



Whitepaper da AWS

Recuperação de desastres de workloads na AWS: recuperação na nuvem



Recuperação de desastres de workloads na AWS: recuperação na nuvem: Whitepaper da AWS

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

As marcas comerciais e o visual comercial da Amazon não podem ser usados em conexão com nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa causar confusão entre os clientes ou que deprecie ou desacredite a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, conectados ou patrocinados pela Amazon.

Table of Contents

Recuperação de desastres de workloads na AWS	1
Resumo	1
Introdução	2
Recuperação de desastres e disponibilidade	2
Modelo de responsabilidade compartilhada para resiliência	5
Responsabilidade da AWS: “Resiliência da nuvem”	5
Responsabilidade do cliente: “Resiliência na nuvem”	5
O que é um desastre?	7
Alta disponibilidade não é recuperação de desastres	8
Plano de continuidade dos negócios	9
Análise de impacto sobre os negócios e avaliação de riscos	9
Objetivos de recuperação (RTO e RPO)	10
A recuperação de desastres é diferente na nuvem	13
Região única da AWS	14
Várias regiões da AWS	15
Opções de recuperação de desastres na nuvem	16
Backup e restauração	16
Serviços da AWS	17
Luz piloto	21
Serviços da AWS	22
CloudEndure Disaster Recovery	24
Standby passivo	24
Serviços da AWS	25
Ativo/ativo em vários locais	26
Serviços da AWS	27
Detecção	30
Teste de recuperação de desastres	32
Conclusão	33
Colaboradores	34
Leitura adicional	35
Revisões do documento	36
Avisos	37

Recuperação de desastres de workloads na AWS: recuperação na nuvem

Data de publicação: 12 de fevereiro de 2021 ([Revisões do documento](#))

Resumo

Recuperação de desastres é um processo em que você se prepara para se recuperar de um desastre. Um evento que impede que uma workload ou um sistema cumpra os respectivos objetivos de negócios no principal local em que está implantado é considerado um desastre. Este whitepaper descreve as práticas recomendadas para planejar e testar a recuperação de desastres de qualquer workload implantada na AWS e oferece diferentes abordagens para atenuar riscos e atender a metas de objetivo de tempo de recuperação (RTO) e objetivo de ponto de recuperação (RPO) para a workload em questão.

Introdução

Sua workload precisa executar a função pretendida de forma correta e consistente. Para conseguir isso, você deve ter um plano para obter resiliência. Resiliência refere-se à capacidade de uma workload se recuperar de interrupções de infraestrutura ou serviço, adquirir dinamicamente recursos de computação para atender à demanda e mitigar interrupções, como configurações incorretas ou problemas transitórios de rede.

A recuperação de desastres (DR) é uma parte importante de sua estratégia de resiliência e está relacionada a como sua workload responde quando ocorre um desastre ([desastre](#) é um evento que causa um sério impacto negativo em seus negócios). Essa resposta deve se basear nos objetivos de negócios de sua organização que especificam a estratégia da workload para evitar a perda de dados, o que é conhecido como [objetivo de ponto de recuperação \(RPO\)](#), e para reduzir o tempo de inatividade quando a workload não estiver disponível para uso, o que é conhecido como [objetivo de tempo de recuperação \(RTO\)](#). Portanto, para atender aos seus objetivos de recuperação ([RPO e RTO](#)), você deve implementar resiliência no design de suas workloads na nuvem para um evento de desastre ocasional. Essa abordagem ajuda sua organização a manter a continuidade dos negócios como parte do [plano de continuidade dos negócios \(BCP\)](#).

Este artigo mostra principalmente como planejar, projetar e implementar arquiteturas na AWS que atendam aos objetivos de recuperação de desastres de sua empresa. As informações compartilhadas aqui são destinadas a pessoas que desempenham funções de tecnologia, como diretores de tecnologia (CTOs), arquitetos, desenvolvedores e membros da equipe de operações.

Recuperação de desastres e disponibilidade

A recuperação de desastres pode ser comparada com a disponibilidade, que é outro componente importante de sua estratégia de resiliência. Enquanto a recuperação de desastres mede os objetivos para eventos únicos, os objetivos de disponibilidade medem os valores médios ao longo de um período.

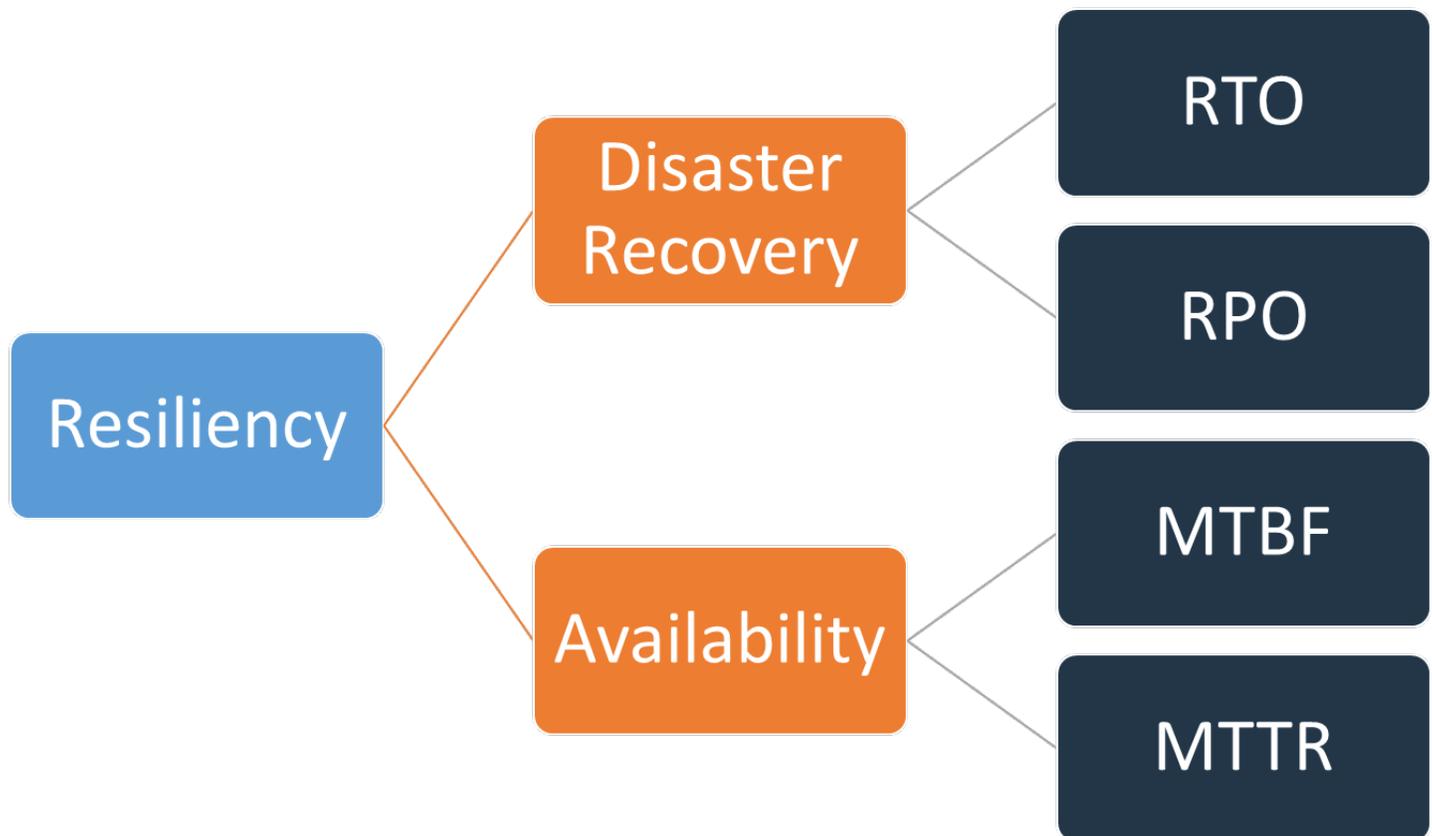


Figura 1: objetivos de resiliência

A disponibilidade é calculada usando o tempo médio entre falhas (MTBF) e o tempo médio de recuperação (MTTR):

$$\textit{Availability} = \frac{\textit{Available for Use Time}}{\textit{Total Time}} = \frac{\textit{MTBF}}{\textit{MTBF} + \textit{MTTR}}$$

Essa abordagem geralmente é chamada de “noves”, em que uma meta de disponibilidade de 99,9% é chamada de “três noves”.

Para sua workload, pode ser mais fácil contar solicitações bem-sucedidas e com falha em vez de usar uma abordagem baseada em tempo. Nesse caso, o seguinte cálculo pode ser usado:

$$\textit{Availability} = \frac{\textit{Successful Responses}}{\textit{Valid Requests}}$$

A recuperação de desastres concentra-se em eventos de desastres, enquanto a disponibilidade se concentra em interrupções mais comuns de menor escala, como falhas de componentes, problemas de rede e picos de carga. O objetivo da recuperação de desastres é a continuidade dos negócios, enquanto a disponibilidade diz respeito à maximização do tempo em que uma workload fica disponível para executar a funcionalidade empresarial pretendida. Ambas devem fazer parte de sua estratégia de resiliência.

Modelo de responsabilidade compartilhada para resiliência

A resiliência é uma responsabilidade compartilhada entre a AWS e você, o cliente. É importante entender como a recuperação de desastres e a disponibilidade funcionam nesse modelo compartilhado como parte da resiliência.

Responsabilidade da AWS: “Resiliência da nuvem”

A AWS é responsável pela resiliência da infraestrutura que executa todos os serviços oferecidos na Nuvem AWS. Essa infraestrutura abrange o hardware, o software, as redes e as instalações que operam os serviços de nuvem da AWS. A AWS empreende esforços comercialmente razoáveis para disponibilizar esses serviços de nuvem da AWS, garantindo que a disponibilidade do serviço atenda ou supere os [Acordos de Nível de Serviço \(SLAs\) da AWS](#).

A [infraestrutura de nuvem global da AWS](#) foi projetada para permitir que os clientes criem arquiteturas de workloads altamente resilientes. Todas as regiões da AWS são totalmente isoladas entre e cada uma consiste em várias [zonas de disponibilidade](#), que são partições fisicamente isoladas da infraestrutura. As zonas de disponibilidade isolam falhas que possam afetar a resiliência da workload, evitando que elas afetem outras zonas da região. Porém, ao mesmo tempo, todas as zonas de disponibilidade em uma região da AWS são interconectadas por redes de alta largura de banda e baixa latência, usando fibra metropolitana dedicada e totalmente redundante para proporcionar redes de alta taxa de transferência e baixa latência entre as zonas de disponibilidade. Todo tráfego entre as zonas é criptografado. A performance da rede é suficiente para realizar a replicação síncrona entre as AZs. As zonas de disponibilidade simplificam o processo de particionamento de aplicações para oferecer alta disponibilidade.

