

Table of Contents

Build With EC2

- 3..... Basic EC2
- 4..... AMIs
- 5..... Shared Tenancy
- 6..... AWS Networking Fundamentals
- 7..... Managing Instance States
- 8..... Storage Options Amazon EC2
- 9..... Elastic Network Interfaces
- 10..... Resilience EC2
- 11..... Comparing Purchasing Options

Build With EC2

Basic EC2

- Más de 400 tipos de instancias
- Son VMs (Máquinas Virtuales)

Tipos de Instancias

- **Concepto:** La combinación de los componentes de hardware virtuales.
- Se agrupan en "INSTANCE FAMILIES" (Familias de instancias). Cada familia tiene optimización para casos de usos específicos.
- Cada familia tiene subfamilias, acorde a la combinación de procesador y almacenamiento.
- **Cálculo de vCPU (Virtual Central Processing Units):** Medida de procesamiento.
 - 2 CPUs y 2 hilos por core = 4 vCPUs.
 - 2 CPUs con dos hilos cada uno.

Arquitectura de Instancias

- **Intel:** c5, m5, r5
- **AMD:** c5a, m5a, t3a
- **ARM (Graviton):** c6g, m6g

Dependiendo de la carga de trabajo, seleccionar una u otra trae beneficios.

- **C** (Compute optimized): Ejemplo: c5.2xlarge – 16 GiB memory y 8vCPU (2:1) -> **Ratio 2:1:** Ratio de memoria a vCPU es de 2 GiB por vCPU.
- **M** (General purpose): Ejemplo: m5.2xlarge – 32 GiB memory y 8vCPU (4:1)
- **R** (Memory optimized): Ejemplo: r5.2xlarge – 64 GiB memory y 8vCPU (8:1)

Familias de Instancias

- **General Purpose (M y T):** Balance entre recursos. Para una amplia gama de tipos de trabajo (ej. Web Server, Repositorio de código).
- **Burstable Instance Options (T):** Cargas de trabajo que no están ocupadas todo el tiempo y no requieren demasiado rendimiento de la CPU. Usar instancias demasiado grandes y de alto rendimiento sería un desperdicio. Pueden usarse instancias de propósito general para estas cargas. Son económicas y tienen capacidad de ráfaga (cubrir picos de carga).



Compute optimized (C)

- Dedicadas al procesamiento con alto rendimiento, cargas de trabajo con mucho procesamiento.
 - Ejemplos:
 - Procesamiento batch: Procesamiento de grandes volúmenes de datos de una vez.
 - Servidores web con mucho tráfico.
 - Crear modelos científicos y simulaciones.
 - Servidores de juegos con alto rendimiento.
 - Inferencias ML para obtener resultados rápidamente.

Memory optimized (R)

- Dedicadas al procesamiento de datos en memoria de grandes volúmenes de datos. Rendimiento rápido para ejecutar operaciones en los sets de datos.
 - Ejemplos:
 - DB relacionales y NoSQL (Para grandes volúmenes, ambos casos).
 - DB en cache, Redis, Memcache, etc.
 - Procesamiento en tiempo real con Spark, Hadoop.
 - Procesos que requieran mucha memoria para cálculos complejos.

Storage optimized (I)

- Optimizadas para cargas de trabajo que requieran un alto rendimiento en lectura y escritura de almacenamiento local para operaciones con conjuntos de datos.
 - Ejemplos:
 - Procesamiento masivo paralelo.
 - Procesamiento distribuido con MapReduce o Hadoop.
 - DB NoSQL: Acceso rápido a grandes volúmenes de transacciones.
 - Cache en memoria: Ej. Redis.
 - DataWarehouses.

Accelerate compute (P y G)

- Utilizan aceleradores de hardware (co-procesadores) (Tarjetas gráficas).
 - Ejemplos:
 - Cálculos con comas flotantes.
 - Procesamiento de gráficos.
 - Coincidencia con patrones de datos.
 - Procesamiento paralelo en alto rendimiento.

Casos de uso

- **Renderizado de video**
- **Procesamiento de video**
- **Big Data:** Análisis de grandes volúmenes de datos
- **IA y Machine Learning:** Entrenamiento e inferencia para modelos que requieran alto rendimiento.

Decodificación del Nombre de la Instancia

Ejemplo: **m5zn.xlarge**

- **Familia (Family):** **m** - Indica el tipo de instancia. En este caso, **m** -> instancias de propósito general.
- **Generación (Gen):** **5** - Indica la generación de la instancia.
- **Opcional:** **z** - Alta frecuencia (High Frequency), mejor rendimiento en los hilos/Threads.
- **Opcional:** **n** - Características de red (Network feature) para alto rendimiento y baja latencia en red.
- **Tamaño (Size):** **xlarge** - Indica el tamaño de la instancia, lo cual afecta la cantidad de RAM y núcleos de CPU. AWS utiliza distintos tamaños como small, medium, large, xlarge, etc.

AMIs

- Plantillas que contienen la configuración inicial de una instancia EC2.
- Son personalizables.
- Con los volúmenes EBS es posible guardar archivos de configuración.
- Una AMI tiene uno o varios snapshots EBS.
- Una AMI se puede adjuntar a varias instancias para generar múltiples configuraciones iguales.

Elementos clave en una instancia EC2

- Hardware físico:** Componentes como almacenamiento, red y cómputo.
- AMI (Amazon Machine Image):** Plantilla de configuración de software que incluye el sistema operativo y aplicaciones.
- Hypervisor:** Software que virtualiza los recursos físicos y permite que múltiples sistemas operativos se ejecuten en un solo servidor físico.

