

Arquitecturas de nube con AWS

Ing. Fernando Lichtschein

Ing. Mora Villa Abrille

3. Servicios de cómputo

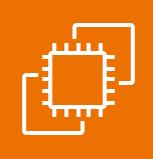
Amazon EC2

Objetivos

- Identificar cómo usar Amazon Elastic Compute Cloud (Amazon EC2) en una arquitectura.
- Explicar el valor de usar Amazon Machine Images (AMI) para acelerar la creación y la repetibilidad de la infraestructura.
- Recomendar tipos de instancias de EC2 según los requerimientos.
- Recomendar soluciones de almacenamiento para Amazon EC2.
- Reconocer cómo configurar instancias de Amazon EC2 con datos de usuario.
- Describir las opciones de precios de EC2 y hacer recomendaciones basadas en el costo.
- Lanzar una instancia de Amazon EC2.
- Utilizar los principios de AWS Well-Architected Framework al diseñar una capa de cómputo con Amazon EC2.

Opciones de cómputo en tiempo de ejecución de AWS

Diferentes servicios de cómputo para varios casos de uso

Máquinas virtuales (VMs)	Contenedores	Servidores Virtuales Privados(VPS)	Platform as a Service (PaaS)	Serverless
 Amazon Elastic Compute Cloud (Amazon EC2)	 Amazon Elastic Container Service (Amazon ECS) 	 Amazon Lightsail	 AWS Elastic Beanstalk	 AWS Lambda
	  AWS Fargate			

Servicios de cómputo

Cada categoría de servicio ofrece distintos niveles de control de infraestructura y velocidad de despliegue de aplicaciones

VMs

Contenedores

VPS

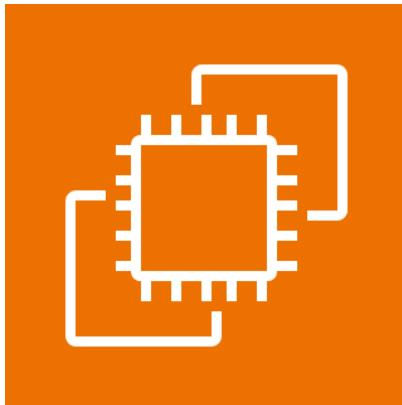
PaaS

Serverless

Mayor control y personalización de la infraestructura

Despliegues de aplicaciones más rápidos

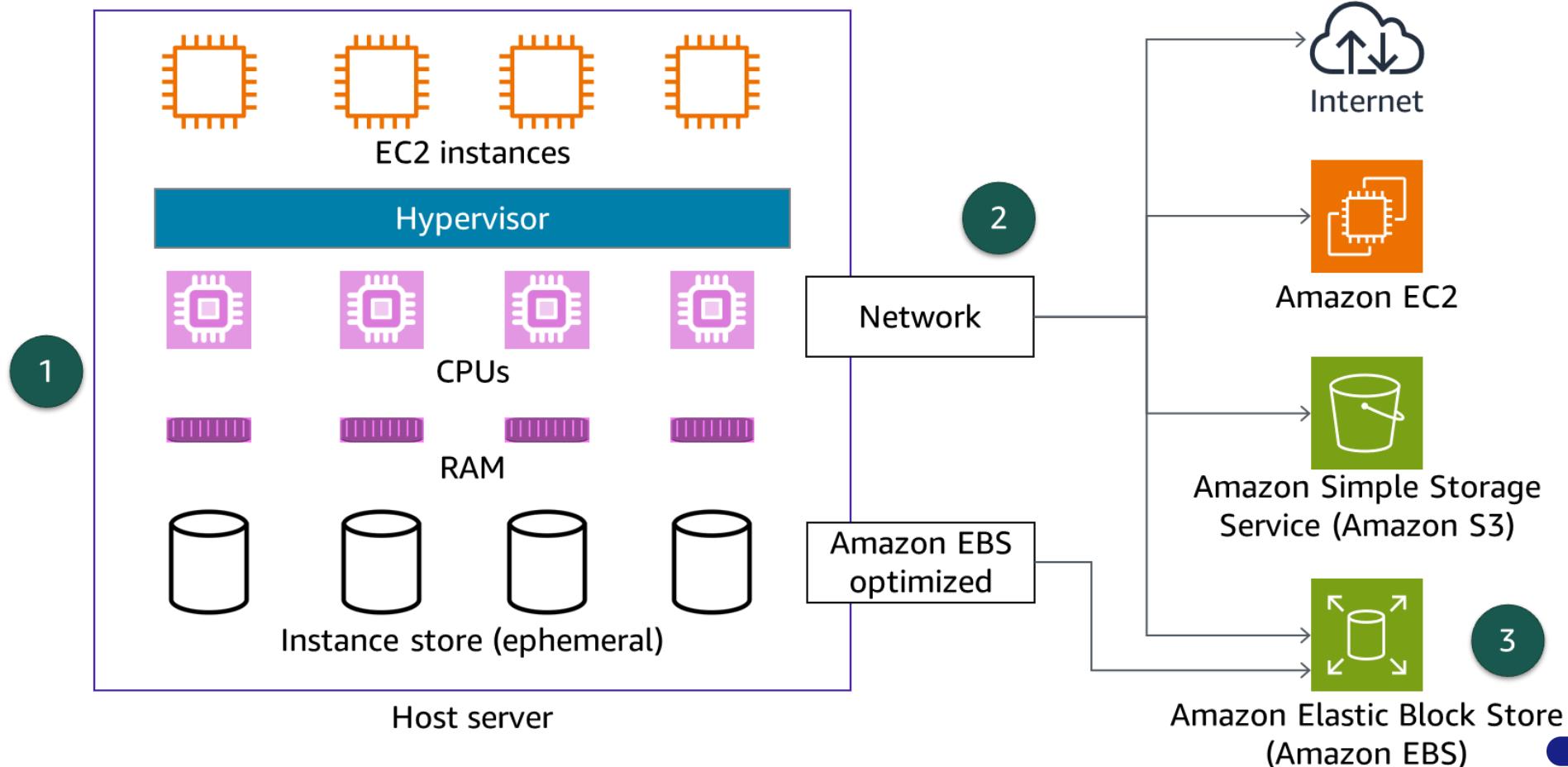
Amazon EC2



- Proporciona máquinas virtuales (servidores) en la nube.
- Aprovisiona servidores en minutos.
- Puede aumentar o reducir automáticamente la capacidad según sea necesario.
- Permite pagar sólo por la capacidad que se utiliza.

