

Curso Online de Cloud Computing con Amazon Web Services



Introducción a AWS

Germán Moltó

Instituto de Instrumentación para
Imagen Molecular

gmolto@dsic.upv.es

<http://www.grycap.upv.es/cursocloud>



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Resultados de Aprendizaje

- Se espera que una vez acabes esta presentación seas capaz de ...
 - Conocer una panorámica de los servicios de Amazon Web Services (AWS).
 - Conocer el objetivo fundamental de cada uno de los servicios de AWS.
 - Comprender la infraestructura global de AWS.
 - Conocer las diferentes regiones de AWS.
 - Entender la diferencia entre región y zona de disponibilidad

Introducción a Amazon Web Services

- Amazon Web Services (AWS) ofrece servicios de infraestructuras para ejecutar aplicaciones en la nube.
- Es pionero en Cloud Computing, desde 2006 ofreciendo:
 - Bajo Coste
 - Agilidad y elasticidad instantánea
 - Abierto y flexible
 - Seguro



Principales Servicios de AWS (I)

- Procesamiento de datos
 - Amazon Elastic Compute Cloud (EC2)
 - Amazon Elastic MapReduce
 - Auto Scaling
 - Elastic Load Balancing
- Almacenamiento
 - Amazon Simple Storage Service (S3)
 - Amazon Glacier
 - Amazon Elastic Block Store (EBS)
 - AWS Import/Export
 - AWS Storage Gateway

Principales Servicios de AWS (II)

- Base de datos
 - Amazon SimpleDB
 - Amazon Relational Database Service (RDS)
 - Amazon ElastiCache
 - Amazon DynamoDB
- Despliegue y Gestión
 - AWS Identity and Access Management (IAM)
 - Amazon CloudWatch
 - AWS Elastic Beanstalk
 - AWS CloudFormation
 - AWS Data Pipeline

Principales Servicios de AWS (III)

- Servicios de Aplicaciones
 - Amazon CloudSearch
 - Amazon Simple Workflow Service (SWF)
 - Amazon Simple Queue Service (SQS)
 - Amazon Simple Notification Service (SNS)
 - Amazon Simple Email Service (SES)
- Redes
 - Amazon Route 53
 - Amazon Virtual Private Cloud (VPC)
 - AWS Direct Connect

Principales Servicios de AWS (IV)

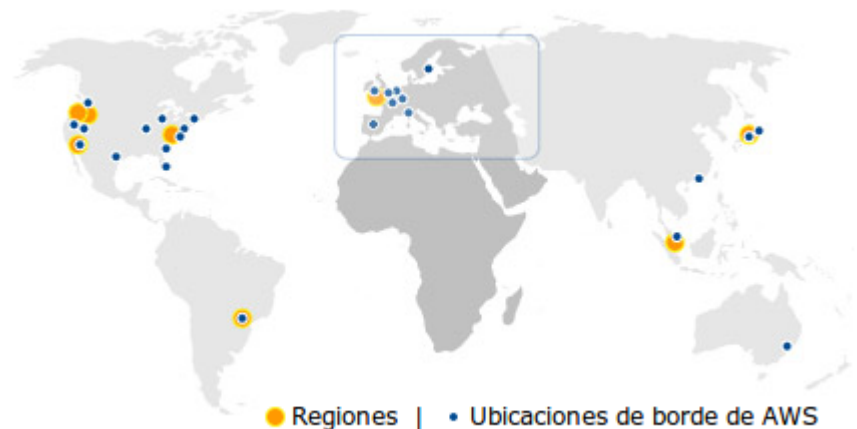
- Entrega de contenido
 - Amazon CloudFront
- Pagos y facturación
 - Amazon Flexible Payment Service (FPS)
 - Amazon DevPay
- Personal
 - Amazon Mechanical Turk
- Software
 - AWS Marketplace

Principales Servicios de AWS (V)

- Amazon Redshift
- Amazon Elastic Transcoder
- AWS OpsWorks

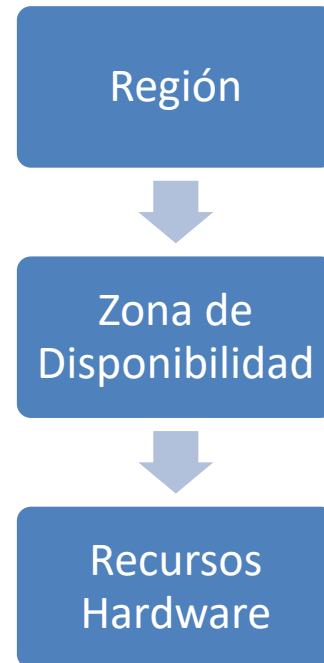
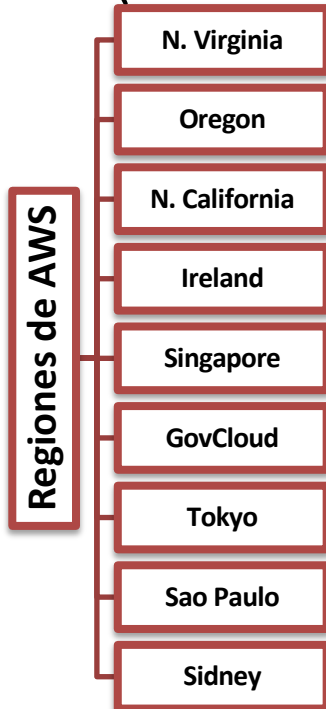
Infraestructura Global

- AWS está disponible en más de 190 países y su infraestructura está distribuida a lo largo del mundo para facilitar la proximidad con sus clientes.
 - Reducción de latencia y costes variables por región.



Regiones en AWS

- Los recursos de AWS están distribuidos por el mundo por Regiones (9 regiones).
 - Cada Región incluye diferentes zonas de disponibilidad (diferentes CPDs, por tolerancia a fallos).



Conclusiones

- Amazon Web Services (AWS) es el proveedor Cloud público pionero y líder.
- Incluye numerosos servicios para el despliegue de aplicaciones en el Cloud.
- Consta de diferentes regiones (datacenters en diferentes localizaciones geográficas)
 - Con múltiples zonas de disponibilidad en cada región.

Curso Online de Cloud Computing con Amazon Web Services



Introducción a Amazon S3

Germán Moltó

Instituto de Instrumentación para
Imagen Molecular

gmolto@dsic.upv.es

<http://www.grycap.upv.es/cursocloud>



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Resultados de Aprendizaje

- Se espera que una vez acabes esta presentación seas capaz de ...
 - Conocer el objetivo principal de Amazon S3.
 - Conocer las características de replicación del servicio Amazon S3.
 - Comprender el concepto de *bucket*.
 - Conocer los diferentes mecanismos de redundancia ofrecidos.

Introducción a S3

- “Amazon S3 es almacenamiento para Internet. Está diseñado para facilitar a los desarrolladores la informática a escala web.”
- Consiste en un sistema de almacenamiento de un número ilimitado de objetos para el entorno Cloud.
 - Cada objeto entre 1 byte y 5 Terabytes.
- Accesible mediante protocolos estándar (HTTP, Bittorrent etc.).
- Confiabilidad (*Reliability*)
 - Almacenamiento de un 99.999999999% de durabilidad, 99.99% de disponibilidad.

