Computation Institute

US ATLAS Facility Projects

ATLAS Connect, FaxBox, T3 flocking

Rob Gardner
Computation and Enrico Fermi Institutes
University of Chicago

US ATLAS Workshop on Distributed Computing
University of Arizona
Tucson, December 11-12, 2013

THE UNIVERSITY OF CHICAGO

efi.uchicago.edu
ci.uchicago.edu

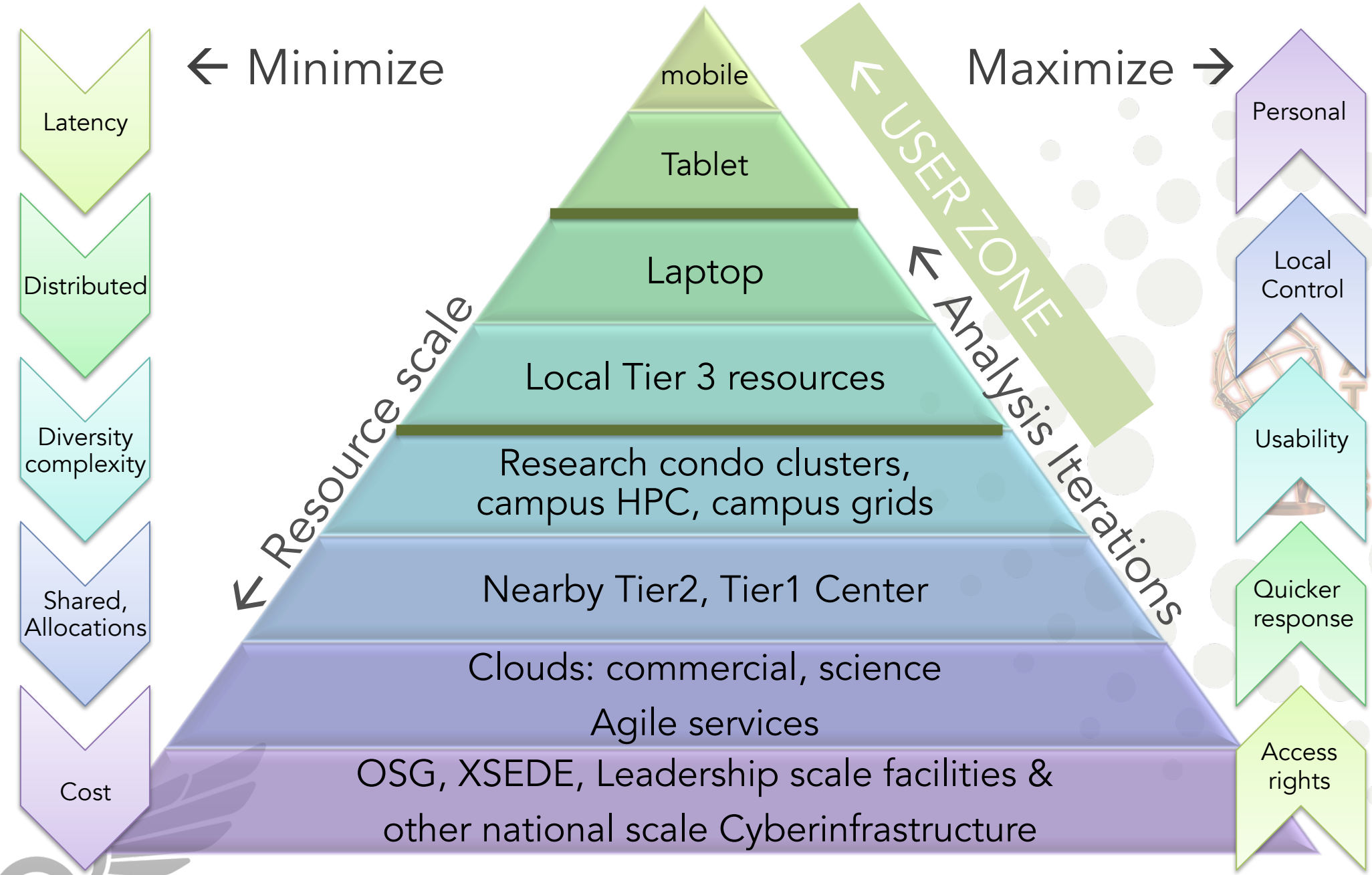
Background and Motivation

- Try out new approaches for delivering analysis resources for end-users in Run2, given limited and aging Tier3
  - Logically extending local Tier3 resources in some cases
  - Reaching institutional resources that might be locally available
    - E.g. a shared condo cluster in the research computing center
  - Complement existing Panda ANALY queues
- Informed by Tier 3 implementation committee (more later this AM)
- Leverage related work
  - In OSG (campus grids, BOSCO)
  - U of C (sustainable computing-as-service)
    - Globus Online, CI Connect
  - And FAX (i.e. WAN access)

