

Série de Certificação GCP: Seção 4: Garantindo o funcionamento bem- sucedido de uma solução em nuvem

4.1: 4.1 Gerenciando recursos do Compute Engine



Prashanta Paudel

6 de novembro de 2018 · 25 minutos de leitura

Hoje vamos começar a seção 4. Esta seção lida basicamente com várias tarefas que você deve ser capaz de fazer no GCP para a função de Associate.

Muitas das porções teóricas já estão feitas, então vamos nos aprofundar diretamente em como concluir essas tarefas.

Gerenciando uma única instância de VM (por exemplo, iniciar, parar, editar a configuração ou excluir uma instância)

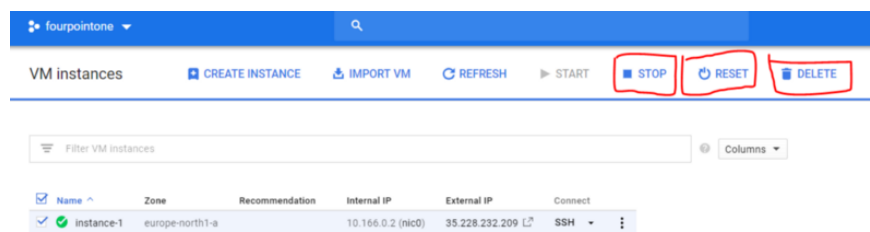
Quando a VM estiver no ar e funcionando, você poderá executar várias operações, como iniciar, parar, editar e excluir.

Inicialmente, quando acionamos a VM pela primeira vez, ela começará a funcionar assim que estiver pronta.

As etapas são bem simples e diretas. Vamos fazer no console e no Shell.

No console

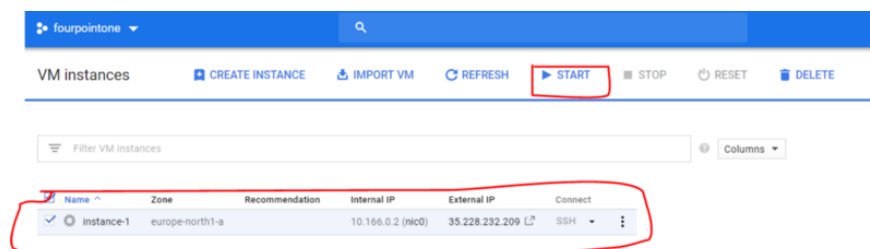
Ir para a instância da VM. Quando mostrado pronto, ele estará em execução para que você possa PARAR ou EXCLUIR VM neste momento.



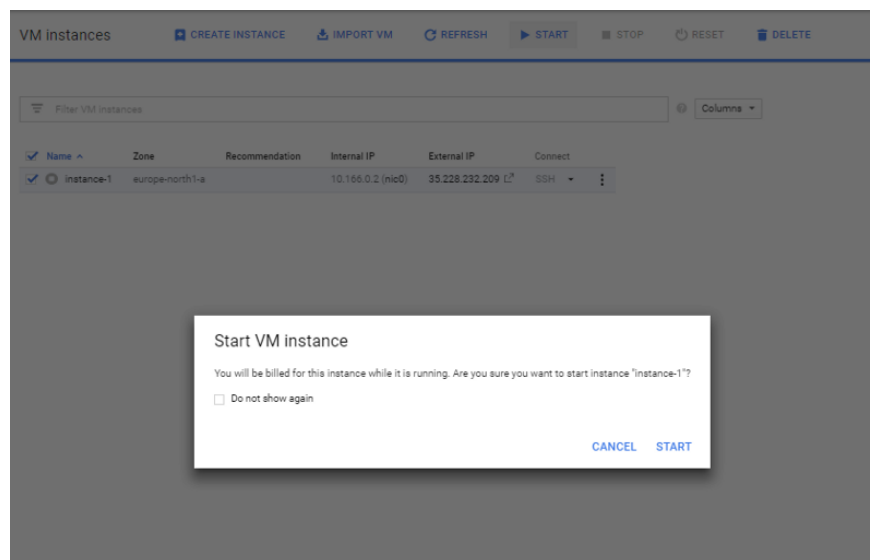
VM

quando você parar a VM

então você verá a opção START ativada.



começar ativado



Iniciar VM

Quando você quiser excluir a VM, basta selecionar a VM e clicar em EXCLUIR na parte superior da página.

Se você quiser editar a configuração, clique no nome da VM e, em seguida, você verá

VM Instance details

[EDIT](#) [RESET](#) [CREATE SIMILAR](#) [STOP](#) [DELETE](#)

Details Monitoring

instance-1

Remote access

SSH [Connect to serial console](#)

☐ Enable connecting to serial ports

Logs

[Stackdriver Logging](#)

Serial port 1 (console)

[More](#)

Machine type

n1-standard-1 (1 vCPU, 3.75 GB memory)

CPU platform

Intel Skylake

Zone

europe-north1-a

Labels

None

Creation time

Nov 6, 2018, 9:47:47 AM

Network interfaces

Name	Network	Subnetwork	Primary internal IP	Alias IP ranges	External IP	Network Tier	IP forwarding	Network details
nic0	default	default	10.166.0.2	—	35.228.156.106 (ephemeral)	Premium	Off	View details

Public DNS PTR Record

None

Firewalls

☒ Allow HTTP traffic

☒ Allow HTTPS traffic

Network tags

http-server, https-server

Deletion protection

☐ Enable deletion protection

When deletion protection is enabled, instance cannot be deleted. [Learn more](#)

Boot disk and local disks

Name	Size (GB)	Type	Encryption	Mode
instance-1	10	Standard persistent disk	Google managed	Boot, read/write

☒ Delete boot disk when instance is deleted

Additional disks

None

Editar VM

Agora basta clicar em Editar e você pode alterar as configurações, anexar um disco rígido à VM, atualizar a memória e assim por diante.

Additional disks (Optional)

Name	Mode	When deleting instance
disk-1	Read/write	Keep disk

[+ Add item](#)

Shielded VM

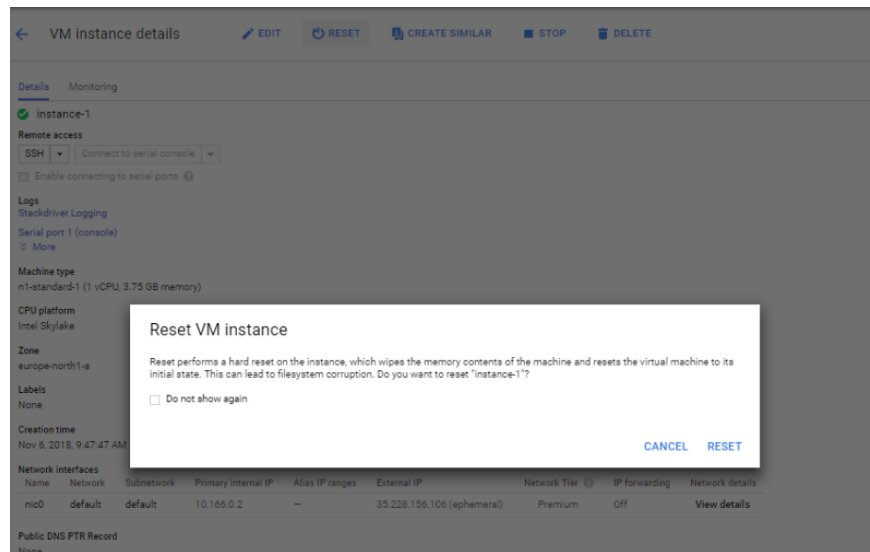
Select a shielded image to use shielded VM features.

Turn on all settings for the most secure configuration.

☐ Turn on Secure Boot

anexar disco

Há uma opção para redefinir a VM, isso apagará todos os dados e reverterá para as configurações padrão.



Redefinição de VM.

Na Shell

```
prashantagcppaudel @ cloudshell: ~ (fourpointone-221707) $
gcloud instâncias de computação create vm1
Você quis dizer zone [europe-west4-a] por exemplo: [vm1] (Y
/ n)? y

Criado em
[https://www.googleapis.com/compute/v1/projects/fourpointone-
221707/zones/europe-west4-a/instances/vm1].
NOME DA ZONA MACHINE_TYPE
PREEPTIBLE INTERNAL_IP EXTERNAL_IP STATUS vm1 europe-west4-a
n1-standard-1 10.164.0.2 35.204.71.54 RUNNING

prashantagcppaudel @ cloudshell: ~ (fourpointone-221707) $
gcloud instâncias de computação stop instance-1 --zone =
europe-north1-a
Parando instância (s) instance-1 ... pronto.
Atualizado
[https://www.googleapis.com/compute/v1/projects/fourpointone-
221707/zones/europe-north1-a/instances/instance-1].

prashantagcppaudel @ cloudshell: ~ (fourpointone-221707) $
gcloud instâncias de cálculo start instance-1 --zone =
europe-north1-a
Iniciando instância (s) instance-1 ... pronto.
Atualizado
[https://www.googleapis.com/compute/v1/projects/fourpointone-
221707/zones/europe-north1-a/instances/instance-1].
prashantagcppaudel @ cloudshell: ~ (fourpointone-221707) $

prashantagcppaudel @ cloudshell: ~ (fourpointone-221707) $
gcloud instâncias de computação delete instance-1 --zone =
```

europa-north1-a

As instâncias a seguir serão excluídas. Todos os discos conectados configurados para serem excluídos automaticamente serão excluídos, a menos que estejam anexados a outras instâncias ou que o sinalizador `--keep-disks` seja fornecido e sejam especificados para manutenção. A exclusão de um disco é irreversível e todos os dados no disco serão perdidos.

```
- [instance-1] in [europa-north1-a]
```

Deseja continuar (S / n)? y

Excluído

```
[https://www.googleapis.com/compute/v1/projects/fourpointone-221707/zones/europa-north1-a/instances/instance-1].
```

```
prashantagcpaude @ cloudshell: ~ (fourpointone-221707) $
```

```
=====
=====
```

SSH / RDP para a instância

O SSH é o meio de acesso mais freqüente e comumente usado para acessar a VM na plataforma linux. RDP é comumente usado para máquinas com Windows

