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

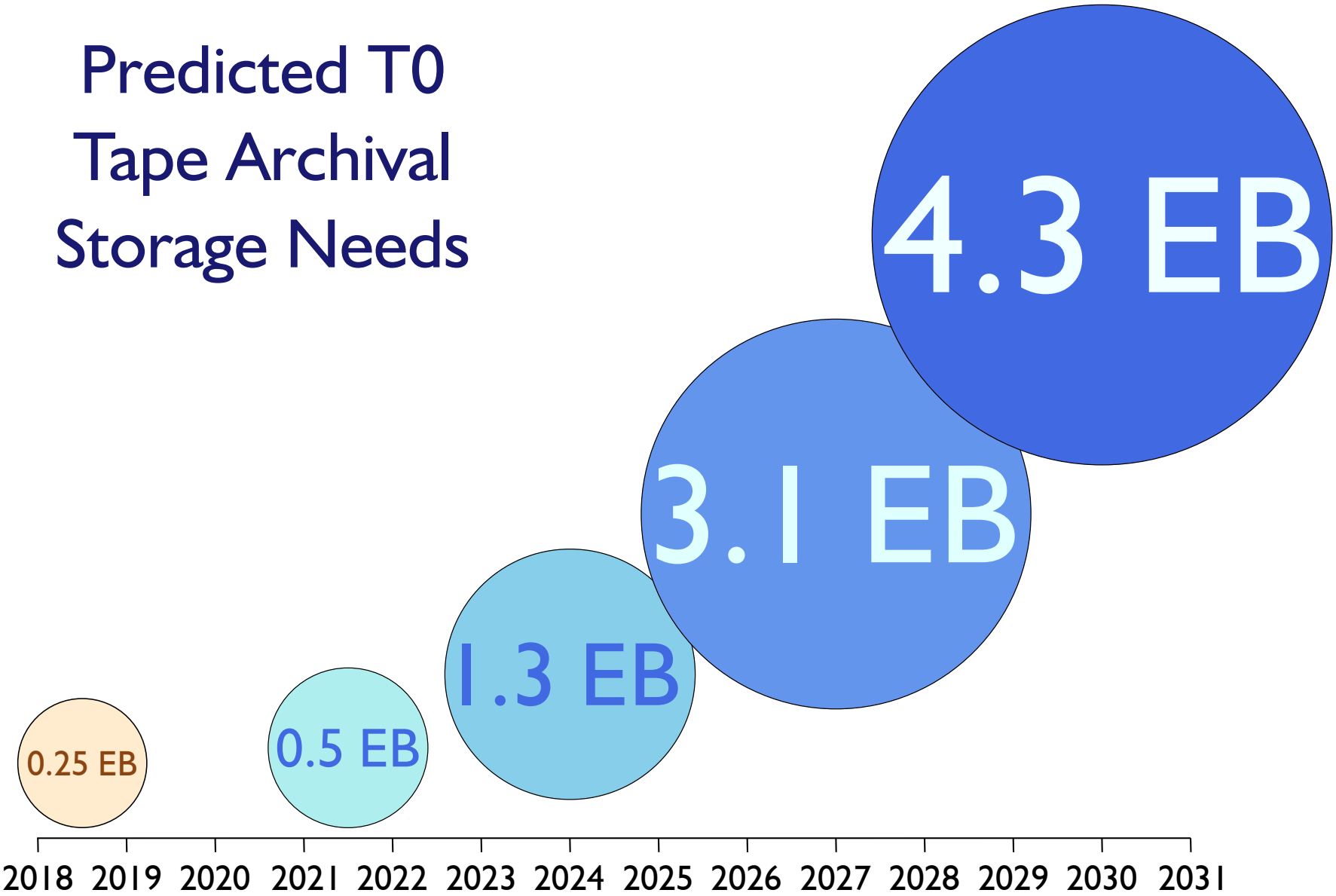
Michael Davis, Vladimír Bahyl, Germán Cancio,
Eric Cano, Julien Leduc and Steven Murray

