

Building Data Engineering Teams

Wouter de Bie

Engineering Director - Data Engineering



DATADOG

Hi! So Wouter...

... what are you going to talk about in the next 40 minutes? I just got coffee and breakfast and I'm dying to listen to your talk! Tell me more!

Let's talk about
how I've built
data engineering
teams!

- Data Engineering
- Common patterns
- Horizontal vs vertical
- High performing teams

Alright Wouter, nice to meet
you...

... but before you start, who are you? Why are
you talking to us?



Who am I?

Wouter de Bie

- Engineering Director at Datadog
- Formerly Engineering Director at Spotify
- Dutch, living in Brooklyn NY until next week

OK Wouter...

... we're here to talk Data and not about where you're moving. But I can't help asking the question.. Where to?

New Orleans, Louisiana!

It's the internet age! Datadog let's
me work remote!

Mardi Gras!



Ok, ok Wouter...

... let's get to business. You've been at a few companies that have gone through hyper-growth. What did that mean ?

Hyper-growth numbers!

- Spotify: 250 people when I started, 5500 when I left.
- Datadog: 100% growth year over year.

But Wouter...

... apart from employee growth, did you have any scalability challenges?

Scale is a challenge!

What scale?

Datadog

- Observability as a service.
- Metrics, Traces and Logs.
- Multiple clouds.
- Billions of points a day.

Spotify

- Music streaming service
- Data about playback
- 100's of millions of plays a day
- 100's of millions of users

Data Engineering



So Wouter...

... you did a bunch of Data Engineering. Tell me about that. What did you do?

Data Engineering definition

Techtarget.com: *"A data engineer is a worker whose primary job responsibilities involve preparing data for analytical or operational uses."*

O'Reilly: *"Build and maintain the organization's data pipeline systems... Clean and wrangle data into a usable state"*

Datacamp.com: *"Ensure clean, reliable, and performative access to data and databases."*

Wouter's definition

Data engineering takes care of storing and processing large volumes of data, coming from different sources and makes it available for downstream applications (like ML, Analytics, Product, etc)

But Wouter...

... these Data Engineers process Data and such,
but who runs the systems that they use?

Data Engineering

!=

Data Infrastructure
Engineering

Wouter's definition

Data Infrastructure Engineering build, maintains and operates systems that enable Data Engineers to do their work.

Skills required

Data Engineering

- Write Spark/MR/Beam/etc pipelines
- Understand concepts like sorting, grouping, filtering, etc
- Understand the domain
- Understand scalability of the data

Data Infrastructure Engineering

- Operate data processing systems like YARN/Mesos/HDFS
- Write tools that make life easier for Data Engineers
- Understand scalability of systems

Of course there is
overlap

So Wouter...

... you've been at a few companies? Any common patterns you have observed?

Common patterns

Big Data Starts Small

- There is some interesting data somewhere
- Small team uses the data to do multiple things with it
- Small team operates infrastructure besides data engineering and analytics

Big Data Becomes Valuable

- Small team becomes a bottleneck
- Team splits up into Analytics, Data Engineering and Data Infrastructure

Big Data Becomes Part of Day-to-Day

Data Engineering becomes a competence rather than a team, since every domain needs Data Engineers

Datadog's Evolution

Single team focussed on Analytics, Data Engineering and Infrastructure

Single team split up into Data Engineering, Data Infra

Currently:

- Data Infrastructure team
- Data Engineering for Historical Metrics team
- Internal Analytics
- Revenue Engineering
- Data Science

Spotify's Evolution

Single team focussed on Analytics, Data Engineering and Infrastructure

Single team split up into Analytics and Data Infrastructure.

Currently:

- Data Infrastructure org
- Data Engineering is a competence in many teams
- Analytics is a competence in many teams
- Data Science is a competence in many teams

But Wouter...

... when Data Engineering becomes a competence that can be found in many teams, how do we ensure knowledge is shared?

Two things I've done

Data Eng Guild

Organizational structure that involves all data engineers. Focussed on knowledge sharing, but sometimes work on things together.

Data University

Curriculum focussed on teaching Data Engineering skills. Used for on-boarding or training engineers in the art of Data Engineering.

Right Wouter...

... so now I have a bunch of teams. How did you decide how to split them up? Any tips?



Horizontal vs Vertical teams

Horizontal vs vertical

Vertical teams

Narrow focus on one group of constituents or a single domain.

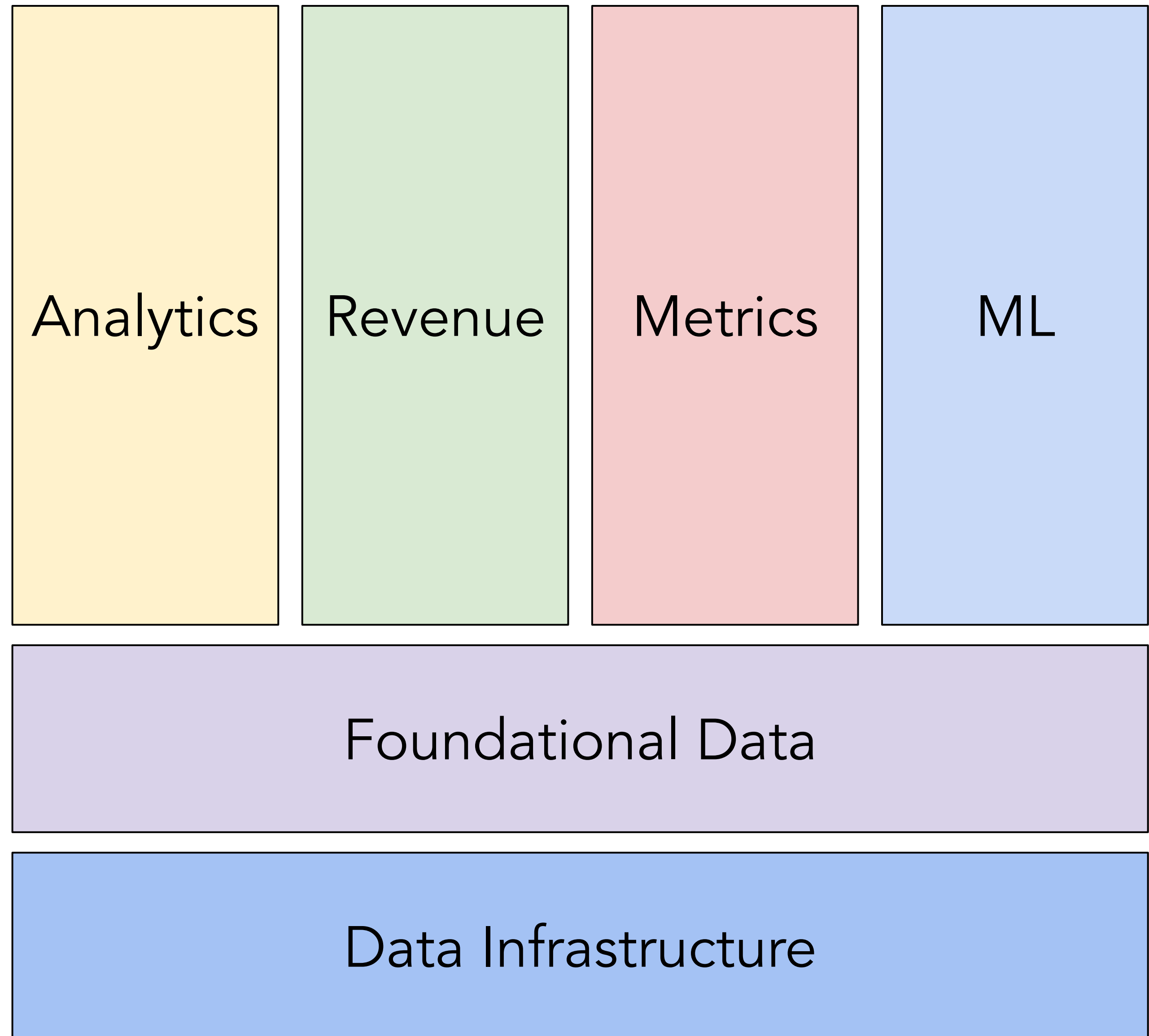
E.g. Historical Metrics, Revenue Engineering, Recommendations.

Horizontal teams

Wide focus on providing data to many other (downstream) teams.

(Infrastructure is horizontal by definition)

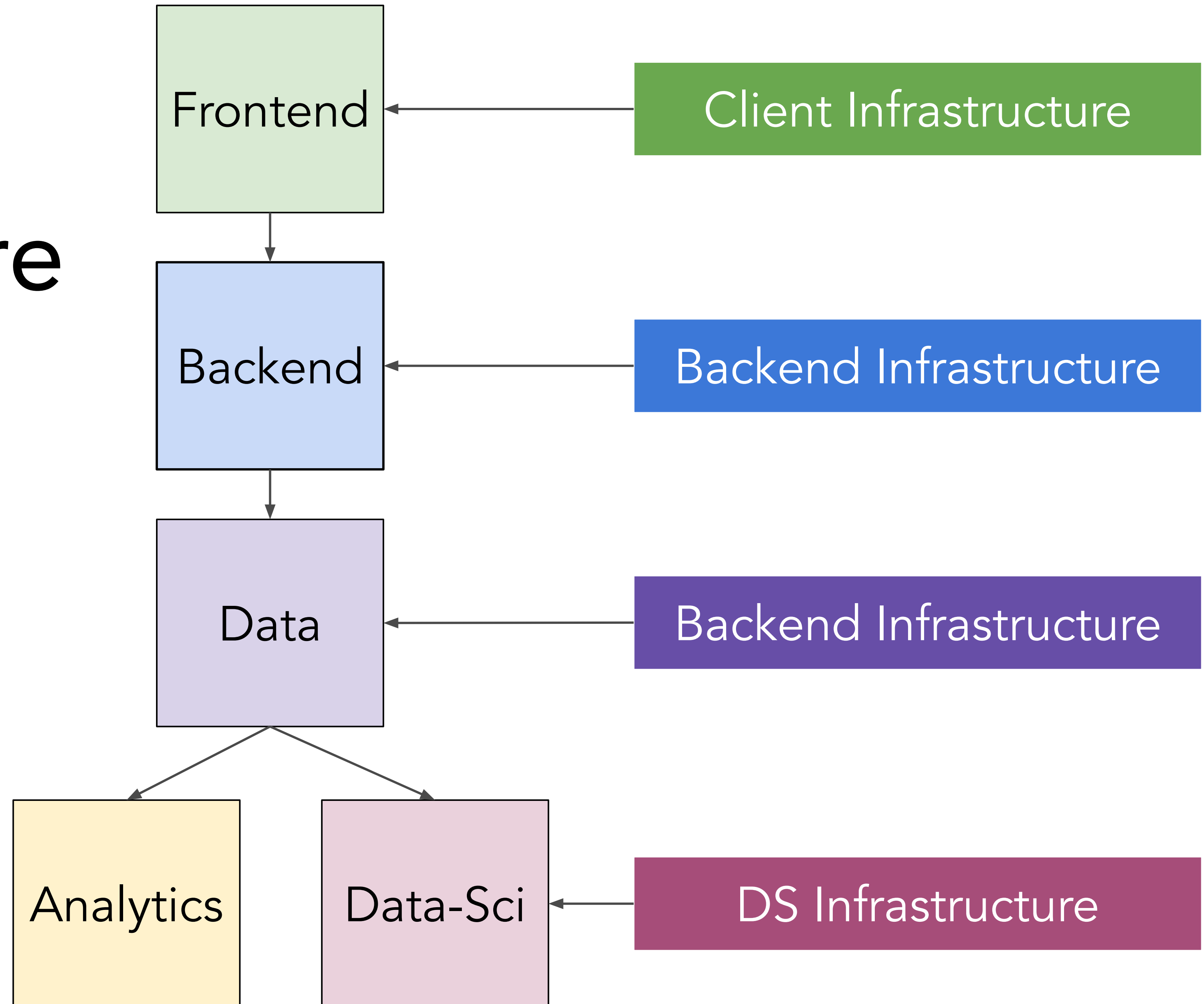
Horizontal / Vertical example



I see Wouter...

... but now I have all these teams building the same thing over and over again. Any thoughts?

The role of infrastructure



Allright, alright Wouter...

... so I've grown my org, split my teams, hired the right folks. Now how do I make sure they perform?

High performing teams

Autonomy, Mastery, Purpose

Autonomy

Our desire to be self directed.
It increases engagement over compliance.

Mastery

The urge to get better skills

Purpose

The desire to do something that has meaning and is important.

Yes, yes Wouter...

... that makes sense. I see you read "Drive" by Daniel H. Pink. But how do I make sure my teams value Autonomy, Mastery and Purpose?

Hiring is key!

- Specialized interview process for both Data Eng and Infra
- Focus on skills and culture
- Enthusiasm about the product is really important

Alright man...

... I learned something and I'm feeling the coffee. Can we wrap this up?

Let's wrap up!

Takeaways

- Data Engineering != Data Infra
- Teams and orgs evolve
- Horizontal vs Vertical thinking
- Build solid Infrastructure
- Hiring is key!

Lastly,
a shameless plug

We're hiring! Come see us at our booth!

Thank you!

Come see me at office hours!



DATADOG

Large text can
be used to make
a single
important point.

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Standard slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Slide with two points

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Slide with three points

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

Slide with four points

Identify and convince great people to join

Great = able + driven + grown-up

Hiring is what makes/breaks us

Expect to spend 10-20% of your time on hiring

Big Statement

1

Geometric
top-line growth

2

Strong
product-market fit

3

Financially efficient
go-to-market

Use iconography to support content



Cras mauris purus,
varius in justo in, varius
vulputate eros.

Secondary text, if needed



Curabitur pharetra
dictum mattis curabitur
dictum.

Secondary text, if needed



Proin iaculis nulla
turpis, a fermentum
orci mattis in.

Secondary text, if needed

The Pillars of Observability



Logs



Traces



Metrics

White background slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Text+image slide

