

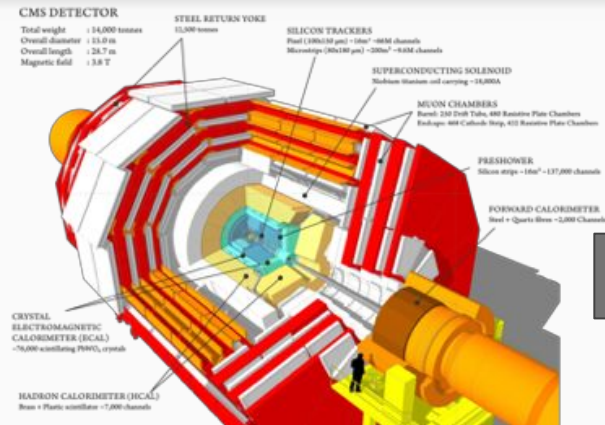
CMS Computing, Offline and Data Access

Tommaso Boccali
CMS Computing Coordinator
INFN-Pisa/CERN

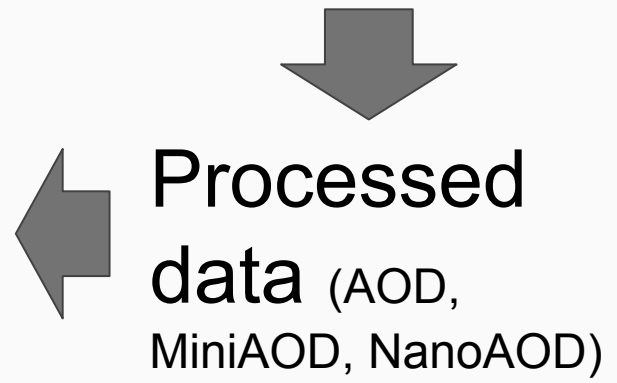
outline

- The CMS Computing and Offline environment
- Resources, services, facilities
- Analysis workflow, from A to Z

Data flows #1 - data

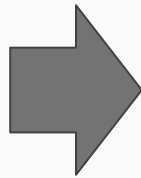


User



Data flows #2 - DT

MC contact,
Analysis contact



Processed
data (AOD,
MiniAOD, NanoAOD)



User



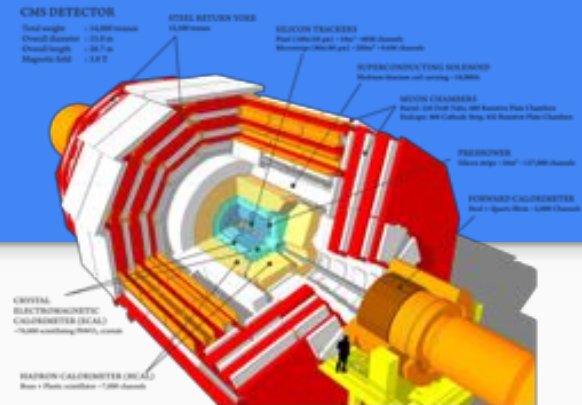
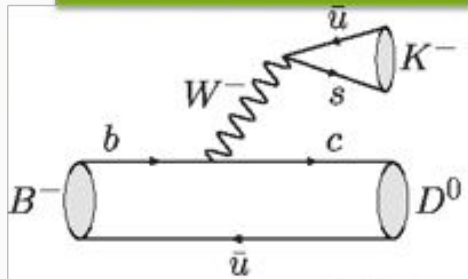
Analysis data
(ntuples,)



Reality



Decay of unstable particles

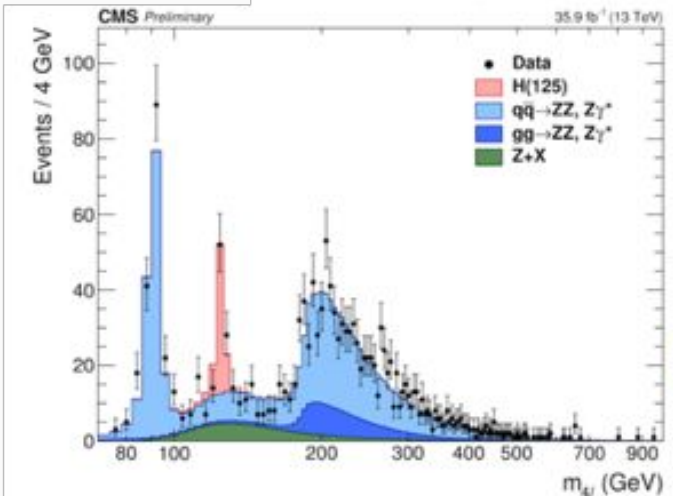


Detector electronics

Trigger (selection)

