

DANIEL NORIO TAKASU REBELO
LUCAS ARTHUR FELGUEIRAS
LUIZ GUSTAVO DOS SANTOS
VICTOR FRANÇA FERREIRA

COMPUTAÇÃO EM NUVEM: INTRODUÇÃO

São Paulo
2018

**DANIEL NORIO TAKASU REBELO
LUCAS ARTHUR FELGUEIRAS
LUIZ GUSTAVO DOS SANTOS
VICTOR FRANÇA FERREIRA**

COMPUTAÇÃO EM NUVEM: INTRODUÇÃO

Trabalho apresentado à Escola Politécnica
da Universidade de São Paulo para a
disciplina de Engenharia de Sistemas.

São Paulo
2018

**DANIEL NORIO TAKASU REBELO
LUCAS ARTHUR FELGUEIRAS
LUIZ GUSTAVO DOS SANTOS
VICTOR FRANÇA FERREIRA**

COMPUTAÇÃO EM NUVEM: INTRODUÇÃO

Trabalho apresentado à Escola Politécnica
da Universidade de São Paulo para a
disciplina de Engenharia de Sistemas.

Área de Concentração:
Engenharia de Sistemas

Orientador:
Prof^ª. Dr^a. Selma Shin Shimizu
Melnikoff

RESUMO

A tecnologia de computação em nuvem é uma tendência para as novas aplicações do mundo. Grandes empresas estão migrando suas estruturas para a nuvem, outras constroem ambientes para abrigar essas novas aplicações. O objetivo do trabalho consiste em compreender melhor como funciona essas tecnologias, sua motivação e as principais opções disponíveis no mercado para uso.

Palavras-Chave – Nuvem, Arquitetura, Computação, Engenharia, Programa.

ABSTRACT

Cloud computing technology is a trend for new applications around the world. Large companies are migrating their structures to the cloud, others build environments to house these new applications. The objective of the work is to better understand how these technologies work, their motivation and the main options available in the market to use.

Keywords – Cloud, Architecture, Computing, Enginnering, Software.

LISTA DE FIGURAS

| | | |
|----|---|----|
| 1 | Arquitetura Básica de Computação em Nuvem(1) | 10 |
| 2 | Histórico do Amazon EC2(2) | 12 |
| 3 | Algumas soluções de Computação em Nuvem(3) | 13 |
| 4 | Adesão de serviços de computação em nuvem segundo a RightScale(4) . . . | 13 |
| 5 | Arquitetura de referência para Cloud Computing | 21 |
| 6 | Serviços da Amazon Web Services(5) | 31 |
| 7 | Console da Amazon Web Services | 32 |
| 8 | Serviços do Google Cloud(6) | 33 |
| 9 | Exemplo do <i>Dashboard</i> do Google Cloud com seus recursos | 34 |
| 10 | Arquitetura Básica do Heroku(7) | 37 |
| 11 | Ligação entre <i>router</i> e <i>dynos</i> do Heroku(7) | 38 |
| 12 | Exemplo do <i>Dashboard</i> do Heroku com seus recursos | 39 |
| 13 | Regiões disponibilizadas pela Azure | 41 |
| 14 | Camadas de Segurança da Azure | 42 |

LISTA DE TABELAS

SUMÁRIO

| | |
|--|-----------|
| Parte I: INTRODUÇÃO | 9 |
| 1 Objetivo | 10 |
| 2 Histórico | 11 |
| 2.1 Contexto | 11 |
| 2.2 Problemas | 11 |
| 2.3 Primórdios | 11 |
| 2.4 Impactos | 12 |
| Parte II: ASPECTOS CONCEITUAIS | 14 |
| 3 Conceitos Básicos | 15 |
| 3.1 Definição de Cloud Computing | 15 |
| 3.2 Características básicas de Cloud Computing | 15 |
| 4 Modelos para Cloud Computing | 17 |
| 4.1 Modelos de serviços | 17 |
| 4.2 Modelos de implantação | 18 |
| 5 Arquitetura | 20 |
| 5.1 Arquitetura de referência | 20 |
| 5.1.1 Cloud Consumer | 21 |
| 5.1.2 Cloud Auditor | 21 |
| 5.1.3 Cloud Provider | 22 |
| 5.1.4 Cloud Broker | 22 |

| | | |
|--------------------------------------|---|-----------|
| 5.1.5 | Cloud Carrier | 22 |
| Parte III: ASPECTOS TÉCNICOS | | 23 |
| 6 | Estrutura | 24 |
| 6.1 | Hypervisor | 24 |
| 6.1.1 | Mapeador de páginas virtuais | 24 |
| 6.1.2 | Agendador | 25 |
| 7 | Consideração técnicas sobre requisitos | 26 |
| 7.1 | Disponibilidade | 26 |
| 7.1.1 | Stateless | 26 |
| 7.1.2 | Instâncias reservas | 27 |
| 7.1.3 | Dados em zonas distintas | 27 |
| 7.1.4 | Degradação silenciosa | 27 |
| 7.2 | Performance | 27 |
| 7.3 | Segurança | 28 |
| 7.3.1 | Multi-tenancy (multilocação) | 28 |
| 7.3.2 | Ataques de multilocação | 28 |
| Parte IV: SOLUÇÕES DO MERCADO | | 30 |
| 8 | Amazon Web Services | 31 |
| 8.1 | Introdução | 31 |
| 8.2 | Funcionamento básico | 32 |
| 8.2.1 | EC2 | 32 |
| 8.2.2 | S3 | 32 |
| 8.2.3 | RDS | 32 |
| 8.2.4 | Route 53 | 32 |

| | | |
|-----------|------------------------------------|-----------|
| 8.3 | Outros Serviços | 32 |
| 8.4 | Conclusão | 32 |
| 9 | Google Cloud | 33 |
| 9.1 | Introdução | 33 |
| 9.2 | Funcionamento básico | 34 |
| 9.2.1 | Cloud Functions | 34 |
| 9.2.2 | App Engine | 35 |
| 9.2.3 | Kubernetes Engine | 35 |
| 9.2.4 | Compute Engine | 35 |
| 9.3 | Outros serviços | 36 |
| 9.4 | Conclusão | 36 |
| 10 | Heroku | 37 |
| 10.1 | Introdução | 37 |
| 10.2 | Funcionamento básico | 37 |
| 10.3 | Tarifação | 38 |
| 10.4 | <i>Containers</i> | 39 |
| 10.5 | Conclusão | 40 |
| 11 | Azure | 41 |
| 11.1 | Introdução | 41 |
| 11.2 | Vantagens e diferenciais | 42 |
| 11.3 | Desvantagens | 43 |
| 11.4 | Conclusão | 43 |
| | Referências | 44 |

PARTE I

INTRODUÇÃO

1 OBJETIVO

First to mind when asked what 'the cloud' is, a majority respond it's either an actual cloud, the sky, or something related to weather.

-- Citrix Cloud Survey Guide(8)

Com o desenvolvimento da computação, novos programas foram criados, cada vez mais consumindo recursos e sendo hospedados em máquinas pessoais ou dedicadas, porém sem um gerenciamento inteligente de toda a infraestrutura. Além disso, houve a nítida migração de programas **distribuídos** pela internet para programas que **rodem** na internet.

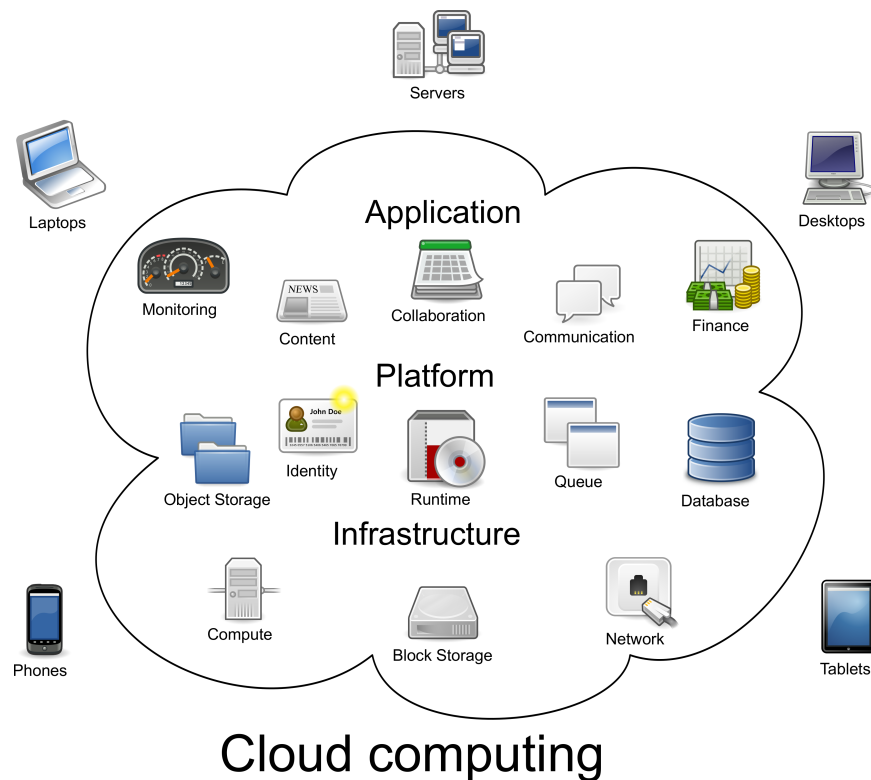


Figura 1: Arquitetura Básica de Computação em Nuvem(1)

Essas mudanças de paradigma motivaram a criação e consolidação do que conhecemos hoje como Computação em Nuvem, área fortalecida e necessária no desenvolvimento de programas modernos. Ao longo dos próximos capítulos, haverá a explanação em detalhes de como funciona uma arquitetura em nuvem básica, seus desafios técnicos e as principais soluções existentes no mercado.