CHEP 2018, Sofia, Bulgaria

9 July 2018

Changing Use Cases for Archival Storage

1. Scaling up for Run 3 and HL–LHC



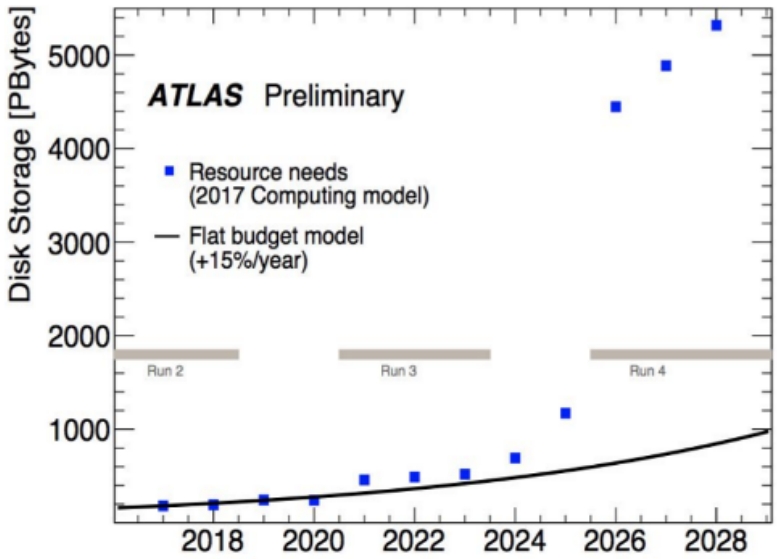
Changing Use Cases for Archival Storage

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

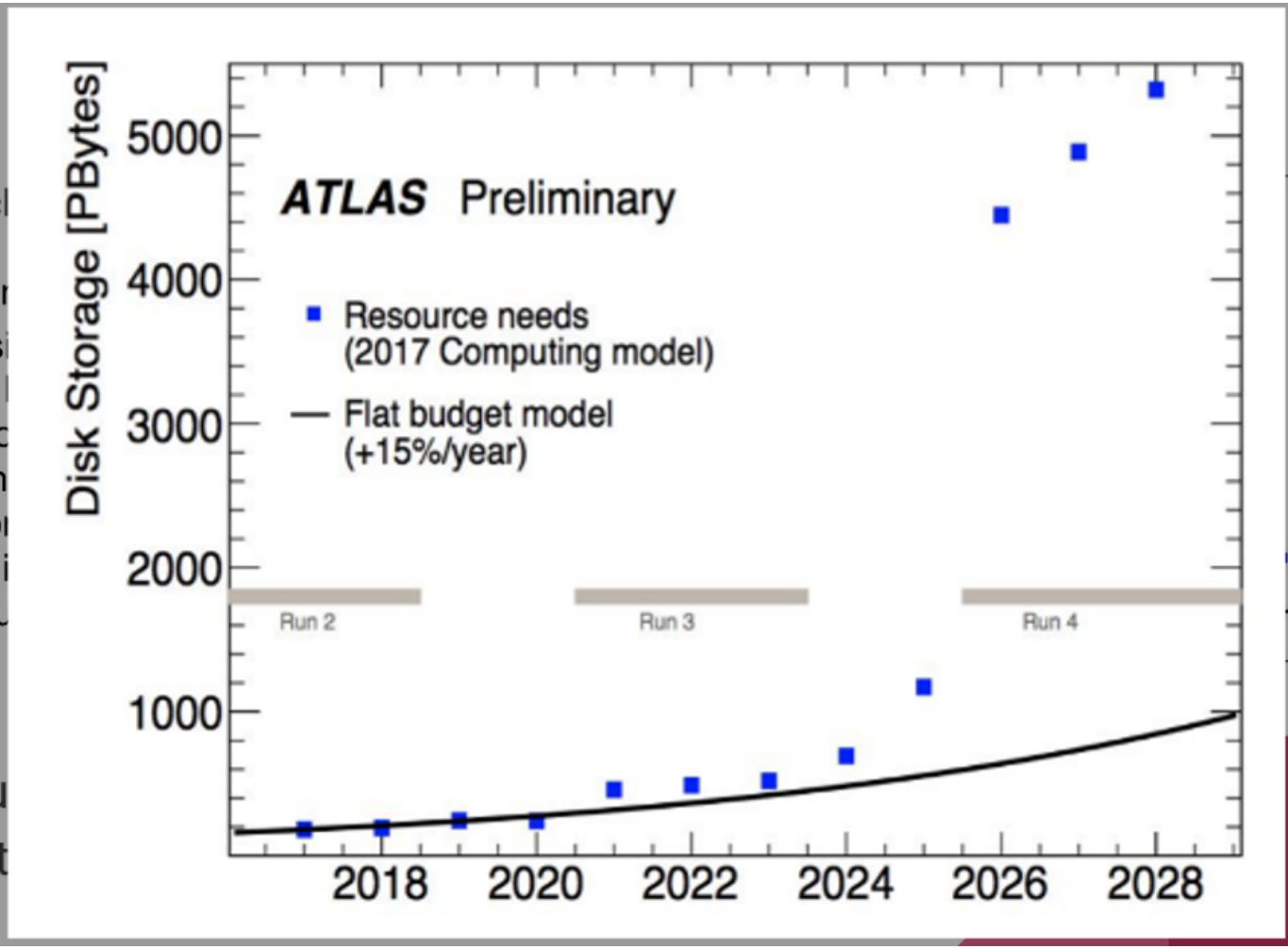
2. Data for online analysis stored on tape ("Data Carousel")

