

Git & GitHub Mastery for PhD Researchers

A Comprehensive 4-Part Lecture Series on Reproducible Science



LECTURE 01

The Foundations

Introduction to Version Control & Terminal Basics

The Researcher's Core Git Loop



git init

Initialize your local 'lab bench'. Git starts tracking every character change in this folder.



git add

Move changes to the 'Staging Area'. Like preparing a sample for a microscope snapshot.



git commit

Save a permanent snapshot of your logic with a detailed note on 'Why' it changed.



git push

Synchronize your local history to GitHub for backup and academic collaboration.

Mastering the CLI Environment

Terminal as the 'Backstage'

While GUIs exist, the Command Line Interface (CLI) is the gold standard for research automation.

- Speed: Perform complex file operations instantly.
- Precision: Explicit control over branching and merging.
- Scripting: Automate repetitive git workflows.

```
dministrator: Command Prompt

kdir git-test && pushd git-test && git init
ed empty Git repository in c:/work/git-test/.git/

it-test>git config --local user.name "Żołta Pszczołka" && git config --local user.email "test@email"

it-test>git config --local -l | grep user
=A3A,ta Pszcza3A,ka
l=test email

it-test>touch a && git add a && git commit -m "Test commit by Marcin Kłopotek"
root-commit) 238b0ff] Test commit by Marcin Kłopotek
hanged, 0 insertions(+), 0 deletions(-)
ode 100644 a

it-test>git log
8b0ffd8c97fdfbc21ec62d3cca5213ff32bae
C5><BB><C3><B3><C5><82>ta Pszcza<C3><B3><C5><82>ka <test@email>
ed Dec 14 09:41:02 2016 +0100
commit by Marcin K<C5><82>opotek

it-test>chcp
de page: 437

it-test>
```

LECTURE 02

Writing & Data

Manuscript Tracking & Handling Large Files

The Art of the .gitignore File

In a PhD, we produce lots of "noise" files. Use .gitignore to keep your repo focused on content, not clutter.

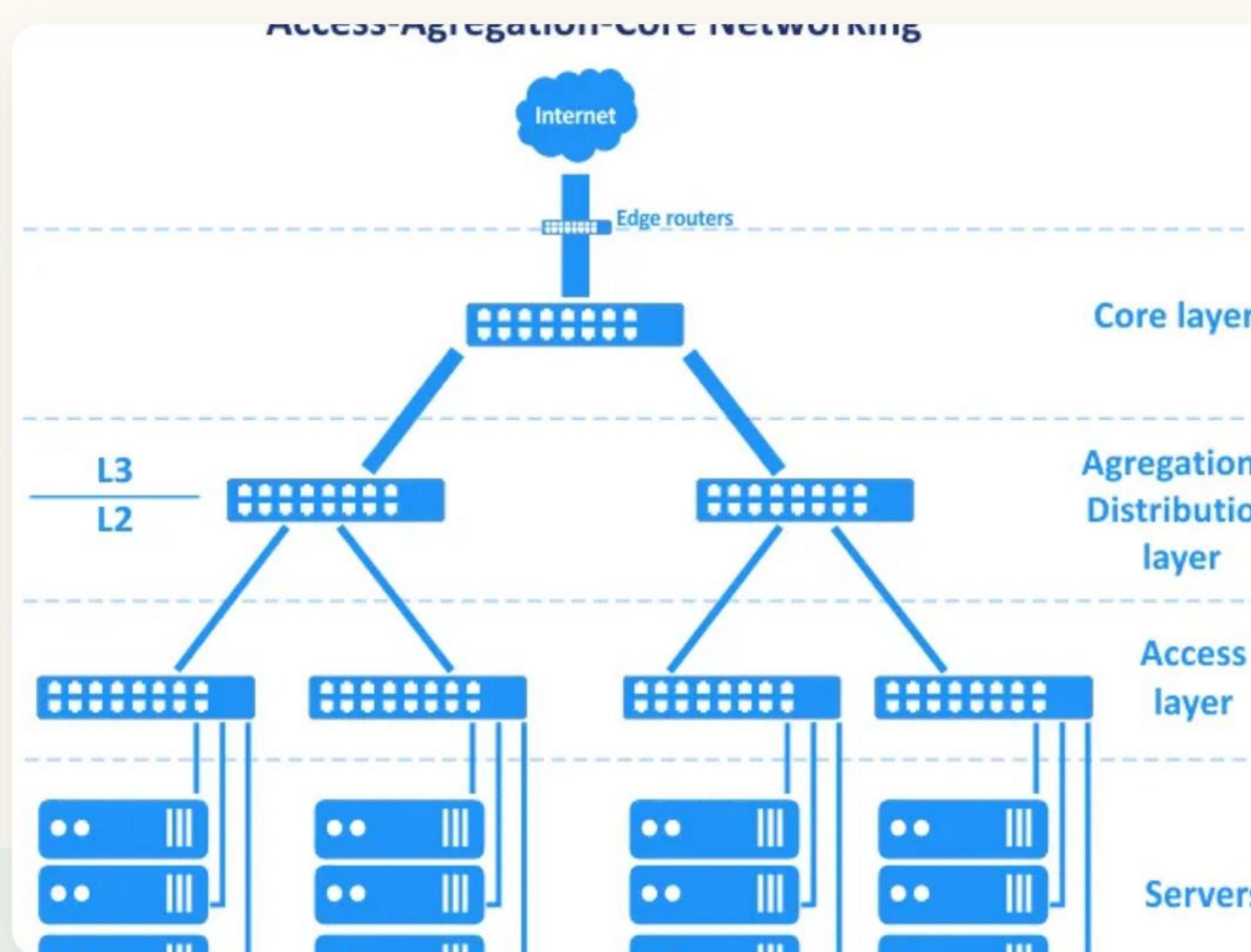
 **LaTeX Noise:** Ignore .aux, .log, .toc, and .out files.