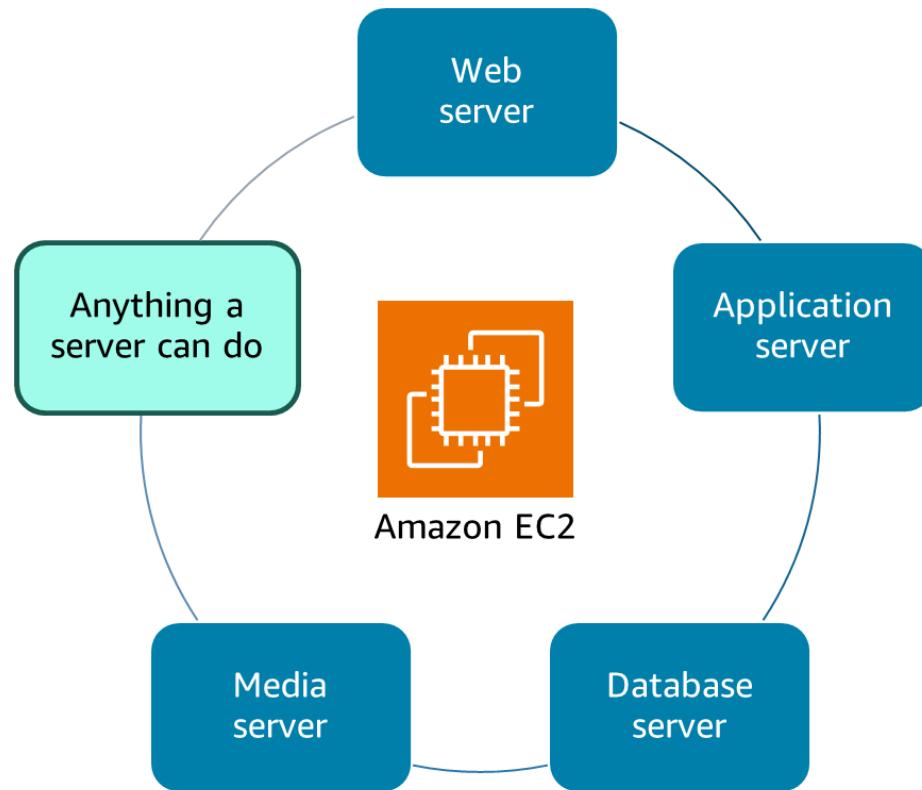
Amazon EC2

Una instancia EC2 es una VM que corre en un servidor físico



Amazon EC2

Casos de uso

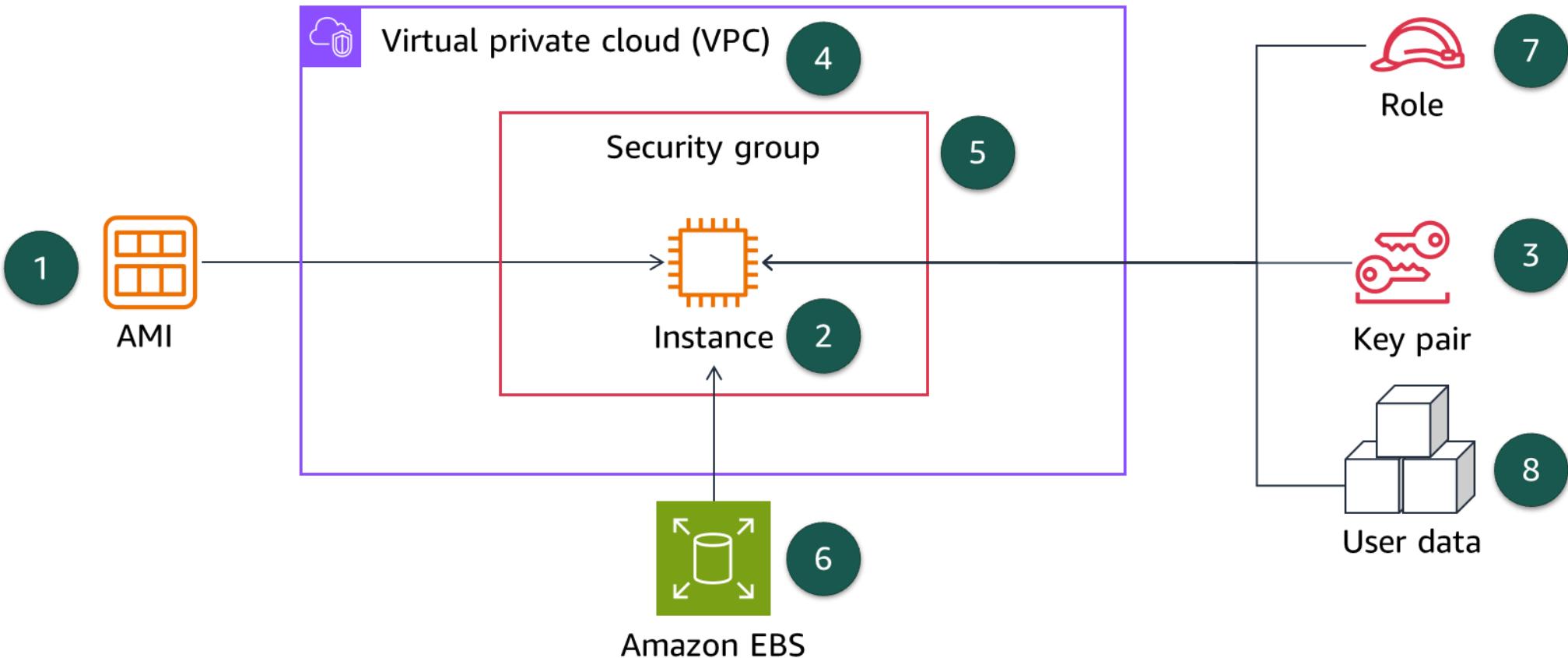


Siempre que se necesite:

- Control completo de sus recursos de cómputo, incluyendo sistema operativo y tipo de procesador.
 - Opciones para optimizar sus costos de cómputo
 - Instancias On-Demand , reservadas y Spot
 - Savings Plans
 - Capacidad para correr cualquier tipo de carga de trabajo
 - Sitios Web simples
 - Aplicaciones empresariales
 - Aplicaciones de inteligencia artificial generativa

Amazon EC2

Pasos para aprovisionar una instancia

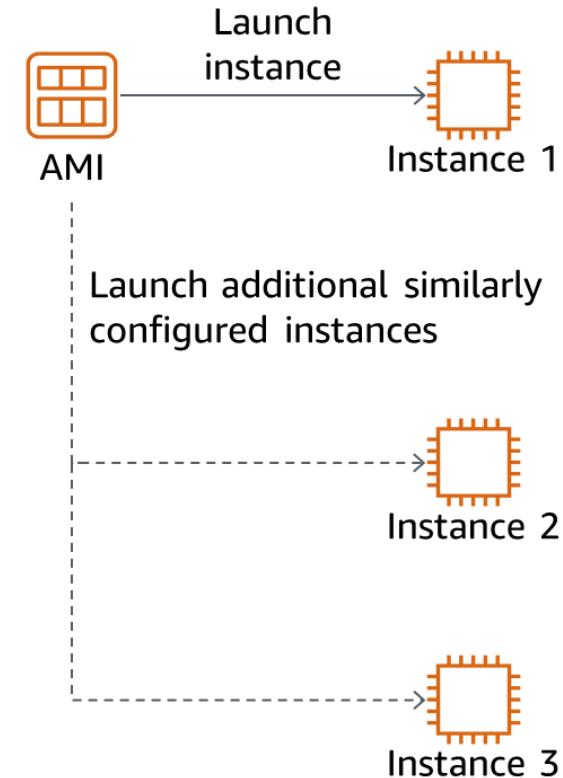


Amazon Machine Image (AMI)

Una AMI provee la información necesaria para lanzar una instancia, incluyendo:

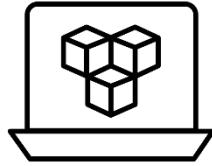
- Un template para el volumen raíz: contiene el sistema operativo anfitrión y opcionalmente otro software instalado
- Permisos: Controla quién puede acceder la AMI
- Mapeos a dispositivos de bloque: Especifica los volúmenes de almacenamiento conectados a la instancia

Create multiple instances from the same AMI



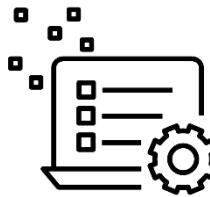
Amazon Machine Image (AMI)

Beneficios



Repetibilidad

Una AMI se puede usar en forma repetida para lanzar instancias con eficiencia y precisión.



Reusabilidad

Las instancias lanzadas con la misma AMI están configuradas de forma idéntica.



Recuperabilidad

Se puede crear una AMI a partir de una instancia configurada como un backup que se puede restaurar.

Se puede reemplazar una instancia que falló lanzando una nueva desde la misma AMI.

Amazon Machine Image (AMI)

Selección



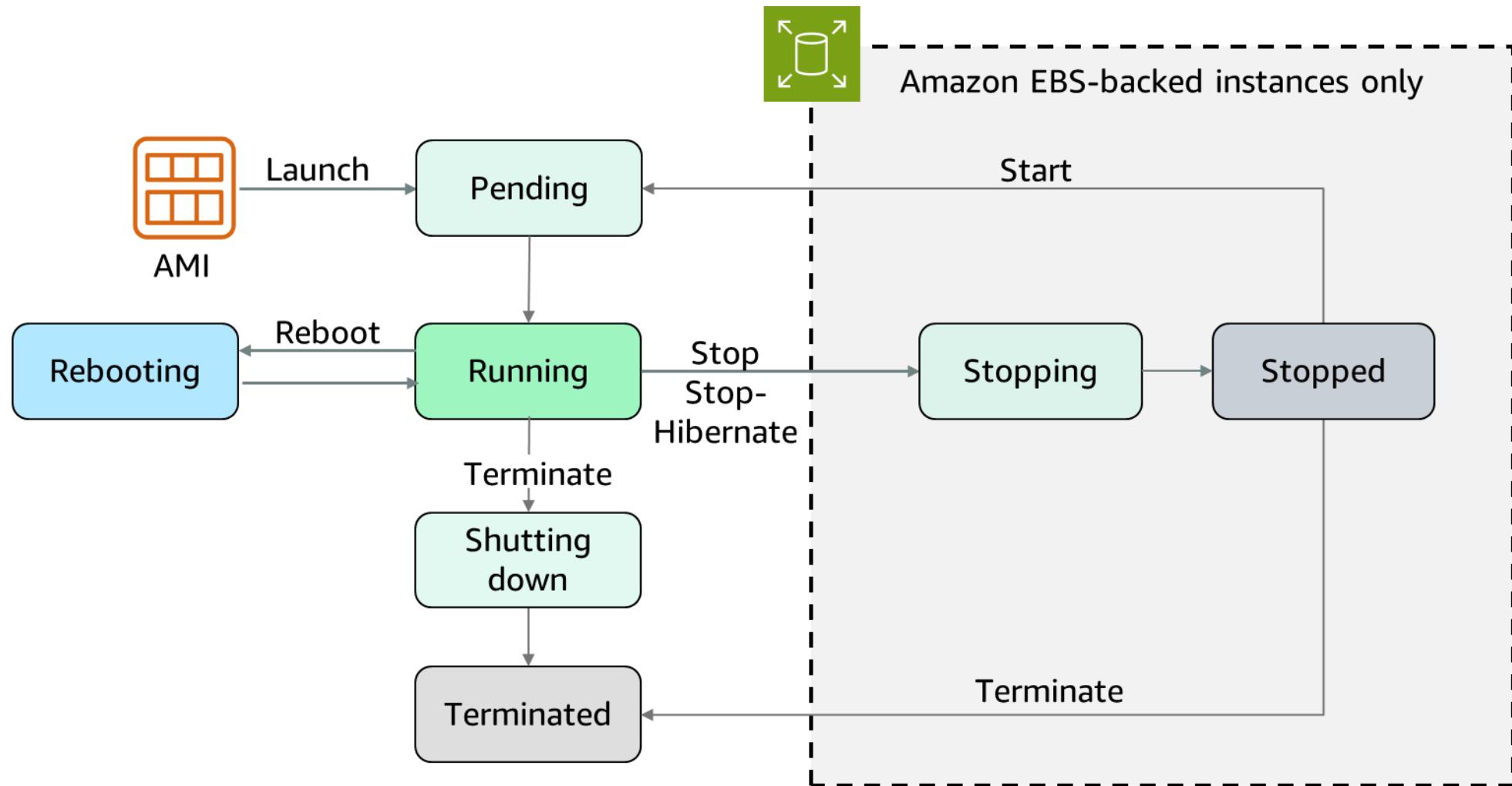
Criterios	AMI source
Región	Quick Start: AMIs con Linux y Microsoft Windows provistas por AWS.
Sistema Operativo	My AMIs: AMIs creadas por el usuario.
Tipo de almacenamiento del dispositivo raíz	AWS Marketplace: templates pre-configurados por terceros.
Arquitectura	Community AMIs: AMIs compartidas por otros. Usar bajo su propio riesgo.
Tipo de virtualización: para mejor desempeño, use AMIs con tipo de virtualización Hardware Virtual Machine (HVM)	

