Science & Technology Facilities Council | UK Research and Innovation

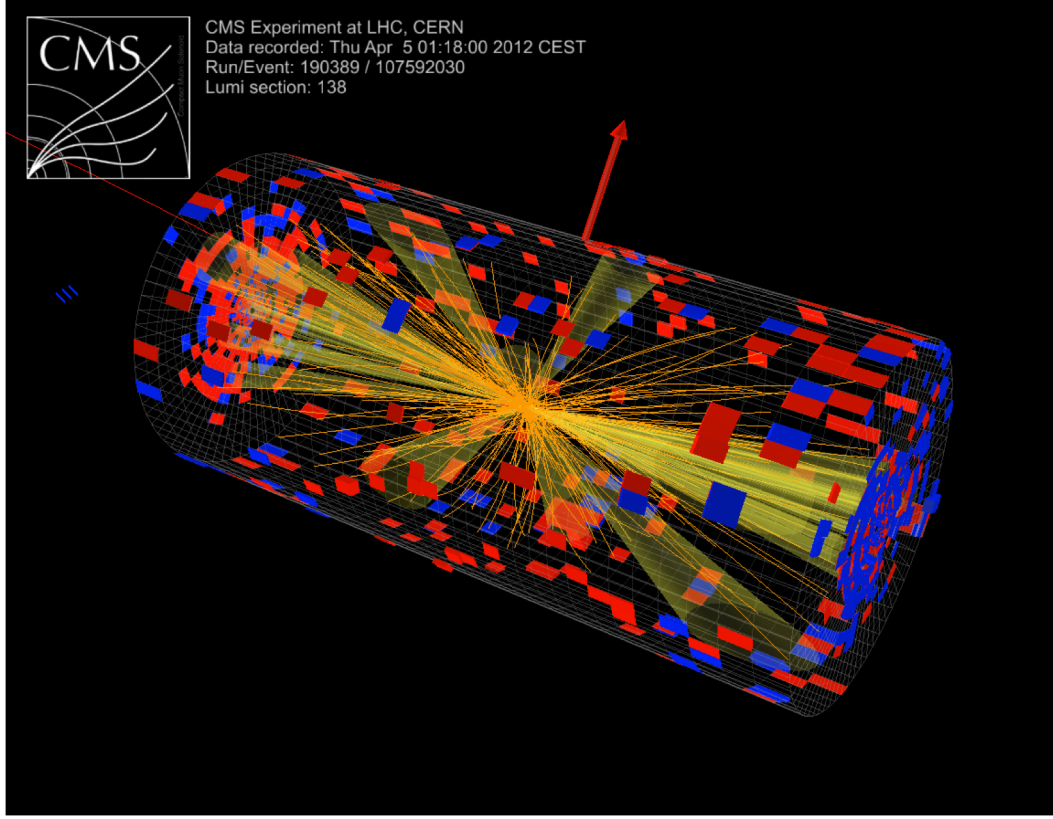
CMS

Katy Ellis

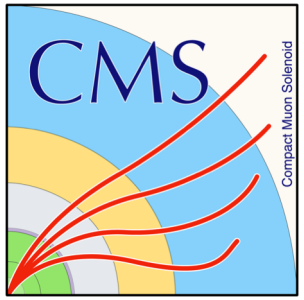
25th April 2019

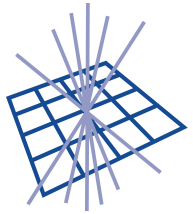
GriddPP42

Introduction



- Katy Ellis
- New CMS / RAL Tier 1 Liaison
- Started in September 2018
- New to CMS (PhD on ATLAS, 2012)
- New to computing/site admin-type role
- Previous experience in computer modelling including testing, writing code and as a user.

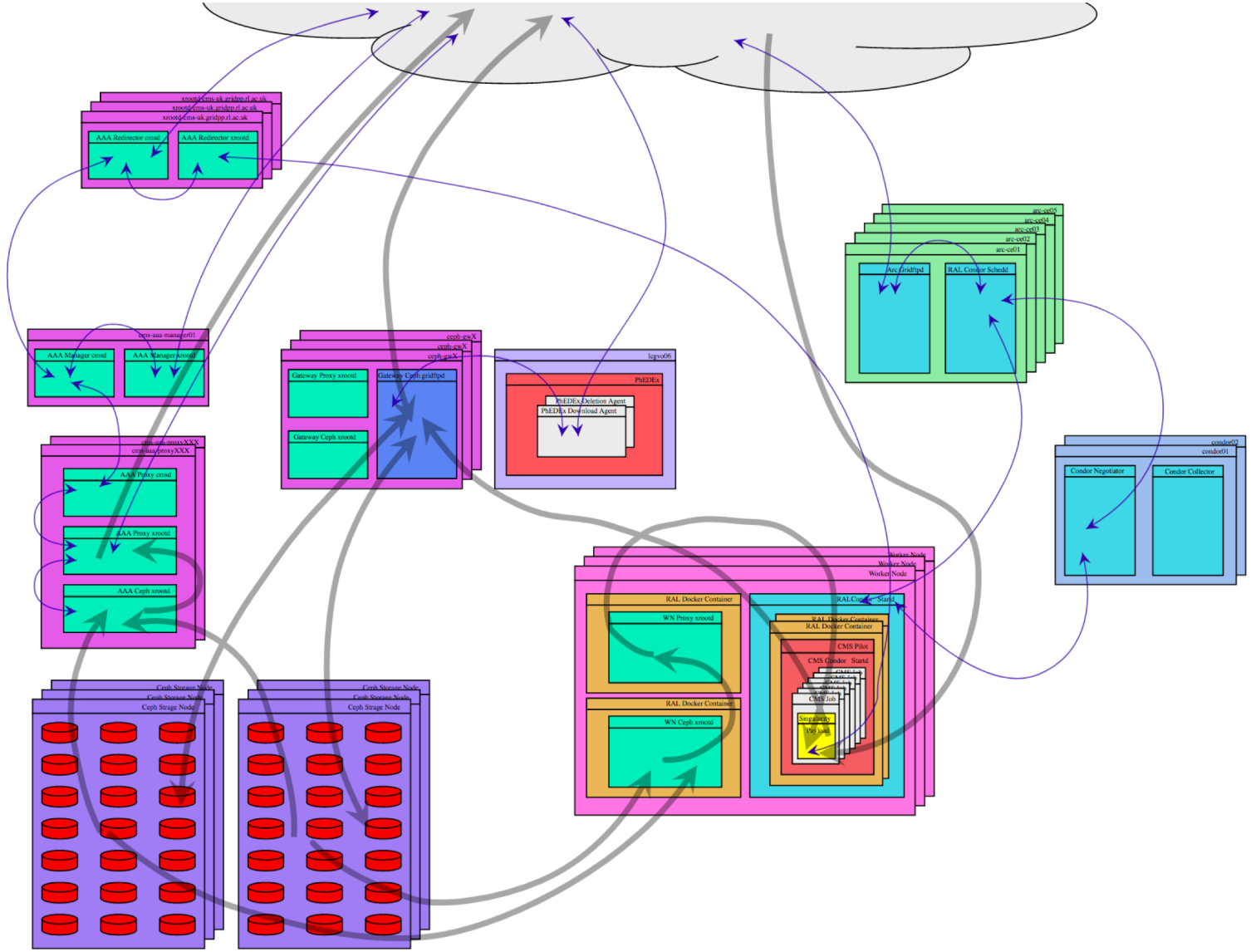




GridPP
UK Computing for Particle Physics

Why is it so complicated?

Why is it so complicated?

