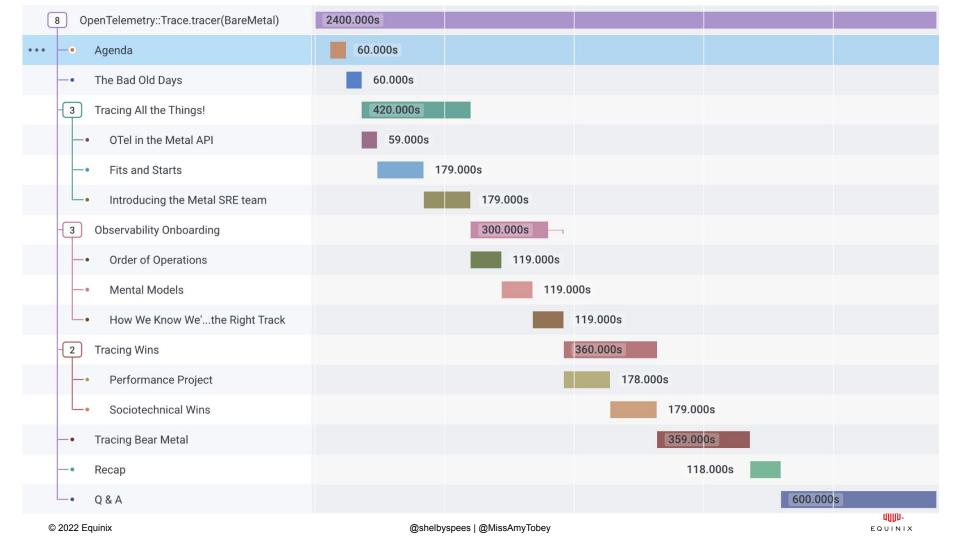
# EQUINIX

# **Tracing Bare Metal** with **OpenTelemetry**

Amy Tobey, Sr. Principal Engineer Shelby Spees, SRE





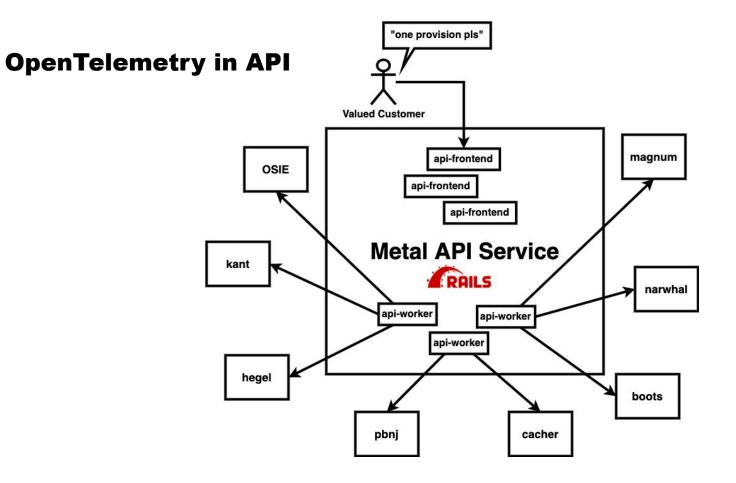
# **The Bad Old Days**

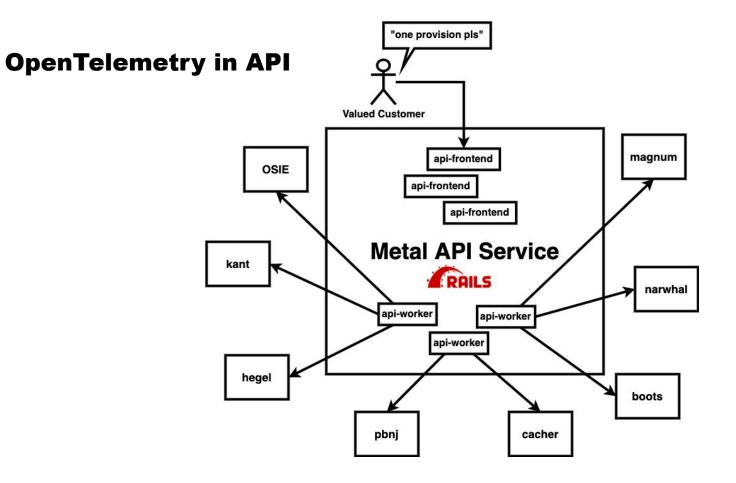
long incidents, inconclusive findings learned helplessness rhyming incidents team siloization & heroes low autonomy