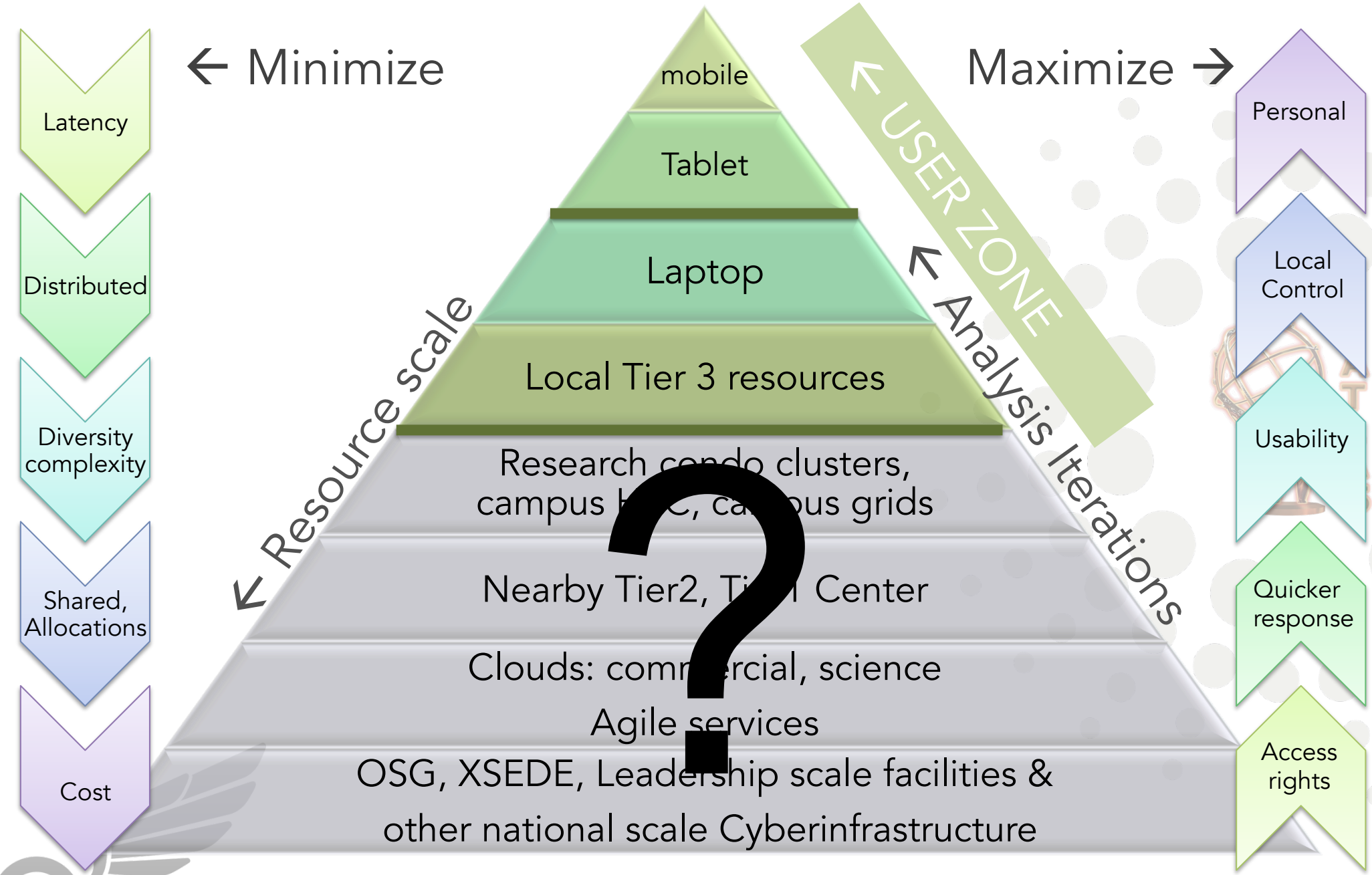
The Local Analysis Batch

- Local Tier3 resources are precious
  - They cannot be replaced by the grid, at least not now
  - Especially with regard to highly iterative processing, interactive processing with fast turn-around
  - **Locally controlled**, **low network latency** resources proven the best match for end-user analysis
- However, analysis batch job processing closely follows the classic high throughput computing (HTC) model
  - Including distributed HTC over a wide area network
- Therefore distributed resources can be considered to meet some of the of the Tier3 use cases

