



INFRAESTRUTURA E APLICATIVOS COM GOOGLE CLOUD

Redes em ambientes Google Cloud

Toda infraestrutura necessita de uma rede para comunicação entre aplicações, seja internamente (componentes que estão na mesma infraestrutura), seja externamente (componentes que estão em infraestruturas diferentes, por exemplo, outro servidor ou provedor de nuvem). Isso não é diferente em ambientes de nuvem. Para todo ambiente alocado em Google Cloud deve haver uma rede configurada.

Na plataforma do Google Cloud, por se tratar de uma virtualização de um ambiente físico disponibilizado como serviço para o usuário, utilizamos uma rede virtual privada para configurações, conhecida como VPC (virtual private cloud).

Uma **VPC** é um modo de rede seguro, individual e privado, hospedado em Google Cloud e, por meio dela, os usuários podem disponibilizar suas aplicações, armazenar dados e hospedar sites, por exemplo. Uma cloud privada virtual conecta os recursos do Google Cloud a qualquer outro recurso na internet, e por meio dela é possível configurar políticas de firewall, IPs (internet protocol), portas e protocolos.

A VPC do Google Cloud é global. Isso significa que, ao criá-la, podemos ser atendidos em qualquer região disponível. Para isso, dentro das redes privadas virtuais, são configuradas as sub-redes, conhecidas como subnets (da palavra em inglês subnetwork), que podem estar alocadas em qualquer região onde sejam necessárias e, com isso, todas as políticas e regras podem ser feitas regionalmente. Utilizar uma rede global e subnets alocadas em regiões diferentes possibilita uma grande diversidade de arquiteturas. Uma estratégia muito utilizada é a de disponibilidade por regiões. Entenda melhor:

1. Acesso e tempo de resposta

Imagine uma empresa que presta serviços globais, para todos os continentes. Para um usuário da Europa acessar a região do Brasil no Google Cloud é mais complexo, pois a distância é grande, e isso impacta no tempo de entrega e resposta.

2. Subnets: acesso por geolocalização

Então, ao traçarmos um planejamento de subnets em regiões diferentes, possibilitamos o acesso por geolocalização, ou seja, quem está na América do Sul acessa o servidor do Brasil, quem está na Europa

acessa o servidor mais próximo, e o mesmo ocorre com a América do Norte.

3. Agilidade em comunicação

Isso traz agilidade no tempo de resposta e latência da comunicação e a empresa consegue prestar um serviço com uma boa experiência para o usuário.

Essa mesma estrutura de ter a aplicação em regiões diferentes funciona para uma estratégia de recuperação de desastres (disaster recovery). Isso funciona para o caso de uma indisponibilidade de uma região inteira, como, por exemplo, um desastre meteorológico como um furacão ou tsunami, ou até mesmo algo mais simples, como um incêndio. Nesse caso, teremos uma segunda opção de acesso em outra região disponível e, por mais que a experiência não seja a melhor possível, ainda assim as aplicações e serviços estarão disponíveis.

Máquinas virtuais em ambientes Google Cloud

Uma máquina virtual, também conhecida como VM (virtual machine), não é diferente de um computador ou laptop físico tradicional. Ela possui CPU, memória, armazenamento e rede, de maneira que pode se conectar à internet, caso necessário. A grande diferença é que essa máquina está sendo “emulada”, ou virtualizada, em um ambiente físico. Um único hardware pode virtualizar várias máquinas utilizando seus recursos e, com isso, ter uma grande variedade de aplicações e diferentes sistemas operacionais, todos utilizando apenas um hardware.

O Google Cloud oferece máquinas virtuais como IaaS, uma infraestrutura como serviço, e essa oferta é chamada de Compute Engine, de forma que é possível selecionar a máquina que mais encaixar com os requisitos de recursos. As máquinas virtuais estão divididas em algumas categorias no Google Cloud:

Família de máquinas

Um conjunto selecionado de configurações de processador e hardware otimizadas para cargas de trabalho (aplicações) específicas. Ao criar uma VM, podemos escolher um tipo de máquina predefinido.

Série

As famílias de máquinas são classificadas também por série e geração. Por exemplo, a série N1 dentro da família de máquinas de uso geral é a versão mais antiga da série N2. Geralmente, usa-se um número maior para descrever as gerações mais atuais. Sendo assim, a N3 é mais nova que a N2 que, por sua vez, é mais nova que a N1.

Tipo de máquina

Cada série tem tipos de máquinas predefinidas, que possuem um conjunto de recursos para a VM. Se nenhum deles atender sua necessidade, é possível criar uma máquina personalizada.

As máquinas disponibilizadas pelo Google Cloud são apresentadas na tabela a seguir:

Após selecionarmos o tipo de máquina e provisionarmos essa máquina utilizando o serviço de IaaS do Google Cloud, ela se torna uma estação de trabalho utilizando o sistema operacional que foi escolhido, pronto para implantação da sua aplicação.

Uma das vantagens de uma VM é a facilidade de gerenciamento. Além de ser flexível, do ponto de vista de recursos computacionais, podemos facilmente deletá-la e recriá-la de maneira muito simples. Outro ponto é a flexibilidade em customizações, de forma que podemos instalar a versão que melhor nos atender de um software e customizar as bibliotecas.



