Amazon EC2 (Elastic Compute Cloud)

Definición

Amazon EC2 (Elastic Compute Cloud) es un servicio web que proporciona capacidad de cómputo escalable en la nube. EC2 permite a los usuarios lanzar y gestionar instancias de servidores virtuales, conocidas como instancias EC2, para ejecutar aplicaciones en la infraestructura de Amazon Web Services (AWS).

Características

1. **Elasticidad:**
    - Permite escalar la capacidad de cómputo hacia arriba o hacia abajo según las necesidades.
    - Fácil de aumentar o reducir el número de instancias en respuesta a la demanda.
2. **Flexibilidad:**
    - Amplia variedad de tipos de instancias optimizadas para diferentes casos de uso.
    - Soporte para múltiples sistemas operativos como Linux, Windows, y macOS.
3. **Pago por Uso:**
    - Modelo de precios basado en el uso, donde se paga solo por los recursos consumidos.
    - Opciones de precios como On-Demand, Reserved Instances y Spot Instances.
4. **Seguridad:**
    - Integración con Amazon VPC (Virtual Private Cloud) para aislar recursos.
    - Soporte para AWS Identity and Access Management (IAM) para gestionar el acceso y permisos.
5. **Integración:**
    - Fácil integración con otros servicios de AWS como S3, RDS, EBS y más.
    - Compatible con herramientas de desarrollo y gestión de AWS como AWS CLI, SDKs, y CloudFormation.

Componentes de Amazon EC2

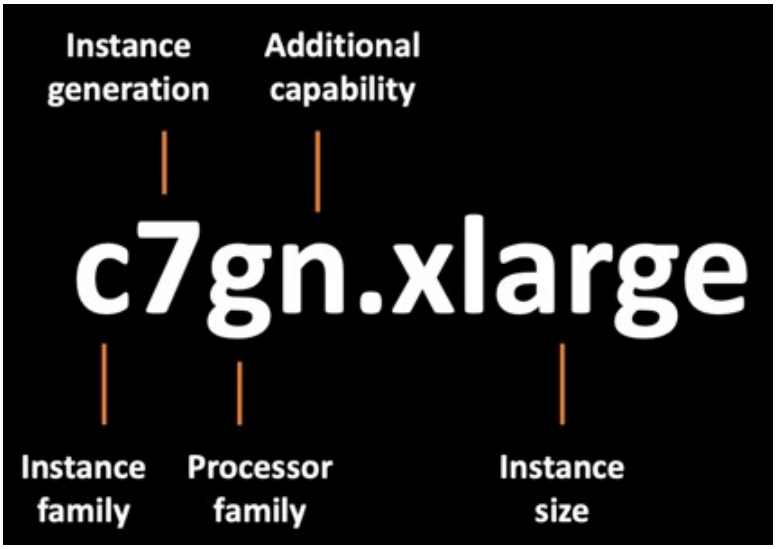
1. **Instancias EC2:**
    - Servidores virtuales que ejecutan aplicaciones.
    - Configurables con diferentes tipos de instancias para ajustar el cómputo, la memoria y el almacenamiento.
2. **Amazon Machine Images (AMIs):**
    - Plantillas preconfiguradas que incluyen un sistema operativo y aplicaciones.
    - Usadas para lanzar nuevas instancias EC2 rápidamente.
3. **Tipos de Instancias:**
    - Variedad de instancias optimizadas para diferentes casos de uso:
        - **Propósito General:** T3, T4g, M5, M6g.
        - **Cómputo Optimizado:** C5, C6g.
        - **Memoria Optimizada:** R5, R6g, X1, X2.
        - **Almacenamiento Optimizado:** I3, I4i.
        - **Aceleradas por GPU:** P3, P4, G4, G5.
        - **Propósito Especial:** Inf1, F1.

4. **Almacenamiento de Bloques (Amazon EBS):**
    - Almacenamiento persistente para instancias EC2.
    - Ofrece opciones de rendimiento y capacidad para diferentes necesidades de almacenamiento.
5. **Almacenamiento de Instancias:**
    - Almacenamiento temporal que se asocia con la instancia EC2.
    - Se pierde cuando la instancia se detiene o termina.
6. **Grupos de Seguridad:**
    - Actúan como firewall virtual para controlar el tráfico de red entrante y saliente.
    - Configurables con reglas para permitir o denegar tráfico en función de protocolos, puertos y direcciones IP.
7. **Elastic IPs:**
    - Direcciones IP estáticas para instancias EC2.
    - Facilitan la reasignación de direcciones IP en caso de fallo de instancia.
8. **Amazon VPC (Virtual Private Cloud):**
    - Red virtual en la nube donde se pueden lanzar recursos de AWS.
    - Proporciona control sobre el entorno de red, incluyendo la selección de rangos de IP, subredes y configuración de gateways.
9. **Key Pairs:**
    - Claves criptográficas para acceder a instancias EC2.
    - Facilitan la autenticación segura mediante SSH para Linux o RDP para Windows.
10. **Auto Scaling:**
    - Permite escalar automáticamente el número de instancias EC2 en respuesta a la demanda.
    - Asegura que se ejecuten las instancias necesarias para manejar la carga.
11. **Elastic Load Balancing (ELB):**
    - Distribuye automáticamente el tráfico entrante entre múltiples instancias EC2.
    - Mejora la disponibilidad y tolerancia a fallos de las aplicaciones.
12. **CloudWatch:**
    - Servicio de monitoreo que proporciona datos y métricas de las instancias EC2.
    - Permite configurar alarmas y acciones automáticas basadas en las métricas monitoreadas.

