Brief introduction to version control with Git

Dr. Maximilian Hindermann

February 17, 2023
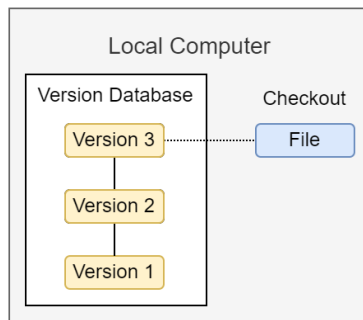
This section's goal

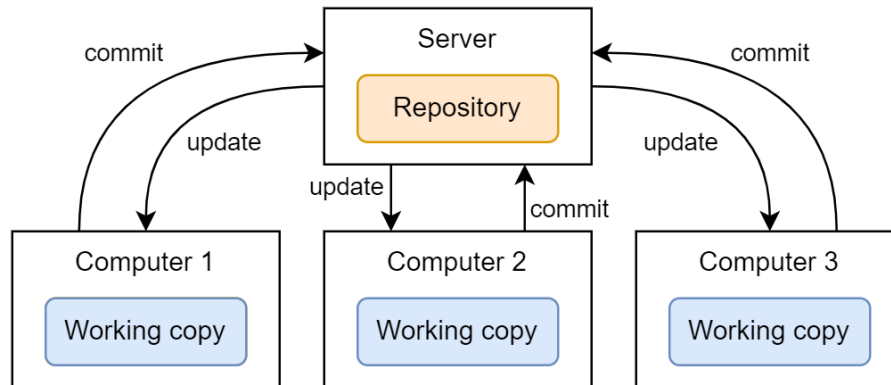Get you in a position to decide if Git might be useful for your work.

What is Git?

- Git is a distributed version control system (VCS).
- A VCS is "a system that records changes to a file or set of files over time so that specific versions can be recalled later" (Chacon & Straub 2014: 10).
- Git is very well suited for text-like data such as code, but of course also for text-like data such as:
    - an ALTO-XML file encoding the transcript of a medieval manuscript
    - a CSV file recording numerical weather data for Basel
    - a JSON-LD file containing a Tropy database
    - an SVG file of a research project's fancy new logo
    - a TXT file containing a novel from Project Gutenberg
- Git is less well suited for binary data such as audio data, images or video (changes in these files are opaque), especially if these files are large (>100MB); use Git Large File Storage extensionin such cases.
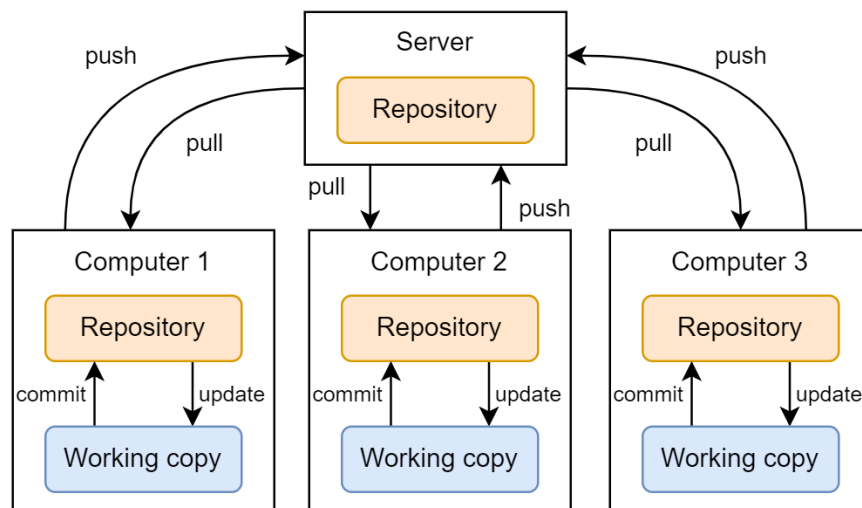
Local version control

- Solves the problem of error-prone manual versioning of files.

Centralized version control

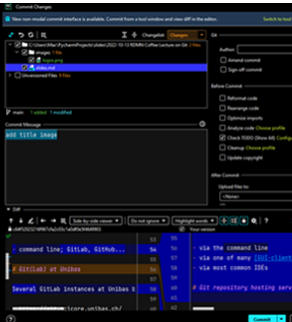- Solves the problem of how to work on the same files together on different computers.

Distributed version control

- Solves the problem of being dependent on central server for work and backup.

Using Git


```
$ git init
Initialized empty Git repository in /tmp/tmp.IMBYSY7R8Y/.git/
$ cat > README << 'EOF'
> Git is a distributed revision control system.
> EOF
$ git add README
$ git commit
[master (root-commit) e4dcc69] You can edit locally and push
to any remote.
 1 file changed, 1 insertion(+)
 crate mode 100644 README
$ git remote add origin git@github.com:cdown/thats.git
$ git push -u origin master
```

After installation on your computer, you can use Git:

- via the command line
- via one of many GUI-clients
- via most common IDEs

Git repository hosting services

But where do I host my Git repository? Do I have to configure a server myself?

- No. In practice, most users rely on a commercial Git repository hosting service such as GitHub, GitLab, Bitbucket, or others.

- Git repository hosting services do not replace but extend Git!

GitLab at Unibas

In contrast to GitHub, GitLab can be installed on your own infrastructure (e.g., required for sensitive data).

There are several GitLab instances at Unibas but none run as official university-wide service:

- https://git.scicore.unibas.ch/

- https://gitlab.psychologie.unibas.ch/
- https://gitlab.switch.ch/ub-unibas/
- https://gitlab.its.unibas.ch/

National infrastructure

Switch offers a GitLab instance, c4science supports Git:

- https://gitlab.switch.ch/
- https://c4science.ch/

Looking at a sample GitHub repository

GitHub repo for these slides: https://github.com/RISE-UNIBAS/clean-code/git

Image credit: Alex Eylar, "Inception", CC BY-NA-SA 2.0.

Further reading

Pro Git
By Scott Chacon and Ben Straub

Version Control with Git, 2nd ed.
By Jon Loeliger & Matthew McCullough

Git in Practice
By Mike McQuaid

Mastering Git
By Jakub Narębski

Pragmatic Guide to Git
By Travis Swicegood

Git in the Trenches
By Peter Savage

- Tons of material online and in print
- Scott Chacon and Ben Straub (2014): Pro Git, Apress
- Bernd Öggl and Michael Kofler (2022): Git. Projektverwaltung für Entwickler und DevOps-Teams, Rheinwerk Computing