
Gestión de Cómputo y Ficheros en la Nube

Amazon EC2, S3 y Auto Scaling

Germán Moltó

© Universitat Politècnica de València, 2013-2024.

© Germán Moltó, 2013-2024

Queda prohibida la reproducción total o parcial de esta obra sin la autorización por escrito de los propietarios del copyright por cualquier tipo de medio o procedimiento, incluida la reprografía y el tratamiento informático, así como la distribución de ejemplares mediante alquiler, venta o préstamo público, ni siquiera para ámbito educativo.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

FICHA DE LA UNIDAD

TÍTULO: GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE: Amazon EC2, S3 y Auto Scaling

INTRODUCCIÓN	4
ESQUEMA	4
CONTENIDOS	5
1. Introducción a Amazon Web Services (AWS).....	5
Vídeo-lección: Introducción a Amazon Web Services (AWS).....	5
Actividad: Panorámica de los Servicios de AWS	7
Actividad: La Misteriosa Región GovCloud	8
2. Resumen del servicio Amazon S3	8
Vídeo-lección: Introducción a Amazon S3	8
Pregunta: Coste de Subida de Datos al Cloud.....	10
3. Resumen del servicio Amazon EC2	11
Vídeo-lección: Introducción a Amazon EC2	11
Actividad: Tipos de instancias de Amazon EC2	12
Práctica: Despliegue de Instancias de Máquinas Virtuales con Amazon EC2	13
Vídeo-lección: Otros Servicios de Amazon EC2 para Aplicaciones Cloud	14
4. Resumen del servicio EBS (Elastic Block Store).....	15
5. Resumen del servicio ELB (Elastic Load Balancing).....	16
6. Resumen del servicio Auto Scaling	18
Vídeo-lección: Auto-Escalado de Aplicaciones Cloud	18
Vídeo-lección: Utilizando Amazon Web Services (AWS).....	20
Práctica: Despliegue de Grupos Elásticos de Instancias y Gestión de Datos con AWS	20
7. La Calculadora de Precios de AWS.....	20
Actividad: AWS Simple Monthly Calculator	21
Actividad: AWS Total Cost of Ownership (TCO) Calculator	¡Error! Marcador no definido.
8. Límites de AWS	21
9. Otros Servicios de AWS Relacionados.....	21
10. Modelo de Responsabilidad Compartida.....	22
Consejos y Buenas Prácticas	23
ANEXOS.....	23



REFERENCIAS 23

INTRODUCCIÓN

La demanda de recursos variables de las aplicaciones precisa de herramientas y tecnologías que puedan aprovisionar recursos dinámicamente. Una aplicación o servicio web con un aumento repentino de usuarios (traducido en un incremento del número de peticiones por segundo que recibe dicho servicio) debe poder desplegar nuevas máquinas (virtuales), sobre las que se ejecuta la aplicación que atiende el servicio, para poder satisfacer dicha carga de trabajo, minimizando el impacto en la calidad del servicio ofrecido (por ejemplo, tiempo medio de respuesta del servicio).

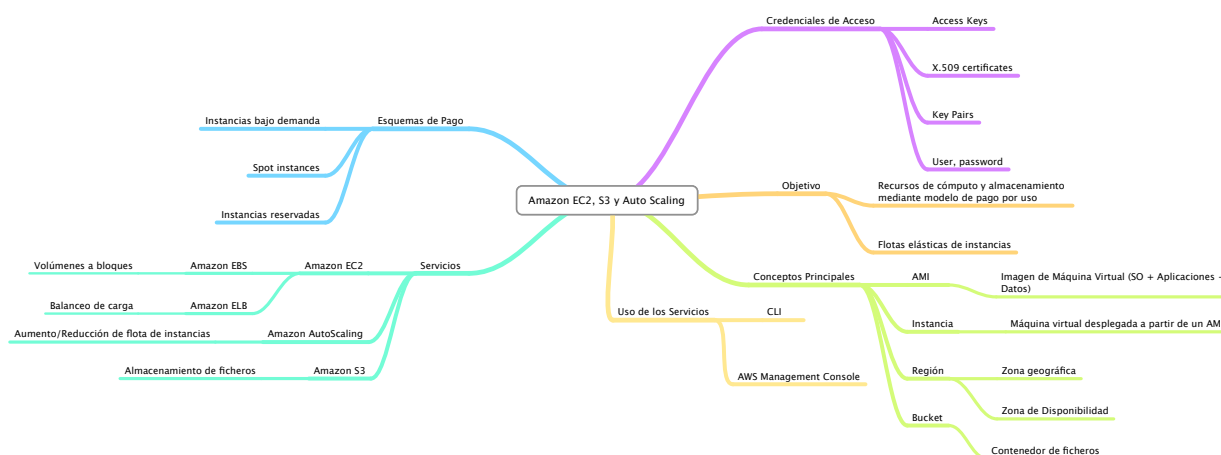
En este sentido, Amazon EC2 (*Elastic Compute Cloud*) [1] permite el despliegue de máquinas virtuales sobre la infraestructura Cloud de AWS. Por su parte, el servicio Auto Scaling [2], en conjunción con el servicio de monitorización Amazon CloudWatch [3], permite la definición de reglas de escalado horizontal mediante las cuales es posible aumentar o reducir el tamaño de una flota de instancias para adaptarse a la carga de trabajo. Además, el servicio ELB (*Elastic Load Balancing*), dentro de EC2, permite la creación de balanceadores de carga que distribuyen las peticiones a un servicio entre las diferentes instancias (máquinas virtuales) que pueden atender dicho servicio. La combinación de estas herramientas facilita la creación de aplicaciones elásticas que puedan auto-aprovisionar (y auto-liberar) recursos en función de la carga de trabajo, pagando únicamente por aquellos recursos utilizados en cada momento.

Por otra parte, el servicio Amazon S3 (*Simple Storage Service*) [4] permite el almacenamiento de ficheros en la nube con replicación automática de los datos para tener tolerancia a fallos. Los ficheros pueden ser accedidos por medio de protocolos estándares (como HTTP o BitTorrent). Adicionalmente, existe la posibilidad de crear dispositivos orientados a bloques, llamados volúmenes EBS (*Elastic Block Store*) [5], dentro del servicio EC2, para conectarlos a instancias, al igual que si de un disco duro o un USB Stick se tratase. Esto permite guardar ficheros y carpetas en dichos volúmenes.

La orquestación de los servicios anteriormente mencionados permite el diseño o arquitectura de aplicaciones elásticas con gestión eficiente de datos en la nube. Todos los servicios anteriormente mencionados se cubren en esta unidad.

ESQUEMA

Este es el esquema de los principales contenidos que trata esta unidad. Se cubren los principales conceptos y la funcionalidad de los servicios de Amazon EC2, S3 y Auto Scaling, los esquemas de pago, las credenciales de acceso necesaria para acceder a los mismos y otros detalles de uso de los servicios.



Esta unidad consta de diferentes recursos didácticos:

- Vídeo-lecciones donde se describen los servicios anteriormente mencionados.
 - VL211: Introducción a AWS (12'25")
 - <https://media.upv.es/player/?id=004333de0-021e-11e6-851a-656f7e06a374&autoplay=true>
 - VL212: Introducción a Amazon S3 (9'24")
 - <https://media.upv.es/player/?id=00028b60-021e-11e6-851a-656f7e06a374&autoplay=true>
 - VL213: Introducción a Amazon EC2 (10'29")

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

- <https://media.upv.es/player/?id=01e729e0-021e-11e6-851a-656f7e06a374&autoplay=true>
- VL214: Otros Servicios de Amazon EC2 para Aplicaciones Cloud (7'43")
 - <https://media.upv.es/player/?id=02ee39a0-021e-11e6-851a-656f7e06a374&autoplay=true>
- VL215: Auto-Escalado de Aplicaciones Cloud (7'33")
 - <https://media.upv.es/player/?id=030e92e0-021e-11e6-851a-656f7e06a374&autoplay=true>
- VL216: Utilizando Amazon Web Services (9'42")
 - <https://media.upv.es/player/?id=027a6930-021e-11e6-851a-656f7e06a374&autoplay=true>
- Transparencias de PowerPoint utilizadas en las grabaciones anteriores.
- Prácticas de uso de los servicios Amazon EC2, S3 y Auto Scaling sobre la infraestructura real de AWS.
- Resumen de los aspectos más importantes abordados en los vídeos (descritos en esta misma sección a modo de apuntes)

Los apuntes incluidos en este documento no deben considerarse como una explicación exhaustiva de los servicios Amazon EC2, S3 y Auto Scaling. Por el contrario, los vídeos incluyen una explicación más detallada de los mismos, que se complementa con las prácticas propuestas sobre la infraestructura de AWS. Las prácticas de esta unidad deben ser realizadas para poder conocer de primera mano la tecnología que subyace a los servicios presentados. Ten en cuenta que, dada la velocidad a la que evoluciona AWS, las novedades se recogen principalmente en este documento, por lo que siempre encontrarás información más actualizada en las guías (como la de este documento) que la expuesta en los vídeos.

CONTENIDOS

1. Introducción a Amazon Web Services (AWS)

Vídeo-lección: Introducción a Amazon Web Services (AWS)

Este vídeo realiza una breve introducción a Amazon Web Services (AWS), el proveedor de Cloud público pionero y actual líder en aprovisionamiento de infraestructura virtual. Incluye una breve descripción del objetivo de los numerosos servicios que ofrece para el despliegue de aplicaciones en la nube. Finalmente, el vídeo aborda las diferentes regiones de AWS y su división en múltiples zonas de disponibilidad, como mecanismo de tolerancia a fallos.

Disponible en: <https://media.upv.es/player/?id=00433de0-021e-11e6-851a-656f7e06a374&autoplay=true>

Duración: 12'25"

Amazon Web Services (AWS) es un proveedor de Cloud público pionero en el campo de las tecnologías Cloud (ofreciendo servicio desde 2006). Además, es líder en el ámbito de los proveedores de infraestructura (IaaS – *Infrastructure as a Service*), permitiendo el acceso a recursos virtuales (cómputo, almacenamiento, red, servicios, etc.) bajo demanda mediante un modelo de pago por uso. A fecha de 2013, no hay ningún otro proveedor que ofrezca un catálogo de servicios tan amplio para el despliegue de aplicaciones Cloud como AWS, por lo que suele ser el proveedor utilizado por la gran mayoría de nuevas aplicaciones Cloud (como Dropbox [8], FourSquare [9] o Pinterest [10]). Tienes una perspectiva histórica de los últimos 10 años de servicios de AWS en [81].

A continuación, se enumeran algunos de los servicios disponibles en AWS. Dada la rapidez con la que los ingenieros de AWS desarrollan nuevos servicios, es posible que cuando leas este documento ya hayan aparecido nuevos servicios. En ese caso, es siempre recomendable que consultes la fuente original con el catálogo de servicios actualizado [11]:

Algunos de los principales productos y servicios de AWS:

- Computación
 - EC2 (Elastic Compute Cloud)
 - EC2 Container Service
 - Elastic Beanstalk
 - Lambda
- Almacenamiento y Entrega de Contenido
 - S3 (Simple Storage Service)

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

- CloudFront
 - Elastic File System
 - Amazon Glacier
 - Import/Export Snowball
 - Storage Gateway
- Base de datos
 - RDS (Relational Database Service)
 - DynamoDB
 - ElastiCache
 - RedShift
 - DMS (Database Migration Service)
 - SimpleDB
- Redes
 - VPC (Virtual Private Cloud)
 - Direct Connect
 - Route 53
- Herramientas para Desarrolladores
 - CodeCommit
 - CodeDeploy
 - CodePipeline
- Herramientas de Gestión
 - CloudWatch
 - CloudFormation
 - CloudTrail
 - Config
 - OpsWorks
 - Service Catalog
 - Trusted Advisor
- Seguridad e Identidad
 - IAM (Identity & Access Management)
 - Directory Service
 - Inspector
 - WAF (Web Application Firewall)
 - Certificate Manager
- Analítica
 - EMR (Elastic MapReduce)
 - Data Pipeline
 - Elasticsearch Service
 - Kinesis
 - Machine Learning
- IoT (Internet of Things)
 - AWS IoT
- Desarrollos de Juegos
 - GameLift
- Servicios para Móviles
 - Mobile Hub
 - Cognito
 - Device Farm
 - Mobile Analytics
 - SNS (Simple Notification Service)
- Servicios de Aplicaciones
 - API Gateway
 - AppStream
 - CloudSearch
 - Elastic Transcoder

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

- SES (Simple Email Service)
- SQS (Simple Queue Service)
- SWF (Simple Workflow Service)
- Aplicaciones Empresariales
 - WorkSpaces
 - WorkDocs
 - WorkMail
- Personal
 - Amazon Mechanical Turk
- Software
 - AWS Marketplace

La Figura 1 muestra una panorámica de los principales servicios de AWS organizados por categorías, obtenida de [70].

