Investir na engenharia do caos como uma necessidade estratégica

AWS Orientação prescritiva

AWS Orientação prescritiva: Investir na engenharia do caos como uma necessidade estratégica

Table of Contents

Investir na engenharia do caos como uma necessidade estratégica

Adrian Hornsby, Amazon Web Services

Janeiro de 2025 (histórico do documento)

As práticas de engenharia do caos usam interrupções controladas para identificar problemas e oportunidades do sistema para evitar interrupções e outros incidentes. A engenharia do caos tornou-se essencial para melhorar sistemas resilientes, mas a adoção generalizada enfrenta obstáculos relacionados a equívocos, resistência cultural, recursos e como quantificar o valor comercial. Definir metas iniciais ajuda a iniciar os esforços de engenharia do caos, enquanto a quantificação do retorno sobre o investimento (ROI) justifica o investimento contínuo, especialmente em meio a pressões econômicas.

Este documento estratégico descreve uma abordagem holística para capturar tanto as melhorias operacionais quantitativas quanto os benefícios organizacionais qualitativos. O objetivo final é tratar a engenharia do caos como uma necessidade estratégica semelhante à cibersegurança e não como um exercício contínuo de justificação de custos.

Custos de inatividade e o surgimento da engenharia do caos

A Consultoria de Inteligência em Tecnologia da Informação (ITIC) estima que 90 por cento das empresas enfrentam custos superiores a 300 mil dólares por hora de inatividade, com 41 por cento excedendo de 1 a 5 milhões de dólares por hora. Além da perda imediata de receita, o tempo de inatividade pode levar a problemas de longo prazo, incluindo falhas de conformidade, preços de ações reduzidos, custos significativos de mitigação e até danos à marca.

Embora o tempo de inatividade seja comumente associado a sistemas on-line geradores de receita, o impacto negativo vai muito além disso. Todas as grandes empresas e organizações, independentemente de seu modelo de receita principal, dependem criticamente da disponibilidade de seus sistemas internos, como RH e folha de pagamento.

O tempo de inatividade que afeta esses principais serviços internos pode inibir a capacidade de funcionamento de uma empresa, levando a interrupções operacionais substanciais e repercussões financeiras. Os problemas resultantes podem incluir o seguinte:

- Atrasos no pagamento de funcionários e fornecedores
- Incapacidade de processar pedidos ou transações de clientes
- Violações de dados confidenciais permitidas por sistemas de segurança comprometidos
- Perda de produtividade e oportunidades de receita
- Penalidades regulatórias por não conformidade
- Danos à reputação da marca

A engenharia do caos introduz intencionalmente interrupções controladas. Usar a engenharia do caos para entender ou verificar a resposta do sistema às deficiências tornou-se uma prática crítica para melhorar a resiliência do sistema. A engenharia do caos permite que sua organização descubra problemas de forma proativa, valide mecanismos de resiliência e, por fim, reduza o risco de tempo de inatividade não planejado e os custos associados. Os benefícios da engenharia do caos incluem o seguinte:

- Expondo a dívida técnica
- Exercício dos músculos operacionais
- Construindo confiança nos sistemas
- Identificação de pontos de falha

- Melhorando o monitoramento e a observabilidade
- Apoiando o aprendizado baseado em experimentos
- Oferecendo maior resiliência para reduzir o tempo de inatividade

À medida que os sistemas se tornam mais complexos e as expectativas dos clientes aumentam, a importância da engenharia do caos aumenta. A Gartner recomenda a engenharia do caos como uma prática essencial para que as organizações reduzam o tempo de inatividade não planejado e melhorem a resiliência.

Os desafios da adoção da engenharia do caos

Embora a engenharia do caos seja uma prática cada vez mais importante para melhorar a resiliência do sistema, sua adoção pode enfrentar os seguintes obstáculos:

- Percepções errôneas sobre risco – Uma percepção errônea comum é que a engenharia do caos é conduzida somente em ambientes de produção, o que gera preocupações com riscos excessivos. Essa percepção decorre da falta de compreensão sobre a natureza sistemática e controlada das práticas de engenharia do caos. Conforme observado no AWS Well-Architected Framework, conduza primeiro a simulação de falhas em um ambiente que não seja de produção.
- Valor comercial a longo prazo – Os benefícios da engenharia do caos se acumulam gradualmente, dificultando a quantificação do valor comercial e a justificação do investimento inicial. O ROI mais lento dificulta que as organizações priorizem e continuem com a engenharia do caos.
- Lacunas de habilidades e conhecimentos – A engenharia do caos exige um conjunto exclusivo de habilidades e conhecimentos que podem não estar prontamente disponíveis em sua organização. Construir ou adquirir essa experiência pode ser uma barreira significativa, especialmente para organizações que são novas na prática e aquelas com recursos limitados.

O restante deste documento estratégico se concentrará principalmente no segundo desafio, que é demonstrar o valor comercial da engenharia do caos.

Os efeitos acumulados da engenharia do caos

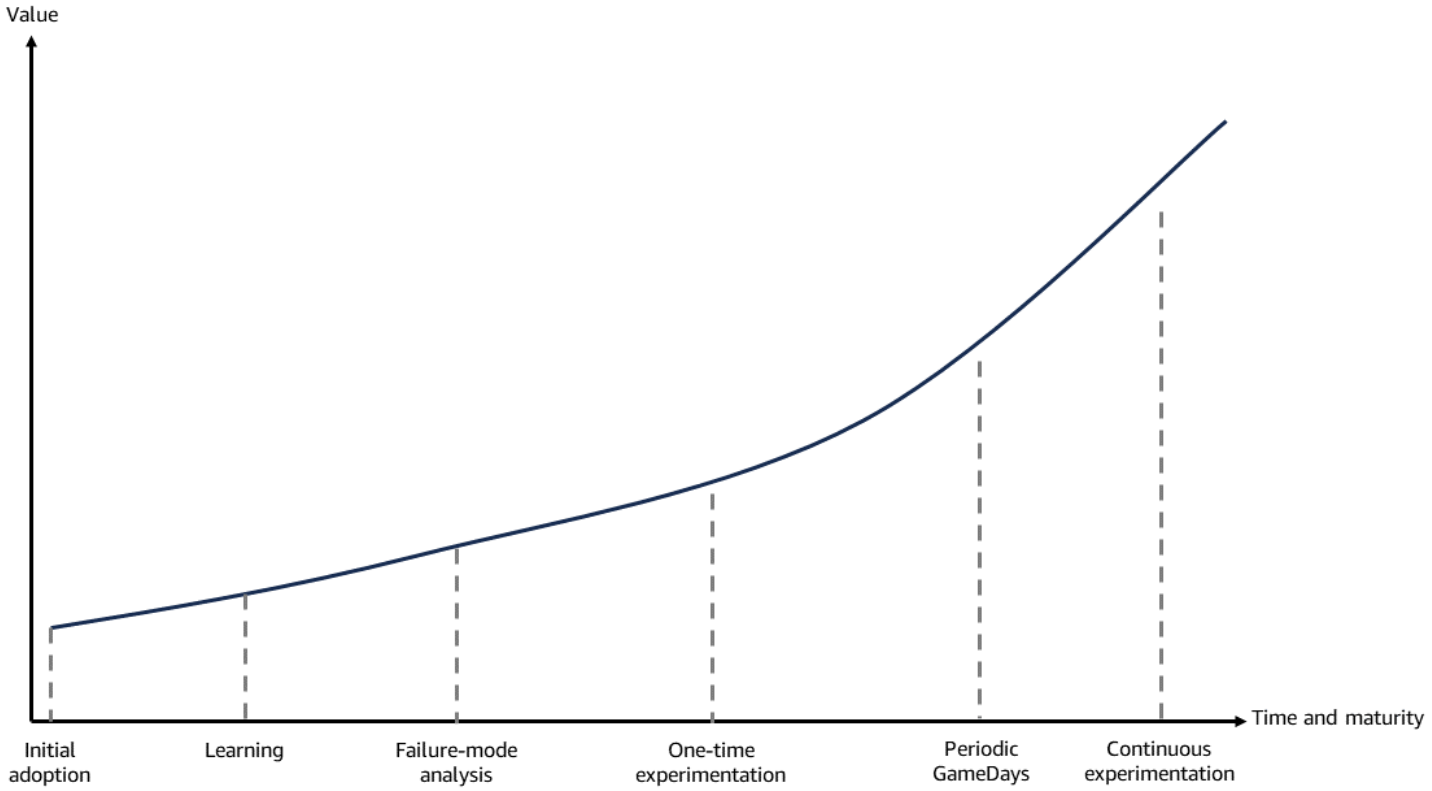
Diferentemente dos projetos de tecnologia tradicionais com datas de início e término bem definidas, a engenharia do caos é uma prática contínua de aprendizado contínuo e melhorias contínuas na resiliência do sistema. Os benefícios da engenharia do caos aumentam com o tempo.

À medida que os sistemas evoluem e se tornam mais complexos, surgem novos modos de falha. São necessários mais experimentos de caos para identificar possíveis problemas. A solução de um problema pode levar meses, especialmente em grandes empresas com sistemas e processos complexos ou quando as falhas são de propriedade de provedores de serviços externos.

A mudança cultural para encarar o fracasso como uma oportunidade de aprendizado e aprimoramento cresce com o passar dos anos e se torna arraigada na organização. Os investimentos na automação de experimentos de engenharia do caos e no desenvolvimento de ferramentas de apoio continuam a simplificar e aprimorar a prática da engenharia do caos. Construir esse conhecimento institucional e entender a resiliência do sistema é um processo gradual que se acumula ao longo do tempo. O conhecimento, os processos e as ferramentas desenvolvidos por meio da engenharia do caos aumentam de valor à medida que a prática amadurece junto com os sistemas em constante evolução.

O diagrama a seguir mostra como o valor aumenta com o tempo à medida que a adoção do caos progride nos seguintes estágios:

- Adoção inicial
- Aprendendo
- Análise do modo de falha
- Experimentos únicos
- Periódico GameDays
- Experimentação contínua



Conforme mostrado no diagrama, os benefícios da engenharia do caos geralmente começam antes que qualquer falha seja injetada no sistema. O processo de planejar e projetar experimentos de caos em si fornece valor imediato. Identificar possíveis cenários de falha, pontos únicos de falha e áreas de incerteza no sistema leva a melhorias.

Por exemplo, escrever cenários de falha e discutir os possíveis efeitos em cascata, um processo chamado modo de falha e análise de efeitos (FMEA), ajuda a descobrir fraquezas ou lacunas óbvias que podem ter sido negligenciadas. Sua organização pode resolver esses problemas de forma proativa, mesmo antes de sujeitar o sistema a qualquer interrupção intencional. Para obter mais informações, consulte a estrutura de análise de resiliência.

Além disso, o maior foco na observabilidade e monitoramento do sistema, que geralmente acompanha as iniciativas de engenharia do caos, começa a gerar benefícios imediatamente. Melhorar a visibilidade do comportamento do sistema e dos modos de falha ajuda a equipe a entender melhor as condições operacionais normais do sistema. Uma maior visibilidade também ajuda a equipe a entender como as condições operacionais se degradam, se adaptam e falham quando levadas ao limite.

Tanto o experimento único quanto o GameDay modo periódico são abordagens mais manuais em comparação com o modo de experimentação contínua. Eles exigem um processo mais prático e

exploratório, no qual os engenheiros moldam e refinam ativamente as hipóteses por meio de suas observações e experimentos.

O modo de experimentação contínua é, por outro lado, mais automatizado por natureza. Esse modo se concentra na execução de hipóteses aprovadas e validadas de forma controlada e iterativa. Ele usa automação e integração no processo de desenvolvimento por meio de um pipeline de caos dedicado para ajudar a garantir experimentos consistentes e repetíveis.

Iniciativas de engenharia do caos de base

A jornada da engenharia do caos geralmente começa no nível básico, onde as equipes de engenharia identificam as necessidades e começam a experimentar a engenharia do caos de forma independente.

Nessa abordagem de base, as equipes experimentam, aprendem e refinam suas práticas de engenharia do caos. O valor da engenharia do caos pode ser demonstrado por meio dos seguintes resultados tangíveis:

- Incidentes reduzidos
- Melhor observabilidade
- Tempos de recuperação mais rápidos
- Resiliência aprimorada e contínua do sistema

As iniciativas de engenharia do caos de base geralmente surgem sob condições organizacionais específicas. Eles exigem um ambiente com alto grau de autonomia de engenharia, onde as equipes tenham a liberdade de experimentar e inovar sem barreiras burocráticas excessivas. A experiência local em engenharia de resiliência ou sistemas distribuídos é crucial, pois fornece a base técnica para entender e implementar experimentos de caos. Mais importante ainda, essas iniciativas geralmente contam com campeões do caos — indivíduos apaixonados que entendem o valor da engenharia do caos. Os campeões do caos estão dispostos a defender a adoção da engenharia do caos, educar seus colegas e conduzir experimentos iniciais. Sem liberdade organizacional, conhecimento técnico e campeões motivados, os esforços populares de engenharia do caos raramente se enraízam, independentemente de seus benefícios potenciais.

O papel das metas na adoção da engenharia do caos

É comum que as metas iniciais surjam organicamente dos esforços de engenharia do caos de base dentro de uma organização. Impulsionadas pela necessidade de resolver seus próprios problemas recorrentes, essas equipes ou grupos geralmente exploram práticas de engenharia do caos sem aprovação explícita ou priorização de níveis mais altos.

As equipes podem usar esses resultados para criar um argumento convincente para uma adoção organizacional mais ampla, tornando-se efetivamente um campo de provas para outras equipes.

Depois que os benefícios dos esforços de base se tornam significativos demais para serem ignorados, essas equipes podem elevar seus esforços e conhecimentos à liderança e estabelecer metas. Essa maior visibilidade pode facilitar a adoção de objetivos de resiliência em toda a organização e levar ao suporte e aos recursos necessários para a implementação da engenharia do caos.

As metas, especialmente aquelas orientadas pela liderança e estabelecidas em resposta a interrupções significativas, desempenham um papel crucial na catalisação da adoção de práticas de engenharia do caos. Os tipos comuns de metas incluem o seguinte:

- Metas de disponibilidade para identificar e reduzir pontos únicos de falha (SPOF)
- Metas de recuperação de serviços para melhorar a capacidade de recuperação de interrupções ou falhas
- Metas de experiência do usuário para atender aos objetivos específicos de nível de serviço () SLOs
- Metas orientadas por métricas para monitorar o progresso na mitigação dos riscos de disponibilidade conhecidos e na implementação de medidas de resiliência recomendadas
- Metas regulatórias e de conformidade para demonstrar resiliência operacional

Para obter mais informações sobre alguns desses tipos de metas e como a Amazon e outras organizações usaram metas durante a adoção da engenharia do caos, consulte o Apêndice A.

Essas metas servem como uma justificativa convincente e fornecem uma abordagem direcionada e acionável para impulsionar a adoção da engenharia do caos. No início, as metas servem como um indicador das métricas tradicionais de ROI. As metas oferecem uma justificativa convincente quando cálculos quantificáveis de ROI de resiliência podem ser difíceis de obter. Sem essas metas logo no

início da adoção, a prática de engenharia do caos corre o risco de não demonstrar sua eficácia e obter uma adesão organizacional mais ampla.

