PagerDuty

4 Steps to Prepare for an Outage

Prepare your teams and systems for an inevitable outage

The quest for 100% reliability is a battle that can't be won, but it is possible to get very close. While completely eliminating failure isn't an option, there are steps you can take to prepare for an eventual outage. The most important weapons in your arsenal are your internal monitoring and alerting systems. The way you set up and define your critical metrics is key in making sure you're well prepared and able to quickly restore availability when an outage occurs. This four-step process for setting up your metrics will ensure your team is ready to manage the unexpected.

Contents

PagerDuty

STEP ONE:

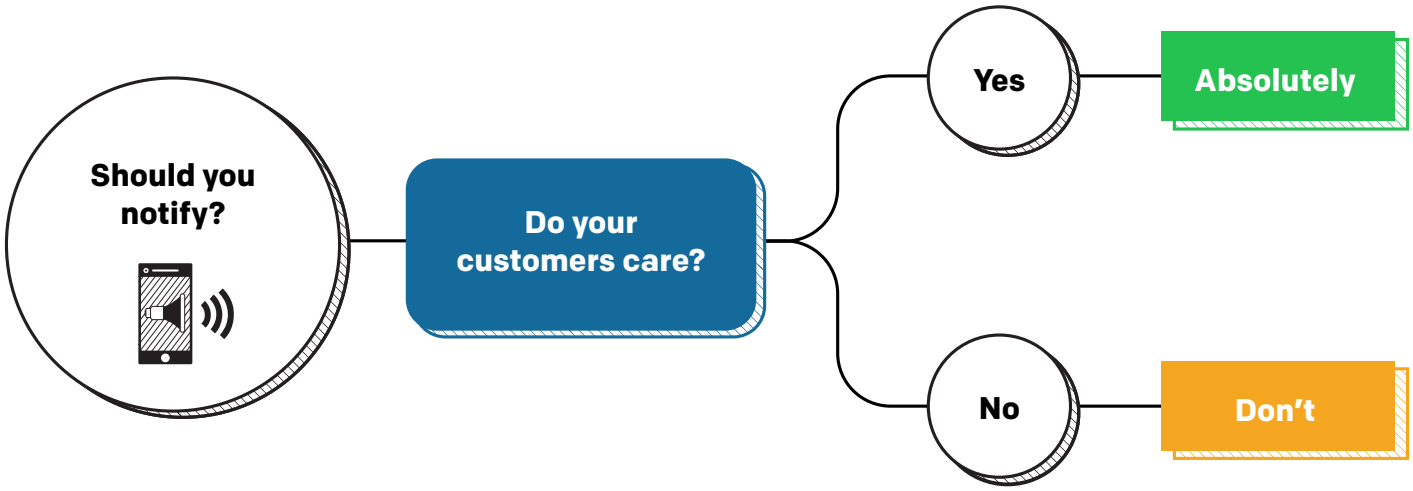
Define Your Business Critical Metrics

Before you can go to war, you've got to know what you're fighting for. When does a simple outage become a serious issue? When is it worth waking teams up in the middle of the night?

Not every IT incident results in a critical failure.

Your critical metrics should align with what matters most to your customers. You'll want to determine what functionality must remain available for them in the event of downtime. A single server going down may not have a customer-facing impact if you have several more to carry the load.

Customers and supporters don't need to be aware of every issue your business may face, but your customer-facing reliability should be a top priority. If it matters to your target audience, it should matter to you. If they don't care, you should retain that information for forensics but not notify your on-call staff about it, so they can focus on what requires an immediate response. As your business matures, your monitoring will evolve along with it.



STEP TWO:

Set Up Your Monitoring and Alerting

Now that you know exactly which metrics matter, it's time to set up your monitoring. This is essential; the repercussions for not monitoring will cost you much more than what you spend setting up your monitoring tools. When you're building your product, you need to think about where customer-facing failure points may be, then set-up appropriate monitoring before launching. Don't wait until something happens that takes your entire service down before you set up alerts.

Follow these best practices for setting up essential monitoring before you roll out new applications and services:

01

Analyze logs to provide valuable insight into what's going on with your systems at a granular level. They'll help you uncover and analyze bugs and even help you when testing new features in development. Correlated against captured metrics, scheduled attacks allow you to be proactive, going beyond just fixing problems to preventing them from occurring in the first place.