Tipos de Instancias

1. **Instancias de Propósito General:**
    - **T3, T4g (Burstable Performance):** Balance entre cómputo, memoria y red. Ideal para aplicaciones de uso general.
    - **M5, M6g (Balanceadas):** Ofrecen un equilibrio entre cómputo, memoria y recursos de red. Adecuadas para una amplia variedad de cargas de trabajo.
2. **Instancias de Cómputo Optimizado:**
    - **C5, C6g:** Diseñadas para cargas de trabajo con altos requerimientos de CPU como servidores web, procesamiento de datos y aplicaciones científicas.

3. **Instancias de Memoria Optimizada:**
    - **R5, R6g:** Optimizadas para aplicaciones que requieren gran cantidad de memoria, como bases de datos y análisis en memoria.
    - **X1, X2:** Ofrecen grandes cantidades de memoria para cargas de trabajo en memoria como SAP HANA.
4. **Instancias de Almacenamiento Optimizado:**
    - **I3, I4i:** Ofrecen almacenamiento en disco local de alta velocidad, ideales para bases de datos NoSQL, sistemas de archivos distribuidos y otras aplicaciones que necesitan acceso rápido a almacenamiento en disco.
5. **Instancias Aceleradas por GPU:**
    - **P3, P4:** Diseñadas para aprendizaje profundo, simulaciones científicas y aplicaciones de gráficos.
    - **G4, G5:** Optimizadas para inferencia de aprendizaje automático y procesamiento de gráficos.
6. **Instancias de Propósito Especial:**
    - **Inf1:** Optimizadas para inferencia de machine learning.
    - **F1:** Diseñadas para aplicaciones de computación acelerada con FPGA.



Usos Comunes

1. **Alojamiento de Aplicaciones Web y Servidores:**
    - Hospedaje de sitios web, aplicaciones web y servidores backend.
    - Implementación de servidores de aplicaciones y servidores web como Apache, Nginx, etc.
2. **Desarrollo y Pruebas:**
    - Ambientes de desarrollo y prueba escalables para aplicaciones de software.
    - Entornos de integración continua y despliegue continuo (CI/CD).
3. **Procesamiento de Datos y Análisis:**
    - Procesamiento y análisis de grandes volúmenes de datos.
    - Ejecución de cargas de trabajo de Big Data con herramientas como Hadoop y Spark.
4. **Aprendizaje Automático e Inteligencia Artificial:**
    - Entrenamiento e inferencia de modelos de aprendizaje automático.
    - Ejecución de aplicaciones de IA y análisis predictivo.

5. **Aplicaciones Empresariales:**
    - Implementación de aplicaciones empresariales como ERP, CRM y más.
    - Migración de aplicaciones empresariales existentes a la nube.
6. **Computación Científica y de Alto Rendimiento:**
    - Ejecución de simulaciones científicas y tareas de computación de alto rendimiento (HPC).
    - Uso de instancias con GPU y FPGA para cálculos complejos.

Opciones de Compra

1. **On-Demand Instances:**
    - Pago por segundo o por hora sin compromiso a largo plazo.
    - Ideal para aplicaciones con cargas de trabajo impredecibles y en fase de prueba.
2. **Reserved Instances:**
    - Compromiso a uno o tres años con un costo menor comparado con On-Demand.
    - Tipos: Standard RIs (descuento significativo a cambio de compromiso a largo plazo), Convertible RIs (permiten cambiar atributos de la RI si las necesidades cambian), y Scheduled RIs (reservadas para periodos específicos).
3. **Spot Instances:**
    - Permiten pujar por capacidad de cómputo EC2 no utilizada.
    - Ofrecen hasta un 90% de descuento en comparación con las instancias On-Demand.
    - Ideal para cargas de trabajo interrumpibles como procesamiento por lotes y análisis de datos.
4. **Savings Plans:**
    - Compromiso a un monto de uso por hora a uno o tres años.
    - Ofrecen flexibilidad para cambiar instancias, regiones y tamaños mientras se mantienen los descuentos.
5. **Dedicated Hosts:**
    - Proporcionan servidores físicos dedicados a un solo cliente.
    - Ayudan a cumplir con requisitos regulatorios y de licencias específicas.
    - Mayor control sobre la ubicación de las instancias.
6. **Dedicated Instances:**
    - Instancias EC2 que se ejecutan en hardware dedicado al cliente.
    - Mayor aislamiento en comparación con las instancias estándar.