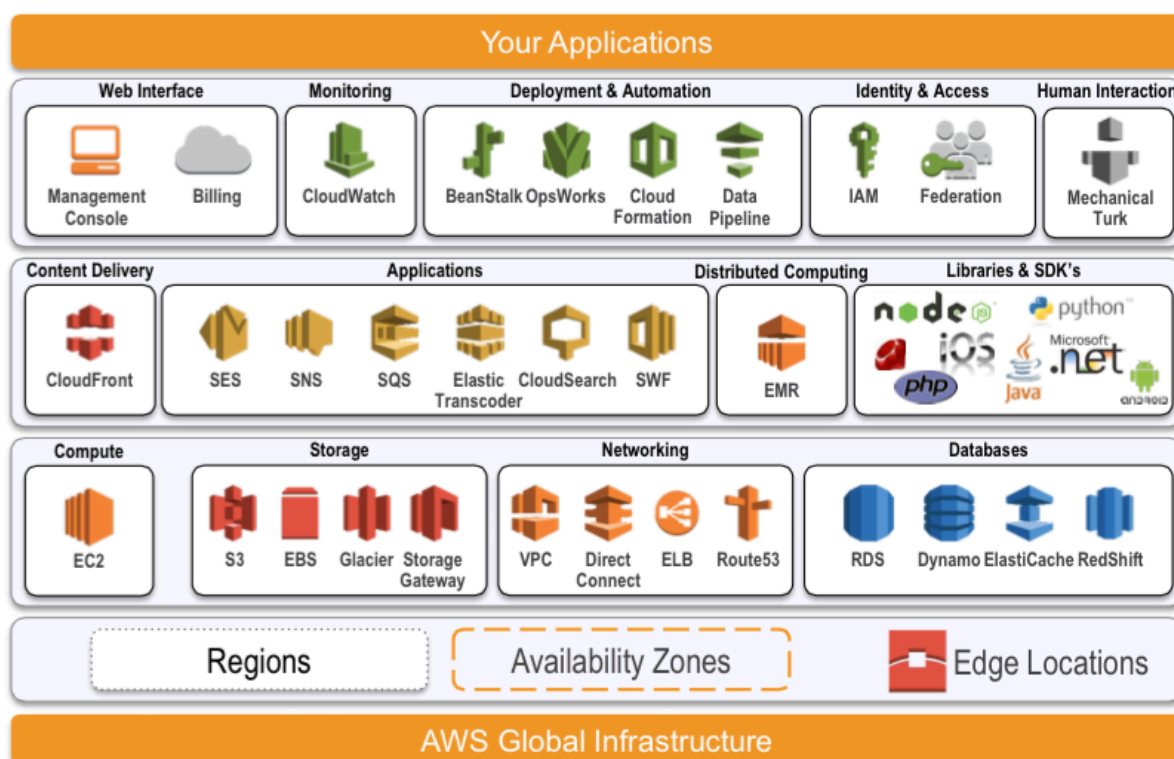


Figura 1. Panorámica de los principales servicios de AWS

+ Actividad: Panorámica de los Servicios de AWS

Visita la siguiente dirección <http://aws.amazon.com/es/products/> si quieres tener una idea más detallada del propósito de cada uno de los servicios de AWS.

Si quieres una explicación complementaria sobre qué es AWS, te aconsejo que revises la web “¿Qué es la informática en la nube?”, disponible en: https://aws.amazon.com/es/what-is-cloud-computing/?nc1=h_ls

Como habrás podido observar, la cantidad de servicios que ofrece AWS para el soporte a la ejecución de aplicaciones en la nube es abrumador. En efecto, AWS ofrece un catálogo de servicios que cubre todo el espectro de modelos de servicio (IaaS, PaaS y FaaS).

AWS opera en diferentes regiones a lo largo del mundo, aunque no todos los servicios se ofrecen en todas las regiones. El hecho de disponer de una infraestructura global, utilizable por cualquier usuario, permite la construcción de aplicaciones

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

distribuidas, que toleren fallos ocurridos en una región, además de poder acercar tanto el cómputo como los datos a los usuarios, según su ubicación geográfica. A continuación, se enumeran algunas de las regiones disponibles, aunque la información más actualizada siempre está disponible en [12] y el mapa de regiones está disponible en [113]:

- US East (Ohio)
- EE. UU. Este (Norte de Virginia)
- EE.UU. Oeste (Norte de California)
- EE. UU. Oeste (Oregón)
- Africa (Cape Town)
- Asia Pacific (Hong Kong)
- Asia Pacific (Jakarta)
- Asia Pacific (Mumbai)
- Asia Pacific (Osaka)
- Asia Pacific (Seoul)
- Asia Pacifico (Singapur)
- Asia Pacifico (Sídney)
- Asia Pacifico (Tokio)
- Canada (Central)
- Europe (Frankfurt)
- Europa (Irlanda)
- Europe (London)
- Europe (Milan)
- Europe (Paris)
- Europe (Stockholm)
- Middle East (Bahrain)
- América del Sur (São Paulo)
- GovCloud



→ Región para uso de Gobiernos estatales → +++ seguridad en los datos
Ningún dato puede ser filtrado



Actividad: La Misteriosa Región GovCloud

Todas las regiones anteriormente mencionadas se corresponden con ubicaciones geográficas a lo largo del mundo salvo la última (GovCloud). ¿A qué crees que corresponde esta región? Puedes averiguar más información sobre la región GovCloud en el siguiente enlace (en inglés). <https://aws.amazon.com/es/govcloud-us/>

Además de las regiones existen formas de entregar aplicaciones con una latencia ultrabaja para dispositivos 5G mediante AWS Wavelength [114]. También existen zonas locales (AWS Local Zones) [115] para ejecutar aplicaciones sensibles a la latencia más cerca de los usuarios finales.

No todos los servicios de AWS están disponibles actualmente en todas las regiones (aunque es previsible que, eventualmente, lleguen a homogeneizarse). Puedes encontrar un listado de los servicios disponibles en cada región en [44].

De entre todos los servicios que ofrece AWS, en esta unidad se abordan los siguientes servicios:

- Amazon S3 (Simple Storage Service)
- Amazon EC2 (Elastic Compute Cloud)
- Amazon Auto Scaling

2. Resumen del servicio Amazon S3



Vídeo-lección: Introducción a Amazon S3

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

Este vídeo realiza una introducción al servicio **Amazon S3 (Simple Storage Service)**, para el **almacenamiento de ficheros** en la nube de AWS. Se aborda el concepto de *bucket* y los esquemas de redundancia ofrecidos. Se tratan las herramientas de interacción con S3. También se menciona el servicio Amazon Glacier para datos perdurables con restricciones en el tiempo de acceso a los datos.

Disponible en: <https://media.upv.es/player/?id=00028b60-021e-11e6-851a-656f7e06a374&autoplay=true>

Duración: 9'24"

Errata: En la última transparencia del vídeo se dice que "El servicio Amazon Glacier ofrece mejores precios a costa de reducir el tiempo de acceso a los datos" cuando debería decir "a costa de aumentar el tiempo de acceso a los datos".

Modificación: En el vídeo se indica que se trabajará con la herramienta "aws" de Tim Kay [13] pero actualmente utilizamos en el curso la herramienta oficial de línea de comandos de AWS (AWS CLI) [49]. De hecho, únicamente utilizaremos la herramienta de Tim Kay para interactuar con Amazon SimpleDB ya que todavía no hay soporte no experimental desde AWS CLI. Esta modificación es extensible para todos los vídeos donde se mencione dicha herramienta.

Amazon S3 [4] permite almacenar ficheros en la nube (de hasta 5 Tbytes cada uno) para que puedan ser accedidos por medio de protocolos estándar (HTTP, Bittorrent, etc.). Los ficheros se almacenan en buckets (o depósitos, según la traducción oficial de Amazon) que van ligados a una región concreta, para que los ficheros/datos no salgan de dicha región¹. Es posible llevar un control de acceso y permisos de acceso al bucket.

Los ficheros en Amazon S3 se replican en diferentes servidores de manera que puedan sobrevivir hasta a dos fallos en dos instalaciones diferentes (con una durabilidad del 99.99999999%). También es posible almacenar los ficheros con almacenamiento de redundancia reducida (Reduced Redundancy Storage) para tener menos réplicas de los datos pero pagando un precio más bajo (durabilidad del 99.99%).

Las **actualizaciones de los ficheros son siempre atómicas**, por lo que si una operación no termina satisfactoriamente no deja los datos en un estado inconsistente. Además, **S3 ofrece consistencia fuerte**, por lo que los cambios realizados serán inmediatamente visibles. *[CONSISTENCY - AVAILABILITY - PARTITIONS]*

Es posible interactuar con Amazon S3 desde la AWS Management Console, una herramienta web que permite gestionar los buckets, así como subir y descargar ficheros. También se puede utilizar alguna herramienta de línea de comandos como por ejemplo *aws* [13], una utilidad que consta de un único script en Perl, por lo que es fácilmente instalable en cualquier equipo. Sin embargo, también está disponible la interfaz oficial de línea de comandos de AWS, llamada AWS CLI [36], que es la herramienta que se utilizará principalmente en las prácticas para interactuar con los servicios de AWS. **Los ficheros en un bucket pueden almacenarse como públicos para que puedan ser accedidos por cualquiera mediante protocolo http.** Sin embargo, **también es posible** indicar listas de **control de acceso (ACL)** [82] que restrinjan el acceso a los mismos. En ese caso, es necesario estar autenticado y autorizado para acceder a los ficheros, para ello, es necesario indicar el Access Key ID y Secret Access Key en la configuración de la herramienta cliente utilizada para acceder a dichos ficheros.

Los esquemas de precios en AWS pueden cambiar a lo largo del tiempo (y también cambian según la región), por lo que siempre es recomendable acudir directamente a la fuente para obtener la información más actualizada de los precios de Amazon S3 [14]. A continuación, tienes algunos datos de referencia:

- Costes de almacenamiento:
 - 0.095 \$ por GB (estándar), y 0.080\$ por GB (redundancia reducida)
 - Los precios por GB bajan conforme aumenta el almacenamiento
- Costes de acceso
 - 0.01\$ por 1000 peticiones (PUT, COPY, POST, LIST)
- Coste por transferencia
 - Sin cargo por subir datos al Cloud
 - Descargar el primer GB / mes no cuesta dinero
 - 0.120\$ por GB (hasta 10 TB/mes)
 - Precio por GB baja conforme aumenta las transferencias
 - Sin coste de transferencias entre instancias de la misma región

¹ Esto es relevante para poder cumplir ciertas normativas, como por ejemplo que los datos almacenados en Europa no salgan de Europa.

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

Es posible acelerar el acceso a datos en una aplicación distribuida mediante el uso del servicio de entrega de contenido Amazon CloudFront [28]. Este es un servicio que permite la copia automática de datos a una serie de ubicaciones de borde distribuidas geográficamente a lo largo del mundo de manera que los clientes, distribuidos a lo largo del mundo, siempre acceden a la copia de los datos más cercana a su ubicación. Por ejemplo, si una distribuidora quiere promocionar el tráiler de una nueva película, puede utilizar Amazon CloudFront para que dicho fichero se transfiera automáticamente a dichas ubicaciones de borde. Esto permite que una persona en Tokio acceda automáticamente a la copia más cercana de dicho tráiler, reduciendo así la latencia y aumentando el ancho de banda en el acceso a dicho fichero. Actualmente existen más de 300 ubicaciones de borde a lo largo del mundo [29] [30].