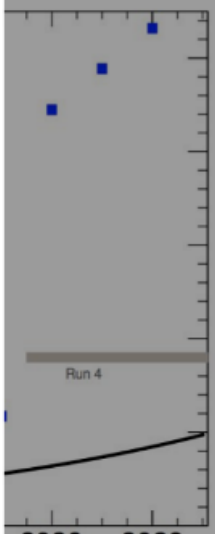
What is

Data storage cl

- ➔ 'Opportu
- ➔ Format si are both efforts fr computin
- ➔ Tape sto storage, way to cu HL-LHC

'Data Carou as the input





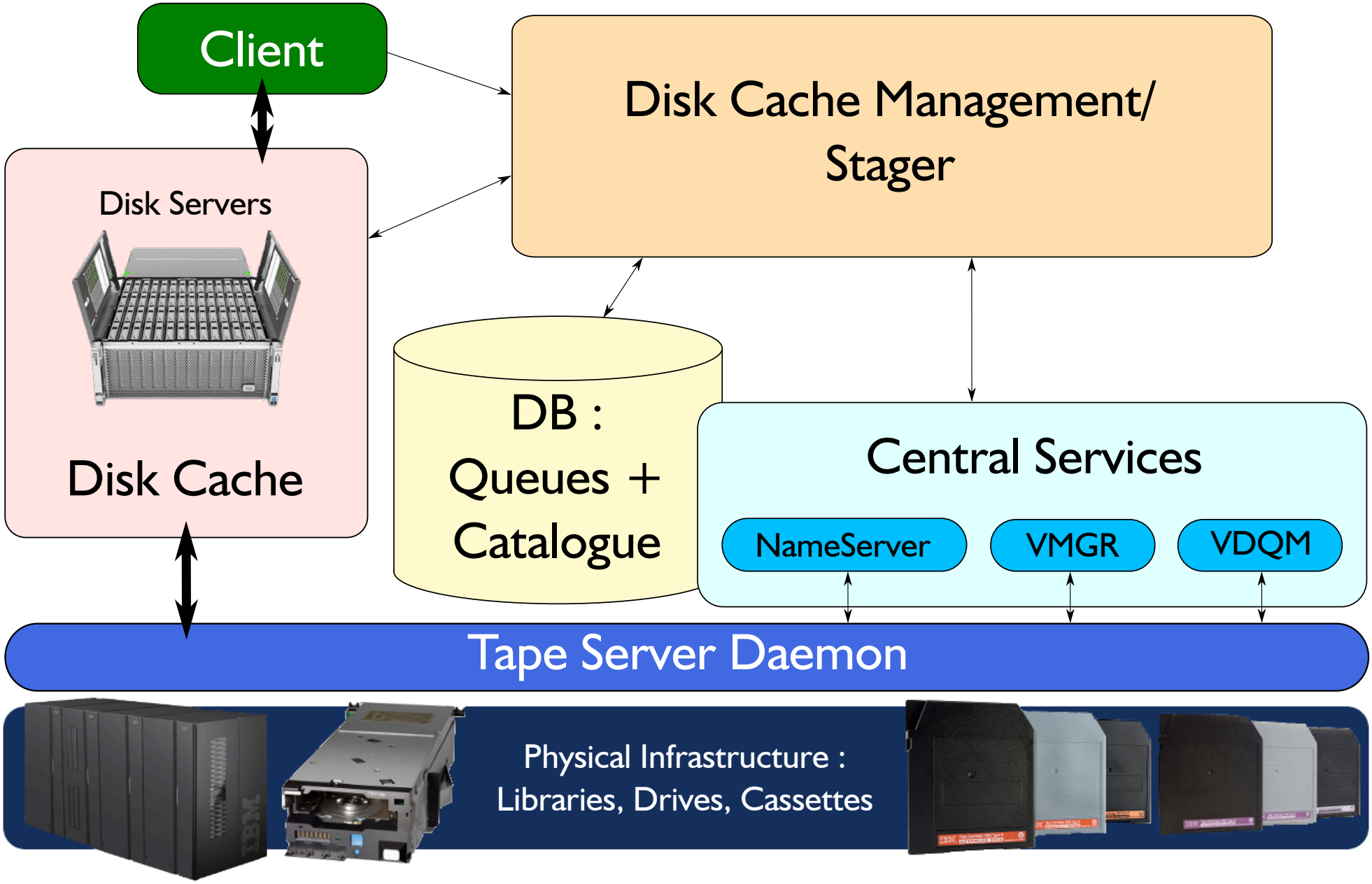
Source: Tape Usage, Xin Zhao (Brookhaven National Laboratory), ADC Technical Coordination Board Meeting, 28 May 2018

