Melhores práticas para simplificar a observabilidade do Amazon EKS

AWS Orientação prescritiva

AWS Orientação prescritiva: Melhores práticas para simplificar a observabilidade do Amazon EKS

Table of Contents

Melhores práticas para simplificar a observabilidade do Amazon EKS

Ishwar Chauthaiwale, Naveen Suthar e Pratap Kumar Nanda, da Amazon Web Services ()AWS

Março de 2026 (histórico do documento)

O Amazon Elastic Kubernetes Service (Amazon EKS) exige soluções abrangentes de observabilidade para monitorar e solucionar problemas de cargas de trabalho em contêineres de forma eficaz. Sistemas distribuídos e microsserviços têm arquiteturas complexas nos ambientes Amazon EKS, portanto, implementar práticas adequadas de observabilidade é crucial para manter operações confiáveis. A observabilidade efetiva nos ambientes Amazon EKS permite que as equipes obtenham insights profundos sobre o desempenho do aplicativo, solucionem problemas com eficiência e mantenham a integridade ideal do cluster.

O desafio está em navegar pelo vasto ecossistema de ferramentas e técnicas disponíveis para a observabilidade do Amazon EKS e, ao mesmo tempo, aderir às melhores práticas que se alinham às metas organizacionais e aos padrões do setor. Estratégias eficazes de observabilidade devem equilibrar a coleta abrangente de dados com considerações de desempenho, custo-benefício e escalabilidade.

Este guia foi criado para ajudar as organizações a otimizar a observabilidade do Amazon EKS nas seguintes áreas:

- Estabelecendo mecanismos de registro eficientes
- Implementando soluções robustas de monitoramento
- Usando rastreamento distribuído para arquiteturas complexas
- Implementando estratégias de alerta e resposta a incidentes

Ao adotar essas melhores práticas, sua organização pode aprimorar sua capacidade de obter insights profundos sobre o ambiente Amazon EKS, o que leva a uma maior confiabilidade, desempenho e eficiência operacional. Essa abordagem simplificada de observabilidade ajuda na solução de problemas e na manutenção, além de apoiar a tomada de decisão baseada em dados para a melhoria contínua dos aplicativos e da infraestrutura baseados em Kubernetes. (Para obter informações detalhadas sobre o Amazon EKS, consulte a documentação do serviço.)

Este guia se aprofunda em cada aspecto da observabilidade do Amazon EKS e explora as ferramentas e estratégias que você pode personalizar para atender às necessidades específicas de suas implantações do Amazon EKS, desde aplicativos de pequena escala até arquiteturas de microsserviços grandes e complexas.

Neste guia:

Objetivos

Este guia pode ajudar você e sua organização a alcançar os seguintes objetivos comerciais:

- Visibilidade operacional aprimorada — Obtenha uma visão abrangente de seus clusters e aplicativos do Amazon EKS por meio de práticas eficazes de observabilidade.

  Esse objetivo enfatiza a importância de manter a visibilidade completa em todo o seu ambiente Amazon EKS. Ferramentas como AWS X-RayAmazon CloudWatch Container Insights e AWS Distro OpenTelemetry ajudam você a entender o comportamento do sistema, identificar problemas rapidamente e manter o desempenho ideal.
- Maior eficiência na solução de problemas — Reduza o tempo médio de detecção (MTTD) e o tempo médio de resolução (MTTR) por meio de estratégias eficazes de rastreamento e monitoramento.

  Esse objetivo se concentra na implementação de práticas de observabilidade que permitem a rápida identificação e resolução de problemas. Técnicas como rastreamento distribuído, registro efetivo e coleta abrangente de métricas são fundamentais para alcançar esse objetivo.
- Gerenciamento proativo do desempenho — Permita a detecção precoce de possíveis problemas antes que eles afetem os usuários finais.

O monitoramento proativo é crucial para manter a alta disponibilidade e o desempenho do serviço. Esse objetivo aborda a importância de implementar alertas, análises de tendências e monitoramento preditivo adequados para evitar interrupções no serviço.

- Observabilidade econômica — otimize os custos de observabilidade enquanto mantém uma visibilidade abrangente do sistema.

  A otimização de custos engloba a implementação de estratégias de amostragem eficientes, políticas apropriadas de retenção de dados e abordagens de instrumentação ideais. O objetivo é equilibrar as necessidades de observabilidade com as considerações de custo e, ao mesmo tempo, garantir o monitoramento eficaz do sistema.
- Arquitetura de monitoramento escalável — Certifique-se de que suas soluções de observabilidade escalem perfeitamente com seu ambiente Amazon EKS.

  Esse objetivo se concentra na implementação de soluções de monitoramento que possam crescer com seu aplicativo. Se você estiver executando um único cluster ou uma implantação multirregional em vários clusters, sua estratégia de observabilidade deve ser dimensionada adequadamente

Fazendo login no Amazon EKS

O registro em log é um aspecto essencial do gerenciamento e manutenção de aplicativos executados no Amazon EKS. Práticas eficazes de registro em ambientes Amazon EKS ajudam desenvolvedores, equipes de operações e administradores de sistemas a obter informações valiosas sobre o comportamento, o desempenho e a integridade de seus aplicativos em contêineres e de sua infraestrutura subjacente.

A implementação de uma estratégia de registro robusta no Amazon EKS é essencial por vários motivos:

- Solução de problemas: os registros ajudam a identificar e diagnosticar problemas rapidamente, o que reduz o tempo de inatividade e melhora a confiabilidade geral do sistema.
- Conformidade: muitos setores exigem registros abrangentes para fins de auditoria e regulamentação.
- Segurança: a análise de registros pode ajudá-lo a detectar e investigar possíveis ameaças ou violações de segurança.
- Otimização do desempenho: os registros fornecem informações sobre o desempenho do aplicativo e do sistema, para que você possa identificar gargalos e otimizar a utilização dos recursos.
- Monitoramento e alertas: os dados de registro podem ser usados para configurar sistemas de monitoramento e acionar alertas para eventos ou condições específicas.

Nesta seção:

Tipos de registro no Amazon EKS

No Amazon EKS, o registro em log envolve capturar, armazenar e analisar vários tipos de dados de log gerados por diferentes componentes do cluster Kubernetes, incluindo:

- Registros do sistema: informações sobre as instâncias ou nós subjacentes do Amazon Elastic Compute Cloud (Amazon EC2) AWS Fargate

- Registros de componentes do Kubernetes : dados dos principais componentes do Kubernetes, como o servidor da API, o agendador e o gerenciador do controlador
- Registros de tempo de execução do contêiner: informações do tempo de execução do contêiner, como Docker ou containerd
- Registros de aplicativos: saída de aplicativos em contêineres

Para gerenciar registros em seu ambiente Amazon EKS de forma eficaz, você normalmente emprega uma combinação de Serviços da AWS ferramentas de terceiros e melhores práticas. Isso pode incluir o uso da Amazon CloudWatch, Fluent Bit, Elasticsearch, Kibana e outras ferramentas de registro e análise para coletar, armazenar e visualizar dados de log.

As seções a seguir exploram vários aspectos do registro no Amazon EKS, incluindo melhores práticas, ferramentas e técnicas para implementar uma estratégia de registro abrangente em seus clusters do Kubernetes em. AWS

Logs do sistema.

O registro em log para instâncias EC2 subjacentes ou nós Fargate no Amazon EKS envolve abordagens diferentes, dependendo do tipo de nó.

Para implementar o registro em log para instâncias do EC2 no Amazon EKS, você pode usar as seguintes ferramentas:

- CloudWatch agente: instale e configure o CloudWatch agente em suas instâncias do EC2. Configure-o para coletar registros do sistema, como `/var/log/messages` `/var/log/secure` e. Você pode usar scripts de dados do usuário ou ferramentas de gerenciamento de configuração para automatizar esse processo.
- Fluent Bit: implante o Fluent Bit como um DaemonSet para coletar registros de todos os nós. Configure-o para encaminhar registros para o CloudWatch Logs ou outros sistemas de registro centralizados.
- Container Insights: habilite o Container Insights em seu cluster EKS para coletar automaticamente métricas e registros de instâncias do EC2.
- Scripts personalizados: desenvolva scripts personalizados para coletar registros específicos e enviá-los ao seu destino de registro preferido.
- Agente SSM: Use o AWS Systems Manager Agente (Agente SSM) para coletar e encaminhar registros para CloudWatch o Logs.

Para implementar o registro em log para nós do Fargate no Amazon EKS, use estas ferramentas:

- Registro em Fargate: o Fargate coleta `stdout` e `stderr` registra automaticamente de seus contêineres. Configure seu perfil do Fargate para enviar esses registros para CloudWatch o Logs.
- Fluent Bit for Fargate AWS : fornece uma imagem Fluent Bit especificamente para o registro em Fargate. Implante-o como um contêiner auxiliar em seus pods do Fargate para coletar e encaminhar troncos.
- Container Insights para Fargate: habilite o Container Insights para coletar métricas e registros dos nós do Fargate.

