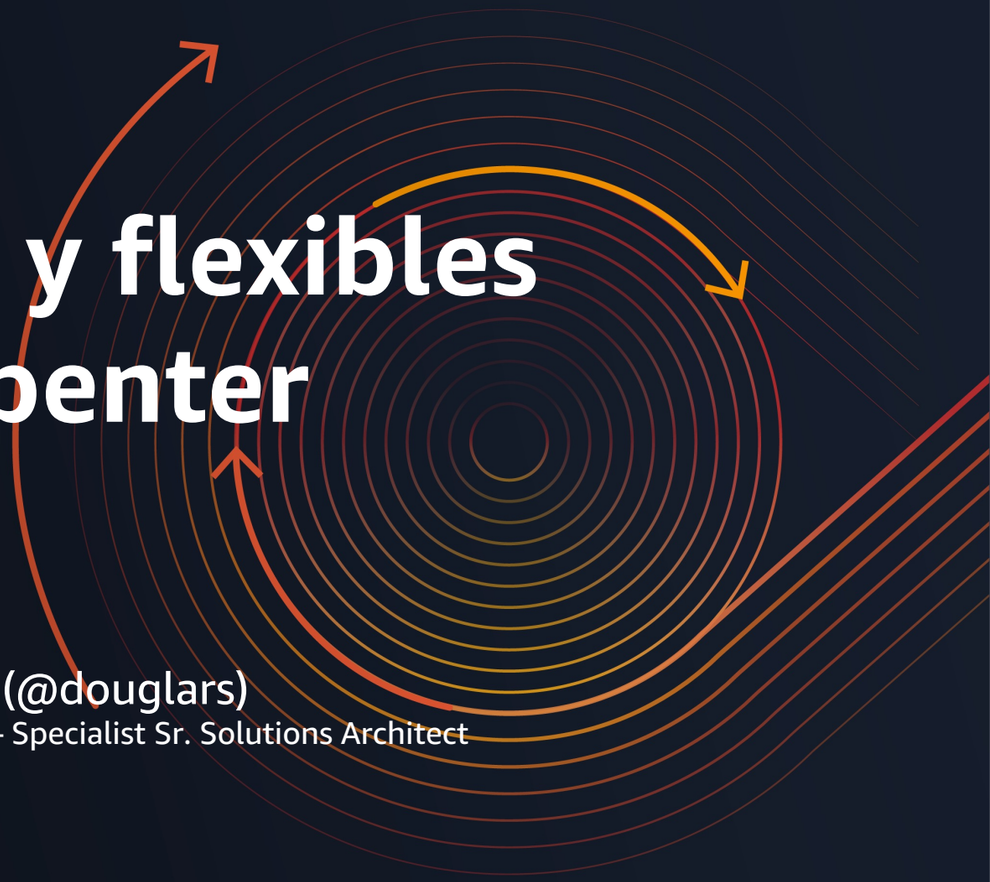
aws LATAM Compute Roadshow

Optimice sus cargas stateless y flexibles con EC2 Spot, Graviton y Karpenter

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



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

aws

aws LATAM Compute Roadshow

Optimice las estrategias de compra de EC2

aws

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

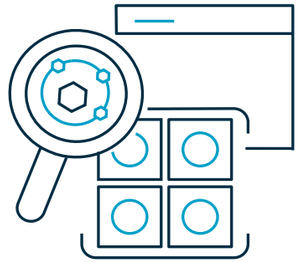
Capacidad de recersva con ahorros de hasta el 90% con respecto a los precios bajo demanda

