CMS Data Quality Monitoring: Offline Workflow and Physics

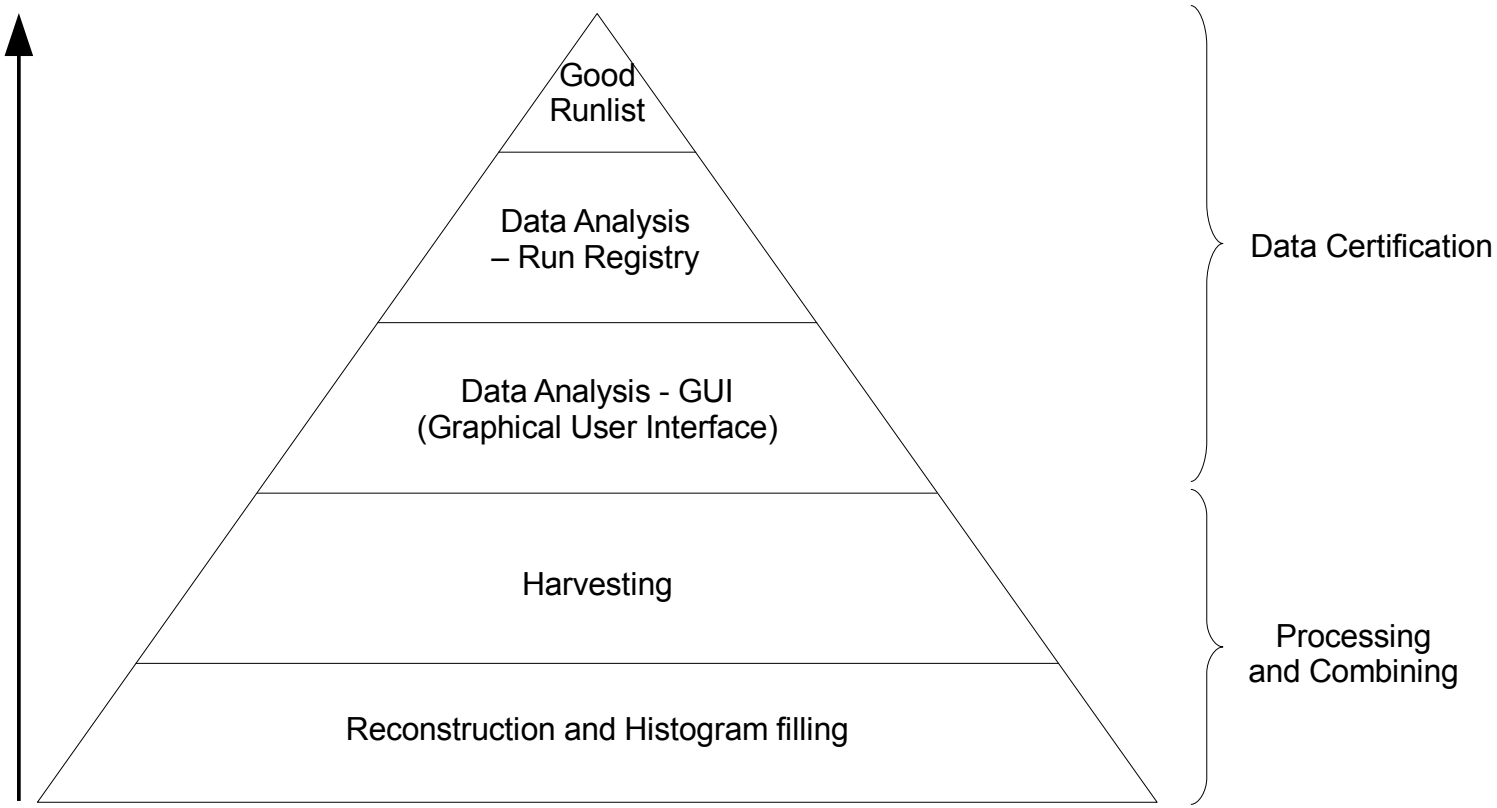
Alexander Chaushev, Sarah Lindner

September 7^{th}, 2011

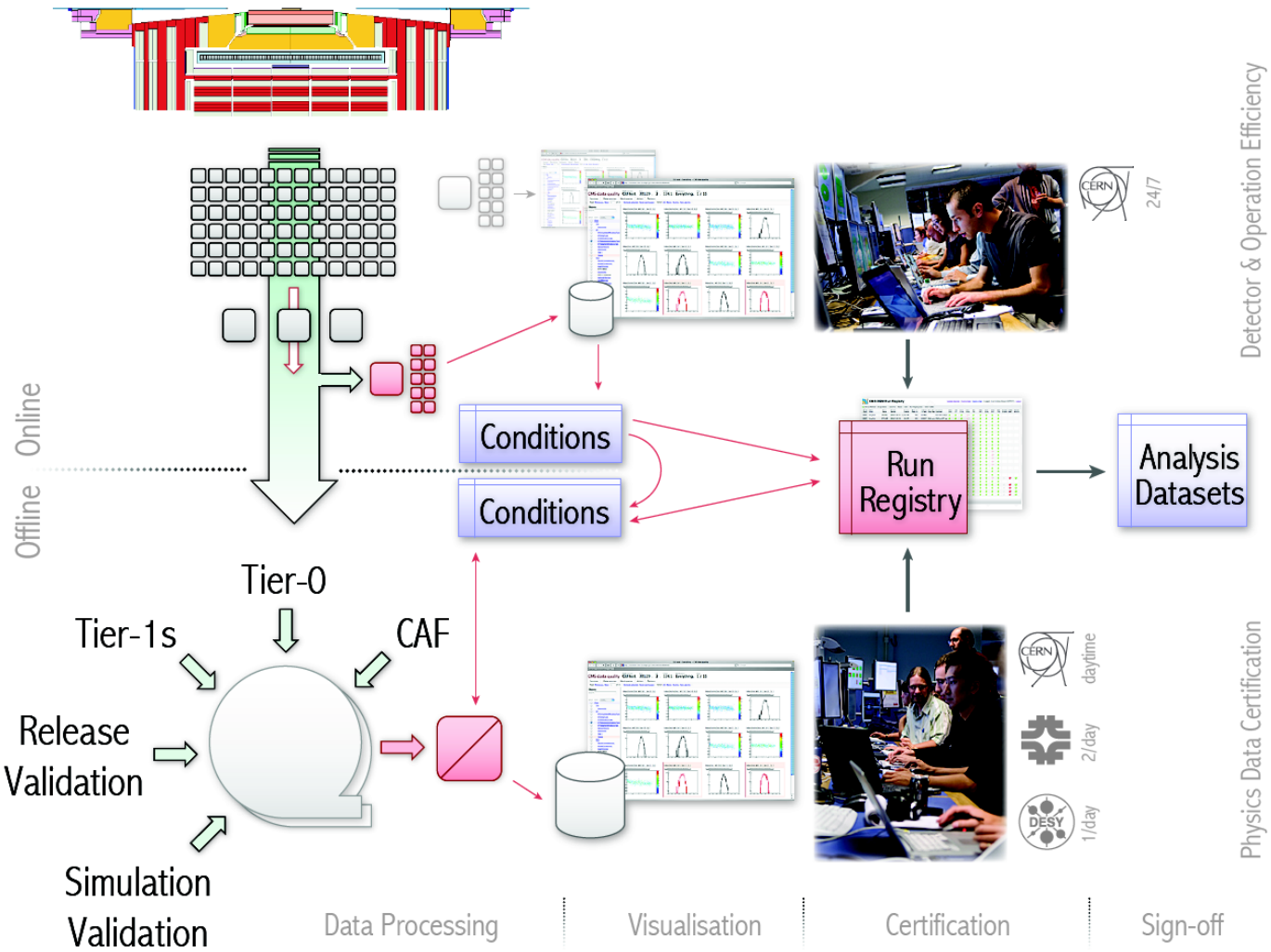
What is Data Quality Monitoring?

