

USENIX SREcon21

# Evolution of Incident Management at Slack

Brent Chapman

@brent\_chapman

bchapman@slack-corp.com





# Brent Chapman

Staff Engineer / Reliability Pillar / Slack





Join Extra Cr

Login

Search Q

Disrupt

Startups

Videos

Audio

Newsletters

Extra Crunc

EC-1s

Advertise

Events

More



Join Extra Crunch

Login

Search Q

Disrupt

Startups

Videos

Audio

Newsletters

Extra Crunch

EC-1s

Advertise

Events

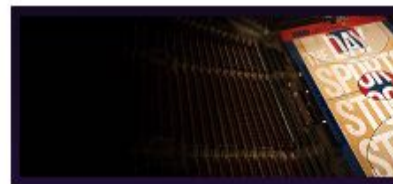
More

# Slack is down, so enjoy your three-

## Slack is down for some users (Update: Slack is back!)



NEWS JOBS EVENTS RESOURCES ABOUT



Trending: Are we just ants in a death spiral? Research

## Now that's scary: S

BY TOM KRAZIT on October 31, 2017 at 4:51 pm

Share 19 Tweet Share Reddit



### Connectivity

Today at 3:58 PM PDT

As of this story, workplace collaboration app Slack was do

A major outage took out Slack's services for v its userbase Tuesday afternoon.



