Getting Out The Data - Fighting The Latency Dragon

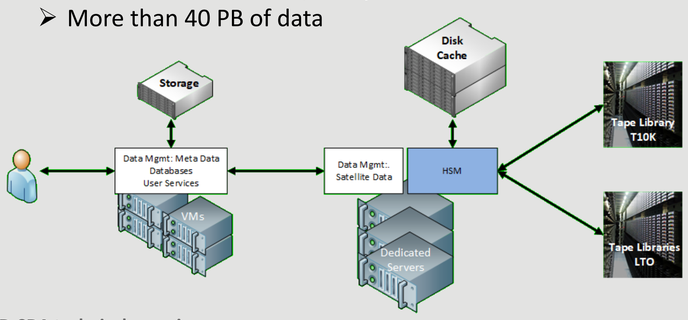
Charlotte Wehn

Deutsches Zentrum für Luft- und Raumfahrt (DLR), Earth Observation Center, Oberpfaffenhofen
charlotte.wehn@dlr.de

The D-SDA Long Term Archive

D-SDA is the German Satellite Data Archive, a long term archive for earth observation data established in the 1990s.

- Operated by DLR (German Aerospace Center) Earth Observation Center
- More than 40 national, European and international missions
- More than 40 PB of data



D-SDA technical overview

Long term storage on tape

- Cost: tape is cheaper than disk
- Energy efficiency: idle tape does not use energy
- Reliability: tape is designed for longer life than disk
- Security: It's harder to delete or encrypt data on tape

Putting the Data to Work

The D-SDA is an active archive, with data being used for scientific and third party projects and demand growing with new ideas.

Time series, e.g. for

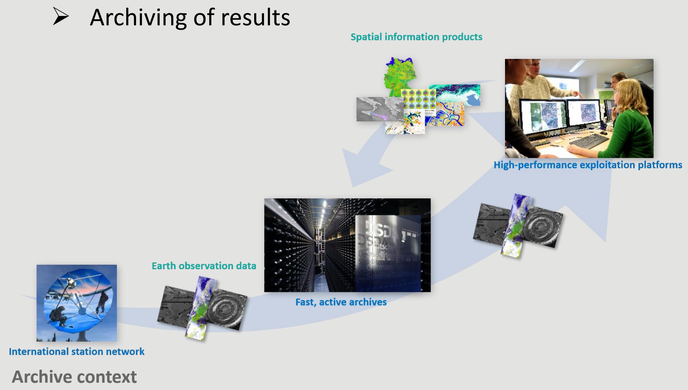
- Urban growth
- Land use
- Ozone hole changes
- Climate change

New exploitation platforms for data access and processing

- Processing services on platforms instead of downloads ("user to the data")
- Large amounts of data online, usually more recent data
- Processing optimized data formats (analysis ready data)
- Often very high processing power

Platforms interact with archives:

- Initial load from archive to platform
- Gap filling
- Reload of historical or evicted data
- Archiving of results



Archive context

Higher processing power and algorithmic maturity enable new ways of working with earth observation data

- **More data is needed …**
- **… faster**

The Problem

Latency!

Tape speed increases, but latency is hard to reduce and makes overall access slow.

Startup time for tape reads

- Tape mounts
- Tape positioning → can take a very long time

Implications:

- Serial reads of many files, same write and read order → OK
- Random reads from many tapes → Slow

Write and read patterns differ frequently.

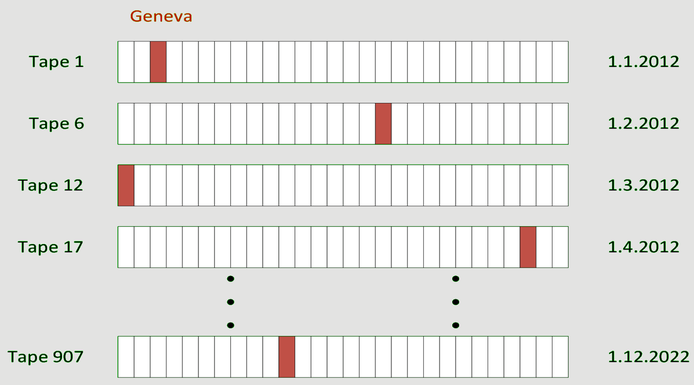
Exploitation Platforms

- Initial load: certain types of data or regions (e.g. Europe) → selective reads over large number of tapes
- Reloads: by user demand → random reads

Time series over a certain region

- Writes by archiving time
- Reads by location

Example: Time series for images taken over Geneva, using monthly image over 10 years.



