Jupyter Hubs for Education

5 years

2000+ students a semester (600% increase)

26 California Community Colleges

8 California State Universities and UC

Here is what we did …

Infrastructure

What courses need:

- Zero-Configuration by Student or Instructor
- Single-Sign On
- Consistent Environment for all students
- Funded Compute(!)
- User Telemetry (A goal!)

What users need:

- **Professors:** Minimize steps needed to teach
- **Students:**
  - No configuration
  - No package management
  - Eliminate anything that requires experience outside of what they are learning

What we offer:

- **CPUs:** 2
- **RAM :** 1 - 2 GB
- **Dev Environments:**
  - Jupyter Lab and Jupyter Notebook
  - VS Code
  - RStudio
- **Languages:**
  - Python, R, C++

Jupyter Hub Infrastructure

- **JH Deployments:**
  - Managed by 2i2c and Shane Knapp
  - GH Actions: Automated deployments via hubploy
  - 5 minutes of my time; 10 minutes to deploy
- **User Images:**
  - repo2docker configured user images

Jupyterhub Authentication

- **CiLogon:**
  - Creates Single-sign on either via:
    - Institutional Registration
    - Google or Microsoft IDP
- **GitHub OAuth:**
  - Rare but sometimes needed

Users

Who are we working with:

- Data 8 Adopters at Community Colleges
- Statistics, Math, and Computer Science:
  - Private and Public (out of and in state)
  - California State Universities
  - University of California

Instructors: Two Types

- **Instructor A:**
  - Experience with notebooks/browser-based computing
  - Already has course and materials
  - Used alternatives - looking for sustainable solution
- **Instructor B:**
  - Unfamiliar with Jupyter Notebooks, cloud infrastructure, Git/GitHub
  - Lots of experience teaching - making the move to computational-based teaching and learning

Students

- Beginning Data Science or Computer Science
- Need libraries and packages handled
- Learning file management
- No or little experience git/github experience

⇒ **Providing a smooth and efficient path to learning content with computing tools**

Experienced Instructors: Done

✅ Jupyter Hub infrastructure
✅ Customizable Images
✅ Authentication

⇒ Support:

- Package installation
- User issues related to browser-based computing

New Instructors: Course Delivery

- Our Focus: Data 8 (and Data 6)
  - Notebooks (31 in Data 8!)
  - Canvas Shells
  - Automatic Grading (otter-grader)
  - Lecture Slides
  - Discussion Slides
  - Online Textbook

Course Canvas Shell

- Lays out entire course - week-by-week
- Slides, videos, and readings are embedded
- nbgitpuller links tied to notebooks are configured to institutions hub

⇒ **Support:**

- One-time configuration of links per institution

Distribution of Notebooks

✅ nbgitpuller: url to open a notebook in a hub
✅ Backed by GitHub(Fork our repo!)
✅ Links are embedded in Canvas Shell

⇒ **Support:**

- Any package problem difficult to fix mid-term
- Notebook updates require syncing repo!

Automatic Grading: Otter-Grader

- Web-based service for batch grading otter-grader configured Jupyter Notebooks

⇒ **Support:**

- Largest amount of our time(!)
- Debugging student notebooks
- Supporting new teachers use the system

Last Mile: Browser-based Teaching

- **Problem Handling:** Needs Immediate Attention
- **Instructor Time:** Extremely tight and pulled on constantly
- **Non-Traditional set of tools:** everything needs to work to deliver the course smoothly

Instructor Support Structure

- One-on-One Orientations
- Support for notebook maintenance and updates
- Support for grading
- Support for handling hub-related issues
- Community Of Practice: Meets once-a-month
- National DS Workshop at Berkeley every summer

Future

Open Source Education?

- How do we empower instructors to create, modify and contribute to open source education materials - Jupyter notebooks, R, or Python?
- Lots of Documentation and tried various forms of training

The Future

- Developing Community of Practice
- Collaboration Tools
- AI Tutors (by instructor choice!)
- Perfecting and Refining our Trainings
- Building up Open Source Notebooks and Courses via Modules

⇒ **Scale Nationally - Spread the word!**

The Team

- Funding From:
  - CloudBank
  - California Learning Lab
- 2i2c - Yuvi Panda, Chris Holdgraf
- UC Berkeley Team:
  - Shane Knapp : Hub Development
  - Eric Van Dusen : Project Development
  - Sean Morris: Support and Outreach
  - Kseniya Usovich: Community Development
  - Jonathan Ferrari/Edwin Vargas Navarro: DS Modules

Useful References

- Course Adoption Documentation
- repo2docker
- hubploy
- nbgitpuller
- 2i2c.org
- otter-grader
- California Hubs: cal-icor.org
- CloudBank Classroom : National Hubs