02

Discover bottlenecks by monitoring and managing the performance and availability of your applications. Is your system running slowly because of increased network traffic or is something else going on a little deeper? An Application Performance Management (APM) tool will help you identity these root causes.

03

Correlate metrics between customer and non-customer facing systems in order to find the root cause of an outage.

04

Limit the number of monitoring tools in your infrastructure. Every tool consumes computing resources, and you can only go so deep. It's impractical to monitor your monitoring tools and you can always adjust your monitoring later.

05

Funnel all your systems into a single view. Have one platform to plug into all of your monitoring tools, giving you the ability to scan the complete health of your system at a glance.

06

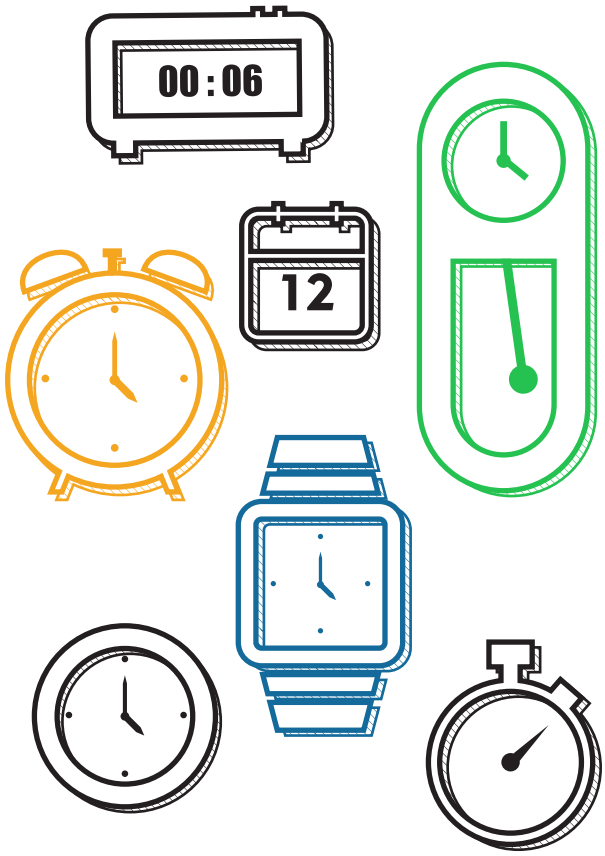
Connect your systems with people. Any metrics you monitor should be tied to alert the specific person or team best suited to handle an issue, should one occur.

07

Alert and schedule your people the way that will be most effective for your team and the way your individual team members prefer to be notified. This customization will lead to quicker response times and ultimately a faster mean time to resolution for all incidents.

08

Communicate. Make sure that the issue's status and progress towards resolution is properly communicated throughout the organization and your support team updates your customers via email, social media or other communication method.



Each incident of downtime is an opportunity to make necessary adjustments to your monitoring and alerting procedures.

If you take no action on an alert, adjust your monitoring to avoid unnecessary noise for something that isn't critical. At the same time, if you don't receive an alert for something that required action, you'll want to adjust the thresholds in your monitoring and alerting to make sure you're notified for those things in the future.

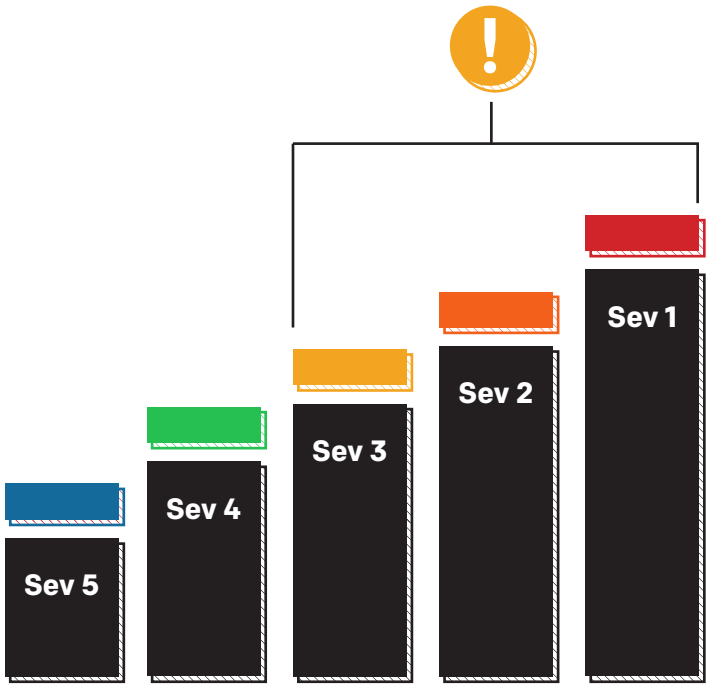
When an outage does occur, analyzing your alerting history is also helpful in determining incident severity. Look at how many alerts are received each week, and for each alert, ask: Was any action taken? Was a customer affected? Was this fully in my control? Perhaps low-level alerts at 3 a.m. initially require an engineer to acknowledge the incident, validate that it is not critical, go back to bed and resolve the root cause the next day. By tagging alerts with a severity level, or suppressing non-actionable alerts via rules, you can ensure that the on-call engineer only gets woken up for customer-impacting issues that need a response now. This leaves room for only high severity alerts and will help prevent alert fatigue.