A mudança das metas para a medição do ROI

À medida que as práticas amadurecem e as metas iniciais são alcançadas, o foco acaba mudando do estabelecimento de metas para a quantificação dos benefícios financeiros tangíveis da engenharia do caos — o retorno sobre o investimento (ROI). A mudança decorre de dois motivos principais:

- Considerações econômicas
- Preservando a experiência e a confiança do cliente

Considerações econômicas

Em tempos de crescimento econômico e finanças saudáveis, as empresas geralmente não precisam de uma justificativa extensa para definir metas específicas para estratégias de engenharia do caos. No entanto, mudanças no cenário financeiro levaram muitas organizações a reavaliar seus investimentos, e as implementações de engenharia do caos precisam fornecer ROI quantificado.

Essas empresas agora têm a tarefa de definir métricas claras e tradicionais de ROI para demonstrar o valor e o impacto das práticas de engenharia do caos. Esse desafio é ainda mais complicado pelo paradoxo da prevenção. O paradoxo da prevenção ocorre quando a prevenção bem-sucedida de incidentes torna mais difícil justificar o investimento, porque as partes interessadas tendem a subestimar as catástrofes evitadas. Até mesmo organizações com uma cultura profundamente arraigada de excelência operacional enfrentam pressão para usar métricas de ROI para justificar a adoção contínua da engenharia do caos.

Preservando a experiência e a confiança do cliente

Manter a resiliência orientada por metas pode ser um desafio a longo prazo. Depois que uma meta inicial, como atingir uma meta de tempo de recuperação, é atingida, justificar o investimento contínuo em engenharia de caos se torna difícil até a próxima grande interrupção. O fluxo e o refluxo do investimento criam um ciclo reativo de dentes de serra. Para cada nova interrupção, o investimento em resiliência aumenta com uma nova meta que aborda a causa raiz. Depois que a nova meta é atingida, o investimento cai até o próximo incidente, reiniciando o ciclo reativo.

As interrupções que impulsionam essa abordagem reativa afetam negativamente os clientes. A questão-chave: quantas grandes interrupções os clientes tolerarão antes de abandonarem um provedor de serviços em favor de um concorrente mais resiliente?

Quantificando o ROI da engenharia do caos

Atualmente, poucos recursos publicados fornecem metodologias abrangentes ou dados do mundo real para quantificar o retorno sobre o investimento (ROI) de longo prazo para a engenharia do caos.

No paper The Business Case for Chaos Engineering, a Netflix oferece uma equação valiosa para calcular o ROI da engenharia do caos. Essa equação fornece um ponto de partida para as organizações que embarcam em sua jornada de engenharia do caos.

A equação exige que você estime com precisão os custos do seguinte:

- Interrupções evitáveis e não evitáveis
- Custos de implementação do programa de engenharia de caos
- Custos de danos induzidos pelo caos

O dano induzido pelo caos se refere ao impacto negativo ou à interrupção causada pela injeção deliberada de falhas ou condições turbulentas em um sistema como parte de experimentos de engenharia do caos. A equação exige estimar os custos de interrupções evitáveis e não evitáveis, os custos de implementação do programa de engenharia do caos e os custos de danos induzidos pelo caos.

Determinar com certeza quais problemas poderiam ter sido evitados por um programa de engenharia do caos é uma tarefa difícil. Isso requer uma análise hipotética que envolva examinar as causas básicas dos problemas e especular como os experimentos de engenharia do caos podem ter ajudado a identificá-los. Essa análise é desafiadora porque os sistemas modernos são altamente complexos, com inúmeras interdependências e interações entre vários componentes, serviços e bibliotecas de terceiros. Além disso, as falhas nos sistemas geralmente não são determinísticas, e as condições que causam falhas podem ser difíceis de entender completamente em retrospectiva.

Embora a abordagem sugerida pela Netflix tenha algumas limitações, ela serve como uma boa base para organizações começarem a explorar a engenharia do caos. A equação pode orientá-lo na estimativa de custos e benefícios potenciais, o que ajuda você a tomar decisões sobre a implementação desse programa. No entanto, à medida que as organizações avançam em sua jornada de engenharia do caos, é importante expandir a avaliação do ROI para incorporar uma perspectiva mais holística.

Essa abordagem holística não apenas capturará os benefícios diretos da redução de interrupções e custos de engenharia, mas também destacará os efeitos transformadores de longo prazo na

resiliência geral da organização. Ele captura os benefícios combinados e os efeitos organizacionais mais amplos da engenharia do caos para fornecer uma representação mais precisa do verdadeiro valor e impacto da engenharia do caos.

Uma abordagem holística para quantificação do ROI

Uma avaliação holística do ROI deve levar em conta não apenas as medidas quantitativas, mas também os fatores qualitativos. A abordagem holística requer dados do mundo real de organizações que praticam engenharia do caos em grande escala por longos períodos de tempo. Você pode usar dados a partir dos projetos e metas de base por meio de quaisquer dados de ROI de abordagem de equação que você coletou.

As medidas quantitativas se concentram em quantidades ou frequências. As medidas são objetivas e podem ser analisadas estatisticamente. Os exemplos incluem pesquisas, experimentos e análise de dados. As medidas quantitativas podem incluir o seguinte:

- Métricas de incidentes
- Custos
- Melhorias
- Compliance
- Taxas de adoção
- Satisfação de cliente

O rastreamento de medidas quantitativas pode demonstrar os benefícios operacionais diretos da engenharia do caos.

As medidas qualitativas são descritivas e se concentram na compreensão de experiências e opiniões. Eles geralmente são subjetivos e não podem ser facilmente medidos numericamente. Para a engenharia do caos, as medidas qualitativas capturam os impactos organizacionais mais amplos. As medidas qualitativas podem incluir o seguinte:

- Confiança dos funcionários
- Mudança cultural
- Colaboração
- Eficácia do treinamento
- Retenção de talentos

- Reputação da marca
- Vantagem competitiva

Ao considerar os impactos financeiros quantitativos e os benefícios organizacionais qualitativos, você pode tomar decisões mais informadas sobre o investimento contínuo em engenharia do caos e, ao mesmo tempo, promover uma cultura de resiliência.

Para obter mais informações sobre essas medidas e sua estrutura de classificação de incidentes associada, consulte o Apêndice B e o Apêndice C.

A transição do ROI para a engenharia do caos como uma necessidade estratégica

Embora seja tentador monitorar o ROI, os desafios em medir o valor da engenharia do caos geralmente levam as organizações a priorizar eficiências imediatas e de curto prazo em relação aos investimentos estratégicos em resiliência. Essa abordagem ignora a engenharia do caos como um dos principais impulsionadores da resiliência e as vantagens competitivas de evitar interrupções. O valor real da engenharia do caos está em evitar futuras falhas. A engenharia do caos oferece suporte à continuidade dos negócios a longo prazo.

Em vez de focar no ROI, trate a engenharia do caos como cibersegurança. Conforme explicado no artigo da Forbes Cibersegurança como investimento estratégico: como a otimização do ROI pode levar a um futuro mais seguro, a cibersegurança não deve ser vista como um centro de custos ou despesa obrigatória para as organizações porque essa mentalidade não reconhece o valor estratégico que medidas robustas de cibersegurança podem fornecer ao longo do tempo. Em vez disso, o autor argumenta que, ao mudar as perspectivas para tratar a segurança cibernética como um investimento de longo prazo que gera vantagens competitivas, as organizações podem abrir novos caminhos para inovação, eficiência operacional e diferenciação em seus respectivos mercados. Ao adotar essa abordagem, o autor conclui que os Diretores de Segurança da Informação (CISOs) podem garantir melhor a adesão e o financiamento da liderança. Eles podem então posicionar suas empresas para superar os concorrentes em um cenário cibernético cada vez mais arriscado. Essa criação de valor estratégico e de longo prazo da segurança cibernética é paralela às melhorias contínuas inerentes às práticas de engenharia do caos.

Enquanto a segurança protege a capacidade de uma organização de operar e proteger ativos, a engenharia do caos ajuda a garantir a disponibilidade, a confiabilidade e a capacidade de recuperação dos principais sistemas e serviços. Para obter valor a longo prazo e vantagem competitiva, trate a engenharia do caos como uma capacidade essencial e um imperativo estratégico, não como uma iniciativa que exige justificativa constante.

O diagrama a seguir mostra a evolução da engenharia do caos desde a base até as metas e o ROI, até se tornar uma estratégia.

No nível de base, as equipes individuais geralmente experimentam de forma independente, orientadas pelas necessidades locais. Esses experimentos são promovidos por engenheiros apaixonados que demonstram valor por meio da redução de incidentes e da melhoria da observabilidade.

Quando esses esforços são bem-sucedidos, as equipes podem elevar seu aprendizado à liderança. Com essa visibilidade, os esforços passam para uma fase orientada por metas. A organização define objetivos formais de resiliência e recuperação, apoiados por recursos e suporte para uma implementação mais ampla.

Por fim, a engenharia do caos amadurece além de exigir uma justificativa constante de ROI para ser reconhecida como uma necessidade estratégica, semelhante à segurança cibernética. Nesse estágio, a engenharia do caos se torna totalmente integrada aos processos organizacionais. A implementação se concentra na resiliência de longo prazo, em vez de métricas de curto prazo. A engenharia do caos é tratada como uma capacidade essencial para manter a vantagem competitiva e a confiança do cliente.

Integrando a engenharia do caos em sua organização

Para elevar a engenharia do caos ao mesmo nível de importância da segurança, considere as seguintes sugestões:

- Estabeleça a engenharia do caos como uma prática inegociável – Assim como a segurança cibernética é considerada um requisito fundamental para as organizações, veja a engenharia do caos como uma prática obrigatória para garantir a resiliência e a confiabilidade do sistema. Integre a engenharia do caos aos processos, ferramentas e cultura da sua organização, em vez de considerá-la uma atividade opcional ou discricionária. Para obter mais informações, consulte o guia da estrutura do ciclo de vida de resiliência.
- Apoio e suporte seguros em nível executivo – Assim como acontece com as iniciativas de segurança, os esforços de engenharia do caos devem contar com a adesão e o apoio ativo da liderança executiva. Isso inclui alocar recursos, orçamento e pessoal dedicados para implementar e manter práticas de engenharia do caos em toda a organização.

- Implemente governança e supervisão – Semelhante a uma estrutura de CISO e governança de segurança, estabeleça uma equipe dedicada de engenharia do caos ou um diretor de resiliência. Essa equipe ou função é responsável por supervisionar e coordenar os esforços de engenharia do caos em diferentes equipes e unidades de negócios.
- Integre a engenharia do caos aos ciclos de desenvolvimento e operações – Assim como as práticas de segurança são integradas aos processos de desenvolvimento e implantação de software, faça da engenharia do caos uma parte perfeita do ciclo de vida de desenvolvimento e entrega de software.
- Conduza exercícios e simulações regulares de engenharia do caos – Semelhante às simulações de violação de segurança e exercícios de resposta a incidentes, conduza experimentos regulares de engenharia do caos para validar as capacidades de resposta a incidentes e identificar possíveis pontos cegos de forma proativa.
- Use a engenharia do caos para manter os runbooks – Assim como na realização de análises de segurança, use experimentos de engenharia do caos para validar a eficácia e a precisão dos runbooks para resposta e recuperação de incidentes. Além disso, os experimentos de engenharia do caos podem servir como simulações realistas para engenheiros de plantão praticarem a execução de procedimentos do runbook. As simulações ajudam os engenheiros a manter sua memória muscular operacional e sua preparação para lidar com incidentes do mundo real.
- Promova uma cultura de resiliência – Assim como no treinamento de conscientização sobre segurança, invista em educação sobre engenharia do caos e iniciativas de compartilhamento de conhecimento para promover uma cultura de resiliência. Inclua programas de treinamento, colaboração interfuncional e incentivos para equipes que adotam práticas de engenharia do caos.
- Meça e relate as métricas de resiliência – Monitore regularmente as métricas de resiliência e informe-as às partes interessadas. Use as métricas quantitativas e qualitativas discutidas neste documento como ponto de partida.
- Trate a resiliência como uma vantagem competitiva – As medidas de cibersegurança podem fornecer uma vantagem competitiva. Da mesma forma, veja seus recursos de engenharia de caos e resiliência como um diferencial que ajuda você a oferecer serviços mais confiáveis e confiáveis aos seus clientes.

Obter a adesão executiva

A engenharia do caos geralmente carece de um proprietário claro das responsabilidades tradicionais da diretoria. O CEO se preocupa com crescimento, lucratividade e liderança de mercado. O CFO se concentra no desempenho financeiro, controle de custos e gerenciamento de riscos. O CTO

prioriza a estratégia de tecnologia, os roteiros de produtos e a excelência em engenharia. O CISO supervisiona a segurança e a conformidade.

Sem um único executivo que realmente possua resiliência, muitas vezes é difícil obter adesão e apoio. No entanto, as falhas do sistema afetam a receita, a satisfação do cliente e a reputação da marca, que são preocupações do CEO e do CFO. O CTO e o CISO têm a tarefa de implementar medidas de resiliência, mas podem não ter mandato organizacional. Essa ambigüidade pode atrapalhar a realização de investimentos estratégicos e o alinhamento da organização em direção a uma estratégia comum de resiliência.

Essa ambigüidade também dificulta a adesão de executivos a iniciativas de resiliência, como a engenharia do caos. Afinal, os líderes de nível C estão lidando com uma infinidade de prioridades estratégicas: crescimento, inovação, experiência do cliente, conformidade e muito mais.

Para comunicar com eficácia o valor da engenharia do caos aos executivos de nível C, considere as seguintes abordagens:

- Determine as principais preocupações e os fatores de decisão de seus executivos de alto escalão.

  Por exemplo, os executivos da diretoria estão preocupados com a rotatividade de clientes, a conformidade regulatória, a redução de custos ou as pressões competitivas? Posicione a engenharia do caos como um multiplicador de forças que se alinha aos desafios e metas exclusivos da empresa.
- Identifique objetivos compartilhados e resultados estratégicos.

  Como sua estratégia de engenharia do caos apoia a estratégia geral de crescimento, a experiência do cliente, as oportunidades de mercado e a eficiência operacional da organização? Priorize as iniciativas com base nas metas, no impacto nos negócios, no ROI e no risco de não realizar as iniciativas.
- Comunique a eficácia de sua estratégia de engenharia do caos em termos quantificáveis usando os principais indicadores de resiliência.

  Comece com esses quatro indicadores-chave de resiliência: disponibilidade, tempo para detectar, tempo para responder e tempo para se recuperar. Vincule isso diretamente aos resultados comerciais, como receita, economia de custos e reputação da marca.
- Não se perca nos detalhes técnicos.

Concentre-se no sentimento geral e no impacto mensurável nos negócios. A diretoria se preocupa com os resultados que impulsionam o crescimento, aumentam a confiança do cliente e promovem a inovação.

