freedesktop.org

Status of freedesktop.org GitLab/cloud hosting

Benjamin Tissoires (bentiss / benjamin.tissoires@gmail.com)

https://bentiss.pages.freedesktop.org/xdc-2021-slides

Hello

I am Benjamin Tissoires (bentiss)

I am here because I am a freedesktop admin.

This is a followup presentation of the one I gave last year.

(I also work at Red Hat)

In the previous episode...

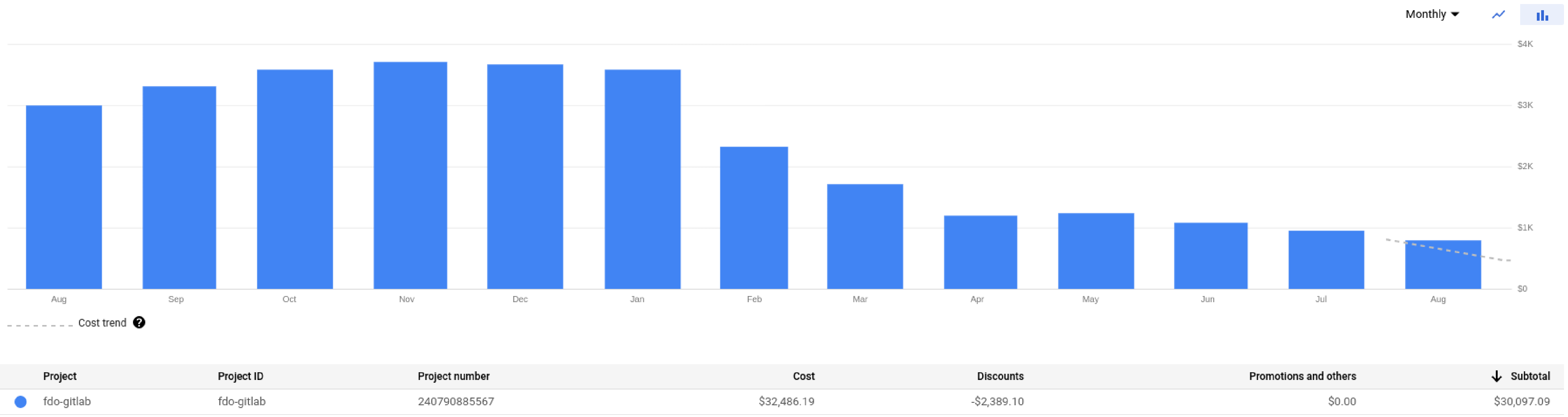
In the previous episode…

- $6000 a month
- aarrggh
- docker-free-space
- artifacts were chased down
- git cache for mesa
- ~$3000 a month 🎉

Clickbait:

How to go from $3000 a month to $800 in 12 months?

Current numbers



	Project	Project ID	Project number	Cost	Discounts	Promotions and others	↓ Subtotal
●	fdo-gitlab	fdo-gitlab	240790885567	$32,486.19	-$2,389.10	$0.00	$30,097.09

Month	Cost	Month	Cost
Sep 20	$3305.54	Mar 21	$1700.66
Oct 20	$3572.59	Apr 21	$1192.14
Nov 20	$3702.38	May 21	$1232.31
Dec 20	$3665.04	Jun 21	$1076.71
Jan 21	$3585.00	Jul 21	$953.04
Feb 21	$2314.87	Aug 21	$799.05

Sep 2020 - Feb 2021

The heat was gone, time to prototype a better solution:

- host our own cluster on Equinix Metal (formerly Packet.net)

constraints:

- keep the container registry on GCS (6TB of data)
- have something equivalent of GCP
- do a migration to the new cluster without much downtime

Initial deployment (Jan 2021)

- k3s as a base
- 3 machines from Equinix Metal (one server + 2 agents)
- wireguard for inter-node communication
- *light* storage through local SSDs on Ceph (git + db)
- **heavier** storage through Equinix Metal Elastic Storage (artifacts)
- kilo to join both clusters
- custom gobetween tunnels to link services together

Initial deployment (Jan 2021)