STEP THREE:

Define Your Severity Levels

Just like the government's alerting system, a labeling system for outages will help your team better understand the severity of a specific incident, allowing you to prioritize and take the appropriate action. Severity levels should be determined based on your company's unique business model and customer reliability needs. Here is an example, with Severity 5 being the least impactful and easiest to fix, and Severity 1 being the most urgent.

Severity Level	Example
Severity 5	A bug that doesn't immediately impact customers
Severity 4	A minor bug or performance issue that has a limited impact on usability.
Severity 3	A delay or stop in usability for a limited number of customers.
Severity 2	A customer facing, critical metric, affecting the usability of your product or service.
Severity 1	A critical metric is impacted for a large number of customers over an extended period of time.



PagerDuty

STEP FOUR:

Create a Game Plan

Regardless of the severity of the incident, having a plan in place will make correcting it far easier, and eliminate the chaos.

A Standard Operating Procedure, or SOP, is a generic procedure manual that can be used in almost any major failure scenario. It doesn't do any good to create your SOP after the proverbial fan gets hit, of course. You need to have one in place well in advance.

Set up a dedicated conference bridge phone line or other method of communication in advance. Share the login and ID information so that everyone who might work on a problem has a quick and easy communication link, or even better, have the conferencing information directly embedded in any incident. This way, when there's a failure, the entire team can come together seamlessly on a call.

Designate an "Incident Commander." This person should be a seasoned veteran who can keep everyone on track, make sure balls don't get dropped and resolve any disputes.

Pinpoint a set of diagnostics the Incident Commander should share out while the call is being set up and people are joining. Diagnostics might include monitoring data, relevant graphs, related problems in other systems, etc., that will be useful when the conference call begins.

Identify a designated chat system for sharing non-verbal information such as data, links and code snippets to facilitate real-time collaboration.

Appoint someone on the critical response team to communicate directly with business stakeholders. This will ensure that both business and stakeholders can receive real-time updates on the progression of a customer-impacting outage, so lines of business (Support, Marketing, Legal, etc.) can immediately take the appropriate actions.

Having an SOP for high-severity issues reduces the variability of response and resolution times, and gets everyone up to speed quickly, including business decision makers. A clear procedure reduces stress, uncertainty and confusion when dealing with serious incidents.

There's No Such Thing as 100% Reliability

While there is no such thing as 100% reliability, it is possible to get very close. If you're prepared for an outage before it happens, with the appropriate metrics, strong monitoring and alerting systems, defined severity levels and an incident response plan in place, you'll be as close to 100% as you can get. And that's a strategy that will save time, money... and possibly your job.

Try PagerDuty Free for 14 Days

About PagerDuty

PagerDuty is the leading digital operations management platform for businesses, that integrates with ITOps and DevOps monitoring stacks to improve operational reliability and agility. From enriching and aggregating events to correlating them into actionable alerts, PagerDuty provides insights so you can intelligently respond to critical disruptions for exceptional customer experience. With hundreds of native integrations with operations tools, automated scheduling, advanced reporting, and guaranteed reliability, PagerDuty is trusted by thousands of organizations globally to increase business and employee efficiency.

For a free trial or to learn more, visit www.pagerduty.com/freetrial.

PagerDuty