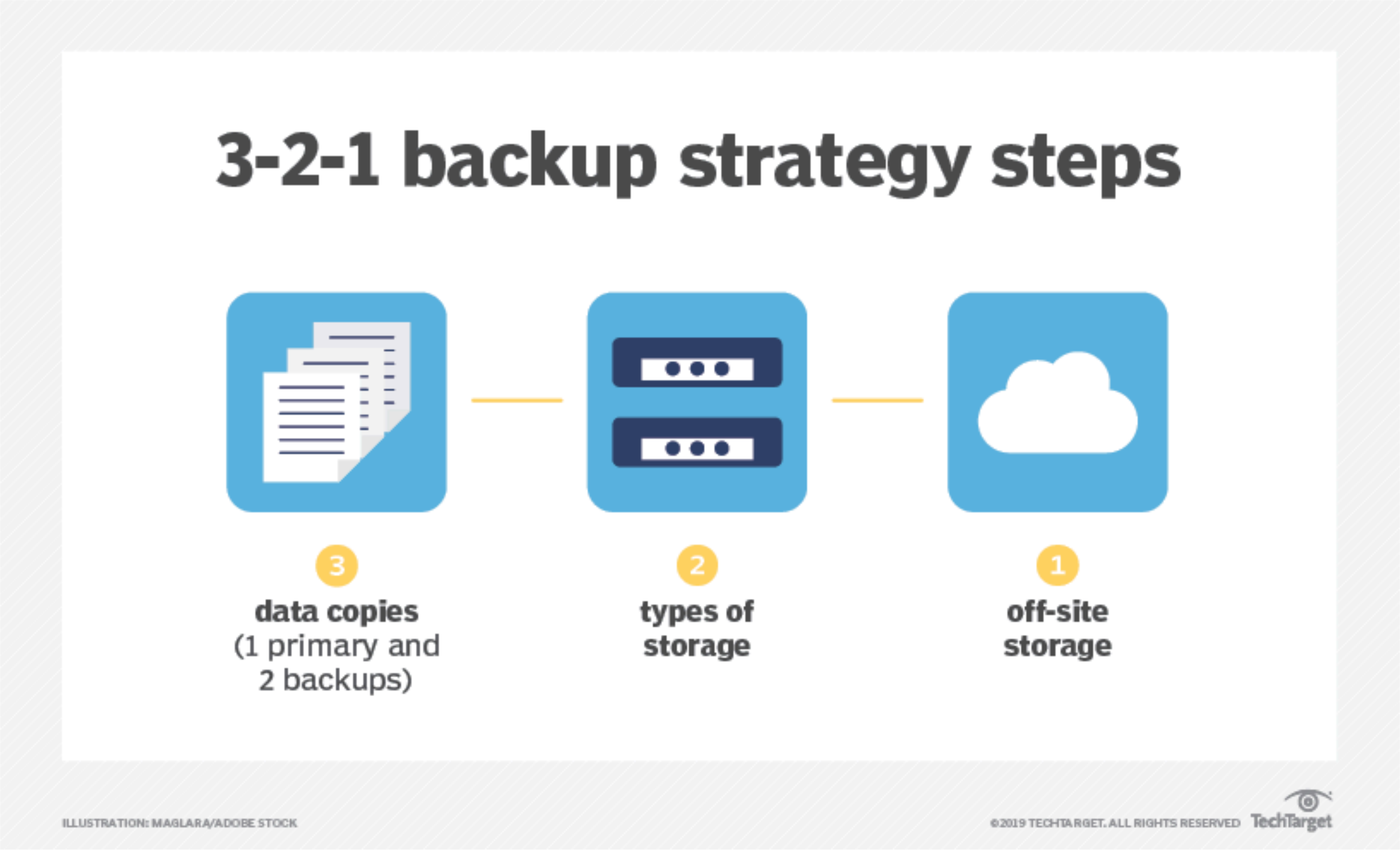
Data/code management.git

NCML lab meeting

2024-08-07 Seung-Goo KIM

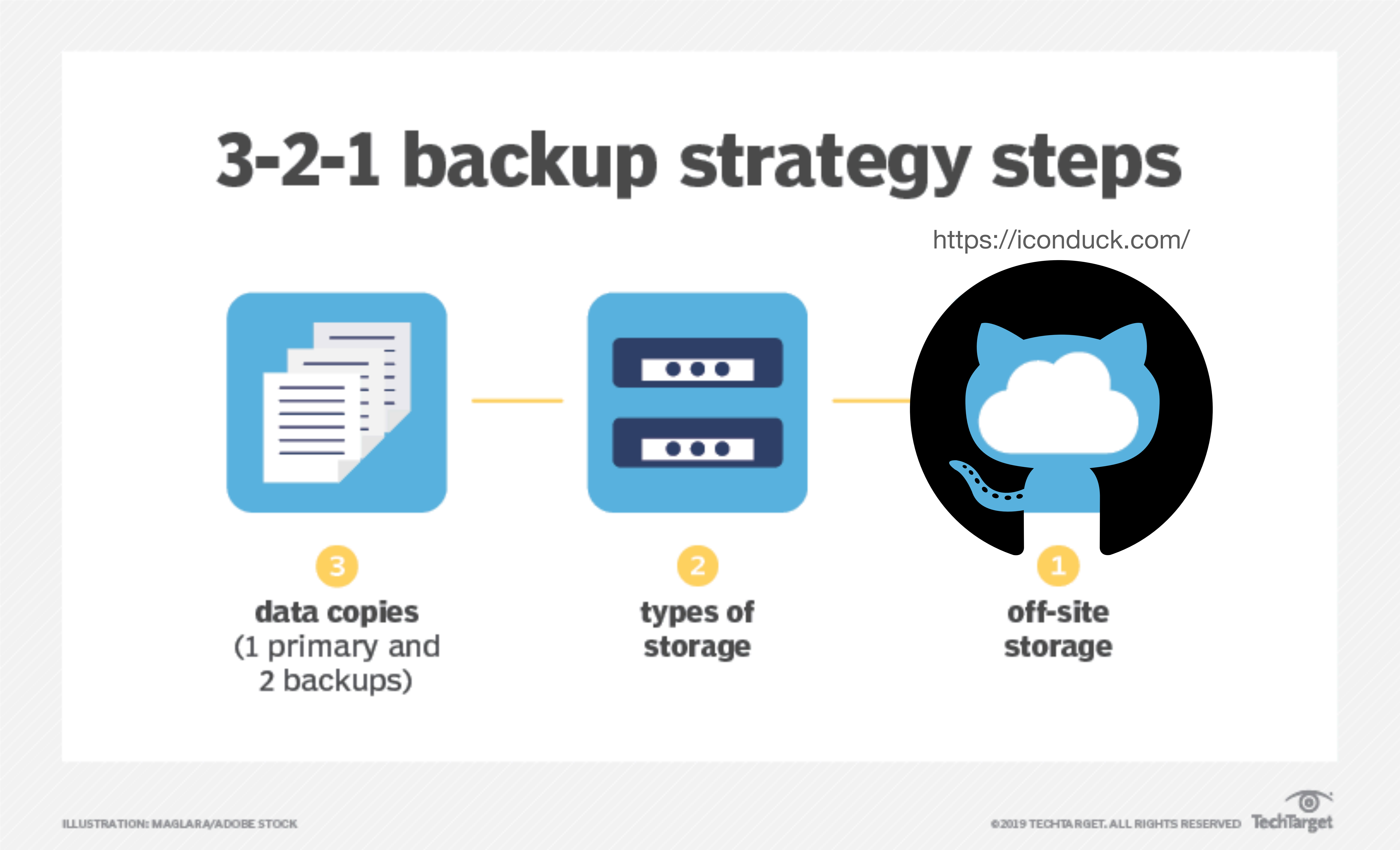
Code backup


3-2-1 backup strategy steps
3
data copies
(1 primary and
2 backups)
2
types of
storage
1
off-site
storage
ILLUSTRATION: MAGLARA/ADOBE STOCK
©2019 TECHTARGET. ALL RIGHTS RESERVED
TechTarget

- 1. Original: laptop HDD
- 2. Copy in the same HDD
- 3. Copy in the "cloud"

