Ambiente de computação em nuvem – Google Cloud

Utilização das funcionalidades do Google Cloud para benefícios de performance e disponibilidade de aplicações e infraestrutura em ambientes de computação em nuvem.

Prof. Iago Leoni

Propósito

Conhecer as soluções providas pelo Google para ambientes de computação em nuvem é importante para performance de aplicações e infraestruturas, entendendo as melhores práticas de segurança, alta disponibilidade e desenvolvimento.

Objetivos

- Reconhecer os diferenciais dos ambientes como serviços do Google Cloud.
- Reconhecer os principais pontos de configuração e implantação de infraestrutura e aplicativos na plataforma do Google Cloud.
- Identificar as melhores práticas de segurança e operações em ambientes de nuvem do Google.
- Analisar soluções práticas com uso do Google Cloud.

Introdução

Empresas de todos os nichos estão buscando migrar suas infraestruturas e aplicações para ambientes de nuvem, que oferecem tecnologias como serviço, ou seja, sem a necessidade de instalações e gerenciamento, pagando apenas pelos recursos utilizados.

Esses novos serviços trazem inúmeras arquiteturas diferentes para todos os tipos de aplicativos, assim como para qualquer infraestrutura que seja necessária. Por isso, é muito importante estar a par de todas essas possibilidades e ser capaz de identificar a melhor solução para cada um dos desafios.

Neste conteúdo, exploraremos as principais características do Google Cloud, um dos principais fornecedores do serviço de computação em nuvem.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Entendendo computação em nuvem

A ideia de computação em nuvem não é algo novo, e vem se destacando nos últimos anos, devido às possibilidades e facilidades que agrega em todo o processo de jornada tecnológica.

Tudo o que utilizamos para o desenvolvimento de aplicações, seja a hospedagem, banco de dados ou ferramentas, deve estar localizado em algum lugar, mais tradicionalmente “no local” (do termo em inglês, on-premises).

Isso significa que toda a infraestrutura é localizada em um ambiente físico gerenciado por quem a utiliza. E essa infraestrutura inclui, além de servidores, também os sistemas de refrigeração, discos rígidos, energia elétrica, espaço físico, internet, cabeamento e mão de obra para suportar todas essas necessidades.

Servidores.

Além da preocupação com o que é tangível, existe todo o processo de gerenciamento dos softwares que suportam os aplicativos e servidores. Isso inclui instalações, configurações e atualizações de sistemas e versões.

> A computação em nuvem provê um ambiente sob demanda, ou seja, toda a capacidade de processamento e desenvolvimento de acordo com o que for necessário, sem a necessidade de se preocupar com o gerenciamento de uma infraestrutura local, tudo a um clique de distância para ter acesso.

Imagine em um data-center on-premises, no qual a quantidade de servidores chegou no limite de processamento e, para conseguir suprir, será necessário comprar mais uma unidade. Nesse processo, existem inúmeras preocupações além da compra de um novo servidor, como também onde ele ficará localizado. Ainda existe espaço físico que suporte? Existem tomadas para isso? Quem instalará e gerenciará todos os softwares necessários, como sistemas operacionais? E quando não se precisar mais desse servidor, o que acontecerá com o investimento que já foi feito?

Com a computação em nuvem, essas preocupações não existem, pois todos os recursos necessários, máquinas com diferentes programações de CPU e memória estão disponíveis sob demanda, alugando recursos somente quando necessário e, se um dia não forem mais necessárias, podemos simplesmente excluir essas máquinas, sem a necessidade de continuar pagando.

Por que aprender nuvem e Google Cloud?

Neste vídeo, daremos um panorama geral e fala sobre o impacto tecnológico que a nuvem vem causando nos últimos 10 anos na inovação de TI.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Soluções como serviço

A chegada da computação em nuvem trouxe novas ofertas para o mundo da tecnologia. Algumas delas são a infraestrutura como serviço (infrastructure as a service, ou IaaS), a plataforma como serviço (platform as a service, ou PaaS) e o software como serviço (software as a service, ou SaaS).

Entenda a seguir:

Iaas (infraestrutura como serviço)

É o modelo que oferece infraestrutura sob demanda, alugando recursos computacionais como armazenamento, rede e virtualização. Esse modelo ajuda a eliminar grande parte da complexidade e custos associados à criação e manutenção de infraestrutura física de um datacenter próprio.Mesmo utilizando IaaS do Google Cloud, o cliente ainda tem a responsabilidade de gerenciar a camada de sistema operacional, dados e aplicação.

PaaS (plataforma como serviço)

É um modelo que engloba a IaaS, mas com uma camada de abstração maior, tirando toda responsabilidade do ponto de vista de operações e infraestrutura, ou seja, o usuário não precisa se preocupar com nada além da sua aplicação. Enquanto na IaaS ainda existe uma responsabilidade do ponto de vista de sistema operacional e dados, por exemplo, na PaaS isso é desnecessário, visto que tudo é provido sob demanda.

SaaS (software como serviço)

É o serviço mais gerenciado possível, no qual o usuário vai apenas utilizar o software, sem necessidade de gestão, instalação ou até mesmo o desenvolvimento. O SaaS é uma oferta que visa acelerar ainda mais o processo de adequação à nuvem, já que o tempo para começar a utilizar esse serviço é praticamente imediato. Por outro lado, trata-se de um serviço com pouquíssima flexibilidade, no qual o usuário não tem controle e não pode opinar sobre nenhum componente de infraestrutura ou desenvolvimento.

Cada uma dessas ofertas possui pontos positivos e negativos, sobre os quais devemos identificar a necessidade da aplicação, do time ou da empresa para utilização de algum provedor como serviço. Quanto mais se abstrai de responsabilidades de gerenciamento, menos flexibilidade se tem no ambiente.

