Incident report: 2014-09-26

Description

A storage node in our Cape Town cluster failed at around 09:15 on Friday 26 September 2014. This failure was due to a corrupted ext4 file system. Ordinarily such failures are not be an issue due to the redundant nature of our storage solution, however in this instance, the failure cascaded into a cluster wide problem that affected the bulk of our customer VMs.

Analysis

The Ceph storage layer utilised by isoho.st makes use of a probabilistic function to distribute customer data roughly evenly over the available storage nodes in the cluster. When a storage node fails, a backfilling process is kicked off to redistribute the cluster data to take into account the changed topology due to the missing node. Unfortunately, this distribution function allocated too much data to a single node, thus causing it to fill beyond 80% full. At this point a safety mechanism kicks in to freeze all writes to the storage cluster as a whole to prevent the cluster falling into an unrecoverable state.

Impact

While customer VMs did not go down, they were paused in execution until the storage issue was resolved. The net effect is that any customer VMs that were writing to disk were inaccessible (effectively frozen) for approximately 40 minutes until full service was restored at around 09:55.

Actions

After examining server logs to ascertain that there were no hardware faults, the failed storage node was brought back into service after rebuilding the corrupted file system.

As a preventative measure, the distribution function's weights were also adjusted to take into account the present data load across storage nodes.

We will also follow up with the Ceph project upstream to find out if there is safe mechanism for handling a full storage node that is less disruptive to the whole cluster.

Please direct any questions or comments to support@isoho.st