

dCache as open-source project showcase for education

Tigran Mkrtchyan for dCache team

CHEP2018, Sofia, 11.07.2018



Nordic e-Infrastructure
Collaboration



eXtreme DataCloud



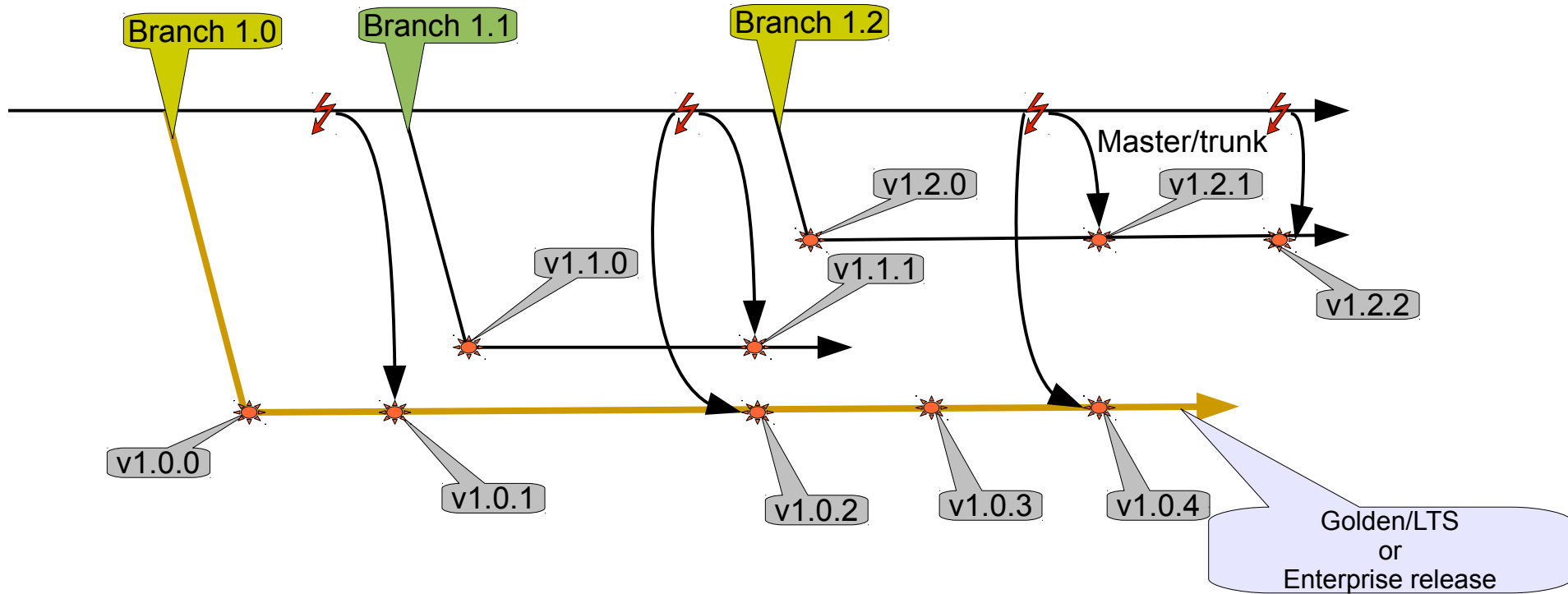


- 76 installations for WLCG
- ~10 no GRID installations
- ~50% of LHC data

Some numbers

- ~60 developers/contributors in 18 years
 - DESY
 - FNAL
 - NDGF
- 25 modules
- 61 production branches (since 01.11.2007)
- 926 releases (since 01.11.2007)
- ~400K lines of code in ~2600 files
- ~100 external dependencies
- RPM, DEB, TAR, Solaris and Docker packages

Version management



Release Policy

- Yearly long-term branch
 - 2+ years support
- Time-based feature released
 - ~every 3 months
 - supported until next LTS release
- Weekly maintenance releases
 - release for all supported branches

The “**Release Process**”
gives a common framework to
get stuff done with a
**predictable and
reproducible result.**

Code policy

- master/dev
 - **ALWAYS** compilable
 - **ALMOST** deployable
- Branch
 - **ALWAYS** releasable

