

O'REILLY®



Uma introdução aos bancos de dados na nuvem

Guia para administradores

Wendy A. Neu, Vlad Vlasceanu,
Andy Oram e Sam Alapati

RELATÓRIO

Liberte-se de bancos de dados ultrapassados

A AWS oferece a mais ampla solução de bancos de dados com propósito específico para que você possa **economizar, crescer e acelerar a inovação**



Escala empresarial por um décimo do custo de bancos de dados comerciais



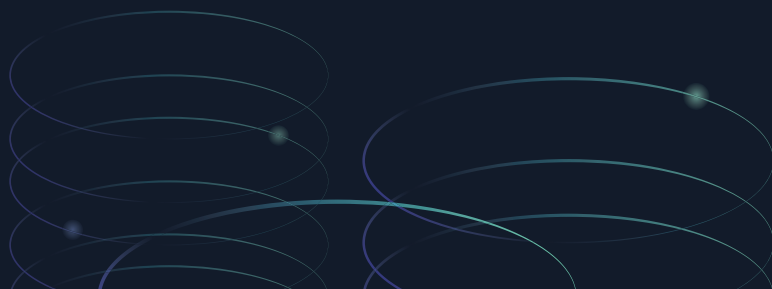
Performance três a cinco vezes melhor que as alternativas populares



Mais de 14 mecanismos de banco de dados: mais que qualquer outro provedor



Saiba mais: aws.amazon.com/databases



Uma introdução aos bancos de dados na nuvem

Guia para administradores

*Wendy A. Neu, Vlad Vlasceanu, Andy Oram
e Sam Alapati*

Pequim • Boston • Farnham • Sebastopol • Tóquio

O'REILLY®

Uma introdução aos bancos de dados na nuvem

por Wendy A. Neu, Vlad Vlasceanu, Andy Oram e Sam Alapati

Copyright © 2019 O'Reilly Media Inc. Todos os direitos reservados.

Impresso nos Estados Unidos da América.

Publicado por O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, EUA.

Os livros da O'Reilly podem ser adquiridos para fins de educação, negócios ou promoção de vendas. Edições online também estão disponíveis para a maioria dos títulos (<http://oreilly.com>). Para obter mais informações, entre em contato com o nosso departamento de vendas corporativo/institucional: 800-998-9938 ou corporate@oreilly.com.

Editor de desenvolvimento: Jeff Bleiel

Editor de aquisições: Jonathan Hassell

Editora de produção: Katherine Tozer

Revisor: Octal Publishing, LLC

Designer da parte interna: David Futato

Designer da capa: Karen Montgomery

Ilustradora: Rebecca Demarest

Setembro de 2019:

Primeira edição

Histórico de revisões da primeira edição

19/08/2019:

Primeira impressão

O logotipo da O'Reilly é uma marca registrada da O'Reilly Media, Inc. *Uma introdução aos bancos de dados na nuvem*, a imagem da capa e a apresentação comercial são marcas registradas da O'Reilly Media, Inc.

As opiniões expressadas neste trabalho são dos autores e não representam as opiniões da editora. Embora a editora e os autores tenham empreendido seus melhores esforços para garantir que as informações e as instruções contidas neste trabalho sejam precisas, a editora e os autores se isentam de qualquer responsabilidade por erros ou omissões, incluindo, entre outras, a responsabilidade por danos resultantes do uso deste trabalho ou da confiança nele. O uso das informações e das instruções contidas neste trabalho é por sua conta e risco. Se qualquer exemplo de código ou outra tecnologia contida ou descrita neste trabalho estiver sujeita a licenças de código aberto ou for propriedade intelectual de terceiros, é sua responsabilidade garantir que o seu uso desse código ou tecnologia esteja em conformidade com essas licenças e/ou direitos.

Este trabalho é parte de uma colaboração entre a O'Reilly e a AWS. Veja a nossa [declaração de independência editorial](#).

978-1-492-04482-6

[LSI]

Índice

1. Opções de banco de dados na nuvem.....	1
Resumo dos efeitos da migração para a nuvem	2
Bancos de dados autogerenciados e gerenciados	5
Bancos de dados nativos da nuvem	6
Tipos de bancos de dados gerenciados	7
Quando um banco de dados gerenciado pode não ser adequado	9
Função do DBA em um banco de dados gerenciado	9
2. A evolução da função do DBA na nuvem.....	11
Como as tarefas do DBA mudam na nuvem	13
Segurança para dados e aplicações na nuvem	19
Infraestrutura como código: como obter o máximo valor da nuvem	23
3. Como migrar os bancos de dados para a nuvem.....	25
Planejamento	26
Migração dos dados	32
Otimização	35
Conclusão	38
4. Conclusão.....	39

Opções de banco de dados na nuvem

Muitas vezes, a corrida para a nuvem é avaliada em termos puramente empresariais, e sua crescente popularidade é amplamente reconhecida. Por exemplo, um [relatório recente do Gartner](#) (que exigia o envio de informações empresariais para visualização) constatou que os bancos de dados estão crescendo a uma taxa de 68% na nuvem. Em contrapartida, o crescimento on-premise é bem reduzido e observa-se um aumento nos preços e no que eles denominam “upgrades forçados”.

Mas essa mudança radical no acesso à computação e aos bancos de dados também aparece nos tipos de serviços oferecidos e na evolução de trabalhos profissionais em computação. Hoje, os bancos de dados gerenciados de diversos tipos fazem parte das ofertas de todos os principais fornecedores de nuvem. O uso desses bancos de dados removerá várias tarefas executadas pelos administradores de banco de dados (DBAs) em ambiente tradicionais, nos quais a organização é proprietária de seu próprio hardware, e que chamaremos de ambientes *on-premise*. A migração para a nuvem criará novas tarefas, alterará algumas tarefas atuais e oferecerá um contexto sutilmente diferente para compreender muitas delas.

Os fornecedores de nuvem acompanham obstinadamente as alterações na área de banco de dados e variam suas ofertas para essa área de acordo com as diversas necessidades dos clientes. Os bancos de dados relacionais, nas formas transacional e data warehouse, aparecem ao lado de bancos de dados não relacionais, como armazenamentos de chave-valor e documentos.

Esse relatório ajuda os DBAs e as equipes relacionadas, como cientistas de dados, arquitetos de dados e desenvolvedores de aplicações, a escolher entre as diversas ofertas de nuvem. Ele explica resumidamente quais as responsabilidades que um DBA pode assumir no ambiente de nuvem. Por fim, ele oferece diretrizes para a migração.

O relatório não descreve argumentos a favor ou contra a migração para a nuvem porque há outros recursos para apoiar essa decisão e porque ela é fortemente vinculada às características particulares dos seus bancos de dados e à forma como são utilizados. Além disso, não analisamos ou recomendamos ofertas de nuvem específicas, embora mencionemos as ofertas dos atuais fornecedores de nuvem predominantes: Amazon Web Services (AWS), Microsoft Azure e Google Cloud Platform.

Este capítulo inclui:

- Uma breve recapitulação das diferenças entre computação on-premise e computação na nuvem, destacando o que é relevante para os DBAs.
- A definição de banco de dados gerenciado.
- Os tipos de bancos de dados gerenciados oferecidos por provedores de nuvem.
- Uma visão geral da função do DBA na nuvem.

Muitas vezes, as ofertas de nuvem espelham o software disponível para implantações tradicionais on-premise. Isso significa que você pode migrar dados do seu próprio equipamento para um banco de dados comparável oferecido por um fornecedor de nuvem. Por exemplo, você pode executar o MySQL ou o Oracle na nuvem, em vez de rodá-lo no seu próprio datacenter, ou usar um banco de dados nativo da nuvem, como o Amazon Aurora.

Você também deve ser capaz de manter os dois ambientes (on-premise e na nuvem), migrando os dados para a nuvem em um ritmo confortável para a organização. Além disso, podem-se manter permanentemente as duas implantações, uma on-premise e a outra na nuvem. A nuvem é frequentemente um bom local para experimentar algo totalmente novo, como um projeto de big data que você nunca criou antes.

Esperamos que a maioria dos leitores estejam familiarizados com a nuvem, seus benefícios e os diversos motivos para usá-la ou evitá-la, mas a próxima sessão resume as características da nuvem que embasam as ideias deste relatório. Depois disso, poderemos nos concentrar nos bancos de dados.

Resumo dos efeitos da migração para a nuvem

À medida que você migra dados para a nuvem, encontra diferenças em relação às implantações on-premise que afetam suas decisões. Estas são as principais mudanças relevantes para um DBA:

- Você paga conforme o uso, em incrementos horários ou mensais, para executar cada servidor (denominado uma *instância* no jargão de computação virtual). Dessa forma, se você tem uma aplicação de inteligência de negócios (BI) executado todas as noites, pode acionar uma instância de banco de dados por oito horas durante a noite e deixar de pagar pelas outras 16 horas do dia. Com alguns serviços (como as **instâncias reservadas da AWS**) você também pode executar um banco de dados por tempo limitado pagando um preço com desconto.
- O hardware e a infraestrutura relacionada são administrados pelo fornecedor. Isso pode livrar os administradores do sistema e da base de dados de um enorme volume de trabalho. Em vez de provisionar o seu próprio hardware, você escolhe a quantidade de capacidade de CPU, memória e armazenamento desejada para a instância de banco de dados. Muitas vezes, os fornecedores têm ofertas especiais que podem aprimorar a performance de maneiras relacionadas às estratégias usadas on-premise. Por exemplo, o Microsoft Azure SQL Data Warehouse oferece armazenamento em Solid State Drive (SSD – Unidades de estado sólido) e o Amazon DynamoDB tem uma versão especial com armazenamento em cache na memória.
- Além de oferecer o isolamento de workloads em máquinas virtuais (VMs), os principais fornecedores de nuvem (AWS, Azure e Google Cloud) permitem isolar redes corporativas completas na nuvem, um serviço denominado **Virtual Private Cloud (VPC – Nuvem privada virtual)**. Da mesma forma que as VMs permitem que cada um dos servidores use recursos computacionais com eficiência e segurança, uma VPC permite que você configure sua rede corporativa na nuvem, possivelmente reduzindo custos e transferindo para o fornecedor de nuvem uma boa parte do gerenciamento da segurança da rede. Os servidores de bancos de dados podem participar dessas VPCs. Você também pode configurar uma Virtual Private Network (VPN – Rede privada virtual) para comunicação entre a VPC na nuvem e os computadores on-premise.
- Os fornecedores de nuvem oferecem várias localizações, o que é importante para diversas finalidades: oferecer servidores na proximidade geográfica das bases de usuários dos clientes, proporcionar redundância em caso de desastres e permitir o balanceamento de carga, entre outros recursos. Alguns países, particularmente na União Europeia, exigem que certos tipos de dados sejam armazenados dentro de suas jurisdições legais por questões de privacidade ou outros motivos jurídicos. O conceito de regiões, áreas que podem abranger grande parte de um continente, e de *zonas de disponibilidade* (AZs) dentro dessas regiões aparece na **AWS**, no **Azure** e no **Google Cloud**. Assim, você pode distribuir workloads em diversas AZs de uma região,

replicando dados e serviços, e também configurar balanceamento de carga entre cópias redundantes de dados para obter resiliência. Alguns fornecedores de nuvem oferecem regiões com recursos para atender a necessidades regulatórias, particularmente as relacionadas à segurança. Além disso, todos os principais provedores de nuvem oferecem serviços de Content Delivery Network (CDN – Rede de entrega de conteúdo).

