Wait for Us!

Evolving On-Call as Your Company Grows



usenix LISA17

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

Agenda

- About myself and Datadog
- Observations of the journey from startup to large company for on-call teams
- Tips and tools to ensure your on-call teams are not forgotten
- Review the takeaways



Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

About me - Chris Hoey

- Wireless Generation → Amplify (10y)
  - QA Lead
  - Linux Sysadmin
  - Senior IT Manager
- Mortar Data → Datadog (5y)
  - Director of Engineering, Ops
  - SRE
  - Director of SRE

Member of and managed on-call teams from small startup days through 800 person organizations

First LISA → LISA'08



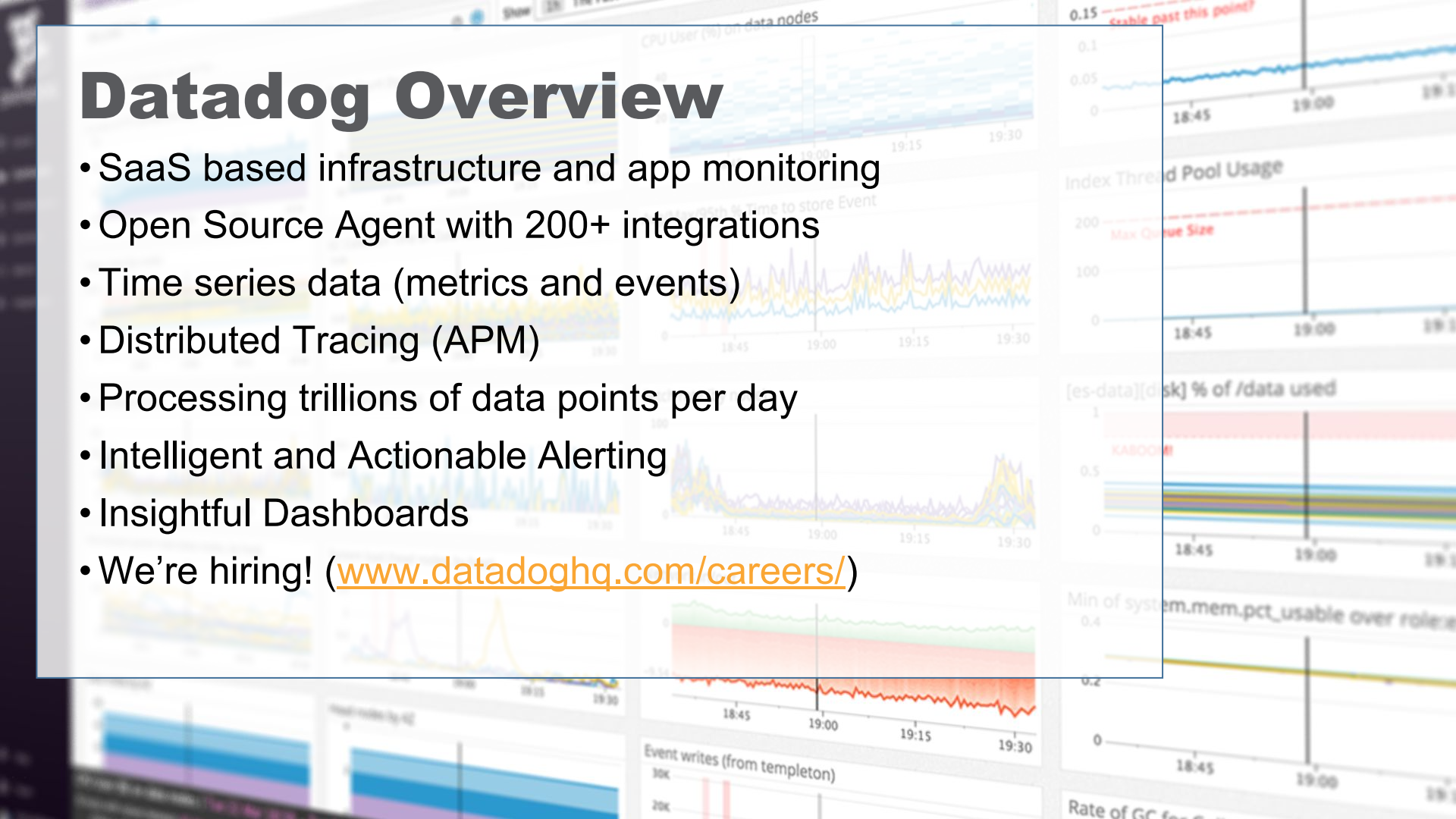
usenix LISA17

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

Datadog Overview

- SaaS based infrastructure and app monitoring
- Open Source Agent with 200+ integrations
- Time series data (metrics and events)
- Distributed Tracing (APM)
- Processing trillions of data points per day
- Intelligent and Actionable Alerting
- Insightful Dashboards
- We're hiring! (www.datadoghq.com/careers/)



The early startup years

- Pretty much everyone is on-call while wearing many hats
- Trivial for one human to reason about the entire system
- Little to no customers
- Product focus
  - Build, ship, repeat → get the MVP out asap!
- Security
  - what? 🙈 🙉 🙊
- Tech Debt
  - Do we even know what we are doing? Try all the things.



