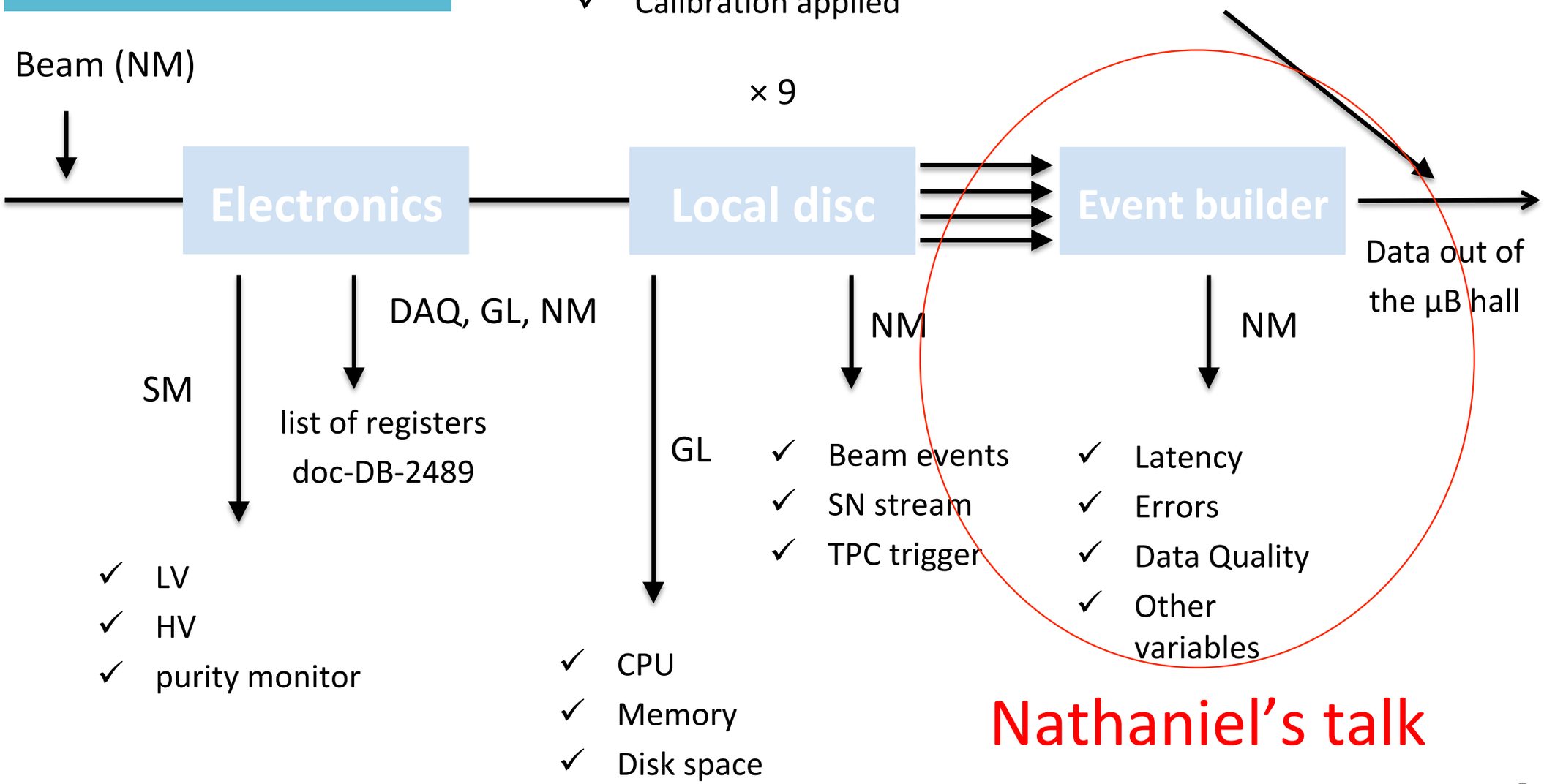
Ideas
for
Data Quality Monitoring

Kazuhiro Terao @ Nevis, Columbia

Online ideas as March 2013

Three different systems :
- ✓ Ganglia – GL
- ✓ NearLine – NM
- ✓ Slow Monitor - SM

