BOEING

Upgrading an SGI Origin system from 8P to 64P to 256P in 275 days

Barry Sharp, Jim Glidewell and Chuck Keagle

The Boeing Company Shared Services Group

Barry.Sharp@boeing.com

Tuesday May 21, 2002 (Lord Mayor's Parlour at 11:30am)

We had an emerging need in Jan 2001

- Fast Track, High Visibility Project
- Initial forecast 372,600 Origin3800 cpu-hrs
  - Equaled an Origin3800 w/43Ps for one year
  - Forecasts could be off by 2x to 3x
- Easy to justify a system upgrade
- Clearly - current Origin2000 w/8P was inadequate

Improvements to CFD programs and other drivers justifying an upgrade

- Many codes becoming ***parallel savvy***
- Code developers required better turn-around
- Scaling evaluations required more processors
- Production codes required better turn-around
- Origin2000 reliability had been excellent
- NASA Ames reporting excellent scaling results
- Boeing R&D +ive experience at NASA Ames

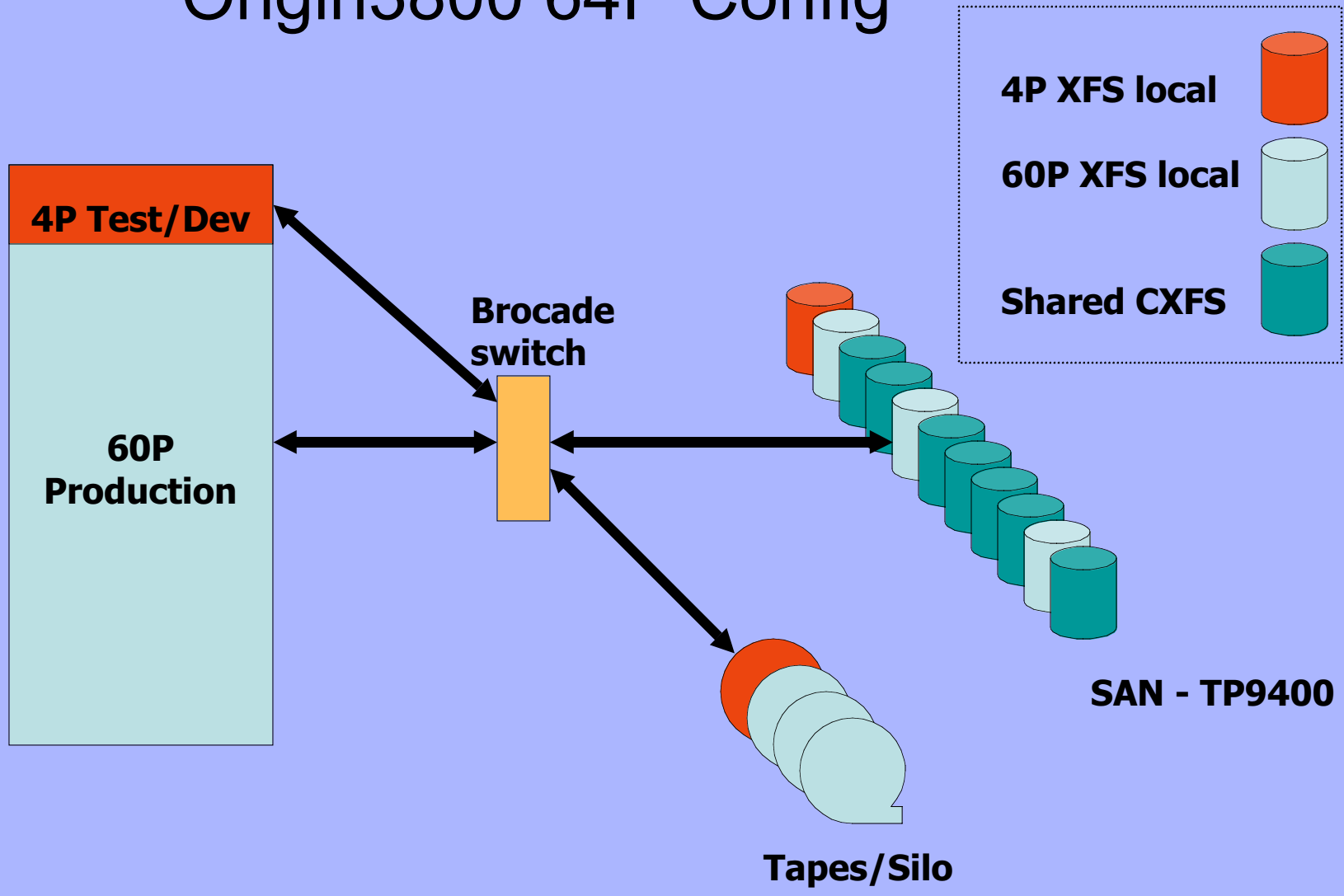
BOEING

To help design a Boeing Near Sonic Cruiser

The Origin3800 64P

- **Late 2000, Early 2001 activities/issues**
  - Negotiations with SGI
  - Procurement and Approval activities
  - System configuration complicated by
    - Partitioning 60P + 4P - ***SGI agreed to support***
    - CXFS req^{d} for SAN experience and the future
    - DMF w/CXFS uncertainties
    - Introduction of new TP9400 FC RAID

Origin3800 64P Config



We had a great deal to learn

- CXFS
- CXFS w/DMF managed file systems
- IRIX cpuset feature (w/64P we needed this)
- TP9400 RAID subsystem
- SAN w/Brocade switch
- NQE using cpusets ???
- Onsite classes were setup
  - SGI tech specialists identified for classes
  - Many hands-on config examples conducted

After Origin3800 64P placed into production

- CXFS and CXFS w/DMF worked as advertised
- NQE
  - Cpuset-s not supported
  - Kernel & Job limits w/NQE very problematic
  - SGI finally said no Bugfixes nor Support
  - Decision made to transition NQE to PBS Pro ASAP
- TP9400 RAID subsystem performed well
- Process memory allocations a mystery (pmem ?)
- Engineers happy but needed even better turn-around

