ATLASSIAN
Opsgenie

As 5 fases do gerenciamento de incidentes e como melhorá-las

Sumário

Vamos começar

Simplificando, o gerenciamento eficaz de incidentes é uma parte essencial de todo o sistemas de negócios de empresas. Por quê? Porque como ferramentas tecnológicas e fluxos de trabalho estão se tornando cada vez mais complexas e interligadas, os sistemas tornam-se vulneráveis à inatividades não planejadas As chamadas “quedas” podem atingir qualquer sistema a qualquer momento - impactando as operações comerciais internas e externas. Os custos dos incidentes são, em sua maioria, medidos em dezenas, senão centenas, de milhares de dólares por minuto.

Com esse impacto potencial na linha, as organizações procuram evoluir rapidamente em práticas de resposta a incidentes, para garantir que eles possam ser gerenciados efetivamente o quanto antes possível. Isso significa adotar uma abordagem holística para um incidente, entender como ele evolui e como melhorar constantemente a resiliência dos sistemas. Do ponto de vista acadêmico, existem várias opiniões sobre como muitos estágios estão associados a um fluxo de trabalho típico de resposta a incidentes. Embora isso possa ser diferente para cada organização, vamos nos concentrar em seguir essas cinco etapas que representam o ciclo de vida de um incidente:

1. Preparação
2. Detecção & Alarme
3. Contenção
4. Remediação
5. Análise

Sem considerar cada uma dessas etapas, as organizações estão expondo-se ao risco de que os incidentes sejam mal administrados, resultando em atrasos desnecessários e altos custos. A seguir, vamos olhar para cada uma dessas etapas e oferecer recomendações sobre práticas que ajudarão as equipes a resolver incidentes de forma mais eficiente.

Preparação

Até mesmo os profissionais de TI mais experientes dirão que a preparação é uma parte essencial do gerenciamento de incidentes. É o palco onde as equipes exploram os cenários “e se” e depois definem os processos para abordá-los.

As organizações líderes fazem questão de se concentrar na preparação da mesma maneira que os atletas praticam um esporte. O objetivo é construir memória muscular ao redor da resposta a incidentes, de modo que as reações possam ser mais rápidas quando, de fato, algo acontecer.

> As metodologias de resposta a incidentes tipicamente enfatizam a preparação - não apenas estabelecendo uma capacidade de resposta a incidentes, mas também para preveni-los, garantindo que sistemas, redes e aplicativos estejam suficientemente seguros.”
>
> NIST

Ideias para melhorias

- **Esteja sempre preparado.**

Um “paraquedas” para os respondentes de incidentes é um backup de informações críticas que as equipes precisam para solucionar problemas com o mínimo de atraso. Ao centralizar esse material em um único local, as equipes têm o conhecimento na ponta dos dedos em vez de precisar procurá-lo. Dependendo da estrutura das equipes e dos sistemas da organização, isso pode incluir uma variedade de coisas:

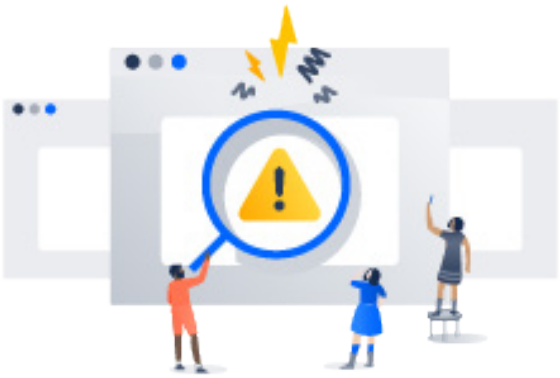
- Planos de resposta a incidentes
- Listas de contatos
- Cronograma(s) de plantão
- Políticas de escalonamento
- Links para ferramentas de conferência
- Códigos de acessos
- Documentos de apólice
- Documentação técnica e “runbooks”

- **Não fuja dos runbooks.**

Os runbooks oferecem aos membros da equipe uma orientação essencial sobre as etapas a serem seguidas em um determinado cenário. Isso é especialmente importante para equipes que trabalham em horários rotativos e/ou quando um especialista do sistema não pode ser imediatamente contatado. Sem os runbooks disponíveis, os respondentes não familiarizados com o sistema ficam correndo em círculos, tentando determinar quais etapas precisam ser seguidas para começar a remediação. Um conjunto de runbooks atualizado e disponível não apenas permite que equipes respondam mais rápido, mas também constrói coletivamente uma base de conhecimento que promove a melhoria contínua das práticas de resposta a incidentes.

