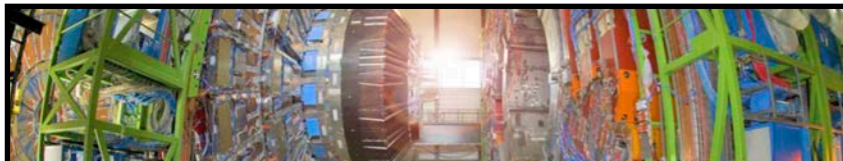


# Exercise: Physics Performance and Datasets

Norraphat SRIMANOBHAS  
(Chulalongkorn U., CERN)  
On behalf of PPD

**CMSDAS 2019 Pisa,  
INFN Pisa,  
28 Jan - 1 Feb 2019**

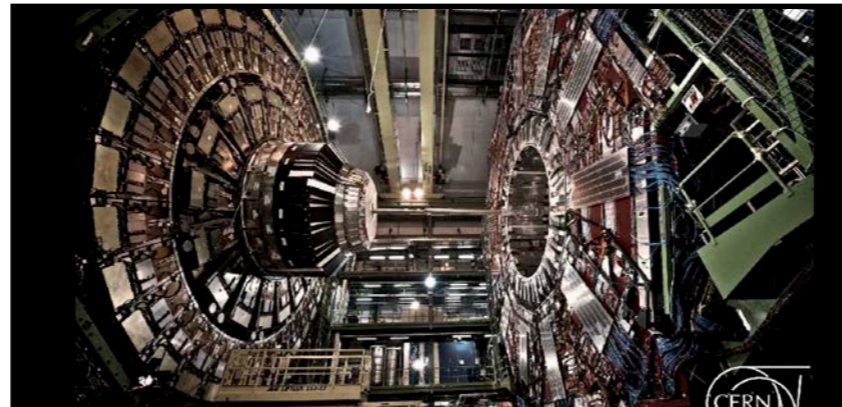
# Credits



## Data Preparation at CMS

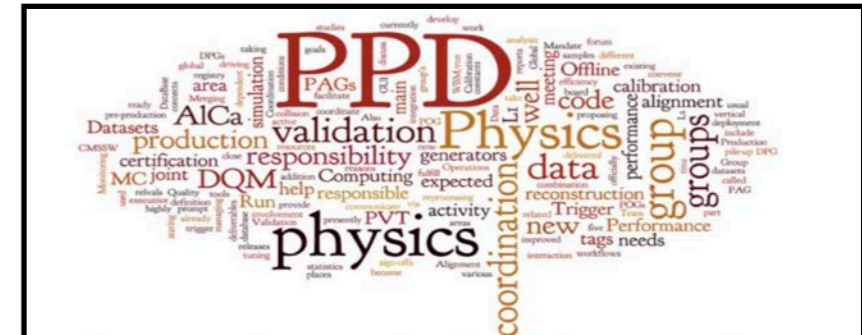
G. Franzoni (CERN)  
17 February 2016  
CMS Data Analysis School @Taipei

Acknowledgements: G. Cerminara (CERN)  
N. Srimanobhas (U Chulalongkorn), M. Musich (U. Louvain)



## OFFLINE DATA PREPARATION

G. Cerminara (CERN)  
CMS Induction Session - 15th of June 2017 - CERN




## Overview of PPD Operations

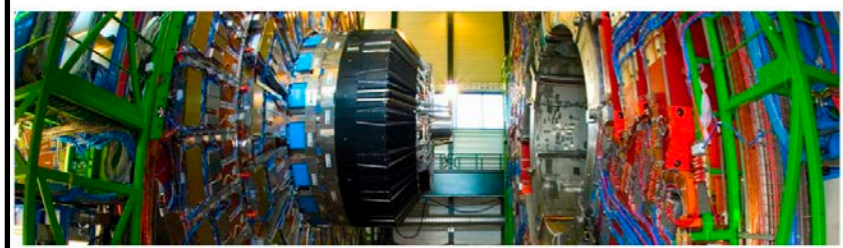
Acknowledgements: Giovanni Franzoni, Broen Besien, Kaori Maeshima, Virginia Azzolini, Gurpreet Singh Chahal, Francesco Fabozzi

Arun Kumar  
National Taiwan University

CMS Induction Course  
15-16 June 2017




## Offline Data Preparation



Gurpreet Singh Chahal (Imperial College London and Durham University)  
on behalf of PPD and Offline & Computing teams

CMS Induction Course, 1 February 2018



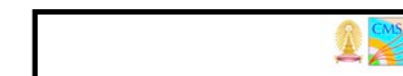
## Exercises about Physics Performance and Datasets

Gurpreet Singh Chahal (Imperial College and IPPP Durham University)  
Tongguang Cheng (Purdue University Northwest)

on behalf of PPD and Generator teams

CMS DAS, DESY, 10 Sept 2018

and all feedbacks to my previous talks ...



### Exercise: Monojet analysis

CMSDAS 2015, Bari


Norraphat Srimanobhas  
Chulalongkorn University, Thailand



### Matching in Analysis Tools

Norraphat SRIMANOBHAS, Chulalongkorn University (TH)  
(srimanob@mail.cern.ch)

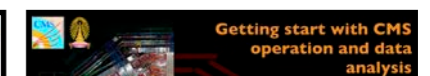
PAT Tutorial  
29 June - 3 July 2015



### Data-Monte Carlo preparation at CMS

Norraphat SRIMANOBHAS  
(Chulalongkorn U, Thailand, CMS Collaboration)

Tata Institute of Fundamental Research  
Mumbai, India  
November 13, 2016



### Getting start with CMS operation and data analysis

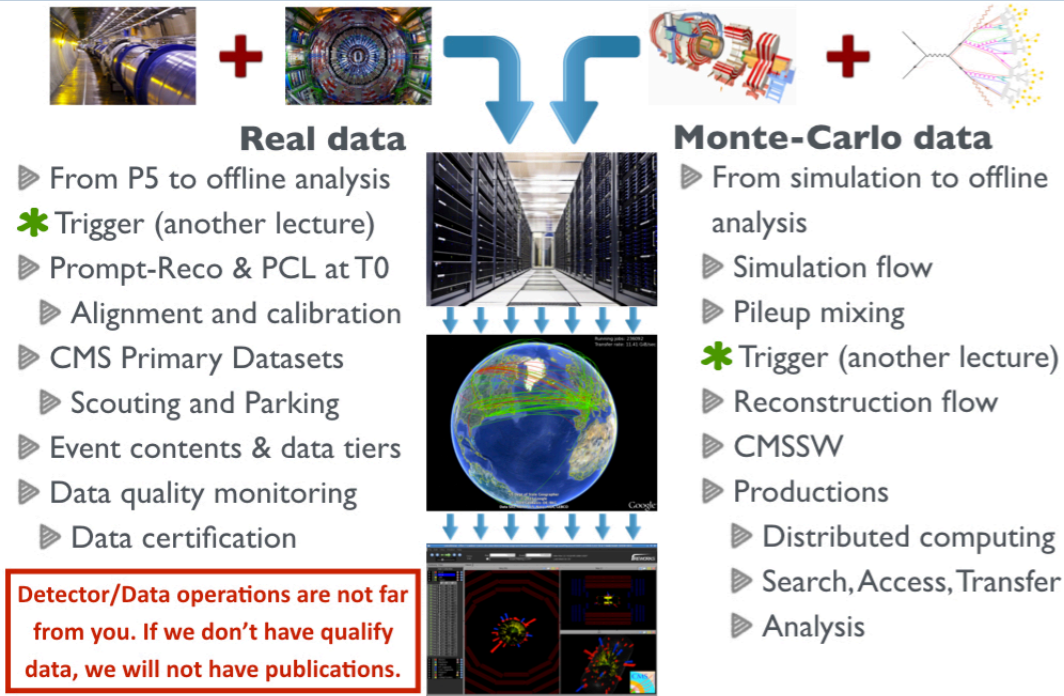
PART I: CMS Operation and Data-MC Preparation

Norraphat SRIMANOBHAS  
Chulalongkorn U, Thailand, CMS Collaboration

Sultan Qaboos University,  
Muscat, Oman  
March 2-4, 2016

# Outline

## Outline: PPD + O&C



N. SRIMANOBHAS (Norrapat.Srimanobhas@cern.ch)

Exercise: Physics Performance and Datasets

4

## Outline: Exercises

In this set of exercises, we have

- Exercise 0: Starting your RunII analysis
- Exercise 1: Exploring dataset using DAS : searching 2018 EGamma dataset as an example
- Exercise 2: Explore information for a Monte Carlo miniAODSIM sample from DAS, McM and pMp
- Exercise 3: Miscellaneous details about datasets
- Exercise 4: Generating MC NanoAOD events from scratch
- Exercise 5: Explore GlobalTag through cmsDBbrowser
- Exercise 6: Compute the integrated luminosity collected by CMS in RunII

N. SRIMANOBHAS (Norrapat.Srimanobhas@cern.ch)

Exercise: Physics Performance and Datasets

29

## Introduction

- ▶ Real data flow
- ▶ Monte-Carlo flow
- ▶ Analysis

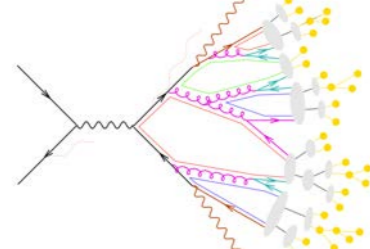
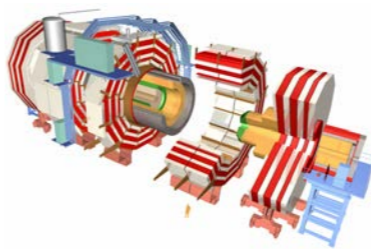


**Our target in CMSDAS**

## Exercises

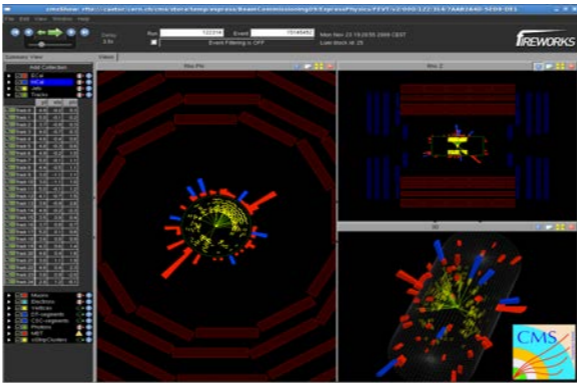
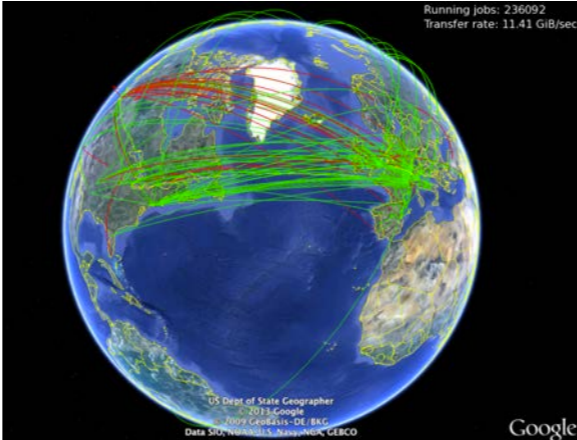
- ▶ Exploring CMS datasets
- ▶ MC production from scratch
- ▶ Conditions
- ▶ Compute integrated luminosities used in your analysis

# Outline: PPD + O&C



## Real data

- ▶ From P5 to offline analysis
- \* Trigger (another lecture)
- ▶ Prompt-Reco & PCL at T0
  - ▶ Alignment and calibration
- ▶ CMS Primary Datasets
  - ▶ Scouting and Parking
- ▶ Event contents & data tiers
- ▶ Data quality monitoring
  - ▶ Data certification



## Monte-Carlo data

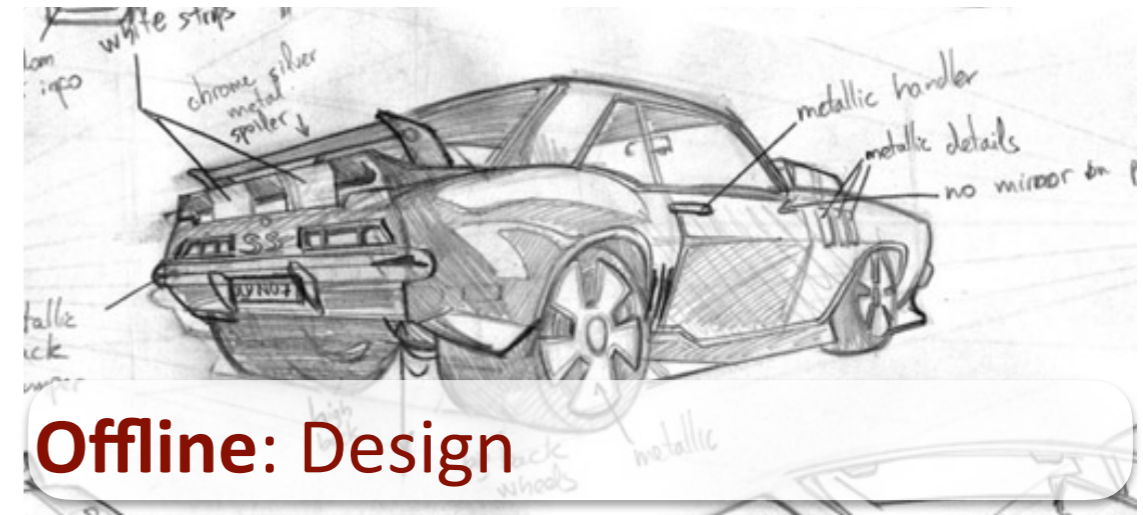
- ▶ From simulation to offline analysis
- ▶ Simulation flow
- ▶ Pileup mixing
- \* Trigger (another lecture)
- ▶ Reconstruction flow
- ▶ CMSSW
- ▶ Productions
  - ▶ Distributed computing
  - ▶ Search, Access, Transfer
  - ▶ Analysis

**Detector/Data operations are not far from you. If we don't have quality data, we will not have publications.**

# Data preparation and coordination areas

In this talk we focus on 2 main CMS coordination areas:

- **Offline & Computing (O&C)**
  - CMSSW software development, event reconstruction and simulation
  - data processing and Simulated events generation, events storage and management
- **Physics Performance and Datasets (PPD)**
  - data quality & certification
  - alignment & calibrations
  - software validation
  - management of Monte Carlo requests
  - organization and configuration of datasets and data processing



# From P5 to offline



Events collected by CMS reach the Tier-0 at CERN for tape archival, organization, and processing.

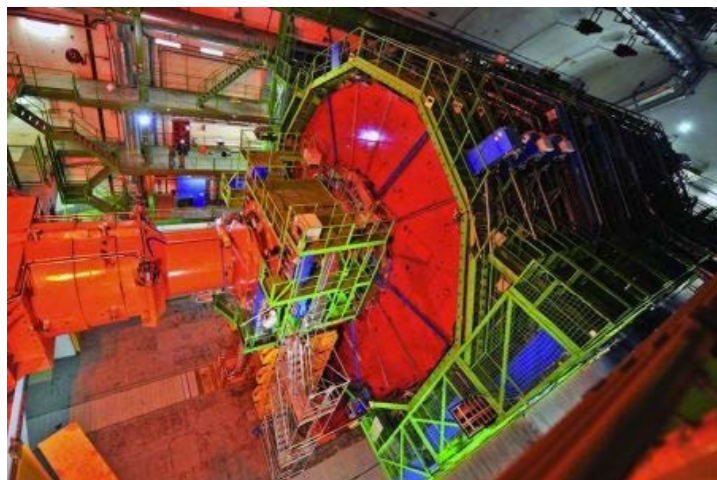
## Data steams

- **Express**: Available ~2h after data collection for prompt feedback and calibration. About 100 Hz bandwidth shared by Calibrations, Detector, Physics monitoring.
- **Alignment & Calibration**: Dedicated event selection & event content devised for calibration process.
- **Physics**: Split into primary datasets and promptly reconstructed for physics analysis.
- **Other specialized streams**: Scouting, Parking.



## Data rates

- **Run I**: 300 Hz Prompt-Reco + 300-600 Hz of parked data
- **Run II**: 1 kHz of Prompt-Reco + high rate of scouting data with reduced event content + parking



# From P5 to offline



[cmsonline.cern.ch](http://cmsonline.cern.ch) (1) Basic DAQ, (2) Data streams and their rates

# Prompt reconstruction and PCL at Tier-0

$t=0$

CMS aims at providing analysts with reconstructed data within days after data collections.

