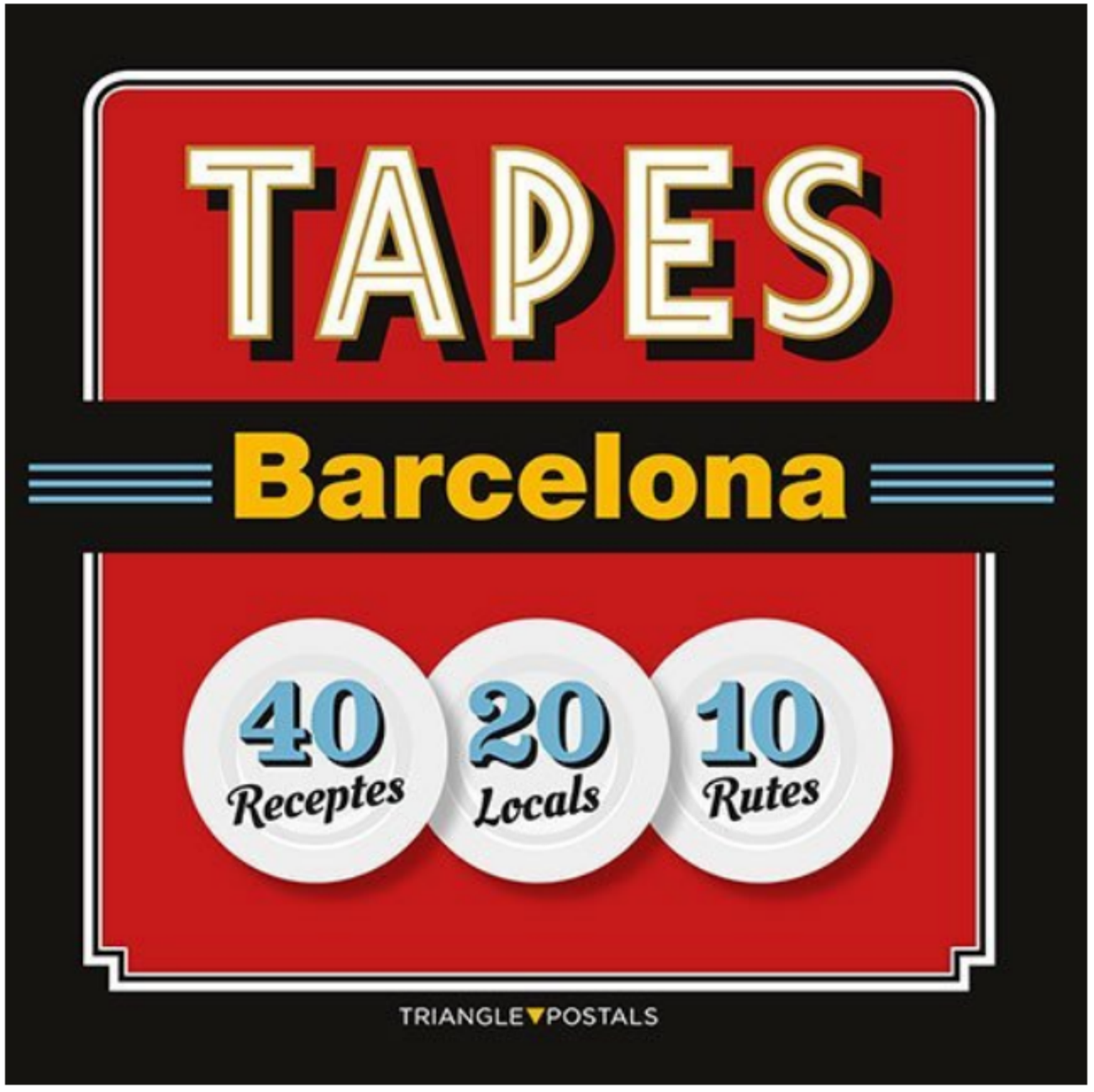
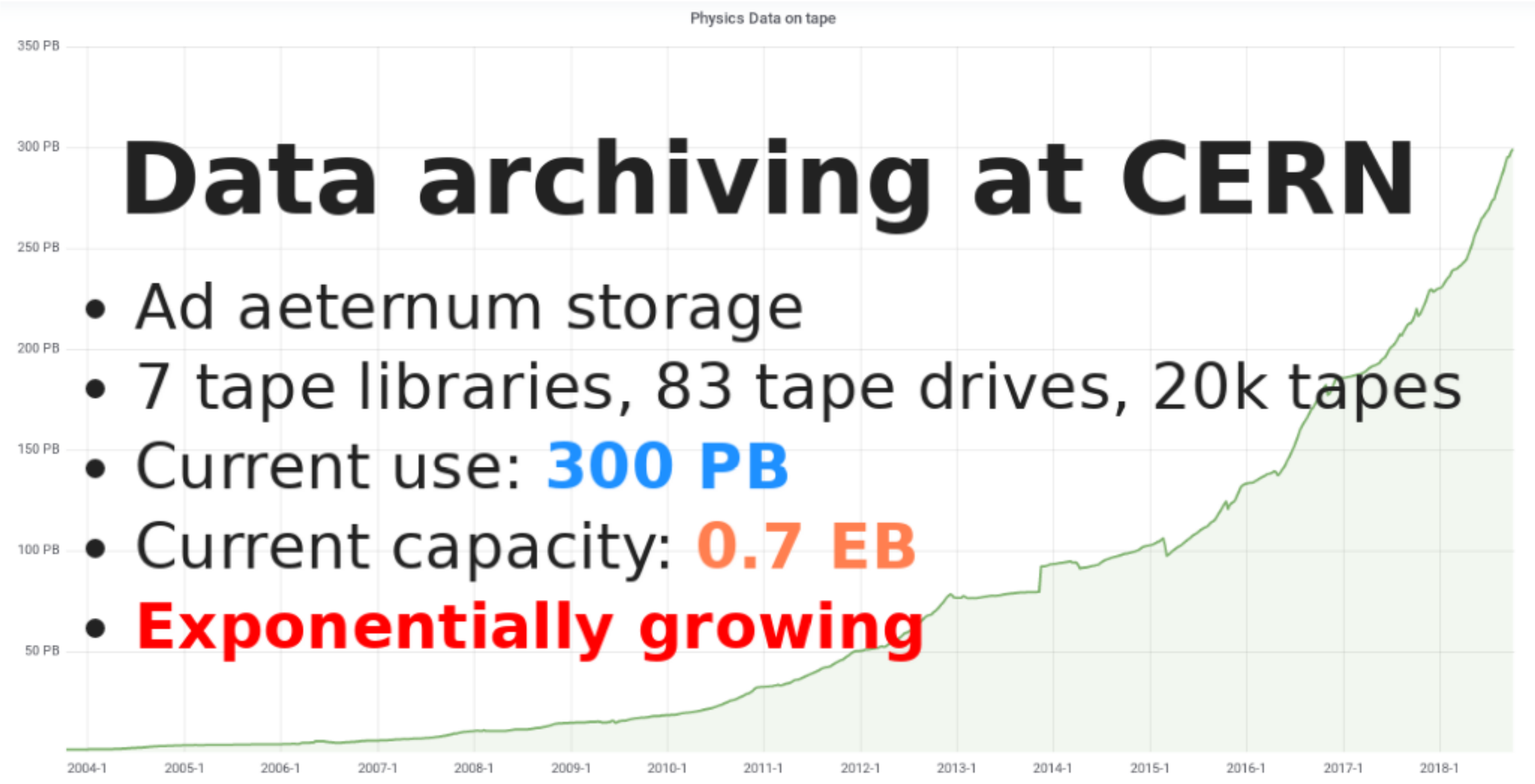
initial deployment

Julien Leduc


TAPES
Barcelona
40 Receptes
20 Locals
10 Rutes
TRIANGLE POSTALS

Data archiving at CERN

- Ad aeternum storage
- 7 tape libraries, 83 tape drives, 20k tapes
- Current use: **300 PB**
- Current capacity: **0.7 EB**
- **Exponentially growing**