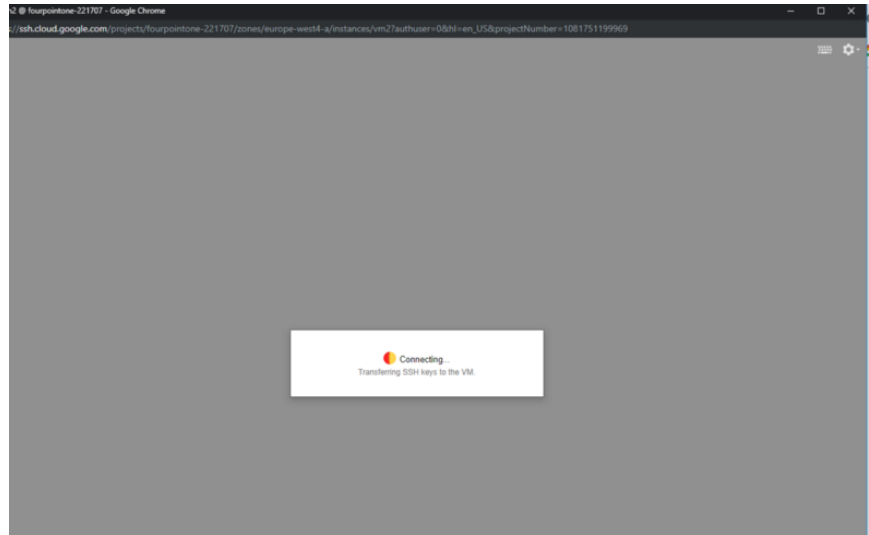
Uma vez que a VM é criada, você verá SSH ou RDP na coluna mais à direita, dependendo de ter janelas ou VMs Linux criadas.

VM instances						
CREATE INSTANCE IMPORT VM REFRESH START STOP RESET DELETE						
<input type="text" value="Filter VM instances"/> Columns						
<input type="checkbox"/>	Name	Zone	Recommendation	Internal IP	External IP	Connect
<input type="checkbox"/>	instance-1	us-east1-b		10.142.0.2 (nic0)	35.231.207.255	RDP
<input type="checkbox"/>	vm1	europa-west4-a		10.164.0.2 (nic0)	35.204.71.54	SSH
<input type="checkbox"/>	vm2	europa-west4-a		10.164.0.3 (nic0)	35.204.215.50	SSH

VM e tipo de acesso

A maneira mais fácil de acessar a VM é apenas clicar em SSH e esperar que a conexão seja estabelecida.

Demorará um pouco para o SSH estabelecer, dependendo da velocidade da sua conexão



Conexão SSH

Uma vez terminado, você verá o prompt da CLI da VM.

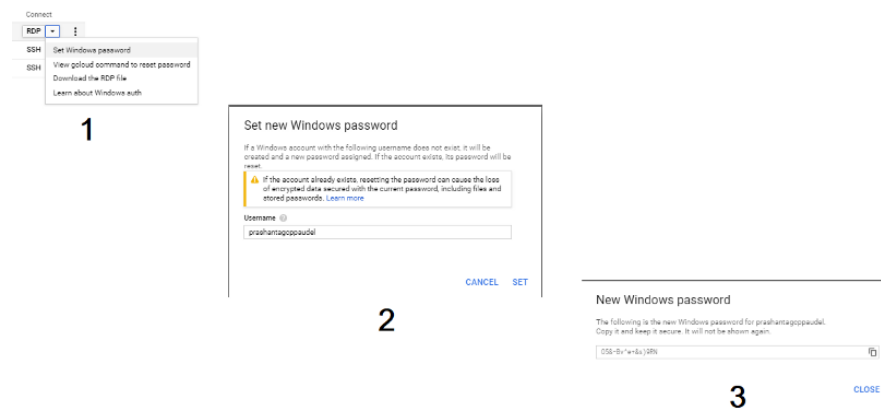
```
https://ssh.cloud.google.com/projects/fourpointone-221707/zones/europe-west4-a/instances/vm2?authuser=0&hl=en_US&projectNumber=1081751199969
connected, host fingerprint: ssh-rsa 2048 8B:20:56:CC:87:C8:76:F3:63:16:44:35:21
:67:44:68:A0:BD:D3:F6:88:1E:25:94:2A:51:F2:BC:28:E3:46:2B
Linux vm2 4.9.0-8-amd64 #1 SMP Debian 4.9.110-3+deb9u6 (2018-10-08) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

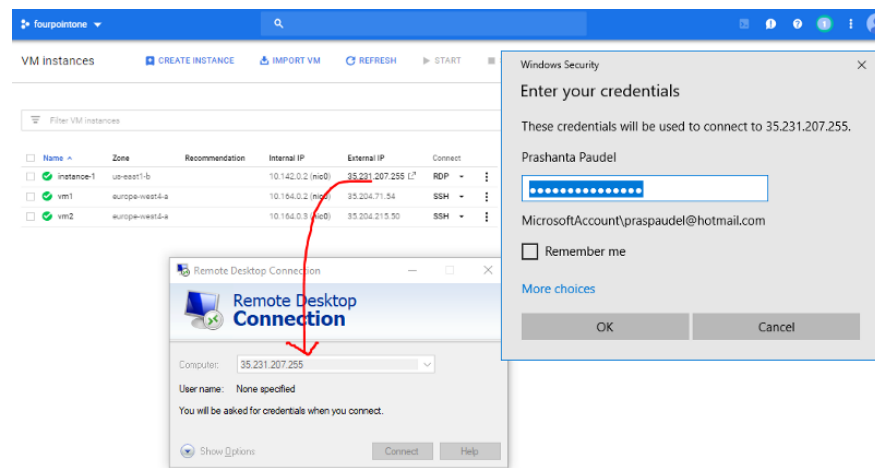
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
prashantagcpaudel@vm2:~$
```

CLI da VM

Enquanto no RDP o processo é um pouco diferente. Você precisa redefinir a senha antes de acessar o servidor.



Agora use a senha e o nome de usuário para acessar o servidor remotamente usando o RDP.



Área de trabalho RDP

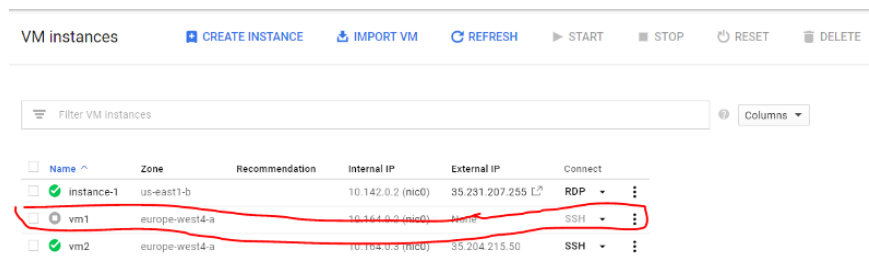
=====

=====

Anexando uma GPU a uma nova instância e instalando bibliotecas CUDA

Podemos instalar a GPU de duas maneiras, seja selecionando enquanto criamos uma VM ou instalamos depois da instalação da VM.

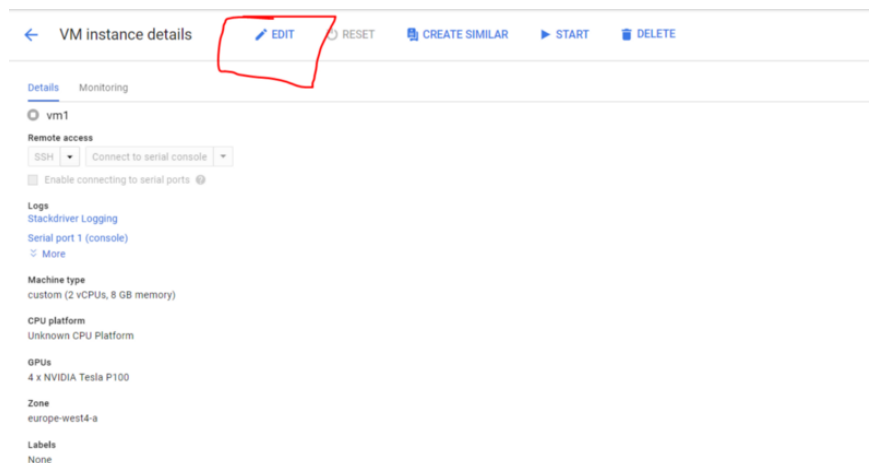
Para instalar a GPU depois, você precisa PARAR a VM e editar as configurações para adicionar a GPU.



Name	Zone	Recommendation	Internal IP	External IP	Connect
instance-1	us-east1-b		10.142.0.2 (nic0)	35.231.207.255 L	RDP
vm1	europa-west4-a		10.164.0.2 (nic0)	None	SSH
vm2	europa-west4-a		10.164.0.3 (nic0)	35.204.215.50	SSH

Pare a VM

Em seguida, edite a VM para adicionar GPU



VM instance details

EDIT RESET CREATE SIMILAR START DELETE

Details Monitoring

vm1

Remote access

SSH Connect to serial console

Enable connecting to serial ports

Logs

Stackdriver Logging

Serial port 1 (console)

More

Machine type

custom (2 vCPUs, 8 GB memory)

CPU platform

Unknown CPU Platform

GPUs

4 x NVIDIA Tesla P100

Zone

europa-west4-a

Labels

None

Adicionando GPU

Adicione GPU de acordo com a escolha e clique em salvar. Em seguida, apenas inicie a VM.

Para adicionar bibliotecas CUDA, você precisa ter uma conta NVIDIA e baixar o repo do site.

A instância do GCP deve ter a Nvidia ativada para K80 ou P100, o que não está ativado na conta gratuita.

Agora, comece baixando o repositório da Nvidia.

```
$ curl -O
http://developer.download.nvidia.com/compute/cuda/repos/ubuntu1604/x86_64/cuda-repo-ubuntu1604_8.0.61-1_amd64.deb
```


Quando você tiver concluído o download, instale com o dpkg e atualize o repositório. Em seguida, limpe depois de nós mesmos.

```
$ sudo dpkg -i cuda-repo-ubuntu1604_8.0.61-1_amd64.deb
$ sudo apt-get atualizar
$ rm cuda-repo-ubuntu1604_8.0.61-1_amd64.deb
```

=====

=====

Visualizando o inventário atual de VM em execução

Verificando o status da instância

Ao criar uma instância pela primeira vez, você deve verificar o status da instância para ver se ela está em execução antes de esperar que ela responda a solicitações. Pode levar alguns segundos até que sua instância esteja totalmente em operação após a solicitação inicial. Você também pode verificar o status de uma instância a qualquer momento após a criação da instância.

