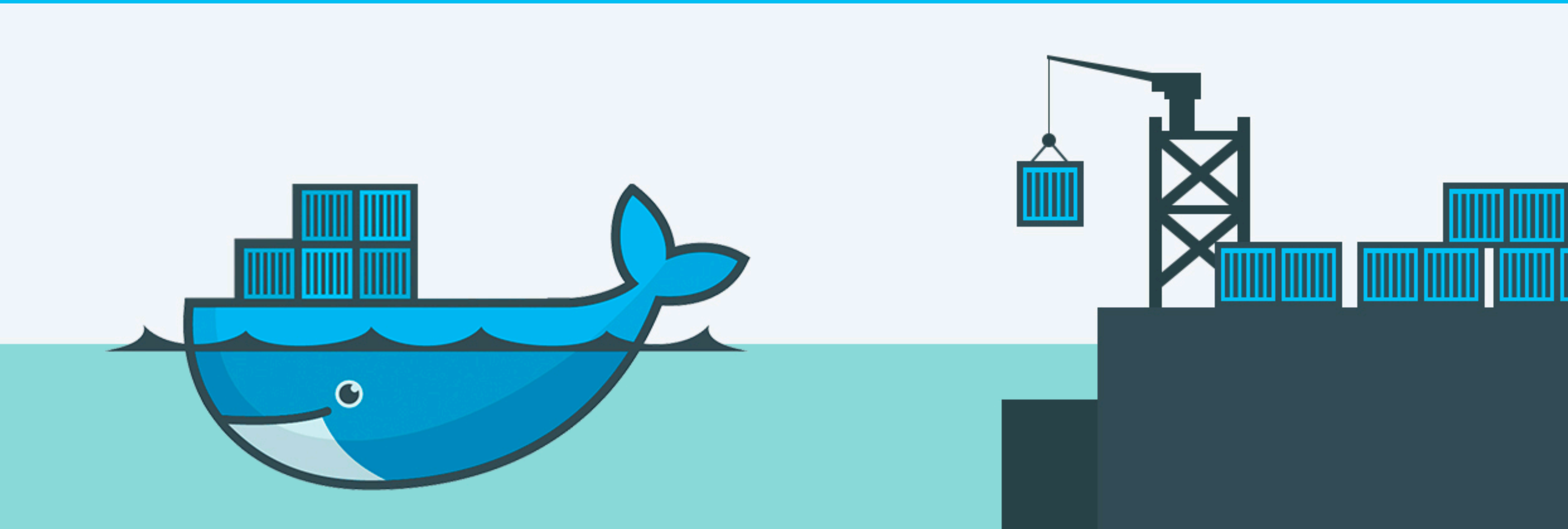
DOCKER AND YOUR RESEARCH

PRESENTED BY – ZEYI WANG



TOPICS

1. What is Docker?
2. What can Docker do?
3. Why Docker?
4. How to use Docker for your research project?
5. Demo time!

WHAT IS DOCKER?

A bunch of tools for containerization.