Reconstruction

Analysis



SW

SW

SW

Simulation - all SW

Theoretical model

Simulation of decays of unstable particles

Simulation of interactions particle-detector



$$L_{QCD} = \sum_f \bar{\psi}_f (i\gamma_\mu D^\mu - m_f) \psi_f - \frac{1}{2} \text{Tr} [\bar{G}_\mu \bar{G}^{\mu\nu}]$$

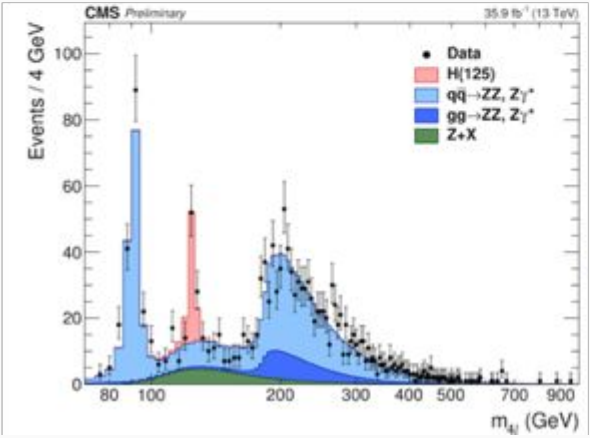


Simulation of detector electronics

Trigger Simulation

Reconstruction

Analysis



CMS Computing - how much?

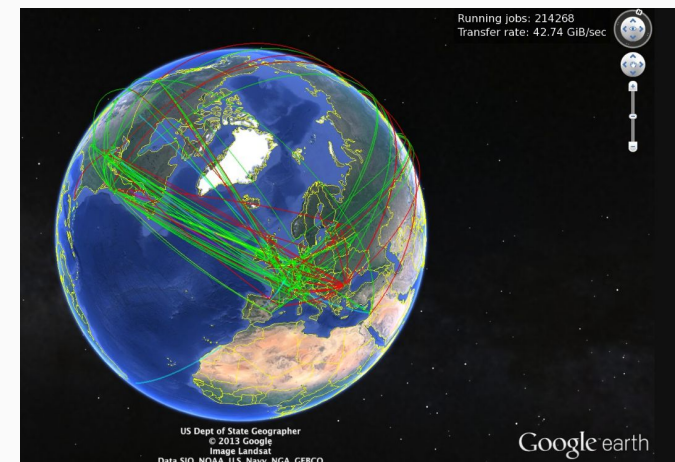
- Basic parameters for (e.g.) 2018:
 - Data taking @ 1 kHz + 2 kHz parked → 3 kHz of events, 6 months a year
 - ~ 20B events/y
 - Each event ~ 1 MB
 - ~20 PB x2 (custodial copies) RAW
 - For each data event, 1 to 2 MC events to process (~40 sec/ev)
 - Processing RAW → AOD ~ 30 sec/ev
 - CPUs for analysis etc...
- All in all, resources needed for 2018 are:
 - ~250k computing cores
 - ~130 PB Disk
 - ~250 PB Tape

How to handle such a large system?

WLCG - the Worldwide LHC Computing GRID

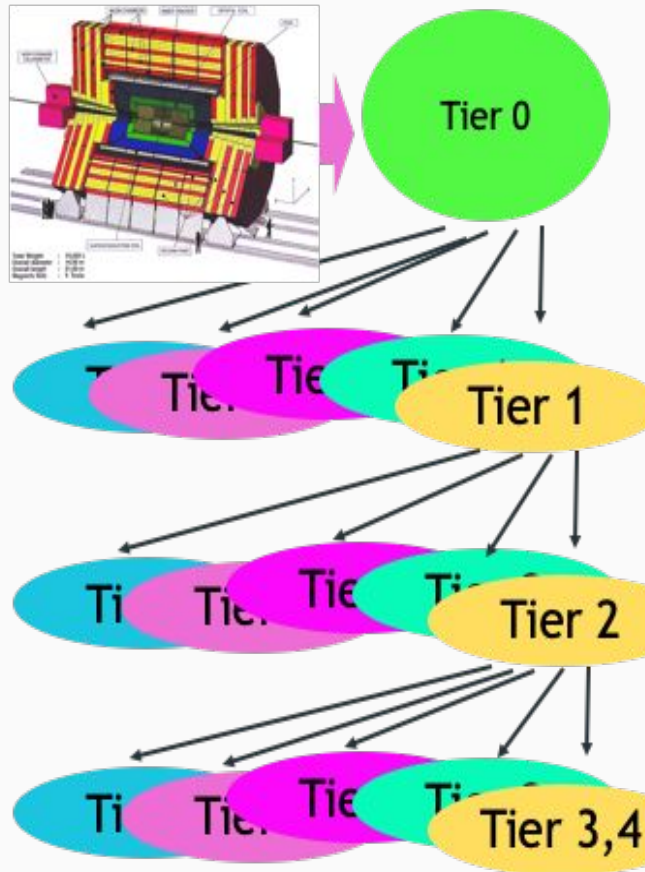
- 200+ sites
 - Storage already at Exabyte level
 - **1000 PB**
 - ~ 1 million cores
 - Expected to increase ~20x in the next 10y
- > 40 GB/s average rate 24x7
 - Now embracing also Dune, BelleII, ...

