

User and group EOS storage management at the CMS CERN Tier-2

G. Cerminara, G. Franzoni, A. Pfeiffer

CERN CMG-CO
for the CMS collaboration



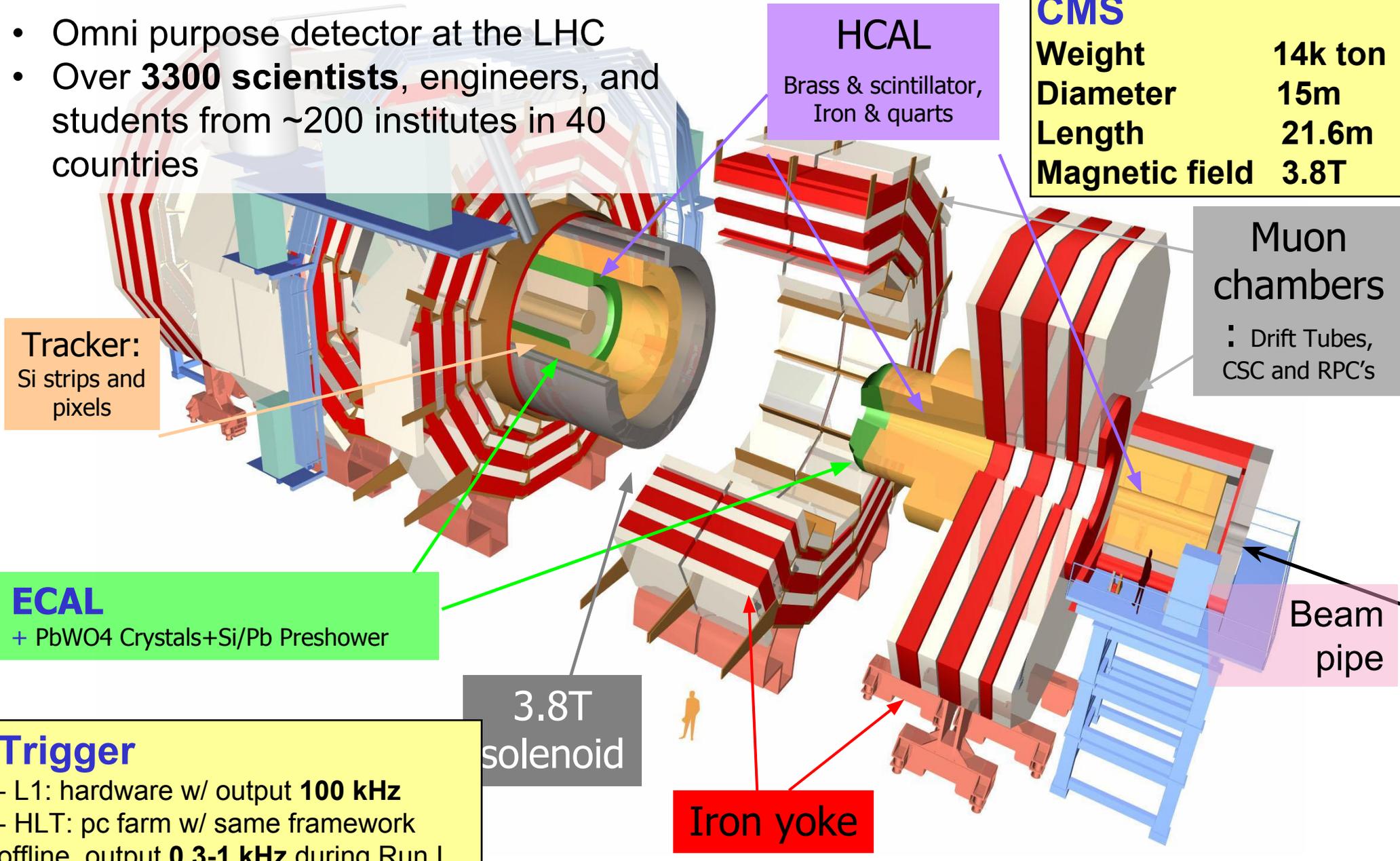
- The CMS collaboration
- Data processing and analysis workflows at CMS

- EOS for analysis at the T2_CH_CERN
- working teams and roles

Compact Muon Solenoid

- Omni purpose detector at the LHC
- Over **3300 scientists**, engineers, and students from ~200 institutes in 40 countries

