

# Status of freedesktop.org gitlab/cloud hosting

# Hello

I am Benjamin Tissoires (bentiss)

I am here because I am a (recent)  
freedesktop admin

(I also work at Red Hat)



Why?



<https://lists.freedesktop.org/archives/wayland-devel/2020-February/041232.html>

“The good news: gitlab.fd.o has become very popular with our communities, and is used extensively. This especially includes all the CI integration. Modern development process and tooling, yay!

Daniel Vetter,  
in a very alarmous email

<https://lists.freedesktop.org/archives/wayland-devel/2020-February/041232.html>

“The good news: gitlab.fd.o has become very popular with our communities, and is used extensively. This especially includes all the CI integration. Modern development process and tooling, yay!

The bad news: The cost in growth has also been tremendous, and it's breaking our bank account.

Daniel Vetter,  
in a very alarmous email

<https://lists.freedesktop.org/archives/wayland-devel/2020-February/041232.html>

“The good news: gitlab.fd.o has become very popular with our communities, and is used extensively. This especially includes all the CI integration. Modern development process and tooling, yay!

The bad news: The cost in growth has also been tremendous, and it's breaking our bank account.

[...]

Assuming that we want cash flow reserves for one year of gitlab.fd.o (without CI support) and a trimmed XDC and assuming no sponsor payment meanwhile, we'd have to cut CI services somewhere between May and June this year.”

*TL;DR: we are in a better shape, and we have a plan*

Daniel Vetter,  
in a very alarmous email

# A little bit of history of gitlab.fd.o

Let us start with the beginning of everything



# Before December 2017

<https://www.fooishbar.org/blog/gitlab-fdo-introduction/>

- Hardware was reaching EOL
- Multiple services independent of each other
- A lot of manual admin interventions for every tiny bit of operation

Projects were fleeing fd.o hosting





# Admins are lazy people (it's a good thing)

- Gitlab CE was solving many of these issues
  - Can run on Google Cloud Platform
  - Integrates issues, code, wiki and CI
  - No more admin bottlenecks, project leaders are "admins" of their own projects
- Gitlab Inc. could help @daniels setting things up
- GNOME went through a similar process



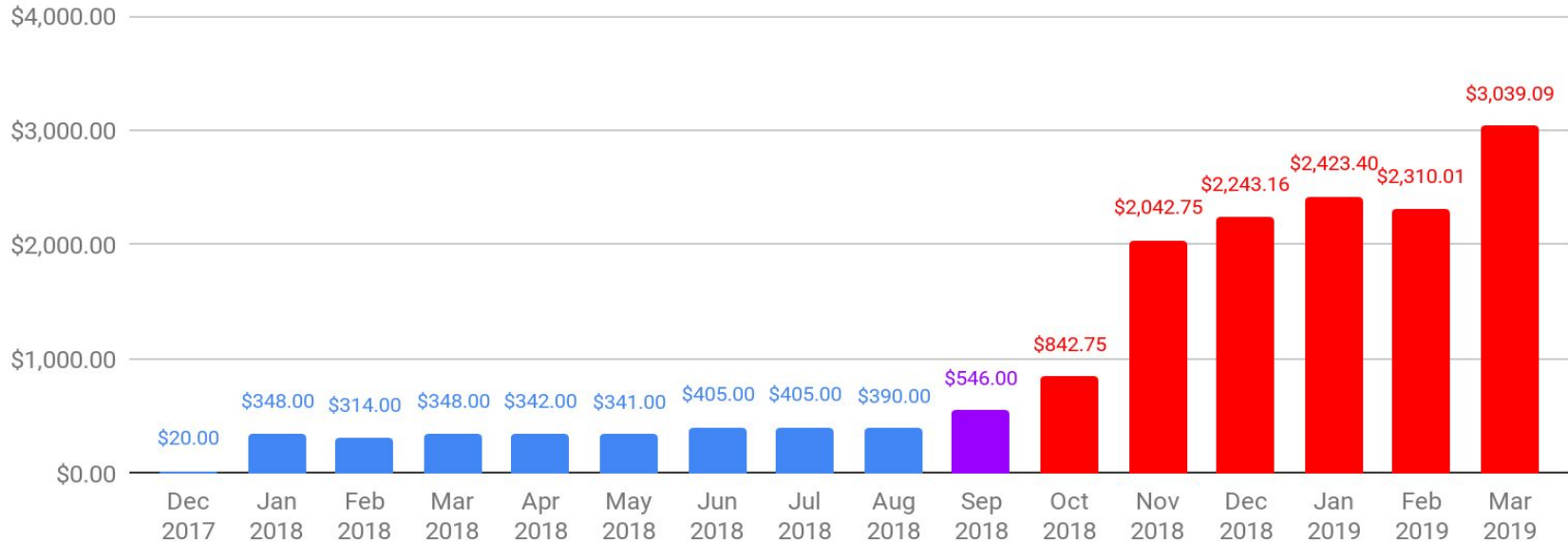
# Dec 2017 - Sep 2018: initial deployment of gitlab.fd.o

- Few volunteered projects who could deal with some downtime
- CI runners deployed, but not very production ready
- Billed directly on daniels' bank account, but reimbursed by Gitlab inc.
- Costs were reasonable



