



WLCG Service Report

Andrea.Valassi@cern.ch
(for the SCOD team)

~ ~ ~

WLCG Management Board, 22nd January 2013

Thanks to Maria Dimou, Mike Kenyon, David Tuckett, Eddie Karavakis for the GGUS and reliability reports

Summary (1)

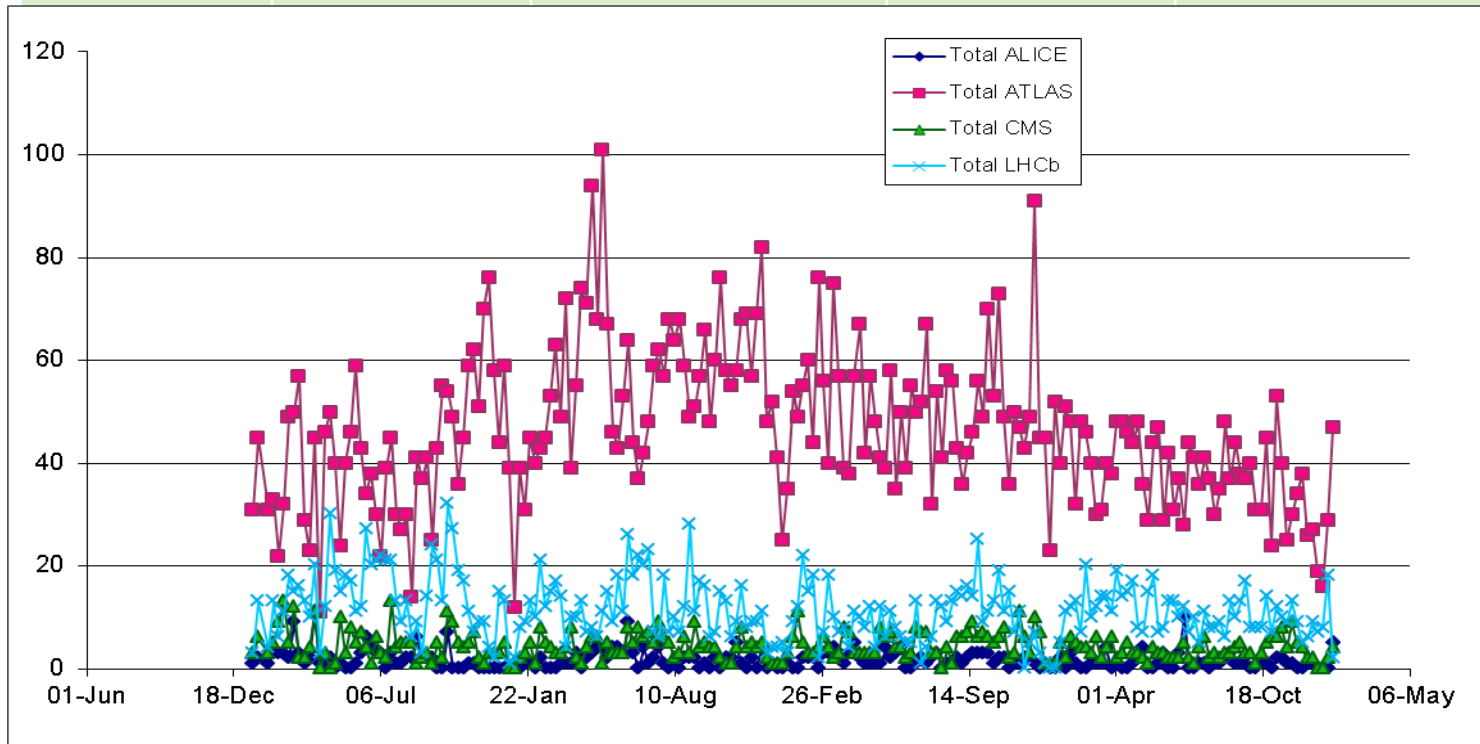
- 9 weeks since the last MB report on 2012 November 20th
 - Including the two week-long Christmas break at CERN
 - Smooth LHC operations (proton-proton in 2012, now proton-ion)
- Four Service Incident Reports received:
 - LHCb file loss at PIC during tape migration/repack (Oct 20: [SIR](#))
 - RAL power incident due to UPS failure (Nov 7: [SIR](#))
 - RAL power incident due to human error during work on UPS (Nov 20: [SIR](#))
 - CERN CA CRL expired (Jan 19: [SIR](#))
- 8 real GGUS ALARMS, all at CERN – details in the next pages
 - 6 for ATLAS, 2 for CMS (3 about storage, 3 about CA CRL)

Summary (2)

- Many other issues reported at the daily meetings, most notably:
 - *IN2P3 dCache freezes and needs to be restarted due to SRM bug with long proxies ([GGUS:88984](#)), affecting ATLAS and CMS – Nov through Jan, no fix yet*
 - KIT tape system overloaded, affecting ATLAS and CMS
 - Wrong information in BDII queues, affecting LHCb
 - FTS transfer failures to Gridka for LHCb due to DB problems ([GGUS:88906](#))
 - CERN EOS quotas increased for ALICE and CMS
 - Network issues in CNAF to FNAL transfers for CMS ([GGUS:88752](#))
 - ATLAS job failures due to insufficient space in Taiwan ([GGUS:89253](#))
 - CMS HammerCloud test jobs failing on CERN CREAM CEs ([GGUS:89124](#))
 - SRM instabilities/overload at PIC affecting LHCb ([GGUS:89664](#))
 - Network outages/glitches at KIT, RAL, CERN, NDGF
 - PIC reducing CPU power to 70% until February due to electricity cost in winter
 - Several upgrades and maintenance operations, especially in December
 - LHCb tests of online farm usage for offline reprocessing over Christmas
 - LFC schema upgrade at CERN after Christmas (EMI-2 upgrade) – ongoing
 - CMS T0 changes (ProdAgent/WMAgent, streamer files) during proton-ion run

GGUS summary (9 weeks)

| VO | User | Team | Alarm | Total |
|--------|------|------|-------|-------|
| ALICE | 2 | 6 | 0 | 8 |
| ATLAS | 26 | 234 | 6 | 266 |
| CMS | 19 | 7 | 2 | 28 |
| LHCb | 7 | 70 | 0 | 77 |
| Totals | 54 | 317 | 8 | 379 |



Support-related events since last MB

- **There have been 8 real ALARMS since the 2012/11/20 MB.**
- **2 were submitted by CMS and 6 by ATLAS. 6 of these ALARMS were submitted during the weekend.**
- **All concerned the CERN site.**
- **2 GGUS Releases took place since the last MB, on 2012/11/28 & 2012/12/12. All ALARM tests were successful (operators received notification, reacted within minutes, interfaces worked, experts closed promptly).**

Support-related events since last MB

II

- The issue of CERN CA CRL expiration of Saturday evening 2013/01/19 had a big impact. 6 GGUS tickets (of which 3 are ALARMS) were opened on this issue and its side-effects. The security and middleware evolution groups in the WLCG operations' area will discuss the CRL update frequency policy and the clarity of the error messages.
- On EGI request we shall hold a meeting on alternative ways (without certificates) to authenticate to GGUS and other Grid applications on 2013/02/05 at 14:30 CET in CERN room 28-1-025 with remote tel. connectivity arranged. Details in <https://savannah.cern.ch/support/?132872> and (still draft) agenda here. Today we offer a username and passwd login but we require the registration of the user's DN.

ATLAS ALARM->CERN AFS REL.

AREA INACCESSIBLE GGUS:88856

| What time UTC | What happened |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2012/11/25 16:31 SUNDAY | GGUS ALARM ticket opened, automatic email notification to atlas-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem: File Access. |
| 2012/11/25 16:40 | Operator records in the ticket that AFS piquet is working on the issue. |
| 2012/11/25 17:21 | Supporter explains that the <i>physical afs server had a hard disk issue, solved by a reboot.</i> |
| 2012/11/25 17:59 | Submitter confirms service quality is improving. |
| 2012/11/26 08:14 MONDAY | Another ATLAS supporter reports atlas.web.cern.ch problems in the same ticket, because the site is afs-hosted. |
| 2012/11/26 09:59 | Ticket 'solved' after exchange of 8 comments, where afs experts insisted to distinguish the ATLAS afs file access problems from the web ones - hosts involved afs154.cern.ch vs afs140.cern.ch. |

ATLAS ALARM->CERN CASTOR FILE EXPORT PROBLEMS GGUS:89107

| What time UTC | What happened |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2012/12/02 00:17 SUNDAY | GGUS ALARM ticket opened, automatic email notification to atlas-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem: File Access. |
| 2012/12/02 00:26 | Expert confirms in the ticket that investigation started. |
| 2012/12/02 00:28 | Operator records in the ticket that CASTOR piquet was called. |
| 2012/12/02 01:15 | Expert puts the ticket to status <i>'solved' after identifying a problem on the node where the files reside and rebooting it</i> (4 comments exchanged). |
| 2012/12/02 05:49 MONDAY | Ticket 're-opened' because 1 of the 2 files was not transferred. |
| 2012/12/02 11:18 | Ticket 'solved' after exchange of 6 comments and migration of the file-to-transfer to a more stable machine. |

CMS ALARM->CERN SRM

UNREACHABLE GGUS:89186

| What time UTC | What happened |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2012/12/04 14:27 | GGUS ALARM ticket opened, automatic email notification to cms-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem: Storage Systems. |
| 2012/12/04 14:27 | Expert confirms in the ticket that investigation started. |
| 2012/12/04 14:38 | Operator records in the ticket that CASTOR piquet was called. |
| 2012/12/04 14:42 | Expert and submitter agree the symptoms lasted for 1.5 hrs and disappeared, also from SLS. |
| 2012/12/04 16:22 | Ticket 'solved' and cause fully understood. <i>A bug was revealed by a chain of srmAbort & srmReleaseFiles requests. The patch will be applied</i> during an agreed quiet LHC operations' period. |
| 2012/12/04 17:09 | Ticket 'verified' by the submitter, although he doubts the problem is really solved (he wrote that transfer errors still persisted). |