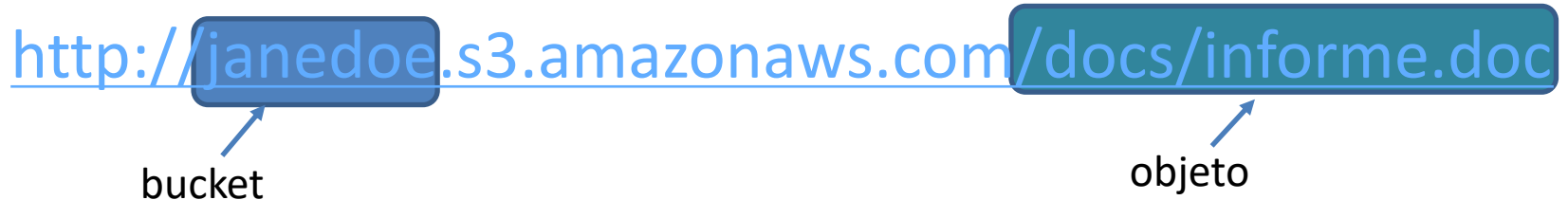
Buckets en S3

- Datos almacenados como objetos en *buckets* (traducidos como *depósitos*).
- Un bucket
 - Es un contenedor de objetos.
 - Lleva control de acceso establecido por el usuario (quién puede crear, borrar y listar el bucket).
 - Puede ser auditado (registros de acceso) y versionado.
 - Puede ser almacenado en una región.
- Un objeto
 - es un fichero + [metadatos de descripción (opcional)]
 - Tiene un tamaño entre 1 byte y 5 Terabytes

Sobre los Buckets

- Si el objeto llamado *docs/informe.doc* está guardado en el bucket *janedoe*, entonces, el documento está accesible en la dirección:

<http://janedoe.s3.amazonaws.com/docs/informe.doc>



- También se utiliza la nomenclatura:
`http://s3.amazonaws.com/janedoe/docs/informe.doc`
- Es posible utilizar carpetas (folders) para organizar a los objetos dentro un bucket.
 - docs/ podría ser una carpeta situada en el bucket janedoe.

Sobre Replicación y Consistencia (I)

- Los objetos se replican entre diferentes servidores de los centros de datos de Amazon.
- Reduced Redundancy Storage (RRS), Almacenamiento con Redundancia Reducida
 - Almacenamiento en S3 de ficheros que no son críticos (o se pueden reproducir).
- Las actualizaciones de un objeto son atómicas.
 - La escritura no se traduce en datos corruptos o incompletos si la operación se ejecuta.

Sobre Replicación y Consistencia (II)

- Según la región es posible disponer de:
 - eventual *consistency*
 - El nuevo contenido del objeto no tiene porqué estar disponible inmediatamente después de la escritura. Hasta que los cambios no se propagan a todas las réplicas, es posible obtener resultados inesperados (datos borrados que todavía son accesibles, datos recién subidos inexistentes, etc.)
 - read-after-write *consistency*
 - Garantía de lectura del valor correcto tras operación de escritura.

Uso de S3

- Existen diferentes formas de interactuar con S3:
 - Por medio de la AWS Management Console (entorno gráfico basado en Web)
 - Mediante herramientas de línea de comandos
 - Usando las APIs públicas de S3 desde un lenguaje de programación.
 - Python, Ruby, Java, etc.

AWS Management Console con S3

- Accesible vía web, permite crear buckets, carpetas y gestionar ficheros y permisos.

The screenshot displays the AWS Management Console interface for an S3 bucket named 'alucoud00'. The top navigation bar includes the AWS logo, 'Services' dropdown, 'Edit' dropdown, and user information 'Germán Moltó' with 'Global' and 'Help' links. Below the navigation bar, there are buttons for 'Upload', 'Create Folder', and 'Actions'. The main content area shows a table of objects in the bucket. The table has columns for 'Name', 'Storage Class', 'Size', and 'Last Modified'. One object is listed: 'hello2.txt' with a 'Standard' storage class, '18 bytes' size, and a last modified date of 'Thu Jul 05 18:51:23 GMT'. To the right of the table, a modal window displays the details for the selected file 'hello2.txt'. This modal includes tabs for 'None', 'Properties', and 'Transfers'. The 'Properties' tab is active, showing metadata such as 'Bucket: alucoud00', 'Name: hello2.txt', 'Link: https://s3.amazonaws.com/alucoud00/hello2.txt', 'Size: 18 bytes', 'Last Modified: Thu Jul 05 18:51:23 GMT+200 2012', 'Owner: Me', 'ETag: a7c7af82727092e76bb60d10b8ac69ba', 'Expiry Date: None', and 'Expiration Rule: N/A'. Below the metadata, there are sections for 'Details' and 'Permissions'. The 'Details' section shows 'Storage Class' set to 'Standard' (selected) and 'Reduced Redundancy', and 'Server Side Encryption' set to 'None' (selected) and 'AES-256'. At the bottom right of the modal, there are 'Save' and 'Cancel' buttons.

Name	Storage Class	Size	Last Modified
hello2.txt	Standard	18 bytes	Thu Jul 05 18:51:23 GMT

hello2.txt

Bucket: alucoud00
Name: hello2.txt
Link: <https://s3.amazonaws.com/alucoud00/hello2.txt>
Size: 18 bytes
Last Modified: Thu Jul 05 18:51:23 GMT+200 2012
Owner: Me
ETag: a7c7af82727092e76bb60d10b8ac69ba
Expiry Date: None
Expiration Rule: N/A

▼ Details

Storage Class: ☒ Standard ☐ Reduced Redundancy

Server Side Encryption: ☒ None ☐ AES-256

Save **Cancel**

► Permissions

► Metadata

Acceso a S3 desde línea de comandos

- Existen diversas herramientas desarrolladas por terceras personas para permitir el acceso a S3 desde línea de comandos:
 - <http://aws.amazon.com/code/1710>
 - <http://aws.amazon.com/code/2124>
- Una de las herramientas más populares es:
 - Simple Command-Line Access to Amazon EC2 and Amazon S3
 - <http://aws.amazon.com/code/739> [Top-rated community code]
 - <http://timkay.com/aws/>
 - Se trata de un único fichero de código Perl que permite facilitar el acceso tanto a EC2 como a S3 (entre otros servicios de AWS).
 - Tan solo depende de Perl y de CURL.

Esquema de Precios de S3

- Los precios dependen de la región. Ejemplo para US Standard:
- Coste 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)
- Costes de 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

Amazon Glacier

- Servicio de almacenamiento de muy bajo coste (desde 0.01\$/GB/mes).
- Pensado para datos perdurables en el tiempo pero cuya tasa de acceso no es elevada:
 - Por ejemplo copias de seguridad.
- La recuperación de los datos tarda entre 3 y 5 horas.



Conclusiones

- Amazon S3 permite el almacenamiento de ficheros en AWS.
- Ofrece diferentes opciones de redundancia para reducir los costes (almacenando menos réplicas de datos)
- El servicio Amazon Glacier todavía ofrece mejores precios a costa de aumentar el tiempo de acceso a los datos.

Curso Online de Cloud Computing con Amazon Web Services



El Servicio Amazon EC2

Germán Moltó

Instituto de Instrumentación para
Imagen Molecular

gmolto@dsic.upv.es

<http://www.grycap.upv.es/cursocloud>



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

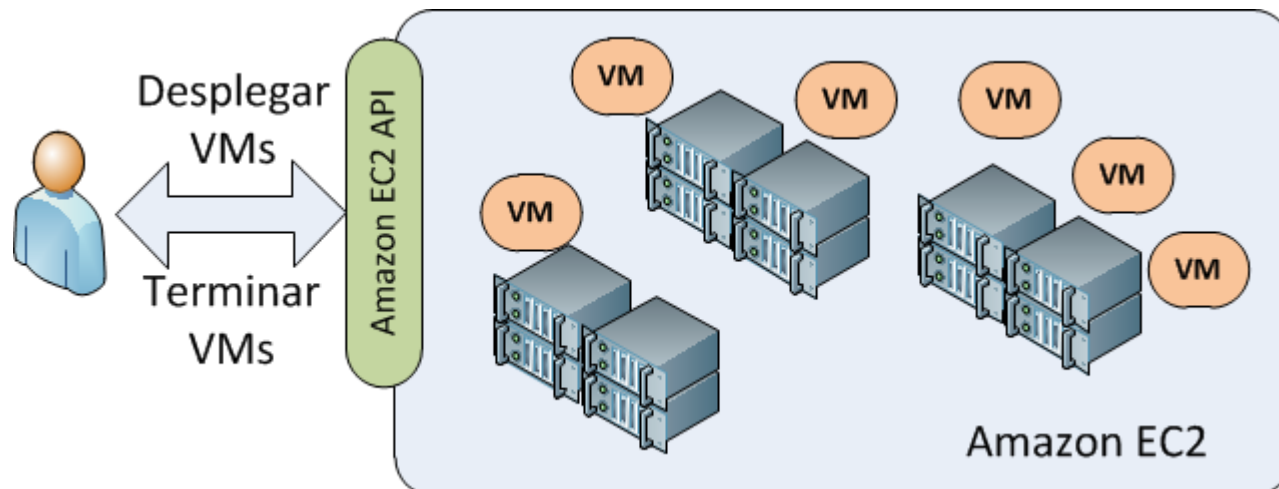


Resultados de Aprendizaje

- Se espera que una vez acabes esta presentación seas capaz de ...
 - Comprender la funcionalidad ofrecida por EC2.
 - Entender los conceptos de AMI, Región, Par de Clave, Instancia y Grupo de Seguridad, en EC2.
 - Conocer los diferentes tipos de instancias de cómputo disponibles y una idea de su coste.
 - Comprender la funcionalidad del servicio EBS.
 - Conocer las diferencias entre AMIs basadas en S3 y AMIs basadas en EBS.