? Pregunta: Coste de Subida de Datos al Cloud

Amazon S3 no cobra por subir los datos al Cloud (sí cobra por las descargas de ficheros). ¿A qué crees que es debido que no se cobre las subidas?

Amazon S3 no cobra por subir datos al Cloud porque el objetivo de AWS es atraer numerosos clientes a su Cloud. La idea es que si subes datos al Cloud luego los utilizarás, por lo que probablemente utilizarás algún otro servicio de AWS por el que sí se te cobrará.

Amazon S3 también soporta replicación automática entre regiones (Cross-Region Replication) [71] de objetos situados en un bucket, de forma asíncrona. Para ello, hay que añadir una configuración de replicación al bucket de origen en el que se indican qué objetos deben ser replicados (se puede indicar un patrón de nomenclatura para que únicamente afecte, por ejemplo, a un subconjunto de los objetos guardados en el bucket). Esta característica puede ser interesante para obtener un nivel de tolerancia a fallos mayor, al introducir mayor distancia entre las diferentes réplicas o para reducir la latencia de acceso a los ficheros, al poder disponer los ficheros en diferentes regiones, acercando así los ficheros a las ubicaciones de los posibles clientes.

Existen varias clases de almacenamiento de Amazon S3 [78], todas ellas con un SLA de durabilidad de 99.99999999%, si bien las más relevantes son:

- **Standard**
 - SLA de disponibilidad del 99,9% (esto significa que los datos pueden no estar accesibles durante unas 8.77 horas al año sin incumplir dicho SLA, tal y como se muestra en la Availability Calculator [109]).
 - Baja latencia y alto rendimiento → bajo retraso entre subida de datos y su disponibilidad
- **Standard - IA (Infrequent Access – Acceso Poco Frecuente)**
 - SLA de disponibilidad del 99%
 - Precio por GB reducido, pero se cobra 0.01\$ por GB en concepto de recuperación de datos
 - Los objetos menores de 128 KB se cobran como 128 KB de almacenamiento
 - Útil para almacenamiento a largo plazo, copias de seguridad, recuperación ante desastres en las que sea necesario acceder rápidamente a los datos
- **Glacier**
 - Almacenamiento a largo plazo de datos que no precisen acceso inmediato
 - La opción más barata
 - Requiere esperar 3-5 horas hasta recuperar los datos solicitados (aunque se puede realizar una recuperación Expedited con un coste superior y obtener los datos en 1-5 minutos)).
 - Útil para copias de seguridad de acceso infrecuente, datos archivados históricos, etc.

Además, existe la posibilidad de definir reglas de transición automáticas para que los datos se muevan entre las diferentes clases de almacenamiento. Por ejemplo, se puede comenzar almacenando los datos en la clase Standard, moverlo a la clase Standard - IA pasados 30 días y, finalmente, depositarlos en Amazon Glacier una vez pasados 60 días. Independientemente del cambio de clase, los datos se siguen almacenando en el mismo bucket y se acceden con las mismas URLs, por lo que no es necesario introducir ningún cambio en las aplicaciones cliente para beneficiarse de la reducción en el coste de almacenamiento, tal y como se describe en [78].

Para copiar grandes volúmenes de datos a Amazon S3 es posible utilizar S3 Multi-Part Upload [92], que permite cargar un solo objeto como un conjunto de partes de manera que se combinan posteriormente en S3. También está el servicio AWS DataSync [93], un servicio de transferencia de ficheros que permite subir y descargar datos entre un sistema on-premises y

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

AWS (por ejemplo, entre Amazon S3 y un sistema de almacenamiento local basado en NFS). Por último, AWS Snowball [94] posibilita el transporte de datos físicos entre una organización y AWS, permitiendo copiar los datos en discos duros y mandándolos a AWS para conectarlos físicamente allí y realizar la carga de datos).

También es posible montar un bucket de S3 de forma remota y utilizarlo como si fuera un sistema de archivos utilizando *mountpoint-s3* [118], si bien no implementa toda la funcionalidad de un sistema de archivos POSIX.

Finalmente, es importante tener en cuenta los costes derivados de la descarga de datos, que pueden ser significativamente elevados para grandes transferencias de datos, tal y como le ocurrió a la NASA [101].

* + Actividad (Opcional): Búsqueda de Tesoros para Profundizar con Amazon S3 (Línea de Comandos)

Si quieres profundizar sobre las características de Amazon S3 como acceso público, metadatos, etiquetas, versionado, consultas SQL a ficheros en S3, URLs temporales, clases de almacenamiento, cifrado, etc. te recomiendo que realices el siguiente juego:

<http://s3game-level1.s3-website.us-east-2.amazonaws.com/>

Desarrollado por Vasily Pantyukhin, tendrás que utilizar la línea de comandos ofrecida por AWS CLI para profundizar en las características de Amazon S3 mencionadas. Deberás encontrar las pistas para poder avanzar hacia la siguiente fase. Hay 11 niveles y su resolución te puede llevar más de una hora, pero creo que es un buen ejercicio de aprendizaje, además de entretenido. Utiliza la máquina de prácticas para realizar el ejercicio, ya que tienes los permisos necesarios para acceder a los buckets involucrados en el juego (aunque estén en la región us-east-2). Se trata de una actividad opcional. No es fundamental pero sí recomendable.

3. Resumen del servicio Amazon EC2

+ Vídeo-lección: Introducción a Amazon EC2

Este vídeo realiza una introducción al servicio Amazon EC2 (Elastic Compute Cloud) para el despliegue de instancias (máquinas virtuales) de cómputo. Se abordan los principales conceptos como región, AMI, grupo de seguridad y par de claves. Se tratan los diferentes tipos de AMIs y se clasifican en función del coste y sus capacidades. Se cubre el servicio EBS de creación de dispositivos virtuales orientados a bloques. Finalmente, se proporciona información sobre el coste de Amazon EC2, incluyendo opciones de ahorro en la forma de instancias reservadas y también *spot instances*.

Disponible en: <https://media.upv.es/player/?id=01e729e0-021e-11e6-851a-656f7e06a374&autoplay=true>

Duración: 10'29"

Amazon EC2 (Elastic Compute Cloud) [1] es un servicio que proporciona recursos de cómputo en la nube. Permite el despliegue de imágenes de máquinas virtuales (llamadas AMI o Amazon Machine Image [15]) sobre instancias de tamaño predefinido para disponer de cómputo bajo demanda. Una AMI es como si fuera una imagen ISO (como un DVD) que contiene una configuración de hardware (virtual), la instalación de un Sistema Operativo (SO), unas aplicaciones y unos datos. Por lo tanto, es la instantánea de una determinada configuración de un equipo.

AWS utiliza el término *instancia* para referirse a una máquina virtual (MV), donde cada tipo de instancia puede tener un número diferente de CPUs virtuales (vCPUs), cantidad de memoria y capacidad y número de discos. Antiguamente se indicaba que cada tipo de instancia ofrecía una cantidad predecible de capacidad de cómputo dedicada medida en ECUs (Elastic Compute Unit), donde 1 ECU proporciona el equivalente a una CPU Opteron de 1.0-1.2 GHz 2007 o un Procesador Xeon de 2007. Sin embargo, desde Abril de 2014 se ha abandonado esta nomenclatura para usar únicamente vCPUs [77].

AWS dispone de centros de datos a lo largo del mundo, distribuidos en Regiones, tal y como se ha comentado anteriormente. Por ejemplo, AWS dispone de varias regiones en la Unión Europea (por ejemplo, en Irlanda y en Frankfurt),

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

varias en Estados Unidos de América, en Japón, en Brasil, etc. Esto permite tener una **distribución geográfica de los recursos** para cumplir **dos objetivos básicos**. En **primer** lugar, por **tolerancia a fallos** ya que los fallos de hardware en una región no afectan a otras regiones. En **segundo** lugar, para **acercar las aplicaciones y los datos a los usuarios**. Una región consta de múltiples zonas de disponibilidad que deben entenderse como diferentes centros de datos dentro de la misma región (para que un fallo en una zona de disponibilidad no afecte a los recursos de otra zona de disponibilidad, dentro de la misma región).

Para desplegar una instancia se debe especificar:

Instancia = Presta en marcha = Máquina Virtual funcionando en de una AMI la Cloud AWS

- **El tipo de instancia**
 - Determina las **prestaciones de la máquina virtual** (instancia), **en términos de cantidad de memoria RAM, número de núcleos (cores) virtuales, espacio de almacenamiento en disco, arquitectura y rendimiento de E/S**. El precio por unidad de tiempo varía según las prestaciones del tipo de instancia elegida y la región elegida.
- **La AMI** *→ tipo de "plantilla" con la conf. para lanzar una instancia*
 - Se trata de una **imagen de máquina virtual**, que encapsula un Sistema Operativo (SO), unas aplicaciones y unos datos.
- **Grupo de seguridad (Security Group)**
 - La **configuración de cortafuegos de la instancia** (qué tráfico puede recibir la instancia (TCP, UDP) y a qué puerto).
- **Par de claves (Key Pair)** *→ Clave privada + Clave pública*
 - Permite la **conexión a una cuenta** de usuario de una instancia GNU/Linux mediante SSH **sin tener que especificar la contraseña**.
- **La región o zona de disponibilidad** (Opcional)
 - AWS dispone de centros de datos a lo largo del mundo por lo que interesa escoger aquella región más cercana al usuario. **La región por defecto es us-east-1** (en Virginia, USA).

+ Actividad: Tipos de instancias de Amazon EC2

Existen **numerosos tipos de instancias de cómputo**, que **determinan las capacidades de una máquina virtual** (instancia) en ejecución sobre Amazon EC2. Revisa la siguiente web para conocer más detalles sobre los tipos de instancias:

<http://aws.amazon.com/es/ec2/instance-types/>

Tienes además una descripción detallada sobre cómo elegir el tipo de instancia más apropiado en función del tipo de aplicación en [83].

Por ejemplo, es posible desplegar una AMI con Ubuntu 20.04 y un servidor web Apache configurado para iniciarse en el arranque de la instancia) sobre una instancia de tipo *t3.small* (que consta de 2 GiB de RAM, 2 vCPUs) en la zona de disponibilidad us-east-1e (una de las zonas de disponibilidad de la región de Estados Unidos de América, en la zona de Virginia).

El coste del servicio se establece en un precio por segundo [91] dependiente tanto del tipo de instancia elegido (a mayores prestaciones mayor coste) como del número de segundos de uso.

Es importante conocer que hay **dos tipos de AMIs**:

- **Basadas en S3** (denominadas *instance-store*)
 - Cualquier **fichero creado o modificado** en la instancia **desaparecerá** cuando se apague la máquina virtual. Por tanto, el almacenamiento es efímero.
- **Basadas en EBS** (denominadas *EBS*)
 - El dispositivo raíz ya **no es efímero**. Los **cambios** en la instancia **se almacenan en un volumen EBS asociado a la instancia**. La instancia puede detenerse por lo que dejará de generar un coste por segundos de uso, pero sí que **generará coste por el almacenamiento del volumen EBS asociado**.

S3
- Efímero
- Coste de uso

EBS
- No efímero
- Coste de uso
- Coste de almacenamiento - tamaños Disco ++

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

Las AMI basadas en EBS presentan muchas más ventajas que las AMIs basadas en S3 y es la opción preferida hoy en día. Salvo contadas excepciones (por ejemplo, si quieres construir una AMI de pago [26]) siempre debes elegir AMIs basadas en EBS. Por ejemplo, las instancias de AMIs basadas en EBS pueden detenerse y volver a ser iniciadas sin perder los archivos que se hayan creado en la instancia (puesto que estos se almacenan en un volumen EBS). También es posible modificar el tipo de instancia una vez detenida, para incrementar sus prestaciones y capacidades. Además, permiten tamaños más grandes de discos. Por ejemplo, las AMIs de Windows siempre son AMIs basadas en EBS, ya que la instalación del Sistema Operativo ya requiere prácticamente los 10 GBytes de tamaño, que es el límite de una AMI basada en S3 (instance-store). Puedes encontrar más ventajas de las AMIs basadas en EBS en [25] y en [39].