Registros de componentes do Kubernetes

A coleta de registros de componentes do Kubernetes, como o servidor de API, o agendador e o gerenciador de controladores no Amazon EKS, exige uma abordagem um pouco diferente da geração de registros de aplicativos. Esses componentes são executados como parte do plano de controle do Amazon EKS, que é gerenciado pelo AWS. Veja como você pode coletar e acessar esses registros:

- Ativar o registro do plano de controle: você pode ativar o registro do plano de controle para seu cluster EKS por meio das ferramentas Console de gerenciamento da AWS, AWS Command Line Interface (AWS CLI) ou infraestrutura como código (IaC), como AWS CloudFormationo Terraform. Quando você ativa o registro no plano de controle, os registros são enviados para o Amazon CloudWatch Logs. Você pode visualizá-los no CloudWatch console no grupo de `/aws/eks/<cluster-name>/cluster` registros. Dentro desse grupo de registros, cada componente do plano de controle tem seu próprio fluxo de registros da seguinte forma:

Nome do fluxo	Description
servidor kube-api	Registros do servidor da API Kubernetes
programador de cubos	Registros de decisão do agendador
kube-controller-manager	Registros do gerenciador do controlador
autenticadora	Registros do autenticador do IAM

Nome do fluxo	Description
audit	Registros de auditoria do Kubernetes (devem ser ativados explicitamente)

Para visualizar os registros de um componente específico, navegue até o grupo de registros do cluster e filtre pelo nome do fluxo de registros de destino.

- Use o CloudWatch Logs Insights: você pode usar o CloudWatch Logs Insights para realizar consultas complexas em seus registros.
- Exportar registros para o Amazon S3: Para armazenamento a longo prazo ou análises adicionais, você pode exportar registros para o Amazon Simple Storage Service (Amazon S3).
- Use ferramentas de terceiros: você pode usar ferramentas como o Fluent Bit para coletar esses registros e encaminhá-los para outros sistemas de registro, como Elasticsearch ou Splunk.
- Uso AWS CloudTrail: o AWS CloudTrailserviço pode fornecer informações adicionais sobre as chamadas de API feitas para seu cluster EKS.

Registros de tempo de execução do

O registro de registros de tempo de execução do contêiner no Amazon EKS envolve a captura e o gerenciamento de registros do tempo de execução do contêiner, o que normalmente é `containerd` para o Amazon EKS. Veja como você pode abordar o registro de registros de tempo de execução de contêineres no Amazon EKS:

- Acesse diretamente os registros nos nós do Amazon EC2. Para nós EC2 autogerenciados, você pode acessar diretamente os registros de tempo de execução do contêiner no host a partir destes locais:
  - `containerd`registros: `/var/log/containers/`
  - Registros do Docker (se você estiver usando o tempo de execução do Docker): `/var/log/docker.log`
- Use um DaemonSet para coleta de registros.
- Implante um agente de coleta de registros (como o Fluent Bit) DaemonSet para coletar registros de todos os nós.
- Configure o CloudWatch agente para coletar registros de tempo de execução do contêiner.

- Ative o Container Insights para coletar métricas e registros de tempo de execução do contêiner.
- Use o Fargate. Para os nós do Fargate, os registros de tempo de execução do contêiner são coletados automaticamente e podem ser acessados por meio CloudWatch de registros.
- Implemente soluções de registro personalizadas usando ferramentas como Fluent Bit ou Logstash. Configure CloudWatchalarmes ou use ferramentas como o Prometheus para monitorar padrões ou problemas específicos nos registros de tempo de execução do contêiner. Considere usar soluções de registro de terceiros que se integrem bem ao Kubernetes e ao Amazon EKS, como Datadog, Splunk ou Elastic Stack (ELK Stack). Use ferramentas de agregação de registros para coletar registros de várias fontes e encaminhá-los para um sistema de registro centralizado.

Logs de aplicações

Os registros de aplicativos no Amazon EKS são uma parte crucial da manutenção e solução de problemas de seus aplicativos. Para implementar o registro de aplicativos no Amazon EKS, você pode escolher entre estas opções:

- Grave registros em`stdout`/`stderr`: A maneira mais simples e nativa do Kubernetes de lidar com registros de aplicativos é gravá-los em e. `stdout` `stderr` O Kubernetes captura automaticamente esses streams.
- Implemente a agregação de registros: use um agregador de registros, como o Fluent Bit, para coletar registros de todos os seus pods.
- Configurar o roteamento de registros: configure seu agregador de registros para rotear os registros para o destino desejado (como CloudWatch Logs ou Elasticsearch).
- Use o CloudWatch Container Insights: habilite o Container Insights para registro e monitoramento abrangentes.

Melhores práticas para fazer login no Amazon EKS

As melhores práticas a seguir ajudam a criar um sistema de registro robusto, escalável e eficiente para seu ambiente Amazon EKS e fornecem melhores soluções de problemas, monitoramento e gerenciamento geral de seus clusters Kubernetes.

- Centralize a coleta de registros: use uma solução de registro centralizada, como CloudWatch Logs, Elasticsearch ou um serviço de terceiros, para agregar registros de todos os componentes. Isso fornece um único ponto de acesso para análise de registros e simplifica o gerenciamento.

- Implemente registros estruturados: use formatos de registro estruturados, como JSON, para que os registros possam ser analisados e pesquisados com mais facilidade. Inclua metadados relevantes, como registros de data e hora, níveis de registro e identificadores de origem.
- Use os níveis de log adequadamente: implemente níveis de log adequados (como `DEBUGINFO`,`WARN`, e`ERROR`) em seus aplicativos. Configure ambientes de produção para registrar em níveis apropriados para evitar registros excessivos.
- Ativar o registro de contêineres: configure seus contêineres para fazer login em `stdout` `stderr` e. Isso permite que o Kubernetes capture e encaminhe esses registros para a solução de registro escolhida.
- Habilitar o registro de aplicativos: configure aplicativos para gravar registros `stderr` em `stdout` e em vez de gravar em arquivos de log. Isso segue a metodologia de aplicativo de 12 fatores e se alinha às melhores práticas nativas da nuvem.
- Use o Kubernetes DaemonSets para coleta de registros: implante agentes de coleta de registros (como o Fluent Bit) DaemonSets para garantir que eles sejam executados em todos os nós do seu cluster.
- Implemente políticas de retenção: defina e aplique políticas de retenção de registros para cumprir as normas e gerenciar os custos de armazenamento.
- Dados de registro seguros: criptografe registros em trânsito e em repouso. Implemente controles de acesso para restringir quem pode ver e gerenciar registros.
- Monitore a ingestão de registros: configure alertas para falhas ou atrasos na ingestão de registros para garantir o registro contínuo.
- Use anotações e rótulos do Kubernetes: use anotações e rótulos do Kubernetes para adicionar metadados aos seus registros e melhorar a capacidade de pesquisa e a filtragem.
- Implemente rastreamento distribuído: use ferramentas de rastreamento distribuído, como AWS X-Rayo Jaeger, para correlacionar registros em microsserviços.
- Otimize o volume de registros: seja seletivo quanto ao que você registra para evitar custos desnecessários e problemas de desempenho. Use a amostragem para registros de alto volume e baixo valor.
- Implemente a agregação de registros: use ferramentas como o Logstash para agregar registros de várias fontes antes de enviá-los ao seu sistema de registro central.
- Use Serviços da AWS quando possível: serviços como CloudWatch Logs e Container Insights fornecem integração perfeita com outros Serviços da AWS.

- Implemente análise e visualização de registros: use ferramentas como CloudWatch Logs Insights, Elasticsearch com Kibana ou soluções de terceiros para análise e visualização de registros.
- Implemente a análise automatizada de registros: use ferramentas de aprendizado de máquina e inteligência artificial para detectar automaticamente anomalias e padrões em seus registros.
- Documente sua estratégia de registro: mantenha uma documentação clara de sua arquitetura, práticas e ferramentas de registro para sua equipe.

Considerações importantes para fazer login no Amazon EKS

Esta seção discute considerações importantes que você deve ter em mente ao implementar o registro no Amazon EKS.

- Impacto no desempenho: o registro excessivo pode afetar o desempenho do aplicativo. Esteja atento ao volume e à frequência dos registros gerados.
- Gerenciamento de custos: o armazenamento e o processamento de registros podem gerar custos significativos, especialmente em grande escala. Implemente políticas de retenção de registros e considere o uso da agregação de registros para reduzir custos.
- Segurança e conformidade: certifique-se de que os registros não contenham informações confidenciais, como senhas ou dados pessoais. Implemente criptografia para registros em trânsito e em repouso. Considere os requisitos de conformidade, como o Regulamento Geral de Proteção de Dados (GDPR) ou a Lei de Portabilidade e Responsabilidade de Seguros de Saúde (HIPAA), ao lidar com registros.
- Escalabilidade: certifique-se de que sua solução de registro possa ser dimensionada de acordo com o tamanho do cluster e o volume de registros. Considere o uso de buffering e agrupamento em lotes para transmissão de registros.
- Retenção de registros: defina e implemente períodos apropriados de retenção de registros. Equilibre os requisitos de conformidade com os custos de armazenamento.
- Controle de acesso: implemente funções e políticas AWS Identity and Access Management (IAM) adequadas para acesso aos registros. Siga o princípio do menor privilégio para o gerenciamento de registros.
- Consistência de registros: use formatos de registro consistentes em diferentes aplicativos e serviços. Use registros estruturados para facilitar a análise e a análise.
- Sincronização de horário: sincronize o tempo em todos os nós para obter registros de data e hora consistentes.