2 HISTÓRICO

I don't need a hard disk in my computer if I can get to the server faster... carrying around these non-connected computers is byzantine by comparison.

-- Steve Jobs(8)

2.1 Contexto

A computação, em meados dos anos 90, consistia na evolução do computador pessoal e na consolidação da internet pelo mundo todo. A essência das aplicações construídas estava na distribuição de programas pela internet. Com o advento de aplicações web, torna-se necessário hospedar essas aplicações em servidores. Geralmente, era uma máquina física dedicada ao serviço, onde a aplicação ficaria hospedada, recebendo suas requisições.

2.2 Problemas

No contexto da época, esses servidores dedicados falhavam com certa frequência. Em casos de falha, o administrador do sistema precisava colocar uma nova máquina para atender a demanda, enquanto entendia a falha de *software* ocorrida. Porém, toda a provisão de novos servidores para gerar redundância e tolerância a falhas aumentava a probabilidade de máquinas físicas falharem, gerando maiores problemas e dificuldades para manutenção.

2.3 Primórdios

Em 2006, a Amazon, percebendo a oportunidade de diversificar seu negócio atuando com provisionamento de infraestrutura, gerenciando máquinas ociosas de acordo com os horários possíveis e cuidando de toda a manutenção trabalhosa pelo lado do desenvolvedor, formulou e lançou a primeira versão do Amazon Elastic Compute Cloud (EC2), o primeiro produto voltado para a computação em nuvem, com o propósito de atender essa necessidade encontrada e vender infraestrutura para o usuário final.

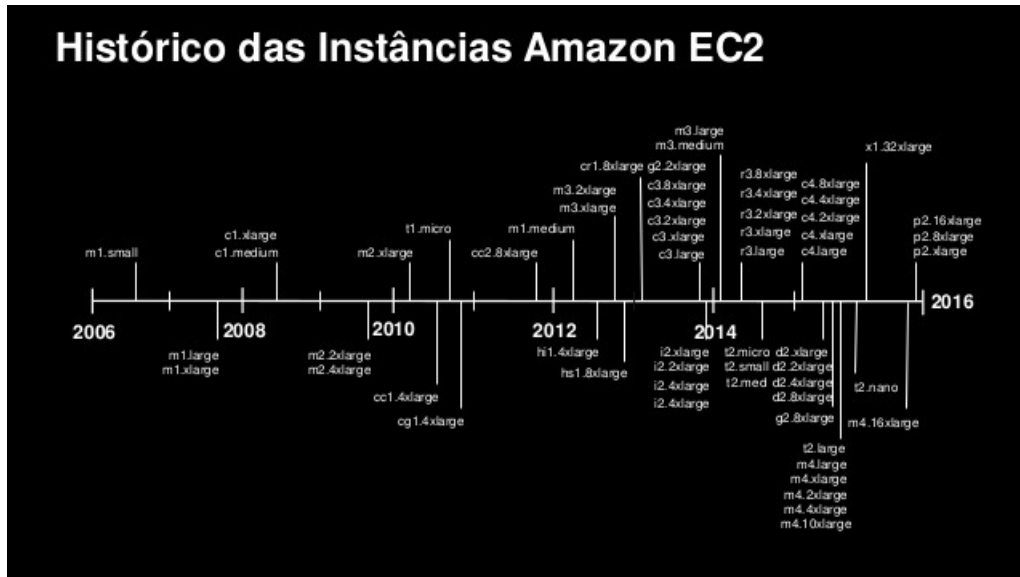


Figura 2: Histórico do Amazon EC2(2)

A postagem original no site da Amazon anunciava o serviço (em inglês)(10):

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Just as Amazon Simple Storage Service (Amazon S3) enables storage in the cloud, Amazon EC2 enables “compute” in the cloud. Amazon EC2’s simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon’s proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use.

2.4 Impactos

Essencialmente, a Amazon comercializou o conceito de infraestrutura como serviço (*IaaS*). A revolução iniciada na época permitiu que muitas aplicações hospedadas em servidores físicos fossem migrados para a infraestrutura da Amazon. A cobrança realizada por hora de instância virtual carregada era bem inferior aos custos (tanto reais quanto de trabalho) de manter uma infraestrutura física equivalente, fora o fato de que a ociosidade foi drasticamente reduzida, pois o gerenciamento inteligente do EC2 permitia que uma mesma máquina atendesse dois serviços em horários distintos.

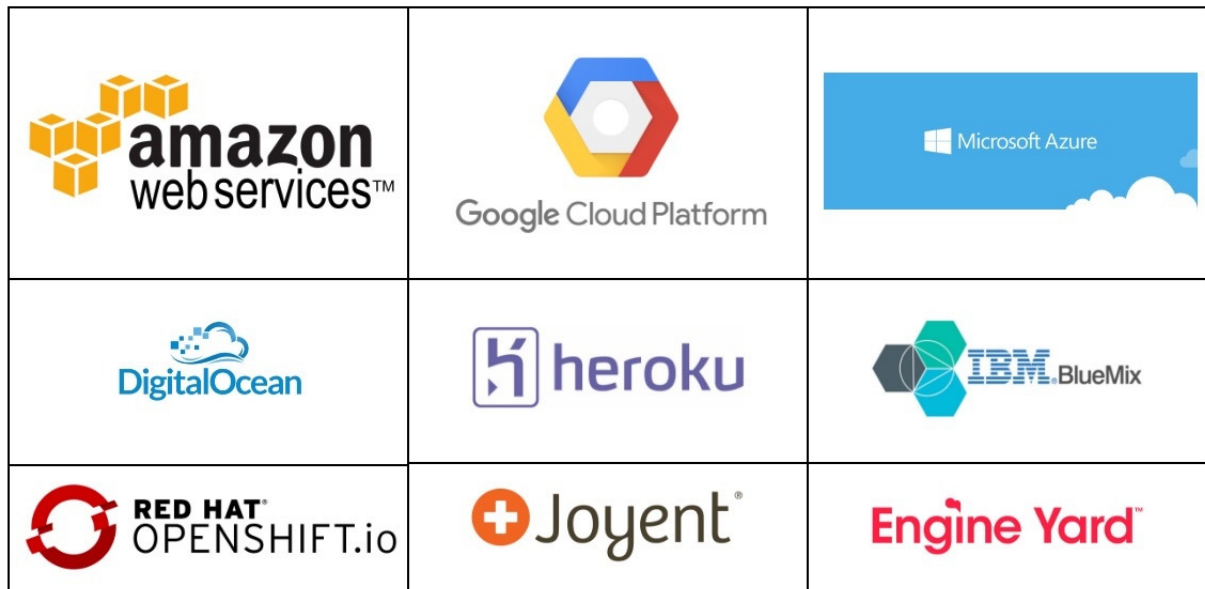
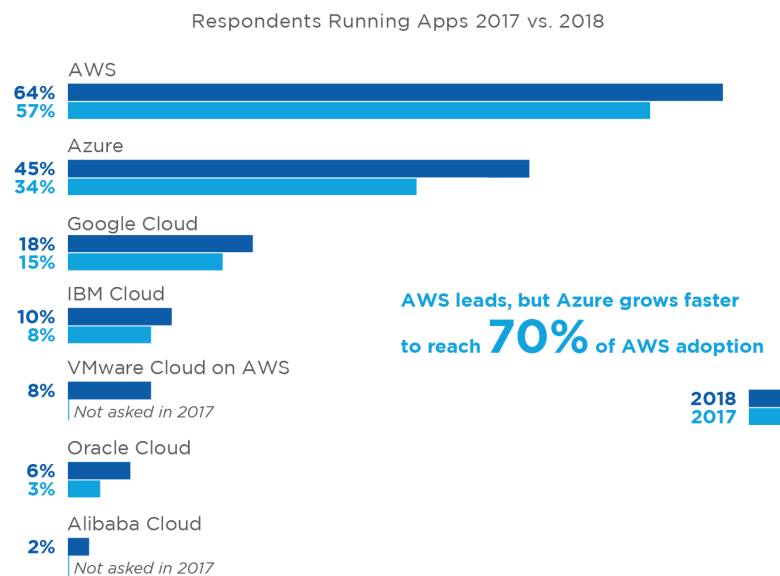


Figura 3: Algumas soluções de Computação em Nuvem(3)

Com o trunfo do EC2, outras soluções ganharam força, merecendo destaque para: **Salesforce Heroku**, **Google Cloud** e **Microsoft Azure**, que serão exploradas adiante. Atualmente, a adoção, segundo o relatório anual de Computação em Nuvem da RightScale(4), está em grande parte para a AWS:



Source: RightScale 2018 State of the Cloud Report

Figura 4: Adesão de serviços de computação em nuvem segundo a RightScale(4)

PARTE II

ASPECTOS CONCEITUAIS

3 CONCEITOS BÁSICOS

Este capítulo introduz as definições e os conceitos básicos quando se trata de Cloud Computing.

