Medium-Term Evolution: Storage (Site Positions)

Sam Skipsey
(with input from UK T2s, RAL [Brian Davies, Alastair Dewhurst], France [Catherine Biscarat], Netherlands [Jeff Templon], Italy [Alessandra Doria], CIEMAT [Antonio Delgado])

WLCG Workshop 1 Feb 2016

Itinerary

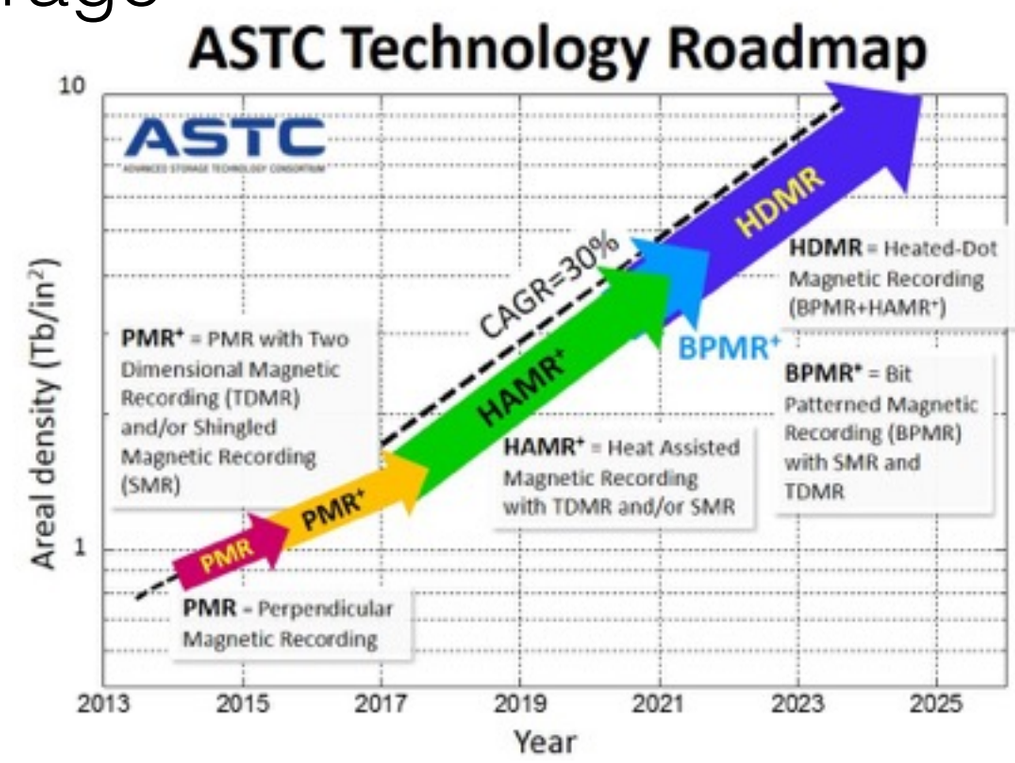
- Technology trends
- Political/Economic trends
- Collated site feedback.

A note

- The allowed scope for “medium term” is the 10 years before the HL-LHC.
- Most IT forecasters caution against making predictions more than 5 years in the future.
- Both the iPhone and Android are <10 years old
  - disrupted the entire phone industry
  - killed Symbian, Palm, almost killed RIM

Trends: Hardware

- Hard disk **density** scaling still exponential
  - **Performance** scaling much slower
  - RAID6 turn over point is in next 5 years
- SSDs won't catch HDDs for mass storage
  - but NVRAM caching layers?
- New experimental techs
  - Seagate Kinetic etc



Trends: Software

- Modern mass storage trends to (chunk) distributed, parallel object storage.
- HDFS, Ceph, DDN WOS et al.
- Improved operational efficiency by distributing files over servers, allowing automatic error repair, dynamic load balancing etc.
- Most modern systems support erasure-coding style resilience, not just dumb replication.
- “hash-based object location” (no metadata bottleneck)

Trends:PoliticoEconomic

- Funding is not increasing (it is decreasing)
- Emphasis on collaboration + consolidation
  - not just serving WLCG
  - New Exciting Science is also out there.
- Some? funders are “zeitgeist aware”
  - pressure to use Cloud services (or at least be seen to be looking at them)
  - “efficient manpower”

Storage Complexity

- Storage is currently hard to maintain.
  - “Many protocols”, “complex systems” (*not* number of machines!)
- Funding decline means less manpower in future per site.
- Request: simplify storage where possible.
  - “Killing SRM” is a good first step.
  - Getting rid of some protocols is next.
  - (Supporting modern filesystem backends would also be good.)

Storageless sites?

- Is there a minimum supportable storage per site?
- UK plans: “storageless T2s” below certain threshold.
- Experiment “cache-like”, “networking leveraging” data models imply similar idea.
- Concerns from several that this approach has not been sufficiently tested or explored.
- Do we know how to make networks work for this?
    - UK evidence implies network is the limit for concurrency.

Storageless sites (2)?

- Storage is currently “what HEP is good at” wrt other communities.
- By decreasing storage at sites, do we make them less interesting/ useful, remove expertise?
- Funding for sites which would become storageless needs to be planned *sufficiently* in advance.
  - Procurements take time.
  - Disk pools last 5+ years!
- Pledges and so forth need properly adjusted to reflect new zeitgeist.

Storage as caches?

- A true cache should be able to incur a miss without breaking the user code. (And also cache on misses.)
- Can experiments cope with “catalog misses” at T2 sites which are cache?
- (If sites were caches, would we need to have any resilience at all in that storage? Could we reduce operational effort needed for storage?)

(HTTP) Federation

- Italy are experimenting with a “remote pool” approach (single endpoint sees geographically distributed pool storage nodes).
- More generally, lot of support explicitly for HTTP/WebDAV federation.
  - (No one actually mentioned xrootd federations!)
- Possibility to reduce operational overheads, as “soft” version of caching.

Working with nonWLCG

- Need to support non-WLCG communities
  - These people don't care about xrootd, or even gridftp.
  - Need to support 'standard' protocols (http?, nfs4.1?, s3?)
  - And need to support 'standard' (Federated) Authentication mechanisms.
    - (Grid certs are *not standard,* although X509 can be.)
    - Kerberos? Shibboleth?

Future Growth

- Funding is decreasing.
- HDD capacity is growing.
- But not as fast as post-Run 2 Experiment requirements.
- Experiments need to plan to cope with having much less storage than they would like.
- (and having to share it with other communities!)

Conclusions

- Currently WLCG storage is Complex, Esoteric.
- Sites are pressured to support more diverse communities, with less manpower and funding.
- WLCG Storage should aim to be Simple, Familiar.