



En este módulo, analizaremos las instancias de máquinas virtuales, o VM.

Las VM son el componente de infraestructura más común y, en el caso de GCP, las proporciona Compute Engine. Una VM es similar, pero no idéntica, a una computadora de hardware. Las VM constan de una CPU virtual, una cantidad determinada de memoria, almacenamiento en disco y una dirección IP.

Compute Engine es el servicio de GCP que crea VM. Es muy flexible y cuenta con muchas opciones, incluso algunas que no pueden existir en el hardware físico. Por ejemplo, las micro VM comparten CPU con otras máquinas virtuales, por lo que usted puede obtener una VM con menos capacidad a un costo más bajo. Otro ejemplo de una función que no puede existir en el hardware es que algunas VM cuentan con capacidad de aumento de actividad, es decir, la CPU virtual se ejecutará por sobre la capacidad por la que se paga durante un período breve a través de la CPU física compartida disponible. Las opciones principales de VM son CPU, memoria, discos y redes.

Opciones de procesamiento de Google Cloud

	Compute Engine	GKE	Entorno estándar de App Engine	Entorno flexible de App Engine	Cloud Functions
Idiomas admitidos	Cualquiera	Cualquiera	Python Node.js Go Java PHP	Python Node.js Go Java PHP Ruby .NET Entornos de ejecución personalizados	Python Node.js Go
Modelo de uso	IaaS	IaaS PaaS	PaaS	PaaS	Arquitectura de microservicios
Escalamiento	Ajuste de escala automático del servidor	Clúster	Servidores administrados con ajuste de escala automático		Sin servidores
Caso de uso principal:	Cargas de trabajo generales	Cargas de trabajo de contenedores	Aplicaciones web escalables Aplicaciones de backend móvil		Acciones de eventos ligeros

Como se mencionó en la introducción del curso, existe una gran variedad de opciones para el procesamiento en Google Cloud. Nos enfocaremos en las instancias tradicionales de máquina virtual.

La diferencia es que Compute Engine le entrega la máxima flexibilidad: realice ejecuciones en cualquier lenguaje que desee; es su máquina virtual. Este es simplemente un modelo de infraestructura como servicio (IaaS).

Usted cuenta con una VM y un sistema operativo y puede elegir cómo administrarlos y manejar factores como el ajuste de escala automático, en el que configurará las reglas para agregar más máquinas virtuales en situaciones específicas. Más adelante en el curso hablaremos sobre el ajuste de escala automático.

El caso de uso principal de Compute Engine es cualquier carga de trabajo general, sobre todo aplicaciones empresariales que se diseñaron para ejecutarse en una infraestructura de servidores. Esto hace que Compute Engine sea muy portátil y fácil de ejecutar en la nube.

Es posible que otros servicios, como Google Kubernetes Engine, que consta de cargas de trabajo alojadas en contenedores, no sean tan fáciles de transferir como las cargas de trabajo locales que suele usar.

Compute Engine

Infraestructura como servicio (IaaS)



Compute Engine

Tipos personalizados o predefinidos de máquinas:

- CPU virtuales (núcleos) y memoria (RAM)
- Almacenamiento
 - Disco persistente zonal o regional (HDD o SSD)
 - SSD local
 - Cloud Storage
- Redes
- Linux o Windows

¿Qué es Compute Engine? En términos sencillos, se trata de los servidores físicos que ya conoce, pero que se ejecutan en el entorno de Google Cloud con varios parámetros de configuración diferentes.

Los tipos personalizados de máquinas y los predefinidos le permiten elegir cuánta memoria y cuánta CPU desea. Usted elige el tipo de disco que desea, sin importar si se trata de discos persistentes respaldados por discos duros estándar o unidades de estado sólido, SSD locales, Cloud Storage o una mezcla. Incluso puede configurar interfaces de redes y ejecutar una combinación de máquinas de Linux y Windows. Analizaremos estas opciones de manera más detallada más adelante en el módulo.

Características de Compute Engine

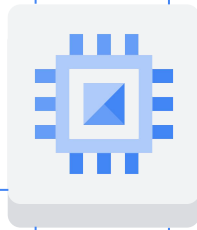
Redimensionamiento de máquinas

- Motor de recomendaciones para un tamaño de máquina óptimo
- Estadísticas de Cloud Monitoring
- Recomendaciones nuevas 24 horas después de la creación o el ajuste de las VM

- Metadatos de la instancia
- Secuencia de comandos de inicio

Políticas de disponibilidad:

- Migración en vivo
- Reinicio automático



Balanceo de cargas global:

- Varias regiones para la disponibilidad

Administración de parches de SO:

- Crear aprobaciones de parches
- Configurar la programación flexible
- Aplicar la configuración avanzada de parches

- Facturación por segundo
- Descuentos por uso continuo
- Descuentos por compromiso de uso

Interrumpible:

- Hasta un 80% de descuento
- Sin ANS

En este módulo se abordarán muchas características diferentes, como el redimensionamiento de máquinas, las secuencias de comandos de inicio, los metadatos, las políticas de disponibilidad, la administración de parches de SO y los descuentos en el precio y por uso.

Compute

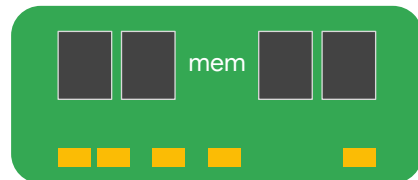
Muchos tipos de máquinas

- La capacidad de procesamiento de la red escala 2 Gbps por CPU virtual (pocas excepciones)
- Máximo teórico de 32 Gbps con 16 CPU virtuales o 100 Gbps con GPU T4 o V100

Una CPU virtual equivale a 1 hipersubproceso de hardware



CPU virtual



Para comenzar, analicemos las opciones de procesamiento.

Compute Engine proporciona muchos tipos de máquinas diferentes que analizaremos más adelante en este módulo. Si estas no satisfacen sus necesidades, puede personalizar su propia máquina.

La CPU que elija influirá en la capacidad de procesamiento de su red.

Específicamente, su red escalará a 2 Gbits por segundo por cada núcleo de CPU, excepto las instancias con 2 o 4 CPU, que recibirán hasta 10 Gbits por segundo de ancho de banda.

Al momento de esta grabación, la capacidad de procesamiento tenía un máximo teórico de 32 Gbits por segundo para una instancia con 16 o más CPU, y una capacidad de procesamiento con un máximo de 100 Gbits por segundo para instancias específicas que tienen GPU T4 y V100 conectadas.

Cuando migra desde una configuración local, está acostumbrado a los núcleos físicos, que tienen hipersubprocesos. En Compute Engine, cada CPU virtual se implementa como un hipersubproceso de hardware único en una de las plataformas de CPU disponibles.

Para obtener una lista actualizada de las plataformas de CPU disponibles, consulte la

Almacenamiento

Discos

- Estándar, SSD o SSD local
- Los PD SSD estándar y locales reducen la escala del rendimiento por cada GB de espacio asignado

Cambie el tamaño de los discos o migre instancias sin tiempo de inactividad.



Después de que elija sus opciones de procesamiento, le recomendamos elegir su disco.

