Resource exhaustion on us-west2 for HHMI

Status: Closed

Overview

GCP zone repeatedly ran out of resources

What Happened

The autoscaler failed to bring up new nodes to meet demand, due to exhaustion of resources in GCP zone. We saw ZONE_RESOURCE_POOL_EXHAUSTED errors in our cloud logs for the user-pod nodepool

Resolution

Subsequently we reduced the resource requirements of the user-nodes (fitting fewer pods per node) so that we're more likely to acquire one.

Where we got lucky

No comments added

What Went Well?

No comments added

What Didn't Go So Well?

We didn't get alerts for exhaustion events, only the symptom (repeated failed server starts)

Action Items

- https://github.com/2i2c-org/infrastructure/issues/7681
- https://github.com/2i2c-org/infrastructure/pull/7755
- https://github.com/2i2c-org/infrastructure/pull/7691

OWNER OF REVIEW PROCESS

Angus

IMPACT TIME

Feb 12 at 01:48 to Feb 13 at 16:37

DURATION

1d 14h 48m 54s

*All times listed in this report are in Edinburgh, London.

Timeline

Feb 12, 2026

1:48 AM

Triggered through the API.
Description: [FIRING:1] Two servers failed to start in the last 30m hhmi staging (immediate action needed) (View Message)
INCIDENT #1871

[FIRING:1] Two servers failed to start in the last 30m hhmi staging (immediate action needed)

7:21 PM   Engineer observes that GCP is short on resources in zone

Note added by Angus.
Resolution Note: GCP ran out of resources in us-west2-b
INCIDENT #1871

[FIRING:1] Two servers failed to start in the last 30m hhmi staging (immediate action needed)

1:48 AM

Triggered through the API.
Description: [FIRING:1] Two servers failed to start in the last 30m hhmi staging (immediate action needed) (View Message)
INCIDENT #1871

[FIRING:1] Two servers failed to start in the last 30m hhmi staging (immediate action needed)

7:21 PM   Engineer observes that GCP is short on resources in zone