# Simply Restart Everything

# **Tracing All The Things!**







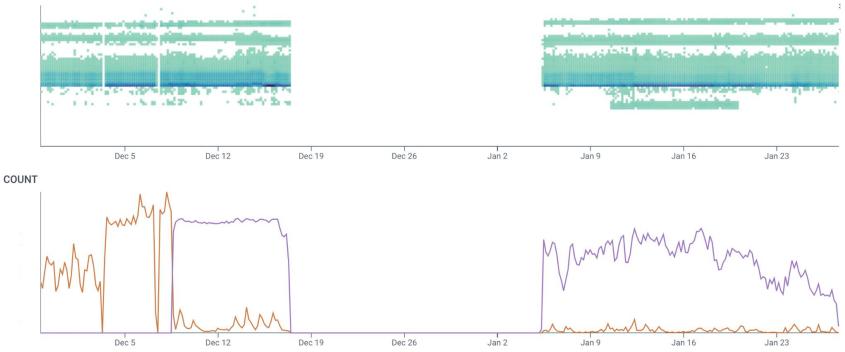
# **Fits and Starts**

[SRE-168] Reenable OTEL tracing for jobs (Reverts PR 7182)  #7212 by merged on Jan 7 • Approved	otel code is executing, just the sending part is disabled
<ul> <li>Disabling otel for activejob </li> <li>#7182 by was merged on Dec 17, 2021 • Approved</li> </ul>	□ Solution Disable OTEL ✓ #363 by was merged on Jun 4, 2021 • Approved
Revert "Re-enable ActiveRecord OTEL span integration" • #6666 by was merged on Jun 2, 2021 • Approved	[ENG-15475] Re-enable OTEL Tracing in API codebase  #360 by
[ENG-15475] Re-enable ActiveRecord OTEL span integration • #6665 by was merged on Jun 2, 2021 • Approved	□
Disable ActiveRecord OTEL span integration • #6648 by was merged on May 26, 2021 • Approved	



# **December 2021: Disabled OTel in API Service**



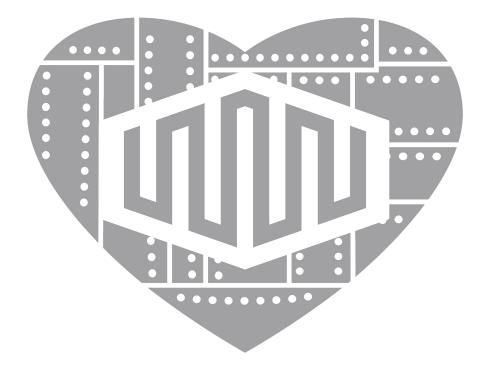


# **Blog: "OpenTelemetry and the Ghost of Provisioning Past"**



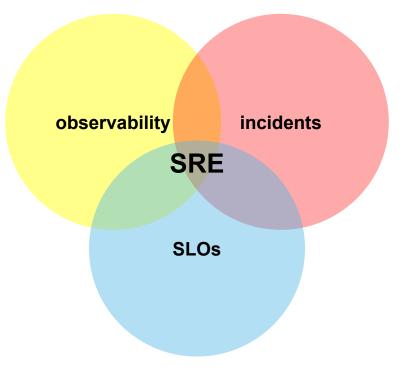


# One SRE Wasn't Enough, We Needed a Team





# The Very Model of a Modern SRE Program





# **Observability Onboarding**



# **Order of Operations**

- (always) add plumbing first!
- (probably) OpenTelemetry Collector
- (maybe) tracing initializers
- (almost certainly) propagation



Blog: "OpenTelemetry: What's a Collector and Why Would I Want One?"