**Express processing:** Data reconstructed for

- ▶ Monitoring
- ▶ ECAL pedestal
- ▶ Calibration
- ▶ Pixel counting for luminosity

**Prompt Calibration Loop (PCL):** Express data is used as input to automated calibration workflows running at Tier-0 (or online)

- ▶ beamspot (LS by LS)
- ▶ ECAL transparency corr
- ▶ SiStrip bad-channels/gains
- ▶ SiPixel alignment

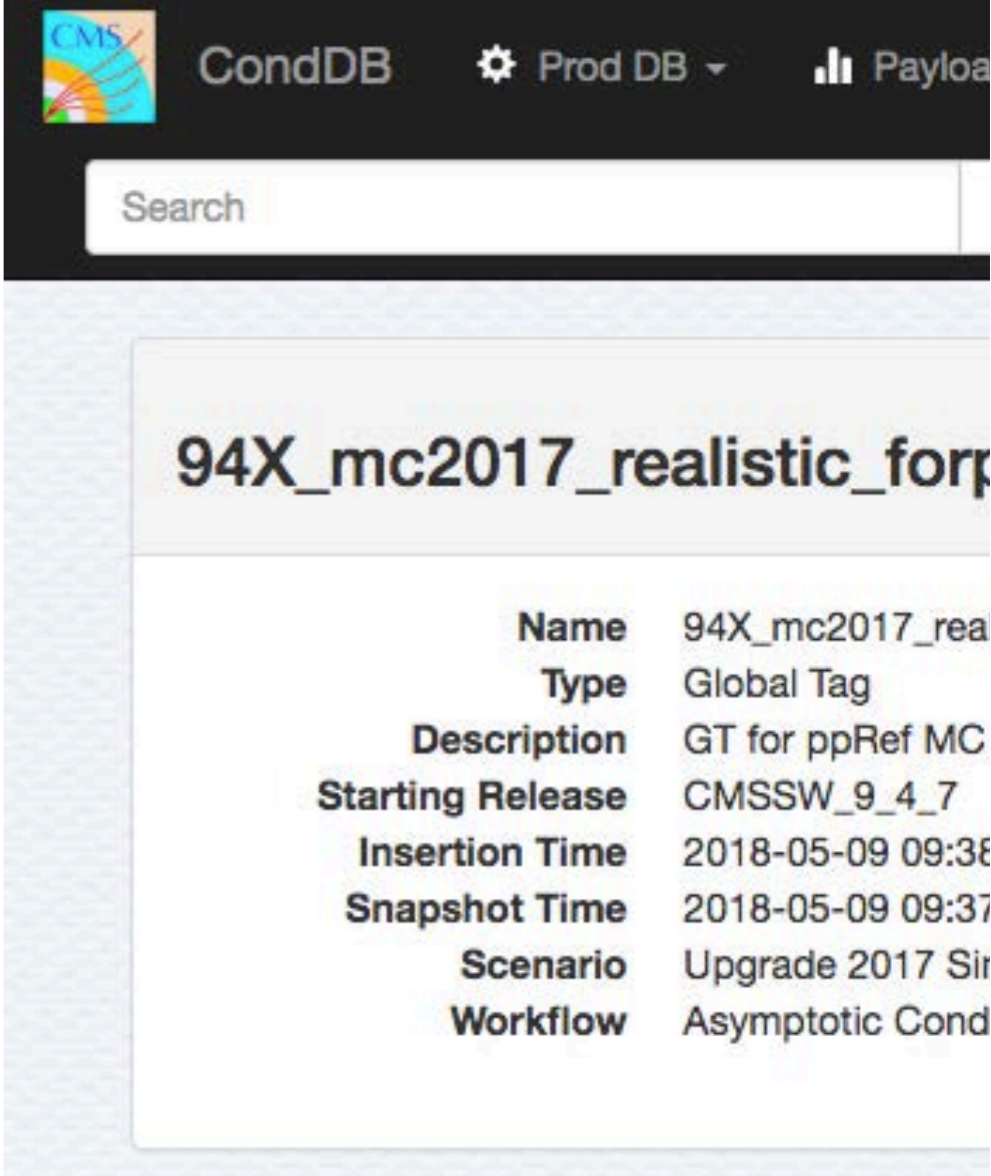
**Prompt Reconstruction:** Physics streams are reconstructed consuming calibrations computed by PCL. These are datasets for analysis. We normally start prompt reconstruction within 48 hour.

48h



# Calibration workflow

- Providing the most up-to-date alca conditions at all stages of the data and Monte Carlo processing is a major challenge involving all DPGs, POGs, AlCa team in PPD
- Critical to sustain the quick pace of analysis turn-around
  - ▶ Prompt-reconstruction must be of near-perfect quality;
- Few level of workflows depending on the time scale of updates
  - ▶ **Quasi-online calibrations** for HLT and express processing
  - ▶ **Prompt calibrations:** monitor and update conditions expected to vary run-by-run, or even more frequently - essential to guarantee performance of prompt reach
  - ▶ **Offline workflows** for data reprocessing and analysis level conditions



The screenshot shows the CMS CondDB interface. At the top, there is a navigation bar with the CMS logo, 'CondDB', a gear icon for 'Prod DB', and a bar chart icon for 'Payload'. Below the navigation bar is a search input field. The main content area displays the details for a specific Global Tag, '94X\_mc2017\_realistic\_forp'. The details are presented in a table-like format with labels on the left and values on the right.

<b>Name</b>	94X_mc2017_rea
<b>Type</b>	Global Tag
<b>Description</b>	GT for ppRef MC
<b>Starting Release</b>	CMSSW_9_4_7
<b>Insertion Time</b>	2018-05-09 09:38
<b>Snapshot Time</b>	2018-05-09 09:37
<b>Scenario</b>	Upgrade 2017 Sir
<b>Workflow</b>	Asymptotic Cond

# Primary datasets

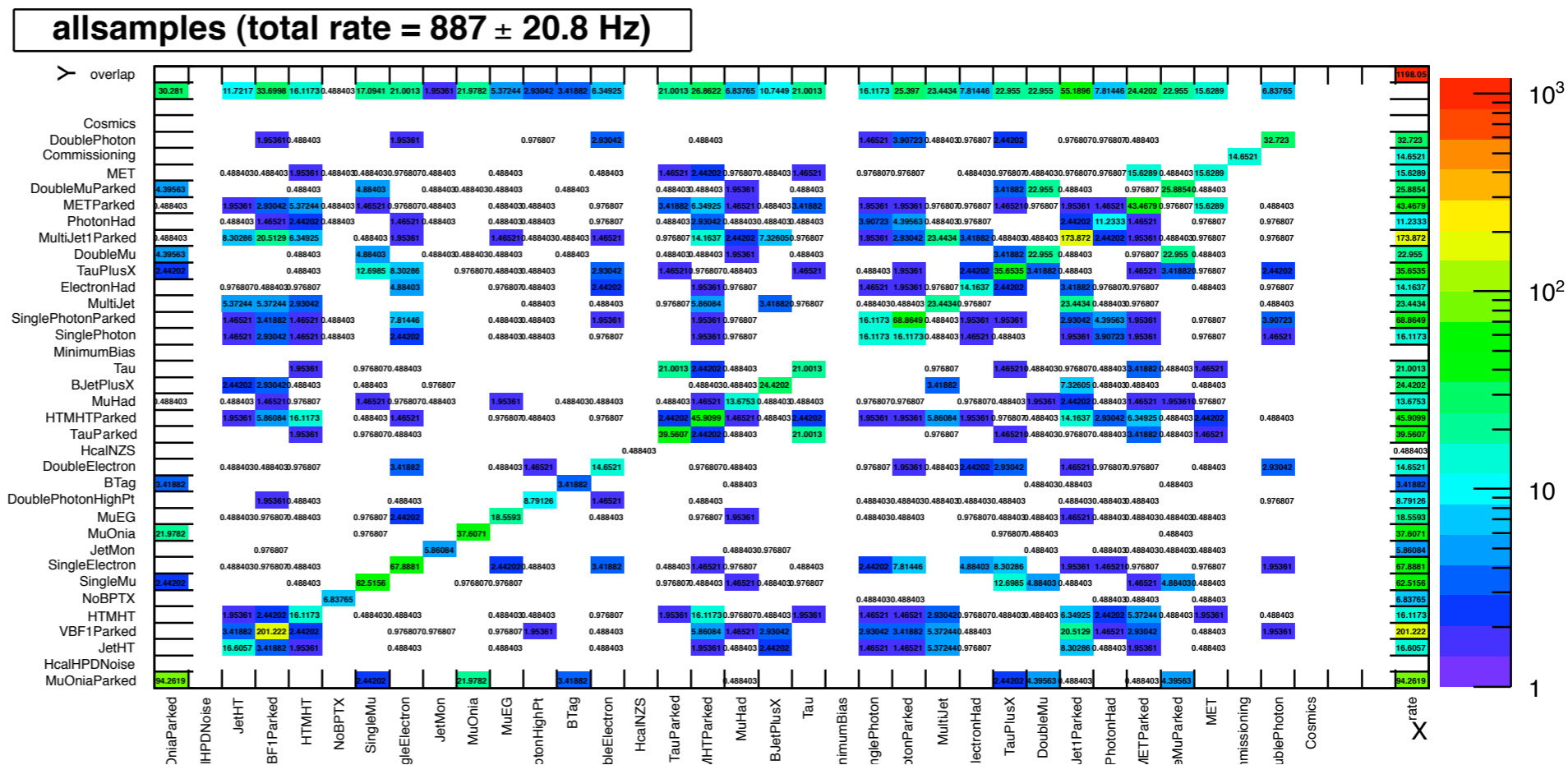
The physics streams from P5 are split to Primary Datasets (PD) on the basis of HLT results in order to group events with related topology and limit replication of events (PD's overlap).

## Constraints

- Physics: Definition centered on physics objects (i.e. SingleElectron, JetMET)
- Processing & Handling: Proper event rates for each PD, to be able to handle by Tier-2.

On top of the primary datasets we can deploy “central skims” to

- customise event content
- Reduce rate using also RECO quantities.



# Data scouting and data parking

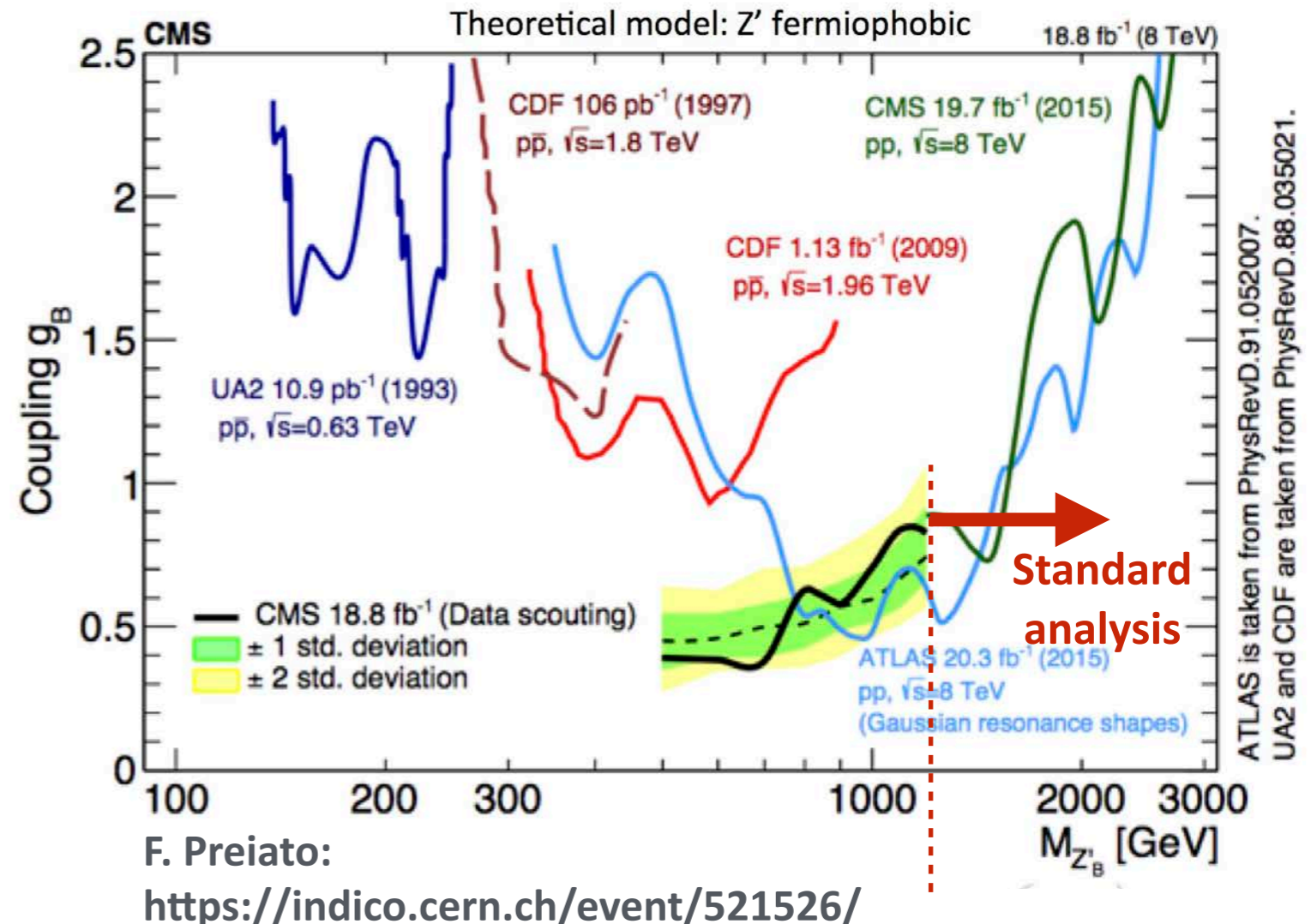
As LHC luminosity rises, increasing trigger thresholds pose a challenge for analyses in CMS

■ Trigger rates are constrained by the CMS prompt reconstruction system, which cannot process much more than 1 kHz of events.

■ Getting around this limit:

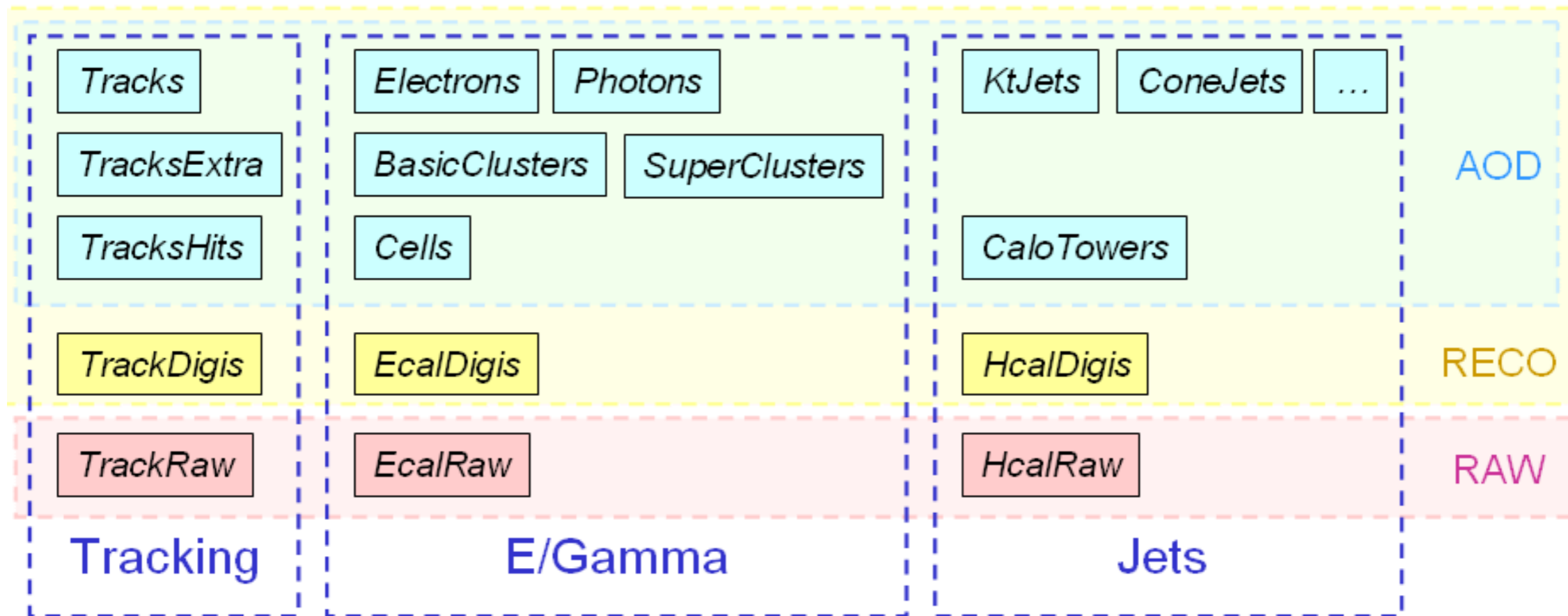
☑ **Data parking:** send events from the HLT to tape without reconstructing them

☑ **Data scouting:** save only a small subset of the event content (e.g., only the HLT-level jet objects)



# Event contents and data tiers

- Event information from each step in the simulation and reconstruction chain is logically grouped into what we call a **data tier**.
- Examples of what are in RAW/RECO/AOD



- Currently, two step further in data reduction:
  - ✓ **MiniAOD**: ~10-15% of AOD size, designed for 90% analyses at CMS.
  - ✓ **NanoAOD**: design for analysis with bare root, cover 30-50% of analyses.

