The great migration west
Decommissioning the Wigner Data Centre
Oliver Keeble on behalf of the EOS team

IT Overview ▾

⏱ Last 24 hours

COMPUTING		STORAGE		NETWORK	
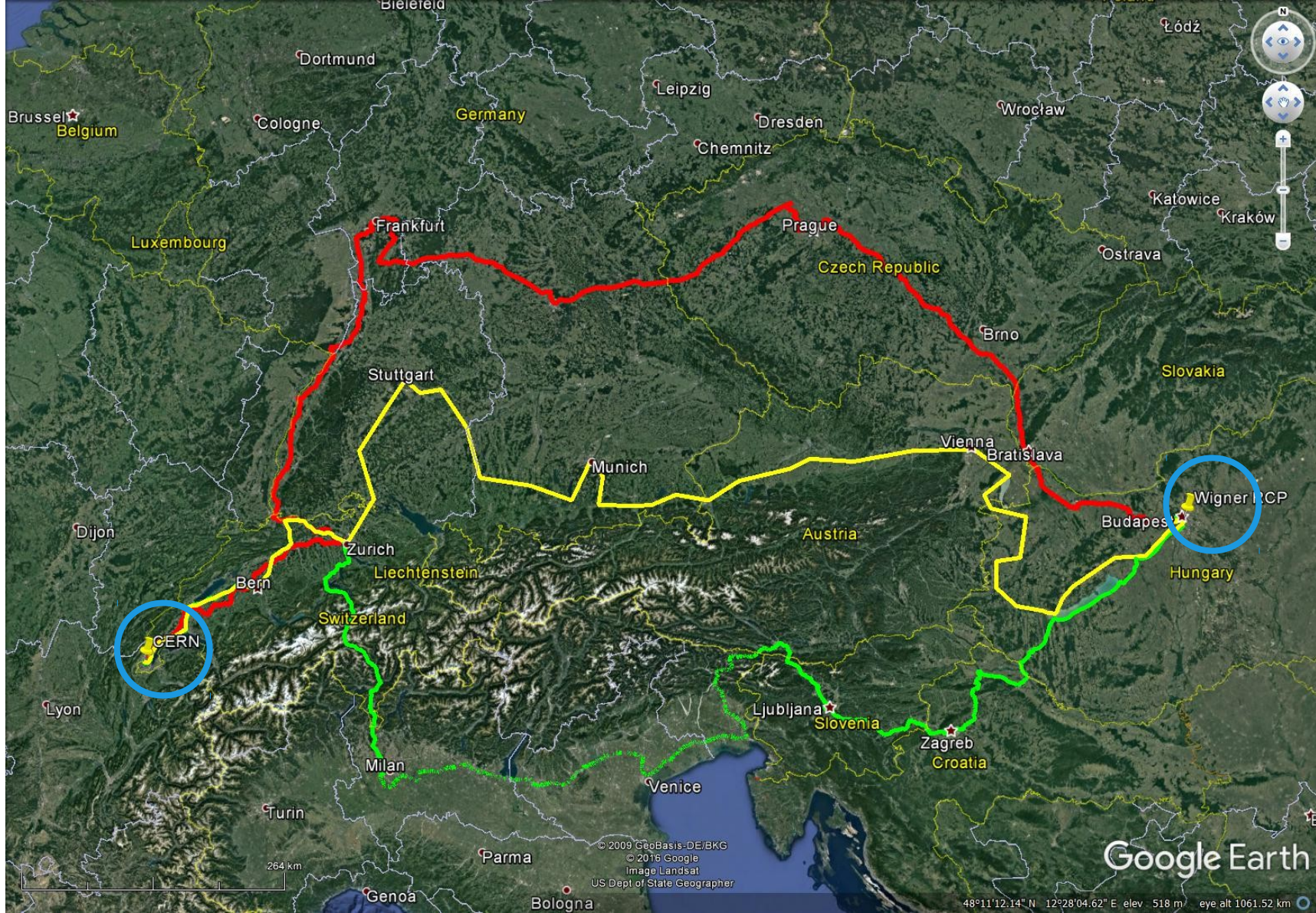
COMPUTING

Servers (Meyrin)	Cores (Meyrin)	Disks (Meyrin)	Tape Drives	Routers	Star Points
11.5 K	**174.3 K**	**61.9 K**	**102**	**258**	**701**

