

## 2 Computação na Nuvem

### 2.1 Definição

A ideia essencial da computação na nuvem é permitir um novo modelo onde o consumo de recursos computacionais, e.g., armazenamento, processamento, banda entrada e saída de dados, são realizados através de serviços.

Podemos traçar um paralelo entre o cenário atual com o final do século XIX, durante o período da Revolução Industrial, quando grandes fabricas eram responsáveis pela produção de sua própria energia. Hoje as fábricas consomem energia como um serviço e pagam pela quantidade utilizada aos fornecedores. A computação na nuvem tem uma proposta similar [1]. Recursos computacionais passarão a ser de responsabilidade de algumas empresas especializadas, que ficarão responsáveis por sua gestão e comercialização através de serviços [2].

Evidente que esta proposta representa uma grande quebra de paradigma, pois, atualmente, tanto empresas quanto pessoas físicas, em sua maioria, utilizam recursos computacionais de forma proprietária, ou seja, são os donos e responsáveis pela gestão, manutenção e atualização dos recursos computacionais que dispõem.

O termo nuvem é uma metáfora em relação à forma como a Internet é usualmente mostrada nos diagramas de rede - como uma nuvem. Nesses diagramas, o ícone da nuvem representa todas as tecnologias que proporcionam o funcionamento da Internet, abstraindo a infraestrutura e a complexidade que ela engloba. No caso de computação na nuvem, refere-se, de modo geral, a uma combinação de tecnologias e paradigmas arquitetada e utilizada de forma a abstrair os seus detalhes do usuário (virtualização, computação utilitária, computação em grade, arquitetura orientada a serviços, dentre outros). Utilizar os serviços dessa nuvem (armazenamento, banco de dados, processamento, entre outros), é fazer computação na nuvem. [3]

O termo computação na nuvem foi introduzido em 2006, pelo CEO da Google, Eric Schmidt, quando este o utilizou para descrever os serviços da própria empresa, e, pouco tempo depois, quando a Amazon utilizou o mesmo termo para lançar seu serviço EC2 (*Elastic Compute Cloud*). De fato, o termo foi

popularizado em artigo da Wired, de George Gilder, intitulado “The Information Factories” (As Fábricas de Informação) [4].

Neste texto optamos, sempre que possível, por apresentar a tradução dos termos para o Português, formatado com letras minúsculas. Desta forma, o termo *Cloud Computing* foi traduzido para computação na nuvem.

Prover uma única definição de computação na nuvem é muito difícil. No entanto, optamos neste trabalho pela definição do grupo Gartner, que define computação na nuvem como “*um estilo de computação onde capacidades de Tecnologia da Informação elásticas e escaláveis são providas como serviços para usuários através da Internet.*” [5]

Segundo Armbrust et al. [6] este novo paradigma oferece as seguintes novidades ligada as características de hardware:

1. Ilusão de recursos computacionais infinitos, disponibilizados sob demanda, eliminando a necessidade do planejamento para a provisão de recursos no longo prazo.

2. Eliminação da necessidade de se fazer grandes investimentos iniciais em infraestrutura, permitindo com que negócios sejam iniciados com um parque computacional pequeno e que aumentem sua infraestrutura na medida em que suas necessidades demandarem.

3. Possibilidade da contratação de recursos computacionais por curto prazo, e.g. processadores por hora, armazenagem por um dia. Uma vez que estes não são mais necessários, capacidade de finalizar os contratos.

Vaquero et al. [7] realizaram um amplo estudo onde foram consideradas dezenas de diferentes definições para o conceito de computação na nuvem. Portanto, sugerimos a consulta deste material para outras definições, pois é interessante notar que os autores remarcam que não encontraram um denominador comum entre todas as definições estudadas, i.e., uma característica comum a todas as definições.

O conjunto de hardware, software, tecnologias e paradigmas que viabilizam a computação na nuvem é o que chamamos de nuvem [5]. Quando a nuvem é fornecida para o público em geral e sob um contrato onde se paga pelo montante utilizado, chamamos a mesma de nuvem pública. Os serviços comercializados são geralmente chamados de computação utilitária. Alguns exemplos de fornecedores

deste serviço são: Amazon, Google AppEngine e Microsoft Azure. Nesse trabalho estamos interessados exclusivamente neste modelo de computação em nuvem.

A Tabela 1 a seguir, adaptada de Cearley [5], faz um resumo das diferenças entre a computação tradicional e a computação na nuvem.

