

Pilar da sustentabilidade



Pilar da sustentabilidade: AWS Well-Architected Framework

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Resumo e introdução	i
Introdução	1
Sustentabilidade da nuvem	2
O modelo de responsabilidade compartilhada	3
Sustentabilidade de nuvem	4
Sustentabilidade na nuvem	4
Sustentabilidade por meio da nuvem	4
Princípios de design para a sustentabilidade na nuvem	5
Processo de melhoria	7
Cenário de exemplo	8
Identificar os alvos para melhoria	8
Recursos	9
Avaliar melhorias específicas	9
Métricas de proxy	9
Métricas de negócios	10
Indicadores-chave de performance	10
Calcular a melhoria	11
Avaliar as melhorias	11
Priorizar e planejar as melhorias	13
Testar e validar as melhorias	13
Implantar alterações na produção	15
Medir os resultados e replicar os sucessos	15
Sustentabilidade como um requisito não funcional	18
Práticas recomendadas para a sustentabilidade na nuvem	20
Seleção de 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 com as metas de sustentabilidade	26
SUS02-BP03 Interromper a criação e a manutenção de ativos não utilizados	27
SUS02-BP04 Otimizar o posicionamento geográfico das workloads com base nos respectivos requisitos de rede	29
SUS02-BP05 Otimizar os recursos dos membros da equipe para as atividades realizadas	32

SUS02-BP06 Implementar armazenamento em buffer ou controle de utilização para nivelar a curva da demanda	33
Software e arquitetura	36
SUS03-BP01 Otimizar o software e a arquitetura para trabalhos assíncronos e programados	36
SUS03-BP02 Remover ou refatorar componentes da workload subutilizados ou não utilizados	39
SUS03-BP03 Otimizar as áreas de código que consomem mais tempo ou recursos	40
SUS03-BP04 Otimizar o impacto sobre dispositivos e equipamentos	42
SUS03-BP05 Usar arquiteturas e padrões de software que atendam melhor aos padrões de armazenamento e acesso aos dados	44
Gerenciamento de dados	46
SUS04-BP01 Implementar uma política de classificação de dados	47
SUS04-BP02 Usar tecnologias compatíveis com seus padrões de acesso e de armazenamento de dados	49
SUS04-BP03 Usar políticas para gerenciar o ciclo de vida de seus conjuntos de dados	54
SUS04-BP04 Usar elasticidade e automação para expandir o armazenamento em bloco ou o sistema de arquivos	56
SUS04-BP05 Remover dados desnecessários ou redundantes	58
SUS04-BP06 Usar sistemas de arquivos compartilhados ou armazenamento para acessar dados comuns	60
SUS04-BP07 Minimizar a movimentação de dados entre redes	62
SUS04-BP08 Fazer backup de dados somente quando for difícil recriar	65
Hardware e serviços	67
SUS05-BP01 Usar a quantidade mínima de hardware para atender às suas necessidades ...	67
SUS05-BP02 Usar tipos de instância com o mínimo de impacto	69
SUS05-BP03 Usar serviços gerenciados	72
SUS05-BP04 Otimizar o uso de aceleradores de computação baseados em hardware	75
Processo e cultura	77
SUS06-BP01 Adotar métodos que podem apresentar melhorias na sustentabilidade rapidamente	77
SUS06-BP02 Manter a workload atualizada	78
SUS06-BP03 Aumentar a utilização de ambientes de desenvolvimento	81
SUS06-BP04 Usar parques de dispositivos gerenciados para testes	82
Conclusão	85
Colaboradores	86

Leitura adicional	87
Revisões do documento	88
Avisos	89
AWS Glossary	90

Pilar Sustentabilidade: AWS Well-Architected Framework

Data de publicação: 3 de outubro de 2023 ([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 de 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 confiáveis, seguras, eficientes, econômicas e sustentáveis na Nuvem AWS. Ele fornece uma maneira de você avaliar consistentemente suas arquiteturas em relação às melhores práticas e identificar áreas de aprimoramento. As workloads bem-arquitetadas aumentam muito sua habilidade de oferecer suporte aos resultados da empresa.

A estrutura é baseada em seis pilares:

- Excelência operacional
- Segurança
- Confiabilidade
- Eficiência de performance
- Otimização de custos
- Sustentabilidade

Este documento se refere ao pilar Sustentabilidade e ao escopo correspondente, com foco na sustentabilidade ambiental. Ele é direcionado para 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 da nuvem

A disciplina da sustentabilidade aborda o impacto ambiental, econômico e social de longo prazo de suas atividades empresariais. O [Comissão Mundial das Nações Unidas sobre Meio Ambiente e Desenvolvimento](#) define o desenvolvimento sustentável como o “desenvolvimento que atende às necessidades do presente sem comprometer a capacidade das gerações futuras de atender 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 focaliza 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 focalizar os impactos ao ambiente, você deve compreender como esses impactos são normalmente levados em consideração e os impactos subsequentes à contabilidade de emissões da sua própria empresa. O [Protocolo de Gases de Efeito de Estufa](#) organiza as emissões de carbono nos escopos a seguir, junto com exemplos de emissões relevantes em cada escopo de um provedor de nuvem, como a AWS:

- Escopo 1: todas as emissões diretas das atividades de uma organização ou sob seu controle. Por exemplo, combustão de combustíveis por geradores de backup do datacenter.
- Escopo 2: emissões indiretas de eletricidade comprada e usada para energizar datacenters e outras instalações. Por exemplo, emissões de geração de energia comercial.
- Escopo 3: todas as outras emissões indiretas de atividades de uma organização de fontes não controladas por ela. Os exemplos da AWS incluem as emissões relacionadas à construção do datacenter, à manufatura e ao transporte de hardware de TI implantados em datacenters.

Da perspectiva de um cliente da AWS, as emissões de suas workloads em execução na AWS são consideradas como emissões indiretas e, portanto, 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 por workload e depende de vários fatores, incluindo os serviços da AWS usados, a energia consumida por esses serviços, a intensidade do carbono das redes

elétricas que servem os datacenters da AWS onde são executados, e a aquisição de energia renovável da AWS.

Este documento descreve primeiro um modelo de responsabilidade compartilhada para a sustentabilidade ambiental, e depois fornece as práticas recomendadas da arquitetura para que seja possível minimizar o impacto das suas workloads com a redução do total de recursos necessários para a execução delas nos datacenters 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 de nuvem, como entrega eficiente, infraestrutura compartilhada, gerenciamento responsável de água e a obtenção de energia renovável.
- Os clientes são responsáveis pela sustentabilidade in nuvem, com a otimização das workloads e da utilização de recursos e a minimização do total de recursos necessários para a implantação de suas workloads.



Modelo de responsabilidade compartilhada

Sustentabilidade de nuvem

Os provedores de nuvem têm uma pegada de carbono mais baixa e são mais eficientes em termos de energia do que as alternativas on-premises típicas, porque investem em energia eficiente e em tecnologia de resfriamento, operam populações de servidores eficientes em energia e obtêm altas taxas de utilização de servidores. As workloads de nuvem reduzem o impacto por se beneficiarem dos recursos compartilhados, como redes, energia, refrigeração e instalações físicas. É possível migrar suas workloads de nuvem para tecnologias mais eficientes, à medida que são disponibilizadas, e usar serviços baseados na nuvem para transformar suas workloads para obterem uma melhor sustentabilidade.

Recursos

- [The Carbon Reduction Opportunity of Moving to Amazon Web Services \(A oportunidade de redução de carbono com a mudança para a Amazon Web Services\)](#)
- [A AWS viabiliza soluções de sustentabilidade](#)

Sustentabilidade na nuvem

A sustentabilidade na nuvem é um esforço contínuo direcionado principalmente à redução do consumo de energia e à eficiência energética em todos os componentes de uma workload por meio da maximização dos benefícios dos recursos provisionados e da minimização do total de recursos necessários. Esse esforço pode envolver vários fatores, como 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 em uma infraestrutura de computação eficiente e corretamente dimensionada e a minimização dos requisitos de hardware com alto consumo de energia para o usuário final.

Sustentabilidade por meio da nuvem

Além de minimizar o impacto das workloads implantadas, é possível usar a Nuvem AWS para executar workloads projetadas para compatibilidade com seus maiores desafios de sustentabilidade. Os exemplos desses desafios incluem a redução de emissões de carbono, a minimização do consumo de energia, a reciclagem de água ou a redução de desperdício em outras áreas de sua empresa ou organização.

Sustentabilidade por meio da nuvem significa usar a tecnologia da AWS para solucionar um desafio maior de sustentabilidade. Por exemplo, é possível usar um serviço de machine learning, como o [Amazon Monitron](#), para detectar comportamento anormal em máquinas industriais. Com essa detecção de dados, é possível conduzir manutenção preventiva para reduzir o risco de incidentes ambientais causados por falhas inesperadas nos equipamentos e garantir que as máquinas continuem operantes com eficiência de pico.

Princípios de design para a sustentabilidade na nuvem

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

- **Compreender seu impacto:** Meça o impacto da seu workload na nuvem e modele seu impacto futuro. Inclua todas as fontes de impacto, inclusive aquelas resultantes do uso de seus produtos pelo cliente e da desativação e descontinuação deles. 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.
- **Estabelecer metas de sustentabilidade:** Para cada workload em nuvem, estabeleça metas de sustentabilidade de longo prazo, tais como 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.
- **Maximizar 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.
- **Antecipar e adotar novas ofertas de hardware e de software mais eficientes:** Apoie as melhorias preventivas que seus parceiros e fornecedores fazem para ajudar você a reduzir o impacto das workloads em nuvem. Monitore e avalie continuamente as ofertas de software e hardware novos e

mais eficientes. Projete visando a flexibilidade para permitir a adoção rápida de novas tecnologias eficientes.

- Usar serviços gerenciados: Compartilhar 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 em nuvem. Por exemplo, os clientes podem compartilhar o impacto de componentes comuns de um datacenter, como energia e redes, migrando as workloads para a Nuvem AWS e adotando serviços gerenciados, como o AWS Fargate, para contêineres com tecnologia sem servidor, operados em escala pela AWS, que é responsável pela eficiência da operação. 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 Amazon S3 Lifecycle ou o Amazon EC2 Auto Scaling para ajustar a capacidade de acordo com a demanda.
- Reduzir o impacto downstream de suas workloads em nuvem: Reduza a quantidade de energia ou recursos necessários para usar seus serviços. Reduza ou elimine a necessidade de os clientes fazerem upgrade de seus dispositivos para usar seus serviços. Teste o uso de farms 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:

- Para eliminar desperdício, baixa utilização e recursos não utilizados ou ociosos
- Para maximizar o valor dos recursos consumidos

Note

Use todos os recursos que você provisiona e realize o mesmo trabalho com o mínimo possível dos recursos necessários.