Exemplo

Em uma IaaS, se toda a infraestrutura é provida sob demanda e gerenciada pelo Google, como podemos escolher qual o hardware utilizar? Isso não é possível, visto que é responsabilidade do provedor. Da mesma maneira, em uma PaaS, além de estar em uma IaaS, também agrega os softwares de plataforma, sendo assim, como vou alterar algo como módulos, bibliotecas ou versão do sistema operacional? Também não é possível. Então, imagine que uma aplicação só pode ser implantada em um sistema operacional em uma versão específica ou com bibliotecas customizadas e, nesse caso, utilizar uma plataforma como serviço não seria o ideal.

Sempre devemos identificar primeiro as necessidades para depois escolher em qual modelo de serviço o aplicativo estará hospedado.

O que é a computação em nuvem do Google Cloud?

Neste vídeo, falaremos sobre a história e os conceitos básicos da computação em nuvem, diferenças entre IaaS, PaaS e SaaS, abordando seus pontos positivos tecnicamente, e o quão impactante ela é para o negócio.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

A computação em nuvem do Google

O Google é uma empresa de tecnologia fundada em 1998 e que hoje atua em diversos segmentos além do mais conhecido, o buscador. O provedor de nuvem, conhecido como Google Cloud Platform, ou GCP, foi lançado em 2008. Desde então, vem agregando novas tecnologias ao mercado.

> Um provedor de nuvem disponibiliza uma infraestrutura como serviço. Isso significa que o Google precisa ter datacenters físicos em algum lugar para que possa prover esses recursos computacionais. As chamadas regiões são os locais onde estão localizados esses datacenters, o que pode ser em países ou estados diferentes.

Por exemplo, no Brasil existe um datacenter localizado em São Paulo, conhecido como southamerica-east1. Dentro de cada região, existe uma divisão chamada de “zonas”, o que representa uma divisão de servidores, denominados “a”, “b” e “c”, e isso visa garantir a disponibilidade da sua aplicação dentro de um único servidor, assim podemos ter redundância do nosso software, tendo uma maior disponibilidade em caso de falhas de servidores na região.

Na imagem a seguir, podemos ver as regiões disponíveis no Google Cloud Platform.

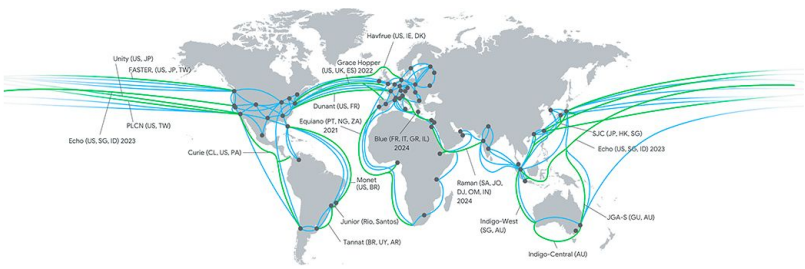
Regiões do Google Cloud Platform em dezembro de 2022.

Com todos esses datacenters espalhados pelo mundo, talvez você esteja se perguntando: como é a comunicação entre eles? Diferentemente de outros provedores de nuvem, o Google possui rede própria de cabos submarinos interligando todas as regiões, isso agrega muita segurança e agilidade na entrega de recursos. Pense que todos os dados que trafegam entre os datacenters do Google estão em uma rede privada, sem a necessidade de passar pela rede tradicional da internet.

Curiosidade

Todos os sites e produtos do Google utilizam a mesma infraestrutura do Google Cloud. Aplicações que possuem bilhões de usuários, como o YouTube, Gmail e o Google.com, estão hospedados e utilizam essa mesma rede que é ofertada para outros clientes.

A imagem a seguir ilustra os cabos de conexão entre datacenters do Google ao redor do mundo.



Rotas de cabos do Google em dezembro de 2022.

Por meio de toda essa infraestrutura, o Google consegue disponibilizar todos os seus produtos e ofertar seu provedor de nuvem **Google Cloud Platform**, com tecnologias como serviço para locação de recursos computacionais, processamento de aprendizado de máquinas, inteligência artificial e plataformas para hospedagem e implantação de aplicativos.

Como são os servidores do Google Cloud?

Neste vídeo, daremos um panorama de como funcionam os servidores do Google Cloud, qual a função de uma região e uma zona dentro da computação em nuvem.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

A computação em nuvem trouxe um novo modelo de utilizar recursos computacionais, no qual

A

os recursos como CPU e memória devem ser adquiridos por pacote.

B

os recursos como CPU e memória são alugados e pagos somente pelo que é utilizado.

C

deve-se ter um servidor local gerenciado pelo usuário.

D

os recursos como CPU e memória são alugados, mas a energia e instalação do hardware é responsabilidade do usuário.

E

o hardware é do usuário, mas é gerenciado pelo provedor de nuvem.

A alternativa B está correta.

A computação em nuvem trouxe o modelo de pagar apenas pelo que se é utilizado. Todo o recurso computacional, como CPU e memória, é alugado, podendo adquirir mais ou "devolver" conforme necessário.

Questão 2

Utilizar o Google Cloud possibilita usufruir de tecnologias como serviço, por exemplo:

A

DaaS (domain as a service) e PaaS (platform as a service).

B

PaaS (program as a service) e IaaS (infrastructure as a service).

C

PaaS (platform as a service) e IaaS (infrastructure as a service).

D

IaaS (internet as a service) e PaaS (platform as a service).

E

CaaS (cloud as a service) e IaaS (internet as a service).

A alternativa C está correta.

O Google Cloud, assim como outros provedores de nuvem, disponibiliza tecnologias como serviço, entre elas está a PaaS, plataforma como serviço (platform as a service) e a IaaS, infraestrutura como serviço (infrastructure as a service), além do Saas, software como serviço (software as a service).

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:

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.

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.

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.

Redes, infraestrutura e aplicações: como funcionam na nuvem?

Neste vídeo, falaremos sobre rede, infraestrutura e aplicações na nuvem, explicando por que é essencial entender e executar as melhores práticas.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

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:

1

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.

2

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.