# Data Quality Monitoring (DQM)

DQM is the tool to produce plots while running RECO (or any CMSSW workflow). There are two main areas of DQM application:

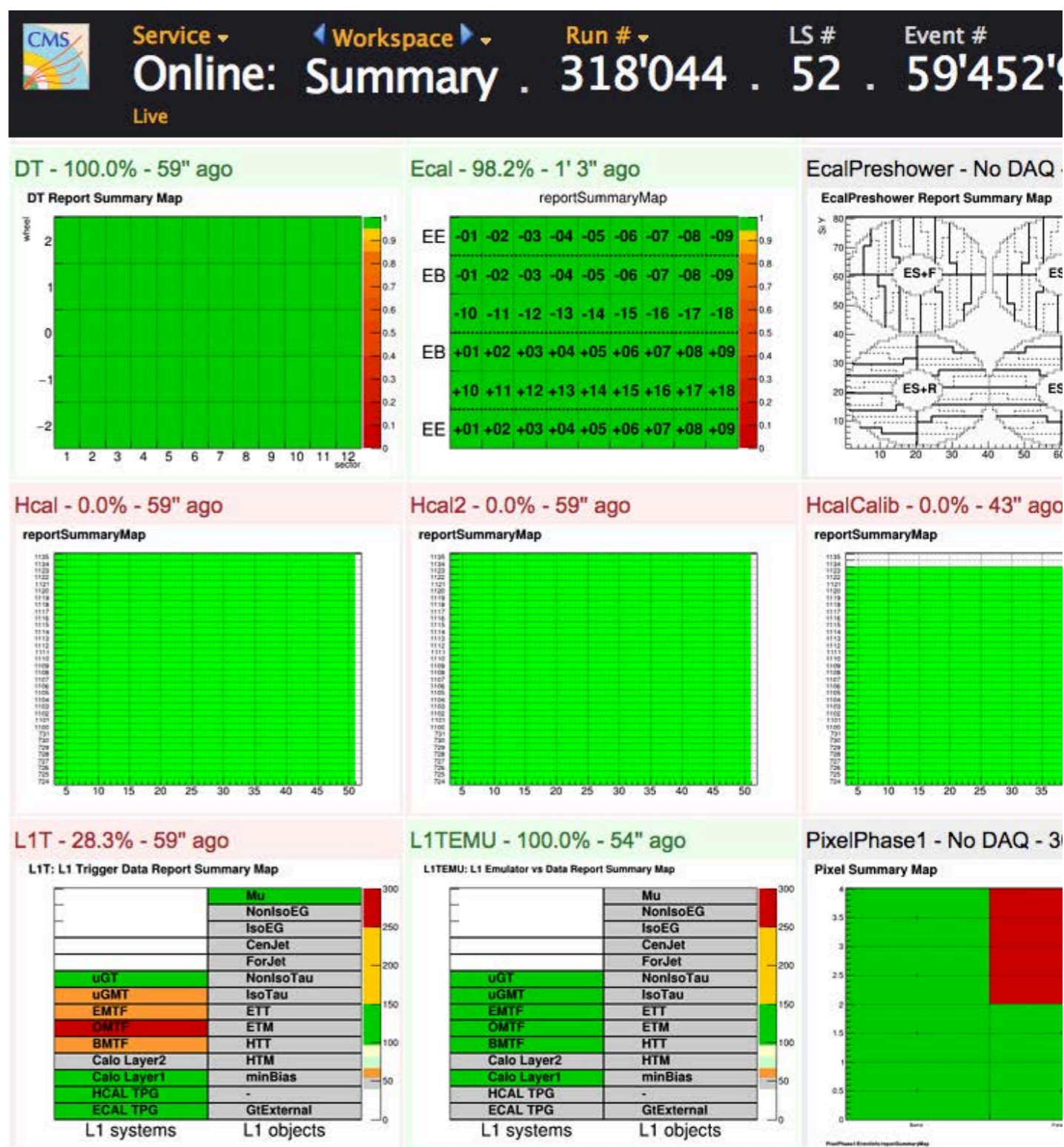
■ **Online:** process events selected by HLT to display variables in the control room with very low latency

☑ Live monitoring of detector performance during data taking

■ **Offline:** process all events while they are simulated or reconstructed and fill diagnostic plots for detailed monitoring of the performance

☑ Data certification

☑ Validation+verification



<https://cmsweb.cern.ch/dqm/online>  
<https://cmsweb.cern.ch/dqm/offline>

# Data Certification (DC)

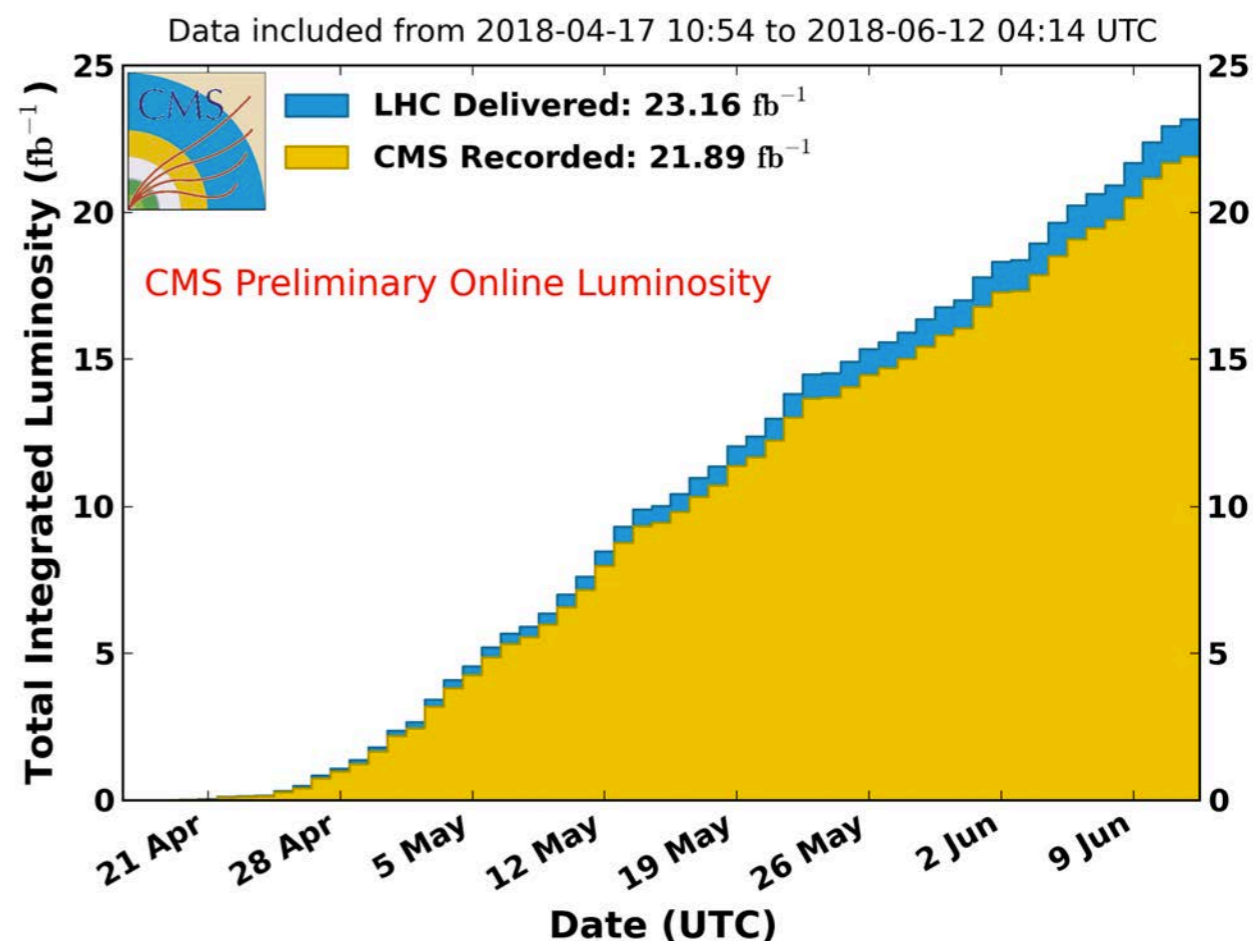
We run analysis only LumiSections (Sub-section of a run during which time the instantaneous luminosity is unchanging, ~23s of RUN) which we consider to be good. This task is done by team of experts in detector and physics objects.

- **Golden**: require all sub-detectors/POGs to be "GOOD"
- **Muon-only**: no requirements on calorimeters
- **DCS-only**: require only tracker to be powered

The format of file is in JSON format

- Weekly for PromptReco
- After each major reprocessing

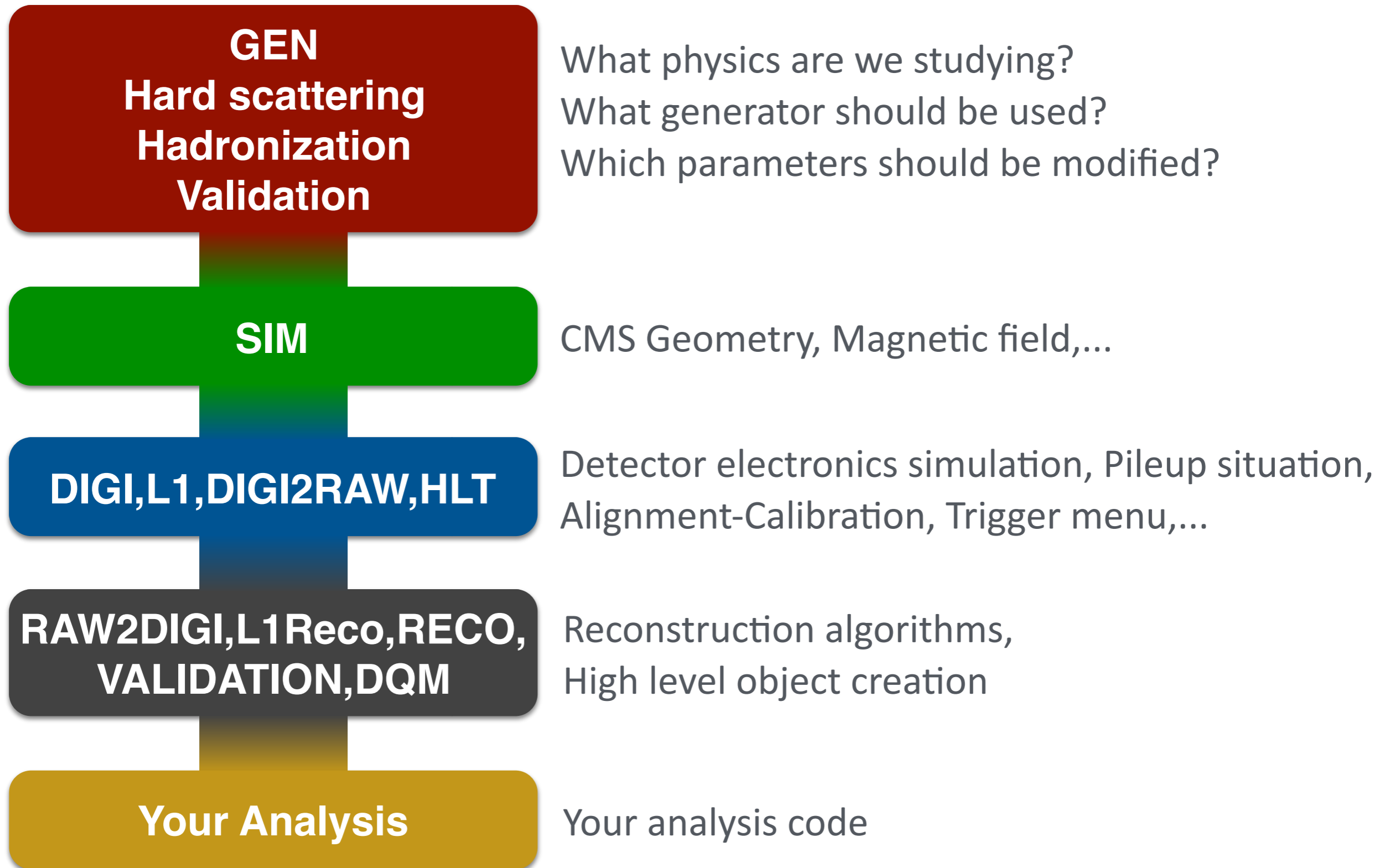
CMS Integrated Luminosity, pp, 2018,  $\sqrt{s} = 13$  TeV



```
{"273158": [[1, 1279]], "273302": [[1, 45  
"273408": [[1, 6]], "273409": [[1, 309]],  
113], [115, 412]], "273448": [[1, 391]],  
"273493": [[1, 233]], "273494": [[1, 192]  
[[1, 173]], "273725": [[83, 252], [254, 2  
"274159": [[1, 43]], "274160": [[1, 207]]  
"274240": [[1, 40], [42, 82]], "274241":
```