Instancias store-backed vs Amazon EBS-backed AMI

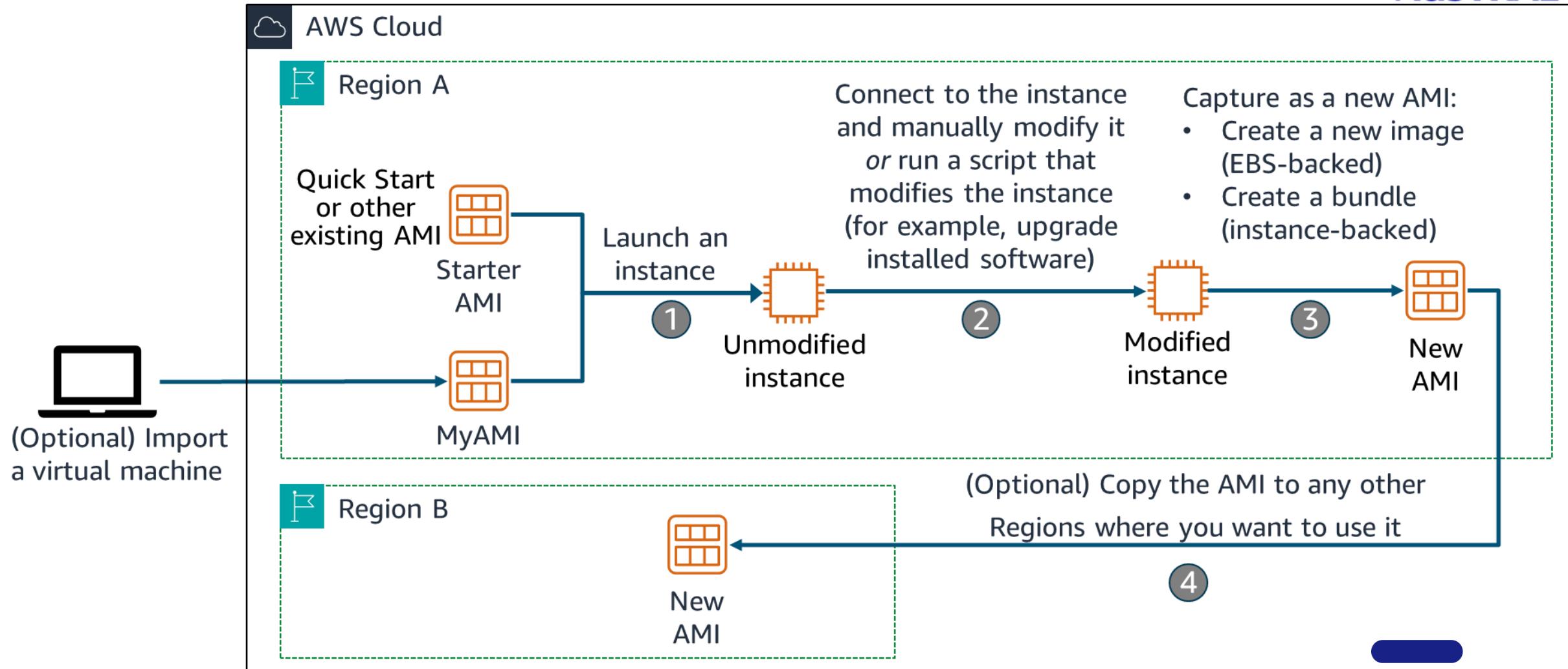
Characteristic	Amazon EBS-Backed Instance	Instance Store-Backed Instance
Boot time for the instance	Boots faster	Takes longer to boot
Maximum size of root device	16 TiB	10 GiB
Ability to stop the instance	Can stop the instance	Cannot be in a stopped state; instances are running or terminated
Ability to change the instance type	Can change the instance type by stopping instance	Can't change the instance type because the instance can't be stopped
Instance charges	You are charged for instance usage, EBS volume usage, and storing your AMI as an EBS snapshot	You are charged for instance usage and storing your AMI in Amazon S3
Use case	Persistent storage	Temporary storage

Amazon EC2

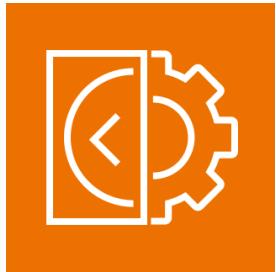
Ciclo de vida de una instancia



Creación de una AMI



EC2 Image Builder



EC2 Image Builder automatiza la creación, gestión e implementación de imágenes de VMs.

- Brinda una interfaz gráfica para crear *pipelines* de creación de imágenes
- Crea y mantiene AMIs de Amazon EC2 e imágenes de VM on-prem.
- Produce imágenes seguras, validadas y actualizadas
- Asegura el control de versiones

Amazon EC2

Configuración del tipo de instancia

El tipo de instancia de EC2 define la configuración de CPU, memoria, almacenamiento y performance de red.

Instance type	vCPU	Memory	Storage	Network performance
m5d.large	2	4 GiB	1 x 50 NVMe SSD	Up to 10 Gbps
m5d.xlarge	4	8 GiB	1 x 100 NVMe SSD	Up to 10 Gbps
m5d.8xlarge	32	128 GiB	2 x 600 NVMe SSD	10 Gbps

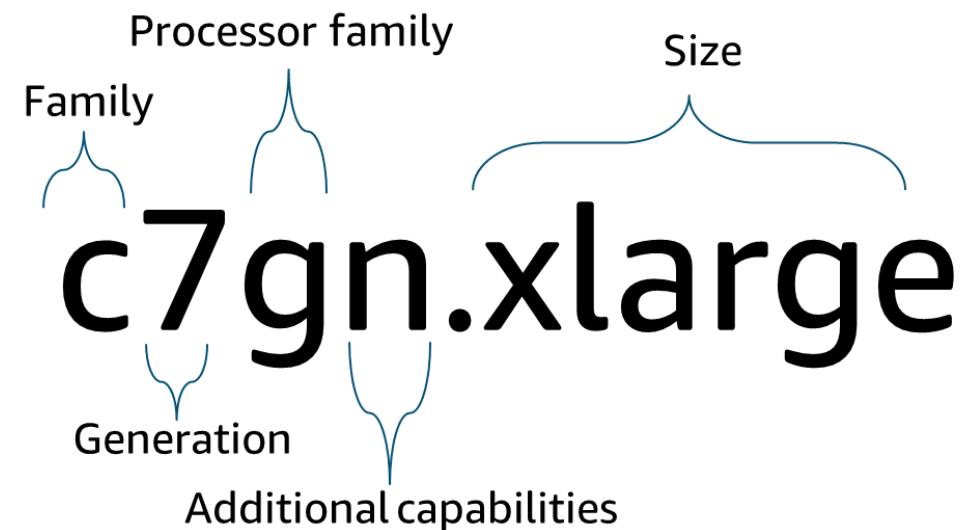
Amazon EC2

Nombres de los tipos de instancias

Componentes de los nombres

- Familia
- Generación
- Familia de procesador
- Capacidades adicionales
- Tamaño

Instance Type Naming:



Amazon EC2

Tipos de instancias según carga de trabajo

Tipo	Ejemplos de cargas	Ejemplos de tipos de instancia
Instancias de propósito general	Servidores de aplicación o Web Aplicaciones empresariales Servidores para juegos Ambientes para desarrollo o prueba	M7, Mac, M6, M5, M4, T4, T3, T2
Instancias optimizadas para cómputo	Procesamiento por lotes Aplicaciones analíticas distribuidas High performance computing (HPC)	C7, C6, C5, C4
Instancias optimizadas para almacenamiento	Bases de datos de alto rendimiento Analítica en tiempo real Cargas transaccionales	I4, Im4, Is4, I3, D2, D3, H1
Memory optimized instance types	In-memory caches Bases de datos de alto rendimiento Aplicaciones analíticas "Big Data"	R7, R6, R5, R4, X2, X1, Z1
Instancias con aceleración de cómputo	Machine learning, inteligencia artificial (IA) HPC	P5, P4, P3, P2, DL1, Trn1, Inf2, Inf1, G5, G4, G3, F1, VT1
Instancias optimizadas para High performance computing (HPC)	Cargas de aprendizaje profundo Cargas de HPC intensivas en cómputo	Hpc7, Hpc6

Amazon EC2

Selección de la instancia más adecuada

Con 270 tipos de instancia disponibles, ¿cómo seleccionar el adecuado?

