Usando o Amazon Web Services para recuperação de desastres

Outubro de 2011

Atualizado: janeiro de 2012

Glen Robinson, Ianni Vamvadelis e Attila Narin

Índice

Resumo

Em caso de desastre, você pode iniciar rapidamente os recursos na Amazon Web Services (AWS) para garantir a continuidade dos negócios. O artigo destaca os recursos e serviços relevantes da AWS que você pode utilizar para os seus processos de recuperação de desastre e mostra exemplos de cenários sobre como recuperar do desastre. Fornece ainda recomendações sobre como é possível melhorar seu plano de recuperação de desastre e utilizar o potencial completo da AWS para os seus processos de Recuperação de Desastre.

Introdução

A Recuperação de desastres (DR) envolve preparar-se para poder se recuperar de um desastres. Qualquer evento que tenha um impacto negativo na continuidade dos seus negócios ou finanças poderia ser chamado de um desastre. Isso poderia ser uma falha de hardware ou software, uma interrupção na rede, uma queda de energia, danos físicos a um edifício como fogo ou inundações, erro humano ou algum outro desastre significativo.

Para minimizar o impacto de um desastre nos negócios as empresas investem tempo e recursos para planejar, preparar, ensaiar, documentar, treinar e atualizar processos para lidar com tais eventos. O valor do investimento no planejamento de recuperação de desastres de um determinado sistema pode variar muito, dependendo do custo de uma paralisação em potencial. Este artigo descreve algumas abordagens comuns que variam desde os investimentos mínimos até a disponibilidade em grande escala e a tolerância a falhas.

A preparação adequada para DR é uma necessidade e este artigo descreve algumas das práticas recomendadas para melhorar os seus processos e planos de DR.

A recuperação de desastres é um processo contínuo de análise e melhoria, à medida que os negócios e sistemas evoluem. Para cada serviço de negócios, os clientes precisam determinar um tempo e um ponto de recuperação aceitável e, em seguida, construir uma solução adequada de DR.

Em um ambiente físico tradicional, uma abordagem comum normalmente envolve a duplicação da infraestrutura para garantir a disponibilidade de capacidade não utilizada em um cenário de desastre. Essa infraestrutura precisa ser adquirida, instalada e mantida para que esteja pronta para lidar com os requisitos de capacidade previstos. Em circunstâncias normais de funcionamento, essa infraestrutura seria normalmente subutilizada ou excessivamente provisionada.

A AWS permite a expansão de sua infraestrutura conforme a sua necessidade. Você tem acesso à mesma infraestrutura altamente escalável, confiável, segura, rápida e barata que a Amazon utiliza para executar sua própria rede global de websites e só paga pelo que você usar. Para uma solução de recuperação de desastres (DR) isso resulta em uma redução significativa de custos. Isso também permite mais agilidade para alterar e otimizar recursos durante um cenário de DR.

O erro humano é a uma das principais causas de inatividade do sistema. AWS fornece ferramentas para permitir a separação de funções para permitir um design de *privilégio mínimo* [1]. A AWS também permite que você automatize a implantação de ambientes inteiros, permitindo configurações previsíveis e que possam ser reproduzidas. Os ambientes de teste de DR podem ser configurados rapidamente e você pode, em seguida, tratá-los como um recurso disponível. Isso permite que as empresas testem as alterações de configuração em um ambiente duplicado antes de colocarem a configuração em produção, sem a necessidade de um ambiente de teste dedicado em grande escala, que se tornaria muitas vezes subutilizado.

[1] http://en.wikipedia.org/wiki/Principle_of_least_privilege

Objetivo de tempo de recuperação e objetivo de ponto de recuperação

Este documento usa dois termos comuns do setor para o planejamento de desastres:

Objetivo de tempo de recuperação (RTO)[2]— Este é o período do tempo e o nível de serviço para que um processo de negócios seja restaurado após um desastre (ou interrupção) para evitar consequências inaceitáveis associadas a uma quebra na continuidade de negócios. Por exemplo, se ocorresse um desastre às 12h (meio dia) e o RTO fosse de 8 horas, o processo de DR garantiria a recuperação em nível de serviço aceitável por volta das 20h.

Objetivo de ponto de recuperação (RPO)[3]— Descreve a quantidade aceitável de perda de dados medida em tempo. Por exemplo, se o RPO foi de 1 hora, depois que o sistema foi recuperado, ele conteria todos os dados até um momento determinado que não seria antes das 11h porque o desastre ocorreu ao meio-dia.

Uma empresa decide normalmente por uma RTO e RPO aceitáveis baseando-se no impacto financeiro para os negócios, quando os sistemas estão disponíveis. O impacto financeiro é tipicamente avaliado por muitos fatores, como a perda de negócios e danos à sua reputação devido ao tempo de inatividade e a falta de disponibilidade de sistemas.

As empresas de TI então planejam soluções econômicas para disponibilizar a recuperação do sistema com base no RPO dentro do cronograma e de um nível de serviço estabelecido pelo RTO.

Práticas tradicionais de investimento em DR

Uma abordagem tradicional à DR envolve diferentes níveis de duplicação fora do local de dados e de infraestrutura. Serviços essenciais aos negócios são criados e mantidos nesta infraestrutura e são testados em intervalos regulares. A localização do ambiente de recuperação de desastres e de infraestrutura de origem devem estar a uma distância física significativa para garantir que o ambiente de recuperação de desastres esteja isolado das falhas que poderiam afetar o site de origem.

A infraestrutura necessária para suportar o ambiente duplicado inclui, mas não se limita ao seguinte:

- Instalações para abrigar a infraestrutura, incluindo fonte de alimentação e resfriamento.
- Segurança para garantir a proteção física dos ativos.
- Capacidade adequada para dimensionar o ambiente.
- Suporte para reparar, substituir e atualizar a infraestrutura.
- Acordos contratuais com um provedor de serviços de Internet (ISP) para fornecer conectividade com a Internet que pode suportar a utilização de largura de banda para o ambiente sob carga completa.
- Infraestrutura de rede, como firewalls, roteadores, switches e balanceadores de carga.
- Suficiente capacidade de servidor para executar todos os serviços de missão crítica, incluindo dispositivos de armazenamento dos dados de suporte e servidores para executar aplicativos e serviços de back-end, como autenticação de usuário, sistema de nome de domínio (DNS), Protocolo de configuração de host dinâmico (DHCP), monitoramento e alertas.

Dependendo da criticidade dos serviços, o ambiente duplicado pode ser configurado de forma tolerante a falhas. Isso normalmente envolve a duplicação de toda a infraestrutura listada acima.

---

[2] Retirado de http://en.wikipedia.org/wiki/Recovery_time_objective
[3] Retirado de http://en.wikipedia.org/wiki/Recovery_point_objective

Serviços e recursos AWS essenciais para Recuperação de desastres

Antes de discutir as diferentes abordagens de DR, é importante analisar os serviços e recursos AWS que são mais relevantes para a recuperação de desastres. Esta seção fornece um resumo.

Na fase de preparação de recuperação de desastres, é essencial considerar a utilização de serviços e recursos que oferecem suporte à migração de dados e ao armazenamento durável, pois eles permitem que você restaure dados armazenados na AWS quando ocorrer um desastre. Para alguns dos cenários que envolvem tanto uma implantação reduzida quanto uma totalmente dimensionada de seu sistema na AWS, recursos de computação também serão necessários.

