Data architecture challenges

for CERN and the High Energy Physics community in the coming years

Alberto Pace

Head, Data and Storage Services

CERN, Geneva, Switzerland

Storage infrastructure at CERN

- Physics data and File services for the physics community
  - More than 2’000 servers, 80’000 disks,
  - 2.2 billion files, 40 PB of online storage
  - 100 PB of offline storage on > 50’000 tape cartridges, 150 tape dtives
  - Sustaining write rates of 6 GB/s
- Important grow rates
  - Doubled between 2012 and 2013
  - Expected to double every 18 months

The mission of data management

- We are the repository of the world data of High Energy Physics
- Mandate:
  - Store
  - Preserve
  - Distribute (i.e. make accessible)

DSS

Challenges ahead

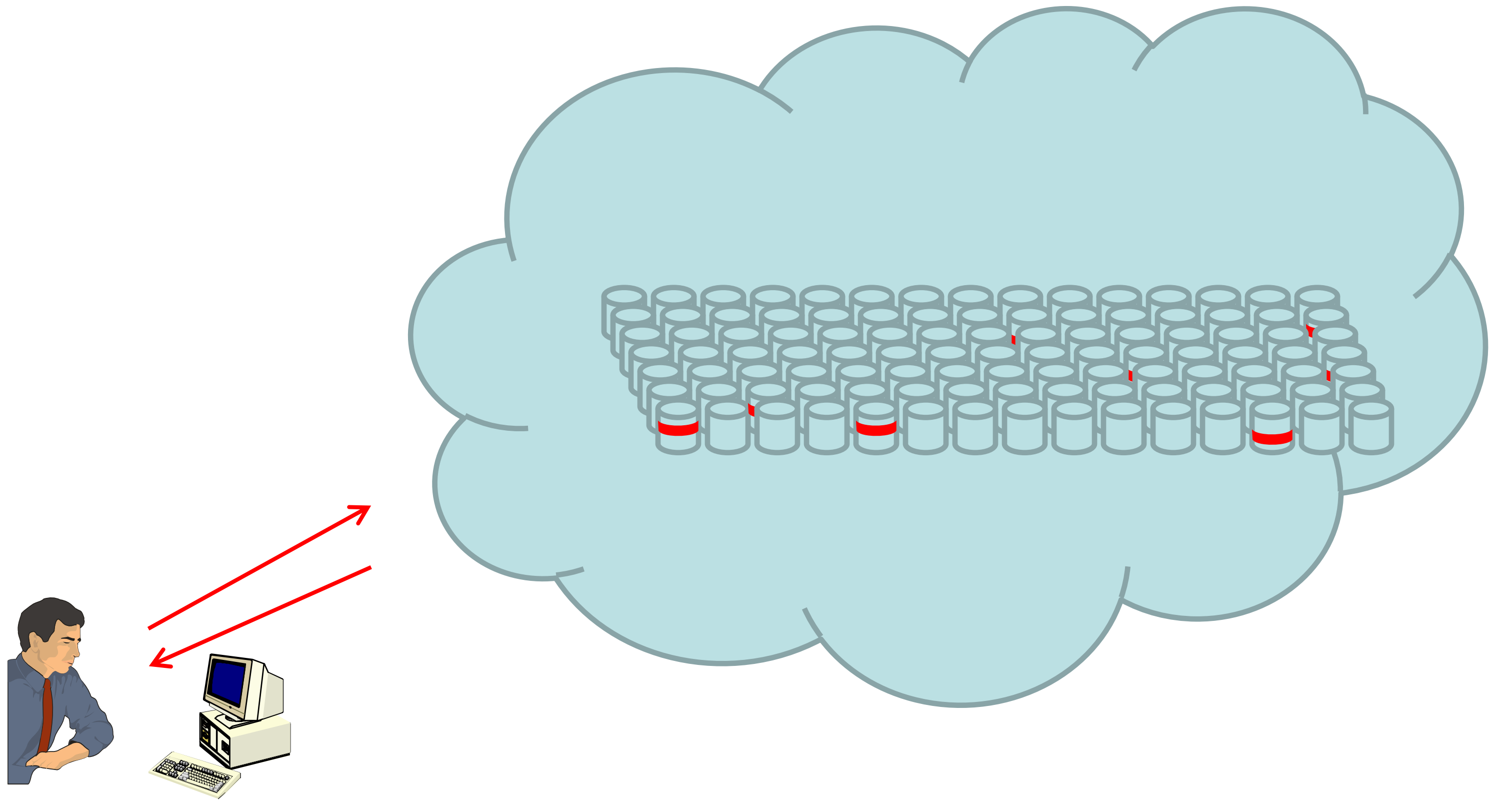
CERN IT Department

- Scalability
  - Ensuring that the existing infrastructure can scale up by a factor of 10 (exabyte scale)
- Reliability
  - Ensure that we can deliver an arbitrary high level of reliability
- Performance
  - Ensure that we can deliver an arbitrary level of performance

- Why ?