Changing Use Cases for Archival Storage

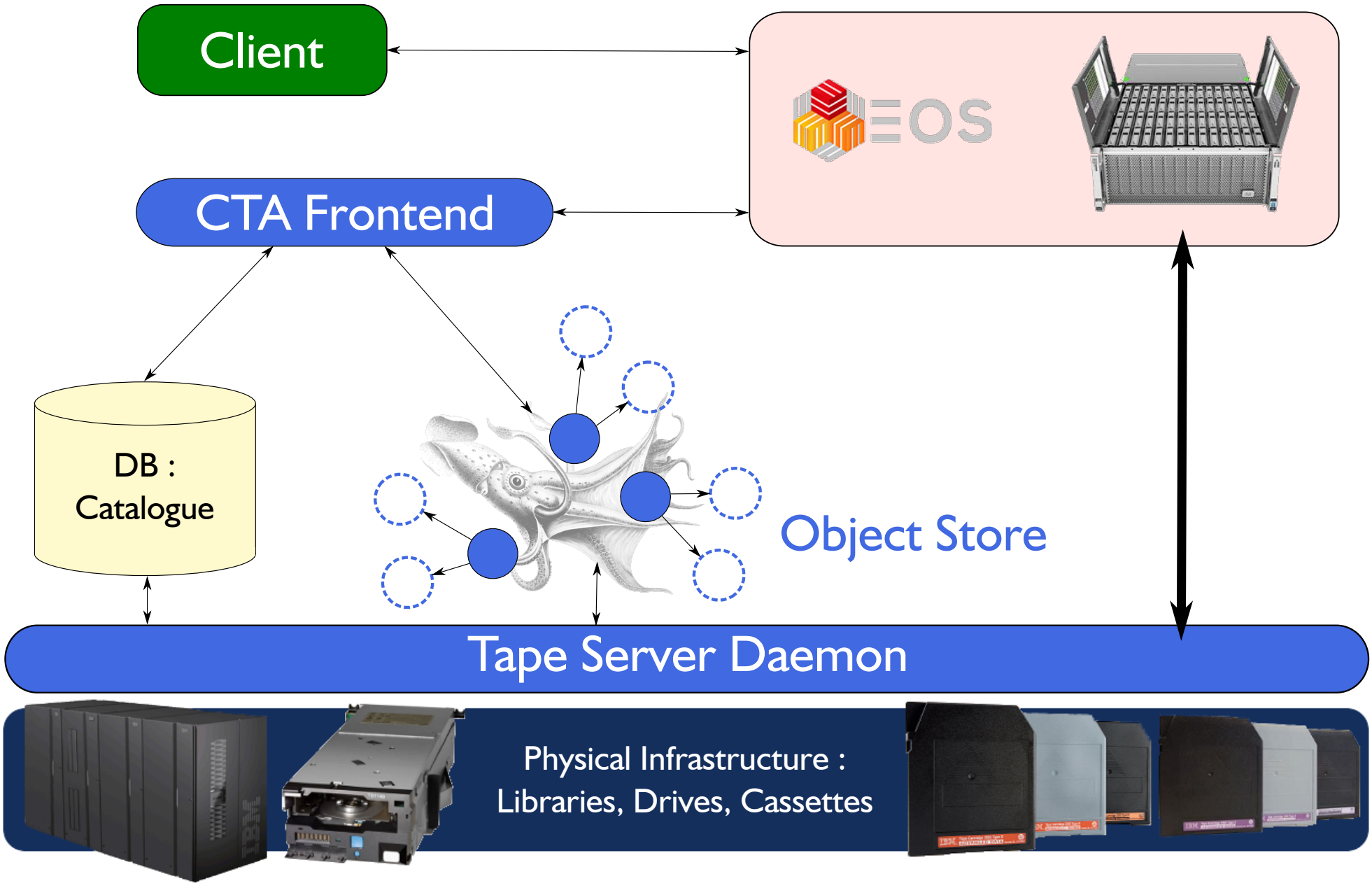
2. Data for online analysis stored on tape ("Data Carousel")

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

CASTOR Architecture



CERN Tape Archive Architecture



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

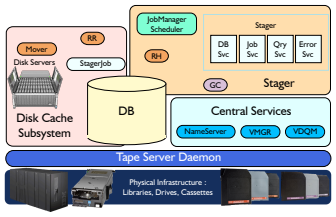
CTA Architecture

CASTOR	CTA
Scheduling decisions made at time of user request.	Scheduling decisions made at time of tape mount.
Tape drive may not be available when job reaches the front of the queue.	Tape drive allocated when job reaches the front of the queue. Reduced latency for users.