Ao reagir a um desastre, é essencial delegar rapidamente recursos computacionais para executar seu sistema na AWS ou orquestrar o failover para recursos já em execução na AWS. As peças de infraestrutura essenciais aqui incluem o DNS, recursos de rede e várias funcionalidades do Amazon Elastic Compute Cloud (Amazon EC2) descritas abaixo.

Regiões

Os Amazon Web Services estão disponíveis em vários Regiões, para que você possa escolher o local mais apropriado como sendo seu local de recuperação de desastres, além do local onde o sistema está totalmente implantado. Até o momento, a AWS está disponível em cinco regiões: Leste dos EUA (Norte da Virgínia), Oeste dos EUA (Norte da Califórnia), UE (Irlanda), Ásia-Pacífico (Cingapura) e Ásia-Pacífico (Tóquio).

Armazenamento

O Amazon Simple Storage Service (Amazon S3) fornece uma infraestrutura de armazenamento altamente durável projetada para armazenamento de dados de missão crítica e primário. Os objetos são armazenados redundantemente em vários dispositivos de múltiplas instalações dentro de uma região. A AWS fornece mais proteção para a retenção de dados e o arquivamento por meio de controle de versão no Amazon S3, AWS Multi-Factor Authentication, políticas de bucket e do Identity and Access Management (IAM).

O Amazon Elastic Block Store (Amazon EBS) também fornece a habilidade de criar snapshots de volumes de dados em um determinado momento. Esses Snapshots podem ser usados como ponto inicial para novos volumes Amazon EBS e para proteger dados para uma durabilidade a longo prazo. Após um volume ser criado, ele pode ser ligado a qualquer instância do Amazon EC2. Os volumes Amazon EBS oferecem armazenamentos fora da instância que persistem independentemente da vida de uma instância.

O AWS Import/Export acelera a movimentação de grandes volumes de dados para dentro e para fora da AWS usando dispositivos de armazenamento portáteis para transporte. A AWS transfere seus dados diretamente para os dispositivos de armazenamento e a partir deles, usando a rede interna de alta velocidade da Amazon, sem a necessidade de passar pela Internet. Para conjuntos de dados significativos, o AWS Import/Export é muitas vezes mais rápido do que a transferência através da Internet e mais rentável do que atualizar a sua conectividade. Você pode usar o AWS Import/Export para migrar dados dentro e fora de baldes do Amazon S3 baldes ou em snapshots do Amazon EBS.

O AWS Storage Gateway permite a migração contínua de dados de um lado para outro entre o armazenamento em nuvem da AWS e aplicativos no local. O AWS Storage Gateway armazena dados de volume localmente em sua infraestrutura e na AWS. Isso permite que os aplicativos existentes no local armazenem dados continuamente na infraestrutura de armazenamento econômica, segura e durável da AWS preservando o acesso de baixa latência a esses dados.

Computacional

O Amazon Elastic Compute Cloud (Amazon EC2) é um serviço web que fornece uma capacidade de computação redimensionável na nuvem. Em questão de minutos, você pode criar instâncias EC2, que são máquinas virtuais sobre as quais você tem o controle completo. No contexto da DR, a capacidade de rapidamente criar máquinas virtuais que você pode controlar é essencial. Descrever todos os recursos do Amazon EC2 está fora do escopo deste documento. Nos concentraremos nos aspectos do Amazon EC2 que são mais relevantes para a DR.

As Amazon Machine Images (AMIs) são pré-configuradas com sistemas operacionais e algumas AMIs pré-configuradas também podem incluir pilhas de aplicativo. Você também pode configurar suas próprias AMIs. No contexto da DR, é altamente recomendável que você tenha suas próprias AMIs configuradas e identificadas para que possam ser iniciadas como parte de seu processo de recuperação. Tais AMIs devem ser pré-configuradas com o sistema operacional de sua escolha além de peças adequadas de pilha de aplicativo.

As instâncias reservadas do Amazon EC2, que muitas vezes são usadas para receber um desconto significativo sobre o custo da execução de uma instância EC2, tem outra vantagem particularmente relevante para a DR. As instâncias reservadas ajudam a garantir que a capacidade de que você precisa esteja disponível quando necessário

As Zonas de disponibilidade são as posições distintas que são projetadas para serem isoladas das falhas em outras Zonas da disponibilidade e fornecem rede de conectividade acessível e de baixa latência para outras Zonas de disponibilidade da mesma região. Ao iniciar as instâncias em Zonas de disponibilidade separadas, você pode proteger seus aplicativos de falha de um único local. As regiões consistem de um ou mais disponibilidade zonas.

O Amazon EC2 VM Import permite que você facilmente importe imagens de máquina virtual do ambiente existente para instâncias do Amazon EC2.

Rede

Quando se lida com um desastre, é muito provável que você tenha que modificar as configurações de rede, pois está sendo realizado um failover para outro local.

O Amazon Route 53 é um serviço web de Domain Name System (DNS) altamente disponível e escalável. Ele é projetado para dar aos desenvolvedores e empresas uma maneira extremamente econômica e confiável de direcionar os usuários finais para aplicativos da Internet.

Os Endereços Elastic IP são endereços de IP estáticos projetados para computação em nuvem dinâmica. Ao contrário dos tradicionais endereços de IP estáticos, os endereços Elastic IP permitem que você filtre a instância ou falhas da Zona de disponibilidade por meio de remapeamento programado de seus endereços de IP públicos para qualquer instância em sua conta. Para recuperação de desastres, você pode pré-alocar também alguns endereços IP para os sistemas mais importantes para que seus endereços IP já sejam conhecidos antes que desastres aconteçam. Isso pode simplificar a execução do plano de DR.

O Elastic Load Balancing distribui automaticamente o tráfego de entrada dos aplicativos em várias instâncias do EC2. Ele permite que você atinja uma maior tolerância a falhas em seus aplicativos, fornecendo a capacidade de equilíbrio de carga necessária em resposta ao tráfego de entrada dos aplicativos. Assim como você pode pré-alocar endereços do Elastic IP, você pode pré-alocar seu Elastic Load Balancer para que seu nome DNS já seja conhecido, o que pode simplificar a execução de seu plano de DR.

O Amazon Virtual Private Cloud (Amazon VPC) permite-lhe aproveitar uma seção privada e isolada da nuvem da Amazon Web Services onde você pode executar recursos AWS em uma rede virtual que você mesmo define. Você tem controle total sobre seu ambiente de rede virtual, incluindo a seleção do seu próprio intervalo de endereços IP, criação de

sub-redes e configuração de tabelas de roteamento e gateways de rede. Isso permitirá que você crie uma conexão de VPN entre seu datacenter corporativo e o VPC e aproveite a nuvem AWS como uma extensão do seu datacenter corporativo. No contexto de DR, você pode usar o Amazon VPC para estender sua topologia de rede existente para a nuvem. Isso pode ser especialmente apropriado ao recuperar aplicativos corporativos que estejam essencialmente na rede interna.

O AWS Direct Connect facilita a configuração de uma conexão de rede dedicada entre a sua premissa e a AWS. Em muitos casos, isso pode reduzir seus custos de rede, aumentar a taxa de transferência de largura de banda e fornecer uma experiência de rede mais consistente do que conexões baseadas em Internet.