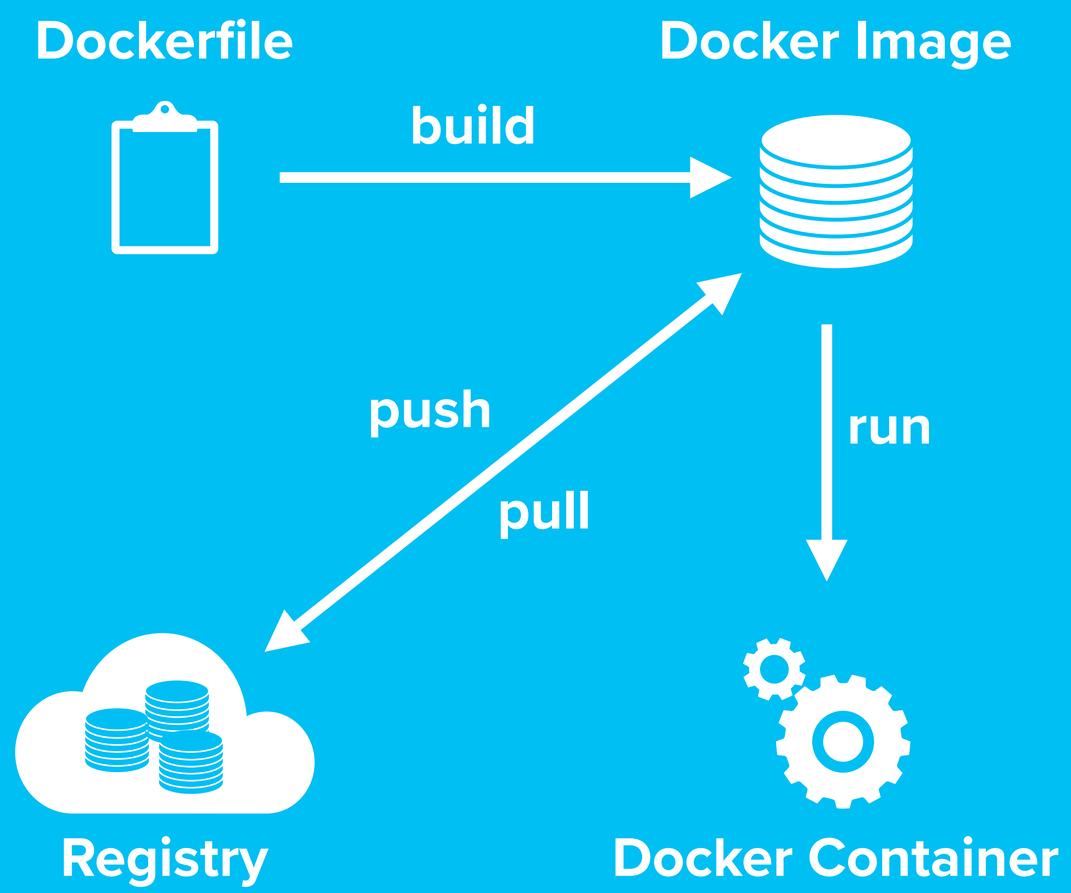
WHAT IS CONTAINERIZATION?

- (I assume you know what is virtualization)
- Just a better way to do virtualization.
  - Minimum (only pack things you need)
  - Easy to setup (can be built from a text file)
  - Efficient (faster on everything)
- (technically incorrect but you get the point)

WHAT CAN DOCKER DO?

- docker build
- docker run
- docker push/pull

- Dockerfile – source file
- Docker Image – class object
- Docker Container – instance



WHY DOCKER?