Nos estágios iniciais da otimização, concentre-se primeiro em áreas com desperdício ou baixa utilização e continue para otimizações mais direcionadas que se ajustem 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 como um processo iterativo que avalia, prioriza, testa e implanta melhorias voltadas para a sustentabilidade de workloads de nuvem.

1. Identificar os alvos para melhoria: analise suas workloads em relação às práticas recomendadas de sustentabilidade descritas neste documento e identifique os alvos para melhoria.
2. Avaliar melhorias específicas: Avalie alterações específicas para melhorar em potencial, custo projetado e risco empresarial.
3. Priorizar e planejar as melhorias: Priorize as alterações que oferecem as maiores melhorias com o menor custo e risco, e estabeleça um plano de testes e de implementação.
4. Testar e validar as melhorias: Implemente as alterações em ambientes de teste para validar seus potenciais de melhoria.

5. Implantar alterações na produção: Implemente as alterações nos ambientes de produção.
6. Medir os resultados e replicar os sucessos: Procure oportunidades de replicar os sucessos nas workloads e reverta as alterações com resultados inaceitáveis.

Cenário de exemplo

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

Sua empresa tem uma workload que executa manipulações de imagens complexas em instâncias do Amazon EC2 e armazena os arquivos modificados e os originais para acesso dos usuários. O processamento das atividades apresenta consumo excessivo de CPU, e os arquivos de saída são extremamente grandes.

Identificar os alvos para melhoria

Compreenda as práticas recomendadas que podem ajudar você a atingir as metas de sustentabilidade. É possível encontrar as descrições dessas [práticas recomendadas](#) e recomendações para melhoria posteriormente neste documento.

Analise suas workloads e os recursos usados. Identificar pontos de acesso, como grandes implantações e recursos usados com frequência. Avalie esses pontos de acesso para obter oportunidades de melhorar a utilização eficaz de seus recursos e reduzir o total de recursos necessários para atingir os resultados de seus negócios.

Analise sua workload em relação às práticas recomendadas e identifique as possibilidades de melhoria.

Ao aplicar esta etapa ao [Cenário de exemplo](#), você identifica as seguintes práticas recomendadas como 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

- [Optimizing your AWS Infrastructure for Sustainability, Part I: Compute \(Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte I: Computação\)](#)
- [Optimizing your AWS Infrastructure for Sustainability, Part II: Storage \(Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte II: Armazenamento\)](#)
- [Optimizing your AWS Infrastructure for Sustainability, Part III: Networking \(Otimizar a sua infraestrutura da AWS para sustentabilidade, Parte III: Redes\)](#)

Avaliar melhorias específicas

Compreenda os recursos provisionados por sua workload para concluir uma unidade de trabalho. Avalie melhorias potenciais e calcule seu impacto potencial, o custo da implementação e os riscos associados.

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

Comece com uma visão geral do seu uso da AWS e use os Relatórios de Custos e Uso da AWS para ajudar a identificar pontos de acesso. Use este [código de exemplo da AWS](#) para ajudar a rever e analisar seu relatório com a ajuda do Amazon Athena.

Métricas de proxy

Ao avaliar alterações específicas, avalie também as métricas que quantificam melhor o efeito dessa alteração no recurso associado. Essas métricas são chamadas de métricas de proxy. Selecione as métricas de proxy que refletem melhor o tipo de melhoria que você está avaliando e os recursos alvos da melhoria. Essas métricas podem evoluir ao longo do tempo.

Os recursos provisionados para comportar a workload incluem computação, armazenamento e recursos de rede. Avalie os recursos provisionados usando suas métricas de proxy para verificar como esses recursos são consumidos.

Use suas métricas de proxy para medir os recursos provisionados para atingir os resultados dos negócios.

Recurso	Métricas de proxy de exemplo	Metas de melhoria
Computação	Minutos de vCPU	Utilização máxima de recursos provisionados
Armazenamento da AWS	GB provisionados GB ou pacotes transferidos	Reduzir o total provisionado Reduzir o total transferido e a distância transferida

Métricas de negócios

Selecione as métricas de negócios para quantificar a aquisição dos resultados dos negócios. As métricas dos negócios devem refletir o valor fornecido pela workload, por exemplo, o número de usuário ativos simultâneos, as chamadas de API atendidas ou o número de transações concluídas. Essas métricas podem evoluir ao longo do tempo. Tenha cuidado ao avaliar métricas financeiras dos negócios, 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 dos negócios obtidos, para determinar os recursos provisionados por unidade de trabalho.

$$\text{Recursos provisionados por unidade de trabalho} = \frac{\text{Métrica de proxy para recurso provisionado}}{\text{Métrica empresarial para o resultado}}$$

Fórmula do KPI

Use os recursos por unidade de trabalho como seus KPIs. Estabeleça padrões baseados nos recursos provisionados para fazer as comparações.

Recurso	KPIs de exemplo	Metas de melhoria
Computação	Minutos de vCPU por transação	Utilização máxima de recursos provisionados

Recurso	KPIs de exemplo	Metas de melhoria
Armazenamento	GB por transação	Reduzir o total provisionado
da AWS	GB transferido por transação ou pacotes transferidos por transação	Reduzir o total transferido e a distância transferida

Calcular a melhoria

Calcule a melhoria como a redução quantitativa dos recursos provisionados (conforme indicado pelas métricas de proxy) e a porcentagem da alteração de seus recursos provisionados de linha de base por unidade de trabalho.

Recurso	KPIs de exemplo	Metas de melhoria
Computação	% de redução de minutos de vCPU por transação	Maximize a utilização
Armazenamento	% de redução de GB por transação	Reduzir o total provisionado
da AWS	% de redução de GB transferido por transação ou pacotes transferidos por transação	Reduzir o total transferido e a distância transferida

Avaliar as melhorias

Avalie as melhorias potenciais em relação ao benefício líquido antecipado. Avalie o tempo, o custo e o nível de esforço da implementação e da manutenção e os riscos dos negócios, como impactos não antecipados.

Com frequência, as melhorias direcionadas representam concessões entre os tipos de recursos consumidos. Por exemplo, para reduzir o consumo de computação, você pode armazenar um resultado ou limitar os dados transferidos e processar os dados antes de enviar o resultado a um cliente. No entanto, [concessões](#) serão discutidas em mais detalhes posteriormente.

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

Ao aplicar esta etapa ao [Cenário de exemplo](#), você avalia as melhorias alvo com os seguintes resultados:

Melhores práticas	Melhoria direcionada	Potencial	Custos	Risco
Use a quantidade e mínima de hardware para atender às suas necessidades	Implemente escalabilidade preditiva para reduzir períodos de baixa utilização	Médio	Baixo	Baixo
Use tecnologias que melhor ofereçam suporte aos seus padrões de acesso e armazenamento de dados	Implemente compactação mais efetiva de mecanismos para reduzir o armazenamento total e o tempo de obtenção dele	Alto	Baixo	Baixo

A implementação de programação preditiva reduz as horas de vCPU consumidas por instâncias subutilizadas ou não utilizadas, fornecendo benefícios moderados sobre os mecanismos de escalabilidade existentes com uma redução estimada de 11% em recursos consumidos. Os custos envolvidos são baixos e incluem a configuração dos recursos de nuvem e a operação de escalabilidade preditiva para o Amazon EC2 Auto Scaling. O risco é performance restringida ao aumentar a escala horizontalmente de forma reativa em resposta a previsões que excedem a demanda.

A implementação de compactação mais efetiva terá um impacto significativo em grandes reduções de tamanho de arquivos em todas as suas imagens originais e manipuladas, com uma redução

estimada em 25% nos requisitos de armazenamento na produção. A implementação do novo algoritmo é uma substituição de pouco esforço com pouco risco envolvido.

Priorizar e planejar as melhorias

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

Decida em quais melhorias focalizar inicialmente e inclua-as no planejamento de recursos e no roteiro de desenvolvimento.

Ao aplicar esta etapa ao [Cenário de exemplo](#), você prioriza as melhorias alvo da seguinte maneira:

Prioridade	no KPI	Potencial	Custos	Risco
1	Implementar compactação mais efetiva para mecanismos	Alto	Baixo	Baixo
2	Implementar a escalabilidade preditiva	Médio	Baixo	Baixo

O alto potencial, o baixo custo e o risco de atualização da compactação de arquivos a tornam um alvo de alto valor para a empresa e uma prioridade sobre a implementação de escalabilidade preditiva. Determine se a implementação de escalabilidade preditiva com seu impacto de potencial médio, baixo custo e baixo risco deve ser a melhoria prioritária após a conclusão da compactação de arquivos.

Atribua um membro da equipe para implementar a compactação de arquivos melhorada e adicione a escalabilidade preditiva à sua lista de pendências.

Testar e validar as melhorias

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

Implemente uma cópia representativa de sua workload no ambiente de testes para limitar o custo e o risco da execução de testes e de 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 testes.

Implemente a melhoria alvo no ambiente de testes e repita o teste usando a mesma metodologia sob as mesmas condições. Meça os recursos provisionados e os usados por unidade de trabalho com a melhoria implementada.

Calcule a porcentagem de alteração na sua linha de base dos recursos provisionados por unidade de trabalho e determine a redução quantitativa esperada nos recursos provisionados no ambiente de produção. Compare esses valores em relação aos valores previstos. Determine se o resultado tem um nível aceitável de melhoria. Avalie se quaisquer concessões em recursos adicionais consumidos tornam o benefício líquido da melhoria inaceitável.

Determine se a melhoria é bem-sucedida e se devem ser investidos recursos na implementação da alteração na produção. Se a alteração for avaliada como bem-sucedida desta vez, redirecione seus recursos para testar e validar seu próximo alvo e continuar com o ciclo de melhorias.

% de redução em recursos provisionados por unidade de trabalho	Redução quantitativa em recursos provisionados	Action (Ação)
Expectativas atendidas	Expectativas atendidas	Prosseguir com a melhoria
Expectativas não atendidas	Expectativas atendidas	Prosseguir com a melhoria
Expectativas atendidas	Expectativas não atendidas	Buscar melhoria alternativa
Expectativas não atendidas	Expectativas não atendidas	Buscar melhoria alternativa

Ao aplicar esta etapa ao [Cenário de exemplo](#), você realiza testes para validar o sucesso.

Depois de realizar os testes no algoritmo de compactação melhorado, a porcentagem de redução dos recursos provisionados por unidade de trabalho (o armazenamento requerido para a imagem original e a imagem modificada) atendeu às expectativas com uma média de 30% de redução do armazenamento provisionado e um aumento insignificante da carga de computação.