Not a flat list of sites, but an ideal hierarchy:



Resource needs as in 2018

- Resources accessed via WLCG Distributed Computing
 - Initially MONARC hierarchical model, now more "cloudy"



MONARC

CERN

- Master copy of RAW data
- Fast calibrations
- Prompt Reconstruction

A second copy of RAW data (Backup)
Re-reconstructions with better calibrations

Analysis Activity
On average dimensioned to help ~ 50 physicists in their analysis activities

Anything smaller, from University clusters to your laptop

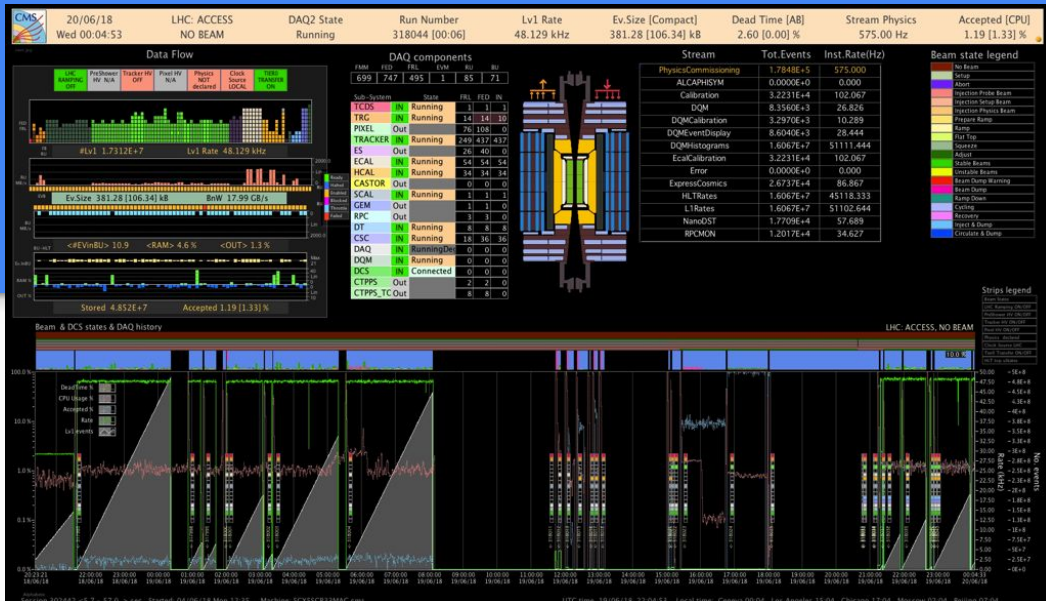
Some rough numbers on CMS Offline + Computing...

- Some **50MEur / y** annual expense on hardware
- **x2** for personnel and electric costs (which are starting to look AWFUL)
- Some **300 persons** participating activities (data taking, processing, CMSSW writing, R&D, ...)
- **1M jobs / day** (processing) + **2M jobs / day** (analysis)

How to find / recognize data?

- **The path / naming / meaning is quite different between data and simulation**
- Examples follow, but in a nutshell
 - Data: naming based on the Physics Trigger
 - Simulation: naming based on the physics process simulated

Real Data



Dataset Name	nLS	nEvents	<Rate> [Hz]
AlCaLumiPixels	1594	69967912	1883.0
AlCaP0	1594	286037474	7698.0
AlCaPhiSym	1594	56562262	1522.0
AlcaLumiPixelsExpress	1594	13872986	373.4
BTagMu	1594	2823820	76.0
Charmonium	1594	2849154	76.68
Commissioning	1594	422268	11.36
DQMONlineBeamspot	1594	2145709	57.75
DisplacedJet	1594	142048	3.823
DoubleMuon	1594	2223653	59.85
DoubleMuonLowMass	1594	1302240	35.05
EGamma	1594	10464644	281.6

- During data taking, data is selected and labelled by the High Level Trigger, and ends up in a series of **Primary Datasets (PDs)**
- Name is “usually” explicative:
 - **DoubleMuon**
 - **DisplacedJet**
 - **EGamma**
 - ...
- What is saved and sent to “Offline” are RAW events
 - **ZeroSuppressed readings of the CMS detector**
 - **~ 1 MB/ev**
- Primary datasets are processed in Prompt at the Tier-0 (CERN), and the results contains “physics objects”
 - **Tracks, electrons, jets, vertices, ...**
- Primary Datasets are NOT exclusive:
 - **Overlap at level of 10%**

