

Using Git For Development

Shantanu Pavgi, pavgi@uab.edu

UAB IT Research Computing

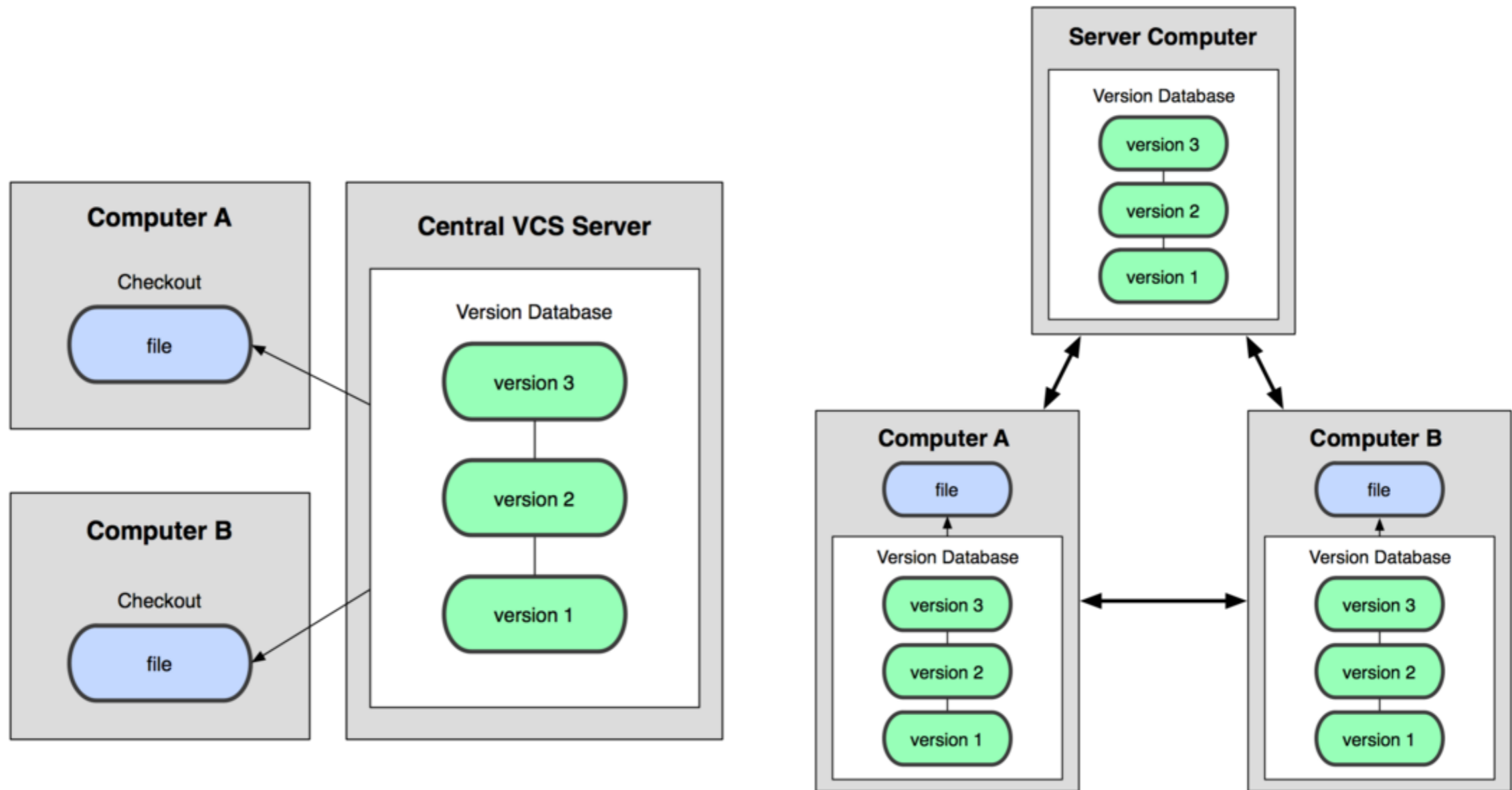
Outline

- Version control system
- Git
 - Branching and Merging
 - Workflows
 - Advantages

Version Control System (VCS)

- Recording changes to files and maintaining associated metadata
- Copy or tar command with timestamps
- Types
 - Centralized
 - Distributed

Centralized vs Distributed VCS



Reference: <http://progit.org/book/ch1-1.html>

Git

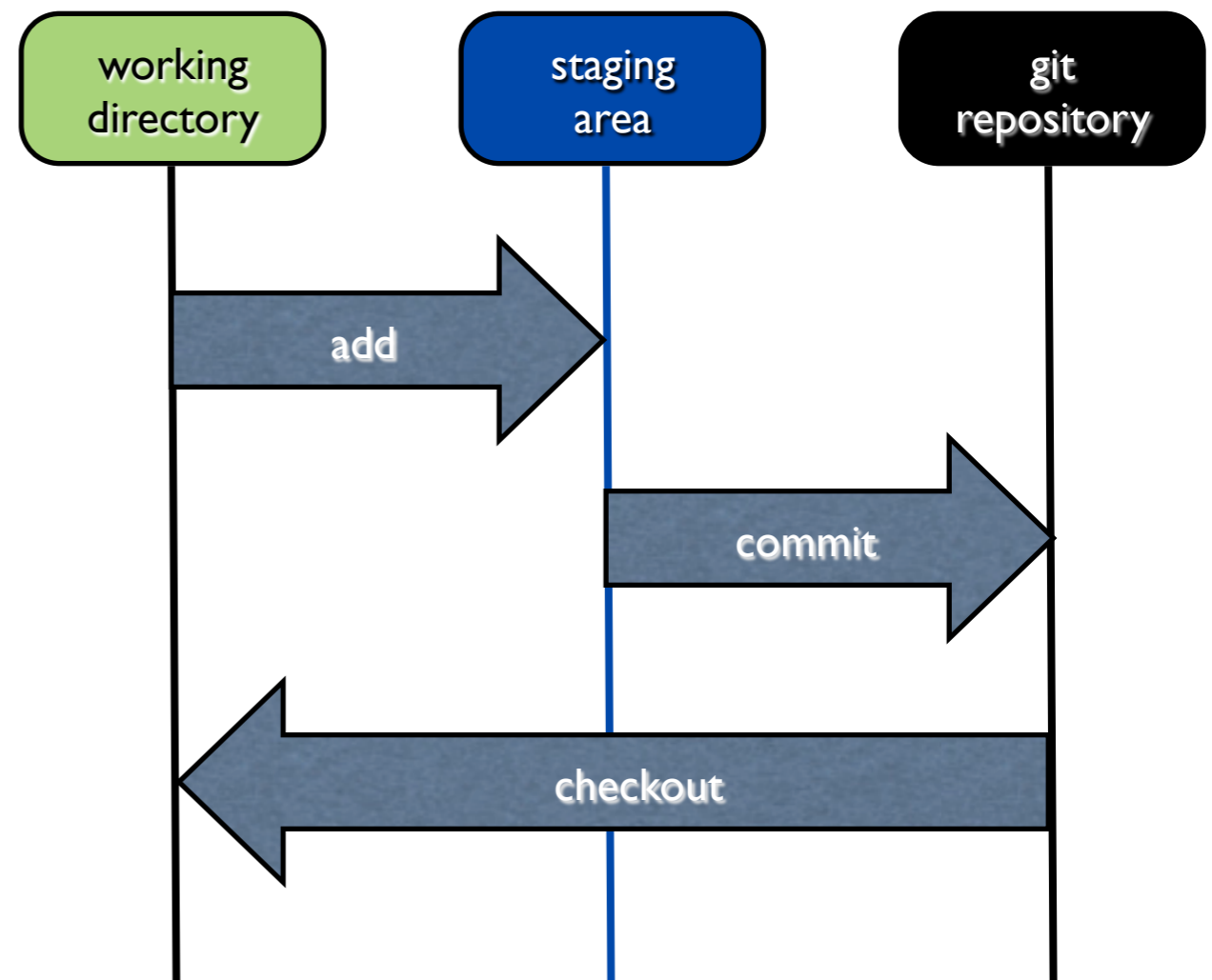
- Git is a free & open source, distributed version control system designed to handle everything from small to very large projects with speed and efficiency.
- Tool for maintaining your work history
- Used by:
 - Linux Kernel
 - Android
 - Ruby on Rails
 - Reference: <http://git-scm.org>