3.1 Definição de Cloud Computing

Computação em Nuvem (do Inglês *Cloud Computing*) é segundo a NIST (*National Institute of Standards and Technology*) um modelo para prover um acesso de rede a um grupo compartilhado (*shared pool*) de recursos computacionais (recursos como capacidade de rede, servidores, armazenamento, aplicações e serviços) de forma ubíqua, prática e sob demanda.

Tal acesso deve ser rapidamente provisionado e lançado com o mínimo de gerenciamento e interação com o provedor de serviços por parte da aplicação.

O modelo de computação em nuvem deve possuir algumas características básicas, que estão descritas na seção seguinte.

3.2 Características básicas de Cloud Computing

A NIST define 5 características essenciais do modelo de Cloud Computing:

- Serviço sobre demanda: Um consumidor pode provisionar unilateralmente capacidades computacionais, como tempo de servidor e armazenamento de rede, conforme for necessário, sem qualquer interação humana com o provedor de serviço;
- Agrupamento de recursos: Os recursos de computação do provedor são agrupados para atender a vários consumidores usando um modelo "multi inquilino", com diferentes recursos físicos e virtuais dinamicamente atribuídos e reatribuídos de acordo com a demanda do consumidor. Existe uma sensação de independência de loca-

lização pois o cliente geralmente não tem controle ou conhecimento sobre a localização exata dos recursos fornecidos, mas pode ser capaz de especificar a localização em um nível de abstração (por exemplo, país, estado ou datacenter). Exemplos de recursos incluem armazenamento, processamento, memória e largura de banda de rede.

- **Amplio acesso à rede:** Os recursos estão disponíveis na rede e são acessados por meio de mecanismos que promovam o uso por plataformas heterogêneas por clientes em diversos dispositivos (por exemplo, telefones celulares, tablets, laptops e estações de trabalho). Esta característica promove o conceito de computação ubíqua, isto é, em toda parte, onipresente.
- **Elasticidade rápida:** Os recursos podem ser provisionados e liberados elasticamente, em alguns casos automaticamente, proporcionando uma escalabilidade crescente ou decrescente conforme a demanda. Os recursos disponíveis normalmente aparentam ser ilimitados para o consumidor, podendo ser requisitados em qualquer quantidade e a qualquer momento.
- **Serviço mensurável:** Os sistemas em nuvem controlam e otimizam automaticamente o uso de recursos, aproveitando-se de uma capacidade de medição em um nível de abstração apropriado ao tipo de serviço (por exemplo, armazenamento, processamento, largura de banda e contas de usuário ativas). O uso de recursos pode ser monitorado, controlado e reportado, gerando transparência tanto para o fornecedor e consumidor do serviço utilizado.

4 MODELOS PARA CLOUD COMPUTING

Este capítulo introduz brevemente os modelos de serviço de cloud computing que podem ser adotados por um provedor e os modelos de *deployment*, nas secções seguintes.

4.1 Modelos de serviços

A teoria por trás dos serviços de computação em nuvem abrange três elementos principais: software, plataforma e infraestrutura. Temos os seguintes modelos de serviço:

- **SaaS, Software as a Service** ("Software como um serviço")

Neste modelo é oferecido ao consumidor o uso de aplicações de um provedor que rodam sobre uma infraestrutura em nuvem. Estas aplicações são acessíveis a partir de vários dispositivos clientes por meio de uma interface simples, como um navegador da web, ou uma interface de por meio de um programa (mobile ou desktop).

O consumidor não gerencia ou controla a infraestrutura de nuvem por trás da aplicação, incluindo rede, servidores, sistemas operacionais, armazenamento ou capacidade da aplicação individual. São disponibilizadas apenas configurações do aplicativo específicas para aquele usuário individual.

Alguns exemplos de SaaS são serviços de *webmail*, *streaming* de vídeos, conversão de arquivos e trabalho colaborativo com arquivos. Este modelo não será mais detalhado neste trabalho.

- **PaaS, Platform as a Service** ("Plataforma como um serviço")

A capacidade fornecida ao consumidor é de implantar na nuvem aplicações criadas por meio de linguagens de programação, bibliotecas, serviços e ferramentas suportadas pelo provedor. Tais aplicações podem ser criadas pelo próprio consumidor, ou consumidas por este.

Assim como no modelo de SaaS, o consumidor geralmente não gerencia ou têm controle sobre a infraestrutura de nuvem por trás, incluindo rede, servidores, sistemas operacionais, ou armazenamento. Porém o consumidor as tem controle sobre os aplicativos implantados e possivelmente sobre definições de configuração para o ambiente de hospedagem do aplicativo.

Como exemplos de provedores no mercado temos *IBM Bluemix*, *Heroku*, e *Windows Azure Cloud*.

- **IaaS, Infrastructure as a Service** ("Infraestrutura como um serviço")

A capacidade oferecida ao consumidor é provisionar processamento, armazenamento, redes e outros recursos fundamentais de computação onde o consumidor é capaz de implantar e executar software arbitrário, incluindo-se sistemas e aplicações. O consumidor não gerencia nem controla a infraestrutura de nuvem por trás, mas tem controle sobre sistemas operacionais, armazenamento e aplicativos implantados. Possivelmente possui também um controle limitado de componentes de rede (por exemplo, firewalls de host). Geralmente acompanha serviços de máquinas virtualizadas.

Alguns exemplos de provedores no mercado são *Amazon Web Services*, *Microsoft Azure*, *Google Cloud* e *VMware Cloud on AWS*.

4.2 Modelos de implantação

Os modelos de implantação (*deployment models*) a seguir delimitam as formas possíveis de tarifação dos serviços, as possíveis localizações da infraestrutura física e o público de escopo. São eles:

- **Nuvem privada:** A infraestrutura de nuvem é disponibilizada para uso exclusivo por uma única organização, compreendendo vários consumidores (*business units*). Pode ser propriedade e gerenciada pela própria organização, um terceiro ou alguma combinação de ambos, e pode existir dentro ou fora das instalações da organização.
- **Nuvem comunitária:** A infraestrutura em nuvem é de uso exclusivo por uma comunidade de consumidores de organizações que compartilham mesmos interesses (por exemplo, missão, requisitos de segurança, políticas e considerações de conformidade). Pode ser de propriedade e administrada por uma ou mais organizações da

comunidade, um terceiro, ou alguma combinação de ambos, e pode existir dentro ou fora das instalações.

- **Nuvem pública:** A infraestrutura de nuvem é para o uso aberto pelo público em geral. Pode ser propriedade ou gerenciada por uma organização comercial, acadêmica ou governamental, ou alguma combinação destes. Existe dentro das instalações do provedor de nuvem.
- **Nuvem híbrida:** A infraestrutura de nuvem é uma composição de duas ou mais infraestruturas de nuvens distintas (privadas, comunitárias ou públicas) que permanecem como entidades únicas, mas unidas por tecnologia padronizada ou proprietária permitindo portabilidade de dados e aplicações (por exemplo, *cloud bursting* para balanceamento de carga entre nuvens).

5 ARQUITETURA

Este capítulo apresenta uma arquitetura de referência para cloud computing na primeira seção. Na seção final apresenta boas práticas e métodos para arquitetar uma aplicação que deve ser colocada em um ambiente de nuvem.

