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!

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

Incident documentation/20190913-maps

< Incident documentation

document status: draft

Contents [hide]

1 Summary
  1.1 Impact
  1.2 Detection
2 Timeline
3 Conclusions
  3.1 What went well?
  3.2 What went poorly?
  3.3 Where did we get lucky?
  3.4 How many people were involved in the remediation?
4 Links to relevant documentation
5 Actionables

Summary

On Friday September 13, map servers were saturating CPU due to some badly formed requests that were not validated properly by the service. This led to partial unavailability of maps from ~4:30 UTC to ~14:30 UTC. Situation was resolved by validating traffic at the caching layer.

Impact

Service was degraded for ~9h.

Thanks to tiles high cache hit ratio, only ~2% of requests were affected according to Turnilo. Given the high number of tiles seen by a single user during a session, it is probably that most users were affected to some extend.

Detection

- HTTP availability for Varnish was flapping starting 4:26 UTC, getting worse by 6:49 UTC
- No page was sent, no direct alert pointing to maps / kartotherian explicitly

Timeline

All times in UTC.

- ~04:20 **OUTAGE BEGINS** CPU saturated on maps servers (maps[12]00[1-4])
- 04:26 icinga alert about HTTP availability for Varnish
- 04:27 recovery of HTTP availability for Varnish
- 05:40 icinga alert about HTTP availability for Varnish
- 05:42 recovery of HTTP availability for Varnish
- 06:49 icinga alert about HTTP availability for Varnish, starts falling regularly from now on
- 06:52 maps identified as the cause of the above alert
- 07:15 Icinga alert for kartotherian LVS endpoint
- 08:33 kartotherian restarted on maps1003, with no effect
- 08:37 rolling restart of karotherian
- 08:45 stop tilerator on maps to help reduce load - no effect
- 08:57 kartotherian eqiad depooled, problem moves to codfw
- 08:57 identified increased occurrence of issue about parsing geojson in logs (can't actually find that again, the graph now looks flat)

- 09:11 kartotherian eqiad repooled
- 09:24 deny access to /geoline on maps1004 - limited effect
- 09:38 deny access to /geoshape on maps1004 - seems to reduce CPU load
- 09:46 re-enabling /geoline on maps1004
- 09:54 /geoshape heavily throttled on varnish - seems to be effective (536549)
- 10:55 icinga alert for maps100[12] kartotherian endpoints health on maps1001 is CRITICAL
- 12:37 temp ban of class of urls on maps1003 nginx
- 12:56 banning more urls on maps1003
- 13:12 ban problematic URLs at varnish (536583)
- 13:38 ban problematic URLs at varnish (536588)
- 14:20 ban problematic URLs at varnish (536595)
- 14:30 **OUTAGE ENDS**

Conclusions

A bug was introduced when fixing linting issues to introduce the CI into the CI pipeline, this created a failure in the HTTP error handler making Kartotherian unable to validate request parameters that leads to high CPU cost and timeout. This needs to be addressed in Kartotherian itself (536641).

The deploy of the code containing the bug occurred September 12 at 21:09 UTC.

The amount of support we have on maps does not match the exposure of the service. While the few people working on maps are dedicated to their work and doing their best, we have too many (bad) surprises. The technical stack has many known and unknown issues and our knowledge of that stack is insufficient.

The majority of maps traffic comes from other websites or apps reusing our tiles. This is allowed (at least to some extend) by Maps Terms of Use and was the original intent of the project. Given the amount of support we have at the moment, this might need to be revisited.

What went well?

- high caching ratio mitigated the visibility of the issue

What went poorly?

- knowledge of the stack is insufficient
- logs were not very helpful and somewhat misleading (e.g. phab:T158657).

Where did we get lucky?

- ?

How many people were involved in the remediation?

- 6 SRE spending significant time during the incident
- 2 SWE during the second half of the incident

Links to relevant documentation

Documentation is minimal (Maps), but this specific problem is being addressed and unlikely to occur again.

Actionables

- Fix HTTP error handler in kartotherian - 536641 (code merged, but needs to be tested and deployed)
- Improve testing tin kartotherian endpoints
- Review the amount of support Maps has in regard of its visibility and use cases

Categories: Incident documentation | Incident documentation drafts | Maps outages, 2019

This page was last edited on 28 April 2020, at 19:55.

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