



# Improvements to the OpenOffice.org release engineering process

**Jens-Heiner Rehtien**

- > Software developer, RE
  - Sun Microsystems, Inc.

# Overview

- OpenOffice.org SCM: migration to Mercurial
- Build 2.0, or how to build fast – even on Windows
- Future plans
- Q&A

# The history of dealing with history

- 2001-2008: CVS
  - > tagging and branching are  $O(n)$  operations
  - > history preserving renames are awkward
  - > binary file handling is error prone
  - > commit operations are not atomic
  - > manual merge tracking required
  - > the central repository was rapidly deteriorating

but also

- > all quirks well known, sort of reliable
- > easy to work with subtrees

# A first look for an alternative

- Q4 2007: Evaluation of alternatives
  - > Subversion (<http://subversion.tigris.org>)
  - > Git (<http://git.or.cz>)
  - > Bazaar (<http://bazaar-vcs.org/>)
  - > Mercurial (<http://www.selenic.com/mercurial/wiki>)
- ESC meeting Feb. 2008
  - > DSCMs were not ready for OOo (Git was near, though)
  - > migrate to SVN as an intermediate step
  - > re-evaluate the DSCMs within a year

# A detour?

- 2008-2009: Subversion
  - > fixed many of the perceived weaknesses of CVS
  - > but there were some serious disappointments, too
    - merge tracking broken: svn clients ver. 1.5.0-1.5.4
    - unreliable:
      - broken working copies
      - reverse merges
    - slow merges
    - file moves: can result in loss of fixes
    - SSH authentication takes 3s for round trip
    - SVN URLs are clumsy to work with

# One year later: a fresh look at DSCMs

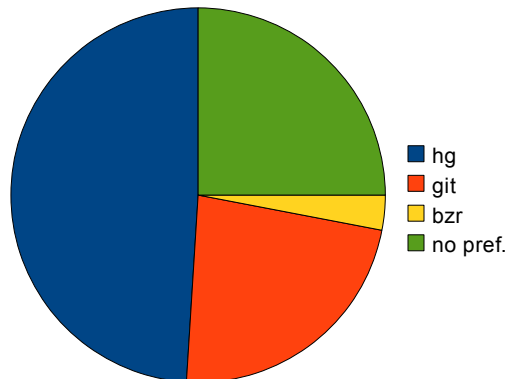
1Q 2009: DSCM

improvements all over  
the place

- all
  - > performance
  - > functionality
- Git
  - > some commands rewritten in C
  - > better windows support
  - > vastly more avail. docs
- Mercurial (hg)
  - > impr. copy/rename handl.
  - > named branches
  - > extens. to deal with CR/LF
  - > impr. import from SVN
- Bazaar (bzd)
  - > stacked branches
  - > new repository format
  - > improved merge
  - > updated server protocols
  - > improved hooks

# But which one to choose?

- ESC meeting Mar. 2009
  - > declares git & hg fit for OOo and bzd nearly so (still import problems – new rep. format came late)
  - > git vs. hg: about as easy to decide as vi vs. emacs
  - > evaluation team slightly prefers hg due to ease of use
  - > developer survey and pilot



Contributors	149
hg	49%
git	23%
bzd	03%
no preference	25%

# The Mercurial pilot was succesful

- Oct. 2009-... hopefully for a long time: Mercurial
  - > DEV300 (OOo 3.3) code line first migrated
  - > OOO320, OOO310 ... will follow soon

# The Mercurial pilot was successful

- Oct. 2009-... hopefully for a long time: Mercurial
  - > DEV300 (OOo 3.3) code line first migrated
  - > OOO320, OOO310 ... will follow soon

Many thanks to the Mercurial pilot team!!

Great discussions, many suggestions  
and lots of testing.

# The source tree is still huge ...

- Some facts about the repository (DEV300)
  - > more than 260,000 revisions
  - > 69,000 (active) files
  - > 1.3 GiB in size (just the store)

# The source tree is still huge ...

- Some facts about the repository (DEV300)
  - > more than 260,000 revisions
  - > 69,000 (active) files
  - > 1.3 GiB in size (just the store)

→ transfer the bulk of changesets only once!

# The source tree is still huge ...

- Some facts about the repository (DEV300)
  - > more than 260,000 revisions
  - > 69,000 (active) files
  - > 1.3 GiB in size (just the store)

→ transfer the bulk of changesets only once!

