

Async processing in the WMF cluster



WIKIMEDIA
FOUNDATION

Async processing for the wikis

- Jobqueue
- Cron scripts (wikidata)
- Change-propagation
- HTCP (via jobqueue)
- ...



**We don't really need
all these mechanisms.**



WIKIMEDIA
FOUNDATION

Issues with the current jobqueue

- Scalability
- Durability
- Multi-DC replica
- Debugging
- Interoperability with SOA



Some action is needed

Maintaining various different mechanisms for propagating changes and performing asynchronous processing is mostly inconvenient, but might or might not be a good idea (let's discuss it!). What we can't keep around is the redis-based jobqueue.

Do we have good candidates?

**Requirement: We want
to continue supporting
pure-PHP simple
installations**



WIKIMEDIA
FOUNDATION

Jobqueue model

- Any part of MediaWiki can submit jobs
- Retry logic
- Throttling/backoff
- Execute once
- Dependency
- Delayed execution
- Deduplication



Differences in models

Change Propagation is an event-based system that is external to MediaWiki and has its own rules coded in.

This clashes with the current model where jobs are coded into MediaWiki, and messages are php-serialized objects/data structures.

**Which features is
Change Propagation
missing?**



WIKIMEDIA
FOUNDATION

Possible strategies

Full transition to EventBus / CP

- Deprecate the Jobqueue in MediaWiki
 - Provide a simple pure-php processor
 - Convert current extensions to use it
 - Deduplication logic should go in the job itself
- CP should call the MW api

Is it worth it?

While the idea to move all jobs processing in production to event-based rules in ChangeProp may seem interesting, it is a lot of work. Basically, every job that uses the current jobqueue model would need to be rewritten.

What other options do we have?

Use EventBus as a backend for the jobqueue

- Make individual jobs special “events” in EventBus
- Use CP to just rely such jobs to a specific load-balanced endpoint (as we do now!)
- Convert some specific jobs to use CP



Use EventBus as a backend for the jobqueue

Converting jobs to CP means dropping them from the eventbus-based jobqueue and set up rules in CP for those.

Example: CirrusSearch updates



Write a kafka backend for the jobqueue

- This is *relatively* simple
- Doesn't require a lot of changes to MediaWiki itself
- What use of EventBus then?
- Convert EventBus/CP back to use the jobqueue?



Use a third-party job executor

- No need to maintain the software ourselves
- Maybe not tailored to our scale/needs
- Gearman
- Chronos
- Nomad

Discussion



WIKIMEDIA
FOUNDATION

**Copyright © Giuseppe Lavagetto,
Wikimedia Foundation Inc.**

**This work is licensed under a Creative
Commons Attribution-ShareAlike 3.0
International License.**



**The Wikimedia presentation template is CC by SA 3.0,
Wikimedia Foundation Inc. All brands and marks are
property of the respective owner.**