https://www.computerweekly.com/de/definition/3-2-1-Backup-Strategie

Code backup



https://iconduck.com/

ILLUSTRATION: MAGLARA/ADOBE STOCK

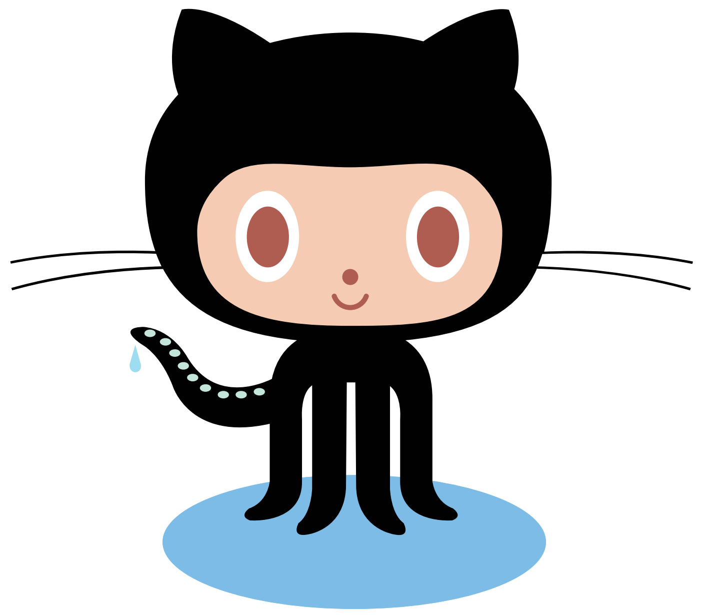
©2019 TECHTARGET. ALL RIGHTS RESERVED TechTarget

- 1. Original: laptop HDD
- 2. Copy in the same HDD
- 3. Copy in the Github or Gitlab server

Code backup & version control

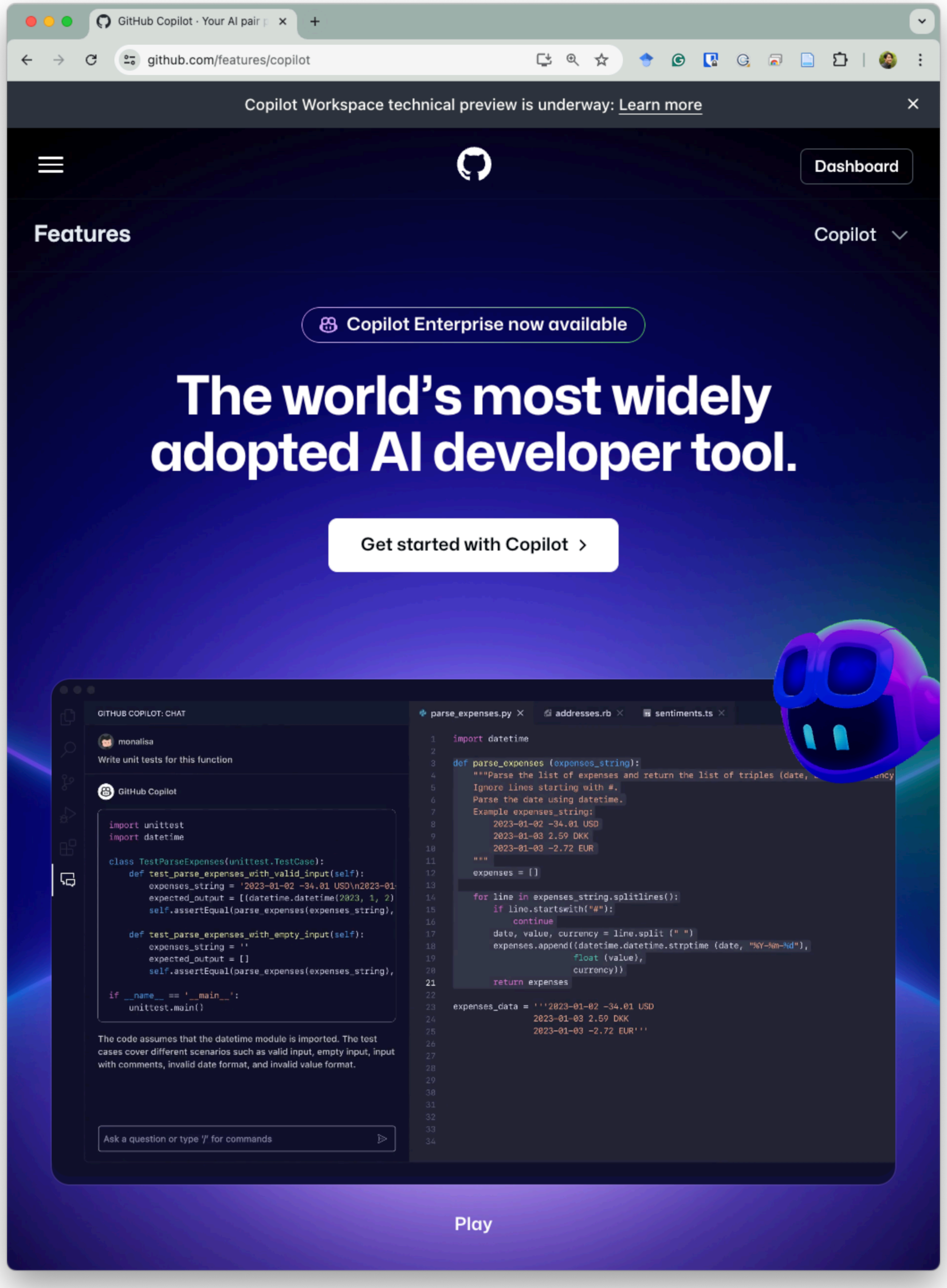
It's a good idea to always upload it on cloud