Determine se os recursos de computação adicionais necessários para aplicar o algoritmo de compactação aos arquivos existentes na produção é insignificante em comparação com a redução atingida no armazenamento. Você confirmou sucesso com a redução quantitativa nos recursos necessários (TBs de armazenamento), e a melhoria é aprovada para implantação na produção.

Implantar alterações na produção

Implemente as melhorias testadas, validadas e aprovadas na produção. Implemente usando implantações limitadas, confirme a funcionalidade da sua workload, teste a redução real nos recursos provisionados e consumidos por unidade de trabalho dentro da implantação limitada e verifique se há consequências não intencionais por causa da alteração. Prossiga com implantações completas após testar com sucesso.

Reverta as alterações, se os testes falharem ou se houver consequências inaceitáveis e não intencionais por causa da alteração.

Ao aplicar esta etapa ao [Cenário de exemplo](#), execute as seguintes ações:

Implemente as alterações na produção usando uma implantação limitada por meio da metodologia azul-verde de implantação. Os testes de funcionalidade nas instâncias recém-implantadas são bem-sucedidos. Você vê uma redução média de 26% no armazenamento provisionado de arquivos de imagens originais e manipuladas. Não há nenhuma evidência de aumento na carga de computação dos novos arquivos compactados.

Você observa uma redução não prevista no tempo transcorrido para compactar os arquivos de imagem, e atribui isso ao código altamente otimizado do novo algoritmo de compactação.

Prossiga com a implantação completa da nova versão.

Medir os resultados e replicar os sucessos

Meça os resultados e replique os sucessos das seguintes maneiras:

- Meça a melhoria inicial aos recursos provisionados por unidade de trabalho e a redução quantitativa nos recursos provisionados.
- Compare as estimativas iniciais e os resultados dos testes nas medições da produção. Identifique os fatores que podem ter contribuído para as diferenças e atualize sua estimativa e as metodologias de testes quando apropriado.

- Determine os sucessos e o grau dos sucessos e compartilhe os resultados com as partes interessadas.
- Se for necessário reverter as alterações devido a testes com falha ou com consequências não intencionais negativas, identifique os fatores contribuintes. Itere quando viável ou avalie novas abordagens para atingir as metas da alteração.
- Use o que aprendeu, estabeleça padrões e aplique melhorias bem-sucedidas a outros sistemas que possam se beneficiar com a mesma alteração. Capture e compartilhe sua metodologia, os artefatos relacionados e os benefícios líquidos, entre as equipes e as organizações para que seu sucesso possa ser adotado e replicado para por elas.
- Monitore os recursos provisionados por unidade de trabalho e rastreie as alterações e o impacto total ao longo do tempo. As alterações nas workloads ou a maneira como os cliente consomem sua workload podem ter impacto na efetividade de sua melhoria. Reavalie as oportunidades de melhorias ao observar reduções de curto prazo significativas na efetividade de sua melhoria ou uma redução acumulada na efetividade ao longo do tempo.
- Quantifique o benefício líquido da sua melhoria ao longo do tempo (incluindo os benefícios recebidos por outras equipes que a aplicaram, se disponíveis) para mostrar o retorno sobre o investimento das atividades da melhoria.

Ao aplicar esta etapa ao [Cenário de exemplo](#), você mede os seguintes resultados.

Sua workload mostra uma melhoria inicial de 23% de redução dos requisitos de armazenamento depois da implantação e da aplicação do novo algoritmo de compactação a arquivos de imagens existentes.

O valor medido está amplamente de acordo com as estimativas iniciais (25%), e a diferença significativa comparada aos testes (30%) é resultado dos arquivos de imagens usados nos testes não serem representativos dos arquivos de imagens presentes na produção. Modifique o conjunto de imagens dos testes para refletir as imagens na produção de maneira mais apropriada.

A melhoria é considerada um sucesso total. A redução total no armazenamento provisionado é 2% menor que os 25% estimados, mas 23% ainda são uma enorme melhoria do impacto na sustentabilidade, acompanhados por uma economia de custos equivalente.

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

Estabeleça um projeto de código aberto interno onde você compartilha seu código, os artefatos associados, as orientações de como implementar a alteração e os resultados de sua implementação. O projeto de código aberto interno facilita a adoção do código pelas suas equipes para todos os casos de uso de armazenamento de arquivos persistentes. Suas equipes adotam a melhoria como um padrão. Os benefícios secundários do projeto de código aberto interno são que todos que adotam a solução se beneficiam das melhorias na solução, e todos podem contribuir com melhorias para o projeto.

Publique seu sucesso e compartilhe o projeto de código aberto em toda a organização. Cada equipe que adota a solução replica o benefício com investimento mínimo e aumenta o benefício líquido recebido do investimento que você fez. Publique esses dados como uma história de sucesso contínua.

Continue a monitorar o impacto da melhoria ao longo do tempo e faça alterações no projeto de código aberto interno conforme necessário.

Sustentabilidade como um requisito não funcional

A adição da sustentabilidade à sua lista de requisitos empresariais pode resultar em soluções mais econômicas. O foco na obtenção de mais valor dos recursos usados e de usar menos recursos é diretamente convertido em economias de custos na AWS pois você paga apenas pelo que usa.

Atender às metas da sustentabilidade pode não exigir concessões equivalentes em uma ou mais das outras métricas tradicionais, como tempo de atividade, disponibilidade ou tempo de resposta. É possível obter ganhos significativos na sustentabilidade sem nenhum impacto mensurável nos níveis de serviço. Quando concessões secundárias são necessárias, as melhorias na sustentabilidade obtidas por essas concessões podem superar a alteração na qualidade do serviço.

Incentive os membros da equipe a fazer experimentos continuamente com melhorias na sustentabilidade ao desenvolverem 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.

Os itens a seguir são concessões de exemplo que podem reduzir os recursos de nuvem que você consome.

Ajuste a qualidade do resultado: É possível trocar a qualidade dos resultados (QoR) por uma redução na intensidade da workload com computação aproximada. A prática da computação aproximada procura oportunidades para explorar a lacuna entre o que os clientes precisam e o que você realmente produz. Por exemplo, se você colocar os dados em uma estrutura de dados definida, poderá descartar o operador ORDER BY no SQL para remover processamento desnecessário, economizando recursos e ainda produzindo uma resposta aceitável.

Ajustar o tempo de resposta: Uma resposta com um tempo de resposta mais lento pode reduzir a emissão de carbono minimizando a sobrecarga compartilhada. O processamento de tarefas ad hoc e temporárias pode incorrer em sobrecarga de inicialização. Agrupe e processe as tarefas em lotes, em vez de pagar pela sobrecarga sempre que uma tarefa for recebida. O processamento em lote troca o tempo de resposta maior por uma redução na sobrecarga compartilhada do desenvolvimento de uma instância, do download do código-fonte e da execução do processo.

Ajustar a disponibilidade: Com a AWS, é possível adicionar redundância facilmente e atender às metas de alta disponibilidade com apenas alguns cliques. É possível aumentar a redundância por meio de técnicas, como a estabilidade estática, provisionando recursos ociosos que sempre resultam em utilização reduzida. Avalie as necessidades da empresa ao definir alvos. Concessões

relativamente menores na disponibilidade podem resultar em melhorias muito maiores na utilização. Por exemplo, um padrão de arquitetura comum envolve o provisionamento da capacidade de failover ocioso para tratar da carga imediatamente após uma falha de componente. Reduzir o requisito de disponibilidade pode remover a necessidade de capacidade on-line ociosa, permitindo que a automação implante os recursos. A adição da capacidade de failover sob demanda aumenta a utilização geral sem afetar os negócios durante as operações normais e tem o benefício secundário de reduzir os custos.

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

Seleção de região

A escolha da região para sua workload afeta significativamente seus KPIs, incluindo desempenho, 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 desempenho, custo e pegada de carbono.

Antipadrões comuns:

- Selecione a região da workload com base em sua localização.
- Você consolida todos os recursos da workload em uma única localização geográfica.

Benefícios de estabelecer esta prática recomendada: Colocar uma workload próxima a projetos de energia renovável da Amazon ou 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 é estabelecida: médio

Orientações para a implementação

O Nuvem AWS é uma rede em constante expansão de regiões e pontos de presença (PoP), com uma infraestrutura de rede global que os conecta. A escolha da região para sua workload afeta significativamente seus KPIs, incluindo desempenho, 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 da implementação

- Siga estas etapas para avaliar e selecionar possíveis regiões para sua workload com base em seus requisitos de negócios, incluindo conformidade, recursos disponíveis, custo e latência:
 - Confirme se essas regiões estão em conformidade, com base nos regulamentos locais exigidos.
 - Use as [Listas de serviços regionais do 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 Pricing Calculator](#).
 - Teste a latência de rede entre as localizações de seus usuários finais e cada Região da AWS.
- 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 [GHG Protocol](#) (métodos baseados no mercado e baseados na localização).
 - Escolha a região com base no método que você usa para rastrear as emissões de carbono. Para obter mais detalhes sobre como escolher uma região com base em suas diretrizes de sustentabilidade, consulte [Como selecionar uma região para sua workload com base nas metas de sustentabilidade](#).

Recursos

Documentos relacionados:

- [Compreensão de suas estimativas de emissão de carbono](#)
- [Amazon em todo o mundo](#)
- [Metodologia de energia renovável](#)
- [O que considerar ao selecionar uma região para suas workloads](#)

Vídeos relacionados:

- [Arquitetura sustentável e redução de sua pegada de carbono do 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 as 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 com as 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.

Antipadrões comuns:

- Você não dimensiona sua infraestrutura de acordo com a carga de usuários.
- Você dimensiona sua infraestrutura manualmente o tempo todo.
- Você deixa a capacidade aumentada após um evento de escalabilidade, em vez de reduzir novamente.

Benefícios do estabelecimento dessa prática recomendada: configurar e testar a elasticidade da workload ajuda a corresponder de maneira eficiente a oferta de recursos de nuvem à demanda e evitar a capacidade superprovisionada. Você pode 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 é estabelecida: médio

Orientações para a implementação

A nuvem fornece a flexibilidade de 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. Os aplicativos podem aumentar ou diminuir a escala verticalmente ao modificar o tamanho da instância e horizontalmente ao modificar o número de instâncias, ou uma combinação de ambos.

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

- Abordagem de monitoramento de meta: monitore sua métrica de escalabilidade e aumente ou diminua automaticamente a capacidade conforme necessário.
- Escalabilidade preditiva: escale antecipadamente em relação às tendências diárias e semanais.
- Abordagem com base na programação: defina sua própria programação de escalabilidade de acordo com as alterações de carga previsíveis.