Responsabilidade do cliente: “Resiliência na nuvem”

Sua responsabilidade será determinada pelos serviços de nuvem da AWS que você selecionar. Isso define a quantidade de trabalho de configuração que você deve executar como parte de suas responsabilidades de resiliência. Por exemplo, um serviço como o Amazon Elastic Compute Cloud (Amazon EC2) requer que o cliente execute todas as tarefas de configuração e gerenciamento de resiliência necessárias. Os clientes que implantam instâncias do Amazon EC2 são responsáveis pela [implantação de instâncias do EC2 em vários locais](#) (como zonas de disponibilidade da AWS), pela [implementação de autorrecuperação](#) por meio de serviços como o AWS Auto Scaling e pelo uso de [práticas recomendadas de arquitetura de workload resiliente](#) para aplicações instaladas

nas instâncias. Com relação a serviços gerenciados, como o Amazon S3 e o Amazon DynamoDB, a AWS opera a camada de infraestrutura, o sistema operacional e as plataformas, e os clientes acessam os endpoints para armazenar e recuperar dados. Você é responsável por gerenciar a resiliência de seus dados, bem como estratégias de backup, versionamento e replicação.

A implantação de sua workload em várias zonas de disponibilidade em uma região da AWS faz parte de uma estratégia de alta disponibilidade que protege as workloads isolando os problemas em uma zona de disponibilidade e usa a redundância das outras zonas para continuar atendendo às solicitações. A arquitetura multi-AZ também faz parte de uma estratégia de DR projetada para isolar e proteger ainda mais as workloads contra problemas como quedas de energia, quedas de raios, tornados, terremotos etc. As estratégias de DR também podem usar várias regiões da AWS. Por exemplo, em uma configuração ativo/passivo, o serviço para a workload fará failover de sua região ativa para sua região de DR se a região ativa não puder mais atender às solicitações.

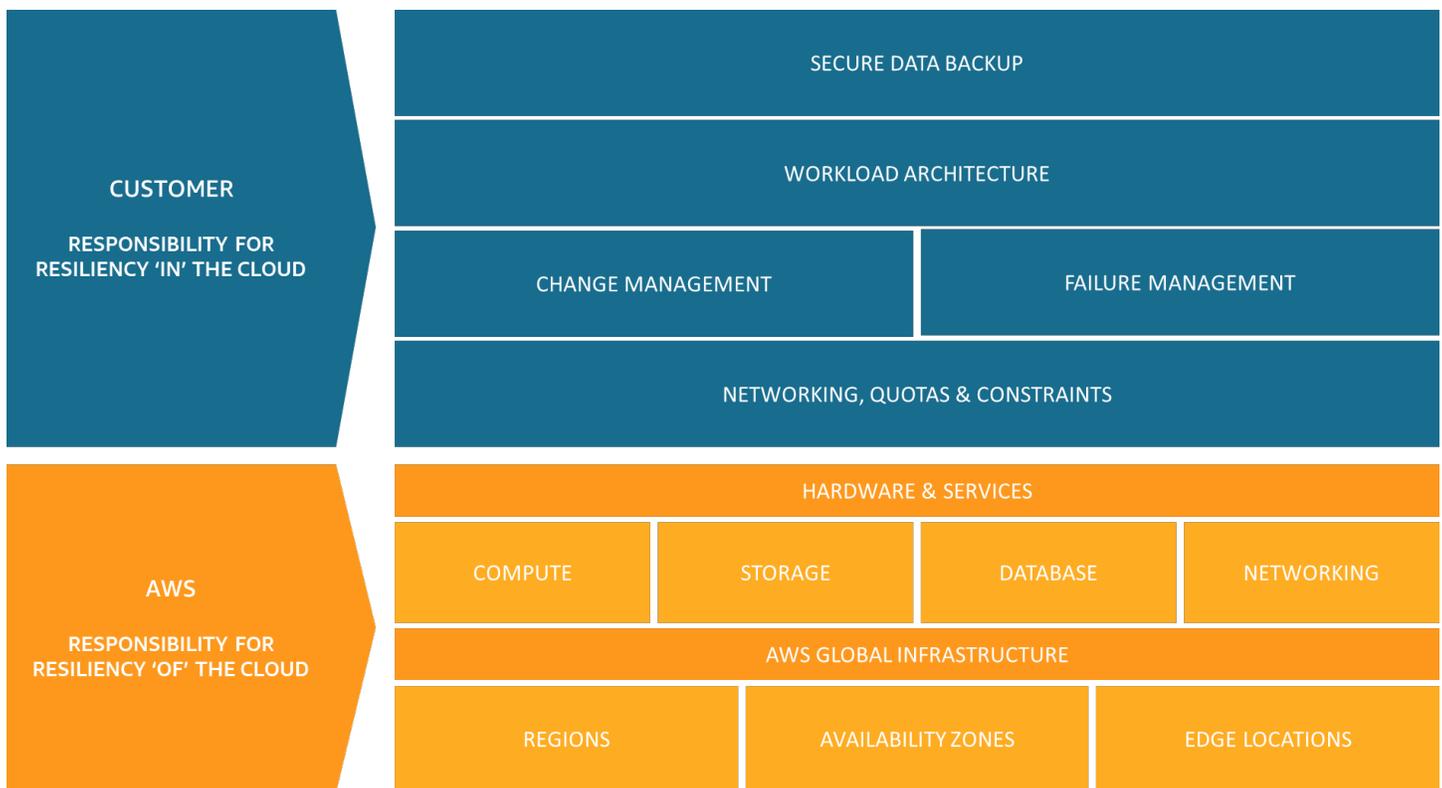


Figura 2: a resiliência é uma responsabilidade compartilhada entre a AWS e o cliente

O que é um desastre?

Ao planejar a recuperação de desastres, avalie seu plano para estas três principais categorias de desastre:

- Desastres naturais, como terremotos ou inundações
- Falhas técnicas, como falha de energia ou conectividade de rede
- Ações humanas, como configuração incorreta inadvertida ou acesso ou modificação não autorizada/externa

Cada um desses possíveis desastres também terá um impacto geográfico que pode ser local, regional, nacional, continental ou global. Tanto a natureza do desastre quanto o impacto geográfico são importantes ao considerar sua estratégia de recuperação de desastres. Por exemplo, você pode atenuar um problema de inundação local que causa uma paralisação no datacenter empregando uma estratégia multi-AZ, já que isso não afetaria mais de uma zona de disponibilidade. No entanto, um ataque aos dados de produção exigiria que você invocasse uma estratégia de recuperação de desastres que aplicasse failover para fazer backup dos dados em outra região da AWS.

Alta disponibilidade não é recuperação de desastres

Tanto a disponibilidade quanto a recuperação de desastres dependem de algumas das mesmas práticas recomendadas, como monitoramento de falhas, implantação em vários locais e failover automático. No entanto, a disponibilidade se concentra nos componentes da workload, enquanto a recuperação de desastres se concentra em cópias distintas de toda a workload. A recuperação de desastres tem objetivos diferentes da disponibilidade, medindo o tempo de recuperação após os eventos de maior escala que se qualificam como desastres. Primeiro, é necessário garantir que sua workload atenda aos seus objetivos de disponibilidade, pois uma arquitetura altamente disponível permitirá que você atenda às necessidades dos clientes no caso de eventos que afetem a disponibilidade. Sua estratégia de recuperação de desastres requer abordagens diferentes das de disponibilidade, concentrando-se na implantação de sistemas distintos em vários locais, para que você possa fazer failover de toda a workload, se necessário.

Você deve considerar a disponibilidade de sua workload no planejamento de recuperação de desastres, pois isso influenciará a abordagem adotada. A workload executada em uma única instância do Amazon EC2, em uma única zona de disponibilidade, não tem alta disponibilidade. Se um problema de inundação local afetar essa zona de disponibilidade, esse cenário exigirá failover para outra AZ para atender aos objetivos de recuperação de desastres. Compare esse cenário com o de uma workload altamente disponível, com implantação ativo/ativo em vários locais, que é implantada em várias regiões ativas e todas as regiões estão atendendo ao tráfego de produção. Nessa circunstância, mesmo no caso improvável de um desastre maciço destruir uma região inteira, a estratégia de DR é executada por meio do encaminhamento de todo o tráfego para as regiões restantes.

Na disponibilidade e na recuperação de desastres, a maneira como você aborda os dados também é diferente. Considere uma solução de armazenamento que se replica continuamente para outro local para obter alta disponibilidade (como uma workload ativa/ativa em vários locais). Se um ou mais arquivos forem excluídos ou corrompidos no dispositivo de armazenamento primário, essas alterações destrutivas poderão ser replicadas para o dispositivo de armazenamento secundário. Nesse cenário, apesar da alta disponibilidade, a capacidade de failover em caso de exclusão ou corrupção de dados será comprometida. Em vez disso, um backup em um ponto anterior no tempo também é necessário como parte da estratégia de DR.

Plano de continuidade dos negócios

Seu plano de recuperação de desastres deve ser um subconjunto do plano de continuidade dos negócios (BCP) de sua organização, e não um documento independente. Não faz sentido manter metas agressivas de recuperação de desastres para restaurar uma workload se os objetivos empresariais referentes a essa workload não puderem ser alcançados devido ao impacto do desastre em outros elementos de seus negócios que não sejam sua workload. Por exemplo, um terremoto pode impedir você de transportar produtos comprados em sua aplicação de comércio eletrônico. Mesmo que uma recuperação de desastres eficaz mantenha sua workload funcionando, seu BCP precisa atender às necessidades de transporte. Sua estratégia de recuperação de desastres deve se basear nos requisitos, nas prioridades e no contexto dos negócios.

Análise de impacto sobre os negócios e avaliação de riscos

Uma análise de impacto sobre os negócios deve quantificar o impacto comercial de uma interrupção em suas workloads. Essa análise deve identificar o impacto sobre os clientes internos e externos, pelo fato de não conseguirem usar suas workloads, e o efeito que isso tem sobre seus negócios. Ela deve ajudar a determinar a rapidez com que a workload precisa ser disponibilizada e a quantidade de perda de dados tolerável. No entanto, é importante observar que os objetivos de recuperação não devem ser realizados isoladamente. A probabilidade de interrupção e o custo da recuperação são fatores-chave que ajudam a revelar o valor empresarial de fornecer recuperação de desastres para uma workload.

O impacto sobre os negócios pode ser uma questão de tempo. É aconselhável levar isso em consideração em seu planejamento de recuperação de desastres. Por exemplo, a interrupção de seu sistema de folha de pagamento provavelmente terá um impacto muito maior sobre os negócios um pouco antes de todos serem pagos, mas pode ter pouco impacto logo depois que todos já tiverem sido pagos.

Uma avaliação de riscos em relação a desastres e impacto geográfico, associada a uma visão geral sobre a implementação técnica de sua workload, determinará a probabilidade de ocorrência de interrupção para cada tipo de desastre.

Para workloads extremamente essenciais, você pode manter a alta disponibilidade em várias regiões com backups contínuos para minimizar o impacto sobre os negócios. Para workloads menos essenciais, uma estratégia válida pode ser não implementar nenhuma recuperação de desastres.

