

Fundamentos de Computação Nuvem para Governos

Adriano Martins

Serviço Federal de Processamento de Dados (SERPRO)

Brasília – DF – Brasil

adriano.martins@serpro.gov.br

Abstract. *This paper describes the history of the emergence of the term cloud computing, its concepts from the bare metal to utility computing models and also include its services models and clouds models practiced by much of the current IT market worldwide. It also warns those involved in its implementation on the implications for IT as well as the forces of influence for a new concept that comes to the formation of this new model. Finally, discusses essentials standards for those who need to implement it.*

Resumo. *Este artigo descreve o histórico do surgimento do termo computação em nuvem, seus conceitos desde o bare metal até a computação utilitária passando também pelos seus modelos de serviços e de nuvens praticado pela grande parte do atual mercado de TI mundial. Além disso, alerta os envolvidos em sua implementação sobre as implicações para TI bem como as forças de influência de um novo conceito que surge para formação desse novo modelo. Por fim, trata de padrões de tecnologia essenciais para quem necessita começar sua implementação.*

1. O que é a nuvem?

O termo nuvem tem sido usado historicamente como uma metáfora para a internet. Seu uso foi originalmente derivado de sua descrição em diagramas de rede como um delineamento de uma nuvem, usados para representar transportes de dados através backbones de rede até um outro ponto final do outro lado da nuvem. Esse conceito é datado do início do ano de 1961 quando o Professor John McCarthy sugeriu que a computação de compartilhamento de tempo poderia levar a um futuro onde o poder computacional e até aplicações específicas seriam vendidas através de um modelo de negócios utilitário.

Essa idéia era muito popular no final de década de 60. Entretanto, no meio da década de 70 ela foi abandonada quando se tornou claro que as tecnologias da informação da época não estavam aptas a sustentar um modelo desses de computação futurística.

Entretanto, com a virada do milênio, esse conceito foi revitalizado e foi neste momento de revitalização que a computação em nuvem começou a emergir nos círculos de tecnologia.

2. O surgimento da computação em nuvem.

Computação utilitária pode ser definida como o provisionamento de recursos computacionais e de armazenamento como um serviço que pode ser medido, similar àqueles providos pelas empresas de públicas que prestam esses tipos de serviços. E é claro que isso não é uma nova idéia nem mesmo algo com um nível alto de inovação ou outro mesmo quebra de paradigma. Esse conceito novo de computação tem crescido e

ganhado popularidade, tanto que empresas têm estendido o modelo de computação em nuvem provendo servidores virtuais em que departamentos de TI e usuários podem requerer acesso sob demanda.

Algumas empresas que estão entrando agora no ramo da computação utilitária usam principalmente para necessidades não críticas; mas isso está mudando rapidamente, já que questões de segurança e confiabilidade estão sendo resolvidas.

Algumas pessoas pensam que computação em nuvem é o próximo grande boom do mundo de TI. Outros acreditam que é somente outra variação de computação utilitária que foi repaginada na década passada como uma nova tendência.

Entretanto, não é somente a buzzword “computação em nuvem” que está causando confusão entre as massas. Atualmente, com poucos players de mercado praticando esta forma de tecnologia e mesmo analistas de diferentes companhias definindo o termo diferentemente, o significado do termo tem se tornado nebuloso.

3 Evolução das Máquinas

É importante entender a evolução da computação para se ter a contextualização do ambiente que se fala hoje de computação em nuvem. Olhando para a evolução do hardware, desde a primeira geração até a quarta e atual, mostra como nós chegamos até aqui.

A primeira geração é de 1943, quando os computadores Mark I e Colossus foram desenvolvidos, eles foram construídos usando circuitos hard-wired e tubos a vácuo ambos usados durante a guerra.

A segunda geração é de 1946, ocasião em que foi construído o famoso ENIAC. Esse foi o primeiro computador reprogramável e capaz de resolver um grande leque de problemas computacionais. Foi construído com tubos termiônicos e teve aplicabilidade durante a guerra. Mas o que marcou a segunda geração foram os computadores com transistores, o que dominou o final dos anos 50 e início dos 60. Apesar de usarem transistores e circuitos impressos, eles eram caros e pesados.

A terceira geração foi consagrada pelos circuitos integrados e microchips e foi nesta era que começou a miniaturização dos computadores e eles puderam ser portados para pequenos negócios.

A quarta geração é a que estamos vivendo neste momento, que utilizamos um microprocessador que põe a capacidade de processamento computacional em um único chip de circuito integrado.

O hardware, entretanto, é somente parte desse processo revolucionário. Dele fazem parte, também, o software, redes e regras/protocolos de comunicação.