- Escalabilidade de serviços: escolha serviços (como tecnologia sem servidor) que são escalados nativamente por design ou fornecem escalabilidade automática 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 da 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 a escalabilidade automática ou como um recurso do serviço. A AWS fornece uma variedade de mecanismos de escalabilidade automática para garantir que as workloads possam reduzir a escala verticalmente de forma rápida e fácil durante períodos de baixa carga de usuário. Veja alguns exemplos de mecanismos de escalabilidade automática:

Auto scaling mechanism	Where to use
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 o seu aplicativo.
Application Auto Scaling	Use para escalar automaticamente os recursos para serviços individuais da AWS além do Amazon EC2, como funções do Lambda ou serviços do Amazon Elastic Container Service (Amazon ECS).
o dimensionador automático de cluster do Kubernetes	Use para escalar automaticamente os clusters do Kubernetes na AWS.

- A escalabilidade geralmente é discutida em relação a serviços de computação, como instâncias do Amazon EC2 ou funções do AWS Lambda. Considere a configuração de serviços não relacionados a computação, como o [Amazon DynamoDB](#), e grave unidades de capacidade ou fragmentos do [Amazon Kinesis Data Streams](#) para corresponder à demanda.
- Verifique se as métricas para aumentar ou reduzir a escala verticalmente são validadas em relação ao tipo de 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. Você pode usar uma [métrica personalizada](#) (como utilização de memória) para a política de escalabilidade, 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 ou diminuir proporcionalmente com o número de instâncias no grupo do Auto Scaling.
- Use a [escalabilidade dinâmica](#) em vez da [escalabilidade manual](#) para o seu grupo do Auto Scaling. Também recomendamos que você use as [políticas de escalabilidade de monitoramento de meta](#) na sua escalabilidade dinâmica.
- 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). Você também pode usar o [histórico de atividades](#) para verificar a atividade de escalabilidade para um grupo do Auto Scaling.
- Avalie sua workload com relação a padrões previsíveis e, ao antecipar alterações previstas e planejadas na demanda, escale proativamente. Com a escalabilidade preditiva, é possível eliminar a necessidade de superprovisionar a capacidade. Para obter mais detalhes, consulte [Escalabilidade preditiva com o Amazon EC2 Auto Scaling](#).

Recursos

Documentos relacionados:

- [Conceitos básicos do Amazon EC2 Auto Scaling](#)
- [Escalabilidade preditiva para o EC2 com Machine Learning](#)
- [Analisar o comportamento dos usuários usando o Amazon OpenSearch Service, o Amazon Data Firehose e o Kibana](#)
- [O que é o Amazon CloudWatch?](#)
- [Monitorar a carga do banco de dados com o Performance Insights no Amazon RDS](#)
- [Introdução ao suporte nativo para escalabilidade preditiva com o Amazon EC2 Auto Scaling](#)
- [Apresentando o Karpenter: um dimensionador automático de clusters do Kubernetes de código aberto e alta performance](#)
- [Aprofundamento do Amazon ECS Cluster Auto Scaling](#)

Vídeos relacionados:

- [Build a cost-, energy-, and resource-efficient compute environment](#) (Criar um ambiente de computação eficiente em termos de custo, energia e recursos)
- [Better, faster, cheaper compute: Cost-optimizing Amazon EC2 \(CMP202-R1\)](#) (Computação melhor, mais rápida e mais barata: otimização de custos com o Amazon EC2)

Exemplos relacionados:

- [Laboratório: Exemplos de grupos do Amazon EC2 Auto Scaling](#)
- [Laboratório: Implementação de escalabilidade automática com o Karpenter](#)

SUS02-BP02 Alinhar os SLAs com as 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.

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: o alinhamento dos SLAs com as metas de sustentabilidade ocasiona o uso ideal dos recursos e a concretização das necessidades empresariais.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: baixo

Orientações para a 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 de 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 da implementação

- Defina ou remodele SLAs que apoiem suas metas de sustentabilidade e, ao mesmo tempo, atendam aos seus requisitos empresariais, não os excedendo.
- Faça 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: o uso de mais recursos para impulsionar a performance pode causar um maior impacto ambiental.
 - Sustentabilidade e segurança: workloads excessivamente seguras podem ter um maior impacto ambiental.
- 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.

Recursos

Documentos relacionados:

- [Acordos de nível de serviço da AWS \(SLAs\)](#)
- [Importance of Service Level Agreement for SaaS Providers](#)

Vídeos relacionados:

- [Delivering sustainable, high-performing architectures](#) (Entregar arquiteturas sustentáveis e de alta performance)
- [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.

Antipadrões comuns:

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

Benefícios do estabelecimento desta prática recomendada: a remoção de ativos ociosos libera recursos e melhora a eficiência geral da workload.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: baixo

Orientações para a 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 da implementação

- Use ferramentas de monitoramento para identificar ativos estáticos que não são mais necessários.
- Antes de remover qualquer ativo, avalie o impacto da remoção sobre a arquitetura.
- Desenvolva um plano e remova os ativos que não são mais necessários.
- Consolidar ativos gerados sobrepostos para remover o processamento redundante.
- Atualize suas aplicações para que não produzam nem armazenem mais ativos que não são necessários.
- Instrua terceiros a interromper a produção e o armazenamento de ativos gerenciados em seu nome que não sejam mais necessários.
- Instrua terceiros a consolidar ativos redundantes produzidos em seu nome.
- Avalie 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 verificar se há recursos ativos dos quais não preciso mais na minha Conta da AWS?](#)

Vídeos relacionados:

- [Como verifico e depois removo recursos ativos dos quais não preciso mais na minha Conta da AWS?](#)

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.

Antipadrões comuns:

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

Benefícios de estabelecer 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 datacenters 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 da implementação

- Analise os padrões de acesso à rede em sua workload para identificar como os usuários utilizam sua aplicação.

- Use ferramentas de monitoramento, como [Amazon CloudWatch](#) e o [AWS CloudTrail](#), para coletar dados sobre as atividades da 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 região](#).
 - A localização dos seus dados: 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.
 - A localização dos usuários: para aplicações voltadas ao usuário, escolha uma região (ou regiões) próxima dos clientes de sua workload.
 - Outras restrições: Leve em conta restrições, como custo e conformidade, conforme explicado em [O que considerar ao selecionar uma região para suas workloads](#).
- Use armazenamento em cache local ou [soluções de armazenamento em cache da AWS](#) para ativos usados com frequência a fim de aumentar a performance, reduzir a movimentação de dados e reduzir o impacto ambiental.

Service	Quando usar
Amazon CloudFront	Use para armazenar conteúdo estático em cache, como imagens, scripts e vídeos, bem como 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:

Service	Quando usar
o Lambda@Edge	Use para operações com uso intenso de computação que são iniciadas quando objetos não estão no cache.
Funções do Amazon CloudFront	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](#)

Vídeos relacionados:

- [Demystifying data transfer on AWS \(Desmistificação da transferência de dados na AWS\)](#)
- [Escalar a performance da rede em instâncias do Amazon EC2 de última geração](#)

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.

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: otimizar recursos para os membros da equipe melhora a eficiência geral das aplicações habilitadas para a nuvem.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: baixo

Orientações para a 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 da implementação

- Provisione estações de trabalho e outros dispositivos para alinhar a maneira como eles são usados.
- Use áreas de trabalho virtuais e a transmissão de aplicações para limitar os requisitos de upgrade e dispositivos.
- Migre para a nuvem as tarefas do processador e as com uso intenso de memória a fim de utilizar a respectiva elasticidade.

- 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 empresariais.
- 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 (UI) que ajuda você a gerenciar remotamente os nós em execução na AWS ou no ambiente on-premises.

Recursos

Documentos relacionados:

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

Vídeos relacionados:

- [Gerenciamento de 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.

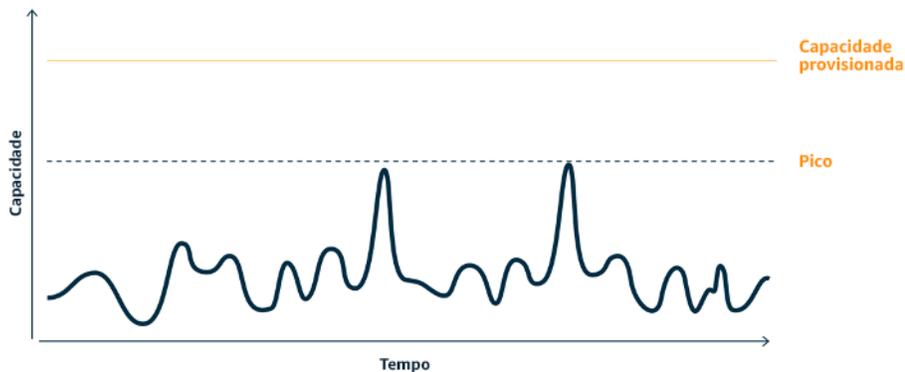
Antipadrões comuns:

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

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

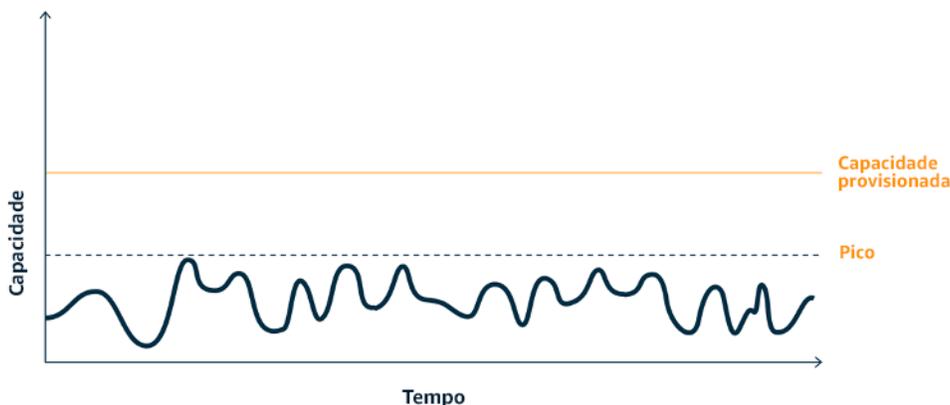
Nível de exposição a riscos quando esta prática recomendada não é estabelecida: baixo

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 da 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.



Efeito do controle de utilização sobre a curva da demanda e a capacidade provisionada.