# **Trace Propagation**

```
traceparent=00-7cf6adc22888fb0b798bfa180a2ba21f-43d43ae150985538-01
tracestate=vendor=stuff,keys=values
baggage=keys=values,out-of-band=signaling
```

00-7cf6adc22888fb0b798bfa180a2ba21f-43d43ae150985538-01

# **Auto-instrumentation**

name 🗸 🖕	0s 0	0.05s 0	.1s 0.1	15s 0.2	.s 0.2	5s 0.3s	s 0.3516s
••• 5 IpReservationRequestsController#create	351.6ms						
HTTP POST		189.6n	ns				
• HTTP GET					6.573ms		
-1 UsagesJob send						0.3468ms	⊢
UsagesJob process						7.782ms	
HTTP POST						5.948ms	L
- 1 InternalEventToSegmentJob send						0.1499	ms
InternalEventToSegmentJob process						20.31	ms
<ul> <li>action_view render_template</li> </ul>						15.3	31ms

# **Mental Models**

Pattern matching requires experience, but experience doesn't scale.

Code tracing > Mental tracing

- data outside of people's brains
- visual reference
- linkable and shareable







# **Custom Instrumentation**

What attributes to add?

- focus on business-critical domain data
- for Equinix Metal: instance, server, operating system, etc.
- reflect what's in the code

fine print: ask your Telemetry vendor if high Cardinality is for you!

str instance.os	str server.arch
VMware ESXi 6.5	x86_64
str instance.os_distro	str server.id
str] instance.plan m3.large.x86	str server.state

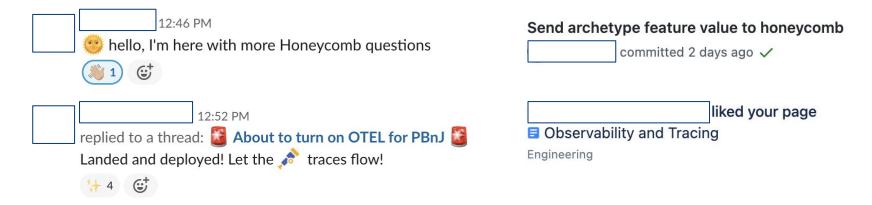


# How We Know We're on the Right Track

developers asking for access to tracing tools & pairing time

PRs being merged to add custom instrumentation

people (and SRE) sharing links to traces and query results during incidents people exclaiming "the system does WHAT!?"





# **Tracing Wins**

# **February 2022: Performance Issues in Customer Portal**

# Long-running issues needed to be addressed

Name	St.	. Ту.		I S	Size	Time 🔻	Waterfall	
ips?state=all&&include=assignments,assignments.assigned_to&exclude=	200	) fe.	<u>c</u>	<u>d</u>	439 B	8.52 s		
ips?&include=assignments,assignments.assigned_to&exclude=	200	) fe.	. <u>c</u>	<u>d</u>	430 B	7.84 s		
plans?include=labels&deployment_type=on_demand	200	) fe.	<u>c</u>	<u>d</u>	411 B	7.26 s		
	200	) fe.	🧕	<u>d</u>	517 B	1.57 s		
projects?page=1&per_page=200&include=transfers&excd,updated_at,customdata,event_alert_configuration	200	) fe.	. 🧕	<u>d</u>	413 B	1.38 s		
□ organizations?per_page=25	200	) fe.	. <u>c</u>	<u>d</u>	9.3 kB	1.18 s		
virtual-networks?&include=instances,internet_gateway,facility,devices&exclude=	200	) fe.	. <u>c</u>	<u>d</u>	320 B	949 ms		
capacity	200	) fe.	<u>c</u>	<u>d</u>	420 B	922 ms		
user?include=emails,number_of_devices,features,messenger_hash	200	) fe.	. <u>c</u>	<u>d</u>	2.4 kB	650 ms		
facilities?include=labels,address&sort_by=name&sort_direction=asc	200	) fe.	. <u>c</u>	<u>d</u>	291 B	526 ms		
operating-systems?include=labels,default_operating_system_version_id	200	) fe.	<u>c</u>	<u>d</u>	413 B	479 ms		
memberships?&include=user&exclude=&per_page=100	200	) fe.	. <u>c</u>	<u>d</u>	423 B	398 ms		
memberships?&include=user&exclude=&per_page=100	200	) fe.	. <u>c</u>	<u>d</u>	411 B	272 ms		
metros	200	) fe.	. <u>c</u>	<u>d</u>	507 B	259 ms		
🗌 ssh-keys	200	) fe.	. <u>c</u>	<u>d</u>	1.6 kB	239 ms		



