

Microsoft ❤️ Linux

# Open Source in the Cloud Expert Academy

2019

# Agenda

Contenido Dia 2



docker



- Presentación Audiencia
- Speaker
- Lluvia de ideas
- Open Source en el Mundo
- Microsoft Loves Linux
- Lab 1: OSS en Azure
- OSS en Azure
- Lab 2: Scaling OSS on Azure
- Mundo Containers
- Lab 3: Containers on Azure
- Kubernetes on Azure
- Lab 4: Kubernetes on Azure
- CosmosDB on Azure
  - Particionamiento
    - Lab 5: Consulta de partición cruzada
    - Lab 6: Solución multi-partición
  - Querying
    - Lab 7: Consultas a CosmosDB empleando SQL
  - Programación
    - Lab 8: Stored Procedures
  - Aplicaciones
- Conclusiones

<https://www.microsoftazurepass.com/>

# Speaker

## Open Source en Azure



Nombre

**Manuel Alejandro Peña Sánchez**

Experiencia

Pre-Sales Data Architect  
bSide

Ex - Especialista de Seguridad Sr  
Grupo Salinas

Ex - IT Security Specialist  
IBM.

Correo

mpena@bside.com.mx

+17 Años

MS

+17 Años

OSS

+17 Años

TI

+11 Años

Seguridad

# Speaker

## Modern Apps



Nombre

**Julio Cesar Zepeda Alvarez**

Experiencia

Pre-Sales Modern Apps  
bSide

Ex - Especialista de desarrollo de aplicaciones  
Grupo Salinas

jzepeda@bside.com.mx

Correo

+7 Años

MS

+8 Años

Desarrollo

+7 Años

TI

+4 Años

Azure

# Speaker

Cosmos DB



## Nombre

David Sánchez Almanza

## Experiencia

Technology Solutions Professional Data / AI  
bSide

Cloud Technical Specialist  
Microsoft

OSS Developer  
Eucledian

## Correo

dsanchez@bside.com.mx

+4 Años

MS

+4 Años

Azure

+4 Años

TI

+3 Años

AI

+3 Años

Data

# ENCUESTA

Open Source en Azure

URL: [live.voxvote.com](https://live.voxvote.com)

PIN: 19379





# OSS en el Mundo



# ¿Qué es Open Source?

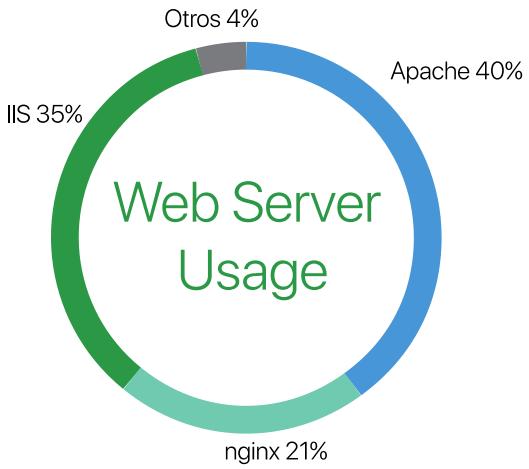
Open Source en Azure



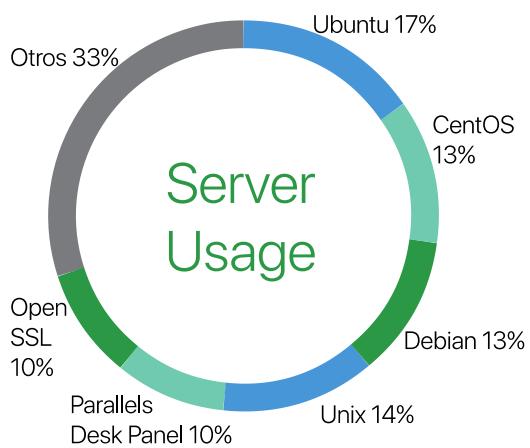
# OSS en el mundo

## Open Source en Azure

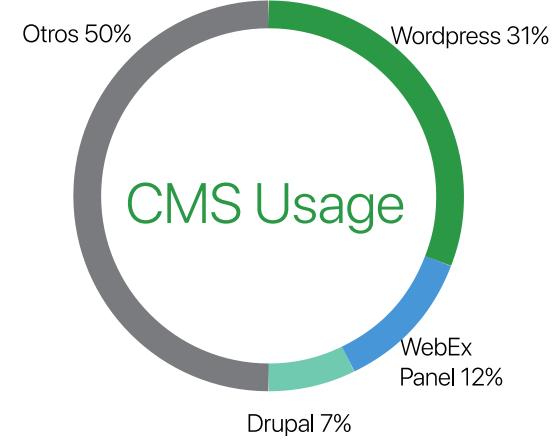
Statistics for websites using Web Server technologies



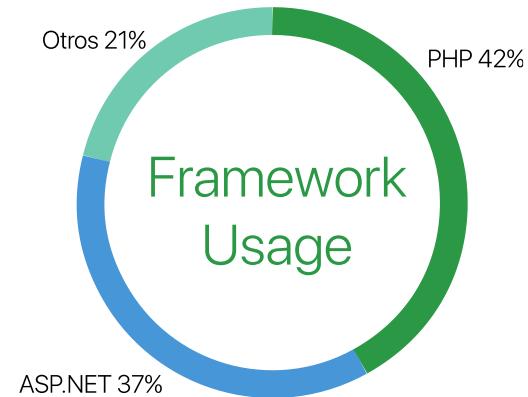
Statistics for websites using Server technologies



Statistics for websites using CMS technologies



Statistics for websites using Framework technologies



Referencia: <https://trends.builtwith.com/>

# Mitos

Open Source en Azure

## Mitos actuales de Open Source



Incompatible



Menos Seguro



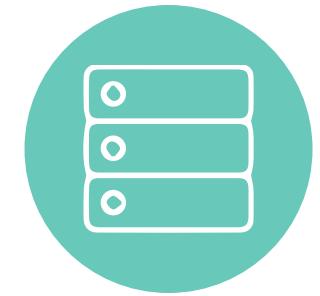
Gratis



Difícil



Sin soporte

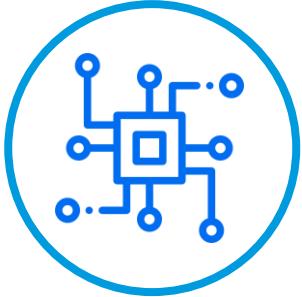


Solo es  
Infraestructura

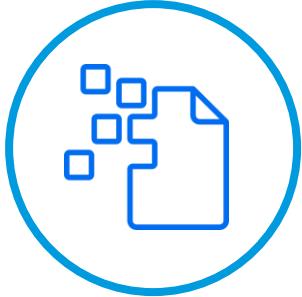
# Tecnologías

Open Source en Azure

Computo



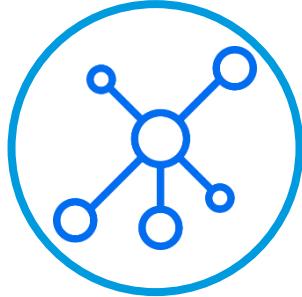
Almacenamiento



Base de Datos



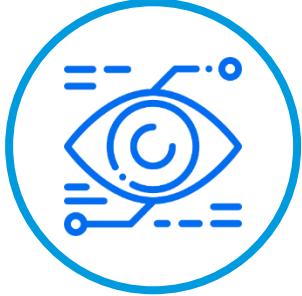
Redes



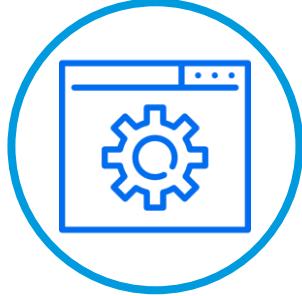
Seguridad



Análisis



Aplicaciones



Móviles



# Microsoft Loves Linux

Open Source en Azure

Microsoft ❤️ Linux

# Microsoft Loves Linux



MuyLinux  
@muylinux

Seguir

Red Hat y Microsoft, juntos para impulsar las tecnologías en la nube



Red Hat y Microsoft, juntos para impulsar las tecnologías en la nube » MuyLi...

Red Hat y Microsoft anunciaron en 2015 una de las alianzas más sorprendentes del siglo en el mundo de la computación. Ver a la compañía Linux líder colabora

muylinux.com

Microsoft SQL Server 2017 para Linux, lo más destacado de la presentación oficial en España



Microsoft SQL Server 2017 para Linux, lo más destacado de la presentación ...

Si te interesa saber más acerca de Microsoft SQL Server 2017 en Linux, no te pierdas ese artículo especial.

muylinux.com



MuyLinux  
@muylinux

Seguir

Linux se impone en Azure y Microsoft se autoproclama el número uno del código abierto



Linux se impone en Azure y Microsoft se autoproclama el número uno del có...

Linux ya es el sistema más utilizado en Microsoft Azure. "Somos el mayor partidario del código abierto en el mundo", dice la compañía.

muylinux.com



MuyLinux  
@muylinux

Seguir

Debian para Windows 10 también está disponible desde la Microsoft Store



Debian para Windows 10 también está disponible desde la Microsoft Store » ...

Debian para Windows 10 también está disponible en la Microsoft Store para su utilización sobre WSL, el subsistema de Linux que Microsoft ha introducido en últim...

muylinux.com

Seguir

Linux sigue gobernando el 100% de las supercomputadoras del TOP500



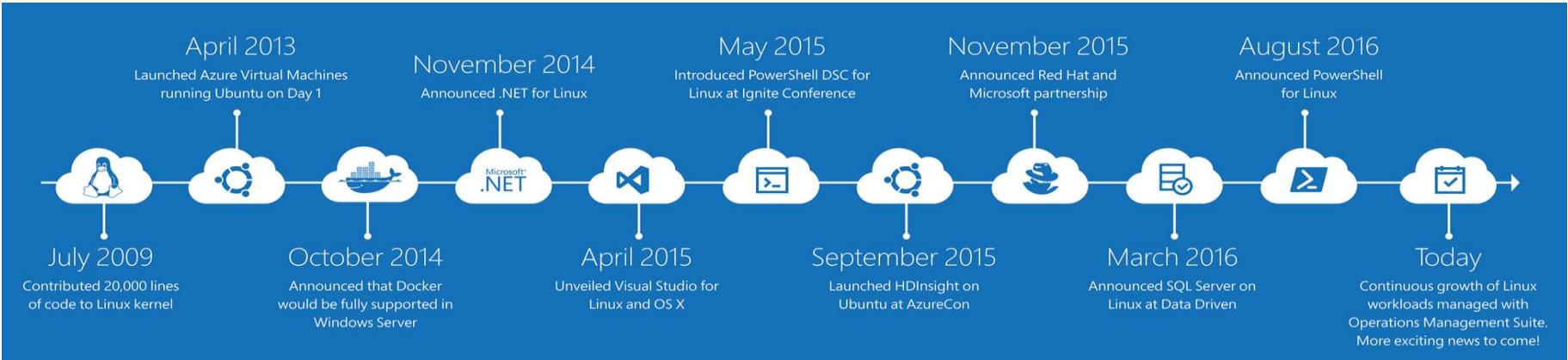
Linux sigue gobernando el 100% de las supercomputadoras del TOP500 » Mu...