- **Manual**: `git` using github.com or gitlab.gwdg.de
  - The server may not be secured (officially, GitHub don't use PRIVATE repo for AI training).
  - Still, many people use (and are familiar with) GitHub...



(But can we really trust this "Octocat" guy? 🧐)

(or this cyber Origami fox?)



Copilot Workspace technical preview is underway: Learn more

Copilot Enterprise now available

The world's most widely adopted AI developer tool.

Get started with Copilot >

GITHUB COPILOT: CHAT

monalisa
Write unit tests for this function

GitHub Copilot

```
import unittest
import datetime

class TestParseExpenses(unittest.TestCase):
    def test_parse_expenses_with_valid_input(self):
        expenses_string = '2023-01-02 -34.01 USD\n2023-01-
        expected_output = [(datetime.datetime(2023, 1, 2)
        self.assertEqual(parse_expenses(expenses_string),

    def test_parse_expenses_with_empty_input(self):
        expenses_string = ''
        expected_output = []
        self.assertEqual(parse_expenses(expenses_string),

if __name__ == '__main__':
    unittest.main()
```

The code assumes that the datetime module is imported. The test cases cover different scenarios such as valid input, empty input, input with comments, invalid date format, and invalid value format.

Ask a question or type '/' for commands

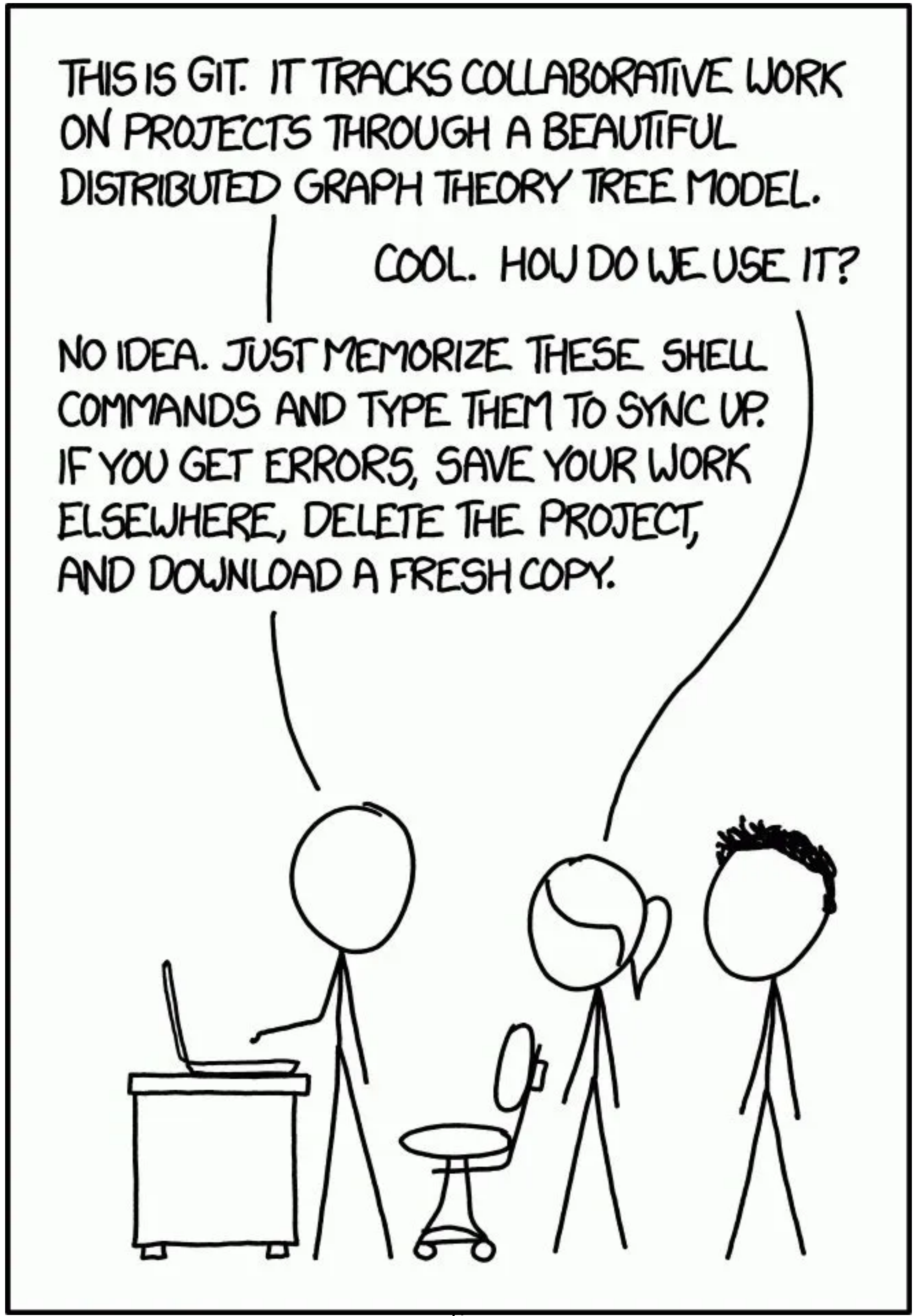
parse_expenses.py × addresses.rb × sentiments.ts ×

```
import datetime

def parse_expenses (expenses_string):
    """Parse the list of expenses and return the list of triples (date,
    Ignore lines starting with #.
    Parse the date using datetime.
    Example expenses_string:
        2023-01-02 -34.01 USD
        2023-01-03 2.59 DKK
        2023-01-03 -2.72 EUR
    """
    expenses = []

    for line in expenses_string.splitlines():
        if line.startswith("#"):
            continue
        date, value, currency = line.split (" ")
        expenses.append((datetime.datetime.strptime (date, "%Y-%m-%d"),
                         float (value),
                         currency))
        return expenses

expenses_data = '''2023-01-02 -34.01 USD
                   2023-01-03 2.59 DKK
                   2023-01-03 -2.72 EUR'''
```

Agenda

- What is Git? git
- What it GitHub? What is GitLab?
- How do we use Git?
    - Your laptop: GitHub Desktop (Graphic User Interface; GUI)
    - Servers: How do we use Terminal (Command Line Interface; CLI)?
    - How do we set up automatic authentication? (SSH key)

What is Git?