- To **discover problems** with data and detector and provide feedback **during data taking**
- Check for problems with physics objects once they have been **reconstructed** from detector measurements
- To **certify** that data coming from the detector is reliable and free of known systematic errors

DQM Pyramid


Good
Runlist
Data Analysis
– Run Registry
Data Analysis - GUI
(Graphical User Interface)
Harvesting
Reconstruction and Histogram filling
Data Certification
Processing
and Combining

DQM Workflow

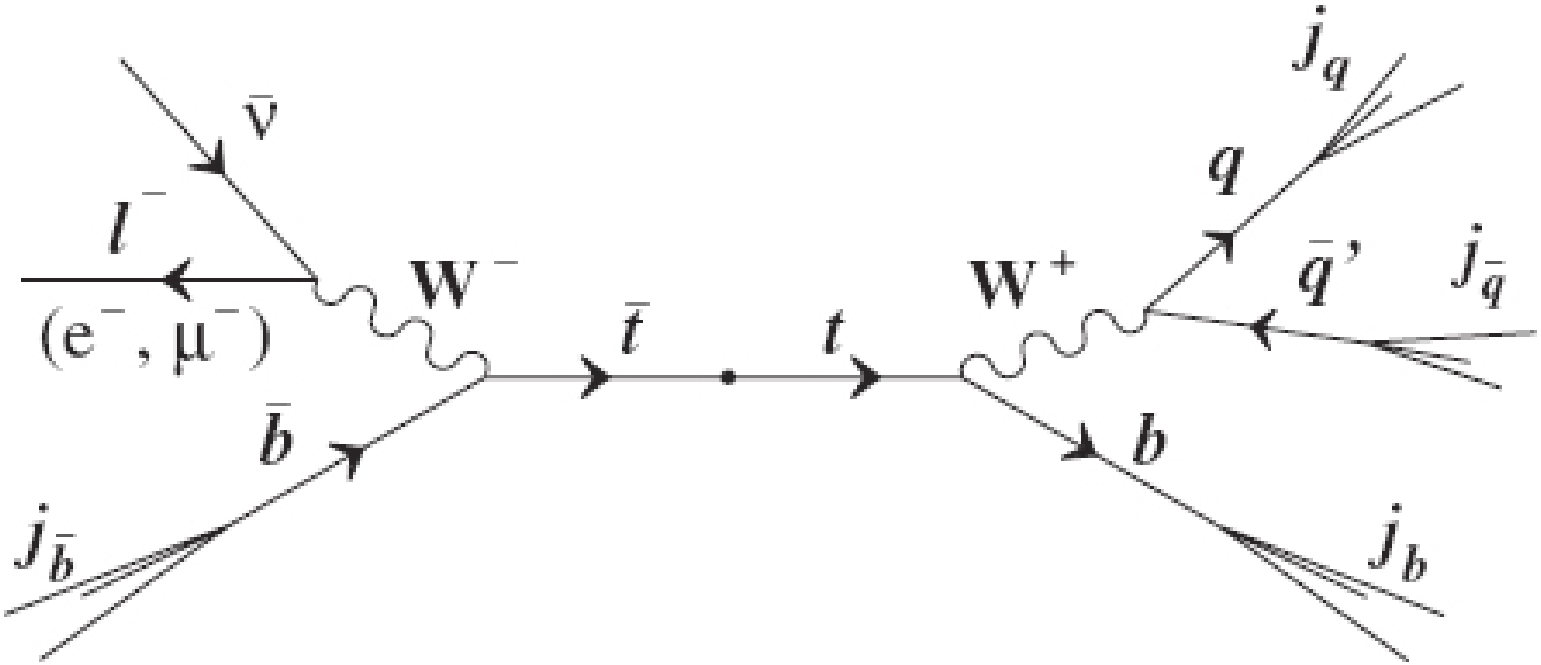


L. Tuura, A. Meyer, I. Segoni, G. Della Ricca: "CMS Data Quality Monitoring: systems and experiences"