** generalizations not specific to any employer*

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

The growth startup years

- Directors and possibly founders on-call
- Still can reason about the entire system but getting harder
- Gaining trust from first customers
- Product focus
  - Ship the features, all of them 
- Security
  - maybe next sprint? 🙈 🙉 🙊
- Tech Debt
  - Those other shortcuts seemed to be ok so these new ones will do for now. When we get around to hiring more people that will make a first great ship for them. 😱



** generalizations not specific to any employer*

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

The hyper-growth years



- Team leads and individuals on-call, trying out dedicated SRE on-call
- Reasoning about the entire system takes significant effort
- Lots of customers, some very large demanding ones
- Product focus
  - new features/products
  - perf fixes and tech debt rewrites



- Security
  - The start of secure all the things!



- Tech Debt
  - That new tech looks like the new hotness, ehhh not sure how or when to fit it in. We will revisit that later.

** generalizations not specific to any employer*

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

The enterprise chasing years

- Core on-call is crushed, dedicated SRE and team based coverage for their respective services is increasing
- Nearly impossible to reason about the entire system as an individual
- Large number of customers, many adding you to their critical path
- Product focus
  - more new features/products
  - rolling acquisitions into the fold
- Security
  - compliance and audits ++++




- Tech debt
  - Greenfield rewrites, Performance Engineering is becoming a thing, cost savings a focus



** generalizations not specific to any employer*

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

But what about the on-call teams?

How are they doing?

What are they doing?



Christopher Hoey
Director SRE @ Datadog
mrchoey

Measure on-call pain

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1			total for hour	173	244	290	150	134	254	115	386	572	257	262	418	357	461	336	359	1467	4208	2518	482	326	221	548	282		
2		total for day		0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00		hours without high priority alert
3	1/1/2017	25	Sunday	1	4	3	2	0	1	0	2	3	4	2	0	1	0	0	0	0	0	0	0	0	1	0	1		12
4	1/2/2017	21	Monday	0	1	1	0	0	0	0	0	1	0	8	1	0	3	0	0	0	1	0	1	0	1	2	1		13
5	1/3/2017	48	Tuesday	1	0	2	1	2	0	2	1	2	1	6	2	0	8	7	4	3	1	0	3	1	0	1	0		6
6	1/4/2017	136	Wednesday	4	29	1	4	4	0	1	0	1	23	0	5	1	5	19	2	16	6	2	4	2	1	4	2		3
7	1/5/2017	73	Thusday	1	0	1	4	1	0	1	3	2	2	1	2	5	3	2	28	3	2	3	1	3	1	3	1		2
8	1/6/2017	1893	Friday	0	1	1	1	3	0	0	4	0	2	2	4	1	4	6	1	613	918	250	7	7	54	10	4		4
9	1/7/2017	56	Saturday	2	6	2	1	1	1	2	3	0	2	1	0	1	8	3	2	2	2	1	11	1	1	2	1		2
10	1/8/2017	38	Sunday	2	4	2	3	1	1	1	0	1	1	1	1	4	1	1	1	1	2	2	2	2	2	0	2		2
11	1/9/2017	1676	Monday	3	2	4	2	0	4	1	3	2	5	13	7	7	10	12	7	52	54	606	245	133	9	367	128		1
12	1/10/2017	4101	Tuesday	53	52	3	2	4	9	6	4	142	7	7	138	6	4	9	54	219	2926	423	4	4	22	2	1		0
13	1/11/2017	748	Wednesday	2	2	1	0	2	120	7	268	223	49	4	10	4	9	4	3	4	10	7	9	6	1	3	0		2
14	1/12/2017	240	Thusday	2	3	8	2	2	3	3	3	11	3	9	3	8	15	13	17	11	58	52	6	2	4	1	1		0
15	1/13/2017	403	Friday	1	4	1	4	3	2	1	1	1	3	4	5	5	12	12	7	252	57	6	2	3	11	2	4		0
16	1/14/2017	62	Saturday	4	5	6	6	3	2	5	9	2	4	1	2	8	0	2	0	0	1	0	1	0	0	1	0		7
17	1/15/2017	46	Sunday	0	3	4	3	4	0	3	2	2	1	2	0	3	1	1	1	3	2	3	2	1	1	2	2		3
18	1/16/2017	79	Monday	2	3	2	6	2	0	1	0	6	2	6	4	3	0	3	1	4	1	0	3	16	1	11	2		4
19	1/17/2017	165	Tuesday	0	2	4	4	6	3	4	8	0	0	7	15	13	9	23	16	10	12	10	5	3	1	8	2		3
20	1/18/2017	1082	Wednesday	0	8	4	7	0	1	5	0	6	0	2	1	4	4	10	7	24	6	966	14	4	7	1	1		4
21	1/19/2017	288	Thusday	3	4	3	1	4	2	4	3	35	7	9	10	61	26	25	20	12	8	8	11	15	8	4	5		0
22	1/20/2017	262	Friday	4	5	4	6	4	5	2	1	5	13	4	17	16	137	4	7	5	3	5	0	1	1	0	13		2
23	1/21/2017	79	Saturday	5	8	2	2	1	1	2	1	3	3	2	3	5	3	6	3	3	2	3	4	1	2	10	4		0
24	1/22/2017	58	Sunday	5	2	3	3	2	3	1	3	3	2	4	3	2	1	5	3	1	1	1	2	1	2	2	3		0
25	1/23/2017	94	Monday	2	1	3	4	3	1	2	2	2	3	4	18	4	3	9	4	9	4	2	2	4	1	4	3		0
26	1/24/2017	173	Tuesday	1	1	1	4	2	0	1	2	8	11	17	19	13	12	9	19	17	13	6	3	6	2	3	3		1

