



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/20150503-JobQueueRedis

[< Incident documentation](#)

## Contents [\[hide\]](#)

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

## Summary

Jobrunner deployment by Aaron Schulz on Saturday 2 May hammered down Redis and made the jobrunners basically ineffective, and causing 2+ second delays on any action that needed to submit a job to the queue and/or to fetch any data from redis. Editors noticed the slowness and notified us. A rollback of the deployment (once correctly done) solved all the issues.

## Timeline

- 2015-05-02 20:25: [Jobrunner deployment](#) by Aaron
- Editors start noticing slowness and report it a few hours later [T97930](#)
- 2015-05-03 12:00: Alex Monk and Giuseppe begin investigating quite vague reports of slowness in editing. Database issues are pretty quickly ruled out.
- 2015-05-03 12:15: Alex notices redis connection timeouts spamming fluorine.eqiad.wmnet:/a/mw-log/hhvm.log; we also notice the jobrunners seem dead (which, with redis not working, is expected)
- 2015-05-03 12:20: Giuseppe identifies the problem with Redis maxing out one single CPU, and suggests, given the timing of the jobrunners problems, it may be caused by Aaron's deployment.
- 2015-05-03 12:38: Revert of jobrunner deployment, [Jobrunner#Deployment](#) followed. Following the instructions on wikitech at the time, Giuseppe just restarts the jobrunner service
- 2015-05-03 13:44-13:54: Brandon joins the investigation, and Aaron joins too, explains new extra step necessary for jobrunner changes - namely to restart jobchron (not documented anywhere at that moment)
- 2015-05-03 13:54: Giuseppe attempts restart of jobchron, but gets it wrong (wrong name used?)
- 2015-05-03 14:16-14:37ish?: Investigation on redis leads ops to think redis is stuck in a single-thread lua interpreter for most of the time, guess some bug has created too many keys. Attempted deletions of keys by Brandon to try to fix the problem
- 2015-05-03 15:17: Giuseppe notices his error earlier, and jobchron gets properly restarted, service appears to begin to return to normal

## Conclusions

There are a lot of takeaways from this incident, the most prominent of which are:

- For no reason but an Unbreak Now! ticket one should deploy (by auto-merging) a change on a Saturday and in general during non-working days. Auto-merging is harmful too
- The monitoring on the jobrunners and the jobqueue is inadequate and has been broken several times by changes in the software in the last year.
- We abuse redis badly. The EVAL and EVALSHA commands that use Lua on the redis server create an horrible performance bottleneck and shouldn't be used as much as we do (a series of tcpdumps on a client basically proved that we mostly use those commands only)
- We mimick a queue with redis, and we do that badly too. Maybe it's time we use a real queue instead of redis that, according to its author, should not be used as more than a toy queue.
- The jobchron service was added in production months prior (<https:// Gerrit.wikimedia.org/r/#/c/192207/>), but no documentation on it existed whatsoever. We should be more disciplined in those cases.

[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](#)

## Actionables

*Explicit next steps to prevent this from happening again as much as possible, with Phabricator tasks linked for every step.*

- Status: ■ **Done** [Document](#) that jobchron needs to be restarted for jobrunner changes [Special:Diff/157315](#)
- We should cleanly separate the jobqueues for critical (as in, data integrity) vs non-critical (refreshlinks, rootjobs for parsoid/restbase?), so that the latter can't kill the former, and so that it's easier to clean up one without affecting the other, etc...
- Review deployment practices: Don't release on a non-working day if not strictly necessary, don't release by auto-merging a pretty significant and untested change, don't release if you don't want to inspect possible unintended consequences, and most importantly don't do the three things at the same time.
- Monitoring/alerts of job queue/runners: [T62105](#)
- Monitoring/alerts of rate of hhvm.log entries
- ...

Category: [Incident documentation](#)

This page was last edited on 4 May 2015, at 18:47.

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](#)