Es posible que te plantees la posibilidad de transformar una AMI basada en S3 a una AMI basada en EBS (o viceversa). Aunque técnicamente es posible (ver [40], [41] ó [42]), casi siempre resulta mucho más operativo partir de una AMI ya existente (del tipo deseado) y desplegar las aplicaciones y los datos sobre una instancia en ejecución desplegada a partir de dicha AMI usando servicios de configuración automatizada como CloudFormation u OpsWorks o incluso herramientas de DevOps como Ansible, Puppet o Chef.

AWS utiliza varias credenciales de acceso para el uso de su infraestructura por lo que es importante saber diferenciarlas:

- **Claves de Acceso** (Access Key ID y Secret Access Key)
 - Sirven para autenticar al usuario ante los servicios de AWS. Es la opción preferida y la que utilizan las herramientas cliente más recientes para acceder a AWS.
- **Certificados X.509**
 - Sirven para autenticar al usuario ante los servicios de AWS. Hay menos herramientas que los utilicen, pero las herramientas oficiales de acceso a EC2 (EC2 API Tools [6]) los usan como mecanismo de autenticación.
- **Pares de Clave (Key Pairs)**
 - Los crea el usuario de manera que AWS guarda la clave pública y el usuario guarda la clave privada. Cuando una instancia de EC2 arranca, AWS introduce en ella la clave pública. Esto permite al usuario acceder mediante un cliente SSH a la instancia (basada en GNU/Linux) con una determinada cuenta de usuario sin necesidad de especificar la contraseña.
- **Usuario y Contraseña**
 - Se utilizan para poder acceder a la consola de administración de AWS (AWS Management Console [16]), una herramienta web para utilizar la mayoría de los servicios de AWS.

Es posible exportar una AMI de manera que quede almacenada en un bucket de S3 para ser posteriormente descargada y ejecutada en los recursos de una organización. Para ello es necesario seguir el proceso descrito en [31].

En Amazon EC2 es posible desplegar instancias bajo demanda (on-demand instances) pero también se pueden pre-comprar instancias a 1 y 3 años vista de manera que el precio por segundo se reduce significativamente tras desembolsar un pago inicial. Para cargas de trabajo estables a varios meses vista puede ser una opción interesante para reducir el coste de la infraestructura. Estas instancias se conocen como instancias reservadas (reserved instances) y permiten realizar un desembolso inicial para conseguir un precio por segundo significativamente inferior al de las instancias bajo demanda, pudiendo llegar a conseguir descuentos de hasta el 75% en comparación con el precio de las instancias bajo demanda (ver [34] para más detalles). Además, también existe la posibilidad de revender parte de las instancias reservadas en un mercado de instancias reservadas operado por AWS [35]. Esto permite recuperar parte de la inversión si en realidad no se aprovechan las instancias reservadas adquiridas inicialmente.



Práctica: Despliegue de Instancias de Máquinas Virtuales con Amazon EC2

En este punto, debes realizar la práctica “Despliegue de Instancias de Máquinas Virtuales con Amazon EC2” (Práctica básica de EC2). En el boletín de la práctica tienes toda la información detallada que te hace falta para seguirla. En ella aprenderás a utilizar el servicio Amazon EC2 para aprovisionar máquinas virtuales y te familiarizarás con los conceptos anteriormente tratados (regiones, zonas de disponibilidad, pares de clave, etc.).

Habrás recibido por correo electrónico o mediante un mensaje en el Foro o en el Chat de PoliformaT la máquina a la que debes conectarte para poder realizar la práctica. De lo contrario, contacta con el profesor vía correo electrónico.

Más adelante en la unidad se te planteará que hagas otra práctica en la que utilizarás funcionalidad avanzada (creación de

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

volúmenes EBS, creación de grupos de auto-escalado, etc.)

Existen dos versiones de dicha práctica:

- Versión AWS Management Console: Esta es la versión recomendada. Esta versión está orientada para aquellos alumnos que tengan dificultades con el manejo de línea de comandos en Linux. Permite usar principalmente la interfaz web de administración de los servicios de AWS (AWS Management Console) mediante un navegador web para interaccionar con dichos servicios. También permite conocer algunos aspectos de configuración adicionales de los servicios empleados.
- Versión AWS CLI: En esta versión de la práctica se utiliza la interfaz de línea de comandos oficial de AWS (AWS CLI) para interaccionar con los servicios de AWS.

Los resultados de aprendizaje de ambas prácticas son equivalentes. Se plantean las mismas actividades, pero se resuelven usando herramientas diferentes. Por ello, la idea es que realices una única versión de la práctica (aquella en la que te vayas a sentir más cómodo). En cualquier caso, siempre puedes consultar la otra versión de la práctica para conocer cómo se realiza determinada acción con otra herramienta alternativa. Si tienes tiempo y deseas hacer las dos versiones de la práctica, no hay ningún problema en ello.

Como habrás podido comprobar, Amazon EC2 soporta diferentes tipos de instancia (*t2.nano*, *t2.micro*, *m1.small*, etc.) con diferentes prestaciones y un determinado coste por segundo asociado. Quizá tengas la duda de cuál es el mejor tipo de instancia a elegir en cada momento. Para ello, lo más razonable es conocer los requisitos de consumo de recursos (memoria, procesador, disco) de la aplicación que se va a ejecutar dentro de la instancia para ajustar las necesidades de la aplicación a las prestaciones que ofrece la instancia. Para ello, se pueden utilizar técnicas de perfilado de aplicaciones (*application profiling*) para conocer dicho consumo de recursos [64].

Vídeo-lección: Otros Servicios de Amazon EC2 para Aplicaciones Cloud

Este vídeo describe otros sub-servicios incluidos en el servicio Amazon EC2 que posibilitan la creación en la nube. Se aborda el mecanismo de instantáneas proporcionado por EBS (*Elastic Block Store*), como aproximación a copias de seguridad, pero también para la compartición de conjuntos de datos en la nube. Se trata el servicio ELB (*Elastic Load Balancing*) para la creación de balanceadores de carga. También se habla sobre IPs elásticas.

Disponible en: <https://media.upv.es/player/?id=02ee39a0-021e-11e6-851a-656f7e06a374&autoplay=true>

Duración: 7'43"

ACTUALIZACIÓN: AWS incrementa periódicamente los límites soportados por sus servicios. En particular, el límite de 1 TB de volumen EBS que se indica durante el vídeo ha sido superado. En el vídeo se describen los balanceadores de carga clásicos (*Classic Load Balancers*) [84]. Sin embargo, recientemente ELB también incluye soporte para balanceadores de carga de aplicación (*Application Load Balancers*) [85], que serán usados en las prácticas más adelante.

Es posible asociar una IP elástica [55] a una instancia en ejecución de Amazon EC2. Esto permite detener la instancia, volver a arrancarla más tarde y asignar nuevamente la IP elástica a la instancia para que mantenga la misma IP pública. Esto permite que los clientes de dicha instancia no vean modificada la IP de acceso a la misma. También permitiría asociar dicha IP elástica a una nueva instancia, sin necesidad de que los clientes deban realizar ninguna modificación.

Sin embargo, es importante que sepas que, al asignar una IP elástica a una instancia, la IP pública y el nombre DNS que tenía antes de dicha asignación cambian para pasar a ser los correspondientes a la IP elástica y un nombre DNS relativo a dicha IP elástica. Detener una instancia implica desasignar la IP elástica (salvo que la instancia de EC2 esté desplegada en un VPC), aunque puede volver a ser asociada nuevamente por el usuario cuando la instancia vuelva a iniciarse. Por último, dada la escasez, las IPs públicas (IPv4) tienen un cierto coste, por lo que es importante identificar y optimizar el uso de estas direcciones IPs en AWS [119].

4. Resumen del servicio EBS (Elastic Block Store)

EBS (Elastic Block Store) [5], dentro del servicio Amazon EC2, proporciona volúmenes orientados a bloques [57] (como si fuera un disco duro o un disco USB) para ser conectados dinámicamente a instancias de EC2. Los volúmenes EBS se crean para una zona de disponibilidad concreta y sólo pueden conectarse a instancias desplegadas en la misma zona de disponibilidad. Es posible conectar varios volúmenes a una instancia, pero un volumen EBS solo puede ser conectado a una única instancia en un momento dado. Sí que es posible desconectar y conectar el volumen EBS de una instancia a otra (por ejemplo, para compartir datos).

Un volumen EBS debe ser particionado y formateado antes de poder utilizarlo desde el SO de la instancia y su tamaño debe estar comprendido entre 1 GB y 16 TB. Es posible superar el límite máximo usando RAID 0 (o LVM) a través de múltiples volúmenes, lo que además permite mejorar las prestaciones de E/S. Es posible obtener más información sobre esta aproximación en [17].

EBS proporciona actualmente tres tipos de volúmenes: Los volúmenes de propósito general basados en SSD (*General Purpose (SSD) Volumes*), volúmenes con aprovisionamiento de operaciones de entrada/salida (*Provisioned IOPS (SSD) Volumes*) y volúmenes basados en discos magnéticos (*Magnetic Volumes*). IOPS es la abreviatura de *Input/Output Operations per Second*, es decir, operaciones de entrada/salida por segundos. Esto permite escalar las prestaciones de un volumen de forma predecible indicando un nivel de prestaciones específico, medido en IOPS, y pudiendo llegar hasta 20.000 IOPS. Puedes encontrar más información sobre esta característica en [46]. De hecho, se puede llegar hasta 48.000 IOPS uniendo varios volúmenes EBS en configuración RAID, tal y como se explica en [50]. Es posible conocer el número de operaciones por segundo (*transfers per second*) desde dentro de una instancia con la herramienta *iostat* [47], lo que puede ayudar a estimar las prestaciones del volumen EBS necesarias para soportar a una aplicación.

La opción más económica es usar los volúmenes basados en discos magnéticos mientras que la opción que ofrece más prestaciones a mayor precio son los volúmenes con IOPS aprovisionadas. Tienes más información sobre los diferentes precios de volúmenes EBS en [51] y sobre las características y prestaciones específicas de cada tipo de volumen en [52].

Se pueden hacer copias de seguridad de un volumen EBS mediante la creación de *instantáneas (snapshots)* [107], que es el mecanismo de copia de seguridad incremental que ofrece EBS, lo que significa que solo se guardan los bloques que han cambiado en el dispositivo después de la instantánea más reciente. Esto disminuye el tiempo necesario para crearlo y ahorra costos de almacenamiento, ya que no se duplican los datos. Cada instantánea contiene toda la información necesaria para restaurar los datos (del momento en que se tomó) en un volumen de EBS nuevo. Cuando se elimina una instantánea, solo se borran los datos que son únicos de dicha instantánea [108].

Aunque es un detalle de optimización, conviene que sepas que existe una penalización en las prestaciones de primera lectura de volúmenes EBS creados a partir de una instantánea (*snapshot*). Por ello, si no deseas que la aplicación incurra en dicha penalización (que solo ocurriría en la primera lectura de cada bloque del volumen), se puede forzar la lectura previa de todos los bloques del volumen, antes de poner en marcha la aplicación que vaya a usar dicho volumen [20]. Alternativamente, es posible activar la opción de "Fast Snapshot Restore" [110] y para un snapshot concreto, indicar en qué AZs (*Availability Zones*, Zonas de Disponibilidad) se crearán volúmenes a partir de dicho snapshot. De esta forma, el volumen creado soportará las máximas prestaciones. A cambio, se incurre en un coste horario por snapshot y por el tiempo en que dicha característica está activada.

Ten en cuenta que como un volumen EBS tradicional únicamente puede estar conectado a una única instancia en cada momento, no es posible utilizarlo como un sistema de archivos compartido de lectura/escritura entre múltiples instancias. Si necesitas tener un sistema de archivos compartido entre varias instancias existen varias alternativas que son dependientes de las necesidades de las aplicaciones que se vayan a ejecutar. Por ejemplo, si únicamente necesitas compartir un volumen en modo solo lectura entre todas las instancias, existe EBS Multi-Attach [116], que posibilita compartir un conjunto de datos entre diferentes instancias. Si necesitas que el sistema de archivos esté compartido en modo lectura/escritura entre todas las instancias, existen herramientas como NFS [65], Gluster [66], OrangeFS [67] o Ceph [68]. De hecho, el servicio Amazon EFS [117] permite crear sistemas de archivos NFS de forma totalmente gestionada por el proveedor. También se podría tratar de externalizar el almacenamiento de ficheros a Amazon S3 o, en función de la aplicación, sacar el almacenamiento de datos a una base de datos externa (en caso de datos no basados en ficheros).

