



## **Infraestrutura de TI e Computação em Nuvem**

Rafael Kelles Vieira Lage

2020

## **Infraestrutura de TI e Computação em Nuvem**

Rafael Kelles Vieira Lage

© Copyright do Instituto de Gestão e Tecnologia da Informação.

Todos os direitos reservados.

## Sumário

---

Capítulo 1. Infraestrutura de TI .....	5
A definição de infraestrutura de TI .....	5
Modelo de infraestrutura de TI.....	6
Características da infraestrutura de TI .....	8
A importância da infraestrutura de TI .....	9
Desafios do gestor de TI.....	9
Capítulo 2. Computação em nuvem .....	11
Conceitos e características.....	11
Modelo de serviço .....	12
Modelos de entrega .....	14
Benefícios da computação em nuvem .....	16
Riscos da computação em nuvem .....	18
Terceirização de TI e computação em nuvem .....	20
Seleção de um provedor de computação em nuvem .....	22
Capítulo 3. Amazon Web Services .....	25
Sobre a AWS.....	25
Computação na AWS .....	27
Armazenamento.....	28
Network .....	29
Arquitetura de tolerância a falhas.....	32
Capítulo 4. Microsoft Azure .....	35
Sobre o Microsoft Azure .....	35
Computação no Azure .....	36
Banco de dados SQL no Azure .....	39
Armazenamento no Azure .....	40

Network no Azure.....	42
Arquitetura de tolerância a falhas.....	45
Capítulo 5. Cloud Privada .....	49
Aspectos gerais.....	49
Solução hiperconvergente .....	51
Referências.....	53

## Capítulo 1. Infraestrutura de TI

---

O que é infraestrutura de TI? A resposta para esta pergunta simples é bem ampla, com diferentes respostas por diferentes interlocutores. Um diretor comercial tem uma visão de infraestrutura de TI completamente diferente de um profissional de redes, que tem uma visão diferente de um analista de suporte do ambiente virtual, que também tem uma resposta diferente de um administrador de banco de dados.

Na verdade, a infraestrutura de TI permeia todos os elementos que dão sustentação às aplicações e, disto, sua relevância no ambiente de TI como um todo é fundamental, pois é a fundação de toda a “casa de TI” que se constrói em cima desta fundação (sistemas e aplicações).

Nesta disciplina vamos tratar destes elementos, não todos, mas os principais, apresentando suas características e analisando aspectos técnicos e de arquitetura para que um entendimento mínimo e rápido de cada elemento possa garantir um bom gerenciamento da infraestrutura de TI nas empresas.

Nesta unidade, estudaremos os elementos que compõem a infraestrutura de TI, sua relevância e a importância do alinhamento das iniciativas de TI com o negócio.

### A definição de infraestrutura de TI

---

Segundo Manoel Veras, em seu livro sobre Data Center, a infraestrutura de TI consiste nas instalações físicas, serviços e gestão que dão suporte a todos os recursos de computação compartilhados na organização. Dessa forma, Veras destaca que a infraestrutura de TI representa os recursos computacionais que são usados por toda a organização, de forma compartilhada.

Seguindo uma mesma linha de entendimento, o site cio.gov afirma que infraestrutura de TI consiste nos equipamentos, sistemas, software e serviços usados em comum por toda a organização, independentemente da missão, programa ou projeto.

Já o ITIL considera infraestrutura de TI como todo o hardware, software, redes e instalações físicas que são necessários para desenvolver, testar, entregar, monitorar, controlar ou suportar os serviços de TI. Nesta definição, a caracterização de suporte aos serviços de TI aparece de forma contundente.

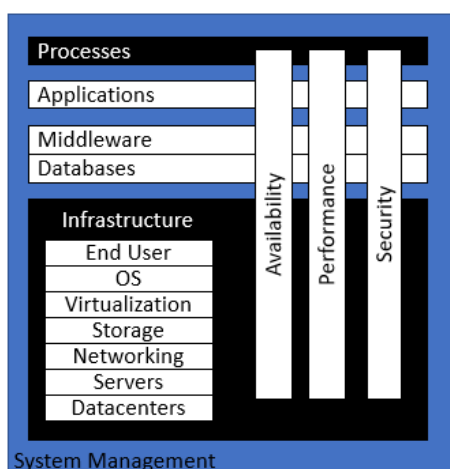
De fato, independentemente da definição, a infraestrutura de TI existe para dar suporte aos processos da empresa, como elemento de sustentação e suporte às aplicações. É a infraestrutura básica de TI que oferece recursos computacionais para hospedar e processar os serviços de TI. Esses sim, são vistos e usados de forma sistemática pelos usuários de toda a organização.

## Modelo de infraestrutura de TI

Para melhor explorar os aspectos mais importantes da infraestrutura de TI, vamos utilizar o modelo apresentado por Sjaak Laan em seu livro Infrastructure Architecture – Infrastructure building blocks and concepts, reforçando os aspectos relacionados aos elementos e sua relação com outras áreas de TI e da própria empresa.

A figura abaixo apresenta o modelo de Laan.

**Figura 1 – Modelo de infraestrutura de TI.**



Pelo modelo de Laan, observamos que os principais elementos de infraestrutura de TI são:

- **Data centers:** compreendendo a “casa de máquinas” com seus principais sistemas de energia, ar condicionado e proteção contra incêndio.
- **Servidores:** a camada computacional da infraestrutura de TI.
- **Rede:** a camada de conectividade.
- **Storage:** a camada de armazenamento.
- **Virtualização:** a camada que hoje oferece flexibilidade e otimiza os recursos de computação, rede e armazenamento.
- **Sistema operacional:** a camada de hospeda os serviços de TI (banco de dados, middleware e aplicações).
- **Usuário final:** toda a computação que ocorre fora do ambiente de data center.

É importante notar que o modelo de Laan mostra, visualmente, que a infraestrutura de TI dá sustentação aos processos de negócio que são, em última instância, o objetivo da tecnologia da informação. Por isso, os processos de negócio estão representados no topo do modelo.

Para sustentar os processos de negócios existem as aplicações e sistemas, que são os elos fortes entre o negócio e a TI, já que são as aplicações que oferecem serviços de TI para o negócio.

Mais abaixo no modelo está o middleware, entendido como os servidores de aplicação (como exemplo, Tomcat, JBOSS, IBM WebSphere, Oracle WebLogic, entre outros), e o banco de dados, onde as transações de negócio são armazenadas. Nestes dois elementos residem a lógica e os dados das aplicações, e são sustentados pela infraestrutura de TI.

A vantagem do modelo de Laan é mostrar, claramente, que a infraestrutura de TI provê recursos de TI e é tipicamente compartilhada por todas as aplicações, como veremos quando falarmos das características da infraestrutura.

## Características da infraestrutura de TI

---

As principais características da infraestrutura de TI são:

- **Provê recursos para as aplicações.**

As aplicações são efetivamente os serviços de TI e que oferecem produtividade e consistência no dia a dia das empresas. Não haveria infraestrutura de TI se não houvessem aplicações e a demanda por novas aplicações, que criam tecnologias de infraestrutura de TI. Computadores evoluem em processamento porque as aplicações estão mais complexas e elaboradas. Da mesma forma, a capacidade de armazenamento cresce exponencialmente porque as aplicações hoje são capazes de tratar grande volume de informações.

Da mesma forma, quando há uma falha em algum elemento de rede ou servidor, as aplicações são impactadas gerando impacto ao negócio.

Todas estas situações deixam claro que a principal característica da infraestrutura de TI é prover recursos para as aplicações.

- **Geralmente compartilhada por várias aplicações.**

- **Mais estática que as aplicações.**

O dinamismo do negócio reflete no dinamismo das aplicações que dão suporte tecnológico as atividades empresariais. A infraestrutura de TI, por outro lado, como sustentação às aplicações, tem um comportamento mais estático já que, em tese, ele permanece estática enquanto a aplicação estiver servindo as necessidades empresariais, sem crescimento de usuários.

De fato, especialmente no modelo tradicional de infraestrutura local, dentro da empresa, esta característica estática da infraestrutura de TI é bem observada com o ciclo de vida de 3 a 8 anos, dos principais elementos de infraestrutura (como PCs, servidores, storage ou rede).

Importante notar que mais estática não significa permanente, pois estes equipamentos têm vida útil e, depois de um tempo, mesmo que a aplicação não



demande mais recursos, é necessário a troca para reduzir o risco de falha e indisponibilidade para as aplicações.

- **Infraestrutura e aplicações são departamentos diferentes na empresa.**

Infraestrutura de TI e aplicações têm características bem distintas e por isso mesmo são departamentos diferentes na empresa, já que são tratativas e abordagens únicas. Estando em departamentos distintos, mas que tem uma relação de proximidade, o importante é buscar o entendimento comum quanto aos objetivos e, neste momento, os aspectos de disponibilidade, segurança e desempenho são a “linguagem comum” para que haja entendimento mútuo.

### A importância da infraestrutura de TI

---

Embora a infraestrutura de TI esteja presente e percebida por todos os usuários como fundamental para as atividades da empresa, a realidade na hora de priorizar os investimentos de TI nem sempre segue a mesma lógica.

O modelo de Laan ajuda a explicar esta diferença, já que a infraestrutura de TI está distante, na maioria das vezes, do processo de negócio em si e, portanto, os investimentos devem ser justificados com maior ênfase. No entanto, para que a infraestrutura de TI seja vista como elemento de sustentação e geração de valor para o negócio, é necessário desenvolver iniciativas que procurem através da disponibilidade, desempenho e segurança, suportar novas iniciativas de negócio que gerem receita ou promovam a redução de custo.

### Desafios do gestor de TI

---

O grande desafio do gestor de TI é buscar o alinhamento entre as iniciativas de TI e o negócio.

A questão fundamental é que, para a área de negócio, o importante é o resultado e não os componentes que irão suportar o resultado. Do outro lado,

sabemos da complexidade que é lidar com certas iniciativas de negócio, pois qualquer nova aplicação requer uma série de recursos de TI (rede, servidores, armazenamento, virtualização, banco de dados e middleware). O importante é caracterizar a solução como elemento que provê recursos de TI para o negócio, evitando os aspectos de complexidade inerentes à solução.

Um grande exemplo desta questão é o sistema de iluminação de uma casa. Quando acionamos o interruptor, a luz simplesmente acende. Por trás, há todo um aparato tecnológico de energia que vai desde a geração, transmissão e distribuição de energia, envolvendo torres, subestações, postes, medição de consumo e infraestrutura de cabeamento para suportar o acionamento da luz através de um interruptor. A TI, como elemento agregador para o negócio, deve ter um papel parecido: eliminar a complexidade e oferecer resultados.

Para tanto, é preciso entender melhor as iniciativas de negócio e procurar oferecer alternativas usando uma linguagem comum: desempenho, disponibilidade e segurança, conforme propõe o modelo Laan.

Como um breve exercício de alinhamento, a tabela abaixo mostra os principais objetivos de iniciativas de negócio e os objetivos equivalentes de um gestor de infraestrutura para uma mesma iniciativa.

**Figura 2 – Alinhamento de negócio e TI.**

Visão de Negócio	Gestor de TI
Aumentar a receita	Flexibilidade dos serviços de TI
Reduzir as despesas	Otimizar os serviços de TI
Eficiência Operacional	Qualidade dos serviços de TI

Como exemplo, em iniciativas de aumento de receita uma boa estratégia para o gerenciamento de infraestrutura de TI é oferecer uma plataforma, que ofereça flexibilidade para os novos serviços de TI. Neste sentido, os serviços de computação em nuvem são interessantes para oferecer esta flexibilidade.

## Capítulo 2. Computação em nuvem

---

Não há qualquer dúvida que a evolução da capacidade de processamento do hardware, as iniciativas de virtualização e o aumento da oferta de serviços de telecomunicações são a base para a computação em nuvem (Cloud Computing). Grandes provedores de datacenter oferecem um ambiente de infraestrutura de TI seguro e confiável para suportar os negócios.

Este capítulo apresenta diferentes aspectos da computação em nuvem como conceito, modelos de serviço, modelos de entrega, benefícios, riscos e seleção do provedor de computação em nuvem.

### Conceitos e características

---

A definição do NIST (National Institute of Standards and Technology) afirma que Cloud computing é um modelo para permitir acesso via rede sob demanda a um conjunto configurável de recursos computacionais (exemplo: redes, servidores, storage, aplicações e serviços), que podem ser rapidamente provisionados e liberados com um esforço mínimo de gerenciamento ou interação, com o provedor de serviços. O NIST também define cinco características essenciais, três modelos de serviço e quatro modelos de entrega.