Hypervisor

Función: Asigna recursos y gestiona el hardware físico en un entorno virtualizado.

Tipos en AWS:

- Original Hypervisor:** Utilizado en los inicios de EC2.
- Nitro Hypervisor:** Parte de AWS Nitro System, que mejora la innovación y la seguridad.

AWS Nitro System

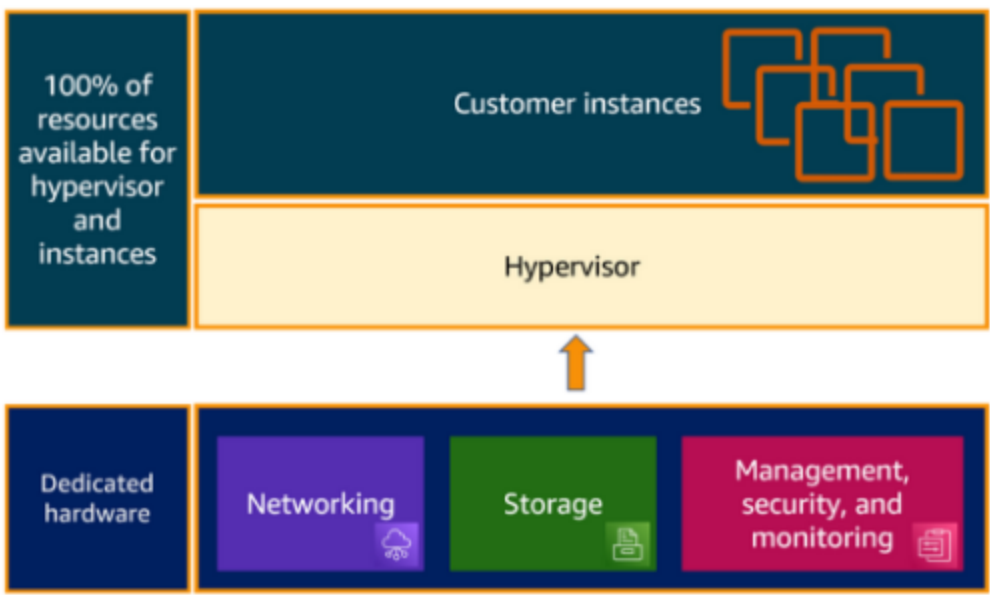
Combina hardware dedicado y un hypervisor liviano para optimizar el rendimiento de las instancias.

Funciones

- Offloading:** Desvía operaciones de entrada/salida (I/O) y gestión de seguridad a hardware dedicado, liberando más recursos para las instancias.
- Rendimiento:** Mejora el rendimiento al proporcionar casi todos los recursos del servidor a las instancias.

Mejoras con AWS Nitro System

- Optimización de recursos:** En el lanzamiento original de EC2, hasta el 30% de los recursos del host se dedicaban al hypervisor, dejando solo el 70% para las instancias.
- Nitro System:** Las funciones del hypervisor se separan y se manejan con hardware y software dedicados, aumentando el rendimiento y reduciendo costos.



Network Interface

Elastic Network Interface (ENI)

- Componente lógico que representa una tarjeta de red virtual en una instancia.
- Permite la comunicación entre instancias, servidores y el internet.
- Similar a la tarjeta de red en laptops, desktops o servidores.

Bare-Metal Instances

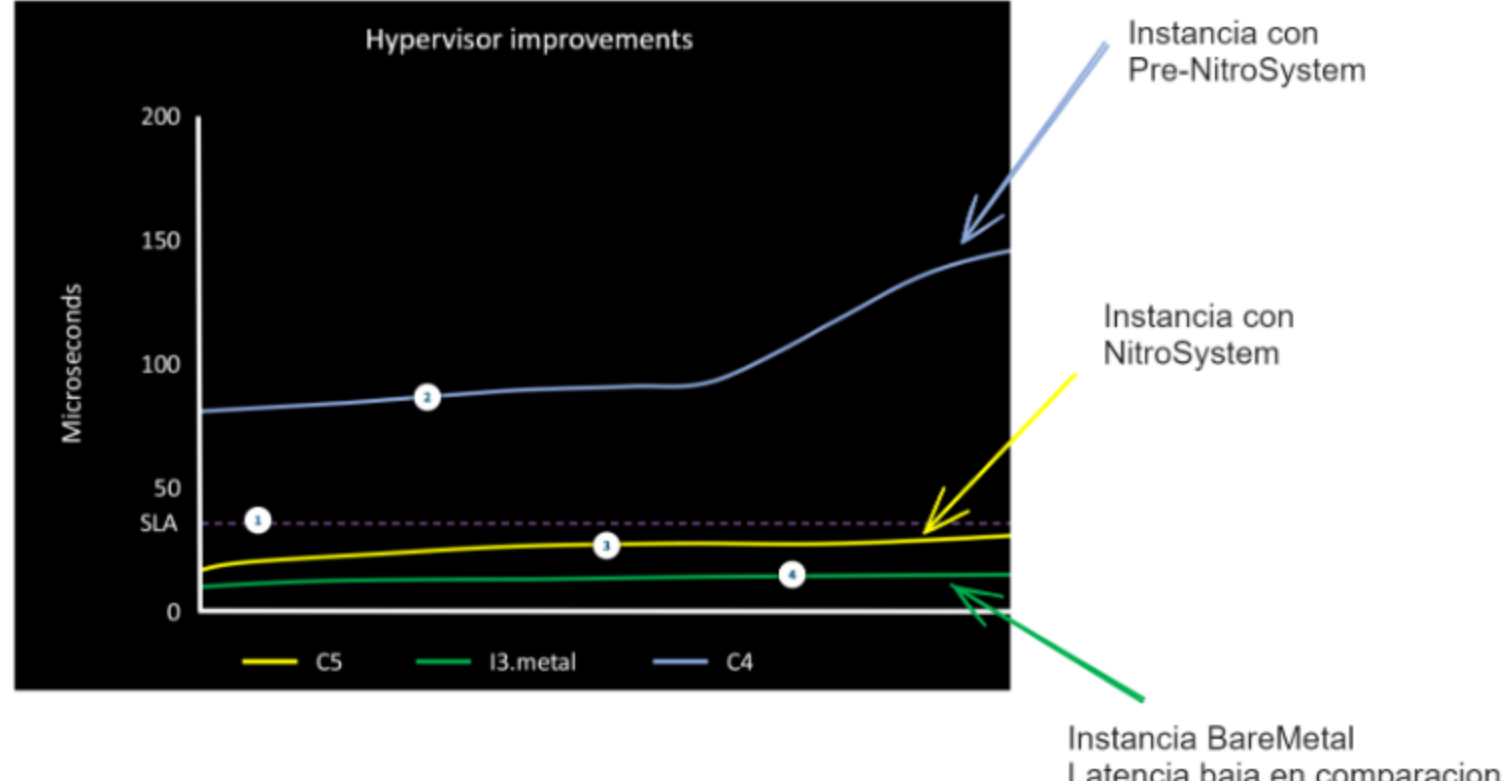
- Instancias Bare-Metal: No tienen software de virtualización preinstalado (hypervisor).
- Diferencia con Dedicated Host:** Un Dedicated Host tiene el hypervisor preinstalado.

