
CERN

Storage evolution at CERN

Alberto Pace, alberto.pace@cern.ch

Roles Storage Services

- Three main roles
  - Storage (store the data)
  - Distribution (ensure that data is accessible)
  - Preservation (ensure that data is not lost)

Size in PB + performance

Availability

Reliability

“Why” data management ?

- Data Management solves the following problems
  - Data reliability
  - Access control
  - Data distribution
  - Data archives, history, long term preservation
  - In general:
    - Empower the implementation of a workflow for data processing

CERN Computing Infrastructure

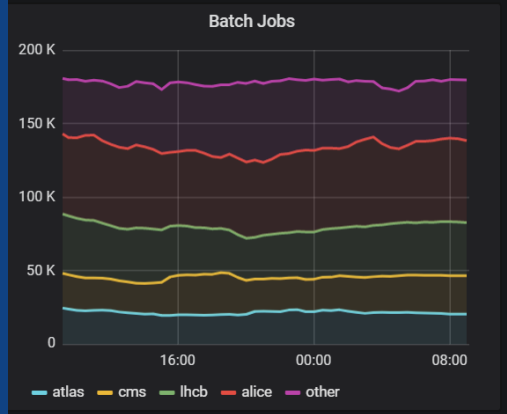
CPUs Network Databases Storage Infrastructure

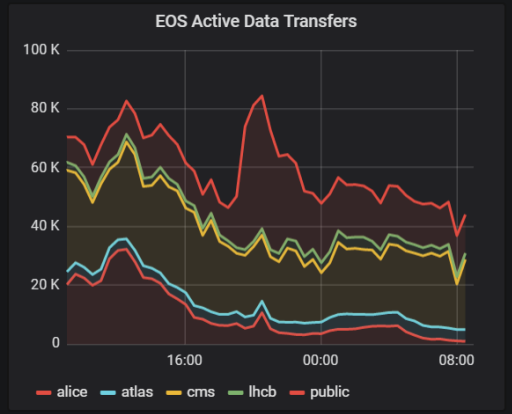
CERN Computing Infrastructure

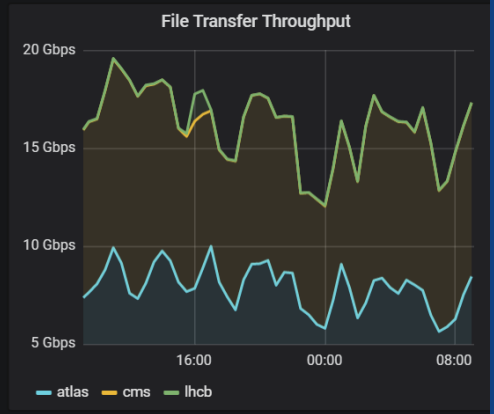
January 2019

Servers (Meyrin)	Cores (Meyrin)	Disks (Meyrin)	Tape Drives	Routers	Star Points
11.5 K	174.3 K	61.9 K	104	251	697

Servers (Wigner)	Cores (Wigner)	Disks (Wigner)	Tape Cartridges	Switches	Wifi Points
3.5 K	56.0 K	29.7 K	33.0 K	4.1 K	912





