Go to the following link and download pig
http://mirrors.estointernet.in/apache/pig/pig-0.16.0/

To untar pig-0.16.0.tar.gz file run the following command:
$ tar xvzf pig-0.16.0.tar.gz

To create a pig folder and move pig-0.16.0 to the pig folder, execute the following command:
$ sudo mv /home/hadoop/pig-0.16.0 /usr/local/hadoop/pig

Now open the .bashrc file to edit the path and variables/settings for pig. Run the following command:
$ sudo nano .bashrc

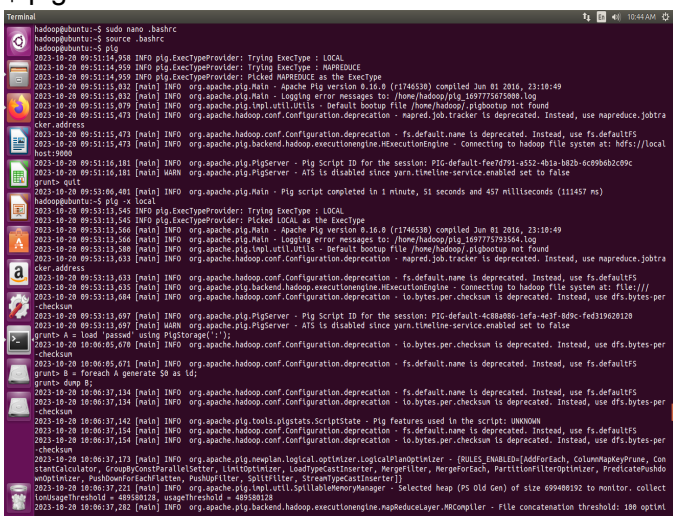
Add the below given to .bashrc file at the end and save the file.

#PIG settings
export PIG_HOME=/usr/local/hadoop/pig-0.16.0
export PATH=$PATH:$PIG_HOME/binexport PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_INSTALL/etc/hadoop/export PIG_CONF_DIR=$PIG_HOME/confexport JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
#PIG setting ends
(to exit ctrl +x)

Run the following command to make the changes effective in the .bashrc file:
$ source .bashrc

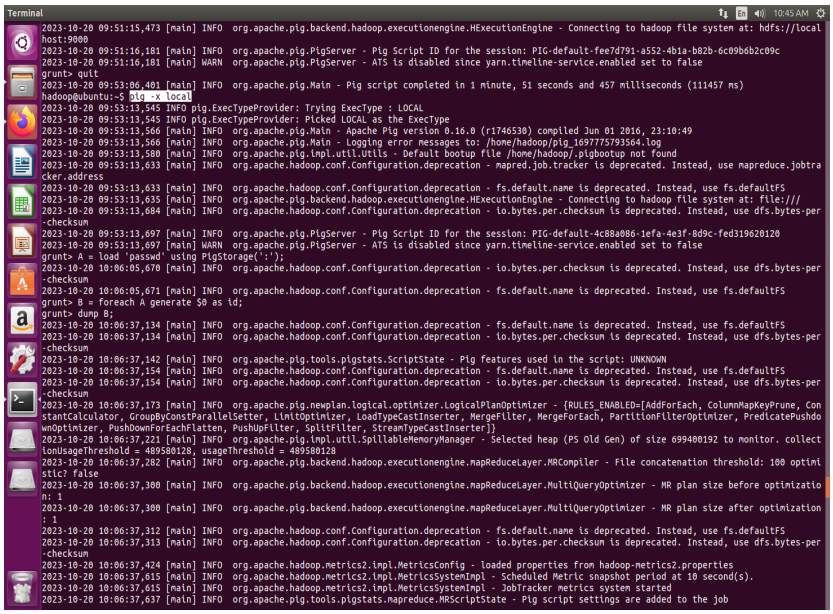
start all Hadoop daemons
cd /usr/local/hadoop/bin
start-all.sh
jps

Now you can launch pig by executing the following command:
$ pig



Now you are in pig and can perform your desired tasks on pig. You can come out of the pig by the quit command:
> quit;

Quit and run pig in local mode
pig -x local



You can run Pig in interactive mode using the Grunt shell. Invoke the Grunt shell using the "pig" command and then enter your Pig Latin statements and Pig commands interactively at the command line.

