

The Data Quality Monitoring Software for the CMS experiment at the LHC

2020-07-31, ICHEP 2020 Presented by Andrius Kirilovas on behalf of DQM Team

CMS Data Quality Monitoring (DQM)

The Data Quality Monitoring (DQM) is a subgroup of the Physics Performance & Dataset (PPD) organization. The primary goal of the group is to provide the following services to the CMS collaboration:

- 1. **Online Data Quality Monitoring:** The goal is to spot problems in the CMS detector while it is running
- 2. Offline Data Quality Monitoring and Data Certification (DC): The goal is to determine which data can be used for data analysis
- 3. The DQM framework and DQM GUI are also used for **release validation**. The goal is to test specific software releases and validate that they produce correct results

DQM produces monitor elements (MEs)

Monitor elements (MEs) are the bread and butter of the CMS DQM

ME usually is 1D or 2D distribution of an observable from a CMS subdetector (eg.: occupancy maps, efficiencies or physics distributions)

All DQM MEs are accumulated per run* but some are also available per lumisection*. Up to LS 2 we had no means to conveniently work with per lumisection plots

* Lumisection is 23s of datataking (atomic, constant unit)* Run is an arbitrary number of lumisections



The DQM system within CMS

- An overall schema of CMS data processing
- Parts in grey are DQM tools and services
- CMS has a centralised DQM system for all subdetectors



Goldet JSON - certified data used for physics analysis RelVal - Release Validation samples

DQM within the CMS event processing framework (CMSSW)

- Central DQM code that provides a framework to analyze events and produce MEs (histograms)
- White components above represent reconstruction
- Monitoring code that uses the DQM framework is developed and maintained by the subsystems



DQM within the CMS event processing framework CMSSW (Online)

- Running on dedicated machines on site of the experiment (P5)
- We get 1/10 data from HLT, the rest is sent to process Offline
- Each client (subsystem) runs on its own process which adds flexibility
- Snapshots of histograms are sent to the DQM GUIs over the DQMNet network protocol (more on that later)



DQM within the CMS event processing framework CMSSW (Offline)

- DQM code running as part of the full CMS data processing in dedicated computing sites
- Or as part of simulations or release validation process
- Data is processed on 1000s of machines and on 1000s of CPUs in parallel
- We recently modernised and cleaned up the DQM framework for readability and maintainability



The DQM GUI

- Receives ROOT files containing MEs, stores them and provides access to each ME via web interface
- Online DQM GUI also accepts snapshots of each ME of the ongoing run to provide live overview of the detector
- Provides APIs* for other systems to interact with the DQM data



The new DQM GUI (currently being developed)

- Has never been used in data taking yet, in the final stage of development currently
- 2. Significantly reduces the complexity of the code
- Makes sure that raw ROOT files containing the MEs are always available
- 4. New instance (with full data) can be deployed in the matter of seconds
- 5. Is part of CMSSW* for easier deployments
- Provides higher granularity (per lumisection*) data. Old tool showed only per run* data

* Lumisection is 23s of datataking

- * Run is an arbitrary number of lumisections
- * CMSSW main CMS event processing framework



Run Registry

• Web service that aggregates certification information



Run Registry

- Web service that aggregates certification information
- Provides APIs to produce a final set of data that is ready to be used for physics analysis (golden JSON)
- The new version of this tool (upgraded for maintainability, usability and new features) is already in the final stage of commissioning
- Supports lumisection flags