THIS IS GIT. IT TRACKS COLLABORATIVE WORK ON PROJECTS THROUGH A BEAUTIFUL DISTRIBUTED GRAPH THEORY TREE MODEL.
COOL. HOW DO WE USE IT?
NO IDEA. JUST MEMORIZE THESE SHELL COMMANDS AND TYPE THEM TO SYNC UP. IF YOU GET ERRORS, SAVE YOUR WORK ELSEWHERE, DELETE THE PROJECT, AND DOWNLOAD A FRESH COPY.

https://xkcd.com/1597/

Git Overview - Computerphile https://youtu.be/92sycL8ij-U?si=9J1Q9kQ_k4tPv45j

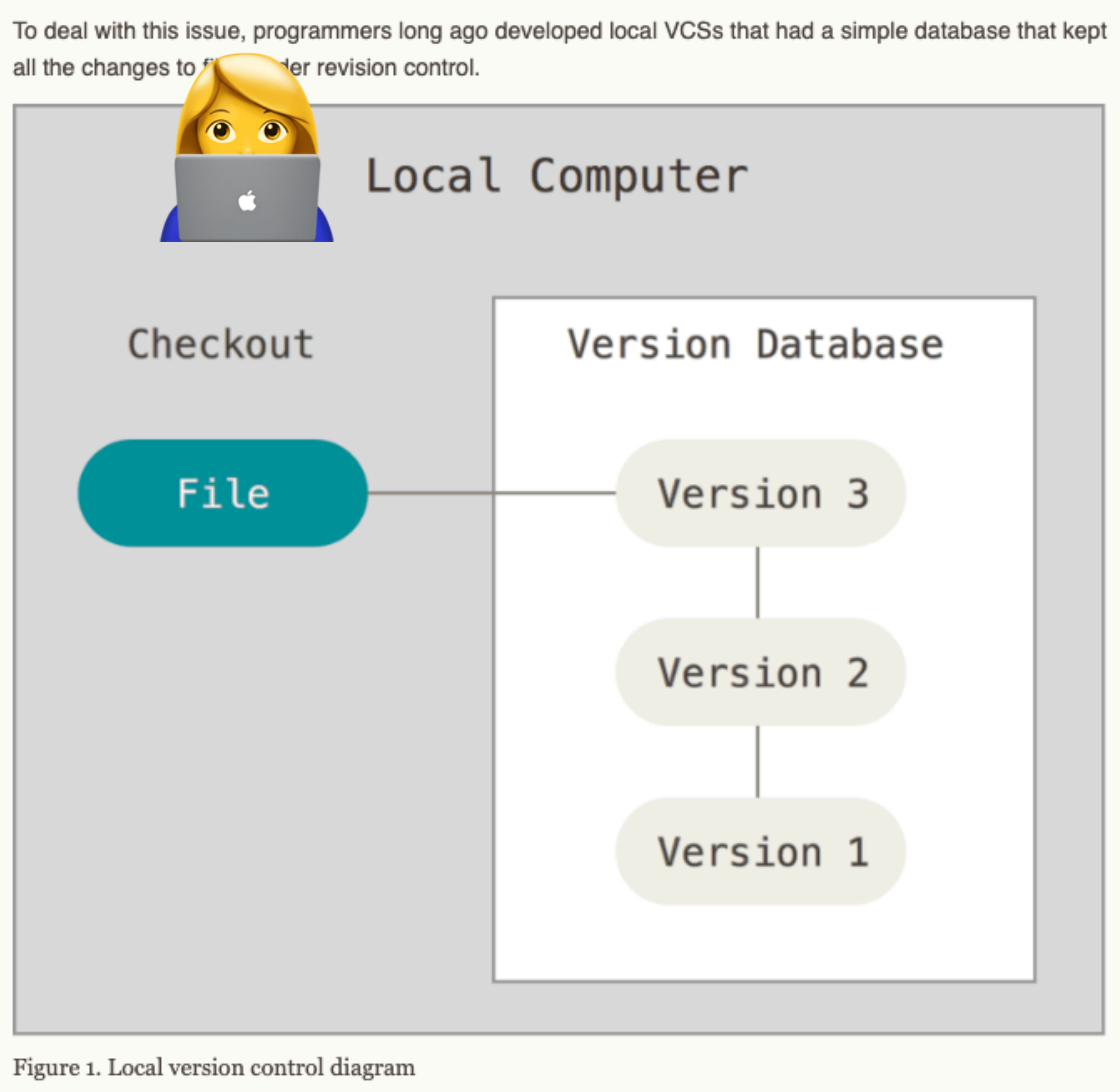
Speaker: Dr. Max Wilson (School of Computer Science, University of Nottingham)
Producer: Brady Haran (creator of “Numerphile”)

https://git-scm.com/book/en/v2

What is Git?

