



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/20171128-ORES

< [Incident documentation](#)

Contents [\[hide\]](#)

- [1 Summary](#)
- [2 Timeline](#)
- [3 Conclusions](#)
- [4 Actionables](#)

Summary

The ORES API cluster lost worker nodes due to memory starvation caused by an excess of requests. As worker nodes died, the service overloaded and failed API responses climbed. A second issue appeared in which Extension:ORES would cause a cache stampede on every ORES-enabled pageview, as per-wiki thresholds expired out of memcache, resulting in tens of thousand times more internal API requests than expected. These factors combined to cause ORES API failures ranging up to 100%, until the cache stampede was fixed 7 hours later.

Tracked in [Phabricator](#)
Task T181538

The worst impact was that most of the nine other services hosted in the same cluster stopped responding to live traffic requests. On the ORES side, its features were unavailable for users during much of this incident.

Timeline

[17:33:21] <icinga-wm> PROBLEM - ORES worker production on ores.wikimedia.org is CRITICAL: CRITICAL - Socket timeout after 10 seconds

(Drop in available memory on scb1001-2 causes OOMs, <https://grafana.wikimedia.org/dashboard/db/ores?panelId=24&fullscreen&orgId=1&from=1511888358404&to=1511892210665>)

[17:34:21] <icinga-wm> RECOVERY - ORES worker production on ores.wikimedia.org is OK: HTTP OK: HTTP/1.1 200 OK - 848 bytes in 9.546 second response time. *(Service continues to flap for the entire incident: <https://grafana-admin.wikimedia.org/dashboard/db/ores?orgId=1&from=1511868777076&to=1511985197076&panelId=9&fullscreen>)*

[17:52:05] <awight> We're losing ORES servers, looks like we'll have to do some manual restarting...

[17:52:31] <icinga-wm> PROBLEM - ores on scb1001 is CRITICAL: CRITICAL - Socket timeout after 10 seconds

[17:58:51] akosiaris lowers LB weights to scb1001-2, which are less powerful than scb1003-4.

[18:04:04] <akosiaris> now on scb1001 OOM showed up

(We discuss and decide to failover from eqiad to codfw, which has much more total processing power.)

[18:33:53] <akosiaris> ok traffic from all varnishes should start flowing to codfw

[18:47:04] <awight> A whole lot of fork(): Cannot allocate memory [core/master_utils.c line 724]

(It becomes clear that codfw can't handle the "traffic" either, and in fact is 100% disabled.)

19:17:31.012 * Asynchronous AOF fsync is taking too long (disk is busy?). Writing the AOF buffer without waiting for fsync to complete, this may slow down Redis.

(codfw Redis has disappeared—this is a single point of failure for ORES [see [FIXME https://phabricator.wikimedia.org/T181559](#)].)

[19:20:21] <awight> akosiaris: mutante: We were dealing with T181538 and failed over to codfw. Now codfw is gone.

[19:27:00] <awight> halfak: Where are the massive number of test_stats requests coming from? UA is MediaWiki which don't make any damn sense. *(We figure out the main culprit, an own-goal cache stampede, <https://logstash.wikimedia.org/goto/36c5fdfa3b3bf983b3a16c2acd91ccbe>,*

[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](#)

<https://logstash.wikimedia.org/goto/c5c525c1a715e97a45fc2e6745111c06>

[19:44:35] <awight> It's a chain-reaction, when a wiki's thresholds cache expires, if the request fails, we fallback to old thresholds. Which now returns a 400, and the old code doesn't cache the error. Instead, every Special:RecentChanges page requests test_stats.

[20:17:54] <halfak> Looks like we're having an event again. We see a cascade of requests for test_stats from MW and then our web workers start to lock up (maybe)

(Failover back to eqiad.)

[20:28:57] <mutante> !log forcing puppet run on cache misc to revert "failover ORES to codfw"

[21:17:45] <akosiaris> it does seem to be rewriting the AOF very often. *(We decide that disabling cache persistence for the Celery queue is a good idea, but not expected to solve our issues.)*

[21:41:44] <akosiaris> !log disable ORES queue redis persistency by config set appendonly no on oresrdb1001
(Ops notices that the self-requests should be going to an internal URL, where they won't put load on Varnish.)

[22:31:34] <akosiaris> !log deploy wmf-config/CommonSettings.php for ORES internal discovery URL, <https://gerrit.wikimedia.org/r/#/c/393924/> T181538

[00:32:18] <logmsgbot> !log awight@tin Synchronized php-1.31.0-wmf.8/extensions/ORES: Hotfix to mitigate cache stampeding, T181567 (duration: 00m 50s)


Case closed.

Conclusions

- In hindsight, once the likely problem was identified as a cache stampede, we could have changed configuration to hardcode all thresholds and this would have saved 5 hours of ORES downtime.
- The decision to not disable ORES on the wikis looks like it was correct, although I don't know whether we were harming services outside of ORES. The impact of disabling entirely would have been to increase the number of users who couldn't use ORES functionality.

Actionables

NOTE: Please add the [#wikimedia-incident](#) Phabricator project to these follow-up tasks and move them to the "follow-up/actionable" column.

- Fix cache stampede in extension thresholds code ([phab:T181567](#))  **Done**
- Host workers on dedicated hardware ([phab:T168073](#))
- Make Redis not a single point of failure ([phab:T181559](#))
- Graph to monitor number of API requests per client user-agent ([phab:T182222](#))

Category: [Incident documentation](#)

This page was last edited on 5 November 2018, at 18:32.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. See [Terms of Use](#) for details.

[Privacy policy](#) [About](#)
[Wikitech](#)

[Disclaimers](#) [Code of Conduct](#) [Developers](#) [Statistics](#) [Cookie statement](#) [Mobile view](#)