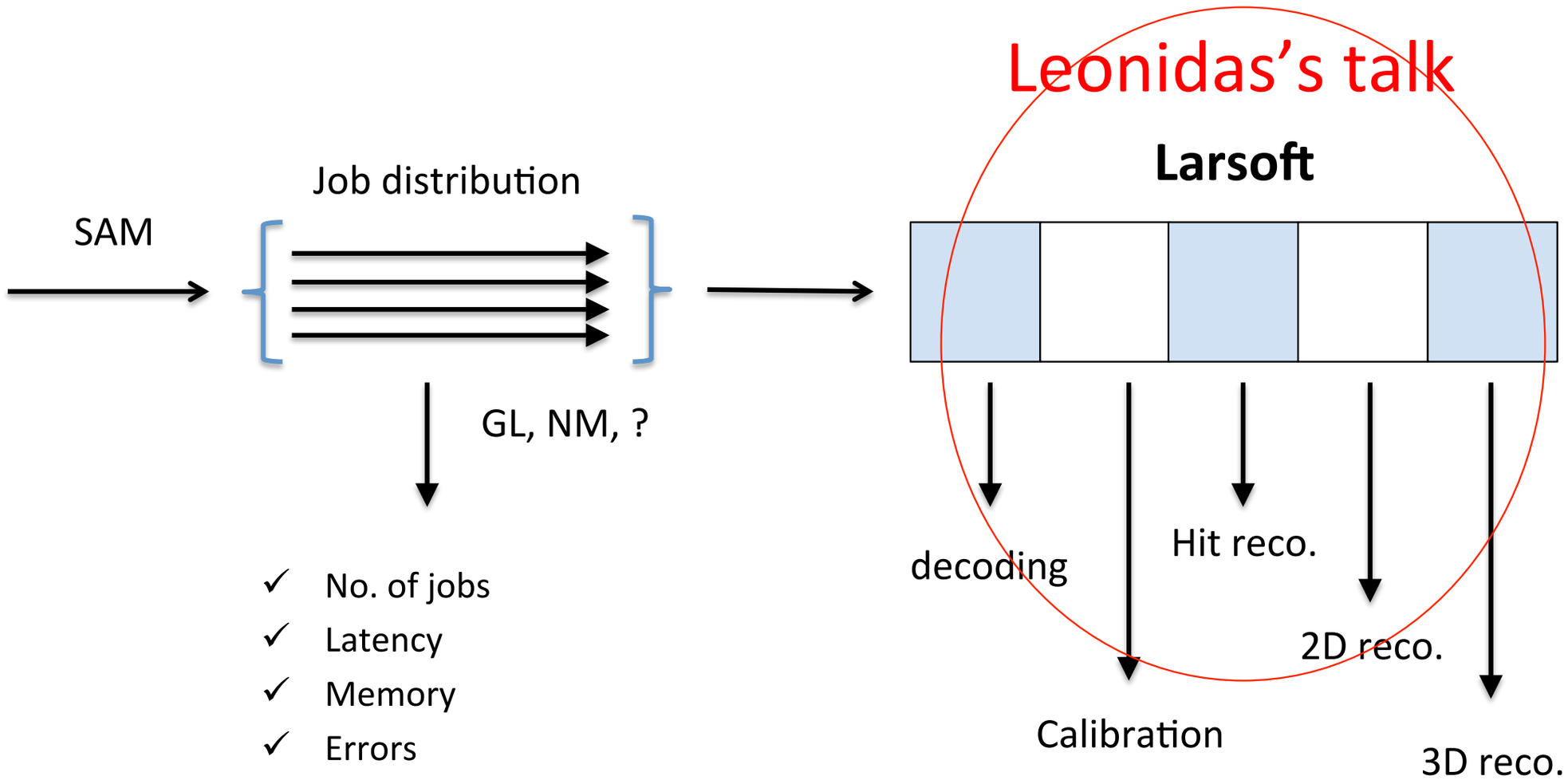
NearLine monitor (NM) display (doc-DB-2474/2385)
- ✓ Event display: DB-2385
- ✓ Data decoded and histograms plotted (including history plots)
- ✓ Calibration applied



