CYGNO Analysis Meeting 2023, 7-8 June
University of Coimbra, Portugal

Workflow for reconstruction, data monitoring and simulation processing

ROMA TRE
UNIVERSITÀ DEGLI STUDI

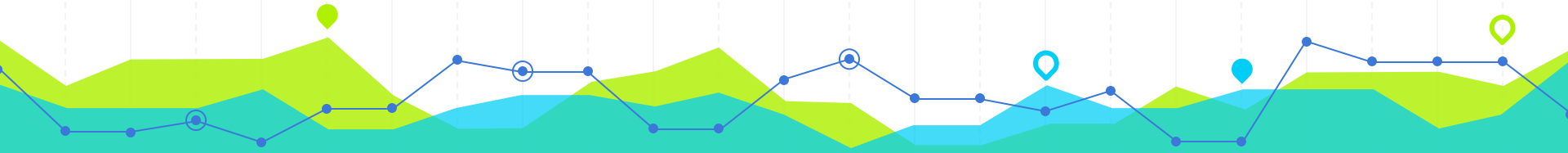
Igor Abritta

Summary

- Status
- Grafana Monitor
- Automation of data reco

To the discussion:

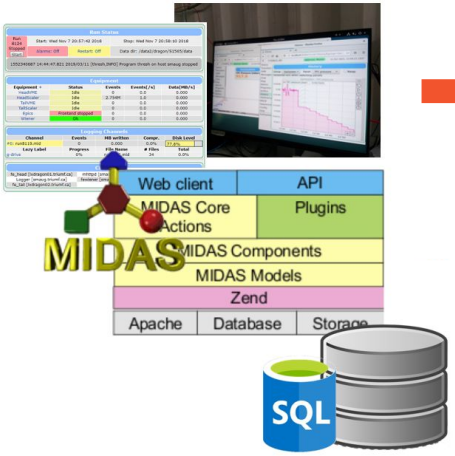
- Online Data Quality Monitor
- Post Processing



Status

data flow - schematic view

from detector to cloud (end of 2022 status)



local storage

DAQ running locally

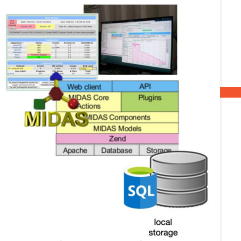
Reconstruction running locally

timestamp	Run	Event	Nclu	Size	nhits	Integral	xmean	ymean	xmax	xmin	ymax	ymin
1653989775	846	479	1	112	16	181.76470588235293	30.28095238095238	880.2346560846561	39	20	891	872
1653989775	846	479	2	128	20	265.52941176470588	30.775254441795523	209.04920523609164	51	28	215	204
1653989775	846	479	3	160	18	133.64705882352948	84.52822291317867	771.867017459938	95	76	779	764
1653989775	846	479	4	192	20	134.52941176470588	2248.870026292726	1335.6790534618758	2259	2236	1343	1328
1653989776	846	480	1	96	14	300.52941176470586	9.835922090729781	1813.0751972386588	15	8	1823	1804
1653989776	846	480	2	64	16	285.64705882352945	28.386762238828727	1729.1178892786334	31	24	1735	1724
1653989776	846	480	3	128	16	156.64705882352595	32.63340049760727	1007.1864406779661	43	24	1015	1000
1653989776	846	480	4	144	18	346.70588235294093	35.553263972009944	749.2419566697358	47	24	755	744
1653989776	846	480	5	144	21	230.47058823529412	42.592031182330004	592.2504330879169	51	36	607	584
1653989776	846	480	6	96	15	297.3529411764704	45.43109540636042	1853.155477031802	51	40	1859	1844
1653989776	846	480	7	176	23	329.41176470588215	46.96754199996738	2049.1464687652915	55	36	2059	2044
1653989776	846	480	8	144	18	189.82352941176467	52.8510526315789	820.8087368421054	59	44	827	812
1653989776	846	480	9	192	22	393.7647058823529	56.19953188999414	783.9046957284962	63	48	795	772
1653989776	846	480	10	160	23	159.23529411764707	99.18919814446654	1092.1045946542965	111	88	1099	1088
1653989776	846	480	11	176	31	235.88235294117644	1125.3636363636365	999.3466249319544	1135	1116	1007	992
1653989776	846	480	12	128	17	259.94117647058823	2191.3359281437124	1287.9009580838324	2199	2184	1299	1280
1653989776	846	480	13	352	52	480.41176470588255	2206.910145761411	1204.6896816263902	2227	2188	1227	1184
1653989776	846	480	14	160	21	324.2941176470589	2204.8646630576814	222.23595295874554	2211	2196	231	212
1653989777	846	481	1	128	14	258.70588235294093	13.376329649901892	2038.2439326654967	19	8	2047	2028
1653989777	846	481	2	320	45	800.3529411764705	28.612665219964496	2180.7683961333596	43	12	2199	2164
1653989777	846	481	3	144	20	417.4117647058825	31.037538410431033	846.1266506104145	47	20	851	840

