

Source Control

- Duane Strong
- Strong Engineering LLC
- duanes@strongenging.com



What is it?

- Manages what can easily become total chaos.
- Part of a larger concern called Software Configuration Management

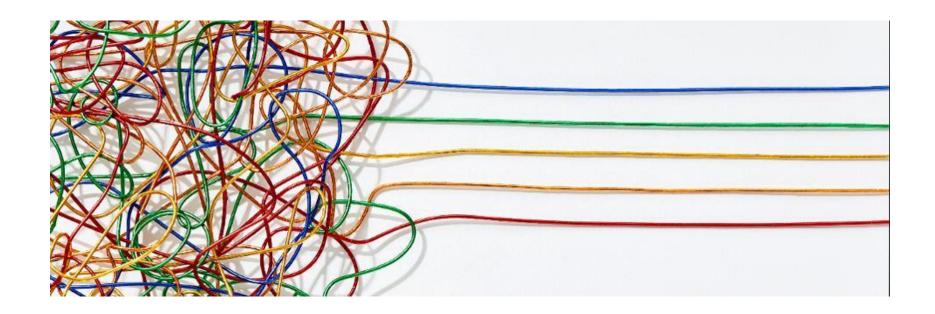
The task of tracking and controlling changes in the software. SCM practices include revision control and the establishment of baselines. If something goes wrong, SCM can determine what was changed and who changed it. If a configuration is working well, SCM can determine how to replicate it across many hosts.

- Provides traceability from feature or change request to implementation.
- Establishes tested baselines
- Establishes a set of requirements for change acceptance

Why do we need it?

Any sufficiently large project becomes (mostly?) an exercise in containing complexity.

The nature of software development is such that it constantly tries to become just more complicated than humans can understand. Entropy is constant.



Why do we need it?

- Team Co-ordination
- Provides an authoritative source
- Time Machine, can always get back to a prior version
- Back up
- Historian Why did we do that?
 - Commit logs
 - Issue reference
- Branches Multiple Development Timelines
 - Allows feature development and bug fixes without disturbing the main time line

What do you put in it?

Anything

- Not just software
- This presentation
- Maybe not huge things
- Once you get used to the safety net you want it

Everything

- Information radiator
- Never have to ask someone for a document
- Other tools might be better (wiki, ERP)
- Not generated items
- Not 3rd party SDKs, etc.

A Little History

SCCS (1972) RCS (1982)

Like a library book

Where the term 'checkout' comes from.

- Files locked by one person at a time
- Hard to collaborate, usually not working on the same part of a file
- Promotes cheating



A Little History

- CVS (1986) Multiple concurrent checkouts
 - No locking
 - Must 'update' prior to commit
 - Diffing, merging. Context aware diff
 - What is a conflict?
- SVN (2000) Improved CVS
 - Atomic commits
 - Change sets have versions not files
 - Database
 - Improved network transports

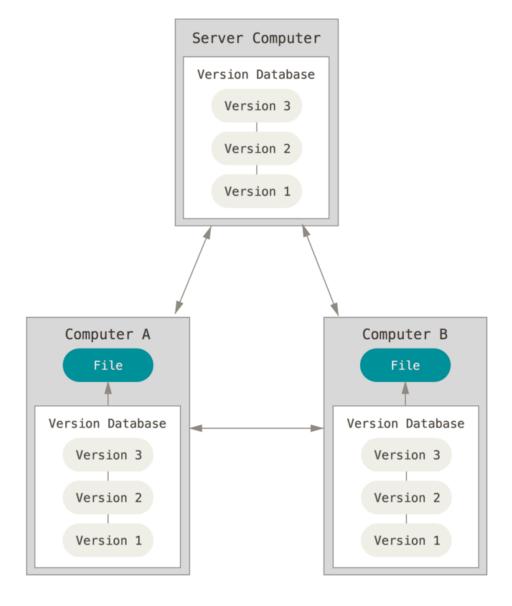
A Little History

- Git (2005) Linus Torvalds
 - Distributed everyone has an entire repository copy
 - No diffs, whole versions (storage is cheap)
 - Open source projects have special requirements
 - Distributed teams
 - Replicated repository, push/pull



Git – how does it work?

- Clone local copy of the repo (.git)
- Central server not required
- Not delta based
- Every change is an entire new copy

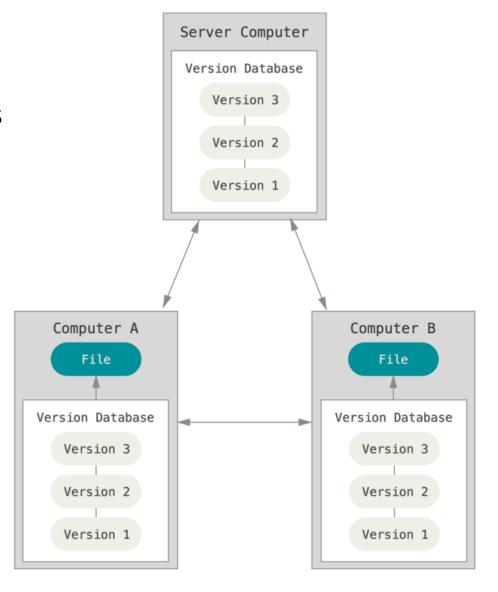


Git – how does it work?