Course goal

- Data intensive science, Big Data
- Introduction to Open Source project
- Use of standard tools
 - Students should be ready to join any project
 - github
 - Maven
- Developer team interaction
 - Code review
- Release management

Project adjustments

- Standard project structure
 - Convention vs. Configuration
- Standard build environment
 - Make project IDE ready
- Standard workflow
- Simple instruction how to build
 - BUILDING.md

Development effort guideline

Effort

	Easy to do	Difficult to do
Impact		
Major	1	3
Minor	2	4

1. Implement immediately
2. Implement immediately
3. Long term plans and development
4. Discarded

Development effort guideline

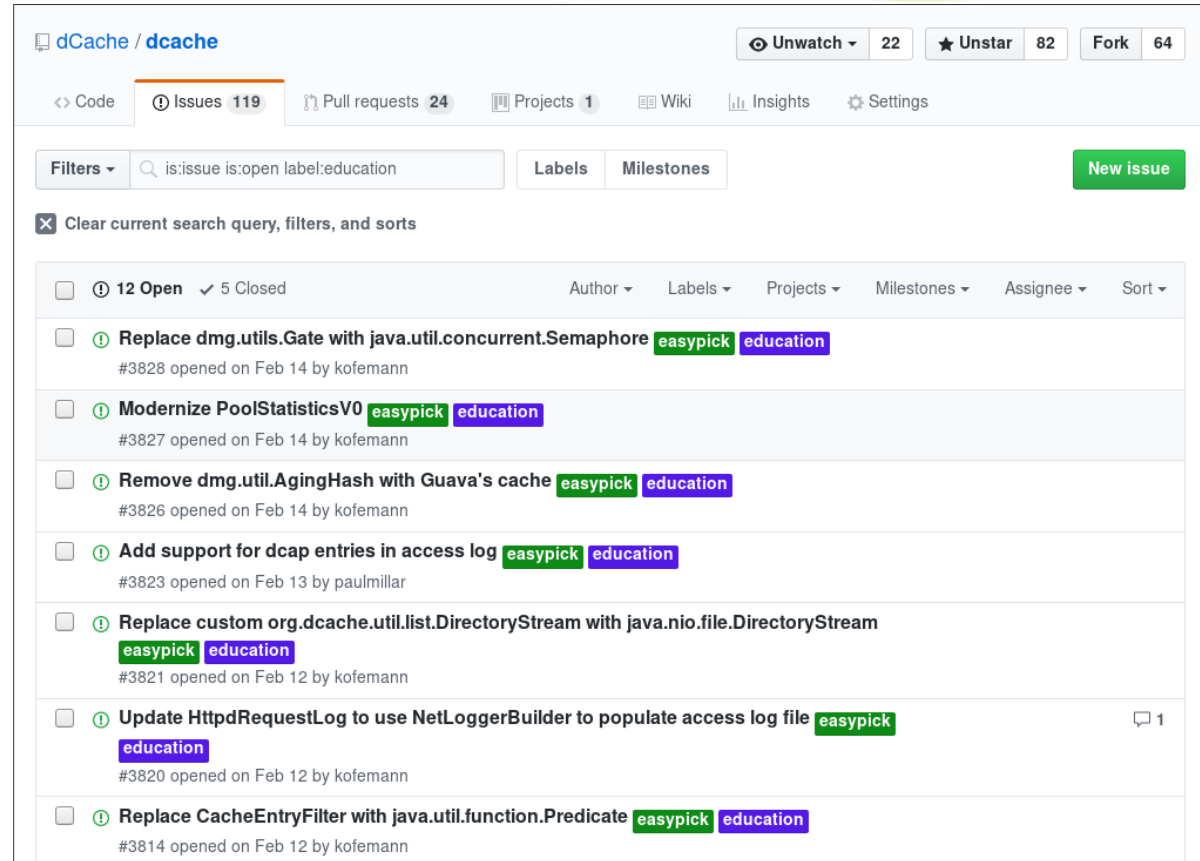
		Effort	
		Easy to do	Difficult to do
Impact	Major	1	3
	Minor	2	4