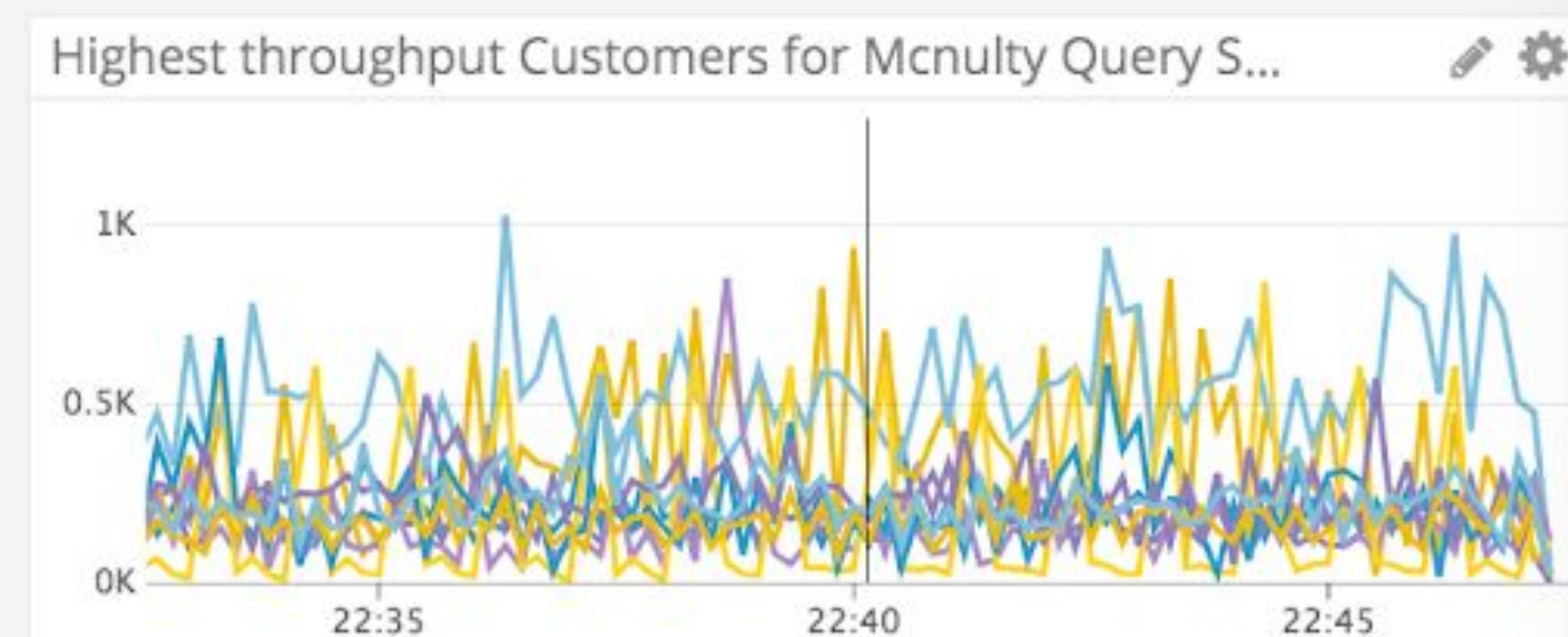
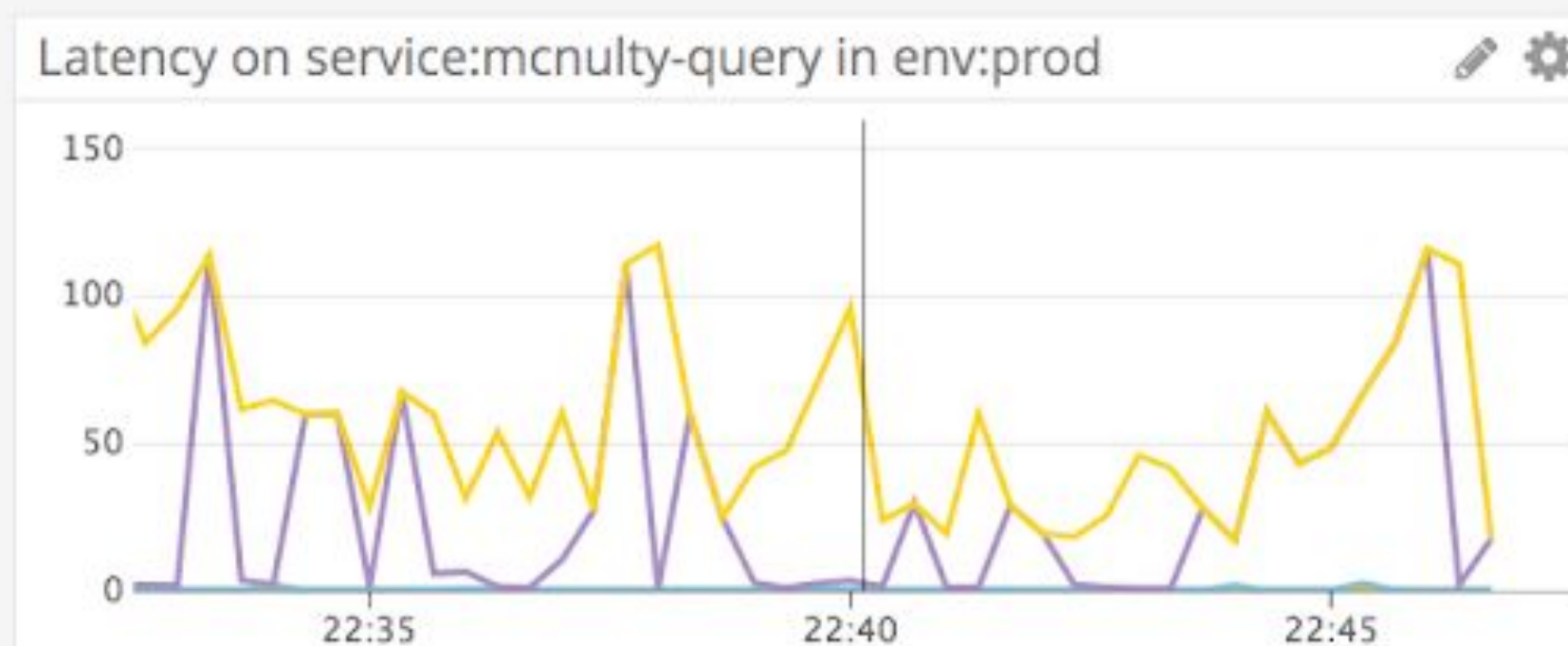
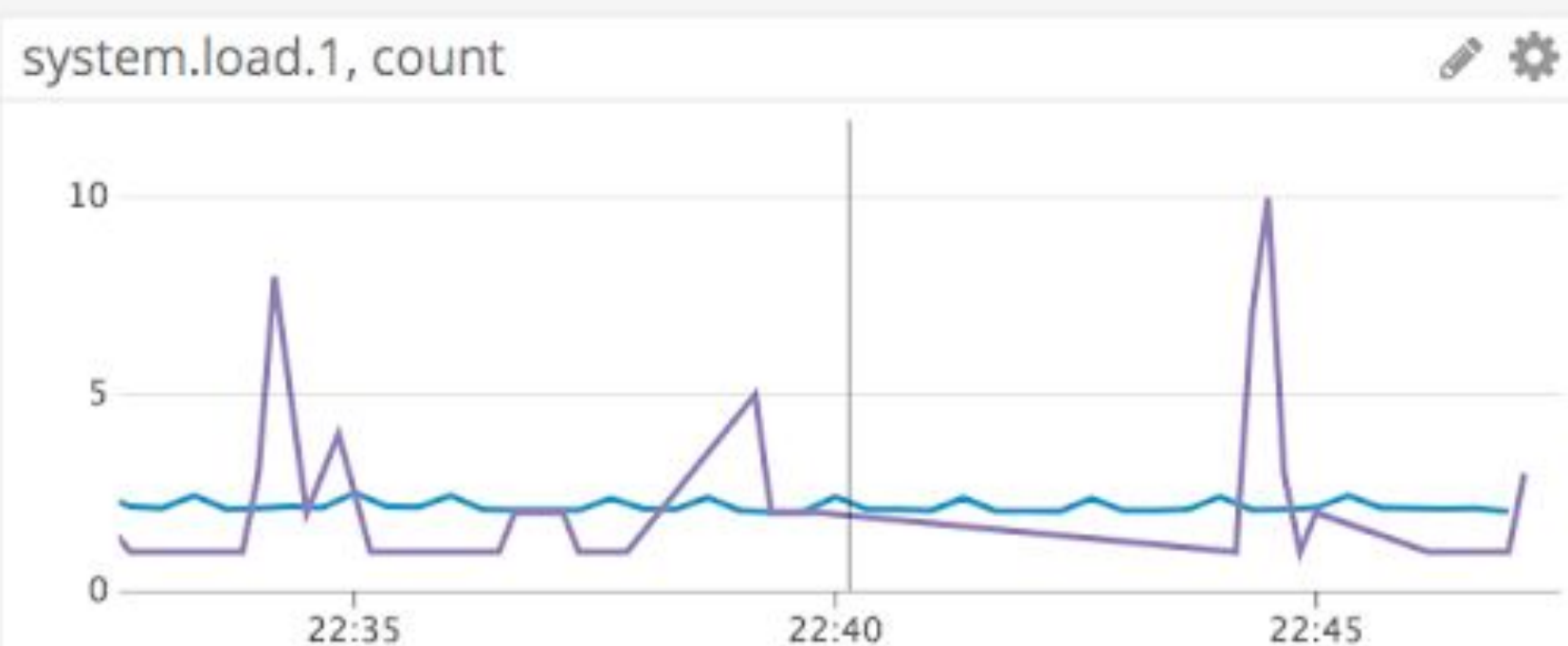
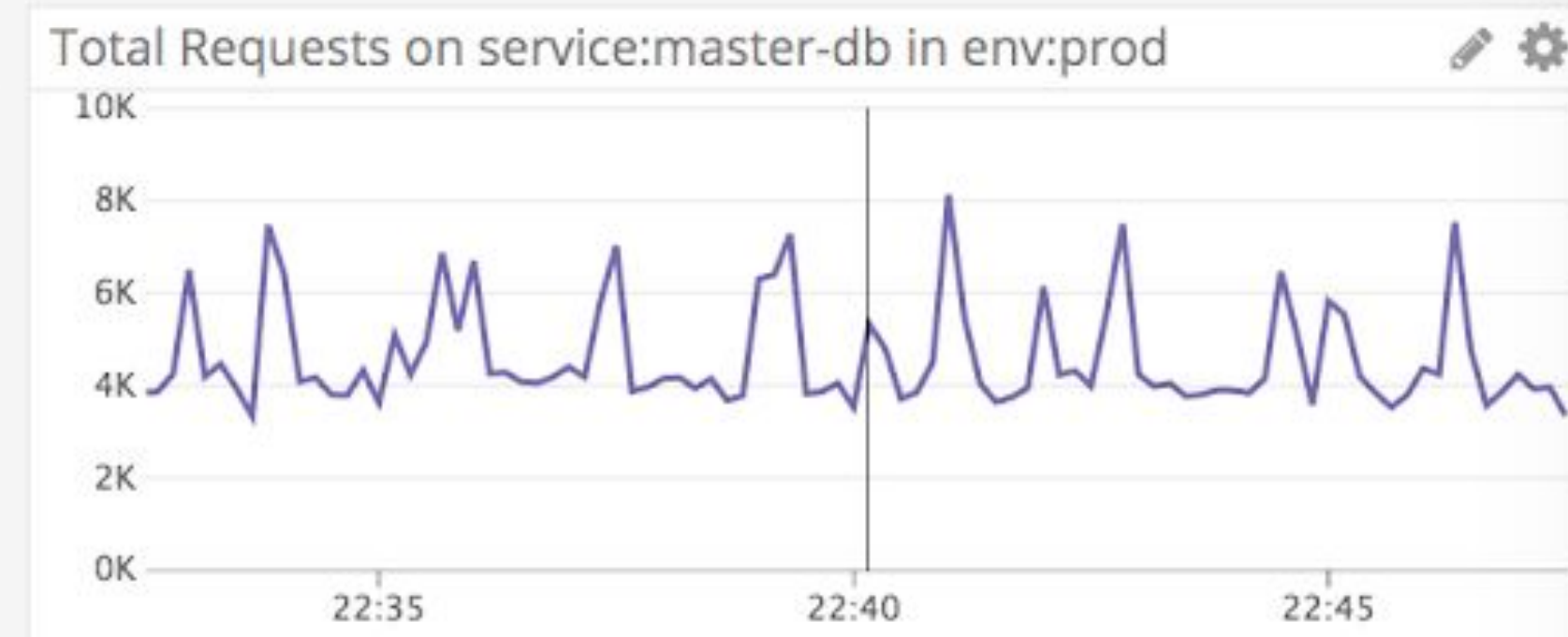
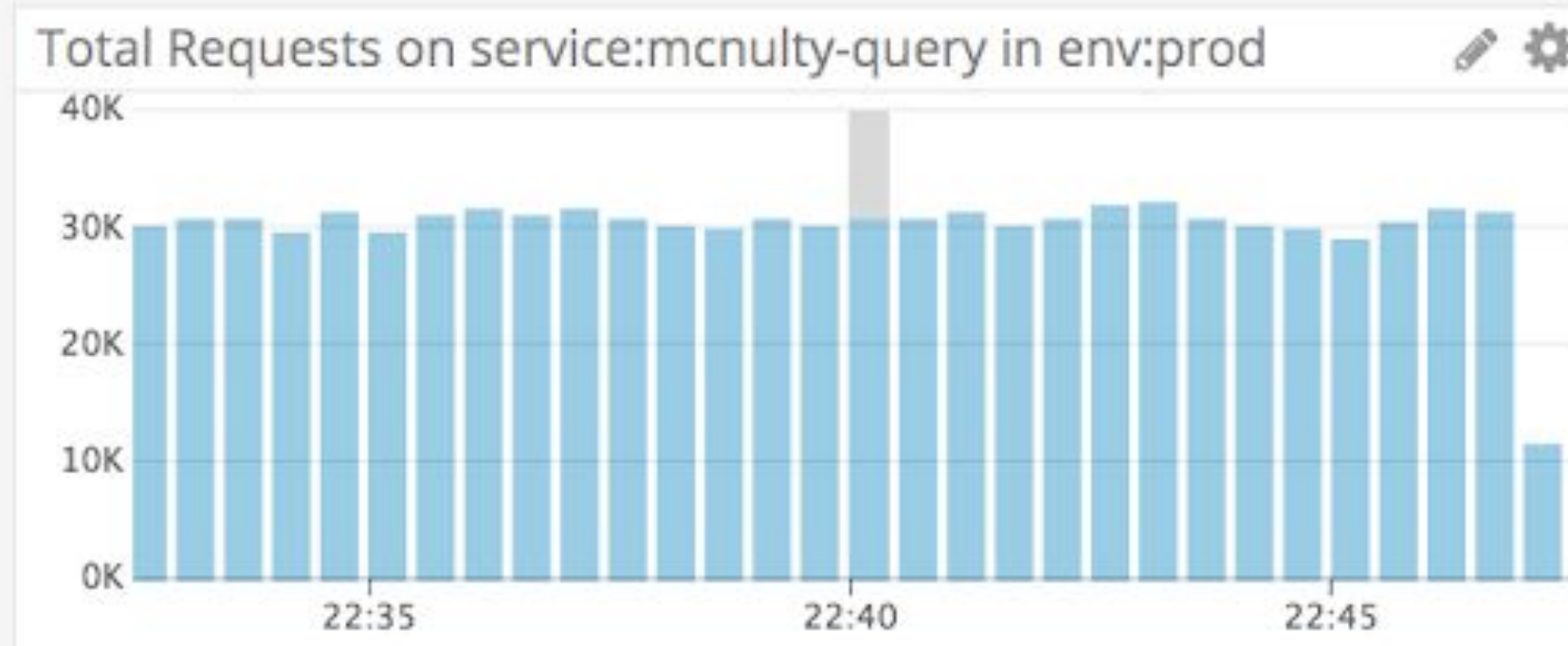
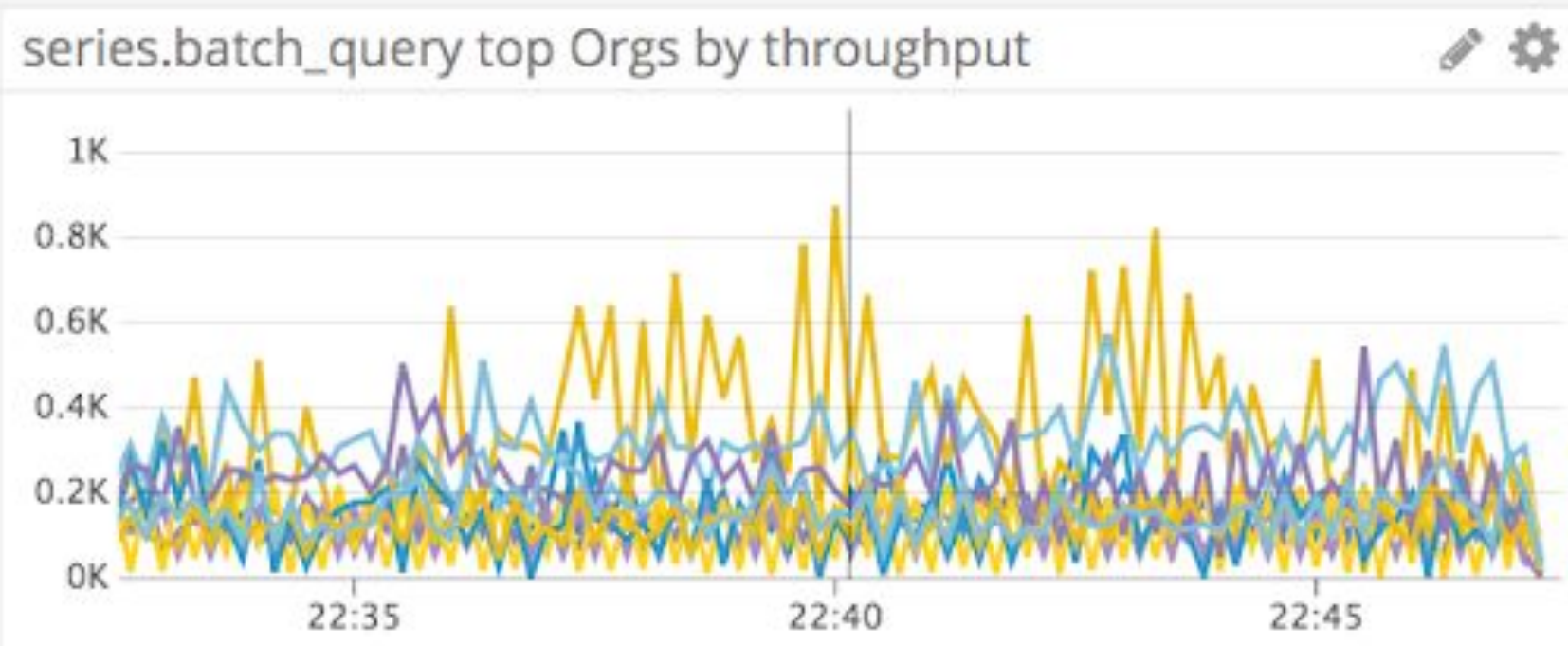
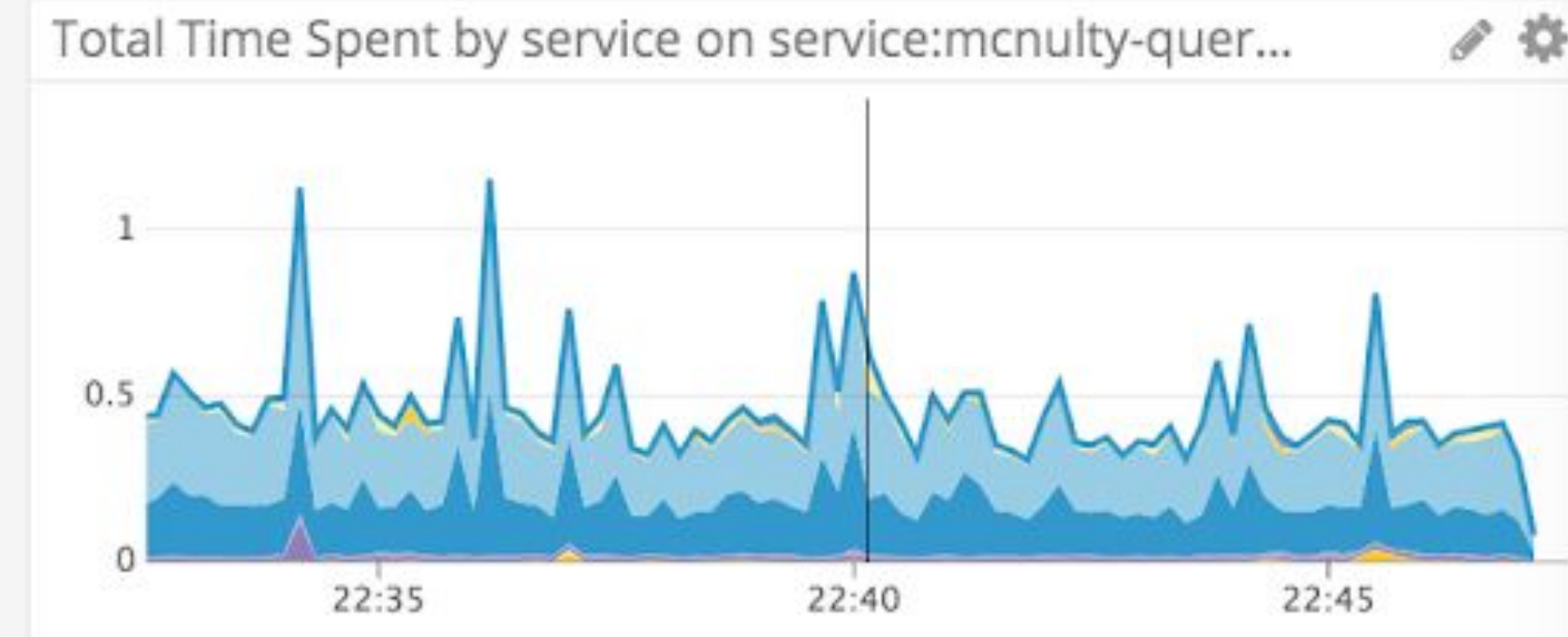
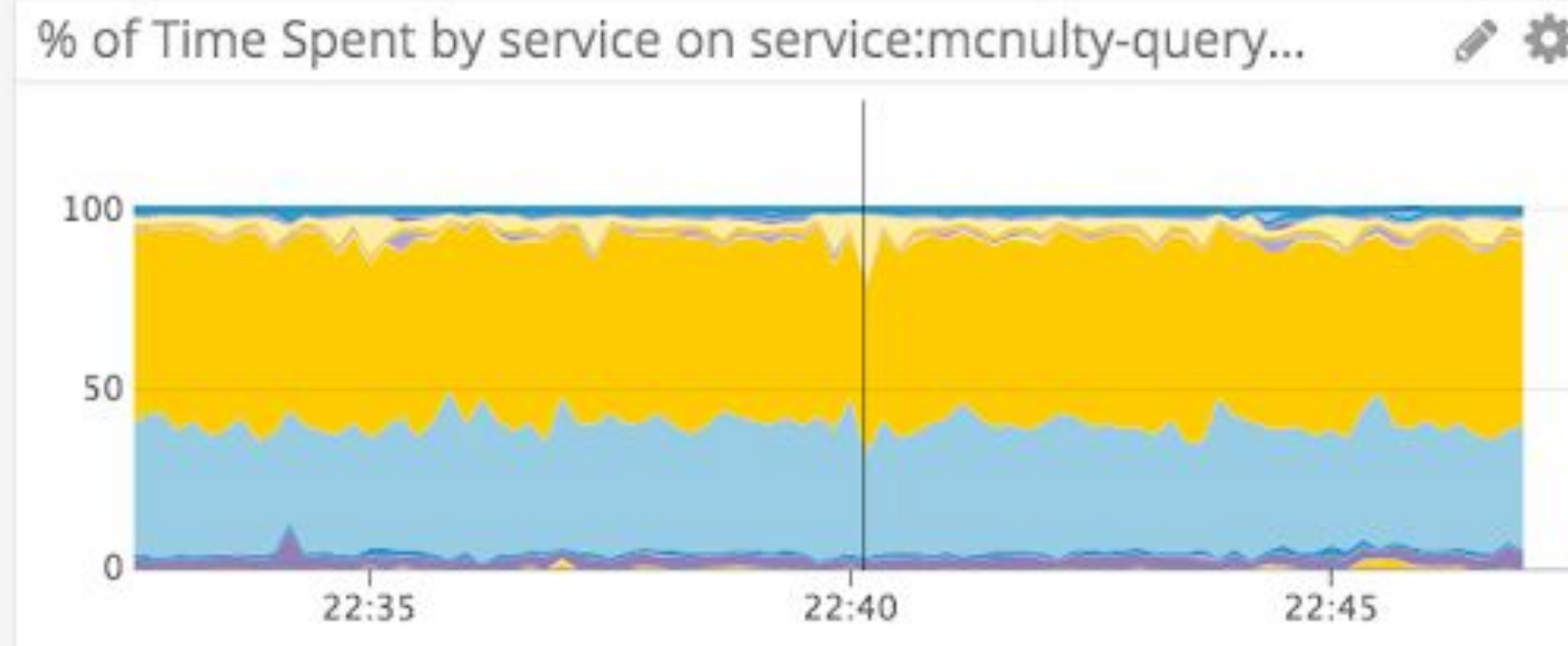
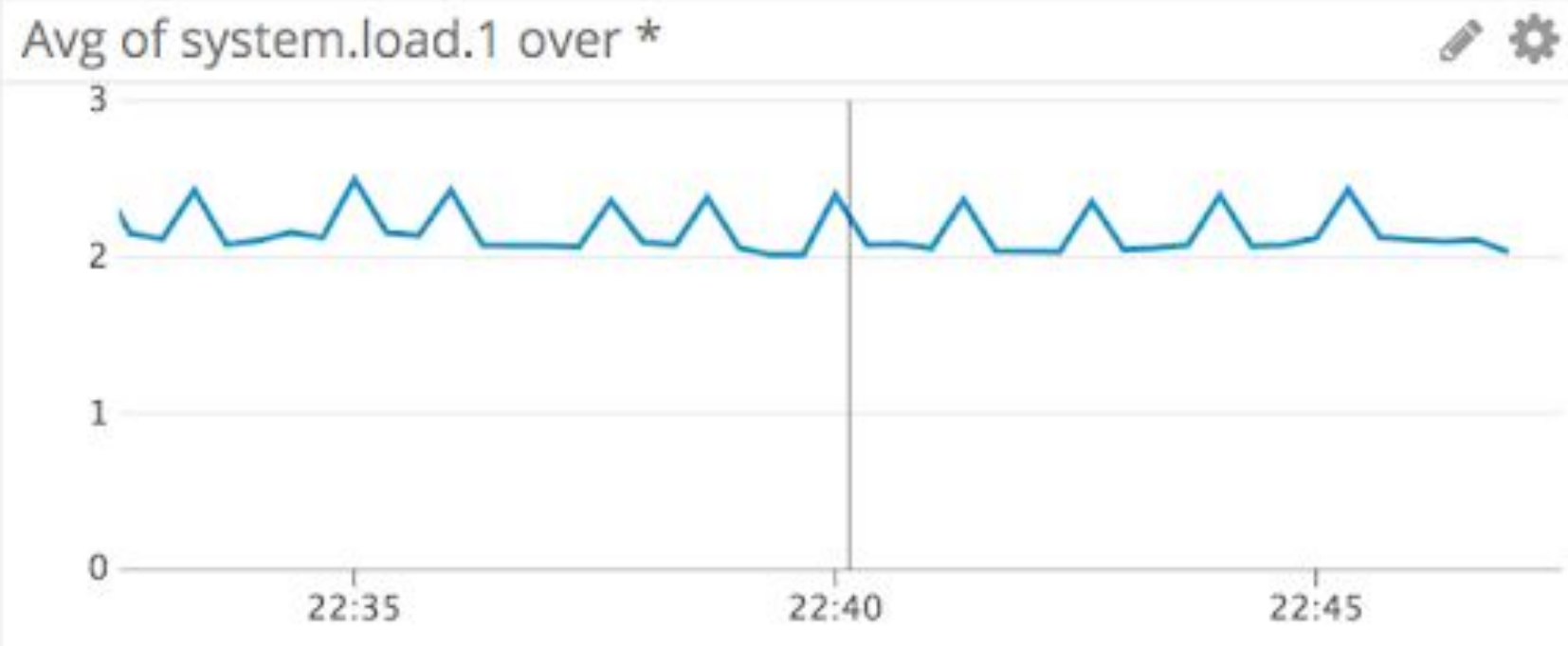
- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Image placeholder

Images are to be anchored to the right side of the slide and centered vertically.

Full-screen images should fill the page. If a border is needed, use matching section color. See next slide for example.



\$var * 📏 ?Show 15m The Past 15 Minutes

1 Make a request

GET

Advanced options

Locations

- All Locations (9)
- Americas (4)
 - US West (Oregon)
 - US East (N. Virginia)
 - South America (São Paulo)
 - Canada (Central)
- Europe (2)
 - EU (Frankfurt)
 - EU (London)
- Asia (2)
 - Asia Pacific (Tokyo)
 - Asia Pacific (Mumbai)
- Oceania (1)
 - Asia Pacific (Sydney)

How often should we send the request?



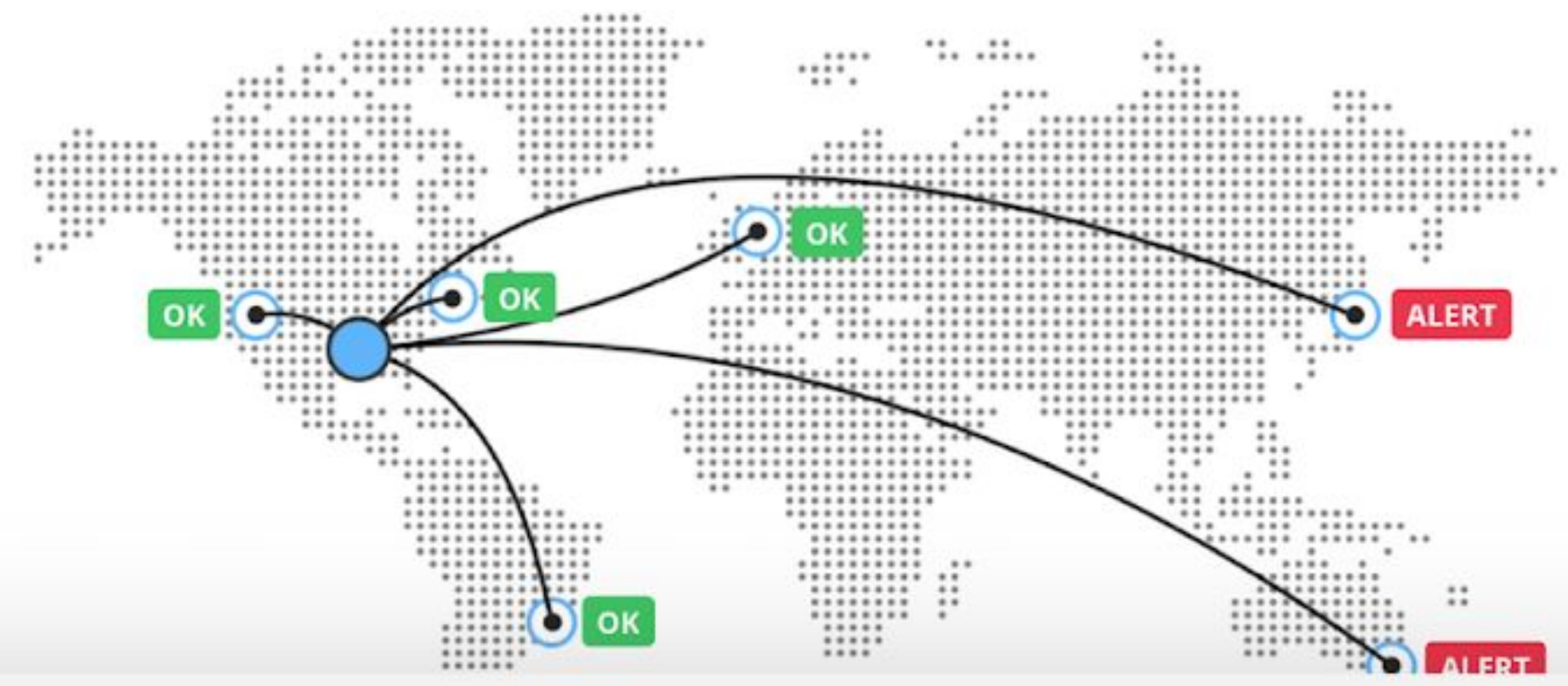
Test the request ✓

2 Validate Results

3 Say what's happening

4 Notify your team

Get started with Synthetics



Properties

Synthetics Check
Created on Oct 30, 8:40 am by Gabin Marignier
Modified on Nov 16, 1:53 pm by Albert Wang

OVERVIEW GET <https://api.datadoghq.com/api/v0/synthetics/checks> from 1 location every 60 s

3 assertions have been configured

- Header - content-type should be application/json
- Response Time should be less than 450ms
- Status Code should be 200

TAGS team:synthetics



Nov 16, 2:20 pm 0.39 s GET <https://api.datadoghq.com/api/v0/synthetics/...> 403 FORBID... America 1 of 3 assertions **FAILED**

Assertions

Property	Key	Comparison	Expected Value	Actual Value	Result
