



U N I V E R S I D A D
COMPLUTENSE
M A D R I D

Big Data en la nube

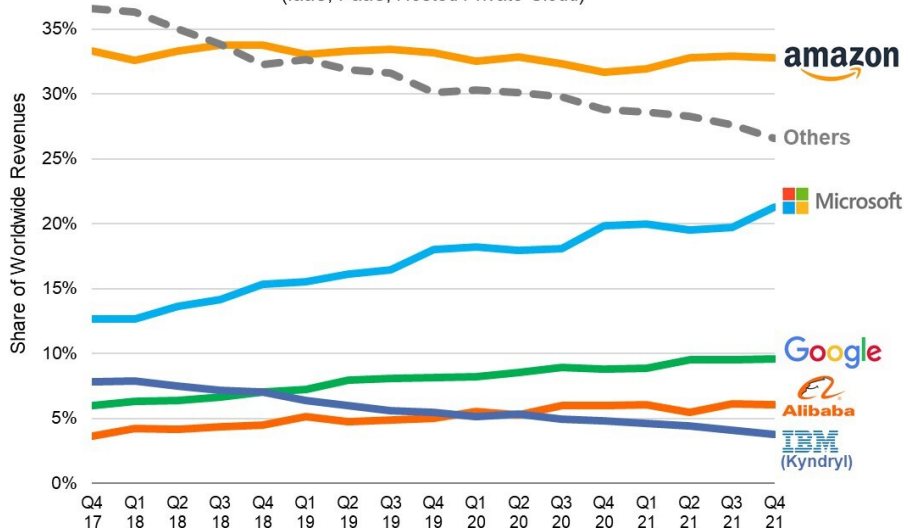
Enrique Martín (emartinm@ucm.es)
Sistemas de Gestión de Datos y de la Información
Master Ing. Informática
Fac. Informática

- A la hora de realizar cálculos Big Data, no siempre se tiene disponible un **clúster físico** del tamaño requerido
- En estos casos se acude a servicios en la nube que proporcionan máquinas que se pueden **crear y eliminar** de manera rápida
- Permite escalar los recursos a las necesidades de proceso (**elasticidad**)
- Hay varios proveedores de Big Data (almacenamiento y procesamiento) como servicio en la nube:
 - Amazon Web Services (AWS)
 - Microsoft Azure
 - Google Cloud Platform
- La tarificación depende principalmente del consumo (horas de cómputo, transferencia de datos, cantidad de almacenamiento)

Proveedores de cómputo en la nube

Cloud Provider Market Share Trend

(IaaS, PaaS, Hosted Private Cloud)



Source: Synergy Research Group

Amazon Web Services (AWS)

- Amazon EC2: servicio de infraestructura (IaaS), permite disponer de máquinas virtuales con Linux o Windows
- Amazon EC2 Container Service: servicio para ejecutar contenedores en instancias EC2
- Amazon Simple Storage Service (S3): almacenamiento en la nube seguro, duradero y escalable
- Amazon DynamoDB: BD NoSQL de clave-valor y documentos
- Amazon EMR: instancias configuradas Spark, Hive, etc.

Precios AWS EC2 en octubre de 2022

Nombre de la instancia ▲	Tarifa por hora bajo demanda ▼	vCPU ▼	Memoria ▼	Almacenamiento ▼	Rendimiento de la red ▼
a1.medium	0,0288 USD	1	2 GIB	Solo EBS	Hasta 10 gigabits
a1.large	0,0576 USD	2	4 GIB	Solo EBS	Hasta 10 gigabits
a1.xlarge	0,1152 USD	4	8 GIB	Solo EBS	Hasta 10 gigabits
a1.2xlarge	0,2304 USD	8	16 GIB	Solo EBS	Hasta 10 gigabits
a1.4xlarge	0,4608 USD	16	32 GIB	Solo EBS	Hasta 10 gigabits
a1.metal	0,461 USD	16	32 GIB	Solo EBS	Hasta 10 gigabits
t4g.nano	0,0046 USD	2	0,5 GIB	Solo EBS	Hasta 5 gigabits
t4g.micro	0,0092 USD	2	1 GIB	Solo EBS	Hasta 5 gigabits
t4g.small	0,0184 USD	2	2 GIB	Solo EBS	Hasta 5 gigabits
t4g.medium	0,0368 USD	2	4 GIB	Solo EBS	Hasta 5 gigabits
t4g.large	0,0736 USD	2	8 GIB	Solo EBS	Hasta 5 gigabits
t4g.xlarge	0,1472 USD	4	16 GIB	Solo EBS	Hasta 5 gigabits

