Nimbus Research: Platforms to Fit

Kate Keahey
keahey@mcs.anl.gov
Computation Institute
Argonne National Laboratory
University of Chicago

Advanced Photon Source

- Brightest synchrotron X-rays
- 66 independent beamlines: unique purpose and software
- Operates 24 hours per day / 6 days per week
- Data reconstruction/analytics performed during experiment
- Mira for on-demand analysis
- Requirements
  - On-demand access
  - Control over the environment
  - The utilization concern

Smart Cities

- Exploring data about the urban environment
  - From sensors to social networks, and other dynamic data sources
  - Understanding a range of physical and social phenomena
- Requirements
  - Handling data volatility
  - Data-dependent processing
- Collaborators: AoT, NCSA

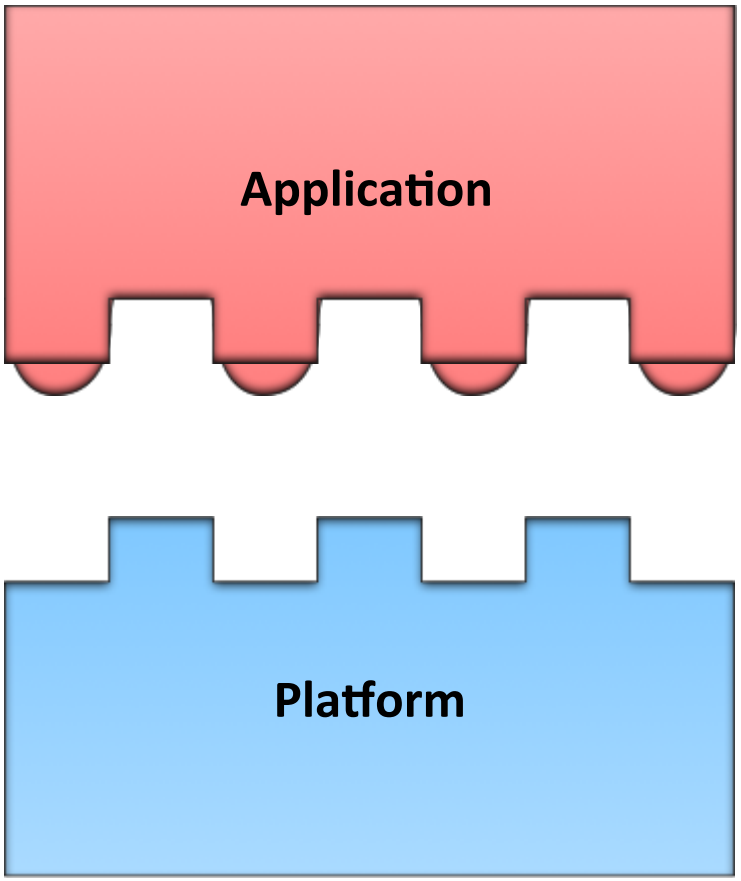
Internet of SMART Things

- Sensors: New Moore's Law changing the playing field
  - Smaller, smarter, more accurate, energy-efficient
- CPUs everywhere
  - Wearable, swallowable, etc.
- "Planetary Instrument": a Cyber-Physical System
  - Data Analysis
  - From observing to predicting
- Projects
  - Waggle, IFC