Advantages

- Works off-line
- Fast
- Easy or cheap branching
- Public and private work histories
- Rewrite or curate history before making it public
- Repositories can talk to each other
- Flexibility in workflows

Commits

- Tracks content and not files
- Allows you to stage changes
- Each commit is a complete snapshot of all staged files
- Unique 40 character hash for each commit



Local operations

Commits

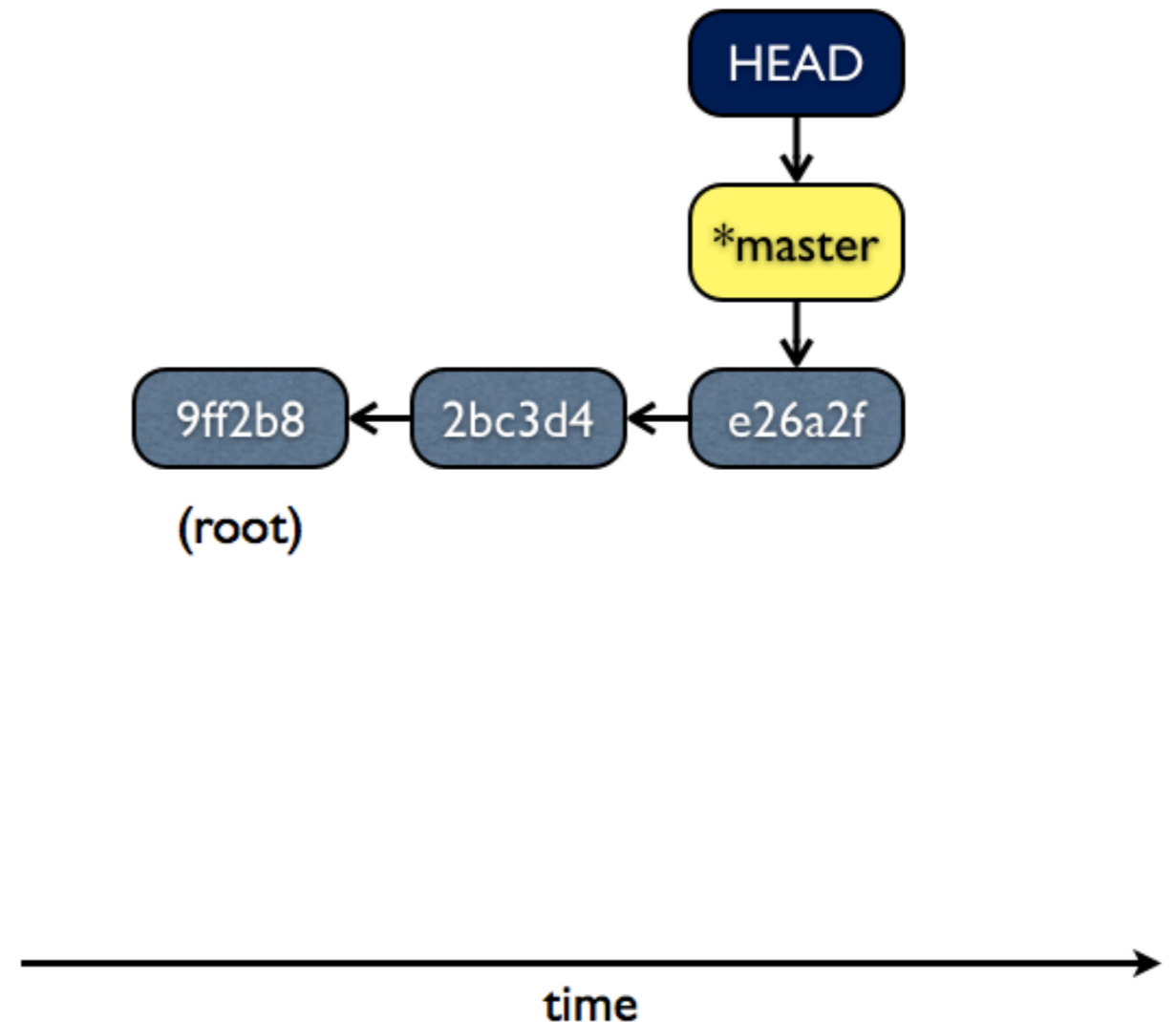
```
commit aae44037761d1d3a7871c1bf20524fd5ca930fb0  
Author: Shantanu Pavgi <pavgi@uab.edu>  
Date: Fri Sep 2 16:39:53 2011 -0500
```

```
Fixed [ticket:74]. Displays env command output in history/web-browser.
```

- Snapshot
- Committer's name and email address
- Parent commits
- Commit message