usenix LISA17

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

Find alert patterns - volume of alerts that resolve within 60 seconds

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1				total for hour	34	65	20	22	13	10	3	5	2	43	36	65	27	40	29	42	39	87	132	18	22	15	23	18		
2		total for day	% of total resolved in 60s		0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00		hours without a high priority alert that resolves in under 60 seconds
3	1/1/2017	2	8.00%	Sunday	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		22
4	1/2/2017	1	4.76%	Monday	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		23
5	1/3/2017	7	14.58%	Tuesday	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	2	0	0	0	1	0	0	0	0		18
6	1/4/2017	17	12.50%	Wednesday	0	9	0	1	0	0	0	0	0	0	0	0	0	1	5	0	0	1	0	0	0	0	0	0		19
7	1/5/2017	11	15.07%	Thusday	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	5	3	1	0	0	0	0	0	0		19
8	1/6/2017	35	1.85%	Friday	0	0	1	0	0	0	0	0	0	0	2	0	0	1	0	0	1	25	4	1	0	0	0	0		17
9	1/7/2017	2	3.57%	Saturday	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		22
10	1/8/2017	1	2.63%	Sunday	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		23
11	1/9/2017	80	4.77%	Monday	0	0	0	0	0	0	0	0	0	0	0	1	0	3	1	1	0	6	56	0	3	0	5	4		15
12	1/10/2017	204	4.97%	Tuesday	25	44	1	0	3	2	0	0	0	1	4	55	5	4	5	21	1	3	28	0	2	0	0	0		8
13	1/11/2017	39	5.21%	Wednesday	0	0	0	0	1	0	2	0	0	28	0	0	0	4	0	0	2	1	0	1	0	0	0	0		17
14	1/12/2017	43	17.92%	Thusday	0	0	1	1	1	0	0	0	0	0	4	0	2	1	1	3	0	24	5	0	0	0	0	0		14
15	1/13/2017	29	7.20%	Friday	0	1	0	1	0	0	0	0	0	0	1	0	2	4	0	0	2	16	1	0	0	1	0	0		15
16	1/14/2017	10	16.13%	Saturday	1	1	1	0	0	0	0	0	1	2	0	2	2	0	0	0	0	0	0	0	0	0	0	0		17
17	1/15/2017	5	10.87%	Sunday	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1		19
18	1/16/2017	18	22.78%	Monday	0	2	0	1	1	0	0	0	0	0	1	1	0	0	1	0	1	1	0	1	2	0	5	1		12
19	1/17/2017	18	10.91%	Tuesday	0	0	0	0	0	0	0	0	0	0	1	1	0	3	4	1	1	4	1	0	1	0	0	1		14
20	1/18/2017	45	4.16%	Wednesday	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	2	9	0	25	5	0	0	0	0		18
21	1/19/2017	6	2.08%	Thusday	0	0	1	0	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	1	0	1	0	0		19
22	1/20/2017	14	5.34%	Friday	0	0	1	0	0	1	0	0	0	8	1	0	0	2	0	0	0	0	0	0	0	0	0	1		18
23	1/21/2017	3	3.80%	Saturday	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0		21
24	1/22/2017	3	5.17%	Sunday	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0		21
25	1/23/2017	5	5.32%	Monday	0	0	1	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0		19