# 23,000 spans in one trace???

Scale problems:

- Rails (and other ORMs) tend to introduce N+1 queries
- tons of data for large customers, internal Metal orgs

SRE-26 some API endpoints are querying Postgres more than 20,000 times per request relates to

ENC-7615	Some API Endpoints are Excessively Chatty with Storage Resources (DB, Memcache)	=	MA	
ENG-15474	Limit the number of DB queries per API endpoint to << 50		^	
ENG-17851	Fetching plans executes thousands of queries per request	0	JS	
ENC-18001	Fetching IP reservations executes too many queries	0	JS	



# How to minimize the blast radius?

#### def with\_active\_record\_tracing

# LaunchDarkly requires a user, so only enable ActiveRecord tracing on requests with a user.
return yield unless current\_user && LaunchDarklyClient.instance

#### begin

```
user_record = { key: current_user.id }
Trace::ActiveRecordSubscriber.enable(
   LaunchDarklyClient.instance.variation(ACTIVE_RECORD_TRACING_KEY, user_record, false)
)
yield
```

#### ensure

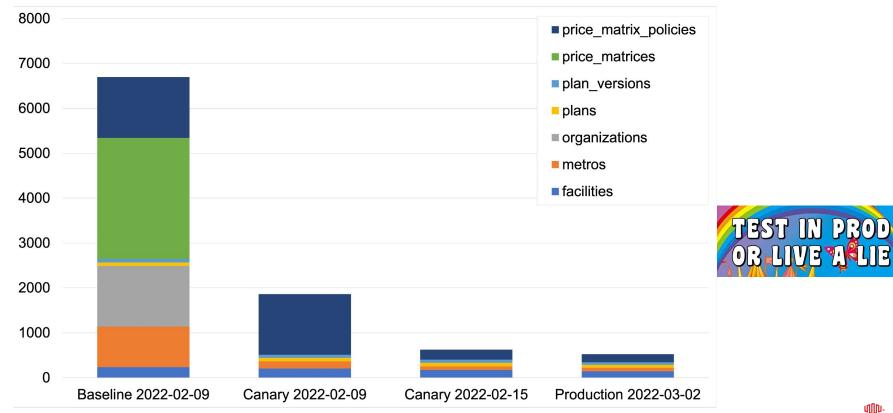
```
# restore thread-local state
```

```
Trace::ActiveRecordSubscriber.enable(false)
```

end

end





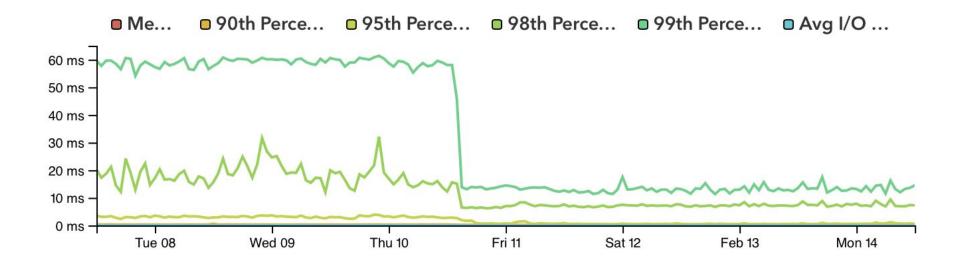
# Number of ActiveRecord Lookups Per Request to /plans

© 2022 Equinix

@shelbyspees | @MissAmyTobey



## **Database optimizations**

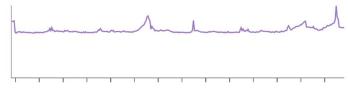




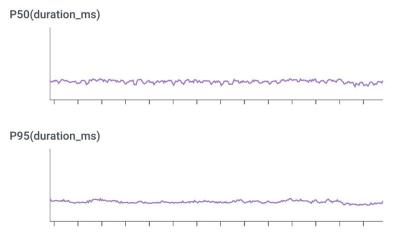
# **API Performance: /ips**

### End of January:

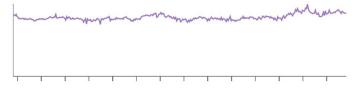
#### P50(duration\_ms)