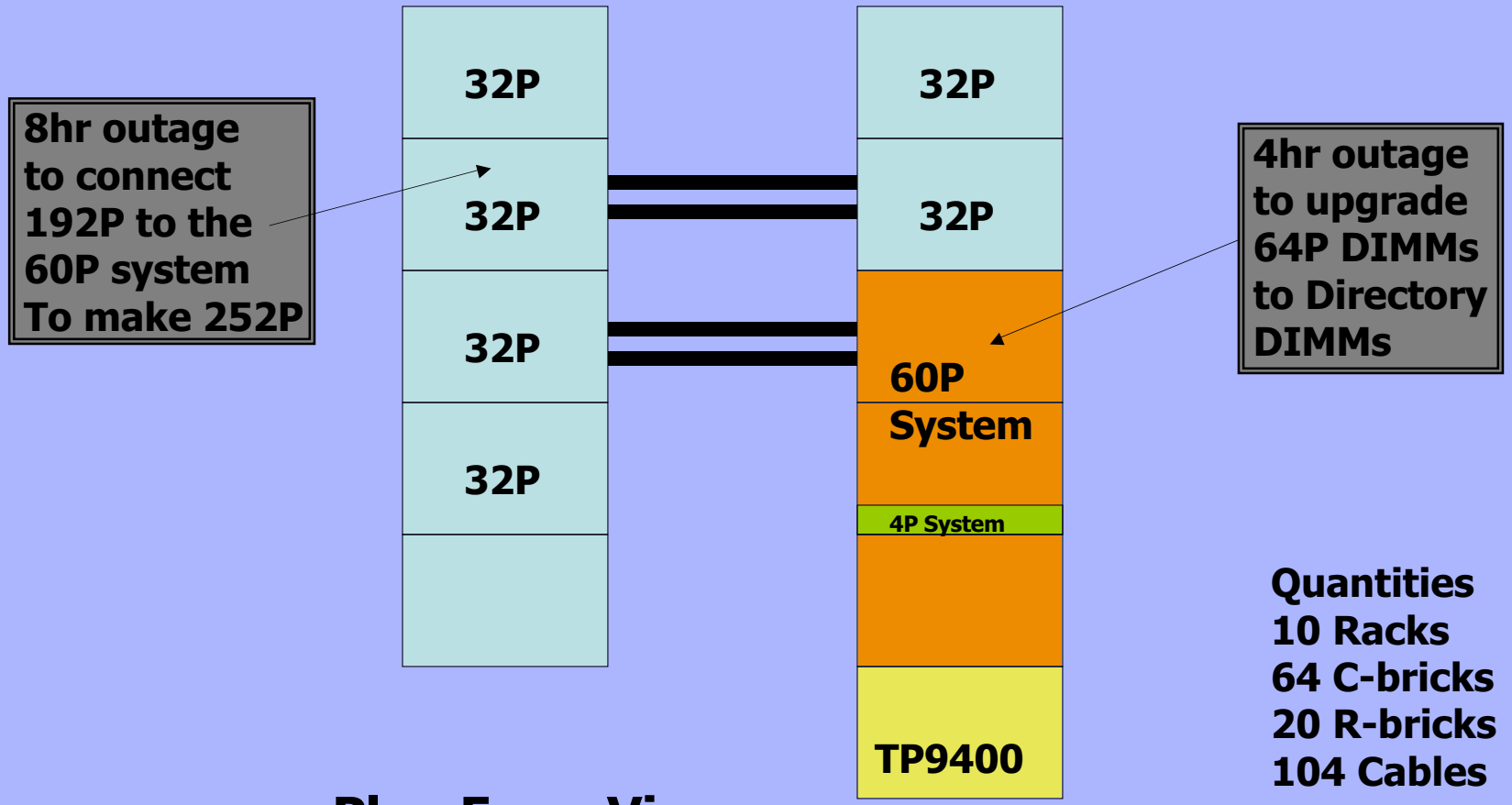
Surprise, surprise -- Geeesh it's too small

- Engineers demanding more than 32 CPUs/job
- New workload forecasts arrived (remember the 2x to 3x)
- Upgrade to 128P, 192P, 256P, … or 2nd machine
- 64P system non-directory DIMMs unusable above 128P
- SGI's support for 124P+4P, 188P+4P, 252P+4P ?
- Minimize interruption to 64P Production during upgrade
- Footprint surprise for a 256P system

Origin3800 w/256P and w/384 Gbyte approved

- SGI agreed to support the 252P + 4P partitioning
- Found home(s) for the 64P's 64Gbyte DIMMs
- Two 128P systems dismissed… much debate
- No mixing 64P/400MHz with 192P/500MHz
- 4x TP9400 I/O bandwidth + Doubled Capacity
- Second Brocade switch needed
- 3rd P-brick reqd for 3 partitions (4+60+192) for transition
- Planned for minimizing interruptions during upgrade
- Twin Towers September 11th ~ on hold for about 3 weeks