Bancos de dados

Para suas necessidades de banco de dados, considere utilizar esses serviços da AWS:

O Amazon Relational Database Service (Amazon RDS) facilita a configuração, a operação e o dimensionamento de seu banco de dados relacional na nuvem. Você pode usar o Amazon RDS na fase de preparação para recuperação de desastres para armazenar seus dados críticos em um banco de dados já em execução e/ou em fase de recuperação para executar seu banco de dados de produção.

O Amazon SimpleDB é um armazenamento de dados altamente disponível, flexível e não relacional que minimiza o trabalho da administração do banco de dados. Ele também pode ser usado na preparação e na fase de recuperação de DR.

Você também pode instalar e executar a sua escolha do software de banco de dados no Amazon EC2 e pode escolher entre uma variedade dos principais sistemas de banco de dados.

Para obter mais detalhes sobre as opções de banco de dados na AWS, consulte Execução de bases de dados na AWS.

Orquestração da implantação

Ferramentas e processos de instalação/configuração software de implantação de automação e de post-startup. Os investimentos nesta área são altamente recomendados. Isso pode ser muito útil para a fase de recuperação para criar o conjunto necessário de recursos no modo automatizado.

O AWS CloudFormation oferece aos desenvolvedores e administradores de sistemas uma maneira fácil de criar um grupo de recursos relacionados à AWS e fornecê-los de uma forma organizada e previsível. Você pode criar modelos para seus ambientes e implantar coleções associadas de recursos (chamadas de pilha) conforme necessário.

Segurança

Há muitos recursos relacionados a segurança entre os serviços AWS. Recomendamos que os clientes consulte o whitepaper sobre Práticas recomendadas de segurança. A AWS também fornece informações adicionais de risco e de conformidade no Centro de segurança AWS. Uma discussão completa da segurança está fora do escopo deste documento.

Exemplo de cenários de recuperação de desastres com a AWS

Esta seção examinará quatro cenários de recuperação de desastres que destacam o uso da AWS e a comparam com métodos tradicionais de recuperação de desastres:

- Backup e restauração
- Luz piloto para recuperação simples na AWS
- Soluções de espera passiva
- Solução de multissite

A Amazon Web Services permite que os clientes executem cada uma dessas estratégias de exemplo de DR de maneira econômica. É importante perceber que esses são apenas alguns exemplos de abordagens possíveis, e que variações e combinações podem ser possíveis.

Backup e restauração

Em ambientes mais tradicionais, os dados são armazenados em fita e são enviados para um local externo regularmente. Seu tempo de recuperação será o mais longo usando esse método. O Amazon S3 é um destino ideal para backup de dados, pois é projetado para fornecer durabilidade de 99.999999999% (11 noves) de objetos ao longo de um determinado ano. Transferência de dados de e para Amazon S3 normalmente é feito através da rede e, portanto, está acessível a partir de qualquer local. Existem muitas soluções de backup comerciais e de código aberto que utilizam o Amazon S3. O serviço AWS importação/exportação permite transferências de conjuntos de dados muito grandes por dispositivos de armazenamento diretamente a AWS de envio.

O serviço AWS Storage Gateway permite criar snapshots de seus volumes de dados no local, dados estes que serão copiados de forma transparente no Amazon S3 para backup. Posteriormente, você pode criar volumes locais ou volumes AWS EBS desses snapshots.

Para sistemas em execução na AWS, os clientes também podem fazer backup no Amazon S3. Os snapshots de volumes do Elastic Block Store (EBS) e backups do Amazon RDS são armazenados no Amazon S3. Como alternativa, você pode copiar arquivos diretamente no Amazon S3, ou pode optar por criar arquivos de backup e copiá-los no Amazon S3. Existem muitas soluções de backup que armazenam dados de backup no Amazon S3, e essas também podem ser usadas a partir de sistemas do Amazon EC2.

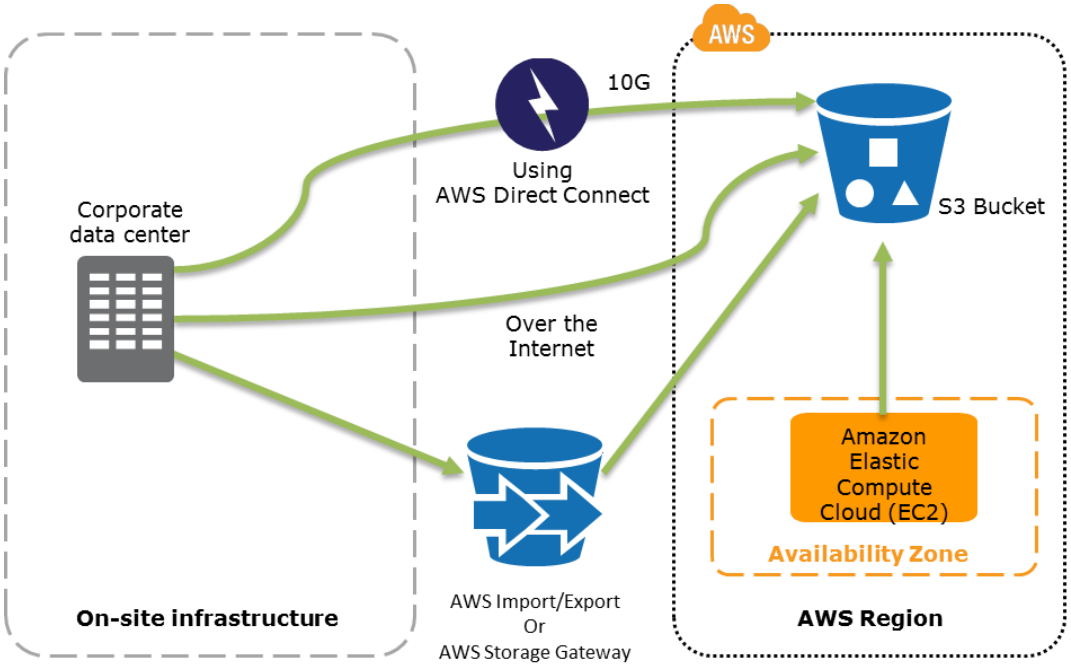

Figura 1: opções de backup de dados no S3 a partir de infraestrutura no local ou da AWS.

O backup de seus dados é apenas metade da história. A recuperação de dados em um cenário de desastre deve ser testada e alcançada de forma rápida e confiável. Os clientes devem garantir que seus sistemas estejam configurados para retenção adequada de dados, segurança de dados e que tenham sido testados para seus processos de recuperação de dados.