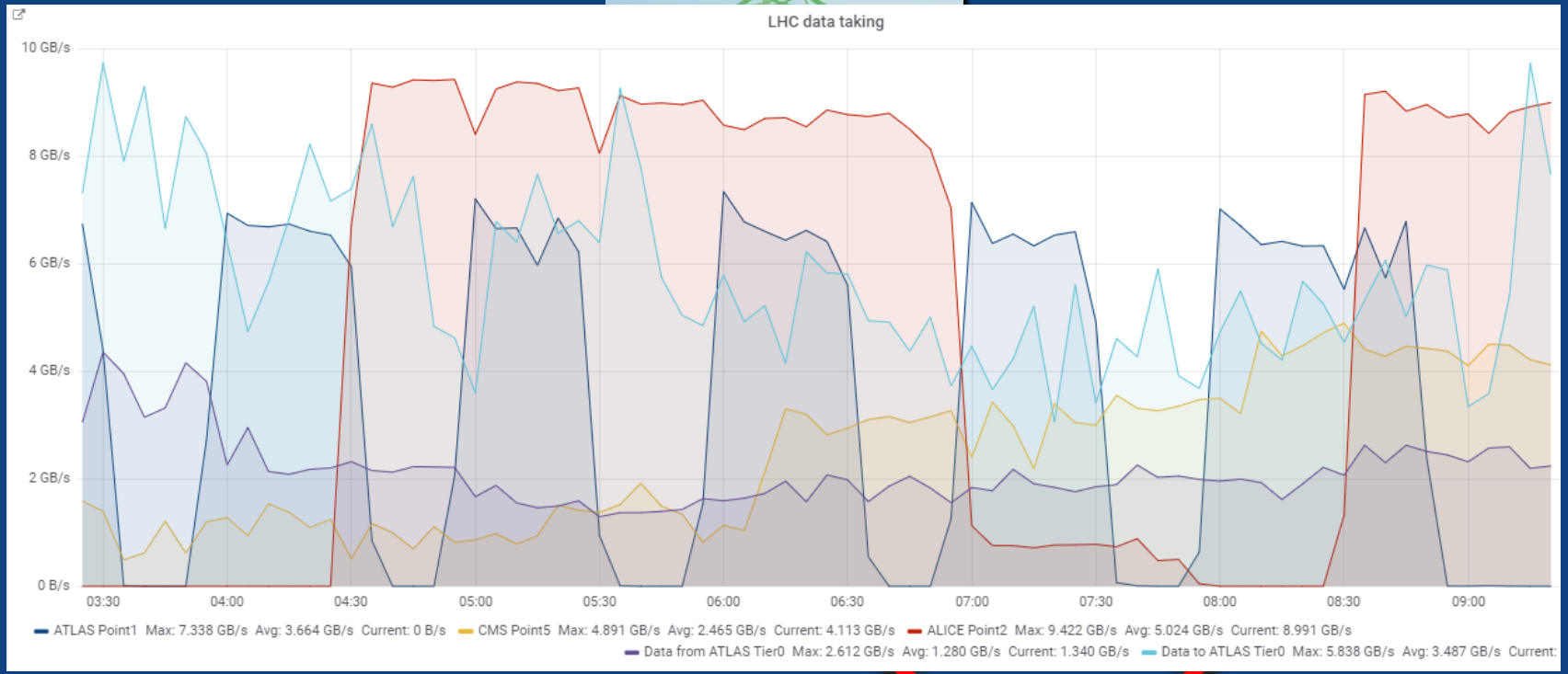


CPUs Network Databases Storage Infrastructure

http://monit-grafana-open.cern.ch/d/000000884/it-overview?orgId=16

CERN Computing Infrastructure

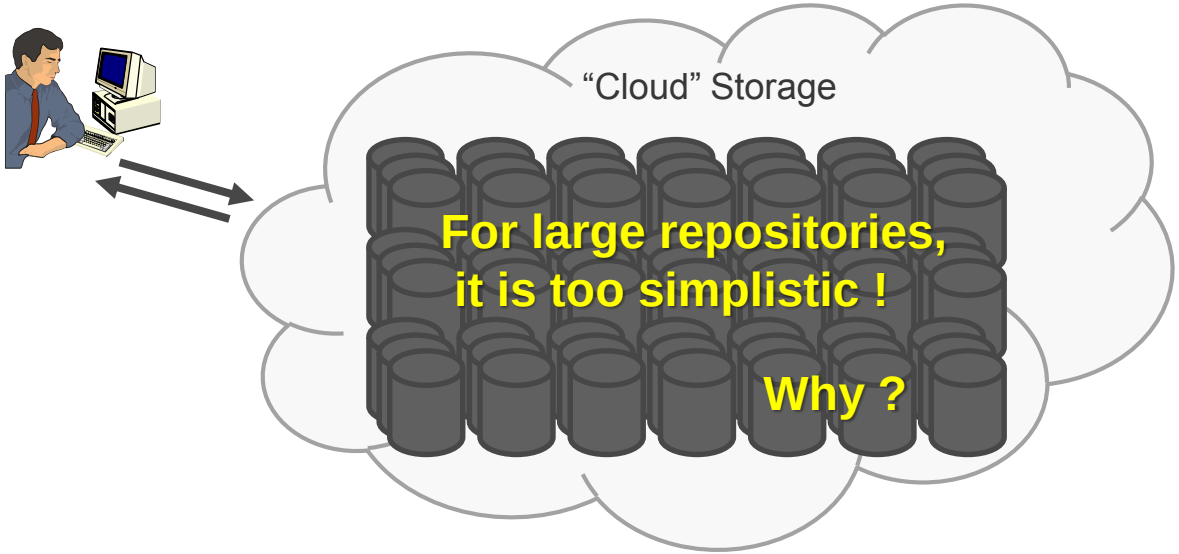
Tue Nov 27th, 2018 at 11:00



CPUs Network Databases Storage Infrastructure

Can we make it simple ?