- Os fornecedores de nuvem oferecem formas sofisticadas de escalar serviços, obter tolerância a falhas e executar balanceamento de cargas. Se uma instância de banco de dados falhar e estiver configurada para alta disponibilidade, uma nova instância poderá estar disponível. Caso o cliente precise reiniciar sessões em andamento na instância anterior, quando ocorreu a falha, os esforços administrativos se concentrarão na configuração da tolerância a falhas, e não em acordar de madrugada para reiniciar o banco de dados. De forma semelhante, se a carga de um banco de dados ficar alta, uma nova instância poderá ser iniciada para compartilhar a carga e ser encerrada quando não for mais necessária. Esse recurso é denominado *escalabilidade elástica*.
- Normalmente, os fornecedores oferecem opções de linha de comando, interfaces gráficas e APIs, cada uma delas com os pontos fortes e fracos já conhecidos pelos profissionais de computação. As APIs permitem a automação e, portanto, assumem parte do seu trabalho.
- Parte da segurança com a qual você se preocupa on-premise é assumida pelo fornecedor. Obviamente, ele controla o acesso físico aos servidores. Dependendo da forma como você usa os serviços, ele também pode proteger você contra ataques de baixo nível pela Internet, como varreduras de porta e Denial-of-Service (DoS – Negação de serviço). Os capítulos seguintes descrevem os aspectos de segurança que permanecem sob a responsabilidade do DBA.
- Os fornecedores de nuvem incorporam ferramentas de monitoramento e performance que você pode conectar aos bancos de dados com bastante facilidade. Além disso, eles disponibilizam continuamente novos recursos. As ferramentas de monitoramento de terceiros também podem conseguir aproveitar e ampliar os recursos de monitoramento incorporados à plataforma.

Muitos critérios para a escolha de um provedor são genéricos e não específicos de banco de dados: definição de preço, ecossistema de desenvolvedores terceirizados para cada oferta, estabilidade, suporte a regimes de conformidade legal e valor agregado em ferramentas sofisticadas de inteligência artificial, entre outros.

Conforme você avalia as ofertas de nuvem, pense sobre as futuras direções da sua organização. Por exemplo, o processamento de streaming e dados em tempo real terá mais importância no futuro? Quais ofertas de nuvem continuarão a apoiar a sua organização daqui a dois anos?

Bancos de dados autogerenciados e gerenciados

Vamos falar agora especificamente sobre banco de dados. Suas opções na nuvem se dividem em duas categorias importantes:

Bancos de dados autogerenciados

Nesta oferta de nuvem, o fornecedor oferece apenas o hardware, o hipervisor para executar sua VM e a API ou outras ferramentas para gerenciar a implantação. Você precisa criar uma VM que execute um sistema operacional e a sua aplicação. Obviamente, é possível executar o que quiser na VM, inclusive um mecanismo de banco de dados à sua escolha. Você precisa executar a maioria das tarefas administrativas, como instalação de atualizações e configuração de todas as opções de rede.

Bancos de dados gerenciados

Nesta oferta de nuvem, além do hardware, o fornecedor oferece o próprio software de servidor. A maioria dos fornecedores oferecem bancos de dados tradicionais (como Oracle e MySQL) e bancos de dados nativos da nuvem específicos do fornecedor.

Os bancos de dados autogerenciados são bastante semelhantes à execução do seu próprio banco de dados on-premise. Os bancos de dados gerenciados alteram bastante a sua forma de trabalho. Estas são as principais diferenças:

- Se você usa um banco de dados gerenciado, não precisa executar manualmente atividades como fazer download, instalar, atualizar, configurar ou fazer backup do banco de dados. O provedor de nuvem faz tudo isso. Você consegue modificar parâmetros que orientam a execução do banco de dados (intercalação, cursores, conexões etc.) e controlar decisões como a instalação de novas versões ou o prazo de retenção de backups.
- Ocasionalmente, o fornecedor de nuvem pode oferecer uma licença de um banco de dados proprietário e incluir o custo no preço da utilização do serviço de nuvem. Em outros casos, a responsabilidade de obter uma licença é sua. Em alguns casos, se você já possuir uma licença de um banco de dados proprietário, poderá aplicá-la ao banco de dados gerenciado.
- Como o fornecedor instala e executa o banco de dados gerenciado, você fica limitado às escolhas feitas pelo fornecedor. Você não pode solicitar um mecanismo ou até mesmo uma versão de banco de dados não oferecida pelo fornecedor. Mas os três maiores fornecedores (AWS, Azure e Google) oferecem diversos bancos de dados de código aberto, além de seus próprios bancos de dados nativos da nuvem.

- Os bancos de dados autogerenciados e gerenciados oferecem parâmetros diferentes para configuração.
- Nos bancos de dados gerenciados, o fornecedor é responsável por boa parte da segurança. Mas até mesmo um banco de dados gerenciado deixa que você tome decisões de segurança importantes, como quem recebe contas e quais limitações são aplicadas a tabelas ou colunas. Normalmente, o monitoramento e a auditoria de tentativas de acesso são sua responsabilidade. Você também pode controlar as configurações de segurança, como dar acesso a endereços IP específicos.

Portanto, os bancos de dados gerenciados oferecem várias vantagens em relação a implantações on-premise ou bancos de dados autogerenciados. No entanto, talvez você tenha seus motivos para executar seu próprio banco de dados. O restante deste relatório abrange bancos de dados gerenciados porque oferecem uma oportunidade única para que os DBAs se dediquem a atividades com impacto maior e mais duradouro na empresa.

Bancos de dados nativos da nuvem

Os bancos de dados gerenciados também podem ser divididos em duas categorias: tradicionais e nativos da nuvem. Muitas vezes, bancos de dados tradicionais, como Oracle, SQL Server, MySQL e PostgreSQL, são oferecidos como bancos de dados gerenciados pelos provedores de nuvem. Se você estruturou sua organização usando um desses bancos de dados on-premise, a migração para o mesmo banco de dados na nuvem será simples. É menos provável que você precise alterar aplicações e será possível usar ferramentas conhecidas para gerenciar o banco de dados. Além disso, você pode combinar ofertas de fornecedores diferentes ou manter a sua versão do banco de dados on-premise.

No entanto, os fornecedores investiram um grande esforço em novas ofertas próprias, denominadas ocasionalmente bancos de dados *nativos da nuvem*. Os fornecedores demonstram que bancos de dados nativos da nuvem têm performance superior, escalabilidade facilitada e custos inferiores no longo prazo. Assim, depoimentos da **Autodesk** e da **InfoScout** sugerem que os engenheiros da AWS resolveram muitos dos problemas de escalabilidade e eficiência resultantes do gerenciamento de bancos de dados relacionais na nuvem com seu próprio banco de dados, o Amazon Aurora. Além disso, os bancos de dados nativos são projetados para oferecer uma escala gigantesca, uma tarefa que sempre foi difícil para bancos de dados relacionais.

Tipos de bancos de dados gerenciados

A maioria dos bancos de dados disponíveis também é oferecida como bancos de dados gerenciados. Além disso, os fornecedores de nuvem desenvolveram seus próprios bancos de dados nativos seguindo as tendências comuns do setor e oferecendo benefícios de performance. Os principais tipos de banco de dados compatíveis incluem:

Bancos de dados relacionais

Como já mencionamos, alguns bancos de dados na nuvem são versões gerenciadas de bancos de dados populares amplamente utilizados. Por exemplo, o Azure oferece o tradicional SQL Server da Microsoft. A Amazon oferece suporte aos bancos de dados MariaDB, MySQL, Oracle, PostgreSQL e SQL Server por meio do seu [Amazon Relational Database Service \(RDS\)](#). Essas ofertas ajudam a facilitar a migração de bancos de dados a partir de instalações on-premises.

Além disso, os fornecedores criaram seus próprios bancos de dados. Por exemplo, o Azure oferece o [Azure Cosmos DB](#), o Google oferece o [Cloud Spanner](#) e a AWS oferece o [Amazon Aurora](#).

Data warehouses

Normalmente, são bancos de dados relacionais, mas apresentam diferenças internas e em suas ofertas em relação aos bancos de dados orientados por transação. Por exemplo, os bancos de dados orientados por transação geralmente armazenam todas as colunas de uma única linha juntas, o que permite recuperar rapidamente várias colunas sobre um cliente ou produto. Por outro lado, os data warehouses na nuvem tendem a ser colunares. Isso quer dizer que armazenam os dados por coluna e não por linha. Isso acelera consideravelmente consultas comuns de data warehouse, como “quero saber a idade de cada um dos clientes que vivem em São Paulo”. Os fornecedores de nuvem com data warehouses oferecem ferramentas voltadas à rapidez de consumo e extração para facilitar sua utilização por aplicações de big data. O [Amazon Redshift](#), o [Google BigQuery](#) e o [Azure SQL Data Warehouse](#) são exemplos dessas ofertas.

Bancos de dados não relacionais

Geralmente, esse termo é utilizado para abranger diversos datastores diferentes que, ao contrário dos bancos de dados tradicionais, são criados para casos de uso de aplicações especiais. Os fornecedores de nuvem oferecem vários desses bancos de dados para finalidades diferentes:

- Os bancos de dados de *chave-valor* oferecem rapidez de armazenamento e recuperação de valores, com suporte a operações mais sofisticadas.

- Os bancos de dados de *documentos* armazenam dados como documentos JSON com um esquema flexível, o que permite armazenar e consultar os dados no mesmo formato usado pelas aplicações.
- Os bancos de dados de *grafos* armazenam relacionamentos entre objetos, facilitando a execução de algoritmos como encontrar o objeto com o maior número de conexões.
- Os bancos de dados de *pesquisas* otimizam a localização de documentos que contêm palavras específicas.
- Os bancos de dados de *séries temporais* registram eventos com time stamps e são otimizados para tarefas relacionadas a tempo, como colocar em gráficos ocorrências de eventos ao longo do tempo, o que é útil para monitorar eventos como leituras ou publicações de dispositivos na Web.
- Os bancos de dados de *ledger* são como arquivos de log aprimorados e seguros que registram atividades como transações financeiras, assinam essas atividades e as tornam imutáveis.