A padronização de um único protocolo para a Internet aumentou significativamente o crescimento de usuários on-line. Isso motivou muitos tecnólogos a fazerem melhorias nas atuais formas de comunicação e criação de outras em alguns casos. Hoje falamos de IPv6 a fim de mitigarmos preocupações de endereçamento na rede e incrementar o uso da comunicação na internet. Com o tempo, foi-se abstraindo os conceitos e criou-se uma interface comum de acesso à Internet que se usou de facilidade de hardware e software: o uso de web browser.

Com isso, começava a popularização do uso da rede e a disseminação em massa de conhecimentos para a humanidade, mesmo que de forma não intencional era o começo da “migração” do modelo tradicional para um modelo de nuvem. O uso de tecnologias como virtualização de servidores, processamento de vetores, multiprocessamento simétrico e processamento paralelo massivo impulsionaram ainda mais a mudança.

4. Definição de Cloud Computing

Cloud computing é um modelo que habilita de forma simplificada o acesso on-demand a uma rede, a qual possui um pool de recursos computacionais configuráveis (por exemplo, redes, servidores, storages, aplicações e serviços) que podem ser rapidamente provisionados, configurados e liberados com um esforço de gerenciamento mínimo e automatizado. Esse modelo de cloud provê alta disponibilidade e é composto de cinco características essenciais, três modelos de serviços e quatro modelos de implantação.

5. Características Essenciais

On-demand self-service: um consumidor pode unilateralmente provisionar recursos computacionais, como servidor dns ou storage, de acordo com sua necessidade, sem a obrigatoriedade de interação humana com o provedor de serviço.

Acesso a rede: acesso a rede permitida por diferentes mecanismos e heterogeneidade de plataformas clientes: Mobiles, laptops e Pdas.

Pool de Recursos: os recursos computacionais de um provedor são agrupados, a fim de servirem múltiplos consumidores num modelo multiuso, com recursos físicos e virtuais diferentes, provisionados e reprovisionados de acordo com a demanda do cliente. Há um senso de localização independente; o cliente não sabe exatamente onde estão localizados os recursos aprovigionados e nem tem o controle e conhecimento desse local. Os recursos normalmente são: processador, memória, banda de rede e máquinas virtuais.

Rápida elasticidade: capacidade de rapidamente e elasticamente provisionar recursos, e em alguns casos automaticamente, para rapidamente aumentar os seus recursos e logo após o término voltar ao estado inicial. Para o usuário final, esta capacidade de crescer e provisionar mais recursos parece ser ilimitada e pode ser conseguida em qualquer quantidade e a qualquer tempo.

Serviço mensurado: automaticamente, sistemas em cloud controlam e otimizam recursos levando em conta a capacidade de medir em algum nível de abstração apropriado pelo cada tipo de serviço (p.ex: storage, processamento, banda de rede e usuários ativos). Recursos usados podem ser monitorados, controlados e reportados com transparência, tanto para o provedor quanto o consumidos dos serviços usados.

6. Modelos de serviços

A Computação em Nuvem está ligada a três áreas da TI: infraestrutura, plataforma e software. O que muitas vezes pode ser referenciado como formas, segmentos, estilos, tipos, níveis ou camadas de computação em nuvem.

Ao invés de se falar em diferentes funcionalidades providas, é melhor pensar em diferentes camadas, porque infraestrutura, plataforma e software logicamente

construídas e subsequentemente interligados dão um caráter mais arquitetural e de integração entre os níveis.

Como a entrega dos recursos de TI ou capacidades como um serviço é uma característica importante de Cloud Computing, as três camadas de arquitetura de Cloud Computing são:

1. Infraestrutura como serviço (IaaS);
2. Plataforma como Serviço (PaaS);
3. Software como Serviço (SaaS).

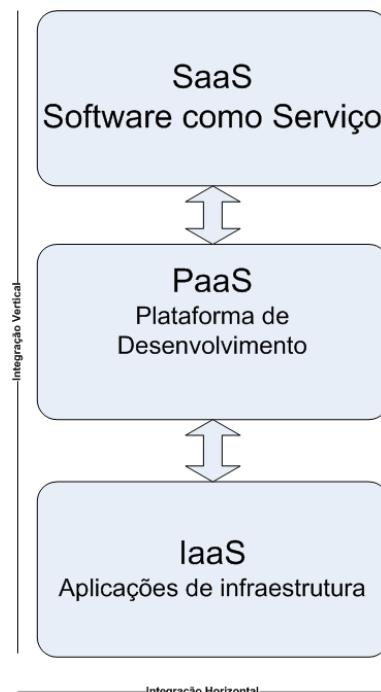


Figura 1: As três camadas de computação em nuvem: SaaS, PaaS, and IaaS

7. Infraestructure as a Service – IaaS (Infraestrutura como serviço)

Característica provida para o cliente que provisiona, processamento, storage, rede e outros recursos computacionais fundamentais onde o cliente está apto a implantar e rodar qualquer software, o que pode incluir sistemas operacionais e aplicações. O cliente não gerencia ou controla os recursos por trás dessa infraestrutura; contudo, tem controle sobre o sistema operacional, storage, aplicações e possibilidade de controle limitada a alguns tipos de componentes de rede como, por exemplo, firewall.

IaaS oferece recursos computacionais como processamento ou armazenamento, os quais podem ser obtidos como se fossem um serviço. Exemplos são a Amazon Web Services com seu Elastic Compute Cloud (EC2) para processamento e Simple Storage Service (S3) para armazenamento e Joyent o qual provê uma infraestrutura sob demanda escalável para rodar web sites aplicações web de interface ricas. Provedores de PaaS and SaaS podem recorrer a ofertas de IaaS baseadas em interfaces padronizadas.

Ao invés de vender infraestrutura de hardware, provedores de IaaS oferecem infraestrutura virtualizada como um serviço.

Foster *et al* (2008) denomina o nível de hardware puro, como computação, armazenamento e recursos de rede como camada de fábrica. Virtualizando, recursos de hardware que são abstraídos e encapsulados e consequentemente ser expostos à próxima camada e aos usuários finais através da padronização de interfaces como recursos unificados na forma de IaaS.

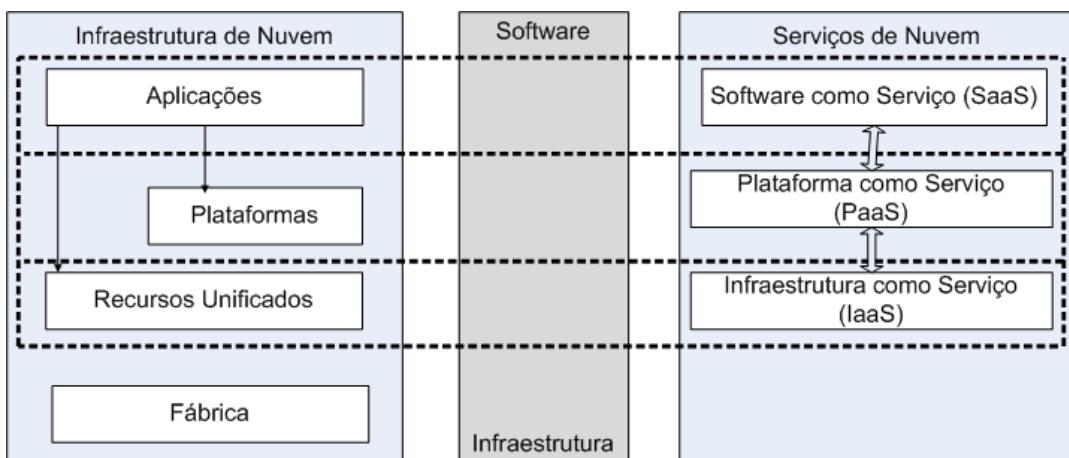


Figura 2: Arquitetura de nuvem relacionada com os Serviços de Nuvem

Já antes do advento da computação em nuvem, a infraestrutura já tinha sido colocada à disposição como um serviço por um bom tempo. Ela era referenciada como computação utilitária, o que é usada por muitos para denotar a camada de infraestrutura de computação em nuvem.

Entretanto, comparada aos recentes modelos de computação utilitária, IaaS denota sua evolução em direção ao suporte integrado dos três layers (IaaS, PaaS e SaaS) na nuvem.

Para as recentes ofertas de mercado de computação utilitária, ficou claro que para seus provedores terem sucesso, eles precisarão fornecer uma interface fácil de acessar, entender, programar e usar, como, por exemplo, uma API que habilita a fácil integração com a infraestrutura de clientes potenciais e desenvolvedores de aplicações SaaS. Os centros de dados de provedores de computação utilitária serão utilizados suficientemente somente se tiverem abaixo de si massas críticas de dados de clientes e provedores de SaaS.

Como uma consequência de requisitos para o fácil e abstrato acesso à camada física da nuvem, virtualização da camada física e plataformas programáveis para desenvolvedores emergem como principais características das nuvens.