Simulation

- Each Physics group has personnel expert in the tuning / setting of Particle generators, and in the definition of processes
 - **PDs can be $TT\bar{b}$, DrellYan, QCD, ...**
- Additional settings are needed to make explicit the running conditions you want to simulate
 - Which detector configuration (which pixel detector, w or w/o forward detectors, ...)
 - Which LHC parameters (PU configuration, CM energy, ...)
 - These are of course well defined in the real data

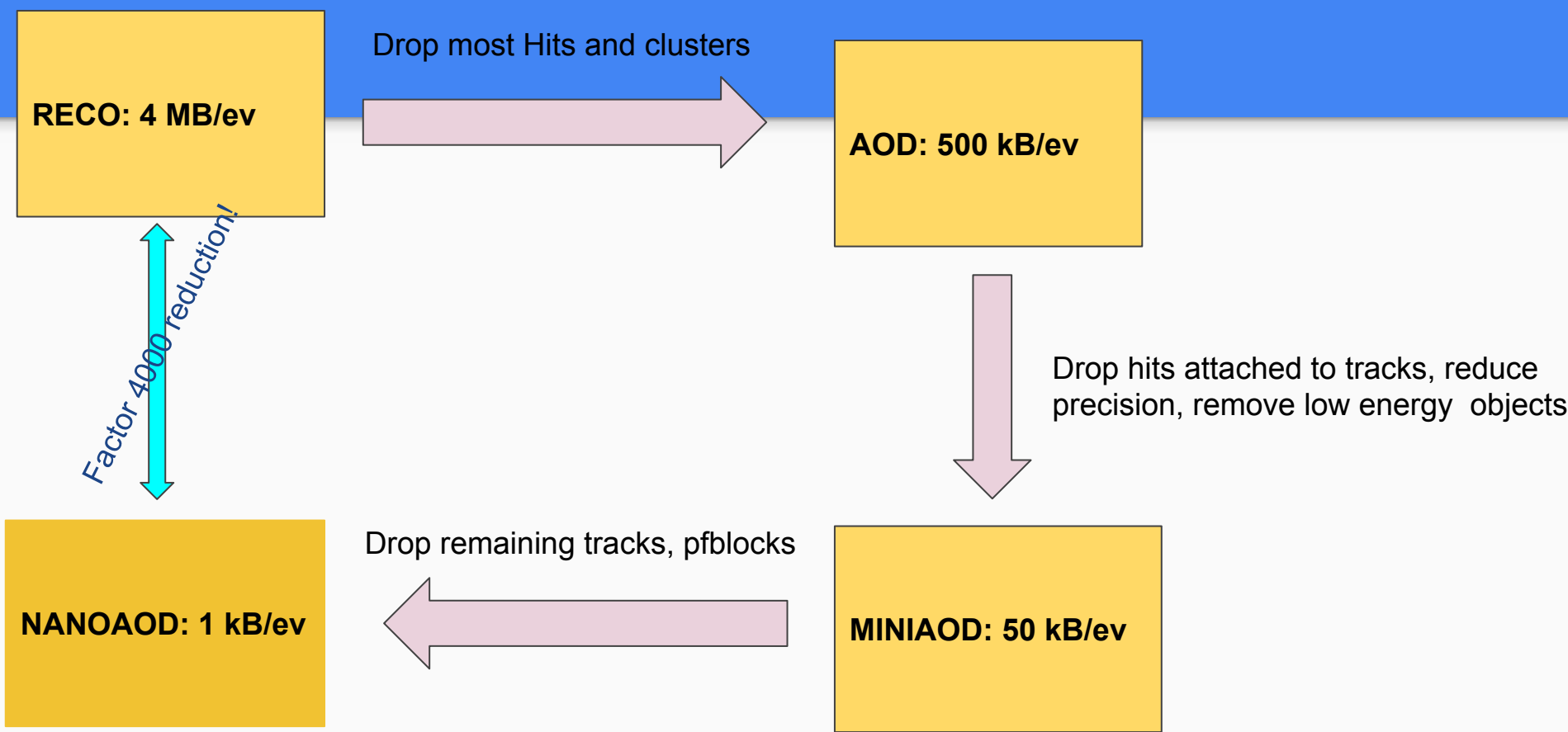
Then comes the reconstruction ... and in general: data tiers!

What is a Data Tier?