Servers (Wigner)	Cores (Wigner)	Disks (Wigner)	Tape Cartridges	Switches	Wifi Points
3.5 K	**56.0 K**	**29.7 K**	**31.7 K**	**4.1 K**	**772**

Batch Jobs

200 K, 150 K, 100 K, 50 K, 0 — 16:00 20:00 00:00 04:00 08:00 12:00

— atlas — cms — lhcb — alice — other
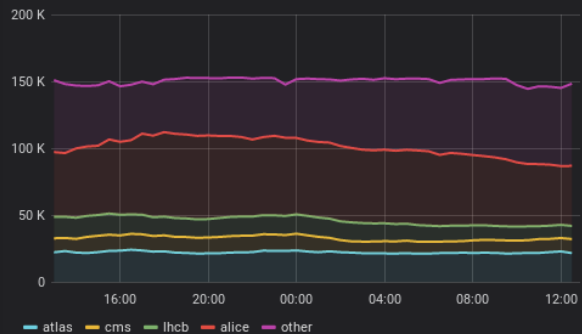
EOS Active Data Transfers

100 K, 80 K, 60 K, 40 K, 20 K, 0 — 16:00 20:00 00:00 04:00 08:00 12:00

— alice — atlas — cms — lhcb — public
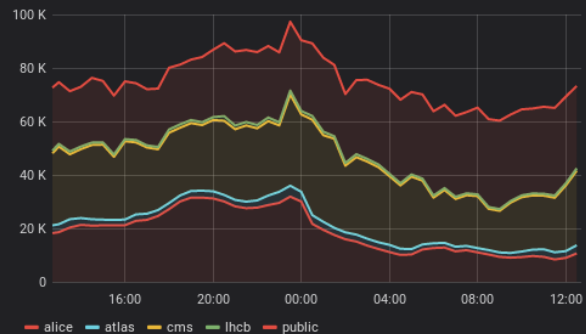
File Transfer Throughput

25 Gbps, 20 Gbps, 15 Gbps, 10 Gbps, 5 Gbps, 0 bps — 16:00 20:00 00:00 04:00 08:00 12:00

— atlas — cms — lhcb

Cloud Virtual Machines

1.0 K, 800, 600, 400, 200 — 16:00 20:00 00:00 04:00 08:00 12:00

— vms created per half hour
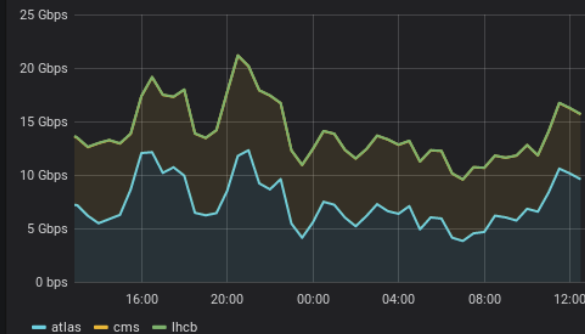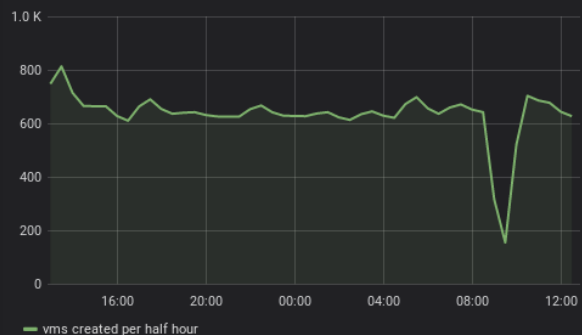
Databases Transactions

