pagerduty

POST-MORTEM TEMPLATE INCLUDED

The PagerDuty Post-Mortem Handbook

Lessons learned from the trenches and how you can conduct better post-mortems

Contents

Post-Mortems Are Necessary

No major incident is ever truly resolved without a post-mortem. Post-mortems are a great way for development teams to identify and analyze elements of a project that were successful or unsuccessful. It's a way to look back and review the incident in detail to determine exactly what went wrong, why it went wrong, and what can be done in the future to make sure it doesn't happen again.

A post-mortem can also be referred to as an after-action review, incident review, or follow-up review. While the name may be different, the process and goal is the same.

Sharing Our Incident Response Process

Reliability has always been one of the primary design considerations at PagerDuty. But what do we do when the unexpected happens and something does go wrong? It's of the utmost importance that we are prepared and can get our systems back into full working order as quickly as possible. We pride ourselves on being able to quickly resolve issues that arise and keep our systems working within their SLA. We've worked very hard to accomplish this, and our incident response process is where it all begins.

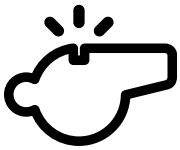
Our internal incident response documentation is something we've built up over the last few years as we've learned from our mistakes. It details the best practices of our process, from how to prepare new employees for on-call responsibilities, to how to handle major incidents, both in preparation and after-work. Few companies seem to talk about their internal processes for dealing with major incidents. It's sometimes considered taboo to even mention the word "incident" in any sort of communication. We would like to change that.

To that end, we'd like to share how we here at PagerDuty conduct post-mortems internally. It is our hope that others will use the documentation as a starting point to formalize their own processes. This guide provides information on what to do after a major incident and shares PagerDuty's follow-up and after-action review procedures.

Check out the rest of our incident response documentation to learn how we prepare for and handle incidents, as well as how we prep our teams to go on-call effectively.

First & Foremost, Create Response Roles

Creating response roles for individuals on your team gives each person specific follow-up tasks to be accountable for. These are generally lightweight tasks that ensure information is organized and customers are followed-up with accordingly. Below are the five response roles we assign.



Incident Commander

An Incident Commander acts as the single source of truth of what is currently happening and helps drive major incidents to resolution.

TASKS INCLUDE:

- Create the post-mortem page from the template, and assign an owner to the post-mortem for the incident.
- Send out an internal email to the relevant stakeholders explaining that we had an incident and provide a link to the post-mortem page.
- Check on the progress of the post-mortem to ensure that it's completed within the desired time frame.

Deputy

A Deputy is a direct support role for the Incident Commander. They support the Incident Commander so that the Incident Commander can focus on the incident at hand.

TASKS INCLUDE:

- There are no steps for a Deputy after an incident is resolved, however, the Incident Commander may ask for your help with their steps.

Scribe

A Scribe documents the timeline of an incident as it progresses and makes sure all important decisions and data are captured for later review.

TASKS INCLUDE:

- Review the chat communications and extract any relevant items from key events.
- Collect all to-do items and add them to the post-mortem.

Subject Matter Experts

A Subject Matter Expert is a domain expert or designated owner of a component or service that is part of the software stack that can quickly help identify and fix issues.

TASKS INCLUDE:

- Add any notes you think are relevant to the post-mortem.

Customer Liaison

A Customer Liaison is the person responsible for interacting with customers, either directly or via public communication channels with due care and attention.

TASKS INCLUDE:

- Reply to any customer inquiries received about the incident.
- Follow the progress of the post-mortem and update the company status page with the external message once it is available.

The Post-Mortem Process

For every major incident, a post-mortem is necessary. A blameless, detailed description of exactly what went wrong, along with a list of steps to take to prevent a similar incident from occurring again in the future is critical for ensuring you don't repeat mistakes; which could lead to burnout among your team or a loss of customer trust.

Read on and learn how to conduct a better post-mortem for your organization.

> *The incident response process itself should also be included as part of the post-mortem.*

Designate an Owner

At the end of a major incident call, or very shortly after, the Incident Commander should designate a post-mortem owner. The post-mortem owner is responsible for populating the post-mortem page, looking up logs, managing the follow-up investigation, and keeping all interested parties in the loop. The owner isn't required to do everything (tasks can be delegated out), but they are responsible for making sure they get done.

RESPONSIBILITIES OF POST-MORTEM OWNER INCLUDE:

- Scheduling the post-mortem meeting with all the relevant stakeholders within 5 business days of the incident
- Keeping the post-mortem page updated with all of the necessary content.
- Investigating the incident
- Creating follow-up tickets (only responsible for creating the tickets, not following up to resolution)
- Running the post-mortem meeting (these generally run themselves, but it's important to get people back on topic if the conversation starts to wander)
- Creating and reviewing a blog post with appropriate parties (if a public blog post is necessary)

Create a Post-Mortem Page

Once a post-mortem owner has been designated, it's time to start updating an internal wiki page with all the relevant information (this could also be as simple as a Google Doc).

01

Create a new post-mortem page for the incident (if not already done by the Incident Commander).

02

Populate the page with all of the information you have:

- The timeline should be the main focus to begin with.
  - The timeline should include important changes in status, impact, and key actions taken by responders.
  - Mark the start of the incident in red and the resolution in green to represent when you went into and out of severity.
- Identify the responders and add them to the page.
  - Identify the Incident Commander and Scribe in this list.

03

Populate the page with more detailed information.

- For each item in the timeline, identify a metric, or some third-party page where the data came from. This could be a link to a graph, a search, a Tweet; basically anything which shows the data point you're trying to illustrate in the timeline.

04

Perform an analysis of the incident.