Tiene tres opciones: Estándar, SSD o SSD local. En pocas palabras, ¿desea usar unidades de disco duro rotativos (HDD) estándar o de estado sólido (SSD) con memoria flash? Ambas opciones otorgan la misma cantidad de capacidad respecto al tamaño del disco cuando se elige un disco persistente. Por lo tanto, la pregunta en realidad apunta a la comparación del rendimiento y el costo, ya que cada opción tiene una estructura de precios diferente.

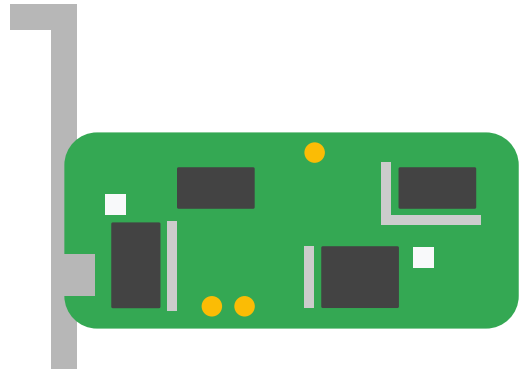
En términos sencillos, las SSD están diseñadas para entregarle una mayor cantidad de IOPS por dólar, mientras que los discos estándar le otorgan más capacidad por la inversión que realizó.

Las SSD locales incluso tienen una mayor capacidad de procesamiento y menos latencia que los discos persistentes SSD, ya que están conectados al hardware físico. Sin embargo, los datos que se almacenan en las SSD locales persisten solo hasta que detiene la instancia o la borra. Por lo general, las SSD locales se usan como discos de intercambio, del mismo modo que si se quisiera crear un disco RAM, pero, si se necesita más capacidad, se pueden almacenar en una SSD local. Puede crear instancias con un máximo de ocho particiones separadas de SSD local de

Redes

Características de las redes sólidas

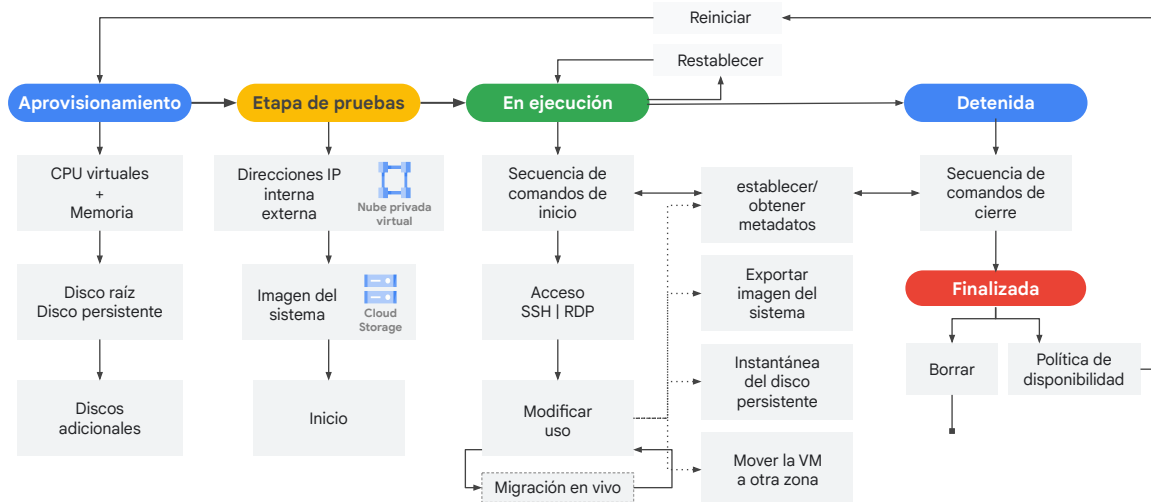
- Redes predeterminadas y personalizadas
- Reglas de firewall entrantes y salientes
 - Basadas en IP
 - Etiquetas de instancia y grupo
- Balanceo de cargas HTTPS regionales
- Balanceo de cargas de red
 - No necesita preparación previa
- Subredes globales y multirregionales



En cuanto a las redes, ya vimos las características que se aplican a Compute Engine en el lab del módulo anterior. Analizamos los distintos tipos de redes y creamos reglas de firewall con direcciones IP y etiquetas de red.

También notará que puede realizar balanceo de cargas de HTTPS regional y de red. Para esto no se necesita ninguna preparación previa, ya que los balanceadores de cargas no son dispositivos de hardware que necesiten analizar su tráfico. En términos simples, un balanceador de cargas es un conjunto de reglas de ingeniería de tráfico que llegan a la red de Google y la VPC aplica sus reglas destinadas al rango de subred de su dirección IP. Aprenderemos más sobre los balanceadores de cargas en un próximo curso de la serie Architecting with Google Compute Engine.

Ciclo de vida de las VM



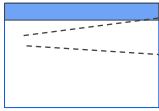
El ciclo de vida de una VM se representa con distintos estados. Abordaremos este ciclo de vida en términos generales, pero recomendamos que use este diagrama como referencia.

Cuando define todas las propiedades de una instancia y hace clic en Crear, la instancia entra al estado de aprovisionamiento. Aquí, los recursos como la CPU, la memoria y los discos se reservan para la instancia, pero la instancia todavía no se está ejecutando. A continuación, la instancia avanza al estado de etapa de pruebas, en la que los recursos se adquirieron y la instancia está lista para iniciarse. Específicamente, en este estado, Compute Engine agrega direcciones IP, inicia la imagen del sistema y también inicia el sistema.

Después de que la instancia comienza a ejecutarse, se pasa por secuencias de comandos de inicio configuradas previamente y habilitará el acceso a SSH o RDP. Ahora puede realizar varias acciones mientras se ejecuta su instancia. Por ejemplo, puede migrar en vivo su máquina virtual a otro host en la misma zona en lugar de solicitar que se reinicie la instancia. Esto le permite a Google Cloud realizar un mantenimiento integral para que la infraestructura esté protegida y sea confiable, sin interrumpir ninguna de las VM. Mientras se ejecuta la instancia, también puede mover su VM a otra zona, obtener una instantánea del disco persistente de la VM, exportar la imagen del sistema o volver a configurar los metadatos. Exploraremos

Cree una VM


1



Instancias de VM [+ CREAR INSTANCIA](#) [IMPORTAR VM](#)

console.google.com

2



\$ `gcloud compute instances create [instance-name]`

línea de comandos
incluido Cloud Shell

3

API de REST

Muchas
opciones
de VM

- Proyecto
- Región
- Zona
- Subred
- Tipo de máquina
- Opciones de discos
- Imagen
- Opciones de IP

Tiene tres opciones para crear y configurar una VM. Puede usar GCP Console (como lo hizo en el lab anterior), la línea de comandos de Cloud Shell o la API de RESTful. Si desea automatizar y procesar parámetros de configuración muy complejos, le recomendamos que los configure de manera programática con la API de RESTful. Para hacerlo, defina todas las opciones de su entorno.

Si pretende usar la línea de comandos o la API de RESTful, le recomendamos que primero configure la instancia mediante GCP Console y, luego, le pida a Compute Engine la solicitud de REST o línea de comandos equivalente, como le explicamos en la demostración anterior. Así, evita los errores tipográficos y obtiene listas desplegables con todas las opciones de CPU y memoria disponibles.

Hablando de opciones de CPU y memoria, veamos los distintos tipos de máquinas disponibles actualmente.