- Alocação de recursos: aloque recursos apropriados (como CPU e memória) para agentes de registro. Monitore o uso de recursos dos componentes de registro.
- Considerações sobre o Fargate: O Fargate tem mecanismos de registro específicos que diferem dos nós baseados em EC2. Entenda as limitações e os recursos do registro em log do Fargate.
- Clusters de vários inquilinos: em ambientes com vários inquilinos, certifique-se de que os registros estejam devidamente isolados entre os inquilinos.
- Análise e análise de registros: considere as ferramentas e habilidades necessárias para uma análise eficaz de registros. Implemente a análise de registros para extração estruturada de dados.
- Monitoramento do sistema de registro: configure o monitoramento da própria infraestrutura de registro. Gere alertas para registrar falhas ou atrasos no sistema.
- Impacto na rede: esteja ciente da largura de banda da rede usada pela transmissão de registros. Considere usar a compactação para dados de log.
- Eventos do Kubernetes: não negligencie os eventos do Kubernetes como fonte de informações importantes.
- Registro do plano de controle: entenda as implicações e os custos de habilitar o registro do plano de controle.
- Capacidades de depuração: certifique-se de que sua solução de registro permita a fácil depuração e solução de problemas.
- Integração com ferramentas existentes: considere como sua solução de registro do Amazon EKS se integra às ferramentas existentes de monitoramento e alerta.
- Teste: teste regularmente sua configuração de registro, especialmente após as atualizações do cluster.
- Documentação: mantenha uma documentação clara de sua arquitetura e práticas de registro.
- Latência de agregação de registros: esteja ciente de qualquer latência na agregação de registros e como ela pode afetar o monitoramento em tempo real.

Monitoramento no Amazon EKS

O monitoramento no Amazon EKS fornece visibilidade crítica sobre a integridade, o desempenho e a segurança de suas cargas de trabalho do Kubernetes. Sem o monitoramento adequado, você corre o risco de interrupções no serviço, violações de segurança e utilização ineficiente de recursos que podem afetar as operações comerciais e aumentar os custos. O monitoramento eficaz permite que você identifique e resolva problemas de forma proativa, otimize o uso de recursos e mantenha os requisitos de conformidade em seus aplicativos em contêineres. Ao implementar soluções de monitoramento abrangentes, você pode garantir alta disponibilidade, detectar anomalias precocemente e tomar decisões baseadas em dados para escalar e melhorar sua infraestrutura do Amazon EKS.

Esta seção explora os vários aspectos do monitoramento do Amazon EKS, incluindo diferentes tipos de monitoramento, ferramentas disponíveis e melhores práticas para ajudá-lo a criar uma estratégia de monitoramento robusta para seu ambiente Kubernetes.

Nesta seção:

Tipos de monitoramento no Amazon EKS

A observabilidade efetiva no Amazon EKS envolve atividades de monitoramento de infraestrutura, aplicativos e segurança.

Monitoramento da infraestrutura

O monitoramento da infraestrutura é um componente fundamental da observabilidade do Amazon EKS, que fornece insights profundos sobre a integridade e o desempenho dos elementos fundamentais do seu cluster Kubernetes. Basicamente, envolve rastrear os sinais vitais dos componentes do plano de controle e dos nós de trabalho e garantir que a plataforma subjacente permaneça estável e eficiente.

- O monitoramento do plano de controle é crucial porque supervisiona os principais componentes, como o servidor de API, o banco de dados etcd e o agendador. Ao monitorar a latência do servidor da API, você pode identificar rapidamente os gargalos de desempenho que podem afetar as implantações de aplicativos ou as operações de escalabilidade. O monitoramento de desempenho do Etcd valida se o banco de dados de estado do cluster opera com eficiência e evita problemas de consistência de dados que podem afetar todo o cluster.
- O monitoramento em nível de nó é igualmente essencial porque se concentra nos recursos computacionais que executam suas cargas de trabalho em contêineres. Isso inclui rastrear a utilização da CPU, o consumo de memória, a E/S de disco e o desempenho da rede em todos os nós de trabalho. A compreensão dessas métricas ajuda a evitar o esgotamento de recursos, otimizar as decisões de escalonamento de nós e garantir o planejamento adequado da capacidade.
- O monitoramento de rede desempenha um papel vital na manutenção da comunicação confiável entre pods, serviços e recursos externos. Ao monitorar a taxa de transferência, a latência e os estados de conexão da rede, você pode identificar problemas de conectividade com antecedência e garantir uma comunicação fluida com os aplicativos. O monitoramento do armazenamento complementa o monitoramento da rede rastreando o desempenho do volume, a utilização da capacidade e I/O os padrões, para ajudar a evitar gargalos relacionados aos dados.

O monitoramento da infraestrutura serve como um sistema de alerta antecipado para possíveis problemas, permite a manutenção proativa e garante a alocação ideal de recursos. Sem um monitoramento robusto da infraestrutura, você corre o risco de tempo de inatividade inesperado, desempenho degradado e uso ineficiente de recursos que podem afetar significativamente as operações e os custos dos negócios.

Monitoramento de aplicações

O monitoramento de aplicativos é essencial para manter aplicativos em contêineres saudáveis, com desempenho e confiabilidade em seu ambiente Amazon EKS. Esse nível de monitoramento se concentra nas cargas de trabalho reais que são executadas em seu cluster e fornece informações essenciais sobre como seus aplicativos se comportam, funcionam e interagem com outros serviços.

O monitoramento de aplicativos inclui monitoramento em nível de contêiner, monitoramento em nível de serviço e rastreamento distribuído.

- No nível do contêiner, o monitoramento de aplicativos rastreia métricas cruciais, como status de integridade do contêiner, contagens de reinicializações e padrões de consumo de recursos. Essas

métricas ajudam a identificar contêineres problemáticos que podem estar consumindo recursos excessivos ou enfrentando reinicializações frequentes, o que pode indicar problemas subjacentes, como vazamentos de memória ou problemas de configuração. Ao monitorar os eventos do ciclo de vida do contêiner, você pode garantir o comportamento adequado do aplicativo e solucionar rapidamente os problemas de implantação.

- O monitoramento em nível de serviço fornece visibilidade das métricas de desempenho e confiabilidade do aplicativo, como tempos de resposta, taxas de erro e taxa de transferência de solicitações. Essas métricas são vitais para manter os objetivos de nível de serviço (SLOs) e garantir uma experiência positiva para o usuário final. Você pode monitorar a latência em diferentes endpoints de serviço, identificar gargalos de desempenho e monitorar padrões de erro para manter a confiabilidade do aplicativo.
- O rastreamento distribuído é outro aspecto crítico do monitoramento de aplicativos, especialmente em arquiteturas de microsserviços. Ao implementar o rastreamento, você pode acompanhar as solicitações à medida que elas fluem por diferentes serviços, entender as dependências e identificar gargalos de desempenho. Essa end-to-end visibilidade ajuda você a otimizar as interações de serviço e solucionar problemas complexos que abrangem vários componentes.

As métricas personalizadas de aplicativos desempenham um papel crucial no fornecimento de insights específicos de negócios. Isso pode incluir métricas como taxas de processamento de pedidos, frequências de login de usuários ou taxas de sucesso de transações. Você pode correlacionar essas métricas personalizadas com métricas de infraestrutura e contêiner para entender melhor como o desempenho da infraestrutura afeta as operações comerciais e tomar decisões baseadas em dados para escalabilidade e otimização.

A importância do monitoramento de aplicativos está em sua capacidade de fornecer uma visão abrangente da integridade e do desempenho dos aplicativos. Esse monitoramento permite que você mantenha a alta qualidade do serviço, resolva problemas rapidamente e otimize continuamente seus aplicativos para atender aos objetivos de negócios.

Monitoramento de segurança

O monitoramento de segurança no Amazon EKS é uma atividade crítica que ajuda as organizações a manter a integridade, a confidencialidade e a conformidade de seus ambientes Kubernetes. Essa abordagem de segurança abrangente combina vigilância contínua, detecção de ameaças e monitoramento de conformidade para proteger cargas de trabalho em contêineres contra possíveis

riscos de segurança e acesso não autorizado. Inclui monitoramento de autenticação e autorização, monitoramento de segurança de rede e monitoramento de configuração e conformidade.

- O monitoramento de autenticação e autorização forma a primeira linha de defesa ao rastrear todas as tentativas de acessar o cluster. Isso inclui monitorar solicitações do servidor de API, rastrear tentativas de login bem-sucedidas e malsucedidas e auditar alterações no controle de acesso baseado em função (RBAC). Ao manter registros de auditoria detalhados de quem acessou quais recursos e quando, você pode detectar rapidamente possíveis violações de segurança, tentativas de acesso não autorizado ou atividades de escalonamento de privilégios. Isso é particularmente crucial em ambientes com vários inquilinos, onde manter controles de acesso rígidos é essencial.
- O monitoramento da segurança de rede se concentra em detectar e impedir a comunicação não autorizada entre pods e serviços. Ao monitorar violações de políticas de rede e padrões de tráfego incomuns, você pode identificar possíveis ameaças à segurança, como tentativas de fuga de contêineres ou movimentação lateral dentro do cluster. Isso inclui rastrear a comunicação interna do cluster e os padrões de tráfego externo para garantir que os contêineres se comuniquem somente com endpoints autorizados e sigam as políticas de segurança definidas.
- O monitoramento da configuração e da conformidade é essencial para manter as linhas de base de segurança e atender aos requisitos regulatórios. Ela envolve a varredura contínua de imagens de contêineres em busca de vulnerabilidades, o monitoramento da segurança do tempo de execução e o rastreamento de alterações na configuração que possam afetar a postura de segurança. As auditorias regulares de conformidade garantem a adesão aos padrões do setor e às políticas de segurança organizacionais, e a detecção de desvios de configuração ajuda a evitar alterações não autorizadas que podem introduzir riscos à segurança.

