



www.chameleoncloud.org

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

Principal Investigator: Kate Keahey

Co-PIs: J. Mambretti, D.K. Panda, P. Rad, W. Smith, D. Stanzione

*NSFCloud Workshop
December 11-12, 2014,
Arlington, VA*

DECEMBER 12, 2014

I



WHY EXPERIMENT?

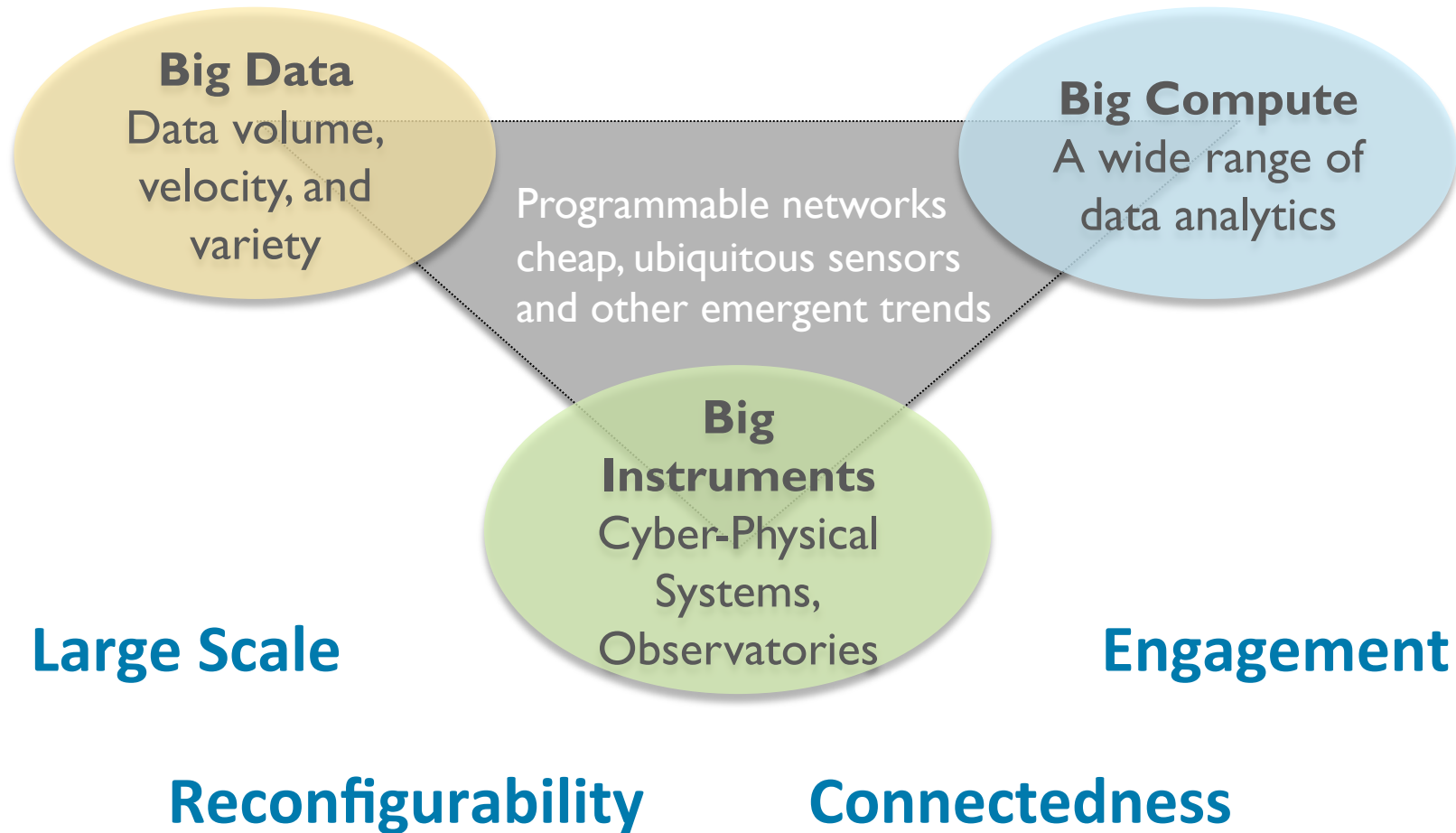


*“Beware of bugs in the above code;
I have only proved it correct, not tried it”*
(Donald Knuth)

