

Offline Computing (Tier-0) Monitoring

Jaroslav GUENTHER (CERN)

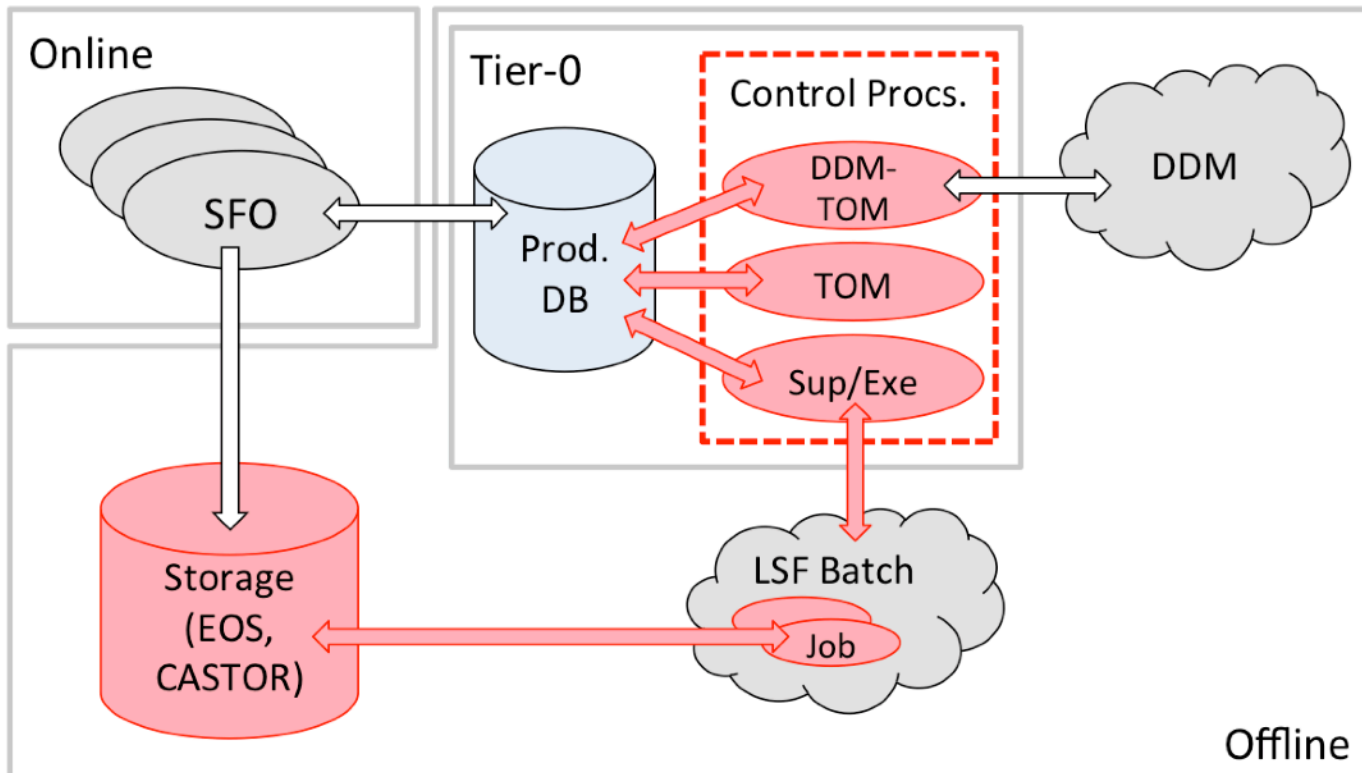
Armin NAIRZ (CERN)

Data Quality Shift Training, 29th March 2018

Offline Computing (Tier-0) Monitoring

(Some) Tier-0 responsibilities:

- Archival of RAW data from the SFOs
- Management, orchestration, execution of first-pass processing
- Registration of all data products, preparation for export to Tier-1 centres