Aplicaciones Ideales

- Acceso a características de hardware específicas.
- Aplicaciones que requieren entornos no virtualizados.
- Cientes que desean usar su propio hypervisor.

Beneficios

- Cargas de trabajo especializadas que necesitan acceso directo al hardware.
- Cargas de trabajo heredadas no compatibles con entornos virtuales.
- Aplicaciones críticas de negocio con restricciones de licencias.
- Menor latencia: No tiene la sobrecarga de un hypervisor.



AMIs

Una Amazon Machine Image (AMI) es una plantilla que contiene la configuración necesaria para lanzar una instancia en AWS. Esto incluye el sistema operativo (OS), software preinstalado y archivos de configuración. Una AMI puede ser simple (solo el OS) o compleja según las necesidades.

Uso de AMIs

- Lanzar nueva instancia:** Utiliza una AMI para lanzar una nueva instancia.
- Duplicación:** Una AMI puede ser usada para lanzar múltiples instancias idénticas.
- Fuentes:** Las AMIs pueden ser provistas por AWS, la comunidad de AWS, comparadas en el AWS Marketplace, o creadas por el usuario.

Ejemplo

Puedes preinstalar instancias idénticas para una campaña. Creas una nueva instancia EC2, la personalizas y luego con esa AMI puedes lanzar las instancias necesarias.

Modos de Arranque

- UEFI (Unified Extensible Firmware Interface):**
 - Protocolo de arranque para instancias Graviton.
 - Permite verificar la identidad de los módulos de software.
 - UEFI Secure Boot:** Asegura que solo el software firmado con claves criptográficas pueda ser ejecutado.
- Legacy BIOS:**
 - Usado por defecto en tipos de instancia Intel y AMD.
 - Método tradicional de arranque del sistema.
 - Compatible con versiones antiguas y nuevas de Windows OS.

AMIs Compartidas y Públicas

- AMIs Compartidas:** AMIs disponibles para otros usuarios. Se pueden personalizar y luego crear una nueva AMI.
- AMIs Públicas:** AWS no garantiza la integridad o seguridad de AMIs públicas. Es recomendable obtener AMIs de fuentes confiables, como las de Amazon.

Consideraciones al Compartir AMIs

- Recursos Regionales:** AMIs deben buscarse y compartirse desde la misma región.
- Componentes Restringidos:** AMIs con volúmenes cifrados, snapshots cifrados o códigos de producto no pueden hacerse públicas.
- Datos Sensibles:** Evitar exponer datos sensibles siguiendo las recomendaciones de seguridad.
- Fecha de Deprecación:** Por defecto, la fecha de deprecación de AMIs públicas es de dos años desde la creación. Se puede ajustar según sea necesario.
- Facturación:** El uso de AMIs por otras cuentas no genera costos adicionales para el propietario de la AMI; las cuentas que lanzan instancias son las que asumen los costos.

Resumen de Paid AMIs

Una AMI de pago es una imagen que puedes comprar en el AWS Marketplace, donde encuentras software para AWS, incluyendo AMIs organizadas en categorías. Lanzar una instancia desde una AMI de pago es igual que con cualquier otra AMI y se cobra según las tarifas del propietario de la AMI, además de las tarifas estándar de AWS.

Permisos de Lanzamiento

- Permiso Público:** Permite la instancia para todas las cuentas de AWS.
- Explícito:** Permite instancias para cuentas específicas, organizaciones o unidades organizacionales específicas.

EC2 Image Builder

AWS EC2 Image Builder es una herramienta que automatiza el proceso de creación y mantenimiento de AMIs. Es ideal para mantener las AMIs actualizadas con los últimos parches de seguridad y actualizaciones de software.

- Características:**
 - Automatiza la creación y el mantenimiento de AMIs.
 - Mantiene las AMIs seguras y actualizadas automáticamente.
 - No requiere pasos manuales ni la creación de una canalización de automatización propia.
 - Image Builder es gratuito y se integra con AWS Organizations.

