



KLS

Computação em Nuvem

Computação em Nuvem

Neumar Costa Malheiros

© 2019 por Editora e Distribuidora Educacional S.A.

Todos os direitos reservados. Nenhuma parte desta publicação poderá ser reproduzida ou transmitida de qualquer modo ou por qualquer outro meio, eletrônico ou mecânico, incluindo fotocópia, gravação ou qualquer outro tipo de sistema de armazenamento e transmissão de informação, sem prévia autorização, por escrito, da Editora e Distribuidora Educacional S.A.

Presidência

Rodrigo Galindo

Vice-Presidência de Produto, Gestão e Expansão

Julia Gonçalves

Vice-Presidência Acadêmica

Marcos Lemos

Diretoria de Produção e Responsabilidade Social

Camilla Veiga

Gerência Editorial

Fernanda Migliorança

Editoração Gráfica e Eletrônica

Renata Galdino

Supervisão da Disciplina

Marcilyanne Moreira Gois

Revisão Técnica

Marcio Aparecido Artero

Marcilyanne Moreira Gois

Dados Internacionais de Catalogação na Publicação (CIP)

Malheiros, Neumar Costa

M249c Computação em nuvem / Neumar Costa Malheiros. –
Londrina : Editora e Distribuidora Educacional S.A., 2019.
200 p.

ISBN 978-85-522-1590-5

1. Computação em nuvem. 2. Modelo de serviços em
nuvem. 3. Nuvem privada. 4. Nuvem pública. 5. PaaS.
6. SaaS. 7. IaaS. I. Malheiros, Neumar Costa. II. Título.

CDD 004

Thamiris Mantovani CRB-8/9491

2019

Editora e Distribuidora Educacional S.A.

Avenida Paris, 675 – Parque Residencial João Piza

CEP: 86041-100 — Londrina — PR

e-mail: editora.educacional@kroton.com.br

Homepage: <http://www.kroton.com.br/>

Sumário

Unidade 1

Fundamentos de computação em nuvem	7
--	---

Seção 1

Conceitos básicos em computação em nuvem	9
--	---

Seção 2

Modelos de serviço em computação em nuvem	21
---	----

Seção 3

Modelos de implantação em computação em nuvem	35
---	----

Unidade 2

Tecnologias e soluções de computação em nuvem	51
---	----

Seção 1

Tecnologias de suporte à nuvem.....	53
-------------------------------------	----

Seção 2

Provedores de computação em nuvem.....	68
--	----

Seção 3

Migração de aplicações para a nuvem	82
---	----

Unidade 3

Ofertas de serviço em Computação em Nuvem.....	99
--	----

Seção 1

Serviços de processamento de dados	101
--	-----

Seção 2

Serviços de armazenamento e análise de dados	115
--	-----

Seção 3

Soluções em nuvem.....	129
------------------------	-----

Unidade 4

Arquitetura de aplicações em nuvem	151
--	-----

Seção 1

Modelos de arquitetura em nuvem	153
---------------------------------------	-----

Seção 2

Qualidade de serviço em nuvem	170
-------------------------------------	-----

Seção 3

Segurança e privacidade em nuvem.....	183
---------------------------------------	-----

Palavras do autor

Atualmente, no mercado de Tecnologia da Informação (TI), o modelo de computação em nuvem é muito utilizado para a disponibilização de soluções de software escaláveis na internet. Portanto, é imprescindível conhecer as características e as tecnologias envolvidas. O conceito de nuvem é uma metáfora para um cenário no qual as aplicações podem ser acessadas remotamente, de qualquer lugar e dispositivo com conexão à Internet. O paradigma de computação em nuvem revolucionou a maneira como as organizações alocam e gerenciam recursos de Tecnologia da Informação (TI). Nesse novo modelo, recursos computacionais, como capacidade de processamento e armazenamento de dados, disponíveis em provedores podem ser alocados sob demanda, com pagamento de acordo com a quantidade e o tempo de uso. A possibilidade de redução de custos e flexibilidade no uso dos recursos de TI motivaram uma rápida migração de servidores e aplicações para provedores de computação em nuvem. Nesse contexto, é essencial o entendimento sobre como os serviços de computação em nuvem são provisionados, suas características e benefícios.

Este livro vai além dos conceitos básicos e discute as mudanças de paradigma introduzidas com o conceito de computação em nuvem. Antes, era muito difícil garantir a escalabilidade das aplicações, mas com a computação em nuvem, a capacidade computacional pode ser dinamicamente ampliada para atender um eventual crescimento de demanda. Serão discutidas, também, as tecnologias que viabilizam a implementação de soluções nesta nova forma de alocação de recursos computacionais. Ao concluir os estudos, você será capaz de analisar comparativamente os diferentes modelos de serviço em ambientes de computação nuvem. Além disso, vai adquirir as habilidades necessárias para especificar a arquitetura de uma aplicação em ambiente de computação em nuvem e entender como a arquitetura influencia no desempenho, confiabilidade e segurança da aplicação. O conteúdo aqui apresentado permitirá o contato com diversos tipos de serviços oferecidos por importantes provedores de computação em nuvem no mercado de TI.

Este material aborda diversos aspectos do paradigma de computação em nuvem. Na Unidade 1, vamos discutir os conceitos fundamentais, os modelos de serviço e os modelos de implantação de infraestruturas de Computação em Nuvem. Na Unidade 2, serão apresentados os principais avanços tecnológicos que viabilizaram as soluções em nuvem, os fatores envolvidos na migração de dados e aplicações para a nuvem e os modelos de tarifação usados pelos provedores. Em seguida, na Unidade 3, vamos caracterizar

e apresentar exemplos para os principais tipos de serviços de computação em nuvem oferecidos no mercado, entre os quais pode-se destacar: servidores virtuais e serviços de armazenamento e análise de dados. Por fim, na Unidade 4, o foco será a arquitetura das aplicações em nuvem. Nessa parte, serão discutidos diversos modelos de arquitetura, fatores que impactam no desempenho e na confiabilidade das aplicações e, também, as importantes questões relacionadas com a segurança de aplicações em Nuvem.

O assunto abordado aqui é de extrema relevância para sua atuação profissional. É preciso foco, disciplina e determinação para aproveitar ao máximo as atividades de aprendizagem.

Bons estudos!

Unidade 1

Neumar Costa Malheiros

Fundamentos de computação em nuvem

Convite ao estudo

Caro estudante, você utiliza alguma aplicação em nuvem? Um exemplo típico é o uso de serviços em nuvem para sincronização de dados entre diversos dispositivos, mas existem muitas outras aplicações. Provedores de computação em nuvem oferecem diversos serviços, como aplicações de gestão corporativa e servidores virtuais para processamento e armazenamento de dados. Para muitas organizações, é mais interessante alocar recursos computacionais de algum provedor em nuvem, do que implantar uma infraestrutura de TI nas suas próprias instalações, pois assim, evitam-se custos com aquisição, operação e manutenção de equipamentos de alto desempenho. Entretanto, essa estratégia pode envolver algumas desvantagens. Nesta seção, vamos caracterizar o que são serviços de computação em nuvem e discutir suas vantagens e desvantagens.

Considere a demanda da empresa que fabrica peças para aeronaves. Ela comprou um novo software de simulações a fim de avaliar o projeto inicial das peças que serão fabricadas. Então, a empresa necessita de servidores de alto desempenho para realizar simulações. A ideia inicial da empresa era instalar sua própria infraestrutura de TI. Porém, o plano de investimento inicial apresentou custos muitos elevados. Sendo assim, você foi acionado pela empresa para avaliar a viabilidade do uso de um ambiente em nuvem. Seu primeiro passo será elaborar um parecer técnico sobre as vantagens e possíveis dificuldades que a fabricante teria ao adotar soluções de computação em nuvem para alocação e operação dos recursos de TI. Dessa forma, você propõe realizar uma análise comparativa das vantagens e desvantagens do uso de computação em nuvem para esse caso em particular. Assim, surgem os seguintes questionamentos: é necessário sigilo sobre os dados da empresa? O tempo e a taxa de utilização dos servidores necessários para execução das simulações? Qual a natureza de softwares de simulação de resistência e aerodinâmica das peças?

Para lhe auxiliar na resolução desse problema, vamos discutir alguns aspectos fundamentais sobre o paradigma de Computação em Nuvem a fim de ajudá-lo a construir um parecer com um sólido embasamento técnico.

Não se esqueça de que você deverá elaborar um parecer técnico com uma fundamentação sólida que justifique a escolha da abordagem mais adequada.

Bons estudos!

Conceitos básicos em computação em nuvem

Diálogo aberto

Caro estudante, você utiliza alguma aplicação em nuvem? Um exemplo típico é o uso de serviços em nuvem para sincronização de dados entre diversos dispositivos, mas existem muitas outras aplicações. Provedores de computação em nuvem oferecem diversos serviços, como aplicações de gestão corporativa e servidores virtuais para processamento e armazenamento de dados. Para muitas organizações, é mais interessante alocar recursos computacionais de algum provedor em nuvem, do que implantar uma infraestrutura de TI nas suas próprias instalações, pois assim, evitam-se custos com aquisição, operação e manutenção de equipamentos de alto desempenho. Entretanto, essa estratégia pode envolver algumas desvantagens. Nesta seção, vamos caracterizar o que são serviços de computação em nuvem e discutir suas vantagens e desvantagens.

Considere a demanda da empresa que fabrica peças para aeronaves. Ela comprou um novo *software* de simulações a fim de avaliar o projeto inicial das peças que serão fabricadas. Então, a empresa necessita de servidores de alto desempenho para realizar simulações. A ideia inicial da empresa era instalar sua própria infraestrutura de TI. Porém, o plano de investimento inicial apresentou custos muitos elevados. Sendo assim, você foi acionado pela empresa para avaliar a viabilidade do uso de um ambiente em nuvem. Seu primeiro passo será elaborar um parecer técnico sobre as vantagens e possíveis dificuldades que a fabricante teria ao adotar soluções de computação em nuvem para alocação e operação dos recursos de TI. Dessa forma, você propõe realizar uma análise comparativa das vantagens e desvantagens do uso de computação em nuvem para esse caso em particular. Assim, surgem os seguintes questionamentos: é necessário sigilo sobre os dados da empresa? O tempo e a taxa de utilização dos servidores necessários para execução das simulações? Qual a natureza de softwares de simulação de resistência e aerodinâmica das peças?

Para lhe auxiliar na resolução desse problema, vamos discutir alguns aspectos fundamentais sobre o paradigma de Computação em Nuvem a fim de ajudá-lo a construir um parecer com um sólido embasamento técnico. Não se esqueça de que você deverá elaborar um parecer técnico com uma fundamentação sólida que justifique a escolha da abordagem mais adequada.

Bons estudos!

Não pode faltar

A evolução das tecnologias de rede permitiu a criação de aplicações inovadoras que representam soluções *disruptivas*. Um exemplo claro são os novos negócios na área de transporte coletivo. Muitas pessoas que moram em regiões metropolitanas têm optado por abandonar o carro próprio em favor do uso de transporte coletivo, principalmente, por meio de aplicativos de transporte. Os usuários desses aplicativos podem se beneficiar com a redução de custos e com a maior conveniência e flexibilidade para se deslocar pela cidade. Um aspecto importante nesta mudança de modelo de negócio é que o usuário deixa de comprar um produto (o carro próprio) para usar um serviço (cobrado pelo deslocamento contratado). Você já parou para analisar que isso é uma mudança de setor da economia: do setor de bens de consumo para o setor de serviços.

E no mercado de tecnologia da informação, os clientes consomem produtos ou serviços? Tradicionalmente, recursos computacionais eram ofertados como produtos. Uma empresa comprava servidores, notebooks, ou um software, por exemplo. No entanto, há alguns anos, uma grande mudança de paradigma foi introduzida a partir do conceito de computação em nuvem. Nesse novo modelo, recursos computacionais, como capacidade de processamento ou aplicações, são oferecidos com serviços, não como produtos. Isso significa que o cliente, em vez de adquirir um recurso computacional, passa a pagar pelo seu uso. Nesse contexto, a computação em nuvem pode ser entendida como um paradigma no qual recursos de TI são ofertados como serviços. Por exemplo, em vez de comprar um servidor para hospedar uma aplicação, uma empresa pode alugar um servidor de um provedor de serviços de computação em nuvem.

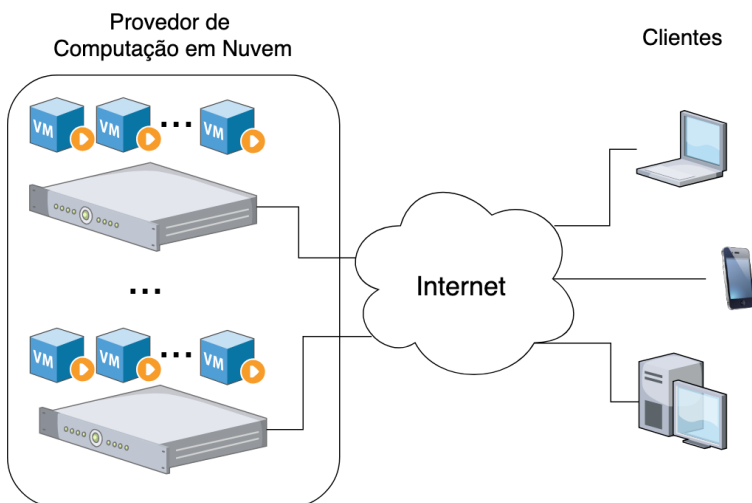
Veja a seguir uma das mais completas e difundidas definições de Computação em Nuvem:

Computação em nuvem é um modelo para possibilitar acesso remoto, de modo ubíquo, conveniente e sob demanda, a um conjunto compartilhado de recursos computacionais configuráveis (por exemplo, redes, servidores, armazenamento, aplicações e serviços) que podem ser rapidamente alocados e liberados com mínimo esforço gerencial ou interação com o provedor de serviços. (MELL; GRANCE, 2011, p. 2)

Observe que os recursos computacionais disponíveis na nuvem podem ser os mais diversos e a alocação ou liberação dos recursos deve ser dinâmica. Isso foi possível, principalmente, em decorrência dos avanços nas tecnologias de virtualização e de redes de comunicação de dados. O conceito de virtualização permite que, em um mesmo computador, sejam criadas uma ou mais máquinas virtuais. Uma máquina virtual é um software que emula o funcionamento de um computador, ou seja, ela é capaz de executar programas como um computador real emulando, inclusive os componentes de uma máquina física, como disco, processador, monitor e placa de rede.

A Figura 1.1 ilustra um provedor de serviços de computação em nuvem com diversas máquinas virtuais (VMs – *virtual machines*) instanciadas na infraestrutura de máquinas físicas. O uso de virtualização permite o compartilhamento da infraestrutura entre vários clientes e também viabiliza a alocação dinâmica. Não seria possível instanciar rapidamente um servidor físico, mas isso pode ser feito em instantes para uma máquina virtual. Os avanços nas tecnologias de rede e de serviços Web resultaram num modelo de amplo acesso aos serviços. Os clientes podem usar os recursos computacionais remotamente a qualquer momento.

Figura 1.1 | Máquinas Virtuais



Fonte: Elaborado pelo autor.

Sousa, Moreira e Machado (2009) apresentam uma perspectiva interessante, que mostra a computação em nuvem como uma evolução do conceito de *Utility Computing*. Um serviço de utilidade é aquele distribuído

amplamente ao público geral com pagamento baseado no uso, por exemplo, a distribuição de energia elétrica e outros serviços de utilidade pública. Assim, a *Utility Computing* é uma visão da TI como um serviço de utilidade pago de acordo com a quantidade de recursos utilizados. Nesse contexto, temos serviços de TI sob demanda com pagamento baseado no uso. Esses serviços abrangem vários níveis, desde aplicações para usuários finais, até grandes infraestruturas computacionais para empresas.

Assimile

Atente-se para o fato de que a computação em nuvem não é uma tecnologia específica, nem é uma arquitetura. Ela é, na verdade, um novo modelo de negócio no qual recursos computacionais são provisionados na forma de serviços. O cliente não compra o recurso, ele paga pelo uso do recurso sob demanda.

Em geral, os serviços em nuvem podem ser acessados e gerenciados por uma aplicação Web. No entanto, não podemos afirmar que toda aplicação disponível na Web é uma aplicação em nuvem. A seguir, você irá conhecer algumas das características essenciais de serviços de computação em nuvem, conforme explicado em (SOUSA; MOREIRA; MACHADO, 2009):

- **Self-service:** o próprio cliente gerencia a alocação dos recursos, com mínima interação com o provedor, na verdade, a alocação e uso dos serviços pode ser automatizada.
- **Ampla acesso:** os recursos podem ser acessados remotamente, de qualquer lugar, por meio de tecnologias baseadas em padrões abertos, consequentemente, têm soluções multiplataforma, ou seja, soluções que não dependem do tipo do dispositivo ou sistema operacional, por exemplo.
- **Pooling de recursos:** o provedor mantém um robusto conjunto de recursos que são compartilhados entre os clientes de forma que a localização e a manutenção dos equipamentos é transparente para os clientes.
- **Elasticidade Rápida:** os clientes podem alocar mais recursos diante de um aumento de demanda ou liberar recursos em uso no caso de diminuição da demanda; a alocação ou liberação de recursos deve ser feita rapidamente, inclusive de forma automatizada, por exemplo, por meio de um script que monitora a demanda e reage de acordo com as mudanças.

- **Serviço medido:** o uso dos recursos pelos clientes é detalhadamente contabilizado para fins de tarifação e também para monitoramento da qualidade do serviço.

Conforme Coutinho (2013), a elasticidade é uma das principais características dos serviços disponibilizados em ambientes de computação em nuvem. Essa característica cria a ilusão de que os recursos na nuvem são ilimitados, uma vez que é possível aumentar ou diminuir a quantidade de recursos sob demanda.

Exemplificando

Considere uma aplicação de comércio eletrônico na Web. O número de clientes que acessa a aplicação varia muito, por exemplo, entre o horário comercial e o período da noite. Se essa aplicação está replicada em máquinas virtuais (VMs) em um provedor de computação em nuvem, a elasticidade representa a possibilidade de alocar mais VMs durante o dia para atender a uma maior demanda de clientes. Durante a noite, a maior parte das VMs pode ser desligada devido à diminuição na carga de acessos, evitando-se subutilização dos recursos.

A elasticidade rápida depende das tecnologias de virtualização. Por exemplo, devido a um aumento brusco nas requisições a um serviço, diversas máquinas virtuais podem ser rapidamente instanciadas para atender ao aumento na carga de trabalho. Essa facilidade de alterar dinamicamente a quantidade de recursos alocados facilita o planejamento de capacidade computacional para o cliente, o que representa um dos principais benefícios que fomentaram o rápido desenvolvimento de soluções de computação em nuvem. Além disso, existe uma série de outros benefícios, entre os quais pode-se destacar a redução de custos e a abstração da complexidade (ERL, PUTTINI, MAHMOOD; 2013).

A redução de custos pode ser observada de várias formas. Uma delas é que o uso de recursos computacionais na nuvem sob demanda evita a necessidade do investimento inicial para montar uma infraestrutura de TI. A flexibilidade da alocação dinâmica de recursos permite que os investimentos em TI sejam realizados de acordo com o crescimento do negócio do cliente. Além disso, o compartilhamento de recursos entre os vários clientes de um provedor diminui os custos relacionados com a manutenção dos equipamentos e com equipe de profissionais especializados. Outro ponto importante que você não pode esquecer é que componentes de software e hardware ficam obsoletos rapidamente. Quando um cliente acessa os recursos na forma de serviço em nuvem, os custos de atualização ficam por conta do provedor.

De fato, a gestão de infraestrutura de equipamentos e profissionais de TI é complexa e tem alto custo. As soluções em nuvem permitem que um cliente tenha acesso aos recursos computacionais com pagamento de acordo com o uso, em vez de ter que implantar sua própria infraestrutura de TI.

Pesquise mais

O artigo indicado a seguir apresenta um histórico bastante interessante, mostrando a evolução de modelos de negócios em TI até o surgimento da computação em Nuvem.

TIGRE, P. B.; NORONHA, V. B. Do mainframe à nuvem: inovações, estrutura industrial e modelos de negócios nas tecnologias da informação e da comunicação. **Revista de Administração**, v. 48, n. 1, p. 114–127, 2013.

Por fim, outra importante vantagem do uso de serviços em ambientes de Nuvem é a possibilidade de abstrações que ocultam a complexidade dos recursos computacionais, uma vez que, quando utilizamos alguma aplicação em nuvem, os detalhes da infraestrutura subjacente são transparentes. Toda a complexidade de instalação, configuração e atualização dos equipamentos e softwares de sistema são responsabilidades do provedor. Até o conceito de versionamento do software perde a importância. Você já parou para verificar a versão de algum software em nuvem que você utiliza? Provavelmente não, até porque, você sempre estará usando a versão mais nova, pois a atualização é feita de forma transparente pelo provedor. Essa vantagem não é privilégio de usuários finais. Equipes de profissionais de TI também fazem uso de ambientes de desenvolvimento de software na nuvem. Esses ambientes são gerenciados pelo provedor de forma que os desenvolvedores possam focar nos requisitos da aplicação e fazer uso de mecanismos sofisticados de projeto, teste e implantação (*deploy*) automática de aplicações.

Refleta

Entre os benefícios da Computação em Nuvem, a redução de custos é um dos destaques. Por exemplo, em vez de investir na instalação de sua própria infraestrutura de TI, um cliente pode “alugar” servidores em um provedor. Mas, você já verificou o custo de “alugar” um servidor em algum provedor em nuvem? Será que esta opção é mesmo sempre a mais vantajosa em termos financeiros? Ou a solução em nuvem pode resultar em um custo final mais alto?

Você aprendeu até aqui as diversas vantagens que o uso de serviços de computação em nuvem pode proporcionar. No entanto, toda estratégia tem

suas desvantagens. Antes de recomendar o uso de um serviço de Computação em Nuvem, você precisará realizar uma avaliação minuciosa da relação custo-benefício. Vamos então discutir eventuais barreiras que podem dificultar a adoção de soluções em nuvem. Entre os principais desafios da computação em nuvem podemos citar:

- **Segurança e privacidade:** uma vez que o acesso aos serviços em nuvem é remoto, existe uma maior vulnerabilidade devido à exposição dos dados dos clientes na Internet. Outro fator importante neste contexto é o compartilhamento de recursos do provedor por diversos clientes. Isso exige mecanismos sofisticados de isolamento e proteção aos dados, para que um cliente não consiga ter acesso aos dados de outro cliente sem a devida autorização.
- **Disponibilidade:** se não houver uma infraestrutura de comunicação de alto desempenho para acesso aos dados e serviços, o desempenho das aplicações em nuvem fica comprometido. A confiabilidade da rede também é importante, pois, sem conexão, o cliente não consegue acessar seus dados no provedor.
- **Portabilidade:** a maior parte das aplicações que utilizamos possui diversas dependências de componentes de software (bibliotecas, frameworks) ou plataformas. Quando ocorre a necessidade de migrar uma aplicação de um provedor para outro pode haver problemas de portabilidade do software ou bases de dados, se o novo provedor não oferecer suporte adequado.
- **Questões jurídicas:** pode haver complicações em relação ao licenciamento de software em função do compartilhamento de recursos do provedor entre vários clientes, pois a licença do provedor tem que permitir a utilização do software por terceiros. Além disso, existem fatores relacionados com a jurisdição considerada em um eventual litígio. A jurisdição depende da localização física dos dados do cliente. Os grandes provedores possuem instalações em diversos países, então o cliente deve estar atento para alocar seus recursos em um país cuja legislação não contrapõe ao seu negócio.

Exemplificando

Um problema típico em provedores de computação em nuvem é o chamado aprisionamento (*vendor lock-in*), ou seja, o cliente fica “trancado” no provedor. Esse termo é usado para descrever o cenário no qual se cria uma dependência entre uma aplicação e os recursos de um provedor específico, dificultando a migração de provedor. Nesse

caso, é muito difícil para um cliente mudar de provedor, em geral, porque o cliente está usando uma tecnologia proprietária do provedor que não é oferecida nos concorrentes. Se houver um reajuste de preços no provedor no qual está o cliente, pode ser mais caro ainda para ele migrar seus serviços para outro provedor, pois os custos envolvidos na migração de bases de dados e na portabilidade de aplicações para outras tecnologias, em geral, são muito altos.

Vimos que há algumas preocupações relevantes que precisam ser consideradas quando da migração de uma aplicação para uma nuvem ou da contratação de um serviço. Por outro lado, esses desafios podem ser vistos como oportunidades (ARMBRUST, 2010). O mercado de computação em nuvem continua crescendo e novas tecnologias e modelos têm sido utilizados para atenuar essas barreiras. Depois da apresentação dos conceitos básicos nesta seção, nas próximas vamos aprofundar nosso estudo e descrever os principais modelos de serviço em nuvem, assim como as abordagens para implantação da infraestrutura dos provedores.

Sem medo de errar

A empresa de fabricação de peças de aeronaves solicitou um parecer sobre os benefícios e desvantagens que ela poderia ter ao adotar o uso de serviços de computação em nuvem para sua demanda de recursos computacionais. Essa demanda se refere à necessidade de realizar simulações para avaliar o projeto das peças antes de iniciar a produção em escala. Simulações dessa natureza exigem alta capacidade de processamento de dados, então a empresa pode adquirir diversos servidores e montar sua própria infraestrutura ou alocar servidores de um provedor de computação em nuvem. Precisamos fazer um comparativo entre as vantagens e desvantagens da opção pela nuvem. Para isso, é importante refletir sobre alguns aspectos como: é necessário sigilo sobre os dados da empresa? Qual é o tempo e a taxa de utilização dos servidores para execução das simulações? Qual a natureza de softwares de simulação de resistência e aerodinâmica das peças?

Durante esta seção, discutimos alguns benefícios da computação em nuvem: 1) facilidade de planejamento de capacidade; 2) redução de custos; e 3) abstração de complexidade. Vamos analisar cada um deles para identificar quais se aplicam a este caso.

- Se a empresa optar por adquirir os servidores, ela corre o risco de comprar um número de servidores maior do que o necessário e perder com o desperdício de recursos, ou de comprar um número menor do que o necessário e atrasar os resultados das simulações. De

fato, a alocação dinâmica de recursos virtualizados em um ambiente de nuvem possibilita que a empresa possa ajustar o número inicial de servidores virtuais alocados de acordo com a demanda computacional das simulações. Se houver alteração na demanda, o número de servidores pode ser rapidamente reajustado de acordo com a demanda e o capital disponível. Portanto, é claro que a empresa pode se beneficiar das facilidades de planejamento de capacidade.

- Mesmo que a capacidade esteja de acordo com a demanda, ainda existe subutilização dos recursos caso a empresa adquira seus próprios servidores. Isso se explica pelo fato de que, quando a etapa de simulação se encerrar para início da fabricação das peças em escala, os servidores ficarão parados até nova etapa de simulações. Além disso, há uma depreciação inevitável dos equipamentos com o passar do tempo. Ao optar pelo modelo de computação em nuvem, a empresa elimina esse problema, pois os servidores inicialmente alocados para as simulações podem ser liberados assim que os resultados forem obtidos. A empresa pagará somente pelo tempo efetivo de alocação dos servidores. Quando uma nova etapa de simulações for necessária, a empresa pode novamente alocar os servidores de acordo com a demanda. Além disso, a empresa não terá custos associados com manutenção de infraestrutura, nem equipe de profissionais especializados para isso. Também não é necessário, neste caso, um capital inicial suficiente para adquirir todos os servidores necessários. Portanto, podemos concluir que o espaço para redução de custos é significativo.
- Se optar por adquirir seus próprios servidores, a empresa terá o trabalho de instalar e configurar todos os equipamentos, incluindo nobreaks, cabeamento e software de sistema. Por outro lado, se escolher o ambiente de nuvem, a empresa simplesmente vai escolher a plataforma desejada sem se preocupar com detalhes e complexidades da infraestrutura subjacente. Dessa forma, a empresa se beneficia também da possibilidade de seus profissionais focarem na execução das simulações, em vez de se preocuparem com detalhes dos recursos computacionais.

Sabe-se que as principais barreiras para adoção de Computação em nuvem são: segurança e privacidade; desempenho de rede; portabilidade e questões jurídicas. Vamos analisar cada uma delas separadamente:

- Segurança e privacidade são fatores críticos nesse caso, devido a questões de sigilo industrial. O valor de propriedade intelectual dos projetos é importante para qualquer empresa. Assim, vulnerabilidades de segurança no provedor representariam um sério problema para a empresa. O uso de mecanismos de segurança adequados são, portanto, imprescindíveis neste cenário.

- Desempenho da rede não é um fator crítico neste caso, pois não se trata de uma aplicação interativa e também não são grandes os volumes de transferência de dados, uma vez que simulações dessa natureza são intensivas em CPU (devidos aos cálculos complexos).
- A princípio, não há com o que se preocupar em relação à portabilidade, pois a empresa precisa apenas de servidores com o sistema operacional exigido para executar o software de simulação.
- Também não há problemas aparentes relacionados a questões jurídicas, já que a empresa adquiriu o software que será utilizado.

A partir dessa análise, pode-se concluir de forma favorável pelo uso de ambiente de computação em nuvem, em função dos benefícios em relação a custos, facilidade de gestão e dimensionamento dos recursos, com apenas uma ressalva em relação à necessidade de preocupação com o uso de ferramentas de segurança que possam garantir o sigilo dos dados no provedor de computação em nuvem. Essa análise representa um importante passo em relação ao objetivo de sermos capazes de avaliar de forma comparativa os diversos serviços em nuvem, enfatizando suas características, benefícios e desvantagens.

Avançando na prática

Desafios para sistema de prontuário eletrônico na Nuvem

Considere uma rede pública de hospitais que faz uso de um sistema de prontuário eletrônico para atender os pacientes de forma mais ágil e eficiente. A rede conta com profissionais de TI e infraestrutura própria na qual está implantado o sistema há alguns anos. O sistema inclui uma base com grande volume de dados sobre o histórico médico dos pacientes.

A fim de reduzir os gastos com os recursos de TI, a rede hospitalar decidiu migrar o sistema para um provedor de computação em nuvem. No entanto, como analista de TI da rede hospitalar, você sugere a necessidade de especificar um plano de contingência para o caso de ocorrer algum problema, pois os hospitais não podem interromper o atendimento aos pacientes. Dessa forma, um dos diretores do hospital lhe faz o seguinte questionamento: quais são os riscos envolvidos na migração para o ambiente de nuvem?

Resolução da situação-problema

Existem diversos riscos, entre os quais podemos destacar:

- **Risco de violação de privacidade dos dados médicos dos pacientes.** Como os dados serão transmitidos pela rede dos dispositivos de acesso para o provedor, existe a possibilidade de que eles sejam interceptados por agentes maliciosos, o que seria um problema crítico. Outra possibilidade ligada à segurança é o eventual compartilhamento de recursos no provedor, por exemplo, se as máquinas virtuais alocadas para a rede hospitalar estiverem no mesmo servidor físico no qual estejam também máquinas virtuais de outros clientes. Caso não haja isolamento e proteção dos dados de forma adequada, os dados médicos podem ser violados.
- **Risco de os dados serem armazenados fora de região permitida.** Em geral, os provedores de computação em nuvem possuem vários centros de dados em diferentes regiões, inclusive em países diferentes. Como se trata de uma rede pública e de dados médicos, pode haver restrições legais sobre os dados serem armazenados em outras regiões. Por exemplo, existem dados de órgãos públicos federais que não podem ser armazenados em outros países. Mesmo que a rede hospitalar escolha um centro de dados em uma região permitida, é preciso estar atento para que réplicas dos dados não sejam criadas em outras regiões.
- **Risco de indisponibilidade do serviço.** Se houver falha no provedor ou nos enlaces de comunicação, o sistema não poderá ser acessado remotamente. O desempenho da rede e a confiabilidade do provedor são críticos para operação do sistema.

Faça valer a pena

1. Em relação às características essenciais de serviços de computação em nuvem, avalie as afirmativas a seguir.

- I. A contabilização do uso dos recursos é imprescindível para viabilizar a tarifação dos serviços em nuvem.
- II. O amplo acesso aos serviços em nuvem exige o uso de software livre gratuito.
- III. Mecanismos de virtualização são uma forma eficiente de permitir a elasticidade dos serviços em nuvem.

Considerando o contexto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. I e II, apenas.
- c. I e III, apenas.

- d. II e III, apenas.
- e. I, II e III

2. Entre as características essenciais de serviços de computação em nuvem, pode-se destacar a elasticidade rápida. Em relação a essa característica, avalie as afirmativas a seguir:

- I. A elasticidade pode ser entendida como a habilidade dos clientes de aumentarem ou diminuir sua capacidade computacional na nuvem de forma rápida e até mesmo automática.
- II. A elasticidade é fundamental para viabilizar a redução de custos e o planejamento da capacidade.
- III. A elasticidade faz com que os recursos do provedor sejam infinitos.

Considerando o contexto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. I e II, apenas.
- c. I e III, apenas.
- d. II e III, apenas.
- e. I, II e III.

3. Vamos supor que você vai iniciar uma jornada empreendedora. Junto com alguns sócios, decidiu criar um negócio a partir da ideia de uma aplicação Web inovadora. Considere que recursos financeiros representam um fator crítico, pois vocês ainda não têm um volume alto de capital para desenvolver um Mínimo Produto Viável (MVP) para divulgar o negócio.

Considerando os possíveis benefícios e riscos da opção por um ambiente de nuvem para implantação da aplicação web que será desenvolvida, assinale a alternativa correta.

- a. Haverá redução de custos pois não será preciso pagar licenças de eventuais softwares necessários para o desenvolvimento ou hospedagem da aplicação web.
- b. Não há risco de problemas de portabilidade, pois trata-se de aplicação Web.
- c. Não há necessidade de se preocupar com mecanismos de segurança, pois isso é responsabilidade do provedor de computação em nuvem.
- d. Como o serviço, neste caso, é apenas de hospedagem da aplicação Web, não haverá o benefício de abstração de complexidade da infraestrutura subjacente.
- e. A escolha de um SGBD proprietário do provedor para armazenar dados da aplicação Web cria um risco de dificuldade de migrar para outro provedor.

Modelos de serviço em computação em nuvem

Diálogo aberto

Nesta seção, vamos descrever os principais modelos de serviço no contexto de computação em nuvem: Infraestrutura como Serviço (IaaS), Plataforma como Serviço (PaaS) e Software como Serviço (SaaS). O modelo de serviço define o nível de controle que o cliente tem sobre os recursos alocados no provedor. No modelo IaaS, o cliente pode alocar capacidade computacional na forma de recursos computacionais virtualizados. Assim, o cliente pode alocar, no provedor, uma infraestrutura, por exemplo, com a capacidade desejada de processamento e armazenamento de dados. A alocação desses recursos possibilita que o cliente tenha controle da escolha do sistema operacional que será utilizado nos servidores alocados. No modelo PaaS, o cliente não tem controle sobre a infraestrutura. Ele recebe do provedor um ambiente já configurado, pronto para o desenvolvimento de aplicações. O cliente escolhe apenas a plataforma de desenvolvimento, tais como, linguagem de programação e banco de dados. No modelo SaaS, o cliente acessa aplicações e não tem nenhum controle sobre os recursos computacionais disponíveis no provedor. Um dos aspectos importantes para escolher o modelo de serviço é entender a relação entre nível de abstração e de controle, conforme os requisitos do cliente. Uma das principais características da computação em nuvem é a abstração da complexidade dos recursos computacionais alocados. O nível de abstração da complexidade está relacionado ao nível de controle do cliente, por exemplo, ao escolher um modelo de serviço com menor controle do cliente sobre a infraestrutura subjacente, maior é o nível de abstração da complexidade.

Vamos voltar à demanda da fabricante de peças de aeronaves, pela consultoria na qual já foi feita uma análise das vantagens e desvantagens do uso de computação em nuvem para gerenciamento dos recursos computacionais necessários. Considerando uma opção pelo ambiente de computação em nuvem, nesta etapa da consultoria, você precisa elaborar um documento no qual argumenta sobre qual o modelo de serviço de computação em nuvem seria mais adequado para o ambiente de alto desempenho de que a empresa necessita. Vale lembrar que esse ambiente será usado para as simulações de aerodinâmica dos projetos das peças, que exigem grande capacidade de processamento, mas não de armazenamento. Além disso, o software de simulação utilizado, fornecido por terceiros, exige privilégios administrativos para customização e configuração das simulações. Considerando o cenário apresentado, quais são as vantagens e desvantagens de cada modelo de serviço de computação em nuvem?

Para alcançar esse objetivo, nesta seção, você vai aprender as características dos principais modelos de serviços de computação em nuvem. Vamos discutir, também, alguns modelos de serviços especializados importantes no mercado. Assim, com sua capacidade de análise crítica, você poderá realizar um estudo comparativo entre os modelos e determinar a melhor solução.

Não pode faltar

Você já observou a diversidade de serviços e aplicações em nuvem que estão disponíveis para acesso via Internet? Os provedores de computação em nuvem são as empresas responsáveis por manter a infraestrutura necessária para processar e armazenar dados. Os provedores oferecem vários tipos de serviço, que, em geral, são organizados em modelos de acordo com o nível de controle oferecido aos clientes em relação à gestão e configuração dos recursos. Por exemplo, um provedor pode provisionar servidores nos quais um cliente pode instalar um banco de dados ou um servidor web. Outros provedores oferecem aplicações para usuários finais.

A taxonomia mais amplamente aceita, como apresentada em (MELL, 2011), define três modelos de serviço:

- Infraestrutura como Serviço (IaaS – *Infrastructure as a Service*)
- Plataforma como Serviço (PaaS – *Platform as a Service*).
- Software como Serviço (SaaS – *Software as a Service*).

No modelo IaaS, os clientes podem alocar dinamicamente recursos computacionais como capacidade de processamento, armazenamento ou rede (ERL; PUTTINI; MAHMOOD, 2013). Para tanto, os provedores fazem uso de técnicas sofisticadas de virtualização. Em geral, os provedores mantêm centros de dados com um grande número de servidores interligados por redes de alto desempenho. Em cada servidor (máquina física), podem ser criadas várias máquinas virtuais. O modelo IaaS permite que o cliente possa alocar essas máquinas virtuais a fim de montar uma infraestrutura completa de TI de acordo com suas necessidades de processamento e armazenamento de dados. O cliente pode determinar o sistema operacional das máquinas virtuais, sobre as quais ele tem privilégios administrativos para configurar o ambiente como quiser. Além disso, os clientes podem configurar serviços de rede (como firewall e DNS). Dessa forma, o cliente tem um maior controle sobre os recursos, em comparação com os demais modelos de serviço.

A gestão dos recursos virtualizados pelo cliente pode ser feita por uma aplicação Web ou por meio de uma API (*Application Programming Interface*) (SOUZA; MOREIRA; MACHADO, 2009). Utilizando a API fornecida pelo provedor, os clientes podem desenvolver programas para automatizar a

alocação e a customização dos recursos disponíveis. Assim, os clientes se beneficiam de flexibilidade e da agilidade para ajustar a quantidade e a configuração dos recursos de acordo com suas necessidades. Nesse contexto, o modelo IaaS viabiliza o fornecimento de recursos de infraestrutura para o cliente montar um ambiente de TI virtualizado sob demanda.

Exemplificando

O caso típico de um serviço no modelo IaaS é a alocação de máquinas virtuais em um provedor. Nesse caso, um dos principais exemplos no mercado é o serviço EC2 (*Elastic Compute Cloud*) do provedor Amazon Web Services (AWS) (AMAZON, 2019). O cliente pode criar instâncias de máquinas virtuais com diferentes configurações, por exemplo, em termos do número de núcleos de processamento e da quantidade de memória. O serviço é tarifado de acordo com o tempo pelo qual cada instância permanece em execução.

No modelo PaaS, o cliente se beneficia do provisionamento dinâmico de um ambiente completo para desenvolvimento, teste e implantação de aplicações em nuvem (SOUZA; MOREIRA; MACHADO, 2009). O provisionamento dinâmico significa que o provedor aloca os recursos computacionais necessários de forma automática. Nesse caso, o cliente não precisa se preocupar com criação e gerenciamento de máquinas virtuais. O ambiente provisionado já é uma plataforma completa para a implementação de aplicações, com sistema operacional, servidores de aplicação, sistemas de gerenciamento de banco de dados (SGBD), compiladores, entre outros. Além disso, a plataforma pode oferecer também ferramentas de colaboração e gerenciamento de projetos.

Os grandes provedores no mercado oferecem plataformas customizadas para as principais tecnologias usadas no desenvolvimento de aplicações, como Java, NodeJS e Python. A customização da plataforma significa que o provedor é o responsável pela alocação da infraestrutura subjacente, a instalação das tecnologias e as ferramentas necessárias para cada ambiente de desenvolvimento oferecido. Em outras palavras, o cliente não tem controle sobre a infraestrutura subjacente. Assim, esse modelo de serviço tem um nível de abstração maior que o modelo IaaS, ou seja, os detalhes da infraestrutura são transparentes para o cliente. No entanto, a infraestrutura ainda é necessária, pois a plataforma é configurada em instâncias virtuais, mas de forma automática e dinâmica pelo provedor. Assim, o cliente pode se concentrar no desenvolvimento de aplicações sem ter que se preocupar com o gerenciamento de servidores. Por outro lado, o cliente corre o risco de problemas para migrar suas aplicações, caso ele faça uso de plataformas com tecnologias proprietárias. Por exemplo, se a aplicação desenvolvida pelo

cliente faz uso de um SGBD proprietário do provedor, ele corre o risco de não encontrar um SGBD compatível no provedor destino para o qual ele deseja migrar. Nesse caso, o cliente teria o trabalho de portar sua base de dados para outro SGBD, o que pode ser uma tarefa complexa e lenta.

Exemplificando

Dois exemplos importantes de plataforma como serviço são: o *Azure App Service* da Microsoft Azure (MICROSOFT, 2019) e o *App Engine* da Google Cloud (GOOGLE, 2019). Ambos os serviços oferecem um ambiente completamente gerenciado pelo provedor, com escalonamento automático para garantir o desempenho das aplicações e suporte a diversas linguagens, frameworks e mecanismos de autenticação e bancos de dados. Em particular, o *App Engine* tem suporte completo para as seguintes linguagens: Node.js, Java, Ruby, C#, Go, Python e PHP. Basta o cliente submeter o código, testar a aplicação e, ao executar, a aplicação já fica disponível na Internet.

O modelo com maior nível de abstração é o SaaS, que consiste em sistemas de software com propósitos específicos, que estão disponíveis para usuários finais por meio de acesso remoto (Internet) (SOUZA; MOREIRA; MACHADO, 2009). Em geral, o provedor oferece o software na forma de uma aplicação Web. Exemplos típicos desse modelo são aplicações office, como o Google Docs e o Microsoft Office 365, e aplicações de gestão corporativa, como o CRM (*Customer Relationship Management*) da Salesforce, que foi uma das soluções pioneiras em *software* como serviço levando a empresa a ser uma das líderes em soluções corporativas no mundo.

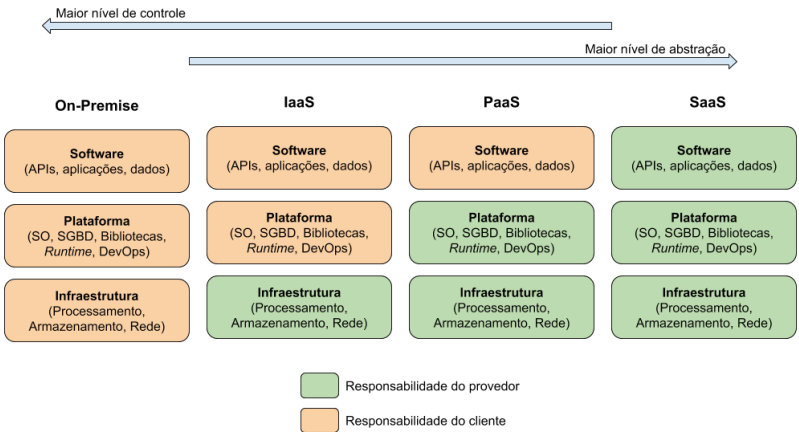
No modelo SaaS, o cliente não tem controle sobre a infraestrutura ou a plataforma. Em geral, ele configura apenas as suas preferências e customiza a aplicação. A evolução das aplicações é transparente para os clientes, que não precisam se preocupar com instalar atualizações (pois o acesso é remoto), nem com espaço de armazenamento dos seus dados. A capacidade computacional necessária é gerenciada pelo provedor de acordo com a demanda dos clientes. Por exemplo, você em algum momento se preocupa com o backup dos documentos que você já criou usando o Google Docs? Não, pois no modelo SaaS, os recursos para processamento, armazenamento de dados, backup, entre outros, são responsabilidade do provedor.

A Figura 1.2 ilustra a relação entre os modelos, além da opção do cliente de criar e manter um ambiente completo de TI em suas próprias instalações (on-premise). Como discutido em (ERL; PUTTINI; MAHMOOD, 2013), o nível de controle administrativo sobre os recursos é completo no modelo

IaaS é limitado no modelo PaaS, e diminui mais ainda no modelo SaaS. Por outro lado, com mais controle, o cliente também tem mais responsabilidade. Por exemplo, no modelo IaaS, o cliente é responsável por configurações do Sistema Operacional (SO) da máquina virtual, o que tem implicações no desempenho e na segurança do ambiente computacional.

Pode-se observar na Figura 1.2 que o cliente tem a responsabilidade de gerenciar todos os elementos do seu ambiente de TI quando ele escolhe manter sua própria infraestrutura. Ao optar por um serviço em nuvem no modelo IaaS, o cliente deixa de se preocupar com a aquisição e manutenção de equipamentos para atentar-se apenas à plataforma e aplicações. Se quiser se concentrar apenas no desenvolvimento de software, o cliente pode, ainda, optar pelo modelo PaaS. No último nível de abstração, estão os clientes de aplicações em nuvem que usam software como serviço.

Figura 1.2 | Modelos de serviço em computação em nuvem



Fonte: elaborada pelo autor.

Outro aspecto importante é a caracterização dos clientes típicos de cada modelo, o que nos permite especificar “papéis” na nuvem (SOUZA, 2009). Em geral, serviços no modelo IaaS são utilizados por clientes que precisam manter uma robusta infraestrutura de TI em ambiente de nuvem. Esse é o caso de organizações que lidam com grandes volumes de dados ou que precisam de recursos computacionais de alto desempenho e preferem realizar isso junto a um provedor em nuvem, devido aos benefícios de redução de custos e flexibilidade administrativa. Existe, também, o cenário no qual um provedor PaaS oferece seus serviços usando infraestrutura no modelo IaaS alocada de outro provedor. Como o pagamento é pelo uso, sem custos iniciais fixos,

é comum que mesmo empresas de pequeno porte e até desenvolvedores independentes usem serviços IaaS, por exemplo, uma máquina virtual para hospedar uma aplicação ainda em desenvolvimento com a confiabilidade e flexibilidade do ambiente de nuvem.

Por outro lado, os clientes de serviços no modelo PaaS são aqueles que necessitam de um ambiente de desenvolvimento gerenciado, para não se preocupar com detalhes da infraestrutura e focar na implementação de aplicações. Um exemplo são as empresas de desenvolvimento de aplicações Web e aplicativos para dispositivos móveis. Por fim, temos que os serviços no modelo SaaS são direcionados para usuários finais que apenas precisam fazer uso de softwares diversos sem qualquer interesse em lidar com as tecnologias ou equipamentos envolvidos.

Assimile

Um dos benefícios importantes da computação em nuvem é a abstração da complexidade e flexibilidade na alocação de recursos computacionais. A abstração é caracterizada pela transparência em relação aos detalhes de implantação e manutenção dos equipamentos. Mesmo no modelo IaaS, que representa o menor nível de abstração, esse benefício é aparente, pois o cliente gerencia recursos virtuais e, assim, ganha agilidade na instanciação e liberação de recursos. Além disso, o cliente não precisa se preocupar com falhas nos equipamentos, pois a reposição ou manutenção é responsabilidade do provedor. Porém, o modelo IaaS oferece o maior controle, o que pode ser importante para clientes que precisam de configurações específicas dos recursos computacionais. Clientes que desenvolvem aplicações usando tecnologias populares podem facilmente encontrar serviços PaaS já configurados, o que significa agilidade no desenvolvimento do software. A plataforma é gerenciada pelo provedor, o qual cuida, por exemplo, da configuração de mecanismos de segurança e do balanceamento de carga. De fato, esses modelos básicos de serviços representam uma relação entre o nível de controle e o nível de abstração, o que acaba por definir as responsabilidades do provedor e do cliente. Optar por mais facilidade e abstração implica abrir mão de algum nível de controle.

Bem, acabamos de compreender os modelos de serviços clássicos. Contudo, com o crescimento da computação em nuvem, foram criados vários modelos de serviços especializados. Tanto que se fala até em XaaS (*Everything as a Service*), que poderia ser entendido como “qualquer coisa como serviço” ou “tudo como serviço” (DUAN *et al.*, 2015, [s.p.]).

Como não seria possível descrever tantos modelos especializados, vamos discutir alguns dos mais importantes no mercado atualmente. Não poderíamos deixar de falar do modelo denominado Banco de Dados como Serviço (DBaaS – *Database as a Service*) (SOUSA *et al.*, 2010). Este é um caso de especialização do modelo PaaS, no qual o cliente pode criar um banco de dados sem ter que instalar e configurar o SGBD. A infraestrutura computacional necessária e o gerenciamento do SGBD são responsabilidades do provedor. O gerenciamento de dados na nuvem envolve grandes desafios, por exemplo, em termos de segurança dos dados, escalabilidade e qualidade de serviço (SOUSA *et al.*, 2010). Nesse contexto, um problema comum é quando uma aplicação em um provedor precisa acessar um banco de dados que está em outro provedor. Nesse caso, a latência de acesso aos dados pode comprometer o desempenho do sistema caso a aplicação precise consultar o banco com muita frequência. O ideal é que a aplicação e o banco de dados sejam implantados em máquinas virtuais hospedadas na mesma máquina física, ou, pelo menos, na mesma rede local, de forma que a latência seja mínima.

Refleta

Com a migração de dados e aplicações para a nuvem, o uso de banco de dados como serviço tem aumentado nos últimos anos. O fato de não ter que se preocupar com o gerenciamento da infraestrutura é algo muito importante para administradores de bancos de dados. Porém, um dos aspectos mais críticos no uso do modelo DBaaS são a segurança e o sigilo dos dados. Algumas questões importantes para refletir são: é recomendável ter bancos de dados de clientes diferentes hospedados na mesma máquina virtual? Será que as tecnologias de virtualização são capazes de prover isolamento e proteção dos dados de clientes diferentes que compartilham os mesmos equipamentos? Não é fácil para empresas tradicionais aceitarem que seus dados (e os dados de seus clientes) serão armazenados em instalações externas à empresa, ou seja, nos provedores.

Os provedores de serviços no modelo DBaaS oferecem recursos avançados como suporte a replicação e balanceamento de carga automático (SOUSA, 2010). Dessa forma, os provedores são capazes de, automaticamente, criar réplicas do banco de dados a fim de atender a eventual aumento no número de consultas ao banco. Assim, as requisições de acesso aos dados são divididas entre as réplicas do banco, o que permite melhorar o desempenho do sistema. Além disso, os serviços incluem suporte tanto para bancos relacionais como para bancos não relacionais (NoSQL). Um exemplo do primeiro caso é o Azure SQL Server da Microsoft. Entre os bancos não relacionais, podemos

destacar o MongoDB que também é oferecido como plataforma na nuvem por vários provedores, como o MongoDB Atlas. Em geral, o cálculo de tarifas relacionadas a serviços de suporte a banco de dados não é uma tarefa simples, pois os custos envolvem os volumes de dados armazenados, a quantidade de transações no banco e o tempo de uso das máquinas virtuais utilizadas para executar uma ou mais instâncias do SGBD.

Pesquise mais

As plataformas que oferecem bancos de dados como serviços são muito populares, já que praticamente todas as aplicações fazem uso de bases de dados. Entre os principais SGBDs no mercado, podemos citar o MySQL. Verifique a descrição dos dois provedores indicados a seguir que oferecem o MySQL como serviço e compare as funcionalidades, benefícios e o preço mensal por GB de dados armazenados em cada provedor.

