Práticas comprovadas para desenvolver uma estratégia multicloud

AWS Orientação prescritiva

AWS Orientação prescritiva: Práticas comprovadas para desenvolver uma estratégia multicloud

Table of Contents

Práticas comprovadas para desenvolver uma estratégia multicloud

Tom Godden e Ellie Tamari, Amazon Web Services

Setembro de 2025 (histórico do documento)

Atualmente, as organizações enfrentam mensagens conflitantes sobre a adoção da multicloud. Alguns desaconselham totalmente isso, enquanto outros afirmam que todos estão migrando para um ambiente multicloud. A realidade está entre esses extremos: existem razões legítimas a favor e contra as estratégias de multicloud, e o sucesso depende do equilíbrio entre o valor potencial do negócio e a complexidade e o risco inerentes.

Na AWS, nosso compromisso com a interoperabilidade é um dos principais motivos pelos quais muitos clientes escolhem nossa plataforma. Acreditamos em dar a você a liberdade de inovar onde quer que estejam suas cargas de trabalho e capacitar você a escolher a tecnologia que melhor atenda às suas necessidades. Na AWS, estamos na vanguarda do desenvolvimento de soluções que permitem criar e implantar aplicativos em qualquer ambiente. Essa abordagem centrada no cliente é fundamental para a Nuvem AWS, na qual milhões de clientes em todo o mundo confiam.

Entendemos que os clientes precisam de plataformas de nuvem que funcionem perfeitamente com as ferramentas existentes e com as futuras opções de tecnologia. Você não precisa reconstruir tudo ao adicionar recursos de outro provedor. Sua nuvem deve ajudá-lo a conectar, proteger e gerenciar cargas de trabalho em todos os ambientes sem forçar você a se tornar um especialista em todas as plataformas. AWS cria pontos de conexão diretamente em seus serviços para ajudá-lo a operar com eficiência, independentemente de sua estratégia ser usar AWS exclusivamente ou seguir uma abordagem multicloud seletiva.

Reconhecemos que cada organização tem requisitos comerciais exclusivos que orientam suas decisões estratégicas de nuvem. Se você está executando cargas de trabalho principalmente em AWS, executando-as em várias nuvens ou usando AWS como parte de uma arquitetura multicloud mais ampla, estamos comprometidos em ajudar você a ter sucesso. AWS fornece a profundidade e a amplitude de ferramentas e recursos para ajudá-lo a criar, migrar e operar com maior facilidade e velocidade, onde quer que suas cargas de trabalho residam. AWS as ferramentas simplificam o gerenciamento entre provedores e, ao mesmo tempo, maximizam o desempenho e o valor de seus investimentos em nuvem.

Este paper se concentra em princípios comprovados para o sucesso com uma estratégia multicloud, incluindo quando e onde uma abordagem multicloud faz sentido e como AWS ajuda as empresas a terem sucesso com suas estratégias multicloud. Ele fornece orientação prescritiva para ajudar os executivos a fazer escolhas informadas de estratégia e tomada de decisão relacionadas à adoção da multicloud. Este paper não oferece uma discussão técnica aprofundada sobre implementações de multicloud. Para suporte técnico de implementação e assistência com seus desafios específicos, recomendamos que você trabalhe com seu arquiteto de AWS soluções.

Este paper apresenta nove princípios comprovados para o sucesso da multicloud com base em nossas experiências com clientes AWS corporativos. Cada princípio aborda um aspecto crítico da estratégia multicloud, desde o alinhamento das metas de negócios até a implementação da segurança. Ao aplicar esses princípios, as organizações podem lidar com a complexidade da multicloud com confiança.

Princípio 1. Alinhe as metas de multicloud à sua estratégia

Pesquisas do Gartner e tendências do setor mostram que as organizações estão adotando cada vez mais abordagens multicloud para atender às necessidades comerciais específicas. Os cenários a seguir demonstram quando uma infraestrutura multicloud pode ser estrategicamente vantajosa.

Fusões e aquisições

As fusões e aquisições (M&A) criam decisões imediatas sobre a estratégia de nuvem. Embora operar várias nuvens possa aumentar os custos e a complexidade, a rápida consolidação pode atrasar o valor da integração e interromper as operações comerciais. Suas decisões de nuvem se tornam fundamentais para obter os benefícios de fusões e aquisições.

O planejamento da integração deve levar em conta todo o cenário tecnológico. Cada carga de trabalho exige uma avaliação dentro do contexto do cronograma de integração e das prioridades de negócios.

Nossa orientação:

- Desenvolva uma estratégia de consolidação orientada aos negócios que equilibre as necessidades imediatas de integração com a eficiência operacional de longo prazo. Mantenha várias nuvens inicialmente em circunstâncias em que a consolidação rápida possa interromper operações comerciais críticas ou atrasar a realização do valor de fusões e aquisições.
- Crie critérios claros de posicionamento da carga de trabalho que se alinhem ao seu cronograma de integração. Priorize os aplicativos geradores de receita e os principais processos de negócios e, ao mesmo tempo, contabilize as dependências técnicas e os requisitos operacionais.

Desejo de aproveitar os recursos diferenciados de longo prazo de outro CSP

O medo de perder algo faz com que algumas empresas queiram um pouco de cada nuvem. As decisões de posicionamento da carga de trabalho afetam toda a organização, das equipes de engenharia às operações financeiras e de segurança.

Portanto, as organizações precisam examinar seu raciocínio para buscar várias nuvens. Alguns argumentam que cada carga de trabalho deve residir no provedor de serviços em nuvem (CSP)

que melhor atenda às suas necessidades. No entanto, a otimização individual da carga de trabalho deve ser equilibrada em relação ao impacto organizacional mais amplo. Cada provedor de nuvem adicional corre o risco de aumentar a complexidade operacional, criar novos requisitos de talentos e introduzir considerações de segurança que afetam toda a organização de tecnologia.

Nossa orientação:

- Siga uma abordagem 80/20: selecione um provedor principal para a maioria das cargas de trabalho e considere fornecedores adicionais somente para casos de uso específicos e de alto valor. Essa estratégia maximiza a eficiência e a retenção de talentos, ao mesmo tempo que reduz a complexidade.
- Considere o custo total da operação nas nuvens. Inclua ferramentas de segurança, produtos de governança, sistemas de gerenciamento financeiro e sobrecarga operacional em sua análise.
- Avalie as dependências e interações de cada carga de trabalho. As cargas de trabalho raramente operam isoladamente; elas compartilham dados, controles de segurança e processos operacionais.
- Conduza uma análise completa de preço-desempenho em todos os fornecedores. Compare não apenas os custos diretos, mas também a sobrecarga do gerenciamento de vários ambientes.

Multinuvem na holding e nuvem primária na empresa operacional ou linha de negócios

Firmas de capital privado e holdings enfrentam considerações exclusivas sobre a estratégia de nuvem. As empresas de seu portfólio geralmente mantêm estratégias de nuvem independentes, frequentemente resultantes de atividades anteriores de fusões e aquisições. Essa estrutura reduz a complexidade normalmente associada às operações multicloud, porque cada unidade de negócios opera de forma independente. No entanto, essa independência pode limitar as oportunidades de aproveitar os descontos por volume e os incentivos de compra em toda a empresa.

A eficácia da estratégia de nuvem no nível da holding depende da autonomia das empresas do portfólio e de suas necessidades tecnológicas individuais. Embora a consolidação possa criar alavancagem de compra, ela pode entrar em conflito com o modelo de operação independente típico de holdings e portfólios de capital privado.

