HadoopHackathon20110518

Apache Hadoop Hackathon, May 18, 2011

Hosted at Cloudera's San Francisco and Palo Alto offices.

This page is aliased at: http://bit.ly/hadoop-hack-may18

Useful resources

- HowToContribute
- EclipseEnvironment
- Previous hackathon notes: http://bit.ly/hadoop-hack-may11
- Eli's build scripts: https://github.com/elicollins/hadoop-dev

Quick Start

Checking out Hadoop:
Git:

```
mkdir hadoop-git ; cd hadoop-git

git clone https://github.com/apache/hadoop-common.git
git clone https://github.com/apache/hadoop-hdfs.git
git clone https://github.com/apache/hadoop-mapreduce.git

(or if we fix ssh: #git clone git://git.apache.org/hadoop-common.git
#git clone git://git.apache.org/hadoop-mapreduce.git
#git clone git://git.apache.org/hadoop-hdfs.git
)
```

svn:

```
mkdir hadoop-svn ; cd hadoop-svn
svn co https://svn.apache.org/repos/asf/hadoop/common/trunk
svn co https://svn.apache.org/repos/asf/hadoop/mapreduce/trunk
svn co https://svn.apache.org/repos/asf/hadoop/hdfs/trunk
(for trunk -- for branches, use /repos/asf/hadoop/common/branches/branch-0.22 )
```

Running tests

```
ant test-core -Dtest.output=yes -Dtestcase=TestEditLog
```

test.output will print output to console, useful for hanging tests

Eclipse: see EclipseEnvironment

Submitting a patch

```
Open a jira
Make change
Run tests
git diff --no-prefix > /tmp/HADOOP-1234.txt
```

Review queues

- Common: https://issues.apache.org/jira/secure/IssueNavigator.jspa?requestId=12311124&mode=hide
- HDFS: https://issues.apache.org/jira/secure/IssueNavigator.jspa?requestId=12313301&mode=hide
- MapReduce: https://issues.apache.org/jira/secure/IssueNavigator.jspa?requestId=12313302&mode=hide

Suggestions for what to work on

Infrastructure improvements

- Create a Hudson job that produces a release tarball: https://builds.apache.org/hudson/view/G-L/view/Hadoop/job/Hadoop-22-Build/
- Include 32-bit and 64-bit native libraries in Jenkins tarball builds: https://issues.apache.org/jira/browse/HADOOP-7283

Make it easier for others to contribute

- Improve documentation at HowToContribute, EclipseEnvironment
- Write instructions for other IDEs
- What's the most confusing thing you found about the contribution process? How can we improve it?

Help get 0.22 out the door

- Close out 0.22 blockers. Perhaps more appropriate for people with context.
- If those are too hard check out the other jiras for 0.22 Common, HDFS, MapReduce

Try to use the release (or build from trunk)

- Work on the documentation
    - Try out the current documentation
    - File jiras and submit fixes for bugs and improvements.
    - Eg config options that should be in the docs but are not..
    - Or have been deprecated and should be removed or updated.
    - Write new documentation that's needed (eg on FS config)
- Setup a small cluster on your laptop or in VMs or using Apache Whirr and bang on it.

Help get trunk in shape

- Help out with the SVN unsplit: https://issues.apache.org/jira/browse/HADOOP-7106. Git expertise is welcome!
- Review/commit patches in the review queues:
    - Common
    - HDFS
    - MapReduce
- Work out the kinks of HBase trunk on HDFS trunk
    - Eg HDFS-1103, HDFS-1152, HDFS-1139, HDFS-1056, HDFS-1060.
- Improve error and log messages
- Improve command line usability (eg error messages)
- Newbie JIRAs