Outras opções, inclusive bancos de dados de armazenamento em cache na memória, também podem estar disponíveis. Os fornecedores de nuvem oferecem serviços para armazenamento menos estruturado (isto é, BLOBs ou arquivos). Entre os exemplos estão o [Amazon Simple Storage Service \(Amazon S3\)](#), o [Google Cloud Storage](#) e o [Microsoft Azure Blob Storage](#). Os serviços oferecem opções de armazenamento de arquivamentos ainda mais econômicas e ferramentas para transferir facilmente dados para dentro ou para fora do repositório.

Este capítulo, necessariamente, definiu recursos de bancos de dados em termos genéricos, com muitas palavras generalistas, como “alguns” ou “podem”. Essa relutância em usar termos mais específicos resulta das diferenças sutis que podem ser percebidas em um exame atento das ofertas dos fornecedores. Algumas ofertas evoluíram mais que as outras. Alguns bancos de dados oferecem administração ou migração mais fácil que os demais. Você deve considerar todos esses fatores quando escolher uma oferta de nuvem.

Quando um banco de dados gerenciado pode não ser adequado

Um serviço de banco de dados gerenciado não é adequado em alguns cenários.

O tamanho do banco de dados e as IOPS excedem os limites do banco de dados

Os bancos de dados nativos da nuvem são projetados para aumentar a escala e atender a demandas extremas. Mas os bancos de dados gerenciados baseados em mecanismos mais tradicionais de bancos de dados frequentemente impõem alguns limites superiores para seu tamanho e o número de Input/Output Operations per Second (IOPS – Operações de entrada/saída por segundo) que conseguem comportar. Esses limites podem ser suficientes para muitas organizações. Mas, se o banco de dados for realmente enorme e você precisar de mais IOPS que as oferecidas pelo provedor, os serviços de banco de dados gerenciado que você escolheu poderão se tornar inviáveis. Você deveria executar um banco de dados nativo da nuvem ou suas próprias instâncias na nuvem alugando as VMs e o armazenamento realmente necessários.

Você precisa ter controle total dos bancos de dados

Se ter controle máximo sobre os bancos de dados for importante para sua organização, você deve gerenciar seus próprios bancos de dados na nuvem. Se você precisar de uma versão específica de banco de dados não oferecida pelo fornecedor de nuvem ou tiver de usar recursos ou opções de banco de dados não disponibilizados pelo provedor, terá de encontrar sua própria solução.

Função do DBA em um banco de dados gerenciado

Como mencionado no início deste relatório, a migração para a nuvem mudará o seu trabalho, algumas vezes de maneira sutil. O [capítulo 2](#) abrange isso com mais detalhes. Resumindo, depois de transferir a responsabilidade pela manutenção de hardware, configuração do sistema e outras tarefas que um banco de dados gerenciado executa por você, ainda sobra bastante trabalho para você. Estas são algumas tarefas que o DBA costuma executar em um banco de dados gerenciado:

- Determinação de requisitos como capacidade de CPU e memória.
- Configuração de execuções de banco de dados inclusive a automação de escalabilidade, recuperação de falhas e obtenção de snapshots do banco de dados.
- Escolha das regiões e zonas de disponibilidade que hospedarão as instâncias de banco de dados.
- Criação e manipulação de bancos de dados e tabelas usando a linha de comando, o console ou a API oferecida pelo fornecedor.
- Autorização de contas de usuários, configuração de grupos e controle do acesso aos bancos de dados e suas partes. Alguns bancos de dados oferecem esse acesso por meio de Identity and Access Management (IAM – Gerenciamento de identidade e acesso) e outros usam as interfaces tradicionais oferecidas pelo banco de dados via implantação on-premise.
- Definição da divisão de dados em fragmentos quando for necessário distribuí-los entre diversos sistemas.
- Especificação de parâmetros de banco de dados que afetam a performance ou a resiliência.
- Auditoria de acesso e monitoramento de atividades do banco de dados para finalidades de segurança, performance e resiliência.
- Execução de consumo e transferência de dados por meio de extração, transformação e carga (ETL) ou ferramentas de dados de streaming mais recentes; essas transferências de dados podem envolver a aceitação de dados de fontes externas, o fornecimento de dados a ferramentas de big data como Spark ou Kafka, o armazenamento de dados em um data warehouse na nuvem ou on-premise e qualquer outro relacionamento que for preciso estabelecer com outra ferramenta para processar os dados.
- Determinação de esquemas ou formatos envolvendo arquitetos de dados e desenvolvedores de aplicações, bem como sua implantação em datastores relacionais ou não relacionais.

O mais interessante é que você terá mais tempo para examinar o contexto empresarial do banco de dados e pensar em formas de aprimorar o valor dos dados que você gerencia para a organização. Falaremos mais detalhadamente sobre esses tópicos no **capítulo 2**.

A evolução da função do DBA na nuvem

Tradicionalmente, um Database Administrator (DBA – Administrador de banco de dados) era responsável por praticamente todas as operações relacionadas ao armazenamento e ao uso dos dados de uma organização. A lista de tarefas de um DBA incluía:

- instalação de racks e cabeamento
- instalação, aplicação de correções e upgrades do software de banco de dados
- criação e configuração de instâncias do banco de dados
- gerenciamento de usuários, funções e permissões
- backup e recuperação
- proteção dos dados, inclusive a criptografia de dados cruciais
- migração, consumo e exportação de dados
- ajustes de performance
- monitoramento e solução de problemas de banco de dados
- administração da alta disponibilidade
- auxílio a desenvolvedores e analistas de banco de dados em suas tarefas relacionadas ao banco de dados

Como podemos ver nessa lista, o trabalho do DBA normalmente envolvia operações de banco de dados, gerenciamento de desastres, gerenciamento de usuários e ajustes de performance. Na maioria das vezes, o monitoramento e a solução de problemas eram manuais e, geralmente, ações reativas, em vez de planejadas.

Sobrava pouco tempo para a execução de tarefas mais elevadas na cadeia de valor, como projeto de bancos de dados e otimização das aplicações que os utilizam.

Com a falta de automação e a abordagem dos problemas de forma reativa, os DBAs tinham dificuldade para acompanhar todos os bancos de dados sob sua responsabilidade. Na maioria dos dias de trabalho, eles acabavam assumindo um modelo de triagem, no qual sempre executavam as tarefas mais urgentes.

A computação na nuvem altera profundamente as funções e as formas como as organizações fazem negócio e, consequentemente, reestrutura o trabalho tradicional de um DBA. Na nuvem, muitas tarefas tradicionais de um DBA simplesmente desaparecem ou são reduzidas substancialmente. Naturalmente, o provedor de nuvem lida com todo o trabalho de infraestrutura, como montagem dos racks e instalação física dos servidores, da rede e do armazenamento. Os backups e a segurança também são terceirizados, principalmente para o provedor de nuvem. Considerando tudo isso, o que os DBAs fazem na nuvem?

Se você estiver executando o seu próprio banco de dados na nuvem, ainda terá de realizar muitas tarefas tradicionais de um DBA, como instalação e aplicação de correções de software, backup dos bancos de dados etc. O foco deste relatório consiste em bancos de dados gerenciados, conforme definidos em **“Bancos de dados autogerenciados e gerenciados” na página 5**, pois eles oferecem diversos benefícios adicionais em relação aos bancos de dados autogerenciados, bem como alterações mais substanciais na função dos DBAs. Além disso, este capítulo também destaca bancos de dados relacionais, que normalmente exigem uma variedade maior de tarefas de gerenciamento.

Quando você usa um banco de dados gerenciado na nuvem, pode dedicar mais tempo à arquitetura de dados e às aplicações que usam o banco de dados. Agora, as principais preocupações de um DBA são prever necessidades e aprimorar o banco de dados. Você estará em uma situação muito melhor para extrair mais valor dos ativos de dados da organização, ajudando as equipes a entregar novos recursos com mais rapidez e ajustando proativamente a performance das aplicações.

Entre as outras tarefas estão ajudar a definir pipelines de dados em lotes e em tempo real para consumir e transformar dados em lotes e de streaming. A criação, a manutenção e o ajuste de pipelines de dados altamente distribuídos são funções importantes dos administradores que trabalham com bancos de dados baseados na nuvem. Além disso, você também pode dedicar mais tempo ao fortalecimento da segurança do banco de dados e à garantia do cumprimento dos requisitos de conformidade impostos por órgãos governamentais e setoriais.

Quando você usa um serviço de banco de dados gerenciado na nuvem, a função tradicional do DBA é invertida. A **figura 2-1** resume a diferença na forma como DBAs gastam seu tempo com um banco de dados on-premise e com um serviço de banco de dados gerenciado.

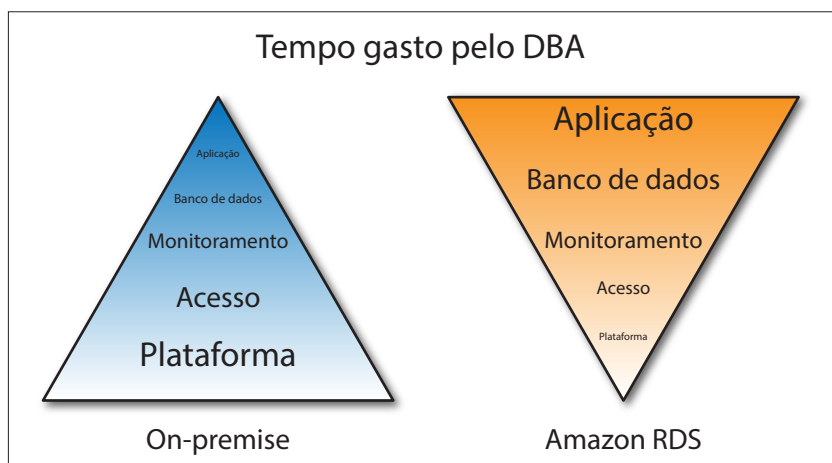


Figura 2-1. A função invertida do DBA na nuvem

Resumindo, com a migração de bancos de dados para a nuvem, administradores e desenvolvedores podem dedicar mais tempo a outras tarefas mais importantes. Como a manutenção, os backups e outras tarefas rotineiras são executadas pelo provedor de nuvem, você pode dedicar mais tempo a metas empresariais, como a otimização dos principais processos de negócios.

O fator crucial que determina o sucesso no novo ambiente de nuvem é familiarizar-se com todos os serviços relevantes do provedor de nuvem. Esses serviços incluem diversas opções de armazenamento, ferramentas de monitoramento (que frequentemente podem ser usadas sem custo adicional), recursos de segurança na nuvem e outros serviços de nuvem integrados, que podem ser explorados para executar o banco de dados na nuvem com economia, alta performance e segurança.