O paradoxo da prevenção

Quando as falhas são mitigadas com sucesso antes de se manifestarem, torna-se difícil convencer as partes interessadas do valor e da necessidade das medidas preventivas tomadas. Esse fenômeno é conhecido como paradoxo da prevenção. O paradoxo da prevenção é o maior obstáculo para integrar a engenharia do caos como uma necessidade estratégica e decorre dos preconceitos inerentes à cognição humana.

O bug do Y2K serve como uma ótima ilustração desse paradoxo. Anos de preparação e bilhões de dólares foram investidos na atualização de sistemas de computador em todo o mundo. No entanto, a transição suave para 2000 foi interpretada por muitos como uma prova da natureza exagerada das preocupações do Y2K. O sucesso dos esforços preventivos realizados raramente foi reconhecido.

Esse paradoxo da prevenção continua desafiando as organizações que investem na engenharia do caos atualmente. Quando possíveis interrupções são evitadas com sucesso por meio de medidas proativas, a própria ausência de catástrofe pode, paradoxalmente, dificultar a justificação dos recursos gastos na prevenção.

A causa raiz desse fenômeno está na forma como nossas mentes são programadas para processar informações. Os processos cognitivos humanos são voltados para responder e lembrar eventos reais e resultados visíveis. Quando um desastre é evitado, não há uma narrativa dramática para guardar ou compartilhar. Outro aspecto do paradoxo da prevenção é o viés retrospectivo. Depois de um não-evento, as pessoas tendem a concluir que nada aconteceu, então não foi um problema real. A possibilidade de que as precauções apropriadas tenham evitado um problema real não é reconhecida. Esse ponto cego psicológico cria um desafio perpétuo para as organizações. Quanto mais bem-sucedido você for em prevenção e resiliência, mais seus esforços parecerão desnecessários em retrospectiva.

Para lidar com o paradoxo da prevenção, sua organização pode tomar medidas específicas para tornar visível, mensurável e valorizado o trabalho invisível da prevenção. As etapas possíveis incluem o seguinte:

- Documente e simule o que poderia ter acontecido sem medidas preventivas.

- Compartilhe histórias de eventos nos quais medidas preventivas evitaram possíveis desastres.
- Indique organizações semelhantes que não se prepararam e que sofreram consequências como resultado.
- Apresente os custos de prevenção no contexto dos impactos potenciais que eles estão prevenindo.
- Divida os esforços de prevenção em marcos e conquistas visíveis.
- Construa uma memória institucional sobre por que existem medidas preventivas e sua importância histórica.
- Eduque regularmente as partes interessadas sobre o valor das práticas de resiliência e engenharia do caos.

Conclusão

A engenharia do caos é um imperativo estratégico para as organizações. Embora sua jornada de adoção possa enfrentar desafios como equívocos, resistência cultural e restrições de recursos, estabelecer metas claras e orientadas pela liderança pode catalisar o processo. À medida que as práticas amadurecem, quantifique o retorno sobre o investimento por meio de uma abordagem holística que captura tanto melhorias operacionais quantitativas quanto benefícios organizacionais qualitativos. A abordagem holística é especialmente importante durante as pressões econômicas.

Para transformar essa necessidade estratégica em realidade, comece avaliando o nível atual de maturidade da sua organização. Sua organização está no estágio de experimentação de base, na fase orientada por metas ou em algum lugar intermediário? Com base nessa avaliação, crie um roteiro personalizado para realizar o seguinte:

- Estabeleça a governança da engenharia do caos (por exemplo, nomeie um diretor de resiliência).
- Integre práticas de caos aos fluxos de trabalho de desenvolvimento.
- Implemente programas de treinamento regulares.
- Desenvolva métricas abrangentes de resiliência.

Essa transformação não acontecerá da noite para o dia. No entanto, tomar essas medidas concretas e, ao mesmo tempo, garantir o apoio executivo contínuo, ajudará a elevar a engenharia do caos ao mesmo nível estratégico da segurança cibernética. Semelhante à cibersegurança, a engenharia do caos pode se tornar parte integrante do DNA e dos processos operacionais da sua organização.

Recursos

- Resultados da pesquisa global de confiabilidade de hardware e sistema operacional de servidor ITIC 2021
- O caso de negócios da engenharia do caos
- Segurança cibernética como investimento estratégico: como a otimização do ROI pode levar a um futuro mais seguro
- O Guia do Líder de I&O para Engenharia do Caos
- Como usar a pontuação do AWS Resilience Hub
- Implementando experimentos recomendados usando o console do AWS Resilience Hub

Apêndice A – Tipos de metas para engenharia do caos

As seguintes descrições dos tipos de metas incluem exemplos reais de como a Amazon e outras organizações criaram metas para a engenharia do caos.

Metas de arquitetura resiliente

Um dos fatores iniciais para a adoção da engenharia do caos é identificar e reduzir pontos únicos de falha (SPOF) em sistemas e infraestrutura. As metas são definidas para validar a resiliência de sistemas e arquiteturas essenciais, especialmente para novos serviços ou aplicativos.

As metas de arquitetura resiliente envolvem a execução de experimentos de caos que simulam falhas nas dependências do serviço. Os experimentos confirmam se os tempos limite, as novas tentativas, o comportamento do cache e as configurações do disjuntor estão funcionando corretamente. Esses experimentos ajudam a descobrir problemas para remediação, evitando incidentes que afetem o cliente. Por exemplo, consulte Criação de serviços resilientes no Prime Video com engenharia do caos.

Metas de recuperação de serviços

As metas de recuperação de serviços se concentram em melhorar a capacidade de recuperação de interrupções operacionais ou falhas na infraestrutura. Por exemplo, sua organização pode ter como objetivo atingir um objetivo de tempo de recuperação (RTO) específico para seus serviços principais no caso de uma interrupção. As equipes podem criar experimentos de caos para validar e otimizar estratégias de evacuação, mecanismos de failover e processos de recuperação automatizados. Em última análise, as otimizações reduzem o tempo necessário para a restauração do serviço. Para obter um exemplo, consulte AWS Lambda: Resiliência. under-the-hood

Metas de experiência do usuário

Manter uma experiência de usuário consistente e confiável é fundamental, especialmente durante períodos de alto tráfego ou eventos críticos. Nesses casos, defina metas centradas no cumprimento de objetivos específicos de nível de serviço (). SLOs Essa abordagem centrada no cliente garante que os esforços de resiliência estejam diretamente alinhados com a entrega de uma experiência superior ao usuário, mesmo em face de falhas ou condições degradadas. Por exemplo, consulte Resiliência de engenharia: lições da jornada de engenharia do caos da Amazon Search.

Metas orientadas por métricas

Você pode estabelecer metas com base em métricas quantitativas, como uma pontuação de resiliência calculada atribuindo pontos aos serviços que adotam as melhores práticas comprovadas de resiliência. Em seguida, você pode usar experimentos de caos específicos para determinar a pontuação de resiliência. Essa pontuação pode servir como uma medida para as equipes acompanharem seu progresso na mitigação dos riscos de disponibilidade conhecidos e na implementação de medidas de resiliência recomendadas. No entanto, é crucial interpretar essas pontuações com cautela e evitar enfatizar demais uma única métrica em detrimento de objetivos mais amplos de resiliência. Por exemplo, consulte Entendendo as pontuações de resiliência.

Metas de conformidade regulatória

O setor de serviços financeiros emergiu como pioneiro na adoção da engenharia do caos, impulsionado principalmente por requisitos regulatórios rigorosos que exigem recursos robustos de resiliência. As regulamentações exigirão que as instituições financeiras identifiquem, testem e corrijam proativamente as vulnerabilidades em seus sistemas e processos críticos. Esses regulamentos incluem o seguinte:

- O documento interinstitucional sobre boas práticas para fortalecer a resiliência operacional emitido por agências federais dos EUA
- As diretrizes do Banco Central Europeu sobre resiliência operacional
- A proposta da Comissão Europeia para uma Lei de Resiliência Operacional Digital (DORA)

Se sua organização for uma instituição financeira, cumpra essas regulamentações definindo metas explícitas para demonstrar resiliência operacional por meio de estratégias abrangentes de testes e validação. Por exemplo, veja o London Stock Exchange Group usa a engenharia do caos AWS para melhorar a resiliência.

Apêndice B – Medidas quantitativas e qualitativas

Esta seção descreve métricas quantitativas para rastrear melhorias operacionais e medidas qualitativas para avaliar resultados organizacionais mais amplos das práticas de engenharia do caos.

Medidas quantitativas

As medidas quantitativas a seguir fornecem uma estrutura para rastrear as principais métricas que podem demonstrar os incidentes diretos e as melhorias operacionais alcançadas por meio de práticas de engenharia do caos:

- Incidentes:
  - Frequência de incidentes – Rastreie o número de incidentes em uma estrutura de classificação de incidentes e classifique-os por sua criticidade (crítica, importante, secundária) durante um período de tempo. Para obter mais informações sobre a estrutura de classificação de incidentes, consulte o Apêndice C.
  - Tempo de inatividade e degradação – Meça a duração total do tempo de inatividade ou da degradação do serviço para cada classificação de incidente.
  - Métricas de resposta a incidentes – Para entender os incidentes, meça o tempo de detecção, o tempo de identificação, o tempo de mitigação, o tempo de recuperação, o tempo de escalonamento e outras métricas relacionadas para cada classificação de incidente.
  - Incidentes que afetam o cliente – Acompanhe o número de incidentes que afetam os clientes ou a porcentagem de incidentes que foram contidos antes do impacto no cliente.
  - Mudanças no runbook – Monitore o número de atualizações ou revisões do runbook resultantes de insights obtidos por meio de experimentos de caos. Um runbook fornece instruções detalhadas para realizar uma operação ou procedimento específico para se recuperar de um determinado tipo de incidente.
- Custos:
  - Custos de infraestrutura – Colete dados sobre os custos de infraestrutura, incluindo recursos de computação em nuvem e medidas de redundância que são exigidas pelas ações tomadas para melhorar a resiliência.
  - Impacto no cliente – Meça os impactos na experiência do cliente, nas taxas de rotatividade e na perda de receita associados a falhas do sistema ou tempo de inatividade.

  - Produtividade da equipe – Monitore o tempo gasto pelas equipes de engenharia e operações na resposta a incidentes, combate a incêndios, redação de autópsias e outras tarefas reativas relacionadas a falhas do sistema.
- Melhorias contínuas do sistema – Conte o número de melhorias de processos, mudanças arquitetônicas ou mecanismos de recuperação automatizados implementados como resultado direto de insights de experimentos de caos.
- Conformidade – Acompanhe os custos e trabalhe para atender aos requisitos regulatórios ou aos padrões do setor relacionados à resiliência operacional.
- Adoção – Acompanhe a taxa de adoção de práticas caóticas em toda a organização.
- Satisfação do cliente – Meça as mudanças nas métricas de satisfação do cliente para avaliar como a maior confiabilidade do sistema afeta os negócios.

Medidas qualitativas

As medidas qualitativas a seguir fornecem uma estrutura para rastrear os resultados organizacionais mais amplos alcançados por meio de práticas de engenharia do caos:

- Confiança e preparação dos funcionários:
  - Pesquise as equipes periodicamente para medir seus níveis de confiança no tratamento de incidentes do mundo real e sua percepção de prontidão para rotações de plantão.
  - Monitore a porcentagem de engenheiros de plantão que participaram de experimentos de caos como parte de seu treinamento.
- Mudança cultural:
  - Avalie o grau em que uma mentalidade de resiliência permeou a organização por meio de pesquisas, sessões de feedback ou auditorias.
  - Monitore o número de equipes que defendem e defendem ativamente as práticas de engenharia do caos.
- Colaboração multifuncional e compartilhamento de conhecimento:
  - Monitore a frequência e a frequência de sessões de compartilhamento de conhecimento entre equipes ou workshops relacionados ao aprendizado de engenharia do caos.
  - Acompanhe o número de iniciativas conjuntas de engenharia do caos envolvendo várias equipes ou departamentos.
- Eficácia do treinamento:

  - Avalie a eficácia dos programas de treinamento em engenharia do caos conduzindo pesquisas ou avaliações pós-treinamento.
  - Monitore o número de engenheiros que participam de programas de treinamento em engenharia do caos e leia autópsias.
- Atração e retenção de talentos:
  - Avalie se o programa de engenharia do caos ajuda a atrair e reter os melhores talentos de engenharia, reduzindo o tempo e o esforço gastos na correção de interrupções.
- Reputação da marca:
  - Acompanhe quaisquer mudanças na percepção ou reputação da marca relacionadas ao comprometimento demonstrado da organização com a resiliência operacional.
- Vantagem competitiva:
  - Acompanhe a vantagem competitiva em relação aos colegas do setor em termos de disponibilidade do sistema.

Apêndice C – Classificação de incidentes

O rastreamento de incidentes em uma estrutura de classificação é crucial porque a estrutura fornece uma visão holística dos tipos de falhas e problemas que afetam o sistema. Se sua organização monitora incidentes somente dentro de uma única classe, como falhas de infraestrutura, você pode perder insights e oportunidades de melhoria em outras áreas. Ao rastrear incidentes em várias classes, você obtém uma melhor compreensão da diversidade de experimentos de caos a serem conduzidos. Essa perspectiva ajuda a identificar possíveis pontos cegos e apoia a expansão do escopo de engenharia, o que leva a um sistema mais resiliente e tolerante a falhas.

A estrutura de classificação de incidentes sugerida foi projetada para ajudar a categorizar os incidentes com base em sua natureza e impacto potencial. Ele usa uma classificação de alto nível que agrupa os incidentes em oito categorias principais:

- Problemas de implantação:
  - Implantações com falha
  - Falhas de reversão
  - Problemas de configuração durante a implantação
- Erros e regressões de software:
  - Erros funcionais
  - Problemas de integração
  - Problemas de desempenho
  - Problemas de cota
  - Problemas do mecanismo de resiliência (novas tentativas, tempos limite)
  - Problemas de integridade de dados
- Problemas de teste:
  - Testes faltantes
  - Testes ineficazes
  - Testes escamosos
- Falhas na infraestrutura:
  - Falhas de hardware (servidores, dispositivos de rede, armazenamento)
  - Problemas de escalabilidade
  - Falhas de dependência (serviços de terceiros) APIs

    - Problemas de conectividade de rede
- Problemas operacionais:
    - Erros humanos (configuração incorreta, alterações acidentais)
    - Falhas de monitoramento e alerta
    - Problemas de planejamento de capacidade
    - Falhas de backup e restauração
- Incidentes de segurança:
    - Tentativas de acesso não autorizado
    - Violações de dados
    - Ataques de negação de serviço (DoS)
- Interrupções no serviço de terceiros:
    - Interrupções no provedor de nuvem
    - Falhas de DNS
    - Interrupções externas de API e serviços
- Fatores ambientais:
    - Desastres naturais (terremotos, incêndios, inundações, quedas de energia)
    - Problemas relacionados ao clima

Este é um exemplo de estrutura de classificação não conclusivo que você pode adaptar para atender às suas necessidades e organização específicas. Recomendamos revisar e atualizar a estrutura de classificação periodicamente à medida que seu sistema evolui ou surgem novos tipos de incidentes.

