

Storage Task Force

Intermediate pre report

History

- GridKa Technical advisory board needs storage numbers: Assemble a team of experts. 04/05
- At HEPiX 05/05
 - many sites indicate similar question(s)
 - What hardware needed for which profile
 - HEPiX community, GDB chair and LCG project leader agree on task force.
- First (telephone) meeting on 17/6/05
 - Since then 1 face to face, several phone conf., numerous emails

Mandate

- Examine the current LHC experiment computing models.
- Attempt to determine the data volumes, access patterns and required data security for the various classes of data, as a function of Tier and of time.
- Consider the current storage technologies, their prices in various geographical regions and their suitability for various classes of data storage.
- Attempt to map the required storage capacities to suitable technologies.
- Formulate a plan to implement the required storage in a timely fashion.
- Report to GDB and/or HEPiX

Members

Roger Jones (convener, ATLAS)

- Tech experts

Martin Gasthuber (DESY)

Andrei Maslennikov (CASPUR)

Helge Meinhard (CERN)

Andrew Sansum (RAL)

Jos van Wezel (FZK)

- Experiment experts

Peter Malzacher (Alice)

Vincenzo Vagnoni (LHCb)

nn (CMS)

- Report due 2nd wk October at HEPiX (SLAC)

Methods used

- Define hardware solutions
 - storage block with certain capacities and capabilities
- Perform simple trend analysis
 - costs, densities, CPU v.s. IO throughput
- Follow storage/data path in computing models
 - ATLAS, CMS and Alice used at the moment
 - assume input rate is fixed (scaling?)
 - estimate inter T1 data flow
 - estimate data flow to T2
 - Attempt to estimate file and array/tape contentions (hard!)
- Define storage classes
 - reliability, throughput (=> costs function)
- Using CERN tender as the basis for an example set of requirements

Trying to deliver

- Type of storage fitted to the specific req.
 - access patterns
 - applications involved
 - data classes
- Amount of storage at time t_0 to t_{+4} years
 - what is needed when
 - growth of data sets is ??
- (Non exhaustive) list of hardware
 - via web site (feed with recent talks)
 - experiences
 - need maintenance
- Determination list for disk storage at T1 and T2
 - IO rates via connections (ether, scsi, i-band)
 - Availability types (RAID)
 - Management, maintenance and replacement costs

For T1 only

- Tape access and throughput specs.
 - involves disk storage for caching
 - relates to the amount of disk (cache/data ratio)
- CMS and ATLAS seem to have different tape access predictions.
- Tape contention for
 - raw data
 - reconstruction
 - Atlas 1500 MB/s in 2008, 4000 MB/s in 2010 all T1's
 - T2 access

Conclusions

- Disk size is leading factor. IO throughput is of minor importance (observed 1 MB/s / cpu) but rates at production are not known.
- Cache to data ratio is not known.
- Probably need yearly assessment of hardware. (especially important for those that buy infrequently)
- Experiment models differ on several points: more analysis needed.
- Disk server scaling difficult because network to disk ratio is not useful.
- Analysis access pattern is not well known but will hit databases most – should be revisited in a year.
- Non-posix access requires more CPU