Como as tarefas do DBA mudam na nuvem

Quando os bancos de dados migram para a nuvem, várias tarefas operacionais desaparecem; algumas permanecem, mas com modificações; e diversas tarefas novas surgem entre as atribuições do DBA. No entanto, praticamente tudo o que era feito on-premise será alterado de alguma forma com a disponibilização de novas ferramentas e opções.

Tarefas do DBA assumidas pelo fornecedor

Uma diferença importante no uso de um serviço de banco de dados gerenciado na nuvem em relação a um banco de dados on-premise é que os DBAs executam tarefas administrativas sem ter acesso físico aos servidores ou controle direto sobre as instalações. Inicialmente, alguns DBAs podem levar certo tempo para se habituar à nova realidade. É importante que eles aproveitem essa oportunidade para ampliar seu conhecimento das novas tarefas discutidas neste capítulo.

Tarefas do DBA que permanecem, mas com modificações

Várias responsabilidades convencionais de DBA continuam sendo necessárias na nuvem, mas com diferenças sutis, tanto em um serviço de banco de dados gerenciado quanto executando o seu próprio banco de dados.

Provisionamento

Os serviços de banco de dados gerenciado, como o Amazon RDS, oferecem uma grande variedade de tipos de instância para os servidores de banco de dados. Em vez de comprar um sistema de hardware com requisitos específicos de memória e CPU para um datacenter, você escolhe um ou mais tipos de instância que oferecem a memória e a capacidade de CPU necessárias para cumprir os requisitos. A associação do tipo de instância às workloads pode ajudar a reduzir a possibilidade de pagar por recursos não utilizados.

A escalabilidade é outra tarefa de provisionamento. A escalabilidade automática, que pode ser aplicada a VMs ou a bancos de dados, permite que o sistema adicione uma instância quando as cargas aumentam e desative uma instância, quando a carga diminui, para evitar desperdiçar capacidade em excesso. Embora a escalabilidade automática seja uma forma de automação extremamente valiosa, você pode ser vítima de um ataque de DoS ou erro de configuração que gera muitas instâncias e aumenta os seus custos.

Além disso, você pode escolher regiões próximas aos usuários do banco de dados para oferecer um tempo de resposta ideal. O uso inteligente das ofertas de um provedor de nuvem global é um aspecto fundamental do provisionamento na nuvem.

Como já discutimos, os bancos de dados tradicionais, particularmente os usados para Online Transaction Processing (OLTP – Processamento de transações online), são conhecidos pelas dificuldades de escalabilidade. Com o crescimento do número de usuários, do tamanho dos conjuntos de dados e da complexidade das operações do banco de dados, aumentam

os tempos de resposta das consultas. O crescimento da empresa pode ser refletido em todos esses fatores, bem como o crescimento das aplicações, que aumentam a sobrecarga do banco de dados. Um desenvolvedor SQL e um DBA qualificados podem mitigar esses problemas até certo ponto mediante a definição de parâmetros de configuração ideais para o banco de dados e a otimização do código SQL.

No entanto, as otimizações de código e de configurações do banco de dados têm um alcance limitado. Assim, em muitos casos, a escalabilidade do banco de dados permanece como um problema sem solução, pois não é fácil alterar dinamicamente a quantidade de servidores, RAM ou CPUs para atender à evolução dos requisitos das workloads. Todos esses recursos envolvem gastos do orçamento de capital, que precisam ser aprovados com muita antecedência. Em muitos datacenters tradicionais, a inclusão de um só servidor físico leva semanas.

ETL, consumo de dados e exportação de dados

Uma das tarefas mais importantes para um DBA é a manipulação de dados em formas variadas, muitas vezes transformando-os de alguma maneira para extrair campos importantes ou adaptar os dados a outros formatos. Há cerca de 40 anos, esse campo vem sendo dominado por ferramentas de ETL, mas diversas formas recentes de processamento de streams, como **Hadoop** e **Spark**, resultaram na disponibilização de diversas novas ferramentas de consumo de dados. Agentes de mensagens, como o **Apache Kafka** e o **Amazon Kinesis**, direcionam os streams de dados para vários consumidores em um modelo de *publicar e assinar* (pub/sub). A maioria dos sites usa versões diferentes dessas ferramentas para finalidades diferentes. Você será responsável por dominar cada uma delas e configurá-las para que coexistam sem causar problemas.

O streaming de dados cria sobrecargas nos bancos de tradicionais, forçando os DBAs a ajustar os parâmetros de forma diferente. O streaming costuma introduzir mais processos que fazem muitas atualizações pequenas e frequentes, em vez de poucas atualizações maiores. Os controles de simultaneidade também sofrem uma sobrecarga.

As tarefas gerais continuam sendo as mesmas para dados enviados a um banco de dados on-premise ou a um armazenamento na nuvem. Mas os fornecedores de nuvem oferecem suas próprias versões de dados de streaming, bem como ferramentas para consumo e exportação de dados. Por exemplo, se usar um data warehouse baseado na nuvem da Microsoft, o **SQL Data Warehouse**, você poderá usar o software existente na sua caixa de ferramentas ou as ferramentas oferecidas pelo **Microsoft Azure Data Factory (ADF)**. A Amazon oferece várias ferramentas para consumo, como o Kinesis Data Firehose e o Snowball. O toolkit de consumo do Google combina opções de várias fontes.

Backups e snapshots

Os bancos de dados na nuvem oferecem dois métodos básicos para backup e restauração de bancos de dados. Normalmente, os backups automatizados são ativados por padrão e podem ser usados para recuperação point-in-time de bancos de dados.

Por outro lado, os snapshots são backups iniciados por você. Nos ambientes on-premises, os DBAs criam rotineiramente snapshots (descarregando dados do banco de dados enquanto estão em um estado consistente) para fazer backups ou replicar o datastore. Na nuvem, continua sendo possível usar um snapshot para backups de longo prazo (como backups para armazenamento em fita), replicação para outra região na nuvem e retenção para manter a conformidade regulamentar.

Os serviços de banco de dados gerenciado na nuvem, como o Amazon RDS, facilitam a criação de snapshots. Você pode criar snapshots de volumes de uma instância de banco de dados (que pode consistir em vários bancos de dados) e criar uma instância de banco de dados restaurando o snapshot do banco de dados.

Acesso

Na nuvem, você é responsável por tudo que é exclusivo dos seus dados e usuários: distribuição de contas, criação de grupos ou funções para combinar contas de usuários e atribuição de acesso a diversas partes do banco de dados. Falaremos mais sobre esse tópico em [“Segurança para dados e aplicações na nuvem” na página 19](#).

Manutenção de ambientes de desenvolvimento e teste

Uma tarefa comum dos DBAs é a configuração de ambientes de desenvolvimento e teste. Muitas vezes, essa tarefa envolve muito trabalho do DBA. Se você cria snapshots de banco de dados com regularidade, pode criar um ambiente de desenvolvimento ou teste que reflete o estado do banco de dados em um momento específico. Você pode encontrar exemplos desses serviços para [Amazon](#), [Google Cloud](#) e [Azure](#).

Logs

Os DBAs precisam de logs de banco de dados para solucionar problemas, manter a conformidade e fazer auditorias. O mesmo acontece na nuvem. A [figura 2-2](#) mostra como você pode visualizar facilmente os logs de uma instância de banco de dados usando o Console AWS. Quando você precisa monitorar e gerenciar vários bancos de dados, a utilidade desse recurso fica bem evidente.

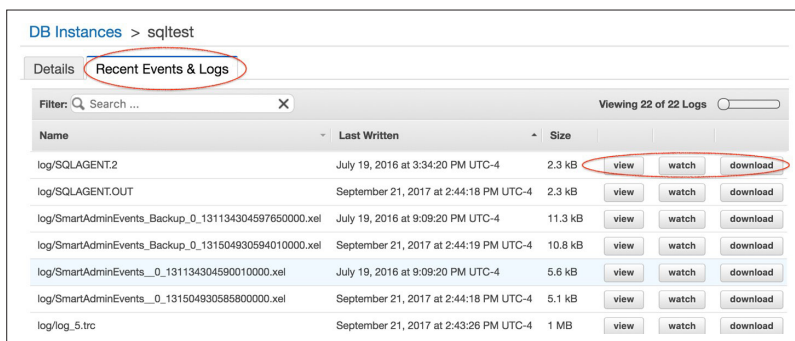


Figura 2-2. Visualização de logs de banco de dados no console do Amazon RDS

Monitoramento e notificações

Os fornecedores de nuvem oferecem muitas informações sobre métricas como utilização de CPU, latência de consultas, throughput de consultas e utilização de E/S e memória. Vários recursos de monitoramento incorporados são oferecidos, gratuitos ou pagos. Você também pode definir alertas para saber rapidamente se essas medições extrapolarem os limites previstos.

Muitas plataformas de nuvem expõem eventos como o momento em que ocorre um failover, um backup ou uma interrupção ou reconfiguração de banco de dados. As plataformas disponibilizam ganchos que permitem definir notificações ou executar outras ações automatizadas quando ocorrem eventos.

Otimização de performance

A otimização da performance do banco de dados é uma das principais tarefas de um DBA tradicional. Ela diferencia bons DBAs e administradores comuns. Os ajustes para obtenção de uma boa performance exigem uma compreensão detalhada de armazenamento, rede, memória e código de aplicações. Colocar os bancos de dados na nuvem não exime o DBA da completa responsabilidade sobre o banco de dados. O ajuste de performance, reativo e proativo, continua a ser uma responsabilidade primordial do DBA. No entanto, os fornecedores de nuvem oferecem alguns meios para obter facilmente essas informações, como o AWS [RDS Performance Insights](#).

Na nuvem, você não gastará muito tempo se preocupando com problemas de performance comuns relacionados a servidores e armazenamento, pois poderá aumentar facilmente a escala vertical e horizontal dos bancos de dados de acordo com as workloads. É mais provável que você dedique seu tempo à análise dos tempos de espera do banco de dados para aprimorar os tempos de resposta do banco de dados.

Em ambientes on-premises tradicionais, o próprio banco de dados oferece algumas ferramentas de performance, como o comando EXPLAIN e o log de consultas lentas do MySQL. Outras ferramentas de monitoramento de performance são executadas fora do banco de dados. Você pode configurar um sistema na nuvem para executá-las. No entanto, os fornecedores de nuvem desenvolvem continuamente ferramentas de performance executadas de forma nativa na nuvem, normalmente sem custo adicional. Você certamente deve utilizar as ferramentas oferecidas pelo fornecedor.