Grupos de Seguridad en Amazon EC2

Definición

Los grupos de seguridad en Amazon EC2 son conjuntos de reglas que actúan como firewall virtual para controlar el tráfico de red entrante y saliente de las instancias EC2. Los grupos de seguridad determinan qué tráfico está permitido hacia y desde las instancias asociadas.

Características

1. **Filtrado de Tráfico:**
    - Filtrado de tráfico entrante y saliente basado en reglas definidas por el usuario.
    - Reglas configurables para permitir o denegar tráfico en función de protocolos, puertos y direcciones IP de origen y destino.

2. **Estado:**
    - Los grupos de seguridad son con estado (stateful), lo que significa que si se permite una solicitud entrante, la respuesta saliente correspondiente se permite automáticamente y viceversa.
3. **Reglas Permitidas:**
    - Solo se pueden crear reglas que permitan tráfico (allow). No hay reglas explícitas de denegación (deny).
    - Cualquier tráfico que no esté explícitamente permitido por una regla se bloquea automáticamente.
4. **Aplicación Inmediata:**
    - Los cambios en las reglas de los grupos de seguridad se aplican de inmediato a todas las instancias asociadas.
5. **Asignación Flexible:**
    - Se pueden asignar múltiples grupos de seguridad a una instancia EC2.
    - Un grupo de seguridad se puede asociar a múltiples instancias.

Componentes de un Grupo de Seguridad

1. **Reglas de Ingreso:**
    - Controlan el tráfico entrante a las instancias.
    - Configurables por protocolo (TCP, UDP, ICMP), rango de puertos y rango de direcciones IP de origen.
2. **Reglas de Egreso:**
    - Controlan el tráfico saliente desde las instancias.
    - Configurables por protocolo, rango de puertos y rango de direcciones IP de destino.

Ejemplo de Configuración

Regla de Ingreso

Protocolo	Rango de Puertos	Origen	Descripción
TCP	22	0.0.0.0/0	Permitir SSH desde cualquier lugar
TCP	80	0.0.0.0/0	Permitir HTTP desde cualquier lugar
TCP	443	0.0.0.0/0	Permitir HTTPS desde cualquier lugar
TCP	3306	192.168.1.0/24	Permitir MySQL desde una subred específica

Regla de Egreso

Protocolo	Rango de Puertos	Destino	Descripción
TCP	0-65535	0.0.0.0/0	Permitir todo el tráfico saliente

Usos Comunes

1. **Seguridad en Redes:**
    - Protección de instancias EC2 al controlar el acceso de red.
    - Restricción de acceso a servicios críticos solo desde direcciones IP de confianza.

2. **Segmentación de Aplicaciones:**
    - Separación de diferentes capas de una aplicación (por ejemplo, capa web, capa de aplicación y capa de base de datos) mediante grupos de seguridad específicos.
    - Control del tráfico entre las capas para mejorar la seguridad.
3. **Cumplimiento y Auditoría:**
    - Implementación de reglas de seguridad para cumplir con políticas de seguridad y requisitos regulatorios.
    - Monitoreo y registro del tráfico de red permitido y bloqueado.
4. **Flexibilidad y Escalabilidad:**
    - Ajuste dinámico de reglas de seguridad a medida que cambian las necesidades de la aplicación.
    - Asociación de múltiples grupos de seguridad a instancias para una configuración más granular.

Buenas Prácticas

1. **Principio de Menor Privilegio:**
    - Definir reglas de seguridad para permitir solo el tráfico necesario y nada más.
    - Minimizar las aperturas de puertos y las direcciones IP permitidas.
2. **Uso de Descripciones Claras:**
    - Proporcionar descripciones detalladas para cada regla de seguridad para facilitar la gestión y el mantenimiento.
3. **Revisión Periódica:**
    - Revisar y actualizar regularmente los grupos de seguridad para asegurar que solo el tráfico necesario está permitido.
    - Eliminar reglas obsoletas o innecesarias.
4. **Monitoreo y Alerta:**
    - Utilizar AWS CloudTrail y Amazon CloudWatch para monitorear cambios en los grupos de seguridad y generar alertas ante actividades sospechosas.