

Máquinas virtuais na nuvem

Fundamentos do GCP: infraestrutura principal



Primeiros passos com o Compute Engine



Google Cloud

Última modificação 2018-08-13

© 2017 Google Inc. Todos os direitos reservados. Google e o logotipo da Google são marcas registradas da Google Inc. Todos os outros nomes de empresas e produtos podem ser marcas registradas das respectivas empresas de seus países respectivos.

© 2017 Google Inc. Todos os direitos reservados. Google e o logotipo da Google são marcas registradas da Google Inc. Todos os outros nomes de empresas e produtos podem ser marcas registradas das respectivas empresas de seus países respectivos.

Agenda

Rede de nuvem privada virtual (VPC)

Compute Engine

Recursos importantes da VPC

Teste e laboratório

Rede de nuvem privada virtual

- Cada rede VPC está contida em um projeto do GCP.
- Você pode provisionar recursos do Cloud Platform, conectá-los entre si e isolá-los uns dos outros.

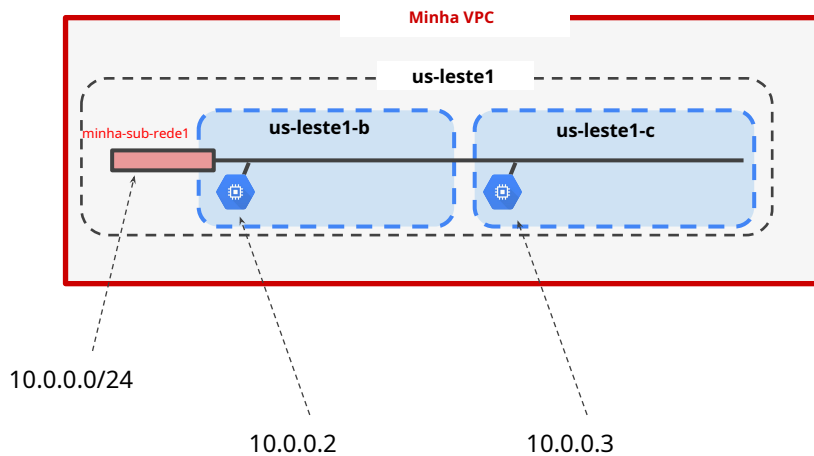


 Google Cloud

Suas redes VPC conectam seus recursos do Google Cloud Platform entre si e com a Internet. Você pode segmentar suas redes, usar regras de firewall para restringir o acesso a instâncias e criar rotas estáticas para encaminhar tráfego para destinos específicos.

Muitos usuários começam a usar o GCP para definir sua própria nuvem privada virtual dentro do primeiro projeto do GCP. Ou eles podem simplesmente escolher a VPC padrão e começar com isso.

As redes VPC do Google Cloud são globais; as sub-redes são regionais



As redes do Google Virtual Private Cloud que você define têm escopo global. Eles podem ter sub-redes em qualquer região do GCP em todo o mundo. As sub-redes podem abranger as zonas que compõem uma região. Essa arquitetura facilita a definição de seu próprio layout de rede com escopo global. Você também pode ter recursos em diferentes zonas na mesma sub-rede.

Você pode aumentar dinamicamente o tamanho de uma sub-rede em uma rede personalizada expandindo o intervalo de endereços IP alocados a ela. Isso não afeta as VMs já configuradas.

Neste exemplo, sua VPC tem uma rede. Até agora, ele tem uma sub-rede definida, na região us-east1 do GCP. Observe que ele tem duas VMs do Compute Engine anexadas a ele. Eles são vizinhos na mesma sub-rede, embora estejam em zonas diferentes! Você pode usar esse recurso para criar soluções resilientes, mas ainda com layouts de rede simples.