*“In theory there is no difference between
theory and practice. In practice there is.”*
(Yogi Berra)



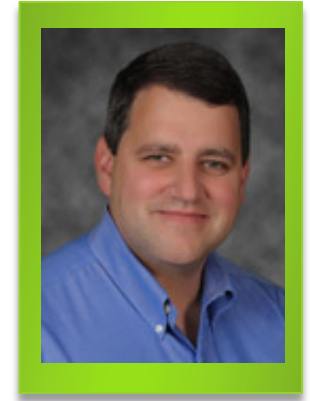
SCALING TO THE CHALLENGE



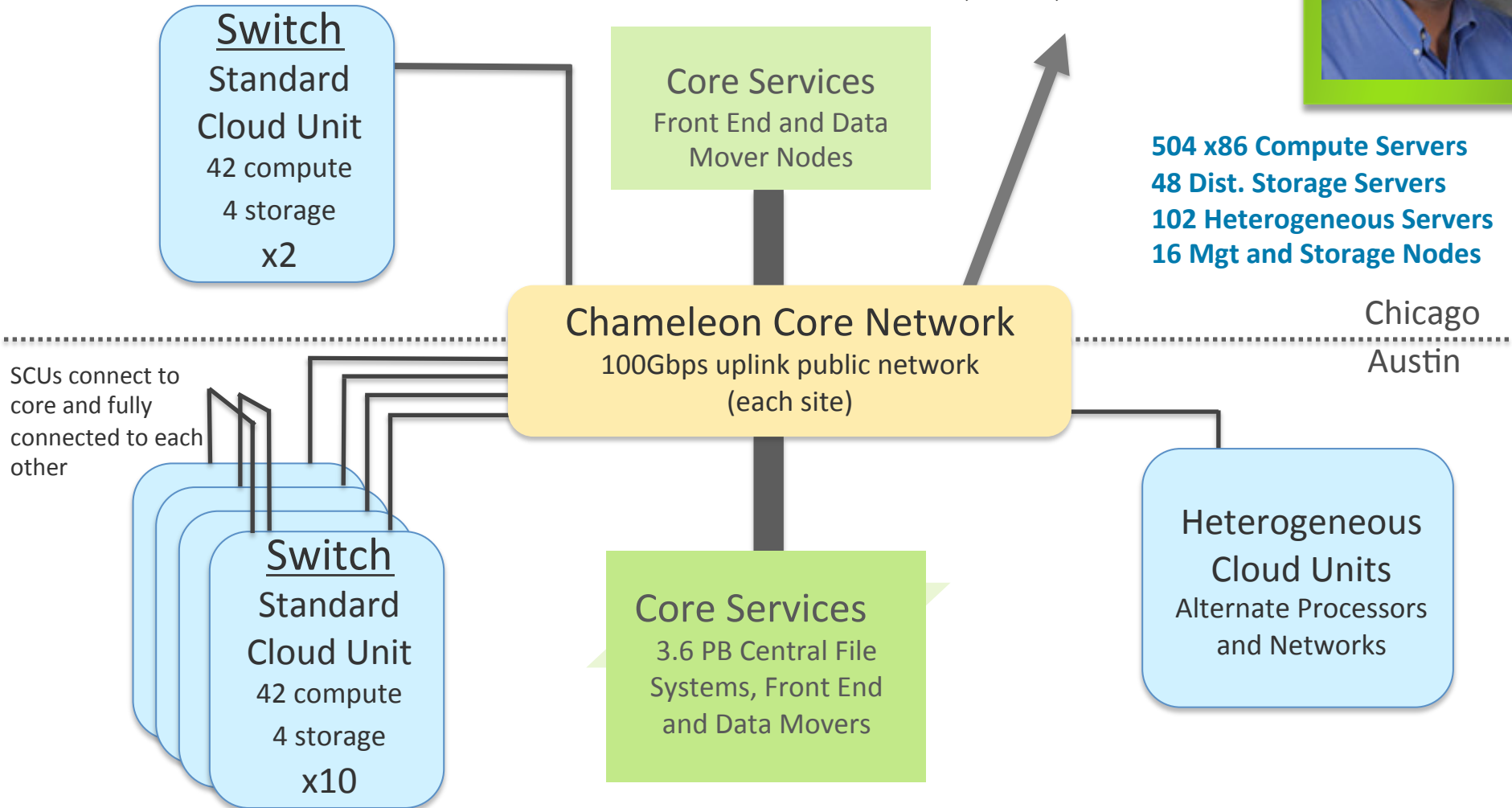
CHAMELEON: A POWERFUL AND FLEXIBLE EXPERIMENTAL INSTRUMENT

- ▶ Large-scale instrument
 - ▶ Targeting 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 instrument
 - ▶ Bare metal reconfiguration, operated as single instrument, graduated approach for ease-of-use
- ▶ Connected instrument
 - ▶ Workload and Trace Archive
 - ▶ Partnerships with production clouds: CERN, OSDC, Rackspace, Google, and others
 - ▶ Partnerships with users
- ▶ Complementary instrument
 - ▶ Complementing GENI, Grid'5000, and potentially other testbeds

CHAMELEON HARDWARE



To UTSA, GENI, Future Partners



STANDARD CLOUD UNIT

- ▶ Each of the 12 SCUs is comprised of a single 48U rack
 - ▶ Allocations can be an entire SCU, multiple SCUs, or within a single one.
- ▶ A single 48 port Force10 s6000 OpenFlow-enabled switch connects all nodes in the rack (with an additional network for management/control plane).
 - ▶ 10Gb to hosts, 40Gb uplinks to Chameleon core network
- ▶ An SCU has 42 Dell R630 compute servers, each with dual-socket Intel Xeon (Haswell) processors and 128GB of RAM
