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/20190407-Zotero

< Incident documentation

Contents [hide]

Summary

On the morning of April 7, most of the zotero pods in codfw started using too much memory, and stopped responding to requests.

Impact

Given restbase and citoid are active/active, any user coming from eqsin, codfw or ulsfo and trying to get a citation would have seen a degraded service.

Detection

The LVS level check of zotero, and the service-checker test of citoid both reported the issue.

Timeline

All times in UTC.

- 04:58 sudden increase in the memory used by the first zotero pod
- 05:05 all zotero pods in codfw have now high memory watermark. Service checker on citoid reports a problem
- 05:12 The zotero LVS endpoint has become unresponsive to monitoring. A page is sent. **OUTAGE BEGINS**
- 05:13 A recovery page arrives. The service will keep flapping and more pages are sent out 5:23 (recovery at 5:37), at 5:41 (recovery at 5:42) at 6:00
- 06:01 Alexandros, Giuseppe and marostegui respond to the page that is now being sent to people in EU timezones
- 06:03 Giuseppe depools zotero in codfw, even if the recovery arrives, so that the issue can be better analyzed. **OUTAGE ENDS**
- 06:23 After some log analysis, it is decided to kill the pods which still show a high memory watermark.

Conclusions

There isn't much to conclude, apart from the fact we still have situations where zotero can fail because of sudden memory increases. This is a bug in the software and unless we can create a reproducible test case (as this seems to happen because of some user request), there aren't many chances to see it fixed. On the positive side, this hadn't happened in a long time.

The root cause of this specific incident is still TBD.

What went well?

- The problem is known, the monitoring adequate, the response clear. This isn't a new kind of outage and we're

well equipped to respond to it.

What went poorly?

- No reponse to repeated pages/recoveries for almost 1 hour.
- No runbook exists for this

Where did we get lucky?

- It's fair to say a malfunctioning zotero doesn't have a huge impact. Requests rates average at 0.05 req/s. See https://grafana.wikimedia.org/d/NJkCVermz/citoid?refresh=5m&panelId=46&fullscreen&orgId=1&var-dc=codfw%20prometheus%2Fk8s&var-service=citoid&from=now-7d&to=now
- That some of us have a light sleep, probably?

Links to relevant documentation

AIUI the documentation on how to operate on kubernetes for debugging should be written this quarter.

Actionables

- Document kubernetes debugging procedures (TODO: Create task)
- Identify and reproduce the situation that caused the memory leak (TODO: Create task)

Category: Incident documentation