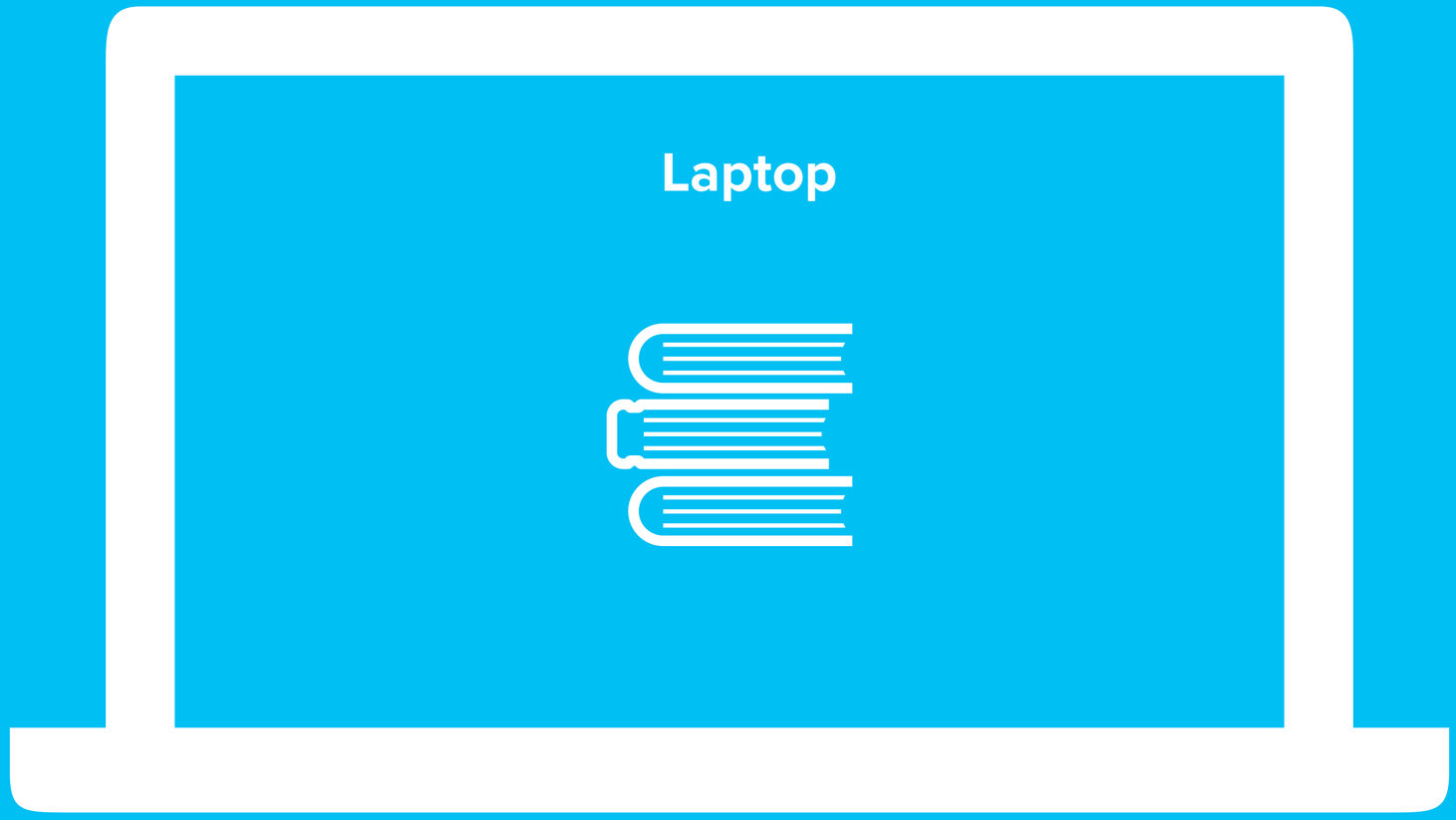
- Research Debt
  - Reproducibility
    - Should be easy for CS projects. Reality? Head scratching… even for your own projects
  - Interpretive Labor (either on the deliverer side, or on the receiver side)
  - I promise I will make my best effort for reducing such debt in my future research
- Because you have to (cirrus)

HOW TO USE DOCKER FOR YOUR RESEARCH PROJECT?

- Things to use
  - Version Control System (Git)
  - Source Code Hosting (GitHub)
  - Continuous Integration Service (GitHub Actions)
  - Docker Registry Service (GitHub Packages)
- Best Practices
  - Keep it simple (ignore other best practices)
  - Define your environment (Pipfile, Dockerfile)
  - Separate implementation with configuration (have a separate “experiments” directory)
  - Make things easy to change (use “docker run -it <image>”)