Para ampliar un volumen EBS es posible hacer un snapshot y luego restaurarlo sobre un volumen de tamaño mayor y, al conectarlo nuevamente a la instancia, será necesario extender el sistema de archivos para poder aprovechar el nuevo tamaño del volumen, tal y como se indica en [89]. También se puede redimensionar un volumen de forma dinámica, para algunos tipos de volúmenes, sin necesidad de realizar un snapshot previo, en función del tipo de instancia utilizado [90], si

bien será necesario redimensionar la partición (o crear una nueva partición en el nuevo espacio) para poder aprovechar el tamaño ampliado. Se puede reemplazar el volumen EBS de arranque de una instancia, ¡incluso sin detenerla! [112].

5. Resumen del servicio ELB (Elastic Load Balancing)

ELB (Elastic Load Balancing) [7], dentro del servicio Amazon EC2, permite la construcción de balanceadores de carga para distribuir las peticiones dirigidas a un determinado puerto del balanceador entre diferentes instancias de distintas zonas de disponibilidad (ver Figura 2) dentro de la misma región. Esto permite realizar un adecuado reparto de la carga para que múltiples instancias puedan atender las peticiones. Los balanceadores de carga están internamente replicados y son gestionados por AWS por lo que no constituyen un punto único de fallo [103]. En casos excepcionales de aplicaciones web que requieran una replicación en múltiples regiones, puedes usar balanceo de carga a nivel de DNS (usando Route 53) para repartir peticiones entre diferentes instancias de balanceadores de carga [104] que puedan estar ubicados en múltiples regiones.

El caso más típico suele ser disponer de un balanceador de carga tras el cual se agrupan diferentes instancias en varias zonas de disponibilidad (con un reparto equitativo de instancias entre zonas de disponibilidad) de la misma región. Los clientes se conectan al puerto 80 del balanceador de carga y éste reenvía cada petición al puerto 80 de una instancia, para que atienda dicha petición. De esta manera, las peticiones se distribuyen entre las diferentes instancias, que pueden absorber mayor carga de trabajo que una única instancia. Lo lógico es que las instancias sean réplicas de la misma aplicación. De esta manera, no importa qué instancia procesa cada petición http.

Un balanceador ELB clásico (*Classic Load Balancer* [84]) puede repartir peticiones entre instancias en la misma zona de disponibilidad y entre diferentes zonas de disponibilidad, pero no entre diferentes regiones². Puede monitorizar el estado de las instancias (conectándose periódicamente a un puerto de las instancias) para descartar aquellas instancias inaccesibles (que volverán a ser integradas en el balanceador cuando vuelvan a estar disponibles). Además, una misma instancia puede estar registrada en diferentes balanceadores de carga [45].

Ten en cuenta que en el momento de despliegue de una instancia de EC2 se le puede indicar la zona de disponibilidad en la que se desea desplegarla. Si no se especifica, EC2 elige una zona de disponibilidad. Puedes averiguar la zona de disponibilidad en la que se está ejecutando una instancia desde la consola de administración de AWS (campo *availability zone*) y desde AWS CLI, revisando las propiedades de la instancia.

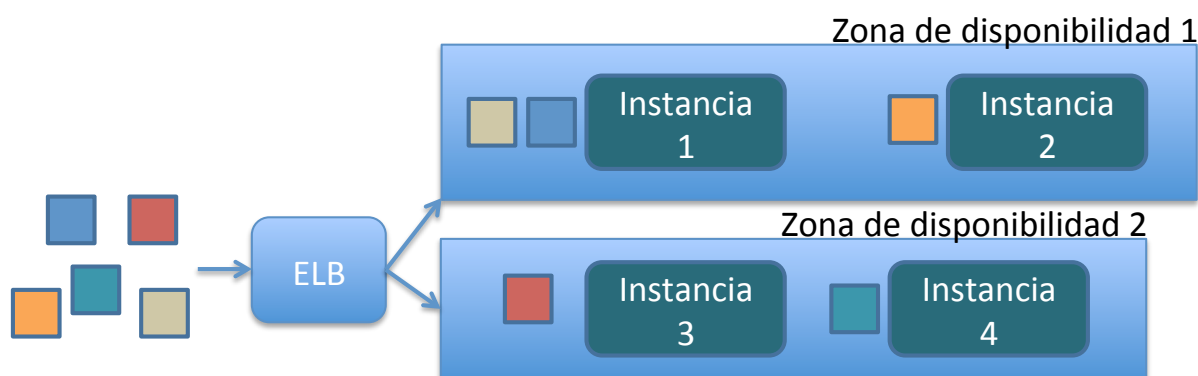


Figura 2. Esquema de funcionamiento de un balanceador de carga creado con ELB.

Un balanceador ELB usa por defecto un esquema *round-robin* [18] para repartir las peticiones recibidas entre zonas de disponibilidad (una petición a cada zona de disponibilidad, de forma circular). Por tanto, se recomienda que las instancias estén equitativamente distribuidas entre las diferentes zonas de disponibilidad. ELB no monitoriza la carga de la CPU de las instancias para dirigir las peticiones a las menos cargadas. En noviembre de 2019 se introdujo soporte para el algoritmo LOR (*Least Outstanding Requests*) [100] con el objetivo de realizar una mejor distribución de la carga para aquellas peticiones que puedan requerir más tiempo de proceso que otras. Con este algoritmo los nodos destino que están más cargados no se les sobrecarga con peticiones adicionales.

² Si quieres balancear peticiones entre diferentes regiones puedes utilizar HAProxy. <http://haproxy.1wt.eu>

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

ELB también soporta *sticky sessions*, de manera que las peticiones de un mismo cliente siempre se dirijan a la misma instancia de aplicación, en lugar de que cada petición vaya a una instancia diferente. Esto puede ser interesante en una aplicación de comercio electrónico, donde el proceso de compra requiere realizar una transacción que involucra múltiples peticiones que deberían ser atendidas por la misma instancia de aplicación. Para ello, hay que utilizar cookies de sesión, tal y como se explica en [95].

ELB soporta una característica denominada *cross-zone load balancing* [24], que permite que el balanceador de carga distribuya el tráfico de forma equitativa entre las instancias (en lugar de equitativa entre zonas de disponibilidad), independientemente del número de instancias desplegadas en cada zona de disponibilidad. Esta característica evita tener que mantener un número equitativo de instancias en cada zona de disponibilidad y mejora la capacidad de una aplicación para gestionar la pérdida de una o más instancias. Sin embargo, todavía se recomienda mantener un número equilibrado de instancias en cada zona para disponer de mayor tolerancia a fallos (por ejemplo ante una caída de una zona de disponibilidad completa).

ELB soporta una característica denominada *connection draining* [37] [38] que permite a las instancias que están registradas en el balanceador dejar de recibir nuevas peticiones, pero seguir estando activas durante un determinado número de segundos (configurable para cada balanceador) hasta que finalicen las conexiones que están en progreso. Esto permite evitar cortes en la comunicación derivados de la desconexión de instancias (de-registro) del balanceador para, por ejemplo, transferencias de ficheros que pudieran estar activas (de la instancia de EC2 al cliente, a través del balanceador de carga).

Ten en cuenta que los balanceadores de carga creados con el servicio ELB no pueden ser detenidos (tan solo puedes destruirlos/terminarlos). Cuando creas un nuevo balanceador recibe una IP y un nombre DNS diferentes. Tampoco puedes asociar una IP elástica al balanceador. Por ello, puedes utilizar este servicio en combinación con un servicio de DNS (como Amazon Route 53) para establecer la correspondencia entre un nombre DNS y la dirección IP del balanceador de carga. Aunque Amazon no proporciona demasiados detalles sobre cómo están implementados sus servicios, en el caso de ELB hay que asumir que el balanceador puede absorber de forma elástica incrementos en el número de peticiones entrantes, de manera que el balanceador no se convierta en el cuello de botella.

De acuerdo con [58], el coste de ELB depende del número de horas en ejecución (las horas parciales se cobran como horas completas) y del número de GBs transferidos a través del balanceador de carga.

Además de los balanceadores clásicos, ELB también soporta los balanceadores de carga a nivel de aplicación (*Application Load Balancer* [85]), de manera que el balanceador de carga pueda tomar decisiones en base al contenido del tráfico de la aplicación en los mensajes http (no únicamente una redirección en base al puerto utilizado, tal y como ocurre en los balanceadores clásicos). Las principales ventajas de esta aproximación son, tal y como se describen en [85]:

- Soporte para enrutamiento en función de la URL especificada en la petición. Esto permite estructurar la aplicación como servicios más pequeños y reenviar las peticiones al servicio correcto en función del contenido de la URL.
 - Por ejemplo, las peticiones que lleven **/monitoring** en la URL pueden ser redirigidas a un grupo objetivo (*target group*).
- Soporte para enrutar peticiones a múltiples servicios en una misma instancia de EC2 (registrando la misma instancia de EC2, pero usando múltiples puertos de manera que cada servicio escuchará en un puerto diferente).
 - Esto es útil especialmente para aplicaciones containerizadas donde se despliegan múltiples contenedores en una misma máquina virtual y deben recibir tráfico de un balanceador de carga de aplicación.
- Soporte para monitorización de la salud de cada servicio de forma independiente.
- Útil para distribuir peticiones a aplicaciones basadas en microservicios [88] de manera que múltiples servicios se pueden estar ejecutando en la misma instancia de EC2.

Las diferencias entre los balanceadores de carga clásicos y los de nivel de aplicación están explicadas en [86]. Los balanceadores de carga clásicos trabajan a nivel 4 del modelo OSI (Open Systems Interconnection model) [87], de manera que no ven el contenido de los paquetes de red. Sin embargo, los balanceadores de carga de aplicación trabajan en el nivel 7 del modelo OSI, de manera que inspeccionan paquetes, tienen acceso a las cabeceras http y pueden realizar tareas más inteligentes de distribución de la carga. El coste horario de estos balanceadores es un 10% más barato que el de los balanceadores clásicos si bien también se cobra por el número de conexiones por segundo (más detalles en [86]).

Para casos puntuales de crecimiento muy rápido de tráfico que puede ser determinado de antemano es posible avisar a AWS para iniciar un proceso de pre-calentador (*warm up*) [96] del balanceador de carga, con el objetivo de obtener las mejores prestaciones. Existe una guía de buenas prácticas para evaluar un balanceador de carga ELB en [97].

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

Los ALB también soportan Grupos de Destino (Target Groups) [98] ponderados, para poder implementar fácilmente despliegues de tipo Blue/Green [99] con el objetivo de cambiar de forma progresiva el tráfico de una versión de la aplicación a otra más reciente, de forma que se puedan monitorizar métricas operacionales, y de negocio, al cambiar a la nueva versión. Aunque tradicionalmente esto se ha implementado usando balanceo de carga a nivel de DNS (y utilizando dos balanceadores de carga), actualmente esto se puede ya implementar exclusivamente usando ALBs, tal y como se muestra en la Figura 3, pudiendo definir la cantidad de tráfico que va a cada versión del despliegue. Tienes más información al respecto en [98].