6.0 K, 5.0 K, 4.0 K, 3.0 K, 2.0 K, 1.0 K, 0 — 16:00 20:00 00:00 04:00 08:00 12:00

— transactions per second

Wigner Network Links

300 Gbps, 250 Gbps, 200 Gbps, 150 Gbps, 100 Gbps, 50 Gbps, 0 bps — 16:00 20:00 00:00 04:00 08:00 12:00

— to cern — to wigner
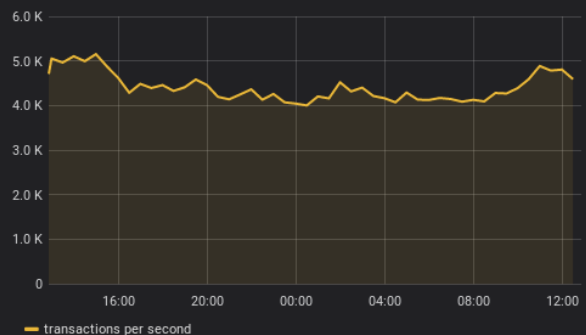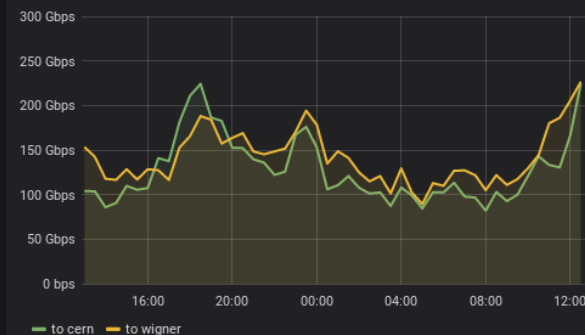
The mission

- Repatriate all data and most capacity before the end of 2019
  - 12 months
  - 68PB of data, 90PB of capacity
- Maintain availability of the service and data
- Respect the pledges
  - We guarantee a certain capacity to the LHC experiments
- Nodes would be shipped back here ("Meyrin") in batches
  - Then added back to the system

Capacity by EOS instance

Instance	Wigner Used	Meyrin Free	Difference
Alice	12.4PB	10.8PB	-1.7PB
Atlas	22.1PB	10.0PB	-12.2PB
Backup	5.7PB	6.8PB	+1.1PB
CMS	11.4PB	11.1PB	-0.2PB
Genome	62.1TB	66.6TB	+4.5TB
LHCb	5.1PB	5.2PB	+0.1PB
PPS	2.2TB	108.5TB	+100TB
Public	11.5PB	2.4PB	-9.0PB

Procurements : +20PB +50PB (servers) + 10PB (bare disk)
Wigner repatriation : +40PB (tbc…)
513 retirement. : -8 PB

Wigner : All_Batches
Generated : 16:46 24 Jan

- All6TB
- All9918Remaini…
- Second4TB
- NetHubPlus
- WD_4TB

The tools

eos node config fst.cern.ch:1095
configstatus=drain

Considerations

- We operate double replica, so a WAN transfer should often be avoidable

- Geo scheduling policies
  - Avoid draining back to Wigner
  - Block writing to Wigner
    - but… what about the pledge?

- Free space per group
  - One EOS instance is 192 small EOS instances

- Assignment of new capacity
  - Large machines (up to 2.3PB, 192 fs)
  - One fs (==sched group) per machine
  - Not much room for manoeuvre