# Sept 2018: SPI got billed directly

## SPI billing



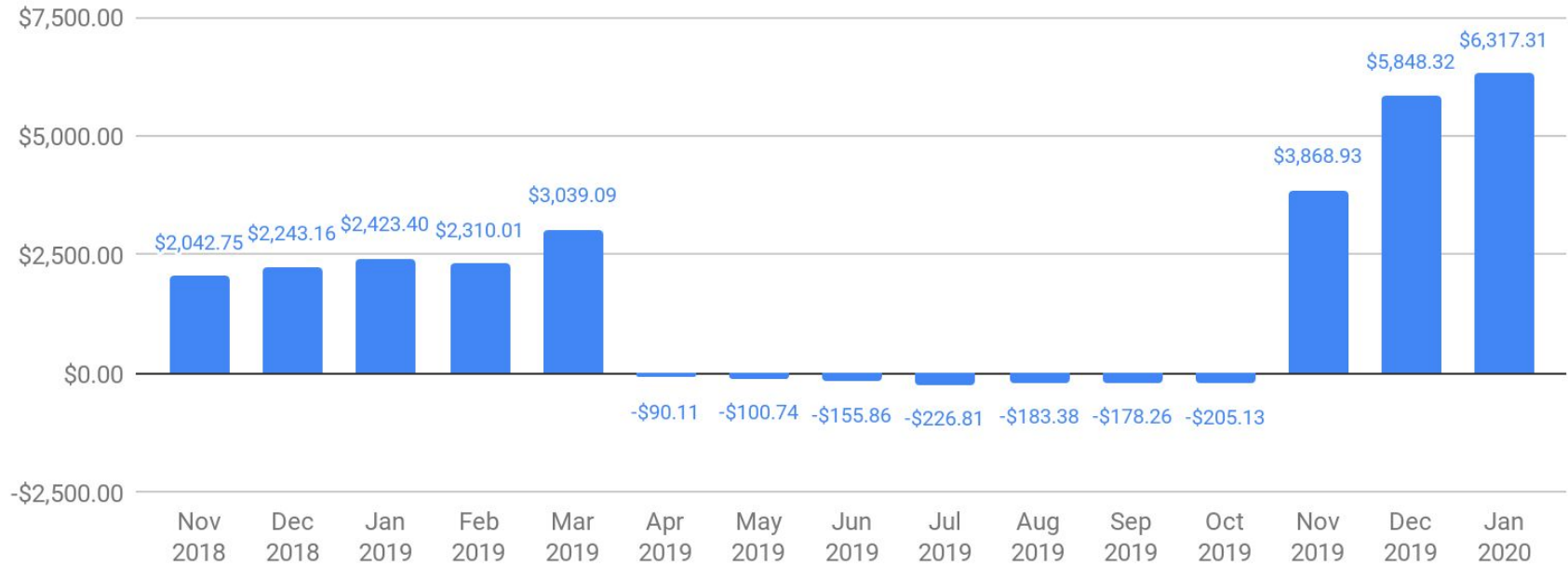
# Reasons for growth

- New projects are added (mesa)
- CI starts to kick in
- \$15K \$20K is roughly the yearly estimate
- Partly reimbursed by Gitlab Inc. (\$9K in total)



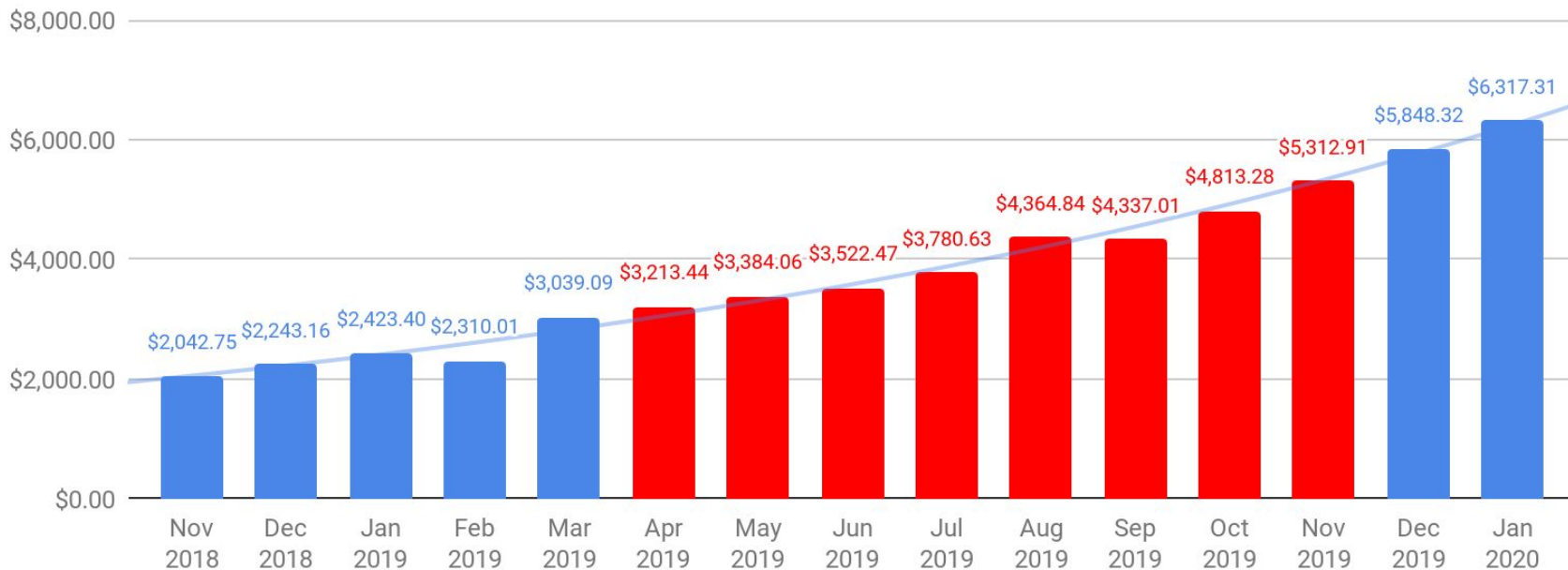
# Then came the \$30K grant from Google

## SPI billing



# Real numbers

## SPI billing



How do we get to  
\$6317.31 a month?

Jan 2019 example



SKU	Product	Usage	Cost	Discounts	Promotions and others	Subtotal
Download Worldwide Destinations (excluding Asia & Australia)	Cloud Storage	15,896.59 gibibyte	\$1,589.05	\$0.00	\$0.00	\$1,589.05
Network Internet Egress from Americas to EMEA	Compute Engine	12,393.96 gibibyte	\$1,308.84	\$0.00	\$0.00	\$1,308.84
N1 Predefined Instance Core running in Americas	Compute Engine	35,711.73 hour	\$1,128.88	-\$338.66	\$0.00	\$790.22
Network Internet Egress from Americas to Americas	Compute Engine	5,164.15 gibibyte	\$578.18	\$0.00	\$0.00	\$578.18
SSD backed PD Capacity	Compute Engine	2,430.71 gibibyte month	\$413.22	\$0.00	\$0.00	\$413.22
N1 Predefined Instance Ram running in Americas	Compute Engine	133,919 gibibyte hour	\$567.41	-\$170.22	\$0.00	\$397.19
Custom Instance Core running in Americas	Compute Engine	16,368 hour	\$542.99	-\$162.90	\$0.00	\$380.09
Log Volume	Stackdriver Logging	484.01 gibibyte	\$217.00	\$0.00	\$0.00	\$217.00
Custom Instance Ram running in Americas	Compute Engine	50,592 gibibyte hour	\$224.93	-\$67.48	\$0.00	\$157.45
Standard Storage US Regional	Cloud Storage	7,162.82 gibibyte month	\$143.16	\$0.00	\$0.00	\$143.16
SSD backed Local Storage	Compute Engine	1,500 gibibyte month	\$120.00	\$0.00	\$0.00	\$120.00
Network Load Balancing: Data Processing Charge in Americas	Compute Engine	11,244.14 gibibyte	\$89.95	\$0.00	\$0.00	\$89.95
Network Inter Zone Egress	Compute Engine	3,581.86 gibibyte	\$35.82	\$0.00	\$0.00	\$35.82
Network Internet Egress from Americas to APAC	Compute Engine	193.37 gibibyte	\$23.08	\$0.00	\$0.00	\$23.08
Network Internet Egress from Americas to China	Compute Engine	82.82 gibibyte	\$19.05	\$0.00	\$0.00	\$19.05
Network Load Balancing: Forwarding Rule Minimum Service Charge in Americas	Compute Engine	744 hour	\$18.60	\$0.00	\$0.00	\$18.60
Network Internet Egress from Americas to Australia	Compute Engine	49.96 gibibyte	\$9.49	\$0.00	\$0.00	\$9.49
Static Ip Charge	Compute Engine	744 hour	\$7.43	\$0.00	\$0.00	\$7.43
Download Australia	Cloud Storage	35.45 gibibyte	\$6.74	\$0.00	\$0.00	\$6.74
Regional Standard Class A Operations	Cloud Storage	1,181,031 count	\$5.88	\$0.00	\$0.00	\$5.88
Download APAC	Cloud Storage	31.89 gibibyte	\$3.71	\$0.00	\$0.00	\$3.71
Download China	Cloud Storage	7.77 gibibyte	\$1.79	\$0.00	\$0.00	\$1.79
Regional Standard Class B Operations	Cloud Storage	2,513,244 count	\$0.99	\$0.00	\$0.00	\$0.99
Network Inter Region Egress from Americas to Netherlands	Compute Engine	4.13 gibibyte	\$0.25	\$0.00	\$0.00	\$0.25
Network Inter Region Egress from Americas to Americas	Compute Engine	12.67 gibibyte	\$0.12	\$0.00	\$0.00	\$0.12
Network Egress via Carrier Peering Network - Americas Based	Compute Engine	0.38 gibibyte	\$0.02	\$0.00	\$0.00	\$0.02
Inter-region GCP Storage egress within NA	Cloud Storage	1.65 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01
Network Inter Region Egress from Americas to Virginia	Compute Engine	0.59 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01