- ▶ In addition, each SCU has 4 DellFX2 storage servers, each with a connected JBOD of 16 2TB drives.
 - ▶ Can be used as local storage within the SCU, or allocated separately (48 total available for Hadoop configurations); SCU storage nodes will not be used for permanent storage.

HETEROGENEOUS CLOUD UNITS

- ▶ One of the SCUs will also contain an Infiniband network, and (hopefully) an OmniScale network
- ▶ Additional HCUs will contain:
 - ▶ 48 Intel Atom microservers
 - ▶ ARM microservers
 - ▶ A mix of servers with:
 - ▶ High RAM
 - ▶ FPGAs (Xilinx/Convey Wolverine)
 - ▶ NVidia K40 GPUs
 - ▶ Intel Xeon Phi

CHAMELEON CORE HARDWARE

▶ Shared Infrastructure:

- ▶ In addition to distributed storage nodes, Chameleon will have 3.6PB of central storage, for a *persistent* object store and shared filesystem.
- ▶ An additional dozen management nodes will provide data movers, user portal, provisioning services, and other core functions within Chameleon.

▶ Core Network

- ▶ Force10 OpenFlow-enabled switches will aggregate the 40Gb uplinks from each unit and provide multiple links to the 100Gb Internet2 layer 2 service.

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, full bare metal reconfiguration

SOFTWARE: CORE CAPABILITIES

Persistent Clouds
(OpenStack)