Além disso, com relação a algumas circunstâncias de desastre, também é válido não ter nenhuma estratégia de recuperação de desastres em vigor que se fundamente em uma baixa probabilidade de ocorrência de desastre. Lembre-se de que as zonas de disponibilidade dentro de uma região da AWS já foram projetadas com uma distância significativa entre elas e um planejamento cuidadoso de localização, para que os desastres mais comuns afetem apenas uma zona e não as demais. Portanto, uma arquitetura multi-AZ em uma região da AWS talvez já atenda às suas necessidades de atenuação de riscos.

O custo das opções de recuperação de desastres deve ser avaliado para garantir que a estratégia para isso forneça o nível correto de valor empresarial com base no risco e impacto comerciais.

Com todas essas informações, você pode documentar a ameaça, o risco, o impacto e o custo de diferentes cenários de desastre e as opções de recuperação correspondentes. Essas informações devem ser usadas para determinar seus objetivos de recuperação para cada uma de suas workloads.

Objetivos de recuperação (RTO e RPO)

Ao criar uma estratégia de recuperação de desastres (DR), as organizações geralmente planejam o objetivo de tempo de recuperação (RTO) e o objetivo de ponto de recuperação (RPO).

How much data can you afford to recreate or lose?

**How quickly must you recover?
What is the cost of downtime?**

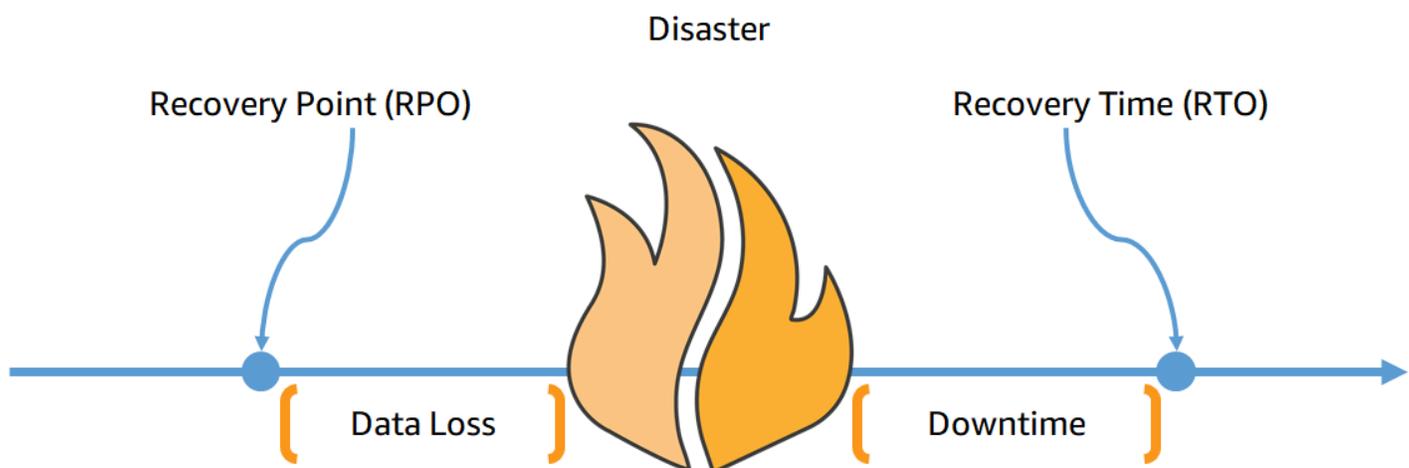


Figura 3: objetivos de recuperação

Objetivo de tempo de recuperação (RTO) refere-se ao atraso máximo aceitável entre a interrupção e a restauração de um serviço. Esse objetivo é definido pela organização e determina o que é considerado uma janela de tempo aceitável quando o serviço está indisponível.

Quatro estratégias de DR são discutidas neste documento: backup e restauração, luz piloto, standby passivo e ativo/ativo em vários locais (consulte [Opções de recuperação de desastres na nuvem](#)). No diagrama a seguir, a empresa determinou seu RTO máximo permissível, bem como o limite ela pode gastar em sua estratégia de restauração de serviços. Em vista dos objetivos da empresa, as estratégias de DR luz piloto e standby passivo atenderão ao RTO e aos critérios de custo.

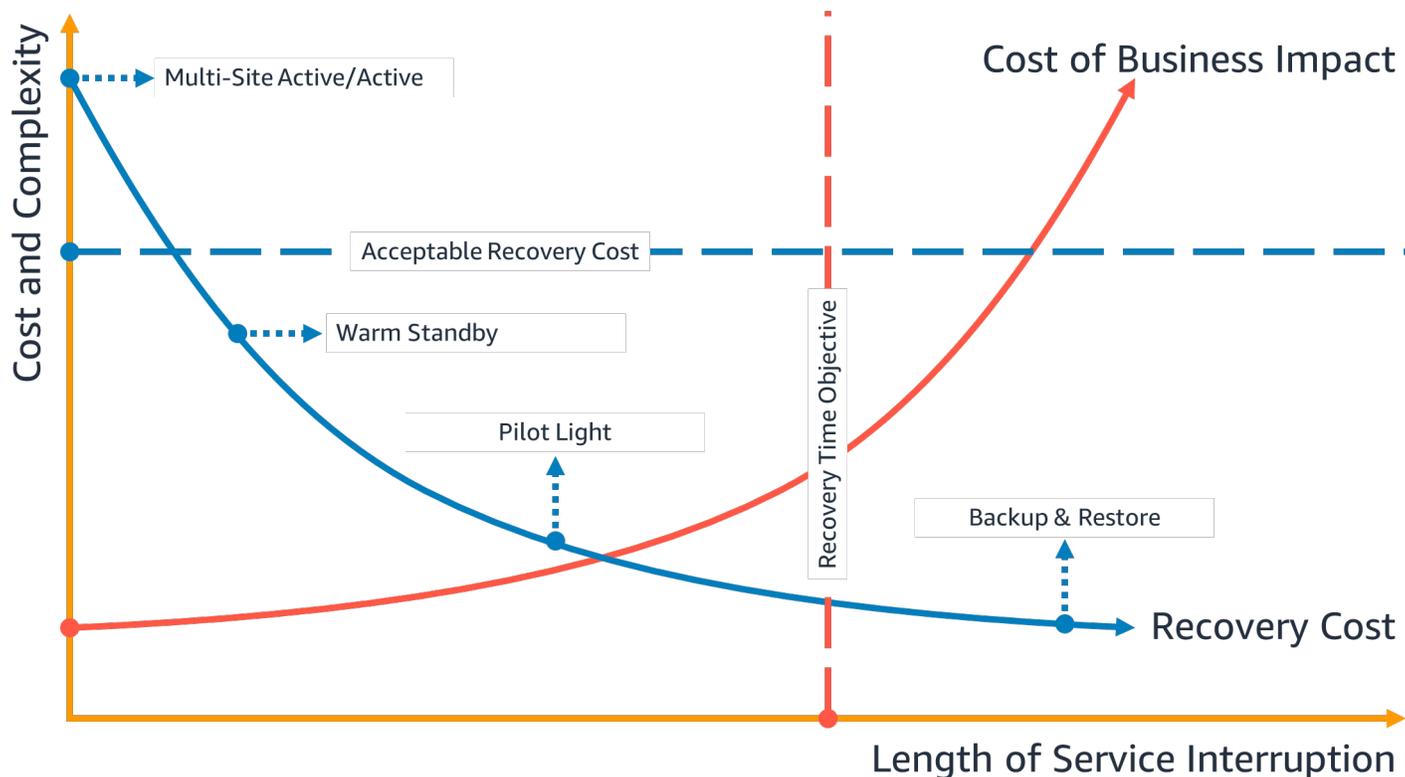


Figura 4: objetivo de tempo de recuperação

Objetivo de ponto de recuperação (RPO) refere-se ao tempo máximo aceitável desde o último ponto de recuperação de dados. Esse objetivo é definido pela organização e determina o que é considerado uma perda de dados aceitável entre o último ponto de recuperação e a interrupção do serviço.

No diagrama a seguir, a empresa determinou seu RPO máximo permissível, bem como o limite do que ela pode gastar em sua estratégia de recuperação de dados. Das quatro estratégias de DR, a de luz piloto ou a de standby passivo atende aos critérios de RPO e custo.

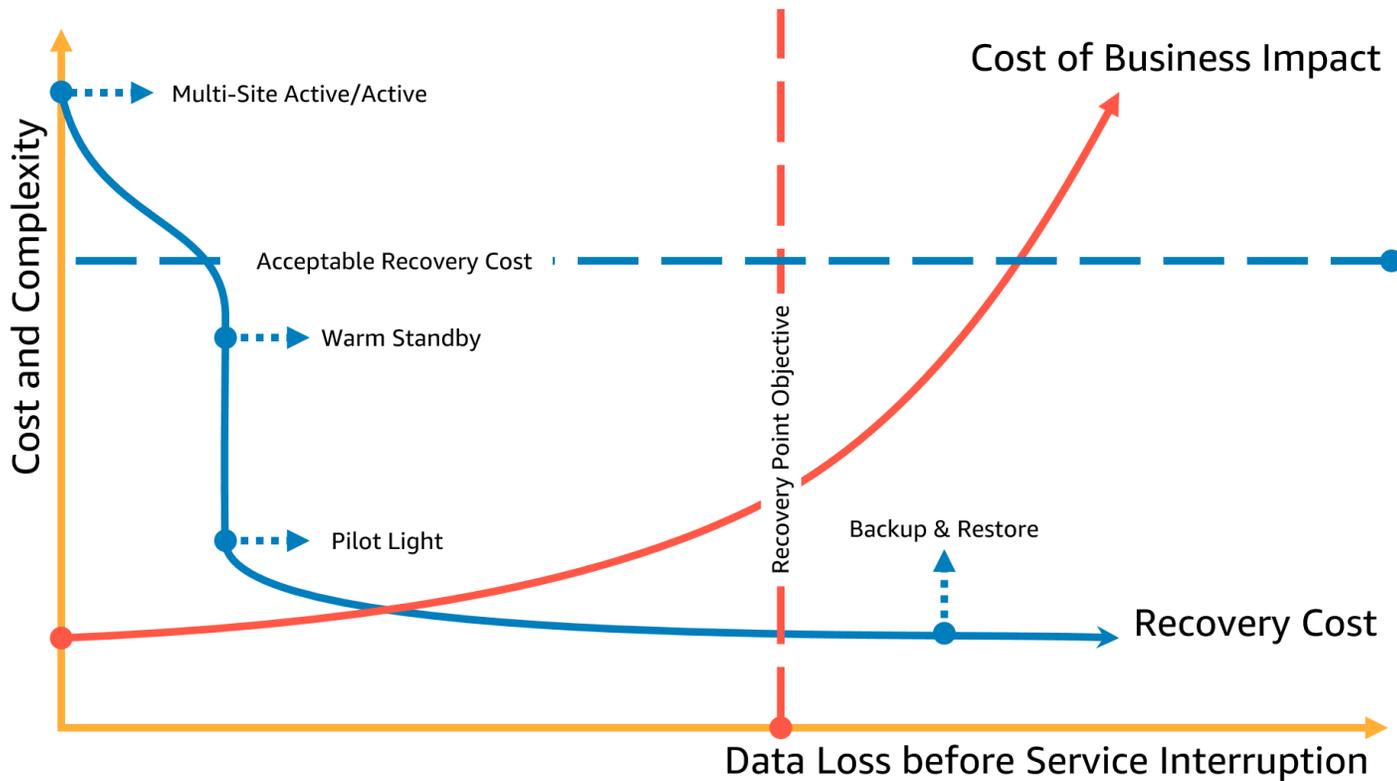


Figura 5: objetivo de ponto de recuperação

Note

Se o custo da recuperação for maior do que o custo da falha ou perda, a opção de recuperação não deve ser implementada, a menos que haja um determinante secundário, como requisitos regulatórios.