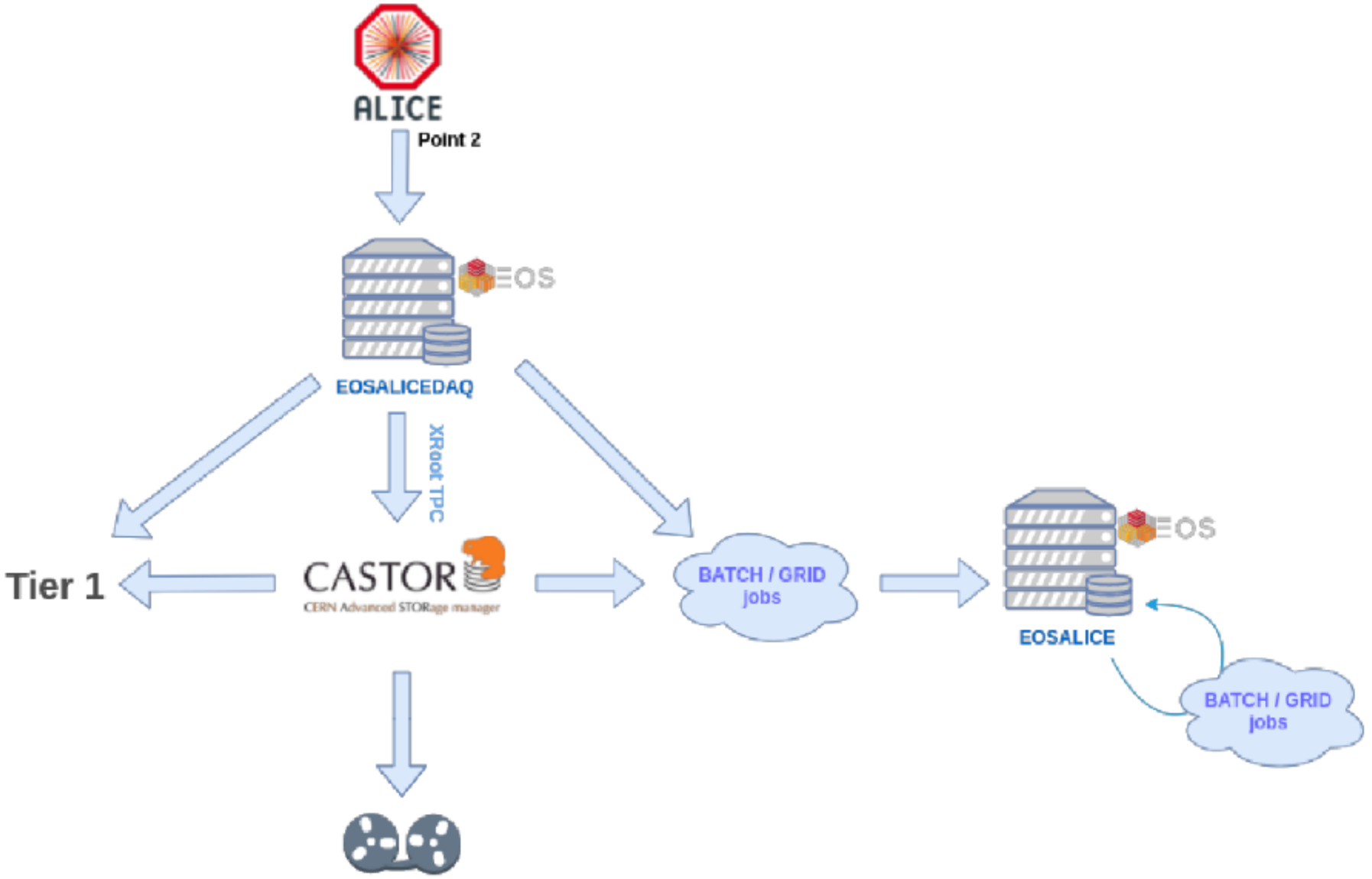


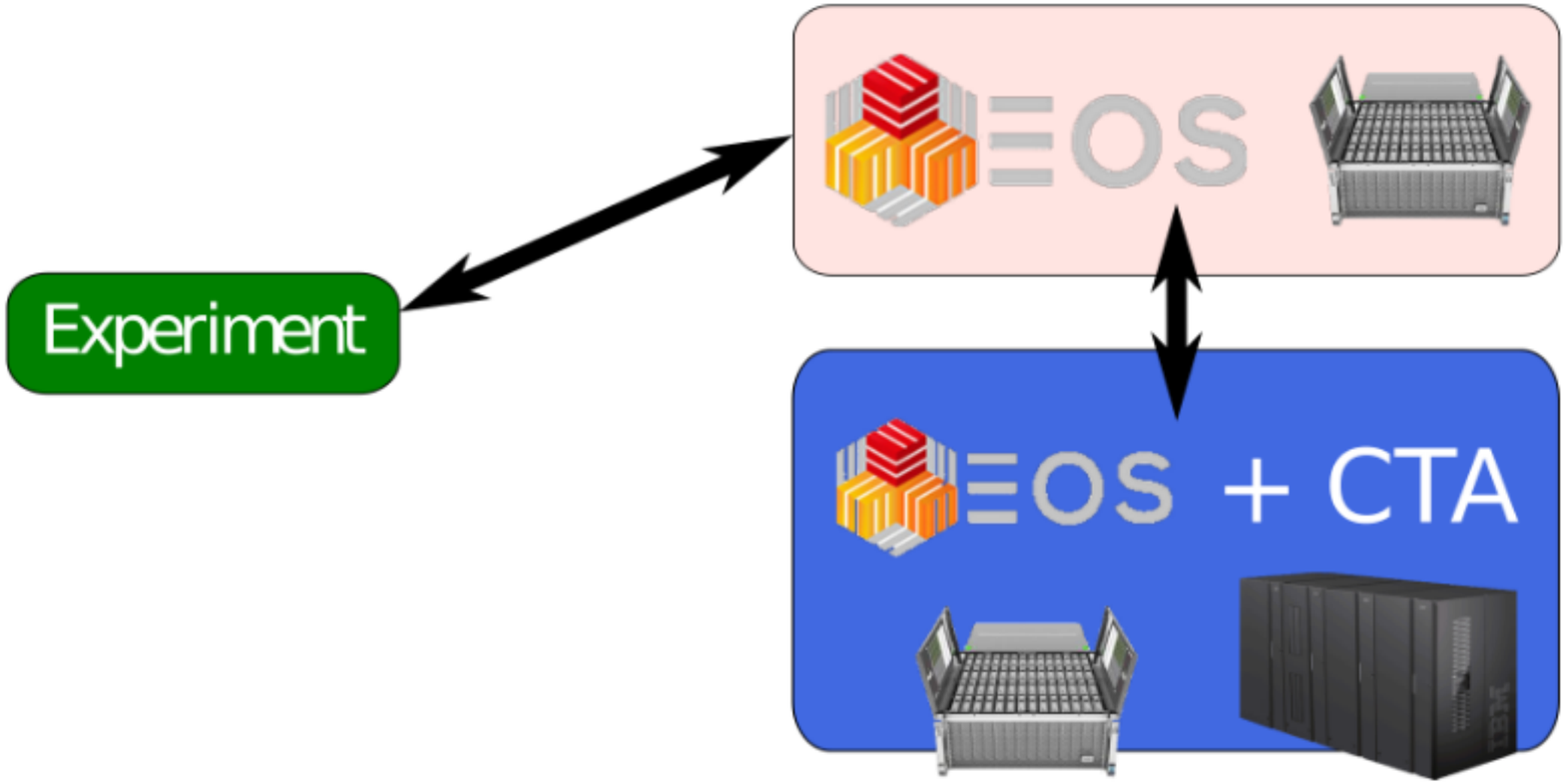
Data Archiving at CERN

Evolution

- EOS + tapes...
  - EOS is CERN strategic storage platform
  - tape is the strategic long term archive medium
- EOS + tapes = ♥
  - Meet CTA: CERN Tape Archive
  - Streamline data paths, software and infrastructure

EOS+CTA *Deployment*


ALICE
Point 2
EOS
EOSALICEDAQ
XRoot TPC
Tier 1
CASTOR
CERN Advanced STORage manager
BATCH / GRID
jobs
EOS
EOSALICE
BATCH / GRID
jobs


