Shoreline

2022 Market Research: Benchmarking Production Operations

THE SURPRISING COST OF ON-CALL

Shoreline

“

With a sophisticated environment using managed Kubernetes, experienced engineers were necessary to ensure that all remediations were handled quickly and safely. Some of this work became quite repetitive, as the same issues cropped up daily, or more often.

Louis-Philippe Kronek
GM Dataiku Online

Contents

THERE'S A PROBLEM

The status quo is unsustainable

Cost of on-call is high

The average team spends over 2,000 hours dealing with incidents

Average cost: $2.5M/yr

Only 27% find on-call work fulfilling

Major incidents are frequent

38% or respondents have experienced 6 or more major incidents in the last 12 months

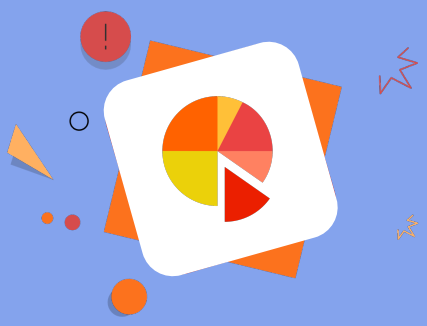
Average 8.7 in last 12 months

62% of these escalated to C-suite

Production ops is getting harder

Infrastructure is growing faster than headcount making it critical that you continuously improve the productivity

Cloud footprints grew 38% last 12 months



SRE teams only grew 26% last 12 months

Shoreline

Key recommendations for improving on-call operations

Generate incident insights

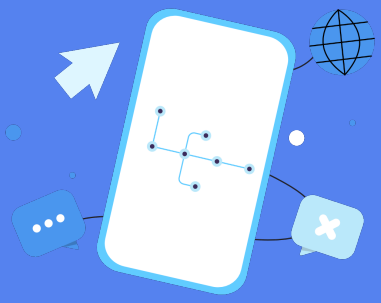
Too many companies can't measure the basics when it comes to incidents

98% struggle with incident ticketing and reporting

"If you can't measure it, you can't manage it" applies more than ever to on-call

Prevent escalations

Because escalated incidents are so frequent and take so long, they represent 78% of the total time spent on incidents



55% of incidents are escalated

3x time to resolve escalated issues

Work to eliminate toil

Eliminating low value, repetitive incidents with scripts and automations can dramatically free up your team's time

48% of incidents are repetitive

5x more human errors than automation errors

Introduction

Shoreline.io was built (and is currently run) by a team of developers passionate about improving the way companies manage on-call operations.

Shoreline was founded by Anurag Gupta, who built a multi-billion dollar database and analytics business from zero to over 5 million nodes at Amazon Web Services.

During his eight years at AWS, Anurag and his teams developed many best practices that proved to be essential in scaling more than 15 different cloud services. His experiences led him to create Shoreline. He knew companies of every size could benefit from what he learned about actually making a developer's life easier and optimizing the way others run their businesses.

It's that same spirit that led us to conduct this market research in partnership with Dimensional Research, with the goal of helping technology executives and cloud operations leaders make better decisions about how they run their businesses.

Most survey reports with titles like The State of DevOps provide a lot of raw practitioner data.

Shoreline and Dimensional Research have worked hard to go beyond the raw data to uncover detailed learnings, present key recommendations to run your business better, and tell the story of how businesses can manage on-call operations in an innovative and improved way.

This report is presented in five sections:

- High level observations on the average on-call operation and the most obvious opportunities to improve.
- The priorities and challenges around improving reliability.
- Shoreline's recommendations on the biggest opportunities most companies have to improve reliability.
- Benchmark data designed to help operations leaders understand how their team compares with similar companies.
- High level demographic data on the survey respondents to provide detailed insights into the source of our survey data.

High Level Observations

In this report, we surveyed more than 300 on-call practitioners, managers, and executives running businesses that ranged from managing fewer than 20 nodes to more than 10,000. A lot can be learned about the industry simply by looking at a theoretical "average company," as this data uncovers a number of fascinating challenges and opportunities for the DevOps industry.

# of Nodes Under Management:	1,772
# of SREs:	13.1
# of engineers available on-call at a time (support/L1/L2/L3:)	30.8
Hours spent on incidents per month:	2,084
Cost of on-call per year:	$2.5M [1]
# of incidents per month per engineer:	17.8
# of major incidents in the last year:	8.7*
Time to resolve non-escalated incidents:	3.6 hours**
Time to resolve escalated incidents:	10.7 hours***
% of issues that get escalated:	55%
% of issues that are straightforward and repetitive	48%
% of employees that find this work fulfilling:	27%

* 62% were escalated to the C-suite | ** 81% resolved in 2 hours or less) | *** 32% resolved in 2 hours or less

[1] Calculated by multiplying 2,084 hours on-call per month by $100/hour in employee cost by 12 months

The data above illustrates that customers are spending millions, and yet still suffer from major outages and thousands of hours of degraded service (in addition to stressing their SRE and on-call teams).

We estimate the cost of on-call operations at the average company is over $2.5 million.

On top of that, an average of 278 incidents occur each month and take a total of 2,084 hours to resolve. These issues could be as small as a page in an app running slow and missing the SLO for a single customer, or as large as widespread downtime that lasts hours. Even more surprising is the fact that the average customer experienced 8.7 major incidents in the last 12 months, with 62% of these incidents reaching the C-suite. In the end, the impact on customer experience and employee productivity is too big to ignore.

Later in this report, we'll cover the most common challenges that make improving reliability so difficult. But just for you, here's a sneak peek: Reducing the complexity of a product or service and giving operations teams more time to proactively invest in reliability stand out as areas companies should prioritize if they truly want to improve reliability.

But our focus right now is that when it comes to improving the productivity of on-call operations, reducing escalations and eliminating low value, repetitive work have the potential to dramatically improve productivity and the customer experience. Companies experiencing rapid growth of their cloud infrastructure will likely find their current approaches to on-call unsustainable. Over half of all incidents are escalated, and these incidents take three times longer to resolve. Finding new, different ways to empower support and L1 on-call to fix more incidents without escalation will be a huge win for team productivity.

Priorities and Challenges in Reliability

Reliability is growing in importance, with no signs of slowing down. We've been told by 97% of on-call stakeholders that their companies have priorities around reliability, including reducing the number of incidents, decreasing costs, and shortening time to recover.

What's interesting is how priorities change as maturity levels increase. Cost reductions are a top priority for companies at every level of maturity. But early stage companies are also prioritizing building a culture of constant, incremental improvement along with reducing complexity.

As companies get better at continuous improvement, their priorities adjust to

include reducing the number of incidents and shortening time to recover. The most mature companies seem to maintain focus on the number of incidents and time to resolve, but have an increased focus on end customer satisfaction.

97% report that their leadership has reliability priorities for cloud infrastrcuture

What are the top production cloud infrastructure reliability priorities for your organizations leadership?