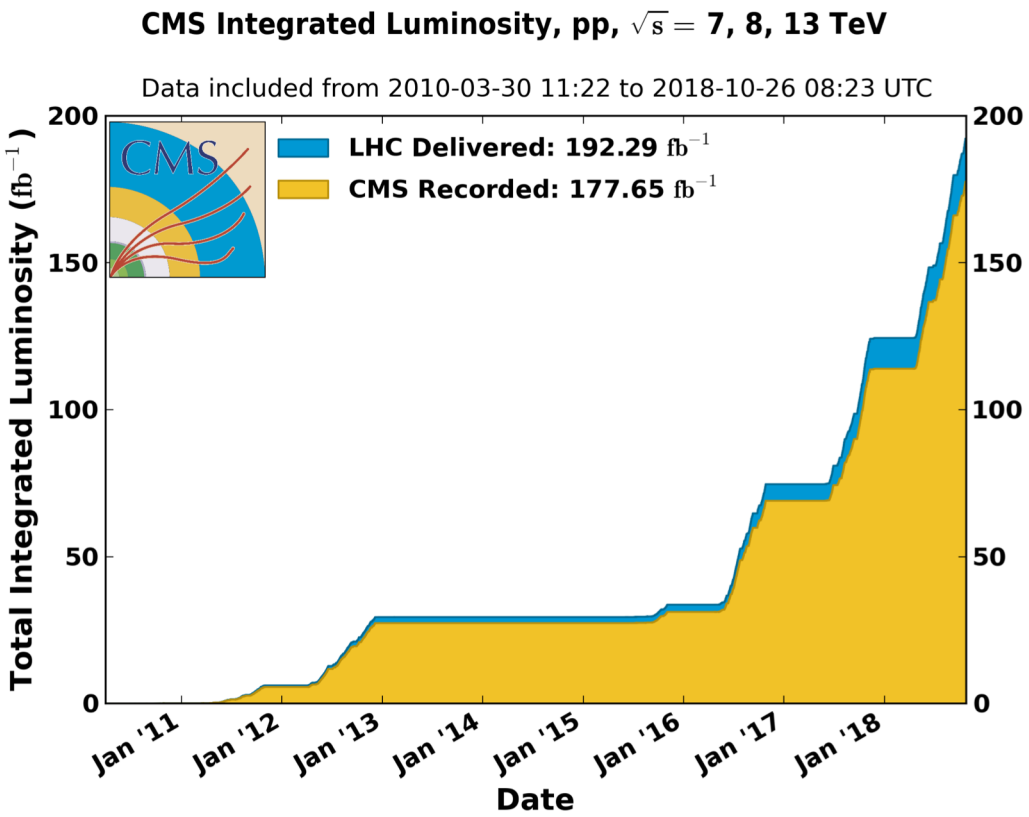


Contents

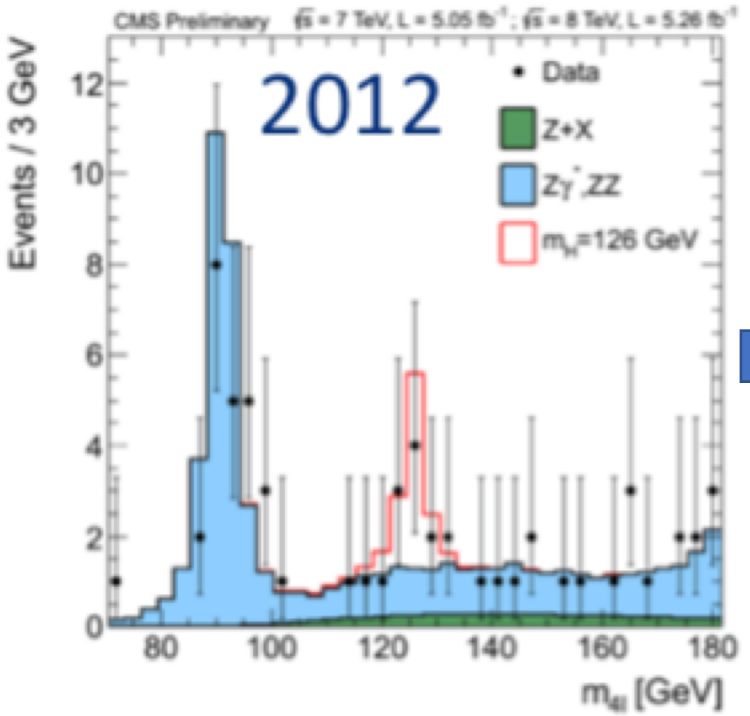
- Overview of CMS status during Run 2
- Plans for Long Shutdown 2 and beyond
- Move to Rucio for data management
- Tier 1 issues
- News from Tier 2s

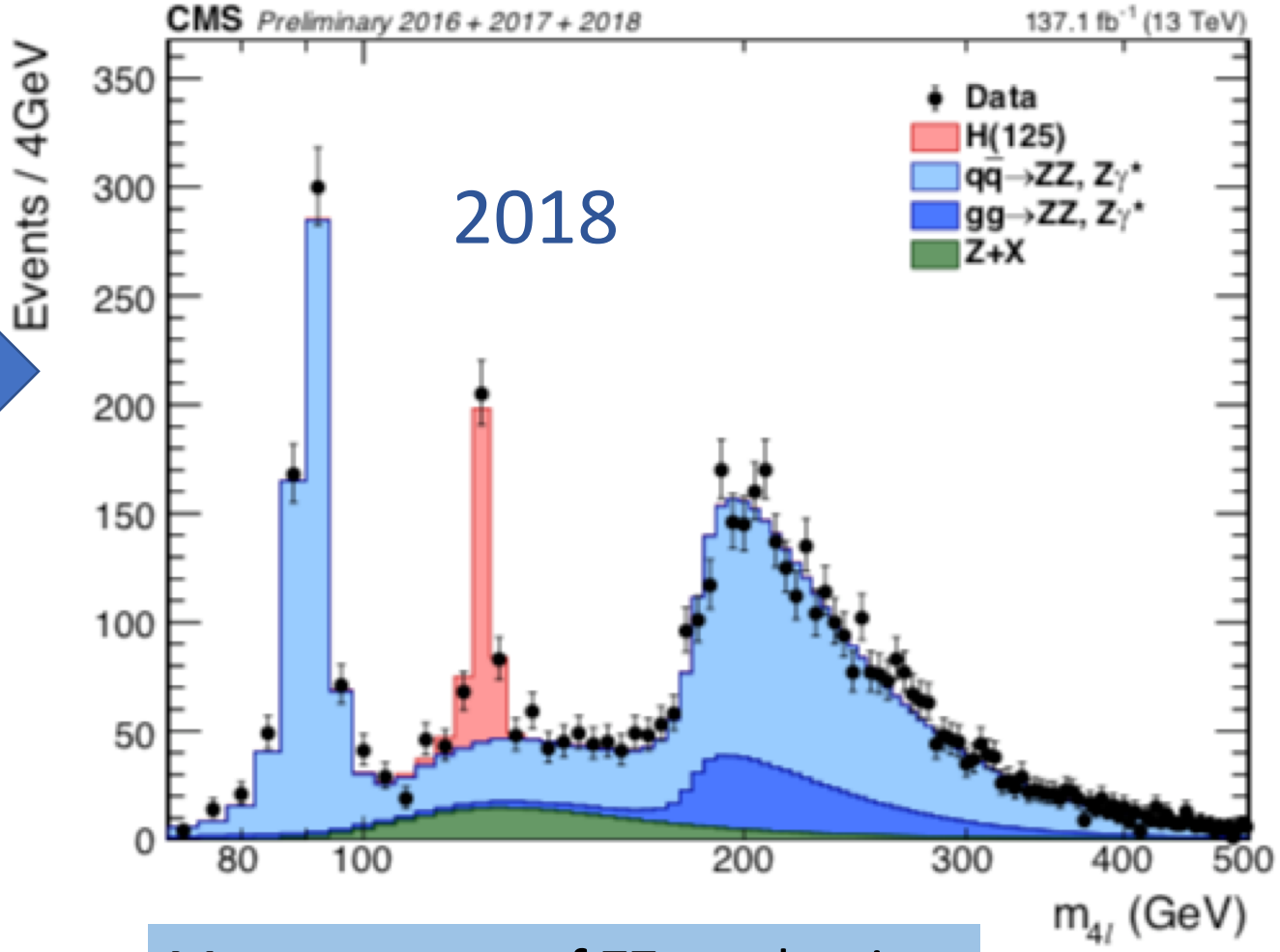
CMS status after Run 2

- Successful data-taking, a productive physics program and good progress on the upgrade projects.