A recuperação de desastres é diferente na nuvem

As estratégias de recuperação de desastres evoluem com a inovação técnica. Um plano de recuperação de desastres on-premises pode envolver o transporte físico de fitas ou a replicação de dados para outro local. Sua organização precisa reavaliar o impacto sobre os negócios, o risco e o custo de suas estratégias anteriores de recuperação de desastres para cumprir seus objetivos de DR na AWS. A recuperação de desastres na Nuvem AWS inclui as seguintes vantagens em relação aos ambientes tradicionais:

- Possibilidade de recuperar-se rapidamente de um desastre sem muita complexidade.
- Testes simples e repetíveis que permitem que você teste com maior facilidade e frequência.
- Menor sobrecarga de gerenciamento, o que reduz a carga operacional.
- Oportunidades para automatizar, diminuir as chances de erro e melhorar o tempo de recuperação.

A AWS permite que você troque as despesas de capital fixo de um datacenter de backup físico pelas despesas operacionais variáveis de um ambiente do tamanho certo na nuvem, o que pode reduzir significativamente os custos.

Para muitas organizações, a recuperação de desastres on-premises baseava-se no risco de interrupção de uma workload ou de workloads em um datacenter e na recuperação de dados copiados ou replicados para um datacenter secundário. Quando as organizações implantam workloads na AWS, elas podem implementar uma workload bem projetada e contar com o design da infraestrutura de nuvem global da AWS para ajudar a atenuar o efeito dessas interrupções. Consulte o whitepaper [Pilar Confiabilidade: AWS Well-Architected Framework](#) para obter mais informações sobre as práticas recomendadas de arquitetura para projetar e operar workloads confiáveis, seguras, eficientes e econômicas na nuvem.

Se suas workloads estiverem na AWS, você não precisará se preocupar com conectividade de datacenter (com exceção de sua capacidade de acessá-lo), energia, ar condicionado, hardware e supressão de incêndio. Além de tudo isso ser gerenciado para você, é possível ter acesso a várias zonas de disponibilidade isoladas de falhas (cada uma composta de um ou mais datacenters separados).

Região única da AWS

Para um evento de desastre que envolva interrupção ou perda de um único datacenter físico, a implementação de uma workload com alta disponibilidade em várias zonas de disponibilidade em uma única região da AWS ajuda a atenuar desastres naturais e técnicos e reduz o risco de ameaças humanas, como erros ou atividades não autorizadas que podem provocar a perda de dados. Cada região da AWS é composta de várias zonas de disponibilidade e cada uma delas está isolada de falhas nas outras zonas. Cada zona de disponibilidade, por sua vez, consiste em vários datacenters físicos. Para isolar melhor os problemas impactantes e obter alta disponibilidade, você pode particionar as workloads em várias zonas na mesma região. As zonas de disponibilidade são projetadas para redundância física e fornecem resiliência, o que permite uma performance ininterrupta, mesmo em caso de quedas no fornecimento de energia, tempo de inatividade da Internet, inundações e outros desastres naturais. Consulte [Infraestrutura de nuvem global AWS](#) para descobrir como a AWS faz isso.

Com a implantação em várias zonas de disponibilidade em uma única região da AWS, sua workload fica mais protegida contra falhas de um único (ou mesmo vários) datacenters. Para ter uma garantia extra em sua implantação em uma única região, você pode fazer backup dos dados e da configuração (bem como da definição da infraestrutura) para outra região. Essa estratégia reduz o escopo de seu plano de recuperação de desastres para incluir apenas backup e restauração de dados. Utilizar a resiliência multirregional fazendo backup para outra região da AWS é simples e barato em relação às outras opções multirregionais descritas na seção a seguir. Por exemplo, o backup para o [Amazon Simple Storage Service \(Amazon S3\)](#) oferece acesso à recuperação imediata de seus dados. No entanto, se sua estratégia de DR para parcelas de seus dados tiver requisitos de tempo de recuperação mais flexíveis (de minutos para horas), o uso do [Amazon S3 Glacier](#) ou [Amazon S3 Glacier Deep Archive](#) reduzirá significativamente os custos de sua estratégia de backup e recuperação.

Algumas workloads podem ter requisitos regulatórios de residência de dados. Se isso se aplicar à sua workload em uma localidade que no momento tem apenas uma região da AWS, além de projetar workloads multi-AZ para alta disponibilidade, conforme discutido anteriormente, você também pode usar as AZs dentro dessa região como locais distintos, o que pode ser útil para atender aos requisitos de residência de dados aplicável à sua workload dentro dessa região. As estratégias de DR descritas nas seções a seguir usam várias regiões da AWS, mas também podem ser implementadas usando zonas de disponibilidade em vez de regiões.

Várias regiões da AWS

Para um evento de desastre que apresenta o risco de perda de vários datacenters a uma distância significativa um do outro, você deve considerar opções de recuperação para atenuar desastres naturais e técnicos que afetem uma região inteira na AWS. Todas as opções descritas nas seções a seguir podem ser implementadas como arquiteturas multirregionais para proteção contra esses desastres.

Opções de recuperação de desastres na nuvem

As estratégias de recuperação de desastres disponíveis para você na AWS podem ser amplamente categorizadas em quatro abordagens, que vão desde baixo custo e baixa complexidade para fazer backups a estratégias mais complexas que envolvem o uso de várias regiões ativas. É fundamental testar regularmente sua estratégia de recuperação de desastres para que tenha confiança em invocá-la caso seja necessário.

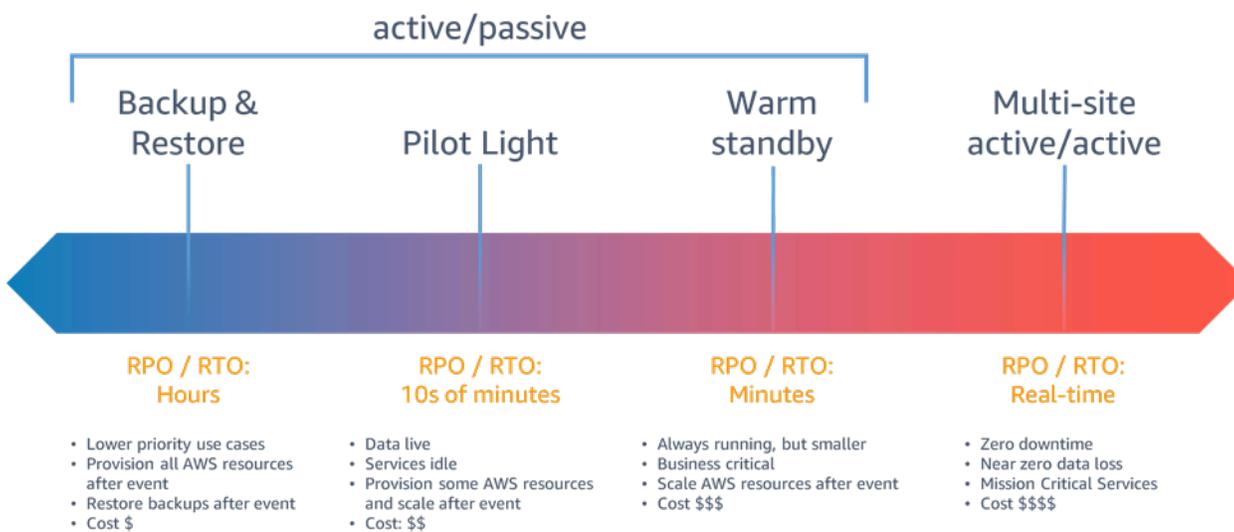


Figura 6: estratégias de recuperação de desastres

Para um evento de desastre que envolva interrupção ou perda de um único datacenter físico para uma workload [bem projetada](#) e altamente disponível, você pode precisar apenas de uma abordagem de backup e restauração para a recuperação de desastres. Se sua definição de desastre for além da interrupção ou perda de um datacenter físico, mas de uma região, ou se você estiver sujeito a requisitos regulatórios que exijam uma estratégia, considere o modo luz piloto, standby passivo ou ativo/ativo em vários locais.

Backup e restauração

Backup e restauração são uma abordagem adequada para atenuar a perda ou corrupção de dados. Essa abordagem também pode ser usada para mitigar um desastre regional por meio da replicação de dados para outras regiões da AWS ou para atenuar a falta de redundância para workloads implantadas em uma única zona de disponibilidade. Além dos dados, você deve reimplantar a infraestrutura, a configuração e o código da aplicação na região de recuperação. Para permitir que

a infraestrutura seja reimplantada rapidamente sem erros, você deve sempre utilizar a infraestrutura como código (IaC) na implantação usando serviços como o [AWS CloudFormation](#) ou o [AWS Cloud Development Kit \(AWS CDK\)](#). Sem a IaC, pode ser difícil restaurar workloads na região de recuperação. Com isso, você pode ter tempos de recuperação maiores e, possivelmente, ultrapassar seu RTO. Além dos dados do usuário, você também deve fazer backup do código e da configuração, bem como das [imagens de máquina da Amazon \(AMIs\)](#) usadas para criar instâncias do Amazon EC2. Você pode usar o [AWS CodePipeline](#) para automatizar a reimplantação do código e da configuração da aplicação.