Etapas da 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 pode 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 vários 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](#)
- [Integração de aplicações usando filas e mensagens](#)

Vídeos relacionados:

- [Escolha do serviço de mensagem correto para sua aplicação distribuída](#)

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 programados](#)
- [SUS03-BP02 Remover ou refatorar componentes da workload subutilizados ou não utilizados](#)
- [SUS03-BP03 Otimizar as áreas de código que consomem mais tempo ou recursos](#)
- [SUS03-BP04 Otimizar o impacto sobre dispositivos e equipamentos](#)
- [SUS03-BP05 Usar arquiteturas e padrões de software que atendam melhor aos padrões de armazenamento e acesso aos dados](#)

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

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

Antipadrões comuns:

- 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 do estabelecimento desta 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 exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a implementação

Use padrões de arquitetura eficientes, como [arquitetura orientada por eventos](#), que resultam na utilização uniforme de componentes e minimizam o provisionamento em excesso 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 da implementação

- Analise a demanda de 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 escalonamento automático para maximizar a utilização. Aqui estão alguns exemplos de quando você pode considerar a arquitetura orientada por filas:

Queuing mechanism	Description
Filas de trabalho do AWS Batch	Os trabalhos do AWS Batch são enviados para uma fila de trabalhos onde permanece até que possam ser agendados para execução em um ambiente de computação.
Instâncias spot do Amazon Simple Queue Service e Amazon EC2	Emparelhamento de 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:

Scheduling mechanism	Description
Agendador do Amazon EventBridge	A capacidade do Amazon EventBridge que permite criar, executar e gerenciar tarefas agendadas em escala.
Programação baseada em tempo do AWS Glue	Defina uma programação baseada em tempo para seus crawlers e trabalhos no AWS Glue.
Tarefas agendadas do Amazon Elastic Container Service (Amazon ECS)	O Amazon ECS permite a criação de tarefas agendadas. Tarefas agendadas usam regras do Amazon EventBridge para executar tarefas em um agendamento ou em resposta a um evento do EventBridge.
Programador de Instâncias	Configure programações de início e parada para suas instâncias do Amazon EC2 e Amazon Relational Database Service.

- Se você usar mecanismos de pesquisa e webhooks em sua arquitetura, substitua-os por eventos. Use [arquiteturas orientadas por eventos](#) para criar workloads altamente eficientes.
- Aproveite a [a computação sem servidor no AWS](#) para eliminar a infraestrutura provisionada em excesso.
- Dimensione corretamente componentes individuais de sua arquitetura para evitar recursos ociosos aguardando entrada.

Recursos

Documentos relacionados:

- [O que é o Amazon Simple Queue Service?](#)
- [O que é o Amazon MQ?](#)
- [Escalabilidade baseada no Amazon SQS](#)
- [O que é o AWS Step Functions?](#)
- [O que é o AWS Lambda?](#)
- [Usando o AWS Lambda com o Amazon SQS](#)

- [O que é o Amazon EventBridge?](#)

Vídeos relacionados:

- [Moving to event-driven architectures \(Mudando para arquiteturas orientadas a eventos\)](#)

SUS03-BP02 Remover ou refatorar componentes da workload subutilizados ou não utilizados

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

Antipadrões comuns:

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

Benefícios do estabelecimento desta prática recomendada: a remoção de ativos ociosos minimiza o desperdício e melhorar a eficiência geral da workload de nuvem.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a implementação

Avalie sua workload para identificar componentes ociosos ou não utilizados. Esse é um processo de melhoria iterativo que pode ser acionado 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 ser uma indicação de que é necessário 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 da implementação

- Monitore e capture métricas de utilização de componentes essenciais de sua workload (como utilização de CPU, utilização de memória ou throughput de rede nas [métricas do Amazon CloudWatch](#)).
- Para workloads estáveis, confira regularmente as ferramentas de dimensionamento correto da AWS, como o [AWS Compute Optimizer](#), para identificar componentes ociosos, não usados ou subutilizados.
- Para workloads efêmeras, avalie as métricas de utilização para identificar componentes ociosos, não usados ou subutilizados.
- Retire componentes e ativos associados (como imagens do Amazon ECR) que não são mais necessários.
- Refatore ou consolide os componentes subutilizados com outros recursos para melhorar a eficiência da utilização. Por exemplo, você pode 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.
- Saiba quais [recursos são provisionados por sua workload para concluir uma unidade de trabalho](#).

Recursos

Documentos relacionados:

- [AWS Trusted Advisor](#)
- [O que é o Amazon CloudWatch?](#)
- [Limpeza automatizada de imagens não utilizadas no Amazon ECR](#)

Exemplos relacionados:

- [Laboratório do Well-Architected: Dimensionamento correto com o AWS Compute Optimizer](#)
- [Laboratório do Well-Architected: Otimizar padrões de hardware e observar KPIs de sustentabilidade](#)

SUS03-BP03 Otimizar as 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.

Antipadrões comuns:

- 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 estabelecer esta prática recomendada: O uso de código eficiente 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 da implementação

- Ao desenvolver suas workloads, adote um processo de revisão de código automatizada para melhorar a qualidade e identificar erros e antipadrões.
 - [Análises de código automatizadas com o Amazon CodeGuru Reviewer](#)
 - [Detecção de erros simultâneos com o Amazon CodeGuru](#)
 - [Elevação da qualidade do código para aplicações Python usando o Amazon CodeGuru](#)
- À medida que você executa suas workloads, monitore os recursos para identificar componentes com altos requisitos de recurso por unidade de trabalho como alvos para revisões de código.
- Para revisões 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 sobre o 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](#)

- Use a linguagem de programação e o sistema operacional mais eficientes para a workload. Para obter detalhes sobre linguagens de programação com eficiência energética (incluindo Rust), consulte [Sustentabilidade com Rust](#).
- 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.

Recursos

Documentos relacionados:

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

Vídeos relacionados:

- [Improve Code Efficiency Using Amazon CodeGuru Profiler \(Como melhorar a eficiência do código usando o Amazon CodeGuru Profiler\)](#)
- [Automate Code Reviews and Application Performance Recommendations with Amazon CodeGuru \(Como automatizar revisões de código e recomendações de performance de aplicação com o Amazon CodeGuru\)](#)

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 de nuvem.

Antipadrões comuns:

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

Benefícios do estabelecimento desta prática recomendada: implementar padrões e recursos de software que são otimizados para o dispositivo do cliente pode reduzir o impacto ambiental de modo geral da workload de nuvem.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a 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 da implementação

- 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.
- Otimize a aplicação executada 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, você pode usar o [Amazon Elastic Transcoder](#) ou o [AWS Elemental MediaConvert](#) para converter arquivos de mídia digital grandes e de alta qualidade em formatos que os usuários possam reproduzir em dispositivos móveis, tablets, navegadores Web e televisores conectados.
 - 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.
- Use um mecanismo sem fio automatizado para implantar atualizações em um ou mais dispositivos.
 - Você pode usar um [pipeline de CI/CD](#) para atualizar aplicativos móveis.
 - Você pode usar o [AWS IoT Device Management](#) para gerenciar remotamente dispositivos conectados em escala.
- 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-BP04 Usar parques de dispositivos gerenciados para testes](#).

Recursos

Documentos relacionados:

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

Vídeos relacionados:

- [Introdução ao AWS Device Farm](#)

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

Antipadrões comuns:

- Você pressupõe que todas as workloads têm 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 do estabelecimento desta 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 a aumentar a utilização de modo 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 exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a implementação

Use padrões de software e arquitetura que melhor se alinhem às características dos dados e aos padrões de acesso. Por exemplo, use uma [arquitetura de dados moderna na AWS](#) que permita que você use serviços com propósito específico otimizados para seus casos de uso exclusivos de análise. Esses padrões de arquitetura possibilitam um processamento de dados eficiente e reduzem o uso de recursos.

Etapas da implementação

- Analise as características dos dados e os padrões de acesso para identificar a configuração correta para seus recursos de nuvem. Principais características a serem consideradas:
 - Tipo de dados: estruturados, semiestruturados e não estruturados
 - Crescimento dos dados: delimitado, não delimitado
 - Durabilidade dos dados: persistentes, efêmeros, transitórios
 - Padrões de acesso: leituras ou gravações, frequência de atualização, com picos ou consistente
- Use padrões de arquitetura que comportem melhor os padrões de armazenamento e acesso aos dados.
 - [Vamos arquitetar! Arquiteturas de dados modernas](#)

- [Bancos de dados na AWS: a ferramenta certa para o trabalho certo](#)
- Use tecnologias que funcionam nativamente com dados compactados.
- Use [serviços de análise](#) com propósito específico para processamento de dados em sua arquitetura.
- Use o mecanismo de banco de dados que melhor comporta seu padrão de consulta dominante. Gerencie seus índices de bancos de dados para garantir a execução eficiente de consultas. Para ter mais detalhes, consulte [Bancos de dados da AWS](#).
- Escolha protocolos de rede que reduzam a quantidade de capacidade de rede consumida em sua arquitetura.

Recursos

Documentos relacionados:

- [Suporte a compactação no Athena](#)
- [COPY de formatos de dados colunar com o Amazon Redshift](#)
- [Converter o formato de registro de entrada no Firehose](#)
- [Opções de formato de dados para entradas e saídas no AWS Glue](#)
- [Melhorar a performance de consultas no Amazon Athena com a conversão em formatos colunares](#)
- [Carregar arquivos de dados compactados do Amazon S3 com o Amazon Redshift](#)
- [Monitorar a carga de banco de dados com o Performance Insights no Amazon Aurora](#)
- [Monitorar a carga de banco de dados com o Performance Insights no Amazon RDS](#)
- [Classe de armazenamento do Amazon S3 Intelligent-Tiering](#)

Vídeos relacionados:

- [Criar arquiteturas modernas de dados 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 diminuïrem, 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 sistemas de arquivos compartilhados ou armazenamento para acessar dados comuns](#)
- [SUS04-BP07 Minimizar a movimentação de dados entre redes](#)
- [SUS04-BP08 Fazer backup de dados somente quando for difícil recriar](#)

SUS04-BP01 Implementar uma política de classificação de dados

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

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: a implementação de uma política de classificação de dados permite determinar o nível de armazenamento eficiente em termos de energia para os dados.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a 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. 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 da implementação

- Realize um inventário dos vários tipos de dados que existem para sua workload.
 - Para obter mais detalhes sobre categorias de classificação de dados, consulte o [whitepaper Data Classification](#) (Classificação de 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.
 - Por exemplo, consulte [Four simple steps to classify your data and secure your startup](#) (Quatro etapas simples para classificar seus dados e proteger sua startup).
- Audite periodicamente seu ambiente em busca de dados que não estejam etiquetados ou classificados e classifique-os e etiquete-os apropriadamente.
 - Por exemplo, consulte [Data Catalog e crawlers no AWS Glue](#).
- Estabeleça um catálogo de dados que forneça recursos de auditoria e governança.
- Determine e documente procedimentos de manipulação para cada classe de dados.
- Use automação para auditar periodicamente seu ambiente em busca de dados que não estejam etiquetados ou classificados e classifique-os e etiquete-os apropriadamente.

