

Framework Well-Architected da AWS

Pilar Sustentabilidade



Pilar Sustentabilidade: Framework Well-Architected da AWS

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens comerciais da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestige a Amazon. Todas as outras marcas comerciais que não são propriedade da Amazon pertencem aos respectivos proprietários, os quais podem ou não ser afiliados, estar conectados ou ser patrocinados pela Amazon.

Table of Contents

Resumo e introdução	i
Introdução	1
Sustentabilidade na nuvem	2
O modelo de responsabilidade compartilhada	3
Sustentabilidade da nuvem	3
Sustentabilidade na nuvem	4
Sustentabilidade via nuvem	4
Princípios de design para sustentabilidade na nuvem	5
Processo de melhoria	7
Exemplo de cenário	8
Identificar metas para melhoria	8
Recursos	8
Avaliar melhorias específicas	9
Métricas proxy	9
Métricas empresariais	10
Indicadores-chave de performance	10
Estimar as melhorias	10
Avaliar as melhorias	11
Priorizar e planejar melhorias	12
Testar e validar as melhorias	13
Implantar alterações na produção	14
Avaliar os resultados e replicar os sucessos	15
Sustentabilidade como um requisito não funcional	18
Práticas recomendadas para a sustentabilidade na nuvem	20
Seleção da região	20
SUS01-BP01 Escolher a região com base nos requisitos empresariais e nas metas de sustentabilidade	20
Alinhamento com a demanda	22
SUS02-BP01 Escalar a infraestrutura da workload dinamicamente	23
SUS02-BP02 Alinhar os SLAs às metas de sustentabilidade	26
SUS02-BP03 Interromper a criação e a manutenção de ativos não utilizados	29
SUS02-BP04 Otimizar o posicionamento geográfico das workloads com base nos respectivos requisitos de rede	30
SUS02-BP05 Otimizar os recursos dos membros da equipe para as atividades realizadas	34

SUS02-BP06 Implementar armazenamento em buffer ou controle de utilização para nivelar a curva da demanda	35
Software e arquitetura	38
SUS03-BP01 Otimizar o software e a arquitetura para trabalhos assíncronos e agendados	39
SUS03-BP02 Remover ou refatorar componentes da workload com pouco ou nenhum uso	42
SUS03-BP03 Otimizar áreas de código que consomem mais tempo ou recursos	44
SUS03-BP04 Otimizar o impacto sobre dispositivos e equipamentos	46
SUS03-BP05 Usar padrões e arquiteturas de software que atendam melhor aos padrões de armazenamento e acesso a dados	48
Gerenciamento de dados	51
SUS04-BP01 Implementar uma política de classificação de dados	52
SUS04-BP02 Usar tecnologias compatíveis com seus padrões de acesso e de armazenamento de dados	53
SUS04-BP03 Usar políticas para gerenciar o ciclo de vida de seus conjuntos de dados	59
SUS04-BP04 Usar elasticidade e automação para expandir o armazenamento em bloco ou o sistema de arquivos	62
SUS04-BP05 Remover dados desnecessários ou redundantes	63
SUS04-BP06 Usar armazenamento ou sistemas de arquivos compartilhados para acessar dados comuns	66
SUS04-BP07 Minimizar a movimentação de dados entre redes	68
SUS04-BP08 Fazer backup de dados somente quando for difícil recriá-los	70
Hardware e serviços	72
SUS05-BP01 Usar a quantidade mínima de hardware para atender às suas necessidades ...	73
SUS05-BP02 Usar tipos de instância com o mínimo de impacto	75
SUS05-BP03 Usar serviços gerenciados	78
SUS05-BP04 Otimizar o uso de aceleradores de computação baseados em hardware	81
Processo e cultura	83
SUS06-BP01 Comunicar e distribuir as metas de sustentabilidade em cascata	83
SUS06-BP02 Adotar métodos que podem apresentar melhorias na sustentabilidade rapidamente	86
SUS06-BP03 Manter a workload atualizada	88
SUS06-BP04 Aumentar a utilização de ambientes de compilação	90
SUS06-BP05 Usar parques de dispositivos gerenciados para testes	92
Conclusão	95

Colaboradores	96
Outras fontes de leitura	97
Revisões do documento	98
Avisos	100
Glossário da AWS	101

Pilar Sustentabilidade: AWS Well-Architected Framework

Data de publicação: 6 de novembro de 2024 ([Revisões do documento](#))

Este whitepaper destaca o pilar Sustentabilidade do Amazon Web Services (AWS) Well-Architected Framework. Ele fornece princípios de design, orientações operacionais, práticas recomendadas, concessões potenciais e planos de melhoria que podem ser usados para atender às metas de sustentabilidade das suas workloads da AWS.

Introdução

O AWS Well-Architected Framework ajuda a compreender os prós e os contras das decisões tomadas ao criar workloads na AWS. O uso do Framework ajuda você a aprender as práticas recomendadas de arquitetura para projetar e operar workloads seguras, confiáveis, eficientes, econômicas e sustentáveis na Nuvem AWS. Ele fornece uma maneira de você avaliar consistentemente suas arquiteturas em relação às práticas recomendadas e identificar áreas de aprimoramento. As workloads bem-arquitetadas aumentam muito sua habilidade de oferecer suporte aos resultados da empresa.

O framework é baseado em seis pilares:

- Excelência operacional
- Segurança
- Confiabilidade
- Eficiência de desempenho
- Otimização de custo
- Sustentabilidade

Este documento se refere ao pilar Sustentabilidade e ao escopo correspondente, com foco na sustentabilidade ambiental. Ele é destinado a pessoas que ocupam cargos de tecnologia, como diretores de tecnologia (CTOs), arquitetos, desenvolvedores e membros da equipe de operações.

Depois de ler este documento, você compreenderá as recomendações e as estratégias atuais da AWS para projetar arquiteturas de nuvem com foco em sustentabilidade. Ao adotar as práticas descritas neste documento, é possível criar arquiteturas que maximizam a eficiência e reduzem o desperdício.

Sustentabilidade na nuvem

A disciplina da sustentabilidade aborda o impacto ambiental, econômico e social de longo prazo de suas atividades empresariais. A [Comissão Mundial sobre Meio Ambiente e Desenvolvimento das Nações Unidas](#) define desenvolvimento sustentável como o "desenvolvimento que atende às necessidades do presente sem comprometer a possibilidade de as gerações futuras atenderem às suas próprias necessidades". Sua empresa ou organização pode ter impactos ambientais negativos, como emissões de carbono diretas ou indiretas, lixo não reciclável e danos a recursos compartilhados, como água potável.

Ao criar workloads em nuvem, a prática da sustentabilidade abrange ações como: compreender o impacto dos serviços usados, quantificá-lo ao longo de todo o ciclo de vida da workload e aplicar princípios e práticas recomendadas de design para reduzi-lo. Este documento foca os impactos ambientais, especialmente a eficiência e o consumo de energia, que são fatores importantes para fundamentar ações diretas dos arquitetos visando reduzir o uso de recursos.

Ao se concentrar nos impactos ambientais, você deve entender como esses impactos são normalmente contabilizados e compreender os impactos subsequentes na contabilidade de emissões da própria organização. O [Protocolo de Gases de Efeito Estufa](#) organiza as emissões de carbono nos seguintes escopos, junto com exemplos de emissões relevantes dentro de cada escopo para um provedor de nuvem como a AWS:

- Escopo 1: todas as emissões diretas das atividades de uma organização ou sob seu controle. Por exemplo, queima de combustível por geradores de backup do data center.
- Escopo 2: emissões indiretas da eletricidade comprada e usada para alimentar data centers e outras instalações. Por exemplo, emissões da geração comercial de energia.
- Escopo 3: todas as outras emissões indiretas das atividades de uma organização provenientes de fontes que ela não controla. Exemplos da AWS incluem emissões relacionadas à construção de data centers e à fabricação e transporte de hardware de TI implantado em data centers.

Do ponto de vista do cliente da AWS, as emissões de suas workloads em execução na AWS são contabilizadas como emissões indiretas e fazem parte de suas emissões do Escopo 3. Cada workload implantada gera uma fração do total de emissões da AWS de cada um dos escopos anteriores. A quantidade real varia de acordo com a workload e depende de vários fatores, incluindo os serviços da AWS usados, a energia consumida por esses serviços, a intensidade de carbono das

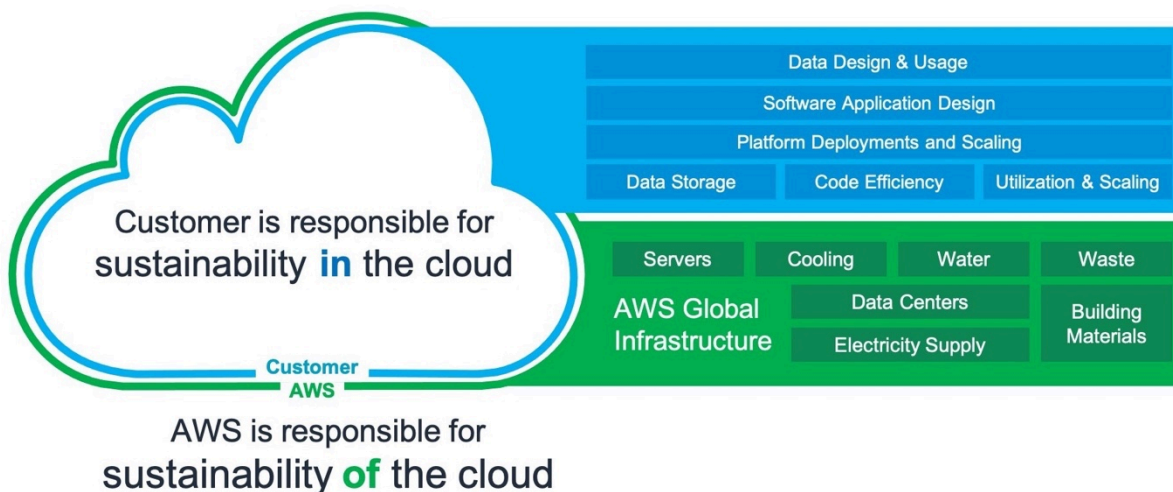
redes elétricas que atendem aos data centers da AWS em que operam e a aquisição de energia renovável pela AWS.

Este documento primeiro descreve um modelo de responsabilidade compartilhada para a sustentabilidade ambiental e, em seguida, apresenta práticas recomendadas de arquitetura para que você possa minimizar o impacto de suas workloads reduzindo o total de recursos necessários para que elas sejam executadas em data centers da AWS.

O modelo de responsabilidade compartilhada

A sustentabilidade ambiental é uma responsabilidade compartilhada entre os clientes e a AWS.

- A AWS é responsável por otimizar a sustentabilidade da nuvem — fornecendo infraestrutura eficiente e compartilhada, administração de recursos hídricos e fornecimento de energia renovável.
- Os clientes são responsáveis pela sustentabilidade na nuvem, otimizando as workloads e a utilização de recursos e minimizando o total de recursos necessários para serem implantados em suas workloads.



Modelo de responsabilidade compartilhada

Sustentabilidade da nuvem

Os provedores de nuvem têm uma pegada de carbono menor e são mais eficientes em termos de energia do que as alternativas on-premises típicas porque investem em tecnologia eficiente de energia e resfriamento, operam populações de servidores com eficiência energética e alcançam altas

taxas de utilização do servidor. As workloads na nuvem são responsáveis por uma fração do impacto dos recursos compartilhados, como redes, energia, refrigeração e instalações físicas. Você pode migrar suas workloads na nuvem para tecnologias mais eficientes à medida que são disponibilizadas e usar serviços baseados na nuvem para transformar suas workloads a fim de obter uma melhor sustentabilidade.

Recursos

- [A oportunidade de redução de carbono proporcionada pela migração para a Amazon Web Services](#)
- [A AWS viabiliza soluções de sustentabilidade](#)

Sustentabilidade na nuvem

A sustentabilidade na nuvem é um esforço contínuo focado principalmente na redução e eficiência de energia em todos os componentes de uma workload, obtendo o máximo benefício dos recursos provisionados e minimizando o total de recursos necessários. Esse esforço pode abranger desde a seleção inicial de uma linguagem de programação eficiente, a adoção de algoritmos modernos, o uso de técnicas eficientes de armazenamento de dados, a implantação de uma infraestrutura computacional de tamanho correto e eficiente e a minimização dos requisitos de hardware de alta potência para o usuário final.

Sustentabilidade via nuvem

Além de minimizar o impacto das workloads que você implantou, é possível usar a Nuvem AWS para executar workloads projetadas para acomodar seus desafios mais amplos de sustentabilidade. Exemplos desses desafios incluem reduzir as emissões de carbono, reduzir o consumo de energia, reciclar água ou reduzir o desperdício em outras áreas de sua empresa ou organização.

Sustentabilidade via nuvem é quando você usa tecnologia da AWS para resolver um desafio mais amplo de sustentabilidade. Por exemplo, você pode usar um serviço de machine learning para detectar comportamentos anormais em equipamentos industriais. Usando esses dados de detecção, é possível realizar a manutenção preventiva para reduzir o risco de incidentes ambientais causados por falhas inesperadas do equipamento e garantir que o maquinário continue operando com eficiência máxima.

Princípios de design para sustentabilidade na nuvem

Aplique esses princípios de design ao arquitetar suas workloads na nuvem para maximizar a sustentabilidade e minimizar o impacto.

- **Entenda seu impacto:** meça o impacto da sua workload na nuvem e modele o impacto futuro que ela poderá causar. Inclua todas as fontes de impacto, inclusive aquelas resultantes do uso dos seus produtos pelo cliente e da desativação e descontinuação dos mesmos. Compare o resultado produtivo com o impacto total de suas workloads em nuvem analisando os recursos e as emissões exigidas por unidade de trabalho. Use esses dados para estabelecer indicadores-chave de performance (KPIs), avaliar maneiras de melhorar a produtividade enquanto reduz o impacto e estimar o impacto das mudanças propostas ao longo do tempo.
- **Defina metas de sustentabilidade:** para cada workload na nuvem, estabeleça metas de sustentabilidade de longo prazo, por exemplo, reduzir os recursos de computação e armazenamento exigidos por transação. Modele o retorno sobre o investimento para as melhorias de sustentabilidade das workloads e ofereça aos proprietários os recursos de que eles precisam para investir em metas de sustentabilidade. Planeje-se para o crescimento e projete suas workloads de forma que seu desenvolvimento resulte em uma intensidade de impacto menor com relação a uma unidade apropriada, como por usuário ou por transação. As metas ajudam você a respaldar os objetivos de sustentabilidade mais amplos de sua empresa ou organização, identificar regressões e priorizar áreas para possível melhoria.
- **Maximize a utilização:** dimensione as workloads corretamente e implemente um design eficiente que garanta uma alta utilização e maximize a eficiência de energia do hardware subjacente. Dois hosts com 30% de utilização são menos eficientes do que um host com 60% devido ao consumo de energia de referência por host. Ao mesmo tempo, elimine ou minimize recursos, processamento e armazenamento ociosos para reduzir a energia total necessária para suprir a workload.
- **Antecipe e adote ofertas de hardware e software novas e mais eficientes:** apoie as melhorias preventivas que seus parceiros e fornecedores disponibilizam para ajudar você a reduzir o impacto das workloads na nuvem. Monitore e avalie continuamente ofertas de software e hardware novos e mais eficientes. Projete visando a flexibilidade para permitir a adoção rápida de novas tecnologias eficientes.
- **Use serviços gerenciados:** compartilhe serviços com uma ampla base de clientes ajuda a maximizar a utilização de recursos, o que reduz a quantidade de infraestrutura necessária para comportar as workloads na nuvem. Por exemplo, os clientes podem compartilhar o impacto de componentes comuns de um data center, como a energia e as redes, migrando workloads para a Nuvem AWS e adotando serviços gerenciados, como o AWS Fargate para contêineres

sem servidor, onde a AWS trabalha em escala e é responsável por sua operação eficiente. Use serviços gerenciados que possam ajudar a minimizar seu impacto, como a migração automática de dados acessados com pouca frequência para o armazenamento com pouco acesso com as configurações do ciclo de vida do Amazon S3 ou o Amazon EC2 Auto Scaling para ajustar a capacidade de acordo com a demanda.

- Reduza o impacto downstream das suas workloads na nuvem: reduza a quantidade de energia ou recursos necessários para usar seus serviços. Reduza ou elimine a necessidade de os clientes fazerem upgrade de dispositivos para usar seus serviços. Teste o uso de parques de dispositivos para saber qual é o impacto esperado e teste com os clientes para entender o impacto atual do uso de seus serviços.

Processo de melhoria

O processo de melhoria da arquitetura inclui a compreensão do que você tem e o que pode ser feito para melhorar, definir o que precisa ser melhorado, testar as melhorias, adotar melhorias bem-sucedidas, quantificar seu êxito e compartilhar o que você aprendeu para que possa ser replicado em outras partes e, então, repetir o ciclo.

As metas de melhoria podem ser:

- Eliminar desperdício, baixa utilização e recursos ociosos ou não utilizados
- Maximizar o valor dos recursos que você consome

Note

Use todos os recursos que você provisiona e conclua o mesmo trabalho com o mínimo de recursos possível.

Nos estágios iniciais da otimização, concentre-se primeiro em áreas com resíduos ou baixa utilização e, depois, passe para otimizações mais direcionadas que se ajustam à sua workload específica.

Monitore as mudanças no consumo de recursos ao longo do tempo. Identifique onde as mudanças acumuladas resultam no consumo ineficiente ou significativamente maior de recursos. Determine a necessidade de melhorias para resolver os desafios no consumo e implemente aquelas que você prioriza.

As etapas a seguir foram projetadas para ser um processo iterativo que avalia, prioriza, testa e implementa melhorias focadas na sustentabilidade para workloads na nuvem.

1. Identifique metas de melhoria: revise suas workloads em relação às práticas recomendadas de sustentabilidade identificadas neste documento e identifique metas de melhoria.
2. Avalie melhorias específicas: avalie alterações específicas quanto a possíveis aprimoramentos, custo projetado e risco empresarial.
3. Priorize e planeje melhorias: priorize as mudanças que oferecem as maiores melhorias com o menor custo e risco e estabeleça um plano para testes e implementação.
4. Teste e valide melhorias: implemente alterações nos ambientes de teste para validar seu potencial de melhoria.

5. Implemente alterações na produção: implemente alterações em todos os ambientes de produção.
6. Avalie os resultados e replique os sucessos: procure oportunidades de replicar os sucessos em todas as workloads e reverta as alterações com resultados inaceitáveis.

Exemplo de cenário

O exemplo de cenário a seguir é mencionado posteriormente neste documento para ilustrar cada etapa do processo de melhoria.

Sua empresa tem uma workload que executa manipulações complexas de imagens em instâncias do Amazon EC2 e armazena os arquivos modificados e originais para acesso do usuário. As atividades de processamento consomem muita CPU e os arquivos de saída são extremamente grandes.

Identificar metas para melhoria

Entenda as práticas recomendadas que podem ajudar a atingir suas metas de sustentabilidade. Descrições detalhadas dessas [práticas recomendadas](#) e outras recomendações para melhorias podem ser encontradas posteriormente neste documento.

Analise suas workloads e os recursos usados. Identifique pontos críticos, como grandes implantações e recursos usados com frequência. Avalie esses pontos críticos em busca de oportunidades de melhorar a utilização efetiva de seus recursos e reduzir o total de recursos necessários para alcançar seus resultados comerciais.

Analise sua workload em relação às práticas recomendadas e identifique candidatos para melhorias.

Ao aplicar essa etapa ao [Exemplo de cenário](#), você identifica as seguintes práticas recomendadas como prováveis alvos de melhoria:

- Use a quantidade mínima de hardware para atender às suas necessidades
- Use tecnologias que melhor ofereçam suporte aos seus padrões de acesso e armazenamento de dados

Recursos

- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte I: Computação](#)

- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte II: Armazenamento](#)
- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte III: Redes](#)

Avaliar melhorias específicas

Entenda os recursos provisionados pela sua workload para concluir uma unidade de trabalho. Avalie possíveis melhorias e estime seu impacto potencial, o custo de implementação e os riscos associados.

Para medir as melhorias ao longo do tempo, primeiro entenda o que você provisionou na AWS e como esses recursos estão sendo consumidos.

Comece com uma visão geral completa do seu uso da AWS e use os relatórios de custo e uso da AWS para ajudar a identificar pontos críticos. Use este [código de exemplo da AWS](#) para obter ajuda para revisar e analisar seu relatório com o Amazon Athena.

Métricas proxy

Ao avaliar alterações específicas, você também deve avaliar quais métricas melhor quantificam o efeito dessa alteração no recurso associado. Essas métricas são chamadas de métricas proxy. Selecione métricas proxy que melhor reflitam o tipo de melhoria que você está avaliando e os recursos visados pela melhoria. Essas métricas podem evoluir com o tempo.

