BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

Cyberinfrastructure Update

Patrick J. Clemins
Manager, Cyberinfrastructure and Partnerships, Vermont EPSCoR
Assoc Research Professor, Dept of Computer Science

This material is based upon work supported by the National Science Foundation under Grant No. OIA-1556770.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Your Cyber Team

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

- Manager, Cyberinfrastructure and Partnerships
  - Patrick Clemins
- Software Developers
  - Scott Turnbull
- Technologist
  - Steve Exler
- Collaborators
  - UVM Enterprise Technology Services (ETS)
  - CISL @ NCAR

Our Support Roles

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

- **Data Collection:** Support the collection and management of data from the soil, stream, and lake sensor networks and social systems surveys
- **BREE Domain Scientists:** Provide software development and data management expertise for building and calibrating the component models of the IAM
- **The IAM Effort:** Build the connectivity between the component models and the necessary cyberinfrastructure to effectively run the IAM
- **Collaboration:** Develop and maintain video conferencing capabilities, data management and sharing policies, and our Macroinvertebrates mobile app

Core CI Resources - Hardware

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

- Compute
  - Babbage – Dell PowerEdge R820 compute server
    - 32/64 CPU cores, 256 GB RAM
  - Pascal – NVIDIA DXG-1 GPU compute server
    - 40/80 CPU cores, 512 GB RAM, 28672 NVIDIA CUDA cores (8x Telsa GP100)
  - Cheyenne / XSEDE (i.e. Xstream)
    - National high performance computing (HPC) resources
- Data Management
  - Aquarius – Aqaurius hydrology database server
  - Aquariusws (Leopold) – LoggerNet Admin and Aquarius workstation
- Data Storage
  - epscorfs – 100TB network storage

Past Year's Progress

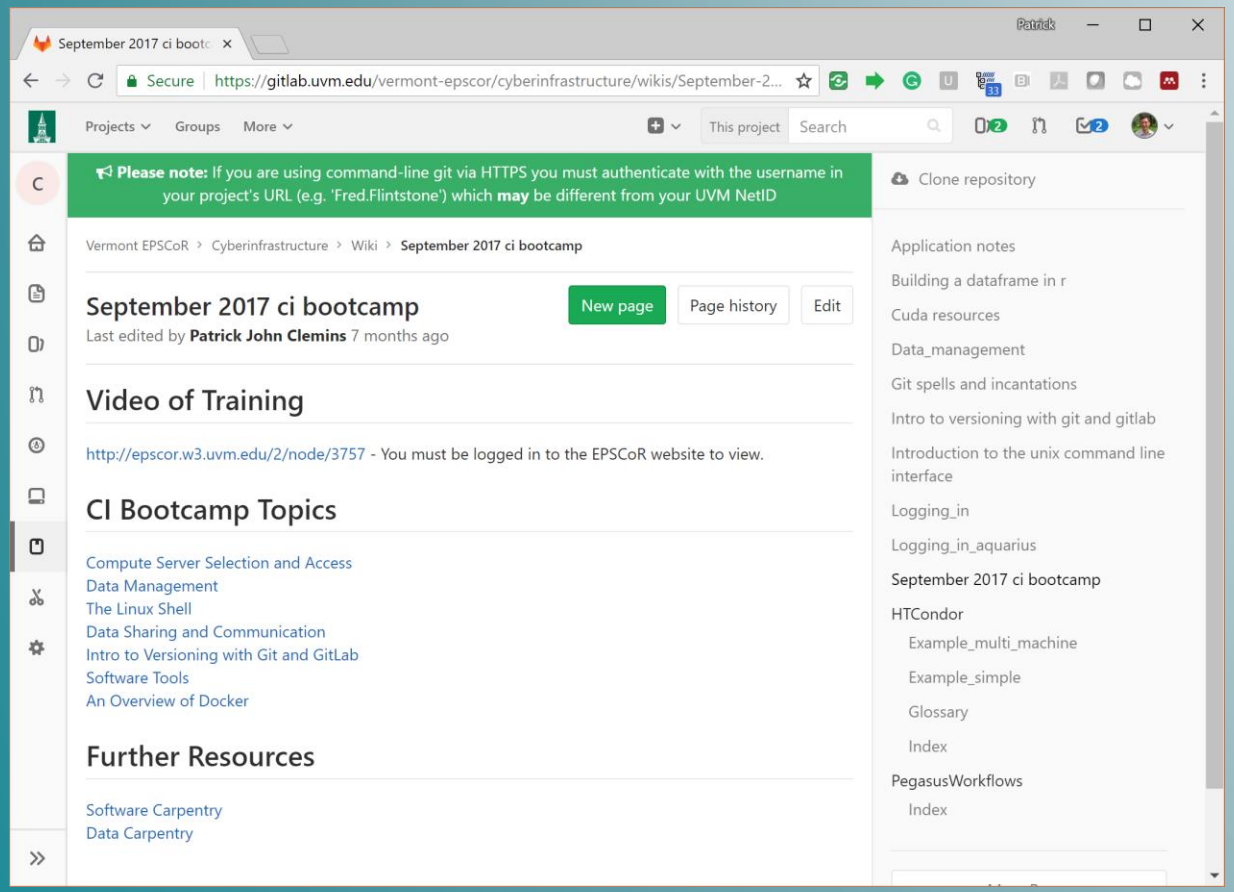
BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