3

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:

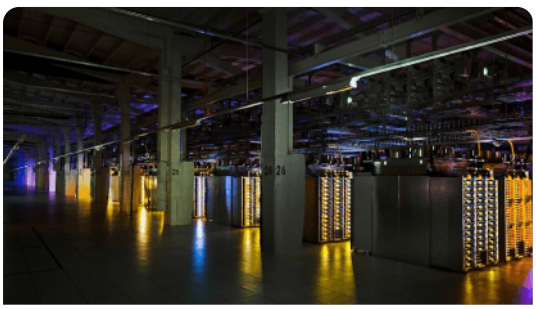
Tipo de uso	Econômico	Equilibrado	Escalonamento horizontal otimizado	Otimização de memória	Otimizad comput
Família	E2	N2, N2D, N1	Tau T2D, Tau T2A	M3, M2, M1	C2, C
Objetivo	Computação básica a um custo menor	Desempenho e preços equilibrados	Melhor desempenho para cargas que precisam de escalabilidade horizontal	Cargas de trabalho com memória ultraelevada	Ultradeser para carg trabalho demanda intensivo

Tipo de uso	Econômico	Equilibrado	Escalonamento horizontal otimizado	Otimização de memória	Otimizad comput
Família	E2	N2, N2D, N1	Tau T2D, Tau T2A	M3, M2, M1	C2, C
Exemplos	• Aplicação web • Front-end • Banco de dados pequenos • Ambientes para desenvolvimento • API • Microsserviços	• Streaming de mídia • Banco de dados médios e grandes • Aplicações web • Cache	• Aplicativos Java em grande escala • Microsserviços em containers	• Banco de dados de análises em memória • Bancos de dados como Microsoft SQL Server	• J • Com d dese (• Inte ar • Apl web dese

Tabela: Famílias e classificações de máquinas virtuais Google Cloud. cloud.google.com.

Após selecionarmos o tipo de máquina e aprovisionarmos 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 aprovisionamos 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 aprovisionamos 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.

Qual a função de uma máquina virtual?

Neste vídeo, falaremos sobre como funciona uma máquina virtual do ponto de vista de arquitetura, melhores práticas e operação, fazendo um panorama geral dos diferentes tipos de máquinas e qual a usabilidade de cada uma.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

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.

Como desenvolver aplicações para a nuvem?

Neste vídeo, falaremos sobre as especificações de uma aplicação nativa de nuvem e quais seus benefícios, explicando por que é necessário desenvolver uma aplicação preparada para um ambiente de nuvem e como isso pode impactar todo o processo de desenvolvimento.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

Qual a sigla e o nome dado à rede virtual em ambiente de nuvem do Google Cloud?

A

VPC – virtual private connect.

B

VPN – virtual private network.

C

VPC – virtual public cloud.

D

VPC – virtual private cloud.

E

VPN – virtual public network.

A alternativa D está correta.

VPC é a abreviação da expressão em inglês virtual private cloud (nuvem virtual privada), que é responsável por ser uma interface do ambiente Google Cloud com componentes externos.

Questão 2

Qual o nome dado para o ambiente em nuvem virtualizado em hardware físico?

A

Virtual machine.

B

Virtual cloud.

C

Virtual components.

D

Virtual network.

E

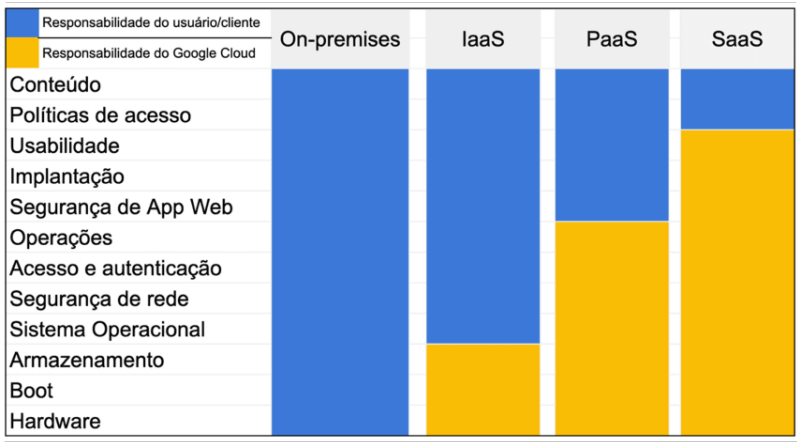
Virtual notebooks.

A alternativa A está correta.

Virtual machine (máquina virtual) ou VM é o ambiente em nuvem no qual virtualiza-se um servidor com recursos de CPU, memória e armazenamento, sendo utilizado para hospedagem de aplicações.

Segurança em Google Cloud

O Google Cloud, por mais que seja um provedor de serviços gerenciados, necessita estar alinhado com boas práticas de segurança. Dentro de uma matriz de responsabilidades, em um provedor, existem diferentes níveis para cada tipo de serviço, e cada parte terá sua responsabilidade pela segurança. A imagem a seguir nos mostra essa matriz:



Matriz de responsabilidades de segurança.

Como vemos na imagem, quanto mais gerenciado for o serviço, menor a responsabilidade do usuário com a segurança. Assim, ao utilizar uma infraestrutura como serviço (IaaS), por exemplo, toda a responsabilidade de segurança com hardware e boot é do provedor. Por outro lado, a segurança da aplicação, da implantação dessa aplicação, sistema operacional e rede é totalmente do usuário. Caso utilize uma plataforma como serviço (PaaS), a responsabilidade passa a ser somente em nível de aplicação.

> Em um provedor de nuvem, é importante estarmos atentos às normas nas quais ele está adequado. As normas regulamentadoras servem para assegurar um serviço prestado, e podem ser internacionais ou nacionais.

No Brasil, por exemplo, temos a Lei Geral de Proteção de Dados, também conhecida como LGPD, que visa normatizar a utilização e processamento de dados de maneira correta, impondo regras na transferência e circulação de dados pessoais.

Outra norma muito comum é a ISO/IEC 27001, na qual a Organização Internacional de Normatização (ISO) descreve os requisitos de um sistema de gestão de segurança e especifica um conjunto de práticas recomendadas, mostrando detalhes sobre os controles exigidos. O Google Cloud está adequado a essas e inúmeras outras normas, com servidores certificados.