Tipos de máquina

Tipos predefinidos de máquinas: proporción de GB de memoria por CPU virtual

- Estándar
- Capacidad de memoria alta
- Capacidad de CPU alta
- Con optimización de memoria
- Optimizada para procesamiento
- Núcleo compartido

Tipos personalizados de máquinas:

- Usted especifica la cantidad de memoria y de CPU virtuales.

Un tipo de máquina especifica un conjunto particular de recursos de hardware virtualizados disponibles para una instancia de VM, incluidos la cantidad de memoria del sistema, el recuento de CPU virtuales y la capacidad máxima de los discos persistentes. GCP cuenta con muchos tipos de máquinas que se pueden agrupar en 2 categorías:

- Tipos predefinidos de máquinas: Estos tienen un conjunto de recursos fijo, los administra Compute Engine y están disponibles en varias clases diferentes. Cada clase tiene una proporción de GB de memoria por CPU virtual definida previamente. Incluye los siguientes tipos de máquinas:
 - Tipos de máquinas estándar
 - Tipos de máquinas con alta capacidad de memoria
 - Tipos de máquinas con alta capacidad de CPU
 - Tipos de máquinas con optimización de memoria
 - Tipos de máquinas con optimización de procesamiento
 - Tipos de máquinas de núcleo compartido
- Tipos personalizados de máquinas: Estos le permiten especificar la cantidad de CPU virtuales y de memoria para su instancia.

Exploremos cada uno de estos tipos de máquinas, pero recuerde que ellos y sus opciones disponibles pueden cambiar.

Tipos de máquinas estándar

3.75 GB de memoria
1 CPU virtual

Nombre de la máquina	CPU virtuales	Memoria (GB)	Cant. máx. de PD	Tamaño total máx. de PD
n1-standard-1	1	3.75	128	257 TB
n1-standard-2	2	7.50		
n1-standard-4	4	15		
n1-standard-8	8	30		
n1-standard-16	16	60		
n1-standard-32	32	120		
n1-standard-64	64	240		
n1-standard-96	96	360		

Los tipos de máquinas estándar son adecuados para las tareas que tienen necesidades equilibradas de CPU y memoria. Los tipos de máquinas estándar tienen 3.75 GB de memoria por CPU virtual. Los parámetros de configuración de CPU virtuales vienen en intervalos diferentes desde 1 CPU virtual hasta 96 CPU virtuales, como se muestra en esta tabla. Cada una de estas máquinas admite un máximo de 128 discos persistentes con un tamaño total de 257 TB para cada PD, que también es el caso en los tipos de máquina con alta capacidad de memoria, alta capacidad de CPU, con optimización de memoria y optimizados para procesamiento.

Tipos de máquinas con alta capacidad de memoria

6.5 GB de memoria

1 CPU virtual

Nombre de la máquina	CPU virtuales	Memoria (GB)	Cant. máx. de PD	Tamaño total máx. de PD
n1-highmem-2	2	13	128	257 TB
n1-highmem-4	4	26		
n1-highmem-8	8	52		
n1-highmem-16	16	104		
n1-highmem-32	32	208		
n1-highmem-64	64	416		
n1-highmem-96	96	624		

Los tipos de máquinas con alta capacidad de memoria son ideales para las tareas que requieren más memoria en relación con las CPU virtuales. Los tipos de máquinas con alta capacidad de memoria tienen 6.50 GB de memoria de sistema por CPU virtual. De manera similar a los tipos de máquinas estándar, los parámetros de configuración de CPU virtuales vienen en intervalos diferentes, desde 2 hasta 96 CPU virtuales, como se muestra en esta tabla.

Tipos de máquinas con alta capacidad de CPU

0.9 GB de memoria

1 CPU virtual

Nombre de la máquina	CPU virtuales	Memoria (GB)	Cant. máx. de PD	Tamaño total máx. de PD
n1-highcpu-2	2	1.80	128	257 TB
n1-highcpu-4	4	3.60		
n1-highcpu-8	8	7.20		
n1-highcpu-16	16	14.4		
n1-highcpu-32	32	28.8		
n1-highcpu-64	64	57.6		
n1-highcpu-96	96	86.4		

Los tipos de máquinas con alta capacidad de CPU son ideales para las tareas que requieren más CPU virtuales en relación con la memoria. Los tipos de máquinas con alta capacidad de CPU tienen 0.90 GB de memoria por CPU virtual.

Tipos de máquinas con optimización de memoria

Más de 14 GB de memoria

1 CPU virtual

Nombre de la máquina	CPU virtuales	Memoria (GB)	Cant. máx. de PD	Tamaño total máx. de PD
m1-ultramem-40	40	961	128	257 TB
m1-ultramem-80	80	1922		
m1-megamem-96	96	1433.6		
m1-ultramem-160	160	3844		

Los tipos de máquinas con optimización de memoria son ideales para las tareas que requieren un uso intensivo de la memoria y una proporción memoria-CPU virtual más alta que los tipos de máquinas con alta capacidad de memoria. Estos tipos de máquinas son perfectos para las bases de datos y estadísticas en la memoria, como SAP HANA, cargas de trabajo de almacenamiento empresarial, análisis genómico y servicios de análisis de SQL. Los tipos de máquinas con optimización de memoria tienen más de 14 GB de memoria por CPU virtual.

Estas máquinas tienen 4 parámetros de configuración, como se muestra en esta tabla. Actualmente, solo m1-megamem-96 admite una SSD local.

Tipos de máquinas con optimización de procesamiento

El rendimiento más alto por CPU virtual (turbo de núcleo completo continuo de hasta 3.8 GHz)

Nombre de la máquina	CPU virtuales	Memoria (GB)	Cant. máx. de PD	Tamaño total máx. de PD
c2-standard-4	4	16	128	257 TB
c2-standard-8	8	32		
c2-standard-16	16	64		
c2-standard-30	30	120		
c2-standard-60	60	240		

Los tipos de máquinas optimizados para procesamiento son ideales para cargas de trabajo que hacen un uso intensivo del procesamiento. Ofrecen el mayor rendimiento por núcleo de Compute Engine. Los tipos de máquinas C2, basados en los procesadores escalables Intel de última generación (Cascade Lake), ofrecen hasta 3.8 GHz de turbo de núcleo completo continuo y proporcionan transparencia total en la arquitectura de las plataformas de servidores subyacentes, lo que permite un ajuste avanzado del rendimiento. Los tipos de máquinas C2 cuentan con mucha más potencia de procesamiento, se ejecutan en una plataforma más reciente y, en general, son más confiables para lidiar con cargas de trabajo de procesamiento intensivo que los tipos de máquinas con alta capacidad de CPU N1.

Tipos de máquinas de núcleo compartido

Nombre de la máquina	CPU virtuales	Memoria (GB)	Cant. máx. de PD	Tamaño total máx. de PD
f1-micro	0.2	0.60	16	3 TB
g1-small	0.5	1.70		

Los tipos de máquinas de núcleo compartido proporcionan una CPU virtual que puede ejecutarse en menos tiempo del habitual en un único hipersubproceso de hardware de la CPU host que ejecuta su instancia. Las instancias con núcleo compartido pueden ser más rentables para ejecutar aplicaciones pequeñas que no requieren de más recursos que otros tipos de máquinas. Existen solo dos tipos de máquinas de núcleo compartido entre las que elegir:

- f1-micro
- g1-small

Los tipos de máquinas f1-micro cuentan con funciones de aumentos de actividad que permiten que las instancias usen capacidad adicional de CPU física por períodos breves. Los aumentos de actividad se activan automáticamente cuando su instancia necesita más capacidad de CPU física que la asignada en un principio. Durante estos períodos de actividad máxima, su instancia aprovechará la capacidad de CPU física disponible en aumentos de actividad. Tenga en cuenta que los aumentos de actividad no son permanentes y solo pueden usarse de forma periódica.