Database on Cloud

data flow - schematic view

from detector to cloud (today's status)



DAQ running

Database storage and replication

Midas2Cloud

R2Console (Kafka producer)

Running locally

S3 Storage

SQL Replica on Cloud

Database ODB

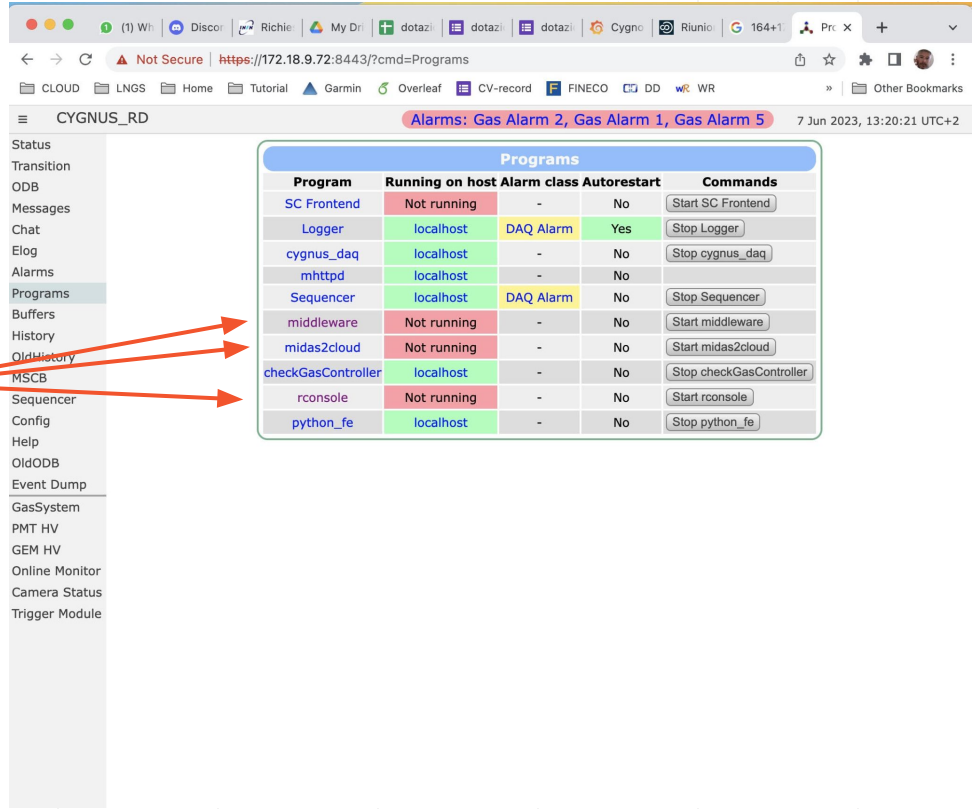
AutoReco

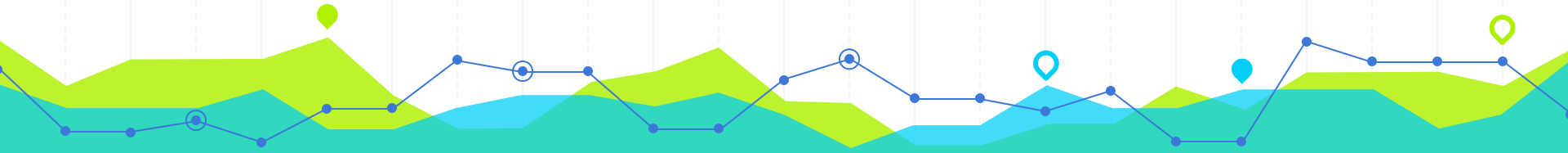
R2Console (Kafka Consumer)

