ICE 2.0: Restructuring and Growing an Instructional HPC Cluster

Nov 12, 2023
HPC SysPros Workshop – SC23
Denver, CO

Eric Coulter, PhD

PACE Research Computing Facilitation Team Lead

Georgia Tech

ICE 2.0: Restructuring and Growing an Instructional HPC Cluster

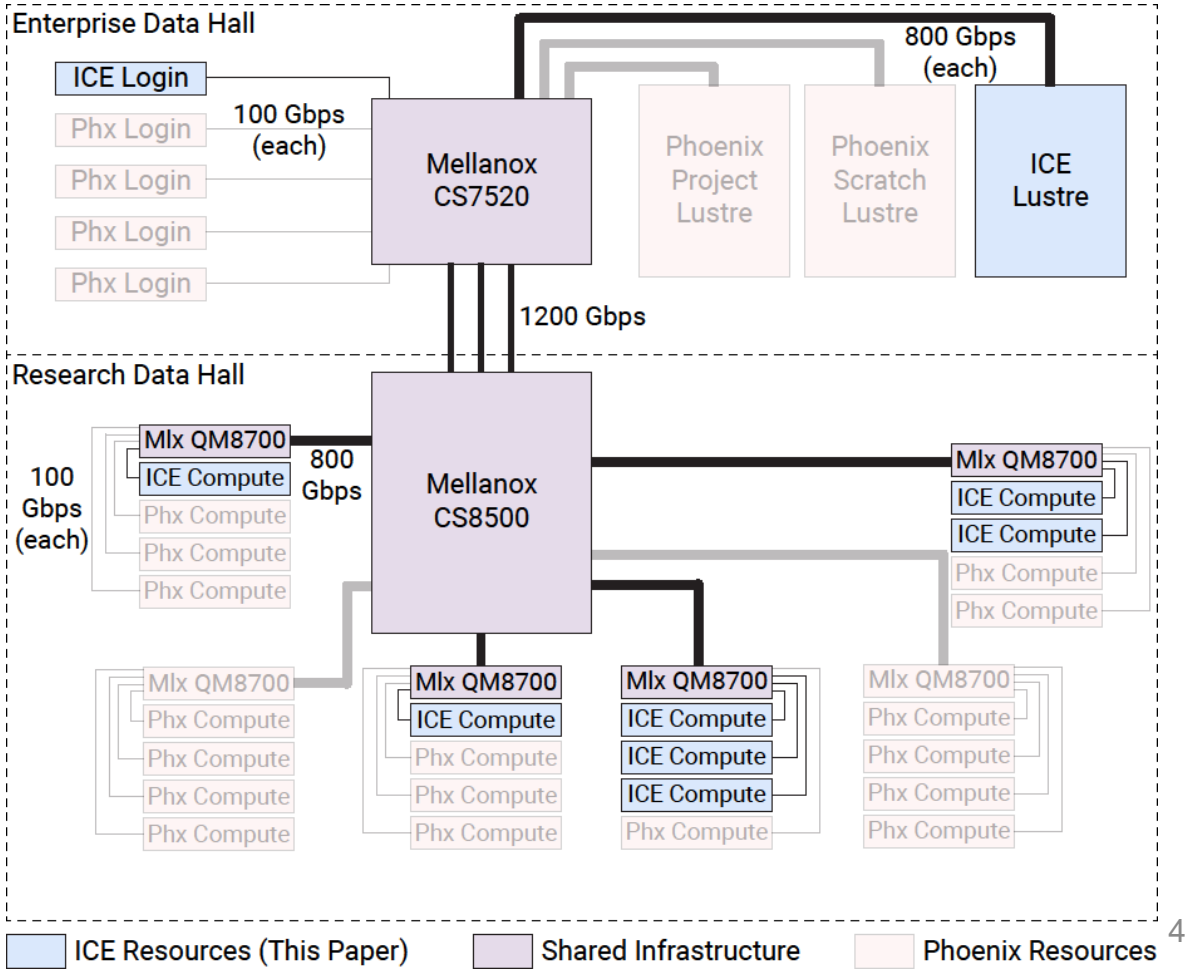
Credit to the whole team:

J. Eric Coulter, Michael D. Weiner, Aaron Jezghani
Matt Guidry, Ruben Lara, Fang (Cherry Liu)
Alan Metts, Ronald Rahaman, Kenneth J. Suda
Peter N. Wan, Gregory Willcox, Deirdre Womack
Dan (Ann) Zhou

Georgia Tech

What is ICE? (Instructional Cluster Environment)