Para obtener información actualizada sobre todos estos tipos de máquinas, consulte la [página de documentación](#).

Cree tipos personalizados de máquinas

Cuándo seleccionar los tipos personalizados:

- Cuando los requisitos se ajustan entre los tipos predefinidos
- Cuando se necesita más memoria o más CPU

Personalice la cantidad de memoria y CPU virtuales para su máquina:

- 1 CPU virtual o un número par de CPU virtuales
- 0.9 GB hasta 6.5 GB por CPU virtual (predeterminado)
- La memoria total debe ser un múltiplo de 256 MB

Tipo de máquina
Para seleccionar núcleos, memoria y GPU, haz clic en Personalizar.

[Vista básica](#)

Núcleos
1 CPU virtual 1-96

Memoria
3.75 GB 1-6.5

☐ Extender memoria ?

Plataforma de CPU ?
Automática

GPU
La cantidad de bloques de GPU se vincula con la cantidad de núcleos de CPU y memoria seleccionados para esta instancia. Con la configuración actual, debes seleccionar, como mínimo, 1 bloque de GPU de este tipo. [Más información](#)

Cantidad de GPU **Tipo de GPU**
Ninguna NVIDIA Tesla K80

Si ninguno de los tipos predefinidos de máquinas satisface sus necesidades, puede especificar de manera independiente la cantidad de CPU virtuales y de memoria para su instancia. Los tipos personalizados de máquinas son ideales para las situaciones siguientes:

- Cuando tiene cargas de trabajo que no se ajustan bien a los tipos predefinidos de máquinas disponibles para usted
- Cuando tiene cargas de trabajo que requieren más capacidad de procesamiento o memoria, pero no necesitan todas las mejoras que proporciona el tipo predefinido de máquina que le sigue en tamaño

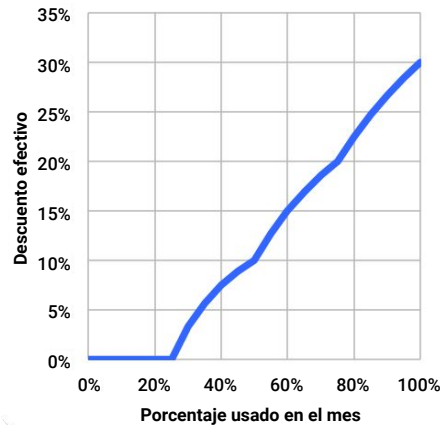
El uso de un tipo personalizado de máquina cuesta un poco más que un tipo predefinido de máquina equivalente y, aun así, existen algunas limitaciones relacionadas con la cantidad de memoria y CPU virtuales que puede seleccionar:

- Solo se pueden crear tipos de máquina con 1 CPU virtual o un número par de CPU virtuales.
- La memoria debe ser entre 0.9 GB y 6.5 GB por CPU virtual (de forma predeterminada).
- La memoria total de la instancia debe ser un múltiplo de 256 MB.

De forma predeterminada, una máquina personalizada puede tener hasta 6.5 GB de memoria por CPU virtual. Sin embargo, es posible que esta cantidad de memoria no

Descuentos por uso continuo

Nivel de uso (porcentaje del mes)	Porcentaje en el que se cobra el incremental
0%-25%	100% de la tarifa base
25%-50%	80% de la tarifa base
50%-75%	60% de la tarifa base
75%-100%	40% de la tarifa base



Hasta un 30% de descuento neto en las instancias que se ejecutan el mes completo

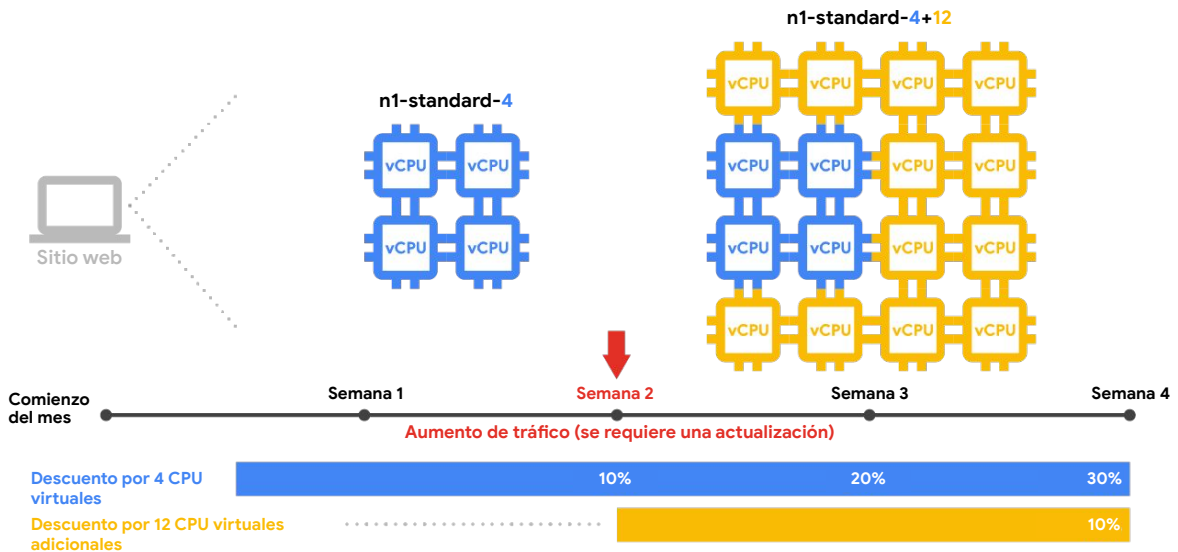
Los descuentos por uso continuo son descuentos automáticos que obtiene por ejecutar recursos específicos de Compute Engine (CPU virtuales, memoria o dispositivos de GPU) durante una parte significativa del mes de facturación. Por ejemplo, cuando ejecuta uno de estos recursos durante un período total superior al 25% de un mes, Compute Engine le ofrece automáticamente un descuento por cada minuto incremental que usa en esa instancia. El descuento aumenta con el uso, y puede obtener hasta un 30% de descuento neto para instancias que se ejecutan todo el mes.

En la tabla de esta diapositiva, se describe el descuento que obtiene en cada nivel de uso de una instancia de VM. Para aprovechar el descuento completo del 30%, debe crear sus instancias de VM el primer día del mes, ya que los descuentos se restablecen al comienzo de cada mes.