Lanzamiento de AMIs

- Costos:** No se cobra por el uso de AMIs públicas, pero las AMIs de pago incluyen costos adicionales según el proveedor.

Shared Tenancy

Tenencia Compartida (por defecto):

- **Descripción:** Múltiples cuentas de AWS comparten el mismo hardware físico.
- **Uso:** Por defecto, cualquier instancia de EC2 se lanza en un entorno compartido a menos que se especifique lo contrario.
- **Costo:** Es la opción más económica.
- **Seguridad:** Cada instancia está aislada y segura del resto, aunque compartan el mismo hardware físico.
- **Limitaciones:** No soporta instancias que utilicen el modelo Bring Your Own License (BYOL).
- **Soporte:** Soporta instancias Spot y tipos de instancias burstable.

Instancia Dedicada

- **Descripción:** Las instancias se ejecutan en hardware dedicado a un solo cliente.
- **Uso:** Permite asegurar que solo tus instancias se ejecuten en el hardware dedicado, aunque las instancias dedicadas pueden compartir hardware con otras instancias de la misma cuenta AWS que no sean dedicadas.
- **Control:** No ofrece visibilidad ni control sobre el hardware subyacente o cómo se colocan las instancias en el hardware.
- **Costo:** Tiene dos componentes de precio, una tarifa por uso por HORA por instancia y una tarifa por REGIÓN dedicada.
- **Seguridad:** Las instancias están dedicadas a un solo cliente, proporcionando un entorno más aislado en comparación con la tenencia compartida.

Host Dedicado

- **Descripción:** Un servidor físico donde toda la capacidad de instancia está completamente dedicada al caso del cliente.
- **Uso:** Ideal para cargas de trabajo que requieren cumplimiento, seguridad, o razones de licencias que necesitan servidores dedicados.
- **Control:** Ofrece visibilidad y control sobre cómo se colocan las instancias en el servidor físico y el hardware subyacente. Permite usar el modelo BYOL sin cargos adicionales por el uso de software.
- **Configuración:** Se puede usar el editor de afinidad de instancias del host dedicado, permitiendo un control detallado sobre la infraestructura física.
- **Costo:** Más alto debido al nivel de dedicación y control proporcionado.

Diferencias Clave

- **Visibilidad y Control**
 - **Instancia Dedicada:** No proporciona visibilidad ni control sobre el hardware subyacente.
 - **Host Dedicado:** Proporciona visibilidad y control sobre el hardware subyacente y la colocación de instancias.
- **Uso de Licencias**
 - **Instancia Dedicada:** No permite usar el modelo BYOL.
 - **Host Dedicado:** Permite usar el modelo BYOL.
- **Costo**
 - **Instancia Dedicada:** Cobro por uso por hora y tarifa por región.
 - **Host Dedicado:** Generalmente más costoso debido al control y dedicación total del hardware.
- **Flexibilidad y Aislamiento**
 - **Instancia Dedicada:** Aislamiento a nivel de hardware para las instancias de un solo cliente, pero sin control sobre la infraestructura.
 - **Host Dedicado:** Aislamiento completo y control sobre la infraestructura física, adecuado para entornos con requisitos estrictos de cumplimiento y licencias.

Modelo BYOL en AWS

BYOL (Bring Your Own License) es un modelo que permite usar utilizar licencias de software previamente adquiridas en las instancias de AWS.

Ventajas

- **Ahorro de Costos:** Evita la necesidad de comprar nuevas licencias específicas para la nube.
- **Continuidad:** Utiliza software ya familiar y configurado según las necesidades.
- **Escalabilidad:** Aprovecha la infraestructura escalable de AWS sin cambiar las herramientas y aplicaciones existentes.

AWS Networking Fundamentals

Regiones y Zonas de Disponibilidad

- Una Región de AWS está compuesta por múltiples Zonas de Disponibilidad, cada una aislada pero conectada por conexiones de alta velocidad y baja latencia.
- **Disponibilidad y Durabilidad:** El diseño de múltiples Zonas de Disponibilidad ofrece fiabilidad y durabilidad de los servicios.
- **Soberanía de Datos:** Los datos almacenados en una Región de AWS no salen de esa Región a menos que se otorgue explícitamente el permiso, cumpliendo con las leyes y normativas locales.
- **Elección y Flexibilidad:** Los usuarios tienen la flexibilidad de elegir en qué Región ejecutar sus aplicaciones.

Select REGION

- **Normativas de Datos Locales:** Algunas empresas deben cumplir con regulaciones específicas sobre la localización de datos.
 - Ejemplo: Si tu empresa requiere que todos los datos residan dentro de los límites del Reino Unido, se elige la Región de Londres.
- **Latencia y Entrega de Contenidos:** Seleccionar una Región cercana a la mayoría de tus clientes reduce la latencia y permite una entrega de contenidos más rápida.
- **Disponibilidad de Servicios:** Algunos servicios pueden no estar disponibles en todas las Regiones.
- **Costos Regionales:** Algunas Regiones pueden tener costos diferentes debido a distintos niveles de infraestructura.

AZs (Availability Zones)

- Las Zonas de Disponibilidad (Availability Zones) son un componente clave dentro de las Regiones de AWS, diseñadas para ofrecer alta disponibilidad.
- **Elección Flexible:** La ubicación de las Zonas de Disponibilidad está pensada para ofrecer tolerancia a fallos.

! [Aquí img]

Infraestructura Independiente

- **Componentes:** Cada Zona de Disponibilidad cuenta con alimentación eléctrica, refrigeración y seguridad física independiente.
- **Alta Disponibilidad:** Diseñar aplicaciones para que funcionen en múltiples Zonas de Disponibilidad aumenta la tolerancia a fallos y la disponibilidad.