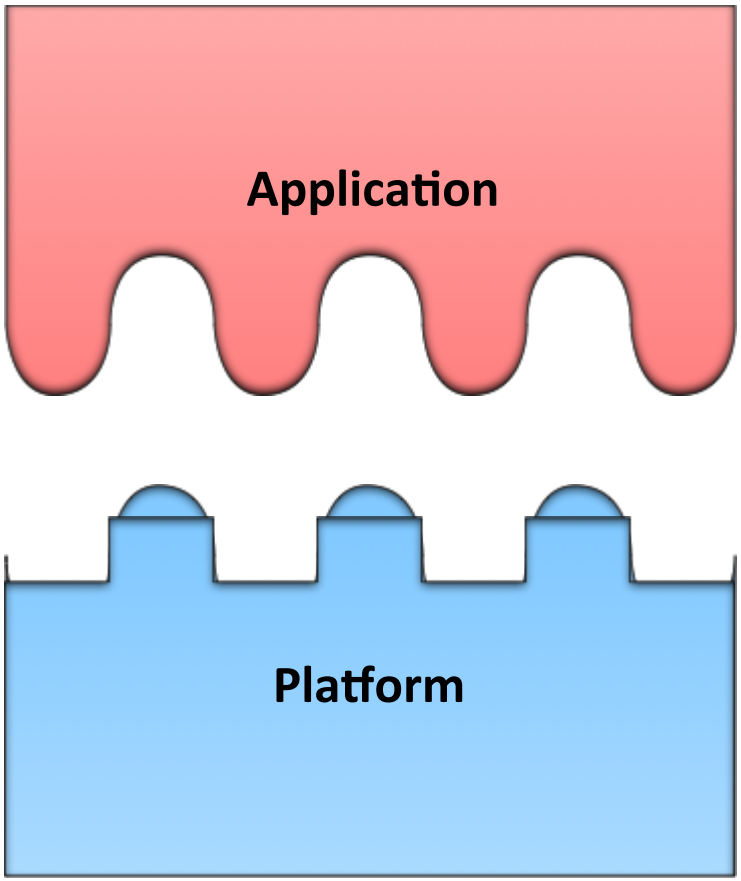
Figure 1: Gene-Z™: Hand-held genetic diagnostics (LAMP or PCR), iPod user interface, disposable chips.

(exploring) Microfluidics device for fast detection of genetic markers

Platform to Fit: a Tale of Two Questions



This is my platform, how can you adapt the application to best use the platform?



This is my application, what is the best platform you can give me?

Challenges

- Coarse-grained platform model
- Manual discovery and provisioning
- Application Portability
- Programming models optimizing to fixed assumptions
- Fault tolerance is expensive
- Emphasizes quantitative factors
- Inefficiency is on the application side

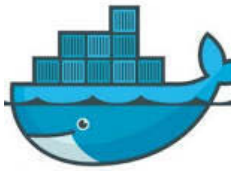
- Finer-grained platform model
- Automated discovery and provisioning
- Performance isolation
- Programming models that adapt to platform change
- Fault-tolerance friendly
- Emphasizes qualitative factors
- Significant understanding of application requirements and modeling
- Inefficiency is on the platform/ container side

Building a Platform to Fit

- Isolation
  - From security concerns to performance isolation
- Environment ownership
  - E.g., Different OS? Different kernel?
- The need for speed
  - Performance as near to native as possible, support for fast interconnects, performance isolation, noise, startup time
- Functionality
  - E.g., Snapshotting, suspend/resume, live migration
- Adaptability: scale out/scale up

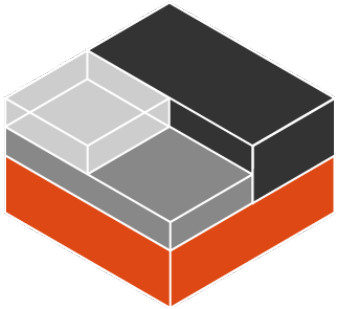
Platform to Fit: Virtualization and Containers

Palacios



Docker

LXD



Bare Metal

Platform to Fit: How Containers Compare

Feature	KVM	Palacios	LXD	Docker
Guest OS and Kernel	Any OS and kernel	Any OS and kernel	Linux based, uses host kernel	Linux based, uses host kernel
Snapshot	Yes	Yes	Yes	Not yet.
Spin-up/Start-up time	Less than one Minute	Less than one Minute	Seconds	Seconds
Security/ Isolation	Strong guest isolation	Assumes guest OS will not subvert VMM	Security features of Linux; by default containers are unprivileged	Security features of Linux; by default containers are unprivileged
Performance	Good	Better	Best	Best

