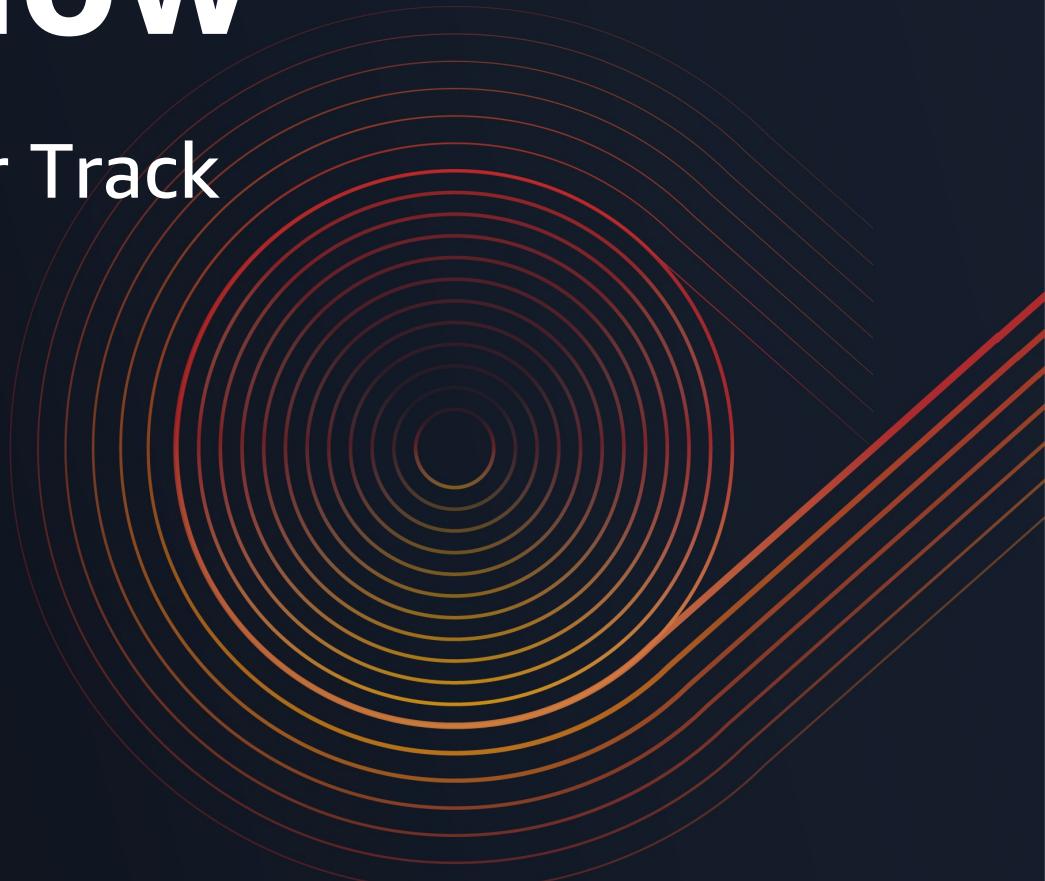




Latam Compute Roadshow

Optimización con servicios de Computación, Customer Track

Darren Koenig
Amer Head of Compute



Quienes Somos

El equipo de Compute trabaja em colaboración com los equipos de cuenta de AWS para acelerar adopción de arquitecturas modernas com **Servicios Avanzados de Compute**

Quienes Somos

Servicios dentro de nuestro portafolio

- Flexible Compute (EC2 Core, Spot, Graviton, Auto-Scaling)
- Serverless (EDA, API Gateway, Lambda, etc)
- Containers (EKS, ECS, Fargate)
- Edge (Outposts e Local Zones)