The cloud model of the past

CERN IT Department



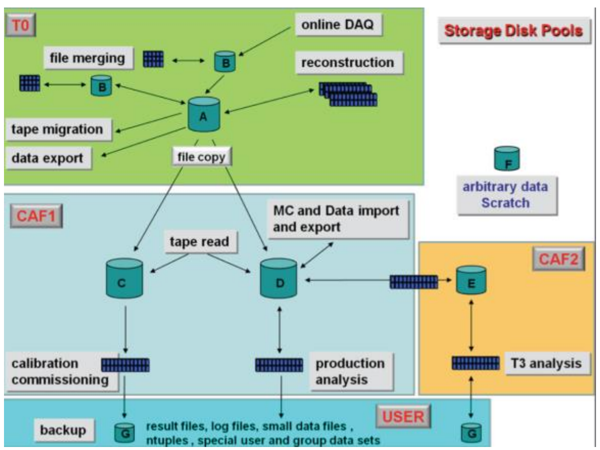
- User would store the data in a “unique pool of data”
- No need to move the data, it is managed “in the cloud”
- Simple and unique quality of service

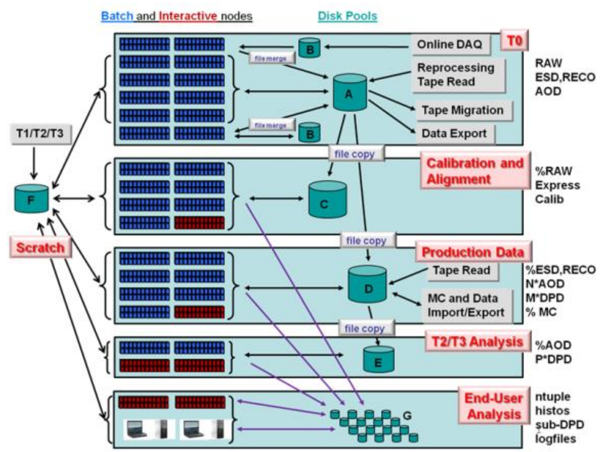
Cloud storage ≠ big data

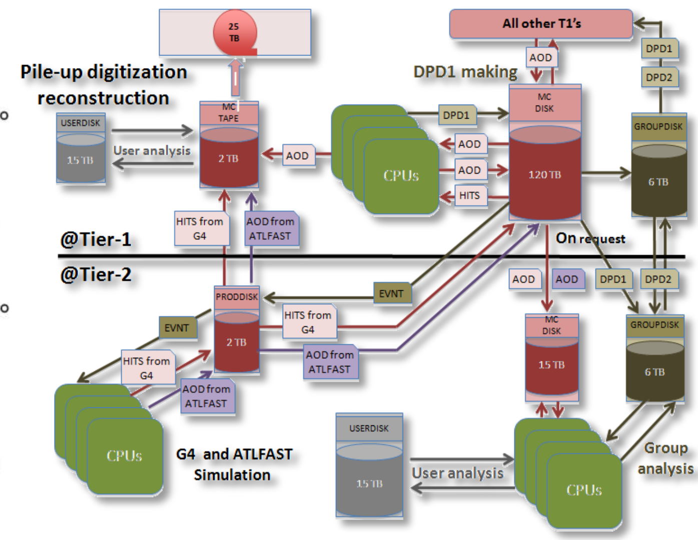
- Unique quality of service becomes too simplistic when dealing with big data
  - Recent data to be analyzed require very high storage performance. *A generic solution can be too slow for this data*
  - Data that has been processed must be archived reliably at minimum cost. *A generic solution can be too expensive for this data*
  - Derived data can, in some case, be recomputed. *A generic solution can be too reliable for this data* (and therefore too expensive)

Real life data management ?