5.1 Arquitetura de referência

O objetivo da arquitetura de referência apresentada a seguir é de providenciar uma taxonomia simples e sem ambiguidade para os três modelos de serviço:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)

Tal arquitetura deve disponibilizar uma visão unificada das cinco características essenciais da NIST:

- Serviço sobre demanda
- Amplo acesso à rede
- Agrupamento de recursos
- Elasticidade rápida (escalabilidade)
- Serviço mensurável

Após pesquisar informações disponibilizadas por provedores, instituições de pesquisa e de consultoria na área achamos interessante a seguinte arquitetura de referência:

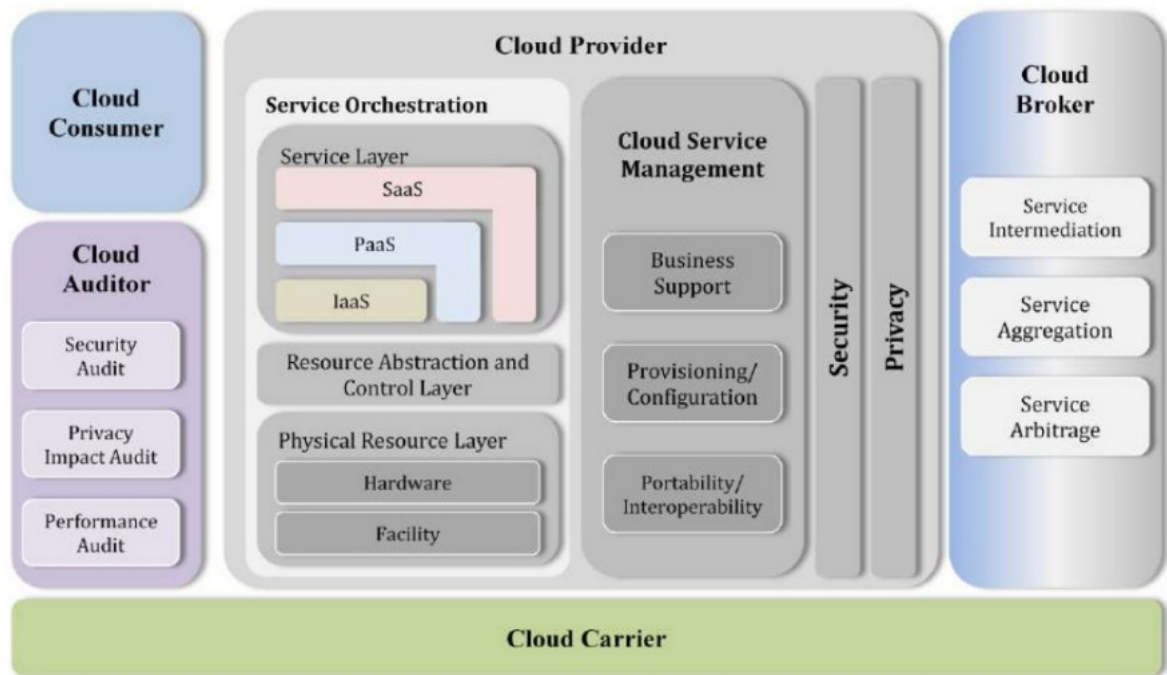


Figura 5: Arquitetura de referência para Cloud Computing

5.1.1 Cloud Consumer

Representa classes de consumidores para cada modelo possível de serviço.

No caso de SaaS são diversos departamentos operacionais de uma empresa, consumindo serviços de ERP, CRM, vendas, finanças, RH, redes sociais, webmail etc.

No caso de PaaS são os setores de análise de dados e desenvolvimento, que consomem serviços de BI, banco de dados, *deployment* de aplicações, integração de sistemas e desenvolvimento e testes.

Por fim, os consumidores de IaaS são os setores da empresa responsável pela infraestrutura e IT, que consomem serviços de backup e recuperação, gerenciamento de serviços, armazenamento, hospedagem de plataformas etc.

5.1.2 Cloud Auditor

Cloud Auditor é o componente responsável por relatar e registrar as ações relativas a segurança, privacidade e performance da nuvem, tanto para fins de transparência para os consumidores quanto de regulamentação governamental para rastrear atividades suspeitas e irregulares. Pode ser tanto um órgão governamental ou uma entidade fiscalizadora global na Internet.

5.1.3 Cloud Provider

Cloud Provider é o provedor de nuvem como a *Amazon Web Services*, *Microsoft Azure*, *Netflix*, *Youtube*, dentre outros. Dependendo das categorias de modelo de serviço em que o provedor se encaixa, ele poderá ter camadas de serviço de SaaS, PaaS ou IaaS.

As principais atividades do Cloud Provider são a implantação de serviços, a orquestração de serviços, o gerenciamento de serviços, a segurança e por fim, a privacidade. Deve gerenciar e abstrair toda lógica de controle e a camada física de recursos de hardware, bem como as instalações e locais físicos.

5.1.4 Cloud Broker

O Cloud Broker é responsável por uma espécie de serviço de corretagem. Este ator no ecossistema de nuvem é responsável por oferecer serviços de proteção jurídica durante a negociação de contratos com provedores de nuvem.

Oferece também serviços de diminuição de custos com o provedor, agregando diversos serviços a fim de obter o melhor custo-benefício.

Por fim, pode oferecer serviços diferenciados, como um serviço adicional de segurança, para proteger o consumidor de vazamento de dados na nuvem.

5.1.5 Cloud Carrier

O Cloud Carrier são todas entidades responsáveis por disponibilizar a conexão e transporte de serviços na nuvem entre consumidores e provedores (Companhias de redes, telecomunicações, dispositivos de acesso e IoT).

PARTE III

ASPECTOS TÉCNICOS

6 ESTRUTURA

Este capítulo descreve alguns dos mecanismos de baixo nível que os provedores de nuvem utilizam para prover seus serviços. No modelo de IaaS por exemplo a nuvem provisiona ao cliente uma máquina virtual carregada com uma imagem. O conceito de virtualização é antigo porém é extremamente importante no contexto de cloud computing. Nas secções seguintes descreveremos os mecanismos principais.

6.1 Hypervisor

Um *hypervisor* é uma espécie de sistema operacional usado para criar e gerenciar máquinas virtuais. Como cada máquina tem seu próprio sistema operacional, uma aplicação de um cliente é gerenciada por duas camadas de sistema operacional: o *hypervisor* e o sistema operacional da máquina virtual. O *hypervisor* administra o SO da máquina virtual e o SO da máquina virtual gerencia a aplicação do cliente.

Os principais serviços usados pelo *hypervisor* para suportar o gerenciamento de máquinas virtuais são o mapeador de páginas virtuais (*virtual page mapper*) e um agendador (*scheduler*). Um *hypervisor* também possui outros serviços, mas não o discutiremos neste trabalho.

6.1.1 Mapeador de páginas virtuais

Usar memória virtual como um mecanismo de virtualização envolve adicionar outro nível de indireção. Os processadores modernos contêm muitas otimizações para tornar esse processo mais eficiente.

Um aplicativo de usuário gera a próxima instrução com seu endereço de destino. Este endereço de destino está dentro da VM na qual o aplicativo está sendo executado. A tabela de páginas da máquina mapeia esse endereço de destino para um endereço dentro da máquina virtual com base no endereço de destino como no caso normal.

O que é novo neste processo para nuvem é que o endereço dentro do VM é convertido em um endereço físico por uma tabela de páginas do *hypervisor* que gerencia as VMs atuais.

6.1.2 Agendador

O agendador do *hypervisor* opera como qualquer agendador de sistema operacional. Sempre que o *hypervisor* tiver o controle de execução ele decide para qual máquina virtual passará o controle. Um simples agendamento de pelo algoritmo de *round robin* atribui o processador a cada VM, mas muitas outras opções de algoritmos são possíveis.

A escolha do algoritmo de agendamento correto requer que você faça suposições sobre as características de demanda das diferentes VMs hospedadas em um único servidor. Uma área de pesquisa é a aplicação de algoritmos de programação em tempo real para *hypervisors*. Agendadores em tempo real é apropriado para o uso de virtualização dentro de sistemas embarcados, mas não necessariamente na nuvem ainda.

7 CONSIDERAÇÃO TÉCNICAS SOBRE REQUISITOS

Neste capítulo apresentamos algumas considerações técnicas e táticas para cumprir com êxito três principais requisitos não funcionais claros para Cloud Computing: disponibilidade, performance e segurança.

7.1 Disponibilidade

Assume-se que a nuvem sempre estará disponível, 24 horas por dia, 7 dias por semana. Porém nenhum de seus componentes é a prova de falhas. O host físico de uma máquina virtual pode falhar, bem como a rede virtual com menor probabilidade, ao sofrer um ataque de negação de serviço (DDoS), por exemplo.