Reconstruction and Histogram Filling

Example: top-antitop decay

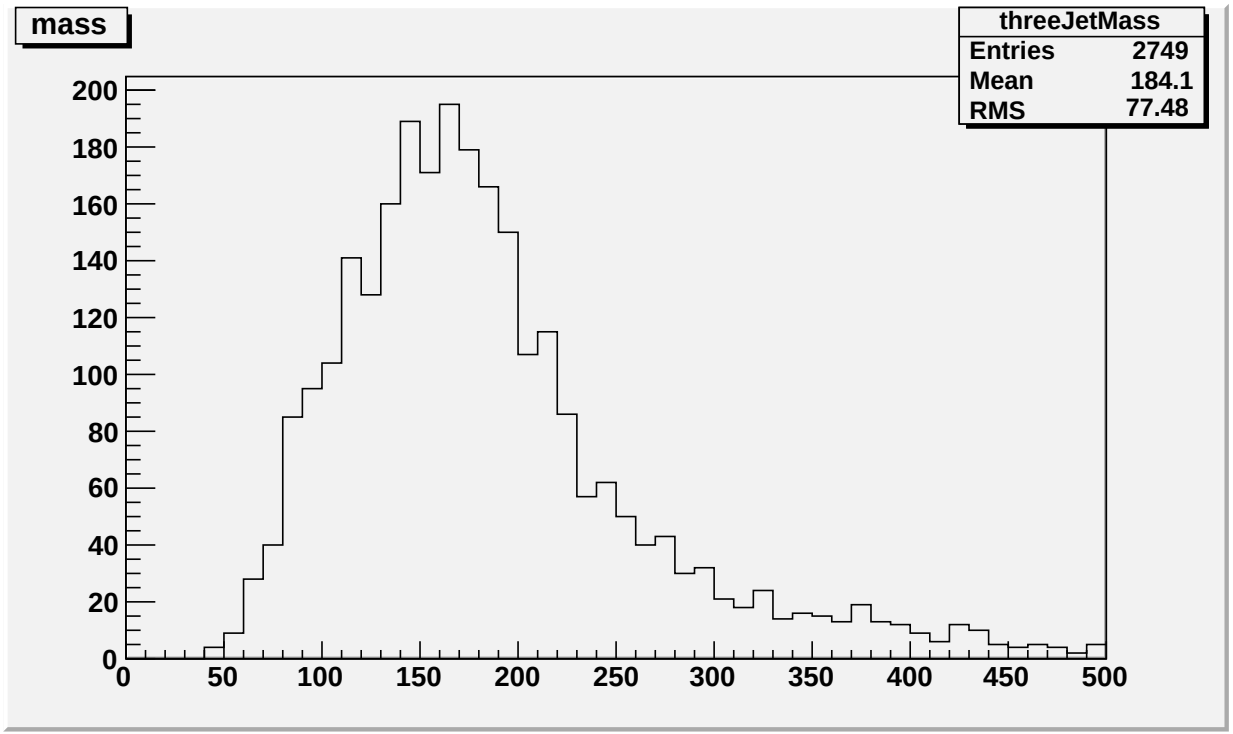
- Reconstruct top quantities
- Single muon, four jets



J. A. R. Cembranos, A. Rajaraman and F. Takayama: "Searching for CPT violation in $t\bar{t}$ production"

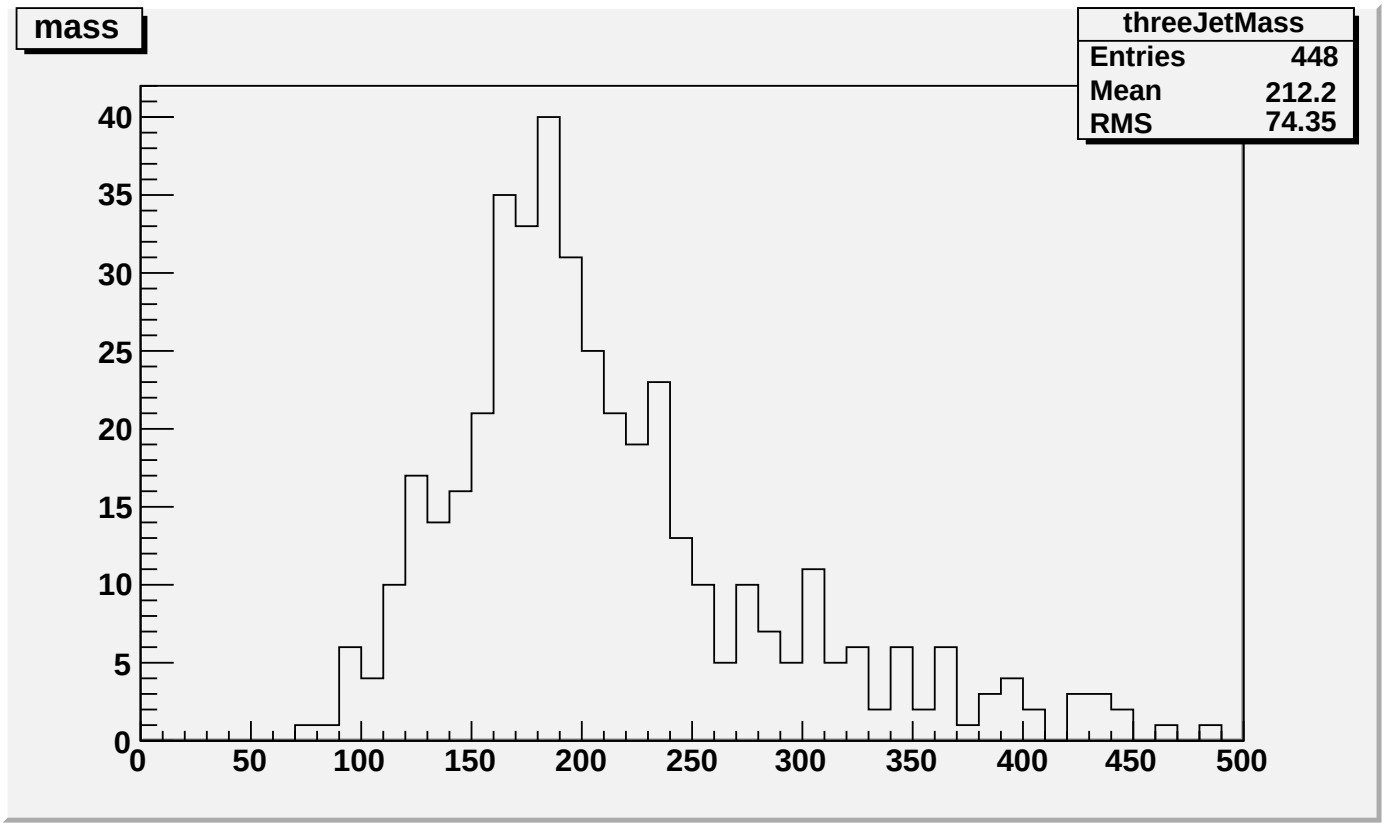
Reconstruction and Histogram Filling

- ROOT program processing Monte Carlo Data, $\approx$ 8000 events
- Back-to-back signature of top pair (leading order in transverse plane)
- Maximum vectorial sum of the transverse momentum of three jets
- Loose selection