8. Platform as a Service – PaaS (Plataforma como serviço)

Característica provida pela nuvem que possibilita ao usuário portar dentro da nuvem aplicações produzidas pelo cliente ou de terceiros, usando linguagens de programação e ferramentas suportadas pela nuvem. O cliente não gerencia ou mesmo controla os ativos que compõem essa infraestrutura; entretanto, tem controle sobre a aplicação implantada dentro da nuvem e configurações de ambiente dentro da mesma.

Plataformas são camadas de abstração entre aplicações de software (SaaS) e a infraestrutura virtualizada. As ofertas de PaaS são alvo dos desenvolvedores de software que podem escrever suas aplicações de acordo com as especificações de uma plataforma em particular sem a necessidade de se preocuparem com a camada subjacente de infraestrutura de hardware. (IaaS).

Os desenvolvedores fazem o upload de seus códigos para a plataforma, o que deve aumentar o alerta pela monitoração e gerenciamento automático quando o uso da aplicação cresce. As funcionalidades providas pelo PaaS podem cobrir todas as fases de desenvolvimento de software ou talvez especializada em uma dada área como o gerenciamento de conteúdo.

A camada PaaS da nuvem tem como base a padronização de interface da camada IaaS que virtualiza o acesso a recursos disponíveis, provê interfaces padronizadas e plataforma de desenvolvimento para a camada SaaS.

9. Software as a Service – SaaS (Software como serviço)

Serviço disponibilizado aos clientes que permite o uso de aplicações no provedor que rodam dentro da infraestrutura de nuvem. As aplicações estão acessíveis a qualquer cliente através de vários tipos de dispositivos, como uma interface web. O consumidor não gerencia ou controla o que há por baixo da infraestrutura como rede, servidores, sistemas operacionais, storage ou até mesmo algumas aplicações específicas.

SaaS é software de um provedor que é possuído, entregável e gerenciável por este de forma remota e negociado de forma pay-per-use. SaaS é a camada mais visível em cloud computing para usuários finais, porque são as aplicações de software que são acessadas e usadas.

Da perspectiva dos usuários, obter um software como serviço é mais motivador pelas vantagens de custo ao modelo de pagamento baseado em utilitário.

Os usuários mais comuns do SaaS normalmente não têm conhecimento nem controle sobre a camada abaixo, seja ela a imediatamente abaixo à plataforma como serviço ou os hardwares da infraestrutura como serviço. Entretanto, essas camadas abaixo são de grande relevância para o provedor de SaaS, porque elas são a base da infraestrutura, podendo ser vendidas e terceirizadas.

Como exemplo típico, cita-se que uma aplicação pode ser desenvolvida em uma plataforma qualquer e rodar em infraestrutura de terceiros. Ter-se uma plataforma e infraestrutura como serviço é um atrativo para os provedores de SaaS, pois pode aliviá-los de pesadas licenças de software e custos de investimentos em infraestrutura, além da flexibilidade. Isso também possibilita a corporação a focar em suas competências principais, que estão intimamente relacionadas ao negócio da empresa.

De acordo com os analistas de mercado, o crescimento pela inserção dentro do modelo de SaaS pelas empresas e a alta pressão de reduzir custos de TI são os maiores drivers pela alta demanda e crescimento do SaaS, também pelo crescimento por Cloud Computing nos próximos anos.

10. Modelos de Nuvens

Nuvem Privada: a infraestrutura de cloud é operada por uma organização e pode ser gerida pela própria organização ou por empresa terceira.

Nuvem comunitária: a infraestrutura de cloud é compartilhada por algumas organizações e abrange uma comunidade específica que tem os mesmos valores. (missão, requisitos de segurança, políticas e considerações de conformidade). Pode ser administrada pelas organizações ou por empresa terceira.

Nuvem Pública: a infraestrutura de cloud está disponível para o público geral ou um grupo de indústrias, ou é de propriedade de uma organização que vende os serviços da nuvem.

Nuvem Híbrida: a infraestrutura de cloud é composição de uma ou mais nuvens (privada, comunitária ou pública) que se mantêm como entidades únicas; entretanto, são ligadas pela padronização ou propriedade tecnológica, que permite portabilidade de aplicações e de dados.

11. Implicações para TI

Com este novo modelo, muitas implicações para uma empresa de TI podem surgir. Tais implicações derivam desta nova maneira de comunicação e estão direcionando mudanças na interatividade dos negócios. Hoje, os negócios precisam de respostas na velocidade da internet com novos serviços, funcionalidades diferentes e a quase obrigatoriedade de estar à frente de seu tempo e principalmente dos concorrentes.