Os provedores de serviço na nuvem fornecem dados aos consumidores sobre a porcentagem de tempo que a rede pode ficar disponível. Apesar de ser um número pequeno (para alguns planos de serviço da AWS variam entre 0,05% e 0,001%), é de responsabilidade do arquiteto do sistema que consumidor projetá-lo de tal forma que seja incluído um plano de contingência para essa indisponibilidade.

Discutiremos algumas táticas a seguir:

7.1.1 Stateless

Arquitelar sistemas *stateless* (Sem estado) permite que qualquer instância de serviço possa servir igualmente uma requisição do cliente. Assim, se um servidor falha as requisições são redirecionadas para outra instância de serviço.

7.1.2 Instâncias reservas

O sistema deve incluir instâncias reservas da aplicação, para onde o tráfego de rede será redirecionado caso as instâncias principais venham a falhar.

7.1.3 Dados em zonas distintas

Um provedor na nuvem pode disponibilizar zonas distintas, que correspondem a *data centers* distintos. O sistema então pode implementar uma redundância com diversas cópias dos dados espalhados por essas zonas. Se a rede de uma zona cair, uma zona que está ociosa ou a mais próxima ativa assume seu lugar.

7.1.4 Degradação silenciosa

Os princípios para lidar com falhas em aplicações se baseiam em táticas de renovação ou remoção de serviços:

- **Restauração rápida:** Crie *timeouts* curtos para testar periodicamente o status de cada componente, restabelencendo-o antes que o sistema sinta essa falha.
- **Downgrades:** Cada feature é projetada com capacidade de ser degradada ou reduzida para uma representação de menor qualidade. Ex.: *streaming* de vídeos.
- **Remoção de features:** Se a rede estiver mais lenta que o normal e a *feature* não for crítica, ela deve ser removida da página.

7.2 Performance

Aplicações com boas performances computacionais são aquelas que respondem de forma instantânea ou com o tempo que lhe foi especificado, dentro de um limite tolerável que não prejudique uma atividade do usuário. Essa capacidade em uma máquina virtual ou física vai depender do que está sendo executado. Assim, cada aplicação deve monitorar a si mesma para determinar os recursos que recebe e quais irá precisar em um dado momento.

Uma vantagem da computação em nuvem é o provisionamento de *hosts* elásticos, isto é, que permitem que recursos adicionais sejam adquiridos conforme for necessário. Por meio do mecanismo de clusterização uma máquina virtual adicional pode prover essa capacidade

computacional extra. Alguns provedores em nuvem alocam recursos automaticamente, enquanto outros consideram a alocação responsabilidade do usuário.

Mesmo se o provedor faça essa alocação de recursos de forma automática, a aplicação deve ser arquitetada de forma a estar ciente do uso de recursos atual e futuro, em qualquer instante de tempo.

O provedor não pode fazer nada mais do que usar alguns algoritmos genéricos para determinar se há necessidade de alocar ou liberar mais recursos. Uma aplicação precisa ter um modelo de seu comportamento e estar preparada para fazer uma alocação sozinha de recursos. No pior caso, a aplicação pode comparar suas previsões com a do provedor para ter uma noção do que pode ocorrer.

7.3 Segurança

Segurança é uma temática que geralmente abrange aspectos técnicos e não-técnicos. Muitos esquemas de segurança não podem ser provados teoricamente como modelos seguros, mas empiricamente trazem resultados.

Aspectos não-técnicos são, por exemplo, o nível de confiança que é depositada no provedor de nuvem, as seguranças físicas que o servidor utiliza, e os padrões de governança implementado entre seus funcionários. Nesta seção focaremos apenas nos aspectos técnicos.

7.3.1 Multi-tenancy (multilocalização)

Multilocalização significa que seu aplicativo está utilizando uma máquina virtual em um computador físico que hospeda várias máquinas virtuais. Se um dos outros inquilinos (*tenancies*) da sua máquina for mal-intencionado ele pode atacar sua aplicação. Este é o único elemento extra introduzido pela arquitetura em nuvem que introduz problemas exclusivos de segurança. Como os outros tipos de ataques também ocorrem em outras arquiteturas Web, eles não serão discutidos aqui.

7.3.2 Ataques de multilocalização

O arquiteto responsável deve considerar as quatro possíveis formas de ataque abaixo utilizando *multi-tenancy*:

- **Acesso indevido de informações:**

Para cada *tenant* é dado um conjunto de recursos virtuais. Cada recurso virtual é mapeado para um recurso físico. É possível que uma informação de um recurso físico vaze de um **tenant** para outro, com ataques de memory overflow e exploração de bugs de execução especulativa, por exemplo.

- **”Fuga” de máquina virtual:**

Uma máquina virtual é isolada de outras máquinas virtuais através de espaços de endereços de memória distintos. É bem difícil e raro porém possível que um atacante explore erros de software no *hypervisor* para acessar informações sem autorização, através de exploits na memória virtual.

- **Ataque por canal lateral** (*Side-channel attack*):

Um atacante malicioso pode deduzir partes de chaves ou informações sensíveis monitorando os horários do ativamento de cache. Isso atualmente é um estudo acadêmico mas pode se tornar um problema futuro. Outras formas de ataques laterais em criptoanálises já foram comprovadas em alguns estudos, como por exemplo, a recuperação de uma chave a partir da análise sonora a partir da captura por um microfone de ruídos de um processador primitivo.

- **Ataque de negação de serviço:**

Um atacante malicioso pode criar artificialmente milhões de requisições simultâneas em alguns segundos, forçando a todos recursos serem consumidos até que não seja possível servir uma nova requisição. Essa forma de ataque de negação de serviço é comum em qualquer arquitetura web, porém em um cloud *multi-tenant*, outras formas são possíveis.

Outros *tenants* podem usar recursos suficientes no computador host para que seu aplicativo não seja capaz de fornecer serviço. Alguns provedores permitem que os clientes reservem máquinas inteiras para uso exclusivo. Embora isso encareça o custo do uso da nuvem, é um mecanismo efetivo contra ataques de multilocação. Uma organização deve considerar possíveis ataques ao decidir quais aplicativos hospedar na nuvem, exatamente como deveriam ao considerar qualquer opção de hospedagem.

PARTE IV

SOLUÇÕES DO MERCADO

8 AMAZON WEB SERVICES

8.1 Introdução

A Amazon Web Services (AWS) é uma plataforma de serviços em nuvem segura, oferecendo poder computacional, armazenamento de banco de dados, distribuição de conteúdo e outras funcionalidades para ajudar as empresas em seu dimensionamento e crescimento. Segundo pesquisas recentes (4), é o provedor de nuvem que detém a maior parcela do mercado em termos de adoção, com aproximadamente 64%. Sua gama de serviços é bastante ampla, como mostrado abaixo:



Figura 6: Serviços da Amazon Web Services(5)

A AWS possui quatro serviços essenciais de nuvem para hospedagem de aplicações, sendo eles: **EC2**, **S3**, **RDS** e **Route 53**. Além disso, possui soluções para outros variados campos da computação, como inteligência artificial, desenvolvimento *mobile*, realidade virtual e internet das coisas.

8.2 Funcionamento básico

No contexto do documento, serão destacados os quatro serviços já citados anteriormente: j1ec2shd

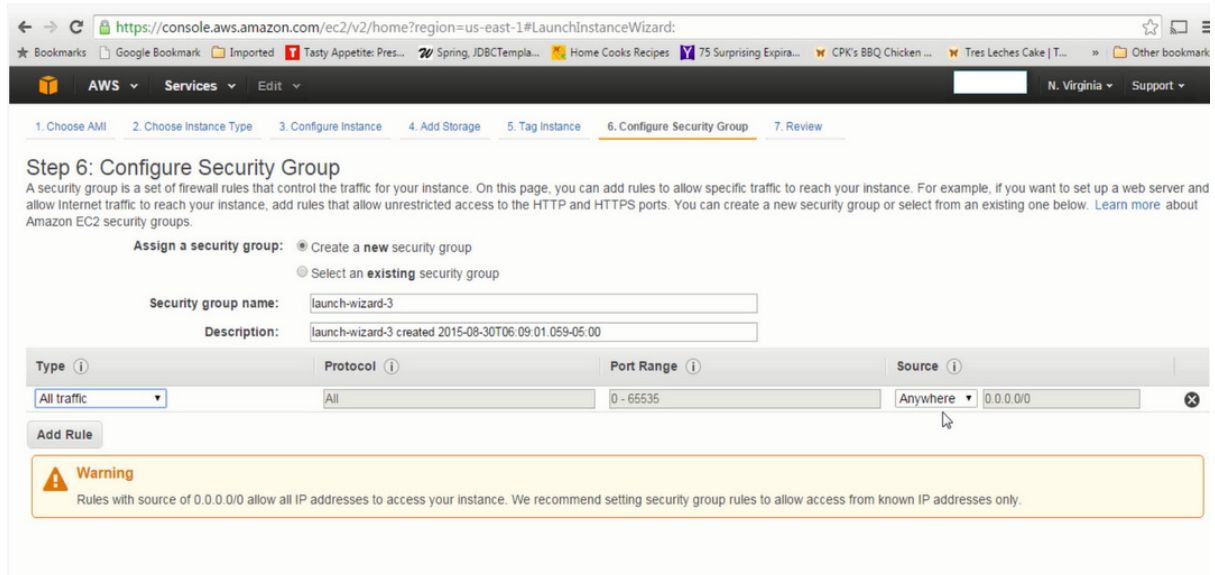


Figura 7: Console da Amazon Web Services

8.2.1 EC2

8.2.2 S3

8.2.3 RDS

8.2.4 Route 53

8.3 Outros Serviços

8.4 Conclusão

9 GOOGLE CLOUD

9.1 Introdução

O Google Cloud é uma das soluções mais completas atualmente no mercado, com diversas possibilidades e serviços disponíveis aos usuários, de acordo com suas necessidades. Possui toda o conjunto de hospedagem, desde uma aplicação pronta até toda a parte de infraestrutura de máquinas para escalabilidade.