Experiment
EOS
EOS + CTA

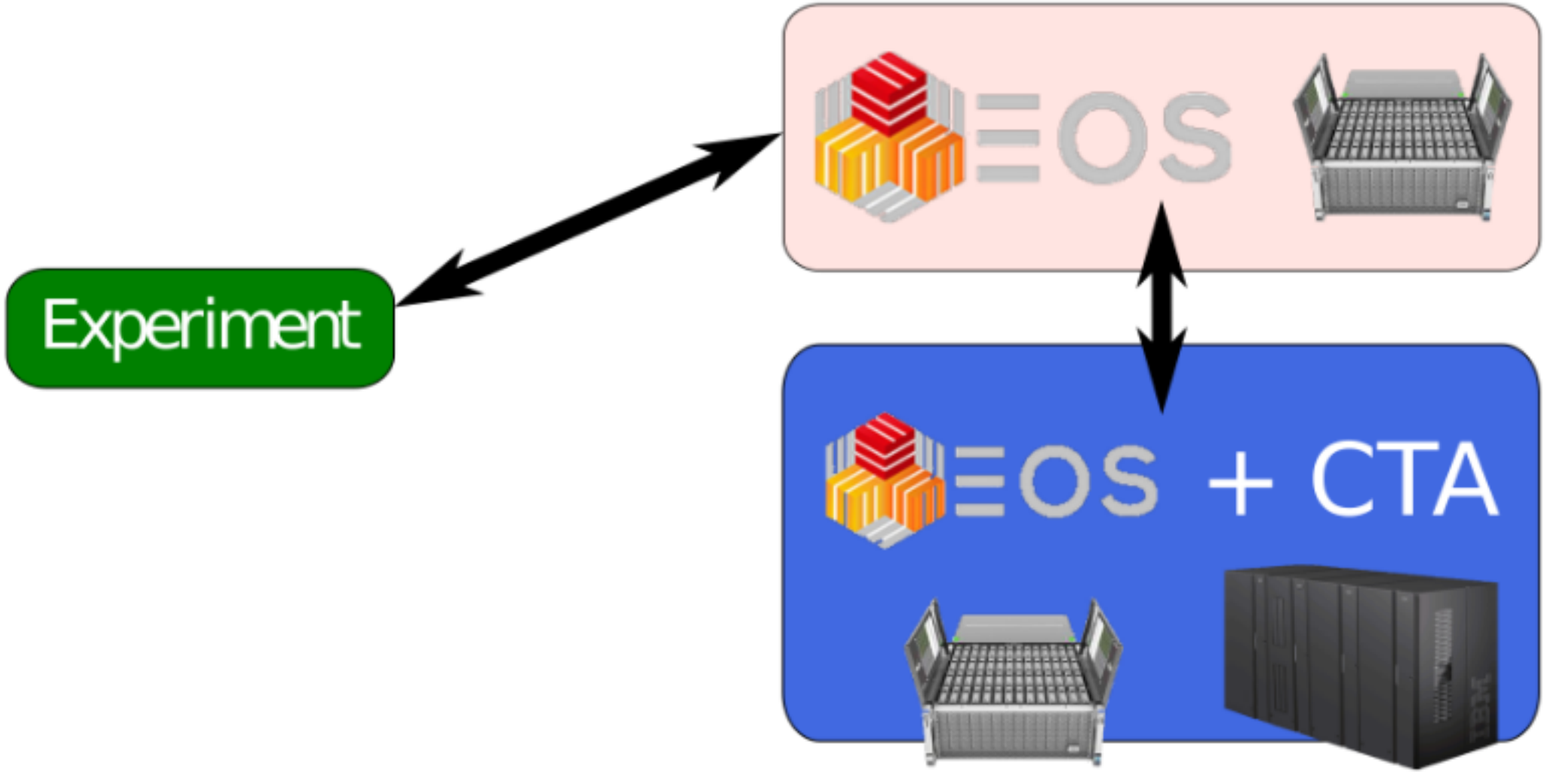
EOS+CTA
ALICE

EOS+CTA
ATLAS

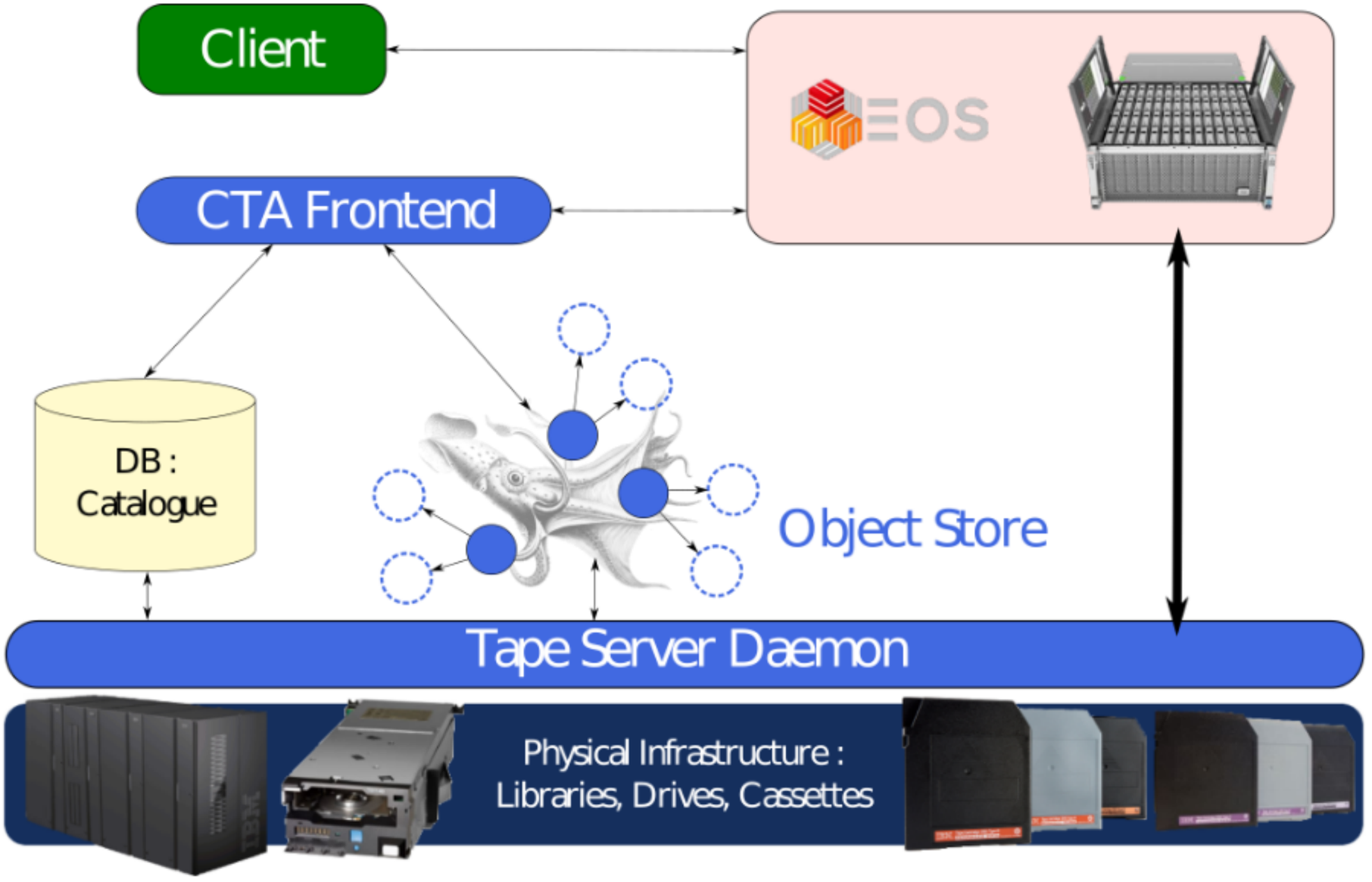
EOS+CTA
CMS

EOS+CTA
LHCb

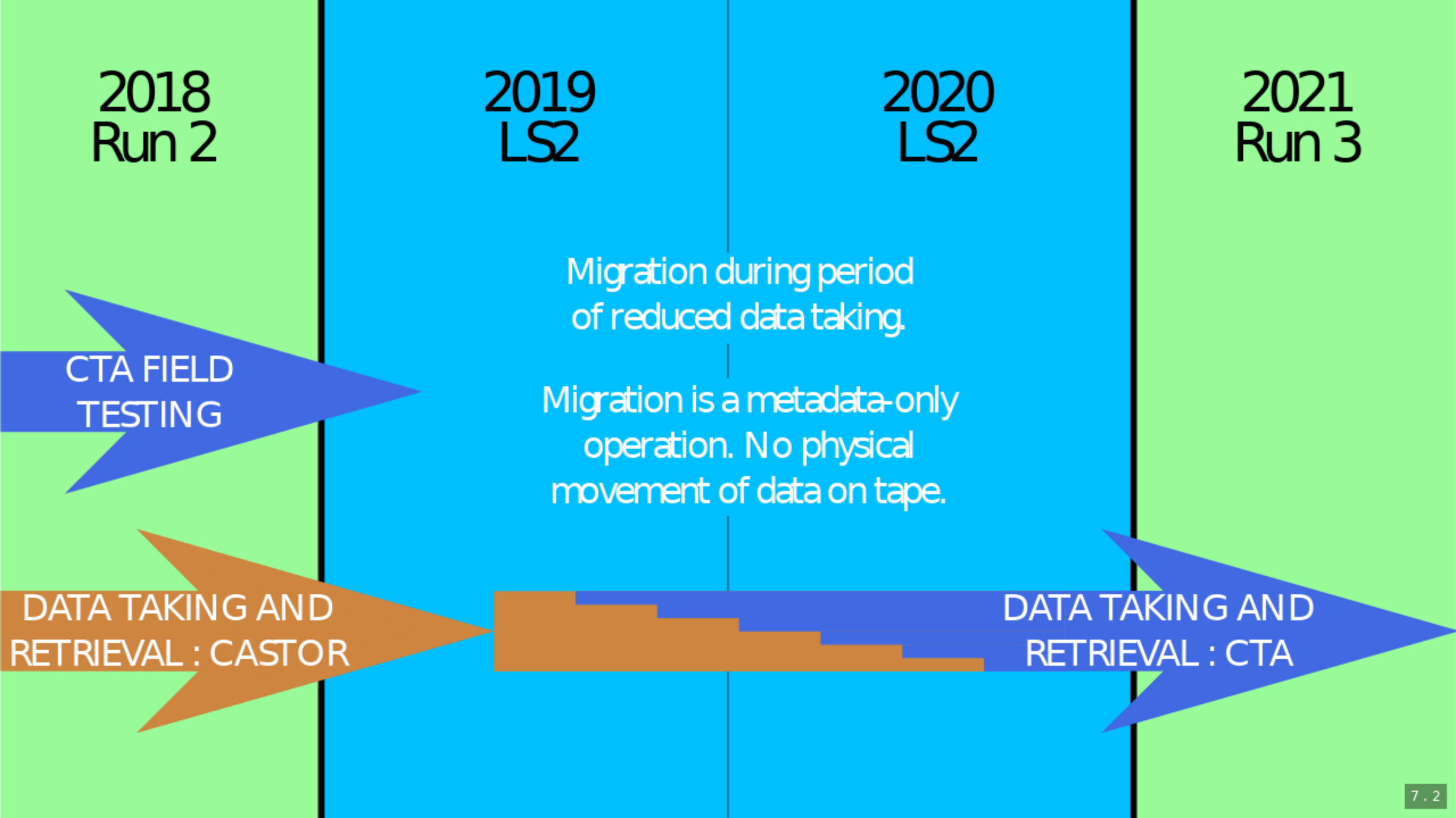
EOS+CTA
PUBLIC



EOS+CTA *Architecture*


Client
EOS
CTA Frontend
DB :
Catalogue
Object Store
Tape Server Daemon
Physical Infrastructure :