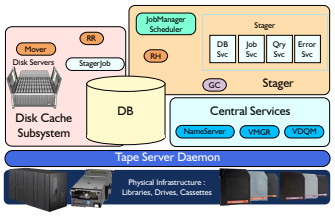
CTA Architecture

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

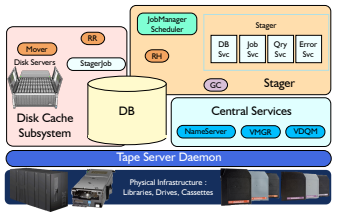
CASTOR Deployment Model



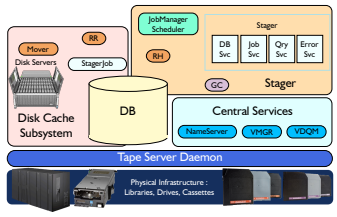
CASTOR ALICE



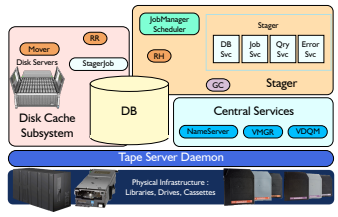
CASTOR ATLAS



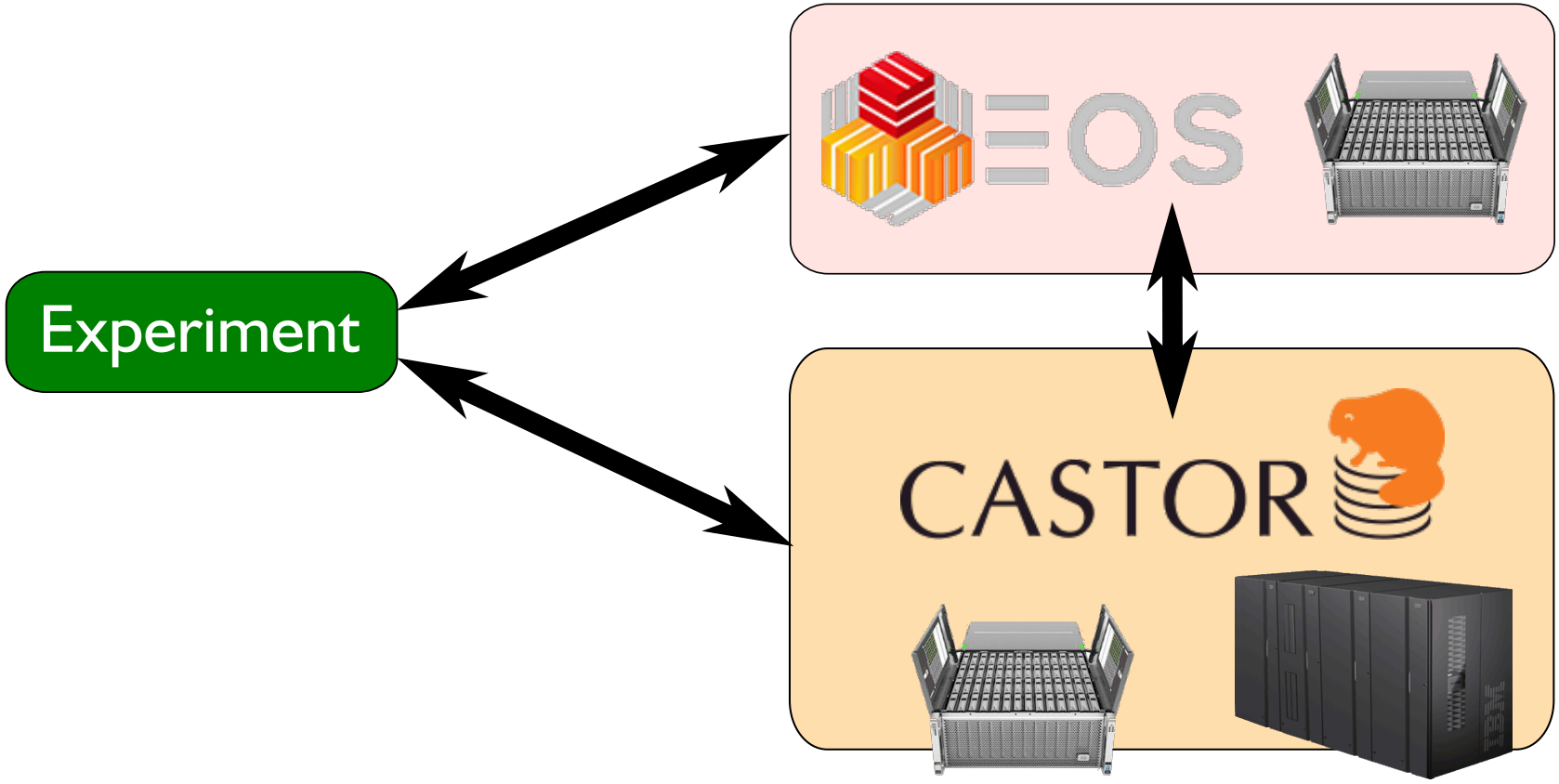
CASTOR CMS



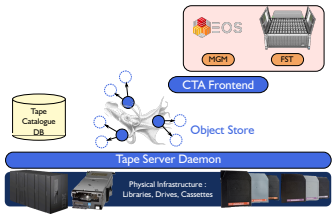
CASTOR LHCb



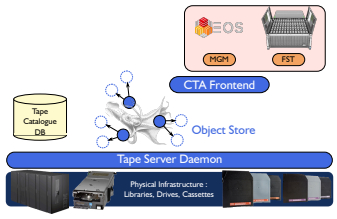
CASTOR PUBLIC



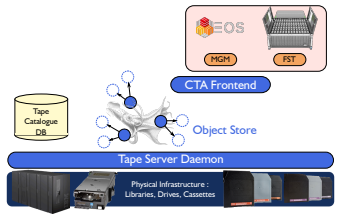
CTA Deployment Model



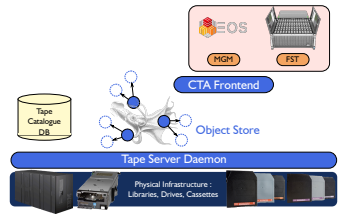
EOS+CTA
ALICE



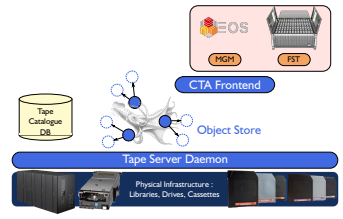