Conoce a nuestro equipo Latam Compute Go-to-market

Ricardo Brognoli – Head of Compute Latam

Saul Carranza – Container Solutions Architect Latam

Peterson Larentis – Serverless Solutions Architect Latam

Douglas Ramiro – Flexible Compute Solutions Architect Latam

Juan Mestre – Flexible Compute GTM Specialist Latam

David Monroy – Container Solutions GTM Specialist Latam

Agenda

Time	Track	Speaker	Session
8h00 - 9h00	Coffee		Welcome Coffee
9h00 - 9h15	Open Session		Juan Mestre
9h15 - 10h00	Flex Compute	Juan Mestre	Optimización con EC2, Spot y Graviton
10h00 - 10h45	Containers	Iago Henrique Banov	
10h45 - 11h00	Break		Break
11h00 - 11h45	Serverless	Peterson Larentis	Mejores prácticas de optimización de costos y performance Serverless
12h00 - 13h30	Lunch		Lunch
14h00 - 14h45	Containers	Iago Henrique Banov	
14h45 - 15h30	Flex Compute	Douglas Ramiro	Karpenter para Kubernetes y funcionalidades avanzadas
15h30 - 16h00	Coffee-Break		Coffee-Break
16h00 - 16h45	Serverless	Alfredo Peña	Mejores prácticas y patrones de flujo de trabajo Serverless avanzados
16h45 - 17h	Closing / Sorteo / Kahoot		Juan Mestre



Compute Latam Roadshow



Agenda

9:00 – 9:15 Open

9:15 – 10:00 Flexible Compute

10:00 – 10:45 Containers

10:45 – 11:00 Break

11:00 – 11:45 Serverless

11:45 – 12:00 Kahoot

Meet the team

Compute SA Latam Head, Enrico Bergamo

Flexible Compute MDU – Douglas SA, Juan BDM

Containers MDU – Saul SA, Silvio BDM

Serverless MDU – Peterson SA, Pablo/Enrico (Covering)

Optimice la computación de EC2 siendo flexible

Introduction to AWS Flexible Compute WWSO

Juan Mestre (@juanmest)
EC2 Flexible Compute – Principal Go to market specialis

Douglas Ramiro (@douglars)
EC2 Flexible Compute – Specialist Sr. Solutions Architect



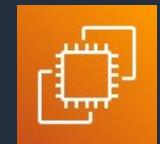
Agenda

- ¿Qué es Flexible Compute?
- Plataforma informática para prácticamente todas las cargas de trabajo
- Attribute based instance selection
- Capacity Optimization
- Cost Optimization
- Performance Optimization

Nuestra Misión es...

Ayudarle a adoptar **prácticas recomendadas** de arquitectura y **servicios de AWS específicos** que optimicen y mejoren el rendimiento de sus cargas de trabajo a la vez que minimizan el desperdicio mediante el **tamaño, el escalado y la flexibilidad**.