SKU	Product	Usage	Cost	Discounts	Promotions and others	Subtotal
Download Worldwide Destinations (excluding Asia & Australia)	Cloud Storage	15,896.59 gibibyte	\$1,589.05	\$0.00	\$0.00	\$1,589.05
Network Internet Egress from Americas to EMEA	Compute Engine	12,393.96 gibibyte	\$1,308.84	\$0.00	\$0.00	\$1,308.84
N1 Predefined Instance Core running in Americas	Compute Engine	35,711.73 hour	\$1,128.88	-\$338.66	\$0.00	\$790.22
Network Internet Egress from Americas to Americas	Compute Engine	5,164.15 gibibyte	\$578.18	\$0.00	\$0.00	\$578.18
SSD backed PD Capacity	Compute Engine	2,430.71 gibibyte month	\$413.22	\$0.00	\$0.00	\$413.22
N1 Predefined Instance Ram running in Americas	Compute Engine	133,919 gibibyte hour	\$567.41	-\$170.22	\$0.00	\$397.19
Custom Instance Core running in Americas	Compute Engine	16,368 hour	\$542.99	-\$162.90	\$0.00	\$380.09
Log Volume	Stackdriver Logging	484.01 gibibyte	\$217.00	\$0.00	\$0.00	\$217.00
Custom Instance Ram running in Americas	Compute Engine	50,592 gibibyte hour	\$224.93	-\$67.48	\$0.00	\$157.45
Standard Storage US Regional	Cloud Storage	7,162.82 gibibyte month	\$143.16	\$0.00	\$0.00	\$143.16
SSD backed Local Storage	Compute Engine	1,500 gibibyte month	\$120.00	\$0.00	\$0.00	\$120.00
Network Load Balancing: Data Processing Charge in Americas	Compute Engine	11,244.14 gibibyte	\$89.95	\$0.00	\$0.00	\$89.95
Network Inter Zone Egress	Compute Engine	3,581.86 gibibyte	\$35.82	\$0.00	\$0.00	\$35.82
Network Internet Egress from Americas to APAC	Compute Engine	193.37 gibibyte	\$23.08	\$0.00	\$0.00	\$23.08
Network Internet Egress from Americas to China	Compute Engine	82.82 gibibyte	\$19.05	\$0.00	\$0.00	\$19.05
Network Load Balancing: Forwarding Rule Minimum Service Charge in Americas	Compute Engine	744 hour	\$18.60	\$0.00	\$0.00	\$18.60
Network Internet Egress from Americas to Australia	Compute Engine	49.96 gibibyte	\$9.49	\$0.00	\$0.00	\$9.49
Static Ip Charge	Compute Engine	744 hour	\$7.43	\$0.00	\$0.00	\$7.43
Download Australia	Cloud Storage	35.45 gibibyte	\$6.74	\$0.00	\$0.00	\$6.74
Regional Standard Class A Operations	Cloud Storage	1,181,031 count	\$5.88	\$0.00	\$0.00	\$5.88
Download APAC	Cloud Storage	31.89 gibibyte	\$3.71	\$0.00	\$0.00	\$3.71
Download China	Cloud Storage	7.77 gibibyte	\$1.79	\$0.00	\$0.00	\$1.79
Regional Standard Class B Operations	Cloud Storage	2,513,244 count	\$0.99	\$0.00	\$0.00	\$0.99
Network Inter Region Egress from Americas to Netherlands	Compute Engine	4.13 gibibyte	\$0.25	\$0.00	\$0.00	\$0.25
Network Inter Region Egress from Americas to Americas	Compute Engine	12.67 gibibyte	\$0.12	\$0.00	\$0.00	\$0.12
Network Egress via Carrier Peering Network - Americas Based	Compute Engine	0.38 gibibyte	\$0.02	\$0.00	\$0.00	\$0.02
Inter-region GCP Storage egress within NA	Cloud Storage	1.65 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01
Network Inter Region Egress from Americas to Virginia	Compute Engine	0.59 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01



# Breakdown

- Networking costs: \$3,699.01
  - ???
- Compute costs: \$2,258.17
  - "Hardware" choices
- Kubernetes costs: \$360.16
  - Mandatory (logging and cloud storage)
- Total: \$6,317.34