ATLAS ALARM->CERN SLOW LSF

GGUS:89202

| What time UTC | What happened |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2012/12/04 23:46 | GGUS ALARM ticket opened, automatic email notification to atlas-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem: Local Batch Systems. |
| 2012/12/04 23:50 | Submitter attaches to the ticket plots showing number of pending jobs & bsub average time. |
| 2012/12/05 00:02 | Operator records in the ticket that Local Batch System piquet was called. |
| 2012/12/05 00:06 Time is always UTC! | Expert comments that the <i>reason for this slowness is a reconfiguration that started at 0:00 CET and took too long to finish.</i> |
| 2012/12/05 07:59 | Ticket 'solved' by the expert with a note for pending action: "we need to understand why the system took so long to reconfigure". |
| 2012/12/05 09:29 | Ticket 'verified' by the submitter → no way to get further updated with the answer to the above question on the slow reconfiguration. |

ATLAS ALARM->CERN WEB SERVER DOWN GGUS:89334

| What time UTC | What happened |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2012/12/08 09:02 SATURDAY | GGUS ALARM ticket opened, automatic email notification to atlas-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem: Monitoring. |
| 2012/12/08 09:13 | Operator records in the ticket that the responsible of the concerned host webafs10.cern.ch is checking. |
| 2012/12/08 09:32 | Service mgr asks if situation has improved and explains that reason is related to an intervention on the power infrastructure that took longer. |
| 2012/12/08 09:39 | Service mgr asks the submitter twice if the service is restored, then, having no answer, leaves the ticket with web 3 rd Line Support. |
| 2012/12/08 09:47 | Ticket 'solved' by the expert. <i>Reason was "power cut"</i> |
| 2012/12/09 18:49 | Ticket set to 'solved' again by the supporter doing these drills, as an acknowledgment by the submitter had caused a re-opening. 'verified' on 2012/12/11. |

ATLAS ALARM->CERN LFC

CONNECTION ERRORS GGUS:90602

| What time UTC | What happened |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2013/01/19 20:12 SATURDAY | GGUS ALARM ticket opened, automatic email notification to atlas-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem:Other. |
| 2013/01/19 20:18 | IT-ES-VOS expert refers to GGUS:90599 & GGUS:90600 which explain the problem, namely that the <i>CERN CA CRL appeared expired.</i> |
| 2013/01/19 20:23 | Operator records in the ticket that email was sent to the IT-PES-PS service. |
| 2013/01/19 20:29 | Another shifter shows that more than LFC suffers, e.g. voms-proxy fails. |
| 2013/01/19 20:40 | After 4 comments' exchange the ticket was 'solved' as soon as the CERN CA CRL was updated. T |

CMS ALARM->CERN CRL BLOCKS

ALL JOBS GGUS:90604

| What time UTC | What happened |
|------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2013/01/19 20:34 SATURDAY | GGUS ALARM ticket opened, automatic email notification to cms-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem: Middleware. |
| 2013/01/19 20:36 | Submitter pastes the voms-proxy-init VERY misleading error msg saying "certificate verify failed" which makes users think there is something wrong with THEIR OWN certificate. |
| 2013/01/19 20:41 | IT-ES-VOS expert fixed the afs UI and prompts the submitter to try again. At the same time, the operator records in the ticket that sms was sent to IT-PES-PS. |
| 2013/01/19 21:07 | Expert and submitter agree the problem is gone. <i>The fetch-crl cron job was run manually for the VO boxes and the ticket was set to 'solved'.</i> |

ATLAS ALARM->CERN SRM ALARM ON MONITOR GGUS:90605

| What time UTC | What happened |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2013/01/19 20:50 SATURDAY | GGUS ALARM ticket opened, automatic email notification to atlas-operator-alarm@cern.ch AND automatic assignment to ROC_CERN. Automatic SNOW ticket creation successful. Type of Problem: Storage systems. |
| 2013/01/19 21:05 | Operator records in the ticket that the CASTOR piquet was called. |
| 2013/01/19 21:19 | Service expert comments the problem is, indeed, due to <i>missing CRL update being done manually.</i> |
| 2013/01/19 21:33 | Problem set to 'solved' as the SLS rebecame green after the CRL update. |
| 2013/01/20 23:24 | The ticket was set to 'verified'. |

Analysis of the reliability plots : Week 19/11/2012

- All VOs

- *0.1: RAL (20-22/11/2012). Emergency shutdown due to electrical problem.*

- ATLAS

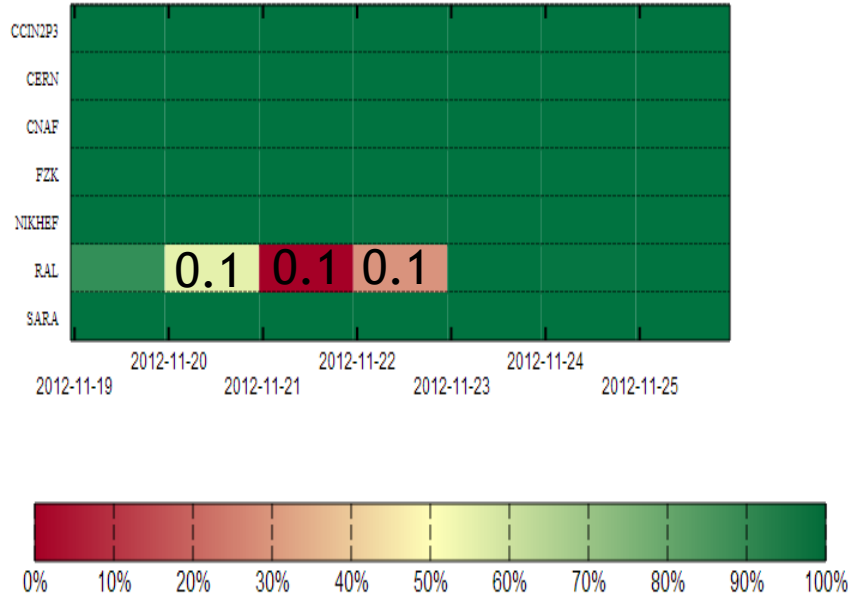
- 1.1: INFN (19 & 20/11/2012). SRM-put tests failing for storm-fe.cr endpoint's ATLASDATADISK space token. No downtime registered.
- NB. All sites exhibit reduced reliability on 23/11 as a result of an expired testing proxy.

- LHCb

- 2.1: GridKA (23 & 24/11/2012). SRM Del & LS tests failing for gridka-dCache.fzk.de endpoint. No downtime registered.

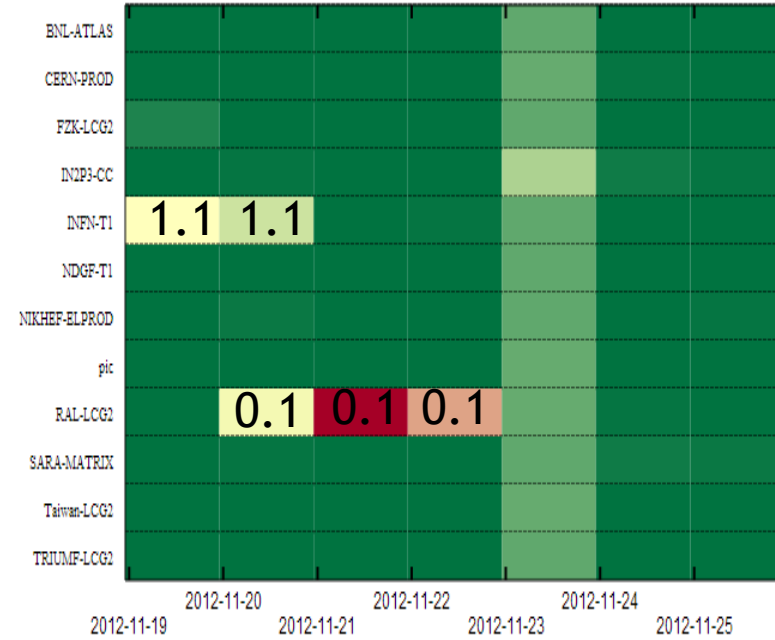
Site reliability using ALICE_CRITICAL

168 hours from 2012-11-19 00:00 to 2012-11-26 00:00



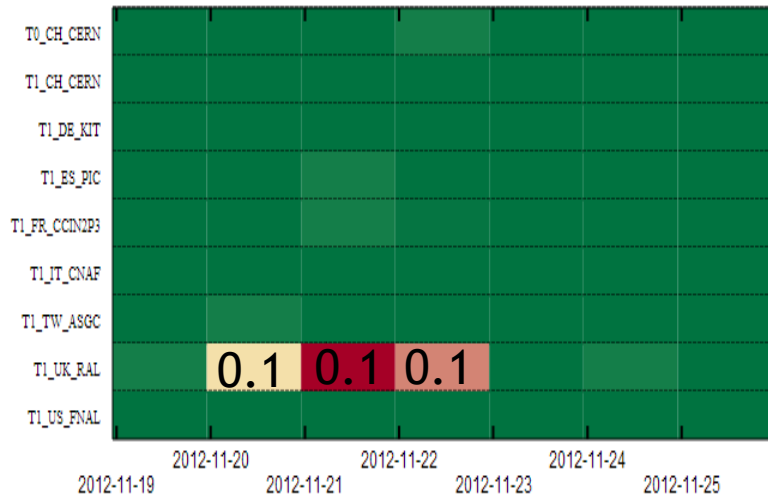
Site reliability using ATLAS_CRITICAL

168 hours from 2012-11-19 00:00 to 2012-11-26 00:00



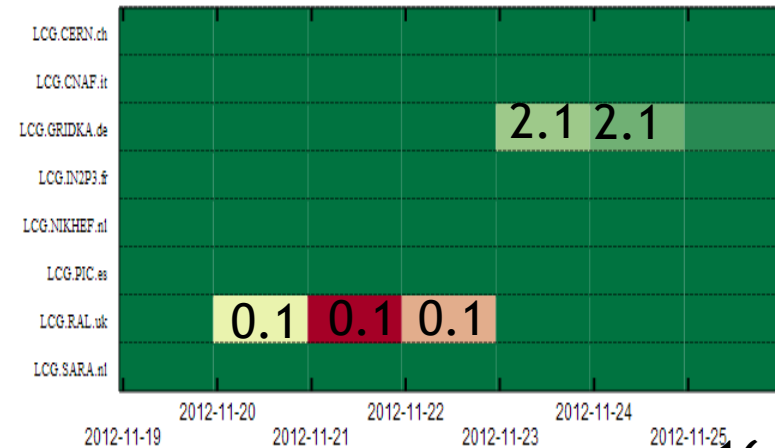
Site reliability using CMS_CRITICAL_FULL