Servicios EC2 que le ayudan con flexibilidad



Core EC2



EC2 Auto Scaling



EC2 Spot



Graviton

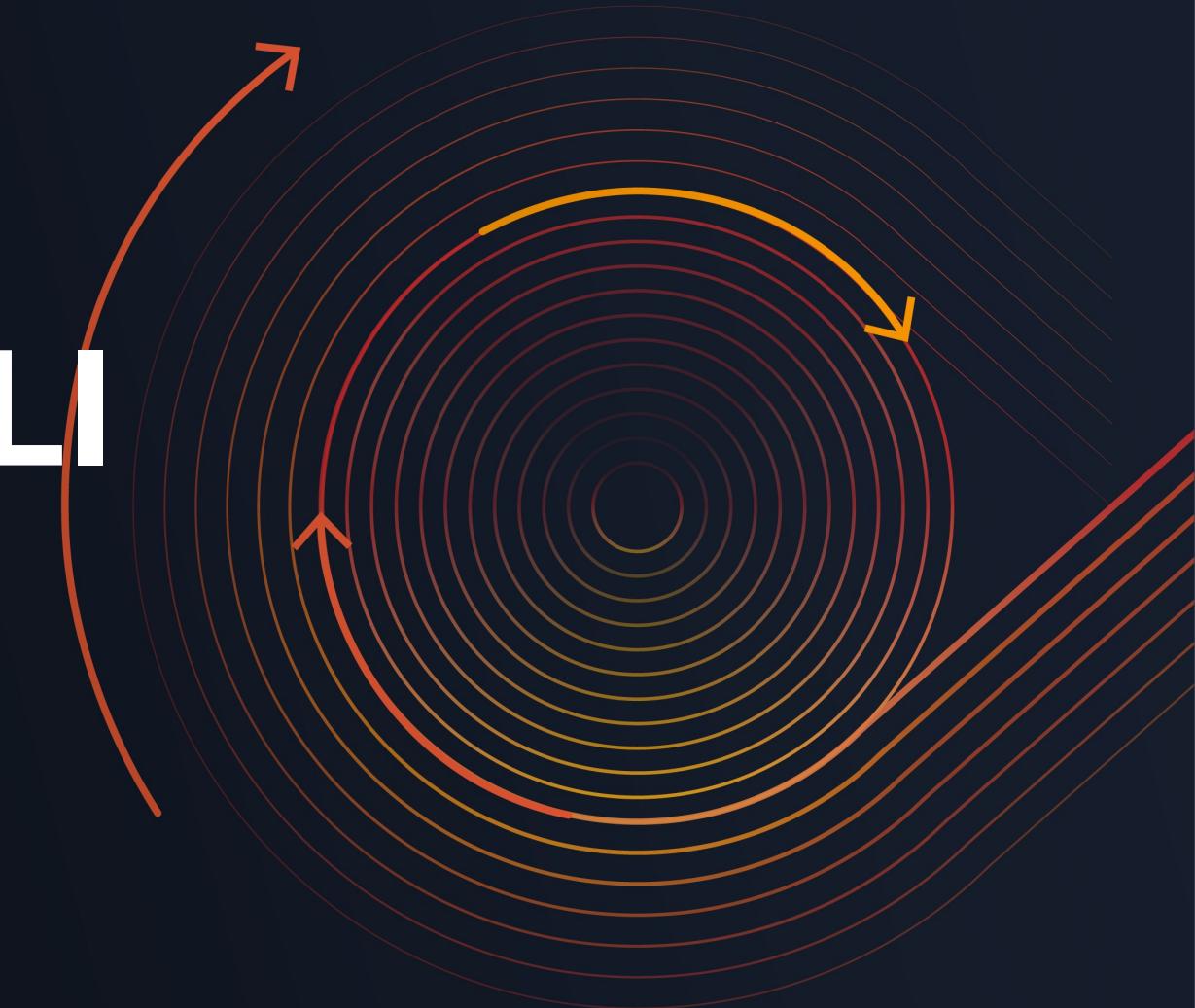


Compute Optimizer

Spot Invaders



Public Reference: MELI





Saved **75% per month** by changing four lines of code



Was able to **save 74%** on their K8s cluster



Saved 40% over Reserved Instances and increased ETL **processing time 100%**

Western Digital.

Completed **2.5 million** tasks in 8 hours by spinning up an EC2 cluster with over **1 million vCPUs**



Reduced grid infrastructure **costs by 60%**



Uses Spot Instances and auto scaling for it's rendering-as-a-service workload to spend **less and scale more**



A containerized environment that would have taken more than a year to create previously only took **a few months.**

CITADEL | Securities

75% reduction in compute cost with Fleet



Reduced monthly compute **costs by 75%** while gaining more compute power



Processes over **100 billion** requests per day with an average response time of 90ms, saving over **\$3M per year**



FRED HUTCH
CURES START HERE™

Decreased the time it took to analyze 10,000 biological samples from **7 years to 7 days**

Delivery Hero

Saved 70% running Kubernetes workloads on AWS



Optimice las estrategias de compra de EC2

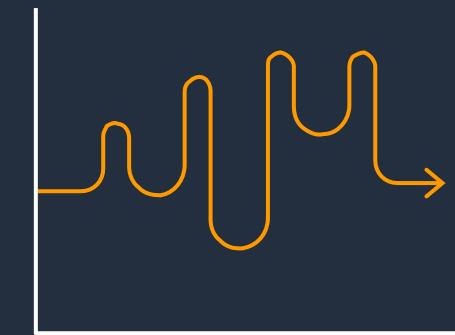


Opciones de compra de Amazon EC2

Piense con flexibilidad y concéntrese en estos para optimizar el rendimiento y el costo

Instancias On-Demand

Pague por la capacidad de cómputo **por segundo** sin compromisos a largo plazo



Cargas de trabajo puntuales

Lanza Instancias

Savings Plans

Los mismos grandes descuentos que los RI de Amazon EC2 con más flexibilidad



Acceso flexible a Computo

Aplica ahorros a instancias bajo demanda

Spot Instances

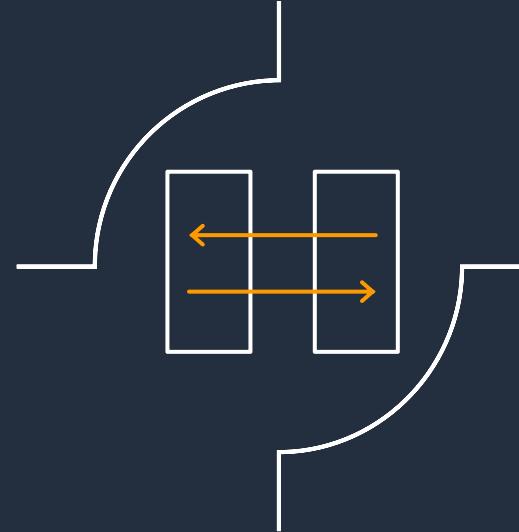
Capacidad de recersva con ahorros de hasta el 90% con respecto a los precios bajo demanda