Libraries, Drives, Cassettes

EOS+CTA *Timeline*

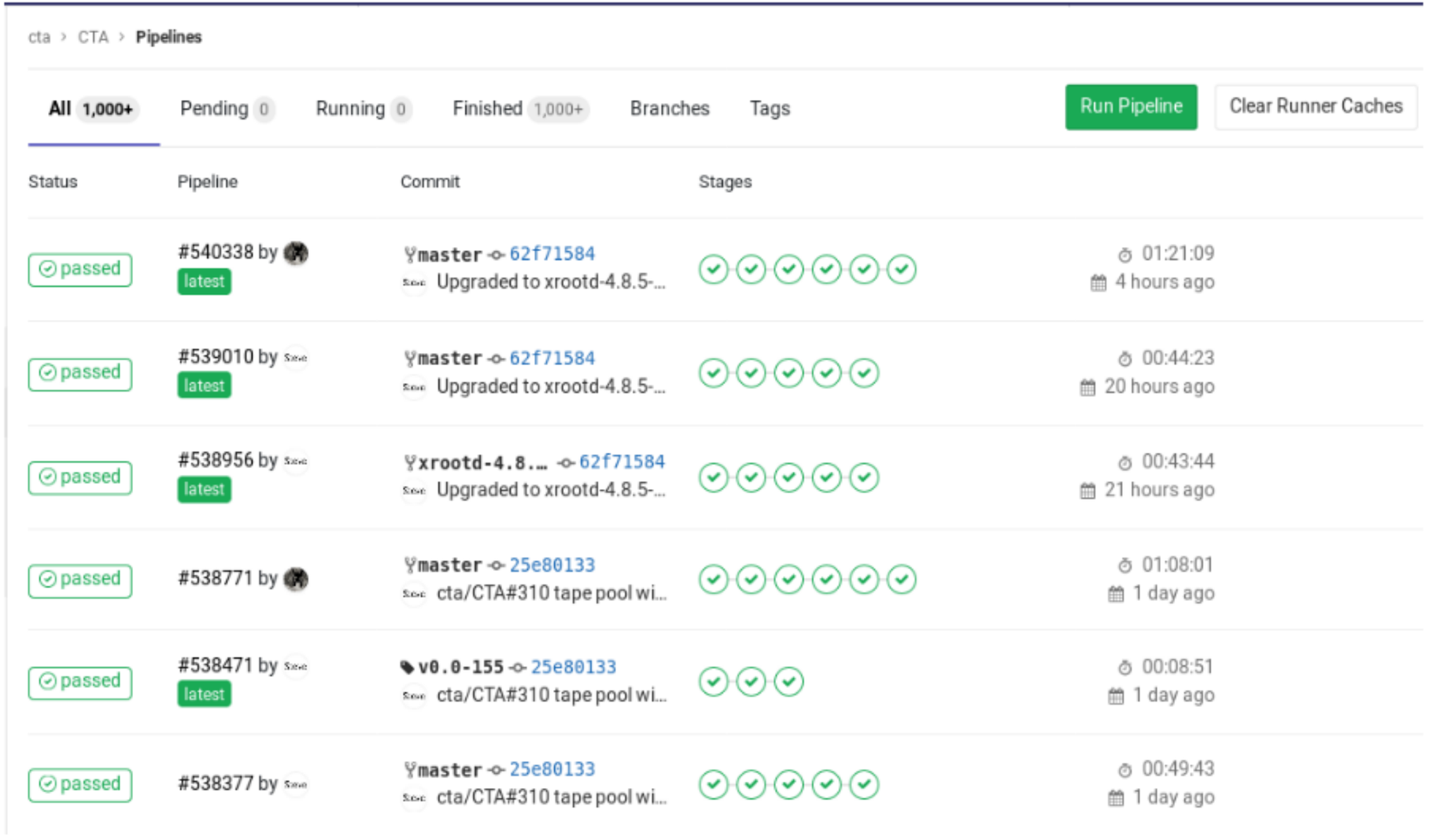

2018
Run 2
2019
LS2
2020
LS2
2021
Run 3
Migration during period
of reduced data taking.
Migration is a metadata-only
operation. No physical
movement of data on tape.
CTA FIELD
TESTING
DATA TAKING AND
RETRIEVAL : CASTOR
DATA TAKING AND
RETRIEVAL : CTA