Tape storage pattern for data of a region over time

With large amounts of data, reading data like this takes forever!

Workarounds

Optimize access patterns

Data extraction from the archive can be optimized within processing chains, done e.g. for the DLR TIMELINE project (see PV presentation by M. Wolfmüller [1]).

- Start extracting data before processing
- Build a pipeline reading data from tape, extracting from the archive and processing data in parallel bulks

Drawbacks:

- Need to know data needs in advance
- Needs sufficient amount of disk cache / processing cache
- Considerable effort to adjust processing chains, cache sizes and processing resources
- With more processing power, archive resources still remain the bottleneck
- Hard for external systems (e.g. platforms): no knowledge about archiving environment
- User driven / non-systematic data requirements not predictable

Keep everything online

With larger disks and falling disk prices, more data can be kept online.

Drawbacks:

- Data volume growing exponentially
- Cost + energy consumption of disk still higher than tape
- For reliable archive: 2 disk copies in different locations, needs twice the disk capacity

Our Solution

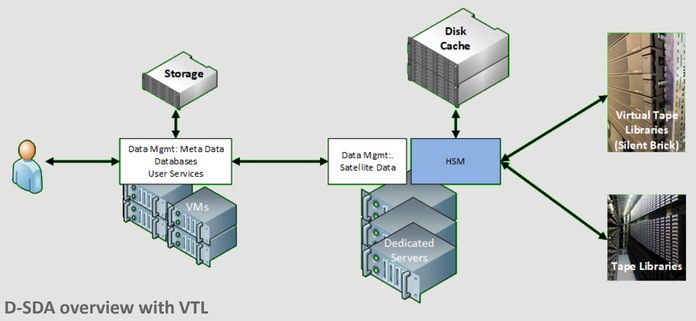
We now store our primary copy on a virtual tape library, which emulates a tape library, drives and tapes.

Virtual tape library (VTL): FastLTA Silent Brick System

- Disk based → **Fast positioning, less latency**
- Virtual tape = Disk shelf with erasure coding
  4 out of 12 disks **redundancy → data safety**
- Powers down disks for "unmounted tapes"
  Max. number of active disks controlled by virtual drives
  → Limit and control energy consumption!

Integration in archiving environment

- Configured with 26 disk shelves / 2 virtual drives per VTL
- Integrated as "tape library" with existing hierarchical storage management software (Oracle HSM)
  → Established mechanisms for data steering
  → No change for archiving application
- Keep one real tape copy in different location (building)



D-SDA overview with VTL

Experience

FastLTA Silent Brick Systems have been in production since 2021.

Operations

- Mostly stable
- Occasional disk failures and firmware updates
- Two unplanned downtimes of an entire system (central part failure), less than five unplanned downtimes for a single Shelf
- No data loss

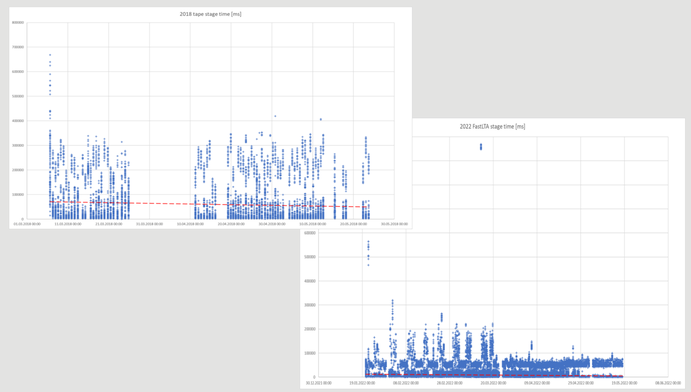
Energy consumption

- Vendor information: typical usage for one system with 2 active Shelves: ~1,2kW
- Power measurement currently only possible for entire room including other servers and storage systems
- Live measurement for traditional tape library: ~ 5kW
- Systematic measurements still outstanding

Performance

- Power-on time: slightly faster than tape mounts
- "Spooling" time: none
- Bandwidth higher than our current tape drives
- Overall read times reduced, especially long reads

Read times extracted from archiving application logs for data deliveries for AC-SAF project:



Application read times: tape vs. virtual tape library

References

1. M. Wolfmüller, S. Holzwarth, S. Asam, S. Kiemle, D. Krause, A. Scherbachenko: Optimized Data Access from and to a Long-term Archive for the Processing of Time Series, PV 2023, CERN, Geneva, Switzerland