# End of February:



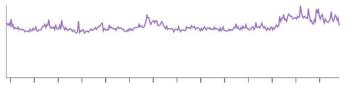
P95(duration\_ms)



P99(duration\_ms)



P99(duration\_ms)





# **Sociotechnical wins**

Teams want tracing for new services

More engineers are canary testing *(*) Iterating on SLOs



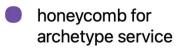
**spees** 11:55 AM

the following SLOs will now send an alert here as well as in #sre-bots:

- Customer provisions
- Customer deprovisions
- API Latency
- API Error Rate
- Authentication

(edited)

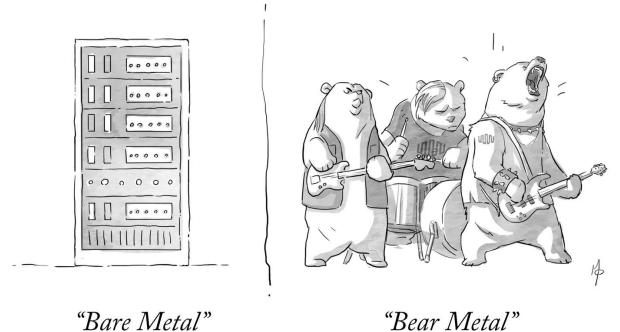




- Wednesday, February 23, 2022 3:00 PM to 3:30 PM (30 minutes)
- https://equinix.zoom.us/
- Otel in github.com/ tinkerbell/dhcp
- Thursday, February 24, 2022 4:00 PM to 5:00 PM (1 hour)
- https://equinix.zoom.us/



# **Tracing Bear Bare Metal**





# **Provisioning a New Instance**

name 🗸	service.name 🗸 🖕	0s 20s 40s 60s 80s 100s 120.6s
••• 6 InstancesController#create	PacketApi	1.185s
- 1 Cacher::PushJob send	PacketApi	0.2730ms
Cacher::PushJob process	PacketApi	452.8ms
- 1 ProvisionDeviceJob send	PacketApi	-0.2859ms
-507 ProvisionDeviceJob process	PacketApi	119.771s
- 1 ReclaimSpotInstanceJob send	PacketApi	0.1346ms
ReclaimSpotInstanceJob process	PacketApi	534.3ms
- InternalEventToSegmentJob send	PacketApi	0.2547ms
InternalEventToSegmentJob process	PacketApi	13.34ms
- InternalEventToSegmentJob send	PacketApi	0.2606ms
InternalEventToSegmentJob process	PacketApi	31.11ms
action_view render_template	PacketApi	143.9ms



# Inside the ProvisionDeviceJob

-137 ProvisionDeviceJob process	PacketApi	119.771s
—● HTTP POST	PacketApi	7.789ms
github.com.tinkerbellv1.Machine/BootDevice	PacketApi	⊣ 300.4ms
github.com.tinkerbeMachine/BootDevice	pbnj	<b>О</b> 54.7µs
client.SetBootDevice	pbnj	3.015s
github.com.tinkerbell.pbnj.api.v1.Task/Status	PacketApi	300.9ms
• github.com.tinkerbelnj.api.v1.Task/Status	pbnj	<b>О</b> 76.9µs
- 1 github.com.tinkerbell.pbnj.api.v1.Task/Status	PacketApi	75.11ms
• github.com.tinkerbelnj.api.v1.Task/Status	pbnj	<b>♦</b> 74.1µs
github.com.tinkerbell.pbnj.api.v1.Task/Status	PacketApi	75.15ms
• github.com.tinkerbelnj.api.v1.Task/Status	pbnj	🔶 78.5µs
github.com.tinkerbell.pbnj.api.v1.Task/Status	PacketApi	75.88ms
github.com.tinkerbelnj.api.v1.Task/Status	pbnj	🔶 85.1µs