Amazon EC2

- Amazon EC2 es un servicio que proporciona recursos de cómputo en la nube.
 - Permite el despliegue de imágenes de máquinas virtuales sobre instancias de tamaño predefinido para disponer de cómputo bajo demanda mediante un modelo de pago por uso.



Concepto de Instancia en EC2

- AWS utiliza el término *instancia* para referirse a una máquina virtual (MV).
 - Una instancia proporciona una cantidad predecible de capacidad de cómputo dedicada.
- Cada tipo de instancia proporciona un número determinado de *ECUs* (Elastic Compute Unit)
 - Un ECU proporciona el equivalente a una CPU Opteron de 1.0-1.2 GHz 2007 o un Procesador Xeon de 2007.

Despliegue de Instancias en EC2

- El despliegue de una instancia requiere indicar:
 - Tipo de instancia
 - Prestaciones de la MV (tipos predefinidos de instancia cuyo precio por hora varía según las prestaciones)
 - AMI (Amazon Machine Image)
 - Imagen de la Máquina Virtual. Determina el SO y las aplicaciones disponibles en la instancia cuando arranque.
 - Grupo de seguridad (*security group*)
 - La configuración de cortafuegos de la instancia (qué tráfico puede recibir la instancia)
 - Par de claves (*keypair*)
 - Permite la conexión a la instancia mediante SSH sin contraseña
 - Región (y zona de disponibilidad) [Opcional]
 - La región por defecto es us-east-1 (Virginia, USA)

EC2: Tipos de Instancias (I)

- Distintos tipos de instancias preconfiguradas.
- Instancias específicas para:
 - HPC, GPUs
- Spot instances vs Reserved instances.

Instancias Estándar

Nombre	Mem.	Disco	Cores(*)	Plataforma	Precio por hora
m1.small	1.7 Gb	160 Gb.	1 (1x1)	32 ó 64 bits	0.060\$
m1.medium	3.75 Gb	410 Gb	2 (1x2)	32 ó 64 bits	0.120\$
m1.large	7.5 Gb	850 Gb.	4 (2x2)	64 bits	0.240\$
m1.xlarge	15 Gb	1690 Gb.	8 (4x2)	64 bits	0.480\$

(*) Amazon mide en n procesadores de m EC2 Cores.
1 EC2 core ~ Xeon 1-1.2 GHz (de 2007).

Instancias con Gran Cantidad de Memoria

Nombre	Memoria	Disco	Cores(*)	Plataforma	Precio por hora
m2.xlarge	17.1 Gb	420 Gb.	6.5 (2x3.25)	64 bits	0.410\$
m2.2xlarge	34.2 Gb	850 Gb.	13 (4x3.25)	64 bits	0.820\$
m2.4xlarge	68.4 Gb	1690 Gb.	26 (8x3.25)	64 bits	1.640\$

Precios de la región: UE (Irlanda), Linux, Marzo, 2013

Amazon Machine Images

- AMIs por:
 - Proveedor
 - IBM, Oracle, etc.
 - Región
 - US East, Asia Pacific, etc.
 - Arquitectura
 - i386, x86_64.
 - Plataforma
 - Ubuntu, Windows, Amazon Linux, etc.
 - Dispositivo raíz
 - instance-store
 - ebs

The screenshot shows the Amazon Machine Images (AMIs) console. The browser address bar displays <https://aws.amazon.com/amis>. The page title is "Amazon Machine Images (AMIs)". A sidebar on the left lists categories for browsing AMIs: Provider (Community, IBM, Oracle, Amazon Web Services, Sun Microsystems, Novell, Microsoft), Region (US East, EU, Asia Pacific, US West, South America), Architecture (i386, x86_64), Root Device Type (instance-store, ebs), and Platform (Ubuntu, Other Linux, Windows, Red Hat, Cent OS). The main content area shows a list of AMIs, with the first four being BitNami stacks (Subversion, Pootle, Pootle 64bits, EnanoCMS) and the fifth being Searchdaimon Enterprise Search 2.5. Each AMI entry includes the provider logo, the AMI name, a brief description, and a "Launch AMI" button. The list is sorted by "Date - newest first" and shows "Showing 1-25 of 1062 AMIs".

Amazon Machine Images (AMIs)

An Amazon Machine Image (AMI) is a special type of pre-configured operating system and virtual application software which is used to create a virtual machine within the Amazon Elastic Compute Cloud (EC2). It serves as the basic unit of deployment for services delivered using EC2.

Read the Amazon EC2 Developer Guide for information on [safely using shared AMIs](#).

Showing 1-25 of 1062 AMIs Sort by: Date - newest first

ubuntu **BitNami Subversion Stack 1.7.1-0 EBS (Ubuntu 12.04)**
BitNami Subversion Stack Amazon Machine Image packages Subversion and all of it required dependencies, such as Apache and SQLite and the Ubuntu 12.04.
Listed on Jun 19, 2012 02:03 GMT [Launch AMI](#) [More Info](#)

ubuntu **BitNami Pootle Stack 2.1.6-1 EBS (Ubuntu 12.04)**
BitNami Pootle Stack Amazon Machine Image packages Pootle and all of it required dependencies, such as Apache, MySQL, PostgreSQL, SQLite and Django and the Ubuntu 12.04.
Listed on Jun 19, 2012 02:02 GMT [Launch AMI](#) [More Info](#)

ubuntu **BitNami Pootle Stack 2.1.6-1 64bits (Ubuntu 12.04)**
BitNami Pootle Stack Amazon Machine Image packages Pootle and all of it required dependencies, such as Apache, MySQL, PostgreSQL, SQLite and Django and the Ubuntu 12.04.
Listed on Jun 19, 2012 02:02 GMT [Launch AMI](#) [More Info](#)

ubuntu **BitNami EnanoCMS Stack 1.1.8-1 EBS (Ubuntu 12.04)**
BitNami EnanoCMS Stack Amazon Machine Image packages EnanoCMS and all of it required dependencies, such as Apache, MySQL, PHP and phpMyAdmin and the Ubuntu 12.04.
Listed on Jun 19, 2012 02:02 GMT [Launch AMI](#) [More Info](#)

fedora **Searchdaimon Enterprise Search 2.5**
Searchdaimon ES is a free Enterprise search appliance configured, ready to run. It can index websites and several common enterprise systems like SharePoint, Exchange, SQL databases, Windows file shares etc.
Listed on Jun 19, 2012 02:02 GMT [Launch AMI](#) [More Info](#)

ubuntu **BitNami JBoss Stack 6.0.0-0 EBS 64bits (Ubuntu 12.04)**
BitNami JBoss Stack Amazon Machine Image packages Ubuntu 12.04 operating system and a fully configured JBoss environment, including
Listed on Jun 19, 2012 02:02 GMT [Launch AMI](#) [More Info](#)

Sobre las AMIs

- Una AMI es como un CD de arranque, contiene toda la información necesaria para arrancar la instancia.
- Hay dos tipos de AMIs:
 - Almacenadas en S3 (denominadas *instance-store*)
 - Cualquier fichero creado o modificado en la máquina virtual desaparecerá cuando se apague la máquina virtual (almacenamiento efímero).
 - Almacenadas en EBS (denominadas *ebs*)
 - El dispositivo raíz ya no es efímero. Los cambios en la máquina virtual se almacenan en un volumen EBS asociado a la instancia.
- Las AMIs se crean para una región concreta, pero pueden copiarse entre regiones.