Platform to Fit: Availability

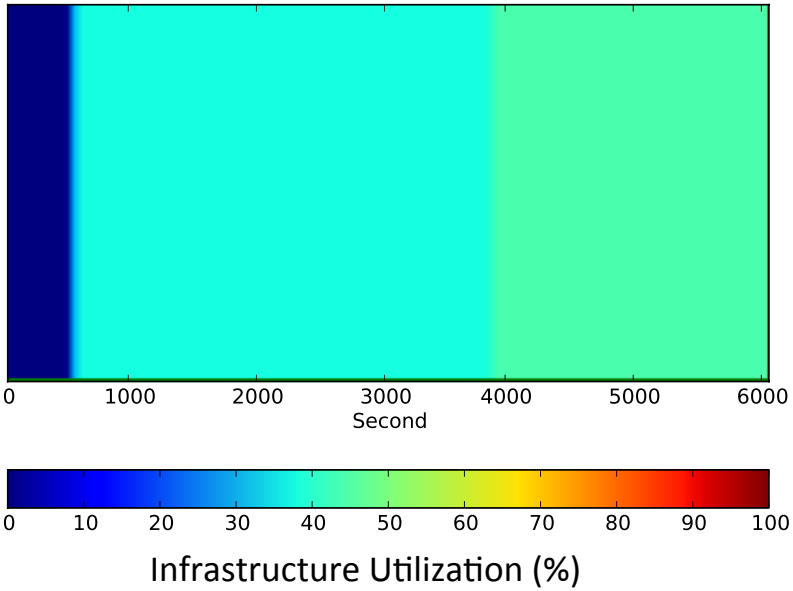
- Challenge: availability versus cost/utilization
- Example lease types:
  - On-demand: user decides availability
  - On-availability: provider decides availability, e.g., spot leases
- Lease parameters