Os recursos provisionados para oferecer suporte à sua workload incluem os recursos de computação, armazenamento e rede. Avalie os recursos provisionados usando suas métricas proxy para ver como esses recursos são consumidos.

Use suas métricas proxy para medir os recursos provisionados para alcançar resultados comerciais.

Recurso	Exemplo de métricas proxy	Metas de melhoria
Computação	Minutos de vCPU	Maximize a utilização dos recursos provisionados
Armazenamento	GB provisionados	Reduza o total provisionado
Rede	GB transferidos ou pacotes transferidos	Reduza o total transferido e a distância da transferência

Métricas empresariais

Selecione métricas de negócios para quantificar a obtenção dos resultados comerciais. Suas métricas de negócios devem refletir o valor fornecido pela sua workload, por exemplo, o número de usuários ativos simultâneos, as chamadas de API atendidas ou o número de transações concluídas. Essas métricas podem evoluir com o tempo. Tenha cuidado ao avaliar métricas de negócios baseadas em finanças, pois a inconsistência no valor das transações invalida as comparações.

Indicadores-chave de performance

Usando a fórmula a seguir, divida os recursos provisionados pelos resultados comerciais alcançados para determinar os recursos provisionados por unidade de trabalho.

$$\text{Resources provisioned per unit of work} = \frac{\text{Proxy metric for provisioned resource}}{\text{Business metric for outcome}}$$

Fórmula de KPI

Use seus recursos por unidade de trabalho como KPIs. Estabeleça linhas de base para os recursos provisionados como base para comparações.

Recurso	Exemplos de KPIs	Metas de melhoria
Computação	Minutos de vCPU por transação	Maximize a utilização dos recursos provisionados
Armazenamento	GB por transação	Reduza o total provisionado
Rede	GB transferidos por transação ou pacotes transferidos por transação	Reduza o total transferido e a distância da transferência

Estimar as melhorias

Estime a melhoria com base na redução quantitativa nos recursos provisionados (conforme indicado por suas métricas de proxy) e na variação percentual dos recursos básicos provisionados por unidade de trabalho.

Recurso	Exemplos de KPIs	Metas de melhoria
Computação	% de redução de minutos de vCPUs por transação	Maximize a utilização
Armazenamento	% de redução de GB por transação	Reduza o total provisionado
Rede	% de redução de GB transferidos por transação ou pacotes transferidos por transação	Reduza o total transferido e a distância da transferência

Avaliar as melhorias

Avalie possíveis melhorias em relação ao benefício líquido previsto. Avalie o tempo, o custo e o nível de esforço para implementar e manter, além dos riscos comerciais, como impactos imprevistos.

Melhorias direcionadas geralmente representam compensações entre os tipos de recursos consumidos. Por exemplo, para reduzir o consumo de computação, é possível armazenar um resultado ou, para limitar os dados transferidos, você pode processar os dados antes de enviar o resultado para um cliente. Essas [compensações](#) serão discutidas em mais detalhes posteriormente.

Inclua requisitos não funcionais ao avaliar os riscos de sua workload, incluindo segurança, confiabilidade, eficiência de performance, otimização de custos e o impacto das melhorias em sua capacidade de operar sua workload.

Aplicando essa etapa ao [Exemplo de cenário](#), você avalia as melhorias desejadas com os seguintes resultados:

Prática recomendada	Melhoria direcionada	Potencial	Custos	Risco
Use a quantidade e mínima de hardware para atender às suas necessidades	Implemente e ajuste de escala preditivo para reduzir	Médio	Baixo	Baixo

Prática recomendada	Melhoria direcionada	Potencial	Custos	Risco
	os períodos de baixa utilização			
Use tecnologias que melhor ofereçam suporte aos seus padrões de acesso e armazenamento de dados	Implemente mecanismos de compactação mais eficazes para reduzir o armazenamento total e o tempo necessário para alcançá-lo	Alta	Baixo	Baixo

A implementação do ajuste de escala preditivo reduz as horas de vCPU consumidas por instâncias subutilizadas ou não utilizadas, oferecendo benefícios moderados em relação aos mecanismos de ajuste de escala existentes com uma redução estimada de 11% nos recursos consumidos. Os custos envolvidos são baixos e incluem a configuração dos recursos de nuvem e a operação de ajuste de escala preditivo para o Amazon EC2 Auto Scaling. O risco é a limitação de performance quando a expansão é executada de forma reativa em resposta à demanda que excede as previsões.

A implementação de uma compactação mais eficaz pode ter um impacto significativo com grandes reduções no tamanho do arquivo em todas as imagens originais e manipuladas, gerando uma redução estimada de 25% nos requisitos de armazenamento na produção. A implementação do novo algoritmo é uma substituição de baixo esforço com pouco risco envolvido.

Priorizar e planejar melhorias

Priorize suas melhorias identificadas com base no maior impacto previsto com os menores custos e riscos aceitáveis.

Decida em quais melhorias focar inicialmente e inclua-as em seu roteiro de planejamento e desenvolvimento de recursos.

Ao aplicar essa etapa ao [Exemplo de cenário](#), você prioriza as melhorias desejadas da seguinte forma:

Priority	Melhoria	Potencial	Custos	Risco
1	Implementar mecanismos de compactação mais eficazes	Alta	Baixo	Baixo
2	Implementar o ajuste de escala preditivo	Médio	Baixo	Baixo

O alto potencial, o baixo custo e o risco de atualizar a compactação de arquivos a tornam uma meta de alto valor para sua empresa e uma prioridade em relação à implementação do ajuste de escala preditivo. Você determina que a implementação do ajuste de escala preditivo com seu impacto potencial médio, baixo custo e baixo risco deve ser a melhoria prioritária após a conclusão da compactação do arquivo.

Você define um membro da equipe para implementar uma compactação aprimorada de arquivos e adicionar ajuste de escala preditivo à sua lista de pendências.

Testar e validar as melhorias

Realize pequenos testes com investimento mínimo para reduzir o risco de um esforço em grande escala.

Implemente uma cópia representativa de sua workload em seu ambiente de teste para limitar o custo e o risco de realizar testes e validação. Execute um conjunto predefinido de transações de teste, meça os recursos provisionados e determine os recursos usados por unidade de trabalho para estabelecer uma linha de base de teste.

Implemente sua meta de melhoria no ambiente de teste e repita o teste usando a mesma metodologia sob as mesmas condições. Em seguida, meça os recursos provisionados e os recursos usados por unidade de trabalho com sua melhoria implementada.

Calcule a variação percentual de sua linha de base dos recursos provisionados por unidade de trabalho e determine a redução quantitativa esperada nos recursos provisionados em seu ambiente de produção. Compare esses valores com os valores previstos. Determine se o resultado é um nível

aceitável de melhoria. Avalie se alguma compensação nos recursos adicionais consumidos torna o benefício líquido da melhoria inaceitável.

Determine se a melhoria foi um sucesso e se os recursos devem ser investidos na implementação da mudança na produção. Se a mudança for avaliada como malsucedida no momento, redirecione seus recursos para testar e validar sua próxima meta e continuar seu ciclo de melhoria.

% de redução nos recursos provisionados por unidade de trabalho	Redução quantitativa nos recursos provisionados	Ação
Atendeu às expectativas	Atendeu às expectativas	Prossiga com a melhoria
Não atendeu às expectativas	Atendeu às expectativas	Prossiga com a melhoria
Atendeu às expectativas	Não atendeu às expectativas	Buscar melhorias alternativas
Não atendeu às expectativas	Não atendeu às expectativas	Buscar melhorias alternativas

Aplicando essa etapa ao [Exemplo de cenário](#), você executa testes para validar o sucesso.

Depois de realizar os testes no algoritmo de compactação aprimorado, a redução percentual nos recursos provisionados por unidade de trabalho (o armazenamento necessário tanto para a imagem original quanto para a imagem modificada) atendeu às expectativas com uma redução média de 30% no armazenamento provisionado e um aumento insignificante da carga computacional.

Você determina que os recursos computacionais adicionais necessários para aplicar o algoritmo de compactação aprimorado aos arquivos existentes em produção são insignificantes em comparação com a redução no armazenamento alcançada. Você confirmou o sucesso com a redução quantitativa dos recursos necessários (TBs de armazenamento), e a melhoria foi aprovada para implantação em produção.

Implantar alterações na produção

Implemente melhorias testadas, validadas e aprovadas na produção. Implemente usando implantações limitadas, confirme a funcionalidade de sua workload, teste a redução real nos recursos provisionados e nos recursos consumidos por unidade de trabalho dentro da implantação limitada e

verifique as consequências não intencionais da mudança. Prossiga com as implantações completas após o teste bem-sucedido.

Reverta as alterações se os testes falharem ou se você encontrar consequências não intencionais inaceitáveis da sua alteração.

Aplicando essa etapa ao [Exemplo de cenário](#), você executa as ações a seguir.

Você implementa as mudanças na produção usando uma implantação limitada por meio de uma metodologia de implantação azul-verde. Os testes de funcionalidade nas instâncias recém-implantadas foram bem-sucedidos. Você observa uma redução média de 26% no armazenamento provisionado para arquivos de imagem originais e manipulados. Você não observa nenhuma evidência de um aumento na carga de computação compactando novos arquivos.

Você percebe uma diminuição inesperada no tempo decorrido para compactar arquivos de imagem e atribui isso ao código altamente otimizado do novo algoritmo de compactação.

Você prossegue com a implantação completa da nova versão.

Avaliar os resultados e replicar os sucessos

Avalie os resultados e replique os sucessos das seguintes formas:

- Meça a melhoria inicial dos recursos provisionados por unidade de trabalho e a diminuição quantitativa nos recursos provisionados.
- Compare as estimativas iniciais e os resultados dos testes com suas medições de produção. Identifique os fatores que possam ter contribuído para as diferenças e atualize suas metodologias de estimativa e teste quando apropriado.
- Determine o sucesso e o grau de sucesso e compartilhe os resultados com as partes interessadas.
- Se você precisou reverter as alterações devido a falhas nos testes ou a consequências negativas não intencionais da alteração, identifique os fatores contribuintes. Faça iterações onde for viável ou avalie novas abordagens para atingir os objetivos da alteração.
- Pegue o que você aprendeu, estabeleça padrões e aplique melhorias bem-sucedidas em outros sistemas que também podem se beneficiar. Capture e compartilhe sua metodologia, artefatos relacionados e benefícios líquidos entre equipes e organizações para que outras pessoas possam adotar seu padrão e replicar seu sucesso.
- Monitore os recursos provisionados por unidade de trabalho e acompanhe as alterações e o impacto total ao longo do tempo. Mudanças em sua workload, ou na forma como seus

clientes consomem a workload, podem ter um impacto na eficácia de sua melhoria. Reavalie as oportunidades de melhoria se notar reduções significativas de curto prazo na eficácia de sua melhoria ou uma redução acumulada na eficácia ao longo do tempo.

- Quantifique o benefício líquido de sua melhoria ao longo do tempo (incluindo os benefícios recebidos por outras equipes que aplicaram sua melhoria, se disponível) para mostrar o retorno sobre o investimento das suas atividades de melhoria.

Ao aplicar essa etapa ao [Exemplo de cenário](#), você mede os resultados a seguir.

Sua workload mostra uma melhoria inicial de 23% de redução nos requisitos de armazenamento após a implantação e aplicação do novo algoritmo de compactação aos arquivos de imagem existentes.

O valor medido está amplamente de acordo com as estimativas iniciais (25%), e a diferença significativa em relação ao teste (30%) é determinada como resultado do fato de os arquivos de imagem usados nos testes não serem representativos dos arquivos de imagem presentes na produção. Você modifica o conjunto de imagens de teste para refletir mais adequadamente as imagens em produção.

A melhoria é considerada um sucesso total. A redução total no armazenamento provisionado é 2% menor do que os 25% estimados, mas 23% ainda é uma grande melhoria no impacto da sustentabilidade e é acompanhada por uma redução de custos equivalente.

As únicas consequências não intencionais da alteração são a redução benéfica no tempo decorrido para realizar a compactação e uma redução equivalente no consumo de vCPU. Essas melhorias são atribuídas ao código altamente otimizado.

Você estabelece um projeto interno de código aberto no qual compartilha seu código, artefatos associados, orientações sobre como implementar a mudança e os resultados de sua implementação. O projeto interno de código aberto facilita que suas equipes adotem o código para todos os casos de uso persistente de armazenamento de arquivos. Suas equipes adotam a melhoria como padrão. Os benefícios secundários do projeto interno de código aberto são que todos que adotam a solução se beneficiam das melhorias na solução, e qualquer pessoa pode contribuir com melhorias no projeto.

Você publica seu sucesso e compartilha o projeto de código aberto em toda a sua organização. Cada equipe que adota a solução replica o benefício com um investimento mínimo e aumenta o benefício líquido recebido do seu investimento. Você publica esses dados como uma história de sucesso contínua.

Você continua monitorando o impacto da melhoria ao longo do tempo e fará alterações no projeto interno de código aberto conforme necessário.

Sustentabilidade como um requisito não funcional

Adicionar sustentabilidade à sua lista de requisitos empresariais pode gerar resultados mais econômicos. Concentrar-se em obter mais valor dos recursos que você usa e usar menos deles se traduz diretamente em redução de custos na AWS, pois você paga apenas pelo que usa.

O cumprimento das metas de sustentabilidade pode não exigir compensações equivalentes em uma ou mais métricas tradicionais, como tempo de atividade, disponibilidade ou tempo de resposta. Você pode obter ganhos significativos em sustentabilidade sem causar impacto mensurável nos níveis de serviço. Quando pequenas compensações são necessárias, as melhorias de sustentabilidade obtidas por essas compensações podem compensar a mudança na qualidade do serviço.

Incentive os membros da sua equipe a experimentar continuamente melhorias de sustentabilidade à medida que desenvolvem requisitos funcionais. As equipes também devem incorporar métricas de proxy ao definir metas para garantir que avaliem a intensidade dos recursos ao desenvolver suas workloads.

Veja a seguir exemplos de compensações que podem reduzir os recursos de nuvem que você consome:

Ajuste a qualidade do resultado: você pode trocar a qualidade dos resultados (QoR) por uma redução na intensidade da workload usando computação aproximada. A prática da computação aproximada procura oportunidades para explorar a lacuna entre o que é absolutamente necessário e o que você realmente produz. Por exemplo, se você colocar seus dados em uma estrutura de dados definida, poderá eliminar o operador ORDER BY no código SQL para remover o processamento desnecessário, economizando recursos e, ao mesmo tempo, fornecendo uma resposta aceitável.

Ajuste o tempo de resposta: uma resposta com um tempo de resposta mais lento (RT) pode economizar carbono minimizando a sobrecarga compartilhada. O processamento de tarefas ad hoc e temporárias pode incorrer em sobrecarga de inicialização. Agrupe tarefas e processe-as em lotes, em vez de pagar pela sobrecarga sempre que uma tarefa chegar. O processamento em lote troca maior tempo de resposta por uma redução na sobrecarga compartilhada da geração de uma instância, do download do código-fonte e da execução do processo.

Ajuste a disponibilidade: com a AWS, é possível adicionar redundância e atingir metas de alta disponibilidade com apenas alguns cliques. A redundância pode ser aumentada por meio de técnicas como estabilidade estática com o provisionamento de recursos ociosos que sempre resultam em menor utilização. Avalie as necessidades comerciais ao definir metas. Compensações relativamente

pequenas na disponibilidade podem resultar em melhorias muito maiores na utilização. Por exemplo, o padrão de arquitetura com estabilidade estática envolve o provisionamento de capacidade de failover ocioso para carregar imediatamente após uma falha de componente. Relaxar o requisito de disponibilidade pode eliminar a necessidade de capacidade online ociosa fazendo com que haja tempo para a automação implantar recursos de substituição. A adição de capacidade de failover sob demanda aumenta a utilização geral sem afetar os negócios durante as operações normais.

Práticas recomendadas para a sustentabilidade na nuvem

Otimize o posicionamento da workload e otimize sua arquitetura de acordo com a demanda, o software, os dados, o hardware e o processo para aumentar a eficiência energética. Cada uma dessas áreas oferece a oportunidade de empregar as práticas recomendadas a fim de reduzir o impacto na sustentabilidade da workload em nuvem por meio da maximização do uso e da minimização do desperdício e do total de recursos implantados e alimentados para comportar a workload.

Tópicos

- [Seleção da região](#)
- [Alinhamento com a demanda](#)
- [Software e arquitetura](#)
- [Gerenciamento de dados](#)
- [Hardware e serviços](#)
- [Processo e cultura](#)

Seleção da região

A escolha da região para sua workload afeta significativamente seus KPIs, incluindo performance, custo e pegada de carbono. Para melhorar efetivamente esses KPIs, você deve escolher regiões para suas workloads com base em requisitos empresariais e metas de sustentabilidade.

Práticas recomendadas

- [SUS01-BP01 Escolher a região com base nos requisitos empresariais e nas metas de sustentabilidade](#)

SUS01-BP01 Escolher a região com base nos requisitos empresariais e nas metas de sustentabilidade

Escolha uma região para sua workload com base em seus requisitos empresariais e metas de sustentabilidade para otimizar seus KPIs, incluindo performance, custo e pegada de carbono.

Práticas comuns que devem ser evitadas:

- Selecionar a região da workload com base em sua localização.
- Consolidar todos os recursos da workload em uma única localização geográfica.

Benefícios de implementar esta prática recomendada: colocar uma workload próxima aos projetos de energia renovável da Amazon ou às regiões com baixa intensidade de carbono publicada pode ajudar a reduzir a pegada de carbono de uma workload na nuvem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A Nuvem AWS é uma rede em constante expansão de regiões e pontos de presença (PoP) com uma infraestrutura de rede global que conecta uns aos outros. A escolha da região para sua workload afeta significativamente seus KPIs, incluindo performance, custo e pegada de carbono. Para melhorar efetivamente esses KPIs, você deve escolher regiões para sua workload com base em seus requisitos empresariais e metas de sustentabilidade.

Etapas de implementação

- Selecione possíveis regiões: siga estas etapas para avaliar e selecionar possíveis regiões para a workload com base em seus requisitos de negócios, incluindo conformidade, recursos disponíveis, custo e latência:
 - Confirme se essas regiões estão em conformidade com as regulamentações locais aplicáveis (por exemplo, soberania de dados).
 - Use as [Listas de serviços regionais da AWS](#) para verificar se as regiões têm os serviços e recursos necessários para executar sua workload.
 - Calcule o custo da workload em cada região usando o [AWS Calculadora de Preços](#).
 - Teste a latência de rede entre as localizações de seus usuários finais e cada Região da AWS.
- Escolha regiões: escolha regiões próximas aos projetos de energia renovável da Amazon e regiões onde a grade de intensidade de carbono publicada esteja abaixo de outros locais (ou regiões).
 - Identifique suas diretrizes de sustentabilidade relevantes para rastrear e comparar as emissões de carbono ano a ano com base no [Protocolo de Gases de Efeito Estufa](#) (métodos baseados no mercado e na localização).
 - Escolha a região com base no método que você utiliza para rastrear as emissões de carbono. Para obter mais detalhes sobre como escolher uma região em função das suas diretrizes de

sustentabilidade, consulte [Como selecionar uma região para sua workload com base em metas de sustentabilidade](#).

Recursos

Documentos relacionados:

- [Noções básicas das suas estimativas de emissão de carbono](#)
- [Amazon ao redor do mundo](#)
- [Metodologia de energia renovável](#)
- [O que considerar ao selecionar uma região para suas workloads](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Inovação sustentável na infraestrutura global da AWS](#)
- [AWS re:Invent 2023: Arquitetura sustentável: passado, presente e futuro](#)
- [AWS re:Invent 2022: Como entregar arquiteturas sustentáveis e de alta performance](#)
- [AWS re:Invent 2022: Arquitetar de forma sustentável e reduzir sua pegada de carbono da AWS](#)
- [AWS re:Invent 2022: Sustentabilidade na infraestrutura global da AWS](#)

Alinhamento com a demanda

A maneira como os usuários e as aplicações consomem suas workloads e outros recursos pode ajudar você a identificar melhorias para atingir metas de sustentabilidade. Escale a infraestrutura de forma que ela corresponda à demanda e use apenas os recursos mínimos necessários para oferecer suporte aos usuários. Alinhe os níveis de serviço às necessidades do cliente. Posicione os recursos a fim de limitar a rede necessária para que usuários e aplicações os consumam. Elimine ativos não utilizados. Forneça aos membros da sua equipe dispositivos compatíveis com suas necessidades e minimize o impacto na sustentabilidade.

Práticas recomendadas

- [SUS02-BP01 Escalar a infraestrutura da workload dinamicamente](#)
- [SUS02-BP02 Alinhar os SLAs às metas de sustentabilidade](#)
- [SUS02-BP03 Interromper a criação e a manutenção de ativos não utilizados](#)

- [SUS02-BP04 Otimizar o posicionamento geográfico das workloads com base nos respectivos requisitos de rede](#)
- [SUS02-BP05 Otimizar os recursos dos membros da equipe para as atividades realizadas](#)
- [SUS02-BP06 Implementar armazenamento em buffer ou controle de utilização para nivelar a curva da demanda](#)

SUS02-BP01 Escalar a infraestrutura da workload dinamicamente

Use a elasticidade da nuvem e escale sua infraestrutura de forma dinâmica para corresponder a oferta de recursos de nuvem à demanda e evitar capacidade superprovisionada em sua workload.