atlas_fullgroups.txt : 09:17 07 Oct
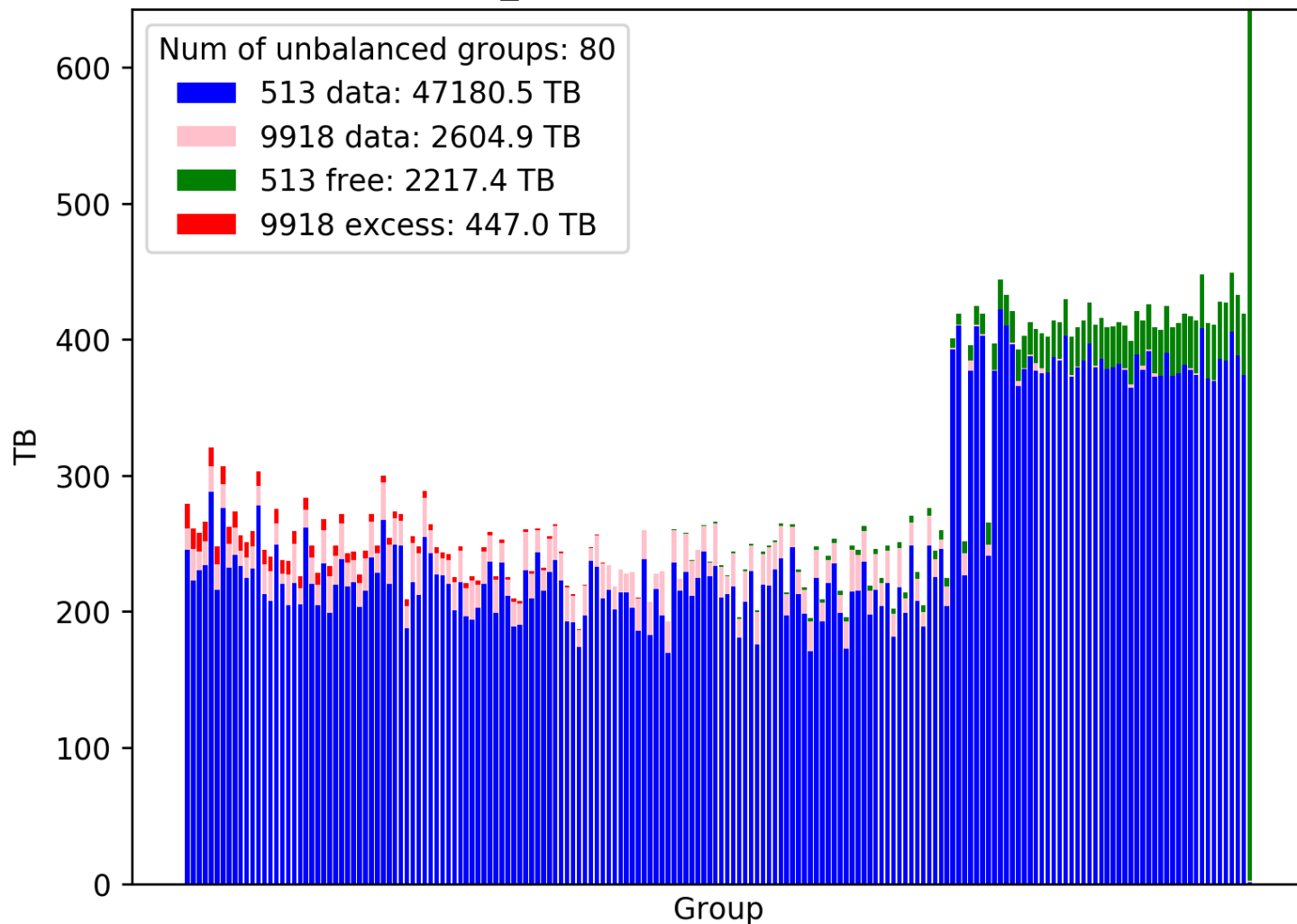
Num of unbalanced groups: 80
- 513 data: 47180.5 TB
- 9918 data: 2604.9 TB
- 513 free: 2217.4 TB
- 9918 excess: 447.0 TB
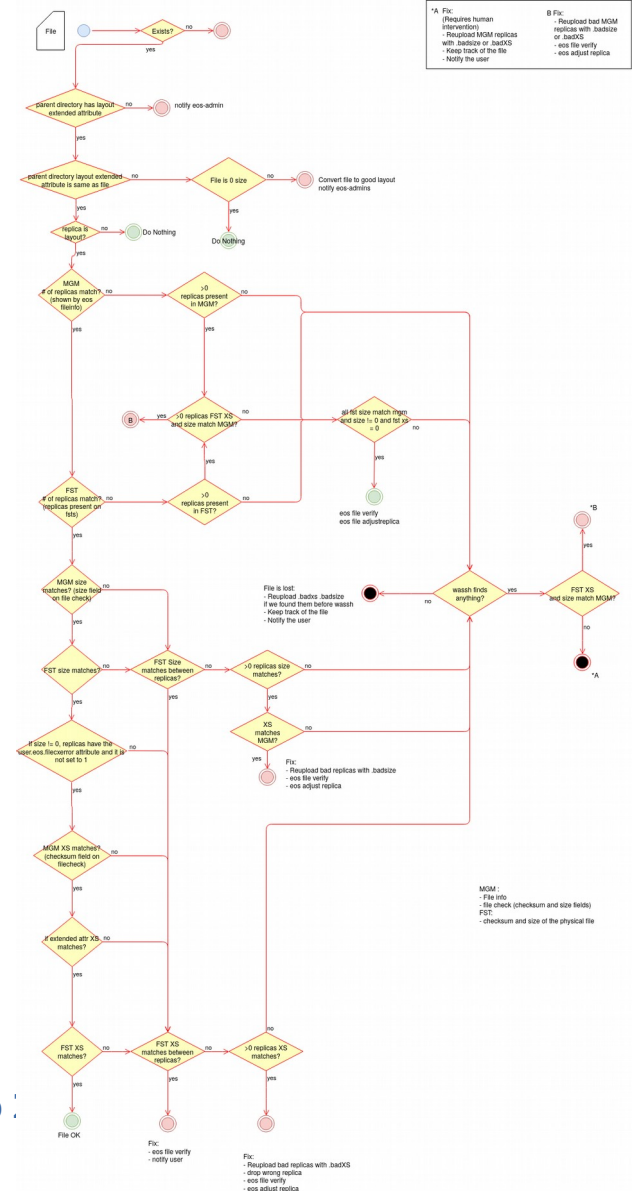
513 == Meyrin
9918 == Wigner