Join Extra Crunch

Login

Search Q

Disrupt

Startups

Videos

Audio

Newsletters

Extra Crunch

EC-1s

Advertise

Events

More

## Slack is still down and it's past 5 o'clock, so go home (Update: It's back)

Devin Coldewey @techcrunch / 5:17 PM PDT • October 31, 2017

Comment



MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV

TECH

## Messaging app Slack is back online after suffering a worldwide service outage

PUBLISHED TUE, OCT 31 2017-8:29 PM EDT | UPDATED WED, NOV 1 2017-12:44 AM EDT

Saheli Roy Choudhury @SAHELIRC

SHARE f t in e

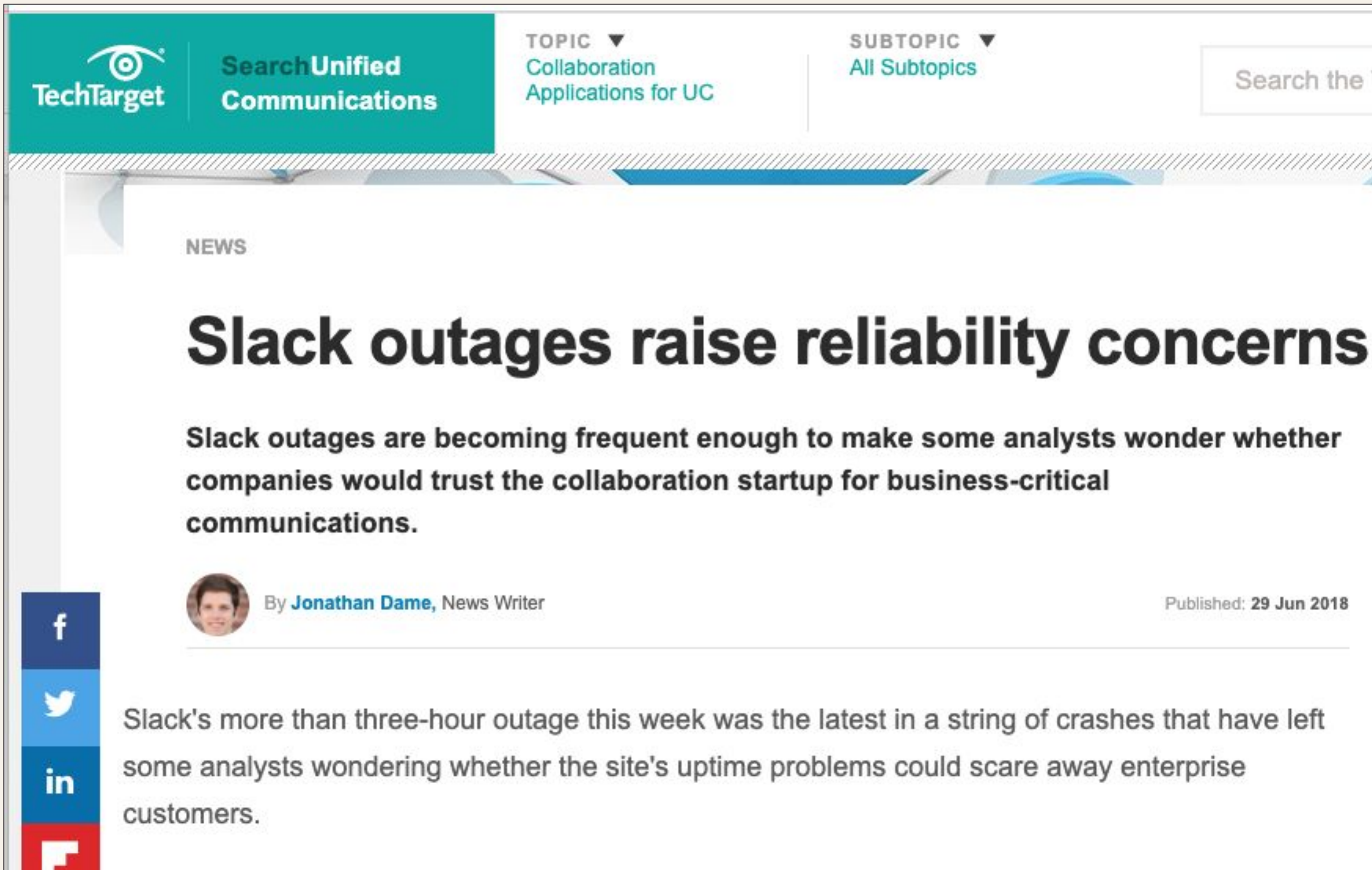
### KEY POINTS

- Slack, a popular messaging service, suffered a massive outage late Tuesday evening
- At about 9:35 p.m. ET (6:35 p.m. PDT), the company said that services were back online

Update: Slack is back, fo too late, though: everyon

Slack has been down for just over an hour, according to its official status page — which itself was down when I first checked a few minutes after the widely-used productivity tool bit it for us. At this point the end of the workday is pretty much shot, so you can go home. I would but I work from home so I have to sit here.





The screenshot shows a web page from TechTarget's SearchUnified Communications section. The page features a teal header with the TechTarget logo and navigation options for 'TOPIC' (Collaboration Applications for UC) and 'SUBTOPIC' (All Subtopics). A search bar is visible on the right. The main content area is titled 'NEWS' and contains an article with the headline 'Slack outages raise reliability concerns'. The article's lead paragraph states: 'Slack outages are becoming frequent enough to make some analysts wonder whether companies would trust the collaboration startup for business-critical communications.' The author is identified as Jonathan Dame, News Writer, and the article was published on 29 Jun 2018. On the left side of the article, there is a vertical stack of social media sharing buttons for Facebook, Twitter, LinkedIn, and Print. The beginning of the article's body text is visible below the author information: 'Slack's more than three-hour outage this week was the latest in a string of crashes that have left some analysts wondering whether the site's uptime problems could scare away enterprise customers.'



[Risks & Definitions](#)

## THE UTILITIES SECTOR OF THE S&P 500 IN ONE ETF

LEARN MORE



TECHNOLOGY NEWS MARCH 21, 2018 / 1:48 PM / UPDATED 3 YEARS AGO

# Slack Technologies builds engineering team to combat outages

By Salvador Rodriguez

3 MIN READ



SAN FRANCISCO (Reuters) - Collaboration software provider Slack Technologies Inc is building a safety engineering team that will develop methods to help reduce the disruptions that have been more frequent on Slack's service than rival systems, the company told Reuters.



