

\$ What broke, why, and what changes

Postmortem

Sam Whitfield

Site Reliability • September 2, 2026

\$ Timeline

14:02 **ALERT** error rate crosses threshold

14:06 **PAGE** on-call acknowledges

14:31 **MITIGATED** traffic shifted to the previous version

15:10 **RESOLVED** root cause identified and patched

\$ How long it actually lasted

68 min

from first alert to resolution, of which 29 were before anybody looked at the right dashboard.

Contributing factors

\$ Not a root cause, several

1. the change was correct and the config for it was not
2. the alert fired on a symptom two systems away
3. the rollback needed a person who was asleep

\$ Actions

OWNED alert on the cause, not the symptom

due next sprint

OWNED rollback runnable by anyone on call

due next sprint

OPEN config validated in CI

needs a decision

Open questions

\$ What is still unknown

OPEN the thing nobody has measured yet

OPEN the assumption this plan rests on

NEXT who finds out, and by when

\$ Appendix

Graphs, raw output and the commands to reproduce any of it. Kept after the close so the talk stays short and the evidence stays available.