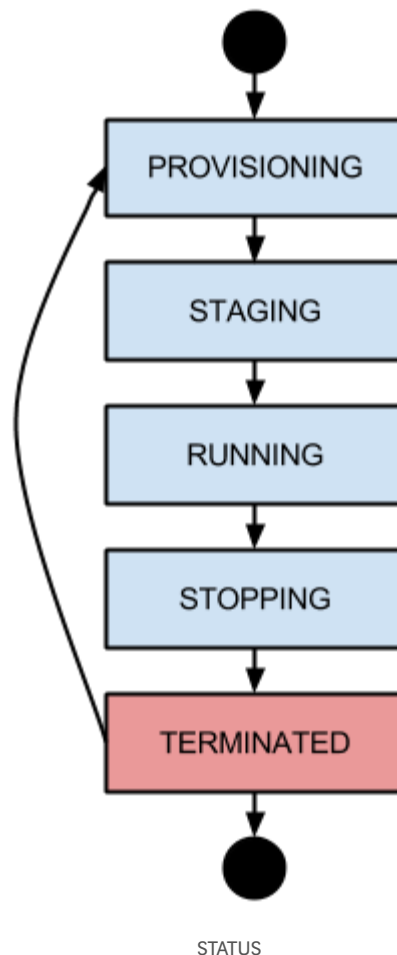
Status das instâncias

As instâncias podem estar nos seguintes estados:

- **PROVISIONING** - Os recursos estão sendo reservados para a instância. A instância ainda não está em execução.
- **STAGING** - Recursos foram adquiridos e a instância está sendo preparada para o lançamento.
- **RUNNING** - A instância está sendo inicializada ou em execução. Você deve poder enviar ssh para a instância em breve, embora não imediatamente após entrar nesse estado.
- **STOPPING** - A instância está sendo interrompida devido a uma falha ou a instância está desligada. Este é um status temporário e a instância será movida para **TERMINATED**.

- **TERMINATED** - A instância foi encerrada ou encontrou uma falha, seja por meio da API ou de dentro do convidado. Você pode optar por reiniciar a instância ou excluí-la.
- **Aviso:** **TERMINATED** instâncias com SSDs locais conectados não podem ser reiniciadas neste momento. Planejamos dar suporte a essa funcionalidade no futuro.

O diagrama a seguir descreve a progressão desses status.



Depois que uma instância é **RUNNING**, você pode tentar se conectar a ela. Se você não conseguir se conectar à instância inicialmente, tente conectar-se novamente em alguns minutos, pois o sistema operacional ainda pode estar sendo inicializado.

Instâncias do Windows experimentam um tempo de inicialização mais longo devido ao processo sysprep, portanto, você deve executar uma

verificação adicional para verificar se a instância do Windows foi iniciada.

Verificando o status de uma instância

Listar todas as instâncias e seu status:

```
gcloud compute instances list
```

Descreva o status de uma única instância:

```
gcloud compute instances describe example-instance
```

```
=====
=====
```

Trabalhando com instantâneos (por exemplo, criar um instantâneo de uma VM, exibir instantâneos, excluir um instantâneo)

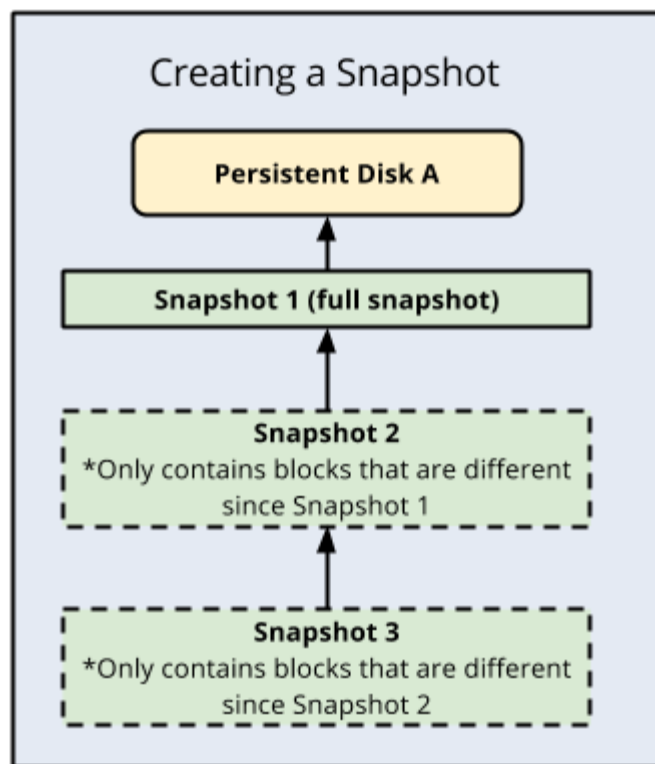
Os instantâneos são diferentes das imagens públicas e das imagens personalizadas, que são usadas principalmente para criar instâncias ou configurar modelos de instância. Os instantâneos são úteis para backup periódico dos dados em seus discos permanentes e você pode usar instantâneos para criar uma imagem personalizada quando necessário. Você pode criar instantâneos de discos permanentes, mesmo quando estiverem conectados a instâncias em execução.

Os instantâneos são incrementais e compactados automaticamente, para que você possa criar instantâneos regulares em um disco permanente com mais rapidez e a um custo muito menor do que se você criasse regularmente uma imagem completa do disco. Instantâneos incrementais funcionam da seguinte maneira:

- O primeiro instantâneo bem-sucedido de um disco permanente é um instantâneo completo que contém todos os dados no disco permanente.

- O segundo instantâneo contém apenas novos dados ou dados modificados desde o primeiro instantâneo. Os dados que não foram alterados desde a captura instantânea 1 não estão incluídos. Em vez disso, o instantâneo 2 contém referências ao instantâneo 1 para quaisquer dados inalterados.
- O Instantâneo 3 contém dados novos ou alterados desde o instantâneo 2, mas não conterá dados inalterados do instantâneo 1 ou 2. Em vez disso, o instantâneo 3 contém referências a blocos no instantâneo 1 e instantâneo 2 para quaisquer dados inalterados.

Isso se repete para todos os instantâneos subsequentes do disco permanente. Os instantâneos são sempre criados com base no último instantâneo bem-sucedido tirado.



O Compute Engine armazena várias cópias de cada instantâneo de maneira redundante em vários locais com somas de verificação automáticas para garantir a integridade de seus dados. Use funções do IAM para compartilhar instantâneos entre projetos.

Para ver uma lista de capturas instantâneas disponíveis para um projeto, use o `gcloud compute snapshots list` comando:

```
gcloud compute snapshots list
```

Para listar informações sobre um instantâneo específico, use o `gcloud compute snapshots describe` comando:

```
gcloud compute snapshots describe example-snapshot
```

Entendendo as melhores práticas de captura instantânea

Você pode criar quantos instantâneos precisar a qualquer momento, mas pode criar instantâneos mais rapidamente e com maior confiabilidade, se usar as seguintes práticas recomendadas:

Prepare seu disco permanente para obter a melhor consistência de instantâneo

Na maioria das situações, você pode criar um instantâneo a partir de discos permanentes mesmo quando seus aplicativos estiverem gravando dados para eles e ainda assim esperar que o instantâneo tenha boa consistência. A qualidade do instantâneo depende da capacidade de seus aplicativos se recuperarem de instantâneos que você cria durante cargas de trabalho de gravação pesada.

Se seus aplicativos exigirem consistência estrita, você poderá executar uma ou mais etapas para garantir que um instantâneo seja consistente com o estado desejado do disco permanente.

- Pausa aplicativos ou processos do sistema operacional que gravam dados nesse disco permanente. Em seguida, libere os buffers de disco antes de criar o instantâneo.

- Desmonte completamente o disco permanente para garantir que nenhum dado seja gravado enquanto você cria o instantâneo. Isso geralmente é desnecessário, mas melhora a consistência do instantâneo.
- Se seus aplicativos exigirem consistência entre vários discos permanentes, você deverá congelar ou desmontar todos os sistemas de arquivos em cada disco e concluir todos os instantâneos desses discos antes de retomar seus aplicativos. O Compute Engine não garante a consistência entre instantâneos simultâneos em execução em vários discos permanentes.
- Use sistemas de arquivos de registro no diário `ext4` para reduzir o risco de que os dados sejam armazenados em cache sem realmente serem gravados no disco permanente.
- Para discos permanentes conectados a instâncias do Windows Server, use os instantâneos de VSS para ajudar a preservar a integridade dos dados.

Use instantâneos existentes como uma linha de base para instantâneos subsequentes

Se você tiver instantâneos existentes em um disco permanente, o sistema os usará automaticamente como uma linha de base para os instantâneos subsequentes criados nesse mesmo disco.

- Crie um novo instantâneo a partir de um disco permanente antes de excluir o instantâneo anterior desse mesmo disco permanente. O sistema pode criar o novo instantâneo mais rapidamente se puder usar o instantâneo anterior e ler somente os dados novos ou alterados do disco permanente.
- Aguarde a conclusão de novos instantâneos antes de obter instantâneos subsequentes do mesmo disco permanente. Se você executar dois instantâneos simultaneamente no mesmo disco permanente, eles serão iniciados a partir da mesma linha de base e do esforço duplicado. Se você esperar que o novo instantâneo seja concluído, todos os instantâneos subsequentes serão executados mais rapidamente porque eles precisam apenas obter os dados que foram alterados desde o último instantâneo concluído.

Programar instantâneos durante horários fora de pico

Se você programar instantâneos regulares para seus discos permanentes, poderá reduzir o tempo necessário para concluir cada instantâneo, criando-os fora do horário de pico, quando possível.

- Agende instantâneos automatizados durante o dia útil na zona onde seu disco permanente está localizado. A criação de instantâneos normalmente atinge o pico no final do dia útil.
- Programar instantâneos automatizados no início da manhã na zona onde seu disco permanente está localizado e não imediatamente à meia-noite. A criação de instantâneos normalmente atinge o pico à meia-noite.