- Examples from LHC experiment data models





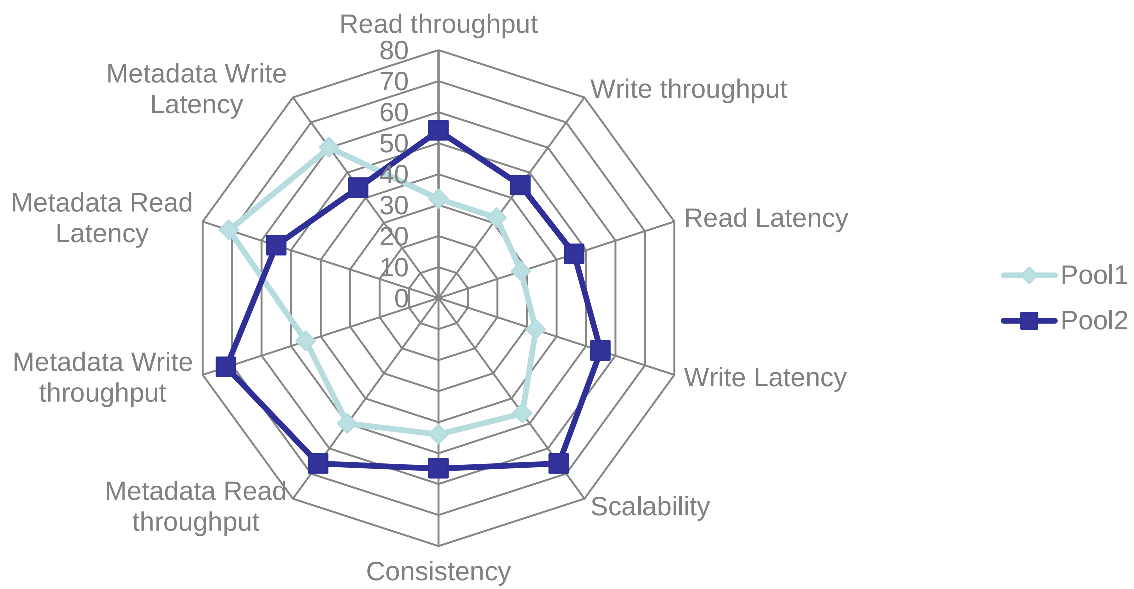


- **Two building blocks to empower data processing**
  - Data pools with *different quality of services*
  - Tools for *data transfer* between pools

DSS

Challenges ahead

- 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”



Challenges: what is needed

- Scalability
  - Ensuring that the existing infrastructure can *scale up by a factor of 10* (exabyte scale)
- Reliability
  - Ensure that we can deliver an *arbitrary level* of reliability
- Performance
  - Ensure that we can deliver an *arbitrary level* of performance
- Cost
  - Ensure that for a given level of reliability / performance, we are “*cost efficient*”