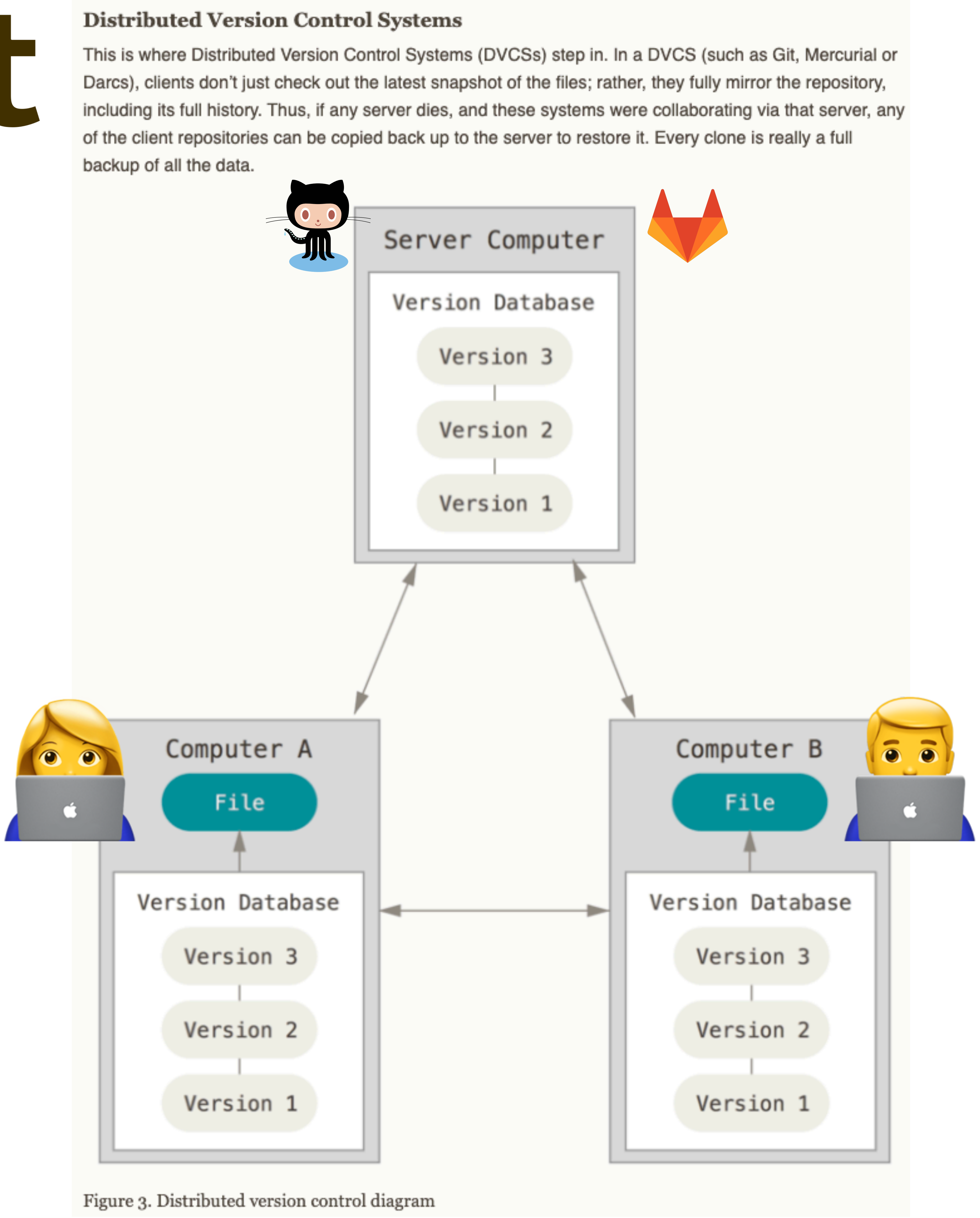
A version control software

Developing Linux kernel since 2005



To deal with this issue, programmers long ago developed local VCSs that had a simple database that kept all the changes to fi[illegible]der revision control.

Figure 1. Local version control diagram



Distributed Version Control Systems

This is where Distributed Version Control Systems (DVCSs) step in. In a DVCS (such as Git, Mercurial or Darcs), clients don't just check out the latest snapshot of the files; rather, they fully mirror the repository, including its full history. Thus, if any server dies, and these systems were collaborating via that server, any of the client repositories can be copied back up to the server to restore it. Every clone is really a full backup of all the data.

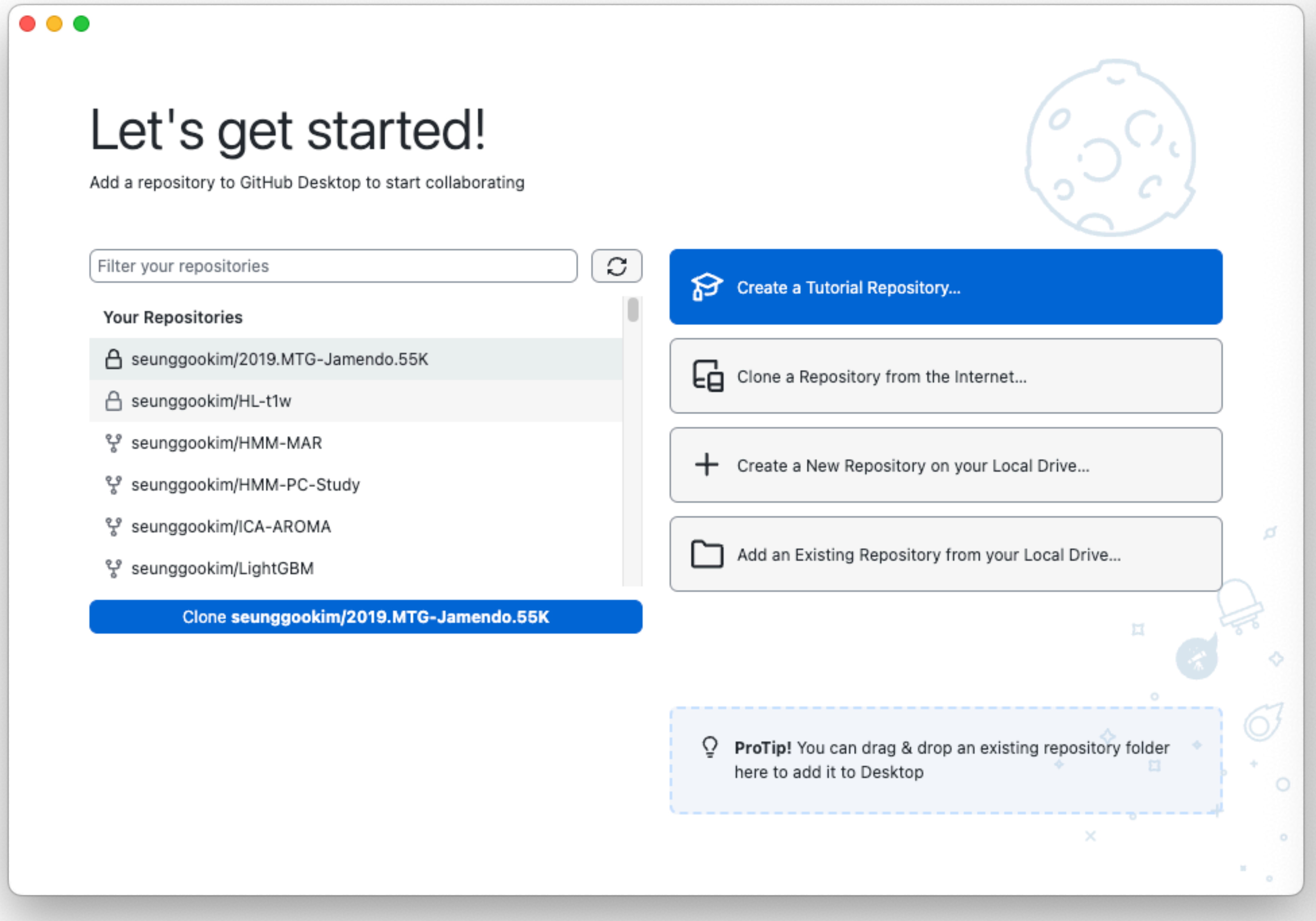
Figure 3. Distributed version control diagram