O monitoramento de segurança no Amazon EKS fornece a visibilidade e o controle necessários para ajudar a proteger contra ameaças de segurança modernas e, ao mesmo tempo, garantir a conformidade com os requisitos regulatórios. Ao implementar um monitoramento de segurança abrangente, sua organização pode manter uma postura de segurança forte, responder rapidamente a incidentes de segurança e demonstrar conformidade com vários padrões regulatórios.

Ferramentas de monitoramento para Amazon EKS

Esta seção discute três categorias de ferramentas de monitoramento do Amazon EKS: serviços de AWS monitoramento, soluções de código aberto ou proprietárias e ferramentas especializadas.

AWS serviços

- Amazon CloudWatch: serviço abrangente de monitoramento e registro

  CloudWatch forma a espinha dorsal das soluções de AWS monitoramento e fornece recursos abrangentes para ambientes Amazon EKS. Ele fornece o Container Insights para métricas granulares de contêineres e clusters, para que você possa monitorar o desempenho, a utilização de recursos e a integridade do aplicativo. O serviço é excelente em agregação e análise de registros e oferece suporte ao registro centralizado em contêineres e nós. CloudWatch se integra naturalmente com Serviços da AWS. Ele fornece configuração automática de alarmes e oferece suporte a métricas e painéis personalizados, o que o torna uma ferramenta essencial para o monitoramento do Amazon EKS.

- AWS X-Ray: Plataforma avançada de rastreamento distribuído

  O X-Ray aumenta a observabilidade ao fornecer recursos sofisticados de rastreamento distribuído. Sua visualização do mapa de serviços oferece uma visão clara sobre a arquitetura e as dependências do aplicativo, e o rastreamento detalhado de solicitações ajuda a identificar gargalos de desempenho em todos os serviços. O X-Ray pode rastrear solicitações por meio de arquiteturas complexas de microsserviços, o que o torna inestimável para solução de problemas e otimização, especialmente em sistemas distribuídos que abrangem vários. Serviços da AWS

- AWS Distro para OpenTelemetry: estrutura unificada de observabilidade

  O Distro for OpenTelemetry fornece recursos unificados de coleta de dados com suporte multiplataforma, o que o torna ideal para ambientes híbridos. Esse serviço se integra a outros Serviços da AWS, oferece suporte a instrumentação personalizada e oferece flexibilidade na implementação de soluções abrangentes de monitoramento, mantendo a compatibilidade com os padrões do setor.

- Amazon Managed Grafana: visualização de nível corporativo

  O Amazon Managed Grafana fornece um serviço totalmente gerenciado para visualização e análise de dados. Ele oferece integração perfeita com outros Serviços da AWS recursos de segurança integrados e escalabilidade de nível corporativo. O serviço simplifica a criação e o gerenciamento do painel, ao mesmo tempo em que fornece recursos avançados, como acesso a fontes de dados entre contas e integração com. Centro de Identidade do AWS IAM

- Amazon Managed Service para Prometheus: monitoramento gerenciado, seguro e altamente disponível

O Amazon Managed Service for Prometheus é um serviço de monitoramento totalmente gerenciado e compatível com o Prometheus. Ele fornece escalabilidade automatizada, alta disponibilidade e ingestão e consulta seguras de métricas. O serviço se integra perfeitamente ao Amazon EKS e elimina a sobrecarga operacional do gerenciamento dos servidores Prometheus.

Soluções de código aberto ou proprietárias

As AWS ferramentas descritas na seção anterior oferecem integração perfeita e serviços gerenciados. As ferramentas de código aberto listadas nesta seção complementam Serviços da AWS fornecendo flexibilidade e amplas opções de personalização. Compreender os recursos e os casos de uso de cada ferramenta ajuda você a projetar estratégias de monitoramento que melhor atendam aos seus requisitos específicos.

- Prometheus: kit de ferramentas de coleta de métricas

  O Prometheus é uma solução de código aberto para coleta de métricas em ambientes Kubernetes. Seu banco de dados de séries temporais e a linguagem de consulta PromQL permitem análises métricas sofisticadas. Os recursos de descoberta de serviços da plataforma se adaptam automaticamente aos ambientes dinâmicos do Kubernetes, e seu sistema de gerenciamento de alertas mantém você informado sobre problemas críticos. O Prometheus oferece amplas opções de integração, o que o torna uma opção versátil para monitoramento abrangente de métricas.

- Grafana: mecanismo de visualização avançado

  A Grafana transforma dados de monitoramento complexos em insights acionáveis por meio de seus recursos de visualização. A plataforma cria painéis personalizados que combinam dados de várias fontes e fornecem uma visão unificada das métricas de infraestrutura e aplicativos. Seu suporte para várias fontes de dados e recursos de gerenciamento de alertas fornecem monitoramento abrangente. O Grafana pode ajudá-lo a visualizar dados históricos e em tempo real, para que você possa identificar tendências e tomar decisões informadas.

- Fluent Bit: camada de registro unificada

  Essa solução de registro fornece coleta e gerenciamento de registros para ambientes Kubernetes. Sua integração nativa com o Kubernetes garante a coleta perfeita de registros de contêineres e nós, e seu suporte para vários destinos de saída oferece flexibilidade no armazenamento e análise de registros. Recursos avançados, como análise e filtragem de registros, permitem processar

e rotear registros com base em requisitos específicos. A natureza leve do Fluent Bit o torna particularmente adequado para ambientes em contêineres.

- Datadog: observabilidade em pilha completa

  O Datadog fornece recursos abrangentes de monitoramento com suporte nativo ao Kubernetes. Ele oferece monitoramento de infraestrutura, monitoramento de desempenho de aplicativos (APM), gerenciamento de registros e análises em tempo real. Você pode usar a descoberta automática de serviços e o extenso catálogo de integração da plataforma para o monitoramento do Amazon EKS e seus recursos de aprendizado de máquina para detectar anomalias e prever possíveis problemas.

- New Relic: monitoramento do desempenho de aplicativos

  A New Relic oferece visibilidade do desempenho do aplicativo e da integridade da infraestrutura. Sua integração com o Kubernetes fornece informações detalhadas sobre contêineres, rastreamento distribuído e painéis personalizados. A plataforma ajuda você a correlacionar o desempenho do aplicativo com as métricas da infraestrutura, para que você possa identificar e resolver problemas rapidamente.

- Elastic Stack (ELK Stack): análise e pesquisa de registros

  O ELK Stack combina Elasticsearch, Logstash e Kibana para fornecer recursos de gerenciamento e análise de registros. Ele oferece funcionalidade avançada de pesquisa, ferramentas de visualização e recursos de aprendizado de máquina. Você pode usar a pilha para lidar com grandes volumes de dados de log de seus ambientes Amazon EKS.

Ferramentas especializadas

Você pode combinar as seguintes ferramentas com base em seus requisitos específicos de monitoramento, escala de operações e preferências organizacionais. A chave é criar uma pilha de monitoramento que forneça visibilidade abrangente e, ao mesmo tempo, permaneça gerenciável e econômica.

- kube-state-metrics (KSM): monitoramento do estado do Kubernetes

  Esse serviço complementar escuta o servidor da API Kubernetes e gera métricas sobre o estado dos objetos. Ele fornece informações sobre a integridade das implantações, pods e outros recursos do Kubernetes.

- Kubernetes Metrics Server: métricas de recursos

Esse servidor de métricas coleta métricas de recursos dos kubelets e as expõe por meio da API de métricas do Kubernetes. Ele fornece escalonamento automático horizontal de pods e métricas básicas de CPU e memória.

- Kubecost: monitoramento de custos do Kubernetes

  Ferramentas como o Kubecost fornecem análises detalhadas de custos e recomendações de otimização para clusters EKS. Eles ajudam você a entender e otimizar os gastos com a nuvem em diferentes namespaces, implantações e serviços.

Implementando alta disponibilidade para soluções de monitoramento do Amazon EKS

Uma estratégia robusta de alta disponibilidade (HA) para o monitoramento do Amazon EKS é crucial para garantir a visibilidade contínua do seu ambiente Kubernetes. Esta seção discute uma abordagem abrangente para implementar HA em diferentes aspectos de sua infraestrutura de monitoramento.

Redundância e escalabilidade arquitetônicas

A construção de um sistema de monitoramento altamente disponível começa com um projeto arquitetônico adequado. Os componentes de monitoramento devem ser distribuídos em várias zonas de AWS disponibilidade para se protegerem contra falhas na zona. Isso inclui a implementação de escalabilidade horizontal para componentes críticos de monitoramento, como servidores Prometheus, coletores de registros e gerenciadores de alertas. Você pode usar serviços AWS gerenciados, como o Amazon Managed Service for Prometheus e o Amazon Managed Grafana, para ajudar a reduzir a sobrecarga operacional e, ao mesmo tempo, garantir a alta disponibilidade. Configure mecanismos de failover automático para manter a continuidade do serviço durante falhas nos componentes, com verificações de integridade e procedimentos de recuperação automatizados em vigor.