- The output of the Reconstruction is a set of **low/medium/high level physics objects**. Currently, the total amounts to **4 MB/ev** of content, and is defined as the **RECO** data tier
 - This is way too much for our disk capabilities, and mostly useless: only very few analysis need all the Tracker Silicon Hits, all Ecal Clusters etc
 - We do not save **RECO** any more apart from for a few debug samples
- Analysis objects are selected out of **RECO**, and saved as the **AOD** Tier
 - **< 500 kB/ev**
 - Still contains most of the objects needed for analysis
 - For those FEW analyses which need more, CMS defines SKIMS → very specific samples defined specifically for one analysis.
- **AOD** was the most widespread data tier used in RunI analysis
- Still, 500 kB/ev turned out to be too big when projected to the RunII, so we had to invent something ... →

Mini / Nano

- **MINIAOD** are the default / advised data tier for RunII analysis
 - Wrt to AOD, they
 - drop track hits other smaller stuff
 - Drop very small Pt tracks
 - reduce precision of objects (you do not need 12 digits on an off diagonal term of an error matrix)
 - They are @ 50 kB/ev, they would be “technically ok” for storage space for the next 10 y
- They can currently cover 95% of the analyses; the rest (5%) still needs either
- **They are ~ 50 kB/ev (10x smaller than AOD)**
- **AOD** or even Skims
- **NANO AOD** are the last incarnation of centrally produced data tiers. They are a ROOT FLAT NTUPLE which mostly drops tracks (you know only their number!) and go to 1 kB/ev
 - It is our bet for a faster / easier / less error prone analysis global effort



Hot to identify the content from the dataset name

- Dataset names are like (in data)

/ SingleMu / Run2012A-23May2012-v2 / AOD

- You should interpret this as

/ PRIMARYDATASET / ADDITIONALINFO / DATATIER

The PD is defined by the Trigger configuration; in this case events where a single muon with high enough Pt was found

This contains the **year** and the **Era** ("A") of data taking, the **date** of the reprocessing, the **version** (for example if the first one was stopped since broken)

The data tier as explained

For MC, more complicated

/ **DYToLL-M-50_1J_14TeV-madgraphMLM-pythia8** /
PhasellTDRFall17DR-PU200_93X_upgrade2023_realistic_v2-v1

Process simulated (DrellYan to lepton lepton etc...),
generator name ("madgraph"), **hadronizer name** ("pythia")

Campaign (Phasell for the TDR), **PileUp** (200), **CMSSW Release** (93X), **detector configuration** (upgrade2023), **type of conditions** (ideal, realistic_v2, after irradiation)

/ **GEN-SIM-RECO**
Data Tier (this means does contain RECO + simulation and generator information)

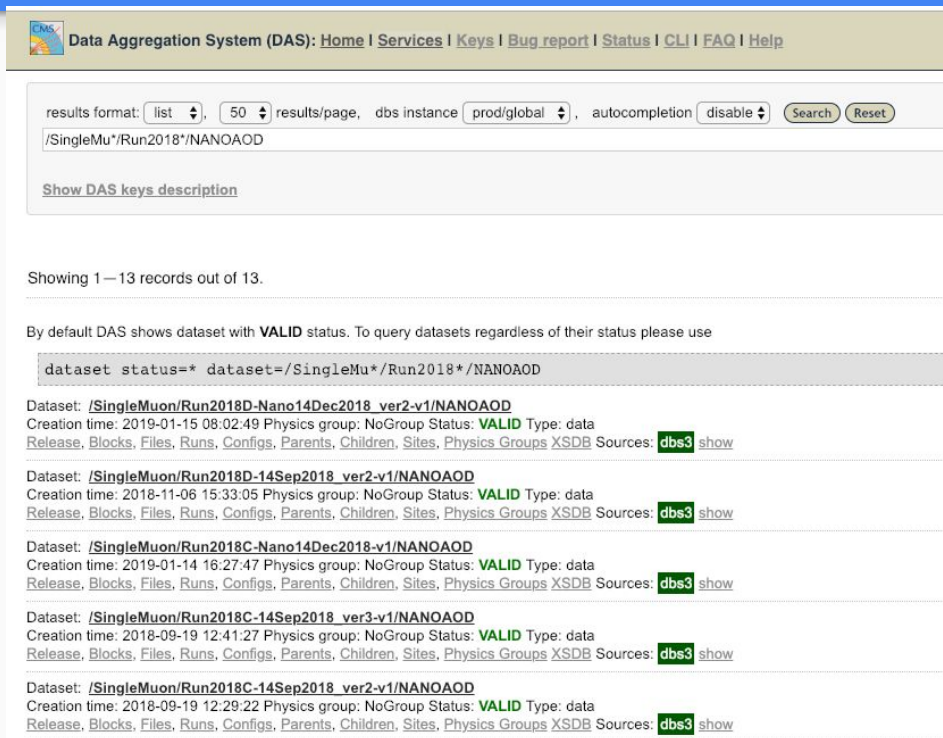
So, how to find the datasets for your analysis?