Figura 8: Serviços do Google Cloud(6)

O Google Cloud possui quatro serviços essenciais de nuvem para aplicações, sendo eles: **App Engine**, **Compute Engine**, **Kubernetes Engine** e **Cloud Functions**. Além disso, possui toda a infraestrutura de gerenciamento de rede, banco de dados, *big data*, *machine learning*, entre outros.

9.2 Funcionamento básico

No contexto do documento, serão destacados os quatro serviços já citados anteriormente:

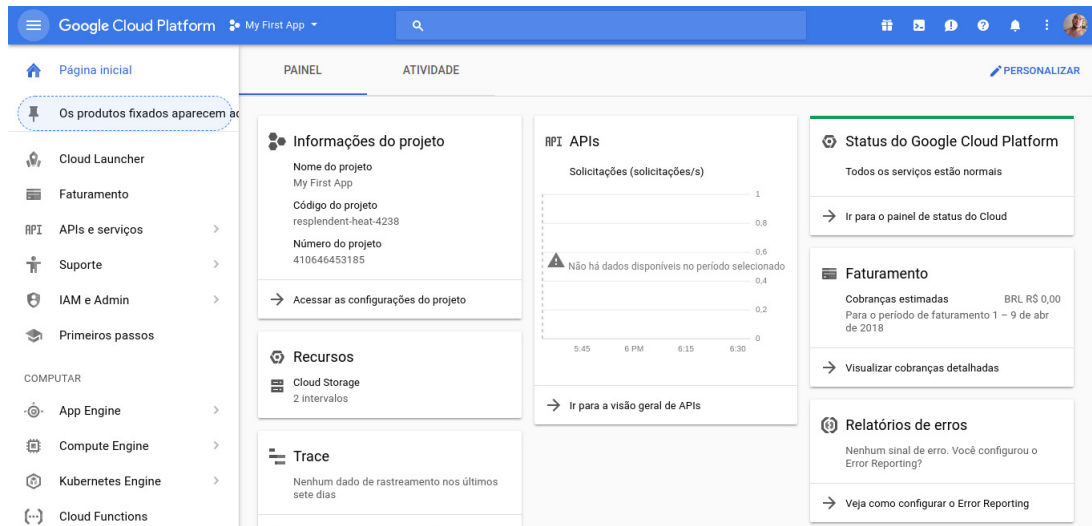


Figura 9: Exemplo do *Dashboard* do Google Cloud com seus recursos

9.2.1 Cloud Functions

As **Cloud Functions** têm como principal funcionalidade atender pequenos trechos de código, gerando *endpoints* prontos para o usuário final utilizar. O produto atende em especial pessoas que não necessitam de grandes aplicações para solucionarem seus problemas, onde um microserviço seria algo elevado demais para o contexto.

O funcionamento básico dele consiste em implementar soluções simples no ambiente de microserviços sobre monólitos, explicado da seguinte forma pelo Google(11):

A agilidade do desenvolvedor vem de sistemas de construção compostos por pequenas unidades de funcionalidade independentes que executam uma única coisa muito bem. O Cloud Functions permite criar e implantar serviços no nível de uma única função, e não no nível de aplicativos, contêineres ou VMs inteiros.

Contextualizando, é possível subir simples funções e deixar o gerenciamento da escalabilidade com o próprio Google Cloud, facilitando a vida do desenvolvedor. Uma solução similar é implementada pela suíte de soluções da Amazon Web Services: as funções *lambda*.

9.2.2 App Engine

O **App Engine** é um serviço intermediário, permitindo que aplicações completas sejam hospedadas em seu serviço, porém com os recursos necessários apenas e sem grandes preocupações em gerenciamento da escalabilidade.

Seu funcionamento consiste em hospedar aplicações completas no estilo *PaaS (Platform as a Service)*, deixando de lado a parte de gerenciamento de infraestrutura para o Google Cloud. O depoimento abaixo, de Stefan Hauk, Desenvolvedor da empresa de games Rovio, ilustra bem o funcionamento do serviço(12):

Nossos jogos da Web tendem a ficar famosos imediatamente, então não temos a opção de escalonamento ao longo do tempo. O Google App Engine simplifica esse processo, já que ele pode iniciar instantaneamente quantos servidores forem necessários.

A vantagem do serviço está no escalonamento automático simples, o que facilita o trabalho do desenvolvedor. Além disso, há uma integração tranquila com as principais linguagens de programação usadas, atendendo os mais diversos problemas de maneira simples. A solução é bem parecida com a implementada por padrão pelo Heroku.

9.2.3 Kubernetes Engine

O **Kubernetes Engine** é uma solução mais robusta que permite o uso de *containers*, dando maior poder ao desenvolvedor para escalar sua aplicação em máquinas virtuais. A solução adotada pelo Google usa seu próprio orquestrador de *containers*, o Kubernetes, definido pelo Google (em inglês)(13):

Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. It groups containers that make up an application into logical units for easy management and discovery.

Seu funcionamento consiste em gerenciar *containers* de acordo com as necessidades dos usuários, onde as aplicações colocadas em produção se encontram dentro deles. Essa ferramenta permite maior controle do desenvolvedor para determinar quantos e quais *containers* serão escalados, para atender a demanda necessária.

9.2.4 Compute Engine

Por fim, o **Compute Engine** é a solução mais poderosa fornecida pelo Google, onde é disponibilizada uma máquina completa para o usuário subir suas aplicações. A solução

é da categoria *IaaS* (*Infrastructure as a Service*), provisionando infraestruturas de acordo com a necessidade do desenvolvedor.

Seu funcionamento se assemelha ao EC2 da Amazon Web Services, onde a essência está em fornecer máquinas completas ao usuário final, com os mais diversos sistemas operacionais (tanto Linux quanto Windows). A escalabilidade é possível, expandindo recursos da máquina de acordo com as necessidades, mas todo o trabalho fica ao encargo do desenvolvedor.

9.3 Outros serviços

Além dos serviços básicos de aplicações, merecem destaque outros serviços:

- Armazenamento: Bigtable (NoSQL), Cloud SQL (MySQL e PostgreSQL).
- Rede: Rede VPC (Gerenciamento de IP), Balanceamento de Carga.
- *Stackdriver*: Monitoração, Depuração e Registros das Aplicações.
- Big Data: BigQuery (Análise de dados massivos), Dataflow (*Pipeline* para processamento de dados), IoT Core (Mensagens em tempo real).

9.4 Conclusão

O Google Cloud possui uma suíte bem completa de produtos, para os mais diversos públicos e problemas a serem solucionados, tanto soluções auto-escaláveis quanto máquinas completas para aplicações mais complexas, tudo com a infraestrutura do Google e a preços mais acessíveis do que concorrentes como o Heroku.

10 HEROKU

10.1 Introdução

O Heroku é uma plataforma de computação em nuvem conhecida no mercado, com a possibilidade de subir aplicações nas linguagens Ruby, Node.js, Java, Python, Clojure, Scala, Go e PHP. O grande destaque do Heroku está na facilidade em subir uma aplicação com facilidade e de maneira gratuita, o que possibilita testar e validar ideias básicas antes de escalar de fato.

A estrutura básica do Heroku funciona com o uso de *dynos*, que servem tanto para hospedar sua aplicação principal quanto máquinas auxiliares para serviços externos e/ou paralelos. Porém, recentemente o Heroku disponibilizou também a estrutura de *containers* em Docker (tecnologia de gerenciamento de *containers*) (assim como seus principais concorrentes), gerando maior flexibilidade para o serviço.