ISO 27001: referência internacional para gestão da segurança da informação.

Toda infraestrutura física do Google Cloud, em todos os servidores, tem um protocolo de segurança extremamente rígido. Para acesso aos datacenters, por exemplo, existem seis camadas de segurança, desde a autorização para entrar no local, passando por câmeras de detecção térmica, segurança 24h, sistemas de alarme, até chegar no acesso ao piso de servidores. Um fato interessante é que apenas funcionários autorizados

possuem acesso aos datacenters, e esses funcionários representam menos de 1% do total de funcionários do Google.

Como é a segurança em ambientes Google Cloud?

Neste vídeo, falaremos sobre as responsabilidades de cada persona a respeito da segurança no ambiente de nuvem em cada tipo de serviço (IaaS, PaaS e SaaS). Além disso, ele aborda as melhores práticas executadas atualmente.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Segurança na nuvem

A computação em nuvem trouxe grandes facilidades e agilidade para a inovação. Da mesma maneira, com as ofertas como serviço, tirou do usuário grande parte da responsabilidade de segurança. Mesmo assim, ainda é necessário adotar medidas para garantir a seguridade das aplicações, dados e infraestrutura.

Ameaças

A grande realidade é que, atualmente, existem incontáveis tipos de ameaças, e isso é uma grande preocupação e ponto de atenção para todas as empresas. Qualquer tipo de ataque, seja para sequestros de dados ou para gerar indisponibilidade, impacta diretamente, desestabilizando o negócio.

Segurança na nuvem

Devido a todas essas preocupações, surgiu um termo chamado de “segurança na nuvem” (cloud security), que se refere às melhores práticas de gerenciamento de segurança na nuvem, incluindo tecnologias, políticas, controles e serviços, que mantém dados, aplicativos e infraestrutura sob a maior proteção possível contra ameaças externas e internas.

A segurança na nuvem está muito relacionada a tecnologias e processos que podem ser utilizados, como: gestão de acesso e identidade, segurança contra ataques, lista de bloqueio. Vamos ver como funcionam?

Gestão de acesso e identidade

Também conhecido como IAM (identity and access management), é um processo que visa controlar o acesso às informações, recursos e ações do ambiente. O Google Cloud oferece sua solução de IAM integrada à plataforma e, com isso, é possível gerir todos os usuários e seu nível de acesso.

Exemplo

Se um usuário tem acesso de desenvolvedor, estará autorizado a somente visualizar ou talvez editar configurações relacionadas à sua alçada, não podendo visualizar, utilizar ou configurar nada de infraestrutura, tecnologias ou aplicações que não são de sua responsabilidade.

O login pode ser combinado com mais uma camada de segurança, com múltiplos fatores de autenticação, ou seja, além do padrão usuário e senha, é necessário mais um fator para comprovar a identidade do usuário como, por exemplo, um SMS no telefone celular, ou um token físico como um pen drive.

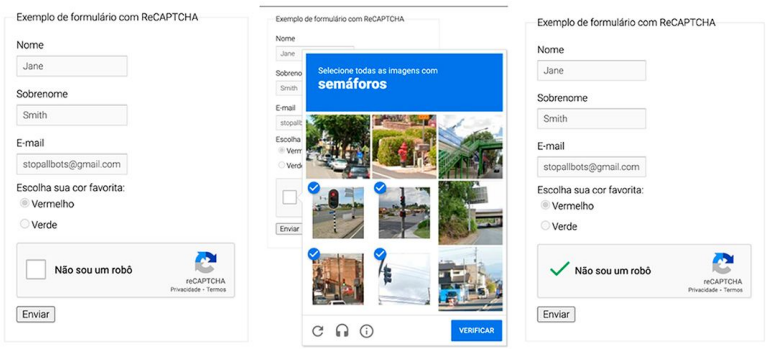
Com esse nível de gestão, utilizando políticas e limites de acesso, evita-se qualquer possibilidade de alguém não autorizado utilizar algo. Ou até mesmo, caso alguma credencial seja roubada, as limitações de acessos dão maior segurança.

Segurança contra ataques

Empresas vêm regularmente sofrendo ataques e, para se proteger contra isso, é necessário uma boa estratégia e ferramentas. O Google Cloud possui uma oferta chamada de Cloud Armor, que consiste em uma tecnologia que utiliza inteligência artificial e aprendizado de máquinas para mitigar ataques contra aplicações e servidores dos clientes.

> Por meio de inúmeros dados processados pelo Google, o Cloud Armor consegue entender quando um comportamento é uma ameaça e evitá-lo. Por exemplo, um ataque de negação de serviço distribuída (DDoS), que ocorre por meio de um grande volume de requisições falsas para o serviço, gerando indisponibilidade.

Outra tecnologia que é muito utilizada para segurança contra ataques às aplicações é o reCAPTCHA, uma tecnologia do Google que permite distinguir entre um acesso humano ou automatizado por meio do uso de identificações visuais ou auditivas. Isso é importante pois, a cada dia, cresce o número de bots maliciosos, que visam buscar brechas em sistemas.



Exemplos de reCAPTCHA.

Lista de bloqueio

Outra estratégia muito utilizada é a de lista de bloqueios (blocklist) e lista de permitidos (allowlist). Seu objetivo é ter uma lista de bloqueios ou permissões de acessos ao sistema, e isso pode ser baseado em IPs ou regiões.

Essa estratégia é muito útil para podermos negar o acesso de IPs externos da nossa rede, por exemplo, a uma aplicação. Outra possibilidade é a de bloquear todo acesso que venha de outro país, o que mitiga ataques em massa que utilizam servidores, bots e computadores do mundo todo para atingir nossos ambientes.

Quais as responsabilidades de segurança na nuvem?

Neste vídeo, falaremos sobre a responsabilidade do servidor e do cliente quanto a cada serviço prestado (IaaS, PaaS e SaaS) e qual a diferença de manter a segurança em um servidor próprio.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Segurança nas aplicações

Como vimos, em ambientes de nuvem utilizando soluções como serviço IaaS ou PaaS, a camada de aplicação é responsabilidade do usuário e, por isso, devemos tomar todas as medidas de segurança possíveis.