Platform to Fit: Availability

On-Demand Only

Average utilization: 36.36%

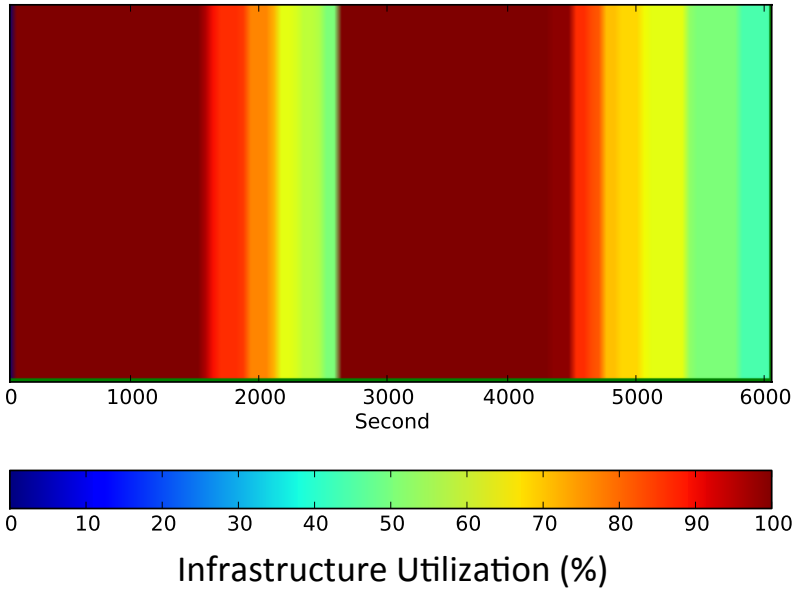
Maximum utilization: 43.75%



On-Demand and On-Availability

Average utilization: 83.82%

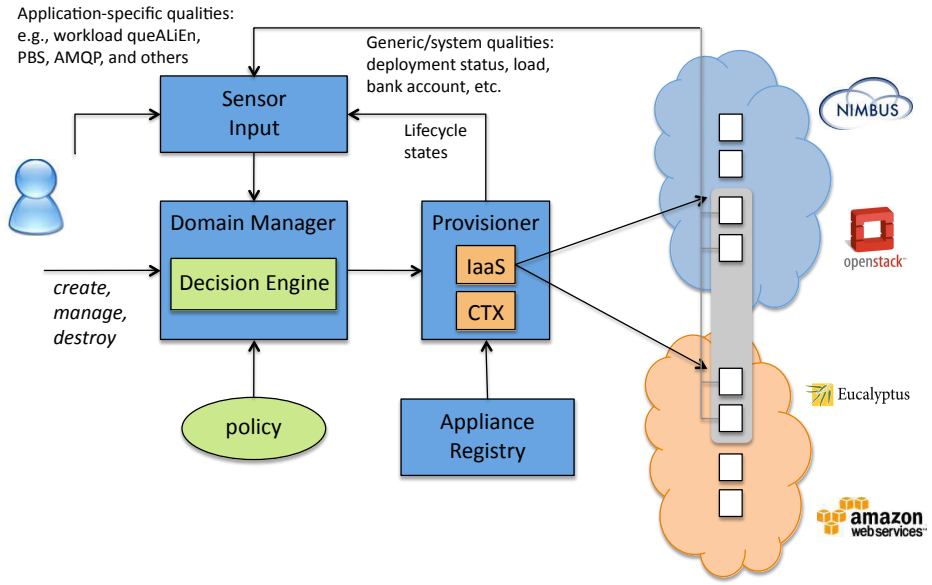
Maximum utilization: **100%**



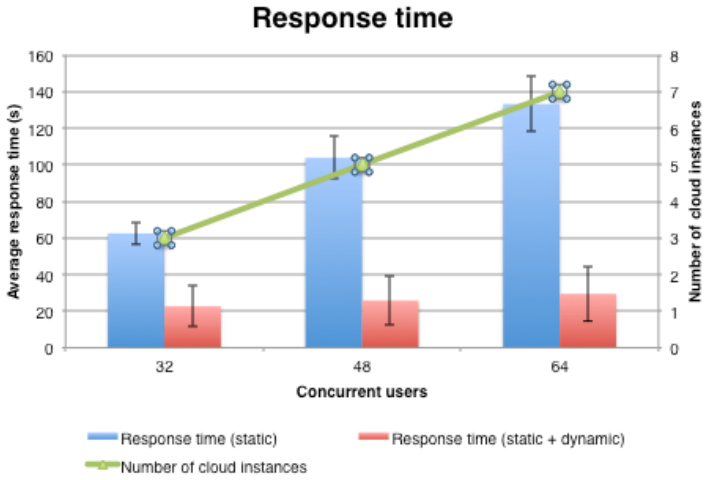
Marshall et al., CCGrid'11

Adaptable Platform: Computation to Fit

Compute environment that automatically scales to fit the evolving need of application or community