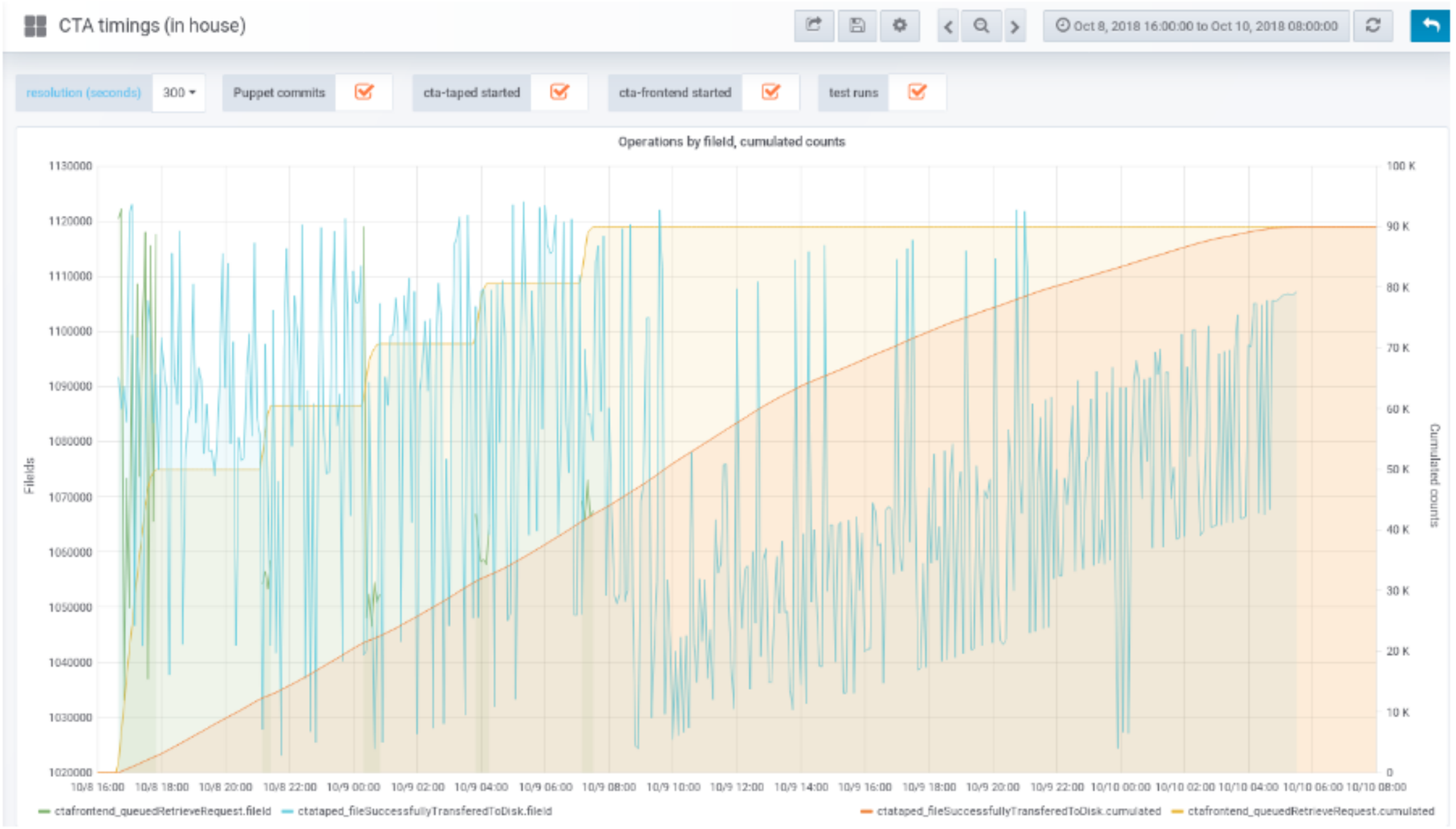
EOS+CTA *Dev&oper*

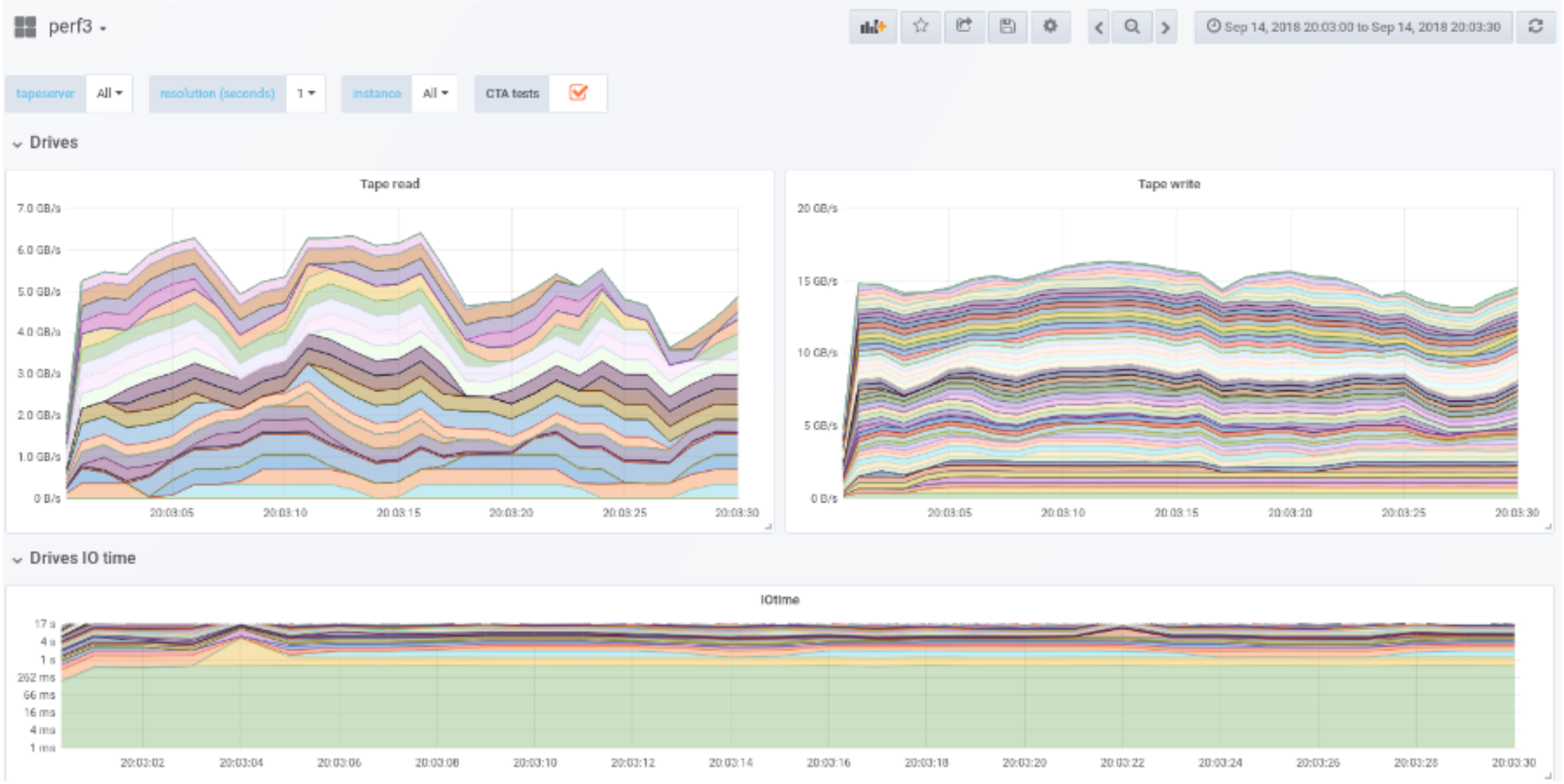
Tightly coupled software ⇒ tightly coupled developments

Extensive and systematic testing is paramount to limit regressions

Extensive monitoring in place to ease debugging and target high performance from day 1


cta › CTA › Pipelines
All 1,000+
Pending 0
Running 0
Finished 1,000+
Branches
Tags
Run Pipeline
Clear Runner Caches
Status
Pipeline
Commit
Stages
passed
#540338 by
latest
master 62f71584
Upgraded to xrootd-4.8.5-...
01:21:09
4 hours ago
passed
#539010 by
latest
master 62f71584
Upgraded to xrootd-4.8.5-...
00:44:23
20 hours ago
passed
#538956 by
latest
xrootd-4.8.... 62f71584
Upgraded to xrootd-4.8.5-...
00:43:44
21 hours ago
passed
#538771 by
master 25e80133
cta/CTA#310 tape pool wi...
01:08:01
1 day ago
passed
#538471 by
latest
v0.0-155 25e80133
cta/CTA#310 tape pool wi...
00:08:51
1 day ago
passed
#538377 by
master 25e80133
cta/CTA#310 tape pool wi...
00:49:43
1 day ago


CTA timings (in house)
Oct 8, 2018 16:00:00 to Oct 10, 2018 08:00:00
resolution (seconds)
300
Puppet commits
cta-taped started
cta-frontend started
test runs
Operations by fileId, cumulated counts
Fileids
Cumulated counts
1130000
1120000
1110000
1100000
1090000
1080000
1070000
1060000
1050000
1040000
1030000
1020000
100 K
90 K
80 K
70 K
60 K
50 K
40 K
30 K
20 K
10 K
0
10/8 16:00
10/8 18:00
10/8 20:00
10/8 22:00
10/9 00:00
10/9 02:00
10/9 04:00
10/9 06:00
10/9 08:00
10/9 10:00
10/9 12:00
10/9 14:00
10/9 16:00
10/9 18:00
10/9 20:00
10/9 22:00
10/10 00:00
10/10 02:00
10/10 04:00
10/10 06:00
10/10 08:00
ctafrontend_queuedRetrieveRequest.fileId
ctataped_fileSuccessfullyTransferedToDisk.fileId
ctataped_fileSuccessfullyTransferedToDisk.cumulated
ctafrontend_queuedRetrieveRequest.cumulated


perf3
Sep 14, 2018 20:03:00 to Sep 14, 2018 20:03:30
tapeserver
All
resolution (seconds)
1
instance
All
CTA tests
Drives
Tape read
7.0 GB/s
6.0 GB/s
5.0 GB/s
4.0 GB/s
3.0 GB/s
2.0 GB/s
1.0 GB/s
0 B/s
20:03:05
20:03:10
20:03:15
20:03:20
20:03:25
20:03:30
Tape write
20 GB/s
15 GB/s
10 GB/s
5 GB/s
0 B/s
20:03:05
20:03:10
20:03:15
20:03:20
20:03:25
20:03:30
Drives IO time
IOtime
17 s
4 s
1 s
262 ms
66 ms
16 ms
4 ms
1 ms
20:03:02
20:03:04
20:03:06
20:03:08
20:03:10
20:03:12
20:03:14
20:03:16
20:03:18
20:03:20
20:03:22
20:03:24
20:03:26
20:03:28
20:03:30