3

Courtesy of Camillo M. (DocDB 2639)

Online ideas as March 2013



- ✓ No. of jobs
- ✓ Latency
- ✓ Memory
- ✓ Errors

Need more discussion/work ← Obtain fundamental histograms through larsoft scripts - send them back to the Online monitor displayer

4

Courtesy of Camillo M. (DocDB 2639)

... just as an example ...

I show what we did in Double Chooz

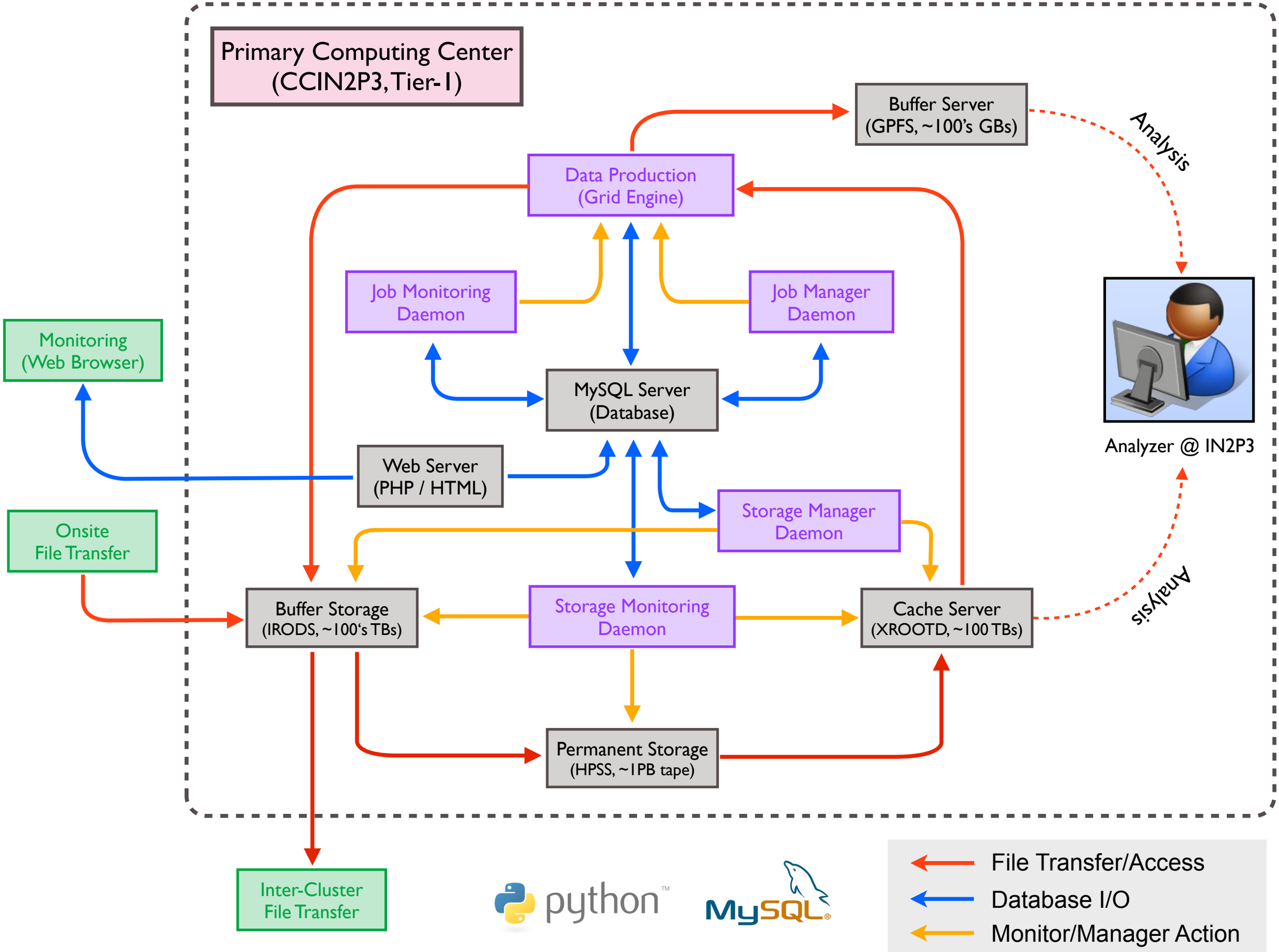
Double Chooz Data Processing & Monitoring