Choose up to 3 of the following:

Other	Reduce onprem footprint; Security; reduce redundant data centre costs

Despite this growing focus on reliability, our survey results highlight several major impediments to improving reliability. At the top of the list is the complexity of the environments companies are managing. As the company's product complexity increases, it becomes harder and harder to find SRE and DevOps professionals that have the breadth of experience needed. So, it's no surprise that training and hiring are fourth and fifth on the list of challenges.

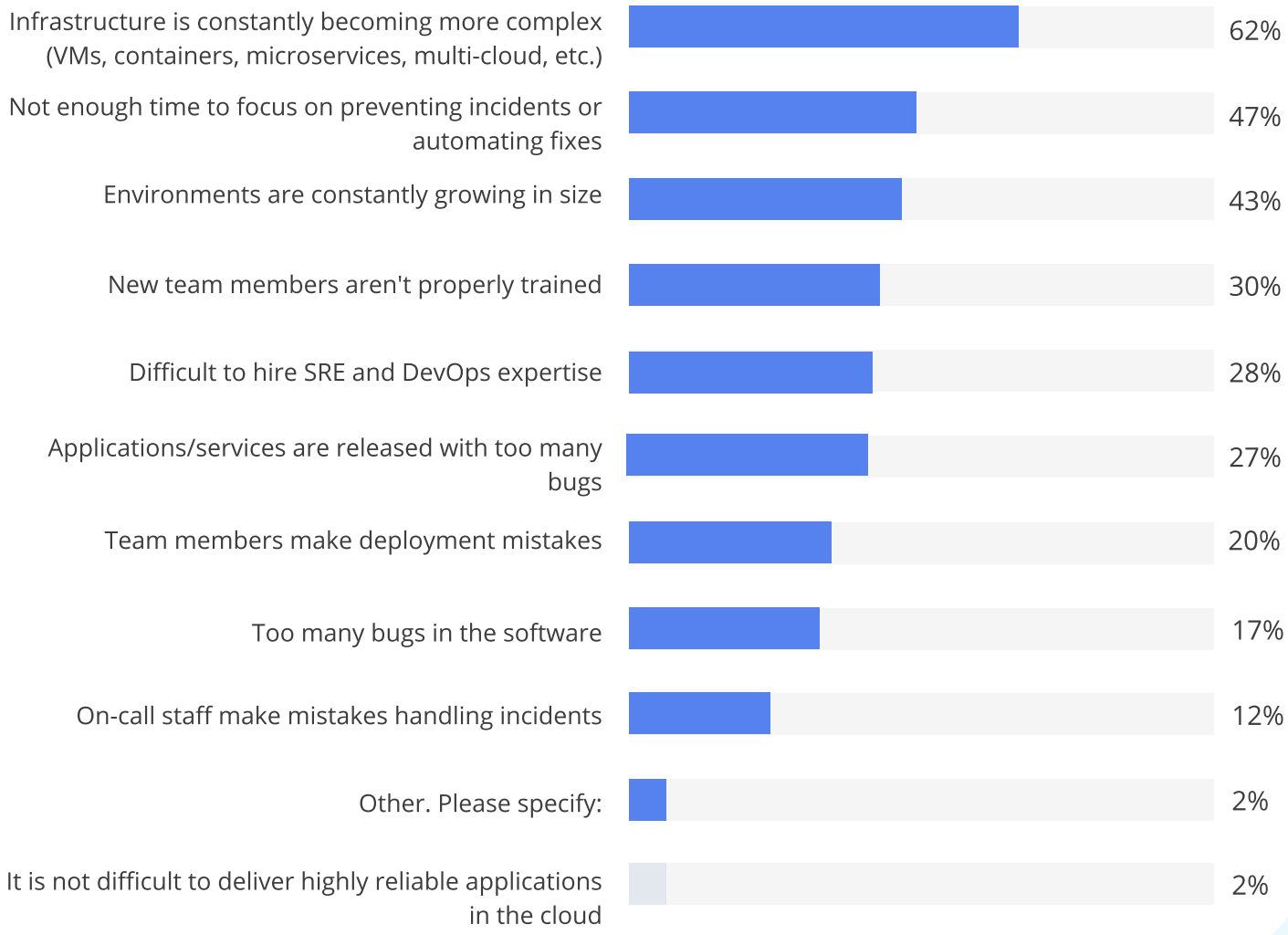
The second biggest issue is the lack of time to focus on preventing incidents or automating fixes. This can become a vicious cycle where the less time a team has, the less they can invest in

improvements, while the product continues to grow in size and complexity. As the workload on operations teams increases, people leave, causing the burden to be shared by fewer people. To avoid falling victim to this cycle, your organization should start investing in incident prevention and repair automation right away, no matter where you are in your product journey.

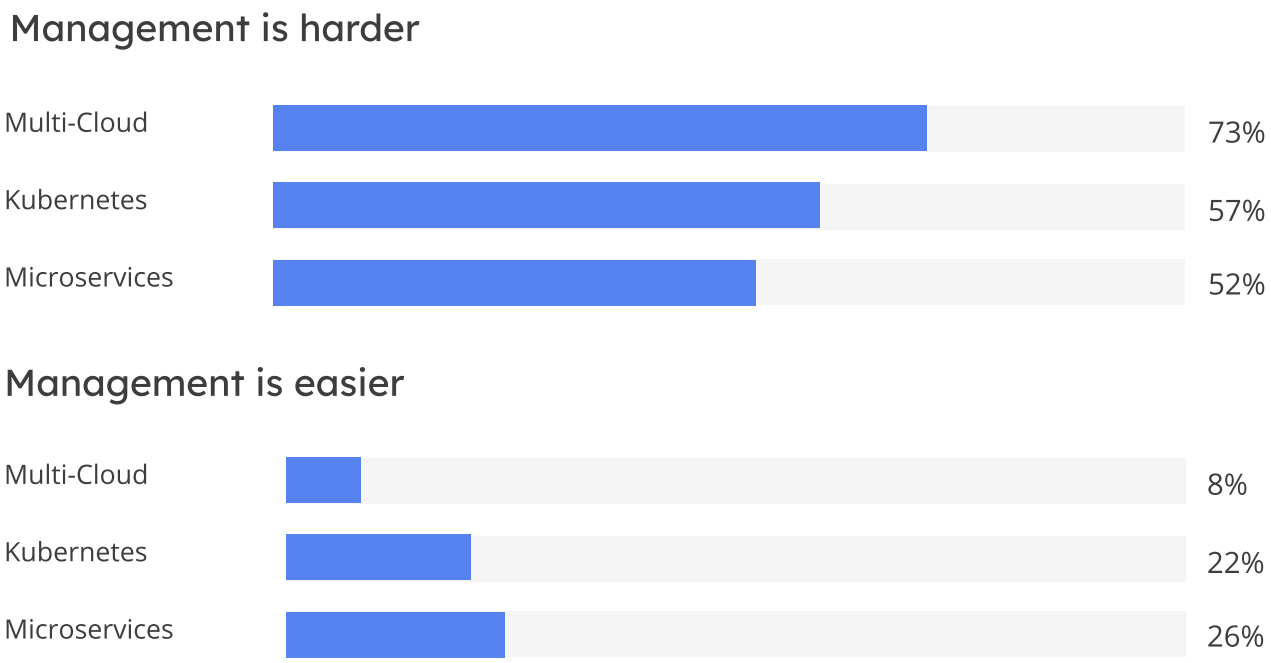
98% report they face challenges delivering highly reliable cloud applications

Why is it difficult to deliver highly reliable applications running in the cloud infrastructure?