> Um consenso básico de segurança das aplicações é sempre manter as versões de linguagens de programações e frameworks de código livre (open source) atualizados para a versão estável, o que não significa necessariamente que essa seja a última disponível. Sempre que uma nova versão de algo é lançada, normalmente ela não é estável, pois ainda não teve o período de amadurecimento e, devido a isso, pode possuir brechas de segurança que ainda não foram descobertas.

Outro conceito bem maduro e útil é o DevOps, que é um processo de entrega de software que envolve desenvolvimento e operações juntos. Esses dois times possuem visões diferentes. Enquanto desenvolvedores se preocupam com regras de negócio, linguagens de programação e construir a aplicação, os operadores estão visando à performance, segurança geral e monitoramento do ambiente e, assim, o desempenho de entrega de softwares acaba tornando complexo todo o processo.

Por isso, a teoria do DevOps busca envolver ambos os lados e compartilhar a responsabilidade utilizando, para isso, tecnologia para um desenvolvimento em ciclos. Essa técnica agrega a segurança da aplicação, pois cada persona se preocupa com a segurança de sua especialidade. Desse modo, o desenvolvedor não precisa se preocupar com as configurações de infraestrutura dentro da aplicação, pois as tecnologias padronizam os ambientes. Todo processo de entrega de novas versões de software acaba se tornando mais seguro e documentado, pois existe uma prática por trás, e isso torna-se mais nítido para os envolvidos.

DevOps: utilização de tecnologia para desenvolvimento em ciclos.

Todas essas práticas são recomendadas e adotadas pelo Google Cloud, que também fornece ferramentas e tecnologias para sua execução, como: Cloud Repository, um versionador de códigos fontes GIT; o Cloud Build, um integrador e construtor de aplicações com base em códigos-fontes, que têm como função automatizar todo o processo; e Cloud Deploy, ferramenta de entrega de aplicações em ambientes finais (produtivos) em nuvem.

Como garantir a segurança das aplicações?

Neste vídeo, falaremos sobre as melhores práticas para garantir a segurança de aplicações em ambiente de nuvem pensando na infraestrutura que será utilizada e nas ferramentas (linguagem de programação, ferramentas utilitárias, plataformas) que serão adotadas.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Operações em Google Cloud

As operações em nuvem consistem no processo de operar ou executar uma infraestrutura em ambiente de nuvem. Isso tudo é complementado com o gerenciamento geral para garantir e assegurar que todos os processos e recursos estejam normatizados, a fim de manter essas operações funcionando, garantindo o máximo desempenho e a disponibilidade para satisfazer as necessidades e expectativas dos clientes.

> Uma das principais responsabilidades de operações é o monitoramento e, para que isso seja feito, é necessário identificar as métricas que devem ser analisadas e as tecnologias que podem ser utilizadas.

O monitoramento mais básico possível ocorre por meio de LOGS que, por definição, consistem no registro de eventos relevantes no sistema ou infraestrutura, de maneira geral. Ou seja, é todo o histórico de acontecimentos importantes, com data e hora registrada. Com o uso dos LOGS, é possível identificar erros, acessos e alterações realizadas, para que, a partir dessas informações, seja possível fazer um trabalho preventivo e corretivo, quando necessário.

Cloud Monitoring: monitoramento de aplicativos e da infraestrutura.

O Google Cloud disponibiliza ferramentas do **Cloud Monitoring** para monitoramento de aplicações e infraestrutura de maneira mais minuciosa, coletando um grande volume de dados para disponibilização, pois o monitoramento de um ambiente é um dos papéis mais importantes em nuvem.

Por meio dessa ferramenta é possível traçar estratégias para melhoria contínua, tomar uma ação preventiva para evitar qualquer problema ou empecilho que possa prejudicar a operação do sistema e, caso ocorra, identificar a falha por meio do registro histórico e agir o mais rápido possível.

O Cloud Monitoring permite coletar dados e métricas de latência, por exemplo, o que nos indicará qual o tempo médio de transações e comunicação entre servidores. Indica também a quantidade de acessos e requisições em nossas aplicações, para que possamos entender e programar a infraestrutura para o volume que está sendo monitorado.

Saiba mais

O Google acredita que a prevenção é sempre mais vantajosa que a remediação. Qualquer problema impacta diretamente o negócio, receita e claramente o time. Estar constantemente analisando as métricas de operações previne esse tipo de problema e auxilia na hora de tomar uma decisão final, que será embasada em dados reais.

Operações e gerenciamento em ambientes Google Cloud

Neste vídeo, falaremos sobre as principais métricas analisadas em um ambiente de nuvem (consumo de memória, CPU, latência e requisições) e como essas métricas impactam nossas aplicações e infraestrutura.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

Considerando a matriz de responsabilidade do Google Cloud, quem é o responsável por gerenciar a segurança do hardware em um ambiente on-premises (“no local”)?

A

Inteiramente o Google Cloud.

B

Parcialmente o cliente/usuário.

C

Inteiramente o cliente/usuário.

D

Parcialmente o Google Cloud.

E

Compartilhada entre Google Cloud e cliente/usuário.

A alternativa C está correta.

Em um ambiente on-premises, a responsabilidade de gerenciamento de segurança do hardware como atualizações, versões e acesso é inteiramente do cliente/usuário, diferentemente de uma opção de infraestrutura como serviço (IaaS) fornecida pelo Google, na qual a responsabilidade seria inteiramente do Google Cloud.

Questão 2

Qual o nome e a sigla do processo que visa controlar o acesso a informações, recursos e ações em ambientes Google Cloud?

A

Controle de acesso ao ambiente/IAA.

B

Controle de acesso Google Cloud/IAM.

C

Controle de acesso e identidade/IMA.

D

Controle de identidade Google Cloud/GCP.

E

Controle de acesso e identidade/IAM.