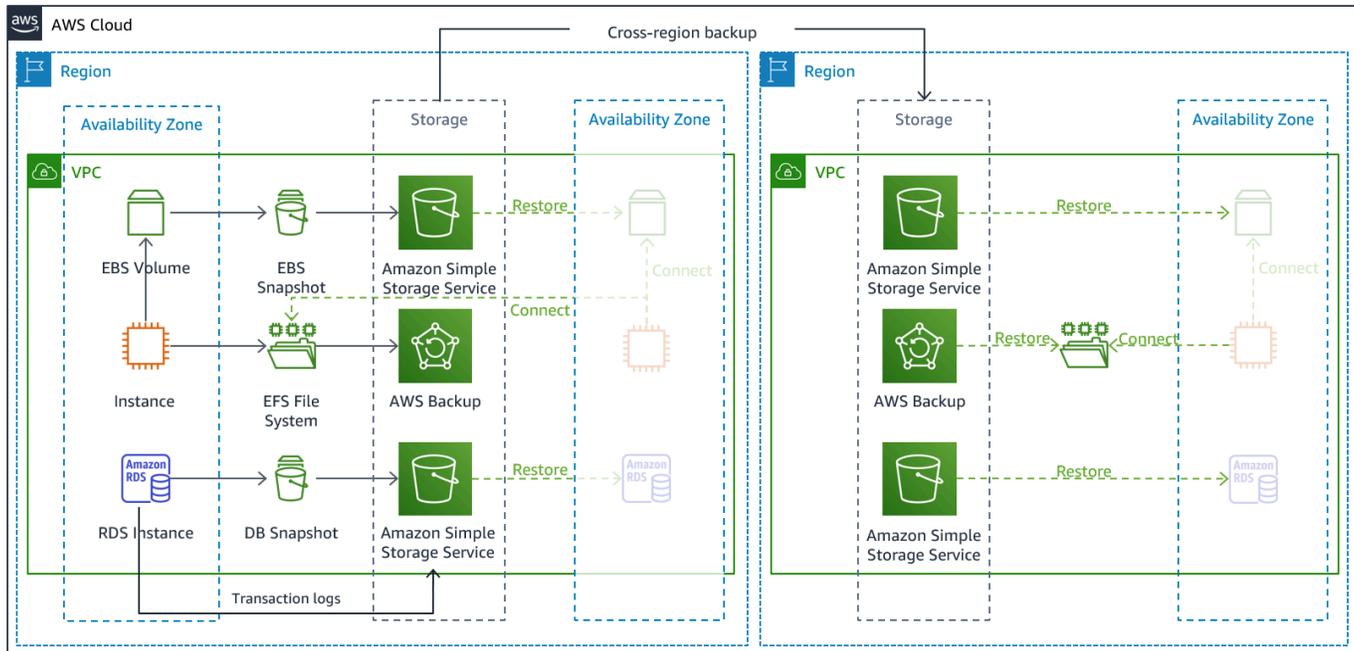


Figura 7: arquitetura de backup e restauração

Serviços da AWS

Os dados da workload exigirão uma estratégia de backup que seja executada periodicamente ou seja contínua. A frequência com que você executa o backup determinará o ponto de recuperação alcançável (que deve estar alinhado para atender ao seu RPO). O backup também deve oferecer um meio para ser restaurado até o ponto em que foi realizado. O backup com recuperação em um ponto anterior no tempo é disponibilizado por meio dos seguintes serviços e recursos:

- [Snapshot do Amazon Elastic Block Store \(Amazon EBS\)](#)
- [Backup do Amazon DynamoDB](#)
- [Snapshot do Amazon RDS](#)

- [Snapshot do banco de dados do Amazon Aurora](#)
- [Backup do Amazon EFS](#) (ao usar o AWS Backup)
- [Snapshot do Amazon Redshift](#)
- [Snapshot do Amazon Neptune](#)

Para o Amazon Simple Storage Service (Amazon S3), você pode usar a [replicação entre regiões \(CRR\) do Amazon S3](#) para copiar objetos continuamente de forma assíncrona para um bucket do S3 na região de DR e, ao mesmo tempo, fornecer versionamento para os objetos armazenados para que possa escolher o ponto de restauração. A replicação contínua de dados tem a vantagem de oferecer o menor tempo de backup (quase zero), mas pode não oferecer proteção contra eventos de desastre, como corrupção de dados ou ataques mal-intencionados (por exemplo, exclusão não autorizada de dados), bem como backups em um ponto anterior no tempo. A replicação contínua é abordada na seção [Serviços da AWS para luz piloto](#).

O [AWS Backup](#) fornece um local centralizado para configurar, programar e monitorar os recursos de backup da AWS para os seguintes serviços e recursos:

- Volumes do [Amazon Elastic Block Store \(Amazon EBS\)](#)
- Instâncias do [Amazon EC2](#)
- Bancos de dados do [Amazon Relational Database Service \(Amazon RDS\)](#) (incluindo bancos de dados do [Amazon Aurora](#))
- Tabelas do [Amazon DynamoDB](#)
- Sistemas de arquivos do [Amazon Elastic File System \(Amazon EFS\)](#)
- Volumes do [AWS Storage Gateway](#)
- [Amazon FSx for Windows File Server](#) e [Amazon FSx for Lustre](#)

Com o AWS Backup, é possível fazer cópia de backups entre regiões; por exemplo, para uma região de recuperação de desastres.

Para ter outra estratégia de recuperação de desastres para seus dados do Amazon S3, habilite o [versionamento de objetos do S3](#). O versionamento de objetos protege seus dados no S3 contra as consequências das ações de exclusão ou modificação, mantendo a versão original antes da ação. O versionamento de objetos pode ser um método útil para atenuar desastres provocados por erros humanos. Se você estiver usando a replicação do S3 para fazer backup de dados em sua região de DR, quando um objeto for excluído no bucket de origem, o [Amazon S3 adicionará um marcador de](#)

[exclusão somente no bucket de origem](#) por padrão. Essa abordagem protege os dados na região de DR contra exclusões mal-intencionadas na região de origem.

Além dos dados, você também deve fazer backup da configuração e da infraestrutura necessárias para reimplantar sua workload e atender ao objetivo de tempo de recuperação (RTO). O [AWS CloudFormation](#) fornece infraestrutura como código (IaC) e permite que você defina todos os recursos da AWS em sua workload para que possa realizar implantações e reimplementações de forma confiável em várias contas e regiões da AWS. Você pode fazer backup das instâncias do Amazon EC2 usadas por sua workload como imagens de máquina da Amazon (AMIs). A AMI é criada de snapshots do volume raiz da instância e de quaisquer outros volumes do EBS anexados à instância. Você pode usar essa AMI para executar uma versão restaurada da instância do EC2. É [possível copiar uma AMI](#) dentro ou entre regiões. Ou você pode usar o [AWS Backup](#) para copiar backups entre contas e para outras regiões da AWS. O recurso de backup entre contas ajuda a oferecer proteção contra eventos de desastres que incluem ameaças internas ou comprometimento da conta. O AWS Backup também adiciona outros recursos para backup do EC2, além dos volumes individuais do EBS da instância. Além disso, o AWS Backup armazena e monitora os seguintes metadados: tipo de instância, nuvem privada virtual (VPC) configurada, grupo de segurança, [função do IAM](#), configuração de monitoramento e etiquetas. No entanto, esses metadados adicionais só são usados na restauração do backup do EC2 para a mesma região da AWS.

Todos os dados armazenados na região de recuperação de desastres como backups devem ser restaurados no momento do failover. O AWS Backup oferece recurso de restauração, mas atualmente não permite restauração programada ou automática. Você pode implementar a restauração automática para a região de DR usando o AWS SDK para chamar APIs para o AWS Backup. Você pode definir essa configuração como um trabalho regularmente recorrente ou acionar a restauração sempre que um backup for concluído. A figura a seguir mostra um exemplo de restauração automática usando o [Amazon Simple Notification Service \(Amazon SNS\)](#) e o [AWS Lambda](#).

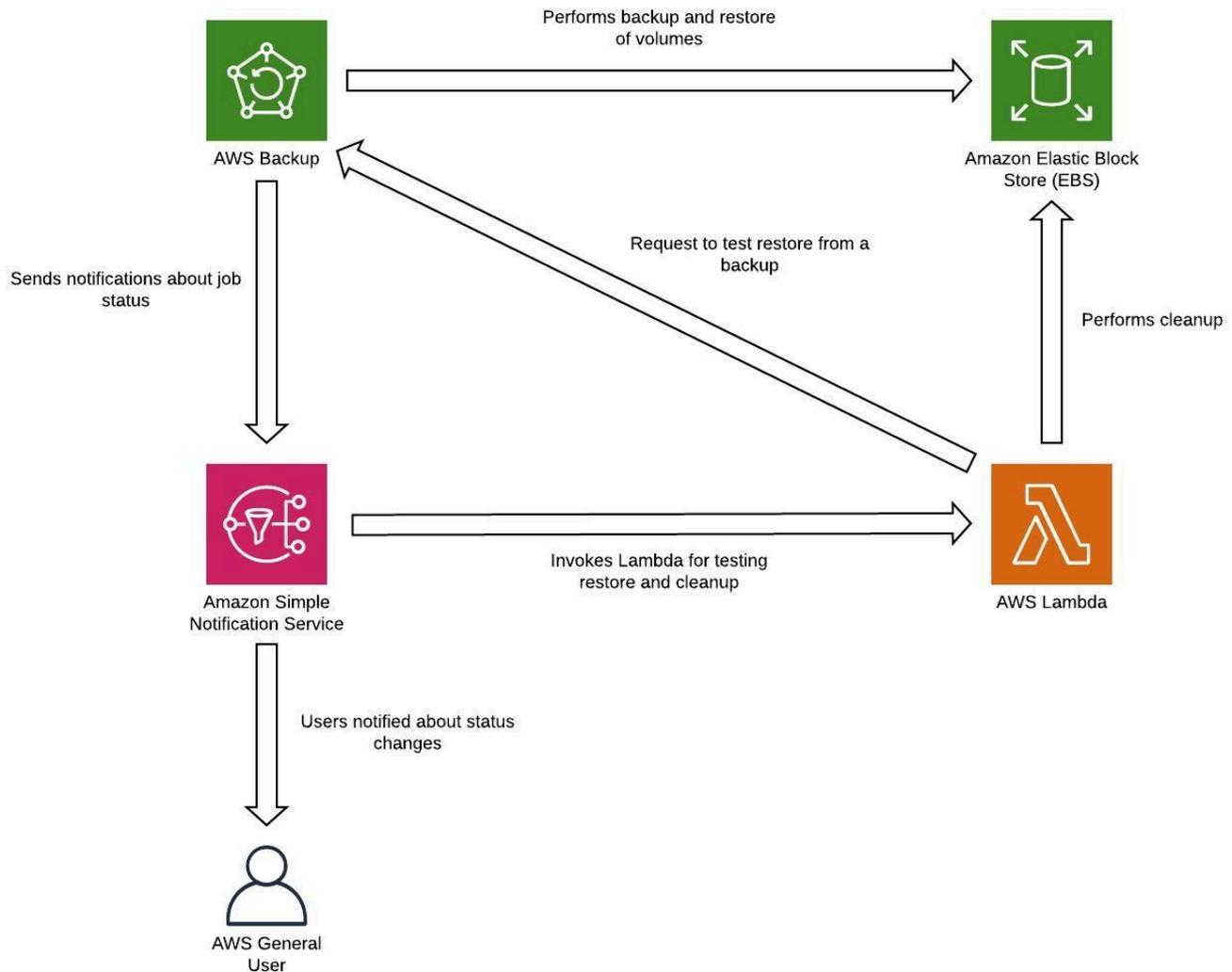


Figura 8: restauração e teste de backups

Note

Sua estratégia de backup deve incluir o teste de seus backups. Consulte a seção [Testes de recuperação de desastres](#) para obter mais informações. Consulte [AWS Well-Architected Lab: Testing Backup and Restore of Data](#) para obter uma demonstração prática da implementação.