Choose all that apply:



Modern Technologies Are Making Infrastructure Management Harder



One dimension of complexity is the technologies companies are using to build their products and services. The cloud, Kubernetes, and microservices all bring tremendous benefits like increased flexibility and agility. But, they also make it harder to maintain, diagnose, and repair incidents as they arise.

Not surprisingly, moving to a multi-cloud environment tops the lists of challenges, with 73% of on-call stakeholders saying that multi-cloud makes their job harder while only 8% say it makes the job easier. Next up on the list of challenges is Kubernetes, with 52% saying the technology makes their job harder and 22% saying it makes their job easier. Last but not least is microservices, with 52% saying it makes their job harder and 26% saying it makes their job easier.

Now these results do not mean companies shouldn't adopt these technologies. Instead, it means that companies should recognize the potential impact these technologies will have on their reliability team.

Key Recommendations for Improving On-call Operations

1) Ensure your incident management system provides detailed insights

While it may be the most boring recommendation in this report, it's probably the most impactful. While 63% of the industry considers itself to be good at ticketing, tracking, and management, 98% still report struggles with their incident management approach. This is because while most companies are doing an adequate job of assigning and routing work, very few are actually able to use their ticketing data to get insight into on-call operations. This data is critical for understanding opportunities to improve productivity.

The old adage "if you can't measure it, you can't manage it" has never been more true than in on-call operations.

To better understand your opportunities for improving productivity, make sure that all of your incidents are being tracked in a single location. All too often, some incidents are managed in email, others in Slack, and then the rest are in a separate system of record. From there, we recommend you get very clear on what you want to track and how you track it.

Here are key data points that we recommend you track with your incident reporting:

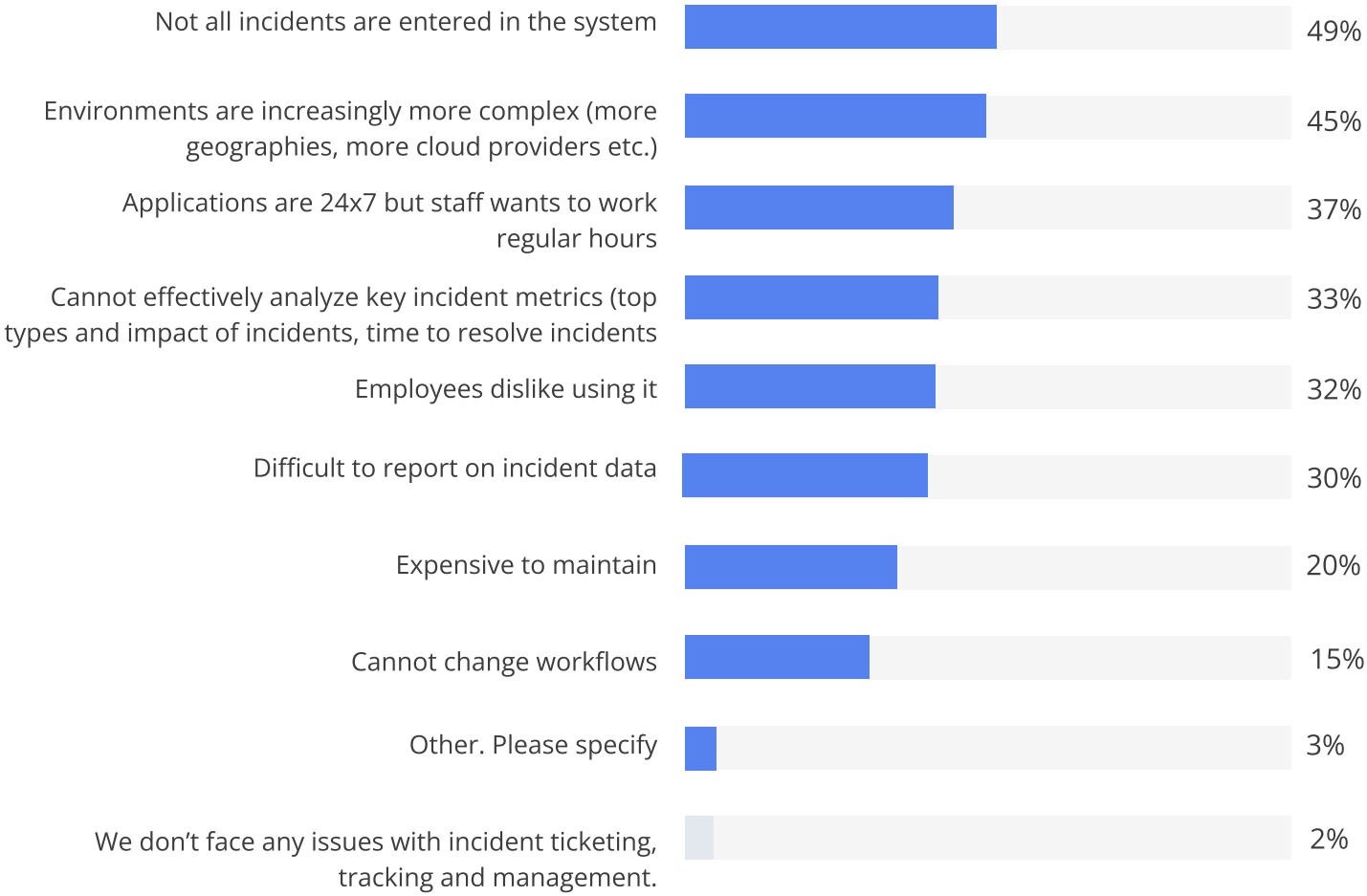
- Number of incidents per week
- Number of high severity incidents
- Categorize incidents by: service down, service degraded, no customer impact.
- Track the number of customers affected per incident
- Track the number of incidents that were escalated
- Track the number of people who touched each incident

Getting to this level of insight typically uncovers problems that have a bigger effect on the customer experience than your team realizes. Prioritizing these issues will likely deliver some important, quick wins for your customer experience and on-call teams.

98% report struggles with their incident management approach

What challenges does your organization face with your current approach to incident ticketing, tracking and management?