EXAMPLE WORKFLOW

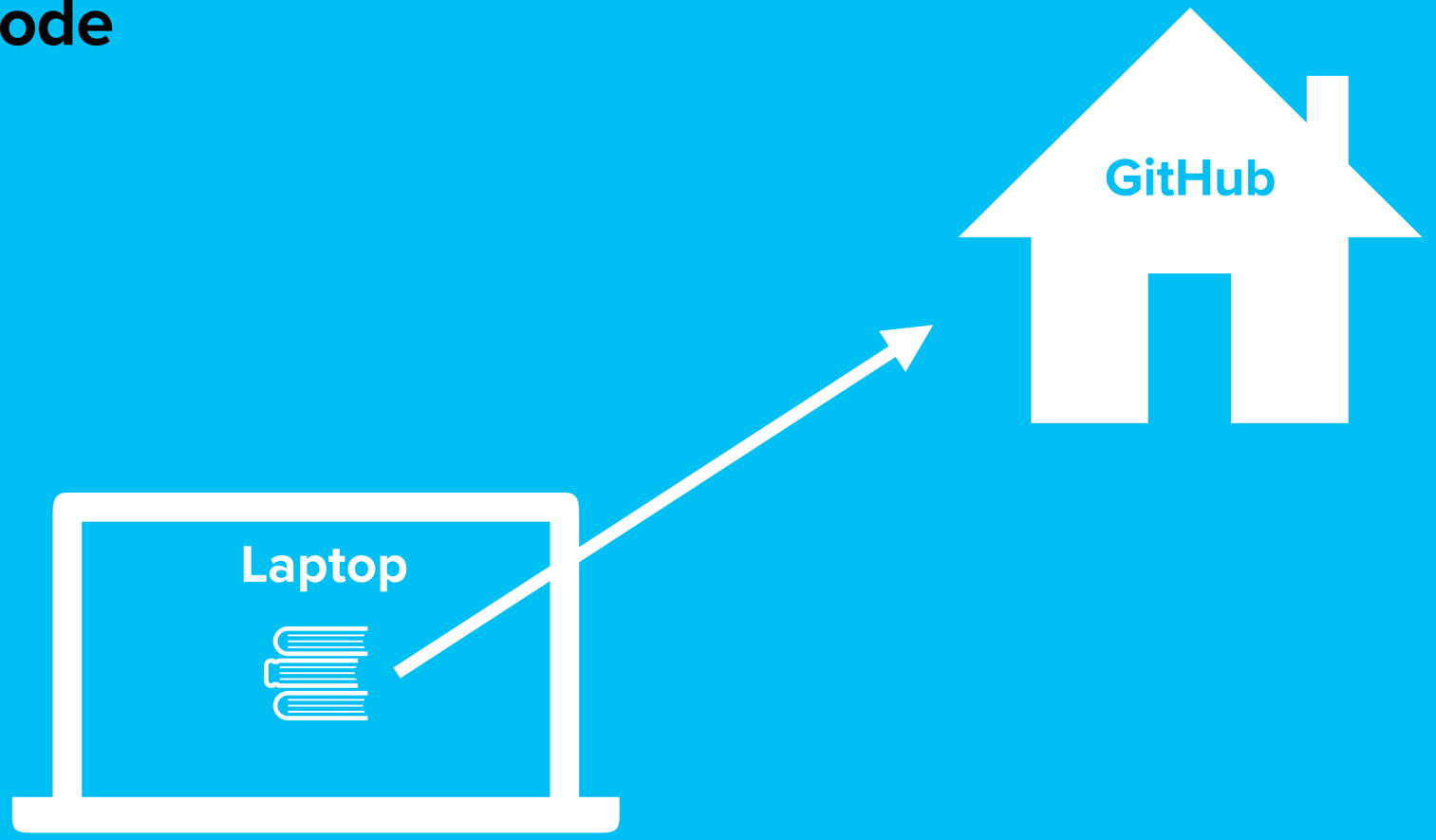
Local: develop, run



EXAMPLE WORKFLOW

Local: develop, run