Branching

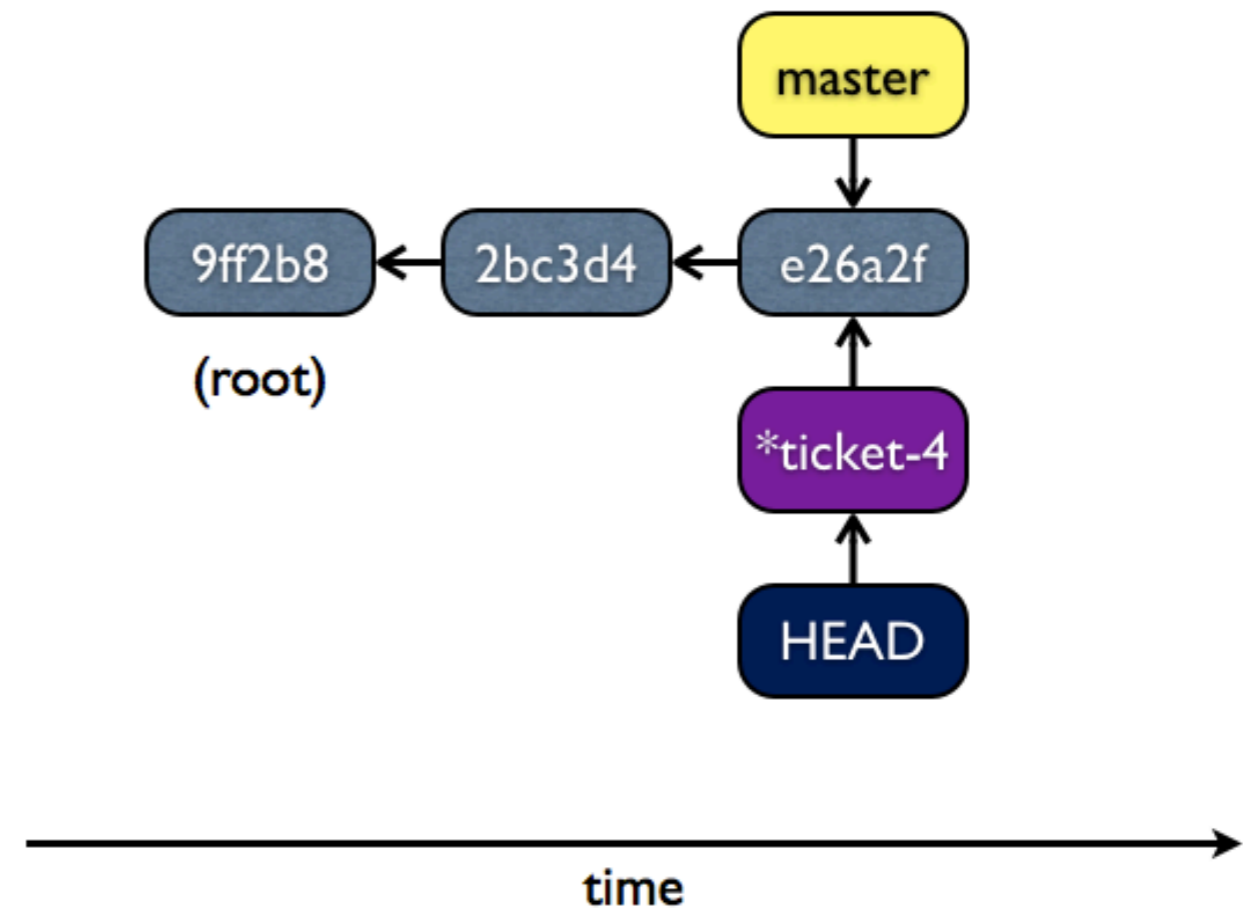
- Branch: Line of development
- 'master' branch is the default
- Create branches to have isolated lines of development



A 'head' is a named reference to a commit object – e.g. branches and tags. A repository can contain any number of heads, however it points to only one 'head' at a time. This 'current head' is called HEAD.

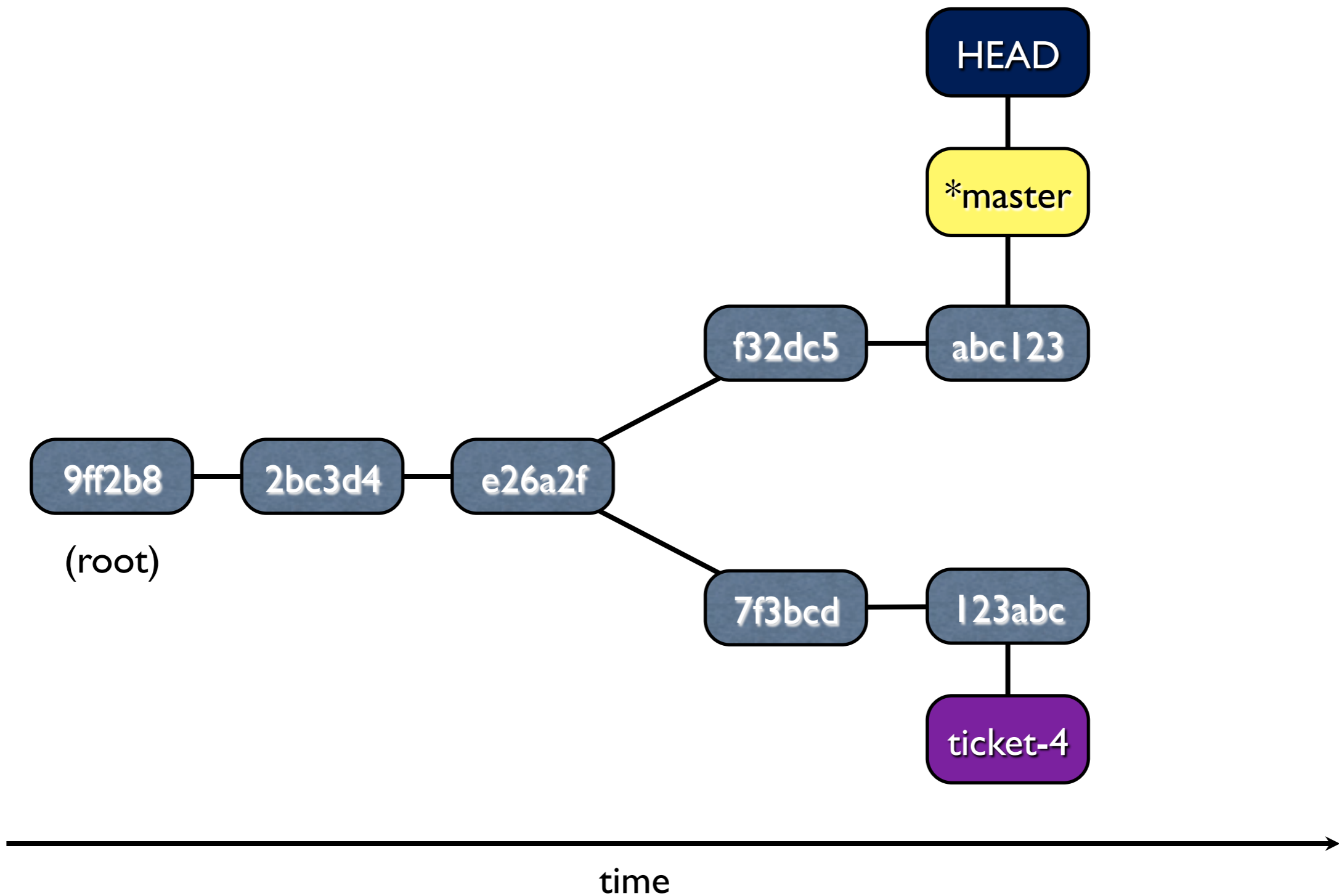
Branching

- Branch: Line of development
- 'master' branch is the default
- Create branches to have isolated lines of development