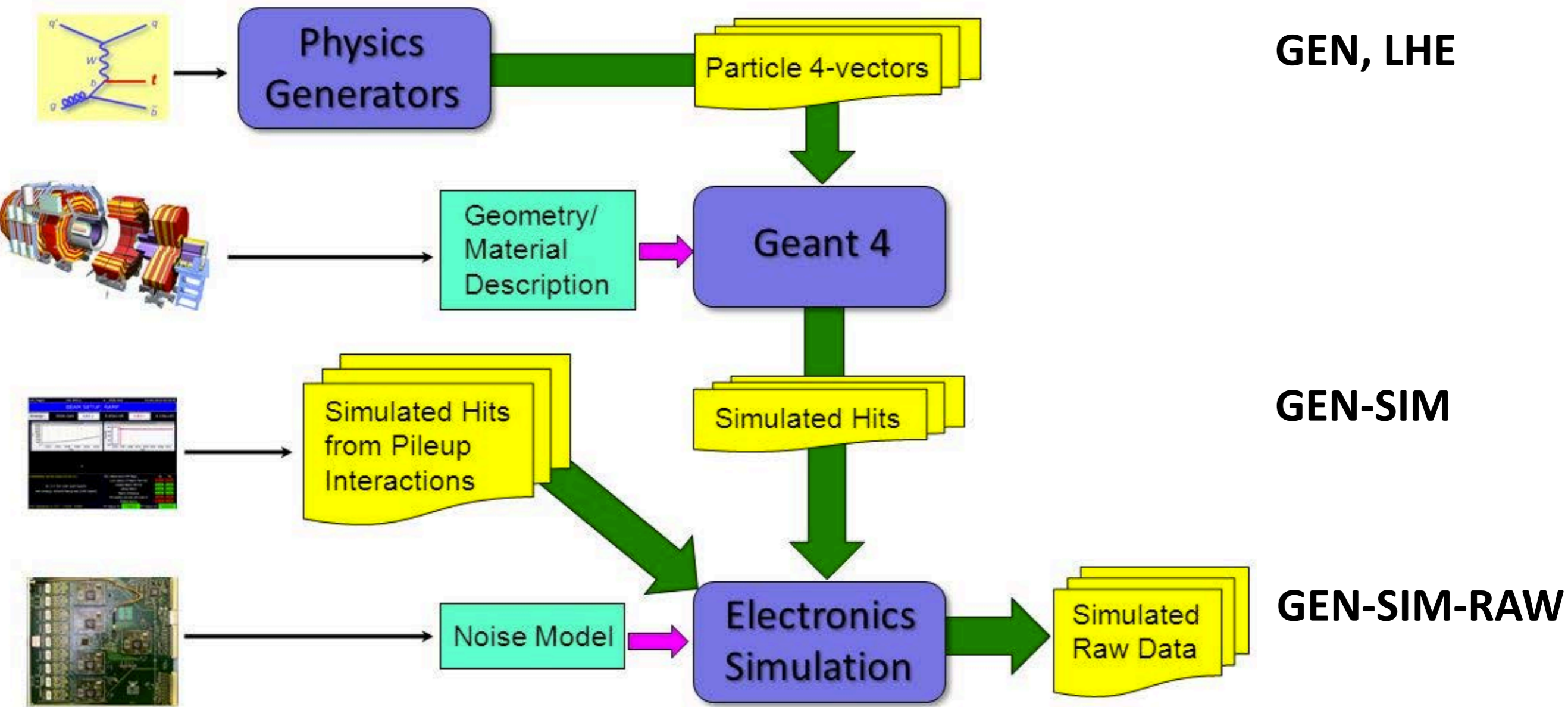
↑ Run no.      ↑ Range of LumiSections

# From collision/simulation to physics analysis



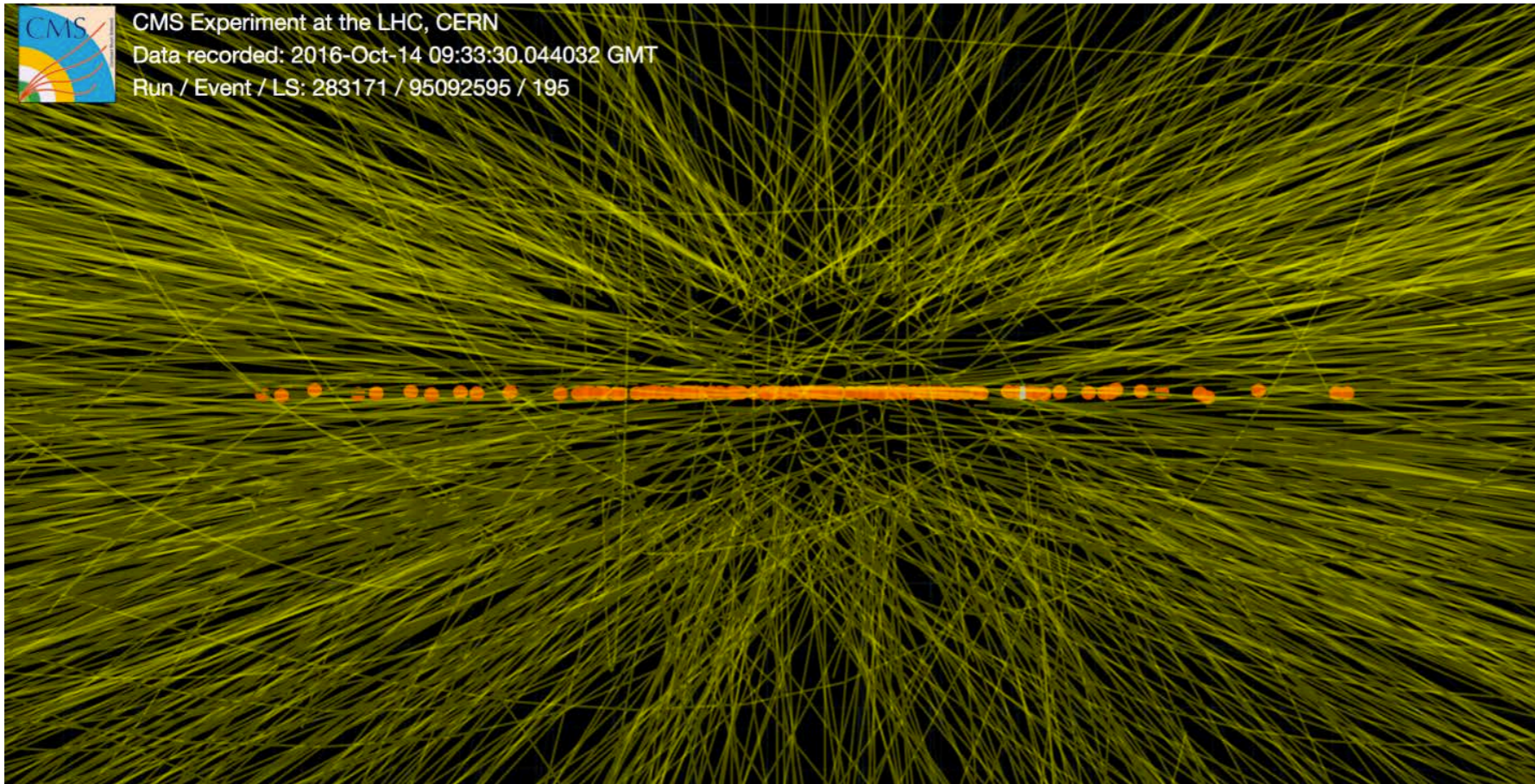
# Simulation framework

- Event simulation algorithms are implemented as module, communicating via the Event.
- The simulation sequence aims at producing MC truth and RAWDATA as it comes from point 5.





# Pileup

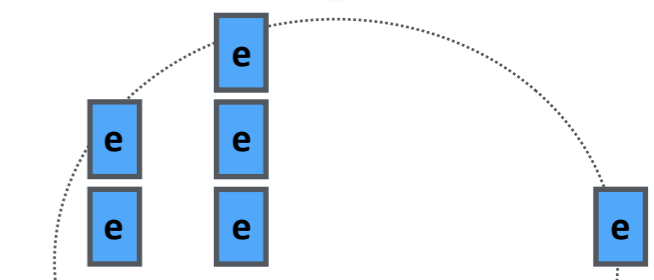
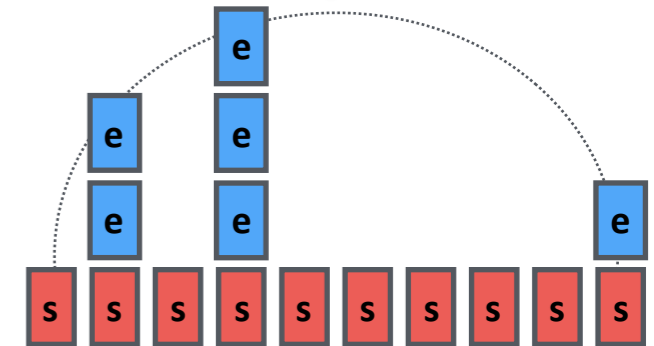
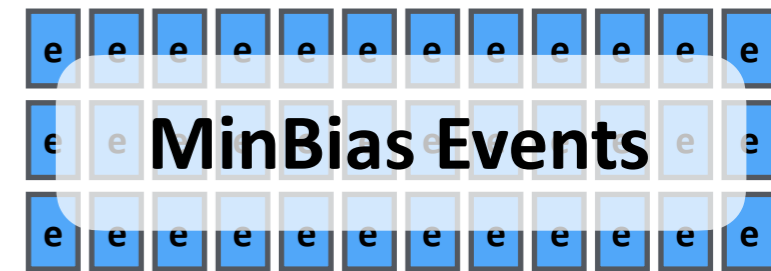


## Classic mixing

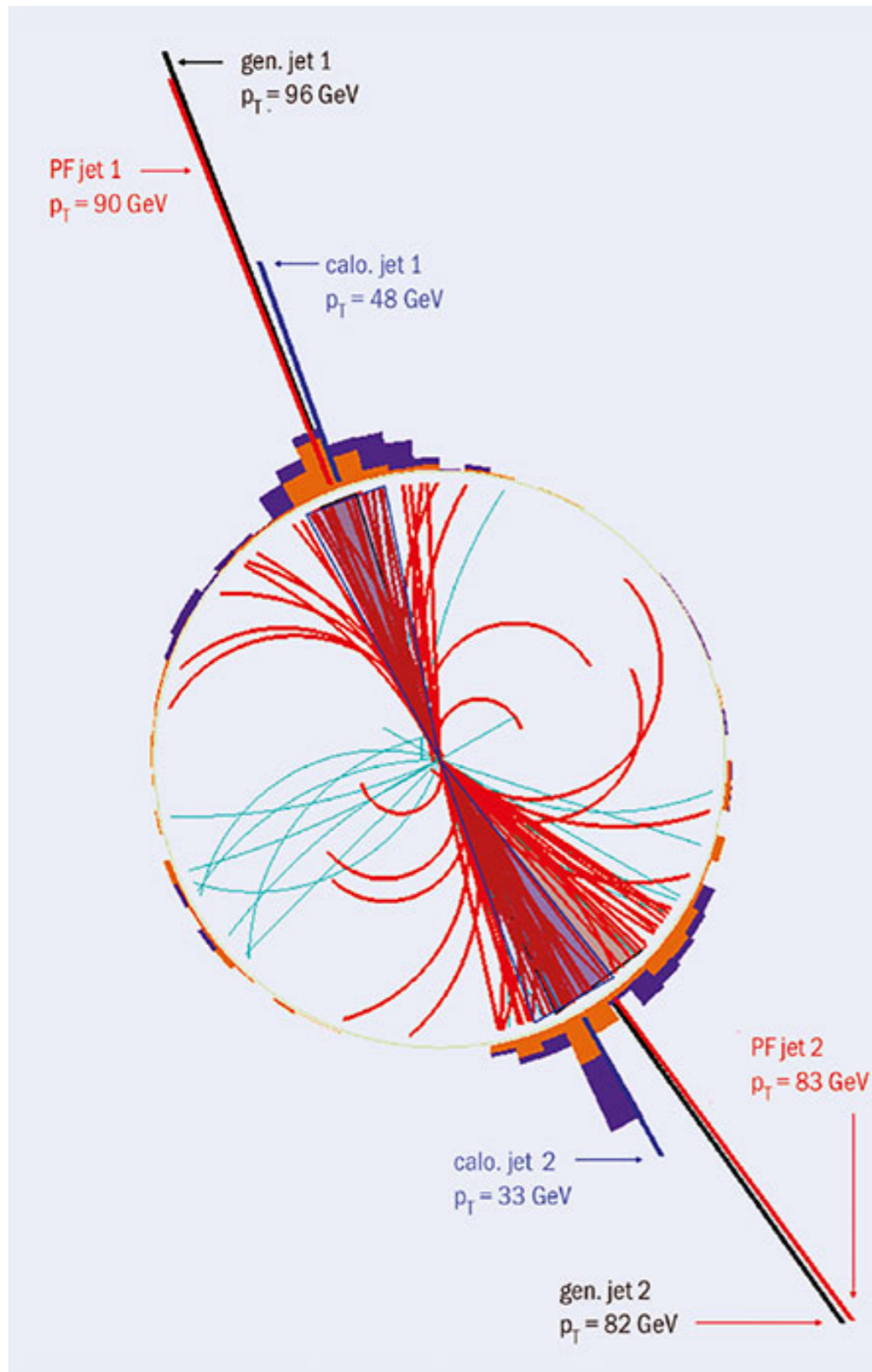
- GENSIM Signal (MC Hard-scatter event) is overlaid with GENSIM MinBias with chosen pileup configuration.

## Pre-mixing

- MinBias events in RAWSIM format are overlaid on empty single neutrino events using a chosen pileup configuration. Digis made in this step are converted to RAW.
- 1-1 combination of PreMixed event - signal event. RawToDigi is done on-the-fly to premixed events before overlay.



# Reconstruction



- The reconstruction sequence turns the binary output (RAW) from CMS/DIGI into physically interpretable quantities ready for data analysis
- Hits in the detector are aggregated in cluster and tracks, which in turn are matched to create particle candidates (Particle Flow): Tracks, muons, electrons, photon, jet, ...

## ■ Data tiers include

- ▶ RECO
- ▶ AOD
- ▶ MINIAOD
- ▶ NANOAOD

Data Tier	Size (kB)
RAW	1000
GEN	< 50
SIM	1000
DIGI	3000
RECO(SIM)	3000
AOD(SIM)	400 (8x reduction)
MINIAOD(SIM)	50 (8x reduction)
NANOAOD(SIM)	1 (50x reduction)

Analysis data formats

# CMSSW

- CMSSW: one release to rule them all
  - ☑ GENERator, SIMulation, HLT, RECOstruction ANALYSIS workflows...
- C++ code and configuration handled via Python
  - ☑ “git” used for code versioning and integration
- Release schedule follows a “train model”:  
dear developer: catch this train or wait for the next one
  - ☑ regular timetable of ~6 months (slightly tuned for major conferences or physics needs)
  - ☑ pre-releases are regularly produced while the release is under development

GitHub, Inc. github.com/cms-sw/cmssw/tree/CMSSW\_10\_2\_X

191,549 commits    104 branches    1,719 releases

Branch: CMSSW\_10\_2\_X    New pull request

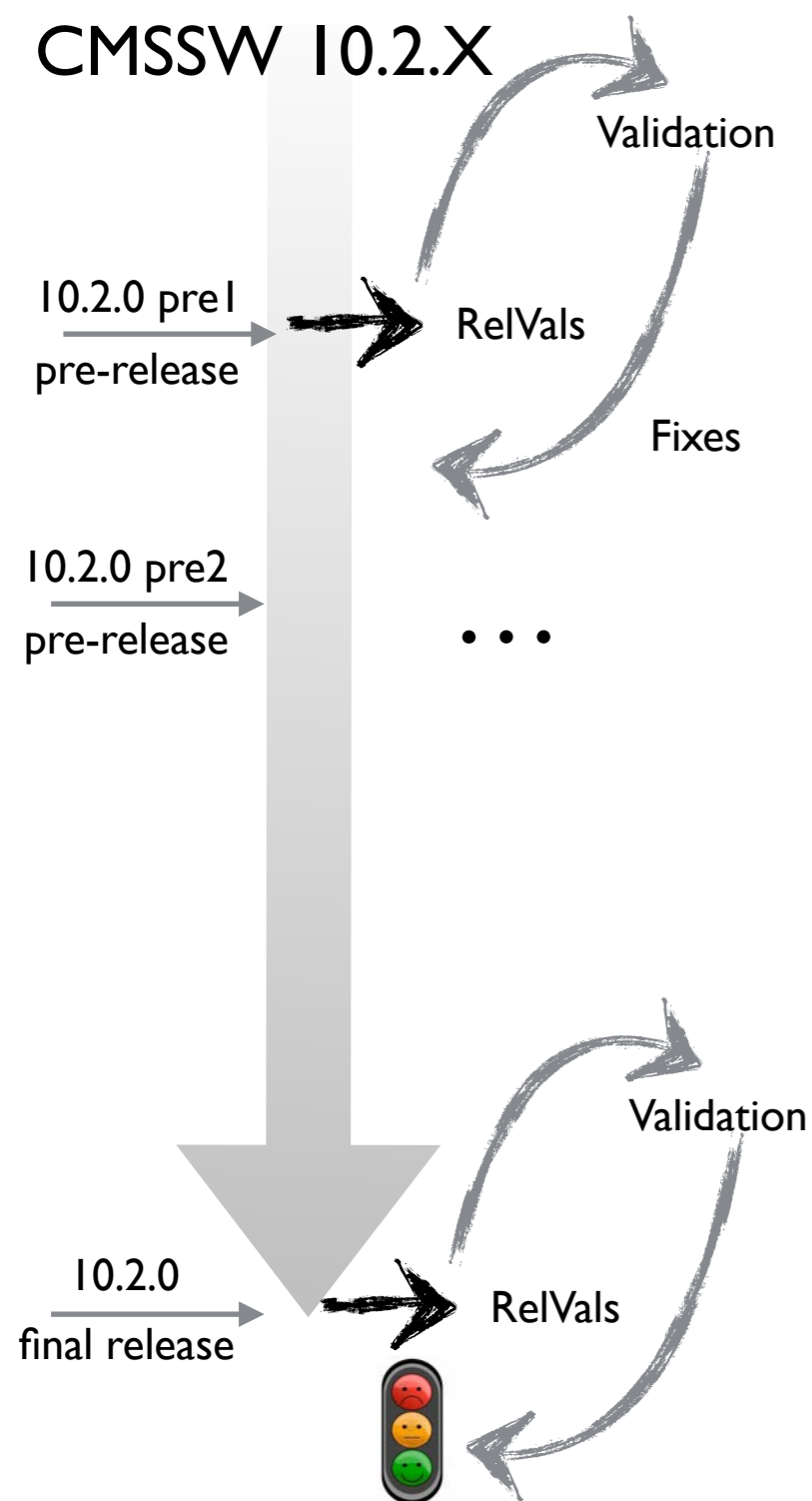
This branch is even with master.

cmsbuild Merge pull request #23590 from slava77/patch-37

Alignment	Merge pull request #23301 from usarica/HIPAlignment/...
AnalysisAlgos	clang check AnalysisAlgos
AnalysisDataFormats	AnalysisDataFormats/TopObjects: Fix bug found by cla...
BigProducts/Simulation	* Add SimG4Core/PrintGeomInfo to Big products
CalibCalorimetry	fix hdrs CalibCalorimetry/HcalPlugins
CalibFormats	Fixed calls to sprintf to avoid buffer problems.
CalibMuon	python 2to3 tool lib2to3.fixes.fix_basestring
CalibTracker	Merge pull request #23454 from davidlange6/python2t...
Calibration	Tune vertex error scale factor for BS fit on ZeroBias
CaloOnlineTools	fix hdr CaloOnlineTools/EcalTools
CommonTools	Merge pull request #23414 from ahinzmann/linScalePu...
CondCore	Merge pull request #23517 from depasse/PayloadInspe...
CondFormats	Merge pull request #23589 from ggovi/lhcinfo-o2o-1-1
CondTools	Merge pull request #23589 from ggovi/lhcinfo-o2o-1-1
Configuration	Merge pull request #23363 from angirar/Phase1-Fastsi...