Práticas comuns que devem ser evitadas:

- Você não escala sua infraestrutura de acordo com a carga de usuários.
- Você escala sua infraestrutura manualmente o tempo todo.
- Manter a capacidade aumentada após um evento de ajuste de escala, em vez de reduzi-la novamente.

Benefícios de implementar esta prática recomendada: configurar e testar a elasticidade da workload ajuda a adequar eficientemente a oferta de recursos de nuvem à demanda e evitar o excesso de capacidade provisionada. É possível aproveitar a elasticidade na nuvem para escalar automaticamente a capacidade durante e depois de picos de demanda para garantir que esteja usando apenas o número exato de recursos necessários para atender aos requisitos do seu negócio.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A nuvem fornece a flexibilidade necessária para expandir ou reduzir seus recursos dinamicamente por meio de diversos mecanismos para atender a mudanças na demanda. O equilíbrio ideal entre a oferta e a demanda oferece o menor impacto ambiental para uma workload.

A demanda pode ser fixa ou variável, exigindo métricas e automação para garantir que o gerenciamento não se torne um gasto excessivo. As aplicações podem aumentar ou diminuir a escala verticalmente ao modificar o tamanho da instância e horizontalmente ao modificar o número de instâncias, ou ainda uma combinação de ambos.

Você pode usar diversas abordagens diferentes para corresponder a oferta de recursos com a demanda.

- Abordagem de rastreamento de métricas: monitore a métrica de ajuste de escala e aumente ou diminua automaticamente a capacidade conforme necessário.
- Ajuste de escala preditivo: aumente ou reduza a escala em antecipação às tendências diárias e semanais.
- Abordagem baseada em cronograma: defina seu próprio cronograma de ajuste de escala de acordo com as mudanças de carga previsíveis.
- Dimensionamento de serviços: escolha serviços (como sem servidor) cuja escala seja modificada de forma nativa por design ou forneçam o ajuste de escala automático como um recurso.

Identifique períodos de utilização baixa ou sem utilização e escale os recursos para eliminar a capacidade em excesso e melhorar a eficiência.

Etapas de implementação

- A elasticidade corresponde à oferta de recursos que você tem face à demanda por estes recursos. Instâncias, contêineres e funções fornecem mecanismos para elasticidade, seja em combinação com o ajuste de escala automático ou como um recurso do serviço. A AWS fornece uma variedade de mecanismos de ajuste de escala automático para garantir que as workloads possam reduzir a escala verticalmente de forma rápida e fácil durante períodos de baixa carga de usuários. Veja alguns exemplos de mecanismos de ajuste de escala automático:

Mecanismo de ajuste de escala automático	Onde usar
Amazon EC2 Auto Scaling	Use para verificar se você tem o número correto de instâncias do Amazon EC2 disponíveis para processar a carga de usuário para a sua aplicação.
Application Auto Scaling	Use para escalar automaticamente os recursos para serviços da AWS individuais além do Amazon EC2, como funções do Lambda ou serviços do Amazon Elastic Container Service (Amazon ECS).

Mecanismo de ajuste de escala automático	Onde usar
Kubernetes Cluster Autoscaler	Use para escalar automaticamente os clusters do Kubernetes na AWS.

- O ajuste de escala geralmente é discutido em relação a serviços de computação, como instâncias do Amazon EC2 ou funções do AWS Lambda. Considere a configuração de serviços não computacionais, como unidades de capacidade de leitura e gravação do [Amazon DynamoDB](#) ou fragmentos do [Amazon Kinesis Data Streams](#) para atender à demanda.
- Verifique se as métricas para aumentar ou reduzir a escala verticalmente são validadas em relação ao tipo da workload que está sendo implantada. Se você estiver implantando uma aplicação de transcodificação de vídeo, espera-se que a utilização da CPU seja de 100%, e essa não deve ser sua métrica principal. É possível usar uma [métrica personalizada](#) (como utilização de memória) para sua política de ajuste de escala, se necessário. Para escolher as métricas certas, considere a seguinte orientação para o Amazon EC2:
 - A métrica deve ser uma métrica de utilização válida e descrever o quanto uma instância está ocupada.
 - O valor da métrica deve aumentar e diminuir em proporção ao número das instâncias no grupo do Auto Scaling.
- Use o [ajuste de escala dinâmico](#) em vez do [ajuste de escala manual](#) para seu grupo do Auto Scaling. Também recomendamos usar [políticas de ajuste de escala de rastreamento de metas](#) em seu ajuste de escala dinâmico.
- Verifique se as implantações da workload podem lidar com eventos de aumento e redução horizontal da escala. Crie cenários de teste para eventos de redução horizontal da escala para verificar se a workload se comporta conforme o esperado e não afeta a experiência do usuário (como perda da sessão persistente). É possível usar o [Histórico de atividades](#) para verificar uma atividade de ajuste de escala para um grupo do Auto Scaling.
- Avalie sua workload em relação a padrões previsíveis e, ao antecipar alterações previstas e planejadas na demanda, ajuste a escala proativamente. Com o ajuste de escala preditivo, é possível eliminar a necessidade de superprovisionar a capacidade. Para obter mais informações, consulte [Ajuste de escala preditivo com o Amazon EC2 Auto Scaling](#).

Recursos

Documentos relacionados:

- [Conceitos básicos do Amazon EC2 Auto Scaling](#)
- [Ajuste de escala preditivo para o EC2 com Machine Learning](#)
- [Analisar o comportamento do usuário usando o Amazon OpenSearch Service, o Amazon Kinesis Data Firehose e o Kibana](#)
- [O que é o Amazon CloudWatch?](#)
- [Monitorar a workload de banco de dados com o Performance Insights no Amazon RDS](#)
- [Introdução de suporte nativo para ajuste de escala preditivo com o Amazon EC2 Auto Scaling](#)
- [Introdução ao Karpenter: um dimensionador automático de escala de clusters do Kubernetes de código aberto e alta performance](#)
- [Mergulho profundo no ajuste de escala automático de clusters do Amazon ECS](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Escalar na AWS para seus primeiros 10 milhões de usuários](#)
- [AWS re:Invent 2023: Arquitetura sustentável: passado, presente e futuro](#)
- [AWS re:Invent 2022: Criar um ambiente de computação eficiente em termos de custo, energia e recursos](#)
- [AWS re:Invent 2022: Escalar contêineres de um para milhões de usuários](#)
- [AWS re:Invent 2023: Escalar a inferência de FM para centenas de modelos com o Amazon SageMaker AI](#)
- [AWS re:Invent 2023: Aproveitar o poder do Karpenter para escalar, otimizar e atualizar o Kubernetes](#)

Exemplos relacionados:

- [Ajuste de escala automático](#)

SUS02-BP02 Alinhar os SLAs às metas de sustentabilidade

Analise e otimize os acordos de serviço (SLA) com base em suas metas de sustentabilidade para minimizar os recursos necessários a fim de oferecer compatibilidade com sua workload enquanto continua a atender às necessidades empresariais.

Práticas comuns que devem ser evitadas:

- SLAs de workload são desconhecidos ou ambíguos.
- Você define seu SLA apenas para disponibilidade e performance.
- Você usa o mesmo padrão de design (como arquitetura multi-AZ) para todas as suas workloads.

Benefícios de implementar esta prática recomendada: alinhar os SLAs às metas de sustentabilidade leva ao uso ideal dos recursos e, ao mesmo tempo, atende às necessidades de negócios.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Os SLAs definem o nível de serviço esperado de uma workload de nuvem, como tempo de resposta, disponibilidade e retenção de dados. Eles influenciam a arquitetura, o uso de recursos e o impacto ambiental de uma workload na nuvem. Em uma cadência regular, analise os SLAs e faça compensações que reduzam significativamente o uso de recursos em troca de reduções aceitáveis em níveis de serviço.

Etapas de implementação

- Entenda as metas de sustentabilidade: identifique metas de sustentabilidade em sua organização, como redução de carbono ou melhoria na utilização de recursos.
- Analise os SLAs: avalie seus SLAs para avaliar se eles atendem aos requisitos da sua empresa. Se estiver superando os SLAs, faça uma análise adicional.
- Entenda as vantagens e desvantagens: entenda as vantagens e desvantagens da complexidade da sua workload (como alto volume de usuários simultâneos), performance (como latência) e impacto na sustentabilidade (como os recursos necessários). Normalmente, priorizar dois dos fatores inviabiliza o terceiro.
- Ajuste os SLAs: ajuste os SLAs fazendo compensações que reduzam significativamente os impactos na sustentabilidade em troca de reduções aceitáveis em níveis de serviço.
 - Sustentabilidade e confiabilidade: workloads altamente disponíveis tendem a consumir mais recursos.
 - Sustentabilidade e performance: usar mais recursos para aumentar a performance pode ter um impacto ambiental maior.
 - Sustentabilidade e segurança: workloads excessivamente seguras podem ter um impacto ambiental maior.

- Defina SLAs de sustentabilidade, se possível: inclua SLAs de sustentabilidade para sua workload. Por exemplo, defina um nível mínimo de utilização como um SLA de sustentabilidade para as instâncias de computação.
- Use padrões de design eficientes: use padrões de design, como microsserviços na AWS, que priorizem funções essenciais aos negócios e permita níveis de serviço mais baixos (como objetivos de tempo de resposta ou de tempo de recuperação) para funções não essenciais.
- Estabeleça comunicação e responsabilidade: compartilhe os SLAs com todas as partes interessadas relevantes, incluindo sua equipe de desenvolvimento e seus clientes. Use relatórios para rastrear e monitorar os SLAs. Atribua responsabilidades para alcançar as metas de sustentabilidade dos SLAs.
- Use incentivos e recompensas: use incentivos e recompensas para alcançar ou superar os SLAs alinhados às metas de sustentabilidade.
- Revise e repita: revise e ajuste regularmente seus SLAs para garantir que estejam alinhados às metas de sustentabilidade e performance em evolução.

Recursos

Documentos relacionados:

- [Entender padrões de resiliência e compromissos para arquitetar de forma eficiente na nuvem](#)
- [Importância do acordo de nível de serviço para provedores de SaaS](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Capacidade, disponibilidade, eficiência de custos: escolha três](#)
- [AWS re:Invent 2023: Arquitetura sustentável: passado, presente e futuro](#)
- [AWS re:Invent 2023: padrões de integração avançados e compromissos para sistemas com acoplamento fraco](#)
- [AWS re:Invent 2022: Como entregar arquiteturas sustentáveis e de alta performance](#)
- [AWS re:Invent 2022: Criar um ambiente de computação eficiente em termos de custo, energia e recursos](#)

SUS02-BP03 Interromper a criação e a manutenção de ativos não utilizados

Desative ativos em sua workload para reduzir o número de recursos necessários para atender à sua demanda e minimizar o desperdício.

Práticas comuns que devem ser evitadas:

- Você não analisa sua aplicação com relação a ativos redundantes ou não mais necessários.
- Você não remove ativos redundantes ou que não mais necessários.

Benefícios de implementar esta prática recomendada: a remoção de ativos não utilizados libera recursos e melhora a eficiência geral da workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Os ativos ociosos consomem recursos de nuvem como espaço de armazenamento e potência computacional. Com a identificação e eliminação desses ativos, você pode liberar esses recursos e aumentar a eficiência da arquitetura de nuvem. Analise regularmente os ativos de aplicações (como relatórios pré-compilados, conjuntos de dados e imagens estáticas) e os padrões de acesso aos ativos para identificar redundâncias, subutilização e possíveis alvos de desativação. Remova esses ativos redundantes para diminuir o desperdício de recursos em sua workload.

Etapas de implementação

- Faça um inventário: faça um inventário abrangente para identificar todos os ativos em sua workload.
- Analise o uso: use ferramentas de monitoramento para identificar ativos estáticos que não são mais necessários.
- Remova ativos não utilizados: desenvolva um plano para remover ativos que não são mais necessários.
 - Antes de remover qualquer ativo, avalie o impacto da remoção sobre a arquitetura.
 - Consolide ativos gerados sobrepostos para remover o processamento redundante.
 - Atualize suas aplicações para que não produzam nem armazenem mais ativos que não são necessários.

- **Comunique-se com terceiros:** instrua terceiros a interromper a produção e o armazenamento de ativos gerenciados em seu nome que não sejam mais necessários. Peça que consolidem ativos redundantes.
- **Use políticas de ciclo de vida:** use políticas de ciclo de vida para excluir automaticamente ativos não utilizados.
 - Você pode usar o [Amazon S3 Lifecycle](#) para gerenciar seus objetos durante todo o ciclo de vida de cada um.
 - É possível usar o [Amazon Data Lifecycle Manager](#) para automatizar a criação, a retenção e a exclusão de snapshots do Amazon EBS e de AMIs apoiadas pelo Amazon EBS.
- **Revise e optimize:** revise regularmente sua workload para identificar e remover ativos ociosos.

Recursos

Documentos relacionados:

- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte II: Armazenamento](#)
- [Como faço para encerrar recursos ativos dos quais não preciso mais em minha Conta da AWS?](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Arquitetura sustentável: passado, presente e futuro](#)
- [AWS re:Invent 2022: Preservar e maximizar o valor de ativos de mídia digital usando o Amazon S3](#)
- [AWS re:Invent 2023: Otimizar custos em seus ambientes com várias contas](#)

SUS02-BP04 Otimizar o posicionamento geográfico das workloads com base nos respectivos requisitos de rede

Selecione locais e serviços de nuvem para sua workload que reduzam a distância que o tráfego de rede deve percorrer e diminua o total de recursos de rede necessários para comportar a workload.

Práticas comuns que devem ser evitadas:

- Selecione a região da workload com base em sua localização.
- Consolidar todos os recursos da workload em uma única localização geográfica.
- Todo o tráfego flui por meio dos data centers existentes.

Benefícios de implementar esta prática recomendada: implantar uma workload perto dos clientes proporciona a latência mais baixa enquanto reduz a movimentação de dados pela rede e reduz o impacto ambiental.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A infraestrutura da Nuvem AWS é construída em torno de opções de local, como regiões, zonas de disponibilidade, grupos de posicionamento e locais da borda, como [AWS Outposts](#) e [zonas locais da AWS](#). Essas opções de local são responsáveis por manter a conectividade entre componentes da aplicação, serviços de nuvem, redes da borda e data centers on-premises.

Analise os padrões de acesso à rede em sua workload para identificar como usar essas opções de local de nuvem e reduzir a distância que o tráfego de rede precisa percorrer.

Etapas de implementação

- Analise os padrões de acesso à rede em sua workload para identificar como os usuários utilizam sua aplicação.
 - Use ferramentas de monitoramento, como o [Amazon CloudWatch](#) e o [AWS CloudTrail](#), para coletar dados sobre atividades de rede.
 - Analise os dados para identificar o padrão de acesso à rede.
- Selecione as regiões para implantação da workload com base nos seguintes elementos fundamentais:
 - Sua meta de sustentabilidade: conforme explicado em [Seleção de regiões](#).
 - Onde seus dados estão localizados: para aplicações com uso intenso de dados (como big data e machine learning), o código da aplicação deve ser executado o mais perto possível dos dados.
 - Onde seus usuários estão localizados: para aplicações voltadas ao usuário, escolha uma ou mais regiões perto dos clientes da workload.
 - Outras restrições: considere restrições como custo e conformidade, conforme explicado em [O que considerar ao selecionar uma região para suas workloads](#).
- Use o armazenamento em cache local ou [soluções de armazenamento em cache da AWS](#) para dados usados com frequência a fim de aumentar a performance, reduzir a movimentação de dados e diminuir o impacto ambiental.

Serviço	Quando usar
Amazon CloudFront	Use para armazenar conteúdo estático em cache, como imagens, scripts e vídeos, além de conteúdo dinâmico como respostas de API ou aplicações Web.
Amazon ElastiCache	Use para armazenar conteúdo em cache para aplicações Web.
DynamoDB Accelerator	Use para adicionar aceleração na memória às suas tabelas do DynamoDB.

- Use serviços que podem ajudar você a executar código mais perto dos usuários da workload:

Serviço	Quando usar
Lambda@Edge	Use para operações com uso intenso de computação que são iniciadas quando objetos não estão no cache.
Amazon CloudFront Functions	Use para casos de uso simples, como solicitações HTTP(s) ou manipulações de resposta que podem ser iniciadas por funções de curta duração.
AWS IoT Greengrass	Use para executar computação local, mensagens e armazenamento de dados em cache para dispositivos conectados.

- Use o agrupamento de conexões para permitir a reutilização de conexões e reduzir os recursos necessários.
- Use datastores distribuídos que não dependem de conexões persistentes e atualizações síncronas para fins de consistência com o objetivo de atender a populações regionais.
- Substitua a capacidade de rede estática pré-provisionada por capacidade dinâmica compartilhada e divida o impacto sobre a sustentabilidade da capacidade de rede com outros assinantes.

Recursos

Documentos relacionados:

- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte III: Redes](#)
- [Documentação do Amazon ElastiCache](#)
- [O que é o Amazon CloudFront?](#)
- [Principais recursos do Amazon CloudFront](#)
- [Infraestrutura global da AWS](#)
- [Zonas locais da AWS e AWS Outposts: como escolher a tecnologia certa para sua workload de borda\)](#)
- [Grupos de posicionamento](#)
- [Zonas locais da AWS](#)
- [AWS Outposts](#)

Vídeos relacionados:

- [Desmistificar a transferência de dados na AWS](#)
- [Escalar a performance da rede em instâncias do Amazon EC2 de última geração](#)
- [Vídeo explicativo de zonas locais da AWS](#)
- [Visão geral do AWS Outposts e como ele funciona](#)
- [AWS re:Invent 2023: Uma estratégia de migração para workloads periféricas e on-premises](#)
- [AWS re:Invent 2021: AWS Outposts: como trazer a experiência da AWS para ambientes on-premises](#)
- [AWS re:Invent 2020: AWS Wavelength:executar aplicações com latência ultrabaixa na borda 5G](#)
- [AWS re:Invent 2022: Zonas locais da AWS: como criar aplicações para uma borda distribuída](#)
- [AWS re:Invent 2021: Criar sites de baixa latência com o Amazon CloudFront](#)
- [AWS re:Invent 2022: Aprimorar a performance e a disponibilidade com o AWS Global Accelerator](#)
- [AWS re:Invent 2022: Criar sua rede de longa distância usando a AWS](#)
- [AWS re:Invent 2020: Gerenciamento de tráfego global com o Amazon Route 53](#)

Exemplos relacionados:

- [Workshops de redes da AWS](#)
- [Arquitetura para a sustentabilidade: reduza a movimentação de dados entre redes](#)

SUS02-BP05 Otimizar os recursos dos membros da equipe para as atividades realizadas

Otimize os recursos fornecidos aos membros da equipe para minimizar o impacto sobre a sustentabilidade ambiental e, ao mesmo tempo, atender às suas necessidades.

Práticas comuns que devem ser evitadas:

- Você ignora o impacto dos dispositivos usados pelos membros da equipe sobre a eficiência geral de sua aplicação de nuvem.
- Você gerencia e atualiza manualmente os recursos usados pelos membros da equipe.

Benefícios de implementar esta prática recomendada: a otimização dos recursos dos membros da equipe melhora a eficiência geral das aplicações habilitadas para a nuvem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Conheça os recursos que os membros da equipe usam para consumir seus serviços, o ciclo de vida esperado e o impacto financeiro e na sustentabilidade. Implemente estratégias para otimizar esses recursos. Por exemplo, realize operações complexas, como renderização e compilação, em infraestrutura escalável com alta utilização em vez de em sistemas de usuário único subutilizados com alto consumo de energia.

Etapas de implementação

- Use estações de trabalho energeticamente eficientes: forneça aos membros da equipe estações de trabalho e periféricos com baixo consumo de energia. Use recursos eficientes para gerenciamento de energia (como o modo de baixo consumo de energia) nesses dispositivos para reduzir o uso de energia
- Use virtualização: use áreas de trabalho virtuais e a transmissão de aplicações para limitar os requisitos de upgrade e dispositivos.

- Incentive a colaboração remota: incentive os membros da equipe a usar ferramentas de colaboração remota como o [Amazon Chime](#) ou o [AWS Wickr](#) para reduzir a necessidade de viagens e as emissões de carbono associadas.
- Use software energeticamente eficiente: forneça aos membros da equipe software de baixo consumo de energia removendo ou desativando recursos e processos desnecessários.
- Gerencie os ciclos de vida: avalie o impacto de processos e sistemas no ciclo de vida de seus dispositivos e escolha soluções que minimizem o requisito de substituição de dispositivos e, ao mesmo tempo, atendam aos requisitos comerciais. Mantenha e atualize regularmente estações de trabalho e o software para manter e melhorar a eficiência.
- Gerenciamento remoto de dispositivos: implemente o gerenciamento remoto de dispositivos para reduzir as viagens de negócios.
 - O [AWS Systems Manager Fleet Manager](#) é uma experiência de interface do usuário (IU) unificada que ajuda você a gerenciar remotamente os nós em execução no AWS ou on-premises.