Nossa orientação:

- Entenda as estruturas de descontos por volume do CSP. Cada provedor oferece mecanismos para adicionar ou remover subsidiárias de contratos corporativos e dividir unidades de negócios em entidades separadas. Elas representam decisões bidirecionais.
- Planeje cuidadosamente os compromissos de compra da nuvem. Envolva a equipe de contas do seu CSP com antecedência ou entre em contato AWS Partner com alguém da competência AWS em operações de nuvem para obter assistência.
- Equilibre independência com eficiência. Considere serviços compartilhados ou contratos de compra que beneficiem as empresas do portfólio sem restringir suas operações.
- Concentre-se primeiro nos objetivos de negócios. Desenvolva estratégias de tecnologia que suportem seu modelo operacional, em vez de adotar uma estratégia de multicloud por si só.
- Avalie as estratégias de nuvem sob a ótica do gerenciamento de portfólio. Considere como as opções de nuvem afetam possíveis alienações ou aquisições futuras.

Princípio 2. Esteja atento aos equívocos sobre a multicloud

Ao desenvolver sua estratégia de multicloud, evite os equívocos comuns discutidos nas seções a seguir.

Todos estão adotando estratégias multicloud

Empresas de consultoria e empresas de mídia traçam um quadro complexo da adoção da multicloud. Pesquisas mostram amplo interesse em abordagens multicloud, mas os padrões de gastos geralmente contam uma história diferente. Na prática, muitas empresas mantêm ambientes de nuvem únicos ou relacionamentos claros com o primary/secondary CSP. Essa desconexão destaca a importância de olhar além das manchetes e focar nas necessidades específicas de sua organização.

Nossa orientação:

- Tome decisões de nuvem com base em seus requisitos comerciais específicos, em vez de seguir as tendências do setor. Concentre-se em custos e riscos mensuráveis para sua organização.
- Examine casos de uso de multicloud dentro do contexto do seu setor. As estratégias de nuvem que funcionam para empresas de tecnologia de consumo podem não se traduzir em serviços financeiros, manufatura ou ambientes de jogos.
- Considere a gravidade dos dados como um fator principal nas decisões de posicionamento da carga de trabalho. A localização e a movimentação dos dados geralmente determinam a arquitetura de nuvem mais eficaz.
- Veja além das estatísticas de adoção para entender os padrões de gastos. As altas taxas de adoção da multicloud relatadas geralmente mascaram os padrões reais de gastos.
- Avalie as restrições técnicas antes de se comprometer com um ambiente multicloud. Algumas cargas de trabalho têm melhor desempenho quando seus componentes permanecem em um único ambiente de nuvem.

A multicloud reduz o risco de dependência de um fornecedor

A flexibilidade do fornecedor é uma consideração legítima no desenvolvimento da estratégia de nuvem. As organizações valorizam a capacidade de adaptar suas opções de tecnologia à medida que as necessidades dos negócios evoluem. Essa preocupação reflete experiências anteriores

com investimentos tradicionais em TI que criaram compromissos vinculativos de longo prazo. Os serviços em nuvem oferecem dinâmicas diferentes em relação à flexibilidade do provedor. AWS fornece serviços compatíveis de código aberto e opções de portabilidade de dados que reduzem as barreiras técnicas à migração. No entanto, a compensação entre flexibilidade e eficiência operacional continua sendo importante. As organizações devem avaliar o valor comercial de manter as opções do provedor em relação às vantagens técnicas de uma integração profunda com serviços especializados de um provedor primário.

Alguns clientes tentam evitar o aprisionamento projetando soluções independentes de nuvem que usam contêineres. Essa abordagem geralmente os restringe a serviços básicos de computação e armazenamento e ignora as vantagens dos recursos avançados de nuvem. Nossa experiência mostra que essa estratégia adiciona uma complexidade considerável devido ao aumento do tempo de desenvolvimento e dos recursos necessários, em comparação com o uso de serviços nativos.

Nossa orientação:

- Considere o custo total das arquiteturas independentes de nuvem. A sobrecarga adicional de engenharia pode não justificar os benefícios da portabilidade.
- Use recursos nativos da nuvem para obter o máximo valor. Os serviços básicos de computação e armazenamento, por si só, geralmente sacrificam vantagens significativas em segurança, escalabilidade e inovação.
- Planeje estratégias de nuvem com base nos requisitos de negócios. Quando uma implementação multicloud agrega um valor claro, como a capacidade de atender usuários em várias plataformas, o investimento adicional em engenharia vale a pena.
- Avalie cenários e custos de saída realistas. Compare a probabilidade e as despesas de mudar de fornecedor com os benefícios de usar o conjunto completo de Serviços da AWS.
- Construa sobre as bases de código aberto do AWS. AWS serviços gerenciados, como o Amazon Relational Database Service (Amazon RDS), oferecem flexibilidade e excelência operacional, além de oferecer suporte aos mecanismos de banco de dados que você usa atualmente.
- Aproveite as ferramentas abrangentes de migração fornecidas pela AWS. Ajudamos você a mover as cargas de trabalho em qualquer direção e fornecemos a saída gratuita de dados se você deixar de AWS usar outros provedores. Para obter mais informações, consulte a postagem do AWS blog Transferência gratuita de dados para a Internet ao sair de AWS.

A multicloud melhora a disponibilidade e a resiliência

A crença na troca perfeita da carga de trabalho entre provedores de nuvem durante interrupções leva algumas organizações a adotar estratégias multicloud. Essa mentalidade cria uma visão simplificada da resiliência da infraestrutura de nuvem que ignora as realidades técnicas fundamentais.

Com base em anos de experiência trabalhando com clientes multinuvem AWS, vimos que manter a portabilidade total da carga de trabalho entre provedores geralmente cria uma complexidade substancial sem oferecer todos os benefícios esperados. Os aplicativos com uso intensivo de dados enfrentam desafios insuperáveis devido às restrições da gravidade dos dados. Na verdade, em nossa opinião, é quase impossível para as organizações implementarem com sucesso um failover multicloud verdadeiramente perfeito para cargas de trabalho com muitos dados.

Lydia Leong, renomada vice-presidente de análise do Gartner, reforça essa perspectiva em um post nas redes sociais: “O failover multicloud é complexo e caro a ponto de quase sempre ser impraticável, e não é uma forma especialmente eficaz de lidar com os riscos de resiliência da nuvem”. A diferenciação inerente entre provedores de rede, armazenamento, bancos de dados, aprendizado de máquina e segurança torna a verdadeira portabilidade quase impossível. A distribuição das cargas de trabalho entre os provedores pode aumentar o risco, pois uma falha em qualquer um dos ambientes pode provocar uma interrupção em todos os ambientes.