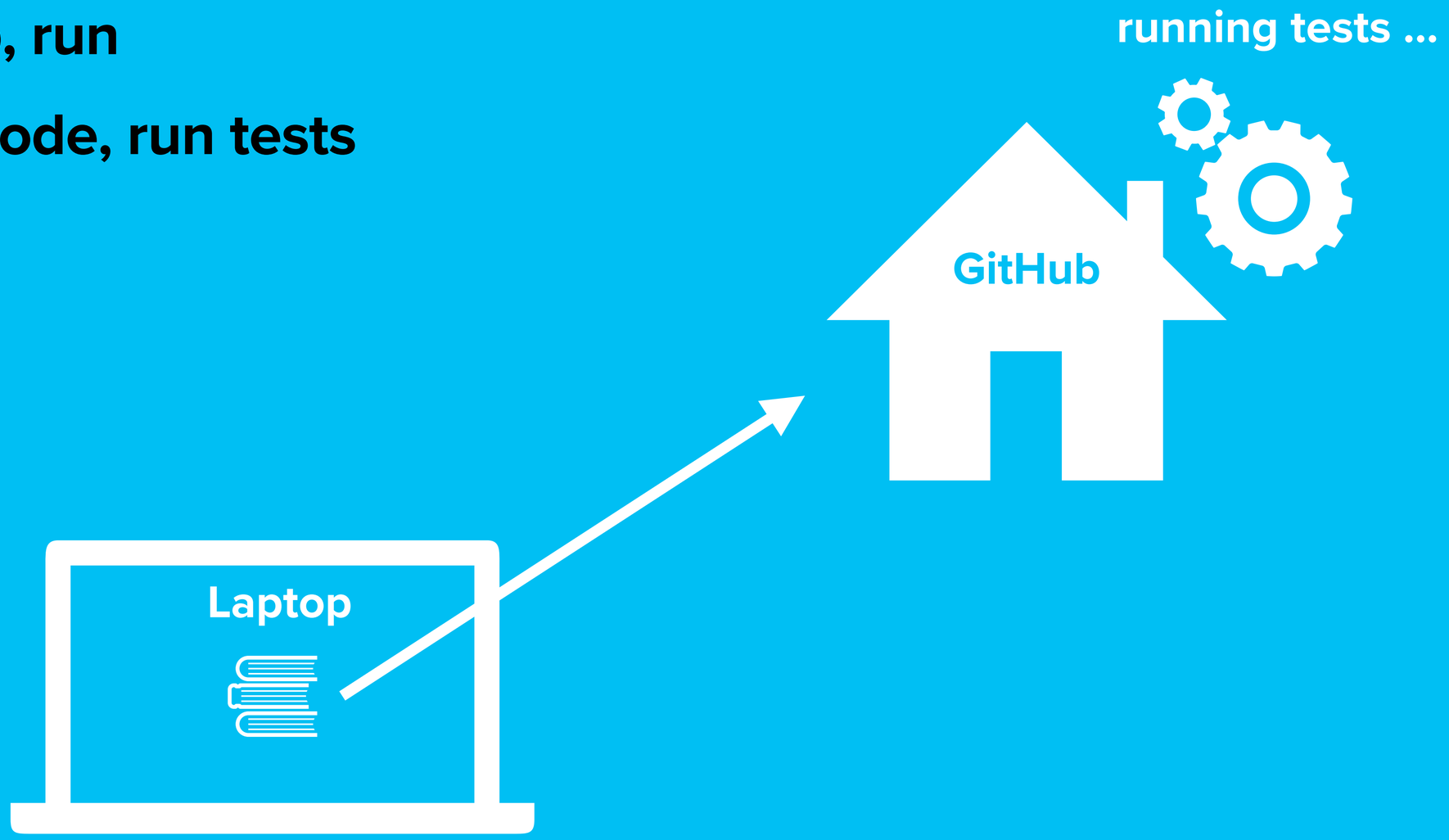
GitHub: host code



EXAMPLE WORKFLOW

Local: develop, run

