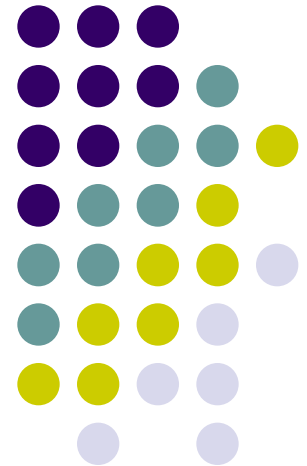
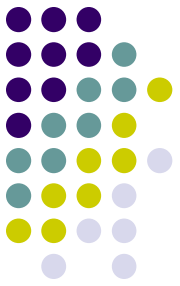




---

Roberto Díaz Morales  
rdiazm@tsc.uc3m.es  
MLG 24/06/2013





# Indice

- Amazon AWS
  - Servicios
  - Capa gratuita
- Amazon S3
- Amazon Elastic MapReduce
- Amazon Elastic Compute Cloud
- Amazon EBS

# Amazon AWS







- Amazon Web Services es un conjunto de servicios de infraestructura y aplicaciones en “la nube”.
- Las principales ventajas son:
  - La facturación por recursos utilizados y horas de uso (no hay consumo mínimo), lo que evita un desembolso inicial en infraestructuras.
  - Escalabilidad: Se pueden ampliar o reducir las infraestructuras utilizadas de una forma muy sencilla.
  - La existencia de una capa gratuita.

# Servicios:



## Informática y redes

-  [AWS Direct Connect »](#)
-  [Amazon EC2 »](#)
-  [Amazon EMR »](#)
-  [Amazon Route 53 »](#)
-  [Amazon VPC »](#)





## Almacenamiento y entrega de contenido

-  [Amazon CloudFront »](#)
-  [Amazon Glacier »](#)
-  [Amazon S3 »](#)
-  [AWS Storage Gateway »](#)

## Base de datos

-  [Amazon DynamoDB »](#)
-  [Amazon ElastiCache »](#)
-  [Amazon RDS »](#)
-  [Amazon Redshift »](#)

## Implementación y gestión

-  [AWS CloudFormation »](#)
-  [Amazon CloudWatch »](#)
-  [AWS Data Pipeline »](#)
-  [AWS Elastic Beanstalk »](#)
-  [AWS IAM »](#)
-  [AWS OpsWorks »](#)

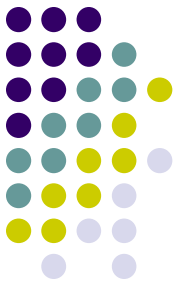
## Servicios de aplicaciones

-  [Amazon CloudSearch »](#)
-  [Amazon SQS »](#)
-  [Amazon SNS »](#)
-  [Amazon SES »](#)
-  [Amazon SWF »](#)
-  [Amazon Elastic Transcoder »](#)



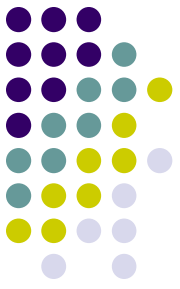
# Capa gratuita:

- Amazon Elastic Compute Cloud (EC2): 750 horas mensuales de instancias linux.
- Amazon Elastic Block Storage: 30 GB de almacenamiento y dos millones de E/S.
- Simple Storage Service: 5 GB de almacenamiento estándar.
- Muchos otros: Bases de datos, notificaciones, ....



# ¿Por donde empezar?

- Se debe crear una una cuenta AWS:
  - <http://aws.amazon.com/es/>
- Datos personales, cuenta bancaria, ... Al terminar el proceso recibirás una llamar llamada telefónica donde tendrás que indicar un código que han enviado a tu dirección de correo.



# Amazon S3

- **La capa gratuita da 5 GB de almacenamiento, 15 GB de transferencia, 2000 operaciones PUT, 20000 operaciones GET.**
- **A partir de ahí ahí los precios varían dependiendo de la región y el sistema de almacenamiento:**
  - **Aprox. 0.09\$/mes cada GB si quieres realizar muchas operaciones de subida/bajada.**
  - **Aprox. 0.01\$/mes si el contenido no va a sufrir apenas cambios.**
- **Es necesaria su utilización para otros servicios como elastic mapreduce.**

# Amazon S3



- La información se almacena en “buckets”
- Dentro de cada “bucket” se pueden crear directorios y subir/bajar ficheros.
- Vamos a subir ahora los ficheros necesarios para poder ejecutar más adelante elastic mapreduce...

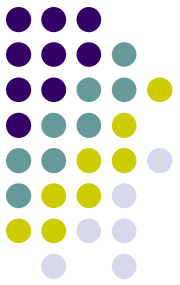


# AMAZON EMR



- Es la implementación mapreduce de amazon.
- Se utilizan las instancias de amazon EC2, su precio es aproximadamente un 25% que reservar esas mismas instancias en amazon EC2.
- Permite sistema de subastas para adquirir instancias.

# AMAZON EMR



- **Se puede utilizar a través de la consola web.**
- **1-Crear un nuevo “Job Flow”**
  - **Introducir el nombre**
  - **Seleccionar si se desea la distribución de amazon u otra versión de hadoop diferente.**
  - **Seleccionar “run your own application”**
  - **Job Type:**
    - **Streaming->Ruby,PHP,R,C,Python,...**
    - **Jar-> Java**
    - **...**

# AMAZON EMR



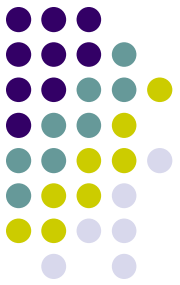
- Input location:  
s3n://**bucket**/dirDatos/ficheroDatos
- Output location: s3n://**bucket**/dirOutput/
  - Esta carpeta debe estar vacía.
- Mapper s3n://**bucket**/dirFunction/**map.py**
- Reducer: aggregate

# AMAZON EMR



- Se selecciona el tipo de instancia maestro (m1.small)
- Se selecciona el número y tipo de instancias esclavos.
- Se pueden pedir instancias spot: Se pone un precio de puja, las instancias libres se reparten entre los mayores pujadores.