168 hours from 2012-11-19 00:00 to 2012-11-26 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2012-11-19 00:00 to 2012-11-26 00:00

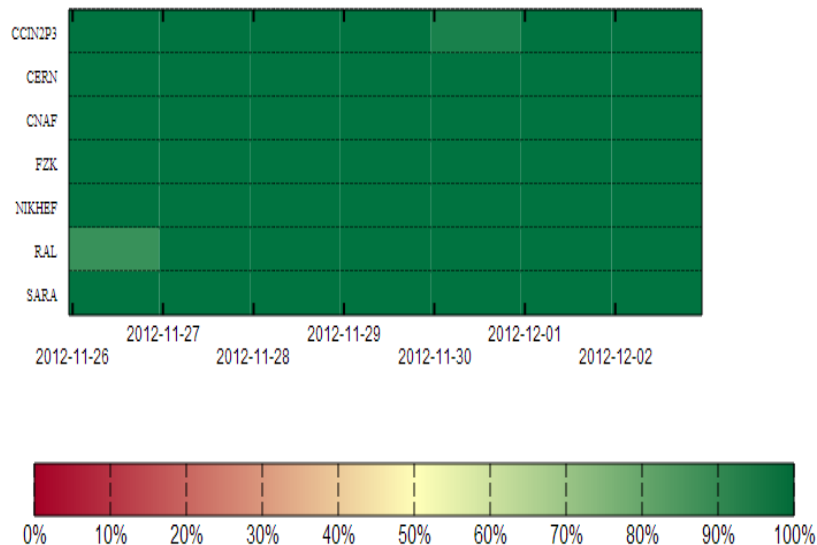


Analysis of the reliability plots : Week 26/11/2012

- ALICE: Nothing to report.
- ATLAS
 - 1.1: IN2P3 (26,27 & 30/11/2012). SRM-VOPut tests failing for ccsrm.in2p3 endpoint. No downtime registered.
 - 1.2: INFN (29/11/2012). SRM-VOPut test intermittently failing for storm-fe endpoint. No downtime registered.
- CMS
 - 2.1: RAL (29/11/2012). SRM-VOPut test failing for srm-cms endpoint. Related tickets: GGUS:[89003](#) and [89004](#). No downtime registered.
- LHCb: Nothing to report.

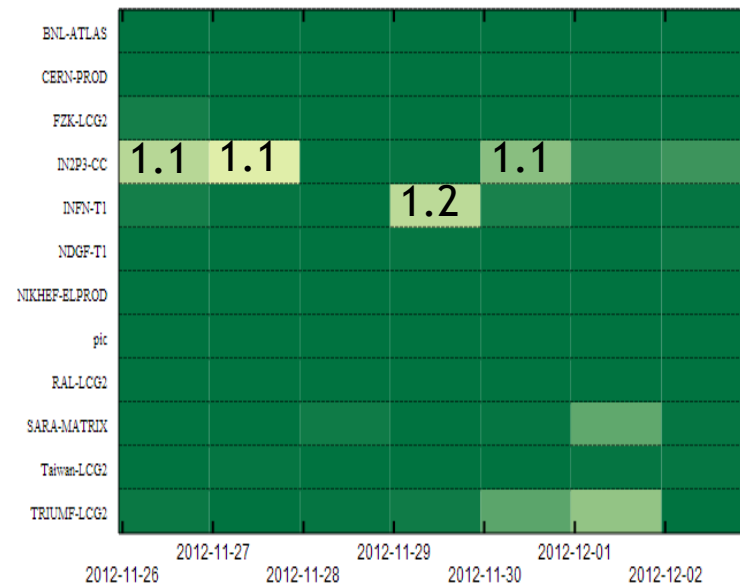
Site reliability using ALICE_CRITICAL

168 hours from 2012-11-26 00:00 to 2012-12-03 00:00



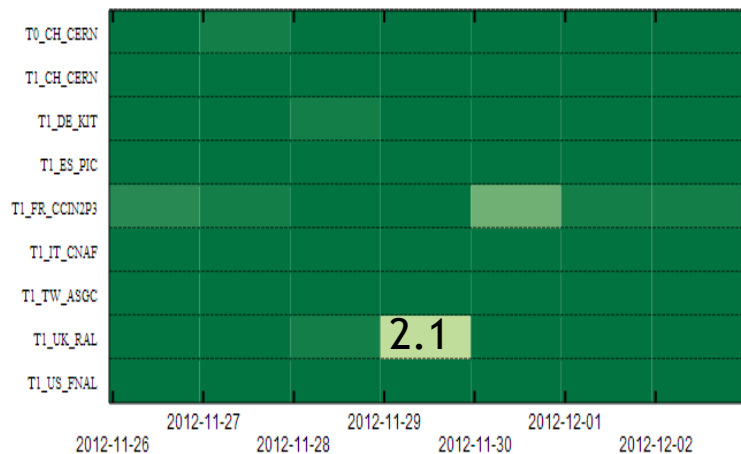
Site reliability using ATLAS_CRITICAL

168 hours from 2012-11-26 00:00 to 2012-12-03 00:00



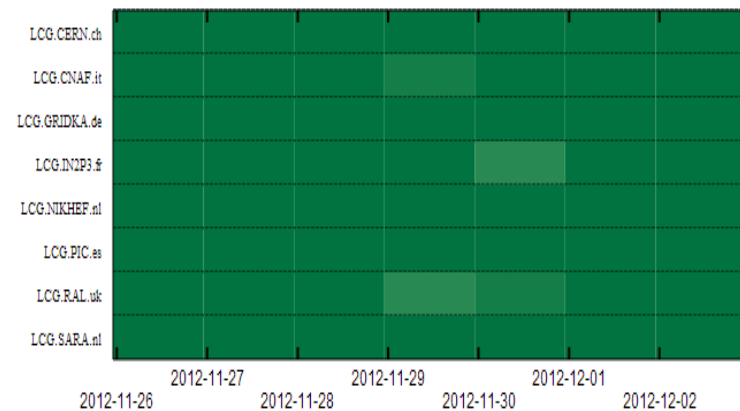
Site reliability using CMS_CRITICAL_FULL

168 hours from 2012-11-26 00:00 to 2012-12-03 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2012-11-26 00:00 to 2012-12-03 00:00

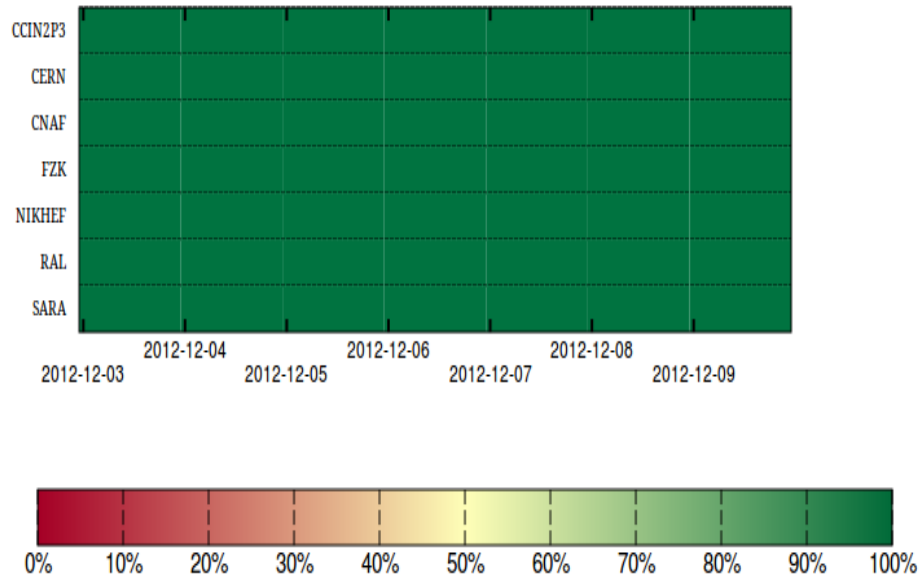


Analysis of the reliability plots : Week 03/12/2012

- ALICE: Nothing to report.
- ATLAS:
 - 2.1: SARA (04/12/2012). Monitoring Issue. Low priority CREAMCE-JobSubmit tests timing out due to high load of production jobs.
- CMS: Nothing to report.
- LHCb: Nothing to report.