Agenda

Rede de nuvem privada virtual (VPC)

Compute Engine

Recursos importantes da VPC

Teste e laboratório

O Compute Engine oferece máquina virtual gerenciada S

- Alta CPU, alta memória, tipos de máquina padrão e de núcleo compartilhado
- Discos permanentes
 - Padrão, SSD, SSD local
 - Instantâneos
- Redimensione discos sem tempo de inatividade
- Metadados de instância e scripts de inicialização



 Google Cloud

As máquinas virtuais têm o poder e a generalidade de um sistema operacional completo em cada uma. Você configura uma máquina virtual da mesma forma que cria um servidor físico: especificando suas quantidades de energia e memória da CPU, suas quantidades e tipos de armazenamento e seu sistema operacional. O Compute Engine permite criar e executar máquinas virtuais na infraestrutura do Google. Não há investimentos iniciais e você pode executar milhares de CPUs virtuais em um sistema projetado para ser rápido e oferecer desempenho consistente.

Você pode reconfigurar com flexibilidade as máquinas virtuais do Compute Engine. E uma VM em execução na nuvem do Google tem conectividade de rede mundial incomparável.

Você pode criar uma instância de máquina virtual usando o Console do Google Cloud Platform ou a ferramenta de linha de comando gcloud. Uma instância do Compute Engine pode executar imagens do Linux e do Windows Server fornecidas pelo Google ou qualquer versão personalizada dessas imagens. Você também pode criar e executar imagens de outros sistemas operacionais.

O Compute Engine oferece preços acessíveis ao cliente

- Faturamento por segundo, descontos por uso prolongado, descontos por uso contínuo
- Instâncias preemptivas
- Alta taxa de transferência para armazenamento sem custo extra
- Tipos de máquina personalizados: pague apenas pelo hardware que você precisa



 Google Cloud

O Compute Engine cobra por segundo pelo uso de máquinas virtuais, com um mínimo de um minuto. E os descontos se aplicam automaticamente a máquinas virtuais que funcionam por frações substanciais de um mês. Para cada VM executada por mais de 25% de um mês, o Compute Engine oferece automaticamente um desconto para cada minuto incremental. Você pode obter um desconto líquido de até 30% para VMs que executam o mês inteiro.

O Compute Engine oferece a possibilidade de comprar contratos de uso contínuo em troca de preços com grandes descontos para uso de VM. Esses descontos são conhecidos como descontos por uso contínuo. Se sua carga de trabalho for estável e previsível, você poderá comprar uma quantidade específica de vCPUs e memória com até 57% de desconto em relação aos preços normais em troca de se comprometer com um período de uso de 1 ou 3 anos.

Suponha que você tenha uma carga de trabalho que nenhum ser humano está esperando para terminar. Digamos, um trabalho em lote analisando um grande conjunto de dados. Você pode economizar dinheiro escolhendo VMs preemptivas para executar o trabalho. Uma VM preemptiva é diferente de uma VM comum do Compute Engine em apenas um aspecto: você concedeu ao Compute Engine permissão para encerrá-la se seus recursos forem necessários em outro lugar. Você pode economizar muito dinheiro com VMs preemptivas, mas certifique-se de que seu trabalho possa ser interrompido e reiniciado.

Você não precisa selecionar uma opção específica ou tipo de máquina para obter alto rendimento

entre seu processamento e seus discos permanentes. Esse é o padrão.

Você pode escolher as propriedades de máquina de suas instâncias, como o número de CPUs virtuais e a quantidade de memória, usando um conjunto de tipos de máquina predefinidos ou criando seus próprios tipos de máquina personalizados.

Escale verticalmente ou horizontalmente com o Compute Engine



Use grandes VMs para aplicativos com uso intensivo de memória e computação



*Use o escalonamento automático para resiliência,
aplicativos escaláveis*

Você pode criar VMs muito grandes no Compute Engine. Na época em que este deck foi produzido, o número máximo de CPUs virtuais em uma VM era de 96, e o tamanho máximo de memória era de 624 na versão beta. Verifique o site do GCP para ver onde estão esses máximos hoje.