Cargas de trabajo tolerantes a errores, flexibles y sin estado

Lanza instacias

Amazon EC2 Spot Instances, Mejor descuento – sin compromiso



EC2 Spot Instances

Capacidad de reserva de EC2 de la misma infraestructura



Precios

Cambios suaves e infrecuentes; Más predecible



Uso

Elegir diferentes tipos de instancias, tamaños y zonas de disponibilidad en un grupo de Amazon EC2 Auto Scaling



Capacidad

AWS puede reclamar con 2 minutos de antelación; las interrupciones ocurren cuando EC2 necesita recuperar la capacidad de reserva

EC2 Spot Instances have standard integrations for many AWS services, AWS partner solutions, and open-source software

Cargas de trabajo de spot ideales



Tolerante a fallos



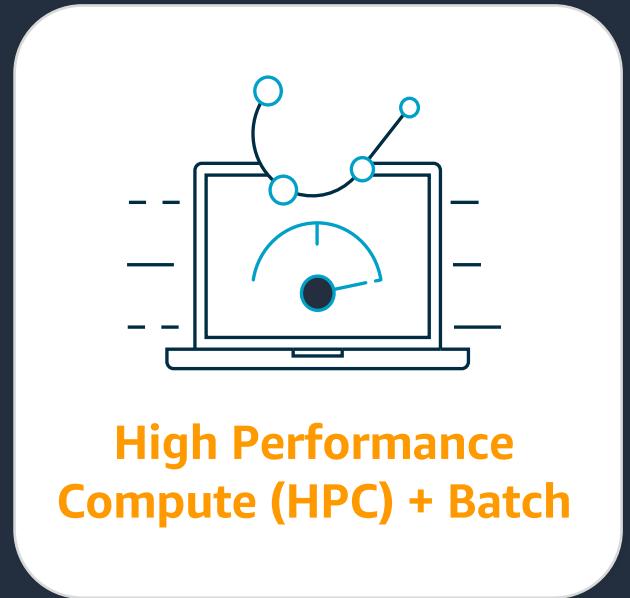
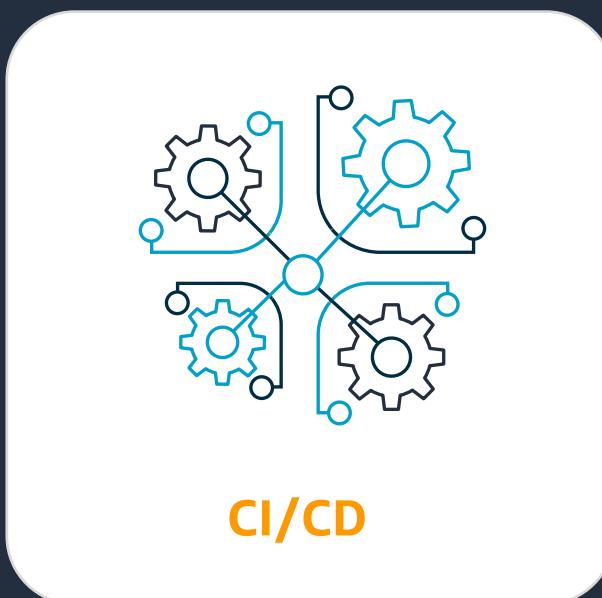
Flexible



