LIS hub cannot scale

Status: Reviewed

Overview

At 10AM UK time the B zone in Google Cloud's europe-west2 region (London) ran out of resources to allocate. While we deploy regional clusters, we restrict the nodes to a specific zone colocated with our NFS Filestore to maximise performance. This resulted in the cluster not being able to scale and a "backoff after failed scale-up" message reported to the user.

What Happened

- User reported "backoff after failed scale-up" error
- This is usually related to quotas so engineer checked those and increased the quotas for Persistent SSD disk (which was red) and the CPUs. Neither of these worked.
- Eventually the ZONE_RESOURCE_POOL_EXHAUSTED error was found in logs
- Engineers confirmed this meant Google had no more resources in the requested zone
- The affected node pool was edited in the GCP console to allow node creation in the A and C zones within the region
- This resolved the problem

Resolution

Incident resolved by allowing the node pool to create nodes in the A and C zones which weren't exhausted. This is a temporary fix as it now means that the NFS Filestore is working across zones, which will impact performance.

Where we got lucky

No comments added

What Went Well?

No comments added

What Didn't Go So Well?

Single engineer working on this issue wearing many hats (incident commander, comms, debugging) for large period of the incident. A team response would've provided more support. Pages also didn't get through to response engineers which compounded the above issue

OWNER OF REVIEW PROCESS

Sarah Gibson

IMPACT TIME

Nov 14 at 02:00 to Nov 14 at 08:30

DURATION

6h 30m

*All times listed in this report are in Pacific Time (US & Canada).

Action Items

Do EITHER (1) OR (2)
1) Restrict node pool back to only B zone after Black Friday event is over
2) Move the cluster and the NFS to the A zone
GitHub issue: https://github.com/2i2c-org/infrastructure/issues/1944
3) Investigate an enterprise plan which allows for regional NFS (would cost more)
GitHub issue: https://github.com/2i2c-org/infrastructure/issues/1945
4) Ensure PagerDuty notifications can always get through to engineers
GitHub issue: https://github.com/2i2c-org/team-compass/issues/574

Timeline

Nov 14, 2022

5:40 AM

Triggered by Sarah Gibson.
Description: LIS hub cannot scale (View Message)
INCIDENT #36

LIS hub cannot scale

5:45 AM

Sarah Gibson #managed_jupyte_inc_36

there's an incident on LIS that I don't know how to fix
> Warning FailedScheduling 3m29s http://gke.io/optimize-utilization-scheduler 0/2 nodes are available: 2 node(s) didn't match Pod's node affinity/selector. preemption: 0/2 nodes are available: 2 Preemption is not helpful for scheduling.
> Normal NotTriggerScaleUp 3m27s cluster-autoscaler pod didn't trigger scale-up: 1 node(s) didn't match Pod's node affinity/selector, 1 in backoff after failed scale-up
"backoff after failed scale-up" usually means a quota issue. I already doubled the quota of the Persistent Disk SSD quota which was red, but that doesn't seem to have helped the problem.
Are the SSDs used for home dir storage? I tried looking at the grafana but I couldn't find a graph that showed storage per user https://grafana.uk.2i2c.cloud/dashboards
Matthew has some upcoming classes
I opened an incident ticket: https://github.com/2i2c-org/infrastructure/issues/1915

5:55 AM

Sarah Gibson #managed_jupyte_inc_36

Found a pvc with 100GB on it, it's from the prometheus server https://github.com/2i2c-org/infrastructure/issues/1915#issuecomment-1313742288

5:56 AM

Sarah Gibson #managed_jupyte_inc_36

But I don't know why that would cause a failure to scale? I feel like I'm down the wrong rabbit hole

6:10 AM

Sarah Gibson #managed_jupyte_inc_36

hub logs

6:13 AM

Sarah Gibson #managed_jupyte_inc_36

```[W 2022-11-14 14:06:44.540 JupyterHub user:824] sgibson91's server failed to start in 600 seconds, giving up.

Common causes of this timeout, and debugging tips:

1. Everything is working, but it took too long.
```

To fix: increase `Spawner.start_timeout` configuration
to a number of seconds that is enough for spawners to finish starting.
2. The server didn't finish starting,
or it crashed due to a configuration issue.
Check the single-user server's logs for hints at what needs fixing.