Essas enormes VMs são ótimas para cargas de trabalho como bancos de dados na memória e análises com uso intensivo de CPU. Mas a maioria dos clientes do GCP começa expandindo, não aumentando. O Compute Engine tem um recurso chamado Autoscaling que permite adicionar e remover VMs do seu aplicativo com base nas métricas de carga. A outra parte de fazer esse trabalho é equilibrar o tráfego de entrada entre as VMs. E o Google VPC oferece suporte a vários tipos diferentes de balanceamento de carga! Vamos considerá-los na próxima seção.

Agenda

Rede de nuvem privada virtual (VPC) do Google

Google Compute Engine

Recursos importantes da VPC

Teste e laboratório

Você controla a topologia da sua rede VPC

- Use sua tabela de rotas para encaminhar o tráfego dentro do rede, mesmo entre sub-redes.
- Use seu firewall para controlar qual tráfego de rede é permitido.
- Use a VPC compartilhada para compartilhar uma rede ou sub-redes individuais, com outros projetos do GCP.
- Use o emparelhamento de VPC para interconectar redes em Projetos GCP.

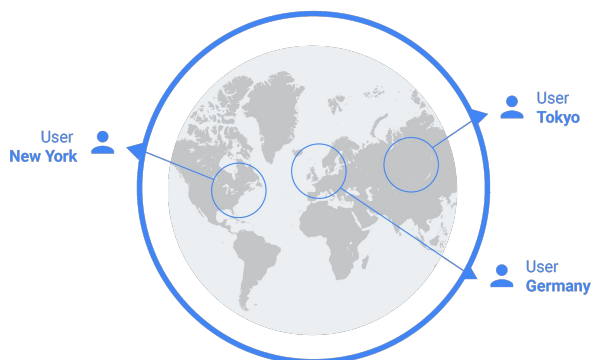


Assim como as redes físicas, as VPCs têm tabelas de roteamento. Eles são usados para encaminhar o tráfego de uma instância para outra na mesma rede, mesmo entre sub-redes e até mesmo entre zonas do GCP, sem exigir um endereço IP externo. As tabelas de roteamento das VPCs são incorporadas; você não precisa provisionar ou gerenciar um roteador.

Outra coisa que você não precisa provisionar ou gerenciar para o GCP: um firewall. As VPCs oferecem um firewall distribuído global que você pode controlar para restringir o acesso a instâncias, tanto para tráfego de entrada quanto de saída. Você pode definir regras de firewall em termos de tags de metadados em instâncias do Compute Engine, o que é muito conveniente. Por exemplo, você pode marcar todos os seus servidores da Web com, digamos, "WEB" e escrever uma regra de firewall dizendo que o tráfego nas portas 80 ou 443 é permitido em todas as VMs com a tag "WEB", independentemente do endereço IP deles. ser.

Lembre-se de que as VPCs pertencem a projetos do GCP. Mas e se sua empresa tiver vários projetos do GCP e as VPCs precisarem conversar entre si? Se você deseja simplesmente estabelecer uma relação de emparelhamento entre duas VPCs, para que elas possam trocar tráfego, configure o emparelhamento de VPC. Por outro lado, se você quiser usar todo o poder do IAM para controlar quem e o que em um projeto pode interagir com uma VPC em outro, configure a VPC compartilhada.

aplicativo apresenta um



- Os usuários obtêm um único endereço IP anycast global.
- O tráfego passa pelo backbone do Google do ponto de presença mais próximo ao usuário.
- Os back-ends são selecionados com base na carga.
- Apenas back-ends íntegros recebem tráfego.
- Não é necessário pré-aquecimento.

Alguns slides atrás, falamos sobre como as máquinas virtuais podem ser dimensionadas automaticamente para responder à mudança de carga. Mas como seus clientes chegam ao seu aplicativo quando ele pode ser fornecido por quatro VMs em um momento e quarenta VMs em outro? O Cloud Load Balancing é a resposta.