Scaling based on system and application factors

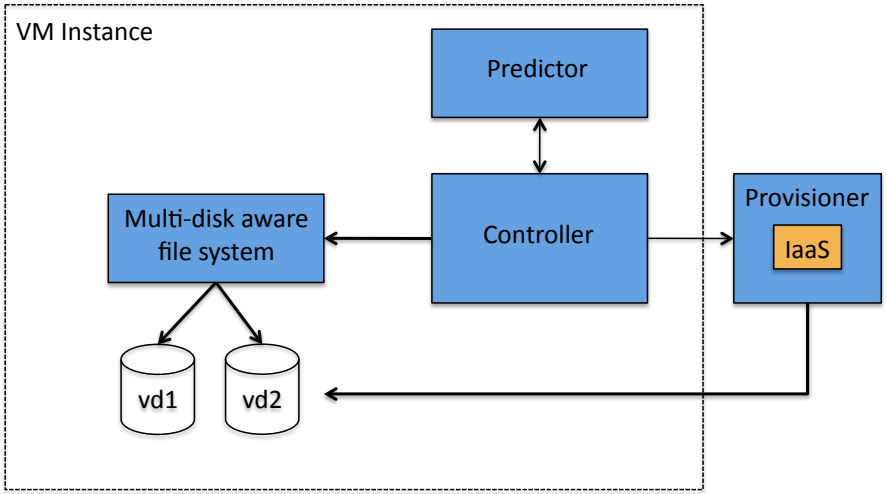


Response time of a CyberGIS application

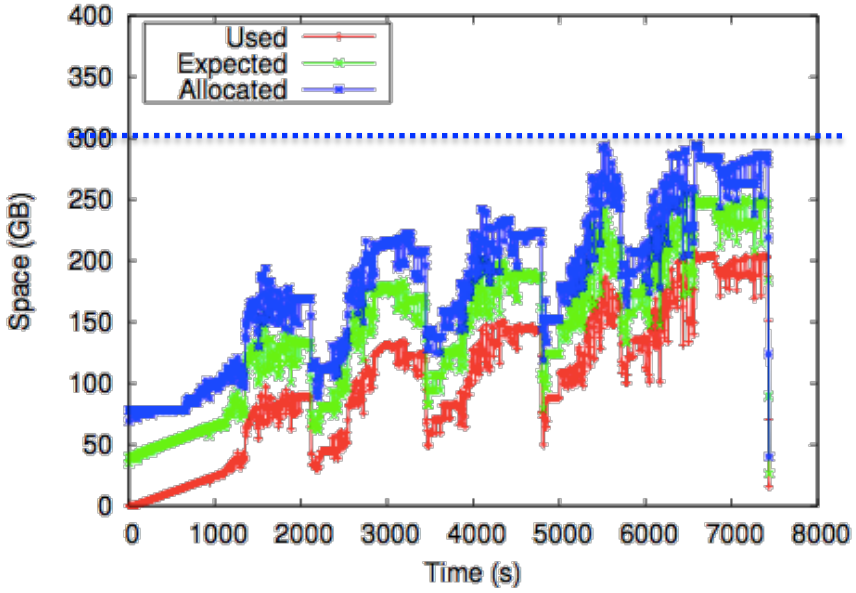
Riteau et al., ScienceCloud 2014)

Adaptable Platform: Storage to Fit

Storage that automatically scales to fit the evolving application needs both in terms of size and type



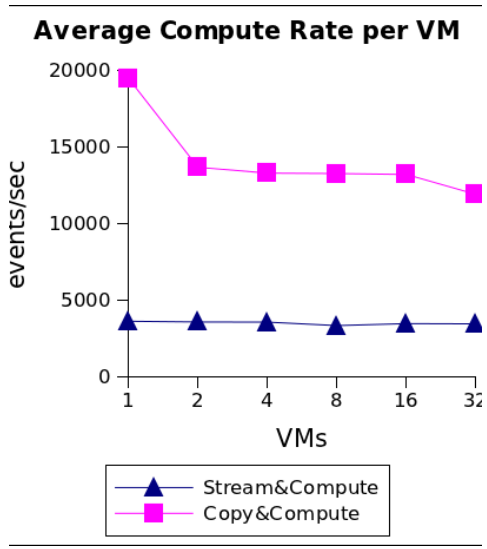
Adaptive storage scaling system



Predictive storage scaling for K-means
Nicolae et al., IPDPS 2014

Adaptable Platform: Network to Fit

- Programmable networks: SDN and OpenFlow
- Reservations, e.g., OSCARS
- Monitoring network congestion to adjust global provisioning strategies



Comparison of compute rates resulting from different streaming scenarios

Tudoran et al., CCGrid'14

Parting Thoughts

- With new opportunities, usage changes
- Is on-demand access the “new network” for the datacenter?
- Leveraging industry trends
  - Joining discussion, proposing solutions
- The power of ∃: demonstrations and challenges