La lista TOP500 de junio de 2018 muestra que Linux sigue gobernando el 100% de las 500 supercomputadoras más rápidas del mundo.

muylinux.com

# Microsoft Loves Linux

## Open Source en Azure



PowerShell is now **open source** and available on Linux

Hundreds of contributors to the **Azure Resource Manager** community on **GitHub**

**Linux Agent (GA) & Container Solution (Preview)** for Operations Management Suite

Over **50%** of VMs created in **new** deployments in Azure today run **Linux!**

Nearly **50%** VMs in Azure run **Linux** today

**4x** growth in **container** customers in Azure since January, all with Docker on Linux!

Redmond top man Satya Nadella: 'Microsoft LOVES Linux'



The Register

# Microsoft Loves Linux

## Productos



SQL Server en Linux



Adquisición



HD Insight managed service en Linux

Windows Subsystem for Linux

```
C:\Users\markhill> bash  
root@localhost: #
```

Corre Linux en Windows nativamente

## Azure Marketplace



+ 50 %

60% de las imágenes en Azure Marketplace están basadas en Linux/OSS

+ 50% VMS en Azure corren Linux, y más del 50% de las **nuevas** VMs corren Linux.

## Ofertas



eclipse  
Microsoft joins Eclipse Foundation



Jenkins project en Azure



Asociación con Linux Foundation para Certificación de Linux en Azure

## Alianzas



Ross Gardler  
President Apache SW Foundation



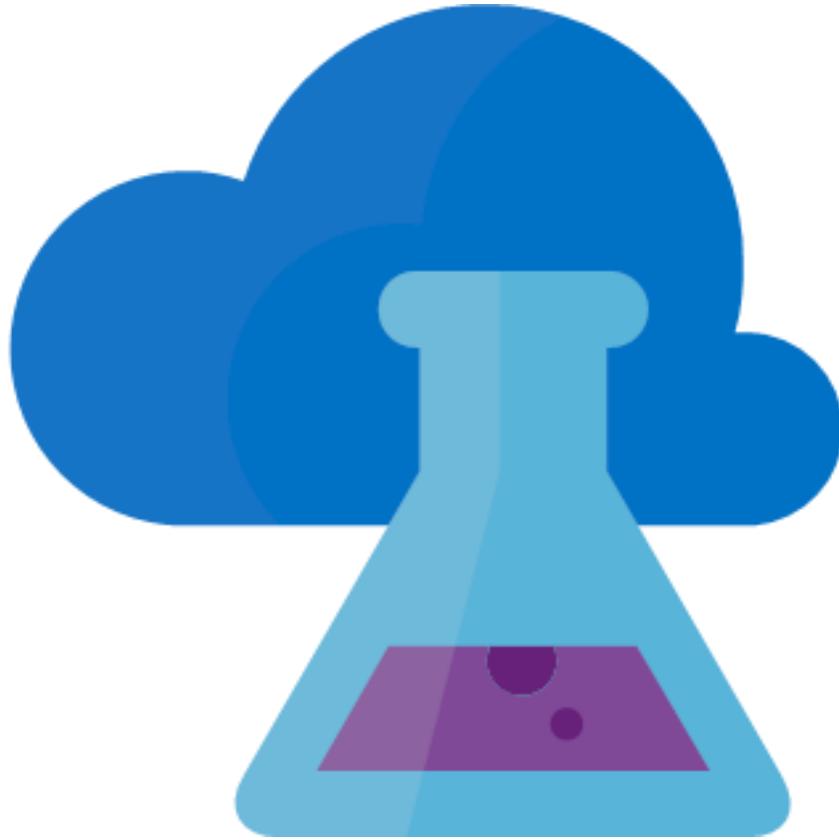
Miguel de Icaza  
Founder Xamarin



600 Million+  
Lines of open source code  
[Microsoft Open Source Hub](#)



Gold Microsoft Partner  
Microsoft



# Lab 1

## OSS on Azure

¡Descarga tu material!

<https://github.com/psmapad/ExpertAcademy>

<https://www.microsoftazurepass.com/>

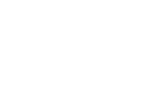
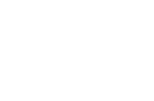
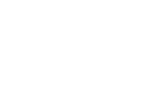


# OSS en Azure



# Microsoft Loves Linux

Open Source en Azure

Management	System Center OMS	     
DevOps & PaaS	Team Foundation Services	       
Applications	Dynamics SharePoint Exchange	     
Frameworks & Tools	.NET Visual Studio	        
Data	SQL Server	          
Infrastructure	Windows Server	       
	Microsoft Traditional Monetization	

# Enfoque de integración Microsoft en el Open Source

## Open Source en Azure



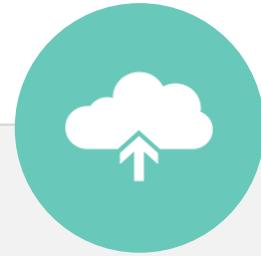
### Habilitar

Permitir a Linux y tecnología Open Source ser ciudadanos de primera clase en Azure



### Integrar

Adoptar tecnología Open Source líder e integrarlo al ecosistema de productos Microsoft con agilidad y consistencia.



### Liberar

Liberar tecnologías Microsoft clave, en el mundo de open source para construir un ecosistema mas fuerte.

R Server  
.NET Core  
Roslyn  
TypeScript  
F#  
  
autorest  
PowerBI Visuals  
Office UI Fabric  
Tools plugins



### Participar

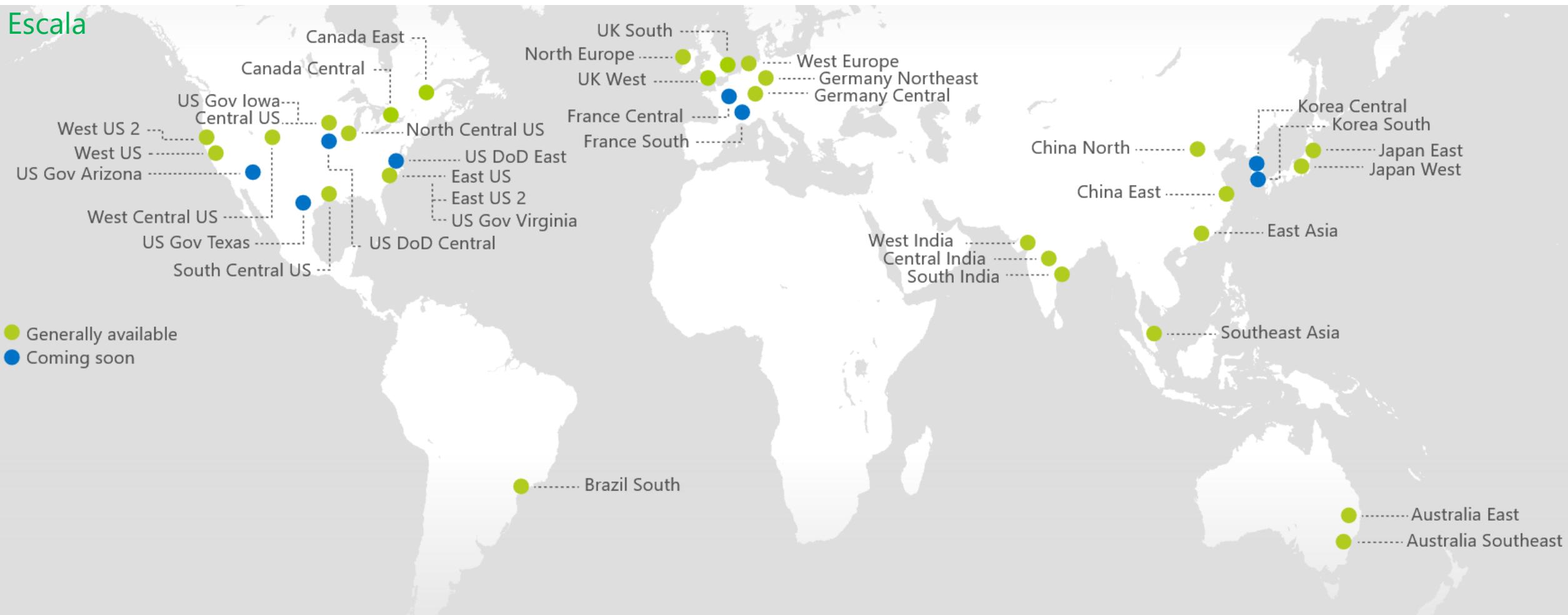
Ingenieros de Microsoft participa y contribuyen en proyectos clave de Open Source.



