UCLA & UC Research Cyberinfrastructure, the Grid and the Network

Jim Davis
Associate Vice Chancellor & CIO
Bill Labate
Director Research Computing Technologies
UCLA
Acknowledgements

Rosio Alvarez, LBNL
Steve Beckwith, UCOP
Fran Berman, UCSD
Art Ellis, UCSD
David Ernst, UCOP
Warrant Mori, IDRE
Larry Smarr, Calit2
Mike Van Norman, UCLA
Margo Reveil, UCLA
Tammy Welcome, LBNL
Many Faculty
CENIC
UC CIOs and VCRs
UC CPG, RCG, DC WG
Why UC and UCLA CI: The Researcher Perspective

UC CI

P2P Experimentation among researchers

Capability & capacity not readily supported by researcher unit or institution

Research pipeline: Capability available as research needs grow

Team based intra/inter university research

Researcher pipeline: Capability available as researcher expertise grows
Why UC and UCLA CI: The VCR and CIO Perspective

• P2P (not centralized) driven research
  – Independently owned and administered (faculty, center, institutional) resources to ensure research-driven need, capability, and capacity
  – Resources joined together on a software and hardware infrastructure based on value

• A CI designed to balance and manage competing dimensions
  – Individual vs. team research,
  – Research autonomy vs. capacity,
  – Disciplinary need vs. standardization,
  – Ownership vs. sharing,
  – Specialized vs. scale,
  – Grant funding vs. institutional investment
  – Sustainability vs. short life cycles

• Dependable, consistent and policy-based resource sharing
• Unused capacity repurposed based on agreed upon policies
Why A Shared Cluster Model: The Operational Perspective

- More efficient use of scarce people resources
- Standalone clusters have separate “everything”. Storage, head/interactive nodes, network, user space, configuration
- Higher overall performance than a standalone Cluster
- Recovery of compute cycles wasted on non-pooled clusters – 50% in some cases
- More efficient data center operations
- Better security
- Dedicated system admin, application support, research personnel to manage efficiently & correctly
  - seven, 32 node clusters @ .2 FTE = 1.4 FTE vs. one, 200 node cluster @ 1.4 FTE vs. one, 400 node cluster @ 2.5 FTE
- Better machine performance
  - Estimated 30%+ of cycles lost to I/O wait state for parallel jobs running on GigE versus Infiniband
    - Faster scratch and home directory space increases efficiency
    - OS, applications, compilers, libraries and queuing system are optimized
- Better data center efficiency
  - Data centers 3 – 4 x more efficient than ad hoc space
  - Regional data centers more efficient than distributed
UC CI and the Grid – Current Status

Potentially 60 Tflops Available

14.2 Tflops
19.2 Tflops
23 Tflops(1)
3.1 Tflops(2)
19.2 Tflops
10GB CalREN/CENIC Network
UC CI Data Center South At SDSC
UC CI Data Center North At LBL
UC CI and the Grid – Current Status

(1) Former AAP Resources
(2) Includes New Broadcom Cluster
The Shared Cluster Concept – UCLA Illustrated

Campus General Purpose Cluster (160 Cores)

Base Cluster (96 Cores)
Contributed Clusters (space for ~12,000 Cores)

Research Virtual Shared Cluster

Infiniband Interconnect

Parallel

Open to all campus researchers for commercial parallel apps & parallel jobs

Researchers have guaranteed access to equivalent number of their contributed nodes for serial or parallel jobs with access to additional pooled serial or parallel cycles.

Serial

Open to all campus researchers for commercial serial apps & serial jobs
Value to Researchers

- Administration of the cluster hardware, OS, queuing system and applications by a dedicated, professional staff
- High performance network, interconnect, and home/scratch storage (not cost-effective for individual clusters)
- Dedicated data center facility
- Ability to use surplus cycles across the entire cluster
- Access to a highly optimized applications-only cluster
  - pool licenses with others users
  - access to additional commercial as well as open source applications
- Web access to cluster without knowledge of the command line interface
Computational Needs (managed by policy on single facility)
  • General Purpose Campus Cluster
    • Periodic, infrequent, those with no dedicated resources
  • Pooled Cluster
    • Shared Cluster model
  • Surge – Local campus, UC, or external resources
    • Harvested cycles, special arrangement, Grid, “Cloud”

Concentration of Physical Resources in Data Centers

UCLA Research Desktop Concept
  • Applications – Available via the Grid (storage & computation not at desktop)
  • Connectivity – 1GB - monitor for applicability
  • Visualization – Local install, Grid, or cluster based visualization.
    • Scale down for desktop, scale up for formal presentation, higher resolution.
    • Scale with individual requirements and support capability available
  • Monitor for special need, HD, latency
Emerging UCLA Business Model with Researchers for Virtual Shared Cluster

