CERN Tape Archive (CTA) :
From Development to Production Deployment

Michael Davis, Vladimír Bahyl, Germán Cancio, Eric Cano, Julien Leduc and Steven Murray

CHEP 2018, Sofia, Bulgaria
9 July 2018
Changing Use Cases for Archival Storage

1. Scaling up for Run 3 and HL–LHC

Predicted T0 Tape Archival Storage Needs

- 2018: 0.25 EB
- 2019: 0.5 EB
- 2020: 1.3 EB
- 2021: 3.1 EB
- 2022: 4.3 EB
- 2023: 4.3 EB
- 2024: 4.3 EB
- 2025: 4.3 EB
- 2026: 4.3 EB
- 2027: 4.3 EB
- 2028: 4.3 EB
- 2029: 4.3 EB
- 2030: 4.3 EB
- 2031: 4.3 EB
Changing Use Cases for Archival Storage

2. Data for online analysis stored on tape ("Data Carousel")

What is ‘data carousel’ and why?

Data storage challenge of HL-LHC:

- ‘Opportunistic storage’ basically doesn’t exist
- Format size reduction and data compression are both long-term goals, require significant efforts from the software and distributed computing teams
- Tape storage is 3~5 times cheaper than disk storage, increasing tape usage is a natural way to cut into the gap of storage shortage for HL-LHC

‘Data Carousel’ R&D → to study the feasibility to use tape as the input to various I/O intensive workflows.

Source: Tape Usage, Xin Zhao (Brookhaven National Laboratory), ADC Technical Coordination Board Meeting, 28 May 2018
Changing Use Cases for Archival Storage

2. Data for online analysis stored on tape ("Data Carousel")

What is "Data Carousel"?

Data storage challenge:

- "Opportunity" for tape usage
- Format shift: As data formats are both becoming more complex and efforts from the data analysis and computing areas increase
- Tape storage is a promising solution for archival storage, it has been considered as the input to the HL-LHC

Source: Tape Usage, Xin Zhao (Brookhaven National Laboratory), ADC Technical Coordination Board Meeting, 28 May 2018
Changing Use Cases for Archival Storage

2. Data for online analysis stored on tape (“Data Carousel”)

Source: Fascinating Vintage 20 Cassette Carousel from 1972 : Panasonic RS–296US
CASTOR Architecture

**Use Cases**
- CASTOR to CTA
- Testing and Deployment
- Summary

**CASTOR Architecture**

- **Client**
- **Disk Servers**
- **Disk Cache**
- **DB:** Queues + Catalogue
- **Disk Cache Management/Stager**
- **Central Services**
  - NameServer
  - VMGR
  - VDQM
- **Tape Server Daemon**

**Physical Infrastructure:**
- Libraries, Drives, Cassettes
CERN Tape Archive Architecture

- **Client**
  - Physical Infrastructure: Libraries, Drives, Cassettes
  - Tape Server Daemon

- **CTA Frontend**

- **DB**: Catalogue

- **Object Store**

- **Tape Server Daemon**

- **Physical Infrastructure**: Libraries, Drives, Cassettes
CTA Architecture

CTA offers the “Best of Both Worlds”

- User interface, file access and disk pool management from EOS
- Tape system management from CASTOR
- New scalable, robust queuing system to link the two

CTA design principles

- Simplicity
- Scalability
- Performance

Full details: An efficient, modular and simple tape archiving solution for LHC Run 3, Steven Murray et al. (CERN), CHEP 2016
## CTA Architecture

<table>
<thead>
<tr>
<th>CASTOR</th>
<th>CTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheduling decisions made at time of user request.</td>
<td>Scheduling decisions made at time of tape mount.</td>
</tr>
<tr>
<td>Tape drive may not be available when job reaches the front of the queue.</td>
<td>Tape drive allocated when job reaches the front of the queue. Reduced latency for users.</td>
</tr>
<tr>
<td>High-priority jobs cannot interrupt running jobs.</td>
<td>High-priority jobs can preempt lower-priority jobs.</td>
</tr>
<tr>
<td>Can switch from repack to data taking and back without operator intervention. System operates at full capacity at all times.</td>
<td></td>
</tr>
</tbody>
</table>
CASTOR Deployment Model

CASTOR ALICE
CASTOR ATLAS
CASTOR CMS
CASTOR LHCb
CASTOR PUBLIC

EOS

Experiment
CTA Deployment Model

<table>
<thead>
<tr>
<th>Experiment</th>
<th>EOS+CTA</th>
<th>ALICE</th>
<th>EOS+CTA</th>
<th>ATLAS</th>
<th>EOS+CTA</th>
<th>CMS</th>
<th>EOS+CTA</th>
<th>LHCb</th>
<th>EOS+CTA</th>
<th>PUBLIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Infrastructure:</td>
<td>Libraries, Drives, Cassettes</td>
<td>Tape Server Daemon</td>
<td>FST</td>
<td>CTA Frontend</td>
<td>MGM</td>
<td>Tape</td>
<td>Catalogue</td>
<td>DB</td>
<td>Object Store</td>
<td>+ CTA</td>
</tr>
</tbody>
</table>

Use Cases

CASTOR to CTA

Testing and Deployment

Summary
System Testing

Scale tests and stress tests:
10 million files archived in ∼27 hours

Files archived to tape per second

Time (hours)
Field Testing

- Goal of user testing is to ensure that all use cases are covered
- Rucio/File Transfer Service (FTS) tests with ATLAS have started

Transfer 'c2f71c26-761b-11e8-ae8-02163e01826d' FINISHED

<table>
<thead>
<tr>
<th>VO: atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total size</td>
</tr>
<tr>
<td>976.56 KiB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>File ID</th>
<th>File State</th>
<th>File Size</th>
<th>Throughput</th>
<th>Start Time</th>
<th>Finish Time</th>
</tr>
</thead>
</table>

Next:
- Agree schedule for field testing with all CERN experiments
Migration Schedule

2018 Run 2
2019 LS2
2020 LS2
2021 Run 3

Migration during period of reduced data taking.

Migration is a metadata-only operation. No physical movement of data on tape.
CERN Tape Archive: Summary

Use cases for tape archival are changing

- Increased rate of data taking for Run 3 and HL-LHC
- Data for online analysis accessed via “Data Carousel”

CTA is the “Best of Both Worlds” — EOS disk and CASTOR tape

- Simplicity
- Scalability
- Performance

Deployment

- Now: Field test instances with redundant copies of data
- LS2: Migration from CASTOR to CTA