Page   **Discussion**   Read   **View source**   **View history**   Search Wikitech

Incident documentation/20171120-Ext:ORES

<   Incident documentation

Contents [hide]

Summary

An ORES service deployment included a fix for bug T179711, which changed the ORES API behavior for impossible threshold requests. Previously, we would return a 500 when any of the thresholds could not be calculated, but the new code would return a successful response supplying all thresholds except the impossible one, which simply holds a "null" value. This new response format was handled incorrectly by Extension:ORES, throwing an uncaught RuntimeException after failing to interpret the null value.

Since threshold configuration is different for each wiki, only ruwiki and frwiki were affected. The Special:RecentChanges and Special:Watchlist pages were completely unusable on these wikis during the outage.

A MediaWiki train deployment went out during the outage period and was rolled back because of the apparent alignment. It was not involved in the bug, just another casualty.

Timeline

- 21:37 <awight@tin> Started deploy [ores/deploy@5084251]: Updating ORES to revscoring 2.0.10, T179711
- 22:05 - 22:13 ORES services are restarted with new code.
- 22:10 Sharp rise in HTTP 500 errors is visible in Grafana: https://grafana.wikimedia.org/dashboard/file/varnish-http-errors.json?orgId=1&from=1511206917882&to=1511228187762
- 22:27 <awight@tin> Finished deploy [ores/deploy@5084251]: Updating ORES to revscoring 2.0.10 (duration: 49m 54s)
- 22:49 Krinkle: Sharp rise in HTTP 500 errors as of 22:05 (45 minutes ago)
- 22:50 Krinkle: MW HTTP 500 spike tracked as https://phabricator.wikimedia.org/T181006
- 22:54 <awight@tin> Started deploy [ores/deploy@5084251]: Rollback ORES; T179711. *Here I grabbed the wrong revision, tried to roll back to the *new* version.*
- 22:55 <awight> rolling back ORES to fix T181006
- 22:55 <awight@tin> Finished deploy [ores/deploy@5084251]: Rollback ORES (duration: 01m 05s)
- 23:35 awight: purge cache keys for ORES thresholds on frwiki and ruwiki
- 23:11 <awight> purged memcache key 'ruwiki:ORES:threshold_statistics:goodfaith:1',
- 23:18 awight@tin: Started deploy [ores/deploy@95cd523]: Rollback ORES (take 2); 181006. *This was an old stable revision, but turned out to not be cached yet, so I aborted the rollback because it would have taken unacceptably long (45 min).*
- 23:19 awight: aborted ORES rollback
- 23:25 <awight@tin> Started deploy [ores/deploy@82a13ae]: Rollback ORES (take 3); T181006
- 23:35 legoktm@tin: Synchronized wmf-config/InitialiseSettings.php: emergency disable ORES on frwp/ruwp T181006 (duration: 00m 49s)
- 23:37 HTTP 500 errors drop back to normal levels.
- 23:38 Finished deploying ORES rollback to 82a13ae.

Conclusions

See also bug T181010.

- awight was only monitoring server-side graphs and logs during the deployment, whereas I should have been looking at the client side as well. Need to update deployment documentation to mention this, and follow myself during future deployments.
- Ext:ORES shouldn't be able to kill any of the pages it's used on. Any type of failure should be caught and the feature gracefully degraded. If only some models can be used, proceed without the others. If no models can be used, proceed without ORES. Log like bloody hell, though.
- The rollback tree-ish was not easy to figure out. We were using tin to deploy to both production and to our new cluster for stress testing, so scap logs were too messy to be useful. Eventually, I had to "ls -ltr" the deployment cache on the server machine, which is also error-prone. The biggest thing to fix here is that we shouldn't be deploying non-production machines out of the same directory as production, IMO.
- It's not feasible to manually verify every wiki when deploying ORES changes. Not all wikis are available on the beta cluster, and the sheer number of combinations of page and language is out of range for humans. We could possibly have automated UI testing on beta for a few known hot spots, but that's also a slow and expensive way to QA.
- Deploying and especially rolling back ORES takes too long. The worst pain points for rollback can be solved with: a) parallel deployment across hosts, and b) caching the built virtualenv corresponding to each source revision.
- ORES configuration in Beta cluster was such that we never could have detected this bug. The failing threshold was on the "goodfaith" model, which was accidentally disabled for all wikis but English. Keep beta config in sync with production.

Actionables

- ✅ **Done** phab:T181191 - Make MediaWiki pages robust to ORES or Ext:ORES failures.
- ✅ **Done** phab:T181183 - Deployment documentation and protocol to cover what awight missed here.
- (pending review) phab:T181071 - Cache virtualenv for faster rollback.
- ✅ **Done** phab:T181067 - Parallelize ORES deployment.
- ✅ **Done** phab:T181187 - Always make ORES beta cluster config the same as production.

Related

Meeting Minutes

Category: Incident documentation

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

WIKIMEDIA
a project

Powered By
MediaWiki