Nossa orientação:

- Concentre-se em dominar os AWS recursos para cargas de trabalho individuais em vez de buscar arquiteturas multicloud complexas.
- Crie resiliência por meio de zonas Regiões da AWS de disponibilidade em vez de tentar o failover entre fornecedores. Para uma análise técnica aprofundada sobre como AWS fazer o failover automático de cargas de trabalho entre data centers físicos, consulte a postagem do AWS blog, Mudança automática zonal — afaste automaticamente seu tráfego das zonas de disponibilidade quando detectamos possíveis problemas.
- Migre cargas de trabalho estrategicamente e concentre-se em um aplicativo por vez para maximizar o sucesso. AWS

Multicloud oferece melhores preços

A competitividade de preços pode ser o argumento mais fraco de todos para ambientes multicloud. As experiências das organizações com contratos complicados e caros de software ou data center

que as prendem a contratos de vários anos as tornaram cautelosas ao adquirir serviços de TI. As abordagens tradicionais de aquisição não se adaptaram às pay-as-you-go compras, aos descontos por volume ou à realidade da concorrência de preços na nuvem. (Em janeiro de 2025, AWS reduziu os preços 151 vezes desde a sua criação.)

O maior fator único de redução de custos é um ambiente de nuvem bem gerenciado e otimizado. Uma empresa vê uma melhor otimização de custos trabalhando principalmente com um provedor cujos serviços oferecem vantagens de preço-desempenho (como instâncias de computação baseadas em chips personalizados, como AWS Graviton) e tem soluções superiores de gerenciamento financeiro em nuvem. De acordo com um estudo de 2022 do Hackett Group com mais de 1.000 organizações, os gastos com infraestrutura como porcentagem do total de gastos com TI foram 20% menores para AWS os clientes em comparação com organizações multicloud.

Nossa experiência mostrou que as empresas não prevêem o custo adicional e a complexidade de operar em várias nuvens, nem avaliam adequadamente esse custo em relação ao ganho percebido em um head-to-head contrato de fornecimento.

Nossa orientação:

- Crie sua estratégia de otimização de custos no pilar de otimização de custos do AWS Well-Architected Framework. Há cinco princípios de design:
  - Implemente o gerenciamento financeiro na nuvem: para alcançar o sucesso financeiro e acelerar a realização do valor comercial na nuvem, você deve investir no gerenciamento financeiro na nuvem. Sua organização deve dedicar o tempo e os recursos necessários para criar aptidão nesse novo domínio de gerenciamento de utilização e tecnologia. Assim como sua capacidade de segurança ou operações, você precisa aumentar as capacidades por meio da construção de conhecimento, programas, recursos e processos para ajudar a se tornar uma organização econômica.
  - Adotar um modelo de consumo: pague apenas pelos recursos de computação que você consome e aumente ou diminua o uso dependendo dos requisitos da empresa. Por exemplo, ambientes de desenvolvimento e teste normalmente são usados apenas oito horas por dia durante a semana de trabalho. Você pode interromper esses recursos quando eles não estiverem em uso para uma economia potencial de 75% (40 horas versus 168 horas).
  - Avalie a eficiência geral: meça a produção comercial de sua carga de trabalho e os custos associados à entrega. Use esses dados para entender os ganhos obtidos com o aumento da saída, o aumento da funcionalidade e a redução de custos.

  - Pare de gastar dinheiro em trabalhos pesados indiferenciados: CSPs faça o trabalho pesado das operações do data center, como montar, empilhar e alimentar servidores. Eles também eliminam a carga operacional de gerenciar sistemas operacionais e aplicativos usando serviços gerenciados. Isso permite que você se concentre em seus clientes e projetos de negócios em vez de na infraestrutura de TI.
  - Analisar e atribuir gastos: a nuvem facilita a identificação precisa do custo e uso das workloads, o que permite a atribuição transparente de custos de TI para fluxos de receita e proprietários de workloads individuais. Dessa forma, a medição do retorno sobre o investimento (ROI) é facilitada e os proprietários de workloads têm a oportunidade de otimizar recursos e reduzir custos.
- Dada a sobrecarga financeira de operar em diferentes fornecedores, orientamos os clientes a investirem pesadamente em ferramentas de automação e otimização de custos. Cada CSP oferece ferramentas nativas abrangentes nessa área, como o. Hub de Otimização de Custos da AWS A maioria das ferramentas nativas fornece excelentes recursos para os clientes em seu ambiente de nuvem. No entanto, para entender os gastos em vários CSPs, você pode escolher entre um rico conjunto de produtos ISV e software como serviço (SaaS) que ampliam esses recursos para fornecer uma experiência única de otimização de custos.
- Diluir o poder de compra por meio de uma estratégia de equidade de gastos não gera valor comercial. Isso pode prejudicar possíveis descontos por volume e potencialmente prejudicar o design técnico. A maneira mais eficiente de consumir serviços em nuvem é usar um provedor primário para a maior parte de suas operações e usar outro CSPs somente quando ele agrega valor comercial.

Princípio 3. Tenha uma estratégia e governança claras para apoiá-la

A decisão de adotar uma estratégia de multicloud é insuficiente; você deve estabelecer uma estratégia para cumprir seus objetivos, incluindo uma governança clara para quais cargas de trabalho irão para onde e por quê. Os critérios de avaliação devem ser usados para otimizar as cargas de trabalho e suas dependências. Se a avaliação for deixada para indivíduos, uma dispersão descoordenada provavelmente CSPs corroerá o valor da estratégia multicloud. Recomendamos que você avalie o desempenho da carga de trabalho do CSP regularmente e use sua avaliação como uma entrada importante para a seleção, os critérios e o uso futuro do CSP.

Uma estratégia de governança eficaz exige visibilidade do número total de serviços, aplicativos e componentes usados em toda a empresa. Parte integrante disso é uma estratégia de marcação robusta que abrange CSPs e estabelece propriedade, uso e ambiente claros (como desenvolvimento, controle de qualidade, preparação e produção) para todos os recursos implantados. Tudo deve ser marcado para um proprietário; se não estiver marcado ou se o proprietário não puder ser identificado, ele deve ser removido. Trabalhamos em estreita colaboração com uma grande organização de serviços financeiros que encontra e remove automaticamente todos os recursos não marcados e considera isso uma prática recomendada, independentemente do inconveniente que isso represente às equipes de desenvolvimento. Essa abordagem de marcação codifica as regras de governança e automatiza a fiscalização em vez de criar obstáculos ao progresso (ou seja, implementa grades de proteção, não barreiras). Custo, operações e segurança devem ser monitorados, monitorados e aplicados da mesma forma, com a mesma profundidade de dados e transparência. CSPs

Quando você implementa uma estratégia multicloud, estabelecer uma estrutura de contas clara e consistente entre os provedores de nuvem é crucial para manter o controle operacional e a segurança. Recomendamos a adoção de um hub-and-spoke modelo em que você crie unidades de negócios separadas Contas da AWS para diferentes unidades de negócios. Elas são ancoradas por duas contas centrais críticas: uma security/audit conta para monitoramento consolidado de conformidade e segurança e uma conta de rede central para gerenciar a interconectividade. (Essa abordagem é codificada no design do AWS Control Tower. No entanto, os princípios de privilégio mínimo e separação de deveres são igualmente aplicáveis a outras nuvens. O AWS Well-Architected Framework discute esses conceitos detalhadamente e é altamente recomendado para o público técnico.) Essa abordagem fundamental deve ser espelhada em todos os provedores de nuvem para manter a consistência na governança e nas operações. As contas de carga de trabalho devem ser

organizadas por ambiente (desenvolvimento, preparação, produção) ou função, com processos claros estabelecidos para criação e exclusão de contas.