O Cloud Load Balancing é um serviço totalmente distribuído, definido por software e gerenciado para todo o seu tráfego. E como os balanceadores de carga não são executados em VMs que você precisa gerenciar, você não precisa se preocupar em escaloná-los ou gerenciá-los. Você pode colocar o Cloud Load Balancing na frente de todo o seu tráfego: HTTP(S), outros tráfegos TCP e SSL e também tráfego UDP.

Com o Cloud Load Balancing, um único front-end de IP anycast faz o front-end de todas as suas instâncias de back-end em regiões ao redor do mundo. Ele fornece balanceamento de carga entre regiões, incluindo failover multirregional automático, que move suavemente o tráfego em frações se os back-ends se tornarem íntegros. O Cloud Load Balancing reage rapidamente a mudanças nos usuários, tráfego, rede, integridade do back-end e outras condições relacionadas.

E se você antecipar um grande aumento na demanda? Digamos que seu jogo online já seja um sucesso; você precisa enviar um tíquete de suporte para avisar o Google sobre a carga recebida? Não. Não é necessário o chamado “pré-aquecimento”.

O Google VPC oferece um conjunto de opções de balanceamento de carga

HTTP(S) global	SSL global <i>Procuração</i>	TCP global <i>Procuração</i>	Regional	Regional interno
Carga da camada 7 baseado em balanceamento carregando	Carga da camada 4 balanceamento de SSL não HTTPS tráfego baseado em carregar	Carga da camada 4 balanceamento de TCP não SSL tráfego	Balanceamento de carga de qualquer tráfego (TCP, UDP)	Balanceamento de carga de tráfego dentro de um VPC
Pode rotear URLs diferentes para costas diferentes termina	Compatível com porta específica números	Compatível com porta específica números	Suportado em qualquer número da porta	Uso para o camadas internas de multicamadas formulários

Se você precisar de balanceamento de carga entre regiões para um aplicativo Web, use o balanceamento de carga HTTP(S). Para tráfego de Secure Sockets Layer que não seja HTTP, use o balanceador de carga Global SSL Proxy. Se for outro tráfego TCP que não usa Secure Sockets Layer, use o balanceador de carga Global TCP Proxy.

Esses dois serviços de proxy funcionam apenas para números de porta específicos e funcionam apenas para TCP. Se você quiser balancear a carga do tráfego UDP ou de qualquer número de porta, ainda poderá balancear a carga em uma região do GCP com o balanceador de carga regional.

Por fim, o que todos esses serviços têm em comum é que eles são destinados ao tráfego que entra na rede do Google a partir da Internet. Mas e se você quiser balancear a carga do tráfego dentro do seu projeto, digamos, entre a camada de apresentação e a camada de negócios do seu aplicativo? Para isso, use o balanceador de carga interno. Ele aceita tráfego em um endereço IP interno do GCP e faz o balanceamento de carga nas VMs do Compute Engine.

O Cloud DNS é altamente disponível e escalável

- Crie zonas gerenciadas e adicione, edite, exclua registros DNS
- Gerencie zonas e registros programaticamente usando a API RESTful ou a interface de linha de comando



 Google Cloud

Um dos serviços do Google mais famosos pelos quais as pessoas não pagam é o 8.8.8.8, que fornece um serviço de nome de domínio público para o mundo. O DNS é o que traduz os nomes de host da Internet em endereços e, como você pode imaginar, o Google possui uma infraestrutura de DNS altamente desenvolvida. Ele disponibiliza o 8.8.8.8 para que todos possam tirar proveito dele.

Mas e os nomes de host e endereços da Internet dos aplicativos que você cria no GCP?