Run Registry	ONLINE	OFFLINE	JSON CR	EATION	JSON PORTA	L	LOG					And	rius Kiril	ovas Lo	g out
WORKSPACES	Offline / Dat	tasets / Global					G Show Cycle	s		Fee	edback is	welcome!	(JIRA)	Configura	ition v
GLOBAL	Datasets wait	ting to appear in	DQM GUI (17	'):						Exp	ort to CSV	Get /	API Call	Open in	HDQM
BTAG	AND Y	+Rule +	Group												
	Run Number	Dataset Name	Class	Manage / LS	GUI Link	OMS	global State	LS Duration	btag	castor	cms	CSC	ctpps	lowlumi	dt
CASTOR	334280	/Express/Comm	Commissioning	1 Manage / LS	GUI Link V	OMS	waiting dqm gui / move	8	No Lumisec	No Lumisec	No Lumisec	No Lumisec	No Lumise	c No Lumised	: No Lumise
	334279	/Express/Comm	Commissioning	1 Manage / LS	GUI Link V	OMS	waiting dqm gui / move	6	No Lumisec	No Lumisec	No Lumisec	No Lumisec	No Lumise	c No Lumised	No Lumise
CMS	334277	/Express/Comm	Commissioning	1 Manage / LS	GUI Link V	OMS	waiting dqm gui / move	1430	No Lumisec	No Lumisec	No Lumisec	No Lumisec	No Lumise	c No Lumised	No Lumise
	334274	/Express/Comm	Commissioning	1 Manage / LS	GUI Link V	OMS	waiting dqm gui / move	106	No Lumisec	No Lumisec	No Lumisec	No Lumisec	No Lumise	c No Lumised	No Lumise
CSC	334269	/Express/Comm	Commissioning	1 Manage / LS	GUI Link∀	OMS	waiting dqm gui / move	87	No Lumisec	No Lumisec	No Lumisec	No Lumisec	No Lumise	c No Lumised	No Lumise
CTPPS		Previou	IS		Page	1	of 4	5 ro	vs Y				Next		
DC	Editable data	sets (25550):								Exp	ort to CSV	Get /	API Call	Open in	HDQM
DT	AND ~	+Rule +	Group												
	Run Number	Dataset Name	Class	Manage / LS	GUI Link	OMS	global State	LS Duration	btag	castor	cms	csc	ctpps	lowlumi	dt
ECAL	327744	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	723	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
	327743	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	812	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
EGAMMA	327740	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link∀	OMS	COMPLETED / move	353	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
	327696	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link∨	OMS	COMPLETED / move	70	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
HCAL	327693	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link∀	OMS	COMPLETED / move	480	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
	327692	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	163	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
HLT	327692	/TestRun/HICos	Cosmics18	Manage / LS	Not in GUI yet	OMS	COMPLETED / move	163	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
	327676	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	23	GOOD	EXCLUDED	BAD	EXCLUDED	BAD	BAD	GOOD
JETMET	327618	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	666	GOOD	EXCLUDED	BAD	GOOD	BAD	BAD	GOOD
	327604	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	219	GOOD	EXCLUDED	BAD	GOOD	BAD	BAD	GOOD
L1T	327601	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link 🗸	OMS	COMPLETED / move	29	GOOD	EXCLUDED	BAD	GOOD	BAD	BAD	GOOD
	327600	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	39	GOOD	GOOD	BAD	GOOD	BAD	BAD	GOOD
LUMI	327596	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link	OMS	COMPLETED / move	140	GOOD	GOOD	BAD	GOOD	BAD	BAD	GOOD
	327593	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	12	GOOD	GOOD	BAD	GOOD	BAD	BAD	GOOD
	327592	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link	OMS	COMPLETED / move	30	GOOD	GOOD	BAD	GOOD	BAD	BAD	GOOD
<	327590	/PromptReco/HI	Cosmics18	Manage / LS	GUI Link V	OMS	COMPLETED / move	91	GOOD	GOOD	BAD	GOOD	BAD	BAD	GOOD
	327589	/PromptReco/HI	Cosmics18	Manage /1 S	GUILLink	OMS	COMPLETED / move	44	6000	6000	BAD	6000	BAD	BAD	6000

Historic DQM (HDQM)

- HDQM allows to monitor the trends of DQM quantities over long periods of time
- It extracts a single value from a DQM ME and plots it as a function of run number



Historic DQM (HDQM)

- HDQM allows to monitor the trends of DQM quantities over long periods of time
- It extracts a single value from a DQM ME and plots it as a function of run number
- Used to debug slow cooking issues with the detector



DQM Users

- Shifter constantly monitors the DQM MEs as data is coming in (online monitoring)
- Validators monitor DQM MEs to validate new software releases (release validation)
- Shifter also use DQM MEs to distinguish good data from bad data on run by runs basis (offline monitoring and certification)
 - Ongoing ML studies might simplify their work and provide data certification with higher granularity on a lumisection basis
- Detector experts consult various DQM tools in case their subsystem malfunctions



Some numbers about CMS DQM

- We estimate that all per run DQM ROOT files add up to about 100TB
- We estimate that higher granularity per lumisection data adds up to 100TB per year!
- $5.5 \cdot 10^{10}$ MEs in the Offline DQM GUI, covering $1.2 \cdot 10^{9}$ events processed by DQM
- Online 1MB @ 100Hz stream for processing



Future plans

- Finalizing the development and integration of the new DQM GUI
 - Both frontend and backend
 - Horizontally scalable deployment in Kubernetes environment
- Completing the work on the new Run Registry
- Integrating machine learning based feedback into data certification process



Conclusion

- We are happy with the state and the performance of DQM during Run 2
 - During the last MWGR all shifts were carried out remotely and very successfully
- We have new tools in the commissioning stage and seeking for new ideas/developments especially on the ML side
- With current developments approaching the end, we are confident that the DQM system is ready for the challenges of Run 3