Nossa orientação:

- Implemente uma estratégia abrangente de marcação para manter padrões claros de propriedade e uso em todos os recursos da nuvem. Monitore ambientes, centros de custos, aplicativos e unidades de negócios por meio de políticas de etiquetagem consistentes. Remova os recursos que não têm etiquetas adequadas para aplicar os padrões de governança e manter a clareza do ambiente.
- Estabeleça uma estrutura de conformidade unificada que mapeie os requisitos regulatórios em seu ambiente multicloud. Mantenha uma documentação clara de como os controles e certificações de cada provedor de nuvem apoiam suas obrigações de conformidade.
- Automatize a fiscalização da governança por meio da automação, em vez de usar processos de aprovação manual. Codifique suas regras de governança em sistemas automatizados que evitam violações de políticas antes que elas ocorram. Isso elimina o erro humano e, ao mesmo tempo, mantém a velocidade de desenvolvimento.
- Estruture contas em um hub-and-spoke modelo com segurança centralizada e controle de rede. Crie contas dedicadas para auditoria de segurança e gerenciamento de rede para centralizar funções críticas. Essa base permite políticas de segurança e conectividade de rede consistentes em toda a organização.
- Para manter os limites operacionais, crie contas, assinaturas ou projetos separados (dependendo da nomenclatura do seu CSP) para diferentes ambientes e funções. Divida as cargas de trabalho por ambientes de desenvolvimento, preparação e produção. Essa separação evita que incidentes de segurança se espalhem e mantém domínios operacionais claros.
- Monitore custos, operações e segurança por meio de métricas consistentes em todo o ambiente. Implemente monitoramento unificado para utilização de recursos, eventos de segurança e padrões de gastos. Use esses dados para otimizar o posicionamento da carga de trabalho e as decisões de alocação de recursos.
- Evite o uso não autorizado da nuvem por meio de políticas organizacionais e controles automatizados. Defina processos claros para criação de contas e provisionamento de recursos. Implemente políticas de controle de serviços (SCPs) para garantir a conformidade com os padrões organizacionais em todas as contas.
- Estabeleça controles preventivos e de detetive para evitar que a TI paralela surja por meio de contas de fornecedores não autorizados. Monitore o uso não autorizado da nuvem por meio de

relatórios de despesas e tráfego de rede. Bloqueie o acesso não autorizado de fornecedores e, ao mesmo tempo, mantenha caminhos aprovados para inovação.

Princípio 4. Não distribua cargas de trabalho contíguas pelas nuvens

A distribuição de cargas de trabalho contíguas entre vários provedores de nuvem cria complexidade, risco e custo desnecessários. Quando as cargas de trabalho que processam e analisam dados em conjunto abrangem vários fornecedores, as organizações enfrentam desafios na movimentação, sincronização e consistência dos dados. As equipes devem navegar por interfaces de gerenciamento APIs, modelos de segurança e processos operacionais diferentes para cada provedor, o que aumenta a probabilidade de erros e aumenta a sobrecarga operacional. Essa complexidade aumenta as chances de erros e a sobrecarga operacional e pode prejudicar a agilidade e a escalabilidade.

No entanto, em alguns cenários práticos, as organizações podem precisar distribuir cargas de trabalho contíguas nas nuvens devido a requisitos comerciais ou técnicos específicos. Nesses casos, recomendamos que você estabeleça critérios claros e princípios orientadores para avaliar as compensações e garantir que a abordagem esteja alinhada com a estratégia multicloud geral da sua organização.

Quando as organizações optam por distribuir cargas de trabalho em várias nuvens, adotar uma arquitetura centrada em mensagens e acoplamento flexível pode aliviar muitos dos desafios associados. Essa é a melhor maneira de separar as preocupações entre as nuvens e reduzir o escopo do impacto se um provedor for prejudicado. O ideal é que as operações com maior limite de tempo, como transações financeiras, sejam mantidas em um único ambiente. Nunca se deve permitir que uma interrupção em um ambiente coloque em risco as cargas de trabalho em outro ambiente.

Nossa orientação:

- Projete cargas de trabalho na nuvem para obter independência operacional e minimizar as dependências em tempo real entre os fornecedores. Quando a distribuição da carga de trabalho for necessária, implemente mecanismos eficientes de transferência de dados em massa em vez de manter conexões constantes entre nuvens.
- Avalie cada carga de trabalho distribuída proposta com base em critérios comerciais claros. Considere os benefícios estratégicos e a complexidade operacional introduzidos pela distribuição.

Princípio 5. Tenha uma estratégia de integração de longo prazo

Tenha cuidado ao mover grandes volumes de dados entre aplicativos em diferentes nuvens, especialmente se seus recursos computacionais e aplicativos estiverem implantados em um CSP e seus recursos de armazenamento de dados estiverem implantados em outro. Essa situação pode aumentar a complexidade e a latência que podem compensar os benefícios percebidos. Conversamos com muitos clientes que têm um data lake em uma nuvem, mas desejam realizar aprendizado de máquina (ML) ou análises com ferramentas de outro CSP. Decidir onde colocar as cargas de trabalho em um ambiente multicloud é uma das decisões mais cruciais — e muitas vezes mais desafiadoras — que as organizações enfrentam. Recomendamos que você avalie cada decisão de posicionamento da carga de trabalho por meio de três dimensões críticas: requisitos técnicos, necessidades comerciais e pontos fortes do fornecedor.

Inicie as avaliações técnicas mapeando as características essenciais de cada carga de trabalho: poder computacional, operações de dados, necessidades de tempo de resposta e requisitos de crescimento. Naturalmente, os aplicativos têm melhor desempenho quando estão localizados perto de seus dados. Afastar os aplicativos de suas fontes de dados cria obstáculos técnicos desnecessários e diminui o desempenho.

As decisões de negócios devem levar em conta os preços do provedor, os requisitos de residência de dados e os contratos do fornecedor. Cada posicionamento da carga de trabalho afeta as operações, a segurança e a produtividade de toda a organização. Analisar as cargas de trabalho isoladamente leva a decisões abaixo do ideal.

Nossa orientação:

- Implemente a transferência de dados em massa entre nuvens em vez do acesso em tempo real. Agende a atualização periódica de dados usando operações em massa eficientes em vez de usar chamadas de API constantes entre nuvens. Essa abordagem reduz custos, melhora a confiabilidade e mantém um desempenho consistente. Por exemplo, exporte dados de vendas diárias resumidos em vez de consultar transações individuais nas nuvens.
- Considere a gravidade dos dados ao projetar o posicionamento da carga de trabalho. Mantenha os aplicativos próximos às suas fontes de dados primárias para manter o desempenho e reduzir os custos. Modelos de ML, mecanismos de análise e sistemas de processamento de transações se beneficiam do acesso direto aos seus dados. Afastar essas cargas de trabalho de seus dados cria latência e complexidade de rede desnecessárias.

- Avalie as decisões de carga de trabalho dentro do contexto de sua estratégia de nuvem completa, em vez de analisá-las isoladamente. Considere como cada opção de posicionamento afeta os processos operacionais, os controles de segurança e as capacidades da equipe em toda a organização. Uma decisão que parece ideal para uma única carga de trabalho pode complicar o monitoramento ou aumentar os riscos de segurança quando vista de forma holística.
- Defina políticas claras de propriedade e governança de dados que especifiquem onde os diferentes tipos de dados podem residir. Crie uma estrutura de classificação de dados que conduza decisões consistentes sobre o posicionamento dos dados entre os provedores de nuvem.