1. Implement immediately
2. Implement immediately
3. Long term plans and development
4. Discarded

Students work target

Student assignments

- Simple development task
 - Mostly in a single file
 - No project prior knowledge is required
- Collected over the time
 - (easy with minor impact)
- Formed as GitHub issues
 - **easypick** / **education**
- Goal: merge of a **pull request**

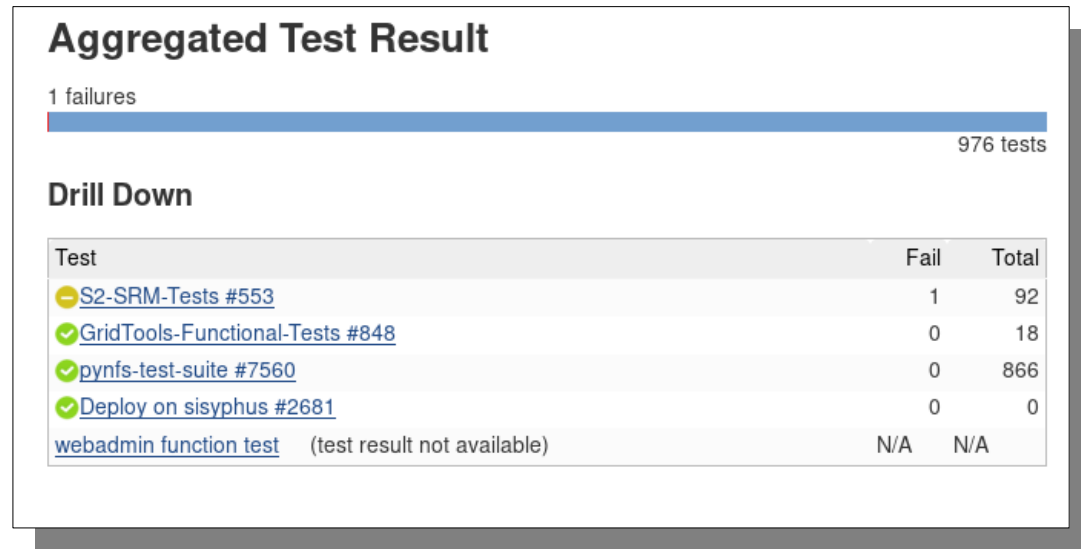


The screenshot shows the GitHub interface for the dCache repository. At the top, it displays the repository name 'dCache / dcache' and navigation options like 'Code', 'Issues 119', 'Pull requests 24', 'Projects 1', 'Wiki', 'Insights', and 'Settings'. Below this, there are search filters and a 'New issue' button. The main content is a list of 12 open issues, each with a title, a status icon, and labels. The issues are:

- 12 Open** ✓ 5 Closed
- Replace dmg.util.Gate with java.util.concurrent.Semaphore** **easypick** **education** #3828 opened on Feb 14 by kofemann
- Modernize PoolStatisticsV0** **easypick** **education** #3827 opened on Feb 14 by kofemann
- Remove dmg.util.AgingHash with Guava's cache** **easypick** **education** #3826 opened on Feb 14 by kofemann
- Add support for dcap entries in access log** **easypick** **education** #3823 opened on Feb 13 by paulmillar
- Replace custom org.dcache.util.list.DirectoryStream with java.nio.file.DirectoryStream** **easypick** **education** #3821 opened on Feb 12 by kofemann
- Update HttpRequestLog to use NetLoggerBuilder to populate access log file** **easypick** **education** #3820 opened on Feb 12 by kofemann
- Replace CacheEntryFilter with java.util.function.Predicate** **easypick** **education** #3814 opened on Feb 12 by kofemann

Commit/PR workflow

- Full Build/Deploy/Test
 - Code analysis
 - Unit tests
 - Functional tests
- Fail Fast



Good COMMIT

- Does only one thing
 - easy to review
 - easy to move around
 - easy to revert
- As small as possible
 - “your diff does not fit on my screen”
- Informative log message
 - ``git blame`` is your friend

Contributor agreement

- External contributors may sneak proprietary code
- Have clear contribution policy
- dCache uses linux kernel model
 - **`Signed-off-by:`** annotation
- Check with your legal department!