- Yes we do talk about "the ICE Cluster" and I die a little bit every time
- HPC environment purely for supporting courses and student projects at Georgia Tech
- Mix of CPU and GPU hardware, collaboration between PACE and the College of Computing
- Available to any course at GT, priority access to some hardware based on funding source



Georgia Tech

Goals of this Project (Problems to solve)

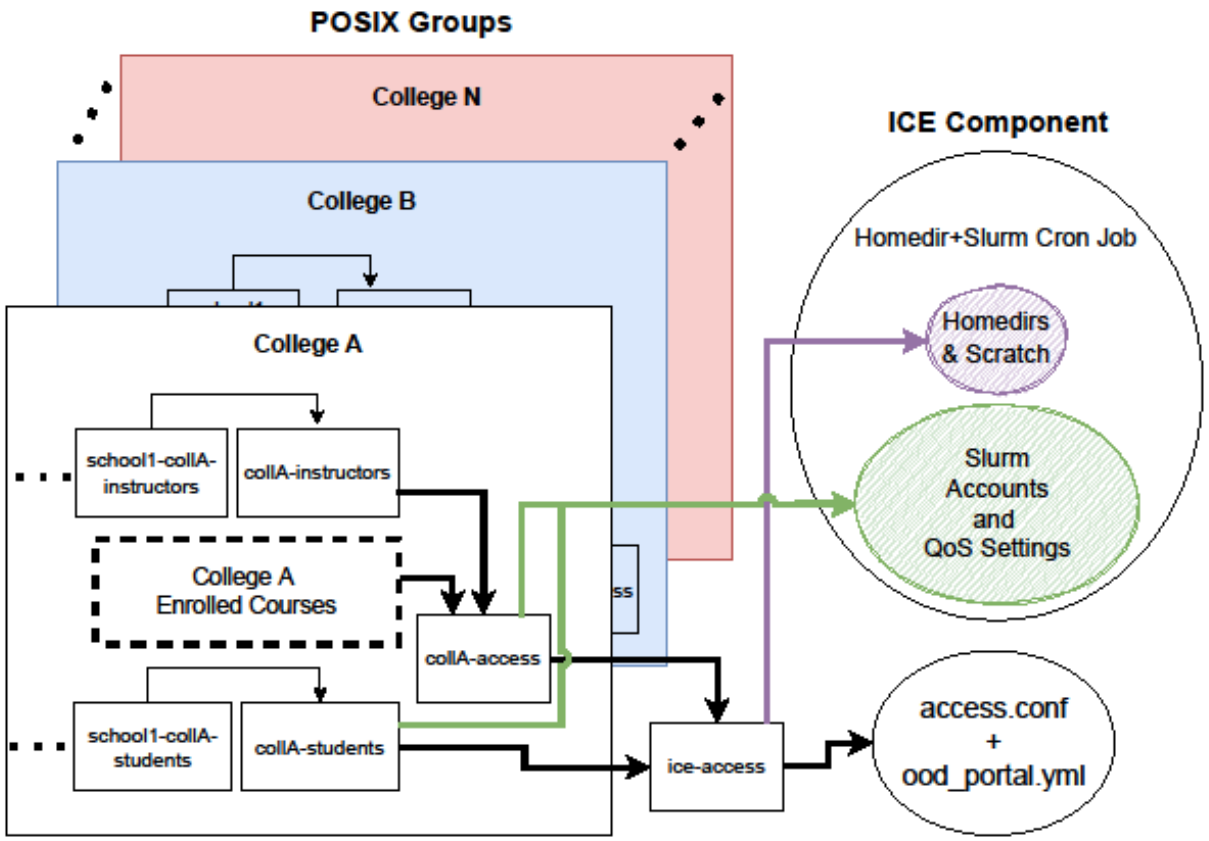
- In the beginning...
  - There were TWO ICEs (one for PACE, one for College of Computing (CoC)
- Merge the ICEs
  - Including user account data
  - Course data
  - Accounts
    - Naturally this is a perfect time to also figure out which accounts need to be pruned
- Migrate to Slurm (Previously on Torque)

How ICE Works

- Instructors request access
- We add the course to a special group, which we call "entitlements" to an internal roles system
- Everyone enrolled in that course now gains access (and loses it accordingly)
- This extends to special workshops, PACE-led training, etc.
- In the process of extending to departmental membership

Accounting Design

- Cron jobs regularly check membership in those entitlements
- Several things happen as a result:
  - Home and scratch directories are created
  - Scheduler accounts are created
    - At an appropriate QoS level
    - Includes departmental affiliation for metrics
  - Login access is granted
  - Lots of error checking