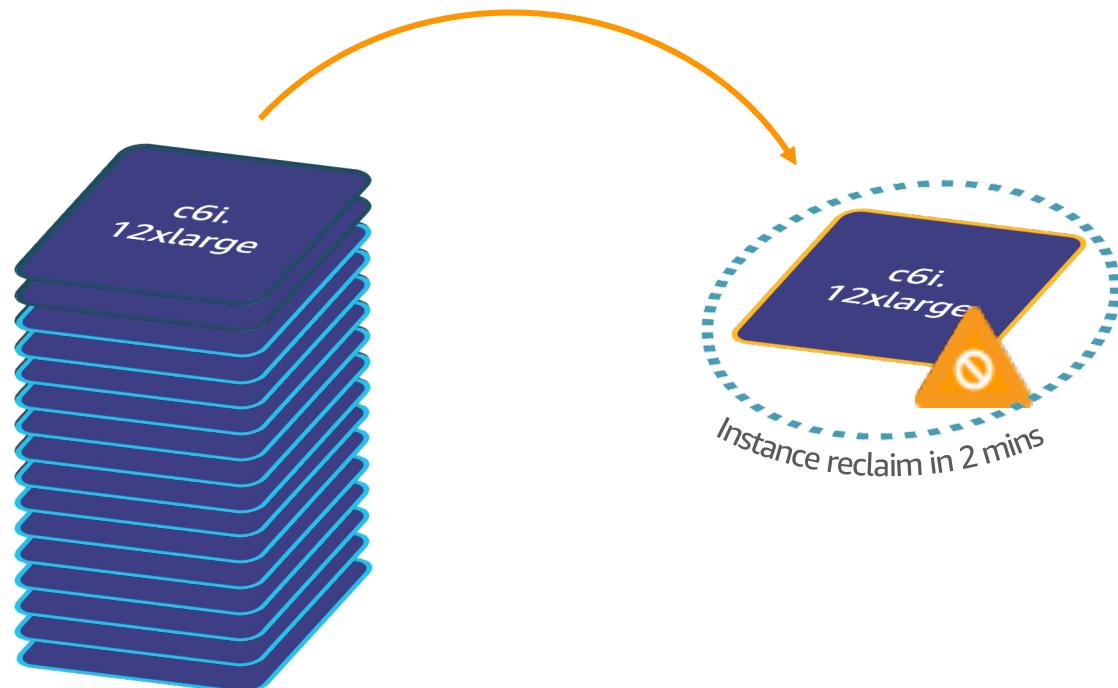
Loosely coupled



Stateless



Interrupciones



Una instancia de Spot se puede interrumpir si la instancia es necesaria para On-Demand.

AWS proporciona dos tipos de notificaciones para gestionar la respuesta de forma **automatizada**:

EC2 instance rebalance recommendation (proactive)

- Actúe cuando su instancia de Spot tenga un riesgo elevado de interrupción
- Soporte integrado para integraciones como **EC2 Auto Scaling** y **EKS Managed Node Groups**



Spot instance termination notice (reactive)

- 2 minutos antes de que se interrumpa la instancia de Spot
- Soporte integrado para los mismos servicios de AWS
- DIY Gestión de interrupciones (**AWS tiene recetas** Para caso uso recomendados)



Históricamente

El 95% de las instancias de Spot lanzadas en los últimos 3 meses se completaron sin interrupción