 **Python/R Junk:** Filter out __pycache__, .Rhistory, and .DS_Store.

 **Sensitive Data:** Ensure raw, identifiable data stays local.

 **Local Configs:** Block IDE settings like .vscode or .idea.

Handling Massive Datasets with LFS



Breaking the 100MB Barrier

GitHub limits file sizes, but PhD data (Genomics, LiDAR, Surveys) can be huge. Git LFS (Large File Storage) is the solution.

- ⌚ Pointer Files: Git tracks a tiny text pointer in the repo.
-]=] Remote Storage: The 1GB CSV lives on a dedicated LFS server.
- ⬇ Smart Fetch: Only download large files when you checkout.

LECTURE 03

Collaborative Science

Project Management & Scientific Presence

Project Management for Chapters

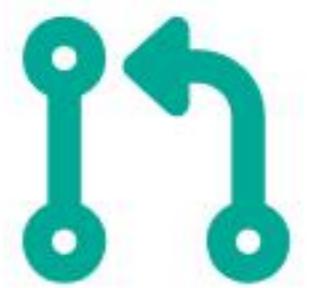


GitHub Issues

Track bugs in code or 'To-Do' lists for Chapter 1 revisions. Link them directly to lines of code.

Projects (Boards)

Visualize your dissertation timeline. Drag tasks from 'To-Do' to 'In-Progress' to 'Submitted'.



Pull Requests

The standard for co-author review. Supervisors can leave comments on specific lines of your manuscript.

From Code to Citation: Zenodo DOI

Phase	Platform Tool	Research Outcome
Active Analysis	GitHub Repo (Private/Public)	Complete audit trail of logic.
Submission	GitHub "Release" v1.0	Snapshot of code for peer-review.
Publication	Zenodo Integration	Generates a citable DOI for your code.

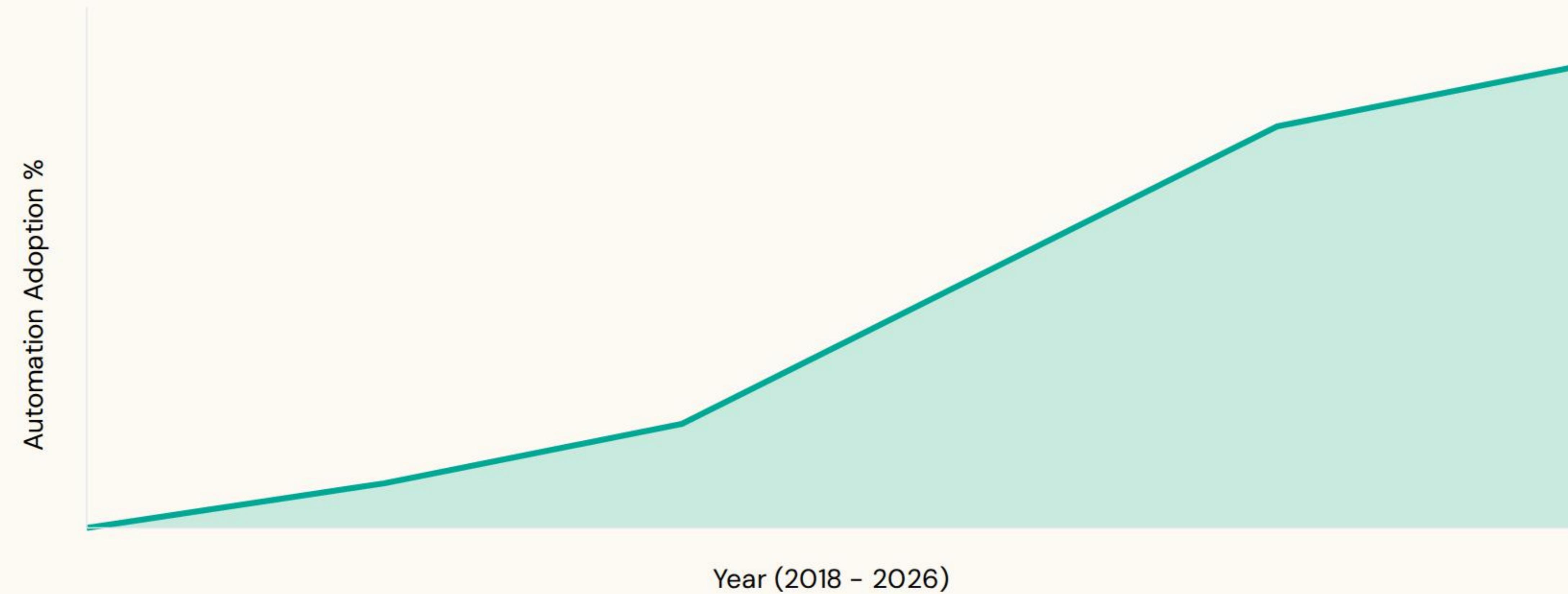
$$\text{Reproducibility Index} = \frac{\text{Citable Code (DOI)+Open Data}}{\text{Rigorous Peer Review}}$$