usenix LISA17

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

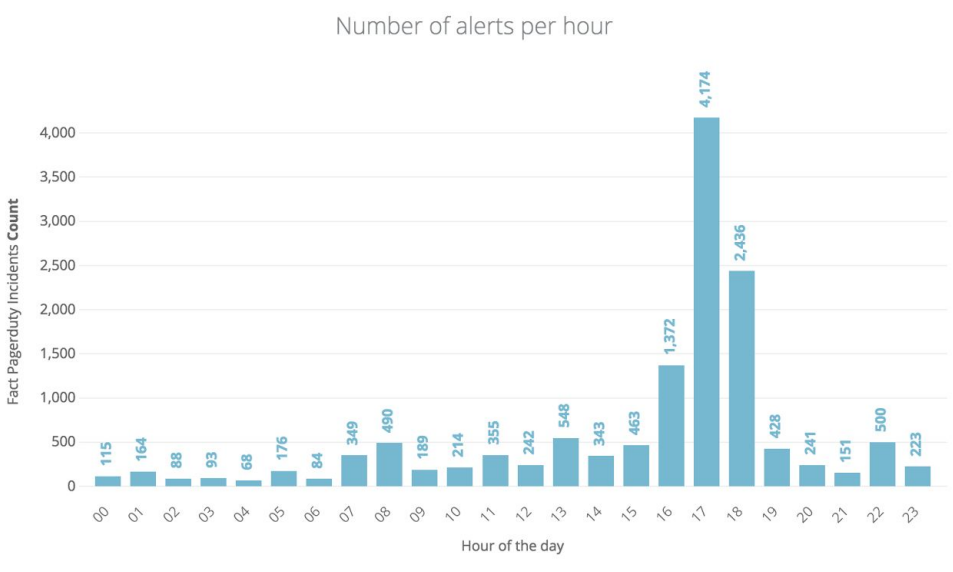
Find alert patterns - volume of alerts that resolve within 300 seconds