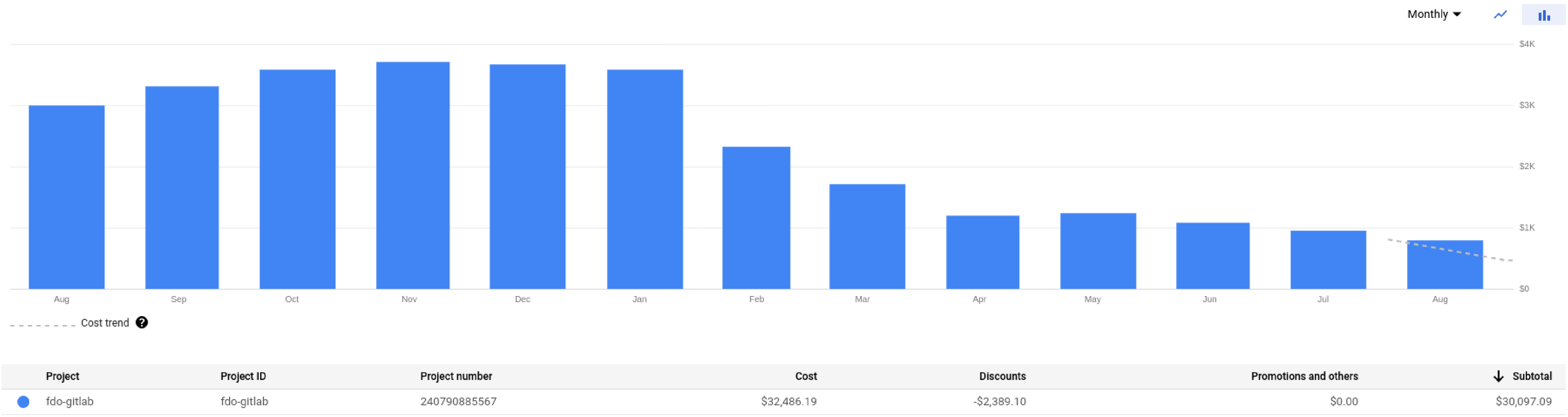
MISSING PIECES FROM GCP

- logging
  - replaced with EKS (Elastic Search + Kibana)
- redundancy of the main server
  - well, we had to move on and I couldn’t make a high-availability cluster at that time

End of Jan 2021, beginning of Feb 2021: migration from GCP

- we moved all the git repos to Equinix Metal
- we moved most of the artifacts (couple of months before and all the job logs)
- on Feb 7, 2021 -> shut down of the service, and db migration, and update DNS
- then gradually killed the services on GCP, and removed nodes (VMs)

First migration



	Project	Project ID	Project number	Cost	Discounts	Promotions and others	↓ Subtotal
●	fdo-gitlab	fdo-gitlab	240790885567	$32,486.19	-$2,389.10	$0.00	$30,097.09

Month	Cost	Month	Cost
Sep 20	$3305.54	**Mar 21**	**$1700.66**
Oct 20	$3572.59	Apr 21	$1192.14
Nov 20	$3702.38	May 21	$1232.31
Dec 20	$3665.04	Jun 21	$1076.71
Jan 21	$3585.00	Jul 21	$953.04
⚠️**Feb 21**⚠️	**$2314.87** 🎉	Aug 21	$799.05

What next?

March 2021 -> May 2021

worried about non High-Availability control plane

- the control plane (server) had no backups (GitLab has)
- we would lose the entire cluster if it comes down

March 2021 -> May 2021

worried about non High-Availability control plane

- discovered kube-vip
- understood a little bit more BGP
- k3s got improved
- started prototyping a k3s with HA control plane

Oops, I killed the entire cluster

May 11 2021

OR THEREABOUTS

while testing the automated deployment of the cluster:

- I followed the fresh docs and deployed the test cluster on the production cluster…
- luckily the data was **not** lost, only hidden and not used
- I managed to recover it after a little bit of time to calm down

Oh, BTW, we almost lost all of our storage

(SORT OF)

May 2021

In an email to our Equinix Metal sponsor:

> However, I realized last Saturday, that the elastic storage is going to be decommissioned *very* soon (on June 1st 2021).

Answer:

> I think your approaching of expanding your cluster to include more local disk is the right option. Feel free to leverage some s1 machines or additional m1 to help you make it happen.

May 2021

In an email to our Equinix Metal sponsor:

> However, I realized last Saturday, that the elastic storage is going to be decommissioned *very* soon (on June 1st 2021).

Answer:

> I think your approaching of expanding your cluster to include more local disk is the right option. Feel free to leverage some s1 machines or additional m1 to help you make it happen.