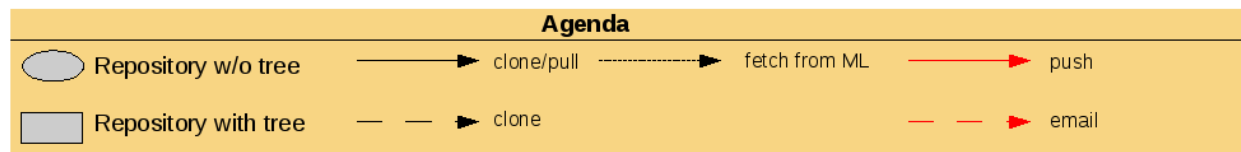
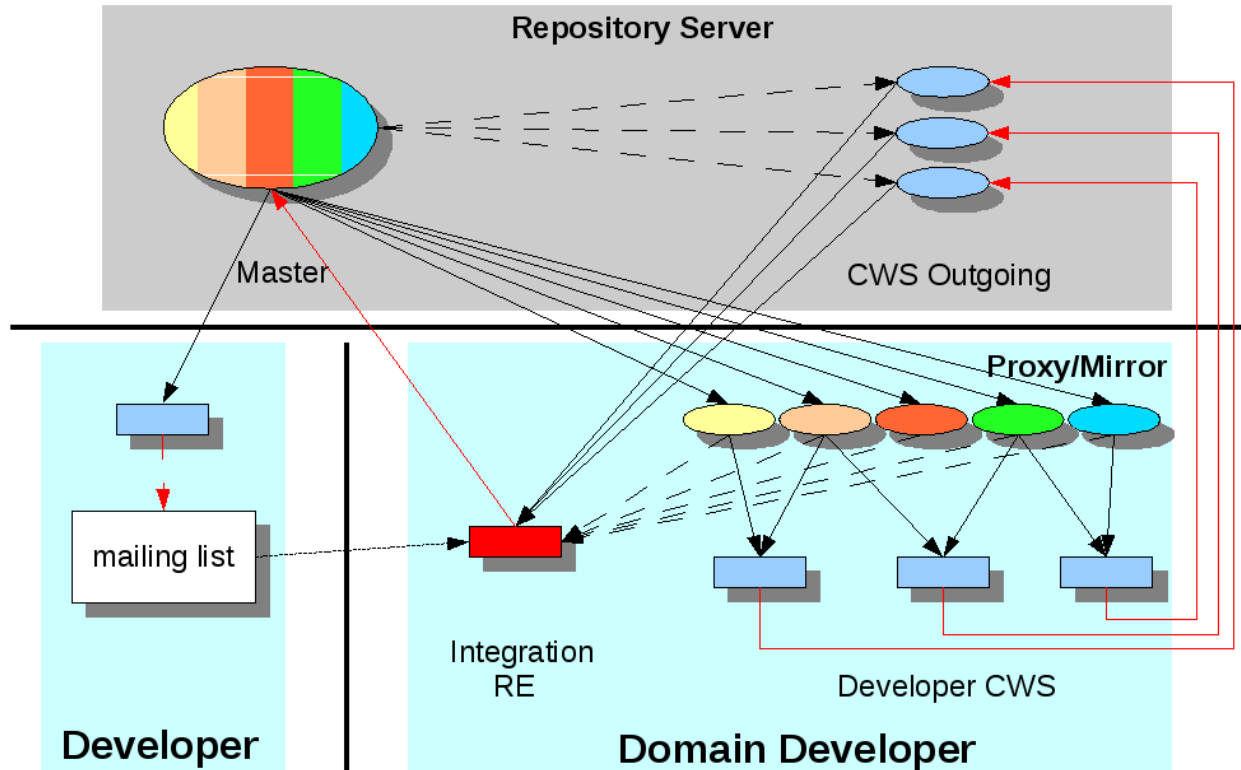
Preferably by downloading the nightly bundle (800MB):

<http://hg.services.openoffice.org/bundle/DEV300.hg>

```
hg init DEV300; cd DEV300
```

```
hg unbundle <path_to_bundle>/DEV300.hg
```