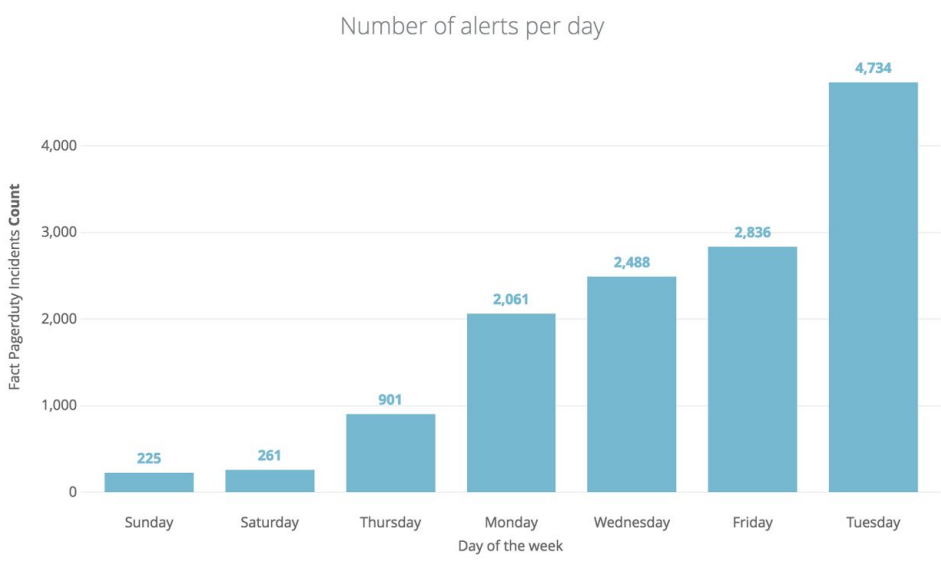
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1				total for hour	83	136	61	63	53	154	21	277	377	118	92	217	77	118	82	137	715	523	764	306	180	58	422	179		
2		total for day	% of total resolved in 300s		0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00		hours without a high priority alert that resolves in under 300 seconds
3	1/1/2017	6	24.00%	Sunday	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1		18
4	1/2/2017	2	9.52%	Monday	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0		22
5	1/3/2017	23	47.92%	Tuesday	1	0	2	0	0	0	0	0	0	1	2	2	0	2	4	3	2	1	0	1	1	0	1	0		11
6	1/4/2017	57	41.91%	Wednesday	0	12	1	2	2	0	0	0	0	23	0	0	0	1	7	0	2	3	1	1	0	0	2	0		12
7	1/5/2017	39	53.42%	Thusday	0	0	1	2	0	0	0	0	0	1	0	0	0	2	0	28	3	1	0	0	0	0	1	0		16
8	1/6/2017	997	52.67%	Friday	0	1	1	1	2	0	0	0	0	0	2	2	0	2	2	0	506	318	144	2	0	12	2	0		10
9	1/7/2017	13	23.21%	Saturday	2	2	2	1	1	0	1	1	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0		15
10	1/8/2017	7	18.42%	Sunday	1	2	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		19
11	1/9/2017	1221	72.85%	Monday	0	0	1	0	0	1	0	0	0	1	2	2	4	4	1	3	51	52	234	245	130	5	359	126		7
12	1/10/2017	670	16.34%	Tuesday	52	49	1	0	4	7	1	0	125	2	4	129	5	4	5	45	1	25	205	2	2	2	0	0		4
13	1/11/2017	683	91.31%	Wednesday	0	1	0	0	1	119	5	256	221	46	2	5	3	6	1	1	2	4	4	5	1	0	0	0		6
14	1/12/2017	145	60.42%	Thusday	0	2	5	1	2	0	1	0	2	0	6	1	3	4	6	9	2	49	50	2	0	0	0	0		8
15	1/13/2017	110	27.30%	Friday	0	2	1	2	0	0	0	0	0	2	2	1	3	7	1	2	47	29	2	1	0	3	2	3		7
16	1/14/2017	34	54.84%	Saturday	3	2	5	3	1	0	2	5	1	3	1	2	3	0	1	0	0	1	0	0	0	0	1	0		9
17	1/15/2017	11	23.91%	Sunday	0	1	3	2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1		16
18	1/16/2017	41	51.90%	Monday	2	3	2	4	1	0	0	0	0	1	4	2	0	0	1	0	1	1	0	2	4	0	11	2		9
19	1/17/2017	54	32.73%	Tuesday	0	1	2	3	4	0	0	3	0	0	1	6	3	4	7	5	2	6	3	0	1	0	1	2		7
20	1/18/2017	132	12.20%	Wednesday	0	5	3	6	0	1	0	0	1	0	1	1	2	0	1	2	18	3	75	12	0	1	0	0		9
21	1/19/2017	70	24.31%	Thusday	1	1	1	0	1	0	1	0	11	3	0	1	3	17	4	4	3	2	1	5	7	2	0	2		5
22	1/20/2017	56	21.37%	Friday	0	2	1	1	1	1	0	0	0	8	1	7	1	20	0	1	1	0	1	0	0	0	0	10		10
23	1/21/2017	13	16.46%	Saturday	3	4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	1		17
24	1/22/2017	9	15.52%	Sunday	2	0	2	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	2		18
25	1/23/2017	19	20.21%	Monday	0	0	1	2	0	0	0	0	0	0	1	8	1	0	0	2	2	0	0	0	0	0	0	2		16