Princípio 6. Use contêineres estrategicamente

Os contêineres podem desempenhar um papel valioso no suporte de uma estratégia de multicloud, mas também é importante reconhecer suas limitações. Usar contêineres geralmente é uma boa ideia para qualquer aplicativo moderno e nativo da nuvem, pois eles oferecem benefícios de portabilidade e consistência em diferentes ambientes. Os contêineres são independentes de plataforma, o que significa que podem ser executados em qualquer plataforma ou infraestrutura de nuvem que ofereça suporte à tecnologia de conteinerização, como o Kubernetes. As organizações que usam contêineres podem desenvolver e empacotar seus aplicativos uma vez e depois implantá-los de forma consistente em vários provedores de nuvem ou ambientes locais, sem a necessidade de modificações significativas. Ao encapsular o código do aplicativo, as dependências e o ambiente de execução em um contêiner, você pode obter um alto grau de portabilidade, o que permite mover cargas de trabalho sem problemas entre provedores de nuvem ou entre a nuvem e os data centers locais.

No entanto, os contêineres podem não resolver todos os casos de uso ou eliminar todos os desafios que uma organização pode enfrentar ao adotar uma estratégia multicloud. Os contêineres funcionam melhor com arquiteturas modernas baseadas em microsserviços, mas podem não ser tão adequados para aplicativos grandes e monolíticos. Além disso, embora os contêineres possam abordar certos aspectos da portabilidade, como o tempo de execução do aplicativo, eles não resolvem automaticamente problemas relacionados ao gerenciamento de dados, políticas de segurança e outras dependências entre nuvens. As organizações ainda precisam planejar e arquitetar cuidadosamente suas soluções multicloud para garantir gerenciamento consistente de dados, controles de segurança unificados e integração perfeita entre componentes hospedados na nuvem e no local.

Nossa orientação:

- Use os recursos nativos de gerenciamento de contêineres de cada provedor de nuvem para maximizar o valor comercial e acelerar a entrega. Essa abordagem garante um desempenho ideal e, ao mesmo tempo, evita a complexidade de criar soluções independentes da nuvem que raramente oferecem retornos significativos.
- Desenvolva estratégias de contêineres que abordem o quadro operacional completo, incluindo gerenciamento de dados, segurança e dependências entre nuvens. Concentre-se nos resultados comerciais ao tomar decisões de arquitetura de contêineres.

Princípio 7. Tenha um único CCo E, mas se especialize nele

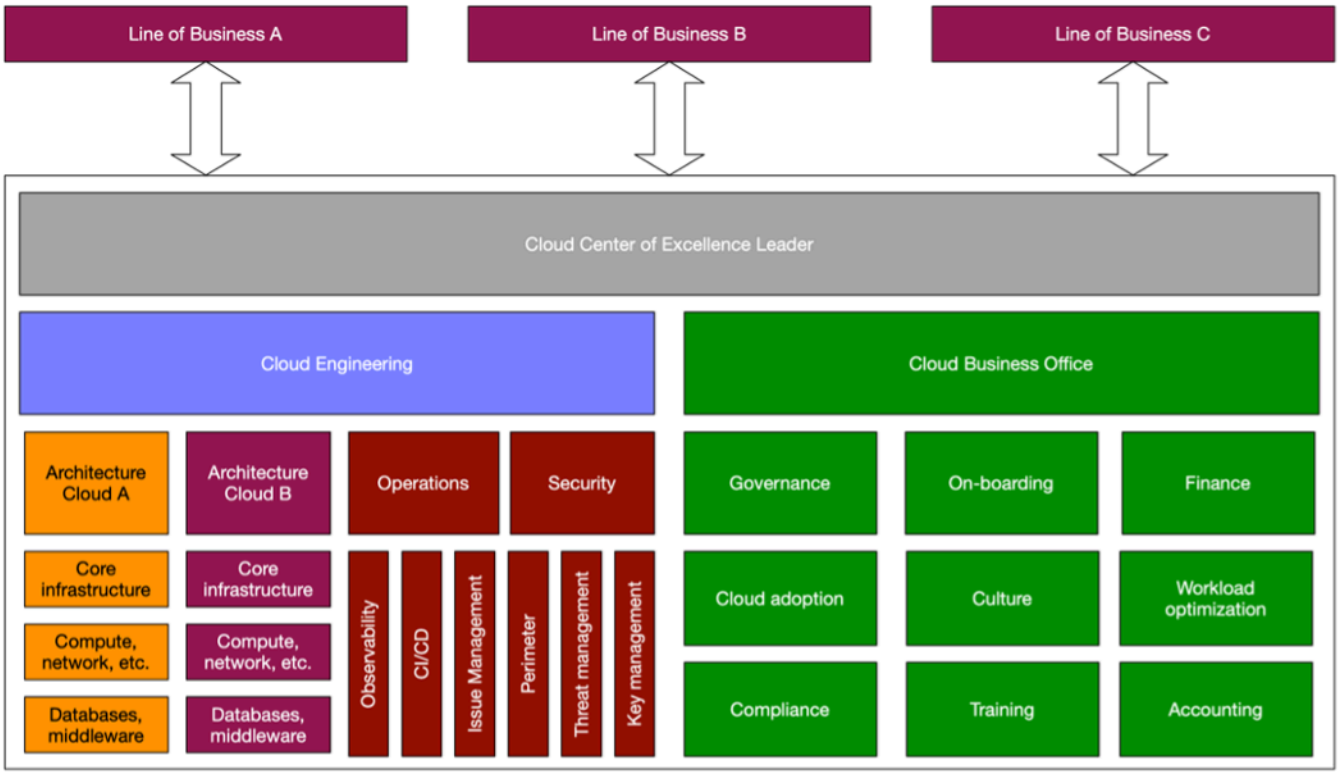
Como aconselhamos muitos AWS clientes, você deve criar um Centro de Excelência em Nuvem (CCoE) em sua organização para fornecer liderança, padronização e aceleração de sua jornada na nuvem. Quando se trata de ambientes multicloud, descobrimos que as empresas mais bem-sucedidas adotam uma abordagem equilibrada com seu E. CCo

Em vez de estabelecer CCo Es separados para cada CSP, recomendamos que você tenha um CCo E único e unificado que supervisione a estratégia multicloud da organização. Isso ajuda a garantir uma abordagem coordenada e consistente, em vez de esforços isolados que podem levar à divergência, reengenharia e desperdício. Certifique-se de que as equipes de seu único CCo E tenham as habilidades, ferramentas e mecanismos especializados necessários para cada CSP que sua organização usa. Esse conhecimento especializado permite que a CCo E governe, apoie e acelere o uso das diferentes plataformas de nuvem de forma eficaz.

Por exemplo, a CCo E deve ter especialistas AWS específicos que entendam profundamente os serviços e as melhores práticas, bem como especialistas CSPs que possam orientar o uso dessas tecnologias de nuvem pela organização. Nuvem AWS Essa experiência especializada em um único CCo E pode ajudar sua organização a se beneficiar da coordenação e padronização de uma abordagem centralizada e, ao mesmo tempo, garantir que cada plataforma de nuvem seja usada de forma otimizada.

O único CCo E deve servir como o órgão regulador central que estabelece padrões, políticas e melhores práticas para a estratégia multicloud da organização. A implementação real de cargas de trabalho e projetos na nuvem pode ser distribuída para equipes especializadas ou unidades de negócios, enquanto o CCOE fornece supervisão, suporte e coordenação. Essa abordagem equilibrada ajuda a garantir uma estratégia multicloud coesa e, ao mesmo tempo, fornece o grau necessário de flexibilidade e autonomia dentro da organização.