🕎 Run Regis	ry ONLINE DOFTLINE JSON CREATION JSON PORTAL LOG Andrius Kirilavas Legiout	CNS Run: < 317520 >> Lurri: Al >> Dataset Select >> [ZeroStas@Run20185-PromptRecs-y1,DQM >>> Q Workspace: Everything (Enter plot name Q)	DQM Help PitePhase1	PromptRico
WORKSPACES *	Offline / Datasets / Global Feedback is welcome! (JRA) Configuration ~	/> PiestPhase1 > Phase1_Mechanicat/law (\$ Settings	Select runs and search for plots	
GLOBAL	Datasets welling to appear in DOM GUI (17): Expert to CSV Get API Call Open in HDOM	> Options	Latest v 50	Search query Options Plot list Submit
BTAG	AND V +Rule +Group	PXBarrel PXForward	Periduals V EBY per disk - (BMS)	Pariduals V CPIV par dick - (Maap)
CASTOR	Pun-Humber Defaset Name Dass Manage / LS GUILLA GV/S global Data LIS Duration bing center ons co diges Invitant dt Station / topores/Cencer, Cencres/cencer, Cencres/cencer, Stations/ Dots Webling driving / Invite 3 No Londers In Duration Duration (LS Londers In Londers Nummer Terms Cencer/Cencer Stations) Manage / LS GUILLA OUS Webling driving / Invite S No Londers Duration Durations Duration Durations Duration	sizeX_PXBarrel 0 num_clusters_per_Lumi Size_per_LumiBlock_per 0 sdc_per_LumiBlock_PXF 0 sizeyveeta_PXBarrel 0 Sizeyveeta_PXBarrel 0 Sizeyveeta_PXBarrel 0 Sizeyveeta_PXBarrel 0	Kesidudi X (PR), peritudi k = (Mild) Kesidudi X (PR), peritudi kilis, ms, prosp Mater 8.1235e-3, 1945: 1.64521e-3 0.022	Residual; (1774); per kusik – (1766); j Residual; (1776); per kusik (1876); mest, prosp Mear: -3.08766-5, 8MS: 3.59526-5
CMS	334277 (Tepress/Comm. Commissioning) Manage / (5 GKLink+ OUS waiking don gul / now 1430 No Lumise:		8 0.021	l univ
CSC	1912169 (Supress/Comm. Commissioning) Manage (15 011 Linky OV5 waiking days gal / mone 87 No Lumine: No Lumine			· · · ·
CTPPS	Previous Poge 1 of 4 Sincers V Next		1	
DC	Editable datasets (25550): Expert to CSV Get API Call Open in HDQM		a 0000	*** X
DT	AND V +Rule +Group	chana per i umilioni a chananalita ya OYEA A aparterata per BY A nindernovalita per bia		g -0.0016 12k
ECAL	Pur Number Defaset Neme Class Menopy / (3 CUI Link OV/S global Stats IS Dunition Integ caster ema coc etipps lowlum dt S20744 ,/PermptRecolf Ceanter 16 Manopy / (3 GLI Link) OV/S GOMPLETICI / Intere 723 COCO DXXLECIO 840 E500000 840 E600 COCO	Construction of the second secon		
EGAMMA	327743 (PrompRecolf Centicn18 Memory / LS GUI Date: OVER COMP Date: DAD ERXUSIO ERXUSIO ERXUSIO ERXUSIO ERXUSIO ERXUSIO ERXUSIO ERXUSIO <td></td> <td>6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7</td> <td></td>		6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	
HCAL	33766 (ParageSeco)H. Connects1 Manage /LS Ball Line: DUS Connects10 Jinov 70 GOOD Description BAD BAD <th< td=""><td></td><td>Residuals X FPRC perMinusDisk, rms Residuals X FPRC Dm2, rms Residuals X FPRC Dm1, rms Residuals X FPRC Dm1, rms File</td><td>Residuals_Y_FPRX_perMinusDisk_mean Residuals_Y_FPIX_Dm1_mean Fils</td></th<>		Residuals X FPRC perMinusDisk, rms Residuals X FPRC Dm2, rms Residuals X FPRC Dm1, rms Residuals X FPRC Dm1, rms File	Residuals_Y_FPRX_perMinusDisk_mean Residuals_Y_FPIX_Dm1_mean Fils
HLT	22760 //Www.994404. Clowersta Manager/13 00 Data 000 Collectria Januar 163 0000 Potection 442 (2000) 000 000 0000 20000 20000 000 000 0		Show / Change ranges	Show / Change ranges
JETMET	337876 (PermpRecoNL) Countin 18 Manage / LS GUI Link v OVE COMPUTED / Imave 2.3 GOOD DECLESS BAD EMO			
117	3272001 (PrampRecold	size_per_LumiBlock.per 6 sizeX_PXForward 6 size_per_LumiBlock.PX 6 adc_per_LumiBlock.PX 6 adc_per_LumiBlock.px 6 pXDisk PXForward 9XDisk PXForward 9	Residuals Y FPIX per disk - (RMS)	Residuals X FPIX per disk + (Mean)
LUMI	327160 (Panyudheo)M Canica's Manage (1.5 dot is Manage (1.5 </td <td></td> <td>Residuality V, PPR, perditivatilisk, mis, prosp Minar: 1.3265e-2, 1945: 3.3127e-3 606</td> <td>Roskkaals, 2, PPU, perflisz004, enem, prosp Mear: 3.1241a-5, IMS: 2.2044e-5 0.00018</td>		Residuality V, PPR, perditivatilisk, mis, prosp Minar: 1.3265e-2, 1945: 3.3127e-3 606	Roskkaals, 2, PPU, perflisz004, enem, prosp Mear: 3.1241a-5, IMS: 2.2044e-5 0.00018
	Access Presentation Control of the second s		8 0.024 48k	0.00012 48k
×	1972/00 (Prosperiore) Constant Property Constant Org. Constant / 2011 COCO COCO COCO COCO COCO COCO COCO C		a the second the second s	i the second

17