May 2021: new prototype

- 6 machines (3 servers + 3 agents)
- k3s-HA
- kube-vip
- BGP
- fully deployed with a Python script
- MinIO cluster for serving the artifacts
- dedicated elastic IPs for GitLab and other services

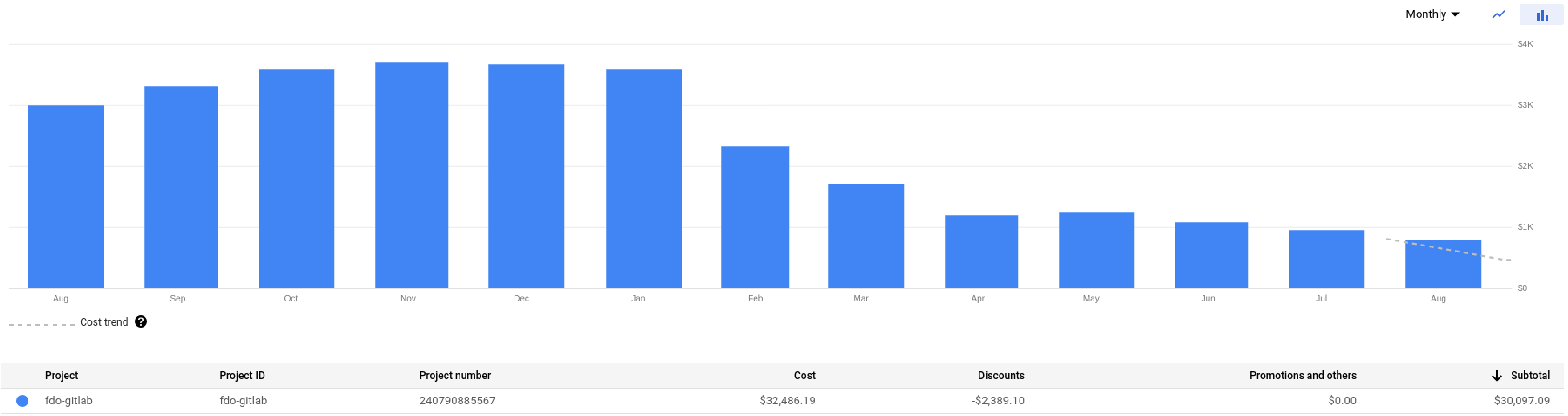
May 24 2021: new cluster in production

- git migration couple of weeks before
- shutdown, db migration, back up
- 🎉
- MinIO immediately started to show weaknesses (artifacts)
- go back to drawing board
- selected Ceph Object Storage as a replacement
- put it in prod after one week

First week of June 2021

- cluster back up
- we lost all of our artifact data for the previous week
- seems much better now

Second migration



	Project	Project ID	Project number	Cost	Discounts	Promotions and others	↓ Subtotal
●	fdo-gitlab	fdo-gitlab	240790885567	$32,486.19	-$2,389.10	$0.00	$30,097.09

Month	Cost	Month	Cost
Sep 20	$3305.54	Mar 21	$1700.66
Oct 20	$3572.59	Apr 21	$1192.14
Nov 20	$3702.38	May 21	$1232.31
Dec 20	$3665.04	⚠️**Jun 21**⚠️	**$1076.71**
Jan 21	$3585.00	Jul 21	$953.04
Feb 21	$2314.87	Aug 21	$799.05

Next step: reduce the registry cost

- deployed a docker cache
  - registry-mirror.freedesktop.org
- fdo runners on Equinix Metal are told to use the cache
- does not work completely as expected
- sub-$1000 a month

What now?

Machines sponsored by Equinix Metal

operation	#	type	proc	ram	disk
k8s cluster	3	c2.medium.x86	1 x AMD EPYC 7401P 24-Core Processor @ 2.0GHz	64GB	2 x 120GB SSD + 2 x 480GB SSD
	3	s1.large.x86	2 x Intel Xeon E5-2620V4 @2.1GHz	256GB	2 x 480 GB SSD + 12 x 2TB HDD + 1 x 128GB SSD
runner	3	m1.xlarge.x86	2 x Intel Xeon E5-2650 v4 @2.2GHz	256GB	6 x 480GB SSD
	2	c2.large.arm	1 x Ampere eMAG 8180 32-core @ 3.0Ghz	128GB	1 x 480GB SSD
minio-packet.fd.o	1	m1.xlarge.x86	2 x Intel Xeon E5-2650 v4 @2.2GHz	256GB	6 x 480GB SSD

monthly cost:

~$12,700 (sponsored)

Current issues

- logging/metrics not really stable
- we still have a lot of 502s
- can we ditch entirely GCS for registry?
- integrate minio-packet.freedesktop.org into the cluster
- include runners in the cluster

Current issues

- logging/metrics not really stable
- we still have a lot of 502s
- can we ditch entirely GCS for registry?
- integrate minio-packet.freedesktop.org into the cluster
- include runners in the cluster

Logging/metrics not really stable

THE ADMIN IS NOT HAPPY :(

- deployed beginning of Sep 21 loki to replace Elastic Search
- still need to implement data retention rules