Cloud Services

data flow

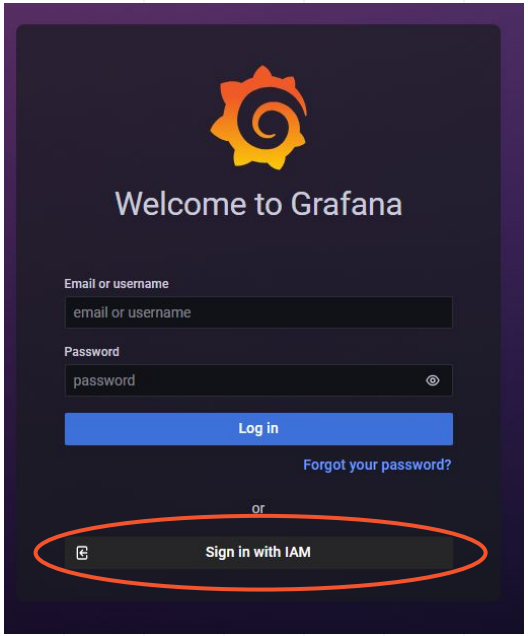
These are the process that must be running on the DAQ machine in order to feed the services showed in the previous slide



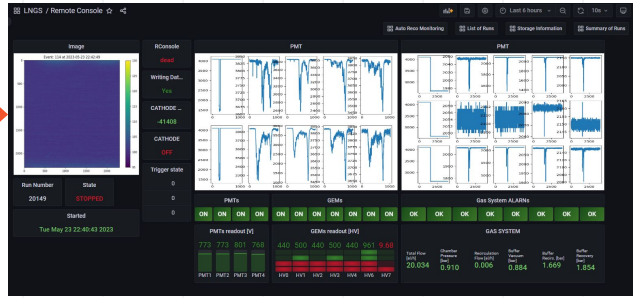


Grafana Monitor

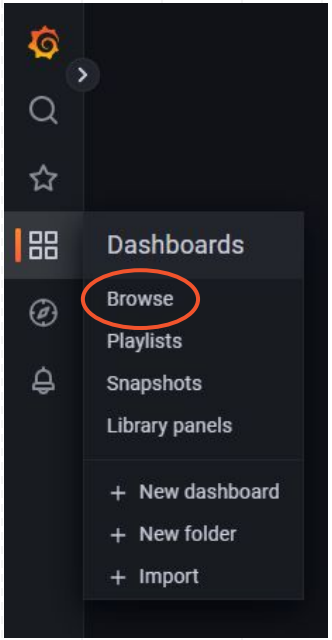
Grafana (https://grafana.cygno.cloud.infn.it/)