- **One Time Costs to Researchers**
  - Researchers fund nodes & storage
  - Storage - $3K per TB, includes backup
    - *Some pushback on price – looking at different cost/performance tiers*
  - Infiniband interconnect- card and cable approximately $470 per node
    - *Most see benefit of IB, especially those with parallel code*

- **Harvesting and use of unused cycles**
  - Computing resources returned in 24 hours or less
    - *General acceptance although some want a shorter period. Looking at a variable policy*

- **Adherence to basic, minimum, system standards**
  - *No real issue as our standards are based on the current price/performance sweet spot.*
Emerging UCLA Investment in the Shared Cluster

- UCLA furnishes
  - System Administration and HPC Applications support
    - Universal approval
    - Applications support highly desirable
  - Infiniband and Ethernet infrastructure
    - Highly supported, generally higher quality and performance than researchers would buy
  - High performance scratch space
    - Very desirable, seen as necessary to the overall performance of the cluster
  - The data center including environmentals and racks
    - Expected. Seen almost as a “given”
  - High Performance Networking to the Data Centers
UCLA Shared Cluster Build Out

10 Projects 264 nodes > 1000 cores 350 TB

• Current
  • Brad Hansen, Astrophysics, 22 nodes, 2TB storage
  • Moshe Buchinsky, Economics, 10 nodes, 1TB of storage
  • John Miao, Physics, 47 nodes, 5TB of storage
  • Eleazar Eskin, Computer Science, 32 nodes, 5TB of storage
  • Neil Morley, Physics, 21 nodes, 2TB of storage
  • Mark Cohen, Neuroimaging, 8 nodes, 2TB storage
  • David Teplow, Neurology, 8 nodes, 1TB of storage
  • David Saltzberg, Astrophysics, 5TB of storage

• Pending
  • Stan Nelson, Human Genetics, 96 nodes, 300TB+ storage
  • Various, Atmospheric Sciences, 20+ nodes, 20-30TB storage
UC Cyberinfrastructure Initiative

- 10 campuses, 5 medical centers, SDSC, LBL
- High potential for regional and system capability and capacity
- Production prototype for UC Grid in operation - 3 campuses connected - 3 in progress
- Variation of need, capability, investment, policy
- Requires integrated networking, data centers, grid, computation & storage, management, investment, policy and governance,
- Proposed UC CI Pilot
  - How to work as a UC system – non-trivial
  - Build the experience base on a system shared resources
  - Build the experience based with a shared regional data centers
  - Build the business model
  - Build the trust of the faculty researchers
Proposed UC Research Virtual Shared Clusters
North & South UC CI Clusters

Research Virtual Shared Cluster

Base Cluster
(3,200 Cores)
Phase 1

Contributed Clusters
(space for ~ TBD Cores)
Phase 2

Infiniband Interconnect

Resources allocated for UC CI Projects. (3,200 cores at each location)

Researchers have guaranteed access to equivalent number of their contributed nodes for jobs with access to additional pooled surplus cycles.

Phased to Build Researcher Trust
UC CI Project Interest - all campuses

- Phylogenomics Cyberinfrastructure for Biological Discovery
- Optimized Materials and Nanostructures From Predictive Computer Simulations
- Space Plasma Simulations
- Nano-system modeling and design of advance materials
- Study organic reaction mechanisms and selectivities, enzyme design, and material and molecular devices
- Oceanic Simulation of Surface Waves and Currents from the Shoreline to the Deep Sea
- Particle-in-cell simulations of Plasmas
- Dynamics and Allosteric Regulation of Enzyme Complex
- Functional Theory for Multi-Scaling of Complex Molecular Systems and Processes
- Development and mathematical analysis of computational methods
- Computational Chemistry and Chemical Engineering Projects
- Study of California Current System
- Physics-Based Protein Structure Prediction
- Speeding the Annotation and Analysis of Genomic Data for Biofuels and Biology Research
- Application of Community Climate System Model (CCSM) to study the interactions of new biofuels with carbon cycles
- Research in the physics of real materials at the most fundamental level using atomistic first principles (or ab initio) quantum-mechanical calculations.
- Universe-Scale Simulations for Dark Energy Experiments
Distributed Storage Driven by Need

- “Workflow” – output is manipulated in multiple locations
  - Multiple computational facilities
  - Output data is prepared in one location, visualization resources are in another
  - Creation and greater usage of data preprocessing services
  - Closely coupled with a backup and/or hierarchical storage management system. Disaster recovery