- Banco de Dados do Azure para MySQL: MySQL da comunidade pronto para empresas e totalmente gerenciado. Microsoft Azure.
- Amazon RDS for MySQL: Configure, opere e escale um banco de dados relacional na nuvem com apenas alguns cliques.

Outro importante modelo especializado é o *Backend* como Serviço (BaaS – *Backend as a Service*) (LANE, 2015). Esse tipo de plataforma é importante para agilizar e aprimorar o desenvolvimento de aplicações Web e aplicativos para dispositivos móveis (BATSCHINSKI, 2016). Nesse sentido, o serviço do provedor inclui o provisionamento de diversos recursos, tais como armazenamento de dados e objetos, mecanismos de autenticação, etc. Esses recursos são comuns no *backend* de sistemas para usuários finais. Com isso, o cliente pode focar no desenvolvimento do *frontend*, ou seja, a parte da aplicação com a qual os usuários interagem diretamente. Essa abordagem reduz o tempo de desenvolvimento dos sistemas. Um exemplo bastante representativo desse modelo é a plataforma Firebase da Google (FIREBASE, 2019). O Firebase inclui bancos de dados com sincronização em tempo real, serviços de armazenamento de objetos (como imagens e vídeos), soluções de autenticação baseadas em e-mail ou número de telefone dos usuários e recursos para hospedagem de aplicações web.

Por fim, vamos destacar um último exemplo de modelo de serviço especializado: Processos de Negócio como Serviço (BPaaS – *Business Process as a Service*). Esse modelo de serviço consiste no provisionamento

de ferramentas para modelagem de fluxos de trabalho, integração de dados e aplicações corporativas em vários segmentos (LYNN, 2014). Isso permite às grandes organizações implementar soluções de gestão corporativa que são flexíveis e capazes de responder prontamente a mudanças nos ambientes de negócios, sem que essas organizações tenham que se preocupar com o gerenciamento da infraestrutura de TI necessária para operar tais soluções.

Caro aluno, chegamos ao final de mais uma seção, na qual descrevemos os modelos de serviços básicos e também alguns modelos especializados. É importante entender os benefícios de cada modelo e o nível de abstração oferecido por cada um. Assim, o cliente pode fazer a escolha certa de acordo com sua demanda e considerando o nível de controle necessários sobre os serviços e os custos para operar e também migrar aplicações.

Sem medo de errar

Vimos que a empresa de fabricação de peças para aeronaves precisa de um ambiente de TI com recursos computacionais de alto desempenho para a execução de um software de simulação fornecido por uma empresa parceira. Na seção anterior, foi mostrado que a opção por um ambiente de nuvem é uma abordagem atrativa devido à redução de custos e abstração da complexidade da infraestrutura. O próximo passo é determinar qual dos modelos de serviço é mais apropriado para este caso: SaaS, PaaS ou IaaS. Você já refletiu sobre quais são as vantagens e desvantagens de cada modelo? Vamos analisar caso a caso.

O modelo SaaS seria interessante caso o fornecedor do software de simulação disponibilizasse esse software como serviço. No entanto, o alto nível de abstração e a falta de controle sobre os recursos não são apropriados em um cenário no qual o desempenho é um fator crítico. Sabemos que simulações de projetos dessa natureza são processos que impõem carga de processamento intensiva. O modelo SaaS é mais interessante para aplicações interativas voltadas para usuários finais.

O modelo PaaS poderia então ser uma alternativa interessante, pois oferece um certo nível de controle administrativo, embora limitado, sobre os recursos computacionais subjacentes. No entanto, não é comum encontrar provedores que oferecem plataformas com softwares muito específicos. Em geral, as plataformas como serviço incluem pilhas de tecnologias amplamente usadas para desenvolvimento de aplicações web ou aplicativos para dispositivos móveis. Além disso, foi mencionado que o software de simulação utilizado exige privilégios administrativos sobre os servidores para customização do ambiente de execução e ajustes de desempenho.

Dessa forma, o modelo mais apropriado seria IaaS. Nesse caso, a empresa poderia alocar dinamicamente máquinas virtuais nas quais poderia instalar o SO adequado e o software de simulação com eventuais dependências. Assim, a empresa reduziria os custos iniciais para montar a infraestrutura necessária e não teria problemas com falta ou desperdícios de servidores, pois poderiam ser alocados ou liberados conforme a demanda de projetos da empresa. Além disso, a empresa não teria que arcar com despesas ligadas à manutenção e operação da infraestrutura.

Outro fator importante é o nível de controle que, neste caso, é completo, pois no modelo IaaS o cliente é responsável por instalar e configurar o sistema. Por outro lado, todos os aspectos de segurança e desempenho da solução seriam de competência do cliente. O Quadro 1.1 resume as vantagens e desvantagens de cada modelo para o caso analisado.

Quadro 1.1 | Resumo comparativo entre os modelos de serviço para o caso da fabricante de peças

Modelo de Serviço	Vantagens	Desvantagens
SaaS	- Atualizações de software automáticas	- Falta de controle sobre os recursos computacionais.
PaaS	- Ambiente pré-configurado.	- Inexistência de ambiente customizado para a plataforma de simulação específica. - Falta de controle sobre os recursos computacionais.
IaaS	- Nível de controle suficiente sobre os recursos para implantação e configuração do software de simulação. - Alocação dinâmica de servidores. - Redução de custos com aquisição e manutenção da infraestrutura.	- Responsabilidade sobre administração dos sistemas e mecanismos de segurança.

Fonte: elaborado pelo autor.

Portanto, podemos concluir que a melhor alternativa para a fabricante de peças é usar um ambiente de comutação em nuvem conforme o modelo IaaS. Essa abordagem atende aos requisitos de controle e desempenho e reduz custos. A única desvantagem seria o baixo nível de abstração do modelo, o que exige do cliente responsabilidade de configuração da infraestrutura. Enfim, apesar de existir uma dificuldade, o IaaS é o modelo que atende os requisitos e apresenta mais vantagens para o cenário considerado. Este estudo

comparativo entre os modelos básicos de serviços em nuvem nos permitiu um melhor entendimento das suas características, vantagens e desvantagens.

Quadro 1.2 | Resumo comparativo entre os modelos IaaS e PaaS para o caso da empresa de desenvolvimento de software

Modelo de Serviço	Vantagens	Desvantagens
PaaS	<ul style="list-style-type: none">- Ambiente pronto para implementação e teste de software, já configurado com sistema operacional, bibliotecas, compiladores, entre outras ferramentas.- Maior nível de abstração: agilidade no provisionamento e transparência em relação aos detalhes da infraestrutura.- Mecanismos automáticos para balanceamento de carga para prover escalabilidade das aplicações.	<ul style="list-style-type: none">- Risco de problemas caso seja necessário migrar para outro provedor.
IaaS	<ul style="list-style-type: none">- Maior nível de controle sobre a infraestrutura (possibilidade de customização do ambiente).	<ul style="list-style-type: none">- Responsabilidade sobre alocação, configuração e gerenciamento dos servidores e mecanismos de segurança.- Necessidade de instalação e gerenciamento do ambiente de desenvolvimento.

Fonte: elaborado pelo autor.

Avançando na prática

Desenvolvimento de software na Nuvem para empresas de pequeno porte

Foi solicitada a uma pequena empresa de desenvolvimento de software a criação de um sistema de gestão de eventos. O sistema deve ter uma aplicação Web e um aplicativo para dispositivos móveis. A aplicação deve ter funcionalidades como: criar um evento, verificar a programação de um evento e fazer inscrição em uma atividade. Deve incluir, ainda, funcionalidades adicionais para organizadores de evento, tais como validar a inscrição e registrar a presença dos participantes por meio de *QR code*. Como se trata de uma empresa nova e pequena, não há grande disponibilidade de recursos computacionais e o número de funcionários é pequeno. A empresa então vai optar por um ambiente de nuvem para desenvolvimento do sistema. Nesse caso, qual seria o modelo de serviço mais adequado: SaaS, PaaS ou IaaS?

Resolução da situação-problema

Primeiramente, já podemos descartar a opção de SaaS, pois, na verdade, a empresa não precisa usar um software específico disponível na nuvem, mas desenvolver a solução conforme requisitado.

Precisamos, então, fazer uma escolha entre IaaS e PaaS. Pode-se perceber que pela demanda apresentada, não há necessidade do nível de controle oferecido no modelo IaaS. Portanto, pode-se fazer uma opção por um modelo com maior nível de abstração. Além disso, se a empresa escolhesse o modelo IaaS, ela teria que se preocupar com a instalação e configuração de todo o ambiente necessário para desenvolvimento, teste e implantação do sistema. Essas tarefas demandam um tempo considerável e profissionais habilitados, o que representaria um alto investimento inicial para uma empresa pequena.

Podemos concluir que o modelo PaaS é a melhor opção. Os provedores de PaaS oferecem ambientes para desenvolvimento e hospedagem de sistemas e bancos de dados com as tecnologias e ferramentas mais importantes para desenvolvimento de aplicação web e aplicativos. O ambiente é rapidamente provisionado e dispensa a necessidade de gerenciamento da infraestrutura, o que agiliza e reduz os custos para desenvolvimento das soluções. Além disso, os provedores de PaaS oferecem mecanismos de replicação e balanceamento de carga automáticos para facilitar o trabalho de garantir a escalabilidade e desempenho da solução desenvolvida. Essas facilidades são muito importantes, principalmente para empresas menores que não dispõem de equipes especializadas para lidar com esses desafios. O Quadro 1.2 resume as vantagens e as desvantagens dos modelos IaaS e PaaS para empresa de desenvolvimento de software.

Faça a valer a pena

1. Existem três modelos básicos de serviço no paradigma de computação em nuvem: IaaS, PaaS e SaaS. Esses modelos oferecem diferentes níveis de controle sobre os recursos alocados dos provedores. Considerando as características desses modelos, avalie as afirmativas a seguir:

- I. O modelo IaaS é adequado para clientes que precisam de um alto nível de controle sobre os recursos alocados em ambiente de computação em nuvem.
- II. Ao usar um serviço PaaS, o cliente é responsável por instalar as ferramentas necessárias para o desenvolvimento de aplicações no ambiente de nuvem.
- III. Um Software como Serviço é uma aplicação na nuvem na qual o usuário final não tem qualquer controle administrativo sobre a plataforma subjacente.

Considerando o contexto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas
- c. I e II, apenas.
- d. II e III, apenas.
- e. I e III, apenas.

2. Um dos modelos de serviço especializado mais importantes na computação em nuvem é o Banco de Dados como Serviço (DBaaS). Esse serviço permite que um cliente crie instâncias de bancos de dados nos provedores.

Considerando o modelo (DBaaS), avalie as afirmativas a seguir:

- I. Uma das vantagens do serviço DBaaS para pequenas e médias empresas é que elas podem manter bancos de dados escaláveis e confiáveis sem ter que gerenciar servidores.
- II. Como o acesso ao banco de dados é remoto, o desempenho da rede de acesso pode comprometer a qualidade do serviço e mecanismos de segurança devem ser usados na transmissão de dados críticos.
- III. Um dos benefícios do serviço DBaaS é a facilidade para o cliente migrar o banco de dados de um provedor para outro, pois o banco de dados já está em ambiente de nuvem.

Considerando o contexto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

3. Um provedor de serviços de computação em nuvem pode ter diferentes responsabilidades, de acordo com o modelo de serviço oferecido. Além disso, o mesmo provedor pode oferecer serviços em modelos diferentes, por exemplo, serviços IaaS e serviços PaaS.

Considerando o papel dos provedores no contexto de diferentes modelos de serviço, assinale a alternativa correta:

- a. No modelo SaaS, o provedor não precisa monitorar o uso de recursos computacionais pelos clientes, pois esses clientes não alocam infraestrutura.
- b. Os provedores são responsáveis por todos os aspectos de segurança dos serviços em ambiente de computação em nuvem.
- c. Um provedor de serviços PaaS não precisa ter sua própria infraestrutura, pois ele pode provisionar esses serviços em máquinas virtuais alocadas de algum provedor IaaS.
- d. No modelo PaaS, o provedor delega controle administrativo completo dos recursos para o cliente.
- e. Um provedor DBaaS não precisa de infraestrutura virtualizada.

Modelos de implantação em computação em nuvem

Diálogo aberto

Existem diversas abordagens e tecnologias que podem ser utilizadas pelos provedores no provisionamento de serviços de computação em nuvem. Nesta seção, vamos estudar quais são os principais modelos e plataformas para implantação da infraestrutura de ambientes de computação em nuvem. Os principais modelos são: nuvem privada e nuvem pública. No modelo de nuvem privada, o ambiente de computação em nuvem é de uso exclusivo dos usuários de uma única organização, por exemplo, uma empresa ou uma faculdade. Por outro lado, no modelo de nuvem pública, a infraestrutura gerenciada por um provedor é compartilhada entre diversas organizações clientes, que podem ser empresas, instituições ou até mesmo indivíduos. É importante entender como os modelos de implantação **impactam** os usuários de serviços em nuvem em termos, principalmente, de custo, controle de acesso aos recursos e nível de segurança. Como analista de TI, você deve ser capaz de compreender as vantagens e desvantagens de cada modelo, considerando os requisitos das aplicações dos clientes.

Para compreender melhor esse assunto, vamos retomar o contexto inicial no qual lidamos com a demanda da empresa de fabricação de peças para aeronaves. A empresa precisa de uma infraestrutura de alto desempenho para execução de simulações computacionais. Já analisamos os benefícios que a empresa teria ao adotar um ambiente de computação em nuvem e também determinamos o modelo de serviço mais apropriado: IaaS. Nesta etapa, vamos considerar especificamente a questão da abordagem a ser utilizada para implantação da infraestrutura de computação em nuvem. Na tapa da consultoria, sua responsabilidade envolve avaliar qual é o modelo de implantação mais adequado para essa empresa. Nesse sentido, você deve elaborar um parecer para a empresa, com uma justificativa sobre o modelo de implantação mais apropriado para o caso descrito. Você precisa avaliar questões como: qual modelo de implantação envolve maiores custos? Qual modelo oferece maior nível de privacidade dos dados?

Para ajudá-lo com este último desafio, esta seção descreve os principais modelos de implantação de serviços de computação em nuvem, destacando quais são as vantagens e desvantagens de cada modelo. Além disso, serão descritas as principais plataformas de software livre para gerenciamento de

infraestruturas de computação nuvem. Vamos discutir as características e os recursos oferecidos por essas plataformas.

Ao completar esta unidade, você terá uma compreensão abrangente sobre os fundamentos de serviços de computação em nuvem e será capaz de fazer uma análise comparativa dos diferentes modelos de serviço e modelos de implantação.

Bons estudos!

Não pode faltar

Caro aluno, nas seções anteriores, discutimos os benefícios da computação em nuvem e os seus principais modelos de serviço. Agora, vamos aprender quais são as principais abordagens e tecnologias para a implantação dos ambientes de computação em nuvem? Vamos estudar as principais soluções para implantação de uma infraestrutura de recursos de TI a fim de viabilizar o provisionamento de serviços de computação em nuvem.

Um fator extremamente relevante no uso de serviços de computação em nuvem é escolher o modelo de implantação adequado. Esse modelo determina como será o gerenciamento da infraestrutura de TI. Consequentemente, os modelos de implantação definem aspectos de controle de acesso, segurança e disponibilidade dos recursos computacionais ofertados como serviços no ambiente de computação em nuvem. Existem quatro modelos de implantação: nuvem privada, nuvem comunitária, nuvem pública e nuvem híbrida (MELL; GRANCE, 2011).

Uma nuvem privada é um ambiente de nuvem de propriedade de uma única organização (ERL; PUTTINI; MAHMOOD, 2013). Assim, os recursos computacionais são utilizados exclusivamente pela organização que é proprietária da infraestrutura. O modelo de implantação nuvem privada permite que uma empresa ou instituição utilize tecnologias de computação em nuvem para gerenciar uma infraestrutura de TI própria. Dessa forma, os recursos computacionais de uma organização podem ser compartilhados como serviços sob demanda entre diversas unidades e usuários internos à organização. Nesse caso, é necessário implementar políticas que restringem o acesso aos recursos disponíveis para os membros da organização, de acordo com os privilégios de cada um. Para isso, são necessárias tecnologias de autenticação e controle de acesso e também a configuração adequada dos serviços de rede. Um exemplo de nuvem privada é o caso de uma empresa que adquire uma ampla infraestrutura de TI e usa tecnologias de computação em nuvem para gerenciar essa infraestrutura de forma centralizada. A empresa então permite que seus diversos setores ou filiais possam alocar

recursos computacionais sob demanda. Os níveis de acesso aos serviços e a quantidade de recursos permitidos para cada setor da empresa podem ser diferentes.

Uma organização pode ter uma nuvem privada administrada por terceiros, neste caso, uma empresa especializada no provisionamento de serviços de computação em nuvem (SOUSA; MOREIRA; MACHADO, 2009). O que caracteriza a nuvem privada é o uso exclusivo dos recursos por uma determinada organização. Nesse caso, não importa quem é o responsável pela administração da infraestrutura. Além disso, a infraestrutura não precisa, necessariamente, ser implantada nas instalações da própria organização. Os equipamentos podem estar instalados em espaço geográfico fora da organização com o acesso aos serviços realizado remotamente pela Internet.

Assimile

Você aprendeu que um dos principais benefícios da computação em nuvem é a redução de custos, pois o usuário dos serviços não precisa se preocupar com a aquisição, instalação e manutenção da infraestrutura de TI. Porém, esse benefício não existe no caso do modelo de nuvem privada, uma vez que o ambiente de nuvem é propriedade da organização que utiliza os serviços, assim como os custos para administração desse ambiente. Mesmo assim, esse modelo é interessante em diversos cenários, pois, além de manter outros benefícios da computação em nuvem, como a flexibilidade e a eficiência no gerenciamento dos recursos computacionais, o modelo de nuvem privada evita alguns problemas como aqueles relacionados com segurança e privacidade quando uma empresa faz uso de serviços de um provedor externo. Na nuvem privada, o controle sobre a infraestrutura é completo e exclusivo da organização proprietária. Esse fator é importante para instituições governamentais ou empresas do setor financeiro, por exemplo, por se tratar de organizações que lidam com dados críticos ou confidenciais que, muitas vezes, não podem ser hospedados em provedores externos. Exemplos desses tipos de dados são: dados bancários ou prontuário médico de indivíduos.

O modelo de nuvem privada é mais apropriado para organizações que querem se beneficiar de tecnologias de nuvem para ter flexibilidade e eficiência no compartilhamento de recursos de TI, sem abrir mão do controle administrativo de uma infraestrutura dedicada. Porém, esse modelo de implantação pode exigir custos altos relacionados à propriedade e exclusividade da infraestrutura. Uma alternativa para isso, é o modelo de implantação

denominado nuvem comunitária, no qual um conjunto de organizações compartilham um mesmo ambiente de computação em nuvem (SOUSA; MOREIRA; MACHADO, 2009). Assim, a infraestrutura é dedicada para um conjunto de organizações.

Essa operação conjunta da infraestrutura pode ser motivada por interesses comuns e requer um alinhamento em termos de requisitos de segurança e políticas de acesso. Empresas com políticas ou interesses conflitantes poderiam ter problemas para compartilhar um ambiente de nuvem. Em geral, a administração do ambiente é responsabilidade de um consórcio ou empresa especializada e o acesso aos recursos é remoto. Um exemplo típico de nuvem comunitária é o caso de um grupo de universidades que podem cooperar com recursos financeiros e profissionais especializados para implantar um ambiente de nuvem para um projeto de pesquisa que exige uma infraestrutura computacional de alto desempenho.

Um outro modelo de implantação é a nuvem pública, em que os recursos do ambiente de computação em nuvem são mantidos por um provedor e disponibilizados sob demanda para qualquer empresa cliente (SOUSA; MOREIRA; MACHADO, 2009). Isso representa um modelo de negócio no qual um provedor faz o investimento na infraestrutura e assume a operação com o intuito de faturar com a oferta de serviços de computação em nuvem ao público em geral.

O modelo de nuvem pública estabelece uma clara distinção entre o provedor e o cliente. Os clientes podem alocar recursos computacionais sob demanda de diversos provedores sem ter que assumir a responsabilidade e os custos associados com a gestão da infraestrutura. Além disso, os clientes se beneficiam de escalabilidade “ilimitada”, pois eles podem alocar ou liberar recursos dos provedores de forma rápida conforme a necessidade de suas aplicações.

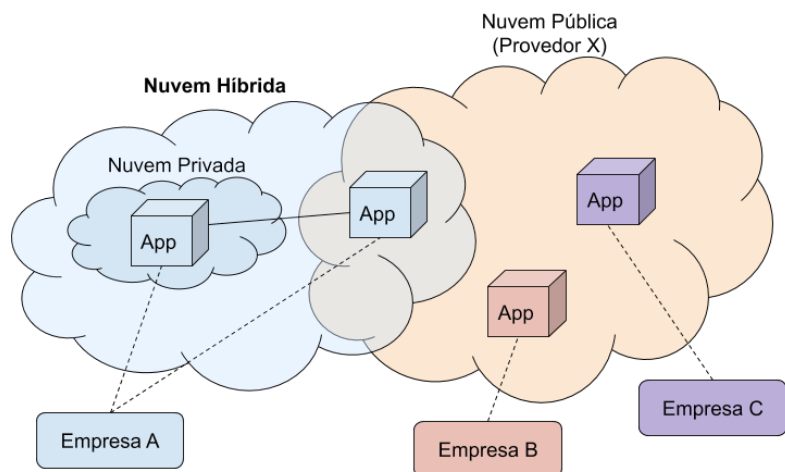
Exemplificando

O modelo de nuvem pública permite que qualquer empresa, profissional e até mesmo usuários finais tenham acesso a aplicações e serviços de TI em nuvem. Como o pagamento é baseado no uso, os custos podem ser minuciosamente controlados. Assim, mesmo pequenas empresas podem ter acesso a servidores robustos e soluções de TI sofisticadas, em áreas como IoT, Big Data e Inteligência de Negócios. Entre os mais importantes provedores de computação em nuvem que oferecem serviços ao público em geral podemos citar: Amazon Web Services (AWS) (AMAZON, 2019), Google Cloud Platform (GOOGLE, 2019), Microsoft Azure (MICROSOFT, 2019) e IBM Cloud (IBM, 2019).

É importante salientar que não é necessário optar por um modelo de implantação específico. Se for conveniente, uma organização pode, por exemplo, fazer uso de uma nuvem privada e de uma nuvem pública ao mesmo tempo. Esse modelo caracteriza o que se denomina uma nuvem híbrida, ou seja, um ambiente de computação em nuvem que é a combinação de modelos de implantação diferentes (ERL; PUTTINI; MAHMOOD, 2013). Mesmo havendo a composição de duas ou mais infraestruturas de nuvem com modelos de implantação diferentes, a nuvem híbrida é vista como um sistema único. Por exemplo, uma empresa pode manter diversos componentes de uma mesma aplicação em ambientes diferentes da nuvem híbrida. Componentes que envolvem manipulação de dados críticos ou sigilosos podem ficar em uma nuvem privada. Os demais componentes da aplicação podem ser hospedados em provedor de nuvem pública para fins de escalabilidade.

Como discutido em (ERL; PUTTINI; MAHMOOD, 2013), nuvens híbridas representam cenários complexos uma vez que envolvem a divisão de responsabilidades administrativas entre os responsáveis por cada ambiente e também devido a eventuais problemas de compatibilidade, pois cada ambiente da nuvem híbrida pode ter configurações e tecnologias específicas. A Figura 1.3 ilustra um cenário no qual a Empresa A aloca recursos do Provedor X (nuvem pública) para ampliar a capacidade de sua nuvem privada formando então uma nuvem híbrida. Parte da aplicação da Empresa A está na nuvem privada e outra parte está na nuvem pública, compartilhada com aplicações de outras empresas clientes do provedor. Esse cenário é interessante para o suporte a aplicações ou serviços suscetíveis a sazonalidade, ou seja, quando a quantidade de acessos (carga de trabalho) aumenta significativamente em determinados períodos. Por exemplo, como ocorre com aplicações de comércio eletrônico em datas comemorativas ou na Black Friday. Neste caso, recursos de nuvem pública podem ser temporariamente alocados para atender ao aumento repentino na demanda, evitando-se lentidão ou indisponibilidade dos sistemas devido à grande quantidade de acessos dos clientes (sobrecarga).

Figura 1.3 | Cenário de caracterização de nuvem híbrida



Fonte: elaborada pelo autor.

Assimile

Além da nuvem híbrida, a Nuvem Privada Virtual (VPC – Virtual Private Cloud) é outro modelo que busca aproveitar as vantagens tanto de nuvens públicas quanto privadas (WOOD, 2009). Uma VPC consiste na alocação de recursos computacionais em um provedor de nuvem pública de forma dedicada para uma organização. Dessa forma, é como se uma nuvem privada fosse criada em um provedor de nuvem pública. Para isso, são usados mecanismos de autenticação e criptografia para isolar os recursos alocados dos demais clientes da nuvem pública. Assim, a VPC envolve também virtualização de recursos de rede para criar uma rede privada virtual no ambiente de nuvem pública. Esse modelo de nuvem privada virtual também é considerado uma forma menos complexa de migração para ambiente de nuvem (ZHANG, 2010).

Nesse contexto de heterogeneidade, no qual um cliente pode alocar recursos computacionais em diferentes ambientes de computação em nuvem, um aspecto muito importante é a gerência de identidades federadas (FELICIANO, 2011). Nuvens federadas são ambientes de computação em nuvem que, mesmo administradas por provedores diferentes, compartilham recursos entre si (MORENO-VOZMEDIANO; MONTERO; LLORENTE, 2012). Esse compartilhamento pode se dar em vários níveis, mas sempre é necessário um mecanismo comum para autenticação e controle de acesso aos recursos. Dessa forma, um cliente pode usar um mesmo identificador para acessar recursos de diferentes provedores. O Quadro 1.3 apresenta sucintamente a descrição de cada modelo de implantação.

Modelo de Implantação	Descrição
Nuvem privada	Ambiente de computação em nuvem dedicado a uma única organização.
Nuvem comunitária	Ambiente de computação em nuvem dedicado a um conjunto de organizações que compartilham os recursos.
Nuvem pública	Ambiente de computação em nuvem mantido por um provedor que então oferta serviços sob demanda para diversos clientes.
Nuvem híbrida	Ambiente de computação em nuvem que envolve mais de um modelo de implantação.

Fonte: elaborado pelo autor.

Você aprendeu sobre os principais modelos conceituais para implantação de nuvens, mas uma questão que ainda não foi explicada é a seguinte: como gerenciar a infraestrutura de ambientes de computação em nuvem? A resposta é: utilizando uma plataforma de computação em nuvem (CARISSIMI, 2015). Essa plataforma consiste no conjunto de ferramentas de software utilizado para gerenciar os servidores e equipamentos de rede que compõem a infraestrutura sobre a qual serão provisionados os serviços de computação em nuvem. Para dar suporte a esses serviços, a plataforma tem que oferecer tecnologias de virtualização dos recursos computacionais de processamento, armazenamento e transmissão de dados, assim como mecanismos para alocação desses recursos virtualizados de forma remota, ou seja, por meio de acesso via Internet. Em geral, os mecanismos de acesso remoto utilizam o protocolo SSH (*Secure Shell*) ou o protocolo HTTP (*Hypertext Transfer Protocol*). O protocolo SSH oferece uma conexão por terminal de linha de comando. Com o uso do HTTP, o gerenciamento e a alocação dos recursos podem ser feitos por meio de uma aplicação Web.

As plataformas de computação em nuvem são também conhecidas como software de servidor para nuvem ou ainda sistema operacional de nuvem. Os provedores de nuvem pública muitas vezes utilizam plataformas proprietárias para gerenciamento de sua infraestrutura. No entanto, existem muitas plataformas de computação em nuvem distribuídas como softwares livres. Essas plataformas podem ser utilizadas para a implantação de nuvens públicas

ou privadas. Entre as principais plataformas, pode-se destacar as seguintes: OpenStack, CloudStack, Eucalyptus e OpenNebula (THOMÉ; HENTGES; GRIEBLER, 2013).

Refleta

Uma das vantagens de um ambiente de computação em nuvem é flexibilidade na alocação dos recursos computacionais, que viabiliza a elasticidade rápida. Os recursos podem ser rapidamente alocados ou liberados de forma automática. Quais seriam as funcionalidades importantes em uma plataforma de computação em nuvem para viabilizar essa alocação dinâmica de recursos? Como a alocação de recursos pode ser efetivamente automatizada, ou seja, ser realizada sem a intervenção humana?

A plataforma OpenStack pode ser definida como “uma combinação de ferramentas open source (conhecidas como projetos) que usam um pool de recursos virtuais para criar e gerenciar nuvens privadas e públicas” (REDHAT, 2019). O OpenStack funciona como um sistema operacional para Nuvem que permite controlar recursos de processamento, armazenamento e rede (OPENSTACK, 2019).

A plataforma OpenStack foi criada a partir de uma colaboração entre a NASA e a empresa Rackspace a partir de 2010. As principais funcionalidades da plataforma incluem gerenciamento de máquinas virtuais, orquestração de contêineres, balanceamento de carga, virtualização de funções de redes e controle de acesso aos recursos. Com uma arquitetura modular, a plataforma OpenStack combina diversos módulos funcionais (chamados de projetos) (REDHAT, 2019). Os projetos básicos do OpenStack são:

- **Nova:** gerenciamento de instâncias computacionais (por exemplo, criação de máquinas virtuais).
- **Neutron:** gerenciamento de conectividade e virtualização de serviços de rede.
- **Swift:** armazenamento distribuído de alta disponibilidade para objetos e dados não estruturados como vídeos, backups, imagens, etc.
- **Cinder:** gerenciamento de armazenamento em bloco (discos virtuais), por exemplo, para criar dispositivo lógico de armazenamento persistente para máquinas virtuais.
- **Keystone:** gerenciamento de identidades, autenticação e controle de acesso.

- **Glance:** gerenciamento de imagens de máquinas virtuais.

Essas e outras ferramentas combinadas na plataforma OpenStack são disponibilizadas por meio de um painel de bordo que oferece controle administrativo e funcionalidades para provisionamento dos recursos, por meio de uma interface Web. Além da aplicação Web, as funcionalidades podem ser acessadas por terminal de comando, SDK (*Software Development Kit*) para linguagem python ou API (*Application Programming Interface*). O OpenStack provê um conjunto consistente de APIs para abstrair os recursos virtuais e viabilizar mecanismos de gerenciamento da infraestrutura do ambiente de computação em nuvem (REDHAT, 2019).

Outra plataforma relevante é a CloudStack da Fundação Apache (APACHE, 2019). Essa plataforma pode ser usada para implantação de Nuvens privadas, públicas e híbridas. Ela foi projetada para conseguir gerenciar redes de máquinas virtuais em larga escala e criar um ambiente de computação em nuvem de alta disponibilidade e alta escalabilidade. Uma característica interessante é que as tecnologias usadas nessa plataforma são compatíveis com serviços de provedores comerciais de computação em nuvem, como o AWS EC2, o que permite a integração de serviços para implantação de nuvens híbridas. A CloudStack permite o gerenciamento dos recursos por meio de interface Web, ferramentas de linha de comando e API Web.

Outra solução para gerenciamento de infraestrutura de nuvem é a plataforma Eucalyptus (EUCALYPTUS, 2019). Ela também oferece compatibilidade com os serviços de nuvem públicas, como a AWS, de forma que é possível criar nuvens privadas ou híbridas. Essa plataforma suporta diferentes tecnologias de virtualização e o gerenciamento dos recursos pode ser feito remotamente via ferramentas de linha de comando ou interface Web. A arquitetura dessa plataforma inclui nós controladores que reúnem informações de monitoramento da infraestrutura e permitem o acesso aos recursos da nuvem (CARISSIMI, 2015).

A plataforma OpenNebula é direcionada ao gerenciamento de nuvens privadas (OPENNEBULA, 2019), mas também permite a criação de ambientes híbridos com nuvens públicas AWS e Microsoft Azure. O foco dessa plataforma é prover uma solução leve, simples e flexível para o uso de virtualização no gerenciamento da infraestrutura. Uma característica de destaque dessa plataforma é o suporte para criação de nuvens federadas.

Pesquise mais

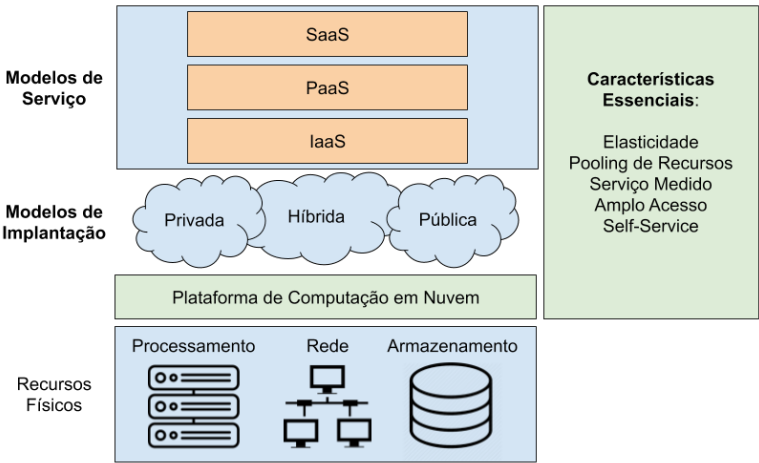
Vimos que existem diversas plataformas para gerenciamento de infraestruturas de computação em nuvem. Essas plataformas suportam diferentes modelos de implantação e permitem a alocação de recursos virtualizados

sob demanda. O artigo indicado a seguir apresenta uma análise comparativa entre várias plataformas, considerando aspectos como o suporte a mecanismos de economia de energia e o balanceamento de carga.

- THOMÉ, B.; HENTGES, E.; GRIEBLER, D. Computação em Nuvem: análise comparativa de ferramentas open source para IaaS. In: **11th Escola Regional de Redes de Computadores (ERRC)**, 2013.

Ao final desta seção, completamos a Unidade 1. Aprendemos muito sobre os benefícios e os desafios no uso de ambientes em nuvem e as características essenciais dos serviços de computação em nuvem. Além disso, descrevemos os principais modelos de serviço e modelos de implantação, assim como uma visão geral de plataformas para gerenciamento de infraestruturas de computação em nuvem. A Figura 1.4 ilustra como todos esses conceitos se relacionam. Uma plataforma é usada para virtualização dos recursos físicos a fim de se implantar um ambiente de nuvem, podendo ser privada, pública ou híbrida. Nesse ambiente, podem ser oferecidos serviços de software, plataforma ou infraestrutura, que, sob demanda, têm características em comum, como a elasticidade rápida, que caracterizam o ambiente em Nuvem.

Figura 1.4 | Arquitetura conceitual da Computação em Nuvem.



Fonte: elaborada pelo autor.

Você completou um estudo abrangente sobre os conceitos fundamentais do paradigma de computação em nuvem. Agora que concluiu esta primeira unidade, você já é capaz de analisar criteriosamente as vantagens e desvantagens dos diferentes modelos de computação em nuvem.

Sem medo de errar

A empresa que fabrica peças para aeronaves optou por usar um ambiente de computação em nuvem para alocar a capacidade computacional necessária para a execução das simulações do projeto das peças. O modelo de serviços escolhido foi o IaaS, para que se tenha o completo controle sobre os recursos de processamento e armazenamento de dados. Assim, a empresa pode alocar servidores virtuais e instalar e configurar as ferramentas de simulação conforme suas necessidades. Nesta última etapa, a tarefa consiste em avaliar qual é o melhor modelo para implantação dessa infraestrutura de computação em nuvem. Seu trabalho é elaborar um parecer justificando qual dos modelos de implantação deve ser utilizado pela empresa.

Para resolver essa tarefa, pode-se fazer uma análise caso a caso, considerando as vantagens e desvantagens de cada um dos modelos para o cenário descrito. Os modelos de implantação são: nuvem privada, nuvem comunitária, nuvem pública e nuvem híbrida.

A nuvem híbrida é adequada quando a capacidade instalada de uma nuvem privada precisa ser expandida a partir de recursos de uma nuvem pública. Essa necessidade de expansão geralmente ocorre devido a um crescimento inesperado da demanda por recursos computacionais. No cenário em questão, a empresa necessita de alta capacidade de processamento para realizar simulações da aerodinâmica de peças como parte do processo de fabricação. Em geral, não se observa alterações na demanda para um processo planejado como é o caso de projetos de novas peças industriais. Portanto, não se mostra clara a necessidade de implantação de uma nuvem híbrida. A infraestrutura exigida para a execução de ferramentas de simulação é alta, mas previsível.

Também não se apresenta nenhum motivo para implantação da infraestrutura conforme o modelo de Nuvem Comunitária. Esse modelo é adequado para aplicações nas quais diversas organizações trabalham em conjunto por um objetivo comum. A fabricante de peças não tem qualquer interesse em compartilhar a tarefa, pois, na verdade, o mais importante é o sigilo do processo, pois o segredo industrial do processo pode ser uma vantagem competitiva no mercado.

Restam, assim, duas alternativas. A empresa pode implantar sua própria infraestrutura, seguindo o modelo de nuvem privada, ou alocar a capacidade computacional necessária em um provedor de nuvem pública. As vantagens da nuvem privada são: maior controle sobre a infraestrutura para mais flexibilidade na customização do ambiente e maior nível de segurança e privacidade, pois os recursos não são compartilhados. As vantagens da nuvem pública são: maior redução de custos, pois a infraestrutura é compartilhada

com os demais clientes do provedor e pagamento será somente pelo tempo de uso dos recursos e mais escalabilidade, pois o provedor dispõe de larga escala de recursos que podem ser alocados sob demanda.

Dessa forma, a recomendação para a empresa é que se o critério mais importante para ela for o sigilo dos dados das simulações, por questões de propriedade intelectual, deve-se optar pelo modelo de nuvem privada. Por outro lado, se o fator mais importante é o custo, deve-se optar pelo modelo de nuvem pública.

Com essa análise, completa-se a avaliação da viabilidade e dos benefícios do uso de um ambiente de computação em nuvem pela fabricante de peças, assim como os modelos de serviço e de implantação mais adequados para a aplicação descrita. Essa capacidade de avaliar as vantagens e desvantagens de cada modelo é muito importante para um analista de TI, visto que é crescente a demanda por recursos em nuvem por parte de empresas de diversos setores do mercado.

Avançando na prática

Cloud Bursting

Uma rede de ensino superior com faculdades em várias cidades deseja criar uma aplicação web na qual os estudantes podem consultar a matriz curricular dos cursos e a oferta de disciplinas obrigatórias e eletivas, assim como realizar a matrícula em disciplinas a cada semestre. Como o número de estudantes nesta rede é muito grande, a escalabilidade da aplicação é um fator relevante. A instituição decidiu utilizar uma infraestrutura em nuvem para hospedar a aplicação. Sua tarefa, como analista de TI da rede, é avaliar qual o modelo de implantação de ambiente de nuvem deve ser utilizado para o cenário apresentado.

Resolução da situação-problema

Neste caso, a solução mais adequada seria o modelo de nuvem híbrida. Dessa forma, a rede de faculdades poderia manter a base de dados da aplicação em uma nuvem privada, em suas próprias instalações. Com o controle completo sobre os recursos dedicados, não haveria o risco de manter os dados dos alunos em um provedor público, com recursos compartilhados com outros clientes. Além disso, essa nuvem privada poderia ter uma capacidade reduzida, para se evitarem altos custos com implantação e operação da infraestrutura. Porém, essa nuvem privada, com poucos recursos, poderia não ser suficiente

para manter a escalabilidade do sistema com o aumento de carga de trabalho nos períodos de matrícula a cada semestre. Para isso, a rede de ensino poderia alocar servidores em uma nuvem pública para hospedar apenas réplicas da interface web do sistema, a base de dados continuaria na nuvem privada. Os servidores no provedor de nuvem pública seriam mantidos somente durante os períodos de matrícula, a fim de evitar eventual degradação de desempenho da aplicação. Fora dos períodos de matrícula, quando as consultas ao sistema são reduzidas, os recursos na nuvem pública poderiam ser liberados para redução de custos. Esse procedimento, característico de nuvem híbrida, que consiste em usar recursos adicionais de uma nuvem pública para aumentar temporariamente a capacidade de uma aplicação inicialmente hospedada em nuvem privada é denominado *Cloud Bursting*.

Faça valer a pena

1. Os modelos de implantação de computação determinam a responsabilidade sobre o gerenciamento da infraestrutura. Nesse contexto, avalie as afirmativas a seguir:

- I. O modelo de nuvem híbrida consiste em um ambiente, cujo gerenciamento é compartilhado por vários provedores.
- II. No modelo de nuvem privada, os custos de gerenciamento da infraestrutura são repassados para o provedor.
- III. No modelo de nuvem pública, o compartilhamento da infraestrutura entre vários clientes contribui para a redução dos custos de alocação dos recursos.

Considerando o contexto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

2. Cada modelo de implantação é mais adequado para determinados tipos de aplicações. O modelo de nuvem privada resulta em um grau maior de segurança e um grau menor de escalabilidade, quando comparado com os demais modelos. Além disso, a nuvem privada envolve maior custo dos serviços.

Entre os cenários descritos a seguir, para qual é apropriado o uso de uma nuvem privada:

- a. Um aplicativo de gestão de tarefas desenvolvido por uma empresa de TI de pequeno porte.
- b. Integração de sistemas de órgãos da Secretaria de Segurança Pública (SSP) do Estado.
- c. Consórcio entre universidades para projeto de pesquisa na área de Ciência dos Dados (*Data Science*).
- d. Sistema e-commerce para uma grande loja de departamentos.
- e. Serviço de web e-mail.

3. Os dois principais modelos de implantação, nuvem pública e nuvem privada, se diferem no nível de controle, confiabilidade e segurança à disposição dos usuários da infraestrutura. Eles também se opõem em relação aos limites de escalabilidade e aos custos dos serviços.

Considerando o contexto apresentado, é correto afirmar que a nuvem pública:

- a. Oferece maior grau de segurança que a nuvem privada.
- b. Oferece menor grau de escalabilidade que a nuvem privada.
- c. Tem custo relativamente maior que a nuvem privada.
- d. Oferece menor grau de controle que a nuvem privada.
- e. Oferece menor grau de confiabilidade que a nuvem privada.

Referências

- AMAZON EC2. Amazon Web Services (AWS). Disponível em: <https://aws.amazon.com/pt/ec2/>. Acesso em: 28 jun. 2019.
- APACHE CloudStack. About CloudStack. Disponível em: <https://cloudstack.apache.org/>. Acesso em: 24 abr. 2019.
- ARMBRUST, M. *et al.* A View of Cloud Computing. **Communications of the ACM**, v. 53, n. 4, p. 50–58, abr. 2010.
- BATSCHINSKI, G. Backend as a Service: Prós e Contras. **Revista InfoQ**. [S.l.], 2016. Disponível em: <https://www.infoq.com/br/news/2016/07/backend-pros-e-contras>. Acesso em: 28 jun. 2019.
- CARISSIMI, A. **Desmistificando a Computação em Nuvem**. ROSE, Cesar de; PASIN, Marcia; SCHNORR, Lucas. (Org.). Escola Regional de Alto Desempenho. Porto Alegre: SBC, p. 3–24, 2015.
- COUTINHO, E. *et al.* Elasticidade em computação na nuvem: Uma abordagem sistemática. **XXXI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2013) - Minicursos**, p. 1-44, 2013. Disponível em: <http://sbrc2013.unb.br/files/anais/minicursos/minicurso-5.pdf>. Acesso em: 27 jun. 2019.
- DUAN, Y. *et al.* Everything as a Service (XaaS) on the Cloud: Origins, Current and Future Trends. In: **The IEEE 8th International Conference on Cloud Computing**, 2015.
- ERL, T.; PUTTINI, R.; MAHMOOD, Z. **Cloud Computing: Concepts, Technology & Architecture**. [S.l.] Prentice Hall, 2013.
- EUCALYPTUS. Disponível em: <https://www.eucalyptus.cloud/index.html>. Acesso em: 28 jun. 2019.
- FELICIANO, G. *et al.* Gerência de identidades federadas em nuvens: Enfoque na utilização de soluções abertas. In: **XI Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg)**, 2011.
- FIREBASE. Google Firebase. Disponível em: <https://firebase.google.com/>. Acesso em: 28 jun. 2019.
- GOOGLE Cloud. Google App Engine. Portal do Google Cloud. Disponível em: <https://cloud.google.com/appengine/>. Acesso em: 28 jun. 2019.
- IBM. IBM Cloud. Disponível em: <https://www.ibm.com/br-pt/cloud>. Acesso em: 28 jun. 2019.
- LANE, K. Overview of the backend as a service (BaaS) space. **API Evangelist**, 2015.
- LYNN, T. *et al.* Towards a framework for defining and categorising business Process-As-A-Service (BPaaS). In: **21st International Product Development Management Conference**, 2014.
- MELL, P.; GRANCE, T. **The NIST definition of cloud computing**. [s.l.] National Institute of Standards and Technology, 2011. Disponível em: <https://doi.org/10.6028/NIST.SP.800-145>. Acesso em: 27 jun. 2019.

MICROSOFT Azure. SERVIÇO de Aplicativo do Azure – hospedagem de aplicativo. **Portal do Microsoft Azure**. Disponível em: <https://azure.microsoft.com/pt-br/services/app-service/>. Acesso em: 28 jun. 2019.

MORENO-VOZMEDIANO, R.; MONTERO, R. S.; LLORENTE, I. M. IaaS Cloud Architecture: From Virtualized Datacenters to Federated Cloud Infrastructures. **Computer**, v. 45, n. 12, p. 65–72, 2012.

OPENNEBULA. Discover OpenNebula. Disponível em: <https://opennebula.org/discover/>. Acesso em: 28 jun. 2019.

OPENSTACK. What is OpenStack? Disponível em: <https://www.openstack.org/software/>. Acesso em: 28 jun. 2019.

REDHAT. **Introdução ao OpenStack**. Disponível em: <https://www.redhat.com/pt-br/topics/openstack>. Acesso em: 28 jun. 2019.

SOUSA, F. R. C.; MOREIRA, L. O.; MACHADO, J. C. Computação em nuvem: Conceitos, tecnologias, aplicações e desafios. **II Escola Regional de Computação Ceará, Maranhão e Piauí (ERCCEMAPI)**, p. 150–175, 2009. Disponível em: https://www.researchgate.net/profile/Javam_Machado/publication/237644729_Computacao_em_Nuvem_Conceitos_Tecnologias_Aplicacoes_e_Desafios/links/56044f4308aea25fce3121f3.pdf. Acesso em: 27 jun. 2019.

SOUSA, F. R. *et al.* **Gerenciamento de dados em nuvem**: Conceitos, sistemas e desafios. Tópicos em sistemas colaborativos, interativos, multimídia, web e bancos de dados. Sociedade Brasileira de Computação, p. 101–130, 2010.

THOMÉ, B.; HENTGES, E.; GRIEBLER, D. Computação em Nuvem: análise comparativa de ferramentas open source para IaaS. In: 11th Escola Regional de Redes de Computadores (ERRC), 2013. Disponível em: http://larcc.setrem.com.br/wp-content/uploads/2017/03/THOME_ERRC_2013.pdf. Acesso em: 28 jun. 2019.

TIGRE, P. B.; NORONHA, V. B. Do mainframe à nuvem: inovações, estrutura industrial e modelos de negócios nas tecnologias da informação e da comunicação. **Revista de Administração**, v. 48, n. 1, p. 114–127, 2013. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0080210716302709>. Acesso em: 27 jun. 2019

WOOD, T. *et al.* The Case for Enterprise-ready Virtual Private Clouds. In: **Proceedings of the 2009 Conference on Hot Topics in Cloud Computing**, Berkeley, CA, USA: USENIX Association, 2009.

ZHANG, Q.; CHENG, L.; BOUTABA, R. Cloud computing: state-of-the-art and research challenges. **Journal of internet services and applications**, v. 1, n. 1, p. 7–18, 2010.

Unidade 2

Neumar Costa Malheiros

Tecnologias e soluções de computação em nuvem

Convite ao estudo

Caro aluno, já aprendemos alguns conceitos básicos sobre a computação em nuvem, como os modelos de serviço e os modelos de implantação. Nesta unidade, vamos discutir os fatores envolvidos na migração de aplicações para a nuvem. Assim, precisamos entender os custos envolvidos, as formas de acessar os serviços e como controlar as permissões de acesso aos recursos. Você conhece os desafios na hospedagem de aplicações na nuvem? É preciso escolher o modelo de serviço e as tecnologias mais apropriadas.

Para compreender melhor esses tópicos, vamos considerar o seguinte cenário. Você é um analista de TI em uma empresa de comércio eletrônico de médio porte. Essa empresa mantém uma aplicação web por meio da qual são realizadas as vendas dos produtos. Esse sistema de vendas é mantido no setor de TI da própria empresa. Com o objetivo de reduzir custos e melhorar a escalabilidade e a confiabilidade da sua aplicação, essa empresa deseja migrar essa aplicação para um provedor de computação em nuvem. A aplicação foi desenvolvida há alguns anos e o banco de dados que mantém registros das vendas e dos clientes já ocupa um grande volume de dados, da ordem de 4 terabytes, incluindo os backups. Além disso, a aplicação web de comércio eletrônico foi implementada utilizando um framework cujo desenvolvimento foi descontinuado. Você foi escolhido para liderar a equipe de transição, responsável por migrar a aplicação web das instalações da empresa para um ambiente de nuvem pública. Entretanto, você precisa compreender quais são as variáveis utilizadas na precificação dos serviços em nuvem e também determinar como configurar as permissões de acesso da equipe de TI que trabalha na manutenção da aplicação e do banco de dados. Além disso, você terá que determinar o modelo e tecnologia para hospedar a aplicação no provedor.

No decorrer desta unidade, vamos explicar os diversos fatores relacionados com os provedores de nuvem pública, para que possamos compreender os custos, desafios e alternativas para migrar aplicações para esses provedores.

- Na Seção 1, vamos estudar quais são as principais tecnologias que viabilizaram o surgimento do paradigma de computação em nuvem, com foco em modelos de virtualização.
- Na Seção 2, vamos explorar aspectos relacionados com os modelos de tarifação usados pelos provedores para determinar os custos de serviços em nuvem.
- Por fim, na Seção 3, vamos aprender sobre a migração de aplicações para a nuvem. Nesse contexto, abordaremos tópicos como as diferentes formas de acessar os serviços em nuvem e os mecanismos de autenticação e controle de acesso utilizados pelos provedores.

Ao concluir esta unidade, você será capaz de avaliar os custos envolvidos na hospedagem de aplicações em um provedor de computação em nuvem, assim como as formas de acesso e controle de permissões para gerenciar as aplicações.

Bons estudos!

Tecnologias de suporte à nuvem

Diálogo aberto

Como já discutimos, a computação em nuvem não é uma tecnologia específica, mas sim um modelo de negócio no qual recursos de TI são ofertados como serviços. Porém, quais são as novas tecnologias ou quais as tecnologias existentes que evoluíram para viabilizar os serviços em nuvem? Nesta seção, faremos uma discussão sobre as tecnologias fundamentais que permitem as características próprias de serviços em nuvem, como a elasticidade rápida e o amplo acesso.

Vamos retomar o caso da empresa de comércio eletrônico em que você é analista de TI responsável pela migração do sistema de vendas para um ambiente de nuvem pública. O sistema consiste em uma aplicação web cuja implementação fez uso de um framework que já foi descontinuado. Isso significa que não há novas atualizações desse framework e, portanto, ele caiu em desuso. O sistema inclui também um banco de dados e requer integração com uma aplicação de terceiros credenciada para efetivação dos pagamentos eletrônicos. Como líder da equipe, seu papel é definir a estratégia de migração do sistema para a nuvem. Você deve escolher o modelo de serviço e as tecnologias para hospedar a aplicação. Nesse contexto, surgem questões importantes como: qual é o melhor modelo de serviço para a aplicação web de comércio eletrônico? Como lidar com a dependência em relação a um sistema legado, isto é, ao framework escolhido quando do desenvolvimento da aplicação? Como essas escolhas podem afetar a replicação do sistema de comércio eletrônico para questões de escalabilidade e desempenho? Sua responsabilidade é elaborar um parecer técnico sobre a estratégia inicial de migração. Você deve entregar um documento escrito que determina qual deve ser o modelo de serviço e a tecnologia de virtualização adotados para implantação da aplicação no ambiente de computação em nuvem. O documento deve conter uma justificativa técnica para a escolha.