CMS	
Weight	14k ton
Diameter	15m
Length	21.6m
Magnetic field	3.8T



Tracker:
Si strips and pixels

ECAL
+ PbWO4 Crystals + Si/Pb Preshower

HCAL
Brass & scintillator,
Iron & quarts

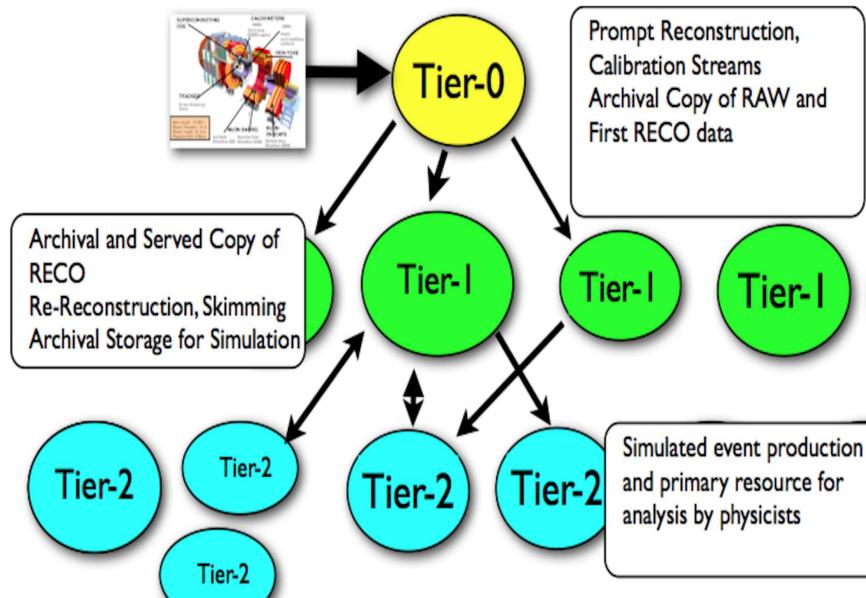
Muon chambers
: Drift Tubes,
CSC and RPC's

3.8T solenoid

Iron yoke

Beam pipe

Trigger
- L1: hardware w/ output **100 kHz**
- HLT: pc farm w/ same framework offline, output **0.3-1 kHz** during Run I



- The dataset recorded by CMS is processed and distributed worldwide :
 - 10 major computing (Tier1) centers
 - ~100 smaller regional centers (Tier2 and Tier3)
- Simulated events: produced (T1,T2) & distributed to T1,2,3

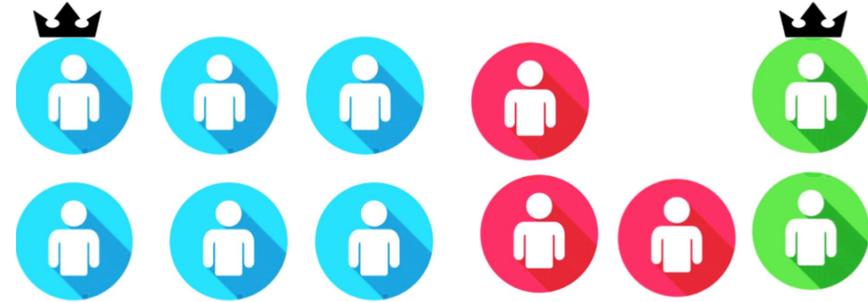
- Complex analysis workflows involve the creation of analysis-specific datasets
 - derived from those produced centrally by CMS
 - specifically targeting an analysis technique or measurement
 - with reduced event content
 - with customised event content
 - in several versions

- **T2_CH_CERN caters for unique analysis needs** in the CMS collaboration
 - **prompt commissioning and calibration** workflows
 - mostly done on site at CERN, to guarantee low latency
 - grant access to official CMS datasets specifically made for commissioning and not distributed world-wide
 - many **collaborators from remote labs/universities reside at CERN** for extended periods of time and prefer local storage resources to those offered by their home institute
 - aggregation of large and leading working teams on CERN premises working on the same measurement
 - activities recognized by the official detector, physics object and physics analysis groups, convened by L2 managers
- A pool of **EOS storage has been dedicated** to such needs
 - currently 2.3 PB in total, 500 users



drwxr-sr+	aguy	zh	Feb-7-2015	Apr-2-2016	/eos/cms/ecal/f1.txt
drwxr-sr+	otherguy	zh	Jan-6-2014	Feb-9-2015	/eos/cms/ecal/tupl_a.root
drwxr-sr+	otherguy	zh	Oct-4-2012	Feb-9-2016	/eos/cms/ecal/tupl_b.root
drwxr-sr+	trdguy	zh	May-3-2013	Mar-5-2015	/eos/cms/muon/eff1.root
drwxr-sr+	fthguy	zh	Jun-1-2012	Jun-2-2016	/eos/cms/exo/plot.eps
drwxr-sr+	fthguy	zh	Jun-1-2012	Mar-5-2016	/eos/cms/exo/plot.root