Implementación de EC2

- Al lanzar una instancia de EC2, se usa una máquina virtual en el hardware físico ubicado dentro de una o más zonas de datos que componen una Zona de Disponibilidad.
- **Replicación:** Por defecto, las instancias de EC2 no replican copias a través de las Zonas de Disponibilidad. Sin embargo, puedes crear instancias separadas en cada Zona de Disponibilidad para asegurar que, en caso de que una instancia falle, otra instancia en una Zona de Disponibilidad diferente pueda gestionar las solicitudes de los usuarios.

Servicios de AWS y Redundancia

- Algunos servicios de AWS, como Amazon RDS, replican automáticamente copias entre múltiples Zonas de Disponibilidad para tolerancia a fallos y redundancia.
- **Diseño Personalizado:** Aunque EC2 no replica automáticamente, puedes diseñar tu arquitectura para incluir instancias en múltiples AZs, mejorando la disponibilidad y el rendimiento de tu aplicación.

Resumen Virtual Private Cloud (VPC)

Definición de VPC

- Una VPC es una red virtual privada dedicada a tu cuenta de AWS.
- **Configuración Basada en Software:** Se configura y mapea directamente a infraestructura física.
- **Aislamiento Lógico:** Cada VPC está lógicamente aislada de otras redes virtuales en la nube de AWS.

Características de la VPC

- Una VPC es como un datacenter en la nube: segmentada en subredes, cada apartamento es una red separada.
- **Gateway:** Permite el acceso dentro y fuera de la VPC.
- **Lanzamiento de Recursos:** Los recursos dentro de la VPC pueden ser lanzados en una AZ, incluso instancias EC2.

Tipos de VPC

- **Default VPC:** Se crea automáticamente al abrir una cuenta AWS en cada región.
 - Incluye una subred pública en cada zona de disponibilidad, un Internet Gateway y configuraciones de DNS.
 - Ideal para comenzar rápidamente y lanzar instancias públicas como sitios web simples.
 - Solo puedes tener una Default VPC por región.
- **Nondefault VPC:** Son adicionales y deben ser creadas manualmente.
 - Están aisladas por defecto, sin permitir tráfico de entrada o salida sin configuración explícita.
 - Pueden tener tenancy predeterminado (recursos en hardware compartido) o dedicado (hardware exclusivo para tu cuenta).
 - Puedes tener hasta cinco Nondefault VPCs por región, aunque este límite puede ser aumentado mediante una solicitud a AWS Support.

Subnet

- Una subred (o subnet) es una subdivisión lógica de una red IP más grande. Se utiliza para organizar y segmentar una red en partes más pequeñas.
 - **Funcionamiento:** Un subred consiste en una porción más pequeña del rango de direcciones IP de una VPC.
 - **Ubicación:** Cada subred tiene que estar dentro de una única zona de disponibilidad (Availability Zone) y no puede abarcar varias zonas.
 - **Tipos:**
 - **Subred Pública:** Para recursos que necesitan estar conectados a internet, como servidores web.
 - **Subred Privada:** Para recursos que no necesitan acceso directo a internet, como bases de datos.
 - **Dual Stack:** AWS VPC usa IPv4 y también puedes optar por usar IPv6.

Network ACLs (Access Control Lists)

- Un Network ACL es un filtro de seguridad que controla el tráfico que entra y sale de una subnet.
 - **Funcionamiento:**
 - Se aplica a nivel de la subnet.
 - El ACL por defecto permite todo el tráfico, pero se pueden crear ACLs personalizados que bloquean todo el tráfico hasta que se configuren reglas.
 - **Reglas:**
 - No recuerdan rutas de tráfico, por lo que se necesitan reglas explícitas para tráfico entrante y saliente.
 - Las reglas se procesan en orden desde el número de regla más bajo al más alto, y se detienen al encontrar una coincidencia.

Security Groups

- Un Security Group es una característica de seguridad asignada a la interfaz de red elástica (ENI) de los recursos de AWS.
 - **Funcionamiento:**
 - Se aplica a nivel de la instancia.
 - Las reglas de los Security Groups se evalúan permitiendo automáticamente el tráfico saliente correspondiente.
 - No tienen reglas explícitas de denegación; si no está explícitamente permitido, está denegado.

Managing Instance States

Encendido y Ciclo de Encendido

Un servidor físico pasa por varios estados durante su ciclo de vida.

Estados:

- **Apagado:** El servidor está apagado.
- **Encendido (Cold Booting):** El servidor se enciende y pasa al estado de funcionamiento (running).
- **Reinicio (Rebooting):** Si se reinicia, cambia al estado de reinicio.

Modos de Ahorro de Energía

- **Sleep (Suspensión):** Se detienen todas las acciones y se guarda todo en la memoria, usando solo energía mínima.
- **Hibernate (Hibernación):** Guarda los archivos y aplicaciones abiertas en el disco duro, permitiendo apagar completamente.

Ciclo de Vida de una Instancia EC2

- **Pending:** Al lanzarla, brevemente está en estado pendiente mientras se registra.
- **Running:** Luego pasa al estado de funcionamiento. Permanece activa hasta que se detiene o termina.
- **Stopping/Stopped:** Puedes detener una instancia, la cual puede ser reiniciada posteriormente.
- **Terminated:** Una vez terminada, no puede reiniciarse.