SKU	Product	Usage	Cost	Discounts	Promotions and others	Subtotal
Download Worldwide Destinations (excluding Asia & Australia)	Cloud Storage	15,896.59 gibibyte	\$1,589.05	\$0.00	\$0.00	\$1,589.05
Network Internet Egress from Americas to EMEA	Compute Engine	12,393.96 gibibyte	\$1,308.84	\$0.00	\$0.00	\$1,308.84
N1 Predefined Instance Core running in Americas	Compute Engine	35,711.73 hour	\$1,128.88	-\$338.66	\$0.00	\$790.22
Network Internet Egress from Americas to Americas	Compute Engine	5,164.15 gibibyte	\$578.18	\$0.00	\$0.00	\$578.18
SSD backed PD Capacity	Compute Engine	2,430.71 gibibyte month	\$413.22	\$0.00	\$0.00	\$413.22
N1 Predefined Instance Ram running in Americas	Compute Engine	133,919 gibibyte hour	\$567.41	-\$170.22	\$0.00	\$397.19
Custom Instance Core running in Americas	Compute Engine	16,368 hour	\$542.99	-\$162.90	\$0.00	\$380.09
Log Volume	Stackdriver Logging	484.01 gibibyte	\$217.00	\$0.00	\$0.00	\$217.00
Custom Instance Ram running in Americas	Compute Engine	50,592 gibibyte hour	\$224.93	-\$67.48	\$0.00	\$157.45
Standard Storage US Regional	Cloud Storage	7,162.82 gibibyte month	\$143.16	\$0.00	\$0.00	\$143.16
SSD backed Local Storage	Compute Engine	1,500 gibibyte month	\$120.00	\$0.00	\$0.00	\$120.00
Network Load Balancing: Data Processing Charge in Americas	Compute Engine	11,244.14 gibibyte	\$89.95	\$0.00	\$0.00	\$89.95
Network Inter Zone Egress	Compute Engine	3,581.86 gibibyte	\$35.82	\$0.00	\$0.00	\$35.82
Network Internet Egress from Americas to APAC	Compute Engine	193.37 gibibyte	\$23.08	\$0.00	\$0.00	\$23.08
Network Internet Egress from Americas to China	Compute Engine	82.82 gibibyte	\$19.05	\$0.00	\$0.00	\$19.05
Network Load Balancing: Forwarding Rule Minimum Service Charge in Americas	Compute Engine	744 hour	\$18.60	\$0.00	\$0.00	\$18.60
Network Internet Egress from Americas to Australia	Compute Engine	49.96 gibibyte	\$9.49	\$0.00	\$0.00	\$9.49
Static Ip Charge	Compute Engine	744 hour	\$7.43	\$0.00	\$0.00	\$7.43
Download Australia	Cloud Storage	35.45 gibibyte	\$6.74	\$0.00	\$0.00	\$6.74
Regional Standard Class A Operations	Cloud Storage	1,181,031 count	\$5.88	\$0.00	\$0.00	\$5.88
Download APAC	Cloud Storage	31.89 gibibyte	\$3.71	\$0.00	\$0.00	\$3.71
Download China	Cloud Storage	7.77 gibibyte	\$1.79	\$0.00	\$0.00	\$1.79
Regional Standard Class B Operations	Cloud Storage	2,513,244 count	\$0.99	\$0.00	\$0.00	\$0.99
Network Inter Region Egress from Americas to Netherlands	Compute Engine	4.13 gibibyte	\$0.25	\$0.00	\$0.00	\$0.25
Network Inter Region Egress from Americas to Americas	Compute Engine	12.67 gibibyte	\$0.12	\$0.00	\$0.00	\$0.12
Network Egress via Carrier Peering Network - Americas Based	Compute Engine	0.38 gibibyte	\$0.02	\$0.00	\$0.00	\$0.02
Inter-region GCP Storage egress within NA	Cloud Storage	1.65 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01
Network Inter Region Egress from Americas to Virginia	Compute Engine	0.59 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01



SKU	Product	Usage	Cost	Discounts	Promotions and others	Subtotal
Download Worldwide Destinations (excluding Asia & Australia)	Cloud Storage	15,896.59 gibibyte	\$1,589.05	\$0.00	\$0.00	\$1,589.05
Network Internet Egress from Americas to EMEA	Compute Engine	12,393.96 gibibyte	\$1,308.84	\$0.00	\$0.00	\$1,308.84
N1 Predefined Instance Core running in Americas	Compute Engine	35,711.73 hour	\$1,128.88	-\$338.66	\$0.00	\$790.22
Network Internet Egress from Americas to Americas	Compute Engine	5,164.15 gibibyte	\$578.18	\$0.00	\$0.00	\$578.18
SSD backed PD Capacity	Compute Engine	2,430.71 gibibyte month	\$413.22	\$0.00	\$0.00	\$413.22
N1 Predefined Instance Ram running in Americas	Compute Engine	133,919 gibibyte hour	\$567.41	-\$170.22	\$0.00	\$397.19
Custom Instance Core running in Americas	Compute Engine	16,368 hour	\$542.99	-\$162.90	\$0.00	\$380.09
Log Volume	Stackdriver Logging	484.01 gibibyte	\$217.00	\$0.00	\$0.00	\$217.00
Custom Instance Ram running in Americas	Compute Engine	50,592 gibibyte hour	\$224.93	-\$67.48	\$0.00	\$157.45
Standard Storage US Regional	Cloud Storage	7,162.82 gibibyte month	\$143.16	\$0.00	\$0.00	\$143.16
SSD backed Local Storage	Compute Engine	1,500 gibibyte month	\$120.00	\$0.00	\$0.00	\$120.00
Network Load Balancing: Data Processing Charge in Americas	Compute Engine	11,244.14 gibibyte	\$89.95	\$0.00	\$0.00	\$89.95
Network Inter Zone Egress	Compute Engine	3,581.86 gibibyte	\$35.82	\$0.00	\$0.00	\$35.82
Network Internet Egress from Americas to APAC	Compute Engine	193.37 gibibyte	\$23.08	\$0.00	\$0.00	\$23.08
Network Internet Egress from Americas to China	Compute Engine	82.82 gibibyte	\$19.05	\$0.00	\$0.00	\$19.05
Network Load Balancing: Forwarding Rule Minimum Service Charge in Americas	Compute Engine	744 hour	\$18.60	\$0.00	\$0.00	\$18.60
Network Internet Egress from Americas to Australia	Compute Engine	49.96 gibibyte	\$9.49	\$0.00	\$0.00	\$9.49
Static Ip Charge	Compute Engine	744 hour	\$7.43	\$0.00	\$0.00	\$7.43
Download Australia	Cloud Storage	35.45 gibibyte	\$6.74	\$0.00	\$0.00	\$6.74
Regional Standard Class A Operations	Cloud Storage	1,181,031 count	\$5.88	\$0.00	\$0.00	\$5.88
Download APAC	Cloud Storage	31.89 gibibyte	\$3.71	\$0.00	\$0.00	\$3.71
Download China	Cloud Storage	7.77 gibibyte	\$1.79	\$0.00	\$0.00	\$1.79
Regional Standard Class B Operations	Cloud Storage	2,513,244 count	\$0.99	\$0.00	\$0.00	\$0.99
Network Inter Region Egress from Americas to Netherlands	Compute Engine	4.13 gibibyte	\$0.25	\$0.00	\$0.00	\$0.25
Network Inter Region Egress from Americas to Americas	Compute Engine	12.67 gibibyte	\$0.12	\$0.00	\$0.00	\$0.12
Network Egress via Carrier Peering Network - Americas Based	Compute Engine	0.38 gibibyte	\$0.02	\$0.00	\$0.00	\$0.02
Inter-region GCP Storage egress within NA	Cloud Storage	1.65 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01
Network Inter Region Egress from Americas to Virginia	Compute Engine	0.59 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01