Recursos

Documentos relacionados:

- [O que é o Amazon WorkSpaces?](#)
- [Otimizador de custos para Amazon WorkSpaces](#)
- [Documentação do Amazon AppStream 2.0](#)
- [NICE DCV](#)

Vídeos relacionados:

- [Gerenciar os custos do Amazon WorkSpaces na AWS](#)

SUS02-BP06 Implementar armazenamento em buffer ou controle de utilização para nivelar a curva da demanda

O armazenamento em buffer e o controle de utilização nivelam a curva da demanda e reduzem a capacidade provisionada necessária para sua workload.

Práticas comuns que devem ser evitadas:

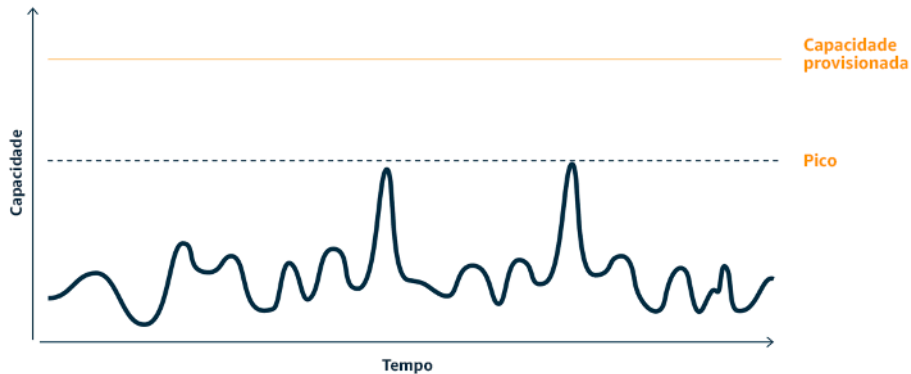
- Você processa imediatamente as solicitações de cliente embora isso não seja necessário.
- Você não analisa os requisitos das solicitações de cliente.

Benefícios de implementar esta prática recomendada: nivelar a curva de demanda reduz a capacidade provisionada necessária para a workload. Reduzir a capacidade provisionada significa diminuir o consumo de energia e o impacto ambiental.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

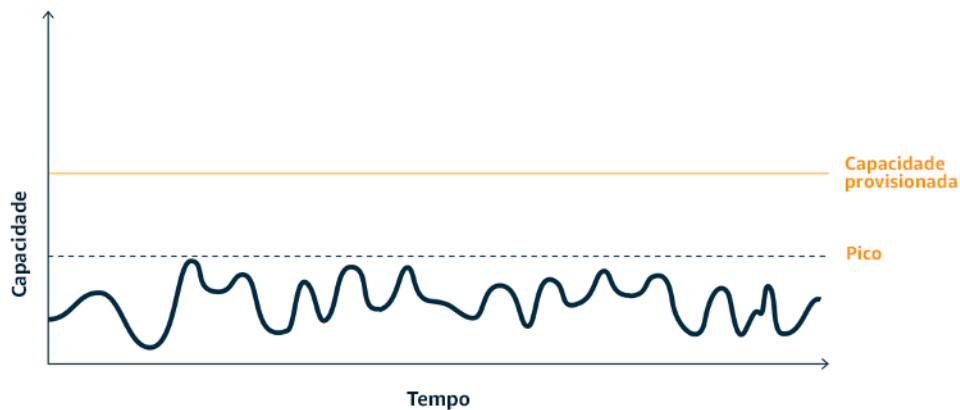
Orientação para implementação

Nivelar a curva da demanda pode ajudar você a reduzir a capacidade provisionada para uma workload e a diminuir o respectivo impacto ambiental. Considere a workload com a curva de demanda mostrada na figura a seguir. Essa workload tem dois picos e, para lidar com eles, é provisionada a capacidade de recurso mostrada pela linha laranja. Os recursos e energia usados para essa workload não são indicados pela área abaixo da curva da demanda, mas pela área abaixo da linha da capacidade provisionada, visto que é preciso ter capacidade provisionada para lidar com esses dois picos.



Curva de demanda com dois picos distintos que exigem alta capacidade provisionada.

Você pode usar o armazenamento em buffer ou o controle de utilização para modificar a curva da demanda e atenuar os picos, o que significa menor capacidade provisionada e menor consumo de energia. Implemente o controle de utilização quando seus clientes puderem realizar novas tentativas. Implemente o armazenamento em buffer para armazenar a solicitação e adiar o processamento até um momento posterior.



O efeito do controle de utilização na curva de demanda e na capacidade provisionada.

Etapas de implementação

- Analise as solicitações dos clientes para determinar como responder a elas. As perguntas a serem consideradas incluem:
 - Essa solicitação pode ser processada assincronamente?
 - O cliente tem capacidade de repetição?
- Se o cliente tiver capacidade de repetição, você poderá implementar o controle de utilização, que informa à origem que, se ela não puder atender à solicitação naquele momento, deverá tentar novamente mais tarde.
 - Você pode usar o [Amazon API Gateway](#) para implementar o controle de utilização.
- Para clientes que não podem realizar novas tentativas, é necessário implementar um buffer para nivelar a curva da demanda. O buffer adia o processamento de solicitações, permitindo que as aplicações executadas em diferentes taxas se comuniquem com eficácia. Uma abordagem baseada em buffer usa uma fila ou um fluxo para aceitar mensagens de produtores. As mensagens são lidas pelos consumidores e processadas, permitindo que as mensagens sejam executadas na taxa que atenda aos requisitos de negócios dos consumidores.
 - O [Amazon Simple Queue Service \(Amazon SQS\)](#) é um serviço gerenciado que fornece filas que permitem que um único consumidor leia mensagens individuais.
 - O [Amazon Kinesis](#) oferece um fluxo que permite que muitos consumidores leiam as mesmas mensagens.
- Analise a demanda geral, a taxa de alteração e o tempo de resposta necessário para dimensionar adequadamente o controle ou buffer necessário.

Recursos

Documentos relacionados:

- [Conceitos básicos do Amazon SQS](#)
- [Integrar aplicações usando filas e mensagens](#)
- [Gerenciar e monitorar o controle de utilização de APIs em suas workloads](#)
- [Controlar a utilização de uma API REST em camadas e multilocatário em grande escala usando o API Gateway](#)
- [Integrar aplicações usando filas e mensagens](#)

Vídeos relacionados:

- [AWS re:Invent 2022: padrões de integração de aplicações para microsserviços](#)
- [AWS re:Invent 2023: economias inteligentes: estratégias de otimização de custos no Amazon EC2](#)
- [AWS re:Invent 2023: padrões de integração avançados e compromissos para sistemas com acoplamento fraco](#)

Software e arquitetura

Implemente padrões que suavizem os picos de carga e mantenham a alta utilização consistente de recursos implantados para minimizar os recursos consumidos. Os componentes podem ficar ociosos devido à falta de uso por conta das mudanças no comportamento do usuário ao longo do tempo. Revise os padrões e a arquitetura para consolidar os componentes subutilizados a fim de aumentar a utilização geral. Retire os componentes que não são mais necessários. Saiba qual é a performance dos componentes de sua workload e otimize os componentes que consomem a maioria dos recursos. Esteja ciente dos dispositivos que seus clientes usam para acessar seus serviços e implemente padrões a fim de minimizar a necessidade de upgrades de dispositivos.

Práticas recomendadas

- [SUS03-BP01 Otimizar o software e a arquitetura para trabalhos assíncronos e agendados](#)
- [SUS03-BP02 Remover ou refatorar componentes da workload com pouco ou nenhum uso](#)
- [SUS03-BP03 Otimizar áreas de código que consomem mais tempo ou recursos](#)
- [SUS03-BP04 Otimizar o impacto sobre dispositivos e equipamentos](#)

- [SUS03-BP05 Usar padrões e arquiteturas de software que atendam melhor aos padrões de armazenamento e acesso a dados](#)

SUS03-BP01 Otimizar o software e a arquitetura para trabalhos assíncronos e agendados

Use software eficiente e padrões de arquitetura, como orientado a filas, para manter uma alta e consistente utilização dos recursos implantados.

Práticas comuns que devem ser evitadas:

- Provisione em excesso os recursos em sua workload na nuvem para atender a picos imprevistos na demanda.
- Sua arquitetura não separa remetentes e destinatários de mensagens assíncronas por um componente de sistema de mensagens.

Benefícios de implementar esta prática recomendada:

- Padrões eficientes de software e arquitetura minimizam os recursos não utilizados em sua workload e melhoram a eficiência geral.
- Você pode dimensionar o processamento independentemente do recebimento de mensagens assíncronas.
- Por meio de um componente de mensagens, você relaxou os requisitos de disponibilidade que podem ser atendidos com menos recursos.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Use padrões de arquitetura eficientes, como [arquitetura orientada por eventos](#), que resultam na utilização uniforme dos componentes e minimizam o excesso de provisionamento em sua workload. A utilização de padrões de arquitetura eficientes minimiza recursos ociosos por falta de uso devido a mudanças na demanda ao longo do tempo.

Entenda os requisitos de seus componentes de workload e adote padrões de arquitetura que aumentam a utilização geral dos recursos. Retire os componentes que não são mais necessários.

Etapas de implementação

- Analise a demanda da sua workload para determinar como responder a ela.
- Para solicitações ou trabalhos que não exigem respostas síncronas, use arquiteturas orientadas por filas e operadores de escalação automática para maximizar a utilização. Aqui estão alguns exemplos de quando você pode considerar a arquitetura orientada por filas:

Mecanismo de filas	Descrição
Filas de trabalhos do AWS Batch	Os trabalhos do AWS Batch são enviados para uma fila de trabalhos onde permanecem até que possam ser agendados para execução em um ambiente de computação.
Amazon Simple Queue Service e instâncias spot do Amazon EC2	Emparelhar o Amazon SQS e instâncias spot para criar uma arquitetura eficiente e tolerante a falhas.

- Para solicitações ou trabalhos que podem ser processados a qualquer momento, use mecanismos de agendamento para processar trabalhos em lote para maior eficiência. Aqui estão alguns exemplos de mecanismos de agendamento no AWS:

Mecanismos de agendamento	Descrição
Agendador do Amazon EventBridge	Um recurso do Amazon EventBridge que permite criar, executar e gerenciar tarefas agendadas em grande escala.
Agendamento do AWS Glue baseado em tempo	Defina um agendamento baseada em tempo para seus crawlers e trabalhos no AWS Glue.
Tarefas agendadas do Amazon Elastic Container Service (Amazon ECS)	O Amazon ECS oferece suporte à criação de tarefas agendadas. As tarefas programadas usam as regras do Amazon EventBridge para executar tarefas em uma programação ou em uma resposta a um evento do EventBridge.

Mecanismos de agendamento	Descrição
Agendador de instâncias	Configure agendamentos de início e término para suas instâncias do Amazon EC2 e do Amazon Relational Database Service.

- Se você usa mecanismos de pesquisa e webhooks em sua arquitetura, substitua-os por eventos. Use [arquiteturas orientadas por eventos](#) para criar workloads altamente eficientes.
- Aproveite a [tecnologia sem servidor na AWS](#) para eliminar a infraestrutura provisionada em excesso.
- Dimensione corretamente componentes individuais da sua arquitetura para evitar recursos ociosos aguardando entrada.
 - As [Recomendações de dimensionamento correto em no AWS Cost Explorer](#) ou o [AWS Compute Optimizer](#) podem ser usados para identificar oportunidades de dimensionamento correto.
 - Para obter mais detalhes, consulte [Dimensionamento correto: provisionamento de instâncias para corresponder a workloads](#).

Recursos

Documentos relacionados:

- [O que é o Amazon Simple Queue Service?](#)
- [O que é o Amazon MQ?](#)
- [Ajuste de escala baseado no Amazon SQS](#)
- [O que é AWS Step Functions?](#)
- [O que é AWS Lambda?](#)
- [Usar o AWS Lambda com o Amazon SQS](#)
- [O que é o Amazon EventBridge?](#)
- [Gerenciar fluxos de trabalho assíncronos com uma API REST](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Navegar na jornada rumo à arquitetura orientada a eventos sem servidor](#)
- [AWS re:Invent 2023: Usar a tecnologia sem servidor para arquitetura orientada a eventos e design orientado por domínio](#)

- [AWS re:Invent 2023: Padrões avançados orientados a eventos com o Amazon EventBridge](#)
- [AWS re:Invent 2023: Arquitetura sustentável: passado, presente e futuro](#)
- [Padrões de mensagens assíncronas | Eventos da AWS](#)

Exemplos relacionados:

- [Arquitetura orientada a eventos com processadores AWS Graviton e instâncias spot do Amazon EC2](#)

SUS03-BP02 Remover ou refatorar componentes da workload com pouco ou nenhum uso

Remova os componentes que não são mais utilizados nem necessários e refatore os componentes pouco usados para minimizar o desperdício em sua workload.

Práticas comuns que devem ser evitadas:

- Você não verifica regularmente o nível de utilização de componentes individuais da sua workload.
- Você não verifica as recomendações de ferramentas de dimensionamento correta da AWS, como o [AWS Compute Optimizer](#).

Benefícios de implementar esta prática recomendada: a remoção de componentes não utilizados minimiza o desperdício e melhora a eficiência geral da sua workload na nuvem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Componentes não utilizados ou subutilizados em uma workload na nuvem consomem recursos desnecessários de computação, armazenamento ou rede. Remova ou refatore esses componentes para reduzir diretamente o desperdício e melhorar a eficiência geral de uma workload na nuvem. Esse é um processo de melhoria iterativo que pode ser iniciado por alterações na demanda ou pelo lançamento de um novo serviço de nuvem. Por exemplo, uma queda significativa no tempo de execução da função do [AWS Lambda](#) pode indicar a necessidade de reduzir o tamanho da memória. Além disso, à medida que a AWS lança novos serviços e recursos, a arquitetura e os serviços ideais para sua workload podem mudar.

Monitore continuamente a atividade da workload e procure oportunidades para melhorar o nível de utilização de componentes individuais. Com a remoção de componentes ociosos e a execução de atividades de dimensionamento correto, você atende aos seus requisitos empresariais com menos recursos de nuvem.

Etapas de implementação

- Faça um inventário dos recursos da AWS: crie um inventário dos recursos da AWS. Na AWS, é possível ativar o [Explorador de recursos da AWS](#) para explorar e organizar seus recursos da AWS. Para obter mais detalhes, consulte [AWS re:Invent 2022: Como gerenciar recursos e aplicações em grande escala AWS](#).
- Monitore a utilização: monitore e capture as métricas de utilização de componentes críticos da workload (como utilização de CPU, utilização de memória ou throughput de rede nas [métricas do Amazon CloudWatch](#)).
- Identifique componentes não utilizados: identifique componentes não utilizados ou subutilizados na arquitetura.
 - Para workloads estáveis, verifique as ferramentas de dimensionamento correto da AWS, como o [AWS Compute Optimizer](#), em intervalos regulares para identificar componentes ociosos, não utilizados ou subutilizados.
 - Para workloads efêmeras, avalie as métricas de utilização para identificar componentes ociosos, não usados ou subutilizados.
- Remova componentes não utilizados: retire componentes e ativos associados (como imagens do Amazon ECR) que não são mais necessários.
 - [Limpeza automatizada de imagens não utilizadas no Amazon ECR](#)
 - [Excluir volumes não utilizados do Amazon Elastic Block Store \(Amazon EBS\) usando AWS Config e AWS Systems Manager](#)
- Refatore os componentes subutilizados: refatore ou consolide os componentes subutilizados com outros recursos para melhorar a eficiência da utilização. Por exemplo, é possível provisionar vários bancos de dados pequenos em uma única instância de banco de dados do [Amazon RDS](#) em vez de executar bancos de dados em instâncias individuais subutilizadas.
- Avalie as melhorias: entenda os [recursos provisionados pela workload para concluir uma unidade de trabalho](#). Use essas informações para avaliar as melhorias obtidas com a remoção ou refatoração de componentes.
 - [Measure and track cloud efficiency with sustainability proxy metrics, Part I: What are proxy metrics?](#)

- [Measure and track cloud efficiency with sustainability proxy metrics, Part II: Establish a metrics pipeline](#)

Recursos

Documentos relacionados:

- [AWS Trusted Advisor](#)
- [O que é o Amazon CloudWatch?](#)
- [Dimensionamento correto: provisionamento de instâncias para corresponder às workloads](#)
- [Como otimizar seus custos com as recomendações de redimensionamento](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Capacidade, disponibilidade, eficiência de custos: escolha três](#)

SUS03-BP03 Otimizar áreas de código que consomem mais tempo ou recursos

Otimize o código que é executado em diferentes componentes de sua arquitetura para minimizar o uso de recursos e, ao mesmo tempo, maximizar a performance.

Práticas comuns que devem ser evitadas:

- Você ignora a otimização de seu código para uso de recursos.
- Normalmente, você responde a problemas de performance aumentando os recursos.
- Seu processo de revisão e desenvolvimento de código não monitora alterações na performance.

Benefícios de implementar esta prática recomendada: o uso eficiente de código minimiza o uso de recursos e melhora a performance.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

É essencial examinar toda área funcional, incluindo o código referente a uma aplicação projetada para a nuvem, para otimizar o uso de recursos e a performance. Monitore continuamente a

performance da workload em ambientes de compilação e na produção e identifique oportunidades para melhorar os trechos cujo uso de recursos é particularmente alto. Adote um processo de revisão regular para identificar erros ou antipadrões dentro do código que usa os recursos ineficazmente. Utilize algoritmos simples e eficientes que produzem os mesmos resultados para seu caso de uso.

Etapas de implementação

- Use uma linguagem de programação eficiente: use um sistema operacional e uma linguagem de programação eficientes para a workload. Para obter detalhes sobre linguagens de programação com eficiência energética (incluindo Rust), consulte [Sustentabilidade com o Rust](#).
- Use um companheiro de codificação de IA: considere usar um companheiro de codificação de IA, como o [Amazon Q Developer](#), para escrever código com eficiência.
- Automatize as revisões de código: ao desenvolver suas workloads, adote um processo de revisão de código automatizado para melhorar a qualidade e identificar erros e práticas não recomendadas.
 - [Automatize as revisões de código com o Amazon CodeGuru Reviewer](#)
 - [Detectar bugs de simultaneidade com o Amazon CodeGuru](#)
 - [Melhorar a qualidade do código para aplicações Python com o Amazon CodeGuru](#)
- Use um criador de perfil de código: use um criador de perfil de código para identificar as áreas de código que gastam mais tempo ou usam mais recursos e as defina como alvos de otimização.
 - [Reduzir a pegada de carbono de sua organização com o Amazon CodeGuru Profiler](#)
 - [Conceitos básicos do uso de memória em sua aplicação Java com o Amazon CodeGuru Profiler](#)
 - [Melhorar a experiência do cliente e reduzir os custos com o Amazon CodeGuru Profiler](#)
- Monitore e otimize: use recursos de monitoramento contínuo para identificar componentes com altos requisitos de recursos ou configuração abaixo do ideal.
 - Substitua os algoritmos com uso intenso de computação por uma versão mais simples e mais eficiente que produza o mesmo resultado.
 - Remova códigos desnecessários, como classificações e formatações.
- Use refatoração ou transformação de código: explore a possibilidade da [transformação de código do Amazon Q](#) para manutenção e atualizações de aplicações.
 - [Atualizar as versões de idioma com o Amazon Q Code Transformation](#)
 - [AWS re:Invent 2023: Automatizar as atualizações e a manutenção de aplicações usando o Amazon Q Code Transformation](#)

Recursos

Documentos relacionados:

- [O que é o Amazon CodeGuru Profiler?](#)
- [Instâncias FPGA](#)
- [Os AWS SDKs em Ferramentas para desenvolver na AWS](#)

Vídeos relacionados:

- [Melhorar a eficiência do código usando o Amazon CodeGuru Profiler](#)
- [Automatizar revisões de código e recomendações de performance de aplicações com o Amazon CodeGuru](#)

SUS03-BP04 Otimizar o impacto sobre dispositivos e equipamentos

Conheça os dispositivos e equipamentos usados em sua arquitetura e use estratégias para reduzir o respectivo uso. Isso pode minimizar o impacto ambiental de modo geral de sua workload na nuvem.

Práticas comuns que devem ser evitadas:

- Você ignora o impacto ambiental dos dispositivos usados por seus clientes.
- Você gerencia e atualiza manualmente os recursos usados pelos clientes.