Estratégia resiliente de armazenamento de dados

A resiliência do armazenamento de dados é fundamental para manter a confiabilidade do sistema de monitoramento. A implementação de soluções de armazenamento distribuído garante que os dados e registros métricos permaneçam acessíveis mesmo se os nós de armazenamento individuais falharem. Isso inclui configurar a replicação adequada de dados em várias zonas de disponibilidade

e usar diferentes back-ends de armazenamento para redundância. Estabeleça procedimentos regulares de backup para dados históricos, com processos de recuperação documentados para vários cenários de falha. Para bancos de dados de séries temporais, como o Prometheus, a implementação de soluções de armazenamento remoto ajuda a separar as preocupações de armazenamento da coleta de dados e melhora a confiabilidade geral do sistema.

Gerenciamento redundante de alertas

O gerenciamento de alertas requer atenção especial em uma configuração de HA. A implantação de gerenciadores de alertas redundantes garante que as notificações críticas cheguem aos destinatários pretendidos mesmo durante falhas no sistema. Configure vários canais de notificação, como e-mail, SMS, Slack, e PagerDuty forneça caminhos alternativos de comunicação. Use mecanismos de desduplicação de alertas para evitar tempestades de alertas durante falhas parciais do sistema e métodos de notificação alternativa para garantir que alertas críticos nunca sejam perdidos. A implementação da correlação de alertas ajuda a manter o contexto durante cenários de failover e evita notificações duplicadas de sistemas redundantes.

Balanceamento de carga e descoberta de serviços

O balanceamento de carga adequado é essencial para manter os serviços de monitoramento estáveis. AWS Os Application Load Balancers distribuem o tráfego de monitoramento de entrada em vários endpoints, e as verificações de integridade garantem que o tráfego seja roteado somente para instâncias íntegras. Os mecanismos de descoberta de serviços ajudam os componentes de monitoramento a se adaptarem automaticamente às mudanças no ambiente, como a adição de novos nós ou serviços. Implemente agentes de monitoramento de forma consistente em todos os nós usando DaemonSets para garantir uma cobertura abrangente à medida que o cluster se expande.

Considerações adicionais de HA

Resiliência de rede:

- Implemente caminhos de rede redundantes.
- Configure o design adequado da sub-rede em todas as zonas de disponibilidade.
- Use AWS Direct Connectcom rotas de backup.
- Configure grupos de segurança e listas de controle de acesso à rede (rede ACLs) apropriados.

Monitorando os monitores:

- Implante sistemas de monitoramento secundários.
- Implemente o monitoramento entre regiões.
- Configure alertas para sistemas que não respondem.
- Teste os procedimentos de failover regularmente.

Planejamento de capacidade:

- Monitore as tendências de uso de recursos.
- Implemente o escalonamento preditivo.
- Teste o desempenho regularmente.

Gerenciamento de dados:

- Implemente políticas de retenção de dados.
- Configure a agregação métrica.
- Planeje o gerenciamento do ciclo de vida dos dados.
- Otimize o armazenamento regularmente.

Procedimentos de recuperação:

- Processos de recuperação de documentos.
- Teste a recuperação de desastres regularmente.
- Implemente a recuperação automatizada sempre que possível.
- Identifique e implemente caminhos claros de escalonamento.

Ao implementar essas práticas de alta disponibilidade, você pode garantir que sua infraestrutura de monitoramento do Amazon EKS permaneça confiável e resiliente e que você tenha visibilidade contínua de seus ambientes Kubernetes, mesmo durante vários cenários de falha. Testes e atualizações regulares dessas configurações de HA garantem que elas permaneçam eficazes à medida que o ambiente evolui.

Melhores práticas para monitoramento no Amazon EKS

Abordagem de implementação estratégica

Uma estratégia de monitoramento bem-sucedida do Amazon EKS começa com uma abordagem de implementação em fases e bem planejada.

- Comece identificando e monitorando métricas críticas que afetam diretamente suas operações comerciais e a confiabilidade dos aplicativos. Essa base deve incluir métricas essenciais de infraestrutura, indicadores-chave de desempenho de aplicativos e métricas críticas de segurança. Expanda gradualmente a cobertura de monitoramento com base nas necessidades operacionais e nas lições aprendidas e garanta que cada adição forneça um valor significativo.
- Implemente processos de implantação automatizados usando ferramentas de infraestrutura como código (IaC), como o Terraform, ou CloudFormation para garantir consistência e repetibilidade.
- Teste e valide sistemas de monitoramento para ajudar a manter a confiabilidade e a precisão.
- Refine os parâmetros de monitoramento continuamente de acordo com as necessidades comerciais em evolução.

Gerenciamento eficaz de dados

O gerenciamento adequado de dados é crucial para manter uma solução de monitoramento eficiente e econômica.

- Implemente políticas claras de retenção de dados que equilibrem as necessidades de análise histórica com os custos de armazenamento.
- Configure taxas de amostragem apropriadas para diferentes tipos de métricas: maior frequência para métricas críticas e menor frequência para métricas menos críticas.
- Use a agregação métrica para reduzir o volume de dados e, ao mesmo tempo, manter insights significativos, especialmente para análise de tendências de longo prazo.
- Implemente procedimentos sistemáticos de retenção e arquivamento de registros em sistemas de registro centralizados (como CloudWatch registros) para gerenciar os custos de armazenamento e manter o acesso a dados importantes acessível.

> **Note**
>
> A rotação de registros em nível de contêiner é feita automaticamente pelo kubelet na versão 1.21 ou posterior do Amazon EKS.

- Considere implementar uma hot-warm-cold arquitetura para armazenamento de registros para otimizar a velocidade de acesso e a eficiência de custos.

Configuração e gerenciamento de alertas

A configuração do alerta exige uma análise cuidadosa para manter a eficácia sem causar fadiga do alerta.

- Defina limites claros e acionáveis com base nos objetivos de nível de serviço (SLOs) e nos padrões históricos de desempenho.
- Implemente um sistema hierárquico de severidade de alertas que diferencie claramente entre problemas críticos que exigem atenção imediata e assuntos menos urgentes.
- Certifique-se de que os alertas forneçam contexto suficiente e informações práticas para facilitar a rápida resolução de problemas.
- Estabeleça procedimentos claros de escalonamento com propriedade e tempos de resposta definidos para diferentes severidades de alerta.
- Revise e refine as configurações de alerta regularmente para ajudar a manter sua relevância e eficácia.

Otimização de recursos

O monitoramento contínuo da utilização de recursos é essencial para manter as operações econômicas.

- Implemente um monitoramento abrangente de recursos em todos os componentes do cluster, incluindo nós, pods e volumes persistentes.
- Configure o escalonamento automático com base nos padrões reais de uso e nos requisitos de desempenho para garantir a utilização eficiente dos recursos e, ao mesmo tempo, manter o desempenho.

- Use tags de alocação de custos para monitorar o consumo de recursos por diferentes equipes, aplicativos ou ambientes.
- Analise regularmente as métricas de eficiência de recursos para identificar oportunidades de otimização e implementar melhorias.
- Considere implementar ferramentas de gerenciamento de custos para monitorar e otimizar os gastos com a nuvem.

Segurança

As considerações de segurança devem ser parte integrante de sua estratégia de monitoramento.

- Implemente princípios de acesso com privilégios mínimos para todos os componentes de monitoramento para garantir que usuários e serviços tenham somente as permissões de que precisam.
- Habilite um registro de auditoria abrangente para rastrear todos os acessos e alterações nos sistemas de monitoramento.
- Realize análises regulares de segurança das configurações de monitoramento e dos padrões de acesso para identificar possíveis vulnerabilidades.
- Implemente criptografia para dados de monitoramento confidenciais em trânsito e em repouso.
- Integre o monitoramento de segurança aos sistemas existentes de gerenciamento de eventos e informações de segurança (SIEM) para uma visibilidade abrangente da segurança.

Considerações de monitoramento avançado no Amazon EKS

Otimização do desempenho:

- Otimize os intervalos de coleta de métricas.
- Configure padrões de consulta eficientes.
- Implemente a pré-agregação métrica.
- Use soluções de armazenamento apropriadas.

Conformidade e governança:

- Mantenha trilhas de auditoria.

- Implemente o monitoramento de conformidade.
- Forneça relatórios regulares de conformidade.
- Procedimentos de monitoramento de documentos.

Recuperação de desastres:

- Faça backup das configurações de monitoramento regularmente.
- Procedimentos de recuperação de documentos.
- Teste os processos de recuperação.

Melhoria contínua:

- Monitore as sessões de revisão regularmente.
- Otimize os ciclos de desempenho.
- Atualize o monitoramento com base em incidentes.
- Incorpore o feedback do usuário.

Essas melhores práticas fornecem uma estrutura para implementar e manter soluções de monitoramento eficazes para ambientes Amazon EKS. Revise e atualize regularmente essas práticas para que elas permaneçam alinhadas às suas necessidades organizacionais e aos padrões do setor. O monitoramento não é uma configuração única — é um processo contínuo que requer atenção e refinamento regulares.

Rastreamento no Amazon EKS

O rastreamento é um componente essencial da observabilidade de aplicativos no Amazon EKS. O rastreamento fornece visibilidade detalhada dos fluxos de solicitações e das interações de serviços, coletando, processando e visualizando o caminho das solicitações à medida que elas percorrem vários microsserviços implantados em clusters EKS. Esse recurso ajuda você a entender o comportamento do sistema, identificar gargalos e solucionar problemas de forma eficaz em seu ambiente Amazon EKS. O rastreamento eficaz elimina a complexidade da depuração de sistemas distribuídos, fornecendo end-to-end visibilidade dos fluxos de solicitações. Isso possibilita rastrear transações entre os limites do serviço e identificar problemas de desempenho ou falhas nas cargas de trabalho do Amazon EKS.