header	content-type	is	application/json	application/json	OK
responseTime	-	lessThan	450	390	OK
statusCode	-	is	200	403	FAILED

Trace

mcnulty-web | pylons.request api/synthetics.get_all_checks

i-00eadfed030fcbd80 14.2 ms (100.0% of total time)

Span Metadata **Host Info** Logs

CPU usage (%) [Bar chart]

Load Averages 1-5-15 [Line chart]

Chef
env:prod role:mcnulty-web

Datadog
datacenter:us1.prod.dog service:mcnulty-web site:datadoghq.com

Section divider 2

Large text can
be used to make
a single
important point.

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Standard slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Slide with two points

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Slide with three points

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

Slide with four points

Identify and convince great people to join

Great = able + driven + grown-up

Hiring is what makes/breaks us

Expect to spend 10-20% of your time on hiring

Big Statement

1

Geometric
top-line growth

2

Strong
product-market fit

3

Financially efficient
go-to-market

Use iconography to support content



Cras mauris purus,
varius in justo in, varius
vulputate eros.

Secondary text, if needed



Curabitur pharetra
dictum mattis curabitur
dictum.

Secondary text, if needed



Proin iaculis nulla
turpis, a fermentum
orci mattis in.

Secondary text, if needed

The Pillars of Observability



Logs



Traces



Metrics

White background slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Text+image slide

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Image placeholder

Images are to be anchored to the right side of the slide and centered vertically.

Full-screen images should fill the page. If a border is needed, use matching section color. See next slide for example.