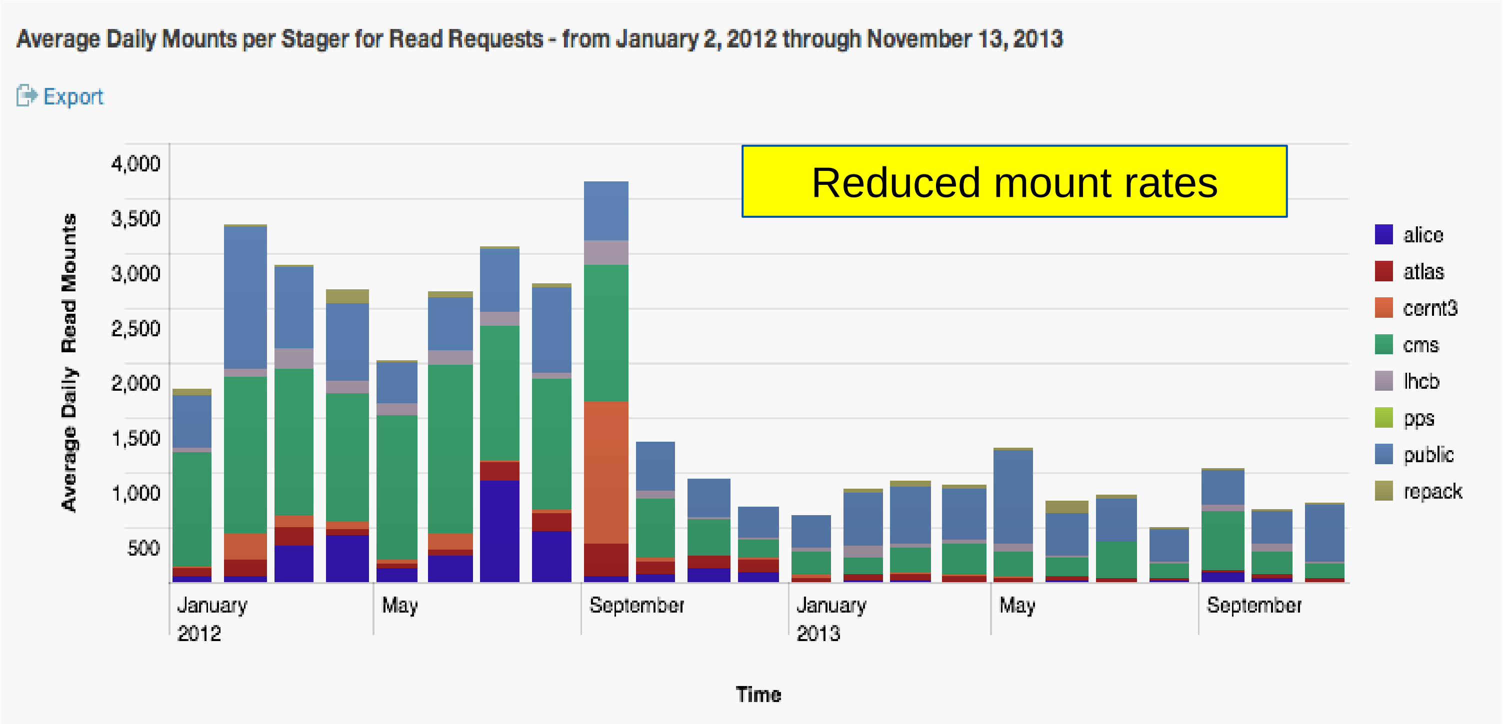
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Examples of challenges