Os bancos de dados com performance insatisfatória on-premise não começarão magicamente a ter uma boa performance só porque foram migrados para a nuvem. No entanto, em alguns casos, a migração pode ser útil por si só. Por exemplo, quando o ambiente de banco de dados on-premise não tem capacidade suficiente para fornecer o grande número de IOPS exigidas por uma aplicação.

Administração da alta disponibilidade

Atualmente, esta tarefa é crucial: as pessoas abandonam um site que não **carrega em dois ou três segundos**. Em ambientes on-premises, essa tarefa exige a compra antecipada de capacidade de processamento adequada e a configuração de load balancers. Na nuvem, você pode escolher a escalabilidade automática e configurar o banco de dados para execução em várias zonas. Um banco de dados em várias zonas pode executar **automaticamente um failover para outra AZ** quando a primeira zona falha. Da mesma forma, você pode oferecer os dados em várias regiões para obter alta disponibilidade e armazenamento local.

Novas tarefas do DBA na nuvem

Quando você executa bancos de dados em um ambiente na nuvem, encontra tarefas com as quais, na sua maioria, não precisava se preocupar em um datacenter. Essas tarefas incluem uma nova orientação de custos e o monitoramento de licenças de banco de dados na nuvem.

Planejamento de custos operacionais

Os cálculos básicos do Total Cost of Ownership (TCO – Custo total de propriedade) e Return on Investment (ROI – Retorno sobre o investimento) mudam radicalmente na nuvem. Como mencionado em **"Provisionamento" na página 14**, você normalmente paga por instâncias e quantidades de tempo específicas na nuvem, em vez de comprar infraestrutura fixa antecipadamente. No entanto, é preciso considerar muitos outros fatores não diretamente relacionados ao banco de dados. Por exemplo, o posicionamento de bancos de dados em regiões mais próximas dos usuários pode reduzir os custos de rede e talvez até gerar novos negócios.

Os fornecedores oferecem várias opções para que você equipare o uso às necessidades e reduza consideravelmente as despesas. Por exemplo, a compra de instâncias reservadas na AWS pode oferecer reduções significativas de custo. Geralmente, um planejamento cuidadoso pode permitir que você preveja as necessidades em qualquer momento específico e elimine o provisionamento em excesso em datacenters on-premise.

Efeitos sobre o projeto de arquitetura e de aplicações

A arquitetura de bancos de dados baseados na nuvem é inerentemente muito diferente da arquitetura de um banco de dados tradicional on-premise em um datacenter. Você deve considerar as várias regiões geográficas e AZs, diferentes tipos de projeto de rede e outros recursos específicos da nuvem durante o planejamento das implantações de banco de dados. O DBA deve manter um estreito contato com os desenvolvedores de aplicações para compreender suas necessidades e oferecer os recursos mais adequados na nuvem com o menor custo possível.

Monitoramento da conformidade de licenças

Os bancos de dados proprietários são oferecidos na nuvem em dois modelos de licenciamento básicos: licença inclusa ou Bring Your Own License (BYOL – Traga a sua própria licença). O modelo BYOL é útil se você já tiver uma licença que permite usá-lo. Pode ser necessário consultar um especialista em licenciamento ou o fornecedor para saber se a licença existente pode ser usada na nuvem.

Segurança para dados e aplicações na nuvem

Normalmente, as introduções à computação na nuvem organizam as funções do administrador em uma pilha, com a instalação de servidores físicos e cabeamento na parte inferior da pilha e o gerenciamento de aplicações na parte superior. Os bancos de dados gerenciados designam a responsabilidade pela maioria dessas tarefas ao provedor de nuvem. Você, o cliente, fica apenas com as tarefas relacionadas a aplicações. Os bancos de dados autogerenciados mantêm a maioria das tarefas nas mãos do cliente. As tarefas atribuídas ao fornecedor são principalmente o fornecimento de servidores físicos e redes básicas.

As tarefas de segurança também se adaptam a esse conceito de pilha, resultando em um modelo geralmente conhecido como segurança compartilhada. Nos bancos de dados gerenciados, o fornecedor assume mais tarefas de segurança que nos bancos autogerenciados. Resumindo, o provedor de nuvem supervisiona a segurança da infraestrutura, mas você continua responsável pela segurança dos seus dados e informações de usuários na nuvem. Sendo assim, estas são as

principais tarefas executadas pelo fornecedor de nuvem incluem:

- Segurança física do datacenter e dos equipamentos, incluindo a supervisão da equipe.
- Garantia de redundância por meio de replicação e backups de acordo com as diretrizes do cliente.
- Correções e atualizações de software, incluindo o próprio banco de dados no modelo autogerenciado.
- Execução de ferramentas de monitoramento e auditoria de acesso.
- Fornecimento de firewalls de rede e aplicações.
- Proteção contra ataques DoS.
- Serviços de gerenciamento de identidade, que você mesmo configura.

As tarefas de segurança do cliente são estas:

- Definir usuários e funções por meio do sistema de Identity and Access Management (IAM – Gerenciamento de identidade e acesso) do fornecedor.
- Atribuição de direitos de acesso aos usuários no banco de dados, inclusive a definição de contas e funções e a manutenção das contas atualizadas mediante a remoção de usuários que não precisam mais ter acesso.
- Especificação de regras de firewall, backups e outros parâmetros oferecidos pelo fornecedor de nuvem.
- Configuração de alertas e revisão de logs para detectar acessos não autorizados.
- Segurança do sistema operacional, caso você use sua própria VM em vez de um banco de dados gerenciado.
- Segurança de aplicações, como a prevenção de ataques de injeção de SQL.
- Criptografia de dados ociosos e em trânsito.
- Autenticação e autorização adequadas de usuários que tentam acessar os bancos de dados ou as aplicações.
- Coleta e inspeção de logs.
- Configuração de alertas de eventos e monitoramento de eventos para detectar anomalias.

A familiaridade com todos os recursos de segurança na nuvem é crucial para executar bancos de dados na nuvem. Nas subseções a seguir, descrevemos os principais recursos de segurança da nuvem nos quais os DBAs devem concentrar seus estudos: controle de acesso e IAM, isolamento de redes e criptografia de dados.

Controle de acesso e gerenciamento de identidade e acesso

Geralmente, os provedores de nuvem se destacam pelos sólidos mecanismos de controle de acesso. Os fornecedores mencionados neste relatório usam IAM centralizado para gerenciar usuários, credenciais de segurança (senhas, chaves de acesso e permissões) e políticas de autorização que controlam quais recursos e serviços usuários podem acessar. Os administradores precisam dominar o IAM para obter acesso à nuvem para si mesmos e para os usuários. Além disso, os bancos de dados nativos da nuvem ocasionalmente são integrados a ferramentas gerais de IAM na nuvem. A vinculação das proteções do banco de dados ao IAM da nuvem oferece as políticas de acesso mais fáceis e seguras.

Você usará o IAM para definir contas de usuário e adicionar a elas regras de acesso específicas do banco de dados. Usando o IAM, você pode conceder permissões de usuário diferentes para executar operações distintas no banco de dados. Você também pode instituir controles de acesso detalhados (oferecidos por muitos bancos de dados, como restrições para linhas ou colunas específicas) por meio do IAM.

Os bancos de dados tradicionais usam na nuvem as mesmas ferramentas que você usa on-premise, como instruções GRANT no SQL. Talvez seja possível **conectá-las ao IAM** para que você possa usar as mesmas contas de usuário na nuvem e no banco de dados, além de se beneficiar da segurança e comodidade extras oferecidas pela integração.

Isolamento de rede

As VPCs, os firewalls e as listas de controle de acesso (ACLs) são alguns recursos de nuvem comuns para a proteção de sistemas.

Como discutido em [“Resumo dos efeitos da migração para a nuvem” na página 2](#), uma VPC é uma rede privada dentro da nuvem para comunicação entre seus servidores. Em uma VPC, você pode isolar instâncias de banco de dados especificando o intervalo de IPs que podem acessar cada banco de dados. A organização que cria uma VPC tem controle total sobre o ambiente de redes virtuais e pode selecionar seus próprios intervalos de endereços IP, criar sub-redes e configurar suas próprias tabelas de rotas e gateways de rede.

Além disso, você pode configurar um gateway privado virtual que estende a rede corporativa à VPC e permite acesso às instâncias de banco de dados nessa VPC por meio de uma VPN que você escolher.

A maioria dos provedores de nuvem oferece firewalls incorporados, que ajudam a controlar o acesso às instâncias de computação pela rede. Além disso, o provedor de nuvem pode oferecer uma opção de conectividade privada ou dedicada para conectar o escritório e os ambientes on-premises do cliente de nuvem ao ambiente de nuvem. Você pode configurar grupos de segurança de banco de dados para proteger instâncias de banco de dados dentro de uma VPC. Os grupos de segurança são regras de firewall que controlam o acesso a bancos de dados na nuvem pela rede. Também é possível permitir ou negar a entrada e saída de tráfego da rede em uma sub-rede por meio de ACLs de rede. Todo tráfego de rede que entra ou sai da VPC por meio da VPN pode ser inspecionado pela infraestrutura de segurança on-premise (como um firewall) e por sistemas de detecção de invasão.

OBSERVAÇÃO

Conexões diretas

Em vez de uma VPN, você pode conectar sistemas fora da nuvem a sistemas dentro da nuvem por meio do **Direct Connect** na AWS ou do **ExpressRoute** no Azure. Eles exploram links de redes privadas fornecidos por operadoras de telecomunicações para criar uma conexão direta. A criptografia de ponta a ponta ainda é recomendada e geralmente é realizada por meio de Secure Sockets Layer (SSL)/Transport Layer Security (TLS).

Criptografia de dados

Na nuvem, é fácil criptografar os dados para oferecer proteção adicional aos dados ociosos. Por exemplo, quando você habilita a criptografia em um cluster do Amazon RDS, o banco de dados armazena todos os dados em um formato criptografado nas tabelas criadas. A criptografia também se aplica a backups de bancos de dados. Além disso, é fácil configurar a criptografia para a transmissão de dados com origem ou destino na nuvem.

A criptografia é particularmente necessária para organizações que devem cumprir requisitos de conformidade de sua área de atuação, como as normas Health Insurance Portability and Accountability Act (HIPAA – Lei de Portabilidade e Responsabilidade de Seguros de Saúde) para a área de saúde, a Sarbanes-Oxley Act (SOX – Lei Sarbanes-Oxley) para relatórios financeiros e o Payment Card Industry Data Security

Standard (PCI DSS – Padrão de Segurança de Dados do Setor de Cartões de Pagamento) para empresas de comércio eletrônico e varejo. Se você proteger as chaves de criptografia, a perda de dados criptografados não é geralmente considerada um evento de segurança que deve ser comunicado.