- Considere los requerimientos tanto de desempeño como de costo.
- Use los recursos disponibles para obtener recomendaciones.

Tarea	Solución
Creación de una nueva instancia	<p>En la consola de EC2, use la página de Tipos de Instancia para filtrar según las características seleccionadas.</p> <p>Recomendación: La última generación de cada familia generalmente tiene la mejor relación costo/rendimiento.</p>
Optimización de una instancia existente	<p>Puede obtener recomendaciones para optimizar el tipo de instancia mediante AWS Compute Optimizer.</p> <p>Puede evaluar recomendaciones y modificar la instancia de acuerdo a ellas.</p>

AWS Compute Optimizer

AWS Compute Optimizer > Dashboard > Recommendations for EC2 instances

Recommendations for EC2 instances (8) Info
Recommendations for modifying current resources for better cost and performance.

Action ▾ View detail

Filter by one or more Regions: 090765505187 | Over-provisioned | < 1 > | ⌂

Region: US East (N. Virginia) X | Clear filters

Instance ID	Instance name	Finding	Current instance type	Current On-Demand price	Recommended instance type
i-0218a45abd8b53658	-	Over-provisioned	m5.xlarge	\$0.192 per hour	r5.large
i-069f6e837890db127	-	Over-provisioned	c5.xlarge	\$0.17 per hour	t3.large
i-07084b94d1bcf391b	-	Over-provisioned	c5.xlarge	\$0.17 per hour	t3.large
i-0af9322ff627d7e8f	-	Over-provisioned	m5.xlarge	\$0.192 per hour	r5.large
i-0ceb95ed248026d24	-	Over-provisioned	m5.xlarge	\$0.192 per hour	r5.large
i-0f277818dfef522e9	-	Over-provisioned	c5.xlarge	\$0.17 per hour	t3.large
i-0f4f4c06ad8afe81a	-	Over-provisioned	m5.2xlarge	\$0.384 per hour	r5.xlarge
i-0fb9323080785de1e	-	Over-provisioned	c5.xlarge	\$0.17 per hour	t3.large

Recomienda el tipo y tamaño de instancia, y configuración de grupo de Auto Scaling óptimos

Analiza patrones de carga de trabajo y hace recomendaciones.

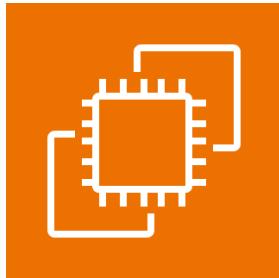
Clasifica los hallazgos en las instancias como aprovisionamiento insuficiente, excesivo, optimizado o Ninguno

Actividad

Selección de tipos de instancia

Actividad

Consigna



Elegir el tipo de instancia EC2 adecuado para cada uno de los casos de uso de ejemplo.

Usar la información en
<https://aws.amazon.com/ec2/instance-types/>

Actividad

Casos

Carga de trabajo
Bases de datos transaccionales
Entornos de desarrollo pequeños
Servidores de juegos
<i>In-memory caches</i>
Generación de imágenes y video
Machine learning
Procesamiento batch

Familias de instancias

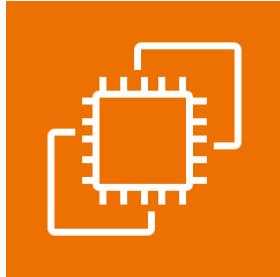
C	M	I	P
Inf2	R	T	

Actividad

Respuestas

Carga de trabajo	Familia de instancias
Bases de datos transaccionales	I
Entornos de desarrollo pequeños	T
Servidores de juegos	M
In-memory caches	R
Generación de imágenes y video	Inf2
Machine learning	P
Procesamiento batch	C

Puntos clave



Un tipo de instancia EC2 define una configuración de características de rendimiento de CPU, memoria, almacenamiento y red.

Como recomendación, elija tipos de instancias de nueva generación en una familia porque generalmente tienen mejores relaciones precio-rendimiento.

Utilice la página Tipos de instancia en la consola de Amazon EC2 y AWS Compute Optimizer para encontrar el tipo de instancia adecuado para su carga de trabajo.

Almacenamiento en instancias EC2

Recurso de almacenamiento de EC2	Volumen raíz	Volúmenes de datos para una sola instancia	Volúmenes de datos que se pueden acceder desde varias instancias de Linux	Volúmenes de datos que se pueden acceder desde varias instancias de Windows
Amazon EBS (Sólo SSD)	Sí	Sí	No	No
Instance store	Sí	Sí	No	No
Amazon Elastic File System (Amazon EFS) [Linux]	No	No	Sí	No
Amazon FSx for Windows File Server	No	No	No	Sí

Una instancia EC2 siempre va a tener un volumen raíz, y opcionalmente uno o más volúmenes de datos.

Instance stores

Un *instance store* proporciona almacenamiento no persistente a una instancia. El volumen de datos se almacena en el mismo servidor físico donde se ejecuta la instancia.

Características

Almacenamiento temporal orientado a bloque

Usa HDD o SSD

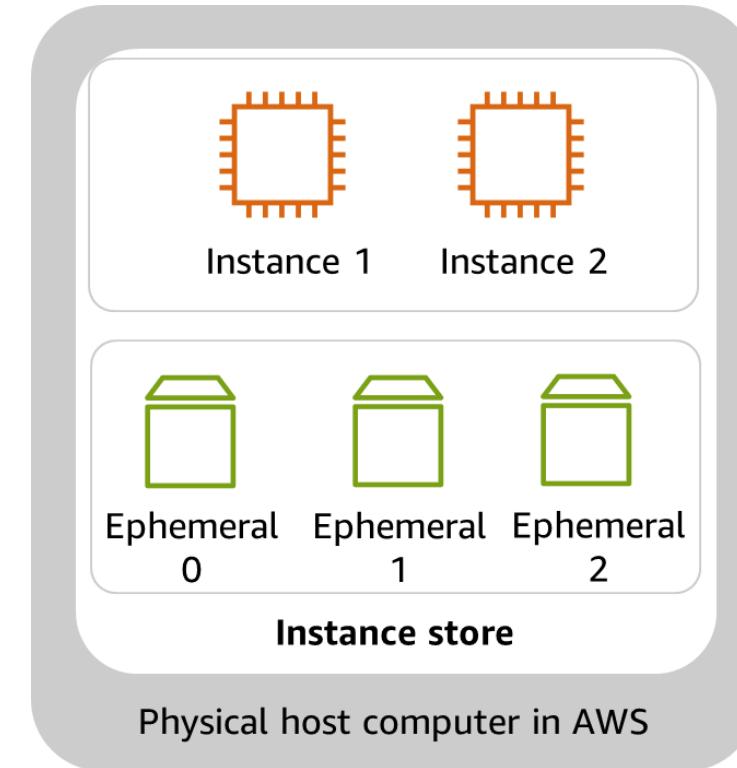
Los datos del *instance store* se pierden cuando la instancia es detenida o terminada.