Introducción a EBS (I)

- Amazon Elastic Block Store (EBS) proporciona volúmenes orientados a bloques (como si fuera un disco duro o un disco USB) para ser conectados dinámicamente a instancias de EC2.
- Límites de hasta 5000 volúmenes EBS por cuenta ó hasta 20 TB de espacio total de almacenamiento en EBS. [1GB, 1 TB] para cada volumen.
- 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
- Existen los volúmenes EBS con IOPS (*Input/output Operations per Second*) aprovisionadas, para ofrecer alto rendimiento.

Introducción a EBS (II)

- Una vez conectado el volumen, hay que particionarlo y formatearlo para poder montarlo a la instancia.
 - Tal y como se haría con cualquier disco conectado a un equipo.
- Un volumen EBS solo puede estar conectado a una misma instancia en un momento dado.
 - Analogía con un disco duro, que no puede estar conectado a dos equipos de forma simultánea.
 - Además, el volumen y la instancia deben estar en la misma zona de disponibilidad.
- Comparación con S3
 - S3 es un servicio de objetos (ficheros), mientras que EBS es un servicio de volúmenes (discos).

Instancias Basadas en EBS

- La puesta en marcha de una AMI basada en EBS provoca la creación automática de un volumen EBS para almacenar los datos de la misma.
- El volumen queda conectado (y los datos preservados) incluso aunque se detenga la instancia (no es el caso si se termina).
 - Al detener la instancia ya no se produce gasto por consumo de horas de instancia, pero sí por el almacenamiento del volumen de datos EBS asociado.
- La terminación de la instancia basada en EBS provoca la destrucción del volumen EBS asociado.


Sobre los Tipos de AMI

- Las instancias basadas en EBS son, por lo general, más convenientes que las basadas en S3.
 - Principalmente porque soportan mayor tamaño de la partición raíz.
- Esquema de pago
 - En las instancias basadas en S3 se cobra por el almacenamiento de la AMI (si es del usuario) y por el uso.
 - En las instancias basadas en EBS:
 - No se cobra si la instancia está detenida.
 - Se cobra por el espacio de almacenamiento.
 - Se cobra por cada transición de detenida a iniciada.

AMIs Basadas en S3 vs AMIs basadas en EBS

	AMI basada en S3 (Amazon EC2 instance store-backed)	AMI basada en EBS (Amazon EBS-backed)
Tamaño de la Partición Raíz	Hasta 10 GiB	Hasta 16 TiB
Posibles estados	Solo se pueden crear y terminar, no parar	Se pueden parar y reiniciar, sin perder el estado.
Tiempo de Arranque	Más lento. Hay que copiar toda la partición de arranque desde S3	Más rápido. Solo las porciones necesarias para arrancar deben ser obtenidas desde un snapshot
Creación de AMIs	Crear imagen desde la propia instancia. Posterior registro. Difícil.	Crear imagen desde fuera. Single-click para duplicar volumen y registrar (válido para Windows y Linux). Sencillo.

AWS Management Console con EC2

 **Services** ▾ **Edit** ▾

Germán Moltó ▾ N. Virginia ▾ Help ▾

EC2 Dashboard

Events

INSTANCES

Instances

Spot Requests

Reserved Instances

IMAGES

AMIs

Bundle Tasks

ELASTIC BLOCK STORE

Volumes

Snapshots

NETWORK & SECURITY

Security Groups

Elastic IPs

Placement Groups

Load Balancers

Key Pairs

Network Interfaces

Resources

You are using the following Amazon EC2 resources in the US East (N. Virginia) region:

0 Running Instances

0 Elastic IPs

0 Volumes

0 Snapshots

6 Key Pairs

0 Load Balancers

0 Placement Groups

8 Security Groups

Create Instance

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

Launch Instance

Note: Your instances will launch in the US East (N. Virginia) region

Service Health

Service Status:

✓ US East (N. Virginia): This service is operating normally

Availability Zone Status:

✓ us-east-1a

us-east-1a

Availability zone is operating normally

✓ us-east-1b

us-east-1b

Availability zone is operating normally

✓ us-east-1c

us-east-1c

Availability zone is operating normally

✓ us-east-1d

us-east-1d

Availability zone is operating normally

✓ us-east-1e

us-east-1e

Availability zone is operating normally

[Service Health Dashboard](#)

Scheduled Events

US East (N. Virginia):

No events

Additional Information

Getting Started Guide

Documentation

All EC2 Resources

Find software on AWS Marketplace

Forums

Pricing

Feedback

Feedback

Report an Issue

38

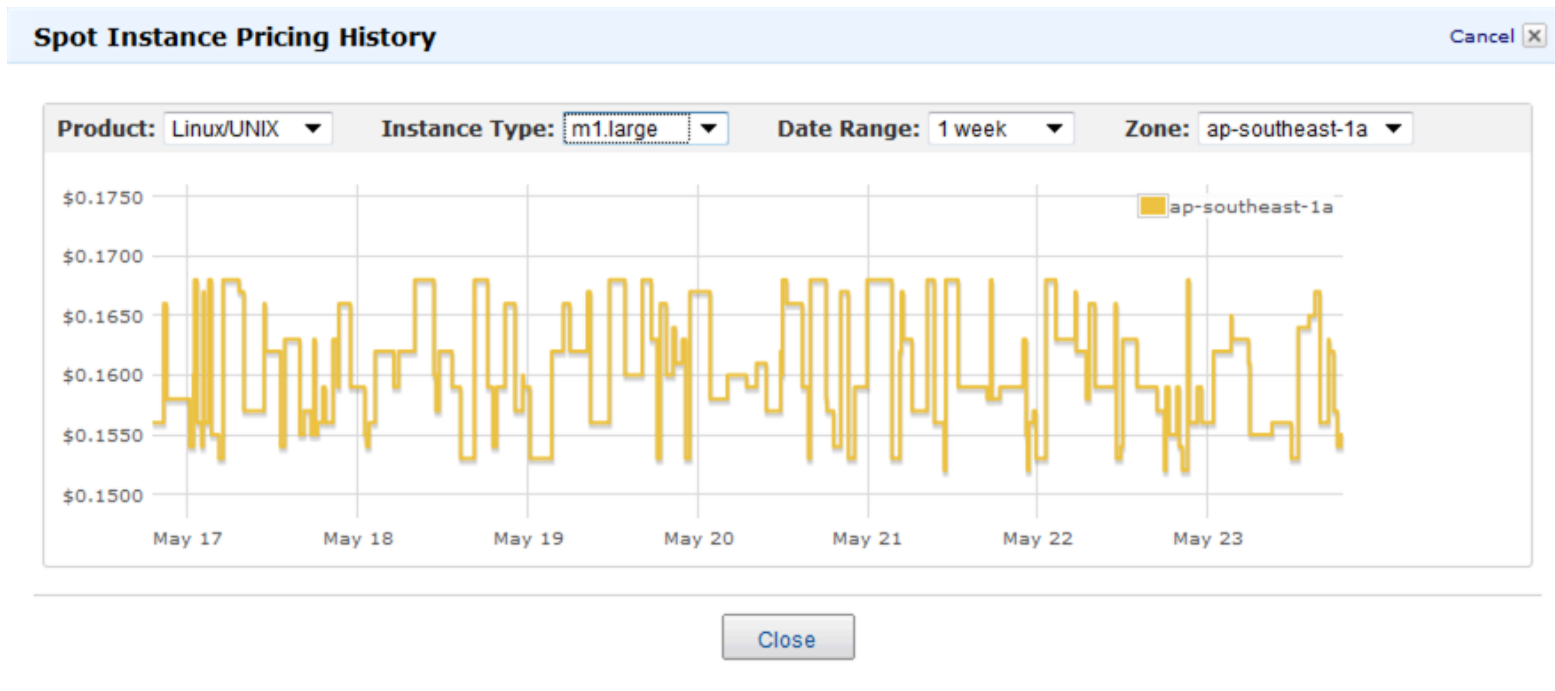
Esquema de Precios de EC2

(I)

- Los precios dependen de la región. Ejemplo para EU Ireland:
- Instancias bajo demanda:
 - Desde 0.020\$/hora para una instancia t1.micro hasta 3.410\$/hora para una instancia *Quadruple Extra Large High-I/O On-Demand instance*.
- Las instancias basadas en Windows cuestan algo más (entre un 10% y un 70%) que las basadas en Linux.
- Instancias reservadas:
 - A 1 año y a 3 años, permite obtener reducción significativa (más del 50%) del precio por hora de las instancias.
- <http://aws.amazon.com/ec2/pricing/>

Esquema de Precios de EC2 (II)

- Es posible reducir todavía más el precio usando *spot instances*:
 - Precio por hora variable a lo largo del tiempo.