## Open Source Partners & Ecosystem

# Microsoft Azure

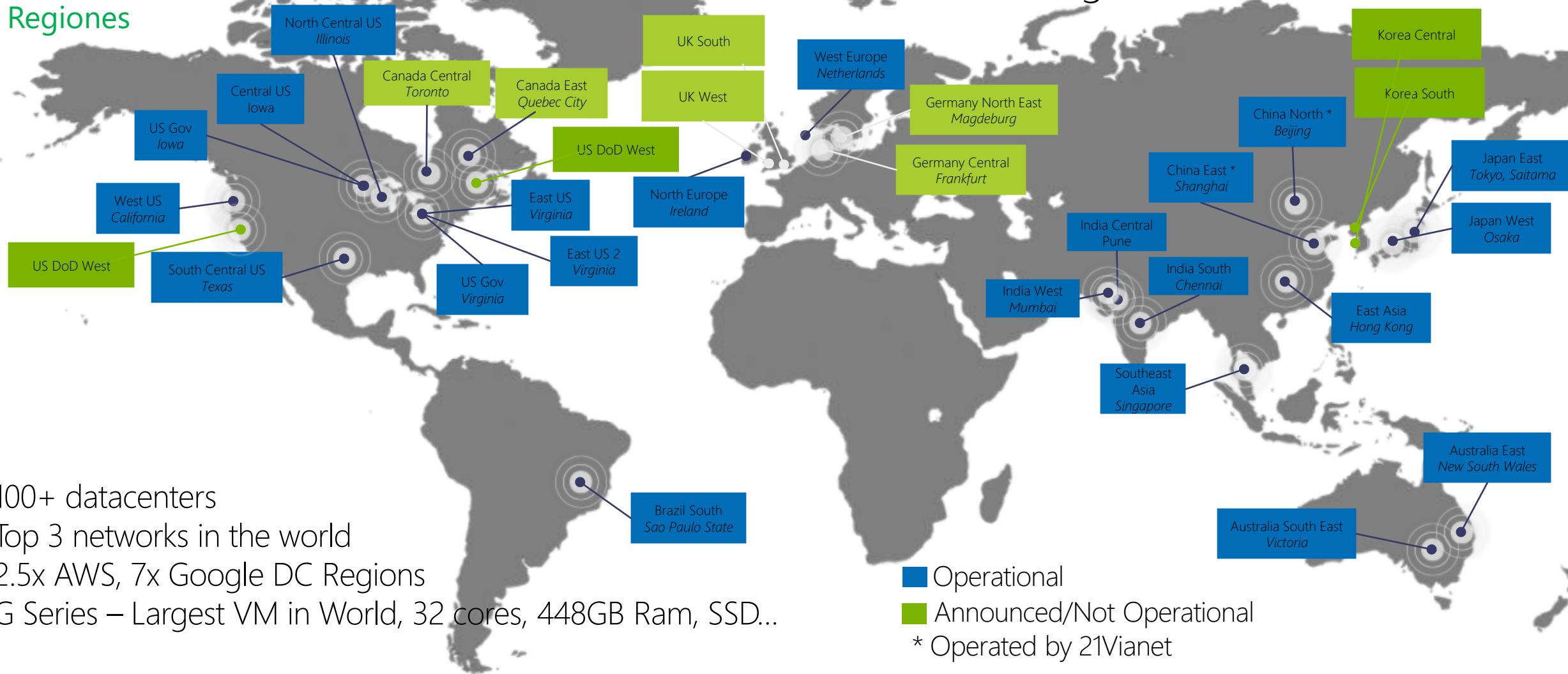
Escala



42 Regions - More than AWS and Google Combined

# Microsoft Azure

## Regiones



# Microsoft Azure

## Familias de Computo



Entry  
Level



General  
Purpose VMs



Compute  
Optimized VMs



Large Memory  
VMs



>80,000 IOPs  
Premium Storage

Dev/Test and  
entry-level workloads

Earliest generation, HDD

Good combination of  
memory, SSD for most  
common production  
applications

Memory-intensive  
variants

Compute-intensive  
apps like Gaming,  
Analytics

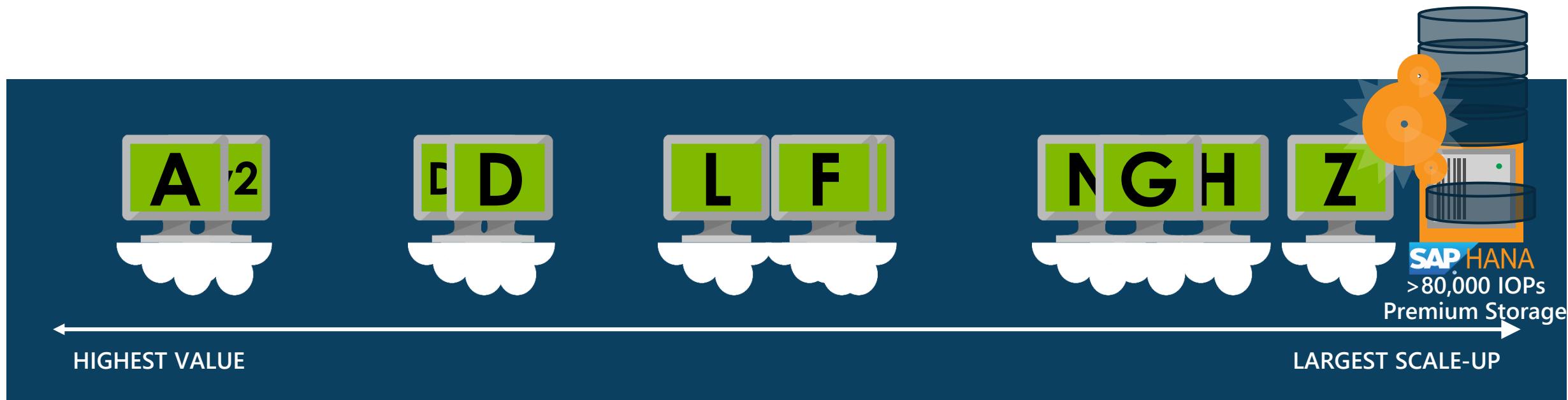
More CPU to  
memory ratio

Large VMs for large  
databases requiring  
fast Storage

Intel Haswell processor

# Microsoft Azure

## Familias de Computo



Dev/Test and entry-level workloads  
Earliest generation, HDD

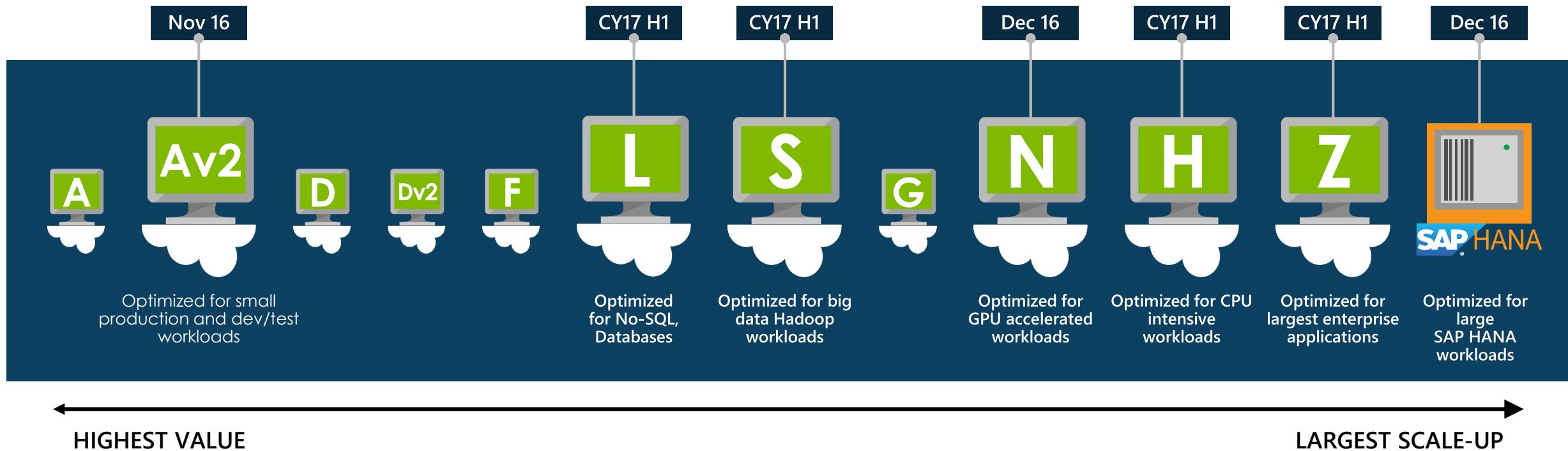
Good combination of memory, SSD for most common production applications  
Memory-intensive variants

Compute-intensive apps like Gaming, Analytics  
More CPU to memory ratio

Large VMs for large databases requiring fast Storage  
Intel Haswell processor

# Microsoft Azure

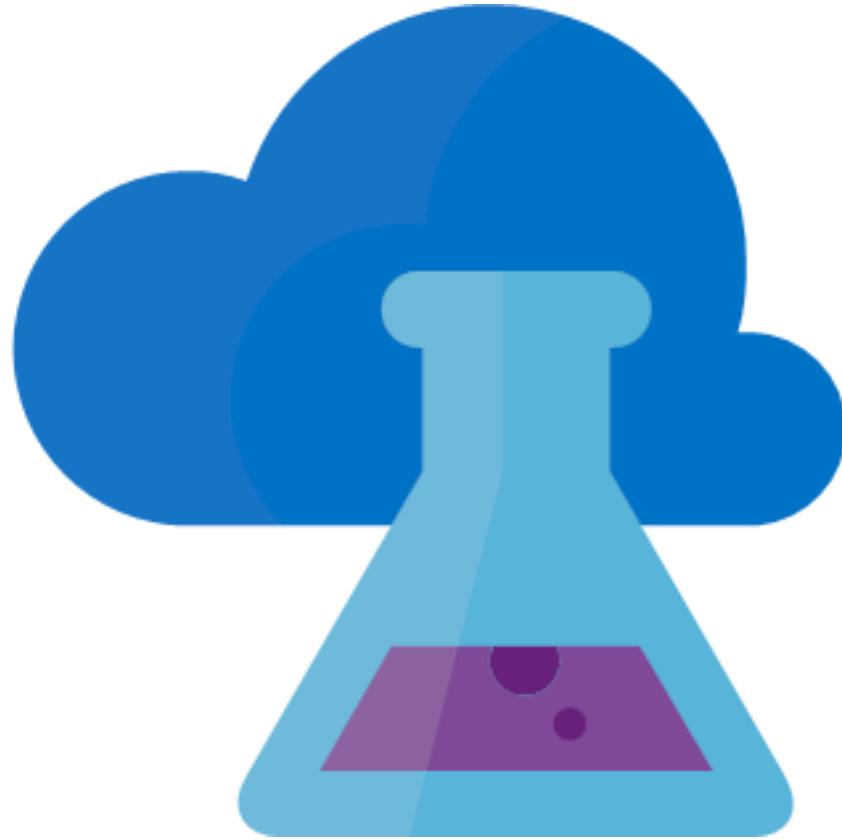
## Familias de Computo



# Open Source Software en Azure

## Distros Soportadas

Distribution	Version	Drivers	Agent
CentOS by OpenLogic	CentOS 6.3+, 7.0+	CentOS 6.3: <a href="#">LIS Download</a> CentOS 6.4+: In Kernel	Package: In <a href="#">OpenLogic repo</a> under "WALinuxAgent" Source Code: <a href="#">GitHub</a>
CoreOS	494.4.0+	In Kernel	Source Code: <a href="#">GitHub</a>
Debian	Debian 7.9+, 8.2+	In Kernel	Package: In repo under "waagent" Source Code: <a href="#">GitHub</a>
Oracle Linux	6.4+, 7.0+	In Kernel	Package: In repo under "WALinuxAgent" Source Code: <a href="#">GitHub</a>
Red Hat Enterprise Linux	RHEL 6.7+, 7.1+	In Kernel	Package: In repo under "WALinuxAgent" Source Code: <a href="#">GitHub</a>
SUSE Linux Enterprise	SLES 11 SP4, SLES 12 SP1+ and SLES for SAP 11 SP3+	In Kernel	Package: In <a href="#">Cloud:Tools</a> repo under "python-azure-agent" Source Code: <a href="#">GitHub</a>
openSUSE	openSUSE 13.2+	In Kernel	Package: In <a href="#">Cloud:Tools</a> repo under "python-azure-agent" Source Code: <a href="#">GitHub</a>
Ubuntu	Ubuntu 12.04, 14.04, 16.04, 16.10	In Kernel	Package: In repo under "walinuxagent" Source Code: <a href="#">GitHub</a>



# Lab 2

## Scaling on Azure

¡Descarga tu material!