10.2 Funcionamento básico

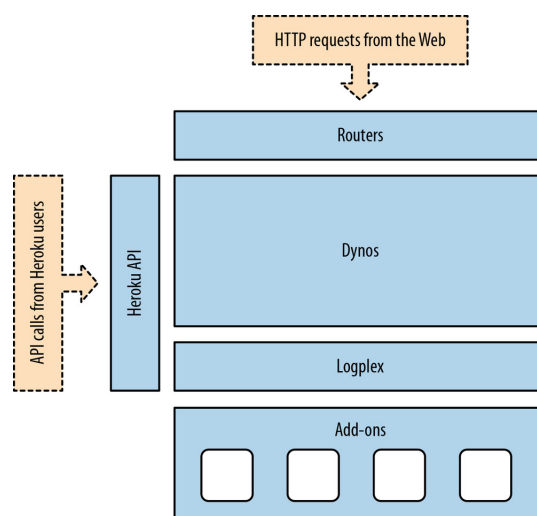


Figura 10: Arquitetura Básica do Heroku(7)

O funcionamento do Heroku consiste no uso de *dynos* para hospedar as aplicações e nos *routers* para tratar e encaminhar as requisições dos usuários. Além disso, o próprio Heroku disponibiliza extensões para gerenciar sua aplicação, como por exemplo o gerenciador de IP estático, ou o banco de dados necessário para a aplicação.

Um *dyno* é um *container* pronto, com 512MB de memória RAM, responsável por abrigar uma ou mais instâncias de sua aplicação, permitindo escalabilidade e tolerância de erros. Além disso, os *dynos* permitem rodar tarefas de sua aplicação à parte, como filas, requisições assíncronas, entre outros.

Já os *routers* são responsáveis por gerenciar os acessos dos usuários às aplicações correspondentes, dado que não é padrão do Heroku estruturar IP estático para cada *dyno*. Ou seja, a estrutura é responsável por fazer sua aplicação funcionar corretamente, sem acessar aplicações alheias.

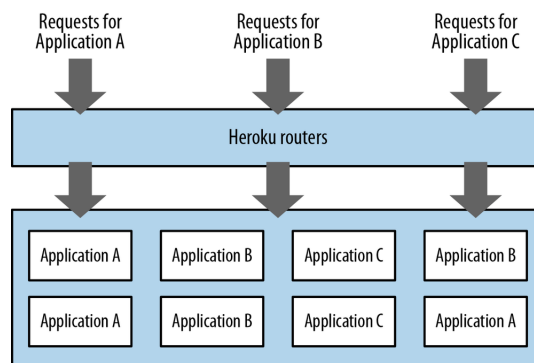


Figura 11: Ligação entre *router* e *dynos* do Heroku(7)

Graças a essa estrutura, é possível, com facilidade, gerenciar versionamento e escalar a aplicação de maneira horizontal, com novos *dynos* gerenciados pelo *router*, dando uma popularidade considerável ao Heroku, em especial nas aplicações onde velocidade de validar hipóteses é o principal foco.

10.3 Tarifação

O Heroku possui um diferencial em relação à tarifação: possui um plano básico gratuito que possibilita testar aplicações de maneira fácil e sem dificuldades de expansão. Essencialmente, a cobrança do Heroku ocorre via uso de seus *dynos*-hora. Além disso, há a cobrança pelas extensões usadas dentro da aplicação, onde, no geral, há um plano gratuito de testes ou até para projetos pequenos funcionarem com tranquilidade sem necessidades de escalabilidade.

Para qualquer usuário do Heroku, é disponibilizado um pacote de 750 *dynos*-hora durante um mês, ou seja, uma aplicação gratuita usando um simples *dyno* pode durar tranquilamente. Além disso, o Heroku de uso gratuito limita o tempo em que um *dyno* fica ligado de maneira ociosa: passando trinta minutos desde a última requisição feita pelo cliente (e repassada pelo *router*), o *dyno* é derrubado, sendo religado após uma nova requisição.

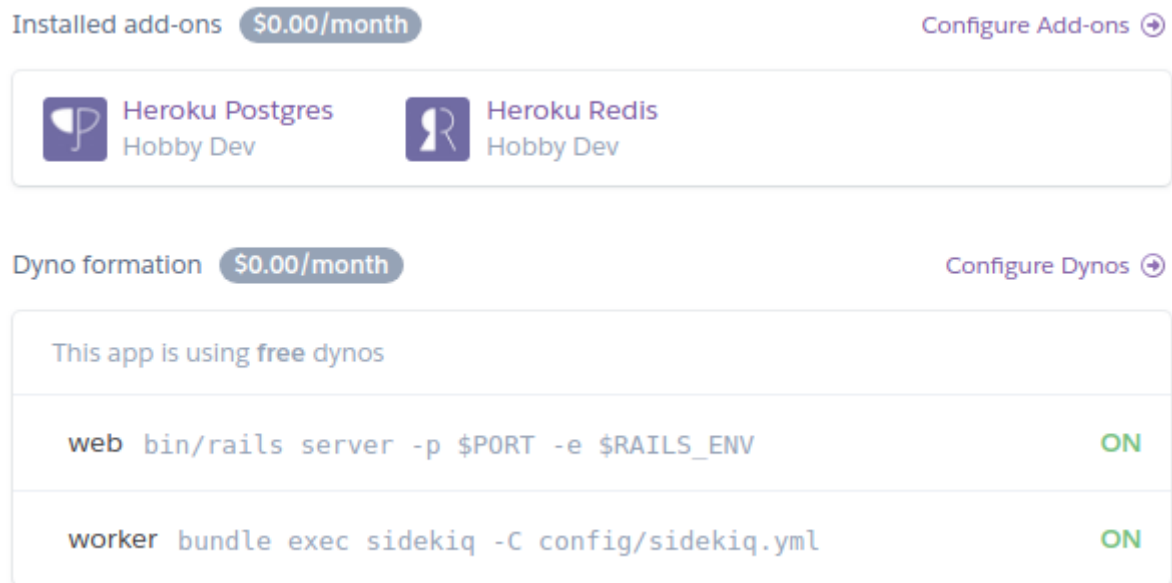


Figura 12: Exemplo do *Dashboard* do Heroku com seus recursos

Além da cobrança do uso dos *dynos*, há também a cobrança para cada extensão utilizada por sua aplicação, o que de fato encarece o custo final para escalar uma aplicação. Por exemplo, o banco de dados relacional padrão do Heroku é gratuito até certo limite de dados, passando desse limite, há a extensão do plano que permite, além de mais dados armazenados, aumenta o fluxo possível de acesso ao banco.

10.4 *Containers*

Recentemente, o Heroku possibilitou em seu catálogo de soluções o uso de estruturas isoladas customizadas: os *containers*. *Containers*, segundo o site do Docker⁽¹⁶⁾ (em inglês):

A container image is a lightweight, stand-alone, executable package of a piece of software that includes everything needed to run it: code, runtime, system tools, system libraries, settings.

Ou seja, são imagens autossuficientes e virtualizadas que permitem rodar aplicações prontas nos mais diversos dispositivos, nos moldes da virtualização.

O Heroku possibilitou, em 2017, uma maneira eficiente de receber *containers* em seu serviço, com o **Heroku Container Registry**(17). Esse serviço facilita colocar em produção as máquinas isoladas contruídas com o uso do Docker, aumentando as possibilidades de operação com a plataforma. A tarifação do serviço segue o mesmo padrão do Heroku tradicional, dado o fato de que *dynos* são uma espécie de *containers*.

10.5 Conclusão

O Heroku é uma solução que se destaca pela velocidade e facilidade de colocar uma ideia em produção, de maneira gratuita, com possibilidades eficientes de escalabilidade. Porém, financeiramente, não é a melhor solução, com uma tarifação elevada para escalar a aplicação. Além disso, sua estrutura atende apenas aplicações prontas, não permitindo outras estruturas avançadas como o uso de *containers*, sendo incorporado ao catálogo de produtos recentemente.

11.2 Vantagens e diferenciais

O primeiro atributo positivo a considerar em relação aos serviços de nuvem da Azure é a confiabilidade da infraestrutura do provedor. Seu Contrato de Nível de Serviço (em inglês, *SLA*) estima uma porcentagem de disponibilidade de 99,95%, o que equivale a pouco mais de quatro horas anuais de tempo de *downtime*. Essa é uma conquista que até mesmo uma solução *on premises* simplesmente não consegue atingir de forma consistente. No entanto, hospedar um aplicativo ou armazenar dados na nuvem geralmente está associado a uma alta disponibilidade - certamente um dos maiores apelos de *cloud computing*. Porém a alta disponibilidade dos Serviços de Nuvem do Azure da *Microsoft* continua a impressionar, inspirando muitas vezes *CIOs* a migrar pelo menos uma parte de suas soluções mais pesadas de dados *on-line*.