A fim de ajudá-lo a resolver esse desafio, esta seção vai abordar a evolução das tecnologias que contribuíram para o advento da computação em nuvem. Vamos discutir modelos de computação distribuída predecessores dos serviços sob demanda na nuvem. Em seguida, descreveremos as duas principais abordagens de virtualização: máquinas virtuais e contêineres. Por fim, vamos ressaltar o papel das tecnologias web no sentido de facilitar a integração entre aplicações distribuídas. A compreensão da evolução dessas tecnologias é muito importante para avaliar e comparar as soluções oferecidas

pelos provedores. Um estudo minucioso desses tópicos vai lhe proporcionar as condições necessárias para resolver o problema proposto. Bom estudo!

Não pode faltar

Computação em nuvem

Atualmente, a computação em nuvem tem um papel fundamental na indústria de TI, a expectativa é que o mercado de serviços em nuvem continue crescendo nos próximos anos. O estudo reportado na revista Forbes (COLUMBUS, 2019) estima um crescimento médio mundial de 12,6% ao ano até 2022, com destaque para serviços IaaS, cuja previsão de crescimento para 2019 é de 27,5% alcançando um volume total de quase 40 bilhões de dólares.

Contexto histórico

Você sabe como foi a evolução até a consolidação do paradigma de computação em nuvem? Dada a sua importância no mercado de TI, é importante entender as tecnologias que viabilizaram os serviços nessa área.

- **1990:** Conforme breve histórico descrito em (ERL; PUTTINI; MAHMOOD, 2013), as primeiras iniciativas disponíveis na Internet, para provisionamento remoto de recursos de TI como serviços, surgiram no final dos anos 90. O primeiro caso de sucesso foi o sistema de CRM (*Customer Relationship Management*) (SALESFORCE, 2019) da empresa Salesforce, caracterizado como o primeiro serviço SaaS de sucesso para o mundo corporativo.
- **2002:** Alguns anos depois, em 2002, a Amazon lançaria a plataforma *Amazon Web Services* (AWS) (AMAZON, 2019) que, atualmente, é a líder no mercado de provedores de nuvens públicas para serviços IaaS (DIGNAN, 2019).
- **2006:** O termo computação em nuvem foi usado comercialmente somente em 2006, com o lançamento do serviço IaaS da AWS denominado EC2 (*Elastic Compute Cloud*) (AMAZON, 2019b).
- **2009:** O primeiro serviço PaaS de destaque foi o *Google App Engine* criado em 2009 (GOOGLE, 2019).

Além dos aspectos históricos, precisamos compreender as tecnologias precedentes, que viabilizaram os modelos de serviços em nuvem. Nesse contexto, é relevante a definição de computação em nuvem como **uma forma de computação distribuída que introduziu modelos de provisionamento**

remoto de serviços escaláveis (ERL; PUTTINI; MAHMOOD, 2013). Sob essa perspectiva, a computação em nuvem representa uma inovação em relação a outras formas de computação distribuída, entre as quais podemos destacar: computação em cluster e computação em grade.

Cluster

Um *cluster* é um conjunto de computadores similares interconectados por uma rede de alto desempenho (TANENBAUM; STEEN, 2008). O desenvolvimento das tecnologias utilizadas na computação em cluster contribuiriam principalmente para o avanço em mecanismos de balanceamento de carga e de recuperação de falhas que também são usados nos provedores de computação em nuvem. O que caracteriza a computação em cluster é a homogeneidade, já que todos os computadores do *cluster* devem ter a mesma plataforma (sistema operacional e outros componentes de software) (TANENBAUM; STEEN, 2008).

Computação em grade

Por outro lado, a computação em grade consiste em reunir recursos computacionais de diferentes organizações, mesmo que as tecnologias de hardware e software sejam diferentes (TANENBAUM; STEEN, 2008). Dessa forma, uma grade computacional envolve sistemas computacionais heterogêneos e geograficamente dispersos. As tecnologias de computação em grade também influenciaram o desenvolvimento de novos mecanismos que viabilizaram a computação em nuvem, em particular, tecnologias relacionadas com portabilidade das aplicações e gerenciamento de infraestrutura (ERL; PUTTINI; MAHMOOD, 2013).

Virtualização

Os mecanismos usados em *clusters* e grades computacionais serviram de base para viabilizar os serviços em nuvem. Além disso, o crescimento no uso desses serviços também foi fomentado pelo aprimoramento do desempenho das tecnologias de rede, pois o acesso aos serviços é feito através da Internet.

No entanto, você sabe qual foi a evolução tecnológica chave para o advento da computação em nuvem?

Foram os avanços nas tecnologias de virtualização! Como explicado em (ERL, 2013), os softwares de virtualização permitem a criação de múltiplas instâncias lógicas de um recurso computacional de forma que esse recurso possa ser compartilhado entre diversos usuários. O conceito de virtualização

não é recente, mas, somente com os ganhos em termos de desempenho e confiabilidade das ferramentas de virtualização modernas é que foi possível viabilizar características como a elasticidade rápida e self-service sob demanda, próprias dos serviços de computação em nuvem.

Pesquise mais

Tipos de virtualização

O trabalho indicado a seguir descreve tipos de virtualização e apresenta uma comparação de desempenho entre importantes ferramentas de virtualização gratuitas disponíveis no Mercado: VMware ESXi (VMWARE, 2019) e XenServer da Citrix (CITRIX, 2019).

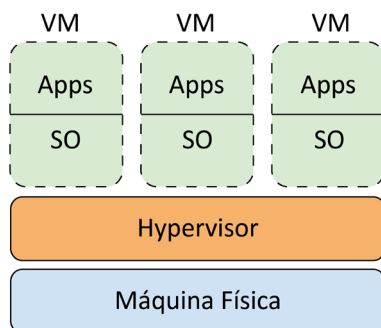
GROSMANN, D. *et al.* Estudo comparativo sobre o uso do VMware e Xen Server na virtualização de Servidores. In: **V Encontro Regional de Computação Ceará, Maranhão e Piauí - ERCEMAPI**, 2011.

As plataformas de computação em nuvem, como o OpenStack, utilizam as ferramentas de virtualização para gerenciar o compartilhamento de recursos computacionais disponíveis na infraestrutura do ambiente de computação em nuvem. Por exemplo, a criação de máquinas virtuais no OpenStack pode ser realizada por meio de diversas ferramentas de virtualização como VMware ESXi, XenServer ou Hyper-V (MICROSOFT, 2019).

Os centros de dados (*data centers*) modernos fazem uso de tecnologias de virtualização para obter eficiência e flexibilidade na administração da infraestrutura de TI. As ferramentas de virtualização, também denominadas Hypervisor, abstraem os recursos computacionais (como servidores e equipamentos de rede) na forma de componentes virtualizados que podem então ser facilmente alocados ou liberados (ERL; PUTTINI; MAHMOOD, 2013). Isso é fundamental para a elasticidade rápida dos ambientes de computação em nuvem.

A virtualização pode ser definida como “uma tecnologia que permite criar vários ambientes simulados ou recursos dedicados a partir de um único sistema de hardware físico” (REDHAT, 2019). Como ilustrado na Figura 2.1, o ambiente lógico criado sobre o equipamento físico é denominado máquina virtual (VM – *Virtual Machine*). A virtualização permite, por exemplo, que um mesmo servidor seja compartilhado por várias aplicações (Apps) de diferentes usuários de forma isolada e segura, pois cada aplicação pode ser executada em uma máquina virtual diferente no mesmo servidor. Observe que cada máquina virtual tem seu próprio sistema operacional (SO), então é possível ter máquinas virtuais com sistemas operacionais diferentes no mesmo servidor.

Figura 2.1 | Máquinas virtuais em um servidor físico



Fonte: elaborada pelo autor.

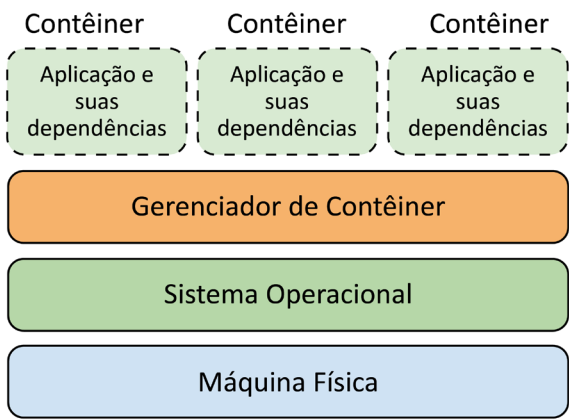
A virtualização viabiliza três fatores fundamentais para a computação em nuvem: independência de hardware, possibilidade de consolidação de servidores e facilidade de replicação de recursos (ERL, 2013).

- **Independência de hardware:** A ferramenta de virtualização abstrai as peculiaridades dos recursos físicos, de forma que problemas de compatibilidade são minimizados. Assim, a migração de uma aplicação em uma máquina virtual não depende das características do hardware do equipamento de destino, desde que o formato da máquina virtual seja suportado pelo hypervisor.
- **Consolidação de servidores:** A consolidação de servidores é um processo para aumentar a taxa de utilização dos servidores em um centro de dados a fim reduzir custos e economizar energia (AHMAD, 2015). Uma das formas de consolidação de servidores é migrar as máquinas virtuais para o menor número possível de servidores. Por exemplo, se existe apenas uma máquina virtual em um servidor, ela pode ser migrada para outro servidor que ainda tem recursos disponíveis para que, o primeiro servidor, agora sem nenhuma máquina virtual, possa ser desligado.
- **Facilidade de replicação:** O terceiro fator importante é a facilidade na replicação das instâncias de máquinas virtuais. Isso decorre do fato de que a máquina virtual é software e pode ser replicada com operações simples de manipulação de arquivos. Assim, é mais fácil instanciar e replicar máquinas virtuais do que servidores físicos.

Contêiner de aplicação

Um modelo alternativo à virtualização baseada em hypervisor é a virtualização baseada em contêiner, que ocorre no nível do sistema operacional (BACHIEGA; SOUZA; BRUSCHI, 2017). Neste caso, um conceito importante é o contêiner de aplicação (*Application Container*) que pode ser entendido como um componente de software autossuficiente, no sentido em que ele encapsula uma aplicação e todas as suas dependências (como bibliotecas, arquivos de configuração, etc.) (SILVA, 2017). Diferentemente de uma máquina virtual, o contêiner não inclui um sistema operacional. Na verdade, o sistema operacional do servidor é compartilhado entre os contêineres em execução. A Figura 2.2 ilustra o esquema de contêineres em um servidor físico. Com uma ferramenta de gerenciamento de contêineres (*container engine*), pode-se instanciar vários contêineres em uma única máquina física. O hardware e o sistema operacional dessa máquina são compartilhados entre os contêineres (SILVA, 2017).

Figura 2.2 | Contêineres em um servidor físico



Fonte: elaborada pelo autor.

Refleta

Os contêineres são amplamente utilizados nos provedores de computação em nuvem, pois a virtualização baseada em contêiner viabiliza mecanismos sofisticados para implantação e replicação de aplicações distribuídas na nuvem. Vários contêineres podem ser rapidamente instanciados em um servidor. É possível que um contêiner com uma aplicação implementada para ambiente Linux possa compartilhar o mesmo servidor que um contêiner com uma aplicação implementada

para ambiente Windows? Quais seriam as alternativas para viabilizar esse cenário?

Uma vez que a imagem de uma máquina virtual inclui um sistema operacional, ela é maior que a imagem de um contêiner que envolve somente software no nível de usuário (aplicações, bibliotecas, etc.). Portanto, em termos de desempenho, o uso de contêineres pode oferecer vantagens em relação ao uso de máquinas virtuais (SILVA, 2017). Por ser mais leve, um contêiner demora menos tempo para inicializar ou para ser transmitido de um servidor para outro. Entretanto, uma vez que os contêineres compartilham o sistema operacional subjacente, o nível de isolamento e segurança é menor, quando comparado com as máquinas virtuais, que emulam um ambiente computacional completo e independente, com seu próprio sistema operacional.

Assimile

É importante salientar que contêineres e máquinas virtuais não são, necessariamente, tecnologias concorrentes. Na verdade, essas tecnologias podem ser utilizadas de forma complementar para compartilhar a infraestrutura dos provedores entre vários clientes. Isso é possível porque um gerenciador de contêineres, como o Docker (DOCKER, 2019c), pode ser instalado em uma máquina virtual, que então poderia suportar a execução de contêineres. Além disso, existem aplicações para as quais a implantação em máquina virtual é mais interessante, por exemplo, devido a requisitos de segurança. Em outros cenários, a implantação de contêineres é mais indicada, por exemplo, no caso da replicação automática de aplicações web em provedores de Plataforma como Serviço (SILVA, 2017; PEREIRA, 2019).

A principal característica do contêiner é permitir que o ambiente de execução da aplicação seja sempre o mesmo, pois tudo que a aplicação precisa está encapsulada no contêiner, independentemente da plataforma subjacente (DOCKER, 2019). Isso favorece a portabilidade e facilita a replicação da aplicação em vários servidores. Basta, para isso, que um gerenciador de contêiner compatível esteja instalado em cada servidor e uma cópia do contêiner seja copiada e executada em cada um deles. Essa agilidade para replicar aplicações ou migrar aplicações de um servidor para outro é muito importante em um ambiente de computação em nuvem, do qual se espera o escalonamento dinâmico de aplicações.

Exemplificando

Os provedores de computação em nuvem usam diversas tecnologias de contêiner de aplicações para hospedar serviços PaaS e SaaS (BERNSTEIN, 2014). Entre as principais tecnologias de contêineres, podemos citar o LXC (CANONICAL, 2019) e o Docker (DOCKER, 2019). A plataforma Docker se tornou a mais popular para virtualização baseada em contêiner. Um contêiner Docker pode ser definido como “uma unidade de software padronizada” (DOCKER, 2019). A plataforma usa o conceito de imagem, a partir da qual o contêiner é instanciado. A imagem é um arquivo que empacota tudo o que é necessário para executar uma aplicação, por exemplo, código executável, bibliotecas, ferramentas de sistema, configurações, etc. A plataforma Docker oferece, ainda, ferramentas como o Docker Compose para serviços que resultam da composição de vários contêineres; e o Docker Swarm, para gerenciamento de clusters de contêineres distribuídos por várias máquinas físicas. Além disso, o Docker Hub é considerado o maior repositório mundial de imagens de contêineres (DOCKER, 2019b). Nesse repositório, podem ser encontradas imagens de diversas aplicações, ferramentas, SGBDs, entre outros.

A evolução das tecnologias de rede e virtualização foram fundamentais para a consolidação dos provedores de computação em nuvem. Essas tecnologias viabilizaram as principais características dos serviços em nuvem, como:

- Self-service sob demanda.
- Acesso remoto.
- Elasticidade rápida.

Arquitetura Orientada a Serviços (SOA)

Já para a característica de amplo acesso, um fator relevante foi o desenvolvimento de Arquitetura Orientada a Serviços (SOA – *Service Oriented Architectures*) (ERL; PUTTINI; MAHMOOD, 2013). Essa arquitetura consiste em decompor as funcionalidades de um sistema em serviços que podem ser reutilizados (CONCEIÇÃO, 2014). O objetivo principal desse estilo arquitetural é promover a interoperabilidade entre aplicações. Neste caso, os serviços devem ser especificados de forma abstrata, sem dependências em relação a plataformas ou linguagens de programação.

Além disso, as requisições aos serviços devem ser feitas por meio de tecnologias e padrões abertos. Neste contexto, foram introduzidos os

Serviços Web (WS – *Web Services*). Esses serviços não são aplicações Web para usuários finais. Eles são componentes de software cujas funcionalidades podem ser invocadas por outras aplicações por meio de requisições HTTP. Os dois modelos principais de Serviços Web são: **SOAP Web Services** e **RESTful Web Services** (ERL; PUTTINI; MAHMOOD, 2013). O primeiro modelo representou a primeira geração de serviços Web, em que as requisições aos serviços eram especificadas conforme o protocolo denominado SOAP (Simple Object Access Protocol) e encapsuladas em mensagens HTTP. As mensagens de requisição SOAP, a descrição dos serviços e a representação dos dados são todas baseadas em esquemas XML (*Extensible Markup Language*) (BECKER, 2001).

Diferentemente do SOAP, que é um protocolo, o REST (REpresentational State Transfer) é um estilo arquitetural para sistemas distribuídos (RODRIGUES, 2014). Os serviços web que seguem os princípios e as restrições REST são então denominados RESTful Web Services. Entre as restrições REST (DAL MORO; DORNELES; REBONATTO, 2011), podemos destacar: o sistema deve seguir o modelo cliente-servidor; não há manutenção de informações de estado por parte do servidor (*stateless*) (TANENBAUM; STEEN, 2008, p. 54) e o sistema deve usar uma interface uniforme (padronizada) para acesso aos recursos disponíveis. O conceito de recurso é fundamental neste modelo de serviço web. Como discutido em (SALVADORI, 2015) sobre os conceitos e princípios do REST, um recurso é qualquer informação, acessível pelo serviço web, que pode ser endereçada através de um identificador padronizado (URI – *Uniform Resource Identifier*) (TANENBAUM; STEEN, 2008, p. 344). Os recursos devem ser representados em um formato textual, sendo o mais popular o JSON (*JavaScript Object Notation*) (SALVADORI, 2015). Para manipulação dos recursos, são utilizados os métodos padronizados no protocolo HTTP. Os principais métodos são descritos no Quadro 2.1.

Quadro 2.1 | Principais métodos HTTP usados em RESTful Web Services

Método	Descrição
GET	Obter a representação textual de um recurso.
POST	Criar um recurso com a representação textual enviada.
PUT	Atualizar os dados de um recurso com a representação textual enviada.
DELETE	Apagar um recurso.

Fonte: adaptado de Rodrigues (2014, p.173).

O estilo arquitetural REST condiciona o uso de tecnologias Web abertas, como o HTTP e o JSON, e o uso de interface padronizada para os serviços, a

partir da URI que identifica cada recurso e os métodos HTTP utilizados para manipular esses recursos. Esse modelo favorece a interoperabilidade entre as aplicações e a independência em relação a tecnologias e plataformas específicas. Esses aspectos são muito importantes para serviços de computação em nuvem, principalmente no que diz respeito à característica de amplo acesso. Assim, os RESTful Web Services são uma das principais formas utilizadas pelos provedores de nuvem pública para disponibilizar recursos de TI como serviço para aplicações clientes.

Caro aluno, finalizamos aqui esta seção que introduz nosso estudo sobre tecnologias e soluções usadas em provedores de computação em nuvem. Foram discutidas as principais tecnologias que viabilizaram o surgimento dos serviços em nuvem, como as ferramentas de virtualização baseadas em hypervisor e os modelos de Serviços Web para integração de aplicações. No decorrer desta unidade, ainda vamos estudar as métricas de custos de serviços em nuvem e os aspectos envolvidos na migração e hospedagem de aplicações em um provedor de nuvem pública.

Sem medo de errar

Como analista de TI da empresa de comércio eletrônico, você deve coordenar a migração do sistema para um provedor de nuvem pública. Precisamos recordar que o sistema está em operação há alguns anos e o banco de dados já tem um significativo volume de dados. Além disso, existe a questão de framework, já obsoleto, sobre o qual a aplicação web foi desenvolvida. Outro requisito importante é que a migração foi motivada não só pela redução de custos, mas também pela possibilidade de melhorar a escalabilidade e a confiabilidade do sistema. Sua tarefa é justificar a escolha do modelo de serviço e da tecnologia para gerenciamento das instâncias da aplicação no provedor se for necessário.

Vimos na unidade anterior alguns modelos de serviço como: Infraestrutura como Serviço (IaaS), Plataforma como Serviço (PaaS), Software como Serviço (SaaS) e Banco de Dados como Serviço (DBaaS).

Em primeiro lugar, podemos assumir que o banco de dados do sistema de comércio eletrônico seria migrado para um serviço DBaaS. A outra opção seria contratar um servidor virtual, no modelo IaaS, e migrar o banco de dados para esse servidor. No entanto, essa opção envolve a responsabilidade de instanciar o servidor, instalar, configurar e manter o SGBD correspondente, o que não ajuda na redução de custos. Como há um serviço especializado, ele acaba sendo a melhor opção, pois dispensa gerenciamento e já é customizado para o propósito específico em questão.

Por outro lado, a migração da aplicação web envolve muitas alternativas. A princípio, podemos descartar o modelo SaaS, pois aqui não há necessidade de algum software disponível no provedor, mas sim de um ambiente para hospedagem do software da própria empresa. Podemos então escolher entre os modelos IaaS e PaaS. Os dois modelos são factíveis, então é necessário ponderar sobre as vantagens e desvantagens de cada um.

Vejamos o caso do modelo PaaS. A grande vantagem desse tipo de serviço é que ele oferece um ambiente completo pronto para o desenvolvimento ou implantação de aplicações, sem a necessidade de gerenciamento de servidores. Basta enviar a aplicação para o provedor. Os mecanismos de instanciação, replicação e balanceamento de carga são automáticos. No entanto, como a aplicação web da empresa de comércio eletrônico usa uma tecnologia obsoleta, pode ser pequeno o número de provedores no mercado com plataformas que suportem essa tecnologia. Isso representa um risco, haveria poucas alternativas para mudar de provedor, caso necessário. Existe inclusive a possibilidade de nenhum provedor oferecer plataforma com o suporte à tecnologia em questão. Portar a aplicação para uma tecnologia moderna também não é uma solução viável pois grande parte do código da aplicação teria que ser reescrito. Em geral, esse tipo de tarefa leva muito tempo e tem custos altos.

O modelo IaaS então é o que pode atender melhor a empresa no cenário descrito. Com maior nível de controle desse tipo de serviço, a equipe de TI pode configurar o ambiente necessário para executar a aplicação Web de comércio eletrônico. No entanto, a empresa seria a responsável pelo gerenciamento da infraestrutura. Neste caso, é muito importante fazer uma escolha adequada em relação ao modelo de virtualização utilizado para implantar o ambiente de execução da aplicação web. O uso de máquinas virtuais é importante para aplicações cujo requisito principal envolve segurança e sigilo. Seria o caso de hospedar o banco de dados, se não houvesse um serviço especializado. Para o caso da aplicação web de comércio eletrônico, o mais apropriado é uma infraestrutura com virtualização baseada em contêiner. Dessa forma, as dependências da aplicação web podem ser encapsuladas no contêiner. Por ser mais leve os contêineres também facilitariam o trabalho e o custo de replicação de instâncias da aplicação, o que atende aos requisitos de escalabilidade e confiabilidade. As operações com contêineres são mais rápidas em comparação com máquinas virtuais. Isso é importante para aplicações web como neste caso que podem ter um acréscimo na demanda repentinamente, por exemplo devido a períodos de compras ou promoções. As ferramentas de gerenciamento de contêiner podem escalar rapidamente o número de instâncias para o sistema reagir de forma satisfatório ao aumento na demanda usuá-rios.

Com esta análise, você já deu um primeiro passo na estratégia de migrar o sistema. Conhecendo o modelo de serviço, a tecnologia de virtualização e as demandas de armazenamento de dados, será então possível trabalhar em estimativas de custo. Outro ponto importante será definir políticas de controle de acesso para a equipe de TI responsável pela manutenção do sistema que acessará os recursos remotamente.

Avançando na prática

Nível de virtualização

Uma empresa vai lançar um novo aplicativo para gestão de tarefas colaborativas. O objetivo é promover aumento de produtividade para pequenas empresas e profissionais liberais. O sistema inclui um serviço web e um banco de dados no qual são compartilhados os dados das tarefas. O aplicativo se conecta no serviço web por meio de requisições HTTP para manipular os recursos do sistema. A empresa está com grande expectativa de sucesso. Para atender muitos clientes, já escolheu um grande provedor de nuvem pública para hospedar o sistema. Você foi contratado como consultor para definir a melhor estratégia de implantação do serviço web e do banco de dados nesse provedor específico que oferece apenas serviços nos modelos PaaS e IaaS. Você precisa determinar qual a melhor opção para implantação do serviço web e do banco de dados em termos do modelo de serviço e tecnologia de virtualização.

Resolução da situação-problema

O Quadro 2.2 apresenta um resumo comparativo das tecnologias de virtualização. A máquina virtual é considerada um modelo de virtualização ao nível de hardware. Já o contêiner é considerado um modelo de virtualização ao nível de sistema operacional. Por isso, a imagem da máquina virtual é muito grande comparada com a do contêiner, pois a imagem da máquina virtual precisa conter seu próprio sistema operacional. Por outro lado, os contêineres compartilham o sistema operacional da máquina física no qual executam. A desvantagem é que a aplicação incluída no contêiner tem que ter sido implementada para o mesmo sistema operacional no qual o gerenciador de contêiner está executando.

	Máquina Virtual	Contêineres
Abordagem	Virtualização no nível de hardware	Virtualização no nível de sistema operacional
Denominação da ferramenta de virtualização	Hypervisor	Container Engine
Vantagens	<ul style="list-style-type: none">• Maior nível de isolamento e segurança.• Permite mais de um sistema operacional no mesmo hardware.	<ul style="list-style-type: none">• Permite compartilhamento do sistema operacional.• Mais “leve” (exige menos recursos e ocupa menos espaço).

Fonte: elaborado pelo autor.

Vamos avaliar cada componente do sistema para determinar qual dos dois modelos de virtualização é mais adequado. O banco de dados pode conter dados sobre os quais os clientes desejam privacidade, portanto proteção e isolamento dos dados é um critério relevante. Além disso, o banco de dados possui uma ou apenas algumas réplicas, não há necessidade de escalar para um número elevado de instâncias dos SGBD. Portanto, a melhor estratégia é implantar o banco de dados em uma máquina virtual alocada no modelo IaaS.

O outro componente do sistema é o serviço web, que pode necessitar de replicação em escala para balanceamento de carga. O serviço web se beneficia de tecnologias abertas e padronizadas disponíveis em plataformas pré-configuradas na maioria dos provedores, sem necessidade de gerenciamento de infraestrutura. Nesse caso, a opção mais adequada para implantar o serviço web seria um serviço PaaS baseado em contêiner para agilizar a replicação das instâncias.

Faça valer a pena

1. A consolidação do modelo de computação em nuvem foi viabilizada por uma série de avanços em diversas tecnologias de redes e sistemas distribuídos. Nesse contexto, considere as afirmativas a seguir:

- I. A evolução das tecnologias de transmissão de dados melhorou a qualidade do acesso à internet, o que beneficiou o acesso remoto a serviços de computação em nuvem.

- II. A escalabilidade de serviços em nuvem foi possível na medida em que a virtualização baseada em contêineres substituiu a virtualização baseada em hypervisor.
- III. A especificação de modelos de serviços baseados em tecnologias web favoreceu os mecanismos de integração de aplicações em nuvem.

Considerando o texto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. I e II, apenas.
- c. I e III, apenas.
- d. II e III, apenas.
- e. I, II e III.

2. A computação em nuvem oferece vários benefícios como a abstração da complexidade de gerenciamento de infraestrutura. No caso dos modelos de serviços SaaS e PaaS, o cliente do serviço não precisa se preocupar com gerenciamento de servidores e ainda pode contar com mecanismos de replicação automática de aplicações.

Dentre as evoluções tecnológicas que viabilizaram os serviços de computação em nuvem, a que contribui para a abstração de complexidade é

- a. O aprimoramento dos mecanismos de balanceamento de carga em clusters.
- b. Interfaces padronizadas por meio de tecnologias web abertas.
- c. A melhoria no desempenho das redes de acesso à internet.
- d. O aprimoramento da segurança e desempenho das tecnologias de virtualização.
- e. Novas técnicas de consolidação de servidores.

3. Os dois principais modelos de virtualização são: virtualização baseada em hypervisor e virtualização baseada em contêiner. Ambos os modelos podem ser usados para provisionamento dinâmico de recursos em ambientes de computação em nuvem. Entretanto eles têm características muito diferentes.

Em relação aos modelos de virtualização, assinale a alternativa correta.

- a. As máquinas virtuais compartilham o sistema operacional da máquina física no qual estão sendo executadas.

- b. O processo de inicializar um contêiner é mais lento que o processo de inicializar uma máquina virtual.
- c. O uso de contêiner provê um maior nível de proteção aos dados de uma aplicação, quando comparado com o uso de máquina virtual.
- d. Um contêiner com aplicação implementada para sistema Windows pode ser executado em uma máquina física com sistema Linux.
- e. Contêineres e máquinas virtuais podem conter mais de uma aplicação.

Provedores de computação em nuvem

Diálogo aberto

Caro aluno, vamos começar mais uma etapa de nosso estudo sobre as tecnologias e soluções empregadas nos provedores de computação em nuvem. Um dos aspectos mais importantes no momento de criar uma aplicação na nuvem ou de migrar para a nuvem uma aplicação já existente é a estimativa dos custos dos serviços. Você conhece os modelos utilizados pelos provedores para determinar os preços dos serviços? Já aprendemos que uma das características da computação em nuvem é o pagamento baseado no uso. Mas quais são as métricas utilizadas para contabilizar uso dos recursos de TI?

Para entender melhor esses tópicos, vamos voltar ao cenário descrito no início da Unidade 2. Você é um analista de TI em uma empresa que mantém uma aplicação web de comércio eletrônico. Sua responsabilidade é coordenar o processo de migração dessa aplicação para um provedor de nuvem pública, com isso, a empresa espera reduzir os custos e incrementar a escalabilidade da aplicação. Na seção anterior, vimos que a melhor solução seria usar serviço no modelo IaaS, com virtualização baseada em contêiner para a aplicação web. O próximo passo é você escolher um provedor para o qual migrar a aplicação. O critério de escolha será o custo. Então, você deve elaborar um documento no qual especifica os serviços necessários para migrar a aplicação e mostrar qual provedor oferece os melhores preços para esse caso. Considere que o banco de dados da empresa é um banco MySQL com um volume total de dados de 4 TB. Para executar o banco de dados, será necessária uma máquina virtual alocada sob demanda com 4 núcleos de processamento e 16 GB de memória. Além disso, a aplicação web precisa ser executada em um contêiner alocado sob demanda com 2 núcleos de processamento e 8 GB de memória. Você deve calcular o custo médio mensal para hospedar o sistema em pelo menos dois provedores entre os líderes de mercado. Tanto a aplicação como o banco de dados devem ser hospedados em datacenter no território brasileiro. Desconsidere os custos relacionados com a transmissão de dados e requisições.

- Para resolver essas questões, é preciso entender como os provedores determinam os custos dos diversos tipos de serviços. Esta seção:
- Iniciará com exemplos entre os principais provedores de computação em nuvem no mundo.

- Em seguida, serão abordadas as métricas de custo e os modelos de tarifação.
- Será concluída com uma breve discussão sobre opções de treinamento e certificação para profissionais que atuam na área de computação em nuvem.

Ao concluir esta etapa, você será capaz de estimar os custos envolvidos no uso de serviços em Nuvem. Bom estudo!

Não pode faltar

Caro aluno, na seção anterior, aprendemos sobre as principais tecnologias que viabilizaram os serviços de computação em nuvem, assim como os modelos de serviço e de implantação. Vimos que uma empresa pode implantar sua própria nuvem privada utilizando uma plataforma de computação em nuvem como o OpenStack. Essa abordagem permite à empresa um controle completo sobre o ambiente, mas implica um investimento inicial alto para aquisição e implantação da infraestrutura, além da responsabilidade e dos custos adicionais de gerenciamento e operação do ambiente computacional. Como discutido anteriormente, a alternativa para maior redução de custos é optar pelo modelo de nuvem pública, no qual o provedor é responsável pela implantação e pela manutenção da infraestrutura de recursos computacionais.

- Você sabe quais são os principais provedores no mercado de computação em nuvem?
- Conhece os custos e os modelos de tarifação empregados por esses provedores?

Agora, vamos aprender os principais fatores que determinam o custo dos serviços de computação em nuvem, assim como as principais métricas utilizadas na precificação desses serviços. Além disso, vamos descrever a política dos provedores para que o mercado possa reconhecer os profissionais habilitados no projeto e desenvolvimento de soluções de computação em nuvem.

Provedores

Considerando os provedores de nuvem pública de serviços de infraestrutura e plataforma (IaaS e PaaS), podemos mencionar a AWS (Amazon Web Services), o Microsoft Azure e o Google Cloud Platform (RANGER, 2019). Em se tratando de soluções corporativas de software como serviço (SaaS), podemos citar a Microsoft Salesforce, Adobe, Oracle e SAP (COLUMBUS, 2018). É importante mencionar que a Salesforce, pioneira em software corporativo em nuvem, também mantém o Heroku, um provedor PaaS, assim como o Openshift da Red Hat e o Cloud Foundry, que é uma plataforma livre originalmente criada

pela VMWare e hoje mantida pela Cloud Foundry Foundation. O Quadro 2.3 apresenta um resumo dos principais provedores e respectivos serviços.

Quadro 2.3 | Principais provedores de serviços de computação em nuvem

Modelo de Serviço	Provedor	Exemplo de serviço ou produto
IaaS	AWS (AWS, 2019c)	AWS Elastic Cloud Computing (EC2): máquinas virtuais (AWS, 2019e).
	Microsoft Azure (AZURE, 2019)	Azure Virtual Machines: máquinas virtuais (AZURE, 2019c).
	Google Cloud Platform (GOOGLE, 2019)	Google Compute Engine: máquinas virtuais (GOOGLE, 2019c).
PaaS	AWS	AWS Elastic Beanstalk: ambiente gerenciado para desenvolvimento e implantação de aplicações (AWS, 2019d).
	Microsoft Azure	Azure App Service: ambiente gerenciado para desenvolvimento e implantação de aplicações (AZURE, 2019b).
	Google Cloud Platform	Google App Engine: ambiente gerenciado para desenvolvimento e implantação de aplicações (GOOGLE, 2019b).
	Salesforce (SALESFORCE, 2019)	Heroku Platform: ambiente gerenciado para desenvolvimento e implantação de aplicações (HEROKU, 2019),
	Red Hat OpenShift (OPENSIFT, 2019)	Red Hat OpenShift: ambiente gerenciado para desenvolvimento e implantação de aplicações.
	Cloud Foundry (CLOUD FOUNDRY, 2019)	Cloud Foundry: uma plataforma baseada em contêineres, distribuída como software livre, para desenvolvimento e implantação de aplicações.
SaaS	Microsoft Office 365 (MICROSOFT, 2019)	Microsoft Office 365: aplicativos de produtividade e colaboração.
	Salesforce	Salesforce Essentials: aplicações de CRM (Customer Relationship Management) (SALESFORCE, 2019b).
	Adobe Experience Cloud (ADOBE, 2019)	Adobe Experience Cloud: aplicações para marketing corporativo.
	Oracle Cloud (ORACLE, 2019)	Oracle Cloud Applications: aplicações para gestão corporativa (ORACLE, 2019b).
	SAP Cloud Platform (SAP, 2019)	Cloud ERP (Enterprise Resource Planning): aplicações para gestão corporativa (SAP, 2019b).

Fonte: elaborado pelo autor.

Gerenciamento de custos

Em geral, aplicações em nuvem (SaaS) são utilizadas por usuários finais. A cobrança por esses serviços é, tipicamente, na forma de uma assinatura, com pagamento mensal ou anual. Por outro lado, serviços IaaS e PaaS são utilizados por empresas ou profissionais de TI e o cálculo do preço associado a esses serviços envolve diversas variáveis (SOUSA; MOREIRA; MACHADO, 2009). Pesquisas de mercado recentes mostram que o principal desafio no uso de serviços em nuvem é o gerenciamento de custos (DIGNAN, 2019), sendo isso uma maior preocupação até em relação aos aspectos de segurança.

O gerenciamento dos custos de serviços em nuvem é uma tarefa difícil, uma vez que o cálculo do custo total envolve várias métricas e os preços dos recursos computacionais podem variar de acordo com a localização da infraestrutura ou a forma de tarifação (RODAMILANS, 2014). Além disso, as empresas podem utilizar serviços de diferentes provedores para implementar uma solução, o que caracteriza um cenário denominado *multicloud* (PETCU, 2013). Assim, mesmo que cada provedor ofereça uma ferramenta de gerenciamento de custos, faz-se necessária uma solução adicional centralizada para agregar os custos dos diferentes provedores.

Exemplificando

A dificuldade de monitoramento de custos de serviços em nuvem, principalmente para clientes com serviços em diversos provedores, motivou o desenvolvimento de soluções de gerenciamento de custos de nuvem (*Cloud Cost Management*). Entre as principais soluções nesse segmento, pode-se destacar as seguintes plataformas: CloudHealth (VMWare, 2019) e Cloudability (CLOUDABILITY, 2019). Além do monitoramento centralizado dos gastos, essas plataformas permitem também a otimização do uso dos recursos, definição de orçamentos (limites de gastos, por exemplo, para cada tipo de recurso), análise de despesas, mecanismos para alinhamento contábil com políticas de governança corporativa, etc.

Uma das características dos serviços de computação em nuvem é o pagamento baseado no uso. Na prática, existem diversas métricas que são utilizadas para se calcular os custos para o tempo de uso de cada tipo de recurso computacional. Em geral, três métricas básicas são considerados na determinação do custo de serviços em nuvem:

Uso de recursos de processamento: O uso de recursos de processamento é uma métrica calculada a partir do tempo de uso de núcleos de

processamento de máquinas virtuais. Em geral esse custo é proporcional à quantidade de memória RAM.

Uso de recursos de armazenamento: Essa métrica serve para tarifar o espaço utilizado para armazenamento persistente de dados e, em geral, é uma cobrança mensal para cada gigabyte de dados armazenados.

Uso de recursos de transmissão de dados (ERL; PUTTINI; MAHMOOD, 2013): Essa métrica serve para tarifar os dados transferidos entre a rede do provedor e outras redes. Como a transmissão pode ocorrer nos dois sentidos, o provedor pode definir preços diferentes para tráfego de saída e de entrada. Assim, se você aloca uma máquina virtual, além de pagar pelo uso da máquina e do armazenamento, também poderá ser cobrado pelos dados que foram enviados para essa máquina ou dessa máquina para outro computador qualquer. O Quadro 2.4 apresenta um resumo das métricas básicas para tarificação de serviços de computação em nuvem.

Quadro 2.4 | Principais provedores de serviços de computação em nuvem

Métrica	Descrição	Tipo de cobrança	Exemplo
Uso de recursos de processamento.	Medida da capacidade de processamento alocada. Representa o pagamento pelo uso de núcleos de processamento e memória RAM.	Por tempo (em geral, por hora).	R\$ 0,80 por hora de uso de máquina virtual com 2 núcleos de processamento e 4 GB de memória RAM
Uso de recursos de armazenamento.	Medida da capacidade de armazenamento alocada. Representa o pagamento pelo uso espaço de memória secundária (armazenamento persistente).	Por tempo (em geral, por mês).	R\$ 0,60 por mês para cada 1 GB de dados armazenados em uma unidade de armazenamento (disco virtual).
Uso de recursos de transmissão de dados.	Medida do volume de dados transferidos. Representa o pagamento pela transferência de dados para ou da rede do provedor.	Por evento (em geral, por GB transferido).	R\$ 0,05 para cada 1 GB de dados transferido de/para uma máquina virtual.

Fonte: elaborado pelo autor.

Para cada tipo de serviço, o provedor pode combinar as três métricas para determinar o custo final do serviço. Vejamos o caso do modelo IaaS, no qual um cliente pode alocar uma máquina virtual no provedor.

Modelos IaaS

Nesse caso, o custo depende do tipo de instância da máquina virtual e não há uma padronização das instâncias entre os provedores (RODAMILANS, 2014). Em geral, os tipos de instâncias de máquinas virtuais são caracterizados em função do número de núcleos de processamento e do tamanho da memória RAM. Por exemplo, um provedor pode tarifar o cliente em R\$0,50 por hora de uso de uma máquina virtual com 1 núcleo de processamento e 2 GB de memória e em R\$1,00 por hora de uso de uma máquina com 1 núcleo de processamento e 2 GB de memória. Além da capacidade de processamento, o custo de uso da máquina virtual pode incluir custos relacionado ao uso de recursos de armazenamento e de transmissão de dados. Por exemplo, o provedor pode tarifar o cliente em R\$ 0,30 por GB de dados armazenados em disco por mês ou em R\$0,15 para cada 1 GB de dados transferidos pela rede. Alguns provedores não cobram a transmissão de dados entre máquinas virtuais dentro da rede do provedor. Nesse sentido, são tarifadas apenas as transferências de dados entre uma máquina no provedor e redes externas.

E no caso de serviços PaaS? Você conhece como são calculados os custos?

Modelo PaaS

Nesse caso, o cliente faz uso de um ambiente customizado para o desenvolvimento e execução de aplicações. Em geral, a cobrança é feita proporcionalmente aos recursos computacionais alocados para o ambiente. Por exemplo, o provedor pode cobrar do cliente PaaS o correspondente ao uso de espaço de armazenamento para dados e código da aplicação e também o tempo de uso da máquina virtual ou contêiner onde a aplicação estiver sendo executada. Enfim, o cálculo detalhado e preciso dos custos associados a um serviço de Computação em Nuvem é uma tarefa complexa. Para lidar com isso, muitos provedores disponibilizam aplicações que permitem calcular estimativas de custos para seus serviços (AZURE, 2019d; GOOGLE, 2019d; AWS, 2019f).

Assimile

Além das métricas básicas, associadas aos recursos computacionais da infraestrutura de um provedor, existem também métricas adicionais, associadas ao uso de componentes de software. Nesse caso, é usual a

precificação por evento, e não pelo tempo de uso. Assim, um provedor poderia tarifar o cliente pelo número de operações de escrita em um SGBD ou pelo número de requisições a uma aplicação web. Por exemplo, um provedor poderia cobrar do cliente R\$ 2,00 para cada 1.000 requisições a um serviço web.

Em geral, os custos envolvidos no gerenciamento de uma infraestrutura de TI podem ser divididos em duas categorias: os custos iniciais (*up-front costs*) e os custos recorrentes (*on-going costs*) (ERL; PUTTINI; MAHMOOD, 2013).

Os **custos iniciais** correspondem ao capital inicial para adquirir os recursos computacionais e montar a infraestrutura. Os custos **recorrentes** são aqueles relacionados ao gerenciamento e à operação da infraestrutura ao longo do tempo. Se uma empresa escolhe manter sua própria infraestrutura, então os custos iniciais são altos. Para reduzir esses custos, a empresa pode optar por alocar os recursos em ambiente de nuvem, pois a infraestrutura já está disponível pelo provedor e pode ser compartilhada de forma segura entre vários clientes. Porém, os custos iniciais para uso de um ambiente em nuvem não são nulos, pois ainda é necessário o investimento em redes de acesso e treinamento ou contratação de profissionais especializados. Ainda assim, os custos iniciais para ambiente de nuvem pública são menores.

Por outro lado, os custos recorrentes podem ser mais altos no ambiente de nuvem, pois o provedor será remunerado pela manutenção da infraestrutura. Não existe uma resposta definitiva sobre qual abordagem é mais vantajosa em termos de custo. Nesse sentido, cada caso deve ser analisado individualmente, de acordo com as necessidades de cada empresa, o período pelo qual os recursos serão utilizados e os preços praticados pelos provedores. É importante, entretanto, observar que o crescimento do mercado de nuvem pública mostra que essa é uma opção atraente em termos financeiros. Além disso, existe uma série de custos indiretos envolvidos na manutenção de uma infraestrutura própria, como a depreciação e subutilização dos equipamentos, gastos com energia elétrica, etc.

Pesquise mais

No artigo indicado a seguir, é feito um estudo que compara o custo da aquisição de servidores físicos com o custo do uso de máquinas virtuais em um provedor de nuvem pública. O cenário hipotético apresentado ilustra diversas variáveis envolvidas na análise dos custos, por exemplo, a taxa de utilização dos servidores.

- SOUSA, F. R. C.; MOREIRA, L. O.; MACHADO, J. C. **Computação em nuvem: Conceitos, tecnologias, aplicações e desafios**. II Escola

Mecanismos de monitoramento

Em resumo, os provedores tarificam o acesso a software como serviço por meio de plano de assinaturas ou em função do número de requisições e tarificam plataforma ou infraestrutura como serviço em função do tempo de uso de recursos computacionais. De qualquer forma, os provedores precisam implementar mecanismos de monitoramento de uso dos serviços em nuvem (ERL; PUTTINI; MAHMOOD, 2013). **Esses mecanismos servem para contabilizar de forma minuciosa as requisições e tempo de alocação de recursos por parte dos clientes.** As informações resultantes são utilizadas principalmente para determinar o valor a ser cobrado de cada cliente, e também servem para análise e otimização da utilização dos recursos.

Em geral, os provedores de computação em nuvem oferecem uma interface web por meio da qual os clientes podem gerenciar o uso dos serviços disponíveis. Essas ferramentas também incluem várias funcionalidades relacionadas com o gerenciamento de custos como: consulta do preço dos serviços utilizados; verificação dos pagamentos já realizados e relatório detalhado de gastos; estimativa das próximas faturas; definição de limites e alerta de gastos. Para isso, os provedores precisam elaborar modelos de tarificação (ERL; PUTTINI; MAHMOOD, 2013). Esses modelos determinam custos unitários para uso dos recursos de TI disponíveis no provedor. O modelo consiste em métricas e custos padronizados que especificam como o custo de um serviço em nuvem deve ser medido e calculado. Não há um padrão comum entre os provedores; as métricas e preços variam muito entre provedores. Alguns provedores possuem taxas variáveis em função da quantidade de recursos alocados, com descontos para grandes volumes de utilização. Por exemplo, a AWS oferece desconto na alocação de instâncias de máquinas virtuais se essa alocação for feita com reserva e pagamento antecipado pelos recursos, ao contrário da alocação sob demanda (AWS, 2019).

Outra variável para determinação do preço é a localização dos recursos. Os grandes provedores possuem diversos *datacenters* em países diferentes. O preço de uma máquina virtual em um *datacenter* no Brasil, por exemplo, pode ser diferente do preço de uma máquina virtual em um *datacenter* no Reino Unido, mesmo que o provedor e a configuração da máquina sejam os mesmos. Essas variações nos preços refletem as diferenças de impostos e custos de equipamentos e recursos humanos nas diferentes regiões de operação dos provedores.

Assimile

Um provedor PaaS não precisa ter sua própria infraestrutura. Ele pode alocar infraestrutura de um provedor IaaS e, sobre essa infraestrutura de terceiro, implantar as plataformas que serão oferecidas como serviço a seus clientes. O mesmo vale para um provedor SaaS. Nesse caso, a aplicação oferecida como serviço pode estar em execução sobre a infraestrutura de outro provedor. O provedor que usa infraestrutura de outro pode optar por duas formas de tarifação: integrada ou separada (ERL; PUTTINI; MAHMOOD, 2013). Na tarifação integrada, os custos da infraestrutura estão embutidos no preço final do serviço. Na tarifação separada, o cliente pode identificar as diferenças de valores de acordo com o provedor de infraestrutura envolvido. Por exemplo, o cliente de um serviço PaaS pode inclusive escolher o provedor de infraestrutura no qual será hospedada a plataforma. Esse é o caso do MongoDB Atlas, um provedor de banco de dados MongoDB como serviço. Os preços de um banco de dados na nuvem dependem do servidor IaaS escolhido pelo cliente, pois o provedor MongoDB Atlas não possui infraestrutura própria. O cliente possui três opções de provedor IaaS para hospedar o banco: AWS, Azure ou Google Cloud. Os preços são estipulados de acordo com o provedor escolhido pelo cliente (MONGODB, 2019).

A tarefa de projetar soluções em nuvem ou de migrar soluções existentes para a nuvem inclui também o planejamento dos custos. Um custo relevante para o cliente de serviços em nuvem são os profissionais necessários, especialistas em computação em nuvem.

Especialista em computação em nuvem

O profissional que deseja atuar no desenvolvimento e gerenciamento de aplicações e serviços em nuvem deve conhecer bem não apenas as tecnologias e os modelos de serviço, mas também os modelos de tarifação. Para qualificação dos profissionais, a maioria dos provedores oferece programas de treinamento e certificação em diversas áreas de atuação no contexto do paradigma de computação em nuvem.

- Por exemplo, o provedor AWS estabelece três papéis principais de atuação na área de computação em nuvem: arquiteto de soluções em nuvem, desenvolvedor de software para nuvem e administrador de sistemas (AWS, 2019b).
- O arquiteto é o profissional com habilidades técnicas para elaborar o projeto de sistemas escaláveis e altamente disponíveis na nuvem.

- O desenvolvedor é o profissional capaz de implementar serviços e aplicações eficientes em ambientes de computação em nuvem.
- O administrador de sistemas é o profissional responsável por gerenciar a operação das soluções. Ele deve possuir habilidades para automatizar a implantação de aplicações e configurar sistemas e serviços de rede em nuvem.

Existem também certificações por área de atuação, como IoT, aprendizado de máquina ou segurança, por exemplo. A maioria dos provedores oferece programas de treinamento e exames que certificam as habilidades dos profissionais. Essas certificações podem contribuir para uma carreira de sucesso na área de computação em nuvem.

Refleta

Muitos provedores incentivam o uso de tecnologias proprietárias e também estipulam o preço de transferência de dados de forma que a migração de uma aplicação de um provedor para outro envolva custos muito altos. Isso é uma forma de inibir um cliente de mudar para outro provedor concorrente. Quais seriam as estratégias que as empresas poderiam adotar para lidar com esse problema? Uma vez que não é fácil calcular os preços de serviços em nuvem, será que existem soluções ou serviços para estimativas e comparação de preços entre provedores?

Nesta seção foram apresentados alguns dos principais provedores de computação em nuvem e os conceitos básicos sobre como são determinados os custos dos serviços oferecidos. Com o conteúdo abordado nesta seção, você será capaz de estimar custos de serviços de acordo com os preços vigentes para provedores de interesse.

Sem medo de errar

Calcular os custos associados a serviços em nuvem não é uma tarefa fácil, pois existem muitas métricas e variáveis envolvidas. Grandes empresas podem fazer uso de ferramentas de gerenciamento de custos para acompanhar os gastos com serviços em nuvem, inclusive no caso de serviços em provedores diferentes.

Antes de migrar uma aplicação para um provedor, é importante estimar quais seriam os custos resultantes. Como analista de TI da empresa, você precisa estimar o custo dos serviços necessários para executar o sistema de comércio eletrônico. O sistema consiste em uma aplicação web e um banco

de dados MySQL. A aplicação web precisa ser executada em um contêiner alocado sob demanda com 2 núcleos de processamento e 8 GB de memória. O banco de dados precisa de 4 TB de capacidade de armazenamento e uma máquina virtual alocada sob demanda com 4 núcleos de processamento e 16 GB de memória para executar o SGBD. Já foi determinado anteriormente que o melhor modelo para hospedagem da aplicação web seria o modelo IaaS com virtualização baseada em contêiner. Vamos então analisar o custo médio mensal para migrar esse sistema para um ambiente de nuvem.

Existem diversas soluções para essa questão de acordo com preços praticados pelos provedores. A título de exemplo, vamos considerar dois provedores: Provedor A e Provedor B. Suponha que ambos têm datacenters no Brasil e oferecem instâncias com virtualização baseada em hypervisor e contêiner.

Os serviços necessários são:

- Para executar o banco de dados MySQL:
 - Máquina virtual (VM) sob demanda com 4 CPUs e 16 GB de memória.
 - 4 TB de armazenamento.
- Para executar a aplicação web:
 - Contêiner sob demanda com 2 CPUs e 8 GB de memória.

Considere como apresentado a seguir os preços do Provedor A para cada serviço:

- VM 4 CPUs e 16 GB RAM: US\$ 0,564 por hora = US\$ 406,08 por mês.
- 4 TB de armazenamento magnético: US\$ 0,19 por GB/mês = US\$ 778,24 por mês.
- Contêiner 2 CPUs e 8 GB RAM: US\$ 0,1344 por hora = US\$ 96,768 por mês.

A seguir, os preços praticados para cada serviço pelo Provedor B:

- VM 4 CPUs e 16 GB RAM: US\$ 0,579 por hora = US\$ 416,88 por mês.
- 4 TB de armazenamento magnético: US\$ 0,135 por GB/mês = US\$ 552,96 por mês.
- Contêiner 2 CPUs e 8 GB RAM: US\$ 0,1508 por hora = US\$ 108,576 por mês.