Apesar disso, muitas corporações ainda não estão aptas a responder nesta velocidade e possuem o modelo tradicional de aquisições para compras de ativos de infraestrutura, o que lhes traz implicações negativas e compromete a agilidade de provisionamento. Muitas vezes, o equipamento está disponível; entretanto, há um processo burocrático de preparação e disponibilidade para uso; os recursos precisam estar prontos para uso com o aval técnico positivo. O que acontece é que se tem muitos processos que envolvem pessoal do storage, rede, segurança e algumas outras facilidades.

Normalmente essas dificuldades estão relacionadas a:

Planejamento de Capacidade: Para a maioria das organizações, não existe um planejamento de capacidade consistente, com planos de provisionamento de dados, serviços onde deve-se por tal equipamento ou mesmo como será o crescimento de tal aplicação. Isso é um impeditivo grande para implantação de um modelo baseado em cloud computing.

Equilíbrio das forças que o mercado demanda versus a utilização de ativos: A TI deve estar em sinergia em controlar seus gastos e ser responsável pelo negócio. A assertiva de que mais um servidor resolve o problema não colabora para o processo de construção desse novo modelo.

A empresa sempre quer algo rápido e consistente. O pessoal da área de negócio sempre vem com demandas para a área de TI que não têm orçamento aprovado, esperando que a TI consiga dar um jeito de produzir de qualquer maneira. As questões relacionadas com segurança da informação são um elo que deve ser bem fortalecido, principalmente por se tratar de uma nova área de negócio e tecnologia que tem muitos processos e lacunas não preenchidas, sendo passíveis de construções muitas vezes não

muitos sólidas e prejudiciais à empresa, exigindo cautela nos detalhes de serviços que serão prestados e nas transações executadas.

Outras implicações é que empresas dispostas a entrar nesse nicho e modificar a forma de orientação dentro do seu data center devem estar atentas à infraestrutura de TI, que deverá maximizar o gerenciamento e eficiência. E se gerenciar uma quantidade em massa de um data center não é uma das competências principais da empresa, ela deve delegar isso a uma empresa que:

Tem superioridade Econômica: Os grandes provedores de aplicações e serviços de TI, que compram tantos servidores, storages e outros muitos equipamentos de data center e que têm um enorme poder de negociação quando se fala de preço de hardware, licença de software e contratos de suporte.

Melhores Práticas: As maiores corporações têm investido não apenas em melhores processos, mas também investiram na construção de ferramentas de gerenciamento e administração que permitem a elas espalhar aplicações através de milhares de servidores de forma rápida.

Expertise em gerenciamento de capacidade dinâmica: Para grandes empresas, a produtividade de seus ativos é fundamental, assim como o custo de seus serviços é diretamente proporcional às despesas correntes do centro de dados. Quanto maior a produtividade que se pode tirar de cada metro quadrado de espaço, maior é a rentabilidade de um serviço. Por isso, é necessário uma monitoração de perto do consumo de recursos por cada aplicação dentro da infraestrutura disponibilizada.

12. Forças de influência

Quando se olha para um data center contemporâneo e moderno, não há como negar que eles são muito diferentes dos data centers dos de 10 ou 5 anos atrás. Certamente muitos o hardware existentes são de diferentes fabricantes, muitos têm heterogeneidade de servidores, como plataformas baixas e mainframes e aplicações. Nesta diversidade, provavelmente algum nível de organização deve-se ter notado ao longo do tempo. Entretanto, essa organização, para muitas empresas, é sinônimo de caos, pois somente alguns profissionais têm o *modus operandi* em mente, e não é suficientemente sustentável para um data center dentro do paradigma de cloud computing.

Algumas forças sustentam esse novo conceito e são elas que vão diferenciar a tradicional maneira de hosting do novo modelo preconizado dentro de cloud:

Comoditização: É lamentável que para muitos o termo *commodity* tenha uma conotação diferente; ela é a evolução de produtos feitos a mão, um sinal de avanço de produção e existência de mercados com liquidez – para encurtar, progresso econômico. Comoditização depende de uma vasta rede integrada de infraestrutura de compradores, distribuidores, fornecedores e montadores. Quando se vê uma commodity, pode-se ver por trás uma rede complexa para sustentar a produção, o que inclui produzi-la, distribuí-la, apoiá-la e entregá-la. O processo de comoditização em um mercado maduro gradualmente leva o foco da competição entre organizações da funcionalidade para a qualidade, serviços complementares e, por último, o preço.