- Not trivial, but a few guidelines:
 - For data, you need to know / guess / test which datasets could contribute to your selection; sorry, no general rule
 - For MC, you need to be aware of the signals which can contribute, AND which is the best / most appropriate generator to use
 - For the mid (“ADDITIONALINFO”) part, there are guidelines from the PPD project, which suggest the best processing per year; for example
 - <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV2018Analysis>
- But where to see what is available ??? → **DAS**

Data Aggregation System

(<https://cmsweb.cern.ch/das/>)

- Data Aggregation Service (DAS) is the place where you can look for samples
 - A Certificate is needed (CERN, or mapping to CERN account)
- DAS lists datasets and their properties (**requestID, sites, run # and LS #....**) aggregating information from various services; you can even get the **configuration** used to process it, the **parent** datasets etc
- **Wildcards** can be used!



The screenshot shows the Data Aggregation System (DAS) interface. At the top, there is a navigation bar with links for Home, Services, Keys, Bug report, Status, CLI, FAQ, and Help. Below this, there is a search bar with a dropdown menu for 'results format' set to 'list', a '50' results/page selector, a 'dbS instance' dropdown set to 'prod/global', and an 'autocompletion' dropdown set to 'disable'. There are 'Search' and 'Reset' buttons. The search query is '/SingleMu*/Run2018*/NANOAOOD'. Below the search bar, there is a link to 'Show DAS keys description'. The results section shows 'Showing 1 — 13 records out of 13.' Below this, there is a text box containing the query 'dataset status=* dataset=/SingleMu*/Run2018*/NANOAOOD'. The results list shows four datasets, each with a link to 'Show DAS keys description' and a 'dbS3 show' button. The datasets are:

- Dataset: /SingleMuon/Run2018D-Nano14Dec2018_ver2-v1/NANOAOOD
Creation time: 2019-01-15 08:02:49 Physics group: NoGroup Status: VALID Type: data
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups XSDB Sources: dbS3 show
- Dataset: /SingleMuon/Run2018D-14Sep2018_ver2-v1/NANOAOOD
Creation time: 2018-11-06 15:33:05 Physics group: NoGroup Status: VALID Type: data
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups XSDB Sources: dbS3 show
- Dataset: /SingleMuon/Run2018C-Nano14Dec2018-v1/NANOAOOD
Creation time: 2019-01-14 16:27:47 Physics group: NoGroup Status: VALID Type: data
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups XSDB Sources: dbS3 show
- Dataset: /SingleMuon/Run2018C-14Sep2018_ver3-v1/NANOAOOD
Creation time: 2018-09-19 12:41:27 Physics group: NoGroup Status: VALID Type: data
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups XSDB Sources: dbS3 show
- Dataset: /SingleMuon/Run2018C-14Sep2018_ver2-v1/NANOAOOD
Creation time: 2018-09-19 12:29:22 Physics group: NoGroup Status: VALID Type: data
Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups XSDB Sources: dbS3 show

So let's start...

- Assume you know **which are the samples** (data and MC) you will need to process; you also know that MINIAOD (+SIM) are enough for you
- Assume “you know the Physics” (*I need to select events with 2 Muons with $P_t > 20$ GeV and 1 b-tagged jet*)

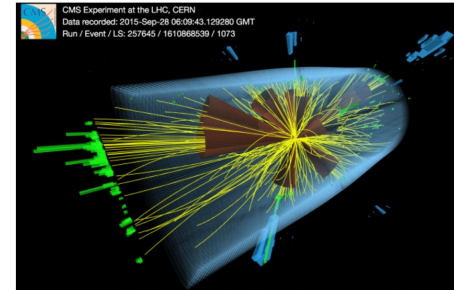
What to do?????

- **Answer from 2012:** write C++ code inside **CMSSW**, and use **CRAB** to process it
- **Answer from today:** you can do that, or use ROOT/C++ or Python macros to access the same quantities
- For simple analyses accessing small datasets, your choice
- For complex workflows (re-reconstruction of stuff) or very large datasets, better stay in the old schema

CMSSW (in one slide)