A implementação geral do rastreamento no Amazon EKS permite que você entenda o comportamento do sistema, otimize o desempenho e mantenha a confiabilidade de seus aplicativos em contêineres. Em última análise, os recursos de rastreamento melhoram a visibilidade operacional e a capacidade de manutenção do sistema nos ambientes Amazon EKS.

AWS X-Ray desempenha um papel importante no rastreamento de dados sobre seu aplicativo. O rastreamento envolve o monitoramento de vários aspectos das interações do serviço, incluindo os seguintes:

- Os caminhos de solicitação e as dependências fornecem informações cruciais sobre o comportamento do seu sistema distribuído. Eles acompanham a jornada completa das solicitações à medida que elas percorrem diferentes microsserviços e componentes. O mapeamento das dependências do serviço ajuda você a entender os padrões de comunicação e identificar caminhos críticos na arquitetura do seu aplicativo. Para obter detalhes sobre a implementação, consulte Usando o mapa AWS X-Ray de rastreamento do serviço na documentação do X-Ray.
- Latências e gargalos do serviço são métricas essenciais para manter o desempenho ideal do sistema. Ao medir e analisar os tempos de resposta entre os serviços, você pode identificar problemas de desempenho de forma eficaz. Esses dados permitem identificar serviços ou operações específicos que estão causando atrasos na cadeia de solicitações e possibilitar esforços de otimização direcionados. Para saber mais sobre a análise de latência, consulte Interagindo com o console do Analytics na documentação do X-Ray.
- Os padrões de propagação de erros ajudam você a entender a confiabilidade do sistema e a tolerância a falhas. Ao entender como as falhas se espalham pelo sistema rastreando os caminhos de erro nos serviços, você pode arquitetar melhor seus aplicativos. Essa visibilidade ajuda a

identificar a causa raiz dos erros e seu impacto nos serviços dependentes, o que leva a sistemas mais resilientes. Para obter detalhes de implementação, consulte Traces na documentação do X-Ray.

- A utilização de recursos em todos os serviços fornece informações sobre a eficiência do sistema e a otimização de custos. Você pode monitorar padrões de uso de CPU, memória e rede que estão correlacionados com dados de rastreamento para entender as demandas de recursos. Esses dados ajudam você a analisar as tendências de consumo de recursos para otimizar o desempenho e o custo do serviço em seu cluster EKS. Para configuração de monitoramento, consulte Monitore o desempenho do seu cluster e visualize os registros na documentação do Amazon EKS.
- Os fluxos de transações do usuário final são essenciais para entender e melhorar a experiência do usuário. Ao rastrear as interações completas do usuário, dos serviços de front-end a back-end, você pode garantir o desempenho ideal do aplicativo. Você pode medir e otimizar os tempos de end-to-end resposta para jornadas críticas do usuário, o que afeta diretamente a satisfação do cliente. Para implementar o monitoramento do usuário final, use o AWS X-Ray SDK para sua linguagem de programação.
- As interações do gateway de API formam a linha de frente do desempenho e da segurança do seu aplicativo. Você pode monitorar os padrões e o desempenho das solicitações nos pontos de entrada da API para garantir a entrega ideal do serviço. Essa visibilidade ajuda você a monitorar os impactos de autenticação, autorização e limitação de taxa nos fluxos de solicitações, para manter os requisitos de segurança e desempenho. Saiba mais sobre o rastreamento de API na documentação do Amazon API Gateway with X-Ray.

O rastreamento eficaz no Amazon EKS vai além da coleta de extensões e traços. Ela exige uma estratégia bem estruturada que equilibre as necessidades de observabilidade com o desempenho do sistema. Essa estratégia deve se concentrar em:

- Implementando taxas de amostragem apropriadas: configure regras de amostragem com base em padrões de tráfego e prioridades de negócios para otimizar os custos e, ao mesmo tempo, manter a visibilidade das transações críticas. Para saber mais, consulte Configurando regras de amostragem na documentação do X-Ray.
- Definindo caminhos e serviços críticos a serem rastreados: identifique e priorize os serviços essenciais e as jornadas do usuário que exigem rastreamento detalhado para garantir o monitoramento ideal do desempenho. Para obter mais informações, consulte Enviar dados métricos e de rastreamento com o operador ADOT na documentação do Amazon EKS.

- Estabelecendo políticas adequadas de retenção de dados: configure regras de gerenciamento do ciclo de vida dos dados para equilibrar as necessidades de observabilidade com os custos de armazenamento e os requisitos de conformidade. Para ver as políticas CloudWatch de retenção, consulte Trabalho com grupos e fluxos de CloudWatch registros na documentação de registros.
- Configurando ferramentas eficazes de visualização e análise: implante e configure ferramentas de visualização, como o console AWS X-Ray Analytics ou o Amazon Managed Grafana, para analisar dados de rastreamento de forma eficaz. Para obter mais informações, consulte Interagindo com o console do Analytics na documentação do X-Ray.

Nesta seção:

Ferramentas de rastreamento para Amazon EKS

O Amazon EKS oferece suporte a várias opções AWS e de terceiros para implementar o rastreamento distribuído.

Serviços da AWS

- AWS X-Ray: Plataforma avançada de rastreamento distribuído

  O X-Ray é totalmente gerenciado AWS service (Serviço da AWS) que fornece recursos end-to-end de rastreamento. Ele instrumenta Serviços da AWS e fornece automaticamente mapas e análises de serviços detalhados para seus aplicativos que são executados no Amazon EKS. O X-Ray é integrado a outros Serviços da AWS, incluindo a Amazon CloudWatch, e oferece correlação automática de rastreamentos com AWS service (Serviço da AWS) chamadas.
- AWS Distro para OpenTelemetry: estrutura unificada de observabilidade

  O Distro for OpenTelemetry é uma distribuição segura, pronta para produção e com AWS suporte para aplicativos nativos da nuvem. OpenTelemetry Ele oferece recursos de instrumentação independentes do fornecedor, mantendo a AWS service (Serviço da AWS) integração nativa, o que o torna ideal para ambientes de nuvem híbrida. O Distro for OpenTelemetry suporta vários back-ends de observabilidade e fornece integração perfeita com serviços de monitoramento. AWS

Soluções de código aberto

- OpenTelemetry: Estrutura de observabilidade de código aberto

  OpenTelemetry fornece uma estrutura de observabilidade padronizada com bibliotecas de instrumentação abrangentes que oferecem suporte a várias linguagens de programação. Suas opções flexíveis de back-end e sua abordagem independente do fornecedor o tornam ideal para cargas de trabalho que exigem consistência em diferentes ambientes. O amplo ecossistema da estrutura garante ampla compatibilidade com várias soluções de monitoramento.

- Jaeger: plataforma de rastreamento distribuído de código aberto

  A Jaeger oferece recursos abrangentes de rastreamento com propagação de contexto distribuído em tempo real. Ele fornece análise da causa raiz e otimização do desempenho por meio da visualização detalhada da dependência do serviço. A arquitetura da Jaeger foi projetada para alta escalabilidade e oferece suporte a vários back-ends de armazenamento, o que a torna adequada para implantações em grande escala do Amazon EKS. Veja a configuração do Jaeger para EKS

- Grafana Tempo: Rastreamento distribuído

  O Tempo é uma solução da Grafana Labs que fornece armazenamento de rastreamento em alta escala e integração perfeita com as métricas do Prometheus. Seu modelo econômico de retenção de traços e a integração nativa com o Grafana o tornam adequado para organizações que já usam o Grafana para visualização. A arquitetura do Tempo foi projetada especificamente para ambientes nativos da nuvem, como o Amazon EKS.

Melhores práticas para rastreamento no Amazon EKS

Esta seção fornece uma lista abrangente das melhores práticas e técnicas para criar um sistema de rastreamento eficaz que aprimora a observabilidade e a solução de problemas de seus aplicativos baseados em Kubernetes no Amazon EKS.

- Amostragem estratégica: configure diferentes taxas de amostragem com base nos padrões de tráfego do seu aplicativo e na importância dos serviços que você está usando. Implemente taxas de amostragem mais altas para caminhos críticos e, ao mesmo tempo, reduza a amostragem para rotas menos críticas e de alto volume para otimizar custos. Para obter orientação, consulte Configuração de regras de amostragem na AWS X-Ray documentação.