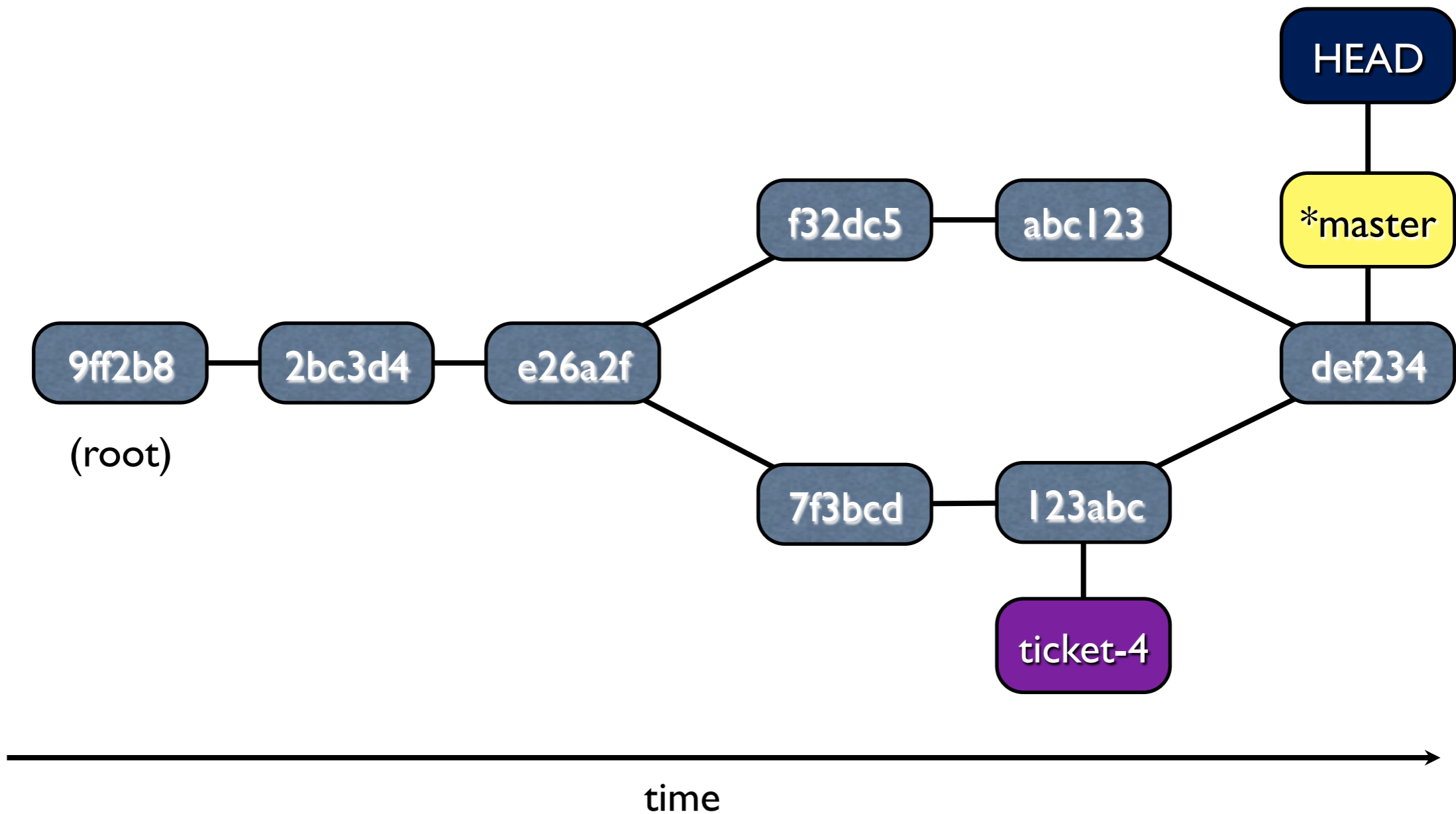


A 'head' is a named reference to a commit object – e.g. branches and tags. A repository can contain any number of heads, however it points to only one 'head' at a time. This 'current head' is called HEAD.

Branching



Merging



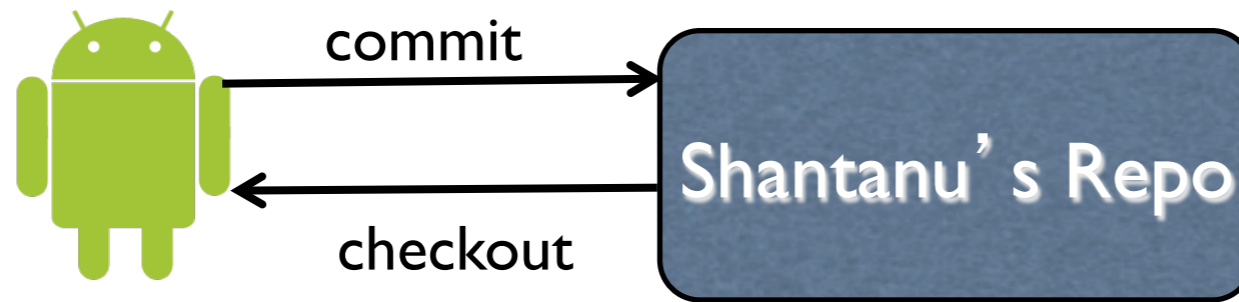
Branching and Merging

- Allows switching contexts
- Diverge from main line of development or stable code in a clean manner
- Separate space for experimentation and bug fixes
- “Local” - private commit history
- Rewrite or curate history before publishing it or merging into other branches

