

Bridging the Gap Between the CIO and the Scientist

Michael A. Pearce – Deputy CIO
University of Southern California (USC)

Workshop
April, 2006

USC General Information

- Private R1 institution founded in 1880, now celebrating our 125th anniversary
- A highly diverse and early adopter of interdisciplinary research and instruction
 - USC has 19 academic units consisting of 17 professional schools, the Graduate School, and the USC College of Letters, Arts and Sciences
 - Focus on creative solutions with revenue center management
- USC's Schools of Medicine and Pharmacy were the first in Southern California
 - USC physicians serve more than one million patients each year
- Research
 - 2th fastest supercomputing cluster in the country among academic institutions
 - USC ranks 9th among all private universities in the United States in dollar volume of federal research support
 - USC is the only university in the west coast that has two National Science Foundation Engineering Research Centers:
 - ▲ Integrated Media Systems Center (IMSC)
 - ▲ Biomimetic MicroElectronic Systems (BMES) Center

USC Demographics

- As of the 2003-2004 academic year, USC has approximately:
 - 31,000 students
 - ▶ 16% of students represent over 115 countries
 - ▶ The incoming freshman's SAT score average is 1,345 and mean GPA is 3.99
 - 4,300 faculty
 - ▶ More than 200 faculty members have received prestigious academic and professional awards including Nobel Prize laureate George A. Olah and 40 members of the National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine
 - 14,000 staff
 - ▶ Largest private employer in the city of Los Angeles

USC Services

- Email and directory services support over 60,000 users with approximately 1,700,000 daily messages
- 7x24 data centers support over 300 applications across 4,000 machines on various platforms (PC, Mac, Unix, Linux)
- Over 1,100,000 top level web pages per day are viewed on the central web
- On average/day, we have 3,000,000 system probes and 40,000 viruses
- Integrated Library System, which contains approximately 1.8 million titles
- High-end technology infrastructure that supports faculty and staff in excess of 500 research grants and projects
- Videoconferencing and multimedia support (150 classrooms / auditoria)
- University-wide telephony

So What Does that Mean?

- Our technical environment today includes...
 - 46,560 network port connections
 - 28,700 phone lines
 - 1,011 network switches
 - 53 network routers
 - 30 miles of fiber cable
 - 36 miles of underground copper cable
 - 35 miles of underground conduit
 - 730 miles of station (building) cable
 - 7,400 wireless cards
 - 100 terabytes of central disk storage
 - Shared memory large-scale computing – Sun E15K, 72 processors, 288GB
 - Linux cluster – 1,726 nodes

HPCC Today

■ Goal

- Provide common facilities and services for a large cross-section of the University that requires leading edge computational, storage, and networking resources
- Leverage USC central resources with externally funded projects
- Leverage University resources through “condo” arrangements

■ Current Resources

- Linux cluster (1,820 nodes, 5384 CPU/cores, 2GB/second Myrinet)
 - ▲ 120TB shared disk, 18GB-40GB local disk per node
 - ▲ Ranks in top few for academic clusters
- Myrinet switch is 1,820 nodes
- Additional nodes funded by USC research groups
- Sun Core Servers (E15k shared memory)
 - ▲ 72 processors, 288GB memory, 30TB shared disk
- Mass Storage Facilities (Unitree)
 - ▲ 18,000 tape capacity (4.5PB)
 - ▲ 1.1PB

■ **Large Scale Storage**

- Visual History Archives – 20TB (disk), 200TB (tape)
 - ▲ Adding 200TB disk
- Institutional Repository – 20TB
- SCEC – Shake Simulation Runs – 30TB (each run)
- Natural Language Translations – 80TB (ISI)

■ **Grid Solutions**

- SCEC Collaborations
- GRID U (Compute to the Classroom)
- NSF – Identity Management – Shibboleth (NMI)
- Visualization Center – (3 technologies ... Immersion, Tile, Access)

■ **Network Initiatives**

- Pacific Wave (University of Washington, CENIC, USC)
- NLR

HPCC Governance

■ **HPCC Faculty Advisory Group**

- Meets 4-5 times per year to provide DCIO and CTP with guidance
- “Final” decisions made by ISD; usual mode is agreement

■ **Time Allocation**

- No recharge
- Large project reviewed by faculty allocation group (some projects over 500K node hours)
- “Condo” users get dedicated nodes and cost-sharing

■ **Research Leverage**

- Condo
- Cost-sharing
- External funding
- Next generation network

HPCC - Bridging The Gaps

■ CIO Perspective

- **Balancing Funding Requirements – 7/24**
 - ▲ **Priorities – Limited Resources, capital, etc... I need it**
- **Understanding the Investment**
 - ▲ **What does it really cost**
- **Communicating the Value - (high value – small population)**
 - ▲ **Perspective - Administration, faculty, etc.**
- **Shared Infrastructure with diverse needs**
 - ▲ **Open environments vs. Security Needs**

■ Its About Relationship / Partnership

- **Adding Value**
 - ▲ **Colo's**
 - ▲ **Shared investment resources, purchases, people**
 - ▲ **Assisting / Facilitating .. leverage projects.. Grants..**
- **Governance and Communications**
 - ▲ **Faculty Advisory Groups..**
- **Responsiveness**
 - ▲ **Doesn't mean always saying yes...**

HPCC Future Resource Goals

■ **Cluster**

- **Continue to expand cluster (central 2,048 nodes, with condo 3000 + nodes)**
 - ▲ **50-80TF in near term (2 years?)**
- **Leverage Funding / Research Projects**
- **Rank near top for academic clusters**

■ **Share memory systems**

- **Replace E15k with large shared memory commodity for that community**
- **Installing 64GB 16 core systems for CMB group**
 - ▲ **Use as model for larger memory users,**
 - **Cluster at ‘larger memory level’ as well as today’s GB/core model.**