- **Autosserviço sob demanda:** um consumidor pode unilateralmente prover capacidade computacional, como tamanho do servidor e espaço de armazenamento em rede, de acordo com sua necessidade, sem requerer intervenção humana com o provedor de serviços.
- **Amplo acesso à rede:** os recursos ficam disponíveis através da rede e são acessados através de mecanismos padrão que permitem o seu uso por plataformas heterogêneas dos clientes.
- **Pooling de recursos:** os recursos computacionais do provedor existem para servir vários consumidores usando um modelo multi-tenant, com recursos físicos ou virtuais diferentes e dinamicamente atribuídos de acordo com a

demanda do consumidor. Como há uma visão da independência de localização, em que o cliente não tem qualquer controle ou conhecimento de onde os serviços contratados estão localizados, o cliente define apenas a maior região onde o serviço será alocado.

- **Elasticidade:** os recursos podem ser elasticamente provisionados e liberados, em alguns casos automaticamente, para mais ou para menos, de acordo com a demanda. Para o cliente, os recursos disponíveis para provisionamento frequentemente parecem ser ilimitados e podem ser utilizados em qualquer quantidade a qualquer tempo.
- **Medição do consumo:** os sistemas de nuvem automaticamente controlam e otimizam o uso de recursos, em função da capacidade de medição em algum nível de abstração apropriada ao tipo do serviço. O uso dos recursos deve ser monitorado, controlado e reportado, garantindo transparência para o provedor e o consumidor do serviço utilizado.

### Modelo de serviço

---

O modelo de serviços é o conjunto de serviços de TI oferecidos por um provedor, que é o responsável por fornecer recursos para um consumidor dentro de um SLA predefinido e mutuamente acordado. Pode-se dizer que o provedor de serviços de TI cuida disso, enquanto o consumidor foca nos aspectos relevantes ao suporte efetivo às áreas de negócio da empresa.

**Software as a Service (SaaS):** o recurso oferecido ao cliente é poder usar a aplicação do provedor rodando em uma infraestrutura em nuvem. As aplicações são acessíveis a partir de vários dispositivos clientes, tipicamente via web. O cliente consumidor não gerencia ou controla a infraestrutura responsável pela disponibilização do recurso incluindo rede, servidores, sistemas operacionais, dispositivos de armazenamento ou até mesmo os recursos disponíveis para cada aplicação, com uma possibilidade de configuração de parâmetros da aplicação por

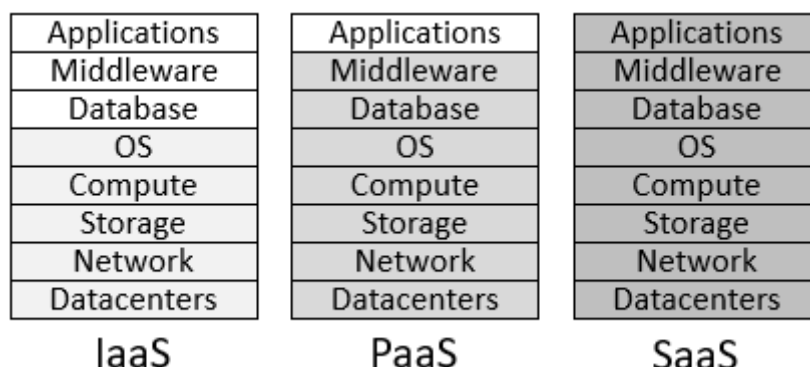
usuário. Ou seja, o serviço oferecido é a própria aplicação que suporta diretamente o processo de negócio.

**Platform as a Service (PaaS):** o recurso oferecido ao cliente consumidor é entregar dentro da infraestrutura da nuvem aplicações desenvolvidas internamente usando linguagens, bibliotecas, serviços e ferramentas suportadas pelo provedor de serviços. O cliente consumidor não gerencia ou controla a infraestrutura necessária para a aplicação, incluindo rede, servidores, sistemas operacionais ou dispositivos de armazenamento, mas tem controle sobre as aplicações entregues e possivelmente sobre os parâmetros de configuração do ambiente de hospedagem da aplicação. Este modelo de serviço oferece um ambiente de desenvolvimento, a partir do qual o cliente desenvolve sua aplicação. Neste sentido, a responsabilidade pela aplicação é do cliente, enquanto o ambiente onde a aplicação reside é responsabilidade do provedor de serviços de computação em nuvem.

**Infrastructure as a Service (IaaS):** nesse caso, o recurso oferecido ao cliente consumidor é prover capacidade de processamento, armazenamento e outros recursos sobre os quais é possível que o cliente rode qualquer aplicação, inclusive podendo incluir o sistema operacional. O cliente não controla a infraestrutura de nuvem que suporta os recursos oferecidos a ele, mas tem controle sobre os sistemas operacionais, plataforma de desenvolvimento e aplicações utilizadas, podendo chegar até ao controle de alguns elementos de rede.

Alguns modelos de serviço que estão sendo vistos no mercado merecem uma menção honrosa, eles acabam, de certa forma, sendo uma especialização dos modelos de serviço anteriores. São eles o CaaS (Communication as a service), o DBaaS (Database as a service), o BPaaS (Business Process as a Service). A figura 3 ilustra a diferença existente entre os três modelos de serviço de computação em nuvem.

**Figura 3 – Modelos de serviço de computação em nuvem.**



Ainda sobre o aspecto do valor para os clientes, pode-se estabelecer que:

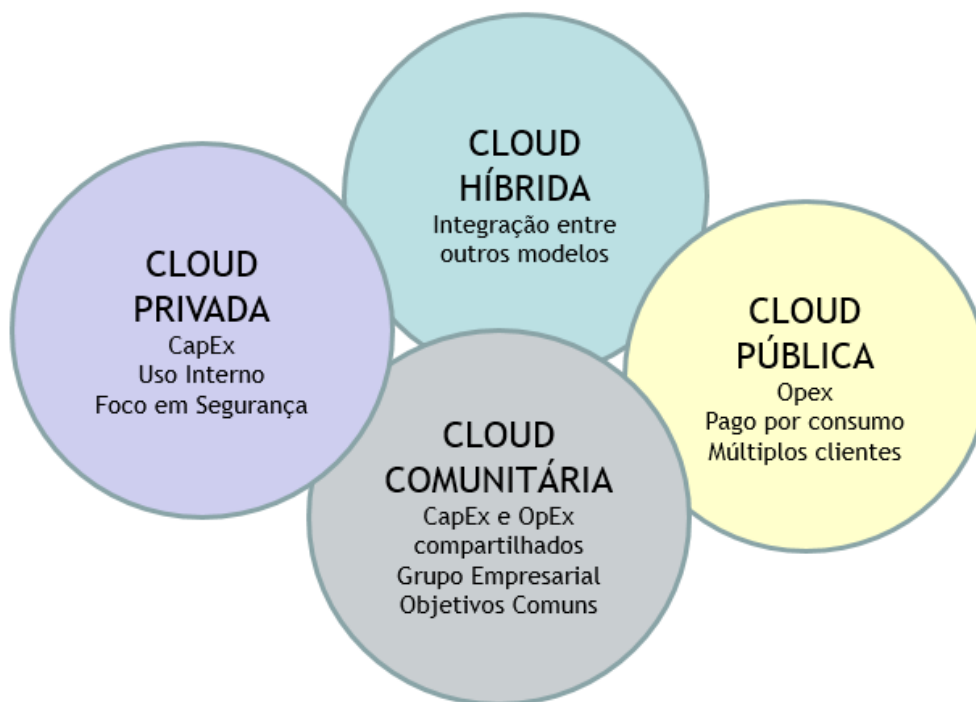
- **IaaS:** a responsabilidade pela disponibilidade da infraestrutura de TI é do provedor de serviços.
- **PaaS:** garante que a propriedade e operação do hardware, sistema operacional e plataforma de desenvolvimento (servidor de aplicação e banco de dados) é do provedor de serviços.
- **SaaS:** garante que a propriedade e operação do hardware, sistema operacional, middleware, banco de dados e aplicativo é do provedor de serviços. Neste caso, toda a gestão de TI no contexto da aplicação específica é responsabilidade do provedor.

Sendo assim, em uma análise superficial pode-se dizer que o modelo SaaS traz mais valor para o cliente, já que permite ao cliente focar diretamente nos processos de negócio.

## Modelos de entrega

Os modelos de entrega estão destacados na figura abaixo e explicados na sequência.

**Figura 4 – Modelos de entrega de computação em nuvem.**



**Private cloud:** a infraestrutura da nuvem é dedicada ao uso exclusivo de uma única organização, que pode atender sua matriz, filiais, departamentos e subsidiárias. Ela pode ser de propriedade, gerenciada e operada pela organização, por terceiros ou alguma combinação dos dois, e pode estar dentro ou fora das instalações da organização.

**Community cloud:** é a infraestrutura de nuvem dedicada para uso exclusivo de uma comunidade específica de consumidores de organizações que possuem preocupações compartilhadas (por exemplo: missão, requisitos de segurança, políticas e aspectos de conformidade). Ela pode ser de propriedade, gerenciada e operacionalizada por uma ou mais organizações que sejam parte da comunidade, por terceiros ou uma combinação dos dois e pode estar dentro ou fora das instalações. É um modelo que permite que empresas do mesmo grupo, por exemplo, utilizem a mesma infraestrutura de nuvem. Um exemplo típico é um grupo empresarial, onde se cria uma infraestrutura comunitária para suportar as atividades comuns da empresa, como um centro de serviços de TI compartilhado.

**Public cloud:** a infraestrutura da nuvem é oferecida de forma aberta e para o público em geral. Ela pode ser de propriedade, gerenciada e operacionalizada por

uma empresa, por uma escola ou organização governamental, ou uma combinação destes. Ela existe dentro das instalações do provedor da nuvem. Uma das suas vantagens é a redução do investimento inicial necessário para que uma empresa inicie suas operações.

**Hybrid cloud:** a infraestrutura da nuvem é uma composição de um ou mais dos modos anteriores que permanecem como entidades únicas, mas estão unidas por tecnologia padronizada ou proprietária que permite a portabilidade dos dados e das aplicações. Para que seja caracterizada como nuvem híbrida, deve haver a troca de informações entre a nuvem privada ou datacenter interno da empresa e a nuvem pública.

### Benefícios da computação em nuvem

---

Quando se pensa em computação em nuvem, o primeiro benefício que surge é o novo modelo financeiro de pagamento pelo uso. Contudo, este é apenas um dos benefícios do uso da computação em nuvem e, em muitos casos, ele nem sempre se materializa de fato.

Alguns benefícios do uso da computação em nuvem são:

**Acesso aos recursos:** um dos grandes benefícios da computação em nuvem é o acesso aos recursos de infraestrutura de TI (processamento, armazenamento e rede) disponíveis através de acesso remoto, permitindo empresas de todo porte ter acesso aos recursos (percebidos como ilimitados) que não seria possível em uma estrutura interna e local, por uma fração do custo da estrutura própria.

**Mobilidade:** as empresas podem acessar os serviços de TI de qualquer lugar no mundo, já que são baseados em serviços web. Mais ainda, com o crescimento dos serviços móveis, a mobilidade dos usuários e dos processos de negócios ganhou nova dimensão e a infraestrutura de computação em nuvem dá suporte nativo a esta tendência, promovendo o acesso a qualquer hora (anytime), de qualquer lugar (anywhere) e usando qualquer tipo de dispositivo (any device).



**Arquitetura escalável:** empresas com processos de negócio com características sazonais ou novas oportunidades que crescem em larga escala, podem se beneficiar de uma plataforma escalável simplesmente requisitando mais recursos (capacidade ou armazenamento) através do pagamento mensal ou pelo uso.

**Segurança de dados:** segurança é um ponto especialmente importante nas organizações, pois uma falha pode resultar em perdas financeiras ou de reputação. Para a maioria das organizações, a segurança de dados oferecida pelos provedores de computação em nuvem é bem superior ao que se pode oferecer internamente, a um custo compatível. Isto se dá pelo ganho de escala dos provedores de computação em nuvem e o nível de especialização exigido para a implementação de práticas de segurança mais sofisticadas nas empresas. Para os provedores de computação em nuvem, esta especialização e o investimento nesta área é um requisito para o negócio, ou seja, uma obrigação.