Benefícios de implementar esta prática recomendada: a implementação de padrões e recursos de software otimizados para o dispositivo do cliente pode reduzir o impacto ambiental geral da workload na nuvem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Implementar padrões e recursos de software que são otimizados para os dispositivos do clientes pode reduzir o impacto ambiental de variadas maneiras:

- Implementar novos recursos que são compatíveis com versões anteriores pode reduzir o número de substituições de hardware.

- Otimizar uma aplicação para ser executada com eficiência nos dispositivos pode ajudar a reduzir o consumo de energia e a estender a duração da bateria (se eles forem alimentados por bateria).
- Otimizar uma aplicação para dispositivos também pode reduzir a transferência de dados ao longo da rede.

Conheça os dispositivos e equipamentos usados em sua arquitetura, o ciclo de vida esperado e o impacto da substituição desses componentes. Implemente padrões e recursos de software que possam ajudar a minimizar o consumo de energia do dispositivo, bem como a necessidade de os clientes substituírem o dispositivo e também atualizá-lo manualmente.

Etapas de implementação

- Conduza um inventário: faça um inventário dos dispositivos usados em sua arquitetura. Os dispositivos podem ser celular, tablet, dispositivos IoT, lâmpada inteligente ou até dispositivos inteligentes em uma fábrica.
- Use dispositivos energeticamente eficientes: considere usar dispositivos com baixo consumo de energia em sua arquitetura. Use as configurações de gerenciamento de energia nos dispositivos para que entrem no modo de baixo consumo quando não estiverem em uso.
- Execute aplicações eficientes: otimize a aplicação em execução nos dispositivos:
 - Use estratégias como execução de tarefas em segundo plano para reduzir o consumo de energia.
 - Considere a largura de banda da rede e a latência ao criar cargas úteis, e implemente recursos que ajudem suas aplicações a funcionar bem em links de baixa largura de banda e alta latência.
 - Converta cargas úteis e arquivos nos formatos otimizados exigidos pelos dispositivos. Por exemplo, é possível usar o [Amazon Elastic Transcoder](#) ou o [AWS Elemental MediaConvert](#) para converter arquivos de mídia digital grande e de alta qualidade em formatos que podem ser reproduzidos em dispositivos móveis, tablets, navegadores da web e televisões conectadas.
 - Realize atividades com computação intensa no lado do servidor (como renderização de imagens) ou use a transmissão de aplicações para melhorar a experiência do usuário em dispositivos mais antigos.
 - Faça a segmentação e a paginação dos dados de saída, especialmente para sessões interativas, a fim de gerenciar cargas úteis e limitar os requisitos de armazenamento local.
- Envolver fornecedores: trabalhe com fornecedores de dispositivos que usam materiais sustentáveis e fornecem transparência em suas cadeias de suprimentos e certificações ambientais.

- Use atualizações sem fios: use um mecanismo sem fio automatizado para implantar atualizações em um ou mais dispositivos.
 - É possível usar um [pipeline de CI/CD](#) para atualizar aplicações móveis.
 - O [AWS IoT Device Management](#) também pode ser usado para gerenciar remotamente dispositivos conectados em grande escala.
- Use parques de dispositivos gerenciados: para testar novos recursos e atualizações, use parques de dispositivos gerenciados com conjuntos representativos de hardware e itere o desenvolvimento para maximizar os dispositivos compatíveis. Para obter mais detalhes, consulte [SUS06-BP05 Usar parques de dispositivos gerenciados para testes](#).
- Continue monitorando e melhorando: acompanhe o uso de energia dos dispositivos para identificar áreas de melhoria. Use novas tecnologias ou práticas recomendadas para melhorar os impactos ambientais desses dispositivos.

Recursos

Documentos relacionados:

- [O que é AWS Device Farm?](#)
- [Documentação do Amazon AppStream 2.0](#)
- [NICE DCV](#)
- [Tutorial OTA para atualização de firmware em dispositivos que executam o FreeRTOS](#)
- [Otimizando seus dispositivos de IoT para sustentabilidade ambiental](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Melhorar a qualidade das suas aplicações móveis e Web com o AWS Device Farm](#)

SUS03-BP05 Usar padrões e arquiteturas de software que atendam melhor aos padrões de armazenamento e acesso a dados

Entenda como os dados são usados com sua workload, consumidos pelos usuários, transferidos e armazenados. Use os padrões e arquiteturas de software ideais para acesso e armazenamento de dados a fim de minimizar os recursos de computação, rede e armazenamento necessários para atender à workload.

Práticas comuns que devem ser evitadas:

- Você pressupõe que todas as workloads tenha, padrões de acesso e armazenamento de dados semelhantes.
- Você usa apenas um nível de armazenamento, supondo que todas as workloads se encaixem nesse nível.
- Você pressupõe que os padrões de acesso aos dados permanecerão consistentes ao longo do tempo.
- Na eventualidade de uma alta expansão no acesso aos dados, sua arquitetura é capaz de comportá-la, mas isso faz com que os recursos fiquem ociosos na maior parte do tempo.

Benefícios de implementar esta prática recomendada: selecionar e otimizar sua arquitetura com base nos padrões de acesso e armazenamento de dados ajudará a diminuir a complexidade do desenvolvimento e aumentar a utilização geral. Compreender quando usar tabelas globais, provisionamento de dados e armazenamento em cache ajuda a reduzir a despesas operacionais indiretas e a escalar com base nas necessidades da workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Para melhorar a sustentabilidade da workload a longo prazo, use padrões de arquitetura que ofereçam suporte às características de acesso e armazenamento de dados para a workload. Esses padrões ajudam você a recuperar e processar dados com eficiência. Por exemplo, você pode usar uma [arquitetura de dados moderna na AWS](#) com serviços personalizados e otimizados para seus casos de uso de analytics exclusivos. Esses padrões de arquitetura possibilitam um processamento de dados eficiente e reduzem o uso de recursos.

Etapas de implementação

- Entenda as características dos dados: analise as características dos dados e os padrões de acesso para identificar a configuração correta para os recursos de nuvem. Principais características a serem consideradas:
 - Tipos de dados: estruturados, semiestruturados e não estruturados
 - Crescimento de dados: limitado, ilimitado
 - Durabilidade dos dados: persistentes, efêmeros, transitórios
 - Padrões de acesso: leituras ou gravações, frequência, com picos ou consistente

- Use padrões de arquitetura otimizados: use padrões de arquitetura que comportem melhor os padrões de armazenamento e acesso aos dados.
 - [Padrões para permitir a persistência de dados](#)
 - [Vamos arquitetar! Arquiteturas de dados modernas](#)
 - [Bancos de dados na AWS: a ferramenta certa para o trabalho certo](#)
- Use serviços específicos: use tecnologias que sejam adequadas à finalidade.
 - Use tecnologias que funcionam nativamente com dados compactados.
 - [Formatos de arquivos compactados compatíveis com o Athena](#)
 - [Opções de formato para entradas e saídas de ETL no AWS Glue](#)
 - [Carregar arquivos de dados compactados do Amazon S3 com o Amazon Redshift](#)
 - Use [serviços de análise](#) específicos para processamento de dados em sua arquitetura. Para obter detalhes sobre serviços de análise da AWS com propósitos específicos, consulte [AWS re:Invent 2022: construir arquiteturas de dados modernas na AWS](#).
 - Use o mecanismo de banco de dados que melhor comporta seu padrão de consulta dominante. Gerencie os índices de bancos de dados para fazer consultas eficientes. Para obter mais detalhes, consulte [Bancos de dados da AWS](#) e [AWS re:Invent 2022: Modernizar aplicações com bancos de dados com propósitos específicos](#).
- Minimize a transferência de dados: escolha protocolos de rede que reduzam a quantidade de capacidade de rede consumida na arquitetura.

Recursos

Documentos relacionados:

- [COPY de formatos de dados colunares com o Amazon Redshift](#)
- [Converter formato de registros recebidos no Firehose](#)
- [Melhorar a performance de consultas no Amazon Athena fazendo a conversão para formatos colunares](#)
- [Monitorar a workload de banco de dados com o Performance Insights no Amazon Aurora](#)
- [Monitorar a workload de banco de dados com o Performance Insights no Amazon RDS](#)
- [Classe de armazenamento do Amazon S3 Intelligent-Tiering](#)
- [Criar uma loja de eventos CQRS com o Amazon DynamoDB](#)

Vídeos relacionados:

- [AWS re:Invent 2022: Construir arquiteturas de data mesh na AWS](#)
- [AWS re:Invent 2023: Mergulho profundo no Amazon Aurora e suas inovações](#)
- [AWS re:Invent 2023: Melhorar a eficiência do Amazon EBS e ser mais econômico](#)
- [AWS re:Invent 2023: Otimizar o preço e a performance do armazenamento com o Amazon S3](#)
- [AWS re:Invent 2023: Criar e otimizar data lakes no Amazon S3](#)
- [AWS re:Invent 2023: Padrões avançados orientados a eventos com o Amazon EventBridge](#)

Exemplos relacionados:

- [Workshop de bancos de dados com propósito específico na AWS](#)
- [Dia de imersão na arquitetura de dados moderna na AWS](#)
- [Criar um data mesh na AWS](#)

Gerenciamento de dados

Implemente práticas de gerenciamento de dados para reduzir o armazenamento provisionado necessário para comportar a workload e os recursos exigidos para usá-la. Entenda seus dados e use as tecnologias e as configurações de armazenamento que melhor promovam o valor empresarial dos dados e a forma como eles são usados. Gerencie o ciclo de vida dos dados e opte por um armazenamento mais eficiente e com menor performance quando os requisitos diminuírem, excluindo os dados que não são mais necessários.

Práticas recomendadas

- [SUS04-BP01 Implementar uma política de classificação de dados](#)
- [SUS04-BP02 Usar tecnologias compatíveis com seus padrões de acesso e de armazenamento de dados](#)
- [SUS04-BP03 Usar políticas para gerenciar o ciclo de vida de seus conjuntos de dados](#)
- [SUS04-BP04 Usar elasticidade e automação para expandir o armazenamento em bloco ou o sistema de arquivos](#)
- [SUS04-BP05 Remover dados desnecessários ou redundantes](#)
- [SUS04-BP06 Usar armazenamento ou sistemas de arquivos compartilhados para acessar dados comuns](#)

- [SUS04-BP07 Minimizar a movimentação de dados entre redes](#)
- [SUS04-BP08 Fazer backup de dados somente quando for difícil recriá-los](#)

SUS04-BP01 Implementar uma política de classificação de dados

Classifique os dados para entender sua importância para os resultados comerciais e selecione o nível de armazenamento eficiente em termos de energia para armazenar os dados.

Práticas comuns que devem ser evitadas:

- Você não identifica ativos de dados com características semelhantes (como sensibilidade, importância empresarial ou requisitos regulatórios) que estão sendo processados ou armazenados.
- Você não implementou um catálogo de dados para criar um inventário de seus ativos de dados.

Benefícios de implementar esta prática recomendada: a implementação de uma política de classificação de dados permite determinar o nível de armazenamento de dados com maior eficiência energética.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A classificação de dados envolve a identificação dos tipos de dados que estão sendo processados e armazenados em um sistema de informação pertencente a uma organização ou operado por ela. Ela também envolve a decisão em relação à importância dos dados e ao impacto provável do comprometimento dos dados, perda ou uso incorreto.

Implemente a política de classificação de dados trabalhando de forma reversa a partir do uso contextual dos dados e criando um esquema de categorização que leve em conta a importância de determinado conjunto de dados para as operações de uma organização.

Etapas de implementação

- Faça o inventário dos dados: faça um inventário dos vários tipos de dados que existem para sua workload.
- Agrupe os dados: determine a importância, a confidencialidade, a integridade e a disponibilidade dos dados com base no risco para a organização. Use esses requisitos para agrupar dados em um

dos níveis de classificação de dados adotados. Como exemplo, consulte [Quatro etapas simples para classificar seus dados e proteger sua startup](#).

- Defina níveis e políticas de classificação de dados: para cada grupo de dados, defina o nível de classificação de dados (por exemplo, público ou confidencial) e as políticas de tratamento. Marque os dados adequadamente. Para obter mais detalhes sobre as categorias de classificação de dados, consulte o whitepaper [Classificação de dados](#).
- Revise periodicamente: revise e audite periodicamente seu ambiente em busca de dados não marcados e não classificados. Utilize automação para identificar esses dados e classificá-los e marcá-los adequadamente. Como exemplo, consulte [Catálogo de dados e crawlers na AWS Glue](#).
- Estabeleça um catálogo de dados: estabeleça um catálogo de dados que forneça recursos de auditoria e governança.
- Documentação: documente as políticas de classificação de dados e os procedimentos de tratamento para cada classe de dados.

Recursos

Documentos relacionados:

- [Utilizar a Nuvem AWS para apoiar a classificação de dados](#)
- [Políticas de tags do AWS Organizations](#)

Vídeos relacionados:

- [AWS re:Invent 2022: Capacitar a agilidade com a governança de dados na AWS](#)
- [AWS re:Invent 2023: Proteção e resiliência de dados com armazenamento na AWS](#)

SUS04-BP02 Usar tecnologias compatíveis com seus padrões de acesso e de armazenamento de dados

Use tecnologias de armazenamento mais adequadas à maneira como seus dados são acessados e armazenados a fim de reduzir os recursos provisionados e, ao mesmo tempo, comportar sua workload.

Práticas comuns que devem ser evitadas:

- Você pressupõe que todas as workloads tenham, padrões de acesso e armazenamento de dados semelhantes.
- Você usa apenas um nível de armazenamento, supondo que todas as workloads se encaixem nesse nível.
- Você pressupõe que os padrões de acesso aos dados permanecerão consistentes ao longo do tempo.

Benefícios de implementar esta prática recomendada: selecionar e otimizar suas tecnologias de armazenamento com base em padrões de armazenamento e acesso aos dados ajudará a reduzir os recursos de nuvem necessários a fim de atender às suas necessidades empresariais e melhorar a eficiência geral da workload de nuvem.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Selecione a solução de armazenamento mais alinhada a seus padrões de acesso ou considere a possibilidade de alterar seus padrões de acesso para alinhamento com a solução de armazenamento a fim de maximizar a eficiência da performance.

Etapas de implementação

- **Avalie dados e acesse características:** avalie suas características de dados e padrão de acesso a fim de reunir as principais características de suas necessidades de armazenamento. Principais características a serem consideradas:
 - Tipos de dados: estruturados, semiestruturados e não estruturados
 - Crescimento de dados: limitado, ilimitado
 - Durabilidade dos dados: persistentes, efêmeros, transitórios
 - Padrões de acesso: leituras ou gravações, frequência, com picos ou consistente
- **Escolha a tecnologia de armazenamento correta:** migre os dados para a tecnologia de armazenamento apropriada que seja compatível com suas características de dados e padrão de acesso. Veja alguns exemplos de tecnologias de armazenamento da AWS e suas principais características:

Tipo	Tecnologia	Características principais
Armazenamento de objetos	Amazon S3	Um serviço de armazenamento de objetos com escalabilidade ilimitada, alta disponibilidade e várias opções de acessibilidade. A transferência e o acesso de objetos dentro e fora do Amazon S3 podem usar um serviço, como o Transfer Acceleration ou Access Points , para oferecer suporte à sua localização, necessidades de segurança e padrões de acesso.
Armazenamento de arquivamento	Amazon S3 Glacier	Classe de armazenamento do Amazon S3 desenvolvida para arquivamento de dados.
Sistema de arquivos compartilhado	Amazon Elastic File System (Amazon EFS)	Sistema de arquivos montável que pode ser acessado por vários tipos de soluções de computação. O Amazon EFS aumenta e reduz automaticamente o armazenamento e possui performance otimizada para oferecer baixas latências consistentes.

Tipo	Tecnologia	Características principais
Sistema de arquivos compartilhado	Amazon FSx	Baseia-se nas soluções de computação mais recentes da AWS para oferecer compatibilidade com quatro sistemas de arquivos usados com frequência: NetApp ONTAP, OpenZFS, Windows File Server e Lustre. A latência, o throughput e as IOPS do Amazon FSx variam de acordo com o sistema de arquivos e devem ser consideradas ao selecionar o sistema de arquivos certo para as necessidades da sua workload.
Armazenamento em bloco	Amazon Elastic Block Store (Amazon EBS)	Serviço de armazenamento em blocos fácil de usar, escalável e de alta performance projetado para o Amazon Elastic Compute Cloud (Amazon EC2). O Amazon EBS inclui armazenamento baseado em SSD para workloads transacionais de alto throughput e em HDD para workloads trabalho de alto throughput.

Tipo	Tecnologia	Características principais
Banco de dados relacional	Amazon Aurora , Amazon RDS , Amazon Redshift	Projetados para oferecer compatibilidade com transações ACID (atomicidade, consistência, isolamento, durabilidade) e manter a integridade referencial e uma forte consistência de dados. Muitas aplicações tradicionais, sistemas de planejamento de recursos empresariais (ERP), de gerenciamento de relacionamentos com o cliente (CRM) e de comércio eletrônico usam bancos de dados relacionais para armazenar seus dados.
Banco de dados de chave-valor	Amazon DynamoDB	Otimizados para padrões de acesso comuns, normalmente visando armazenar e recuperar grandes volumes de dados. Aplicações web de alto tráfego, sistemas de comércio eletrônico e aplicações de jogos são os casos de uso habituais para bancos de dados de chave-valor.

- Automatize a alocação de armazenamento: para sistemas de armazenamento que têm tamanho fixo, como o Amazon EBS ou o Amazon FSx, monitore o espaço de armazenamento disponível e automatize a alocação de armazenamento ao atingir um limite. É possível utilizar o Amazon CloudWatch para coletar e analisar diferentes métricas para o [Amazon EBS](#) e o [Amazon FSx](#).
- Escolha a classe de armazenamento correta: escolha a classe de armazenamento apropriada para seus dados.

- As classes de armazenamento do Amazon S3 podem ser configuradas em nível de objeto. Um único bucket pode conter objetos armazenados em todas as classes de armazenamento.
- É possível usar [políticas de ciclo de vida do Amazon S3](#) para realizar a transição automática de objetos entre classes de armazenamento ou remover dados sem nenhuma alteração na aplicação. Em geral, é necessário fazer uma compensação entre a eficiência dos recursos, a latência de acesso e a confiabilidade ao considerar esses mecanismos de armazenamento.

Recursos

Documentos relacionados:

- [Tipos de volume do Amazon EBS](#)
- [Armazenamento de instâncias do Amazon EC2](#)
- [Amazon S3 Intelligent-Tiering](#)
- [Características de E/S do Amazon EBS](#)
- [Usar classes de armazenamento do Amazon S3](#)
- [O que é o Amazon S3 Glacier?](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Melhorar a eficiência do Amazon EBS e ser mais econômico](#)
- [AWS re:Invent 2023: Otimizar o preço e a performance do armazenamento com o Amazon S3](#)
- [AWS re:Invent 2023: Criar e otimizar data lakes no Amazon S3](#)
- [AWS re:Invent 2022: Construir arquiteturas de dados modernos na AWS](#)
- [AWS re:Invent 2022: Modernizar aplicações com bancos de dados com propósito específico](#)
- [AWS re:Invent 2022: Construir arquiteturas de data mesh na AWS](#)
- [AWS re:Invent 2023: Mergulho profundo no Amazon Aurora e suas inovações](#)
- [AWS re:Invent 2023: Modelagem de dados com o Amazon DynamoDB](#)

Exemplos relacionados:

- [Exemplos do Amazon S3](#)
- [Workshop de bancos de dados com propósito específico na AWS](#)

- [Bancos de dados para desenvolvedores](#)
- [Dia de imersão na arquitetura de dados moderna na AWS](#)
- [Criar um data mesh na AWS](#)

SUS04-BP03 Usar políticas para gerenciar o ciclo de vida de seus conjuntos de dados

Gerencie o ciclo de vida de todos os seus dados e aplique a exclusão automaticamente para minimizar o armazenamento total necessário para sua workload.

Práticas comuns que devem ser evitadas:

- Você exclui dados manualmente.
- Você não exclui nenhum de seus dados de workload.
- Você não move os dados para níveis de armazenamento mais eficientes em termos de energia com base em seus requisitos de retenção e acesso.

Benefícios de estabelecer essa prática recomendada: o uso de políticas de ciclo de vida de dados garante acesso e retenção eficientes de dados em uma workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Os conjuntos de dados geralmente têm diferentes requisitos de retenção e acesso durante seu ciclo de vida. Por exemplo, sua aplicação pode precisar de acesso frequente a alguns conjuntos de dados por um período limitado. Depois disso, esses conjuntos de dados são acessados com pouca frequência. Para melhorar a eficiência da computação e do armazenamento de dados ao longo do tempo, implemente políticas de ciclo de vida, que são regras que definem como os dados são tratados ao longo do tempo.