Estados Adicionales de las Instancias

Hibernation (Hibernación):

- **Proceso:** El sistema operativo guarda el contenido de la memoria RAM en el volumen raíz EBS.
- **Al iniciar:**
 1. El volumen raíz EBS se restaura.
 2. Se recargan los contenidos de la RAM.
 3. Los procesos previamente en ejecución se reanudan.
 4. Los volúmenes de datos adjuntos se vuelven a conectar y la instancia retiene su ID.

Reboot (Reinicio):

- AWS puede programar reinicios para mantenimiento de hardware.
- Equivalente a un reinicio del sistema operativo.
- Conserva: Nombre DNS público, direcciones IPv4 e IPv6, y datos en volúmenes de almacenamiento de instancias.
- **Nota:** No inicia un nuevo periodo de facturación, a diferencia de detener y arrancar la instancia.

Retirement (Retiro):

- Programado cuando AWS detecta fallos irreparables en el hardware.
- **Volumen Raíz EBS:** La instancia se detiene y puede ser reiniciada, migrándose a nuevo hardware.
- **Volumen Raíz de Almacenamiento de Instancia:** La instancia se termina y no puede ser reutilizada.

Storage Options Amazon EC2

Tipos de Almacenamiento de Amazon EC2

Almacenamiento de Instancias (Instance Store):

- Almacenamiento temporal que se asigna a tu instancia de EC2.
- Proporciona almacenamiento de bloques a nivel de hardware, adjunto a la instancia.
- **Usos:**
 - Datos que cambian con frecuencia.
 - Almacenamiento de caché.
 - Datos temporales.
- **Características:**
 - Los datos se pierden cuando la instancia se detiene, termina o falla.
 - No se recomienda para datos persistentes.

Elastic Block Store (EBS):

- Almacenamiento persistente que se puede adjuntar a instancias EC2.
- Proporciona almacenamiento de bloques a nivel de red.
- **Usos:**
 - Bases de datos.
 - Archivos de logs.
 - Aplicaciones que requieren almacenamiento persistente.
- **Características:**
 - Los volúmenes EBS son replicados dentro de su Zona de Disponibilidad.
 - Se pueden crear snapshots para respaldos y restauraciones.
 - Se pueden cambiar los tipos de volúmenes para ajustar el rendimiento y el costo.
 - Los datos persisten incluso cuando la instancia se detiene.

Tipos de Volúmenes EBS

1. General Purpose SSD (gp3/gp2):

- Volúmenes SSD balanceados.
- **Usos:**
 - Volúmenes de arranque.
 - Volúmenes de datos que necesitan un rendimiento equilibrado.
- **Rendimiento:**
 - IOPS: Hasta 16,000 (gp3) / 16,000 (gp2).
 - Throughput: Hasta 1,000 MB/s (gp3) / 250 MB/s (gp2).

2. Provisioned IOPS SSD (io2/io1):

- Volúmenes SSD de alto rendimiento.
- **Usos:**
 - Aplicaciones críticas de negocio.
 - Bases de datos intensivas en I/O.
- **Rendimiento:**
 - IOPS: Hasta 64,000 (io2) / 64,000 (io1).
 - Throughput: Hasta 1,000 MB/s (io2) / 1,000 MB/s (io1).

3. Throughput Optimized HDD (st1):

- Volúmenes HDD optimizados para throughput.
- **Usos:**
 - Big data.
 - Data warehouses.
- **Rendimiento:**
 - Throughput: Hasta 500 MB/s.
 - IOPS: Variado, no tan alto como SSDs.

4. Cold HDD (sc1):

- Volúmenes HDD para datos poco frecuentados.
- **Usos:**
 - Almacenamiento de archivos.
 - Backups a largo plazo.
- **Rendimiento:**
 - Throughput: Hasta 250 MB/s.
 - IOPS: Variado, no tan alto como SSDs.

EBS Snapshots

- **Backups Incrementales:** Solo los bloques modificados desde el último snapshot.
- **Restauración:** Puedes crear volúmenes EBS a partir de snapshots.
- **Copias Regionales:** Se pueden copiar snapshots entre regiones.

Opciones de Almacenamiento de Archivos

Amazon S3:

- Almacenamiento de objetos.
- **Usos:**
 - Archivos estáticos.
 - Backups.
 - Data lakes.
- **Características:**
 - Alta durabilidad.
 - Almacenamiento de bajo costo.
 - Acceso vía HTTP/HTTPS.

Amazon EFS (Elastic File System):

- Almacenamiento de archivos gestionado.
- **Usos:**
 - Aplicaciones que requieren un sistema de archivos compartido.
 - Archivos de aplicaciones.
- **Características:**
 - Escalable.
 - Acceso concurrente desde múltiples instancias EC2.

Amazon FSx:

- Almacenamiento de archivos totalmente gestionado.
- **Usos:**
 - Aplicaciones que requieren sistemas de archivos específicos (Windows, Lustre).
- **Características:**
 - Integración con Active Directory.
 - Rendimiento optimizado para cargas de trabajo específicas.

Consideraciones de Seguridad

- **Encriptación:** Tanto en tránsito como en reposo.
- **Control de Acceso:** Utiliza IAM para gestionar permisos.

Elastic Network Interfaces (ENI)

- Una interfaz de red elástica (ENI) es un componente lógico de red que representa una tarjeta de red virtual.
- Permite que la instancia EC2 se comunique en la red con otros hosts, recursos e internet externo.
- Al crear un grupo de seguridad, este se asocia con la interfaz de red elástica.
- El tráfico que intenta conectarse a la interfaz de red elástica debe tener una regla del grupo de seguridad que permita el acceso entrante a la instancia.