# AMAZON EMR CLI



- Existen SDKs para varios lenguajes de programación y la posibilidad de ejecutar todo por línea de comandos en nuestro PC.
- Es necesario instalar ruby (versión 1.8.7, ni anterior ni posterior). Añadir ejecutables al path.
- Descargar la command line interface:
  - <http://aws.amazon.com/developertools/2264>
- Descargar y descomprimir el zip.

# AMAZON EMR CLI



- **Crear** el fichero credentials.json.
- {  
  "access\_id": "acceso\_id",  
  "private\_key": "clave privada\_id",  
  "keypair": "clave\_com",  
  "key-pairfile": "clave.pem",  
  "log\_uri": "s3n://bucket/logs",  
  "region": "eu-west-1"  
}

# AMAZON EMR CLI



- ACCESS\_ID y PRIVATE\_KEY son las claves publica y privada para autenticación.
- Se obtienen en “My Account” -> “Security credentials”
- EL “Key Pair” se obtiene desde la cuenta de amazon EC2, son la clave publica y privada de comunicación.
  - Depende de la zona (USA, Irlanda, ...)
  - Si se pierde el fichero de clave privada no es recuperable.

# AMAZON EMR CLI



```
>>Ruby elastic-mapreduce --create --alive
```

Created job flow **JobFlowID**

```
>>ruby elastic-mapreduce --j JobFlowID --stream --  
  mapper s3n://.... --input s3n://... --output s3n://... --  
  reducer aggregate
```

```
>>ruby elastic-mapreduce --terminate JobFlowID
```

Para tener más instancias, a la hora de crear el job flow:

```
--instance-group master --instance-type m1.large --  
  instance-group core --instance-type m1.large --instance-  
  count 2 --instance-group task --instance-type m1.small --  
  instance-count 4 --bid-price 0.03
```



# AMAZON EC2



- Proporciona capacidad informática
- Proporciona recursos informáticos escalables.
- En resumen, nos asignan servidores con los cores, gpus, memoria y espacio de almacenamiento que necesitemos (o que podemos pagar).
- La capa gratuita nos da 750 horas/mes de servidores linux con 600 MB de ram y 8 GB de disco.

# AMAZON EC2 Instancias



- Primera generación M1
  - Small: RAM 1.7GB, 1núcleo, 160GB disco 0.06\$
  - Xlarge: RAM 15GB, 4núcleos, 1690GB disco. 0.5\$
- Segunda generación M3
  - Xlarge doble: 30 GB, 8núcleos. 1.1\$
- Memoria elevada
- CPU elevada
- Instancias en clúster 2.7 \$
- De GPU: 2 GPUs NVIDIA Tesla “Fermi” M2050 2.36\$
- Existe sistema de subasta.



# AMAZON EC2 CLI

- Existen SDKs para varios lenguajes de programación.
- Para el manejo mediante línea de comandos está la interfaz CLI:
  - <http://aws.amazon.com/developertools/351>
- En este caso no hay un fichero de configuración, los parámetros se configuran por variables de entorno.

# AMAZON EC2



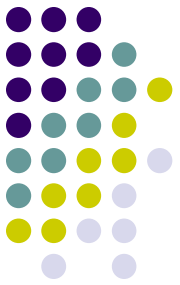
- Para generar una instancia entrar en la consola web, seleccionar Instances en el menú y después “Launch Instance”.
- Seleccionar un Key Pair existente o crear uno nuevo.
- Seleccionar una configuración (La estrella significa que la configuración es gratis).
- Lanzar.

# AMAZON EC2



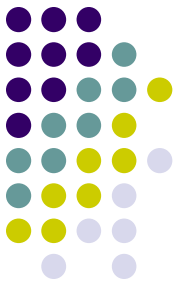
- Para conectarse a la instancia ssh o sftp.
- Programas como Putty o WinSCP no admiten el fichero de clave pública de amazon. Descargar PuttyGen para convertir el fichero .pem en .ppk.
- Para conectar mediante Putty seleccionar el fichero ppk en el menú SSH -> AUTH
- Usuario de la instancia ec2-user

# AMAZON EC2



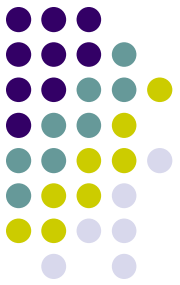
- Tras instalar el software que necesitemos, crearemos una imagen. (Menú Instances, Actions, create Image).
- Para lanzar un cluster de instancias en el menú AMI seleccionamos la instancia que queremos y seleccionamos “launch”.
- Podemos indicar el número de instancias y su tipo (número de cores, ram y disco de cada una).

# AMAZON EBS



- Se puede añadir más capacidad de almacenamiento a las instancias mediante amazon EBS. (Hasta 30 GB la capa gratuita)
- Menú Volumes/Create Volume
- Los snapshots son volúmenes ya creados por alguien con un conjunto de datos de uso público (genoma humano, 1 million songs,...).

# Fin



- Para más información:
  - <http://aws.amazon.com>
  - <http://aws.amazon.com/es/documentation/elastic-mapreduce/>
  - <http://aws.amazon.com/es/documentation/ec2/>