Georgia Tech

Entitlements Design

- Top level “access entitlement”
- College-level entitlements contain:
- Separate Student/Instructor entitlements
  - Separated by School at each of those levels
  - Student entitlements are populated by Course Enrollment rules

- PACE exists at the “College” level for workshops and training sessions
- Easily extensible as new Colleges or Schools participate in ICE

Data Migration+Merger

- Back to the TWO ICES – some 20K accounts to migrate and merge
  - Total of 7TB for homedir data, an additional ~1TB for course data
- Migrated to /newhome/{oldICE_1,oldICE_2} directories to preserve data
  - Since some accounts existed on both ICEs
- Parallel rsync (being very careful about sparse files! –S on initial sync, -W on final) took about 5 days
  - Scripts included in artifact

Storage Design

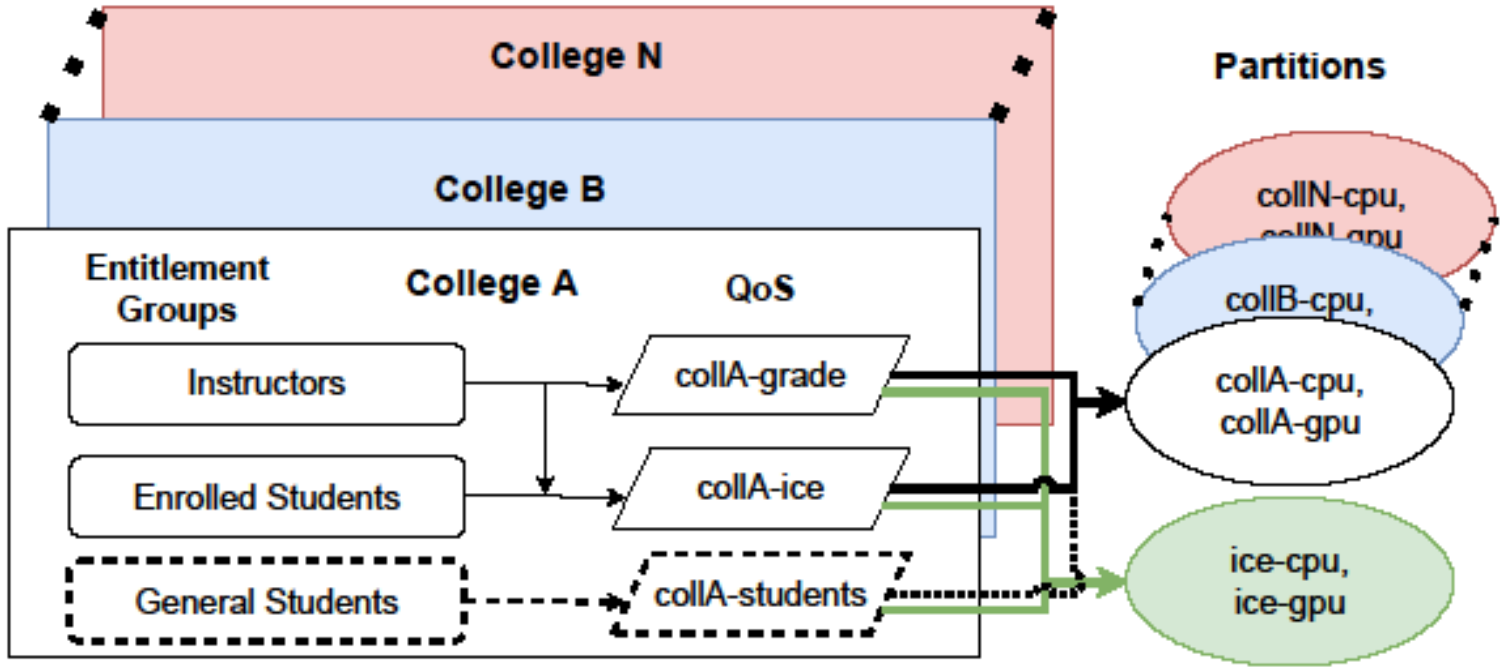
- Home and Scratch space, with separate spaces for course data
- Roughly 16K extant homedirs, planning for all 40K students eventually having access
- Able to trim down to ~14K based on current GT affiliation (long overdue…)
- Homedirs and scratch bucketed by last 2 digits of UID:
  - /home/${uid: -2:1}/${uid: -1}/${username}
  - /scratch/${uid: -2:1}/${uid: -1}/${username}
- Mounted via AutoFS, which has to override homedir provided by SSSD via dynamic map (which was *fun* to discover/set up!)
  - Config included in artifact