O diagrama a seguir ilustra como um CCo E pode fornecer uma abordagem e governança centralizadas em várias linhas de negócios (LOBs), equipes de engenharia de nuvem e equipes do Cloud Business Office (CBO).



Nossa orientação:

- Estruture seu CCo E para manter a supervisão estratégica e, ao mesmo tempo, incorporar conhecimentos especializados para cada provedor de nuvem. Concentre-se em recrutar uma profunda experiência em plataformas de nuvem individuais, em vez de procurar especialistas raros em multicloud, e promova o compartilhamento interno de conhecimento para desenvolver capacidades organizacionais.
- Capacite seu CCo E a estabelecer padrões corporativos para questões transversais, como segurança e observabilidade, ao mesmo tempo em que dá às equipes individuais a autonomia para executar de acordo com essas diretrizes usando ferramentas e serviços nativos da nuvem.
- Desenvolva uma estratégia abrangente de talentos que equilibre a profunda experiência em plataformas de nuvem primárias com um conhecimento arquitetônico mais amplo. Concentre-se na formação de equipes que combinem habilidades sólidas e específicas da nuvem com experiência em arquitetura corporativa.

Princípio 8. Certifique-se de que a segurança seja sempre uma prioridade máxima

Uma abordagem multicloud dificulta a garantia da segurança, aumentando o risco de acesso não autorizado, porque sua postura de segurança deve levar em conta mais superfícies de ataque. Uma estratégia multicloud geralmente força as empresas a lidar com vários modelos de segurança CSPs em áreas como gerenciamento de identidade, segurança de rede, gerenciamento de ativos e registro de auditoria. Essa complexidade pode dificultar a transparência, aumentar a carga sobre as equipes de segurança e elevar os riscos.

A automação da segurança é essencial em ambientes multicloud. O gerenciamento de identidade deve funcionar perfeitamente em todos os ambientes; ele deve conectar os provedores de identidade existentes e, ao mesmo tempo, manter políticas de acesso consistentes. A segurança exige proteção integrada nas camadas de dados, rede e endpoint. A classificação de dados, a criptografia e o gerenciamento do ciclo de vida formam a base. A segurança de rede se baseia em designs e padrões de conexão padronizados. A proteção de terminais completa a estrutura por meio de gerenciamento consistente de patches e controles baseados em host.

Esses elementos fundamentais são essenciais para a adoção segura e bem-sucedida de vários provedores de nuvem e devem ser considerados logo no início de qualquer planejamento estratégico de multicloud.

Nossa orientação:

- Implemente uma estrutura de segurança integrada em seu ambiente multicloud que se concentre em três elementos principais: proteção de dados por meio de classificação e criptografia padronizadas, segurança de rede por meio de padrões de design consistentes e proteção de terminais por meio de controles sistemáticos e gerenciamento de patches.
- Estabeleça um modelo unificado de operações de segurança que aproveite os recursos de segurança nativos de cada provedor de nuvem, mantendo a visibilidade e o controle centralizados por meio de ferramentas e processos padronizados.
- Centralize a coleta e a análise de dados de segurança usando o Amazon Security Lake. Essa plataforma agrega informações de segurança de AWS outros provedores de nuvem, aplicativos SaaS e sistemas locais em uma única visualização. Ele suporta o Open Cybersecurity Schema Framework (OCSF) e permite análises padronizadas em seu ambiente híbrido e multicloud.

Essa abordagem centralizada melhora a detecção e a resposta a ameaças e, ao mesmo tempo, simplifica as operações de segurança.

- Implante as ferramentas de segurança nativas de cada provedor para aprimorar seus recursos de proteção. Esses serviços desenvolvidos especificamente abordam os recursos específicos do provedor e, ao mesmo tempo, fornecem dados à sua plataforma de segurança centralizada. Uma combinação de ferramentas nativas e visibilidade centralizada ajuda a fornecer cobertura de segurança abrangente em toda a sua infraestrutura.
- Implemente uma estratégia unificada de observabilidade que forneça visibilidade abrangente em todo o seu cenário de nuvem, incluindo dados operacionais e de segurança, desde o início. Padronize as abordagens de monitoramento líderes do setor que permitem o rastreamento consistente dos serviços comerciais, independentemente de onde operem.
- Estabeleça padrões corporativos para coleta e visualização de dados operacionais que permitam a rápida identificação e resolução de problemas em seu ambiente multicloud. Concentre-se em criar uma única fonte confiável para insights operacionais que atenda às partes interessadas técnicas e comerciais.

Princípio 9. Adote uma abordagem 80/20 sobre distribuição igualitária

A forma como você distribui as cargas de trabalho entre os provedores determina fundamentalmente seu sucesso na multicloud. Muitas organizações buscam erroneamente a igualdade em sua distribuição na nuvem e tentam distribuir as cargas de trabalho uniformemente entre os provedores. Essa abordagem aumenta a complexidade sem oferecer benefícios proporcionais. A distribuição igualitária fragmenta suas capacidades técnicas, dilui seu poder de compra e cria uma sobrecarga operacional desnecessária. As equipes lutam para desenvolver um profundo conhecimento quando são forçadas a manter a competência em várias plataformas simultaneamente.

A abordagem 80/20 oferece resultados comprovadamente melhores do que a distribuição igualitária nas nuvens. Concentrar 80% do seu investimento em um provedor principal e, ao mesmo tempo, usar seletivamente outros para recursos específicos cria uma estratégia equilibrada que reduz o custo e a complexidade. Essa abordagem concentrada acelera a inovação porque suas equipes podem desenvolver uma profunda experiência com os serviços avançados da sua plataforma principal. Sua equipe técnica pode se tornar especialista em uma arquitetura em vez de manter o conhecimento superficial em vários ambientes. Quando os engenheiros dominam uma plataforma, eles constroem com mais eficiência, solucionam problemas com mais rapidez e implementam soluções mais sofisticadas.

As empresas que seguem a abordagem 80/20 geralmente relatam uma melhor retenção de talentos porque suas equipes desenvolvem conhecimentos valiosos e comercializáveis, em vez de se limitarem a várias tecnologias. Essa estratégia concentrada também ajuda a simplificar o gerenciamento da segurança, limitando a complexidade dos diferentes modelos de segurança entre os fornecedores. A nuvem primária recebe a maior parte do seu investimento em ferramentas de segurança, soluções de monitoramento e processos operacionais. Isso cria uma base de segurança mais forte do que a possível com recursos igualmente divididos.

Nossa orientação:

- Selecione um provedor de nuvem principal que se alinhe à maioria dos seus requisitos comerciais e técnicos. Esse provedor deve suportar pelo menos 80% de suas cargas de trabalho e se tornar a base de sua estratégia de nuvem. Concentre seus investimentos em treinamento, padrões arquitetônicos e processos operacionais na maximização do valor dessa plataforma principal.
- Desenvolva critérios claros para cargas de trabalho que garantam a colocação em nuvens secundárias. Esses critérios devem se concentrar no valor comercial específico que não pode

ser alcançado em seu provedor principal. Resista à colocação de cargas de trabalho em nuvens secundárias simplesmente para manter a equidade de gastos ou o equilíbrio artificial entre os fornecedores.