Choose all that apply:



2) Prevent escalations

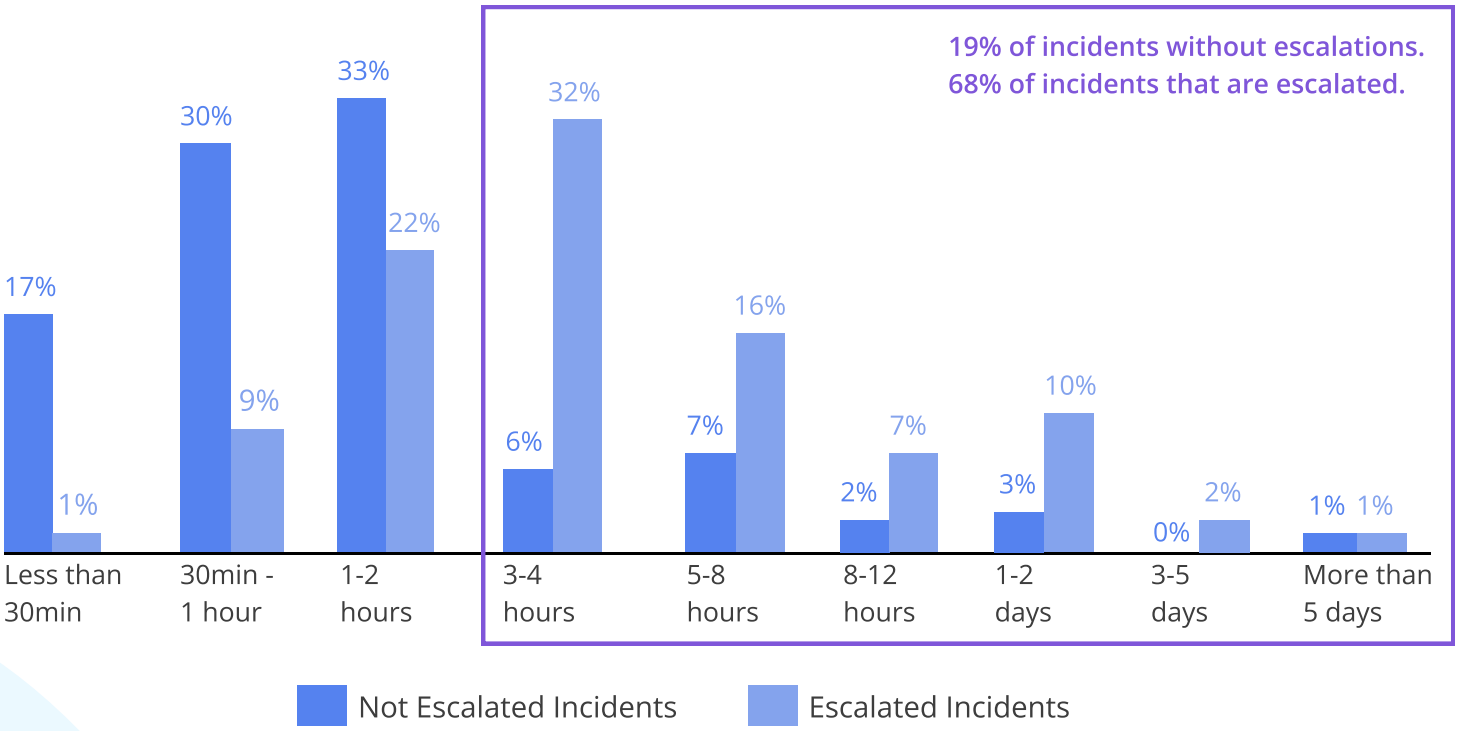
Based on the data from this study, reducing incident escalations represents the biggest opportunity to improve on-call productivity. On average, 55% of incidents are escalated, and they take **three times longer to resolve.** When you combine these two data points, 22% of on-call time is spent on non-escalated issues and a **whopping 78% is spent on escalations.**

Of course some of this is driven by the complexity of the problems that are being escalated, but a good portion is driven by inefficiencies in the handoff process itself. All too often the L1 engineer doesn't know who to

escalate to, so it takes time to pull the team together, identify the true cause, and actually fix the issue.

Investing in self-service tools to empower and support and L1 teams in being self-sufficient is a game-changer. Companies will heavily benefit from utilizing tools that help the initial responder capture diagnostics, identify the root cause, and — in some cases — actually fix the issue. With these tools, the initial responder will not only escalate fewer issues, they'll know who to escalate to. This leads to fewer needless interruptions and allows teams to provide better diagnostic data as part of the handoff, improving the overall time to resolution.

Average Time to Resolve Metrics By Company



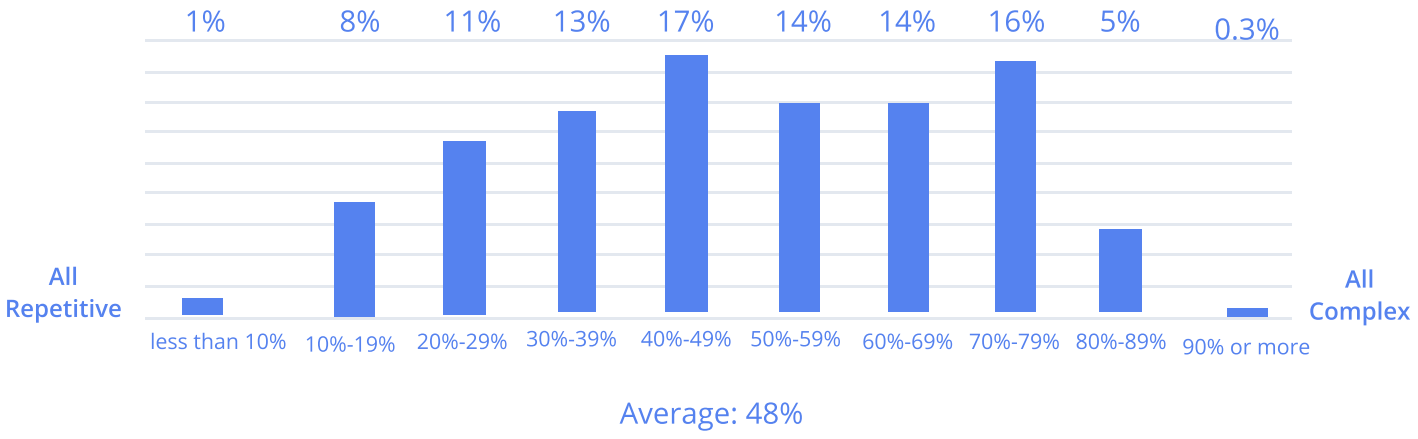
Escalated incidents are more than 3X more likely to take 3 or more hours to resolve.