Example:

dcache: added a decorator for RepositoryChannels to get IO statistics (#3430)

Motivation:

Getting a possibility to generate statistics and spot misbehavior in IO processes of RepositoryChannel.

Modification:

Added package statistics, created the class IoStatisticsChannel as decorator for a RepositoryChannel and created the class IoStatistics

which provides statistics on number of read and write requests, read and write IO block sizes and different measures about the read and write IO speed.

Result:

Possibility to add a decorator to a RepositoryChannel which provides statistics about IO processes.

Signed-off-by: Lotta xxx <xxx@htw-berlin.de>

Make my code better!

```
modules/chimera/src/main/java/org/dcache/chimera/JdbcFs.java
Revision 2b96f0fc15f213bcfa93ab8a40e65ec3551d3040 New Change
+ 400 lines
+20
401 public FsInode createFile(FsInode parent, String name, int owner, int group, int mode, int type) throws ChimeraFsException {
402     try {
403         checkNameLength(name);
404         if (!parent.exists()) {
405             dbConnection
406                 .stat(parent
407                     ._sqlDriver.stat(dbConnection
408                         .if (parent
409                             .throw
410                                 FileNotFoundHimeraFsException("parent=" +
411                                     parent.toString());
412                             }
413                         if (parent.isDirectory()) {
414                             if ((parent
415                                 .UnixPermission.S_IFDIR) !=
416                                     // read/write only
417                                     dbConnection.setAutoCommit(false);
418                                     if ((parent.statCache().getMode()
419                                         & UnixPermission.S_ISGID) != 0) {
420                                         group =
421                                             parent.statCache().getGid();
422                                         }
423                                         inode =
424                                             _sqlDriver.createFile(dbConnection, parent, name,
425                                                 owner, group, mode, type);
426                                     }
427                                     }
428                                     }
```

rb.dcache.org/r/6312/

Status: Open From: All

There are no open issues

Expand All Collapse All

Karsten Schwank Posted 2 months ago (Dec. 5, 2013, 11:44 a.m.)

Wouldn't this be nicely unit-testable?

Add comment

```
modules/chimera/src/main/java/org/dcache/chimera/JdbcFs.java (Diff revision 1)
public FsInode createFile(FsInode parent, String name, int owner, int group, int mode, int type) throws
ChimeraFsException {
    408     if (parent == null) {

```

should this really be parent and not parentStat?

✓ The issue has been resolved. Re-open Show all issues

Tigran Mkrtchyan 2 months ago (Dec. 6, 2013, 11:25 a.m.)

good catch

Add comment

+ Review request changed Updated 2 months ago (Dec. 6, 2013, 11:25 a.m.)