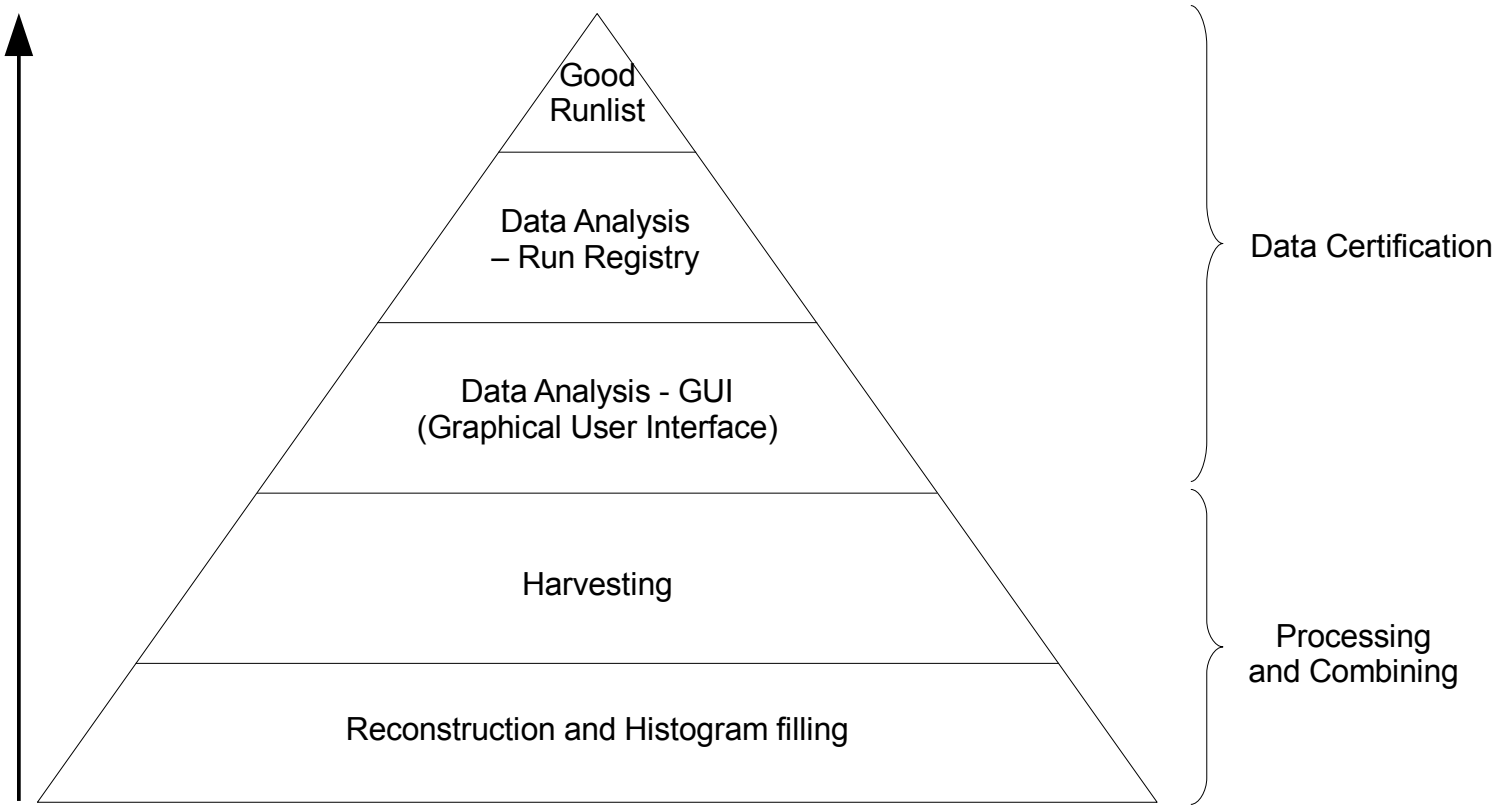


Reconstruction and Histogram Filling

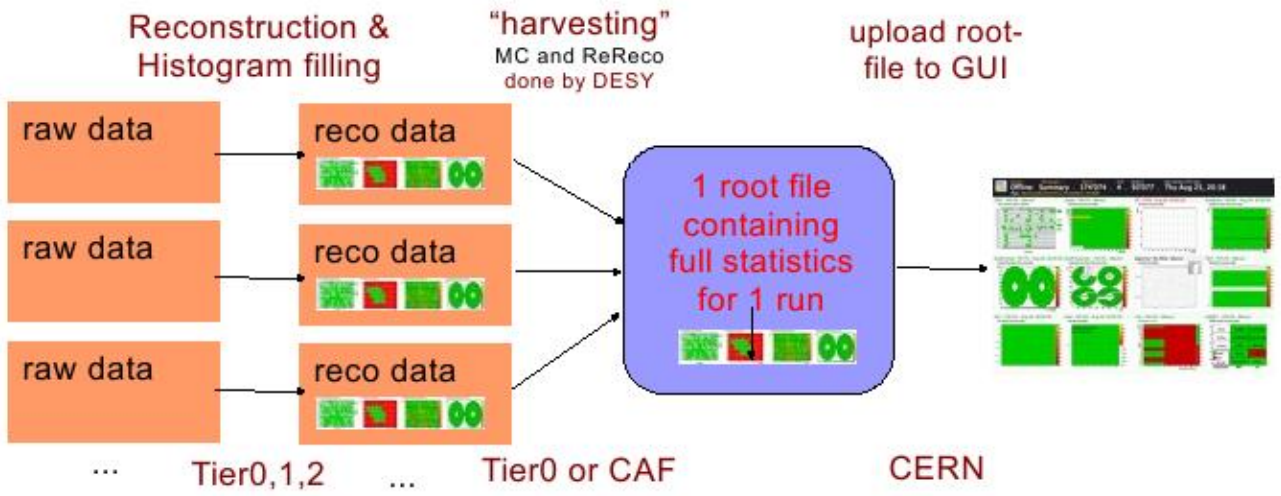
- Additional selection criteria: muon and jet cuts, remove jet candidates which really are electrons, ...