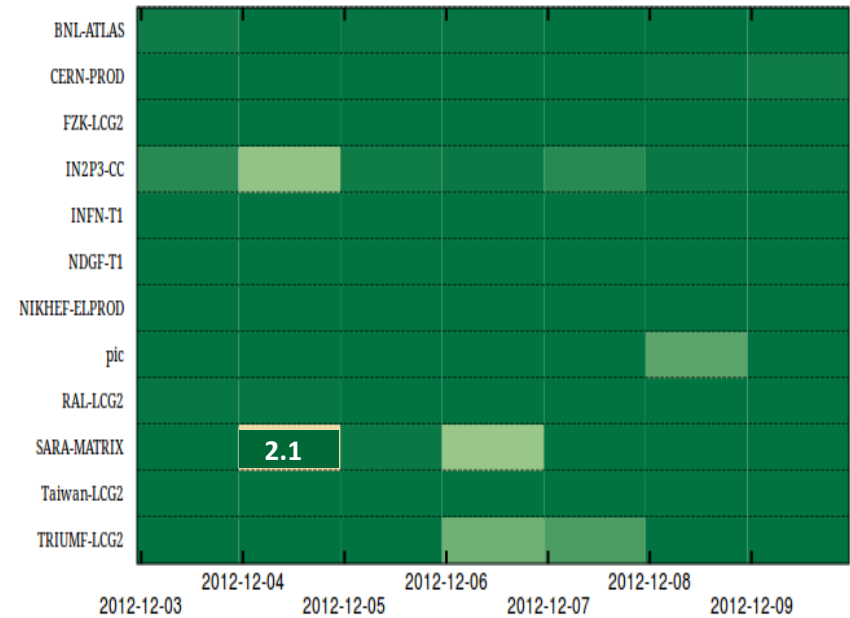
Site reliability using ALICE_CRITICAL

168 hours from 2012-12-03 00:00 to 2012-12-10 00:00



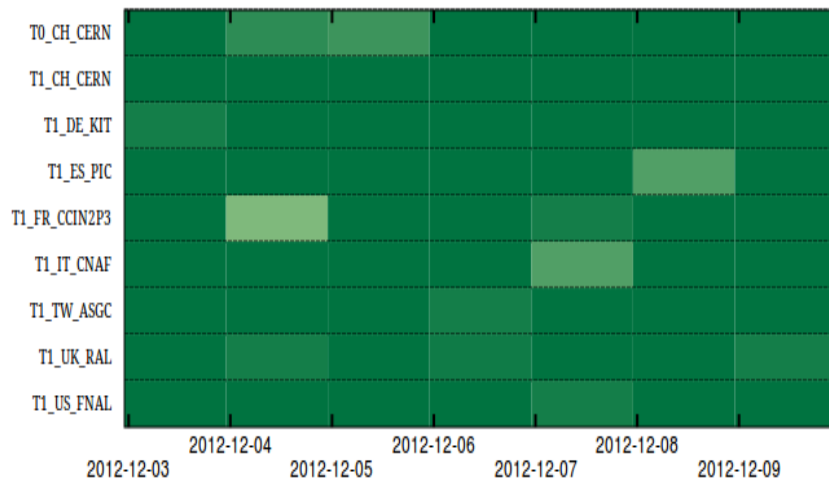
Site reliability using ATLAS_CRITICAL

168 hours from 2012-12-03 00:00 to 2012-12-10 00:00



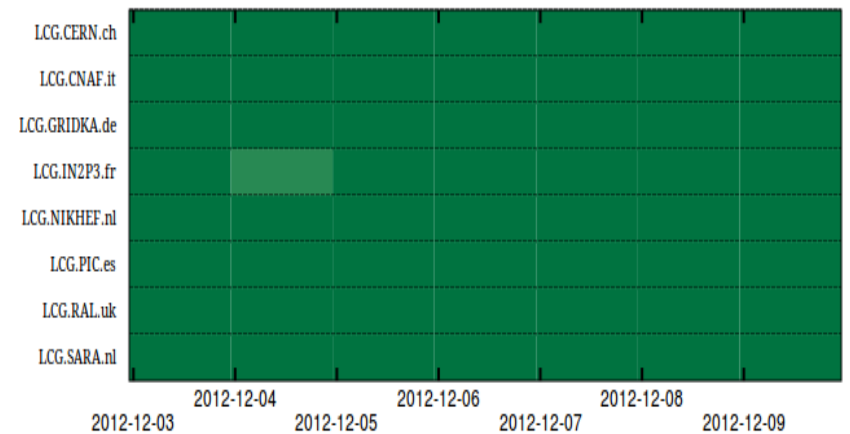
Site reliability using CMS_CRITICAL_FULL

168 hours from 2012-12-03 00:00 to 2012-12-10 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2012-12-03 00:00 to 2012-12-10 00:00



Analysis of the reliability plots : Week 10/12/2012

- ALL:

- *0.1: IN2P3 (10/12/2012 – 12/12/2012). Scheduled downtime "Installation of a second power supply line on site". [GOC:118889](#)*

- ALICE & ATLAS:

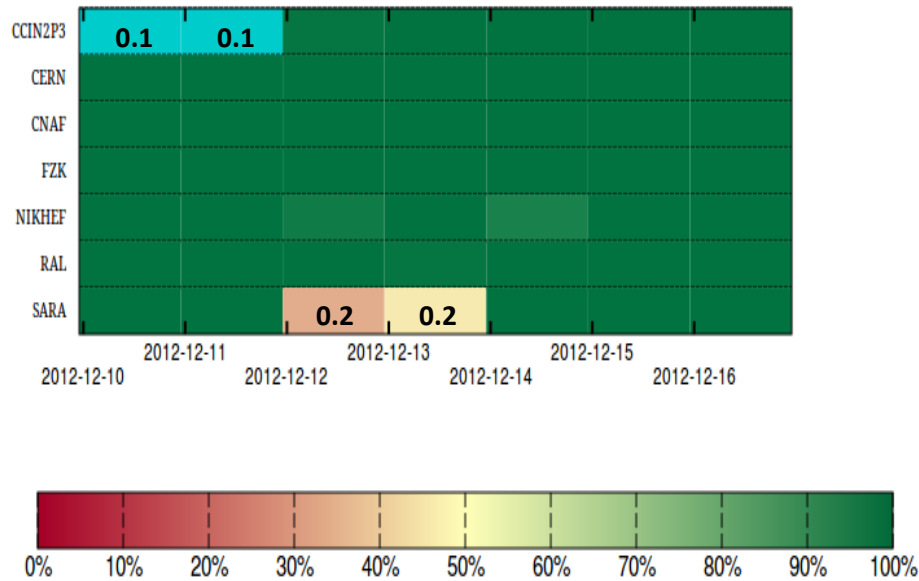
- 0.2: SARA (12/12/2012 – 13/12/2012). CREAMCE-JobSubmit tests failing on creamce.gina.sara.nl. Solved "misconfiguration of the /etc/sudoers of the creamce". [GGUS:89597](#)

- CMS: Nothing to add.

- LHCb: Nothing to add.