- A simple storage model: all data into the same container
  - Uniform, simple, easy to manage, no need to move data
  - Can provide sufficient level of performance and reliability

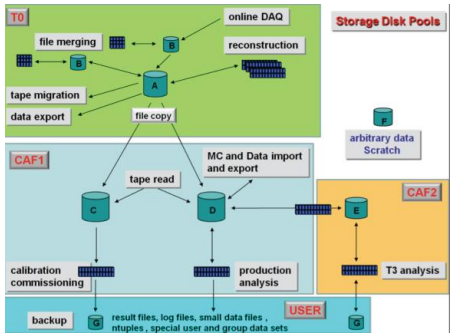


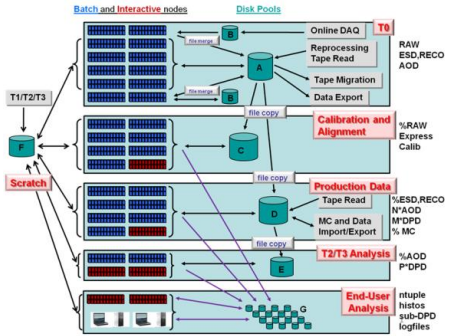
Why multiple pools and quality ?

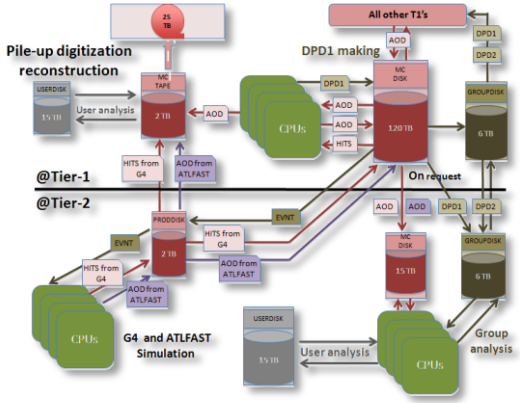
- Derived data used for analysis and accessed by thousands of nodes
  - Need high performance, Low cost, minimal reliability (derived data can be recalculated)
- Raw data that need to be analyzed
  - Need high performance, High reliability, can be expensive (small sizes)
- Raw data that has been analyzed and archived
  - Must be low cost (huge volumes), High reliability (must be preserved), performance not necessary

So, … what is data management ?

- Examples from LHC experiment data models



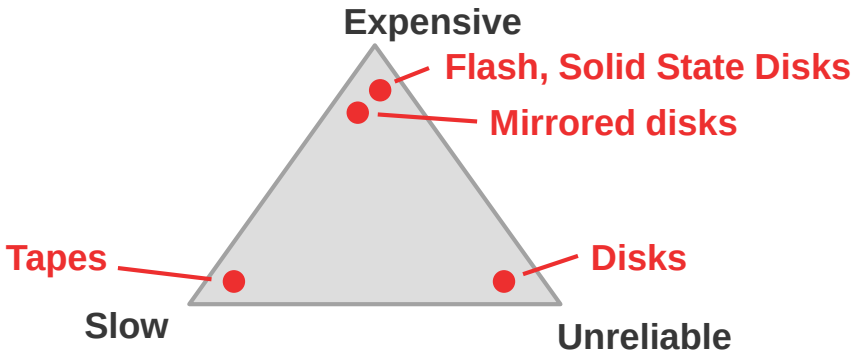




- Two building blocks to empower data processing
  - Data pools with different quality of services
  - Tools for data transfer between pools

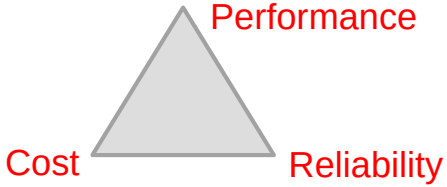
Data pools

- Different quality of services
  - Three parameters: (Performance, Reliability, Cost)
  - You can have two but not three