Ejemplos de caso de uso

Buffers

Cache

Datos temporarios



Amazon EBS

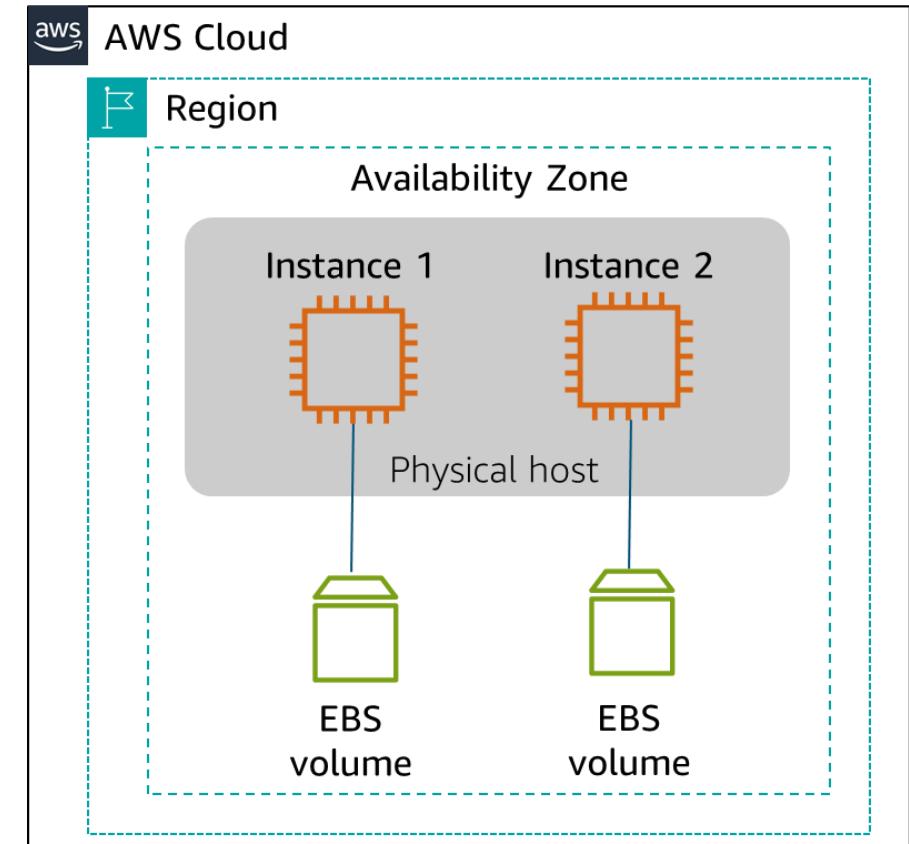
Los volúmenes Amazon EBS proveen almacenamiento persistente conectado por red a instancias EC2.

Características

- Almacenamiento persistente en bloques
- Puede conectarse a cualquier instancia en la misma Zona de Disponibilidad
- Usa HDD o SSD
- Puede estar cifrado
- Soporta snapshots que son persistidos en S3
- Los datos persisten independientemente de la vida de la instancia.

Ejemplos de caso de uso

- Bases de datos stand-alone
- Almacenamiento de datos de aplicación



Amazon EBS

Volúmenes respaldados por HDD

Los volúmenes respaldados por HDD de Amazon EBS funcionan bien cuando el foco está en el rendimiento (*throughput*).

Tipo de volumen	Descripción	Casos de uso
Throughput Optimized HDD (st1)	<ul style="list-style-type: none"> Tipo de volumen de bajo costo Diseñado para cargas de trabajo de alto rendimiento y acceso frecuente 	<ul style="list-style-type: none"> Cargas de <i>streaming</i> Big data Data warehouses Procesamiento de logs No puede ser un volumen de boot
Cold HDD (sc1)	<ul style="list-style-type: none"> Volumen HDD de menor costo Diseñado para cargas de trabajo con acceso menos frecuente 	<ul style="list-style-type: none"> Almacenamiento orientado al <i>throughput</i> para grandes volúmenes de datos de acceso poco frecuente Casos de uso donde el menor costo de almacenamiento es importante No puede ser un volumen de boot

Instancias optimizadas para EBS

Algunos tipos de instancia EC2 pueden ser *EBS-optimized*

Beneficios

Proporciona una conexión de red dedicada a los volúmenes EBS conectados.

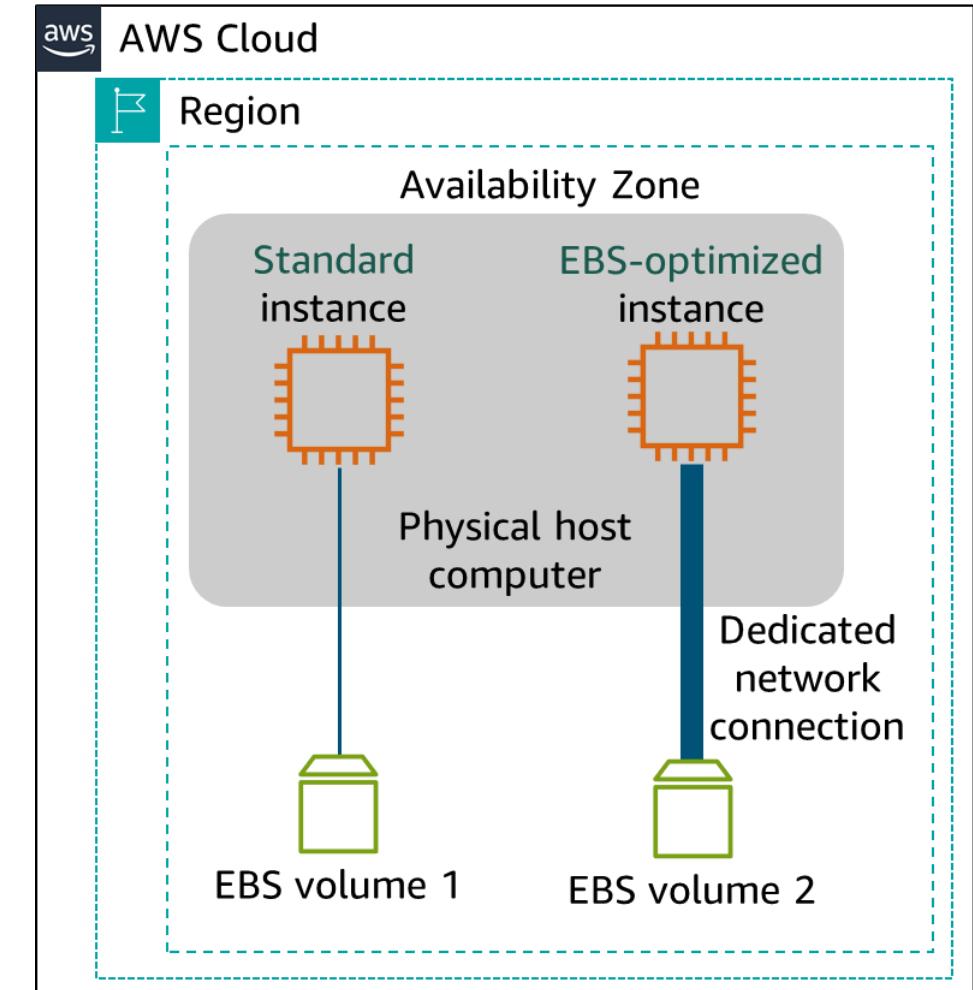
Aumenta el rendimiento de E/S.

Se consigue un rendimiento adicional si se utiliza una instancia basada en el sistema Amazon EC2 Nitro.

Uso

Para los tipos de instancias EBS-optimized, la optimización está habilitada de forma predeterminada

Para otros tipos de instancias compatibles, la optimización debe habilitarse manualmente



Instancias EC2

Sistemas de archivos compartidos

¿Qué sucede si tenemos varias instancias que deben usar el mismo almacenamiento?



No es una opción



Amazon EBS

Se conecta solo a una instancia



No es la opción ideal



Amazon S3

Amazon S3: Es una opción, pero no es ideal



Mejor opción



Amazon EFS
(Linux)



Amazon FSx for
Windows File
Server (Windows)

Amazon EFS and Amazon FSx for Windows File Server: ambos satisfacen el requerimiento

Amazon Elastic File System

Amazon EFS



Proporciona almacenamiento de sistema de archivos para cargas de trabajo basadas en Linux.

Sistema de archivos elástico totalmente gestionado.

Se escala automáticamente a medida que se añaden o eliminan archivos.

Petabytes de capacidad.

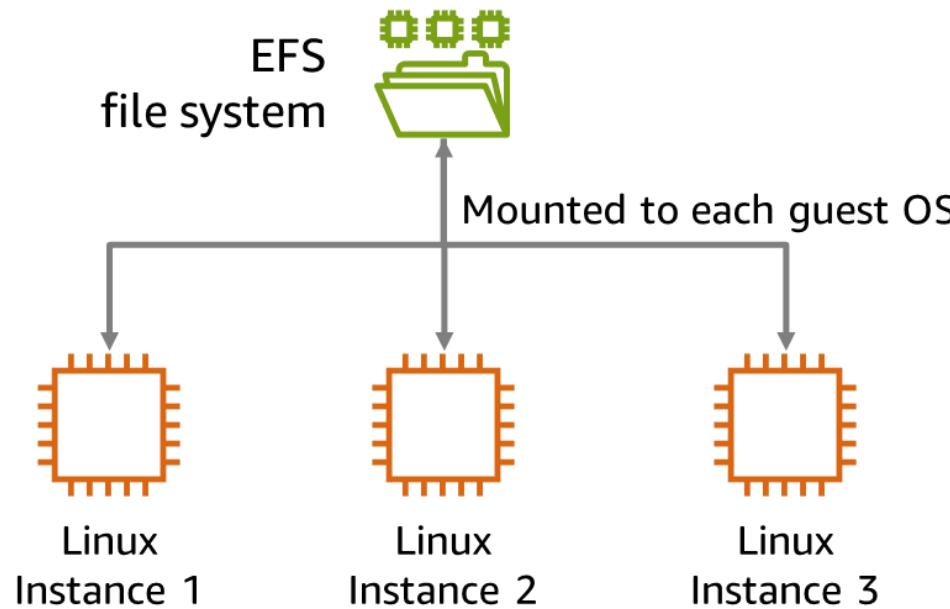
Soporta los protocolos de Network File Systems (NFS).