Data Policies

- 15GB of data in /home (NetApp)
- 100GB in /scratch (Lustre)
- No quota on shared course data directories (backed by NetApp)
- Semesterly cleanup
  - Any files not touched in 120 days at the end of the semester are removed from /scratch dirs.
  - Any home directories without active entitlement OR login activity for 1 year are removed

Scheduler Design

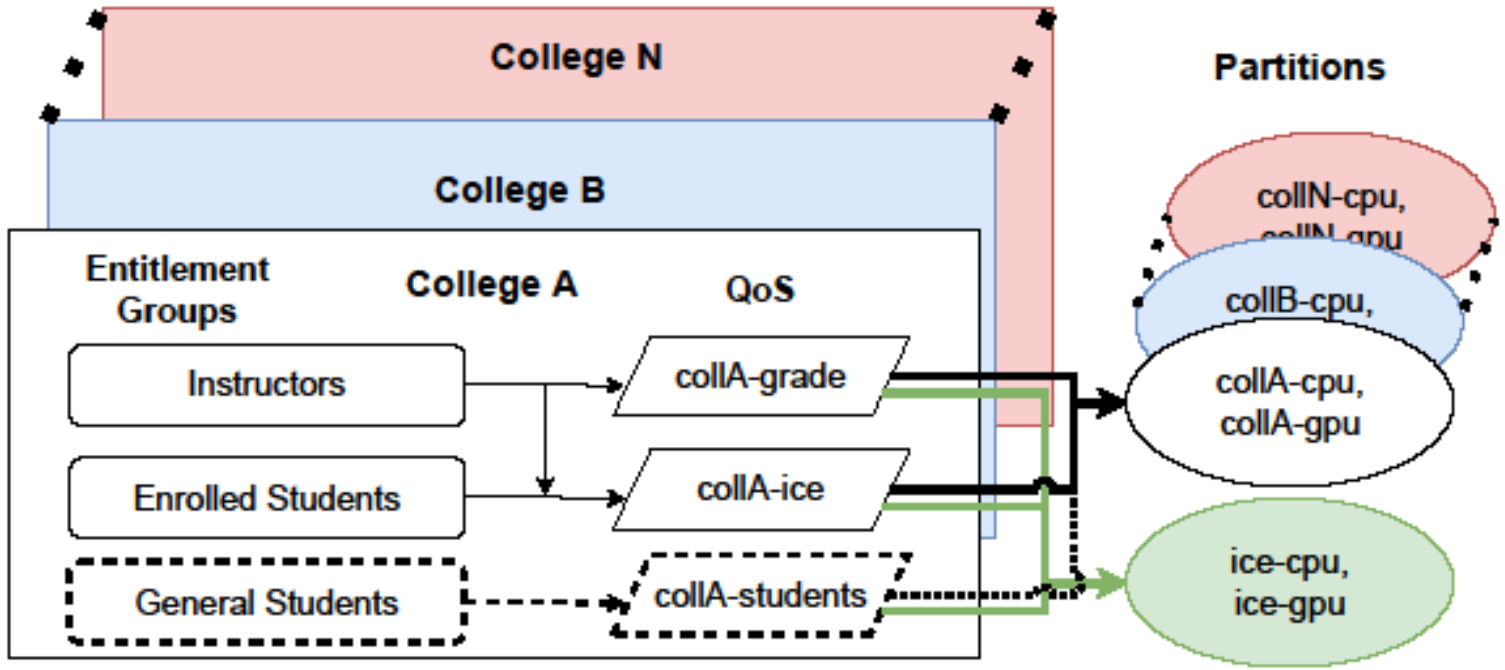
- Allowed QoS determined by entitlement memberships
- People who buy hardware have the option of priority access
- Instructors have priority access for grading needs (repro student submissions)



Georgia Tech

Scheduler Design

- Job routing by job_submit.lua based on QoS and resource needs
- Simplify necessary Slurm scripts as much as possible for student use with sensible defaults
- Configs available in our artifact!



Georgia Tech

Questions?

- **Scripts and config examples available in GitHub: https://github.com/pace-gt/hpcsyspros-SC23-ICE**
- **Previous ICE Paper: (Includes some more details around accounting, policy, and design):** https://dl.acm.org/doi/pdf/10.1145/3219104.3219112
- **Contact us: j.eric@gatech.edu**