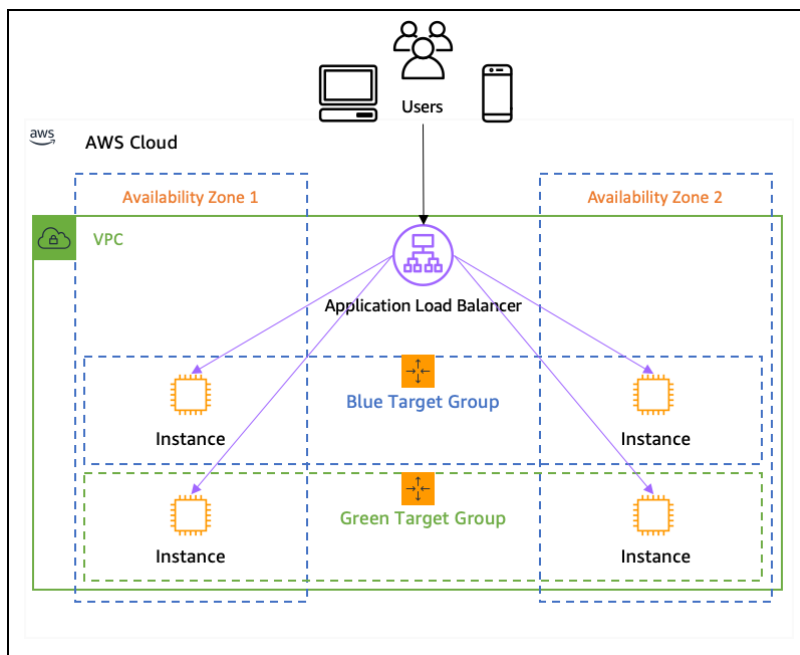


Figura 3. Implementación de despliegues Blue/Green con ALBs.

6. Resumen del servicio Auto Scaling

+ Vídeo-lección: Auto-Escalado de Aplicaciones Cloud

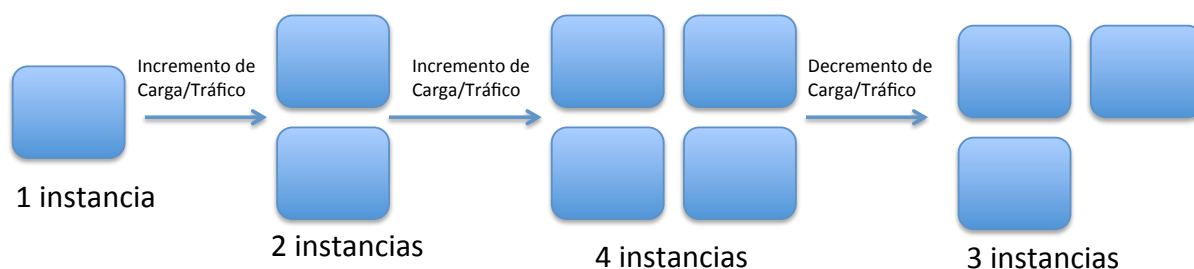
Este vídeo aborda el servicio Auto Scaling que permite la creación de flotas elásticas de instancias de EC2, cuyo número puede crecer y decrecer en función de ciertas reglas de escalado dependientes de la monitorización del estado de la flota. Esto permite la creación de aplicaciones con soporte a elasticidad horizontal en la nube.

Disponible en: <https://media.upv.es/player/?id=030e92e0-021e-11e6-851a-656f7e06a374&autoplay=true>

Duración: 7'33"

El servicio Auto Scaling [2] permite el crecimiento y decrecimiento de flotas de instancias flanqueadas por un balanceador ELB. Esto permite el escalado horizontal en base a reglas de aumento de la flota (*scale out*), así como reglas de reducción del tamaño de la flota (*scale in*). Esto permite ofrecer servicios de elasticidad horizontal a aplicaciones Cloud que experimentan patrones de uso variable a lo largo del tiempo.

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling



Para poder usar Auto Scaling es necesario conocer:

- Grupos de Auto-Escalado, que agrupan instancias bajo un ELB
- Configuraciones de Lanzamiento, que determinan qué tipo de instancia y AMI se debe lanzar.
- Alarmas, que determinan cuando una acción de escalado (denominada Política) debe ocurrir.
- Políticas, que indican si se deben desplegar o terminar instancias.

Mediante este servicio es posible definir reglas de elasticidad (aumento y reducción del tamaño de la flota de instancias) como, por ejemplo: Si el promedio de uso de CPU de las instancias de este grupo de auto-escalado supera el 70% durante al menos 2 periodos de 3 minutos, entonces incrementa el tamaño del grupo en 2 instancias adicionales de las especificadas en esta configuración de lanzamiento. Este servicio se utiliza en conjunción con el servicio Amazon CloudWatch [3] que permite obtener información de consumo de recursos de las instancias. Este servicio es gratuito para frecuencias de monitorización de 5 minutos de instancias de EC2, pero si se desea obtener información de monitorización más frecuente se incurre en gastos (ver precios en [69]). CloudWatch también permite monitorizar métricas relacionadas con las aplicaciones que se ejecutan dentro de las instancias, instalando un agente en las instancias EC2 para que publiquen los logs de la aplicación. Tienes más información al respecto en [75].

Además, también es posible definir reglas de escalado en función del calendario, de manera que una aplicación puede aumentar su capacidad de lunes a viernes y que el fin de semana repliegue algunas instancias para reducir el consumo. En [48] se describe la política de terminación de instancias por defecto, que incluye diversas estrategias como terminar instancias lo más cercanas a la siguiente hora de facturación (para minimizar el coste), aquellas con la configuración de lanzamiento más antigua (para que las nuevas instancias sean más actualizadas) y considerando primero aquellas cuya zona de disponibilidad cuenta con el mayor número de instancias. Esto permite reducir el consumo al tiempo que equilibra el número de instancias de las diferentes zonas de disponibilidad.

La elasticidad es una de las características principales de las plataformas Cloud dado que permite adaptar las capacidades de una aplicación a los requisitos de cómputo de la misma. Sin embargo, no siempre es conveniente que una aplicación escale ante un aumento de los requisitos de cómputo. Por ejemplo, ante un ataque de denegación de servicio económico (EDoS – Economic Denial of Service), un servicio Cloud puede recibir un incremento elevado de peticiones al servicio proveniente de falsos clientes, para forzar su escalado automático, provocando un aumento de los gastos del proveedor del servicio. Ante una situación de peticiones maliciosas, la aplicación no debería escalar.

Por otra parte, existen una serie de escenarios de uso que el auto-escalado no resuelve, tal y como se describen en [72]. Por ejemplo, este es el caso de un incremento rápido de la demanda. El despliegue de una instancia de EC2 seguido de la configuración automatizada para desempeñar un cierto rol dentro de una arquitectura de aplicación requiere un tiempo que puede ser del orden de decenas de minutos. Si la carga de trabajo sigue aumentando (por ejemplo, el número de peticiones de acceso a un servicio), entonces se puede producir denegación de servicio. Para resolver este tipo de limitaciones existen desarrollos centrados en realizar auto-escalado predictivo, como es el caso de la herramienta de código abierto Scryer de Netflix [72].

De hecho, el servicio Auto Scaling soporta *Step Scaling Policies* [73], además de las políticas de escalado tradicionales (*Simple Scaling Policies*) de manera que las políticas de escalado definidas continúan respondiendo a las alarmas incluso cuando todavía se está realizando una actividad de escalado, obviando así el periodo de *cooldown* que soportan las políticas de escalado tradicionales. Esto permite realizar una evaluación continua de las alarmas y no bloquear el grupo de auto-escalado durante estas operaciones, permitiendo así una rápida adaptación a las condiciones cambiantes de carga de trabajo. También existen unas políticas de escalado de seguimiento de destino (*Target Tracking Policy*) [102] que permite ajustar el tamaño del grupo de auto-escalado para tratar de cumplir una determinada métrica, como que el promedio de uso

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

de CPU esté situado entorno a 70%. Esto implicará desplegar instancias adicionales si el valor de dicha métrica sube por encima de dicho valor, desplegando un número de instancias adicionales en función de la discrepancia entre el valor objetivo y el valor actual de la métrica. Alternativamente, se eliminan instancias cuando dicho valor baja del valor objetivo.

Finalmente, he de comentar que también es posible el uso del servicio de Auto Scaling también desde la AWS Management Console, dentro de la consola de administración del servicio EC2. Esto facilita la creación de configuraciones de lanzamiento y grupos de auto-escalado a través de una interfaz gráfica, y no únicamente mediante las herramientas de línea de comandos.

+ Vídeo-lección: Utilizando Amazon Web Services (AWS)

Este vídeo indica los pasos necesarios para comenzar a utilizar Amazon Web Services (AWS). Se introduce la capa de uso gratuita de AWS y se explican las diferentes credenciales de acceso utilizables en AWS. Se menciona también tanto la calculadora de AWS, para calcular el coste derivado del uso del Cloud, en función de los servicios y la tasa de uso de los mismos.

Disponible en: <https://media.upv.es/player/?id=027a6930-021e-11e6-851a-656f7e06a374&autoplay=true>

Duración: 9'42"

+ Práctica: Despliegue de Grupos Elásticos de Instancias y Gestión de Datos con AWS

En este punto de la unidad debes realizar la práctica "Despliegue de Grupos Elásticos de Instancias y Gestión de Datos con Amazon Web Services (Práctica avanzada de Amazon EC2, Amazon S3 y Auto Scaling)". En ella verás aspectos avanzados del servicio Amazon EC2, como la creación de flotas elásticas de instancias usando grupos de auto-escalado y también utilizarás el servicio S3 para el almacenamiento de ficheros en la nube. Con esta práctica aprenderás a crear aplicaciones elásticas que crezcan dinámicamente en función de la carga de trabajo.

Existen dos versiones de dicha práctica:

- Versión AWS Management Console: Esta versión es la recomendada y está orientada para aquellos alumnos que tengan dificultades con el manejo de línea de comandos en Linux. Permite usar principalmente la interfaz web de administración de los servicios de AWS (AWS Management Console) mediante un navegador web para interactuar con dichos servicios. También permite conocer algunos aspectos de configuración adicionales de los servicios empleados.
- Versión AWS CLI: Esta es la versión donde se utiliza la interfaz de línea de comandos oficial de AWS (AWS CLI) para interactuar con los servicios de AWS. Elige esta versión si prefieres manejarte desde línea de comandos. Ante la duda elige la versión que usa la AWS Management Console.

Los resultados de aprendizaje de ambas prácticas son equivalentes. Se plantean las mismas actividades pero se resuelven usando herramientas diferentes. Por ello, la idea es que realices una única versión de la práctica (aquella en la que te vayas a sentir más cómodo). En cualquier caso, siempre puedes consultar la otra versión de la práctica para conocer cómo se realiza determinada acción con otra herramienta alternativa.

7. La Calculadora de Precios de AWS

Amazon ofrece la "Amazon Web Services Simple Monthly Calculator", disponible en [43], para poder estimar el consumo de una determinada arquitectura de aplicación en la nube en función de los recursos desplegados (tipo y características de los mismos así como el tiempo que estarán desplegados). Sin embargo, esta herramienta está comenzando a ser abandonada en favor de la AWS Pricing Calculator [105], con una interfaz más moderna.

Por ejemplo, la siguiente imagen muestra la configuración y el coste total de:

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

- 1 instancia de tipo t1.micro (0.6 GB de RAM, hasta 2 ECUs) basada en Linux
- 1 instancia de tipo m1.medium (3.7 GB de RAM, 2 ECUs) basada en Windows
- 2 volúmenes EBS de 180 GB cada uno.
- 1 IP elástica
- 1 balanceador de carga ELB procesando 100 GB / mes

Services Estimate of your Monthly Bill (\$ 154.07)

Choose region: US-East / US Standard (Virginia) Inbound Data Transfer is Free and Outbound Data Transfer is 1 GB free per month. Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale Amazon Elastic Block Store (EBS) provides persistent storage to Amazon EC2 instances.