Histórico do documento

A tabela a seguir descreve alterações significativas feitas neste guia. Se desejar receber notificações sobre futuras atualizações, inscreva-se em um feed RSS.

Alteração	Descrição	Data
Publicação inicial	—	28 de janeiro de 2025

AWS Glossário de orientação prescritiva

A seguir estão os termos comumente usados em estratégias, guias e padrões fornecidos pela Orientação AWS Prescritiva. Para sugerir entradas, use o link Fornecer feedback no final do glossário.

Números

7 Rs

Sete estratégias comuns de migração para mover aplicações para a nuvem. Essas estratégias baseiam-se nos 5 Rs identificados pela Gartner em 2011 e consistem em:

- Refatorar/rearquitetar: mova uma aplicação e modifique sua arquitetura aproveitando ao máximo os recursos nativos de nuvem para melhorar a agilidade, a performance e a escalabilidade. Isso normalmente envolve a portabilidade do sistema operacional e do banco de dados. Exemplo: migre seu banco de dados Oracle local para a edição compatível com o Amazon Aurora PostgreSQL.
- Redefinir a plataforma (mover e redefinir [mover e redefinir (lift-and-reshape]): mova uma aplicação para a nuvem e introduza algum nível de otimização a fim de aproveitar os recursos da nuvem. Exemplo: Migre seu banco de dados Oracle local para o Amazon Relational Database Service (Amazon RDS) for Oracle no. Nuvem AWS
- Recomprar (drop and shop): mude para um produto diferente, normalmente migrando de uma licença tradicional para um modelo SaaS. Exemplo: migre seu sistema de gerenciamento de relacionamento com o cliente (CRM) para a Salesforce.com.
- Redefinir a hospedagem (mover sem alterações [lift-and-shift])mover uma aplicação para a nuvem sem fazer nenhuma alteração a fim de aproveitar os recursos da nuvem. Exemplo: Migre seu banco de dados Oracle local para o Oracle em uma EC2 instância no. Nuvem AWS
- Realocar (mover o hipervisor sem alterações [hypervisor-level lift-and-shift]): mover a infraestrutura para a nuvem sem comprar novo hardware, reescrever aplicações ou modificar suas operações existentes. Você migra servidores de uma plataforma local para um serviço em nuvem para a mesma plataforma. Exemplo: Migrar um Microsoft Hyper-V aplicativo para o. AWS
- Reter (revisitar): mantenha as aplicações em seu ambiente de origem. Isso pode incluir aplicações que exigem grande refatoração, e você deseja adiar esse trabalho para um

momento posterior, e aplicações antigas que você deseja manter porque não há justificativa comercial para migrá-las.

- Retirar: desative ou remova aplicações que não são mais necessárias em seu ambiente de origem.

A

ABAC

Consulte controle de acesso baseado em atributos.

serviços abstratos

Veja os serviços gerenciados.

ACID

Veja atomicidade, consistência, isolamento, durabilidade.

migração ativa-ativa

Um método de migração de banco de dados no qual os bancos de dados de origem e de destino são mantidos em sincronia (por meio de uma ferramenta de replicação bidirecional ou operações de gravação dupla), e ambos os bancos de dados lidam com transações de aplicações conectadas durante a migração. Esse método oferece suporte à migração em lotes pequenos e controlados, em vez de exigir uma substituição única. É mais flexível, mas exige mais trabalho do que a migração ativa-passiva.

migração ativa-passiva

Um método de migração de banco de dados no qual os bancos de dados de origem e de destino são mantidos em sincronia, mas somente o banco de dados de origem manipula as transações das aplicações conectadas enquanto os dados são replicados no banco de dados de destino. O banco de dados de destino não aceita nenhuma transação durante a migração.

função agregada

Uma função SQL que opera em um grupo de linhas e calcula um único valor de retorno para o grupo. Exemplos de funções agregadas incluem `SUM` e. `MAX`

AI

Veja a inteligência artificial.

AIOps

Veja as operações de inteligência artificial.

anonimização

O processo de excluir permanentemente informações pessoais em um conjunto de dados. A anonimização pode ajudar a proteger a privacidade pessoal. Dados anônimos não são mais considerados dados pessoais.

antipadrões

Uma solução frequentemente usada para um problema recorrente em que a solução é contraproducente, ineficaz ou menos eficaz do que uma alternativa.

controle de aplicativos

Uma abordagem de segurança que permite o uso somente de aplicativos aprovados para ajudar a proteger um sistema contra malware.

portfólio de aplicações

Uma coleção de informações detalhadas sobre cada aplicação usada por uma organização, incluindo o custo para criar e manter a aplicação e seu valor comercial. Essas informações são fundamentais para o processo de descoberta e análise de portfólio e ajudam a identificar e priorizar as aplicações a serem migradas, modernizadas e otimizadas.

inteligência artificial (IA)

O campo da ciência da computação que se dedica ao uso de tecnologias de computação para desempenhar funções cognitivas normalmente associadas aos humanos, como aprender, resolver problemas e reconhecer padrões. Para obter mais informações, consulte O que é inteligência artificial?

operações de inteligência artificial (AIOps)

O processo de usar técnicas de machine learning para resolver problemas operacionais, reduzir incidentes operacionais e intervenção humana e aumentar a qualidade do serviço. Para obter mais informações sobre como AIOps é usado na estratégia de AWS migração, consulte o guia de integração de operações.

criptografia assimétrica

Um algoritmo de criptografia que usa um par de chaves, uma chave pública para criptografia e uma chave privada para descriptografia. É possível compartilhar a chave pública porque ela não é usada na descriptografia, mas o acesso à chave privada deve ser altamente restrito.

atomicidade, consistência, isolamento, durabilidade (ACID)

Um conjunto de propriedades de software que garantem a validade dos dados e a confiabilidade operacional de um banco de dados, mesmo no caso de erros, falhas de energia ou outros problemas.

controle de acesso por atributo (ABAC)

A prática de criar permissões minuciosas com base nos atributos do usuário, como departamento, cargo e nome da equipe. Para obter mais informações, consulte ABAC AWS na documentação AWS Identity and Access Management (IAM).

fonte de dados autorizada

Um local onde você armazena a versão principal dos dados, que é considerada a fonte de informações mais confiável. Você pode copiar dados da fonte de dados autorizada para outros locais com o objetivo de processar ou modificar os dados, como anonimizá-los, redigi-los ou pseudonimizá-los.

Zona de disponibilidade

Um local distinto dentro de um Região da AWS que está isolado de falhas em outras zonas de disponibilidade e fornece conectividade de rede barata e de baixa latência a outras zonas de disponibilidade na mesma região.

AWS Estrutura de adoção da nuvem (AWS CAF)

Uma estrutura de diretrizes e melhores práticas AWS para ajudar as organizações a desenvolver um plano eficiente e eficaz para migrar com sucesso para a nuvem. AWS O CAF organiza a orientação em seis áreas de foco chamadas perspectivas: negócios, pessoas, governança, plataforma, segurança e operações. As perspectivas de negócios, pessoas e governança têm como foco habilidades e processos de negócios; as perspectivas de plataforma, segurança e operações concentram-se em habilidades e processos técnicos. Por exemplo, a perspectiva das pessoas tem como alvo as partes interessadas que lidam com recursos humanos (RH), funções de pessoal e gerenciamento de pessoal. Nessa perspectiva, o AWS CAF fornece orientação para desenvolvimento, treinamento e comunicação de pessoas para ajudar a preparar a organização para a adoção bem-sucedida da nuvem. Para obter mais informações, consulte o site da AWS CAF e o whitepaper da AWS CAF.

AWS Estrutura de qualificação da carga de trabalho (AWS WQF)

Uma ferramenta que avalia as cargas de trabalho de migração do banco de dados, recomenda estratégias de migração e fornece estimativas de trabalho. AWS O WQF está incluído com AWS

Schema Conversion Tool ()AWS SCT. Ela analisa esquemas de banco de dados e objetos de código, código de aplicações, dependências e características de performance, além de fornecer relatórios de avaliação.

B

bot ruim

Um bot destinado a perturbar ou causar danos a indivíduos ou organizações.

BCP

Veja o planejamento de continuidade de negócios.

gráfico de comportamento

Uma visualização unificada e interativa do comportamento e das interações de recursos ao longo do tempo. É possível usar um gráfico de comportamento com o Amazon Detective para examinar tentativas de login malsucedidas, chamadas de API suspeitas e ações similares. Para obter mais informações, consulte Dados em um gráfico de comportamento na documentação do Detective.

sistema big-endian

Um sistema que armazena o byte mais significativo antes. Veja também endianness.

classificação binária

Um processo que prevê um resultado binário (uma de duas classes possíveis). Por exemplo, seu modelo de ML pode precisar prever problemas como “Este e-mail é ou não é spam?” ou “Este produto é um livro ou um carro?”

filtro de bloom

Uma estrutura de dados probabilística e eficiente em termos de memória que é usada para testar se um elemento é membro de um conjunto.

blue/green deployment (implantação azul/verde)

Uma estratégia de implantação em que você cria dois ambientes separados, mas idênticos. Você executa a versão atual do aplicativo em um ambiente (azul) e a nova versão do aplicativo no outro ambiente (verde). Essa estratégia ajuda você a reverter rapidamente com o mínimo de impacto.

bot

Um aplicativo de software que executa tarefas automatizadas pela Internet e simula a atividade ou interação humana. Alguns bots são úteis ou benéficos, como rastreadores da Web que indexam informações na Internet. Alguns outros bots, conhecidos como bots ruins, têm como objetivo perturbar ou causar danos a indivíduos ou organizações.

botnet

Redes de bots infectadas por malware e sob o controle de uma única parte, conhecidas como pastor de bots ou operador de bots. As redes de bots são o mecanismo mais conhecido para escalar bots e seu impacto.

ramo

Uma área contida de um repositório de código. A primeira ramificação criada em um repositório é a ramificação principal. Você pode criar uma nova ramificação a partir de uma ramificação existente e, em seguida, desenvolver recursos ou corrigir bugs na nova ramificação. Uma ramificação que você cria para gerar um recurso é comumente chamada de ramificação de recurso. Quando o recurso estiver pronto para lançamento, você mesclará a ramificação do recurso de volta com a ramificação principal. Para obter mais informações, consulte Sobre filiais (GitHub documentação).

acesso em vidro quebrado

Em circunstâncias excepcionais e por meio de um processo aprovado, um meio rápido para um usuário obter acesso a um Conta da AWS que ele normalmente não tem permissão para acessar. Para obter mais informações, consulte o indicador Implementar procedimentos de quebra de vidro na orientação do Well-Architected AWS .

estratégia brownfield

A infraestrutura existente em seu ambiente. Ao adotar uma estratégia brownfield para uma arquitetura de sistema, você desenvolve a arquitetura de acordo com as restrições dos sistemas e da infraestrutura atuais. Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e greenfield.

cache do buffer

A área da memória em que os dados acessados com mais frequência são armazenados.

capacidade de negócios

O que uma empresa faz para gerar valor (por exemplo, vendas, atendimento ao cliente ou marketing). As arquiteturas de microsserviços e as decisões de desenvolvimento podem

ser orientadas por recursos de negócios. Para obter mais informações, consulte a seção Organizados de acordo com as capacidades de negócios do whitepaper Executar microsserviços conteinerizados na AWS.

planejamento de continuidade de negócios (BCP)

Um plano que aborda o impacto potencial de um evento disruptivo, como uma migração em grande escala, nas operações e permite que uma empresa retome as operações rapidamente.

C

CAF

Consulte Estrutura de adoção da AWS nuvem.

implantação canária

O lançamento lento e incremental de uma versão para usuários finais. Quando estiver confiante, você implanta a nova versão e substituirá a versão atual em sua totalidade.

CCoE

Veja o Centro de Excelência em Nuvem.

CDC

Veja a captura de dados de alterações.

captura de dados de alterações (CDC)

O processo de rastrear alterações em uma fonte de dados, como uma tabela de banco de dados, e registrar metadados sobre a alteração. É possível usar o CDC para várias finalidades, como auditar ou replicar alterações em um sistema de destino para manter a sincronização.

engenharia do caos

Introduzir intencionalmente falhas ou eventos disruptivos para testar a resiliência de um sistema. Você pode usar AWS Fault Injection Service (AWS FIS) para realizar experimentos que estressam suas AWS cargas de trabalho e avaliar sua resposta.

CI/CD

Veja a integração e a entrega contínuas.

classificação

Um processo de categorização que ajuda a gerar previsões. Os modelos de ML para problemas de classificação predizem um valor discreto. Os valores discretos são sempre diferentes uns dos outros. Por exemplo, um modelo pode precisar avaliar se há ou não um carro em uma imagem.

criptografia no lado do cliente

Criptografia de dados localmente, antes que o alvo os AWS service (Serviço da AWS) receba.

Centro de excelência em nuvem (CCoE)

Uma equipe multidisciplinar que impulsiona os esforços de adoção da nuvem em toda a organização, incluindo o desenvolvimento de práticas recomendas de nuvem, a mobilização de recursos, o estabelecimento de cronogramas de migração e a liderança da organização em transformações em grande escala. Para obter mais informações, consulte as publicações CCo E no Blog de Estratégia Nuvem AWS Empresarial.

computação em nuvem

A tecnologia de nuvem normalmente usada para armazenamento de dados remoto e gerenciamento de dispositivos de IoT. A computação em nuvem geralmente está conectada à tecnologia de computação de ponta.

modelo operacional em nuvem

Em uma organização de TI, o modelo operacional usado para criar, amadurecer e otimizar um ou mais ambientes de nuvem. Para obter mais informações, consulte Criar seu modelo operacional de nuvem.

estágios de adoção da nuvem

As quatro fases pelas quais as organizações normalmente passam quando migram para o Nuvem AWS:

- Projeto: executar alguns projetos relacionados à nuvem para fins de prova de conceito e aprendizado
- Fundação — Fazer investimentos fundamentais para escalar sua adoção da nuvem (por exemplo, criar uma landing zone, definir um CCo E, estabelecer um modelo de operações)
- Migração: migrar aplicações individuais
- Reinvenção: otimizar produtos e serviços e inovar na nuvem

Esses estágios foram definidos por Stephen Orban na postagem do blog The Journey Toward Cloud-First & the Stages of Adoption no blog de estratégia Nuvem AWS empresarial. Para obter

informações sobre como eles se relacionam com a estratégia de AWS migração, consulte o guia de preparação para migração.

