

# ORNL Site Report HUF 2003

Randy Burris  
Storage Systems Manager

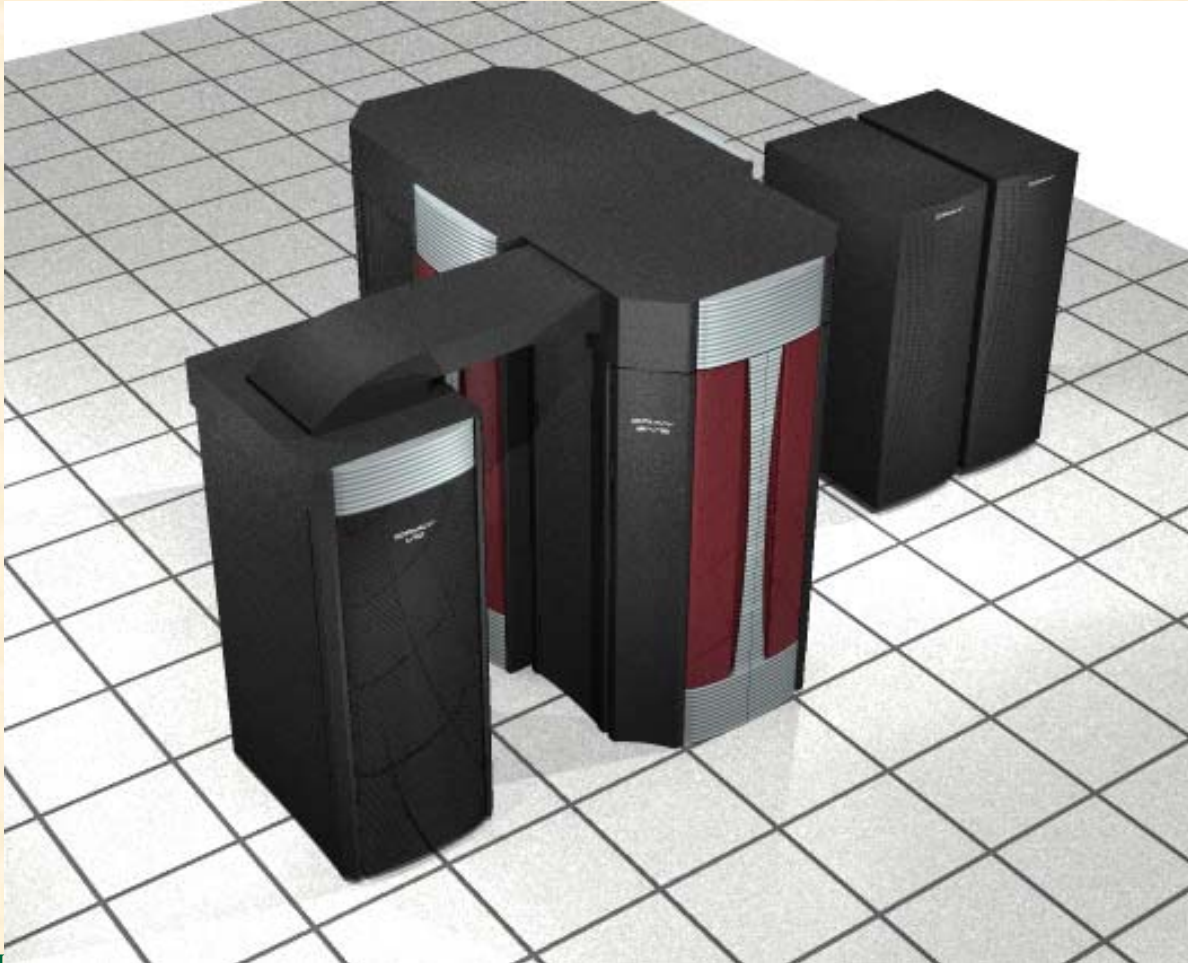
# Topics I'll be covering:

- ❑ Configuration
- ❑ Statistics
- ❑ Problems/solutions
- ❑ Uniqueness
- ❑ Wish list
- ❑ etc