**Baixo custo:** o modelo financeiro, como já mencionamos, é atrativo para a maioria das organizações, pois ele é baseado em pagamentos mensais (seja pelo uso dos recursos ou pela quantidade de ativos). Ao contrário dos investimentos em infraestrutura de TI própria, onde o pagamento é antecipado tanto para os equipamentos adquiridos quanto para as licenças de software, que são baseadas em métricas confusas e difíceis de manter a governança e compliance.

O pagamento mensal, normalmente, inclui o suporte e a manutenção, licenciamento de software, além de várias garantias operacionais relacionadas aos datacenters em si, segurança e elevados níveis de disponibilidade da infraestrutura física onde o serviço contratado reside.

O pagamento por uso é ainda mais atraente para organizações com picos de demanda, pois é mais barato pagar pelo uso nos momentos onde a demanda existe ao invés de adquirir soluções para suportar o pico de utilização e permanecer subutilizado durante o resto do ano.

**Manutenção e suporte:** o provedor de computação em nuvem, normalmente, oferece serviços de suporte. Entretanto, serviços remotos torna o processo mais simples e menos intrusivo nas operações das empresas, já que o provedor oferece

backup, atualização e upgrades da infraestrutura de TI sem afetar os serviços de TI. Isto torna a operação de TI mais simples e rápida, tornando a computação baseada em ideias e inovação para suportar os processos de negócio das empresas.

Além disso, a infraestrutura de TI oferecida pelos provedores parece ilimitada e atualizada constantemente, sem que haja investimentos ou perda de garantia de equipamentos, como tipicamente ocorre em empresas cuja infraestrutura de TI está obsoleta pois foi adquirida há mais de 8 anos.

**Ambiente sustentável:** embora seja um tópico polêmico, os provedores de computação em nuvem operam com servidores com capacidade acima de 60% de utilização, diferentemente do que ocorre em infraestrutura própria, quando os servidores operam abaixo dos 20% de utilização e consomem, praticamente, a mesma energia. Além disso, alguns provedores de computação em nuvem estão optando por localidades com uso de energia renovável (solar ou eólica), reduzindo custos de energia ao mesmo tempo que potencializa a ideia da sustentabilidade.

**Testes gratuitos:** a maioria dos provedores oferece a oportunidade de testar os serviços de TI por um período gratuitamente. Isto torna a adoção e o entendimento da solução mais fácil, enquanto permite aos provedores a oportunidade de conquistar mais um cliente.

### Riscos da computação em nuvem

---

Em se tratando de risco da computação em nuvem, fica evidente o risco de perda do controle interno já que a infraestrutura de TI não é mais propriedade da empresa. Ou seja, há um novo ator no processo de prover serviços de TI e este ator é um agente externo. O que fazer se algo der errado e tivermos um problema no provedor de serviços de TI? De fato, este risco existe, mas precisamos materializar melhor esta possibilidade em ameaças mais concretas e sair da subjetividade da questão colocada. Quando saímos da subjetividade e analisamos as ameaças e os impactos, percebemos, muitas vezes, que o possível impacto é o mesmo de uma infraestrutura própria e, na maioria dos casos, a probabilidade de ocorrência da

ameaça é menor na computação em nuvem, pelo expertise e nível de redundância da infraestrutura de TI contratada.

De uma forma geral, os riscos da computação em nuvem são:

- **Confiabilidade da internet.**

Uma vez que os serviços de TI são, na computação em nuvem, oferecidos com acesso via internet, a confiabilidade dos links é um fator crítico para o sucesso, especialmente em aplicações de missão crítica como e-mail e processos de negócio. Assim, é necessário mitigar este risco com a contratação de links de melhor qualidade e redundantes, diminuindo a possibilidade de ter os serviços de TI interrompidos em função de problemas de desempenho ou disponibilidade do acesso. É importante lembrar que serviços de TI internos também estão expostos às falhas de conectividade pelos mesmos motivos (por exemplo, filiais dispersas) ou por outras dificuldades próprias de uma solução on-premise (problemas de capacidade, obsolescência de equipamentos, falta de suporte, problemas de energia e ar condicionado nos datacenters próprios) que um provedor de computação em nuvem tipicamente gerencia melhor.

- **Dependência do provedor externo.**

Com o uso da computação em nuvem, os serviços de TI dependem de um agente externo, mais do que em uma solução de datacenter próprio (cuja dependência está na manutenção e suporte da infraestrutura de TI em relação aos fabricantes). A preocupação maior é a saúde financeira do provedor, possíveis problemas legais e a existência ou não de um subcontratado pelo provedor, que pode vir a afetar os serviços de TI, especialmente os mais críticos.

Contudo, a dependência de um terceiro no ambiente de TI não é uma novidade da computação em nuvem. Trata-se de algo comum hoje, seja no contexto do desenvolvimento de aplicações, consultoria para implementação de projetos e até mesmo no suporte à infraestrutura de TI. Desta forma, metodologias para identificação dos riscos e elaboração de tratativas para eliminá-las ou reduzi-las, são práticas de mercado bastante conhecidas que podem ser adotadas pelas empresas

para processos envolvendo a adoção de serviços de computação em nuvem. Por exemplo, diligências podem identificar problemas financeiros, além de referências de clientes podem mostrar o nível de confiabilidade de outras empresas neste provedor.

### Terceirização de TI e computação em nuvem

---

Computação em nuvem pode ser vista como uma forma de terceirização de TI, já que significa delegar a responsabilidade de parte da TI (tipicamente a infraestrutura de TI) para um terceiro, caracterizando assim o modelo de terceirização de TI. Desta forma, vamos aqui nos apoiar em alguns aspectos da terceirização de TI para avaliar e entender melhor as alternativas de computação em nuvem.

- **Quanto ao número de provedores:** em terceirização de TI podemos trabalhar com provedor único, conjunto de provedores e consórcio de provedores. Trabalhar com provedor único estabelece uma relação estratégica com o provedor, mas pode limitar as alternativas e as soluções disponíveis. Em termos de computação em nuvem, as ofertas mais interessantes são baseadas em SaaS e, neste cenário, é comum termos mais de um provedor em uma empresa. Por exemplo, o serviço de e-mail com a Microsoft (Office 365), o serviço de ERP com a Totvs e o serviço de CRM com a Salesforce. Percebe-se, claramente, que o cenário baseado em um conjunto de provedores se mostra comum nas empresas, embora crie outras dificuldades, especialmente de integração entre os serviços e aspectos de segurança em ambiente multicloud.
- **Quanto a estratégia de implementação:** novamente são três as alternativas – total, parcial ou incremental. É consenso entre os analistas de mercado e consultorias que a adoção de cloud computing é, na maioria dos casos, uma jornada. Portanto, as estratégias de implementação mais usadas são a implementação parcial, criando o que chamamos de cloud híbrida, ou a implementação incremental, fortalecendo o conceito de jornada. Do ponto de vista prático, já existem empresas maduras e consolidadas em termos de oferta e estudos de casos. Seguindo a regra geral e pensando em implementações

parciais, o uso de correio eletrônico na nuvem ou backup na nuvem são implementações parciais seguras e de baixo risco, pois já foram realizadas por inúmeras empresas com sucesso. Em termos de implementação incremental, poderia mencionar a estratégia Microsoft: começa com o Office 365, evolui para o Exchange e finaliza com a adoção do SharePoint na nuvem.

- **Quanto ao relacionamento:** parcerias estratégicas, co-sourcing e transação. Em termos de computação em nuvem, todas as alternativas têm estudos de caso bem conhecidos e divulgados. Como relação estratégica podemos citar a Netflix com a AWS, onde a Netflix desenvolveu um modelo de negócio novo, apoiando-se no potencial de escalabilidade e confiabilidade da AWS para se tornar um provedor global de filmes e séries de televisão via streaming. O modelo mais comum é o transacional, onde se usa a computação em nuvem para projetos específicos de desenvolvimento de novas aplicações, por exemplo, usando big data, ferramentas analíticas, marketing digital ou mobilidade de forma pontual. O co-sourcing é quando cliente e provedor dividem a responsabilidade em um projeto, é menos explorado, mas possível em provedores de escala local ou regional.

Finalmente, os benefícios da terceirização de TI também se estendem para a computação em nuvem. A seguir destacamos alguns:

- **Reduzir custos operacionais:** em termos de infraestrutura de TI, os custos de uma infraestrutura segura e confiável representa um custo alto para a maioria das empresas. A computação em nuvem oferece uma oportunidade de redução dos custos operacionais de infraestrutura, ao mesmo tempo que simplifica o seu uso, já que ela é compartilhada por vários clientes.
- **Compartilhar riscos:** com o uso da computação em nuvem, aspectos de disponibilidade, segurança e desempenho dos serviços de TI são responsabilidades compartilhadas com os provedores, que são empresas especializadas em tecnologia.
- **Dar foco nas competências centrais:** a área de TI permeia quase todos os departamentos de uma empresa, mas não é, na maioria das indústrias, uma

competência central para o negócio. A tecnologia da informação suporta os negócios, mas é considerado uma área de apoio. Desta forma, usar computação em nuvem significa voltar a TI para novas oportunidades e aplicações de suporte ao negócio, delegando a infraestrutura de TI para uma empresa especializada.

- **Ganhar acesso as melhores práticas da indústria:** um provedor de computação em nuvem é uma empresa especializada em computação, armazenamento e segurança que, para oferecer os melhores serviços, faz uso das melhores práticas da indústria de TI como ferramenta de competitividade.

### Seleção de um provedor de computação em nuvem

---

Quando decide-se fazer a mudança para computação em nuvem, a próxima etapa será selecionar um provedor de serviços de nuvem. É essencial avaliar a confiabilidade e a capacidade de um provedor de serviços no qual se planeja confiar os aplicativos e os dados de sua organização. A seguir, destacamos alguns pontos importantes na seleção de um provedor de computação em nuvem.

**Saúde financeira:** o provedor deve manter um registro da estabilidade e ter uma posição financeira íntegra, com capital suficiente para operar com êxito a longo prazo.

**Organização, governança, planejamento e gerenciamento de riscos:** o provedor deve ter uma estrutura de gerenciamento formal, políticas de gerenciamento de riscos estabelecidas e um processo formal de gerenciamento de riscos e mudanças no ambiente de infraestrutura de TI.

**Confiança:** deve-se gostar da empresa e de seus princípios. Verifique a reputação do provedor e veja quais são seus parceiros. Descubra o nível de experiência dele com a nuvem. Leia avaliações e fale com clientes cuja situação é similar à sua.

**Conhecimento de negócios e técnico:** o provedor deve compreender seus negócios e seus objetivos e conseguir relacionar todas essas informações com o conhecimento técnico dele.

**Auditoria de conformidade:** o provedor deve conseguir validar a conformidade com todos os seus requisitos por meio de uma auditoria de terceiros.

**Contratos de Nível de Serviço (SLAs):** os provedores devem conseguir prometer um nível básico de serviço com o qual você esteja acostumado.

**Relatório de desempenho:** o provedor deve fornecer relatórios de desempenho.

**Monitoramento de recursos e gerenciamento de configurações:** deve haver controles suficientes para o provedor acompanhar e monitorar serviços fornecidos aos clientes e alterações feitas aos seus sistemas.

**Cobrança e contabilização:** este processo deve ser automatizado para que você possa monitorar quais recursos está usando e seus custos, para que você não acumule contas inesperadas. Também deve haver suporte a problemas relacionados a cobrança.

**Facilidade de implantação, gerenciamento e atualização:** certifique-se de que o provedor tenha mecanismos que facilitem a implantação, o gerenciamento e a atualização de seu software e seus aplicativos.

**Gerenciamento de eventos:** o provedor deve ter um sistema formal para gerenciamento de eventos, que seja integrado ao seu sistema de monitoramento/gerenciamento.

**Gerenciamento de alterações:** o provedor deve ter documentos e processos formais para solicitar, registrar em log, aprovar, testar e aceitar mudanças.

**Funcionalidades híbridas:** mesmo se você não planeja usar uma nuvem híbrida de início, deve ter certeza de que o provedor pode dar suporte a este modelo. Ele tem vantagens que você pode querer explorar posteriormente.



**Infraestrutura de segurança:** deve haver uma infraestrutura de segurança abrangente para todos os níveis e tipos de serviços de nuvem.

**Políticas de segurança:** deve haver políticas e procedimentos de segurança abrangentes em vigor para controle de acesso aos sistemas do cliente e do provedor.