EOS+CTA
ATLAS



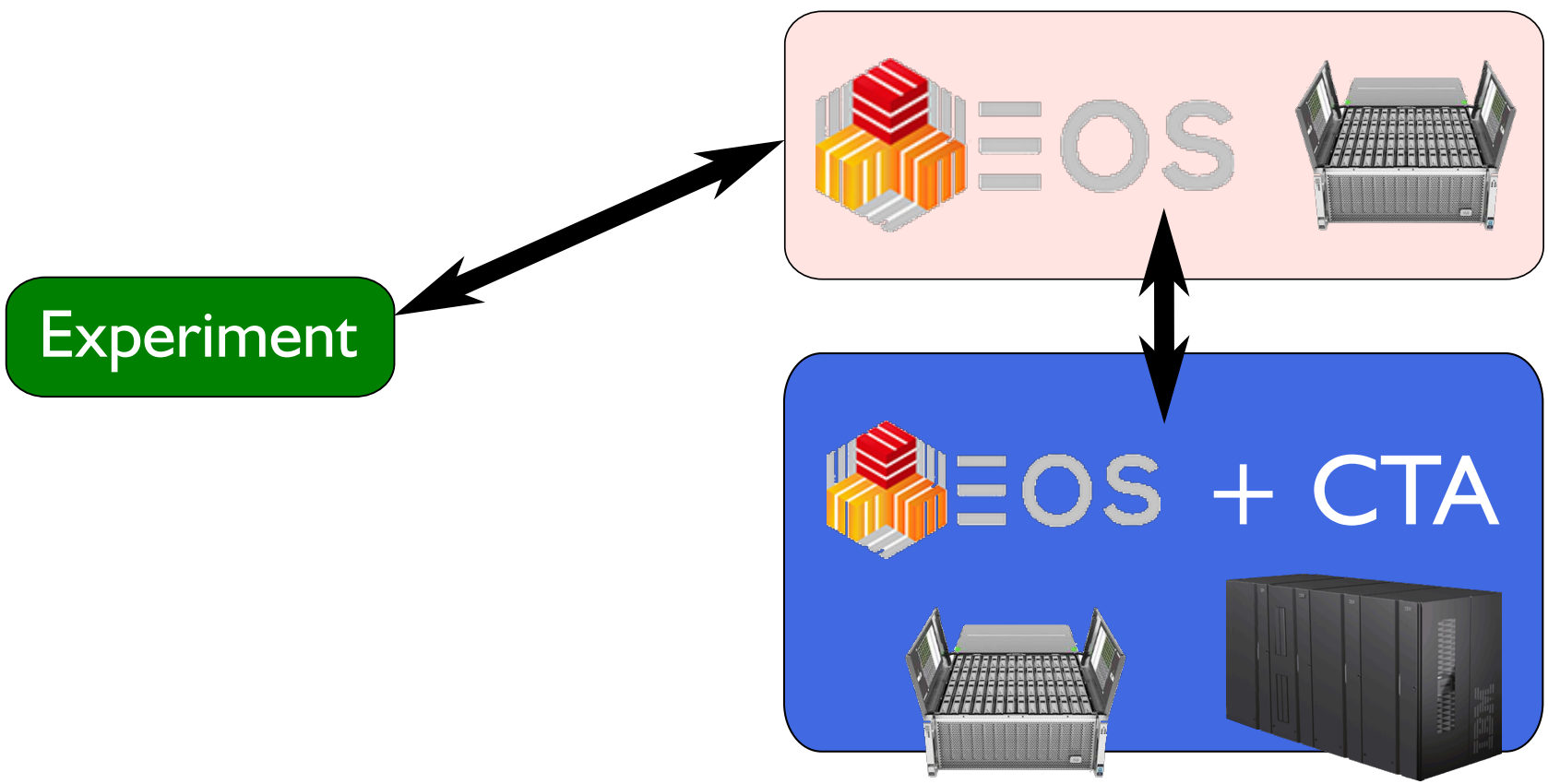
EOS+CTA
CMS



EOS+CTA
LHCb

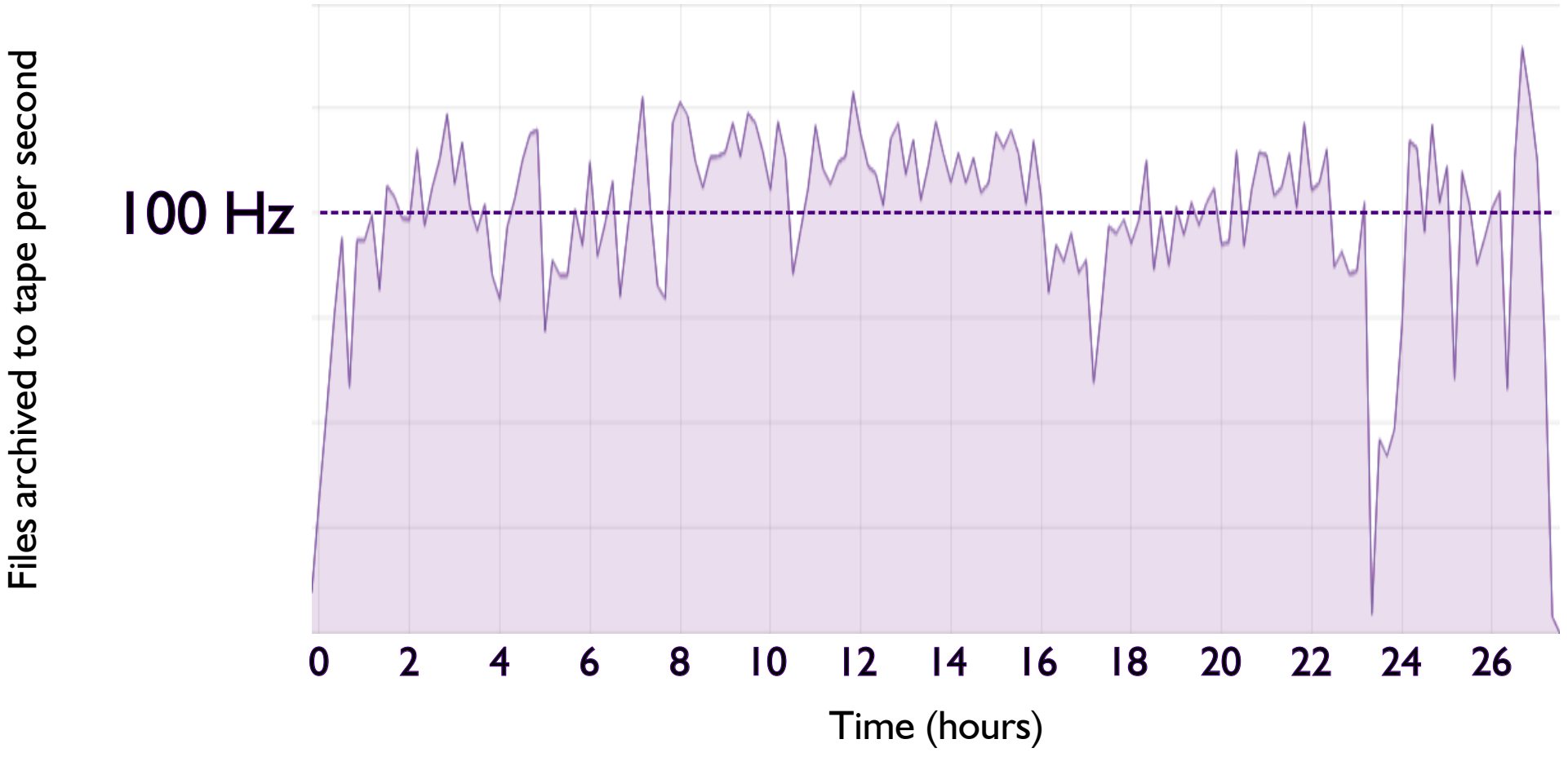


EOS+CTA
PUBLIC



System Testing

Scale tests and stress tests :
10 million files archived in ≈27 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-aee8-02163e01826d' FINISHED