www. chameleoncloud.org

CHAMELEON: A LARGE-SCALE, RECONFIGURABLE EXPERIMENTAL ENVIRONMENT FOR CLOUD RESEARCH

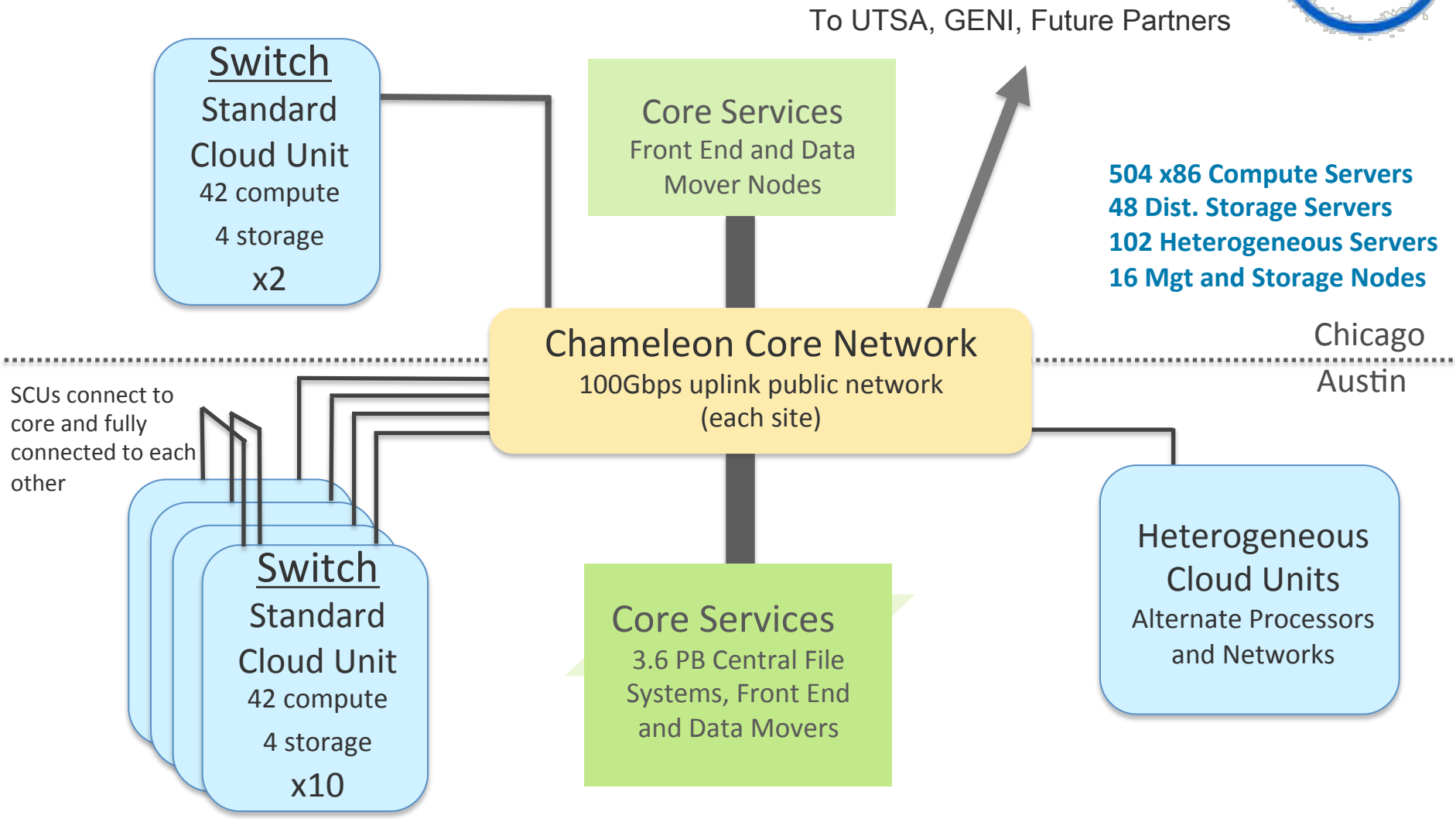
Kate Keahey

keahey@anl.gov

CHAMELEON: A FLEXIBLE AND POWERFUL EXPERIMENTAL INSTRUMENT

- Large-scale: "Big Data, Big Compute, Big Instrument research"
  - ~650 nodes (~14,500 cores), 5 PB disk over two sites, 2 sites connected with 100G network
- Reconfigurable: "As close as possible to having it in your lab"
  - From bare metal reconfiguration to clouds
  - Support for repeatable and reproducible experiments
- Connected: "One stop shopping for experimental needs"