En el gráfico de esta diapositiva, se demuestra como su descuento efectivo aumenta con el uso. Por ejemplo, si usa una máquina virtual durante el 50% del mes, obtendrá un descuento efectivo del 10%. Si la usa durante el 75% del mes, obtendrá un descuento efectivo del 20%. Si la usa durante el 100% del mes, obtendrá un descuento efectivo del 30%. También puede usar la calculadora de precios de GCP a fin de estimar su descuento por uso continuo para cualquier carga de trabajo arbitraria. Para usar la calculadora, consulte la sección de vínculos de este video: [\[https://cloud.google.com/products/calculator/\]](https://cloud.google.com/products/calculator/)

Compute Engine calcula los descuentos por uso continuo según la utilización de CPU

Descuentos por uso continuo (ejemplo)



Veamos un ejemplo en el que tiene dos instancias que están en una misma región, pero cuyos tipos de máquinas son diferentes y se ejecutan en momentos distintos del mes. Compute Engine desglosa la cantidad de CPU virtuales y memoria que se usa en todas las instancias con tipos predefinidos de máquinas y combina los recursos a fin de calificar para los mayores descuentos por uso continuo posibles. Como se muestra en esta diapositiva, usted ejecuta las siguientes instancias en la región us-central1 durante un mes:

- Durante la primera mitad del mes, ejecuta una instancia n1-standard-4 con 4 CPU virtuales y 15 GB de memoria.
- Durante la segunda mitad del mes, ejecuta una instancia n1-standard-16 más grande con 16 CPU virtuales y 60 GB de memoria.

En este caso, Compute Engine reorganiza los tipos de máquinas en recursos de memoria y CPU virtuales individuales, y combina su utilización para crear los siguientes recursos, como se muestra en la parte inferior:

- 4 CPU virtuales y 15 GB de memoria durante un mes completo
- 12 CPU virtuales y 45 GB de memoria durante la mitad del mes

Interrumpible

- Precio más bajo por un servicio interrumpible (hasta el 80%)
- La VM se puede finalizar en cualquier momento
 - No se aplican cargos si se finaliza en el primer minuto
 - Máx. 24 horas
 - Advertencia de finalización de 30 segundos, pero no garantizada
 - Tiempo para una secuencia de comandos de cierre

Recuerde que es posible que estas VM se interrumpan en cualquier momento y, si ocurre en el primer minuto, no se aplicarán cargos. Además, las VM interrumpibles solo existirán por un máximo de 24 horas, y usted solo recibirá una notificación 30 segundos antes de que se interrumpa la máquina.

Interrumpible

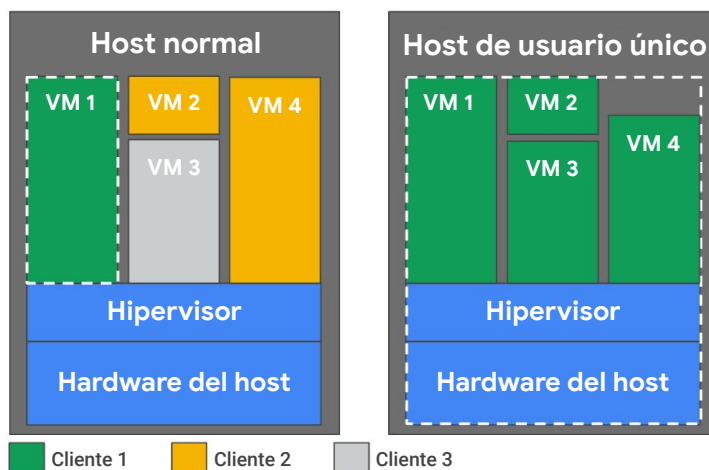
- Precio más bajo por un servicio interrumpible (hasta el 80%)
- La VM se puede finalizar en cualquier momento
 - No se aplican cargos si se finaliza en el primer minuto
 - Máx. 24 horas
 - Advertencia de finalización de 30 segundos, pero no garantizada
 - Tiempo para una secuencia de comandos de cierre
- Sin migración en vivo ni reinicio automático
- Puede solicitar que la cuota de CPU de una región se divida entre normal y con interrupción
 - Predeterminado: las VM interrumpibles se cuentan en función de la cuota de CPU de la región

Google Cloud

Un caso de uso principal de las VM interrumpibles es ejecutar trabajos de procesamiento por lotes. Si algunas de estas instancias finalizan durante el procesamiento, el trabajo se ralentiza, pero no se detiene por completo. Por lo tanto, las instancias interrumpibles completan las tareas de procesamiento por lotes sin colocar cargas de trabajo adicionales en las instancias existentes. Además, usted no debe pagar el precio total por las instancias normales adicionales.

[https://cloud.google.com/compute/docs/instances/preemptible#what_is_a_preemptible_instance]

Los nodos de usuario único aíslan físicamente las cargas de trabajo



Puede usar licencias adquiridas por el usuario.

Google Cloud

Si tiene cargas de trabajo que necesitan aislamiento físico de otras cargas de trabajo o máquinas virtuales para satisfacer los requisitos de cumplimiento, le recomendamos los nodos de usuario único.

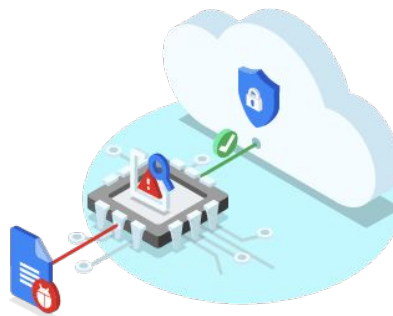
Un nodo de usuario único es un servidor físico de Compute Engine que se usa con el fin de alojar instancias de VM solo para su proyecto específico. Utilícelos a fin de mantener sus instancias separadas físicamente de aquellas en otros proyectos o para agrupar sus instancias en el mismo hardware de host, por ejemplo, si tiene una carga de trabajo de procesamiento de pagos que necesita aislarse para cumplir con los requisitos de cumplimiento.

En el diagrama de la izquierda, se muestra un host normal con varias instancias de VM de varios clientes. Como se puede apreciar en el diagrama de la derecha, un nodo de usuario único también tiene varias instancias de VM, pero todas pertenecen al mismo proyecto. Al día en que se realizó esta grabación, el único tipo de nodo disponible puede alojar instancias de VM de hasta 96 CPU virtuales y 624 GB de memoria. También puede completar el nodo con varias instancias de VM más pequeñas de distintos tamaños, incluidos instancias con memoria extendida y tipos personalizados de máquinas. Además, si tiene licencias de sistema operativo existentes, puede usarlas en Compute Engine con nodos de usuario único y, a su vez, minimizar el uso del núcleo físico con la función de reinicio local.

Si desea obtener información para crear nodos y ubicar sus instancias en ellos, consulte la sección de vínculos de este video.

Las VM protegidas cuentan con integridad verificable

- Inicio seguro
- Módulo de plataforma segura virtual (vTPM)
- Supervisión de integridad



Requiere de una imagen protegida

Google Cloud

Otra opción de procesamiento es crear VM protegidas. Estas cuentan con integridad verificable de sus instancias de VM, por lo que puede confiar en que el software malicioso y los rootkits a nivel de inicio y kernel no vulneraron sus instancias. La integridad verificable de la VM protegida se logra a través de Inicio seguro, el inicio medido habilitado para el módulo de plataforma segura virtual o vTPM y la supervisión de integridad.

Las VM protegidas son la primera oferta de la iniciativa Shielded Cloud. El objetivo de Shielded Cloud es proporcionar una base aún más segura para todo Google Cloud a través de integridad verificable y funciones específicas, como la protección de vTPM o el sellado, que previenen el robo de datos.

Para poder usar estas funciones de la VM protegida, debe seleccionar una imagen protegida. Aprenderemos sobre las imágenes en la próxima sección.