**Gerenciamento de identidade:** mudanças a qualquer serviço de aplicativo ou componente de hardware devem ser autorizadas em uma base de função de grupo ou pessoal, e a autenticação deve ser necessária para qualquer pessoa ao alterar um aplicativo ou dado.

**Backup e retenção de dados:** políticas e procedimentos para garantir a integridade dos dados do cliente devem estar em vigor e em operação.

**Segurança física:** controles garantindo segurança física devem estar em vigor, incluindo para acesso de contratos de colocation. Além disso, datacenters devem ter proteções ambientais para proteger equipamentos e dados contra eventos disruptivos. Deve haver redes e energia redundantes, além de um plano de continuidade de negócios e recuperação de desastre documentado.



## Capítulo 3. Amazon Web Services

---

Este capítulo apresenta as ofertas da AWS (Amazon Web Services) no que diz respeito a infraestrutura de TI.

### Sobre a AWS

---

Amazon Web Services (AWS) é, formalmente, uma subsidiária da Amazon e os serviços da AWS foram criados para dar suporte a infraestrutura de TI da própria Amazon, empresa global de comércio eletrônico que necessitava de uma infraestrutura de TI que tivesse alto desempenho e que fosse escalável, suportando os negócios da empresa.

A plataforma AWS foi lançada em 2002 e, depois, reformulada em 2003, formando a base dos serviços atuais. Já em 2003 a visão de uma nova oportunidade de negócios para a Amazon já era percebida, com a possibilidade de vender os serviços web padronizados e usados internamente como serviço.

Desta forma, a AWS lança os serviços ao mercado de TI em 2006 e se transformou em líder no mercado de computação em nuvem, especialmente na plataforma IaaS.

Com uma ampla oferta de solução, a AWS oferece serviços para soluções completas com base na nuvem para diversas necessidades de TI como sites, backup e recuperação, DevOps, Big Data, serviços financeiros, marketing digital, internet das coisas, jogos, entre outros.

Com uma plataforma padronizada de serviços web, a AWS tornou a computação em nuvem baseada em ideias e não em recursos computacionais, estimulando a inovação por meio de novas aplicações e criou uma forma de infraestrutura de TI, baseada no consumo de TI.

Os serviços AWS apresentam as seguintes características:

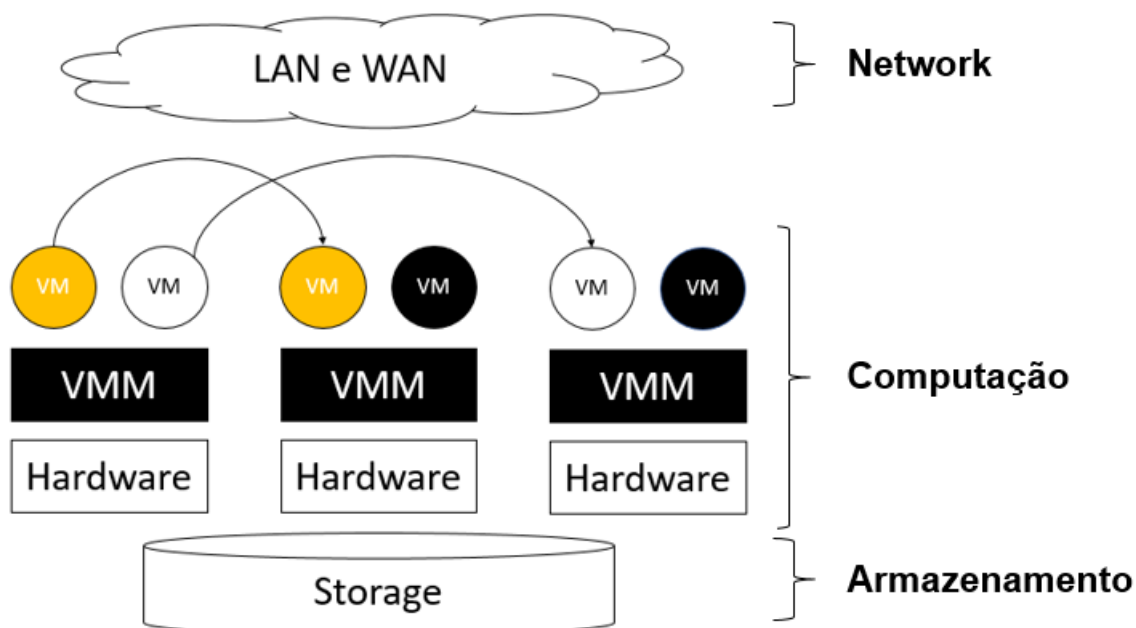
- **Confiabilidade:** foi projetada para tolerar falhas e corrigir os serviços rapidamente.
- **Simplicidade:** construída a partir de conceitos simples e rápida implementação, com grande flexibilidade para os desenvolvedores (consumidores de serviço).
- **Escalabilidade:** um projeto que permite a visão de escalabilidade ilimitada, com ampliação dos serviços web para atender picos de demanda.
- **Baixo custo:** valores competitivos frente a outras soluções de mercado, pessoais e corporativas.

Para suportar os negócios de empresas de todo o porte e setor, a infraestrutura global da AWS conta com certificações, declarações de conformidade, atendimento aos aspectos legais e reguladores, além de alinhamento de estruturas para suportar diferentes normas e padrões específicos de indústria. Com isto, a AWS apoia a conformidade de seus clientes com relação a adoção da nuvem AWS em seus processos de negócio.

Embora o portfólio de produtos seja extenso, como um primeiro passo para entendimento dos serviços AWS, especialmente aqueles voltados para infraestrutura de TI, vamos dividi-los em blocos de construção alinhados com o modelo de Laan, apresentado no capítulo 1.

A figura abaixo mostra os blocos de construção essenciais de uma infraestrutura de TI e que, a partir deles, analisaremos as ofertas básicas de infraestrutura de TI.

**Figura 5 – Blocos de construção de infraestrutura de TI.**



## Computação na AWS

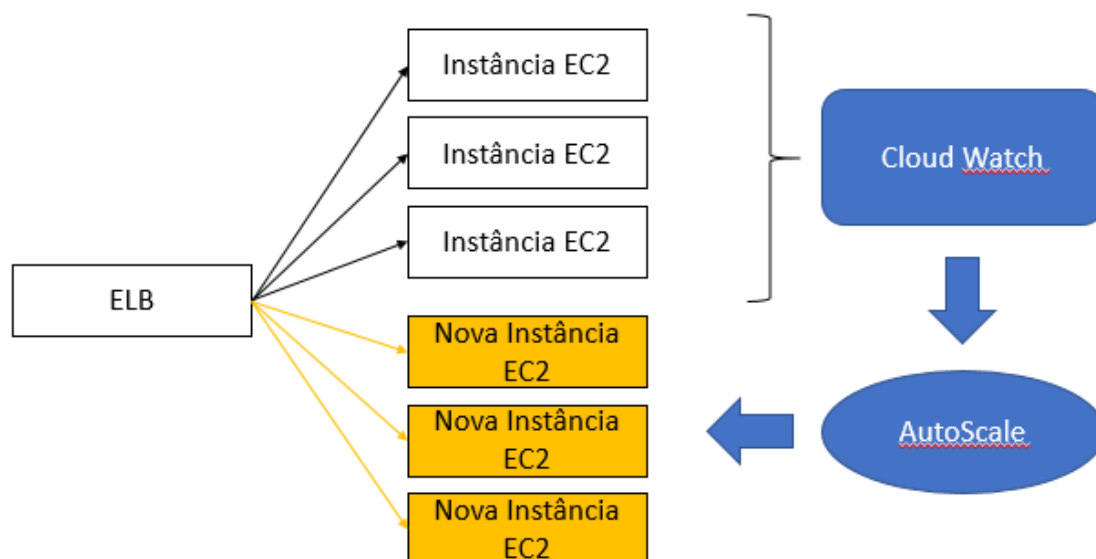
O serviço principal de computação da AWS é conhecido como Amazon EC2 (Amazon Elastic Compute Cloud). Trata-se de um serviço web que disponibiliza capacidade computacional segura e redimensionável na nuvem. Ele foi criado para facilitar para os desenvolvedores a computação em escala web.

Seu elemento principal é a instância EC2, onde se aloca recursos computacionais a partir do tipo da instância (baseado na quantidade de CPU e memória RAM). Em seguida, a instância é criada a partir de uma imagem de sistema operacional (Linux ou Windows), conforme um catálogo de opções de imagens públicas, pagas ou mesmo do próprio cliente.

Uma característica fundamental das instâncias EC2 é a capacidade de se adequar, automaticamente, a demanda de serviços web. Conforme mostra a figura 6, esta capacidade é baseada em uma série de elementos da AWS que atuam em conjunto suportando esta elasticidade. As instâncias EC2 são monitoradas por meio do serviço CloudWatch e, a partir de thresholds, são iniciadas as ações de

crescimento (aumento das instâncias EC2) ou de diminuição (término das instâncias EC2). Esta elasticidade das instâncias EC2 da AWS é conhecida como AutoScale.

**Figura 6 – A elasticidade das instâncias EC2 da AWS.**



## Armazenamento

O armazenamento na nuvem é um componente essencial da cloud computing, que mantêm as informações usadas pelos aplicativos. Análise de big data, data warehouses, Internet das Coisas, bancos de dados e aplicativos de backup e arquivamento, todos esses serviços dependem de algum tipo de arquitetura de armazenamento de dados. Geralmente, o armazenamento na nuvem é mais confiável, escalável e seguro do que sistemas de armazenamento locais tradicionais.

A AWS oferece um leque de ofertas de armazenamento na nuvem para apoiar requisitos de conformidade para aplicativos e arquivamento. Escolha entre os serviços de armazenamento de objetos, de bloco e em arquivos.

**Amazon S3 – Simple Storage Service:** um recurso persistente e escalável para disponibilizar dados por meio de qualquer local com acesso à internet, usado para conteúdo gerado pelo usuário como arquivos, documentos, áudio, vídeo, imagens e backup e recuperação.

**Amazon EBS – Elastic Block Service:** armazenamento local persistente para as instâncias Amazon EC2, como bancos de dados relacionais e NoSQL, data warehousing, aplicações empresariais e processamento de big data. Junto com esta tecnologia, podemos associar snapshots que facilitam a criação de instâncias EC2. Esta capacidade de criação de instâncias a partir de snapshot é uma característica importante para a elasticidade dos serviços AWS.

**Amazon Elastic File System:** interface de sistema de arquivos e semântica de acesso ao sistema de arquivos para disponibilizar dados para uma ou mais instâncias EC2, para distribuição de conteúdo.

**AWS Storage Gateway:** uma nuvem de armazenamento híbrido que amplia seu ambiente local com o armazenamento na nuvem da Amazon para bursting, divisão em camadas ou migração, ou mesmo repositório de backup na nuvem.

## Network

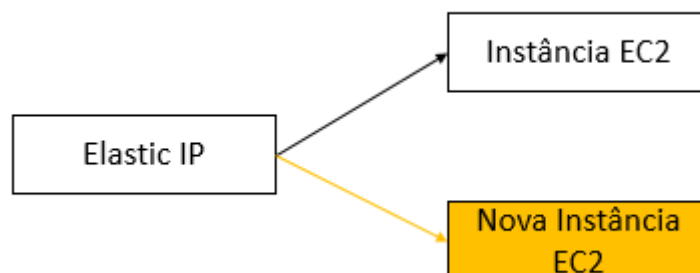
---

Os produtos de redes da AWS permitem que isolar a infraestrutura de nuvem e conectar à rede física ou à rede virtual privada, com possibilidade de segmentação, tal qual em uma infraestrutura local.

Os produtos de redes da AWS trabalham juntos para atender às necessidades das aplicações. Por exemplo, o Elastic Load Balancing trabalha com a Amazon Virtual Private Cloud (VPC) para disponibilizar sólidos recursos de rede e segurança.

