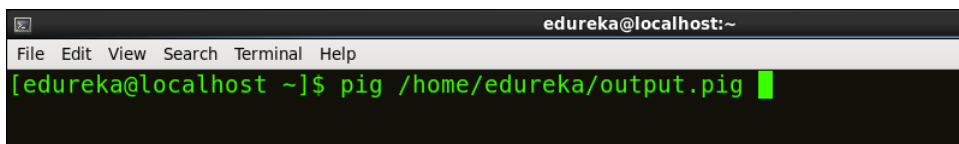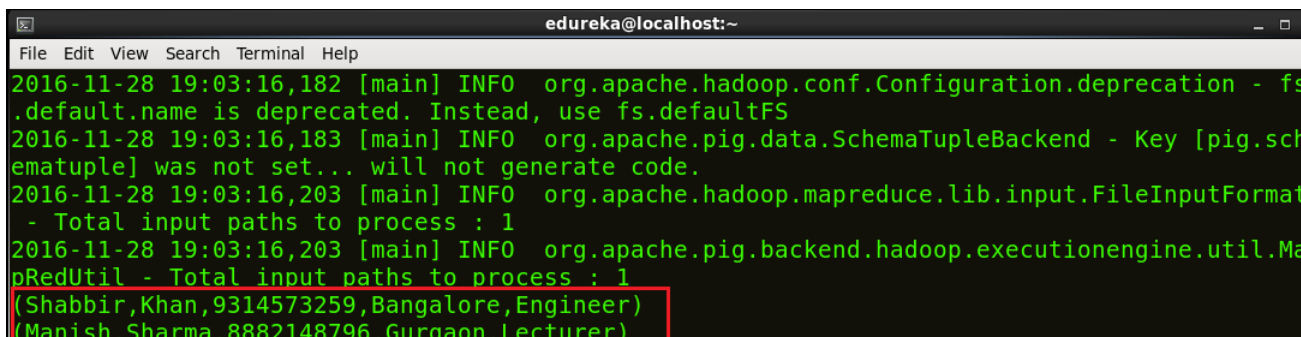
```
(Sampath,Reddy,8347987412,Hyderabad,Engineer)
(Mohan,Reddy,9256458798,Hyderabad,Professor)
2016-11-28 19:03:16,248 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - ma
preduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2016-11-28 19:03:16,249 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - df
s.permissions is deprecated. Instead, use dfs.permissions.enabled
2016-11-28 19:03:16,249 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io
.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
[edureka@localhost ~]$
```

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** (https://www.edureka.co/blog/hadoop-tutorial/) will be covering how to **create UDF (User Defined Functions) in Apache Pig** (https://www.edureka.co/blog/pig-programming-apache-pig-script-with-udf-in-hdfs-mode/) and execute it in MapReduce/HDFS mode.

Now that you have created and executed Apache Pig Script, check out the **Hadoop training** *(https://www.edureka.co/big-data-and-hadoop) by Edureka, a trusted online learning company with a network of more than 250,000 satisfied learners spread across the globe. The Edureka Big Data Hadoop Certification Training course helps learners become expert in HDFS, Yarn, MapReduce, Pig, Hive, HBase, Oozie, Flume and Sqoop using real-time use cases on Retail, Social Media, Aviation, Tourism, Finance domain.*

Got a question for us? Please mention it in the comments section and we will get back to you.

About Shubham Sinha (26 Posts (https://www.edureka.co/blog/author/shubham-sinha/))

Shubham Sinha is a Big Data and Hadoop expert working as a Research Analyst at Edureka. He is keen to work with Big Data related technologies such as Hadoop, Spark, Flink and Storm and web development technologies including Angular, Node.js & PHP.

Share on

(https://www.facebook.com/sharer/sharer.php?u=https://www.edureka.co/blog/pig-programming-create-your-first-apache-pig-script/)
(https://twitter.com/intent/tweet?url=https://www.edureka.co/blog/pig-programming-create-your-first-apache-pig-script/)
(https://plus.google.com/share?url=https://www.edureka.co/blog/pig-programming-create-your-first-apache-pig-script/)
(https://www.linkedin.com/cws/share?url=https://www.edureka.co/blog/pig-programming-create-your-first-apache-pig-script/)

‹ PREVIOUS   NEXT ›

Got your brain cells running?
Stay tuned to latest technology updates

Enter your Email Address

SUBSCRIBE

Related Posts

Pig Tutorial: Apache Pig Architecture & Twitter Case Study
👁 22.7K

Hadoop Tutorial: All you need to know about Hadoop!
👁 92.4K

Apache Pig Installation on Linux
👁 18.3K

Pig Programming: Apache Pig Script with UDF in HDFS Mode
👁 14K

(https://www.edureka.co/blog/pig-tutorial/)
(https://www.edureka.co/blog/hadoop-tutorial/)
(https://www.edureka.co/blog/apache-pig-installation)
(https://www.edureka.co/blog/pig-programming-apache-pig-script-with-udf-in-hdfs-mode/)

Browse Categories

Announcing the Ridiculously Committed Mentor Award   Nominate Now

Business Intelligence (https://www.edureka.co/blog/category/business-intelligence/)   Cloud Computing (https://www.edureka.co/blog/category/cloud-computing/)