Organize seus dados em discos permanentes separados

Se você criar um instantâneo de um disco permanente, todos os dados armazenados no disco serão incluídos no instantâneo. Quantidades maiores de dados criam snapshots maiores, que custam mais e levam mais tempo para serem criados. Para garantir que você esteja capturando apenas os dados necessários, organize seus dados em discos permanentes separados.

- Armazene dados críticos em um disco persistente secundário em vez de em seu disco de inicialização. Isso permite que você tire um instantâneo dos discos de inicialização somente quando necessário ou com uma programação menos frequente.
- Se você criar instantâneos de seus discos de inicialização, armazene partições de troca, arquivos de paginação, arquivos de cache e logs não críticos em um disco permanente separado. Esses arquivos e partições mudam com frequência e é provável que o processo de captura instantânea os identifique como dados alterados que devem ser incluídos em uma captura instantânea incremental.
- Reduza o número de instantâneos que você precisa criar, mantendo dados semelhantes juntos em um disco permanente. Você deseja manter o sistema operacional e os dados voláteis separados dos dados que deseja capturar, mas não há necessidade de distribuir seus dados críticos por vários discos permanentes, como faria em uma máquina física. Um disco permanente grande

é capaz de obter o mesmo desempenho de vários discos persistentes menores com o mesmo tamanho total.

Ative a `discard` opção ou execute `fstrim` em seu disco permanente



Em instâncias do Linux, se você não formatou e montou seu disco permanente com a `discard` opção, execute o `fstrim` comando na instância antes de criar um instantâneo. O comando remove os blocos que o sistema de arquivos não precisa mais para que o sistema possa criar o instantâneo mais rapidamente e com um tamanho menor. Veja a formatação e montagem de um disco permanente para aprender a configurar a `discard` opção em seus discos permanentes.

Para criar um instantâneo para uma VM, vá para Compute engine > snapshots > create Snapshot

The screenshot shows the 'Create a snapshot' interface. A red box highlights the 'Encryption type' section, which includes the option 'Enable VSS'. A red arrow points from this box to the text 'For windows snapshots'.

criar instantâneo

O instantâneo atualizado terá dados atualizados apenas no novo instantâneo

Filter snapshots				
<input type="checkbox"/> Name ^	Source disk	Creation time	Disk size	Snapshot size
<input type="checkbox"/>  snapshot-1	instance-2	Nov 6, 2018, 11:44:12 AM	10 GB	523.85 MB
<input type="checkbox"/>  snapshot-2	instance-2	Nov 6, 2018, 11:50:22 AM	10 GB	2.17 MB

instantâneo atualizado

Excluir um instantâneo

```
$ gcloud compute snapshots delete snapshot-2
Os seguintes instantâneos serão excluídos:
- [snapshot-2]

Deseja continuar (S / n)? y

Excluído
[https://www.googleapis.com/compute/v1/projects/fourpointone-221707/global/snapshots/snapshot-2].
```

Trabalhar com imagens (por exemplo, criar uma imagem a partir de uma VM ou um instantâneo, ver imagens, apagar uma imagem)

Imagens

Use imagens do sistema operacional para criar discos de inicialização para suas instâncias. Você pode usar um dos seguintes tipos de imagem:

- **Imagens públicas** são fornecidas e mantidas pelo Google, comunidades de código aberto e fornecedores terceirizados. Por padrão, todos os projetos têm acesso a essas imagens e podem usá-las para criar instâncias.
- **Imagens personalizadas** estão disponíveis apenas para o seu projeto. Você pode criar uma imagem personalizada a partir de discos de inicialização e outras imagens. Em seguida, use a imagem personalizada para criar uma instância.

Você pode usar a maioria das imagens públicas sem custo adicional, mas há algumas imagens premium que adicionam custos adicionais às suas instâncias. As imagens personalizadas que você importa para o Compute Engine não acrescentam custos às suas instâncias, mas incorrem em uma taxa de armazenamento de imagens enquanto você mantém sua imagem personalizada em seu projeto.

Algumas imagens são capazes de executar contêineres no Compute Engine.

Imagens públicas

O Compute Engine oferece muitas imagens públicas pré-configuradas que possuem sistemas operacionais Linux e Windows compatíveis. Use essas imagens do sistema operacional para criar e iniciar instâncias. O Compute Engine usa sua imagem selecionada para criar um disco de inicialização persistente para cada instância. Por padrão, o disco de inicialização de uma instância tem o mesmo tamanho da imagem selecionada. Se a sua instância exigir um disco de inicialização persistente maior que o tamanho da imagem, redimensione o disco de inicialização.

Para ver a lista completa de imagens públicas com seus nomes de imagens, números de versões e tamanhos de imagens, vá para a página Imagens no console. O Google atualiza imagens públicas regularmente ou quando um patch para um impacto crítico CVE está disponível. Assine as notificações de imagem do GCE para receber notificações de lançamentos de atualização.

Imagens personalizadas

Uma imagem personalizada é uma imagem de disco de inicialização que você possui e controla o acesso. Use imagens customizadas para as seguintes tarefas:

- Importe uma imagem de disco de inicialização para o Compute Engine de seu ambiente local ou importe discos virtuais de VMs em execução em sua estação de trabalho local ou em outra plataforma de nuvem.
- **Observação:** se você planeja migrar várias VMs para o Compute Engine, considere usar o serviço de migração de VMs.

- Crie uma imagem a partir dos discos de inicialização de suas instâncias existentes do Compute Engine. Em seguida, use essa imagem para criar novos discos de inicialização para suas instâncias. Esse processo permite criar novas instâncias que são pré-configuradas com os aplicativos de que você precisa, sem precisar configurar uma imagem pública a partir do zero.
- Copie uma imagem para outra imagem usando a `gcloud` ferramenta ou a API. Use o mesmo processo usado para criar uma imagem, mas especifique outra imagem como a origem da imagem. Você também pode criar uma imagem a partir de uma imagem personalizada em um projeto diferente.

Crie uma imagem da VM

← Create an image

Name ?

image-1

Family (Optional) ?

Description (Optional)

Labels ? (Optional)

[+ Add label](#)

Encryption

Data is encrypted automatically. Select an encryption key management solution.

- ☒ Google-managed key
No configuration required
- ☐ Customer-managed key
Manage via Google Cloud Key Management Service
- ☐ Customer-supplied key
Manage outside of Google Cloud

Source ?

Disk

Source disk ?

instance-2

Image size: 10 GB

- ☐ Keep instance running (not recommended)
Filesystem integrity can't be guaranteed while the instance is running, which may create a corrupted image

You will be billed for this image. [Compute Engine pricing](#)

Create

Cancel

Equivalent [REST](#) or [command line](#)

criar uma imagem da VM

Para excluir uma imagem

Images

CREATE IMAGE

REFRESH

CREATE INSTANCE

DEPRECATE

DELETE

Filter images

Columns

<< Previous

1

2

Next >>

<div></div>	Name	Size	Created by	Family	Creation time
<div><div><input checked="" type="checkbox"/></div><div>✓</div></div>	Image-1	10 GB	fourpointone		Nov 6, 2018, 12:07:31 PM

excluir uma imagem

Para criar uma imagem de um instantâneo

← Create an image

Name [?]
image-1

Family (Optional) [?]

Description (Optional)

Labels [?] (Optional)
+ Add label

Encryption
Data is encrypted automatically. Select an encryption key management solution.

- ☒ **Google-managed key**
No configuration required
- ☐ **Customer-managed key**
Manage via Google Cloud Key Management Service
- ☐ **Customer-supplied key**
Manage outside of Google Cloud

Source [?]
Snapshot

Source snapshot [?]
snapshot-1

Image size: 10 GB

You will be billed for this image. [Compute Engine pricing](#) [?]

Create Cancel

Equivalent [REST](#) or [command line](#)

imagem do instantâneo

=====

=====

Trabalhar com grupos de instâncias (por exemplo, definir parâmetros de dimensionamento automático, atribuir modelo de instância, criar um modelo de instância, remover grupo de instâncias)

O ponto de partida no dimensionamento automático é determinar o modelo da máquina, depois criar um grupo de instâncias e, finalmente, definir o balanceamento de carga.

Escalonamento automático e balanceamento de carga são o cumprimento um do outro.

Grupos de instâncias de escalonamento automático

Managed instance groups offer autoscaling capabilities that allow you to automatically add or delete instances from a managed instance group based on increases or decreases in load. Autoscaling helps your applications gracefully handle increases in traffic and reduces cost when the need for resources is lower. You just define the autoscaling policy and the autoscaler performs automatic scaling based on the measured load.

Autoscaling works by adding more instances to your instance group when there is more load (upscaling), and deleting instances when the need for instances is lowered (downscaling).

Fundamentals

Autoscaling uses the following fundamental concepts and services.

Managed instance groups

Autoscaling is a feature of managed instance groups. A managed instance group is a pool of homogeneous instances, created from a common instance template. An autoscaler adds or deletes instances from a managed instance group. Although Compute Engine has both managed and unmanaged instance groups, only managed instance groups can be used with autoscaler.

To understand the difference between a managed instance group and unmanaged instance group, see the Instance Groups documentation.

Autoscaling policy and target utilization

To create an autoscaler, you must specify the autoscaling policy and a target utilization level that the autoscaler uses to determine when to scale the group. You can choose to scale using the following policies:

- Average CPU utilization
- HTTP load balancing serving capacity, which can be based on either utilization or requests per second.

- Stackdriver Monitoring metrics

The autoscaler will collect information based on the policy, compare it to your desired target utilization, and determine if it needs to perform scaling.

The target utilization level is the level at which you want to maintain your virtual machine instances. For example, if you scale based on CPU utilization, you can set your target utilization level at 75% and the autoscaler will maintain the CPU utilization of the specified group of instances at or close to 75%. The utilization level for each metric is interpreted differently based on the autoscaling policy.

For a brief summary of each policy, see [Autoscaling policies in Overviews](#). For a detailed discussion of each policy, see:

- [Scaling Based on CPU or Load Balancing Serving Capacity](#)
- [Scaling Based on Stackdriver Monitoring Metrics](#)

Instance Groups

You can create and manage groups of virtual machine (VM) instances so that you don't have to individually control each instance in your project. Compute Engine offers two different types of instance groups: **managed** and **unmanaged instance** groups.