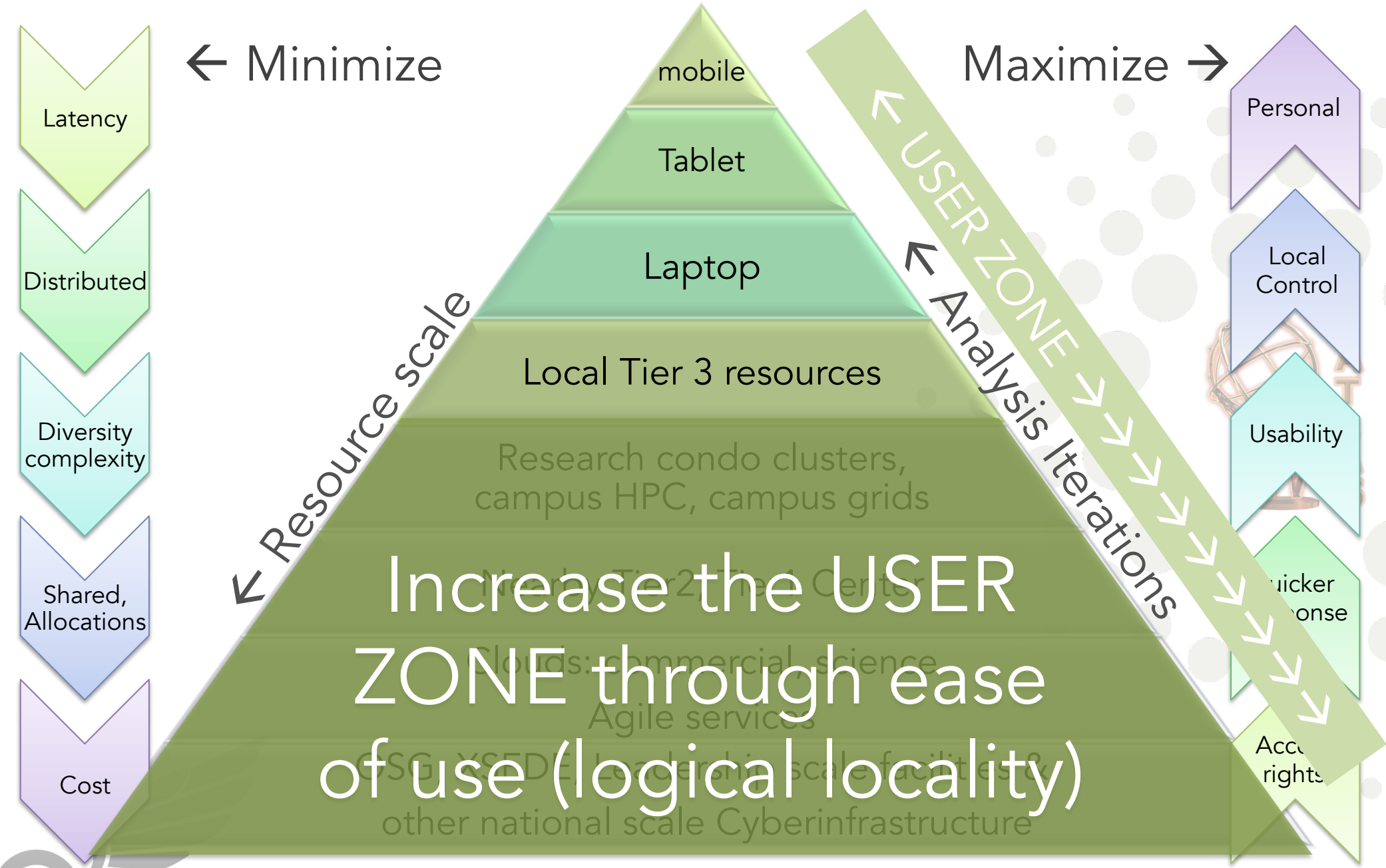
Characteristics of Resources



How best to reach them?



Increase the reach, but keep things "local" where possible



Practical approaches

- Match impedance of local analysis use cases
  - Build on robust, well-supported technologies
- Match impedance of resource providers
  - Choose common, widely adopted technologies, leading industry technologies
- Provide as a service
  - Bring new users, groups, resource targets up simply
  - Roll out upgrades across the board

“Connect” technology

- Campus Grids area in OSG
  - Problem of getting new research communities running on OSG
  - Small groups, many times single individual
  - Developed OSG Connect (http://osgconnect.net) to provide a “retail” service for OSG – rolled out at Duke Workshop in August
- At UC, we built the UC3 campus grid to provide UC Tier3 users with more campus resources
  - And learned why there are so few campus grids in the process
  - Thus developed a new technology CI Connect to make this simple for campuses and collaborations
  - Deployed Duke campus grid: http://duke.ci-connect.net/
    - Bridged with UC campus grid, and OSG
- Then realized we could make a connect service for US ATLAS based on CI Connect:
  - http://connect.usatlas.org/