Conforme modelo de tarifação do Provedor A, o custo mensal médio é US\$ 1.281,088. Conforme modelo de tarifação do Provedor B, o custo mensal

médio é US\$ 1.078,416. Portanto, o Provedor B representa a solução com mais baixo custo. Você pode encontrar provedores com custos diferentes do que os dois analisados como exemplo nesta proposta de solução.

Este estudo ilustra as principais métricas de custo para serviços IaaS, com exceção do uso de recursos de transmissão de dados. Os conhecimentos abordados neste estudo de caso são muito importantes para entendimento do complexo processo de estimar custos de serviços em nuvem, assim como os desafios envolvidos.

Avançando na prática

Infraestrutura própria ou na nuvem

Considere o caso de uma empresa que precisa executar uma aplicação de mineração de dados. A execução vai durar 20 meses e requer 8 servidores. A empresa pode montar sua própria infraestrutura ou alocar os servidores virtuais na nuvem. No primeiro caso, a empresa teria que arcar com a compra dos servidores que custam R\$ 6.000 cada e também com manutenção da infraestrutura que custaria R\$ 2.500 por mês para gastos com equipe técnica e refrigeração do ambiente. Se optar por usar um provedor de serviços em nuvem, a empresa tem opção de alocar servidores virtuais de capacidade compatível pelo valor de R\$ 600 por mês cada. Qual opção representa a solução de menor custo para a empresa? E se o prazo do projeto mudasse para 24 meses?

Resolução da situação-problema

O custo de manter a infraestrutura própria consiste em adquirir 8 servidores por R\$ 6.000 cada, mais um custo de operação mensal de R\$ 2.000 por 20 meses. Isso equivale a um total de R\$ 98.000 de custo de execução do projeto.

O custo de alocar a infraestrutura no provedor consiste em pagar pelo uso de 8 servidores que custam R\$ 600 por mês cada por um período total de 20 meses. Isso equivale a um total de R\$ 96.000 de custo de execução do projeto. Portanto, a solução em nuvem representa a opção de menor custo.

No entanto, se a duração do projeto fosse 24 meses, a situação se invertia. O custo da execução do projeto em infraestrutura própria seria de R\$ 108.000, enquanto o custo na nuvem seria de R\$ 115.200. Isso acontece porque o investimento inicial é maior no caso de infraestrutura própria, mas o custo operacional mensal no caso da nuvem é maior. Com um prazo maior, o custo operacional acaba por representar uma parcela mais significativa

do custo total. Por outro lado, prazos maiores também levam a uma maior depreciação da infraestrutura própria. De fato, cada caso tem que ser analisado isoladamente para se verificar qual é mais vantajoso.

Faça valer a pena

1. Um provedor oferece um serviço de banco de dados em nuvem. O modelo de tarifação determina os seguintes preços: R\$ 20,00 por mês para cada 1 GB de dados armazenado; R\$ 150,00 por mês pela instância (máquina virtual) que hospeda o banco de dados; R\$10 para cada 1.000 operações de leitura realizadas no banco. Suponha que o banco de dados tenha um volume de 20 GB de dados e, em média, sejam realizadas 20.000 requisições no banco por mês.

Considerando o cenário descrito, o custo total pelo uso do banco de dados por 5 meses seria:

- a. R\$ 1.850,00.
- b. R\$ 2.850,00.
- c. R\$ 3.000,00.
- d. R\$ 3.750,00.
- e. R\$ 4.000,00.

2. O cálculo dos custos de serviços em nuvem é uma tarefa complexa, uma vez que envolve muitas métricas e condições variadas. Considerando esse assunto, avalie as afirmativas apresentadas a seguir.

- I. O custo de uma máquina virtual depende, entre outros aspectos, da configuração da instância, da sua localização e do tempo de uso.
- II. Um dos componentes do custo de um banco de dados como serviço é o volume de dados armazenado.
- III. Em termos de custos, é sempre mais vantajoso para uma empresa alocar um servidor na nuvem do que manter um servidor próprio.

Considerando o contexto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.

e. II e III, apenas.

3. As ferramentas de gerenciamento de custos em nuvem são muito importantes para lidar com a complexidade de estimativa e o controle de gastos com serviços em ambiente de nuvem. Nesse contexto, avalie as afirmativas a seguir:

- I. Multicloud é um exemplo de nuvem híbrida.
- II. O monitoramento do uso de recursos computacionais em nuvem é utilizado pelos provedores apenas para cálculo de custos dos serviços.
- III. Os modelos de tarifação determinam métricas e custos unitários padronizados para uso de software e recursos computacionais em um provedor.

Considerando as afirmativas apresentadas, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

Migração de aplicações para a nuvem

Diálogo aberto

Prezado estudante, chegamos à última seção da segunda unidade. Nas seções anteriores, aprendemos sobre os principais avanços que viabilizaram o modelo de computação em nuvem, como as tecnologias de virtualização e contêineres, assim como, diversos aspectos que determinam como são tarifados os serviços em nuvem. Vamos agora dar mais um passo na compreensão das soluções e tecnologias utilizadas pelos provedores para oferta desses serviços. De que forma podemos acessar os provedores para instanciar e configurar serviços? Como as empresas controlam quais usuários podem configurar ou acessar cada serviço? Vamos discutir essas questões e outros tópicos importantes relacionados à migração de dados e aplicações para ambientes de computação em nuvem.

A fim de compreender esses assuntos, vamos retomar o caso no qual você, como analista de TI, está responsável por migrar o sistema de comércio eletrônico de um ambiente de TI tradicional da própria empresa para um ambiente de computação em nuvem. O objetivo é reduzir custos e melhorar a escalabilidade e a confiabilidade. O sistema em questão consiste em dois componentes: uma aplicação web e um banco de dados. A aplicação foi desenvolvida utilizando-se um framework antigo que já foi descontinuado. O banco de dados segue o modelo relacional e já armazena um volume de dados total de 4 terabytes. Nesta última etapa, você precisa elaborar um relatório respondendo à seguinte questão: qual a melhor estratégia de migração neste caso?

O primeiro passo é compreender os conceitos e os mecanismos relacionados com o problema de migração da aplicação. Nesta seção, serão discutidos diversos fatores associados com a migração e acesso a serviços em ambientes de computação em nuvem.

- Primeiramente, serão discutidos alguns desafios na migração de dados e aplicações para provedores de nuvem, assim como estratégias que podem ser utilizadas nesse processo.
- Em seguida, serão descritos o conceito de gerenciamento de identidade, o acesso e suas principais funcionalidades. Vamos ainda caracterizar diversas formas de acesso a serviços na nuvem.

- Por fim, serão ilustrados alguns entre os principais provedores que suportam a hospedagem de aplicações em nuvem por meio de serviço no modelo PaaS.

Bom estudo!

Não pode faltar

Migração para nuvem

Muitas empresas têm adotado soluções de TI baseadas no modelo de computação em nuvem em virtude de vários benefícios como redução de custos e maior escalabilidade. No caso de novas aplicações, as decisões de projeto já consideram características e tecnologias dos ambientes de computação em nuvem. O que pode ser feito no caso de aplicações já existentes que estão em execução em um ambiente tradicional de TI? Essas aplicações também podem se beneficiar das tecnologias de computação em nuvem. Para isso, as aplicações que estão em um ambiente tradicional precisam ser reimplantadas em um ambiente de computação em nuvem. Esse processo é denominado **migração para a nuvem** (PAHL; XIONG, 2013).

A migração de aplicações e dados para a nuvem não é uma tarefa simples. Pelo contrário, trata-se de um processo que envolve muitos desafios e, portanto, exige análise e planejamento cuidadosos para garantir o correto funcionamento dos sistemas na infraestrutura de computação em nuvem e assegurar que não haverá violação dos requisitos de segurança e privacidade (MORAIS, 2015). Para viabilizar a migração, podem ser necessárias, por exemplo, mudanças na arquitetura do sistema ou nas tecnologias utilizadas. No processo de migração de dados e de aplicações para um novo ambiente, o custo é um dos fatores mais relevantes. Entre os principais podemos citar:

- Treinamento de profissionais para gerenciamento do ambiente de computação em nuvem.
- Modificação nas aplicações ou necessidade de implementar novos componentes de software.
- Transferência de grandes volumes de dados.

Se a migração envolve uma aplicação com uma base de dados extensa, a transmissão dos dados pode durar semanas ou até meses, de acordo com o desempenho da rede e com o volume de dados (AWS, 2019).

Pesquise mais

Análise de custos

Um dos passos importantes de uma estratégia de migração bem planejada é a análise de custos. Para verificar a viabilidade financeira de mover a aplicação para ambiente de nuvem, devem ser observados não somente os custos iniciais de implantação, mas também os custos de operação a médio e longo prazo. Nesse sentido, indicamos a leitura do seguinte artigo:

RIBAS, M. *et al.* Tomada de decisão multicritério na migração de aplicativos para ambientes de nuvem do tipo software as a service. **Revista Brasileira de Administração Científica**, v. 5, n. 2, p. 83–94, 2014.

Outro desafio importante na migração para a nuvem são as questões relacionadas com portabilidade e interoperabilidade.

Portabilidade

Como discutido em Silva (2013), a portabilidade pode ser entendida como a facilidade com a qual um sistema pode ser levado de um ambiente para outro. O desejável é que a migração possa ser feita com o mínimo de adaptações necessárias nos sistemas. É importante observar que a compatibilidade não é esperada apenas em termos de ambiente de execução ou padrões de comunicação entre componentes de software, mas também em relação às tecnologias de gerenciamento de dados. Esses aspectos são relevantes no processo de migrar sistemas para a nuvem e também na migração entre provedores de nuvem, caso necessário.

Em geral, o processo de migração para a nuvem pode ser dividido em três etapas (MORAIS, 2015):

- **Planejamento:** O planejamento envolve atividades como: levantamento de requisitos; análise de riscos; escolha de ferramentas de migração, provedores e modelos de serviço e implantação; análise de custos e viabilidade e definição da estratégia de migração.
- **Execução:** A etapa de execução consiste nas ações para levar o sistema para o novo ambiente, conforme planejado. Essa etapa envolve atividades como: extração, conversão e transferência de dados; adaptação na arquitetura do sistema; alterações nas aplicações existentes, substituição ou implementação de novos componentes e bibliotecas.
- **Avaliação:** Por fim, a etapa de avaliação consiste na validação do processo e testes do sistema. Essa etapa inclui atividades como:

avaliação de desempenho e segurança; testes de integridade e validação da qualidade de experiência do usuário.

Assimile

Estratégia de migração

Uma das tarefas da etapa de planejamento é a definição da estratégia de migração. Existem diversas abordagens (AWS, 2018), entre as quais podemos destacar:

- **Re-host:** migrar a aplicação sem qualquer mudança arquitetural ou modificação de código. Isso é possível, por exemplo, quando a aplicação já executa em ambiente virtualizado, e as máquinas virtuais são copiadas para um provedor de nuvem pública. Os custos são reduzidos, mas a solução não é otimizada para ambiente de nuvem, portanto alguns mecanismos, como balanceamento dinâmico de carga, podem ser inviabilizados.
- **Re-platform:** a arquitetura geral do sistema não é alterada, mas os componentes são migrados de forma independente para aproveitar os serviços em nuvem. Por exemplo, instâncias de banco de dados são migradas para um provedor DBaaS (data base as a service) e aplicações são migradas para ambientes PaaS. É uma abordagem intermediária em termos da relação custo-benefício.
- **Re-factor ou Re-architect:** a aplicação é remodelada e reimplementada a partir de tecnologias e mecanismos típicos de computação em nuvem. É a estratégia que envolve maior custo e tempo de implementação. Por outro lado, permite aproveitar todo potencial do ambiente de nuvem.

Gerenciamento de Acesso e Identidade

Uma vez que um sistema é migrado para a nuvem, seus administradores precisam redefinir como o sistema será gerenciado e utilizado no novo ambiente. Em particular, é necessário determinar como será o acesso remoto aos componentes do sistema que passaram a executar em infraestrutura de computação em nuvem. Isso envolve dois aspectos principais: **o controle de permissões e as tecnologias de acesso**. Isso é relevante mesmo nos casos em que um sistema é migrado para ambiente de nuvem privada, pois o controle de acesso é parte das plataformas de computação em nuvem usadas tanto em modelo de nuvem pública quanto de nuvem privada.

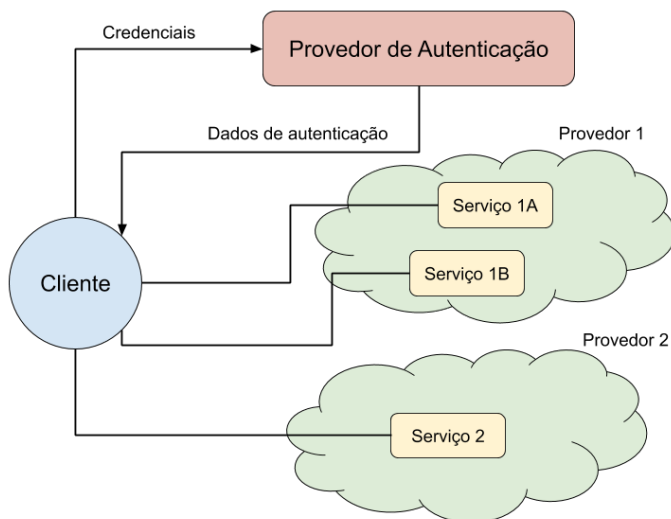
Um dos mecanismos mais importantes para controle é o **Gerenciamento de Acesso e Identidade** (IAM – *Identity and Access Management*). Esse mecanismo consiste nos componentes e políticas necessários para verificar e monitorar a identidade e as permissões de cada usuário em relação ao uso dos recursos computacionais disponíveis no ambiente de nuvem (ERL; PUTTINI; MAHMOOD, 2013). Em geral, um mecanismo IAM inclui três funcionalidades principais:

- **Autenticação:** verificação da identidade do usuário, por exemplo, a partir de senha ou certificado digital. Mecanismos mais sofisticados incluem uso de biometria, assim como a restrição do acesso de acordo com o dispositivo utilizado pelo usuário ou até mesmo sua localização.
- **Autorização:** controle dos níveis de acesso e das permissões concedidas para um usuário em relação aos serviços e recursos disponíveis.
- **Gerenciamento de usuários e credenciais:** permite criar ou alterar cadastro de usuários, assim como grupos de usuários, políticas de segurança ou regras de acesso para facilitar o controle de uso dos recursos e lidar com ameaças de segurança e privacidade.

Esse mecanismo permite que os administradores do sistema especifiquem os níveis ou permissões de acesso para cada profissional ou grupo de profissionais envolvidos no desenvolvimento e operação da uma aplicação em ambiente de computação em nuvem e também dos usuários da aplicação. Esse controle é imprescindível, pois o acesso aos serviços de suporte ou à própria aplicação é feito por meio da rede.

Por outro lado, um rigoroso controle de acesso leva a uma necessidade de autenticação e verificação de permissão de acesso para cada serviço em nuvem utilizado para implantar uma aplicação, ou seja, máquinas virtuais, bancos de dados, componentes de software, etc. A situação pode ser mais complicada no caso em que uma mesma aplicação faz uso de serviços em provedores diferentes. Para lidar com essas questões, foram criados mecanismos de autenticação unificada (SSO – *Single Sign On*). Essa abordagem introduz o papel de um provedor de autenticação, por meio do qual um usuário pode acessar diversos serviços, inclusive em diferentes provedores, mediante um único processo de autenticação (ERL; PUTTINI; MAHMOOD, 2013). A Figura 2.3 ilustra esse cenário. Depois de realizar autenticação, um cliente pode acessar diversos serviços, inclusive em provedores diferentes, por meio da mesma identidade.

Figura 2.3 | Autenticação unificada para acesso a serviços em nuvem



Fonte: elaborada pelo autor.

Pesquise mais

Os mecanismos de autenticação unificada requerem formatos padronizados e abertos para especificação das credenciais dos usuários, assim como para as políticas e regras de controle de acesso.

MARCON JR. *et al.* Aspectos de segurança e privacidade em ambientes de Computação em Nuvem. In: **Minicursos do Simpósio Brasileiro de Segurança da Informação e Sistemas**, p. 20-24. Porto Alegre: Sociedade Brasileira de Computação, 2010.

Acesso remoto

Além do uso de mecanismos de gerenciamento de identidade e controle de acesso, é importante a definição das formas de acesso remoto aos recursos computacionais virtualizados disponíveis na infraestrutura de computação em nuvem. Quando um sistema é migrado para um ambiente de nuvem, ele pode fazer uso de diversos modelos de serviço, como IaaS ou PaaS. Por exemplo, podem ser alocados contêineres ou máquinas virtuais, ambientes de desenvolvimento e testes de aplicações web, serviços de armazenamentos de dados, etc. Para cada modelo, podem existir diferentes formas de utilizar e configurar os serviços.

Em geral, quando o sistema faz uso de nuvem pública, o provedor oferece um portal administrativo com interface web, a partir do qual todos os serviços podem ser instanciados e configurados de forma centralizada. Esse tipo de portal administrativo permite, ainda, o gerenciamento de identidade para controle de acesso e monitoramento de desempenho e custos dos serviços. No entanto, tal sistema web exige interação humana para gerenciamento dos serviços. Dessa forma, são necessárias outras formas de acesso que permitam automatização das tarefas de alocação e configuração dos serviços em nuvem.

Vamos discutir alguns casos típicos. Uma instância de máquina virtual é um exemplo de serviço no modelo IaaS. Você sabe quais são as formas de acessar uma máquina virtual que está em algum provedor de nuvem pública por meio da Internet? Como discutido em (ERL; PUTTINI; MAHMOOD, 2013), as duas formas principais de acessar uma máquina remotamente são:

Ferramenta de Desktop Remoto (CASAS *et al.*, 2013): O desktop remoto é uma tecnologia que permite usuários finais acessarem remotamente conteúdo e aplicações que estão em outro computador por meio de uma conexão de rede.

SSH (Secure Shell) (YLONEN; LONVICK, 2006): O SSH é um protocolo que viabiliza o *login* em um terminal de linha de comando de um computador remoto, de forma segura devido ao uso de técnicas de criptografia.

Com essas ferramentas, é possível acessar a máquina no provedor e realizar tarefas como editar arquivos de configuração e instalar bibliotecas e aplicações.

Instanciar e configurar serviços em nuvem

As tecnologias discutidas anteriormente permitem o acesso a uma máquina virtual, mas existe outra forma de criar a máquina virtual, além da forma interativa por meio da interface web do provedor? Sim, existem outras três formas básicas de instanciar e configurar serviços em nuvem: **Interface de linha de comando (CLI – *Command Line Interface*):** Geralmente, o provedor disponibiliza uma CLI na forma de um software cliente que permite ao usuário executar comandos em um terminal para criar, configurar e acessar serviços no provedor. A execução desses comandos só é possível depois de autenticação do usuário.

Kit de desenvolvimento de software (SDK – *Software Development kit*): De forma análoga, o provedor distribui SDKs para as linguagens de programação para as quais que ele oferece suporte. Dessa forma, o usuário pode criar programas para manipulação dos recursos e serviços disponíveis no provedor. Dessa forma, é possível automatizar a alocação e a configuração

dos serviços em nuvem. Isso também é possível com uso de CLI, mas, em vez de codificar programas, o usuário codifica *shell scripts*. Por exemplo, o usuário pode escrever um programa ou um script para criar um cluster de máquinas virtuais em uma nuvem pública, submeter tarefas para serem executadas no cluster e liberar os recursos depois que as tarefas tiverem sido finalizadas. Essa possibilidade de automatizar as tarefas de acesso e o gerenciamento dos recursos virtualizados em ambiente de computação em nuvem culminou no conceito de Infraestrutura como código (IaC – *Infrastructure as Code*) (ARTAC *et al.*, 2017). Dessa forma, o gerenciamento de servidores em nuvem pode ser automatizado com arquivos de configuração e programas (ou scripts).

CLI e SDK são distribuídos pelos provedores na forma de bibliotecas de software, ou seja, precisam ser previamente instalados no computador local para serem utilizados.

Interface de programação de aplicações (API – *Application Programming Interface*): O uso de API permite a integração dinâmica de aplicações clientes com o provedor para gerenciamento dos serviços. Em geral, o acesso à API dos provedores é disponibilizado por meio do protocolo SSH. Assim, elas são denominadas Web APIs. Isso significa que uma aplicação cliente pode implementar requisições HTTP para acessar funcionalidades especificadas na API do provedor de forma dinâmica. Não é necessário instalar componentes de software específicos para isso.

Exemplificando

Foram discutidas três formas de acesso para instanciação e configuração de serviços em provedores de nuvem: CLI, SDK e API. Mas, quais são as principais funcionalidades que os provedores disponibilizam para acesso remoto por meio dessas tecnologias? A seguir, alguns exemplos das principais funcionalidades (ERL; PUTTINI; MAHMOOD, 2013):

- Acionamento de mecanismos de escalonamento automático e balanceamento de carga.
- Criação de unidades de armazenamento de dados.
- Verificação de métricas de desempenho.
- Criação de máquinas virtuais.
- Monitoramento de custos.
- Versionamento e controle de implantação de aplicações.

Em resumo, de acordo com a estratégia escolhida, a migração de sistemas para ambiente de nuvem envolve tarefas de:

- Planejamento.

- Execução.
- Avaliação.

Mecanismos de gerenciamento de identidade são utilizados para determinar as permissões de acesso aos serviços para cada um dos profissionais que participam na migração do sistema na nuvem. Algumas formas de acesso aos recursos podem ser usadas para instanciar e configurar os serviços necessários para migração das aplicações e dados: uma forma interativa (portal web do provedor) ou as formas de acesso que suportam programação das tarefas (CLI, SDK e API).

Serviços PaaS

Na prática, a estratégia básica de migração inicial ou empresas de pequeno porte é a estratégia Re-Platform, como explicado anteriormente. Uma forma de usar essa estratégia é migrar os componentes do sistema para serviços PaaS. Isso significa enviar o código já existente da aplicação para um ambiente de desenvolvimento compatível na nuvem e implantar (fazer o *deploy*) da aplicação. Adicionalmente, um serviço DBaaS pode ser usado para hospedar o banco de dados da aplicação, se for o caso. As principais dificuldades são a portabilidade do código e o paradigma de gerenciamento de dados. Para aplicações feitas em plataformas populares na indústria de TI, como Java, Python ou JavaScript, a portabilidade não é um problema, pois a maioria dos provedores PaaS suporta essas e outras plataformas como Ruby e PHP. A maioria dos provedores também suporta diversos SGBDs, tanto os que seguem o modelo relacional, quanto os bancos NoSQL (DE SOUZA *et al.*, 2014).

Refleta

A migração de bases de dados para provedores PaaS pode parecer uma tarefa simples, mas não é. Não basta a existência de um banco de dados como serviço com tecnologia compatível. Quanto tempo demoraria para transmitir uma base de dados de 2 TB para um provedor considerando uma taxa de transmissão média de 5 Mbps? Sem contar as tarifas de transferência de dados que podem ser cobradas pelo provedor ou falhas que podem ocorrer durante a transmissão. Existem ainda aspectos de otimização e segurança da base de dados. Quais seriam as abordagens mais interessantes para migração de bancos de dados? Será que existem ferramentas específicas para auxiliar nesse processo?

Essa abordagem de hospedagem de aplicações e bases de dados na nuvem por meio de provedores PaaS permite a redução de custos, escalabilidade

e dispensa a responsabilidade de gerenciamento de máquinas virtuais. O Quadro 2.5 apresenta alguns entre os principais serviços de hospedagem de aplicações em provedores PaaS.

Quadro 2.5 | Principais serviços de hospedagem de aplicações em nuvem pública

Serviço PaaS	Tecnologias Suportadas
AWS Elastic Beanstalk.	Java, .NET, PHP, Node.js, Python, Ruby, Go e Docker.
Google App Engine.	Java, PHP, Node.js, Python, C#, .Net, Ruby e Go.
Heroku Platform.	Node, Ruby, Java, PHP, Python, Go, Scala, Clojure e contêineres customizados.
Azure App Service.	NET, Java, Node.js, PHP, e Python em Windows, .NET Core, Node.js, PHP ou Ruby em Linux e contêineres.

Fonte: Elaborado pelo autor.

Chegamos ao final de mais uma unidade, na qual aprendemos muitos tópicos interessantes, como tecnologias de virtualização, métricas de tarifação para serviços em nuvem e fatores importantes na migração de sistemas para ambientes de computação em nuvem. Com o conhecimento adquirido até aqui, você será capaz de analisar as melhores estratégias para migrar aplicações para provedores considerando aspectos como a escalabilidade e os custos envolvidos.

Sem medo de errar

A empresa na qual você é analista de TI pretende migrar um sistema de comércio eletrônico para ambiente de computação em nuvem a fim de reduzir custos operacionais e aprimorar o grau de escalabilidade do sistema. Seu papel é liderar a equipe responsável pela migração do sistema. Esse sistema inclui uma aplicação web por meio da qual os clientes realizam os pedidos e um banco de dados no qual se registram as informações dos clientes e das compras. A aplicação foi desenvolvida há muitos anos, com um framework que já foi descontinuado.

Você precisa determinar qual é a melhor estratégia de migração neste caso. Foram apresentadas três estratégias de migração. Vamos analisar cada uma delas a fim de justificar qual seria a melhor escolha.

A estratégia Re-host é a migração sem qualquer mudança arquitetural ou modificação nas aplicações ou banco de dados, o que a torna a estratégia mais simples e de menor custo. No entanto, ela só é aplicável quando o sistema já executa em ambiente virtualizado. Nesse caso, as máquinas

virtuais são copiadas para a infraestrutura do provedor e a aplicação é inicializada. Apesar do baixo custo, essa estratégia não permite o aproveitamento dos benefícios da nuvem em todo seu potencial pois o sistema legado não foi adaptado para usar os mecanismos disponíveis. Essa estratégia não é factível para o sistema de comércio eletrônico que se pretende migrar, pois esse sistema está em execução em uma infraestrutura tradicional de TI sem mecanismos sofisticados de virtualização. Além disso, mesmo que isso fosse possível, essa estratégia não atenderia os requisitos de escalabilidade e confiabilidade desejáveis. Esses requisitos são viabilizados por mecanismos nativos da computação em nuvem.

A estratégia Re-platform também não exige alterações significativas na arquitetura geral do sistema, mas os componentes são migrados de forma independente para aproveitar os serviços em nuvem. Isso é feito por meio da alocação de serviços PaaS, ou seja, ambientes pré-configurados com as tecnologias necessárias para implantação das aplicações e bancos de dados. Essa estratégia caracteriza uma solução intermediária, em termos de custos e complexidade. No entanto, ela também não serve para o sistema em questão, pois a aplicação foi implementada com um framework específico que foi descontinuado. Logo, não é provável que haja provedor PaaS com suporte para aquele framework. Existe, portanto, um problema de portabilidade dessa aplicação para um ambiente moderno de computação em nuvem.

A terceira e última estratégia de migração descrita foi a Re-factor, também chamada Re-architect. Essa estratégia consiste em implementar novamente o sistema ou a maior parte de seus componentes a partir de tecnologias e mecanismos típicos de computação em nuvem. Devido a esse trabalho significativo de refatoração e reescrita de código, essa estratégia é a que implica maior custo e tempo de implementação. A vantagem é que, com as adaptações, os benefícios da computação em nuvem serão maximizados. Essa abordagem é, portanto, a melhor escolha para o sistema de comércio eletrônico, devido às necessidades de portar o código legado e aos requisitos de confiabilidade e escalabilidade que a empresa enseja para o sistema. Esses requisitos serão viabilizados uma vez que a migração contemple uma nova implementação da aplicação web a partir de tecnologias modernas dos provedores e então seja possível também fazer uso de mecanismos tais como escalonamento automático e balanceamento de carga.

Podemos concluir que, apesar dos relevantes benefícios da computação em nuvem, a migração de aplicações para esse modelo não é uma tarefa trivial. Uma migração bem-sucedida requer planejamento, profissionais qualificados, uso de novas tecnologias, entre outros. Tudo isso pode tornar o processo caro e demorado. No médio e longo prazo, mesmo uma migração que exige alto investimento pode ser vantajosa, devido aos

inúmeros benefícios da computação em nuvem, tanto em aspectos de custo operacional e abstração da complexidade, como em relação ao desempenho e confiabilidade dos serviços.

As avaliações e análises realizadas durante esta unidade permitiram uma compreensão ampla de diversas soluções e tecnologias usadas pelos provedores de nuvem, de métricas utilizadas no cálculo dos preços de serviços em nuvem e também dos desafios e estratégias de migração de aplicações para infraestruturas de computação em nuvem.

Avançando na prática

Tempo para transferência de um banco de dados

Nesta nova situação-problema, vamos continuar considerando o sistema de comércio eletrônico. Suponha que, com o passar dos anos, a base de dados do sistema já atingiu um volume de 4 terabytes de dados. Nesse contexto, determine quanto tempo demoraria para transferir a base de dados da empresa para um provedor A. Considere que a taxa de transmissão média entre a empresa e o provedor A é de 4 Mbps.

Resolução da situação-problema

A tarefa consiste em calcular quanto tempo seria necessário para completar a transmissão da base de dados do sistema considerando uma taxa de transmissão média de 4 Mbps com um dado provedor. O cálculo requer dividir o volume total de dados pela taxa de transmissão.

Vamos converter os valores dados para bits a fim de facilitar os cálculos. Temos que 4 TB equivale a $4 \times 1024 \times 1024 \times 1024 \times 1024 \times 8$ bits. Agora, precisamos converter a taxa de transmissão de megabits por segundo (Mbps) para bits por segundo (bps). Vale lembrar que taxa de transmissão em redes é medida em potência de 10, então a taxa de 4 Mbps corresponde a $4 \times 1000 \times 1000$ bps. Logo, o resultado da divisão é 8796093,02 segundos. Para encontrar o valor em dias, podemos dividir o valor em segundos por $60 \times 60 \times 24$. O resultado equivale a aproximadamente 101 dias. Por exemplo, para realizar a transferência em no máximo uma semana, a taxa de transferência deveria ser em torno de 15 vezes maior. Portanto, uma rede de alto desempenho é fundamental para migração e operação de bancos de dados na nuvem.

Em resumo, a transmissão demoraria mais de 2 meses para ser completada, o que poderia ser pior se considerássemos a possibilidade de falhas. Além disso, a migração de bases de dados envolve outras questões importantes como verificar as tecnologias de armazenamento e os SGBDs compatíveis, a localização geográfica do provedor (por questões de governança e legislação interna ou governamental), os mecanismos de segurança de dados disponíveis no provedor e a latência entre a aplicação e o banco de dados, pois isso afeta o desempenho geral do sistema.

Faça valer a pena

1. Nos últimos anos, grande parte das empresas decidiram migrar suas aplicações para ambientes de Computação em Nuvem a fim de se beneficiar em aspectos como redução de custos e maior escalabilidade. Avalie as afirmativas a seguir sobre o processo de migração de aplicações de infraestrutura tradicional de TI para uma infraestrutura de Computação em Nuvem.

- I. A migração é um processo complexo que envolve etapas como planejamento, execução e avaliação.
- II. Problemas de portabilidade e interoperabilidade podem elevar os custos do processo de migração de dados e aplicações.
- III. Entre as estratégias de migração, a que implica em mais modificações na arquitetura e código das aplicações é a estratégia Re-host.

No contexto apresentado, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

2. As plataformas de computação em nuvem, como o OpenStack, oferecem suporte ao gerenciamento de acesso e identidade (IAM). Sobre esse mecanismo, avalie as afirmativas a seguir:

- I. Os mecanismos IAM permitem diferentes formas de controlar o acesso dos desenvolvedores e dos usuários às aplicações em nuvem.

- II. A funcionalidade de autenticação consiste em verificar as permissões de um usuário em relação aos serviços disponíveis.
- III. Políticas de controle de acesso podem ser direcionadas para um usuário ou para um grupo de usuários.

No contexto apresentado, é correto o que se afirma em:

- a. III, apenas.
- b. I e II, apenas.
- c. II e III, apenas.
- d. I e III, apenas.
- e. I, II e III.

3. Os provedores de computação em nuvem suportam diferentes formas de gerenciamento e acesso aos serviços. Algumas formas de acesso permitem automatização das ações de gerenciamento. Entre as principais formas de acesso, pode-se destacar: interface web, CLI, SDK e API web.

Considerando as formas de acesso e configuração dos serviços em nuvem, é correto o que se afirma em:

- a. Software como Serviço (SaaS) é acessado sempre como interface web.
- b. A forma de acesso por linha de comando (CLI) dispensa a instalação de qualquer software do provedor, pois pode ser usado um terminal de comando nativo.
- c. Todas as formas de acesso facilitam o modelo de Infraestrutura como código (IaC).
- d. API web pode ser utilizada para acesso e configuração de diversos modelos de serviço, inclusive serviços de armazenamento de dados.
- e. Sem o SDK de um provedor para uma determinada linguagem, não é possível hospedar uma aplicação que usa essa linguagem nesse provedor.

Referências

- AHMAD, R. W. *et al.* A survey on virtual machine migration and server consolidation frameworks for cloud data centers. **Journal of Network and Computer Applications**, v. 52, p. 11–25, 2015.
- AMAZON. **Amazon Elastic Compute Cloud (Amazon EC2)**. Disponível em: <https://aws.amazon.com/pt/ec2/>. Acesso em: 12 jul. 2019.
- AMAZON. **Amazon Web Services (AWS)**. Disponível em: <https://aws.amazon.com/pt/>. Acesso em: 12 jul. 2019.
- ARTAC, M. *et al.* DevOps: Introducing Infrastructure-as-Code. In: **Proceedings of the IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)**, p. 497-498. 2017.
- AWS. **AWS Migration Whitepaper**. AWS Professional Services, 2018. Disponível em: <https://d1.awsstatic.com/whitepapers/Migration/aws-migration-whitepaper.pdf>. Acesso em: 12 jul. 2019.
- AWS. **Migração de dados para a nuvem**. Disponível em: <https://aws.amazon.com/pt/cloud-data-migration/>. Acesso em: 24 maio 2019.
- BACHIEGA, N. G.; SOUZA, P. S. L. de; BRUSCHI, S. M. Avaliação de desempenho de virtualização baseada em contêiner. In: **Escola Regional de Alto Desempenho de São Paulo - ERAD-SP**, SBC, 2017.
- BECKER, A. K.; CLARO, D. B.; SOBRAL, J. B. Web services e XML: um novo paradigma da computação distribuída. **Objetos Distribuídos**, p. 51–66, 2001.
- BERNSTEIN, D. Containers and cloud: From LXC to Docker to Kubernetes. **IEEE Cloud Computing**, v. 1, n. 3, p. 81–84, 2014.
- CANONICAL. **Infrastructure for container projects**. Linux containers. Disponível em: <https://linuxcontainers.org/>. Acesso em: 15 jul. 2019.
- CASAS, P. *et al.* Quality of experience in remote virtual desktop services. In: **Proceedings of the 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)**, 2013.
- DESOUZA, A. M. *et al.* Critérios para Seleção de SGBD NoSQL: o Ponto de Vista de Especialistas com base na Literatura. In: **Anais do X Simpósio Brasileiro de Sistemas de Informação (SBSI)**, p. 131-142. Porto Alegre: Sociedade Brasileira de Computação, 2014.
- ERL, T.; PUTTINI, R.; MAHMOOD, Z. **Cloud Computing: Concepts, Technology & Architecture**. [s.l.] Prentice Hall, 2013.

MARCON JR. *et al.* Aspectos de segurança e privacidade em ambientes de Computação em Nuvem. *In: Minicursos do Simpósio Brasileiro de Segurança da Informação e Sistemas*, p. 20-24. Porto Alegre: Sociedade Brasileira de Computação, 2010. Disponível em: https://www.researchgate.net/publication/262567797_Aspectos_de_Seguranca_e_Privacidade_em_Ambientes_de_Computacao_em_Nuvem. Acesso em: 15 jul. 2019.

MORAIS, N. S. de. **Proposta de modelo de migração de sistemas de ambiente tradicional para nuvem privada para o Polo de Tecnologia da Informação do Exército brasileiro**. 2015. Dissertação (Mestrado Profissional em Computação Aplicada) – Universidade de Brasília, Brasília, 2015.

PAHL, C.; XIONG, H. Migration to PaaS clouds-migration process and architectural concerns. *In: Proceedings of the IEEE 7th International Symposium on the Maintenance and Evolution of Service-Oriented and Cloud-Based Systems*, 2013.

RIBAS, M. *et al.* Tomada de decisão multicritério na migração de aplicativos para ambientes de nuvem do tipo software as a service. **Revista Brasileira de Administração Científica**, v. 5, n. 2, p. 83–94, 2014. Disponível em: <https://doi.org/10.6008/SPC2179-684X.2014.002.0007>. Acesso em: 24 maio 2019.

SILVA, E. A. N. da. **Uma abordagem dirigida por modelos para portabilidade entre plataformas de computação em nuvem**. 2013. Dissertação (Mestrado em Ciências Exatas e da Terra) – Universidade Federal de São Carlos, São Carlos, 2013.

YLONEN, T.; LONVICK, C. **The secure shell (SSH) protocol architecture**. RFC 4251. [s.l.] The Internet Engineering Task Force (IETF), 2006.

Unidade 3

Neumar Costa Malheiros

Ofertas de serviço em Computação em Nuvem

Convite ao estudo

Caro aluno, vamos iniciar mais uma unidade da nossa jornada de aprendizagem sobre Computação em Nuvem. Nosso foco é compreender os tipos de serviços e suas finalidades na implantação de soluções de TI em ambientes de nuvem. Você já verificou na plataforma dos grandes provedores a variedade de serviços existentes? Ao migrar um sistema computacional para a nuvem, temos que escolher os serviços adequados para implantar cada um dos componentes do sistema. Esse desafio também se mostra presente no desenvolvimento de novas aplicações já no ambiente de nuvem, as chamadas aplicações nativas em nuvem. Entre os principais tipos de serviços, vamos nos concentrar naqueles que servem para executar aplicações e armazenar dados. O importante é saber quais são as opções para fazer a melhor escolha no momento de hospedar um componente de software na nuvem.

A fim de trazer a discussão para tópicos e exemplos concretos, vamos considerar o seguinte cenário: suponha que você é um analista de TI em uma empresa de desenvolvimento de aplicativos para dispositivos móveis. Sua responsabilidade é coordenar os processos de integração de aplicativos com soluções em nuvem. A empresa desenvolve diversos tipos de aplicativos, por exemplo, para streaming de áudio, mobilidade urbana e mercado financeiro. Você precisa conhecer as soluções dos provedores de nuvem pública em termos das tecnologias utilizadas e preços praticados. Para cada projeto da empresa, surgem demandas para diversos tipos de serviços em nuvem. Você deverá tomar decisões de projeto que envolvem questões como:

- Determinar o tipo de serviço de armazenamento em nuvem mais apropriado para dados de aplicativos utilizados no monitoramento da qualidade de produção em grandes manufaturas.
- Conhecer serviços disponíveis na área de Inteligência Artificial (IA) para o desenvolvimento de soluções inovadoras em inteligência de negócios ou aplicações financeiras.
- Verificar os serviços em nuvem que oferecem melhor relação custo-benefício para executar aplicações web que suportem a sincronização com dados dos aplicativos.

Para avaliar e comparar as possíveis soluções, é importante conhecer a finalidade, as tecnologias e os custos dos inúmeros serviços em nuvem. Diante disso, veremos, nesta unidade, os diversos tipos de serviços ofertados por provedores. Vamos começar a nossa abordagem na Seção 1 com a descrição de serviços de processamentos de dados, ou seja, instâncias que podem ser alocadas para execução de aplicações ou serviços web. Na seção seguinte, vamos descrever serviços voltados para armazenamento e análise de dados, explicando os principais tipos de armazenamento em nuvem e apresentando alguns exemplos. Além disso, destacaremos alguns serviços específicos para análise de dados, pois trata-se de um assunto muito importante atualmente, por exemplo, para aplicações em Ciência dos Dados (*Data Science*). Por fim, na última seção, descreveremos alguns serviços voltados para aplicações específicas, como *Machine Learning*, IoT e aplicações multimídia. Com o conteúdo desta unidade, você será capaz de compreender os principais tipos de serviços ofertados pelos provedores e em quais cenários eles devem ser utilizados.

Bons estudos!

Serviços de processamento de dados

Diálogo aberto

Caro aluno, será que você conhece os principais tipos de serviços para execução de aplicações em ambientes de Computação em Nuvem? Uma escolha criteriosa da melhor opção para cada aplicação depende do entendimento das características dos principais serviços ofertados pelos provedores de nuvem pública.

Para aprofundar nosso estudo sobre o tema, vamos considerar a seguinte situação: a empresa em que você trabalha está desenvolvendo um novo aplicativo para uma grande loja de departamentos, cujo objetivo é possibilitar que o consumidor faça suas compras por meio do aplicativo e do site oficial da loja. Como a contratante já tem uma grande base de consumidores em todo o território nacional, espera-se um grande volume de acessos ao aplicativo, apesar de ser difícil estimar precisamente o volume total. Diante desse cenário, a empresa solicita que você faça um estudo sobre os serviços de Computação em Nuvem apropriados para integração do aplicativo com os demais componentes do sistema de vendas da loja de departamentos. A integração é feita por meio de uma API web, então é preciso escolher um serviço em nuvem para executar essa API. Além disso, exige que seja um serviço de Computação em Nuvem com alto grau de escalabilidade, no qual as instâncias sejam alocadas sob demanda, sem a necessidade de alocar ou gerenciar os recursos virtualizados. Você também decidiu fazer uma avaliação do modelo de tarifação e buscou orçamento dos valores em pelo menos dois provedores. Ao fim do trabalho, você deve entregar um relatório indicando do tipo de serviço apropriado, os possíveis modelos de tarifação e os dois orçamentos de provedores de grande porte.

Para lidar com essa questão e outros problemas relacionados ao assunto, apresentaremos, nesta seção, uma visão geral dos serviços em nuvem classificados como serviços de computação (*Compute Services*), que representam os recursos virtualizados com capacidade de processamento para executar aplicações. Vamos estudar os três principais tipos de serviços nessa categoria: instâncias de máquinas virtuais, contêineres e Função como Serviço (FaaS). Além disso, vamos discutir alguns exemplos importantes de serviços de rede, que permitem a integração de componentes de software e viabilizam mecanismos de tolerância a falhas e replicação de dados. Ao concluir esta seção, você será capaz de comparar as opções para execução de aplicações em nuvem a fim de avaliar a melhor escolha para diferentes cenários.

Não pode faltar

Os provedores de Computação em Nuvem oferecem muitos serviços diferentes, que podem ser classificados quanto ao modelo em: IaaS, PaaS e SaaS. Além disso, os serviços em nuvem podem ainda ser classificados de acordo com sua finalidade, por exemplo, serviços para armazenamento de dados ou serviços para hospedagem de aplicações. Existem alguns tipos de serviços mais comuns que estão no portfólio de produtos da maioria dos provedores. Esse é o caso dos serviços cuja finalidade é prover capacidade de computação (processamento de dados), ou seja, são os serviços em nuvem que permitem a execução de aplicações. O exemplo típico de serviço de computação é a criação de instâncias de máquinas virtuais (VMs – *Virtual Machines*), também denominadas servidores virtuais. Por exemplo, se uma empresa precisa executar uma aplicação, ela deve alocar um servidor com capacidade computacional para isso, podendo criar uma máquina virtual na nuvem onde essa aplicação será executada.

Para provisionamento de serviços em nuvem, os provedores fazem uso de sistemas de gerenciamento dos recursos computacionais. Um dos mecanismos mais importantes de tais sistemas é o Gerenciador de Infraestrutura Virtual (VIM – Virtual Infrastructure Manager), que coordena a criação de instâncias de máquinas virtuais a partir dos recursos computacionais disponíveis no provedor (ERL; PUTTINI; MAHMOOD, 2013). Em geral, o VIM mantém um repositório com imagens de máquinas virtuais, que são modelos pré-configurados a partir dos quais são criadas as instâncias solicitadas pelos clientes. Cada imagem possui uma determinada capacidade, em termos do número de núcleos de processamento e da quantidade de memória, e um sistema operacional específico já instalado e configurado. Alguns provedores permitem também a customização de VMs, isto é, em vez de usar um modelo predefinido, o cliente pode configurar sua própria VM.

Pesquise mais

Com a alocação dinâmica de máquinas virtuais, o provedor pode chegar a um cenário no qual há servidores físicos sobrecarregados (com muitas máquinas virtuais), enquanto outros servidores permanecem subutilizados (com poucas máquinas virtuais). Então, pode ser necessário realizar a migração de máquinas virtuais entre servidores físicos, a fim de otimizar desempenho e economizar energia.

CARDOSO, L. P. **Migração de Máquinas Virtuais para Economia de Energia**. 2014. Trabalho de Conclusão de Curso (Engenharia de Computação e Informação) - Universidade Federal do Rio de Janeiro, p. 6-15. Rio de Janeiro: UFRJ/POLI, 2014.

Com a finalidade de dar suporte a aplicações com diferentes requisitos, cada provedor oferece vários modelos de máquinas virtuais, além de especificar o preço de cada um deles. Dessa forma, o cliente pode criar quantas instâncias forem necessárias para atender a sua demanda de TI. Um fator que influencia muito os preços é o tipo de alocação, sendo os dois principais: alocação de instâncias sob demanda e alocação de instâncias reservadas (ERL; PUTTINI; MAHMOOD, 2013). A alocação sob demanda é caracterizada pelo provisionamento dinâmico da máquina virtual no instante da solicitação, caso existam recursos disponíveis. Já a alocação de instâncias reservadas exige que o cliente especifique previamente um compromisso de uso dos recursos. Os provedores cobram um preço menor para as instâncias reservadas, pois a solicitação antecipada permite melhor planejamento e otimização do uso da infraestrutura. Além disso, as instâncias sob demanda podem ser liberadas a qualquer momento, enquanto as reservadas têm o tempo de uso previamente especificado. Assim, é oferecido um preço menor para as instâncias reservadas com o objetivo de que os clientes as utilizem por um período de tempo maior, o que resulta em menos recursos ociosos no provedor. Nesse caso, o pagamento é feito de acordo com o tempo reservado e não pelo tempo efetivo de uso.

Suponha que você precise de uma máquina com sistema Linux para executar um servidor de e-mail ou uma máquina com sistema Windows para executar uma aplicação web. Você conhece os serviços de computação disponíveis no mercado para alocação de máquinas virtuais? Entre os serviços para criação de máquinas virtuais em provedores de Computação em Nuvem, podemos citar:

- Amazon Elastic Compute Cloud (Amazon EC2) (AWS, 2019).
- Google Cloud Compute Engine (GOOGLE, 2019a).
- IBM Cloud Virtual Servers (IBM, 2019c).
- Azure Virtual Machines (MICROSOFT, 2019).

As empresas podem usar esses serviços para alocar máquinas virtuais e, assim, construir uma infraestrutura virtualizada na nuvem de acordo com sua demanda de capacidade computacional. Os serviços de instanciação de máquinas virtuais oferecem mecanismos de auto escalonamento. Isso significa que é possível habilitar a alocação automática de mais instâncias quando a demanda aumentar ou diminuir.

No entanto, o uso de máquinas virtuais não é a única opção de serviços de computação. Outro tipo importante de serviço para execução de aplicações é a instanciação de contêineres. Assim, quando uma empresa cliente precisa executar uma aplicação em um provedor, ela pode criar um contêiner para executar essa

aplicação, em vez de criar uma máquina virtual. Além de escalonamento automático, os serviços de contêineres facilitam a replicação das instâncias de execução pois eles exigem menos recursos que as máquinas virtuais. Isso também implica em custos mais baixos. Por outro lado, as máquinas virtuais oferecem nível mais alto de isolamento e proteção. A seguir, são apresentados exemplos de serviços para gerenciamento de contêineres em provedores de nuvem pública:

- Amazon Elastic Container Service (Amazon ECS) (AWS, 2019b).
- Azure Container Instances (MICROSOFT, 2019a).
- Google Kubernetes Engine (GOOGLE, 2019d).
- IBM Cloud Kubernetes Service (IBM, 2019b).

Em geral, as aplicações de TI são sistemas complexos compostos por vários componentes, como bancos de dados, frontend web, entre outros. Cada deles pode ser encapsulado em um contêiner e, para isso, são necessárias ferramentas para o gerenciamento. A partir disso, surgiu o conceito de orquestração de contêineres que corresponde ao processo de automatizar a implantação, o escalonamento e o gerenciamento dessas aplicações (YEGULALP, 2019).

Exemplificando

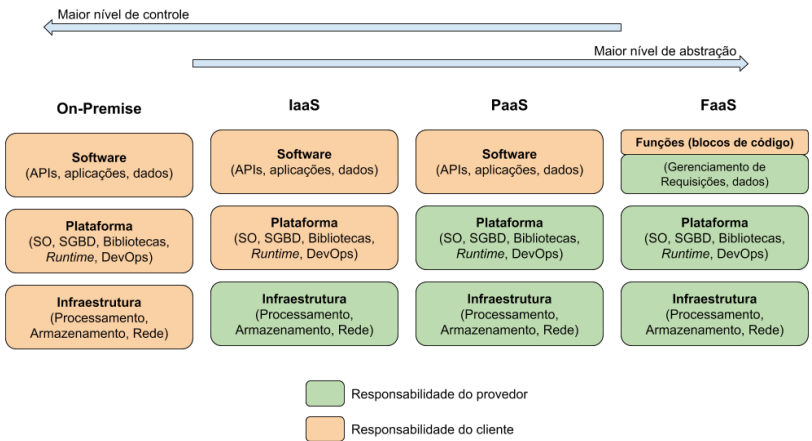
A plataforma Kubernetes é a tecnologia padrão para orquestração de contêineres em nuvens públicas (KUBERNETES, 2019), criada pelo Google e distribuída como um projeto de software livre em 2014. A maioria dos provedores oferece suporte para a plataforma Kubernetes de forma a garantir escalabilidade e facilidade de gerenciamento para aplicações que envolvem contêineres.

As máquinas virtuais e os contêineres são as principais instâncias para execução de aplicações em nuvem e, fazendo uso desses serviços, não há necessidade de gerenciar uma infraestrutura própria de TI. No entanto, ainda é necessário gerenciar os recursos virtualizados, o que pode ser um desafio, principalmente para pequenas e médias empresas ou aquelas cujo foco não é a área de TI. Para lidar com essa questão, foi introduzido o paradigma de Computação Sem Servidor (Serverless Computing) (BALDINI et al., 2017), que corresponde a um modelo de arquitetura de aplicações no qual blocos de código são executados na nuvem sem nenhum controle sobre os recursos computacionais subjacentes. Nesse sentido, alguma instância precisa ser provisionada para executar os blocos de código, mas isso é feito sob demanda pelo provedor.

Uma das formas de implementar solução Serverless é o modelo de Função como Serviço (FaaS – Function as a Service) (BACK; ANDRIKOPOULOS,

2018). Esse modelo corresponde a um nível de abstração da infraestrutura, em que o cliente do serviço não precisa se preocupar com as instâncias nas quais seu software será executado. Portanto, o modelo de serviço FaaS é visto como uma evolução dos modelos para execução de aplicações em nuvem. A Figura 3.1 ilustra a evolução dos modelos de serviço e, a partir dela, é possível observar que o modelo FaaS possui um maior nível abstração quando comparado aos demais modelos, porém, com um menor nível de controle.

Figura 3.1 | Evolução dos modelos de serviço



Fonte: elaborada pelo autor.

No modelo FaaS, um ambiente de execução orientado a eventos permite a execução de blocos de código (funções) sem a necessidade de instanciar máquinas virtuais ou contêineres. A execução das funções é disparada por eventos, como requisições de usuários, alterações em um banco de dados, etc. Quando a execução da função é disparada, o provedor deverá provisionar os recursos computacionais necessários sob demanda. Dessa forma, a escalabilidade da aplicação é transparente para o cliente, pois o provedor é o responsável por provisionar os recursos sempre que for necessário executar uma função. Nesse caso, a tarifação é baseada na invocação das funções e não em termos de recursos computacionais alocados. As métricas de tarifação podem ser:

- O número de invocações, por exemplo, R\$ 0,50 para cada 1 milhão de requisições de execução de alguma função.
- Tempo de execução, por exemplo, R\$ 0,0000002 para cada 100 ms de tempo de execução de alguma função.