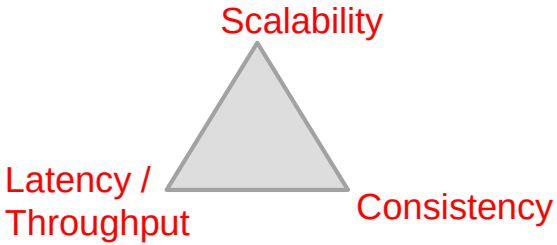


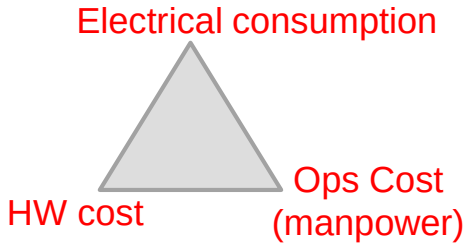
But the balance is not as simple

- Many ways to split (performance, reliability, cost)



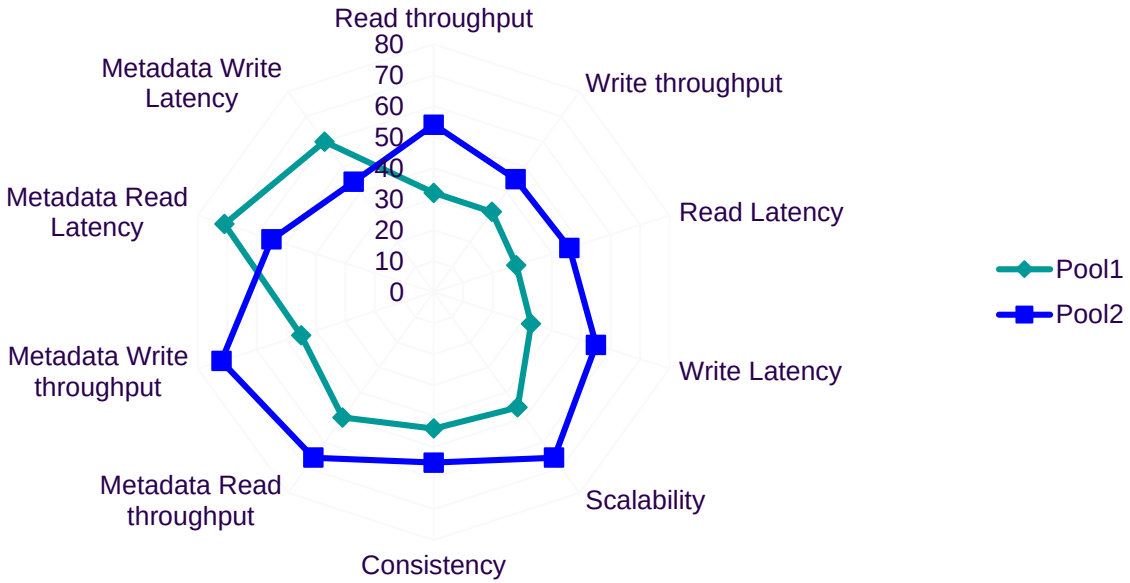
- Performance has many sub-parameters
- Cost has many sub-parameters
- Reliability has many sub-parameters





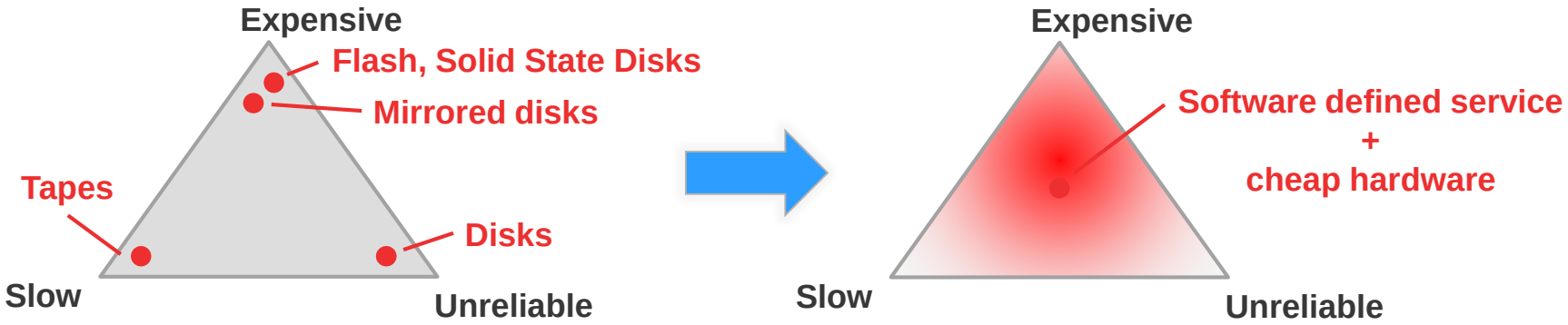
And reality is complicated

- Key requirements: Simple, Scalable, Consistent, Reliable, Available, Manageable, Flexible, Performing, Cheap, Secure.
- Aiming for “à la carte” services (storage pools) with on-demand “quality of service”



Where are we heading ?