Luz piloto

Com a abordagem de luz piloto, você replica seus dados de uma região para outra e provisiona uma cópia de sua infraestrutura de workload principal. Os recursos necessários para auxiliar a replicação e o backup de dados, como bancos de dados e armazenamento de objetos, estão sempre ativos. Outros elementos, como servidores de aplicações, são carregados com o código e as configurações da aplicação, mas são desativados e usados apenas durante testes ou quando o failover de recuperação de desastres é invocado. Ao contrário da abordagem de backup e restauração, sua infraestrutura principal está constantemente disponível e você sempre tem a opção de provisionar rapidamente um ambiente de produção em grande escala ativando e aumentando a escala na horizontal dos servidores de aplicações.

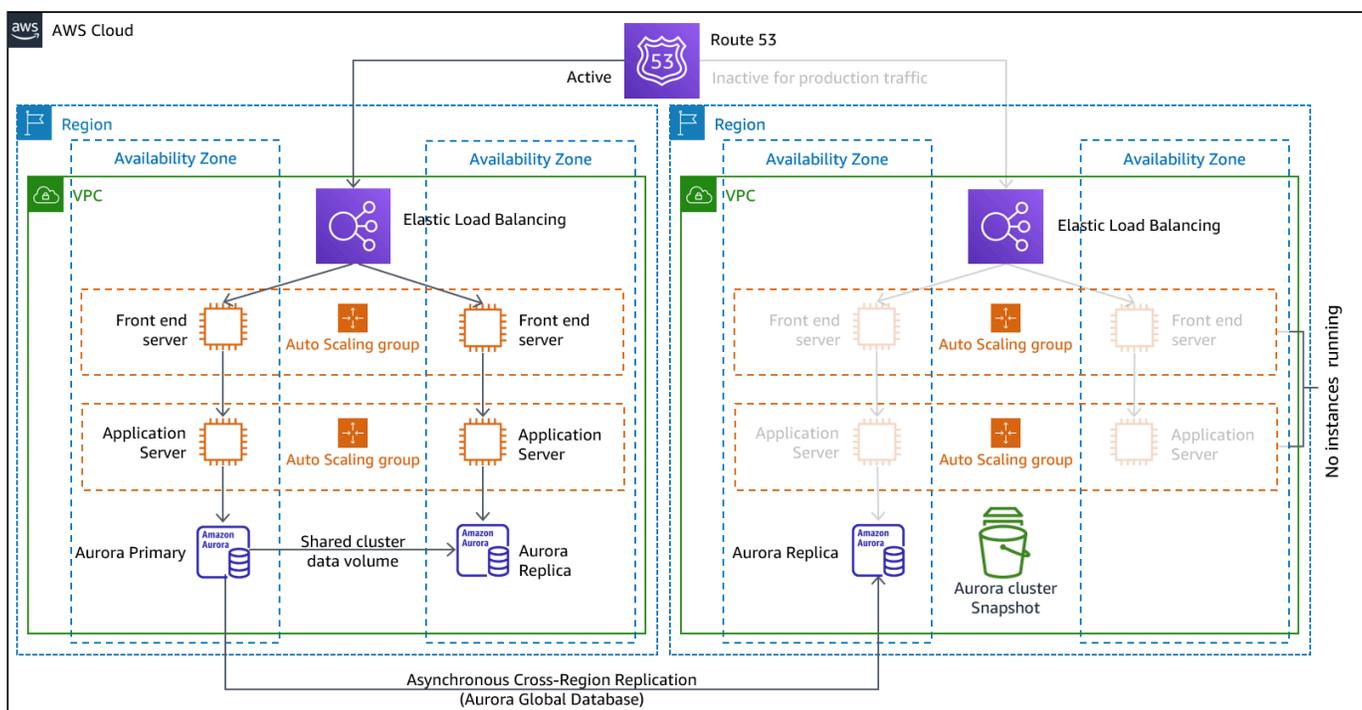


Figura 9: arquitetura de luz piloto

A abordagem de luz piloto reduz o custo contínuo da recuperação de desastres ao minimizar os recursos ativos e simplifica a recuperação no momento de um desastre porque todos os principais requisitos de infraestrutura estão em ordem. Essa opção de recuperação exige que você altere sua abordagem de implantação. Você precisa fazer alterações na infraestrutura principal em cada região e implantar alterações de workload (configuração, código) simultaneamente em cada região. Para simplificar essa etapa, você pode automatizar suas implantações e usar a infraestrutura como código (IaC) para implantar infraestrutura em várias contas e regiões (implantação completa

da infraestrutura na região principal e implantação de infraestrutura reduzida na escala vertical/desativada para regiões de DR). É recomendável usar uma conta diferente por região para fornecer o mais alto nível de isolamento de recursos e segurança (caso seus planos de recuperação de desastres também incluam credenciais comprometidas).

Com essa abordagem, você também precisa atenuar desastres de dados. A replicação contínua de dados protege você contra alguns tipos de desastre, mas pode não oferecer proteção contra corrupção ou destruição de dados, a menos que sua estratégia também inclua versionamento de dados armazenados ou opções para recuperação em um ponto anterior no tempo. Você pode fazer backup dos dados replicados na região do desastre para criar backups em um ponto anterior no tempo nessa mesma região.

Serviços da AWS

Além de usar os serviços da AWS abordados na seção [Backup e restauração](#) para criar backups em um ponto anterior no tempo, considere também os seguintes serviços para sua estratégia de luz piloto.

Com relação à luz piloto, a replicação contínua de dados para bancos de dados e armazenamentos de dados em tempo real na região de DR é a melhor abordagem para RPO baixo (quando usada além dos backups em um ponto anterior no tempo discutidos anteriormente). A AWS fornece replicação de dados contínua, entre regiões e assíncrona para dados por meio dos seguintes serviços e recursos:

- [Replicação do Amazon Simple Storage Service \(Amazon S3\)](#)
- [Réplicas de leitura do Amazon RDS](#)
- [Amazon Aurora Global Database](#)
- [Tabelas globais do Amazon DynamoDB](#)

Com a replicação contínua, as versões de seus dados ficam disponíveis quase imediatamente em sua região de DR. Os tempos reais de replicação podem ser monitorados usando recursos de serviço como o [S3 Replication Time Control \(S3 RTC\)](#) para objetos do S3 e [recursos de gerenciamento do Amazon Aurora Global Database](#).

Ao fazer o failover para executar sua workload de leitura e gravação na região de recuperação de desastres, você deve promover uma réplica de leitura do RDS para que se torne a instância primária. Para [instâncias de banco de dados diferentes do Aurora, o processo](#) leva alguns minutos para ser concluído e a reinicialização faz parte do processo. Para replicação entre regiões (CRR) e failover

com o RDS, o uso do [Amazon Aurora Global Database](#) oferece várias vantagens. O banco de dados global usa uma infraestrutura dedicada que mantém seus bancos de dados totalmente disponíveis para atender à sua aplicação e pode replicar para a região secundária com latência típica de menos de 1 segundo (e com uma latência bem menor que 100 milissegundos dentro de uma região da AWS). Com o Amazon Aurora Global Database, se sua região principal sofrer uma degradação de performance ou interrupção, você poderá promover uma das regiões secundárias para assumir responsabilidades de leitura/gravação em menos de 1 minuto, mesmo no caso de uma paralisação regional completa. A promoção pode ser automática e não há reinicialização.

Uma versão de sua infraestrutura de workload principal reduzida na escala vertical, com menos recursos ou recursos menores, deve ser implantada em sua região de DR. Usando o AWS CloudFormation, você pode definir sua infraestrutura e implantá-la de forma consistente nas contas e regiões da AWS. O AWS CloudFormation usa [pseudoparâmetros](#) predefinidos para identificar a conta da AWS e a região da AWS em que ela está implantada. Portanto, você pode implementar a [lógica de condição em seus modelos do CloudFormation](#) para implantar na região de recuperação de desastres somente a versão de sua infraestrutura reduzida na escala vertical. Para implantações de instâncias do EC2, uma imagem de máquina da Amazon (AMI) fornece informações como configuração de hardware e software instalado. Você pode implementar um pipeline do [Image Builder](#) que cria as AMIs necessárias e copiá-las para as regiões principal e de backup. Isso ajuda a garantir que essas AMIs ouro tenham tudo o que você precisa para reimplantar ou expandir sua workload em uma nova região em caso de desastre. As instâncias do Amazon EC2 são implantadas em uma configuração reduzida na escala vertical (menos instâncias do que em sua região principal). Você pode usar o [hibernate](#) para colocar instâncias do EC2 em um estado interrompido, no qual você não paga pelos custos do EC2, mas apenas pelo armazenamento usado. Para iniciar instâncias do EC2, você pode criar scripts usando a [Command Line Interface \(AWS CLI\)](#) ou o [AWS SDK](#). Para dimensionar a infraestrutura para comportar o tráfego de produção, consulte [AWS Auto Scaling](#) na seção [Standby passivo](#).

Para uma configuração ativo/standby, como a luz piloto, a princípio todo tráfego vai para a região principal e muda para a região de recuperação de desastres se a região principal não estiver mais disponível. Há duas opções de gerenciamento de tráfego a serem consideradas ao usar os serviços da AWS. A primeira é usar o [Amazon Route 53](#). Com o [Amazon Route 53](#), você pode associar vários endpoints de IP em uma ou mais regiões da AWS a um nome de domínio do Route 53. Em seguida, você pode encaminhar o tráfego para o endpoint apropriado com esse nome de domínio. [As verificações de integridade do Amazon Route 53](#) monitoram esses endpoints. Usando essas verificações de integridade, você pode configurar o [failover de DNS](#) para garantir que o tráfego seja enviado para endpoints íntegros.

A segunda opção é usar o [AWS Global Accelerator](#). Usando o AnyCast IP, você pode associar vários endpoints em uma ou mais regiões da AWS com o(s) mesmo(s) endereço(s) IP estático(s). Em seguida, o AWS Global Accelerator encaminha o tráfego para o devido endpoint associado a esse endereço. [As verificações de integridade do Global Accelerator](#) monitoram endpoints. Usando essas verificações de integridade, o AWS Global Accelerator verifica automaticamente a integridade de suas aplicações e encaminha o tráfego do usuário somente para um endpoint íntegro da aplicação. O Global Accelerator oferece latências mais baixas para o endpoint da aplicação, pois utiliza a extensa rede de borda da AWS para introduzir o tráfego na estrutura da rede da AWS o mais rápido possível. O Global Accelerator também evita problemas de armazenamento em cache que podem ocorrer com sistemas DNS (como o Route 53).

