



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/20150803-Kafka

[< Incident documentation](#)

Summary

A gradual [upgrade was planned](#) to expand the Kafka cluster and to bring it to Jessie running Kafka 0.8.2.1 in <https://phabricator.wikimedia.org/T106581>.

Otto tested this upgrade plan in labs, and all was fine. However, the testing was not extensive enough, as it was not tested with varnishkafka or a regular stream of events.

Timeline

At 2015-08-03T17:21, Otto began the process of partition reassignment to balance partitions across 6 nodes, 3 of which were the original Precise with Kafka 0.8.1.1, 3 Jessie with Kafka 0.8.2.1. This seemed fine, but Otto believes that once some partitions were totally moved off of the original 0.8.1.1 brokers, things started going wrong. Otto is still not entirely sure what went wrong. He spent several hours attempting to reconcile the problems. During this time, most varnishkafkas were still producing. However, many of them periodically failed with metadata request timeouts, which caused failed produce requests.

After several hours of trying to fix this, Otto decided the faster way to get everything back to normal would be to revert to only the 4 original 0.8.1.1 Kafka brokers, by wiping the Kafka cluster metadata and data clean and starting from scratch. This was done at 2015-08-03T23:45. After this, the status quo was restored and message production and consumption proceeded as normal.

Webrequest data experienced significant loss during this outage. Data in Hadoop between 2015-08-03T17:00 and 2015-08-04T00:00 should be considered unreliable.

See also: [Analytics/Data/Webrequest](#)

Actionables

- Status: ■ **Done** - Rebuild 0.8.2.1 Kafka .deb package with updated Snappy dependency, and double check other dependencies in gradle.
- Status: ■ **Done** - Figure out what went wrong and/or do more extensive testing in labs.
- Status: ■ **Done** - Augment deployment plan to do this again without incident.

Category: [Incident documentation](#)

This page was last edited on 26 August 2015, at 20:15.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. See [Terms of Use](#) for details.