3) Work to eliminate toil

On average, 48% of incidents are repetitive

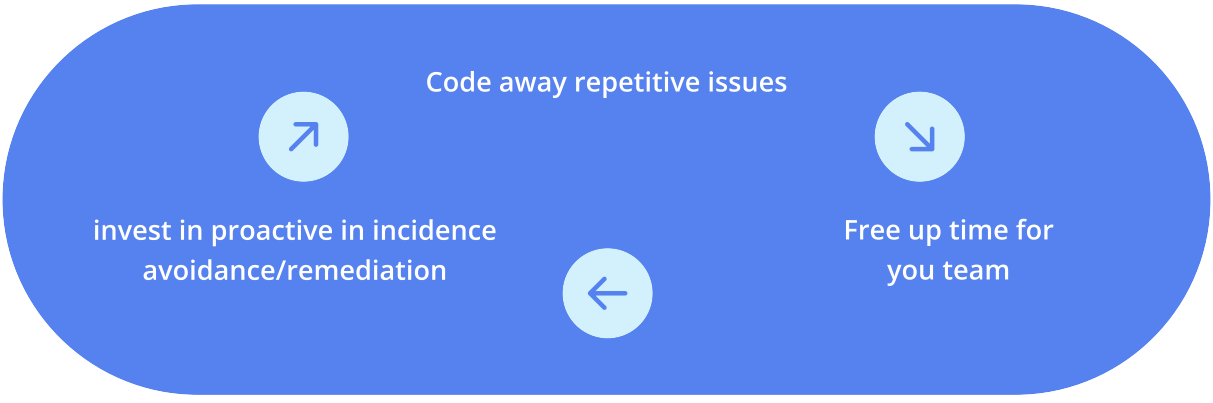
Think across all incidents that a typical on-call employee must deal with. To the best of your knowledge, approximately what percentage of these are straightforward and repetitive to resolve and how many are complex and require judgement to resolve?

Slide the dot on the bar to your answer, where 0%= “All incidents are repetitive” and 100% = “All incidents are complex”



While the average amount of low-value and repetitive incidents is 48%, the percentage is somewhat evenly distributed from 20% to 79%. In this case, to accurately determine the amount of toil affecting your teams, we recommend you conduct your own survey and ensure you’re prepared to effectively support your team in the ways they need.

Repetitive work represents a golden opportunity to improve the productivity of your team. The more you can free your teams from low value work, the more time they have available to improve resiliency, secure environments, lower costs, and further improve productivity. This work can enable a virtuous cycle where the better your team gets, the better it can become.



Automation isn't so scary after all...

It's not uncommon for people to be a little afraid of automation or creating code to automatically fix repetitive incidents. The incidents we often hear about in the news are the ones that happen to the largest tech companies in the world and these are often tied to issues in automation. If they face these kinds of incidents, why bother investing so much in automation?

The answer: Human error causes even more issues!

Our research finds that major incidents are nearly five times more likely to come from human errors (34%) than from automation (7%). So, while many avoid building incident automation due to fear of outages, doing so actually significantly increases the likelihood of their next major outage.

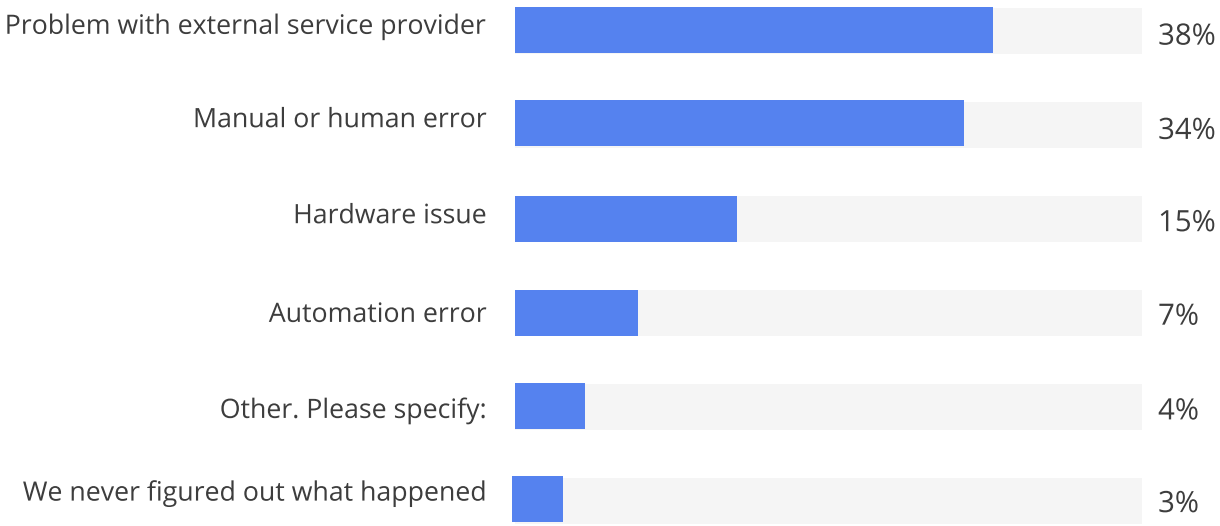
While this may be surprising, it is borne out by industry research. Studies across industries show that humans make between three and six mistakes per hour. Without the proper safeguards, it's just a matter of time before someone makes a very costly mistake.

Note, the number one cause of incidents in this study is a problem with an external service provider at 38%. When we filter down to the 18% of on-call practitioners that describe themselves as mature in on-call, external service providers only cause 20% of major incidents. This is presumably because these companies have invested more in failover and resilience.

Human errors nearly 5X more likely than automation errors to cause a major outage

Think back to your most recent major incident. What was the root cause?

Choose the one answer that most closely applies:



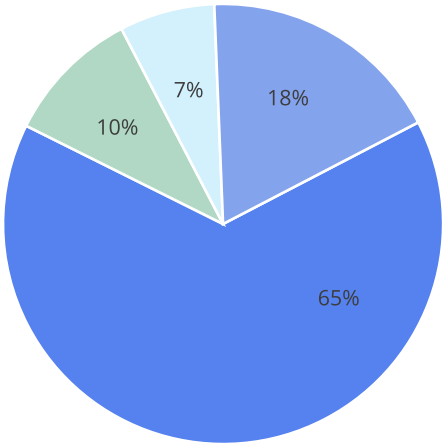
Other	Code issue; Configuration error; Patch broke system; Process error; Didn't realise process would lead to a timeout error; Still haven't determined root cause or even excluded AWS in this case; System upgrade caused unexpected interruption of service; User with too much rights; network issue; Old software version.

Benchmarking Data

This next section is designed to give operations leaders benchmark data to compare their organization against similar companies. We provide a summary of the size, maturity, challenges, and growth rates that companies are currently experiencing.

The on-call function is a work in progress

Which of the following statements best describes the maturity of your organization's on-call function?