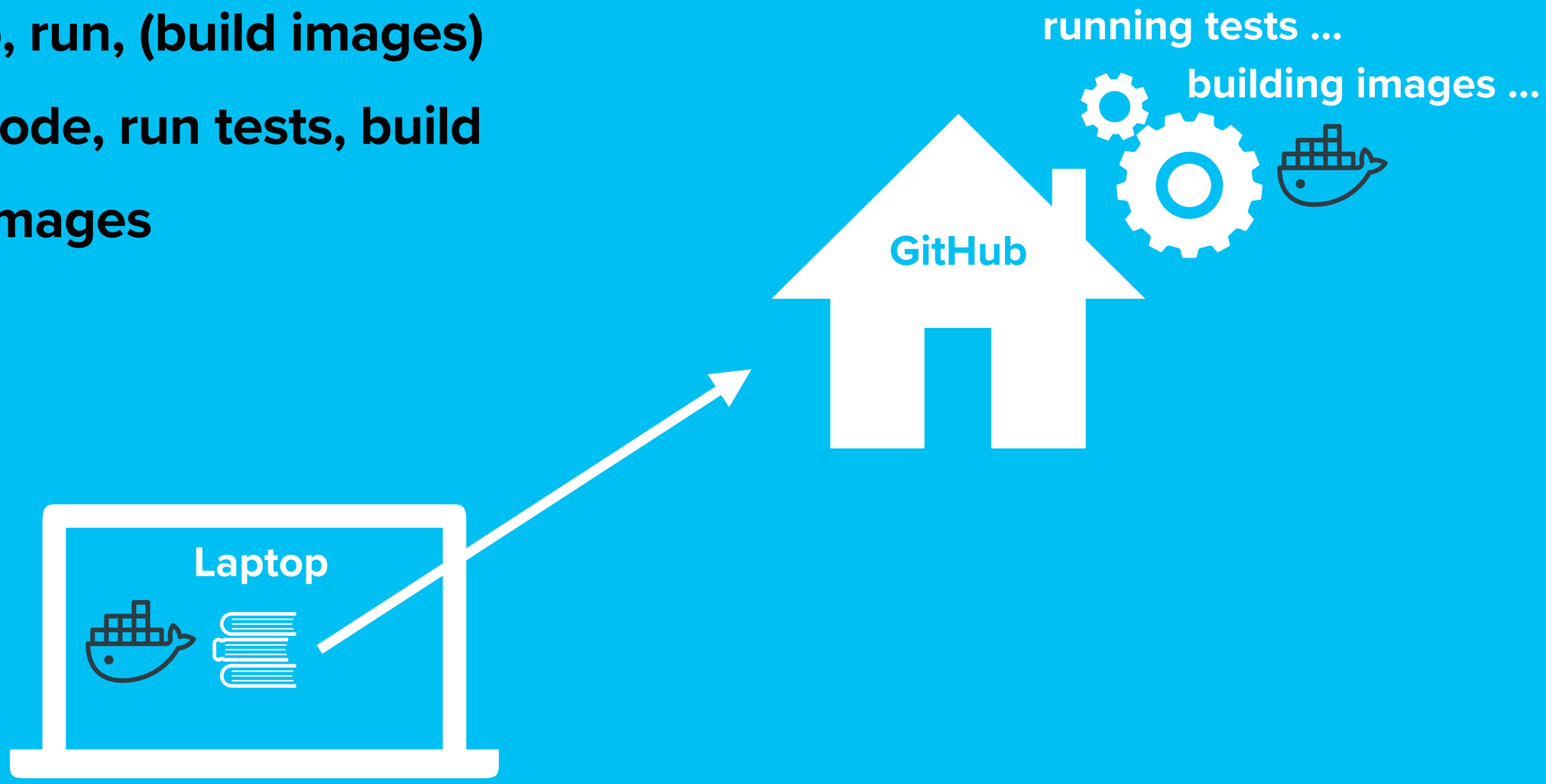
GitHub: host code, run tests



EXAMPLE WORKFLOW

Local: develop, run, (build images)