<b>Modelo</b>	<b>Computação Tradicional</b>	<b>Computação na Nuvem</b>
<b>Aquisição</b>	Aquisição de hardware, espaço físico e infraestrutura de instalação e manutenção	Contrato de serviço
<b>Negócio</b>	Custo afundado, desperdício de ativos e overhead administrativo (manutenção, suporte, etc.)	Pagamento baseado na utilização dos serviços
<b>Técnico</b>	Único “morador”, sem compartilhamento e estático	Condomínio, elástico, escalável e dinâmico
<b>Alocação</b>	Alocação realizada com base na previsão de demanda	Alocação ajustada em relação a demanda

Tabela 1 – Comparação da computação tradicional e na nuvem

## 2.2 Características Essenciais

Apesar da falta de consenso sobre um conjunto essencial a aplicativos de computação na nuvem, existem algumas características centrais citadas na maioria dos textos relevantes sobre o assunto, quais sejam: (a) virtualização de recursos; (b) independência de localização; (c) elasticidade; e (d) modelo de pagamento baseado no consumo [7].

A virtualização de recursos, obtida através do uso de tecnologias já estabelecidas como virtualização de máquinas, de memória, de armazenamento e de rede, separa os serviços de infraestrutura dos recursos físicos. Com essa abstração, a localização real dos recursos é tratada na camada mais baixa da computação na nuvem, sendo transparente para as demais camadas e, conseqüentemente, para seus usuários. Desta forma, os recursos passam a ser disponibilizados e consumidos como utilitários.

Em uma arquitetura baseada em serviços, características do consumidor são abstraídas das características do provedor através de interfaces de serviços bem definidas. Estas interfaces ocultam os detalhes e possibilitam trocas automatizadas entre provedores e consumidores de serviços. Neste modelo, serviços ganham um nível a mais de abstração, ou seja, passam a ser desenhados para suprir necessidades específicas dos consumidores, ao invés de se ater a detalhes tecnológicos.

Na computação na nuvem, os serviços se tornam acessíveis de qualquer lugar que se tenha acesso à Internet. As nuvens aparentam ser o ponto único de acesso para todas as necessidades de computação dos usuários.

A elasticidade é, talvez, a característica mais inovadora de computação na nuvem. Há uma diferença sutil entre elasticidade e escalabilidade. Escalabilidade é a habilidade de satisfazer um requisito de aumento da capacidade de trabalho pela adição proporcional da quantidade de recursos. Uma arquitetura escalável é construída tipicamente com base em uma infraestrutura básica passível de repetição, cujo crescimento pode ser alcançado simplesmente com a adição repetidamente do mesmo conjunto de hardware básico.

Tradicionalmente, a escalabilidade é projetada para garantir que o custo operacional possa crescer linearmente de acordo com a carga de trabalho. Usualmente, não há preocupação com a remoção de recursos nem a preocupação se os recursos são plenamente utilizados, porque os recursos adquiridos já são custo afundado. Por outro lado, a elasticidade é a capacidade de provisionar e desprovisionar rapidamente grandes quantidades de recursos em tempo de execução.

A Figura 1 exemplifica como a elasticidade da nuvem (à direita) permite provisionar e desprovisionar os recursos de forma a acompanhar a demanda, reduzindo ao máximo a ocorrência de escassez (vermelho) e ociosidade (amarelo) de recursos, enquanto o modelo tradicional (à esquerda) é obrigado a escalar a infraestrutura pelos picos de demanda, causando desperdício de recursos.

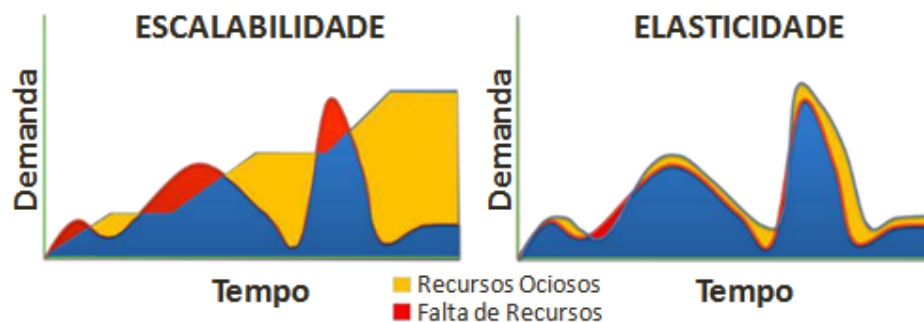


Figura 1 – Escalabilidade x Elasticidade

A elasticidade pode ser medida quantitativamente através dos seguintes fatores: (a) velocidade no provisionamento e desprovisionamento dos recursos; (b) quantidade máxima de recursos que podem ser provisionados; (c) granularidade na contabilidade do uso dos recursos.

É importante notar que mesmo que se consiga elasticidade, quando uma aplicação é colocada em um ambiente de computação na nuvem, a escalabilidade não é uma consequência direta, ou seja, não é um atributo natural da arquitetura do sistema. Para que o sistema seja de fato escalável é preciso que tenha sido desenvolvido levando este requisito não funcional em conta desde o início. Em casos extremos, o fato de colocar o aplicativo em um ambiente de computação na nuvem pode ter efeitos contrários do desejado, i.e., consumo de recursos maior que o esperado, sem benefícios adicionais. Em momento oportuno voltaremos a discutir esses conceitos no âmbito deste trabalho.