- Mature - we have excellent processes, documentation, and tools for identifying, escalating, and resolving incidents, and to ensure issues are not repeated
- In-process - we have a good process, documentation, and tools for managing incidents, but could do better
- Getting started - our team has recognized this as an issue and has started to put in place some processs and tools
- Reactive - issues are reported to the help desk or support team, and dealt with on a case-by-case basis with no standard process

The data above highlights that on-call is a journey that almost everyone must take, but few have achieved excellence. Only 18% of respondents consider themselves to be mature while a whopping 65% rate themselves as "in-process."

Centralized SRE more common for on-call than development handling incidents

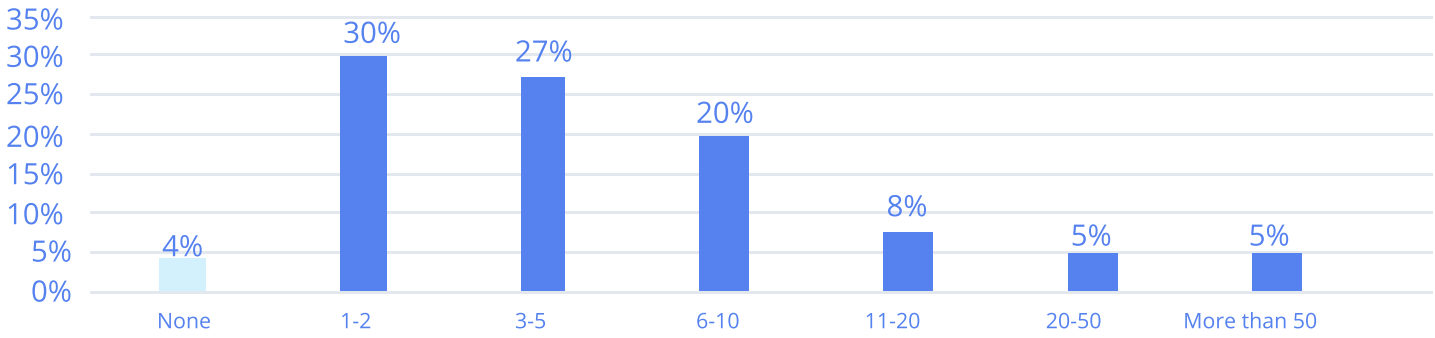
What model does your company use for on-call?

Choose the one answer that most closely applies:

A surprisingly high percentage of companies have a centralized approach to on-call. When you break the data out by industry, the software industry is less centralized, with 34% having a “you build it, you own it” model. Other industries like financial services, insurance, healthcare, and manufacturing come in at an average of 24%.

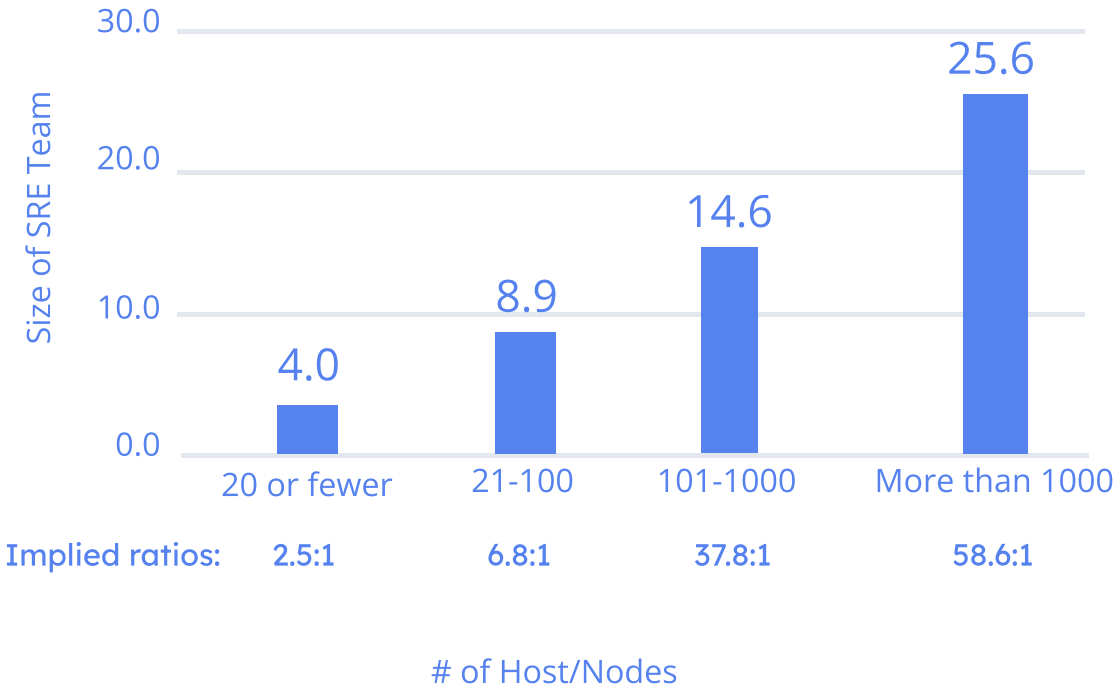
96% have experienced a major incident with their production cloud infrastructure in past year

How many major incidents with your production cloud infrastructure has your organization had in the past 12 months?



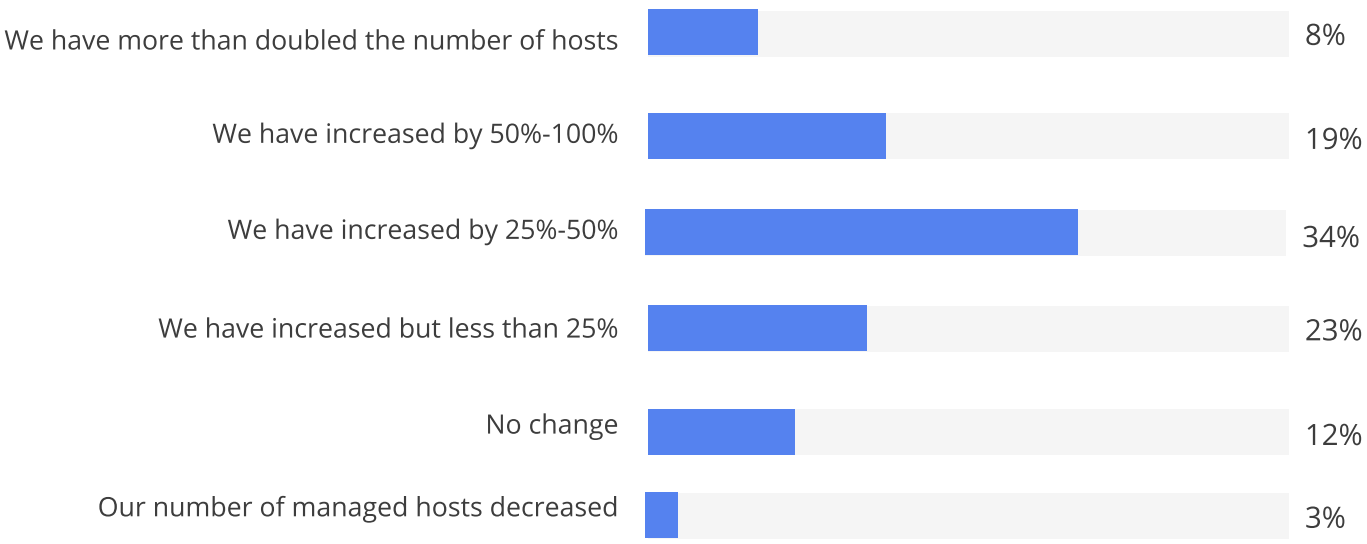
One of the most surprising data points from this survey is the number of major incidents that companies have experienced in the last 12 months. Only 34% of companies have experienced two or less major incidents, while 38% have experienced six or more.