Cyber Security (https://www.edureka.co/blog/category/cyber-security/)   Deep Learning (https://www.edureka.co/blog/category/deep-learning/)

Finance (https://www.edureka.co/blog/category/finance/)   Frameworks (https://www.edureka.co/blog/category/frameworks/)

Marketing (https://www.edureka.co/blog/category/marketing/)   Mobile Development (https://www.edureka.co/blog/category/mobile-development/)

Operations (https://www.edureka.co/blog/category/operations/)   Programming (https://www.edureka.co/blog/category/programming/)

Project Management (https://www.edureka.co/blog/category/project-management/)   Robotic Process Automation (https://www.edureka.co/blog/category/robotic-process-automation/)

Success Story (https://www.edureka.co/blog/category/success-story/)   Systems & Architecture (https://www.edureka.co/blog/category/systems-architecture/)

Systems Engineering (https://www.edureka.co/blog/category/systems-engineering/)   Testing (https://www.edureka.co/blog/category/testing/)

Comments

Comments   **2 Comments**

2 Comments   https://www.edureka.co/blog/   🔵 **Rajiv Chaudhuri** ▾

♡ **Recommend**   ↪ **Share**   Sort by Best ▾

Join the discussion…

srinivas • 5 years ago   ⊟ | ▾
Good example. Its helped me a lot
13 ∧ | ∨ • Reply • Share ›

Subin • a year ago   ⊟ | ▾
Hi,
The second variable B is loaded with only the required data which does not include the LName and city. But the results highlighted shows all the fields.
∧ | ∨ • Reply • Share ›

ALSO ON HTTPS://WWW.EDUREKA.CO/BLOG/

Top 10 DevOps Tools You Must Know In 2018
4 comments • 8 months ago
    Pavan — Oh between where is Shippable? Yes, I work at Shippable and I believe we
    are the best when it comes to DevOps.

Talend Big Data Tutorial – A Revolution In Big Data
2 comments • 8 months ago
    imen elloumi — Thank you very much for this article

Top 50 ServiceNow Interview Questions You Must Prepare In 2018
1 comment • 8 months ago
    upendhar lingireddy — 1. what is discovery and describe its functionality and use-
    cases?2. what is an PPM, your role in ppm?3. describe what is …

Real Time Big Data Applications in Various Domains
2 comments • 7 months ago
    Amber Anen — Yes, I totally agree with this. Industry influencers, academicians, and
    other thought leaders agree that big data is making impact on business …

✉ Subscribe   Ⓓ Add Disqus to your siteAdd DisqusAdd   🔒 Disqus' Privacy PolicyPrivacy PolicyPrivacy

Subscribe
to our newsletter

Enter your Email Address

SUBSCRIBE

Related Blogs

Announcing the Ridiculously Committed Mentor Award   Nominate Now

(https://www.
tutorial/)

Hadoop Tutorial: All you need to know about Hadoop! (https://www.edureka.co/blog/hadoop-tutorial/)

(https://www.
tutorial/)

Apache Pig Installation on Linux (https://www.edureka.co/blog/apache-pig-installation)

(https://www.
pig-
installation)

Pig Programming: Apache Pig Script with UDF in HDFS Mode (https://www.edureka.co/blog/pig-programming-apache-pig-script-with-udf-in-hdfs-mode/)

(https://www.
programming
apache-pig-
script-with-
udf-in-hdfs-
mode/)

Big Data Analytics Courses

Big Data Hadoop Certification Training (/big-data-and-hadoop)

(/big-data-
and-hadoop)

Apache Spark and Scala Certification Training (/apache-spark-scala-training)

(/apache-
spark-scala-
training)

Hadoop Administration Certification Training (/hadoop-admin)

(/hadoop-
admin)

Splunk Power User & Admin Certification Training (/splunk)

(/splunk)

Edureka

About us
(https://www.edureka.co/about-
us)

News & Media
(https://www.edureka.co/allmedia)

Contact us
(https://www.edureka.co/contact-
us)

Blog
(https://www.edureka.co/blog/)

Community
(https://www.edureka.co/community)

Work with us

Careers
(https://www.edureka.co/careers)

Become an Instructor
(https://www.edureka.co/instructors)

Become an Affiliate
(https://www.edureka.co/affiliate-
program)

Hire from Edureka
(https://www.edureka.co/hire-
from-edureka)

Usefull Links

Reviews
(https://www.edureka.co/reviews)

Terms & conditions
(https://www.edureka.co/terms-
and-conditions)

Privacy policy
(https://www.edureka.co/privacy-
policy)

Sitemap
(https://www.edureka.co/sitemap)

Follow us on

f   🐦   in   ▶

(https://www.facebook.com/edurekaIN)(https://twitter.com/edurekaIN)(https://www.linkedin.com/company/edureka)(https://www.youtube.com/user/edurekaIN)

Learn on the GO!

(https://itunes.apple.com/in/app/edureka/id1033145415?
mt=8)

(https://play.google.com/store/apps/details?
id=co.edureka.app)

edureka! (https://www.edureka.co)
© 2014 Brain4ce Education Solutions Pvt. Ltd. All
rights Reserved.

"PMP®","PMI®", "PMI-ACP®" and "PMBOK®" are registered marks of the
Project Management Institute, Inc.
MongoDB®, Mongo and the leaf logo are the registered trademarks of
MongoDB, Inc.