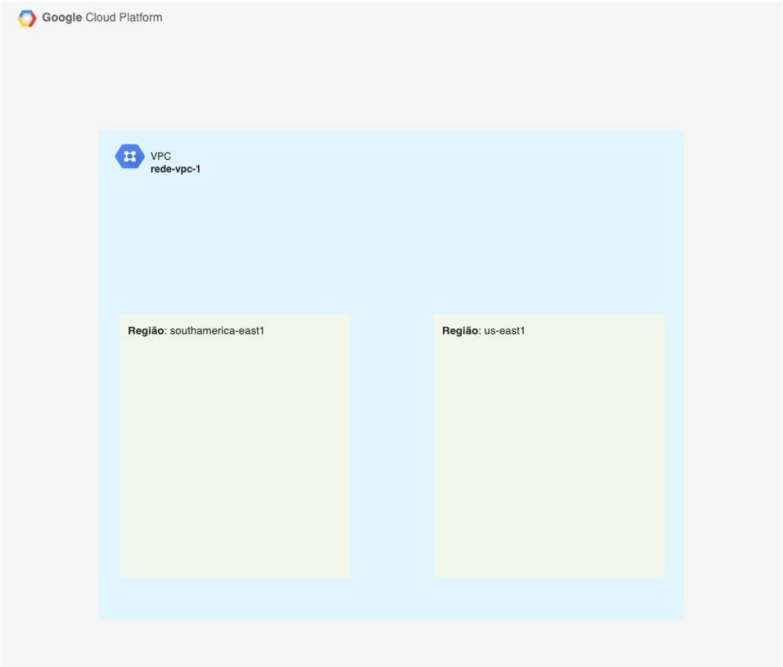
A alternativa E está correta.

O controle de acesso e Identidade, conhecido como IAM (identity and access management), é o responsável por controlar o acesso a informações, recursos e ações em ambientes Google Cloud.

Aplicação da arquitetura de infraestrutura

Como vimos, existem práticas dentro do ambiente de nuvem do Google para desenvolvimento de uma arquitetura de infraestrutura, como redes virtuais em nuvem e máquinas virtuais, e que podem ser divididas em regiões diferentes.

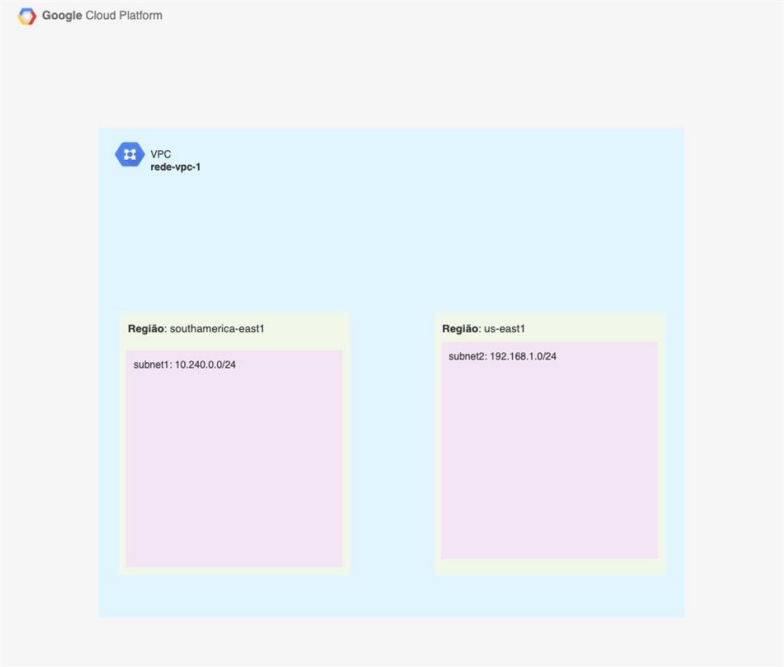
O primeiro ponto, quando aplicamos a prática desses conceitos juntos, é a definição das redes necessárias dentro de um projeto e, por isso, iniciamos com a VPC.



Arquitetura de VPC em Google Cloud.

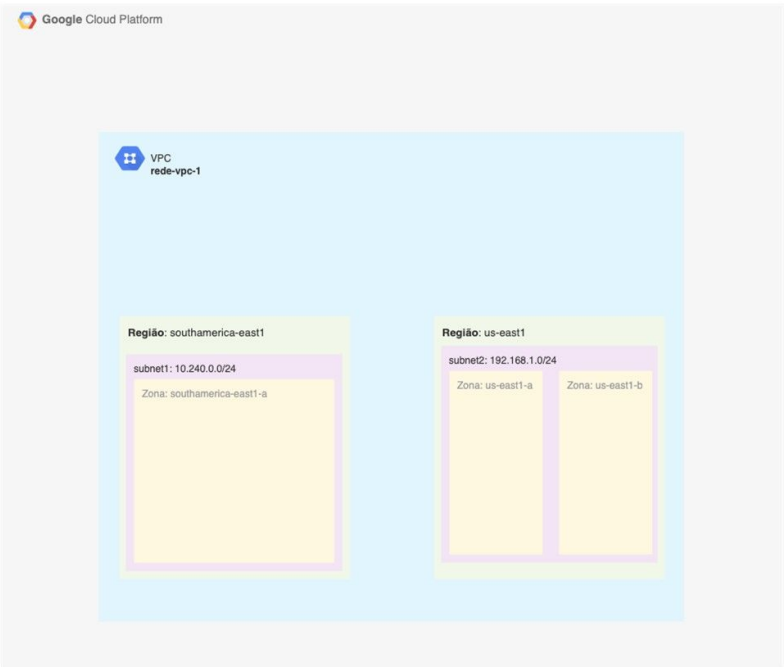
Como vemos na imagem, foi definida uma rede VPC chamada de “rede-vpc-1”, que está expandida em duas regiões: “southamerica-east1”, localizada no Brasil, e “us-east1” localizada nos Estados Unidos.