DQM Pyramid


Good
Runlist
Data Analysis
– Run Registry
Data Analysis - GUI
(Graphical User Interface)
Harvesting
Reconstruction and Histogram filling
Data Certification
Processing
and Combining

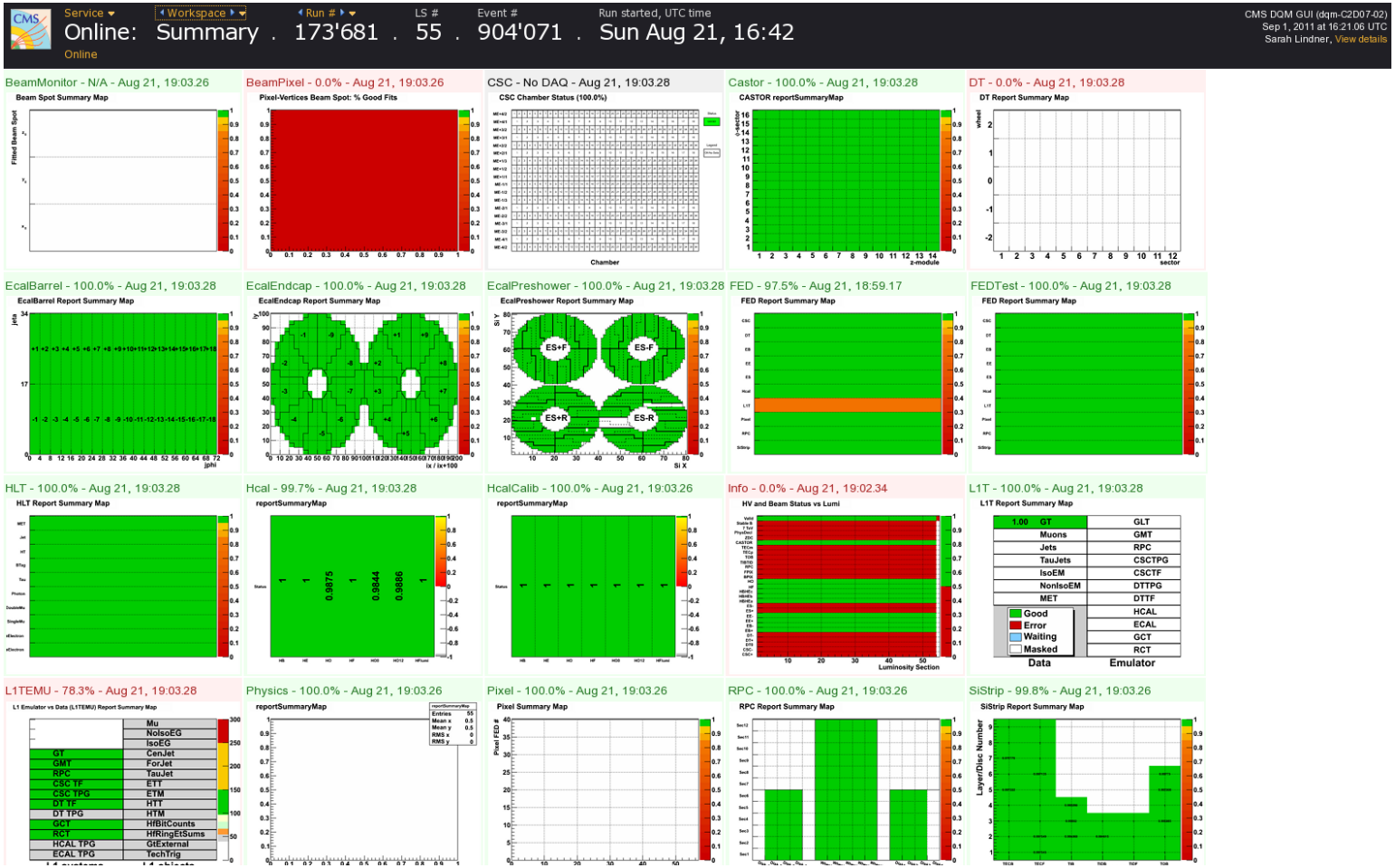
DQM Harvesting



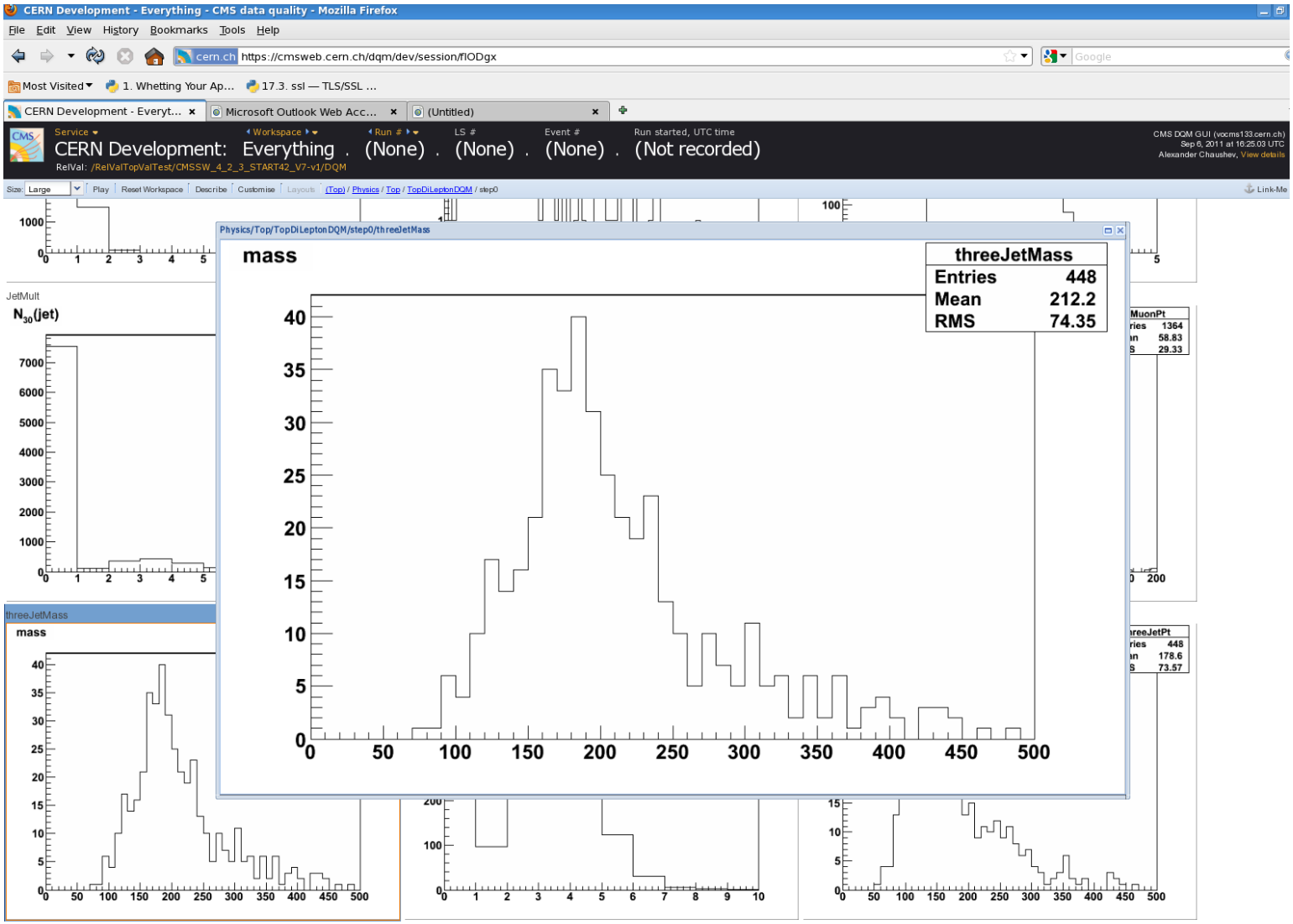
- Improved/revised existing harvester automation script
- The function of the revised script is to run the CMS Harvester, find new data and upload the resulting histograms to the DQM GUI
- At DESY data harvesting is done for Monte Carlo and Re-Reco data sets for CMS experiment