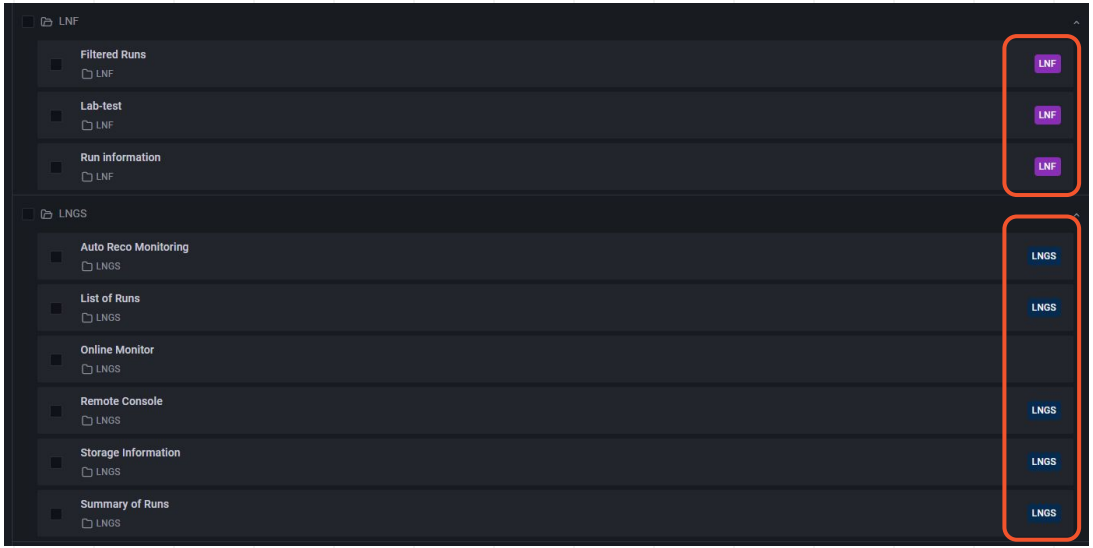


Home Page

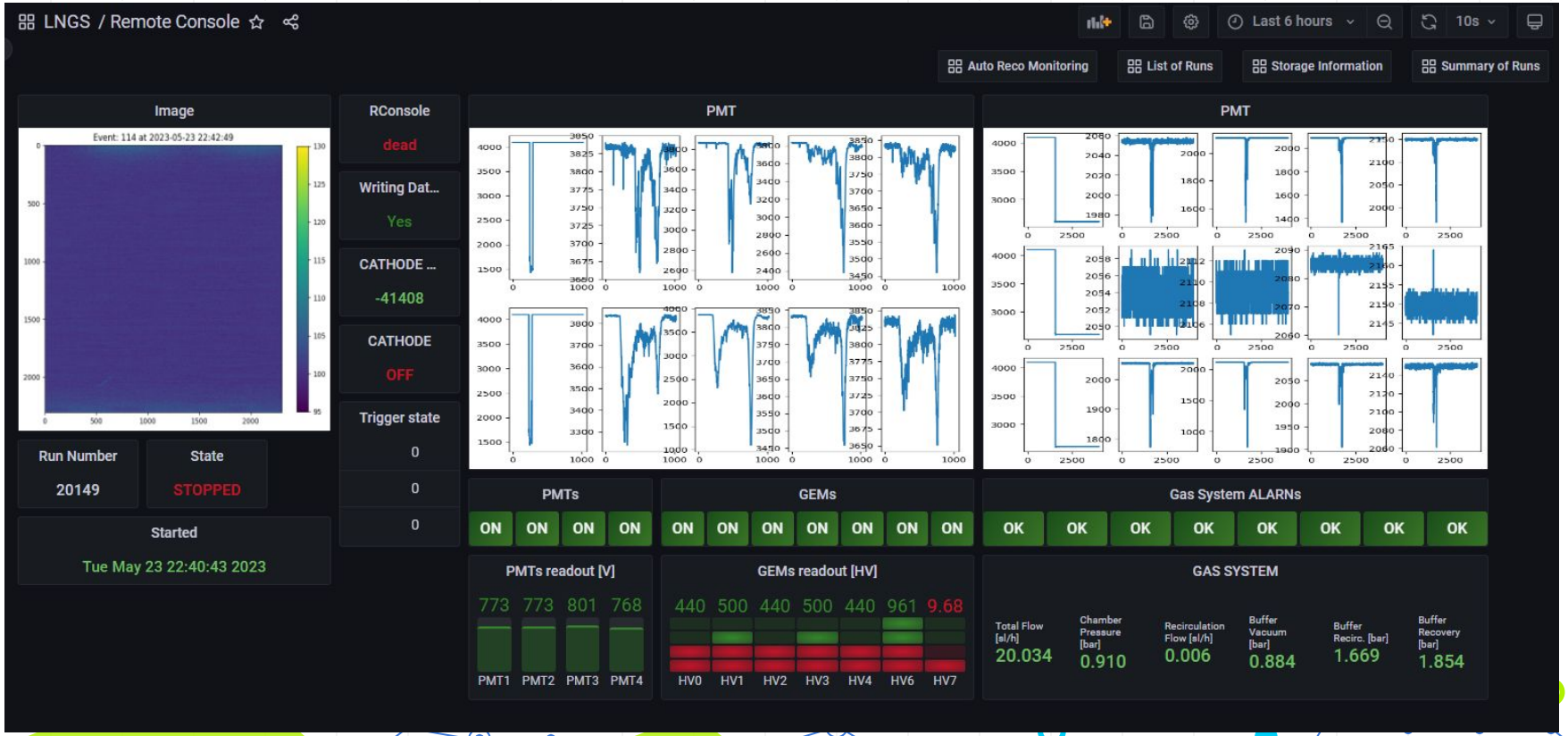


Grafana (https://grafana.cygno.cloud.infn.it/)