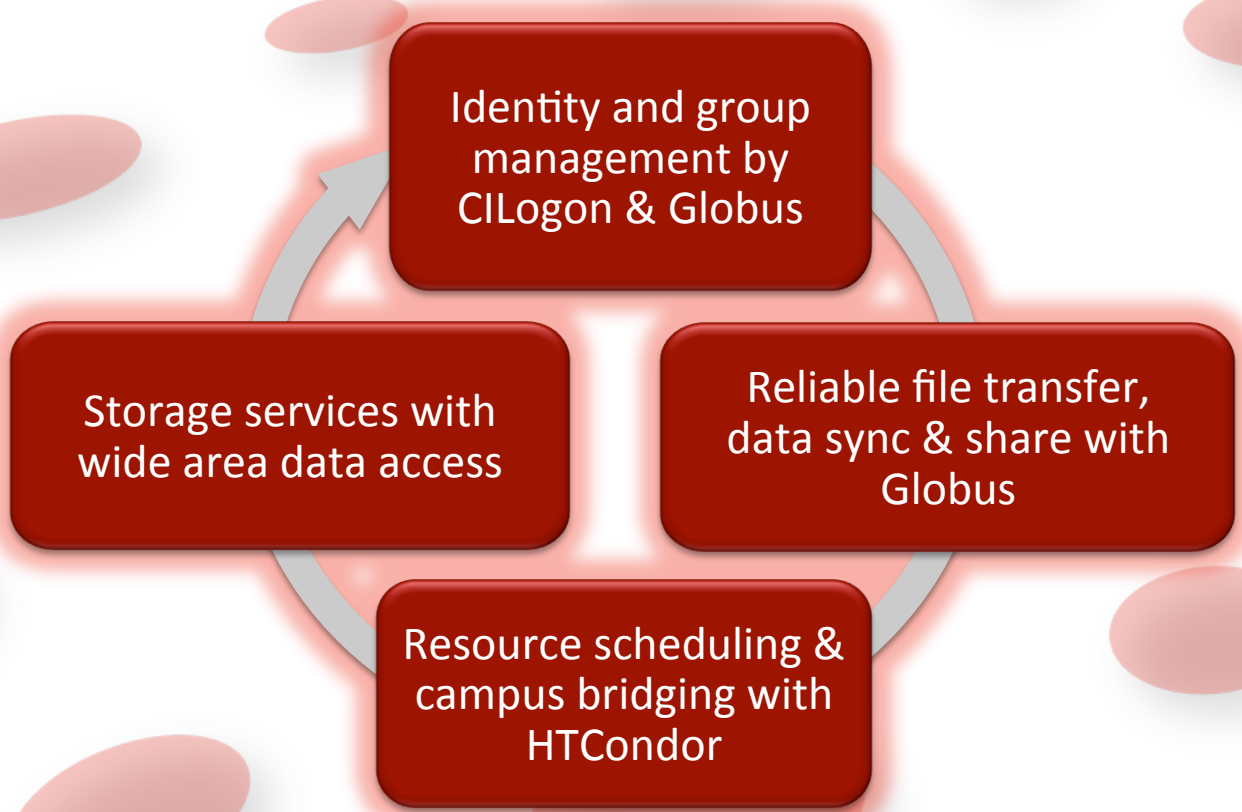
CI Connect

Accelerate research on campus by providing connective services for local, cloud and national cyberinfrastructure

http://ci-connect.net/

ci connect

THE UNIVERSITY OF CHICAGO



osg connect

globus

HTCondor High Throughput Computing

Component choices

- Identity management (campus → anywhere) [CILogon + InCommon + X509 (grid)] which is implemented by Globus
- Connecting CPU resources [HTCondor, BOSCO, cloud provisioning services]
- Software distribution [CERN Virtual Machine File System – CVMFS (OSG OASIS)]
- Distributed software access [Parrot + Chirp]
- Distributed data access [Globus Online, XRootD, domain-specific tools & services, Parrot + Chirp]

Technologies & Infrastructure

HTCondor
High Throughput Computing

CILogon

XRootD

UC3

globus

ATLAS

Open Science Grid

ceph

CYCLECOMPUTING

puppet labs

CONNECT user

connect.usatlas.org

FaxBox

portal

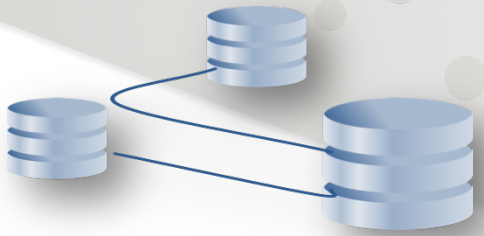
login.usatlas.org

PanDA

Tier1, Tier2, Tier3

Amazon cloud

Campus Grids



globus ci connect

Current ATLAS Connect deployment

- Fresno State, MWT2, and UC3. AGLT2 soon.
- Configurable number of slots

CycleServer | Usage | Jobs

Welcome

Pool Summary

Pool	Total Slots	Running	Idle	Owner	Status	Detailed View	
CSU Fresno Factory	0	0	0	0		Usage	Jobs
Midwest Tier 2 Factory	0	0	0	0		Usage	Jobs
University of Chicago Computing Cooperative	528	0	528	0		Usage	Jobs
Total	528	0	528	0		Usage	Jobs