Monta el sistema de archivos en la instancia EC2.

Compatible con todas las AMI basadas en Linux para Amazon EC2.

Amazon EFS

Casos de uso



Ejemplo de comando para montar el *file system* en cada uno de las instancias

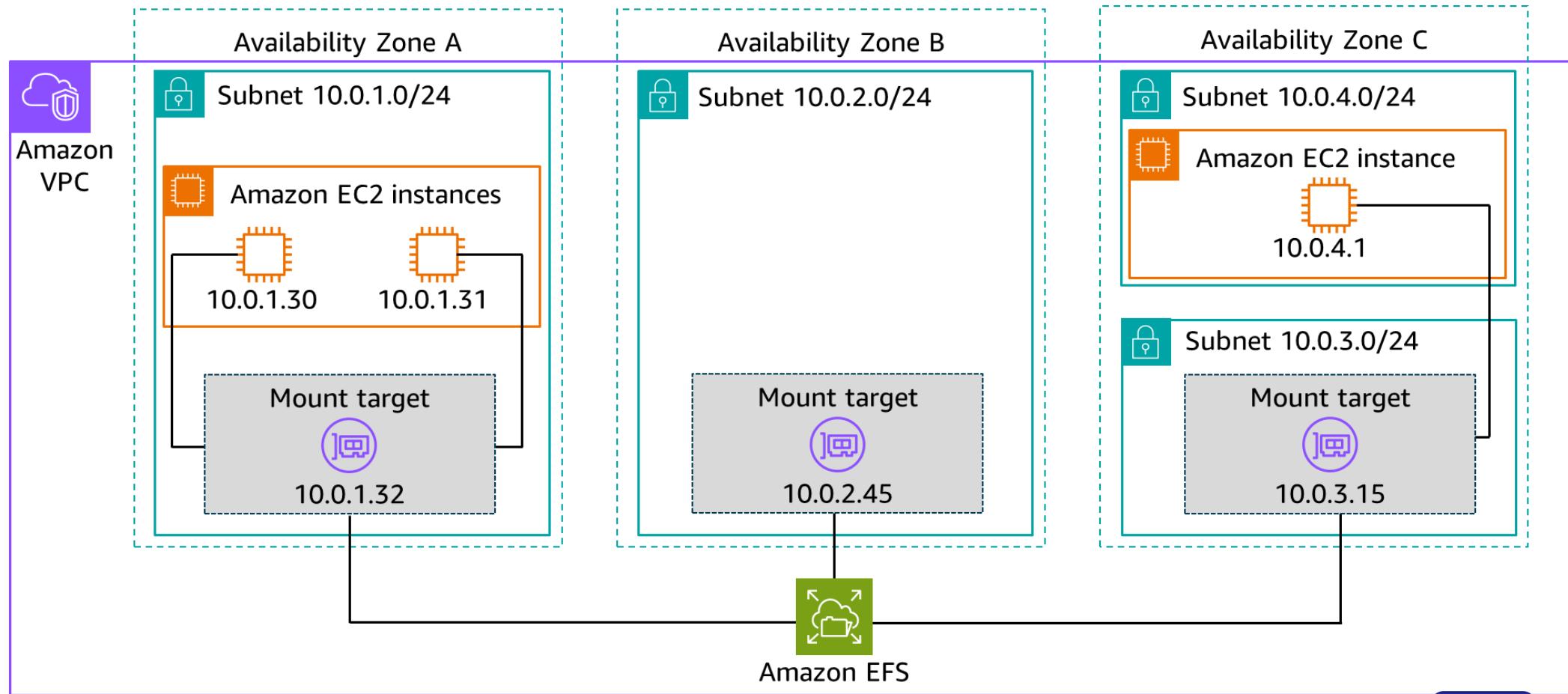
```
$ sudo mount -t nfs4 mount-target-DNS:/ ~efs-mount-point
```

Algunos ejemplos de aplicaciones y cargas de trabajo comunes para Amazon EFS incluyen:

- Home directories
- Sistema de archivos para aplicaciones empresariales
- Pruebas y desarrollo de aplicaciones
- Copias de seguridad de bases de datos
- Servidores web y gestión de contenido
- Flujos de trabajo multimedia
- Análiticos de *big data*

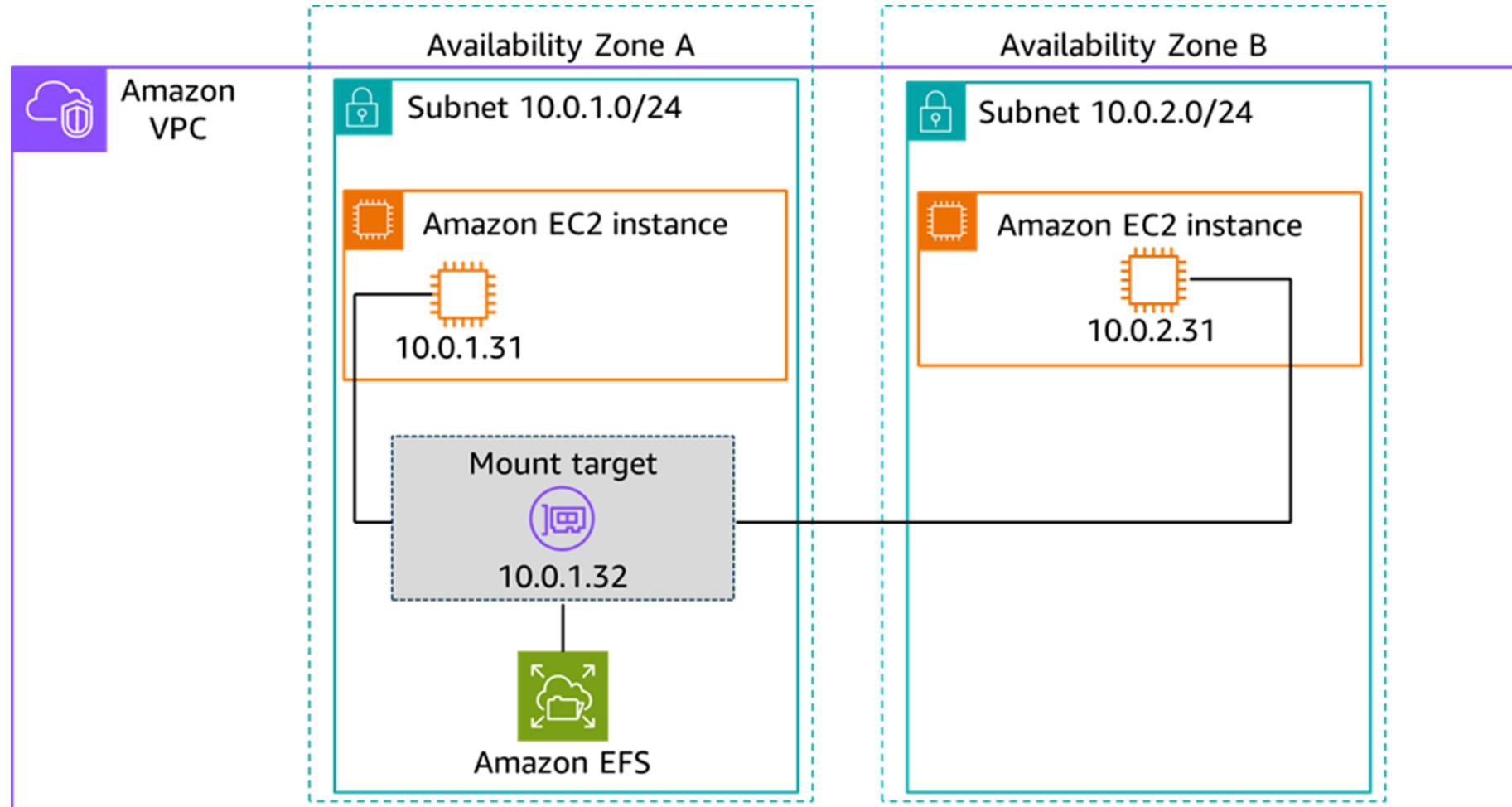
Amazon EFS

Standard class storage



Amazon EFS

One-zone class storage



Amazon FSx

For Windows file server



Proporciona almacenamiento de sistema de archivos compartido totalmente administrado para instancias EC2 de Microsoft Windows.

Compatibilidad nativa con Microsoft Windows.

Sistema de archivos NTFS

Utiliza el protocolo nativo Server Message Block (SMB) versión 2.0 a 3.1.1.

Espacios de nombres y replicación DFS de Distributed File System (DFS).

Se integra con Microsoft Active Directory y admite listas de control de acceso (ACL) de Windows.

Respaldado por almacenamiento SSD de alto rendimiento.