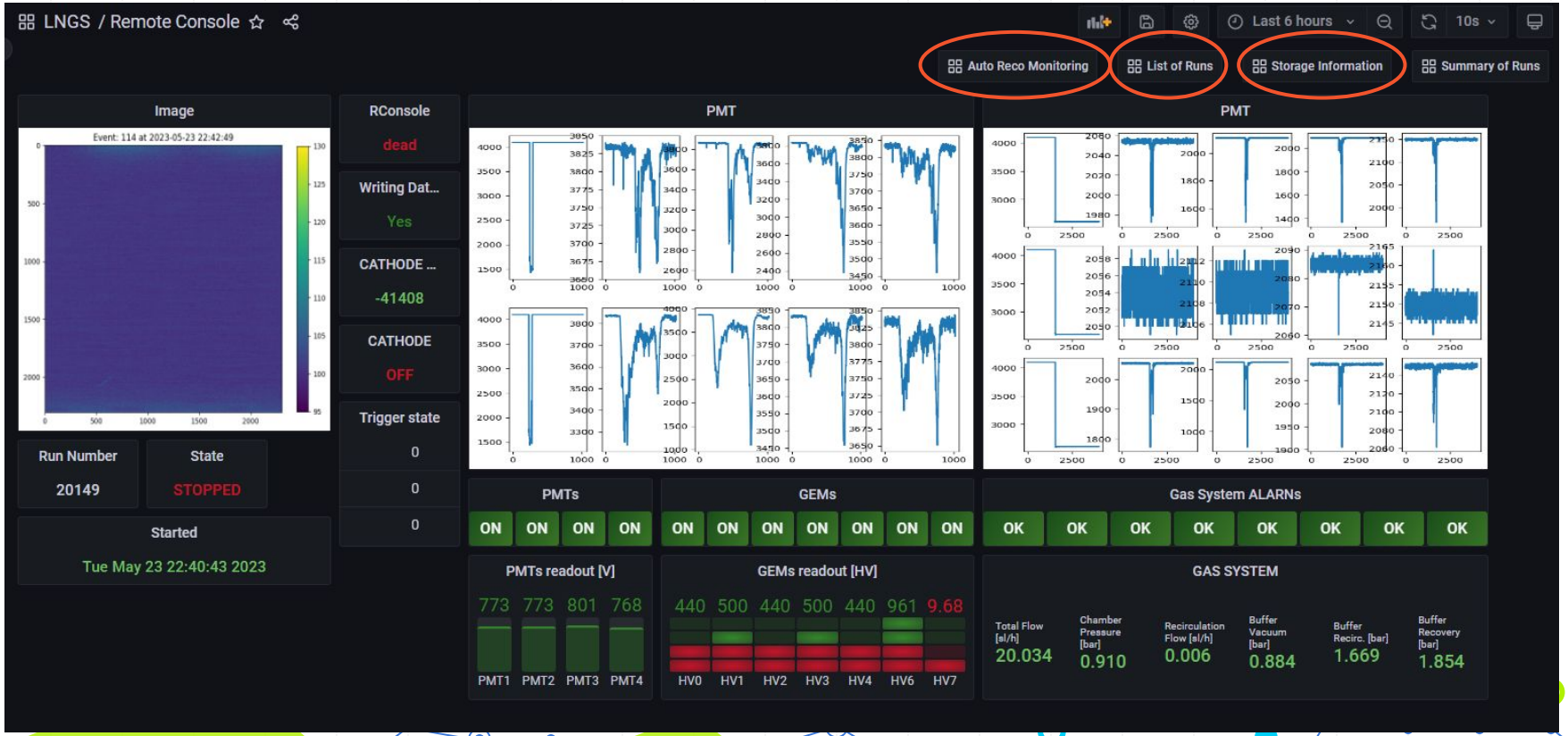




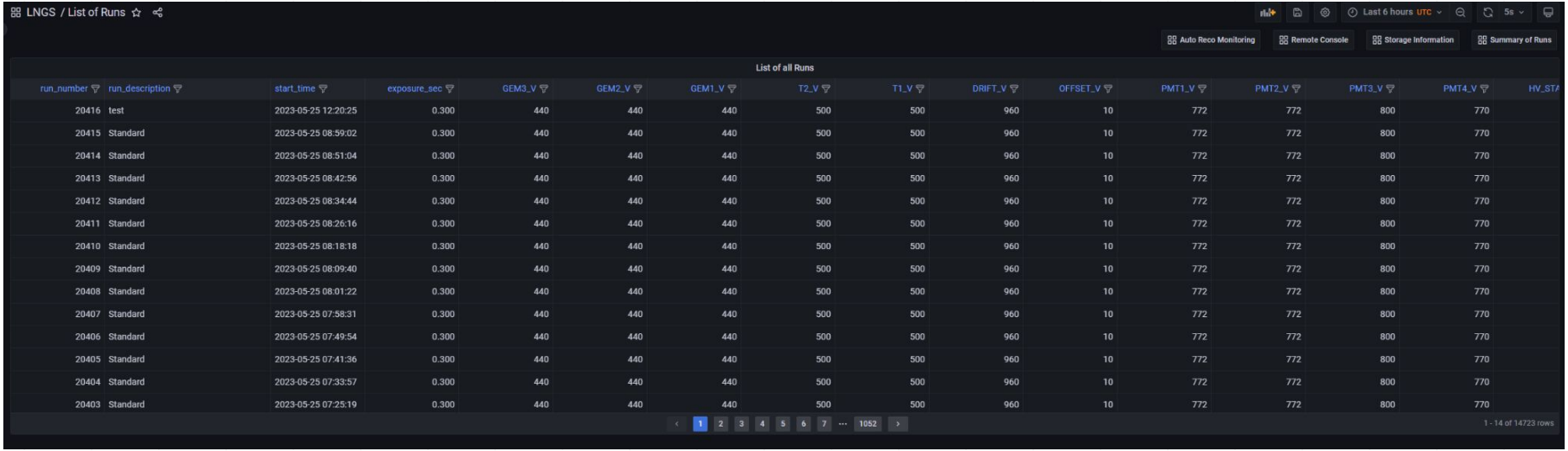
Grafana (https://grafana.cygno.cloud.infn.it/)



Grafana (https://grafana.cygno.cloud.infn.it/)



Grafana (https://grafana.cygno.cloud.infn.it/)



LNGS / List of Runs

Last 6 hours UTC | 5s

Auto Reco Monitoring | Remote Console | Storage Information | Summary of Runs

List of all Runs

run_number	run_description	start_time	exposure_sec	GEM3_V	GEM2_V	GEM1_V	T2_V	T1_V	DRIFT_V	OFFSET_V	PMT1_V	PMT2_V	PMT3_V	PMT4_V	HV_STA
20416	test	2023-05-25 12:20:25	0.300	440	440	440	500	500	960	10	772	772	800	770	
20415	Standard	2023-05-25 08:59:02	0.300	440	440	440	500	500	960	10	772	772	800	770	
20414	Standard	2023-05-25 08:51:04	0.300	440	440	440	500	500	960	10	772	772	800	770	
20413	Standard	2023-05-25 08:42:56	0.300	440	440	440	500	500	960	10	772	772	800	770	
20412	Standard	2023-05-25 08:34:44	0.300	440	440	440	500	500	960	10	772	772	800	770	
20411	Standard	2023-05-25 08:26:16	0.300	440	440	440	500	500	960	10	772	772	800	770	
20410	Standard	2023-05-25 08:18:18	0.300	440	440	440	500	500	960	10	772	772	800	770	
20409	Standard	2023-05-25 08:09:40	0.300	440	440	440	500	500	960	10	772	772	800	770	
20408	Standard	2023-05-25 08:01:22	0.300	440	440	440	500	500	960	10	772	772	800	770	
20407	Standard	2023-05-25 07:58:31	0.300	440	440	440	500	500	960	10	772	772	800	770	
20406	Standard	2023-05-25 07:49:54	0.300	440	440	440	500	500	960	10	772	772	800	770	
20405	Standard	2023-05-25 07:41:36	0.300	440	440	440	500	500	960	10	772	772	800	770	
20404	Standard	2023-05-25 07:33:57	0.300	440	440	440	500	500	960	10	772	772	800	770	
20403	Standard	2023-05-25 07:25:19	0.300	440	440	440	500	500	960	10	772	772	800	770	