- **Aceite o caos e promova a estabilidade.**

O termo “Engenharia do Caos” pode até parecer um oximoro. Mas não é. É, na verdade, à prática de experimentar, inserindo falhas em sistemas, para entender como eles podem ser construídos de forma mais robusta e segura. Um exemplo disso é Chaos Monkey. Originalmente desenvolvido pela Netflix, Chaos Monkey é uma ferramenta que testa a resiliência da rede, intencionalmente levando sistemas de produção a se desligarem. Embora aparentemente perigosa, a prática realmente ajuda os engenheiros, que continuamente testam os sistemas para garantir a capacidade de recuperação. Desta forma, o Chaos Monkey ajudou as equipes da Netflix a construir uma cultura em torno da resiliência do sistema. Com esse sucesso, muitas outras organizações seguiram o exemplo desta prática.



Detecção & Alarme

A detecção de incidentes não se concentra apenas em saber que algo está errado, mas também sobre como as equipes são notificadas sobre isso. Enquanto estes dois processos podem parecer separados, eles são, de fato, muito conectados. O desafio é que enquanto a proliferação de ferramentas de monitoramento de TI disponíveis melhorou a capacidade das equipes de detectar anormalidades e incidentes, elas também podem criar “tempestades em copo d’água” ou falsos positivos que complicam a resposta no processo.

As equipes de TI mais “tops” adicionam uma camada ao processo de monitoramento para garantir que os alertas sejam gerenciados corretamente. Essa camada age para centralizar o processo de alerta, enquanto também desenvolve inteligência adicional para a forma como eles são entregues.

> A detecção deve levar à resposta apropriada. Isso exige, principalmente, a necessidade de identificar e comunicar claramente os papéis, responsabilidades e a abordagem inicial para o tratamento de incidentes. Deve incluir a determinação de identificar o incidente e analisar a sua gravidade como um meio para lidar com o problema de forma eficaz dentro do contexto organizacional.”
>
> MITA

Ideias para melhorias

- **Pense fora do NOC.**

Historicamente, os centros de operações de rede (NOCs) atuavam como um hub de alertas e monitoramento para sistemas de TI de grande escala. O desafio é que um típico engenheiro de NOC pode ser responsável pela triagem e escalada de incidentes de qualquer lugar no sistema. Ferramentas modernas de gerenciamento de incidentes fazem com que esse processo seja simplificado significativamente. Automatizando a entrega de alertas, os fluxos de trabalho, as agendas de equipe e as políticas de escalonamento, o potencial de erro humano e/ou atrasos podem ser evitados.

- **Agregue, não agrave.**

Nada é pior do que receber uma onda contínua de alertas vindos de várias ferramentas de monitoramento. Ao centralizar o fluxo de alertas por meio de uma única ferramenta, as equipes são capazes de filtrar melhor o ruído para que possam focar rapidamente nos assuntos que precisam de atenção.

- **Conhecimento = poder.**

Um alerta básico indica que algo está errado, mas nem sempre expressa o que exatamente é. Isso causa atrasos desnecessários, pois as equipes devem investigar e determinar quais foram as causas. Integrando alertas aos detalhes técnicos do porquê foi acionado, o processo de remediação pode começar mais rápido.

- **Quis custodiet ipsos custodes?**

A frase em latim “Quem está guardando os guardas?” identifica um problema universal enfrentado por todas as equipes de TI. Isso ocorre porque as ferramentas de monitoramento que eles empregam são igualmente vulneráveis a incidentes e paralisações, assim como os sistemas que são projetados para proteger. Sem uma maneira de garantir que essas ferramentas funcionem corretamente, os sistemas podem ficar off-line facilmente, e sem ninguém saber. Os processos de alerta holístico garantem que tanto os sistemas como as ferramentas que eles monitoram sejam constantemente verificados.

Contenção

O processo de triagem para um incidente de TI é semelhante aos processos implantados na área médica. O primeiro passo é identificar a extensão do incidente. Depois, ele precisa ser contido para evitar que a situação piore. Todas as ações tomadas nesta fase devem ser focadas em limitar e evitar que mais danos ocorram.