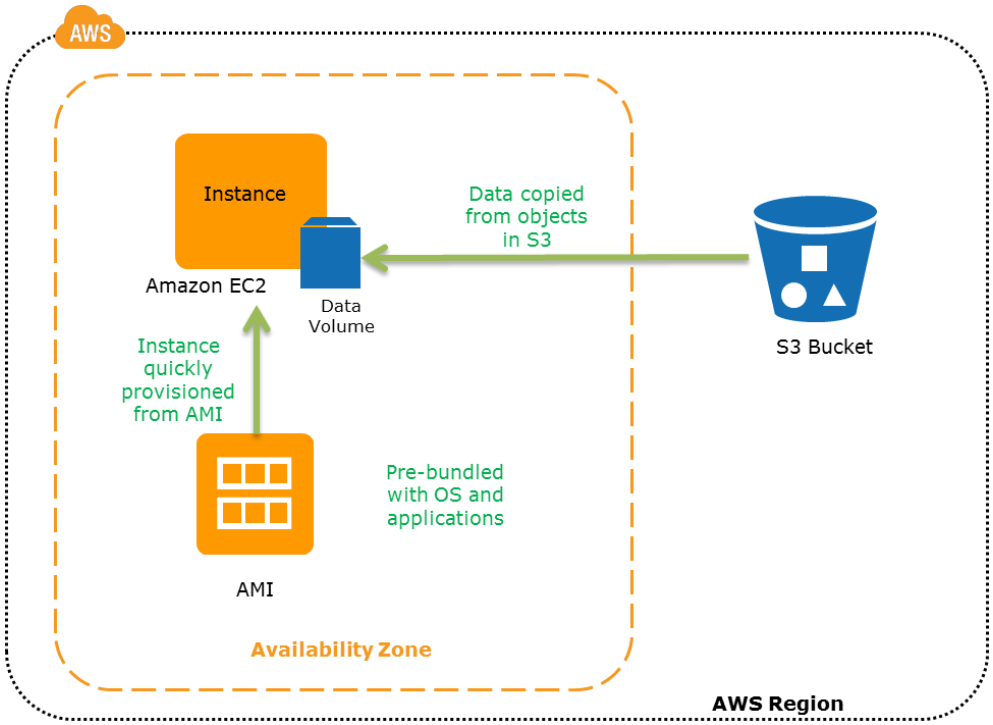

Figura 2: restauração de um sistema de backups do S3 para o EC2 da AWS

Principais etapas para backup e restauração:

- Selecione uma ferramenta apropriada ou método para fazer backup de seus dados na AWS.
- Assegure-se de ter uma política de retenção adequada para esses dados.
- Certifique-se de que as medidas de segurança adequadas estejam em vigor para esses dados, incluindo as políticas de acesso e criptografia.
- Teste regularmente a recuperação desses dados e a restauração do sistema.

Luz piloto para recuperação simples na AWS

A ideia da luz piloto é uma analogia que vem do aquecedor a gás. Em um aquecedor a gás, uma pequena chama ociosa que está sempre ligada rapidamente conduz à ignição de toda a fornalha para aquecer uma casa, conforme necessário. Este cenário é semelhante a um cenário de Backup e restauração, no entanto, você deve assegurar que os mais importantes elementos fundamentais do seu sistema já estejam configurados e em execução na AWS (a luz piloto). No momento de executar a recuperação, você provisionaria rapidamente um ambiente de produção de escala completa ao redor do núcleo crítico.

Elementos da infraestrutura para a luz piloto propriamente dita normalmente incluem seus servidores de banco de dados, que replicariam os dados para o Amazon EC2. Dependendo do sistema, pode haver outros dados críticos fora do banco de dados que precisam ser replicados na AWS. Este é o núcleo fundamental do sistema (a luz piloto) em torno do qual todas as outras peças da infraestrutura podem rapidamente ser configuradas (a fornalha) para restaurar o sistema completo.

Para provisionar o restante da infraestrutura para restaurar serviços críticos de negócios, você normalmente teria alguns servidores pré-configurados empacotados como Amazon Machine Images (AMIs), que estivessem prontos para serem iniciados em qualquer momento. Ao iniciar a recuperação, as instâncias dessas AMIs surgem rapidamente e encontram sua função dentro da implantação em torno a luz piloto. Do ponto de vista da rede, você pode usar os endereços Elastic IP (que podem ser pré-alocados na fase de preparação para recuperação de desastres) e associá-los às suas instâncias, ou usar o Elastic Load Balancing para distribuir o tráfego para várias instâncias. Em seguida, se atualizaria seus registros DNS para apontarem para a instância do Amazon EC2 ou para o Elastic Load Balancing usando um CNAME.

Para sistemas menos críticos, você pode garantir que você tenha quaisquer pacotes de instalação e informações de configuração disponíveis na AWS, por exemplo, sob a forma de um snapshot de EBS. Isso acelerará a configuração do servidor de aplicativo, pois você pode criar rapidamente vários volumes em várias zonas de disponibilidade, para anexar a instâncias EC2. Você pode, em seguida, instalar e configurar adequadamente.

O método de luz piloto disponibilizará um tempo de recuperação mais rápido do que com o cenário “Backup e restauração” acima, pois as peças principais do sistema já estão em execução e são constantemente atualizadas. Existem ainda algumas tarefas de instalação e configuração para que se possa recuperar totalmente os aplicativos. A AWS permite que você automatize o provisionamento e a configuração dos recursos de infraestrutura, que podem ser um benefício significativo para economizar tempo e ajudar a proteger contra erros humanos.

Fase de preparação:

A figura a seguir mostra a fase de preparação, na qual você precisa ter seu regularmente alteração de dados replicado para a luz piloto, o pequeno núcleo em torno do qual o ambiente completo será iniciado na fase de recuperação. Seus dados atualizados com menos frequência, como sistemas operacionais e aplicativos podem ser periodicamente atualizados e armazenados como Amazon Machine Images (AMIs).

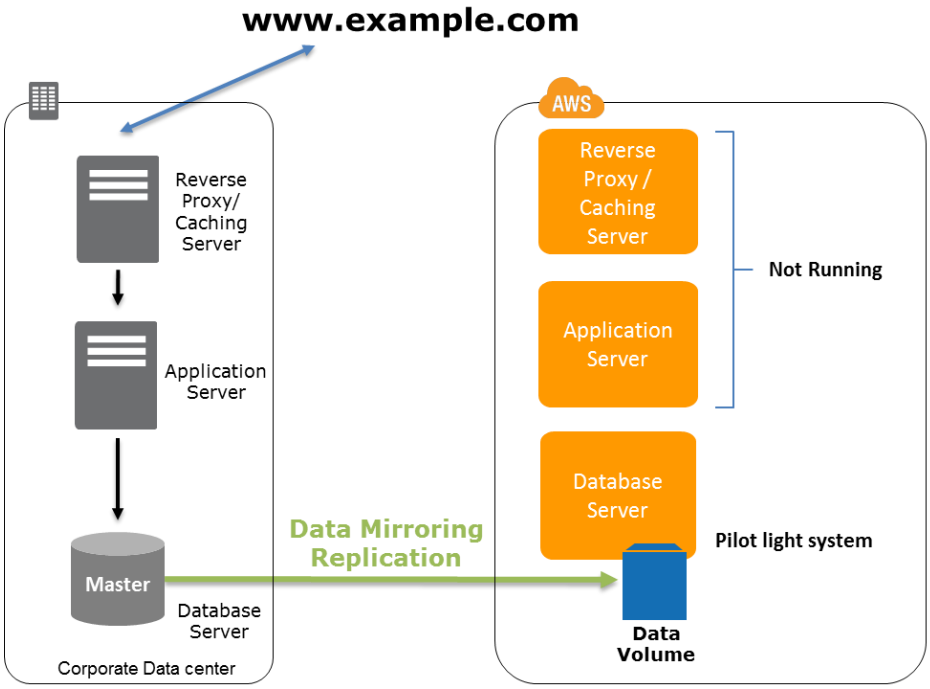

Figura 3: a fase de preparação do cenário de luz piloto

Pontos chave para preparação:

- Configure instâncias EC2 para replicar ou espelhar dados.
- Verifique se você tem todos os pacotes compatíveis de software personalizado disponíveis na AWS.
- Crie e mantenha Amazon Machine Images (AMI) de servidores-chave para os quais a recuperação rápida seja necessária.
- Execute, teste e aplique atualizações de software e alterações de configuração a esses servidores regularmente.
- Considere a possibilidade de automatizar o provisionamento de recursos AWS.