Persistent Cloud

User Clouds

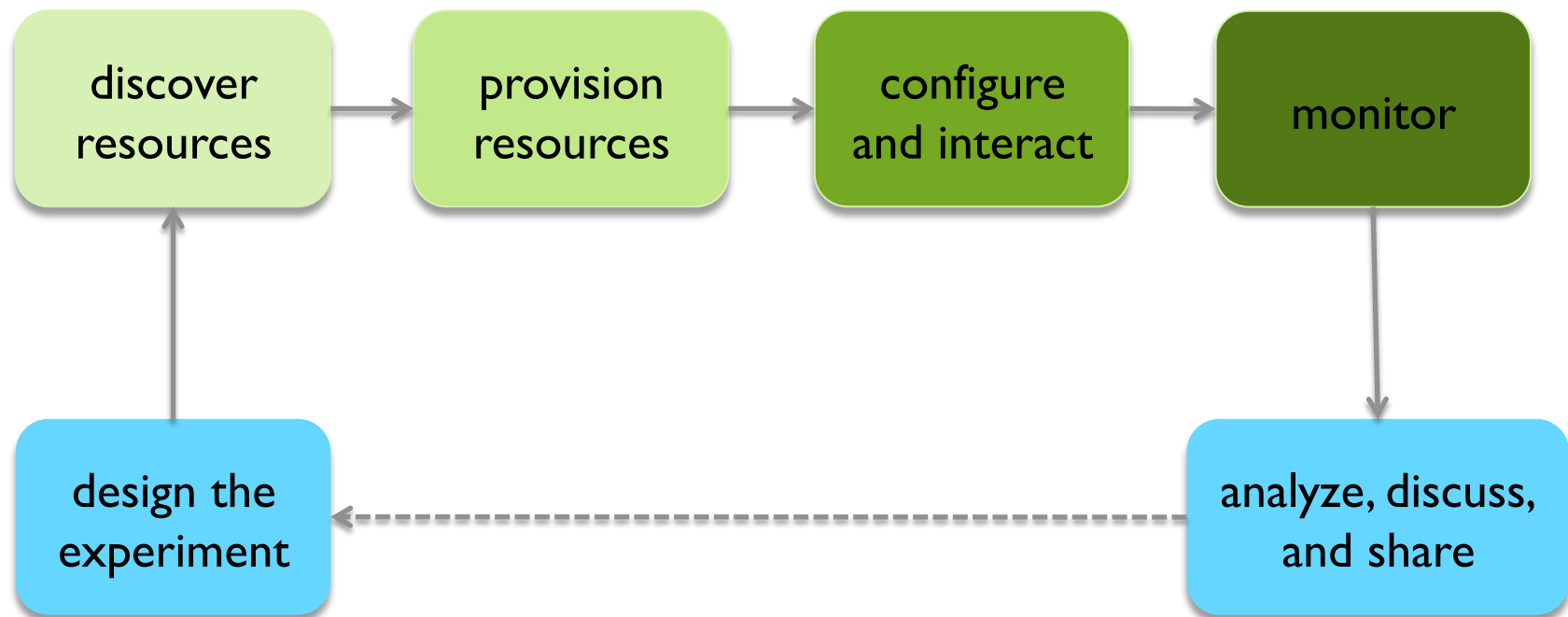
Chameleon Appliance Catalog

A library of generic, special-purpose, and educational environments

Discovery, Provisioning, Configuration, and Monitoring

Testbed representation and discovery (Grid'5000)
Nova/Blazar, Ironic, Neutron, Ceilometer
(OpenStack, Rackspace OnMetal)

SUPPORT FOR EXPERIMENT WORKFLOW



SELECTING AND VERIFYING RESOURCES

- ▶ Complete and current representation of actual testbed resources
- ▶ Fine-grained representation
- ▶ Machine parsable, enables match making
- ▶ Versioned
 - ▶ “What was the drive on the nodes I used 6 months ago?”
 - ▶ Hardware upgrades, maintenance, extensions
- ▶ Dynamically Verifiable
 - ▶ Does reality correspond to description? (e.g., failures)
 - ▶ Can't afford false assumptions!

RESOURCE CATALOG

▶ Grid'5000 Registry

- ▶ Largely automated resource discovery and fine-grained description
- ▶ Browseable: REST, CLI, and web interfaces
- ▶ Match making
- ▶ Automated description export for the Resource Manager

