LARGE USER DATA MIGRATION—WHAT WE'VE LEARNED

Milan Daneček, Jan Horníček

SIG-CISS, Geneva

29. 5. 2019

Agenda

- short intro into CESNET
- and its data-related services
- case study
  - migration of large diverse user data
  - when changing/renewing hierarchical systems
  - in a specific e-Infrastructure environment

Introduction

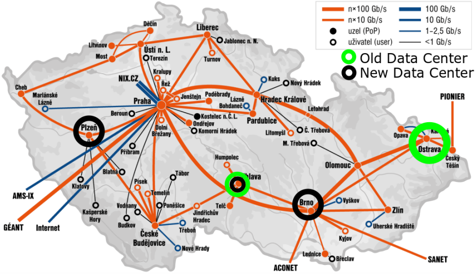
- about CESNET
  - Czech e-Infrastructure provider
  - for Research and Academic sector
  - *Data Storage (DS)*, Networks, Grid&Cloud Computing, Multimedia, etc.
- Data Storage Dpt.
  - data storage for archival, backup, and sharing
  - filesystem and object storage
  - long-term archival storage

Services

- ownCloud
  - CE, default quota 100 GB
  - 13.5k users registered
  - 146M stored objects, 170 TB
- FileSender 2.0
  - within 10 months—up 65 TB/35.1k files, down 76.5 TB/56.5k files
- filesystem access via file transfer protocols (SSH, NFS, SFTP, Globus, . . .)
- long term archival storage
  - dark archive for AIP packages (based on OASIS standard)
  - validation, replicas, periodical check sums, audit logs, autorecovery, . . .

Infrastructure

- HSMs and disk arrays, currently 5 systems
  - 3 HSMs at the end of their life (purchased 2011–2013); total capacity 21 PB
  - 1 new HSM (2018), 1 disk array (2019), total 26 PB
- object storage
  - currently 1 cluster (6.8 PB), tender running for another (est. 20 PB)



Data Migration

- hierarchical storage accessed by users “per system”
  - ftp.du1.cesnet.cz, ssh.du2.cesnet.cz, . . .
- three HSMs reached the end of their life—data migration necessary
- easy way out (for us)—don’t ask, just move all data to a new system, but there is a (big) but
  - all old systems were filled up
  - due to investment schedule (1yr gap between projects)
    - data from at least two systems must fit into a single new one
    - some *data reduction* unavoidable
  - we *don’t want to migrate unnecessary data*

Data Reduction I

- storage facilities were full
  - discussion on regulatory mechanism since 2013 ;)
- how to regulate storage usage?
  - we handle users on individual basis
  - user groups form ad-hoc virtual organisations (managed by user's representatives)
- first, some really bad ideas
  - pay per use: extremely unsuitable for us
    - members: universities, Academy of Sciences
    - CESNET is financed by projects, member fees ($<$ 25%), "commercial" activities ($<$ 10%)
    - members get a bunch of standard services ("for the fee")

Data Reduction II

  - another proposal: moderating member fees by “the ratios of storage usage”
    - member fee is agreed upon by top mgmt of our members
    - users are individuals “in need of storage”
    - wouldn’t solve anything in the end
- what we implemented: dividing the data into categories—backup and archive
  - archive limited by amount of data (quota)
    - we haggle over quotas seriously
  - backup limited by storage time
    - 1 year (reasonable turnaround window for backups)
    - we are entitled to *delete* files later
  - policies applied to new data facilities

Data Reduction III

- back to migration: we *asked the users to migrate their data*
  - no need to migrate backups: users redirect to the new system
  - archives: users must do the transfer
    - at least the users show they still need the data
- users are always free to ask us to migrate the data for them

Technical and Time Aspects

- time necessary
  - migration of large data sets from HSM—recalls from tapes, number of files
  - available network and system throughput
- technical aspects
  - due to investment delays, old facilities no longer under warranty/service
    - extending the service prohibitively expensive
  - we wanted to minimise stress on old systems to avoid catastrophic failures
    - funny stories off the record ;)

Implementation Plan

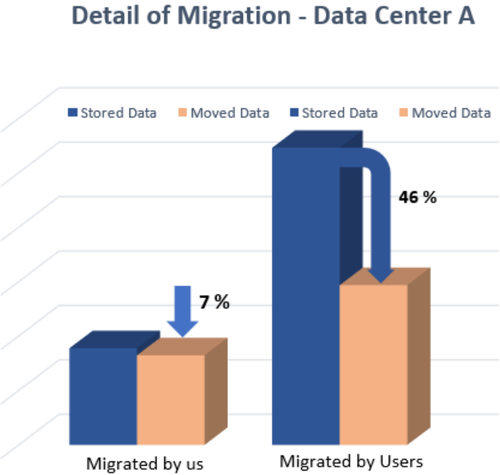
- how to distribute load caused by migration in time
  - users divided into 5 “migration groups”
  - each migration group up to 200 TB
  - time slots—three weeks, then lock up
- user support
  - archives—guides for Globus or rsync
  - backups—just switch the target
  - accounting, mailing, web to confirm data migration
  - on-demand assistance with data migration
    - migration of large groups (above 100 TB)
    - migration of shared directories—permission/ACL integrity (rsync ⇝ lock up ⇝ final rsync ⇝ open on the new data center [to minimise down-time])