Fuente: <https://aws.amazon.com/es/ec2/pricing/on-demand/>

Precios Amazon S3 en octubre de 2022

Región:

Europa (Irlanda) ↕

Precios de almacenamiento

S3 Standard: almacenamiento de propósito general para cualquier clase de datos que se utiliza generalmente para datos a los que se accede con frecuencia

Primeros 50 TB/mes

0,023 USD por GB

Siguientes 450 TB/mes

0,022 USD por GB

Más de 500 TB/mes

0,021 USD por GB

Fuente: <https://aws.amazon.com/es/s3/pricing/>

- Compute Engine: servicio de infraestructura (IaaS), permite disponer de máquinas virtuales ejecutando Linux o Windows
- Container Engine: ejecución de contenedores utilizando Kubernetes
- Google Cloud Storage: servicio para almacenar objetos de manera persistente, con control de acceso

Precios Google Compute Engine en octubre de 2022

Tipos de máquinas estándares N1

Países Bajos (europe-west4) ▾		Por mes <input checked="" type="radio"/> Por hora <input type="radio"/>		
Tipo de máquina	CPU virtuales	Memoria	Precio (en USD)	Precio de las interrumpibles (en USD)
n1-standard-1	1	3.75GB	\$0.0523	\$0.01100
n1-standard-2	2	7.5GB	\$0.1046	\$0.02210
n1-standard-4	4	15GB	\$0.2092	\$0.04410
n1-standard-8	8	30GB	\$0.4184	\$0.08820
n1-standard-16	16	60GB	\$0.8368	\$0.17640
n1-standard-32	32	120GB	\$1.6736	\$0.35280
n1-standard-64	64	240GB	\$3.3472	\$0.70570

Fuente: <https://cloud.google.com/compute/all-pricing?hl=es>

Precios Google Storage en octubre de 2022

Norteamérica Sudamérica <u>Europa</u> Asia Indonesia Australia				
Ubicación	Almacenamiento estándar (por GB al mes)	Almacenamiento RDFS (por GB al mes)	Almacenamiento en frío (por GB al mes)	Almacenamiento de archivos (por GB al mes)
Varsovia (europe-central2)	0,023 USD	0,013 USD	0,006 \$	0,0025 USD
Finlandia (europe-north1)	0,020 USD	0,010 USD	0,004 USD	0,0012 USD
Bélgica (europe-west1)	0,020 USD	0,010 USD	0,004 USD	0,0012 USD
Londres (europe-west2)	0,023 USD	0,013 USD	0,007 USD	0,0025 USD
Fráncfort (europe-west3)	0,023 USD	0,013 USD	0,006 \$	0,0025 USD
Países Bajos (europe-west4)	0,020 USD	0,010 USD	0,004 USD	0,0012 USD
Zúrich (europe-west6)	0,025 USD	0,014 USD	0,007 USD	0,0025 USD
Milán (europe-west8)	0,023 USD	0,013 USD	0,006 \$	0,0025 USD
París (europe-west9)	0,023 USD	0,013 USD	0,006 \$	0,0025 USD

Fuente: <https://cloud.google.com/storage/pricing?hl=es>

- Databricks es una empresa formada por los creadores de Apache Spark
- Proporciona clústeres Spark en la nube con distintas configuraciones y tamaños, e incluso con más herramientas (scikit-learn, keras, TensorFlow...)
- Existe una versión gratuita donde podéis practicar con Spark:
<https://community.cloud.databricks.com>