Tipos de Interfaces de Red Elásticas

Primaria y Secundaria

- **Primaria:** Se crea por defecto cuando se crea la instancia. No se puede separar ni mover de la instancia en la que fue creada.
- **Secundaria:** Es una interfaz adicional que se crea y se adjunta a la instancia.

Nota: El número máximo de interfaces de red elásticas que se pueden usar varía según el tipo de instancia.

Usos y Configuración de ENI

Un ENI es útil para proporcionar flexibilidad y escalabilidad en la configuración de red de las instancias EC2. Permite que una instancia tenga múltiples interfaces de red, lo que puede ser útil para gestionar tráfico de red diverso o proporcionar aislamiento de red.

Ejemplo:

- Tienes una aplicación en una instancia EC2 y necesitas separar el tráfico de administración del tráfico de usuario por razones de seguridad.
- Puedes usar un ENI primario para el tráfico principal de la aplicación y agregar una ENI secundaria para el tráfico administrativo.

Configuración de ENI Secundaria

1. Crea una ENI secundaria y la asocias con la misma instancia EC2.
2. Configuras un grupo de seguridad específico para la ENI secundaria que solo permite acceso desde tu red de administración.

Aquí imagen sobre configuración de ENI

Mover una ENI entre Instancias

Mover una ENI entre instancias es útil para mantener la continuidad de la red y la configuración de seguridad cuando necesitas reemplazar o actualizar una instancia EC2 sin interrumpir el servicio.

Ejemplo:

- Tienes una aplicación crítica en una instancia EC2 y necesitas realizar mantenimiento en esa instancia o actualizarla. No quieres que el tráfico de red se interrumpa durante este proceso.

Pasos:

1. Adjuntar una ENI Secundaria: Crea y adjunta una ENI secundaria a tu instancia actual que tiene configuraciones específicas de red y reglas de seguridad.
2. Mover la ENI a una Nueva Instancia:
 - Antes de realizar el mantenimiento, separa la ENI de la instancia actual.
 - Adjuntas la misma ENI a una nueva instancia que está preparada para tomar el relevo.
3. Redirección Automática de Tráfico:
 - El tráfico de red se redirige automáticamente a la nueva instancia debido a que los atributos y las configuraciones de la ENI se mantienen, lo que garantiza una transición sin interrupciones.

Dirección IP Pública

Una dirección IP pública permite que las instancias EC2 sean accesibles desde internet. Es esencial para aplicaciones y servicios que necesitan ser accesibles públicamente, como servidores web.

Asignación de Dirección IP Pública:

- Las direcciones IP públicas son asignadas desde un pool de direcciones IPv4 públicas de Amazon.
- Estas direcciones no están asociadas con tu cuenta de AWS.
- Cuando una instancia se detiene, hiberna o termina, la dirección IP pública se libera y se asigna una nueva cuando la instancia se reinicia.

Aquí imagen sobre IP Pública Asignada y Elastic IP

Elastic IP (EIP)

Si necesitas una dirección IP pública persistente que puedas asociar y disociar según sea necesario, usa una dirección IP elástica (Elastic IP).

Usos de Elastic IP:

- Mantener la misma dirección IP pública para que los usuarios puedan acceder siempre a la misma URL.
- Asociar la Elastic IP a tu instancia EC2 y esta dirección IP permanecerá igual, incluso si paras y reinicias la instancia.

Costos y Límites:

- Tienes un límite de cinco Elastic IPs por región por cuenta de AWS.
- Una Elastic IP no está asociada a una instancia, se cobra una pequeña tarifa si no está en uso.

Traer Tus Propias Direcciones IP (BYOIP) y IPv6

BYOIP

Permite traer tus propias direcciones IP a AWS si ya tienes un rango de direcciones IP públicamente enrutable. Esto te permite mantener el control y la consistencia en tu infraestructura de red.

Ejemplo:

- Migrar a AWS manteniendo las mismas direcciones IP que tus clientes ya conocen y utilizan para acceder a tus servicios.

IPv6

Asociar bloques CIDR IPv6 con tu VPC y subredes es útil para preparar tu infraestructura para el futuro del direccionamiento IP, ya que IPv6 proporciona un espacio de direcciones mucho más grande que IPv4.

Consideraciones de IPv6:

- Mantener el control sobre tus direcciones IP y seguir utilizándolas sin cambios para tus usuarios finales.
- IPv6 se asigna automáticamente desde el pool de Amazon.

Aquí imagen sobre BYOIP y IPv6

Detalles Clave:

- Las ENI proporcionan flexibilidad y escalabilidad en la red de EC2.
- Es posible mover ENI entre instancias para mantenimiento sin interrumpir el servicio.
- Las direcciones IP públicas permiten la accesibilidad desde internet, mientras que las Elastic IP proporcionan persistencia.
- BYOIP y IPv6 son útiles para mantener el control de la red y prepararse para el futuro.

Resiliencia y configuración en Amazon EC2

La resiliencia asegura que tus aplicaciones y datos puedan recuperarse rápidamente de interrupciones, manteniendo la disponibilidad y minimizando el tiempo de inactividad.

Ejemplos:

Configuración de Terminación y Comportamiento de Apagado

Supongamos que tienes una aplicación crítica en una instancia EC2 que maneja datos sensibles. Para asegurarte de que los datos no se pierdan al terminar la instancia, configuras el atributo **DeleteOnTermination** del volumen raíz a **FALSE**, preservando los volúmenes incluso si la instancia se termina. Además, configuras el comportamiento de apagado a **STOP** para evitar la terminación accidental.