<https://github.com/psmapad/ExpertAcademy>

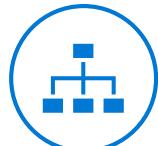


# Mundo Containers



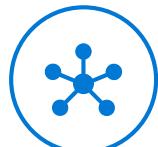
# Microservicios

## ¿Qué son?



### Un estilo arquitectónico de software

Las aplicaciones se componen de pequeños módulos independientes que se comunican entre sí mediante API bien definidas. No específico de la plataforma



### Desacoplados

Estos módulos de servicio son bloques de construcción altamente desacoplados que son lo suficientemente pequeños para implementar una funcionalidad única pero que juntos pueden formar sistemas más grandes



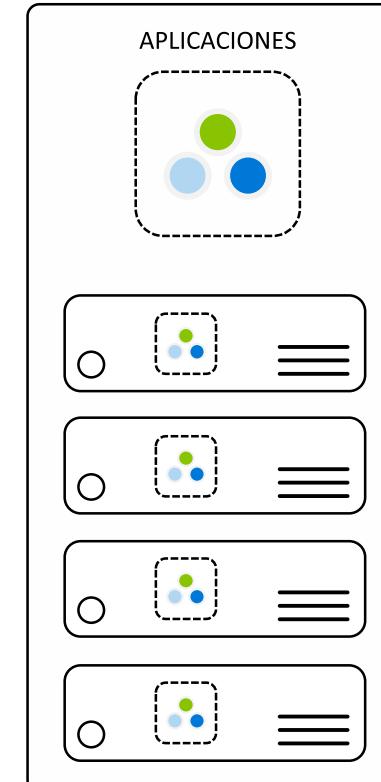
### Independientemente versionado, implementado y escalado

Con una arquitectura de microservicios, los desarrolladores pueden crear, gestionar y mejorar los servicios de aplicaciones de forma independiente, incluso utilizando diferentes idiomas.

Los contenedores proporcionan el formato consistente y el aislamiento deseado por los microservicios

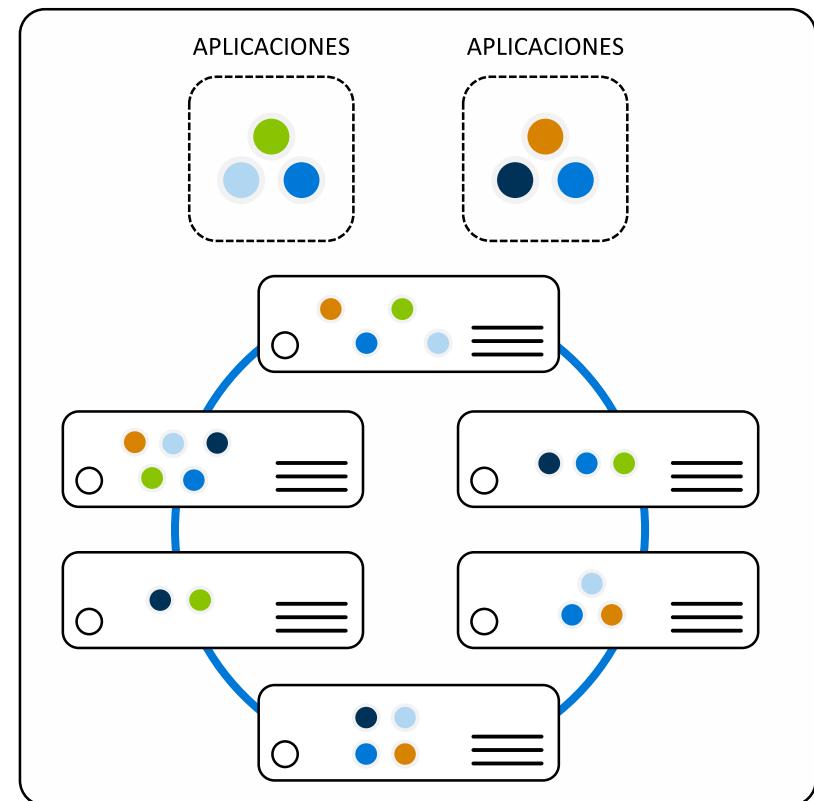
### Monolítico

Aplicación con todo incluido



### Microservicios

Servicios pequeños e independientes



# Microservicios

---



# Contenedores

¿Qué es?

No es un objeto real.

Es un mecanismo de despliegue de aplicaciones con un proceso de aislamiento basado en muchas de las características del kernel de Linux.

Elemento de software

Es un elemento de software estándar que empaqueta el código y las dependencias de un aplicativo para tener la capacidad de funcionar en cualquier ambiente.

# Contenedores

## Problemática

### Desarrollo

- Necesidad de construir sin preocuparse por la infraestructura.
- “Las aplicaciones si corren en mi máquina”.
- La productividad e innovación se detiene cuando infra configura.



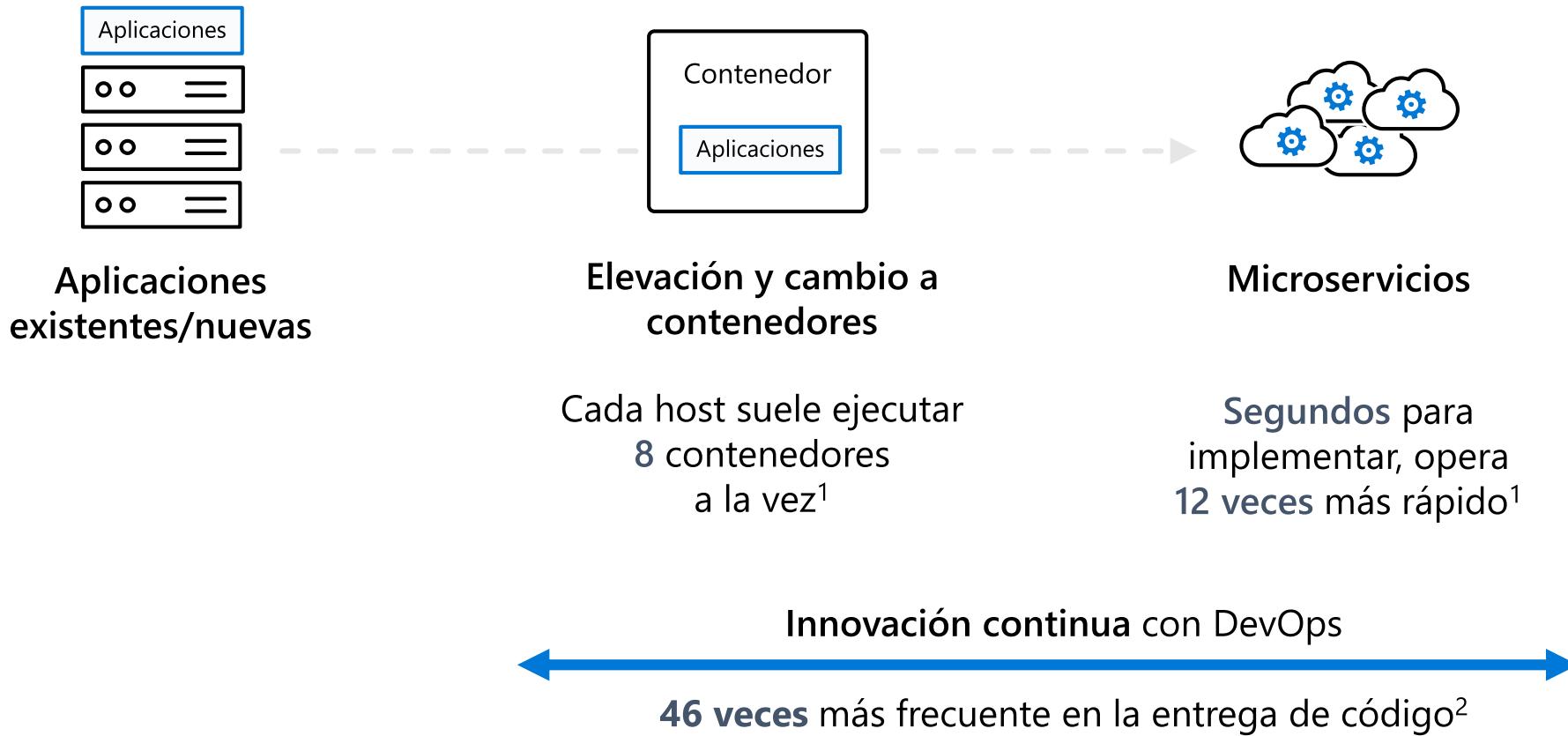
### Operaciones/Infra IT

- Necesidad de organizar servidores y tenerlos disponibles.
- No conozco la aplicación y necesito ayuda del desarrollador.
- No me puedo enfocar en seguridad de la red si no puedo garantizar el funcionamiento.
- “No funciona la aplicación”.



# Contenedores

## Potencial



1: Informe Datalog: 8 hechos sorprendentes sobre la adopción de Docker; 2: Informe sobre el estado de los dispositivos en 2017

# Docker

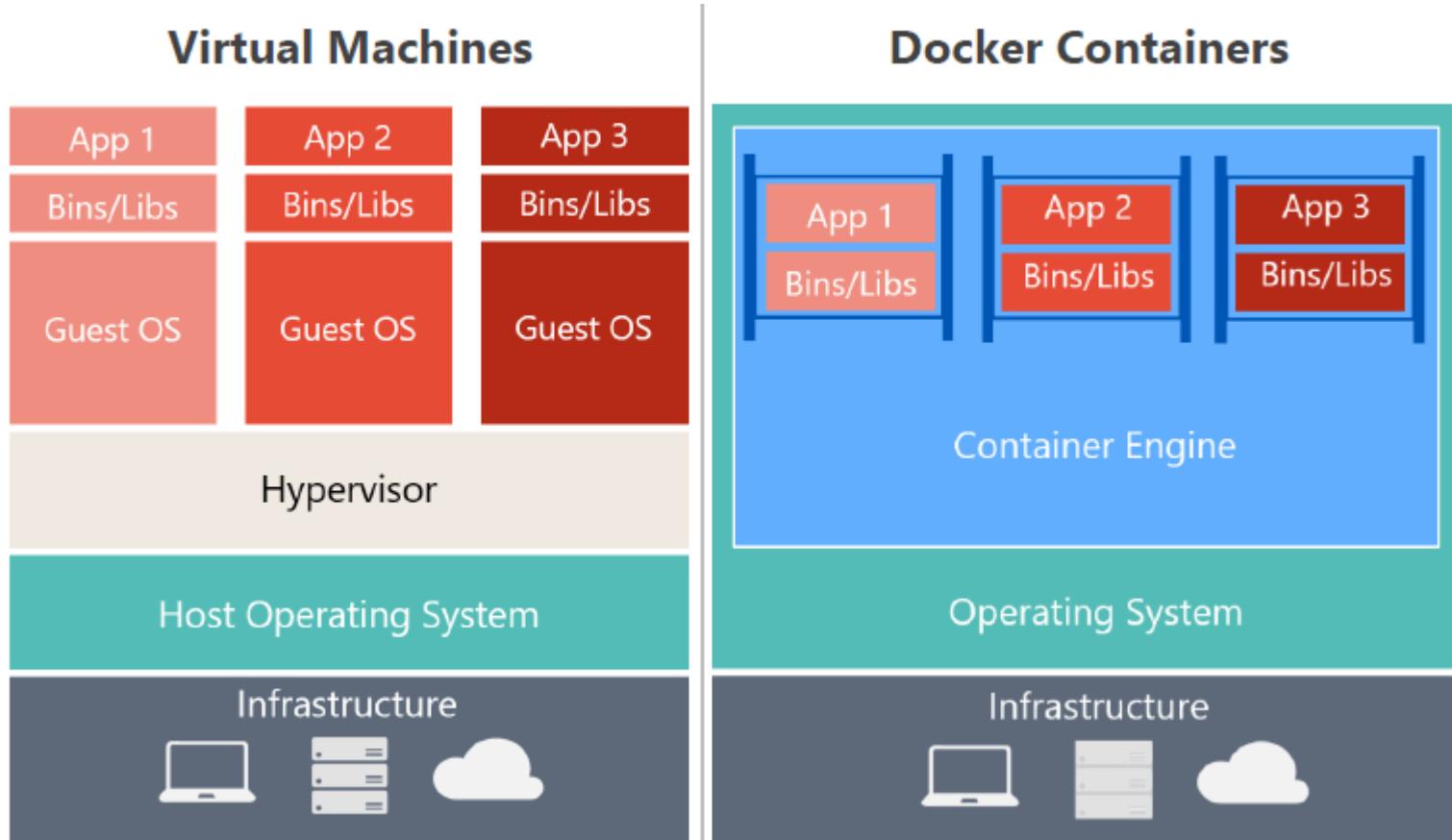
## Concepto

- OOS Container Runtime.
- Soporte para Mac, Linux y Windows.
- Herramienta para línea de comandos.
- Formato “Dockerfile”.
- Docker engine.
- Docker image.