Servidores da Google em Hamina, Finlândia.

A responsabilidade de gerenciamento da VM é mista. Toda a infraestrutura física é provida e gerenciada pelo Google, mas o gerenciamento de utilização dos recursos, instalação e atualizações de softwares é de responsabilidade única do usuário. O que também se aplica para responsabilidades de segurança, nas quais o Google Cloud é responsável pelo hardware. Entretanto,

do lado do software instalado nas VMs, é dever do usuário administrar e monitorar.

Vejamos os exemplos a seguir sobre gerenciamento da VM.

Exemplo

Para exemplificar, imagine que provisionamos uma máquina virtual Linux, com 8 CPUs e 4 GB de memória. Em dado momento, minha VM ficou indisponível, pois o servidor no qual ela estava localizada em São Paulo teve falta de energia elétrica. De quem é a responsabilidade? Nesse caso, é do Google, pois, como vimos, é ele o responsável por gerenciar e disponibilizar a infraestrutura física. Outro exemplo: imagine que provisionamos a mesma VM do exemplo anterior, mas minha aplicação necessita de 8 GB de memória ao invés dos 4 GB que solicitamos. De quem é a responsabilidade de gerenciar isso? Do usuário, pois trata-se do gerenciamento de consumo de recursos no qual o cliente pode, se necessário, solicitar mais memória ou reduzir, por exemplo.

Aplicações para ambientes de nuvem

Existem inúmeras maneiras de desenvolver uma aplicação. Isso porque os requisitos podem ser diferentes e, com isso, podem ser usadas tecnologias diferentes para cada um dos casos, como uma linguagem de programação específica. Um dos requisitos que devem ser levados em conta é o ambiente no qual a aplicação será implementada, já que deve estar adequada aos recursos que consumirá e deve estar alinhada com os requisitos de negócio e de investimento.

Atenção!

Quando pensamos no ambiente de nuvem, lembramos que ele é cobrado por recursos consumidos. Então, como necessidade, devemos adequar nossas aplicações para consumir o mínimo de CPU e memória possíveis, mas que, ao mesmo tempo, se mantenham estáveis. Esse tipo de arquitetura tem um nome próprio, aplicativo nativo da nuvem (cloud native application).

Uma aplicação nativa de nuvem consiste em um software que utiliza recursos como serviço e que é desacoplado, ou seja, seus componentes não dependem de outros componentes externos, é independente. Um dos principais benefícios é se adaptar e agregar valor tecnológico ao negócio, fornecendo uma experiência de desenvolvimento mais ágil e com as melhores práticas.

Duas arquiteturas comuns são a “com estado” (do termo em inglês, stateful) e a “sem estado” (do termo em inglês, stateless). As duas opções têm pontos positivos e negativos e suas complexidades particulares. Entenda a seguir:

Aplicações stateful (com estado)

O termo “estado” refere-se ao armazenamento de informações, assim, uma arquitetura stateful consiste em uma aplicação que, além da lógica, armazena os dados em si.

Aplicações stateless (sem estado)

Tem todos os dados armazenados em um componente externo, ficando apenas responsável pela lógica e processamento.

A arquitetura stateless traz uma possibilidade de implantação em uma infraestrutura “sem servidor” (do termo em inglês, serverless). A ideia não é literalmente implantar em nenhum servidor, mas consiste em ser uma infraestrutura elástica do ponto de vista de escalabilidade. Sendo assim, podemos facilmente escalar para mais de uma “cópia” do aplicativo para atender à demanda e, caso não exista nenhum tipo de requisição, é possível “desligar” o servidor. Como vimos, em ambientes de nuvem a cobrança é feita com base no aluguel de recursos e, ao escalar para zero, essa aplicação não vai gerar consumo, sendo assim, não pagaremos por isso.

Aplicações com estado possuem algumas peculiaridades em algumas arquiteturas. Imagine que tenhamos que duplicar uma aplicação que armazena dados para poder atender a uma alta demanda. Nesse caso, como poderíamos garantir a consistência de informações entre elas? Ou, caso fôssemos utilizar uma aplicação com estado em uma estratégia “sem servidor”, ao escalar para zero ou “desligar” o servidor, como iremos garantir que as informações não serão perdidas?



Estratégia serverless.

Fora todas essas indagações, uma estratégia serverless precisa ter um tempo de resposta extremamente alto, tendo que “ligar” a aplicação no servidor no menor tempo possível. Devido a isso, devem ser leves e portáteis. Arquiteturas stateful tendem a ser mais pesadas e, por isso, possuem um tempo de início mais longo.

O Google Cloud possui um PaaS que entrega uma infraestrutura sem servidor para as aplicações. Essa solução se chama **Cloud Run** e suporta as principais linguagens de programação, como Node.JS, Python, Java, Go, entre outras. Um dos grandes benefícios de uma plataforma como serviço é que podemos nos preocupar somente com a aplicação e sua lógica e isso agrega muito, pois podemos nos concentrar nos processos e regras de negócio, trazendo maior agilidade produtiva.

Dica

Para aplicações com estado, dentro do Google Cloud é possível utilizar o Compute Engine, solução de máquinas virtuais do Google, ou até mesmo tecnologias mais atuais, como contêineres e kubernetes.