DSS

2013 Scalability improvements

CERN IT Department

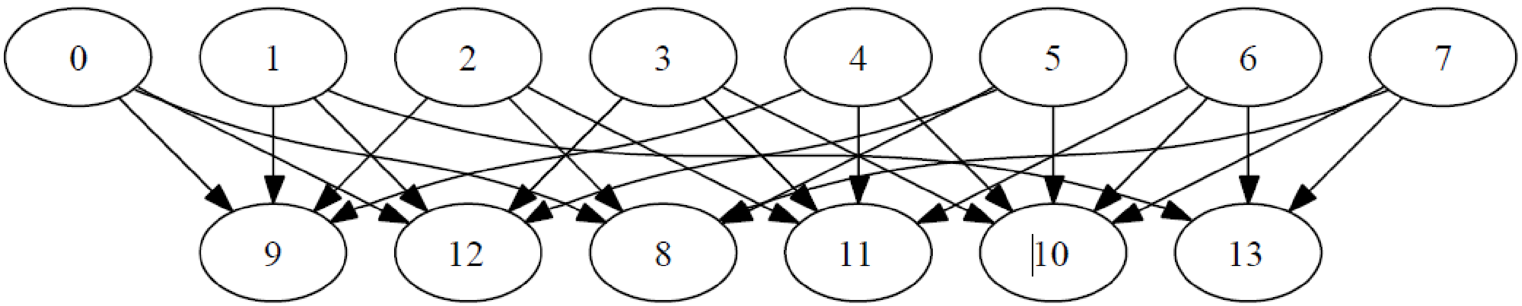


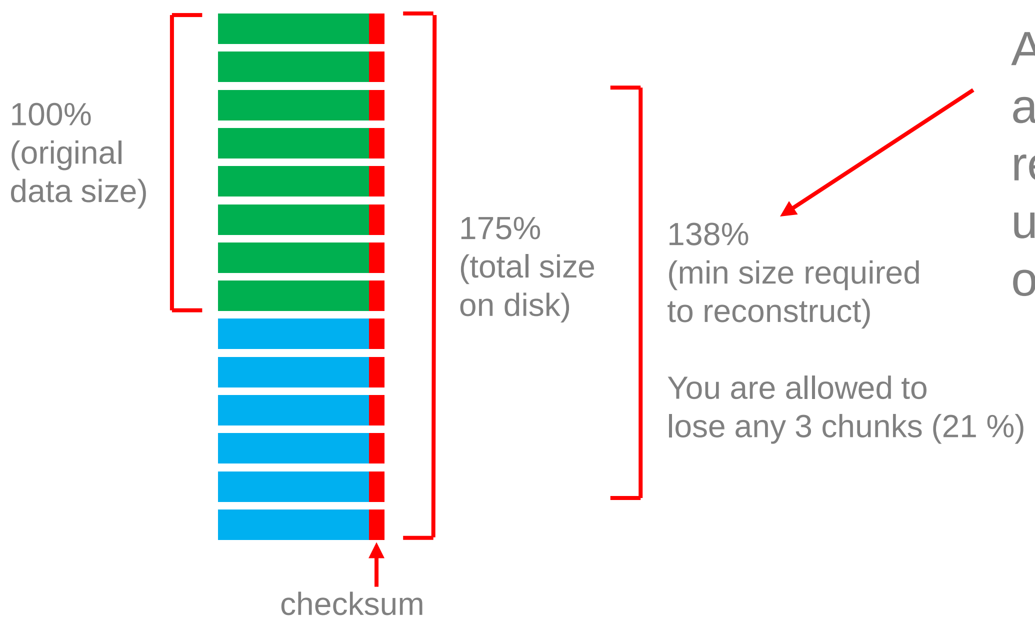
DSS Reliability improvements

- Change “on the fly” reliability of storage pools by supporting multiple data encoding algorithms
  - Plain (reliability of the service = reliability of the hardware)
  - Replication
    - Reliable, maximum performance, but heavy storage overhead
    - Example: 3 copies, 200% overhead
  - Reed-Solomon, double, triple parity, NetRaid5, NetRaid6
    - Maximum reliability, minimum storage overhead
    - Example 10+3, can lose any 3, remaining 10 are enough to reconstruct, only 30 % storage overhead
  - Low Density Parity Check (LDPC) / Fountain Codes / Raptor Codes
    - Excellent performance, more storage overhead
    - Example: 8+6, can lose any 3, remaining 11 are enough to reconstruct, 75 % storage overhead (See next slide)