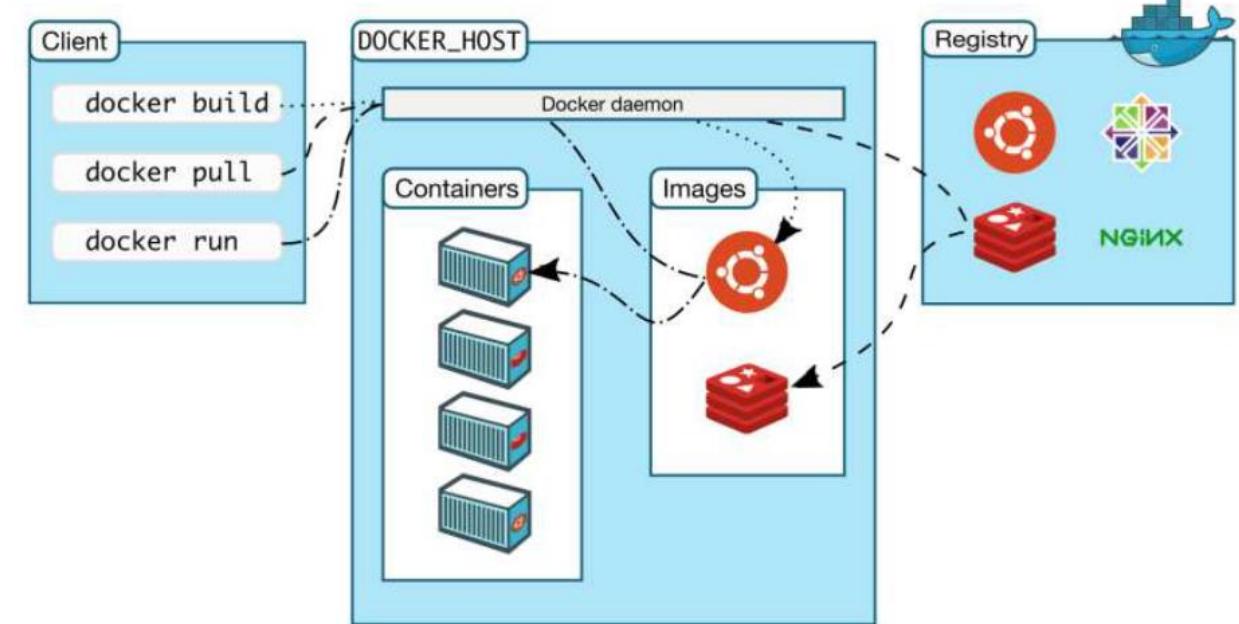
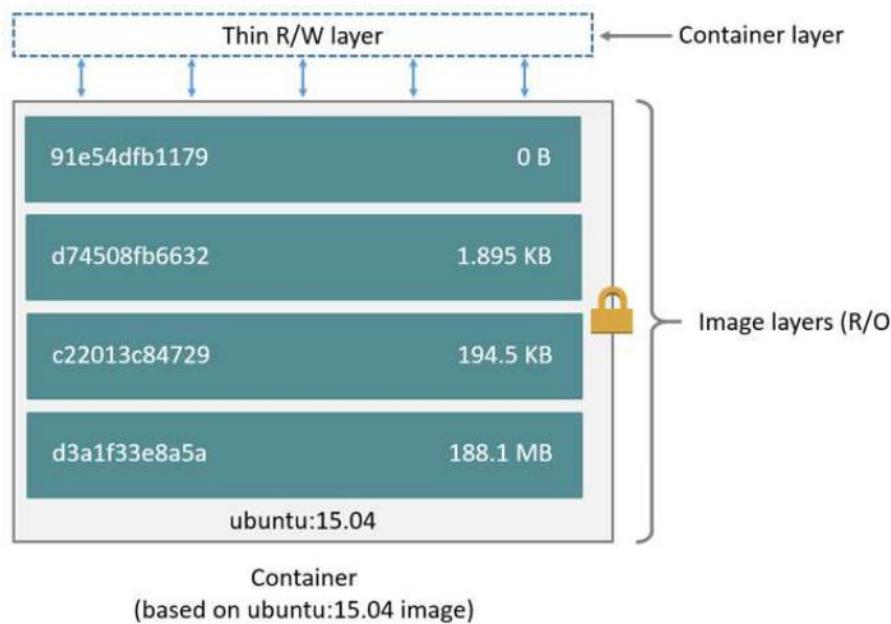
# VMs vs Contenedores

## Infraestructura



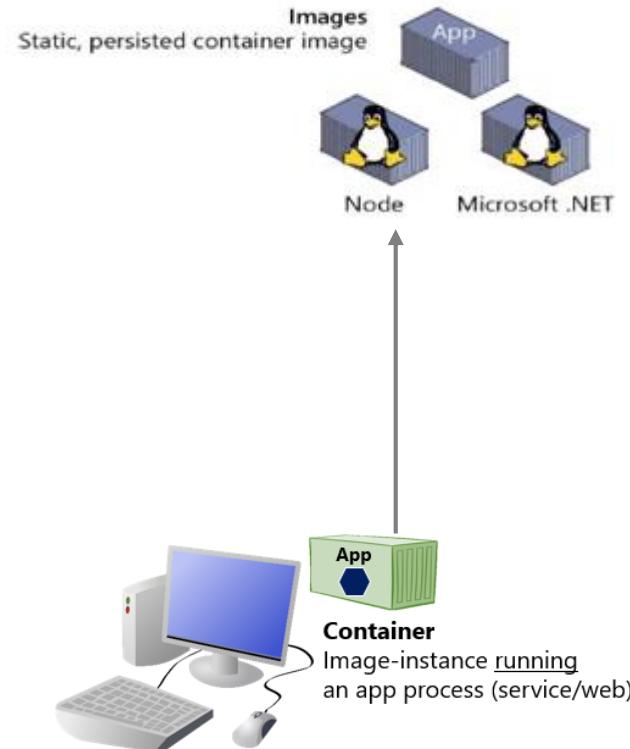
# Docker

## Arquitectura



# Docker

## Taxonomía



Hosted Docker Registry

Docker Trusted Registry on-prem.

Docker Hub Registry

Docker Trusted Registry on-cloud

Azure Container Registry

AWS Container Registry

Google Container Registry

Quay Registry

Other Cloud

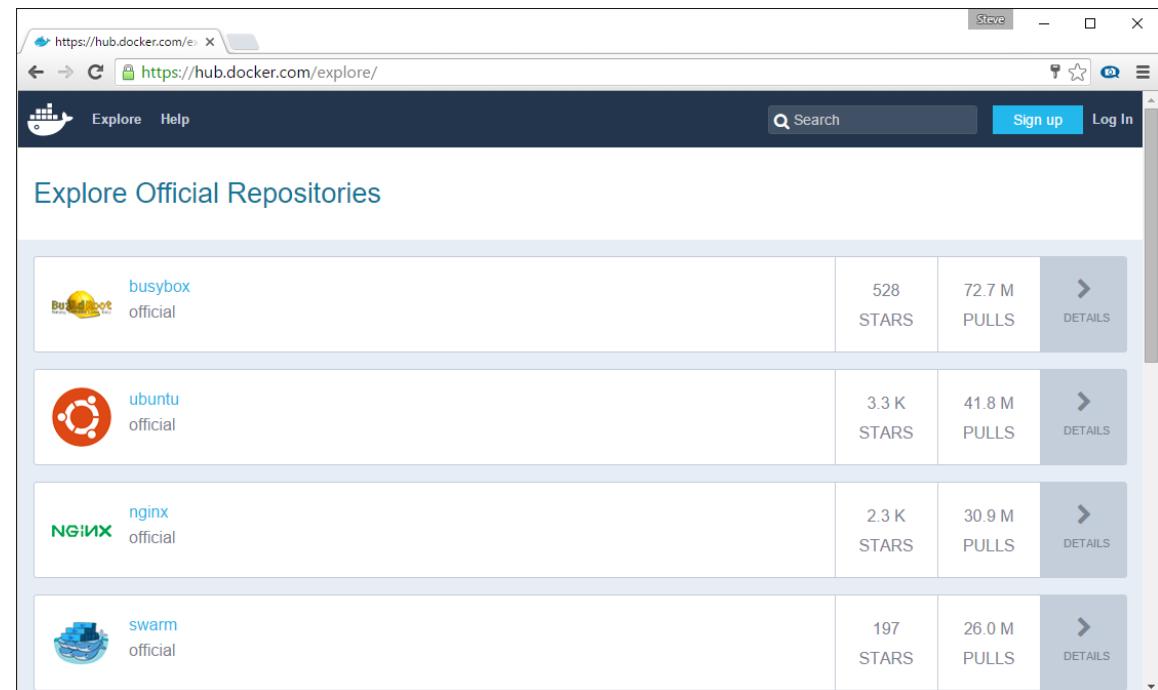
**On-premises**  
(‘n’ private organizations)

**Public Cloud**  
(specific vendors)

# Docker

## Registry

- Almacena las imágenes de Docker
- Imágenes fácilmente localizables.
- Registro publico – Hub.Docker.com
- Registros privados
  - Para instancias personales



# Docker

## Características

- Se pueden empaquetar más aplicaciones que una máquina virtual.
- Aislamiento de aplicaciones con todas sus funcionalidades y dependencias.
- El sistema operativo es compartido por múltiples aplicaciones.
- El Kernel es comparativo.
- Eficiencia en el despliegue.
- Escalamiento ágil.

# Contenedores

¿Cómo ayuda Azure?