> A contenção de curto prazo não se destina a ser uma solução a longo prazo para o problema; destina-se apenas a limitar o incidente antes que piore.”
>
> R. BEJTLICH, O INSTITUTO DE BROOKINGS

Ideias para melhorias

- **Estanque o sangramento.**

Médicos de triagem sabem que estão arriscando aumentar os danos se ficarem presos na tentativa de resolver todas as situações. Seu foco está em ações de curto prazo que estabilizam o paciente o suficiente para movê-lo até um cuidado mais agudo. Nos campos tecnológicos, as ações de contenção se concentram em soluções temporárias (isolando uma rede, regredindo uma compilação, reiniciando servidores etc.) que, no mínimo, limitam as ações do incidente ou, idealmente, trazem os sistemas de volta on-line. Se os esforços de gerenciamento de incidentes focarem puramente na remediação, e não na contenção, uma interrupção pode ser estendida desnecessariamente enquanto a solução permanente ainda está sendo procurada.

- **Não faça tudo isso sozinho.**

A “cultura do herói” nas equipes de TI é uma filosofia que está morrendo. Não está mais na moda ser o engenheiro solitário que trabalha horas e fins de semana intermináveis porque é a única pessoa que pode trazer os sistemas de volta on-line. Em vez disso, as equipes estão trabalhando só como isso: equipes. Colaborando em problemas porque entendem que os incidentes podem ser resolvidos mais rapidamente quando o conhecimento é compartilhado. Linhas de conferência, ferramentas de bate-papo e feeds de vídeo ao vivo tornam-se elementos essenciais da caixa de ferramentas de gerenciamento de incidentes. Estes podem se juntar rapidamente às equipes para que elas possam colaborar em tempo real.

É também comum para as equipes integrarem ferramentas de bate-papo com ferramentas de gerenciamento de incidentes para que estes possam ser acionados, reconhecidos e resolvidos em uma única plataforma.

- **Seja transparente.**

A era digital faz quantidades infinitas de informações disponíveis a qualquer momento. No meio de um colapso de TI, isso pode ser uma vantagem ou desvantagem. Se os usuários sofrerem uma interrupção de serviço, é comum que o incidente seja divulgado para o público rapidamente. Para sempre ficar à frente, as equipes devem ter um plano de comunicação de incidentes no jeito. O objetivo é criar confiança junto aos clientes, reconhecendo publicamente que uma queda está ocorrendo e assegurando que as etapas estão sendo tomadas para resolvê-la. Ferramentas como Twitter, StatusPage e fóruns de usuários são ótimos lugares para compartilhar essa informação. Este processo deve ser projetado para continuar por meio das fases de remediação e análise, aumentando ainda mais a confiança com usuários que, de outra forma, poderiam abandonar o sistema.

Remediação

Intimamente vinculada à contenção é a remediação. Aqui, é onde a longo prazo são implementadas soluções que garantam que o incidente foi abordado completamente e eficazmente. Onde na contenção o objetivo pode ser trazer sistemas de volta on-line, na remediação o objetivo muda para entender o que causou o problema e como ele pode ser corrigido para evitar que incidentes semelhantes ocorram no futuro.

> “Antes da recuperação total do sistema, os esforços de correção devem ser executados para remediar a fonte do problema. O estágio final da recuperação é não apenas restaurar o sistema para onde estava, mas torná-lo melhor e mais seguro. O sistema deve ter as mesmas capacidades operacionais, mas também proteger contra o que causou o incidente em primeiro lugar.”
>
> DEPARTAMENTO DE SEGURANÇA NO DOMICÍLIO DOS EUA

Ideias para melhorias

- **Cynefin.**

Uma estrutura de tomada de decisão, a Cynefin (pronuncia-se “KUN-iv-en”) fornece uma maneira estruturada de abordar problemas que ajuda os agentes de incidentes a determinarem o melhor curso de ação com base na natureza do problema em si. Dependendo do tipo de incidente (simples, complexo, complicado, caótico), uma abordagem para resolvê-lo pode ser definida.

- O incidente tem causa e solução conhecidas?
- Preciso envolver outras pessoas para ajudar a resolver esse incidente?
- Há tempo para investigar o problema e identificar a melhor resposta ou a situação requer ação imediata?

- **Que tal automatizar?**

Ferramentas de bate-papo se tornaram um recurso ideal para as organizações melhorarem a comunicação e a colaboração. No entanto essas ferramentas evoluíram muito além de simplesmente permitir que as equipes enviem mensagens. A equipe de desenvolvimento de software no GitHub foi pioneira na evolução das ferramentas de chat quando lançou a ferramenta de código aberto Hubot, que permite que os usuários acionem ações e scripts diretamente dentro de um ambiente de bate-papo. Isso permite que as equipes simplifiquem as operações, criando bots que automatizam processos (iniciando uma reinicialização do servidor, implantando um trecho de código etc.).

