SPOC to MOOC, Extending Local Training to the HPC Community

Julia Mullen and Lauren Milechin

September 25, 2019

Massachusetts Institute of Technology

Outline

- Challenges in HPC Education and Training
- Introduction to MOOCS
- Case Study: Understanding HPC Workflows and How to Exploit Them
  - The initial one-on-one training
  - The first online course
  - The future refactored courses
- Lessons Learned

HPC Education Challenges

Audience

- Learners want
  - Formal training with certificates
  - Informal training to complete work task
- Diverse learner background
  - Range of ages
  - Range of computer literacy
  - Multi-cultural, multi-lingual

Content Selection

- Workflows vary across domains requiring different solution techniques
- Learners have a range of background and skills
- Content needs include
  - Basic Unix skills
  - Basic HPC concepts
  - HPC software development and troubleshooting

Delivery Mode

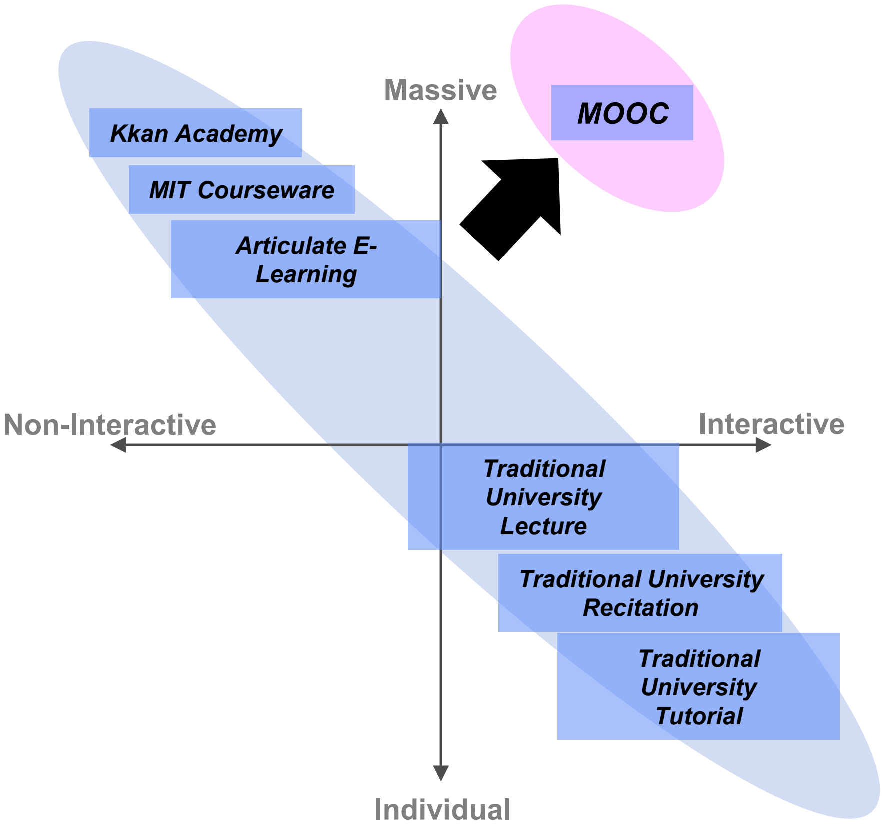
- In person workshops
  - Limited pool of expert trainers
  - HPC system access limited
- Web resources
  - Predominantly text
  - Tutorials for a given technology, e.g. MPI, OpenMP
  - Simplified examples, often tightly coupled to specific systems

Expanding HPC education and creating personalized “Just In Time” education and training is necessary but hard.

Outline

- Challenges in HPC Education and Training
- Introduction to MOOCS
- Case Study: Understanding HPC Workflows and How to Exploit Them
  - The initial one-on-one training
  - The first online course
  - The future refactored courses
- Lessons Learned

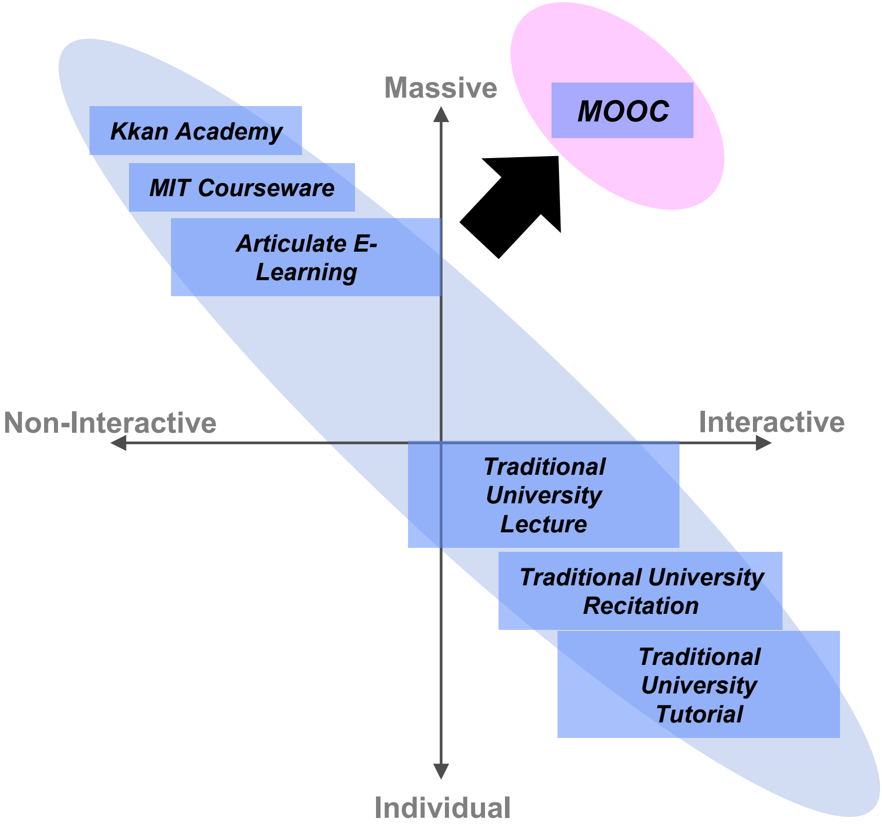
Why MOOCs?