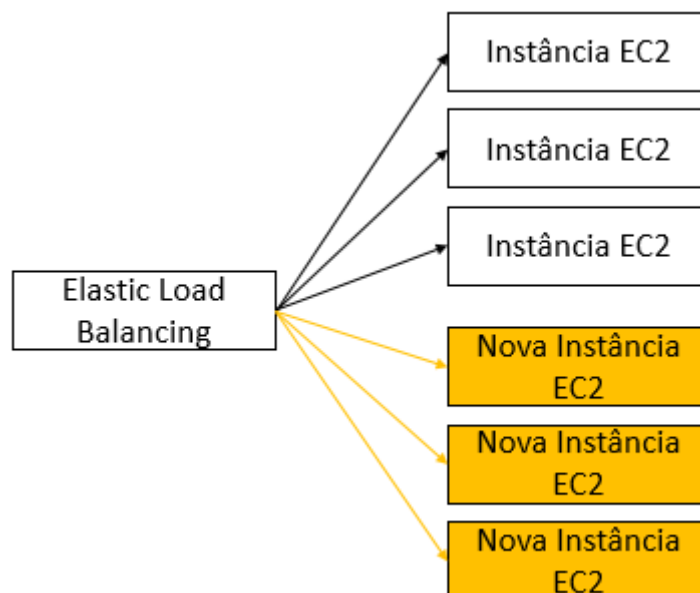
**Amazon Elastic IP:** um endereço IP elástico é um endereço IPv4 estático projetado para computação em nuvem dinâmica. Um endereço IP elástico está associado à conta da AWS. Com um endereço IP elástico, você pode mascarar a falha de uma instância rapidamente, associando o endereço para outra instância.

**Figura 7 – Elastic IP.**



**Amazon ELB (Elastic Load Balancing):** balanceamento de carga em alta escala. O Elastic Load Balancing distribui automaticamente o tráfego de entrada dos aplicativos em várias instâncias do Amazon EC2 na nuvem. Ele permite que você atinja níveis mais altos de tolerância a falhas em seus aplicativos, disponibilizando de forma transparente a capacidade necessária de load balancing para distribuir o tráfego dos aplicativos.

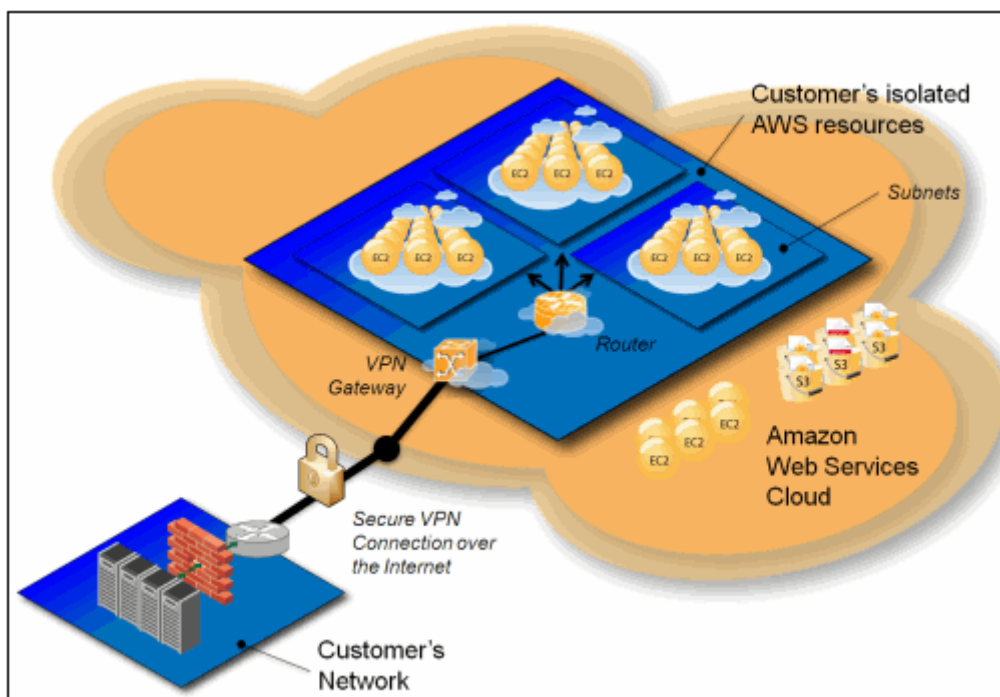
**Figura 8 – Amazon Elastic Load Balancing.**



**Amazon VPC (Virtual Private Cloud):** a Amazon Virtual Private Cloud (Amazon VPC) permite provisionar uma rede isolada logicamente, onde é possível executar recursos da AWS que se conectam via VPN com sua rede própria. Você tem controle total sobre seu ambiente de rede virtual, inclusive a seleção do seu próprio

intervalo de endereços IP, criação de segmentação de rede e configuração de tabelas de roteamento e gateways de rede.

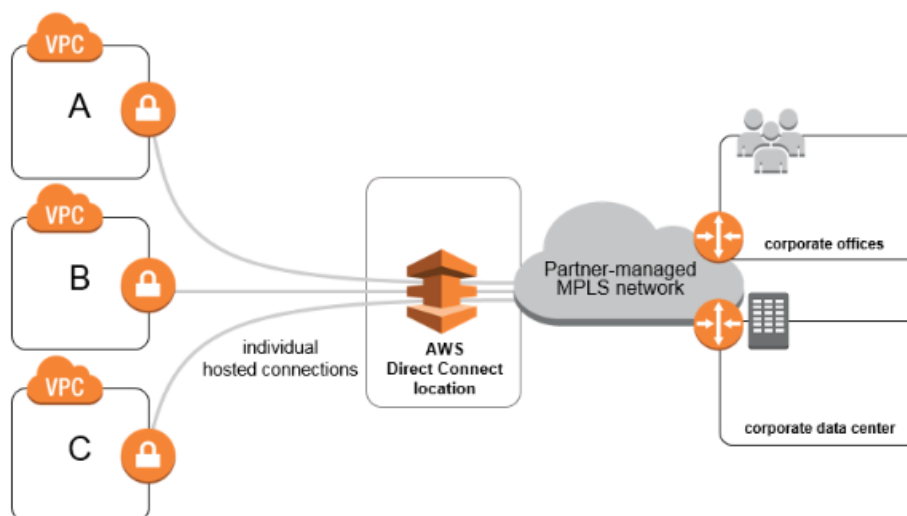
**Figura 9 – Amazon Virtual Private Cloud.**



Fonte: <https://aws.amazon.com/pt/blogs/aws/introducing-amazon-virtual-private-cloud-vpc/>

**AWS Direct Connect:** o AWS Direct Connect facilita o estabelecimento de uma conexão de rede dedicada entre o ambiente local e a AWS. Usando o AWS Direct Connect, você pode estabelecer conectividade privada entre a AWS e o seu datacenter, escritório ou ambiente de compartilhamento, possibilitando, em muitos casos, o aumento da taxa de transferência para a AWS. Além disso, ele proporciona uma experiência de rede mais estável do que a oferecida por conexões de internet.

**Figura 10 – AWS Direct Connect.**



Fonte: <https://aws.amazon.com/pt/answers/networking/aws-network-connectivity-over-mpls/>.

**Amazon Route 53:** o Amazon Route 53 é um serviço da web DNS, com disponibilidade e escalabilidade providas pela AWS. Ele foi projetado para oferecer aos desenvolvedores e empresas uma maneira altamente confiável e econômica de direcionar os usuários finais aos aplicativos de internet, convertendo nomes como `www.example.com` para endereços IP numéricos como `10.10.0.1`, usados pelos computadores para se conectarem entre si.

### Arquitetura de tolerância a falhas

A AWS fornece uma infraestrutura para criar sistemas confiáveis, tolerantes a falhas e altamente disponíveis na nuvem. Os serviços AWS foram projetados com estes requisitos de disponibilidade e, em sua maioria, são persistentes.

O Amazon EC2 fornece blocos de construção de infraestrutura que, por si mesmos, talvez não sejam tolerantes a falhas. Os discos rígidos poderão falhar, assim como as fontes de alimentação e os racks. É importante usar combinações dos recursos disponíveis na AWS para atingir a alta disponibilidade e a tolerância a falhas.



Vários serviços AWS, como o Amazon Simple Storage Service (S3), Amazon SimpleDB, Amazon Simple Queue Service (SQS) e Amazon Elastic Load Balancing (ELB), foram criados com suporte a alta disponibilidade e a tolerância a falhas. A instância EC2 é um serviço efêmero, mas outros serviços AWS como Amazon EBS fornece recursos específicos, como zonas de disponibilidade, endereços Elastic IP e snapshots, dos quais um sistema tolerante a falhas e altamente disponível deve usufruir e usar corretamente. Simplesmente mover um sistema para a nuvem não o torna altamente disponível.

O balanceamento de carga é uma maneira eficiente de aumentar a disponibilidade de um sistema. As instâncias que falharem poderão ser perfeitamente substituídas atrás do balanceador de cargas enquanto outras instâncias continuam operando. Elastic Load Balancing pode ser usado para balancear instâncias em várias zonas de disponibilidade de uma região.

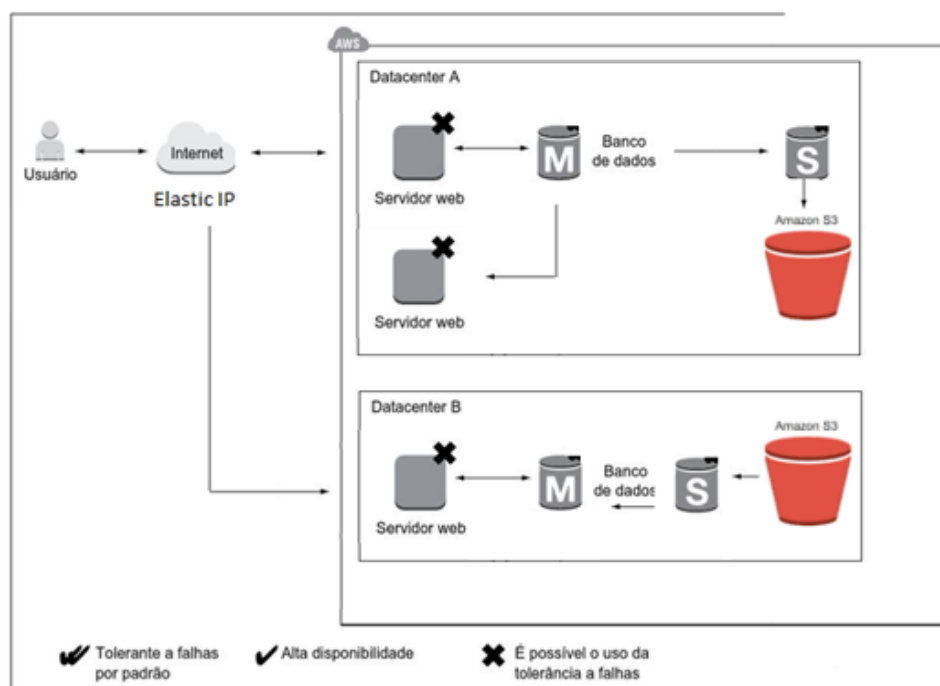
Zonas de disponibilidade são localizações geográficas distintas, projetadas para ser isoladas de falhas em outras zonas. Ao colocar instâncias do EC2 em várias zonas de disponibilidade, um aplicativo pode ser protegido contra falha em uma única localização. É importante executar aplicativos independentes em mais de uma zona de disponibilidade, seja na mesma região ou em outra, de forma que se uma zona falhar, o aplicativo na outra zona poderá continuar sendo executado.

Elastic IP são endereços IP públicos que podem ser mapeados de modo programado entre instâncias em uma região. Eles são associados à conta da AWS e não a uma instância específica ou período de uma instância. Os endereços do Elastic IP podem ser usados para solucionar falhas no host e em zonas de disponibilidade ao remapear rapidamente o endereço para outra instância, sendo executada, ou uma instância substituta que acabou de ser iniciada.

Os dados valiosos nunca devem ser armazenados somente em um armazenamento de instância, sem backups apropriados ou replicação. A Amazon EBS oferece volumes de armazenamento persistentes de instâncias desativadas, que são uma ordem de magnitude mais duráveis do que o armazenamento de instâncias ativas. Os volumes EBS são replicados automaticamente dentro de uma única zona de disponibilidade. Para aumentar a disponibilidade ainda mais, snapshots precisos

podem ser criados para armazenar dados em volumes no Amazon S3 que, em seguida, são replicados para outras zonas de disponibilidade. Embora os volumes de EBS estejam vinculados a uma zona de disponibilidade específica, os snapshots estão vinculados à região. Usando um snapshot, você pode criar volumes de EBS em qualquer outra zona de disponibilidade da mesma região. Essa é uma forma eficiente de lidar com falhas do disco ou outros problemas no nível do host, assim como com problemas que afetam uma zona de disponibilidade. Os snapshots são incrementais, então é aconselhável guardar os mais recentes. A figura 6 apresenta esta arquitetura de referência da AWS para aplicações com alta disponibilidade.