CMDB

Consulte o banco de dados de gerenciamento de configuração.

repositório de código

Um local onde o código-fonte e outros ativos, como documentação, amostras e scripts, são armazenados e atualizados por meio de processos de controle de versão. Os repositórios de nuvem comuns incluem GitHub ouBitbucket Cloud. Cada versão do código é chamada de ramificação. Em uma estrutura de microsserviços, cada repositório é dedicado a uma única peça de funcionalidade. Um único pipeline de CI/CD pode usar vários repositórios.

cache frio

Um cache de buffer que está vazio, não está bem preenchido ou contém dados obsoletos ou irrelevantes. Isso afeta a performance porque a instância do banco de dados deve ler da memória principal ou do disco, um processo que é mais lento do que a leitura do cache do buffer.

dados frios

Dados que raramente são acessados e geralmente são históricos. Ao consultar esse tipo de dados, consultas lentas geralmente são aceitáveis. Mover esses dados para níveis ou classes de armazenamento de baixo desempenho e menos caros pode reduzir os custos.

visão computacional (CV)

Um campo da IA que usa aprendizado de máquina para analisar e extrair informações de formatos visuais, como imagens e vídeos digitais. Por exemplo, a Amazon SageMaker AI fornece algoritmos de processamento de imagem para CV.

desvio de configuração

Para uma carga de trabalho, uma alteração de configuração em relação ao estado esperado. Isso pode fazer com que a carga de trabalho se torne incompatível e, normalmente, é gradual e não intencional.

banco de dados de gerenciamento de configuração (CMDB)

Um repositório que armazena e gerencia informações sobre um banco de dados e seu ambiente de TI, incluindo componentes de hardware e software e suas configurações. Normalmente, os dados de um CMDB são usados no estágio de descoberta e análise do portfólio da migração.

pacote de conformidade

Um conjunto de AWS Config regras e ações de remediação que você pode montar para personalizar suas verificações de conformidade e segurança. Você pode implantar um pacote de conformidade como uma entidade única em uma Conta da AWS região ou em uma organização usando um modelo YAML. Para obter mais informações, consulte Pacotes de conformidade na documentação. AWS Config

integração contínua e entrega contínua (CI/CD)

O processo de automatizar os estágios de origem, criação, teste, preparação e produção do processo de lançamento do software. CI/CD is commonly described as a pipeline. CI/CDpode ajudá-lo a automatizar processos, melhorar a produtividade, melhorar a qualidade do código e entregar com mais rapidez. Para obter mais informações, consulte Benefícios da entrega contínua. CD também pode significar implantação contínua. Para obter mais informações, consulte Entrega contínua versus implantação contínua.

CV

Veja visão computacional.

D

dados em repouso

Dados estacionários em sua rede, por exemplo, dados que estão em um armazenamento.

classificação de dados

Um processo para identificar e categorizar os dados em sua rede com base em criticalidade e confidencialidade. É um componente crítico de qualquer estratégia de gerenciamento de riscos de segurança cibernética, pois ajuda a determinar os controles adequados de proteção e retenção para os dados. A classificação de dados é um componente do pilar de segurança no AWS Well-Architected Framework. Para obter mais informações, consulte Classificação de dados.

desvio de dados

Uma variação significativa entre os dados de produção e os dados usados para treinar um modelo de ML ou uma alteração significativa nos dados de entrada ao longo do tempo. O desvio de dados pode reduzir a qualidade geral, a precisão e a imparcialidade das previsões do modelo de ML.

dados em trânsito

Dados que estão se movendo ativamente pela sua rede, como entre os recursos da rede.

malha de dados

Uma estrutura arquitetônica que fornece propriedade de dados distribuída e descentralizada com gerenciamento e governança centralizados.

minimização de dados

O princípio de coletar e processar apenas os dados estritamente necessários. Praticar a minimização de dados no Nuvem AWS pode reduzir os riscos de privacidade, os custos e a pegada de carbono de sua análise.

perímetro de dados

Um conjunto de proteções preventivas em seu AWS ambiente que ajudam a garantir que somente identidades confiáveis acessem recursos confiáveis das redes esperadas. Para obter mais informações, consulte Construindo um perímetro de dados em. AWS

pré-processamento de dados

A transformação de dados brutos em um formato que seja facilmente analisado por seu modelo de ML. O pré-processamento de dados pode significar a remoção de determinadas colunas ou linhas e o tratamento de valores ausentes, inconsistentes ou duplicados.

proveniência dos dados

O processo de rastrear a origem e o histórico dos dados ao longo de seu ciclo de vida, por exemplo, como os dados foram gerados, transmitidos e armazenados.

titular dos dados

Um indivíduo cujos dados estão sendo coletados e processados.

data warehouse

Um sistema de gerenciamento de dados que oferece suporte à inteligência comercial, como análises. Os data warehouses geralmente contêm grandes quantidades de dados históricos e geralmente são usados para consultas e análises.

linguagem de definição de dados (DDL)

Instruções ou comandos para criar ou modificar a estrutura de tabelas e objetos em um banco de dados.

linguagem de manipulação de dados (DML)

Instruções ou comandos para modificar (inserir, atualizar e excluir) informações em um banco de dados.

DDL

Consulte a linguagem de definição de banco de dados.

deep ensemble

A combinação de vários modelos de aprendizado profundo para gerar previsões. Os deep ensembles podem ser usados para produzir uma previsão mais precisa ou para estimar a incerteza nas previsões.

Aprendizado profundo

Um subcampo do ML que usa várias camadas de redes neurais artificiais para identificar o mapeamento entre os dados de entrada e as variáveis-alvo de interesse.

defense-in-depth

Uma abordagem de segurança da informação na qual uma série de mecanismos e controles de segurança são cuidadosamente distribuídos por toda a rede de computadores para proteger a confidencialidade, a integridade e a disponibilidade da rede e dos dados nela contidos. Ao adotar essa estratégia AWS, você adiciona vários controles em diferentes camadas da AWS Organizations estrutura para ajudar a proteger os recursos. Por exemplo, uma defense-in-depth abordagem pode combinar autenticação multifatorial, segmentação de rede e criptografia.

administrador delegado

Em AWS Organizations, um serviço compatível pode registrar uma conta de AWS membro para administrar as contas da organização e gerenciar as permissões desse serviço. Essa conta é chamada de administrador delegado para esse serviço Para obter mais informações e uma lista de serviços compatíveis, consulte Serviços que funcionam com o AWS Organizations na documentação do AWS Organizations .

implantação

O processo de criar uma aplicação, novos recursos ou correções de código disponíveis no ambiente de destino. A implantação envolve a implementação de mudanças em uma base de código e, em seguida, a criação e execução dessa base de código nos ambientes da aplicação

ambiente de desenvolvimento

Veja o ambiente.

controle detectivo

Um controle de segurança projetado para detectar, registrar e alertar após a ocorrência de um evento. Esses controles são uma segunda linha de defesa, alertando você sobre eventos de segurança que contornaram os controles preventivos em vigor. Para obter mais informações, consulte Controles detectivos em Como implementar controles de segurança na AWS.

mapeamento do fluxo de valor de desenvolvimento (DVSM)

Um processo usado para identificar e priorizar restrições que afetam negativamente a velocidade e a qualidade em um ciclo de vida de desenvolvimento de software. O DVSM estende o processo de mapeamento do fluxo de valor originalmente projetado para práticas de manufatura enxuta. Ele se concentra nas etapas e equipes necessárias para criar e movimentar valor por meio do processo de desenvolvimento de software.

gêmeo digital

Uma representação virtual de um sistema real, como um prédio, fábrica, equipamento industrial ou linha de produção. Os gêmeos digitais oferecem suporte à manutenção preditiva, ao monitoramento remoto e à otimização da produção.

tabela de dimensões

Em um esquema em estrela, uma tabela menor que contém atributos de dados sobre dados quantitativos em uma tabela de fatos. Os atributos da tabela de dimensões geralmente são campos de texto ou números discretos que se comportam como texto. Esses atributos são comumente usados para restringir consultas, filtrar e rotular conjuntos de resultados.

desastre

Um evento que impede que uma workload ou sistema cumpra seus objetivos de negócios em seu local principal de implantação. Esses eventos podem ser desastres naturais, falhas técnicas ou o resultado de ações humanas, como configuração incorreta não intencional ou ataque de malware.

Recuperação de desastres (RD)

A estratégia e o processo que você usa para minimizar o tempo de inatividade e a perda de dados causados por um desastre. Para obter mais informações, consulte Recuperação de desastres de cargas de trabalho em AWS: Recuperação na nuvem no AWS Well-Architected Framework.

DML

Veja a linguagem de manipulação de banco de dados.

design orientado por domínio

Uma abordagem ao desenvolvimento de um sistema de software complexo conectando seus componentes aos domínios em evolução, ou principais metas de negócios, atendidos por cada componente. Esse conceito foi introduzido por Eric Evans em seu livro, Design orientado por domínio: lidando com a complexidade no coração do software (Boston: Addison-Wesley Professional, 2003). Para obter informações sobre como usar o design orientado por domínio com o padrão strangler fig, consulte Modernizar incrementalmente os serviços web herdados do Microsoft ASP.NET (ASMX) usando contêineres e o Amazon API Gateway.

DR

Veja a recuperação de desastres.

detecção de deriva

Rastreando desvios de uma configuração básica. Por exemplo, você pode usar AWS CloudFormation para detectar desvios nos recursos do sistema ou AWS Control Tower para detectar mudanças em seu landing zone que possam afetar a conformidade com os requisitos de governança.

DVSM

Veja o mapeamento do fluxo de valor do desenvolvimento.

E

EDA

Veja a análise exploratória de dados.

EDI

Veja intercâmbio eletrônico de dados.

computação de borda

A tecnologia que aumenta o poder computacional de dispositivos inteligentes nas bordas de uma rede de IoT. Quando comparada à computação em nuvem, a computação de ponta pode reduzir a latência da comunicação e melhorar o tempo de resposta.

intercâmbio eletrônico de dados (EDI)

A troca automatizada de documentos comerciais entre organizações. Para obter mais informações, consulte O que é intercâmbio eletrônico de dados.

Criptografia

Um processo de computação que transforma dados de texto simples, legíveis por humanos, em texto cifrado.

chave de criptografia

Uma sequência criptográfica de bits aleatórios que é gerada por um algoritmo de criptografia. As chaves podem variar em tamanho, e cada chave foi projetada para ser imprevisível e exclusiva.

endianismo

A ordem na qual os bytes são armazenados na memória do computador. Os sistemas big-endian armazenam o byte mais significativo antes. Os sistemas little-endian armazenam o byte menos significativo antes.

endpoint

Veja o endpoint do serviço.

serviço de endpoint

Um serviço que pode ser hospedado em uma nuvem privada virtual (VPC) para ser compartilhado com outros usuários. Você pode criar um serviço de endpoint com AWS PrivateLink e conceder permissões a outros diretores Contas da AWS ou a AWS Identity and Access Management (IAM). Essas contas ou entidades principais podem se conectar ao serviço de endpoint de maneira privada criando endpoints da VPC de interface. Para obter mais informações, consulte Criar um serviço de endpoint na documentação do Amazon Virtual Private Cloud (Amazon VPC).

planejamento de recursos corporativos (ERP)

Um sistema que automatiza e gerencia os principais processos de negócios (como contabilidade, MES e gerenciamento de projetos) para uma empresa.

criptografia envelopada

O processo de criptografar uma chave de criptografia com outra chave de criptografia. Para obter mais informações, consulte Criptografia de envelope na documentação AWS Key Management Service (AWS KMS).

ambiente

Uma instância de uma aplicação em execução. Estes são tipos comuns de ambientes na computação em nuvem:

- ambiente de desenvolvimento: uma instância de uma aplicação em execução que está disponível somente para a equipe principal responsável pela manutenção da aplicação. Ambientes de desenvolvimento são usados para testar mudanças antes de promovê-las para ambientes superiores. Esse tipo de ambiente às vezes é chamado de ambiente de teste.
- ambientes inferiores: todos os ambientes de desenvolvimento para uma aplicação, como aqueles usados para compilações e testes iniciais.
- ambiente de produção: uma instância de uma aplicação em execução que os usuários finais podem acessar. Em um pipeline de CI/CD, o ambiente de produção é o último ambiente de implantação.
- ambientes superiores: todos os ambientes que podem ser acessados por usuários que não sejam a equipe principal de desenvolvimento. Isso pode incluir um ambiente de produção, ambientes de pré-produção e ambientes para testes de aceitação do usuário.

epic

Em metodologias ágeis, categorias funcionais que ajudam a organizar e priorizar seu trabalho. Os epics fornecem uma descrição de alto nível dos requisitos e das tarefas de implementação. Por exemplo, os épicos de segurança AWS da CAF incluem gerenciamento de identidade e acesso, controles de detetive, segurança de infraestrutura, proteção de dados e resposta a incidentes. Para obter mais informações sobre epics na estratégia de migração da AWS , consulte o guia de implementação do programa.

ERP

Veja o planejamento de recursos corporativos.

análise exploratória de dados (EDA)

O processo de analisar um conjunto de dados para entender suas principais características. Você coleta ou agrega dados e, em seguida, realiza investigações iniciais para encontrar padrões, detectar anomalias e verificar suposições. O EDA é realizado por meio do cálculo de estatísticas resumidas e da criação de visualizações de dados.

F

tabela de fatos

A tabela central em um esquema em estrela. Ele armazena dados quantitativos sobre as operações comerciais. Normalmente, uma tabela de fatos contém dois tipos de colunas: aquelas que contêm medidas e aquelas que contêm uma chave externa para uma tabela de dimensões.

falham rapidamente

Uma filosofia que usa testes frequentes e incrementais para reduzir o ciclo de vida do desenvolvimento. É uma parte essencial de uma abordagem ágil.

limite de isolamento de falhas

No Nuvem AWS, um limite, como uma zona de disponibilidade, Região da AWS um plano de controle ou um plano de dados, que limita o efeito de uma falha e ajuda a melhorar a resiliência das cargas de trabalho. Para obter mais informações, consulte Limites de isolamento de AWS falhas.

ramificação de recursos

Veja a filial.

recursos

Os dados de entrada usados para fazer uma previsão. Por exemplo, em um contexto de manufatura, os recursos podem ser imagens capturadas periodicamente na linha de fabricação.

importância do recurso

O quanto um recurso é importante para as previsões de um modelo. Isso geralmente é expresso como uma pontuação numérica que pode ser calculada por meio de várias técnicas, como Shapley Additive Explanations (SHAP) e gradientes integrados. Para obter mais informações, consulte Interpretabilidade do modelo de aprendizado de máquina com AWS.

transformação de recursos

O processo de otimizar dados para o processo de ML, incluindo enriquecer dados com fontes adicionais, escalar valores ou extrair vários conjuntos de informações de um único campo de dados. Isso permite que o modelo de ML se beneficie dos dados. Por exemplo, se a data “2021-05-27 00:15:37” for dividida em “2021”, “maio”, “quinta” e “15”, isso poderá ajudar o algoritmo de aprendizado a aprender padrões diferenciados associados a diferentes componentes de dados.

solicitação rápida

Fornecer a um LLM um pequeno número de exemplos que demonstram a tarefa e o resultado desejado antes de solicitar que ele execute uma tarefa semelhante. Essa técnica é uma aplicação do aprendizado contextual, em que os modelos aprendem com exemplos (fotos) incorporados aos prompts. Solicitações rápidas podem ser eficazes para tarefas que exigem formatação, raciocínio ou conhecimento de domínio específicos. Veja também a solicitação zero-shot.

FGAC

Veja o controle de acesso refinado.