Goals of code review

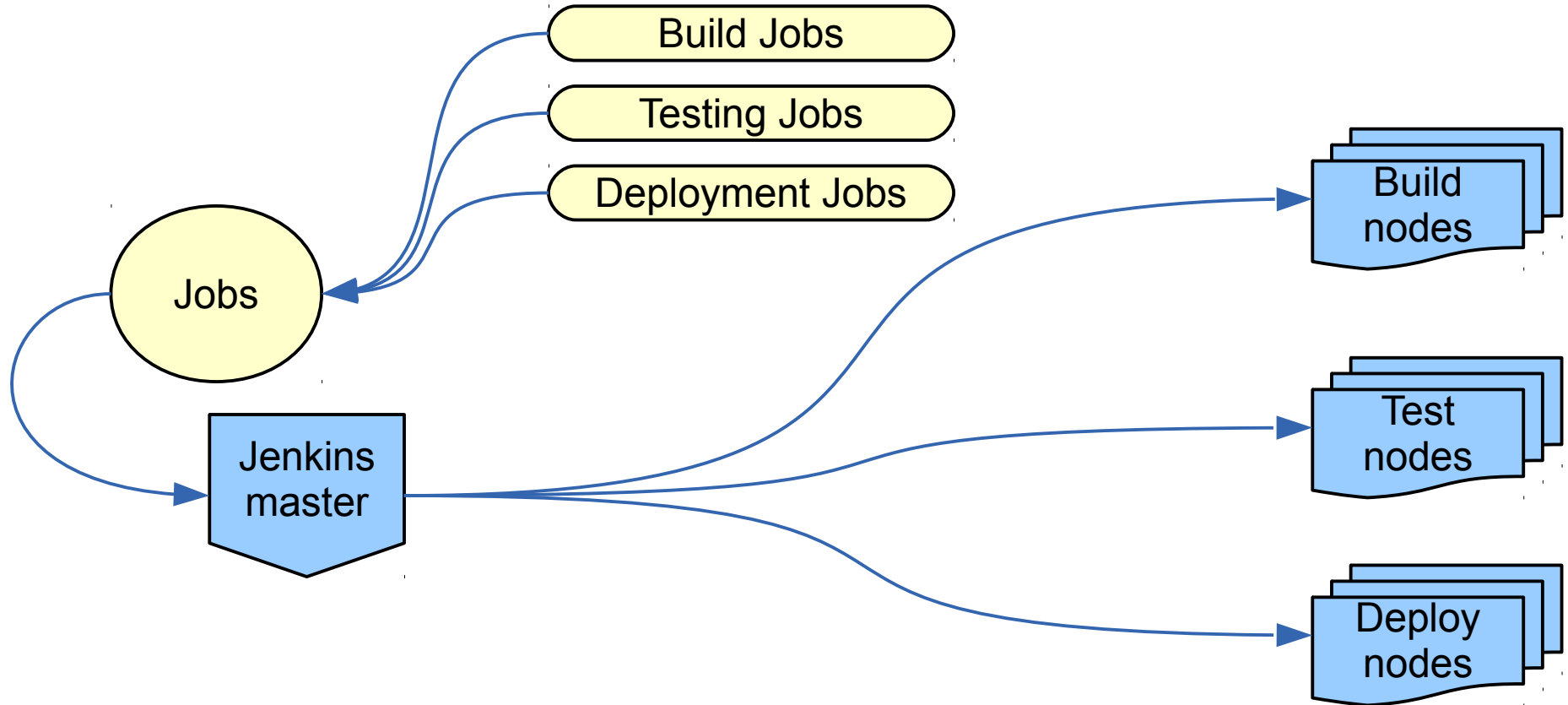
- Let others see my code
 - Forces me to publish better code
- Learn best-practice
 - New libraries used
 - New technique used
- Learn about code changes (by others)
 - reduce 'bus factor'

Conclusions

- Use standard tools to attract students
 - make them benefit from new skills
- Make your infrastructure robust for external contributors
 - ensure reliable testing and deployment
- Clear license and contribution policy
- Be open for new contributors
 - simple tasks for newcomers
 - Document build process

Thank You!

Build system setup




Task: #4
dCache version 2.6 have switch to a new version of Berkeley DB, which was binary incompatible with the used one. A migration script has been added, which is not necessary any more (this task requires git-based archaeology)

Description:
remove file `skel/share/misc/je-4.1.21.jar` with corresponding

Starting point: use `git log` to find the commit v
added `skel/share/misc/je-4.1.21.jar`


Task: #9
To spot misbehaving applications and understand data access profile a low lever statistics on IO requests is required.

Description:
Create a `RepositoryChannel` decorator which can provide statistics on number of requests, IO speed and io block sizes

Starting point: `RepositoryChannel` interface and Decorator design pattern


Task: #12
When inspecting active transfers a 'show transfers' command in the dCache's admin interface is used. On a busy system, the output of the command may return more lines than can be visually inspected.

Description:
Enhancements to NFS door's admin interface to filter active transfers by client, `pnfsid` and pool.

Starting point: `NFSv41Door` class and java8 stream API.

Commit message

module: summary line (up to 60 chars)

(empty line here)

Motivation:

(explain why we need this change)

Modification:

(explain what was change)

Result:

(what will be different now)

Some extra info:

(bug tracker id, branch back-port)

Commit message

ssh: fix handling of idle connection timeout

Motivation:

To preserve idle ssh connection one have to adjust IDLE_TIMEOUT, however sshd has yet an other timeout - READ timeout ...

Modification:

Explicitly set read timeout to prevent it to fire before idle timeout.

Result:

idle ssh session is not closed

Ticket: #9175

Target: master, 3.1