O próximo passo é definir as sub-redes (subnets) com as possibilidades de variações (ranges) de IP e, como demonstramos na imagem a seguir, temos duas subnets: “subnet1” e “subnet2”, cada qual com seu range de IP.



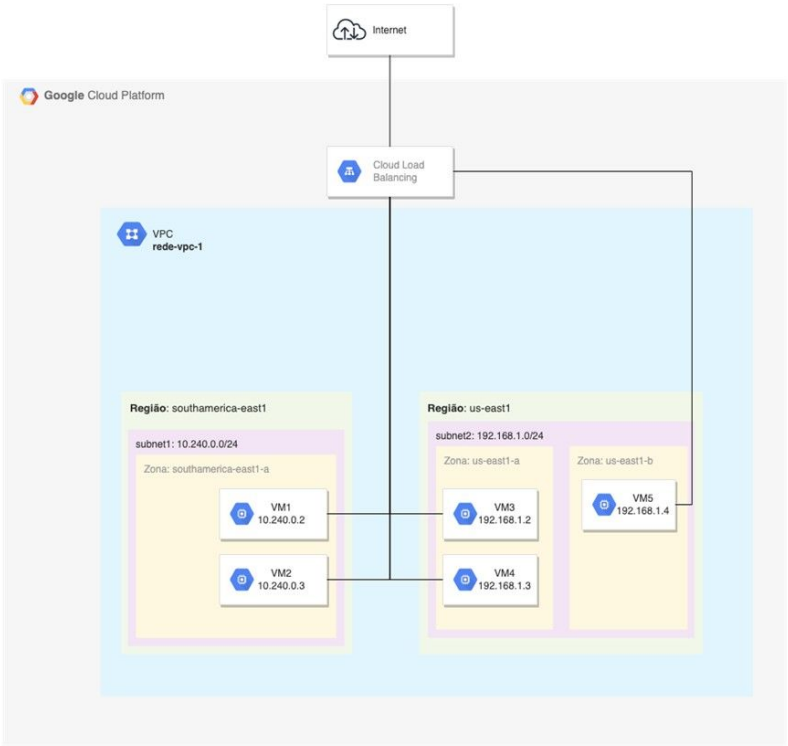
Subnets Google Cloud.

Dentro das nossas subnets podemos definir quais serão as zonas de disponibilidades utilizadas nas regiões. Para esse projeto, definimos utilizar na “subnet1” somente a zona denominada “southamerica-east1-a”, enquanto na “subnet2”, vamos utilizar duas zonas, “us-east1-**a**” e “us-east1-**b**”.



Zonas na arquitetura de infraestrutura.

Com esse desenho básico de rede, já podemos começar a adicionar as funções de infraestrutura para a aplicação, como nossas VMs. As máquinas virtuais ficam diretamente relacionadas a uma subnet de uma VPC, e agregam em si os valores de ranges de IP preestabelecidos para cada uma dessas subnets. A partir disso, já podemos disponibilizar nossas aplicações para a internet.



Arquitetura de rede com máquinas virtuais.

Como um website funciona na prática em Google Cloud?

Neste vídeo, falaremos sobre como funcionaria um website em ambiente Google Cloud. Explica também importância de todo conteúdo aprendido e sua aplicabilidade na prática, expondo que isso é o que o mercado de trabalho desenvolve.

> Conteúdo interativo
>
> Acesse a versão digital para assistir ao vídeo.

Aplicação da arquitetura de aplicativos

Conforme já vimos, as aplicações precisam estar adequadas às infraestruturas que serão executadas da melhor forma possível. Para ambientes de nuvem, utilizamos uma arquitetura chamada de **nativa de nuvem (cloud native architecture)**, cujo objetivo é se adequar às tecnologias como serviço oferecidas pelo provedor de nuvem.

> Dentro do Google Cloud, existem opções para facilitar o desenvolvimento e agregar valor ao negócio, assim como impulsionar e agilizar todo o processo de criação de aplicativos.

Vamos imaginar um cenário: um site web de compras, e-commerce, que possui em sua arquitetura as funções de página do produto e finalização de compras, o check-out, com emissão de nota fiscal. Apenas nessas duas funções já existem inúmeros desafios a se enfrentar do ponto de vista de desenvolvimento. Vamos exercitar alguns a seguir.

Página do produto

A função da página do produto é mostrar seu nome, descrição, preço e o estoque disponível. Porém, devemos saber a origem dessas informações com consistência e disponibilidade, evitando problemas como, por exemplo, ter inúmeras pessoas na mesma página e isso ocasionar uma indisponibilidade do serviço por alta

demanda ou até mesmo apresentar inconsistências. Para sanar isso, precisaremos de um banco de dados organizado em tabelas, sendo necessário escalar, ou seja, tornar mais “disponível” em altas demandas, para cenários de grande volume de acesso. Para essa estratégia, podemos aprovisionar um banco de dados gerenciado pelo usuário, ou como estamos em ambiente de nuvem, utilizar como serviço.

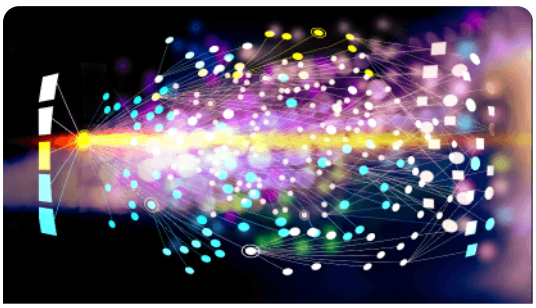
Cloud SQL: banco de dados do Google Cloud.

O Google Cloud oferece o **Cloud SQL**, um banco de dados SQL reacional, completamente gerenciado e com alta disponibilidade, escalável e em nuvem. Com isso, podemos nos concentrar no desenvolvimento da aplicação utilizando alguma linguagem de programação, evitando preocupações com o gerenciamento de infraestrutura para nosso banco de dados.

Finalização de compras e checkout

A finalização da compra é um dos momentos mais importantes de um e-commerce. É o processo final do usuário, e deve ser o mais fluido possível para evitar transtornos. A grande dificuldade desse sistema é depender de outros sistemas para funcionar perfeitamente.