Conclusiones

- Amazon EC2 permite desplegar instancias de cómputo en la nube.
- Acceso *root* (superusuario) a las instancias para posteriormente ser configuradas.
- Gran variedad de tipos de instancia, con diferente coste, para adaptarse a diferentes escenarios.
- Posibilidad de mayores ahorros con *reserved instances* y *spot instances*.

Curso Online de Cloud Computing con Amazon Web Services



Otros Servicios de Amazon EC2 para Aplicaciones Cloud

Germán Moltó

Instituto de Instrumentación para
Imagen Molecular

gmolto@dsic.upv.es

<http://www.grycap.upv.es/cursocloud>



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Resultados de Aprendizaje

- Se espera que una vez acabes esta presentación seas capaz de:
 - Conocer el mecanismo de instantáneas de volúmenes EBS para copias de seguridad.
 - Comprender la importancia de las IPs elásticas para poder cambiar dinámicamente la instancia existente tras una dirección IP.
 - Entender el proceso de balanceo de carga llevado a cabo por ELB.

EBS Snapshots/ Instantáneas (I)

- Amazon EBS (Elastic Block Store) permite la creación de instantáneas (snapshots) de un volumen EBS.
 - Provoca la copia de los datos del volumen en S3, donde se guardan de forma redundante en varias zonas de disponibilidad.
- Se pueden utilizar los snapshots para:
 - Crear un nuevo volumen EBS a partir de alguno existente.
 - Aumentar el tamaño de un volumen
 - Mover volúmenes entre diferentes zonas de disponibilidad.

EBS Snapshots (II)

- Los snapshots son copias de seguridad incrementales.
 - Solo los datos modificados en el volumen desde el último snapshot se almacenan en el nuevo snapshot.
 - Si tenemos un dispositivo con 100 GB de datos, pero solo se han modificado 5 GB de datos desde la última instantánea, solo se restauran estos 5 GB adicionales de la instantánea en Amazon S3

EBS Snapshots (III)

- Eliminación de EBS Snapshots
 - Si se elimina una instantánea solo se eliminan los datos que no se necesitan para ninguna otra instantánea.
 - Independientemente de las instantáneas anteriores que se hayan suprimido, todas las instantáneas activas contienen la información necesaria para restaurar el volumen.
- El tiempo necesario para restaurar el volumen es el mismo para todas las instantáneas.

EBS Snapshots (IV)

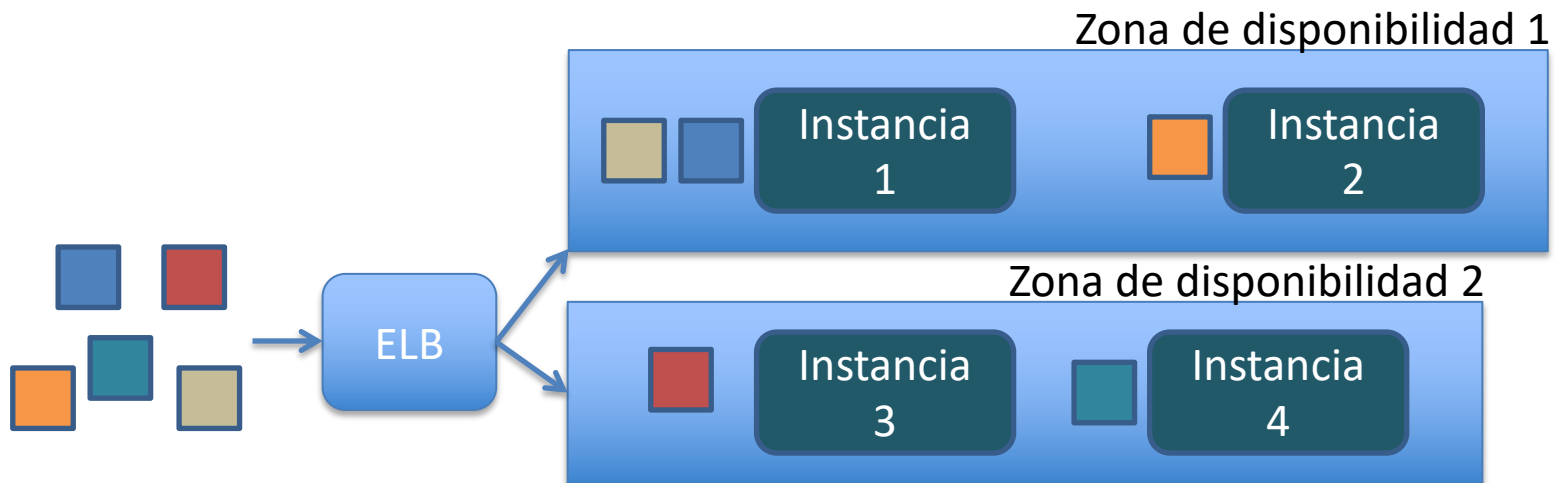
- Las instantáneas compartidas de Amazon EBS permiten la compartición de datos entre usuarios de AWS
 - Los usuarios autorizados pueden utilizar instantáneas compartidas de Amazon EBS como base para crear sus propios volúmenes de Amazon EBS
 - Es posible crear instantáneas públicas.
- Los usuarios a los que ha otorgado acceso pueden crear sus propios volúmenes de EBS partiendo de la instantánea
 - La instantánea original permanece intacta.

IPs elásticas

- Las direcciones IP elásticas son IPs estáticas que van ligadas a una cuenta de usuario AWS y que permiten ser asociadas temporalmente a instancias.
- Esto permite mapear la IP a instancias diferentes cada vez
 - Por ejemplo, utilizar la IP elástica para acceder a una instancia de un sitio web y, ante detección de caída de la instancia, asignar la IP elástica a otra instancia de respaldo, para seguir manteniendo el servicio.
- La IP elástica se resuelve desde dentro de EC2 a la IP privada de la instancia y a la IP pública desde fuera.
 - Permite reducir costes de transferencias entre instancias (las transferencias entre instancias no generan gasto)
- Al parar la instancia, se desasocia la IP elástica.

Balancedores de Carga

- El escalado de una aplicación en Cloud permite aumentar el número de instancias dedicadas a procesar el tráfico de entrada.
- Éste debe ir acompañado de un adecuado balanceo de la carga para distribuir las peticiones entre todas las instancias disponibles.