Controle de acesso refinado (FGAC)

O uso de várias condições para permitir ou negar uma solicitação de acesso.

migração flash-cut

Um método de migração de banco de dados que usa replicação contínua de dados por meio da captura de dados alterados para migrar dados no menor tempo possível, em vez de usar uma abordagem em fases. O objetivo é reduzir ao mínimo o tempo de inatividade.

FM

Veja o modelo da fundação.

modelo de fundação (FM)

Uma grande rede neural de aprendizado profundo que vem treinando em grandes conjuntos de dados generalizados e não rotulados. FMs são capazes de realizar uma ampla variedade de tarefas gerais, como entender a linguagem, gerar texto e imagens e conversar em linguagem natural. Para obter mais informações, consulte O que são modelos básicos.

G

IA generativa

Um subconjunto de modelos de IA que foram treinados em grandes quantidades de dados e que podem usar uma simples solicitação de texto para criar novos conteúdos e artefatos, como imagens, vídeos, texto e áudio. Para obter mais informações, consulte O que é IA generativa.

bloqueio geográfico

Veja as restrições geográficas.

restrições geográficas (bloqueio geográfico)

Na Amazon CloudFront, uma opção para impedir que usuários em países específicos acessem distribuições de conteúdo. É possível usar uma lista de permissões ou uma lista de bloqueios para especificar países aprovados e banidos. Para obter mais informações, consulte Restringir a distribuição geográfica do seu conteúdo na CloudFront documentação.

Fluxo de trabalho do GitFlow

Uma abordagem na qual ambientes inferiores e superiores usam ramificações diferentes em um repositório de código-fonte. O fluxo de trabalho do Gitflow é considerado legado, e o fluxo de trabalho baseado em troncos é a abordagem moderna e preferida.

imagem dourada

Um instantâneo de um sistema ou software usado como modelo para implantar novas instâncias desse sistema ou software. Por exemplo, na manufatura, uma imagem dourada pode ser usada para provisionar software em vários dispositivos e ajudar a melhorar a velocidade, a escalabilidade e a produtividade nas operações de fabricação de dispositivos.

estratégia greenfield

A ausência de infraestrutura existente em um novo ambiente. Ao adotar uma estratégia greenfield para uma arquitetura de sistema, é possível selecionar todas as novas tecnologias sem a restrição da compatibilidade com a infraestrutura existente, também conhecida como brownfield. Se estiver expandindo a infraestrutura existente, poderá combinar as estratégias brownfield e greenfield.

barreira de proteção

Uma regra de alto nível que ajuda a governar recursos, políticas e conformidade em todas as unidades organizacionais ()OUs. Barreiras de proteção preventivas impõem políticas para garantir o alinhamento a padrões de conformidade. Elas são implementadas usando políticas de controle de serviço e limites de permissões do IAM. Barreiras de proteção detectivas detectam violações de políticas e problemas de conformidade e geram alertas para remediação. Eles são implementados usando AWS Config, AWS Security Hub, Amazon GuardDuty AWS Trusted Advisor, Amazon Inspector e verificações personalizadas AWS Lambda .

H

HA

Veja a alta disponibilidade.

migração heterogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que usa um mecanismo de banco de dados diferente (por exemplo, Oracle para Amazon Aurora). A migração heterogênea geralmente faz parte de um esforço de redefinição da arquitetura, e converter

o esquema pode ser uma tarefa complexa. O AWS fornece o AWS SCT para ajudar nas conversões de esquemas.

alta disponibilidade (HA)

A capacidade de uma workload operar continuamente, sem intervenção, em caso de desafios ou desastres. Os sistemas AH são projetados para realizar o failover automático, oferecer consistentemente desempenho de alta qualidade e lidar com diferentes cargas e falhas com impacto mínimo no desempenho.

modernização de historiador

Uma abordagem usada para modernizar e atualizar os sistemas de tecnologia operacional (OT) para melhor atender às necessidades do setor de manufatura. Um historiador é um tipo de banco de dados usado para coletar e armazenar dados de várias fontes em uma fábrica.

dados de retenção

Uma parte dos dados históricos rotulados que são retidos de um conjunto de dados usado para treinar um modelo de aprendizado de máquina. Você pode usar dados de retenção para avaliar o desempenho do modelo comparando as previsões do modelo com os dados de retenção.

migração homogênea de bancos de dados

Migrar seu banco de dados de origem para um banco de dados de destino que compartilha o mesmo mecanismo de banco de dados (por exemplo, Microsoft SQL Server para Amazon RDS para SQL Server). A migração homogênea geralmente faz parte de um esforço de redefinição da hospedagem ou da plataforma. É possível usar utilitários de banco de dados nativos para migrar o esquema.

dados quentes

Dados acessados com frequência, como dados em tempo real ou dados translacionais recentes. Esses dados normalmente exigem uma camada ou classe de armazenamento de alto desempenho para fornecer respostas rápidas às consultas.

hotfix

Uma correção urgente para um problema crítico em um ambiente de produção. Devido à sua urgência, um hotfix geralmente é feito fora do fluxo de trabalho típico de uma DevOps versão.

período de hipercuidados

Imediatamente após a substituição, o período em que uma equipe de migração gerencia e monitora as aplicações migradas na nuvem para resolver quaisquer problemas. Normalmente,

a duração desse período é de 1 a 4 dias. No final do período de hipercuidados, a equipe de migração normalmente transfere a responsabilidade pelas aplicações para a equipe de operações de nuvem.

eu

IaC

Veja a infraestrutura como código.

Política baseada em identidade

Uma política anexada a um ou mais diretores do IAM que define suas permissões no Nuvem AWS ambiente.

aplicação ociosa

Uma aplicação que tem um uso médio de CPU e memória entre 5 e 20% em um período de 90 dias. Em um projeto de migração, é comum retirar essas aplicações ou retê-las on-premises.

IIoT

Veja a Internet das Coisas industrial.

infraestrutura imutável

Um modelo que implanta uma nova infraestrutura para cargas de trabalho de produção em vez de atualizar, corrigir ou modificar a infraestrutura existente. Infraestruturas imutáveis são inerentemente mais consistentes, confiáveis e previsíveis do que infraestruturas mutáveis. Para obter mais informações, consulte as melhores práticas de implantação usando infraestrutura imutável no Well-Architected AWS Framework.

VPC de entrada (admissão)

Em uma arquitetura de AWS várias contas, uma VPC que aceita, inspeciona e roteia conexões de rede de fora de um aplicativo. A Arquitetura de Referência de AWS Segurança recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

migração incremental

Uma estratégia de substituição na qual você migra a aplicação em pequenas partes, em vez de realizar uma única substituição completa. Por exemplo, é possível mover inicialmente

apenas alguns microsserviços ou usuários para o novo sistema. Depois de verificar se tudo está funcionando corretamente, mova os microsserviços ou usuários adicionais de forma incremental até poder descomissionar seu sistema herdado. Essa estratégia reduz os riscos associados a migrações de grande porte.

Indústria 4.0

Um termo que foi introduzido por Klaus Schwab em 2016 para se referir à modernização dos processos de fabricação por meio de avanços em conectividade, dados em tempo real, automação, análise e IA/ML.

infraestrutura

Todos os recursos e ativos contidos no ambiente de uma aplicação.

Infraestrutura como código (IaC)

O processo de provisionamento e gerenciamento da infraestrutura de uma aplicação por meio de um conjunto de arquivos de configuração. A IaC foi projetada para ajudar você a centralizar o gerenciamento da infraestrutura, padronizar recursos e escalar rapidamente para que novos ambientes sejam reproduzíveis, confiáveis e consistentes.

Internet industrial das coisas (IIoT)

O uso de sensores e dispositivos conectados à Internet nos setores industriais, como manufatura, energia, automotivo, saúde, ciências biológicas e agricultura. Para obter mais informações, consulte Criando uma estratégia de transformação digital industrial da Internet das Coisas (IIoT).

VPC de inspeção

Em uma arquitetura de AWS várias contas, uma VPC centralizada que gerencia as inspeções do tráfego de rede entre VPCs (na mesma ou em diferentes Regiões da AWS) a Internet e as redes locais. A Arquitetura de Referência de AWS Segurança recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

Internet das Coisas (IoT)

A rede de objetos físicos conectados com sensores ou processadores incorporados que se comunicam com outros dispositivos e sistemas pela Internet ou por uma rede de comunicação local. Para obter mais informações, consulte O que é IoT?

interpretabilidade

Uma característica de um modelo de machine learning que descreve o grau em que um ser humano pode entender como as previsões do modelo dependem de suas entradas. Para obter mais informações, consulte Interpretabilidade do modelo de aprendizado de máquina com AWS.

IoT

Consulte Internet das Coisas.

Biblioteca de informações de TI (ITIL)

Um conjunto de práticas recomendadas para fornecer serviços de TI e alinhar esses serviços a requisitos de negócios. A ITIL fornece a base para o ITSM.

Gerenciamento de serviços de TI (ITSM)

Atividades associadas a design, implementação, gerenciamento e suporte de serviços de TI para uma organização. Para obter informações sobre a integração de operações em nuvem com ferramentas de ITSM, consulte o guia de integração de operações.

ITIL

Consulte a biblioteca de informações de TI.

ITSM

Veja o gerenciamento de serviços de TI.

L

controle de acesso baseado em etiqueta (LBAC)

Uma implementação do controle de acesso obrigatório (MAC) em que os usuários e os dados em si recebem explicitamente um valor de etiqueta de segurança. A interseção entre a etiqueta de segurança do usuário e a etiqueta de segurança dos dados determina quais linhas e colunas podem ser vistas pelo usuário.

zona de pouso

Uma landing zone é um AWS ambiente bem arquitetado, com várias contas, escalável e seguro. Um ponto a partir do qual suas organizações podem iniciar e implantar rapidamente workloads e aplicações com confiança em seu ambiente de segurança e infraestrutura. Para obter mais

informações sobre zonas de pouso, consulte Configurar um ambiente da AWS com várias contas seguro e escalável.

modelo de linguagem grande (LLM)

Um modelo de IA de aprendizado profundo que é pré-treinado em uma grande quantidade de dados. Um LLM pode realizar várias tarefas, como responder perguntas, resumir documentos, traduzir texto para outros idiomas e completar frases. Para obter mais informações, consulte O que são LLMs.

migração de grande porte

Uma migração de 300 servidores ou mais.

LBAC

Veja controle de acesso baseado em etiquetas.

privilégio mínimo

A prática recomendada de segurança de conceder as permissões mínimas necessárias para executar uma tarefa. Para obter mais informações, consulte Aplicar permissões de privilégios mínimos na documentação do IAM.

mover sem alterações (lift-and-shift)

Veja 7 Rs.

sistema little-endian

Um sistema que armazena o byte menos significativo antes. Veja também endianness.

LLM

Veja um modelo de linguagem grande.

ambientes inferiores

Veja o ambiente.

M

machine learning (ML)

Um tipo de inteligência artificial que usa algoritmos e técnicas para reconhecimento e aprendizado de padrões. O ML analisa e aprende com dados gravados, por exemplo, dados da

Internet das Coisas (IoT), para gerar um modelo estatístico baseado em padrões. Para obter mais informações, consulte Machine learning.

ramificação principal

Veja a filial.

malware

Software projetado para comprometer a segurança ou a privacidade do computador. O malware pode interromper os sistemas do computador, vazar informações confidenciais ou obter acesso não autorizado. Exemplos de malware incluem vírus, worms, ransomware, cavalos de Tróia, spyware e keyloggers.

serviços gerenciados

Serviços da AWS para o qual AWS opera a camada de infraestrutura, o sistema operacional e as plataformas, e você acessa os endpoints para armazenar e recuperar dados. O Amazon Simple Storage Service (Amazon S3) e o Amazon DynamoDB são exemplos de serviços gerenciados. Eles também são conhecidos como serviços abstratos.

sistema de execução de manufatura (MES)

Um sistema de software para rastrear, monitorar, documentar e controlar processos de produção que convertem matérias-primas em produtos acabados no chão de fábrica.

MAP

Consulte Migration Acceleration Program.

mecanismo

Um processo completo no qual você cria uma ferramenta, impulsiona a adoção da ferramenta e, em seguida, inspeciona os resultados para fazer ajustes. Um mecanismo é um ciclo que se reforça e se aprimora à medida que opera. Para obter mais informações, consulte Construindo mecanismos no AWS Well-Architected Framework.

conta de membro

Todos, Contas da AWS exceto a conta de gerenciamento, que fazem parte de uma organização em AWS Organizations. Uma conta só pode ser membro de uma organização de cada vez.

MES

Veja o sistema de execução de manufatura.

Transporte de telemetria de enfileiramento de mensagens (MQTT)

Um protocolo de comunicação leve machine-to-machine (M2M), baseado no padrão de publicação/assinatura, para dispositivos de IoT com recursos limitados.

microsserviço

Um serviço pequeno e independente que se comunica de forma bem definida APIs e normalmente é de propriedade de equipes pequenas e independentes. Por exemplo, um sistema de seguradora pode incluir microsserviços que mapeiam as capacidades comerciais, como vendas ou marketing, ou subdomínios, como compras, reclamações ou análises. Os benefícios dos microsserviços incluem agilidade, escalabilidade flexível, fácil implantação, código reutilizável e resiliência. Para obter mais informações, consulte Integração de microsserviços usando serviços sem AWS servidor.

arquitetura de microsserviços

Uma abordagem à criação de aplicações com componentes independentes que executam cada processo de aplicação como um microsserviço. Esses microsserviços se comunicam por meio de uma interface bem definida usando leveza. APIs Cada microsserviço nessa arquitetura pode ser atualizado, implantado e escalado para atender à demanda por funções específicas de uma aplicação. Para obter mais informações, consulte Implementação de microsserviços em. AWS

Programa de Aceleração da Migração (MAP)

Um AWS programa que fornece suporte de consultoria, treinamento e serviços para ajudar as organizações a criar uma base operacional sólida para migrar para a nuvem e ajudar a compensar o custo inicial das migrações. O MAP inclui uma metodologia de migração para executar migrações legadas de forma metódica e um conjunto de ferramentas para automatizar e acelerar cenários comuns de migração.

migração em escala

O processo de mover a maior parte do portfólio de aplicações para a nuvem em ondas, com mais aplicações sendo movidas em um ritmo mais rápido a cada onda. Essa fase usa as práticas recomendadas e lições aprendidas nas fases anteriores para implementar uma fábrica de migração de equipes, ferramentas e processos para agilizar a migração de workloads por meio de automação e entrega ágeis. Esta é a terceira fase da estratégia de migração para a AWS.

fábrica de migração

Equipes multifuncionais que simplificam a migração de workloads por meio de abordagens automatizadas e ágeis. As equipes da fábrica de migração geralmente incluem operações,

analistas e proprietários de negócios, engenheiros de migração, desenvolvedores e DevOps profissionais que trabalham em sprints. Entre 20 e 50% de um portfólio de aplicações corporativas consiste em padrões repetidos que podem ser otimizados por meio de uma abordagem de fábrica. Para obter mais informações, consulte discussão sobre fábricas de migração e o guia do Cloud Migration Factory neste conjunto de conteúdo.

metadados de migração

As informações sobre a aplicação e o servidor necessárias para concluir a migração. Cada padrão de migração exige um conjunto de metadados de migração diferente. Exemplos de metadados de migração incluem a sub-rede, o grupo de segurança e AWS a conta de destino.

padrão de migração

Uma tarefa de migração repetível que detalha a estratégia de migração, o destino da migração e a aplicação ou o serviço de migração usado. Exemplo: rehospede a migração para a Amazon EC2 com o AWS Application Migration Service.