Recursos

Documentos relacionados:

- [Leveraging Nuvem AWS to Support Data Classification](#) (Utilizar a Nuvem AWS para oferecer compatibilidade com a classificação de dados)
- [Políticas de tag do AWS Organizations](#)

Vídeos relacionados:

- [Enabling agility with data governance on AWS](#) (Ativar a agilidade com governança de dados 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.

Antipadrões comuns:

- Você pressupõe que todas as workloads têm 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 estabelecer 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.

- 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:

- Tipo de dados: estruturados, semiestruturados e não estruturados
- Crescimento dos dados: delimitado, não vinculado
- Durabilidade de dados: persistente, efêmero, transitório
- Padrões de acesso: leituras ou gravações, frequência, picos ou consistentes
- 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 a objetos dentro e fora do Amazon S3 podem usar um serviço, como Aceleração de Transferências ou Pontos de Acesso , para oferecer compatibilidade com seu local, 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 diversos tipos de soluções de computação. O Amazon EFS aumenta e reduz automaticamente

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

Tipo	Tecnologia	Características principais
Banco de dados relacional	Amazon Aurora , o Amazon RDS , o 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	tabelas do 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.

- Para sistemas de armazenamento que têm tamanho fixo, como Amazon EBS ou Amazon FSx, monitore o espaço de armazenamento disponível e automatize a alocação de armazenamento ao atingir um limite. Você pode utilizar o Amazon CloudWatch para coletar e analisar diferentes métricas para o [Amazon EBS](#) e o [Amazon FSx](#).
- As classes de armazenamento do Amazon S3 podem ser configuradas em nível de objeto e um único bucket pode conter objetos armazenados em todas as classes de armazenamento.

- Você também pode 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, você precisa 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](#)
- [Uso de classes de armazenamento do Amazon S3](#)
- [O que é o Amazon S3 Glacier?](#)

Vídeos relacionados:

- [Architectural Patterns for Data Lakes on AWS \(Padrões arquitetônicos para data lakes na AWS\)](#)
- [Deep dive on Amazon EBS \(STG303-R1\)](#)
- [Optimize your storage performance with Amazon S3 \(STG343\)](#)
- [Building modern data architectures on AWS \(Criação de arquiteturas de dados na AWS\)](#)

Exemplos relacionados:

- [Driver CSI do Amazon EFS](#)
- [Driver CSI do Amazon EBS](#)
- [Utilitários do Amazon EFS](#)
- [Escalabilidade automática do Amazon EBS](#)
- [Exemplos do Amazon S3](#)

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.

Antipadrões comuns:

- 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 esta prática recomendada: O uso de políticas de ciclo de vida de dados garante acesso e retenção de dados eficientes em uma workload.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a implementação

Os conjuntos de dados geralmente têm diferentes requisitos de retenção e acesso durante seu ciclo de vida. Por exemplo, seu aplicativo 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 gerenciar com eficiência seus conjuntos de dados ao longo de seu ciclo de vida, configure políticas de ciclo de vida, que são regras que definem como lidar com conjuntos de dados.

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.

Etapas da implementação

- [Classifique conjuntos de dados em sua workload.](#)
- Defina procedimentos de manipulação para cada classe de dados.
- 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:

Storage service	How to set automated lifecycle policies
Amazon S3	<p>Você pode usar o ciclo de vida do Amazon S3 para gerenciar seus objetos durante todo o ciclo de vida. Se seus padrões de acesso forem desconhecidos, mutáveis ou imprevisíveis, você pode usar o Amazon S3 Intelligent-Tiering, que monitora os padrões de acesso e move automaticamente os objetos que não foram acessados para níveis de acesso de custo mais baixo. Você pode aproveitar as métricas da Lente de Armazenamento do Amazon S3 para identificar oportunidades de otimização e lacunas no gerenciamento do ciclo de vida.</p>
Amazon Elastic Block Store	<p>Você pode usar o Amazon Data Lifecycle Manager para automatizar a criação, retenção e exclusão de snapshots do Amazon EBS e AMIs com suporte do Amazon EBS.</p>
Amazon Elastic File System	<p>O gerenciamento do ciclo de vida do Amazon EFS gerencia automaticamente o armazenamento de arquivos para seus sistemas de arquivos.</p>
Amazon Elastic Container Registry	<p>As políticas de ciclo de vida do Amazon ECR automatizam a limpeza de suas imagens de contêiner, expirando imagens com base na idade ou contagem.</p>
AWS Elemental MediaStore	<p>Você pode usar uma política de ciclo de vida do objeto que rege por quanto tempo os objetos devem ser armazenados no contêiner do MediaStore.</p>

- Exclua volumes não utilizados, snapshots e dados que estão fora do período de retenção. Aproveite os recursos do serviço nativo, como o tempo de vida útil do Amazon DynamoDB ou retenção de log do Amazon CloudWatch para exclusão.
- Agregue e compacte dados quando possível com base nas regras do ciclo de vida.

Recursos

Documentos relacionados:

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

Vídeos relacionados:

- [Simplifique o ciclo de vida de seus dados e otimize os custos de armazenamento com o ciclo de vida do Amazon S3](#)
- [Reduza seus custos de armazenamento usando 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 aumentarem e minimizar o total de armazenamento provisionado.

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: minimizar o provisionamento em excesso do sistema de armazenamento reduz os recursos ociosos e melhora a eficiência geral da workload.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a 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 da implementação

- Para armazenamento fixo como o [Amazon EBS](#), verifique se está monitorando a quantidade de armazenamento usada em comparação com o tamanho geral do armazenamento e, se possível, crie 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. A título de exemplo, você pode usar os [Volumes Elásticos do Amazon EBS](#) para alterar o tamanho ou o tipo de volume ou ajustar a performance dos 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.
 - [Desempenho do Amazon EFS](#)
 - [Performance do volume do Amazon EBS em instâncias 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.
- Analise 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:

- [Documentação do Amazon FSx](#)
- [O que é o Amazon Elastic File System?](#)

Vídeos relacionados:

- [Aprofundamento em Volumes Elásticos do Amazon EBS](#)
- [Amazon EBS e estratégias de otimização de snapshots para melhorar a performance e reduzir os custos](#)
- [Otimização de custo e performance do Amazon EFS 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.

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: A remoção de dados desnecessários reduz o tamanho de armazenamento necessário para sua workload e o impacto ambiental da workload.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a implementação

Não armazene dados de que você não precisa. Automatize a exclusão de dados desnecessários. Use tecnologias que eliminem dados duplicados em níveis de arquivo e bloco. Aproveite a replicação de dados nativos e os recursos de redundância dos serviços.

Etapas da implementação

- Avalie se você pode evitar o armazenamento de dados usando conjuntos de dados disponíveis publicamente no [AWS Data Exchange](#) e [Dados abertos no AWS](#).

- Use mecanismos que possam duplicar dados no nível de bloco e objeto. Aqui estão alguns exemplos de como deduplicar dados no AWS:

Storage service	Deduplication mechanism
Amazon S3	Use AWS Lake Formation FindMatches para localizar registros correspondentes em um conjunto de dados (incluindo aqueles sem identificadores) usando o novo FindMatches ML Transform.
Amazon FSx	Habilite a deduplicação de dados no Amazon FSx para Windows.
Snapshots do Amazon Elastic Block Store	Os snapshots são backups incrementais, o que significa que apenas os blocos no dispositivo que foram alterados após o snapshot mais recente são salvos.

- Analise o acesso aos dados para identificar dados desnecessários. Automatize as políticas de ciclo de vida. Aproveite os recursos do serviço nativo, como o [tempo de vida útil do Amazon DynamoDB](#), [ciclo de vida do Amazon S3](#) ou [retenção de log do Amazon CloudWatch](#) para exclusão.
- Use os recursos de virtualização de dados no AWS para manter os dados em sua origem e evitar a duplicação de dados.
 - [Virtualização de dados nativos da nuvem no AWS](#)
 - [Laboratório: Otimizar padrão de dados usando o compartilhamento de dados do Amazon Redshift](#)
- Use a tecnologia de backup que pode fazer backups incrementais.
- Aproveite a durabilidade do [Amazon S3](#) e a [replicação do Amazon EBS](#) para atender às suas metas de durabilidade em vez de tecnologias autogerenciadas (como uma Redundant Array of Independent Disks [RAID – Matriz redundante de discos independentes]).
- Centralize o log e rastreie os dados, elimine a duplicação de entradas de log idênticas e estabeleça mecanismos para ajustar a prolixidade quando necessário.
- Preencha os caches com antecedência somente quando justificável.
- Estabeleça o monitoramento e a automação de cache para redimensioná-lo de forma adequada.

- Remova implantações e ativos desatualizados de depósitos de objetos e caches de borda ao enviar novas versões da sua workload.

Recursos

Documentos relacionados:

- [Retenção de dados do log de alterações no CloudWatch Logs](#)
- [Eliminação de duplicação de dados no Amazon FSx for Windows File Server](#)
- [Recursos do Amazon FSx for ONTAP incluindo a eliminação da duplicação de dados](#)
- [Invalidar 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](#)

Vídeos relacionados:

- [Correspondência difusa e deduplicação de dados com ML Transforms para o AWS Lake Formation](#)

Exemplos relacionados:

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

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

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

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: o uso de armazenamento ou sistemas de arquivos permite que os dados sejam compartilhados com um ou mais consumidores sem precisar copiá-los. Isso ajuda a reduzir os recursos de armazenamento necessários à workload.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a 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 sua 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 da implementação

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

Storage option	When to use
Amazon EBS Multi-Attach	O Amazon EBS Multi-Attach permite que você anexe um único volume de SSD de IOPS provisionados (io1 ou io2) para várias instâncias que estão na mesma zona de disponibilidade.
Amazon EFS	Consulte Quando escolher o Amazon EFS .
Amazon FSx	Consulte Como escolher um Amazon FSx File System .
Amazon S3	As aplicações que não exigem uma estrutura de sistema de arquivos e são projetadas para

Storage option	When to use
	funcionar com armazenamento de objetos podem usar o Amazon S3 como solução de armazenamento de objetos altamente escalável, durável e de baixo custo.

- Copie os dados para ou busque os dados de sistemas de arquivos compartilhados somente quando necessário. A título de exemplo, você pode usar um [sistema de arquivos do Amazon FSx for Lustre respaldado pelo Amazon S3](#) e carregar somente o subconjunto de dados necessário para processar os trabalhos para o Amazon FSx.
- Exclua dados conforme apropriado para os seus padrões de uso, tal como descrito em [SUS04-BP03 Usar políticas para gerenciar o ciclo de vida de seus conjuntos de dados](#).
- Desvincule volumes de clientes que não estão utilizando-os ativamente.

