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.wmflabs.org!

Incident documentation/20170507-labs-dns

< Incident documentation

Contents [hide]

Summary

Around 05:00UTC on Sunday 2017-05-07 new Labs instance creation started to fail. Andrew noticed the issue later that day. Most of the various parts of designate and powerdns were functioning properly but the AXFR sync between the two was failing, so new records were failing to be registered in the public DNS servers.

Andrew switched the active designate server from labservices1002 to labservices1001, at which point normal service was restored (at around 17:30 UTC). The ultimate cause of the issue remains obscure, although labservices1002 seems generally troubled.

Timeline

- 05:03 icinga alerts because the nova-fullstack process is no longer running. This happens after 7 creation attempts had failed (one every 5 minutes), so most likely the original issue appeared about 30 minutes earlier.
- 16:00 while testing an unrelated fix Andrew notes that new instances are failing to get DNS connections and begins investigation
- 17:00 It's clear that most Designate services are working properly. Instance creation is detected, the new records are created properly and registered in the database. Pdns is not noticing the new records, and the syslog on labservices1001 and 1002 is full of timeout messages rather than proper syncs with designate-mdns:

```
May 7 06:44:02 labservices1001 pdns[47578]: Received serial number updates for 0
zones, had 161 timeouts
```

there are also many deadlock warnings in the designate-central log. These are not unheard of during normal operation but are much more frequent than usual:

```
designate-central.log.1:2017-05-07 15:36:15.730 21604 WARNING
designate.central.service [req-5064db6f-5238-491c-a5fa-447464baa1aa - - - - -]
Deadlock detected. Retrying...
```

Puppet runs are also extremely slow on labservices1002, over 5 minutes per run. This is reminiscent of https://phabricator.wikimedia.org/T159835 which was fixed with a reboot.

- 17:20 Andrew gives up on understanding what's happening an begins the failover process to labservices1001 as documented in https://wikitech.wikimedia.org/wiki/Labs_troubleshooting#Fail-over_2
- 17:40 Failover is completed and normal service restored. No DNS interruptions were detected for running services.

Conclusions

This outage was long but probably didn't affect any actual users (other than Andrew.) The outage would have been detected much earlier if the fullstack test paged on failure.

The failover process from labservices1002 to 1001 went very smoothly with little or no additional service interruption.

We still don't know what's going on with labservices1002 -- investigation continues as part of https://phabricator.wikimedia.org/T164675 .

Actionables

- Understand and fix whatever is wrong with labservices1002
- Add paging to the fullstack alert. A single failure is probably not worth paging but if we get 6 in a row and the test switches off I'd like to know about it.

Category: Incident documentation

Privacy policy   About   Disclaimers   Code of Conduct   Developers   Statistics   Cookie statement   Mobile view
Wikitech

WIKIMEDIA
a project

Powered By
MediaWiki