

Project Suggestions

Here are some suggestions for some interesting Hadoop projects. For more information, please inquire on the [Hadoop mailing lists](#). Also, please update and add to these lists.

For good small JIRAs to get started on, see [this list of newbie jiras](#) and [this list of test failures](#). There is also the list of [all open Hadoop Issues with no patch](#).

1. [Test Projects](#)
2. [Research Projects](#)
3. [Tool Investigations](#)
4. [Random Ideas](#)

Test Projects

Rough estimates are given in hours. These estimates assume an existing understanding of Hadoop.

Description	Estimate	Links
write Junit test cases that run the Hadoop examples	16	
MapRed reliability tests	40	HADOOP-2483
HDFS reliability tests	40	HADOOP-2483
refactor TestDFSUpgradeFromImage (once HADOOP-1622 is committed) to auto zip and unzip the supporting DFS image	6	
write compatibility tests for reading the same data set from different HDFS versions	24	
re-write (or drop) flaky TestMiniMRWithDFS unit test	8	
write the "system tests" for DFS Upgrade	20	DFS Upgrade Test Plan
write new unit tests based on code coverage	40	Nightly Code Coverage Report
pipes and libhdfs benchmark tests	30	
review Findbugs warnings and fix the reasonable warnings	18	Nightly Findbugs Warnings
create a distributed JUnit runner on top of Hadoop	80	HADOOP-1257
implement a Map-Reduce application which can be used to reliably launch speculative tasks	40	HADOOP 2214

Research Projects

Check out this page of [Hadoop Research Projects](#).

Tool Investigations

We are always looking for open source testing tools that add value to our development and build process. Here are some that need to be investigated.

Description	Links
evaluate new unit test frameworks (rewriting some existing test in the new framework to show benefits)	JUnit4 , TestNG
evaluate mock object frameworks for unit testing	JMockit , EasyMock , JMock
evaluate PMD	PMD
evaluate concurrency test tools	PathFinder , ConTest , CheckUncontendedLock , MultithreadedTC
evaluate Fortify	Fortify Tools
evaluate dashboards like QALab and Panopticode	QALab , Panopticode
evaluate Faban	Faban , Blog
evaluate NCSS, a source code metrics suite	JavaNCSS
evaluate Java PathFinder, a software model checker	JavaPathFinder
evaluate SA4J, a structural dependency analysis	SA4J
evaluate JDepend, generates design quality metrics	JDepend

evaluate Dependency Finder, generates design quality metrics and dependency graphs	DepFind
evaluate Classycle, finds class and package cyclic dependencies	Classycle
evaluate XRadat, an extensible code report tool	XRadat
evaluate Crap4j, combines cyclomatic complexity and code coverage	Crap4j
evaluate Eclipse TPTP, a test and performance tools platform	TPTP
evaluate JUnitFactory	JUnit Factory
evaluate code review applications like Codestriker and Review Board	Codestriker , Review Board
evaluate test automation frameworks	STAF
evaluate QA management platform	Sonar

Random Ideas

Description	Estimate	Links
Implement a advanced <i>job control</i> framework to help chain multiple Map-Reduce jobs i.e. investigate/improve upon existing org.apache.hadoop.mapred.jobcontrol package.	tbd	
Implement a library/framework to support Genetic Algorithms on Hadoop Map-Reduce.	tbd	
Improve the Eclipse Plugin	tbd	