Avaliação de Portfólio para Migração (MPA)

Uma ferramenta on-line que fornece informações para validar o caso de negócios para migrar para o. Nuvem AWS O MPA fornece avaliação detalhada do portfólio (dimensionamento correto do servidor, preços, comparações de TCO, análise de custos de migração), bem como planejamento de migração (análise e coleta de dados de aplicações, agrupamento de aplicações, priorização de migração e planejamento de ondas). A ferramenta MPA (requer login) está disponível gratuitamente para todos os AWS consultores e consultores parceiros da APN.

Avaliação de Preparação para Migração (MRA)

O processo de obter insights sobre o status de prontidão de uma organização para a nuvem, identificar pontos fortes e fracos e criar um plano de ação para fechar as lacunas identificadas, usando o CAF. AWS Para mais informações, consulte o guia de preparação para migração. A MRA é a primeira fase da estratégia de migração para a AWS.

estratégia de migração

A abordagem usada para migrar uma carga de trabalho para o. Nuvem AWS Para obter mais informações, consulte a entrada de 7 Rs neste glossário e consulte Mobilize sua organização para acelerar migrações em grande escala.

ML

Veja o aprendizado de máquina.

modernização

Transformar uma aplicação desatualizada (herdada ou monolítica) e sua infraestrutura em um sistema ágil, elástico e altamente disponível na nuvem para reduzir custos, ganhar eficiência e aproveitar as inovações. Para obter mais informações, consulte Estratégia para modernizar aplicativos no Nuvem AWS.

avaliação de preparação para modernização

Uma avaliação que ajuda a determinar a preparação para modernização das aplicações de uma organização. Ela identifica benefícios, riscos e dependências e determina o quão bem a organização pode acomodar o estado futuro dessas aplicações. O resultado da avaliação é um esquema da arquitetura de destino, um roteiro que detalha as fases de desenvolvimento e os marcos do processo de modernização e um plano de ação para abordar as lacunas identificadas. Para obter mais informações, consulte Avaliação da prontidão para modernização de aplicativos no. Nuvem AWS

aplicações monolíticas (monólitos)

Aplicações que são executadas como um único serviço com processos fortemente acoplados. As aplicações monolíticas apresentam várias desvantagens. Se um recurso da aplicação apresentar um aumento na demanda, toda a arquitetura deverá ser escalada. Adicionar ou melhorar os recursos de uma aplicação monolítica também se torna mais complexo quando a base de código cresce. Para resolver esses problemas, é possível criar uma arquitetura de microsserviços. Para obter mais informações, consulte Decompor monólitos em microsserviços.

MAPA

Consulte Avaliação do portfólio de migração.

MQTT

Consulte Transporte de telemetria de enfileiramento de mensagens.

classificação multiclasse

Um processo que ajuda a gerar previsões para várias classes (prevendo um ou mais de dois resultados). Por exemplo, um modelo de ML pode perguntar “Este produto é um livro, um carro ou um telefone?” ou “Qual categoria de produtos é mais interessante para este cliente?”

infraestrutura mutável

Um modelo que atualiza e modifica a infraestrutura existente para cargas de trabalho de produção. Para melhorar a consistência, confiabilidade e previsibilidade, o AWS Well-Architected Framework recomenda o uso de infraestrutura imutável como uma prática recomendada.

O

OAC

Veja o controle de acesso de origem.

CARVALHO

Veja a identidade de acesso de origem.

OCM

Veja o gerenciamento de mudanças organizacionais.

migração offline

Um método de migração no qual a workload de origem é desativada durante o processo de migração. Esse método envolve tempo de inatividade prolongado e geralmente é usado para workloads pequenas e não críticas.

OI

Veja a integração de operações.

OLA

Veja o contrato em nível operacional.

migração online

Um método de migração no qual a workload de origem é copiada para o sistema de destino sem ser colocada offline. As aplicações conectadas à workload podem continuar funcionando durante a migração. Esse método envolve um tempo de inatividade nulo ou mínimo e normalmente é usado para workloads essenciais para a produção.

OPC-UA

Consulte Comunicação de processo aberto — Arquitetura unificada.

Comunicação de processo aberto — Arquitetura unificada (OPC-UA)

Um protocolo de comunicação machine-to-machine (M2M) para automação industrial. O OPC-UA fornece um padrão de interoperabilidade com esquemas de criptografia, autenticação e autorização de dados.

acordo de nível operacional (OLA)

Um acordo que esclarece o que os grupos funcionais de TI prometem oferecer uns aos outros para apoiar um acordo de serviço (SLA).

análise de prontidão operacional (ORR)

Uma lista de verificação de perguntas e melhores práticas associadas que ajudam você a entender, avaliar, prevenir ou reduzir o escopo de incidentes e possíveis falhas. Para obter mais informações, consulte Operational Readiness Reviews (ORR) no Well-Architected AWS Framework.

tecnologia operacional (OT)

Sistemas de hardware e software que funcionam com o ambiente físico para controlar operações, equipamentos e infraestrutura industriais. Na manufatura, a integração dos sistemas OT e de tecnologia da informação (TI) é o foco principal das transformações da Indústria 4.0.

integração de operações (OI)

O processo de modernização das operações na nuvem, que envolve planejamento de preparação, automação e integração. Para obter mais informações, consulte o guia de integração de operações.

trilha organizacional

Uma trilha criada por ela AWS CloudTrail registra todos os eventos de todos Contas da AWS em uma organização em AWS Organizations. Essa trilha é criada em cada Conta da AWS que faz parte da organização e monitora a atividade em cada conta. Para obter mais informações, consulte Criação de uma trilha para uma organização na CloudTrail documentação.

gerenciamento de alterações organizacionais (OCM)

Uma estrutura para gerenciar grandes transformações de negócios disruptivas de uma perspectiva de pessoas, cultura e liderança. O OCM ajuda as organizações a se prepararem e fazerem a transição para novos sistemas e estratégias, acelerando a adoção de alterações, abordando questões de transição e promovendo mudanças culturais e organizacionais. Na estratégia de AWS migração, essa estrutura é chamada de aceleração de pessoas, devido à velocidade de mudança exigida nos projetos de adoção da nuvem. Para obter mais informações, consulte o guia do OCM.

controle de acesso de origem (OAC)

Em CloudFront, uma opção aprimorada para restringir o acesso para proteger seu conteúdo do Amazon Simple Storage Service (Amazon S3). O OAC oferece suporte a todos os buckets

S3 Regiões da AWS, criptografia do lado do servidor com AWS KMS (SSE-KMS) e solicitações dinâmicas ao bucket S3. `PUT` `DELETE`

Identidade do acesso de origem (OAI)

Em CloudFront, uma opção para restringir o acesso para proteger seu conteúdo do Amazon S3. Quando você usa o OAI, CloudFront cria um principal com o qual o Amazon S3 pode se autenticar. Os diretores autenticados podem acessar o conteúdo em um bucket do S3 somente por meio de uma distribuição específica. CloudFront Veja também OAC, que fornece um controle de acesso mais granular e aprimorado.

ORR

Veja a análise de prontidão operacional.

OT

Veja a tecnologia operacional.

VPC de saída (egresso)

Em uma arquitetura de AWS várias contas, uma VPC que gerencia conexões de rede que são iniciadas de dentro de um aplicativo. A Arquitetura de Referência de AWS Segurança recomenda configurar sua conta de rede com entrada, saída e inspeção VPCs para proteger a interface bidirecional entre seu aplicativo e a Internet em geral.

P

limite de permissões

Uma política de gerenciamento do IAM anexada a entidades principais do IAM para definir as permissões máximas que o usuário ou perfil podem ter. Para obter mais informações, consulte Limites de permissões na documentação do IAM.

Informações de identificação pessoal (PII)

Informações que, quando visualizadas diretamente ou combinadas com outros dados relacionados, podem ser usadas para inferir razoavelmente a identidade de um indivíduo. Exemplos de PII incluem nomes, endereços e informações de contato.

PII

Veja as informações de identificação pessoal.

manual

Um conjunto de etapas predefinidas que capturam o trabalho associado às migrações, como a entrega das principais funções operacionais na nuvem. Um manual pode assumir a forma de scripts, runbooks automatizados ou um resumo dos processos ou etapas necessários para operar seu ambiente modernizado.

PLC

Consulte controlador lógico programável.

AMEIXA

Veja o gerenciamento do ciclo de vida do produto.

política

Um objeto que pode definir permissões (consulte a política baseada em identidade), especificar as condições de acesso (consulte a política baseada em recursos) ou definir as permissões máximas para todas as contas em uma organização em AWS Organizations (consulte a política de controle de serviços).

persistência poliglota

Escolher de forma independente a tecnologia de armazenamento de dados de um microsserviço com base em padrões de acesso a dados e outros requisitos. Se seus microsserviços tiverem a mesma tecnologia de armazenamento de dados, eles poderão enfrentar desafios de implementação ou apresentar baixa performance. Os microsserviços serão implementados com mais facilidade e alcançarão performance e escalabilidade melhores se usarem o armazenamento de dados mais bem adaptado às suas necessidades. Para obter mais informações, consulte Habilitar a persistência de dados em microsserviços.

avaliação do portfólio

Um processo de descobrir, analisar e priorizar o portfólio de aplicações para planejar a migração. Para obter mais informações, consulte Avaliar a preparação para a migração.

predicado

Uma condição de consulta que retorna `true` ou`false`, normalmente localizada em uma `WHERE` cláusula.

pressão de predicados

Uma técnica de otimização de consulta de banco de dados que filtra os dados na consulta antes da transferência. Isso reduz a quantidade de dados que devem ser recuperados e processados do banco de dados relacional e melhora o desempenho das consultas.

controle preventivo

Um controle de segurança projetado para evitar que um evento ocorra. Esses controles são a primeira linha de defesa para ajudar a evitar acesso não autorizado ou alterações indesejadas em sua rede. Para obter mais informações, consulte Controles preventivos em Como implementar controles de segurança na AWS.

principal (entidade principal)

Uma entidade AWS que pode realizar ações e acessar recursos. Essa entidade geralmente é um usuário raiz para um Conta da AWS, uma função do IAM ou um usuário. Para obter mais informações, consulte Entidade principal em Termos e conceitos de perfis na documentação do IAM.

privacidade por design

Uma abordagem de engenharia de sistema que leva em consideração a privacidade em todo o processo de desenvolvimento.

zonas hospedadas privadas

Um contêiner que contém informações sobre como você deseja que o Amazon Route 53 responda às consultas de DNS para um domínio e seus subdomínios em um ou mais. VPCs Para obter mais informações, consulte Como trabalhar com zonas hospedadas privadas na documentação do Route 53.

controle proativo

Um controle de segurança projetado para impedir a implantação de recursos não compatíveis. Esses controles examinam os recursos antes de serem provisionados. Se o recurso não estiver em conformidade com o controle, ele não será provisionado. Para obter mais informações, consulte o guia de referência de controles na AWS Control Tower documentação e consulte Controles proativos em Implementação de controles de segurança em AWS.

gerenciamento do ciclo de vida do produto (PLM)

O gerenciamento de dados e processos de um produto em todo o seu ciclo de vida, desde o design, desenvolvimento e lançamento, passando pelo crescimento e maturidade, até o declínio e a remoção.

ambiente de produção

Veja o ambiente.

controlador lógico programável (PLC)

Na fabricação, um computador altamente confiável e adaptável que monitora as máquinas e automatiza os processos de fabricação.

encadeamento imediato

Usando a saída de um prompt do LLM como entrada para o próximo prompt para gerar respostas melhores. Essa técnica é usada para dividir uma tarefa complexa em subtarefas ou para refinar ou expandir iterativamente uma resposta preliminar. Isso ajuda a melhorar a precisão e a relevância das respostas de um modelo e permite resultados mais granulares e personalizados.

pseudonimização

O processo de substituir identificadores pessoais em um conjunto de dados por valores de espaço reservado. A pseudonimização pode ajudar a proteger a privacidade pessoal. Os dados pseudonimizados ainda são considerados dados pessoais.

publish/subscribe (pub/sub)

Um padrão que permite comunicações assíncronas entre microsserviços para melhorar a escalabilidade e a capacidade de resposta. Por exemplo, em um MES baseado em microsserviços, um microsserviço pode publicar mensagens de eventos em um canal no qual outros microsserviços possam se inscrever. O sistema pode adicionar novos microsserviços sem alterar o serviço de publicação.

Q

plano de consulta

Uma série de etapas, como instruções, usadas para acessar os dados em um sistema de banco de dados relacional SQL.

regressão de planos de consultas

Quando um otimizador de serviço de banco de dados escolhe um plano menos adequado do que escolhia antes de uma determinada alteração no ambiente de banco de dados ocorrer. Isso pode ser causado por alterações em estatísticas, restrições, configurações do ambiente, associações de parâmetros de consulta e atualizações do mecanismo de banco de dados.

R

Matriz RACI

Veja responsável, responsável, consultado, informado (RACI).

RAG

Consulte Geração Aumentada de Recuperação.

ransomware

Um software mal-intencionado desenvolvido para bloquear o acesso a um sistema ou dados de computador até que um pagamento seja feito.

Matriz RASCI

Veja responsável, responsável, consultado, informado (RACI).

RCAC

Veja o controle de acesso por linha e coluna.

réplica de leitura

Uma cópia de um banco de dados usada somente para leitura. É possível encaminhar consultas para a réplica de leitura e reduzir a carga no banco de dados principal.

rearquiteta

Veja 7 Rs.

objetivo de ponto de recuperação (RPO).

O máximo período de tempo aceitável desde o último ponto de recuperação de dados. Isso determina o que é considerado uma perda aceitável de dados entre o último ponto de recuperação e a interrupção do serviço.

objetivo de tempo de recuperação (RTO)

O máximo atraso aceitável entre a interrupção e a restauração do serviço.

refatorar

Veja 7 Rs.