- Workflow impacts
  - Robust and reliable storage to facilitate workflow
  - Robust and reliable HP inter institution networking and networking to campus data centers
  - “Quality of service” is crucial for proper scheduling of resources
    - Computational resources are available but the data has not moved. Data arrives too late, job falls back into queue
    - Move or Stream
    - On-demand
  - “Good enough” vs. highest quality

- Monitoring other drivers of localized campus need
  - High Definition
  - Instrumentation
UC Data Center Initiative

- Integrated approach to long range computing requirements for UC:
  - Project new 60 – 80,000 sq ft driven mostly by research
  - Increased energy costs > $15 million unless addressed with more efficient data centers
  - Support the technical infrastructure required to support the UC CI
  - Green
  - Fast track needs for additional capacity (UCD, UCDMC, UCLA, UCSB, UCSC)
  - Begin with existing space at SDSC and LBL

- Optimize UC spend
  - Network capabilities
  - Energy efficient expertise
  - Economies of scale
  - Sharing or resources
  - Best procurement practices
  - A change in funding models
The Network

- CENIC HPR upgrade critical inter UC capability and national and international capability
  - 10GB or greater at key aggregation points
- Campuses
  - Focusing connectivity with applicable bandwidth
    - Data centers
    - Large institutes
    - Visualization centers
  - Currently building end-to-end services on installed shared network base
  - CENIC HPR network to each campus border: Layer 3 connectivity at 10Gb/s as well as the new Layer 2 and Layer 1 circuit services.
- Monitoring local, distributed QoS needs – High Definition, low latency, dedicated wave, Layer 1/Layer 2 services, instrument control, medical
  - Monitoring UCSD
Governance/Building Trust – The People Side

Faculty + Staff Oversight

VCR-CIO CI Implementation Team

Investment Functionality Policy Oversight

Dedicated Staff Support + Campus Staff

IDRE Executive Committee

VCR-CIO

Investment Functionality Policy Oversight

Academic Technology Services + Dept Staff
UC Cloud Project

• New project to add a cloud computing capability to the UC Grid
  • Provide an on-demand, customizable environment to compliment the Grid’s fixed environment
• Based on the open source Eucalyptus project out of Rich Wolski’s CS group at UCSB
  • *Elastic Utility Computing Architecture Linking Your Programs To Useful Systems*
• Web services based implementation of elastic/utility/cloud computing infrastructure
  • Linux image hosting ala Amazon
  • Interface compatible with EC2
  • Works with command-line tools from Amazon w/o modification
  • Enables leverage of emerging EC2 value-added service venues

*Graphic and verbiage courtesy of Rich Wolski. Presented at UCSCS 08’*
UC CI and the Grid

UC CI and the Grid Portal

Makes computational clusters at UCLA available from a single web location.

View resource availability and status.

Submit batch jobs and generate program input for them, if desired, by filling out forms.

Visualize data.

ssh to a cluster head node or open an xterm there.

Submit batch jobs

and generate program input for them, if desired, by filling out forms.

Visualize data.

ssh to a cluster head node or open an xterm there.

Single-campus grid architecture

Multiple-campus grid architecture

UC Grid Portal

Makes computational clusters system-wide available from a single web location.

Work with files.

Program.

Submit batch jobs

and generate program input for them, if desired, by filling out forms.

Visualize data.

ssh to a cluster head node or open an xterm there.

Single-campus grid architecture

Multiple-campus grid architecture

UC Grid Portal

Makes computational clusters system-wide available from a single web location.

Work with files.

Submit batch jobs

and generate program input for them, if desired, by filling out forms.

Visualize data.

ssh to a cluster head node or open an xterm there.

Single-campus grid architecture

Multiple-campus grid architecture

UC Grid Portal

Makes computational clusters system-wide available from a single web location.

Work with files.

Submit batch jobs

and generate program input for them, if desired, by filling out forms.

Visualize data.

ssh to a cluster head node or open an xterm there.

Single-campus grid architecture

Multiple-campus grid architecture

UC Grid Portal

Makes computational clusters system-wide available from a single web location.

Work with files.

Submit batch jobs

and generate program input for them, if desired, by filling out forms.

Visualize data.

ssh to a cluster head node or open an xterm there.

Single-campus grid architecture

Multiple-campus grid architecture

UC Grid Portal

Makes computational clusters system-wide available from a single web location.

Work with files.

Submit batch jobs

and generate program input for them, if desired, by filling out forms.

Visualize data.

ssh to a cluster head node or open an xterm there.

Single-campus grid architecture

Multiple-campus grid architecture