CloudEndure Disaster Recovery

O [CloudEndure Disaster Recovery](#), disponível no [AWS Marketplace](#), replica continuamente para a AWS as aplicações hospedadas em servidor e os bancos de dados hospedados em servidor de qualquer origem usando replicação em nível de bloco do servidor subjacente. O CloudEndure Disaster Recovery permite que você use a Nuvem AWS como uma região de recuperação de desastres para uma workload on-premises e o respectivo ambiente. Ele também pode ser usado para recuperação de desastres de workloads hospedadas na AWS se elas consistirem apenas em aplicações e bancos de dados hospedados no EC2 (ou seja, não no RDS). O CloudEndure Disaster Recovery usa a estratégia de luz piloto e mantém uma cópia de dados e recursos desativados em uma Amazon Virtual Private Cloud (Amazon VPC) usada como área de preparação. Quando um evento de failover é acionado, os recursos preparados são usados para criar automaticamente uma implantação de capacidade total na Amazon VPC de destino usada como local de recuperação.

Figura 10: arquitetura do CloudEndure Disaster Recovery

Standby passivo

Na abordagem de standby passivo, é necessário garantir que haja uma cópia reduzida na escala vertical, mas totalmente funcional, de seu ambiente de produção em outra região. Essa abordagem amplia o conceito de luz piloto e diminui o tempo de recuperação porque sua workload fica sempre ativa em outra região. Ela também permite que você realize testes com maior facilidade ou implemente testes contínuos para aumentar a confiança em sua capacidade de se recuperar de um desastre.

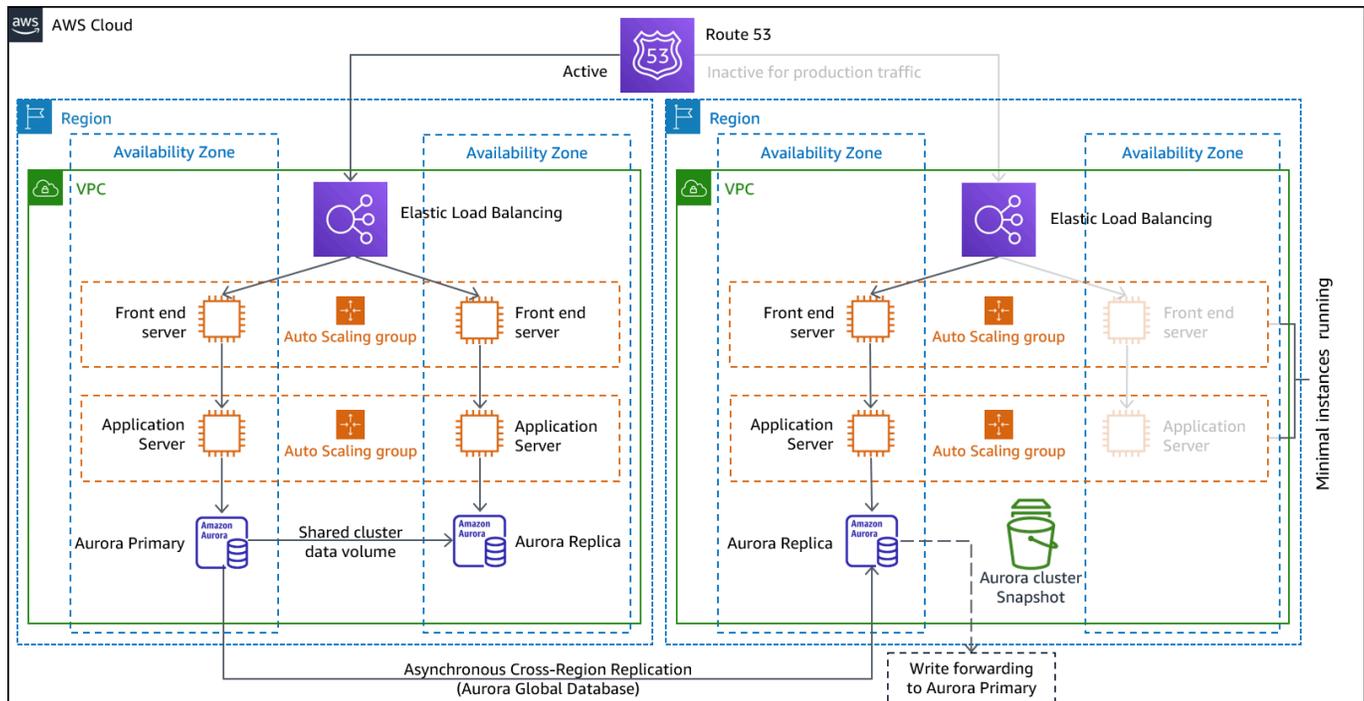


Figura 11: arquitetura de standby passivo

Observação: a diferença entre [luz piloto](#) e [standby passivo](#) às vezes pode ser difícil de entender. Ambos incluem um ambiente em sua região de DR com cópias dos principais ativos da região. A diferença é que a luz piloto não pode processar solicitações sem que outras ações sejam executadas primeiro, enquanto o modo de standby passivo pode lidar imediatamente com o tráfego (em níveis de capacidade reduzidos). A abordagem de luz piloto requer que você “ligue” os servidores, possivelmente implante infraestrutura adicional (não central) e aumente a escala na vertical, enquanto o modo de standby passivo requer apenas que você aumente a escala na vertical (tudo já está implantado e em execução). Use suas necessidades de RTO e RPO para ajudar você a escolher uma dessas abordagens.

Serviços da AWS

Todos os serviços da AWS cobertos por [backup e restauração](#) e [luz piloto](#) também são usados em standby para backup de dados, replicação de dados, roteamento de tráfego ativo/standby e implantação de infraestrutura que inclui instâncias do EC2.

O [AWS Auto Scaling](#) é usado para escalar recursos, incluindo instâncias do Amazon EC2, tarefas do Amazon ECS, taxa de transferência do Amazon DynamoDB e réplicas do Amazon Aurora em uma região da AWS. O [Amazon EC2 Auto Scaling dimensiona](#) escala a implantação da instância

do EC2 nas zonas de disponibilidade dentro de uma região da AWS, fornecendo resiliência dentro dessa região. Como parte de uma estratégia de luz piloto ou standby passivo, use o Auto Scaling para aumentar a escala de sua região de DR na horizontal para a capacidade total de produção. Por exemplo, para o EC2, aumente a configuração da capacidade desejada no grupo do Auto Scaling. Você pode ajustar essa configuração manualmente por meio do AWS Management Console, automaticamente por meio do AWS SDK ou reimplantando seu modelo do AWS CloudFormation usando o novo valor de capacidade desejada. Você pode usar parâmetros do AWS CloudFormation para facilitar a reimplantação do modelo do CloudFormation. Defina as [cotas de serviço](#) em sua região de DR com um valor alto o suficiente para não impedir que você aumente a escala na vertical para a capacidade de produção.

Ativo/ativo em vários locais

Você pode executar sua workload simultaneamente em várias regiões como parte de uma estratégia ativo/ativo em vários locais ou de standby a quente ativo/passivo. A estratégia ativo/ativo em vários locais atende ao tráfego de todas as regiões nas quais está implantada, enquanto o standby a quente atende ao tráfego apenas de uma região, e as outras regiões são usadas somente para recuperação de desastres. Com uma abordagem ativo/ativo em vários locais, os usuários podem acessar sua workload em qualquer uma das regiões em que ela está implantada. Essa é a abordagem mais complexa e cara para a recuperação de desastres, mas pode reduzir o tempo de recuperação para quase zero na maioria dos desastres com as opções de tecnologia e implementação corretas (no entanto, a corrupção de dados pode precisar contar com backups, o que geralmente resulta em um ponto de recuperação diferente de zero). O standby a quente usa uma configuração ativo/passivo em que os usuários são direcionados apenas para uma região e as regiões de DR não recebem tráfego. A maioria dos clientes acredita que, se eles forem criar um ambiente completo na segunda região, faz sentido usá-lo como ativo/ativo. Entretanto, se você não quiser usar as duas regiões para lidar com o tráfego do usuário, o modo de standby a quente é uma abordagem mais econômica e menos complexa do ponto de vista operacional.

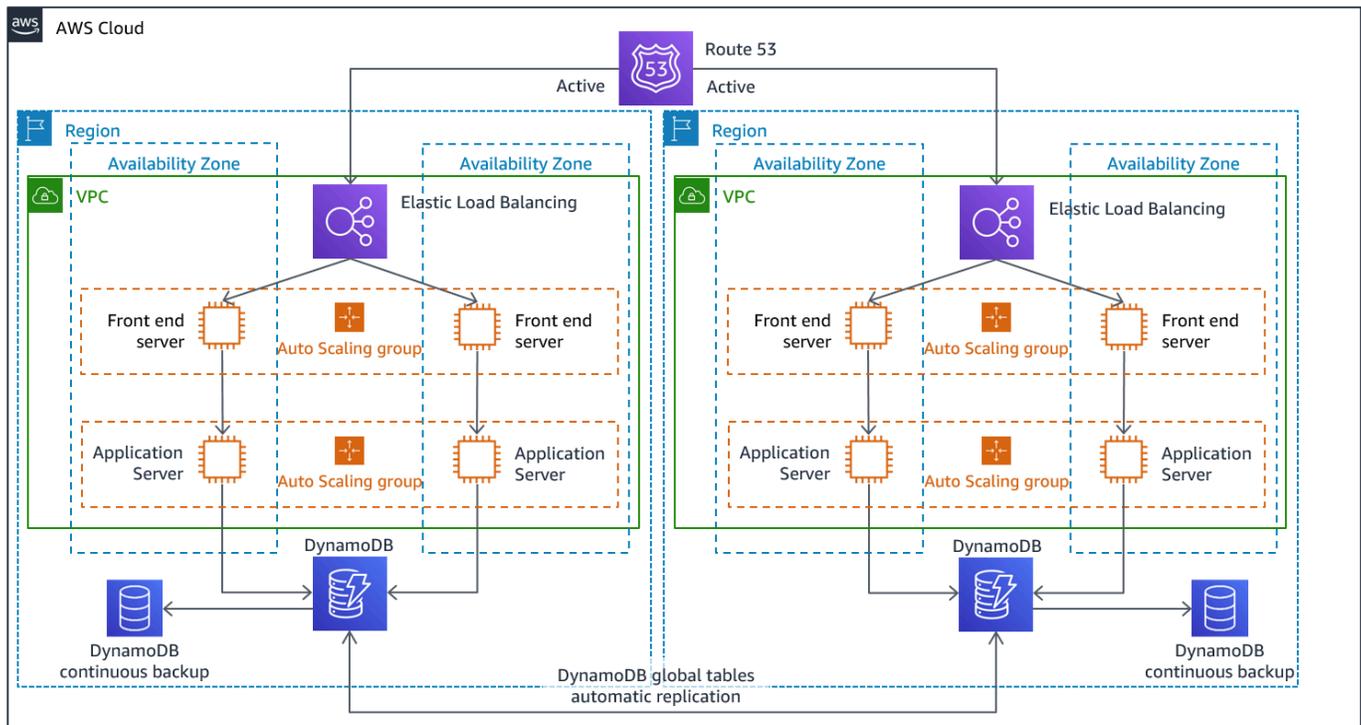


Figura 12: arquitetura ativo/ativa em vários locais (altere um caminho ativo para Inativo para standby a quente)

No cenário ativo/ativo em vários locais, visto que a workload está sendo executada em mais de uma região, não existe failover. Nesse caso, o teste de recuperação de desastres se concentraria em como a workload reage à perda de uma região: o tráfego é encaminhado para fora da região com falha? As outras regiões podem lidar com todo o tráfego? Também é necessário testar um desastre de dados. O backup e a recuperação ainda assim são necessários e devem ser testados regularmente. Também é necessário observar que os tempos de recuperação de um desastre de dados envolvendo corrupção, exclusão ou obscurecimento de dados sempre serão superiores a zero e o ponto de recuperação sempre estará em algum ponto antes da descoberta do desastre. Se a complexidade e custo adicionais de uma abordagem ativo/ativo em vários locais (ou standby a quente) forem necessários para manter tempos de recuperação quase nulos, outros esforços deverão ser feitos para preservar a segurança, bem como evitar erros humanos e atenuar desastres humanos.