Os provedores de nuvem costumam oferecer serviços gerenciados para simplificar a criação, o controle e o gerenciamento de chaves de criptografia. Dessa forma, o AWS Key Management Service (KMS) proporciona uma visão centralizada de todas as chaves de usuário na organização. Ele é integrado ao AWS CloudTrail para fornecer um log que mostra o uso das chaves em toda a organização, cumprindo assim vários requisitos regulatórios e de conformidade importantes.

Infraestrutura como código: como tirar o máximo proveito da nuvem

Ao migrar bancos de dados para a nuvem, alguns clientes ficam tentados a reduzir ao máximo as alterações de gerenciamento porque já estão acostumados a fazer as coisas de determinada maneira. No entanto, essa estratégia significa que você está simplesmente colocando vinho velho em garrafas novas e deixando de aproveitar os imensos benefícios disponibilizados pela nuvem.

Um avanço tecnológico importante que diferencia sistemas baseados na nuvem dos que residem em datacenters on-premise é a fácil disponibilidade de sistemas de configuração que tratam a infraestrutura como código. O tratamento de infraestrutura como código possibilita várias práticas de DevOps que, por sua vez, facilitam uma colaboração estreita entre as áreas de desenvolvimento e operações e a automação da entrega de aplicações em grande escala. No ambiente local, você certamente pode instalar ferramentas de automação como sistemas de gerenciamento de configuração (Puppet e Chef são dois exemplos populares) e ferramentas de integração contínua (CI) (por exemplo, Jenkins). No entanto, a nuvem incorpora ferramentas desse tipo altamente sofisticadas.

Por exemplo, os modelos fornecidos com o AWS CloudFormation ajudam você a modelar toda a infraestrutura como código. Você pode definir uma infraestrutura na nuvem criando e configurando recursos como tabelas de banco de dados e buckets de armazenamento do Amazon S3, e tratar esses recursos como código. É possível registrar modelos do AWS CloudFormation no sistema de controle de código-fonte e gerenciá-los da mesma forma que os desenvolvedores gerenciam arquivos de código.

Uma ferramenta como o CloudFormation oferece os seguintes benefícios:

- Ajuda a estabelecer uma única fonte confiável para todos os recursos da nuvem.
- Pode ser integrada a ferramentas de gerenciamento de código, como um sistema de controle de código-fonte.
- Ajuda a automatizar implantações de desenvolvimento e teste.
- Apoia planos de recuperação de desastres.

O uso de arquivos de modelos para criar recursos de forma programática permite alcançar repetibilidade e consistência em todos os ambientes.

Como migrar os bancos de dados para a nuvem

Como a maioria das transformações organizacionais, a migração para a nuvem não é realizada de um dia para o outro. Você deve escolher um projeto para experimentar no provedor de nuvem escolhido: um projeto existente, que oferece um caso de teste útil, ou um novo projeto livre de práticas herdadas.

Praticamente todas as organizações que migram para a nuvem começam com uma migração do tipo prova de conceito, usando um banco de dados não essencial. Depois que você realizar uma migração fácil ou migrar um projeto pequeno que obtém benefícios do uso da nuvem, poderá estender o que aprendeu para outros bancos de dados em toda a organização.

Este capítulo ajuda a compreender os critérios normalmente associados a uma primeira migração bem-sucedida. Boa parte do nosso material se aplica a bancos de dados relacionais. Embora cada migração seja única, você provavelmente executará a maioria ou todas as seguintes etapas:

Planejamento

- coleta de requisitos
- determinação dos recursos que cumprem os requisitos
- avaliação dos bancos de dados a serem migrados e das possíveis migrações necessárias nos bancos de dados ou nas aplicações que os utilizam
- estabelecimento de critérios de sucesso e de reversão (à prova de falhas)

Migração dos dados

- replicação
- incorporação de mudanças desde a criação da réplica
- testes de aplicações
- transição
- verificações pós-migração

Otimização

- ajustes de performance
- projeto de alta disponibilidade
- determinação dos eventos a registrar em log e monitorar
- criação de um plano de recuperação de desastres

Examinaremos todas as etapas principais nas próximas seções.

Planejamento

Esta etapa ajuda uma organização a avaliar os seguintes problemas:

- Elementos relevantes do ambiente atual: aplicações, bancos de dados e workloads cruciais.
- Execução adequada da aplicação ou workload no ambiente do provedor de nuvem.
- Modos como a migração ajudará a alcançar objetivos empresariais.
- Requisitos do estado final do sistema.
- Custo da execução do ambiente de computação atual na nuvem, que deve representar um amplo retorno sobre o investimento, como explicado em [“Planejamento de custos operacionais” na página 18](#).

Fatores em uma migração

Quando você quer migrar um banco de dados executado on-premise para um serviço gerenciado na nuvem, precisa lidar com problemas físicos, organizacionais e de software. Isso exige atenção a diversos níveis de mudança:

Migração física dos dados

Você precisa colocar os próprios dados na nuvem. A transferência de dados coletados em grandes bancos de dados pode ser demorada e cara, mas há serviços de migração que podem ajudar.

Compatibilidade de infraestrutura

Verifique todos os aspectos do serviço de nuvem em relação aos sistemas herdados on-premise. Um [estudo de caso no blog da AWS](#) oferece um exemplo interessante dos obstáculos que podem surgir durante uma migração de teste, decorrentes de incompatibilidades de software inesperadas e pequenos erros. Essa migração específica foi suspensa por algum tempo devido a diferenças de implantação de SSL entre o sistema on-premise e o provedor de nuvem.

Compatibilidade de banco de dados

Provavelmente, o banco de dados na nuvem terá algumas diferenças em relação ao banco de dados on-premise que afetarão a migração. Sutilezas como conjuntos de caracteres de banco de dados e permissões de procedimentos armazenados podem causar problemas. Uma migração homogênea (como a de um banco de dados Oracle para outro banco de dados Oracle) provavelmente será mais tranquila que uma migração heterogênea (como de um banco de dados Oracle para um banco de dados PostgreSQL ou um dos bancos de dados proprietários do fornecedor de nuvem). Em ["Verificação de incompatibilidades" na página 31](#), há explicações sobre o que você pode fazer para reduzir imprevistos entre os bancos de dados. Para certos tipos de migração, os fornecedores de nuvem podem oferecer ferramentas que facilitam parte do seu trabalho.

Se você estiver executando uma versão muito antiga de um banco de dados on-premise, pode ser o momento de fazer upgrade. Em vez de simplesmente escolher uma nova versão do mesmo Relational Database Management System (RDBMS – Sistema de gerenciamento de banco de dados relacional), o que envolverá alguns problemas, você poderá aproveitar a oportunidade para tentar uma migração heterogênea para um novo RDBMS com futuro promissor.

Mudanças organizacionais

Pessoas de várias equipes, particularmente DBAs e desenvolvedores de aplicações, precisarão dedicar um tempo para aprender as ferramentas de nuvem e o gerenciamento de novas instâncias, sem deixar de cuidar da execução de sistemas herdados on-premise, até que você esteja pronto para migrar tudo para a nuvem. Você pode

optar por levar meses ou até mesmo anos para fazer uma migração completa. Outra opção possível é manter permanentemente on-premise uma parte dos dados (ou uma cópia dos dados).

Principais tarefas de migração

Sugerimos as seguintes etapas para uma migração bem-sucedida.

Crie um plano de migração para a nuvem

O plano de migração para a nuvem deve relacionar todos os bancos de dados e aplicações na ordem em que você quer transferi-los para a nuvem. O plano de avaliação final deve descrever o plano de migração de todos os bancos de dados. Se novos recursos financeiros, pessoais ou de software forem necessários no novo ambiente de banco de dados, deverão estar incluídos no plano.

Determine quem executa a migração

Os DBAs e desenvolvedores devem trabalhar em conjunto no esforço de migração porque cada equipe tem algo a oferecer. Ao escolher a equipe, você também pode optar por utilizar um serviço de migração oferecido pelo fornecedor de nuvem ou empresa terceirizada.

Tarefas educacionais

Incluem tarefas genéricas como compreender o uso das ferramentas do fornecedor de nuvem, os recursos específicos do mecanismo de banco de dados de destino, o escopo da migração do banco de dados e a arquitetura dos bancos de dados na nuvem.

Crie a arquitetura de banco de dados na nuvem

Selecione o tipo e o modelo (gerenciado ou autogerenciado) de banco de dados que deseja utilizar. Ao fazer essa escolha, pondere todos os fatores relevantes como custo, performance, confiabilidade e escalabilidade.

Escolha um processo de migração

Em **"Migração do banco de dados" na página 32**, você encontrará as definições dos critérios para escolher entre replicação, backup/restauração ou um serviço dedicado de migração.

Para um banco de dados autogerenciado, crie a infraestrutura de computação

Se você optar por gerenciar seus próprios bancos de dados na nuvem, crie instâncias virtuais e as implante usando o serviço de computação do fornecedor, como o Amazon Elastic Compute Cloud (Amazon EC2).

Determine oportunidades para redefinir a arquitetura

Você pode perceber que as escolhas feitas no ambiente on-premise não são as mais adequadas para o novo ambiente na nuvem. Por exemplo, você pode conseguir consolidar fragmentos.

Configure as contas de banco de dados na nuvem

Configure as contas gerais de usuários na nuvem usando o IAM antes de configurar usuários, funções e grupos no banco de dados.

Faça upgrade das aplicações e teste-as

Pode ser necessário alterar o código para que funcione com o novo banco de dados. Algumas migrações de aplicações podem ser automatizáveis, como a do próprio banco de dados.

Avaliação da prontidão

Uma avaliação da prontidão ajuda você a estimar os custos, a arquitetura do banco de dados, os planos de migração e o impacto da migração para a nuvem sobre os regulamentos de conformidade. O resultado de uma avaliação de prontidão para a nuvem é um relatório detalhado sobre a preparação da empresa para migrar seus bancos de dados para a nuvem.

Veja a seguir breves descrições das principais etapas de uma avaliação de prontidão.

Entrevistas com as partes interessadas

Conversas com desenvolvedores de aplicações, usuários empresariais e outros envolvidos no uso dos dados na equipe da organização ajudarão a determinar os requisitos de performance e alta disponibilidade, bem como os recursos e as funcionalidades dos bancos de dados na nuvem.

Análise dos bancos de dados on-premise atuais

Volte aos bancos de dados existentes para determinar padrões de crescimento de dados, estratégias de backup e recuperação, exportações e importações contínuas etc. A compreensão dos padrões de uso atuais do banco de dados será útil no momento de decidir quais bancos de dados serão usados na nuvem e se você precisará de uma determinada oferta de banco de dados para obter os recursos necessários.

Priorização dos bancos de dados que serão migrados

Escolha os primeiros bancos de dados que você quer migrar para a nuvem. Um critério importante é a extensão das mudanças que serão necessárias, o que exige coordenação com

os desenvolvedores.