O GCP oferece o Cloud DNS para ajudar o mundo a encontrá-los. É um serviço de DNS gerenciado executado na mesma infraestrutura do Google. Ele tem baixa latência e alta disponibilidade e é uma maneira econômica de disponibilizar seus aplicativos e serviços para seus usuários. As informações de DNS que você publica são fornecidas a partir de locais redundantes em todo o mundo.

O Cloud DNS também é programável. Você pode publicar e gerenciar milhões de zonas e registros DNS usando o Console do GCP, a interface de linha de comando ou a API.

Cloud CDN (rede de entrega de conteúdo)

- Use os caches de borda distribuídos globalmente do Google para armazenar conteúdo em cache próximo aos seus usuários
- Ou use o CDN Interconnect se preferir usar um CDN diferente



 Google Cloud

O Google tem um sistema global de caches de borda. Você pode usar esse sistema para acelerar a entrega de conteúdo em seu aplicativo usando o Google Cloud CDN. Seus clientes terão uma latência de rede menor, as origens do seu conteúdo terão uma carga reduzida e você também poderá economizar dinheiro. Depois de configurar o balanceamento de carga HTTP(S), basta ativar o Cloud CDN com uma única caixa de seleção.

Existem muitos outros CDNs por aí, é claro. Se você já estiver usando um, é provável que ele faça parte do programa de parceiros CDN Interconnect do GCP e você possa continuar a usá-lo.

O Google Cloud Platform oferece muitas opções de interconexão



VPN

Multi-Gbps seguro
conexão sobre
Túneis VPN



Peering direto

Conexão privada
entre você e o Google
para sua nuvem híbrida
cargas de trabalho



Interconexão Dedicada

Conecte circuitos de transporte NX 10G para
tráfego de nuvem privada ao Google Cloud em
POPs do Google
SLAs disponíveis



Emparelhamento de operadora

Conexão através do
maior rede de parceiros de
provedores de serviço



Interconexão de parceiros

Conectividade entre sua rede
local e sua rede VPC por meio de um
serviço compatível
fornecedor
SLAs disponíveis



Muitos clientes do GCP desejam interconectar suas outras redes às VPCs do Google, como redes locais ou suas redes em outras nuvens. Há muitas boas escolhas.

Muitos clientes começam com uma conexão de Rede Privada Virtual pela Internet, usando o protocolo IPsec. Para tornar isso dinâmico, eles usam um recurso do GCP chamado Cloud Router. O Cloud Router permite que suas outras redes e sua Google VPC troquem informações de rota pela VPN usando o Border Gateway Protocol. Por exemplo, se você adicionar uma nova sub-rede à sua Google VPC, sua rede local obterá automaticamente rotas para ela.

Mas alguns clientes não querem usar a Internet, seja por questões de segurança ou porque precisam de largura de banda mais confiável. Eles podem considerar o peering com o Google usando o peering direto. Peering significa colocar um roteador no mesmo datacenter público que um ponto de presença do Google e trocar tráfego. O Google tem mais de 100 pontos de presença em todo o mundo. Os clientes que ainda não estão em um ponto de presença podem contratar um parceiro no programa Carrier Peering para se conectar.

Uma desvantagem do peering, porém, é que ele não é coberto por um Contrato de nível de serviço do Google. Clientes que desejam os maiores tempos de atividade para sua interconexão com

O Google deve usar a Interconexão dedicada, na qual os clientes obtêm uma ou mais conexões privadas diretas com o Google. Se essas conexões tiverem topologias que atendam às especificações do Google, elas poderão ser cobertas por um SLA de até 99,99%. Essas conexões podem ser apoiadas por uma VPN para uma confiabilidade ainda maior.

