Apache pig is a high-level procedural language platform developed to simplify querying large data sets in apache Hadoop and Map reduce. It features a “Pig Latin” language layer that enables SQL-like queries to be performed on distributed datasets within Hadoop applications.  
It allows you to write complex MapReduce transformations using a simple scripting language. Pig Latin defines a set of transformations on a data set such as aggregate, join and sort. Pig translates the pig latin into MapReduce so that it can be executed within Hadoop.

Pig runs on Hadoop and makes use of MapReduce and the Hadoop Distributed File System (HDFS). The language for the platform is called Pig Latin, which abstracts from the Java MapReduce idiom into a form similar to SQL, while Pig Latin allows you to write a data flow that describes how your data will be transformed. Since Pig Latin scripts can be graphs it is possible to build complex data flows involving multiple inputs, transforms and outputs.

Installation:

Requirements:  
1. Java  
2. Hadoop

Steps:  
1. Download a stable release of Pig  
 <http://pig.apache.org/releases.html>

2. Unpack the tarball  
 user$ tar –zvxf pig-x.y.z.tar.gz

3. Unpack hadoop-2.4.1 to your current user as well  
 user$ tar –zxvf hadoop-2.4.1.tar.gz

4. Set up your environment variables in .profile and .bashrc  
 user$ vim ~/.profile  
 user$ vim~/.bashrc  
   
 export PIG\_INSTALL=/home/usr/local/pig-0.11.1  
 export JAVA\_HOME=/usr/lib/jvm/oracle-java8  
 export HADOOP\_INSTALL=/home/usr/local/hadoop-2.4.1  
 export PATH=$PATH:$PIG\_INSTALL/bin:$HADOOP\_INSTALL/bin

5. Launch Pig  
 pig –x local  
 or pig –x mapreduce