

Deploying Pig

Table of contents

| | |
|------------------------------|---|
| 1 Requirements..... | 2 |
| 2 Building Pig | 2 |
| 3 Running Pig Programs | 2 |

1. Requirements

1. **Java 1.5.x.** preferably from Sun. Set JAVA_HOME to the root of your Java installation.
2. **Ant** build tool: <http://ant.apache.org/>.
3. To run unit tests, you also need **JUnit**: <http://junit.sourceforge.net/>.
4. To run pig programs, you need access to a **Hadoop cluster**: <http://lucene.apache.org/hadoop/>.

2. Building Pig

1. Check out pig code from svn: `svn co http://svn.apache.org/repos/asf/hadoop/pig/trunk`.
2. Build the code from the top directory: `ant`. If the build is successful, you should see `pig.jar` created in that directory.

3. Running Pig Programs

There are two ways to run pig. The first way is by using `pig.pl` that can be found in the `scripts` directory of your source tree. Using the script would require having Perl installed on your machine. You can use it by issuing the following command: `pig.pl -cp pig.jar:HADOOPSITEPATH` where `HADOOPSITEPATH` is the directory in which `hadoop-site.xml` file for your Hadoop cluster is located. Example: `pig.pl -cp pig.jar:/hadoop/conf`

The second way to do this is by using java directly: `java -cp pig.jar:HADOOPSITEPATH org.apache.pig.Main`

Regardless of how you invoke pig, the commands that are specified above will take you to an interactive shell called `grunt` where you can run DFS and pig commands. **The documentation about grunt will be posted on wiki soon.** If you want to run Pig in batch mode, you can append your pig script to either of the commands above. Example: `pig.pl -cp pig.jar:/hadoop/conf myscript.pig` or `java -cp pig.jar:/hadoop/conf myscript.pig`.