Summary:

- in progress, but seems to be better now

Current issues

- logging/metrics not really stable
- we still have a lot of 502s
- can we ditch entirely GCS for registry?
- integrate minio-packet.freedesktop.org into the cluster
- include runners in the cluster

We still have (had?) a lot of 502s

THE COMMUNITY IS (WAS) NOT HAPPY :(

- seems to be related to network
    - a heavy traffic on Object Storage tends to kill some disks
    - wireguard seems to be a good candidate for messing up the network
- or Ceph

We still have (had?) a lot of 502s (2/3)

What if it is network related?:

Planned actions:

- create VLANs on Equinix Metal just for kubernetes
- remove wireguard for the cluster and rely on the VLAN for privacy
- see if that helps
- need help

We still have (had?) a lot of 502s (3/3)

What if it is Ceph related?:

- done 🤦 (<- this is a facepam emoji):
  - tweaked the memory requests and limits for Ceph disks

Results are promising:

- 27 errors in the past 8 days
  - (last OOM event gave 35,142 errors over 30 minutes)

But I still want the network to be done!

Current issues

- logging/metrics not really stable
- we still have a lot of 502s
- can we ditch entirely GCS for registry?
- integrate minio-packet.freedesktop.org into the cluster
- include runners in the cluster

Can we ditch entirely GCS for registry?

THE ROAD TO $0 BILL (THE TREASURER WILL BE HAPPIER)

- in theory yes, but…
  - last year: 6TB
  - today: 10TB
- ongoing plan from GitLab to change the registry architecture (use of a proper db, gc while running)
- once that is in place, a registry migration will be required
- we’ll use that oportunity to bring back our data

Summary:

- scheduled for late 2021, if time permits and if GitLab manages to enable that
- scripts needed to chose which images are still valid (help needed)

Current issues

- logging/metrics not really stable
- we still have a lot of 502s
- can we ditch entirely GCS for registry?
- integrate minio-packet.freedesktop.org into the cluster
- include runners in the cluster

Machines sponsored by Equinix Metal

operation	#	type	proc	ram	disk
k8s cluster	3	c2.medium.x86	1 x AMD EPYC 7401P 24-Core Processor @ 2.0GHz	64GB	2 x 120GB SSD + 2 x 480GB SSD
	3	s1.large.x86	2 x Intel Xeon E5-2620V4 @2.1GHz	256GB	2 x 480 GB SSD + 12 x 2TB HDD + 1 x 128GB SSD
runner	3	m1.xlarge.x86	2 x Intel Xeon E5-2650 v4 @2.2GHz	256GB	6 x 480GB SSD
	2	c2.large.arm	1 x Ampere eMAG 8180 32-core @ 3.0Ghz	128GB	1 x 480GB SSD
minio-packet.fd.o	1	m1.xlarge.x86	2 x Intel Xeon E5-2650 v4 @2.2GHz	256GB	6 x 480GB SSD

Integrate minio-packet.fd.o into the cluster

MORE CI!!!

- still in a separate m1xl machine
- use MinIO + OPA

So far:

- MinIO on the new cluster doesn't work
- Ceph Storage can not work with GitLab JWT tokens
- OPA config is crashing Ceph

Solution:

- add KeyCloak to convert GitLab token into Keycloak tokens
- add proxy in front of Ceph Object Storage to talk to OPA (Istio?)
- help needed

Current issues

- logging/metrics not really stable
- we still have a lot of 502s
- can we ditch entirely GCS for registry?
- integrate minio-packet.freedesktop.org into the cluster
- include runners in the cluster

Include runners in the cluster

THE ADMINS WILL BE THANKFUL

why?

- runners are still manually administered (automatically deployed though)
  - a change in the config requires manual intervention
  - an upgrade requires manual intervention (times 5)
- metrics!

Include runners in the cluster

THE ADMINS WILL BE THANKFUL

but?

- can’t have privileged runners anymore
  - docker-in-docker will not work in the near future
    - please use ci-templates
- need minio-packet.fd.o to release its machine

actions needed

- write a k8s deployment for them
  - can’t use k8s deployment from GitLab
- ensure the runners jobs are on a different VLAN than the main cluster

Lessons learned

Lessons learned

- GCP gives a lot of benefits
- GCP has a cost
- still not in a perfect cluster environment
- FDO community is awesome
- Equinix Metal is awesome too

Thank you

Lessons learned

- GCP gives a lot of benefits
- GCP has a cost
- still not in a perfect cluster environment
- FDO community is awesome
- Equinix Metal is awesome too