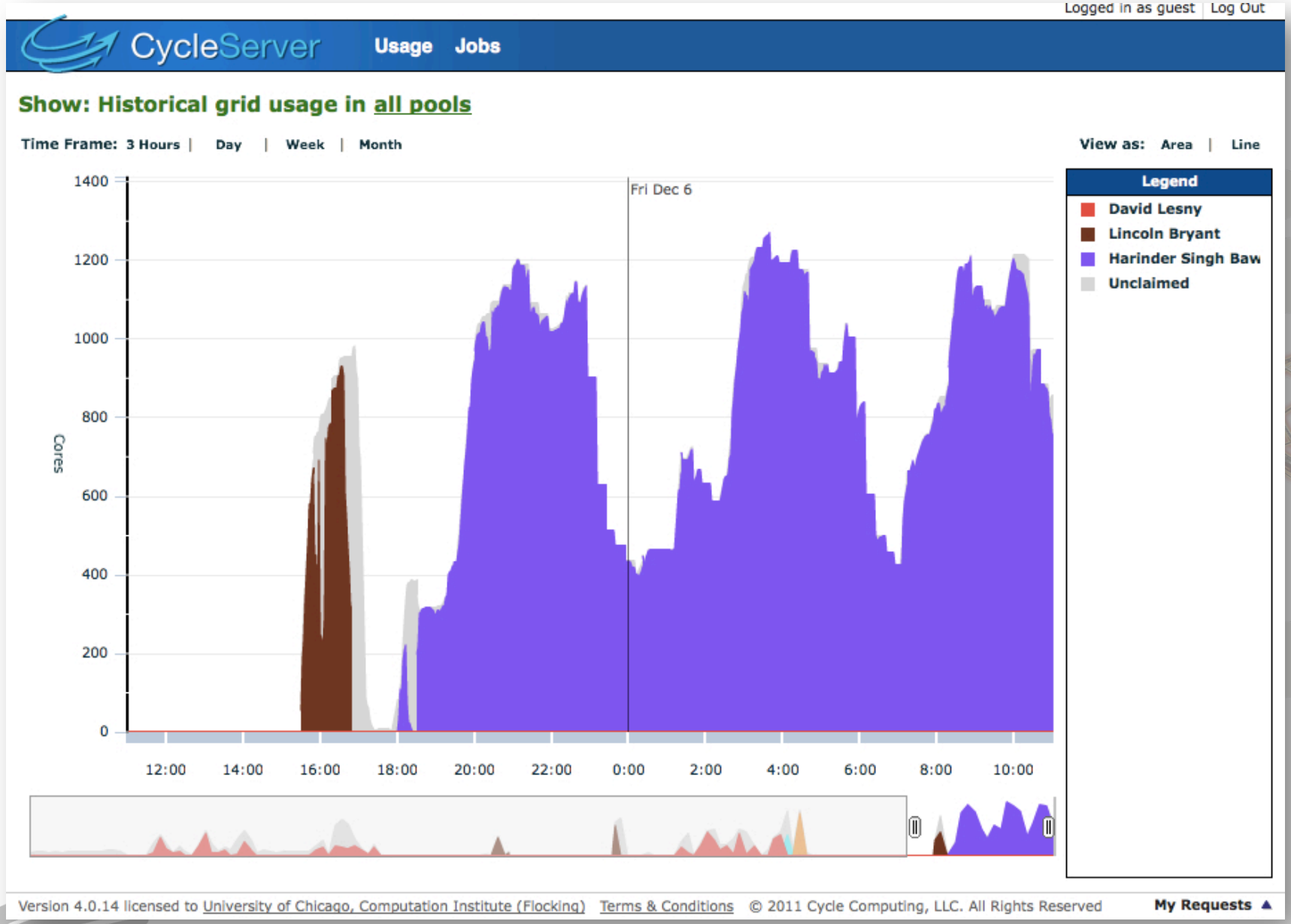
Jobs by State

Jobs by Owner

Slots by State