Serviços da AWS

Todos os serviços da AWS cobertos por [backup e restauração](#), [luz piloto](#) e [standby passivo](#) também são usados aqui para backup de dados em um ponto anterior no tempo, replicação de

dados, roteamento de tráfego ativo/ativo e implantação e escalabilidade de infraestrutura que inclui instâncias do EC2.

Para os cenários ativo/passivo discutidos anteriormente (luz piloto e standby passivo), o Amazon Route 53 e o AWS Global Accelerator podem ser usados para encaminhar o tráfego de rede para a região ativa. Aqui, para a estratégia ativo/ativo, esses dois serviços também permitem a definição de políticas que determinam quais usuários irão para qual endpoint regional ativo. Com o AWS Global Accelerator, você define uma [discagem de tráfego para controlar a porcentagem de tráfego](#) que é direcionada para cada endpoint da aplicação. O Amazon Route 53 comporta essa abordagem de porcentagem e também [várias outras políticas disponíveis](#), inclusive aquelas baseadas em geoproximidade e latência. O [Global Accelerator utiliza automaticamente a extensa rede de servidores de borda da AWS](#) para introduzir o tráfego na estrutura de rede da AWS o mais rápido possível, o que resulta em latências de solicitação mais baixas.

A replicação de dados com essa estratégia permite um RPO próximo de zero. Serviços da AWS como o [Amazon Aurora Global Database](#) usam uma infraestrutura dedicada que mantém seus bancos de dados totalmente disponíveis para atender à sua aplicação e podem realizar a replicação para uma região secundária com latência típica de menos de um segundo. Com as estratégias ativo/passivo, as gravações ocorrem somente na região primária. Na estratégia ativo/ativo, a diferença é delinear de que forma as gravações em cada região ativa serão tratadas. As leituras do usuário normalmente são projetadas para que sejam realizadas na região mais próxima a eles, o que é conhecido como leitura local. Nas gravações, você tem várias opções:

- Na estratégia de gravação global, todas as gravações são encaminhadas para uma única região. Em caso de falha nessa região, outra região é promovida para aceitar gravações. O [Aurora Global Database](#) é uma boa opção para gravação global porque comporta sincronização com réplicas de leitura entre regiões e você pode promover uma das regiões secundárias para assumir responsabilidades de leitura/gravação em menos de 1 minuto.
- Na estratégia de gravação local, as gravações são encaminhadas para a região mais próxima (assim como as leituras). As [tabelas globais do Amazon DynamoDB](#) possibilitam essa estratégia, permitindo leituras e gravações de todas as regiões em que sua tabela global estiver implantada. As tabelas globais do Amazon DynamoDB usam a reconciliação último gravador ganha entre as atualizações simultâneas.
- Na estratégia de gravação particionada, as gravações são atribuídas a uma região específica com base em uma chave de partição (como ID de usuário) para evitar conflitos de gravação. A replicação do Amazon S3 [configurada bidirecionalmente](#) pode ser usada para esse caso e, no momento, comporta replicação entre duas regiões. Ao implementar essa abordagem, habilite a

[sincronização de modificação de réplica](#) nos buckets A e B para replicar alterações de metadados de réplica como listas de controle de acesso a objetos (ACLs), etiquetas de objeto ou bloqueios de objeto nos objetos replicados. Você também pode configurar se deve ou não [replicar marcadores de exclusão](#) entre buckets em suas regiões ativas. Além da replicação, sua estratégia também deve incluir backups em um ponto anterior no tempo como proteção contra eventos de corrupção ou destruição de dados.

O AWS CloudFormation é uma ferramenta avançada para aplicar a infraestrutura implantada de forma consistente entre contas da AWS em várias regiões da AWS. O [AWS CloudFormation StackSets](#) amplia essa funcionalidade ao permitir que você crie, atualize ou exclua pilhas do CloudFormation em várias contas e regiões com uma única operação. Embora o AWS CloudFormation use YAML ou JSON para definir a infraestrutura como código, o [AWS Cloud Development Kit \(AWS CDK\)](#) permite que você defina a IaC usando linguagens de programação conhecidas. Seu código é convertido para o CloudFormation, que é usado para implantar recursos na AWS.

Detecção

É importante estar ciente o mais rápido possível de que suas workloads não estão entregando os resultados de negócios que deveriam. Dessa forma, você pode declarar rapidamente um desastre e se recuperar de um incidente. Com relação a objetivos de recuperação agressivos, esse tempo de resposta, associado às informações apropriadas, é essencial para atingi-los. Se o objetivo de ponto de recuperação for de uma hora, você precisará detectar o incidente, notificar a equipe apropriada, envolver seus processos de escalonamento, avaliar informações (se houver) sobre o tempo esperado para recuperação (sem executar o plano de DR), declarar o desastre e recuperar-se em uma hora.

Note

Se as partes interessadas decidirem não recorrer à DR, mesmo que o RTO esteja em risco, reavalie os planos e objetivos de DR. A decisão de não recorrer a planos de DR pode ser porque os planos são inadequados ou não há confiança na execução.

É fundamental levar em consideração fatores como detecção, notificação, escalonamento, descoberta e declaração de incidentes em no planejamento e nos objetivos para estabelecer objetivos realistas e alcançáveis que forneçam valor empresarial.

A AWS também publica nossas informações mais recentes sobre disponibilidade de serviços no [Service Health Dashboard](#). Verifique a qualquer momento para obter informações de status atuais ou assine um feed RSS para ser notificado sobre interrupções em cada serviço específico. Se estiver enfrentando um problema operacional em tempo real com um de nossos serviços que não é mostrado no Service Health Dashboard, você pode criar uma [solicitação de suporte](#).

O [AWS Health Dashboard](#) fornece informações sobre eventos do AWS Health que podem afetar sua conta. As informações são apresentadas de duas formas: um painel que mostra eventos recentes e futuros organizados por categoria e um log de eventos completo que mostra todos os eventos dos últimos 90 dias.

Para os requisitos de RTO mais rigorosos, você pode implementar failover automatizado com base em [verificações de integridade](#). Projete verificações de integridade representativas da experiência do usuário e baseadas nos indicadores-chave de performance. As verificações de integridade mais detalhadas realizam a principal funcionalidade de sua workload e vão além das verificações

superficiais de pulsação. Use verificações de integridade detalhadas com base em vários sinais. Tenha cuidado com essa abordagem para não acionar alarmes falsos, pois aplicar failover quando não há necessidade pode, por si só, apresentar riscos de disponibilidade.

Teste de recuperação de desastres

Teste a implementação de recuperação de desastres para validar a implementação e teste regularmente o failover para a região de DR de sua workload para garantir que o RTO e o RPO sejam atendidos.

Um padrão que deve ser evitado é o desenvolvimento de caminhos de recuperação que raramente são executados. Por exemplo, você pode ter um repositório de dados secundário utilizado para consultas somente leitura. Quando você grava em um repositório de dados e o repositório de dados primário falha, pode ser necessário fazer o failover para o repositório de dados secundário. Se você não testar esse failover com frequência, poderá descobrir que suas suposições sobre as capacidades do armazenamento de dados secundário são incorretas. A capacidade do secundário, que talvez tenha sido suficiente quando você testou pela última vez, pode não ser mais capaz de tolerar a carga nesse cenário ou as cotas de serviço na região secundária podem não ser suficientes.

Nossa experiência demonstra que a única recuperação de erro que funciona é o caminho que você testa com frequência. É por isso que é melhor ter um pequeno número de caminhos de recuperação.

Você pode estabelecer padrões de recuperação e testá-los regularmente. Se você tiver um caminho de recuperação complexo ou crítico, ainda precisará executar regularmente essa falha na produção para validar o funcionamento desse caminho.

Gerencie o desvio de configuração na região de DR. Garanta que sua infraestrutura, seus dados e sua configuração estejam de acordo com o que é necessário na região de DR. Por exemplo, verifique se as AMIs e as cotas de serviço estão atualizadas.

Você pode usar o [AWS Config](#) para monitorar e registrar continuamente suas configurações de recursos da AWS. O AWS Config pode detectar desvios e acionar o [AWS Systems Manager Automation](#) para corrigir desvios e acionar alarmes. Além disso, o [AWS CloudFormation](#) pode detectar desvios nas pilhas que você implantou.

Conclusão

Os clientes são responsáveis pela disponibilidade de suas aplicações na nuvem. É importante definir o que é um desastre e ter um plano de recuperação de desastres que reflita essa definição e o impacto que ele pode ter sobre os resultados dos negócios. Crie o objetivo de tempo de recuperação (RTO) e o objetivo de ponto de recuperação (RPO) com base na análise de impacto e nas avaliações de risco e, em seguida, escolha a arquitetura apropriada para atenuar desastres. Para garantir que a detecção de desastres seja possível e oportuna, é vital saber quando os objetivos estão em risco. Tenha um plano e valide-o com testes. Os planos de recuperação de desastres que não foram validados correm o risco de não serem implementados devido à falta de confiança ou ao não cumprimento dos objetivos de recuperação de desastres.

Colaboradores

Os colaboradores deste documento incluem:

- Alex Livingstone, líder de prática de operações na nuvem do AWS Enterprise Support
- Seth Eliot, arquiteto-chefe de soluções de confiabilidade da Amazon Web Services

Leitura adicional

Para obter informações adicionais, consulte:

- [Pilar Confiabilidade: AWS Well-Architected Framework](#)
- [Lista de verificação do plano de recuperação de desastres](#)
- [Implementação de verificações de integridade](#)
- [Implementações de soluções da AWS: Multi-Region Application Architecture](#)
- [AWS re:Invent 2018: Architecture Patterns for Multi-Region Active-Active Applications \(ARC209-R2\)](#)

Histórico do documento

Alteração	Descrição	Data
Publicação inicial	Primeira publicação.	12 de fevereiro de 2021

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

Avisos

Os clientes são responsáveis por fazer sua própria avaliação independente das informações neste documento. Este documento é: (a) fornecido apenas para fins informativos, (b) representa as ofertas e práticas de produtos atuais da AWS, que estão sujeitas a alterações sem aviso prévio e (c) não 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 qualquer tipo, explícitas ou implícitas. As responsabilidades e obrigações da AWS 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.