  - Workload and Trace Archive
  - Partnerships with production clouds: CERN, OSDC, Rackspace, Google, and others
  - Partnerships with users
- Complementary: "Can't do everything ourselves"
  - Complementing GENI, Grid'5000, and other experimental testbeds

CHAMELEON HARDWARE



CAPABILITIES AND SUPPORTED RESEARCH

Development of new models, algorithms, platforms, auto-scaling HA, etc., innovative application and educational uses

Persistent, reliable, shared clouds

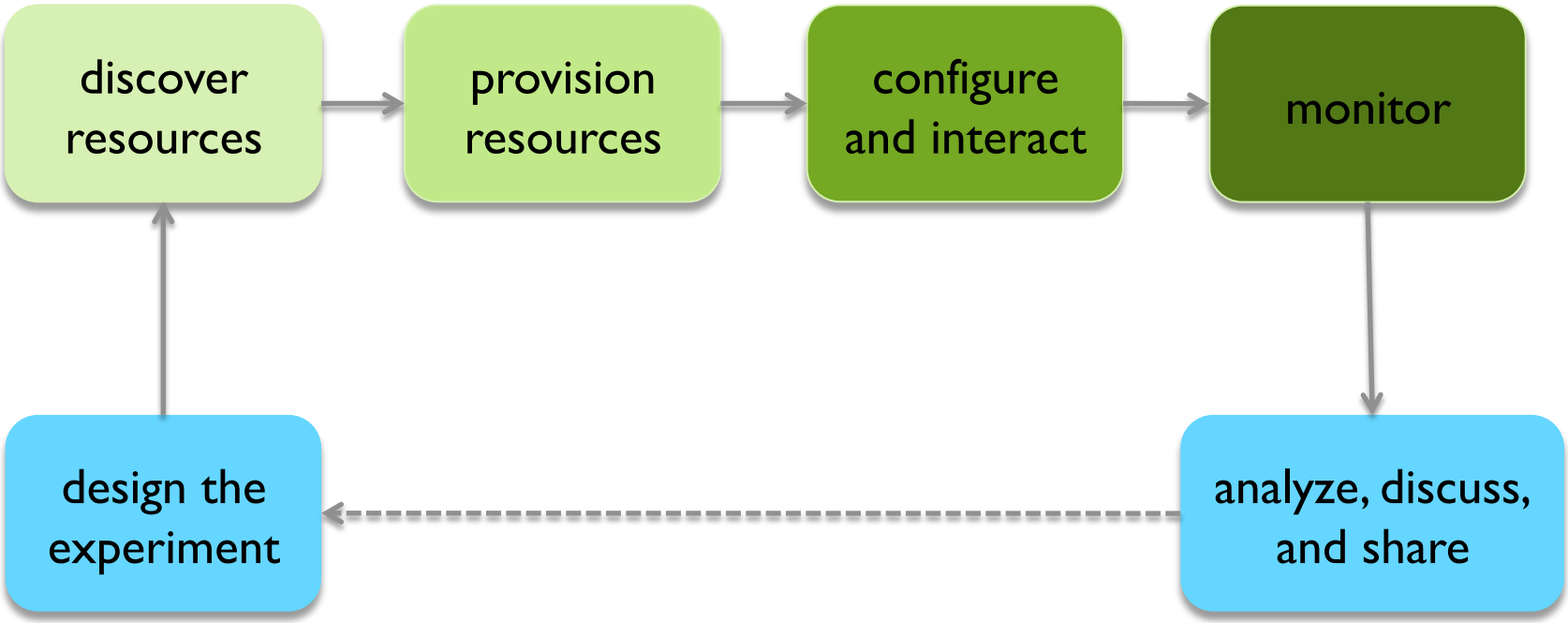
Repeatable experiments in new models, algorithms, platforms, auto-scaling, high-availability, cloud federation, etc.