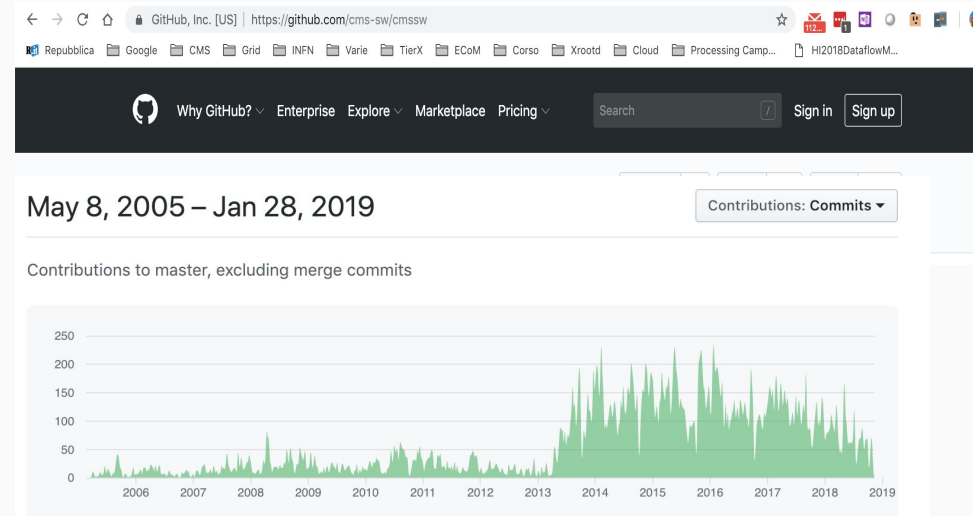
- It is the third major incarnation of CMS software
 - CMSIM: fortran, < Y2000
 - ORCA: C++, Y2000-2005
 - CMSSW: Y2005-
- It is a comprehensive FW used for **trigger**, **reconstruction**, **analysis**
- Runs at all levels in CMS, from the pit to the tier-0 to all the distributed sites
- **5 M lines of code, mostly C++ with some Python**
- For analysis: you will need to write an “Analyzer”, which can access data, process on that, write ntuples / histograms
- You will see at least in some of the exercises ... no time here

Welcome to CMS and CMSSW



The LHC smashes groups of protons together at close to the speed of light: 40 million times per second and with seven times the energy of the most powerful accelerators built up to now. Many of these will just be glancing blows but some will be head on collisions and very energetic. When this happens some of the energy of the collision is turned into mass and previously unobserved, short-lived particles – which could give clues about how Nature behaves at a fundamental level - fly out and into the detector. Our work includes the experimental discovery of the **Higgs boson**, which lead to the award of a Nobel prize for the underlying theory that predicted the Higgs boson as an important piece of the standard model theory of particle physics.

On Github, Apache 2 license (Open Source)



~800 developers, ~200 contributions/week

But, how to actually run your code????

- If you use a simple code, and you need to run on few (say less than 1 M) events, you can probably do on your machine, directly with **CMSSW / Python / ROOT**
- For larger data/CPU utilization, you NEED to use [CRAB](#)
- CRAB can:
 - **Run** your code remotely, on literally thousands on CPUs
 - Automatically **optimize** the # of CPUs to use to reduce the ETA
 - **Choose** the best sites where to run jobs, and revert to slightly worse if that helps the ETA
 - Trigger dataset **recalls** to disk if the samples are only on tape
 - Automatically **retry** in case of errors
 - **Save** in DBS/DAS the results as a new CMS official sample

Documentation / Who can help!

- Always use as entry point the CMS the [Offline Workbook](#)
- More specific explanations on Computing can be found [here](#)
- If you still need help:
 - If it is a code / C++ problem
 - If you can try and identify which is the subsystem affected (Muons, Tracking, ...) you can use the proper Hypernews: Category: Software Development [here](#)
 - If it seems a generic computing problem (access to data, something very basic failing):
 - You can use hn-cms-computing-tools@cern.ch
 - If it about which data to use:
 - you can use hn-cms-physics-validation@cern.ch

Some more links in the next page!

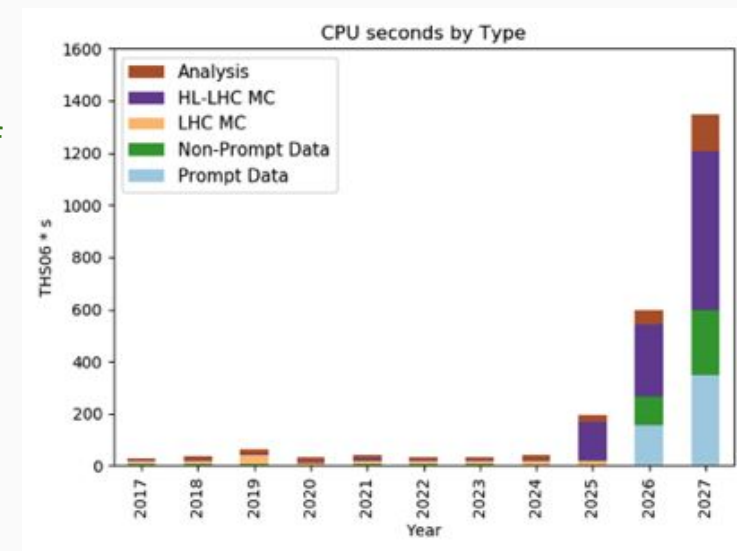
Contacts	Documentation
cms-ppd-coordinator@cern.ch cms-offcomp-coordinator@cern.ch	PPD Main Twiki Offline Main Twiki Computing Main Twiki
hn-cms-dataset-definition@cern.ch	DDT Twiki
hn-cms-computing-tools@cern.ch	DAS
hn-cms-prep-ops@cern.ch	PdmV Twiki McM
	Computing Model Workbook
hn-cms-offlineAnnounce@cern.ch hn-cms-relAnnounce@cern.ch	Offline Workbook SW Guide
hn-cms-physTools@cern.ch	MiniAOD Workbook
hn-cms-phedex@cern.ch	XROOTD doc Phedex - Phedex Workbook
hn-cms-relval@cern.ch hn-cms-physics-validation@cern.ch	PdmV Twiki
hn-cms-evfdqmannounce@cern.ch	DQM Twiki
hn-cms-data-certification@cern.ch	DQM-DC Twiki RunRegistry
hn-cms-data-certification@cern.ch	JSON File Twiki
hn-cms-luminosity@cern.ch cms-dpg-conveners-bril@cern.ch cms-pog-conveners-lum@cern.ch	brilcalc Doc bril dpg lumi pog
hn-cms-alca@cern.ch	AICaDB Twiki GlobalTag Twiki