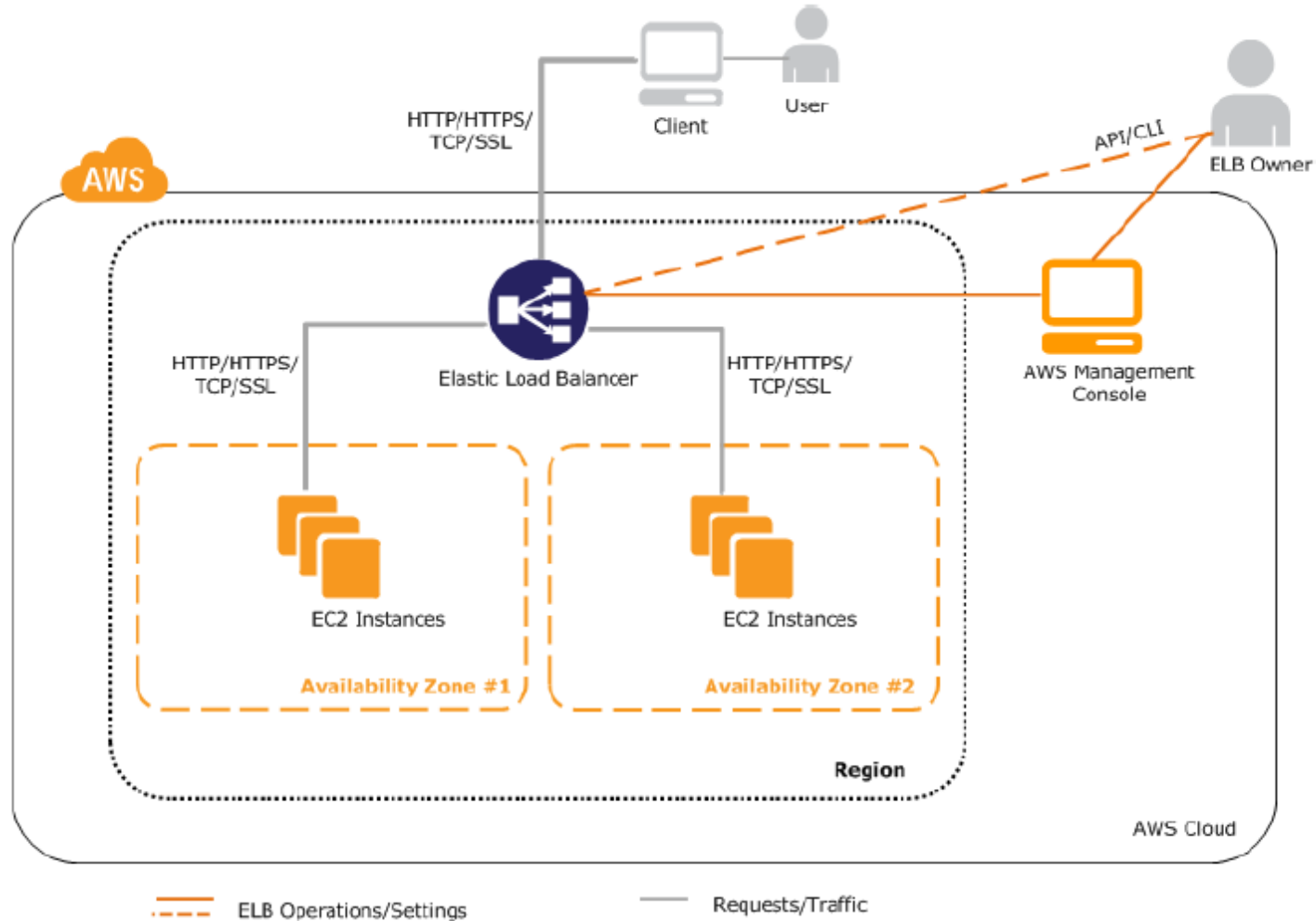
Elastic Load Balancing (ELB)

- Permite distribuir el tráfico entrante entre diferentes instancias de EC2 para repartir la carga de trabajo.
- Ofrece los siguientes servicios:
 - Distribución de peticiones entre instancias en la misma zona de disponibilidad (y a otras zonas de disponibilidad de la misma región, no entre diferentes regiones).
 - Monitorización de la salud de las instancias de EC2 registradas en el balanceador de la carga (las peticiones sólo se envían a las instancias sanas).
 - *Sticky session* (sesiones de usuario mantenidas en la misma instancia de EC2).

Elastic Load Balancing (II)

- El ELB distribuye el tráfico de forma equitativa entre diferentes zonas de disponibilidad (independientemente del número de instancias en cada zona de disponibilidad).
 - Es importante tener un número de instancias aproximadamente similar en cada zona de disponibilidad.
- Utiliza una aproximación tipo *round-robin* entre las zonas de disponibilidad (no entre las instancias)
 - NO usa esquemas sofisticados del estilo “envía las peticiones a aquella instancia con una menor carga de CPU”, etc.
- Una misma instancia puede ser registrada en diferentes balanceadores de carga.

Arquitectura del ELB



Conclusiones

- El servicio Amazon EC2 incluye otros sub-servicios como es el caso de:
 - EBS (Elastic Block Store), para la creación de volúmenes e instantáneas de volúmenes.
 - ELB (Elastic Load Balancing), para la creación de balanceadores de carga que distribuyan peticiones entre zonas de disponibilidad.
 - Elastic IP, para crear IPs públicas que dinámicamente se pueden asociar a instancias desplegadas en EC2.

Curso Online de Cloud Computing con Amazon Web Services



Auto-Escalado de Aplicaciones Cloud

Germán Moltó

Instituto de Instrumentación para
Imagen Molecular

gmolto@dsic.upv.es

<http://www.grycap.upv.es/cursocloud>



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Resultados de Aprendizaje

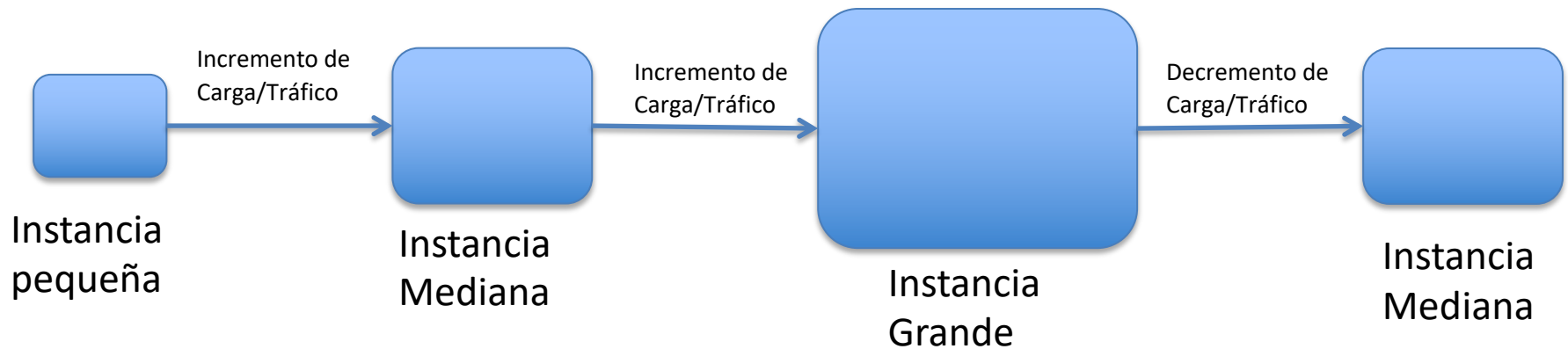
- Se espera que una vez acabes esta presentación seas capaz de:
 - Conocer el procedimiento habitual de escalado de aplicaciones en el Cloud.
 - Diferenciar entre elasticidad vertical y horizontal.
 - Conocer aspectos generales sobre el servicio Auto Scaling
 - Entender el concepto de reglas de elasticidad y su aplicabilidad al escalado automático.