The 256P/384GB System



Plan Form View

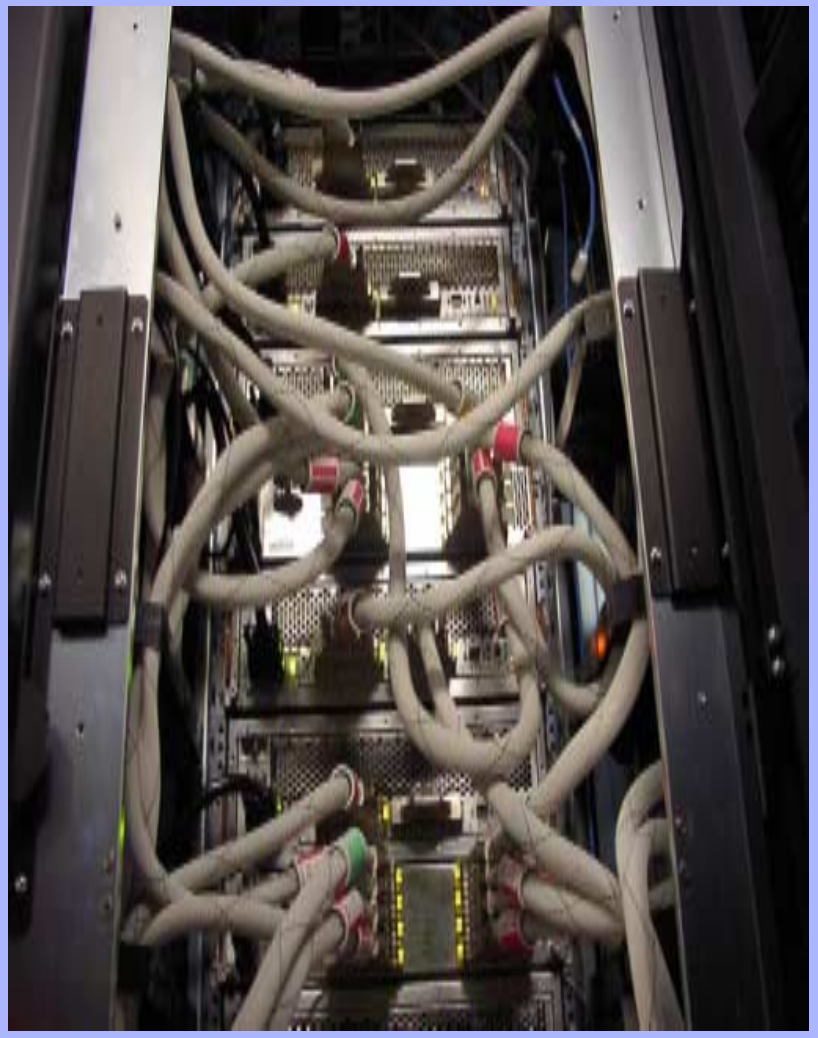
Quantities
10 Racks
64 C-bricks
20 R-bricks
104 Cables

BOEING

Seismic Platforms

BOEING

NUMAflex Cabling



Early 256P System Problems

- **Prod Oct 29th to Nov 27th (1st five weeks)**
  - 10 unscheduled interruptions
    - One was a 5 hour outage during prime-time
  - A few DIMMs failed
  - CPU #81 problems ~ resolve with ***cable wiggle*** test
  - Visibility of cpuset use was lacking
  - Brocade switch boxes getting too hot
  - Not making our 4hr to recovery service level
    - SGI tried… now establishing new process

Current Problems

- Hardware scache problem occurs
  - System begins silent slow death spiral
  - ***Is unnoticed unless someone watching***
  - System daemons start cpu looping (some at 300%)
  - /var/adm/SYSLOG stops updating
  - PBS daemon stalls - Input jobs not scheduled
  - Other cpu-looping daemons not killable or restartable
  - Executing PBS batch work continues
  - System crashes ~ after 5 hours in one case
  - Can't get system dump