Offline Computing (Tier-0) Monitoring

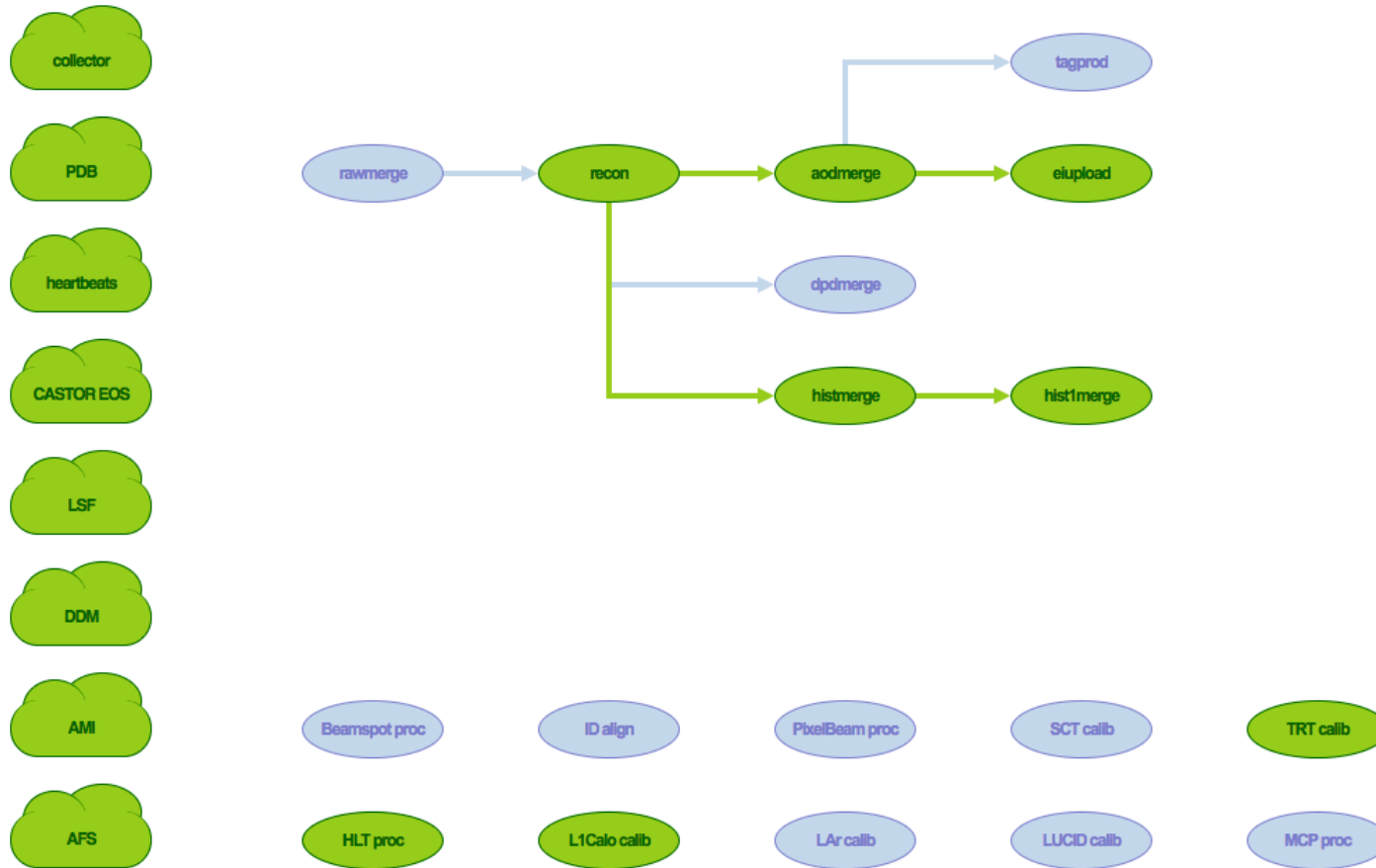
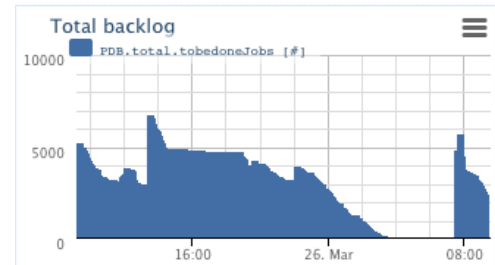
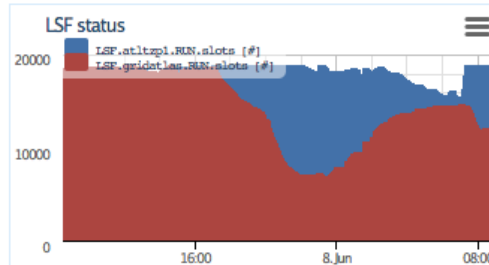
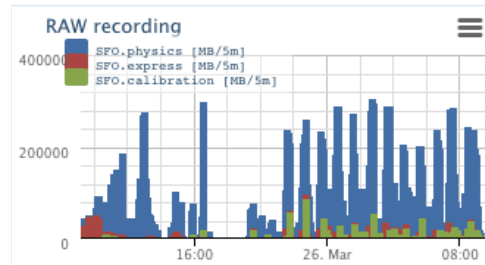
The DQ shifter is asked to monitor the basic functionality of the Tier-0 and offline infrastructure by using one of the web pages