LECTURE 04

The Automated Researcher

GitHub Actions & Personal Portfolios

Trend: The Rise of Scientific Automation



Automation via GitHub Actions is becoming the standard for open-source PhD workflows.

Leveraging GitHub Actions (CI/CD)



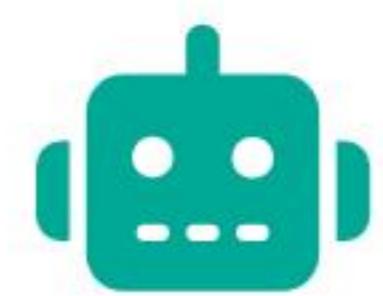
LaTeX Auto-Compile

Every time you push a .tex file, GitHub Actions compiles the PDF and uploads it as an artifact.



Data Validation

Automatically run Python/R tests to ensure a new data update didn't break your statistical assumptions.



Daily Scraping

Schedule a 'Cron Job' to scrape data or fetch news every morning at 9:00 AM automatically.

Personal Portfolios

GitHub Pages: Host your professional academic website for free. Link it to your repo to display dynamic results, interactive plots, and your CV.

Using Jekyll or Hugo, you can turn your research code into a beautiful blog that reaches a global audience beyond the paywall.

Rising senior at the University of Pennsylvania studying Finance Management with career interests in tech, sports, and music.

[Intro](#)

I was born and raised in Southern California, a Lakers town and always had good basketball experiences. These threads of family, basketball, and experiences have led me to my path and experiences. I am entering the Wharton School of Business. As a member of Penn's Women's Basketball Team, I will be able to continue my passion for basketball while also pursuing my studies at one of the best business schools in the country.

Course Completion Checklist

12

CORE COMMANDS MASTERED

0

FILES NAMED "FINAL_V2"

1

CITABLE DOI FOR THESIS

You are now ready to lead the future of Open Science.

Final Q&A

Next Steps: Initialize your first 'Research-Ready' repository today.

Assignment: Create a .gitignore and README.md for your current project.

Contact: github-tutor@university.edu

Image Sources



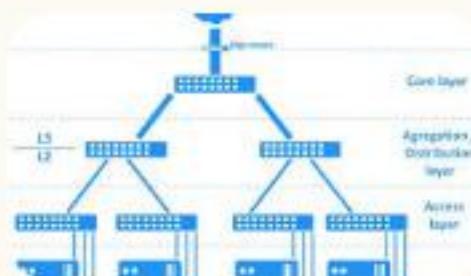
<https://media.slidesgo.com/storage/47143388/0-blue-aesthetic-portfolio.jpg>

Source: slidesgo.com



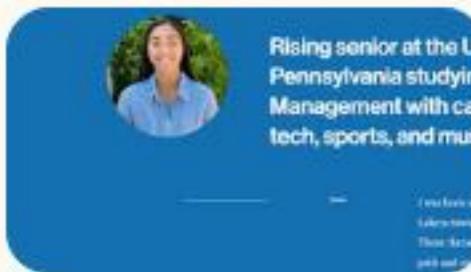
<https://i.sstatic.net/C3IKW.png>

Source: stackoverflow.com



<https://www.nakivo.com/blog/wp-content/uploads/2021/08/1.-The-Access-Aggregation-Core-network-topology-type.webp>

Source: www.nakivo.com



<https://ucarecdn.com/89016ecf-58d5-4da8-8e6a-cca57a3c25c0/>

Source: www.sitebuilderreport.com