GitHub: host code, run tests, build images, host images

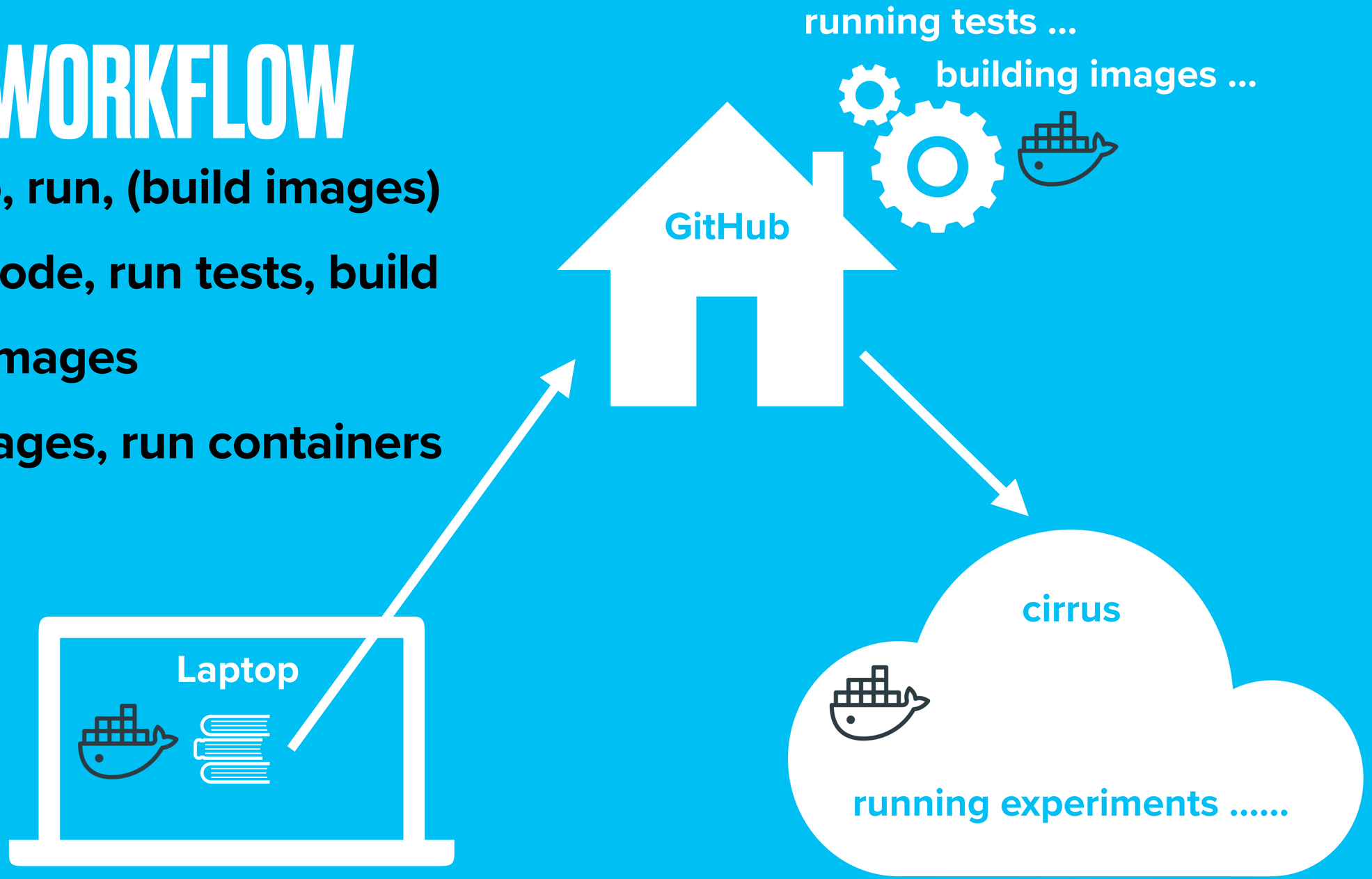


EXAMPLE WORKFLOW

Local: develop, run, (build images)