Managed instance groups

A managed instance group uses an instance template to create a group of identical instances. You control a managed instance group as a single entity. If you wanted to make changes to instances that are part of a managed instance group, you would make the change to the whole instance group. Because managed instance groups contain identical instances, they offer the following features:

- When your applications require additional compute resources, managed instance groups can automatically scale the number of instances in the group.
- Managed instance groups work with load balancing services to distribute traffic to all of the instances in the group.

- If an instance in the group stops, crashes, or is deleted by an action other than the instance groups commands, the managed instance group automatically recreates the instance so it can resume its processing tasks. The recreated instance uses the same name and the same instance template as the previous instance, even if the group references a different instance template.
- Managed instance groups can automatically identify and recreate unhealthy instances in a group to ensure that all of the instances are running optimally.

Types of managed instance groups

You can create two types of managed instance groups:

- A zonal managed instance group, which contains instances from the same zone.
- A regional managed instance group, which contains instances from multiple zones across the same region.

Regional managed instance groups are generally recommended over zonal managed instance groups because they allow you to spread application load across multiple zones, rather than confining your application to a single zone or having to manage multiple instance groups across different zones. This replication protects against zonal failures and unforeseen scenarios where an entire group of instances in a single zone malfunctions. If that happens, your application can continue serving traffic from instances running in another zone in the same region.

Choose zonal managed instance groups if you want to avoid the slightly higher latency incurred by cross-zone communication or if you need fine-grained control of the sizes of your groups in each zone.

Managed instance groups and the network

By default, instances in the group will be placed in the `default` network and randomly assigned IP addresses from the regional range. Alternatively, you can restrict the IP range of the group by creating a custom mode VPC network and subnet that uses a smaller IP range,

then specifying this subnet in the instance template. This can simplify the creation of firewall rules.

Depois de criar um grupo de instâncias gerenciadas, as novas instâncias começam no grupo assim que o sistema puder fornecê-las. Esse processo pode levar um tempo significativo, dependendo do número de instâncias no grupo. Verifique o status das instâncias no seu grupo de instâncias gerenciadas.

Grupos de instâncias não gerenciadas

Grupos de instâncias não gerenciadas são grupos de instâncias diferentes que você pode adicionar e remover arbitrariamente do grupo. Grupos de instâncias não gerenciadas não oferecem escalonamento automático, suporte a atualizações contínuas ou o uso de modelos de instâncias, de modo que o Google recomenda a criação de grupos de instâncias gerenciados sempre que possível. Use grupos de instâncias não gerenciadas somente se você precisar aplicar o balanceamento de carga às configurações pré-existentes ou a grupos de instâncias diferentes.

Se você precisar criar um grupo de instâncias diferentes que não seguem um modelo de instância, consulte Grupos de instâncias não gerenciadas.

Grupos de instâncias e balanceamento de carga

Todas as configurações de balanceamento de carga disponíveis no Google Cloud Platform exigem que você especifique grupos de instâncias ou pools de destino que possam veicular o tráfego distribuído a partir do balanceador de carga.

Para balanceamento de carga HTTP (S), interno e SSL, você deve atribuir um grupo de instâncias a um serviço de back-end. Um serviço de back-end é um serviço centralizado para o gerenciamento de back-ends, que, por sua vez, gerencia instâncias que lidam com solicitações de usuários para seu balanceador de carga. Cada serviço de back-end contém um ou mais back-ends e cada back-end contém um grupo de instâncias. O serviço de back-end sabe quais instâncias ele pode usar, quanto tráfego eles podem manipular e quanto tráfego eles estão

manipulando no momento. Você pode atribuir um grupo de instâncias gerenciado ou não gerenciado a um serviço de back-end.

Para balanceamento de carga de rede, você deve adicionar instâncias de VMs individuais a um pool de destino ou atribuir um ou mais grupos de instâncias gerenciadas a um pool de destino, o que faz com que o servidor adicione todas as instâncias que fazem parte do grupo de instâncias ao pool de destino especificado.

Para obter mais informações sobre diferentes configurações de balanceamento de carga, consulte a documentação sobre balanceamento de carga.

Grupos de instâncias gerenciadas e escalonamento automático

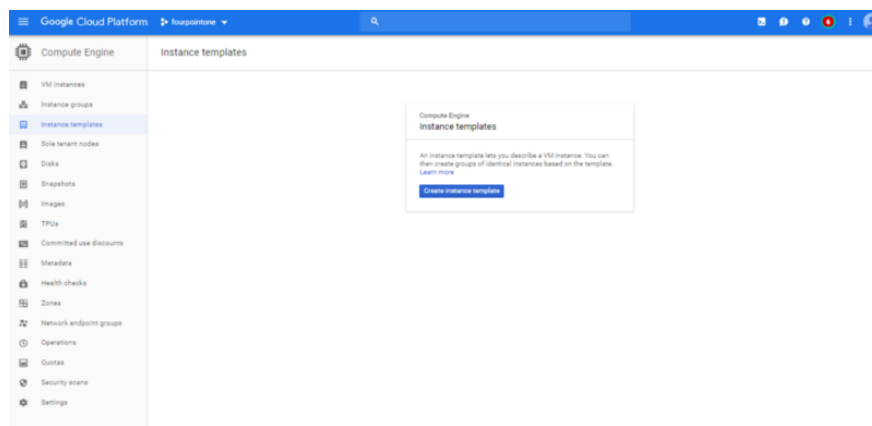
Os grupos de instâncias gerenciadas aceitam o escalonamento automático para que você possa adicionar ou remover dinamicamente instâncias de um grupo de instâncias gerenciadas em resposta a aumentos ou diminuições no carregamento. Você ativa o escalonamento automático e escolhe uma política de escalonamento automático para determinar como deseja escalonar. As políticas de escalonamento automático aplicáveis incluem o escalonamento com base na utilização da CPU, na capacidade de balanceamento de carga, nas métricas de monitoramento do Stackdriver ou em uma carga de trabalho baseada em filas, como o Google Cloud Pub / Sub.

Como o escalonamento automático exige a adição e a remoção de instâncias de um grupo, você só pode usar o escalonamento automático com grupos de instâncias gerenciadas para que o escalonador automático possa manter instâncias idênticas. O escalonamento automático não funciona em grupos de instâncias não gerenciadas, que podem conter instâncias heterogêneas.

Para obter mais informações, leia Grupos de instâncias de escalonamento automático.

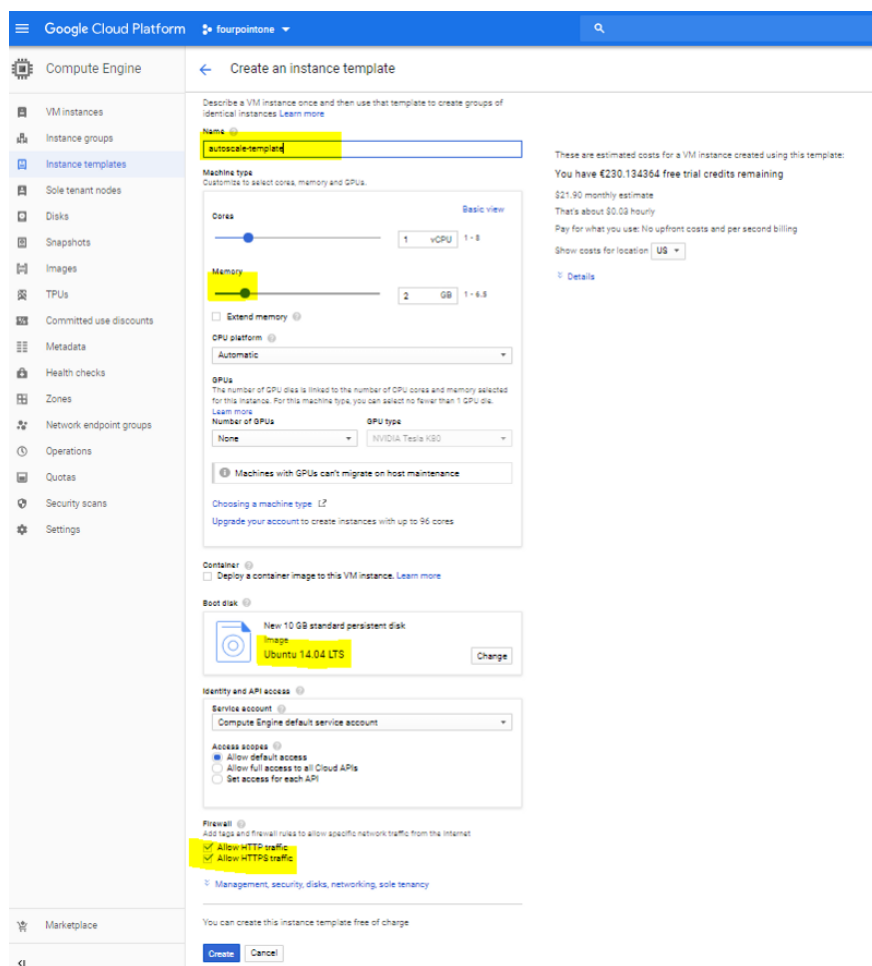
Criar modelo de instância

Para criar um modelo de instância, goto compute engine e instance template



criar modelo de instância

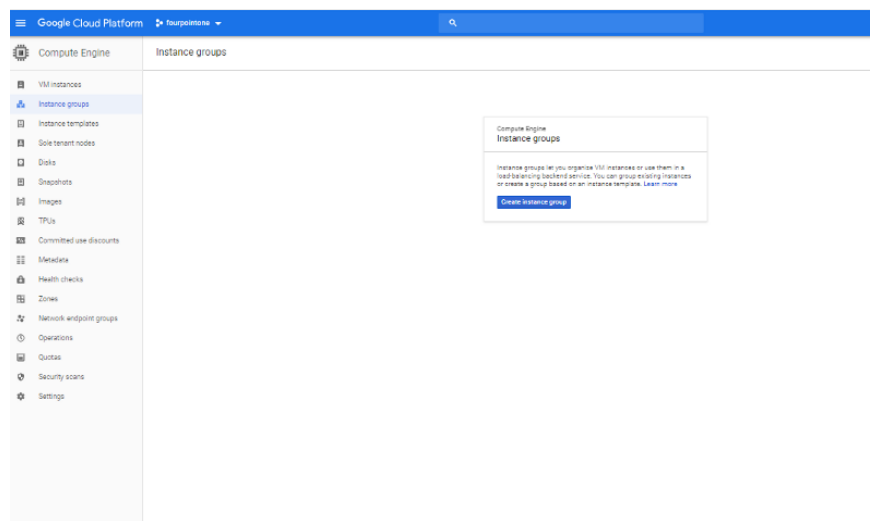
Depois de clicar em criar, você verá o mesmo tipo de página que criar uma VM. Este modelo será replicado para várias máquinas durante o escalonamento automático.