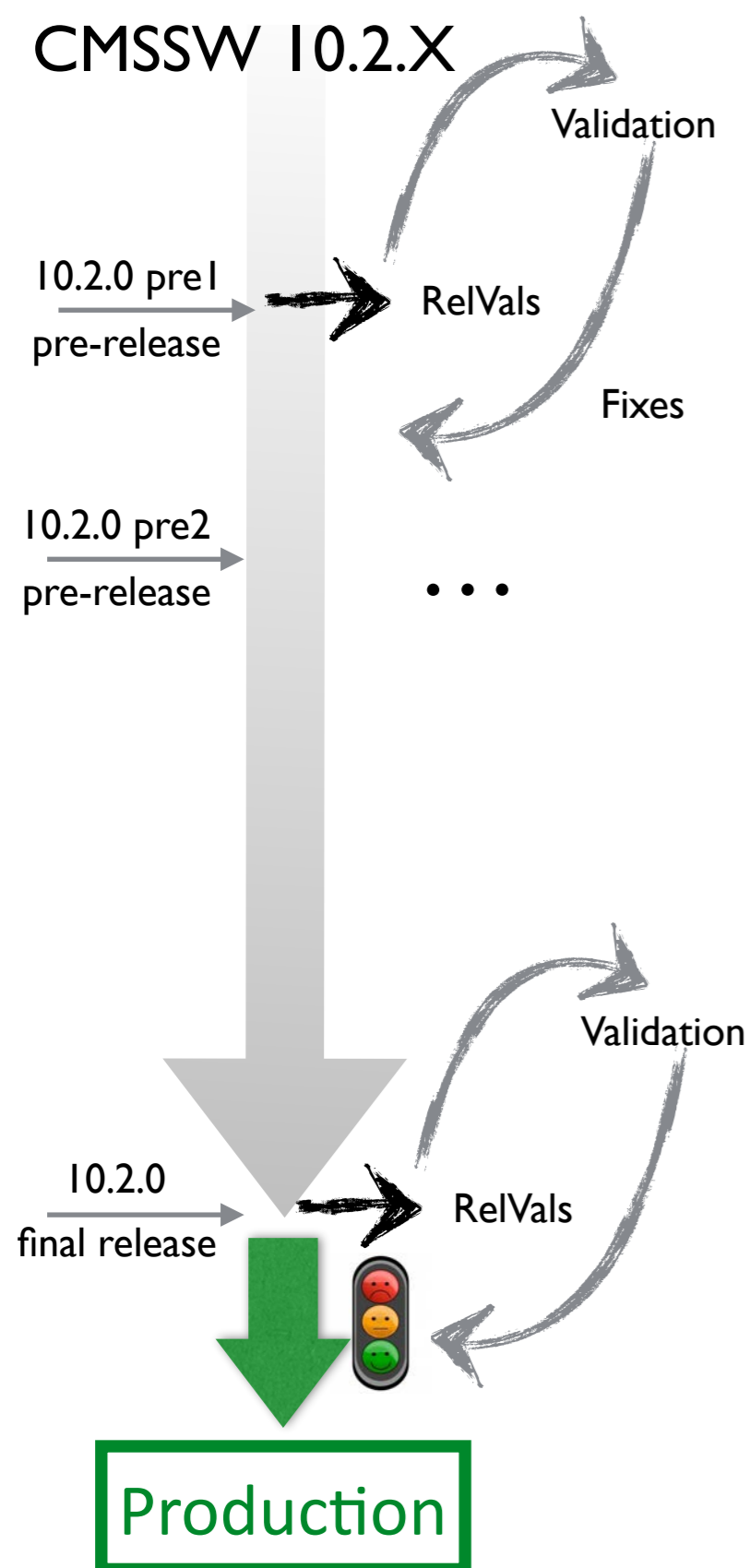
Start from  
<http://cms-sw.github.io/>

# CMSSW release validation



- Release integration bound to Quality Assurance Tests
  - Data Quality Monitoring (DQM)
  - ☑ unit tests & regression tests
  - ☑ small scale production tests: Release Validation Test (RelVal) producing DQM plots
- Validation: iterative process performed all along release cycle (from pre-releases to final version)
  - ☑ DPG, POG and PAG validation experts check the plots
  - ☑ PPD/PdmV group coordinates validation campaigns
  - ☑ sign-off on quality of release and calibrations
- differential validation compares plots of each release w.r.t reference (ex. last validated pre-release)
  - ☑ each cycle takes 4-5 days to have the samples + 1 week for feedback from the validators

# CMSSW release validation & production preparation



- Once a major release (X.Y.Z) is green-lighted  
→ start preparation of the campaign  
(re-reco or MC production)
- ☑ finalization of the alignment and calibration conditions (and their validation)
- ☑ finalization of the parameters for the Pile-Up overlay (PU scenario)
- ☑ preparation of the injection machinery for the central processing by computing

# CMSSW & Production

■ Feature planning:

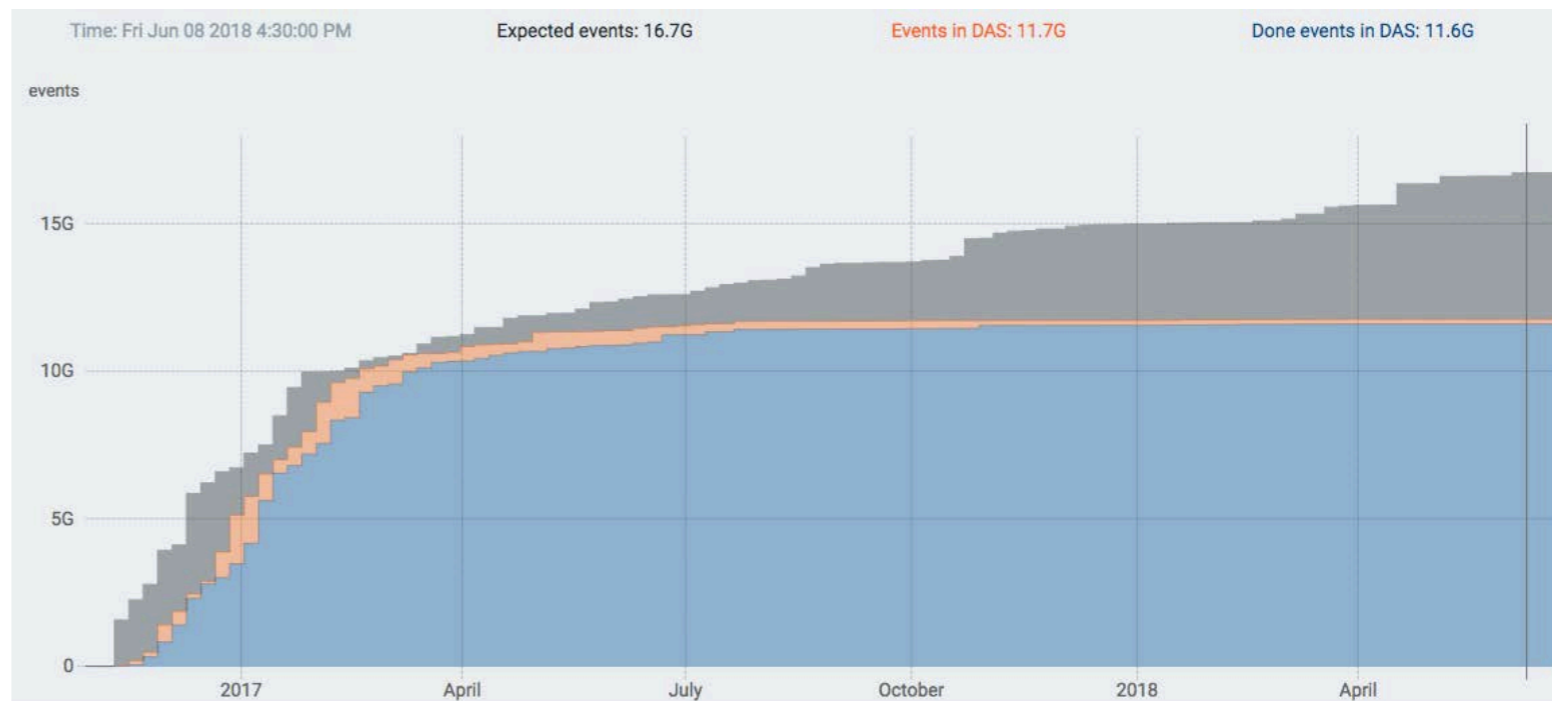
☑ production releases: driven by physics/machine constraints & goals

☑ e.g.  
RunII Summer16 DR80 Premix production is producing MC compatible to 2016 data

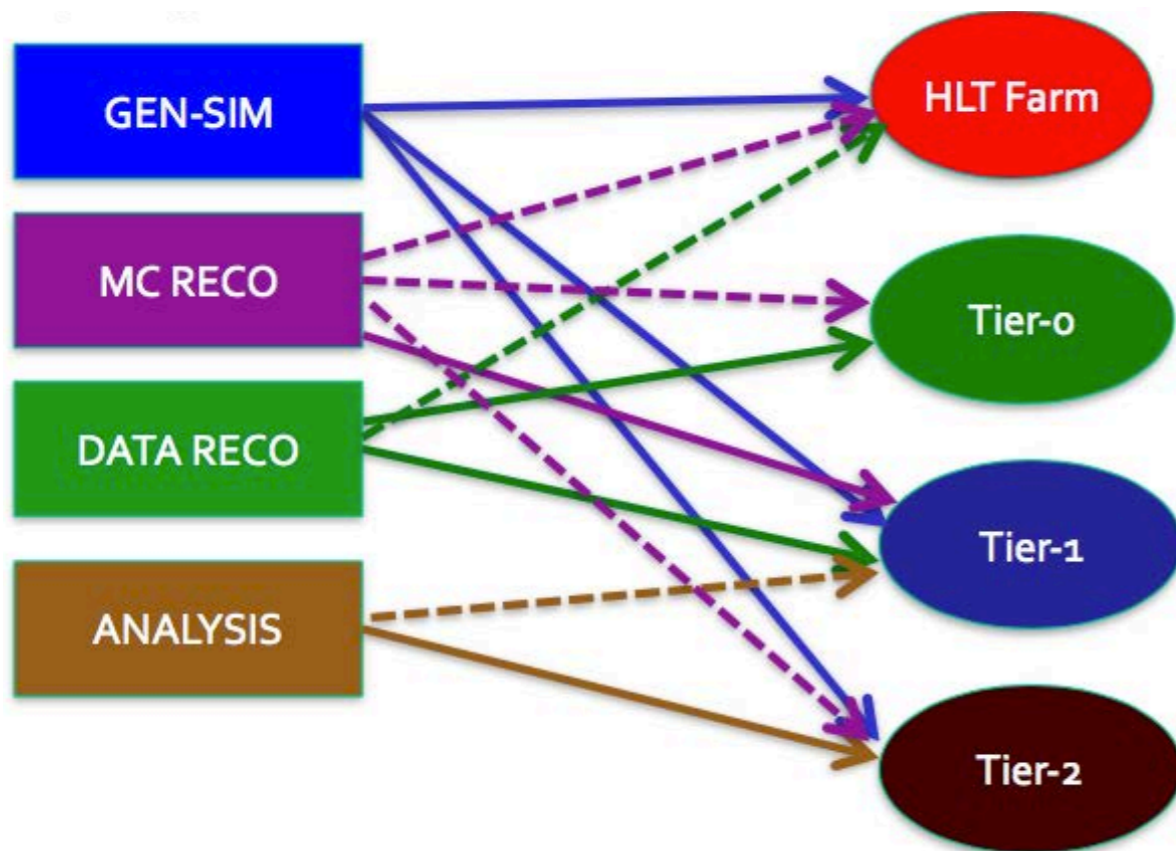
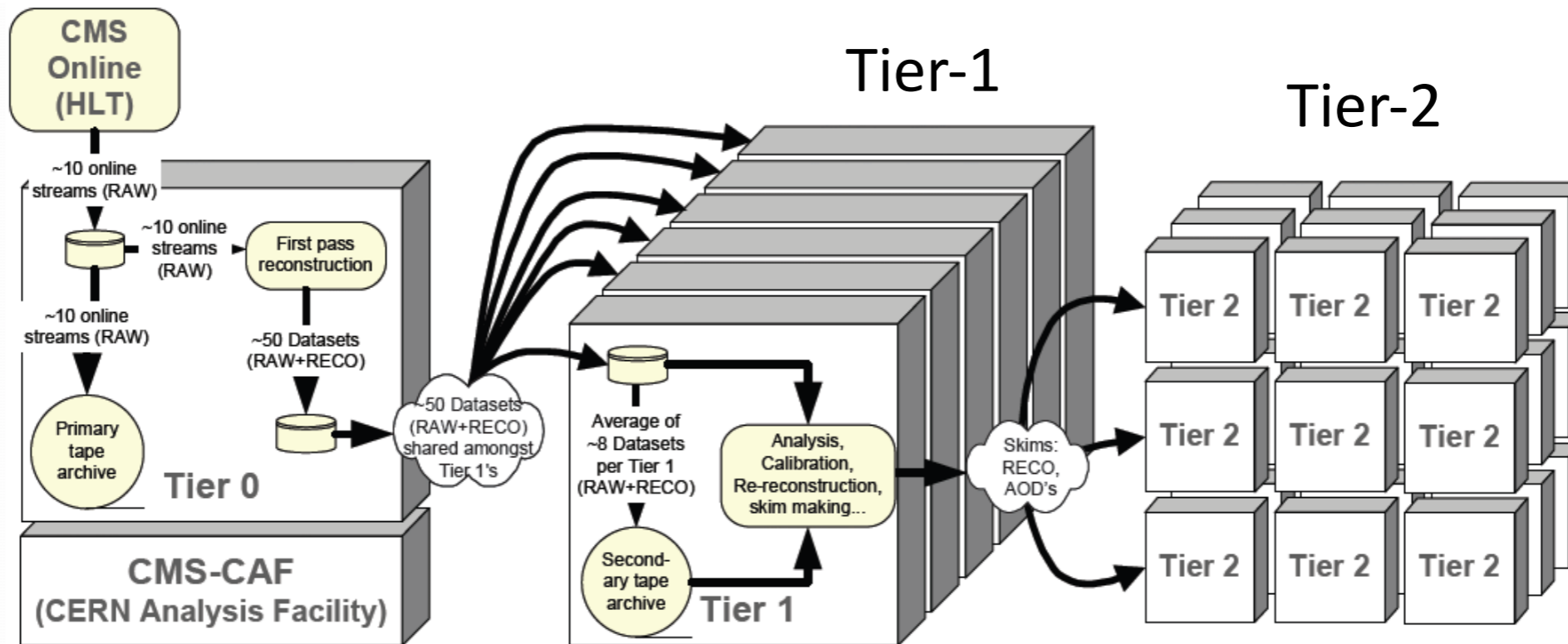
23Sep2016 re-reco data: 4.9G events



RunII Summer16 DR80 Premix: 11.7G events



# Distributed computing



■ Increasing the flexibility for facilities and workflows

☑ More places that jobs can run

— Run-1    - - - Run-2

# Search, access, transfer datasets

**Data Aggregation Service (DAS)** is the place where you can look for samples.

■ <https://cmsweb.cern.ch/das/>

☑ 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.

■ Dataset name structure is mentioned in the backup.

■ PhEDEx service will help to transfer datasets between sites (backup).

The screenshot shows the Data Aggregation System (DAS) interface. At the top, there is a navigation bar with the CMS logo and links for Home, Services, Keys, Bug report, Status, CLI, FAQ, and Help. Below the navigation bar, there is a search bar with the following options: results format: list, 50 results/page, dbs instance: prod/global, and autocompletion: disable. There are Search and Reset buttons. The search results show two records for the dataset /ZeroBias/Run2018A-v1/RAW. The first record shows the dataset name, number of files (15949), number of blocks (91), number of events (116036305), and dataset size (66.1TB). The second record shows the dataset name, type (data), status (VALID), creation time (2018-04-26 21:28:25), physics group (NoGroup), and dataset size (66.1TB). Both records have a 'show' button next to the 'dbs3' source.