Fase de recuperação:

Para recuperar o restante do ambiente ao redor da luz piloto, você inicia seus sistemas a partir da Amazon Machine Images (AMIs) em minutos nos tipos de instância apropriada. Para seus servidores de dados dinâmicos, você pode redimensioná-los para lidar com volumes de produção, conforme necessário, ou adicionar capacidade de acordo com a demanda. O dimensionamento horizontal, se possível, é muitas vezes a forma de abordagem mais econômica e dimensionável de adicionar capacidade a um sistema. No entanto, também é possível escolher tipos maiores de instância do EC2 e, portanto, dimensionar verticalmente. Do ponto de vista da rede, quaisquer atualizações DNS necessárias podem ser feitas em paralelo.

Uma vez recuperado, você deve garantir que a redundância seja restaurada o mais rápido possível. Mesmo sendo improvável ocorrer uma falha de seu ambiente de DR seguida de uma falha de seu ambiente de produção, você precisará estar ciente desse risco. Continue a fazer backups regulares do seu sistema e considere redundância adicional na camada de dados.

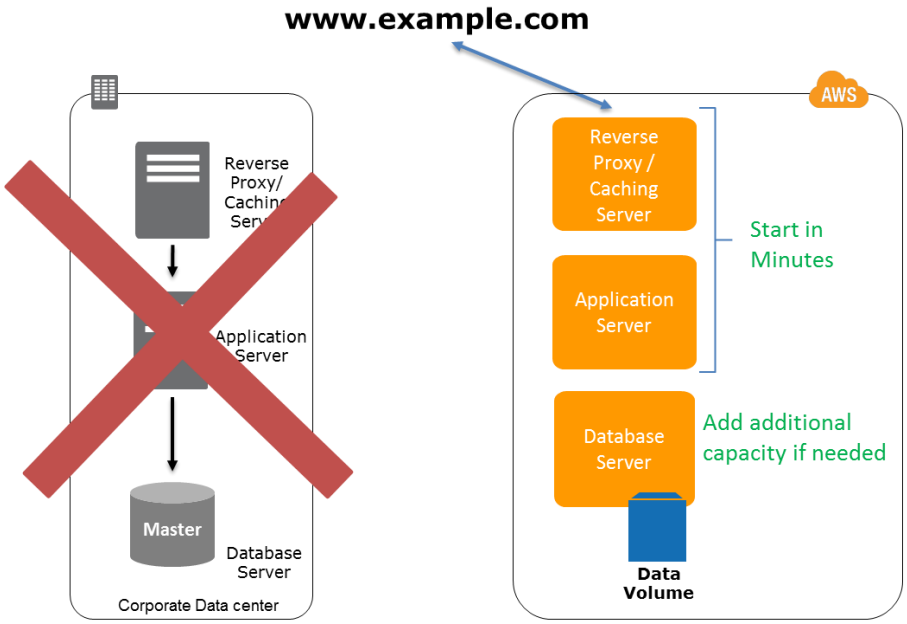

Figura 4: a fase de recuperação do cenário de luz piloto.

Pontos chave para recuperação:

- Inicie suas instâncias do aplicativo EC2 a partir suas AMIs personalizadas.
- Redimensione e/ou dimensione quaisquer bancos de dados / instâncias de dados armazenados, sempre que necessário.
- Altere o DNS para apontar para os servidores do EC2.
- Instale e configure todos os sistemas não baseados em AMI, idealmente no modo automatizado.

Soluções de espera passiva na AWS

Uma solução de espera passiva amplia os elementos de luz piloto e a preparação. Ela diminui ainda mais o tempo de recuperação, pois, neste caso, alguns serviços estão sempre em execução. Através da identificação de seus sistemas críticos para os negócios, seria possível duplicar totalmente esses sistemas na AWS e tê-los sempre em execução.

Esses servidores podem estar sendo executados em frotas de tamanho mínimo de instâncias do EC2 nos menores tamanhos possíveis. Esta solução não é dimensionada para suportar uma carga de produção completa, mas é totalmente funcional. Ela pode ser usada para trabalhos que não sejam de produção, como testes, controle de qualidade, uso interno etc.

No caso de um desastre, o sistema é dimensionado rapidamente para lidar com a carga de produção. Na AWS, isso pode ser feito adicionando mais instâncias ao balanceador de carga e redimensionando os servidores de pequena capacidade para serem executados em tipos de instância EC2 maiores. Como dito acima, o dimensionamento horizontal, se possível, é muitas vezes preferido ao dimensionamento vertical.

Fase de preparação:

O diagrama a seguir mostra a fase de preparação para uma solução de espera passiva, em que uma solução no local e uma solução AWS são executadas lado a lado.

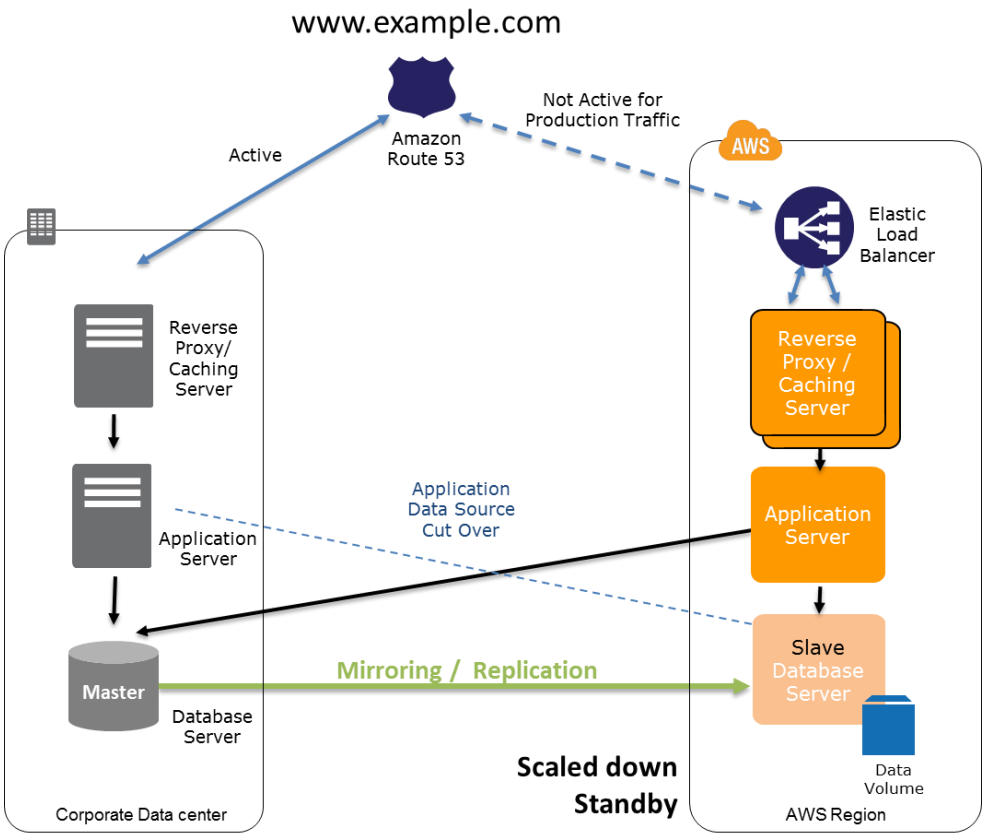

Figura 5: a fase de preparação do cenário de "espera passiva"

Pontos chave para preparação:

- Configure instâncias EC2 para replicar ou espelhar dados.
- Crie e mantenha Amazon Machine Images (AMI).
- Execute o aplicativo usando um espaço mínimo de instâncias EC2 ou de infraestrutura AWS.
- Patch e atualize arquivos de configuração e de software em conformidade com seu ambiente ao vivo.