10

Barrel scraping

- Many instances have accumulated a lot of history
  - Have seen a lot of EOS releases
  - Have suffered a bug or two
  - Data has been moved around
- Draining reveals the residue of all these issues
  - Have to be fixed to complete the drain

Classifying issues

- Always left with some files that didn't drain. Why??
  - How many replicas?
  - Which checksums are there?
  - How big is the file? zero?
  - Size on disk? On MGM?
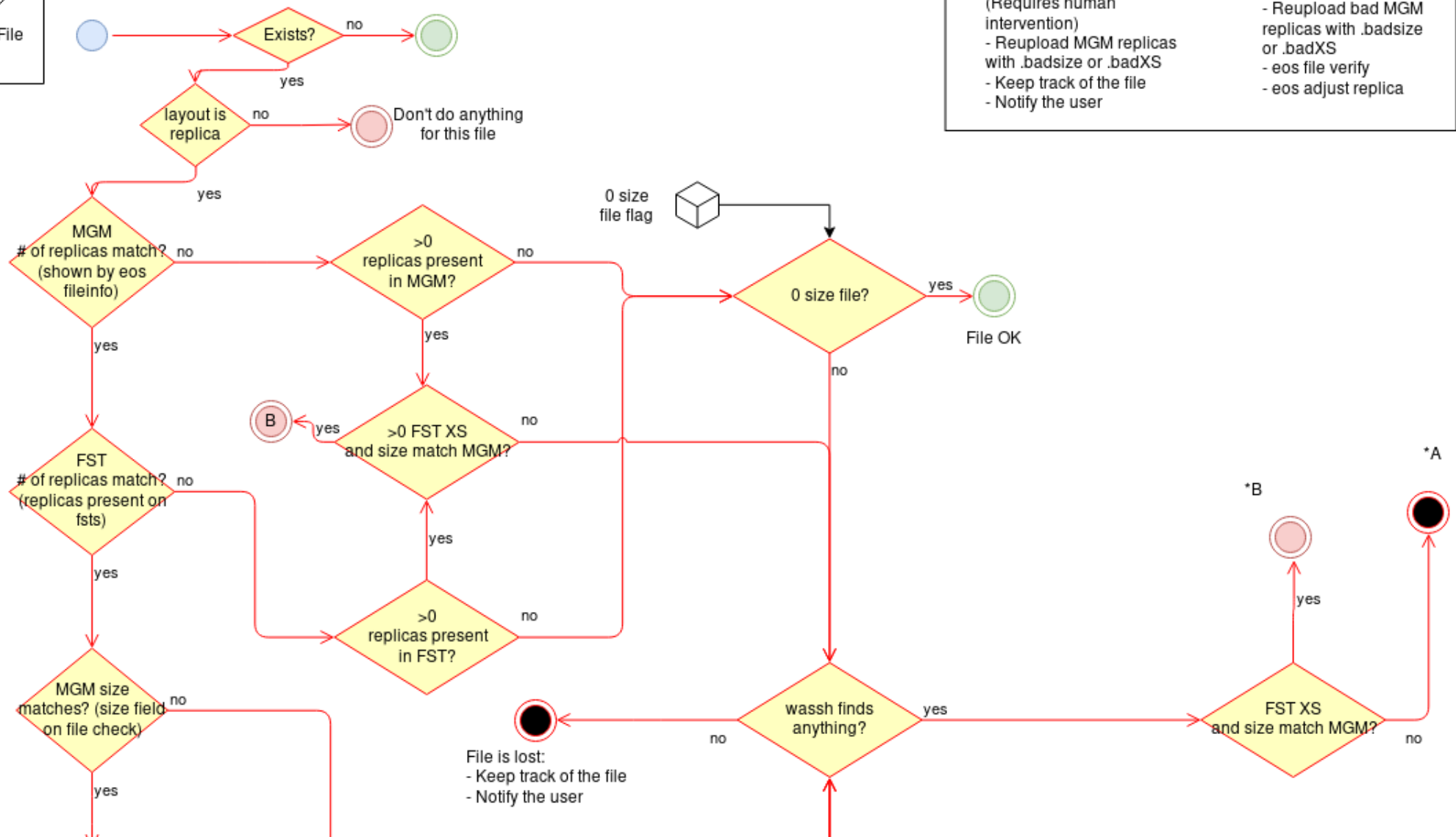- These were individually fixed by the ops team