[I 2022-11-14 14:06:44.541 JupyterHub spawner:2808] Deleting pod lis/jupyter-sgibson91
[W 2022-11-14 14:06:44.669 JupyterHub base:1030] 3 consecutive spawns failed. Hub will exit if failure count reaches 5 before succeeding
[E 2022-11-14 14:06:44.669 JupyterHub gen:630] Exception in Future <Task finished name='Task-2530' coro=<BaseHandler.spawn_single_user.<locals>.finish_user_spawn() done, defined at /usr/local/lib/python3.9/site-packages/jupyterhub/handlers/base.py:954> exception=TimeoutError('Timeout')> after timeout
Traceback (most recent call last):
File "/usr/local/lib/python3.9/site-packages/tornado/gen.py", line 625, in error_callback
future.result()
File "/usr/local/lib/python3.9/site-packages/jupyterhub/handlers/base.py", line 961, in finish_user_spawn
await spawn_future
File "/usr/local/lib/python3.9/site-packages/jupyterhub/user.py", line 850, in spawn
raise e
File "/usr/local/lib/python3.9/site-packages/jupyterhub/user.py", line 747, in spawn
url = await gen.with_timeout(timedelta(seconds=spawner.start_timeout), f)
asyncio.exceptions.TimeoutError: Timeout

[I 2022-11-14 14:06:44.671 JupyterHub log:186] 200 GET /hub/api/users/sgibson91/server/progress (sgibson91@10.154.0.51) 599743.33ms
[E 2022-11-14 14:07:09.968 JupyterHub gen:630] Exception in Future <Task finished name='Task-2531' coro=<ConfiguratorSpawnerMixin.start() done, defined at /home/jovyan/.local/lib/python3.9/site-packages/jupyterhub_configurator/mixins.py:25> exception=TimeoutError('pod lis/jupyter-sgibson91 did not start in 600 seconds!')> after timeout
Traceback (most recent call last):
File "/usr/local/lib/python3.9/site-packages/tornado/gen.py", line 625, in error_callback
future.result()
File "/home/jovyan/.local/lib/python3.9/site-packages/jupyterhub_configurator/mixins.py", line 45, in start
return await super().start(*args, **kwargs)
File "/home/jovyan/.local/lib/python3.9/site-packages/kubespawner/spawner.py", line 2760, in _start
await exponential_backoff(
File "/usr/local/lib/python3.9/site-packages/jupyterhub/utils.py", line 236, in exponential_backoff
raise asyncio.TimeoutError(fail_message)
asyncio.exceptions.TimeoutError: pod lis/jupyter-sgibson91 did not start in 600 seconds!

[I 2022-11-14 14:07:34.339 JupyterHub log:186] 200 GET /hub/metrics (@10.100.2.28) 7.37ms
[I 2022-11-14 14:08:34.342 JupyterHub log:186] 200 GET /hub/metrics (@10.100.2.28) 10.76ms
[I 2022-11-14 14:09:34.339 JupyterHub log:186] 200 GET /hub/metrics (@10.100.2.28) 7.44ms```

6:38 AM

Sarah Gibson #managed_jupyte_inc_36

Tried doubling to CPU quota to 48 for the hell of it

6:40 AM

Sarah Gibson #managed_jupyte_inc_36

Manually resizing the nb-user nodepool to one to at least see if I can get a user server up

6:46 AM

Sarah Gibson #managed_jupyte_inc_36

I don't think that worked, because even though the resizing completed successfully in the console, `k get nodes` didn't show a new node

6:46 AM

Sarah Gibson #managed_jupyte_inc_36

And number of nodes has remained at 0 in the console

6:47 AM

Sarah Gibson #managed_jupyte_inc_36

Will try deleting the hub pod

6:48 AM

Sarah Gibson #managed_jupyte_inc_36

Didn't work either

8:01 AM

Georgiana Dolocan #managed_jupyte_inc_36

The first `ZONE_RESOURCE_POOL_EXHAUSTED` was at 10 am uk time

8:29 AM

Sarah Gibson #managed_jupyte_inc_36

Ooooh, I think that worked!!!!

8:47 AM

yuvipanda #managed_jupyte_inc_36

Looks like they ran out of cloud

8:51 AM

yuvipanda #managed_jupyte_inc_36

I see the new nodes are in `europe-west2-c`

Nov 17, 2022

2:20 AM

Resolved by Sarah Gibson through the website.

INCIDENT #36

LIS hub cannot scale