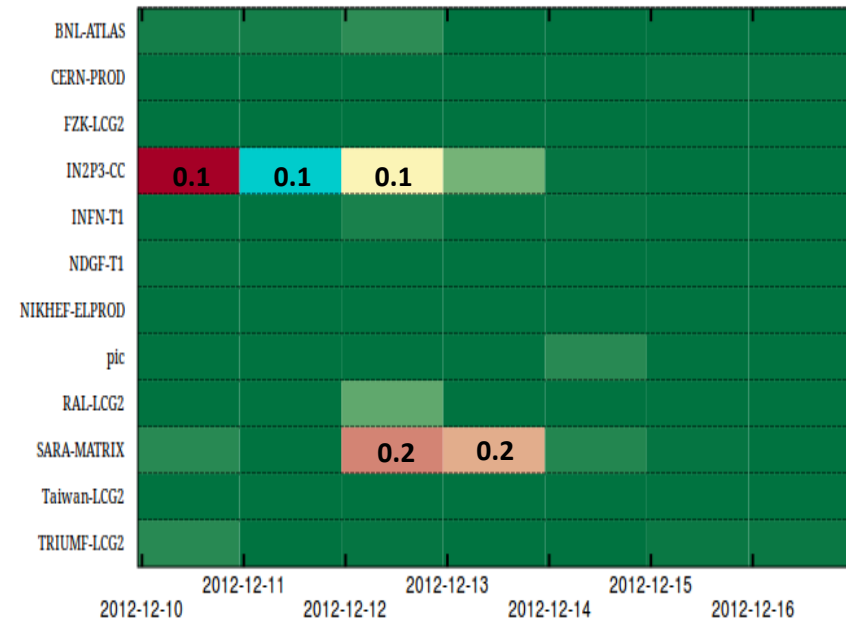
Site reliability using ALICE_CRITICAL

168 hours from 2012-12-10 00:00 to 2012-12-17 00:00



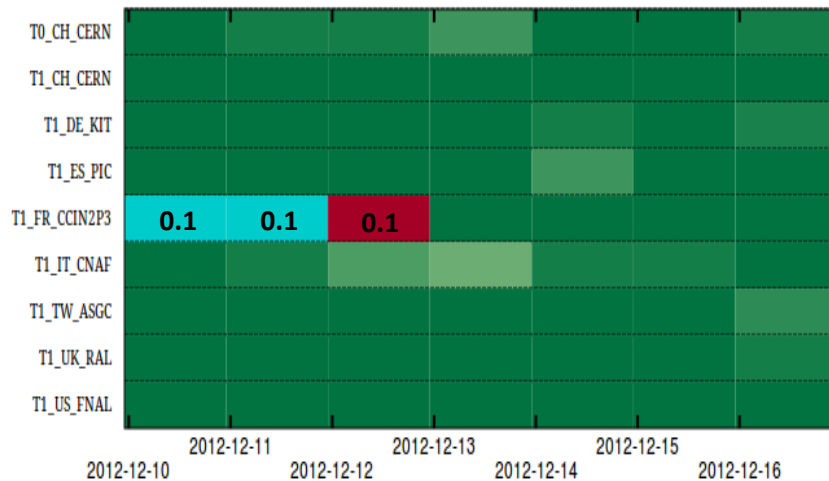
Site reliability using ATLAS_CRITICAL

168 hours from 2012-12-10 00:00 to 2012-12-17 00:00



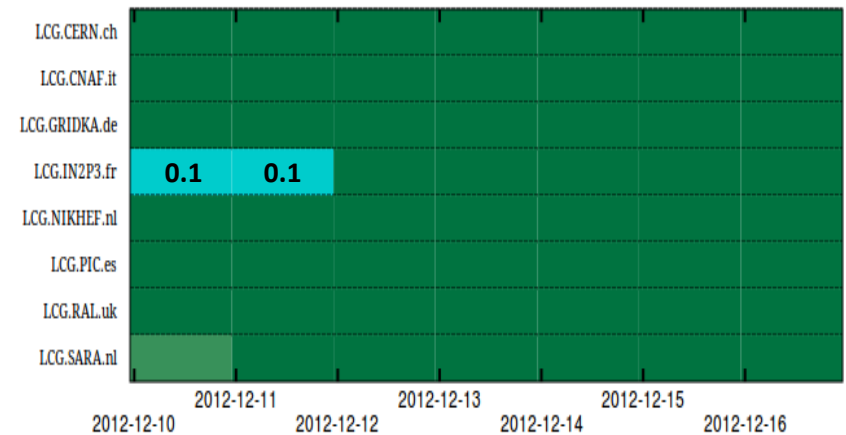
Site reliability using CMS_CRITICAL_FULL

168 hours from 2012-12-10 00:00 to 2012-12-17 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2012-12-10 00:00 to 2012-12-17 00:00



Analysis of the reliability plots :

Week 17/12/2012

- Multiple Virtual Organisations:

- *0.1: FZK. Unscheduled outage "Site wide network outage". [GOCDB:119381](#), [GOCDB:119386](#).*
- 0.2: Taiwan. Scheduled outage "Network intervention, DPM and CASTOR upgrade" [GOCDB:119416](#).

- ALICE: Nothing to report

- ATLAS:

- 2.1: INFN. Monitoring Issue. Low priority org.sam.CREAMCE-JobSubmit tests timing out due to high load of regular jobs.

- CMS:

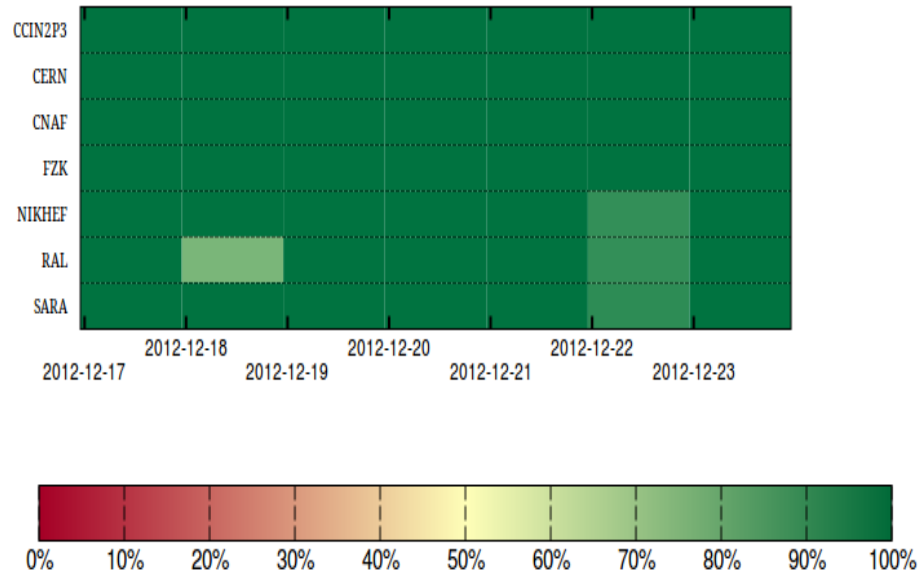
- 3.1: ASGC. org.cms.WN-swinst failing "ERROR: Some CMSSW version published but not present" [Savannah:134733](#) (Assigned to SAM squad not site squad).

- LHCb:

- 4.1: CNAF. Unscheduled outage "extraordinary maintenance operations of the filesystem" [GOCDB:119540](#).

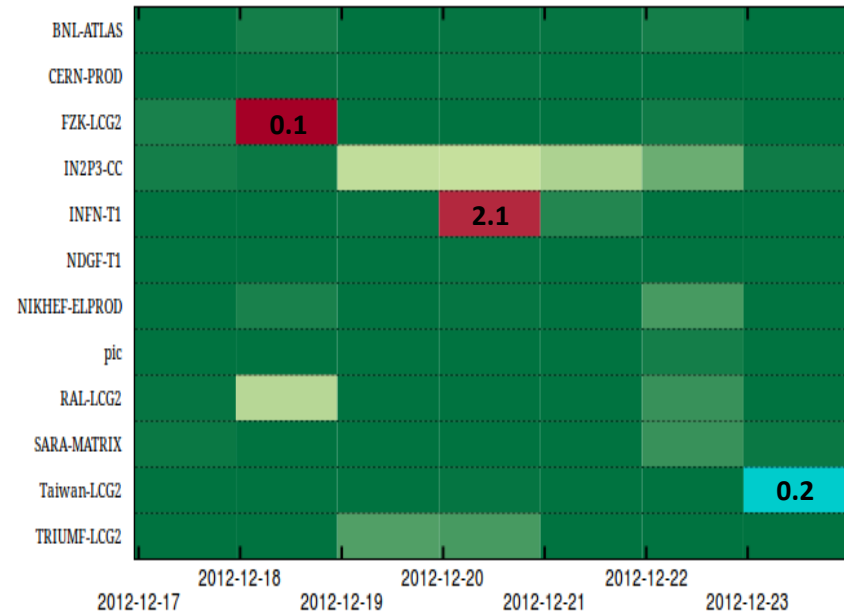
Site reliability using ALICE_CRITICAL

168 hours from 2012-12-17 00:00 to 2012-12-24 00:00



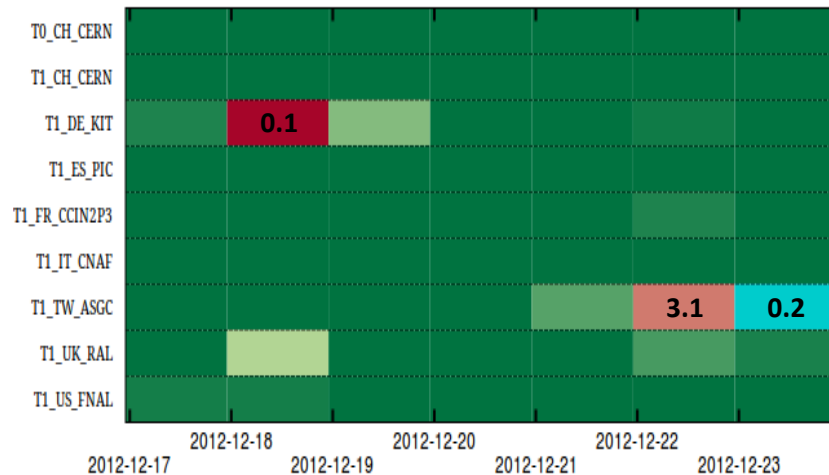
Site reliability using ATLAS_CRITICAL

168 hours from 2012-12-17 00:00 to 2012-12-24 00:00



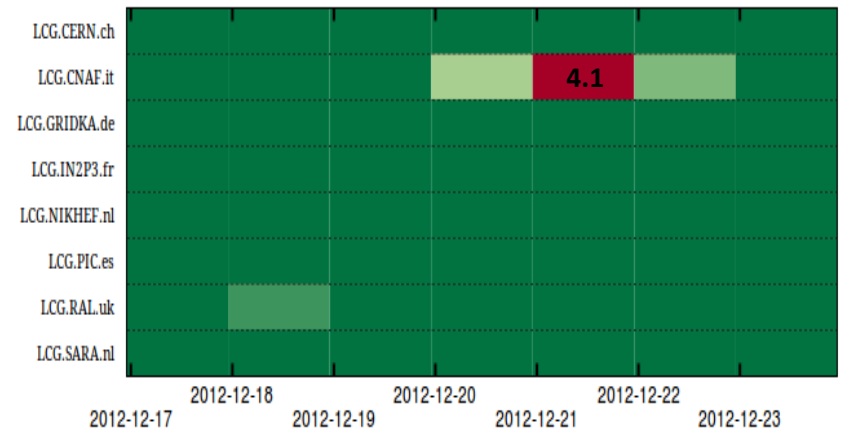
Site reliability using CMS_CRITICAL_FULL

168 hours from 2012-12-17 00:00 to 2012-12-24 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2012-12-17 00:00 to 2012-12-24 00:00



Analysis of the reliability plots :

Week 24/12/2012

- Multiple Virtual Organisations:

- 0.1: Taiwan. Scheduled outage "Network intervention, DPM and CASTOR upgrade" [GOCDB:119416](#).
- *0.2: Taiwan. Unscheduled outage "DC power cut"* [GOCDB:119614](#).

- ALICE: Nothing to report

- ATLAS:

- 2.1: RAL. Unscheduled outage "ATLAS SRM database problems" [GOCDB:119596](#).
- 2.2: TRIUMF. Scheduled warning "Site-Wide power maintenance, however power to Tier-1 centre will be on, create a warning downtime since it may still affect T1" [GOCDB:119014](#).