Download Worldwide Destinations (excluding Asia & Australia)	Cloud Storage	15,896.59 gibibyte	\$1,589.05	\$0.00	\$0.00	\$1,589.05
Download Australia	Cloud Storage	35.45 gibibyte	\$6.74	\$0.00	\$0.00	\$6.74
Regional Standard Class A Operations	Cloud Storage	1,181,031 count	\$5.88	\$0.00	\$0.00	\$5.88
Download APAC	Cloud Storage	31.89 gibibyte	\$3.71	\$0.00	\$0.00	\$3.71
Download China	Cloud Storage	7.77 gibibyte	\$1.79	\$0.00	\$0.00	\$1.79
Regional Standard Class B Operations	Cloud Storage	2,513,244 count	\$0.99	\$0.00	\$0.00	\$0.99
Inter-region GCP Storage egress within NA	Cloud Storage	1.65 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01
<b>total</b>						<b>\$1,608.17</b>
Network Internet Egress from Americas to EMEA	Compute Engine	12,393.96 gibibyte	\$1,308.84	\$0.00	\$0.00	\$1,308.84
Network Internet Egress from Americas to Americas	Compute Engine	5,164.15 gibibyte	\$578.18	\$0.00	\$0.00	\$578.18
Network Load Balancing: Data Processing Charge in Americas	Compute Engine	11,244.14 gibibyte	\$89.95	\$0.00	\$0.00	\$89.95
Network Inter Zone Egress	Compute Engine	3,581.86 gibibyte	\$35.82	\$0.00	\$0.00	\$35.82
Network Internet Egress from Americas to APAC	Compute Engine	193.37 gibibyte	\$23.08	\$0.00	\$0.00	\$23.08
Network Internet Egress from Americas to China	Compute Engine	82.82 gibibyte	\$19.05	\$0.00	\$0.00	\$19.05
Network Load Balancing: Forwarding Rule Minimum Service Charge in Americas	Compute Engine	744 hour	\$18.60	\$0.00	\$0.00	\$18.60
Network Internet Egress from Americas to Australia	Compute Engine	49.96 gibibyte	\$9.49	\$0.00	\$0.00	\$9.49
Static Ip Charge	Compute Engine	744 hour	\$7.43	\$0.00	\$0.00	\$7.43
Network Inter Region Egress from Americas to Netherlands	Compute Engine	4.13 gibibyte	\$0.25	\$0.00	\$0.00	\$0.25
Network Inter Region Egress from Americas to Americas	Compute Engine	12.67 gibibyte	\$0.12	\$0.00	\$0.00	\$0.12
Network Egress via Carrier Peering Network - Americas Based	Compute Engine	0.38 gibibyte	\$0.02	\$0.00	\$0.00	\$0.02
Network Inter Region Egress from Americas to Virginia	Compute Engine	0.59 gibibyte	\$0.01	\$0.00	\$0.00	\$0.01
<b>total</b>						<b>\$2,090.84</b>

*Download*: data counted from Google Cloud Platform to the clients

*Egress*: data counted from the compute nodes to the clients (IOW download)

*Ingress*: data counted from the client to the compute nodes (IOW upload)

*Node*: a machine in the cluster (being a VM or bare metal)

# Network Breakdown

- **Compute Engine** costs: \$2090.84
  - Anything that goes out of gitlab.freedesktop.org
    - Web
    - Git
    - API
    - etc...
- **Cloud Storage** costs: \$1608.17
  - Coming from Google cloud buckets:
    - Artifacts
    - Registry
    - LFS
    - Uploads



# What could be done?

Compute Engine networking costs:

- local addresses in the logs

Can't do much at that time

# What could be done?

Cloud Storage networking costs:

- Well, no useful information from the installation
- There must be a way



# What was done?

Storage:

- Daniels enabled the logs (CSV format)
  - `gsutil logging set on -b gs://example-logs-bucket [-o log_object_prefix ]`  
`gs://example-bucket`

# What was done?

## Storage:

- Daniels enabled the logs (CSV format)

- `gsutil logging set on -b gs://example-logs-bucket [-o log_object_prefix ]`

- `gs://example-bucket`

- Example:

- `"Time_micros", "c_ip", "c_ip_type", "c_ip_region", "cs_method", "cs_uri", "sc_status", "cs_bytes", "sc_bytes", "time_taken_micros", "cs_host", "cs_referer", "cs_user_agent", "s_request_id", "cs_operation", "cs_bucket", "cs_object"`

- `"1582886759757866", "XX.YY.ZZ.AA", "1", "", "GET", "/fdo-gitlab-artifacts/ed/6f/ed6f2060379b3b2192ea06e51b6e771f0ef931e9efc702f767da99869d4e67f7/2020_02_28/1763943/2599416/junit.xml.gz?GoogleAccessId=G00G0EXJKVLVDILPMJ5V2WSY&Signature=qJeNdfQKao4viTBbi7d%2FZBWoH8c%3D&Expires=1582887359", "200", "0", "4865", "16000", "storage.googleapis.com", "", "Ruby", "AEnB2UrF4yH4mEPZI_iGd_GaLmsOUPQ1E3Rlw40yswzQqGBo3yfD0eVCFxkWrJrkLcQIPY9ZKaC9Tti0XHWxYPtESZDBl-_RWQ", "GET_Object", "fdo-gitlab-artifacts", "ed/6f/ed6f2060379b3b2192ea06e51b6e771f0ef931e9efc702f767da99869d4e67f7/2020_02_28/1763943/2599416/junit.xml.gz"`



# then...

<https://gitlab.freedesktop.org/freedesktop/freedesktop/-/issues/247>

“We have a massive growth in usage, and we need to figure out whom to blame for this so that optimization efforts are actually directed where we'd benefit most, instead of just wild guesses.

# then...

<https://gitlab.freedesktop.org/freedesktop/freedesktop/-/issues/247>

“We have a massive growth in usage, and we need to figure out whom to blame for this so that optimization efforts are actually directed where we'd benefit most, instead of just wild guesses.

First priority is google cloud storage downloads (since that's growing the quickest). We'd want to be able to assign used network bw both to projects (to know where to cut artifact sizes) and to external CI runner labs (in order to know where we really need better caching).”

