CERN Disk Storage Services:
report from last data taking, evolution and future outlook towards Exabyte-scale storage

Luca Mascetti
CERN IT-Storage
luca.mascetti@cern.ch
Our goal:
provide and operate storage solution and tools
for data management and data analysis
to experiments and users.
CASTOR and CTA

330PB of data (+10 PB dual copy)
~0.6 EB capacity
Minimal space increase in 2019
“Only” 16 PB of disk cache on CC7

Preparing the migration to CTA

E.Cano: CERN Tape Archive: production status, migration from CASTOR and new features
**EOS**

Production Instances:

- 5 for the LHC experiments
- 9 CERNBox
  - EOSUSER + 5 EOSHOME + 3 EOSPROJECT
- EOSMEDIA (photo/audio/video)
- EOSPUBLIC (Open Data and non-LHC experiments)
- EOSBACKUP (backup for CERNBox)

Total Space: 340 PB

Files Stored: 5.12 Bil

# Storage Nodes: ~1600

# Disks: ~60000

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>2 PB</td>
</tr>
<tr>
<td>2012</td>
<td>12 PB</td>
</tr>
<tr>
<td>2014</td>
<td>40 PB</td>
</tr>
<tr>
<td>2016</td>
<td>150 PB</td>
</tr>
<tr>
<td>2018</td>
<td>270 PB</td>
</tr>
<tr>
<td>2019</td>
<td>340 PB</td>
</tr>
</tbody>
</table>
EOS

• Deployment of a new EOS architecture for CERNBox instances
  • latest generation EOS mounted filesystem (fusex) deployed
• Full deployment of the latest generation namespace (quarkdb)
  • reduced restart time
  • reduced memory requirements
• Wigner Data Centre Decommissioning
  • 90 PB to be decommissioned
  • 80 PB of data to be moved

A.J. Peters: EOS architectural evolution and strategic development directions
CERNBox and SWAN

CERNBox

• 5 years of production
  • still growing service with 18k users
  • 4PB+ data 1B+ files 110k+ shares
• New core daemon (Reva) in production
• Consolidating HOMEs into CERNBox
  • Migration out of DFS ongoing
  • HA SAMBA gateway in production
• Central Hub for CERN data and apps
  • draw.io, onlyoffice, collabora, kopano, …

SWAN

• Turn-key data analysis platform
• Accessible from everywhere via a web browser
• Support for ROOT/C++, Python, R, Octave
• Fully integrated in CERN ecosystem
  • Storage on EOS, Sharing with CERNBox
  • Software provided by CVMFS
  • Massive computations on Spark

H.G.Labrador: Evolution of the CERNBox platform to support collaborative applications and MALT
Ceph, CVMFS and S3

Ceph
- Openstack RBD remain the biggest use-case
  - ~1PB/year
- Openstack Manila Share in production during 2018
  - NFS-like share
- Storage for Kopano

CVMFS
- 49 repositories with 868 M files and 52 TB
  - 43 using block storage 6 using S3 objects

S3
- IT applications, ATLAS+LHC@home
- CVMFS repositories
- CERNBox new backups using Restic

<table>
<thead>
<tr>
<th>CERN Ceph Clusters</th>
<th>Size</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenStack Prod.</td>
<td>6.4 PiB</td>
<td>mimic</td>
</tr>
<tr>
<td>Cinder/ Glance Wigner</td>
<td>1.6-0 PiB</td>
<td>nautilus</td>
</tr>
<tr>
<td>CephFS Prod.</td>
<td>1.09 PiB</td>
<td>luminous</td>
</tr>
<tr>
<td>CephFS Pre-prod.</td>
<td>164 TiB</td>
<td>mimic</td>
</tr>
<tr>
<td>CephFS Hyperconv.</td>
<td>356 TiB</td>
<td>mimic</td>
</tr>
<tr>
<td>CASTOR Disk Buffer</td>
<td>5.5 PiB</td>
<td>nautilus</td>
</tr>
<tr>
<td>S3+SWIFT Prod. (4+2EC)</td>
<td>1.92 PiB</td>
<td>luminous</td>
</tr>
</tbody>
</table>

E.Bocchi: Evolution of the S3 service at CERN as a storage backend for infrastructure services and software repositories
LHC Run2: Heavy Ions Highlights

Data Flow and Data Rates (nominal)
- LHC operations with DC 50-60%
- ALICE: sustain 8GB/s during collision
- ATLAS and CMS: 4GB/s each
- LHCb: 1-2GB/s transferred off-collision
  - tape first workflow direct to CASTOR

Common Readiness tests - Sept. 2018
- wrong OS scheduler policy
- issue with xrdcp checksum computation
  - reading file twice
LHC Run2: Heavy Ions Highlights

EOS Rates during data taking

- 10 GB/s
- 8 GB/s
- 4 GB/s
LHC Run2: Heavy Ions Highlights
LHC Run2: Heavy Ions Highlights

CASTOR Rates during data taking (ALICE)
Future Challenges

Disk Space Optimisation
- Large Scale Erasure Encoding
  - EC4+2 EC8+4 EC10+2

High Density Storage Solutions
- Run3 Storage purchasing
- ALICE O2 Storage hosting in IT?
- LHCb Storage Synergies

A.J. Peters: Erasure Coding for production in the EOS Open Storage system
Future Challenges

New Workflows and New Tools
• CTA instead of CASTOR
• T1s data export via xroot?
• CMS: Phedex- > Rucio

Experiments New Data Rates
• LHCb  10x Bandwidth increase
  • 10GB/s
• ALICE: 90+GB/s ?

Run3 Common Data Challenge
• Starting the Coordination from Q1 2020
• Testing Each Part of the Workflow
  • from Px to EOS to CTA to Export
Thanks for the Attention!

Questions?

Accélératrice de science