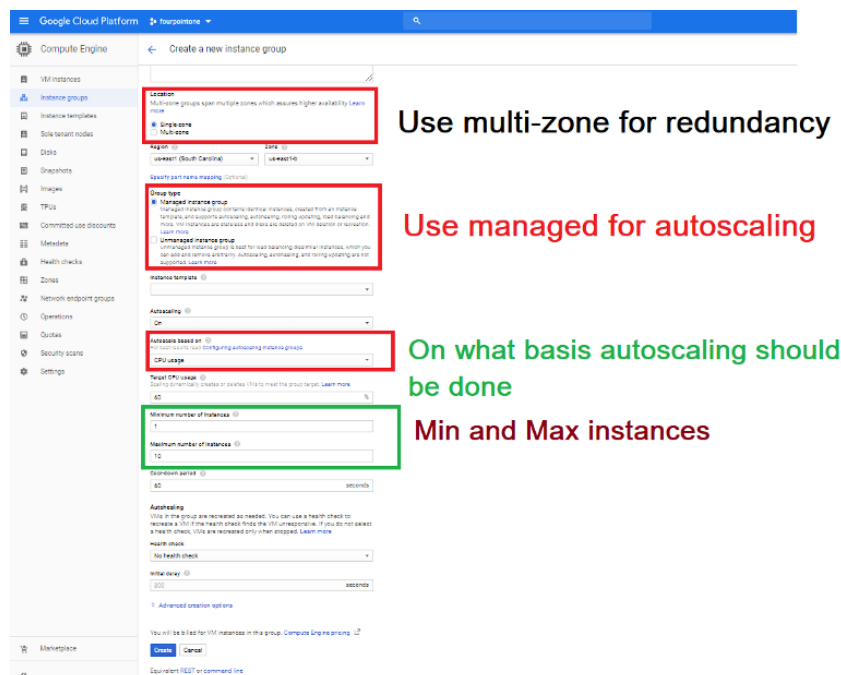
modelo

Crie um grupo de instâncias

clique no mecanismo de computação > grupo de instâncias

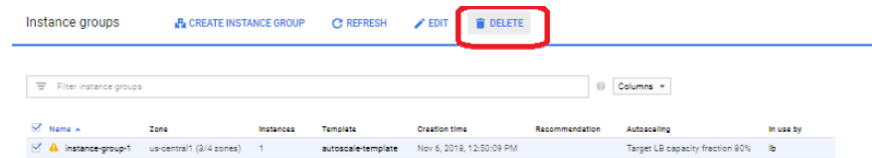


Grupo de instâncias



Crie um grupo

Para remover o grupo de instâncias



excluir grupo de instâncias

Com casca

```
$ gcloud compute instance-groups gerenciados delete
instance-group-1 --region = us-central1
Os seguintes gerenciadores de grupos de instâncias
regionais serão excluídos:
- [instance-group-1] in [us-central1]

Deseja continuar ( S / n)? y

Exclusão do autoescalador ... ï. Excluído
[https://www.googleapis.com/compute/v1/projects/fourpointone
-221707/regions/us-central1/autoscalers/instance-group-1].
Excluindo o autoescalador ... concluído.
Excluindo o grupo de instâncias gerenciadas ... pronto.
```

=====

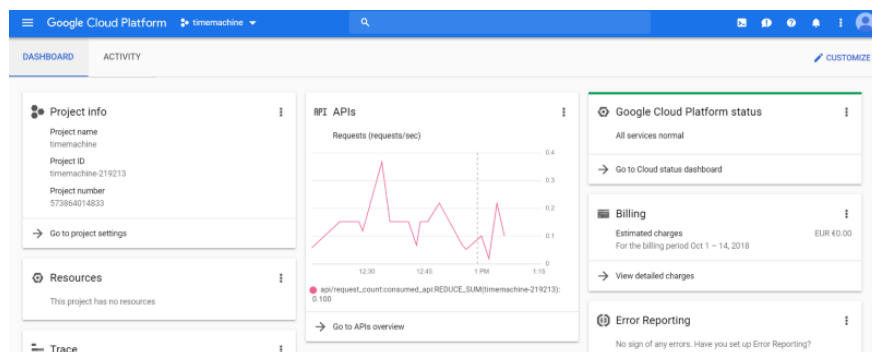
=====

Trabalhando com interfaces de gerenciamento (por exemplo, Cloud Console, Cloud Shell, GCloud SDK)

A plataforma de nuvem do Google pode ser usada de várias maneiras por meio de API, shell de nuvem e console de nuvem. O método mais comum de sujar as mãos enquanto se aprende a plataforma de nuvem do Google é o Cloud Console, enquanto alguém com conhecimento de script ou programação pode usar o shell.

Console de nuvem

O Cloud Console é a maneira visual de trabalhar com a nuvem. Tem quase todos os recursos que o SDK em nuvem como uma máquina virtual fornece ao administrador. Console no gateway para todos os recursos do Google Cloud e é muito útil de uma forma que é organizada visualmente.



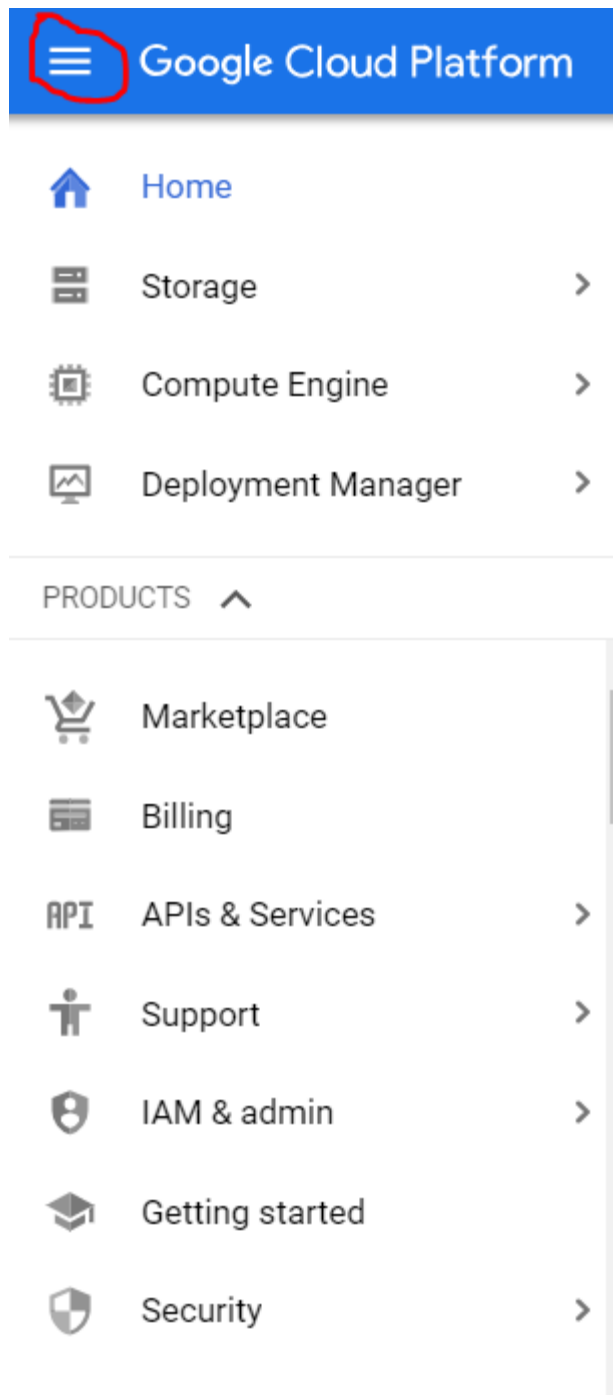
Console do GCP

A parte superior esquerda do console tem um ícone de menu que, quando clicado, deslizará uma lista de menus a partir da esquerda. Mesmo que a interface do GCP tenha mudado frequentemente o lugar para o ícone do menu e o perfil pessoal, notificações, o cloud shell permanece mais ou menos em alguns pontos.

Quando você clicar no ícone do menu (três linhas pequenas), você receberá uma lista de sub-menus. Por padrão, você terá um botão inicial e produtos. O botão "Início" mostrará o painel do seu projeto do GCP selecionado além do menu principal.