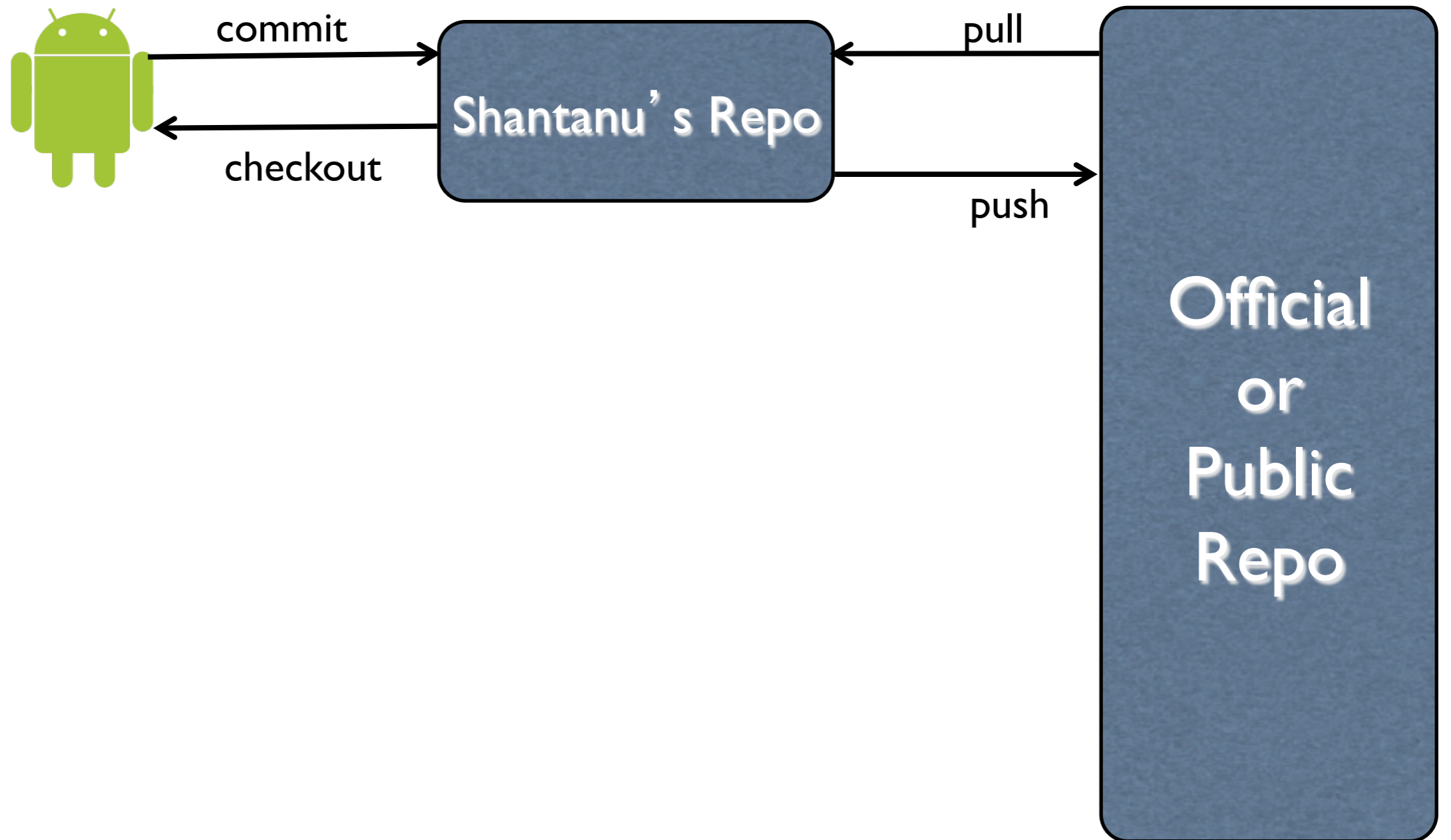
Workflows

- Distributed model and cheap branching allows numerous types of workflows
- Push model: Centralized repository type
- Pull model: Integration manager type
- ‘git workflows --help’
- Read more in ProGit book

