

# Storage I/O, LAN Protocols and Requirements and evolution of storage

Giacinto Donvito (INFN-Bari)  
Wahid Bhimji (University of Edinburgh)

# Storage I/O

- Accessing data from disk with increasing volumes and fewer spindles as we get larger and larger disks we have fewer spindles
  - more cores trying to read: the IO of the application and the sw performances get even hotter topics
    - Could the “analysis train” help here?
  - Could the use of a distributed disk farm configuration (HADOOP like), help to mitigate (in the short term) this problem?
- Spinning magnetic disks will become extinct in some years (5-10), SSDs or something similar will take over. Will probably solve some performance problems but how is the price/TB evolving? Reliability?

# Storage I/O

- Addressing scalability - how to scale to bigger files, exabyte storage, more files...
  - Also other scientific communities are already working to address the problem of exabyte computing. Do we need to start doing the same?
- There is the need to provide a representative benchmark for each of the LHC experiments:
  - Not only in terms of theoretical bandwidth required from each job (at the moment it should be around 5MB/s for Atlas&CMS)
  - But also a “easy-to-run” application that could be used to test storage infrastructure
    - This could be used, for examples by the HEPIX Storage Working Group to provide useful numbers about the performance of each storage system

# LAN Protocols

- In the HEP community there are several approach to the problem:
  - Standard posix, xrootd, dcap, http, etc
  - In the first step of the work, the analysis application is supported by experiments framework, but standard posix could be of help for the end users during the last step of the analysis (~tier3 activities)
    - Using standard posix help also the applications to exploit easily the kernel “VFS caching”
    - but we need the capability of throttling the I/O requests also in case of standard posix interface
- LAN bandwidth and the network in general is a key point here:
  - It is important to study and reduce/remove all bottlenecks
  - The trend is to have always more core in the same motherboard:
    - We surely have to use 10Gbit/s network on the WN very soon

# Requirements and evolution of storage

- Weak error-resilience of storage services (to failures and to legitimate users overloads)
  - This will impact also over the need of man power for managing the site
  - The variety of issues caused by increasing disk sizes, such as
    - a) any disk server interventions results in larger numbers of files being inaccessible to the experiment
    - b) failure of a disk server could result in larger data loss
    - c) reduction in performance (network bandwidth, job slots, ...)
  - We need to have a solution for the typical incident when a disk server or a tape are temporarily (or permanently ) off-line
- Difficult to understand when and how storage issues affect jobs (low efficiency to complete, job failures, etc) due to the lack of relevant global monitoring information.