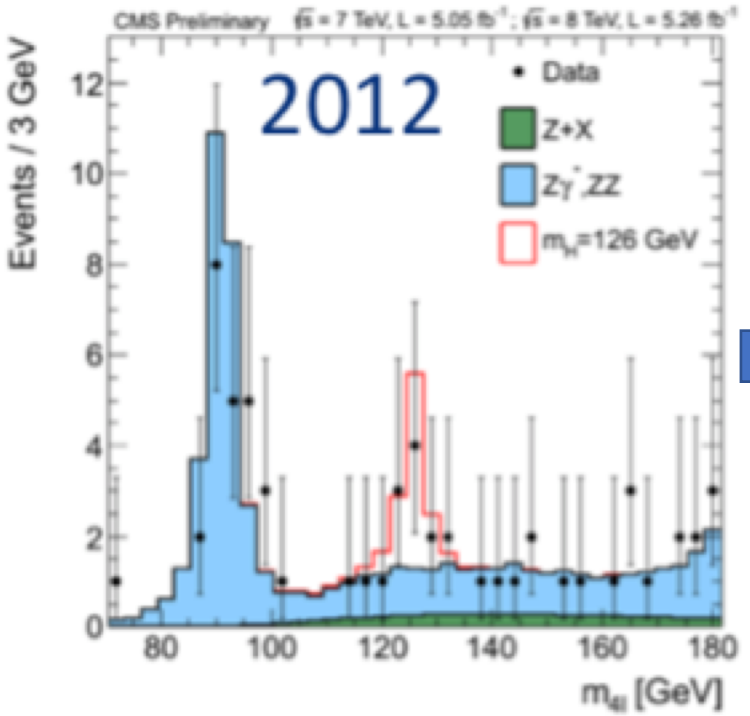
Physics results

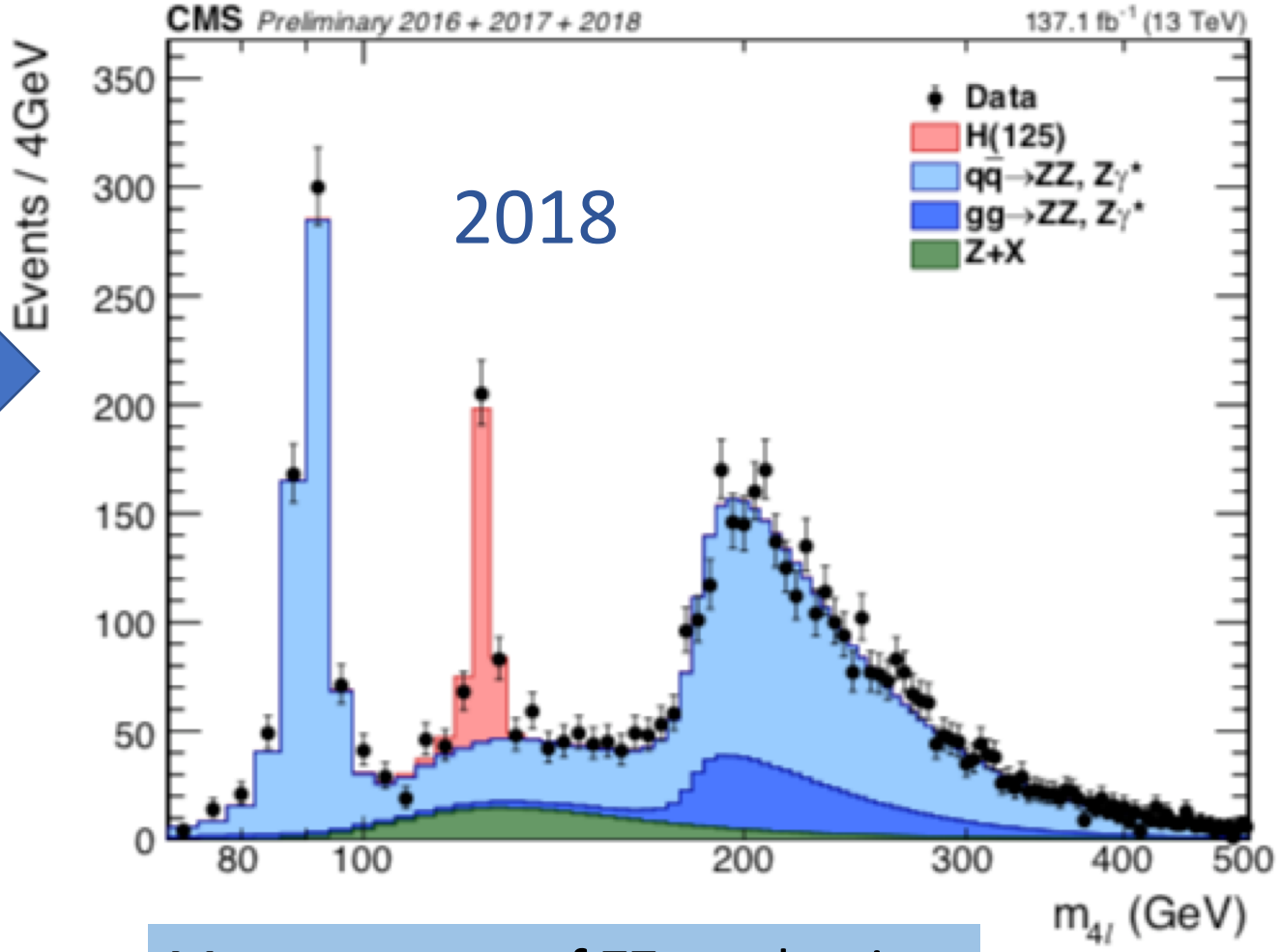




Measurement of ZZ production

Physics results

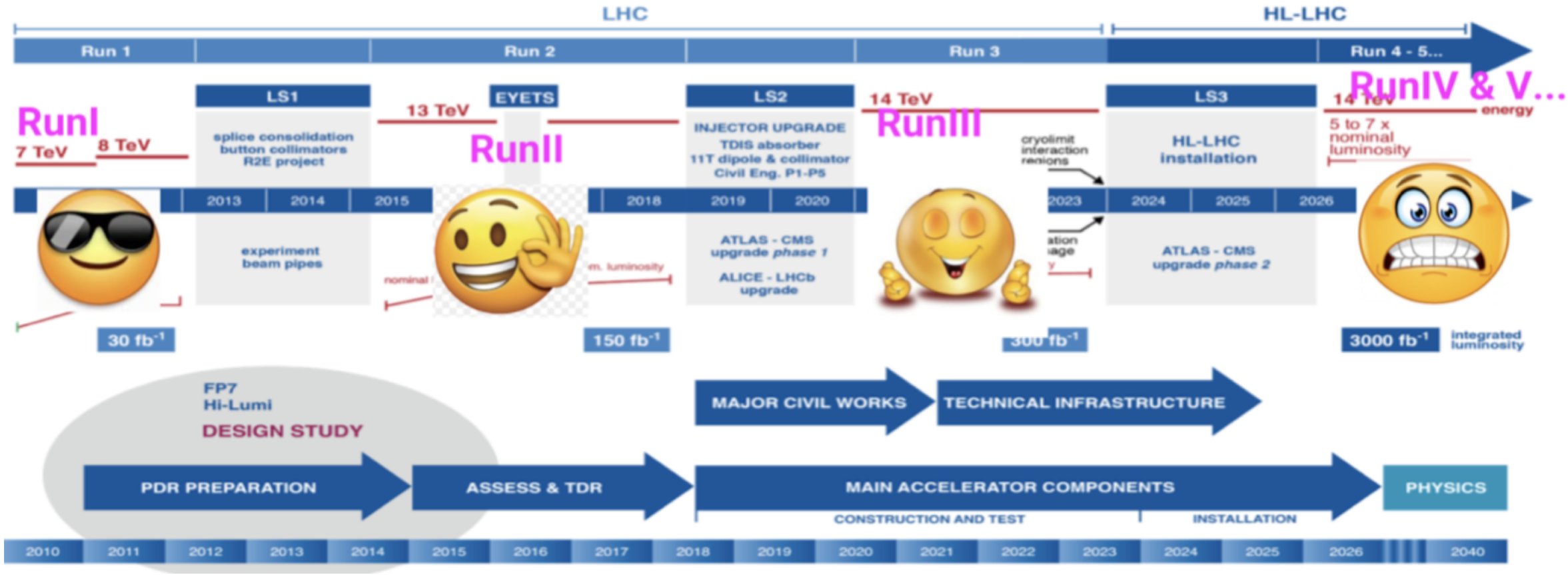




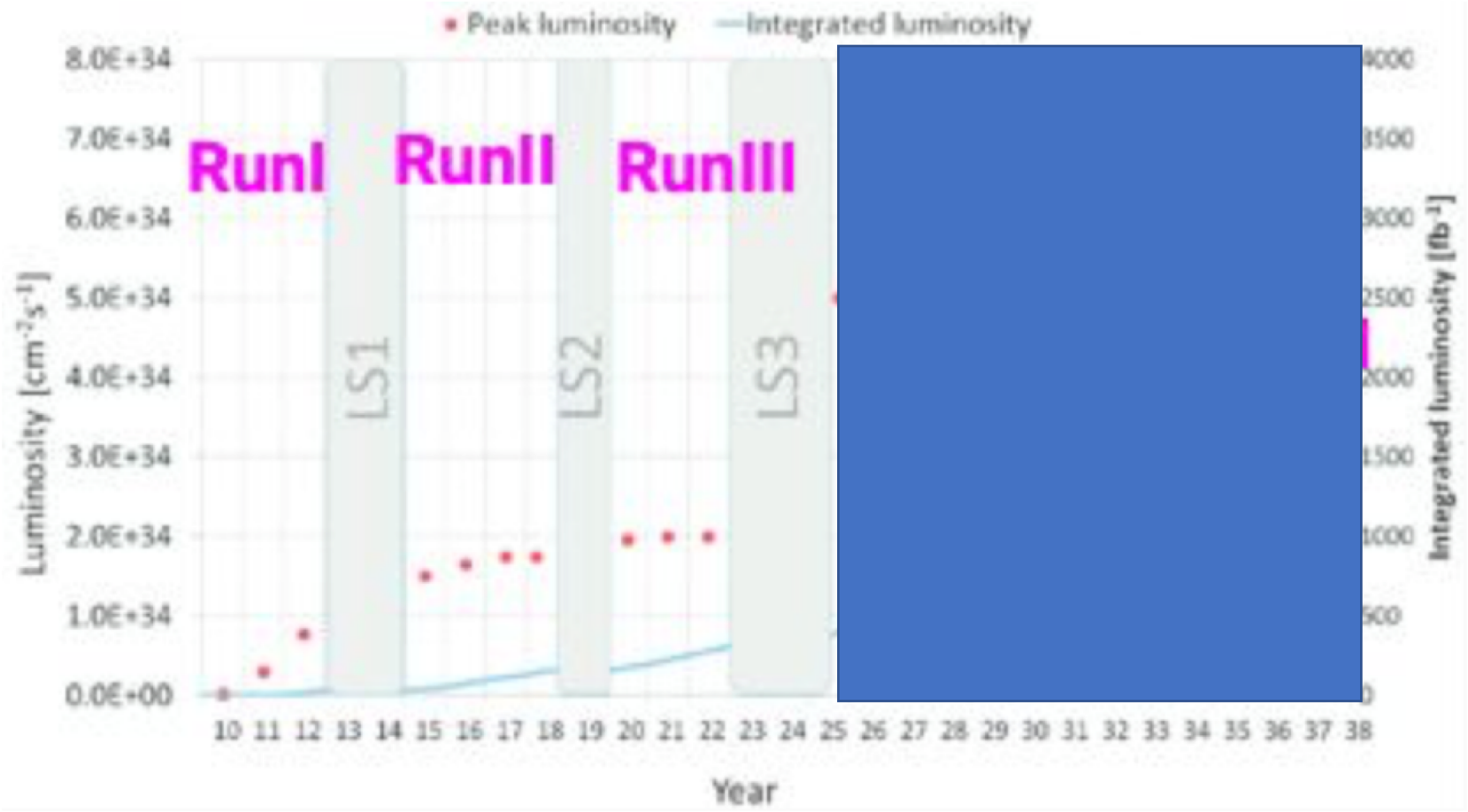
- Observation of:
  - New B-meson states
  - Higgs -> bb
  - ttH
  - H -> $\tau\,\tau$