Fazendo um paralelo com o data center dentro do modelo de cloud computing, pode-se entender que ele será a rede e as demandas estão atreladas à sua atomicidade. Para o mundo externo, ele é uma única manufatura que possibilitará com transparência a

execução de uma tarefa para produzir algum produto ou serviço através da padronização especializada por função.

Virtualização: Em computação, virtualização é um termo genérico utilizado para se referir à abstração dos recursos do computador. Uma definição seria: uma técnica para mascarar as características físicas dos recursos do computador de forma que outros sistemas, aplicações ou usuários finais possam interagir com tais recursos. Atrelado a esse conceito pode-se quebrar em mais duas forças: Miniaturização e Massificação. A primeira permite que em um servidor possa ter inúmeros sistemas operacionais virtualizados e massificados, ou seja, em várias outras equipamentos, referente ao segundo termo.

Portanto, o grau de abstração de uma solução de cloud computing depende também de quanto seu ambiente está virtualizado.

Independência de Aplicações e Sistemas Operacionais: A arquitetura que o ambiente de cloud provê hoje tem que estar habilitada para aceitar qualquer tipo de aplicação que o cliente deseja hospedar, já que ele não precisará acessar diretamente o hardware ou outros elementos internos da estrutura.

Liberdade de instalação de software ou hardware: Provisionamento tem que ser automático e sem burocracia. Não existe o processo de área e transações que é aplicado no modelo tradicional vigente nas empresas.

Integração: Integração é fundamental nessa nova abordagem, pois terá a necessidade de integrar nos vários níveis: software, hardware e middleware de forma global e esquecer a maneira transacional de integrar partições, sistemas em clustering, grids, barramentos e outros. O data center será um objeto atômico, não divisível para o mundo externo.

13. Modelos Tecnológicos

Alguns modelos tecnológicos devem ser padronizados para começo de construção de uma estrutura em nuvem. Os principais serão destacados nas subseções seguintes.

13.1. Modelo de Arquitetura

Neste modelo, deve ser definido de que forma será a integração entre as aplicações, serviços e outras nuvens externas. Deve-se mapear e fazer o projeto da infraestrutura comoditizada de hardware e software com definições macros de software e container dos softwares virtualizadores. Nesse momento, também terá que materializar a questão da disponibilidade e performance da solução através da topologia de rede e ligação dos servidores, questões de segurança como políticas de dados dentro e fora do ambiente produtivo. Enfim, este modelo será um arcabouço da solução com várias definições alto nível desde políticas escritas até garantia de segurança dos dados.

13.2. Modelo de Grid Computing

Computação em grade (do inglês Grid Computing) é um modelo computacional capaz de alcançar uma alta taxa de processamento, dividindo as tarefas entre diversas máquinas, podendo ser em rede local ou rede de longa distância, que formam uma máquina virtual. Esses processos podem ser executados no momento em que as

máquinas não estão sendo utilizadas pelo usuário, assim evitando o desperdício de processamento da máquina utilizada.

Nos anos 90, uma nova infraestrutura de computação distribuída foi proposta, visando auxiliar atividades de pesquisa e desenvolvimento científico. Vários modelos desta infraestrutura foram especificados; dentre eles, a Tecnologia em Grade, em analogia às redes elétricas (“power grids”) se propõe a apresentar-se ao usuário como um computador virtual, mascarando toda a infraestrutura distribuída, assim como a rede elétrica para uma pessoa que utiliza uma tomada, sem saber como a energia chega a ela. Seu objetivo era casar tecnologias heterogêneas (e muitas vezes geograficamente dispersas), formando um sistema robusto, dinâmico e escalável, onde se pudesse compartilhar processamento, espaço de armazenamento, dados, aplicações, dispositivos, entre outros.

Pesquisadores da área acreditam que a tecnologia de grades computacionais seja a evolução dos sistemas computacionais atuais, não sendo apenas um fenômeno tecnológico, mas também social, pois, num futuro próximo, reuniria recursos e pessoas de várias localidades, com várias atividades diferentes, numa mesma infraestrutura, possibilitando sua interação de uma forma antes impossível.

13.3. Modelo de Cluster

Um *cluster* é formado por um conjunto de computadores, que utiliza um tipo especial de sistema operacional classificado como sistema distribuído. Muitas vezes é construído a partir de computadores convencionais (*personal computers*), os quais são ligados em rede e comunicam-se através do sistema, trabalhando como se fossem uma única máquina de grande porte. Há diversos tipos de *cluster*. Um tipo famoso é o cluster da classe Beowulf, constituído por diversos nós escravos gerenciados por um só computador.