Assimile

Quando um cliente aloca uma máquina virtual em um provedor, ele é tarifado de acordo com o período pelo qual a máquina permanece alocada, mesmo que ela não seja usada. Por exemplo, se uma máquina virtual foi alocada para hospedar uma aplicação Web e, por um intervalo de tempo, nenhum usuário acessou a aplicação, ainda sim haverá cobrança de tarifa. Esse tipo de custo não acontece no modelo FaaS, pois se não houver nenhuma requisição, o custo do serviço será zero. Só há cobrança por invocação ou tempo de execução de uma função. Se nenhuma função for executada, não há valor a ser pago.

O modelo FaaS tem ganhado popularidade com os padrões de arquitetura de aplicações distribuídas que visam explorar a modularidade de software para viabilizar mecanismos sofisticados de replicação e tolerância a falhas. Entre os serviços FaaS de provedores de nuvem pública, podemos citar:

- AWS Lambda (AWS, 2019c).
- Google Cloud Functions (GOOGLE, 2019b).
- IBM Cloud Functions (IBM, 2019a).
- Azure Functions (MICROSOFT, 2019b).

Pesquise mais

Além de explicar conceitos básicos sobre *Serverless Computing* e FaaS, o trabalho a seguir apresenta uma comparação abrangente entre os serviços AWS Lambda e Azure Functions. Não deixe de consultá-lo:

PAULA, G. S. de. Avaliação de serviços Serverless: um experimento piloto. 2018. Trabalho de Conclusão de Curso (Graduação) - Universidade Tecnológica Federal do Paraná, p. 11-20. Dois Vizinhos, 2018.

A Computação sem Servidor (*Serverless Computing*) é um conceito importante e deve-se compreender como o modelo FaaS é uma alternativa de serviço de computação para execução de aplicações em nuvem. No caso do serviço FaaS, a aplicação que será executada em nuvem precisa ser modularizada, de forma que uma função no provedor em nuvem será criada para cada funcionalidade específica da aplicação. Por exemplo, em uma aplicação para gestão de eventos, poderia haver uma função para cadastro de evento, uma função para realizar inscrição em evento, etc. Dessa forma, a aplicação é vista como um conjunto de funções disponíveis no provedor. Cada função é considerada um componente de software independente. Portanto, as funções de uma mesma

aplicação podem ser implementadas em linguagens de programação diferentes. Por exemplo, o serviço IBM Cloud Functions dá suporte para funções implementadas em Node.js, Python, Swift, Go, PHP e Ruby. Esse serviço oferece ferramentas para criação, execução e monitoramento de funções na nuvem.

Cada função disponível em um provedor pode ser invocada de forma independente. A forma tradicional para um software cliente invocar uma função é através de uma requisição HTTP. Assim, para cada função criada o provedor gera um URL correspondente. Você acha que, para testar um a função, seria necessário implementar um software que faz uma requisição HTTP? Bem, essa é uma alternativa que tomaria muito tempo do desenvolvedor. Para agilizar o processo existem ferramentas que permitem a execução de requisições Web. Um exemplo é a ferramenta denominada Postman (POSTMAN, 2019), que pode ser usada para testar APIs Web. Com essa ferramenta é possível enviar diversos tipos de requisições HTTP, verificar ocorrência de erros, conteúdo da resposta e tempo de execução da requisição.

Refleta

O modelo denominado Funções como Serviço (FaaS) apresenta diversas vantagens interessantes, como o fato de dispensar o gerenciamento de servidores, a redução de custos com recursos de processamento e a alta escalabilidade das aplicações. No entanto, toda solução tecnológica tem suas desvantagens. Quais seriam as limitações ou desvantagens no uso de FaaS?

Em resumo, os três principais serviços de computação são: instâncias de máquinas virtuais, instâncias de contêineres e FaaS. Como discutido no decorrer desta seção, cada tipo de serviço pode ser mais interessante de acordo com os requisitos de cada aplicação, em termos de escalabilidade, nível de proteção, custos, etc. No Quadro 3.1, é apresentado um resumo dos exemplos desses serviços.

Quadro 3.1 | Exemplos de serviços de processamento de dados

Tipo de Serviço	Provedor	Serviço
Máquinas virtuais	AWS	Amazon Elastic Compute Cloud (Amazon EC2)
	Google Cloud	Google Cloud Compute Engine
	IBM Cloud	IBM Cloud Virtual Servers
	Microsoft Azure	Azure Virtual Machines
Contêineres	AWS	Amazon Elastic Container Service
	Google Cloud	Google Kubernetes Engine
	IBM Cloud	IBM Cloud Kubernetes Service
	Microsoft Azure	Azure Container Instances (ACI)

Tipo de Serviço	Provedor	Serviço
FaaS	AWS	AWS Lambda
	Google Cloud	Google Cloud Functions
	IBM Cloud	IBM Cloud Functions
	Microsoft Azure	Azure Functions

Fonte: elaborado pelo autor.

Em geral, as aplicações envolvem diversos componentes de software que podem ser executados por meio de serviços diferentes, que podem ser utilizados em provedores diversos ou em centros de dados variados de um mesmo provedor. Para lidar com a integração de componentes distribuídos e viabilizar mecanismos que visam melhorar o desempenho e a escalabilidade das aplicações, os serviços de rede são fundamentais.

Os serviços de rede compreendem mecanismos para configurar e monitorar a comunicação entre componentes de software e também para implementar funcionalidades de aplicações distribuídas. Além disso, os serviços de rede influenciam o funcionamento de outros mecanismos implementados no provedor, como balanceamento de carga, descoberta de serviços e tolerância a falhas. Os principais serviços de rede para aplicações em nuvem são:

- **DNS (*Domain Name System*):** serviço de nomes de domínio na Internet que permite identificar recursos na rede e direcionar requisições (TANENBAUM; STEEN, 2008). Com esse serviço, o cliente pode gerenciar os nomes do seu domínio.
- **Distribuição de Conteúdo:** serviço para replicação de conteúdo em diversos servidores de forma automática. O provedor mantém uma Rede de Entrega de Conteúdo (CDN – *Content Delivery Network*) com servidores dispersos por várias regiões (TANENBAUM, 2008). Essa rede é usada para replicar o conteúdo estático de aplicações (como fotos, documentos PDF, etc.). O serviço é muito utilizado para implementação de aplicações de streaming de áudio ou vídeo, o que permite diminuir a latência de acesso aos dados, pois o conteúdo estará replicado em um servidor próximo ao usuário. Como os dados estão replicados, a disponibilidade do serviço também é melhorada.
- **Balanceamento de Carga:** serviço que permite distribuir as requisições entre várias réplicas de uma aplicação. Representa um mecanismo fundamental para melhorar o desempenho e o grau de tolerância a falhas das aplicações.

- Nuvem Privada Virtual (VPC – *Virtual Private Cloud*): serviço que permite a criação de uma rede privada entre os recursos alocados no provedor. É possível configurar aspectos como roteamento, firewall e endereçamento IP das instâncias alocadas.

Para cada um dos serviços de rede explicados, no Quadro 3.2, são comparados alguns exemplos concretos de serviços oferecidos por provedores de nuvem pública conforme descritos em (AWS, 2019d; GOOGLE, 2019b).

Quadro 3.2 | Exemplos de serviços de rede

Categoria	Provedor	Serviço
CDN	AWS	Amazon CloudFront
	Google Cloud	Google Cloud CDN
DNS	AWS	Amazon Route 53
	Google Cloud	Google Cloud DNS
Balanceamento de Carga	AWS	Elastic Load Balancing
	Google Cloud	Google Cloud Load Balancing
VPC	AWS	Amazon Virtual Private Cloud (Amazon VPC)
	Google Cloud	Google Virtual Private Cloud (VPC) Network

Fonte: elaborado pelo autor.

Caro aluno, esse é o primeiro passo no estudo dos tipos de serviços ofertados por provedores de Computação em Nuvem. Foram apresentados os serviços de computação e de rede. Veremos, ainda, serviços de armazenamento e análise de dados e alguns serviços especializados, por exemplo, para aplicações multimídia e de IoT.

Sem medo de errar

Você trabalha como analista de TI em uma empresa de desenvolvimento de aplicativos móveis. Seu desafio é realizar uma avaliação da melhor opção para executar em nuvem uma API web que funciona como o backend de um aplicativo de compras para uma loja de departamento. A empresa não quer assumir responsabilidade de gerenciar instâncias de recursos virtualizados, pois sabe-se que a infraestrutura necessária é de larga escala, portanto, o controle total deve ficar com o provedor. Você é o responsável pela escolha de soluções para integração do aplicativo com a API que deve ser hospedada na nuvem. O resultado do seu trabalho você deve ser um relatório que especifica

o tipo de serviço apropriado para executar a API web, as métricas de custo envolvidas e dois orçamentos de provedores de grande porte.

Para a elaboração do relatório, o primeiro passo é determinar qual tipo de serviço de computação (processamento de dados) seria mais apropriado para executar a API web: máquinas virtuais, contêineres ou FaaS. Os serviços tradicionais de máquinas virtuais e contêineres oferecem os benefícios básicos do modelo de Computação em Nuvem. No entanto, eles exigem algum nível de controle e gerenciamento por parte cliente. Por exemplo, o cliente precisa escolher as imagens de máquinas virtuais que devem ser instanciadas ou configurar algum mecanismo para orquestração de contêineres. Além disso, a empresa precisa de um serviço altamente escalável, uma vez que não é possível estimar com precisão da carga de trabalho resultante dos usuários do aplicativo. Neste caso, modelo de Função como Serviço representa uma alternativa atraente por várias razões. A primeira é a possibilidade de redução de custos nos períodos em que o volume de acessos ao sistema reduz drasticamente, por exemplo durante a madrugada. Como o pagamento é pelo número de execuções, não haveria cobrança por alocação de recursos que podem permanecer ociosos durante esses períodos. Além disso, FaaS oferece alta escalabilidade e dispensa qualquer tarefa de gerenciamento ou controle sobre os recursos computacionais nos quais as funções serão executadas. Esses recursos devem ser alocados sob demanda pelo próprio provedor.

O segundo passo é caracterizar o modelo de tarifação desse serviço escolhido. Conforme discutimos durante esta seção, os preços dos serviços FaaS consideram duas métricas principais: o número de requisições realizadas e o tempo de execução. É importante salientar que muitos provedores estabelecem um tempo de execução mínimo para a cobrança. Por exemplo, se o tempo mínimo de cobrança for 50 ms, mesmo funções simples que executem em menor tempo, serão tarifadas em pelo menos 50 ms. O consumo de memória também é contabilizado em adição ao tempo de execução, de forma que a tarifa é em função do tempo, mas proporcional à quantidade de memória usada para executar a função. Além disso, pode haver cobrança pela transferência de dados resultantes da execução das funções. Por exemplo, se uma função consiste em gerar um relatório a partir de dados em um banco de dados, pode haver tarifas para transmissão dos dados para o dispositivo dos usuários que requisitaram o relatório.

Por fim, é preciso fazer um orçamento em provedores relevantes no mercado. Os resultados dependem do provedor escolhido. A título de exemplo, são apresentados a seguir os preços reais praticados por dois provedores cujos nomes serão preservados. Vamos considerar os provedores A e B. Os preços estão em dólar americano para efeito de comparação, mas os valores foram pesquisados para datacenters no Brasil.

- Custos do provedor A.
 - Preço por tempo de execução: \$0.000016/GB-s.
 - Preço pelo volume de requisições: \$0.20 para cada 1 milhão de requisições.
 - Cota gratuita: 400,000 GB-s e 1 milhão de requisições.
- Custos do provedor B.
 - Preço por tempo de execução: \$0.0000125/GB-s
 - Preço pelo volume de requisições: \$0.40 para cada 1 milhão de requisições.
 - Cota gratuita: 400,000 GB-s e 2 milhões de requisições.

Com os tópicos e questões discutidos nesta seção, você já possui a habilidade de avaliar os cenários e estimativas de custo para os principais tipos de serviços disponíveis para execução de aplicações em nuvem.

Avançando na prática

Comparativo de Preços

Suponha que um provedor ofereça dois serviços para execução de aplicações web: serviço PaaS baseado em contêineres e serviço FaaS. Considere ainda o seguinte modelo de tarifação. O provedor cobra R\$ 120 por mês por contêiner no modelo PaaS. No caso do serviço FaaS, a tarifa é R\$ 2 para cada mil execuções de qualquer função. Para executar a aplicação web no modelo PaaS, um único contêiner é suficiente, pois a aplicação é monolítica. A aplicação recebe uma média de 100 requisições por usuário mensalmente. Analise qual o modelo é mais vantajoso no caso a aplicação ter 100, 500 ou 1.000 usuários.

Resolução da situação-problema

Um ponto positivo do modelo PaaS considerado é que o custo da instância é previsível: R\$ 120, independentemente do número de requisições. No entanto, se o número de usuários é muito pequeno, a instância fica subutilizada e, portanto, há um desperdício de recursos, que o provedor não deixa de cobrar do cliente.

No caso do serviço FaaS, a tarifação é proporcional ao número de requisições, que disparam consequentemente a execução de alguma função da aplicação. Vejamos qual é o custo em função do número de usuários:

- Para 100 usuários, o número de requisições médio é $100 \times 100 = 10.000$. Assim, o custo mensal seria R\$ 20 ($10 \times \text{R\$ } 2$).
- Para 500 usuários, o número de requisições médio é $500 \times 100 = 50.000$. Assim, o custo mensal seria R\$ 100 ($50 \times \text{R\$ } 2$).
- Para 1.000 usuários, o número de requisições médio é $1.000 \times 100 = 100.000$. Assim, o custo mensal seria R\$ 200 ($100 \times \text{R\$ } 2$).

Podemos concluir, para 100 ou 500 usuários é mais econômico usar o serviço FaaS. Quando o número de usuários é 1.000, é mais econômico usar o serviço PaaS. Na verdade, pode-se observar que o modelo PaaS passa a ser mais interessante em relação ao custo quando o número de usuários passa de 600. No entanto, isso poderia significar uma degradação de desempenho, pois um único contêiner seria usado para atender todas as requisições. No caso do serviço FaaS, a escalabilidade é automática uma vez que o provedor é responsável por alocar os recursos necessários para executar as funções (seja qual for o volume de requisições).

Faça valer a pena

1. Máquinas virtuais e contêineres compreendem os serviços básicos de computação para execução de aplicações em nuvem. Neste contexto, avalie as afirmativas a seguir:

- I. Os provedores possuem um repositório de imagens de máquinas virtuais a partir das quais as instâncias podem ser rapidamente criadas.
- II. Depois que uma instância de máquina virtual é criada com a quantidade de memória e de núcleos de processamento solicitados, o cliente precisa instalar o sistema operacional de sua preferência na instância.
- III. A orquestração de contêineres é o mecanismo que permite a comunicação entre os contêineres que formam uma aplicação.

Considerando as afirmativas apresentadas, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.

- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

2. O modelo de Função como Serviço (FaaS) é uma das formas de se implementar Computação sem Servidor. Neste caso, todo o gerenciamento dos recursos necessários para execução da aplicação fica a cargo do provedor. As funções são a unidade escala para replicação dos componentes e tarifação dos serviços.

Sobre as vantagens do modelo FaaS, é correto o que se afirma em:

- a. O uso de FaaS reduz os custos, pois não são necessários recursos computacionais para executar a aplicação.
- b. O uso de FaaS facilita o gerenciamento, pois não é necessário integrar os módulos da aplicação.
- c. O uso de FaaS aumenta o grau de escalabilidade da aplicação, pois não é necessário instanciar recursos para executar a aplicação.
- d. O uso de FaaS permite reduzir os custos, pois não há cobrança por recursos computacionais alocados.
- e. O uso de FaaS aumenta o desempenho da aplicação, pois a execução das funções é mais rápida.

3. Os provedores de Computação em Nuvem oferecem diversos serviços de rede. Esses serviços podem ser utilizados juntamente com instâncias de computação para a implantação de aplicações distribuídas em ambiente de nuvem. Neste contexto, avalie as afirmativas a seguir:

- I. Os serviços de balanceamento de carga aumentam a tolerância a falhas e a segurança das aplicações.
- II. Os serviços de CDN contribuem para melhorar o desempenho e a disponibilidade das aplicações em nuvem.
- III. A nuvem Privada Virtual dispensa o uso de mecanismos de criptografia para garantir o sigilo dos dados das aplicações em nuvem pois os recursos alocados não são compartilhados com outros clientes.

Considerando as afirmativas apresentadas, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.

- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

Serviços de armazenamento e análise de dados

Diálogo aberto

Vamos continuar nossos estudos explorando novos tipos de serviços de Computação em Nuvem. Já discutimos alguns serviços de processamento de dados, ou seja, instâncias para executar aplicações, mas isso não é o bastante, pois a maioria das aplicações exige alguma solução para a persistência de dados, ou seja, para o armazenamento de dados em unidade de memória não volátil, como o disco rígido (HD – hard disk). Além disso, muitas aplicações também podem requerer mecanismos para análise de dados. Você já utilizou algum serviço de gerenciamento de dados em nuvem? Nosso objetivo aqui é conhecer tipos e exemplos de serviços em nuvem que podem ser utilizados para armazenar e analisar dados de aplicações.

Para facilitar a compreensão desses tópicos, vamos considerar um caso concreto, em que você tem um papel importante na empresa de desenvolvimento de aplicativos para dispositivos móveis. Como analista de TI, sua responsabilidade é escolher as tecnologias e serviços em nuvem que serão utilizados para o desenvolvimento de aplicativos na empresa. Um novo projeto sob sua coordenação envolve a implementação de um aplicativo para streaming de vídeo para uma empresa cliente, cujo negócio é a oferta de cursos online. A sua empresa não deseja manter uma infraestrutura de TI própria para armazenamento de dados, assim, você deve escolher soluções de armazenamento em nuvem para os dados do aplicativo de streaming de vídeo. Com isso, será necessário o armazenamento na nuvem dos vídeos que estarão disponíveis para os usuários do aplicativo, assim como as informações de cadastro dos vídeos, como título, descrição, data, entre outros atributos que descrevem cada um deles. Sua tarefa, então, é especificar um serviço de algum provedor de nuvem pública apropriado para armazenar os vídeos e os dados cadastrais correspondentes.

Esse é um grande desafio e, para tomar a decisão correta sobre a melhor solução para os dados em nuvem, é preciso conhecer as características de cada tipo de serviço. Diante disso, vamos descrever os tipos de serviços de armazenamento de dados e apresentar alguns exemplos que envolverão, também, serviços para análise de dados em nuvem. Ao final desta seção, você será capaz de comparar as diversas soluções para gerenciamento de dados em ambientes de Computação em Nuvem.

Bons estudos!

Não pode faltar

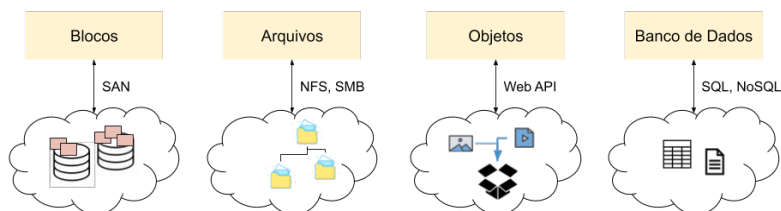
Na seção anterior, estudamos as principais alternativas em relação a recursos computacionais para execução de aplicações em nuvem e vimos que é possível usar, por exemplo, servidores virtuais ou contêineres para hospedar aplicações. Agora, no que diz respeito ao armazenamento de dados, você conhece soluções em nuvem? A utilização de serviços de armazenamento de dados em nuvem traz uma série de vantagens, como redução de custos; delegação (para o provedor) das tarefas de gerenciamento e otimização dos bancos de dados e escalabilidade (SOUSA *et al.*, 2010). Além disso, devido à elasticidade dos serviços em nuvem, a alocação de recursos de armazenamento de dados pode ser feita de forma incremental, com o pagamento pelo uso. Assim, um cliente pode alocar mais capacidade de armazenamento à medida que mais dados são gerados pelas aplicações.

Existe uma grande variedade de serviços de armazenamento de dados em nuvem. Para escolher o mais adequado, é importante avaliar como os dados devem ser estruturados, o volume de dados a ser manipulado e os requisitos da aplicação. Conforme explicado em (ERL; PUTTINI; MAHMOOD, 2013), existem quatro tipos de serviços de armazenamento de dados em ambientes de Computação em Nuvem:

- **Armazenamento em blocos:**
 - Alocação de blocos em discos de armazenamento disponibilizados na infraestrutura do provedor.
 - Permite ao cliente alocar blocos em unidades de armazenamento gerenciadas pelo provedor.
 - As unidades podem ser acessadas remotamente por meio de uma rede de alto desempenho dedicada para interconexão de equipamentos de armazenamento de dados. Esse tipo de rede é denominado Redes de Armazenamento de Dados (SAN- Storage Area Network) (NETO, 2007).
- **Armazenamento de arquivos:**
 - Leitura e escrita em diretórios com sistema de arquivos gerenciado pelo provedor.
 - O cliente do provedor de nuvem pode manipular arquivos remotamente utilizando tecnologias para sistemas de arquivos distribuídos, como o Network File System (NFS) (TANENBAUM, 2008) ou Server Message Block (SMB) (GONÇALVES, 2018).

- **Armazenamento de objetos:**
 - Repositório para itens de dados binários (como vídeos, imagens, etc.).
 - Solução para acesso a arquivos binários, como áudios ou fotos, por meio de um serviço web gerenciado pelo provedor.
- **Armazenamento de bases de dados:**
 - Sistemas de gerenciamento de banco de dados (SGBDs) mantidos pelo provedor. Há suporte tanto para SGBDs relacionais, quanto não relacionais.

Figura 3.2 | Tipos de serviços de armazenamento em nuvem



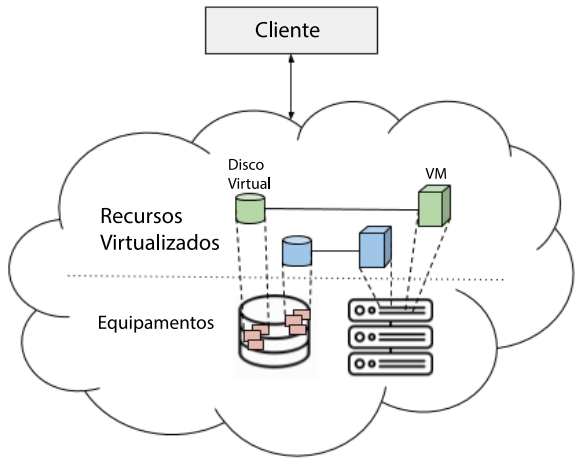
Fonte: elaborada pelo autor.

O armazenamento em blocos é o modelo com menor nível de abstração entre as alternativas para persistência de dados em provedores de nuvem (ERL; PUTTINI; MAHMOOD, 2013). De fato, o bloco é a unidade lógica mínima de armazenamento de dados oferecida por um sistema operacional em se tratando de gerenciamento de unidades de memória secundária. No ambiente de nuvem, o provedor utiliza também técnicas de virtualização para criar unidades de armazenamento virtuais em dispositivos físicos de armazenamento, da mesma maneira como as máquinas virtuais podem ser criadas em servidores físicos. Os dispositivos de armazenamento físicos utilizados pelos provedores incluem HDs (Hard Disks) tradicionais, do tipo disco magnético, e também unidades do tipo SSD (solid-state drive). As unidades virtuais de armazenamento em bloco oferecidas pelos provedores são criadas sobre esses recursos físicos para servir como discos das máquinas virtuais ou contêineres.

Em geral, a principal finalidade dos serviços de armazenamento em bloco na nuvem é servir como discos para instâncias como máquinas virtuais (VMs) ou contêineres, como ilustrado na Figura 3.3. Essas instâncias precisam de capacidade de armazenamento de dados para executar as aplicações. Então, quando a criação de uma instância é solicitada, o provedor cria

automaticamente uma unidade virtual de armazenamento em bloco, que é usada como disco dessa instância, ou seja, a unidade de armazenamento em bloco é o disco da máquina virtual.

Figura 3.3 | Tipos de serviços de armazenamento em nuvem



Fonte: elaborada pelo autor.

No entanto, as unidades de armazenamento em bloco também podem ser usadas em outras soluções, por exemplo, para implantação de algum mecanismo automático de backup de dados na nuvem. No Quadro 3.3, são apresentados alguns serviços em nuvem de armazenamento em bloco.

Quadro 3.3 | Exemplos de serviços de armazenamento em bloco

Provedor	Serviço.
AWS	Amazon Elastic Block Store (Amazon EBS) (AWS, 2019b).
Google Cloud	Persistent Disk (GOOGLE, 2019c).
Microsoft Azure	Azure Disk Storage (MICROSOFT, 2019b).

Fonte: elaborado pelo autor.

O segundo tipo de serviço é o armazenamento de arquivos em nuvem. Nesse caso, a unidade de armazenamento são os arquivos (ERL; PUTTINI; MAHMOOD, 2013). O acesso ao serviço de armazenamento é feito por meio de alguma tecnologia de sistema de arquivos distribuídos, como NFS ou SMB. Esse tipo de serviço é bastante utilizado para compartilhamento de arquivos em rede, por exemplo, para manter o diretório home dos usuários de uma rede corporativa, com as vantagens da escalabilidade e disponibilidade do

armazenamento em nuvem. Assim, os usuários podem acessar seus arquivos de qualquer dispositivo com acesso à Internet. Exemplos de serviços em nuvem para armazenamento de arquivos são relacionados no Quadro 3.4.

Quadro 3.4 | Exemplos de serviços de armazenamento arquivo

Provedor	Serviço.
AWS	Amazon Elastic File System (Amazon EFS) (AWS, 2019c).
Google Cloud	Cloud Filestore (GOOGLE, 2019d).
Microsoft Azure	Azure Files (MICROSOFT, 2019b).

Fonte: elaborado pelo autor.

Assimile

Os modelos de armazenamento (blocos e arquivos) são tipicamente utilizados para implantação de infraestruturas em nuvem, por exemplo, a partir de tecnologias de sistemas de arquivos distribuídos. Os outros dois modelos de armazenamento (objetos e bases de dados), explicados a seguir, apresentam um maior nível de abstração e são geralmente utilizados com plataformas para implementação de aplicações em nuvem. Neste caso, o cliente do serviço não precisa se preocupar com a infraestrutura de recursos de armazenamento, pois isso é responsabilidade do provedor.

O terceiro tipo de serviço de armazenamento em nuvem é o armazenamento de objetos. Neste caso, as unidades de armazenamento são objetos que são gerenciados como recursos web (ERL; PUTTINI; MAHMOOD, 2013). Isso significa que o acesso aos dados pode ser feito na forma de requisições HTTP. Os provedores suportam armazenamento de objetos de vários tipos como imagens, arquivos executáveis, vídeos, entre outros. Além disso, os provedores oferecem uma API para que as aplicações possam manipular os objetos por meio de requisições na web, em vez de usar linguagens de consultas comuns em gerenciadores de bancos de dados. Exemplos de serviços em nuvem para armazenamento de objetos são apresentados no Quadro 3.5.

Quadro 3.5 | Exemplos de serviços de armazenamento de objetos

Provedor	Serviço
AWS	Amazon Simple Storage Service (Amazon S3) (AWS, 2019h)
Google Cloud	Google Cloud Storage (GOOGLE, 2019i)
Microsoft Azure	Azure Blob Storage (MICROSOFT, 2019a)

Fonte: elaborado pelo autor.

Por fim, o quarto tipo de serviço de armazenamento em nuvem é o armazenamento de bases de dados. Esse modelo corresponde aos serviços de banco de dados em nuvem ou Banco de Dados como Serviço (DBaaS - Database as a Service). São serviços que, em geral, suportam algum tipo de linguagem de consulta, além de operações básicas de escrita ou leitura de dados (ERL; PUTTINI; MAHMOOD, 2013). No Quadro 3.6, são apresentados exemplos de serviços de banco de dados em nuvem para duas categorias de Sistema Gerenciador de Banco de Dados (SGBD): relacional e não relacional. Como explicado em (SOUZA et al., 2014), os bancos de dados relacionais são aqueles que utilizam a álgebra relacional para suporte à consistência forte, armazenam dados de forma estruturada e implementam alguma linguagem de consulta estruturada (SQL - Structured Query Language).

Quadro 3.6 | Exemplos de serviços de banco de dados em nuvem

Provedor	Categoria	Serviço
AWS	Relacional	Amazon RDS (AWS, 2019g)
	Não relacional	Amazon DynamoDB (AWS, 2019a)
Google Cloud	Relacional	Cloud SQL (GOOGLE, 2019e)
	Não relacional	Cloud Firestore (GOOGLE, 2019d)
Microsoft Azure	Relacional	Azure SQL Database (MICROSOFT, 2019e)
	Não relacional	Azure Cosmos DB (MICROSOFT, 2019d)
MongoDB	Não relacional	MongoDB Atlas (MONGODB, 2019)

Fonte: elaborado pelo autor.

O baixo desempenho dos bancos de dados relacionais para gerenciamento de grandes volumes de dados motivou o desenvolvimento dos bancos de dados não relacionais, também denominados “não apenas SQL” (NoSQL – not only SQL) (SOUZA et al., 2014). Os bancos não relacionais apresentam desempenho e escalabilidade significativamente superior para lidar com grandes volumes de dados não estruturados, o que é muito importante, por exemplo, para aplicações web em larga escala.

Refleta

Os bancos de dados NoSQL em nuvem, por exemplo o Google BigTable (GOOGLE, 2019a) e o MongoDB Atlas (MONGODB, 2019), têm sido muito utilizados devido ao alto grau de escalabilidade e desempenho. No entanto, os bancos NoSQL não suportam o mesmo nível de consistência dos bancos relacionais. Quais são as implicações disso? Esse modelo de banco de dados em nuvem é apropriado para quais tipos de aplicações?

Existem vários modelos de banco de dados NoSQL, entre os principais podemos citar: orientado a documentos, orientado a grafos, chave-valor e orientado a colunas. Para os bancos NoSQL listados no Quadro 3.7, temos a seguinte classificação: Amazon DynamoDB suporta dois modelos, chave-valor e orientado a documentos; o Google Cloud Firestore é um banco orientado a documentos; o Azure Cosmos DB suporta múltiplos modelos. O MongoDB é um banco orientado a documentos. Muitas tecnologias de bancos de dados NoSQL são proprietárias e mesmo os bancos de dados que seguem um mesmo modelo podem não ser compatíveis. Isso pode resultar em problemas de interoperabilidade e dificultar a portabilidade de dados em ambientes de nuvem (ERL; PUTTINI; MAHMOOD, 2013). Essa é uma preocupação que um arquiteto de soluções em nuvem sempre deve ter em mente.

Ao usar um serviço de banco de dados em nuvem, o cliente delega para o provedor as tarefas de alocar a infraestrutura necessária para executar o SGBD e para armazenar os dados. Assim, a aplicação e o banco de dados serão executados em servidores diferentes. A aplicação acessa o banco de dados por meio da rede, para solicitar a escrita ou leitura de dados. A aplicação precisa então de uma string de conexão, ou seja, de uma URL que serve como endereço para um servidor onde o banco está disponível.

Como foi discutido no decorrer desta seção, existe uma grande variedade de serviços de armazenamento em nuvem para as mais diversas aplicações. Esses serviços apresentam muitos benefícios, como alta escalabilidade e disponibilidade, assim como a alocação incremental de capacidade de armazenamento com pagamento baseado no uso. No entanto, como profissionais, precisamos também estar atentos às dificuldades ou desafios envolvidos no uso desse tipo de serviço (SOUSA et al., 2010).

A principal barreira ao uso de armazenamento em nuvem pública diz respeito a aspectos de segurança e privacidade, pois os dados são acessados por meio da Internet. É imprescindível que os provedores ofereçam mecanismos de segurança para proteger os dados dos clientes. Devem ser implementadas soluções para garantir o sigilo de informações sensíveis, como dados financeiros ou médicos de indivíduos, ou informações com segredo industrial de empresas.

Pesquise mais

O texto a seguir apresenta os conceitos básicos sobre segurança e privacidade de dados em nuvem, assim como uma discussão sobre algumas vulnerabilidades de soluções em nuvem e os principais aspectos de segurança em relação aos dados, que devem ser observados na implementação de aplicações.

BRANCO JR, E. C.; MACHADO, J. C.; MONTEIRO, J. M. Estratégias para proteção da privacidade de dados armazenados na nuvem. *In: Anais do 29 Simpósio Brasileiro de Banco de Dados*. Curitiba: Sociedade Brasileira de Computação, 2014. p. 46–52.

Outro desafio para armazenamento em nuvem é o desempenho, em termos da latência de acesso remoto aos dados. Se uma aplicação precisa acessar dados que estão armazenados em outro provedor, o tempo de transferência pode comprometer o desempenho da solução. Além disso, o processamento de consultas a dados em nuvem são tarefas complexas, o que exige significativa capacidade computacional para garantir um desempenho satisfatório.

Pesquise mais

O texto a seguir discute as principais características do gerenciamento de dados em ambientes de Computação em Nuvem e apresenta uma análise comparativa entre diversos serviços de armazenamento de dados.
SOUSA, F. R. C. *et al.* Gerenciamento de dados em nuvem: Conceitos, sistemas e desafios. *In: PEREIRA, A. C. M. et al. (org.). Tópicos em sistemas colaborativos, interativos, multimídia, web e bancos de dados*. Belo Horizonte: Sociedade Brasileira de Computação (SBC), 2010. Cap. 4, p. 108-121.

Foram apresentados quatro tipos de serviços de armazenamento: blocos, arquivos, objetos e bancos de dados. Cada um desses serviços é mais apropriado para diferentes requisitos das aplicações. O Quadro 3.7 apresenta um resumo desses tipos com suas características e exemplos típicos de aplicação.

Quadro 3.7 | Resumo dos tipos de serviços de armazenamento em nuvem

Serviço	Características	Exemplo de uso
Blocos	Alocação de blocos de unidades de armazenamento, com alto nível de controle sobre os recursos virtualizados.	Os blocos podem ser alocados para montar um disco virtual para uma máquina virtual. O cliente pode, por exemplo, escolher formatar o disco virtual com um sistema de arquivos específico.
Arquivos	Leitura e escrita de arquivos em um sistema de arquivos gerenciado pelo provedor.	Compartilhamento de arquivos em rede. Implementação de servidor de arquivos em uma rede corporativa.

Serviço	Características	Exemplo de uso
Objetos	Armazenamento de itens de dados binários de forma escalável utilizando serviços web.	Em uma aplicação de comércio eletrônico, um serviço de armazenamento de objetos pode ser utilizado para armazenar as fotos dos produtos.
Banco de dados	SGBD gerenciado pelo provedor. O cliente não precisa criar uma máquina virtual e instalar o SGBD. Serviço confiável e escalável, com mecanismos automáticos para replicação e backup dos bancos de dados.	Bancos de dados estruturados ou bancos de dados NoSQL para quaisquer aplicações.

Fonte: elaborado pelo autor.

Em se tratando de armazenamento de dados em nuvem, não podemos deixar de falar no conceito de Big Data, que corresponde ao estudo e desenvolvimento de soluções para armazenamento, transferência e análise de grandes volumes de dados (COSTA, 2012). O que caracteriza o cenário de Big Data é a velocidade que grandes massas de dados são geradas e também a variedades dos tipos de dados, que, na maioria das vezes, não são estruturados. Isso é o resultado do crescimento exponencial de dados na Internet, com cada vez mais aplicações em escala mundial, aplicações de streaming multimídia, soluções de Internet das Coisas, entre outras fontes que geram dados a uma taxa crescente. Por exemplo, o sistema de vendas da maior empresa varejista dos EUA gera um volume de 2,5 PB (Petabyte) de dados por hora (OUSSOUS, 2018).

Exemplificando

Existem aplicações de Big Data em diversas áreas, como inteligência de negócios, fazendas inteligentes e Indústria 4.0. Um exemplo de aplicação é a soluções para planejamento urbano em regiões metropolitanas apresentada em (RATHORE, 2016). Nessa aplicação, um grande volume de dados é captado em tempo real a partir de câmeras de vídeo, smartphones, veículos e diversos sensores dispersos pela cidade. Esses dados são analisados para dar suporte a decisões estratégicas de gerenciamento do transporte coletivo, controle de poluição e gerenciamento de vagas de estacionamento.

O ambiente de nuvem apresenta diversas vantagens para o desenvolvimento de soluções de Big Data, entre as quais podemos destacar: elasticidade, redução de custos e agilidade na alocação de recursos computacionais em larga escala (COSTA, 2012). De fato, há argumentos de que Computação em Nuvem e Big Data são paradigmas inseparáveis (HASHEM *et al.*, 2015). A

razão disso é que as infraestruturas de Computação em Nuvem representam o ambiente computacional que atende aos requisitos para implementação de soluções de Big Data, que exigem o processamento distribuído de grandes volumes de dados de forma eficiente.

Essas soluções, em geral, envolvem três tarefas principais: armazenamento, análise e visualização de dados. Já apresentamos alguns exemplos de serviços em nuvem para armazenamento de dados; vamos, então, conhecer soluções para análise de dados. No Quadro 3.8 são descritos exemplos de serviços para análise de dados oferecidos pelos principais provedores de nuvem pública.

Quadro 3.8 | Exemplos de serviços para análise de dados

Provedor	Serviço	Descrição
AWS	Amazon EMR (AWS, 2019d)	O Amazon Elastic MapReduce (ERM) permite a criação dinâmica de um <i>cluster</i> com instâncias de máquinas virtuais para processamento distribuído de dados utilizando tecnologias como Hadoop (APACHE, 2019), Hive (HUAL <i>et al.</i> , 2014) e Spark (ZAHARIA, 2016).
	Amazon Kinesis (AWS, 2019g)	Solução para análise em tempo real de dados de streaming, como dados de telemetria da IoT, logs de aplicativos e streaming de vídeo.
Google Cloud	BigQuery (GOOGLE, 2019a)	Solução escalável no modelo <i>serverless</i> para aplicações de inteligência de negócios (BI – <i>Business Intelligence</i>).
	Cloud Data Fusion (GOOGLE, 2019b)	Serviço para extração, transformação e integração de dados.
Microsoft Azure	HDInsight (MICROSOFT, 2019g)	Serviço de alta disponibilidade e escalabilidade para processamento distribuído de dados utilizando <i>frameworks</i> de código aberto, como Spark, Kafka e Hadoop.
	Data Factory (MICROSOFT, 2019f)	Serviço para extração, transformação e integração de dados.

Fonte: elaborado pelo autor.

Exemplificando

Outra componente importante em soluções Big Data é a visualização de dados. Os provedores em nuvem também oferecem serviços para essa finalidade. Entre exemplos de serviços para visualização de dados podemos citar o Amazon QuickSight (AWS, 2019f) e o Google Data Studio (GOOGLE, 2019c). Essas ferramentas permitem, por exemplo, a criação de painéis de bordo (*dashboards*) e relatórios interativos.

Chegamos ao fim desta seção, na qual foram discutidos diversos serviços em nuvem para armazenamento e análise de dados. Com este estudo, você aprimorou

suas habilidades na direção de criar soluções em ambientes de Computação em Nuvem e conheceu as importantes opções para o gerenciamento de dados.

Sem medo de errar

Você precisa tomar decisões de projeto em relação ao desenvolvimento do um aplicativo para streaming de vídeo para uma empresa de cursos online. Sua tarefa é escolher soluções de armazenamento em provedores de Computação em Nuvem para os dados do aplicativo. Assim, quando os usuários acessarem o aplicativo, os dados serão buscados no serviço de armazenamento em nuvem. Sabemos que há dois tipos de dados a serem armazenados: os vídeos e os dados cadastrais desses vídeos, que vamos chamar de metadados. Os metadados são descrições textuais que exigem pouco espaço de armazenamento.

Nesse sentido, primeiramente, é importante estabelecer que podemos usar serviços de armazenamento diferentes para os vídeos e os metadados, caso necessário. Essa é de fato a estratégia mais apropriada pois são dados de natureza diferente. Os vídeos são dados binários que em geral ocupam muito espaço.

Vamos primeiro determinar a melhor solução para o armazenamento de vídeos. Conhecemos quatro tipos de serviços de armazenamento: em blocos, arquivo, objetos e banco de dados. Os dois primeiros são mais adequados para sistemas de baixo nível que exigem infraestrutura de armazenamento. Armazenamento de objetos e bancos de dados são soluções voltadas principalmente para desenvolvedores de software. É possível armazenar vídeos em bancos de dados, mas esses não são otimizados para esse fim. A melhor solução é de fato o armazenamento de objetos. Esse tipo de serviço foi criado com foco em armazenamento de dados binários estáticos como fotos, vídeos ou arquivos executáveis, por exemplo. Não existe uma única resposta. Existem várias opções desse tipo de serviço, como aquelas que foram citadas em nosso estudo: Amazon S3, Google Cloud Storage e Azure Blob Storage. A escolha por um serviço em particular dependerá de outros requisitos como custos, funcionalidades disponíveis, formas de acesso ao serviço, SDKs disponíveis, entre outros.

O segundo passo é escolher o serviço de armazenamento para os metadados. Neste caso, a escolha para o desenvolvimento do aplicativo é um serviço de banco de dados. Novamente, não temos uma única opção de resposta correta aqui. Mas, se escalabilidade e desempenho forem requisitos importantes para o aplicativo, assim como agilidade no processo de desenvolvimento, o melhor seria escolher algum serviço de banco de dados NoSQL, por exemplo, o Azure Cosmos DB ou Amazon DynamoDB. É importante ressaltar a necessidade de integração entre os dois serviços de armazenamento. Ou seja, como

valos relacionar o vídeo no serviço de armazenamento de objetos com seus metadados no banco de dados. Em geral, os serviços de armazenamento de objetos geram uma URL que identifica cada objeto. Então, esse problema pode ser resolvido a partir da inclusão da URL do vídeo junto com seus dados cadastrais no banco de dados. Assim, quando o aplicativo recuperar os dados de um vídeo no banco de dados, ele usa a URL para iniciar a streaming do vídeo a partir do serviço de armazenamento de objetos.

A escolha de serviços em nuvem não é uma tarefa fácil, pois a variedade de serviços é extensa, cada aplicação tem seus requisitos específicos e a estimativa de custos é complexa. Mas, com o conteúdo aprendido até aqui, você já é capaz de projetar soluções adequadas em ambientes de Computação em Nuvem, desde a hospedagem de aplicações até a escolha de serviços de armazenamento de dados.

Avançando na prática

Serviços para Backup

Considere que você é um analista de TI em uma empresa que atua no mercado de telecomunicações. Para melhorar a confiabilidade dos serviços, essa empresa pretende implantar um novo software de backup de dados. Para essa solução de backup os requisitos mais importantes são custo e escalabilidade, pois o volume de dados da empresa está crescendo devido a uma expansão da sua área de atuação. Seu papel é determinar um modelo de armazenamento para a software de backup dos dados da empresa.

Resolução da situação-problema

Em geral, as soluções de backup exigem crescente capacidade de recursos computacionais. Isso resulta em altos investimentos, principalmente em dispositivo de armazenamento de dados. Como os principais requisitos são escalabilidade e redução de custos, uma alternativa atraente seria utilizar serviços de armazenamento em nuvem. A razão é que a capacidade de armazenamento pode ser contratada de forma incremental, de acordo com a demanda. Além disso, não seriam necessários investimentos para adquirir e manter a infraestrutura de unidades de armazenamento.

Entre os tipos de serviços de armazenamento em nuvem, deve-se escolher algum caracterizado por um alto nível de controle por parte do cliente. Isso é necessário porque o software de backup precisa formatar as unidades lógicas

de armazenamento. Esta característica é própria dos serviços de armazenamento em blocos. A empresa pode então contratar algum serviço desse tipo, por exemplo, Amazon EBS, Google Persistent Disk ou Azure Disk Storage. O volume contratado seria então disponibilizado como uma unidade lógica de armazenamento para o software de backup. Essa unidade poderia ser formatada e utilizada para armazenar os arquivos de backup gerados pelo software. O volume poderia ter sua capacidade expandida na medida do necessário. Neste caso, é importante preocupar com criptografia para garantir sigilo dos dados de backup de enviados para a unidade de armazenamento no provedor e também com a qualidade da rede para que o backup seja realizado de forma rápida. É importante ainda salientar que existem serviços gerenciados de backup em nuvem, o que pode ser interessantes para muitas empresas que querem um serviço de mais alto nível de abstração, gerenciados pelo provedor. Exemplos desse serviço são: AWS Backup e Azure Backup.

Faça valer a pena

1. Considere uma aplicação em nuvem cujo objetivo é monitorar os equipamentos de uma fábrica de automóveis. A aplicação recebe dados em tempo real de sensores instalados nos equipamentos e precisa analisá-los para fazer um diagnóstico dos equipamentos, indicar possibilidades de falhas e sugerir planos de manutenção.

O serviço em nuvem apropriado para o tipo de análise de dados descrita é:

- a. Amazon Elastic Block Store.
- b. Cloud Filestore.
- c. Azure Blob Storage
- d. Cloud SQL.
- e. Amazon Kinesis.

2. Existe uma grande variedade de serviços em nuvem para armazenamento de dados, desde armazenamento em blocos a bancos de dados completamente gerenciados pelo provedor. Nesse contexto, avalie as afirmativas a seguir:

- I. Ao usar Banco de Dados como Serviço, o cliente do provedor não precisa se preocupar com o gerenciamento da infraestrutura necessária para armazenar os dados.
- II. É adequado uma aplicação usar um serviço de armazenamento de objetos para armazenar fotos de perfil dos usuários cadastrados,

mesmo que os demais dados, como nome, e-mail e telefone, sejam armazenados em um banco de dados relacional.

- III. Os serviços de armazenamento de arquivos devem ser utilizados para prover as unidades (discos) de armazenamento das máquinas virtuais em um provedor de nuvem.

Considerando as afirmativas apresentadas, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

3. Os serviços para armazenamento e análise de dados podem trazer uma série de vantagens para aplicações em nuvem. No entanto, ainda existem alguns desafios no uso de tais serviços. Considerando as possíveis desvantagens e dificuldades relacionadas com esses serviços, avalie as afirmativas a seguir:

- I. A segurança não é um problema no caso de serviços DBaaS, já que o provedor é responsável por gerenciar e configurar os recursos computacionais necessários.
- II. Uma baixa qualidade de acesso à Internet pode ser um sério problema no uso de serviços de armazenamento de dados em nuvem pois pode resultar em maior atraso médio no acesso aos dados que estão no provedor.
- III. O uso de virtualização pelos provedores inviabiliza o uso de mecanismos de replicação e consistência dos bancos de dados em nuvem.

Considerando as afirmativas apresentadas, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

Soluções em nuvem

Diálogo aberto

Caro aluno, já estudamos vários tipos de serviços para execução de aplicações e para armazenamento e análise de dados, que atendem a demandas gerais da maioria das aplicações. No entanto, os provedores disponibilizam também serviços especializados para aplicações em domínios específicos. Por exemplo, se você precisa desenvolver uma solução baseada em Inteligência Artificial (IA) para recomendações de produtos em uma loja de comércio eletrônico ou um aplicativo de transporte, será que os provedores oferecem serviços em suporte a essas finalidades?

Para motivar nosso estudo, vamos reconsiderar o cenário no qual você é um analista de TI em uma empresa cujo foco é o desenvolvimento de aplicativos para dispositivos móveis. Sua atuação principal na empresa é coordenar a avaliação e a implementação de soluções em nuvem para os aplicativos a serem desenvolvidos. Considere que a empresa conseguiu um novo contrato para o desenvolvimento de um aplicativo de streaming de músicas. Considerando que já vários produtos no mercado com essa finalidade, é necessário buscar um diferencial competitivo. Nesse sentido, o objetivo é aprimorar o aplicativo com um sistema sofisticado de recomendação de músicas para o usuário, que deve ser baseado em aspectos que envolverão as seguintes etapas: primeiramente, o sistema vai selecionar as preferências e escolhas do próprio usuário e comparar o seu estilo musical. Em seguida, as recomendações podem considerar a localização dos usuários, pois, na ausência de dados suficientes sobre suas preferências, ao iniciar o aplicativo, pode-se considerar a preferência musical mais popular na região em que ele se encontra. O projeto e implementação de um mecanismo de recomendação pode demorar muito e para agilizar a implementação da solução sem perder a qualidade, a estratégia será buscar serviços disponíveis em provedores de Computação em Nuvem que podem ser utilizados para construir a solução proposta de forma ágil e confiável. Nesse contexto, sua tarefa é entregar um documento com a especificação dos principais requisitos funcionais do aplicativo e a indicação de serviços em nuvem para viabilizar a implementação desses requisitos.

Para ajudá-lo a lidar com essas questões, esta seção dá continuidade ao nosso estudo sobre os diversos tipos de serviços disponibilizados pelos provedores em nuvem, com o objetivo de conhecer aplicações especializadas para algumas áreas importantes. Vamos começar com uma descrição sobre serviços relacionados à Internet das Coisas (IoT – *Internet of Things*). Em

seguida, vamos abordar soluções para o Aprendizado de Máquina (*Machine Learning*) e aplicações multimídia. Por fim, vamos concluir a seção com uma discussão sobre serviços em nuvem para suporte ao desenvolvimento de software, por exemplo, ferramentas em nuvem para automatizar testes e implementar aplicações. Ao final desta seção, você conhecerá uma grande variedade de serviços que podem ser utilizados para enriquecer e sofisticar as funcionalidades, assim como as soluções em nuvem para tornar o processo de desenvolvimento de aplicações mais produtivo.

Bons estudos!

Não pode faltar

Vamos continuar nosso estudo sobre os principais tipos de soluções ofertadas por provedores de Computação em Nuvem. Nas seções anteriores, vimos os serviços de processamento de dados que incluem instâncias para execução de aplicações, como máquinas virtuais e contêineres. Além disso, conhecemos os bancos de dados como serviço e os principais tipos de armazenamento de dados em nuvem, entre eles, o de objetos. Por serem bastante genéricos, todos esses serviços atendem à maioria das aplicações; entretanto, algumas delas possuem propósitos específicos que exigem mecanismos especializados, por exemplo, aquelas que manipulam conteúdo multimídia, aplicações em Internet das Coisas (IoT – *Internet of Things*) ou Inteligência Artificial. Será que os provedores de Computação em Nuvem oferecem caminhos tão especializados? A resposta é sim e nosso objetivo nesta seção é conhecer alguns deles.

Serviços para aplicações de IoT na nuvem

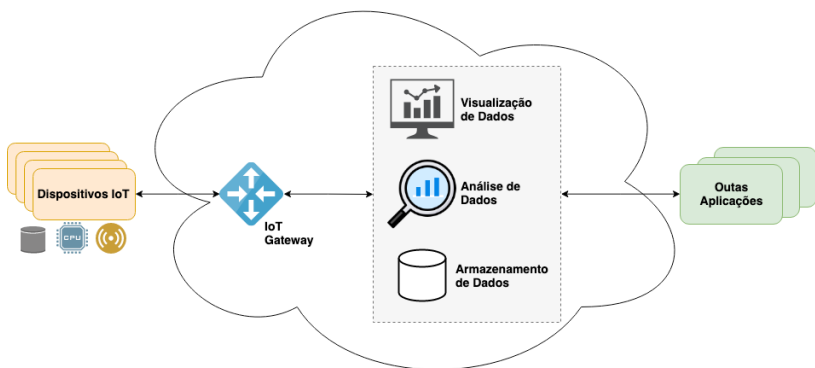
A Internet das Coisas pode ser entendida como um cenário no qual qualquer objeto com capacidade computacional e de comunicação pode se conectar à Internet (SANTOS *et al.*, 2016), o que possibilitou o surgimento de novas aplicações que envolvem a comunicação com ou entre esses objetos inteligentes. Podemos mencionar, inclusive, novos importantes conceitos decorrentes dos avanços em IoT, como Cidades Inteligentes (KON; SANTANA, 2016) e Indústria 4.0 (AZEVEDO, 2017). O conceito de Cidades Inteligentes consiste no uso de tecnologias da informação e comunicação para tornar mais eficiente a gestão dos centros urbanos, por exemplo, com soluções para mobilidade e controle de iluminação pública. De forma análoga, o paradigma da Indústria 4.0 envolve o uso de tecnologias para inovação nos processos industriais, por exemplo, em soluções de monitoramento e manutenção preventiva de equipamentos em fábricas, ou soluções de logística e rastreamento de produtos.