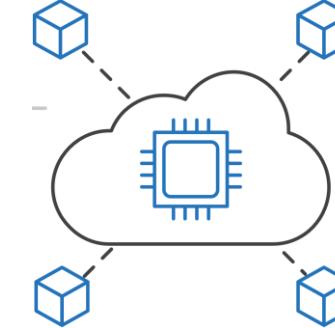
## Flexibilidad

Implemente aplicaciones en contenedores en su entorno preferido



## Productividad

Acelere el desarrollo de aplicaciones en contenedores



## Confianza

Gestione, supervise, y proteja sus contenedores fácilmente

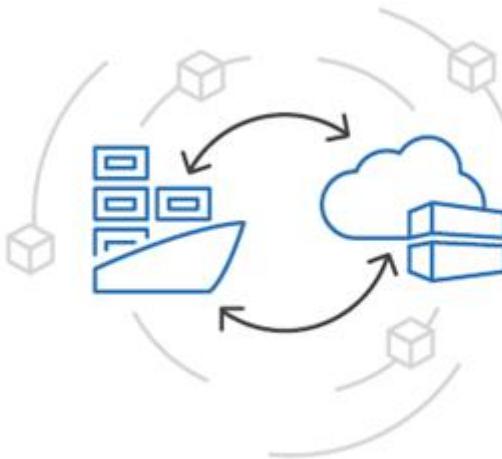
# ACR

Azure Container Registry

## Manejo de un registro de imágenes privado como recurso de Azure



Gestión de imágenes de cualquier tipo de contenedores

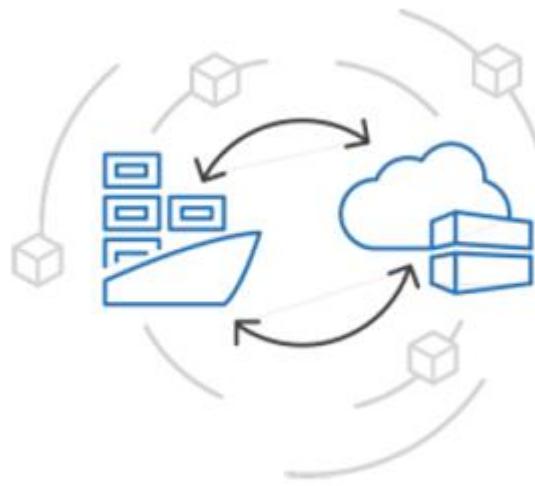


Integración de herramientas de consola

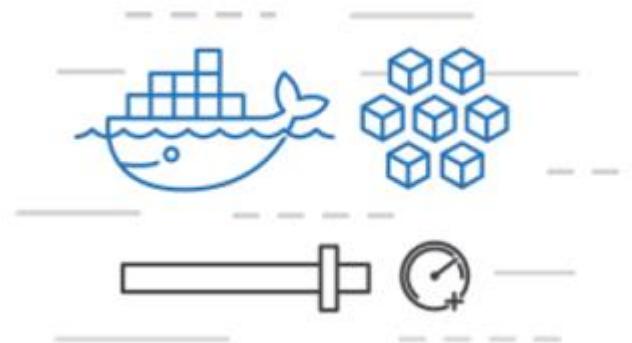


Ventajas de geo replicación

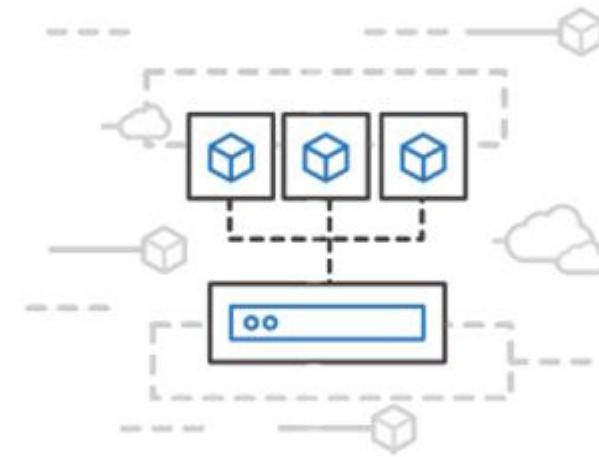
### Ejecución de instancias de contenedores con un simple comando



Iniciación con instancias de contenedores de manera sencilla



Escalamiento aprovechando un servicio de nube



Aislamiento de tipos de contenedores

# Contenedores

## Azure entornos

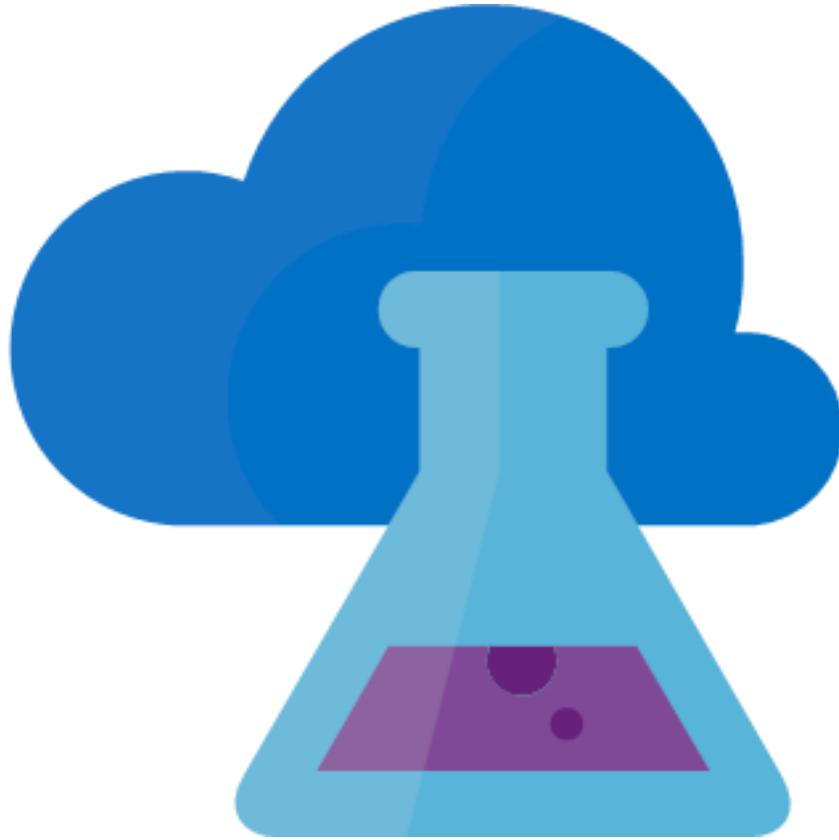


ACR



Hub de Docker

Elección de herramientas y clientes  
para desarrolladores



# Lab 3

## Containers on Azure

¡Descarga tu material!

<https://github.com/psmapad/ExpertAcademy>



# Kubernetes on Azure



# Orquestadores

## Concepto

Google Kubernetes



Azure Service Fabric

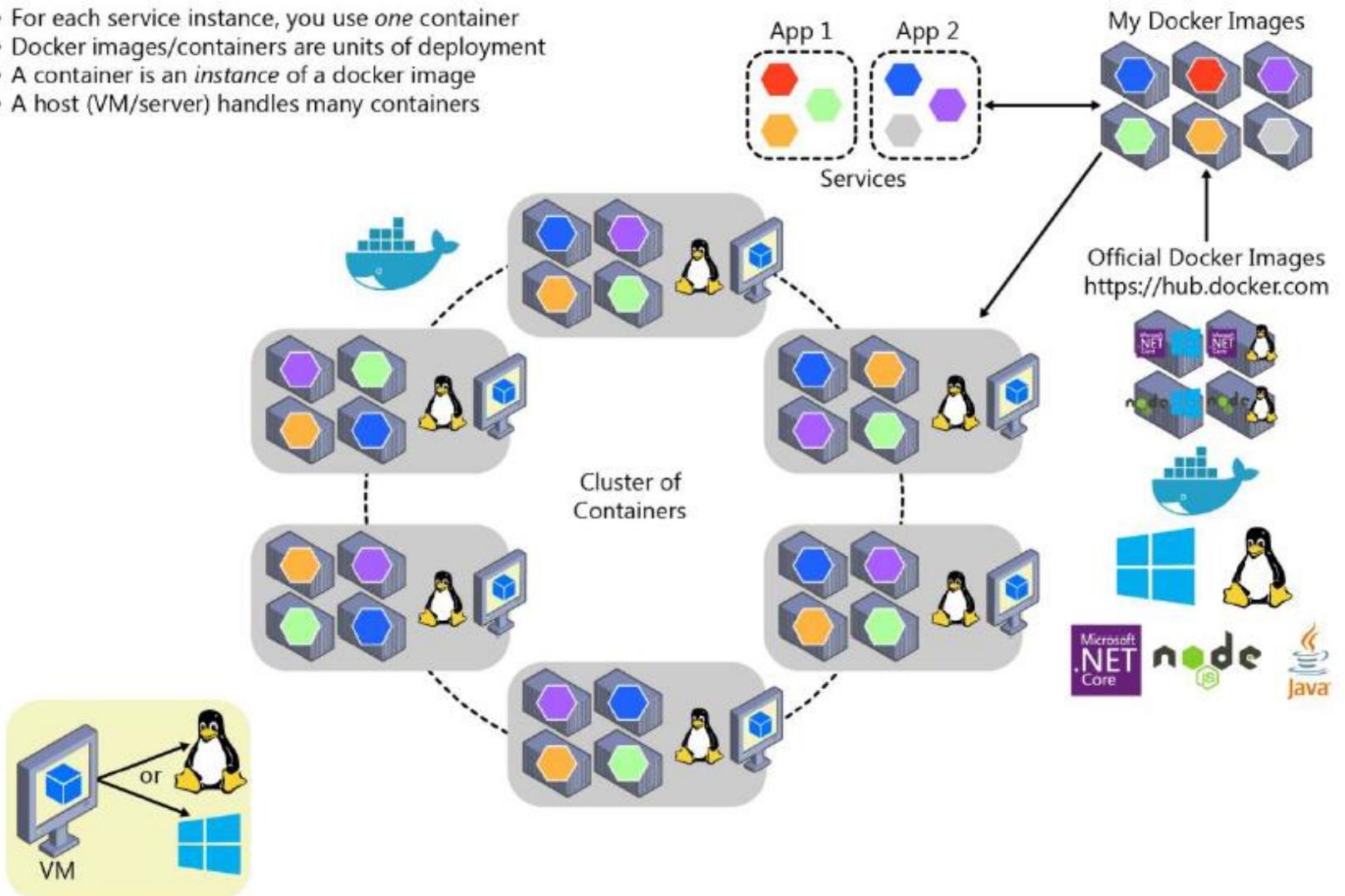


Docker Swarm



 **RED HAT<sup>®</sup>**  
**OPENSHIFT**  
Container Platform

- For each service instance, you use one container
- Docker images/containers are units of deployment
- A container is an *instance* of a docker image
- A host (VM/server) handles many containers



# Kubernetes

## Tendencia

"En 2020, más del 50% de las empresas utilizarán aplicaciones de misión crítica y contenedorizadoras nativas de la nube en la producción".

**Gartner**

El tamaño promedio de una implementación de contenedores ha crecido un 75% en un año.<sup>1</sup>



La mitad del entorno del contenedor está orquestado.<sup>1</sup>

El 77% de las empresas<sup>2</sup> que utilizan orquestadores de contenedores eligen a Kubernetes.