name 🗸	service.name 🗸 🖕	http.url 🗸	0s20s 40s 60s 80s 100s 120.6s
-1 HTTP POST	PacketApi	https://narwhalpacket.net/ports	—314.3ms
5 PortsController#create	narwhal		
• postgresql.Switch_Load	narwhal		0.1891ms
-1 build switch tasks	narwhal		-0.9875ms
10 perform task: create_port	narwhal		7.291s
juniper.edit_config	narwhal		15.65ms
• juniper.edit_config	narwhal		13.58ms
juniper.edit_config	narwhal		8.534ms
• juniper.edit_config	narwhal		11.39ms
juniper.edit_config	narwhal		6.673ms
• juniper.edit_config	narwhal		6.234ms
• juniper.edit_config	narwhal		10.23ms
• juniper.edit_config	narwhal		14.38ms
• juniper.edit_config	narwhal		83.71ms
juniper.edit_config	narwhal		16.20ms
• postgresql.begin	narwhal		0.1356ms
• postgresql.Task_Create.insert	narwhal		0.3422ms
postgresql.commit	narwhal		1.676ms

name 🗸	service.name 🗸 🆕	http.url 🗸	0s20s 40s 60s 80s 100s 120.6s
-1 HTTP POST	PacketApi	https://narwhalpacket.net/acls	—313.6ms
5 AclsController#create	narwhal		— <del>9.12</del> 3ms
postgresql.Switch_Load	narwhal		0.2218ms
- 1 build switch tasks	narwhal		-1.040ms
perform task: create_acl	narwhal		108.6ms
juniper.edit_config	narwhal		47.24ms
juniper.edit_config	narwhal		61.15ms
postgresql.begin	narwhal		0.1419ms
postgresql.Task_Create.insert	narwhal		0.3605ms
<ul> <li>postgresql.commit</li> </ul>	narwhal		1.515ms



# "Is it done yet?"

name 🗸	service.name 🗸 🖕	http.url 🗸	Os2Os 40s 60s 80s 100s 120.6s
-1 HTTP GET	PacketApi	https://narwhal10-28a6def2697e	310.8ms
2 TasksController#show	narwhal		5.499ms
—• postgresql.Task_Load	narwhal		0.2170ms
• postgresql.Switch_Load	narwhal		0.1442ms
-1 HTTP GET	PacketApi	https://narwhalba-759364c5b7a6	309.5ms
2 TasksController#show	narwhal		5.594ms
— • postgresql.Task_Load	narwhal		0.2257ms
• postgresql.Switch_Load	narwhal		0.1390ms