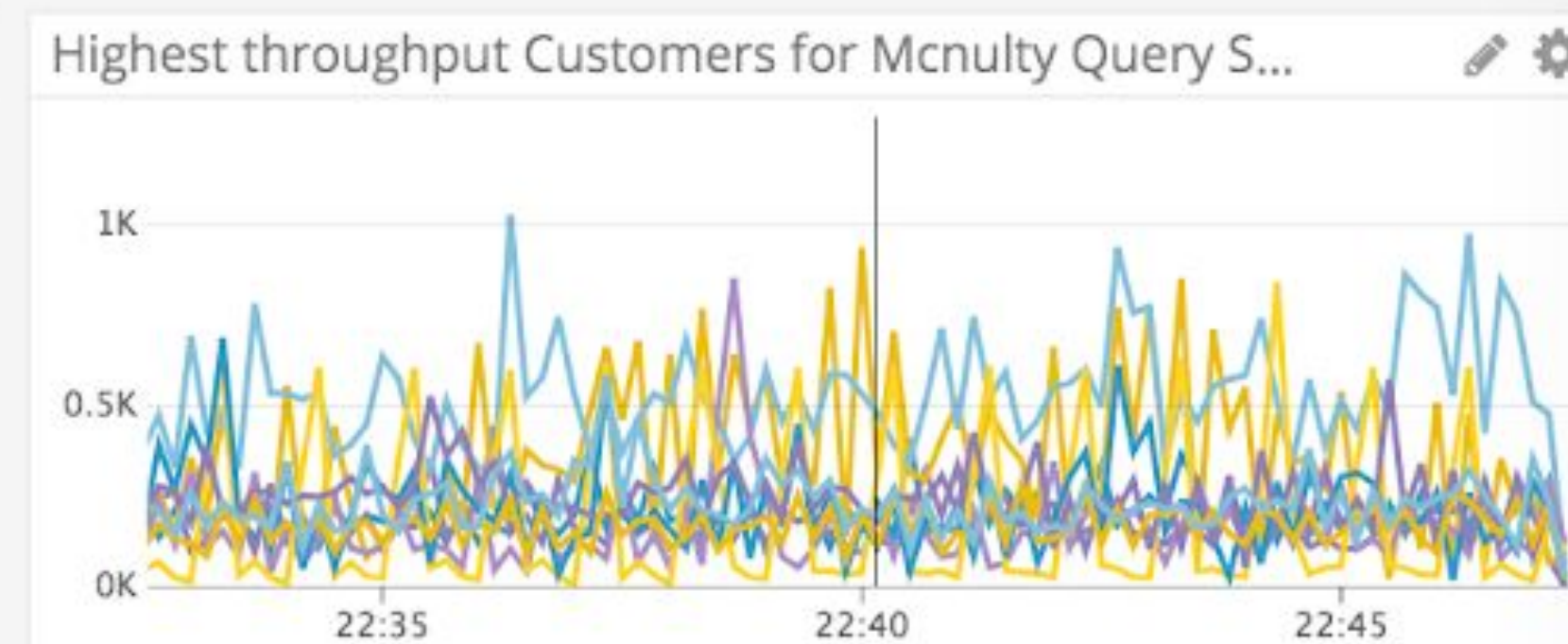
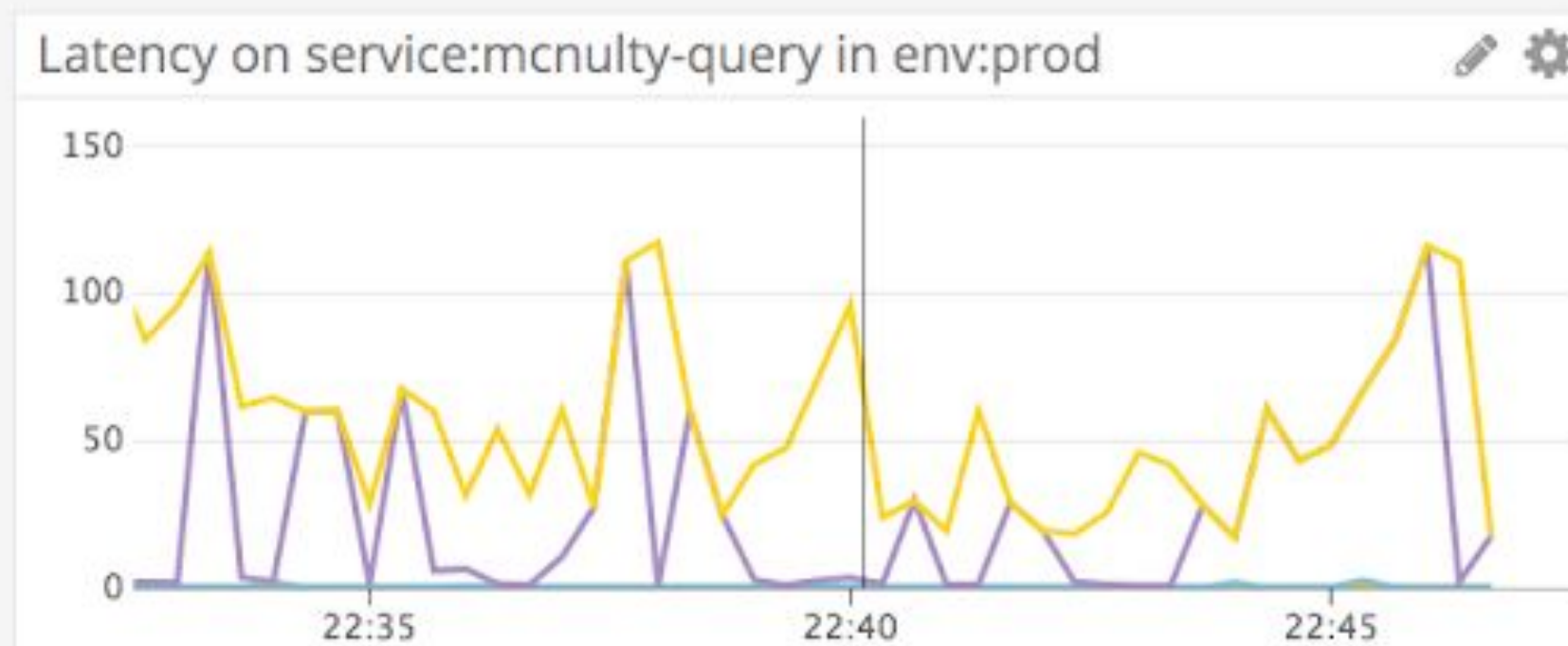
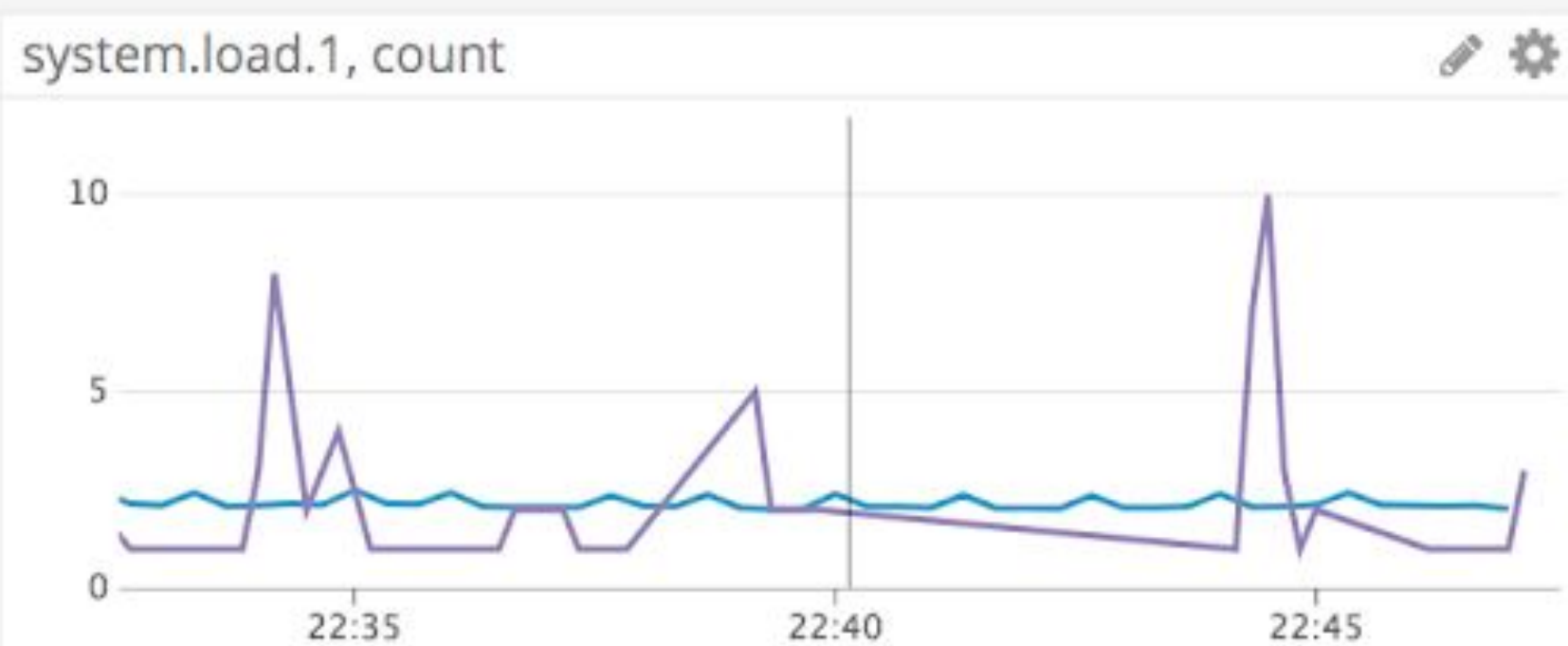
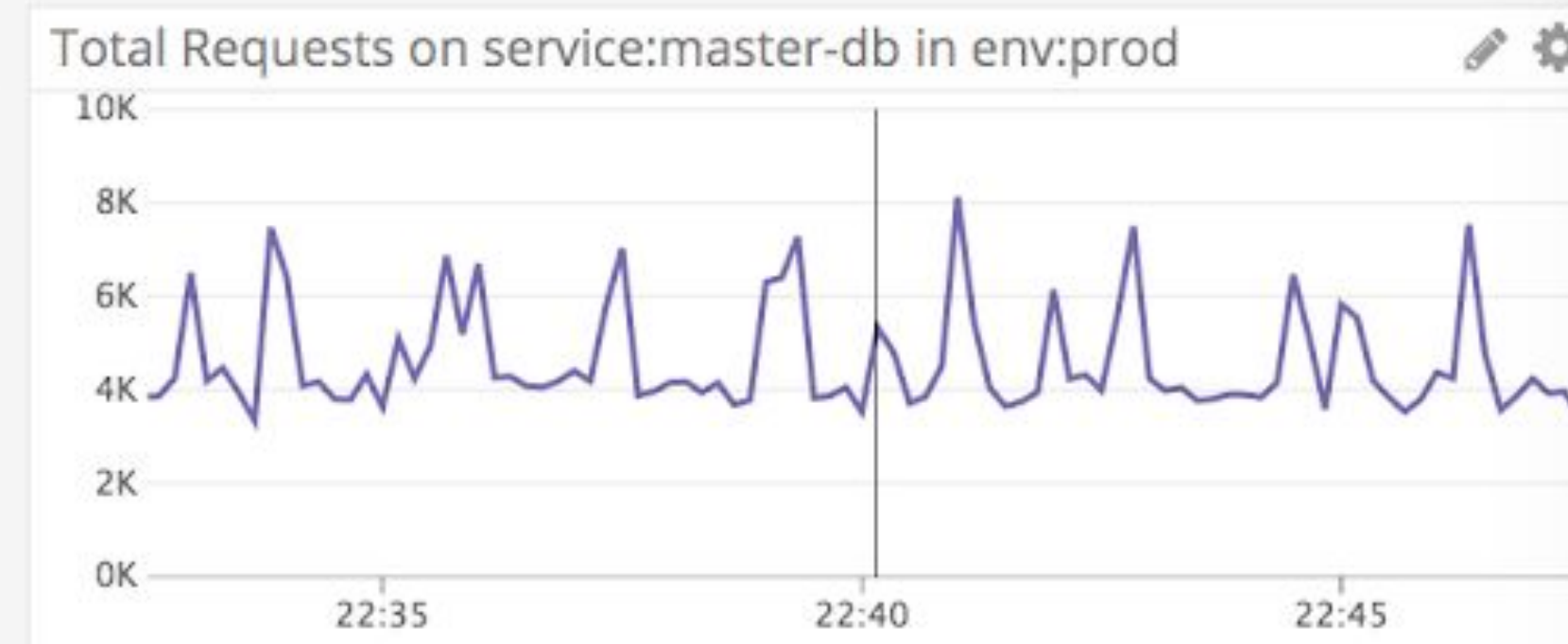
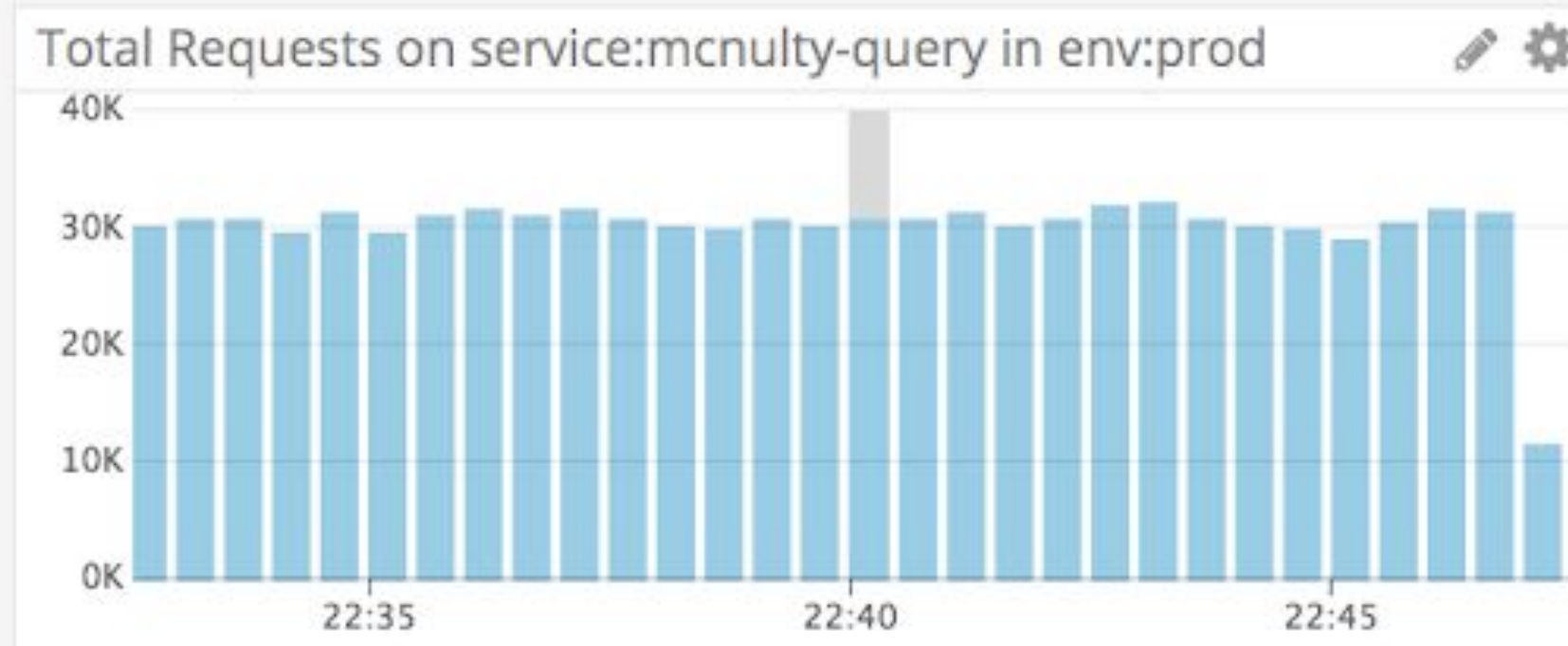
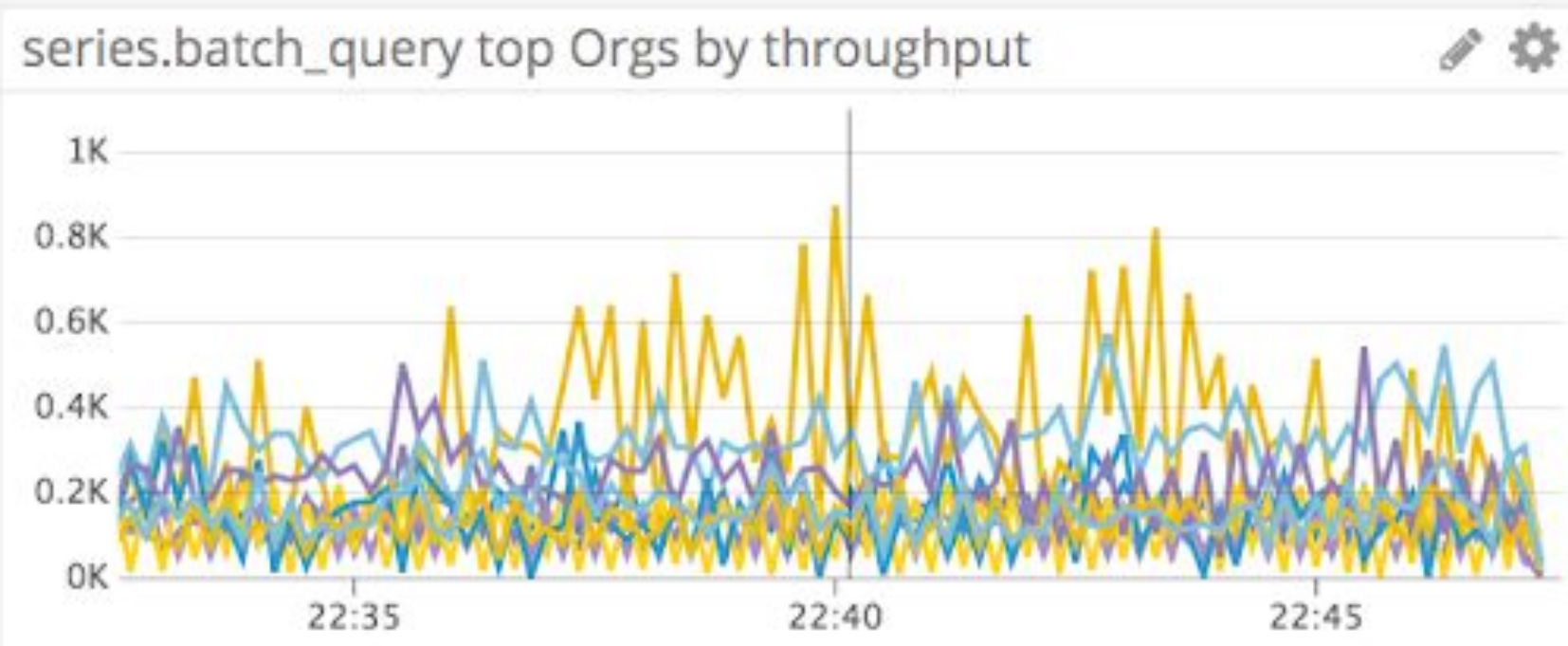
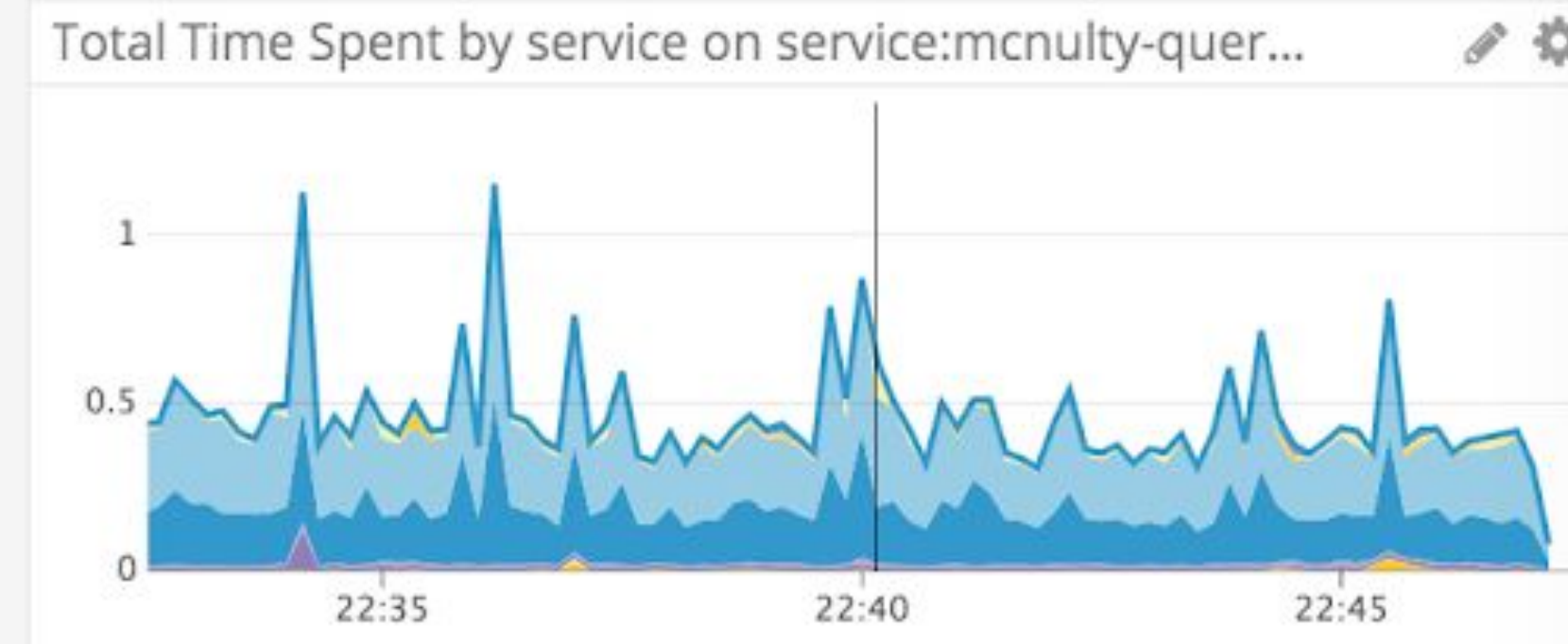
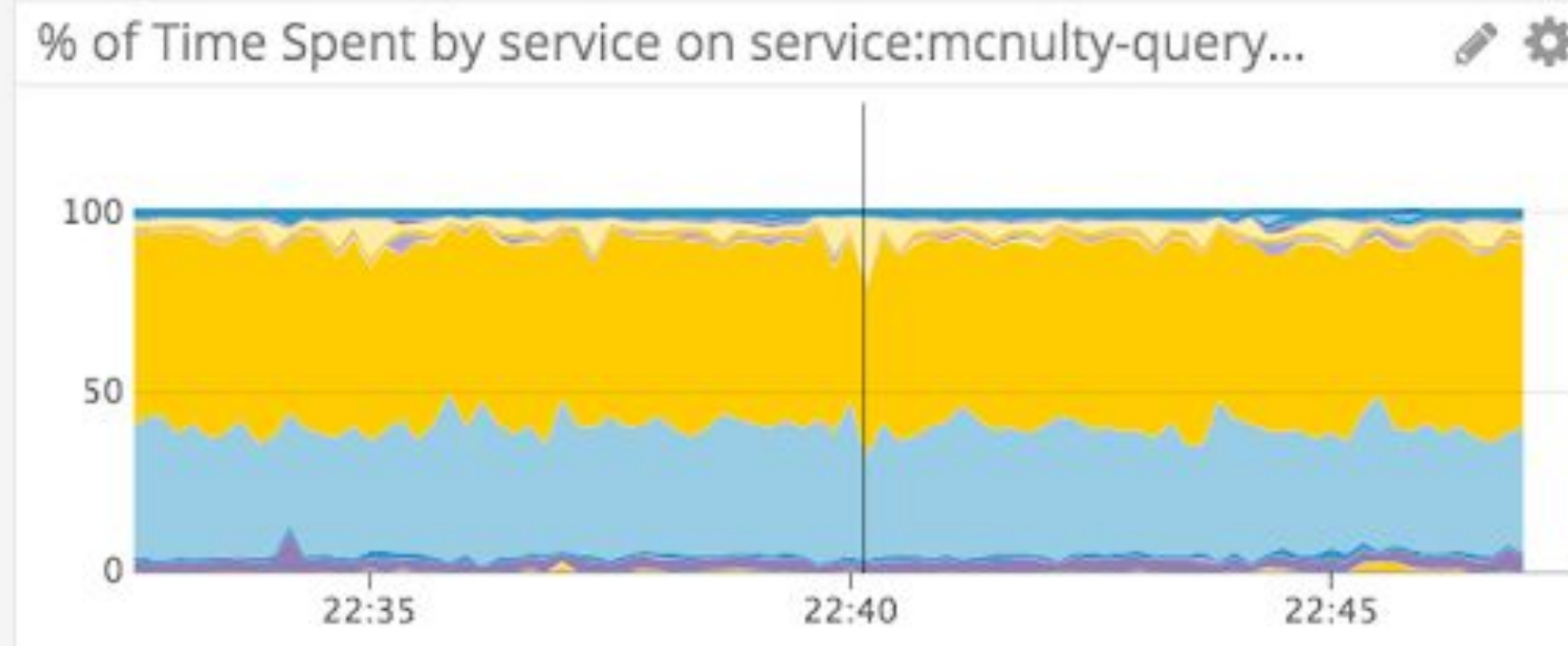
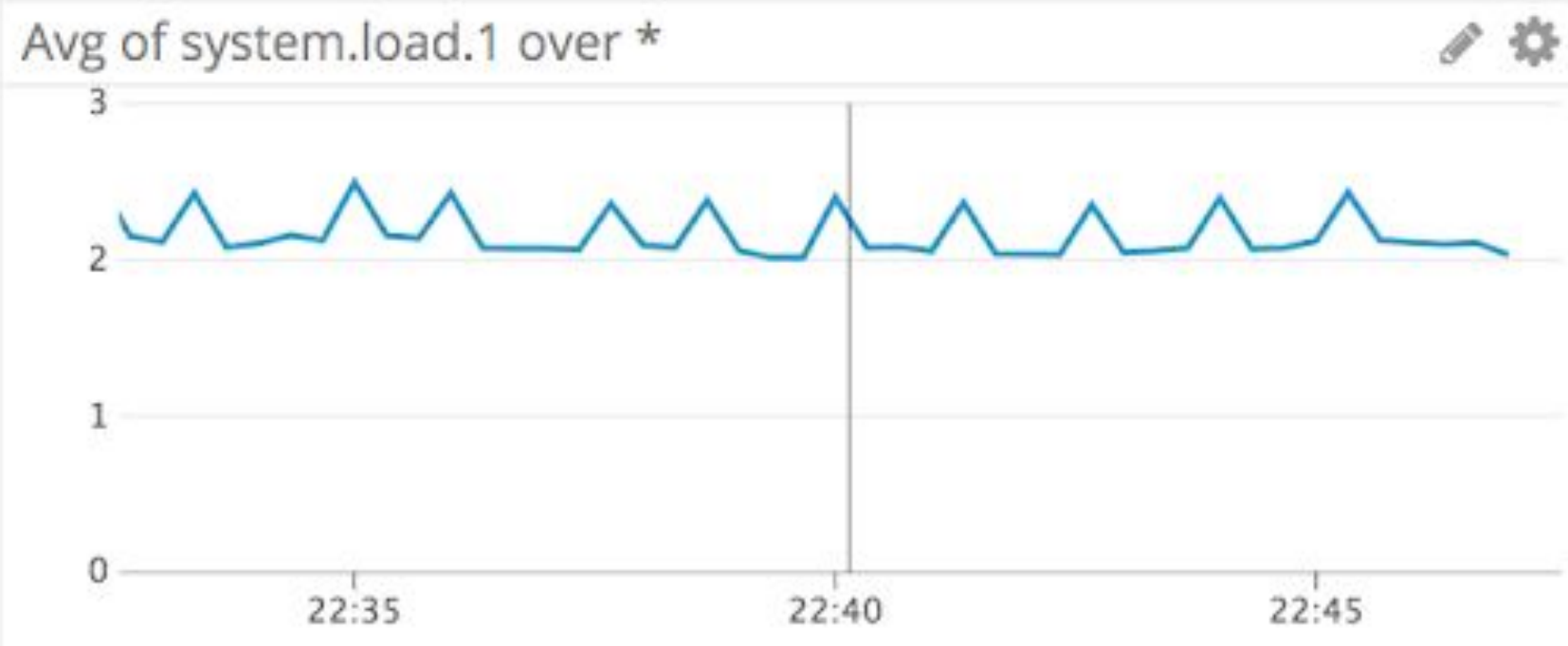