Scaling

- 81 million learners across major providers
- Additional 13 million across independent Open edX sites
- 9+ thousand courses
- 25 languages (primarily Open edX)
- 33 providers worldwide

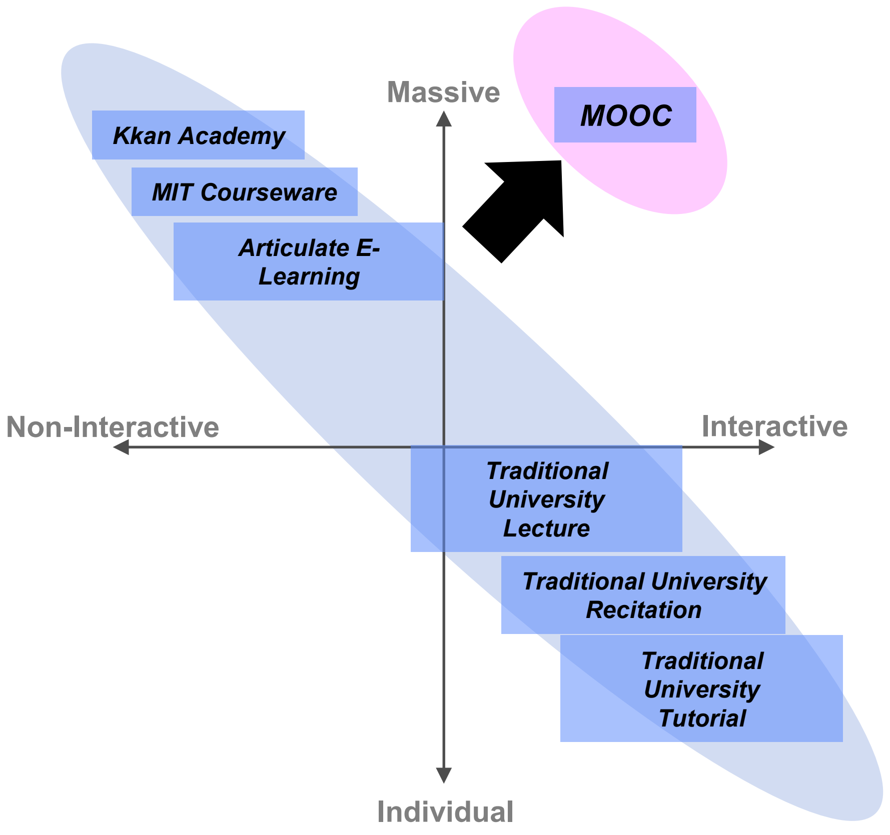
Why MOOCs?



Pedagogy/Andragogy

- Open
  - No pre-requites
  - Range of experience
- Online learning
  - Asynchronous
  - Self-paced
  - Instructor paced
- Social learning – interactions among diverse learner groups
- Built to support theory and practice

Why MOOCs?



Metrics

- Basic demographics
- Engagement with content
- Exercises and grades
- Surveys & feedback comments
- Learning paths
- Data informs course updates

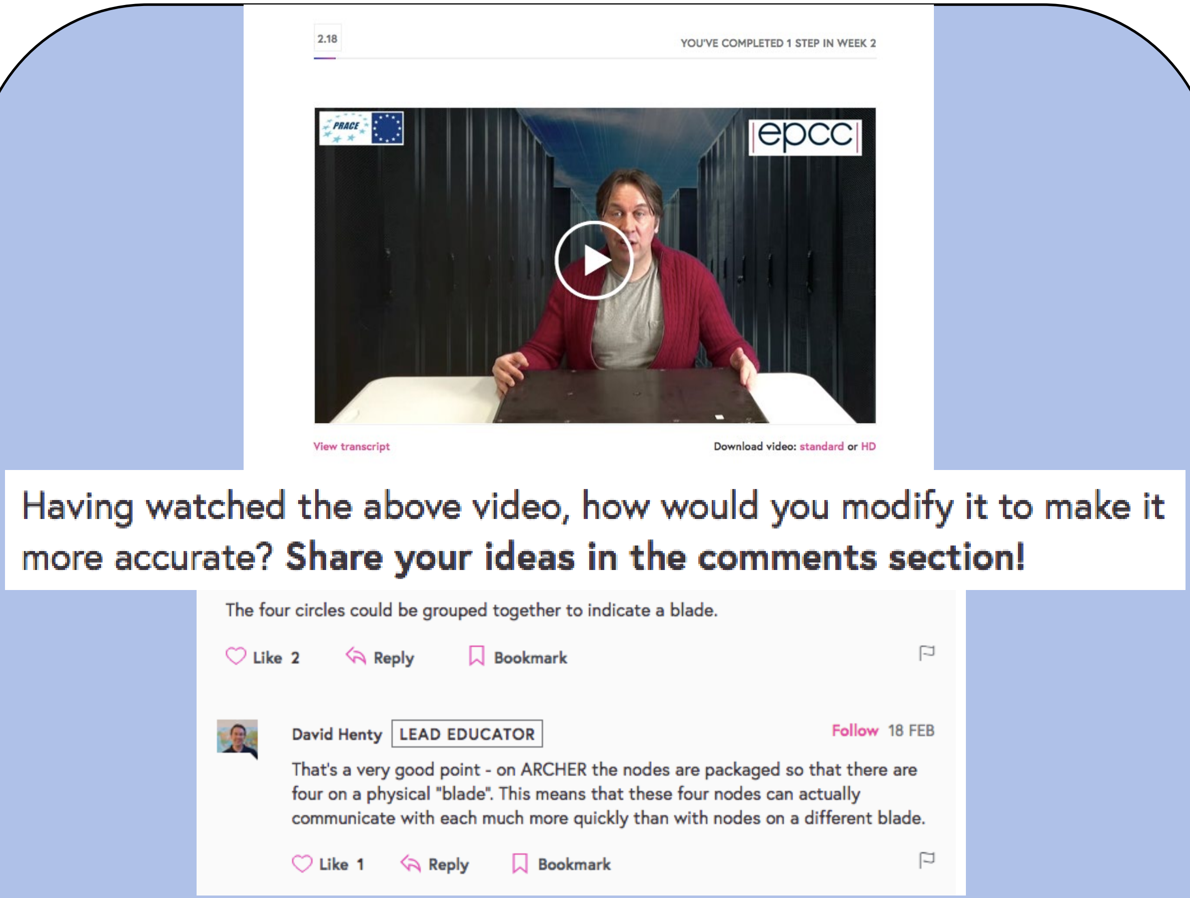
Open edX Platform

- Software stack publicly released in June 2013
- Stack includes integrated CMS and LMS
- Open edX community built around open source platform
- Statistics (as of March 2019)
  - 45 Million Learners
  - 24,000+ courses
  - 2400+ sites
  - 70+ Countries
  - 34 Languages
- Open edX Conferences since 2014