# then...

<https://gitlab.freedesktop.org/freedesktop/freedesktop/-/issues/247>

“We have a massive growth in usage, and we need to figure out whom to blame for this so that optimization efforts are actually directed where we'd benefit most, instead of just wild guesses.

First priority is google cloud storage downloads (since that's growing the quickest). We'd want to be able to assign used network bw both to projects (to know where to cut artifact sizes) and to external CI runner labs (in order to know where we really need better caching).”

bentiss: “If it's not too big of a task, I wouldn't mind volunteering to try to get some info out of it.”

That's how I got on board



**freedesktop.org**

# Cloud Storage

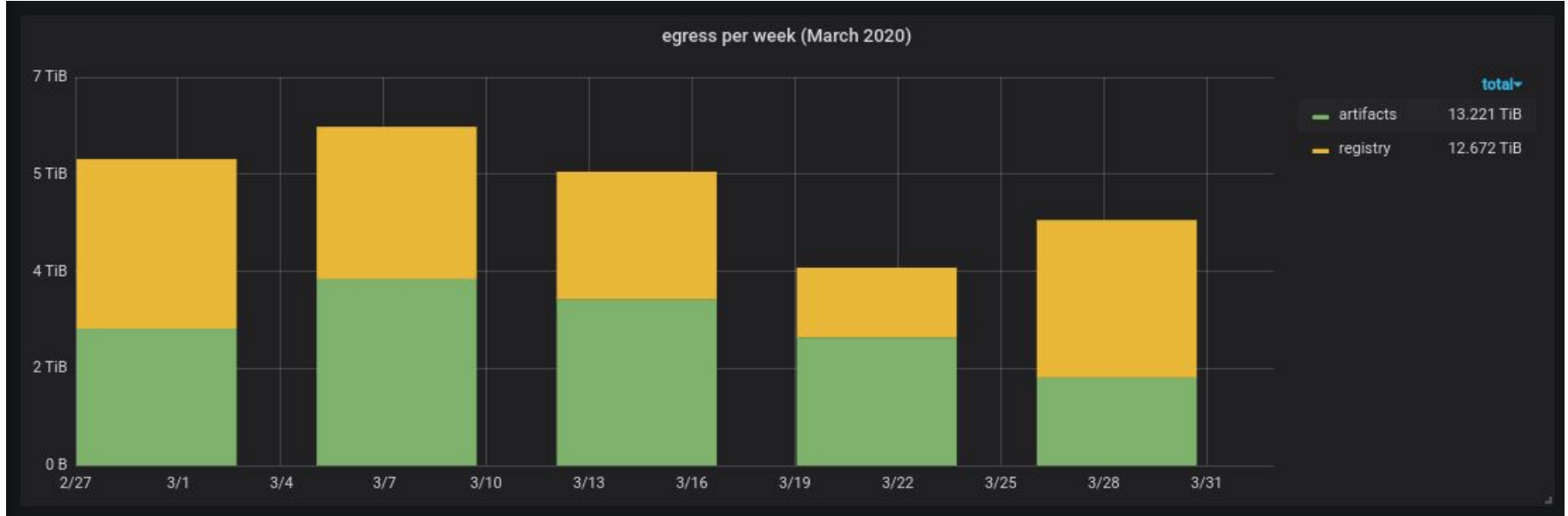


# First: Analysis

- Provide a python script to parse the Google logs
  - <https://gitlab.freedesktop.org/freedesktop/fdo-logs-analysis>
- Installed Grafana to understand and publish the data
  - <https://grafana.freedesktop.org/>
  - Installed InfluxDB to store the data
    - Time series database
    - Initially tried Graphite, didn't work well enough for us



# Analysis



Roughly 50% artifacts, 50% registry



# Registry

Reasons for the high registry pulls:

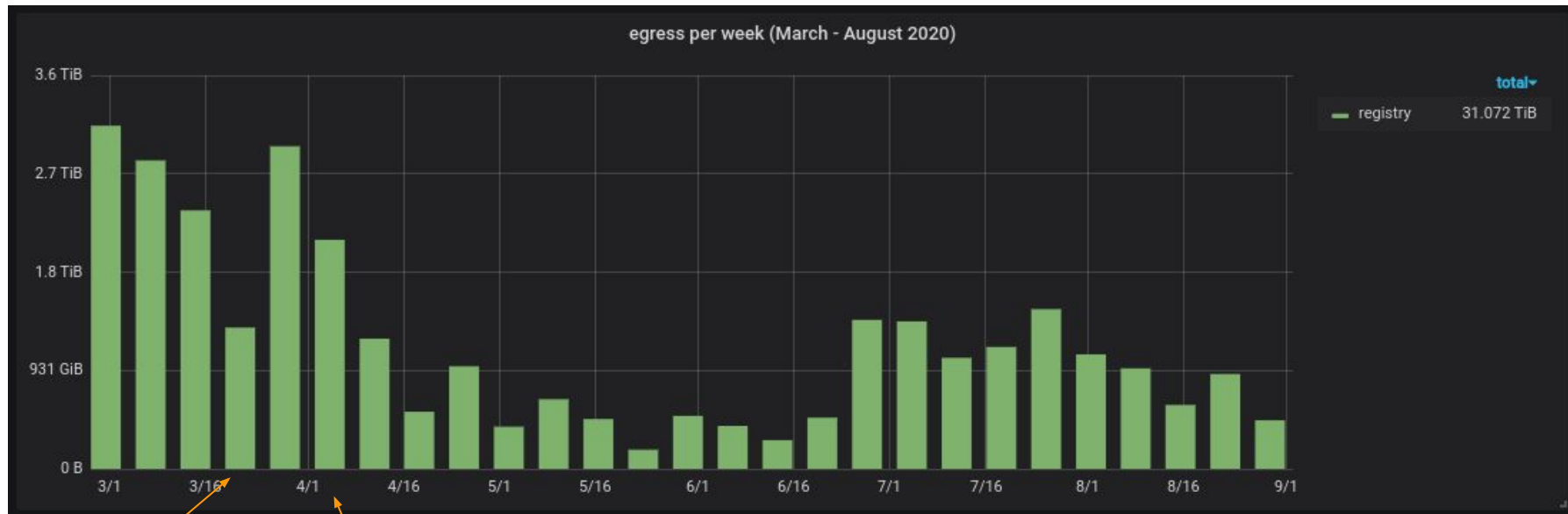
- CI runners not optimized
  - No RAID configured (so small disks)
  - Hard to garbage collect the images on the runners when we are running out of space

# Registry remediations

- No RAID:
  - Configured raid
    - Went from 438G in '/' to dedicated 1.3TB for docker per runner on packet
- Registry images garbage collect:
  - Wrote `docker-free-space.py` with daniels:
    - Take the image usage into account
      - (i.e. often used images are less likely to be deleted)
    - Implement some custom labels support (`fdo.expires-after`, `fdo.upstream-repo`)
    - Run on docker events, not a cron job



# docker-free-space.py impact



RAID activated

d-f-s enabled

Diet period

We don't really care anymore...