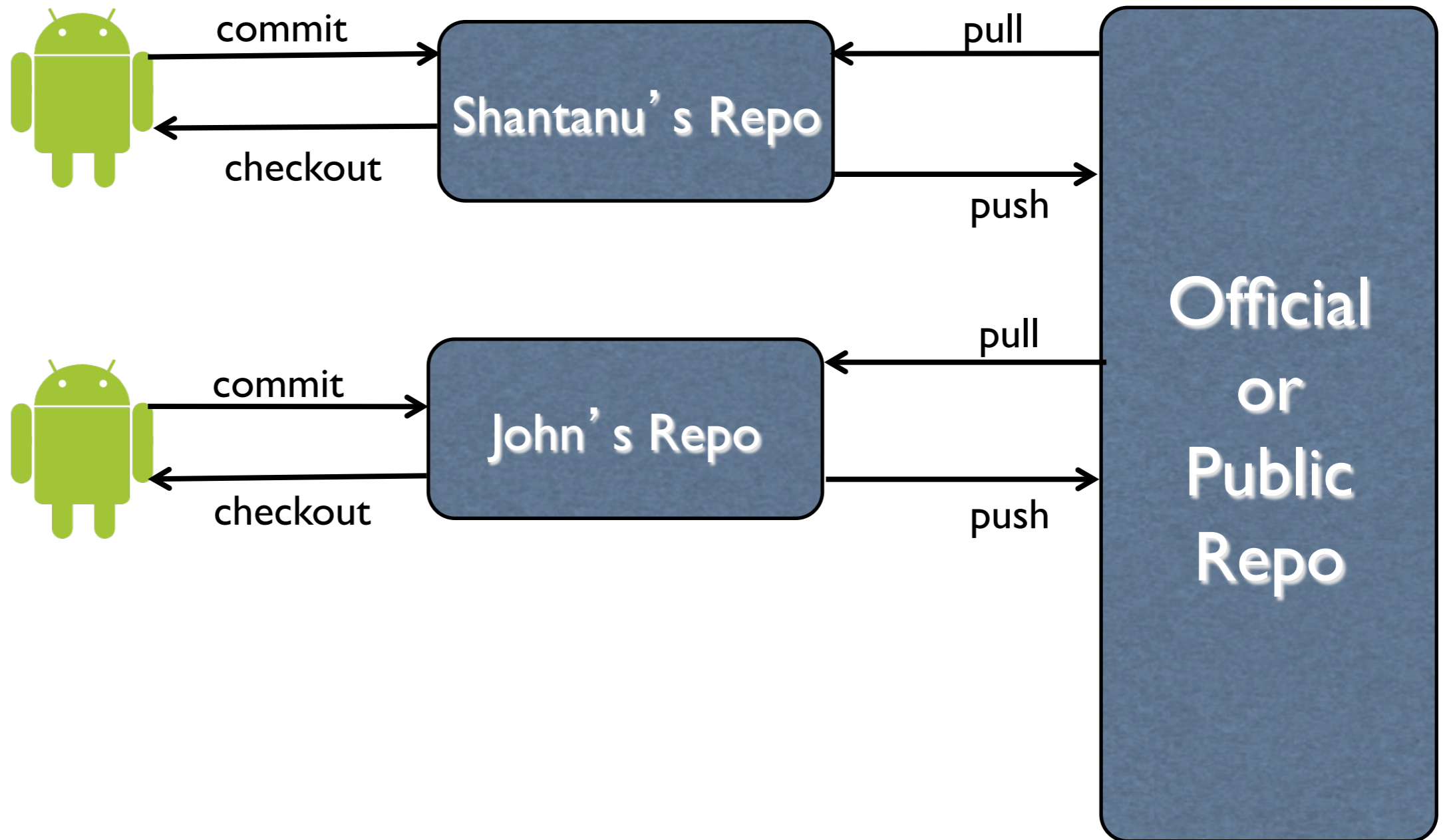
Workflow: Solo



Workflow: Solo



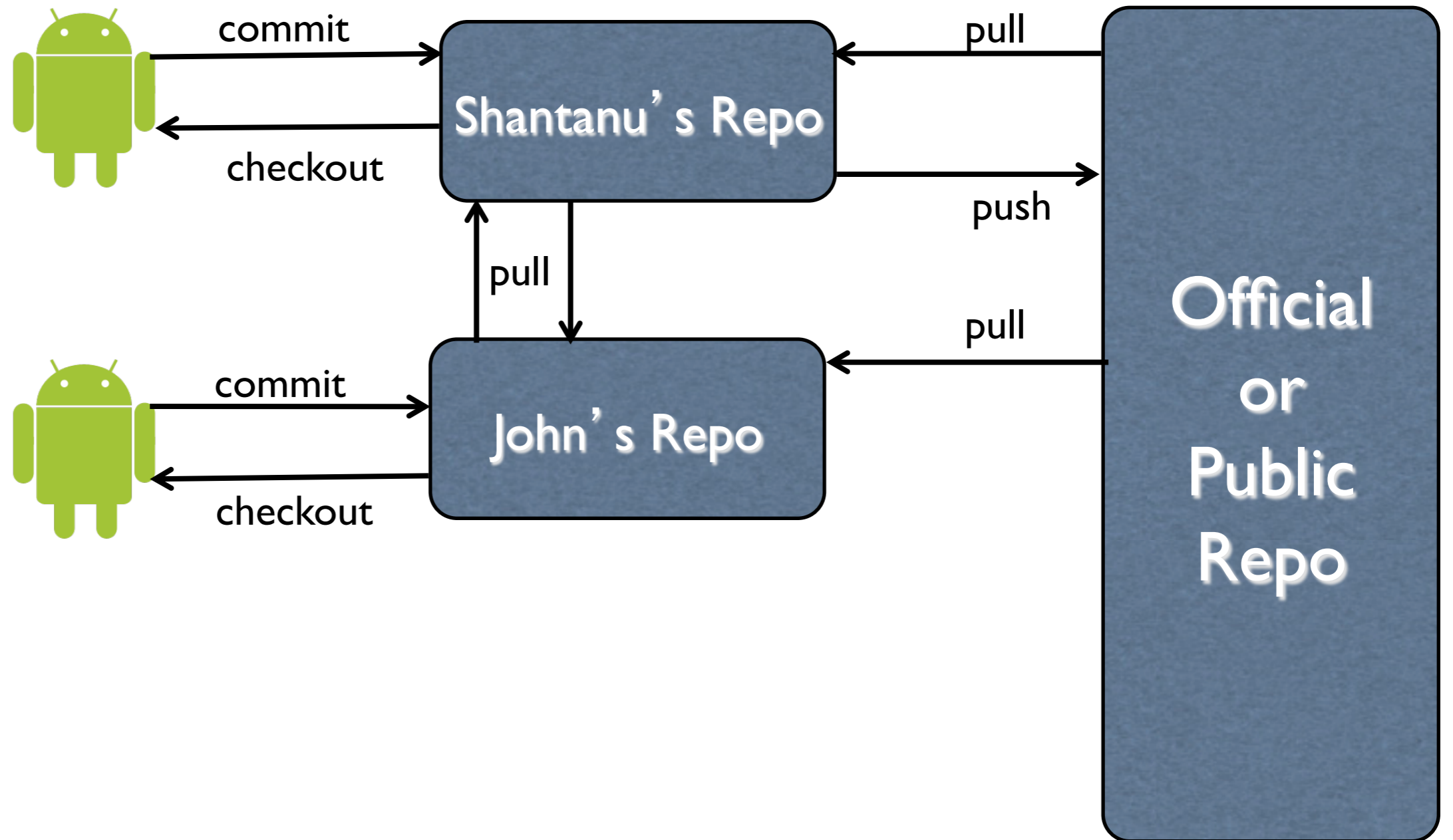
Workflow: Centralized



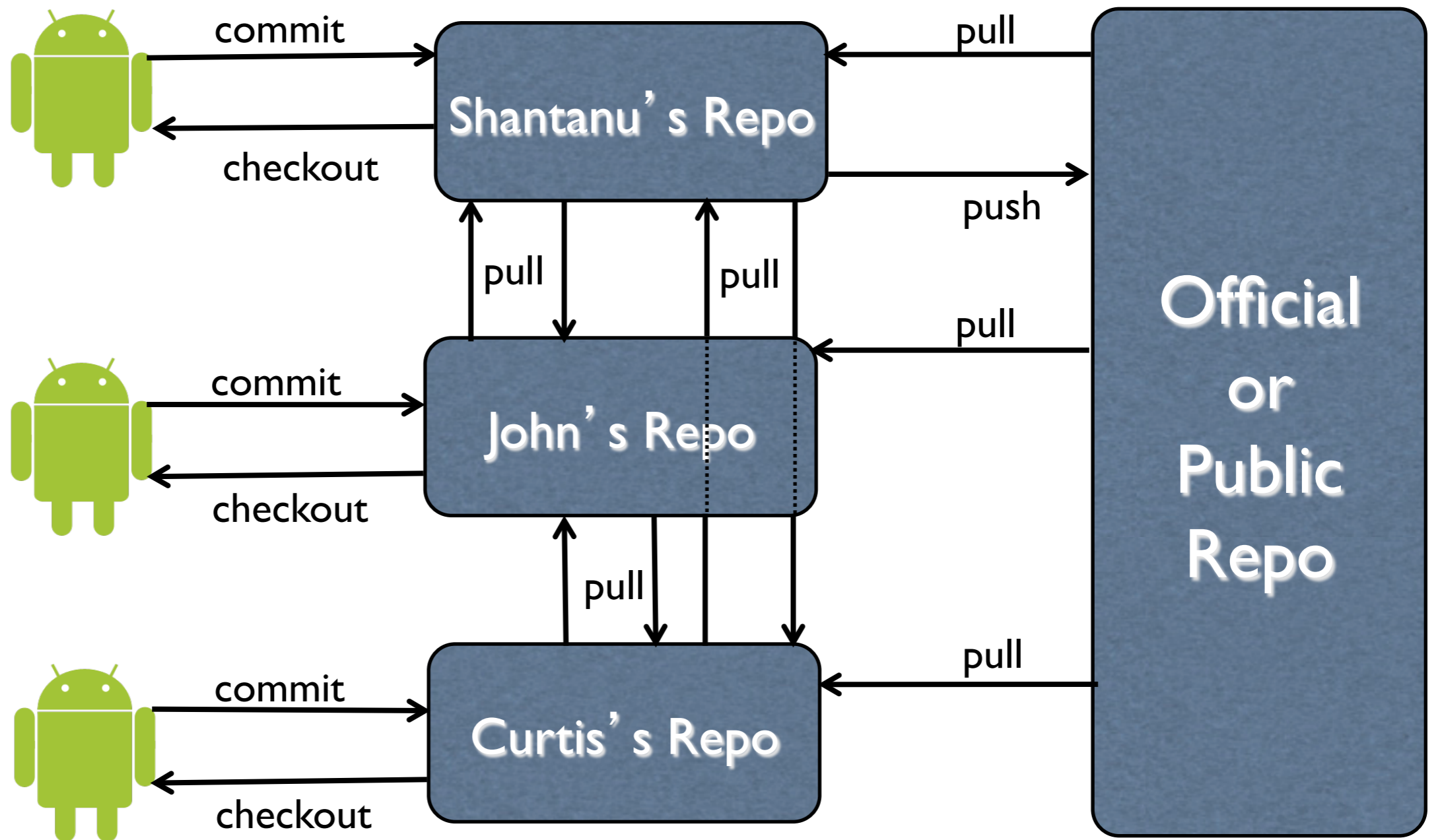
Workflow: Integration Manager



Workflow: Integration Manager



Workflow: Integration Manager



Learn More..