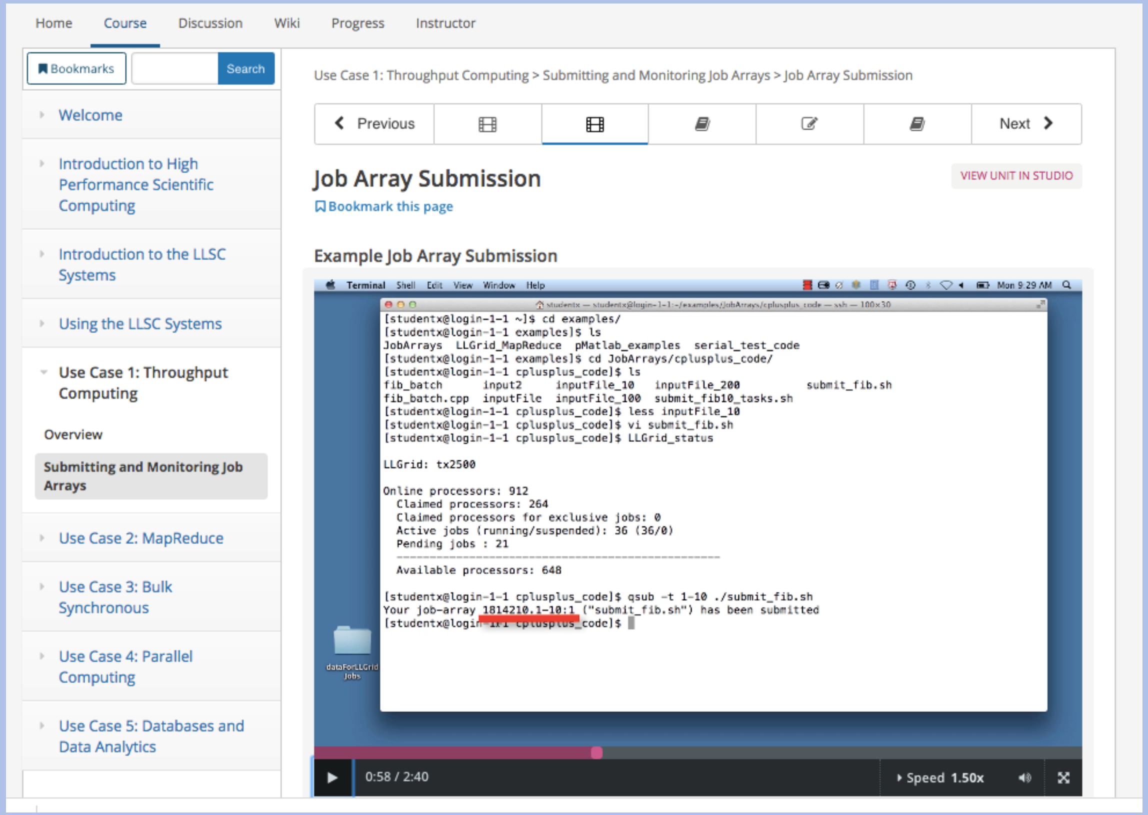
MOOC Design Considerations

- Content Selection
  - Partition material into easily absorbable segments
  - Segments must be self-contained, progression not always linear
  - Content must be clear and simple without unnecessary simplifications
  - Remove all redundant material
- Delivery
  - Vary delivery modes used to present the content, e.g. video, text, simulation
  - Select most suitable medium for content
  - Course structure must be transparent and easy to navigate
  - Hands-on exercises reinforce theory
- Learning experience
  - Provide optional activities and reference to additional information
  - Enable and encourage interactions between the learners
  - Provide learners with a variety of assessments to test their understanding
  - Little-to-no overhead in setting up hands-on exercises

HPC MOOC Examples

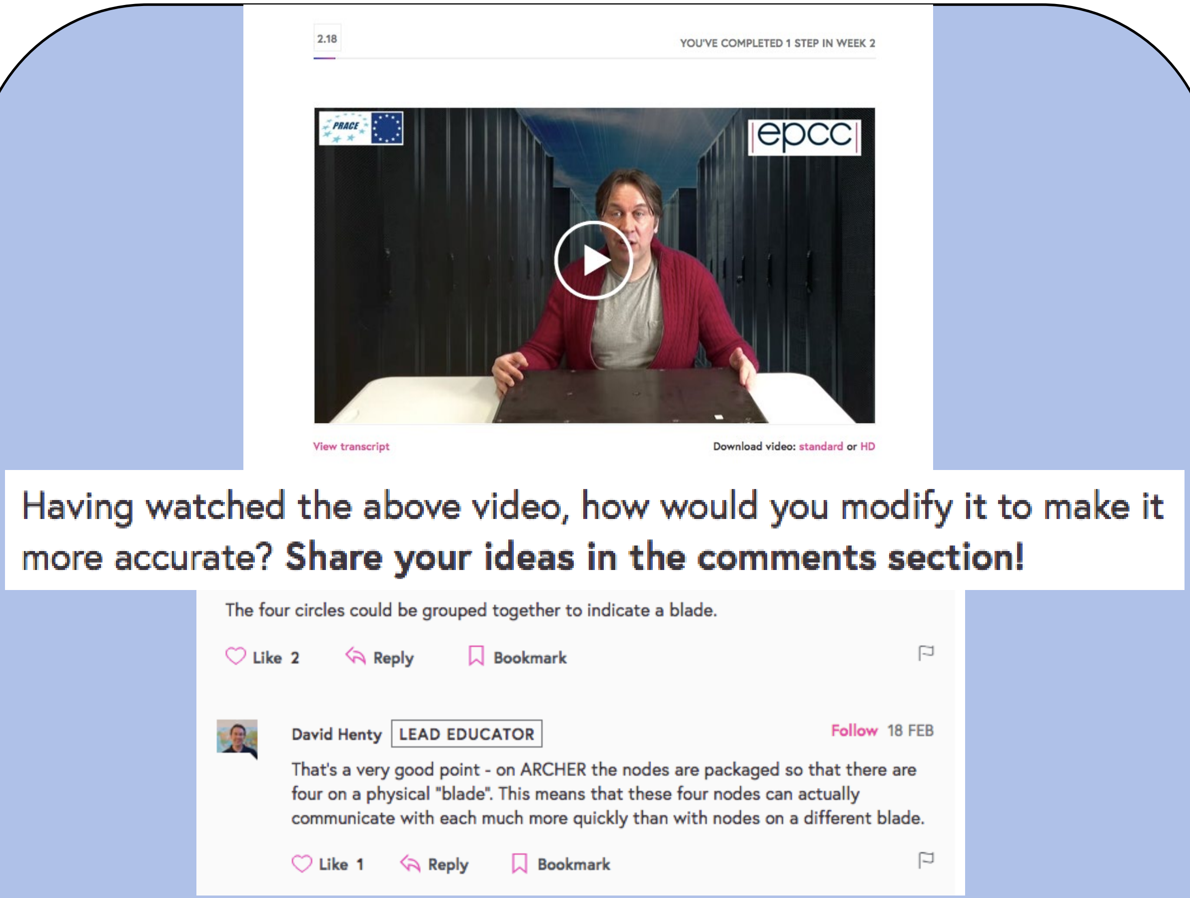


Supercomputing, FutureLearn (EPCC/PRACE)