Size of the SRE team vs. Cloud Footprint



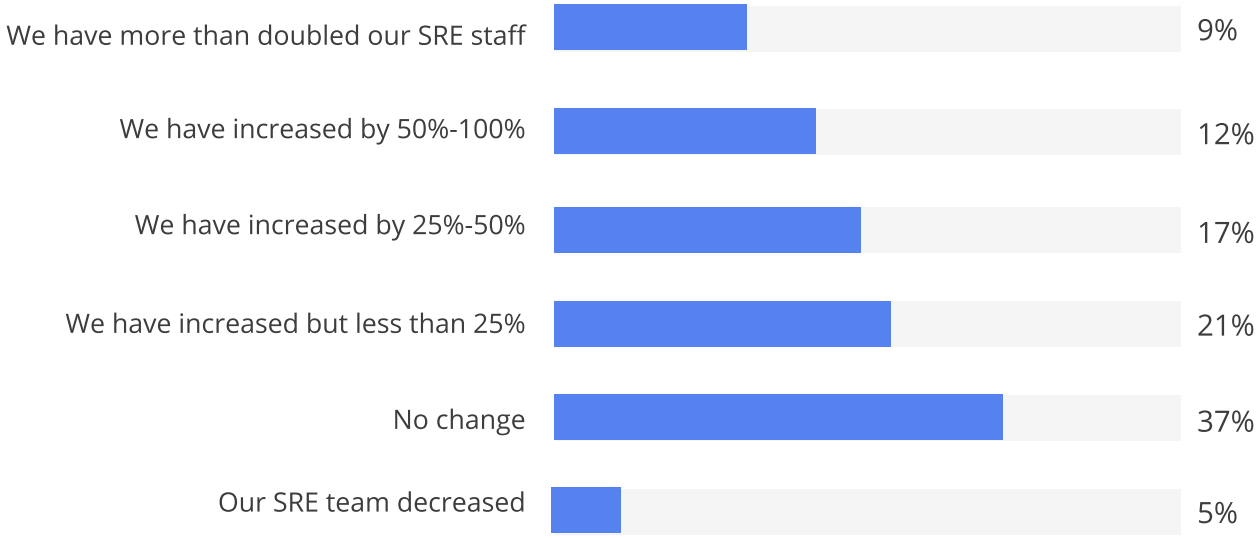
85% Have increased the number of cloud infrastructure hosts they manage

How has the number of cloud infrastructure hosts your organization manages changed in the past 12 months?



58% Grew their SRE team in the past year

How has the size of your organization's SRE team changed in the past year?



It's natural to expect economies of scale as you grow, but with growth comes complexity. More services, more inter-dependencies, and more varied workloads are just a few reasons why the job of on-call and reliability actually gets harder, not easier, as you grow. **You probably won't be able to grow your on-call team and DevOps teams at the same rate as your cloud footprint, so it's critical that you continuously improve the productivity of your team.**

As you look across the industry, the number of hosts grows significantly faster than the size of the SRE team. This leads to improved efficiencies AND puts increasingly more pressure on SRE teams at larger companies. This tradeoff is one that must be balanced carefully.

Even though the average SRE team grew rapidly, increasing 26% over the last year, the average cloud footprint increased even faster: 38% in the last year. All of these statistics point to the potential for significantly increasing stress levels and a likelihood for employees to churn.

72% of on-call employees are either irritated by or resigned to the manual, repetitive parts of on-call work.

In your opinion, how do on-call employees at your organization think of the part of their job that is spent on manual, repetitive tasks (rebooting servers, resizing disks, etc.)?

Choose the one answer that most closely applies:

Other	It depends. 50% resigned and 50% fufilled; Opportunity to write new automation; They are neither angry nor satisfied bit they think that it is their part of work

When it comes to toil, we recommend you survey your team to understand how high the frustration is at your company. 72% of on-call employees are either irritated or resigned by the manual, repetitive parts of on-call work, making turnover a real killer for most DevOps and SRE teams. It's important to have a finger on the pulse of employee satisfaction and to invest in improving it.

In this Shoreline case study, Dataiku points out that investing in incident automations has many soft benefits that can often go overlooked, including:

"A happier development team that enjoys building new software and new automations much more than repeatedly working on the same issues. This builds their coding skills, which is great for everyone."

About Shoreline

Shoreline is helping software engineers and site reliability engineers take the toil out of on-call.

Shoreline makes it easy to quickly build automations to continuously monitor and repair commonplace incidents. For incidents requiring human judgment, Shoreline reduces errors, repair time, and escalations with Jupyter-like notebooks that pre-populate diagnostics and provide step-by-step recipes for repair. For new incidents, Shoreline provides real-time fleetwide debugging, enabling engineers to precisely detect root causes and make repairs without needing to SSH into box after box.

Shoreline also offers a library of pre-built solutions that make it easier to diagnose and repair the most common infrastructure incidents in production cloud environments. Launching with over 35 Op Packs freely available to the community, the solutions library addresses issues like JVM memory leaks, filling disks, rogue processes, and stuck Kubernetes pods, among others.

For more information, visit shoreline.io.

dimensional research

The stakes are high for technology teams needing to deliver reliable and stable systems in a cost effective manner. Dimensional was delighted to partner with Shoreline to conduct this research investigating existing approaches to on-call and incident response for applications and systems running in production cloud environments. We wanted to both understand the current state of on-call, and identify opportunities to do better.

Dimensional's Research Highlights

98%

Face challenges delivering highly reliable cloud applications

96%

Exprerienced at least one major product cloud infrastructure incident in the past year

83%

Lost SRE team members in the past year

86%

Incident remediation automation is key to scaling cloud infrastructure

The data shows that the growth of cloud footprints is outpacing the growth of on-call teams. Cloud environments are becoming increasingly complex at a time when it is very challenging to find staff with the expertise to meet on-call demands. Incident response teams are left struggling to meet business demands for high reliability.