Las Confidential VMs le permiten encriptar datos en uso

- Encripta datos mientras se procesan
- Fácil de usar sin realizar cambios en el código ni afectar el rendimiento
- VM N2D de Compute Engine que se ejecuta en procesadores AMD Epyc de segunda generación

Una máquina virtual confidencial (Confidential VM) es un tipo de VM N2D de Compute Engine que se ejecuta en hosts basados en la segunda generación de procesadores AMD Epyc, cuyo nombre interno es “Rome”.

Las Confidential VMs le permiten encriptar datos en uso

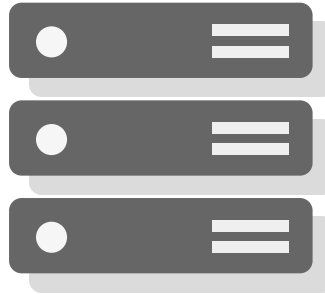
- Encripta datos mientras se procesan
- Fácil de usar sin realizar cambios en el código ni afectar el rendimiento
- VM N2D de Compute Engine que se ejecuta en procesadores AMD Epyc de segunda generación
- Proporciona alta capacidad de memoria y de procesamiento, y admite cargas de trabajo simultáneas y que requieren de mucho procesamiento
- Puede seleccionar el servicio Confidential VM cuando cree una VM nueva

Google Cloud

Puede seleccionar el servicio de Confidential VM cuando cree una VM nueva mediante Google Cloud Console, la API de Compute Engine o la herramienta de línea de comandos de gcloud.

¿Qué contiene una imagen?

- Cargador de inicio
- Sistema operativo
- Estructura del sistema de archivos
- Software
- Personalizaciones



Cuando cree una máquina virtual, puede elegir la imagen de disco de arranque. Esta incluye el cargador de inicio, el sistema operativo, la estructura del sistema de archivos, todo el software preconfigurado y todas las demás personalizaciones.

Imágenes de máquina

Situaciones	Imagen de máquina	Instantánea de disco persistente	Imagen personalizada	Plantilla de instancia
Copia de seguridad de disco único	Sí	Sí	Sí	No
Copia de seguridad de múltiples discos	Sí	No	No	No
Copia de seguridad diferenciales	Sí	Sí	No	No
Clonación y replicación de instancias	Sí	No	Sí	Sí
Configuración de la instancia de VM	Sí	No	No	Sí

Una imagen de máquina es un recurso de Compute Engine que almacena toda la configuración, los metadatos, los permisos y los datos de uno o más discos necesarios para crear una instancia de máquina virtual (VM). Puede usar una imagen de máquina en muchas situaciones de mantenimiento del sistema, como la creación de instancias, la recuperación y la creación de copias de seguridad y la clonación de instancias.

Las imágenes de máquina son los recursos más adecuados para crear copias de seguridad de discos y la clonación y replicación de instancias.

Tenga en cuenta que, al momento de redactar este documento, las imágenes siguen en etapa beta.

Disco de arranque

- La VM incluye un disco persistente raíz.
- La imagen se carga en un disco raíz durante el primer inicio.
 - Inicialable: puede conectarse a una VM y, luego, iniciarse desde ella.
 - Duradero: puede seguir funcionando después de la finalización de la VM.
- Algunas imágenes de SO están personalizadas para Compute Engine.
- Pueden seguir funcionando después de la eliminación de la VM si está inhabilitada la opción “Borrar el disco de arranque cuando se borre la instancia”.

Todas las VM incluyen un disco persistente raíz único, ya que usted elige una imagen base en la que cargarlo.

Esta imagen es inicialable porque puede conectarla a una VM y, luego, iniciarla a partir de ella, y es duradera debido a que puede seguir funcionando si la VM finaliza. Para que un disco de arranque funcione después de la eliminación de una VM, debe inhabilitar la opción “Borrar el disco de arranque cuando se borre la instancia” en las propiedades de la instancia.

Como mencionamos anteriormente, existen diferentes tipos de discos. Explorémoslos de manera más detallada.

Discos persistentes

Almacenamiento de red que aparece como un dispositivo de almacenamiento en bloques

- Conectado a una VM mediante la interfaz de red
- Almacenamiento duradero: puede funcionar después de la finalización de la VM
- Iniciable: puede conectarse a una VM y, luego, iniciarse desde ella
- Instantáneas: copias de seguridad incrementales
- Rendimiento: escala con el tamaño
- Opciones de HDD (magnética) o SSD (más rápida, estado sólido)
- Cambio del tamaño del disco, incluso cuando se está ejecutando y está conectado
- Puede estar conectado en modo de solo lectura a varias VM
- Zonal o regional
 - pd-standard
 - pd-ssd
 - pd-balanced
 - pd-extreme (solo zonal)
- Claves de encriptación:
 - Google-managed
 - Administrada por el cliente
 - Proporcionada por el cliente

El primer disco que creamos es lo que conocemos como disco persistente. Esto significa que se conectará a la VM mediante la interfaz de red. Aunque sea persistente, no está conectado de manera física a la máquina. La separación del disco y el procesamiento permite que el disco siga funcionando si la VM finaliza. También puede obtener instantáneas de estos discos, que son copias de seguridad incrementales que analizaremos más adelante.

Elegir discos HDD o SSD es una decisión que se toma según el costo y rendimiento. Para obtener más información sobre el rendimiento de los discos y cómo se escala con el tamaño, consulte la sección de vínculos de este video:

[\[https://cloud.google.com/compute/docs/disks/performance\]](https://cloud.google.com/compute/docs/disks/performance)

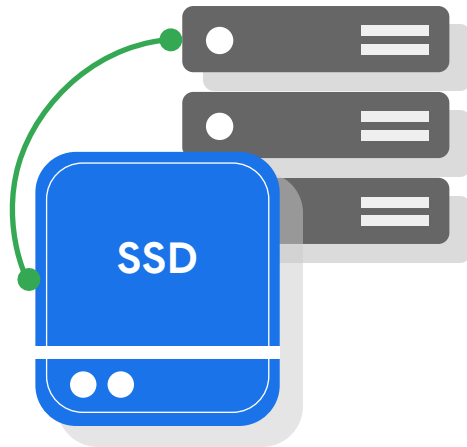
Otra gran función de los discos persistentes es que usted puede cambiarles el tamaño de manera dinámica, incluso cuando se están ejecutando y están conectados a una VM.

También puede conectar un disco en modo de solo lectura a varias VM. Esto le permite compartir datos estáticos entre varias instancias, lo cual es más económico que replicar sus datos en discos únicos para instancias individuales.

Los discos persistentes zonales cuentan con almacenamiento en bloque eficiente y

Los discos SSD locales están conectados físicamente a una VM

- Más IOPS, menos latencia y mayor capacidad de procesamiento que los discos persistentes
- Disco de 375 GB (hasta ocho), total de 3 TB
- Los datos siguen funcionando después del restablecimiento, pero no de la detención ni la finalización de una VM
- Específicos de VM: no se pueden volver a conectar a otras VM



Las SSD locales son diferentes de los discos persistentes, ya que están conectadas físicamente a la máquina virtual. Por lo tanto, estos discos son efímeros, pero otorgan IOPS muy altas. Para obtener cifras actualizadas, le recomendamos que consulte la documentación

[\[https://cloud.google.com/compute/docs/disks/performance\]](https://cloud.google.com/compute/docs/disks/performance).