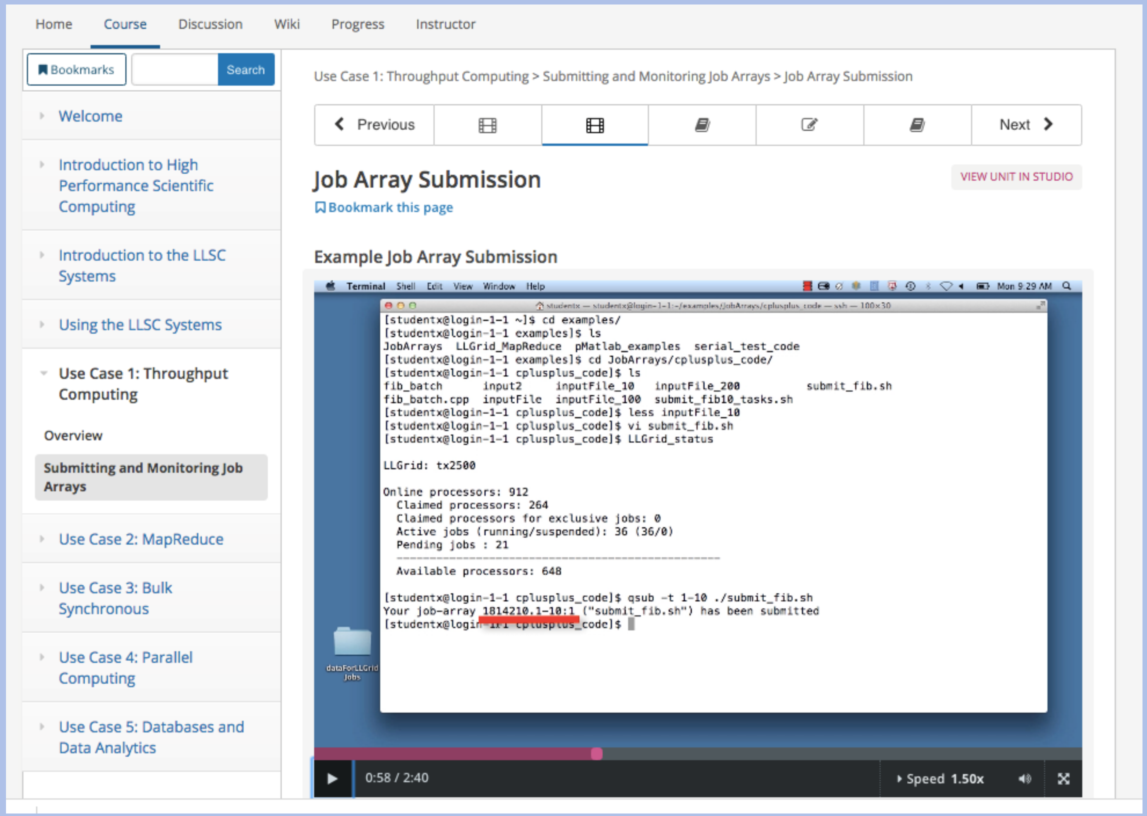


Using MIT Supercloud, LLx (Open edX) (MIT/LLSC)

HPC MOOC Examples



Supercomputing, FutureLearn (EPCC/PRACE)

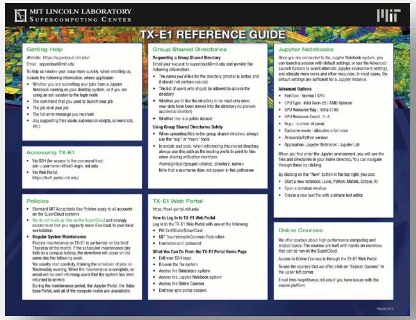


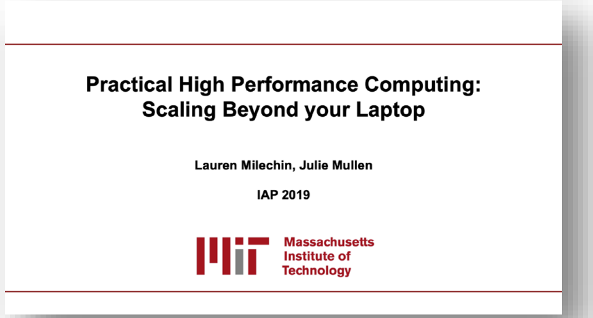
Using MIT Supercloud, LLx (Open edX) (MIT/LLSC)

Outline

- Challenges in HPC Education and Training
- Introduction to MOOCS
- Case Study: Understanding HPC Workflows and How to Exploit Them
  - The initial one-on-one training
  - The first online course
  - The future refactored courses
- Lessons Learned

Pre-MOOC Training

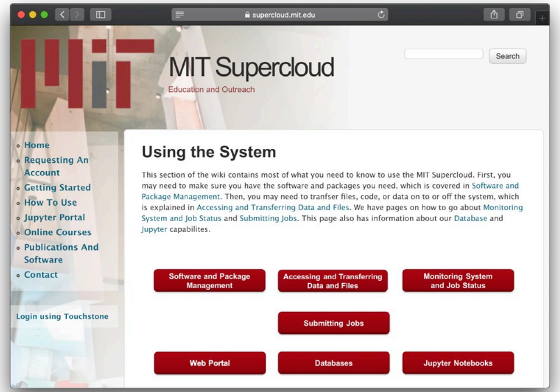


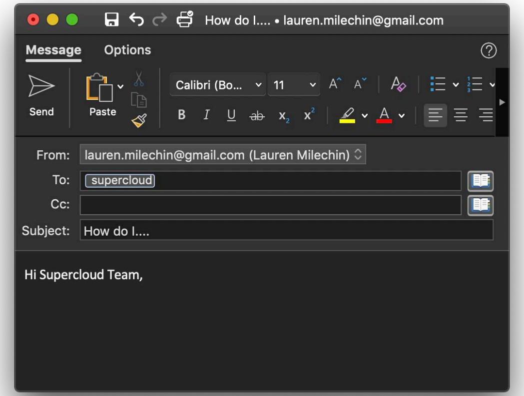


Reference Guides

IAP/Special Workshops

One-on-one Tutorials





Web Pages

Email

Scaling Step 1: Small Private Online Course (SPOC)

Goal: Develop a course for professionals and researchers that teaches strategies for building HPC workflows