- Configuração da instrumentação: use ferramentas automáticas de instrumentação, como o X-Ray SDK ou o AWS Distro para OpenTelemetry coletores, a fim de minimizar o esforço de instrumentação manual. Mantenha convenções de nomenclatura consistentes e propagação de contexto entre os serviços para uma melhor correlação de rastreamento. Para obter mais informações, consulte a documentação do Distro for OpenTelemetry Collector.
- Gerenciamento de dados: implemente períodos de retenção e estratégias de compactação adequados para equilibrar os custos de armazenamento com suas necessidades de observabilidade. Estabeleça controles claros de privacidade de dados e procedimentos de backup para proteger dados confidenciais de rastreamento. Para obter mais informações, consulte Alterar a retenção de dados do registro em CloudWatch Registros na documentação de CloudWatch Registros.
- Otimização do desempenho: monitore e otimize a sobrecarga de rastreamento para minimizar o impacto no desempenho do aplicativo. Use buffer eficiente e processamento assíncrono para reduzir o impacto da latência. Para obter mais informações, consulte Configurando o AWS X-Ray daemon na documentação do X-Ray.
- Controles de segurança: implemente controles de acesso e medidas de proteção de dados adequados usando funções e políticas do IAM. Auditorias regulares de segurança e análises de conformidade ajudam a garantir que os dados de rastreamento permaneçam seguros. Para obter mais informações, consulte Segurança AWS X-Ray na documentação do X-Ray.
- Monitoramento e alertas: configure um monitoramento abrangente da integridade da coleta de rastreamento e configure alertas para problemas de coleta. Acompanhe as taxas de amostragem e as métricas de desempenho do sistema para garantir uma operação ideal. Para obter mais informações, consulte Container Insights na CloudWatch documentação.
- Alta disponibilidade: implante coletores redundantes em todas as zonas de disponibilidade e configure mecanismos de failover adequados. Testes regulares da configuração de alta disponibilidade garantem uma coleta confiável de traços. Para obter mais informações, consulte Usando o AWS Distro OpenTelemetry como coletor na documentação do Amazon Managed Service for Prometheus.

Seguindo essas melhores práticas, você pode criar um sistema de rastreamento robusto, eficiente e eficaz para seu ambiente Amazon EKS. Isso ajudará a garantir observabilidade abrangente, solução de problemas eficiente e desempenho ideal de seus aplicativos baseados em Kubernetes.

Alertas no Amazon EKS

O alerta é um componente essencial do gerenciamento e manutenção de aplicativos executados no Amazon EKS. Ele serve como um sistema de alerta precoce que notifica operadores e desenvolvedores sobre possíveis problemas, anomalias ou degradações de desempenho antes que eles se transformem em problemas graves que possam afetar a disponibilidade do serviço ou a experiência do usuário. O alerta envolve o monitoramento de vários aspectos do cluster Kubernetes, incluindo:

- Saúde da infraestrutura
- Desempenho do aplicativo
- Métricas de contêiner
- Métricas de negócios personalizadas

Os alertas eficazes no Amazon EKS vão além da simples configuração de notificações. Isso requer uma well-thought-out estratégia que equilibre a necessidade de informações oportunas com o potencial de fadiga de alerta. Essa estratégia deve:

- Defina limites e condições significativos.
- Priorize os alertas com base na gravidade e no impacto.
- Implemente procedimentos adequados de roteamento e escalonamento.
- Integre-se às ferramentas de gerenciamento e comunicação de incidentes.

Nesta seção:

Ferramentas de alerta para o Amazon EKS

O Amazon EKS oferece suporte a várias opções AWS e a de terceiros para implementar alertas. Ao escolher uma ferramenta para alertar o Amazon EKS, considere fatores como recursos de integração, escalabilidade, facilidade de uso, custo e recursos específicos que se alinham aos seus requisitos de monitoramento e alerta. Muitas organizações usam uma combinação dessas

ferramentas para criar uma solução abrangente de monitoramento e alerta para seus ambientes Amazon EKS.

- Amazon CloudWatch: AWS service (Serviço da AWS) para monitoramento e observabilidade

  CloudWatch fornece métricas, registros e alarmes para clusters EKS e se integra bem com outros. Serviços da AWS
- Prometheus: ferramenta de monitoramento e alerta de código aberto para Kubernetes

  O Prometheus fornece uma linguagem de consulta poderosa (PromQL) para definir condições de alerta.
- Alertmanager: companheiro do Prometheus para lidar com alertas

  O Alertmanager fornece desduplicação, agrupamento e roteamento de alertas. Ele suporta vários canais de notificação, incluindo e-mail, Slack e. PagerDuty
- Grafana: plataforma de código aberto para monitoramento e observabilidade

  O Grafana fornece recursos de visualização e alerta. Ele pode se integrar a várias fontes de dados, incluindo Prometheus e. CloudWatch
- Elastic Stack (ELK Stack): combinação de Elasticsearch, Logstash e Kibana

  Essa ferramenta é útil para agregação, análise e alertas de registros. Ele pode ser estendido com os recursos de observabilidade da Elastic.
- Soluções de terceiros

  Há muitas ferramentas disponíveis no mercado, incluindo Datadog, New Relic, Sysdig, Dynatrace, Zabbix, Nagios, Splunk, IBM Instana e. AppDynamics

Melhores práticas para alertas no Amazon EKS

Esta seção descreve as melhores práticas para criar um sistema de alerta robusto que aprimora a confiabilidade e o desempenho de seus aplicativos baseados em Kubernetes no Amazon EKS.

Defina limites claros de alerta:

- Defina limites significativos com base em dados históricos e requisitos de negócios.
- Use limites dinâmicos quando apropriado para contabilizar cargas de trabalho variáveis.

Implemente a priorização de alertas:

- Categorize os alertas por gravidade (por exemplo, crítico, alto, médio e baixo).
- Alinhe as prioridades de alerta com o impacto nos negócios.

Evite a fadiga de alerta:

- Reduza o ruído eliminando alertas redundantes ou de baixo valor.
- Correlacione alertas a problemas relacionados ao grupo.

Use alertas em vários estágios:

- Implemente limites de aviso antes que os níveis críticos sejam atingidos.
- Use canais de notificação diferentes para diferentes severidades de alerta.

Implemente o roteamento de alertas adequado:

- Certifique-se de que os alertas sejam enviados para as equipes ou indivíduos certos.
- Use horários e rotações de plantão para a cobertura do dia todo, todos os dias.

Aproveite as métricas nativas do Kubernetes:

- Monitore os principais componentes do Kubernetes (nós, pods, serviços).
- Use kube-state-metrics (KSM) para obter métricas adicionais de objetos do Kubernetes.

Monitore a infraestrutura e os aplicativos:

- Configure alertas para a integridade do cluster, o status do nó e a utilização de recursos.
- Implemente alertas específicos do aplicativo, como taxas de erro e latência.

Use o Prometheus e o Alertmanager:

- Use o Prometheus para coleta de métricas e o PromQL para definir condições de alerta.
- Use o Alertmanager para roteamento e desduplicação de alertas.

Integre com a Amazon CloudWatch:

- Use o CloudWatchContainer Insights para métricas específicas do Amazon EKS.
- Configure CloudWatchalarmes para métricas críticas AWS de recursos.

Implemente alertas contextuais:

- Inclua informações relevantes nas mensagens de alerta, como nome do cluster, namespace e detalhes do pod.
- Forneça links para painéis ou runbooks relevantes em alertas.

Use a detecção de anomalias:

- Implemente a detecção de anomalias baseada em aprendizado de máquina para padrões complexos.
- Use serviços como detecção de CloudWatch anomalias ou ferramentas de terceiros.

Implemente a supressão e o silenciamento de alertas:

- Permita a supressão temporária de problemas conhecidos.
- Implemente janelas de manutenção para reduzir o ruído durante os períodos de inatividade planejados.

Monitore o desempenho do alerta:

- Monitore métricas como frequência de alertas, tempo de resolução e taxas de falsos positivos.
- Revise e refine regularmente as regras de alerta com base nessas métricas.

Implemente procedimentos de escalonamento:

- Defina caminhos claros de escalonamento para alertas não resolvidos.
- Use ferramentas como PagerDuty o Opsgenie para escalonamentos automatizados.

Teste os sistemas de alerta regularmente:

- Realize testes periódicos do seu pipeline de alertas.
- Inclua testes de alerta em exercícios de recuperação de desastres.

Use modelos para consistência de alertas:

- Crie modelos de alerta padronizados para cenários comuns.
- Garanta formatação e informações consistentes em todos os alertas.

Implemente a limitação de taxa:

- Evite tempestades de alertas implementando a limitação de taxa em alertas acionados com frequência.

Use métricas personalizadas:

- Implemente métricas personalizadas para monitoramento específico do aplicativo.
- Use a API de métricas personalizadas do Kubernetes para escalonamento automático com base nessas métricas.

Implemente a integração de registro:

- Correlacione alertas com registros relevantes para agilizar a solução de problemas.
- Use ferramentas como o Grafana Loki ou o ELK Stack em conjunto com seu sistema de alerta.

Considere os alertas de custo:

- Configure alertas para picos inesperados no uso ou nos custos dos recursos.
- Use AWS Budgetsnossas ferramentas de gerenciamento de custos de terceiros.

Use o rastreamento distribuído:

- Integre ferramentas de rastreamento distribuídas, como Jaeger ou. AWS X-Ray
- Configure alertas para padrões de rastreamento ou latências anormais.

Runbooks de alertas de documentos:

- Crie runbooks claros e acionáveis para cada tipo de alerta.
- Inclua etapas de solução de problemas e procedimentos de escalonamento nos runbooks.

Seguindo essas melhores práticas, você pode criar um sistema de alerta robusto, eficiente e eficaz para seu ambiente Amazon EKS. Isso ajudará a garantir alta disponibilidade, resolução rápida de problemas e desempenho ideal de seus aplicativos baseados em Kubernetes.