- Estruture seus contratos corporativos para refletir sua abordagem 80/20. Negocie descontos por volume com seu provedor principal com base em gastos concentrados e mantenha a flexibilidade com fornecedores secundários para casos de uso específicos. Essa abordagem maximiza sua alavancagem de compra e normalmente resulta em melhores preços gerais do que dividir seus gastos igualmente.
- Alinhe sua estratégia de talentos com sua abordagem 80/20. Invista no desenvolvimento de uma profunda experiência com os serviços do seu provedor principal e, ao mesmo tempo, mantenha conhecimento suficiente das plataformas secundárias para suportar cargas de trabalho específicas. Essa estratégia focada em talentos melhora a produtividade, acelera a entrega e reduz o risco de lacunas críticas de habilidades.
- Avalie regularmente os resultados comerciais de sua estratégia de multicloud. Acompanhe métricas que demonstram o valor obtido de cada provedor e ajuste sua distribuição, se necessário. O objetivo não é evitar totalmente a multicloud, mas implementá-la estrategicamente, onde cargas de trabalho específicas realmente se beneficiem de recursos exclusivos de outros provedores.

Conclusão

Este paper descreveu nove princípios fundamentais para o desenvolvimento de uma estratégia multicloud eficaz. As organizações alcançam o maior sucesso por meio de uma abordagem de nuvem primária com o uso estratégico de fornecedores adicionais quando as necessidades comerciais específicas assim o exigirem. A abordagem 80/20 que descrevemos equilibra foco com flexibilidade e permite que as organizações desenvolvam uma experiência mais profunda, mantenham relacionamentos mais fortes com fornecedores e criem talentos mais valiosos, ao mesmo tempo em que atendem aos requisitos legítimos de multicloud.

A implementação bem-sucedida da multicloud exige uma avaliação clara das necessidades comerciais, em vez de seguir as tendências do setor. As empresas devem estabelecer uma governança robusta, manter a segurança como prioridade máxima, evitar a distribuição de cargas de trabalho conectadas entre provedores, manter os aplicativos com seus dados transacionais, reconhecer as limitações dos contêineres e manter um Centro de Excelência em Nuvem unificado, mas especializado.

A AWS abordagem da nuvem é fundamentalmente baseada na escolha e na interoperabilidade do cliente. Projetamos nossas ferramentas e serviços para funcionarem perfeitamente em todos os ambientes porque entendemos que suas necessidades comerciais geralmente vão além de um único fornecedor. De soluções de conectividade híbrida à orquestração de contêineres que abrange ambientes, AWS oferece recursos que ajudam você a operar com eficiência em todo o seu cenário tecnológico.

Em vez de forçar você a se tornar especialista em várias plataformas, AWS simplifica o gerenciamento multicloud por meio de ferramentas intuitivas e interfaces consistentes. Nosso foco é eliminar a complexidade para que você possa se concentrar na inovação. Esses recursos ajudam você a implementar sua estratégia de multicloud em seus próprios termos, seja usando AWS exclusivamente ou usando ambientes específicos Serviços da AWS em conjunto com outros ambientes.

A nuvem deve fortalecer sua estratégia de negócios, não restringi-la. Ao aplicar os princípios descritos neste paper e aproveitar os recursos de AWS interoperabilidade, você pode criar uma abordagem de nuvem que maximize o valor, minimize a complexidade desnecessária e posicione sua organização para o sucesso a longo prazo no ambiente de negócios dinâmico de hoje.

Para saber mais sobre AWS soluções que podem ajudar a simplificar o gerenciamento em ambientes híbridos e multicloud, consulte AWS Soluções para multicloud.

Recursos

Referências

- Usando um Centro de Excelência em Nuvem (CCOE) para transformar toda a empresa (AWS postagem no blog)
- AWS Well-Architected Framework
- Identificação de oportunidades com o Cost Optimization Hub (AWS Cost Management documentação)
- O valor comercial da migração para a Amazon Web Services (The Hackett Group, fevereiro de 2022)
- Transferência gratuita de dados para a Internet ao sair AWS(postagem doAWS blog)

Ferramentas

- Mudança automática zonal — afaste automaticamente seu tráfego das zonas de disponibilidade quando detectamos possíveis problemas (AWS postagem no blog)
- AWS soluções para multicloud

AWS Parceiros

- Nuvem AWS Competência operacional

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	—	3 de setembro 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: migrar seu banco de dados Oracle on-premises para o Amazon Aurora Edição Compatível com 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: migrar seu banco de dados Oracle on-premises para o Amazon Relational Database Service (Amazon RDS) para Oracle na Nuvem AWS.
- Recomprar (drop and shop): mude para um produto diferente, normalmente migrando de uma licença tradicional para um modelo SaaS. Exemplo: migrar seu sistema de gerenciamento de relacionamento com o cliente (CRM) para o 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: migrar seu banco de dados Oracle on-premises para o Oracle em uma instância do EC2 na 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 on-premises para um serviço de 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 atributo.

serviços abstraídos

Veja 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 em que 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.

AGGREGATE FUNCTION

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 inteligência artificial.

AIOps

Veja 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 aplicações

Uma abordagem de segurança que permite o uso somente de aplicações aprovadas 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 malicioso

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

BCP

Veja 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 da aplicação em um ambiente (azul) e a nova versão da aplicação no outro ambiente (verde). Essa estratégia ajuda você a reverter rapidamente com o mínimo de impacto.

bot

Uma aplicação de software que executa tarefas automatizadas na internet e simula a atividade ou interação humana. Alguns bots são úteis ou benéficos, como crawlers da web que indexam informações na internet. Outros bots, conhecidos como bots maliciosos, têm como objetivo causar disrupção ou danos a indivíduos ou organizações.

botnet

Redes de bots infectadas por malware e sob o controle de uma única parte, conhecidas como bot herder ou operador de bots. Os botnets 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 de emergência

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 Implement break-glass procedures nas orientações do AWS Well-Architected.

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

Veja AWS Cloud Adoption Framework.

implantação canário

O lançamento lento e incremental de uma versão para usuários finais. Quando estiver confiante, você implanta a nova versão e substitui a versão atual por completo.

CCoE

Veja Centro de Excelência da Nuvem.

CDC

Veja captura de dados de alteração.

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 integração e 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 corporativa.

computação em nuvem

A tecnologia de nuvem normalmente usada para armazenamento de dados remoto e gerenciamento de dispositivos de IoT. A computação em nuvem é normalmente conectada à tecnologia de computação de borda.

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 ao migrar para a 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

Veja 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 o GitHub ou o Bitbucket 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 de IA que usa machine learning para analisar e extrair informações de formatos visuais, como vídeos e imagens digitais. Por exemplo, a Amazon SageMaker AI fornece algoritmos de processamento de imagem para CV.

desvio de configuração

Em uma workload, uma alteração de configuração em relação ao estado esperado. Isso pode fazer com que a workload 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 é comumente descrito como um pipeline. CI/CD pode 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.

data mesh

Um framework de arquitetura 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 compatível com business intelligence, como analytics. 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

Veja 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 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 normalmente são 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 disastre. Para obter mais informações, consulte Recuperação de desastres de cargas de trabalho em AWS: Recuperação na nuvem no AWS Well-Architected Framework.

DML

Veja 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 recuperação de desastres.

Detecção da oscilação

Rastreamento de desvios de uma configuração de linha de base. 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 mapeamento do fluxo de valor de desenvolvimento.