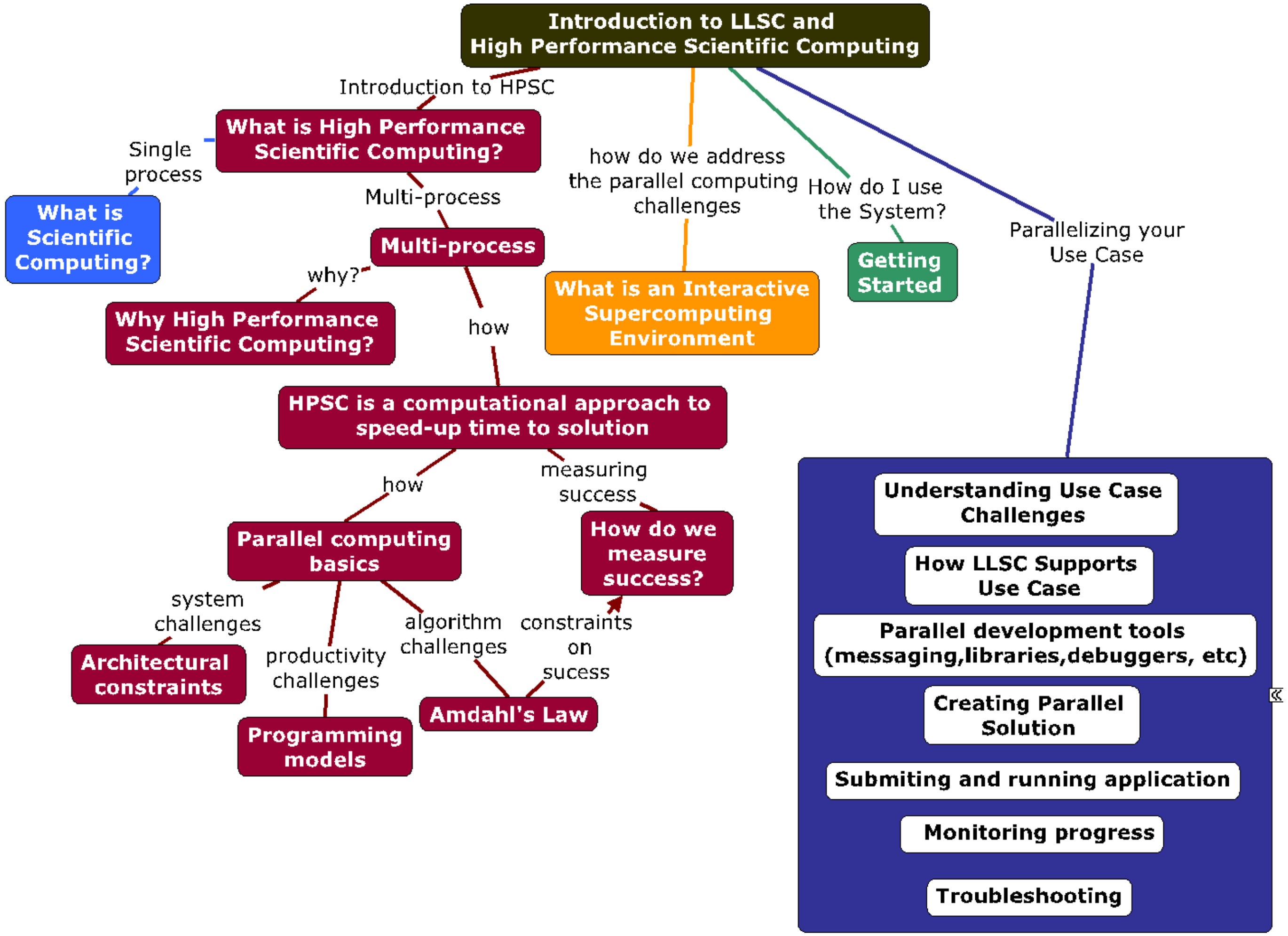
Audience

- Professional engineers and scientists who need to scale scientific workflows
- Diverse Learner background
  - Range of domains
  - Range of problem types
  - Range of computer literacy

Delivery

- A mixture of videos, text, programming exercises and quizzes
- Focus on learning through theory and practice
- Hands-on practice using HPC system
- Self-paced

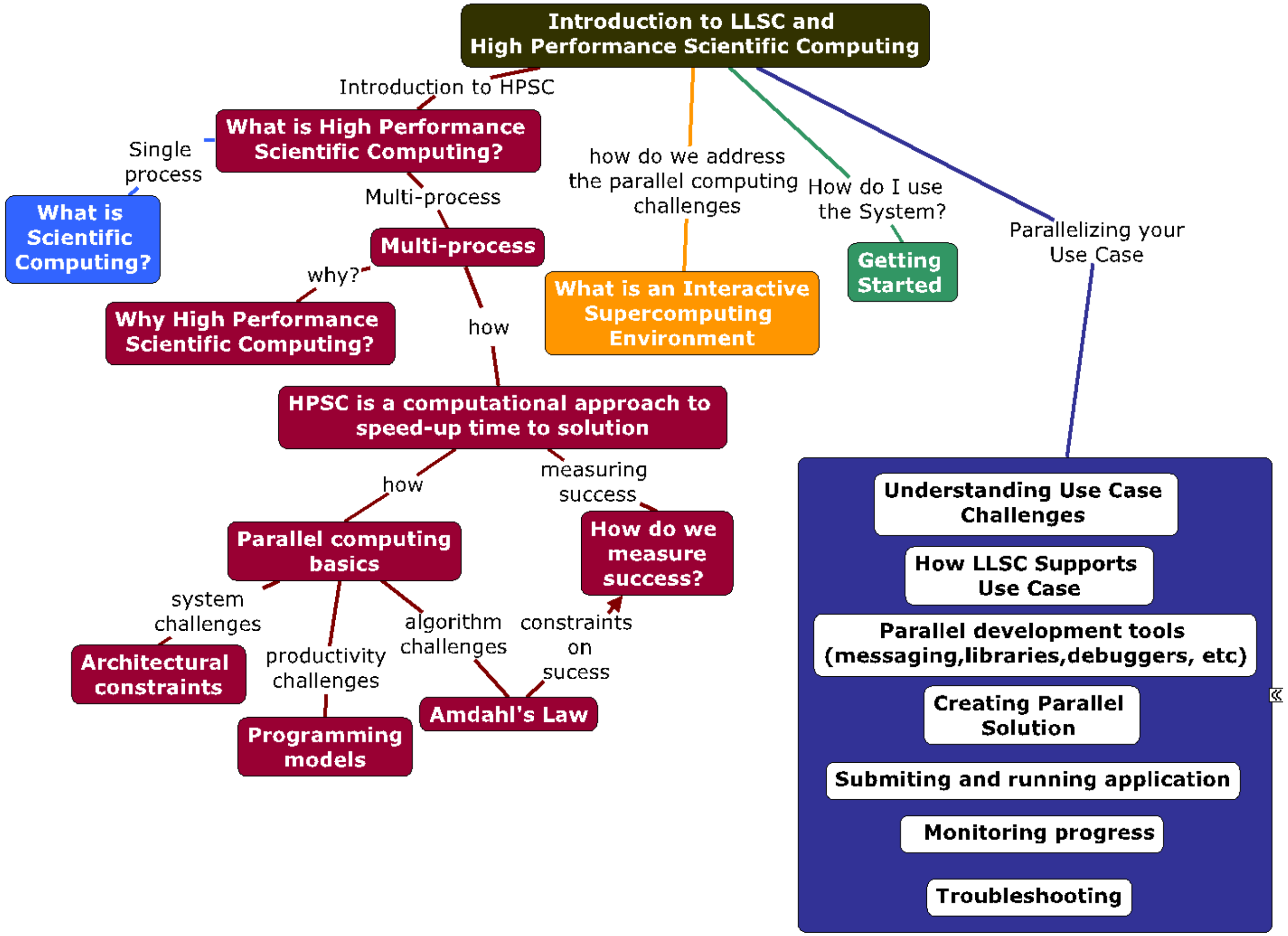
Design of Open Online Course



- **Andragogical Principles**
  - **Interleave Theory and Practice**
  - **Present content in self-contained chunks**
  - **Highlight links between concepts**
  - **Path enables learners to “build their own adventure”**