Offline Goals:

- High level physics data quality
  - Run-by-Run, day-by-day check on static histograms
    - Finding issues that require immediate attention
  - Query for a cumulative information (such as history)
- Data production management
  - Batch job processing
    - Multi-staged job processing (> 1 month, > 100k files, ~200 TBs)
  - Resource usage tracking
    - Useful for future planning
  - File availability

Strategy:

- Use database to store quality control parameters & process status flags
- Multiple daemon processes to log status and/or take action
- PHP scripts to query information through a web browser


Primary Computing Center
(CCIN2P3, Tier-1)
Buffer Server
(GPFS, ~100's GBs)
Analysis
Data Production
(Grid Engine)
Job Monitoring
Daemon
Job Manager
Daemon
Monitoring
(Web Browser)
MySQL Server
(Database)
Analyzer @ IN2P3
Web Server
(PHP / HTML)
Storage Manager
Daemon
Onsite
File Transfer
Analysis
Buffer Storage
(IRODS, ~100's TBs)
Storage Monitoring
Daemon
Cache Server
(XROOTD, ~100 TBs)
Permanent Storage
(HPSS, ~1PB tape)
Inter-Cluster
File Transfer
python
MySQL
File Transfer/Access
Database I/O
Monitor/Manager Action

Static Histogram Check

Description

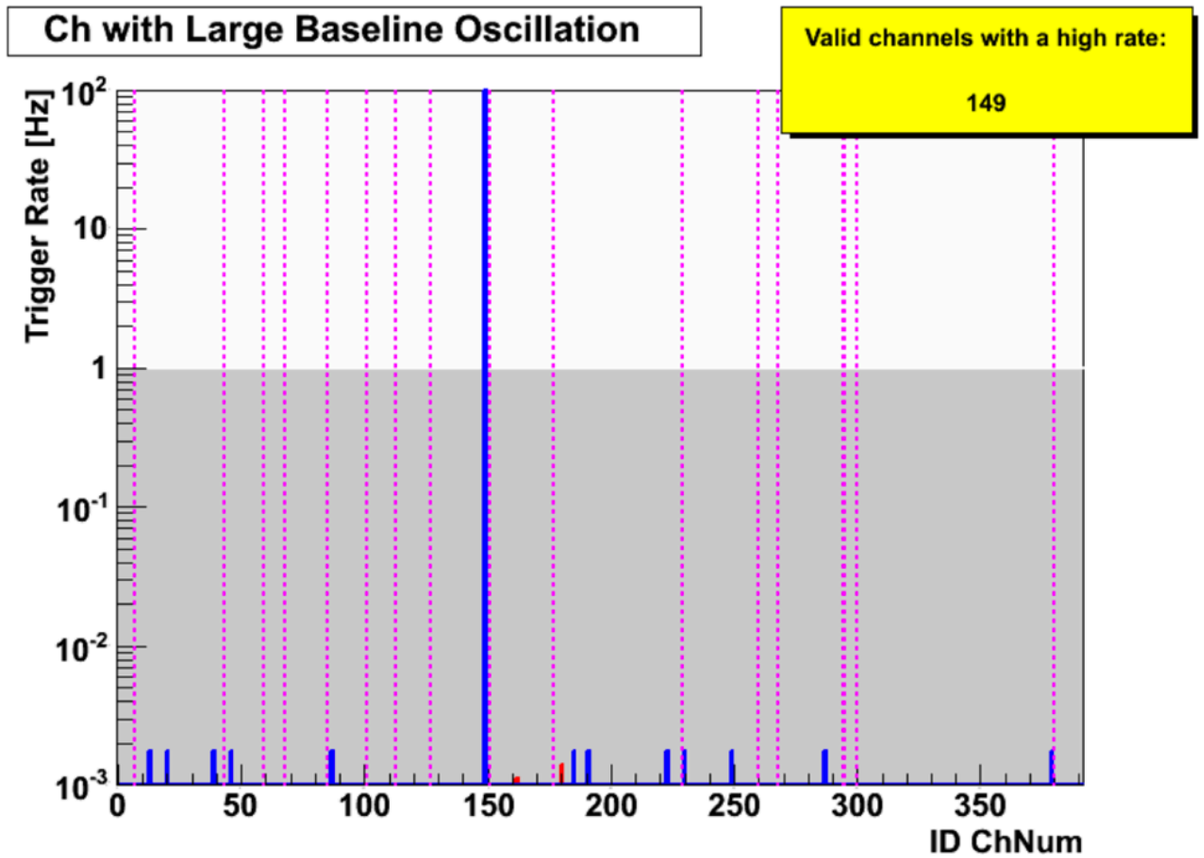
Histogram: _hChNum_BL_OscLarge_ID

Monitor high frequency oscillation with a large amplitude.
Period 1 ~ 10 ns, amplitude > 5 ADC count.