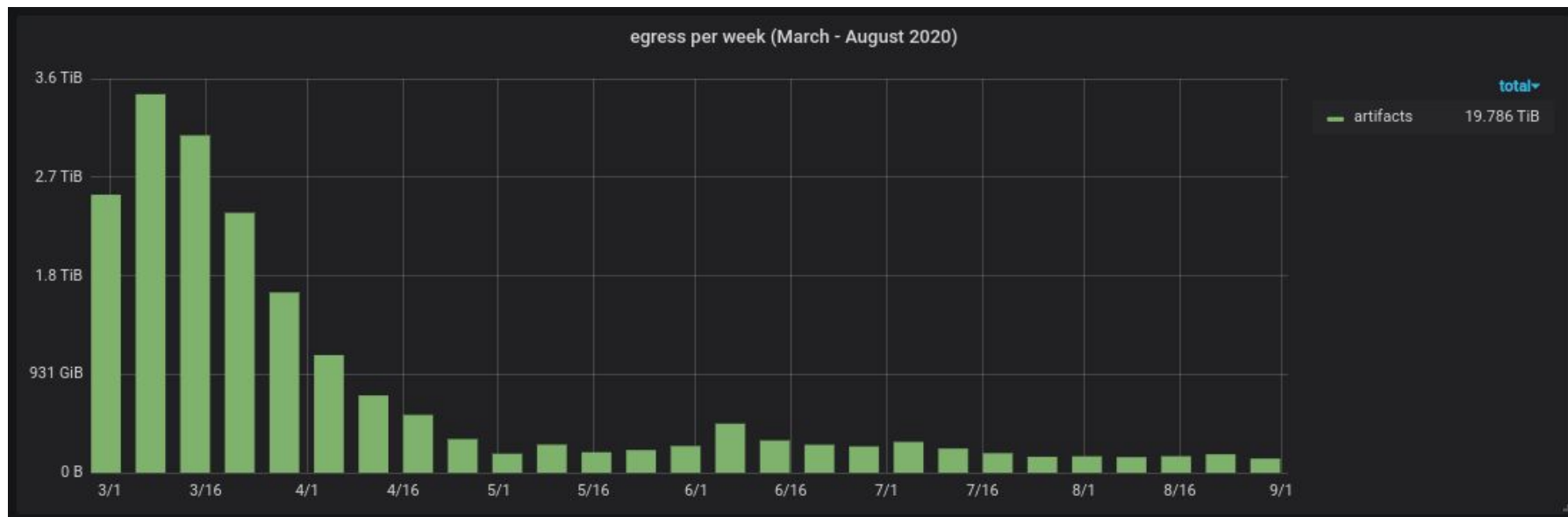
# Artifacts

Reasons:

- Well, nobody cared about it before

# Artifacts remediations

- Kudos to the biggest artifacts project users (mesa + gstreamer):
  - From 3.5 TB per week to a steady 150 GB per week



# Compute Engine

Or cloud infrastructure to the rest of the world



# Some definitions (Cloud 101)

- kubernetes is often written *k8s*
- in k8s, Helm is often used to deploy "applications".
  - Definitions are called "charts"
- Egress: data going from the compute nodes to the clients (IOW download)
- Ingress: data going from the client to the compute nodes (IOW upload)
- Node: a machine in the cluster running multiple pods (being a VM or bare metal)
- Pod: a set of containers running in the same isolated environment



# State as of March 2020

Compute:

- Running an old deprecated chart, with outdated components:
  - Nginx proxy was returning local addresses in the logs
  - Use of a soon to be deprecated API for Let's Encrypt SSL certificates
  - All-in-one Gitlab Omnibus install
    - need a big machine
    - can not split the service across several nodes
- But on the plus side
  - External postgresql, redis, nginx and Let's Encrypt pod definitions



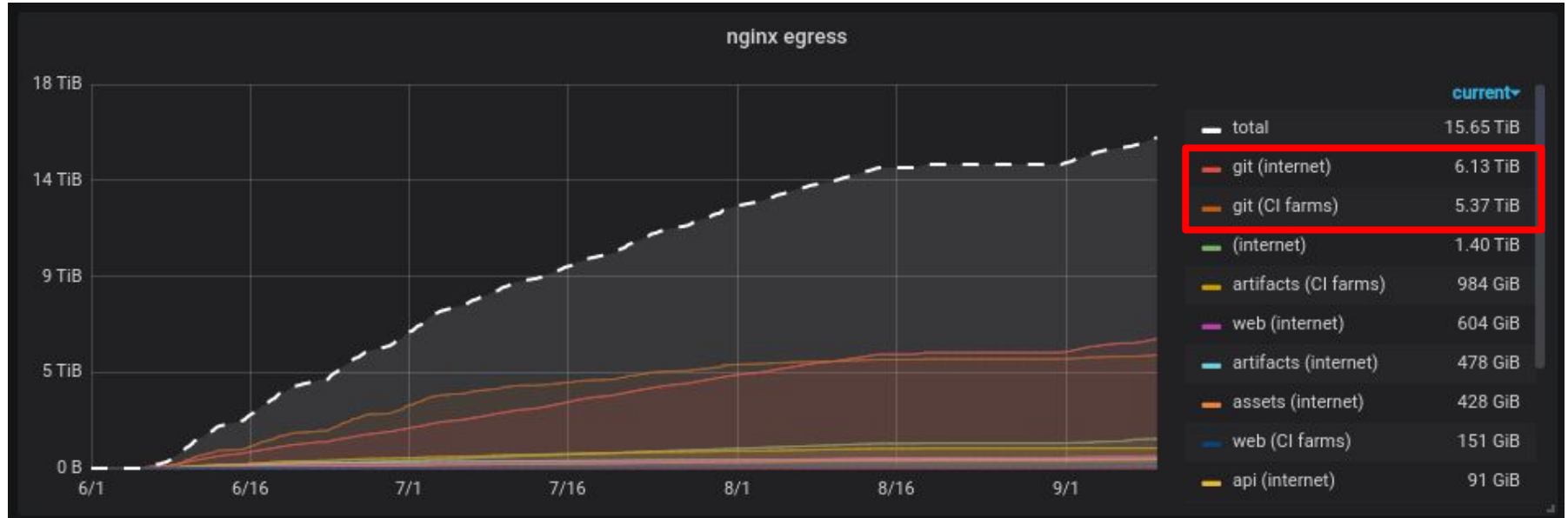


# Steps taken

1. Updated Nginx
  - Now true IP addresses
2. Attempted to rely on fancy domain addresses
  - gitlab-packet-new-dfs.freedesktop.org, gitlab-hetzner.fd.o, etc..
  - Didn't worked as much as we wanted
3. Wrote an other script to parse the logs
  - Now we know which project / operation is using data :-)