- <http://progit.org>
- <http://help.github.com>
- <http://eagain.net/articles/git-for-computer-scientists/>
- Subversion folks: <http://git.or.cz/course/svn.html>
- git man pages

MOTD

- Use Git
 - Works off-line
 - Personal logbook
 - Curate and publish history
 - Flexibility in workflows

- UABgrid docs:

<http://docs.uabgrid.uab.edu/wiki/UsingGitForDevelopment>

Additional Information

Git Installation

- Linux
 - Package managers
 - Source
- Windows
 - <http://code.google.com/p/msysgit/>
- Mac:
 - Source
 - Homebrew: <http://mxcl.github.com/homebrew/>

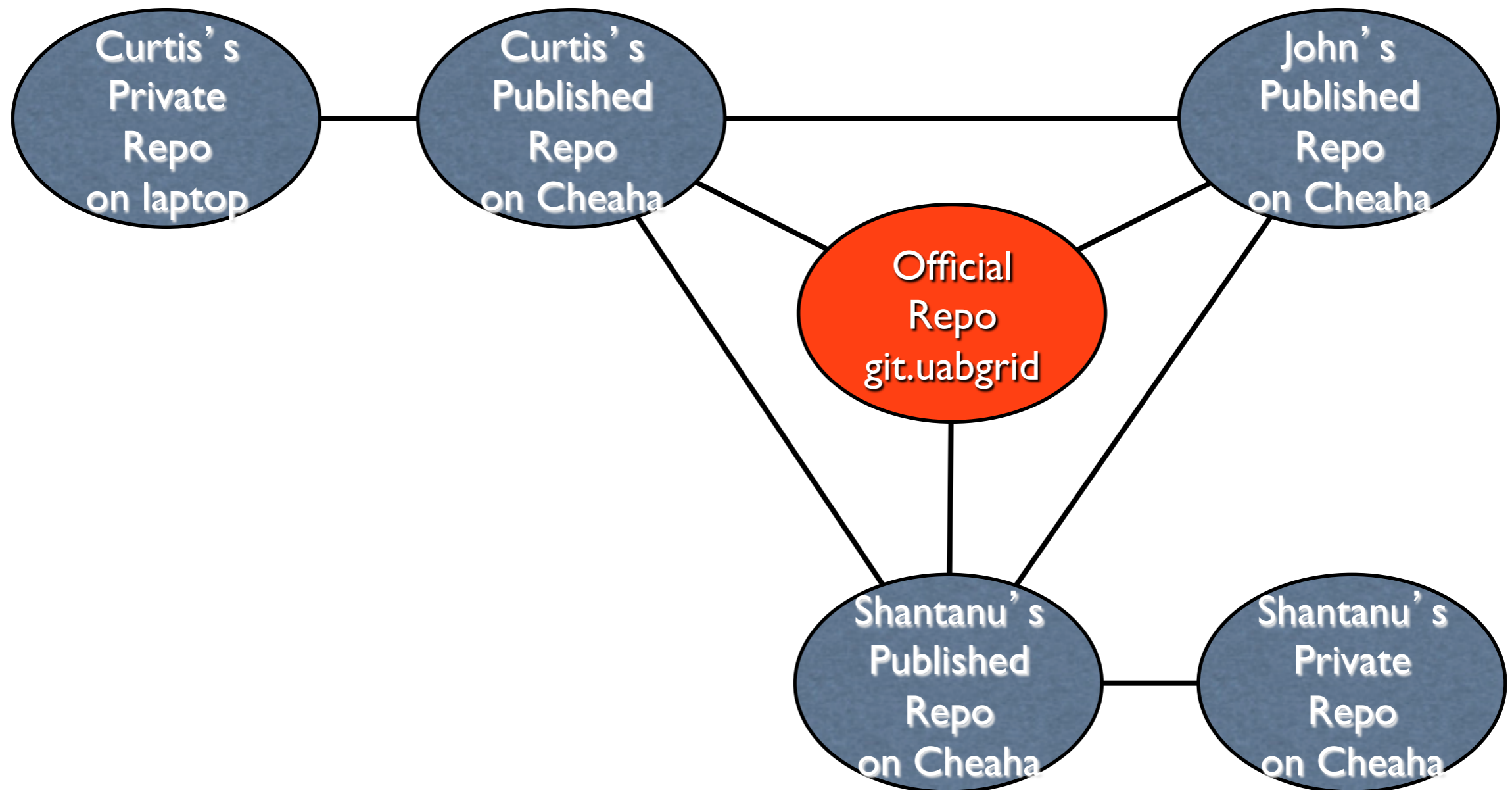
Add-on tools

- Access control: gitosis and gitolite
- GUI: gitk
- Display branch name in shell: <http://www.jonmaddox.com/2008/03/13/show-your-git-branch-name-in-your-prompt/>