Por el momento, puede conectar hasta 8 discos SSD locales con 375 GB cada uno, lo que nos da un total de 3 TB.

Los datos en estos discos seguirán funcionando después del restablecimiento de una VM, pero no de detenerla ni finalizarla, ya que no se pueden volver a conectar a otras VM.

Disco RAM

- tmpfs
- Más rápido que el disco local, pero más lento que la memoria
 - Úselo cuando la aplicación espera una estructura de sistema de archivos y no puede almacenar directamente sus datos en la memoria
 - Disco para instalación nueva rápido o almacenamiento en caché rápido
- Muy volátil: borra en la detención o el reinicio
- Es posible que necesite un tipo de máquina más grande si se ajustó el tamaño de la RAM para la aplicación
- Considere usar un disco persistente para crear una copia de seguridad de los datos del disco RAM

También tiene la opción de usar un disco RAM.

Simplemente puede utilizar tmpfs si desea almacenar datos en la memoria. Este será el tipo de rendimiento más rápido disponible si necesita estructuras de datos pequeñas. Recomendamos una máquina virtual con capacidad de memoria alta si necesita aprovechar esas funciones, además de un disco persistente para crear una copia de seguridad de los datos del disco RAM.

Resumen de las opciones de discos

	Disco persistente HDD	Disco persistente SSD	Disco SSD local	Disco RAM
Redundancia de datos	Sí	Sí	No	No
Encriptación en reposo	Sí	Sí	Sí	N/A
Captura de instantáneas	Sí	Sí	No	No
Iniciable	Sí	Sí	No	No
Caso de uso	Almacenamiento masivo y general de archivos	IOPS muy aleatorias	IOPS altas y latencia baja	Latencia y riesgo de pérdida de datos bajos

En resumen, tiene varias opciones de discos. Los discos persistentes se pueden reiniciar, y es posible obtener instantáneas de ellos, pero los discos SSD y RAM locales son efímeros.

Recomendamos que elija un disco HDD persistente cuando no necesite rendimiento, sino que solo capacidad. Si tiene necesidades de alto rendimiento, consulte las opciones de SSD. Los discos persistentes cuentan con redundancia de datos porque la información en cada disco persistente se distribuye entre muchos discos físicos.

Las SSD locales cuentan con rendimiento incluso más alto, pero sin redundancia de datos.

Por último, los discos RAM son muy volátiles, pero ofrecen el rendimiento más alto.

Diferencias de administración de los discos persistentes



Discos persistentes de la nube

- Un sistema de archivos único es mejor
- Cambia el tamaño de los discos (los aumenta)
- Cambia el tamaño del sistema de archivos
- Servicio de instantáneas integrado
- Encriptación automática



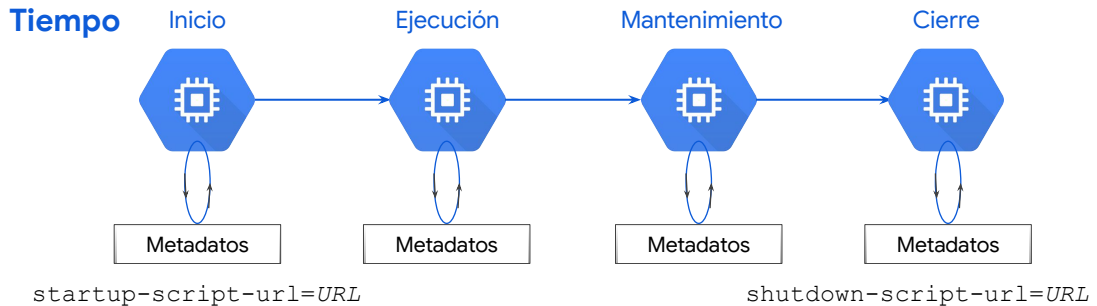
Disco de hardware de la computadora

- Partición
- Vuelve a particionar el disco
- Cambia el formato
- Arrays de disco redundantes
- Instantáneas y administración de subvolumenes
- Encripta archivos antes de escribirlos en el disco

Existen muchas diferencias entre un disco duro físico de una computadora y uno persistente, que es, en términos sencillos, un dispositivo virtual conectado a la red. En primer lugar, recuerde que debe particionar los discos de hardware de computadoras normales. Dicho en términos sencillos, se reserva una parte de la unidad para que el sistema operativo obtenga su propia capacidad. Si desea aumentarla, debe volver a particionar, y si quiere realizar cambios, es posible que incluso deba cambiar el formato. Si desea redundancia, podría crear un array de disco redundante, y si quiere encriptación, debe encriptar archivos antes de escribirlos en el disco.

Con los discos persistentes de la nube, todo es diferente, ya que esa administración se maneja por usted en el backend. Simplemente puede desarrollar discos y cambiar el tamaño del sistema de archivos, ya que los discos son dispositivos virtuales conectados a la red. La redundancia y los servicios de instantáneas están integrados y los discos se encriptan automáticamente. Incluso puede usar sus propias claves y, así, se asegurará de que ninguna parte pueda acceder a los datos, excepto usted.

Metadatos y secuencias de comandos



Cada instancia de VM almacena sus metadatos en un servidor de metadatos. El servidor de metadatos es muy útil junto con las secuencias de comandos de inicio y apagado, ya que puede usarlo para obtener información única sobre una instancia, sin autorización adicional y de manera programática. Por ejemplo, puede escribir una secuencia de comandos de inicio que obtenga el par clave-valor de metadatos de la dirección IP externa de una instancia y usarlo en su secuencia de comandos a fin de configurar una base de datos. Debido a que las claves de metadatos predeterminadas son las mismas en cada instancia, puede volver a usar la secuencia de comandos sin actualizarla para cada instancia. Esto lo ayudará a crear un código menos frágil para las aplicaciones.

Almacenar y recuperar metadatos de instancias es una acción muy común de Compute Engine. Recomendamos que almacene las secuencias de comandos de inicio y apagado en Cloud Storage, como lo descubrirá en el próximo lab de este módulo.