- <https://tzcontzole01.cern.ch/run2/monitor/>
- <https://tzcontzole02.cern.ch/run2/monitor/>
- Two redundant, equivalent implementations

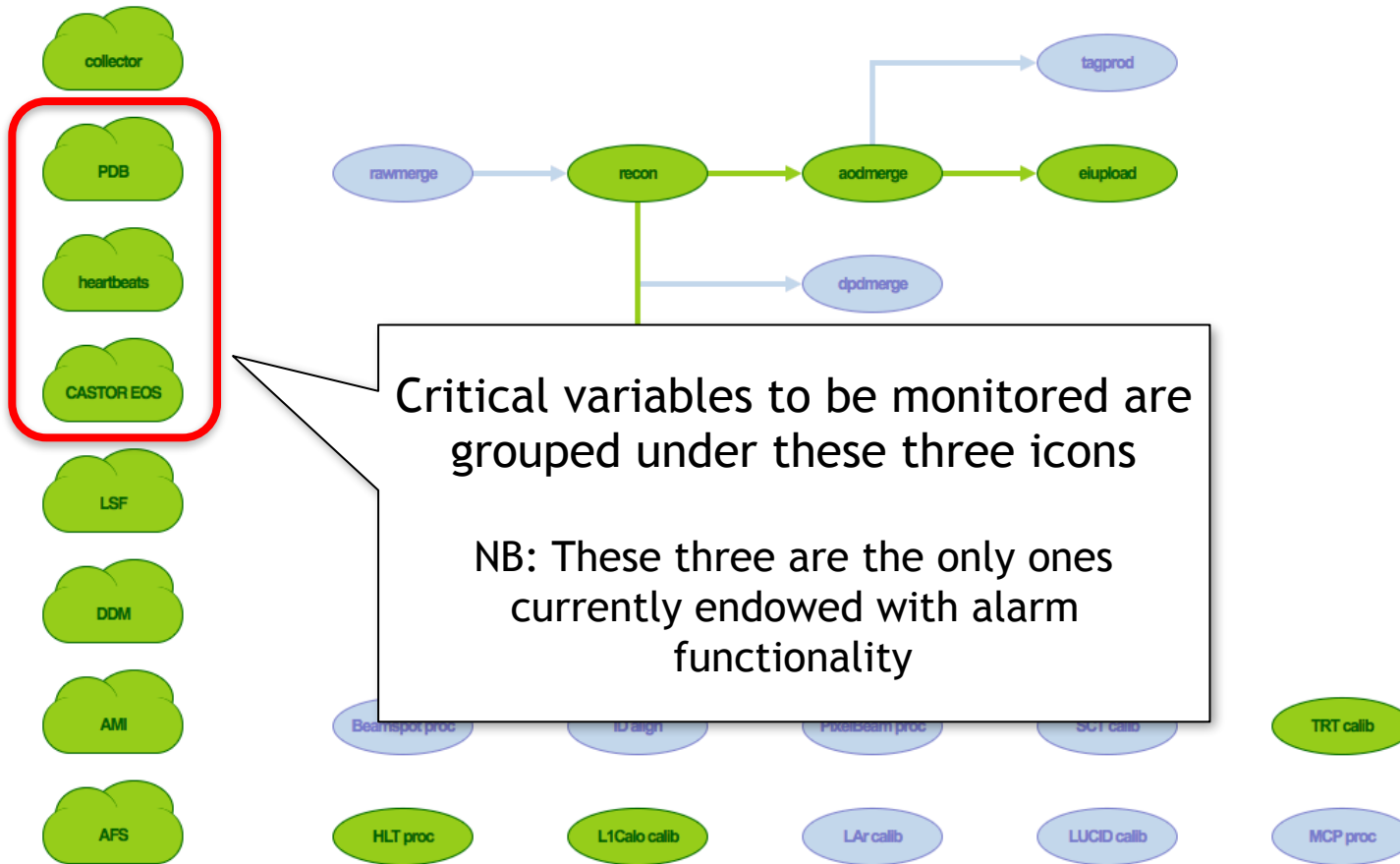
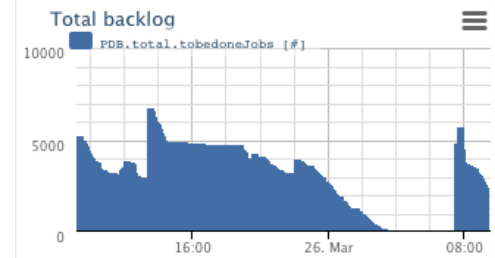
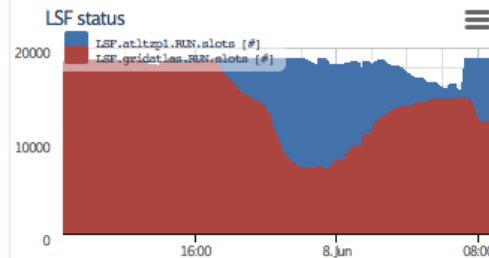
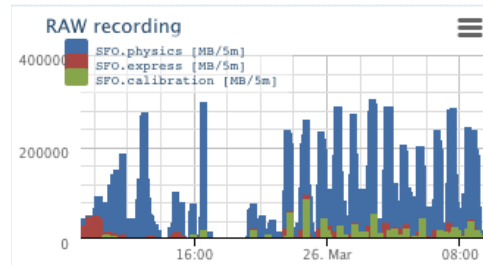
The pages are accessible from the ACR DQ shift desk