Sept 2018

# Reliability Crisis



# Reliability Crisis

- Release tools and processes



# Reliability Crisis

- Release tools and processes
- **Service Ownership**



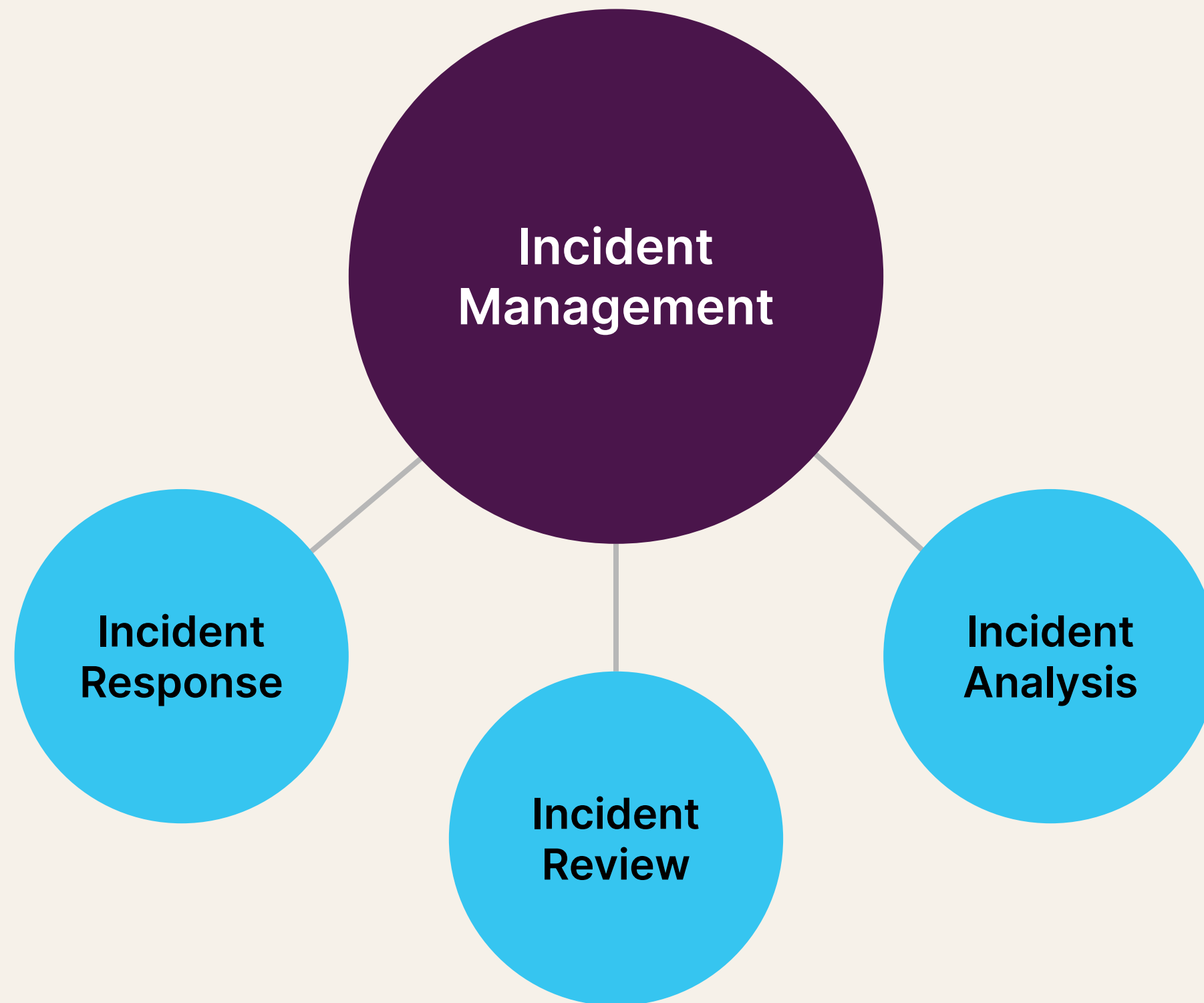


# Reliability Crisis

- Release tools and processes
- Service Ownership
- **Incident Management**



# Incident Management – 3 parts



# **Building support — Vision & Plan doc**



# Vision

1. IM recognized as key capability for company



# Vision

1. IM recognized as key capability for company
- 2. IM practices consistent across company**



# Vision

1. IM recognized as key capability for company
2. IM practices consistent across company
- 3. IM capabilities at all levels of the org**



# Vision

1. IM recognized as key capability for company
2. IM practices consistent across company
3. IM capabilities at all levels of the org
- 4. Effective mechanisms for developing & strengthening IM capabilities**



# Vision

1. IM recognized as key capability for company
2. IM practices consistent across company
3. IM capabilities at all levels of the org
4. Effective mechanisms for developing & strengthening IM capabilities
- 5. Routinely conduct blame-aware incident reviews**





# Vision

- 1. IM recognized as key capability for company**
- 2. IM practices consistent across company**
- 3. IM capabilities at all levels of the org**
- 4. Effective mechanisms for developing & strengthening IM capabilities**
- 5. Routinely conduct blame-aware incident reviews**



# Plan



# Plan

- 1. Awareness training during onboarding (1h)**



# Plan

1. Awareness training during onboarding (1h)
- 2. Responder training for all Engineers (3h)**



# Plan

1. Awareness training during onboarding (1h)
2. Responder training for all Engineers (3h)
- 3. Leadership training for experienced responders (3h)**