La diversificación entre instancias reduce las interrupciones

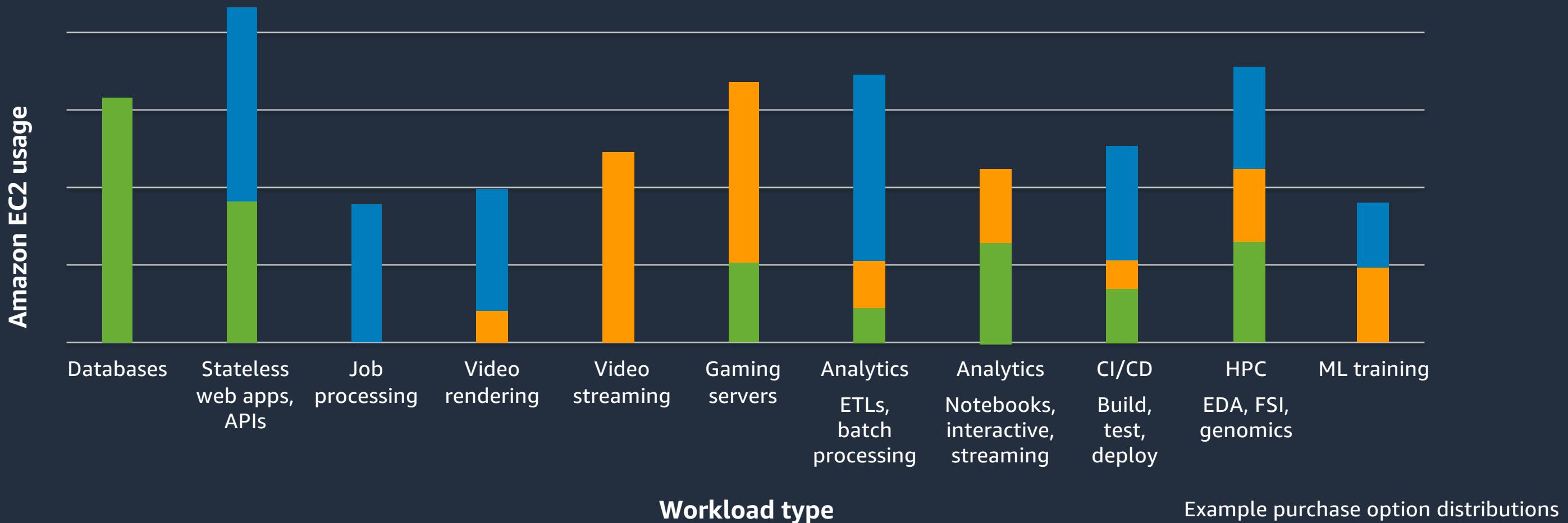


Tipos de carga de trabajo y opciones de compra

Use **Savings Plans** para cargas de trabajo conocidas/de estado estable

Escale usando **On-Demand** Para cargas de trabajo efimeras nuevas o con estado

Escale usando **Spot Instances** para cargas de trabajo flexibles y tolerantes a errores



Amazon EC2 hace 15 años



Una talla única para todos

Paga por lo que usas

Escale verticalmente o reduzca rápidamente según sea necesario

La más amplia selección de procesadores y arquitecturas

PROCESADOR CORRECTO PARA LA CARGA DE TRABAJO



Intel Xeon Scalable
(Skylake and Cascade
Lake) processors



AMD EPYC processor



AWS Graviton2 Processor

Instancias basadas en AWS Graviton2

Hasta un 40 % más de rendimiento de precios que las instancias comparables basadas en x86 de la generación actual

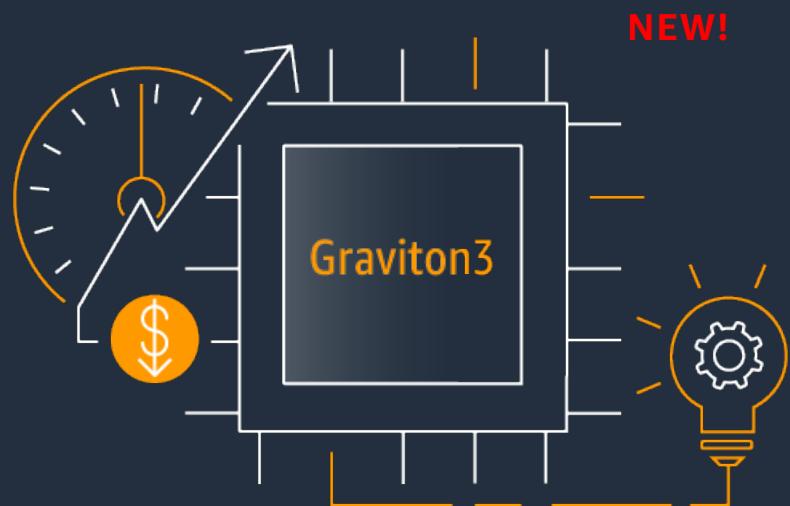
M6g	T4g	C6g	R6g	X2gd
General-purpose workloads	Burstable general-purpose workloads	Compute-intensive workloads	Memory-intensive workloads	Memory-optimized Lowest price/GiB of RAM

También disponible

Opciones de almacenamiento SSD basadas en NVMe local: uso general ([M6gd](#)), Optimizado para cómputo ([C6gd](#)), y memoria optimizada ([R6gd & X2gd](#))
[C6gn](#) instancias con redes de 100 Gbps
Opciones completas para tipos de instancias M, C, R & X: [M6g.metal](#), [M6gd.metal](#), [C6g.metal](#), [C6gd.metal](#), [R6g.metal](#), [R6gd.metal](#), [X2gd.metal](#)

Anuncio de las instancias AWS Graviton3 y Amazon EC2 C7g

Habilitación de la mejor relación calidad-precio para cargas de trabajo en Amazon EC2



Hasta un 25% más de rendimiento en comparación con Graviton2

Hasta 2X más rendimiento de punto flotante, rendimiento de carga de trabajo criptográfica hasta 2X más rápido y hasta 3X mejor rendimiento de MLC en comparación con Graviton2