CTA metadata stress tests

Small EOS instance VTL backed

- Few retired 1Gb/s disk servers and a decent MGM for EOS
- A few VMs for the CTA tape part

Each VTL tape server exposes 20 virtual tape drives and can sustain >1kHz 1kB files

CTA metadata stress tests *Stage in*

Tests until CTA routinely archived 1M files

- **10M file test**:

CTA metadata stress tests *Stage out*

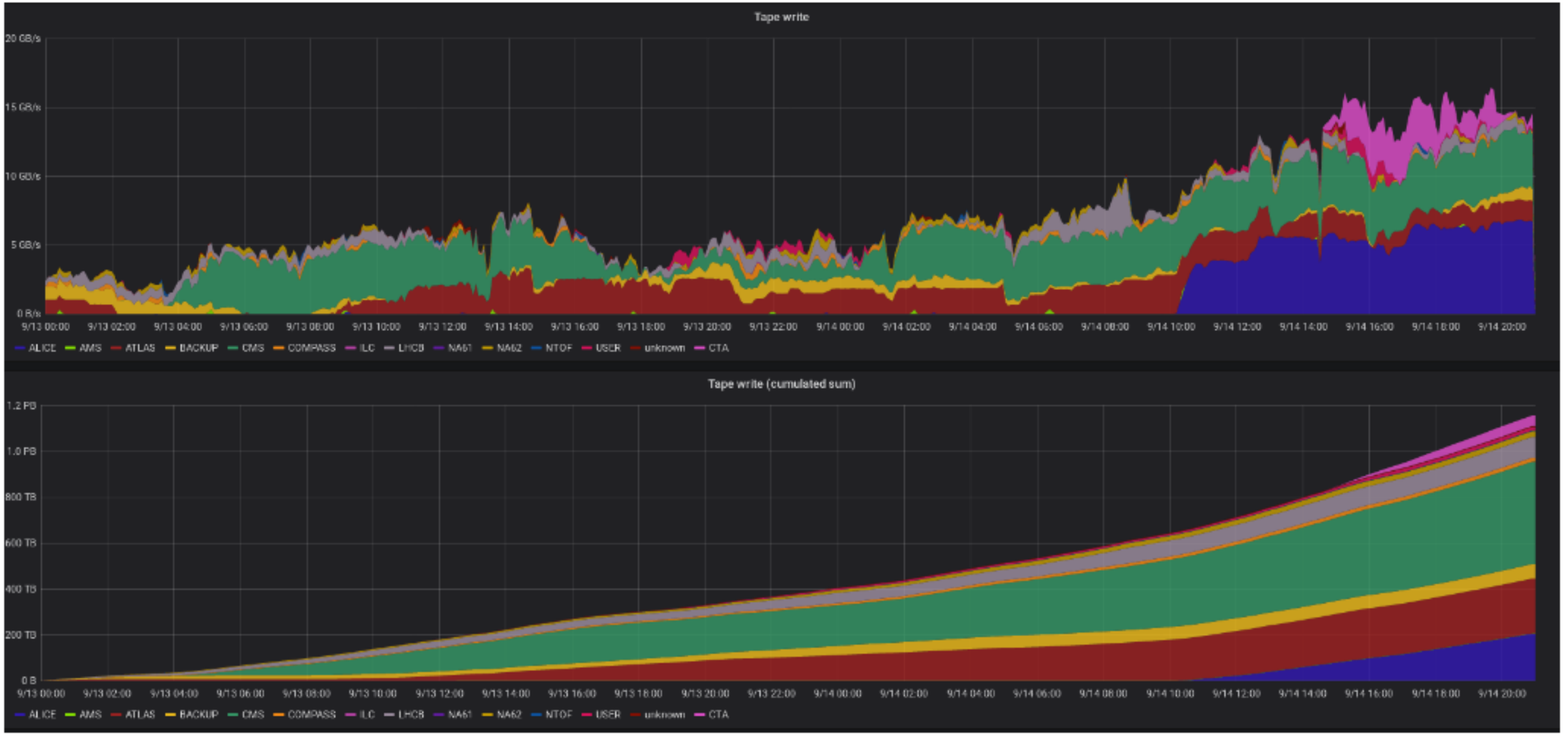
Tests until CTA routinely retrieved 1M files

- **4M file test**:

CTA HI Data challenge

Bigger EOS instance

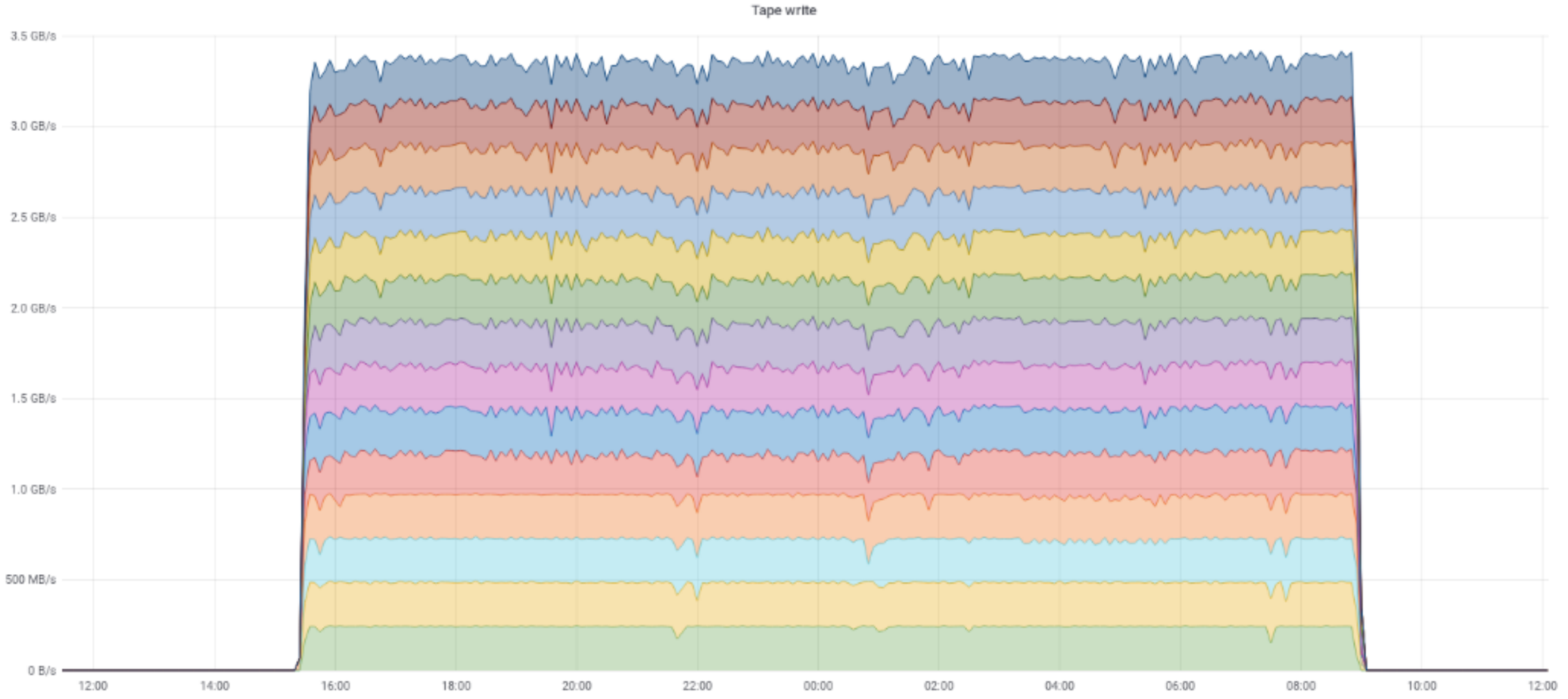
- 10x10Gb/s disk servers and a decent MGM for EOS (~2.4PB)
- 14xT10KD tape drives for the CTA tape part (3.5GB/s)