# Plan

1. Awareness training during onboarding (1h)
2. Responder training for all Engineers (3h)
3. Leadership training for experienced responders (3h)
- 4. Refresher training for experienced responders & leaders**



# Plan

1. Awareness training during onboarding (1h)
2. Responder training for all Engineers (3h)
3. Leadership training for experienced responders (3h)
4. Refresher training for experienced responders & leaders
- 5. Exercises to enable people to practice their IM skills**



# Plan

- ~~1. Awareness training during onboarding (1h)~~
- 2. Responder training for all Engineers (3h)**
- 3. Leadership training for experienced responders (3h)**
- ~~4. Refresher training for experienced responders & leaders~~
- ~~5. Exercises to enable people to practice their IM skills~~





# Training

*Engineering Incident Responder and  
Engineering Incident Commander* classes

- 3 hours each; based on PagerDuty's class

Incident lunch exercise

Incident Commander workshops



# Severity Levels

## Guidelines for

- **what sort of problem is what Sev Level**
- **what sort of response is appropriate**
  - **time expectations (after hours, 24/7, etc.)**
  - **priority relative to other activities**



# Major IC

**First oncall rotation of Incident Commanders**

**Someone on call 24/7**

**Follow-the-sun staffing**

- **Initially San Francisco, Melbourne, Dublin**



# Major IC

**Problem: Melbourne and Dublin teams are smaller**

- **Folks there are on call for Major IC more often**
- **... and also on-call for other services**

**Solutions:**

- **San Francisco handles weekends (low volume)**
- **Increase staffing in Melbourne and Dublin**
- **Reduce hours for both by adding Pune to Major IC**



# Major IC

**Problem: Lots of responsibility for an individual**

**Solutions:**

- **Strong/vocal management and exec support**
- **Peer support from fellow Major IC members**
- **“Bat Signal” pages all Major ICs in current and prior follow-the-sun group**



# Major IC

**Problem: Lots for IC to do; easy to lose track**

**Solutions:**

- **IC Checklist**
- **Incident Bot**



# Major IC

**Problem: More simultaneous incidents**

**Solutions:**

- **Create “Slack IC” oncall rotation to handle less-severe incidents**
- **Slack IC also gives newer ICs a place to gain experience before joining Major IC**



# Major IC

**Problem: Resource contention between incidents**

**Solution: Area Command**

- **Meta-incident, overseeing other incidents**
- **Area Commanders are our most experienced ICs**
- **AC prioritizes & arbitrates between incidents**
- **“Singleton” resources (Deploy Commander, Exec Liaison, etc.) “move up” to AC incident**





# Major IC

**Problem: Tying up Major IC on long-duration (long-tail) incidents**

**Solutions:**

- **Hand off to Slack IC**
- **Train all Eng Managers as ICs, and hand off to EM for the relevant team**



# Major IC

**Problem: Lots of incidents in certain teams/pillars**

**Solution: Pillar-specific IC on-call rotations**

- **Data Engineering**
- **Internal Tools**
- ***... more as we continue to grow ...***



# Incident Review

- Strong “blameless” culture
- Incident Commander is not responsible for leading Incident Review
  - Incident Review is driven by Eng Manager from most-involved team
  - IC is expected to be a key contributor



# **COVID-19 and WFH**

**No big deal for us, since we were already generally working incidents in Slack channels**

**Not being face-to-face didn't really hamper us**

**We do have to consciously work to build and maintain social bonds; no more “break room encounters” and “hallway chats”**



# What's next?

- **Continue to build roster of trained ICs**
- **Establish additional Pillar/Team-specific on-call IC rotations**
- **Improve methods for handling long-running (multi-day, sometimes multi-week) incidents**
- **Continue developing Incident Review capabilities**
- **Evolve an Incident Analysis practice**



# Ongoing challenges

**Slack itself is our key tool for Incident Response**

- **What do we do when Slack itself is down?**



# Ongoing challenges

**How do we recruit and train new incident responders and commanders?**

- **Make incident response involvement part of job ladders and promo expectations**

**How to practice skills and develop confidence?**



# Ongoing challenges

**How do you demonstrate the need for “more” to support and grow the program, when it already appears to be well-functioning?**

**How do you ensure that it remains successful?**







Thank you!

Brent Chapman  
@brent\_chapman  
bchapman@slack-corp.com