Fase de recuperação:

Em caso de falha do sistema de produção, o ambiente de espera será dimensionado para carga de produção e os registros DNS serão alterados para rotear todo o tráfego para a AWS.

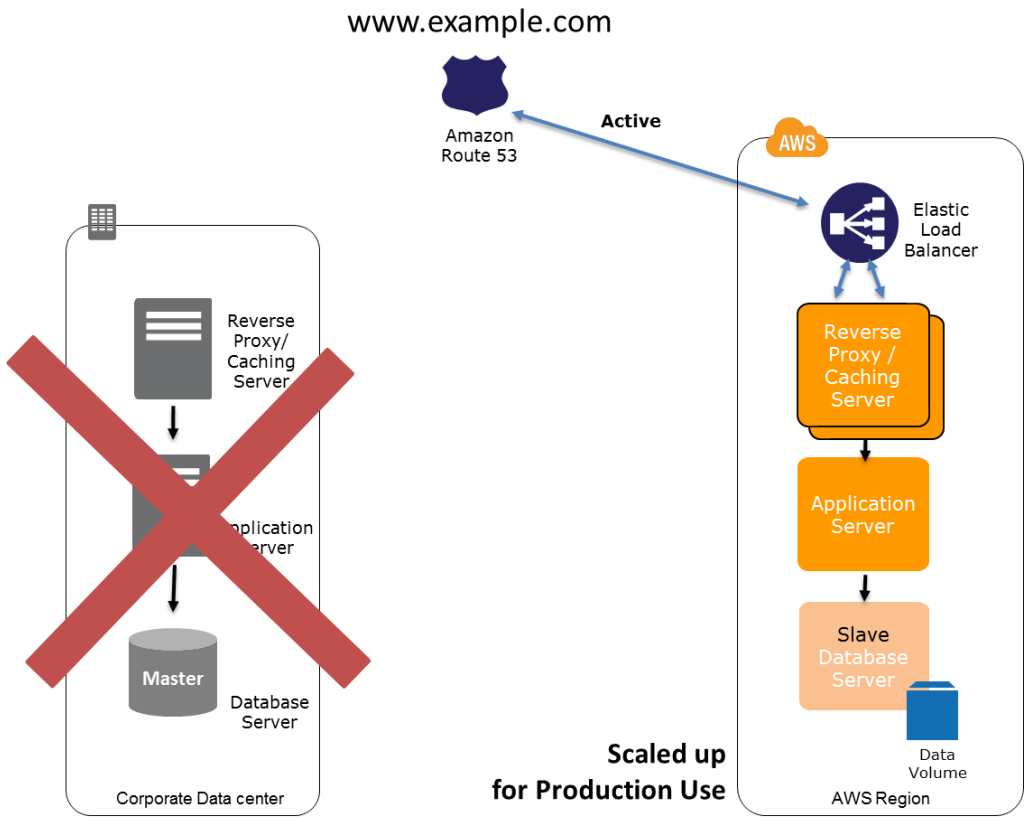

Figura 6: a fase de recuperação do cenário de "espera passiva".

Pontos chave para recuperação:

- Inicie aplicativos em tipos maiores de instância EC2 conforme necessário (dimensionamento vertical).
- Aumente o tamanho das frotas EC2 em serviço com o Load Balancer (dimensionamento horizontal).
- Altere os registros DNS para que todo o tráfego seja roteado para o ambiente AWS.
- Considere o uso de Auto dimensionamento para o tamanho ideal da frota ou acomodar o aumento de carga.

Solução de Multissite implantada na AWS e no local

Uma solução multissite é executada na AWS, como também em sua infra-estrutura existente no local em uma configuração ativa. O método de replicação de dados que você utilizar será determinado pelo ponto de recuperação (consulte RPO acima) que você escolher. Existem vários métodos de replicação (veja abaixo).

Um serviço DNS ponderado, como o Amazon Route 53, é utilizado para rotear o tráfego de produção para diferentes sites. Uma proporção de tráfego se dirige até sua infraestrutura na AWS e o restante até a sua infraestrutura no local.

Em uma situação de desastre no local, você pode ajustar a importância do DNS e enviar todo o tráfego para os servidores AWS. A capacidade do serviço AWS pode ser rapidamente aumentada para lidar com a carga produção completa. O Auto Scaling do EC2 pode ser usado para automatizar esse processo. Pode ser necessário alguma lógica de aplicativo para detectar a falha dos serviços de banco de dados primário e para destinar para os serviços de banco de dados paralelos em execução na AWS.

O custo deste cenário é determinado pela quantidade de tráfego de produção controlada pela AWS em operação normal. Na fase de recuperação, você só paga pelo adicional que você usa e pelo período que o ambiente de DR é necessário em escala completa. Você pode reduzir ainda mais os custos através da compra de Instâncias Reservadas para os seus servidores AWS "sempre em execução".

Fase de preparação:

Na figura abaixo, vemos o uso do DNS para rotear uma parte do tráfego para o site AWS. O aplicativo na AWS pode acessar fontes de dados no sistema de produção no local. Os dados são replicados ou espelhados para infraestrutura AWS.