Recursos

Documentos relacionados:

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

Vídeos relacionados:

- [Otimização de custos de armazenamento com o Amazon EFS](#)

SUS04-BP07 Minimizar a movimentação de dados entre redes

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

Antipadrões comuns:

- 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 estabelecer 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 da implementação

- Considere a proximidade dos dados ou dos usuários como um fator de decisão ao [selecionar uma região para sua workload](#).
- 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 (como Parquet ou ORC) e compacte os dados antes movimentá-los na rede.
- 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 [o compartilhamento de dados entre contas no AWS Lake Formation](#).
- Use serviços que podem ajudar você a executar código mais perto dos usuários da workload.

Service	Quando usar
o Lambda@Edge	Use para operações com uso intenso de computação que são executadas quando objetos não estão no cache.
Funções do CloudFront	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 global de borda 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](#)
- [Distribuição de arquivos compactados com o Amazon CloudFront](#)

Vídeos relacionados:

- [Demystifying data transfer on AWS \(Desmistificação da transferência de dados na AWS\)](#)

Exemplos relacionados:

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

SUS04-BP08 Fazer backup de dados somente quando for difícil recriar

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

Antipadrões comuns:

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

Benefícios do estabelecimento desta prática recomendada: evitar backups de dados não essenciais reduz os recursos de armazenamento necessários à workload e diminui o respectivo impacto ambiental.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a 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 aos requisitos de conformidade. Examine as políticas de backup e exclua armazenamentos temporários que não forneçam valor em um cenário de recuperação.

Etapas da implementação

- Implemente uma política de classificação de dados como descrito em [SUS04-BP01 Implementar uma política de classificação de dados](#).
- Use a criticidade da estratégia de classificação de dados e backup de design 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 os 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 uma solução automatizada ou um serviço gerenciado para fazer backup de dados essenciais aos negócios.
 - O [AWS Backup](#) é um serviço totalmente gerenciado que facilita a centralização e a automatização da proteção de dados nos serviços da AWS, na nuvem e no ambiente on-

premises. Para obter orientações práticas sobre como criar backups automatizados usando o AWS Backup, consulte [Laboratórios do Well-Architected: Teste de backup e restauração de dados](#).

- [Automatize backups e otimize os custos de backup do 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 Realizar o backup de dados automaticamente](#)
- [REL13-BP02 Usar estratégias de recuperação definidas para atender aos 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 do APN: parceiros que podem ajudar com o backup](#)
- [AWS Marketplace: produtos que podem ser usados para backup](#)
- [Como fazer backup do Amazon EFS](#)
- [Como fazer backup do Amazon FSx para Windows File Server](#)
- [Backup e restauração do Amazon ElastiCache for Redis](#)

Vídeos relacionados:

- [AWS re:Invent 2021: Backup, recuperação de desastres e proteção contra ransomware com a AWS](#)
- [AWS Backup Demonstração do AWS Backup: Backup entre contas e entre regiões](#)
- [AWS re:Invent 2019: Deep dive on AWS Backup, ft. Rackspace \(STG341\)](#)

Exemplos relacionados:

- [Laboratório do Well-Architected: Teste de backup e restauração de dados](#)
- [Laboratório do Well-Architected: Backup e restauração com failback para workload do Analytics](#)
- [Laboratório do Well-Architected: Recuperação de desastres: backup e restauração](#)

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.

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: dimensionar corretamente os recursos de nuvem ajuda a reduzir o impacto ambiental da workload, a economizar e a manter os parâmetros de performance.

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

Orientações para a implementação

Selecione do modo mais eficiente o número total de hardware necessário à workload para melhorar a eficiência geral. A Nuvem AWS oferece flexibilidade para expandir ou reduzir dinamicamente a quantidade de recursos 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 o recursos de nuvem e atender às suas necessidades empresariais.

Etapas da implementação

- Escolha os tipos de instância mais adequados às suas necessidades.
 - [Como faço para escolher o tipo de instância do Amazon EC2 apropriado para minha workload?](#)
 - [Seleção de tipo de instância baseada em atributos para frota do Amazon EC2.](#)
 - [Criar um grupo do Auto Scaling usando seleção de tipo de instância baseada em atributos.](#)
- Escale usando pequenos incrementos para workloads variáveis.
- Use várias opções de compra de computação para equilibrar flexibilidade, escalabilidade e economia no uso de instâncias.
 - As [instâncias sob demanda](#) são mais adequadas para workloads novas, com estado e com picos que não podem ter flexibilidade com relação ao tipo de instância, local e tempo.
 - As [instâncias spot](#) são excelentes para complementar as outras opções para aplicações tolerantes a falhas e flexíveis.
 - Utilize [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 uma variedade de instâncias e zonas de disponibilidade para maximizar a disponibilidade da aplicação e aproveitar o excesso de capacidade quando possível.
- Use as recomendações de dimensionamento correto das ferramentas da AWS para fazer ajustes na workload.
 - [AWS Compute Optimizer](#)
 - [AWS Trusted Advisor](#)

- 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 atributo do Auto Scaling para Amazon EC2 Fleet](#)
- [Documentação do AWS Compute Optimizer](#)
- [Otimização do Lambda: otimização da performance](#)
- [Documentação do Auto Scaling](#)

Vídeos relacionados:

- [Criar um ambiente de computação eficiente em termos de custo, energia e recursos](#)

Exemplos relacionados:

- [Laboratórios do Well-Architected: Dimensionamento correto com o AWS Compute Optimizer e utilização de memória habilitada \(Nível 200\)](#)

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.

Antipadrões comuns:

- 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 correta da AWS, como
 - o [AWS Compute Optimizer](#).

Benefícios do estabelecimento desta prática recomendada: Ao usar instâncias com eficiência de energia e dimensionadas corretamente, você consegue 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 da implementação

- Conheça e explore os tipos de instância que podem reduzir o impacto ambiental de sua workload.
 - Inscreva-se nas [Novidades da AWS](#) para ficar por dentro das tecnologias e instâncias mais recentes da AWS.
 - Conheça os diversos tipos de instâncias da AWS.
 - Conheça as instâncias baseadas em AWS Graviton, que oferecem a melhor performance por watt de energia usada no Amazon EC2 assistindo aos vídeos [re:Invent 2020 - Deep dive on AWS Graviton2 processor-powered Amazon EC2 instances \(re:Invent 2020 - aprofundamento em instâncias do Amazon EC2 alimentadas por processadores AWS Graviton2\)](#) e [Deep dive into AWS Graviton3 and Amazon EC2 C7g instances \(Aprofundamento em AWS Graviton3 e instâncias C7g do Amazon EC2\)](#).
- Planeje e migre sua workload para tipos de instância com impacto mínimo.
 - Defina um processo para avaliar novos recursos ou instâncias para sua 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.
 - Considere migrar sua workload para instâncias baseadas em Graviton e melhorar a eficiência da performance da workload.

- [AWS Graviton Fast Start](#)
- [Considerações ao migrar workloads para instâncias do Amazon Elastic Compute Cloud baseadas no AWS Graviton](#)
- [AWS Graviton2 para ISVs](#)
- 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 aos [Amazon EC2 DL1](#). Instâncias do AWS Inferentia, como instâncias Inf2, oferecem performance até 50% melhor por watt em relação a instâncias comparáveis do Amazon EC2.
- Use o [Amazon SageMaker Inference Recommender](#) para dimensionar endpoints de inferência de ML corretamente.
- 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 otimize a instância de sua workload.
 - Para workloads efêmeras, avalie [métricas do Amazon CloudWatch para instâncias](#) , como `CPUUtilization` , a fim de identificar se a instância está ociosa ou é subutilizada.
 - Para workloads estáveis, verifique as ferramentas da AWS para dimensionamento correto, como o [AWS Compute Optimizer](#) , em intervalos regulares a fim de identificar oportunidades para otimizar e dimensionar instâncias corretamente.
 - [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](#)
- [Amazon EC2 DL1](#)
- [Frotas de reserva de capacidade do Amazon EC2](#)
- [Frota spot do Amazon EC2](#)
- [Funções: configuração de função do Lambda](#)
- [Seleção de tipo de instância baseada em atributos para frota do Amazon EC2](#)
- [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:

- [Deep dive on AWS Graviton2 processor-powered Amazon EC2 instances \(Aprofundamento em instâncias do Amazon EC2 alimentadas por processadores AWS Graviton2\)](#)
- [Deep dive into AWS Graviton3 and Amazon EC2 C7g instances \(Aprofundamento em AWS Graviton3 e instâncias C7g do Amazon EC2\)](#)
- [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](#)
- [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](#)
- [Laboratório do Well-Architected: Migração de serviços para o Graviton](#)

SUS05-BP03 Usar serviços gerenciados

Use serviços gerenciados para operar com maior eficiência na nuvem.

Antipadrões comuns:

- 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 do estabelecimento desta 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 por causa do ambiente de gerenciamento de vários locatários.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: Médio

Orientações para a 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 gerenciado de banco de dados. O [Amazon Athena](#), o [Amazon EMR](#) e o [Amazon OpenSearch Service](#) fornecem um serviço gerenciado de análise.

Etapas da implementação

1. Faça um inventário de serviços e componentes para sua workload.
2. Avalie e identifique componentes que podem ser substituídos por serviços gerenciados. Veja aqui alguns exemplos de quando você pode pensar em usar um serviço gerenciado:

Task	What to use on AWS
Hospedagem de banco de dados	Use instâncias do Amazon Relational Database Service (Amazon RDS) gerenciad

Task	What to use on AWS
	as, em vez de manter suas próprias instâncias do Amazon RDS no Amazon Elastic Compute Cloud (Amazon EC2) .
Hospedagem de uma workload containerizada	Use o AWS Fargate , em vez de implementar sua própria infraestrutura de contêiner.
Hospedagem de aplicações Web	Use o AWS Amplify Hosting como CI/CD totalmente gerenciadas e serviço de hospedagem para sites estáticos e aplicações Web renderizadas do lado do servidor.

- Identifique dependências e crie um plano de migração. Atualize runbooks e manuais and playbooks de forma apropriada.
 - O [AWS Application Discovery Service](#) coleta e apresenta automaticamente informações detalhadas sobre dependências e utilização de aplicações para ajudar você a tomar decisões bem fundamentadas ao planejar a migração.
- Teste o serviço antes de migrar para o serviço gerenciado.
- Use o plano de migração para substituir serviços auto-hospedados por serviço gerenciado.
- 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:

- [Operações em nuvem em escala com o AWS Managed Services](#)

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.

Antipadrões comuns:

- 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 desempenho, menor custo e melhor desempenho 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 estabelecer esta prática recomendada: Ao otimizar o uso de aceleradores baseados em hardware, você pode reduzir as demandas de infraestrutura física de sua 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 este hardware somente pelo tempo necessário e desative-as com automação quando não precisar mais delas para reduzir o consumo de recursos.

