Evaluating RAID in the Real World

Andy Kowalski
SURA/Jefferson Lab
Jefferson Lab

• Who are we?
  – SURA/DOE

• What do we do?
  – High Energy Nuclear Physics
  – Operate a 4 GeV continuous electron beam accelerator

• Research
  – quark and gluon
Jefferson Lab
Environment

- Three experimental halls

- Data rates
  - 1 TB/day, 1-100 GB/day, 1-100 GB/day
  - total I/O rate of 3TB/day with batch farm

- Storage Capacities
  - STK SILO with SD3 (Redwood) tape drives
  - Disk Space - 2TB of RAID
Environment cont.

- Fast Ethernet and Fibre Channel
- Batch Farm 350+ SPECint95
  - 6 Dual Sun Ultra2
  - 5 Dual IBM RS6000
  - 11 Dual Pentium II
- Analysis Farm 200+ SPECint95
- Load Sharing Facility (LSF)
- Open Storage Manager (OSM)
Data Path

Data Server

- Sun 4000E
- Sun 3000E
- Sun 2000E

Fibre Channel

Tape Drives

Redwood Tape Drives

Tape

RAID 125GB

RAID 125GB

Fast Ethernet Switch

NFS RAID for “Work” 1TB

Central Batch and Interactive CPU Farms

AIX, HP-UX, Solaris, and Linux

HallA and HallC DAQ

HallA DAQ FY98

HallB DAQ
Why Raid

- High capacities for tape staging and work
  - storage for lots of 2GB files
  - high transfer rate
  - stream to tape at 10MB/sec
  - simultaneous access
- Data Integrity
- Disk management
Considerations

• Access patterns and effects on the data rates
  – simultaneous tape and farm node copies
  – effects on tape transfer rates must be minimal

• Just a Bunch of Disks (JBOD)
  – inexpensive
  – requires software for striping
  – hard to manage
Considerations

- Hardware vs. Software RAID
  - performance
  - dealing with multiple accesses
- Which RAID level?
  - RAID 0 for HallB DAQ
    - needs to be fast
  - RAID 5 for work areas and staging
    - needs to be available
    - needs to be fast
RAID System Evaluations

• Two Procurements
  – direct attached
  – NFS

• Why we wanted to do on site evaluations
  – understand vendor’s numbers and units of measure
  – see how it would work in our environment

• Real comparisons (not just glossies)
Analyze the Data Path

- Determine the uses and locations for RAID
  - tape staging
  - work areas
- Measure data rate for each segment
- Make baseline measurements without RAID
  - compare with the introduction of RAID
Test Setup

Data Server

NFS RAID being tested

Fast Ethernet Switch

Central Batch CPU Farm
Host Attached Tests Performed

- RAID (tests were also run in reverse)
  - Memory to Raid
  - Memory to Raid (3 simultaneously)
  - Memory to Raid and Raid to Memory (simultaneously)
  - Raid to Tape
  - Raid to Tape and Memory to Raid
  - Tape to Raid and Raid to Network
NFS Tests Performed

- NFS RAID (tests were also run in reverse)
  - Memory to Raid
  - Memory to Raid (3 simultaneously)
  - Memory to Raid (2 simultaneously) and Raid to Memory (2 simultaneously)
Procurement

• Host Attached
  – Limited competitive purchase
  – Limited budget
  – Limited price range

• NFS
  – Limited to two vendors for compatibility
Logistics

- **Schedule**
  - 6 vendors for direct attached RAID
  - 2 vendors for NFS RAID
- Vendors were provided tests in advance
- Vendors setup time
- 4 hour test time
NFS RAID Results

![Bar Chart]

- **3 Writes**: 4 MB/sec
- **3 Reads**: 8 MB/sec
- **1 Write**: 5 MB/sec
- **1 Read**: 3 MB/sec
- **2 Writes and 2 Reads**: 6 MB/sec
Conclusions

- Ads do not tell the whole story
  - vendors do not tell you the bad news
- Vendor’s performance numbers are skewed
- Tricks
  - using the outer tracks
  - measuring rates to and from cache
  - turning off redundancies
Conclusions cont.

• On Site Evaluations
  – we learned a lot about RAID
  – well worth the time and effort