Com as regras de configuração do ciclo de vida, é possível orientar o serviço de armazenamento específico a fazer a transição de um conjunto de dados para níveis de armazenamento mais eficientes em termos de energia, arquivá-lo ou excluí-lo. Essa prática minimiza o armazenamento e a recuperação ativos de dados, o que leva a um menor consumo de energia. Além disso, práticas como arquivamento ou exclusão de dados obsoletos apoiam a conformidade normativa e a governança de dados.

Etapas de implementação

- Use a classificação de dados: [classifique os conjuntos de dados na workload](#).
- Defina regras de tratamento: defina procedimentos de tratamento para cada classe de dados.
- Habilite a automação: defina políticas automatizadas de ciclo de vida para aplicar regras de ciclo de vida. Aqui estão alguns exemplos de como configurar políticas de ciclo de vida automatizadas para diferentes serviços de armazenamento do AWS:

Serviços de armazenamento	Como definir políticas de ciclo de vida automatizadas
Amazon S3	Você pode usar o Amazon S3 Lifecycle para gerenciar seus objetos durante todo o ciclo de vida de cada um. Se os padrões de acesso forem desconhecidos, variáveis ou imprevisíveis, você poderá usar o Amazon S3 Intelligent-Tiering para monitorar os padrões de acesso e mover automaticamente os objetos que não foram acessados para níveis de acesso de baixo custo. Também é possível usar as métricas da Lente de Armazenamento do Amazon S3 para identificar oportunidades de otimização e lacunas no gerenciamento do ciclo de vida.
Amazon Elastic Block Store	É possível usar o Amazon Data Lifecycle Manager para automatizar a criação, a retenção e a exclusão de snapshots do Amazon EBS e de AMIs apoiadas pelo Amazon EBS.
Amazon Elastic File System	O gerenciamento de ciclo de vida útil do Amazon EFS gerencia automaticamente o armazenamento de arquivos dos seus sistemas de arquivos.

Serviços de armazenamento	Como definir políticas de ciclo de vida automatizadas
Amazon Elastic Container Registry	As políticas de ciclo de vida do Amazon ECR permitem automatizar a limpeza de imagens de contêineres expirando as imagens com base na idade ou no número.
AWS Elemental MediaStore	Você pode usar uma política de ciclo de vida de objeto que determina por quanto tempo os objetos devem ser armazenados no contêiner MediaStore.

- Exclua ativos não utilizados: exclua volumes não utilizados, snapshots e dados que estão fora do período de retenção. Use recursos nativos de serviços, como a [vida útil do Amazon DynamoDB](#) ou a [retenção de logs do Amazon CloudWatch](#), para exclusão.
- Agregue e compacte: agregue e compacte dados quando possível com base nas regras de ciclo de vida.

Recursos

Documentos relacionados:

- [Otimizar regras de ciclo de vida do Amazon S3 com a análise de classes de armazenamento do Amazon S3](#)
- [Avaliar recursos com o Regras do AWS Config](#)

Vídeos relacionados:

- [AWS re:Invent 2021: Práticas recomendadas do ciclo de vida do Amazon S3 para otimizar seus gastos com armazenamento](#)
- [AWS re:Invent 2023: Otimizar o preço e a performance do armazenamento com o Amazon S3](#)
- [Simplificar o ciclo de vida de dados e otimizar os custos de armazenamento com o Amazon S3 Lifecycle](#)
- [Reduzir seus custos de armazenamento com a Lente de Armazenamento do Amazon S3](#)

SUS04-BP04 Usar elasticidade e automação para expandir o armazenamento em bloco ou o sistema de arquivos

Use a elasticidade e automação para expandir o armazenamento em bloco ou o sistema de arquivos à medida que os dados aumentam para minimizar o total de armazenamento provisionado.

Práticas comuns que devem ser evitadas:

- Você adquire um grande armazenamento em bloco ou sistema de arquivos para necessidades futuras.
- Você provisiona em excesso as operações de entrada e saída por segundo (IOPS) de seu sistema de arquivos.
- Você não monitora a utilização de seus volumes de dados.

Benefícios de implementar esta prática recomendada: minimizar o provisionamento excessivo do sistema de armazenamento reduz os recursos ociosos e melhora a eficiência geral de sua workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Crie armazenamento em bloco ou sistemas de arquivos com alocação de tamanho, throughput e latência apropriados à workload. Use a elasticidade e automação para expandir o armazenamento em bloco ou o sistema de arquivos à medida que os dados aumentarem sem precisar provisionar em excesso esses serviços de armazenamento.

Etapas de implementação

- Para sistemas de armazenamento com tamanho fixo, como o [Amazon EBS](#), certifique-se de monitorar a quantidade de armazenamento usada em comparação com o tamanho geral do armazenamento e, se possível, criar automação para aumentar o tamanho do armazenamento ao atingir um limite.
- Use volumes elásticos e serviços gerenciados de dados em bloco para automatizar a alocação de armazenamento adicional à medida que os seus dados persistentes aumentarem. Por exemplo, é possível usar [Volumes Elásticos do Amazon EBS](#) para alterar o tamanho ou o tipo de volume ou ajustar a performance de seus volumes do Amazon EBS.
- Escolha a classe de armazenamento, modo de performance e modo de throughput corretos para seu sistema de arquivos para atender à necessidade de seus negócios, sem a ultrapassar.

- [Performance do Amazon EFS](#)
- [Performance do volume do Amazon EBS em instâncias do Linux](#)
- Defina os níveis pretendidos de utilização para seus volumes de dados e redimensione os volumes fora dos intervalos esperados.
- Dimensione adequadamente volumes somente leitura para acomodar os dados.
- Migre os dados para depósitos de objetos a fim de evitar o provisionamento de capacidade em excesso que ocorre com os tamanhos de volumes fixos no armazenamento em bloco.
- Revise regularmente volumes elásticos e sistemas de arquivos para encerrar volumes ociosos, reduzir recursos com excesso de provisionamento e se ajustar ao tamanho de dados atual.

Recursos

Documentos relacionados:

- [Estender um sistema de arquivos após redimensionar um volume do EBS](#)
- [Modificar um volume do EBS usando Volumes Elásticos do Amazon EBS](#)
- [Documentação do Amazon FSx](#)
- [O que é o Amazon Elastic File System?](#)

Vídeos relacionados:

- [Mergulho profundo nos Volumes Elásticos do Amazon EBS](#)
- [Amazon EBS e estratégias de otimização de snapshots para melhor performance e redução de custos](#)
- [Otimizar o Amazon EFS para custo e performance usando práticas recomendadas](#)

SUS04-BP05 Remover dados desnecessários ou redundantes

Remova dados desnecessários ou redundantes para minimizar os recursos de armazenamento necessários para armazenar seus conjuntos de dados.

Práticas comuns que devem ser evitadas:

- Você duplica dados que podem ser facilmente obtidos ou recriados.

- Você faz backup de todos os dados sem considerar sua criticidade.
- Você apenas exclui dados irregularmente, em eventos operacionais ou não os exclui.
- Você armazena dados de forma redundante, independentemente da durabilidade do serviço de armazenamento.
- Você ativa o versionamento do Amazon S3 sem qualquer justificativa comercial.

Benefícios de implementar esta prática recomendada: a remoção de dados desnecessários reduz o tamanho de armazenamento necessário para sua workload e o impacto ambiental causado por ela.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Ao remover conjuntos de dados desnecessários e redundantes, você pode reduzir o custo de armazenamento e a pegada ambiental. Essa prática também pode tornar a computação mais eficiente, pois os recursos computacionais processam apenas dados importantes em vez de dados desnecessários. Automatize a exclusão de dados desnecessários. Use tecnologias que eliminam dados duplicados em níveis de arquivo e bloco. Use recursos do serviço para replicação de dados nativos e redundância.

Etapas de implementação

- Avalie conjuntos de dados públicos: avalie se você pode evitar o armazenamento de dados usando conjuntos de dados existentes e publicamente disponíveis em [AWS Data Exchange](#) e [Open Data on AWS](#).
- Elimine duplicações de dados: use mecanismos que possam eliminar duplicações de dados aos níveis do bloco e objeto. Aqui estão alguns exemplos de como eliminar duplicações dados na AWS:

Serviços de armazenamento	Mecanismo de eliminação de duplicações
Amazon S3	Use o AWS Lake Formation FindMatches para encontrar registros correspondentes em um conjunto de dados (incluindo aqueles sem identificadores) usando a nova transformada de ML do FindMatches.

Serviços de armazenamento	Mecanismo de eliminação de duplicações
Amazon FSx	Use a eliminação de duplicação de dados no Amazon FSx para Windows.
Snapshots do Amazon Elastic Block Store	Snapshots são backups incrementais, o que significa que somente os blocos no dispositivo que tiverem mudado depois do snapshot mais recente serão salvos.

- Use políticas de ciclo de vida: use políticas de ciclo de vida para automatizar a exclusão de dados não utilizados. Use recursos nativos de serviços, como a [vida útil do Amazon DynamoDB](#), a [funcionalidade Ciclo de Vida do Amazon S3](#) ou a [retenção de logs do Amazon CloudWatch](#), para exclusão.
- Use a virtualização de dados: use os recursos de virtualização de dados na AWS para manter os dados em sua origem e evitar a duplicação de dados.
 - [Virtualização de dados nativos da nuvem na AWS](#)
 - [Otimizar o padrão de dados usando o compartilhamento de dados do Amazon Redshift](#)
- Use backup incremental: use tecnologia de backup capaz de fazer backups incrementais.
- Use a durabilidade nativa: aproveite a durabilidade do [Amazon S3](#) e a [replicação do Amazon EBS](#), em vez de tecnologias autogerenciadas (como uma matriz redundante de discos independentes, ou RAID), para atingir suas metas de durabilidade.
- Use o registro em log de maneira eficiente: centralize o log e rastreie os dados, elimine a duplicação de entradas de log idênticas e estabeleça mecanismos para ajustar a prolixidade quando necessário.
- Use o armazenamento em cache de maneira eficiente: preencha os caches com antecedência somente quando justificável.
- Estabeleça o monitoramento e a automação de cache para redimensionar o cache de forma adequada.
- Remova versões antigas de ativos: remova implantações e ativos desatualizados dos armazenamentos de objetos e caches de borda ao enviar novas versões da workload.

Recursos

Documentos relacionados:

- [Alterar a retenção de dados de log no CloudWatch Logs](#)
- [Eliminação da duplicação de dados no Amazon FSx para Windows File Server](#)
- [Recursos do Amazon FSx para ONTAP, incluindo a eliminação da duplicação de dados](#)
- [Invalidação de arquivos no Amazon CloudFront](#)
- [Usar o AWS Backup para fazer backup e restaurar sistemas de arquivos do Amazon EFS](#)
- [O que é o Amazon CloudWatch Logs?](#)
- [Trabalhar com backups no Amazon RDS](#)
- [Integrar e eliminar duplicações de conjuntos de dados usando o AWS Lake Formation](#)

Vídeos relacionados:

- [Casos de uso de compartilhamento de dados do Amazon Redshift](#)

Exemplos relacionados:

- [Como analiso meus logs de acesso ao servidor do Amazon S3 usando o Amazon Athena?](#)

SUS04-BP06 Usar armazenamento ou sistemas de arquivos compartilhados para acessar dados comuns

Adote armazenamento ou sistemas de arquivos compartilhados para evitar a duplicação de dados e viabilizar uma infraestrutura mais eficiente para a workload.

Práticas comuns que devem ser evitadas:

- Você provisiona armazenamento para cada cliente específico.
- Você não desanexa o volume de dados dos clientes inativos.
- Você não fornece acesso a armazenamento em plataformas e sistemas.

Benefícios de estabelecer essa prática recomendada: o uso de sistemas de arquivos ou armazenamento compartilhados permite compartilhar dados com um ou mais consumidores sem precisar copiá-los. Isso ajuda a reduzir os recursos de armazenamento necessários à workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Se você tiver vários usuários ou aplicações que acessam os mesmos conjuntos de dados, o uso da tecnologia de armazenamento compartilhado é essencial para viabilizar uma infraestrutura eficiente para a workload. A tecnologia de armazenamento compartilhado oferece um local central para armazenar e gerenciar conjuntos de dados e evitar a duplicação de dados. Ela também impõe a consistência dos dados em sistemas diferentes. Além disso, a tecnologia de armazenamento compartilhado permite o uso mais eficiente da potência computacional, visto que vários recursos podem acessar e processar os dados ao mesmo tempo em paralelo.

Busque dados dos serviços de armazenamento compartilhado somente quando necessário e desanexe os volumes não usados para liberar recursos.

Etapas de implementação

- Use o armazenamento compartilhado: migre os dados para o armazenamento compartilhado quando eles tiverem vários consumidores. Veja aqui alguns exemplos de tecnologia de armazenamento compartilhado na AWS:

Opções de armazenamento	Quando usar
Amazon EBS Multi-Attach	O Amazon EBS Multi-Attach permite que você anexe um único volume de SSD de IOPS provisionadas (io1 ou io2) a várias instâncias que estão na mesma zona de disponibilidade.
Amazon EFS	Consulte Quando escolher o Amazon EFS .
Amazon FSx	Consulte Escolher um sistema de arquivos Amazon FSx .
Amazon S3	As aplicações que não exigem uma estrutura de sistema de arquivos e são projetadas para funcionar com armazenamento de objetos podem usar o Amazon S3 como solução de armazenamento de objetos altamente escalável, durável e de baixo custo.

- Obtenha os dados conforme necessário: copie os dados ou busque-os de sistemas de arquivos compartilhados somente quando necessário. Como exemplo, você pode criar um [sistema de arquivos Amazon FSx para Lustre apoiado pelo Amazon S3](#) e carregar somente o subconjunto de dados necessários para o processamento de trabalhos no Amazon FSx.
- Exclua dados desnecessários: exclua dados conforme apropriado para os seus padrões de uso, conforme descrito em [SUS04-BP03 Usar políticas para gerenciar o ciclo de vida de seus conjuntos de dados](#).
- Desvincule clientes inativos: desvincule volumes de clientes que não os estão usando ativamente.

Recursos

Documentos relacionados:

- [Vincular o sistema de arquivos a um bucket do Amazon S3](#)
- [Usar o Amazon EFS para AWS Lambda em aplicações com tecnologia sem servidor](#)
- [O Amazon EFS Intelligent-Tiering otimiza os custos das workloads com padrões de acesso variáveis](#)
- [Como usar o Amazon FSx com seu repositório de dados on-premises](#)

Vídeos relacionados:

- [Otimização do custo de armazenamento com o Amazon EFS](#)
- [AWS re:Invent 2023: Novidades do armazenamento de arquivos na AWS](#)
- [AWS re:Invent 2023: Armazenamento de arquivos para criadores e cientistas de dados no Amazon Elastic File System](#)

SUS04-BP07 Minimizar a movimentação de dados entre redes

Use armazenamento de objetos ou sistemas de arquivos compartilhados para acessar dados comuns e minimizar os recursos totais de rede exigidos para comportar a movimentação de dados da workload.

Práticas comuns que devem ser evitadas:

- Você armazena todos os dados na mesma Região da AWS independentemente de onde os usuários dos dados estão.

- Você não otimiza o tamanho e o formato dos dados antes de movimentá-los na rede.

Benefícios de implementar esta prática recomendada: otimizar a movimentação de dados na rede reduz os recursos totais de rede necessários à workload e reduz o respectivo impacto ambiental.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

A movimentação de dados em sua organização exige recursos de computação, rede e armazenamento. Use técnicas para minimizar a movimentação de dados e melhorar a eficiência geral da workload.

Etapas de implementação

- Use a proximidade: considere a proximidade dos dados ou dos usuários como um fator decisivo ao [selecionar uma região para a workload](#).
- Particione serviços: particione serviços consumidos regionalmente para que os dados específicos da região sejam armazenados na região em que eles são consumidos.
- Use formatos de arquivo eficientes: use formatos de arquivo eficientes (como Parquet ou ORC) e compacte os dados antes movimentá-los na rede.
- Minimize a movimentação de dados: não movimente dados não usados. Alguns exemplos que podem ajudar você a evitar a movimentação de dados não utilizados:
 - Reduza as respostas de API apenas aos dados relevantes.
 - Agregue os dados onde não houver necessidade de informações detalhadas (em nível de registro).
 - Consulte [Laboratório do Well-Architected: Otimizar padrão de dados usando o compartilhamento de dados do Amazon Redshift](#).
 - Considere [Compartilhamento de dados entre contas no AWS Lake Formation](#).
- Use serviços de borda: use serviços que possam ajudar você a executar código mais perto dos usuários da workload.

Serviço	Quando usar
Lambda@Edge	Use para operações com uso intenso de computação que são executadas quando objetos não estão no cache.
CloudFront Functions	Use para casos de uso simples como solicitações HTTP(s)/manipulações de resposta que podem ser iniciadas por funções de curta duração.
AWS IoT Greengrass	Execute computação local, mensagens e armazenamento de dados em cache para dispositivos conectados.

Recursos

Documentos relacionados:

- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte III: Redes](#)
- [Infraestrutura global da AWS](#)
- [Principais recursos do Amazon CloudFront incluindo a rede de borda global do CloudFront](#)
- [Compactação de solicitações HTTP no Amazon OpenSearch Service](#)
- [Intermediar a compactação de dados com o Amazon EMR](#)
- [Carregar arquivos de dados compactados do Amazon S3 no Amazon Redshift](#)
- [Envio de arquivos compactados com o Amazon CloudFront](#)

Vídeos relacionados:

- [Desmistificar a transferência de dados na AWS](#)

SUS04-BP08 Fazer backup de dados somente quando for difícil recriá-los

Evite fazer backup de dados que não têm valor empresarial para minimizar os requisitos de recursos de armazenamento da workload.

Práticas comuns que devem ser evitadas:

- Você não tem uma estratégia de backup para seus dados.
- Você faz backup de dados que podem ser recriados com facilidade.

Benefícios de implementar esta prática recomendada: evitar o backup de dados não críticos reduz os recursos de armazenamento necessários para a workload e diminui seu impacto ambiental.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Evitar o backup de dados desnecessários pode ajudar a reduzir os custos e os recursos de armazenamento usados pela workload. Faça backup somente de dados com valor empresarial ou que sejam necessários para atender a requisitos de conformidade. Examine as políticas de backup e exclua armazenamentos temporários que não fornecem valor em um cenário de recuperação.

Etapas de implementação

- Classifique dados: implemente a política de classificação de dados conforme descrito em [SUS04-BP01 Implementar uma política de classificação de dados](#).
- Crie uma estratégia de backup: use a importância da classificação de dados e projete uma estratégia de backup com base no [objetivo de tempo de recuperação \(RTO\)](#) e no [objetivo de ponto de recuperação \(RPO\)](#). Evite fazer backup de dados não essenciais.
 - Exclua dados que podem ser recriados com facilidade.
 - Exclua dados temporários dos seus backups.
 - Exclua cópias locais de dados, a menos que o tempo necessário para restaurar esses dados de um local comum exceda seus Acordos de Serviço (SLAs).
- Use backup automatizado: use uma solução automatizada ou um serviço gerenciado para fazer backup de dados essenciais aos negócios.
 - O [AWS Backup](#) é um serviço totalmente gerenciado que ajuda você a centralizar e automatizar a proteção de dados nos serviços da AWS, na nuvem e on-premises. Para obter orientação prática sobre como criar backups automatizados usando o AWS Backup, consulte [Laboratórios do Well-Architected: Testar o backup e a restauração de dados](#).
 - [Automatize backups e otimize os custos de backup para o Amazon EFS usando o AWS Backup](#).

Recursos

Práticas recomendadas relacionadas:

- [REL09-BP01 Identificar e fazer backup de todos os dados que precisam de backup ou reproduzir os dados das fontes](#)
- [REL09-BP03 Fazer backup de dados automaticamente](#)
- [REL13-BP02 Usar estratégias de recuperação definidas para cumprir os objetivos de recuperação](#)

Documentos relacionados:

- [Usar o AWS Backup para fazer backup e restaurar sistemas de arquivos do Amazon EFS](#)
- [Snapshots do Amazon EBS](#)
- [Trabalhar com backups no Amazon Relational Database Service](#)
- [Parceiro da APN: parceiros que podem ajudar com o backup](#)
- [AWS Marketplace: produtos que podem ser usados para backup](#)
- [Fazer backup do Amazon EFS](#)
- [Fazer backup do Amazon FSx para Windows File Server](#)
- [Backup e restauração do Amazon ElastiCache \(Redis OSS\)](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Estratégias de backup e recuperação de desastres para aumentar a resiliência](#)
- [AWS re:Invent 2023: Novidades no AWS Backup](#)
- [AWS re:Invent 2021: Backup, recuperação de desastres e proteção contra ransomware com a AWS](#)

Hardware e serviços

Procure oportunidades para reduzir os impactos na sustentabilidade da workload fazendo mudanças nas suas práticas de gerenciamento de hardware. Minimize a quantidade de hardware necessária para provisionar e implantar e escolha o hardware e os serviços mais eficientes para sua workload específica.