**Figura 11 – Arquitetura de referência AWS para alta disponibilidade.**



## Capítulo 4. Microsoft Azure

---

Este capítulo apresenta as ofertas do Microsoft Azure no que diz respeito a infraestrutura de TI.

### Sobre o Microsoft Azure

---

O Azure é um conjunto de serviços de nuvem que os desenvolvedores e os profissionais de TI usam para criar, implantar e gerenciar aplicativos. Com ferramentas integradas para o desenvolvimento de aplicações (inclusive para dispositivos móveis), o Azure se posiciona como uma oferta de serviços PaaS para o desenvolvimento de aplicações web.

- **O Azure é produtivo para desenvolvedores.**

As soluções do Azure vão de soluções que se baseiam no ambiente de desenvolvimento Microsoft e em soluções de desenvolvimento baseada em software livre, suportando aplicações .NET, Java, PHP, Ruby entre outros.

A plataforma Azure para o desenvolvimento de aplicações, permite que as empresas e desenvolvedores possam desenvolver, testar e publicar aplicações em um ambiente seguro e escalável, garantindo a qualidade das aplicações, incluindo ambientes Windows e Linux.

- **Azure é a única nuvem híbrida coerente.**

Com o Azure, a experiência Microsoft com o mundo corporativo on-premise ganha novas oportunidades, pois os serviços do Azure foram pensados para permitir a integração entre os serviços locais e os serviços na nuvem, criando uma solução híbrida com as mesmas funcionalidades independentemente de localidade.

Esta integração entre ambientes se aplica ao desenvolvimento de aplicações, gerenciamento de máquinas virtuais e gerenciamento de usuários e identidade.

- **O Azure é uma nuvem para soluções corporativas.**

Usando a força financeira da Microsoft, a infraestrutura Azure está presente globalmente em mais de 50 regiões espalhadas pelo mundo e com uma ampla gama de certificações, declarações e conformidades para suportar os negócios independentemente da indústria.

## Computação no Azure

---

Em termos de computação, o Azure oferece várias opções para executar os aplicativos, desde servidores web, máquinas virtuais até ambientes de desenvolvimento e execução de serviços para dispositivos móveis.

Os principais serviços de computação disponíveis no Azure são:

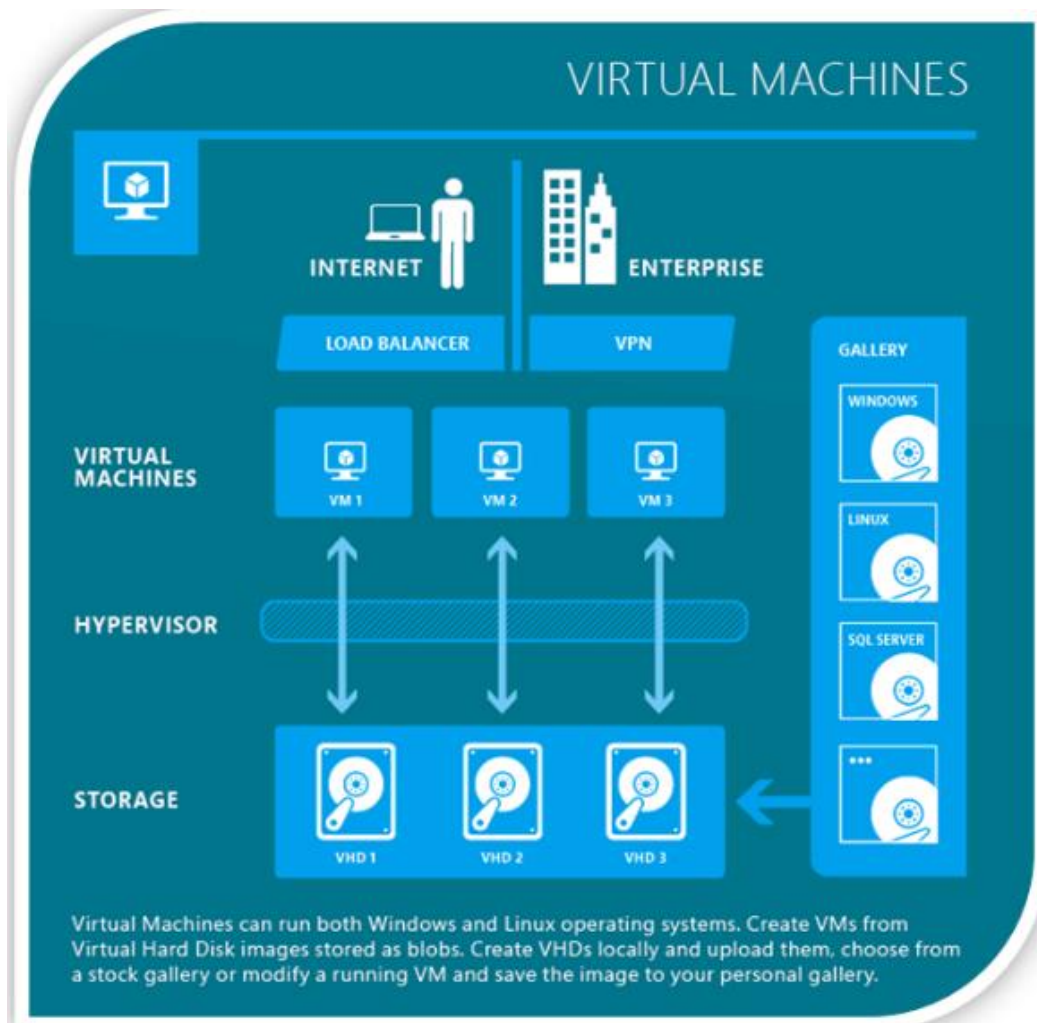
- **Servidores web.**

Servidores web do Azure são uma plataforma escalável, segura e flexível para construir aplicativos web para os negócios, capturando novas oportunidades de receita. Possui um portal onde pode-se optar pelas mais diversas soluções de portal web como WordPress. Como ambiente de desenvolvimento, pode-se trabalhar com .NET, PHP, HTML5 ou mesmo Microsoft Visual Studio.

- **Máquinas virtuais.**

Máquinas virtuais são a solução mais clássica de provisionamento de servidores no Azure. A partir da seleção do tipo (CPU, memória e SO), tem-se um ambiente inicial para a instalação de aplicativos da empresa, tal qual se faria em um ambiente virtual de uma infraestrutura local.

**Figura 12 – Azure Virtual Machines.**



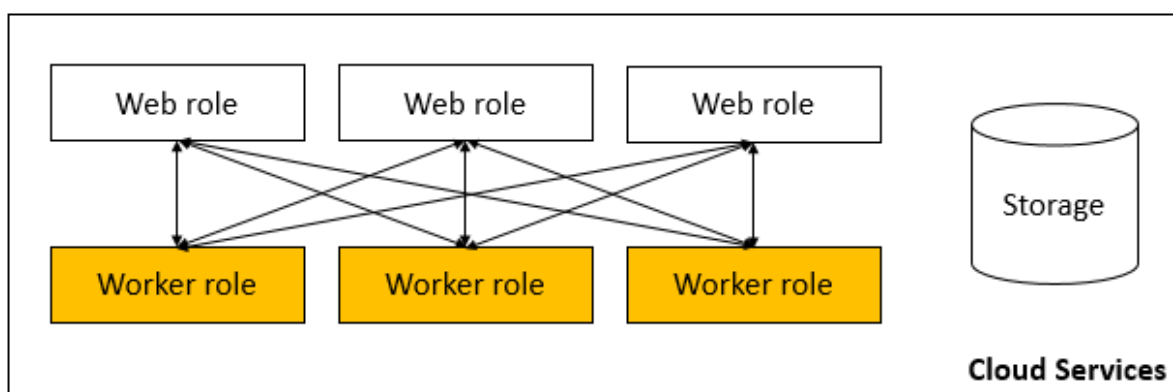
**Fonte:** <https://blogs.msdn.microsoft.com>.

Uma característica da interoperabilidade da Microsoft, está a possibilidade de você copiar um VHD para usar como template para as máquinas virtuais no Azure e vice-versa. O acesso ao ambiente remoto se faz por Remote Desktop Protocol (RDP) ou Windows PowerShell. Máquinas virtuais baseadas em Linux tem seu acesso via Secure Shell (SSH). Discos virtuais podem ser adicionados às máquinas virtuais para armazenar dados das aplicações. Instâncias de máquinas virtuais no Azure são persistentes, isto é, os dados serão preservados após o reboot da máquina virtual. Os discos de boot das máquinas virtuais são armazenados na solução de storage Azure com Azure blobs.

- **Cloud Services.**

Serviços de cloud do Azure permitem a criação de aplicativos em duas camadas: a camada de frontend web e a camada de aplicação, executando a lógica do aplicativo. Com esta arquitetura você consegue explorar a escalabilidade da nuvem, permitindo múltiplas instâncias web em paralelo para suportar um grande volume de usuários e conexões. Da mesma forma, pode-se implementar uma arquitetura horizontal para os workers para ambientes que demandam mais servidores de aplicação.

**Figura 13 – Azure Cloud Services.**



Usando serviços cloud, as empresas podem focar a atenção em desenvolver, testar e publicar aplicações sem se preocupar com os aspectos de preparação da infraestrutura para servidores de aplicação e suas integrações. Mais que isto, a plataforma é toda gerenciada pela Microsoft, como aplicação de patches, falhas de hardware ou análise de problemas de rede que muitas vezes impactam ambientes de aplicação web.

Nestes serviços, é possível usar o AutoScale para permitir a elasticidade da solução quando o volume de demanda aumenta ou diminui, equilibrando melhor o modelo financeiro, que é pelo uso dos serviços.

Diferentemente das máquinas virtuais, as instâncias web roles e worker roles do serviço de cloud do Azure são efêmeras e não persistem após o reboot.

- **Mobile Services.**

Azure mobile services são serviços na nuvem Azure para a construção de aplicações para dispositivos móveis suportadas por estruturas escaláveis no ambiente Azure.

Esta plataforma suporta dispositivos móveis baseados em Windows Phone, Android, Apple, além de HTML e JavaScript.

Trata-se de uma plataforma já preparada para o desenvolvimento de aplicações móveis com os principais serviços (autenticação e notificações) tipicamente necessários neste cenário instalados e configurados, acelerando o desenvolvimento destas aplicações.

### Banco de dados SQL no Azure

---

O SQL Azure é um serviço de banco de dados relacional gerenciado pelo Microsoft Azure, que dá suporte a estruturas como XML, espacial e dados relacionais. O SQL Azure oferece elasticidade e suporte a múltiplas instâncias SQL, com desempenho dinâmico e opções para geração de relatórios e OLTP in-memory para processamento transacional de alto desempenho.

A Microsoft trata todos os patches e a atualização da base de código SQL sem interrupções e abstrai todo o gerenciamento da infraestrutura subjacente.

O Banco de Dados SQL compartilha a sua base de código com o mecanismo de banco de dados do Microsoft SQL Server. Com a estratégia de prioridade de nuvem da Microsoft, os recursos mais recentes do SQL Server são liberados primeiro no Banco de Dados SQL e, em seguida, no próprio SQL Server. Essa abordagem fornece a você os recursos mais recentes do SQL Server sem sobrecarga para aplicação de patch ou atualização, e com esses novos recursos testados em milhões de bancos de dados.



Esses recursos permitem que você se concentre no desenvolvimento rápido de aplicativos e acelere seu tempo de colocação no mercado, em vez de alocar tempo e recursos preciosos para gerenciamento de máquinas virtuais e infraestrutura.

## Armazenamento no Azure

---

O armazenamento do Azure é um serviço gerenciado pela Microsoft que fornece diversas alternativas de armazenamento altamente disponível, seguro e redundante.

O armazenamento do Azure consiste em três serviços de dados: armazenamento de Blobs, armazenamento de filas, tabelas e arquivos.

O armazenamento de blobs dá suporte ao armazenamento standard e premium, com o armazenamento premium usando somente SSDs para que o desempenho seja o mais rápido possível. Outro recurso é o armazenamento frio, que permite que você armazene grandes quantidades de dados raramente acessados para reduzir o custo.