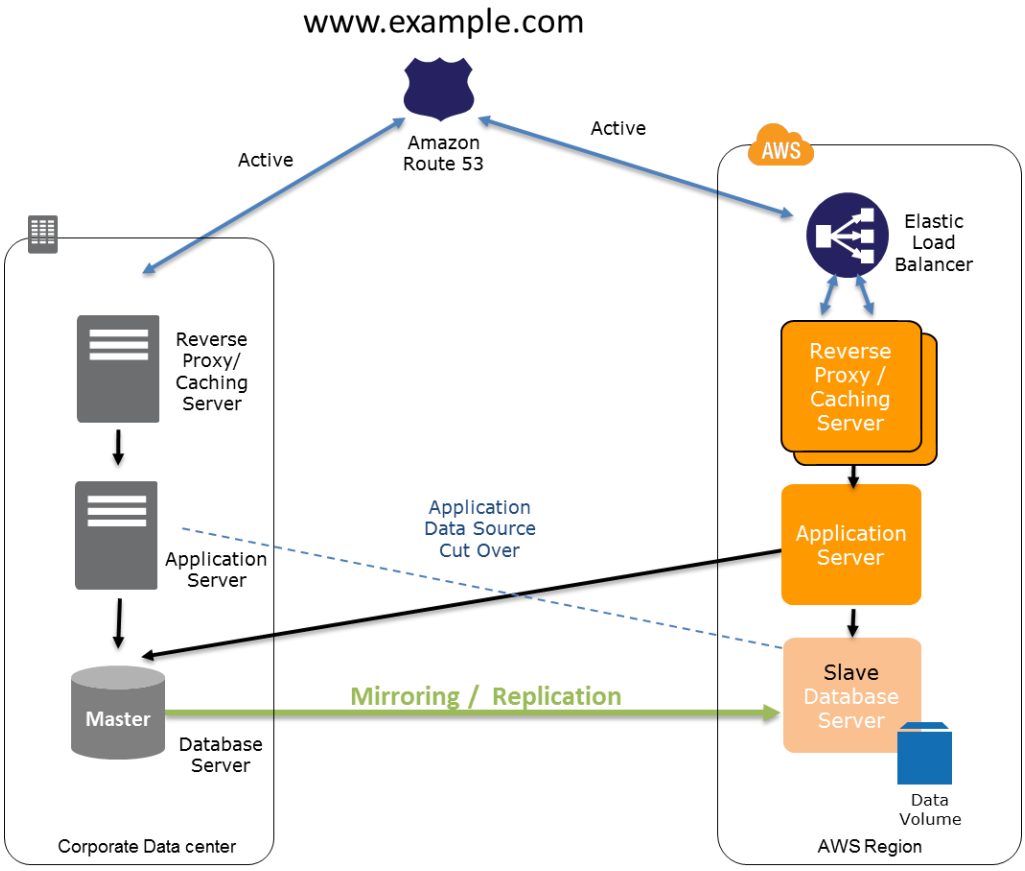

Figura 7: a fase de preparação do cenário "multissite".

Pontos chave para preparação:

- Configure seu ambiente AWS para duplicar seu ambiente de produção.
- Configure a importância do DNS ou tecnologia similar para distribuir as solicitações de entrada para ambos os locais.

Fase de recuperação:

A figura abaixo mostra o que acontece quando ocorrer um desastre no local. O tráfego é redirecionado para a infraestrutura AWS ao se atualizar o DNS.

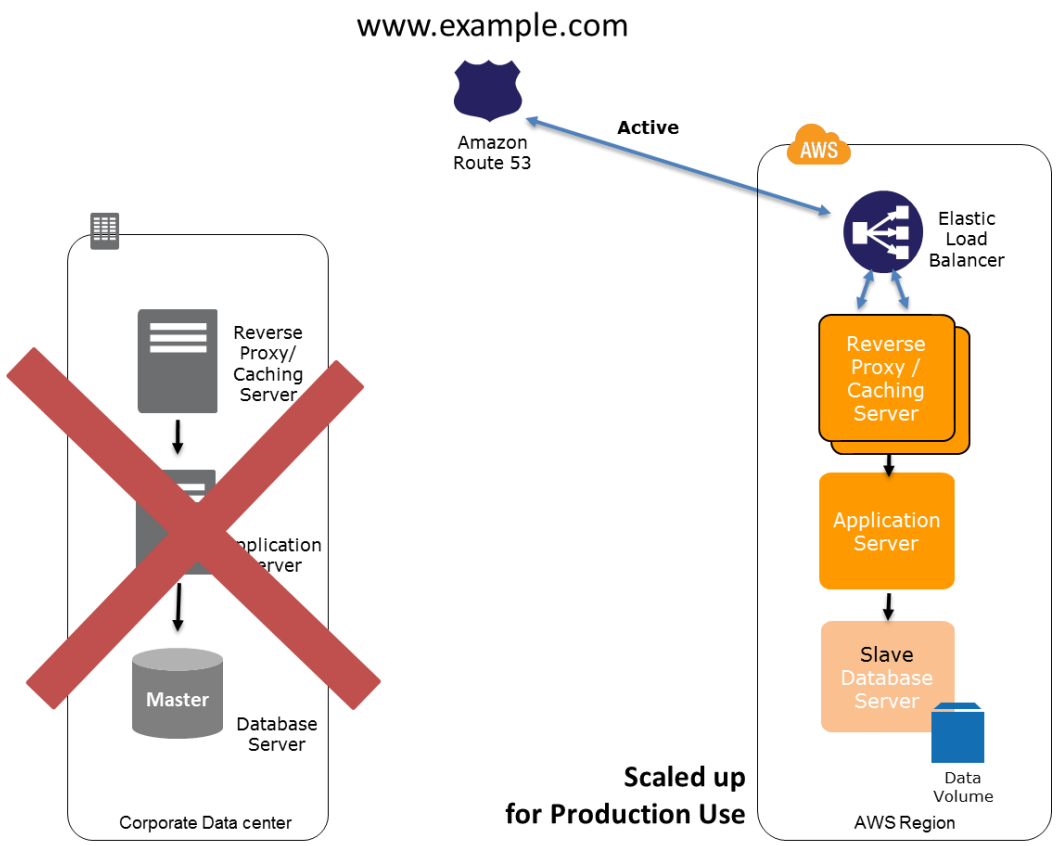

Figura 8: a fase de preparação do cenário "multissite"envolvendo infraestrutura no local e na AWS.

Pontos chave para recuperação:

- Altere a importância do DNS para que todas as solicitações sejam enviadas para o site AWS.
- Disponha de uma lógica de aplicativo para failover para usar os servidores de banco de dados locais AWS.
- Considere o uso de Auto scaling para automaticamente definir o tamanho ideal da frota AWS.

É possível aumentar ainda mais a disponibilidade de sua solução multissite através da concepção de arquiteturas Multi-AZ. Para obter mais informações sobre como criar aplicativos que se estendem em várias zonas de disponibilidade, consulte o whitepaperDesigning Fault-Tolerant Applications in the AWS Cloud (Criando aplicativos tolerantes a falhas na nuvem AWS).

Replicação de dados

Ao replicar dados para um local remoto, existem alguns fatores a serem considerados.

- Distância entre os locais: distâncias maiores normalmente estão sujeitas a mais latência e/ou variação.
- Largura de banda disponível: o quão amplas e variáveis são as interligações?
- Taxa de dados exigida por seu aplicativo: a taxa de dados deve ser menor que a largura de banda disponível.
- A tecnologia de replicação deve ser paralela (para que possa usar a rede efetivamente).

Existem duas abordagens principais ao replicar dados: síncrona e assíncrona.

Replicação síncrona

Os dados são atualizados de forma atômica em vários locais. Isso depende da disponibilidade e do desempenho da rede.

Replicação assíncrona

Os dados não são atualizados de forma atômica em vários locais. Os dados são transferidos conforme o desempenho e a disponibilidade da rede permitem e o aplicativo continua a gravar os dados que ainda não podem ser totalmente replicados.

Muitos sistemas de banco de dados oferecem suporte à replicação de dados assíncrona. A réplica do banco de dados pode ser localizada de forma remota e não necessita estar completamente em sincronia com o servidor de banco de dados primário. Isso é aceitável em muitos cenários, por exemplo, como uma fonte de backup ou em casos de utilização de relatórios/somente leitura.

Aconselhamos os clientes compreenderem bem a tecnologia de replicação usada em sua solução de software. Uma análise detalhada da tecnologia de replicação está fora do escopo deste documento.