Primero en la nube en contar con memoria DDR5

60% más de eficiencia energética que las instancias EC2 comparables

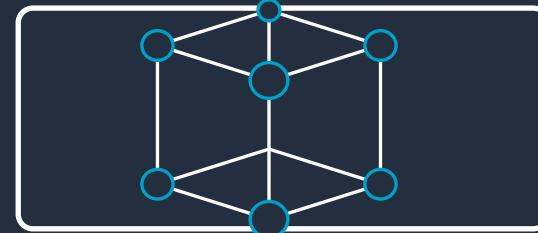
Las instancias C7g, disponibles en versión preliminar, proporcionarán la mejor relación calidad-precio para cargas de trabajo con uso intensivo de informática en Amazon EC2

AWS Graviton: amplia aplicabilidad de cargas de trabajo

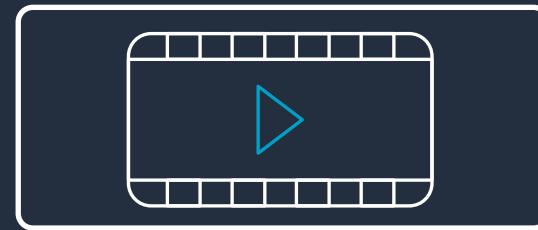
Web and gaming servers



High performance computing



Media encoding



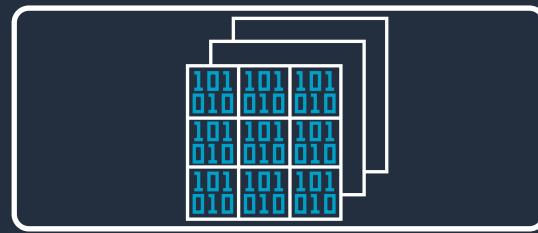
Analytics



Open-source databases



In-memory caches



Electronic design automation



Microservices



Servicios administrados de AWS compatibles con Graviton2

AMPLIACIÓN DE LA RELACIÓN PRECIO-RENDIMIENTO DE GRAVITON2 A LOS SERVICIOS GESTIONADOS

Databases



Amazon
DocumentDB



Amazon
Aurora



Amazon
RDS



Amazon
Elasticache



Amazon
MemoryDB



Amazon
Neptune

NEW!

Analítica



Amazon
OpenSearch



Amazon
EMR

Compute



AWS
Lambda



AWS Fargate
(for ECS)



AWS Elastic
Beanstalk

NEW!



Amazon FSx
for Lustre,
Open ZFS



Optimización de clientes



" El número de publicaciones que la plataforma tiene que manejar puede variar de 5 a 50 por segundo. Con Amazon EC2 Auto Scaling, podemos escalar nuestra infraestructura 10 veces sin ningún problema. . . Como resultado, no pagamos por recursos que no se utilizan, y tampoco lo hacen nuestros clientes".

Piotr Surma

Associate Director, Managing Director



" Al ejecutar nuestra investigación de microbiomas en instancias de spot de Amazon EC2, gastamos menos dinero y menos tiempo para obtener respuestas científicas del análisis: al acelerar nuestros procesos de investigación en AWS, podemos acercarnos al desarrollo de terapias para combatir el cáncer".

Sam Minot

PhD and Staff Scientist



" Con AWS Compute Optimizer, tendremos un único panel para identificar los tipos de instancias EC2 óptimas para nuestras cargas de trabajo en todo nuestro entorno de AWS. La capacidad de visualizar la utilización de recursos proyectada en varios tipos de instancias nos permitirá tomar decisiones más informadas antes de actualizar nuestra flota".

Renan Capaverde

Director of Engineering

Compute LatAm Roadshow 2023 | Post-Event Deep Dive | México

Agradecemos su participación en el LatAm Compute Roadshow en México edición 2023.

Le invitamos a compartir su interés registrándose en la sección adjunta para que nuestros especialistas puedan ponerse en contacto con usted, ya sea para profundizar en sesiones subsecuentes o bien para apoyarle en sus proyectos de Cloud Computing, Modernización de Aplicaciones, tecnologías innovadoras como Contenedores y/o Serverless así mismo de obtener el máximo provecho de nuestros esquemas y servicios de optimización flexible.



¡Gracias!



Title

Text

Tkx!