- To be opened in browser (bookmarked)
- To be watched in addition to other DQ monitoring applications

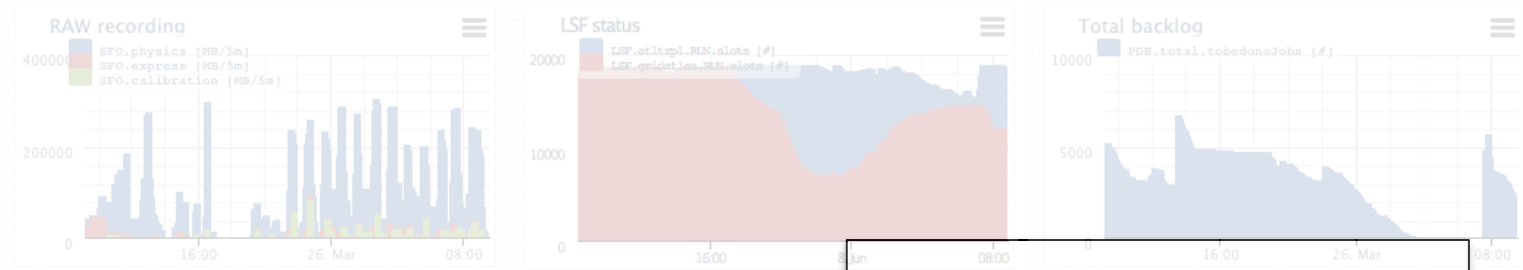
Monitoring Page



Monitoring Page



Monitoring Page

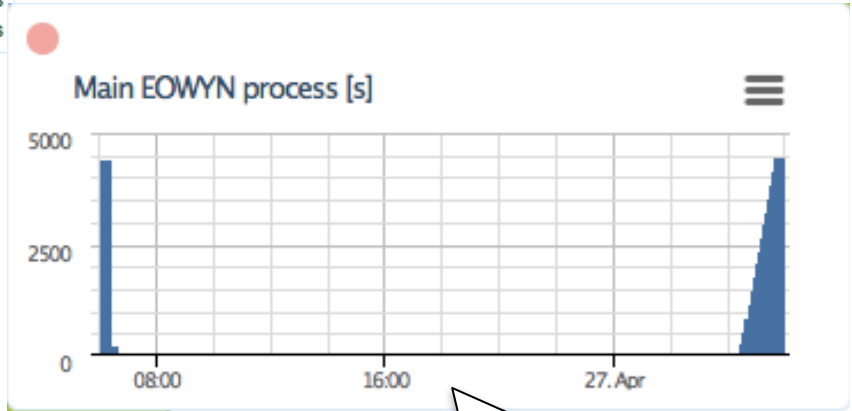
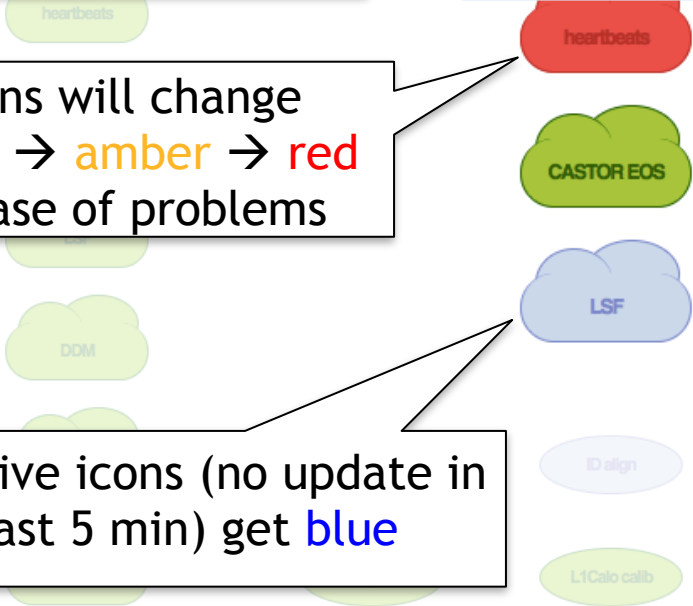