- Capture all available data regarding the incident: what caused it, how many customers were affected, etc.
- Any commands or queries used to look up data should be posted in the page so others can see how the data was gathered.
- Capture the impact to customers.
- Identify the underlying cause of the incident: what happened and why did it happen.

05

Create any follow-up action tickets and note down topics for discussion.

- Go through your history to identify any to do items.
- Label all tickets with their severity level and date tags.
- List any actions which can reduce re-occurrence of the incident.
  - Note: There may be some trade-off here, and that's fine. Sometimes the ROI isn't worth the effort that would go into it.
- Identify any actions which can make our incident response process better.
- ***NOTE:*** *Be careful with creating too many tickets. Try to create things that are P0's or P1's — things that absolutely should be dealt with.*

06

Write the external message that will be sent to customers. This should be reviewed during the post-mortem meeting before it is sent out.

- Avoid using the word "outage" unless it really was a full outage, use the word "incident" instead. Customers generally see "outage" and assume everything was down, when in reality it was likely just some alerts delivered outside of SLA.
- Look at other examples of previous post-mortems to see the kind of thing you should send.

The Post-Mortem Meeting

Post-mortem meetings should generally only last 15-30 minutes (but can last longer for more complex incidents) and are intended to be a wrap up of the post-mortem process. It's important to discuss what happened, what could have been done better, and any follow-up actions that need to be taken. The goal is to suss out any disagreement on the facts, analysis, or recommended actions, and to get some wider awareness of the problems that are causing issues.

INVITE THE FOLLOWING PEOPLE TO THE POST-MORTEM MEETING,

ALWAYS	OPTIONAL
The Incident Commander	Customer Liason
Service owners involved in the incident	
Key engineer(s) and/or responders involved in the incident	

A GENERAL AGENDA FOR THE MEETING MIGHT LOOK SOMETHING LIKE THIS:

1. Timeline recap to make sure everyone agrees and is on the same page.
2. Important points and any unusual items recap.
3. Discuss how the problem could've been caught.
   - Did it show up in a canary release?
   - Could it have been caught in tests or loadtest environment?
4. Analyze customer impact.
5. Review action items that have been created, discuss if appropriate, or if more are needed.

Reviewing the Process

It's not only important to review the incident at hand, it's important to review your process. Ask yourself and your team if the incident was handled well, or are there things you could have done better? Typically, this involves a few of the incident commanders getting together to discuss how things might have been done differently, or if there are any tweaks we can make to the incident response process.

Post-Mortem Template

This is a standard template we use for post-mortems at PagerDuty — feel free to use it for your own!

Guidelines

This page is intended to be reviewed during a post-mortem meeting that should be scheduled within 5 business days of any event. Your first step should be to schedule the post-mortem meeting for within 5 business days after the incident. Don't wait until you've filled in the info to schedule the meeting, however, make sure the page is completed by the meeting.

Post Mortem Owner:

Your name goes here.

Meeting Scheduled for:

Schedule the meeting for within 5 business days after the incident. Put the date/time here.

Call Recording:

Link to the incident call recording.

Overview:

Include a short sentence or two summarizing the root cause, timeline summary, and the impact. (e.g. "On the morning of August 99th, we suffered a 1 minute SEV-1 due to a runaway process on our primary database machine. This slowness caused roughly 0.024% of alerts that had begun during this time to be delivered out of SLA.")

What Happened:

Include a short description of what happened.

Root Cause:

Include a description of the root cause. If there were any actions taken that exacerbated the issue, also include them here with the intention of learning from any mistakes made during the resolution process.

Resolution:

Include a description what solved the problem. If there was a temporary fix in place, describe that along with the long-term solution.

Impact

Be very specific here, include exact numbers.

Time in SEV-1

Time in SEV-2

Notifications Delivered out of SLA

Events Dropped / Not Accepted

Should usually be 0, but always check

Accounts Affected

Users Affected

Support Requests Raised

Include any relevant links to tickets

Timeline

Some important times to include:

Time the root cause began

Time of the page

Time that the status page was updated (i.e. when the incident became public)

Time of any significant actions

Time the SEV-2/1 ended

Links to tools/logs that show how the timestamp was arrived at

Responders

Who was the Incident Commander?

Who was the Scribe?

Who else was involved?

How'd we do?

What went well?

List anything you did well and want to call out. It's OK to not list anything.

What didn't go so well?

List anything you think we didn't do very well. The intent is that we should follow up on all points here to improve our processes.

Action Items

Each action item should be in the form of a ticket, and each ticket should have the same set of two tags: "sev1_YYYYMMDD" and simply "sev1".

Include action items such as:

Any fixes required to prevent the root cause in the future

Remaining post-mortem steps, such as the internal email, as well as the status-page public post

Any preparedness tasks that could help mitigate the problem if it came up again

Any improvements to our incident response process

Messaging

Internal Email

This is a follow-up for employees. It should be sent out right after the post-mortem meeting is over. It only needs a short paragraph summarizing the incident and a link to this wiki page.

External Message

This is what will be included on the status page regarding this incident. What are we telling customers, including an apology (the apology should be genuine, not rote)?

Summary

What happened?

What are we going to do about this?

Don’t Neglect the Post-Mortem

Don’t make the mistake of neglecting a post-mortem after an incident. Without a post-mortem, you fail to recognize what you’re doing right, where you could improve, and most importantly, how to avoid making the same exact mistakes next time around.

A well-designed, blameless post-mortem allows teams to develop comprehensive action plans and serves as a powerful building block in refining your incident response process. We hope that these guidelines and template help you and your team conduct better post-mortems and enable better knowledge share for better organizational learning processes.

pagerduty