Uso de Auto Scaling y Elastic Load Balancing

Para manejar variaciones en el tráfico y asegurar la disponibilidad continua, configuras un grupo de Auto Scaling que lanza nuevas instancias cuando la carga aumenta y las termina cuando disminuye. Usas un Elastic Load Balancer para distribuir el tráfico entre estas instancias, asegurando que el servicio permanezca disponible y eficiente.

Protección contra Terminación y Recuperación Automática

Activar la protección contra terminación para instancias críticas evita terminaciones accidentales mediante la consola o API. Configuras también la recuperación automática simplificada para que cualquier fallo en las comprobaciones de estado del sistema inicie automáticamente la recuperación de la instancia, minimizando el tiempo de inactividad.

Comparing Purchasing Options

On-Demand Instances

- **Descripción:** Pagas por capacidad de cómputo por hora o por segundo, sin compromisos a largo plazo ni pagos iniciales. Ideal para cargas de trabajo de corta duración, irregulares o impredecibles que no pueden ser interrumpidas.
- **Uso común:** Desarrollo y prueba de aplicaciones.

Savings Plans

- **Descripción:** Modelo de precios flexible que ofrece descuentos a cambio de un compromiso de uso (medido en \$/hora) por 1 o 3 años.
- **Tipos:**
 - **Compute Savings Plans:** Se aplica a EC2, AWS Lambda y AWS Fargate.
 - **EC2 Instance Savings Plans:** Se aplica a EC2.
 - **Amazon SageMaker Savings Plans:** Se aplica a Amazon SageMaker.
- **Beneficios:** Mayor flexibilidad para cambiar entre tipos de instancias, regiones y servicios.

Reserved Instances (RIs)

- **Descripción:** Descuento en la facturación por el uso de On-Demand Instances a cambio de un compromiso de 1 o 3 años.
- **Tipos:**
 - **Standard RIs:** Descuentos significativos, ideales para cargas de trabajo constantes.
 - **Convertible RIs:** Descuentos menores, permiten cambiar familias de instancias, tipos de SO y tenencias.
- **Uso común:** Cargas de trabajo constantes y sistemas de larga duración.

Spot Instances

- **Descripción:** Permiten solicitar capacidad de cómputo EC2 no utilizada con hasta un 90% de descuento sobre el precio On-Demand. Ideal para aplicaciones con tiempos de inicio y fin flexibles, que pueden tolerar interrupciones.
- **Uso común:** Procesamiento en segundo plano, pruebas de aplicaciones con costos bajos.

Dedicated Hosts

- **Descripción:** Un servidor físico dedicado para tu uso exclusivo. Ayuda a reducir costos al usar tus licencias de software existentes y cumplir con requisitos de cumplimiento y seguridad. Puede ser comprado On-Demand o como una reserva con hasta un 70% de descuento.

Dedicated Instances

- **Descripción:** Instancias EC2 en servidores físicos dedicados exclusivamente a tu cuenta de AWS.
- **Diferencia clave:** Dedicated Hosts ofrecen mayor visibilidad y control sobre la colocación de las instancias en un servidor físico.

Ejemplos

On-Demand Instances

- **Ejemplo:** Imagina que estás desarrollando una nueva aplicación y necesitas instancias EC2 para pruebas. Usar On-Demand Instances te permite lanzar y detener instancias según necesites, pagando solo por el tiempo de uso sin compromisos a largo plazo.

Savings Plans

- **Ejemplo:** Supón que tienes una aplicación web que necesitas ejecutar continuamente durante los próximos 3 años. Al optar por un Compute Savings Plan, puedes obtener precios más bajos para tus instancias EC2 y servicios relacionados, ahorrando hasta un 66% en comparación con los precios On-Demand.

Reserved Instances (RIs)

- **Ejemplo:** Tienes una base de datos crítica que debe estar en funcionamiento constante. Comprar Standard RIs por 3 años te permite obtener un descuento significativo en comparación con el uso de On-Demand Instances, optimizando costos a largo plazo.

Spot Instances

- **Ejemplo:** Tienes una tarea de procesamiento de datos que no es urgente y puede ser interrumpida. Solicitas Spot Instances, aprovechando los precios bajos para ejecutar tu tarea cuando haya capacidad disponible, reduciendo significativamente los costos.

Dedicated Hosts

- **Ejemplo:** Trabajas en un sector con estrictos requisitos de cumplimiento y necesitas usar tus propias licencias de software. Al elegir Dedicated Hosts, puedes garantizar que tus instancias EC2 cumplan con las normativas y aprovechar tus licencias existentes, reduciendo costos.

Dedicated Instances

- **Ejemplo:** Necesitas asegurar que tus instancias EC2 no compartan hardware con otras cuentas de AWS por motivos de seguridad. Al usar Dedicated Instances, todas tus instancias se ejecutan en servidores físicos dedicados exclusivamente a tu cuenta.

Características y Uso de Spot Instances

- **Descripción:** Spot Instances son ideales para aplicaciones sin estado (stateless), tolerantes a fallos y flexibles.
- **Ejemplos de uso:** Cargas de trabajo en contenedores, integración y entrega continua (CI/CD), servidores web sin estado, computación de alto rendimiento (HPC) y tareas de renderizado.
- **Notas:** Aunque son iguales a las instancias On-Demand en cuanto a funcionalidad, las Spot Instances no garantizan que permanecerán en ejecución el tiempo suficiente para completar tus tareas.
- **Disponibilidad:** La disponibilidad de las Spot Instances puede cambiar según la oferta y la demanda. Esto significa que no siempre puedes obtener la capacidad que solicites de inmediato.
- **Interrupciones:** Debes estar preparado para que las Spot Instances puedan ser interrumpidas en cualquier momento si AWS necesita la capacidad para otras tareas.