- Blobs and the magic of SHA
 - 160 bits long about 10⁴⁸

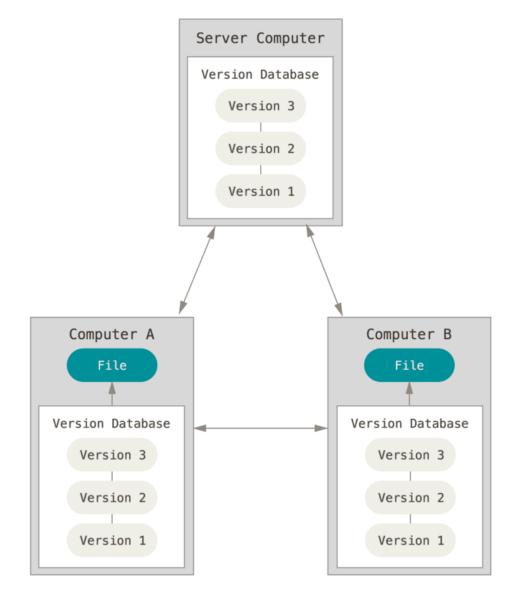
A trillion people generating a trillion new blobs per second for a trillion years is only 10⁴³

- Associative memory
- What is a blob?
 - Files
 - Trees
 - Commits
 - Tags
- Instantaneous branching



Git - terminology

- Cloning
- Origins
- Remote tracking branches
- Master branch
- Feature/Issue branch
- Working directory



Git Terminology

- Staging
- Commit
- Push
- Pull (Fetch + Update)
- Refs
 - HEAD
 - Tags
 - Remotes

Git

- Staging changes
 - git add (new and modified)
 - git rm
 - git mv (not really necessary)
- Commit
 - Atomic
 - git commit
 - Enter meaningful commit messages. Not "updated' or "fixed".
- Push
 - git push