\$var * 📏 ?

🔍 Search events to overlay...



Show 15m The Past 15 Minutes



Section divider 3

Large text can
be used to make
a single
important point.

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Standard slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Slide with two points

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Slide with three points

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

Slide with four points

Identify and convince great people to join

Great = able + driven + grown-up

Hiring is what makes/breaks us

Expect to spend 10-20% of your time on hiring

Big Statement

1

Geometric
top-line growth

2

Strong
product-market fit

3

Financially efficient
go-to-market

Use iconography to support content



Cras mauris purus,
varius in justo in, varius
vulputate eros.

Secondary text, if needed



Curabitur pharetra
dictum mattis curabitur
dictum.

Secondary text, if needed



Proin iaculis nulla
turpis, a fermentum
orci mattis in.

Secondary text, if needed

The Pillars of Observability



Logs



Traces



Metrics

White background slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Text+image slide

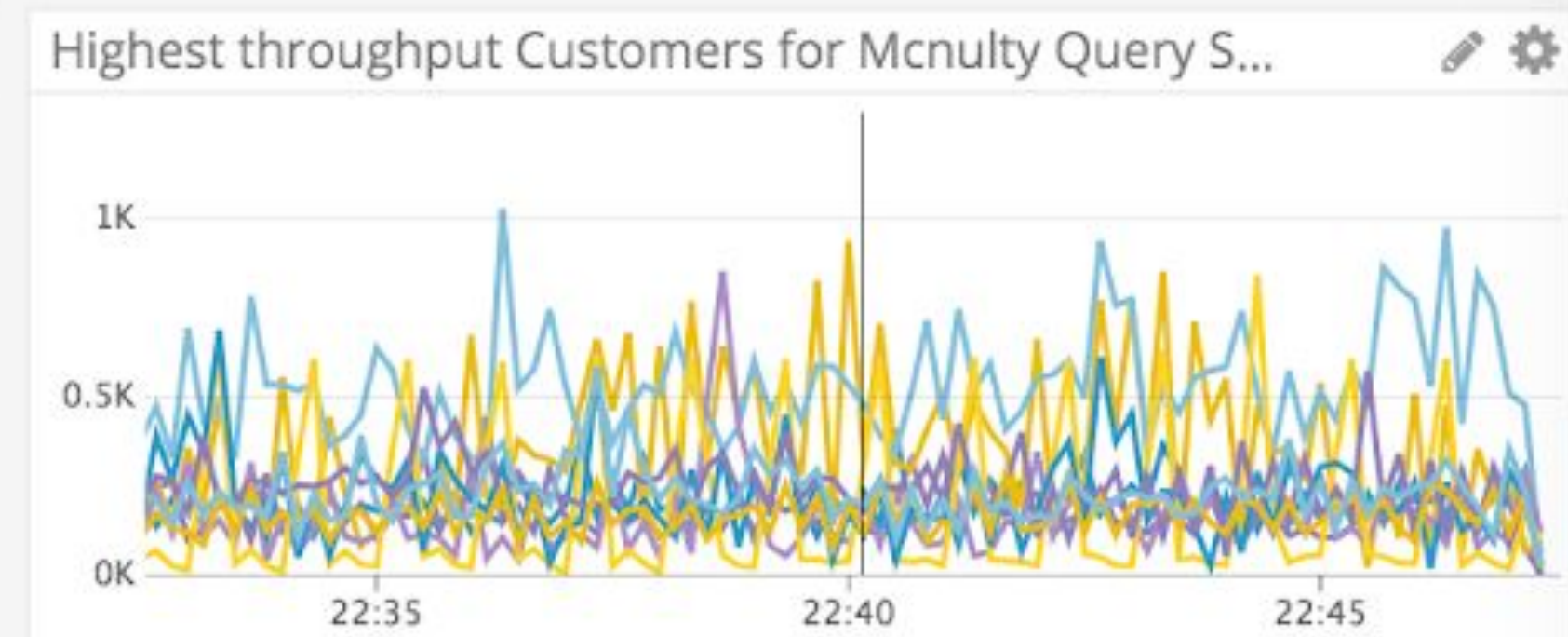
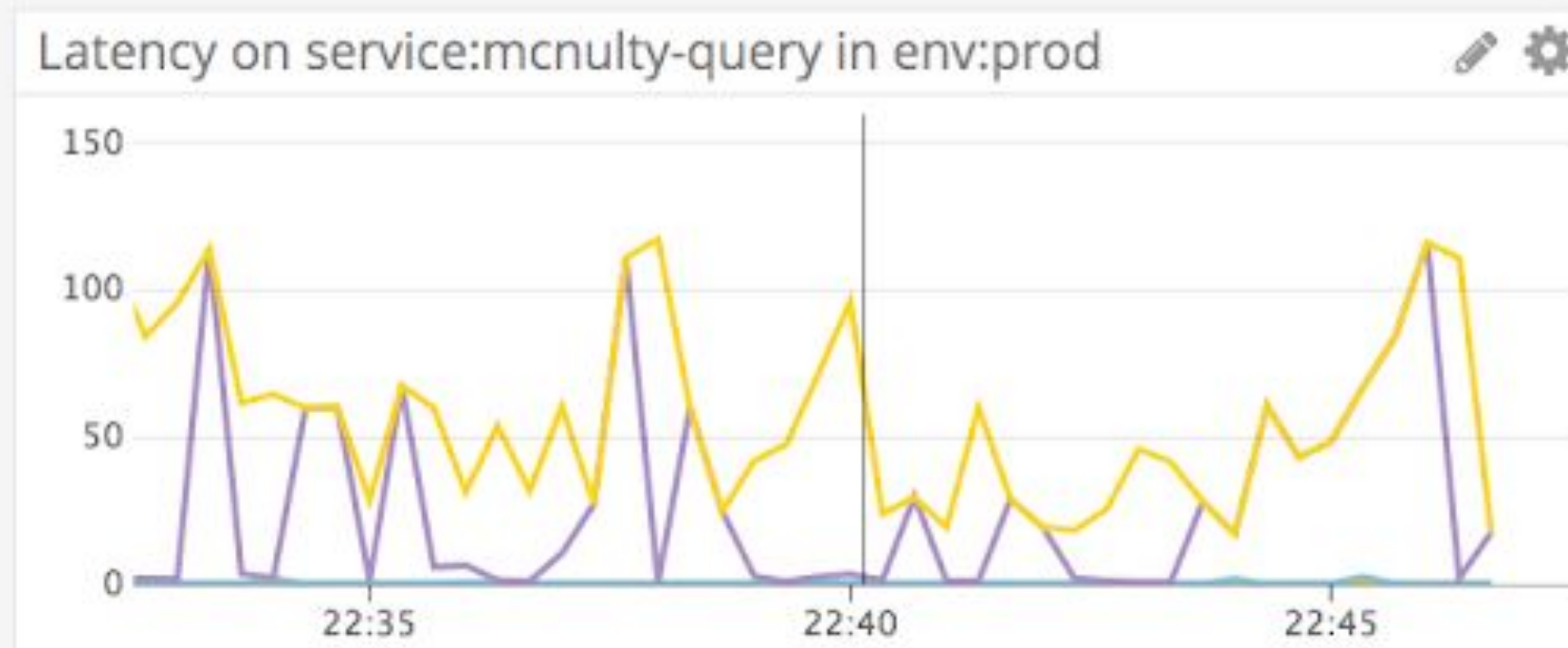
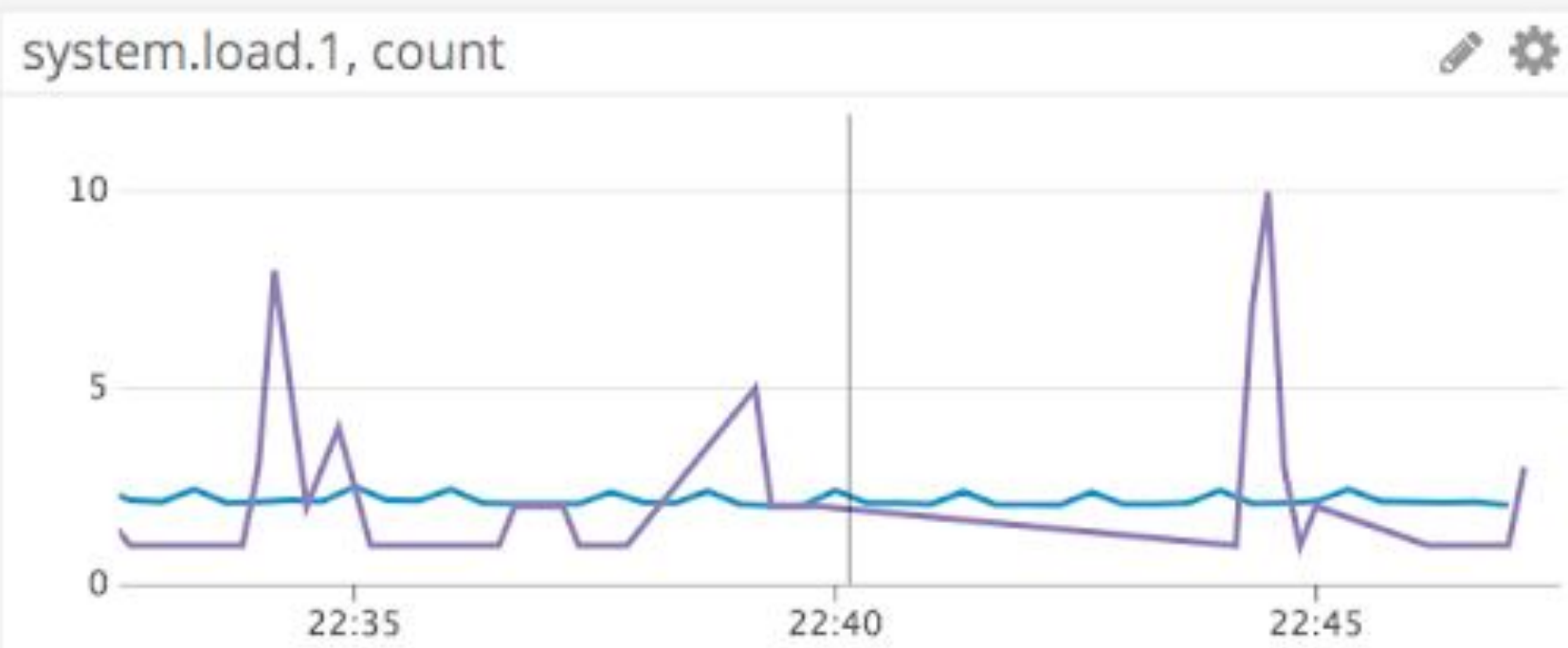
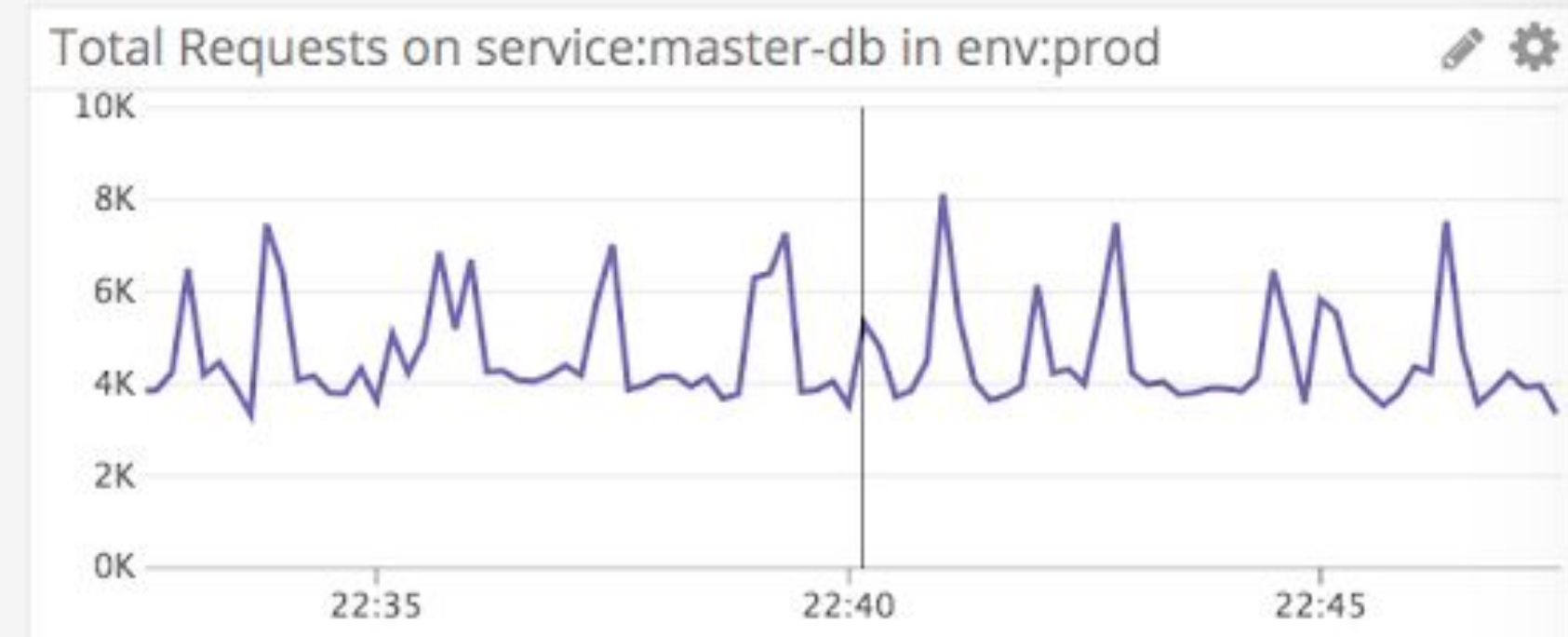
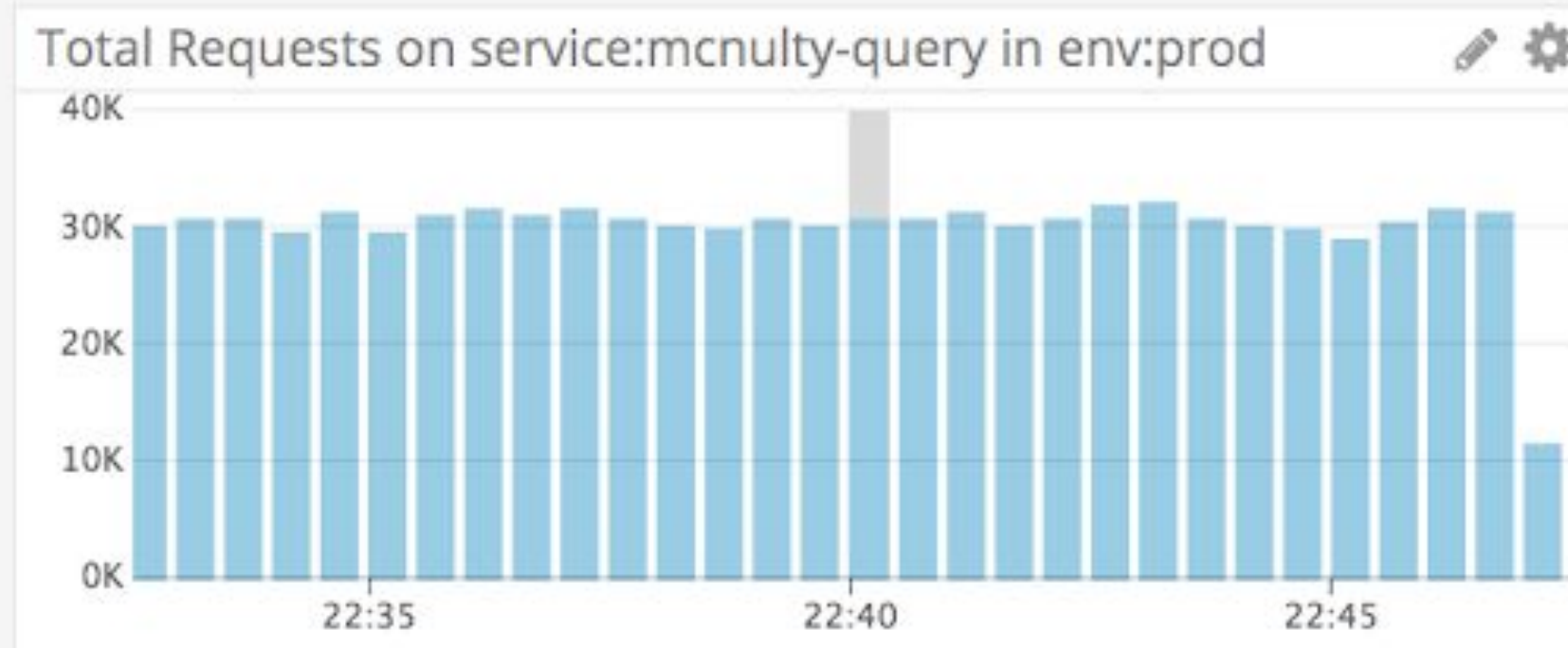
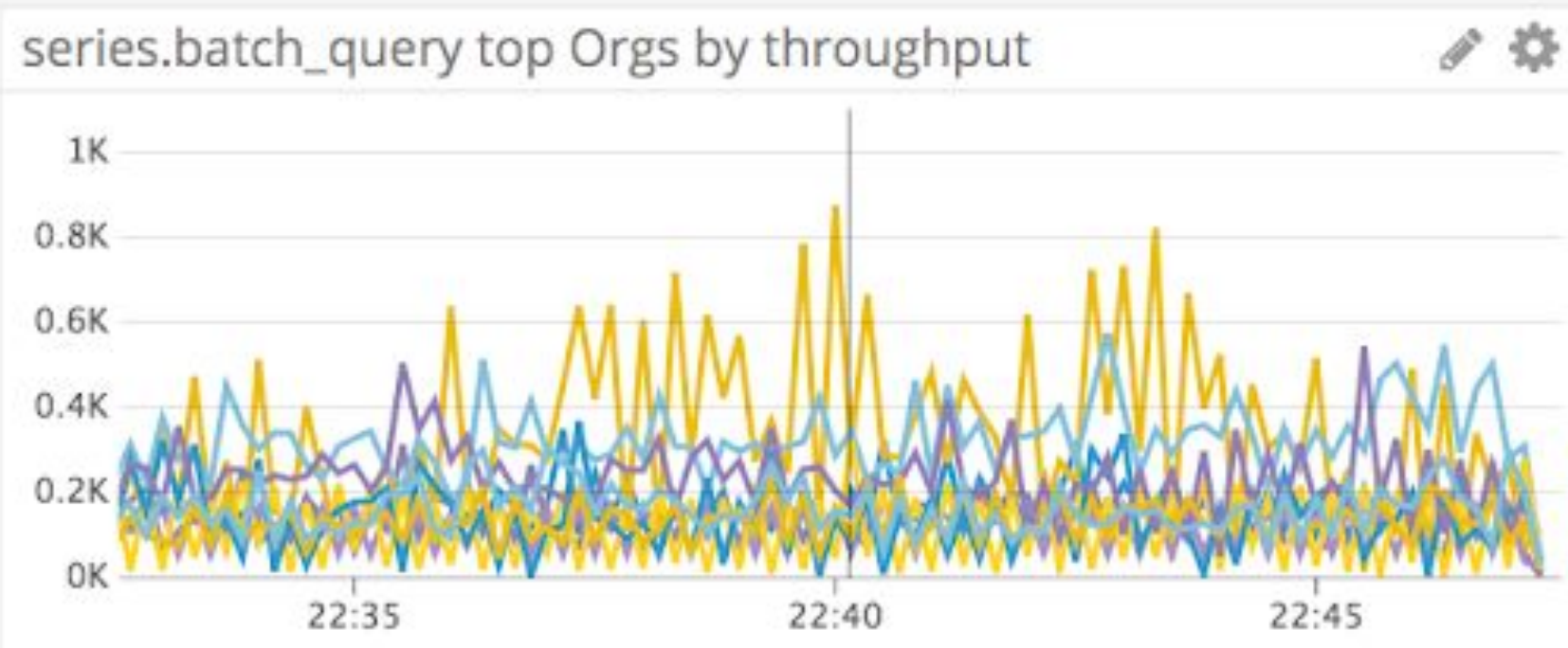
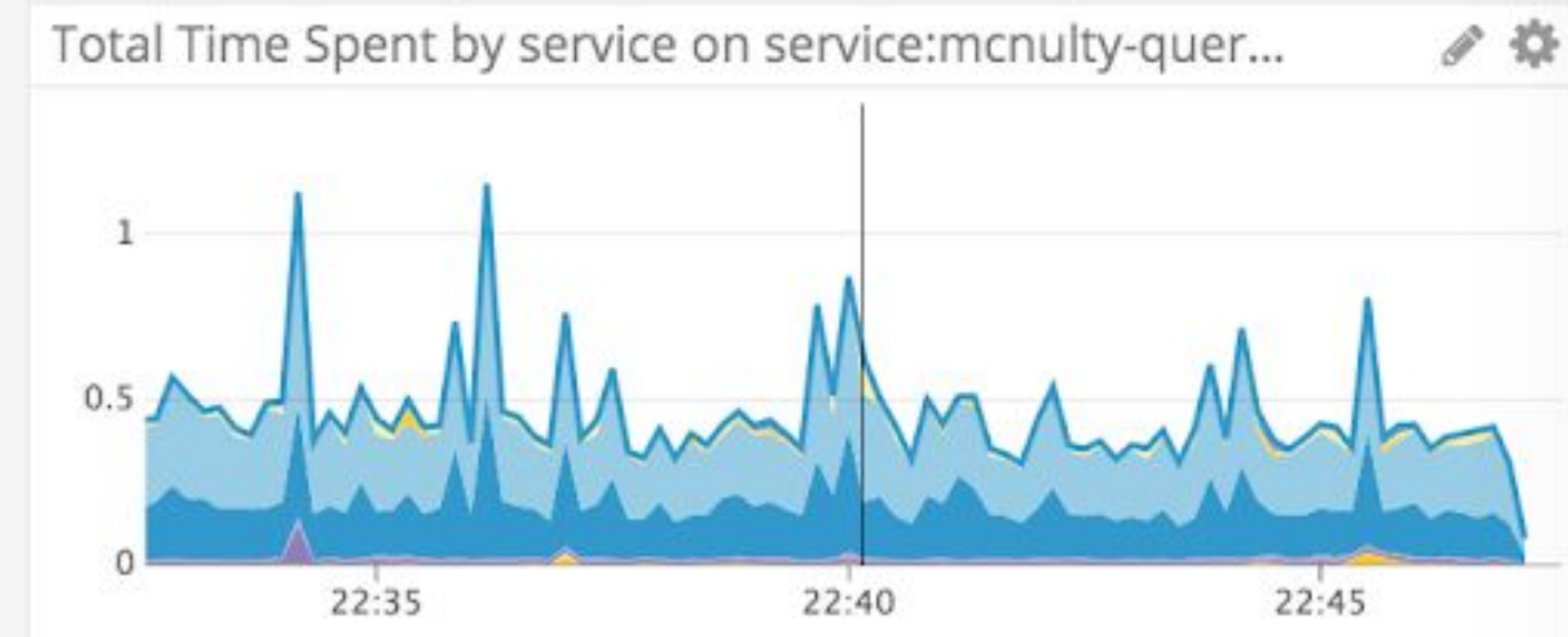
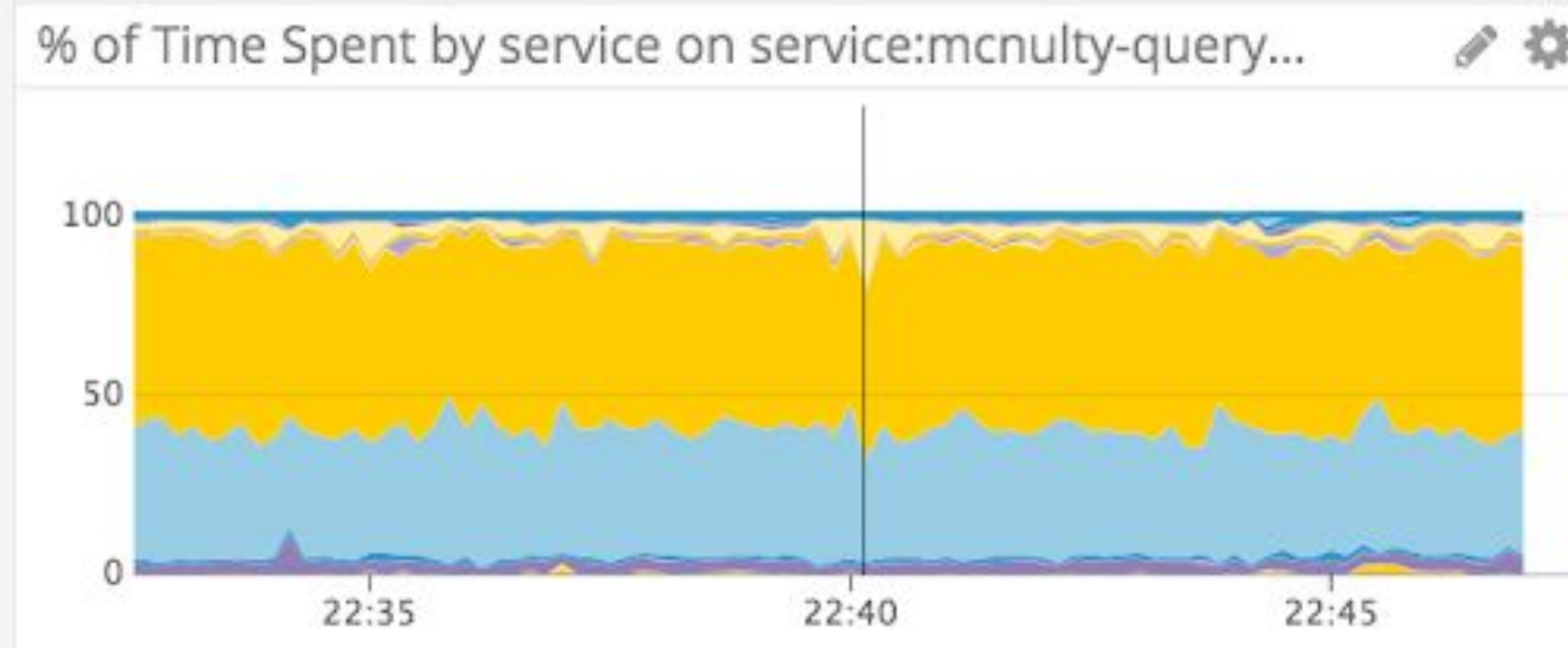
- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Image placeholder