# OpenOffice.org Mercurial setup



# 'Outgoing' repositories

- Collaboration central: development, QA and RE
  - > “treeless” repositories
  - > has all historical changesets up to creation milestone
  - > disk space efficient on the server
  - > available 1 h latest after CWS creation
  - > R/O: <http://hg.services.openoffice.org/cws/foo42>
  - > R/W: `ssh://hg@hg.services.openoffice.org/cws/foo42`
- Never change a published changeset
  - > except you can notify everyone who pulled from the CWS
- Do not add new heads to the outgoing repository
  - > unless you really know what you are doing

# Tooling

- Create a CWS in EIS or with:
  - > `cws create --hg -c foo42`
- Working copy:
  - > with hg client
    - `hg clone <local_DEV300> <work>`
    - `hg --cwd <work> pull -u http://hg.services.openoffice.org/cws/foo42`
  - > or with cws tool: `cws fetch -c foo42 <work>`  
 default-push configured, hook for suppressing CR/LF
- Update to a new milestone
  - > `hg pull; hg merge; hg resolve -m; hg commit;`
  - > `cws setcurrent -m <new_milestone>` or use EIS

# Conventions

- OOo account name in user name
  - > Joe Developer <jdeveloper@openoffice.org>
  - > Joe Developer [jdeveloper] <joe@example.com>
- Enable CR/LF detection
- Commit message:
  - > prefix with CWS name followed by issue number
    - “foo42: #i4711#: implement 'foo' feature”
    - “foo42: merge with DEV300 m70”
  - > long description: stand alone summary line

# Conventions

- OOo account name in user name
  - > Joe Developer <jdeveloper@openoffice.org>
  - > Joe Developer [jdeveloper] <joe@example.com>
- Enable CR/LF detection
- Commit message:
  - > prefix with CWS name followed by issue number
    - “foo42: #i4711#: implement 'foo' feature”
    - “foo42: merge with DEV300 m70”
  - > long description: stand alone summary line
- **do not add heads, do not change publ. changesets**

# Build 2.0

- Team:
  - Mathias Bauer
  - Björn Michaelsen
  - Hans-Joachim Lankenau
  - Hassan Ilter
  - Jens-Heiner Rechtien
- Look thoroughly at each aspect of OOo building
  - > easier
  - > faster
  - > less error prone

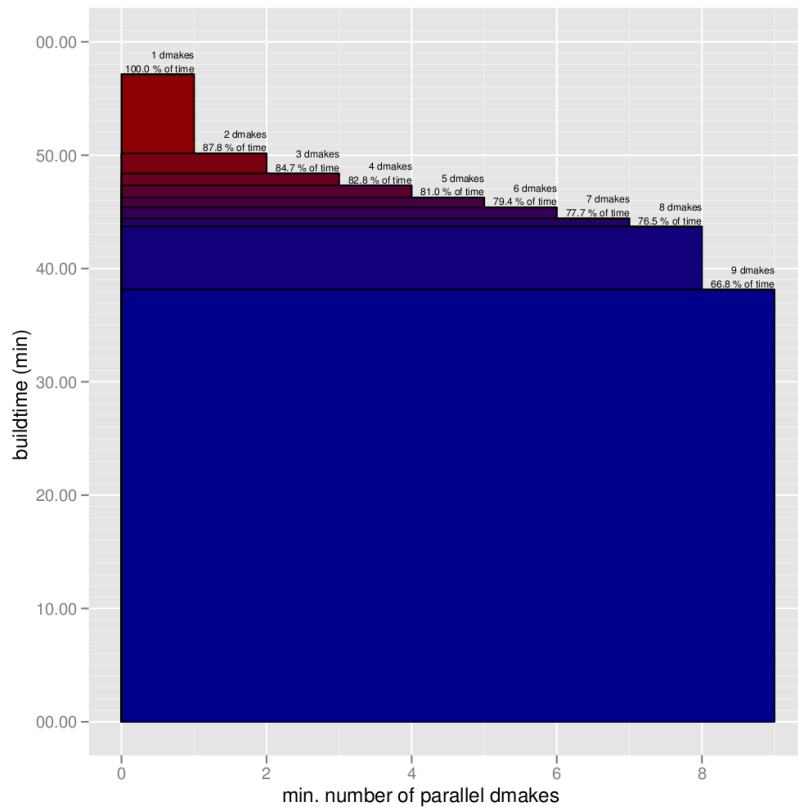
# Build 2.0 efforts

- I10n clean up: move localization to single module
- create language packs without building OOo
- remove 4NT shell support in favor of cygwin
- rework build documentation
- support building with multiple repositories (“source config”)
- build performance