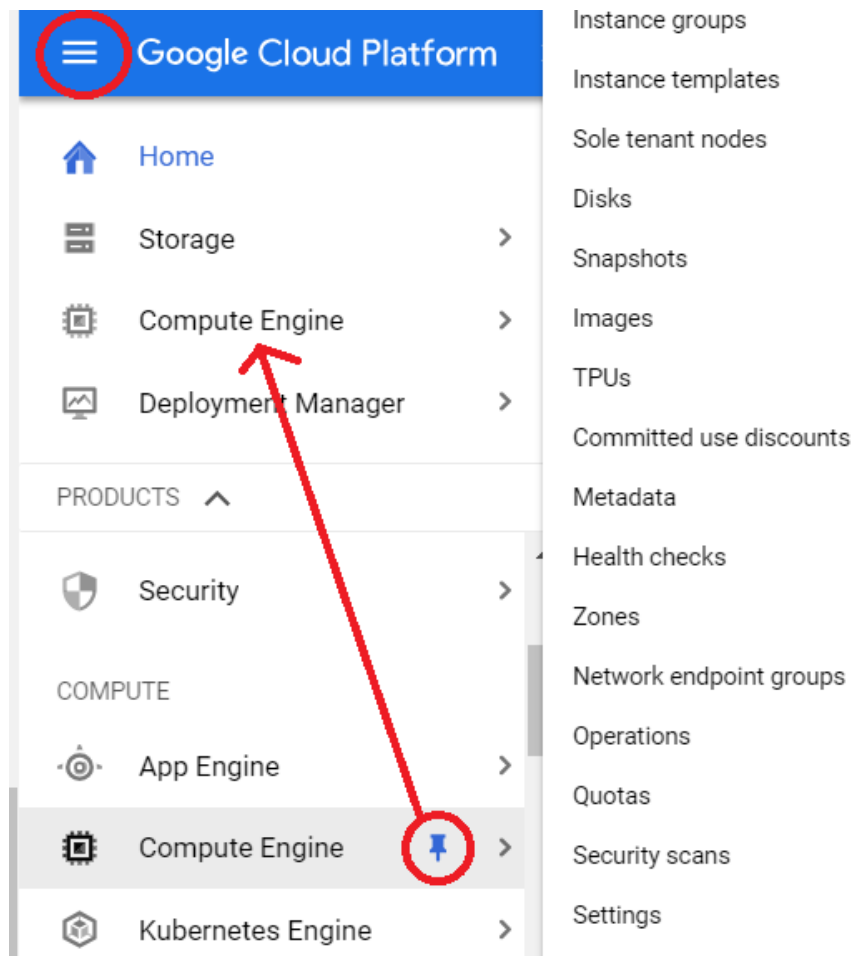
seletor de projeto



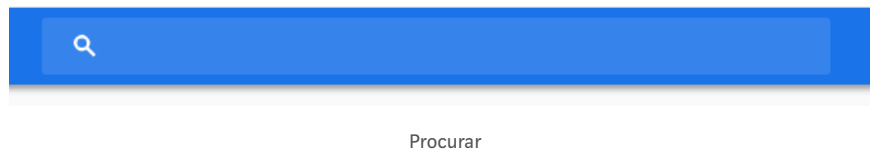
submenus

O submenu consiste em todos os recursos do GCP. Esta lista é a lista que mais muda em toda a página do console. Alguns dos recursos estão em fase beta e outros estão mesclados em outro menu, enquanto alguns aparecem na lista principal.

Na lista principal, há uma opção para fixar o menu abaixo de casa para que você não precise rolar para baixo o tempo todo.



A opção de pesquisa no meio da barra de menu superior facilita o acesso à função de pesquisa para os usuários. Você pode procurar por qualquer projeto, implantação, mercado, etc deste lugar. Se você é rápido, então isso pode ser muito mais fácil do que passar pelo menu principal.



Os itens mais à direita são perfil, configurações, notificação, ajuda, feedback e console na nuvem.



Menu superior direito

Todos os outros botões são gerais, exceto notificação e shell de nuvem.

A notificação é como o registro das atividades que você realiza no GCP. Isso mudará sua aparência quando o GCP estiver fazendo alguma coisa.



criando uma notificação de projeto

Concha nuvem

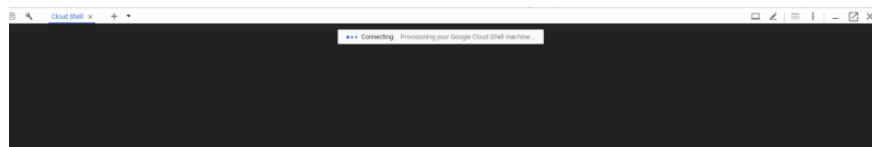
É outra maneira de usar o GCP. O importante a observar ao usar o shell é que você deve ter experiência anterior trabalhando com o shell.

Embora o shell tenha ajuda e manuais para entender seus comandos, é demorado passar por todas as opções e aplicar o comando adequadamente estruturado se a mesma coisa puder ser executada facilmente a partir da GUI.

Você pode acessar o shell da nuvem na barra de menu superior clicando no prompt de comando como um ícone.



Depois de clicar no ícone, uma nova máquina virtual com o SDK em nuvem será inicializada e carregada na plataforma.

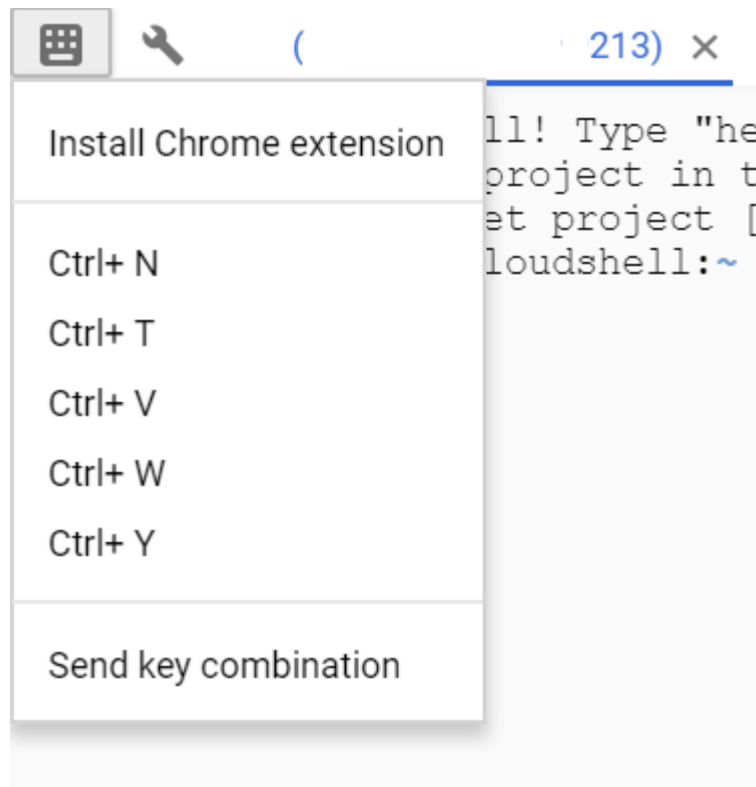


O Cloud Shell fornece um ambiente de linha de comando para acessar os recursos da nuvem diretamente do navegador. Isso equivale ao

download, instalação e conexão ao Google Cloud SDK em seu computador.

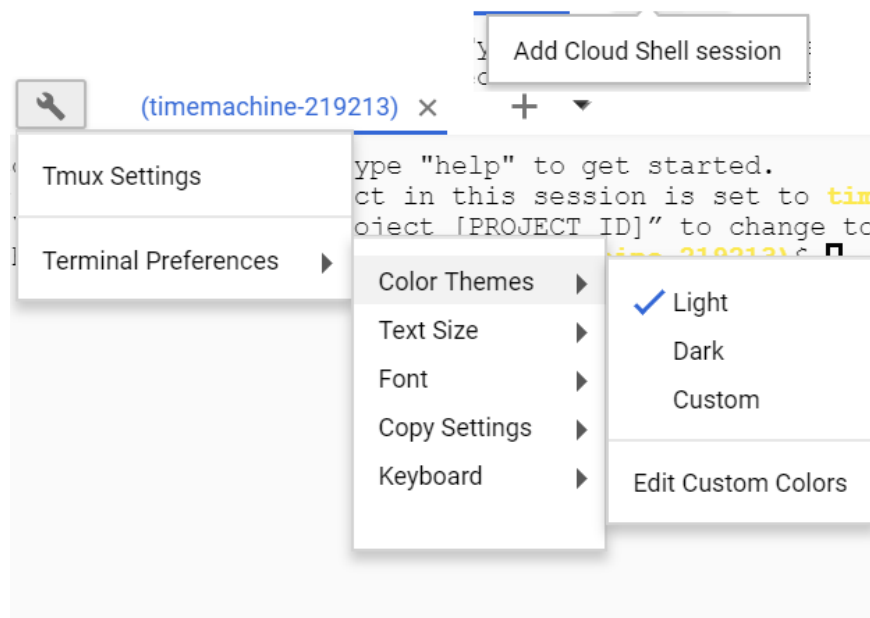
O Cloud Shell também tem opções na barra de menu superior.

A primeira opção permitirá que você instale a extensão chrome para enviar a combinação de teclas



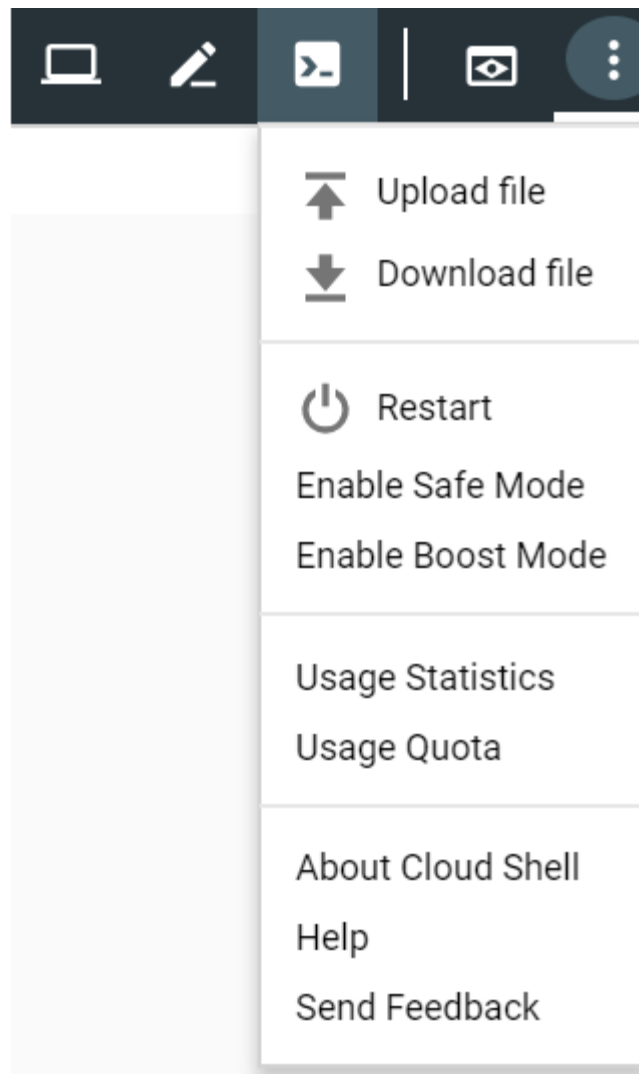
Primeira opção

O segundo ícone dará uma opção para alterar a aparência do shell, bem como as configurações de cópia e teclado



A terceira opção permitirá que você adicione outra sessão ao shell.

O último ícone com três pontos quando pressionado tem várias funções como carregar os arquivos, reiniciar o shell, verificar estatísticas, etc.



última opção

Coisas para ter cuidado

O Cloud Shell oferece 5 GB de armazenamento em disco permanente em seu diretório inicial.

Você deve carregar o projeto no shell da nuvem para começar a usar recursos nele. Nós fazemos isso usando o comando

```
$ gcloud config set project Project_ID
```

se aplicarmos comandos para projetos diferentes, as coisas podem realmente dar errado.