Measurement of ZZ production

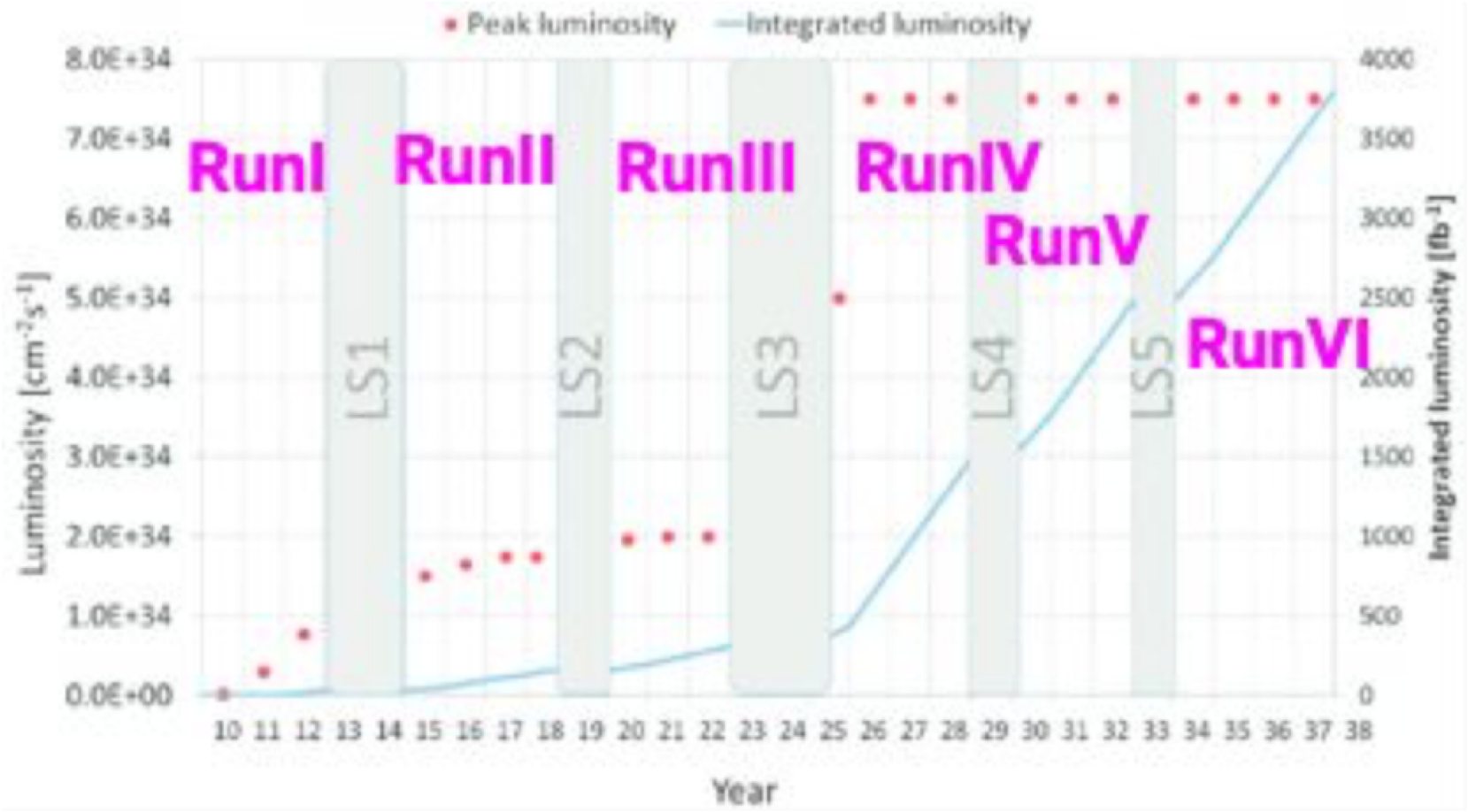
LHC schedule



Increase in data rate



Increase in data rate

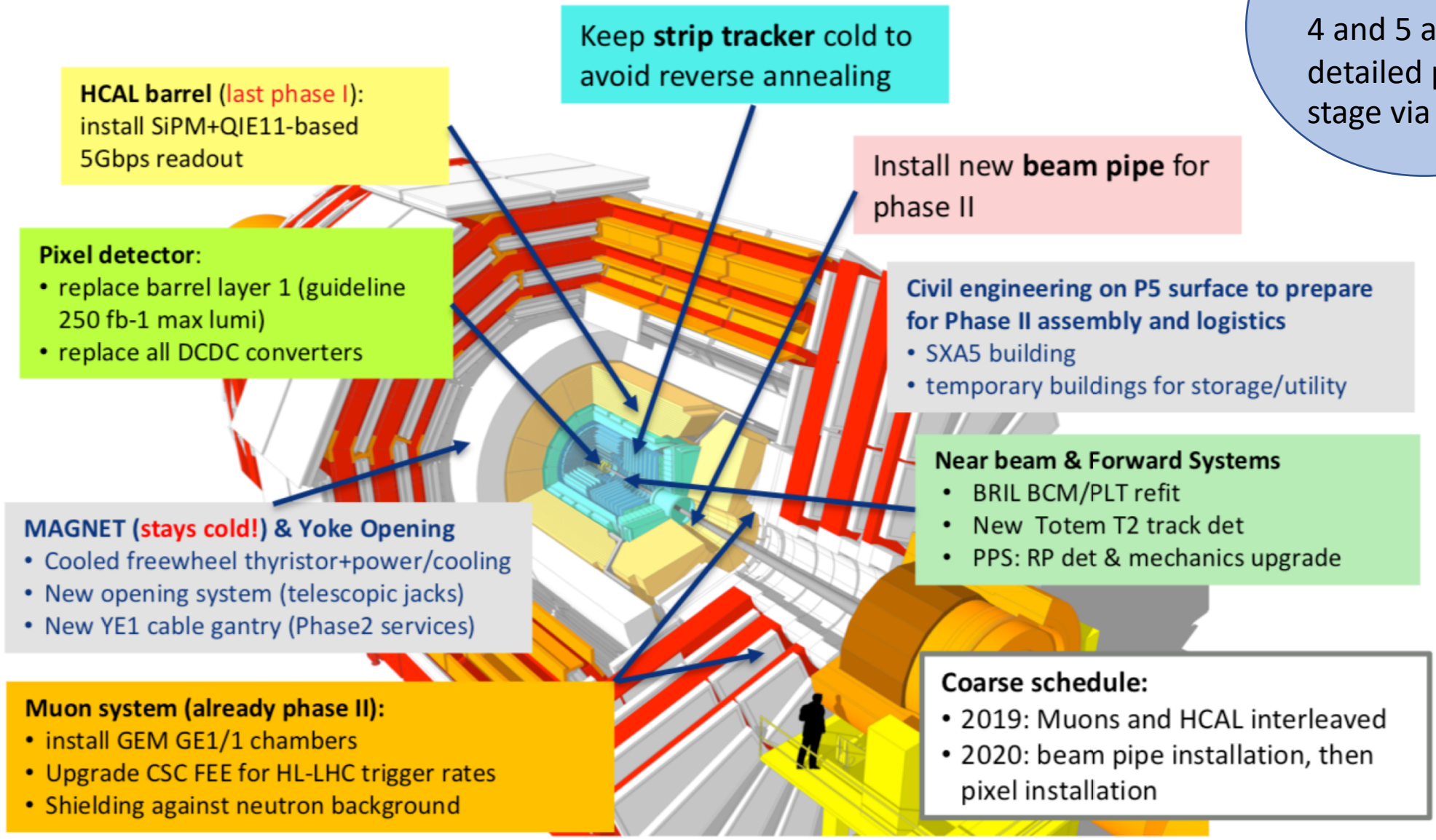


Long shutdown activities

- Detector upgrades
- Production
- Tape and disk cleaning
- CERN Tape Archive
- Rucio

CMS detector upgrades for Run 3

Further upgrades for HL-LHC in Runs 4 and 5 already in detailed planning stage via TDRs.



Keep **strip tracker** cold to avoid reverse annealing

HCAL barrel (last phase I): install SiPM+QIE11-based 5Gbps readout

Install new **beam pipe** for phase II

Pixel detector:
- replace barrel layer 1 (guideline 250 fb-1 max lumi)
- replace all DCDC converters

Civil engineering on P5 surface to prepare for Phase II assembly and logistics
- SXA5 building
- temporary buildings for storage/utility

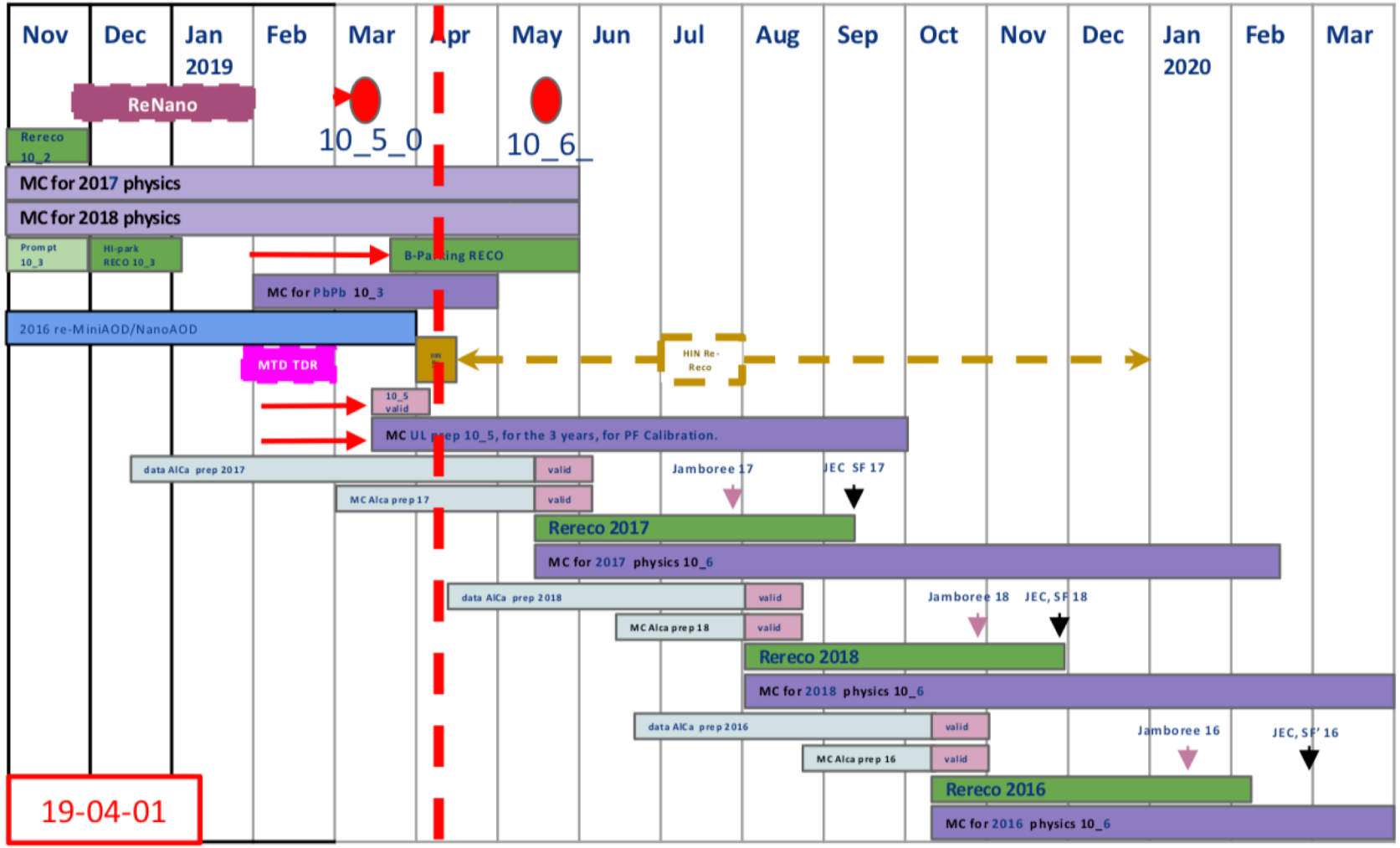
Near beam & Forward Systems
- BRIL BCM/PLT refit
- New Totem T2 track det
- PPS: RP det & mechanics upgrade

MAGNET (stays cold!) & Yoke Opening
- Cooled freewheel thyristor+power/cooling
- New opening system (telescopic jacks)
- New YE1 cable gantry (Phase2 services)

Muon system (already phase II):
- install GEM GE1/1 chambers
- Upgrade CSC FEE for HL-LHC trigger rates
- Shielding against neutron background

Coarse schedule:
- 2019: Muons and HCAL interleaved
- 2020: beam pipe installation, then pixel installation

Production activities



- Monte Carlo (MC) production
  - Event simulation
- Reconstruction
  - Turning electronic signals from the detector into particle candidates with energy and momentum
- Alca
  - Alignment and calibration
- “Rereco”
  - As reconstruction, but with new calibration
- B-parking
  - Raw data waiting for reconstruction

Move to Rucio

- Rucio will replace PhEDEx from Run 3
  - File transfer service
  - File catalogue
  - Highly scalable
  - Heterogeneous storage systems worldwide
  - Run centrally
- Used successfully by ATLAS for several years
- Now open to the wider community
- CMS activities include: integration with Production and User Analysis job submission, setup of databases for e.g. User accounts, data synchronization performance testing, setup of monitoring, etc.