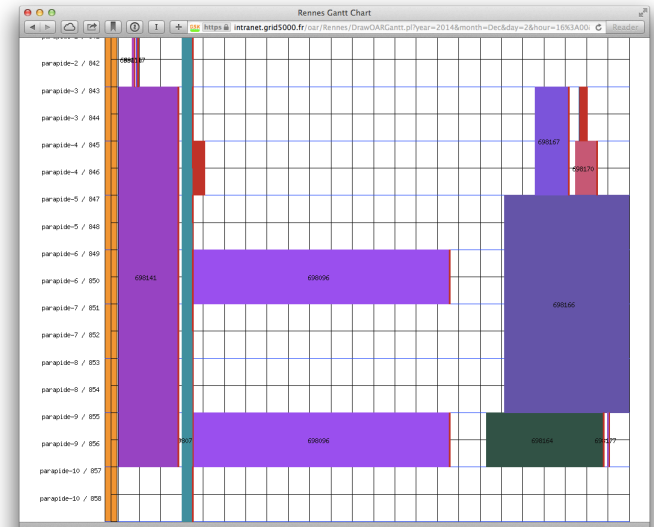
▶ G5K-checks

- ▶ Run at node boot and acquire information on node using ohai, ethtool, etc.
- ▶ Compare with resource catalog description

```
"processor": {
  "cache_l2": 8388608,
  "cache_l1": null,
  "model": "Intel Xeon",
  "instruction_set": "",
  "other_description": "",
  "version": "X3440",
  "vendor": "Intel",
  "cache_ll1": null,
  "cache_llid": null,
  "clock_speed": 2530000000.0
},
"uid": "graphene-1",
"type": "node",
"architecture": {
  "platform_type": "x86_64",
  "smt_size": 4,
  "smp_size": 1
},
"main_memory": {
  "ram_size": 17179869184,
  "virtual_size": null
},
"storage_devices": [
  {
    "model": "Hitachi HDS72103",
    "size": 298023223876.953,
    "driver": "ahci",
    "interface": "SATA II",
    "rev": "JPFO",
    "device": "sda"
  }
],
```

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

NETWORKING CAPABILITIES

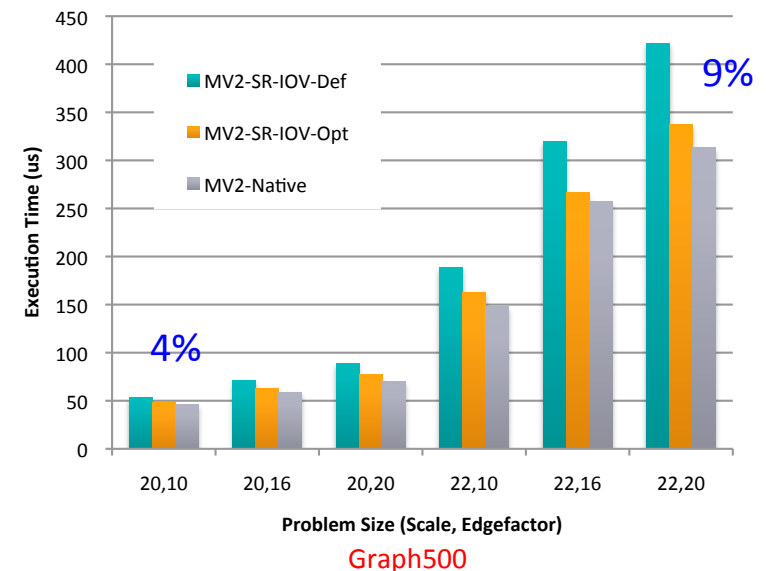
- ▶ Expose SDN, OpenFlow, etc. to users
 - ▶ Isolation
 - ▶ Hybrid network capabilities
 - ▶ Programmable topologies
 - ▶ Integration with other resources within and external to the testbed
- ▶ Pushing 100G network to the limit
 - ▶ Using 100G + SDN optimally is a challenge
 - ▶ Chameleon appliances and services allow experimenters a highly granulated view into -- and control over -- traffic flows
- ▶ Integration with GENI
 - ▶ Data plane integration
 - ▶ Control plane integration
 - ▶ Common policy context



HIGH PERFORMANCE NETWORKS



- ▶ Support virtualization for Big Compute and Big Data
- ▶ Chameleon Appliances:
 - ▶ HPC MPI with IB & SR-IOV
 - ▶ Hadoop with SR-IOV
 - ▶ Integration with OpenStack, etc.
- ▶ Further support for Big Data and Big Compute



Application-Level Performance (8 VM * 8 Core/VM)