Compute: Amazon EC2 Instances:

Description	Instances	Usage	Type	Billing Option	Monthly Cost
	1	100 % Utilized / 1	Linux on t1.micro	On-Demand (No Co)	\$ 14.54
	1	100 % Utilized / 1	Windows on m1.medium	On-Demand (No Co)	\$ 133.23

Storage: Amazon EBS Volumes:

Description	Volumes	Volume Type	Storage	IOPS	Snapshot Storage
	2	Standard	180 GB	0	0 GB-month of Storage

Elastic IP:

Number of Additional Elastic IPs: 1
Elastic IP Non-attached Time: 0 (Hours/Mon)
Number of Elastic IP Remaps: 0 (Per Month)

Data Transfer:

Inter-Region Data Transfer Out: 0 (GB/Month)
Data Transfer Out: 0 (GB/Month)
Data Transfer In: 0 (GB/Month)
Intra-Region Data Transfer: 0 (GB/Month)
Public IP/Elastic IP Data Transfer: 0 (GB/Month)

Elastic Load Balancing:

Number of Elastic LBs: 1
Total Data Processed by all ELBs: 100 (GB/Month)

Services Estimate of your Monthly Bill (\$ 154.07)

Estimate of Your Monthly Bill
Show First Month's Bill (include all one-time fees, if any)

With AWS, You only pay for what you use. Below you will see an estimate of your monthly bill. Expand each line item to see cost breakout of each service. To save this bill and input values, click on 'Save and Share' button. To remove the service from the estimate, jump back to the service and clear the specific service's form.

Amazon EC2 Service (US-East)		\$ 188.63
Compute:	\$ 147.87	
EBS Volumes:	\$ 18.00	
Elastic IPs:	\$ 3.66	
Elastic LBs:	\$ 18.30	
Data Processed by Elastic LBs:	\$ 0.80	
AWS Support (Basic)		\$ 0.00
Free Tier Discount:		\$ -34.56
Total Monthly Payment:		\$ 154.07

Habrás podido comprobar que la instancia más barata, 1 t1.micro desplegado en la región us-east-1, tiene un coste inferior a 15\$ mensuales. Esta herramienta permite obtener la estimación de coste de un determinado despliegue en función de los recursos desplegados, lo que permite dimensionar el tamaño de un despliegue en función de un coste mensual.

+ Actividad: AWS Pricing Calculator

Dedica unos minutos a utilizar la AWS Pricing Calculator, disponible en <https://calculator.aws/> para investigar los costes de los principales servicios de AWS.

Es muy importante llevar un control del gasto en recursos de AWS, estableciendo alarmas de aviso cuando el gasto supere un determinado umbral, así como haciendo uso del servicio Amazon Identity and Access Management (IAM) para limitar qué usuarios pueden usar qué recursos de qué servicios de AWS. Toda protección es poca para evitar casos como el de [62], que dejó accesible por descuido sus credenciales de acceso a AWS (o las de un usuario IAM vinculado a su cuenta) en un repositorio público de GitHub y a la mañana siguiente tenía una factura de más de \$2000. Al parece existen bots que escanean continuamente repositorios públicos de GitHub a la caza de credenciales de acceso a AWS para poder aprovisionar recursos para realizar actividades ilícitas (desde hacer *spam* masivo hasta hacer *Bitcoin mining* [63]).

8. Límites de AWS

AWS incluye por defecto una serie de limitaciones a las cuentas de usuario para que no puedan aprovisionar recursos más allá de ciertos límites. Por ejemplo, una cuenta de usuario de AWS no puede desplegar más de 20 instancias de EC2, aunque los límites varían en función del tipo de instancia [53] ya que, para instancias más potentes, como es el caso del tipo *i2.8xlarge* únicamente se pueden desplegar dos instancias de ese tipo.

Tienes información detallada sobre los límites para cada uno de los servicios en [54].

9. Otros Servicios de AWS Relacionados

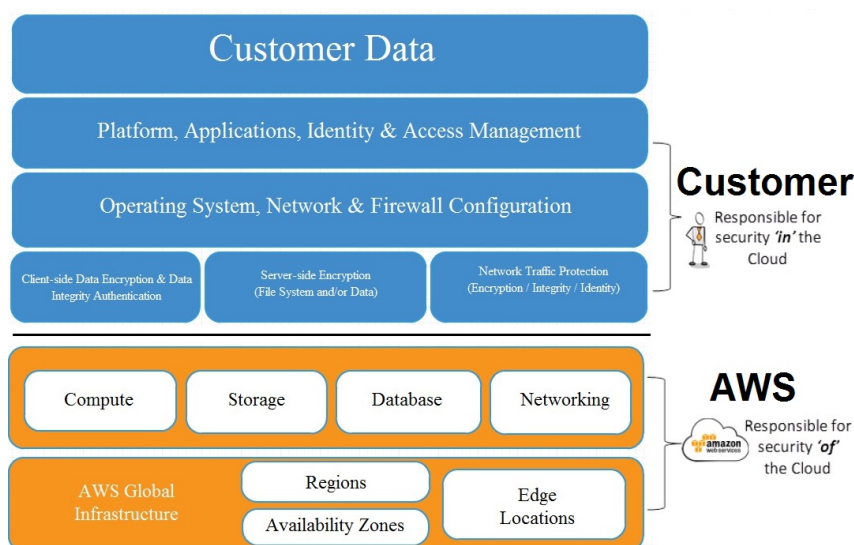
Con Amazon EC2 puedes desplegar instancias que soporten cualquier tipo de servicio web. Un escenario de uso común suele ser montar un servidor web en Amazon EC2. Si se trata de una web estática (código HTML sin necesidad de ejecutar

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

nada en el servidor), ten en cuenta que puedes utilizar la característica de almacenamiento de webs estáticas de Amazon S3 [59], de manera que las páginas se sirven directamente desde un bucket de S3. Si se trata de una web dinámica o necesitas tener un mayor control sobre las características del servidor web, entonces puedes desplegar una instancia de EC2 con un servidor web (como Apache) para alojar la página web. En este caso, generalmente necesitarás un nombre de dominio y ligarlo con la dirección IP de la instancia (que puede ser una IP elástica). Para ello puedes utilizar el servicio Amazon Route 53 [60], que permite tanto el registro del dominio como su vinculación con una o un conjunto de IPs para realizar el proceso de resolución de nombre a direcciones IP.

10. Modelo de Responsabilidad Compartida

Cuando utilizas un proveedor Cloud existe un modelo de seguridad compartida, tal y como ocurre en AWS [76] y que se muestra en la siguiente figura:



El proveedor Cloud, en este caso AWS es responsable de la seguridad *del* Cloud, en particular de los servicios de infraestructura que pone a disposición de los usuarios, mientras que los usuarios son responsables de la seguridad *dentro* del Cloud, es decir de las aplicaciones, sistemas y la gestión de datos desplegados en los recursos del Cloud. Por poner un ejemplo concreto, si despliegas una instancia de EC2 y por una gestión incorrecta de las actualizaciones del sistema operativo o cualquier otra causa, se instala un virus, entonces la responsabilidad, como no cabe de otra manera, recae sobre el usuario y no sobre el proveedor. La siguiente tabla divide la responsabilidad del proveedor de la responsabilidad del cliente, en función de los diferentes recursos:

Responsabilidad del Proveedor	Responsabilidad del Cliente
Instalaciones	Configuración de red
Seguridad física	Grupos de seguridad
Infraestructura de cómputo	Cortafuegos del SO
Infraestructura de red	Configuración del SO
Capa de virtualización	Aplicaciones y Servicios
	Cuentas de usuario

Consejos y Buenas Prácticas

Esta sección recoge algunos consejos y buenas prácticas que debes tener en cuenta a la hora de trabajar con Amazon Web Services.

- No es posible limitar el consumo mensual en AWS a una cierta cantidad, pero sí es posible definir alertas (usando Amazon CloudWatch) para recibir un correo electrónico o un SMS cuando el consumo de recursos supere un cierto umbral de dinero.
- Es posible añadir nuevas reglas a un grupo de seguridad asignado a una instancia de EC2 y esas reglas se aplican automáticamente a todas las instancias en ejecución vinculadas con dicho grupo de seguridad [22].
- Es posible incrementar las prestaciones en el acceso a un volumen EBS realizando un “calentamiento previo” del mismo ya que existe una penalización que puede llegar hasta un 50% de reducción de IOPS (instrucciones de entrada/salida por segundo) la primera vez que se accede a los datos de un volumen. Para ello, se recomienda leer todos los bloques del volumen antes de usarlo por primera vez. En Linux, se puede hacer fácilmente mediante el comando: `dd if=/dev/md0 of=/dev/null` [23].
- Las instancias de AMIs basadas en Ubuntu desplegadas en EC2 suelen impedir la autenticación de usuarios mediante contraseña vía SSH (tan solo mediante clave privada). Para habilitar que una cuenta de usuario creada en la instancia pueda ser accedida mediante SSH es necesario modificar la configuración del servidor SSH. Esto se puede hacer fácilmente cambiando el atributo `PasswordAuthentication` a `yes` en el fichero `/etc/ssh/sshd_config` y reiniciando el servicio SSH con `service sshd restart`.
- Si eliminas el fichero de clave privada correspondiente a un par de claves (keypair) de EC2 no será posible acceder a las instancias desplegadas con dicho par de claves, nuevamente vía SSH, salvo que existan otras cuentas de usuario creadas y la configuración del servidor SSH permita la autenticación mediante contraseña. Sin embargo, es posible modificar el volumen EBS asociado a la instancia para recuperar el acceso a dicha instancia, siguiendo el procedimiento explicado en [27].

Puedes encontrar más consejos relacionados con AWS en los siguientes enlaces [32], [33].

AWS también dispone del servicio AWS Answers [80] que recopila una serie de mejores prácticas para solucionar problemas comunes en áreas tan diversas como la gestión de cuentas de usuarios, la gestión de configuración, gestión de logs, aplicaciones móviles, redes, seguridad y aplicaciones web.

ANEXOS

Esta unidad lleva el siguiente anexo asociado:

- *AWS Command Line Interface (CLI)*. Describe la interfaz de línea de comandos oficial de AWS

REFERENCIAS

- [1] Amazon. Amazon Elastic Compute Cloud (EC2). <http://aws.amazon.com/ec2>
- [2] Amazon. Auto Scaling. <http://aws.amazon.com/autoscaling/>
- [3] Amazon CloudWatch. <http://aws.amazon.com/es/cloudwatch/>
- [4] Amazon. Amazon Simple Storage Service (S3). <http://aws.amazon.com/s3>
- [5] Amazon. Amazon Elastic Block Store (EBS). <http://aws.amazon.com/es/ebs/>
- [6] Amazon EC2 API Tools. <http://aws.amazon.com/developertools/351>
- [7] Elastic Load Balancing. <http://aws.amazon.com/es/elasticloadbalancing/>
- [8] Dropbox. <https://www.dropbox.com/>
- [9] Foursquare. <https://es.foursquare.com>
- [10] Pinterest. <http://pinterest.com>
- [11] Productos y servicios de Amazon Web Services. <http://aws.amazon.com/es/products/>
- [12] AWS Regions. https://aws.amazon.com/es/about-aws/global-infrastructure/regions_az/
- [13] Tim Kay. Simple command-line Access to Amazon EC2 and S3 (and others). <http://timkay.com/aws/>
- [14] Amazon S3 Pricing. <http://aws.amazon.com/es/s3/pricing/>
- [15] Amazon Machine Images. <https://aws.amazon.com/amis/>
- [16] AWS Management Console. <http://aws.amazon.com/es/console/>