Assimile

A Cidade do Cabo, capital da África do Sul, teve em 2018 o pior período de seca já registrado. Uma empresa nessa cidade criou uma solução mais precisa do que as existentes para o monitoramento do consumo residencial de água. Ela foi elaborada com serviços de IoT e serviços de armazenamento e análise de dados de um provedor de Computação em Nuvem. Dessa forma, permite-se que o consumo de água das residências seja acompanhado em tempo real por meio de um aplicativo e de uma aplicação Web, favorecendo a economia de água, uma vez que as pessoas podem verificar o consumo mais facilmente e de forma mais precisa, o que permite reduzir custos e, o mais importante, beneficia o meio-ambiente. Os detalhes técnicos da solução são descritos em (CHARLTON, 2019).

Existem diversos serviços em nuvem para a implementação de aplicações de IoT. Para entender a sua utilidade, precisamos conhecer a arquitetura geral de aplicações IoT que utilizam recursos na nuvem, como ilustrado na Figura 3.4. Os dispositivos IoT são equipados com sensores, processadores e interfaces de comunicação sem fio. Um gateway é utilizado para gerenciar um canal seguro de comunicação entre os dispositivos e os serviços em nuvem. Os dados coletados podem ser utilizados de várias formas, inclusive, por meio de integração com outras aplicações. Os provedores em nuvem oferecem, por exemplo, serviços para coleta, análise e visualização de dados dos dispositivos conectados.

Figura 3.4 | Visão geral da arquitetura de uma aplicação IoT



Fonte: elaborada pelo autor.

O Quadro 3.9 descreve alguns serviços em nuvem para suporte a aplicações de IoT. Por exemplo, o serviço AWS IoT *SiteWise* é voltado para o suporte a soluções para Indústria 4.0, a fim de aprimorar a eficiência de equipamentos e processos. Um caso de uso típico é a coleta de dados de máquinas e robôs em linhas de produção e montagem, enviando-os para análise na nuvem. Relatórios com diagnóstico e indicadores de desempenho podem ser gerados para previsão de falhas nos equipamentos e identificação de oportunidades de otimização dos processos. Outro caso é uma empresa de distribuição de energia elétrica que poderia usar esse serviço para construir uma aplicação de monitoramento remoto dos equipamentos da rede. Para aplicações de IoT que usam dados de localização, uma opção de serviço é o Azure Maps, que pode ser usado na área de mobilidade urbana e sistemas de rastreamento de objetos ou veículos. O serviço Google IoT Core pode ser utilizado como um IoT Gateway (Figura 3.4) para gerenciamento dos dispositivos e coleta de dados.

Quadro 3.9 | Exemplos de serviços para aplicações de IoT

Serviço	Descrição
AWS IoT SiteWise (AWS, 2019f)	Serviço para análise de dados e monitoramento de instalações industriais.
AWS IoT Device Defender (AWS, 2019e)	Solução para configuração de mecanismos de segurança na comunicação de dispositivos de IoT entre si e com a nuvem. Pode ser utilizado para implementação de estratégias de autenticação de dispositivos e criptografia de dados em uma aplicação de IoT.
Cloud IoT Core (GOOGLE, 2019)	Plataforma para coleta, análise e visualização de dados de dispositivos IoT em tempo real.
Azure Maps (MICROSOFT, 2019c)	Mecanismos para manipulação de dados geoespaciais para suporte à inteligência baseada em localização.
Azure Time Series Insights (MICROSOFT, 2019b)	Análise de dados de séries temporais coletados por aplicações IoT em larga escala. Esses recursos são importantes, por exemplo, para aplicações de sensoriamento remoto ou aplicações que trabalham com previsões e identificação de tendências em grandes volumes de dados.

Fonte: elaborado pelo autor.

Exemplificando

Diversos objetos inteligentes podem fazer parte de uma aplicação IoT, como carros, relógios, *smartphones*, entre outros. E o que caracteriza um objeto como inteligente? Objetos inteligentes são aqueles com capacidade de processar, armazenar e transmitir dados. Em geral, os dispositivos IoT também incluem sensores e atuadores para ações de monitoramento e controle. Para tornar um objeto inteligente, podemos acoplar a ele uma placa controladora, como o Arduino, que é uma plata-

forma aberta com diversos componentes que vão desde várias configurações de placas controladoras, até serviços de computação em nuvem para armazenamento e processamento dos dados (ARDUINO, 2019). Uma placa Arduino pode ser usada, por exemplo, para captar e enviar para a nuvem dados sobre a temperatura de um ambiente ou para controlar algum equipamento remotamente.

Esses exemplos mostram como serviços em nuvem podem ser utilizados no suporte a aplicações de IoT. Na verdade, diversos estudos mostram como a IoT e a Computação em Nuvem são complementares e como a integração desses conceitos viabilizou o surgimento de aplicações inovadoras (BOTTA *et al.*, 2016). No Quadro 3.10, alguns aspectos da Computação em Nuvem e da IoT são comparados, para mostrar como elas se relacionam de forma complementar. Enquanto a Computação em Nuvem centraliza uma capacidade computacional virtualmente ilimitada, a IoT viabiliza capilaridade na coleta de dados e controle de dispositivos na borda da rede.

Quadro 3.10 | Relação entre IoT e Computação em Nuvem

	Computação em Nuvem	IoT
Arquitetura	Centralizada	Distribuída
Alcance	Limitado	Ubíquo (pervasivo)
Componentes	Recursos Virtualizados	Objetos do mundo real

Fonte: adaptado de Botta (2016).

Pesquise mais

A segurança na comunicação entre dispositivos IoT e a nuvem é um aspecto crítico no desenvolvimento de aplicações nessa área. O trabalho indicado a seguir discute conceitos básicos e aspectos fundamentais de políticas para garantia de segurança e privacidade na integração de dados entre a nuvem e os dispositivos de IoT.

HUREL, L. M.; LOBATO, L. C. **Segurança e privacidade para a Internet das Coisas**. p. 14-22. Instituto Igarapé, nov. 2018.

Além de serviços de IoT, os provedores de Computação em Nuvem oferecem soluções especializadas para outras áreas. Vamos conhecer mais serviços em três segmentos importantes: Aprendizado de Máquina (*Machine Learning*); aplicações multimídia e soluções para desenvolvimento de software.

Aprendizado de Máquina (Machine Learning)

O uso de tecnologias de Inteligência Artificial (IA) permite o desenvolvimento de aplicações sofisticadas, por exemplo, soluções de reconhecimento facial, inteligência de negócios e robótica. Atualmente, entre as mais relevantes áreas da IA, podemos citar o Aprendizado de Máquina, que pode ser definido como “a construção de sistemas capazes de adquirir conhecimento de forma automática” (MONARD, 2003, p. 39). Os provedores oferecem serviços para modelagem e execução de aplicações baseadas em diversas abordagens de Aprendizado de Máquina, como Aprendizado por Reforço, Redes Neurais e Aprendizado Profundo (*Deep Learning*). No contexto de serviços IaaS, o Google Cloud disponibiliza, por meio do serviço Deep Learning VM Image, imagens de máquinas virtuais customizadas com ferramentas de Aprendizado de Máquina. Por exemplo, o cliente pode criar instâncias para executar aplicações baseadas em Deep Learning. Essa abordagem “objetiva produzir representações hierárquicas de alto nível dos dados de entrada, por meio de camadas de processamento sequencial em uma rede neural artificial” (BEZERRA, 2016). Exemplos de uso são aplicações para reconhecimento visual, reconhecimento de fala e processamento de linguagem natural.

Quadro 3.11 | Exemplos de serviços de Aprendizado de Máquina

Serviço	Descrição	Caso de Uso
Amazon Sage-Maker (AWS, 2019a)	Solução para criar, treinar e implantar modelos de Aprendizado de Máquina para aplicações que lidam com grandes volumes de dados.	Aplicação para diagnosticar fraudes em transações financeiras eletrônicas, como compras com cartão de crédito.
Deep Learning VM Image (GOOGLE, 2019c)	Imagens de máquinas virtuais customizadas para aplicações de Deep Learning (LECUN, 2015).	Criação de máquinas virtuais com modelos de deep learning para aplicações de segurança que envolvem o reconhecimento de faces de criminosos.
Azure Machine Learning (MICROSOFT, 2019e)	Ferramenta para criação e implantação de modelos de Aprendizado de Máquina com suporte a tecnologias abertas como ONNX (ONNX, 2019) e TensorFlow (ABADI <i>et al.</i> , 2016), até mesmo sem usar linguagens de programação.	A empresa British Petroleum utilizou essa ferramenta implementar um modelo de Aprendizado de Máquina capaz de contribuir na predição de falhas nas estações de exploração de gás.

Fonte: elaborado pelo autor.

Pesquise mais

Os provedores de nuvem pública oferecem também serviços para outras áreas da Inteligência Artificial, como mecanismos para sintetização

de áudio, tradução de textos e reconhecimento facial. Esses serviços facilitam a inclusão de soluções inteligentes em aplicações web ou aplicativos. Você pode encontrar mais informações e exemplos visitando o site da AWS Machine Learning e Inteligência Artificial. Pesquise também sobre a plataforma de IA do Azure, disponível no site da Microsoft.

Aplicações multimídia

Vamos estudar agora exemplos voltados para aplicações que incluem conteúdo multimídia. Esses serviços são muito importantes para facilitar o desenvolvimento de aplicações de streaming de vídeo, por exemplo. Vejamos alguns deles:

- Provedor AWS - oferece vários serviços de mídia.
- Elemental MediaConvert - é importante como recurso para promover a portabilidade das soluções, pois permite gerar mídias em diferentes formatos.
- AWS Elemental MediaConnect - é uma plataforma escalável e flexível para transmissão de vídeo ao vivo.

O Microsoft Azure também disponibiliza serviços para aplicações multimídia. O serviço Content Protection inclui mecanismos para garantir que somente dispositivos autorizados possam reproduzir fluxos (streaming) de mídia distribuídos por uma aplicação. Para streaming de vídeo ao vivo ou sob demanda, o Azure oferece o serviço Live and On-Demand Streaming.

Quadro 3.12 | Exemplos de serviços para aplicações multimídia

Serviço	Descrição	Caso de Uso
AWS Elemental MediaConvert (AWS, 2019d)	Mecanismos para converter um vídeo de entrada em diferentes formatos de saída a fim de oferecer suporte para várias plataformas e dispositivos.	Um caso de uso para esses serviços são as aplicações de streaming de áudio, como o Spotify, que precisam gerar arquivos de áudio em vários formatos e qualidades para atender a diferentes dispositivos dos usuários.
AWS Elemental MediaConnect (AWS, 2019c)	Solução para transmissão de vídeo ao vivo escalável, confiável e segura.	Aplicação de transmissão ao vivo de eventos esportivos.
Live and on-demand streaming (MICROSOFT, 2019f)	Serviço com recursos para codificação e distribuição escalável de vídeo (ao vivo ou sob demanda).	Transmissão ao vivo de eventos na forma de webcasts.

Serviço	Descrição	Caso de Uso
Content Protection (MICRO-SOFT, 2019d)	Mecanismos de criptografia e gerenciamento de direitos autorais para mídias digitais a fim de proteger conteúdo distribuído na Internet.	Codificação de canal de TV por assinatura distribuído pela internet para garantir que somente usuários autorizados poderão decodificar os dados para assistir ao canal.

Fonte: elaborado pelo autor.

Assimile

Cada aplicação em Nuvem é um sistema complexo que envolve vários componentes. O projeto de uma aplicação pode incluir diversos serviços, até mesmo serviços oferecidos por provedores diferentes. Em particular, as aplicações de streaming de áudio e vídeo envolvem requisitos severos de desempenho, por exemplo, baixo atraso médio e altas taxas de transmissão de dados. Dessa forma, para desenvolver aplicações desse tipo, podem ser combinados serviços de mídia e serviços de rede. Por exemplo, os serviços CDN podem ser utilizados para replicar o conteúdo multimídia em vários servidores dispersos em escala mundial para otimizar o desempenho. Assim, um usuário de um aplicativo de streaming de áudio, por exemplo, vai obter o conteúdo do servidor mais próximo da sua rede de acesso, o que melhora o desempenho da aplicação.

Gerenciamento de desenvolvimento de software

Por fim, vamos descrever alguns serviços voltados para o gerenciamento de desenvolvimento de software. Os provedores oferecem soluções para diversas etapas do processo de desenvolvimento, desde a codificação até a implantação e o teste de aplicações em ambientes de produção nos provedores de Computação em Nuvem. Essas soluções permitem automatizar todo o ciclo de criação, teste e implantação de aplicações em uma infraestrutura gerenciada pelo provedor.

O Quadro 3.13 relaciona exemplos de serviços voltados para equipes de desenvolvimento de software. A AWS oferece serviços como o AWS X-Ray e o CodePipeline. O AWS X-Ray serve para avaliar o desempenho de aplicações, tanto na fase de testes (para aplicações em desenvolvimento) como na fase de manutenção (para aplicações que já foram implantadas em ambiente de produção). Essa solução permite o diagnóstico de problemas de desempenho nos componentes de aplicações distribuídas. O CodePipeline é um serviço para automatizar o lançamento de novas versões de uma aplicação; nesse caso, uma vez que o código é alterado e testado, o processo de implantação da nova versão no ambiente de nuvem passa a ser automático. Entre

os serviços oferecidos pelo Google Cloud, podemos citar a ferramenta para gerenciamento de código-fonte denominada Cloud Source Repositories, que permite automatizar a alocação de recursos para implantação de aplicações. O provedor Microsoft Azure oferece um conjunto de serviços de gestão de projetos de software agrupados em uma plataforma denominada Azure Devops. Entre esses serviços, podemos citar o Azure Boards, para gerenciamento de tarefas seguindo o paradigma de metodologias ágeis (BALLE, 2011), e o Azure Test Plans, que permite a configuração e a execução de testes de software de forma automatizada.

Quadro 3.13 | Exemplos de serviços para desenvolvimento de software

Serviço	Descrição	Caso de Uso
AWS X-Ray (AWS, 2019g)	Ferramenta para identificar e solucionar problemas de desempenho em aplicações distribuídas.	Identificação automática das partes de uma aplicação web que apresentam problemas de desempenho (alta latência por exemplo), assim como a causa dos problemas.
AWS CodePipeline (AWS, 2019b).	Solução para integração contínua e entrega contínua a fim de automatizar o gerenciamento das atualizações de aplicações.	Especificação de eventos para fazer a implantação (<i>deployment</i>) automático de uma aplicação quando o código-fonte é alterado.
Cloud Source Repositories (GOOGLE, 2019b)	Repositório para gerenciamento de código fonte com suporte a controle de versões e busca.	Controle de diversas versões do código fonte de uma aplicação por uma equipe de desenvolvimento.
Azure Test Plans (MICROSOFT, 2019b)	Ferramenta para automação de testes de software.	Gerar gráficos de resultados de testes em uma aplicação web (por exemplo, número de testes bem-sucedidos e falhas), assim como verificar a compatibilidade com diferentes navegadores web.
Azure Boards (MICROSOFT, 2019a)	Ferramenta para gerenciamento de projetos utilizando metodologias ágeis.	Criar painéis de bordo (<i>dashboards</i>) para acompanhamento das atividades de um projeto de desenvolvimento de software.

Fonte: elaborado pelo autor.

Refleta

Os provedores de Computação em Nuvem oferecem serviços para desenvolvimento, implantação e hospedagem de aplicações. Esses serviços simplificam o ciclo de desenvolvimento de software e evitam a necessidade de gerenciamento da infraestrutura. Em particular, a automação do processo de implantação (*deployment*) das aplicações torna mais ágil e eficiente a atualização das aplicações em nuvem. No entanto, o uso de ferramentas em nuvem para automação de testes e da implantação de aplicações impõe alguns desafios e exige profissio-

nais especializados. Você sabe descrever quais seriam esses desafios? Pense a respeito de portabilidade e compatibilidade de tecnologias, segurança e privacidade dos dados, interoperabilidade das aplicações com sistemas legados.

Caro aluno, chegamos ao final da Unidade 3. Nesta unidade, descrevemos diversos tipos de serviços oferecidos por provedores de Comutação em Nuvem. Vimos como hospedar aplicações, gerenciar e analisar dados, conduzir o desenvolvimento de software, assim como uma variedade de serviços especializados que podem ser utilizados como componentes para aplicações em nuvem. Com o conhecimento adquirido nesta unidade, você é capaz de escolher os serviços apropriados para desenvolver e implantar diversos tipos de aplicações em ambientes de Computação Nuvem.

Sem medo de errar

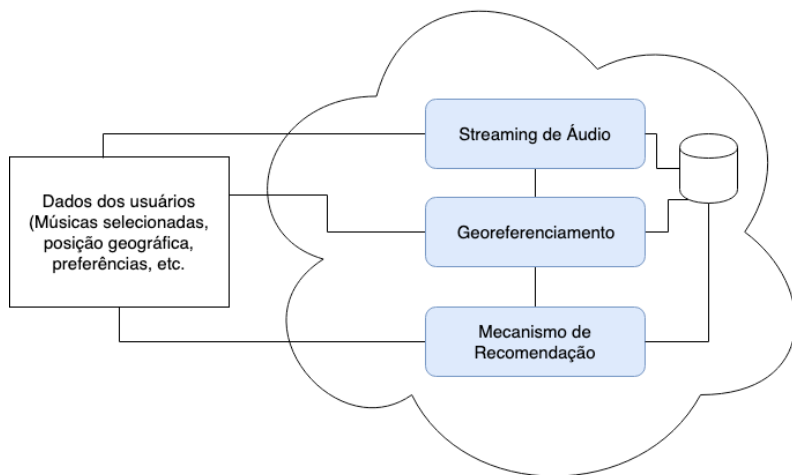
Como analista de TI da empresa de aplicativos, você está responsável pela decisão sobre quais serviços em nuvem serão utilizados como suporte aos principais requisitos do novo projeto da empresa, que consiste na implementação de um aplicativo de streaming de músicas. O diferencial desse produto deve ser um mecanismo de recomendação baseado na localização do usuário e em suas preferências, considerando também as escolhas dos demais usuários. Para resolver essa demanda, vamos considerar duas etapas: a especificação dos principais requisitos e, depois, a escolha de serviços apropriados para cada requisito. Vale salientar que vamos focar a lógica de recomendação do aplicativo sem nos preocuparmos com serviços específicos para armazenamento de dados.

De acordo com o contexto apresentado, podemos elencar os seguintes requisitos principais para o aplicativo:

- Mecanismo eficiente e escalável para streaming de áudio.
- Mecanismo para referenciar por localidade os dados dos usuários (para recomendação baseada em preferências regionais).
- Mecanismo para inferências e decisões de recomendação de músicas para os usuários.

A Figura 3.5 mostra os componentes da aplicação de acordo com os requisitos estabelecidos. Para cada requisito, vamos determinar um ou mais serviços em nuvem adequados.

Figura 3.5 | Componentes da aplicação de streaming de áudio



Fonte: Elaborada pelo autor.

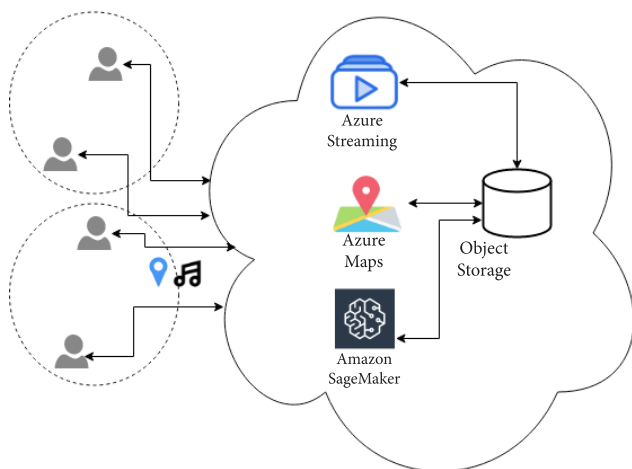
O **primeiro requisito** envolve escolher um serviço para streaming das músicas. Nesta seção, estudamos alguns serviços para aplicações multimídia, em particular, serviços para streaming de mídias digitais. Alguns desses serviços servem para conversão de mídias em diferentes formatos ou proteção de conteúdo para fins de garantia de direitos autorais. Para streaming das músicas, a necessidade corresponde a uma solução de streaming de mídia de forma escalável, pois a base de usuários do aplicativo pode crescer rapidamente. Nesse caso, existem vários serviços adequados, como o serviço *Azure Live and On-demand Streaming*.

O **segundo requisito** corresponde à capacidade de associar informações de localização dos usuários. Assim, torna-se possível uma recomendação georreferenciada. Com essas informações o aplicativo poderia determinar, por exemplo, os gêneros musicais preferidos em uma determinada região para incrementar a gama de recomendações aos usuários. Dentre os serviços estudados, poderíamos considerar aqueles utilizados em aplicações de IoT para suporte à localização dos dispositivos. Novamente, há várias respostas satisfatórias, entre as quais podemos citar o Azure Maps. Com esse serviço, seria possível integrar ao aplicativos funcionalidades de localização avançadas, como a criação de um mapa de densidade dos gêneros musicais ou a verificação de onde há mais usuários que gostam de determinada música.

Por fim, o **terceiro requisito** diz respeito à recomendação propriamente dita. O aplicativo deve ser capaz de tomar uma decisão sobre quais músicas ou gêneros musicais devem ser recomendados para cada usuário. Essa é uma tarefa complexa que envolve reconhecimento de padrões e inferências sobre grandes volumes de dados que representam as preferências, escolhas e localização dos usuários. Dentre as classes de serviços estudadas, podemos concluir que se trata um mecanismo baseado em Inteligência Artificial. Dessa forma, um serviço de Aprendizado de Máquina poderia ser utilizado para esse fim. Como vimos, há diversos desses mecanismos e muitos outros em vários provedores; um exemplo é o serviço Amazon SageMaker.

Considerando os serviços indicados, a solução ficaria como mostra a Figura 3.6. O Azure Maps permite identificar os usuários por região e agrupar suas preferências e estatísticas sobre as músicas escutadas. O Azure Streaming é usado para distribuir os fluxos de áudio para os usuários. O AWS SageMaker permite implementar um modelo de Aprendizado de Máquina para fazer recomendações aos usuários. Os arquivos de áudio podem ficar armazenados em algum serviço de armazenamento de objetos.

Figura 3.6 | Exemplo de serviços para aplicação de streaming de áudio



Fonte: Elaborada pelo autor.

Com o conteúdo abordado nesta unidade, realizamos um estudo abrangente sobre os principais tipos de serviços de Computação em Nuvem e ilustramos os cenários aos quais eles são adequados.

Manutenção preventiva de equipamentos

Observamos nos últimos anos uma revolução na Indústria. Com capacidade computacional embarcada nos mais diversos dispositivos, máquinas e veículos, viabilizaram-se processos avançados de manufatura e logística de produção. Os equipamentos das plantas industriais têm recursos para processamento e transmissão de dados, de forma que podem interagir entre si e com sistemas em nuvem para tornar os processos produtivos mais eficientes e confiáveis.

Considere uma fábrica que opera com equipamentos muito especializados, cuja compra só pode ser feita por encomenda. Assim, se um equipamento for danificado, sua substituição pode demorar muito, o que acarreta em significativo prejuízo financeiro. Por isso, essa fábrica pretende implementar uma solução com sensores e transmissores nos equipamentos a fim de coletar dados para um sistema responsável por monitorar a planta da fábrica, prever falhas e planejar a manutenção e a reposição dos equipamentos a fim de diminuir a probabilidade de um equipamento ficar inutilizado por defeito. Avalie quais serviços em nuvem poderiam ser usados na implementação de tal solução para manutenção preventiva dos equipamentos.

Resolução da situação-problema

Os provedores de nuvem pública oferecem várias soluções gerenciadas para aplicações de IoT, como é o caso de soluções para gerenciamento de instalações industriais. Nesse caso, podemos identificar duas tarefas principais: o gerenciamento da coleta de dados dos sensores e a análise desses dados para diagnóstico e tomada de decisão sobre manutenção dos equipamentos.

Uma primeira estratégia seria escolher um serviço para auxiliar cada uma das tarefas. Podemos utilizar um serviço básico de coleta e armazenamento de dados para aplicações de IoT, como o Cloud IoT Core, e utilizar um serviço de Aprendizado de Máquina, como o Azure Machine Learning para a tarefa de análise de dados.

Uma segunda estratégia poderia ser um serviço para aplicações IoT que já contempla as duas funcionalidades, como é o caso do AWS IoT SiteWise ou do Azure Time Series Insights. Ambos os serviços já incluem mecanismos para gerar estimativas e fazer previsões em função dos dados coletados ao longo do tempo, assim como ferramentas para visualização dos dados que

favorecem o gerenciamento eficiente dos recursos. A Figura 3.7 ilustra como seria o processo utilizando o serviço AWS IoT SiteWise.

Figura 3.7 | Exemplo de serviços para aplicação de streaming de áudio.



Fonte: Adaptada de AWS (2019a).

Faça valer a pena

1. Considere uma empresa que pretende implementar uma aplicação de IoT para monitoramento em tempo real de dados médicos de pacientes com doenças crônicas. Um dos requisitos críticos da aplicação é o sigilo das informações, uma vez que se trata de dados médicos. A empresa precisa encontrar um serviço em nuvem para implementar mecanismos de segurança para a aplicação de IoT.

O serviço em nuvem mais apropriado para a demanda da empresa é o

- a. Content Protection.
- b. Amazon SageMaker.
- c. Cloud IoT Core.
- d. AWS X-Ray.
- e. AWS IoT Device Defende.

2. Existem diversos serviços em nuvem para facilitar o desenvolvimento de aplicações multimídia, principalmente soluções voltadas para streaming de áudio e vídeo a partir de servidores em nuvem. Nesse contexto, analise as afirmativas a seguir.

- I. O serviço Content Protection da Azure pode ser utilizado para lidar com o problema de cópias não autorizadas de conteúdo em aplicações de streaming de áudio e vídeo na Internet.
- II. Além da transmissão de fluxos (*streaming*) de vídeo sob demanda, serviços de mídia em nuvem, como o AWS Elemental MediaCon-

nect e o Azure Live and on-demand streaming, suportam fluxos de vídeo ao vivo em larga escala.

III. O uso de serviços de streaming em provedores de Computação em Nuvem é suficiente para garantir o desempenho das aplicações.

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

3. Os provedores de Computação em Nuvem também oferecem serviços de apoio ao desenvolvimento de software, por exemplo, para gerenciamento de código-fonte e implantação (*deployment*) automática de aplicações. Porém, ainda existem várias dificuldades para o desenvolvimento e a hospedagem de aplicações em nuvem. Nesse contexto, avalie as afirmativas a seguir:

- I. Uma das principais dificuldades é o gerenciamento e o versionamento de código-fonte das aplicações em nuvem, devido ao acesso remoto ao código hospedado em servidores do provedor.
- II. Aspectos de portabilidade, integração e interoperabilidade ficam prejudicados devido às inúmeras ferramentas específicas para automação de testes, implantação, gerenciamento de tarefas e monitoramento de aplicações.
- III. A implantação de aplicações em nuvem exige conhecimentos específicos sobre serviços e protocolos de rede, pois pode ocorrer algum problema na comunicação entre componentes de uma aplicação, mesmo que eles sejam disponibilizados pelo mesmo provedor.

Caracteriza corretamente uma dificuldade no desenvolvimento e implantação de aplicações em nuvem o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. II, apenas.
- e. I, II e III, apenas.

Referências

ABADI, M. *et al.* TensorFlow: A System for Large-scale Machine Learning. In: **Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation**. Berkeley, CA, USA: USENIX Association, 2016. Disponível em: <https://ai.google/research/pubs/pub45166>. Acesso em: 24 jun. 2019.

APACHE. **The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing**. Disponível em: <https://hadoop.apache.org/>. Acesso em: 7 ago. 2019.

ARDUINO. **Arduino IoT Cloud**: Connect the cloud to the world around you. Disponível em: <https://www.arduino.cc/en/IoT/HomePage>. Acesso em: 14 jul. 2019.

AWS. **Amazon DynamoDB**: Serviço de banco de dados NoSQL rápido e flexível para qualquer escala. Disponível em: <https://aws.amazon.com/pt/dynamodb>. Acesso em: 7 ago. 2019.

AWS. **Amazon Elastic Block Store (EBS)**: persistent block storage for Amazon EC2. Disponível em: <https://aws.amazon.com/pt/ebs/>. Acesso em: 7 ago. 2019.

AWS. **Amazon Elastic Compute Cloud (Amazon EC2)** - Capacidade computacional segura e redimensionável na nuvem. Lance aplicativos quando necessário, sem compromissos antecipados. [S.d.]. Disponível em: <https://aws.amazon.com/pt/ec2/>. Acesso em: 7 ago. 2019.

AWS. **Amazon Elastic Container Service (Amazon ECS)** - Execute aplicativos containerizados no ambiente de produção. [S.d.]. Disponível em: <https://aws.amazon.com/pt/ecs/>. Acesso em: 7 ago. 2019.

AWS. **Amazon Elastic File System**: Sistema de arquivos nativo da nuvem, escalável e elástico para Linux. Disponível em: <https://aws.amazon.com/pt/efs>. Acesso em: 7 ago. 2019.

AWS. **Amazon EMR**: Execute e escale facilmente o Apache Spark, o Hadoop, o HBase, o Presto, o Hive e outras estruturas de big data. Disponível em: <https://aws.amazon.com/pt/emr/>. Acesso em: 7 ago. 2019.

AWS. **Amazon Kinesis**: colete, processe e analise facilmente streams de vídeo e dados em tempo real. Disponível em: <https://aws.amazon.com/pt/kinesis>. Acesso em: 7 ago. 2019.

AWS. **Amazon QuickSight**. Disponível em: <https://aws.amazon.com/pt/quicksight>. Acesso em: 7 ago. 2019.

AWS. **Amazon Relational Database Service (RDS)**. Disponível em: <https://aws.amazon.com/pt/rds>. Acesso em: 7 ago. 2019.

AWS. **Amazon S3**: Armazenamento de objetos para armazenar e recuperar qualquer quantidade de dados de qualquer local. Disponível em: <https://aws.amazon.com/pt/s3>. Acesso em: 7 ago. 2019.

AWS. **Amazon SageMaker**: machine learning para todos os desenvolvedores e cientistas de dados. Disponível em: <https://aws.amazon.com/pt/sagemaker/>. Acesso em: 24 jun. 2019.

AWS. **AWS CodePipeline**: automatize pipelines de entrega contínua para oferecer atualizações rápidas e confiáveis. Disponível em: <https://aws.amazon.com/pt/codepipeline/>. Acesso em: 25 jun. 2019.

AWS. **AWS Elemental MediaConnect**: transporte de vídeo ao vivo seguro e confiável. Disponível em: <https://aws.amazon.com/pt/mediaconnect/>. Acesso em: 24 jun. 2019.

AWS. **AWS Elemental MediaConvert**: processe arquivos e clipes de vídeo para preparar conteúdo sob demanda para distribuição ou arquivamento. Disponível em: <https://aws.amazon.com/pt/mediaconvert/>. Acesso em: 24 jun. 2019.

AWS. **AWS IoT Device Defender**: gerenciamento de segurança para dispositivos de IoT. Disponível em: <https://aws.amazon.com/pt/iot-device-defender/>. Acesso em: 24 jun. 2019.

AWS. **AWS IoT SiteWise**: colete e organize facilmente dados de equipamentos industriais. Disponível em: <https://aws.amazon.com/pt/iot-sitewise/>. Acesso em: 24 jun. 2019.

AWS. **AWS Lambda**. Disponível em: <https://aws.amazon.com/pt/lambda/>. Acesso em: 7 ago. 2019.

AWS. **AWS X-Ray**: analise e depure aplicações distribuídas em produção. Disponível em: <https://aws.amazon.com/pt/xray/>. Acesso em: 25 jun. 2019.

AWS. **Redes e entrega de conteúdo da AWS**. Disponível em: <https://aws.amazon.com/pt/products/networking/>. Acesso em: 7 ago. 2019.

AZEVEDO, M. T. de. **Transformação digital na indústria**: indústria 4.0 e a rede de água inteligente no Brasil. 2017. Tese (Doutorado em Sistemas Eletrônicos) - Escola Politécnica, Universidade de São Paulo (USP), São Paulo, 2017. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3142/tde-28062017-110639/pt-br.php>. Acesso em: 24 jun. 2019.

BACK, T.; ANDRIKOPOULOS, V. Using a Microbenchmark to Compare Function as a Service Solutions. (K. Kritikos, P. Plebani, F. de Paoli, Eds.) *In: Service-Oriented and Cloud Computing*, p. 146-160. Cham: Springer, 2018.

BALDINI, I. *et al.* Serverless computing: Current trends and open problems. *In: Research Advances in Cloud Computing*. [S.l.]: Springer, 2017.

BALLE, A. R. **Análise de metodologias ágeis: conceitos, aplicações e relatos sobre XP e Scrum**. 2011. Trabalho de Conclusão de Curso – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

BEZERRA, E. Introdução à aprendizagem profunda. *In: OGASAWARA, E.; VIEIRA, V. Minicursos do XXXI Simpósio Brasileiro de Banco de Dados*. Cap. 3, p. 57-86. Salvador, BA: SBC, 2016. Disponível em: <http://sbbd2016.fpc.ufba.br/sbbd2016/minicursos/minicurso3.pdf>. Acesso em: 14 jul. 2019.

BOTTA, A. *et al.* Integration of Cloud computing and Internet of Things: A survey. **Future Generation Computer Systems**, v. 56, p. 684–700, 2016. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167739X15003015>. Acesso em: 14 jul. 2019.

BRANCO JR, E. C.; MACHADO, J. C.; MONTEIRO, J. M. Estratégias para proteção da privacidade de dados armazenados na nuvem. In: **Anais do 29 Simpósio Brasileiro de Banco de Dados**. Curitiba: Sociedade Brasileira de Computação, 2014. p. 46–52. Disponível em: <http://www.inf.ufpr.br/sbbd-sbsc2014/sbbd/proceedings/artigos/pdfs/14.pdf>. Acesso em: 6 ago. 2019.

CARDOSO, L. P. **Migração de Máquinas Virtuais para Economia de Energia**. 2014. Trabalho de Conclusão de Curso (Engenharia de Computação e Informação) - Universidade Federal do Rio de Janeiro, p. 6-15. Rio de Janeiro: UFRJ/POLI, 2014. Disponível em: <http://monografias.poli.ufrj.br/monografias/monopoli10011239.pdf>. Acesso em: 6 ago. 2019.

CHARLTON, C. Using AWS IoT to Create a Smart Home Water-Monitoring Solution. **The Internet of Things on AWS – Official Blog**. 2019. Disponível em: <https://aws.amazon.com/blogs/iot/using-aws-iot-to-create-a-smart-home-water-monitoring-solution/>. Acesso em: 25 jun. 2019.

COSTA, L. H. M. K. *et al.* Grandes Massas de Dados na Nuvem: Desafios e Técnicas para Inovação. In: **Minicursos do XXX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**. [S.l.], Sociedade Brasileira de Computação (SBC), 2012. cap. 1, p. 1–58. Disponível em: <https://www.gta.ufrj.br/ftp/gta/TechReports/CAC12.pdf>. Acesso em: 15 jun. 2019.

ERL, T.; PUTTINI, R.; MAHMOOD, Z. **Cloud Computing: Concepts, Technology & Architecture**. [S.l.]: Prentice Hall, 2013.

GONÇALVES, M. de O. **Servidor de arquivos SAMBA: definição de um backup automático e controle de uso de disco**. 2018. Trabalho de Conclusão de Curso. Universidade Federal Fluminense, 2018. Disponível em: <https://app.uff.br/riuff/handle/1/9030>. Acesso em: 8 jul. 2019.

GOOGLE. **Cloud Bigtable**: A petabyte-scale, fully managed NoSQL database service for large analytical and operational workloads. Disponível em: <https://cloud.google.com/bigtable/>. Acesso em: 7 ago. 2019.

GOOGLE. **Cloud Data Fusion: Fully managed, code-free data integration at any scale**. Disponível em: <https://cloud.google.com/data-fusion/>. Acesso em: 7 ago. 2019.

GOOGLE. **Cloud Data Studio**. Disponível em: <https://datastudio.google.com/overview>. Acesso em: 7 ago. 2019.

GOOGLE. **Cloud Firestore: store cloud-native app data at global scale**. Disponível em: <https://cloud.google.com/firestore>. Acesso em: 7 ago. 2019.

GOOGLE. **Cloud IoT Core**: a fully managed service to easily and securely connect, manage, and ingest data from globally dispersed devices. Disponível em: <https://cloud.google.com/iot-core/>. Acesso em: 24 jun. 2019.

GOOGLE. **Cloud Source Repositories**: a single place for your team to store, manage, and track code. Disponível em: <https://cloud.google.com/source-repositories/>. Acesso em: 25 jun. 2019.

GOOGLE. **Cloud SQL**: Fully managed relational database services. Disponível em: <https://cloud.google.com/sql>. Acesso em: 7 ago. 2019.

GOOGLE. **Deep Learning VM Image:** preconfigured VMs for deep learning applications. Disponível em: <https://cloud.google.com/deep-learning-vm/>. Acesso em: 24 jun. 2019.

GOOGLE. **Google BigQuery documentation.** Disponível em: <https://cloud.google.com/bigquery/docs/>. Acesso em: 7 ago. 2019.

GOOGLE. **Google Cloud Compute Engine.** Disponível em: <https://cloud.google.com/compute/>. Acesso em: 6 jun. 2019.

GOOGLE. **Google Cloud Filestore:** High-performance, fully managed file storage. Disponível em: <https://cloud.google.com/filestore/>. Acesso em: 7 ago. 2019.

GOOGLE. **Google Cloud Functions.** Disponível em: <https://cloud.google.com/functions/>. Acesso em: 7 ago. 2019.

GOOGLE. **Google Cloud networking products.** Disponível em: <https://cloud.google.com/products/networking/>. Acesso em: 7 ago. 2019.

GOOGLE. **Google Cloud Persistent Disk: reliable, high-performance block storage for virtual machine instances.** Disponível em: <https://cloud.google.com/persistent-disk/>. Acesso em: 7 ago. 2019.

GOOGLE. **Google Cloud Storage: unified object storage for developers and enterprises.** Disponível em: <https://cloud.google.com/storage/>. Acesso em: 7 ago. 2019.

GOOGLE. **Google Kubernetes Engine.** Disponível em: <https://cloud.google.com/kubernetes-engine/>. Acesso em: 7 ago. 2019.

HASHEM, I. A. T. *et al.* The rise of “big data” on cloud computing: Review and open research issues. **Information Systems**, v. 47, p. 98-115, 2015. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0306437914001288>. Acesso em: 7 ago. 2019.

HUAI, Y. *et al.* Major Technical Advancements in Apache Hive. In: **Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data**. New York, NY, USA: ACM, 2014. Disponível em: <http://doi.acm.org/10.1145/2588555.2595630>. Acesso em: 7 ago. 2019.

HUREL, L. M.; LOBATO, L. C. **Segurança e privacidade para a Internet das Coisas.** Instituto Igarapé, p. 14-22, 2018. Disponível em: <https://igarape.org.br/wp-content/uploads/2018/11/Seguranc%CC%A7a-e-Privacidade-para-a-Internet-das-Coisas.pdf>. Acesso em: 25 jun. 2019.

IBM. **IBM Cloud Functions.** Disponível em: <https://www.ibm.com/cloud/functions>. Acesso em: 7 jun. 2019.

IBM. **IBM Cloud Kubernetes Service.** Disponível em: <https://www.ibm.com/cloud/container/-service>. Acesso em: 8 ago. 2019.

IBM. **IBM Cloud Virtual Servers.** Disponível em: <https://www.ibm.com/cloud/virtual-servers/-cloud/>. Acesso em: 7 ago. 2019.

KON, F.; SANTANA, E. F. Z. Cidades Inteligentes: Tecnologias, Aplicações, Iniciativas e Desafios. In: MALDONADO, J. C.; VITERBO, J.; DELAMARO, M.; MARCZAK, S. dos S. (ed.). **Jornadas de Atualização em Informática**. 2016. [S.l.], SBC, 2016. Cap. 1, p. 13-60. Disponível em: <https://doi.org/10.5753/sbc.6.1>. Acesso em: 23 jun. 2019.

KUBERNETES. **What is Kubernetes**. Disponível em: <https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/>. Acesso em: 7 jun. 2019.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, p. 436-444, 2015. Disponível em: <https://doi.org/10.1038/nature14539>. Acesso em: 24 jun. 2019.

MICROSOFT. **Armazenamento de Blob**: armazenamento de objetos extremamente escalonável para dados não estruturados. Disponível em: <https://azure.microsoft.com/pt-br/services/storage/blobs/>. Acesso em: 7 ago. 2019.

MICROSOFT. **Armazenamento de Discos do Azure**: Opções de SSD persistentes, altamente seguras e econômicas para dar suporte às Máquinas Virtuais do Azure. Disponível em: <https://azure.microsoft.com/pt-br/services/storage/disks/>. Acesso em: 7 ago. 2019.

MICROSOFT. **Arquivos do Azure**: compartilhamentos de arquivos em nuvem simples, seguros e totalmente gerenciados. Disponível em: <https://azure.microsoft.com/pt-br/services/storage/files>. Acesso em: 7 ago. 2019.

MICROSOFT. **Azure Boards**: planeje, acompanhe, e discuta sobre trabalhos com suas equipes. Disponível em: <https://azure.microsoft.com/pt-br/services/devops/boards/>. Acesso em: 25 jun. 2019.

MICROSOFT. **Azure Container Instances (ACI)**. Disponível em: <https://azure.microsoft.com/pt-br/services/container-instances/>. Acesso em: 7 ago. 2019.

MICROSOFT. **Azure Cosmos DB**: serviço de vários modelos de banco de dados distribuído globalmente para qualquer escala. Disponível em: <https://azure.microsoft.com/pt-br/services/cosmos-db/>. Acesso em: 16 jun. 2019.

MICROSOFT. **Azure Functions**. Disponível em: <https://azure.microsoft.com/pt-br/services/functions/>. Acesso em: 7 jun. 2019.

MICROSOFT. **Azure Test Plans**: teste e envie com confiança usando um manual e um kit de ferramentas de teste exploratório. Disponível em: <https://azure.microsoft.com/pt-br/services/devops/test-plans/>. Acesso em: 25 jun. 2019.

MICROSOFT. **Azure Time Series Insights**: aumente a inteligência operacional e transforme sua empresa com insights acionáveis em tempo real sobre os dados de série temporal em escala IoT. Disponível em: <https://azure.microsoft.com/pt-br/services/time-series-insights/>. Acesso em: 24 jun. 2019.

MICROSOFT. **Banco de dados SQL do Azure**: SQL inteligente gerenciado na nuvem. Disponível em: <https://azure.microsoft.com/pt-br/services/sql-database/>. Acesso em: 7 ago. 2019.

MICROSOFT. **Data Factory**: Serviço de integração híbrida de dados que simplifica o ETL em escala. Disponível em: <https://azure.microsoft.com/pt-br/services/data-factory/>. Acesso em: 17 jun. 2019.

MICROSOFT. **HDInsight**: Serviço fácil, econômico e de nível empresarial para análise de software livre. Disponível em: <https://azure.microsoft.com/pt-br/services/hdinsight/>. Acesso em: 17 jun. 2019.

MICROSOFT. **Mapas do Azure**: APIs geoespaciais para adição de mapas, análise espacial e soluções de mobilidade para seus aplicativos. Disponível em: <https://azure.microsoft.com/pt-br/services/azure-maps/>. Acesso em: 24 jun. 2019.

MICROSOFT. **Máquinas Virtuais do Microsoft Azure**. Disponível em: <https://azure.microsoft.com/pt-br/services/virtual-machines/>. Acesso em: 6 jun. 2019.

MICROSOFT. **Proteção do conteúdo**: forneça conteúdo altamente seguro usando diversos DRMs (PlayReady, Widevine, FairPlay Streaming) ou criptografia de chave não criptografada AES. Disponível em: <https://azure.microsoft.com/pt-br/services/media-services/content-protection/>. Acesso em: 25 jun. 2019.

MICROSOFT. **Serviço do Azure Machine Learning**: crie modelos rapidamente e operacionalize em escala da nuvem para a borda. Disponível em: <https://azure.microsoft.com/pt-br/services/machine-learning-service/>. Acesso em: 25 jun. 2019.

MICROSOFT. **Streaming sob demanda e ao vivo**: forneça conteúdo para, virtualmente, qualquer dispositivo em escala. Disponível em: <https://azure.microsoft.com/pt-br/services/media-services/live-on-demand/>. Acesso em: 25 jun. 2019.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre Aprendizado de Máquina. In: REZENDE, S. O. (ed.). **Sistemas Inteligentes** - Fundamentos e Aplicações. p. 89-114. Barueri, SP: Editora Manole, 2003. Disponível em: <http://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>. Acesso em: 23 jun. 2019.

MONGODB. **MongoDB Atlas**. Disponível em: <https://www.mongodb.com/cloud/atlas>. Acesso em: 18 jun. 2019.

NETO, A. J. R.; DA FONSECA, N. L. S. Um estudo comparativo do desempenho dos protocolos iSCSI e fibre channel. **IEEE Latin America Transactions**, v. 5, n. 3, p. 151, 2007.

NODE.JS. **About Node.js**. Disponível em: <https://nodejs.org/en/about/>. Acesso em: 7 ago. 2019.

ONNX. **Open Neural Network Exchange Format**: the open ecosystem for interchangeable AI models. Disponível em: <https://onnx.ai/>. Acesso em: 24 jun. 2019.

OUSSOUS, A. *et al.* Big Data technologies: A survey. **Computer and Information Sciences**, v. 30, n. 4, p. 431-448, 2018. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1319157817300034>. Acesso em: 9 jul. 2019.

PAULA, G. S. de. **Avaliação de serviços Serverless**: um experimento piloto. 2018. Trabalho de Conclusão de Curso (Graduação) - Universidade Tecnológica Federal do Paraná, p. 11-20.

Dois Vizinhos, 2018. Disponível em: <http://repositorio.roca.utfpr.edu.br/jspui/handle/1/10033>. Acesso em: 6 ago. 2019.

POSTMAN. **About Postman**. Disponível em: <https://www.getpostman.com/about-postman>. Acesso em: 7 ago. 2019.

RATHORE, M. M. *et al.* Urban planning and building smart cities based on the Internet of Things using Big Data analytics. **Computer Networks**, v. 101, p. 63–80, 2016. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1389128616000086>. Acesso em: 9 jul. 2019.

SANTOS, B. P. *et al.* Internet das coisas: da teoria à prática. In: SIQUEIRA, F. A. *et al.* (org.). **Minicursos do XXXIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)**. Porto Alegre: SBC, 2016. Cap. 1, p. 1-50. Disponível em: <http://www.sbrc2016.ufba.br/downloads/anais/MinicursosSBRC2016.pdf>. Acesso em: 23 jun. 2019.

SOUSA, F. R. C. *et al.* Gerenciamento de dados em nuvem: Conceitos, sistemas e desafios. In: PEREIRA, A. C. M. *et al.* (org.). **Tópicos em sistemas colaborativos, interativos, multi-mídia, web e bancos de dados**. Belo Horizonte: Sociedade Brasileira de Computação (SBC), 2010. Cap. 4, p. 108-121. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/18/81/167-1>. Acesso em: 6 ago. 2019.

SOUZA, A. M. *et al.* Critérios para Seleção de SGBD NoSQL: o Ponto de Vista de Especialistas com base na Literatura. In: **Anais do X Simpósio Brasileiro de Sistemas de Informação (SBSI)**. Porto Alegre: Sociedade Brasileira de Computação (SBC), 2014. p. 149-160. Disponível em: <https://sol.sbc.org.br/index.php/sbsi/article/view/6109>. Acesso em: 17 jun. 2019.

TANENBAUM, A. S.; STEEN, M. V. **Sistemas Distribuídos: Princípios e Paradigmas**. 2. ed. São Paulo: Pearson Prentice Hall, 2008.

TANENBAUM, A. S.; STEEN, M. Van. **Sistemas Distribuídos: Princípios e Paradigmas**. 2. ed. São Paulo: Pearson Prentice Hall, 2008.

YEGULALP, S. What is Kubernetes? Container orchestration explained. **Infoworld**. 3 abr. 2019. Disponível em: <https://www.infoworld.com/article/3268073/what-is-kubernetes-container-orchestration-explained.html>. Acesso em: 7 ago. 2019.

ZAHARIA, M. *et al.* Apache Spark: A Unified Engine for Big Data Processing. **Commun. ACM**, v. 59, n. 11, p. 56–65, out. 2016. Disponível em: <http://doi.acm.org/10.1145/2934664>. Acesso em: 18 jun. 2019.

Unidade 4

Neumar Costa Malheiros

Arquitetura de aplicações em nuvem

Convite ao estudo

Caro estudante, nosso objetivo, nesta seção, é o estudo de modelos de arquitetura para aplicações em nuvem, que, em geral, são complexas, pois envolvem diversos componentes que interagem entre si por meio da Internet. O projeto da arquitetura de uma aplicação afeta diversos aspectos, como seu desempenho, disponibilidade e até mesmo o nível de segurança. Portanto, esse é um tópico relevante e que merece muita atenção. As aplicações em nuvem podem incluir diversos componentes de software, como bancos de dados, serviços web e aplicativos. A especificação da arquitetura consiste em determinar quais são os componentes, como eles devem se comunicar, como podem ser protegidos de ações maliciosas e como garantir que os requisitos da aplicação sejam alcançados.

Para discutir mais profundamente esses assuntos, vamos considerar um cenário no qual você decidiu começar um empreendimento com alguns sócios, que consiste em uma nova aplicação de *streaming* de vídeo. Você é o líder técnico da equipe e tem a responsabilidade de tomar as decisões relacionadas com Tecnologia da Informação. Considerando a fase de projeto da aplicação de *streaming*, você fica encarregado de fazer escolhas estratégicas para a implementação da solução proposta. Dessa forma, é necessário especificar a arquitetura da aplicação, considerando fatores como desempenho e escalabilidade, para que se possa garantir uma experiência satisfatória aos usuários que assistem aos vídeos sob demanda. Além disso, você terá que verificar as principais ameaças e riscos de segurança, a fim de determinar mecanismos de proteção de dados e autenticação dos usuários.

Para lidar com essas questões, precisamos compreender diversos assuntos relacionados aos projetos de aplicações em ambientes de computação. Na Seção 4.1, vamos descrever os principais modelos de arquitetura de aplicações distribuídas. A especificação da arquitetura de uma solução em nuvem tem implicações na sua escalabilidade e no seu desempenho. Além dos conceitos básicos, será discutida a evolução dos modelos de arquitetura, incluindo abordagens como *Microserviços*, *Serverless Computing* e *Edge Computing*. Na Seção 4.2, vamos estudar aspectos específicos de desempenho e de tolerância a falhas de aplicações em nuvem. O objetivo principal será

compreender os mecanismos para garantir a qualidade de serviço nos provedores de Computação em Nuvem. Por fim, na Seção 4.3, vamos concluir nosso estudo com uma ampla discussão sobre segurança de aplicações em nuvem. Serão abordadas as características de sistema seguros, principais ameaças para as aplicações e também mecanismos e soluções para garantia de segurança e gerenciamento de riscos. Ao concluir esta unidade, você será capaz de especificar a arquitetura de aplicações com foco no desempenho, escalabilidade e segurança. Bom estudo!

Modelos de arquitetura em nuvem

Dialógo aberto

Prezado aluno, você já aprendeu que uma solução em nuvem pode fazer uso de vários serviços de armazenamento de dados e de instâncias de execução de aplicações. Como todos esses elementos podem ser organizados a fim de aumentar o desempenho da solução? Para compreender isso, precisamos conhecer os conceitos básicos sobre a arquitetura de aplicações em nuvem. De forma geral, a arquitetura de um sistema é a especificação de quais são os seus componentes e como devem ser conectados. Sabemos que as aplicações em nuvem são sistemas que envolvem vários componentes de software, como aplicativos que se conectam a serviços de armazenamento de objetos ou a um banco de dados que é executado em uma máquina virtual. Outro exemplo é uma aplicação web que recebe requisições dos navegadores web ou interage com serviços de análise de dados. A comunicação entre esses elementos é feita por meio da Internet, então o desempenho e a segurança dependem de um modelo de arquitetura adequado. Além disso, as aplicações em nuvem devem ser escaláveis para lidar com um volume crescente de dados e clientes, o que torna necessário incluir mecanismos de replicação e balanceamento de carga.