Escalado de Aplicaciones en AWS (I)

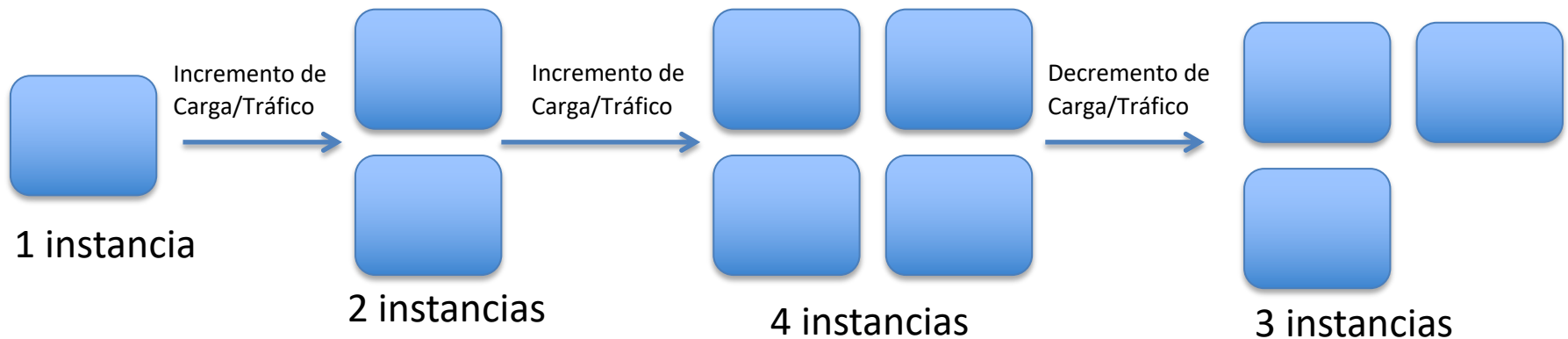
- Las aplicaciones en el Cloud deben ser escalables de manera que se adapten a los incrementos o decrementos de requisitos de cómputo.
 - *Trade-off* entre ofrecer calidad de servicio y ajustar el gasto económico.
- Ejemplos:
 - Incremento de instancias de un servidor web Apache ante un aumento repentino del tráfico recibido.
 - Una aplicación científica de cómputo masivo tipo High Throughput Computing (HTC) debería poder aprovisionar y liberar recursos virtuales de forma dinámica durante la ejecución.

Escalado de Aplicaciones en AWS (II)

- Escalado Vertical (Scale Up / Scale Down)



- Escalado Horizontal (Scale Out / Scale In)



Escalado Vertical (I)

- Permite aumentar/reducir la capacidad de una máquina virtual para proveer:
 - Ejemplo: m1.small <--> m1.large
- No es posible modificar en caliente el tipo de instancia, hay que lanzar una nueva instancia del tipo deseado con la AMI (o simularlo con ELB).
 - Sí es posible detener una instancia basada en EBS, cambiar el tipo y volver a iniciarla.

Escalado Horizontal (I)

- Permite reducir/aumentar el número de instancias de forma automática y transparente para la aplicación, que debe estar diseñada para soportar esta característica.
 - Auto Scaling permite ofrecer servicios de elasticidad a las aplicaciones Cloud, principalmente a aquellas que “experimentan patrones de uso variable a lo largo del tiempo (horario, diario, semanal, etc.)”

Auto Scaling

- Auto Scaling consta de:
 - **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.
- Se utiliza en conjunción con ELB
 - Balanceo de carga entre instancias de las zonas de disponibilidad definidas en el ELB.

Plan de Escalado

- Auto Scaling permite a un grupo de auto-escalado:
 - Mantener un número fijo de instancias EC2 activas.
 - Escalar bajo demanda
 - En función de los requisitos de cómputo de la aplicación.
 - Escalar de acuerdo al calendario
 - Ante eventos extraordinarios, o de forma recurrente (por las noches, etc.).

Escalado Bajo Demanda

- Esto permite definir reglas de elasticidad del estilo:
 - 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.
 - Otros muchos factores: consumo de disco, memoria libre, etc.

Conclusiones

- El servicio Auto Scaling permite la creación de flotas elásticas de instancias cuyo tamaño varíe en función de ciertas reglas de escalado.
- Posibilita la elasticidad horizontal de aplicaciones Cloud, que escalan ante incrementos en la carga de trabajo (o a intervalos específicos o fechas puntuales).
- Auto Scaling depende de CloudWatch, el servicio de monitorización de AWS.

Curso Online de Cloud Computing con Amazon Web Services



Utilizando Amazon Web Services (AWS)

Germán Moltó

Instituto de Instrumentación para
Imagen Molecular

gmolto@dsic.upv.es

<http://www.grycap.upv.es/cursocloud>



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Resultados de Aprendizaje

- Se espera que una vez acabes esta presentación seas capaz de ...
 - Conocer los pasos necesarios para comenzar a utilizar AWS.
 - Entender las ventajas aportadas por la AWS Management Console
 - Distinguir entre las diferentes credenciales de acceso utilizadas en AWS.
 - Conocer herramientas para estimar el coste de migrar al Cloud.

Como Utilizar AWS

- Para poder utilizar AWS es necesario:
 - Crear una cuenta en AWS (Tarjeta de Crédito).
 - Activar los servicios utilizados, Amazon S3, Amazon EC2, etc.
- Comenzar a utilizar el “Cloud”.
 - <http://aws.amazon.com>
- Límite de 20 instancias concurrentes para nuevos usuarios.
 - Contactar con Amazon para incrementar el límite.



AWS Free Usage Tier

- AWS ofrece 1 año de suscripción gratuita para probar todos los servicios con ciertos límites.
- Es posible monitorizar el consumo y establecer alertas para no superar los límites.

AWS Free Usage Tier (Per Month):

- 750 hours of [Amazon EC2](#) Linux Micro Instance usage (613 MB of memory and 32-bit and 64-bit platform support) – enough hours to run continuously each month*
- 750 hours of [Amazon EC2](#) Microsoft Windows Server Micro Instance usage (613 MB of memory and 32-bit and 64-bit platform support) – enough hours to run continuously each month*
- 750 hours of an [Elastic Load Balancer](#) plus 15 GB data processing*
- 30 GB of [Amazon Elastic Block Storage](#), plus 2 million I/Os and 1 GB of snapshot storage*
- 5 GB of [Amazon S3](#) standard storage, 20,000 Get Requests, and 2,000 Put Requests*
- 100 MB of storage, 5 units of write capacity, and 10 units of read capacity for [Amazon DynamoDB](#).**
- 25 [Amazon SimpleDB](#) Machine Hours and 1 GB of Storage**
- 1,000 [Amazon SWF](#) workflow executions can be initiated for free. A total of 10,000 activity tasks, signals, timers and markers, and 30,000 workflow-days can also be used for free**
- 100,000 Requests of [Amazon Simple Queue Service](#)**
- 100,000 Requests, 100,000 HTTP notifications and 1,000 email notifications for [Amazon Simple Notification Service](#)**
- 10 [Amazon Cloudwatch](#) metrics, 10 alarms, and 1,000,000 API requests**
- 15 GB of bandwidth out aggregated across all AWS services*
- 750 hours of [Amazon RDS](#) for SQL Server Micro DB Instance usage (running SQL Server Express Edition in a single Availability Zone) – enough hours to run a DB Instance continuously each month
- 20 GB of database storage
- 10 million I/Os
- 20 GB of backup storage for your automated database backups and any user-initiated DB Snapshots

In addition to these services, the [AWS Management Console](#) is available at no charge to help you build and manage your application on AWS.

AWS Management Console

AWS Service Charges		\$0.01
<div> <div></div> Amazon Elastic Compute Cloud </div>		\$0.00
Download Usage Report »		
EDU_Molto_UPValenciaCourse_Spring2012		Credit
		-7.73
		-7.73
US East (Northern Virginia) Region		
Amazon EC2 running Linux/UNIX		
\$0.080 per Small Instance (m1.small) instance-hour (or partial hour)	4 Hrs	0.32
\$0.020 per Micro Instance (t1.micro) instance-hour (or partial hour)	5 Hrs	0.10
\$2.400 per Cluster Compute Eight Extra Large Instance (cc2.8xlarge) instance-hour (or partial hour)	3 Hrs	7.20
Amazon EC2 EBS		
\$0.10 per GB-month of provisioned storage	0.156 GB-Mo	0.02
\$0.10 per 1 million I/O requests	45,462 IOs	0.01
\$0.125 per GB-Month of snapshot data stored	0.303 GB-Mo	0.04
Elastic Load Balancing		
\$0.008 per GB Data Processed by the LoadBalancer	0.000192 GB	0.01
\$0.025 per LoadBalancer-hour (or partial hour)	1 Hr	0.03
Amazon CloudWatch		
\$0.00 per alarm-month - first 10 alarms	0.611 Alarms	0.00
		7.73
<div> <div></div> Amazon Simple Notification Service </div>		\$0.00
Download Usage Report »		
<div> <div></div> Amazon Simple Storage Service </div>		\$0.00
Download Usage Report »		
EDU_Molto_UPValenciaCourse_Spring2012		Credit
		-0.02
		-0.02
US Standard Region		
\$0.125 per GB - first 1 TB / month of storage used	0.000003 GB-Mo	0.00
\$0.01 per 1,000 PUT, COPY, POST, or LIST requests	19 Requests	0.01
\$0.01 per 10,000 GET and all other requests	24 Requests	0.01
		0.02
<div> <div></div> AWS Data Transfer (excluding Amazon CloudFront) </div>		\$0.00
<div> <div></div> VAT to be collected </div>		\$0.01

- Herramientas para:
 - Gestionar casi todos los diferentes servicios de AWS.
 - El resto por CLI.
 - Monitorizar la actividad
 - Supervisar el gasto.
 - Obtener reportes de uso.
 - Gestionar la identidad
 - Mediante certificados, tokens y *secrets*.
 - Obtener facturas y extractos de gastos.