- The first system (of which we inherited the administration) was:
 - a single EOS node
 - partitioned in directories assigned to the detector/physics group
- Shortcomings and operational hassles:
 - groups could take as much space as available, w/o a budget
 - once the overall quota was reached → general stall → emergency → unspecific calls to cleanup space
- Incidents:
 - Every user had credential to write/remove files anywhere within the EOS node → two accidental massive deletions caused by the *rm* of directory still holding the EOS mount point...



drwxr-sr+ aguy zh Feb-7-2015 Apr-2-2016 [/eos/cms/ecal/f1.txt](#)
 drwxr-sr+ otherguy zh Jan-6-2014 Feb-9-2015 [/eos/cms/ecal/tupl_a.root](#)
 drwxr-sr+ otherguy zh Oct-4-2012 Feb-9-2016 [/eos/cms/ecal/tupl_b.root](#)
 drwxr-sr+ trdguy zh May-3-2013 Mar-5-2015 [/eos/cms/muon/eff1.root](#)
 drwxr-sr+ fthguy zh Jun-1-2012 Jun-2-2016 [/eos/cms/exo/plot.eps](#)
 drwxr-sr+ fthguy zh Jun-1-2012 Mar-5-2016 [/eos/cms/exo/plot.root](#)

- keep personal space to minimum
 - 1 TB allotted only if user is CERN resident >6 m, with overbooking
- **Aggregate** detector/physics group activities in **EOS teams**
 - users focussed on the same specific activity share the same data,
 - one directory per activity, each an eos node with a quota
 - currently **47 teams**: from few TB to ~250TB, 5 to 130 members
- CMS detector/physics group **conveners manage the members** of their EOS teams using e-groups
- **Each team negotiates its quota** and organizes its allotment. No cross-talk between quota-full across different teams
- Users granted credentials to read/write within their EOS team, removal of the permission to write into the overall cms pool



drwxr-sr+	aguy	zh	Feb-7-2015	Apr-2-2016	/eos/cms/ecal/f1.txt
drwxr-sr+	otherguy	zh	Jan-6-2014	Feb-9-2015	/eos/cms/ecal/tupl_a.root
drwxr-sr+	otherguy	zh	Oct-4-2012	Feb-9-2016	/eos/cms/ecal/tupl_b.root
drwxr-sr+	trdguy	zh	May-3-2013	Jun-2-2015	/eos/cms/muon/eff1.root
drwxr-sr+	fthguy	zh	Jun-1-2012	Mar-5-2015	/eos/cms/exo/plot.eps
drwxr-sr+	fthguy	zh	Jun-1-2012	Mar-5-2015	/eos/cms/exo/plot.root

- The team administrators appoint (via an additional egroup) a subset of their team members to the role of “cleaners”
- The cleaners are granted permissions to delete files created and owned by other users of the team - functional to:
 - implement cleanup campaigns
 - remove files of users who have left the team or CMS
- Files’ creation and most-recent-access time (from CERN IT EOS XRootD monitoring) used to distribute popularity reports identifying stale files → guide the work of the cleaners
 - inputs no longer available; could be resumed ?

					drwxr-sr+ aguy zh Feb-7-2015 Apr-2-2015 /eos/cms/ecal/f1.txt
					drwxr-sr+ otherguy zh Oct-4-2012 Feb-9-2015 /eos/cms/ecal/archive/tupl_b.arch
					drwxr-sr+ trdguy zh May-3-2013 Mar-5-2014 /eos/cms/muon/archive/eff1.arch
					drwxr-sr+ fthguy zh Jun-1-2012 Jun-2-2014 /eos/cms/exo/plot.eps

- Frequent use case: analysis in the final stages have very large datasets they likely won't need again, but can't get rid of them just yet
- CMS presented the use case to the CERN IT EOS development team
- Archival from out of EOS to tape:
 - stale data can be cleaned up from disk by archiving it out of EOS onto tape storage (castor)
 - EOS disk quota becomes available from the cleaned up files
- The team cleaners are the only ones granted the credentials to command the archival; restoring back to disk is expected to happen sparsely and infrequently
 - the teams have used the archival extensively, needing occasional support in cases when the archival does not complete

- 2.3 PB of EOS disk space at the T2_CH_CERN to cater for unique and critical analysis needs of the CMS collaboration
 - commissioning and calibration
 - analyzes officially recognized by the physics L2 convenors
- Most of the T2_CH_CERN user/group resources are organized in working teams
 - they share analysis goals and data
 - administration of the working team delegated to the L2 convenors
 - different user patterns: analysis, cleaning & archivist mapped into different EOS access credentials (via e-groups)