Rucio and CERN Tape Archive (CTA)

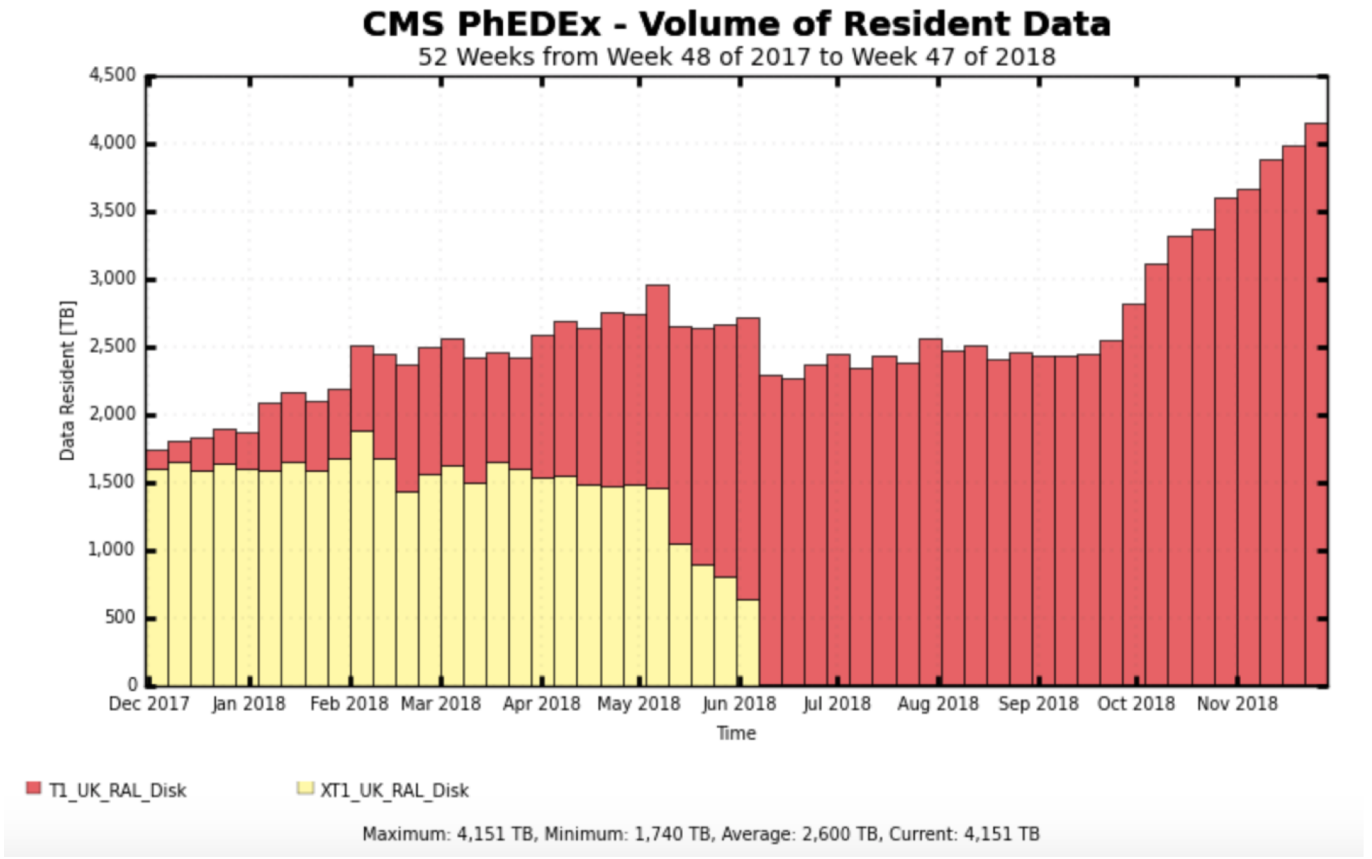
- CTA: CERN Tape Archive, which replaces CASTOR at CERN this summer
  - Meta-data migration only
  - Change to the high-level structure? Possible issue with Rucio
- RAL will also be changing tape system in the medium-term
  - Tender is out
  - Useful to gain expertise integrating Rucio with tape systems
- Pre-production service on CTA, with a test Rucio Storage Element (RSE)

Dynamic resource provisioning

- CMS@home
  - Running jobs successfully
  - Has been absorbed by LHC@home
  - Ivan Reid is now part-time, so he may not have any ‘official’ time to give to this project
- DODAS
  - Pop-up CMS site – running jobs in 20 minutes on any cloud
  - Tested on OpenStack, Azure, AWS, EGI-clouds
- HLT cloud
  - 36,000 cores in between runs and during shutdown

CMS/T1 news and issues

- CMS moved to Ceph-Echo storage approx. 1 year ago
  - On the whole this is working, but Echo is the first large object store CMS has used, and there have been some disconnects between the way CMS requests data and the way Echo provides it.
  - Still finding new ways to get the two systems to mesh.



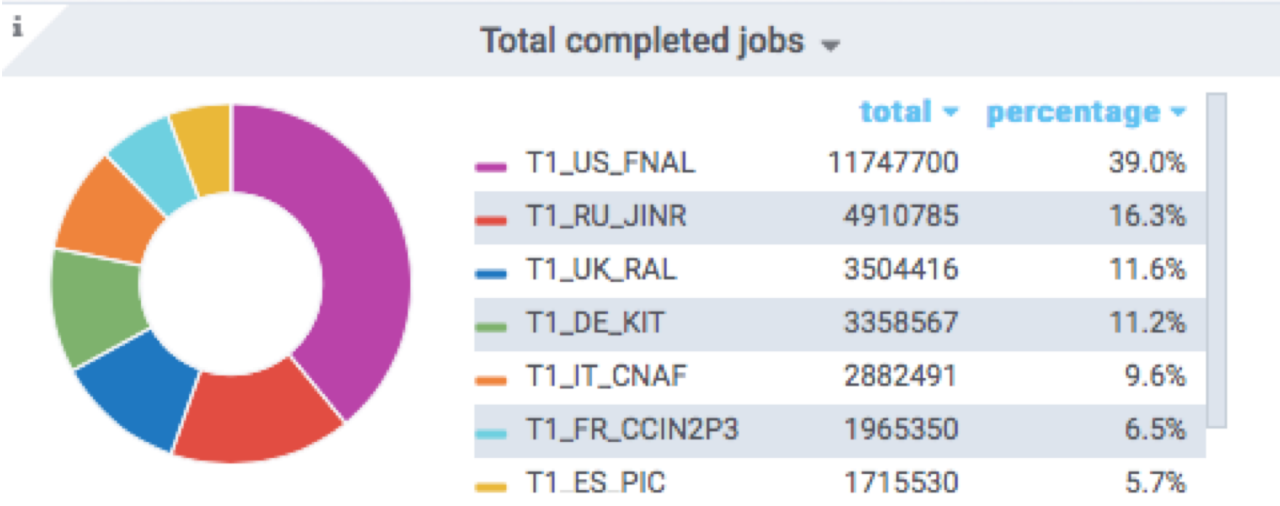
Storage pledge is 5250TB for 2019 – achieved at the end of 2018

Job failures and (in)efficiency

- RAL has high job failures, and hence a low efficiency
- Looking for ways to improve this, e.g.
  - Easier access to storage
  - Investigating problems quickly
- Can be difficult to understand many contributing factors
  - “Bad workflows”

Job failures and (in)efficiency

- RAL has high job failures, and hence a low efficiency
- Looking for ways to improve this, e.g.
  - Easier access to storage
  - Investigating problems quickly
- Can be difficult to understand many contributing factors
  - “Bad workflows”



CMS AAA at RAL T1

- This is a CMS service which allows CMS jobs to access ‘Any data, Anytime, Anywhere’.
- Initially only a fall-back for jobs on sites with broken local storage, but increasingly used to run jobs ‘anywhere’ or on remote datasets.
- This means more and more data is running through the AAA proxies, and optimizing the configuration of this setup as the only major site combining xrootd with object storage is a work in progress!