Images are to be anchored to the right side of the slide and centered vertically.

Full-screen images should fill the page. If a border is needed, use matching section color. See next slide for example.



\$var * 📏 ?Show 15m The Past 15 Minutes



Section divider 4

Large text can
be used to make
a single
important point.

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Standard slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Slide with two points

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Slide with three points

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

Slide with four points

Identify and convince great people to join

Great = able + driven + grown-up

Hiring is what makes/breaks us

Expect to spend 10-20% of your time on hiring

Big Statement

1

Geometric
top-line growth

2

Strong
product-market fit

3

Financially efficient
go-to-market

Use iconography to support content



Cras mauris purus,
varius in justo in, varius
vulputate eros.

Secondary text, if needed



Curabitur pharetra
dictum mattis curabitur
dictum.

Secondary text, if needed



Proin iaculis nulla
turpis, a fermentum
orci mattis in.

Secondary text, if needed

The Pillars of Observability



Logs



Traces



Metrics

White background slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Text+image slide

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Image placeholder

Images are to be anchored to the right side of the slide and centered vertically.

Full-screen images should fill the page. If a border is needed, use matching section color. See next slide for example.

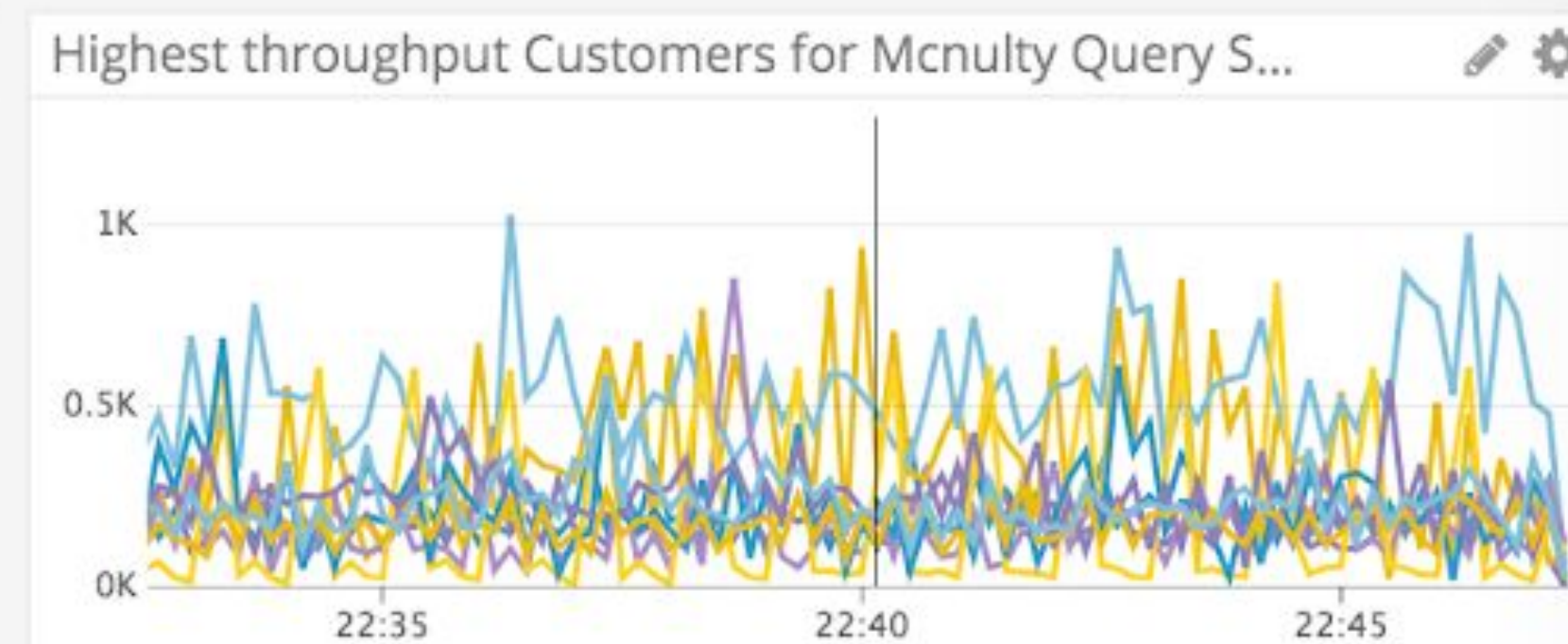
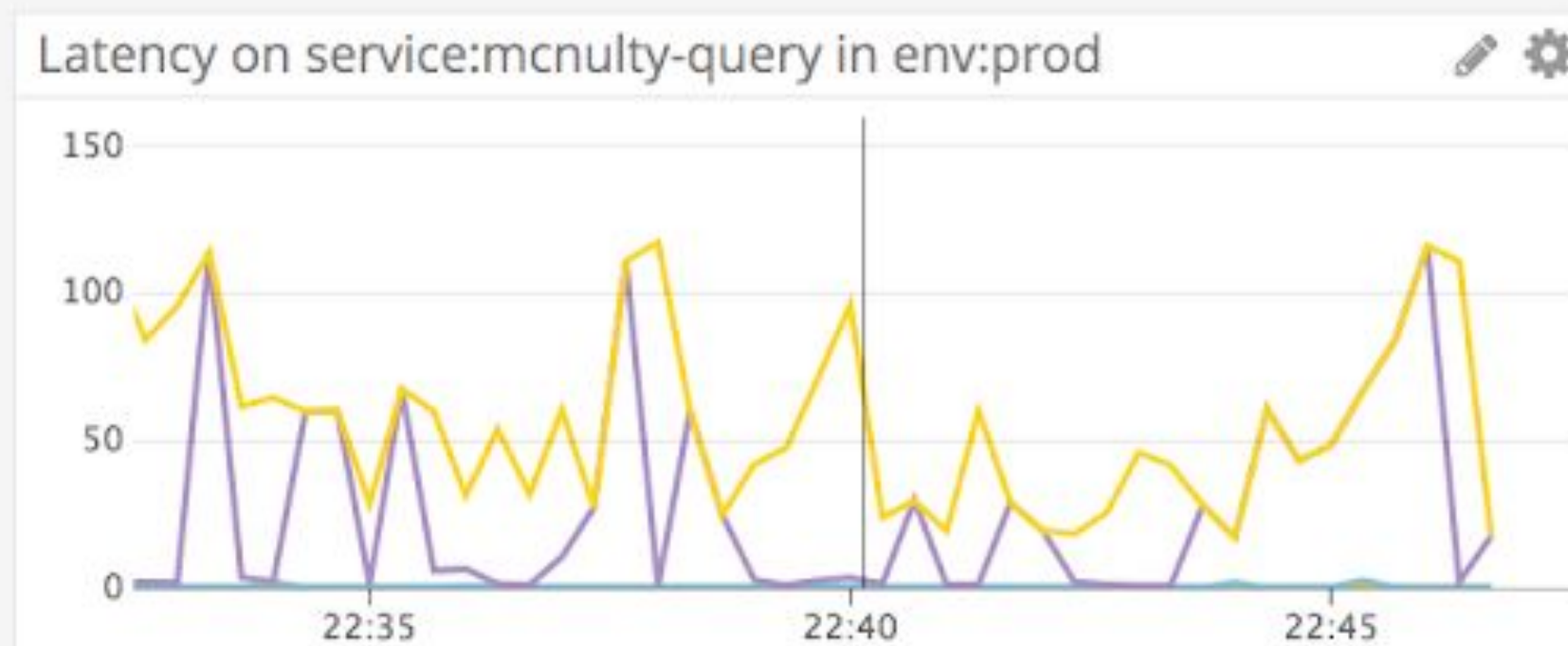
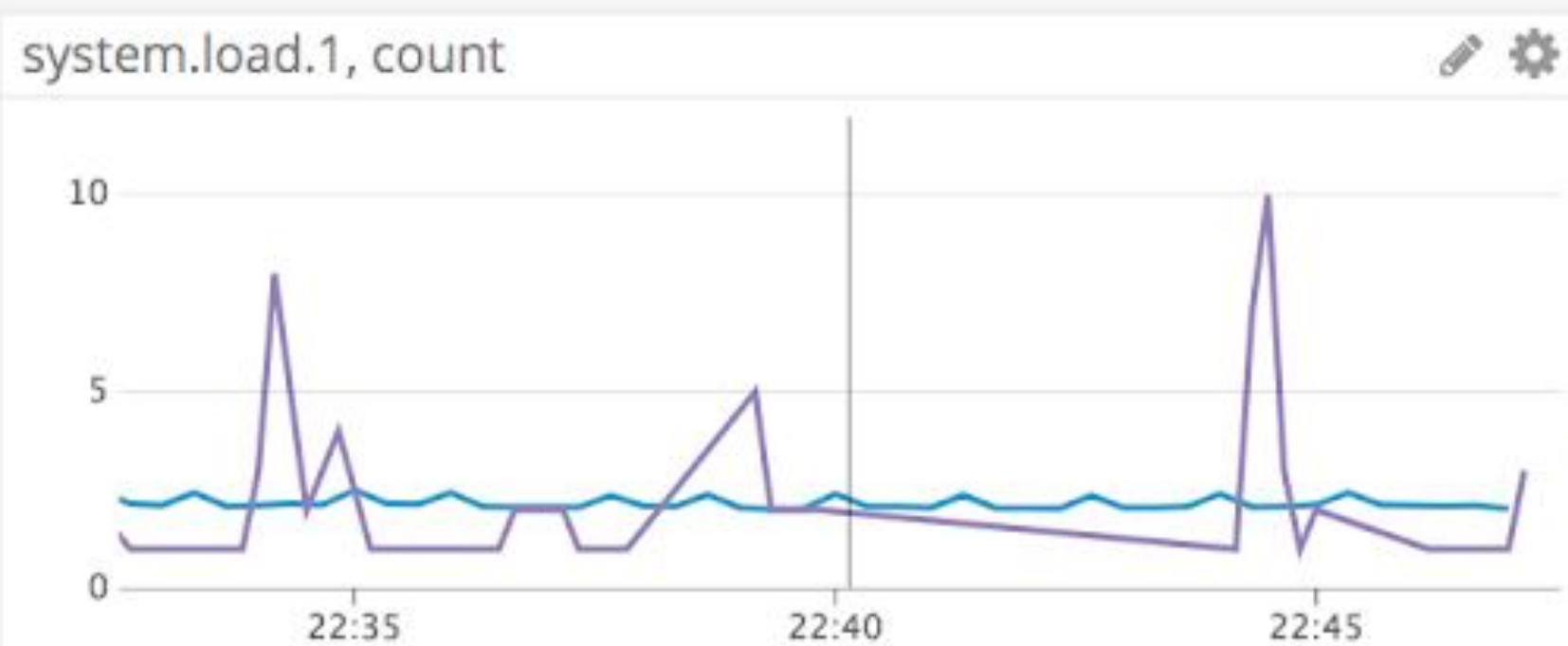
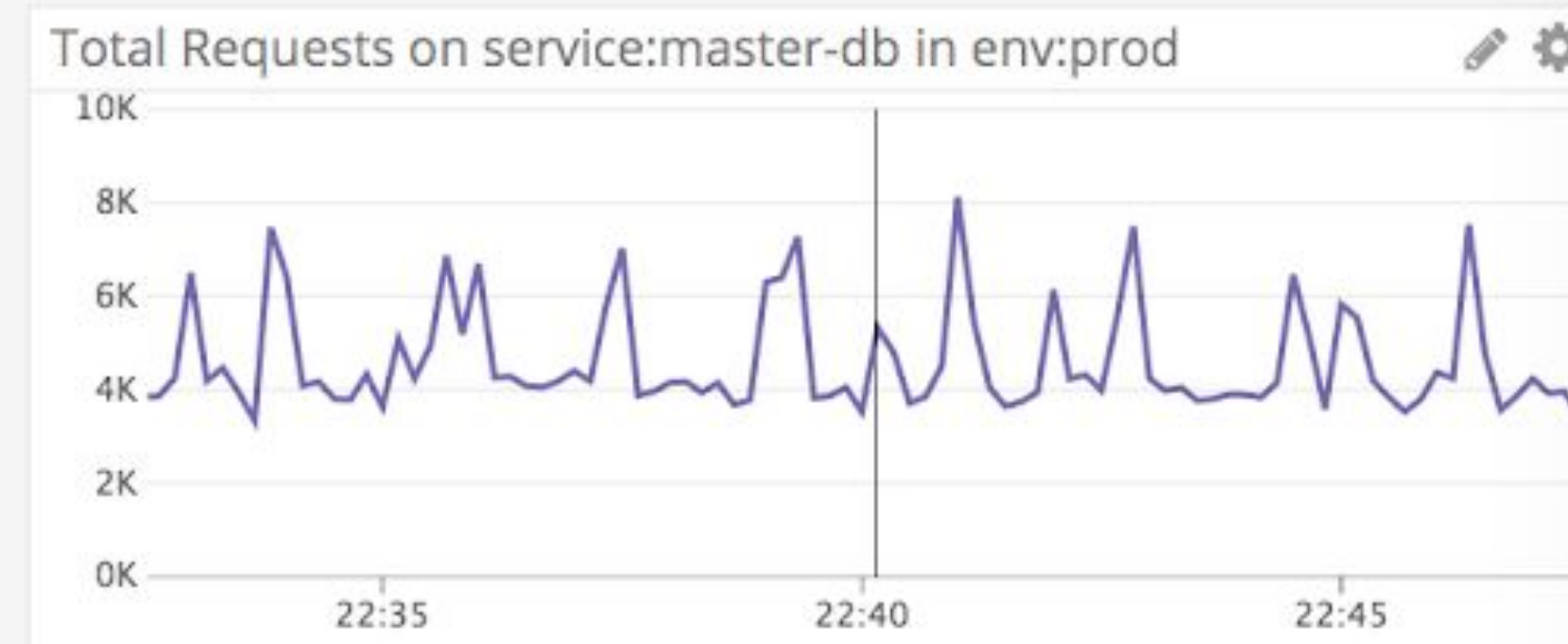
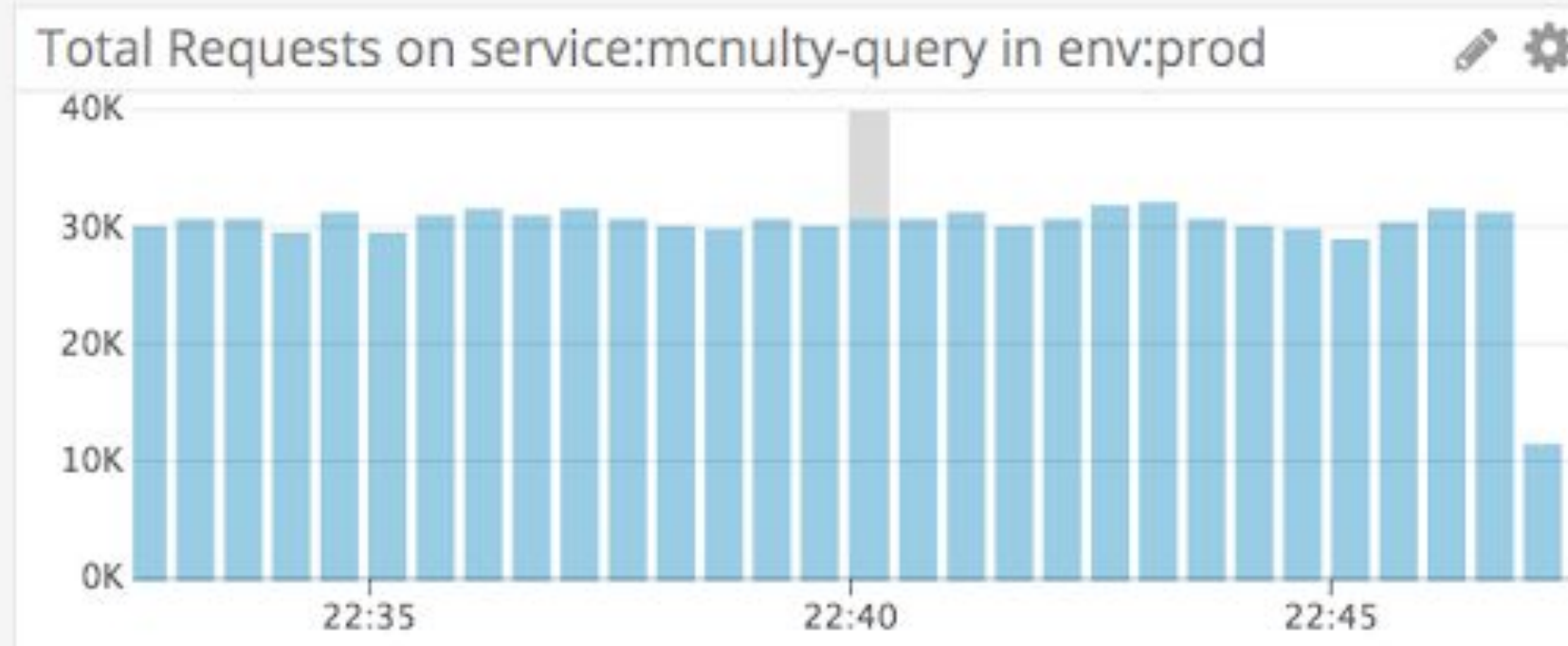
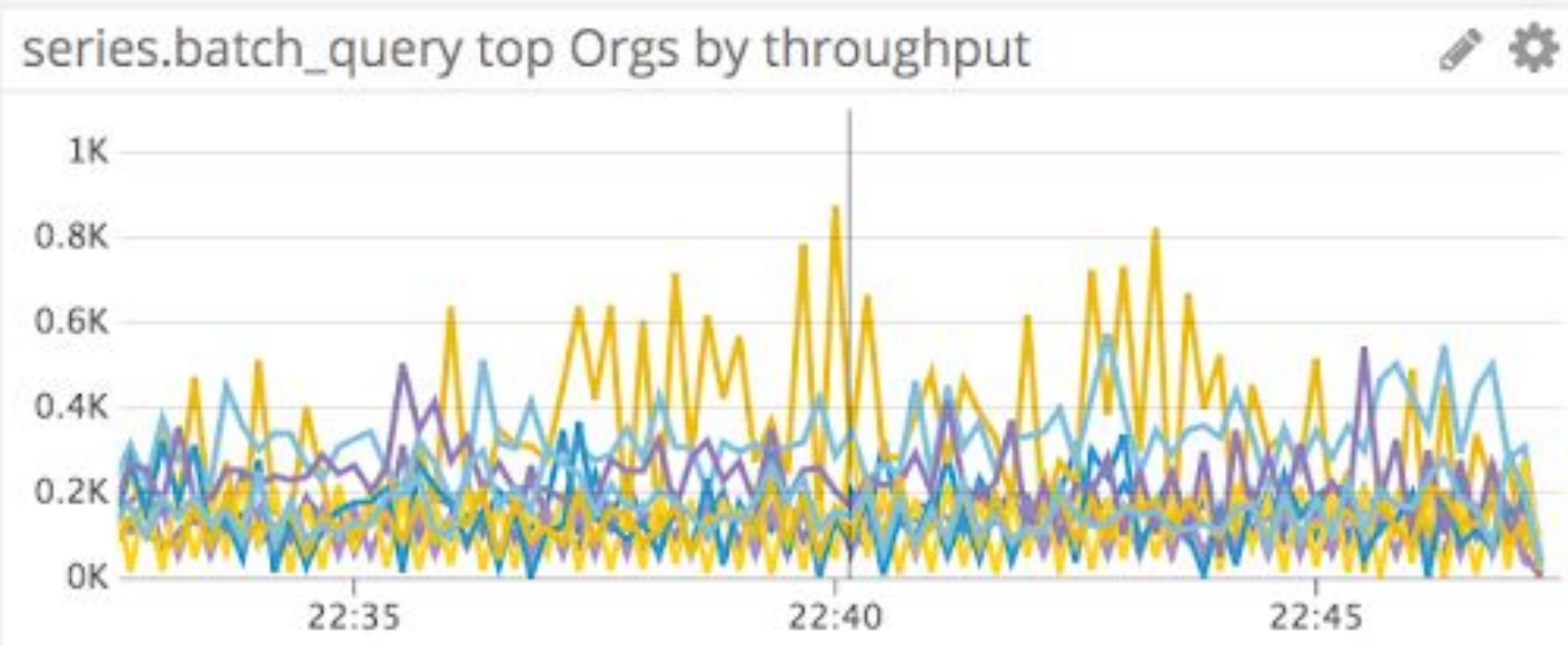
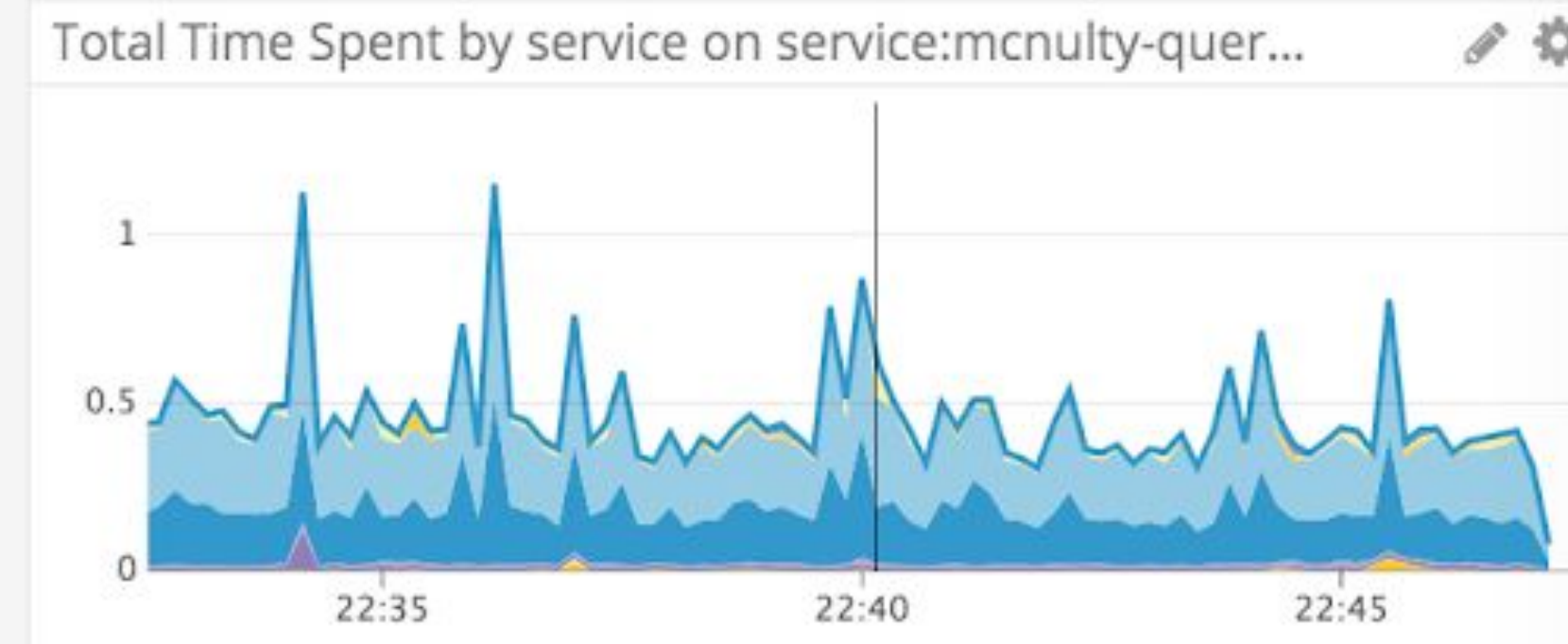
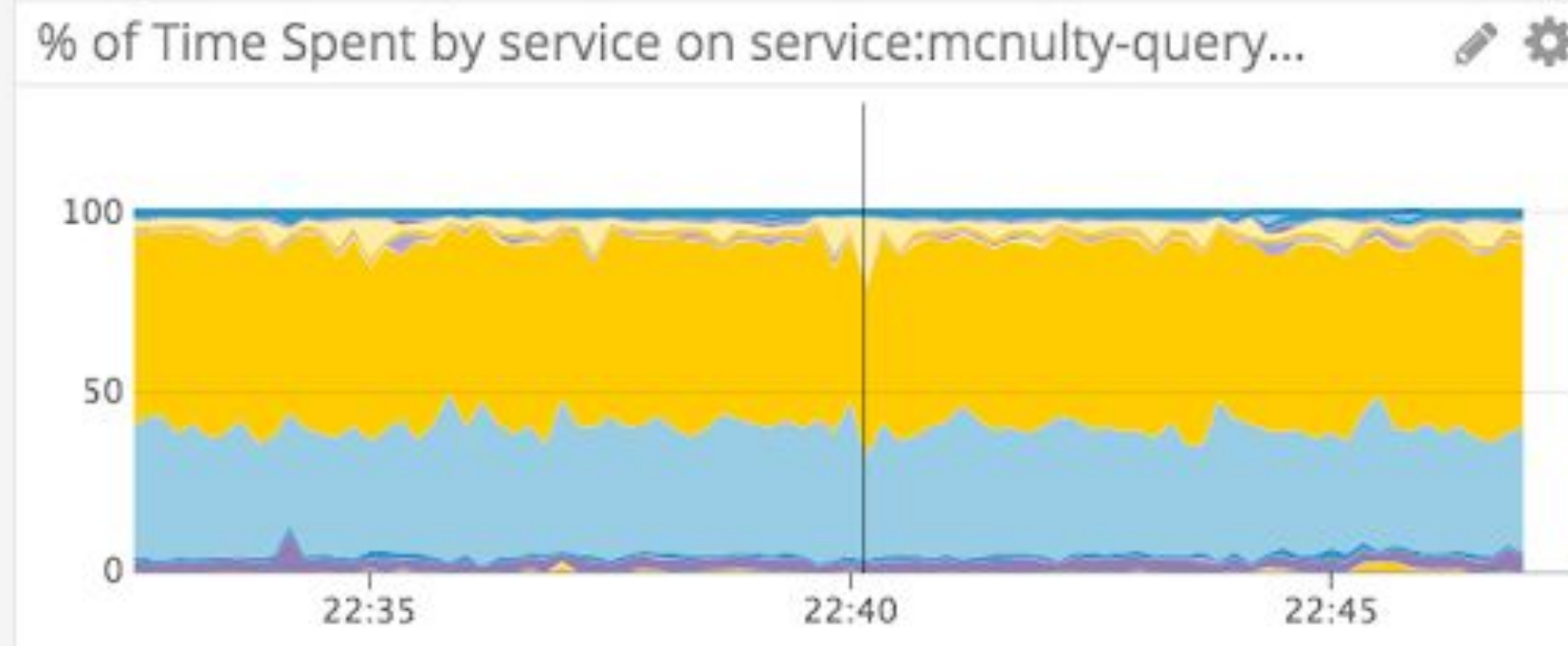
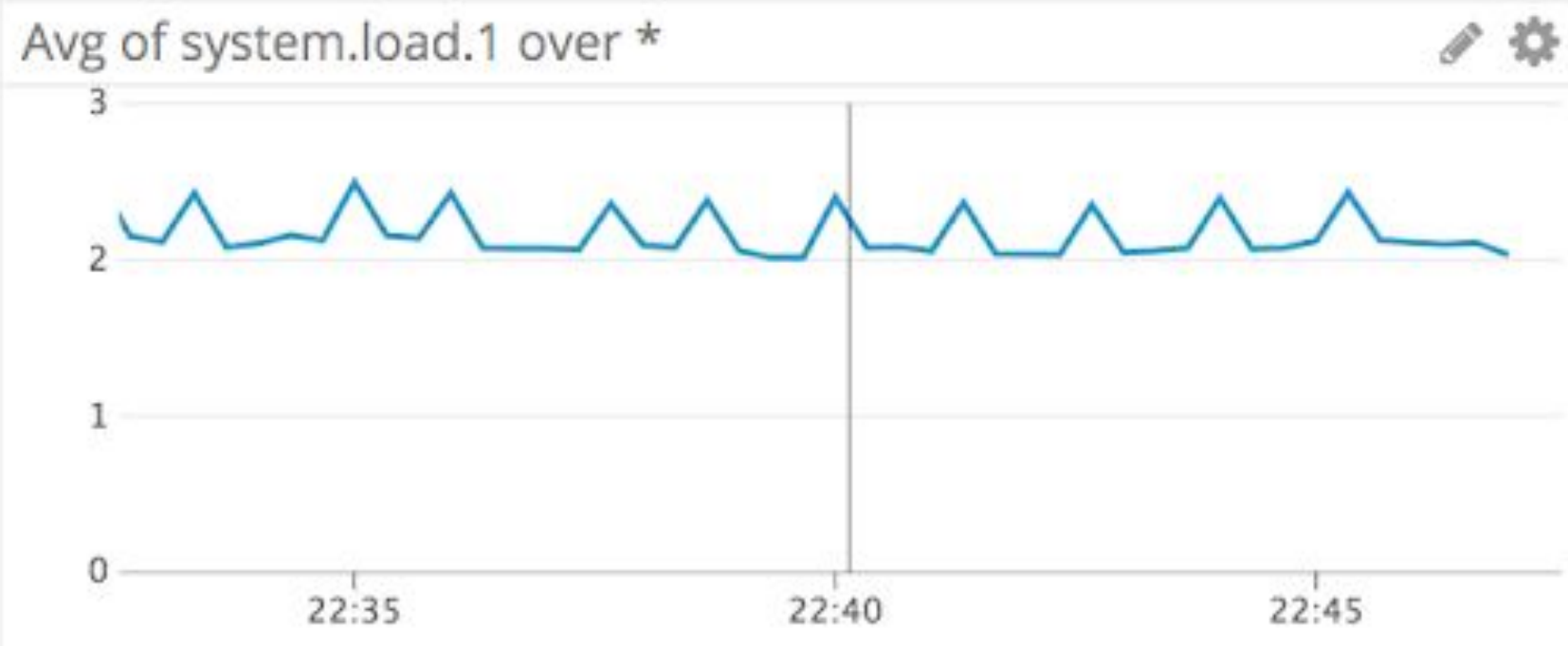


