Page | **Discussion**

Read | **View source** | **View history**

Search Wikitech

Incident documentation/20141130-Eqiad-Rack-C4

< Incident documentation

Summary

Around 03:50 UTC on Nov 30, the network switch for Rack 4 in Row C in our eqiad datacenter went offline. This immediately stopped network traffic to all machines in this rack. None of the affected services were set up for paging ops as critical services, so no pages went out. It was handled as a relatively-low-priority incident. The switch came back online and the affected machines were connected to the network again at about 10:13 UTC, for a total outage time window of about 6.5 hours. Notable affected higher-level services included phabricator and redis job queues. Affected hosts are in the table below:

Hostname
caesium
erbium
gadolinium
gold
hafnium
iridium
labsdb1006
labsdb1007
lead
logstash1001
logstash1002
logstash1003
ms-fe1004
neodymium
osm-web1003
osm-web1004
osmium
platinum
radon
rcs1001
rdb1001
ssl1005
ssl1009
stat1003

Timeline

- ~03:50 UTC - Eqiad Rack C4 Switch goes offline (according to virtual chassis logs).

Sidebar navigation

Main page
Recent changes
Server admin log (Prod)
Server admin log (RelEng)
Deployments
SRE/Operations Help
Incident status

Cloud VPS & Toolforge

Cloud VPS documentation
Toolforge documentation
Request Cloud VPS project
Server admin log (Cloud VPS)

Tools

What links here
Related changes
Special pages
Permanent link
Page information
Cite this page

Print/export

Create a book
Download as PDF
Printable version

- ~03:53 UTC - By this time, all affected machines had been logged by icinga as down hosts and echoed to IRC, but no pages went out (as configured for affected hosts/services).
- ~04:22 UTC - Ori notices the above on IRC + humans complaining about phabricator being down, calls Brandon's phone to report some kind of ops issue of unknown scope.
- ~04:24 UTC - Brandon logs in and starts investigating, noting that no major public services seem to be down. Quickly narrows it down to the affected switch being offline. Spends a lot of time trawling around in Juniper online docs and Wikitech looking for a way to revive it from the asw-c-eqiad virtual chassis console via ssh, doesn't find anything obvious. Noted in IRC that it didn't seem worth waking anyone up in the middle of the night to run out to eqiad physically over it, since phabricator seems to be the most-important fallout noted.
- ~05:54 UTC - Brandon sends email to ops mailing list documenting where things are at with this incident so far.
- ~07:00 UTC - Guiseppe logs in, notes that redis job queues seem to be affected as well, even though they should have failed over to another machine. Manually fails over redis stuff. Guesses we can wait for a fix until a couple more Europeans wake up soon?
- ~10:13 UTC - Mark logs in, connects to serial console of asw-c4-eqiad via the scs service and presses enter, and magically the switch reboots and comes back online.

Conclusions

1. Documentation on how switches and their consoles are laid out and accessed is sub-optimal. The magic reboot may have happened much sooner if one of us knew how/where to connect to the serial port at. In retrospect, this *is* somewhat documented, it's just not easy to find if you don't know what you're looking for.
2. There's some disagreement over the importance of phabricator between our current monitoring configuration and/or various parties. It was noted in the followup at the Ops meeting that phabricator should probably be considered critical and have paging enabled in monitoring, in which case humans and/or the pages should've woken more people up sooner.

Actionables

Being tracked in Phabricator at https://phabricator.wikimedia.org/tag/incident-20141129-network/

Category: Incident documentation