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

Measure, monitor and triage alert trends



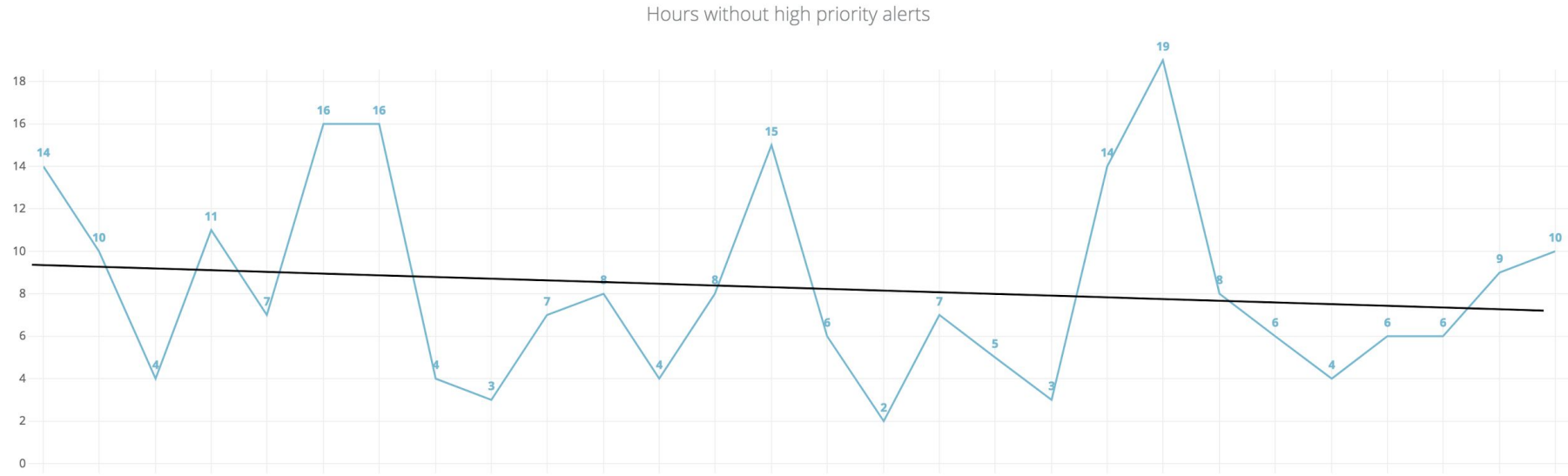


usenix LISA17

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

Measure, monitor and triage alert trends



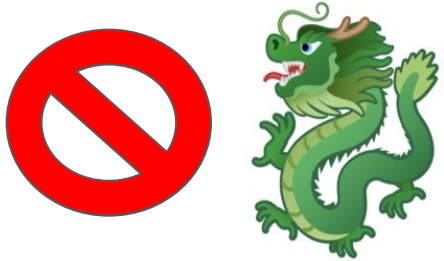
Christopher Hoey
Director SRE @ Datadog
mrchoey