Quando se fala de cluster de um HD (Hard Disk), refere-se ao cruzamento de uma trilha com um setor formatado. Um HDD (hard disk drive) possui vários clusters que serão usados para armazenar dados de um determinado arquivo. Com essa divisão em trilhas e setores, é possível criar um endereçamento que visa facilitar o acesso a dados não contíguos, assim como o endereçamento de uma planilha de cálculos.

Existem vários tipos de *cluster*; no entanto, há alguns que são mais conhecidos, os quais são descritos a seguir:

Cluster de Alto Desempenho: Também conhecido como cluster de alta performance, ele funciona permitindo que ocorra uma grande carga de processamento com um volume alto de gigaflops em computadores comuns e utilizando sistema operacional gratuito, o que diminui seu custo.

Cluster de Alta Disponibilidade: São clusters cujos sistemas conseguem permanecer ativos por um longo período de tempo e em plena condição de uso; sendo assim, pode-se dizer que eles nunca param seu funcionamento. Além disso, conseguem detectar erros se protegendo de possíveis falhas.

Cluster para Balanceamento de Carga: Esse tipo de cluster tem como função controlar a distribuição equilibrada do processamento. Requer um monitoramento constante na sua comunicação e em seus mecanismos de redundância, pois, se ocorrer alguma falha, haverá uma interrupção no seu funcionamento.

13.4. Modelo de Armazenamento

A abstração usada para armazenar dados em sistemas computacionais é o arquivo. Para que esses arquivos sejam acessados, modificados e que sejam criados outros arquivos, é necessária uma estrutura que permita tais operações. Essa estrutura recebe o nome de sistema de arquivos.

A motivação básica dos sistemas distribuídos é o compartilhamento de recursos e, no âmbito dos sistemas de arquivos, o recurso a ser compartilhado são os dados sob a forma de arquivos.

Um Sistema de arquivos distribuído, ou SAD, é um sistema de arquivos no qual os arquivos nele armazenados estão espalhados em hardwares diferentes, interconectados através de uma rede. Eles têm vários aspectos semelhantes aos dos sistemas de arquivos centralizados, além de operações de manipulação de arquivos, preocupações com redundância, consistência, dentre outros atributos desejados de um sistema de arquivos.

O SAD deve prover transparência nos seguintes contextos:

- De acesso: aplicações que acessam os arquivos do SAD não devem estar cientes da localização física deles.
- De localização: todas as aplicações devem ter sempre a mesma visão do espaço de arquivos.
- De mobilidade: com a movimentação dos arquivos, nem programas do cliente e nem tabelas de administração precisam ser modificadas, de modo a refletir essa movimentação.
- De desempenho: programas clientes devem executar satisfatoriamente, independente de variação de carga do serviço de arquivos.
- De escalabilidade: o serviço pode ser expandido por crescimento horizontal, e não vertical, de modo a se adequar à carga demandada e à capacidade da rede disponível.

14. Desafios para a nuvem

Os maiores desafios para as empresas serão o armazenamento de dados seguro, acesso rápido à internet e padronização. Armazenar grandes volumes de dados está diretamente relacionado à privacidade, identidade, preferências de aplicações centralizadas em locais específicos, levanta muitas preocupações sobre a proteção dos dados. Essas preocupações são pertinentes ao framework legal que deve ser implementado para um ambiente de computação em nuvem.

Outra preocupação é a banda larga. Computação em nuvem é impraticável sem uma conexão de alta velocidade, caso tenha-se problema de alta velocidade, a computação em nuvem torna-se inviável para as massas acessarem os serviços com a qualidade desejada.

Além disso, padrões técnicos são necessários para que se tenha todo o arcabouço de computação em nuvem funcionando. Entretanto, eles não estão totalmente definidos, publicamente revistos ou ratificados por um órgão de supervisão. Não existe essa padronização nem da academia nem do mercado. Até mesmo consórcios formados por

grandes corporações necessitam transpor esse tipos de obstáculos e terem uma solução factível para que possam evoluir e gerar novos produtos que contribuam de alguma forma para a nuvem, sem essa definição precisa, ainda espera-se entregas em ritmo não tão acelerado.

Ao lado desses desafios discutidos anteriormente, a confiabilidade da computação em nuvem tem sido um ponto controverso nos encontros de tecnologia pelo mundo afora. Dada a disponibilidade pública do ambiente de nuvem, problemas que ocorrem na nuvem tendem a receber muita exposição pública; por isso, gerência e monitoração desse ambiente de TI são essenciais.

Em outubro de 2008, o Google publicou um artigo online que discute as lições aprendidas no armazenamento de milhões de clientes corporativos no modelo de computação em nuvem.