Teamwork & Git - Computerphile https://youtu.be/RzYJvSnzlMk?si=_NQQMkiYeztcZ4Kx

Speaker: Dr. Max Wilson (School of Computer Science, University of Nottingham)
Producer: Brady Haran (creator of “Numerphile”)

There is a GUI…

https://desktop.github.com/download/

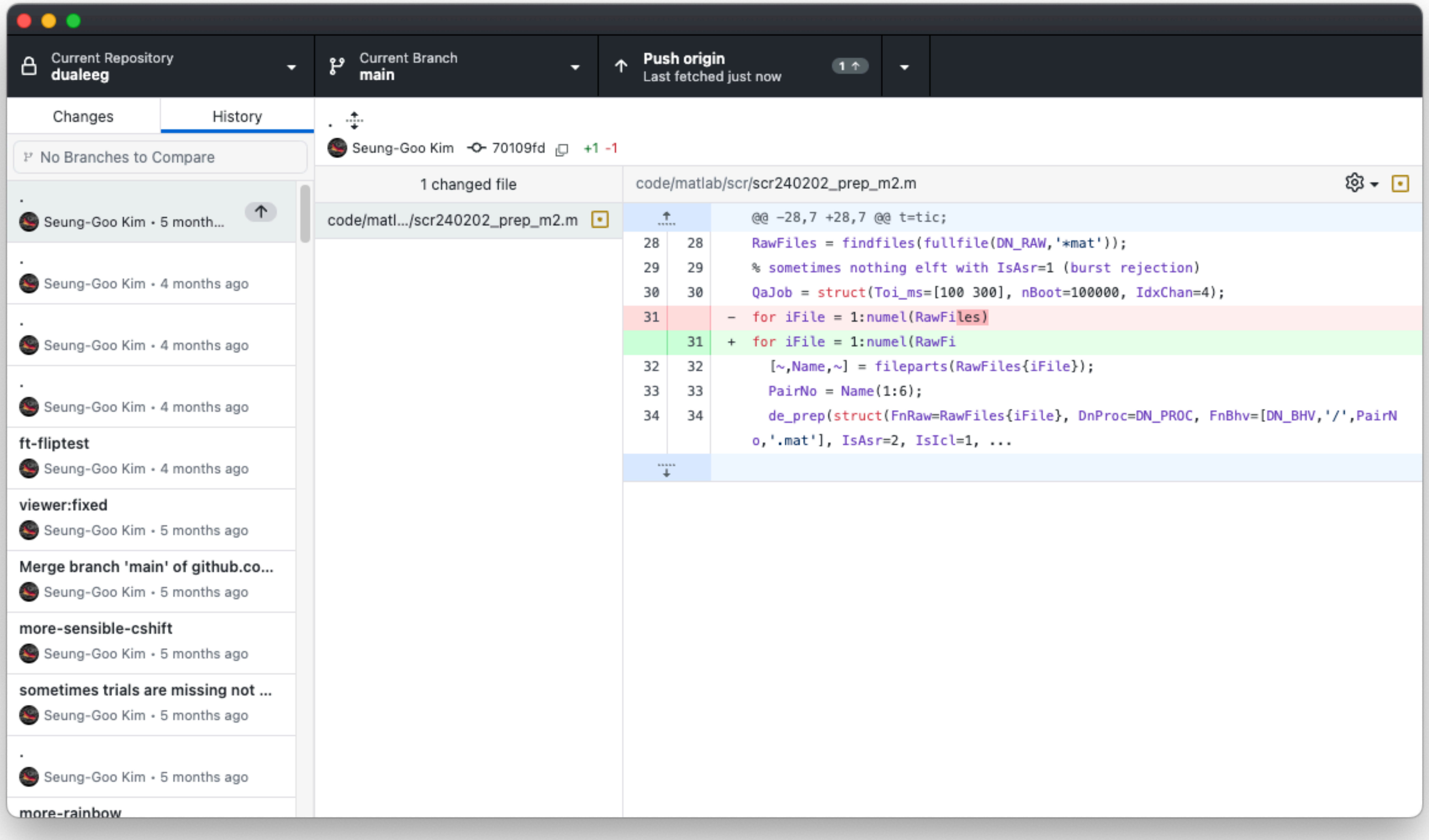


Let's get started!

Add a repository to GitHub Desktop to start collaborating

Filter your repositories
Your Repositories
seunggookim/2019.MTG-Jamendo.55K
seunggookim/HL-t1w
seunggookim/HMM-MAR
seunggookim/HMM-PC-Study
seunggookim/ICA-AROMA
seunggookim/LightGBM
Clone **seunggookim/2019.MTG-Jamendo.55K**
Create a Tutorial Repository...
Clone a Repository from the Internet...
Create a New Repository on your Local Drive...
Add an Existing Repository from your Local Drive...

ProTip! You can drag & drop an existing repository folder here to add it to Desktop



.
Seung-Goo Kim • 5 month...