Etapas da implementação

- Identifique quais [instâncias com computação acelerada](#) podem atender aos seus requisitos.
- Para workloads de machine learning, utilize hardware específico para sua workload, como [AWS Trainium](#), [AWS Inferentia](#) e o [Amazon EC2 DL1](#). Instâncias do AWS Inferentia, como instâncias Inf2, oferecem [performance até 50% melhor por watt em relação a instâncias comparáveis do Amazon EC2](#).

- Colete métricas de uso para suas instâncias com computação acelerada. Por exemplo, você pode usar o agente do CloudWatch para coletar métricas como `utilization_gpu` e `utilization_memory` para suas GPUs, conforme mostrado em [Colete métricas da GPU NVIDIA com o Amazon CloudWatch](#).
- 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 de GPU no Deep Learning AMI](#)
 - [Otimização de E/S para ajuste de desempenho de GPU de treinamento de aprendizado profundo no Amazon SageMaker](#)
- Use as mais recentes bibliotecas de alto desempenho e drivers de GPU.
- Use automação para liberar instâncias de GPU quando não estiverem em uso.

Recursos

Documentos relacionados:

- [Computação acelerada](#)
- [Vamos, arquiteto! Arquitetura 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](#)
- [Escolha o melhor acelerador de IA e compilação de modelo para inferência de visão computacional com o Amazon SageMaker](#)

Vídeos relacionados:

- [How to select Amazon EC2 GPU instances for deep learning \(Como selecionar instâncias de GPU do Amazon EC2 para aprendizado profundo\)](#)
- [Deploying Cost-Effective Deep Learning Inference \(Implantação de inferência de aprendizado profundo econômico\)](#)

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 Adotar métodos que podem apresentar melhorias na sustentabilidade rapidamente](#)
- [SUS06-BP02 Manter a workload atualizada](#)
- [SUS06-BP03 Aumentar a utilização de ambientes de desenvolvimento](#)
- [SUS06-BP04 Usar parques de dispositivos gerenciados para testes](#)

SUS06-BP01 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.

Antipadrões comuns:

- 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 ficou ultrapassada.
- Você não tem mecanismos para melhorar a sustentabilidade de sua workload.

Benefícios do estabelecimento desta prática recomendada: por meio da criação de um processo para introduzir e monitorar melhorias de sustentabilidade, você conseguirá adotar novos recursos e capacidades, eliminar problemas e melhorar a eficiência da workload continuamente.

Nível de risco exposto se esta prática recomendada não é estabelecida: médio

Orientações para a 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 da implementação

- Adicione requisitos de melhoria da sustentabilidade às suas pendências de desenvolvimento.
- Use um [processo de melhoria](#) iterativo para identificar, avaliar, priorizar, testar e implantar essas melhorias.
- Melhore e otimize continuamente seus processos de desenvolvimento. A título de exemplo, [automatize seu processo de entrega de software usando pipelines de integração contínua e entrega contínua](#) a fim de testar e implantar possíveis melhorias para reduzir o nível de esforço e limitar os erros provocados por processos manuais.
- Desenvolva e teste possíveis melhorias usando os componentes representativos mínimos viáveis para reduzir o custo dos testes.
- Avalie continuamente o impacto das melhorias e faça ajustes conforme necessário.

Recursos

Documentos relacionados:

- [A AWS viabiliza soluções de sustentabilidade](#)
- [Práticas de desenvolvimento ágil escaláveis com base no AWS CodeCommit](#)

Vídeos relacionados:

- [Entregar arquiteturas sustentáveis e de alta performance](#)

Exemplos relacionados:

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

SUS06-BP02 Manter a workload atualizada

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

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: ao estabelecer um processo para manter a workload atualizada, você pode adotar novos recursos e capacidades, resolver problemas e aumentar a eficiência da workload.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: baixo

Orientações para a implementação

Sistemas operacionais, tempos de execução, 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 da implementação

- Defina 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 sua workload com o intuito de:
 - Reduzir impactos de sustentabilidade.
 - Obter eficiências de performance.
 - Remover barreiras de melhorias planejadas.
 - Aumentar sua capacidade de medir e gerenciar impactos na sustentabilidade.
- Fazer o inventário de software e arquitetura da workload e identificar os componentes que precisam ser atualizados.
 - Você pode usar o [AWS Systems Manager Inventory](#) 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.
- Entenda como atualizar os componentes de sua workload.

Workload component	How to update
Imagens de máquina	Use o EC2 Image Builder para gerenciar atualizações de imagens de máquina da Amazon (AMIs) para imagens de servidor Linux ou Windows.
Imagens de contêiner	Use o Amazon Elastic Container Registry (Amazon ECR) com seu pipeline para gerenciar Amazon Elastic Container Service (Amazon ECS) imagens .
AWS Lambda	O AWS Lambda inclui recursos de gerenciamento de versão .

- Use automação no processo de atualização para reduzir o nível de esforço para implantar novos recursos e limitar erros causados por processos manuais.
- Você pode usar [CI/CD](#) para atualizar automaticamente AMIs, imagens de contêiner e outros artefatos relacionados à sua aplicação de nuvem.
- Você pode usar ferramentas como o [Gerenciador de Patches do AWS Systems Manager](#) para automatizar o processo de atualizações de sistema e programar a atividade usando as [Janelas de Manutenção do AWS Systems Manager](#).

Recursos

Documentos relacionados:

- [Central de arquitetura da AWS](#)
- [Quais as novidades da AWS](#)
- [Ferramentas do desenvolvedor na AWS](#)

Exemplos relacionados:

- [Laboratórios do Well-Architected: Gerenciamento de inventário e patches](#)
- [Laboratório: AWS Systems Manager](#)

SUS06-BP03 Aumentar a utilização de ambientes de desenvolvimento

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

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: ao aumentar a utilização dos ambientes de compilação, você pode melhorar a eficiência geral de sua workload de nuvem e, ao mesmo tempo, alocar recursos aos compiladores para que eles desenvolvam, testem e compilem com eficiência.

Nível de exposição a riscos quando esta prática recomendada não é estabelecida: baixo

Orientações para a 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 escalabilidade automática, contêineres e tecnologias sem servidor para alinhar a capacidade de desenvolvimento e teste com o uso. Limite o volume de dados apenas para atender os requisitos de teste. Se usar dados de produção no teste, explore possibilidades para compartilhar os dados da produção em vez de movimentá-los.

Etapas da implementação

- Use a infraestrutura como código para provisionar os ambientes de compilação.
- Use a automação para gerenciar o ciclo de vida de seus ambientes de desenvolvimento e teste e maximizar a eficiência dos recursos de compilação.

- 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 o AWS CloudFormation?](#)
- [O que é o AWS CodeBuild?](#)
- [Programador de Instâncias da AWS](#)

Vídeos relacionados:

- [Práticas recomendadas de integração contínua](#)

SUS06-BP04 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.

Antipadrões comuns:

- 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 do estabelecimento desta prática recomendada: usar parques de dispositivos gerenciados para testar aplicações habilitadas para a nuvem oferece inúmeros 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 exposição a riscos quando esta prática recomendada não é estabelecida: baixo

Orientações para a 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 da implementação

- Defina seus requisitos e plano de testes (como tipo de teste, sistemas operacionais e programação dos testes).
 - Você pode usar o [Amazon CloudWatch RUM](#) para coletar e analisar dados do lado do cliente e moldar seu plano de testes.
- Selecione o parque de dispositivos gerenciados capaz de atender aos seus requisitos de teste. Por exemplo, você pode usar o [AWS Device Farm](#) para testar e conhecer o impacto de suas alterações sobre um conjunto representativo de hardware.
- Use a integração contínua/implantação contínua (CI/CD) para programar e executar seus testes.
 - [Integração do AWS Device Farm com seu pipeline de CI/CD para executar testes Selenium entre navegadores](#)
 - [Compilação e teste de aplicativos iOS e iPadOS com DevOps e serviços móveis da AWS](#)
- Avalie continuamente os resultados dos testes e faça as melhorias necessárias.

Recursos

Documentos relacionados:

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

Exemplos relacionados:

- [Aplicativo de exemplo do AWS Device Farm para Android](#)
- [Aplicativo de exemplo do AWS Device Farm para iOS](#)
- [estes Web Appium para AWS Device Farm](#)

Vídeos relacionados:

- [Otimização de aplicações por meio de insights sobre o usuário final com o Amazon CloudWatch RUM](#)

Conclusão

Um número cada vez maior de organizações está definindo metas de sustentabilidade em resposta às alterações nas normas governamentais, na vantagem competitiva e na demanda de clientes, funcionários e investidores. Diretores de tecnologia, arquitetos, desenvolvedores e membros da equipe de operações estão procurando maneiras de colaborar diretamente com as metas de sustentabilidade de suas organizações. Ao usar esses princípios de design e as práticas recomendadas compatíveis com os serviços da AWS, é possível tomar decisões embasadas, equilibrando segurança, custo, performance, confiabilidade e excelência operacional com os resultados de sustentabilidade de suas workloads na Nuvem AWS. Todas as ações realizadas para reduzir o uso de recursos e aumentar a eficiência entre suas workloads contribuem com a redução do impacto ambiental e com as metas de sustentabilidade mais amplas de suas organizações.

Colaboradores

Os colaboradores desse documento incluem:

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

Leitura adicional

Para obter mais informações, consulte:

- [AWS Well-Architected](#)
- [Centro de Arquitetura da 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 deste whitepaper, inscreva-se no RSS feed.

Alteração	Descrição	Data
Níveis de risco atualizados	Pequenas atualizações nos níveis de risco das práticas recomendadas.	October 3, 2023
Orientação sobre práticas recomendadas atualizada	As práticas recomendadas foram atualizadas com novas orientações nas seguintes áreas: Alinhamento com a demanda , Software e arquitetura , dados e hardware e serviços .	July 13, 2023
Atualizado para o novo Framework	Práticas recomendadas atualizadas com orientações prescritivas e novas práticas recomendadas adicionadas.	April 10, 2023
Whitepaper atualizado	Práticas recomendadas atualizadas com novas orientações para implementação.	December 15, 2022
Whitepaper atualizado	Práticas recomendadas ampliadas e planos de melhoria adicionados.	October 20, 2022
Publicação inicial	Publicação do Pilar Sustentabilidade: AWS Well Architected Framework.	December 2, 2021

Avisos

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

© 2021 Amazon Web Services, Inc. ou suas afiliadas. Todos os direitos reservados.

AWS Glossary

For the latest AWS terminology, see the [AWS glossary](#) in the Glossário da AWS Reference.