Page Discussion

Read View source View history

Search Wikitech 🔍

Toolforge webservices are in the final stages of migrating to the toolforge.org domain .
Please help us clean up older documentation referring to tools.wmlabs.org!

Incident documentation/20110926-PowerCable

< Incident documentation

Contents [hide]

What

Access switch power outage at 14:12 UTC for about 18 minutes.

Cause

Our Tampa data center contractor was replacing the broken management switch msw-d2-sdtpa in sdtpa rack D2. Our management network is completely unrelated to our production network, and doesn't carry any important, production traffic. However, the management switch is mounted directly under the production network access switch (asw-d2-sdtpa). During the replacement, at 14:12 UTC, the power cable of asw-d2-sdtpa was accidentally knocked loose as it was very close to the switch and cables being replaced. The technician didn't notice it because switch LEDs are on the other side.

Impact

As a result, the entire rack, consisting of about 40 MediaWiki application servers, lost network connectivity. Our monitoring tool, Nagios, complained about all servers in the rack, and due to all kinds of second order effects, many other services were (briefly) disrupted or overloaded, leading to many other Nagios downtime reports. Several Ops and Engineering members were looking at the downtime, but had difficulty pinpointing exactly what was going down, until Mark noticed that all application servers being reported down were located in the same rack D2 (when he checked in Racktables), and that asw-d2-sdtpa was reported down by Observium (e-mail to root@ and the web interface). A quick check revealed that indeed this switch had gone down. Mark got confirmation from the technician on IRC it was related to his work. As soon as the technician restored power, all network connectivity and services were restored within a minute at 14:30 UTC.

The main problem with a rack of application servers going down, besides the significant reduction in capacity, is the loss of memcached caching - meaning that MediaWiki on the remaining servers will be parsing/regenerating many pages and objects, resulting in a much higher load at reduced capacity for a while. memcached redundancy/resiliency is definitely something that could need improvement for this reason.

Mitigation

In our newer racks and data centers, we also have better redundancy of the network, including redundant power and uplinks for access switches, so this doesn't happen so easily. If we do not intend to move Tampa any time soon, we should look at retrofitting our older racks with that setup as well.

Category: Incident documentation

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