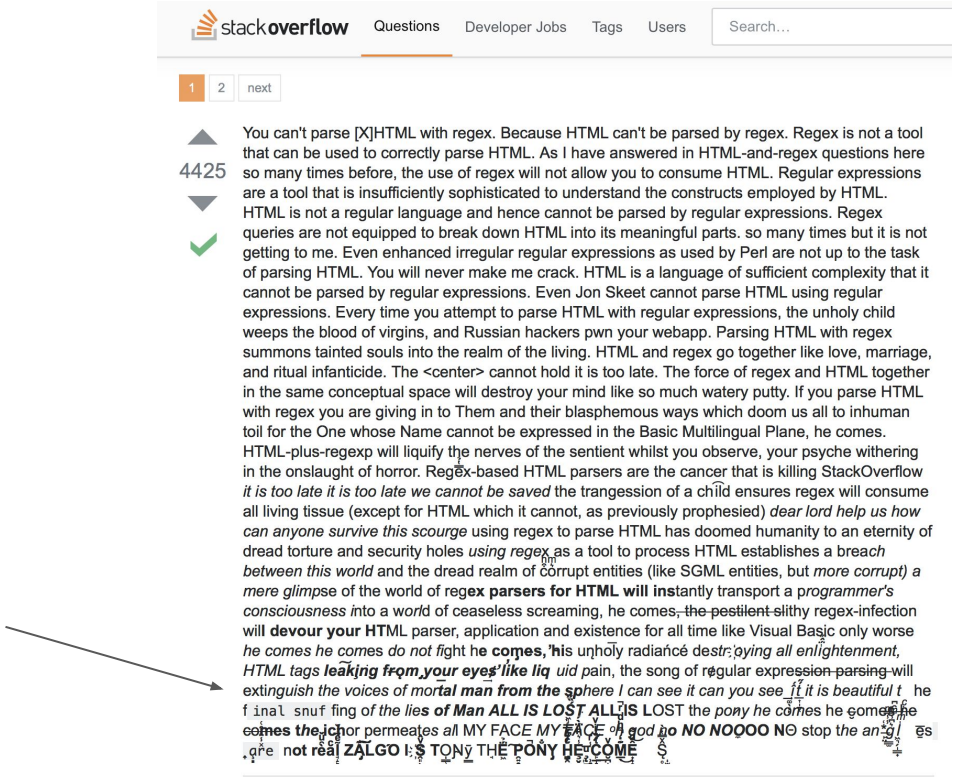
usenix LISA17
DATADOG

Break out your monitors by service

Use a naming convention upfront

Avoid the “Just use a regex on it…” trap





Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

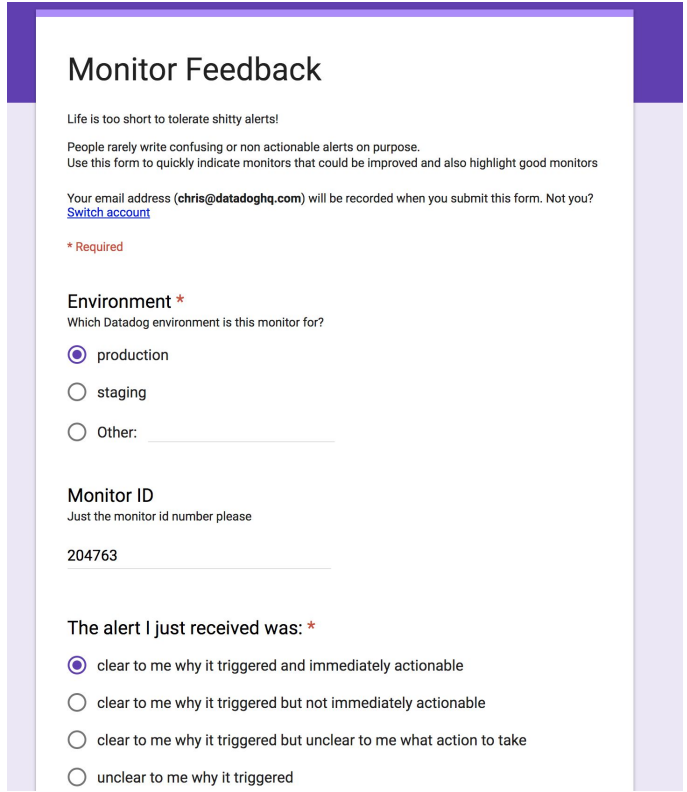
Build monitor feedback loops

In the monitor notification provide a way to give feedback

Monitor Feedback:

This alert was:

- clear to me why it triggered and immediately actionable
- clear to me why it triggered but not immediately actionable
- clear to me why it triggered but unclear to me what action to take
- unclear to me why it triggered **@slack-bourbon-ops**



https://www.slideshare.net/CoryWatson8/building-a-culture-of-observability-at-stripe

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG



We are you putting you into the on-call rotation.

It will be fine…..





Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG



We are preparing you to go into the on-call rotation

Here are some safeties we have in place

Here is how we do shadow ops

Here is how you get help

Lets run some game days together



https://www.usenix.org/conference/srecon15/program/presentation/widdowson

usenix LISA17

Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

Document all the things! -- Runbooks + Checklists + Tech Docs

Runbooks - quick overview of current state of a service as markdown files in a dedicated git repo

- Markdown is easy enough, offline access is nice
- Current work in progress issues can be added as Github Issues on the runbook repo
- Easy to view history of changes
- Can build tools to show what changed since last time a person was on-call

Checklists - the commands and steps to be taken in a specific situation as part of a monitor notification

- Have what to do and where to look as part of the alert
- Do you really want to be searching through wikis at 3am

Techdocs - Google Docs that capture the historical discussion behind a service

- Gives new hires the chance to get some background on why service x is built the way it us or why it scales the way it does
- A chance to in line comment and question sections for a living discussion

http://www.brendangregg.com/blog/2016-05-04/srecon2016-perf-checklists-for-sres.html