Análise de custos de migração

Execute uma análise detalhada dos custos da migração para a nuvem em relação à permanência nos data centers on-premise para compreender totalmente o TCO e o ROI da migração para a nuvem.

A redução dos custos usando um serviço de banco de dados na nuvem pode ser um dos principais objetivos. Mas uma escolha equivocada de serviços de nuvem pode inviabilizar esse objetivo.

Segurança e conformidade

Ocasionalmente, você precisa de regiões ou AZs especiais para cumprir padrões como Service Organization Control 2 (SOC 2 – Controle de organização de serviços 2), PCI DSS ou HIPAA dos EUA. Felizmente, os provedores de nuvem costumam associar regiões e AZs específicas a esses requisitos legais. Além disso, há regiões especializadas como a AWS GovCloud (EUA), uma região da AWS isolada e sujeita às linhas de base High e Moderate do FedRAMP. Por fim, você precisa verificar se pode cumprir os Service Level Agreements (SLAs – Acordos de nível de serviço) da organização na configuração de nuvem escolhida. Geralmente, esses SLAs incluem métricas de manutenção planejada, backup, Recovery Point Objectives (RPOs – Objetivos de ponto de recuperação) e Recovery Time Objectives (RTOs – Objetivos de tempo de recuperação).

Se a sua avaliação de nuvem estiver correta, você também deverá ter uma boa ideia do escopo da migração para a nuvem. Algumas aplicações podem ser tão antigas que precisarão ser totalmente reprojatadas para serem migradas para a nuvem. Se você não tiver os recursos para reprogramar as aplicações, poderá mantê-las nos bancos de dados on-premise ou migrar os servidores inteiros para a nuvem.

Etapas de migração

Você deve planejar exatamente quando e como será feita a migração dos dados, bem como os preparativos para reverter a migração caso testes posteriores mostrem que o banco de dados não está funcionando adequadamente. Estabeleça testes e critérios de sucesso para cada etapa.

Mudanças de aplicações

Esta etapa deve ser realizada de forma coordenada com a equipe

de desenvolvimento. Isso inclui o agendamento das mudanças de aplicações nos processos e testes de desenvolvimento.

Automação da migração

Esse recurso é útil porque permite repetir a migração conforme a necessidade, corrigindo erros à medida que são detectados ou alterando parâmetros. Você também pode adaptar e reutilizar a estrutura de trabalho de automação para migrações posteriores.

Verificação de incompatibilidades

Se você usar o mesmo RDBMS (como MySQL, por exemplo) para ambos os bancos de dados, verifique as versões para determinar se está usando recursos incompatíveis com a versão na nuvem. Obviamente, é mais provável que essas verificações revelem incompatibilidades quando você muda para um novo mecanismo de banco de dados.

Os problemas de conversão encontrados durante uma migração heterogênea para um banco de dados na nuvem são os mesmos de uma migração on-premise. No entanto, você tem acesso a uma maior variedade de soluções na nuvem.

Geralmente, os próprios fornecedores de banco de dados oferecem ferramentas para ajudar a migrar dados para um banco de dados do mesmo tipo. Por exemplo, o Oracle é fornecido com ferramentas de migração de dados como Pump, RMAN e SQL Developer, que ajudam a migrar bancos de dados do Oracle para a nuvem. Além disso, provedores de nuvem e terceiros oferecem serviços para migrar esquemas, como o [AWS Schema Conversion Tool](#). Esses serviços ajudam a determinar se os esquemas podem ser convertidos.; Eles também oferecem sugestões para soluções alternativas, se necessário.

Entretanto, a maioria das migrações heterogêneas de bancos de dados exige ferramentas automatizadas e ações manuais do DBA. Se você estiver transferindo um banco de dados para a AWS, pode aproveitar as [recomendações específicas do site da AWS](#).

Algumas ferramentas de migração usam uma ferramenta para converter os objetos de um banco de dados, como os procedimentos armazenados do Oracle, em objetos de um banco de dados diferente, como o PostgreSQL. No entanto, o trabalho realizado por essas ferramentas pode ser incompleto, pois cada mecanismo de banco de dados usa práticas de codificação exclusivas.

Esse é o motivo da realização de testes rigorosos no banco de dados de destino para verificar a precisão e a performance após a migração para

nuvem. Algumas técnicas de programação utilizadas por um RDBMS específico podem não ter um equivalente exato no banco de dados de destino. A ferramenta de migração pode detectar esses tipos de problemas de conversão durante o processo de migração. Em muitos casos, ela também pode mostrar o código que talvez seja necessário executar para resolver esses problemas e concluir corretamente a migração do banco de dados.

Migração dos dados

Os provedores de nuvem estão em alerta para os requisitos de clientes em potencial para a migração de bancos de dados on-premise para a nuvem com o mínimo de custos e interrupções. Cada fornecedor oferece ferramentas para acelerar a migração. A AWS, por exemplo, oferece vários artigos que descrevem detalhadamente como migrar diversos bancos de dados on-premise para a AWS, incluindo procedimentos para [Oracle](#), [MySQL](#) e [PostgreSQL](#).

Mesmo assim, as migrações para a nuvem podem envolver interrupções dispendiosas. Uma ferramenta de migração, oferecida pelo fornecedor de nuvem ou por um terceiro, deve ser capaz de lidar com todos os aspectos do banco de dados, como esquemas, permissões de usuários, triggers e procedimentos armazenados.

Recomendamos que você documente em um diário as primeiras migrações, pois as lições aprendidas durante o processo ajudarão você e seus colegas em migrações posteriores. Se ocorrerem problemas graves suficientes para que você decida usar outro fornecedor (ou abandonar a migração), o diário terá evidências importantes para apoiar essa decisão.

Migração do banco de dados

Muitos provedores de nuvem e terceiros oferecem serviços de migração, como o [Microsoft Azure Database Migration Service](#) e o [AWS Database Migration Service](#), que valem a pena considerar. Esses serviços são divulgados como rápidos, sem imprevistos e fáceis de usar. No entanto, você pode optar por processos existentes que já conhece, como restaurar um backup para um novo ambiente de banco de dados e usar replicação.

Se usar um serviço de migração de banco de dados, você poderá manter os bancos de dados de origem totalmente ativos durante a migração para a nuvem e minimizar o tempo de inatividade das aplicações que dependem do banco de dados. Isso facilita alcançar o objetivo importante de tempo de inatividade mínimo, citado por muitas organizações que migram para a nuvem. A [figura 3-1](#) mostra como o AWS Database Migration Service cria tabelas, carrega dados e mantém as tabelas sincronizadas com as tabelas do banco de dados de origem.

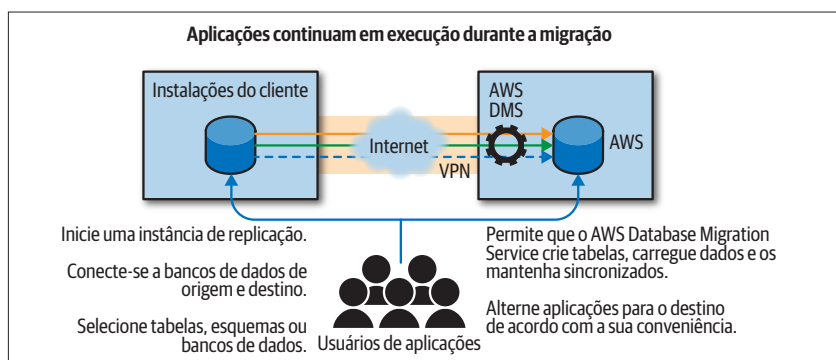


Figura 3-1. Disponibilidade das aplicações durante a migração do banco de dados

Geralmente, o processo de migração é baseado em um processo típico de backup/restauração. As tarefas dependem da possibilidade de desativar o banco de dados durante a transferência. Se for necessário que o banco de dados permaneça em execução, esta será a sequência básica a ser seguida:

1. Faça um backup dinâmico do banco de dados original enquanto está em execução.
2. Copie e restaure o banco de dados na nuvem.
3. Habilite a replicação para incorporar à versão da nuvem todas as alterações efetuadas no banco de dados original após a criação do backup.

Obviamente, o backup e a restauração serão mais simples se for possível desativar o banco de dados. Neste caso, a migração segue este padrão:

1. Interrompa o banco de dados original.
2. Faça um backup.
3. Copie e restaure o banco de dados na nuvem.
4. Inicie a nova versão na nuvem.

No entanto, antes de colocar a nova versão em produção, você provavelmente passará um tempo executando testes para verificar se as aplicações funcionam, se a segurança foi implantada e a performance necessária foi obtida. Portanto, pode ser necessário reiniciar a versão original antes que você possa confiar na nova versão. Nesse caso, você usará a replicação para sincronizar a nova versão com a original.

Migração de aplicações

Há duas estratégias básicas para migrar os bancos de dados e as aplicações relacionadas:

Transferência (lift-and-shift)

Migrar um banco de dados inteiro, sem modificações, com todas as aplicações que o utilizam, para o novo banco de dados na nuvem. Essa é a estratégia de migração de aplicações tipo “vinho velho em garrafas novas”, que usa poucos, ou nenhum, recursos nativos da nuvem.

Reprojetar os bancos de dados

Repensar a forma como bancos de dados on-premise lidam com tarefas administrativas, como escalabilidade e alta disponibilidade, e investigar as ofertas do provedor de nuvem nessas áreas. Essas ofertas ficam cada vez mais sofisticadas. Naturalmente, essa estratégia demanda mais trabalho e esforço. Também pode ser necessário alterar o código de aplicações para ajustá-las às mudanças na estratégia de banco de dados.

Muitas vezes, os clientes começam com a estratégia de transferência e continuam com o reprojeto e outras modernizações.

Verificações pós-migração

Depois que os dados forem totalmente copiados para o banco de dados da nuvem, valide o banco de dados de destino para garantir que todos seus objetos estejam presentes. Vários testes determinam se a migração foi bem-sucedida:

Validar dados

Essa validação pode ser bem simples, como verificar o número de linhas ou realizar uma soma de verificação para conferir se não houve perda ou corrupção de dados. Os processos de validação mais complexos consideram as mudanças de esquema inerentes às migrações heterogêneas. Infelizmente, essa validação básica é frequentemente esquecida.

Funcionalidade básica

Execute testes completos para garantir que a migração foi bem-sucedida e que o sistema opera da forma pretendida. Teste a utilização de recursos que tendem a ser diferentes entre mecanismos e versões de bancos de dados, como triggers e procedimentos.

armazenados. *Testes de performance*

Realize testes de estresse e testes comparativos de performance nos novos sistemas para ter uma ideia de onde você pode aprimorar o tempo de resposta e o throughput.

Avaliação de segurança

Verifique se os sistemas da nuvem estão seguros executando avaliações de vulnerabilidades e testes de intrusão.