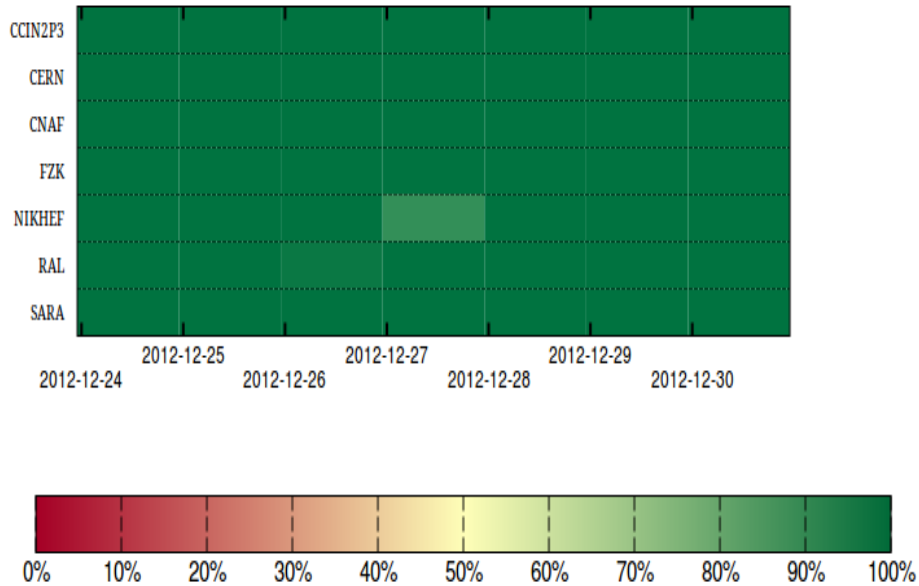
- CMS:

- 3.1: KIT. org.cms.SRM-VOPut failing intermittently. [GGUS:89850](#).
- 3.2: ASGC. org.cms.WN-swinst failing "ERROR: Some CMSSW version published but not present" [Savannah:134733](#) (Assigned to SAM squad not site squad).

- LHCb: Nothing to report

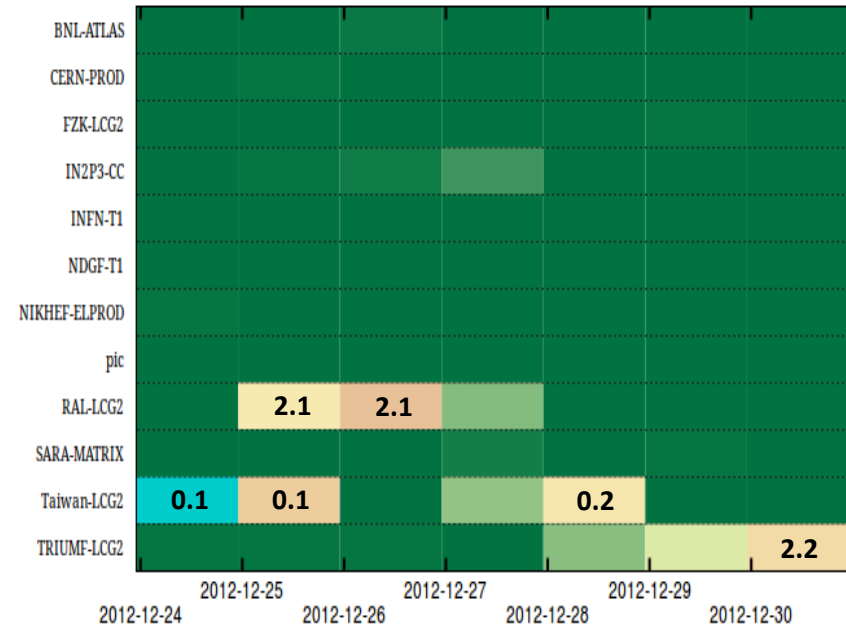
Site reliability using ALICE_CRITICAL

168 hours from 2012-12-24 00:00 to 2012-12-31 00:00



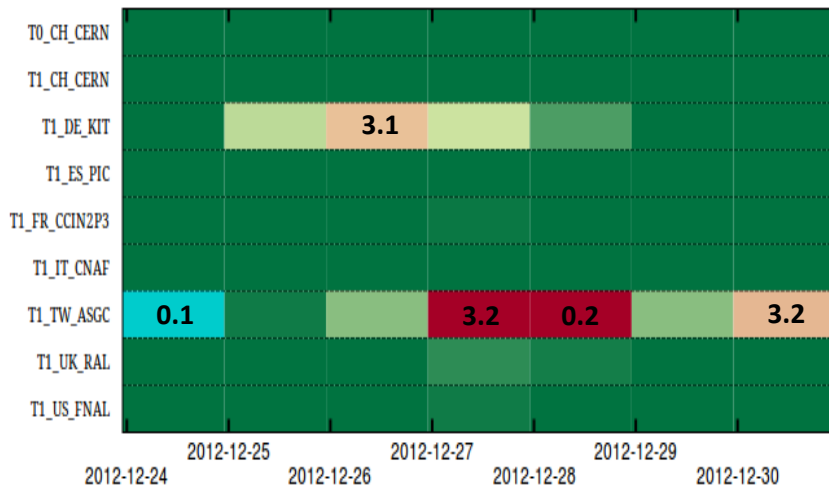
Site reliability using ATLAS_CRITICAL

168 hours from 2012-12-24 00:00 to 2012-12-31 00:00



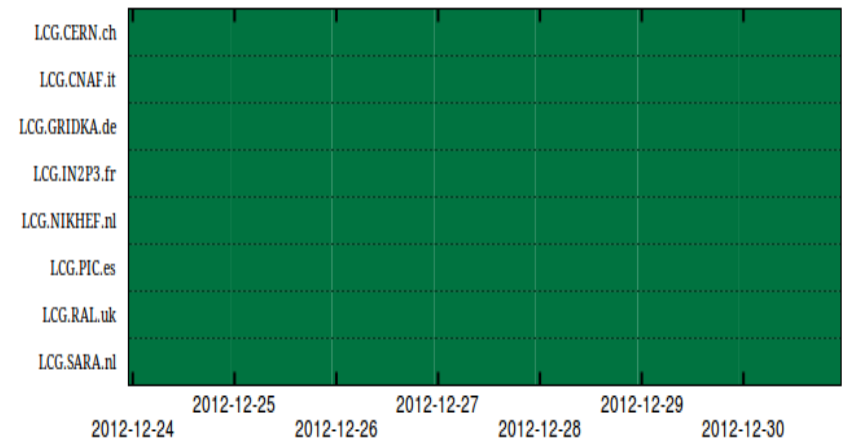
Site reliability using CMS_CRITICAL_FULL

168 hours from 2012-12-24 00:00 to 2012-12-31 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2012-12-24 00:00 to 2012-12-31 00:00



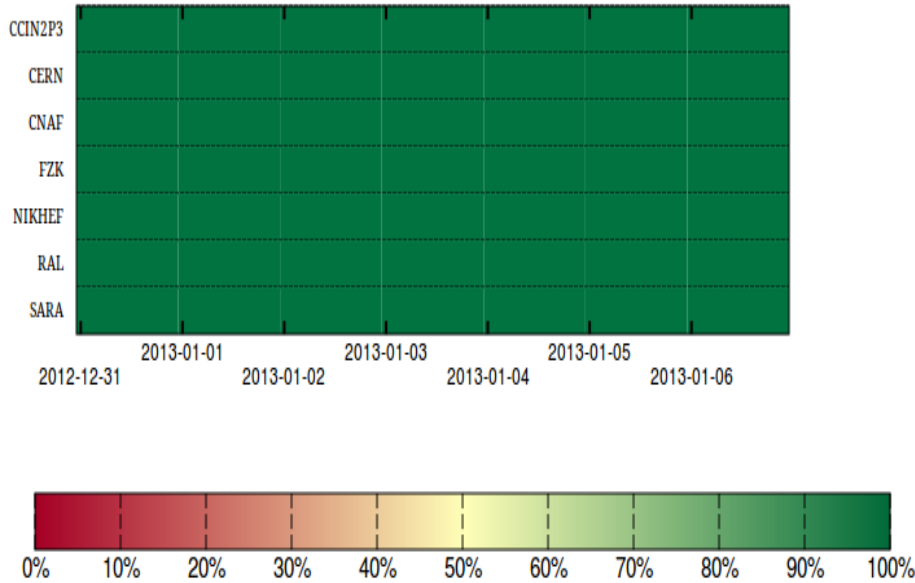
Analysis of the reliability plots :

Week 31/12/2012

- ALICE: Nothing to report
- ATLAS: Nothing to report
- CMS:
 - *3.1: ASGC. org.cms.WN-swinst failing "ERROR: Some CMSSW version published but not present" Savannah:134733 (Assigned to SAM squad not site squad).*
- LHCb:
 - *4.1: Multiple sites. Issue with certificate used for LHCb SAM tests. Not a site-specific issue. Now resolved.*