A métrica de disponibilidade do Google foi a média de uptime por usuário baseado nas taxas de erro do lado do servidor. Eles acreditavam que essa métrica de confiabilidade permitia uma verdadeira comparação com outras soluções. As medidas eram feitas para todas as requisições ao servidor para cada usuário, a cada momento do dia, onde até mesmo um pequeno milissegundo de delay era registrado. O Google analisou os dados coletados do ano anterior e descobriu que sua aplicação Gmail está disponível mais de 99,9% do tempo.

E alguém pode perguntar como 99,9% de métrica de confiabilidade pode se comparar ao modelo convencional usado para email das corporações. De acordo com uma pesquisa da empresa Radicati Group, companhias com soluções de email convencionais têm normalmente de 30 a 60 minutos de parada não programada e mais de 36 a 90 minutos de parada programada por mês, comparados aos 10 a 15 minutos do gmail. Baseado nessas análises o Gmail é duas vezes mais confiável que a solução GroupWise da Novell e quatro vezes mais confiável que a solução da Microsoft, o Exchange, e ambas soluções requerem uma infraestrutura montada específica para suas soluções de email, centralizada e com regras próprias de ambientes.

Baseado nesses dados, Google estava suficientemente confiante para anunciar publicamente em outubro de 2008 que 99,9% de nível de serviço oferecido aos clientes empresariais Premier também se estenderiam ao Google Calendar, Google Docs, Google Sites e Google Talk. Como milhões de negócios usam as Apps Google, ele fez uma série de compromissos para aperfeiçoar a comunicação com seus clientes durante qualquer interrupção e fazer com que todas as questões estejam visíveis e transparentes através de grupos abertos.

Uma vez que o próprio Google roda suas aplicações dentro das plataformas Apps Google, o compromisso que eles fizeram tem sustentabilidade, pois as usam em suas operações do dia a dia. O Google lidera a indústria na evolução do modelo de computação em nuvem para se tornar uma parte do que está sendo chamado de Web 3.0, a próxima geração de Internet.

15. Conclusão

Neste artigo, viu-se a importância de saber a origem do termo nuvem para então poder entrar no ambiente de computação em nuvem. Foram examinadas também questões históricas de evolução do hardware, software, redes e a forma de comunicação

e percebe-se o quanto ajudaram no crescimento da internet nos últimos anos, além de sua popularização com o acesso a browsers.

Após isso, foi tratado de forma mais pragmática o conceito profundo de computação em nuvem, explicitando-se seu modelo de negócio, modelo ser serviços, forças de influência e implicações para a área de TI.

Viu-se que a padronização, virtualização, processamento paralelo, distribuído e um modelo de negócio bem definido são essenciais para sair-se do campo teórico e buscar a aplicação efetiva em modelos que possam ser propagados para toda a sociedade da computação e administração de TI.

Além disso, verificou-se que existem mecanismos de cunho legais e técnicos que precisam ser condensados para que o mercado tenha um modelo único e denso dentro de cada empresa e possa ser dissipado não só entre governo, mas academias, meio privado e organismos internacionais.

Para finalizar, evidenciou-se que o Google com um arcabouço de gerência e monitoração bem definidos; conseguiu usufruir por completo o que a computação em nuvem pode proporcionar e mostrou que, se bem aplicado, esse conceito pode trazer mudanças significativas na maximização de recursos computacionais, além de superar a barreira daquela computação tradicional.

Referências

- ARMBRUST, M.; FOX, A.; GRIFFITH, R.; JOSEPH, A. D.; KATZ, R.; KOWINSKI, A.; LEE, G.; PATTERSON, D.; RABKIN, A.; STOICA, I.; ZAHARIA, M. Above the Clouds: A Berkeley View of Cloud Computing. EECS Department, University of California, Berkeley, fevereiro 2009.
- AMAZON. Elastic Compute Cloud. Disponível em: <<http://aws.amazon.com/ec2/>>. Acesso em: 18 fev. 2010.
- FOSTER, I. What is the Grid? A Three Point Checklist. Argonne National Laboratory & University of Chicago, julho 2002.
- MOHAMED, A. A history of Cloud Computing. ComputerWeekly.com, março 2009.
- SUN MICROSYSTEMS, INC. Introduction to Cloud Computing Architecture. White Paper, 1^a edição, junho 2009a.
- TANENBAUM, A. S.; STEEN, M. Distributed Systems Principles and Paradigms. Nova Jersey: Prentice Hall. 2002.
- WANG, L et al. Cloud Computing: a Perspective Study. RIT Digital Media Library, 2008