77%

Las empresas más grandes están liderando la adopción.<sup>1</sup>

Casi el 50% de las organizaciones<sup>1</sup> que ejecutan 1000 o más hosts han adoptado contenedores.

50%

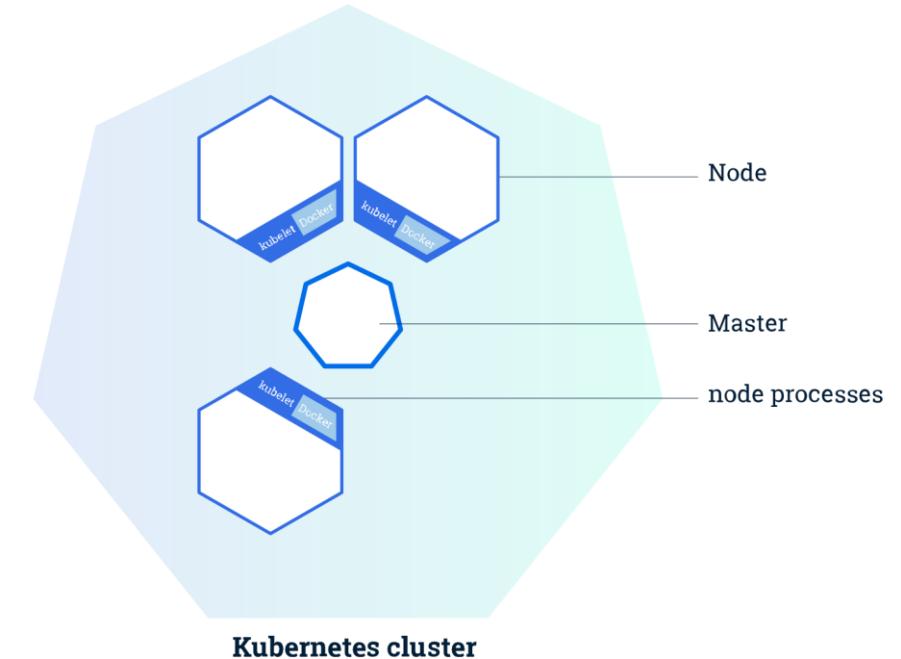
<sup>1</sup> Informe de Datadog: 8 hechos sorprendentes sobre la adopción de Docker

<sup>2</sup> Encuesta CNCF: tecnologías nativas de la nube - escalado - producción - aplicaciones

# Kubernetes

## Concepto

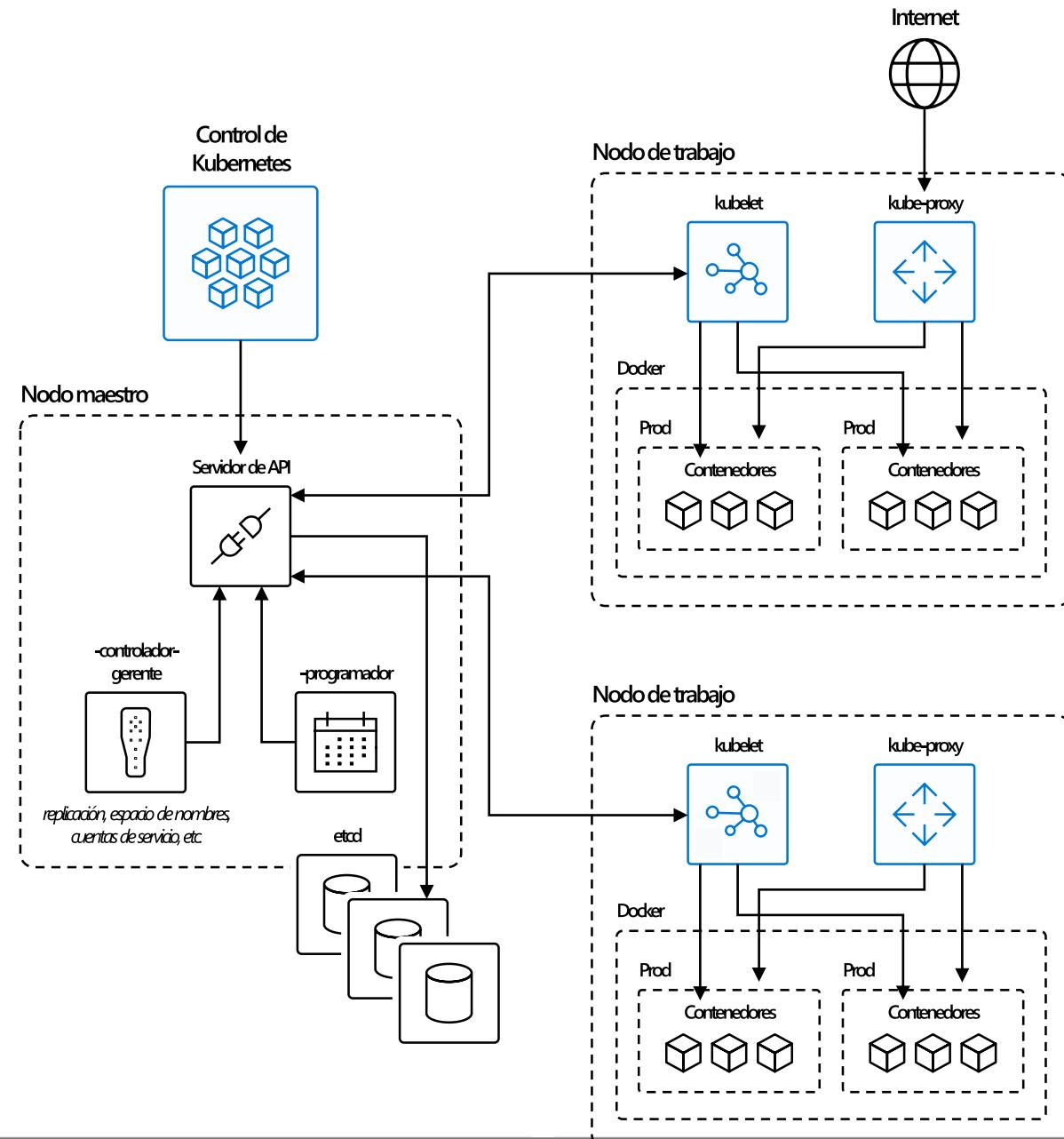
- Sistema OSS
- Aplicaciones containerizadas
- Despliegue automatizado
- Clúster de equipos como unidades funcionales
- Diseñado por Google
- Balanceador de carga integrado
- Escalamiento horizontal



# Kubernetes

## Funcionamiento

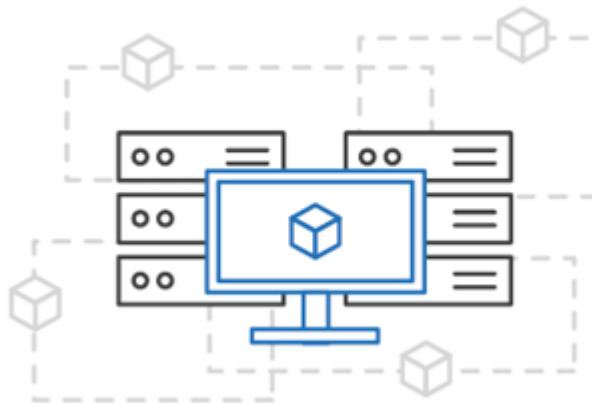
1. Los usuarios de Kubernetes se comunican con el servidor API y aplican el estado deseado.
2. Los nodos maestros aplican activamente el estado deseado en los nodos de trabajo.
3. Los nodos de trabajo soportan la comunicación entre los contenedores.
4. Los nodos de trabajo soportan la comunicación desde Internet.



# AKS

Azure Kubernetes Service

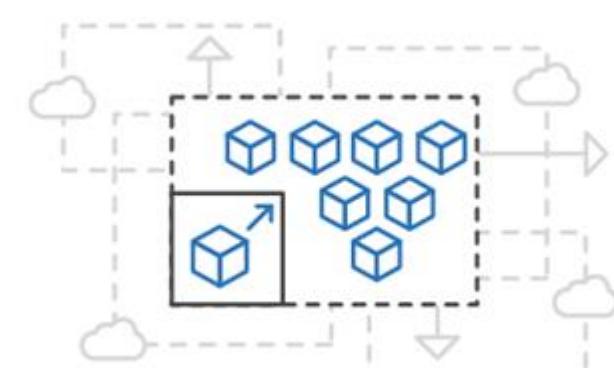
Simplifica el despliegue, gestión y operación de Kubernetes



Enfoque en contenedores y  
no en infraestructura.



Trabajo con cualquier  
tecnología gracias a la  
cantidad de APIs existentes



Escalamiento y ejecución  
de las aplicaciones con  
seguridad

Kubernetes le permite hacer más

Concéntrese en sus contenedores y código, no en sus esquemas

Responsabilidades	"Házlo tú mismo" con Kubernetes	Kubernetes en Azure
Contenedorización		
Iteración de aplicación, depuración		
CI/CD		
Hospedaje de clúster		
Actualización de clúster		
Parches		
Escala		
Monitoreo y registro		

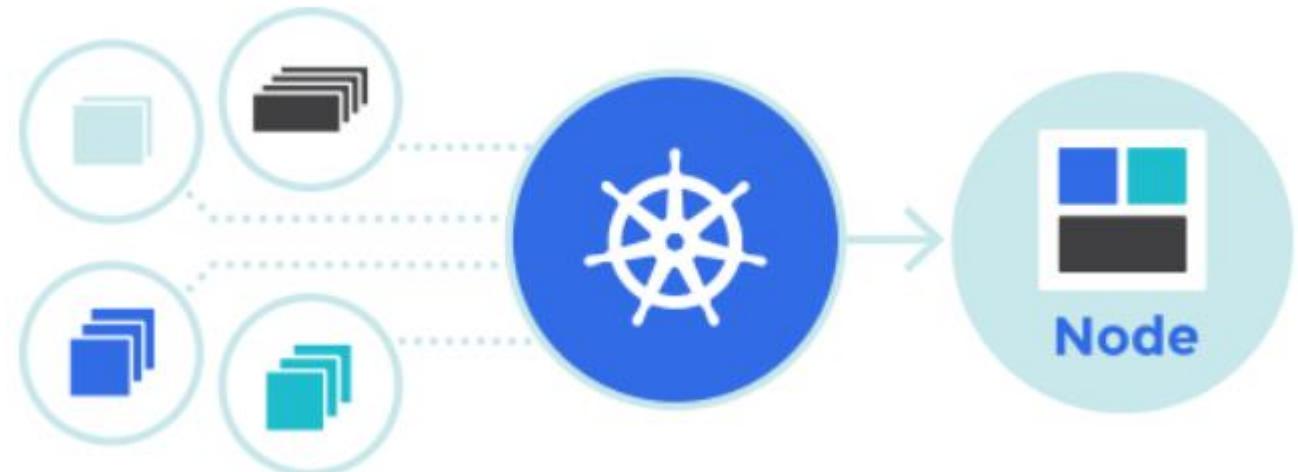
Cliente

Microsoft

# AKS

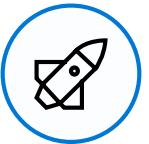
## Características

- Despliegue y gestión de Kubernetes desde el portal de Azure o Azure CLI.
- Proporciona seguridad a clúster con la integración con AAD Azure KeyVault o CNI.
- Optimiza el despliegue de aplicaciones con gran número de contenedores.
- Integración con Azure DevOps dentro del pipeline de CI/CD.
- Ventajas de ser un IaaS