Implementation Reality

- users often ignored emails about data center decommissioning
  - locking users out of data *absolutely* necessary
  - most users have woken up after data lock-up
  - some detective stories—finding users responsible for the data
- users were postponing data migration
  - dividing into groups had positive effect
- active (email-reading) users were cooperating well
- we reduced total amount of migrated data
  - impossible for us to distinguish backups

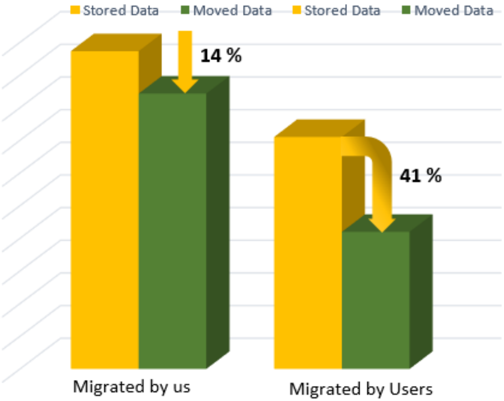
cesnet datacare

Data Migration in Numbers I

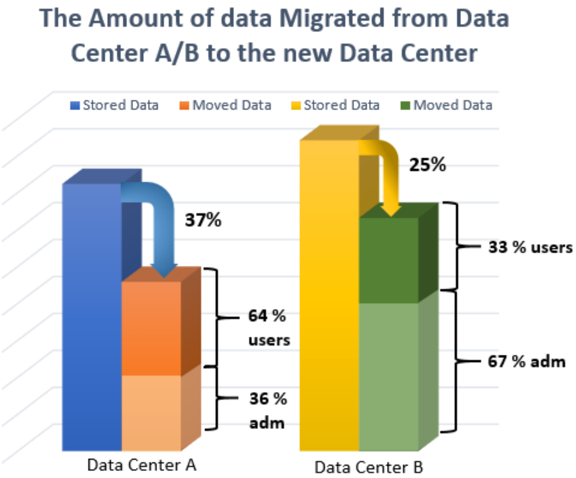


Data Migration in Numbers II

Detail of Migration - Data Center B



Data Migration in Numbers III



Conclusions

- we’ve reduced the amount of migrated data
  - data from center A reduced by 37% after migration
  - data from center B reduced by 25% after migration
    - in B, 67% of data was migrated by admins (upon request)
- users are always the best curators
- prepare for users ignoring mails, hunting them over phone etc.
- dividing users into groups was necessary
- estimate for similar use cases: 1PB/month achievable
  - unless you have extreme numbers of small files

That’s It

- Thank you ⌣
- Questions?

Dark Archive I

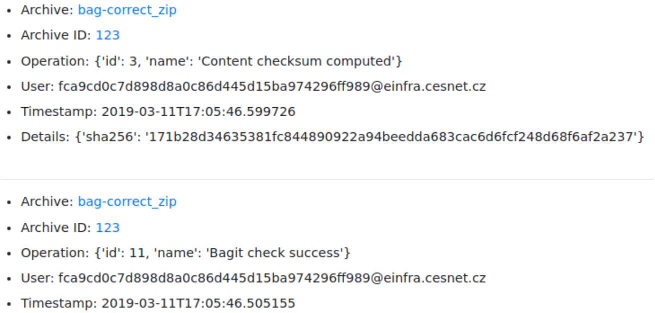
- guaranteed binary storage for valuable data
- storing OAIS Archival Information Packages (metadata, checksum, . . .)
- service was suggested/requested by the community (libraries, uni archives, . . .)
  - users require reliable storage (periodical verification of checksums; restoration from replicas on failure)
- no ambition to provide full LTP including format conversion
  - must be handled by users who understand the information in the data
- plan to interconnect the ‘dark archive’ with the open access repository

Dark Archive II

- primary component is API
- API allows to upload/download packages), check audit logs, searching
- web interface for human access
  - basic functions (up/download, review the audit logs…)

Dark Archive III

example of the audit log



- Archive: bag-correct_zip
- Archive ID: 123
- Operation: {'id': 3, 'name': 'Content checksum computed'}
- User: fca9cd0c7d898d8a0c86d445d15ba974296ff989@einfra.cesnet.cz
- Timestamp: 2019-03-11T17:05:46.599726
- Details: {'sha256': '171b28d34635381fc844890922a94beedda683cac6d6fcf248d68f6af2a237'}

- Archive: bag-correct_zip
- Archive ID: 123
- Operation: {'id': 11, 'name': 'Bagit check success'}
- User: fca9cd0c7d898d8a0c86d445d15ba974296ff989@einfra.cesnet.cz
- Timestamp: 2019-03-11T17:05:46.505155

How It Works

- upload AIP to the system
- AIP is validated, external checksum calculated, internal technical metadata of the AIP checked (internal checksums)
- once validation is done and successful, AIP is stored
- calculated external checksum is stored as an extended attribute
    - only external checksum is used for periodical checks
    - for efficiency reasons
    - we use the same mechanism for general files as well
        - without regular checks, of course