Amazon FSx

Casos de uso



Home directories

Cargas de trabajo de *lift-and-shift* de aplicaciones

Workflows de medios y entretenimiento

Análisis de datos (*Data analytics*)

Servicios web y gestión de contenido

Entornos de desarrollo de software

Almacenamiento para instancias EC2

Resumen

Las opciones de almacenamiento para instancias EC2 incluyen el *instance store*, Amazon EBS, Amazon EFS y Amazon FSx para Windows File Server.

Para un volumen raíz, utilice Amazon EBS con respaldo SSD.

Para un volumen de datos que solo atienda una instancia, utilice el *instance store* o el almacenamiento de Amazon EBS.

Para un volumen de datos que atienda varias instancias de Linux, utilice Amazon EFS.

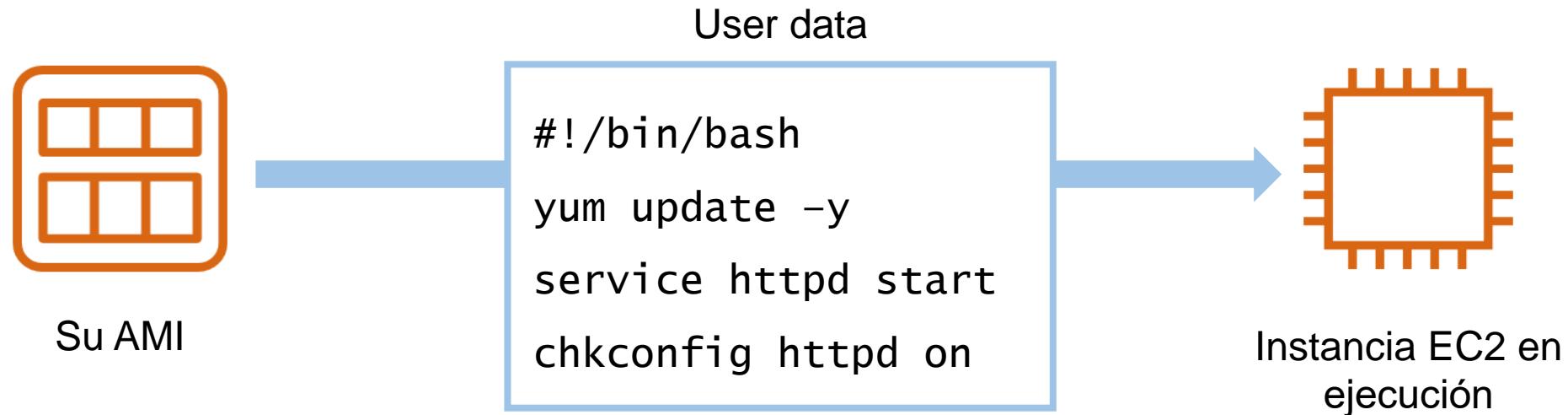
Para un volumen de datos que atienda varias instancias de Microsoft Windows, utilice Amazon FSx para Windows File Server.

Configuración de EC2

Otras consideraciones

EC2 instance user data

Cuando se inicia una instancia EC2, se puede especificar datos de usuario para ejecutar un script de inicialización (shell script o directiva cloud-init).



Recuperación de metadatos de la instancia

Los metadatos de una instancia son información sobre ella.

Se puede acceder a ellos desde su instancia en esta URL:
<http://169.254.169.254/latest/meta-data/>

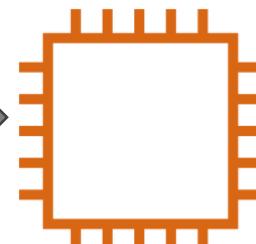
Se pueden recuperar desde un script de datos de usuario.



Su AMI

```
#!/bin/bash
yum update -y
hostname = $(curl -s http://169.254.169.254/latest/meta-data/public-hostname)
```

User data



Instancia
EC2 en
ejecución

Metadato	Valor
instance-id	i-1234567890abcdef0
mac	00-1B-63-84-45-E6
public-hostname	ec2-203-0-113-25.compute-1.amazonaws.com
public-ipv4	67.202.51.223
local-ipv4	10.251.50.12

Datos de usuario en instancias en ejecución

Paso 1

Detenga la instancia.



Current user data:

Example user data script:

```
#!/bin/bash
yum update -y
```

Paso 2

Modificar el script de datos de usuario.



Modified user data:

Example user data script:

```
#!/bin/bash
yum update -y
yum install httpd start
service httpd start
chkconfig httpd on
```

Paso 3

Eliminar el archivo config_scripts_user.



Command to remove config_scripts_user file:

Example code to remove config_scripts_user file:

```
sudo rm /var/lib/
cloud/ Instances/*
/sems/config_scripts
_user
```

Paso 4

Volver a iniciar la instancia o ejecutar un comando para correr el script.



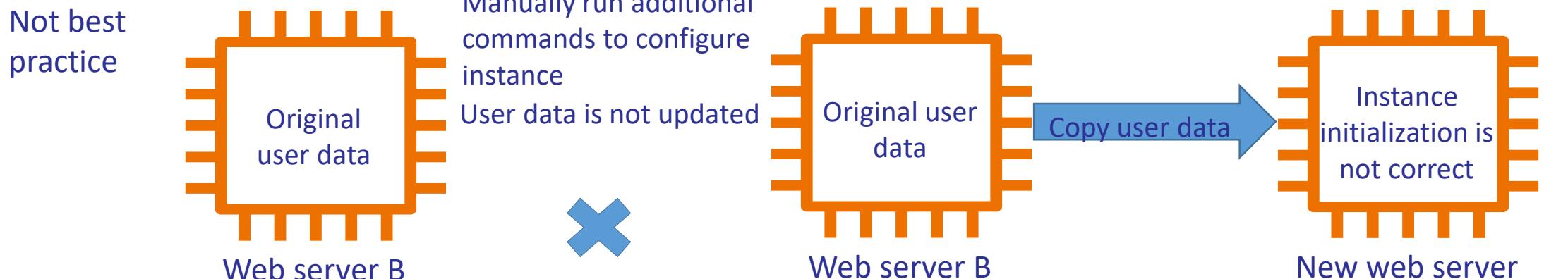
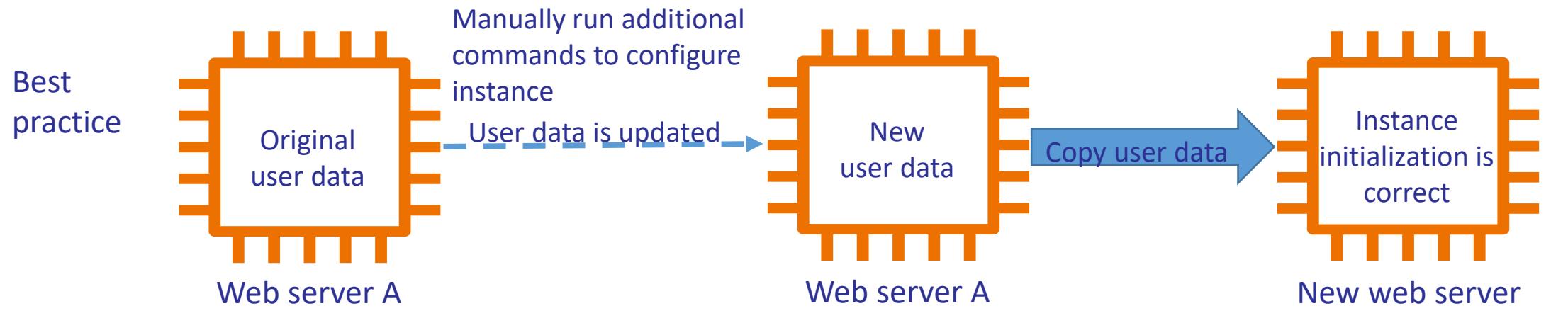
Modified script runs:

Example command to run user data without restarting the instance:

```
/var/lib/cloud/
instance/scripts/
part-001
```

Ejecución manual de comandos

Buenas prácticas



Modelos de implementación de AMI



AMI básica

AMI base

AMI configuradas solo con SO

Totalmente configurables y actualizables

Tiempo de generación más corto

Arranque más lento



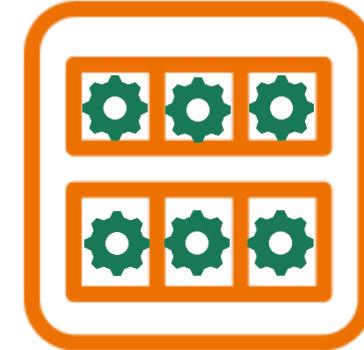
Silver AMI

AWS Managed Services (AMS) provided mutable AMIs

Configurations half baked into the AMI

Some configurations need to be done manually or by user data scripts

Provides a balance between boot speed and build time



Golden AMI

Customized immutable AMIs

Configurations fully baked into the AMI

All instances using the same golden AMI behave the same

Shorter boot times but increases build times

Shorter lifespan of the AMI

Modelos de implementación de AMI

Los *Placement groups* brindan control sobre dónde se ejecuta un grupo de instancias interdependientes en una zona de disponibilidad.