‹ 1 2 3 4 5 6 7 ··· 1052 ›

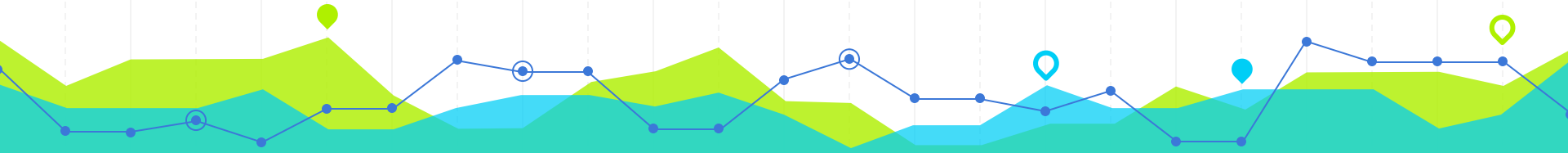
1 - 14 of 14723 rows

Grafana (https://grafana.cygno.cloud.infn.it/)

Grafana (https://grafana.cygno.cloud.infn.it/)

We also want to invite people to use it and give us feedback:

- what is missing?
- what could be modified/improved?
- what is not clear?
- etc.



Auto Reco

Auto Reco overview

Sentinel process with the following tasks:

- Check the SQL database to see if new run have arrived.
- If yes, Submit the Job using the Official Reconstruction code.
- Whole bunch of routines made to monitor all the process of: Submitting job, checking status, retrieve the output from the Queue machine, clean the Queue, update SQL database, upload the Output on the cloud storage, etc.
- Local dataframe create to show in real-time the auto reco status on Grafana Dashboard

HTCondor settings and Reco time:

- We have a dedicated queue only for the auto reco jobs with:
  - the possibility to run 16 jobs in parallel
  - Machines: 4 cores and 8Gb RAM (which give us 1 core per 2gb RAM)
- The reconstruction code is taking 3 seconds (using the cythonize noise reduction) in average for each image (using 4 cores), ~20 minutes per Run.

Auto Reco overview

Sentinel process with the following tasks:

- Check the SQL database to see if new run have arrived.
- If yes, Submit the Job using the Official Reconstruction code.
- Whole bunch of routines made to monitor all the process of: Submitting job, checking status, retrieve the output from the Queue machine, clean the Queue, update SQL database, upload the Output on the cloud storage, etc.
- Local dataframe create to show in real-time the auto reco status on Grafana Dashboard

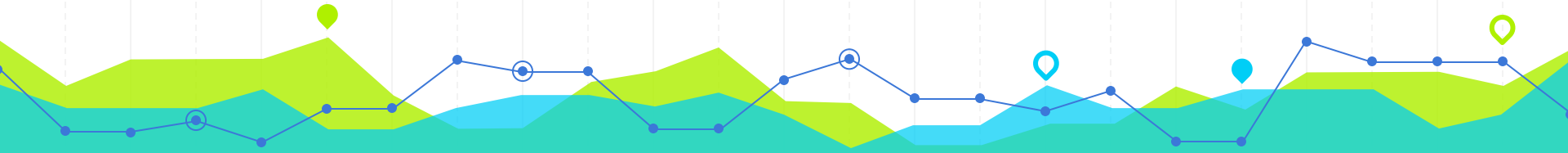
Auto Reco Monitoring

Cluster ID	Run Number	Status	Data Transfered	Cloud Storage	Job In Queue
3729	20416	completed	1	1	1
3728	20415	completed	1	1	1
3727	20414	completed	1	1	1
3726	20413	completed	1	1	1
3725	20412	completed	1	1	1
3724	20411	completed	1	1	1
3723	20410	completed	1	1	1
3722	20409	completed	1	1	1
3721	20408	completed	1	1	1
3720	20406	completed	1	1	1
3719	20407	completed	1	1	1
3718	20405	completed	1	1	1
3717	20404	completed	1	1	1
3716	20403	completed	1	1	1
3715	20402	completed	1	1	1
3714	20401	completed	1	1	1
3713	20400	completed	1	1	1
3712	20399	completed	1	1	1