Práticas recomendadas

- [SUS05-BP01 Usar a quantidade mínima de hardware para atender às suas necessidades](#)
- [SUS05-BP02 Usar tipos de instância com o mínimo de impacto](#)
- [SUS05-BP03 Usar serviços gerenciados](#)
- [SUS05-BP04 Otimizar o uso de aceleradores de computação baseados em hardware](#)

SUS05-BP01 Usar a quantidade mínima de hardware para atender às suas necessidades

Use a quantidade mínima de hardware para sua workload para atender com eficiência às suas necessidades de negócios.

Práticas comuns que devem ser evitadas:

- Você não monitora a utilização de recursos.
- Você tem recursos com baixo nível de utilização em sua arquitetura.
- Você não analisa a utilização de hardware estático para determinar se é necessário redimensioná-lo.
- Você não define metas de utilização de hardware para sua estrutura de computação com base nos KPIs de negócios.

Benefícios de implementar esta prática recomendada: o dimensionamento correto de seus recursos de nuvem ajuda a reduzir o impacto ambiental de uma workload, economizar dinheiro e manter os benchmarks de performance.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Selecione o modo mais eficiente o número total de hardware necessário à workload para melhorar a eficiência geral. A Nuvem AWS fornece a flexibilidade de expandir ou reduzir o número de recursos dinamicamente por meio de diversos mecanismos, como o [AWS Auto Scaling](#), para atender a mudanças na demanda. Ela também fornece [APIs e SDKs](#) que permitem que os recursos sejam modificados com o mínimo de esforço. Use esses recursos para fazer alterações frequentes nas implementações da workload. Além disso, use as orientações sobre dimensionamento correto

das ferramentas da AWS para operar com eficiência os recursos de nuvem e atender às suas necessidades empresariais.

Etapas de implementação

- Escolha o tipo de instância: escolha o tipo de instância certo para melhor atender às suas necessidades. Para saber como escolher instâncias do Amazon Elastic Compute Cloud e usar mecanismos, como a seleção de instâncias com base em atributos, consulte:
 - [Como faço para escolher o tipo de instância do Amazon EC2 apropriado para minha workload?](#)
 - [Seleção de tipo de instância baseada em atributos para Amazon EC2 Fleet.](#)
 - [Criar um grupo do Auto Scaling usando seleção de tipo de instância baseada em atributos.](#)
- Escala: use pequenos incrementos para workloads variáveis.
- Use várias opções de compra de computação: equilibre a flexibilidade, a escalabilidade e a redução de custos da instância com várias opções de compra de computação.
 - As [instâncias sob demanda do Amazon EC2](#) são mais adequadas para workloads novas, dinâmicas e com estado que não podem ser flexíveis em termos de tipo de instância, localização ou horário.
 - As [instâncias spot do Amazon EC2](#) são uma ótima maneira de complementar as outras opções para aplicações que são flexíveis e tolerantes a falhas.
 - Aproveite os [Savings Plans para computação](#) para workloads de estado estável que permitem flexibilidade se suas necessidades (como AZ, região, famílias de instâncias ou tipos de instância) mudarem.
- Use a diversidade de instâncias e zonas de disponibilidade: maximize a disponibilidade das aplicações e aproveite o excesso de capacidade diversificando suas instâncias e zonas de disponibilidade.
- Dimensione as instâncias corretamente: use as recomendações de dimensionamento correto das ferramentas da AWS para fazer ajustes na workload. Para obter mais informações, consulte [Como otimizar seu custo com as recomendações de redimensionamento](#) correto e [Dimensionamento correto: provisionamento de instâncias para corresponder às workloads](#).
 - Use as recomendações de dimensionamento correto no AWS Cost Explorer ou no [AWS Compute Optimizer](#) para identificar oportunidades de dimensionamento correto.
- Negocie acordos de serviço (SLAs): que permitam uma redução temporária na capacidade enquanto a automação implanta recursos de substituição.

Recursos

Documentos relacionados:

- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte I: Computação](#)
- [Seleção de tipo de instância baseada em atributos para Auto Scaling para Amazon EC2 Fleet](#)
- [Documentação do AWS Compute Optimizer](#)
- [Operação do Lambda: otimização da performance](#)
- [Documentação do Auto Scaling](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Novidades no Amazon EC2](#)
- [AWS re:Invent 2023: Economias inteligentes: estratégias de otimização de custos no Amazon Elastic Compute Cloud](#)
- [AWS re:Invent 2022: Otimizar o Amazon Elastic Kubernetes Service para performance e custo na AWS](#)
- [AWS re:Invent 2023: Computação sustentável: reduzir custos e emissões de carbono com a AWS](#)

SUS05-BP02 Usar tipos de instância com o mínimo de impacto

Monitore continuamente e use novos tipos de instância para aproveitar as melhorias de eficiência de energia.

Práticas comuns que devem ser evitadas:

- Você usa apenas uma família de instâncias.
- Você usa apenas instâncias x86.
- Você especifica um tipo de instância em sua configuração do Amazon EC2 Auto Scaling.
- Você usa instâncias da AWS de um modo para o qual elas não foram projetadas (por exemplo, você usa instâncias otimizadas para computação em uma workload com uso intenso de memória).
- Você não avalia os novos tipos de instância regularmente.
- Você não verifica as recomendações de ferramentas de dimensionamento correto da AWS, como o [AWS Compute Optimizer](#).

Benefícios de implementar esta prática recomendada: ao usar instâncias com eficiência de energia e dimensionadas corretamente, é possível reduzir ainda mais o impacto ambiental e o custo da workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Usar instâncias eficientes na workload de nuvem é essencial para reduzir o uso de recursos e os custos. Monitore continuamente o lançamento de novos tipos de instância e aproveite as melhorias de eficiência de energia, incluindo os tipos de instância projetados para comportar workloads específicas, como treinamento e inferência de machine learning e transcodificação de vídeo.

Etapas de implementação

- Conheça e explore tipos de instâncias: encontre tipos de instâncias que podem reduzir o impacto ambiental da sua workload.
 - Assine as [Novidades da AWS](#) para se manter em dia com as tecnologias e instâncias mais recentes da AWS.
 - Conheça os diversos tipos de instâncias da AWS.
 - Saiba mais sobre instâncias baseadas no AWS Graviton que oferecem a melhor performance por watt de uso de energia no Amazon EC2 assistindo ao [re:Invent 2020: Mergulho profundo nas instâncias do Amazon EC2 baseadas no processador AWS Graviton2](#) e [Mergulho profundo no AWS Graviton 3 e instâncias C7g do Amazon EC2](#).
- Use tipos de instâncias com o mínimo de impacto: planeje e migre sua workload para tipos de instância com impacto mínimo.
 - Defina um processo para avaliar novos recursos ou instâncias para a workload. Aproveite a agilidade da nuvem para testar rapidamente como novos tipos de instância podem melhorar a sustentabilidade ambiental de sua workload. Use métricas de proxy para mensurar quantos recursos são necessários para concluir uma unidade de trabalho.
 - Se possível, modifique sua workload para trabalhar com diferentes números de vCPUs e diferentes quantidades de memória para maximizar sua escolha de tipo de instância.
 - Pense em migrar a workload para instâncias baseadas em Graviton e melhorar a eficiência da performance da workload. Para obter mais informações sobre como mover workloads para o AWS Graviton, consulte [Início rápido do AWS Graviton](#) e [Considerações ao fazer a transição de workloads para instâncias do Amazon Elastic Compute Cloud baseadas no AWS Graviton](#).
 - Considere selecionar a opção AWS Graviton em seu uso de [serviços gerenciados da AWS](#).

- Migre sua workload para regiões que ofereçam instâncias com o menor impacto na sustentabilidade e atendam aos seus requisitos de negócios.
- Para workloads de machine learning, utilize hardware específico para sua workload, como [AWS Trainium](#), [AWS Inferentia](#) e [Amazon EC2 DL1](#). AWS Instâncias do Inferentia, como instâncias Inf2, oferecem performance até 50% melhor por watt em relação a instâncias comparáveis do Amazon EC2.
- Use o [Amazon SageMaker AI Inference Recommender](#) para dimensionar corretamente o endpoint de inferência de ML.
- Para workloads com picos (workloads com requisitos irregulares para capacidade adicional), use [instâncias de performance expansível](#).
- Para workloads sem estado e tolerantes a falhas, use [instâncias spot do Amazon EC2](#) para aumentar a utilização geral da nuvem e reduzir o impacto na sustentabilidade de recursos não utilizados.
- Opere e optimize: opere e optimize a instância da sua workload.
 - Para workloads efêmeras, avalie as [métricas da instância do Amazon CloudWatch](#), como CPUUtilization, para identificar se a instância está ociosa ou subutilizada.
 - Para workloads estáveis, verifique as ferramentas de dimensionamento correto da AWS, como [AWS Compute Optimizer](#), em intervalos regulares para identificar oportunidades de otimizar e dimensionar corretamente as instâncias. Para obter mais exemplos e recomendações, consulte os seguintes laboratórios:
 - [Laboratório do Well-Architected: Recomendações de dimensionamento correto](#)
 - [Laboratório do Well-Architected: Dimensionamento correto com o Compute Optimizer](#)
 - [Laboratório do Well-Architected: Otimizar padrões de hardware e observar KPIs de sustentabilidade](#)

Recursos

Documentos relacionados:

- [Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte I: Computação](#)
- [AWS Graviton](#)
- [DL1 do Amazon EC2](#)
- [Frotas de reserva de capacidade do Amazon EC2](#)
- [Frota spot do Amazon EC2](#)

- [Funções: configuração da função do Lambda](#)
- [Seleção de tipo de instância baseada em atributos para Amazon EC2 Fleet](#)
- [Criar aplicações sustentáveis, eficientes e com custo otimizado na AWS](#)
- [Como o Painel de Sustentabilidade da Contino ajuda os clientes a otimizar sua pegada de carbono](#)

Vídeos relacionados:

- [AWS re:Invent 2023: AWS Graviton: a melhor performance de preços para suas workloads da AWS](#)
- [AWS re:Invent 2023: Novos recursos de IA generativa do Amazon Elastic Compute Cloud no AWS Management Console](#)
- [AWS re:Invent 2023: Novidades do Amazon Elastic Compute Cloud](#)
- [AWS re:Invent 2023: Economias inteligentes: estratégias de otimização de custos no Amazon Elastic Compute Cloud](#)
- [AWS re:Invent 2021: Mergulho profundo no AWS Graviton 3 e instâncias C7g do Amazon EC2](#)
- [AWS re:Invent 2022: Criar um ambiente de computação eficiente em termos de custo, energia e recursos](#)

Exemplos relacionados:

- [Solução: orientações sobre como otimizar workloads de aprendizado profundo para sustentabilidade na AWS](#)

SUS05-BP03 Usar serviços gerenciados

Use serviços gerenciados para operar com maior eficiência na nuvem.

Práticas comuns que devem ser evitadas:

- Você usa instâncias do Amazon EC2 com baixa utilização para executar suas aplicações.
- Sua equipe interna gerencia apenas a workload e não tem tempo para se concentrar em inovação ou simplificações.
- Você implanta e mantém tecnologias para tarefas que podem ser executadas com maior eficiência em serviços gerenciados.

Benefícios de implementar esta prática recomendada:

- Com o uso de serviços gerenciados, a responsabilidade é transferida para a AWS, que tem insights referentes a milhões de clientes que podem ajudar a promover inovações inéditas e melhorar a eficiência.
- O serviço gerenciado distribui o impacto ambiental do serviço entre vários usuários em virtude dos ambientes de gerenciamento de vários locatários.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Como os serviços gerenciados, a responsabilidade por manter a alta utilização e otimizar a sustentabilidade do hardware implantado é transferida para a AWS. Os serviços gerenciados também eliminam as despesas operacionais e administrativas da manutenção de um serviço, o que permite que sua equipe tenha mais tempo para se concentrar na inovação.

Avalie sua workload para identificar componentes que podem ser substituídos por serviços gerenciados da AWS. Por exemplo, o [Amazon RDS](#), o [Amazon Redshift](#) e o [Amazon ElastiCache](#) fornecem um serviço de banco de dados gerenciado. O [Amazon Athena](#), o [Amazon EMR](#) e o [Amazon OpenSearch Service](#) oferecem um serviço de análise gerenciado.

Etapas de implementação

1. Faça o inventário da workload: faça um inventário de serviços e componentes para sua workload.
2. Identifique candidatos: avalie e identifique componentes que podem ser substituídos por serviços gerenciados. Veja aqui alguns exemplos de quando considerar usar um serviço gerenciado:

Tarefa	O que usar na AWS
Hospedar um banco de dados	Use instâncias do Amazon Relational Database Service (Amazon RDS) gerenciadas em vez de manter instâncias do Amazon RDS no Amazon Elastic Compute Cloud (Amazon EC2) .
Hospedar uma workload containerizada	Use o AWS Fargate em vez de implementar sua própria infraestrutura de contêineres.

Tarefa	O que usar na AWS
Hospedar aplicações Web	Use o AWS Amplify Hosting como CI/CD totalmente gerenciado e serviço de hospedagem para sites estáticos e aplicações Web renderizadas no lado do servidor.

3. Crie um plano de migração: identifique dependências e crie um plano de migração. Atualize runbooks e playbook de forma apropriada.
 - O [AWS Application Discovery Service](#) coleta e apresenta automaticamente informações detalhadas sobre dependências e utilização de aplicações que ajudam a tomar decisões mais fundamentadas durante o planejamento da migração.
4. Faça testes: teste o serviço antes de migrar para o serviço gerenciado.
5. Substitua os serviços auto-hospedados: use seu plano de migração para substituir os serviços auto-hospedados por serviços gerenciados.
6. Monitore e ajuste: monitore continuamente o serviço após a conclusão da migração para fazer ajustes conforme necessário e otimizar o serviço.

Recursos

Documentos relacionados:

- [Produtos da Nuvem AWS](#)
- [Calculadora de custo total de propriedade \(TCO\) da AWS](#)
- [Amazon DocumentDB](#)
- [Amazon Elastic Kubernetes Service \(EKS\)](#)
- [Amazon Managed Streaming for Apache Kafka \(Amazon MSK\)](#)

Vídeos relacionados:

- [AWS re:Invent 2021: Operações na nuvem em grande escala com o AWS Managed Services](#)
- [AWS re:Invent 2023: Práticas recomendadas para operar na AWS](#)

SUS05-BP04 Otimizar o uso de aceleradores de computação baseados em hardware

Otimize o uso de instâncias com computação acelerada para reduzir as demandas de infraestrutura física de sua workload.

Práticas comuns que devem ser evitadas:

- Você não está monitorando o uso da GPU.
- Você está usando uma instância de finalidade geral para workload, enquanto uma instância criada especificamente pode oferecer maior performance, menor custo e melhor performance por watt.
- Você está usando aceleradores de computação baseados em hardware para tarefas em que são mais eficientes usando alternativas baseadas em CPU.

Benefícios de implementar esta prática recomendada: ao otimizar o uso de aceleradores baseados em hardware, é possível reduzir as demandas de infraestrutura física da workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Se você precisar de alta capacidade de processamento, poderá se beneficiar do uso de instâncias com computação acelerada, que fornecem acesso a aceleradores de computação baseados em hardware, como unidades de processamento gráfico (GPUs) e matrizes de portas programáveis em campo (FPGAs). Esses aceleradores de hardware executam certas funções, como processamento gráfico ou correspondência de padrões de dados, com mais eficiência do que alternativas baseadas em CPU. Muitas workloads aceleradas, como renderização, transcodificação e machine learning, são altamente variáveis em termos de uso de recursos. Execute esse hardware apenas pelo tempo necessário e desative-o com automação quando não precisar mais dele para minimizar o consumo de recursos.

Etapas de implementação

- Explore aceleradores de computação: identifique quais [instâncias com computação aceleradas](#) podem atender às suas necessidades.
- Use hardware específico: para workloads de machine learning, utilize hardware específico para a workload, como [AWS Trainium](#), [AWS Inferentia](#) e [Amazon EC2 DL1](#). As instâncias do AWS

Inferentia, como instâncias Inf2, oferecem desempenho até [50% melhor por watt em relação a instâncias comparáveis do Amazon EC2](#).

- Monitore métricas de uso: colete métricas de uso para as instâncias com computação acelerada. Por exemplo, você pode usar o agente do CloudWatch para coletar métricas como `utilization_gpu` e `utilization_memory` e para suas GPUs, conforme mostrado em [Coletar métricas de GPU NVIDIA com o Amazon CloudWatch](#).
- Faça o dimensionamento correto: otimize o código, a operação de rede e as configurações dos aceleradores de hardware para garantir que o hardware subjacente seja totalmente utilizado.
 - [Otimizar as configurações da GPU](#)
 - [Monitoramento e otimização da GPU na AMI de aprendizado profundo](#)
 - [Otimizar a E/S para ajuste de performance da GPU de treinamento de aprendizado profundo no Amazon SageMaker AI](#)
- Mantenha a atualização: use as mais recentes bibliotecas de alto desempenho e drivers de GPU.
- Libere instâncias não utilizadas: use automação para liberar instâncias de GPU quando não estiverem em uso.

Recursos

Documentos relacionados:

- [Computação acelerada](#)
- [Vamos arquitetar! Como arquitetar com chips e aceleradores personalizados](#)
- [Como faço para escolher o tipo de instância do Amazon EC2 apropriado para minha workload?](#)
- [Instâncias VT1 do Amazon EC2](#)
- [Escolher o melhor acelerador de IA e compilação de modelos para inferência de visão computacional com o Amazon SageMaker AI](#)

Vídeos relacionados:

- [AWS re:Invent 2021: Como selecionar instâncias de GPU do Amazon EC2 para aprendizado profundo](#)
- [AWS Online Tech Talks: Implantar inferência de aprendizado profundo eficiente em termos de custos](#)
- [AWS re:Invent 2023: IA de última geração com a AWS e a NVIDIA](#)

- [AWS re:Invent 2022 \[NOVO LANÇAMENTO!\]: Introdução as instâncias Inf2 do Amazon EC2 baseadas no AWS Inferentia2](#)
- [AWS re:Invent 2022: Acelere o aprendizado profundo e inove com mais rapidez com o AWS Trainium](#)
- [AWS re:Invent 2022: Aprendizado profundo na AWS com a NVIDIA: do treinamento à implantação](#)

Processo e cultura

Procure oportunidades para reduzir seu impacto na sustentabilidade fazendo mudanças nas suas práticas de desenvolvimento, teste e implantação.

Práticas recomendadas

- [SUS06-BP01 Comunicar e distribuir as metas de sustentabilidade em cascata](#)
- [SUS06-BP02 Adotar métodos que podem apresentar melhorias na sustentabilidade rapidamente](#)
- [SUS06-BP03 Manter a workload atualizada](#)
- [SUS06-BP04 Aumentar a utilização de ambientes de compilação](#)
- [SUS06-BP05 Usar parques de dispositivos gerenciados para testes](#)

SUS06-BP01 Comunicar e distribuir as metas de sustentabilidade em cascata

A tecnologia é um dos principais facilitadores da sustentabilidade. As equipes de TI desempenham um papel crucial na condução de mudanças significativas em direção às metas de sustentabilidade da sua organização. Essas equipes devem entender claramente as metas de sustentabilidade da empresa e trabalhar para comunicar e distribuir essas prioridades em todas as operações.

Práticas comuns que devem ser evitadas:

- Você não conhece as metas de sustentabilidade da sua organização e como elas se aplicam à sua equipe.
- Você não tem conhecimento e treinamento suficientes sobre o impacto ambiental das workloads na nuvem.
- Você não tem certeza sobre as áreas específicas a serem priorizadas.
- Você não envolve seus funcionários e clientes em suas iniciativas de sustentabilidade.

Benefícios de implementar essa prática recomendada: da otimização da infraestrutura e dos sistemas ao uso de tecnologias inovadoras, as equipes de TI podem reduzir as emissões de carbono da organização e minimizar o consumo de recursos. A comunicação das metas de sustentabilidade pode permitir que as equipes de TI melhorem e se adaptem continuamente aos desafios de sustentabilidade em evolução. Além disso, essas otimizações sustentáveis geralmente também se traduzem em economia de custos, o que fortalece o caso de negócios.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

As principais metas de sustentabilidade das equipes de TI devem ser otimizar sistemas e soluções para aumentar a eficiência dos recursos e minimizar a pegada de carbono da organização e o impacto ambiental geral. Serviços e iniciativas compartilhados, como programas de treinamento e painéis operacionais, podem apoiar as organizações na otimização das operações de TI e na criação de soluções que podem ajudar a reduzir significativamente a pegada de carbono. A nuvem apresenta uma oportunidade não apenas de transferir a infraestrutura física e as responsabilidades de aquisição de energia para a responsabilidade compartilhada do provedor de nuvem, mas também de otimizar continuamente a eficiência dos recursos dos serviços baseados em nuvem.

Quando as equipes usam a eficiência inerente da nuvem e o modelo de responsabilidade compartilhada, elas podem promover reduções significativas no impacto ambiental da organização. Isso, por sua vez, pode contribuir para as metas gerais de sustentabilidade da organização e demonstrar o valor dessas equipes como parceiras estratégicas na jornada rumo a um futuro mais sustentável.