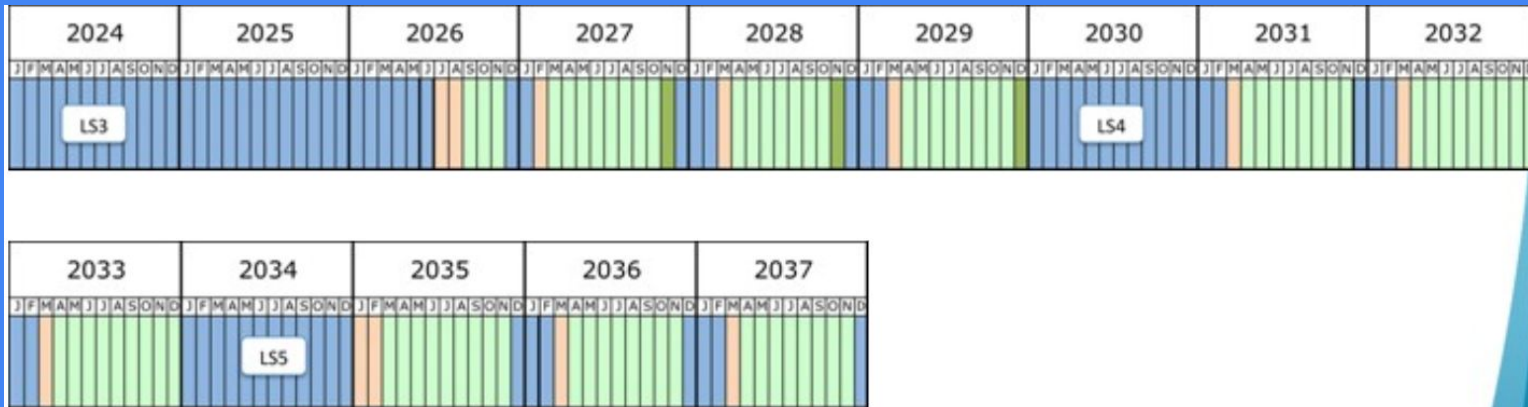
Who can help?

- Clearly it is much easier if you can get help from the colleagues sitting on your desk, but it is not always possible
- If it is a code / C++ problem
 - If you can try and identify which is the subsystem affected (Muons, Tracking,) you can use the proper Hypernews: Category: Software Development [here](#)
- If it seems a generic computing problem (access to data, something very basic failing):
 - You can use hn-cms-computing-tools@cern.ch
- If it about which data to use:
 - you can use hn-cms-physics-validation@cern.ch

The future ...

- As Roberto said, RunIII (2021-2023) is only marginally more complicated than RunII, computing-wise
 - Should be doable with less than 2x the resources of RunII, and it is shifted by 5 years
- PhaseII/RunIV (2026+) is a different beast:
 - Trigger rate 1 kHz → 10 kHz (10x scaling everywhere)
 - Pileup 35 → 200 (and many algorithms are superlinear with <PU>)
 - New detectors (like HGCal and new Pixels) have many more DAQ channels: 1 → 7 MB/ev RAW data
- All in all, easily expect ~50x resource needs increase 2018 vs 2027
- **Clearly not feasible money wise ..**





... where we need TONS of manpower;
directions of work are

1. Machine Learning, Deep Learning
2. Heterogeneous computing
 - Move mission critical fragments of code to FPGA, GPU, TP, (Quantum computing)
3. Novel operating modes
 - No RAW data saved, use BigData tools in collaboration with Google, IBM, Intel
4. New computing infrastructures
 - Use Supercomputers, use Commercial Clouds, ...
5. ...