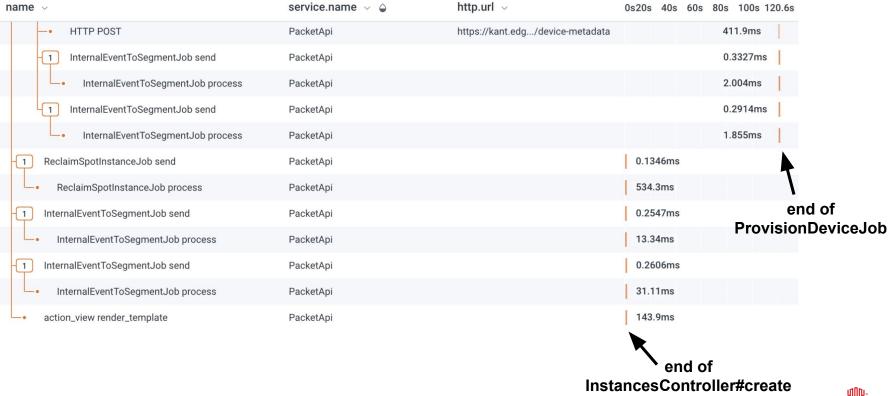
name -	service.name - g	http.url - 0a	254	40a	60a	115	1004	120.4a
HTTP GET	PacketApi	https://narwhal.10-28a6def2697e	100.0ma					
HTTP GET	PackatApi	https://wendual.ba/290264c9b7a6	309.5ms					
II NTTP GET	PackatApi	https://weishal.10.28a6de/2697e	308.5ms					
TIP GET	PackatApi	https://wendral.ba-759364c5b7a6	100.6ms					
HTTP GET	PacketApi	https://narehul.10-28a6def2697e	308.3ma					
E HTTP GET	PackatApi	https://wentul.ba/99264c9b7a6	305 Tem					
I NTTP OLT	PaikatApi	https://wendral.10-28atide/2697e	306.0ms					
T HTTP GET	PackatApi	https://narahal.ba-759364c5b7a6	548.8ma					
I HTTP GET	PacketApi	https://narwhal.10-28a6def2697e	307.7ma					
C HTTP-GET	PackatApl	https://weahal.ba/292264c9b7a6	309.4ms					
I NTTP-GET	PackatApl	https://www.hul.16.28atide/2697e	311.Bea					
T HTTP-GET	PackatApl	https://narahal.ba-759364c5b7a6	307.tma					
E HTTP GET	PacketApi	https://narwhal.10-28a/def2697e	509.2ma					
TIP RTTP OFT	PackatApi	https://weishulbu/202264c9b7a6	306.6ms					
TI NTTP GET	PackatApl	https://narabul.16.28a6de/2697e	210.5ms					
T HTTP-GET	PackatApl	https://wendul.ba/759364c5b7a6	209.4ma					
T HTTP GET	PacketApi	https://narwhal.10-28atide/2697e	311.4em					
TID RTTP OFT	PaikatApi	https://ueidul.iba/202264c9b7a6	308.3ms					
T HTTP-GET	PatkatApl	https://narabul.16.28a6de/2697e	207.4ma					
T HTTP-GET	PackatApl	https://narahal.ba-759364c5b7a6	319.1ma					
C HTTP GET	PackatApi	https://narwhal.10-28atide/2697e	300.5ms					
TIP OFT	PackatApi	https://ueidul.iba/20264c9b7a6	215.2me					
I HTTP-GET	PackatApi	https://nariefull.16-28a6de/2697e	310.5ma					
T HTTP GET	PackatApi	https://wendul.ba-759364c5b7a6	308.3ma					
I HTTP GET	PackatApi	https://www.lite.2012.002	310.4mm					
TID HTTP-GET	PackatApi	https://wendul.ibs/250264c5b7a6	213.4m					
T HTTP-GET	PackatApl	https://narabul.16.28a6de/2697e	201					
T HTTP GET	PackatApl	https://narahal.ba/759364c5b7a6		08.4ma				
TID HTTP-GET	PackatApl	https://weishal.10.28at.def2697e		310.8ms				
TI NTIPOLT	PatkatApi	https://wendral.ba/759264c5b7a6		209.1ms				
I HTTP GET	PacketApi	https://nariehal.16-28a6def2697e		213.4ma				
E HTTP GET	PackatApi	https://newbal.ba/250264c5b7a6		310.6ma				
I HTTP OFT	PackatApi	https://wendral.10.28at.def2697e		309.3ms				
III HTTP-GET	PatkatApi	https://wendral.ba/759364c5b7a6		310.2ma				
I HTTP GET	PacketApi	https://narsehul.16-28a6def2697e		208.0ma				
E HTTP GET	PackatApi	https://wentul.ba759264c5b7a6		309.5ms				
III HTTP GIT	PaikatApi	https://www.hul.10.28a6def2697e		310.0ms				
TTP GET	PackatApi	https://weneful.iba-759364c5b7a6		212.3ma				
HTTP GET	PacketApi .	https://nareful.10-28a6def2697e		306.8ma				
E HTTP GET	PackatApi	https://wentul.ba/550264c9b7a6		306.fms				
I HTTP GIT	PackatApi	https://www.ii.10.28a6def2697e		303.frm				
Ta ette	PackatApi	https://naraful.iba/250364c5b7a6		307.1ma				
TIP GET	PacketApi	https://nareful.10-28a6def2697e		309.1ma				
	PackatApi	https://newful.ibs/250264c9b7a6		100.4ms				
	PackatApi	https://www.ii.10.28a6def2697e						
Ta ette at	PackatApi	https://naraful.ba/759364c5b7a6		319.7	na 06.2ma			
TSD HTTP GET	PacketApi	https://narehul.10-28a6def2697e			08.2ms 309.3ms			
HTTP GET	PackatApi	https://weakut.bs/20264c9b7a6			109.3ms			
RTIP GET	PaiketApi	https://wrebal.10.28atide/2557e			203.3ms			
HTIP GET	Packetépi Packetépi	https://wenkul.ba/759364c5b7a6 https://wenkul.10.28atdef2697e			326.9ma			
RTIP GET		https://www.bat.ito.28ad.de/2697e https://www.bat.ito.759264c9b7a6			306.8ma 310.5ma			
I HTP GET	PaikatApi PackatApi	https://weidul.ibs/2502645957a6 https://weidul.i0.28a5de/2507e			311.5ms 311.5ms			
HTP GET	PacketApi	https://wenduit.ito.250364c507a6			313.5ma 310.5ma			
REPORT					310.3ms			
I RTP GET	PacketApi PacketApi	https://www.hal10.28atide/2697e https://www.hallus.759264c9b7ati			310 JPN 308 JPn			
I HTP-GET	PacketApi	https://wentul.ite/20204/10/26			314.3ma			
TIP GET	PacketApi	https://narwhat.to.759364c5b7a6			214.3ms 308.7ms			
HTP-GET	PacketApi PacketApi	https://narwhal.ba/759264c5b7a6 https://narwhal.10.28atdel/2697e			306.5mm			
I RTP-GET	PacketApi	https://wenduit.ito.2004c007e-			315.3ms			
I HTP GET	PacketApi	https://wendul.ile/20204040746			206.6m			
RTP GET	PacketApi	https://nershul.ito.259384c5b7a6			200.0ms			
TIP GET	PacketApi	https://wenhat.ba/750264c567a6			300.2ms			
RTIP GET	PasketApi	https://wenduit.10.2866462607e			310.0ms			
I HTP GET	PacketApi	https://wentul.ite/10104040746			301. bru			
HTP-GET	PacketApi	https://nareful.16.28a5de/2607e https://nareful.bo/759364c5b7a6			308.2ms			
HTP-GET	PacketApi	https://wendual.ibs/250364c567a6			307.4m			
RTIP GET	PacketApi	https://wendual.it/200546026746			300 Artis			
C HTTP GET	PacketApi	https://weetul.ite/10104010746			209.3m			
C HTIPGET	PacketApi	https://www.ite.250364c5b7a6			209.5mm			
	PacketApi	https://www.l.do-7906458786			307.5mm			
RTIP GET	PacketApi	https://wendul.it/2005462007e			200.1ma			
and the set	- autority							