Credenciales de Acceso

- En AWS hay cuatro tipos de credenciales de acceso y es muy importante saber diferenciarlas.
 - Claves de Acceso (*Access Keys*)
 - Certificados X.509 (*X.509 certificates*)
 - Pares de Clave (*Key Pairs*)
 - Usuario y contraseña
- Las dos primeras (*Access Keys* y *X.509 certificates*) sirven para autenticar al usuario frente a los servicios de AWS.
- Los *Pares de Clave* sirven para poder conectarse mediante SSH sin contraseña a las instancias lanzadas en EC2.
- El usuario y contraseña sirve para conectarse a la AWS Management Console.

Credenciales de Acceso:

Access Keys

- Las Claves de Acceso o *Access Keys* sirven para poder utilizar las APIs de servicios de AWS basados en REST o el protocolo Query.
- Tienen el siguiente aspecto:
 - ACCESS_KEY_ID: AKIAMOJJF5QACV8AQM5R
 - SECRET_KEY_ID:
2ry0B2GM8UVM815MI26JPz+xPhKCYXM6nxbLLVKV
- Existe una tendencia actual a utilizar estas credenciales frente a los certificados X.509
 - En las prácticas del curso se utilizan las *Access Keys* para ser autenticados antes los servicios de AWS.

Credenciales de Acceso: Certificados X.509 (I)

- Los certificados X.509 se utilizan para realizar peticiones seguras mediante el protocolo SOAP a las APIs de servicios de AWS.
 - Constan de un certificado y una clave privada.
- Tienen el siguiente aspecto:
 - Fichero cert-2VMS3RA626TTGG8IUFFFFRWLWGVRQ5Q6W.pem con el contenido:
-----BEGIN CERTIFICATE-----
MIICdjCCAd+gAwIBAgIEGQusEz
...
-----END CERTIFICATE-----

Credenciales de Acceso: Certificados X.509 (II)

- Fichero pk-2VMS3RA626TTGG8IUFFFFRWLWGVRQ5Q6W.pem con el contenido:

```
-----BEGIN PRIVATE KEY-----
```

```
MIICdQIBADANBgkqhkiG9w0B
```

```
...
```

```
-----END PRIVATE KEY-----
```

- Se utilizan para las EC2 API Tools (herramienta de línea de comandos para interactuar con EC2).
 - <http://aws.amazon.com/developertools/351>

Credenciales de Acceso:

Pares de Clave (I)

- Los Pares de Clave (Key Pairs) se usan para poder acceder a las instancias mediante SSH sin contraseña.
- AWS permite la creación de diferentes pares de clave, incluso importar aquellos ya existentes del usuario.
 - Los pares de clave van ligados a una región AWS concreta.
- Consta de una clave pública que se almacena en AWS y que será inyectada en la instancia en el arranque para autorizar la conexión a aquel usuario que disponga de la clave privada.

Credenciales de Acceso:

Pares de Clave (II)

- La clave privada la guarda el usuario en un fichero con el siguiente aspecto:
 - Fichero id_rsa-gmolto_eu_west_1

```
-----BEGIN RSA PRIVATE KEY-----
MIIEpAIBAAKC
...
-----END RSA PRIVATE KEY-----
```

Gestión Apropiable de Credenciales

1. El usuario crea una cuenta de AWS con cargo a su tarjeta de crédito.
2. Posteriormente utiliza el servicio IAM (Identity and Access Management) para crear grupos de usuario
 - Administrador
 - ... (por ejemplo grupo de alumnos del curso de Cloud)
3. Asignación de políticas de permisos a grupos
 - <http://awspolicygen.s3.amazonaws.com/policygen.html>
4. Creación de cuentas de usuario y asignación a grupos
5. Obtención de las *Access Keys* para poder usar los servicios de AWS.

Amazon Simple Monthly Calculator

- <http://aws.amazon.com/es/calculator/>
- Ejemplo:
 - 3 instancias *m1.small* activas durante 8 horas al día con un almacenamiento de 20 GB en S3 donde se presupone una descarga de 5 GB/mes.
 - 64.10\$/mes

Estado del Servicio de AWS

- <http://status.aws.amazon.com>



[Amazon Web Services](#) » Service Health Dashboard

Current Status - Apr 11, 2013

Amazon Web Services publishes our most up-to-the-minute information on service availability in the table below. Check back here any time to get current status information, or subscribe to an RSS feed to be notified of interruptions to each individual service. If you are experiencing a real-time, operational issue with one of our services that is not described below, please inform us by clicking on the "Contact Us" link to submit a service issue report.








North America

South America

Europe

Asia Pacific

Contact Us

Current Status	Details	RSS
<div>✓</div> <div>Amazon CloudFront</div>	Service is operating normally.	
<div>✓</div> <div>Amazon CloudSearch (N. California)</div>	Service is operating normally.	
<div>✓</div> <div>Amazon CloudSearch (N. Virginia)</div>	Service is operating normally.	
<div>✓</div> <div>Amazon CloudSearch (Oregon)</div>	Service is operating normally.	
<div>✓</div> <div>Amazon CloudWatch (N. California)</div>	Service is operating normally.	
<div>✓</div> <div>Amazon CloudWatch (N. Virginia)</div>	Service is operating normally.	
<div>✓</div> <div>Amazon CloudWatch (Oregon)</div>	Service is operating normally.	

Ecosistema alrededor de AWS

- <http://thecloudmarket.com>
 - Clasificación de AMIs por proveedor, con comentarios de usuarios.



the
cloud
market

[Home](#) | [Signup](#) | [Blog](#) | [Login](#)

Featured Owners

Popular Images

Recently Added

Recently Updated

BITNAMI

SECLUD 
SAVING YOUR CLOUD A SAFE HEAVEN



alestic

ANALYTICS(1305)

The Cloud Market is the most complete catalog of Amazon EC2 images

Now indexing **66447** available images

Check the [Statistics](#)  page



Usando AWS

- Es posible utilizar la mayor parte de AWS mediante la AWS Management Console, desde un navegador.
- También es posible utilizar una CLI (interfaz de línea de comandos) para interactuar con los servicios de AWS.
 - Herramienta *aws* desarrollada por Tim Kay
 - Un único script de Perl portable, usa las *Access Keys*, permite interactuar con EC2, S3 y otros servicios de AWS.

Conclusiones

- AWS ofrece diferentes mecanismos de autenticación del usuario: Certificados X.509 y Access Keys.
- Los Key Pairs sirven para acceder mediante SSH sin contraseña a las instancias Linux desplegadas en EC2.
- AWS Free Usage Tier permite probar los servicios de AWS (con restricciones) durante 1 año de forma gratuita.

Referencias

- <http://aws.amazon.com/documentation/ec2/>
- <http://aws.amazon.com/documentation/s3/>
- http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts_micro_instances.html
- http://www.aifb.kit.edu/images/1/17/How_soon_is_eventual.pdf