Slots by Owner

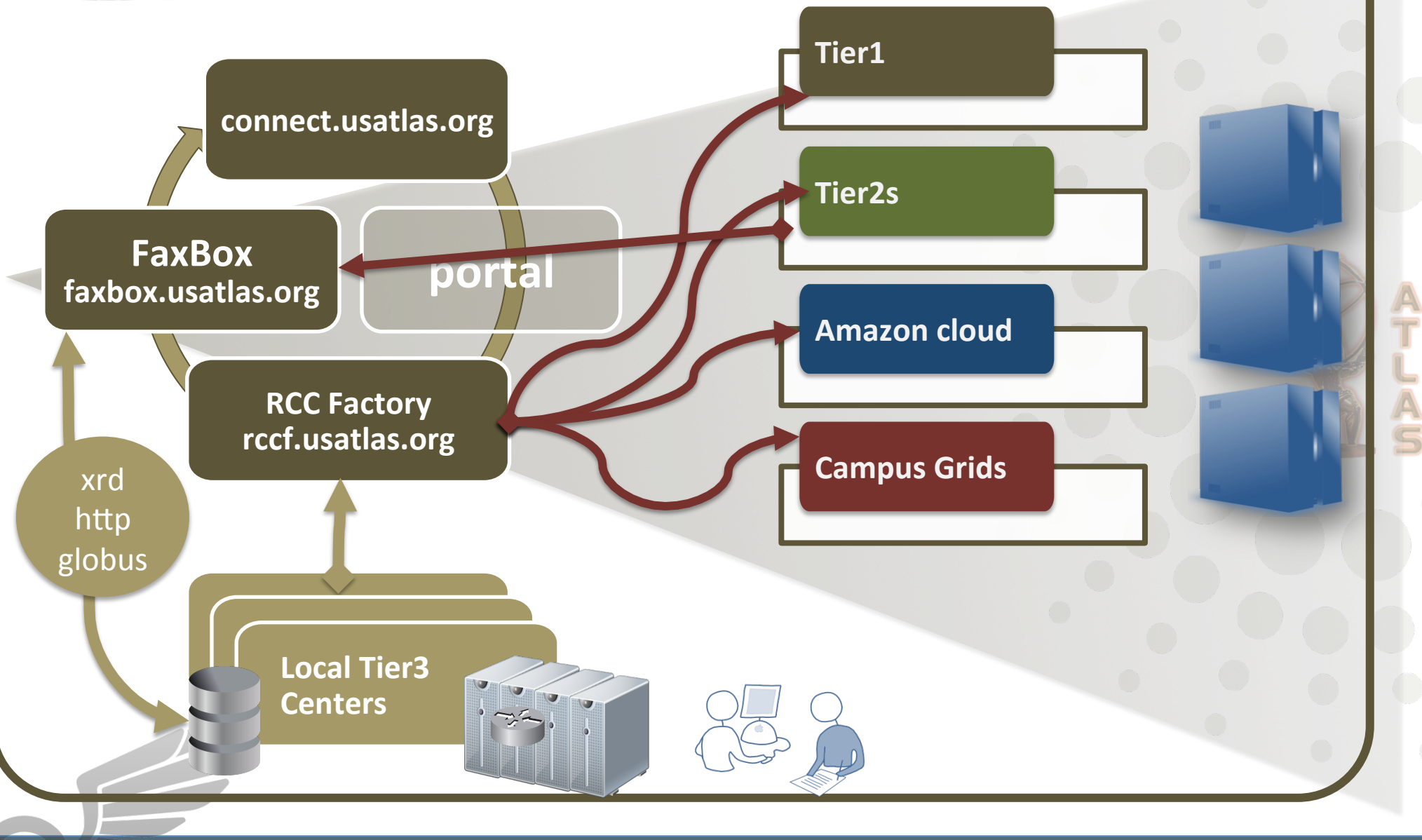
Testing (More in Harinder's talk)