Tape write
20 GB/s
15 GB/s
10 GB/s
5 GB/s
0 B/s
9/13 00:00
9/13 02:00
9/13 04:00
9/13 06:00
9/13 08:00
9/13 10:00
9/13 12:00
9/13 14:00
9/13 16:00
9/13 18:00
9/13 20:00
9/13 22:00
9/14 00:00
9/14 02:00
9/14 04:00
9/14 06:00
9/14 08:00
9/14 10:00
9/14 12:00
9/14 14:00
9/14 16:00
9/14 18:00
9/14 20:00
ALICE AMS ATLAS BACKUP CMS COMPASS ILC LHCB NA61 NA62 NTOF USER unknown CTA
Tape write (cumulated sum)
1.2 PB
1.0 PB
800 TB
600 TB
400 TB
200 TB
0 B
9/13 00:00
9/13 02:00
9/13 04:00
9/13 06:00
9/13 08:00
9/13 10:00
9/13 12:00
9/13 14:00
9/13 16:00
9/13 18:00
9/13 20:00
9/13 22:00
9/14 00:00
9/14 02:00
9/14 04:00
9/14 06:00
9/14 08:00
9/14 10:00
9/14 12:00
9/14 14:00
9/14 16:00
9/14 18:00
9/14 20:00
ALICE AMS ATLAS BACKUP CMS COMPASS ILC LHCB NA61 NA62 NTOF USER unknown CTA

CTA HI Data challenge

Extended

Can we saturate the tape drive speed longer?


Tape write
3.5 GB/s
3.0 GB/s
2.5 GB/s
2.0 GB/s
1.5 GB/s
1.0 GB/s
500 MB/s
0 B/s
12:00
14:00
16:00
18:00
20:00
22:00
00:00
02:00
04:00
06:00
08:00
10:00
12:00

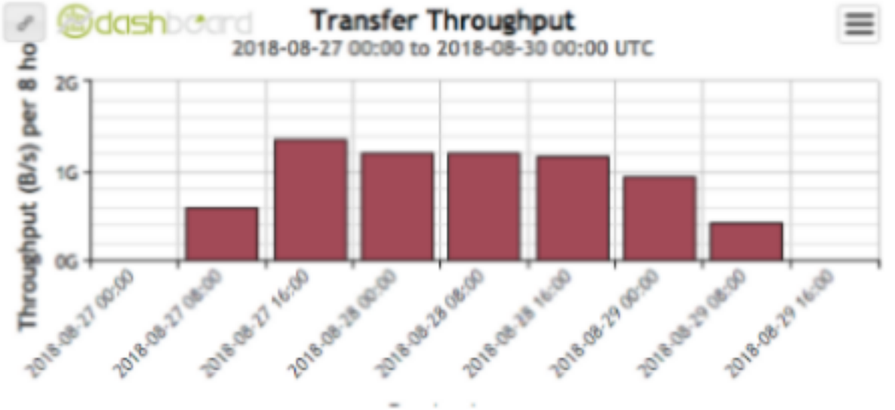
CTA VS experiment data transfers

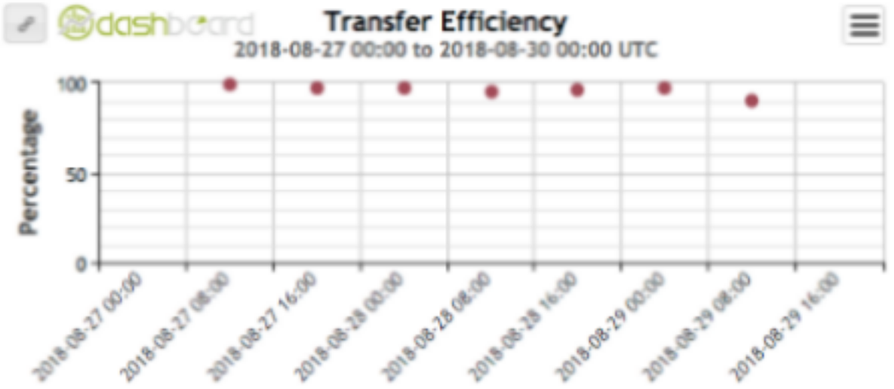
ATLAS stage in

Several tests conducted with Atlas DDM team using Rucio and FTS.

- 2 stage in tests of 200TB each
  - ~90k files of 2.6GB archived to tape
  - sub-optimal EOS instance (2 slow disk servers)

