Post Mortem

Date: 2023-2-2

Authors: Michael, Hongbin, Luis

Status: Complete

Summary: High latency and node crash caused by for loop in interceptor layer

Impact: High latency for approximately 6 minutes, alerting down for about 40 minutes

Root Causes: For loop in interceptor layer adding a lot of unnecessary information to loki

Trigger: Unneeded for loop running 5000 iterations for each request made, causing higher latency and excessive memory usage

Resolution: Deleted for loop in interceptor layer, created new image and updated cluster with new working image. Waited for node to reboot and resolved monitoring pod issues after by updating helm installations for each.

Detection: Michael detected high latency spike

Action Items:

Action Item	Type	Owner	Bug
Rollback deployment to previous image	Mitigate	Hongbin	n/a Done
Fixed Error	corrective	Hongbin	n/a Done
Created new image	corrective	Luis	n/a Done
Applied new image	preventative	Luis	n/a Done
Helm updated for Prometheus and Loki	corrective	Michael	n/a Done
Establish alerting for cluster conditions	preventative	Michael	n/a To do

Lessons Learned

What went well:

- Source code issue was resolved relatively quickly

What went wrong:

- Node crash prevented checking of logs
- While the root cause was caught relatively quickly, the nature of the bug caused excessive damage in a short period of time

Where we got lucky:

- Able to quickly swap to old working image.

Timeline

2023-02-02 (all times PST)

- 8:06 AM New image deployed, initial high latency detected
- 8:10 AM INCIDENT BEGINS
- 8:11 AM Rollback deployment to previous image
- 8:12 AM searched logs and found error originated from interceptors layer
- 8:13 AM fixed error and created new image
- 8:14 AM deployed image, fixing broken method
- 8:16 AM issues with cluster detected, pods stuck in Terminating status
- 8:40 AM Node crash discovered, caused by excessive memory usage by experimental image
- 8:41 AM waiting for Kubernetes to reboot node/pods
- 9:10 AM reboot mostly successful, few pods need replacing for monitoring tools
- 9:20 AM helm updated for prometheus and loki
- 9:30 AM monitoring up and running
- 9:31 AM INCIDENT ENDS, service operating correctly