Icons are clickable, pop-up window shows the list of monitored variables

A green cloud icon labeled 'collector' is at the top. Below it is a red heart icon labeled 'heartbeats'. A pop-up window for 'heartbeats' is shown, containing the following text: 'id: heartbeats', 'last updated: 07:47', 'Main EOWYN process 4432 s', 'Main TOM process 24 s', and 'DDM TOM process 3 s'.

Problematic variables are shown in red

Icons will change green → amber → red in case of problems



Clicking on variables displays the corresponding graph

Inactive icons (no update in last 5 min) get blue

General Instructions

Please have a regular look at the monitoring page
every 60 minutes

The page updates automatically every 5 minutes, but still
make sure that it is up-to-date

- Application in the browser can get stuck
- CERN SSO credentials may expire
- Watch the time labels on the graphs or the last-update timestamp on the pop-up windows
 - NB: time zone on the graphs is GMT (-1h)
- If necessary, refresh the browser window

In a case of emergency (cf. following slides)
call the Tier-0 expert phone

16 1928



Uptime of Tier-0 Processes

There are three critical Tier-0 “control processes”

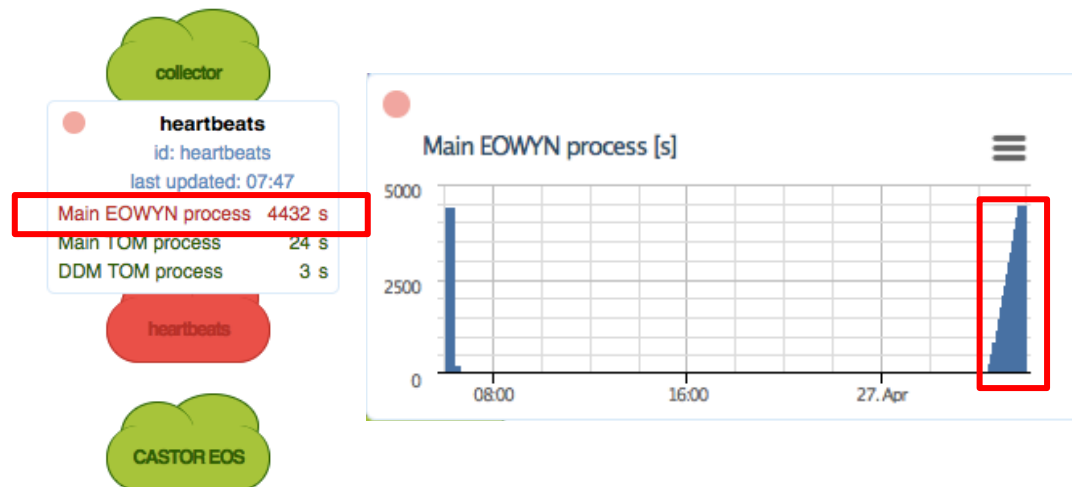
- Supervisor & Executor (“Eowyn”): responsible for submission/running of jobs, associated bookkeeping
- Tier-0 Manager (“TOM”): responsible for handshake with SFO database (at arrival of new data), definition of tasks, jobs, i/o datasets
- A separate TOM instance (“DDM TOM”) responsible for data registration in DDM, AMI, data replication inside CERN (CAF)

The processes send regular “heartbeats”