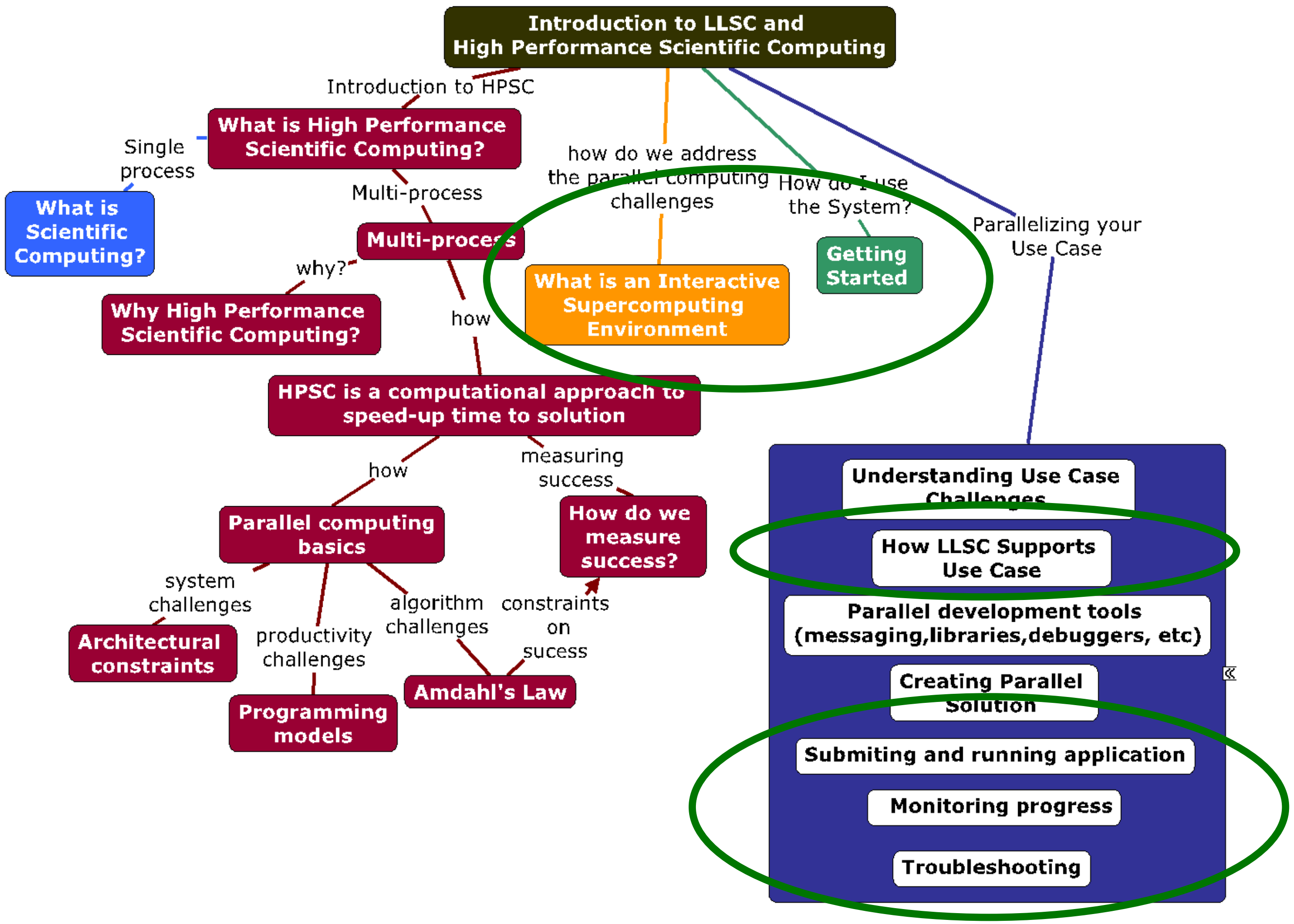
- **Concept Map Design Tool**
  - **Leaf nodes form content units**
  - **Design exposes course components**
    - **Related units form sections**
    - **Related sections form modules**
  - **Links are bi-directional between related concepts**
  - **Concepts with no links are removed**

Initial Results



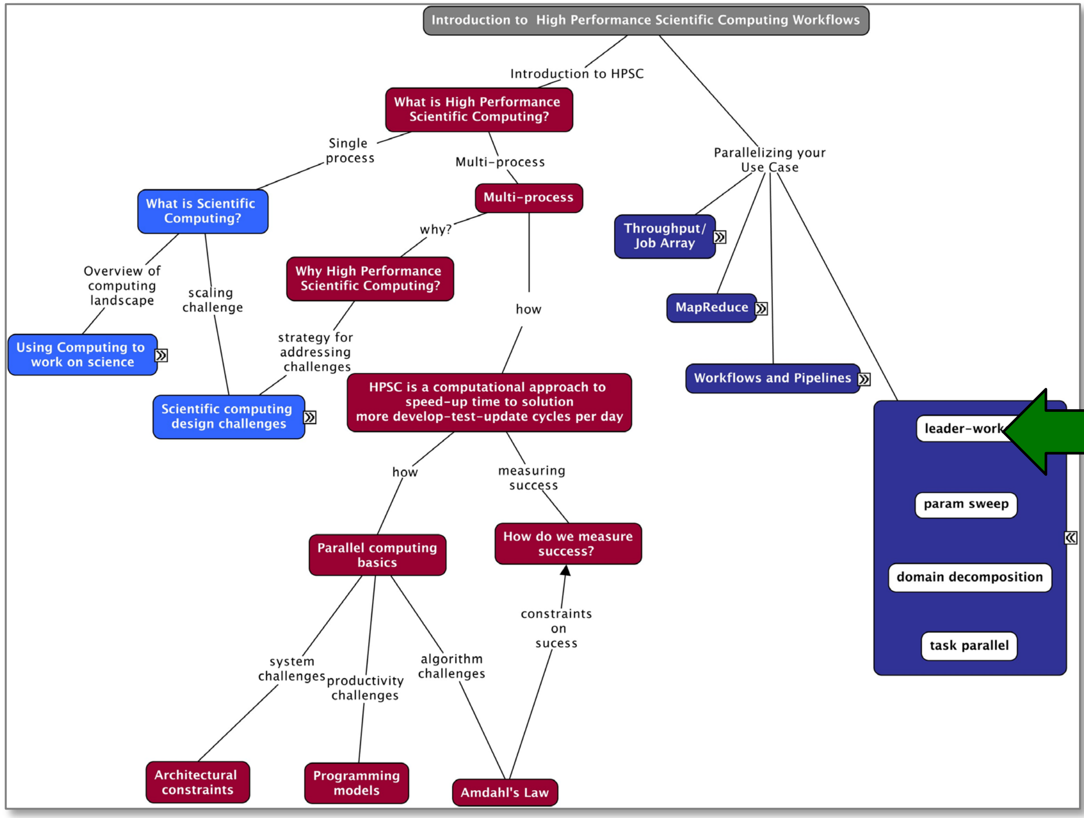
- Design supports learning paths
  - Learners can select content sections relevant to their immediate application
  - Increased likelihood of on the job practice leads to increased retention
  - Supports adult need to match learning to problem or question
- Interleave Theory and Practice
  - Learners gain experience with their application on target system
  - Immediate feedback to assessments minimizes misconceptions