Essas etapas também compõem a transição para a próxima fase, a otimização. Esta fase é uma sequência da migração bem-sucedida dos bancos de dados e aplicações on-premise para a nuvem. Os elementos cruciais dessa fase são gerenciamento de performance e otimização de custos. O monitoramento cuidadoso e contínuo dos novos sistemas é crucial para que você possa reverter rapidamente para os sistemas anteriores caso ocorram problemas inesperados.

Otimização

Neste ponto, você provavelmente está operando sem imprevistos e pode parabenizar toda a equipe por um trabalho bem executado. Esta seção resume as principais tarefas que o DBA e outros membros da equipe devem realizar ao continuar o processo.

Disponibilidade

Naturalmente, este é o requisito fundamental que você deve cumprir em qualquer configuração. Felizmente, a migração para a nuvem remove alguns motivos de falha (se você configurar adequadamente o reinício de serviços que apresentarem falha). Além disso, o fornecedor de nuvem oferece diversas ferramentas para previsão e alertas de falha. Para aproveitá-las, planeje esta etapa (muitas destas atividades também são úteis na otimização de performance):

Plano de recuperação de desastres (DR)

Provedores de nuvem de grande porte e robustos oferecem failover para uma nova AZ em caso de falha da AZ atual. Mas você precisa de um plano e de um processo automatizado para o failover. O plano deve abranger o RTO, o RPO e a redundância geográfica. Explore ao máximo os recursos de failover disponibilizados pelo fornecedor em vez de desenvolver os seus próprios recursos.

Monitoramento de logs e sistemas

Determine os eventos que podem indicar problemas iminentes ou que já ocorreram. Esses eventos podem ser incorporados a um monitoramento automático. O monitoramento deve oferecer informações suficientes para indicar as origens da falha. Por exemplo, se a falha resultou de uma ação do usuário (como reinicialização de um serviço), de um ataque ou de outras alterações no ambiente. Algumas falhas podem ser consideradas normais e resolvidas por ferramentas automatizadas. Nesse caso, elas devem ser registradas, mas não é necessário alertar o administrador.

Monitoramento de mudanças

Os administradores devem sempre saber quais mudanças de configuração, dimensionamento de instâncias ou topologia de cluster de banco de dados afetam a disponibilidade. Os ambientes de desenvolvimento modernos usam processos robustos para o monitoramento de mudanças e controle de versões para que cada migração passe por um processo de triagem e possa ser revertida.

Testes do sistema

Tente determinar os pontos fracos do sistema e prever falhas. Algumas falhas realizam exercícios *pre-mortem* para identificar e remover possíveis fontes de falha. Sites de grande porte podem desativar deliberadamente sistemas e observar se a recuperação é adequada. Esse tipo de teste, denominado *engenharia do caos*, foi popularizado por meio do **Chaos Monkey** da Netflix. Da mesma forma que executa restaurações para verificar o funcionamento dos backups, você deve testar os procedimentos de recuperação.

Otimização de performance

A performance pode se beneficiar das práticas da seção anterior, particularmente o monitoramento. O monitoramento de performance deve permitir determinar as relações entre eventos e mudanças nas métricas do banco de dados, bem como observar as discrepâncias entre as tendências de performance previstas e reais. Além disso, processos adicionais podem realizar manutenção e aprimoramento da performance.

Teste de workloads

O crescimento do tamanho e da complexidade dos dados, assim como as mudanças no comportamento das aplicações, afetam a performance. Teste regularmente a performance para detectar a degradação antes dos clientes para poder alterar a escala ou fazer outras mudanças corretivas. Após uma mudança importante, é necessário algum tempo para a atualização do cache do banco de dados e das estatísticas. Isso quer dizer que os planos de consulta e a performance geral podem demorar um pouco para voltar aos níveis desejados.

Otimização distribuída

Os ambientes de aplicações evoluem rapidamente e os desenvolvedores compreendem as consultas de banco de dados melhor que o DBA. Se eles assumirem os testes e as otimizações de consultas, poderão conseguir uma boa performance mais rápido que o DBA que, por sua vez, terá mais tempo livre para otimizações e planejamento no longo prazo.

Cientes da importância da performance, os fornecedores de nuvem incluíram rastreamentos e disponibilizaram muitas estatísticas aos usuários. Por exemplo, o AWS RDS Performance Insights permite obter dados históricos em um painel e visualizar gráficos que mostram a carga imposta ao banco de dados por operações, instruções SQL ou usuários específicos. Você pode consultar os painéis para responder a perguntas simples, como “quando os hosts ficam sobrecarregados?”, ou mais complexas, como “por que esta consulta tem uma performance abaixo da ideal?”. O Azure oferece monitoramento de performance e **ajustes automáticos** de performance. Terceiros também produzem ferramentas para monitoramento e performance na nuvem.

Adaptação às diferenças da nuvem

Por fim, os próprios fornecedores disponibilizam novos serviços, novas opções de VM, novo hardware (como SSDs) e outras alterações que podem proporcionar a você grandes benefícios. Esteja sempre ciente do que o fornecedor pode fazer por você e transfira seu trabalho para os processos bem testados e padronizados do fornecedor. Quando fizer isso, atualize seus procedimentos operacionais documentados. Finalmente, você pode relatar eventos anormais para o suporte ao cliente do fornecedor, indicando a urgência do evento.

Conclusão

Uma migração para a nuvem é um processo de longo prazo. Comece aos poucos. Você perceberá que tem muito a aprender ao longo da jornada. Mantenha um diário e registre honestamente todos os enganos e problemas encontrados. Não se envergonhe se práticas sofríveis ou se mesmo erros forem detectados em sistemas herdados durante a migração. Isso ocorrerá em praticamente todas as organizações. A documentação dos problemas é a melhor ação que você pode tomar para a empresa.

Uma ou mais das primeiras migrações serão bem-sucedidas e você estará pronto para uma migração importante para a nuvem. Essa migração pode oferecer benefícios de custo, flexibilidade e segurança. Não menos importante, a migração para a nuvem oferecerá um ambiente de computação atualizado que ajudará a atrair para a empresa colaboradores qualificados que queiram trabalhar com tecnologias de vanguarda.

Conclusão

Por muitos anos após a disponibilização da primeira oferta importante de nuvem pela Amazon.com, a imprensa especializada discutiu a pergunta debatida por administradores de sistema e DBAs: "Migrar para a nuvem ou não migrar?". Logo, ofertas de nuvem on-premise e híbrida se juntaram a soluções em nuvem puras como opções a considerar. Mas as escolhas sempre eram mais complicadas. Como demonstramos neste relatório, as ofertas se multiplicaram rapidamente. Os DBAs devem examinar bancos de dados considerando simultaneamente todos estes fatores:

- fornecedor terceirizado, on-premise ou híbrido
- relacional ou uma das diversas variedades não relacionais
- gerenciado ou autogerenciado
- nativo da nuvem (por exemplo, Amazon Aurora) ou em várias plataformas (por exemplo, MySQL)
- utilizar ou não aprimoramentos de performance como SSDs ou caches
- localização física de regiões e zonas de disponibilidade da nuvem
- facilidade de migração
- habilidades relevantes necessárias e disponíveis na equipe
- outros aspectos de suporte e reputação do fornecedor

Não é uma boa ideia se precipitar e tomar uma decisão com base em uma única área sem examinar todas as opções. É bastante provável que você economize muito dinheiro e aprimore a experiência dos clientes fazendo algum treinamento extra ou adotando uma tecnologia ainda pouco conhecida.

Além de definir os critérios básicos para a escolha de bancos de dados, este relatório procurou ajudar você a preparar a migração para a nuvem informando as prováveis alterações em suas responsabilidades e tarefas. A migração para a nuvem remove ou simplifica algumas responsabilidades e tarefas, mas você terá de aprender novas tecnologias e começar a pensar de formas diferentes sobre metas como alta disponibilidade e otimização.

Você aprenderá muito durante a primeira migração ou iniciando um novo projeto na nuvem. Cada projeto esclarecerá o cenário dos bancos de dados na nuvem e oferecerá ideias para o próximo projeto. Esperamos que este relatório seja um alerta para o que você pode esperar durante a jornada.

Sobre os autores

Wendy A. Neu é consultora-chefe do AWS Professional Services e trabalha com os problemas mais difíceis dos clientes, criando sistemas de alta qualidade, escaláveis e com arquitetura correta. Ela contribui regularmente com o blog de banco de dados da AWS e tem certificações da AWS, da Oracle e do Microsoft SQL Server. Antes de ingressar na Amazon, ela trabalhou como consultora em Cincinnati, em Ohio, ajudando clientes a converter necessidades empresariais em soluções de tecnologia práticas.

Vlad Vlasceanu é um arquiteto-chefe de soluções especializado em bancos de dados da AWS e trabalha no escritório de Santa Monica, na Califórnia. Vlad ajuda os clientes a adotar soluções de banco de dados nativas da nuvem, como o Amazon Aurora, e implantar arquiteturas de banco de dados de alta performance em grande escala na AWS. Seu foco está no projeto e implantação de workloads sustentáveis, econômicas e escaláveis que aproveitam as melhores práticas e recursos mais recentes oferecidos pelos produtos e serviços da AWS. Antes de ingressar na AWS, a carreira de Vlad incluiu mais de 15 anos de projeto e desenvolvimento de aplicações Web voltadas a consumidores e aplicações orientadas por dados para o setor de energia. Vlad é mestre em sistemas de informação da Baylor University.

Andy Oram publicou a série sobre Linux da O'Reilly, o livro revolucionário *Peer-to-Peer* e o sucesso de vendas *Beautiful Code*. Andy é autor de vários relatórios sobre tópicos técnicos como data lakes, performance da Web e software de código aberto. Seus artigos foram publicados em veículos como *The Economist*, *Communications of the ACM*, *Copyright World*, *Journal of Information Technology and Politics*, *Vanguardia Dossier* e *Internet Law and Business*. Ele foi palestrante em conferências como O'Reilly's Open Source Convention, FISL (Brasil), FOSDEM, DebConf e LibrePlanet. Andy participa da organização da USTPC, a organização de políticas da Association for Computing Machinery. Ele também escreve para vários sites sobre TI na área de saúde e problemas de computação e políticas.

Sam R. Alapati é administrador de dados na Solera Holdings em Westlake, no Texas. Ele faz parte da equipe de Big Data e Hadoop. Sam é um ACE da Oracle, um reconhecimento concedido pela Oracle Technology Network. Ele é autor de *Modern Linux Administration* (O'Reilly, 2018) e de mais de vinte livros sobre bancos de dados e administração de sistemas. Sam tem experiência de trabalho com todos os três principais provedores de nuvem: AWS, Microsoft Azure e Google Cloud Platform.