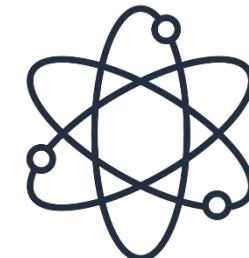
Beneficios

Aumenta el rendimiento de la red entre instancias.
Reduce los fallos correlacionados o simultáneos.



Limitaciones

Una instancia solo se puede lanzar en un *placement group* a la vez
Las instancias con un *tenancy* de host no se pueden lanzar en un *placement group*



Estrategias

Cluster
Partition
Spread

EC2

Opciones de precios

AWS Free Tier: Amazon EC2



12 meses gratis

750 horas por mes de instancias t4g.small dependiendo de la región

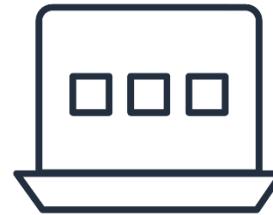
750 horas por mes de instancias Linux, RHEL, o SLES t2.micro o t3.micro dependiendo de la región

750 horas por mes de instancias Windows t2.micro o t3.micro dependiendo de la región

Amazon EC2

Modelos de precios

Amazon EC2 ofrece las siguientes estrategias de compra para ayudar a optimizar sus costos según sus necesidades:



Modelos de compra

El énfasis está en proporcionar grandes ahorros a través de diferentes casos de uso.



Modelos de Capacidad reservada

El énfasis está en proporcionar instancias reservadas para garantizar que estén disponibles cuando se necesiten.



Modelos Dedicados

El énfasis está en proporcionar hardware dedicado que ayude a cumplir con requisitos de cumplimiento y regulatorios.

Amazon EC2

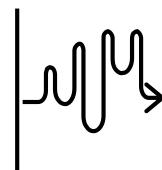
Modelos de compra

On-Demand

Se paga la capacidad de cómputo por segundo o por hora sin compromisos a largo plazo.

Recomendado para:

- Cargas de trabajo con picos
- Cargas de trabajo de experimentación

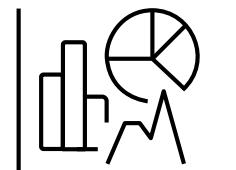


Reservadas

Se compromete uso por 1 o 3 años a cambio de un descuento significativo sobre los precios on-demand.

Recomendado para:

- Cargas de trabajo comprometidas
- Cargas de trabajo estables



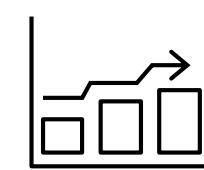
Savings Plans

Mismos descuentos que para instancias reservadas.

Mayor flexibilidad a cambio de un compromiso de \$/hora

Recomendado para:

- Todas las cargas de trabajo de Amazon EC2
- Cargas de trabajo de Amazon EC2 que requieran flexibilidad en el uso comprometido

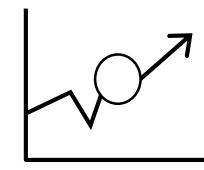


Amazon EC2 Spot

Capacidad de Amazon EC2 ociosa con ahorros substanciales sobre los precios On-Demand.

Recomendado para:

- Cargas tolerantes a fallas
- Cargas flexibles
- Cargas sin estado



Amazon EC2

Reservas de capacidad

Las reservas de capacidad le permiten reservar capacidad de cómputo para instancias de Amazon EC2 en una zona de disponibilidad específica.

On-Demand Capacity Reservations

Esto garantiza que siempre tendrá acceso a la capacidad de EC2 cuando la necesite, durante el tiempo que la necesite.

Casos de uso recomendados:

- Cargas de trabajo que deben cumplir con los requisitos normativos de alta disponibilidad
- Cargas de trabajo que requieren garantía de capacidad

Amazon EC2 Capacity Blocks for ML

Se reservan instancias de GPU para una fecha futura para ejecutar cargas de trabajo de aprendizaje automático (ML).

Casos de uso recomendados:

- Entrenamiento y ajuste de modelos de ML
- Ejecución de experimentos y creación de prototipos
- Planificación para futuros aumentos repentinos de la demanda de aplicaciones de ML

Amazon EC2

Opciones dedicadas

Las opciones dedicadas de Amazon EC2 brindan capacidad de instancia EC2 en servidores físicos dedicados para su uso (*hardware single-tenant*).

Instancias dedicadas

Facturación por instancia
Ubicación automática de instancias
Aísla los hosts donde se ejecutan sus instancias

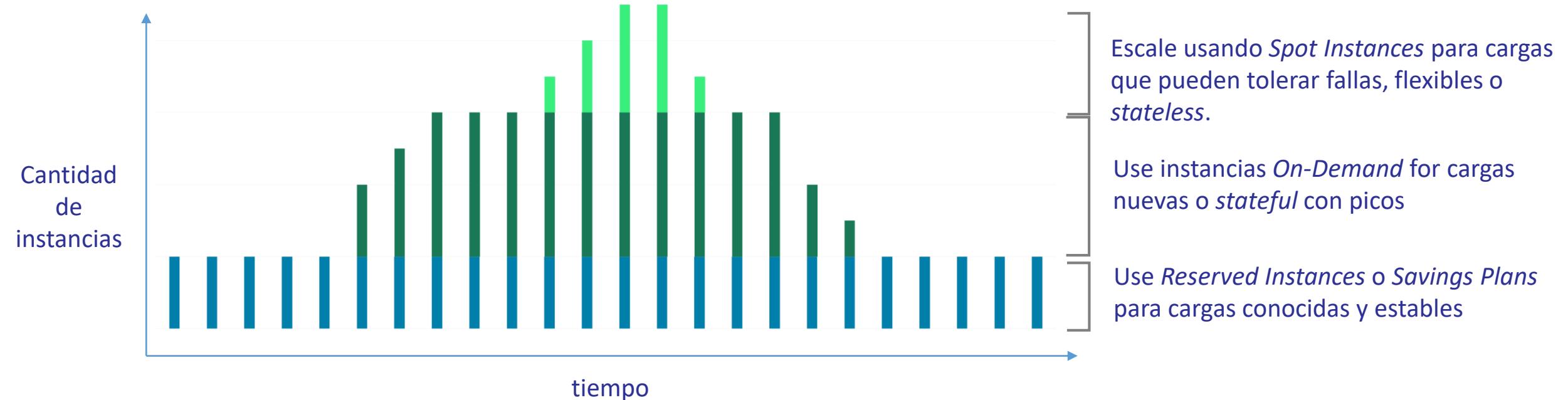
Hosts dedicados

Facturación por host
Visibilidad de sockets, núcleos e ID de host
Afinidad entre un host y una instancia
Ubicación de instancias específica
Aumento de capacidad mediante una solicitud de asignación
Permite usar las licencias de software vinculadas al servidor y cumplir con requisitos de cumplimiento

Amazon EC2

Guía de optimización de costos

Para optimizar el costo de las instancias de Amazon EC2, debemos combinar las opciones de compra disponibles.

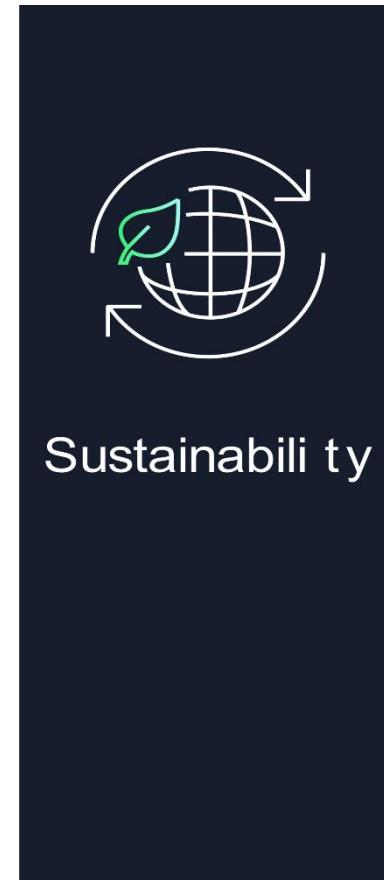


Well-Architected Framework

Aplicación en soluciones de cómputo

Well-Architected Framework

Pilares



Well-Architected Framework

Protección de la infraestructura – Protección del cómputo



Security

Automatice la protección de los recursos
de cómputo.

Well-Architected Framework

Protección de la infraestructura – Protección de redes



Security

Controle el tráfico en todas las capas.

Well-Architected Framework

Performance



Performance
efficiency

Escale las mejores opciones de cómputo para su carga de trabajo.

Configure y dimisione adecuadamente los recursos de cómputo.

Well-Architected Framework

Optimización de costos



Cost
Optimization

Seleccione recursos del tipo, tamaño y cantidad correctos.

Seleccione el mejor modelo de precios.

Well-Architected Framework

Hardware y servicios



Sustainability

Utilizar la mínima cantidad de hardware para satisfacer sus necesidades.

Usar tipos de instancia con el menor impacto.

Usar servicios gerenciados.

Muchas gracias.

www.austral.edu.ar