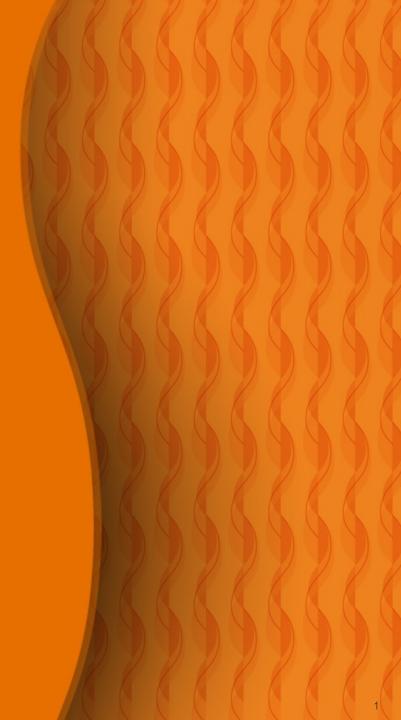# The state of the OpenOffice.org SCM migration project

- **Jens-Heiner Rechtien**
  - Software developer, RE
    - Sun Microsystems, Inc.

Overview

- Why change a running system?
- Evaluating four candidates
- Server and repository setup
- Migration decisions
- CWS-tools revisited
- Problems? Nah …
- Q&A

After 8 years of working with CVS ...

CVS is aging

- tagging and branching are O(n) operations
- history preserving renames are awkward
- binary file handling is error prone
- commit operations are not atomic

8 years working with CVS left us with:

- ~ 22 GB repository size (+5 GB last year)
- ~ 7000 branches (child work spaces!)

... it's time to move to a new SCM

SCM based on change sets instead of file archives

But which one?

- Centralized SCM
    - > Subversion (http://subversion.tigris.org)

- Distributed SCM
    - > Mercurial (http://www.selenic.com/mercurial/wiki)
    - > Git (http://git.or.cz)
    - > Bazaar (http://bazaar-vcs.org/)

After a long evaluation phase ...

- The new SCM
  - > must support the OOo way of development
    - – merge tracking!
  - > be able to import major parts of the history
  - > have a significant better overall performance
  - > known to be stable
  - > should be well documented
  - > should have an understandable user interface
  - > should be scriptable
  - > should integrate with IDEs and other tooling

... we finally chose Subversion ...

- all SCMs evaluated had weaknesses
- Subversion
  - > centralized
  - > merge tracking still new
  - > commands take long URLs
- Git
  - > no partial checkouts
  - > windows support weak
  - > terrible user interface (IMHO)
  - > weak documentation

... at least for now

- Mercurial
  - > no partial checkouts
  - > we didn't manage to do a history import
  - > somewhat slow
  - > "named branches" were still experimental

- Bazaar
  - > no partial checkouts
  - > history import succeeded but late
  - > somewhat slow
  - > "repository branches" were still experimental

We will re-evaluate DSCMs again ...

- need to be careful not to preclude DSCMs by migration choices

- smaller repository sizes help all DSCMs
- limit the number of branches will help bzr, hg

… and if you really dislike SVN
… use a bridge

- Subversion is change set based

    => it's reasonable possible to bridge

- git-svn
    > seems to be already in use

- bzr-svn

- hg-svn

The server

Sun Fire X4150

> 8 cores (3.16GHz)

> 64 GBytes RAM

> 600 GB storage (ZFS).

Second system (same configuration) serves as backup and fail-over system

Subversion server (continued)

- can also be used for DSCM later on
- URL: svn.services.openoffice.org

- Additional services
  - > OpenGrok indexer and source browser
    http://svn.services.openoffice.org/opengrok

Subversion repository

- Subversion 1.5 (current 1.5.4)
  - > merge tracking!

- Repository structure:
  - > .../trunk
    .../tags   DEV300_m18, OpenOffice_3_0_0 ...
    .../branches   DEV300, DEV301 ...
    .../cws   swbugfix03, sb87, ...
    .../dist/...   anyone can get space here for patches etc
    .../contrib   is this needed?
    .../patches   globally useful patches

- path based access needed? Hopefully not!

Scope of migration

- Full migration, including all CWS branches
  - > not really possible, migration will take forever
- Full migr. minus "finished" CWS branches
- Trunk migration only.
- No history migration at all.
  - > probably to radical
- http://wiki.services.openoffice.org/wiki/Scm_migration_scope

Full mig. minus "finished" CWSs

- Pro
  - > Almost everything of relevance there
  - > Historic releases states can be extracted
  - > Current work checked out from new SCM
  - > Can be done, this has been tried out
  - > CVS server can be switched off (or r/o)
- Con
  - > The resulting repository will still be huge
  - > Number of revisions estimated > 500000
  - > The conversion takes about a week
  - > Quiet period for everyone
  - > Accuracy? Legacy workspaces need to be converted

Trunk only (that's what we did)

- Pro
  - > Still almost everything of relevance there
  - > Conversion is easy
  - > Repository is lean and mean
  - > Accuracy is no concern, history is for reference only
  - > No quiet period except for integrations
  - > No hassle with workspace migration
- Con
  - > CVS server remains online
  - > One time patch effort when integrating CWS
  - > Real ancient history might browsing CVS
  - > Historic state constructed from CVS

Modified CWS-tooling

- cws create [-m milest.] <mws> <cws>
  - > creates branch and registers with EIS

- cws fetch [-c CWS] [-m milest] [-s] <workspace>
  - > checks out sources in a configurable way
  - > makes use of pre-checked out source tree
  - > modules are really only top level directories now
  - > includes cwsclone functionality

- cws rebase [-C | -m milestone] <workspace>
  - > thin wrapper around subversion

Modified CWS-tooling (continued)

- cws integrate

- cws query

- cws task

- cws eisclone

- cws ...

Problems?

- Where are my modules?
  - > modules are ingrained to the build system
  - > not possible to mix modules from two or more repositories
- Merge tracking is not as fast as we would like it
  - > use 1.5.4 client
- SVN repository URLs are clumsy

Q & A

The state of the Openoffice.org SCM migration project

- **Jens-Heiner Rechtien**
  - > jens-heiner.rechtien@sun.com