\$var * 📏 ?

🔍 Search events to overlay...



Show 15m The Past 15 Minutes



Section divider 5

Large text can
be used to make
a single
important point.

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Standard slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Slide with two points

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Slide with three points

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

Slide with four points

Identify and convince great people to join

Great = able + driven + grown-up

Hiring is what makes/breaks us

Expect to spend 10-20% of your time on hiring

Big Statement

1

Geometric
top-line growth

2

Strong
product-market fit

3

Financially efficient
go-to-market

Use iconography to support content



Cras mauris purus,
varius in justo in, varius
vulputate eros.

Secondary text, if needed



Curabitur pharetra
dictum mattis curabitur
dictum.

Secondary text, if needed



Proin iaculis nulla
turpis, a fermentum
orci mattis in.

Secondary text, if needed

The Pillars of Observability



Logs



Traces



Metrics

White background slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Text+image slide

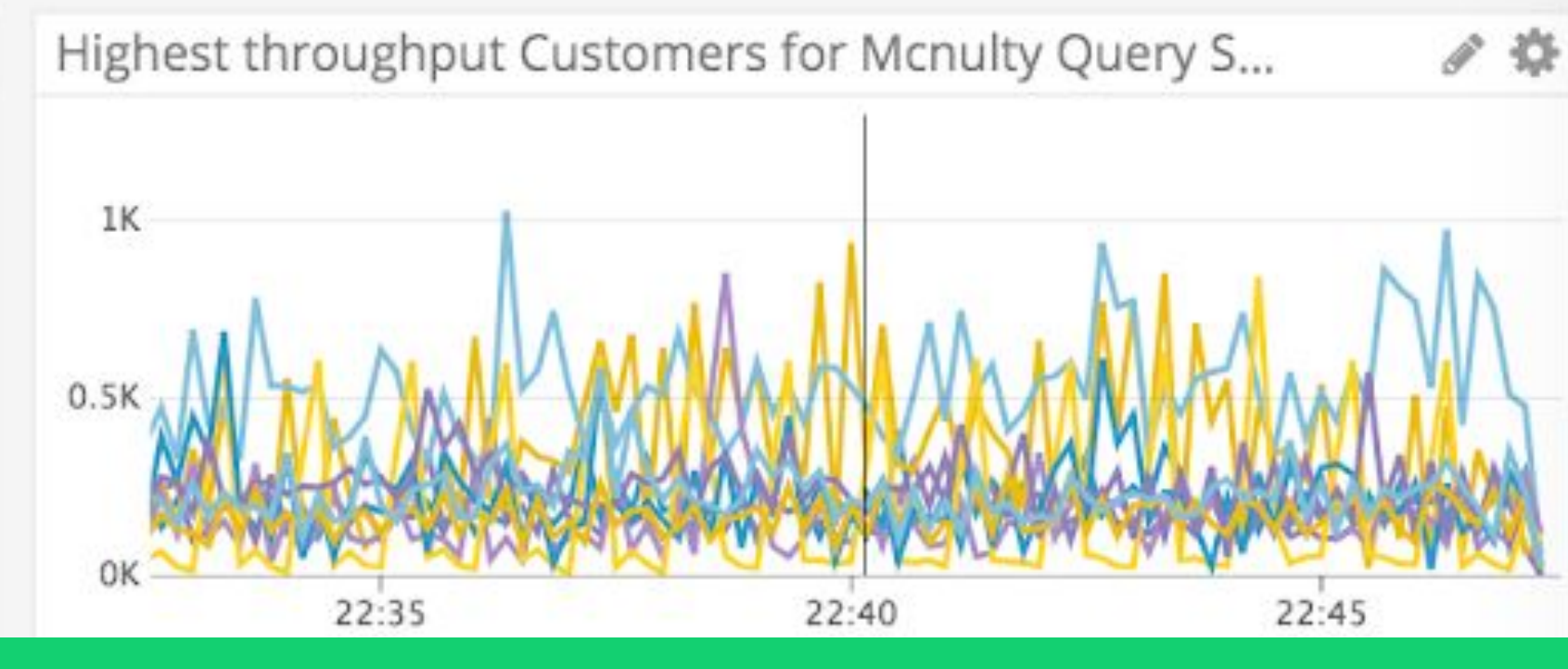
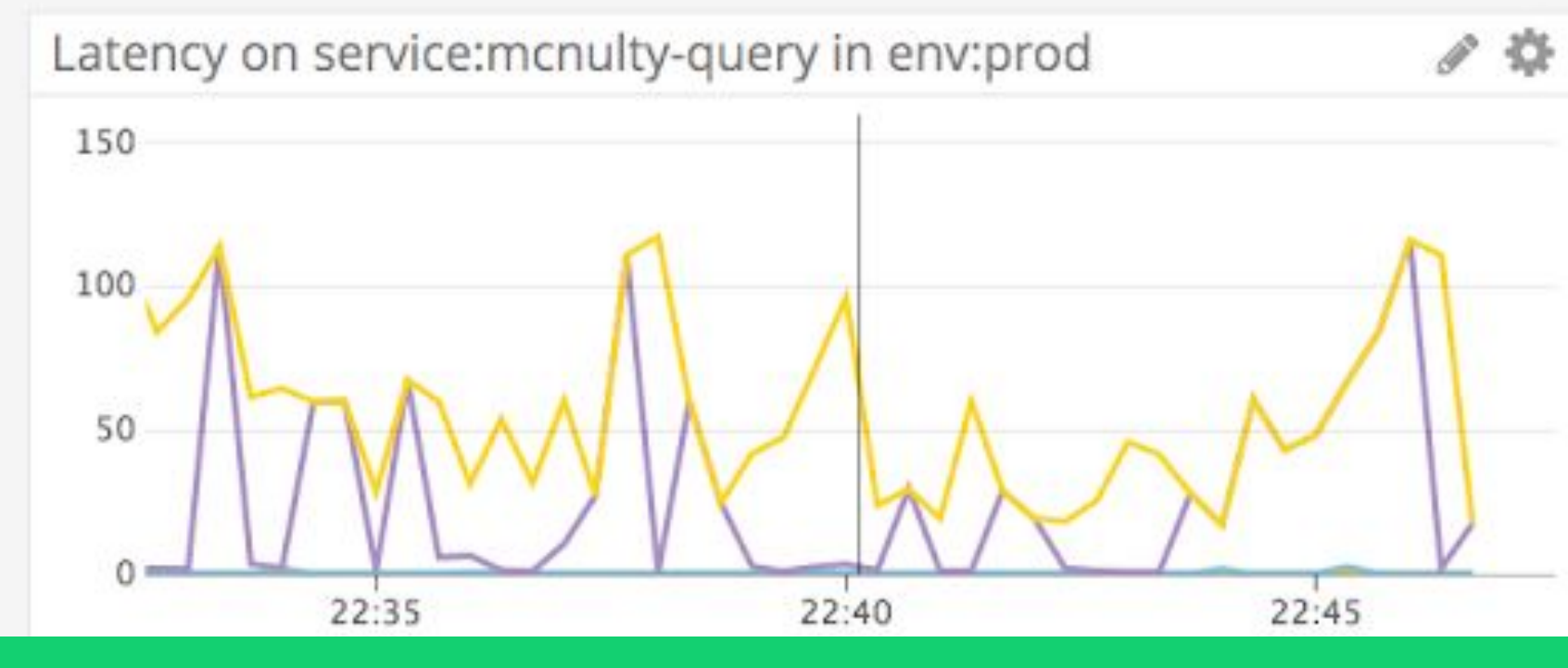
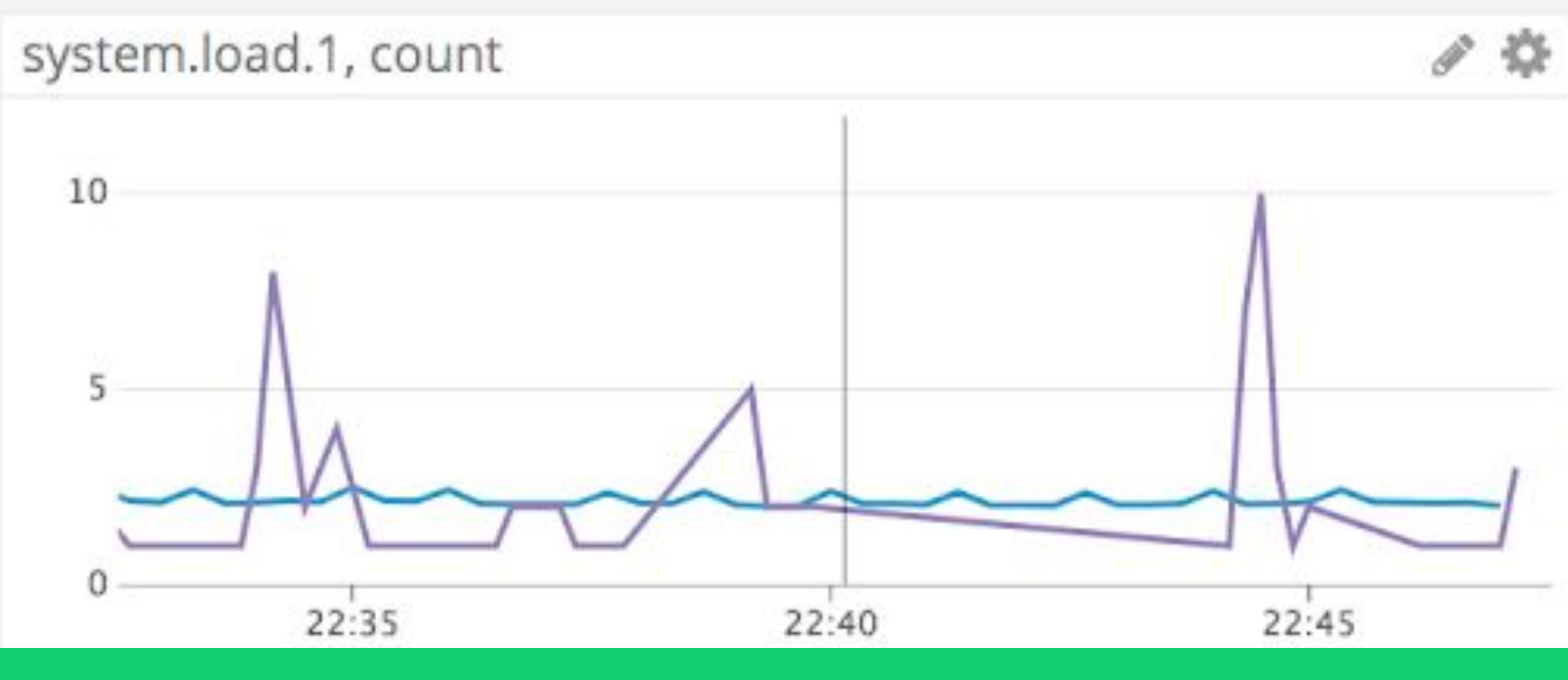
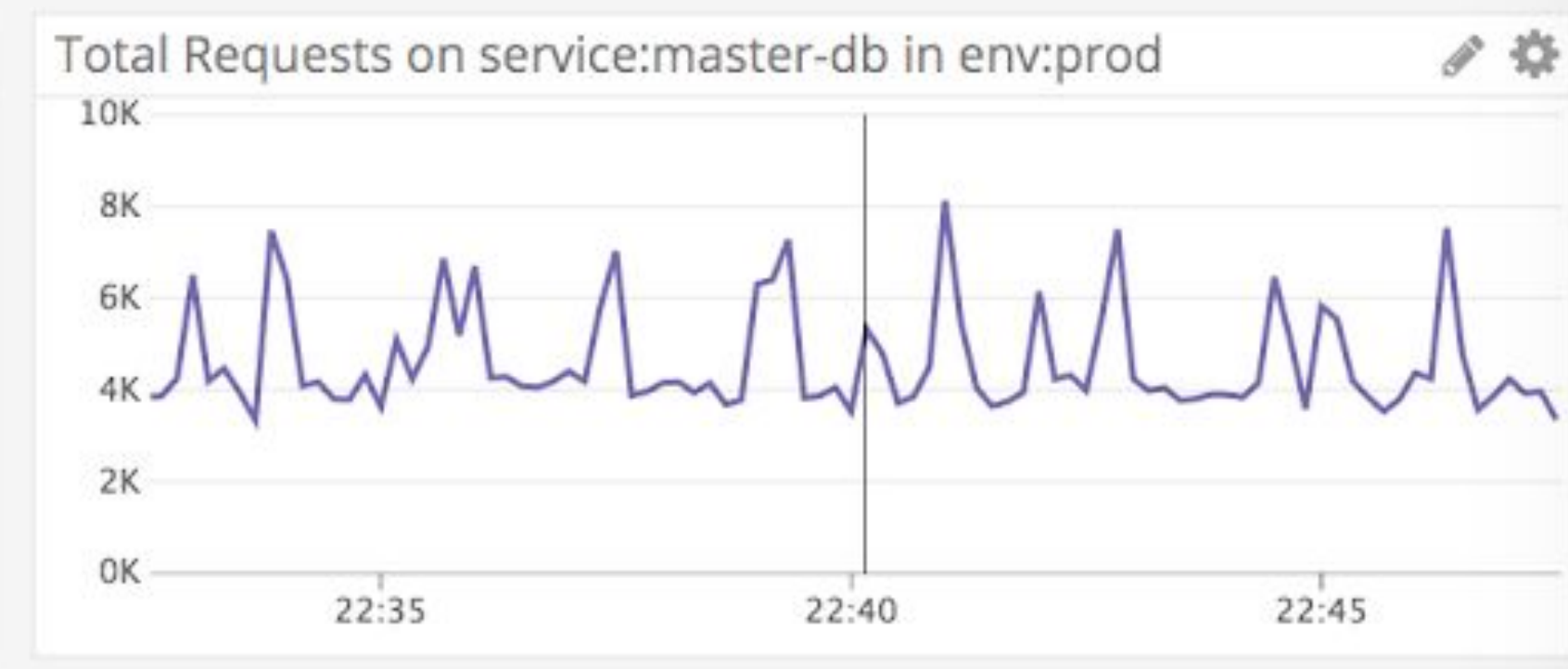
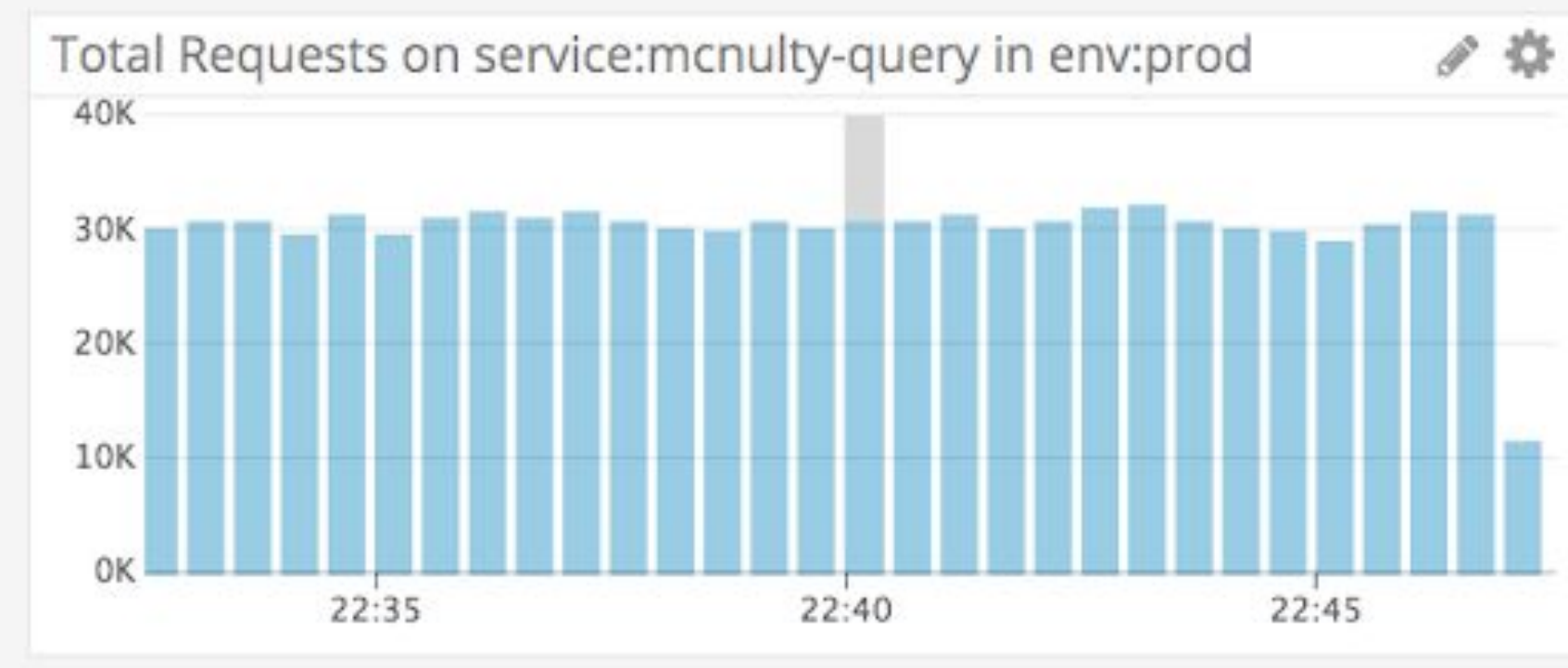
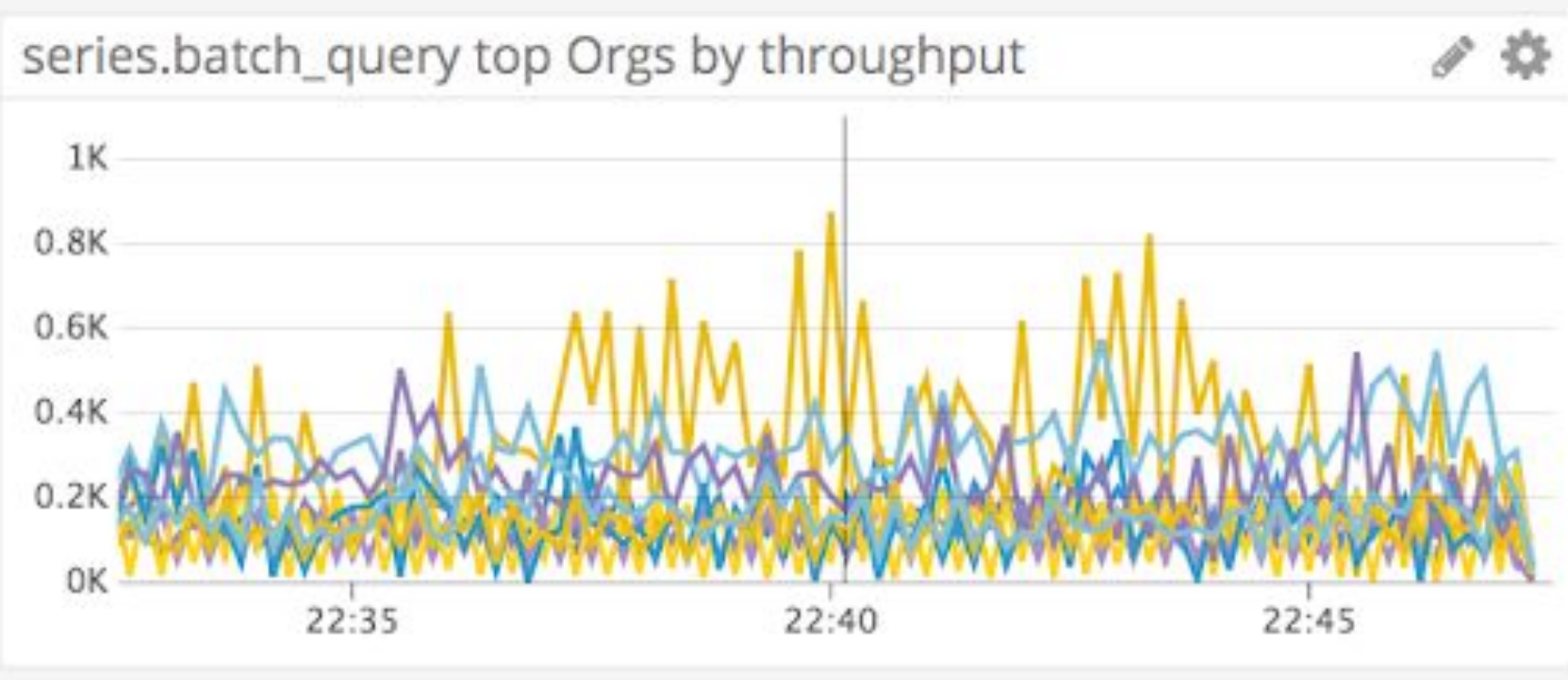
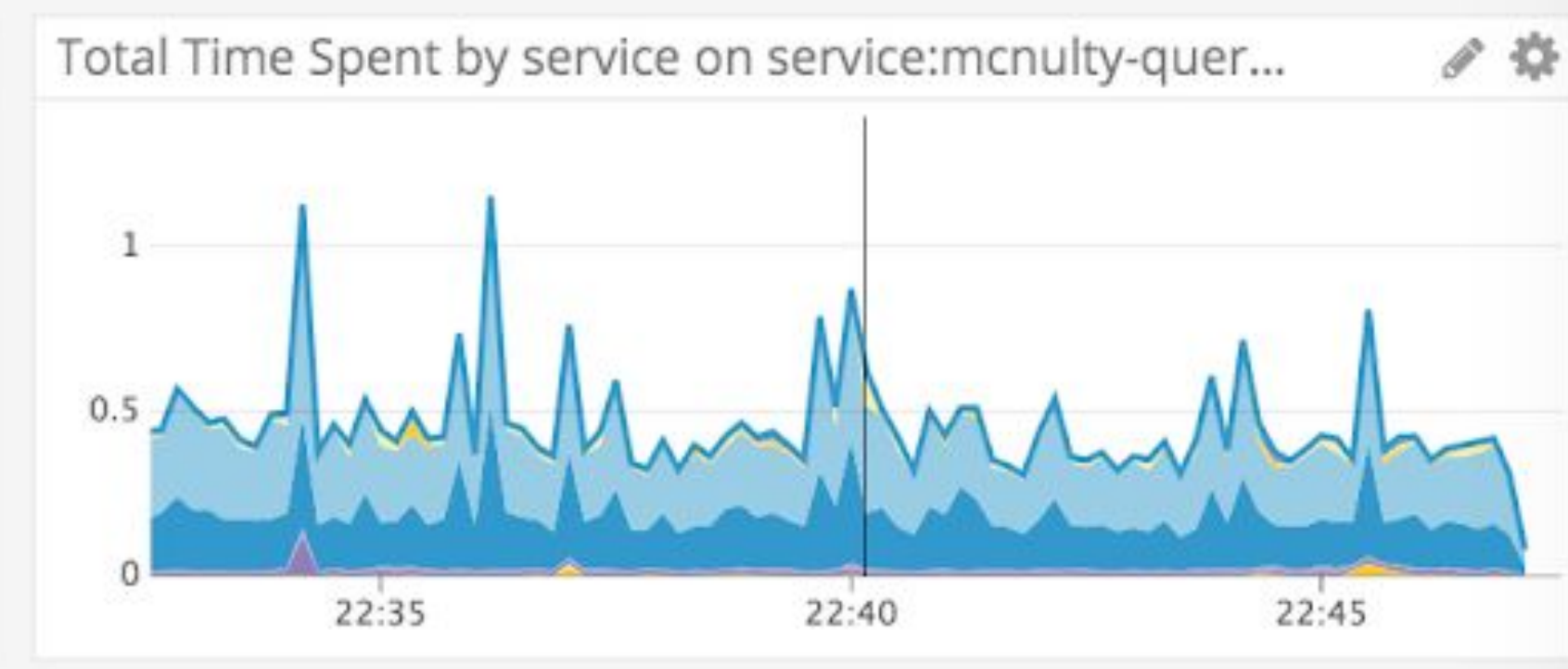
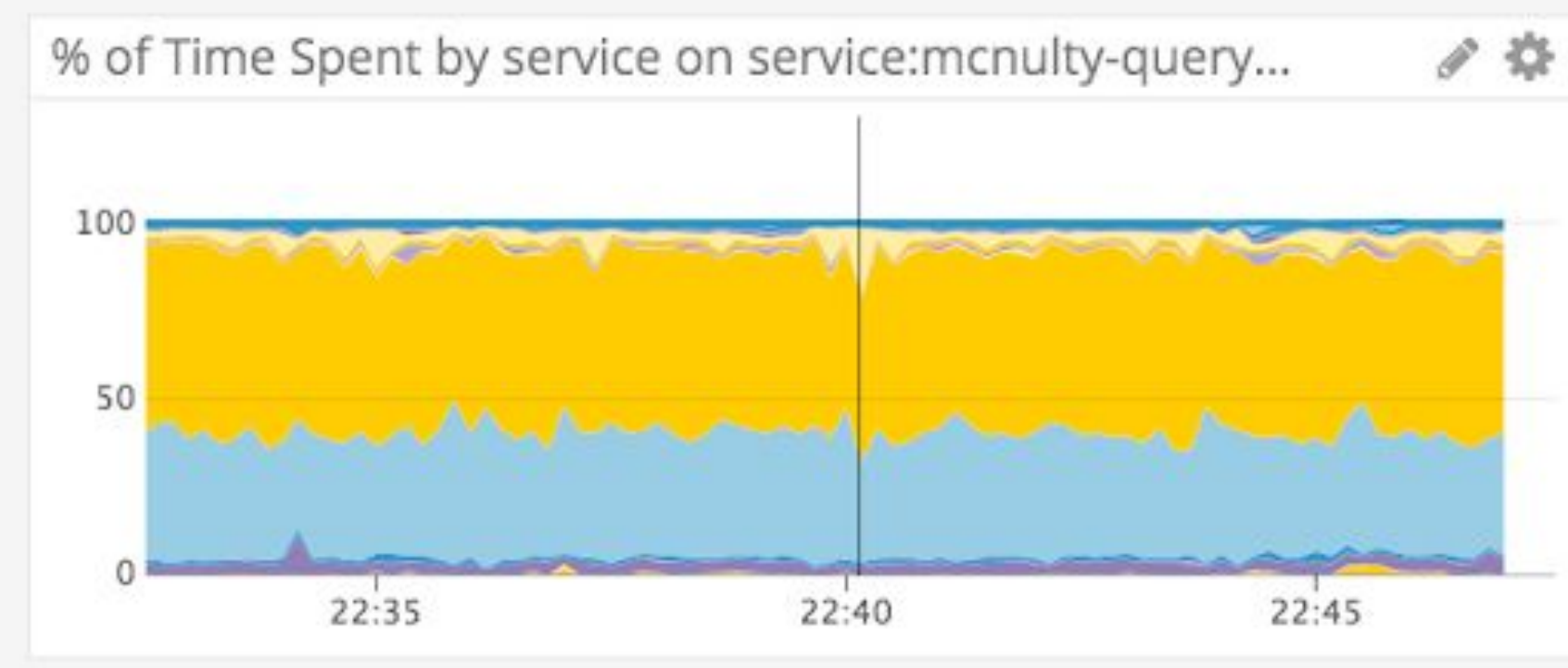
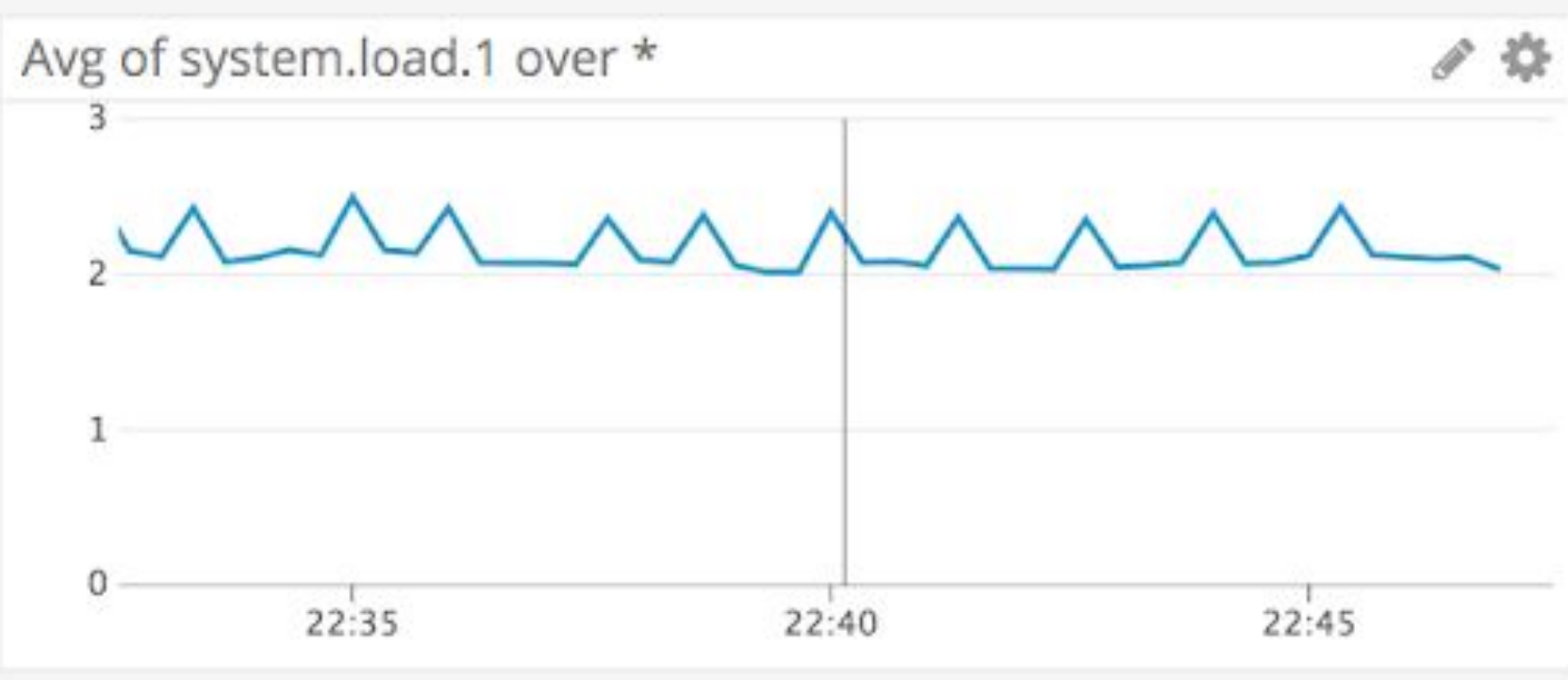
- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Image placeholder

Images are to be anchored to the right side of the slide and centered vertically.

Full-screen images should fill the page. If a border is needed, use matching section color. See next slide for example.



\$var * ▾ 🖋️ ?🔍 Search events to overlay...Show 15m The Past 15 Minutes ▾

Section divider 6

Large text can
be used to make
a single
important point.

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Standard slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Slide with two points

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Cras mauris purus, varius in justo, varius vulputate eros. Cras mauris purus, varius in justo in, varius vulputate eros.

Slide with three points

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

This row should be visually centered on the page.
Adjust vertical position of the entire row together depending on height of text.

Slide with four points

Identify and convince great people to join

Great = able + driven + grown-up

Hiring is what makes/breaks us

Expect to spend 10-20% of your time on hiring

Big Statement

1

Geometric
top-line growth

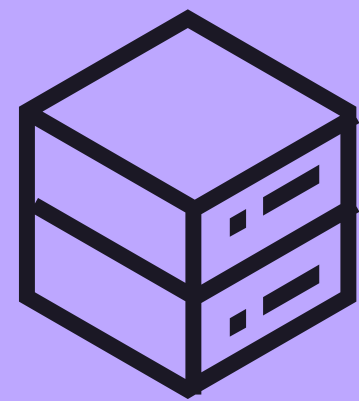
2

Strong
product-market fit

3

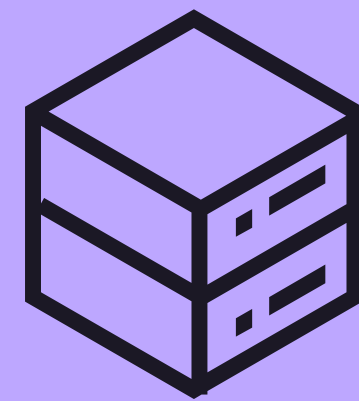
Financially efficient
go-to-market

Use iconography to support content



Cras mauris purus,
varius in justo in, varius
vulputate eros.

Secondary text, if needed



Curabitur pharetra
dictum mattis curabitur
dictum.

Secondary text, if needed



Proin iaculis nulla
turpis, a fermentum
orci mattis in.

Secondary text, if needed

The Pillars of Observability



Logs



Traces



Metrics

White background slide

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec aliquam odio iaculis eros fringilla ultricies. Donec placerat ac lectus nec pellentesque.

Text+image slide

- Supporting text can go here if needed.
- Avoid cramming slides full of text.
- Consider breaking dense content across several slides.

Image placeholder

Images are to be anchored to the right side of the slide and centered vertically.

Full-screen images should fill the page. If a border is needed, use matching section color. See next slide for example.