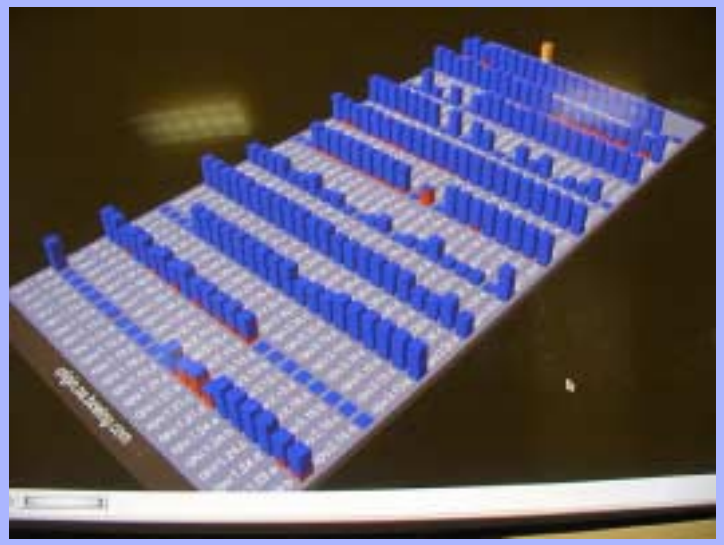
Current Problems

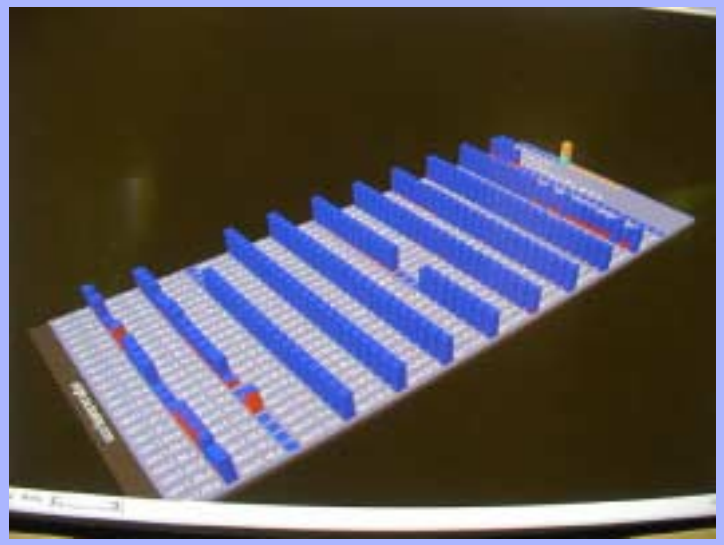
- Poor interactive response when a few processes start performing intensive i/o to a single file system
  - Response times from 6 secs to several mins
  - Batch work appears to be unaffected
  - Overall system i/o rates at 300 to 400 MB/sec
  - Suspect kernel disk i/o sort algorithm laboring