.
Seung-Goo Kim • 4 months ago

.
Seung-Goo Kim • 4 months ago

.
Seung-Goo Kim • 4 months ago

ft-fliptest
Seung-Goo Kim • 4 months ago

viewer:fixed
Seung-Goo Kim • 5 months ago

Merge branch 'main' of github.co...
Seung-Goo Kim • 5 months ago

more-sensible-cshift
Seung-Goo Kim • 5 months ago

sometimes trials are missing not ...
Seung-Goo Kim • 5 months ago

.
Seung-Goo Kim • 5 months ago

more-rainbow

.
Seung-Goo Kim 70109fd +1 -1

1 changed file

code/matl.../scr240202_prep_m2.m

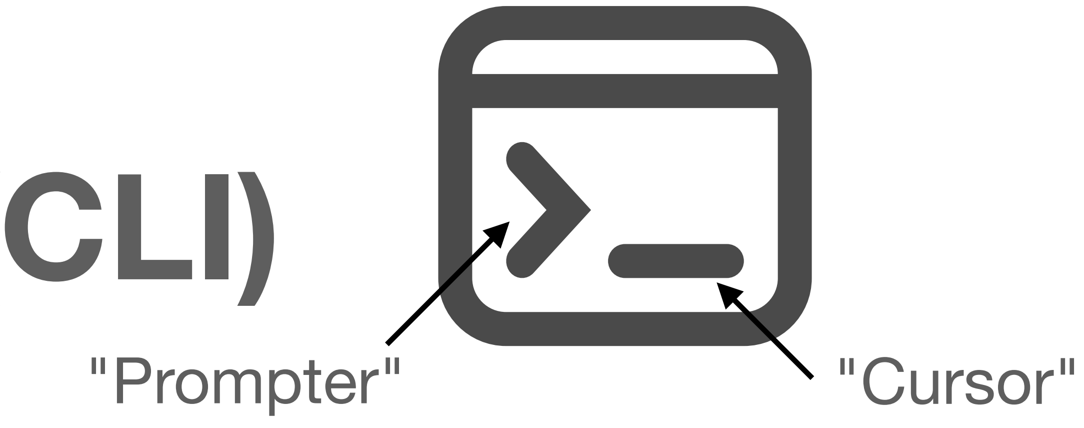
code/matlab/scr/scr240202_prep_m2.m

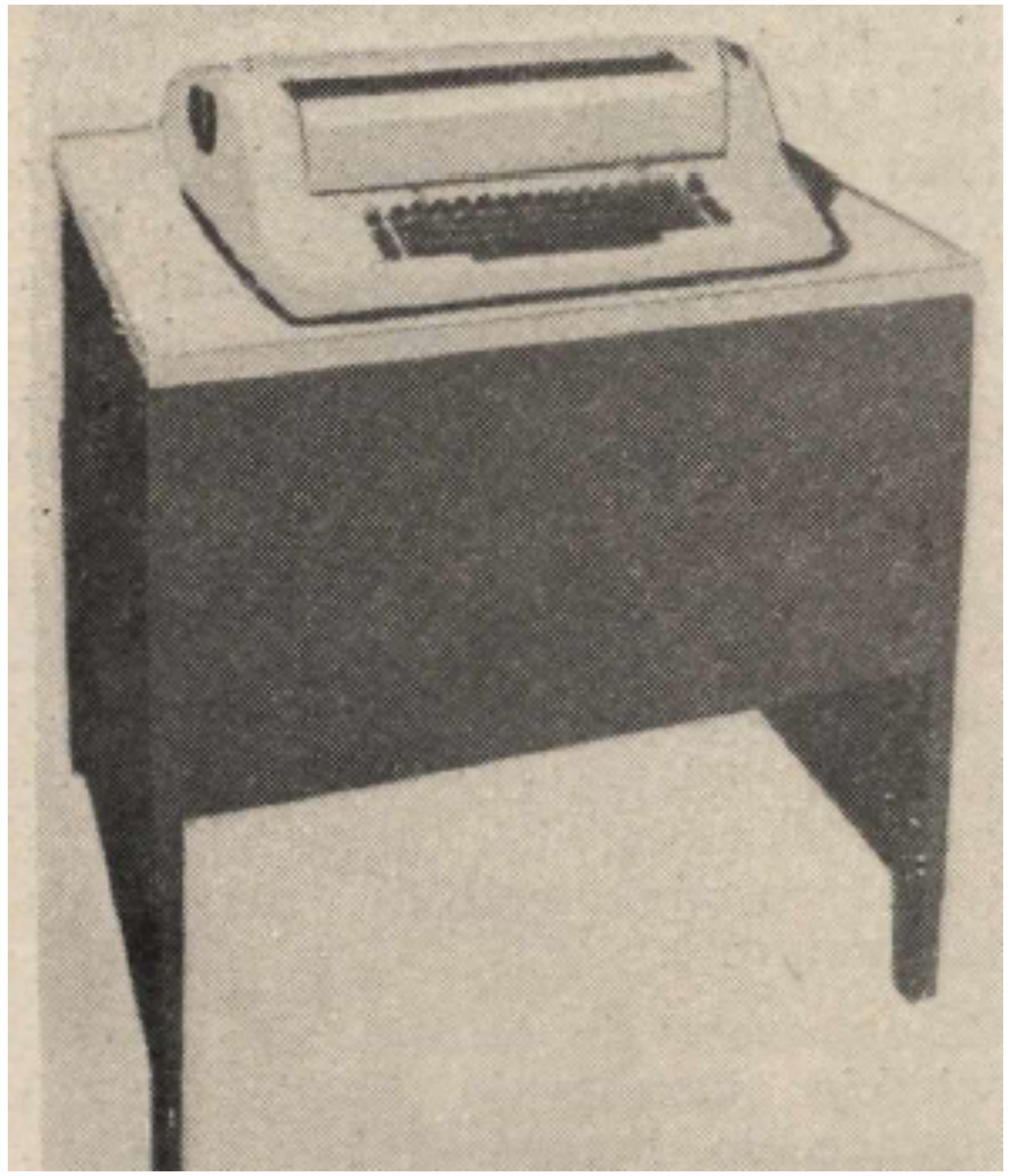
```
    @@ -28,7 +28,7 @@ t=tic;
    RawFiles = findfiles(fullfile(DN_RAW,'*mat'));
    % sometimes nothing elft with IsAsr=1 (burst rejection)
    QaJob = struct(Toi_ms=[100 300], nBoot=100000, IdxChan=4);
  - for iFile = 1:numel(RawFiles)
  + for iFile = 1:numel(RawFi
      [~,Name,~] = fileparts(RawFiles{iFile});
      PairNo = Name(1:6);
      de_prep(struct(FnRaw=RawFiles{iFile}, DnProc=DN_PROC, FnBhv=[DN_BHV,'/',PairN
    o,'.mat'], IsAsr=2, IsIcl=1, ...
```

https://en.wikipedia.org/wiki/Computer_terminal

But terminals…!