Cargas de trabajo tolerantes a errores, flexibles y sin estado

Lanza instacias

Cargas de trabajo de spot ideales

- Tolerante a fallos
- Flexible
- Loosely coupled
- Stateless

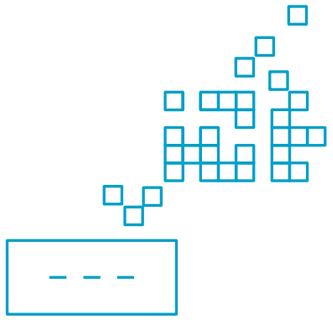


Web services

Containers

High Performance Compute (HPC) + Batch

CI/CD

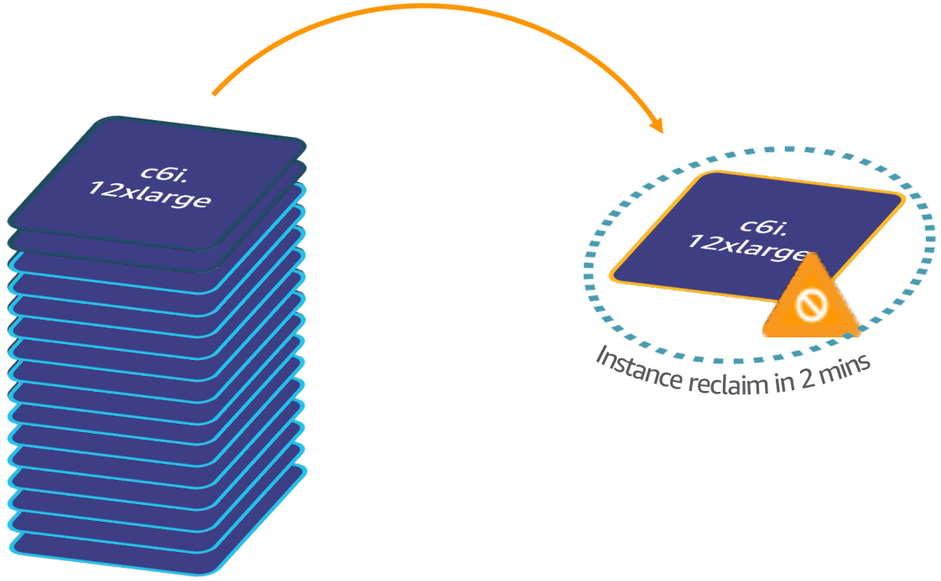


Big data

AI/ML

aws

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 casos de uso recomendados)