Na AWS, as Zonas de disponibilidade dentro de uma região são bem conectadas, mas fisicamente separadas. Por exemplo, quando implantado no modo "Multi-AZ", o serviço de banco de dados relacional da Amazon usa a replicação síncrona para duplicar dados em uma segunda Zona de disponibilidade. Isso garante que os dados não sejam perdidos se a zona primária de disponibilidade se tornar indisponível.

As regiões AWS são completamente independentes entre si, mas não há diferenças na forma como você as acessa e as utiliza. Isso permite que os clientes criem processos de recuperação de desastres que se estendem por distâncias continentais, sem os desafios ou os custos que isso normalmente acarretaria. Os clientes podem fazer backup de dados e sistemas para duas ou mais regiões AWS permitindo restauração de serviço mesmo em face de desastres em extrema grande escala. Os clientes podem usar as regiões AWS para servir seus clientes finais ao redor do mundo com relativamente baixa complexidade para seus processos operacionais.

Como aprimorar seu plano de DR

Alguns passos importantes precisam ser seguidos para se ter um sólido plano de DR. Esta seção descreve algumas das etapas principais.

Teste

Depois que sua solução de recuperação de desastres esteja pronta, ela precisa ser testada. O chamado "Dia D" é quando você exercita um failover para o ambiente de DR. É preciso garantir que documentação suficiente esteja disponível para simplificar ao máximo o processo em caso de um evento real ocorrer. Girar um ambiente duplicado para testar seus cenários durante o chamado dia D é rápido e econômico na AWS, e você não precisa nem tocar em seu ambiente de produção. Você pode usar o AWS CloudFormation para implantar ambientes completos na AWS. Para isso se faz uso de um modelo para descrever os recursos AWS e quaisquer dependências ou parâmetros de tempo de execução associados, necessários para criar um ambiente completo.

Diferenciar seus testes é fundamental para garantir que você esteja protegido contra uma enorme variedade de diferentes tipos de desastres. Os seguintes são exemplo de cenários de "Dia D":

- Perda de energia para um local ou um conjunto de máquinas
- Perda de conectividade de ISP para um único local
- Vírus impactando os principais serviços de negócios que afetam multissites
- Erro de usuário que causou a perda de dados que requer uma recuperação em um momento determinado

Monitoramento e alertas

Você precisa ter um controle regular e monitorização suficientes à disposição para alertá-lo quando seu ambiente de DR for impactado por falha no servidor, problemas de conectividade e de aplicativo. O Amazon CloudWatch fornece acesso a métricas sobre recursos AWS. Os alertas podem ser configurados com base nos limites definidos em qualquer uma das métricas e, quando necessário, podem ser enviadas mensagens de serviço do Amazon Simple Notification Service em caso de um comportamento inesperado. Você pode usar quaisquer soluções de monitoramento na AWS.

Você também pode continuar a usar qualquer existentes ferramentas de monitoramento e alertas que sua empresa usa para monitorar suas métricas de instância, bem como estatísticas de sistema operacional convidado e integridade de aplicativos.

Backups

Uma vez tendo alternado para seu ambiente de DR, é imprescindível continuar a fazer backups regulares. Conduzir testes de backup e de restauração regularmente é essencial como uma solução recuperação.

A AWS disponibiliza a você a flexibilidade de conduzir testes frequentes e econômicos de DR sem a necessidade de infraestrutura de DR para estar "sempre em execução".

Acesso de usuário

Você pode proteger o acesso aos recursos em seu ambiente de DR usando o Identity and AWS Access Management (IAM). Desta maneira você pode criar políticas de segurança baseadas em funções/usuário que dividem as responsabilidades do usuário enquanto trabalham em seu ambiente de DR.

Automação

Você pode automatizar a implantação de aplicativos para os servidores baseados em AWS e seus servidores locais usando o software de gerenciamento ou de orquestração de configuração. Isso permitirá que você lide com aplicativos e com alterações de gerenciamento de configuração de aplicativo em ambos ambientes com facilidade. Há várias opções de software de orquestração populares, além de possíveis fornecedores de soluções que podem ser encontrados em nossa página de provedores de soluções[4]. O AWS CloudFormation funciona em conjunto com várias ferramentas para configurar os serviços de infraestrutura de forma automatizada. Os dados do usuário podem ser passados para a instância no primeiro boot e podem, em seguida, ser direcionados a uma ferramenta de gerenciamento de configuração para determinar o tipo de instância ou a função para garantir que o software e a configuração correta sejam implantados. O objetivo principal deve ser fazer com que suas instâncias finalmente se encontrem no estado final no qual você as necessita o mais automaticamente possível.

O Auto Scaling pode ser usado para garantir que seu pool de instâncias seja adequadamente dimensionado para atender a demanda com base em métricas que você especifica no CloudWatch. Isso significa que em uma situação de recuperação de desastres, à medida que seu usuário base começa a usar mais o ambiente, a solução pode expandir dinamicamente para atender a essa demanda crescente. Após a finalização do evento e a diminuição potencial da utilização, a solução pode ser dimensionada de volta para um nível mínimo de servidores.

Licenciamento de software e DR

Garantir que você esteja licenciado corretamente para seu ambiente AWS é tão importante quanto qualquer outro ambiente de licenciamento. A Amazon fornece uma variedade de modelos para facilitar o gerenciamento de seu licenciamento. Por exemplo, "Trazer sua própria licença" é possível para vários componentes de software ou sistemas operacionais. Como alternativa, há uma variedade de software para a qual o custo da licença está incluído na cobrança horária. Isso é conhecido como "Licença incluída".

"Trazer sua própria licença" permite que você aproveite os investimentos existentes em software durante um desastre. "Licença incluída" minimiza os custos antecipados com licença para um local de DR que não é utilizado diariamente, ou seja, durante um teste de DR.

Se em qualquer etapa, você estiver em dúvida sobre suas licenças e sobre como elas se aplicam a AWS, entre em contato com seu revendedor de licença.

Conclusão

Existem muitas opções e variações para a DR, e este whitepaper destaca alguns dos padrões comuns, que vão desde simples backup e restauração até soluções multissite tolerantes a falhas. A AWS disponibiliza controle refinado e muitos blocos de construção para construir a solução adequada de DR, em vista de seus objetivos de DR (RTO e RPO) e seu orçamento. Os serviços AWS estão disponível on-demand e você só paga pelo que você utiliza. Esta é uma vantagem chave para recuperação de desastres, onde infraestrutura significativa é necessária rapidamente, mas apenas no caso de um desastre.

Este documento mostrou como a AWS fornece soluções de infraestrutura econômica e flexível, permitindo que você tenha um plano mais eficaz de DR.

[4] Os provedores de soluções podem ser encontrados em http://aws.amazon.com/solutions/solution-providers/

Leitura complementar

- Guia de Introdução do S3: http://docs.amazonwebservices.com/AmazonS3/latest/gsg/
- Guia de Introdução do EC2: http://docs.amazonwebservices.com/AWSEC2/latest/GettingStartedGuide/
- Encontre um Provedor de Soluções AWS: http://aws.amazon.com/solutions/solution-providers/
- Como projetar aplicativos tolerantes a falhas no whitepaper da nuvem da AWS: http://aws.amazon.com/whitepapers/
- Centro de conformidade e segurança AWS: http://aws.amazon.com/security/
- Central de arquitetura da AWS: http://aws.amazon.com/architecture
- Whitepapers técnicos da AWS: http://aws.amazon.com/whitepapers