- Training the New Team
  - Held Cyberinfrastructure Bootcamp
  - Deep learning class teams used Pascal for homework and projects
  - Individual and small group training for remote Linux, scripting, parallelism, etc.
- IAM Advancements
  - Transitioned IAM from Yellowstone to Cheyenne (with fallback to Babbage and Pascal)
  - Created ucar_client_worker and ucar_access_server to manage Cheyenne tasks
  - Added support in IAM for Daymet climate data
  - Fully implemented the IAM in Python 3
  - Started to build an IAM testing suite
- Component Models
  - Built a framework to calibrate the ALL ABM
  - Implemented a fix for EFDC scaling issue on Cheyenne
  - Determined source of RCA restart bug

Training

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

- CI Bootcamp – Sept. 28, 2017



- Deep Learning on Pascal

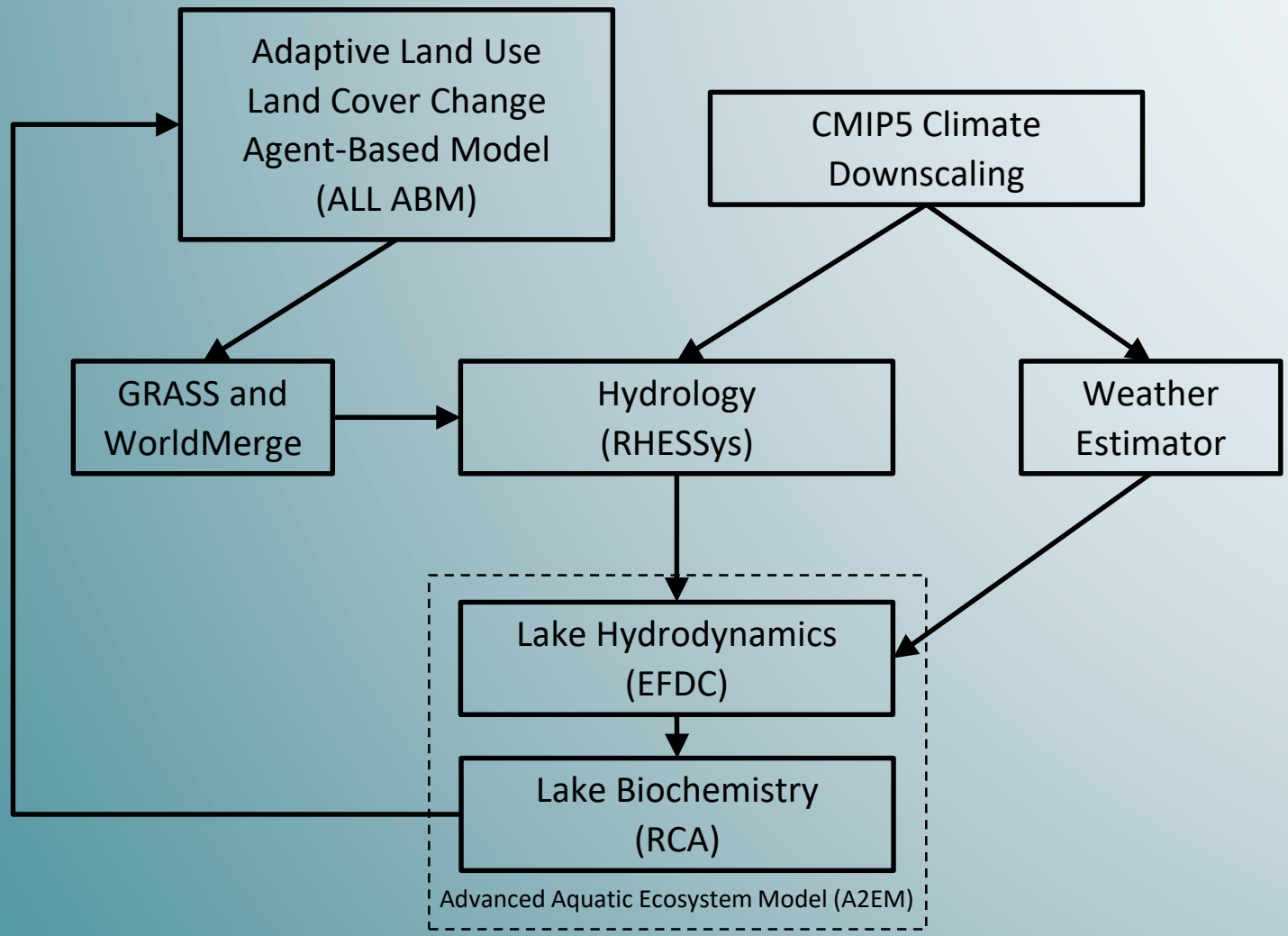
TensorFlow

Keras

docker

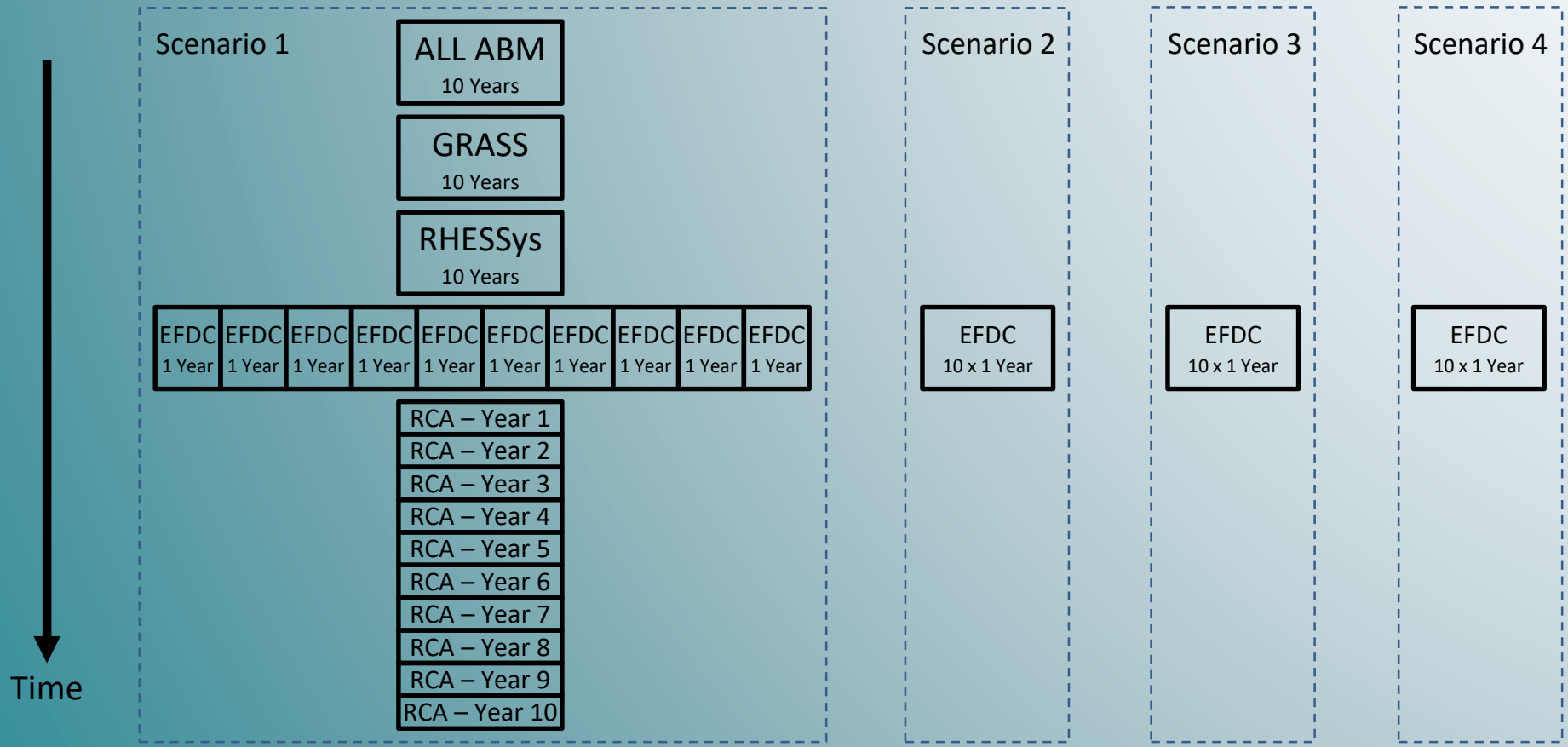
The IAM

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin



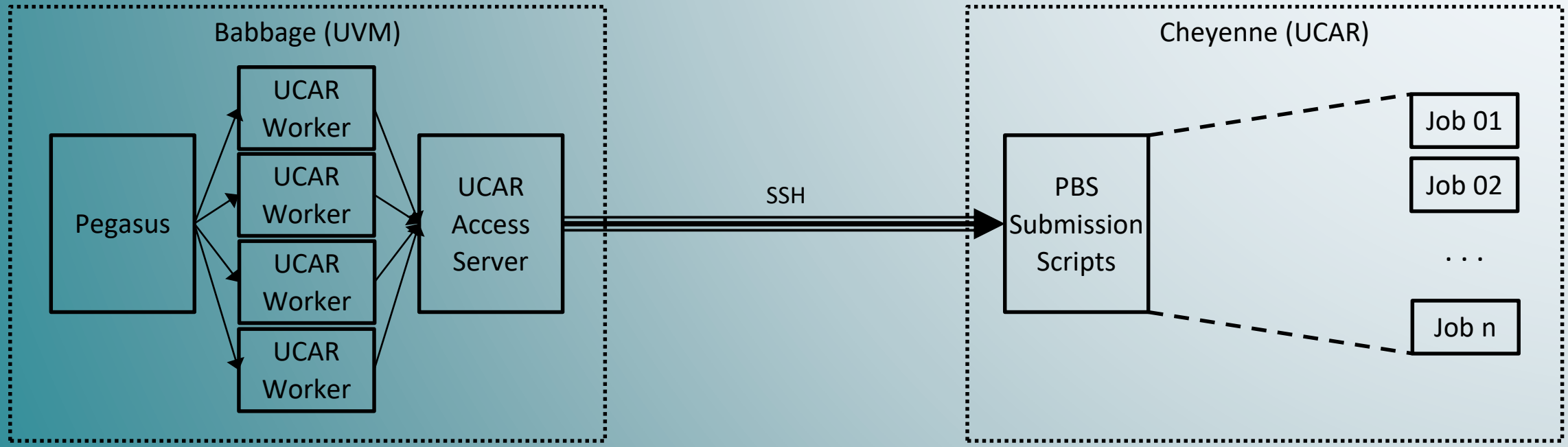
Computation of the IAM

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin



Using Cheyenne

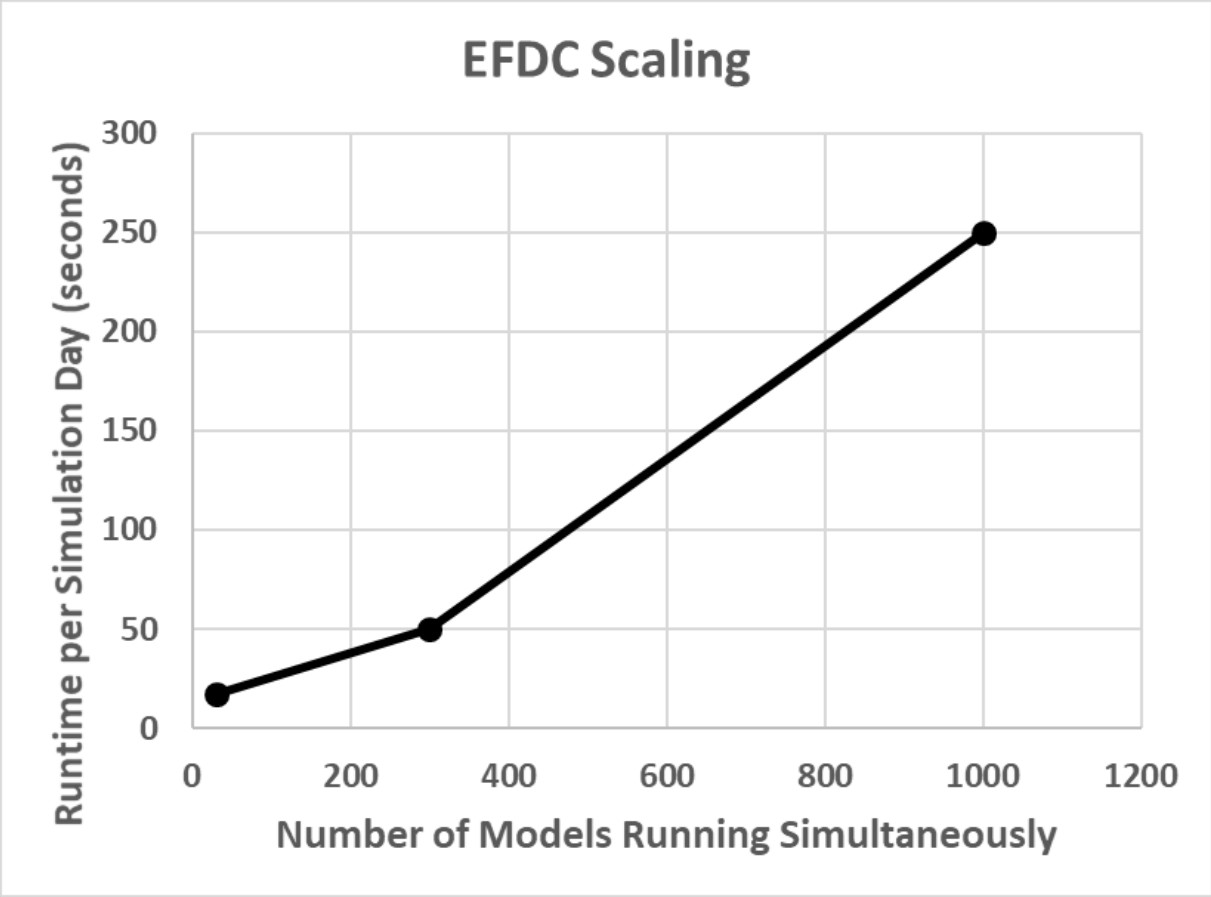
BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin



EFDC Breaks Cheyenne

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

- With the UCAR interface designed to maximize the use of Cheyenne, it became quickly apparent that the more EFDC jobs we sent to Cheyenne, the slower they got.



- Problem
  - EFDC is causing too many disk requests to the same physical platter on GLADE, the 52.7 PB disk array.
- Solution
  - Copy the EFDC code and data to the 32GB /dev/shm (shared RAM) space on each node.

On the Horizon

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

- Aquarius Updates
  - Upgraded Aquarius server and workstation (Leopold)
  - Adding soil sensor network to Aquarius
- Hardware Maintenance
  - Fixing Babbage's broken memory
  - Continue to upgrade Pascal with service updates
- IAM
  - Finish IAM test suite (esp. integration tests)
  - Re-architect ucar_client_worker to dynamically package jobs
  - Implement finer control over dates of IAM execution
- IAM Component Models
  - Deltares is being configured on Cheyenne
  - GEAM is ready to be integrated
  - Moving ALL ABM from AnyLogic to MASON

BREE
Basin Resilience to Extreme Events
in the Lake Champlain Basin

Discussion

IAM / Cyberinfrastructure