Histogram shows oscillation rate (y-axis) per ID channel (x-axis).
Dotted lines indicate known channels that does not carry physics signal.
Ignore a problem associated with the channels indicated by a dotted line.
Report if the oscillation rate is > 1 Hz for any channel.

Status Check: fine

Comment (leave empty if nothing):

Plots From Subject Run



Data Processing Status Query

DOGSifier Monitoring	Common Trunk Monitoring	Software Monitoring	DataChallenge Monitoring
Run Information	Production ROOT Files	DATA Stream Monitoring	DC Web

Double Chooz Homepage

This page was last updated at Tue Aug 27 16:54:12 CEST 2013
PHP version 5.3.8

Data Stream Information Query

Display Run Range

Show Runs Below: 71639
for # runs: 10
Search By Date: date
Refresh Period: 60

Filter Option

Filter PHYSICS Runs
Filter CALIB Runs
Filter TEST Runs
Show ERROR Flags
Show FAILED Flags

Display Option

Bin File Transfer @ DCFar1 => DCFnas
Bin File Deletion @ DCFar1
DOGSification @ DCFar0
Bin File Transfer @ DCFNas => IN2P3 (IRODS)
Bin File Deletion @ DCFnas
DOGSifier ROOT File Transfer @ DCFNas => IN2P3 (IRODS)
DOGSifier ROOT File Deletion @ DCFnas
OV File Transfer @ DCFNas => IN2P3 (IRODS)
OV File Deletion @ Onsite
DOGSifier ROOT File Transfer @ IN2P3 (iRODS=>HPSS)
CT Process @ IN2P3
CT ROOT File Transfer @ IN2P3 (iRODS=>HPSS)