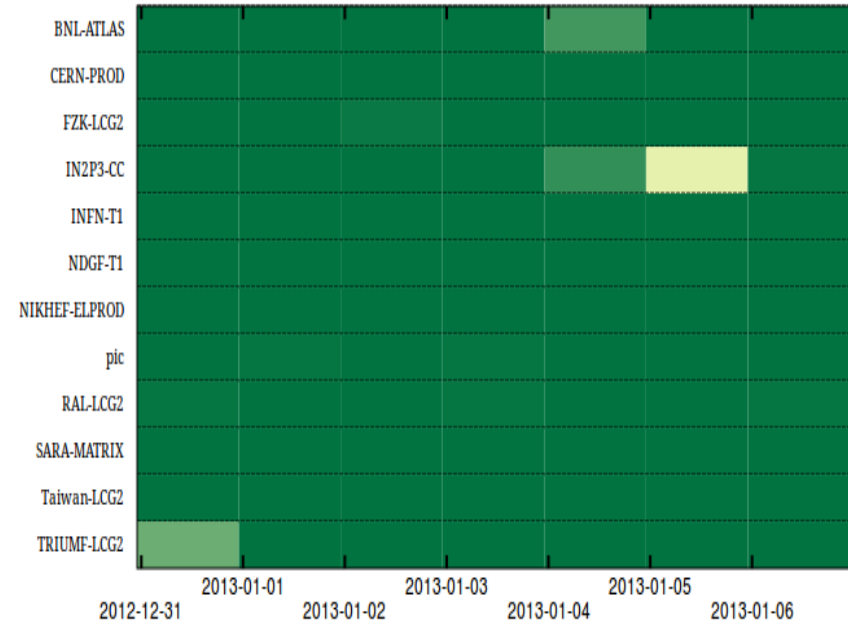
Site reliability using ALICE_CRITICAL

168 hours from 2012-12-31 00:00 to 2013-01-07 00:00



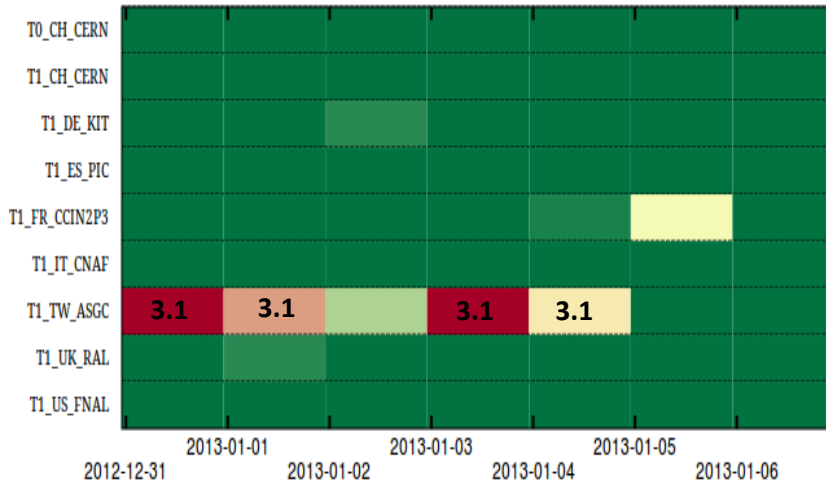
Site reliability using ATLAS_CRITICAL

168 hours from 2012-12-31 00:00 to 2013-01-07 00:00



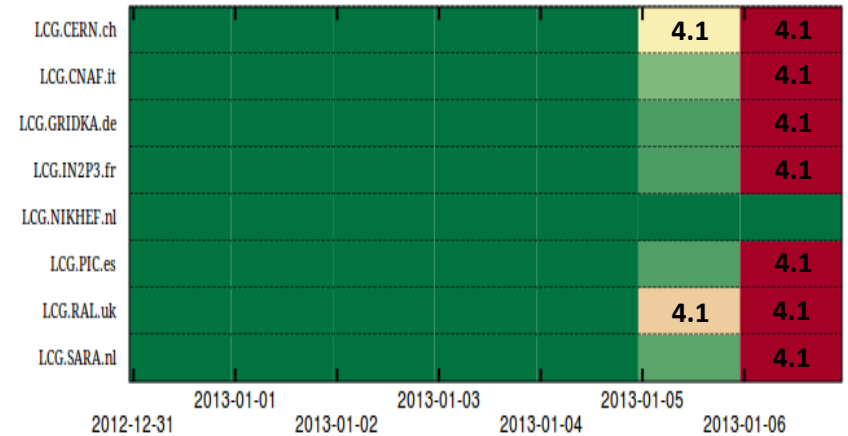
Site reliability using CMS_CRITICAL_FULL

168 hours from 2012-12-31 00:00 to 2013-01-07 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2012-12-31 00:00 to 2013-01-07 00:00



Analysis of the reliability plots :

Week 07/01/2013

- Multiple Virtual Organisations:

- *0.1: INFN. Scheduled Outage from 07/01 08:00 to 10/01 13:27: "Upgrade of the Tier 1 core switch and the largest part of storage systems"* [GOCDB:11951](#).
- 0.2: SARA. Unscheduled Outage: "Relocating the fileserver to new hardware. This downtime is forgotten to be set in GOC" [GOCDB:119831](#).

- ALICE: Nothing to report

- ATLAS: Nothing to report

- CMS:

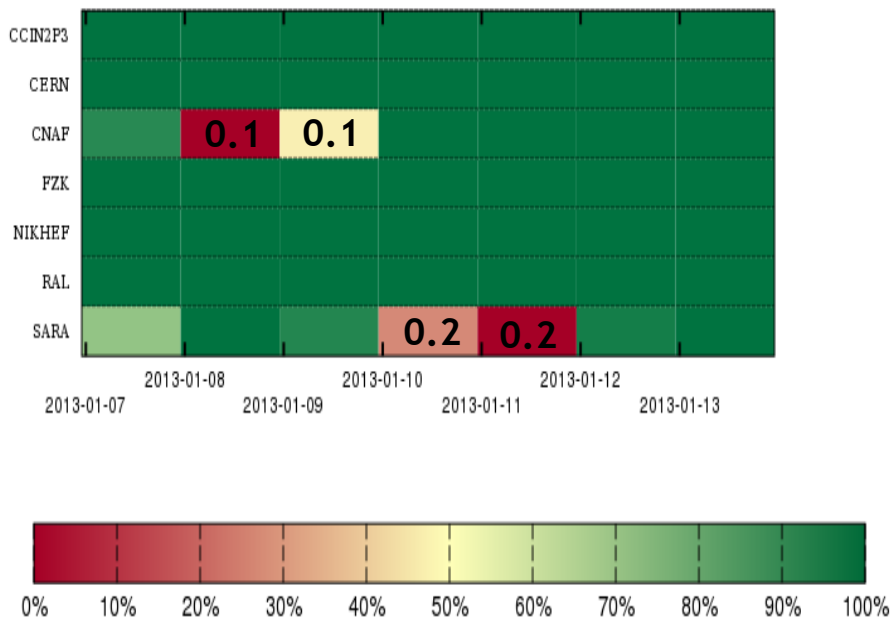
- *3.1: IN2P3. CREAM-CE tests failing: "Possible Network Issue"* [Savannah:134961](#) & *"Not running many jobs, while a lot pending"* [Savannah:134970](#).
- 3.2: ASGC. Not running jobs with the /cms/Role=pilot FQAN: they stay scheduled for up to 23.5 hours at which point SAM kills them.

- LHCb:

- 4.1: Multiple sites. Issue with certificate used for LHCb SAM tests. Not a site-specific issue. Now resolved.