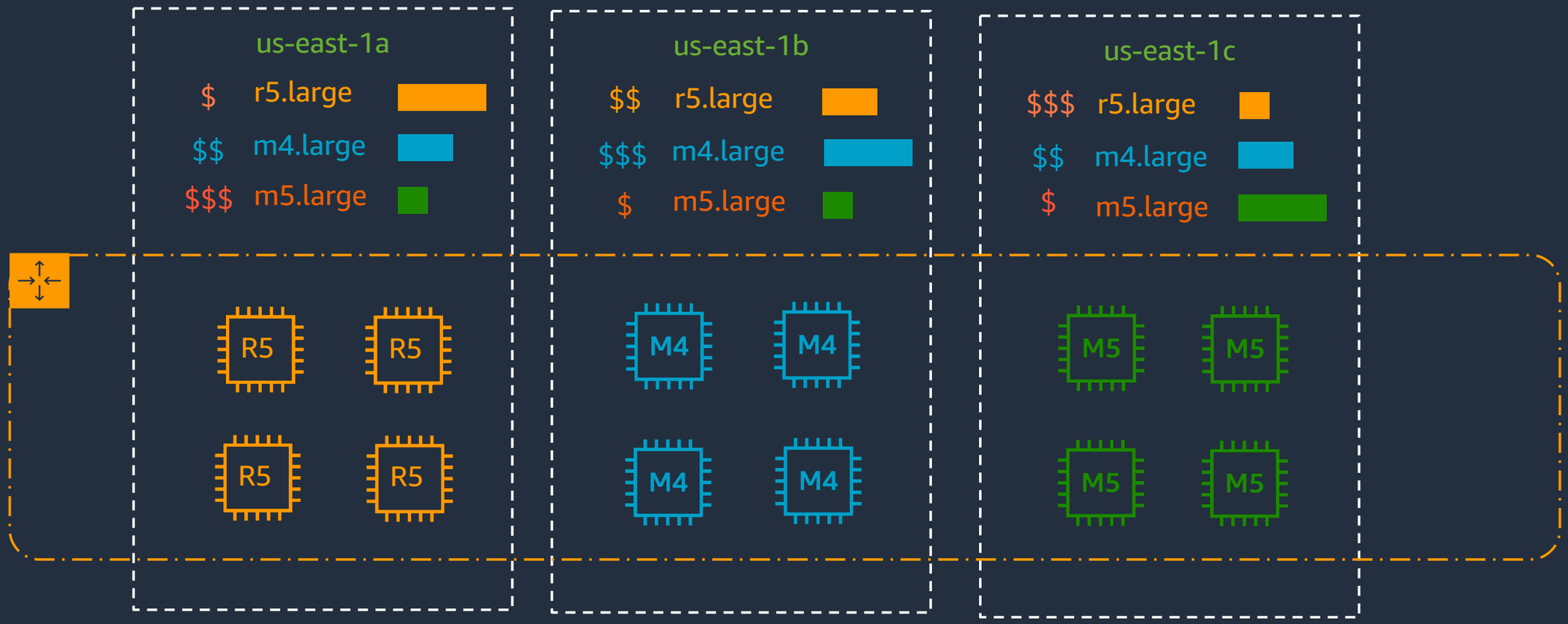
Históricamente

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

La diversificación entre instancias reduce las interrupciones

Estrategia de asignación Capacity-optimized

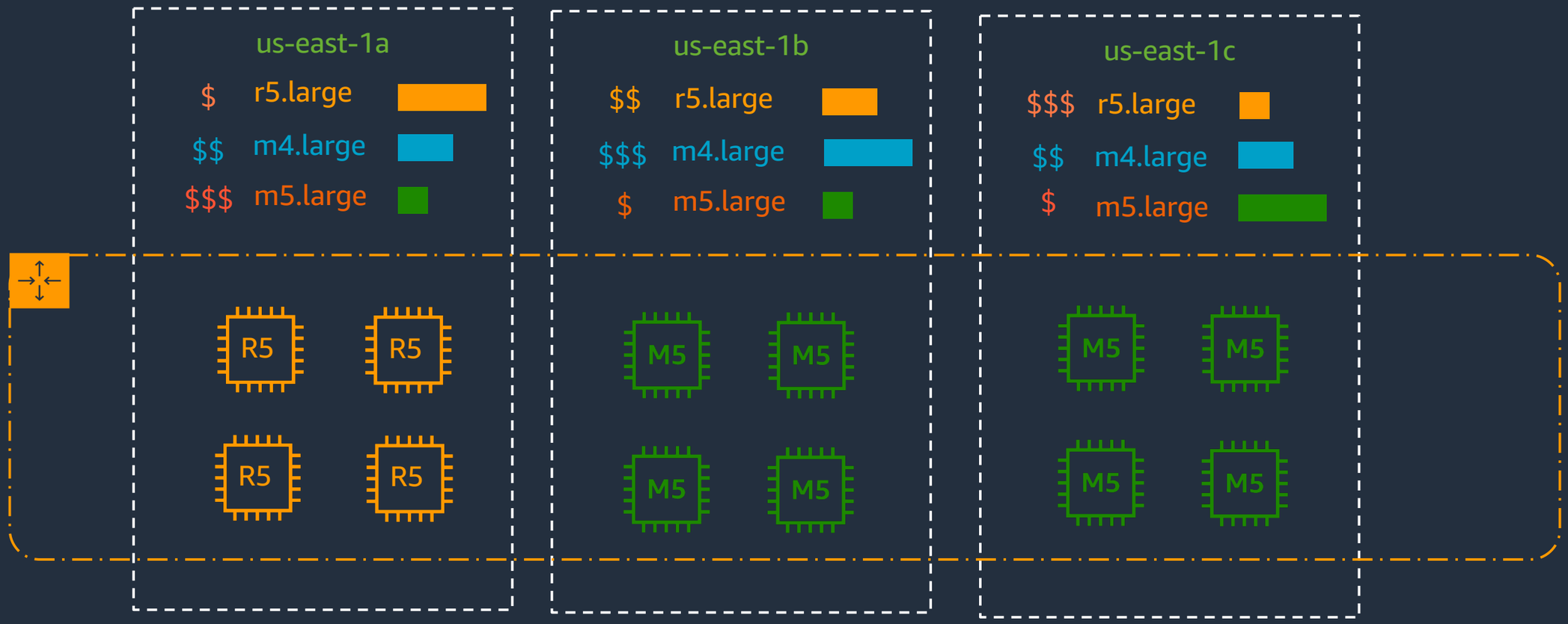
`SpotAllocationStrategy: capacity-optimized`