Time since the last heartbeat is monitored

Alarm if no heartbeat for >30 minutes

- Process may have stopped and need restarting



Action:
call Tier-0 expert

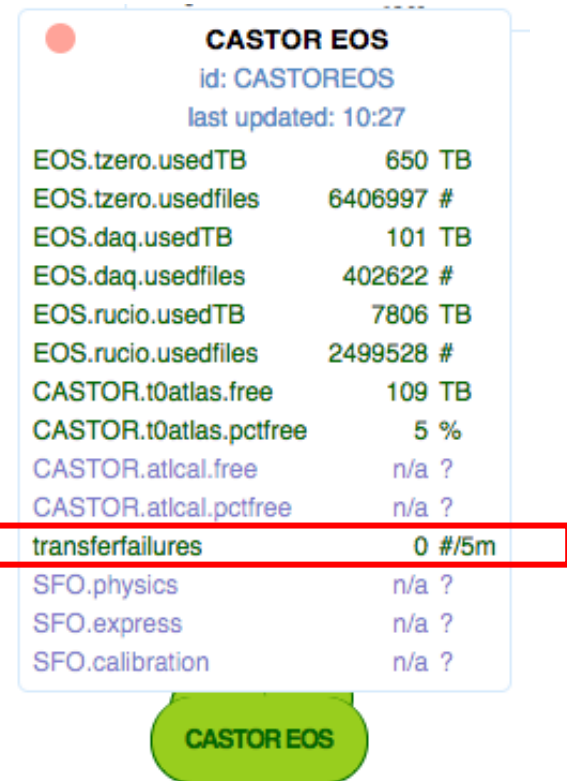
Transfer Errors

Tier-0 system logs failed transfers from/into storage(CASTOR, EOS)
– Indirect monitoring of storage service status

There can be glitches leading to transient spikes in the failure rate

Extended periods may indicate service outage and become dangerous both for Tier-0 and DAQ

Action:
if alarm situation persists
for >60 minutes,
call Tier-0 expert



CASTOR EOS
id: CASTOREOS
last updated: 10:27

EOS.tzero.usedTB	650 TB
EOS.tzero.usedfiles	6406997 #
EOS.daq.usedTB	101 TB
EOS.daq.usedfiles	402622 #
EOS.rucio.usedTB	7806 TB
EOS.rucio.usedfiles	2499528 #
CASTOR.t0atlas.free	109 TB
CASTOR.t0atlas.pctfree	5 %
CASTOR.atlcal.free	n/a ?
CASTOR.atlcal.pctfree	n/a ?
transferfailures	0 #/5m
SFO.physics	n/a ?
SFO.express	n/a ?
SFO.calibration	n/a ?

CASTOR EOS

Job Failures

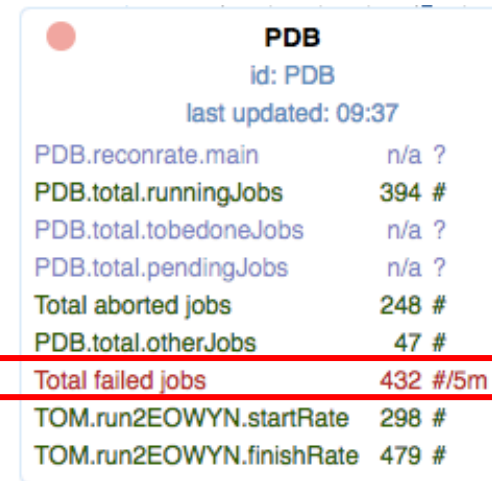
Jobs may fail due to many reasons

- Relevant here are problems of the computing infrastructure
- Software-related issues fall into other domain

There can be glitches leading to transient spikes in the failure rate

Extended periods may indicate batch service outage and result in processing backlog at Tier-0

Action:
if alarm situation persists
for >60 minutes,
call Tier-0 expert



The screenshot shows a status page for 'PDB' (Production DB). It includes a red alarm indicator, the ID 'PDB', and the last update time '09:37'. A table lists various metrics, with 'Total failed jobs' highlighted in a red box.

PDB	
id: PDB	
last updated: 09:37	
PDB.reconrate.main	n/a ?
PDB.total.runningJobs	394 #
PDB.total.tobedoneJobs	n/a ?
PDB.total.pendingJobs	n/a ?
Total aborted jobs	248 #
PDB.total.otherJobs	47 #
Total failed jobs	432 #/5m
TOM.run2EOWYN.startRate	298 #
TOM.run2EOWYN.finishRate	479 #

“PDB” =
Production DB

PDB