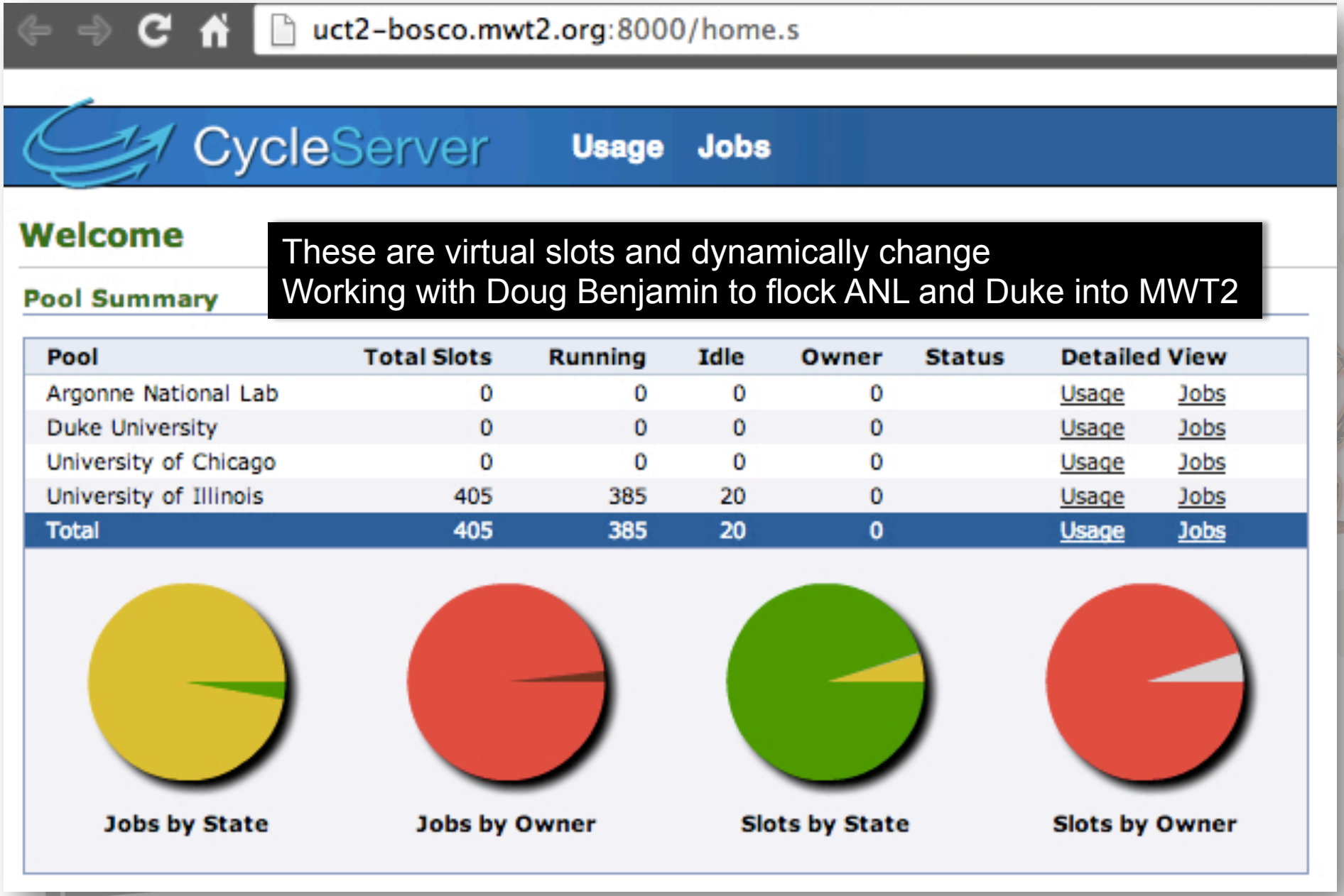
ATLAS Connect Cluster: Flocking Factories

- Logically extend or connect clusters with Condor flocking
- Experience:
  - UIUC site integrated with MWT2 with flocking
    - Now IU, UC, UIUC each flock to the other
  - UChicago Tier 3 experience
    - Large Tier 3 community but medium sized Tier 3 (300 job slots)
    - UCT3 now flocks transparently to the UC campus grid
    - Users can get > 1000 job slots
- ATLAS Connect Cluster extends this for Tier3s managed with Condor