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

- [17] Eric Hammond. Using RAID on EC2 EBS Volumes to Break the 1TB Barrier and Increase Performance. <http://alestic.com/2009/06/ec2-ebs-raid>
- [18] Planificación Round-robin. http://es.wikipedia.org/wiki/Planificaci3n_Round-robin
- [19] Read-After-Write Consistency in Amazon S3. <http://shlomoswidler.com/2009/12/read-after-write-consistency-in-amazon.html>
- [20] <https://forums.aws.amazon.com/message.jspa?messageID=198413>
- [21] AWS Online Training. <http://aws.amazon.com/es/aws-training/aws-online-training/>
- [22] Amazon EC2 Security Groups. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-network-security.html#ec2-classic-security-groups>
- [23] Increasing EBS Performance. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSPerformance.html>
- [24] Cross-zone load balancing. http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/TerminologyandKeyConcepts.html?channel=EM&Campaign_Type=Launch&Campaign_id=47667000&ref_pe=411040_47667000_7&#request-routing
- [25] You Should Use EBS Boot Instances on Amazon EC2. <http://alestic.com/2012/01/ec2-ebs-boot-recommended>
- [26] Paid AMIs. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/paid-amis.html>
- [27] Fixing Files on the Root EBS Volume of an EC2 Instance. <http://alestic.com/2011/02/ec2-fix-ebs-root>
- [28] Amazon CloudFront. <http://aws.amazon.com/es/cloudfront/>
- [29] <http://aws.typepad.com/aws/2013/12/three-more-cloudfront-locations.html>
- [30] <http://aws.amazon.com/es/cloudfront/#details>
- [31] Export EC2 Instances. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ExportingEC2Instances.html>
- [32] AWS Tips, Tricks, and Techniques. <https://launchbylunch.com/posts/2014/Jan/29/aws-tips>
- [33] AWS Tips I Wish I'd Known Before I Started. <http://wblinks.com/notes/aws-tips-i-wish-id-known-before-i-started/>
- [34] Amazon EC2 Reserved Instances. <http://aws.amazon.com/ec2/purchasing-options/reserved-instances/>
- [35] Mercado de instancias reservadas de Amazon EC2. <http://aws.amazon.com/es/ec2/purchasing-options/reserved-instances/marketplace/>
- [36] Interfaz de línea de comandos de AWS. <http://aws.amazon.com/es/cli/>
- [37] Elastic Load Balancing adds support for Connection Draining. <http://aws.amazon.com/about-aws/whats-new/2014/03/20/elastic-load-balancing-supports-connection-draining/>
- [38] Connection Draining. <http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/TerminologyandKeyConcepts.html#connection-drain>
- [39] You Should Use EBS Boot Instances on Amazon EC2. <https://alestic.com/2012/01/ec2-ebs-boot-recommended/>
- [40] Convert Instance-store AMI To EBS-booted AMI. <https://cloudyscripts.com/tool/show/2>
- [41] How to Convert S3-backed (instance store) AMI to EBS-backed AMI. <http://www.eaglegenomics.com/2011/05/how-to-convert-s3-instance-store-amis-to-ebs-amis/>
- [42] HowTo: Convert an Amazon EC2 EBS Image to an Instance Store Image. <http://www.dowdandassociates.com/blog/content/howto-convert-an-amazon-ec2-ebs-image-to-an-instance-store-image/>
- [43] Amazon Web Services Simple Monthly Calculator. <http://calculator.s3.amazonaws.com/index.html>
- [44] Productos y Servicios por Región. <https://aws.amazon.com/es/about-aws/globalinfrastructure/regional-product-services/>
- [45] EC2 Instance Belonging to Multiple ELBs. <http://shlomoswidler.com/2009/07/ec2-instance-belonging-to-multiple-elbs.html>
- [46] Amazon EBS Volume Performance. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSPerformance.html>
- [47] iostat. <https://linux.die.net/man/1/iostat>
- [48] Configure Instance Termination Policy for Your Auto Scaling Group. <http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/us-termination-policy.html>
- [49] AWS Command Line Interface. <http://aws.amazon.com/cli/>
- [50] RAID Configuration. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/raid-config.html>
- [51] EBS Pricing. <http://aws.amazon.com/es/ebs/pricing/>
- [52] Amazon EBS Volume Types. <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSVolumeTypes.html>
- [53] How Many Instances Can I Run in Amazon EC2. http://aws.amazon.com/ec2/faqs/#How_many_instances_can_I_run_in_Amazon_EC2
- [54] AWS Service Limits. http://docs.aws.amazon.com/general/latest/gr/aws_service_limits.html

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

- [55] Elastic IP Addresses (EIP). <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/elastic-ip-addresses-eip.html>
- [56] Amazon Route 53. <http://aws.amazon.com/es/route53/>
- [57] Dispositivos orientados a bloques.
http://es.wikipedia.org/wiki/Archivo_de_dispositivo#Dispositivos_orientados_a_bloques
- [58] Elastic Load Balancing Pricing. <http://aws.amazon.com/elasticloadbalancing/pricing/>
- [59] Hosting a Static Website on Amazon S3. <http://docs.aws.amazon.com/AmazonS3/latest/dev/WebsiteHosting.html>
- [60] Amazon Route 53. <http://aws.amazon.com/es/route53/>
- [61] <http://aws.amazon.com/es/new/reinvent/>
- [62] My \$2375 Amazon EC2 Mistake. <http://www.devfactor.net/2014/12/30/2375-amazon-mistake/>
- [63] Inside a chinese bitcoin mine. <http://www.thecoinsman.com/2014/08/bitcoin/inside-chinese-bitcoin-mine/>
- [64] Profiling (computer programming). [http://en.wikipedia.org/wiki/Profiling_\(computer_programming\)](http://en.wikipedia.org/wiki/Profiling_(computer_programming))
- [65] NFS – Network File System. http://es.wikipedia.org/wiki/Network_File_System
- [66] Gluster. <http://www.gluster.org>
- [67] OrangeFS. <http://orangeefs.org>
- [68] Ceph. <http://ceph.com/docs/master/>
- [69] Precios de Amazon CloudWatch. <http://aws.amazon.com/es/cloudwatch/pricing/>
- [70] AWS Worldwide Public Sector. https://s3.amazonaws.com/awspworkshop/01_DAY1_Introduction-to-AWS.pdf
- [71] Cross-Region Replication. <https://docs.aws.amazon.com/AmazonS3/latest/dev/crr.html>
- [72] Scryer: Netflix's Predictive Auto Scaling Engine. <http://techblog.netflix.com/2013/11/scryer-netflixs-predictive-auto-scaling.html>
- [73] Dynamic Scaling. http://docs.aws.amazon.com/AutoScaling/latest/DeveloperGuide/as-scale-based-on-demand.html?ref=pe_undef
- [74] AWS Total Cost Of Ownership (TCO) Calculator [Abandonada]. <https://awstccalculator.com>
- [75] Getting Started with CloudWatch Logs.
http://docs.aws.amazon.com/AmazonCloudWatch/latest/DeveloperGuide/CWL_GettingStarted.html
- [76] AWS Shared Responsibility Model. <https://aws.amazon.com/es/compliance/shared-responsibility-model/>
- [77] AWS moves from ECU to vCPU. <http://blogs.gartner.com/kyle-hilgendorf/2014/04/16/aws-moves-from-ecu-to-vcpu/>
- [78] AWS Storage Update – New Lower Cost S3 Storage Option & Glacier Price Reduction.
<https://aws.amazon.com/es/blogs/aws/aws-storage-update-new-lower-cost-s3-storage-option-glacier-price-reduction>
- [79] New – AWS Price List API. <https://aws.amazon.com/blogs/aws/new-aws-price-list-api/>
- [80] AWS Answers. <http://aws.amazon.com/answers/>
- [81] A Decade of Innovation. <http://perspectives.mvdirona.com/2016/03/a-decade-of-innovation/>
- [82] Access Control List (ACL) Overview. <http://docs.aws.amazon.com/AmazonS3/latest/dev/acl-overview.html>
- [83] ¿Cómo puedo seleccionar el tipo de instancia correcto?
https://aws.amazon.com/es/ec2/faqs/#How_do_I_select_the_right_instance_type
- [84] What is a Classic Load Balancer? <http://docs.aws.amazon.com/elasticloadbalancing/latest/classic/introduction.html>
- [85] What is an Application Load Balancer?
<http://docs.aws.amazon.com/elasticloadbalancing/latest/application/introduction.html>
- [86] New – AWS Application Load Balancer. <https://aws.amazon.com/es/blogs/aws/new-aws-application-load-balancer/>
- [87] OSI model. https://en.wikipedia.org/wiki/OSI_model
- [88] Microservices. <http://martinfowler.com/articles/microservices.html>
- [89] Modifying the Size, IOPS, or Type of an EBS Volume on Linux.
<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-expand-volume.html>
- [90] Amazon EBS Update – New Elastic Volumes Change Everything. <https://aws.amazon.com/es/blogs/aws/amazon-ebs-update-new-elastic-volumes-change-everything/>
- [91] New – Per-Second Billing for EC2 Instances and EBS Volumes. <https://aws.amazon.com/es/blogs/aws/new-per-second-billing-for-ec2-instances-and-ebs-volumes/>
- [92] Carga de objetos con la API de carga multiparte.
https://docs.aws.amazon.com/es_es/AmazonS3/latest/dev/uploadobjusingmpu.html
- [93] AWS DataSync. <https://aws.amazon.com/es/datasync/>
- [94] AWS Snowball. <https://aws.amazon.com/es/snowball/>
- [95] Configuración de sesiones sticky para balanceador de carga clásico.
https://docs.aws.amazon.com/es_es/elasticloadbalancing/latest/classic/elb-sticky-sessions.html

GESTIÓN DE CÓMPUTO Y FICHEROS EN LA NUBE. Amazon EC2, S3 y Auto Scaling

- [96] Pre-Warming the Load Balancer. <https://aws.amazon.com/articles/best-practices-in-evaluating-elastic-load-balancing/#pre-warming>
- [97] Best Practices in Evaluating Elastic Load Balancing. <https://aws.amazon.com/articles/best-practices-in-evaluating-elastic-load-balancing>
- [98] New – Application Load Balancer Simplifies Deployment with Weighted Target Groups. <https://aws.amazon.com/es/blogs/aws/new-application-load-balancer-simplifies-deployment-with-weighted-target-groups/>
- [99] AWS Blue/Green Deployments. https://d0.awsstatic.com/whitepapers/AWS_Blue_Green_Deployments.pdf
- [100] Application Load Balancer now supports Least Outstanding Requests algorithm for load balancing requests. <https://aws.amazon.com/es/about-aws/whats-new/2019/11/application-load-balancer-now-supports-least-outstanding-requests-algorithm-for-load-balancing-requests/>
- [101] NASA to launch 247 petabytes of data into AWS – but forgot about eye-watering cloudy egress costs before lift-off. https://www.theregister.co.uk/2020/03/19/nasa_cloud_data_migration_mess/
- [102] Políticas de escalado de seguimiento de destino para Auto Scaling de aplicaciones. https://docs.aws.amazon.com/es_es/autoscaling/application/userguide/application-auto-scaling-target-tracking.html
- [103] Single Point of Failure. https://es.wikipedia.org/wiki/Single_point_of_failure
- [104] DNS Load Balancing. <https://www.nginx.com/resources/glossary/dns-load-balancing/>
- [105] AWS Pricing Calculator. <https://calculator.aws/>
- [106] Automating the Amazon EBS snapshot lifecycle. <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/snapshot-lifecycle.html>
- [107] Instantáneas de Amazon EBS. https://docs.aws.amazon.com/es_es/AWSEC2/latest/UserGuide/EBSSnapshots.html
- [108] Eliminación de una instantánea de Amazon EBS. https://docs.aws.amazon.com/es_es/AWSEC2/latest/UserGuide/ebs-deleting-snapshot.html
- [109] Availability Calculator. <https://availability.sre.xyz/>
- [110] Fast Snapshot Restore. <https://aws.amazon.com/es/blogs/aws/new-amazon-ebs-fast-snapshot-restore-fsr/>
- [111] Amazon S3 Update – Strong Read-After-Write Consistency. <https://aws.amazon.com/es/blogs/aws/amazon-s3-update-strong-read-after-write-consistency/>
- [112] Reemplazar un volumen Amazon EBS con una instantánea anterior. https://docs.aws.amazon.com/es_es/AWSEC2/latest/UserGuide/ebs-restoring-volume.html#replace-root
- [113] Infraestructura global. <https://aws.amazon.com/es/about-aws/global-infrastructure/>
- [114] AWS Wavelength. <https://aws.amazon.com/es/wavelength/>
- [115] AWS Local Zones. <https://aws.amazon.com/es/about-aws/global-infrastructure/localzones/>
- [116] Asociar un volumen a varias instancias con Amazon EBS Multi-Attach. https://docs.aws.amazon.com/es_es/AWSEC2/latest/UserGuide/ebs-volumes-multi.html
- [117] Amazon EFS. <https://aws.amazon.com/es/efs/>
- [118] mountpoint-s3. <https://github.com/awslabs/mountpoint-s3/tree/main>
- [119] Identify and optimize public IPv4 address usage on AWS. <https://aws.amazon.com/es/blogs/networking-and-content-delivery/identify-and-optimize-public-ipv4-address-usage-on-aws/>