# Analysis: Any Data, Anytime, Anywhere (AAA)

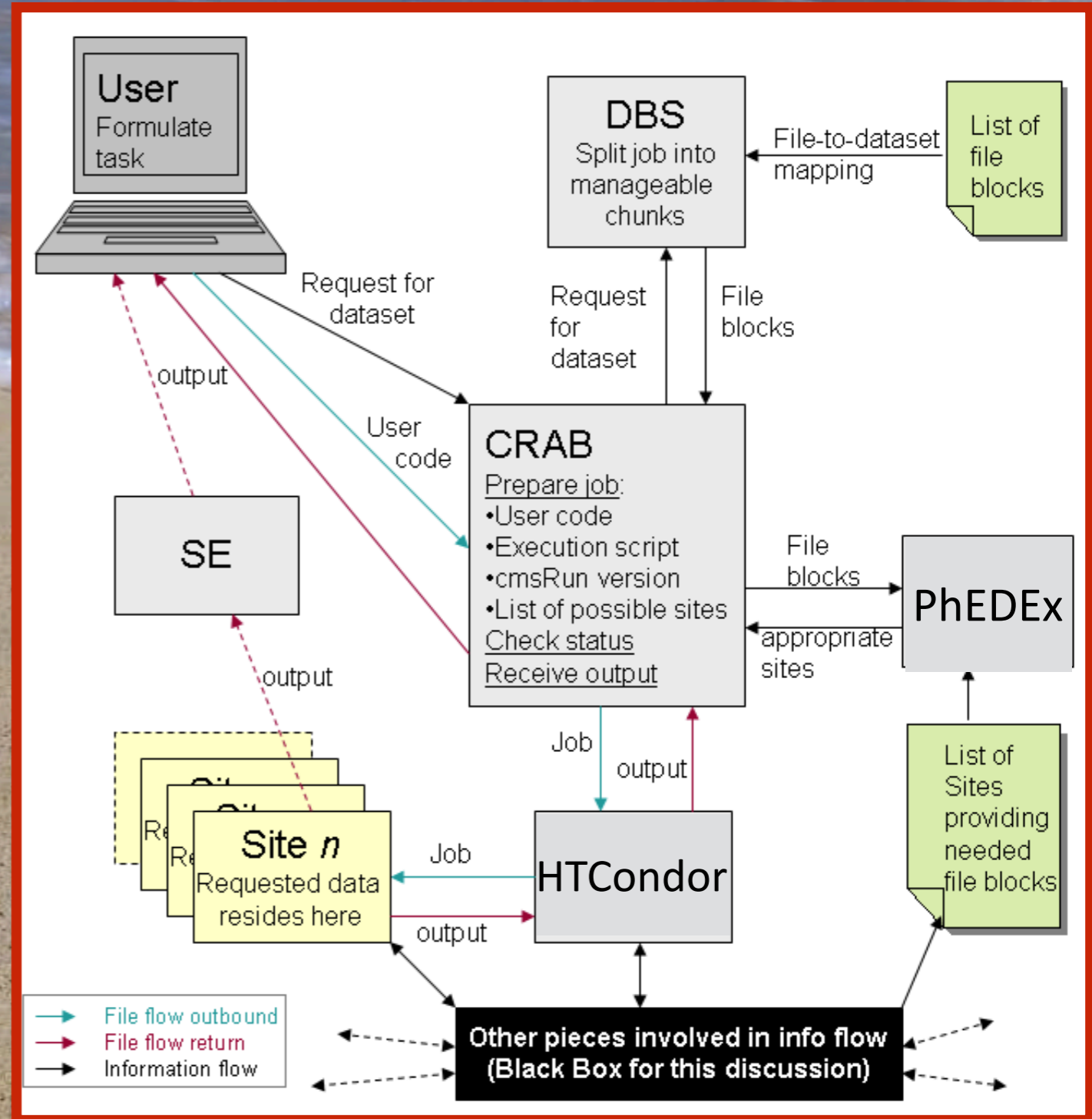
CMS's implementation of a generic xrootd service for analyzing CMS data located at any grid site with bare ROOT or the CMSSW/FWLite environment, without downloading it to your local storage. You are able to analyze data without knowing whether the input file is on your computer or halfway around the world! AAA also allows for greater resilience against damaged or missing input files, and for greater use of opportunistic resources.

If you would like to run analysis jobs in grid environment/resources. → ○

# CRAB

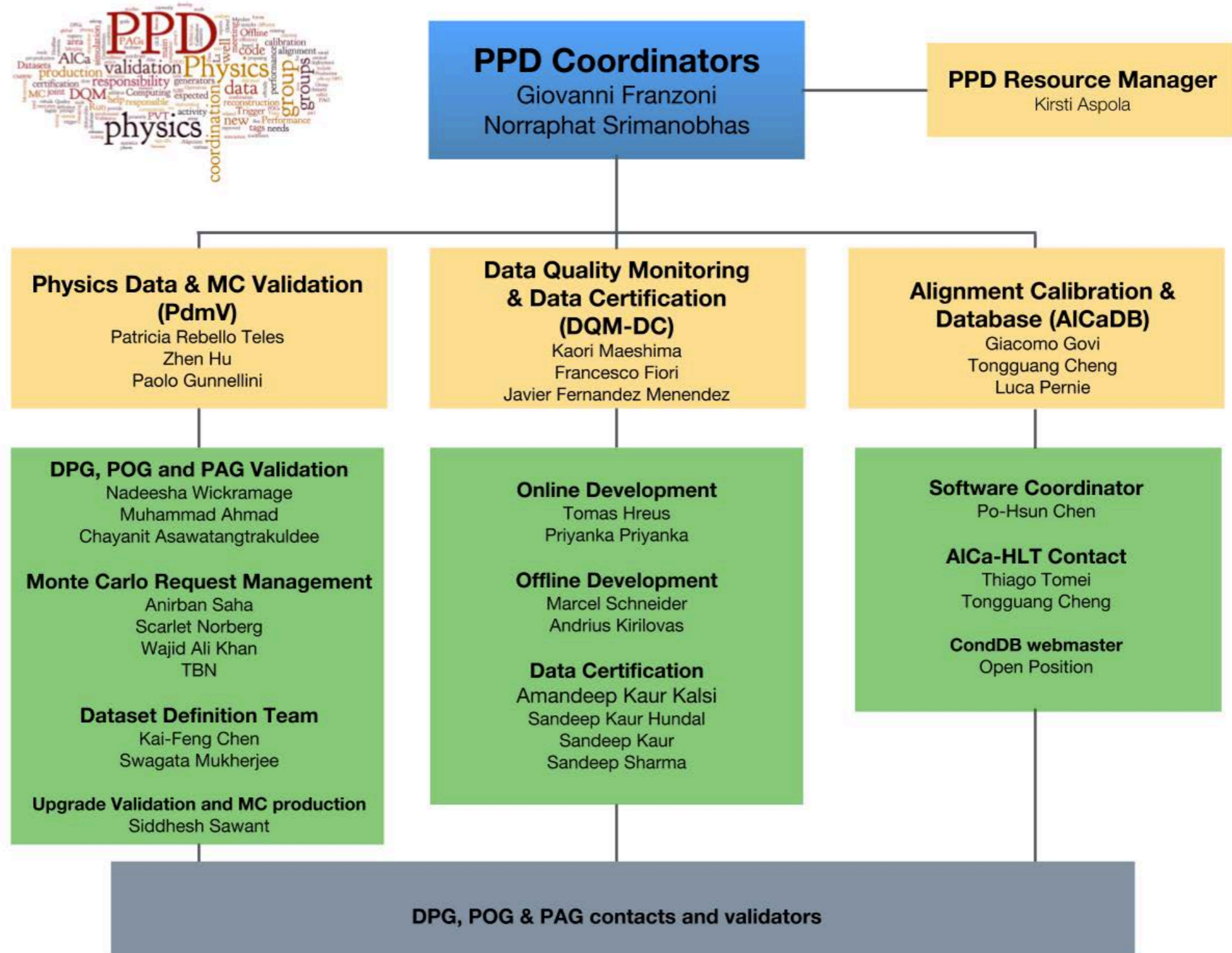
**CRAB** is a Python program intended to simplify the process of creation and submission of CMS analysis jobs into a grid environment.

If datasets you want to use are not on disk everywhere, they will be transfer from tape to disk for you automatically.



# Physics Performance & Datasets (PPD) Organization

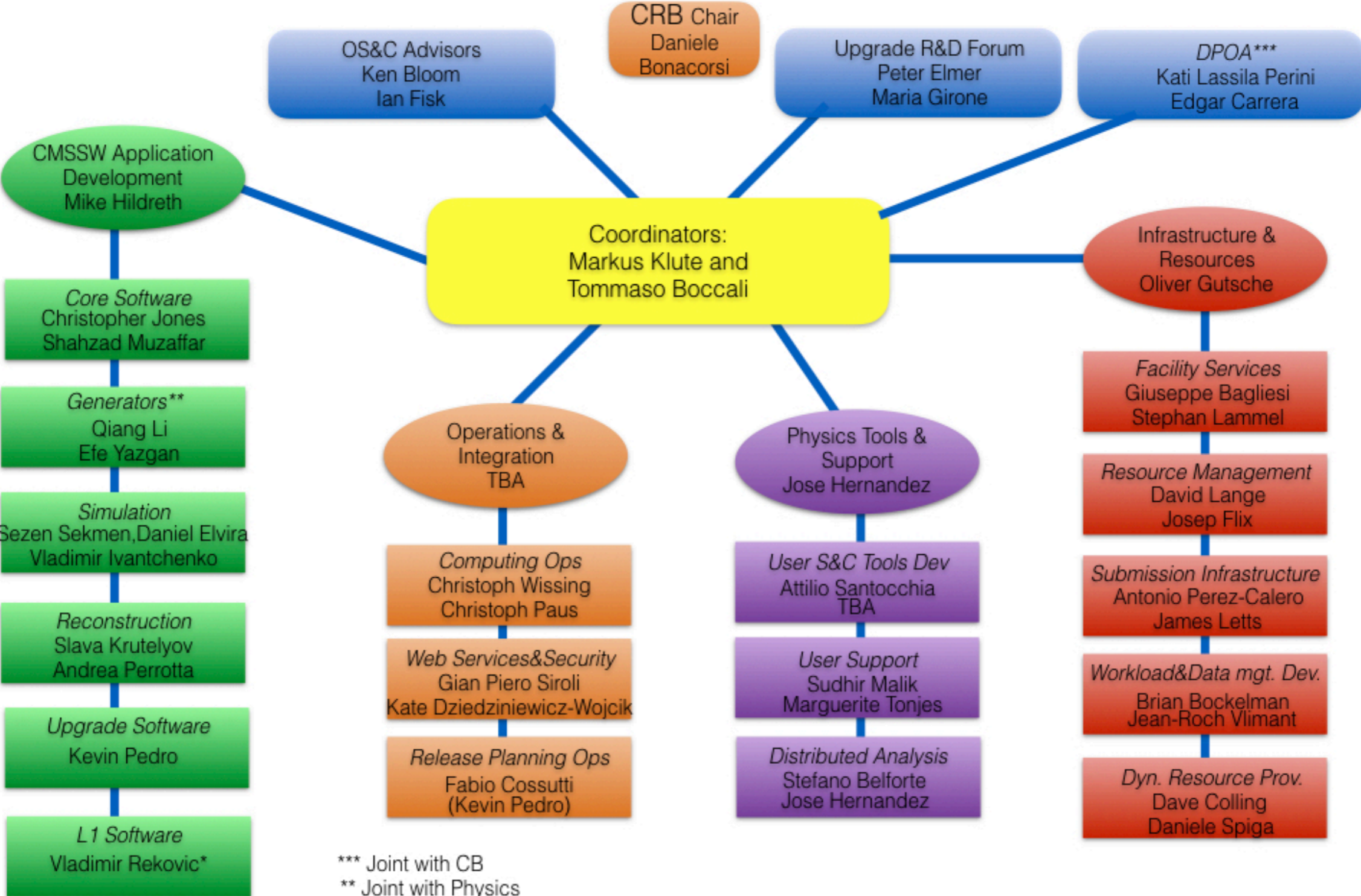
## Physics Performance & Datasets (PPD) organisation 2018



<https://twiki.cern.ch/twiki/bin/view/CMS/PhysicsPerformanceDatasetHome>

# Offline & Computing (O&C) Organization

## 2018 Offline Software and Computing



\*\*\* Joint with CB  
 \*\* Joint with Physics  
 \* Joint with L1 DPG

<https://twiki.cern.ch/twiki/bin/view/CMS/DrupalComputing>

# Outline: Exercises

<https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideCMSDataAnalysisSchoolPisa2019PPDExercise>

In this set of exercises, we have

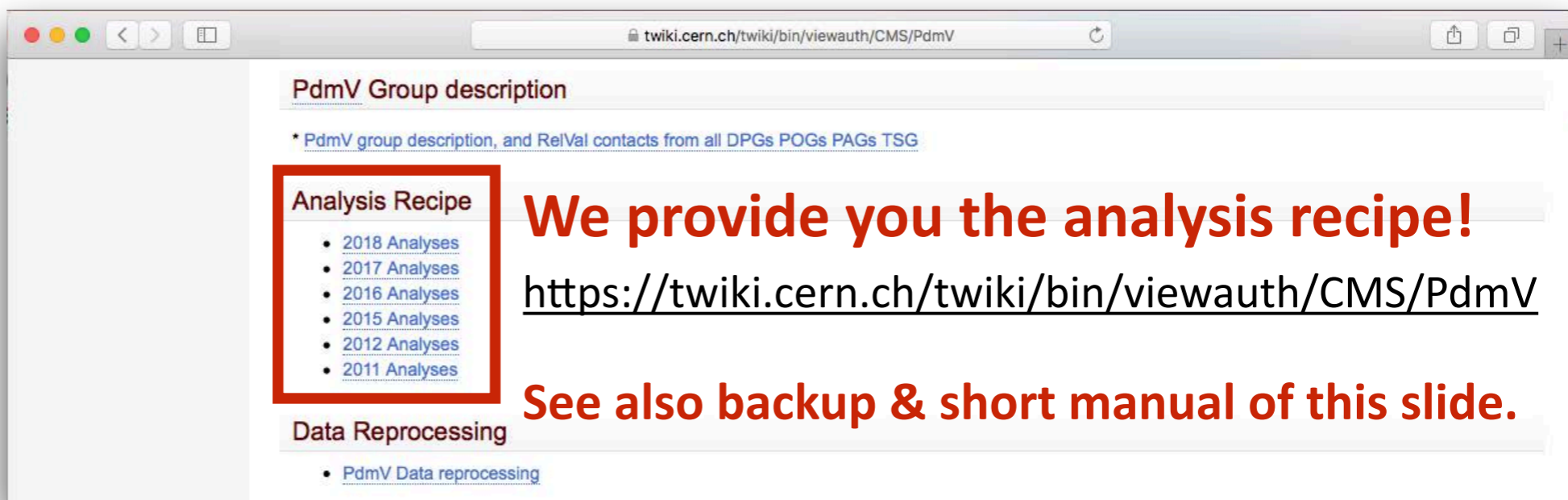
- Exercise 0: Starting your RunII analysis
- Exercise 1: Exploring dataset using DAS : searching 2018 EGamma dataset as an example
- Exercise 2: Explore information for a Monte Carlo miniAODSIM sample from DAS, McM and pMp
- Exercise 3: Miscellaneous details about datasets
- Exercise 4: Generating MC NanoAOD events from scratch
- Exercise 5: Explore GlobalTag through cmsDBbrowser
- Exercise 6: Compute the integrated luminosity collected by CMS in RunII

# Exercise 0: Starting your RunII analysis

# When you start your analysis

Several questions will come to you:

- Which datasets you are using?
- Do you need Monte-Carlo samples for your analysis?
  - ☑ Are they produced centrally, all of them?
  - ☑ Are they currently producing? How can you monitor them.
  - ☑ How to produce it privately?  
(For learning, not suggestion for private production)
- How many fb<sup>-1</sup> of data you are using?
  - ☑ Are you sure that you use all available data? How?



twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV

PdmV Group description

\* PdmV group description, and RelVal contacts from all DPGs POGs PAGs TSG

**Analysis Recipe**

- [2018 Analyses](#)
- [2017 Analyses](#)
- [2016 Analyses](#)
- [2015 Analyses](#)
- [2012 Analyses](#)
- [2011 Analyses](#)

Data Reprocessing

- [PdmV Data reprocessing](#)

**We provide you the analysis recipe!**  
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV>

**See also backup & short manual of this slide.**

# A table for RunII analysis



## PPD RunII Analysis Recipe

Updated: 2018-12-06A

<https://twiki.cern.ch/twiki/bin/view/CMS/PdmV>

	2016	2017	2018
<b>CMSSW</b>	CMSSW_8_0_X (AOD) CMSSW_9_4_X (Mini,Nano)	CMSSW_9_4_X	CMSSW_10_2_X
<b>Data</b>	<b>AOD:</b> <a href="#">07Aug17*</a> <b>MiniAOD:</b> <a href="#">17Jul2018*</a> <b>NanoAOD:</b> <a href="#">22Aug2018*</a>	<b>AOD:</b> <a href="#">17Nov2017*</a> <b>MiniAOD/NanoAOD:</b> <a href="#">31Mar2018*</a>	<b>AOD/MiniAOD:</b> <a href="#">17Sep2018</a> (ABC), <a href="#">Prompt</a> (D) <b>NanoAOD:</b> <a href="#">14Sep2018</a> (ABCD)
<b>Int. lumi. (Golden)</b>	35.92 /fb	41.53 /fb	59.97 /fb
<b>GT for data Analysis</b>	80X_dataRun2_2016SeptRepro_v7 94X_dataRun2_v10	94X_dataRun2_v11	102X_dataRun2_Sep2018Rereco_v1 102X_dataRun2_Prompt_v11
<b>MC Digi-Reco</b>	RunIISummer16DRPremix	RunIIFall17DRPremix	RunIIAutumn18DRPremix
<b>MC MiniAOD</b>	RunIISummer16MiniAODv3	RunIIFall17MiniAODv2	RunIIAutumn18MiniAOD
<b>MC NanoAOD</b>	RunIISummer16NanoAODv3	RunIIFall17NanoAOD	RunIIAutumn18NanoAOD
<b>GT for MC Analysis</b>	80X_mcRun2_asymptotic_2016_TracheIV_v8 94X_mcRun2_asymptotic_v3	94X_mc2017_realistic_v17	102X_upgrade2018_realistic_v12

<https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVAnalysisSummaryTable>



# Exercise 1: Exploring dataset using DAS : searching 2018 EGamma dataset as an example

# Data Aggregation Service (DAS)

**Data Aggregation Service (DAS)** is the place where you can look for samples.

■ <https://cmsweb.cern.ch/das/>

☑ Certificate is needed (CERN, or mapping to CERN account)

☑ Use *dasgoclient*, see in Exercise 3

■ DAS lists datasets and their properties (requestID, sites, run # and LS #....) aggregating information from various services.