File

Exists? — no → (end)

yes

layout is replica — no → Don't do anything for this file

yes

MGM # of replicas match? (shown by eos fileinfo) — no → **>0 replicas present in MGM?** — no → **0 size file?** — yes → File OK

yes ↓ (MGM)   yes ↓   no ↓

>0 FST XS and size match MGM? — yes → B — no →

FST # of replicas match? (replicas present on fsts) — no → **>0 replicas present in FST?** — no →

yes ↓   yes ↑

MGM size matches? (size field on file check) — no →

yes ↓

*A Fix:
(Requires human intervention)
- Reupload MGM replicas with .badsize or .badXS
- Keep track of the file
- Notify the user

B Fix:
- Reupload bad MGM replicas with .badsize or .badXS
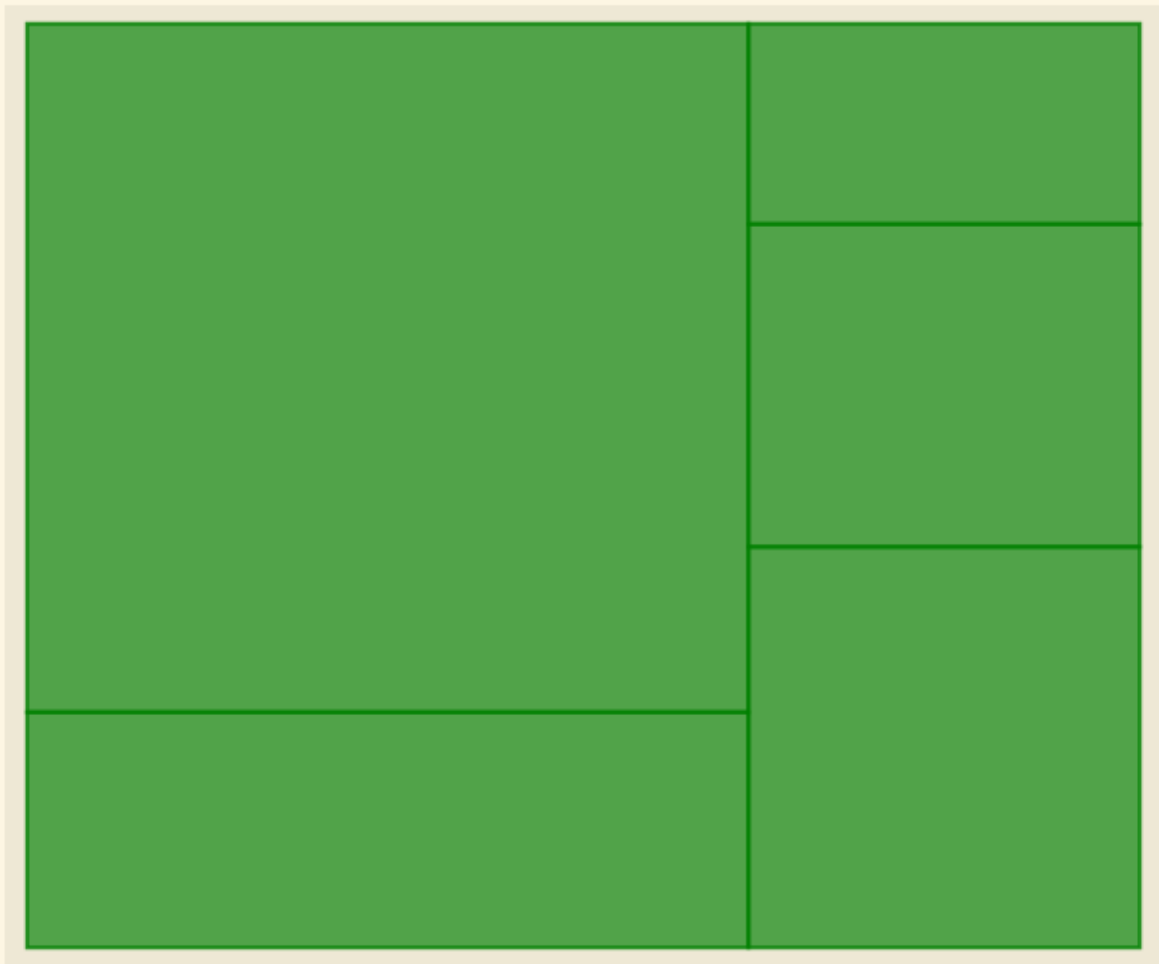- eos file verify
- eos adjust replica