Example

These Pig Latin statements extract all user IDs from the /etc/passwd file. First, copy the /etc/passwd file to your local working directory. Then, enter the Pig Latin statements interactively at the grunt prompt. The DUMP operator will display the results to your terminal screen.

grunt> A = load 'passwd' using PigStorage(':');

```
Terminal
(this is a text file)
grunt> A = load 'passwd' using PigStorage(':');
2023-10-20 10:11:17,967 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per
-checksum
2023-10-20 10:11:17,967 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = foreach A generate $0 as id;
grunt> dump B;
2023-10-20 10:11:23,457 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2023-10-20 10:11:23,464 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-10-20 10:11:23,464 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per
-checksum
2023-10-20 10:11:23,464 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-10-20 10:11:23,465 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, Con
stantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdo
wnOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2023-10-20 10:11:23,465 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimi
stic? false
2023-10-20 10:11:23,466 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimizatio
n: 1
2023-10-20 10:11:23,466 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization
: 1
2023-10-20 10:11:23,471 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-10-20 10:11:23,471 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per
-checksum
2023-10-20 10:11:23,471 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-20 10:11:23,472 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2023-10-20 10:11:23,472 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buff
er.percent is not set, set to default 0.3
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Distributed cache not supported or needed in local mode. Setting key [pi
g.schematuple.local.dir] with code temp directory: /tmp/1697776883473-0
2023-10-20 10:11:23,480 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for s
ubmission.
2023-10-20 10:11:23,481 [JobControl] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-20 10:11:23,487 [JobControl] WARN  org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set.  User classes may not be found. See J
ob or Job#setJar(String).
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.pig.builtin.PigStorage - Using PigTextInputFormat
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2023-10-20 10:11:23,493 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2023-10-20 10:11:23,498 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_local1504266685_0003
2023-10-20 10:11:23,498 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Executing with tokens: []
2023-10-20 10:11:23,533 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8080/
2023-10-20 10:11:23,533 [Thread-11] INFO  org.apache.hadoop.mapred.LocalJobRunner - OutputCommitter set in config null
2023-10-20 10:11:23,538 [Thread-11] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.byte
s-per-checksum
2023-10-20 10:11:23,538 [Thread-11] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.j
obtracker.address
```

grunt> B = foreach A generate $0 as id;

grunt> dump B;

```
Terminal
(this is a text file)
grunt> A = load 'passwd' using PigStorage(':');
2023-10-20 10:11:17,967 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per
-checksum
2023-10-20 10:11:17,967 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> B = foreach A generate $0 as id;
grunt> dump B;
2023-10-20 10:11:23,457 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2023-10-20 10:11:23,464 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-10-20 10:11:23,464 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per
-checksum
2023-10-20 10:11:23,464 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-10-20 10:11:23,465 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, Con
stantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdo
wnOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2023-10-20 10:11:23,465 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimi
stic? false
2023-10-20 10:11:23,466 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimizatio
n: 1
2023-10-20 10:11:23,466 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization
: 1
2023-10-20 10:11:23,471 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2023-10-20 10:11:23,471 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per
-checksum
2023-10-20 10:11:23,471 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-20 10:11:23,472 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2023-10-20 10:11:23,472 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buff
er.percent is not set, set to default 0.3
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2023-10-20 10:11:23,473 [main] INFO  org.apache.pig.data.SchemaTupleFrontend - Distributed cache not supported or needed in local mode. Setting key [pi
g.schematuple.local.dir] with code temp directory: /tmp/1697776883473-0
2023-10-20 10:11:23,480 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for s
ubmission.
2023-10-20 10:11:23,481 [JobControl] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-20 10:11:23,487 [JobControl] WARN  org.apache.hadoop.mapreduce.JobResourceUploader - No job jar file set.  User classes may not be found. See J
ob or Job#setJar(String).
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.pig.builtin.PigStorage - Using PigTextInputFormat
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2023-10-20 10:11:23,492 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2023-10-20 10:11:23,493 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - number of splits:1
2023-10-20 10:11:23,498 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Submitting tokens for job: job_local1504266685_0003
2023-10-20 10:11:23,498 [JobControl] INFO  org.apache.hadoop.mapreduce.JobSubmitter - Executing with tokens: []
2023-10-20 10:11:23,533 [JobControl] INFO  org.apache.hadoop.mapreduce.Job - The url to track the job: http://localhost:8080/
2023-10-20 10:11:23,533 [Thread-11] INFO  org.apache.hadoop.mapred.LocalJobRunner - OutputCommitter set in config null
2023-10-20 10:11:23,538 [Thread-11] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.byte
s-per-checksum
2023-10-20 10:11:23,538 [Thread-11] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.j
obtracker.address
```

```
Terminal
2023-10-20 10:11:23,771 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per
-checksum
2023-10-20 10:11:23,771 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-10-20 10:11:23,774 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-20 10:11:23,774 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(root)
(daemon)
(bin)
(sys)
(sync)
(games)
(man)
(lp)
(mail)
(news)
(uucp)
(proxy)
(www-data)
(backup)
(list)
(irc)
(gnats)
(nobody)
(systemd-timesync)
(systemd-network)
(systemd-resolve)
(systemd-bus-proxy)
(syslog)
(_apt)
(messagebus)
(uuidd)
(lightdm)
(whoopsie)
(avahi-autoipd)
(avahi)
(dnsmasq)
(colord)
(speech-dispatcher)
(hplip)
(kernoops)
(pulse)
(rtkit)
(saned)
(usbmux)
(administrator)
(mysql)
(guest-krd8y3)
(guest-hhk1wu)
(guest-6ilyge)
(guest-88e2nw)
(guest-ortnat)
```