Isolated partition, Chameleon Appliances

Virtualization technology (e.g., SR-IOV, accelerators), systems, networking, infrastructure-level resource management, etc.

Isolated partition, bare metal reconfiguration: OpenStack and Grid'5000

EXPERIMENTAL WORKFLOW



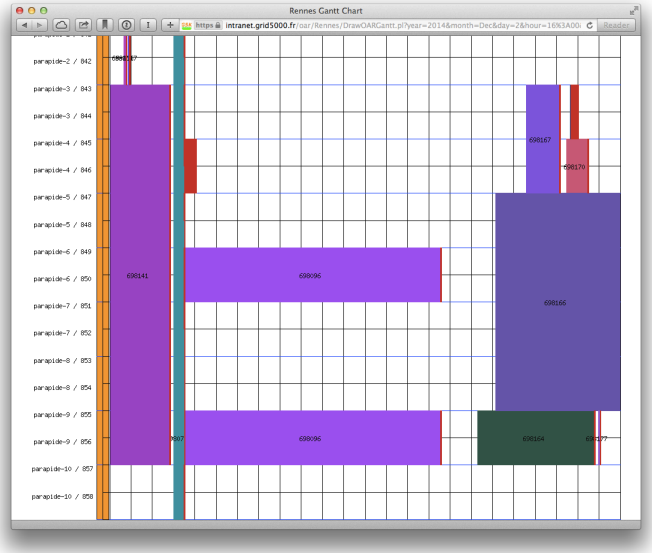
SELECTING AND VERIFYING RESOURCES

- Complete, fine-grained and up-to-date representation
- Machine parsable, enables match making
- Versioned
  - "What was the drive on the nodes I used 6 months ago?"
- Dynamically Verifiable
  - Does reality correspond to description? (e.g., failures)

- Grid'5000 Registry
  - Automated resource description, automated export to RM
- G5K-checks
  - Run at boot, acquire information, compare with resource catalog description

PROVISIONING RESOURCES

- Resource leases
- Allocating a range of resources
  - Different node types, switches, etc.
- Multiple environments in one lease
- Advance reservations (AR)
  - Sharing resources across time
- Eventually: match making, Gantt chart displays



- OpenStack Nova/Blazar
- Extensions to support working with more resources, match making, and displays

CONFIGURE AND INTERACT

- Map multiple appliances to a lease
- Allow deep reconfiguration (incl. BIOS)
- Snapshotting
- Efficient appliance deployment
- Handle complex appliances
  - Virtual clusters, cloud installations, etc.
- Interact: reboot, power on/off, access to console
- Shape experimental conditions

- OpenStack Ironic, Glance, and meta-data servers

MONITORING

- Enables users to understand what happens during the experiment
- Types of monitoring
  - User resource monitoring
  - Infrastructure monitoring (e.g., PDUs)
  - Custom user metrics
- High-resolution metrics
- Easily export data for specific experiments

- OpenStack Ceilometer

PROJECT SCHEDULE

- Now: FutureGrid@Chameleon
  - Chameleon Technology Preview
  - OpenStack FutureGrid-style cloud
  - 43 projects, 81 users, 29 institutions
- Summer 2015: New hardware: large-scale homogenous partitions available to Early Users
- Fall 2015: Large-scale homogenous partitions and bare metal reconfiguration generally available
- 2015/2016: Refinements to experiment management capabilities, higher level capabilities
- Fall 2016: Heterogeneous hardware available

GET ENGAGED

- www.chameleoncloud.org
- Use the FutureGrid@Chameleon KVM cloud
- Technology Preview on FutureGrid hardware
- Early User Program
  - Committed users, driving and testing new capabilities, enhanced level of support
  - Sign up to get access to new hardware

The most important element of any experimental testbed is users and the research they work on