# Configuration – Supercomputers

- ❑ IBM Power-4 SP – 864 processors, 4.5 teraflops
- ❑ IBM Power-3 SP – 720 processors, 1 teraflop
- ❑ Cray X-1 – half-populated node (16 processors) on hand; growing to 3.5 teraflops by September
- ❑ Linux clusters
  - eXtreme TORC – 65 Pentium 4 processors, 129 GF
  - New 256-node SGI Altex – one teraflop – real soon now
  - New Opteron PC visualization cluster – real soon now
- ❑ Two Compaq supercomputers totaling 80 nodes, 500 GF
- ❑ SGI Origin 2000

# The Cray X-1



# HPSS Configuration

## □ Now

- HPSS 4.1.1.4 (blush...)
- Core servers and SFS on one IBM RS/6000 H70
- Movers – one each H70, p660, 43P
- Disks – SSA, RAID Inc. Condors , Compaq StorageWorks
- Tape drives – 18 \* 9840
- Just added two silos (making 4 total)

# HPSS Statistics

- ❑ Over 4 million files
- ❑ Growth around 6 TB/month; max about 1 TB/day
- ❑ Typically move 1-2 TB/day; max over 5 TB/day
- ❑ Stored volume has doubled annually for last six years
- ❑ 143 TB stored as of June 2

# Configuration plans

- Later this fiscal year
  - HPSS 4.5
  - Core servers – p630 (4 GB, 4 processors)
  - Disks – adding FastT-700 (and wishing for others...)
  - Movers – reusing the core-server H70 and adding a p630
  - Tape drives – adding 16 9940Bs
- Will reorganize silos, tapes
  - Two fire zones
  - Segregating the tapes for dual-copy COSs
  - 9840's in one silo, each zone; 9940B's second silo, each zone

# Configuration – how is HPSS being used?

- ❑ hsi is our primary user interface
- ❑ htar will be pushed after we get to HPSS 4.5
- ❑ File transfers:
  - hsi, of course
  - Will support GSI-enabled pftp after we're at HPSS 4.5
- ❑ Testing/developing bulk file transfer mechanisms
  - Hierarchical Resource Manager (Berkeley, SciDAC SDM ISIC)
  - Logistical Networking (University of Tennessee)
  - UDP blasters – comparing, testing, evaluating

# Problems/solutions

- ❑ Implemented File Families, segregating projects to different tape cartridges
- ❑ Retired 3494 robots and 3590 drives
  - Required copying all data from old cartridges
  - Required extensive checking, multiple passes
- ❑ Lost data from single-copy Class of Service
  - Encountered some bad files – media damage
  - Adopted dual-copy default
- ❑ Globus – getting our feet wet.

# Uniqueness

- ❑ New computer room
  - 2 floors, two fire zones, 40,000 square feet
- ❑ Cray X-1
- ❑ Emphasis on bulk data transfers over WAN
- ❑ DFS for user's home directories
  - Will replace with Spinnaker NFS equipment
- ❑ Extensive and diverse test environment
- ❑ SciDAC activities

# SciDAC – Scientific Discovery through Advanced Computation

- ❑ Primary site for running SciDAC Applications
- ❑ Scientific Data Management Integrated Software Infrastructure Center
  - ORNL provides the “place to be” for the ISIC
  - Research and development on optimizing low-level access
- ❑ Terascale Supernova Initiative – impacting
  - storage capacity
  - WAN transfer mechanisms
  - data analysis mechanisms (and how they relate to HPSS)
  - and data rates for visualization
- ❑ Climate – impact on storage capacity, WAN transfers and data rates
- ❑ Earth Sciences Grid II and DOE Science Grid – impact on authentication and middleware interactions
  
- ❑ Net100 – impact on WAN transfers

# Wish List

- ❑ Fast and reliable and cheap bulk wide-area data transfer
  - May actually get only 10% of available bandwidth per stream
  - TCP/IP is not excellent for wide-area bulk data transfers
    - Excessive latency (e.g., 60 ms round-trip time)
    - Congestion
    - Slow error recovery
  - Alternatives
    - Automatic TCP/IP tuning – not fast enough
    - Multiple streams – not fair
    - UDP/IP blasters – not reliable and may not be fast
    - Dedicated line – not cheap
- ❑ More widespread vendor support of DMAPI
- ❑ More bulk-consumer-friendly IDE disk repair mechanisms

# Etcetera