Se você usar privilégios de root, a seleção do projeto deve ser confirmada novamente.

Sempre verifique em qual projeto você está trabalhando, vendo o ID do projeto após `username @ cloudshell ~ (PROJECT_ID) $`.

GCloud SDK no Linux

Primeiro precisamos adicionar informações do repositório no linux e então usar essa informação para baixar o cloud-sdk. Após o download inicialize-o através do navegador ou console e configure a região e o projeto.

```
[prashantapaudel @ localhost ~] $ sudo tee -a
/etc/yum.repos.d/google-cloud-sdk.repo << EOM
> [google-cloud-sdk]
> nome = SDK do Google Cloud
> baseurl = https://packages.cloud.google.com/yum/repos/cloud-
sdk-el7-x86\_64
> enabled = 1
> gpgcheck = 1
> repo_gpgcheck = 1
> gpgkey = https://packages.cloud.google.com/yum/doc/yum-
key.gpg
> https://packages.cloud.google.com/yum/doc/rpm-package-key.gpg
> EOM
[google-cloud-sdk]
name = Google Cloud SDK
baseurl = https://packages.cloud.google.com/yum/repos/cloud-sdk-
el7-x86\_64
enabled = 1
gpgcheck = 1
repo_gpgcheck = 1
gpgkey = https://packages.cloud.google.com/yum/doc/yum-key.gpg
https://packages.cloud.google.com/yum/doc/rpm-package-key.gpg
[prashantapaudel @ localhost ~] $ sudo yum install google-cloud-sdk
Plugins carregados: fastestmirror, langpacks
google-cloud-sdk / signature
| 454 B 00:00:00 Como
recuperar chave de
https://packages.cloud.google.com/yum/doc/yum-key.gpg
Importação da chave GPG 0xA7317B0F: ID do usuário
```

: “Chave de assinatura automática dos pacotes do Google Cloud < gc-team @ google.com > ”

Impressão digital: d0bc 747f d8ca f711 7500 d6fa 3746 c208 a731 7b0f

De: <https://packages.cloud.google.com/yum/doc/yum-key.gpg>

Está tudo bem [y / N]: y

Recuperando a chave de

<https://packages.cloud.google.com/yum/doc/rpm-package-key.gpg>

google-cloud-sdk / signature

| 1,4 kB 00:00:03 !!!

google-cloud-sdk / primary

| 59 kB 00:00:01

Carregando velocidades de espelhamento a partir do hostfile em cache

google-cloud-sdk

380/380

Resolvendo Dependências

→ Executando verificação de transação

--> Pacote google-cloud-sdk.noarch 0: 223.0.0-1.el7 será instalado

→ Resolução de Dependência Concluída

Dependências resolvidas

=====

=====

=====

===== Tamanho do

repositório da

versão do pacote Arch

=====

=====

=====

===== Instalando:

google-cloud-sdk noarch 223.0.0-1.el7 google-cloud-sdk 29 M

Resumo da transação

=====

=====

```
=====
=====
```

Instalar 1 pacote

Tamanho total da transferência: 29 M

Tamanho instalado: 130 M

Está tudo certo [y / d / N]: y Fazendo

download de pacotes:

aviso: / var / cache / yum / x86_64 / 7 / google-cloud-sdk / packages /
441c59a170ea8367941da520d486218a043fb308e4e5d7f64c44c47a6
4603a13-google-cloud-sdk-223.0.0-1.el7.noarch.rpm:

Cabeçalho V4 Assinatura RSA / SHA1, ID da chave 3e1ba8d5: NOKEY

Chave pública para

441c59a170ea8367941da520d486218a043fb308e4e5d7f64c44c47a6
4603a13-google-cloud-sdk-223.0.0-1.el7.noarch. O rpm

não está instalado

441c59a170ea8367941da520d486218a043fb308e4e5d7f64c44c47a6
4603a13-google-cloud-sdk-223.0.

| 29 MB 00:00:09

Recuperando chave de

<https://packages.cloud.google.com/yum/doc/yum-key.gpg>

Importando a chave GPG 0xA7317B0F:

ID do usuário : "Chave de assinatura automática dos pacotes do
Google Cloud < gc-team@google.com >"

Impressão digital: d0bc 747f d8ca f711 7500 d6fa 3746 c208 a731 7b0f

De: <https://packages.cloud.google.com/yum/doc/yum-key.gpg>

Está tudo bem [y / N]: y

Recuperando chave de

<https://packages.cloud.google.com/yum/doc/rpm-package-key.gpg>

Importando chave GPG 0x3E1BA8D5:

Userid: "Google Cloud Pacotes Chave de Assinatura de RPM < gc-team@google.com > "

Impressão digital: 3749 e1ba 95a8 6ce0 5454 6ed2 f09c 394c 3e1b
a8d5

De: <https://packages.cloud.google.com/yum/doc/rpm-package-key.gpg>

Está tudo bem [y / N]: y

Verificação da

transação em execução Teste da transação em execução

Teste de transação bem-sucedido

Executando a transação

Instalando: google-cloud-sdk-223.0.0-1.el7.noarch

1/1

Verificando: google-cloud-sdk-223.0.0-1.el7.noarch

1/1

Instalado:

google-cloud-sdk.noarch 0: 223.0.0-1.el7

Completo!

[prashantapaudel @ localhost ~] \$ gcloud init

Bem-vindo! Este comando irá levá-lo através da configuração do gcloud.

Sua configuração atual foi definida como: [default]

Você pode pular os diagnósticos da próxima vez usando o seguinte sinalizador:

gcloud init - skip-diagnostics

Diagnóstico de rede detecta e corrige problemas de conexão de rede local.

Verificando a conexão de rede ... pronto.

Verificação de Acessibilidade aprovada.

Diagnóstico de rede (1/1 cheques) foi aprovado.

Você deve logar para continuar. Gostaria de fazer login (S / n)? y

Seu navegador foi aberto para visitar:

https://accounts.google.com/o/oauth2/auth?redirect_uri=http%3A%2F%2Flocalhost%3A8085%2F&prompt=select_account&response_type=code&client_id=32555940559.apps.googleusercontent.com&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fuserinfo.email+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-platform+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fappengine.admin+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcompute+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Faccounts.reauth&access_type=offline

(processo: 3530): GLib-CRITICAL **: g_slice_set_config: assertion `sys_page_size == 0` failed

Você está logado como: [prashantagcppaudel@gmail.com].

Escolha o projeto da nuvem a ser usado:

[1] fourpointone-221707

[2] Crie um novo projeto

Insira a opção numérica ou o valor de texto (deve corresponder exatamente ao item da lista): 1

Seu projeto atual foi definido como: [fourpointone-221707].

Deseja configurar uma região e uma zona de computação padrão? (S / n)? y

Qual zona do Google Compute Engine você gostaria de usar como padrão do projeto ?

Se você não especificar uma zona por meio de um sinalizador de linha de comando ao trabalhar com recursos do Compute Engine, o padrão será assumido.

[1] us-east1-b

[2] us-east1-c

[3] us-east1-d

[4] us-east4-c

[5] us-east4-b

[6] us-east4-a

[7] us-central1-c

[8] us-central1-a

[9] us-central1-f

[10] us-central1-b

[11] us-west1-b

[12] us-west1-c

[13] -west1-a

[14] europa-oeste4-a

[15] europa-oeste4-b

[16] europa-oeste4-c

[17] europa-oeste1-b

[18] europa-oeste1-d

[19] europa-oeste1 -c

[20] europe-west3-b

[21] europa-oeste3-c

[22] europa-oeste3-a

[23] europa-oeste2-c
[24] europa-oeste2-b
[25] europa-oeste2-a
[26] ásia-leste1-b
[27]] asia-east1-a
[28] asia-east1-c
[29] Ásia-sudeste1- b
[30] Ásia-sudeste1-a [31] Ásia-sudeste1
-c
[32] Ásia-nordeste1-b
[33] Ásia -northeast1-c
[34] Ásia-northeast1-a
[35] Ásia-south1-c
[36] Ásia-south1-b
[37] Ásia-south1-a
[38] austrália-southeast1-b
[39] austrália-southeast1 -c
[40] austrália-
sudeste1 -a [41] southamerica-east1-b
[42] southamerica-east1-c
[43] southamerica-east1-a
[44] ásia-leste2-a
[45] Ásia-east2-b
[46] Ásia-east2-c
[47] europa-north1-a
[48] europa-north1-b
[49] europa-north1-c
[50] northamerica-northeast1-uma

não Fez imprimir [6] opções.

Muitas opções [56]. Digite "lista" no prompt para imprimir as opções totalmente.

Por favor, digite a opção numérica ou o valor do texto (deve corresponder exatamente ao item da lista): 10

A zona padrão do Compute Engine do seu projeto foi definida como [us-central1-b].

Você pode alterá-lo executando [gcloud config set compute / zone NAME].

A região padrão do Compute Engine do seu projeto foi definida como [us-central1].

Você pode alterá-lo executando [gcloud config set compute / region NAME].

Criado um arquivo de configuração padrão .boto em [/home/prashantapaudel/.boto]. Veja este arquivo e [<https://cloud.google.com/storage/docs/gsutil/commands/config>] para mais informações sobre como configurar o Google Cloud Storage. Seu Google Cloud SDK está configurado e pronto para uso!

* Comandos que requerem autenticação usarão

prashantagcppaudel@gmail.com por padrão

* Os comandos farão referência ao projeto `fourpointone-221707` por padrão

* Os comandos do Compute Engine usarão a região `us-central1` por padrão

* Os comandos do Compute Engine usarão a zona `us-central1-b` por padrão

Execute o comando `gcloud help config` para saber como alterar as configurações individuais

Esta configuração do gcloud é chamada [padrão]. Você pode criar configurações adicionais se trabalhar com várias contas e / ou projetos.

Execute o `gcloud topic configurations` para saber mais.

Algumas coisas para tentar em seguida:

* Execute o comando `gcloud - help` para ver os serviços do Cloud Platform com os quais você pode interagir. E execute o comando `gcloud help COMMAND` para obter ajuda sobre qualquer comando do gcloud .

* Execute o `gcloud topic -h` para aprender sobre recursos avançados do SDK,

como arquivos arg e formatação de saída

[prashantapaudel @ localhost ~] \$