aws

Estrategia de asignación Lowest Price

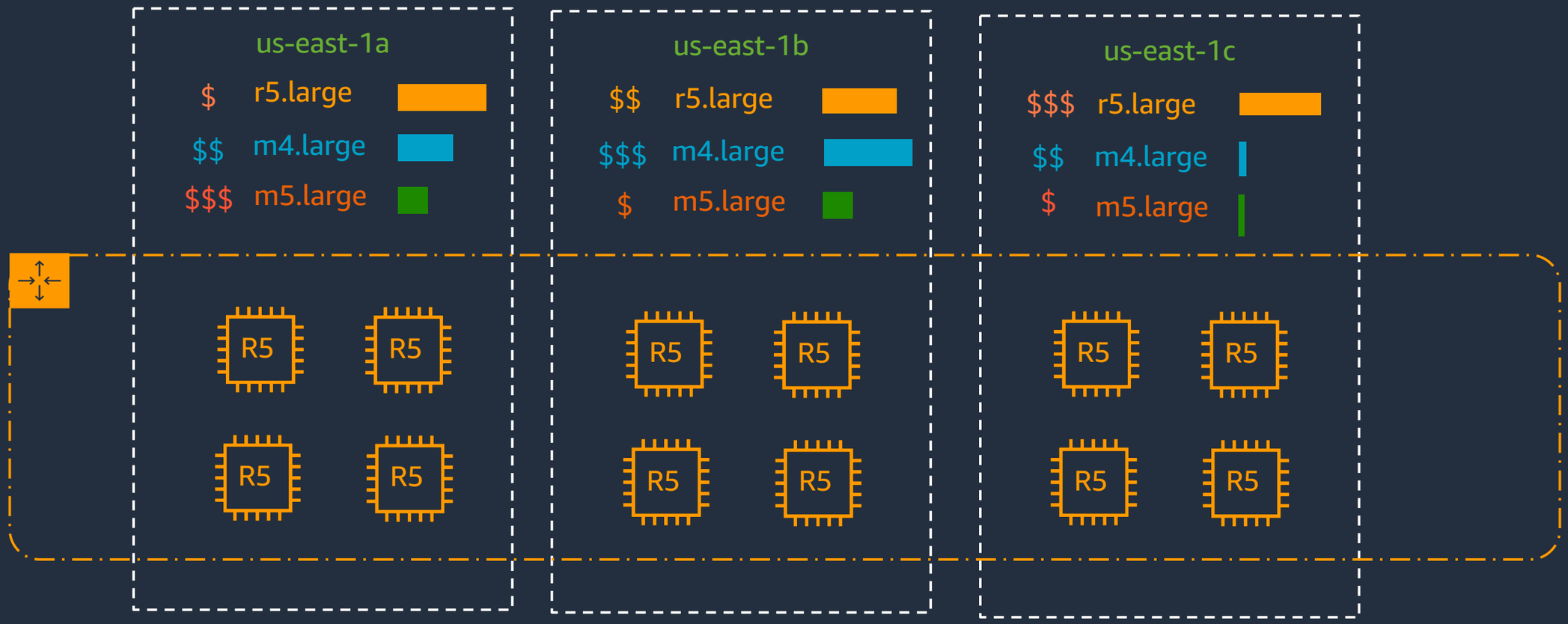
```
SpotAllocationStrategy: lowest-price
```



