**Setup:**

All programs are deployed at:

* /home/akash001/tools
* Logging Directory /home/akash001/log4j

|  |  |  |  |
| --- | --- | --- | --- |
| Machine | Filter Running | Other Cron Jobs | Tools |
| dblab-rack1 | 1pc (One Percent) |  | keywordtool.sh |
| dblab-rack2 | ndq (NASDAQ) |  | keywordtool.sh To manage filters and keywords |
| dblab-rack3 |  | Image Download  (23:59 each day) | keywordtool.sh |
| dblab-rack4 | nse (NYSE) |  | keywordtool.sh |
| dblab-rack5 |  |  |  |

**Note:** Don’t touch /home/akash001 unless you have to. The tools can be deployed in any folder.

**Importing Filter data**

On dblab-rack2: ./keywordtool.sh < filter-name> <file>

1. filter-name: a name for a filter.
2. file: a text file with one keyword per line.

Keywordtool.sh has other options to add, remove and list keywords. Check options by running ./keywordtool.sh

**Running Filters:**

1. Compile the TwitterDownload eclipse project (see instructions below), by running build.xml. This will create a file twitter-download.zip. Copy it onto a free machine (need not be one of the above).
2. Unzip a file into a new directory.
3. Run ./twitterproc.sh <filtername>

Stopping filter :

Use command []: pgrep –u root –f jsvc |xargs kill -9

**Eclipse Workspace Setup**

Make sure you have subeclipse plugin installed: [http://subclipse.tigris.org/](http://subclipse.tigris.org/servlets/ProjectProcess;jsessionid=D01C1010FA3DFBEB47C220BF488C550E?pageID=p4wYuA)

There are two elipse projects, Import them in order:

1. <https://svn1.cs.ucr.edu/akash001/cassandra-apis>
2. <https://svn1.cs.ucr.edu/akash001/twitterDownload>

In eclipse

1. File -> Import -> Checkout Projects from SVN . Click Next
2. Create a new repository location. Click Next
3. Set Url to <https://svn1.cs.ucr.edu/akash001/cassandra-apis> , Click Next
4. Select the project root directory and click Finish.

Repeat the steps for the twitterdownload project.

**Data Model Description**

In Cassandra a “table” a column family. Each column family is a list of columns each of which can be a column family. A column family which contains other column families is called super column columns.