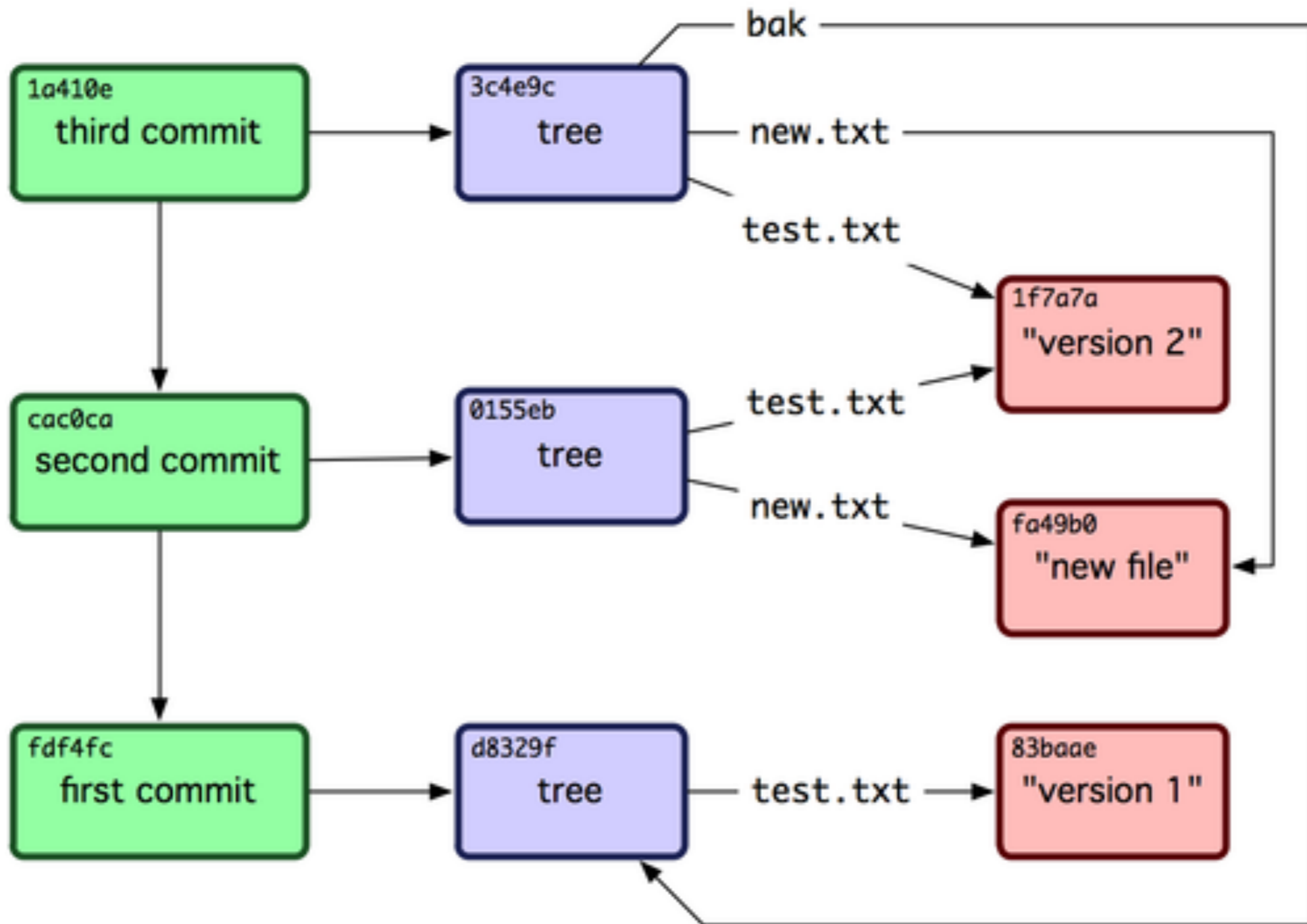
Rewriting History

- Remove sensitive data from repository
- Rewrite commits - commit messages, committer names, timestamps and committed files
- Rewriting changes commit hash
- DON'T rewrite commits of published history unless absolutely necessary
- Commands: rebase and filter-branch

Distributed repository model



Git Objects



Reference: <http://progit.org/book/ch9-2.html>

Repository Size and Packfiles

- Each commit is a complete snapshot
- Packs up several blob objects
 - Creates index file(s) and pack file(s)
 - Latest commit includes a complete copy
 - Older commits contain deltas
- `git gc` - cleanup and optimization
- References:
 - <http://progit.org/book/ch9-4.html>
 - http://book.git-scm.com/7_the_packfile.html

Git with large files

- Limitations:
 - Compression
 - Checksums
 - Deltas
 - Memory
- <http://caca.zoy.org/wiki/git-bigfiles> - few hundred MB
- <http://git-annex.branchable.com/> - tracks content/files without checking it in

Git with binary files

- Configurable with external diff tool
 - Create a new type according to file extension
 - Configure diff tool for the new type
 - Details: <http://progit.org/book/ch7-2.html>

Best Practices

- Work in a private and non-tracking local branch
- Granular commits - easier to squash commits together than split commits
- Fetch and merge
- Create ticket or issue specific branches

Git or Mercurial

- Google's analysis:
 - <http://code.google.com/p/support/wiki/DVCSAnalysis>
- Answer to Google's analysis:
 - <http://felipec.wordpress.com/2011/01/16/mercurial-vs-git-its-all-in-the-branches/>
- Speed and size:
 - <http://draketo.de/proj/hg-vs-git-server/test-results.html>

Git or Mercurial

- Language
 - Git: C and Perl
 - Mercurial: Python
- Syntax
- Branches
- Community

Transport protocols

- git
- SSH
- http(s)

Community Support

- <http://vger.kernel.org/vger-lists.html#git>
- <irc://irc.freenode.net/git>
- <http://letmegoogletthatforyou.com/>

Acknowledgments

- David Shealy
- John-Paul Robinson
- Bill Bradley
- Mike Hanby
- Poornima Pochana,
- Thomas Anthony

Local uncommitted changes, not checked in to index

api — develop — email-module — remotes/origin/develop Increased run-time limit for tophat and bowtie to 18 hours

Fixed [ticket:74]. Displays env command output in history/web-browser.

Fixed [ticket:72]. Galaxy sets metadata on every job output. This configuration change will run set metadata step

Merge branch 'upstream-tracker' into release2-720455407d1c-merge-with-upstream-949e4f5fa03a

remotes/origin/upstream-tracker Added .git directory to .hgignore file

commit after updating (hg pull -u) to upstream galaxy-dist revision 949e4f5fa03a

Merge branch 'develop' of /home/ozborn/projects/galaxy/galaxy into ozborn/develop

Added Sacromyces Cerevisiae genome (build 2) for bwa, bowtie and samtools.

Applied patch from upstream [<https://bitbucket.org/galaxy/galaxy-central/changeset/07de40a5a0b9>] galaxy-central

Changed personalize.sh script to use universe_wsgi.ini.sanitized file

Changed memory limits vf and h_vmem for bowtie, bowtie_color and yophat in universe_wsgi.ini.sanitized

Added paralel environment option to drmaa url for tophat, bowtie and bowtie_color

Untrack universe_wsgi.ini file using 'git rm --cached'

Modified personalize.sh script to use sanitized version of universe_wsgi.ini file

Added universe_wsgi.ini.sanitized file as a copy as deployed galaxy.uabgrid configuration. Any changes to co

Modified personalize.sh script with read command with prompt option than print and read commands. Also, added

Updated drmaa configuration for cuffcompare, cuffdiff, cufflinks, pileup_parser, snpEff, and tophat run-time a

Added TMP and TMPDIR variables in sge and drmaa job submission templates. We had tried adding this to environm

Ignore sqlite database universe.sqlite from git repository - added universe.sqlite to .gitignore

Added a script that generates sed-commands file which modifies universe_wsgi.ini file later.

Added personalization and universe_wsgi.ini.old files to git ignore list; personalization file contains sed co

Added Tree shrew build 62 to bowtie_indices.loc file

Following tool-data files were downloaded by galaxy after starting galaxy process for the first time after mer

remotes/origin/release1-50e249442c5a-merge-with-upstream-720455407d1c Merge branch 'upstream-tracker' into r

commit after updating (hg pull -u) to upstream galaxy-dist revision 720455407d1c

master — remotes/origin/master — remotes/origin/release1-50e249442c5a Modified .gitignore file to exclude t

Updating to run with the latest version of cufflink and tophat. It turns out that they are not compatibile wit

Updated bwa to print out some debug information to resolve the problem where the /tmp space on the compute nod

Updated tool-data file for additional builds

ticket:35 : changed SGE-drmaa job submission parameters

Questions???