### ▪ Armazenamento de blob

Blobs são basicamente arquivos, como aqueles que se armazenam no computador ou smartphone. Eles podem ser imagens, arquivos do Microsoft Excel, discos rígidos virtuais (VHDs), logs, backups de banco de dados – basicamente qualquer tipo de arquivo. Blobs são armazenados em contêineres, que são semelhantes às pastas.

Depois de armazenar arquivos no armazenamento de Blobs, pode-se acessá-los de qualquer lugar no mundo usando URLs, através da interface REST. Eles também podem ser acessados por meio de aplicativos usando as linguagens Node.js, Java, PHP, Ruby, Python e NET.

Há três tipos de blobs – blobs de bloco, blobs de página (usados para arquivos VHD) e blobs de acréscimo.



- Blobs de blocos são usados para armazenar arquivos comuns de até aproximadamente 4.7 TB.
- Blobs de páginas são usados para manter arquivos de acesso aleatório de até 8 TB de tamanho. Eles são usados para os arquivos VHD que dão suporte a máquinas virtuais.
- Blobs de acréscimo são compostos de blocos como blobs de blocos, mas são otimizados para operações de acréscimo. Eles são usados para tarefas como registrar em log informações para o mesmo blob a partir de várias VMs.
- **Armazenamento de filas**

O serviço Fila do Azure é usado para armazenar e recuperar mensagens. Filas são geralmente usadas para armazenar as listas de mensagens a serem processadas de forma assíncrona. Por exemplo, um serviço web pode fazer uma requisição para o worker através da fila e aguardar o retorno em outra fila ou na mesma fila. Desta forma, desacoplamos os serviços, tornando as aplicações mais leves, sem ficar esperando o processamento inteiro. Mais que isto, em caso de falha de algum serviço (web ou worker), os dados na fila podem ser recuperados, sem que haja prejuízo no processamento.

- **Armazenamento de tabela**

O armazenamento de tabela é um serviço mais simples de manipulação de dados no formato tabela, entidade e propriedades, mas sem as propriedades ACID (atomicidade, consistência, isolamento e durabilidade) que encontramos nos bancos de dados relacionais. Desta forma, pode-se desenvolver aplicações escaláveis e leves para o processamento de dados, que são encadeados e possui uma estrutura mínima, mas que não demandaria o uso de um banco de dados relacional.

- **Arquivos do Azure**

A maioria das empresas faz uso do serviço de arquivos da Microsoft (file server) nas empresas, pela facilidade de armazenamento compartilhado de dados entre pessoas, departamentos e até mesmo organizações. O serviço de arquivos do

Azure é exatamente isto, um serviço baseado no protocolo SMB (Server Message Block) padrão e usado internamente para compartilhamento de arquivos. Desta forma, várias máquinas virtuais podem compartilhar os mesmos arquivos para leitura e gravação.

Estando em uma plataforma na nuvem, o serviço de compartilhamento de arquivos do Azure permite o acesso às pastas e arquivos compartilhados de qualquer lugar, usando o endereço do arquivo, de forma segura através de um token.

- **Replicação**

Para garantir que seus dados sejam duráveis, o Armazenamento do Azure manterá (e gerenciará) várias cópias de seus dados. Isso é chamado de replicação ou, algumas vezes, de redundância. Quando você configura sua conta de armazenamento, você seleciona um tipo de replicação. Na maioria dos casos, essa configuração pode ser modificada depois que a conta de armazenamento é configurada. A replicação pode ser com redundância local, com várias cópias no mesmo datacenter; replicação com redundância de zona de disponibilidade e replicação com redundância geográfica, com replicação de dados entre regiões.

## Network no Azure

---

O Azure fornece uma variedade de recursos de rede que podem ser usados juntos ou separadamente.

- **Virtual Network**

Os recursos do Azure, as Máquinas Virtuais, os Cloud Services e os serviços web do Azure, podem comunicar entre si de forma privada por meio de uma VNet (rede virtual) do Azure. Uma VNet é um isolamento lógico da nuvem do Azure dedicada à sua assinatura. É possível implementar várias VNets dentro de cada assinatura e região do Azure. Cada VNet é isolada de outras VNets. Para cada VNet, pode-se:

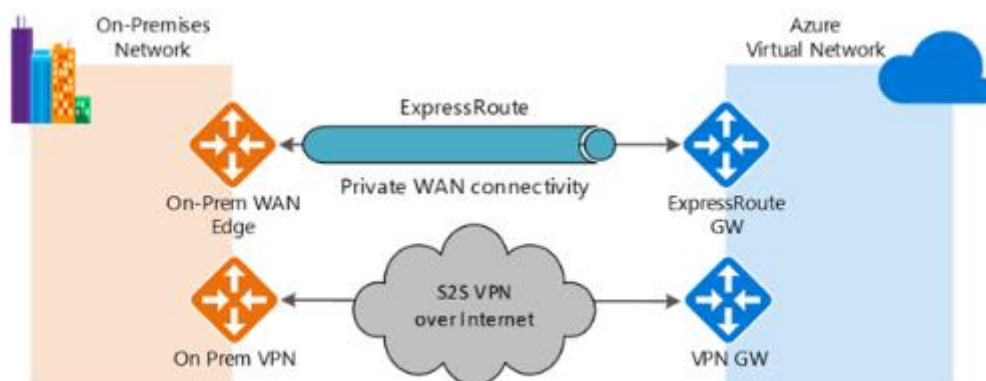
- Especificar um espaço de endereço IP privado.

- Segmentar a Virtual Network em sub-redes e alocar uma parte do espaço de endereço da VNet a cada sub-rede.
- Usar a resolução de nomes fornecida pelo Azure ou especificar seu próprio servidor DNS, que será usado pelos recursos conectados a uma VNet.
- **Emparelhamento:** permite que recursos conectados a diferentes VNets do Azure, na mesma região do Azure, comuniquem entre si. A largura de banda e a latência entre as VNets são as mesmas se os recursos estiverem conectados à mesma VNet.
- **Gateway de VPN:** permite que recursos conectados a diferentes VNets do Azure em diferentes regiões do Azure se comuniquem entre si. O tráfego entre as VNets flui por meio de uma conexão VPN do Azure.
- **Conectividade com a internet:** por padrão, todos os recursos do Azure conectados a uma VNet têm conectividade de saída com a internet. O endereço IP privado do recurso passa por NAT (conversão do endereço de rede de origem), para um endereço IP público pela infraestrutura do Azure.

Para comunicar-se internamente com os recursos do Azure da internet, ou para comunicar-se externamente com a internet sem NAT, é necessário atribuir um endereço IP público ao recurso.

Você pode acessar os recursos de sua VNet com segurança por meio de uma conexão VPN ou de uma conexão privada direta. Para enviar o tráfego de rede entre sua rede virtual do Azure e sua rede local, você deve criar um gateway de rede virtual. Defina as configurações do gateway para criar o tipo de conexão que você deseja, seja ela por VPN ou ExpressRoute.

**Figura 14 – Azure Virtual Network.**



Fonte: <https://azure.microsoft.com/en-us/blog/new-azure-vpn-gateways-now-6x-faster/>.

#### ▪ ExpressRoute

O ExpressRoute é estabelecido entre sua rede e o Azure por meio de um parceiro do ExpressRoute. Essa conexão é privada e o tráfego não atravessa a internet. A latência de uma conexão de ExpressRoute é previsível, pois o tráfego não atravessa a internet. O ExpressRoute pode ser combinado a uma conexão site a site.

#### ▪ Balanceamento de carga de rede

O Azure Load Balancer fornece balanceamento de carga de Camada 4 de baixa latência e alto desempenho para todos os protocolos TCP e UDP. Ele gerencia conexões de entrada e saída. Você pode configurar pontos de extremidade com balanceamento de carga públicos e internos. Você também pode definir regras para mapear conexões de entrada para destinos de pool de backend, usando opções de investigação de integridade TCP e HTTP para gerenciar a disponibilidade do serviço.

#### ▪ Balanceamento de carga do aplicativo

O serviço de Gateway de aplicativo do Azure fornece o ADC (controlador de entrega de aplicativo) como um serviço. O Gateway de aplicativo oferece vários recursos de balanceamento de carga de Camada 7 (HTTP/HTTPS) para seus aplicativos, incluindo um firewall de aplicativo Web para proteger seus aplicativos Web contra vulnerabilidades e explorações. O Gateway de Aplicativo também permite

otimizar a produtividade do farm web descarregando a terminação SSL com uso intensivo de CPU para o Gateway de aplicativo.

- **Gerenciador de tráfego**

O serviço do Gerenciador de tráfego do Azure fornece balanceamento de carga DNS global. O Gerenciador de tráfego responde aos clientes com o endereço IP de um ponto de extremidade íntegro. Os clientes são direcionados para pontos de extremidade específicos (Azure, externo ou aninhado) com base no local geográfico em que as respectivas consultas DNS são originadas. Esse método habilita cenários em que conhecer a região geográfica do cliente e roteá-lo com base nisso, é importante. Os exemplos incluem conformidade com normas de soberania de dados, localização de conteúdo e experiência do usuário, bem como medição do tráfego de diferentes regiões.

O cliente se conecta diretamente ao ponto de extremidade. O Gerenciador de tráfego do Azure detecta quando um ponto de extremidade não está íntegro e redireciona os clientes para outro ponto de extremidade íntegro.

### Arquitetura de tolerância a falhas

---

Para desenvolver uma arquitetura que promova alta disponibilidade ou tolerância a falhas para um aplicativo, é necessário entender a arquitetura por trás da infraestrutura que suporta o aplicativo ou máquina virtual.

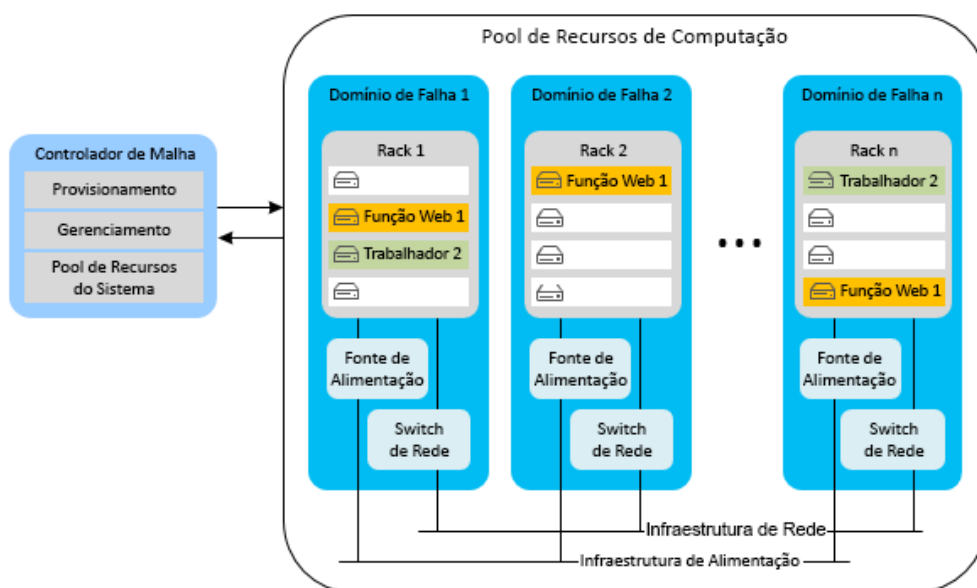
O Azure oferece vários recursos para oferecer uma plataforma de alta disponibilidade para os serviços na nuvem. Neste tópico, vamos descrever alguns destes serviços e discutir arquiteturas de referência Azure para prover alta disponibilidade e tolerância a falhas aos serviços Azure.

- **Fabric Controller:** o controlador de malha (Fabric Controller) é responsável por provisionar e monitorar as instâncias de computação do Azure. Ele monitora a infraestrutura de TI onde ficam hospedadas as instâncias provisionadas, em termos de falha de hardware e software. Quando ele detecta

uma falha, o fabric controller também é responsável por realocar automaticamente a máquina virtual para um outro servidor, garantindo que o serviço virtual continue operacional.

- **Domínios de falha e atualização:** junto com o controlador de malha, o Azure introduziu o conceito de domínios de falha e atualização. São um conjunto de hardware (servidores e rede, tipicamente no mesmo rack). Ao distribuir múltiplas instâncias de um mesmo serviço, o controlador de malha automaticamente aloca estas instâncias em diferentes domínios de falha, garantindo, assim, que a falha de um domínio inteiro não provoque a interrupção completa do serviço. Este mecanismo também permite que o Azure atualize a plataforma automaticamente, sem impactar as instâncias de computação. A figura abaixo mostra o conceito.