Auto Reco overview

Sentinel process with the following tasks:

- Check the SQL database to see if new run have arrived.
- If yes, Submit the Job using the Official Reconstruction code.
- Whole bunch of routines made to monitor all the process of: Submitting job, checking status, retrieve the output from the Queue machine, clean the Queue, update SQL database, upload the Output on the cloud storage, etc.
- Local dataframe create to show in real-time the auto reco status on Grafana Dashboard

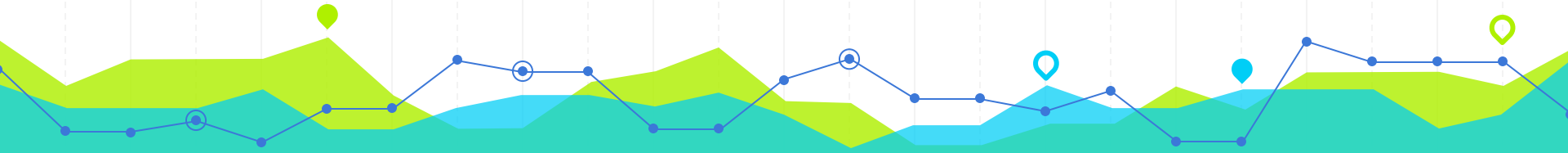


Online Data Quality Monitor

Online Data Quality Monitor

Here we have two options:

1. Develop a process similar to the autoreco to run over the reco output (but also accessing the .mid) and retrieve the relevant information to be shown by Grafana.
2. Add to the Reconstruction code the relevant Data Quality variables, in this way the AutoReco process will do all the job once, save it and Grafana is able to do the work of averaging the information in bunches of N seconds (Runs).

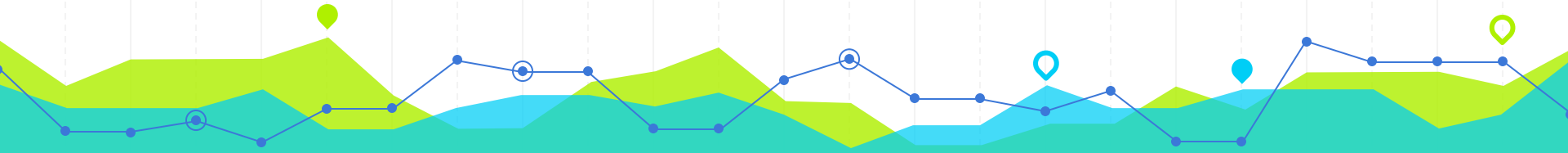


Post Processing

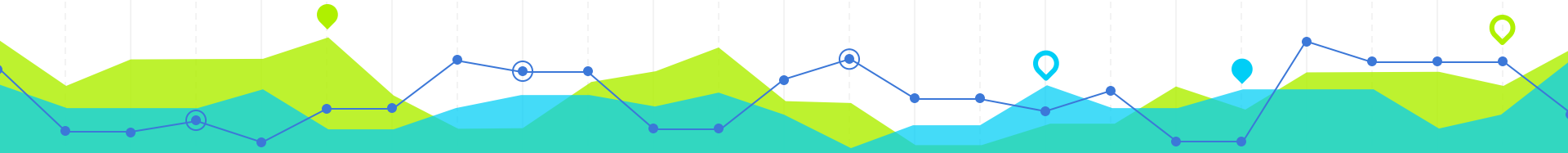
Post Processing
(more details on Emanuele's presentation)

Add to the Automatic Workflow also a second step taking the reco as input:

- This should be done for sure once, asap the reco is done;
- But also have the possibility to run async when we have an update of corrections, calibrations, output of discriminator algorithms that we want to make persistent in the friend ROOT trees.



THANK YOU!



Bonus Slide