■ Dataset name structure is mentioned in the backup.

**Data Aggregation System (DAS):** [Home](#) | [Services](#) | [Keys](#) | [Bug report](#) | [Status](#) | [CLI](#) | [FAQ](#) | [Help](#)

results format:  ,  results/page, dbs instance  , autocompletion

[Show DAS keys description](#)

Showing 1—2 records out of 2.

Dataset: [/ZeroBias/Run2018A-v1/RAW](#)  
Number of files: 15949 Number of blocks: 91 Number of events: 116036305 Dataset size: 66141940194756 (66.1TB)  
[Release](#), [Blocks](#), [Files](#), [Runs](#), [Configs](#), [Parents](#), [Children](#), [Sites](#), [Physics Groups](#) [XSDB](#) Sources: [dbs3](#) [show](#)

Dataset: [/ZeroBias/Run2018A-v1/RAW](#)  
Type: data Status: VALID Creation time: 2018-04-26 21:28:25 Physics group: NoGroup  
[Release](#), [Blocks](#), [Files](#), [Runs](#), [Configs](#), [Parents](#), [Children](#), [Sites](#), [Physics Groups](#) [XSDB](#) Sources: [dbs3](#) [show](#)

Showing 1—2 records out of 2.

# Data Tier

Datasets anatomy for both Data and MC

`/<PrimaryDataset>/<Campaign>-<Processing string>-<Processing version>/<data tier>`

For example

■ Data: `/EGamma/Run2018B-17Sep2018-v1/MINIAOD`

✓ Campaign = CMS Run Period = **Run2018B**

✓ Process string = **17Sep2018**

✓ Processing version = **v1** (Exercise 3)

■ MC: `/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/MINIAODSIM`

✓ Campaign = McM production campaign = **RunIIFall17MiniAODv2**

✓ Process string

▶ **PU2017\_12Apr2018\_new\_pmx**; PU scenario + other string

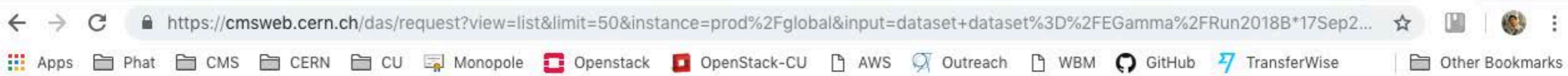
▶ **94X\_mc2017\_realistic\_v14**; Global tag


▶ Sometimes, you may see "**extX**" as sample is extension

✓ Processing version = **v1** (Exercise 3)

■ We use *dash* ("**-**") to separate between <Campaign>, <Processing string>, and <Processing version>. And *underscore* ("**\_**") to link within <Processing string>.

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



 Data Aggregation System (DAS): [Home](#) | [Services](#) | [Keys](#) | [Bug report](#) | [Status](#) | [CLI](#) | [FAQ](#) | [Help](#)

results format:  ,  results/page, dbs instance  , autocompletion

dataset dataset=/EGamma/Run2018B\*17Sep2018\*/MINIAOD

[Show DAS keys description](#)

**Fill in the query**

Showing 1 — 1 records out of 1.

[<first](#) | [prev](#) | [next](#) | [last](#)>

By default DAS shows dataset with **VALID** status. To query datasets regardless of their status please use

**Read this note, we will discuss in Ex. 3**

```
dataset status=* dataset dataset=/EGamma/Run2018B*17Sep2018*/MINIAOD
```

Dataset: [/EGamma/Run2018B-17Sep2018-v1/MINIAOD](#)

Creation time: 2018-10-03 20:24:18 Physics group: NoGroup Status: **VALID** Type: data

[Release](#), [Blocks](#), [Files](#), [Runs](#), [Configs](#), [Parents](#), [Children](#), [Sites](#), [Physics Groups](#) [XSDB](#) Sources: [dbs3](#) [show](#)

Showing 1 — 1 records out of 1.

**Then play with your search result**

[<first](#) | [prev](#) | [next](#) | [last](#)>


processing time: 3.493500548 sec

# Exercise 2: Explore information for a Monte Carlo MiniAODSIM sample from DAS, McM and pMp

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

← → ↻ [https://cmsweb.cern.ch/das/request?view=list&limit=50&instance=prod%2Fglobal&input=dataset+dataset%3D%2FTTJets\\_TuneCP5\\_13TeV-amcatn...](https://cmsweb.cern.ch/das/request?view=list&limit=50&instance=prod%2Fglobal&input=dataset+dataset%3D%2FTTJets_TuneCP5_13TeV-amcatn...) ☆ 📄 👤 ⋮

Apps 📁 Phat 📁 CMS 📁 CERN 📁 CU 📁 Monopole 📁 Openstack 📁 OpenStack-CU 📁 AWS 🗨 Outreach 📁 WBM 🔄 GitHub ⚡ TransferWise 📁 Other Bookmarks

 **Data Aggregation System (DAS):** [Home](#) | [Services](#) | [Keys](#) | [Bug report](#) | [Status](#) | [CLI](#) | [FAQ](#) | [Help](#)

results format:  ,  results/page, dbs instance  , autocompletion

dataset

[Show DAS keys description](#)



**Fill in the query**

Showing 1 — 1 records out of 1.

[<first](#) | [prev](#) | [next](#) | [last](#)>

Dataset: [/TTJets\\_TuneCP5\\_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017\\_12Apr2018\\_new\\_pmx\\_94X\\_mc2017\\_realistic\\_v14-v1/MINIAODSIM](#)

Creation time: 2018-08-25 15:33:37 Physics group: NoGroup Status: **VALID** Type: mc Dataset size: 8610252714647 (8.6TB) Number of blocks: 82 Number of events: 154280331 Number of files: 2540

[Release](#), [Blocks](#), [Files](#), [Runs](#), [Configs](#), [Parents](#), [Children](#), [Sites](#), [Physics Groups](#) [XSDB](#) Sources: [dbs3](#) [show](#)

Showing 1 — 1 records out of 1.

**Then play with your search result**

[<first](#) | [prev](#) | [next](#) | [last](#)>

processing time: 43.498092747 sec

DAS version: git=04.05.01 go=go1.10 date=2019-01-24 13:20:53.783035451 +0100 CET m=+0.006256894

<https://cms-pdmv.cern.ch/mcm/>

The screenshot shows the McM web interface with several callouts:

- 1**: Points to the 'Request' menu item in the top navigation bar.
- 2**: Points to the 'Navigation' tab in the 'Select View' section.
- 3**: Points to the 'prepid' input field in the form, which contains the value 'TOP-RunIIFall17MiniAO'.
- 4**: Points to the 'Search' button above the 'Prepid' input field.

The interface includes a top navigation bar with items like Campaign, Chained campaign, Flow, Priority change, Request, Chained request, Batch, Invalidations, and Mccm. Below this is a secondary bar with Dashboard, User, Settings, Lists, Admin, Help, Support, and Home. The main content area has tabs for Select View, List from file, Navigation, and Output Dataset. A form with various input fields is present, and a table at the bottom shows a list of requests with columns for Prepid, Actions, Approval, Status, and Dataset name.

**Get test command**

Q 2.6: Get the McM test command sequence with the full **digi-reco** configuration.

Navigation: Select View | List from file | Navigation | Output Dataset

status:  member\_of\_chain:  prepid: TOP-RunIIFall17MiniAO

extension:  tags:  energy:


mcdb\_id:  flown\_with:  pwg:

process\_string:  generators:  member\_of\_campaign:

approval:  dataset\_name:

Search:

Prepid	Actions	Approval	Status	Dataset name
TOP-RunIIFall17MiniAODv2-00169		✓ submit	✓ done	TTJets_TuneCP5_13TeV-ar

**I, view chains**

URL: [https://cms-pdmv.cern.ch/mcm/chained\\_requests?contains=TOP-RunIIFall17MiniAODv2-00169&page=0&shown=15](https://cms-pdmv.cern.ch/mcm/chained_requests?contains=TOP-RunIIFall17MiniAODv2-00169&page=0&shown=15)

Navigation: Invalidations | Mccm

User: srimanob (administrator)

Prepid	Approval	Chain
TOP-RunIIFall17wmLHEGS-00010	✓ submit	TOP-RunIIFall17wmLHEGS-00010 → TOP-RunIIFall17DRPremix-00109
TOP-RunIIFall17wmLHEGS-00010	flow	TOP-RunIIFall17wmLHEGS-00010 → TOP-RunIIFall17DRPremix-00109

**Click on DR request, then get test command**

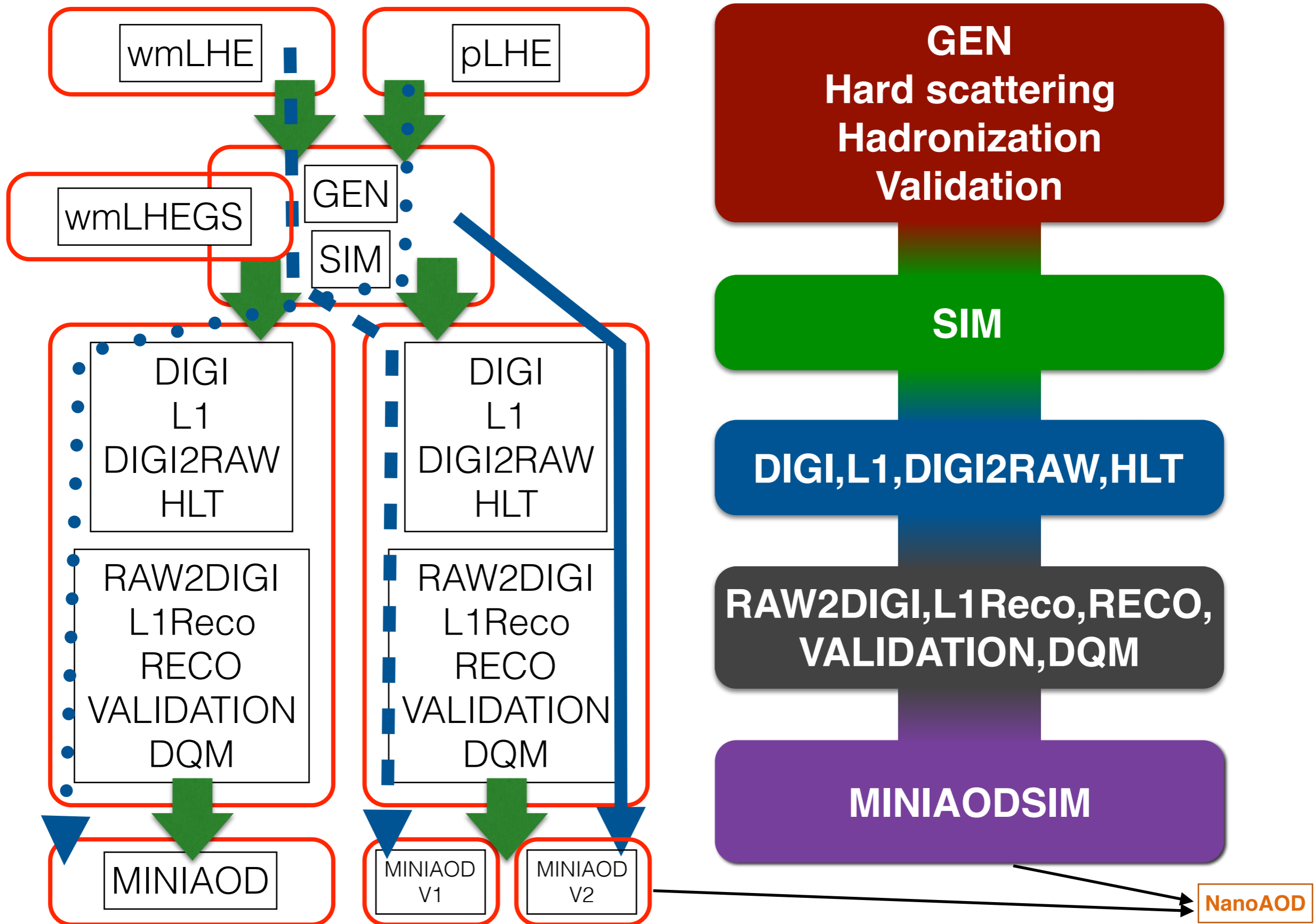




# Exercise 3: Miscellaneous details about datasets

# Exercise 4: Generating MC NanoAOD events from scratch

# How do we organise MC production in CMS?



Browser address bar: <https://cms-pdmv.cern.ch/mcm/requests?prepid=PPD-RunIIFall18GS-00011&page=0&shown=311385131135>

Navigation menu: Campaign, Chained campaign, Flow, Priority change, Request, Chained request, Batch, Invalidations, Mccm

Secondary menu: Dashboard, User, Settings, Lists, Admin, Help, pMp, Support, Home

User: srimanob (administrator)

Select View List from file Navigation Output Dataset

Prepid	Actions	Approval	Status	Dataset name
PPD-RunIIFall18GS-00011		none	new	TTbar_13TeV_TuneCP5_CMSDASPISA19

Get test command

View chains

```

#!/bin/bash
source /cvmfs/cms.cern.ch/cmsset_default.sh
export SCRAM_ARCH=slc6_amd64_gcc700
if [ -r CMSSW_10_2_7/src ] ; then
  echo release CMSSW_10_2_7 already exists
else
  scram p CMSSW CMSSW_10_2_7
fi
cd CMSSW_10_2_7/src
eval `scram runtime -sh`

curl -s --insecure https://cms-pdmv.cern.ch/mcm/public/restapi/requests/get_fragment/PPD-RunIIFall18GS-00011-fragment.py | \
  Configuration/GenProduction/python/PPD-RunIIFall18GS-00011-fragment.py ] || exit $?;

if grep -q "gridpacks" Configuration/GenProduction/python/PPD-RunIIFall18GS-00011-fragment.py; then
  if ! grep -q "/cvmfs/cms.cern.ch/phys_generator/gridpacks" Configuration/GenProduction/python/PPD-RunIIFall18GS-00011-fragment.py; then
    echo "Gridpack inside fragment is not in cvmfs."
    exit -1
  fi
fi

scram b
cd ../../
cmsDriver.py Configuration/GenProduction/python/PPD-RunIIFall18GS-00011-fragment.py --fileout file:PPD-RunIIFall18GS-00011-fragment.py --datatier GEN-SIM --conditions 102X_upgrade2018_realistic_v11 --beamspot Realistic25ns13TeVEarly2018Collision --era Run2_2018 --python_filename PPD-RunIIFall18GS-00011_1_cfg.py --no_exec --customise Configuration/GenProduction/python/PPD-RunIIFall18GS-00011_1_cfg.py || exit $?
  