Git

- git status
- .gitignore
 - Cascades
- Branching
 - git branch
 - git checkout

Checkout changes files in your working directory – tries to not let you mess up

Git

- Stashes
- Merging
 - git merge
- Conflicts

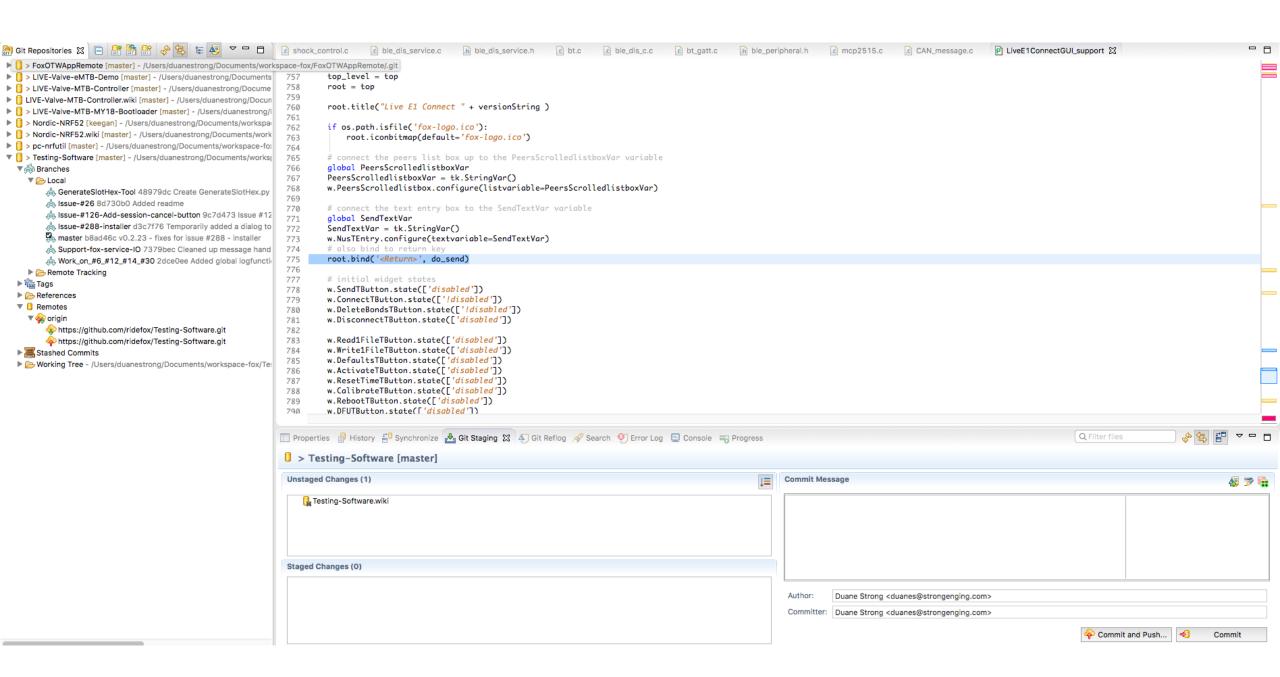
```
<<<<< HEAD:index.html

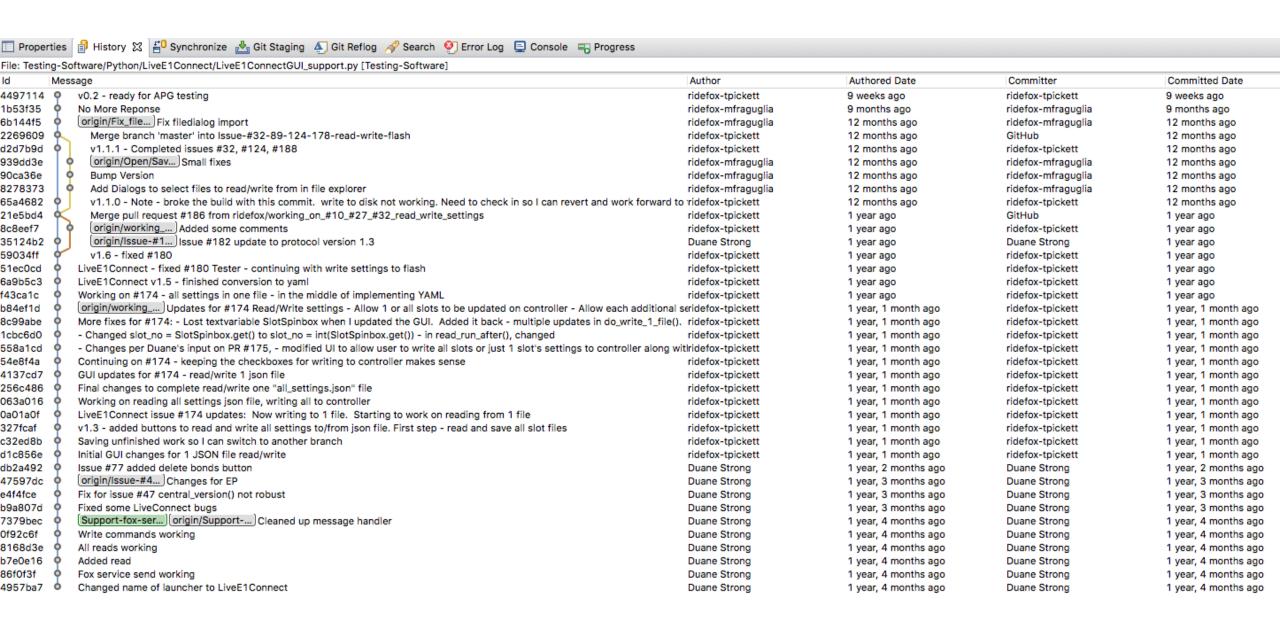
<div id = footer> contact : email.support@github.com</div>
======

<div id = footer>
  please contact us at email.support@github.com</div>
</div>
>>>>> mybranch:index.html
```

Git Frontends

- Command line 'porcelain'
 - I almost never use
- Visual Studio
- Eclipse eGit
- SourceTree
- Tortoise
- Github / Gitlab web interfaces





Workflows

- Never commit directly to master branch.
 - Create branch for an issue
- All work reviewed by at least one other pair of eyes
 - Bugs are exponentially more expensive to fix the farther down the process they live
 - Keeps at least one other team member in the loop
- All work should be traceable back to an issue.
- Merge to master branch
 - Use tools to enforce an acceptance criteria

Don't be one of those people

- Learn your tools well and know what they can do for you.
- Enter meaningful commit messages. Not "updated' or "fixed".
- Don't commit junk files (.o .bin)
 - Use .gitignores to filter out files/directories that should not be checked in
 - gitignore is applied cascading down the directory hierarchy. It is considered rude to filter out generic names at the root folder.
- Don't commit settings that are particular to your workstation (install paths etc.).

Don't be one of those people

- Have healthy respect (fear?) when pushing. You are affecting others lives.
- Don't just add all to git staging. Carefully pick the files to stage. Diff them to head to make sure you didn't leave something in you did not mean to.
- If you break the build you will be in the dog house. Most projects have multiple dependencies.

Mono repo vs. multi repos

- How much do I put into one repo?
- Cross project commits / code sharing / code access
- Atomic commits
- Submodules
 - Another repo as a subdir
 - Tracks at what commit that repo should be
 - Not really connected to each other
- Subtrees
 - Another repo as a subdir
 - Requires user to not mix commits
 - Requires multiple steps for merge
- Side by side repos not a repo

Q & A