# Hurts most: building on Windows

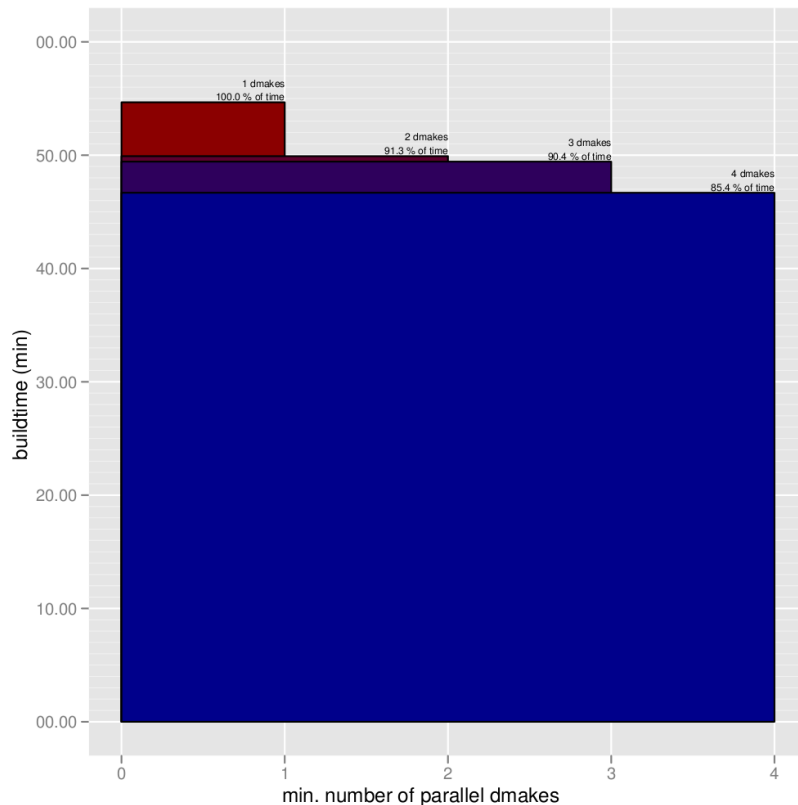
- Build less: good but hard to achieve
  - > smart dependencies vs. correctness
  - > pre-compiled header vs. ccache
  - > split build
- Build faster: parallelize the build
  - > observation: remote builds completely I/O bound
  - > local disk build: CPU or I/O bound?
    - Sun Fire 2200 (2x4 cores, 64GB RAM):  
with 32GB Ramdisk vs. local disk 20% faster (1h 20min)  
→ mostly CPU bound?

# Parallel builds: build -P9 -- -P1



- Linux build
- --with-system-libs
- Sun Ultra 24  
(3.0 GHz Quad, 8 GB RAM)
- serial build: 3:08:01
- th. par. build: 0:47:01
- parallel build: 0:57:09
- overhead: 21.5%

# Parallel builds: build -P4 -- -P4



- Linux build
- --with-system-libs
- Sun Ultra 24  
(3.0 GHz Quad, 8 GB RAM)
- serial build: 3:08:01
- th. par. build: 0:47:01
- parallel build: 0:54:40
- overhead: 16.3%

# Plans for the future

- URL:  
[http://wiki.services.openoffice.org/wiki/Build\\_Environment\\_Effort](http://wiki.services.openoffice.org/wiki/Build_Environment_Effort)
- Soon: move third party libraries out of the repository
  - > 'blobs' are detrimental for DSCM repositories
  - > will use ftp-server and a fetch/cache mechanism
  - > requirement: build without any network connection must still be possible (Debian).
- Split build

# Plans for the future (continued)

- Split repository: maybe after branch off OOo 3.3
  - > more repositories, more overhead
  - > makes only sense if parts can really be separately build
  - > likely candidates: helpcontent, l10n, binfilter
- Plugin for hg as replacement for “cws”

# Q & A



# Improvements to the OpenOffice.org release engineering process

**Jens-Heiner Rehtien**

> [jens-heiner.rehtien@sun.com](mailto:jens-heiner.rehtien@sun.com)