Initial Results

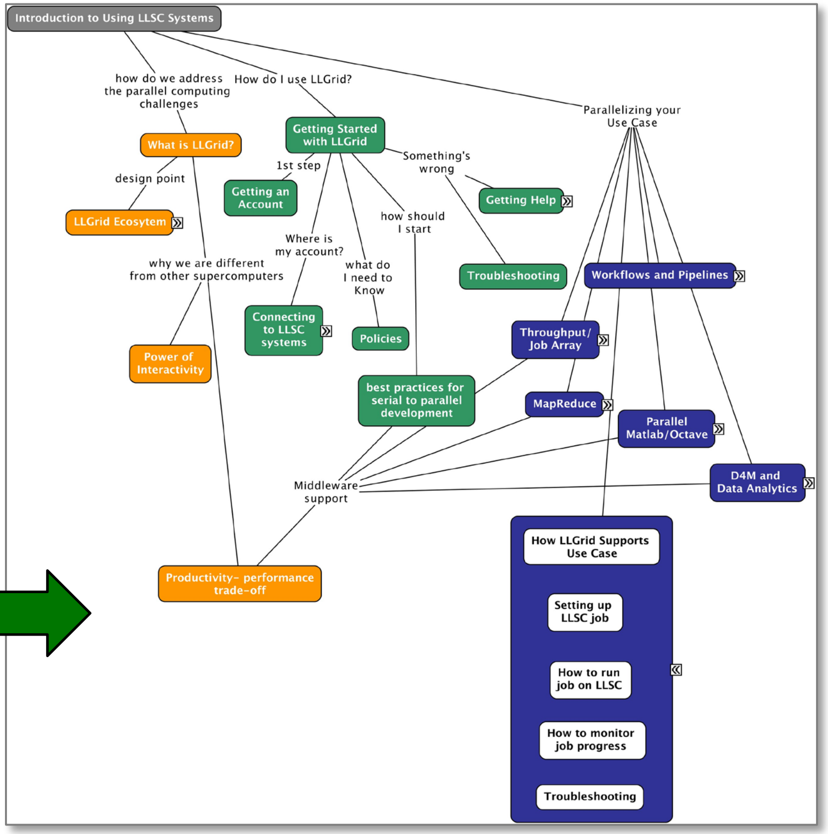


- **Design supports learning paths**
  - **Learners can select content sections relevant to their immediate application**
  - **Increased likelihood of on the job practice leads to increased retention**
  - **Supports adult need to match learning to problem or question**
- **Interleave Theory and Practice**
  - **Learners gain experience with their application on target system**
  - **Immediate feedback to assessments minimizes misconceptions**
- **Not reproducible**
  - **Highlighted areas contain material specific to one supercomputing center**

Scaling Step 2: Redesign as a MOOC and SPOC



- Split course into 2 short courses
  - Introduction to HPC Strategies
  - Using the MIT Supercloud
- Convert programming components to demos and thought exercises
- Maintain hands-on component for center members
- SPOC designed to be easy to edit and adapt as center evolves and changes

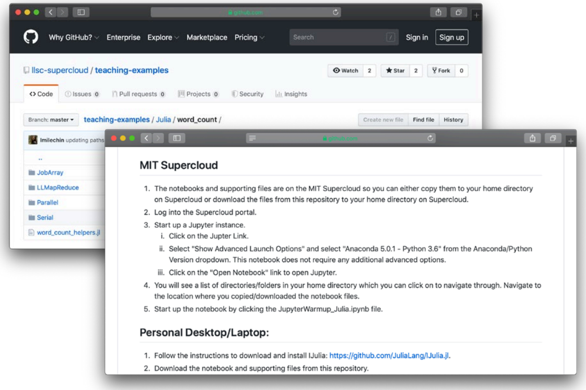


Introduction to High Performance Scientific Workflows (MOOC)

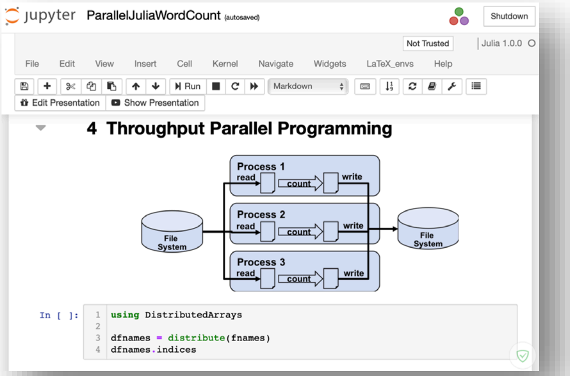
Using the MIT Supercloud System (SPOC)

Hands-On Practice

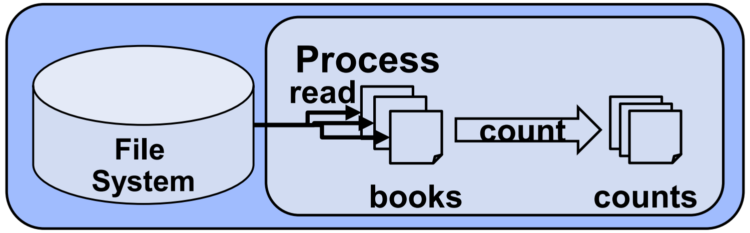
Github Repository



Jupyter Interface



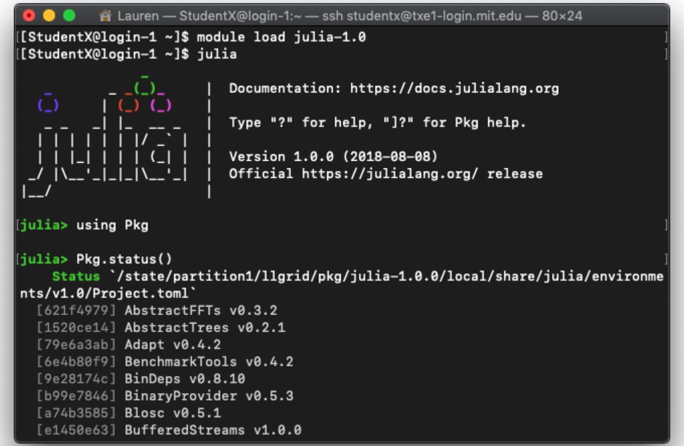
Thought Problems



Questions to Consider:

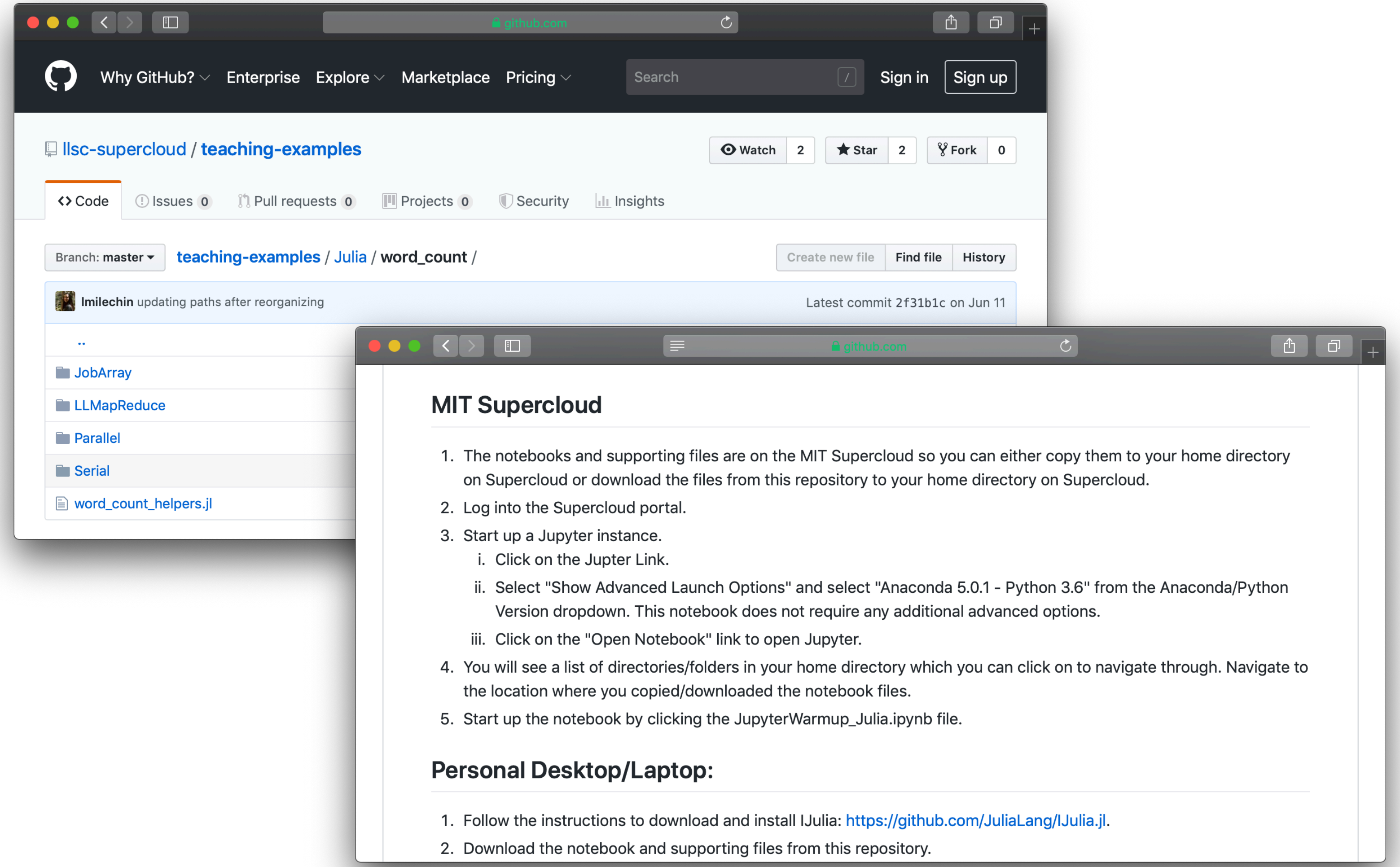
- Where is the independence?
- What data access patterns do you expect?