EDUCATION



- ▶ New courses with new content
 - ▶ Electronic textbooks, multi-media content, and Chameleon Appliances
 - ▶ Graduate courses for Fall 2015: CS6463 (Cloud and Big Data), CS6643 (Parallel Processing), ECE5243 (Data Analytics in Cloud), CS 6393 (Advanced Topics in Computer Security), and others
- ▶ Broaden a Cloud Education Community by Reaching out to the MSI network and other institutes
- ▶ General education: MOOCs and other content
- ▶ Chameleon-specific training and training materials

INDUSTRY OUTREACH



- ▶ Fostering relationship between academia and industry
 - ▶ Industry Board: explore synergy between industry and academia
 - ▶ Facilitating industry-sponsored research projects
 - ▶ Interoperability with industry standards
 - ▶ Commercialization
- ▶ Workload and Track Archive

OUTREACH AND ENGAGEMENT

- ▶ Advisory Bodies
 - ▶ Research Steering Committee: advise on capabilities and priorities needed to investigate upcoming research challenges
 - ▶ Industry Advisory Board: explore synergy between industry and academia
- ▶ Early User Program
 - ▶ Committed users, driving and testing new capabilities, enhanced level of support
- ▶ Chameleon Workshop
 - ▶ Annual workshop to inform, share experimental techniques solutions and platforms, discuss upcoming requirements, and showcase research

PROJECT SCHEDULE

- ▶ Fall 2014: FutureGrid@Chameleon is ready!
- ▶ Spring 2015: Initial bare metal reconfiguration capabilities available on FutureGrid UC&TACC resources for Early Users
- ▶ 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

FUTUREGRID@CHAMELEON

- ▶ Chameleon Portal
 - ▶ FG users can import their projects and accounts
 - ▶ FG user data (accounts, images, volumes, etc.) will be reactivated with account
 - ▶ Available generally by end of year
- ▶ Hotel (UC) and Alamo (TACC) configured FG-style
 - ▶ OpenStack Juno with KVM images
 - ▶ Available via a single interface as OpenStack regions (replicated Keystone)
 - ▶ The same set of images available for both

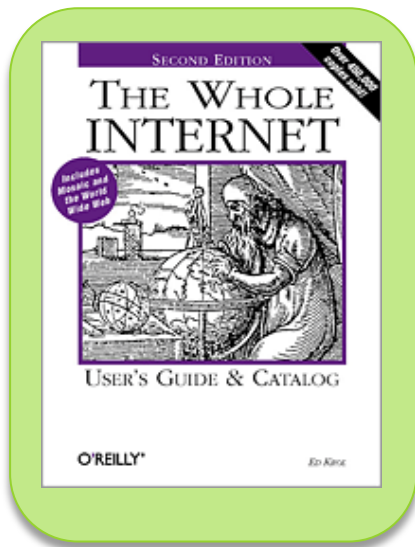
THE TESTBED IS THERE...

- ▶ Large-scale, responsive experimental testbed
 - ▶ Targeting critical research problems at scale
 - ▶ Evolve with the community input
- ▶ Reconfigurable environment
 - ▶ Support use cases from bare metal to production clouds
 - ▶ Support for repeatable and reproducible experiments
- ▶ One-stop shopping for experimental needs
 - ▶ Trace and Workload Archive, user contributions, requirement discussions
- ▶ Engage the community
 - ▶ Network of partnerships and connections with scientific production testbeds and industry
 - ▶ Partnerships with existing experimental testbeds
 - ▶ Outreach activities

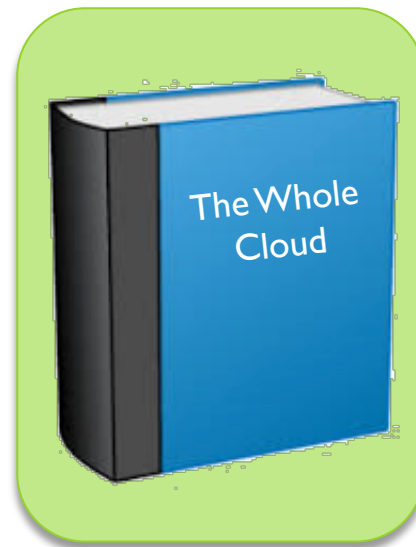
...“JUST” ADD RESEARCH

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

From the Internet...



...to cloud...



...and beyond