Região

Uma coleção de AWS recursos em uma área geográfica. Cada um Região da AWS é isolado e independente dos outros para fornecer tolerância a falhas, estabilidade e resiliência. Para obter mais informações, consulte Especificar o que Regiões da AWS sua conta pode usar.

regressão

Uma técnica de ML que prevê um valor numérico. Por exemplo, para resolver o problema de "Por qual preço esta casa será vendida?" um modelo de ML pode usar um modelo de regressão linear para prever o preço de venda de uma casa com base em fatos conhecidos sobre a casa (por exemplo, a metragem quadrada).

redefinir a hospedagem

Veja 7 Rs.

versão

Em um processo de implantação, o ato de promover mudanças em um ambiente de produção.

realocar

Veja 7 Rs.

redefinir a plataforma

Veja 7 Rs.

recomprar

Veja 7 Rs.

resiliência

A capacidade de um aplicativo de resistir ou se recuperar de interrupções. Alta disponibilidade e recuperação de desastres são considerações comuns ao planejar a resiliência no. Nuvem AWS Para obter mais informações, consulte Nuvem AWS Resiliência.

política baseada em recurso

Uma política associada a um recurso, como um bucket do Amazon S3, um endpoint ou uma chave de criptografia. Esse tipo de política especifica quais entidades principais têm acesso permitido, ações válidas e quaisquer outras condições que devem ser atendidas.

matriz responsável, accountable, consultada, informada (RACI)

Uma matriz que define as funções e responsabilidades de todas as partes envolvidas nas atividades de migração e nas operações de nuvem. O nome da matriz é derivado dos tipos de responsabilidade definidos na matriz: responsável (R), responsabilizável (A), consultado (C) e informado (I). O tipo de suporte (S) é opcional. Se você incluir suporte, a matriz será chamada de matriz RASCI e, se excluir, será chamada de matriz RACI.

controle responsivo

Um controle de segurança desenvolvido para conduzir a remediação de eventos adversos ou desvios em relação à linha de base de segurança. Para obter mais informações, consulte Controles responsivos em Como implementar controles de segurança na AWS.

reter

Veja 7 Rs.

aposentar-se

Veja 7 Rs.

Geração Aumentada de Recuperação (RAG)

Uma tecnologia de IA generativa na qual um LLM faz referência a uma fonte de dados autorizada que está fora de suas fontes de dados de treinamento antes de gerar uma resposta. Por exemplo, um modelo RAG pode realizar uma pesquisa semântica na base de conhecimento ou nos dados personalizados de uma organização. Para obter mais informações, consulte O que é RAG.

alternância

O processo de atualizar periodicamente um segredo para dificultar o acesso das credenciais por um invasor.

controle de acesso por linha e coluna (RCAC)

O uso de expressões SQL básicas e flexíveis que tenham regras de acesso definidas. O RCAC consiste em permissões de linha e máscaras de coluna.

RPO

Veja o objetivo do ponto de recuperação.

RTO

Veja o objetivo do tempo de recuperação.

runbook

Um conjunto de procedimentos manuais ou automatizados necessários para realizar uma tarefa específica. Eles são normalmente criados para agilizar operações ou procedimentos repetitivos com altas taxas de erro.

S

SAML 2.0

Um padrão aberto que muitos provedores de identidade (IdPs) usam. Esse recurso permite o login único federado (SSO), para que os usuários possam fazer login AWS Management Console ou chamar as operações da AWS API sem que você precise criar um usuário no IAM para todos em sua organização. Para obter mais informações sobre a federação baseada em SAML 2.0, consulte Sobre a federação baseada em SAML 2.0 na documentação do IAM.

SCADA

Veja controle de supervisão e aquisição de dados.

SCP

Veja a política de controle de serviços.

secret

Em AWS Secrets Manager, informações confidenciais ou restritas, como uma senha ou credenciais de usuário, que você armazena de forma criptografada. Ele consiste no valor secreto e em seus metadados. O valor secreto pode ser binário, uma única string ou várias strings. Para obter mais informações, consulte O que há em um segredo do Secrets Manager? na documentação do Secrets Manager.

segurança por design

Uma abordagem de engenharia de sistemas que leva em conta a segurança em todo o processo de desenvolvimento.

controle de segurança

Uma barreira de proteção técnica ou administrativa que impede, detecta ou reduz a capacidade de uma ameaça explorar uma vulnerabilidade de segurança. Existem quatro tipos principais de controles de segurança: preventivos, detectivos, responsivos e proativos.

fortalecimento da segurança

O processo de reduzir a superfície de ataque para torná-la mais resistente a ataques. Isso pode incluir ações como remover recursos que não são mais necessários, implementar a prática recomendada de segurança de conceder privilégios mínimos ou desativar recursos desnecessários em arquivos de configuração.

sistema de gerenciamento de eventos e informações de segurança (SIEM)

Ferramentas e serviços que combinam sistemas de gerenciamento de informações de segurança (SIM) e gerenciamento de eventos de segurança (SEM). Um sistema SIEM coleta, monitora e analisa dados de servidores, redes, dispositivos e outras fontes para detectar ameaças e violações de segurança e gerar alertas.

automação de resposta de segurança

Uma ação predefinida e programada projetada para responder ou remediar automaticamente um evento de segurança. Essas automações servem como controles de segurança responsivos ou detectivos que ajudam você a implementar as melhores práticas AWS de segurança. Exemplos de ações de resposta automatizada incluem a modificação de um grupo de segurança da VPC, a correção de uma instância EC2 da Amazon ou a rotação de credenciais.

Criptografia do lado do servidor

Criptografia dos dados em seu destino, por AWS service (Serviço da AWS) quem os recebe.

política de controle de serviços (SCP)

Uma política que fornece controle centralizado sobre as permissões de todas as contas em uma organização em AWS Organizations. SCPs defina barreiras ou estabeleça limites nas ações que um administrador pode delegar a usuários ou funções. Você pode usar SCPs como listas de permissão ou listas de negação para especificar quais serviços ou ações são permitidos ou proibidos. Para obter mais informações, consulte Políticas de controle de serviço na AWS Organizations documentação.

service endpoint (endpoint de serviço)

O URL do ponto de entrada para um AWS service (Serviço da AWS). Você pode usar o endpoint para se conectar programaticamente ao serviço de destino. Para obter mais informações, consulte Endpoints do AWS service (Serviço da AWS) na Referência geral da AWS.

acordo de serviço (SLA)

Um acordo que esclarece o que uma equipe de TI promete fornecer aos clientes, como tempo de atividade e performance do serviço.

indicador de nível de serviço (SLI)

Uma medida de um aspecto de desempenho de um serviço, como taxa de erro, disponibilidade ou taxa de transferência.

objetivo de nível de serviço (SLO)

Uma métrica alvo que representa a integridade de um serviço, conforme medida por um indicador de nível de serviço.

modelo de responsabilidade compartilhada

Um modelo que descreve a responsabilidade com a qual você compartilha AWS pela segurança e conformidade na nuvem. AWS é responsável pela segurança da nuvem, enquanto você é responsável pela segurança na nuvem. Para obter mais informações, consulte o Modelo de responsabilidade compartilhada.

SIEM

Veja informações de segurança e sistema de gerenciamento de eventos.

ponto único de falha (SPOF)

Uma falha em um único componente crítico de um aplicativo que pode interromper o sistema.

SLA

Veja o contrato de nível de serviço.

ESGUIO

Veja o indicador de nível de serviço.

SLO

Veja o objetivo do nível de serviço.

split-and-seed modelo

Um padrão para escalar e acelerar projetos de modernização. À medida que novos recursos e lançamentos de produtos são definidos, a equipe principal se divide para criar novas equipes de produtos. Isso ajuda a escalar os recursos e os serviços da sua organização, melhora a produtividade do desenvolvedor e possibilita inovações rápidas. Para obter mais informações, consulte Abordagem em fases para modernizar aplicativos no. Nuvem AWS

CUSPE

Veja um único ponto de falha.

esquema de estrelas

Uma estrutura organizacional de banco de dados que usa uma grande tabela de fatos para armazenar dados transacionais ou medidos e usa uma ou mais tabelas dimensionais menores para armazenar atributos de dados. Essa estrutura foi projetada para uso em um data warehouse ou para fins de inteligência comercial.

padrão strangler fig

Uma abordagem à modernização de sistemas monolíticos que consiste em reescrever e substituir incrementalmente a funcionalidade do sistema até que o sistema herdado possa ser desativado. Esse padrão usa a analogia de uma videira que cresce e se torna uma árvore estabelecida e, eventualmente, supera e substitui sua hospedeira. O padrão foi apresentado por Martin Fowler como forma de gerenciar riscos ao reescrever sistemas monolíticos. Para ver um exemplo de como aplicar esse padrão, consulte Modernizar incrementalmente os serviços Web herdados do Microsoft ASP.NET (ASMX) usando contêineres e o Amazon API Gateway.

sub-rede

Um intervalo de endereços IP na VPC. Cada sub-rede fica alocada em uma única zona de disponibilidade.

controle de supervisão e aquisição de dados (SCADA)

Na manufatura, um sistema que usa hardware e software para monitorar ativos físicos e operações de produção.

symmetric encryption (criptografia simétrica)

Um algoritmo de criptografia que usa a mesma chave para criptografar e descriptografar dados.

testes sintéticos

Testar um sistema de forma que simule as interações do usuário para detectar possíveis problemas ou monitorar o desempenho. Você pode usar o Amazon CloudWatch Synthetics para criar esses testes.

prompt do sistema

Uma técnica para fornecer contexto, instruções ou diretrizes a um LLM para direcionar seu comportamento. Os prompts do sistema ajudam a definir o contexto e estabelecer regras para interações com os usuários.

T

tags

Pares de valores-chave que atuam como metadados para organizar seus recursos. AWS As tags podem ajudar você a gerenciar, identificar, organizar, pesquisar e filtrar recursos. Para obter mais informações, consulte Marcar seus recursos do AWS.

variável-alvo

O valor que você está tentando prever no ML supervisionado. Ela também é conhecida como variável de resultado. Por exemplo, em uma configuração de fabricação, a variável-alvo pode ser um defeito do produto.

lista de tarefas

Uma ferramenta usada para monitorar o progresso por meio de um runbook. Uma lista de tarefas contém uma visão geral do runbook e uma lista de tarefas gerais a serem concluídas. Para cada tarefa geral, ela inclui o tempo estimado necessário, o proprietário e o progresso.

ambiente de teste

Veja o ambiente.

treinamento

O processo de fornecer dados para que seu modelo de ML aprenda. Os dados de treinamento devem conter a resposta correta. O algoritmo de aprendizado descobre padrões nos dados de treinamento que mapeiam os atributos dos dados de entrada no destino (a resposta que você deseja prever). Ele gera um modelo de ML que captura esses padrões. Você pode usar o modelo de ML para obter previsões de novos dados cujo destino você não conhece.

gateway de trânsito

Um hub de trânsito de rede que você pode usar para interconectar sua rede com VPCs a rede local. Para obter mais informações, consulte O que é um gateway de trânsito na AWS Transit Gateway documentação.

fluxo de trabalho baseado em troncos

Uma abordagem na qual os desenvolvedores criam e testam recursos localmente em uma ramificação de recursos e, em seguida, mesclam essas alterações na ramificação principal. A ramificação principal é então criada para os ambientes de desenvolvimento, pré-produção e produção, sequencialmente.

Acesso confiável

Conceder permissões a um serviço que você especifica para realizar tarefas em sua organização AWS Organizations e em suas contas em seu nome. O serviço confiável cria um perfil vinculado ao serviço em cada conta, quando esse perfil é necessário, para realizar tarefas de gerenciamento para você. Para obter mais informações, consulte Usando AWS Organizations com outros AWS serviços na AWS Organizations documentação.

tuning (ajustar)

Alterar aspectos do processo de treinamento para melhorar a precisão do modelo de ML. Por exemplo, você pode treinar o modelo de ML gerando um conjunto de rótulos, adicionando rótulos e repetindo essas etapas várias vezes em configurações diferentes para otimizar o modelo.

equipe de duas pizzas

Uma pequena DevOps equipe que você pode alimentar com duas pizzas. Uma equipe de duas pizzas garante a melhor oportunidade possível de colaboração no desenvolvimento de software.

U

incerteza

Um conceito que se refere a informações imprecisas, incompletas ou desconhecidas que podem minar a confiabilidade dos modelos preditivos de ML. Há dois tipos de incertezas: a incerteza epistêmica é causada por dados limitados e incompletos, enquanto a incerteza aleatória é causada pelo ruído e pela aleatoriedade inerentes aos dados. Para obter mais informações, consulte o guia Como quantificar a incerteza em sistemas de aprendizado profundo.

tarefas indiferenciadas

Também conhecido como trabalho pesado, trabalho necessário para criar e operar um aplicativo, mas que não fornece valor direto ao usuário final nem oferece vantagem competitiva. Exemplos de tarefas indiferenciadas incluem aquisição, manutenção e planejamento de capacidade.

ambientes superiores

Veja o ambiente.

V

aspiração

Uma operação de manutenção de banco de dados que envolve limpeza após atualizações incrementais para recuperar armazenamento e melhorar a performance.

controle de versões

Processos e ferramentas que rastreiam mudanças, como alterações no código-fonte em um repositório.

emparelhamento da VPC

Uma conexão entre duas VPCs que permite rotear o tráfego usando endereços IP privados. Para ter mais informações, consulte O que é emparelhamento de VPC? na documentação da Amazon VPC.

Vulnerabilidade

Uma falha de software ou hardware que compromete a segurança do sistema.

W

cache quente

Um cache de buffer que contém dados atuais e relevantes que são acessados com frequência. A instância do banco de dados pode ler do cache do buffer, o que é mais rápido do que ler da memória principal ou do disco.

dados mornos

Dados acessados raramente. Ao consultar esse tipo de dados, consultas moderadamente lentas geralmente são aceitáveis.

função de janela

Uma função SQL que executa um cálculo em um grupo de linhas que se relacionam de alguma forma com o registro atual. As funções de janela são úteis para processar tarefas, como calcular uma média móvel ou acessar o valor das linhas com base na posição relativa da linha atual.

workload

Uma coleção de códigos e recursos que geram valor empresarial, como uma aplicação voltada para o cliente ou um processo de back-end.

workstreams

Grupos funcionais em um projeto de migração que são responsáveis por um conjunto específico de tarefas. Cada workstream é independente, mas oferece suporte aos outros workstreams do projeto. Por exemplo, o workstream de portfólio é responsável por priorizar aplicações, planejar ondas e coletar metadados de migração. O workstream de portfólio entrega esses ativos ao workstream de migração, que então migra os servidores e as aplicações.

MINHOCA

Veja escrever uma vez, ler muitas.

WQF

Consulte Estrutura de qualificação AWS da carga de trabalho.

escreva uma vez, leia muitas (WORM)

Um modelo de armazenamento que grava dados uma única vez e evita que os dados sejam excluídos ou modificados. Os usuários autorizados podem ler os dados quantas vezes forem necessárias, mas não podem alterá-los. Essa infraestrutura de armazenamento de dados é considerada imutável.

Z

exploração de dia zero

Um ataque, geralmente malware, que tira proveito de uma vulnerabilidade de dia zero.

vulnerabilidade de dia zero

Uma falha ou vulnerabilidade não mitigada em um sistema de produção. Os agentes de ameaças podem usar esse tipo de vulnerabilidade para atacar o sistema. Os desenvolvedores frequentemente ficam cientes da vulnerabilidade como resultado do ataque.

aviso zero-shot

Fornecer a um LLM instruções para realizar uma tarefa, mas sem exemplos (fotos) que possam ajudar a orientá-la. O LLM deve usar seu conhecimento pré-treinado para lidar com a tarefa. A

eficácia da solicitação zero depende da complexidade da tarefa e da qualidade da solicitação. Veja também a solicitação de algumas fotos.

aplicação zumbi

Uma aplicação que tem um uso médio de CPU e memória inferior a 5%. Em um projeto de migração, é comum retirar essas aplicações.

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.