Análise

Os fluxos de trabalho de gerenciamento de incidentes não terminam quando a poeira abaixa e os sistemas são restaurados. Agora começa uma das fases mais importantes do ciclo de vida do gerenciamento de incidentes: a análise. A intenção de uma “autópsia” é entender claramente as causas sistêmicas de um incidente, juntamente às medidas tomadas para responder a ela.

A partir daqui, as equipes líderes trabalham para identificar oportunidades de melhoria em torno dos sistemas e dos processos definidos para mantê-los. Ao avaliar essas informações, as equipes podem desenvolver novos fluxos de trabalho que ofereçam resiliência e resposta a incidentes mais eficiente.

> “A análise pós-incidente deve ser escrita em forma de relatório para fornecer uma revisão passo a passo de todo o incidente; esse relatório deve ser capaz de responder: quem, o quê, onde, por quê e como. São perguntas que podem surgir durante as lições passadas. O objetivo geral é aprender com os incidentes que ocorreram dentro de uma empresa, a fim de melhorar o desempenho da equipe e fornecer materiais de referência em caso de um incidente similar.”
>
> INSTITUTO SANS

Ideias para melhoria

- **Aprenda com o fracasso.**

Esmagadoramente, as equipes de TI dirão que só perdem tempo para revisar o que chamam de “grandes interrupções”. Embora este seja um bom começo, muitas vezes incidentes que podem ter um impacto persistente são negligenciados. Um relatório de “autópsia” detalhado pode não ser necessário para todos os incidentes, mas uma breve revisão dos detalhes deve sempre ser feita. Desta forma, a consciência de uma situação oferece o avanço do conhecimento comunitário e melhoria contínua.

- **Não existe a causa “raiz”!**

Ou será que existe? Ao analisar um incidente, é raro que uma única causa seja identificada. Segundo o modelo Cynefin, estes se enquadram na categoria de incidentes “simples”, em que a causa e a necessidade de resposta são conhecidas e replicáveis. Mas nunca é assim tão fácil. Frequentemente, os sistemas são muito complexos e interdependentes para definir uma única causa raiz de um incidente. Mesmo que a causa principal pareça aparente (por exemplo, um erro de digitação que falha um aplicativo), normalmente há motivos para entender que fatores externos podem ter permitido que o aplicativo travasse (ou não impedido que acontecesse).

- **Não distribua a culpa.**

O objetivo de cada “autópsia” deve ser entender o que deu errado e o que pode ser feito para evitar incidentes semelhantes no futuro. Importante: esse processo não deve ser usado para atribuir culpas. Isso porque equipes que focam no “quem” e não no “o quê”, deixam as emoções puxarem a análise para longe, e ficam sem entender verdadeiramente o que aconteceu.

Em resumo

Nos ambientes modernos de TI, a mudança é a única constante. Isso significa que os sistemas serão continuamente modificados de maneiras novas e diferentes. Equipes que entendem isso também entendem que não é uma questão de se, mas quando os sistemas falharão. Tomar medidas para se preparar para essas falhas deve ser reconhecido como um elemento crítico do sucesso contínuo e integrado ao DNA das equipes de engenharia.

Sobre o Opsgenie

Opsgenie é uma moderna plataforma de gerenciamento de incidentes para serviços “always-on”, capacitando as equipes de Dev & Ops a planejar interrupções de serviço e no controle durante incidentes. Com mais de 200 integrações profundas e um mecanismo de regras altamente flexível, o Opsgenie centraliza alertas, notifica as pessoas certas de forma confiável e permite que elas colaborem e adotem ações rápidas. Ao longo de todo o ciclo de vida do incidente, o Opsgenie rastreia todas as atividades e fornece insights para melhorar a produtividade e gerar mudanças operacionais contínuas.

Veja o Opsgenie em ação.
Comece gratuitamente hoje!

Recursos

Alerta e Gerenciamento de Incidentes:

- Suportando propriedades de alerta personalizadas
- Como melhorar a colaboração durante um incidente
- 5 problemas comuns de resposta a incidentes (e suas soluções)

Cynefin

- O Framework Cynefin
- O vídeo do Framework Cynefin

ChatOps

- Slack > vídeo de integração da Opsgenie
- ChatOps e Hubot no GitHub

Engenharia do Caos

- http://principlesofchaos.org/
- Chaos Monkey

Análise pós-incidente

- Acompanhamento de incidentes com o Opsgenie

ATLASSIAN

Opsgenie