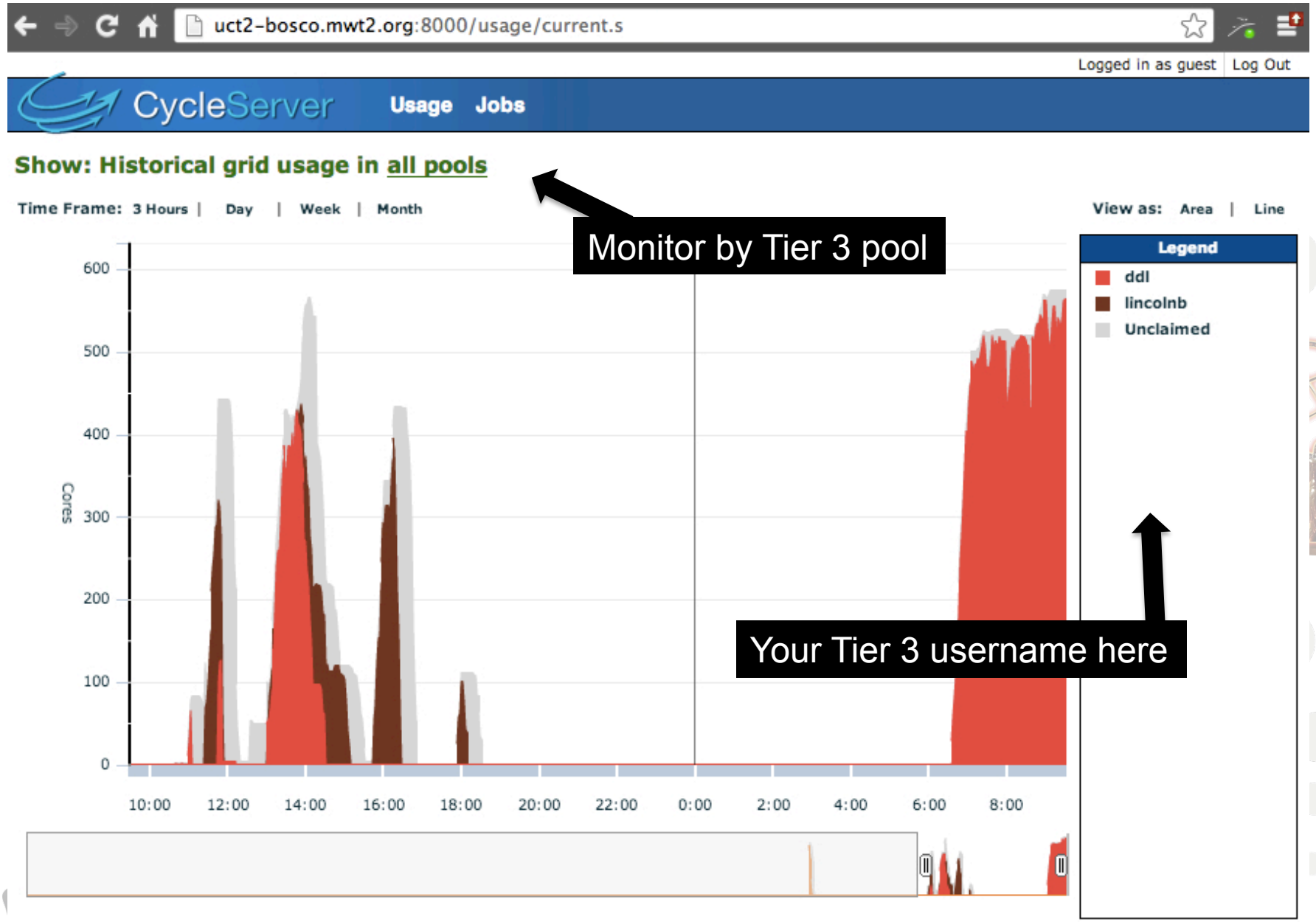
CONNECT cluster



Monitor Tier 3s flocked into MWT2



Where your flock jobs appear



USER Storage close to CPU: **FaxBox**

- To assist with ATLAS Connect User and flocked jobs via ATLAS Connect Cluster
  - Pre-stage data, write outputs for later use, etc.
- Use standard Xrootd tools and protocol
  - root://faxbox.usatlas.org/user/rwg/file
  - Therefore read from anywhere, even a prun job
  - Will include a user quota system, and monitoring tools
- Not a DDM/Rucio endpoint, for now
  - Therefore simple & easy use and manage
- Will provide Globus Online and http access
- See Lightning talk from Lincoln Bryant

Mode shift for users

- Users should not expect their home directories, NFS shares, or to even run jobs as their own user.
- Instead, HTCondor transfer mechanisms, FAX for data access, CVMFS for software
- Make use of LOCALGROUPDISK
- Smaller outputs (on the order of 1GB) can be handled by Condor's internal mechanisms
- For things not in official ATLAS CVMFS, commonly used tools, etc., use MWT2's CVMFS repo
- UC T3 users have made this shift and are happy
- Need to develop best practices and examples
  - Collect at http://connect.usatlas.org/ handbook

Next Steps

- Collect input!
  - Trial users – compare with pure local
  - T3 implementation committee
  - Facility input: enlarging T2 participation (cycles, distributed FaxBox possibly)
- ATLAS Connect User
  - Continue testing in beta service
  - Open to early adopter users who contribute an example of their code to the 'handbook'
  - Aim to have a production service by April (next US ATLAS meeting at SLAC)
  - K.I.S.S.
- ATLAS Connect Cluster
  - Factory submission to AGLT2 (in progress)
  - Distributed HTC batch analysis examples documented
    - E.g. modes: Jobs.. "Stay at home", "Run remote", "Overflow", "Retry"

Computation Institute

Thanks to MWT2 admin staff

Lincoln Bryant (UC)
Dave Lesny (UIUC)
David Champion (UC)
Sarah Williams (IU)

THE UNIVERSITY OF CHICAGO

efi.uchicago.edu
ci.uchicago.edu