aws

Estrategia de asignación price-capacity-optimized

`SpotAllocationStrategy: Price-capacity-optimized`



aws

Demo: SPS

aws

Demo: ABIS

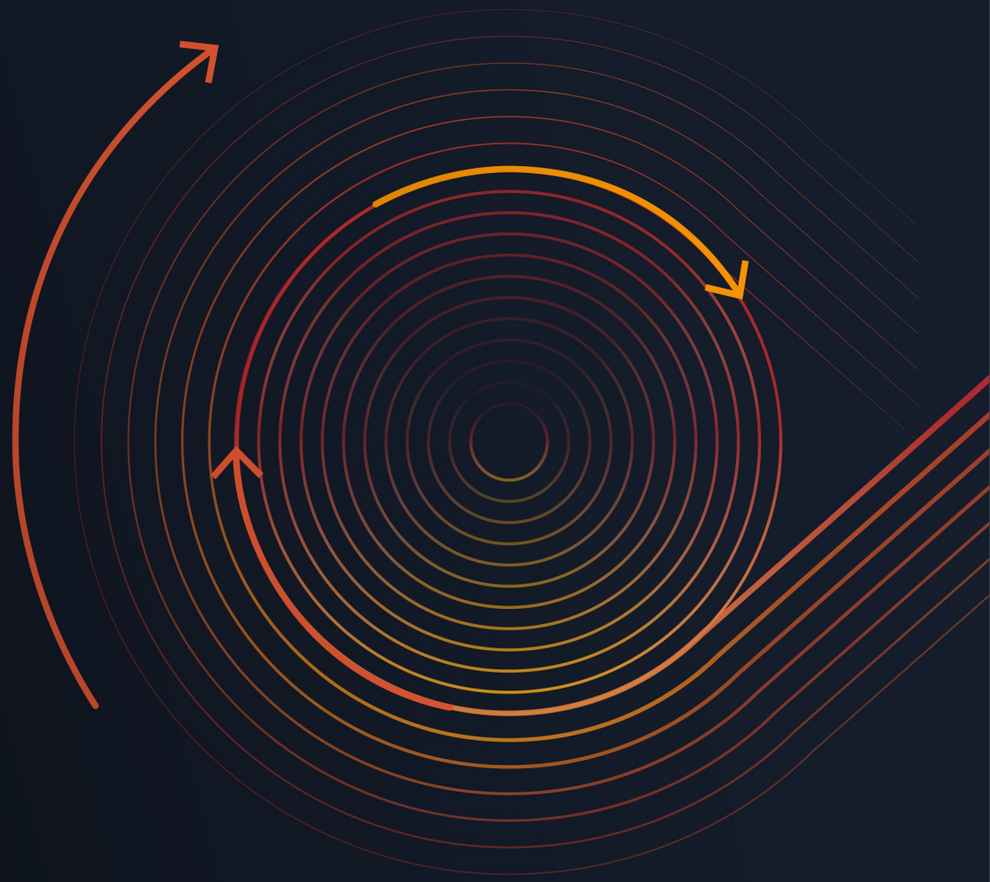
aws

Demo: FIS

aws

aws LATAM Compute Roadshow

Graviton



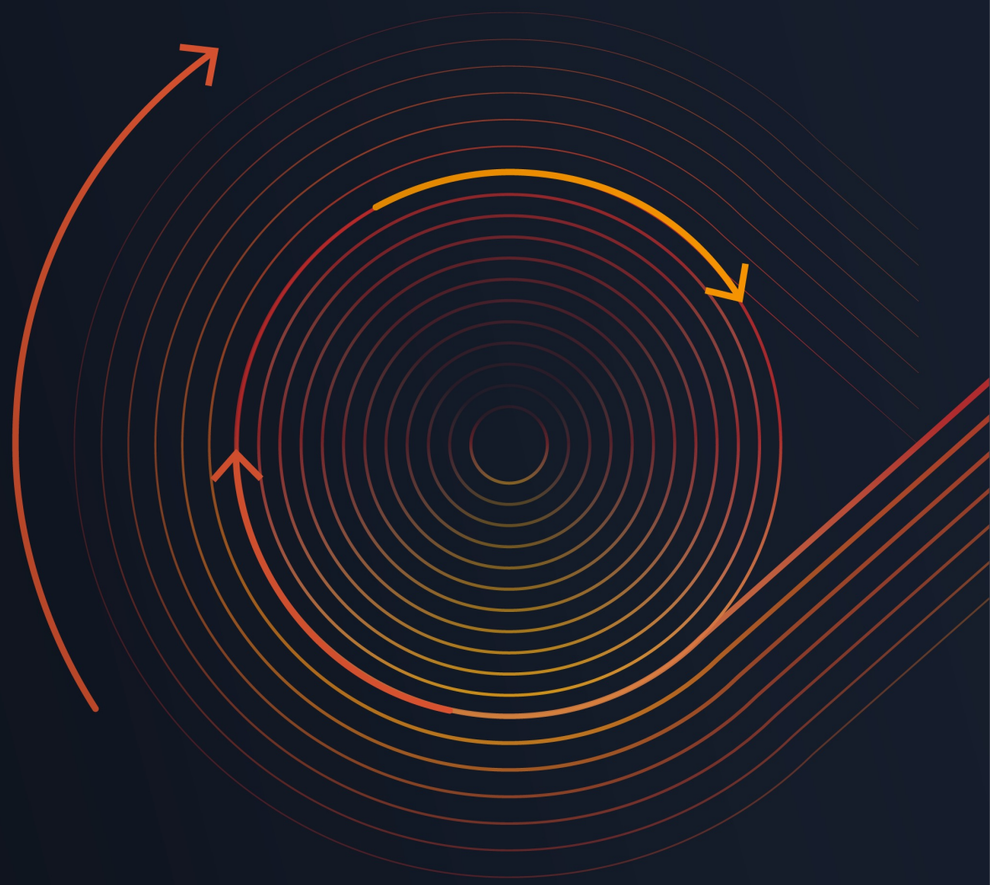
Dificultad	Carga de trabajo	Acciones
Prácticamente ningún esfuerzo	RDS, ElastiCache	Actualice a la versión más reciente y disfrute
Súper fácil	EMR	Por lo general, solo funciona /!\ Módulos nativos de JNI o Python
Bastante fácil	Linux – Lenguajes de alto nivel	Recoger arm64 AMI – Instalar Bonificación si los contenedores /!\ Módulos nativos de JNI o Python
Fácil	Linux – Lenguajes de bajo nivel	Pick up arm64 AMI – Recompilar intrínsecos o assembler para ser portados
Algo de trabajo, alta recompensa	Microsoft Windows – .NET	Migrar a .NET core, ejecutar en Linux/arm64
Lo sentimos, todavía no	Microsoft Windows	Microsoft Windows Server aún no está disponible para arm64