Esse assunto é muito importante diante dos desafios que você e os seus sócios vão enfrentar no desenvolvimento de uma nova aplicação de *streaming* de vídeo. Você é o responsável pelo projeto e precisa especificar como essa solução será implementada em um ambiente de Computação em Nuvem. A aplicação de *streaming* de vídeo inclui três módulos funcionais: o gerenciamento de usuários, com funcionalidades para cadastro, autenticação, entre outras; o gerenciamento dos vídeos, com funcionalidades para inclusão na plataforma, pesquisa pelos usuários e mecanismos de recomendação; e um módulo responsável pelo *streaming* de vídeo sob demanda. Sua primeira tarefa é identificar as principais abordagens utilizadas na definição de arquiteturas de aplicação em Nuvem e determinar qual delas é mais adequada para esse projeto de *streaming* de vídeo. Você deve entregar um relatório de uma a duas páginas, especificando o melhor modelo de arquitetura para essa aplicação, apresentando as devidas justificativas técnicas. Para isso, é importante observar que um dos principais requisitos nesse caso é o desempenho, pois o *streaming* de mídia na Internet é um tipo de aplicação que exige uma conexão de boa qualidade. Para que o usuário possa assistir vídeos em alta resolução, sem falhas ou interrupções, é necessária uma conexão que

suporte, por exemplo, altas taxas de transmissão e um baixo atraso (latência) na entrega dos dados.

Há muitos tópicos relevantes para o projeto de aplicação em ambiente de nuvem. Nesta seção, vamos conhecer os principais modelos de arquitetura. Iniciaremos nosso estudo sobre o tema com uma descrição dos conceitos básicos sobre arquitetura de aplicações distribuídas. Em seguida, vamos conhecer as características dos modelos de arquitetura para aplicações em nuvem, como arquitetura multicamadas, a arquitetura de Microserviços e a Computação sem servidor (Serverless). Além disso, vamos introduzir o conceito de *Edge Computing*, um novo paradigma para lidar com problemas de escalabilidade de soluções em nuvem.

Bom estudo!

Não pode faltar

As aplicações em nuvem são sistemas de software complexos, na medida em que implementam diversas funcionalidades e fazem uso de variados serviços que são acessados por meio da Internet. Para lidar com essa complexidade, as aplicações podem ser divididas em módulos (ou componentes) funcionais. Por exemplo, uma aplicação web de comércio eletrônico pode incluir vários módulos, como um responsável pela autenticação de usuários ou outro responsável pela geração de relatórios de vendas. Cada um deles pode ser um software independente, capaz de ser executado, por exemplo, em uma máquina virtual ou em um contêiner de um provedor de nuvem pública.

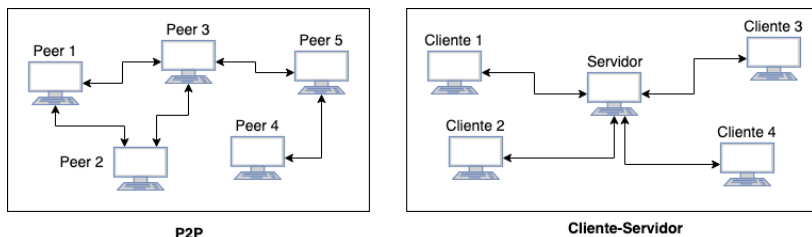
A definição da arquitetura de uma aplicação consiste em definir quais seriam os seus módulos funcionais e como eles devem interagir entre si (COULOURIS *et al.*, 2013). O projeto da arquitetura de aplicações em nuvem é um grande desafio e as decisões nele envolvidas podem influenciar vários aspectos da aplicação, como desempenho, escalabilidade e segurança. A fase de projeto é importante para tentar antecipar a identificação de possíveis problemas como: gargalos de desempenho, ameaças de segurança e impacto da falha de componentes específicos na disponibilidade da aplicação. Componentes críticos podem ser replicados para viabilizar mecanismos de tolerância a falhas e balanceamento de carga.

Existem três modelos básicos para a arquitetura de aplicações distribuídas: arquitetura centralizada, descentralizada e híbrida (TANENBAUM; STEEN, 2008). A arquitetura centralizada é o modelo tradicional cliente-servidor. Quando um componente requisita um serviço de outro, o que faz a requisição é o cliente e o que responde é o servidor. As aplicações web

representam um exemplo típico do modelo de arquitetura centralizada: o servidor web responde as requisições enviadas por navegadores web.

A arquitetura descentralizada é caracterizada pelo modelo Peer-to-Peer (P2P). Nesse caso, não há distinção entre clientes e servidores, e os componentes de software, por sua vez, podem fazer requisições entre si de forma arbitrária. A Figura 4.1 ilustra a lógica dos modelos cliente-servidor e P2P.

Figura 4.1 | Comparação do modelo descentralizado com o centralizado



Fonte: elaborada pelo autor.

Uma alternativa interessante a escolher um modelo centralizado ou descentralizado é a arquitetura híbrida, na qual uma mesma aplicação utiliza os dois modelos, ou seja, algumas funcionalidades da aplicação são implementadas na forma cliente-servidor e outras são implementadas na forma P2P. Assim, pode-se aproveitar o melhor de cada arquitetura.

Exemplificando

Você conhece aplicações com arquitetura híbrida? Esse é o caso de algumas aplicações de compartilhamento de arquivos, como aquelas baseadas no protocolo BitTorrent (COHEN, 2003; LEHMANN *et al.*, 2012). Nessas aplicações, a transferência dos arquivos é implementada conforme o modelo P2P, mas a busca pelos arquivos é no modelo cliente-servidor. Outro exemplo são aplicações de comunicação por voz, como o Skype. Nesse caso, a verificação dos contatos on-line é implementada na forma cliente-servidor, mas as chamadas podem ser P2P.

Apesar da escalabilidade dos sistemas P2P, esse modelo é muito complexo, pois exige a implementação de mecanismos descentralizados. A maior prova dessa complexidade é que a maioria das aplicações que utilizados, como aplicativos de comunicação instantânea e aplicações web, são sistemas cliente-servidor centralizados. Porém, essas aplicações centralizadas apresentam problemas de desempenho e escalabilidade, o que fez com que muitas abordagens surgissem com o intuito de resolvê-los. Uma delas

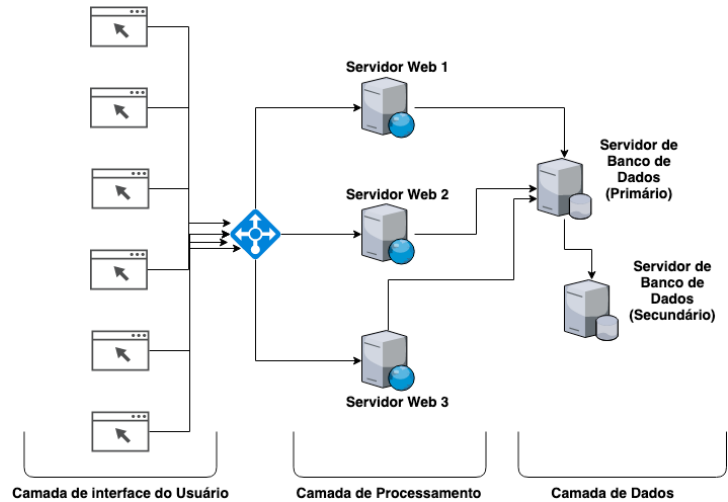
é o modelo cliente-servidor com uma arquitetura em múltiplas camadas (TANENBAUM; STEEN, 2008). Nesse modelo de arquitetura, a aplicação é dividida em várias camadas, sendo cada uma delas responsável por um conjunto específico de funcionalidades. Os componentes de uma camada podem interagir com os componentes das camadas vizinhas, além de serem executados em servidores diferentes para melhorar o desempenho. Nesse caso, temos uma separação física entre as camadas.

Em geral, as aplicações multicamadas são divididas em três:

1. Camada de interface do usuário (controle da interação com o usuário).
2. Camada de processamento (implementação da lógica das principais funcionalidades da aplicação).
3. Camada de dados (armazenamento persistente de dados).

Na Figura 4.2, temos o exemplo de uma aplicação Web baseada na arquitetura em camadas. Nesse caso, o navegador web executa a camada de interface com o usuário, já na camada de processamento, o servidor web responde as requisições de acordo com as regras de negócio. Na camada de dados, um Sistema Gerenciador de Banco de Dados administra o armazenamento das informações do sistema. Em um ambiente de nuvem, podemos usar diversos serviços para hospedar uma aplicação com essa arquitetura. Por exemplo, podemos alocar máquinas virtuais ou contêineres para executar um servidor web (na camada de processamento) e usar serviços de Bancos de Dados em Nuvem para persistências das informações (na camada de dados).

Figura 4.2 | Exemplo de uma aplicação com arquitetura em camadas e componentes replicados

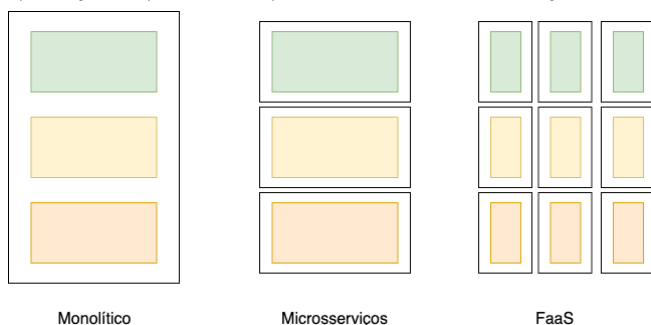


Fonte: elaborada pelo autor.

Os provedores de nuvem também oferecem serviços para replicação dinâmica dos componentes. No exemplo ilustrado na Figura 4.2, há três réplicas do servidor web e duas réplicas do servidor de banco de dados. Esse tipo de replicação pode ser automatizado em ambientes de nuvem. Além disso, os provedores oferecem mecanismos para balanceamento de carga. A Figura 4.2 também demonstra o balanceamento de carga das requisições dos clientes entre as réplicas do servidor Web. Se houver falha em um dos servidores, as demais réplicas continuam a atender as requisições. Além disso, a carga de trabalho é dividida entre as réplicas, o que resulta em menor tempo de resposta aos clientes. Com isso, a aplicação dividida em camadas melhora o grau de desempenho e confiabilidade, mesmo que a arquitetura ainda seja cliente-servidor. Os provedores de Computação em Nuvem permitem a alocação ou liberação automática de réplicas de acordo com as variações no volume de requisições dos clientes.

Em geral, as principais funcionalidades de uma aplicação são implementadas na camada de processamento, por exemplo, na forma de um Serviço Web (*Web Service*). Esse serviço é denominado monolítico quando todos os seus módulos funcionais estão implementados em um único componente de software. Se a aplicação possui muitas funcionalidades, esse serviço pode apresentar alguns problemas como: ocupação de muito espaço de memória; dificuldade de implementar correções e de evolução do software; alto custo de replicação, etc. Diante desses problemas, novas arquiteturas de serviços web foram propostas para melhorar o desempenho e a escalabilidade, assim como facilitar a replicação. Essas abordagens se baseiam na ideia de dividir o serviço web monolítico em diversos componentes de software independentes. Entre as principais abordagens, podemos citar a arquitetura de microsserviços (DRAGONI *et al.*, 2017) e a arquitetura *Serverless* (Computação sem Servidor) (BALDINI *et al.*, 2017), cujo principal exemplo é o modelo Função como Serviço (FaaS – Function as a Service). A Figura 4.3 ilustra a evolução desses modelos, que se reflete no grau de modularização da aplicação.

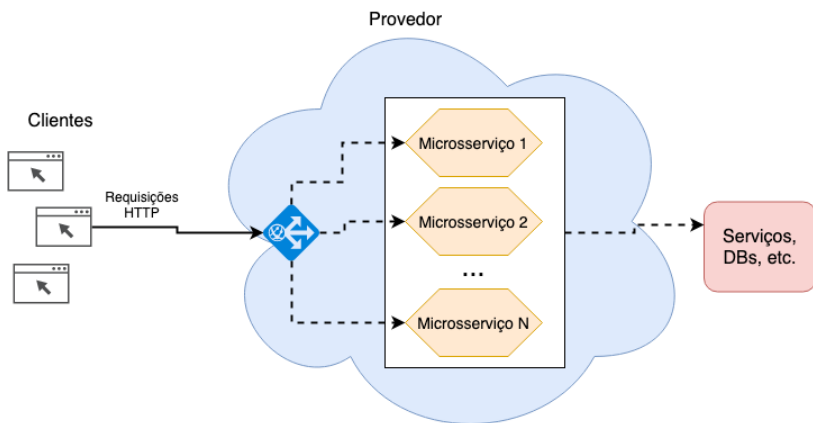
Figura 4.3 | Evolução dos padrões de arquitetura de software como serviço



Fonte: elaborada pelo autor.

No modelo de microsserviços, cada um dos módulos funcionais de uma aplicação monolítica vira um serviço menor e especializado, denominado microsserviço. Cada microsserviço é um componente de software independente. Assim, ele pode ser executado como um serviço web em um contêiner ou máquina virtual. A Figura 4.4 exemplifica a arquitetura de microsserviços. Nesse caso, uma aplicação é modularizada em N microsserviços independentes, que podem se comunicar com serviços remotos, como outra aplicação, com sistemas legados ou com um banco de dados (DB – *Database*). A Figura 4.4 também ilustra o papel de um *API Gateway*, que é um componente responsável por redirecionar as requisições dos clientes para o microsserviço apropriado de acordo com a funcionalidade requisitada. A arquitetura de microsserviços é adequada para aplicações complexas, por exemplo, uma aplicação para consolidação de transações financeiras com cartão de crédito, pois envolve várias etapas e entidades, como lojas, bancos, adquirentes, etc. Nesse caso, a aplicação pode ser modularizada em vários microsserviços específicos que lidam com cada parte do problema.

Figura 4.4 | Exemplo de aplicação com arquitetura de microsserviços

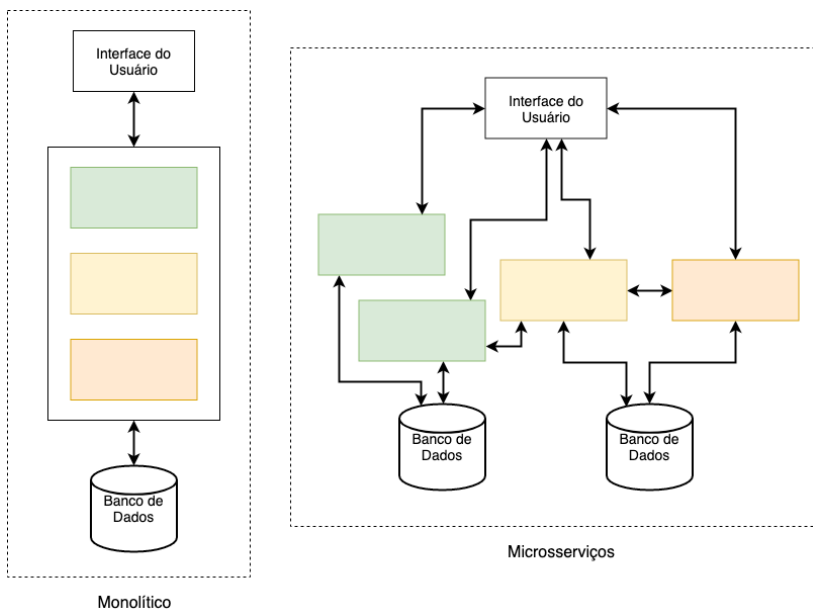


Fonte: elaborada pelo autor.

Como discutido em (DRAGONI *et al.*, 2017), duas características principais definem um microsserviço: o alto grau de coesão das suas funcionalidades e sua capacidade de responder requisições. A coesão significa que um microsserviço implementa um conjunto de funcionalidades relacionadas, que dependem umas das outras. A capacidade de responder requisições implica que os microsserviços, de fato, se comportam como um servidor, ou seja, provêm algum serviço. Eles podem responder requisições de clientes ou requisições de outros microsserviços.

A arquitetura de microsserviços representa aplicações distribuídas que são formadas pela composição de microsserviços independentes. Esse conceito é ilustrado na Figura 4.5. Nesse caso, temos um serviço monolítico formado por três módulos funcionais. Em uma arquitetura de microsserviços, cada um desses módulos poderia ser implementado como um (micro) serviço independente. Como mostra a figura, apenas alguns dos microsserviços podem ser replicados e cada um deles pode ter seu próprio banco de dados.

Figura 4.5 | Comparação entre serviço monolítico e microsserviços



Fonte: elaborada pelo autor.

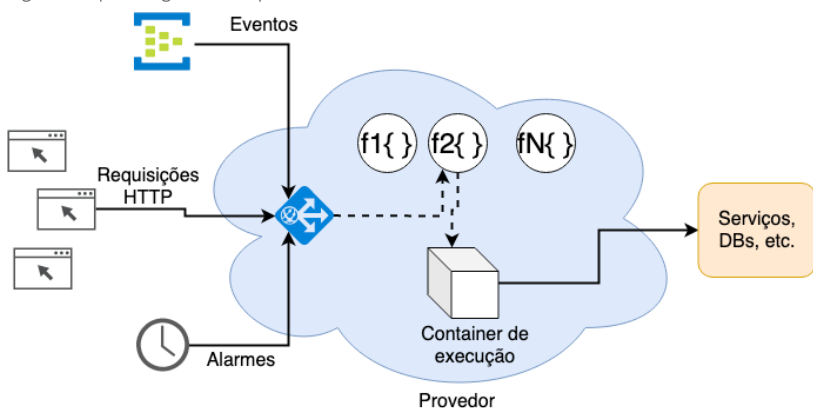
Assimile

Cada microsserviço é um componente de software independente. Assim, tecnologias, linguagens de programação e ferramentas diferentes podem ser utilizadas para cada microsserviço de acordo com seus requisitos. Por exemplo, se um microsserviço precisa de maior desempenho, ele pode ser implementado em uma linguagem compilada de baixo nível.

A arquitetura de microsserviços traz uma série de vantagens. Ao implementar modificações em um microsserviço, somente esse componente específico precisa ser reiniciado. No serviço monolítico, as modificações em apenas um módulo exigem a reinicialização do serviço como um todo. Como o microsserviço é menor, ele exige menos recursos, o que facilita o uso de contêineres e a replicação. Além disso, cada microsserviço pode ser replicado de forma independente. Assim, aquele que apresenta grande demanda pode ser replicado, enquanto outro, com menos demanda, não tem essa necessidade. No caso de um monolítico, o software como um todo (incluindo todos os seus módulos) tem que ser replicado, a despeito da demanda para cada um de seus módulos funcionais, o que pode resultar em desperdício de recursos.

O próximo passo na evolução dos modelos de arquitetura foi o paradigma de Computação sem Servidor (*Serverless*). No modelo *Serverless*, o nível de granularidade é ainda maior que na arquitetura de microsserviços. O objetivo é explorar o fato de que cada microsserviço pode ainda ser dividido em funções específicas e cada função pode ser invocada de forma independente por um software cliente. No ambiente de nuvem, esse modelo pode ser implementado na forma de Função como Serviço (*FaaS – Function as a Service*). O nome “Computação sem Servidor” remete à ideia de que, neste modelo, o cliente não precisa alocar servidores para a execução de uma aplicação. Isso é feito dinamicamente pelo provedor quando alguma função da aplicação é invocada. Assim, a tarifação no modelo *Serverless* é feita de acordo com o tempo de execução das funções e não pela alocação de recursos (BALDINI *et al.*, 2017). A Figura 4.6 mostra as principais características dessa arquitetura. Cada funcionalidade específica da aplicação precisa ser implementada como uma função isolada que pode receber parâmetros e retornar algum resultado. O código das funções de uma aplicação ($f_1 \dots f_N$) são enviados para o provedor. A execução de uma função pode ser disparada por diversas formas, como uma requisição web, eventos em outros componentes (como uma inserção em um banco de dados), ou por um alarme (o que permite agendar a execução de uma função). As requisições são enviadas para um *gateway*, que identifica a função apropriada e aloca os recursos necessários para executá-la (um contêiner, por exemplo). A execução da função pode resultar em resposta a um cliente ou em acesso a outros serviços ou bancos de dados por exemplo.

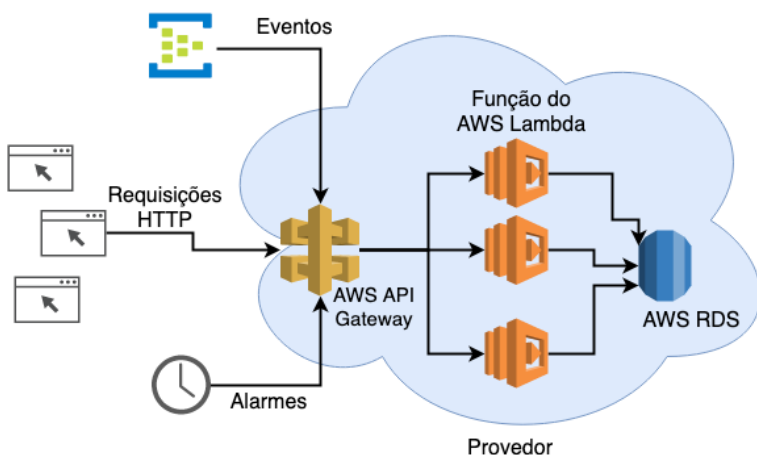
Figura 4.6 | Visão geral da arquitetura *Serverless*



Fonte: elaborada pelo autor.

Os provedores de nuvem pública oferecem vários serviços para implementação de aplicações utilizando o modelo FaaS. Por exemplo, o provedor AWS oferece o serviço AWS Gateway API, que implementa o redirecionamento (*gateway*) de requisições, e o serviço AWS Lambda, que permite a implementação de funções como serviço (AWS, 2015). A Figura 4.7 ilustra esses serviços em uma aplicação em três camadas. O FaaS é usado na segunda (camada de processamento), responsável pela implementação da lógica da aplicação. Na camada três, a persistência de dados é feita com o serviço de banco de dados AWS RDS.

Figura 4.7 | Exemplos de serviços da AWS para Computação Sem Servidor



Fonte: elaborada pelo autor.

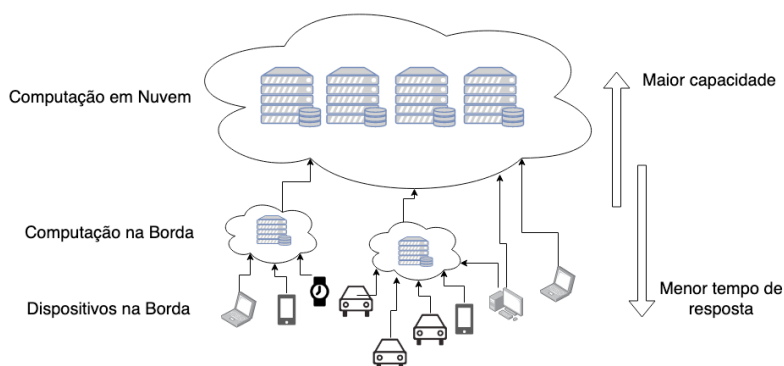
A arquitetura *Serverless* apresenta benefícios para aplicações em nuvem, mas também algumas desvantagens (BALDINI *et al.*, 2017). Entre os benefícios, o mais evidente é que os desenvolvedores não precisam se preocupar em alocar servidores, o que facilita a implantação da aplicação. Além disso, pode haver redução de custos. A alocação de servidores ou contêineres implica custos, mesmo que essas instâncias não recebam nenhuma requisição, pois o pagamento é feito de acordo com a alocação dos recursos. No caso de FaaS, só há tarifação quando uma função é efetivamente executada. Porém, há também algumas possíveis desvantagens, entre as quais podemos destacar: dificuldade de portabilidade das funções devido ao uso de tecnologias proprietárias do provedor; dificuldade de projetar a aplicação como uma composição de funções, etc.

Refleta

Uma das vantagens do modelo de Função como Serviço (FaaS) é que não há necessidade de alocação de instâncias para execução das funções. Isso é feito dinamicamente pelo provedor no momento em que a execução de uma função é requisitada. Assim, pode-se concluir que ocorre um atraso na resposta da requisição, pois os recursos ainda precisam ser alocados antes de que seja processada. Como isso pode afetar o desempenho das aplicações? Como os provedores poderiam lidar com esse problema?

Mesmo com a evolução dos modelos de arquitetura, ainda pode haver problemas de escalabilidade em soluções em nuvem, pois não deixa de ser um modelo centralizado. Por exemplo, nas aplicações de Big Data e Internet das Coisas, um grande volume de dados precisa ser transmitido até o provedor para ser processado e, depois, eventuais respostas necessitam ser enviadas até os dispositivos finais. Essas questões motivaram a consolidação do paradigma de Computação nas Bordas (*Edge Computing*), cuja ideia principal é mover o processamento dos dados para a borda da rede (SHI *et al.*, 2016). Uma visão geral da *Edge Computing* é apresentada na Figura 4.8. A execução de serviços na borda da rede pode beneficiar algumas soluções, por exemplo, aplicações de mobilidade urbana, que precisam de respostas rápidas para análise de dados do trânsito em uma determinada região.

Figura 4.8 | Visão geral da arquitetura de *Edge Computing*



Fonte: elaborada pelo autor.

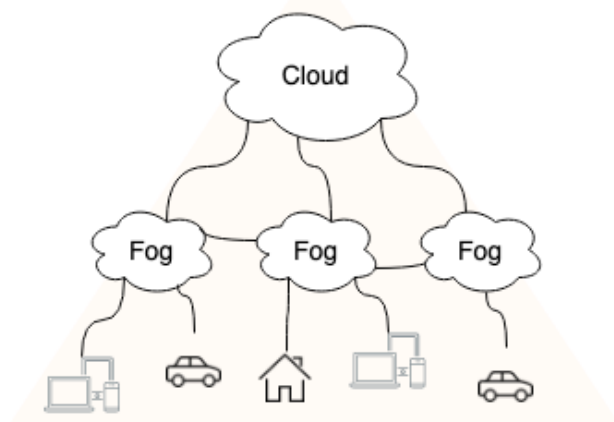
O objetivo desse modelo não é substituir a Computação em Nuvem. Na verdade, são abordagens complementares, pois as aplicações que exigem maior capacidade de armazenamento e processamento de dados continuam a ser executadas de forma centralizada no provedor de Computação em Nuvem. Um bom exemplo disso é um sistema que utiliza Aprendizado de Máquina para identificar fraudes em compras com cartão de crédito. Por outro lado, aplicações cujo requisito principal é o tempo de resposta, podem ser executadas na borda da rede, pelos próprios dispositivos ou por alguma infraestrutura dedicada de Computação nas Bordas. Em geral, é apropriada para aplicações que envolvem análise de dados em tempo real (COUTINHO; CRARNEIRO; GREVE, 2016). A seguir, são descritas algumas importantes aplicações práticas que podem se beneficiar das vantagens do modelo descentralizado da Computação em Bordas:

- Redes de sensores: redes de dispositivos com capacidade de monitoramento (sensoriamento) dispersos por áreas amplas. Essas redes são usadas para implementação de soluções em diversas áreas, como vigilância, monitoramento ambiental, etc. Os dispositivos possuem capacidade limitada, fazem uso de mecanismos de reconhecimento de contexto e topologias hierárquicas, o que torna interessante o uso de Computação nas Bordas.
- Cache de dados: a arquitetura descentralizada da Computação em Bordas permite a implementação de sistemas inteligentes de replicação e distribuição de conteúdo, inclusive para usuários móveis.
- Edifícios inteligentes (*smart buildings*): uso de sensores e atuadores para gerenciamento automático e otimizado da infraestrutura de

edifícios. Essas soluções podem ser usadas para conservar água, energia, vigilância, regulação de temperatura, entre outros. Os requisitos de reconhecimento de contexto, baixa latência e privacidade dos dados tornam mais atrativo para essas aplicações o modelo de Computação nas Bordas.

A Computação em Névoa (do inglês, Fog Computing) é uma das principais abordagens para implementação do paradigma de Computação nas Bordas. A Computação em Névoa pode ser entendida como uma plataforma para provisionamento de recursos computacionais entre os dispositivos finais e os provedores tradicionais de Computação em Nuvem (COUTINHO; CRARNEIRO; GREVE, 2016). A Figura 4.9 ilustra esse conceito de hierarquia no provisionamento de recursos virtualizados. As aplicações de análise de dados em tempo real podem usar recursos provisionados na Névoa (Fog). Aplicações que exigem a integração de dados e serviços e também maior capacidade computacional podem usar recursos provisionados na Nuvem (Cloud).

Figura 4.9 | Visão geral da arquitetura Computação em Névoa



Fonte: elaborada pelo autor.

A arquitetura descentralizada da Computação nas Bordas pode trazer muitas vantagens para o desempenho para aplicações interativa e com restrições de tempo. No entanto, essa arquitetura introduz também alguns desafios, como problemas de segurança (ROMAN; LOPEZ; MAMBO, 2018). Por se tratar de um modelo descentralizado, a Computação em Borda torna mais complexa a implementação de mecanismos de segurança, uma vez que esses mecanismos precisam proteger também cada um dos serviços e dispositivos na borda da rede. Nesse caso, é preciso lidar com a heterogeneidade de plataformas e tecnologias, assim como implantar mecanismos integrados de segurança.

Pesquise mais

O documento a seguir explica diversas tecnologias e abordagens para implementação de Edge Computing, como: Cloudlets (“mini-nuvens”), *Micro Data Centers* (MDC) e *Mobile Edge Computing* (MEC).

- COUTINHO, A. A.; CARNEIRO, E.; GREVE, F. Computação em Névoa: Conceitos, Aplicações e Desafios. **Minicursos do XXXIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, cap. 6, p. 272-284. Porto Alegre: SBC, 2016.

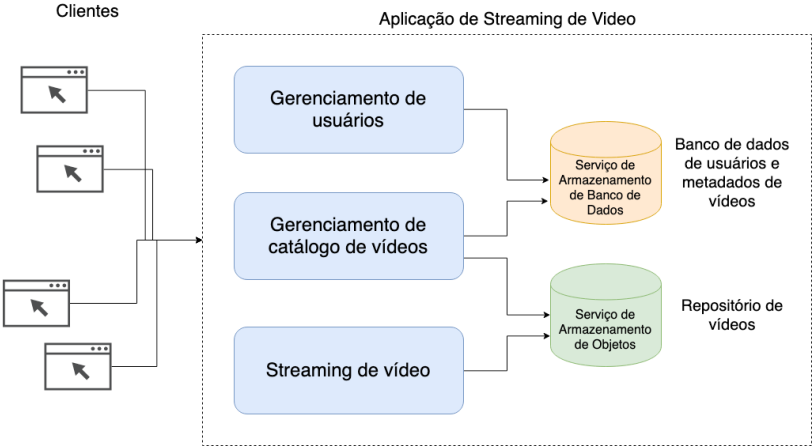
Muito bem, chegamos ao final da seção. Você aprendeu muitos conceitos relacionados com a arquitetura de aplicações em nuvem. Além disso, viu como os diferentes modelos podem beneficiar ou impactar negativamente a escalabilidade ou o desempenho dos serviços.

Sem medo de errar

Você e seus sócios têm o objetivo de criar uma nova aplicação de *streaming* de vídeo. Como líder técnico, você é o responsável pelo projeto da arquitetura da solução. Já que o negócio está em fase inicial, vocês decidiram implantar a aplicação em um provedor de Computação em Nuvem, a fim de reduzir custos e delegar a responsabilidade da administração da infraestrutura de TI. A aplicação de *streaming* de vídeo possui diversos requisitos funcionais que estão agrupados em três módulos. O primeiro consiste nas tarefas de gerenciamento de usuários (cadastro e atualização, mecanismos de autenticação, etc.). O segundo módulo envolve o gerenciamento do catálogo de vídeos da aplicação (cadastro e atualização dos dados, funcionalidades de busca e recomendação para os usuários, etc.). Por fim, o terceiro módulo funcional da aplicação é o mecanismo de distribuição dos fluxos (*streaming*) de vídeos sob demanda.

O desafio aqui é escolher um modelo de arquitetura apropriado para implementar essa aplicação. Ela possui vários módulos funcionais que podem utilizar diferentes serviços de Computação em Nuvem. Existe mais de uma solução possível para o projeto dessa arquitetura, então vamos apresentar aqui um exemplo de resposta satisfatória. A Figura 4.10 ilustra uma visão geral dos módulos funcionais da aplicação de *streaming* de vídeo.

Figura 4.10 | Visão geral dos módulos funcionais da aplicação de *streaming* de vídeo



Fonte: elaborada pelo autor.

É preciso determinar o modelo de arquitetura apropriado para organizar a implementação das funcionalidades da aplicação. Foram discutidos três modelos principais: multicamadas, microsserviços e *serverless* (na forma de FaaS). É preciso avaliar as possibilidades, considerando que foi determinado que o requisito mais importante é a escalabilidade e o desempenho do fluxo de vídeo para os usuários.

Conforme discutido no decorrer da seção, organizar a aplicação em camadas não é suficiente para alcançar o melhor grau de escalabilidade em uma aplicação em Nuvem. Modelos mais recentes como microsserviços e FaaS são mais eficazes para isso. Se o principal requisito da solução fosse em relação aos custos, a arquitetura baseada em FaaS poderia ser mais apropriada pois envolve a tarificação por tempo efetivo de execução das funções (não pelo tempo de alocação de recursos de infraestrutura, como contêineres ou máquinas virtuais). No entanto, como o principal requisito é desempenho, a opção pela arquitetura de microsserviços é mais interessante, pois pode resultar em menor tempo de resposta para as requisições. Isso ocorre porque cada microsserviço pode ser executado em um contêiner que é previamente instanciado. Por outro lado, quando se usa FaaS, os recursos para execução de uma função requisitada são alocados no momento da requisição, o que pode gerar algum atraso. Além disso, um microsserviço é um componente que inclui várias funções relacionadas. As interações (chamadas) entre funções em um mesmo componente de software são feitas por meio de desvios no fluxo de execução do código em memória. Já no modelo FaaS, as funções são componentes independentes. As chamadas entre elas são requisições enviadas pela rede. Então, as chamadas entre funções dentro de um microsserviço são

mais rápidas que as interações por rede entre funções que compõem uma aplicação no modelo FaaS. Essa análise nos permite concluir que a arquitetura de microsserviços é a mais adequada para a aplicação proposta.

Assim, cada um dos três módulos funcionais da aplicação de *streaming* de vídeo poderia ser modelado como um microsserviço. Cada um desses componentes, assim como os repositórios de armazenamento, poderia ser replicado de forma independente de acordo com a demanda para cada um deles, o que contribui para um melhor desempenho e escalabilidade da solução. Com o conteúdo assimilado até aqui, já temos os conhecimentos básicos para trabalhar na modelagem da arquitetura de aplicações utilizando serviços de computação em nuvem.

Avançando na prática

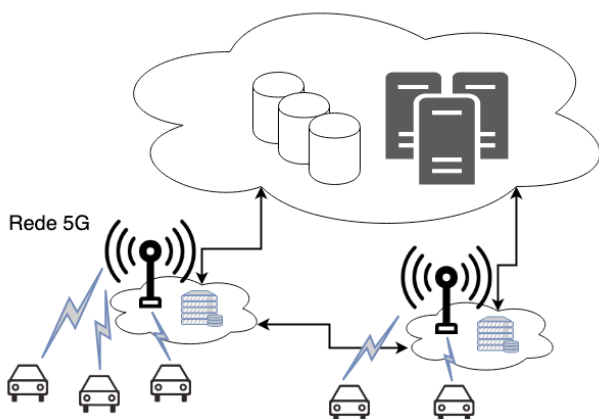
Software de Controle para Veículos autônomos

Você é analista de TI em uma empresa do setor automotivo que decidiu iniciar a fabricação de veículos autônomos, que não precisam de motoristas, pois eles possuem um sistema de controle sofisticado capaz de conduzir o veículo com segurança. Seu papel é liderar a equipe que vai implementar o software de controle para condução automática dos veículos. Você precisa, inicialmente, escolher um modelo de arquitetura para a solução desses veículos.

Resolução da situação-problema

Existem muitos aplicativos de navegação para veículos baseados em soluções em nuvem. No entanto, um software de controle para condução de veículo precisa tomar decisões em tempo real. Além do uso de serviços de inteligência artificial, esse tipo de aplicação requer baixa latência de comunicação e baixo tempo de resposta no processamento de dados. Se os veículos tivessem que se comunicar com servidores na nuvem, esses requisitos poderiam não ser atendidos. O modelo mais adequado nesse caso seria uma abordagem de *Edge Computing*, como ilustrado na Figura 4.11. Os carros poderiam se comunicar com altas taxas de transmissão por meio de uma rede sem fio 5G e aproveitar a capacidade de processamento e armazenamento de dados das estações de transmissão de dados para executar funcionalidades em tempo real. Mesmo assim, serviços em nuvem poderiam ser utilizados para agregar informações, armazenar dados históricos para análise de estatísticas e para cálculos de rotas longas que exigem dados do trânsito em várias regiões.

Figura 4.11 | Cenário de aplicação com veículos autônomos



Fonte: elaborada pelo autor.

Faça valer a pena

1. O modelo básico de arquitetura centralizada (cliente-servidor) evoluiu para um modelo com múltiplas camadas. Nesse caso, o sistema é dividido em camadas que têm responsabilidades específicas. Tradicionalmente, as aplicações Web são divididas em três camadas: camada de interface do usuário, camada de processamento e camada de dados.

Considerando a arquitetura multicamadas, é correto o que se afirma em:

- Na camada de dados, ficam os componentes responsáveis pela entrada de dados da aplicação.
- A lógica da aplicação deve ser implementada na camada de interface de usuários.
- Um componente da camada interface do usuário não deveria interagir diretamente com um componente da camada de dados.
- A separação das camadas é somente lógica, pois todos os componentes de uma aplicação precisam ser executados em um mesmo servidor.
- Por questões de consistência dos dados, não pode haver replicação de componentes da camada de processamento.

2. A arquitetura de microsserviços apresenta uma série de vantagens sobre aplicações monolíticas. Nesse contexto, avalie as afirmativas a seguir:

- I. Caso seja necessário, microsserviços que compõem uma mesma aplicação podem ser implementados em linguagens diferentes em função de requisitos específicos.
- II. Como o nível de modularização é maior nas aplicações baseadas em microsserviços, essa arquitetura permite agilizar as atualizações necessárias.
- III. Como os microsserviços são altamente coesos, testes de integração são desnecessários.

Considerando as afirmativas apresentadas, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

3. As arquiteturas mais modernas para aplicações em nuvem são representadas pelos modelos FaaS (Serverless) e Edge Computing. Sobre esses modelos, avalie as afirmativas a seguir:

- I. Uma desvantagem do modelo FaaS é a possibilidade de maior latência (maior tempo de resposta), pois os recursos precisam ainda ser alocados quando a execução de uma função é requisitada.
- II. O modelo *Edge Computing* seria apropriado para uma aplicação de mineração de dados aplicada à Inteligência de Negócios.
- III. Em uma aplicação implementada com FaaS, uma função pode efetuar requisições para um banco de dados ou outra aplicação, mas não para outra função na mesma aplicação.

Considerando as afirmativas apresentadas, é correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

Qualidade de serviço em nuvem

Diálogo aberto

Caro aluno, um dos obstáculos para migração de aplicações para provedores de Computação em Nuvem é o fato de que o acesso remoto aos serviços possa afetar negativamente o desempenho. De fato, sem uma conexão de rede de boa qualidade é difícil garantir uma experiência satisfatória para os usuários. Além disso, os provedores precisam de soluções para oferecer confiabilidade e escalabilidade para os serviços. Você conhece quais são as métricas para descrever de forma objetiva os requisitos de desempenho de aplicações? Sabe quais são os mecanismos utilizados pelos provedores para lidar com falhas e problemas de desempenho?

Para discutir esses assuntos de forma concreta, vamos relembrar o caso da nova aplicação de *streaming* de vídeo sob demanda que você decidiu começar juntamente com alguns sócios. Como responsável pelas estratégias de tecnologia do empreendimento, você precisa fazer escolhas estratégicas para a implementação da solução proposta em um provedor de nuvem pública. Os principais requisitos para a aplicação de *streaming* de vídeo são desempenho e escalabilidade. Você deve realizar duas tarefas. A primeira é definir pelo menos duas métricas que serão utilizadas para caracterizar os requisitos de desempenho da aplicação. A segunda consiste em determinar dois mecanismos que serão utilizados no ambiente do provedor para aumentar o desempenho e a escalabilidade da aplicação de *streaming*. Você deve entregar um documento no qual especifica as métricas e justifica a escolha dos mecanismos.

Para resolver essas questões, precisamos compreender como os provedores implementam soluções para garantia de Qualidade de Serviço. Nesta seção, vamos iniciar o estudo com a descrição de métricas de desempenho de rede e a definição do conceito de Acordo de Nível de Serviço (SLA). Em seguida, vamos descrever alguns indicadores de Qualidade de Serviço que são utilizados pelos provedores de Computação em Nuvem. Além disso, serão explicados mecanismos que visam aprimorar a escalabilidade, a confiabilidade e a disponibilidade de aplicações em nuvem. Por fim, vamos estudar, também, a recuperação de desastres, que consiste em mecanismos para lidar com falhas. Com esse estudo, você será capaz de compreender como podemos medir a qualidade das aplicações de forma objetiva, assim como as principais soluções para melhoria de desempenho e escalabilidade em ambientes de nuvem.

Não pode faltar

O uso de serviços de Computação em Nuvem na implementação de aplicações distribuídas pode trazer vários benefícios, como a redução de custos e a eliminação da necessidade de gerenciar equipamentos de TI. Por outro lado, existem também algumas barreiras para o uso dos serviços de provedores de nuvem pública. Em particular, o desempenho das aplicações em nuvem pode ser um desafio, se não houver disponibilidade de redes de comunicação de qualidade. Como o acesso aos serviços e dados no provedor é realizado por meio da Internet, pode haver problemas de desempenho ou até mesmo falhas na comunicação entre os componentes. Considere o caso de sistemas com requisitos de tempo real, como jogos ou aplicações de chamada de voz. Esses tipos de aplicações não funcionam de forma adequada se a conexão for de baixa qualidade. Aplicações que armazenam dados na nuvem também podem sofrer degradação de desempenho se houver atraso no acesso remoto aos dados.

Nesse contexto, precisamos utilizar mecanismos para tentar assegurar uma experiência satisfatória dos usuários de aplicações em nuvem. O primeiro ponto a considerar é como caracterizar de forma objetiva a qualidade da comunicação. Isso significa que precisamos de métricas quantitativas para descrever os requisitos mínimos de desempenho. Para lidar com essas questões, há um conceito denominado Qualidade de Serviço (QoS – *Quality of Service*). Em se tratando de redes de computadores, a QoS pode ser entendida como uma abordagem utilizada para especificar parâmetros de desempenho das aplicações, assim como os mecanismos necessários para garantir os requisitos de desempenho estabelecidos (KAMIENSKI; SADOK, 2000). Os modelos de QoS podem ser utilizados para caracterizar objetivamente os requisitos de desempenho de uma aplicação. Além disso, utilizam métricas de desempenho de rede tais como:

- Atraso: é o tempo total de transmissão de um pacote do nó remetente ao nó destinatário. Aplicações de chamadas de voz na Internet, por exemplo, requerem um atraso máximo de 150ms (CHEN; FARLEY; YE, 2004).
- Jitter: é uma medida da variação no atraso na transmissão dos pacotes. Quanto maior o jitter, pior é o desempenho de aplicações multimídia, como *streaming* de músicas na internet.
- Taxa de transmissão: é o volume de dados efetivamente transmitido entre o nó remetente e o nó destinatário. Em geral, é medida em termos de megabits por segundo (Mbps). Por exemplo, a aplicação de *streaming* de vídeo sob demanda Netflix estabelece que, para assistir

vídeos em resolução Ultra HD, é necessária uma conexão com uma taxa de pelo menos 25Mbps (NETFLIX, 2019).

- Taxa de perda: porcentagem dos pacotes que não foram entregues com sucesso para o nó destinatário. Por exemplo, se foram transmitidos 50 pacotes e apenas 40 foram efetivamente entregues ao destinatário, então a taxa de perda é de 20%, ou seja, 10 de 50 pacotes não foram entregues.

Utilizando estas e outras métricas mais específicas, os provedores especificam condições para provisão dos serviços de Computação em Nuvem em um documento chamado Acordo de Nível de Serviço (SLA – *Service Level Agreement*) (ERL; PUTTINI; MAHMOO, 2013). O SLA descreve de forma objetiva as garantias de QoS, a confiabilidade e o desempenho de cada serviço. Uma das principais métricas de QoS utilizadas por provedores de nuvem pública é a disponibilidade de um serviço, que é determinada como a porcentagem do tempo em que um serviço se mantém apto para responder corretamente as requisições de aplicações, clientes ou usuários. Por exemplo, se um serviço web ficou “fora do ar” por 8 dias durante um período de 100 dias, então pode-se afirmar que a sua disponibilidade foi de apenas 92%.

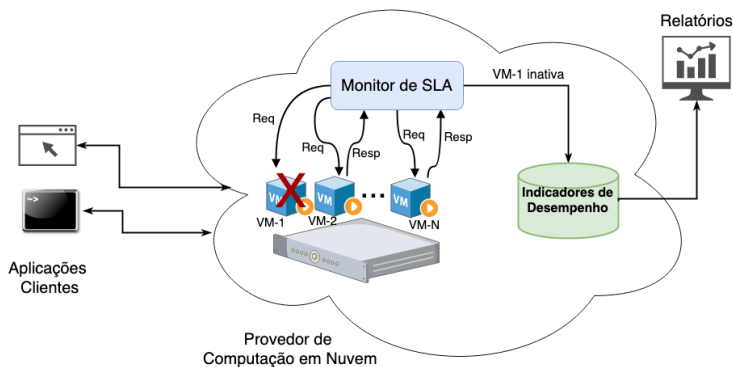
Assimile

O provedor Microsoft Azure estabelece no seu SLA para máquinas virtuais que a disponibilidade mínima para uma instância básica de máquina virtual alocada no provedor é 99,9% (MICROSOFT, [s.d.] a). O acordo define que o cliente terá direito a um ressarcimento na forma de crédito de serviços, caso a métrica de disponibilidade não seja atingida. A verificação é feita mensalmente.

Os provedores utilizam um mecanismo denominado Monitor de SLA para monitorar continuamente os indicadores de desempenho dos serviços em nuvem, a fim de verificar se eles estão de acordo com as métricas de qualidade estabelecidas no SLA (ERL; PUTTINI; MAHMOOD, 2013). O provedor mantém um banco de dados com as informações de desempenho coletadas por esse mecanismo. A Figura 4.15 mostra o papel do Monitor de SLA, considerando um exemplo no qual os serviços monitorados são máquinas virtuais. O monitor de SLA verifica periodicamente se as VMs estão ativas, enviando mensagens de requisição (req) simples. Cada máquina virtual envia uma mensagem de resposta (resp). Nesse caso, a VM-1 não responde ao monitor e então é considerada inativa. As informações são armazenadas em um banco de dados com estatísticas sobre os serviços do provedor. As informações coletadas pelo monitor de SLA podem ser usadas pelo sistema

de gerenciamento do provedor para gerar relatórios de análise de desempenho dos serviços e para fins de tarifação.

Figura 4.12 | Visão geral da atuação do Monitor de SLA



Fonte: elaborada pelo autor.

Além da disponibilidade, os provedores de Computação em Nuvem podem considerar outras métricas de Qualidade de Serviço (ERL; PUTTINI; MAHMOO, 2013). Entre as principais métricas pode-se destacar: escalabilidade, performance, confiabilidade e resiliência. A escalabilidade diz respeito à capacidade de um sistema de ajustar a quantidade de recursos computacionais de acordo com a demanda. Quando um sistema é considerado escalável, isso significa que um aumento na carga de trabalho não compromete o desempenho desse sistema, pois novos recursos podem ser alocados para que não aconteça sobrecarga. A performance é uma medida direta da capacidade de execução do serviço. Pode ser medida, por exemplo, pelo tempo de resposta, que consiste no tempo total para um cliente receber a resposta para uma requisição enviada a um servidor, incluindo o tempo gasto pelo servidor para executar a operação requisitada. A confiabilidade consiste na capacidade do serviço de operar continuamente sem falhas. Ela pode ser medida em termos do tempo médio entre falhas, ou seja, o tempo esperado entre a ocorrência de falhas. A resiliência representa uma medida da robustez de um serviço. Ela está relacionada ao grau de tolerância a falhas do serviço. A resiliência pode ser medida em termos do tempo médio para recuperação, quando da ocorrência de uma falha.

Em geral, o Monitor de SLA tem relação não somente com a disponibilidade, mas também com as demais métricas de QoS mencionadas: performance, escalabilidade, confiabilidade e resiliência. As informações coletadas por esse monitor não são usadas apenas para diagnosticar quais serviços

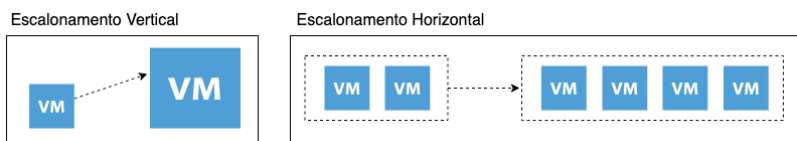
estão disponíveis (ativos). Por exemplo, o monitor de SLA verifica outros dados sobre as VMs, como uso de memória e processador. Portanto, os dados coletados servem para determinar indicadores de desempenho. Além disso, as informações do Monitor de SLA serve, também para disparar eventos que acionam outros mecanismos que visam melhorar a Qualidade de Serviço no provedor de Computação em Nuvem. Entre esses mecanismos, podemos destacar: dimensionamento automático; balanceamento de carga; recuperação de desastres (falhas) (ERL; PUTTINI; MAHMOO, 2013). Por exemplo, uma sobrecarga em um servidor identificada pelo monitor de SLA pode acionar o mecanismo de dimensionamento automático para criar réplicas desse servidor. Todos os mecanismos, em conjunto, contribuem para melhorar o desempenho dos serviços provisionados, o que se reflete nas métricas de QoS.

Exemplificando

A replicação de recursos é uma abordagem fundamental para viabilizar a implementação de diversos mecanismos que visam aumentar os indicadores de QoS. Se um serviço, como um servidor web, por exemplo, é executado em uma máquina virtual, a criação de novas instâncias (réplicas) dessa máquina permite melhorar o desempenho e a disponibilidade do serviço. Como os recursos computacionais na nuvem são virtualizados, a replicação é um processo simples e rápido. Devido a isso, uma das características dos serviços em ambientes de Computação em Nuvem é a elasticidade.

O mecanismo de dimensionamento automático (*automated scaling*) é responsável por ajustar a capacidade de um serviço em função das demandas. Se a carga de trabalho aumenta, o mecanismo aloca mais recursos, para manter a performance do serviço. Por exemplo, esse mecanismo pode automaticamente criar uma réplica de um banco de dados, para lidar com um aumento no número de consultas ao banco. Se a carga de trabalho diminui, o mecanismo libera recursos ociosos para reduzir custos. Assim, esse mecanismo confere escalabilidade aos serviços em nuvem de forma automatizada, buscando otimizar a relação entre custo e performance (MAO; HUMPHREY, 2011). O redimensionamento (escalonamento) pode ser vertical ou horizontal. O vertical corresponde a aumentar a configuração de um recurso, por exemplo, reconfigurar uma máquina virtual com 8GB de memória para 16GB. O escalonamento horizontal corresponde a criar réplicas de uma instância e é bastante utilizado em aplicações Web. Nesse caso, são criadas réplicas do servidor web ou do servidor de bancos de dados para atender a um aumento no número de requisições recebidas pela aplicação. A Figura 4.13 ilustra os conceitos de escalonamento vertical e horizontal.