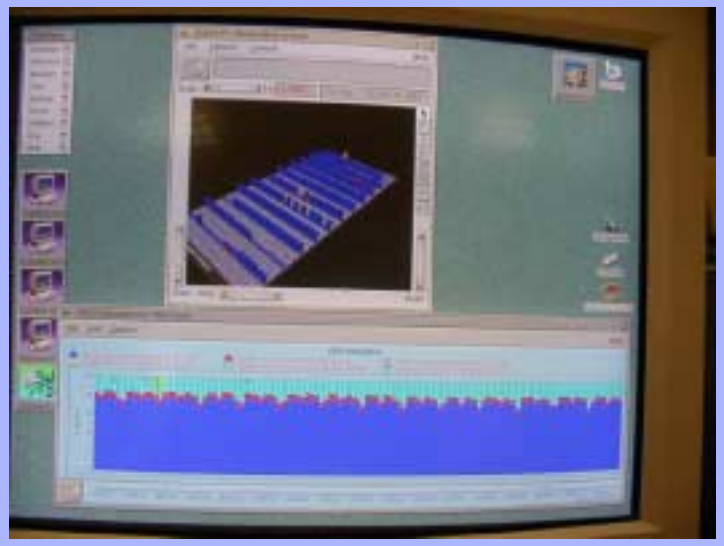
The good news
An Engineer's Perspective

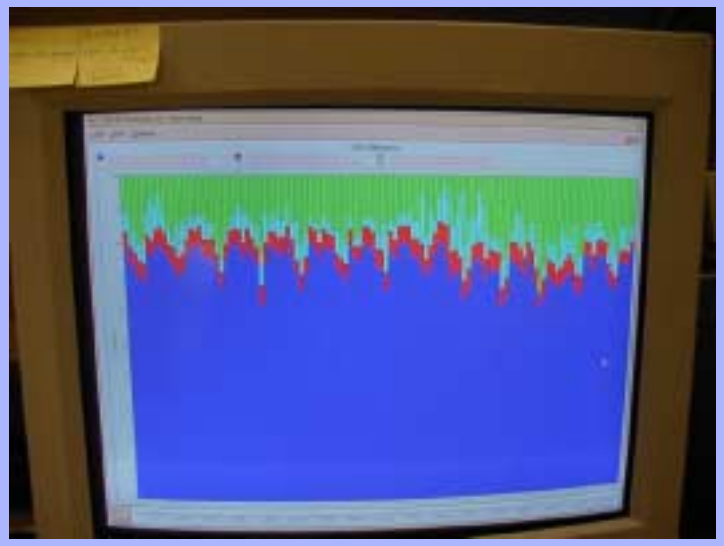
- 4000 steps on a 13 million point CFD problem
  - Impossible to do in timely manner one year ago
  - Can now be done in approx 7 hours wall time
  - Can be started and completed during a workday
- Origin3800/256P meeting current engineering goals

Performance CoPilot Samples







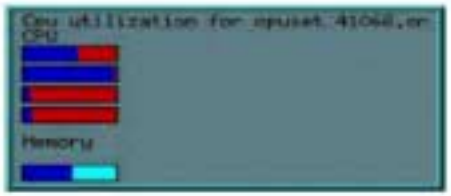


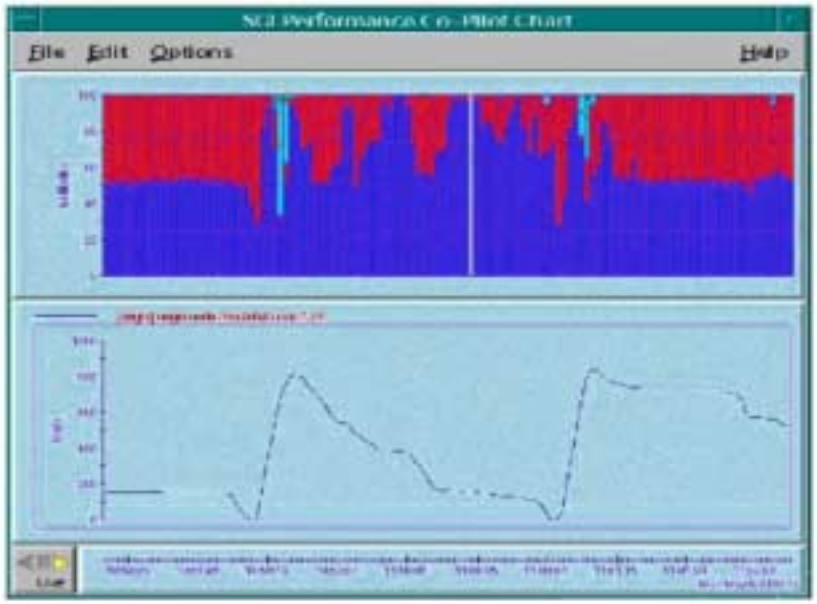
BOEING

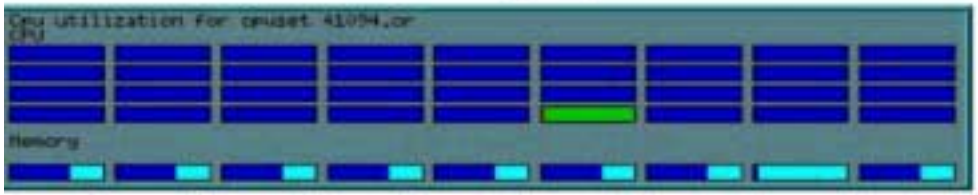
PBS Batch Monitoring

Realtime pbs queue status on the web.

Cpuset aware Performance Co-Pilot tools (pmgcpuset)







BOEING

Yes, we really did do it

January 27th to October 29th (275 Days)