ATLAS stage in



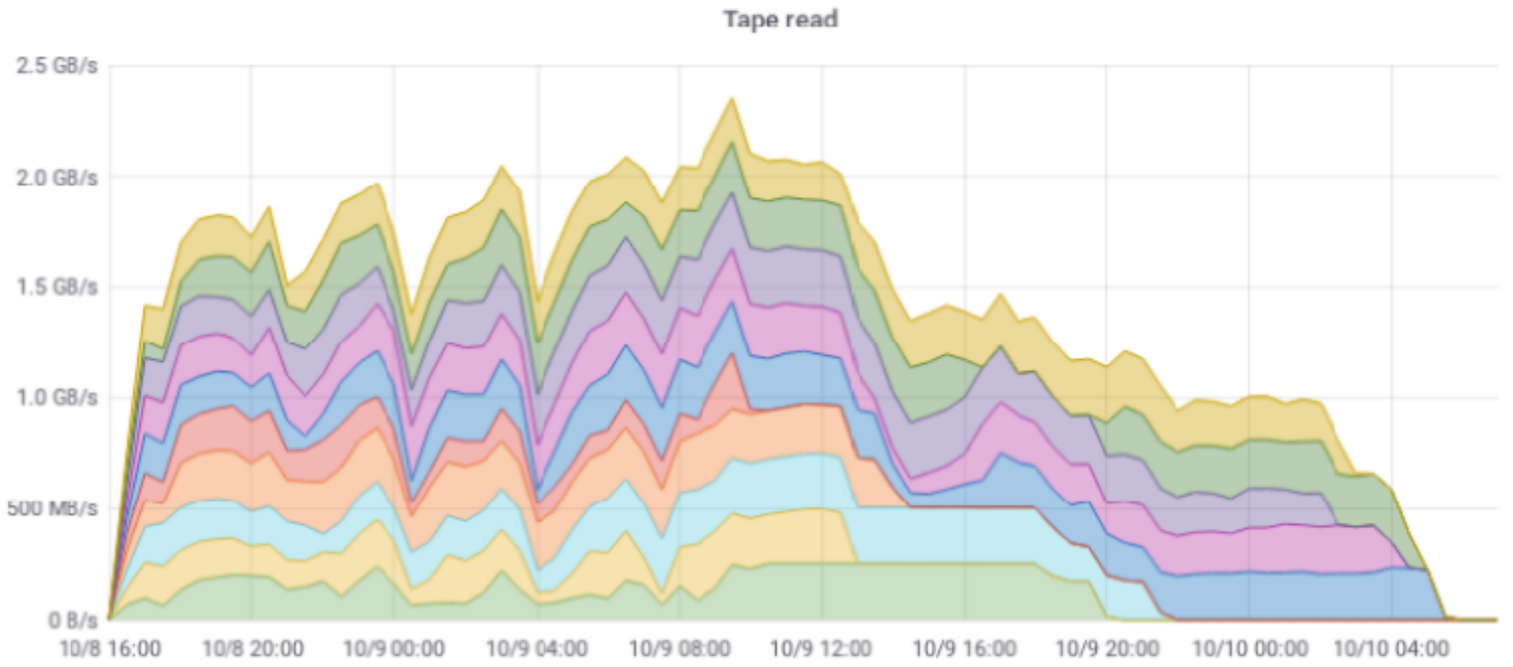


ATLAS stage out

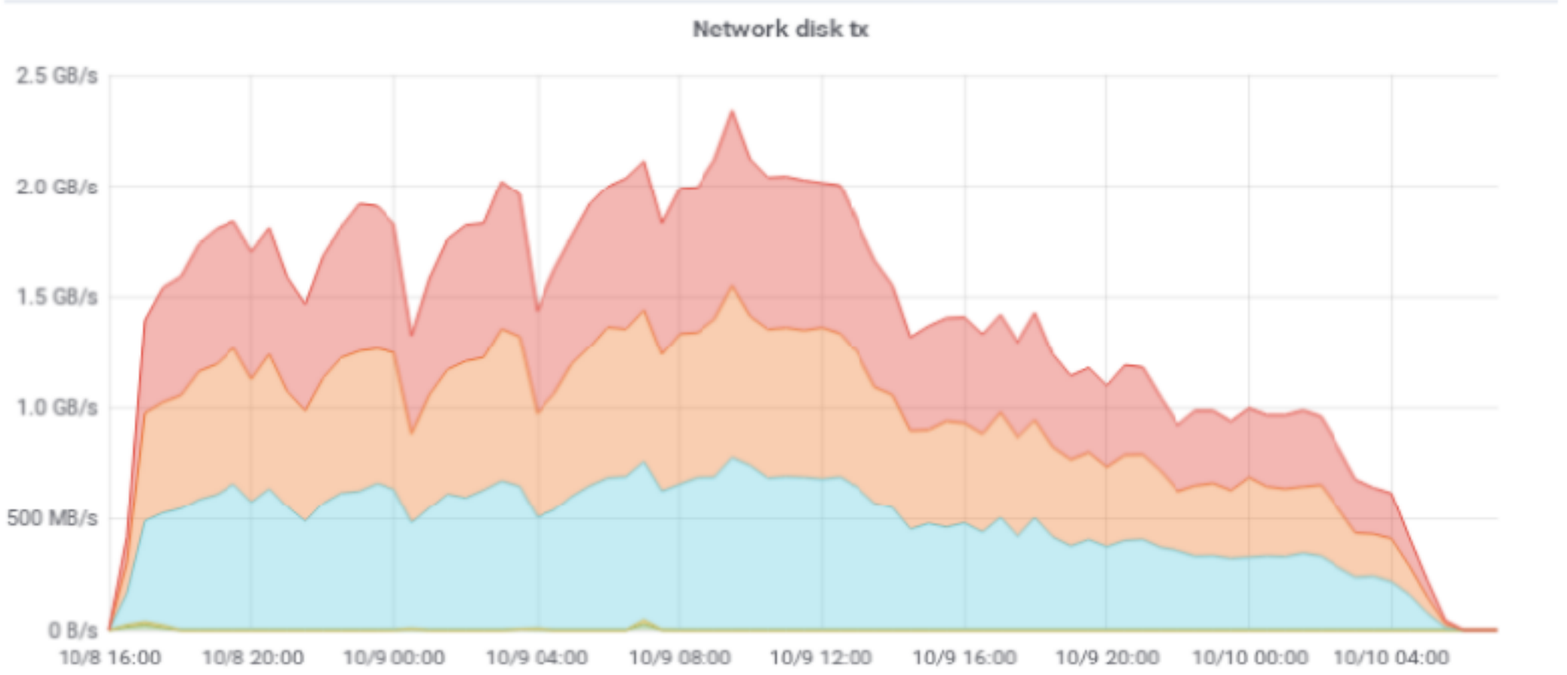
aka *Tape carousel* test took place during Hepix:

- 3 x EOS disk servers (~3x260TB of raw JBOD space)
- 6-10 x T10KD tape drives
- 90k files retrieved from EOSCTAATLASPPS (tape) to EOSATLAS by rucio through FTS

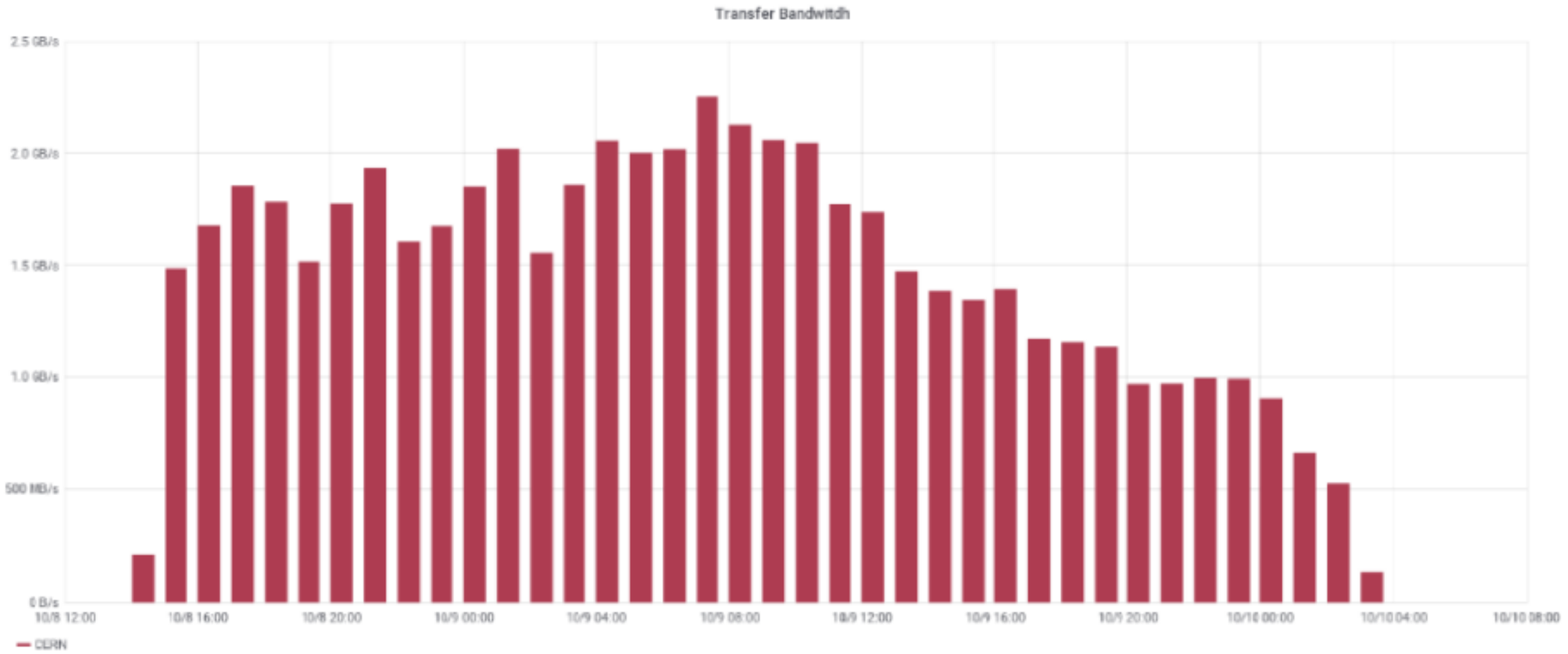
ATLAS stage out

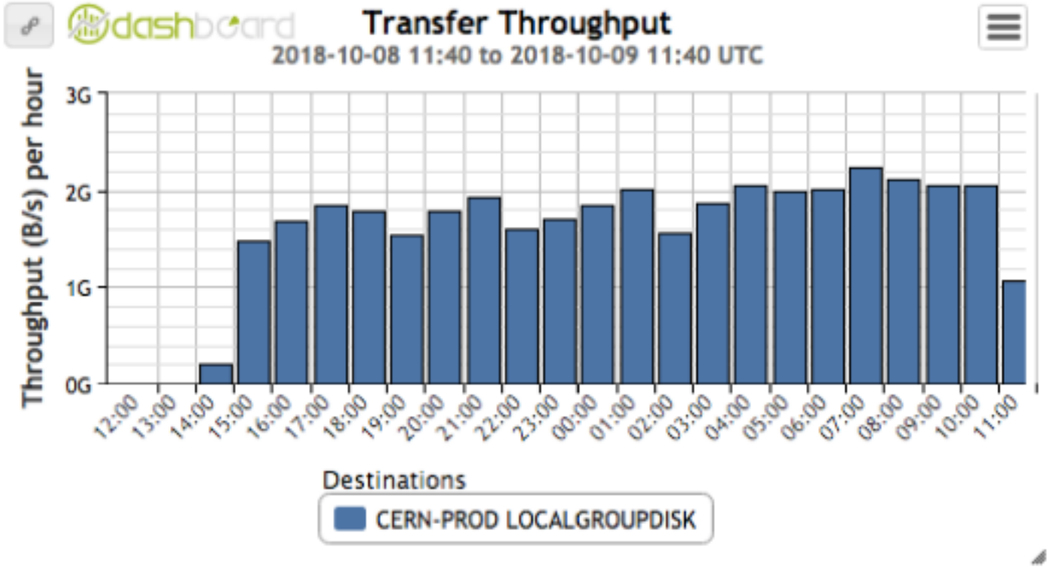


ATLAS stage out



ATLAS stage out DDM




dashboard
Transfer Throughput
2018-10-08 11:40 to 2018-10-09 11:40 UTC
Throughput (B/s) per hour
3G
2G
1G
0G
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
00:00
01:00
02:00
03:00
04:00
05:00
06:00
07:00
08:00
09:00
10:00
11:00
Destinations
CERN-PROD LOCALGROUPDISK