```

Browser address bar: [https://cms-pdmv.cern.ch/mcm/chained\\_requests?contains=PPD-RunIIFall18GS-00011&page=0&shown=15](https://cms-pdmv.cern.ch/mcm/chained_requests?contains=PPD-RunIIFall18GS-00011&page=0&shown=15)

Chain view: PPD-RunIIFall18GS-00011 → PPD-RunIIFall18GS-00011 → PPD-RunIIFall18GS-00011 → PPD-RunIIFall18GS-00011

## RunIIFall18GS-00011

# Exercise 5: Explore GlobalTag through cmsDBbrowser

# Exercise 6: Compute the integrated luminosity collected by CMS in RunII

# BACKUP & SHORT MANUAL



# Important Twiki links for Analysers

- Information about Golden JSONs and recent primary datasets with conditions :  
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV2017Analysis>
  - Previous year twikis also exists with a change of year in url
- Global tags for both data and MC for all the scenarios :  
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideFrontierConditions>
- Documentation for BrilCalc : script to know luminosity which are considering for your analysis :  
<https://cms-service-lumi.web.cern.ch/cms-service-lumi/brilwsdoc.html>
- Information about the pile-up calculation for data :  
[https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#Pileup\\_JSON\\_Files\\_For\\_Run\\_II](https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#Pileup_JSON_Files_For_Run_II)
- A plethora of information about data recorded by CMS is available here :  
<https://cmswbm.cern.ch/> (just enter the run number and press “enter”)
- To know whats inside miniAOD and what are the latest recipes :  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkbookMiniAOD>
- And there are many more from each POG for the latest recipes.





# Important Twiki links for Analysers

- Egamma Recipes for Run-II (id working points, scale factors and other recommendations) :  
<https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2>
- Muon POG main twiki : <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonPOG>
- B-Tag recommendations :  
<https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco>
- JetMET main twiki: <https://twiki.cern.ch/twiki/bin/view/CMS/JetMET>



# References

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



# Accessing Datasets

- Datasets for analysis → they need to be on disk @ Tier2/3 (GRID = “where” doesn’t matter)
  - **AnalysisOps** usually subscribes the most common samples to “central” space → no action needed
  - what if you need something more “exotic”? (are you sure you do need it??)
- You can ask the transfer @ T2s using the Phedex service
  - **users** can request using the interface (complete datasets or “blocks” of files) → requests are associated to “groups” and assigned to a given “site”
  - **data manager** approves/rejects the request (usually evaluating the available quota at the destination site) → “standard” analysis use cases addressed by AnalysisOps group, each PAG/POG group has a person taking care of data-management
  - **Dynamic Data Management** tools → optimal data replication and quota control
- before submitting I want to run on a single file locally → how do I do it
  - most (almost all) datasets accessible remotely via XROOTD protocol → e.g. can run @ CERN reading files @ FNAL

Any data,  
Any time,  
Any where  
(AAA)

[Contact&Doc](#)



# Finding Datasets: anatomy

How do I look for a sample → Data Aggregation Service (DAS)

- list datasets and their properties (requestID, sites, run # and LS #....) aggregating information from various services

Anatomy of the dataset name:

- dataset = /PrimaryDataset/ProcessingVersion/DataTier

Examples:

- data (prompt reco): /SingleElectron/Run2015D-PromptReco-v3/AOD
- data (re-reco): /SingleElectron/Run2015D-16Dec2015-v1/MINIAOD
- MC (RunIIFall15DR76):  
/WprimeToMuNu\_M-1600\_TuneCUETP8M1\_13TeV-pythia8/  
RunIISpring15MiniAODv2-74X\_mcRun2\_asymptotic\_v2-v1/MINIAODSIM



# Finding Datasets: anatomy

How do I look for a sample → Data Aggregation Service (DAS)

- list datasets and their properties (requestID, sites, run # and LS #....) aggregating information from various services

Anatomy of the dataset name:

- dataset = /PrimaryDataset/ProcessingVersion/DataTier

Examples:

- data (prompt reco): /SingleElectron/Run2015D-PromptReco-v3/AOD
- data (re-reco): /SingleElectron/Run2015D-16Dec2015-v1/MINIAOD
- MC (RunIIFall15DR76):  
/WprimeToMuNu\_M-1600\_TuneCUETP8M1\_13TeV-pythia8/  
RunIISpring15MiniAODv2-74X\_mcRun2\_asymptotic\_v2-v1/MINIAODSIM

Event Topology (data)

and physics process simulated (MC)

is indicated in the first segment of the dataset name

[Contact&Doc](#)



# Finding Datasets: anatomy

How do I look for a sample → Data Aggregation Service (DAS)

- list datasets and their properties (requestID, sites, run # and LS #....) aggregating information from various services

Anatomy of the dataset name:

- dataset = /PrimaryDataset/ProcessingVersion/DataTier

Examples:

- data (prompt reco): /SingleElectron/Run2015D-PromptReco-v3/AOD
- data (re-reco): /SingleElectron/Run2015D-16Dec2015-v1/MINIAOD
- MC (RunIIFall15DR76):  
/WprimeToMuNu\_M-1600\_TuneCUETP8M1\_13TeV-pythia8/  
RunIISpring15MiniAODv2-74X\_mcRun2\_asymptotic\_v2-v1/MINIAODSIM

Acquisition Era + PromptReco/reprocessing (data)

Production campaign (MC)

Alignment and Calibration i.e. Global tag (MC)

Dataset Version

[Contact&Doc](#)



# Finding Datasets: anatomy

How do I look for a sample → Data Aggregation Service (DAS)

- list datasets and their properties (requestID, sites, run # and LS #....) aggregating information from various services

Anatomy of the dataset name:

- dataset = /PrimaryDataset/ProcessingVersion/DataTier

Examples:

- data (prompt reco): /SingleElectron/Run2015D-PromptReco-v3/AOD
- data (re-reco): /SingleElectron/Run2015D-16Dec2015-v1/MINIAOD
- MC (RunIIFall15DR76):  
/WprimeToMuNu\_M-1600\_TuneCUETP8M1\_13TeV-pythia8/  
RunIISpring15MiniAODv2-74X\_mcRun2\_asymptotic\_v2-v1/MINIAODSIM

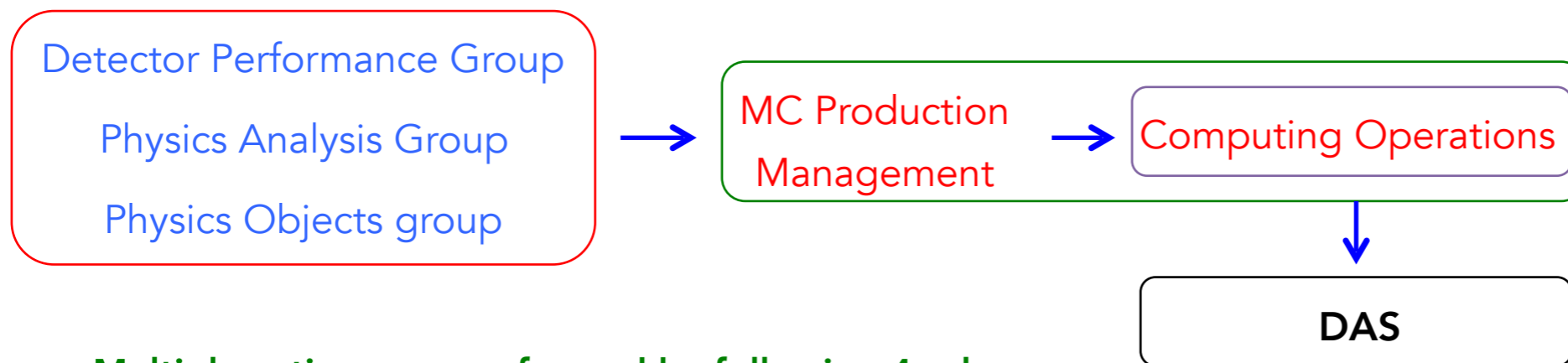
Data tier indicates the collections available at each event

[Contact&Doc](#)



# MC production management

- About 20 groups in CMS: physics analyses (PAGs), detector performance (DPGs), physics object studies (POGs)
- 100s of physics analyses; 1000s of MC samples needed; Billions of events required for ongoing Run-II of LHC
  - CMS has produced over 20 Billion simulated events in 1 year production in 2016-17
- Strong and efficient production infrastructure required: bookkeeping and interface to computing resources



Multiple actions are performed by following 4 roles

**Generator Contact:** collects the needs of simulated samples from detector or physics groups, produces generator configuration and presents them to the production team for execution

**Generator Convener:** scrutinizes and approves event generator configurations

**Request Manager:** configures campaigns and flows, performs requests' chaining, sets their priority and submits requests to the production infrastructure

} McM

**Production Manager:** handle workflows during production and send datasets to DAS

← Computing tools

**CMS Analysts:** performs analyses, and communicates needs with production team





# Data Aggregation System (DAS)

- All completed samples will be available in DAS
- Analysts can find various type of data samples and can transfer files to laptop/Tier-2/3
- General queries about DAS, command line tool, etc. are available at [DAS-FAQs](#)
- Find datasets in DAS: <https://cmsweb.cern.ch/das/>
- Wildcards (\*) can be used but try to be as specific as possible as DAS may show many combinations

results format:  results/page:  dbs instance:  autocompletion:

dataset=/TTJets\_TuneCUETP8M1\_13TeV-madgraphMLM-pythia8/RunIISummer16DR80Premix-PUMoriond17\_80X\_mcRun2\_asymptotic\_2016\_TrancheIV\_v6-v1/AODSIM

[Show DAS keys description](#)

Showing 1—1 records out of 1.

Add filter/aggregator function to the query:

Dataset: [/TTJets\\_TuneCUETP8M1\\_13TeV-madgraphMLM-pythia8/RunIISummer16DR80Premix-PUMoriond17\\_80X\\_mcRun2\\_asymptotic\\_2016\\_TrancheIV\\_v6-v1/AODSIM](#)  
 Creation time: 2017-01-27 04:37:53, Dataset size: 3.7TB, Number of blocks: 55, Number of events: 10199051, Number of files: 1047, Physics group: NoGroup, Status: **VALID**, Type: mc  
[Release](#), [Blocks](#), [Files](#), [Runs](#), [Configs](#), [Parents](#), [Children](#), [Sites](#), [Physics Groups](#), [py](#), [Subscribe to PhEDEx](#), [XsecDB Sources](#): [dbs3](#) [show](#)

↑  
CMSSW

↑  
GEN-SIM

↑  
miniAOD

↑  
To see McM prep-Id

```
DAS service: dbs3 DAS api: datasets
{
  "dataset": {
    "status": "VALID",
    "modified_by": "mcremone",
    "physics_group_name": "NoGroup",
    "acquisition_era_name": "RunIISummer16DR80Premix",
    "prep_id": "TOP-RunIISummer16DR80Premix-00118",
    "processing_version": 1,
    "creation_time": "2017-01-27 04:37:53",
    "created_by": "wagent@vocms0308.cern.ch",
    "processed_ds_name": "RunIISummer16DR80Premix-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1",
    "modification_time": "2017-02-12 12:00:06",
    "datatype": "mc",
    "xtcrosssection": None,
    "dataset_id": 13331380,
    "primary_dataset": {
      "name": "TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8"
    },
    "data_tier_name": "AODSIM",
    "name": "/TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16DR80Premix-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/AODSIM"
  }
}
```



# DAS

- Status **VALID**: production is finished and dataset is available
- Status **PRODUCTION**: Its statistics is still growing due to running production jobs and dataset is not yet announced. But analysts can still run over the existing statistics by using crab parameter "allowNonValidInputDataset" parameter in [CRAB3 configuration parameters](#)

Dataset: /TT JpsiFilter TuneCUETP8M1 mtop166 5 14TeV-powheg-tauola-pythia8/PhaseITDRSpring17MiniAOD-noPU PhaseITDRSpring17 91X upgrade2023 realistic v3-v1/MINIAODSIM  
 Creation time: 2017-06-12 14:24:31, Physics group: NoGroup, Status **PRODUCTION**, Type: mc  
 Release, Blocks, Files, Runs, Configs, Parents, Children, Sites, Physics Groups, py, Subscribe to PhEDEx, XsecDB Sources: **db3** show

Collection of ≥1 files

Global tag and CMSSW release

You can transfer full dataset or only a Block to your Tier-2/Tier3:  
 Subscribe to PhEDEx. But please remember that:

- **DISK SPACE IS VERY CRITICAL**
- **Very high threshold on private transfer requests**

Showing 1 – 3 records out of 3.

Add filter/aggregator function to the query:

Site: <b>T1 ES PIC Disk</b> Block completion: 100%, Block presence: 100%, Dataset presence: 100%, File-replica presence: 100%, Site type: Disk, StorageElement: srmcms.pic.es Datasets, SiteDB Sources: <b>phedex combined</b> show	
Site: <b>T1 US FNAL Buffer</b> Block completion: 100%, Block presence: 100%, Dataset presence: 100%, File-replica presence: 100%, Site type: <b>TAPE no user access</b> , StorageElement: cmsdcatape01.fnal.gov Datasets, SiteDB Sources: <b>phedex combined</b> show	<p>Site type: <b>TAPE no user access</b>, StorageElement: cmsdcatape01.fnal.gov</p>
Site: <b>T1 US FNAL MSS</b> Block completion: 100%, Block presence: 100%, Dataset presence: 100%, File-replica presence: 100%, Site type: <b>TAPE no user access</b> , StorageElement: cmsdcatape01.fnal.gov Datasets, SiteDB Sources: <b>phedex combined</b> show	



# PhEDEx Transfer of Datasets

Subscribe data

**Subscribe to PhEDEx**

Data Items [?]

/TTJets\_TuneCUETP8M1\_13TeV-madgraphMLM-pythia8/RunIISummer16DR80Premix-PUMoriond17\_80X\_mcRun2\_asymptotic\_2016\_TracheIV\_v6-v1/AODSIM

**Replace this data item with a block address if full dataset is not added**

DBS <https://cmsweb.cern.ch/dbs/prod/global/DBSReader> [change](#)

Destination

- |  |   |   |                                      |
|--|---|---|--------------------------------------|
| <input type="checkbox"/> T3_GR_IASA_HG | <input type="checkbox"/> T3_HR_IRB                | <input type="checkbox"/> T3_HU_Debrecen | <input type="checkbox"/> T3_IN_VBU   |
| <input type="checkbox"/> T3_IR_IPM     | <input checked="" type="checkbox"/> T3_IT_Bologna | <input type="checkbox"/> T3_IT_Firenze  | <input type="checkbox"/> T3_IT_MIB   |
| <input type="checkbox"/> T3_IT_Napoli  | <input type="checkbox"/> T3_IT_Perugia            | <input type="checkbox"/> T3_IT_Trieste  | <input type="checkbox"/> T3_KR_KISTI |

Site Custodial [?]

- yes  
 no

Subscription Type [?]

- growing  
 static

Transfer Type [?]

- replica  
 move

Priority [?]

- high  
 normal  
 low

User Group [?]

higgs ▾

Data injected after [?]

YYYY-MM-DD [hh:mm:ss]

Email gurpreet.singh@cern.ch

Comments

enter any additional comments here

Reset

Preview

Accept



# Luminosity: HowTo

- Need to know the luminosity of ALL the LSs you run on and ONLY those
  - CRAB reports the LS successfully processed by your jobs in the same JSON format used for certification

certification **{JSON}**

```
{ "251027": [[1, 60], [61, 61]],  
  "251028": [[1, 18], [19, 21], [141, 141]],  
  "251143": [[1, 16], [17, 19]],  
  "251147": [[1, 24], [29, 29], [30, 187]]  
}
```

CRAB

CRAB output **{JSON}**

```
{ "251027": [[1, 60], [61, 61]],  
  "251028": [[1, 18], [19, 21], [141, 141]],  
  "251143": [[1, 16], [17, 19]],  
  "251147": [[1, 24], [29, 29], [30, 187]]  
}
```

- Lumi POG provides tools to compute luminosity starting from JSON file:
  - `brilCalc.py`: reports measurements of LHC delivered, CMS recorded luminosity for LSs in the JSON
- NOTE: using directly the certification JSON is not correct for several reasons:
  - LS that failed prompt reco or re-reco will appear in the cert. JSON
  - LS that failed in your jobs will appear in the cert. JSON

[Contact&Doc](#)