GitHub: host code, run tests, build images, host images

Cirrus: pull images, run containers

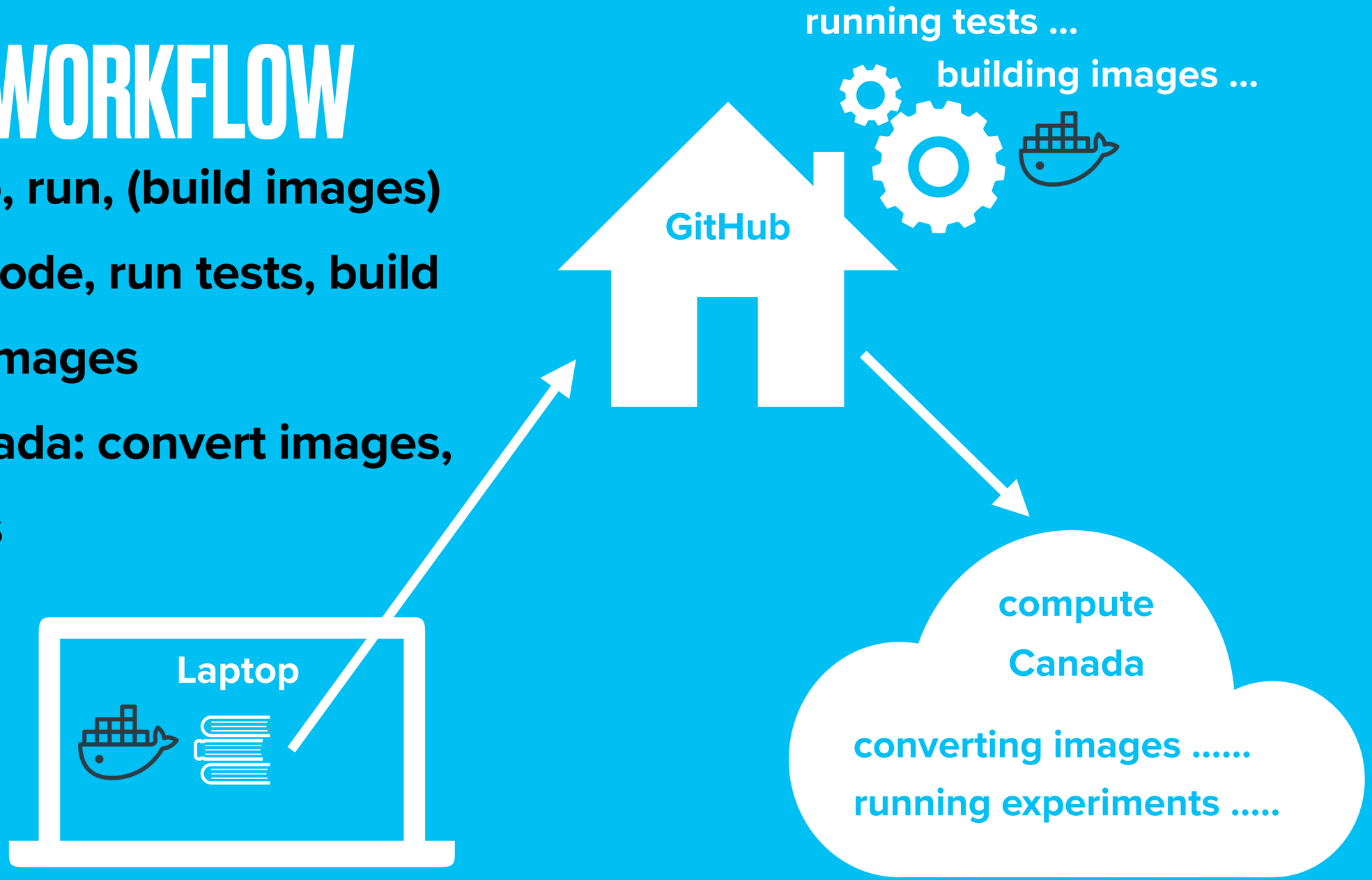


EXAMPLE WORKFLOW

Local: develop, run, (build images)

GitHub: host code, run tests, build images, host images

Compute Canada: convert images, run containers



DEMO TIME

REFERENCES

An introduction to Docker for reproducible research, with examples from the R environment
https://arxiv.org/pdf/1410.0846.pdf

Research Debt
https://distill.pub/2017/research-debt/

Demo Code
https://github.com/uduse/docker-based-research-workflow