Submit Reset

DATA Processing Stream

Run	Bin File Transfer 2 DCFnas	Bin File Deletion @ DCFar1	DOGSification @ DCFar0	Bin File Transfer 2 IN2P3 (IRODS)	Bin File Deletion @ DCFnas	DOGSifier ROOT File Transfer 2 IN2P3 (IRODS)	DOGSifier ROOT File Deletion @ DCFnas	OV File Transfer 2 IN2P3 (IRODS)	OV File Deletion @ Onsite	DOGSifier ROOT File Transfer @ IN2P3(iRODS=>HPSS)	CT Processing @ IN2P3	CT ROOT File Transfer @ IN2P3 (iRODS=>HPSS)
71639	COMPLETED 2013:08:27 10:41:36 33.99 [s]	COMPLETED 2013:08:27 10:43:03 0.59 [s]	COMPLETED 2013:08:27 11:02:05 1099.62 [s]	COMPLETED 2013:08:27 10:43:02 20.73 [s]	NOT YET	COMPLETED 2013:08:27 11:03:47 56.65 [s]	NOT YET	NOT YET	NOT YET	PROCESSING 2013-08-27 11:03:16	NOT YET	NOT YET
71638	COMPLETED 2013:08:27 10:40:03 0.15 [s]	COMPLETED 2013:08:27 10:42:03 0.04 [s]	COMPLETED 2013:08:27 10:42:47 1.14 [s]	COMPLETED 2013:08:27 10:40:58 0.84 [s]	NOT YET	FAILED NO_FILES_ERROR 2013:08:27 10:44:46 0 DataMigro Log	NOT YET	NOT YET	NOT YET	NOT YET	NOT YET	NOT YET
71637	COMPLETED 2013:08:27 10:39:03 0.32 [s]	COMPLETED 2013:08:27 10:41:02 0.11 [s]	FAILED PROCESSING_ERROR 2013:08:27 10:42:06 3.25 DataMigro Log	COMPLETED 2013:08:27 10:39:59 2.26 [s]	NOT YET	NOT YET	NOT YET	COMPLETED 2013:08:17 18:19:50 166.79 [s]	COMPLETED 2013:08:17 18:19:51 0.16 [s]	NOT YET	NOT YET	NOT YET

Resource Usage Monitoring

DOGSifier Monitoring	Common Trunk Monitoring	Software Monitoring	DataChallenge Monitoring
Run Information	Production ROOT Files	DATA Stream Monitoring	DC Web

Double Chooz Homepage

This page was last updated at Tue Aug 27 16:51:13 CEST 2013
PHP version 5.3.8

CommonTrunk Monitoring Web

This website is for shifters to monitor CommonTrunk Processing Status.
And also for DC collaborators to check the CT production progress.

Production Statistics Summary

From left: Total DAQ runs, CT-ed runs, and subset of CT-ed runs recommended for analysis.

	ALL Runs	CT-ed Runs	GOOD CT-ed Runs
# Runs	55512	42214	42144
Run Length [day]	693.88888888889	589.39236111111	587.88541666667

Total disk space usage history on iRODS disk server.
Note that the disk is purged once the available quota is filled.
The numbers shown is total file size written in the past.

	Total Nu-DOGSifier Filesize	Total CT Filesize	Total ROVER Filesize
Space Used [TB]	299.171	70.8919	12.8252