0 size file flag

*B

*A

wassh finds anything? — yes → **FST XS and size match MGM?** — no → *A

no ↓   yes ↑ (to *B)

File is lost:
- Keep track of the file
- Notify the user

Did it work?

Wigner : All_Batches

Generated : 08:36 16 Dec

All6TB
All9918Remaini…
Second4TB
NetHubPlus
WD_4TB

Benefits to the project

- Central draining now battle-tested
- Introduction of proactive replica consistency checking
  - → "Data durability" talk this afternoon
- Improved operational procedures
- Group balancing and conversion engine
  - → Talk this afternoon
- Plus – now we have everything in one datacentre
  - → Erasure Coding possible, "Alice O2" talk (and maybe others)
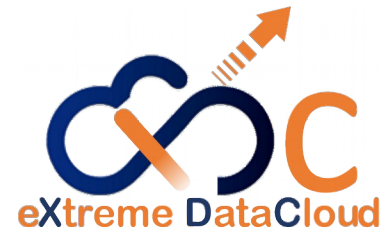
"So long Wigner, and thanks for all the disks"

Extras

eosbackup.cern.ch

- Desperate measures for a special case

- Small files, limited by metadata overheads

- On one occasion, we ran out of time...

```
[root@eosbackup-ns-00 (mgm:master mq:master) ~]$ eos ls -l /eos/backup/wigner_backup
-rw-r--r--   2 daemon    daemon              3972 Jun 24  2019 checksums.tar.gz
-rw-r--r--   2 daemon    daemon     3824630016000 Jun 21  2019 p06253967c40935-016d4a74-7394-41af-ae0a-9592a3aa38d0-data36-1032.tar
-rw-r--r--   2 daemon    daemon     2534791301120 Jun 21  2019 p06253967c40935-02bdcabe-8e5c-49de-ae44-7baa07d10b96-data12-962.tar
-rw-r--r--   2 daemon    daemon     3123901501440 Jun 21  2019 p06253967c40935-0a044cf7-9991-4468-b6ea-3854e3638464-data28-1010.tar
-rw-r--r--   2 daemon    daemon     3899965614080 Jun 21  2019 p06253967c40935-0bab50f1-6e2a-446c-88e9-f4d703534ff0-data13-965.tar
-rw-r--r--   2 daemon    daemon     3913102858240 Jun 21  2019 p06253967c40935-17be1823-db45-4262-b8ac-71a300f2279d-data49-1051.tar
-rw-r--r--   2 daemon    daemon     4449288755200 Jun 21  2019 p06253967c40935-1b06b1e1-4b79-45e9-a854-f1e02d32fe64-data45-1047.tar
-rw-r--r--   2 daemon    daemon     1543242137600 Jun 20  2019 p06253967c40935-1c6405be-3db2-4ed3-8f80-648b35b55ab2-data30-1016.tar
-rw-r--r--   2 daemon    daemon     2547763609600 Jun 20  2019 p06253967c40935-219cb485-3cb0-417e-9795-f0da893d7e8c-data34-1028.tar
-rw-r--r--   2 daemon    daemon     3219341977600 Jun 21  2019 p06253967c40935-23194e13-cbb1-4607-b93a-b45597dc9b68-data16-974.tar
-rw-r--r--   2 daemon    daemon     3730963865600 Jun 21  2019 p06253967c40935-2536695a-e9d1-4fa0-8db1-35ff85414118-data40-1040.tar
-rw-r--r--   2 daemon    daemon     3944267673600 Jun 21  2019 p06253967c40935-271d0ff1-40ce-4377-bfe5-04e9ed7b8264-data23-995.tar
-rw-r--r--   2 daemon    daemon     2783368550400 Jun 21  2019 p06253967c40935-2c723228-793b-4df2-a443-46a382ca2375-data04-931.tar
```