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/20180626-LoadBalancers

< Incident documentation

Summary

An attempt to use a broken functionality of scap, together with an unexpected behaviour of our load balancers made all wikis and all api endpoints fail almost completely for 6 minutes.

Timeline

- 2018-06-26 08:45: Marko and Petr start a deployment window in which an important migration will take place
- 2018-06-26 08:48: A first full scap sync is run, but is aborted as mediawiki emits notices for `Undefined variable: wmgUseEventBus`
- 2018-06-26 09:03: A second full scap sync is attempted, again unsuccessfully (the notices persist)
- 2018-06-26 09:27: Given the situation doesn't improve, it is decided to run scap with the additional flag for restarting HHVM, and an heads-up is given in #-operations
- 2018-06-26 09:28: The scap command is run
- 2018-06-26 09:29-32: while Giuseppe acknowledges that he is sure the HHVM restart from scap is broken, and that it might even be harmful, he advises against stopping the command as the worst-case scenario (all servers get depooled and none is restarted) should be protected by pybal as T184715 is resolved. Moreover, since that feature is still there in scap, it might have been fixed in the meanwhile.
- 2018-06-26 09:33: It is noticed that the worst-case-scenario behaviour from scap is happening. Scap is abruptly stopped, but by this time, it has removed more than 90% of all servers in both clusters from the pool. Comment on IRC from Giuseppe "let's hope pybal saves us"
- 2018-06-26 09:34: first reports of sites being down from users
- 2018-06-26 09:35: A quick glance at the state of the pools on one of the eqiad load balancers confirms that T184715 is indeed not fixed
- 2018-06-26 09:35: A shower of alerts start pouring - our monitoring confirms everywthing's more or less down
- 2018-06-26 09:36: A first mass-repool (for appservers) is issued
- 2018-06-26 09:37: A mass repool of API servers, first in eqiad, then in codfw is issued
- 2018-06-26 09:38: Users report the sites are now working again
- 2018-06-26 09:42: All alerts clear

The real outage, which has been almost complete for non-cached resources, lasted between 9:32 and 9:38, so 6 minutes.

Conclusions

The scap function that caused this issue has long been broken and needs to be removed from the software. That, combined with the pybal bug, caused this outage. While removing the scap function should be easy, most efforts should be spent in really fixing T184715.

Actionables

- Remove --restart from scap (DONE: phab:T198185)
- Make pybal not depool servers if it goes below depool_threshold (TODO: phab:T184715)

Category: Incident documentation