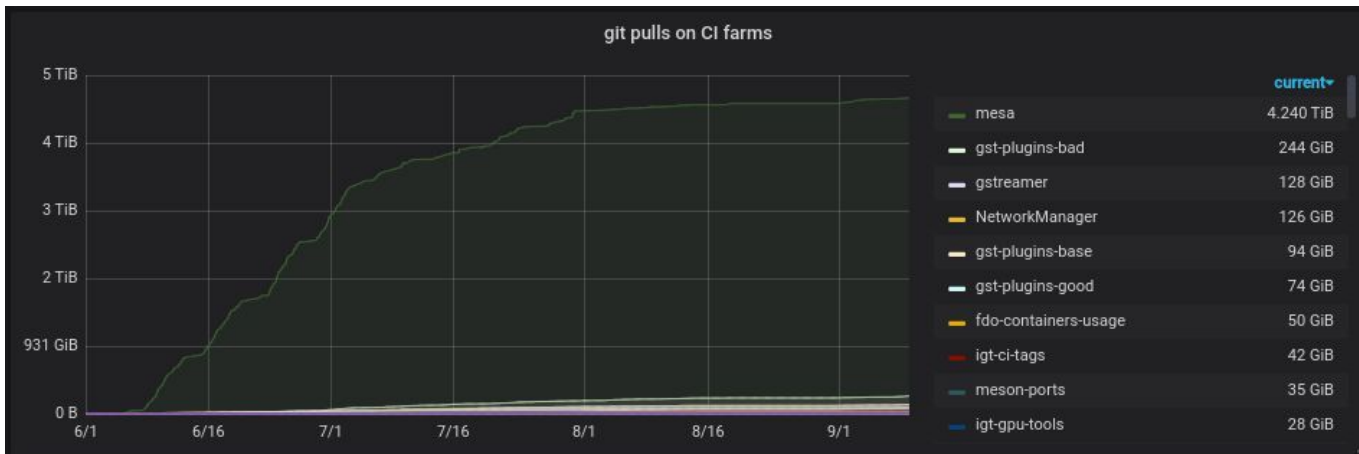
# Results



Script died while I was on vacations

# Git usage

- From "internet":
  - There was a bug in the git transport protocol v2,
    - downloading way too much data on simple git pulls
  - Not much we can do
- From "ci farms":



# Mesa git usage remediations

- Installed on packet.com a file server with S3-like API: MinIO
- Changed the runners config to execute a script before pulling the git tree
- Changed the .gitlab-ci.yaml file to use a cached version of the tree
  - Regenerated every day
  - <http://minio-packet.freedesktop.org/git-cache/mesa/mesa/mesa.tar.gz>

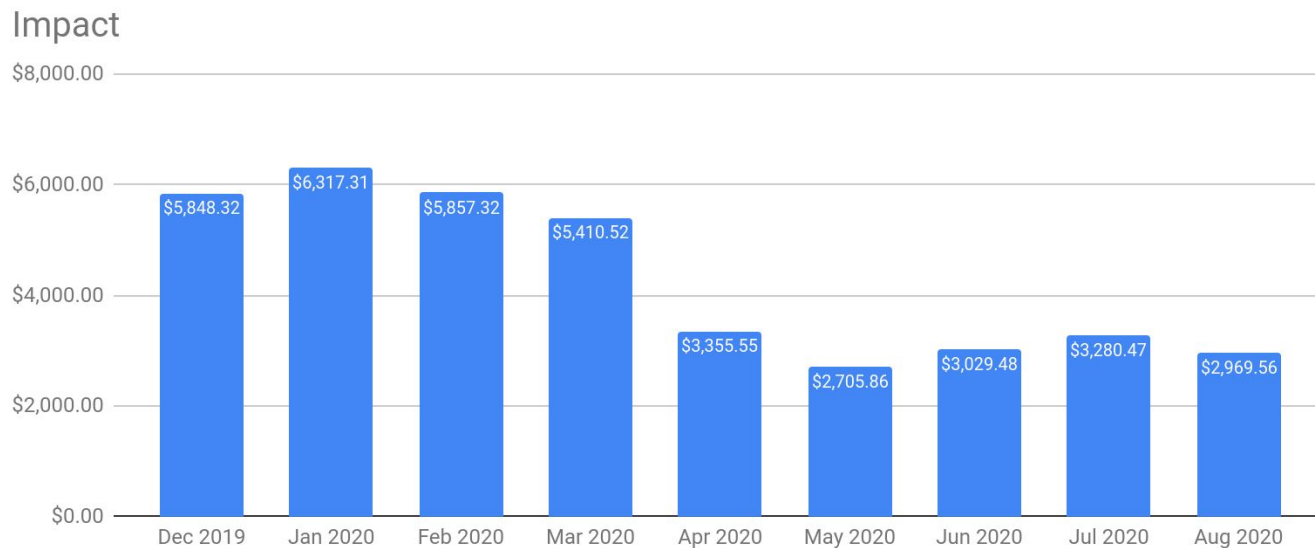
This can be done for any project if you are worried about pulling a lot of git bandwidth



# What now?



# Where we are



- We went for more than \$6000 a month to a regular \$3000
- There is a hard limit of ~\$2500 a month just for the infra



# Continue modernization of our gitlab installation

- Continue migrating from the omnibus installation to the native kubernetes chart:
  - Done:
    - Registry
  - In progress:
    - Webservice
  - Soon to be in progress:
    - Gitaly
    - Sidekiq
    - Task\_runner
    - etc...



# Then (partly) move out from Google Cloud

Gstreamer is paying for CI runners (thanks!)

Packet.com sponsors a lot of our CI (thanks too!).

They offered to host our gitlab infrastructure too





# Take away

- What worked:
  - Grafana!
- What didn't
  - Default metrics tools
- Advice for projects
  - Do not rely too much on artifacts
    - Use MinIO for storing data (ask us on #freedesktop)
  - Use <https://gitlab.freedesktop.org/freedesktop/ci-templates> as soon as you need a custom image (i.e. all the time)



Thank you

# Take away

- What worked:
  - Grafana!
- What didn't
  - Default metrics tools
- Advice for projects
  - Do not rely too much on artifacts
    - Use MinIO for storing data (ask us on #freedesktop)
  - Use <https://gitlab.freedesktop.org/freedesktop/ci-templates> as soon as you need a custom image (i.e. all the time)