- Software solutions + Cheap hardware



Present Strategy, Future Evolution

- Software
  - Software is the most strategic component
  - When you are 'big', using proprietary software is extremely risky
  - It is important that software has a fixed-cost only
- Hardware
  - If the "software" problem is correctly handled, the Hardware + Energy is where variable-costs are concentrated
- Manpower cost
  - Ensure that the 'marginal' cost is a small as possible, maximise automation
- With this approach …
  - the cost of adding a PB of storage is limited to the cost of a PB of HW
  - the cost of operating an additional PB of storage is limited to the cost of the required energy and hardware amortisation

Software

- For the most strategic component, shortcuts are possible but risky
- Example of a heading-for-a-disaster strategy:
  1. Look for the best commercial software available …
  2. Negotiate an outstanding discount, which includes unlimited usage for xx years …
     - Easily done when you are a 'big' customer. You can even get it for free.
  3. Deploy rapidly, grow rapidly, … for xx years.
  4. Pay back all your past savings (and more) at the end of the xx years when you will attempt to renegotiate the contract …
- Does it make sense ? Yes, if you have implemented a clear and tested exit strategy from the beginning

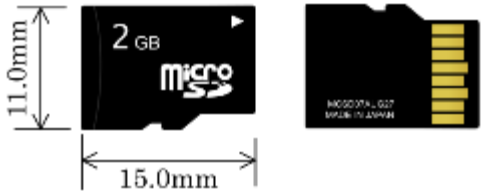
Software strategy

- Three safe scenarios for successful software strategy:
  - Use only commercial software that implements well understood functionalities on well established standard interfaces. There must be implementations from multiple independent vendors with demonstrated interoperability.
    - License cost should be fixed (volume and usage independent) and should not expire.
    - Must have the perpetual right to continue to use the 'old' software in case we would not need or accept or afford to buy renewed version of the software
  - Use Open Source software that has no license cost associated. Fund the necessary software development costs through separated software maintenance or development contracts.
  - Develop core software components ourselves. In open source.
- All three approaches are successfully being applied for the storage service strategy at CERN

Hardware

- In the year 2000, all CERN data (from the previous accelerator - LEP) were filling the datacentre (100 TB)
- Today, all this data can be stored in a drawer of my office
- Will I be able to store all current CERN data in my drawer in 10 years ?

Important digression



- a MicroSD card has a volume of V_{SD} = 15 x 11 x 0.8 = 132 mm^3
  - Available with 512 GB or (soon) 1 TB size
- a 3.5" HDD is V_{HDD} = 101 x 146 x 25.4 = 374'548.4 mm^3
- You can pack many microsd cards in the volume of one hard disk. What storage would you have ?
  - V_{HDD} / V_{SD} = 2837 cards. Capacity = 1.4 PB or (soon) 2.8 PB.
  - 100 PB would require 35 HDD, which fit in my drawer.
  - 100 PB can already fit my drawer **today** using microsd cards

- Will it be slow ? Unreliable ?
  - With striping and erasure encoding you can expect these new storage devices to be arbitrarily reliable (unbreakable) and arbitrarily fast: Always matching the performance of the external interface (Eg: SATA 6 GB/s)
- Media Cost ?
  - Today 250 - 350 K$/PB using microsd. 20 - 30 K$/PB using HDD. 5 - 10 K$/PB using Tapes.
  - So the only question left is :
    - in 10 years, will flash memory match HDD cost ? Will it match tape cost ?
- Intrinsic advantage
  - No power consumption when idle
  - Significant higher performance and reliability

Strategy - Conclusion

- LHC next physics run is expected to deliver 10x today data rates and requires 10x data volumes.
- Must keep fixed cost for software.
  - No license cost proportional to data volumes, or number of nodes, or cores, or disk, or data transferred.
  - (this is why CERN has a Storage group)
- Maximise economy of scale on hardware
  - For storage, this means minimize the cost per PB
  - Many vendors are heavily investing in flash memory to deliver extremely fast storage product that outperform the existing ones at higher cost (bad !)
  - However, that there is a market for low cost, high capacity, flash storage
    - Reliability and performance can be obtained with software
    - Current strategy is to seek for the cheapest possible storage media.

www.cern.ch