CMS AAA at RAL T1

- This is a CMS service which allows CMS jobs to access ‘Any data, Anytime, Anywhere’.
- Initially only a fall-back for jobs on sites with broken local storage, but increasingly used to run jobs ‘anywhere’ or on remote datasets.
- This means more and more data is running through the AAA proxies, and optimizing the configuration of this setup as the only major site combining xrootd with object storage is a work in progress!

Already more reliable compared to CASTOR.

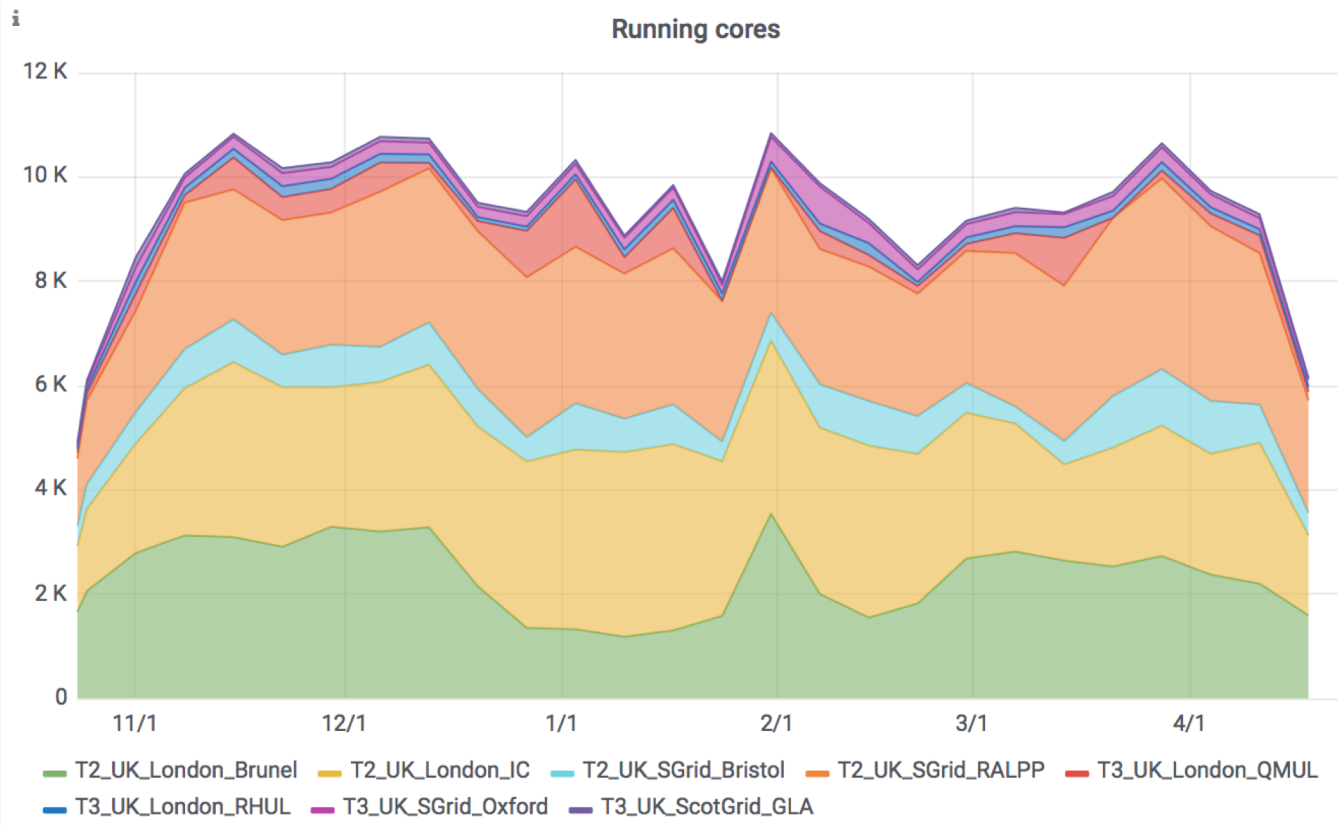
Now dual stack (IPv4/6)

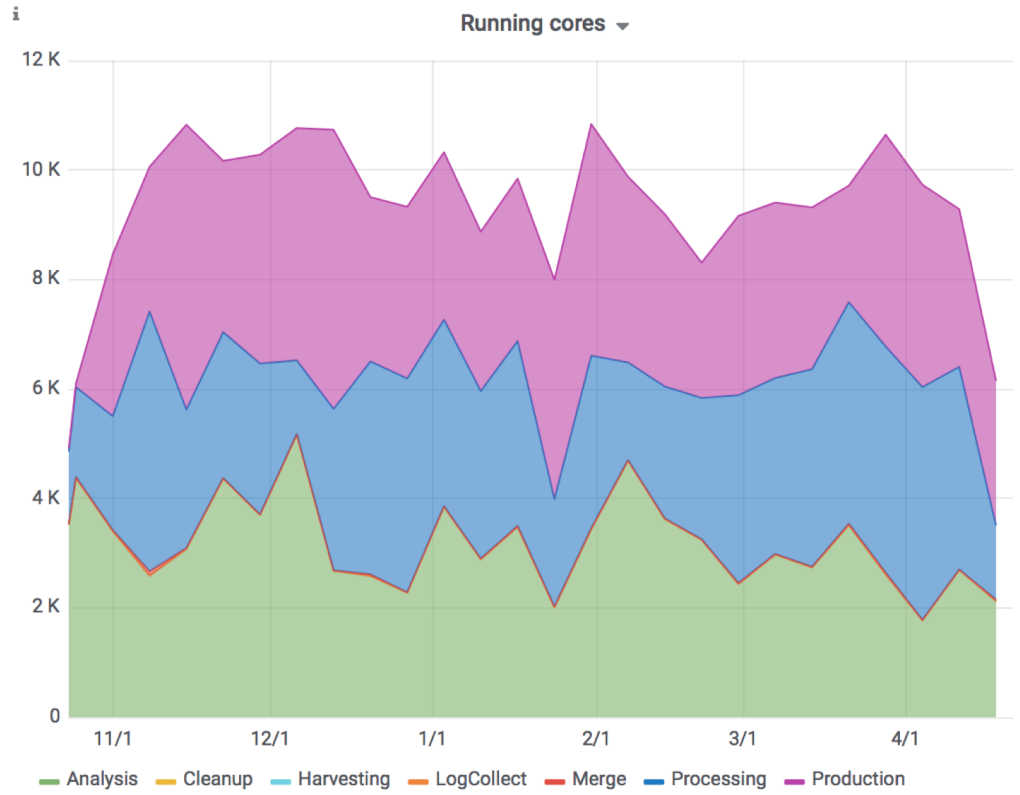
New hardware coming soon.

Further investigations into throttling.