O Partner Interconnect fornece conectividade entre sua rede local e sua rede VPC por meio de um provedor de serviços compatível. Uma conexão de Interconexão por parceiro é útil se seu data center estiver em um local físico que não pode alcançar uma instalação de colocation de Interconexão dedicada ou se suas necessidades de dados não garantirem uma conexão inteira de 10 Gbps. Dependendo de suas necessidades de disponibilidade, você pode configurar o Partner Interconnect para oferecer suporte a serviços ou aplicativos de missão crítica que podem tolerar algum tempo de inatividade. Assim como no Dedicated Interconnect, se essas conexões tiverem topologias que atendam às especificações do Google, elas poderão ser cobertas por um SLA de até 99,99%, mas o Google não é responsável por nenhum aspecto do Partner Interconnect fornecido pelo provedor de serviços terceirizado nem por quaisquer problemas fora da rede do Google.

Agenda

Visão geral do Google Compute Engine

VPC do Google Cloud Platform

Recursos importantes da VPC

[Teste e laboratório](#)

Questionário

Cite 3 serviços de rede robustos disponíveis para seus aplicativos no Google Cloud Platform.

Nome 3 Recursos de preços do Compute Engine.

Verdadeiro ou falso: o Google Cloud Load Balancing permite equilibrar o tráfego HTTP em várias regiões do Compute Engine.

Respostas do questionário

Cite 3 serviços de rede robustos disponíveis para seus aplicativos no Google Cloud Platform.

Cloud Virtual Network, Cloud Interconnect, Cloud DNS, Cloud Load Balancing e Cloud CDN.

Nome 3 Recursos de preços do Compute Engine.

Verdadeiro ou falso: o Google Cloud Load Balancing permite equilibrar o tráfego HTTP em várias regiões do Compute Engine.

Respostas do questionário

Cite 3 serviços de rede robustos disponíveis para seus aplicativos no Google Cloud Platform.

Cloud Virtual Network, Cloud Interconnect, Cloud DNS, Cloud Load Balancing e Cloud CDN.

Nome 3 Recursos de preços do Compute Engine.

Faturamento por segundo, tipos de máquina personalizados, instâncias preemptivas.

Verdadeiro ou falso: o Google Cloud Load Balancing permite equilibrar o tráfego HTTP em várias regiões do Compute Engine.

Respostas do questionário

Nomeie 3 serviços de rede disponíveis para seus aplicativos no Google Cloud Platform.

Cloud Virtual Network, Cloud Interconnect, Cloud DNS, Cloud Load Balancing e Cloud CDN.

Nome 3 Recursos de preços do Compute Engine.

Faturamento por segundo, tipos de máquina personalizados, instâncias preemptivas.

Verdadeiro ou falso: carregamento do Google Cloud

Verdadeiro. O balanceamento permite equilibrar o tráfego HTTP em várias regiões do Compute Engine.

Instruções de laboratório

Neste laboratório, você criará instâncias de máquina virtual (VM) e se conectará a elas. Você também se conectará entre as duas instâncias.

- Crie uma máquina virtual do Compute Engine usando o Console do Google Cloud Platform
- Crie uma máquina virtual do Compute Engine usando a interface de linha de comando gcloud
- Conecte-se entre as duas instâncias

Mais recursos

Google Compute Engine

<https://cloud.google.com/compute/docs/>

VPC do Google Cloud Platform [https://](https://cloud.google.com/compute/docs/vpc/)

cloud.google.com/compute/docs/vpc/

Google Cloud Stackdriver [https://](https://cloud.google.com/stackdriver/docs/)

cloud.google.com/stackdriver/docs/

Repositórios de origem do Google Cloud

gcloudguia de ferramentas

<https://cloud.google.com/source-repositories/docs/>



© 2017 Google Inc. Todos os direitos reservados. Google e o logotipo do Google são marcas registradas da Google Inc. Todos os outros nomes de empresas e produtos podem ser marcas registradas das respectivas empresas de propriedade intelectual.

© 2017 Google Inc. Todos os direitos reservados. Google e o logotipo do Google são marcas registradas da Google Inc. Todos os outros nomes de empresas e produtos podem ser marcas registradas das respectivas empresas de propriedade intelectual.