The STORE operator will write the results to a file (id.out).

/* id.pig */

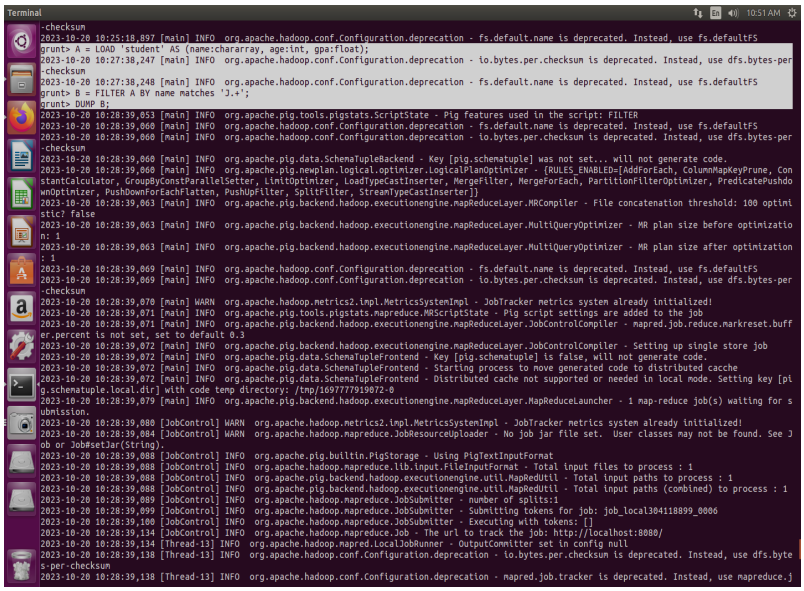
A = load 'passwd' using PigStorage(':'); -- load the passwd file
B = foreach A generate $0 as id; -- extract the user IDs
store B into '/home/hadoop/id.out'; -- write the results to a file name id.out

A = LOAD 'student' AS (name:chararray, age:int, gpa:float);

DUMP A;
(John,18,4.0F)
(Mary,19,3.7F)
(Bill,20,3.9F)
(Joe,22,3.8F)
(Jill,20,4.0F)

B = FILTER A BY name matches 'J.+';

DUMP B;
(John,18,4.0F)
(Joe,22,3.8F)
(Jill,20,4.0F)



A = LOAD 'student' AS (name:chararray, age:int, gpa:float);

B = GROUP A BY name;

C = FOREACH B GENERATE COUNT(A.age);

EXPLAIN C;

FILTER