Além disso, uma das maiores preocupações das empresas ao migrar para a nuvem é a segurança. A Azure segue o modelo de segurança padrão, que consiste em “detectar, avaliar, diagnosticar, estabilizar e fechar”. Esse modelo, alinhado com outras políticas rígidas de segurança, rendeu à plataforma uma série de certificações de *compliance*, a estabelecendo como líder em segurança de *IaaS*. Fornece proteção em vários níveis e serviços simples e fáceis de usar.

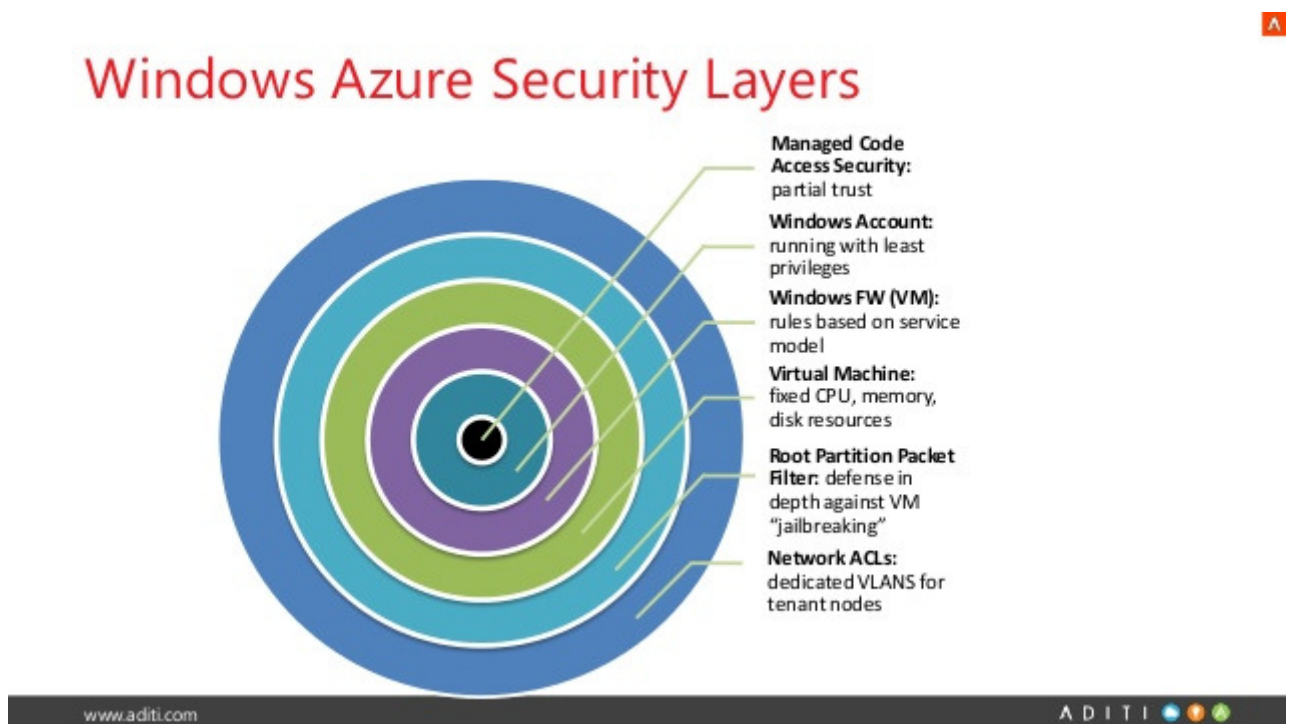


Figura 14: Camadas de Segurança da Azure

Finalmente, os serviços de nuvem da Azure também oferecem a escalabilidade necessária para o crescimento de qualquer solução, fornecendo maior poder computacional conforme necessário. Seja uma pequena empresa, ou uma empresa com a necessidade de um grande *data warehouse*, os serviços de nuvem do Azure têm armazenamento e capacidade computacional para lidar com praticamente qualquer tipo de demanda.

11.3 Desvantagens

Ao contrário das plataformas *SaaS*, nas quais o usuário final está consumindo informações (por exemplo, o *Office 365*), a Azure migra o poder computacional de dentro de sua empresa (*datacenter*) ou escritório para a nuvem. Como ocorre com a maioria dos provedores de serviços em nuvem, o Azure precisa ser gerenciado e mantido cuidadosamente, o que inclui aplicação de *patches* e o monitoramento dos servidores.

Ao contrário dos servidores locais, o Azure exige experiência para garantir que todas as partes móveis trabalhem juntas de forma eficiente. Um erro comum dos administradores de negócios que não estão totalmente envolvidos em quão bem (ou mal) seus servidores de nuvem estão operando é o de fazer o chamado *over-provisioning* dos serviços em nuvem. Embora seja um erro corriqueiro, essa falta de conhecimento dos administradores pode custar às empresas milhares de dólares por ano.

11.4 Conclusão

No geral, a Azure é uma excelente plataforma de *cloud computing*, ficando atrás de seus concorrentes apenas em alguns poucos pontos. Possui serviços robustos e bem consolidados, bastante segurança, e boa disponibilidade e escalabilidade. Além disso, torna-se quase que uma escolha obrigatória para empresas que já utilizam ou tem familiaridade com ambientes *Microsoft*, devido às inúmeras facilidades extras fornecidas (integrações e possíveis cortes de gastos com licenças).

REFERÊNCIAS

- 1 JOHNSTON, S. *Diagrama ilustrando Computação em Nuvem*. 2013. Disponível em: https://upload.wikimedia.org/wikipedia/commons/b/b5/Cloud_computing.svg.
- 2 AMAZON. *Amazon EC2: Boas Práticas e Otimizações de Desempenho*. 2017. Disponível em: <https://www.slideshare.net/AmazonWebServicesLATAM/amazon-ec2-boas-praticas-e-otimizaes-de-desempenho>.
- 3 BALDASSARI, J. *A detailed look at why mabl Chose Google Cloud Platform (GCP) over AWS*. 2017. Disponível em: <https://www.mabl.com/blog/why-we-chose-google-cloud-platform-over-aws-at-mabl>.
- 4 RIGHTSCALE. *RightScale 2018 State of the Cloud ReportTM*. 2018. Disponível em: <https://assets.rightscale.com/uploads/pdfs/RightScale-2018-State-of-the-Cloud-Report.pdf>.
- 5 SERVICES, A. W. *AWS Products*. 2018. Disponível em: <https://aws.amazon.com/pt/products/>.
- 6 LANGIT, L. *Scaling Galaxy on Google Cloud Platform*. 2017. Disponível em: <https://www.slideshare.net/lynnlangit/scaling-galaxy-on-google-cloud-platform>.
- 7 SCHNEEMAN, R.; MIDDLETON, N. *Heroku: Up and Running*. 2013. Disponível em: <https://www.safaribooksonline.com/library/view/heroku-up-and/9781449341381/ch02.html>.
- 8 MCKENDRICK, J. *10 Quotes on Cloud Computing That Really Say it All*. 2013. Disponível em: <https://www.forbes.com/sites/joemckendrick/2013/03/24/10-quotes-on-cloud-computing-that-really-say-it-all/#1e8843877f6c>.
- 9 UKESSAYS. *Life Before Cloud Computing Information Technology Essay*. 2013. Disponível em: <https://www.ukessays.com/essays/information-technology/life-before-cloud-computing-information-technology-essay.php>.
- 10 AMAZON. *Announcing Amazon Elastic Compute Cloud (Amazon EC2) - beta*. 2006. Disponível em: <https://aws.amazon.com/pt/about-aws/whats-new/2006/08/24/announcing-amazon-elastic-compute-cloud-amazon-ec2---beta/>.
- 11 GOOGLE. *Cloud Functions*. 2018. Disponível em: <https://cloud.google.com/functions/?hl=pt-br>.
- 12 GOOGLE. *App Engine*. 2018. Disponível em: <https://cloud.google.com/appengine/?hl=pt-br>.
- 13 GOOGLE. *Kubernetes*. 2018. Disponível em: <https://kubernetes.io/>.

14 GOOGLE. *Container Engine*. 2018. Disponível em: <https://cloud.google.com/kubernetes-engine/?hl=pt-br>.

15 GOOGLE. *Compute Engine*. 2018. Disponível em: <https://cloud.google.com/compute/?hl=pt>.

16 DOCKER. *What is a Container*. 2018. Disponível em: <https://www.docker.com/what-container>.

17 HEROKU. *Container Registry & Runtime (Docker Deploys)*. 2018. Disponível em: <https://devcenter.heroku.com/articles/container-registry-and-runtime>.