Data Certification - Graphical User Interface

Offline shifter and general user view of the system

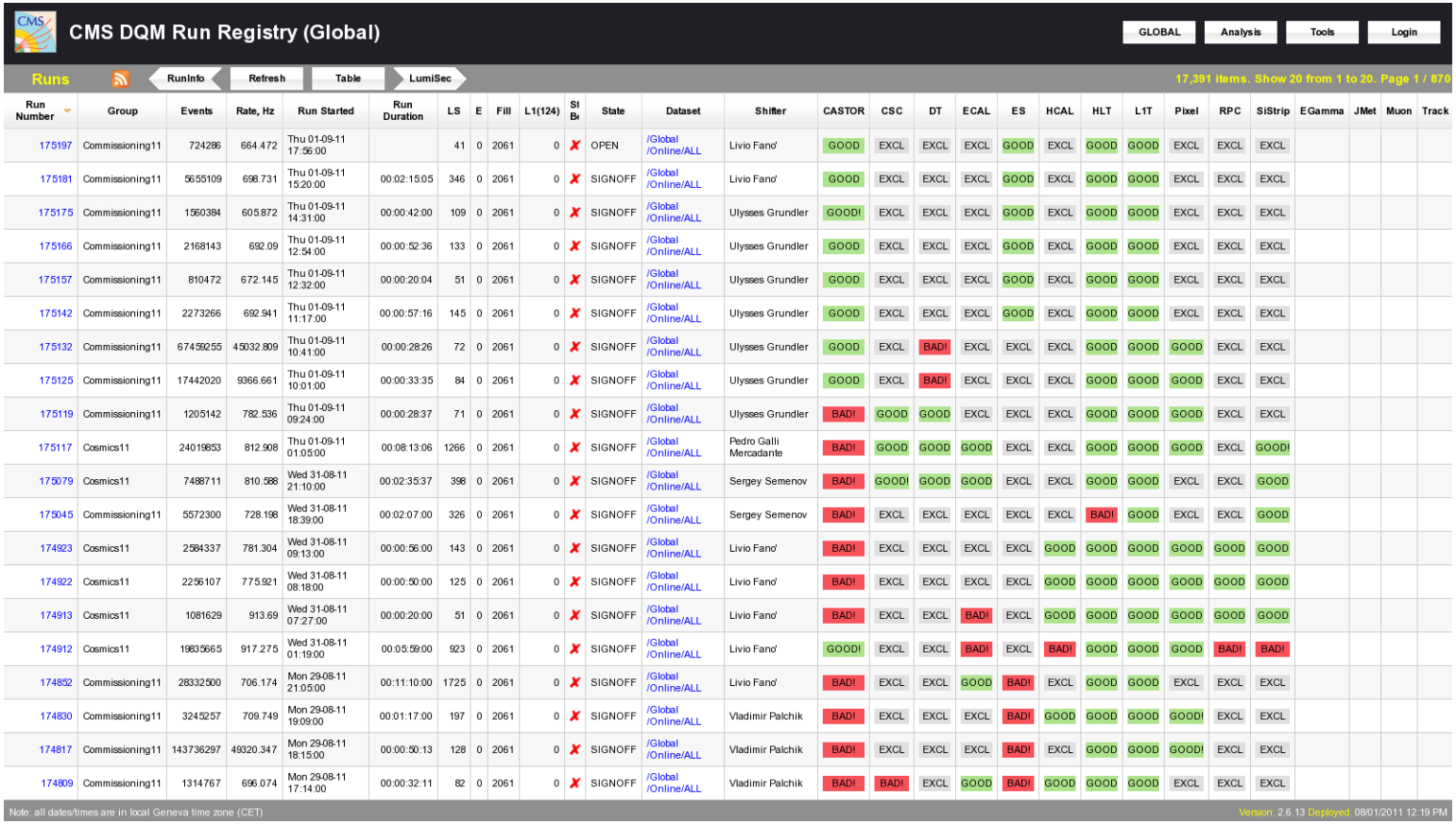


New Histogram on DQM GUI



Data Certification - Run Registry

- Data certification for physics analysis
- For each run good and bad flags are set based on DQM histogram inspection



CMS DQM Run Registry (Global)

