The Carpentries Offline: Teaching Foundational Data Science and Coding Skills with little or no Internet Access
Who are we?

Here today:
➢ Jannetta Steyn (Newcastle University)
➢ Abhishek Dasgupta (University of Oxford)
➢ Colin Sauze (National Oceanography Centre)

The other guys:
➢ Samantha Finnigan (Durham University)
➢ Ethan White (University of Florida)
➢ Virnaliz Cruz (University of Florida)
➢ Frances Turner (Newcastle University)
➢ Andrew Gill (Stellenbosch University)
What is the Carpentries?

“We teach foundational coding and data science skills to researchers worldwide.”

- **Vision**: to be the leading inclusive community teaching data and coding skills.
- **Workshops**: Software Carpentry, Data Carpentry, and Library Carpentry
- **Roles**: Instructors, helpers, Trainers, Maintainers, Mentors, and Core Team
- **Technologies**: Web based course notes, etherpad for shared notes, Github, Jupyter Notebooks
How it all started

• Software Sustainability Institute Collaborations Workshop Hackday 2021
• Running workshops without Internet access
• Use Raspberry Pi as an access point and web server
• Hackday winner
• SSI Fellowship 2022

Original Team:
- Flic
- Alison
- Abhishek
- Emily
- Irma
- Jannetta
- Rebecca
- Sam
- Talia
Offlinedatasci mirrors installers and repositories to enable offline installation

pip install offlinedatasci
offlinedatasci install all /install/path

Developed by a team at University of Florida (Ethan White and Virnaliz Cruz) and us.
What we mirror

➢ Installers for Python and R
➢ Partial mirrors of PyPI, CRAN (packages can be customised)
➢ Carpentries online material
➢ Installers for data science IDEs (RStudio)
Three threads to our project

➢ Using a Raspberry Pi
➢ Using a bootable flash drive
➢ A mini HPC for HPC workshops
Option 1

The Problem

The Solution

- RaspOS
- Gitea
- Etherpad
- Lesson Mirrors
- CRAN and PyPi mirrors
What it looks like
Building the Raspberry Pi Image

➢ Entire build process is scripted.
➢ Building images on the Raspberry Pi is a manual and slow process.
➢ Cloud based GitHub actions build in a Raspberry Pi emulator (Qemu).
  ○ Emulators are slow! Takes 2+ hours to build
  ○ Some hacks to speed things up!
CarpentriesOffline in the Cloud

➢ Docker container using CarpentriesOffline build script
➢ Originally intended for testing.
➢ Much faster than using a Raspberry Pi or an emulator.
➢ Useful for when the Carpentries website/etherpad goes down during your workshop!
➢ Can be hosted in intranet
The Need for Alternatives

➢ RPIs are impossible to get hold of since Covid
➢ RPIs cost money
➢ I already have a laptop
Option 2

- Bootable flash drive
- Slax Linux
- Apache2 (web server)
- Gitea (alternative to GitHub)
- OfflineDataSci – our python package for scraping all Carpentries lessons
The Need For A miniHPC

• Hardware more visible
• Hit resource limitations more easily so more obvious
• No accounts to be setup on a real HPC
• No interfering with real HPC
  – users less afraid to try stuff
  – less likely to break anything important
  – no access to a real HPC
• Access problems
• Networks access
miniHPC Specs

**Pixie the Prototype**
- 3 x Raspberry Pi 4 B
- 1 x Raspberry Pi 4 B head/login node
- Raspberry Pi OS Lite (64 bit)(Debian Bullseye)
- Head node acts as WiFi access point

**RockPi**
- 8 x Rock 4C+ (Dual ARM Cortex-72 @1.5GHz per node)
- 1 x Rock 4SE head/login node
- 8 x Power over Ethernet hats
- *Raxda* build of Debian Bullseye
- Head node acts as WiFi access point
HPC Software

- Slurm
- Lsmod
- Munge
- NFS
- PXE
- EasyBuild

- dnsmasq
  - DHCP
  - tftp
- mpich
- gcc
- python
3D Printing Credit

- https://www.printables.com/@TaylorSteinf_1252185
- https://www.printables.com/model/717134-mini-caliper-10-cm
- https://www.thingiverse.com/thing:2424354
- https://www.printables.com/model/44122-customizable-belt-buckle
Raspberry Pi Image download:
https://github.com/carpentriesoffline/carpentriesoffline-installer/releases

Find Us:
https://carpentriesoffline.org

Slack channel
https://carpentries.slack.com/archives/C03KYQ3PX99