VO: atlas

Total size	Done	Submission time	Start time	Running time	Avg. file throughput
976.56 KiB	976.56 KiB	2018-06-22T12:57:04Z	2018-06-22T12:57:05Z (+1s)	2 s	0.95 MB/s

File ID	File State	File Size	Throughput	Start Time	Finish Time
+ 20024	FINISHED	976.56 KiB	0.95 MB/s	2018-06-22T12:57:05Z	2018-06-22T12:57:07Z

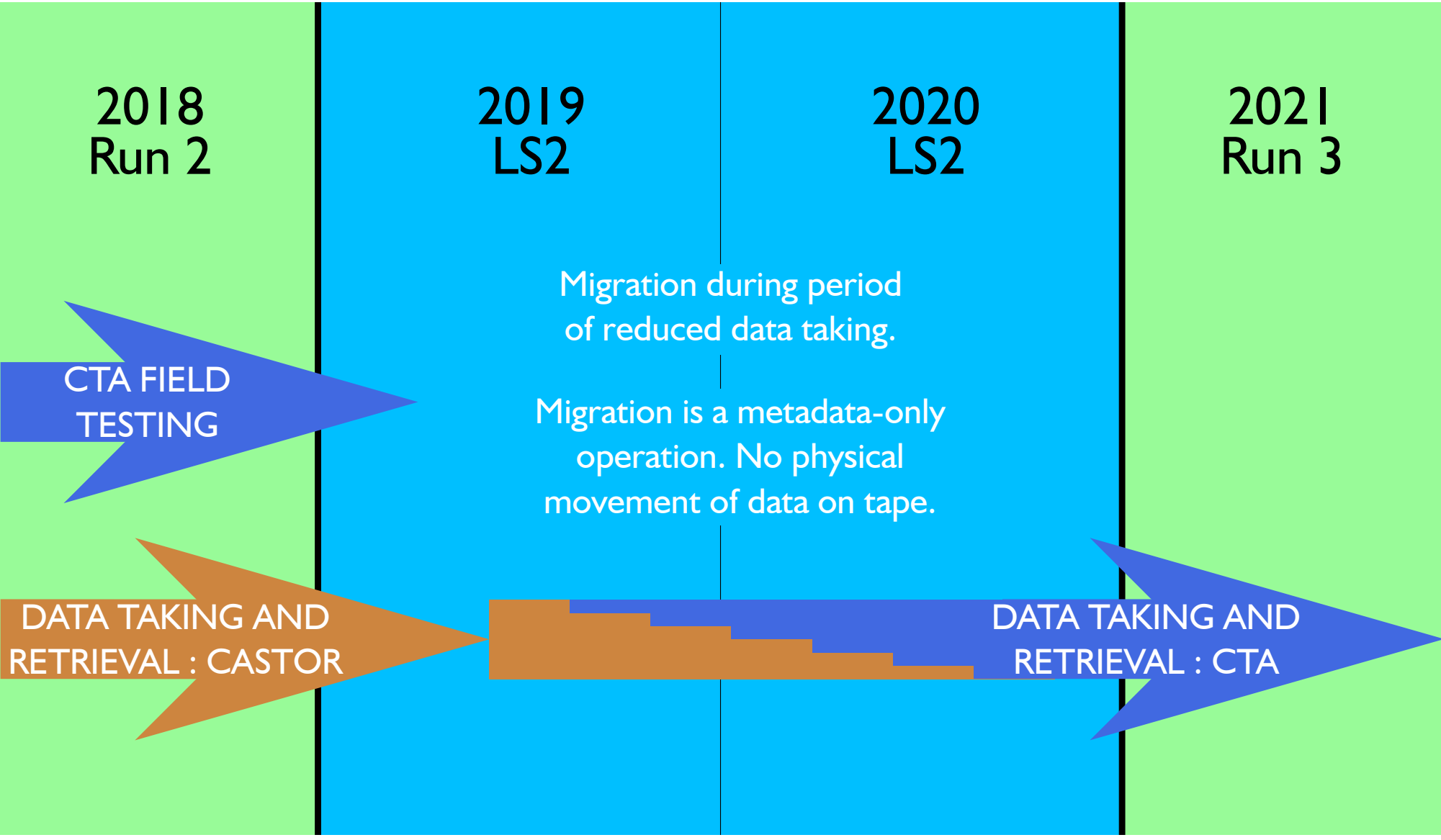
root://eosatlas.cern.ch//eos/atlas/atlasdatadisk/ruciotest/rucio/tests/f8/ee/AOD.9584b376f8c2476688a2c43d39b8e667

root://eosctaatlaspps.cern.ch//eos/ctaatlaspps/preprodvtl/AOD.9584b376f8c2476688a2c43d39b8e6444343

Next :

- Agree schedule for field testing with all CERN experiments

Migration Schedule



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