E

EDA

Veja 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 com a computação em nuvem, a computação de borda 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 EDI (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 endpoint de 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 empresariais (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 CI/CD pipeline, 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 planejamento de recursos empresariais.

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. Ela armazena dados quantitativos sobre as operações comerciais. Normalmente, uma tabela de fatos contém dois tipos de colunas: as que contêm medidas e as que contêm uma chave externa para uma tabela de dimensões.

Antecipar-se à falha

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

delimitação de isolamento contra 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 AWS Fault Isolation Boundaries.

ramificação de recursos

Veja ramificação.

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.

prompt few shot

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 em contexto, em que os modelos aprendem com exemplos (shots) incorporados aos prompts. Prompts few-shot podem ser eficazes para tarefas que exigem formatação, raciocínio ou conhecimento de domínio específicos. Veja também prompts zero-shot.

FGAC

Veja 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 via captura de dados de alteração para migrar os 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 modelo de base.

modelo de base (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 de base?.

G

IA generativa

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

bloqueio geográfico

Veja 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 trunk-based é a abordagem moderna e preferencial.

golden image

Um snapshot de um sistema ou software usado como modelo para implantar novas instâncias desse sistema ou software. Por exemplo, na manufatura, uma golden image 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 CSPM, Amazon GuardDuty AWS Trusted Advisor, Amazon Inspector e verificações personalizadas AWS Lambda .

H

HA

Veja 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 hold-out

Uma parte dos dados históricos rotulados que são retidos de um conjunto de dados usado para treinar um modelo de machine learning. Você pode usar dados de hold-out para avaliar a performance do modelo comparando as predições 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 normal de DevOps lançamento.

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 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 Internet das Coisas Industrial.

infraestrutura imutável

Um modelo que implanta uma nova infraestrutura para workloads de produção em vez de atualizar, aplicar patches ou modificar a infraestrutura existente. Infraestruturas imutáveis são inerentemente mais consistentes, confiáveis e preditivas do que infraestruturas mutáveis. Para obter mais informações, consulte a prática recomendada Implantar usando infraestrutura imutável no AWS Well-Architected 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 manufatura por meio de avanços em conectividade, dados em tempo real, automação, analytics 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

Veja 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

Veja biblioteca de informações de TI.

ITSM

Veja 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.

grande modelo de linguagem (LLM)

Um modelo de IA de aprendizado profundo pré-treinado em uma grande quantidade de dados. Um LLM pode realizar várias tarefas, como responder a 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 rótulo.

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 grande modelo de linguagem.

ambientes inferiores

Veja 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 ramificação.

Malware

Software projetado para comprometer a segurança ou a privacidade do computador. O malware pode interromper os sistemas do computador, vazar informações sensíveis ou obter acesso não autorizado. Exemplos de malware incluem vírus, worms, ransomware, cavalos de Troia, 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 abstraídos.

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

Veja Programa de Aceleração da Migração.

mecanismo

Um processo completo em que 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 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 o Amazon EC2 AWS com o 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 a 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 workload para a Nuvem AWS. Para obter mais informações, veja a entrada 7 Rs neste glossário e consulte Mobilize sua organização para acelerar migrações em grande escala.

ML

Veja machine learning.

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 Strategy for modernizing applications in the 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 Evaluating modernization readiness for applications in the 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.

MPA

Veja Avaliação do Portfólio para Migração.

MQTT

Veja 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 workloads 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 controle de acesso de origem.

OAI

Veja identidade de acesso de origem.

OCM

Veja gerenciamento de alterações 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 integração de operações.

Ola

Veja acordo de 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

Veja Open Process Communications - Unified Architecture.

Open Process Communications - Unified Architecture (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 práticas recomendadas 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 AWS Well-Architected Framework.

tecnologia operacional (TO)

Sistemas de hardware e software que trabalham com o ambiente físico para controlar operações, equipamentos e infraestrutura industriais. Na manufatura, a integração dos sistemas de tecnologia da informação (TI) e tecnologia operacional (TO) 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 análise de prontidão operacional.

OT

Veja 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 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

Veja controlador lógico programável.

PLM

Veja gerenciamento do ciclo de vida do produto.

política

Um objeto que pode definir permissões (veja política baseada em identidade), especificar condições de acesso (veja políta baseada em recurso) ou definir as permissões máximas para todas as contas em uma organização no AWS Organizations (veja 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.

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 cláusula `WHERE`.

pushdown de predicados

Uma técnica de otimização de consultas 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 a performance 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 em engenharia de sistemas 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 desenvolvido para evitar a implantação de recursos não conformes. Esses controles verificam 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 a concepção, o desenvolvimento e o lançamento, passando pelo crescimento e maturidade, até o declínio e a remoção.

ambiente de produção

Veja ambiente.

controlador lógico programável (PLC)

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

encadeamento de prompts

Uso da 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 em que outros microsserviços possam assinar. 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, aprovador, consultado, informado (RACI).

RAG

Veja geração aumentada via 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, aprovador, consultado, informado (RACI).

RCAC

Veja 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.

Redefinir arquitetura

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 informações, consulte Specify which Regiões da AWS your account can use.

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 uma aplicação de resistir ou se recuperar de interrupções. Alta disponibilidade e recuperação de desastres são considerações comuns ao planejar a resiliência na Nuvem AWS. Para obter mais informações, consulte Nuvem AWS Resilience.

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.

Retirada

Veja 7 Rs.

Geração Aumentada de Recuperação (RAG)

Uma tecnologia de IA generativa em que 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 (geração aumentada via recuperação)?.

alternância

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

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 objetivo de ponto de recuperação.

RTO

Veja objetivo de 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 Console de gerenciamento da AWS 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 política de controle de serviço.

secret

Em AWS Secrets Manager, informações confidenciais ou restritas, como uma senha ou credenciais de usuário, que você armazena de forma criptografada. 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 What's in a Secrets Manager secret? na documentação do Secrets Manager.

segurança desde a concepção

Uma abordagem em engenharia de sistemas que leva em consideração 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 primários de controles de segurança: preventivos, detectivos, responsivos e proativos.

hardening 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 aplicação de patches em uma instância do Amazon EC2 ou a alternância 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 avaliação de um aspecto de performance de um serviço, como taxa de erro, disponibilidade ou throughput.

objetivo de nível de serviço (SLO)

Uma métrica alvo que representa a integridade de um serviço, conforme avaliado 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 sistema de gerenciamento de eventos e informações de segurança.

ponto único de falha (SPOF)

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

SLA

Veja cacordo de serviço.

SLI

Veja indicador de nível de serviço.

SLO

Veja objetivo de 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 Phased approach to modernizing applications in the Nuvem AWS.

SPOF

Veja ponto único de falha.

esquema em estrela

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 ser usada 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 supervisório 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 a performance. Você pode usar o Amazon CloudWatch Synthetics para criar esses testes.

prompt do sistema

Uma técnica para fornecer contexto, instruções ou orientações a um LLM a fim de direcionar seu comportamento. Os prompts do sistema ajudam a definir o contexto e a 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 da . 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 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.

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 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 backend.

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.

WORM

Veja gravação única e várias leituras.

WQF

Veja AWS Workload Qualification Framework.

gravação única e várias leituras (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, normalmente malware, que tira proveito de uma vulnerabilidade zero-day.

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.

prompt zero shot

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

eficácia dos prompts zero-shot depende da complexidade da tarefa e da qualidade do prompt. Veja também prompts few-shot.

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.