Imagine uma situação em que, ao finalizar uma compra, é necessário notificar outros sistemas financeiros da empresa, dar baixa no estoque, enviar os dados do cliente para o processo de entrega e utilizar um sistema público para emissão de nota fiscal. Imagine a frustração do cliente ao ter que esperar por todo esse processo na tela de checkout, ou pior, algum desses serviços estar indisponível e o cliente não conseguir efetuar sua compra. Isso impactaria diretamente a experiência do usuário e, consequentemente, o negócio. Esse processo, no qual um sistema depende de outro para ser executado, chama-se “comunicação síncrona”, e necessita que os sistemas estejam sincronizados ao mesmo tempo para conseguir um resultado.


Comunicação assíncrona.

Para mitigar problemas como esse utilizamos, dentro da arquitetura nativa de nuvem, uma estratégia chamada de “comunicação assíncrona” (troca de mensagens de maneira assíncrona, por isso também conhecida como “mensageria assíncrona”) que, como o próprio nome diz, não necessita de sincronicidade entre os sistemas para finalizar uma tarefa.

A ideia principal é utilizar uma tecnologia intermediando as comunicações, e que fica responsável por notificar cada sistema dependente dessas informações. Então, no nosso cenário, ao finalizar uma compra, o cliente final não precisa ficar na tela aguardando o processo interno da compra. As informações são enviadas para essa ferramenta distribuidora, que fica responsável por notificar os outros sistemas. Caso algum deles esteja indisponível no momento dessa transação, a ferramenta armazena a mensagem e transmite quando o sistema voltar a ficar disponível.

Atenção

O Google Cloud oferece uma solução como serviço chamada Pub/Sub, uma ferramenta de mensageria assíncrona, completamente gerenciada, que intermedia as comunicações entre sistemas, tornando-as assíncronas. Utilizar o Pub/Sub em uma arquitetura a torna mais desacoplada e torna os serviços dinâmicos, já que não ficam dependentes uns dos outros.

No nosso exemplo, ao finalizar uma compra no e-commerce, uma mensagem é disparada para o Pub/Sub, que distribui para o sistema de emissão de notas fiscais, entregas e estoque e se, por acaso alguns desses sistemas estiverem indisponíveis, a mensagem poderá continuar a ser consumida quando a aplicação estiver disponível novamente.

Comunicação assíncrona e serverless em Google Cloud

Neste vídeo, falaremos sobre o que são comunicações síncronas e assíncronas no desenvolvimento de software, assim como a arquitetura serverless e quando utilizá-la.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

Uma empresa está criando um projeto novo para desenvolver sua aplicação em VMs. Qual o processo correto para disponibilização da infraestrutura com redes nesse projeto?

A

Criar uma VPC, configurar as zonas e criar as VMs.

B

Criar uma VPC e configurar as VMs diretamente.

C

Criar uma VPC, selecionar uma região, criar as subnets, vincular a VM a essa subnet.

D

Criar a VM e vincular a minha rede local.

E

Criar uma VPC, selecionar uma região, criar as VMs e vincular diretamente a subnet.

A alternativa C está correta.

O processo correto é criar uma VPC, selecionar uma região, criar as subnets, vincular a VM a essa subnet para que, dessa forma, as máquinas virtuais estejam vinculadas a uma subnet com IPs que podem ser utilizados.

Questão 2

Qual o nome dado ao processo que visa intermediar mensagens e entregá-las conforme os sistemas estiverem disponíveis?

A

Mensageria tardia.

B

Mensageria síncrona.

C

Mensageria sincronizada.

D

Mensageria assíncrona.

E

Mensageria descentralizada.

A alternativa D está correta.

Mensageria assíncrona, na qual não é necessário sincronicidade entre os sistemas para finalizar uma tarefa. A ideia principal é utilizar uma tecnologia intermediando as comunicações, de maneira que essa tecnologia fica responsável por notificar cada sistema que depende dessas informações.

Considerações finais

Como vimos, a computação em nuvem é uma tecnologia que auxilia grandemente empresas que estão buscando uma inovação tecnológica com agilidade e facilidade. Poder consumir ferramentas como serviço, sem necessidade de gestão e instalação, é algo que traz uma alta agilidade organizacional.

O principal objetivo, ao se utilizar um provedor de nuvem, é a busca por mudanças de maneira ágil, sem a preocupação com gestão de espaço físico, infraestrutura, instalações, segurança e atualizações. A possibilidade de estar a um clique de distância das principais tecnologias do mundo é uma alteração radical do processo de desenvolvimento e inovação em todos os setores de negócio.

O Google Cloud provê tecnologia para gestão, infraestrutura e aplicações, e agrega valor aos negócios dos clientes, possibilitando uma evolução de inovação extremamente ágil, com investimento atrelado ao que for consumido.

Podcast

Neste podcast o especialista recorda os principais pontos desenvolvidos neste tema.

Conteúdo interativo

Acesse a versão digital para ouvir o áudio.

Explore +

Para aprofundar seus conhecimentos sobre o conteúdo estudado:

Assista ao vídeo **Máquinas Virtuais na nuvem**, disponível no canal do Google Cloud Latam, no YouTube, para entender o que são e como funcionam as máquinas virtuais e VPC do Google Cloud.

Assista ao vídeo **Google Data Center Security: 6 Layers Deep** e veja como é estruturada a segurança física de todos os locais de servidores do Google Cloud. Disponível no canal Google Cloud Tech, no YouTube.

Referências

FERREIRA, A. M. **Introdução ao Cloud Computing**: IaaS, PaaS, SaaS, Tecnologia, Conceito e Modelos de Negócio. 1. ed. Lisboa: FCA, 2015.

TAURION, C. **Cloud Computing** – Computação em Nuvem. 1. ed. Rio de Janeiro: Brasport, 2009.

VERGADIA, P. **Visualizing Google Cloud**: 101 Illustrated References for Cloud Engineers and Architects. 1. ed. New York: Wiley, 2022.