Mueva una instancia a una zona nueva

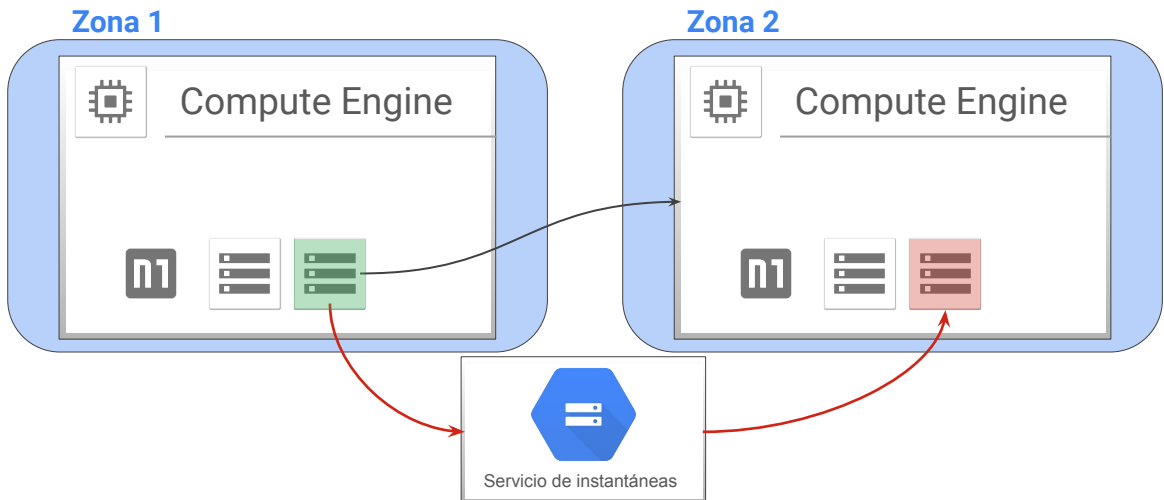
- Proceso automatizado (cambio dentro de una región):
 - `gcloud compute instances move`
 - Actualiza las referencias de la VM: no es automático
- Proceso manual (cambio entre regiones):
 - Cree instantáneas de todos los discos persistentes en la misma VM de origen.
 - Cree discos persistentes nuevos en la zona de destino restablecida a partir de instantáneas.
 - Cree VM nuevas en la zona de destino y conecte discos persistentes nuevos.
 - Asigne una IP estática a la VM nueva.
 - Actualice las referencias de la VM.
 - Borre las instantáneas, los discos originales y la VM original.

Si mueve su instancia dentro de una misma región, puede automatizar el cambio mediante el comando `gcloud compute instances move`.

Si mueve su instancia a una región diferente, debe hacerlo de forma manual siguiendo el proceso que se describe aquí. Este implica crear una instantánea de todos los discos persistentes y discos nuevos en la zona de destino a partir de la instantánea. A continuación, debe crear la VM nueva en la zona de destino y conectar los discos persistentes nuevos, asignar una IP estática y actualizar todas las referencias de la VM. Por último, debe borrar la VM original, sus discos y la instantánea.

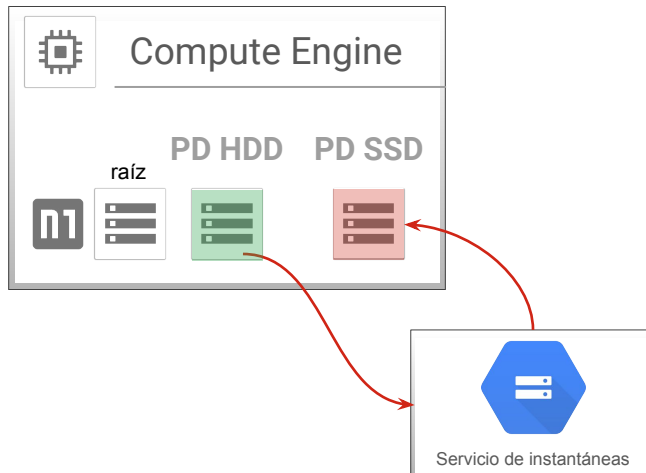
Hablando de instantáneas, veámoslas en detalle.

Instantánea: migre datos entre zonas



Las instantáneas también se pueden usar para migrar datos entre zonas. Analizamos este punto cuando vimos el proceso manual de mover una instancia entre dos regiones, pero también se puede usar para transferir datos de una zona a otra. Por ejemplo, es posible que desee minimizar la latencia migrando datos a una unidad que se pueda conectar de manera local en la zona en que se usa.

Instantánea: transfiera a SSD para mejorar el rendimiento



Lo que me lleva a otro caso de uso de instantáneas para transferir datos a un tipo de disco diferente. Por ejemplo, si desea mejorar el rendimiento del disco, podría usar una instantánea para transferir datos de un disco persistente HDD estándar a un disco persistente SSD.

Instantáneas de discos persistentes

- Las instantáneas no están disponibles para SSD locales.
- Crea una copia de seguridad *incremental* de Cloud Storage.
 - No se puede ver en *sus* buckets; la administra el servicio de instantáneas.
 - Considere los trabajos cron para crear copias de seguridad incrementales de manera periódica.
- Las instantáneas se pueden restablecer a un disco persistente nuevo.
 - El disco nuevo puede estar en otra región o zona del mismo proyecto
 - Base de una migración de VM: “mover” una VM a una zona nueva
 - Las instantáneas no crean copias de seguridad de los metadatos, etiquetas, etc., de las VM

Ahora que abordamos algunos de los casos de uso de las instantáneas, exploremos el concepto de una instantánea de disco.

En primer lugar, el título de la diapositiva es “instantáneas de discos persistentes” porque las instantáneas están disponibles solo para los discos persistentes y no las SSD locales.

Las instantáneas son diferentes de las imágenes públicas y personalizadas, que se usan principalmente a fin de crear instancias o configurar plantillas de instancias. Las instantáneas son útiles para crear copias de seguridad periódicas de los datos de sus discos persistentes.

Las instantáneas son incrementales y se comprimen automáticamente, por lo que puede crear instantáneas frecuentes en un disco persistente más rápido y a un costo mucho menor que si creara una imagen completa del disco con frecuencia. Como vimos en los ejemplos anteriores, las instantáneas se pueden restablecer a un disco persistente nuevo, lo que permite que se mueva a una zona nueva.

Para crear una instantánea de un disco persistente, consulte la sección de vínculos de este video: [\[https://cloud.google.com/compute/docs/disks/create-snapshots\]](https://cloud.google.com/compute/docs/disks/create-snapshots)

Repaso del lab

Cómo trabajar con máquinas virtuales



En este lab, creé una instancia personalizada de una máquina virtual mediante la instalación de un software base (un Java Runtime Environment sin interfaz gráfica) y un software de aplicación (un servidor de juegos de Minecraft).

Preparé y conecté una SSD de alta velocidad con el fin de personalizar la VM, y reservé una IP externa estática para que la dirección siguiera siendo coherente.

Luego, con esa dirección IP, verifiqué la disponibilidad del servidor de videojuegos en línea.

Después, configuré un sistema de copia de seguridad para respaldar los datos del servidor en un bucket de Cloud Storage y probé el sistema de copia de seguridad. Luego, automatizó las copias de seguridad con cron.

Por último, configuré secuencias de comandos de mantenimiento con metadatos para iniciar y apagar el servidor correctamente.

Muchas de estas técnicas, incluida la automatización de secuencias de comandos, se pueden adaptar para administrar servidores de producción en cualquier aplicación.

Puede continuar con un recorrido por el lab, pero recuerde que la interfaz de usuario de GCP puede cambiar, por lo que su entorno podría ser un poco diferente.

Essential Cloud Infrastructure: Core Services

1. Cloud IAM
2. Servicios de almacenamiento de datos
3. Administración de recursos
4. Supervisión de recursos



A continuación, recomendamos que se inscriba en el curso “Essential Cloud Infrastructure: Core Services” de la serie “Architecting with Google Compute Engine”.

Aquí comenzamos abordando Cloud IAM. Luego, llevará a cabo la administración de identidades y accesos para los recursos.

A continuación, analizaremos los distintos servicios de almacenamiento de datos disponibles en GCP, y también implementará algunos de esos servicios.

Luego, exploraremos la administración de recursos, y administrará y examinará la facturación de recursos de GCP.

Por último, hablaremos sobre la supervisión de recursos y usará servicios de Stackdriver para supervisar los recursos de GCP.

Esperamos que disfrute el curso.