CT Production Detailed Summary

# TO-BE CT-ed	# BEING CT-ed	# FILE-XFER QUEUED	# FILE-XFER DONE
2980	0	0	42214

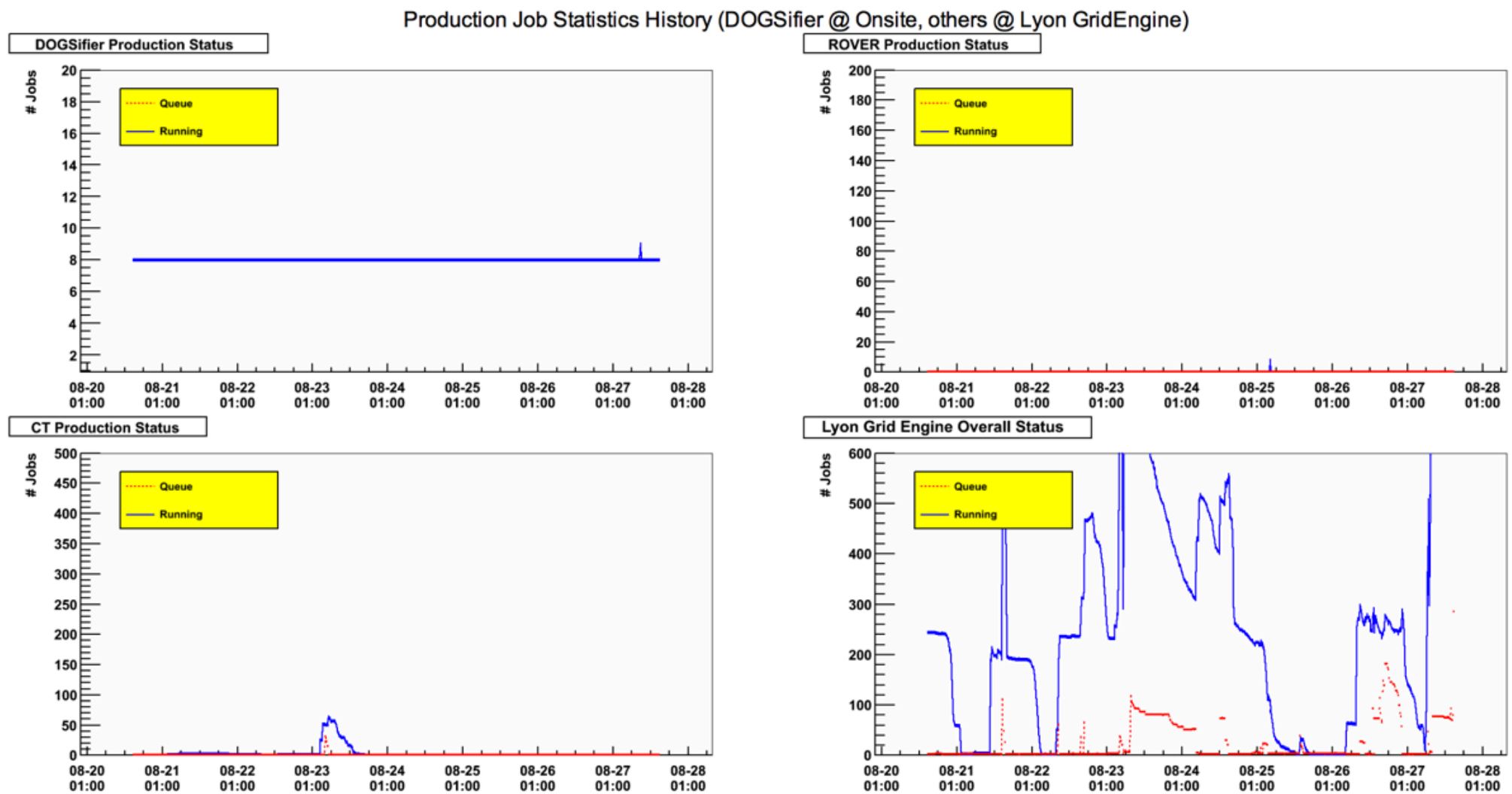
- "TO-BE CT-ed" ... # runs ready to be submitted for CT production.
- "BEING CT-ed" ... # runs currently running as CT production.
- "FILE-XFER QUEUED" ... # runs CT finished and files are on IRODS. Queued to be sent to HPSS.
- "FILE-XFER DONE" ... # runs CT finished and files x-fered to HPSS (ready for analysis access).

CT INPUT ERROR	CT PROCESS ERROR	CT OUTPUT ERROR	FILE-XFER DELAY
302	7	0	0

- "CT INPUT ERROR" ... # runs for which input file(s) missing on IRODS/HPSS.
- "CT PROCESS ERROR" ... # runs in failed in an unexpected manner.
- "CT OUTPUT ERROR" ... # runs CT otuput file transfer to IRODS failed (maybe recoverable)

History Monitoring (grid resource)

Production Process



LOG TIME	DOGSifier Queue	DOGSifier Process	CT Queue	CT Process	ROVER Queue	ROVER Process	GE All Queue	GE All Process
2013-08-27 16:46:14	126	8	0	0	0	0	285	636

Summary

- Example data process & physics monitoring scheme from DC
- Can we implement this for the UBooNE?
  - In principle, yes.
  - What part of this can be useful?
    - I believe some or many similar functions already exists?
  - Requirements include...
    - database server (MySQL would be great for my taste)
    - web server to run cgi scripts
    - permission/knowledge to launch/maintain daemon jobs
    - an advisor knowledgeable about Fermi lab cluster usage