Próximas etapas

Este guia forneceu uma estrutura abrangente para implementar uma observabilidade robusta em ambientes Amazon EKS, com foco na coleta de métricas, infraestrutura de registro, rastreamento distribuído e otimização de custos. Ao entender e aplicar esses componentes principais, você pode criar um ambiente de contêiner altamente observável, sustentável e econômico que forneça insights profundos sobre o comportamento do aplicativo e da infraestrutura. A integração, Serviços da AWS como o Amazon CloudWatch Container Insights AWS X-Ray, combinada com soluções de código aberto, como o Prometheus OpenTelemetry e, cria uma base poderosa para monitorar e solucionar problemas de aplicativos em contêineres.

O sucesso da implementação depende de uma abordagem em fases, começando com a coleta de métricas principais e expandindo gradualmente para recursos abrangentes de registro e rastreamento distribuído. Recomendamos que você comece avaliando seus recursos atuais de monitoramento, identificando lacunas e selecionando combinações de ferramentas apropriadas que se alinhem aos requisitos operacionais e à experiência da equipe. Essa abordagem metódica garante que cada componente da pilha de observabilidade seja implementado e integrado adequadamente, enquanto as equipes desenvolvem as habilidades e os processos necessários para usar essas ferramentas com eficácia.

A sustentabilidade de longo prazo da observabilidade do Amazon EKS depende da otimização regular de custos, recursos e processos. Você deve revisar e ajustar continuamente sua infraestrutura de observabilidade, incluindo políticas de retenção de dados, taxas de amostragem e alocação de recursos, para manter o equilíbrio certo entre monitoramento abrangente e eficiência operacional. Essa abordagem iterativa de melhoria, combinada com o treinamento contínuo da equipe e as atualizações da documentação, permite que sua organização mantenha uma observabilidade eficaz e, ao mesmo tempo, apoie o crescimento dos negócios e se adapte às arquiteturas de aplicativos em evolução.

Recursos

AWS documentação

- Guia de melhores práticas do Amazon EKS
- Amazon CloudWatch Container Insights
- Amazon Managed Service for Prometheus
- Grafana gerenciado pela Amazon
- AWS Distro para e OpenTelemetry AWS X-Ray
- OpenSearch Serviço Amazon

AWS postagens no blog

- O Amazon EKS aprimora a observabilidade do plano de controle do Kubernetes
- Automatizando a coleta de métricas no Amazon EKS com o Amazon Managed Service para raspadores gerenciados Prometheus
- Automatize o monitoramento do seu cluster Amazon EKS usando o CloudWatch Container Insights
- Aprimorando a observabilidade com uma solução de monitoramento gerenciado para o Amazon EKS

Outros recursos

- Documentação do OpenTelemetry
- Documentação do Prometheus
- Documentação do Fluent Bit
- Documentação de monitoramento, registro e depuração no Kubernetes

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
Atualizações	Atualizamos o capítulo Logging in Amazon EKS.	17 de março de 2026
Publicação inicial	—	10 de abril 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:

- Refactor/re-architect — mova um aplicativo e modifique sua arquitetura aproveitando ao máximo os recursos nativos da nuvem para melhorar a agilidade, o desempenho e a escalabilidade. Isso normalmente envolve a portabilidade do sistema operacional e do banco de dados. Exemplo: migre seu banco de dados Oracle local para a Amazon PostgreSQL-Compatible Aurora Edition.
- 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: Migre 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

A2A () Agent-to-Agent

Um protocolo com estado para colaboração entre agentes, apoiando a delegação de tarefas e a transferência de estados.

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.

Agente

Um sistema de IA que pode raciocinar, planejar e realizar ações de forma autônoma usando ferramentas para atingir metas.

Agente Ops

Práticas operacionais para criar, testar, implantar e executar agentes de IA na produção em grande escala.

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 as AIOps são usadas na estratégia de migração para a AWS , 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 implantação

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 Implementar procedimentos de quebra de vidros na AWS Well-Architected orientação.

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.

Desenvolvedor cidadão

Um usuário corporativo que cria aplicativos de IA usando plataformas sem code/low código sem habilidades técnicas especializadas.

criptografia no lado do cliente

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

Centro de Excelência da 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 postagens do CCoE 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: realizar investimentos fundamentais para escalar sua adoção da nuvem (por exemplo, criar uma zona de pouso, definir um CCoE, estabelecer um modelo de operações)
- Migração: migrar aplicações individuais
- Re-invention — Otimizando produtos e serviços e inovando na nuvem

Esses estágios foram definidos por Stephen Orban na postagem do blog The Journey Toward Cloud-First & the Stages of Adoption no blog Nuvem AWS Enterprise Strategy. 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 CI/CD pipeline 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

Uma coleção 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 na AWS Well-Architected Estrutura. 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.

defesa completa

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 abordagem de defesa aprofundada 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 na AWS Well-Architected estrutura.

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 Domain-Driven Design: Tackling Complexity in the Heart of Software (Boston: Addison-Wesley Professional, 2003). Para obter informações sobre como você pode usar o design orientado por domínio com o padrão strangler fig, consulte Modernizando os serviços web legados da Microsoft ASP.NET (ASMX) de forma incremental 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. Big-endian os sistemas armazenam primeiro o byte mais significativo. Little-endian os sistemas armazenam primeiro o byte menos significativo.

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 contextual, em que os modelos aprendem com exemplos (fotos) incorporados aos prompts. Few-shot a solicitação pode ser eficaz 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 treina em grandes conjuntos de dados generalizados e não rotulados. Os 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?.

Gateway FM

Um intermediário centralizado que controla e normaliza o acesso aos modelos de fundação. Também conhecido como gateway LLM.

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 gerenciar recursos, políticas e conformidade em todas as unidades organizacionais (UOs). 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 .

grades de proteção (IA)

Mecanismos de segurança que filtram, validam e restringem as entradas e saídas dos agentes para ajudar a garantir um comportamento de IA responsável e seguro.

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.

humano no circuito (HiTL)

Um padrão de fluxo de trabalho em que a execução do agente é pausada para análise e aprovação humana em pontos críticos de decisão.

migração homogênea de bancos de dados

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

dados quentes

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

hotfix

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

período de hipercuidados

Imediatamente após a substituição, o período em que uma equipe de migração gerencia e monitora as aplicações migradas na nuvem para resolver quaisquer problemas. Normalmente, a duração desse período é de 1 a 4 dias. No final do período de hipercuidados, a equipe de migração normalmente transfere a responsabilidade pelas aplicações para a equipe de operações de nuvem.

eu

IaC

Veja 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 as melhores práticas de implantação usando infraestrutura imutável na AWS Well-Architected Estrutura.

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 segurança da AWS recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação e a Internet em geral.

migração incremental

Uma estratégia de substituição na qual você migra a aplicação em pequenas partes, em vez de realizar uma única substituição completa. Por exemplo, é possível mover inicialmente apenas alguns microsserviços ou usuários para o novo sistema. Depois de verificar se tudo está funcionando corretamente, mova os microsserviços ou usuários adicionais de forma incremental até poder descomissionar seu sistema herdado. Essa estratégia reduz os riscos associados a migrações de grande porte.

Indústria 4.0

Um termo que foi introduzido por Klaus Schwab em 2016 para se referir à modernização dos processos de fabricação por meio de avanços na conectividade, dados em tempo real, automação, análise e. AI/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 das Coisas Industrial (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 Construir uma estratégia de transformação digital para a Internet das Coisas Industrial (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 segurança da AWS recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação 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 é grande modelo de linguagem (LLM)?.

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.

MCP

Consulte Protocolo de contexto do modelo.

Protocolo de contexto para modelos (MCP)

Um protocolo sem estado para comunicação entre agentes e ferramentas.

Servidor MCP

Um serviço que expõe uma ou mais ferramentas por meio do Model Context Protocol.

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 Criação de mecanismos na AWS Well-Architected estrutura.

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, máquina a máquina (M2M), baseado no padrão, para dispositivos de IoT com recursos publish/subscribelimitados.

microsserviço

Um serviço pequeno e independente que se comunica por meio de APIs bem definidas e normalmente pertence a equipes pequenas e autônomas. 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 APIs leves. 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

Cross-functional equipes que simplificam a migração de cargas de trabalho 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, a AWS Well-Architected Estrutura 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.

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

Um protocolo de comunicação máquina a máquina (M2M) para automação industrial. 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) na AWS Well-Architected Estrutura.

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

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 segurança da AWS recomenda configurar sua conta de rede com VPCs de entrada, saída e inspeção para proteger a interface bidirecional entre a aplicação 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 polítca 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 armazena informações sobre como você quer que o Amazon Route 53 responda a consultas ao DNS para um domínio e seus subdomínios dentro de uma 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 no AWS Organizations. As SCPs definem barreiras de proteção ou estabelecem limites para as ações que um administrador pode delegar a usuários ou perfis. É possível usar SCPs como listas de permissão ou 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.

Inteligência artificial sombria

Aplicativos de IA não autorizados criados ou usados fora dos canais controlados dentro de uma organização.

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.

modelo dividir e semear

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 Modernizando os serviços web legados da Microsoft ASP.NET (ASMX) de forma incremental 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

Key-value pares que atuam como metadados para organizar seus AWS recursos. 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.

ferramenta

Uma função ou API que um agente pode invocar para realizar operações em sistemas externos.

gateway de trânsito

Um hub de trânsito de rede que pode ser usado para interconectar as VPCs e as redes on-premises. 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 de VPC

Uma conexão entre duas VPCs que permite rotear 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.