Demo: Porting Advisor for Graviton

aws

LATAM Compute Roadshow

EKS y Karpenter



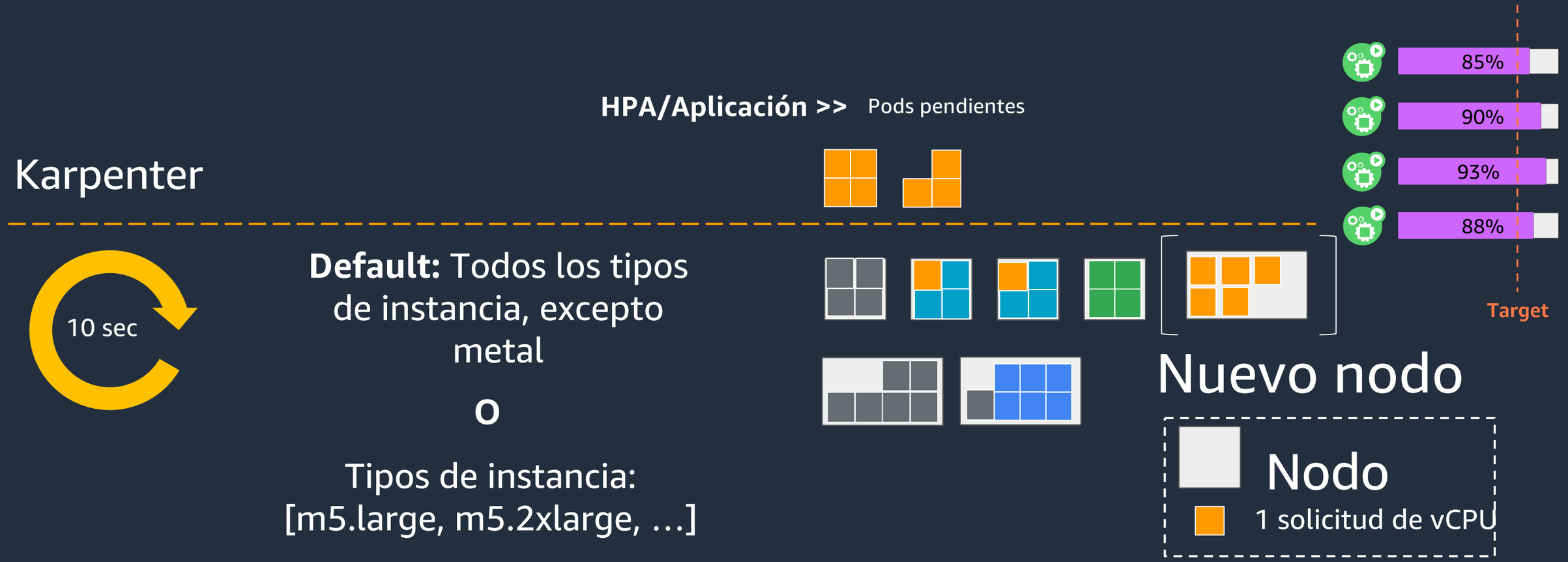
Cómo Karpenter programa los pods

Karpenter trabaja en conjunto con Kube-scheduler y el proveedor de cómputo

- Kube scheduler obtiene la primera solicitación en la programación de pods pendientes en Capacidad existente
-
- Karpenter observa las solicitudes de recursos agregados de pods no programables, calcula y lanza el mejor recurso con Nueva capacidad
-
- Karpenter finaliza nodos vacíos
- Karpenter Consolida nodos infrautilizados

aws

Escalación de Karpenter



Decisiones de aprovisionamiento y programación

Encuadernación anticipada para provisionar nodos vs. placeholder instances

- Quitar la dependencia de la versión del scheduler

Consolidación de Karpenter

```
apiVersion: karpenter.sh/v1alpha5
kind: Provisioner
metadata:
  name: my-provisioner
spec:
  consolidation:
    enabled: true
```

Consolidación

- Elimina un nodo – Cuando los pods pueden ejecutarse con capacidad libre de otros nodos del clúster
- Elimina un nodo – Cuando el nodo está vacío (no es necesario establecer ttSecondsAfterEmpty con consolidación)
- Reemplaza un nodo – Cuando los pods pueden ejecutarse en una combinación de capacidad libre de otros nodos en el clúster + nodo de reemplazo más eficiente

aws

Demo: Karpenter

aws

aws Compute LATAM Roadshow
Optimization through compute services

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

Agradecemos su participáción 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.

aws

aws LATAM Compute Roadshow

¡Gracias!

aws