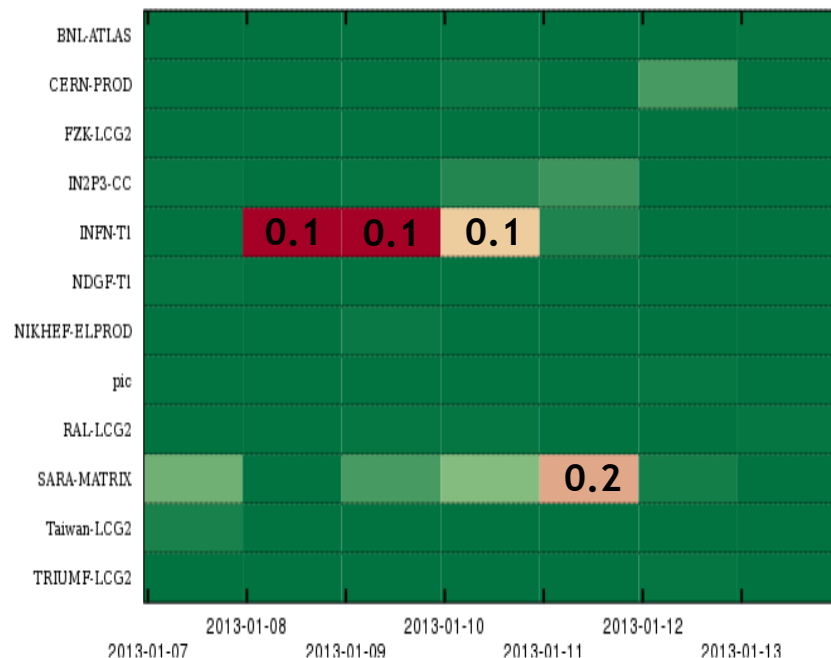
Site reliability using ALICE_CRITICAL

168 hours from 2013-01-07 00:00 to 2013-01-14 00:00



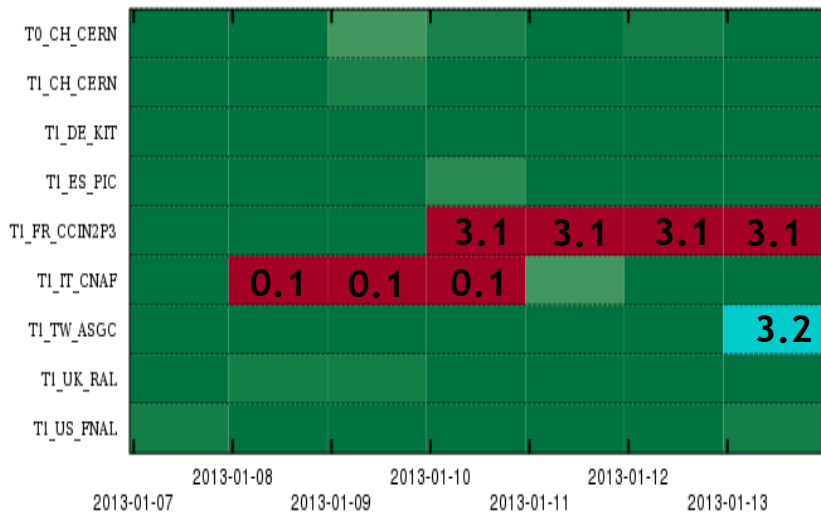
Site reliability using ATLAS_CRITICAL

168 hours from 2013-01-07 00:00 to 2013-01-14 00:00



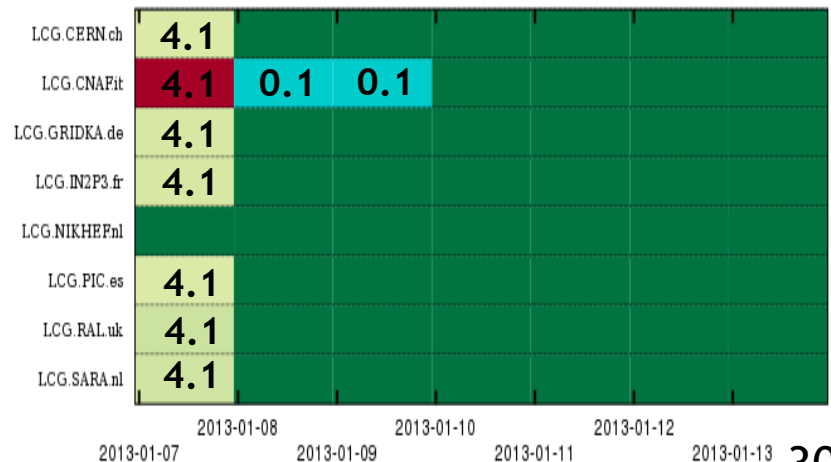
Site reliability using CMS_CRITICAL_FULL

168 hours from 2013-01-07 00:00 to 2013-01-14 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2013-01-07 00:00 to 2013-01-14 00:00



Analysis of the reliability plots :

Week 14/01/2013

- ALICE: Nothing to report

- ATLAS: Nothing to report

- CMS:

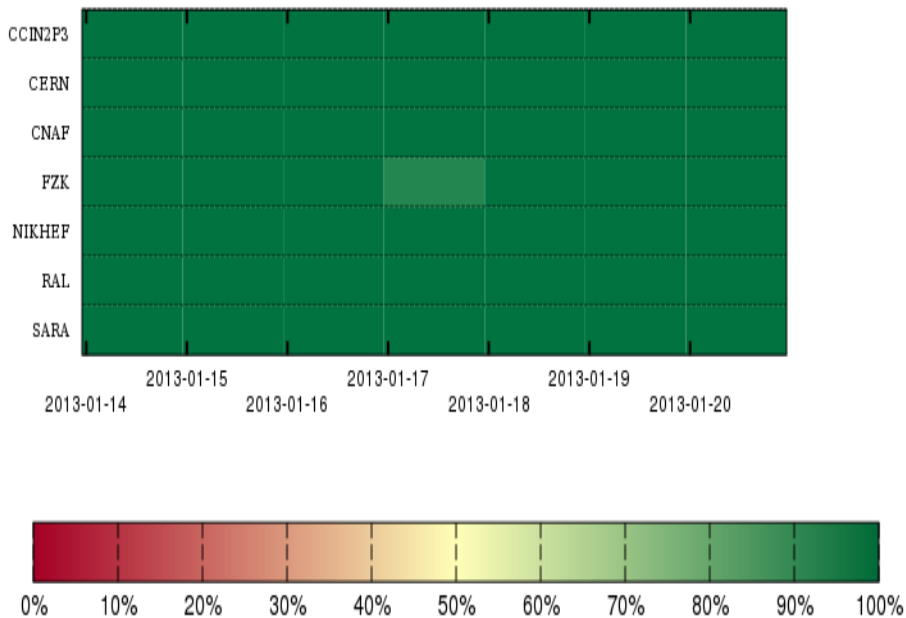
- *3.1: IN2P3. The following CREAM-CE tests were timing out: WN-gLExec, WN-mc and JobSubmit.*

- LHCb:

- *4.1: KIT. Scheduled Outage from 15-01 05:00 until 17-01 08:00: "Separation of LHCb from gridka-dcache.fzk.de SE, injecting it into lhcbstrm-kit.gridka.de; also, migrate from PNFS to Chimera namespace architecture" [GGUS:119125](#). SRM-VODel, VOLs and VOLsDir SAM tests were failing.*

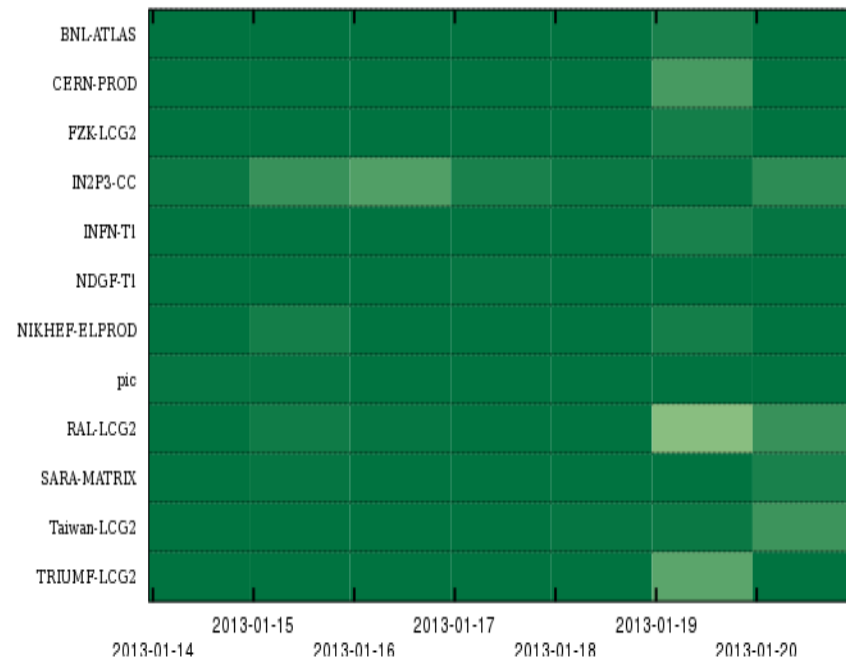
Site reliability using ALICE_CRITICAL

168 hours from 2013-01-14 00:00 to 2013-01-21 00:00



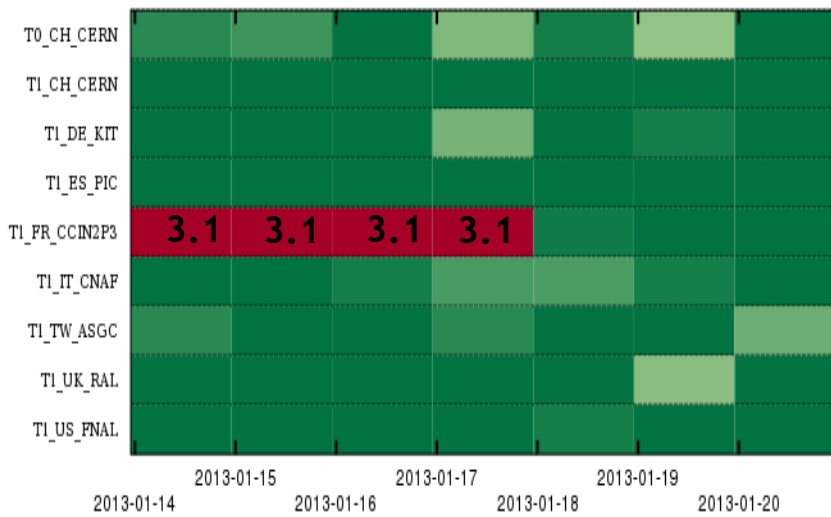
Site reliability using ATLAS_CRITICAL

168 hours from 2013-01-14 00:00 to 2013-01-21 00:00



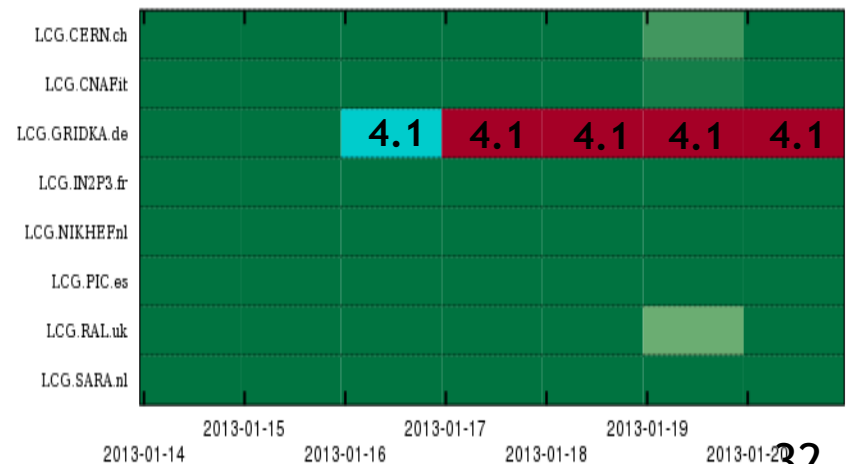
Site reliability using CMS_CRITICAL_FULL

168 hours from 2013-01-14 00:00 to 2013-01-21 00:00



Site reliability using LHCb_CRITICAL

168 hours from 2013-01-14 00:00 to 2013-01-21 00:00



Conclusions

- Business as usual – busy but successful
 - Smooth operation of computing activities over Christmas