# add\_port\_acl

name 🗸	service.name 🗸 🖕	http.url v 0s20s	s 40s 60s 80s 100s 120.6s
1 HTTP POST	PacketApi	https://narwhal1%2F0%2F4/acls	313.1ms
5 PortAclsController#create	narwhal		8.152ms
• postgresql.Switch_Load	narwhal		0.1728ms
- 1 build switch tasks	narwhal		0.8454ms
2 perform task: add_port_acl	narwhal		52.89ms
juniper.edit_config	narwhal		52.62ms
juniper.edit_config	narwhal		29.64ms
— ● postgresql.begin	narwhal		0.1130ms
postgresql.Task_Create.insert	narwhal		0.3265ms
• postgresql.commit	narwhal		1.405ms

# Wrapping up



# **Boots calling PacketAPI**

name 🗸	service.name 🗸 🍦	0s 0.05s 0.1s 0.15s 0.2s	s 0.25s 0.3s 0.3297s
••• 2 ServeDHCP	boots	329.5ms	
-1 HTTP POST	boots	328.9ms	
• Staff::Cacher::HController#create	PacketApi		80.12ms
ServeDHCP Reply	boots		0.1865ms 🗅

```
tracer := otel.Tracer("DHCP")
```

```
ctx, span := tracer.Start(context.Background(), "DHCP Reply",
    trace.WithAttributes(attribute.String("MAC", mac.String())),
    trace.WithAttributes(attribute.String("IP", gi.String())),
    trace.WithAttributes(attribute.String("MessageType", req.GetMessageType().String())),
    trace.WithAttributes(attribute.String("CircuitID", circuitID)),
```



# **Recap Slide!**

SRE team is responsible for enablement

Success relies on cultural adoption

Cultural adoption relies on success!

Observability is the enemy of helplessness

Incident management lets you observe your observability

It's feedback loops all the way down

SLOs are easier with customer-centric trace data

# span.End()

**Reach out!** 

<u>@MissAmyTobey</u> <u>@shelbyspees</u>

**OTel on Equinix Labs** 

equinix-labs/otel-cli equinix-labs/otel-init-go

**OSS Bare Metal Provisioning** 

tinkerbell.org

We're hiring! careers.equinix.com

# Read the blog! metal.equinix.com/blog