GLOBAL | Analysis | Tools | Login

Runs | RunInfo | Refresh | Table | LumiSec

17,391 items. Show 20 from 1 to 20. Page 1 / 870

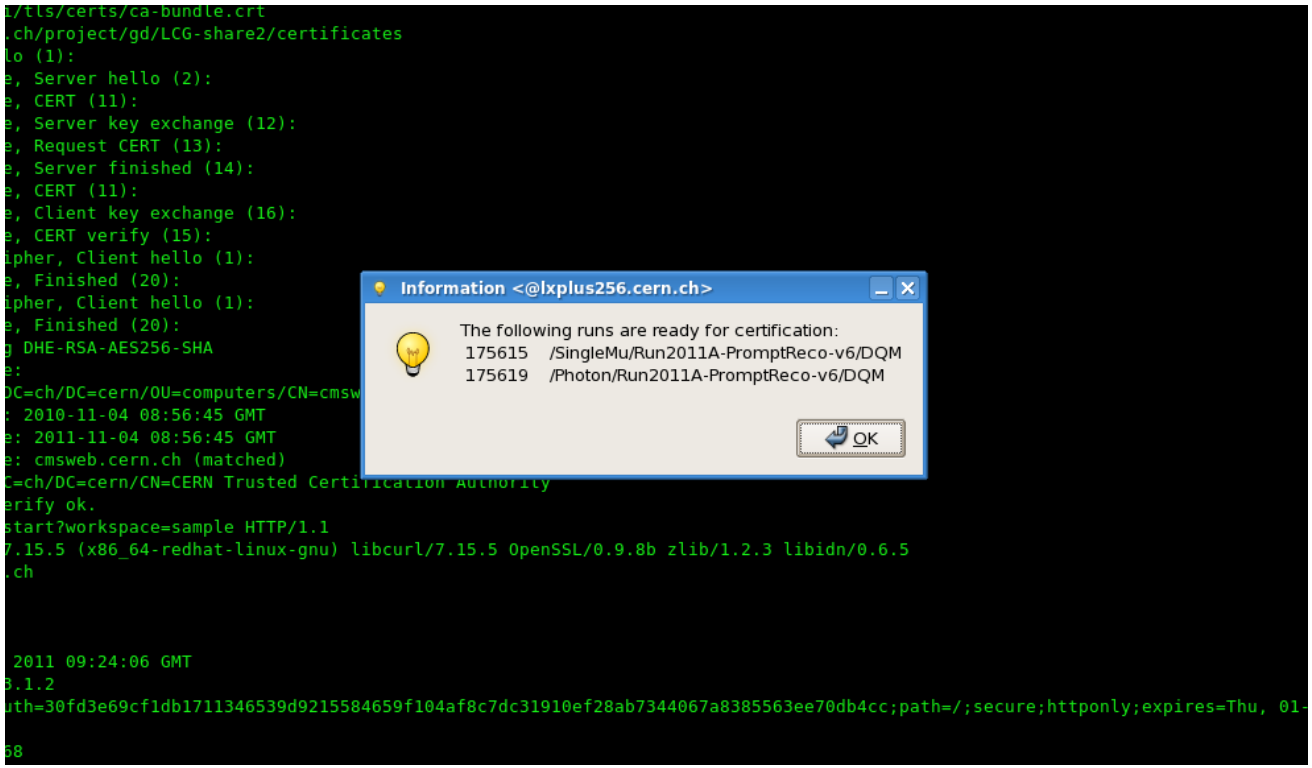
Run Number	Group	Events	Rate, Hz	Run Started	Run Duration	LS	E	Fill	L1(124)	Sl Be	State	Dataset	Shifter	CASTOR	CSC	DT	ECAL	ES	HCAL	HLT	L1T	Pixel	RPC	SiStrip	EGamma	JMet	Muon	Track
175197	Commissioning11	724286	664.472	Thu 01-09-11 17:56:00		41	0	2061	0	✗	OPEN	/Global /Online/ALL	Livio Fano'	GOOD	EXCL	EXCL	EXCL	GOOD	EXCL	GOOD	GOOD	EXCL	EXCL	EXCL				
175181	Commissioning11	5655109	698.731	Thu 01-09-11 15:20:00	00:02:15:05	346	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Livio Fano'	GOOD	EXCL	EXCL	EXCL	GOOD	EXCL	GOOD	GOOD	EXCL	EXCL	EXCL				
175175	Commissioning11	1560384	605.872	Thu 01-09-11 14:31:00	00:00:42:00	109	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Ulysses Grundler	GOOD!	EXCL	EXCL	EXCL	GOOD	EXCL	GOOD	GOOD	EXCL	EXCL	EXCL				
175166	Commissioning11	2168143	692.09	Thu 01-09-11 12:54:00	00:00:52:36	133	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Ulysses Grundler	GOOD	EXCL	EXCL	EXCL	GOOD	EXCL	GOOD	GOOD	EXCL	EXCL	EXCL				
175157	Commissioning11	810472	672.145	Thu 01-09-11 12:32:00	00:00:20:04	51	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Ulysses Grundler	GOOD	EXCL	EXCL	EXCL	GOOD	EXCL	GOOD	GOOD	EXCL	EXCL	EXCL				
175142	Commissioning11	2273266	692.941	Thu 01-09-11 11:17:00	00:00:57:16	145	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Ulysses Grundler	GOOD	EXCL	EXCL	EXCL	GOOD	EXCL	GOOD	GOOD	EXCL	EXCL	EXCL				
175132	Commissioning11	67459255	45032.809	Thu 01-09-11 10:41:00	00:00:28:26	72	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Ulysses Grundler	GOOD	EXCL	BAD!	EXCL	EXCL	EXCL	GOOD	GOOD	GOOD	EXCL	EXCL				
175125	Commissioning11	17442020	9366.661	Thu 01-09-11 10:01:00	00:00:33:35	84	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Ulysses Grundler	GOOD	EXCL	BAD!	EXCL	EXCL	EXCL	GOOD	GOOD	GOOD	EXCL	EXCL				
175119	Commissioning11	1205142	782.536	Thu 01-09-11 09:24:00	00:00:28:37	71	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Ulysses Grundler	BAD!	GOOD	GOOD	EXCL	EXCL	EXCL	GOOD	GOOD	GOOD	EXCL	EXCL				
175117	Cosmics11	24019853	812.908	Thu 01-09-11 01:05:00	00:08:13:06	1266	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Pedro Galli Mercadante	BAD!	GOOD	GOOD	GOOD	EXCL	EXCL	GOOD	GOOD	GOOD	EXCL	GOOD!				
175079	Cosmics11	7488711	810.588	Wed 31-08-11 21:10:00	00:02:35:37	398	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Sergey Semenov	BAD!	GOOD!	GOOD	GOOD	EXCL	EXCL	GOOD	GOOD	EXCL	EXCL	GOOD				
175045	Commissioning11	5572300	728.198	Wed 31-08-11 18:39:00	00:02:07:00	326	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Sergey Semenov	BAD!	EXCL	EXCL	EXCL	EXCL	EXCL	BAD!	GOOD	EXCL	EXCL	GOOD				
174923	Cosmics11	2584337	781.304	Wed 31-08-11 09:13:00	00:00:56:00	143	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Livio Fano'	BAD!	EXCL	EXCL	EXCL	EXCL	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD				
174922	Cosmics11	2256107	775.921	Wed 31-08-11 08:18:00	00:00:50:00	125	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Livio Fano'	BAD!	EXCL	EXCL	EXCL	EXCL	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD				
174913	Cosmics11	1081629	913.69	Wed 31-08-11 07:27:00	00:00:20:00	51	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Livio Fano'	BAD!	EXCL	EXCL	BAD!	EXCL	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD				
174912	Cosmics11	19835665	917.275	Wed 31-08-11 01:19:00	00:05:59:00	923	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Livio Fano'	GOOD!	EXCL	EXCL	BAD!	EXCL	BAD!	GOOD	GOOD	GOOD	BAD!	BAD!				
174852	Commissioning11	28332500	706.174	Mon 29-08-11 21:05:00	00:11:10:00	1725	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Livio Fano'	BAD!	EXCL	EXCL	GOOD	BAD!	EXCL	GOOD	GOOD	EXCL	EXCL	EXCL				
174830	Commissioning11	3245257	709.749	Mon 29-08-11 19:09:00	00:01:17:00	197	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Vladimir Palchik	BAD!	EXCL	EXCL	EXCL	BAD!	GOOD	GOOD	GOOD	GOOD!	EXCL	EXCL				
174817	Commissioning11	143736297	49320.347	Mon 29-08-11 18:15:00	00:00:50:13	128	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Vladimir Palchik	BAD!	EXCL	EXCL	EXCL	BAD!	EXCL	GOOD	GOOD	GOOD!	EXCL	EXCL				
174809	Commissioning11	1314767	696.074	Mon 29-08-11 17:14:00	00:00:32:11	82	0	2061	0	✗	SIGNOFF	/Global /Online/ALL	Vladimir Palchik	BAD!	BAD!	EXCL	GOOD	BAD!	GOOD	GOOD	GOOD	EXCL	EXCL	EXCL				