CMS T2s

- Main Tier 2s: Imperial College, RAL PP, Brunel and (Bristol)
- Tier 2s used as Tier 3s (no CMS storage): QMUL, RHUL, Oxford, Glasgow
  - CMS is open to expansion of this list





Details on CMS T2s (pledge)

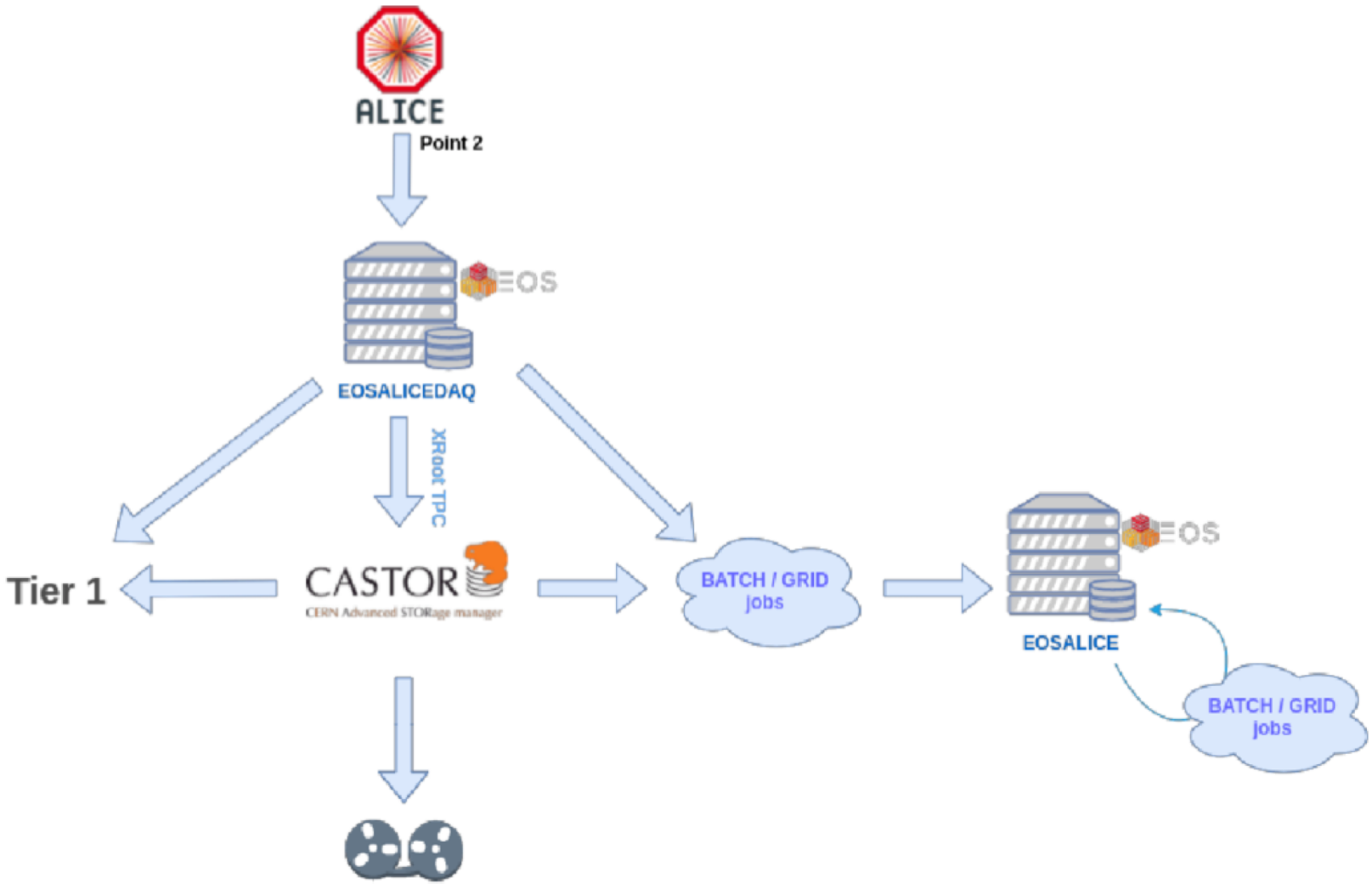
- Imperial (2200TB)
  - Moving the data centre to Slough in June
  - 2 * 100 Gb/s network (one is fallback)
- RAL_PP (1100TB, 1600TB imminently)
  - Connecting to LHCONE soon
- Brunel (500TB)
  - CMS using close to 100% of storage due to a bug
  - Some issues after upgrading to DOME version of DPM. Better testing before deployment would improve this situation.
  - 40Gb/s coming in May

Summary

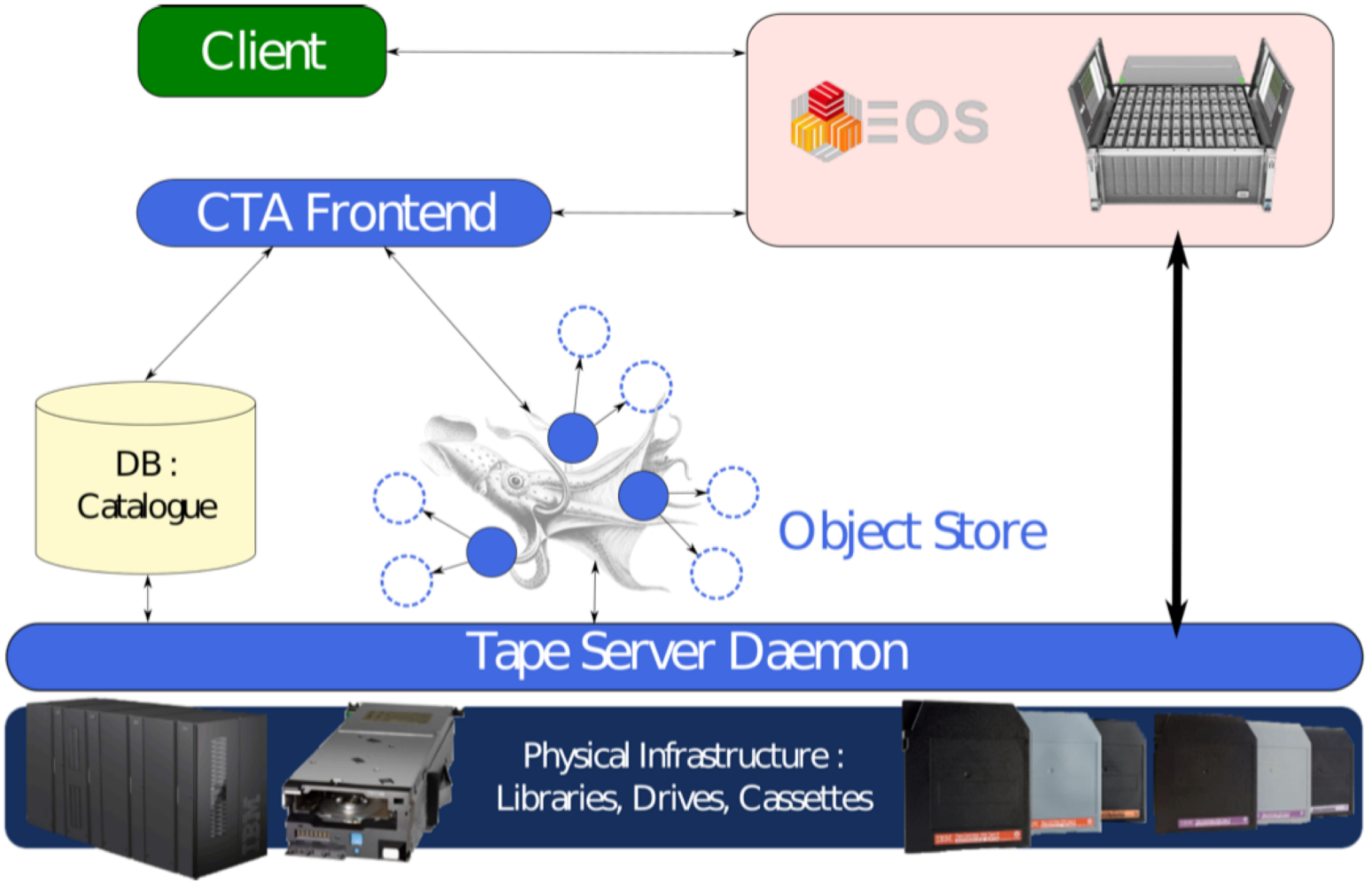
- CMS has had a successful Run 2.
- Long shutdown 2 is a good time for major computing changes…
  - E.g. Rucio
- …and to think about the challenges beyond Run 3.
- Minor changes happening at RAL T1 to improve performance, particularly around data access.
- Continuous improvements at T1.
- T1 and T2s generally working quite well.

Backup

‘Old’ CERN CASTOR tape setup



‘New’ CERN CTA tape setup



Other issues

- Missing/corrupted files – missing chunks. Interaction of phedex not relying on FTS exit statuses and Echo lying about file sizes/checksums (echo still does this, phedex side was fixed – check-gfal-verify)
- Vector reads – change in format, but old files still have this issue. At the start of the job, the meta data is read to avoid any duplicate events