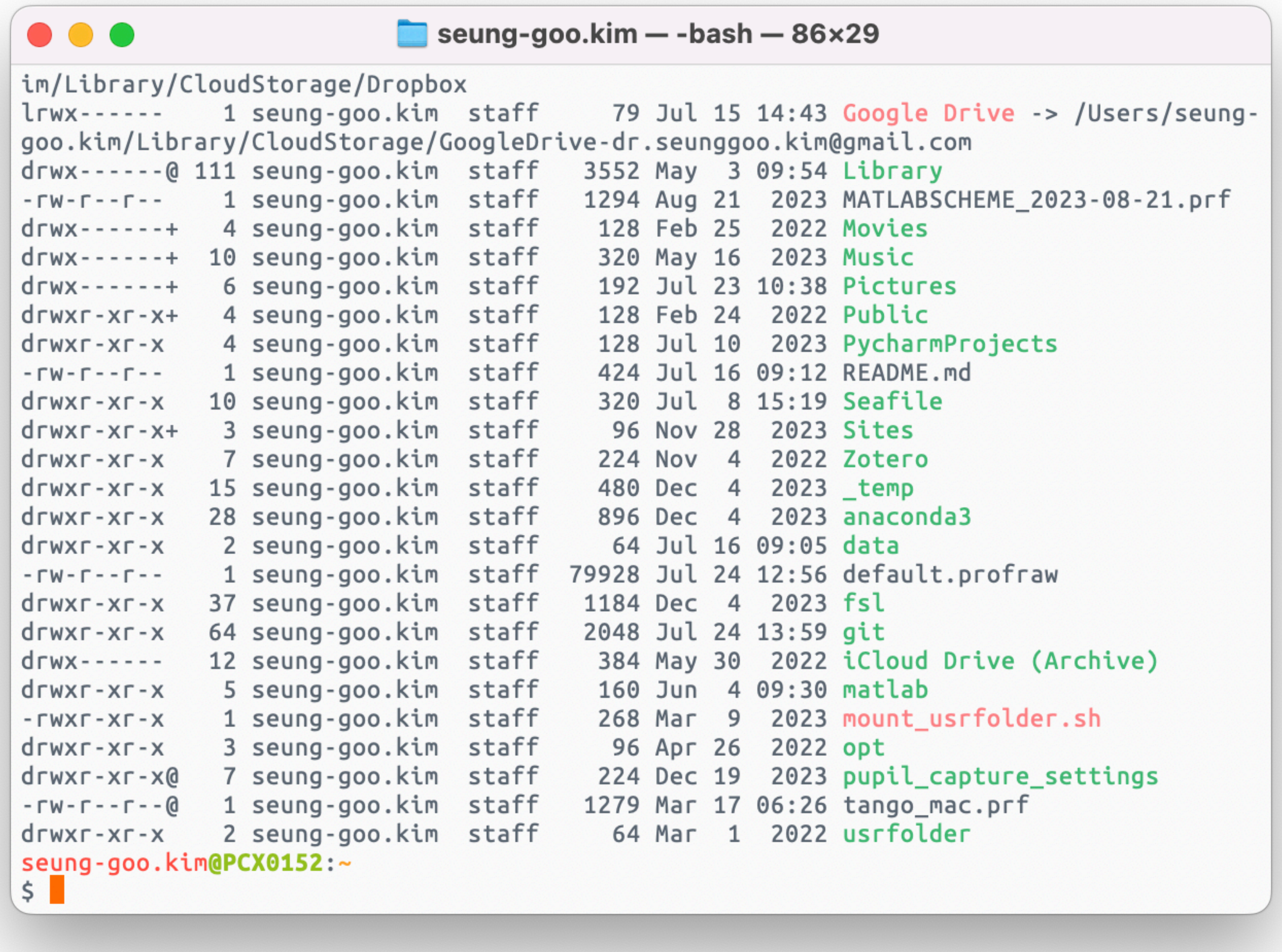
"Command line interface" (CLI)





IBM 2741 (1960s-1970s), "Teleprinters"

DEC VT100, 1978, the first to support cursor control



```
im/Library/CloudStorage/Dropbox
lrwx------    1 seung-goo.kim  staff     79 Jul 15 14:43 Google Drive -> /Users/seung-
goo.kim/Library/CloudStorage/GoogleDrive-dr.seunggoo.kim@gmail.com
drwx------@ 111 seung-goo.kim  staff   3552 May  3 09:54 Library
-rw-r--r--    1 seung-goo.kim  staff   1294 Aug 21  2023 MATLABSCHEME_2023-08-21.prf
drwx------+   4 seung-goo.kim  staff    128 Feb 25  2022 Movies
drwx------+  10 seung-goo.kim  staff    320 May 16  2023 Music
drwx------+   6 seung-goo.kim  staff    192 Jul 23 10:38 Pictures
drwxr-xr-x+   4 seung-goo.kim  staff    128 Feb 24  2022 Public
drwxr-xr-x    4 seung-goo.kim  staff    128 Jul 10  2023 PycharmProjects
-rw-r--r--    1 seung-goo.kim  staff    424 Jul 16 09:12 README.md
drwxr-xr-x   10 seung-goo.kim  staff    320 Jul  8 15:19 Seafile
drwxr-xr-x+   3 seung-goo.kim  staff     96 Nov 28  2023 Sites
drwxr-xr-x    7 seung-goo.kim  staff    224 Nov  4  2022 Zotero
drwxr-xr-x   15 seung-goo.kim  staff    480 Dec  4  2023 _temp
drwxr-xr-x   28 seung-goo.kim  staff    896 Dec  4  2023 anaconda3
drwxr-xr-x    2 seung-goo.kim  staff     64 Jul 16 09:05 data
-rw-r--r--    1 seung-goo.kim  staff  79928 Jul 24 12:56 default.profraw
drwxr-xr-x   37 seung-goo.kim  staff   1184 Dec  4  2023 fsl
drwxr-xr-x   64 seung-goo.kim  staff   2048 Jul 24 13:59 git
drwx------   12 seung-goo.kim  staff    384 May 30  2022 iCloud Drive (Archive)
drwxr-xr-x    5 seung-goo.kim  staff    160 Jun  4 09:30 matlab
-rwxr-xr-x    1 seung-goo.kim  staff    268 Mar  9  2023 mount_usrfolder.sh
drwxr-xr-x    3 seung-goo.kim  staff     96 Apr 26  2022 opt
drwxr-xr-x@   7 seung-goo.kim  staff    224 Dec 19  2023 pupil_capture_settings
-rw-r--r--@   1 seung-goo.kim  staff   1279 Mar 17 06:26 tango_mac.prf
drwxr-xr-x    2 seung-goo.kim  staff     64 Mar  1  2022 usrfolder
seung-goo.kim@PCX0152:~
$ 
```

macOS14, 2023, ayu-light themed

The first Unix in 1969, Bell Lab

The mighty PDP-7

Ken (seated) and Dennis (standing) at a PDP-11 in 1972.

Shell (Terminal) basics

bash or zsh on mac or linux

- Why? **Because everything is "scriptable"!**
- In fact, all shells are **Scripting Languages** like R or Python. You can define variables, call functions, use conditionals, ...
- Most popular shells:
  - "`sh`" (Bourne SHell), 1979 [default in Version 7 Unix]
  - "`bash`" (Bourne-Again SHell), 1989 [default in many linux]
  - "`zsh`" (Z shell), 1990 [default in macOS]

DEMO

- Create a REMOTE repo
- Clone it to local repos
- Create and resolve a conflict
- (Create and merge a branch)