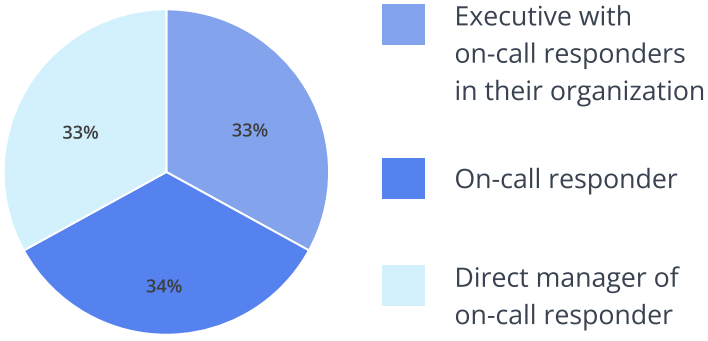
— **Diane Hagglund, Principal, Dimensional Research**

Research Methodology and Participant Profile

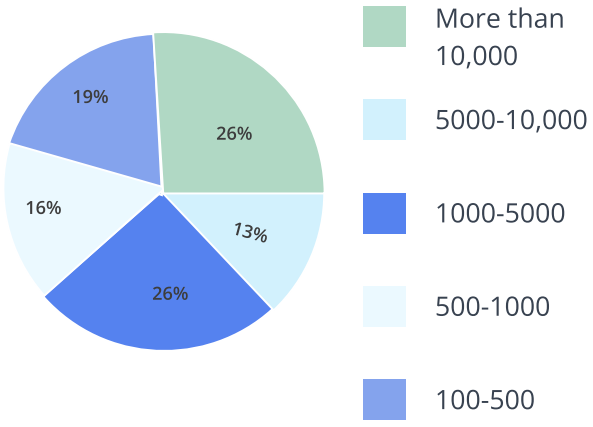
The data in this report is based on an online survey of stakeholders responsible for on-call incident response for production cloud infrastructure conducted by Dimensional Research. A total of 306 qualified individuals from independent sources completed the survey. In addition to responsibility for on-call incident response for production infrastructure in an IaaS environment, all worked at a company with at least 100 employees.

Participants included a mix of company sizes and industries. Quotas were set to ensure participation from a variety of job levels including at least 100 of each of the following three roles: front-line on-call responders (engineer, developer, DevOps, SRE, etc.), direct managers of on-call response staff, and executives with on-call responders in their organization. Shoreline was not revealed as the research sponsor. Survey responses were captured in July 2022.

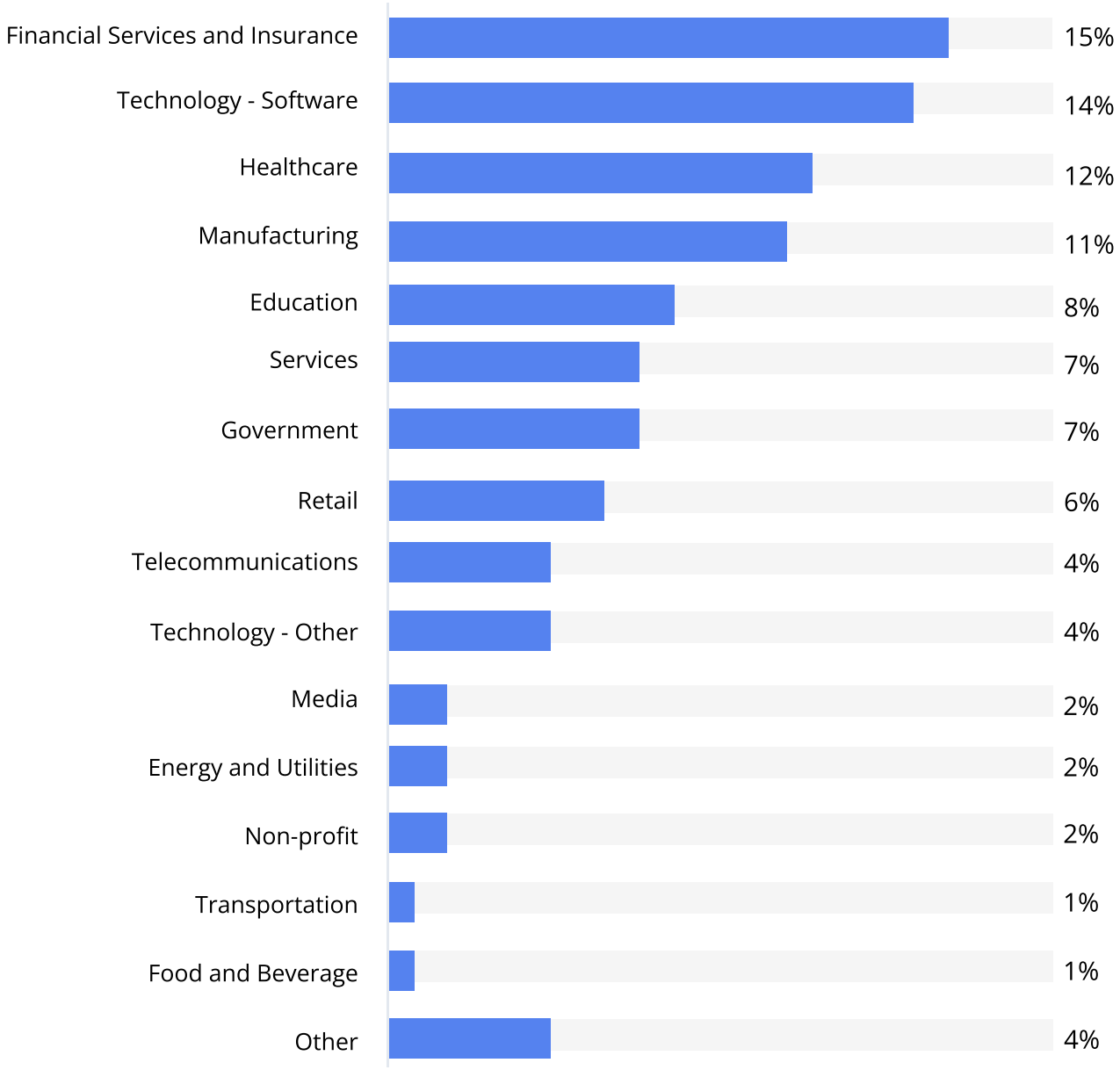
Role



Company size (# of employees)



Industry



About Dimensional Research

Dimensional Research® provides practical market research to help technology companies make their customers more successful. Our researchers are experts in the people, processes, and technology of corporate IT. We partner with our clients to deliver actionable information that reduces risks, increases customer satisfaction, and grows the business. For more information, visit dimensionalresearch.com.

“

Speed and safety are the most important factors when fixing degraded services or outages that directly impact customer experience.

incorta

Sumitha Sampath
VP of Engineering, Cloud SRE

Shoreline

www.shoreline.io

sales@shoreline.io

Learn more