Example: 8+6 LDPC

0 .. 7: original data
8 .. 13: data xor-ed following the arrows in the graph

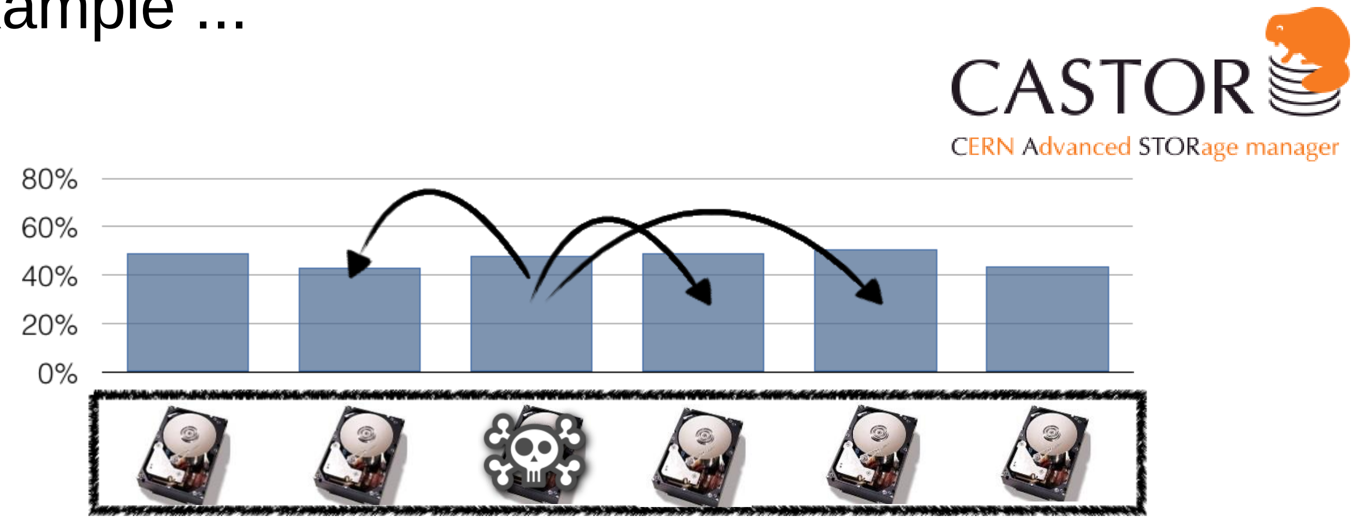


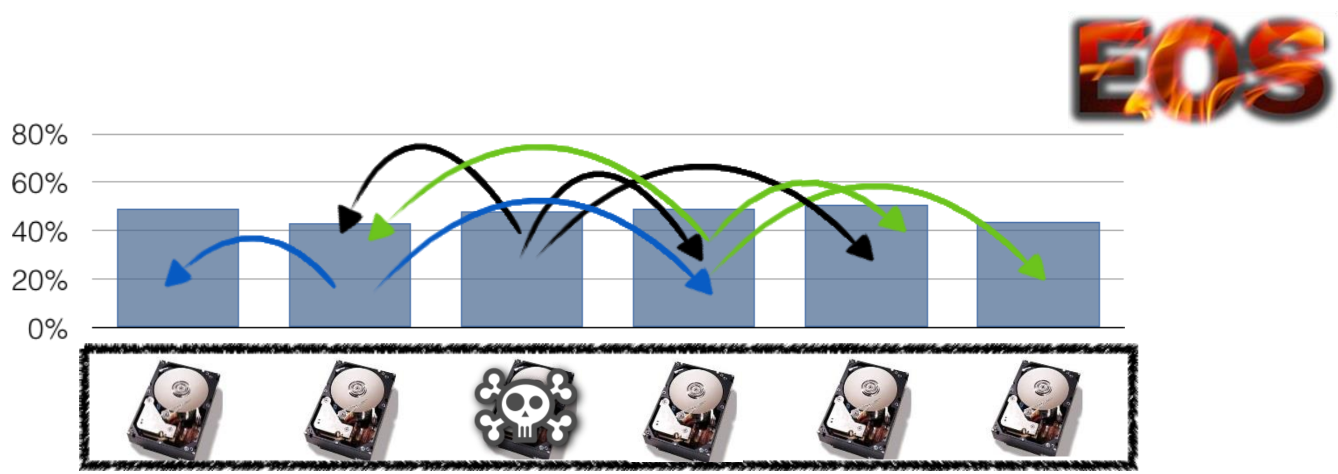


Any 11 of the 14 chunks are enough to reconstruct the data using only XOR operations (very fast)

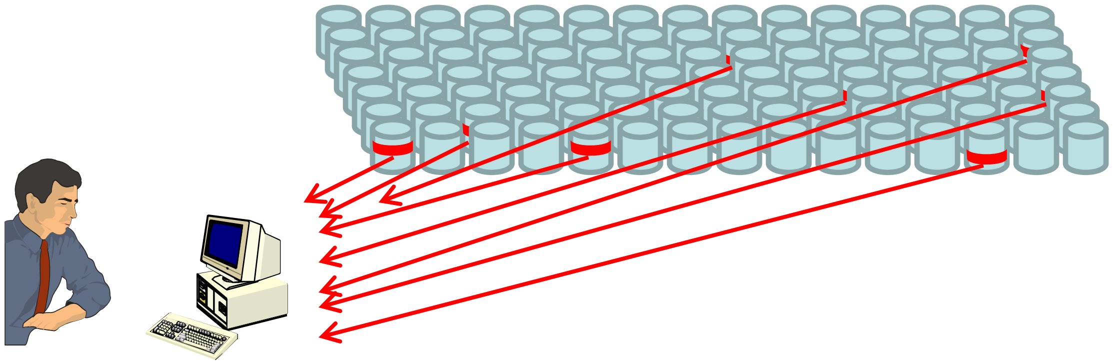
Performance

Draining example ...





High performance parallel read



Data reassembled
directly on the client

Conclusion

CERN IT Department

- We are only at the beginning of what “Data management” services can do
- Efficient “Data management” services still need to be invented !
- Plenty of research opportunities
  - Algorithms (encoding / decoding / error corrections)
  - Storage media, SSD (r)evolution, role of tapes
  - Innovative operation procedure (data verification, data protection, data movements)
  - Architectures (reducing “dependencies of failures, scalability, …)