Note: all dates/times are in local Geneva time zone (CET)

Version: 2.6.13 Deployed: 08/01/2011 12:19 PM

Checking for Available Runs

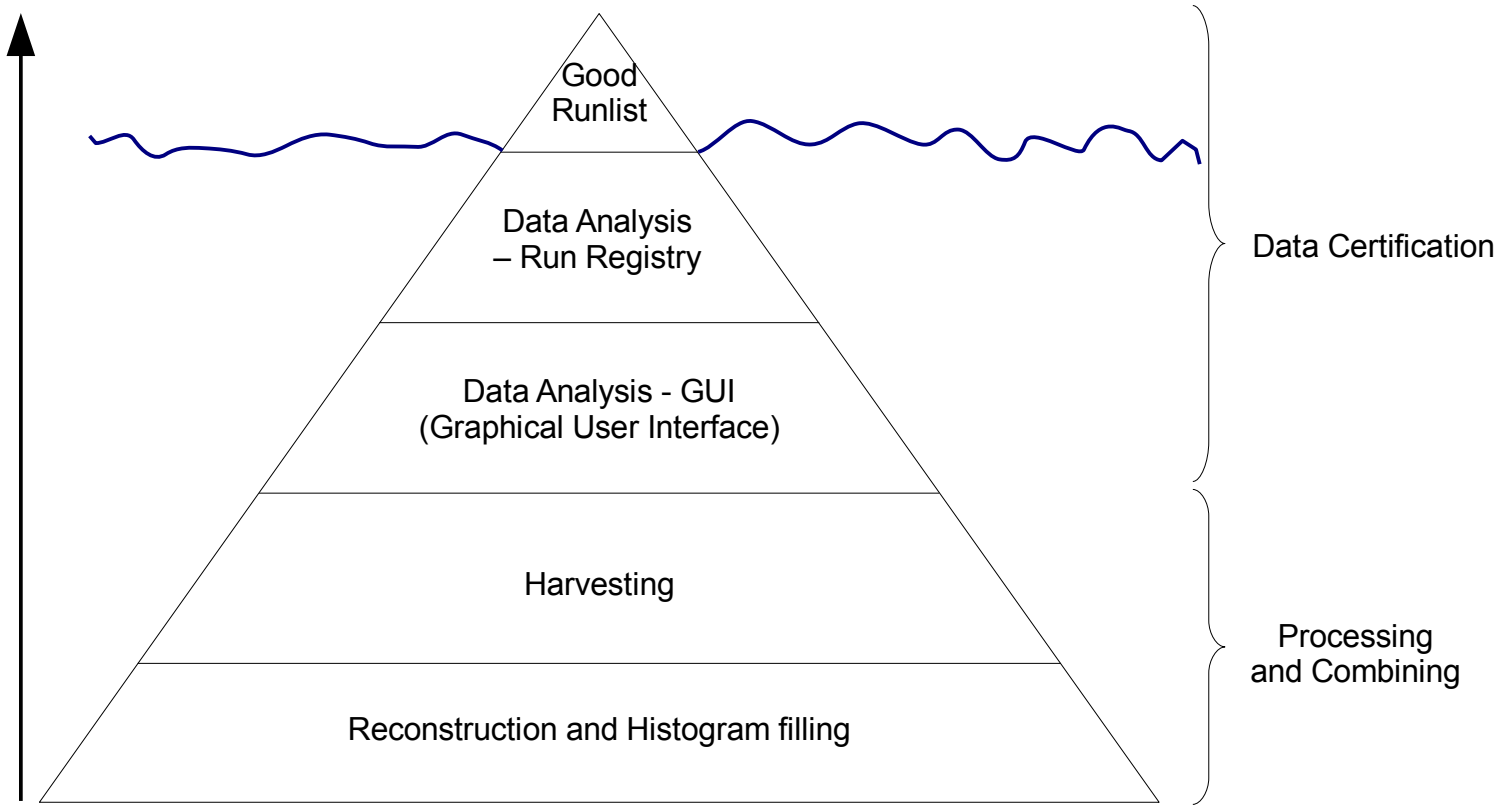
- Checking for new runs for certification is an iterative task which is done by hand during the shift
- Script automatically checks for runs which are available and ready for certification
- Runs are ready for certification once data has arrived in the GUI - script checks for this



Scripting Technicalities - Python to the Rescue!

- Python is good at connecting to and handling servers
- Python also makes string and file handling easier
- This is useful for:
  - Fetching data from many different locations
  - Authentication with GRID and for using https connections
  - Data validation - checking that everything is correct
  - Running the script periodically
  - Connecting to and using API (Application Programming Interface)

DQM Iceberg


Good
Runlist
Data Analysis
– Run Registry
Data Analysis - GUI
(Graphical User Interface)
Harvesting
Reconstruction and Histogram filling
Data Certification
Processing
and Combining

Thank you for the opportunity to take part in the Summer Student Program in the CMS group!

Thank you for your attention!