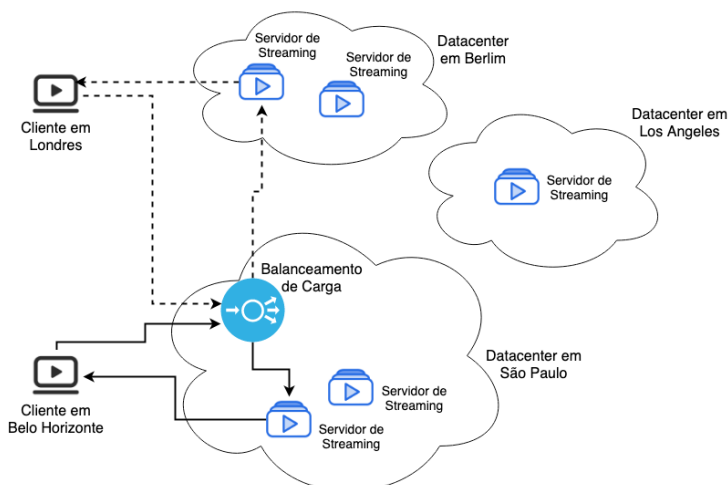
Figura 4.13 | Comparação entre escalonamento vertical e horizontal



Fonte: elaborada pelo autor.

O balanceamento de carga é um mecanismo para distribuir a demanda de trabalho entre as réplicas de um serviço. Por exemplo, cada nova requisição que chega à rede do provedor pode ser encaminhada para a réplica menos sobrecarregada. Isso resulta em melhor performance, pois diminui o tempo de resposta das requisições. Em conjunto com o dimensionamento automático, o balanceamento de carga também favorece a escalabilidade e a disponibilidade do serviço, pois se uma instância de um serviço se tornar inativa (indisponível, devido a uma falha por exemplo) as requisições a esse serviço podem ser encaminhadas para outra réplica. A Figura 4.14 mostra um exemplo de um mecanismo de balanceamento de carga para um serviço de *streaming* de vídeo. Suponha que um provedor tenha três *data centers* em diferentes regiões: Alemanha, Califórnia e São Paulo. O mecanismo de balanceamento de carga escolhe a réplica do servidor de *streaming* mais próxima do cliente para melhorar a performance do serviço. De fato, existem várias técnicas para implementar o balanceamento de carga, como distribuição uniforme, balanceamento ponderado, entre outras (KANSAL; CHANA, 2012).

Figura 4.14 | Balanceamento de carga baseado em localização

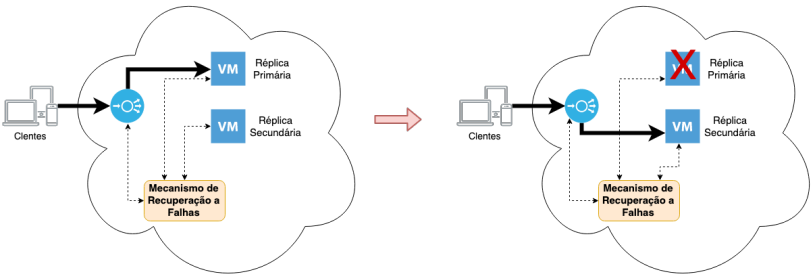


Fonte: elaborada pelo autor.

O mecanismo de recuperação de falhas trabalha em conjunto com os demais mecanismos. O objetivo do mecanismo de recuperação a falhas é identificar a ocorrência de falhas para que as requisições sejam redirecionadas somente para as réplicas do serviço que estejam ativas e funcionando corretamente. Quando uma instância falha, o mecanismo de balanceamento de carga é avisado para não redirecionar requisições para essa instância. Assim, a implementação e recuperação a falhas contribui também para aumentar a confiabilidade e a disponibilidade do serviço.

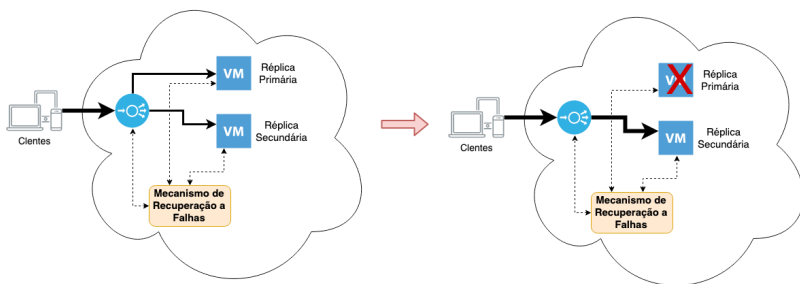
Observe que o mecanismo de recuperação a falhas depende da redundância (replicação) de recursos. Por exemplo, quando uma instância de um serviço falha, deve haver uma réplica (secundária) desse serviço já preparada para receber as requisições. De acordo com a forma como a réplica secundária é utilizada, existem dois modelos básicos de recuperação a falhas: modelo ativo-ativo e modelo ativo-passivo (ERL; PUTTINI; MAHMOO, 2013). No modelo ativo-passivo, a réplica secundária não é utilizada para atender requisições regularmente. Ela só é acionada quando a réplica principal falha. A Figura 4.15 exemplifica essa estratégia. Quando máquina virtual primária falha, as requisições dos clientes são direcionadas para sua réplica. No modelo ativo-ativo, ambas as réplicas recebem requisições (sem distinção). Quando uma das réplicas falha, então todas as requisições são redirecionadas para a outra, até que a réplica que falhou seja corrigida ou uma nova réplica seja instanciada. A Figura 4.16 mostra um cenário baseado nesse modelo.

Figura 4.15 | Visão geral do modelo ativo-passivo



Fonte: elaborada pelo autor.

Figura 4.16 | Visão geral do modelo ativo-ativo



Fonte: elaborada pelo autor.

Reflita

No modelo ativo-passivo, a réplica secundária só é utilizada caso a réplica principal falhe. A princípio, isso representa um desperdício de recursos. Então por que os provedores consideram essa abordagem? De que forma esse modelo pode ser utilizado para prover alguma vantagem para o cliente ou para o provedor?

Em geral, para cada serviço em nuvem que exige alta disponibilidade, são criadas diversas réplicas, inclusive réplicas em outro *data center* do provedor. Dessa forma, mesmo que aconteça um grave desastre em alguma região que comprometa esse *data center*, o serviço continuará disponível, pois existem réplicas em outro.

Assimile

Os mecanismos de recuperação de falhas, escalonamento automático e balanceamento de carga, aliados à escalabilidade dos serviços de Computação em Nuvem, são fundamentais para implementação de soluções de recuperação de desastres (CSA, 2011). Essas soluções são importantes para a garantir a confiabilidade do serviço e a proteção dos dados. Além disso, o provedor deve manter um plano de recuperação de desastres que, entre outras aspectos, especifica as prioridades, serviços e dados mais críticos, além das estratégias de recuperação de falhas que devem ser usadas em cada caso.

Como exemplos de desastres podemos citar falhas nos sistemas de distribuição de energia elétrica que resultam em blackouts, podendo durar horas ou até mesmo dias. Também podem ocorrer catástrofes naturais como

inundações. Nesses casos, as instalações de um *data center* poderiam ficar completamente inoperantes, o que afetaria a continuidade dos negócios que dependem da disponibilidade dos serviços do provedor. Imagine os prejuízos decorrentes da interrupção dos serviços para as aplicações dos clientes do provedor.

Pesquise mais

Em geral, os grandes provedores de nuvem pública possuem vários centros de processamento de dados dispersos por vários países. Quando um cliente usa algum serviço nesses provedores, ele pode habilitar o dimensionamento automático (*auto scaling*) e a recuperação de falhas, permitindo a replicação do serviço em diferentes regiões para aumentar o grau de disponibilidade. Verifique a infraestrutura de alta disponibilidade de alguns provedores:

- MICROSOFT AZURE. **Infraestrutura global do Azure.** [S./l., s.d.].
- AWS. **Infraestrutura global.** [S./l., s.d.].

O projeto de aplicações em nuvem envolve não somente determinar a arquitetura dos componentes da aplicação, mas também os mecanismos necessários para garantir os requisitos de desempenho. Os provedores oferecem mecanismos de balanceamento de carga, recuperação de falhas e dimensionamento automático. Esses mecanismos devem ser habilitados ou incorporados na aplicação para otimização do seu desempenho.

Pesquise mais

Projetar uma aplicação em nuvem é um processo que envolve o uso e a integração de vários serviços em nuvens, tanto para funcionalidades específicas da aplicação, como para aspectos de melhoria de desempenho. Os provedores possuem modelos de arquitetura de referência para ilustrar como os serviços podem ser combinados. Vejam alguns exemplos em:

- ASW. **Centro de Arquitetura da AWS.** [S./l. s.d.].
- AZURE. **Centro de Arquitetura do Azure.** [S./l. s.d.].

Nesta seção, aprendemos muito sobre o desempenho de aplicações em nuvem. Foram descritas várias métricas que permitem especificar de forma objetivo requisitos de Qualidade de Serviço. Também foram explicados alguns mecanismos importantes que são utilizados para aumentar a performance, disponibilidade e resiliência de aplicações em nuvem.

Sem medo de errar

Você vai começar um novo negócio que envolve o projeto, a implementação e a implantação de uma aplicação de *streaming* de vídeo sob demanda em um ambiente de Computação em Nuvem. Os principais requisitos para a aplicação de *streaming* de vídeo são: desempenho e escalabilidade. Você deve definir duas métricas, que seriam utilizadas para especificar quantitativamente os requisitos de desempenho da aplicação, e dois mecanismos que poderiam ser implantados no provedor para aumentar o desempenho e a escalabilidade.

Vamos começar com a especificação de métricas para definir os requisitos de desempenho da aplicação de *streaming* de vídeo. Foram estudadas diversas delas, como atraso, jitter (variação do atraso), taxa de transmissão e taxa de perda de pacotes. Para escolhê-las, o primeiro passo é entender as características da aplicação. Em caso de *streaming* de áudio ou vídeo sob demanda, o mais importante são os requisitos de tempo na comunicação. A razão disso é que, se a transmissão dos dados da mídia atrasar muito, pode haver pausas na reprodução do conteúdo. Nesse contexto, a métrica de desempenho mais indicada para *streaming* de vídeo sob demanda seria o atraso (latência) de transmissão de dados. Estabelecer um valor máximo para o atraso é muito importante, porque *streaming* de mídias depende de um fluxo regular de reprodução do conteúdo, o que pode ser afetado por longos atrasos no tempo de chegada dos dados. Para essas aplicações é muito importante que o atraso médio seja baixo.

Além disso, mídias com conteúdo de alta qualidade correspondem a arquivos muito grandes, o que resulta na necessidade de transmissão de muitos dados em curtos períodos. Assim, uma segunda métrica importante para a aplicação de *streaming* de vídeo seria taxa de transmissão. Nesse caso, é fundamental especificar uma taxa de transmissão mínima de acordo com a qualidade do vídeo que será transmitido. Vídeos em alta resolução não poderiam ser distribuídos sob demanda para conexões que não são capazes de sustentar uma taxa mínima de transmissão de dados.

Portanto as duas métricas utilizadas para especificar os requisitos de desempenho da aplicação de *streaming* de vídeo seriam: atraso máximo e taxa de transmissão mínima. As outras duas métricas estudadas, jitter (variação do atraso) e taxa de perda, não deixam de ser relevantes, mas são menos importantes para uma aplicação de *streaming* de vídeo sob demanda. Por exemplo, o uso de um buffer para carregar uma parte do vídeo antes do início da reprodução é suficiente para evitar uma significativa perda de desempenho devido a um jitter alto. Em relação à perda de pacotes, existem

técnicas capazes de regenerar quadros do vídeo de forma que, se a taxa de perda não for muito alta, a qualidade não é comprometida. Em resumo, todas as métricas são importantes, mas um limite máximo para atraso e um limite mínimo para taxa de transmissão são os mais relevantes para aplicações de *streaming* de vídeo sob demanda.

Foram estudados alguns mecanismos de QoS, como Monitor de SLA, balanceamento de carga, recuperação de falhas e dimensionamento automático. Considerando as prioridades de desempenho e escalabilidade, a escolha mais adequada seria o dimensionamento automático em conjunto com o balanceamento de carga. Dessa forma, as instâncias dos servidores de *streaming* poderiam ser replicadas para atender a uma demanda crescente de usuários e ao balanceamento de carga favorece o tempo de resposta, pois otimiza o uso das réplicas existentes. Novamente, todos os mecanismos são importantes, mas esses dois são imprescindíveis para a escalabilidade de aplicações de *streaming* de vídeo.

O entendimento das métricas de QoS e dos mecanismos usados para otimizar o desempenho é muito importante para o projeto e implantação de aplicações em nuvem. A especificação da arquitetura da aplicação deve contemplar não apenas os serviços em nuvem para as funcionalidades específicas da aplicação, mas também mecanismos oferecidos pelo provedor para melhorar a segurança, o desempenho e a confiabilidade da solução.

Avançando na prática

Abordagens de Recuperação a Falhas

Aplicações em algumas áreas, como no setor financeiro, saúde e governamental, apresentam requisitos bastante rigorosos em relação à confiabilidade e disponibilidade dos serviços. Nesses casos, é importante o uso de mecanismos de recuperação a falhas. Existem dois modelos básicos para isso: ativo-ativo e ativo-passivo. Faça uma análise comparativa desses modelos a fim de destacar suas vantagens e desvantagens, em termos de custo e desempenho.

Resolução da situação-problema

A estratégia básica de recuperação a falhas é a redundância: recursos adicionais são usados em caso de falha dos recursos inicialmente alocados, assim o serviço continuará disponível. Vimos que existem dois modelos

básicos para mecanismos de recuperação a falhas: ativo-ativo e ativo-passivo. Para compreender melhor a diferença entre eles, vamos considerar que, para um dado serviço em nuvem, existe a instância principal do serviço e uma instância secundária (uma réplica criada para viabilizar a implementação da recuperação de falhas). No modelo ativo-passivo, a réplica secundária é utilizada apenas no caso de falha na instância principal. No modelo ativo-ativo, a réplica secundária é utilizada no balanceamento de carga para ajudar no processamento das requisições ao serviço. A vantagem do modelo ativo-ativo é um custo menor, pois não há ociosidade de uma réplica secundária alocada somente para recuperação a falhas. Por outro lado, esse modelo implica em um risco. Se a réplica secundária é utilizada no balanceamento de carga regularmente, ela pode estar sobrecarregada no momento em que a instância principal falhar. Nesse caso, a falta de recursos disponíveis na réplica secundária pode comprometer o desempenho.

Faça valer a pena

1. O sistema de gerenciamento de um provedor de Nuvem Pública gera relatórios anuais sobre o desempenho dos serviços utilizados pelos clientes. Um relatório do ano de 2017 mostrou que uma determinada máquina virtual ficou inativa por um período total de 18 dias e 6 horas durante aquele ano.

Considerando os dados do relatório e o conceito de disponibilidade explicado nesta seção, pode-se afirmar que a disponibilidade dessa máquina virtual no ano de 2017 foi de:

- a. 90%
- b. 95%
- c. 99%
- d. 99,9%
- e. 100%

2. Os provedores de Computação em Nuvem podem utilizar diversas métricas de Qualidade de Serviço, como disponibilidade, escalabilidade, performance e resiliência. Essas métricas servem para determinar de forma objetiva os requisitos de uma aplicação. Considerando as métricas de qualidade de serviço, avalie as afirmativas a seguir:

- I. A resiliência é uma métrica que pode ser usada para caracterizar o grau de tolerância a falhas de um serviço.

- II. A performance descreve os requisitos de QoS de uma aplicação em termos do tempo necessário para se recuperar de uma falha.
- III. Diz-se que uma aplicação tem escalabilidade quando ela responde rapidamente as requisições recebidas.

Está correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

3. Diversos mecanismos são usados pelos provedores em nuvem para medir e melhorar os indicadores de Qualidade de Serviço.

Considerando esses mecanismos, avalie as afirmativas a seguir:

- I. O Monitor de SLA tem um papel limitado ao diagnóstico e sua ausência não afetaria o desempenho de serviços em um provedor.
- II. O balanceamento de carga requer que os serviços sejam replicados e pode utilizar diferentes políticas para distribuir as requisições entre as réplicas.
- III. O dimensionamento automático (*auto scaling*) está diretamente relacionado com a característica de elasticidade rápida dos serviços em nuvem, pois permite adaptar a quantidade de recursos computacionais alocados.

Sobre os mecanismos de QoS, está correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

Segurança e privacidade em nuvem

Diálogo aberto

Prezado estudante, a segurança e a privacidade dos dados é um dos aspectos mais controversos da Computação em Nuvem. Por um lado, algumas empresas e profissionais de TI fazem questionamentos em relação à migração de dados e aplicações para a nuvem, alegando que o uso de provedores públicos aumenta a exposição dos dados e as chances de vulnerabilidades de segurança. Além disso, os recursos dos provedores são compartilhados entre diversos clientes, de forma que falhas de proteção e isolamento poderiam permitir que um cliente tivesse acesso aos dados de outros. Por outro lado, o *data center* dos provedores conta com uma infraestrutura mais robusta e confiável do que a maioria das empresas poderiam manter em suas próprias instalações. Além disso, os provedores contam com profissionais altamente especializados para implementação dos mecanismos de segurança a um custo relativamente menor, e o crescimento de uso da Computação em Nuvem mostra que eles estão tendo sucesso com os problemas de segurança, caso contrário as empresas não estariam migrando suas soluções de TI para a nuvem. De fato, um estudo da Google Cloud em parceria com o MIT mostra um aumento significativo da confiança das empresas em relação à segurança em nuvem (GOOGLE, 2017). O estudo mostra, ainda, que o aumento da segurança é o segundo fator mais importante para as empresas na decisão de utilizar serviços em nuvem.

Vamos retomar o nosso estudo de caso no qual você decidiu começar um empreendimento com alguns sócios com o objetivo de desenvolver uma nova aplicação de *streaming* de vídeo. Como especialista em Tecnologia da Informação, você é responsável pelo projeto e implementação da solução. Um dos problemas de segurança e privacidade para essa aplicação de *streaming* de vídeo é evitar que usuários não autorizados assistam. É preciso garantir que somente os clientes da empresa possam acessar o conteúdo sob demanda. Sua tarefa é caracterizar esse tipo de ameaça e descrever os mecanismos de segurança que poderiam ser utilizados para lidar com o problema.

Para resolver essas questões, é preciso conhecer os fundamentos sobre segurança em Computação em Nuvem. Esse é o objetivo desta seção. Inicialmente, vamos aprender as principais propriedades que caracterizam, de forma geral, um sistema computacional seguro. Depois, vamos explicar as ameaças e as vulnerabilidades para aplicações em ambientes de nuvem, além de alguns dos mecanismos que podem ser usados para melhorar a segurança

das soluções. Por fim, vamos descrever os conceitos básicos sobre de gerenciamento de riscos. Esses assuntos são de extrema relevância, principalmente para aplicações críticas ou para aquelas que envolvem dados sigilosos.

Bons estudos!

Não pode faltar

Embora a Computação em Nuvem possa trazer vários benefícios, existem alguns desafios na adoção de soluções. Você conhece as principais dificuldades apontadas pelas empresas em relação ao uso de serviços em provedores de nuvem pública? Entre essas dificuldades, pode-se destacar as questões de segurança e privacidade, de acordo com um estudo realizado em 2018 (TUCKER, 2019). Isso acontece porque o acesso às aplicações nos provedores é feito por meio da Internet. Assim, existe uma maior exposição dos dados e serviços em comparação ao cenário no qual uma empresa mantém sua própria infraestrutura de TI, com a comunicação entre os componentes realizada através de uma rede local. Dessa forma, a maior exposição dos dados na internet, no modelo de Computação em Nuvem, aumenta as chances de ataques que buscam violar a segurança das aplicações. Portanto, esse tipo de projeto exige muito cuidado com aspectos da segurança dos serviços e da privacidade dos dados. São imprescindíveis o uso de mecanismos de segurança e o estabelecimento de uma estratégia de gerenciamento de riscos.

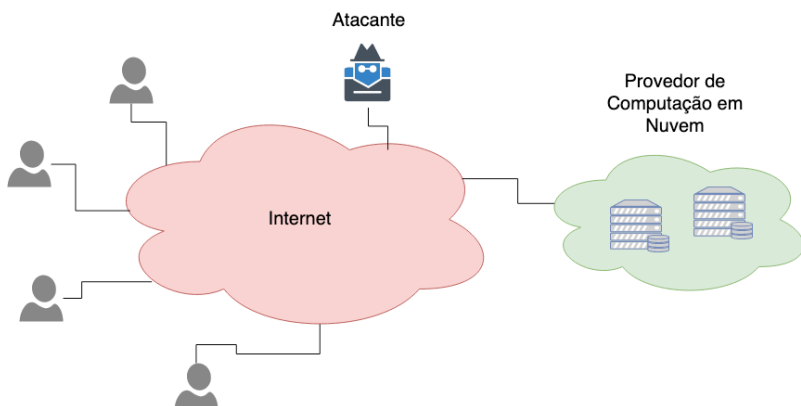
Antes de conhecer as principais vulnerabilidades de aplicações em nuvem e os principais mecanismos de segurança, precisamos compreender os princípios de segurança da informação e quais são as propriedades de um sistema seguro. Conforme apresentado na literatura (KUROSE; ROSS, 2013; TANENBAUM; STEEN, 2008), as propriedades básicas de uma comunicação segura e confiável são:

- **Confidencialidade:** sigilo do conteúdo dos pacotes transmitidos na rede.
- **Integridade:** garantia de que os dados transmitidos não podem ser alterados.
- **Autenticidade:** confirmação da identidade das partes envolvidas na transmissão dos dados.
- **Disponibilidade:** garantia de que um sistema estará apto para realizar as operações de transmissão ou processamento dos dados.

A internet é uma rede que não inclui em sua arquitetura a implementação das propriedades de uma comunicação segura (KUROSE; ROSS, 2013).

Dessa forma, a infraestrutura de rede dos provedores de acesso à internet (ISPs – *Internet Service Providers*) é considerada um meio de comunicação inseguro. Portanto, uma aplicação que utiliza a internet como rede de comunicação precisa ser protegida por meio de mecanismos de segurança, como criptografia e controle de acesso. Nesse contexto, o projeto de aplicações em nuvem precisa contemplar estratégias e ferramentas de segurança para garantia de autenticidade, confidencialidade e integralidade dos dados transmitidos entre os clientes e o provedor. A Figura 4.17 ilustra esse cenário. A comunicação de dados entre os usuários legítimos e os serviços disponíveis no provedor está sujeita às ações de agentes maliciosos que visam realizar ataques que exploram eventuais vulnerabilidades das aplicações.

Figura 4.17 | Comunicação através de um meio inseguro, como a Internet



Fonte: elaborada pelo autor

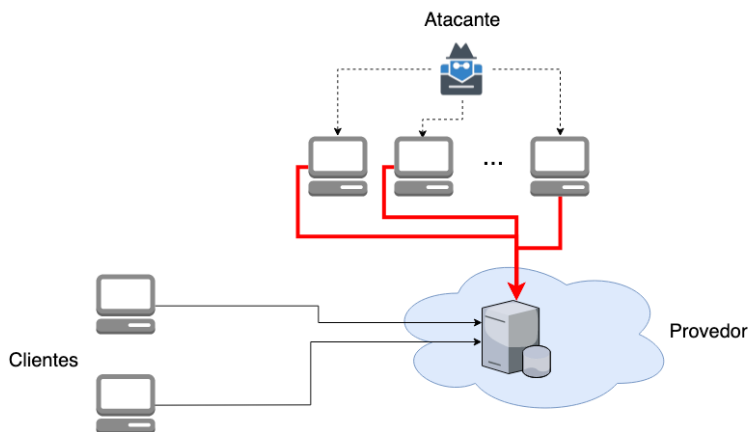
Assimile

A confidencialidade é violada quando um atacante consegue verificar o conteúdo de uma mensagem. Por exemplo, quando um atacante é capaz de interceptar e ler os dados do cartão de crédito de um usuário que faz uma compra em uma aplicação na nuvem. A confidencialidade está diretamente ligada à privacidade, pois ela significa que somente as partes autorizadas podem acessar os dados. A integridade é violada quando os dados são alterados ou corrompidos por uma ação maliciosa. Por exemplo, quando um atacante é capaz de acessar e alterar o banco de dados de uma empresa que está em um provedor de nuvem pública. É importante observar que uma mensagem pode ser alterada sem que seu conteúdo tenha sido conhecido pelo atacante. Ou seja, violar a integridade não implica, necessariamente, em violar a confidencialidade.

Vamos discutir as principais ameaças e vulnerabilidades que podem afetar as aplicações em nuvem para, depois, explicar os alguns mecanismos e políticas que podem ser utilizados para lidar com esses problemas e melhorar o nível de segurança das aplicações. Conforme explicado em (ERL; PUTTINI; MAHMOOD, 2013), as principais ameaças de segurança para aplicações em Nuvem são: interceptação de tráfego, negação de serviço e ataques de virtualização. Vamos conhecer a definição e as características de cada um desses tipos de ataque?

- **Interceptação de tráfego** - ocorre quando uma entidade não autorizada é capaz de obter as informações transmitidas entre provedor em nuvem e clientes, de forma que a confidencialidade dos dados é violada (ERL; PUTTINI; MAHMOOD, 2013). Esse tipo de ataque é difícil de ser detectado, pois é passivo, no sentido em que não há modificações nos dados ou sistemas. O atacante apenas copia os dados transmitidos na rede para ter acesso a informações sigilosas dos provedores e seus clientes. Esse cenário caracteriza uma forma de espionagem.
- **Negação de Serviço (DoS – Denial of Service)** - tem o objetivo de afetar a disponibilidade dos serviços. Esse ataque consiste em sobrecarregar o serviço com um grande volume de requisições de forma que não consiga mais responder aos clientes legítimos com um desempenho satisfatório (ERL; PUTTINI; MAHMOOD, 2013). A Figura 4.18 mostra um exemplo de um ataque de DoS. Nesse caso, um servidor em um provedor de nuvem é sobrecarregado com requisições disparadas de computadores que foram invadidos e controlados por um atacante. Com a sobrecarga de requisições em um curto período, o servidor terá dificuldades em atender as requisições dos clientes legítimos.

Figura 4.18 | Cenário exemplo de um ataque de DoS



Fonte: elaborada pelo autor.

Refleta

Com a característica de elasticidade rápida, os serviços em nuvem são capazes de alocar mais recursos de forma dinâmica para atender a crescimento na demanda. Então, como os ataques de DoS podem afetar serviços em nuvem, já que esses serviços poderiam lidar com a sobrecarga?

Para resistir aos ataques de DoS, os provedores identificavam e bloqueavam os endereços dos nós da rede de onde os atacantes disparavam as requisições. No entanto, esses ataques ficaram mais sofisticados. Em vez de disparar um grande volume de requisições de alguns nós, o que facilitava a identificação da fonte do ataque, os atacantes passaram a disparar poucas requisições de grande número de computadores dispersos pela internet. Essa estratégia é denominada ataque de DoS distribuído (DDoS – *Distributed DoS*). Assim, é muito difícil para o provedor distinguir as requisições enviadas por clientes confiáveis das requisições geradas pelos atacantes em computadores que foram invadidos.

Exemplificando

Os provedores de nuvem pública oferecem serviços para detecção e mitigação de ataques de DDoS para proteger as aplicações hospedadas em nuvem. Entre os exemplos, podemos citar o *AWS Shield* (AWS, 2019) e *Azure DDoS Protection* (MICROSOFT, 2019). Esses serviços visam manter a disponibilidade dos serviços em nuvem, monitorando o padrão de tráfego das aplicações para identificar ataques e atenuar os seus efeitos no desempenho das aplicações. Esses serviços de proteção contra DDoS também ajudam os clientes a reduzir custos, evitando que recursos sejam alocados para atender requisições decorrentes de ataques ou falhas.

Outra ameaça bastante relevante para aplicações em nuvem são os **ataques de virtualização**, que buscam explorar eventuais falhas e vulnerabilidades nas ferramentas de virtualização utilizadas pelos provedores (ERL; PUTTINI; MAHMOOD, 2013). Nesse caso, o atacante pode conseguir algum nível de controle sobre a infraestrutura de TI do provedor. Outro problema é que pode haver violação da privacidade dos dados dos clientes do provedor. Como a virtualização permite que diversos recursos compartilhem um mesmo recurso físico, pode haver dados de diversos clientes em um mesmo equipamento. Os ataques às ferramentas de virtualização podem conseguir violar os mecanismos de proteção que existem para manter um

isolamento entre os recursos e dados de diversos clientes que compartilham a infraestrutura do provedor .

Pesquise mais

Leia o documento referenciado a seguir para conhecer um pouco mais sobre os riscos de aplicações em ambientes de Computação em Nuvem, assim como ações de proteção que devem ser tomadas a fim de melhorar a confiabilidade.

- MELO, K. Segurança na Nuvem: por onde começar? **ComputerWorld**, publicado em 29 fev. 2016.

Além das ameaças relacionadas com segurança de dados e redes, os provedores de Computação em Nuvem também precisam lidar com ameaças relacionadas a outras categorias de segurança, como governança, conformidade e questões legais (GONZALEZ *et al.*, 2013). Entre os principais problemas nessas categorias podemos citar:

- Baixo nível de controle administrativo sobre a segurança dos dados por parte dos clientes de um provedor.
- Dependência das tecnologias e políticas de segurança adotadas pelo provedor
- Problemas relacionados aos requisitos de confiabilidade e políticas de auditoria estabelecidos em acordos de qualidade de serviço (SLA).
- Jurisdição dependente da localização dos provedores onde estão armazenados os dados, pois alguns provedores possuem *data centers* em países diferentes.

Agora que já conhecemos algumas das principais ameaças para aplicações em nuvem, vamos estudar os principais mecanismos de segurança disponíveis. Entre eles, podemos mencionar: criptografia, gerenciamento de acesso e identidade, autenticação unificada e imagens fortalecidas de máquinas virtuais.

A criptografia consiste em técnicas que permitem disfarçar os dados enviados de forma que um atacante não consiga obter nenhuma informação dos dados interceptados (KUROSE; ROSS, 2013). Isso implica codificar os dados de modo que somente o destinatário legítimo poderá decifrá-los, tornando-os, assim, ininteligíveis para terceiros. A principal aplicação da criptografia é assegurar a confidencialidade dos dados armazenados nos provedores de nuvem pública e dos dados transmitidos entre um provedor e seus clientes. Além disso, a criptografia também pode ser usada em mecanismos

para garantia de integridade e autenticidade. Uma vez que a criptografia pode ser usada na implementação de diversos mecanismos, ela é considerada um conceito fundamental para segurança de sistemas computacionais.

Os algoritmos utilizados para codificar os dados são de conhecimento público, mas para decodificá-los é necessário um código secreto denominado chave de criptografia. Existem duas abordagens: a criptografia de chaves simétricas e a criptografia de chave pública (KUROSE; ROSS, 2013). Na primeira abordagem, existe uma única chave que é utilizada pelo remetente para criptografar os dados que serão enviados ou armazenados em nuvem. O receptor precisa de uma cópia dessa chave para decodificar os dados. Na abordagem de chave pública, o processo envolve um par de chaves. Para o destinatário receber os dados, ele precisa gerar o seu par de chaves. Uma delas é denominada pública e não precisa ser mantida em segredo. A outra chave é denominada privada e deve ser mantida em segredo. O remetente usa a chave pública do destinatário para criptografar os dados. Assim, os dados só podem ser decifrados com a chave privada correspondente mantida em segredo pelo destinatário.

Assimile

Comparada com a abordagem criptografia de chave pública, a de chaves simétricas tem a vantagem de exigir menos capacidade computacional na criptografia dos dados. A desvantagem é que, para decodificar os dados, o destinatário precisa da mesma chave que foi usada pelo receptor para criptografá-los. Então, é necessário um mecanismo seguro para compartilhar a chave. A criptografia de chave pública resolve esse problema, pois os dados são criptografados com uma chave que pode ser conhecida por qualquer um, uma vez que não serve para decifrar os dados. Na prática, as duas abordagens são utilizadas em conjunto. A criptografia de chaves públicas, que tem maior custo, é usada somente para compartilhar com segurança uma chave única entre o remetente e o emissor que, então, usam essa chave única para transmitir os dados por meio de criptografia de chave simétrica, que apresenta menor custo.

Os algoritmos de criptografia também podem ser usados em funções de hash para verificar a integridade dos dados (ERL; PUTTINI; MAHMOOD, 2013). Um código hash do conteúdo a ser transmitido é gerado pelo receptor e disponibilizado ao destinatário. Quando recebe os dados, o destinatário gera novamente o código hash e compara com o código gerado pelo receptor. Se os códigos não são iguais é porque o conteúdo foi alterado durante a transmissão e deve, portanto, ser descartado. Se os códigos são iguais então

é porque a integridade dos dados foi mantida. A criptografia também é utilizada em mecanismos de assinatura digital que são importantes na garantia de autenticidade.

Os mecanismos de criptografia podem ser utilizados para garantir a confidencialidade e integridade dos dados transmitidos através da Internet. No caso do uso de nuvem pública, esses mecanismos também podem ser usados para confidencialidade e integridade dos dados armazenados no provedor. No entanto, esse processo pode afetar o desempenho, devido ao tempo gasto para criptografar os dados nas operações de escrita e para descriptografar os dados nas operações de leitura. Essa opção de armazenar os dados de forma criptografada é oferecida pela maioria dos provedores e serve para lidar com ataques de virtualização e garantir a privacidade mesmo que estejam em equipamentos compartilhados.

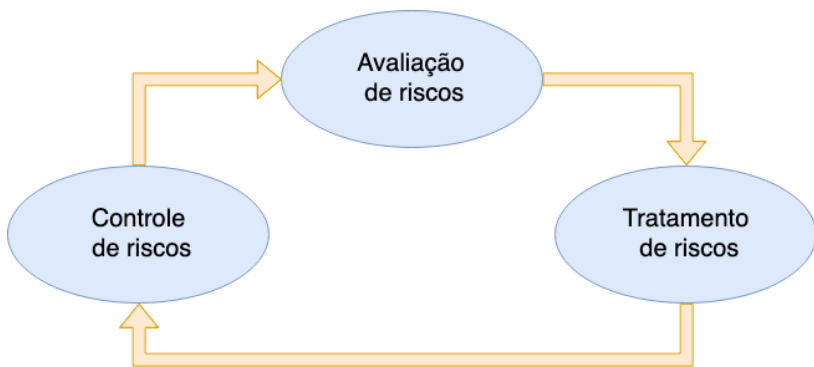
Outra solução de segurança importante nos provedores de nuvem pública é o Gerenciamento de Acesso e Identidade (IAM – *Identity and Access Management*) (ERL; PUTTINI; MAHMOOD, 2013). Esse recurso é utilizado pelos provedores, principalmente, para implementação de políticas de controle acesso. O IAM permite o gerenciamento e a autenticação de usuários, assim como o controle de privilégios para grupos de usuários e gerenciamento de credenciais. Esse tipo de recurso é importante na implementação de soluções de segurança para lidar com ataques de negação de serviços, autenticação fraca e violação de privacidade no acesso a dados e serviços. Se um cliente utiliza serviços em vários provedores, então é importante também para questões de controle de acesso o uso de mecanismos de autenticação unificada (SSO – *Single Sign On*). Esses mecanismos oferecem uma solução segura para autenticação em vários provedores utilizando as mesmas credenciais.

Outro recurso bastante utilizado pelos provedores de Computação em Nuvem para aprimorar a segurança dos serviços é o uso de Imagens fortalecidas de Máquinas Virtuais (*Hardened VM Images*) (ERL; PUTTINI; MAHMOOD, 2013). Sabemos que as instâncias de máquinas virtuais são criadas a partir de imagens disponíveis no provedor. As imagens fortalecidas são aquelas que foram configuradas por especialistas considerando políticas de segurança rigorosas para eliminar possíveis vulnerabilidades. Dessa forma, quando um cliente cria uma VM a partir de uma imagem fortalecida, ele sabe que o sistema operacional dessa VM já foi configurado com as melhores práticas de segurança conhecidas.

Os mecanismos de segurança devem ser usados em conjunto e aprimorados continuamente para lidarem com eventuais ações maliciosas contra as aplicações em ambientes de nuvem. É importante ressaltar que a segurança

não é responsabilidade somente do provedor de Computação em Nuvem; o cliente dos serviços também precisa cooperar. Para assegurar a confiabilidade e a segurança de uma aplicação, é importante a definição de uma estratégia de gerenciamento de riscos, que permitirá que uma empresa possa lidar com problemas de segurança, tanto em relação a segurança de dados e redes como em relação à governança e conformidade dos serviços, ao migrar suas aplicações para um provedor de nuvem pública. Como ilustrado na Figura 4.19, o gerenciamento de riscos envolve três etapas: avaliação, tratamento e controle (ERL; PUTTINI; MAHMOOD, 2013).

Figura 4.19 | Ciclo do processo de gerenciamento de riscos



Fonte: elaborada pelo autor.

A etapa de avaliação de riscos consiste em analisar as possíveis vulnerabilidades e ameaças no ambiente operacional do provedor de nuvem. Devem ser analisados o histórico de incidentes de segurança e as características da rede do provedor. Nesta etapa, os riscos são identificados e classificados em termos do impacto que podem causar. A etapa de tratamento de risco envolve a definição de estratégias e políticas de segurança e a realização de ações para evitar ou atenuar o efeito dos identificados na etapa anterior. Nesta fase, também precisa estar claro quais são as responsabilidades do provedor e dos clientes; algumas delas podem, inclusive, ser delegadas para empresas ou profissionais especializados. O último estágio (controle de riscos) é o momento de diagnosticar a eficácia das ações de segurança implementadas, em que podem ser definidas formas de aprimorar os mecanismos de segurança. Esse processo de gerenciamento de risco deve se repetir periodicamente, pois novas ameaças podem surgir, novas tecnologias e aplicações podem ser incluídas no ambiente do provedor, assim como podem ser introduzidos mecanismos de segurança mais sofisticados.

Nesta seção, aprendemos sobre as propriedades de um sistema seguro e sobre as principais ameaças e soluções de segurança para aplicações em nuvem. O uso de criptografia permite garantir a privacidade e a integridade dos dados, mesmo em provedores de nuvem pública. Além disso, os provedores contam com especialistas em segurança, e os *data centers* são projetados para garantir alto grau de disponibilidade e segurança para as aplicações. É importante conhecer os serviços de segurança e as boas práticas conforme recomendado por cada provedor. O projeto da arquitetura das aplicações deve contemplar mecanismos de segurança para lidar com eventuais ameaças e assegurar a privacidade dos dados.

Sem medo de errar

Você é o responsável pelo projeto e desenvolvimento de uma nova aplicação de *streaming* de vídeo sob demanda. Considerando a fase de projeto da aplicação, seu papel consiste em tomar decisões estratégicas para a implementação da solução visando principalmente desempenho e segurança. Você precisa especificar a arquitetura da solução, levando em conta que o requisito de segurança mais importante para essa aplicação de *streaming* de vídeo é evitar que usuários não autorizados assistam. Ou seja, você deve utilizar soluções de segurança para assegurar que somente os clientes legítimos da empresa possam acessar o conteúdo sob demanda. Sua tarefa é caracterizar esse tipo de ameaça e explicar os mecanismos de segurança que poderiam ser utilizados.

O problema de segurança é que, quando o conteúdo de um vídeo é enviado para um cliente da aplicação, os dados poderiam ser copiados por um agente malicioso, uma vez que esses dados são transmitidos pela internet. Essa cópia não autorizada deve ser impedida para que não sejam violados os direitos autorais do vídeo. Essa ameaça de segurança é caracterizada como interceptação tráfego e viola a confidencialidade dos dados da aplicação de *streaming*, pois um usuário não autorizado teria acesso ao conteúdo. Além da interceptação dos dados transmitidos através da internet, existe a ameaça de que os vídeos possam ser copiados diretamente das unidades de armazenamento no provedor de nuvem por um usuário não autorizado. Isso é possível se houver alguma falha de controle de acesso ou ataque de virtualização.

Para lidar com essas ameaças, pelo menos dois mecanismos de segurança poderiam ser considerados no projeto da aplicação: criptografia e gerenciamento de acesso e identidade. Mecanismos de criptografia seriam utilizados de duas formas. Primeiro, para armazenar os vídeos de forma criptografada. Segundo para criptografar os dados durante o fluxo de distribuição

(*streaming*). Dessa forma, mesmo que os dados armazenados no provedor fossem acessados indevidamente ou interceptados durante a transmissão, o usuário não autorizado não conseguiria decodificar os vídeos, a não ser que fosse capaz de descobrir as chaves de criptografia utilizadas no processo. Além disso, deveria ser configurado o mecanismo de Gerenciamento de Acesso e Identidade (IAM) para implementar políticas de controle de acesso aos dados da aplicação no provedor a partir das credenciais de usuários. O controle de acesso permite assegurar o isolamento e a proteção dos recursos alocados por um cliente mesmo que a infraestrutura física seja compartilhada com outros. O Gerenciamento de Acesso e Identidade é importante para evitar acesso não autorizados aos dados de uma aplicação hospedada no provedor.

Os mecanismos de segurança oferecidos pelos provedores podem ser utilizados no projeto da arquitetura aplicações, a fim de se garantir a segurança e a privacidade dos dados. É importante uma verificação periódica das eventuais vulnerabilidades e ameaças para que seja possível aprimorar os mecanismos de segurança.

Avançando na prática

Segurança no Download de Aplicativos

Considere que você foi contratado como analista de TI por uma empresa que mantém uma loja online de aplicativos. Os usuários acessam a loja para fazer download para seus dispositivos móveis. Existe um problema de segurança que consiste no fato de que o arquivo executável (do aplicativo) recebido usuário seja diferente do arquivo disponível para download na loja. Isso pode acontecer por falhas na transmissão que de alguma forma corromperam o conteúdo do arquivo durante o envio dos dados através da Internet. Outra possibilidade é que um atacante tenha interceptado e alterado intencionalmente o arquivo para injetar um código executável malicioso no arquivo original. Sua tarefa é caracterizar como essa ameaça viola as propriedades de segurança e descrever algum mecanismo que possa ser utilizado para lidar com o problema descrito.

Resolução da situação-problema

Aprendemos sobre quatro propriedades de um sistema seguro: autenticidade, confidencialidade, integridade e disponibilidade. As ameaças descritas não afetam a disponibilidade da aplicação, pois a loja continua ativa para

atender as solicitações de download. Também não afetam a autenticidade, pois o atacante não finge sua identidade para fazer o download de um aplicativo na loja. Também não há que se falar em confidencialidade, pois não há sigilo sobre o arquivo executável dos aplicativos. O problema descrito, de fato, é uma possível violação da integridade do arquivo durante sua transmissão pela rede.

Para lidar com essas ameaças poderia ser utilizado algum mecanismo de *Hashing*. Nesse caso, o mecanismo poderia funcionar da seguinte maneira: a loja utilizaria uma função de hash para gerar um código hash correspondente para cada arquivo executável de aplicativo disponível na loja. O usuário deveria fazer o download do arquivo e do código hash. Quando o download for concluído, o usuário utilizaria a mesma função de Hash para gerar o código e então faria uma comparação dos códigos; Se fossem iguais, é porque a integridade do arquivo foi mantida. Se os códigos fossem diferentes, é porque o arquivo foi alterado de alguma forma e então o download teria que ser feito novamente. Dessa forma, mecanismos de Hashing podem ser utilizados para verificar a integridade de arquivos executáveis transferidos de servidores na internet para evitar a execução de código executável malicioso que, eventualmente, tenha sido incluído no arquivo por um atacante.

Faça valer a pena

1. O uso da Computação em Nuvem traz diversos benefícios para as aplicações. No entanto, é preciso muita atenção aos aspectos de segurança, pois os dados e serviços em ambientes de nuvem estão expostos a muitas ameaças, como ataque de negação de serviço.

O ataque de negação de serviço afeta qual propriedade de segurança de uma aplicação hospedada em nuvem?

- a. Confidencialidade.
- b. Autenticidade.
- c. Integridade.
- d. Disponibilidade.
- e. Privacidade.

2. Existem diversos mecanismos de segurança que podem ser utilizados em ambientes de Computação em Nuvem. Nesse contexto, avalie as afirmativas a seguir:

- I. O uso de criptografia se limita a mecanismos para garantia de confidencialidade dos dados.
- II. Os provedores de nuvem pública usam soluções de Gerenciamento de Acesso e Identidade (IAM) para lidar com ataques que buscam violar confidencialidade dos dados e autenticidade.
- III. Como os provedores disponibilizam imagens fortalecidas de máquinas virtuais, os clientes que usam os serviços em Nuvem não precisam se preocupar com configurações de segurança.

Está correto o que se afirma em:

- a. I, apenas.
- b. II, apenas.
- c. III, apenas.
- d. I e II, apenas.
- e. II e III, apenas.

3. Uma estratégia de gerenciamento de riscos é fundamental para aprimorar o grau de segurança e a confiabilidade de aplicações em nuvem. Em geral, esse gerenciamento envolve pelo menos três etapas: avaliação, tratamento e controle de riscos. Uma empresa cliente de um provedor de nuvem pública precisa, antes de migrar suas aplicações, definir sua estratégia de gerenciamento de riscos.

Sobre gerenciamento de riscos em nuvem, é correto o que se afirma em:

- a. O gerenciamento de riscos é responsabilidade do provedor de nuvem pública.
- b. Na etapa de avaliação, os mecanismos de segurança utilizados são avaliados quanto a sua eficiência em proteger os dados na nuvem.
- c. Na etapa de tratamento, são executadas as ações e políticas de segurança.
- d. Na etapa de controle, são configurados os mecanismos de controle de acesso dos usuários aos serviços em nuvem.
- e. As três etapas são executadas no momento de implantação de uma aplicação no provedor.

Referências

ASW. **Centro de Arquitetura da AWS**. [S.l, s.d.]. Disponível em: <https://aws.amazon.com/pt/architecture>. Acesso em: 15 jul. 2019.

AWS. **Arquiteturas Multicamadas Sem Servidor da AWS**. [S.l.] 2015. Disponível em: https://d0.awsstatic.com/whitepapers/International/pt/AWS_Serverless_Multi-Tier_Architectures_PT. Acesso em: 30 jul. 2019.

AWS. **AWS Shield**: Proteção gerenciada contra DDoS. [S.l., s.d.]. Disponível em: <https://aws.amazon.com/pt/shield/>. Acesso em: 21 jul. 2019.

AWS. **Infraestrutura global**. [S.l., s.d.]. Disponível em: <https://aws.amazon.com/pt/about-aws/global-infrastructure/>. Acesso em: 15 jul. 2019.

AZURE. **Centro de Arquitetura do Azure**. [S.l., s.d.]. Disponível em: <https://docs.microsoft.com/pt-br/azure/architecture/>. Acesso em: 15 jul. 2019.

BALDINI, I. *et al.* Serverless computing: Current trends and open problems. *In: Research Advances in Cloud Computing*. [s.l.], Springer, 2017, p. 1–20.

CHEN, Y.; FARLEY, T.; YE, N. QoS requirements of network applications on the Internet. **Information Knowledge Systems Management**, v. 4, n. 1, p. 55-76, 2004.

COHEN, B. Incentives build robustness in BitTorrent. *In: Proceedings of the Workshop on Economics of Peer-to-Peer systems*. [S.l.], 2003, p. 68-72.

COULOURIS, G. *et al.* **Sistemas Distribuídos**: Conceitos e Projeto. 5 ed. Porto Alegre: Bookman Editora, 2013.

COUTINHO, A. A.; CARNEIRO, E.; GREVE, F. Computação em Névoa: Conceitos, Aplicações e Desafios. *In: Minicursos do XXXIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. Cap. 6, p. 266-315. Porto Alegre: SBC, 2016. Disponível em: https://www.researchgate.net/profile/Antonio_Augusto_Coutinho/publication/309312669_Computacao_em_Neova_Conceitos_Aplicacoes_e_Desafios/links/5809143f08ae040813483c45/Computacao-em-Neova-Conceitos-Aplicacoes-e-Desafios.pdf. Acesso em: 9 de jul. 2019.

CSA. **Guia de segurança para áreas críticas focado em Computação em Nuvem V3.0**. Cloud Security Alliance (CSA), 2011. Disponível em: <https://chapters.cloudsecurityalliance.org/brazil/files/2017/02/Guia-CSA-v-3.0.1-PT-BR-Final.pdf>. Acesso em: 3 ago. 2019.

DRAGONI, N. *et al.* Microservices: Yesterday, Today, and Tomorrow. *In: Present and Ulterior Software Engineering*, p. 195–216. Springer International Publishing, 2017.

ERL, T.; PUTTINI, R.; MAHMOOD, Z. **Cloud Computing: Concepts, Technology & Architecture**. [S.l.] Prentice Hall, 2013.

- GONZALEZ, N. M. *et al.* Segurança das Nuvens Computacionais: Uma Visão dos Principais Problemas e Soluções. **Revista USP**, São Paulo, v. 97, p. 27-42, 2013.
- GOOGLE. **Survey Report: Behind the Growing Confidence in Cloud Security**. MIT SMR Custom Studio, 2017. Disponível em: <https://lp.google-mkto.com/rs/248-TPC-286/images/SMR%20Google%20Cloud%20Security%20Survey%20Report,%20Final%20Approved%20Version,%209.8.17.pdf>. Acesso em: 19 ago. 2019.
- KAMIENSKI, C.; SADOK, D. Qualidade de Serviço na Internet. In: **Minicursos do XVIII Simpósio Brasileiro de Redes de Computadores (SBRC 2000)**. p. 1-40. [S.l.]: Sociedade Brasileira de Computação, 2000. Disponível em: <https://www.researchgate.net/publication/242087340>. Acesso em: 14 jul. 2019.
- KANSAL, N. J.; CHANA, I. Cloud load balancing techniques: A step towards green computing. **IJCSI International Journal of Computer Science Issues**, v. 9, n. 1, p. 238-246, 2012.
- KUROSE, J. F.; ROSS, K. W. **Redes de Computadores e a Internet: uma abordagem top-down**. 6. ed. São Paulo: Pearson Education do Brasil, 2013.
- LEHMANN, M. B. *et al.* Swarming: como BitTorrent revolucionou a Internet. In: **XXXI Jornadas de Atualização em Informática**, cap. 6, p. 209-258. Porto Alegre: SBC, 2012.
- MAO, M.; HUMPHREY, M. Auto-scaling to minimize cost and meet application deadlines in cloud workflows. In: **Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis**. IEEE, 2011. p. 1-12.
- MELO, K. Segurança na Nuvem: por onde começar? ComputerWorld, publicado em 29 fev. 2016. Disponível em: <https://computerworld.com.br/2016/02/29/seguranca-na-nuvem-por-onr-de-comecar/>. Acesso em: 21 jul. 2019.
- MICROSOFT AZURE. **Infraestrutura global do Azure**. [S.l., s.d.] b. Disponível em: <https://azure.microsoft.com/pt-br/global-infrastructure/>. Acesso em: 15 jul. 2019.
- MICROSOFT. **Microsoft Azure: SLA para Máquinas virtuais**. [S.l., s.d.]. Disponível em: https://azure.microsoft.com/pt-br/support/legal/sla/virtual-machines/v1_8/. Acesso em: 14 jul. 2019.
- MICROSOFT. **Proteção contra DDoS**: Proteja seus recursos do Azure contra ataques de DDoS (negação de serviço distribuído). [S.l., s.d.]. Disponível em: <https://azure.microsoft.com/pt-br/services/ddos-protection/>. Acesso em: 21 jul. 2019.
- NETFLIX. **Posso assistir aos títulos da Netflix em Ultra HD?** Centro de ajuda do Portal da Netflix, [s.d.]. Disponível em: <https://help.netflix.com/en/node/13444>. Acesso em: 14 jul. 2019.
- ROMAN, R.; LOPEZ, J.; MAMBO, M. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. **Future Generation Computer Systems**. V. 78, part. 2, p. 680-698, jan. 2018.

SHI, W. *et al.* Edge Computing: Vision and Challenges. **IEEE Internet of Things Journal**, v. 3, n. 5, p. 637–646, 2016.

TANENBAUM, A. S.; STEEN, M. V. **Sistemas Distribuídos: Princípios e Paradigmas**. 2. ed. São Paulo: Pearson Prentice Hall, 2008.

TUCKER, H. **Cloud computing: challenges and solutions**. [S.l., s.d.]. Disponível em: <https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/cloud-computing-challenges-and-solutions.html>. Acesso em: 18 jul. 2019.

ISBN 978-85-522-1590-5



9 788552 215905 >