Etapas de implementação

- Defina metas e objetivos: estabeleça metas bem definidas para seu programa de TI. Isso envolve obter informações de partes interessadas responsáveis de diferentes departamentos, como TI, sustentabilidade e finanças. Essas equipes devem definir metas mensuráveis que se alinhem às metas de sustentabilidade da sua organização, incluindo áreas como redução de carbono e otimização de recursos.
- Entenda os limites da contabilidade de carbono da sua empresa: entenda como os métodos de contabilização de carbono, como o Protocolo de Gases de Efeito Estufa (GHG), se relacionam com suas workloads na nuvem (para obter mais detalhes, consulte [Sustentabilidade da nuvem](#)).
- Use soluções em nuvem para contabilização de carbono: use soluções em nuvem, como [soluções de contabilidade de carbono on](#), AWS para rastrear os escopos um, dois e três das emissões de

GEE em suas operações, portfólios e cadeias de valor. Com essas soluções, as organizações podem agilizar a aquisição de dados de emissões de GEE, simplificar os relatórios e obter insights para embasar suas estratégias climáticas.

- Monitore a pegada de carbono do seu portfólio de TI: acompanhe e relate as emissões de carbono de seus sistemas de TI. Use a [AWS Customer Carbon Footprint Tool](#) para rastrear, medir, analisar e prever as emissões de carbono geradas pelo seu AWS uso.
- Comunique o uso de recursos por meio de métricas de proxy para suas equipes: acompanhe e relate seu [uso de recursos por meio de métricas de proxy](#). Nos modelos de preços sob demanda da nuvem, o uso de recursos está relacionado ao custo, que é uma métrica geralmente compreensível. No mínimo, use o custo como uma métrica intermediária para comunicar o uso de recursos e as melhorias de cada equipe.
- Habilite a granularidade por hora no Explorador de Custos e crie um [Relatório de Custos e Uso \(CUR\)](#): o CUR fornece granularidade de uso, taxas, custos e atributos de uso por dia ou por hora para todos os serviços da AWS. Use [os painéis de inteligência de nuvem](#) e seu painel de métricas de proxy de sustentabilidade como ponto de partida para o processamento e visualização de dados baseados em custo e uso. Para obter mais detalhes, consulte as informações a seguir.
- [Measure and track cloud efficiency with sustainability proxy metrics, Part I: What are proxy metrics?](#)
- [Measure and track cloud efficiency with sustainability proxy metrics, Part II: Establish a metrics pipeline](#)
- Otimize e avalie continuamente: use um [processo de melhoria](#) para otimizar continuamente seus sistemas de TI, incluindo a workload na nuvem para eficiência e sustentabilidade. Monitore a pegada de carbono antes e depois da implementação da estratégia de otimização. Use a redução na pegada de carbono para avaliar a eficácia.
- Promova uma cultura de sustentabilidade: use programas de treinamento (como o [AWS Skill Builder](#)) para educar seus funcionários sobre sustentabilidade. Envolve-os em iniciativas de sustentabilidade. Compartilhe e celebre suas histórias de sucesso. Use incentivos para premiá-los se atingirem as metas de sustentabilidade.

Recursos

Documentos relacionados:

- [Noções básicas das suas estimativas de emissão de carbono](#)

Vídeos relacionados:

- [AWS re:Invent 2023 - Accelerate data-driven circular economy initiatives with AWS](#)
- [AWS re:Invent 2023 - Sustainability innovation in AWS Global Infrastructure](#)
- [AWS re:Invent 2023: Arquitetura sustentável: passado, presente e futuro](#)
- [AWS re:Invent 2022: Como entregar arquiteturas sustentáveis e de alta performance](#)
- [AWS re:Invent 2022: Arquitetar de forma sustentável e reduzir sua pegada de carbono da AWS](#)
- [AWS re:Invent 2022: Sustentabilidade na infraestrutura global da AWS](#)

Exemplos relacionados:

- [Laboratório do Well-Architected: Transformar relatórios de custo e uso em relatórios de eficiência](#)

Treinamento relacionado:

- [Transformação da sustentabilidade em AWS](#)
- [SimuLearn: Relatórios de sustentabilidade](#)
- [Descarbonização com AWS](#)

SUS06-BP02 Adotar métodos que podem apresentar melhorias na sustentabilidade rapidamente

Adote métodos e processos para validar possíveis aprimoramentos, minimizar o custo dos testes e fornecer pequenas melhorias.

Práticas comuns que devem ser evitadas:

- A avaliação da sustentabilidade de sua aplicação é uma tarefa que é feita apenas uma vez no início de um projeto.
- Como o processo de lançamento para introduzir pequenas alterações em prol da eficiência dos recursos é muito trabalhoso, sua workload tornou-se ultrapassada.
- Você não tem mecanismos para melhorar a sustentabilidade de sua workload.

Benefícios de implementar esta prática recomendada: ao estabelecer um processo para introduzir e monitorar melhorias de sustentabilidade, você poderá adotar continuamente novos recursos e capacidades, remover problemas e melhorar a eficiência da workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Médio

Orientação para implementação

Teste e valide as possíveis melhorias de sustentabilidade antes de implantá-las na produção. Considere o custo do teste ao calcular o benefício futuro potencial de uma melhoria. Desenvolva métodos de teste de baixo custo para oferecer pequenas melhorias.

Etapas de implementação

- Entenda e comunique suas metas de sustentabilidade organizacional: entenda suas metas de sustentabilidade organizacional, como redução de carbono ou administração da água. Converta essas metas em requisitos de sustentabilidade para suas workloads na nuvem. Comunique esses requisitos às principais partes interessadas.
- Adicione requisitos de sustentabilidade à sua lista de pendências: adicione requisitos para melhoria da sustentabilidade à sua lista de pendências de desenvolvimento.
- Itere e melhore: use um [processo de melhoria iterativo](#) para identificar, avaliar, priorizar, testar e implantar essas melhorias.
- Teste usando o produto mínimo viável (MVP): desenvolva e teste possíveis melhorias usando os componentes representativos mínimos viáveis para reduzir o custo e o impacto ambiental dos testes.
- Racionalize o processo: melhore e otimize continuamente seus processos de desenvolvimento. A título de exemplo, automatize o processo de entrega de software usando pipelines de integração contínua e entrega contínua (CI/CD) a fim de testar e implantar possíveis melhorias para reduzir o nível de esforço e limitar os erros provocados por processos manuais.
- Treinamento e conscientização: realize programas de treinamento para os membros da sua equipe para educá-los em sustentabilidade e como suas atividades afetam suas metas de sustentabilidade organizacional.
- Avalie e ajuste: avalie continuamente o impacto das melhorias e faça ajustes conforme necessário.

Recursos

Documentos relacionados:

- [A AWS viabiliza soluções de sustentabilidade](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Arquitetura sustentável: passado, presente e futuro](#)
- [AWS re:Invent 2022: Como entregar arquiteturas sustentáveis e de alta performance](#)
- [AWS re:Invent 2022: Arquitetar de forma sustentável e reduzir sua pegada de carbono da AWS](#)
- [AWS re:Invent 2022: Sustentabilidade na infraestrutura global da AWS](#)
- [AWS re:Invent 2023 - What's new with AWS observability and operations](#)

SUS06-BP03 Manter a workload atualizada

Mantenha sua workload atualizada para adotar recursos eficientes, eliminar problemas e melhorar a eficiência geral da workload.

Práticas comuns que devem ser evitadas:

- Você pressupõe que sua arquitetura atual é estática e não será atualizada ao longo do tempo.
- Você não tem nenhum sistema ou ritmo regular para avaliar se software ou pacotes atualizados são compatíveis com sua workload.

Benefícios de implementar esta prática recomendada: ao estabelecer um processo para manter a workload atualizada, você poderá adotar novos recursos e capacidades, resolver problemas e aumentar a eficiência da workload.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Sistemas operacionais, runtimes, middleware, bibliotecas e aplicações atualizados podem melhorar a eficiência da workload e facilitar a adoção de tecnologias mais eficientes. Um software atualizado também pode incluir recursos para medir o impacto na sustentabilidade da workload com maior precisão, pois os fornecedores oferecem recursos para atender às suas próprias metas de sustentabilidade. Adote um ritmo regular para manter a workload atualizada com os recursos e versões mais recentes.

Etapas de implementação

- Defina um processo: use um processo e um cronograma para avaliar novos recursos ou instâncias para sua workload. Aproveite a agilidade da nuvem para testar rapidamente como novos recursos podem melhorar a workload com o objetivo de:
 - Reduzir impactos de sustentabilidade.
 - Obter eficiências de performance.
 - Remover barreiras a melhorias planejadas.
 - Aumentar sua capacidade de medir e gerenciar impactos na sustentabilidade.
- Faça o inventário da workload: faça o inventário de software e arquitetura da workload e identifique os componentes que precisam ser atualizados.
 - É possível usar o [Systems Manager Inventory da AWS](#) para coletar metadados de sistema operacional (SO), aplicação e instância das instâncias do Amazon EC2 e entender rapidamente quais instâncias executam o software e as configurações exigidas pela política de software e quais instâncias precisam ser atualizadas.
- Avalie a nova atualização: entenda como atualizar os componentes da sua workload.

Componente de Workload	Como atualizar
Imagens de máquina	Use o EC2 Image Builder para gerenciar atualizações em imagens de máquina da Amazon (AMIs) para imagens de servidores Linux ou Windows.
Imagens de contêiner	Use o Amazon Elastic Container Registry (Amazon ECR) com seu pipeline atual para gerenciar as imagens do Amazon Elastic Container Service (Amazon ECS) .
AWS Lambda	O AWS Lambda inclui recursos de gerenciamento de versão .

- Use automação: automatize as atualizações para reduzir o nível de esforço para implantar novos recursos e limitar erros causados por processos manuais.

- É possível usar [CI/CD](#) para atualizar automaticamente AMIs, imagens de contêiner e outros artefatos relacionados à aplicação de nuvem.
- Você pode usar ferramentas como o [Gerenciador de Patches do AWS Systems Manager](#) para automatizar o processo de atualizações do sistema e agendar a atividade usando as [Janelas de Manutenção do AWS Systems Manager](#).

Recursos

Documentos relacionados:

- [Centro de Arquitetura da AWS](#)
- [Novidades da AWS](#)
- [Ferramentas de desenvolvedor AWS](#)

Vídeos relacionados:

- [AWS re:Invent 2022: Otimizar suas workloads da AWS com a orientação de práticas recomendadas](#)
- [All Things Patch: AWS Systems Manager](#)

SUS06-BP04 Aumentar a utilização de ambientes de compilação

Aumente a utilização dos recursos para desenvolver, testar e compilar suas workloads.

Práticas comuns que devem ser evitadas:

- Você provisiona ou encerra manualmente seus ambientes de compilação.
- Você mantém seus ambientes de compilação em execução independentemente de atividades de teste, compilação ou lançamento (por exemplo, execução de um ambiente fora do horário de expediente dos membros de sua equipe de desenvolvimento).
- Você provisiona recursos em excesso para seus ambientes de compilação.

Benefícios de implementar esta prática recomendada: ao aumentar a utilização de ambientes de criação, você pode melhorar a eficiência geral da workload na nuvem e, ao mesmo tempo, alocar os recursos para que os criadores desenvolvam, testem e criem com eficiência.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Use a automação e a infraestrutura como código para ativar ambientes de compilação quando necessário e desativá-los quando não forem usados. Um padrão comum é programar períodos de disponibilidade que coincidam com as horas de trabalho dos membros da equipe de desenvolvimento. A configuração dos ambientes de teste deve ser bem semelhante à do ambiente de produção. Entretanto, procure oportunidades para usar tipos de instância com capacidade de expansão, instâncias spot do Amazon EC2, serviços de banco de dados com ajuste de escala automático, contêineres e tecnologias sem servidor para alinhar a capacidade de desenvolvimento e teste ao uso. Limite o volume de dados apenas para atender os requisitos de teste. Ao usar dados de produção no teste, explore possibilidades para compartilhar os dados da produção em vez de movimentá-los.

Etapas de implementação

- Use infraestrutura como código: use a infraestrutura como código para provisionar os ambientes de compilação.
- Use automação: use automação para gerenciar o ciclo de vida de seus ambientes de desenvolvimento e teste e maximizar a eficiência dos recursos de compilação.
- Maximize a utilização: use estratégias para maximizar a utilização de seus ambientes de desenvolvimento e teste.
 - Use ambientes representativos mínimos viáveis para desenvolver e testar possíveis melhorias.
 - Utilize tecnologias sem servidor, se possível.
 - Use instâncias sob demanda para complementar os dispositivos de desenvolvedor.
 - Use tipos de instância com capacidade de expansão, instâncias spot e outras tecnologias para alinhar a capacidade de compilação com o uso.
 - Adote serviços de nuvem nativos para acesso seguro ao shell de instância em vez de implantar frotas de hosts bastion.
 - Escale automaticamente seus recursos de compilação de acordo com seus trabalhos de compilação.

Recursos

Documentos relacionados:

- [Gerenciador de sessões do AWS Systems Manager](#)
- [Instâncias de performance expansível do Amazon EC2](#)
- [O que é AWS CloudFormation?](#)
- [O que é AWS CodeBuild?](#)
- [Agendador de instância na AWS](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Integração e entrega contínuas para AWS](#)

SUS06-BP05 Usar parques de dispositivos gerenciados para testes

Use parques de dispositivos gerenciados para testar com eficiência um novo recurso em um conjunto representativo de hardware.

Práticas comuns que devem ser evitadas:

- Você testa e implanta manualmente sua aplicação em dispositivos físicos individuais.
- Você não usa o serviço de testes de aplicação para testar e interagir com suas aplicações (por exemplo, Android, iOS e aplicações Web) em dispositivos físicos reais.

Benefícios de implementar esta prática recomendada: o uso de parques de dispositivos gerenciados para testar aplicações habilitadas para a nuvem oferece vários benefícios:

- Eles contam com recursos mais eficientes para testar a aplicação em uma ampla variedade de dispositivos.
- Eles eliminam a necessidade de infraestrutura interna para testes.
- Eles oferecem diversos tipos de dispositivo, incluindo hardware mais antigo e menos conhecido, eliminando a necessidade de atualizações de dispositivo desnecessárias.

Nível de risco exposto se esta prática recomendada não for estabelecida: Baixo

Orientação para implementação

Usar parques de dispositivos gerenciados pode ajudar a otimizar o processo de testes de novos recursos em um conjunto representativo de hardware. Os parques de dispositivos gerenciados

oferecem diversos tipos de dispositivo, incluindo hardware mais antigo e menos conhecido, e evita o impacto sobre a sustentabilidade por parte do cliente devido a atualizações desnecessárias de dispositivo.

Etapas de implementação

- Defina seus requisitos de testes: defina seus requisitos e plano de testes (como tipo de teste, sistemas operacionais e programação dos testes).
 - O [Amazon CloudWatch RUM](#) pode ser usado para coletar e analisar dados do lado do cliente e definir seu plano de teste.
- Selecione um parque de dispositivos gerenciados: selecione um parque de dispositivos gerenciados capaz de suportar seus requisitos de teste. Por exemplo, é possível usar o [AWS Device Farm](#) para testar e entender o impacto das suas alterações em um conjunto representativo de hardware.
- Use automação: use a integração contínua/implantação contínua (CI/CD) para agendar e executar seus testes.
 - [Integrar o AWS Device Farm Device Farm com seu pipeline de CI/CD para executar testes de Selenium em vários navegadores](#)
 - [Criar e testar aplicações para iOS e iPadOS com o AWS DevOps e serviços móveis](#)
- Revise e ajuste: revise continuamente os resultados dos testes e faça as melhorias necessárias.

Recursos

Documentos relacionados:

- [Lista de dispositivos do AWS Device Farm](#)
- [Visualizar o painel do CloudWatch RUM](#)

Vídeos relacionados:

- [AWS re:Invent 2023: Melhorar a qualidade das suas aplicações móveis e Web com o AWS Device Farm](#)
- [AWS re:Invent 2021: Otimizar aplicações com base em insights do usuário final com o Amazon CloudWatch RUM](#)

Exemplos relacionados:

- [Exemplo de aplicação do AWS Device Farm para Android](#)
- [Exemplo de aplicação do AWS Device Farm para iOS](#)
- [Appium Web tests for AWS Device Farm](#)

Conclusão

Um número cada vez maior de organizações está estabelecendo metas de sustentabilidade em resposta às mudanças na regulamentação governamental, na vantagem competitiva e na demanda de clientes, funcionários e investidores. CTOs, arquitetos, desenvolvedores e membros das equipes de operações estão buscando maneiras de contribuir diretamente com as metas de sustentabilidade da organização. Ao usar esses princípios de design e as práticas recomendadas apoiadas pelos serviços da AWS, é possível tomar decisões informadas equilibrando segurança, custo, performance, confiabilidade e excelência operacional com resultados de sustentabilidade para suas workloads da Nuvem AWS. Cada ação que empregada para reduzir o uso de recursos e aumentar a eficiência em suas workloads contribui para a redução do impacto ambiental e contribui para as metas mais amplas de sustentabilidade de suas organizações.

Colaboradores

Os colaboradores deste documento incluem:

- Sam Mokhtari, arquiteto líder de soluções de eficiência, Amazon Web Services
- Brendan Sisson, arquiteto líder de soluções de sustentabilidade, Amazon Web Services
- Margaret O'Toole, líder técnica de sustentabilidade, Amazon Web Services
- Steffen Grunwald, arquiteto líder de soluções de sustentabilidade, Amazon Web Services
- Ryan Eccles, engenheiro líder de sustentabilidade, Amazon
- Rodney Lester, arquiteto líder, Amazon Web Services
- Adrian Cockcroft, vice-presidente de arquitetura de sustentabilidade, Amazon Web Services
- Ian Meyers, diretor de tecnologia, arquitetura de soluções, Amazon Web Services

Outras fontes de leitura

Para obter informações adicionais, consulte:

- [AWS Well-Architected](#)
- [Centro de Arquitetura do AWS](#)
- [Sustentabilidade na nuvem](#)
- [A AWS viabiliza soluções de sustentabilidade](#)
- [The Climate Pledge](#)
- [Objetivos de desenvolvimento sustentável das Nações Unidas](#)
- [Protocolo de Gases de Efeito de Estufa](#)

Revisões do documento

Para ser notificado sobre atualizações desse whitepaper, inscreva-se no feed RSS.

Alteração	Descrição	Data
Orientação atualizada sobre práticas recomendadas	As práticas recomendadas foram atualizadas com novas orientações nas seguintes áreas: SUS 1, SUS 3, SUS 4, SUS 5 e SUS 6. A orientação foi aprimorada nessas áreas de práticas recomendadas. A SUS 6 recebeu uma nova prática recomendada, a SUS06-BP01 Comunicar e distribuir as metas de sustentabilidade em cascata. As práticas recomendadas existentes na SUS 6 foram reenumeradas.	6 de novembro de 2024
Orientação atualizada sobre práticas recomendadas	Pequenas alterações em todo o pilar.	27 de junho de 2024
Níveis de risco atualizados	Pequenas atualizações nos níveis de risco das práticas recomendadas.	3 de outubro de 2023
Orientação atualizada sobre práticas recomendadas	As práticas recomendadas foram atualizadas com novas orientações nas seguintes áreas: Alinhamento à demanda , Software e arquitetura , Dados e Hardware e serviços .	13 de julho de 2023

Atualizado para o novo Framework	Atualizações nas práticas recomendadas com recomendações e adição de novas práticas recomendadas.	10 de abril de 2023
Whitepaper atualizado	Práticas recomendadas atualizadas com novas orientações para implementação.	15 de dezembro de 2022
Whitepaper atualizado	Práticas recomendadas ampliadas e planos de melhoria adicionados.	20 de outubro de 2022
Publicação inicial	Publicação do Pilar Sustentabilidade: AWS Well Architected Framework.	2 de dezembro de 2021

Avisos

Os clientes são responsáveis por fazer a própria avaliação independente das informações contidas neste documento. Este documento: (a) é apenas para fins informativos, (b) representa as ofertas e práticas de produtos atuais da AWS, que estão sujeitas a alterações sem aviso prévio e (c) não criam nenhum compromisso ou garantia da AWS e de suas afiliadas, fornecedores ou licenciadores. Os produtos ou serviços da AWS são fornecidos “no estado em que se encontram”, sem garantias, representações ou condições de qualquer tipo, expressas ou implícitas. As responsabilidades e as obrigações da AWS com os seus clientes são controladas por contratos da AWS, e este documento não é parte de, nem modifica, qualquer contrato entre a AWS e seus clientes.

© 2023 Amazon Web Services, Inc. ou suas afiliadas. Todos os direitos reservados.

Glossário da AWS

Para obter a terminologia mais recente da AWS, consulte o [glossário da AWS](#) na Referência do Glossário da AWS.