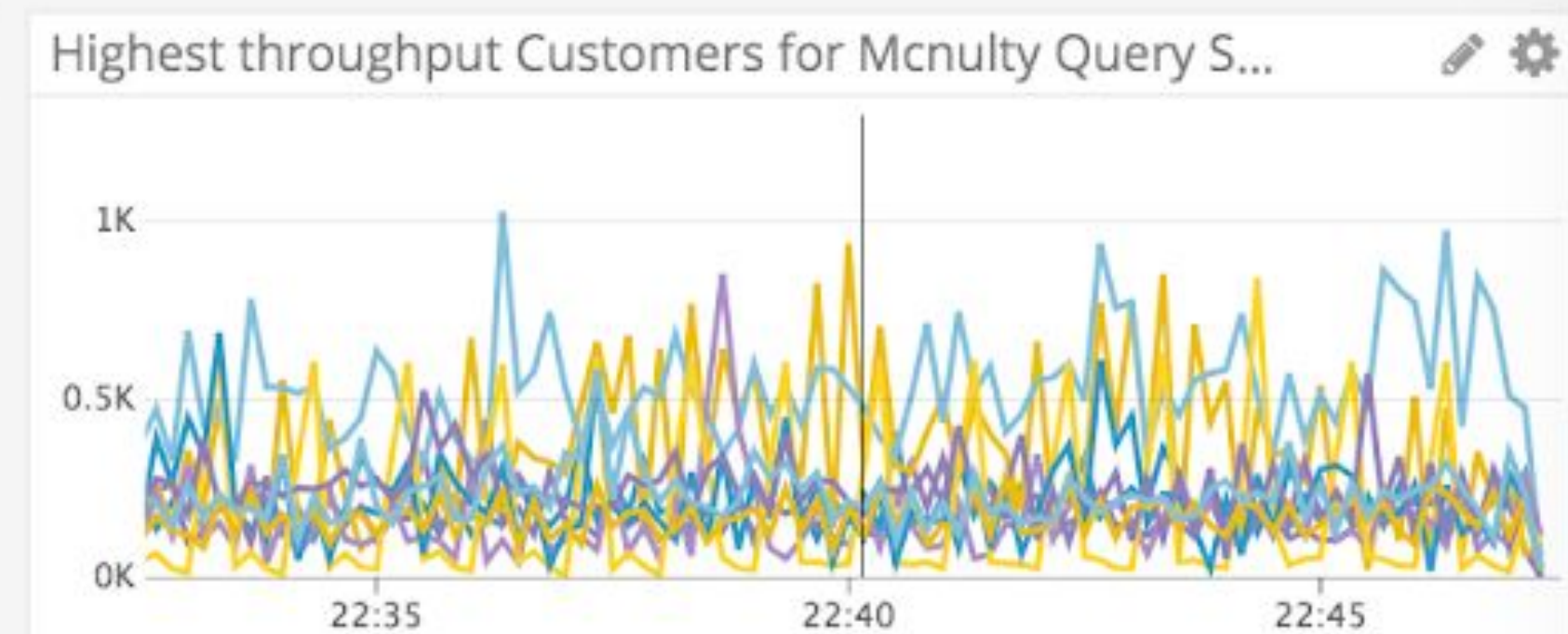
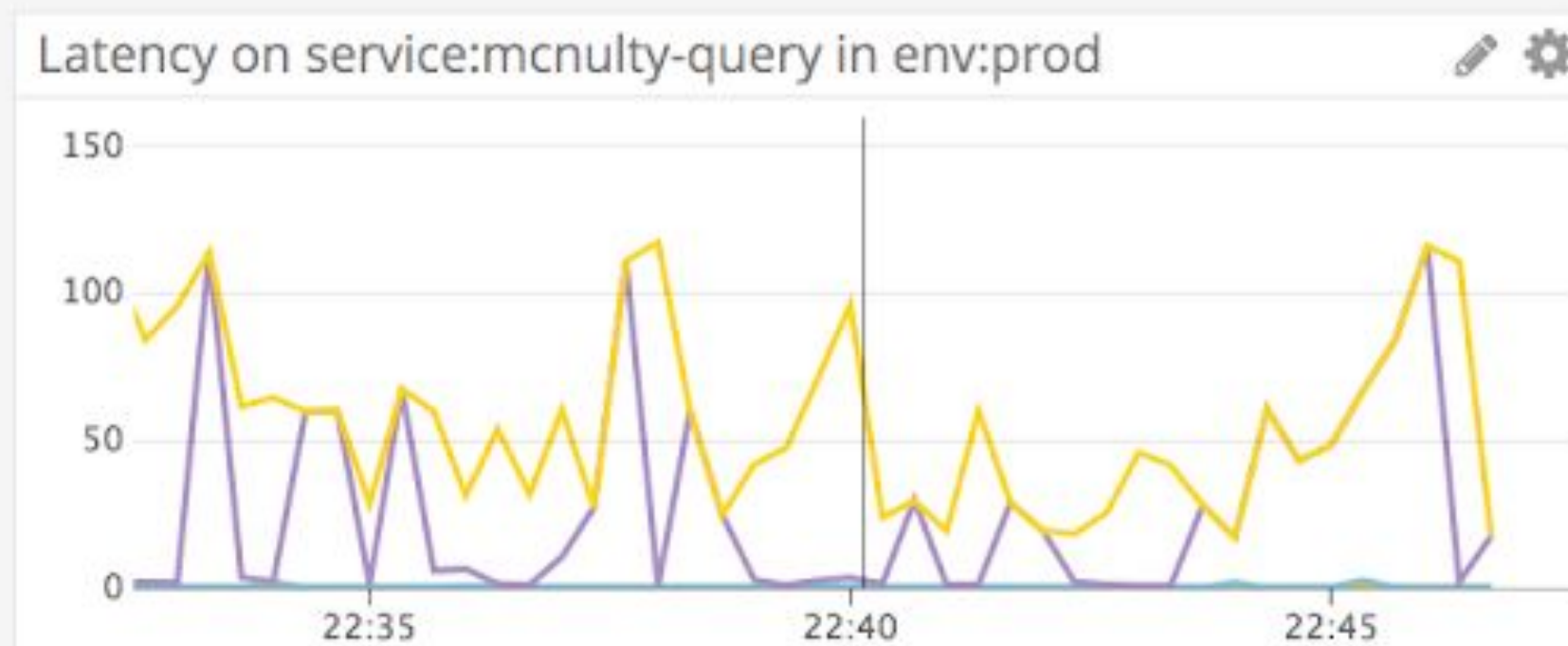
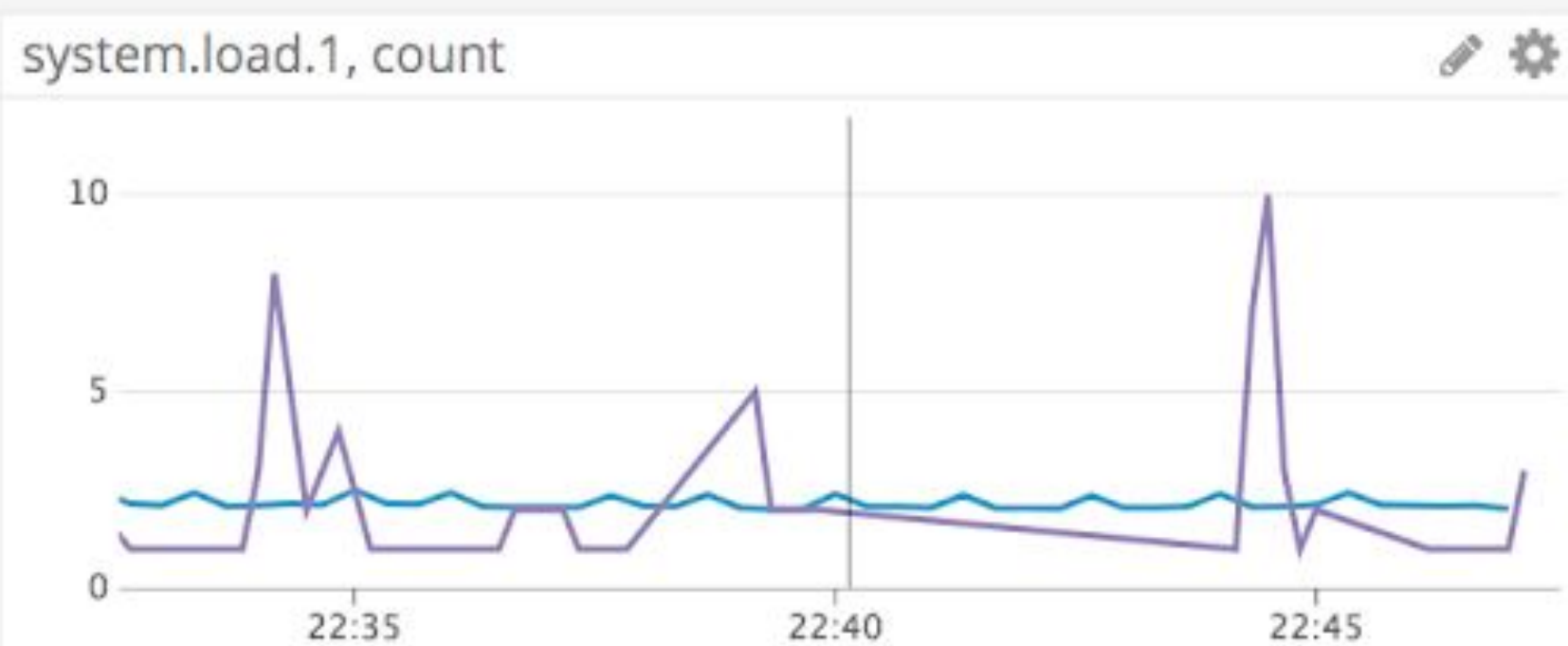
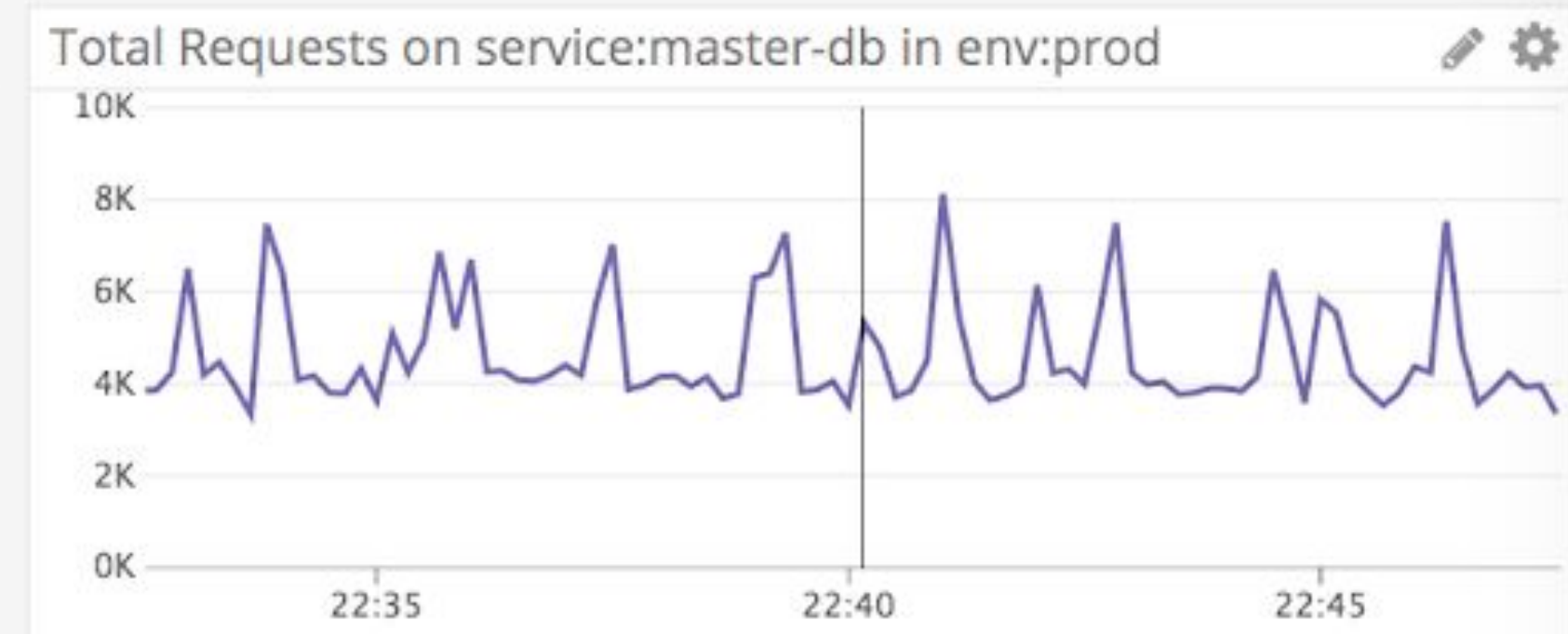
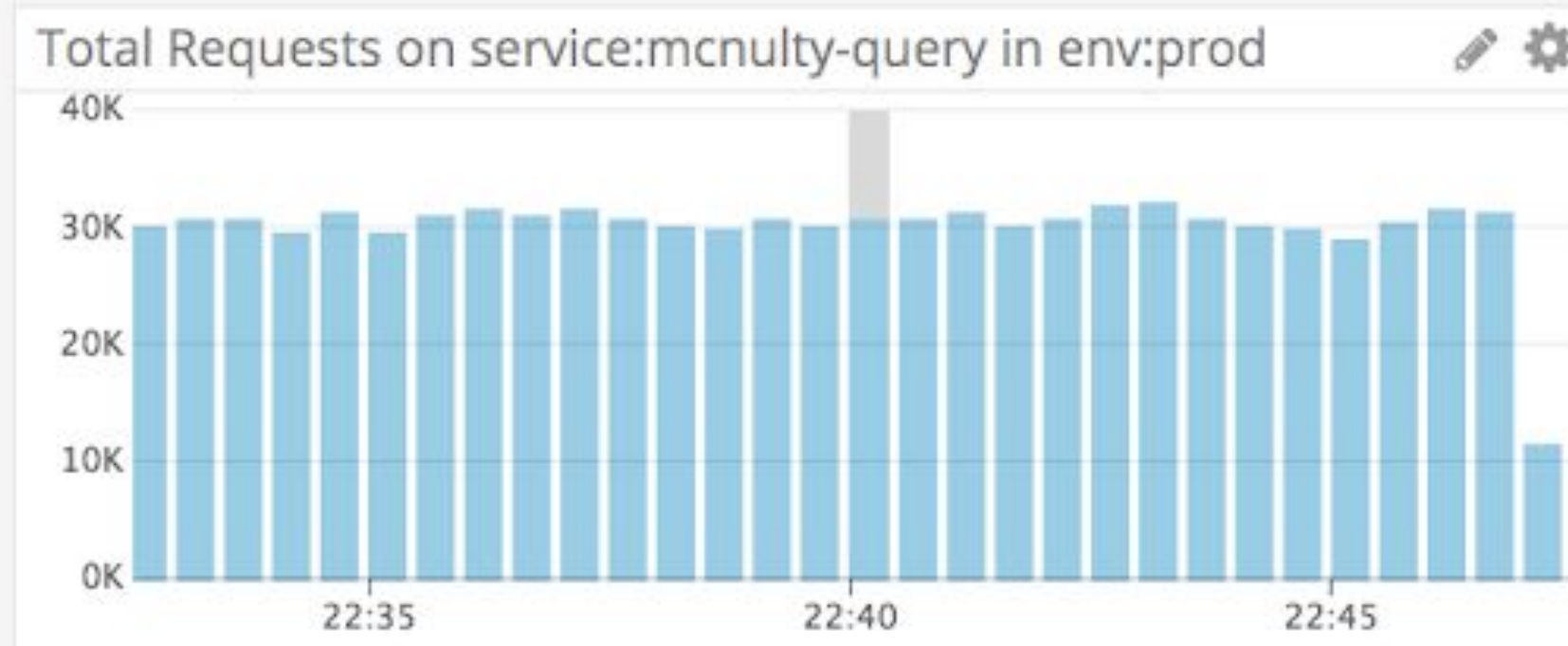
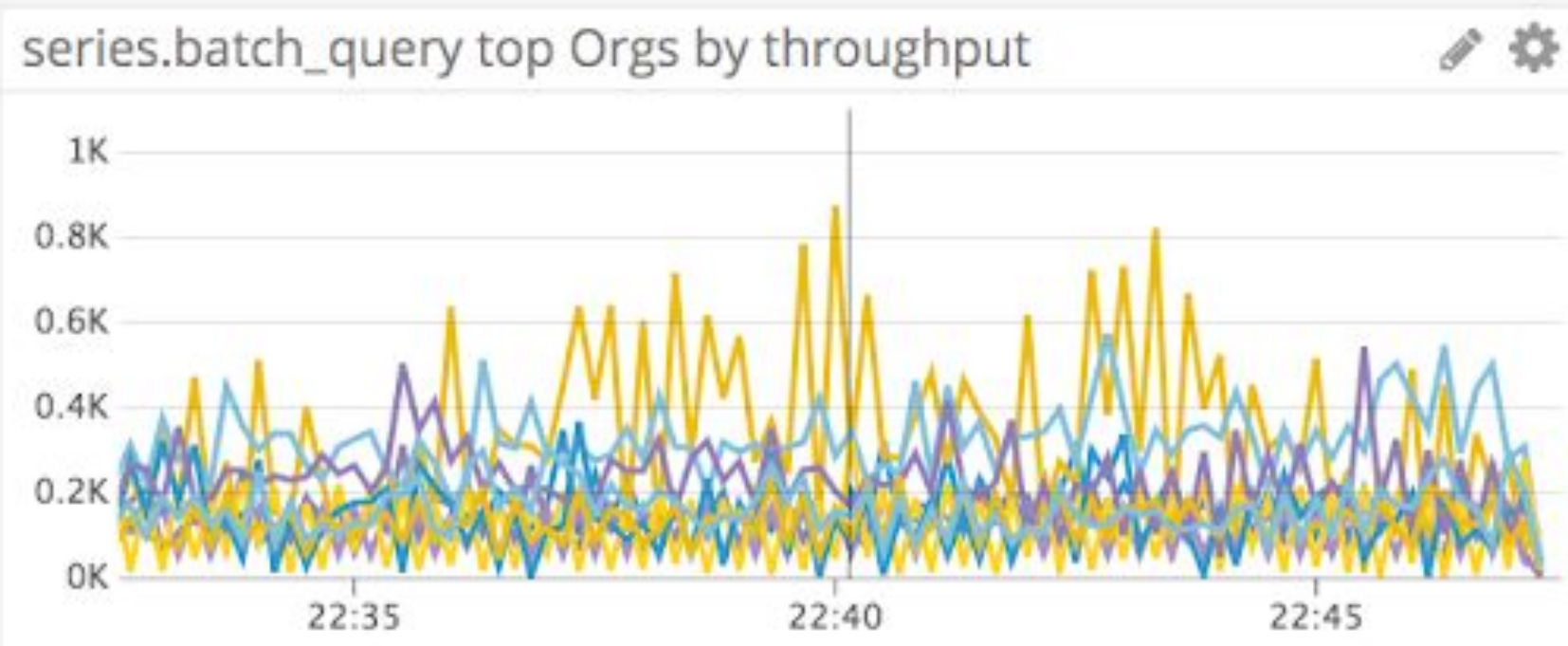
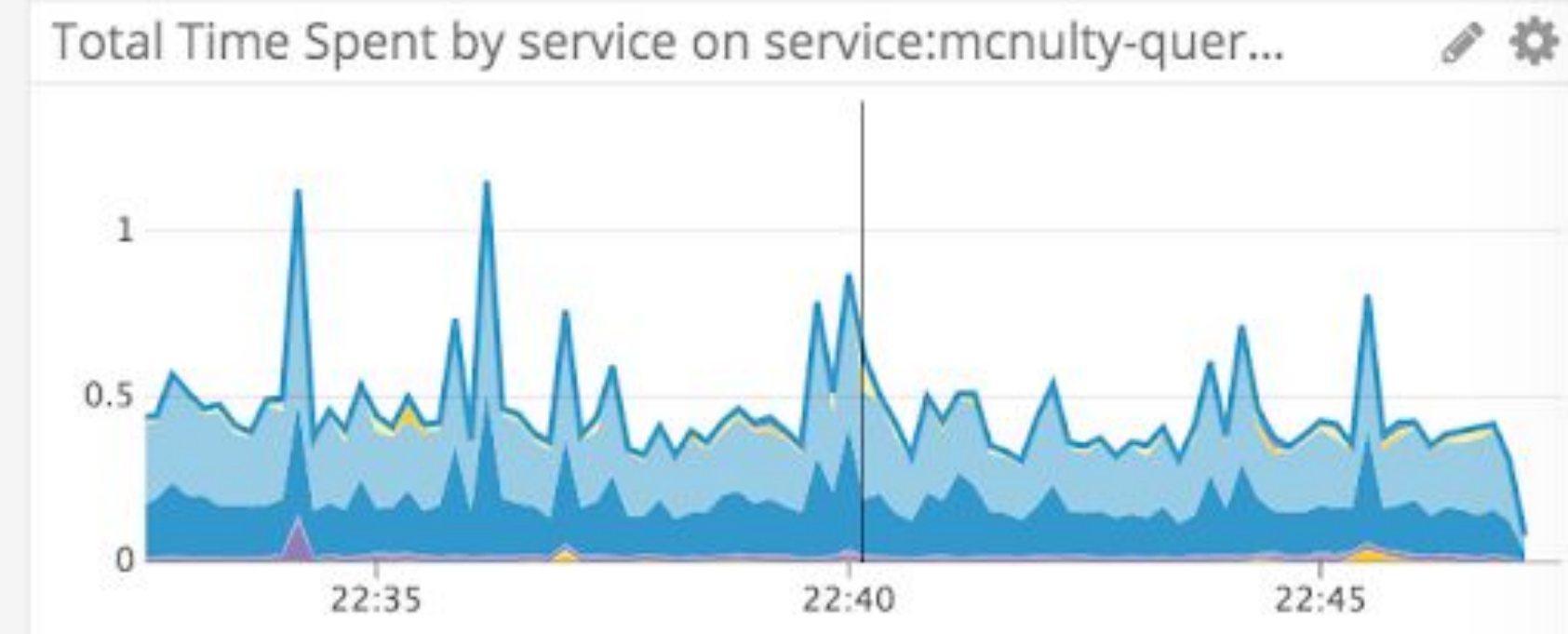
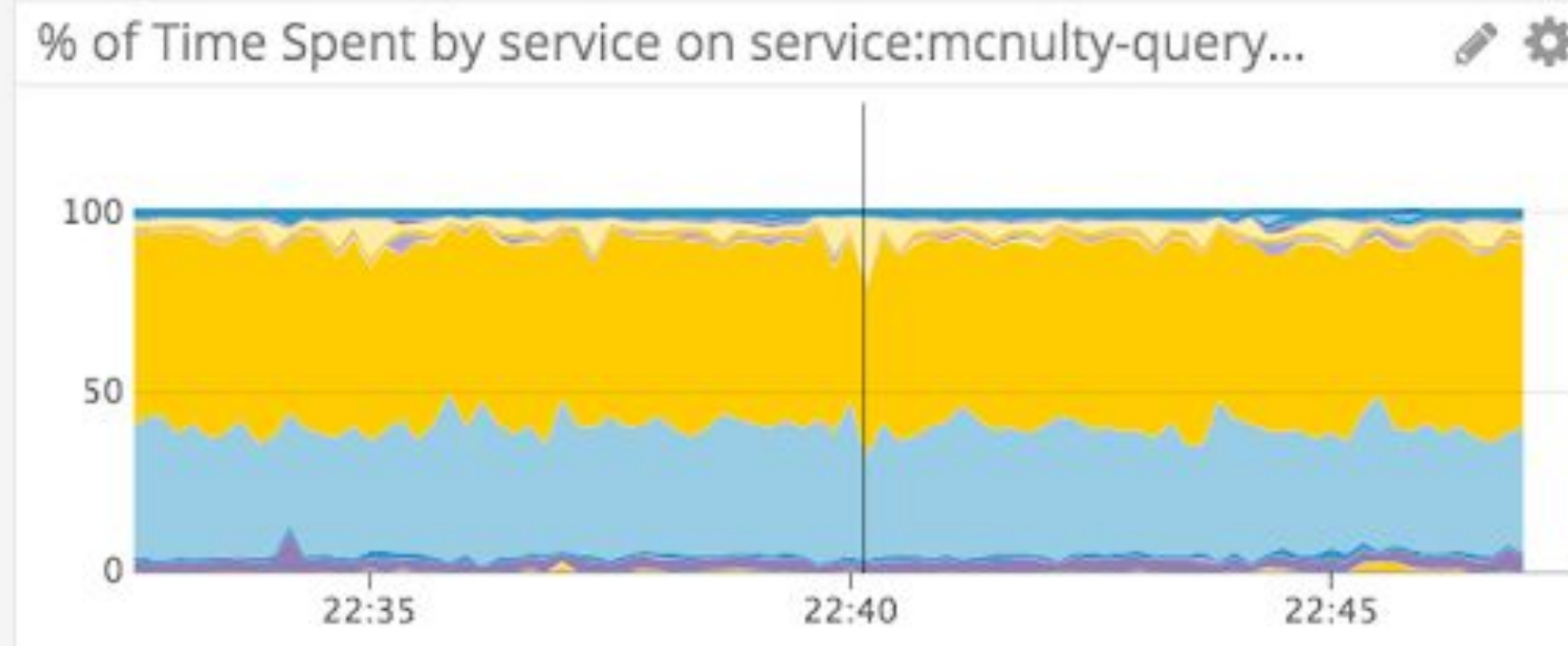
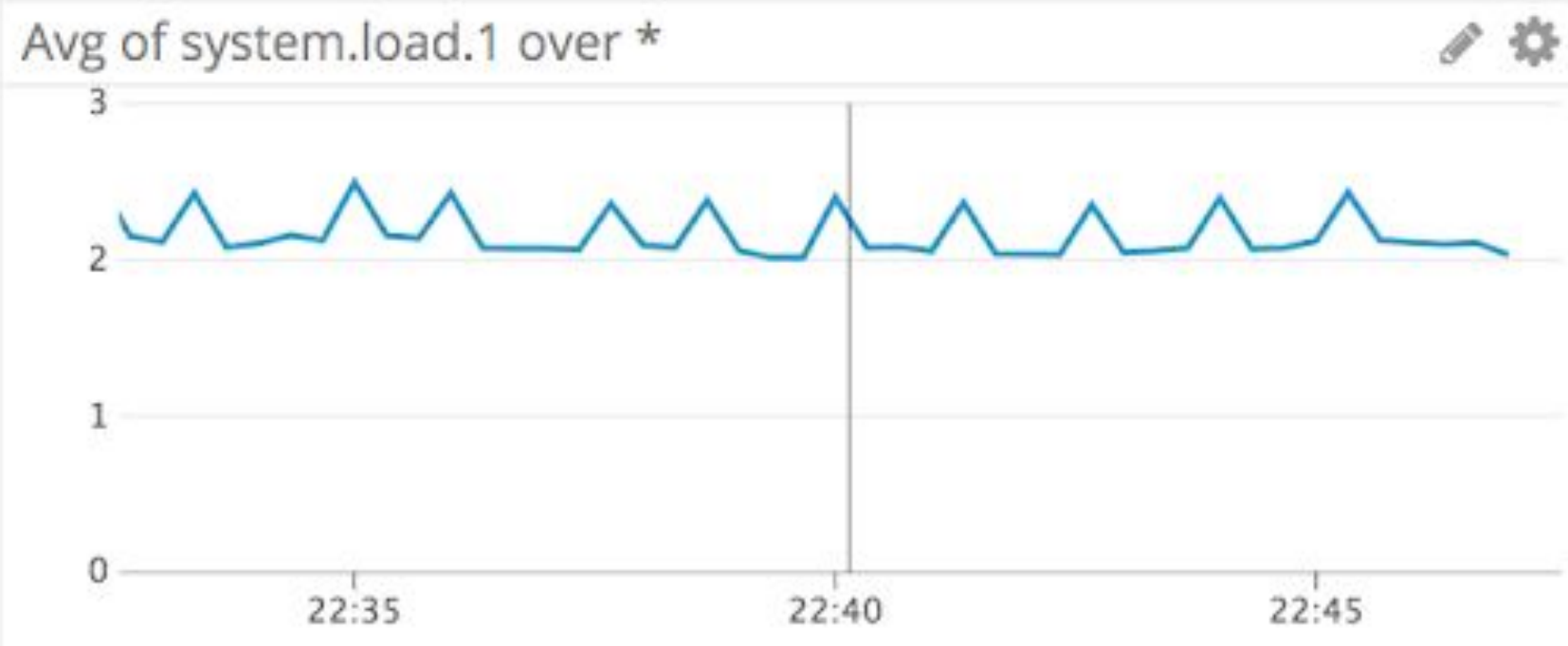


\$var * 📏 ?

🔍 Search events to overlay...



Show 15m The Past 15 Minutes



Thank you