A outra característica central da computação na nuvem é o modelo de pagamento de acordo com a utilização dos recursos. Atualmente, serviços utilitários tais como água, eletricidade, gás e telefonia são considerados essenciais no cotidiano. Estes serviços são utilizados de forma frequente e por isso precisam estar disponíveis para atender a necessidade de seus consumidores. A maior parte destes serviços é oferecida através de modelos de contratos e convênios com provedores onde se paga apenas pela quantidade utilizada dos serviços.

Dentro do paradigma de computação na nuvem, os consumidores de serviços e recursos computacionais pagarão aos provedores apenas quando utilizarem de tais serviços. A grande vantagem do modelo é permitir a contratação de novos recursos na medida em que estes se fizerem necessários, liberando-os quando forem dispensáveis, tornando, deste modo, desnecessários grandes investimentos em infraestrutura.

Do ponto de vista de negócio, há dois direcionadores principais para adoção da computação na nuvem, que também são válidos para adoção de novas tecnologias de um modo geral: (i) redução de custos e (ii) aumento de capacidade.

A redução de custos tem um caráter evolucionário, devido ao seu modelo de cobrança, qual seja, pagamento por uso (nos casos de utilização de um provedor de nuvem público). O aumento da capacidade, por sua vez, tem um caráter revolucionário, em razão de suas características de elasticidade e acesso aos recursos via Internet.

## 2.3 Ambientes de Computação na Nuvem

O uso de ambientes de computação na nuvem viabiliza o surgimento de novas aplicações que se beneficiem, especialmente, da elasticidade no provisionamento de recursos computacionais e do pagamento pela utilização dos mesmos, tornando possível realizar sua alocação inteligente em função da demanda por processamento, custo de alocação e requisitos do negócio.

A Figura 2 ilustra quatro cenários nos quais os ambientes de computação na nuvem podem revolucionar.

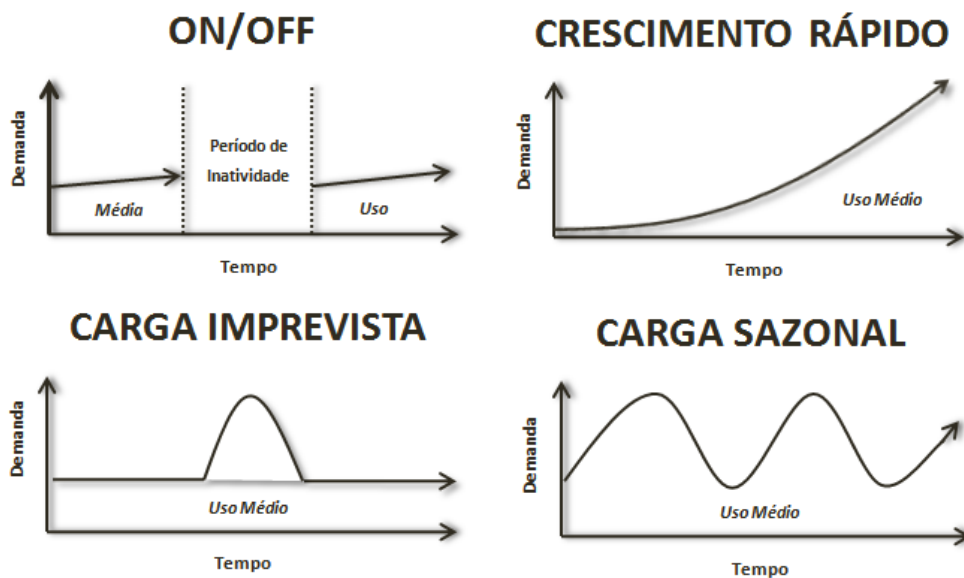


Figura 2 – Cenários típicos de ambientes de computação na nuvem

A descrição de cada cenário é apresentada a seguir.

- (i) Os recursos computacionais podem ser **ligados/desligados** (*On/Off*) para evitar custos enquanto ociosos, como, por exemplo, na realização de grandes rotinas de processamento em segundo plano;
- (ii) **Crescimento rápido** dos recursos computacionais sem um longo processo de aquisição, por exemplo, para atendimento ao crescimento da base de clientes em função do sucesso de um serviço;

- (iii) **Carga imprevista** sobre os recursos computacionais com a necessidade de escalar muitas vezes para atendimento deste pico, podendo, posteriormente, retornar a situação inicial, como, por exemplo, para atender a um imediatismo;
- (iv) **Carga sazonal** sobre os recursos computacionais, cujo provisionamento pelo pico causaria ociosidade de recursos, como, por exemplo, aumento do processamento de dados próximo a datas específicas.