Christopher Hoey
Director SRE @ Datadog
mrchoey

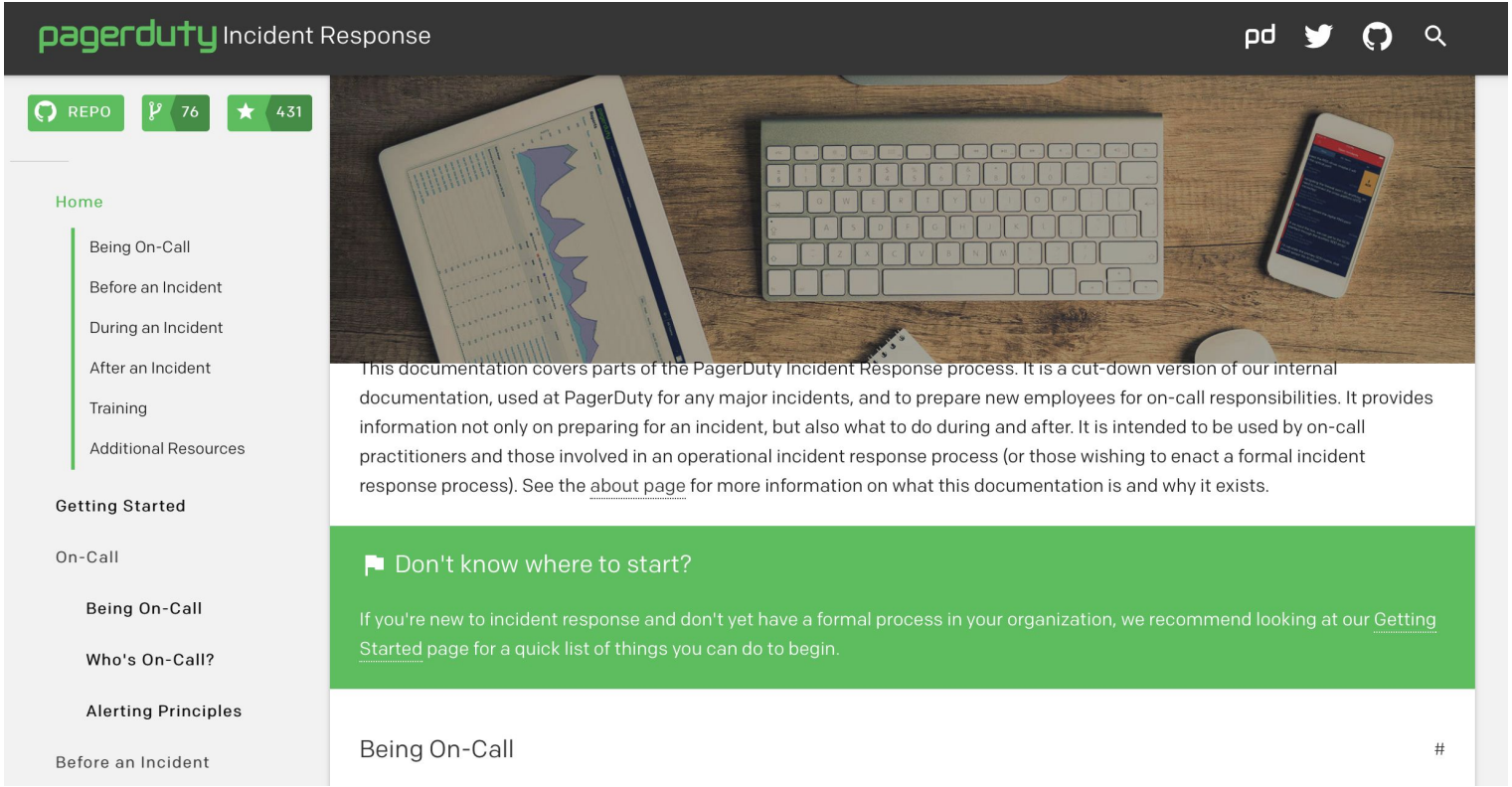
On-call handoffs

- Happen same time, same place, same day each week regardless of holidays
- Third party not on the outgoing or incoming rotation runs them
- Review open issues
- Review alert patterns
- Discuss pain points
- Follow up with teams as needed for recurring issues and toil
- Try to note patterns week over week to discuss with leadership



Christopher Hoey
Director SRE @ Datadog
mrchoey

Incident Response policies and procedures → https://response.pagerduty.com



Christopher Hoey
Director SRE @ Datadog
mrchoey

DATADOG

Takeaways

- Do not forget about your on-call team along your journey of growth
- Just as you would do with your apps measure everything you can about alert volume and on-call quality of life. Plant a solid foundation and use conventions early for ease of analytics later on
- Set and ruthlessly keep on-call handoffs to review alert volume, triage immediate issues, find broader systemic problems but most importantly keep your finger on the pulse of how on-call is going
- Experiment with on-call schedules and rotations. One size does not fit all and what worked yesterday likely won't continue to work tomorrow. Look at what other companies are doing but tailor on-call to your culture and stage of growth
- On-call pain is rarely spread equally. Some teams will be crushed. Be sensitive to their needs and reach out to find ways to help
- As your security and compliance requirements increase make sure on-call members are involved in the discussion. On-call life can be hard enough before all the tools and access gets yanked. Common goals, help us help you.

Christopher Hoey
Director SRE @ Datadog
mrchoey

Image resources

- https://upload.wikimedia.org/wikipedia/commons/e/e2/Amsterdam_-_Hats_-_0924.jpg
- https://ep1.pinkbike.org/p6pb15314668/p6pb15314668.jpg
- https://stackoverflow.com/questions/1732348/regex-match-open-tags-except-xhtml-self-contained-tags/1732454#1732454
- https://cdn.pixabay.com/photo/2013/07/18/10/56/graph-163509_1280.jpg
- https://c.pxhere.com/photos/2f/7f/leaf_growth_seed_plant_green_nature_agriculture_life-1094913.jpg
- https://i.pinimg.com/originals/30/c8/f0/30c8f065c2d2a202f9a387ac27f8d009.jpg
- https://img.purch.com/w/660/aHR0cDovL3d3dy5saXZlc2NpZW5jZS5jb20vaW1hZ2VzL2kvMDAwLzA1Ni82NTYvb3JpZ2luYWwvcmVkd29vZHMuanBn
- https://cdn.pixabay.com/photo/2017/10/18/14/31/box-2864328_1280.png
- https://upload.wikimedia.org/wikipedia/commons/f/f5/U.S.S._Enterprise_NCC_1701-D.jpg
- https://c1.staticflickr.com/5/4091/4976497160_026165c6cd_b.jpg
- https://c.pxhere.com/photos/f8/d5/adorable_pet_animal_breed_canine_curiosity_cute_dog-1198958.jpg

Christopher Hoey
Director SRE @ Datadog
mrchoey