Demonstrations



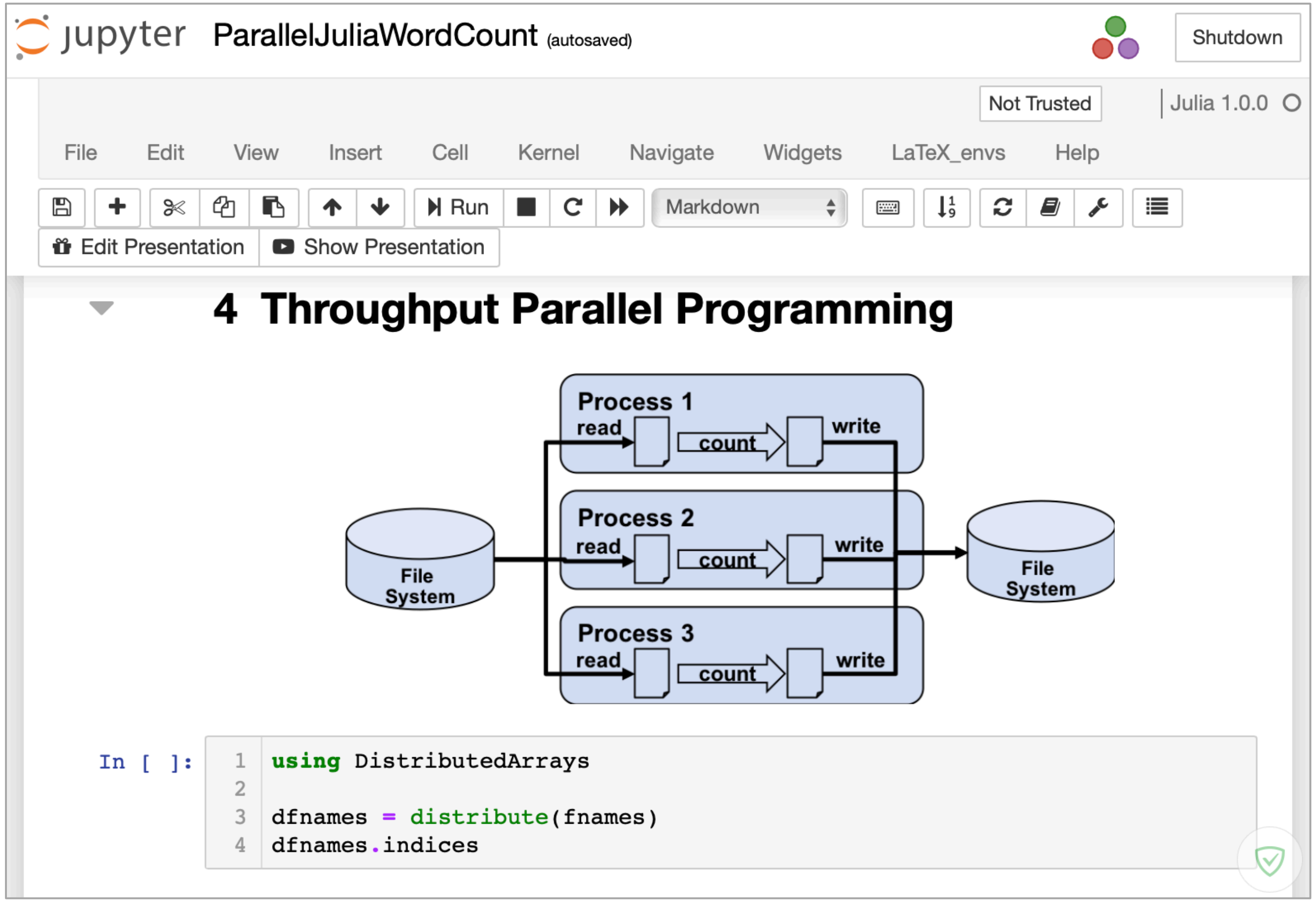
Github Repository of Examples

- **Hands-on examples provided in Github repository**
- **Instructions for setting up and running problems**
  - **Locally**
  - **On Supercloud**
- **Contain or point to sample data where used**



https://github.com/llsc-supercloud/teaching-examples

Incorporating Jupyter



- **In-browser programming environment**
  - **Notebooks contain images, plots, text, executable code blocks**
  - **Text editor with syntax highlighting**
  - **Command line**
- **Support variety of programming languages**
  - **Notebooks: Scripting languages**
  - **Text editor/terminal: Scripting and Compiled languages**
- **Provide easy to use alternative to command line**

Outline

- Challenges in HPC Education and Training
- Introduction to MOOCS
- Case Study: Understanding HPC Workflows and How to Exploit Them
  - The initial one-on-one training
  - The first online course
  - The future refactored courses
- Lessons Learned

Lessons Learned

- **HPC Education and Training can benefit by leveraging MOOCs**
  - Can reach thousands of students
  - Students can
    - Self-select to create their own “Just In Time” experience
    - Learn at their own pace
    - Revisit material for review or deeper understanding
  - Online targeted lectures simplify learning and review for native and non-native speakers
  - Can track student activity to
    - Capture learning interests
    - Content gaps
- **MOOCs aren’t a drop in replacement for existing training materials**
  - Materials need to be refactored with design emphasis on creating stand alone content
  - Diversity of student skills and experience requires inclusion of supporting material
  - Instructor led courses require facilitation
  - Difficult to provide HPC System access to thousands of students for hands-on practice

Acknowledgements

- **MIT Supercloud Team**
- **MIT Lincoln Laboratory Supercomputing Center**
- **Weronika Filinger, EPCC/PRACE**

Backup/Notes