**Figura 15 – Azure Fabric Controller.**



Fonte: <https://docs.microsoft.com/pt-br/azure/architecture/resiliency/high-availability-azure-applications>.

Outros serviços Azure possui mecanismos próprios para garantir a alta disponibilidade.

- O Armazenamento Azure mantém três cópias do dado dentro de um datacenter. Além disso, como já apresentamos, pode-se implementar um mecanismo de replicação de dados, inclusive geograficamente.
- Banco de dados SQL Azure mantém réplicas para alta disponibilidade também, de forma síncrona dentro da mesma região. Além disso, o serviço de banco de dados SQL oferece mecanismo de backup pontual (snapshot) e a possibilidade de replicação geográfica.
- Armazenamento blob e tabelas são replicados para uma outra região, mas estas réplicas são protegidas de acesso no Azure.

Convém observar que estes mecanismos de replicação são síncronos e não impedem situações de corrupção de dados e, para tanto, os snapshots são importantes para recuperar um momento no tempo anterior. Outra preocupação importante é garantir um mecanismo de proteção de informação no contexto da aplicação, pois estes recursos mencionados até agora não monitoram nem protegem a aplicação em si, apenas os serviços oferecidos pelo Azure.

Conjuntos de disponibilidade são estruturas onde residem os domínios de falha já mencionados. Portanto, ao desenvolver uma aplicação escalável e crítica, que requer alta disponibilidade, é fundamental distribuir as instâncias entre conjuntos de disponibilidade, aumentando o SLA desta aplicação.

Outras estratégias são mais adequadas no contexto da aplicação, mas igualmente importantes.

- **Comunicação assíncrona:** com o uso de filas para comunicação entre serviços, ou seja, um aplicativo grava uma informação na fila (como se fosse uma requisição) para o outro aplicativo (um worker, por exemplo), recebe esta requisição e faz o processamento. Ao devolver, ele novamente grava a resposta na fila para que o serviço web que requisitou o processamento possa receber a resposta. O processamento assíncrono torna as aplicações desacopladas e leves. Apesar de introduzir um elemento a mais na cadeia de processamento do aplicativo (a fila), este mecanismo oferece mais

disponibilidade pois em falha do serviço web ou do serviço worker é possível recuperar o processamento no ponto em que ele parou, já que os dados estão armazenados de forma persistente nas filas.

- Uma variação desta implementação de filas ocorre em chamadas diretas para banco de dados. Se ele estiver indisponível, após repetidas tentativas sobre o banco de dados, pode-se armazenar a requisição em uma fila para processamento posterior. Neste caso, só se usa a fila em situação onde a chamada direta (e síncrona) não esteja disponível.

Com várias instâncias em diversas camadas (web e servidores de aplicação), é importante garantir reconexões em caso de falha, pois a reconexão tipicamente permite a continuidade da solução logo em seguida. A plataforma de desenvolvimento Azure já implementa, nativamente, estruturas de repetição para aplicações escaláveis, pois isto é fundamental neste tipo de aplicativo.



## Capítulo 5. Cloud Privada

---

Cloud Privada é a infraestrutura da nuvem que é dedicada ao uso exclusivo de uma única organização, que pode atender sua matriz, filiais, departamentos e subsidiárias. Ela pode ser de propriedade, gerenciada e operada pela organização, por terceiros ou alguma combinação dos dois, e pode estar dentro ou fora das instalações da organização.

Este capítulo apresenta alternativas de infraestrutura de TI, próprias para a criação de uma cloud privada.

### Aspectos gerais

---

A virtualização revolucionou significativamente a natureza da computação nos datacenters, permitindo a consolidação de servidores e mais poder de processamento para as aplicações.

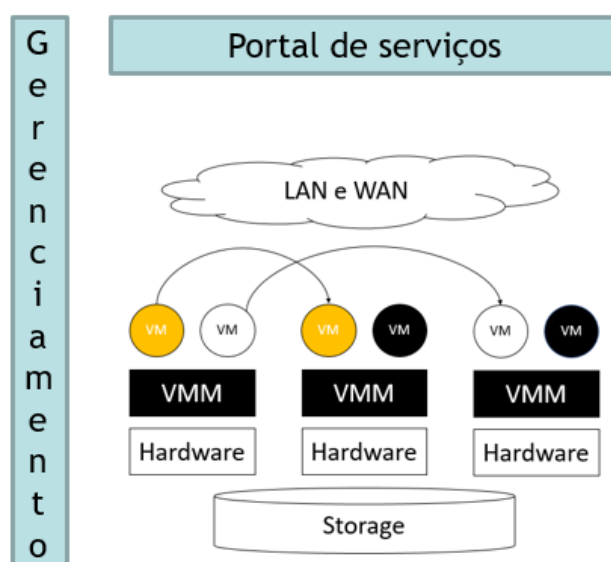
O ambiente virtual aumenta a capacidade de processamento e armazenamento, mas criou um efeito de propagação no número de máquinas virtuais em uma organização, tornando o ambiente de serviços de TI mais complexos de gerenciar e controlar.

Esta complexidade escondeu as melhorias obtidas na flexibilidade dos serviços de TI pois, com a proliferação na quantidade de aplicações e servidores virtuais, o ambiente de rede e a segurança da informação ficou mais complexa de gerenciar. O que promoveria a otimização de recursos e simplificaria a infraestrutura de TI tornou-se um ambiente complexo e com maior custo de gerenciamento.

Com o objetivo de simplificar a infraestrutura de TI, a implementação de computação em nuvem privada incorporou ao ambiente virtual tradicional dois elementos importantes: o portal de serviços e o gerenciamento dos recursos de forma automática.

A cloud privada provê simplicidade e agilidade aos serviços de TI e pode ser implementada em uma infraestrutura tradicional (figura 16), infraestrutura convergente ou infraestrutura hiperconvergente.

**Figura 16 – Arquitetura de referência AWS para alta disponibilidade.**



Em uma infraestrutura tradicional, os servidores, storage, rede, camada de virtualização, portal de serviços e gerenciamento dos serviços de TI são, via de regra, cada um de um fabricante diferente, o que torna o processo de implementação mais lento e sujeitos a dificuldades de integração. Mesmo que um único integrador seja contratado, após a implementação a responsabilidade operacional da solução é da empresa, o que torna o processo mais caro e exige proficiência nas soluções dos diversos fabricantes.

Na tentativa de diminuir o esforço de implementação e os riscos operacionais de gerenciamento, o mercado desenvolveu as soluções convergentes, onde vários fabricantes se uniram para estabelecer guias de melhores práticas para implementação de cloud privada. Alguns fabricantes, inclusive, desenvolveram soluções próprias (como Dell, HPE e Oracle) e outros fizeram alianças para as soluções convergentes (como Cisco e Netapp com o FlexPod). O grande benefício das soluções convergentes é a rápida implementação baseada nos guias de melhores práticas. Contudo, os custos de gerenciamento operacional não foram eliminados.

Infraestrutura convergente é, portanto, uma das muitas mudanças transformacionais no mercado de TI, pois infraestrutura convergente parece ser uma boa escolha para simplificar ambiente de virtualização complexa, graças aos seus problemas de integração. Contudo, os custos de aquisição da solução convergente não mudaram significativamente em relação ao modelo tradicional.

Na tentativa de apresentar uma oferta mais próxima dos serviços de computação em nuvem pública, mas em um formato privado, o mercado desenvolveu as soluções hiperconvergentes, que detalharemos a seguir.

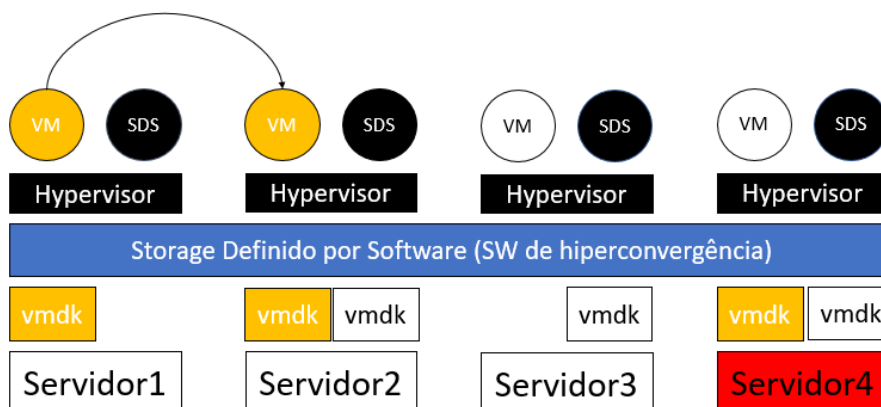
### Solução hiperconvergente

---

Uma solução hiperconvergente pode ser definida como uma solução completa que oferece a camada de computação e dados no mesmo nó. Isto permite a localização da carga de trabalho no nó de armazenamento, minimizando a necessidade de mover os dados no processamento das aplicações. Uma solução de storage definido por software aproveita o armazenamento interno de cada nó para criar um armazenamento central compartilhado, onde residem as máquinas virtuais da estrutura de computação. Nesta arquitetura, tanto os servidores quando o armazenamento de dados é definido por software. Esse recurso garante a migração de máquinas virtuais, garantindo a disponibilidade do ambiente virtual mesmo em caso de falha de um nó.

Fornecedores costumam usar servidores x86 e tomar um sistema distribuído de arquivos (scale-out) ou uma abordagem de armazenamento de objetos, fazendo com que estas soluções sejam rentáveis e altamente escaláveis. Esta infraestrutura segue o modelo usado por grandes datacenters de computação em nuvem e redes sociais, como AWS, Google ou Facebook, mas foram remodeladas para empresas de todo tamanho, já que a escalabilidade pela adição de novos nós é uma característica das soluções hiperconvergentes. A figura 17 mostra um diagrama da solução hiperconvergente.

**Figura 17 – Arquitetura de referência AWS para alta disponibilidade.**



Cópias de vmdk das máquinas virtuais são replicadas entre os nós, garantindo a resiliência do storage virtual mesmo diante da falha de um servidor (no exemplo, servidor4).

Outras características da solução hiperconvergente que se aproximam da nuvem pública são:

- **Simplicidade:** usando apenas servidores com discos internos, o ambiente de infraestrutura de TI é mais simples, pois contém menos equipamentos.
- **Agilidade:** através de uma estrutura virtual é possível disponibilizar novos serviços de TI através das máquinas virtuais.
- **Crescimento sob demanda:** o crescimento de uma solução hiperconvergente se dá pela adição de novos nós, propiciando o conceito de pagamento pelo uso. Embora não seja mensal por subscrição, como na computação em nuvem, ele se aproxima do modelo de crescimento sob demanda.

Evidentemente, a segurança da cloud privada é um diferencial da solução em relação ao modelo de computação em nuvem.

Entre os destaques no mercado hiperconvergente estão a Nutanix, HPE Simplivity, Dell vxRail, VMware vSAN, Cisco HyperFlex, entre outros.

## Referências

---

BCS - The Chartered Institute for IT. *Cloud Computing - Moving IT out of the office*. Disponível em: <<http://www.bcs.org/upload/pdf/cloud-computing.pdf>>. Acesso em: 18 nov. 2020.

HUTTEN, Dennis. *Azure: Microsoft Azure Tutorial The Ultimate Beginners Guide*. Microsoft, 2017.

LAAN, Sjaak. *It Infrastructure Architecture - Infrastructure Building Blocks and Concepts*. Lulu Press Inc, 2011.

NETO, Jocildo Figueiredo Correia. *Decisões de Investimentos Em Tecnologia da Informação*. Rio de Janeiro: Elsevier, 2015.

REZENDE, Denis Alcides; ABREU, Aline França de. *Tecnologia da Informação Aplicada a Sistemas de Informação Empresariais*. São Paulo: Atlas, 2009.

SCHMIDT, Klaus. *High Availability and Disaster Recovery*. Springer, 2006.

SIROHI, Savitra. *Step by Step: Fault Tolerant, Scalable, and Secure AWS Stack*. Amazon Web Services, 2018.

TULLOCH, Mitch. *Introducing Windows Azure for IT Professionals*. Microsoft, 2013.

VERAS, Manoel. *Cloud Computing: Nova Arquitetura de TI*. Brasport, 2012.

VERAS, Manoel. *Datacenter componente central da infraestrutura*. Brasport, 2009.