# AKS

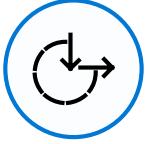
## Características



Implemente y gestione  
Kubernetes con facilidad



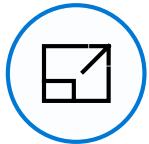
Acelere el desarrollo de  
aplicaciones en  
contenedores



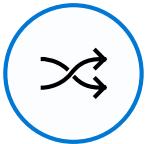
Configurar CI / CD  
en unos pocos clics



Proteja su entorno  
Kubernetes



Escale y ejecute  
aplicaciones con confianza

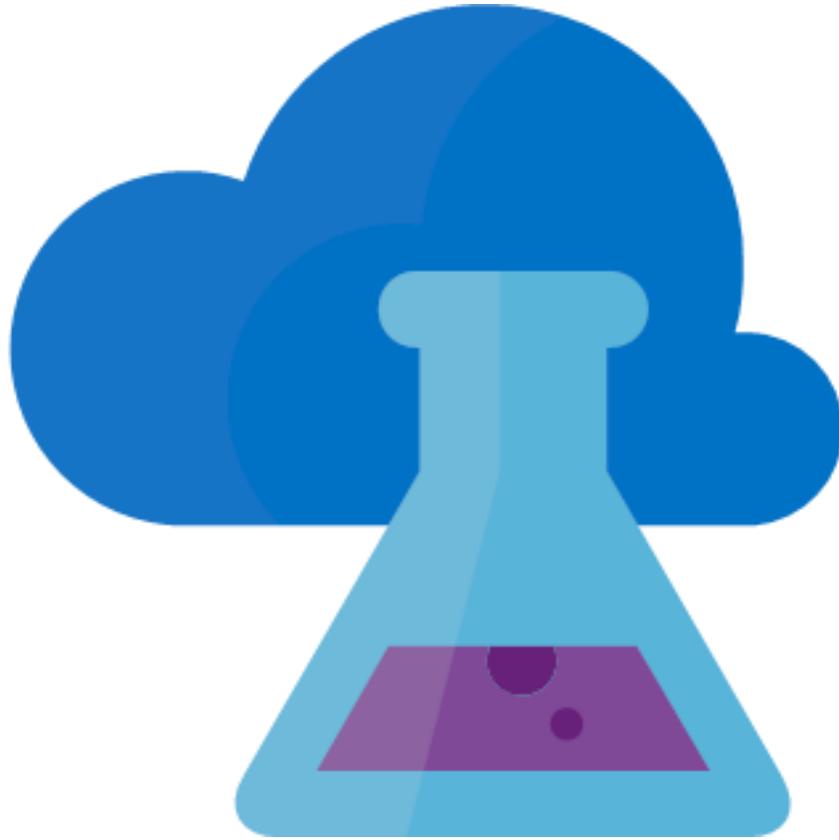


Trabaje como quiera con  
herramientas de código  
abierto y API

# AKS

## Multiplataforma

	Desarrollo	DevOps	Monitoreo	Redes	Almacenamiento	Seguridad
Aproveche los servicios y herramientas del ecosistema Kubernetes.	 	     	     	 	 	   RBAC
...o... Aproveche el creciente soporte de Azure	 Código VS	 Azure DevOps 				Registro de Contenedores Azure AAD Key Vault



# Lab 4

## Kubernetes on Azure

¡Descarga tu material!

<https://github.com/psmapad/ExpertAcademy>



GROW THE BUSINESS TOGETHER

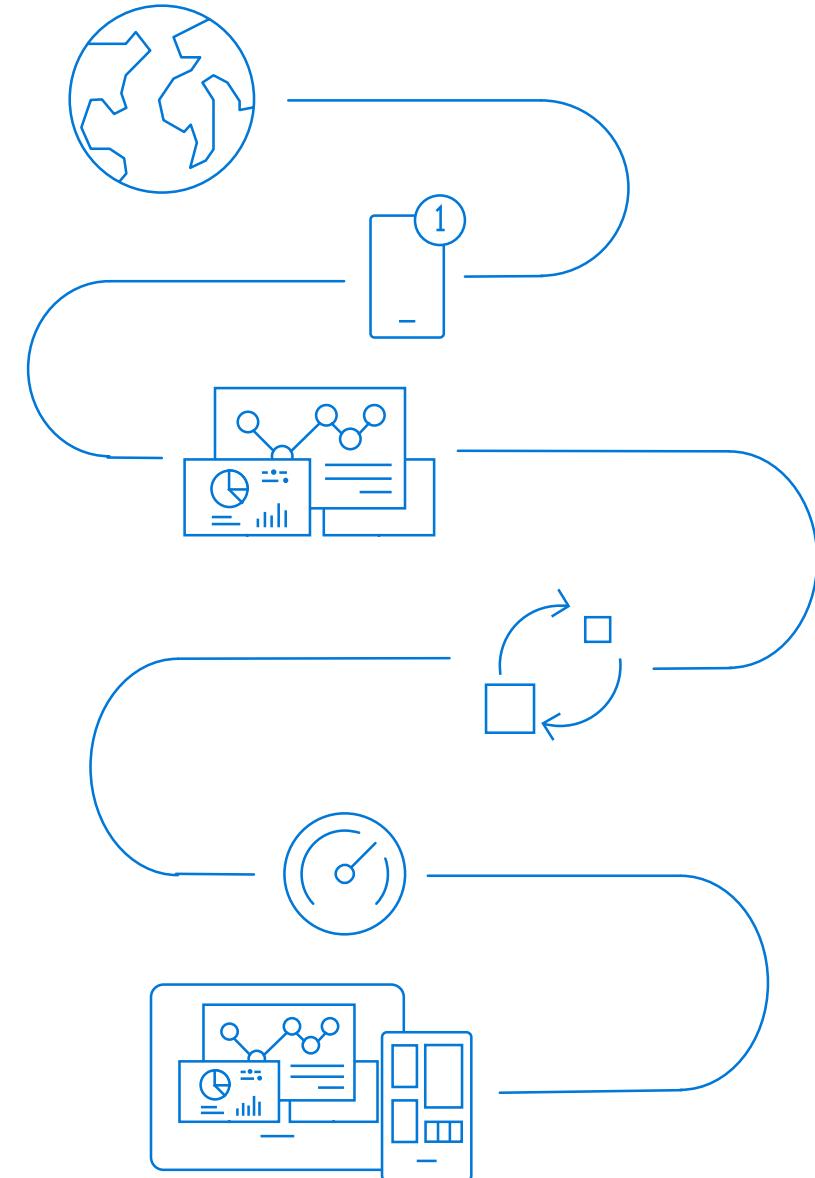
# CosmosDB on Azure



# Lluvia de ideas

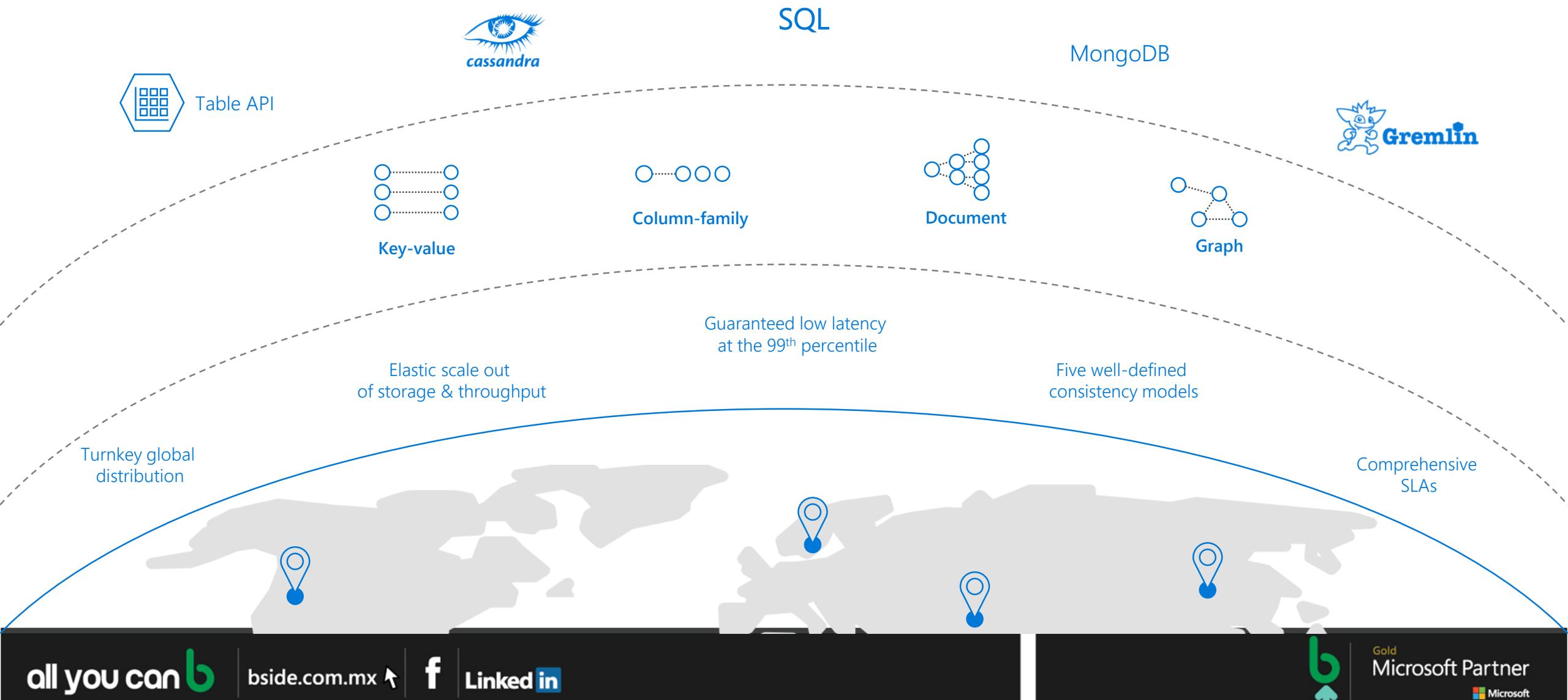
Las aplicaciones modernas enfrentan nuevos retos

- La administración y sincronización de datos distribuidos alrededor del mundo.
- Entrega de personalización de corte tiempo de respuesta, en tiempo real.
- Procesamiento y análisis de enormes cantidades de datos complejos.
- Escalamiento de la capacidad de rendimiento y de almacenamiento basado en la demanda.
- Oferta de acceso a datos con baja latencia a usuarios globales.
- Modernización de aplicaciones existentes y datos.



# ¿Qué es Azure Cosmos DB?

Un servicio de base de datos distribuido globalmente, escalable de forma masiva y multi-modo



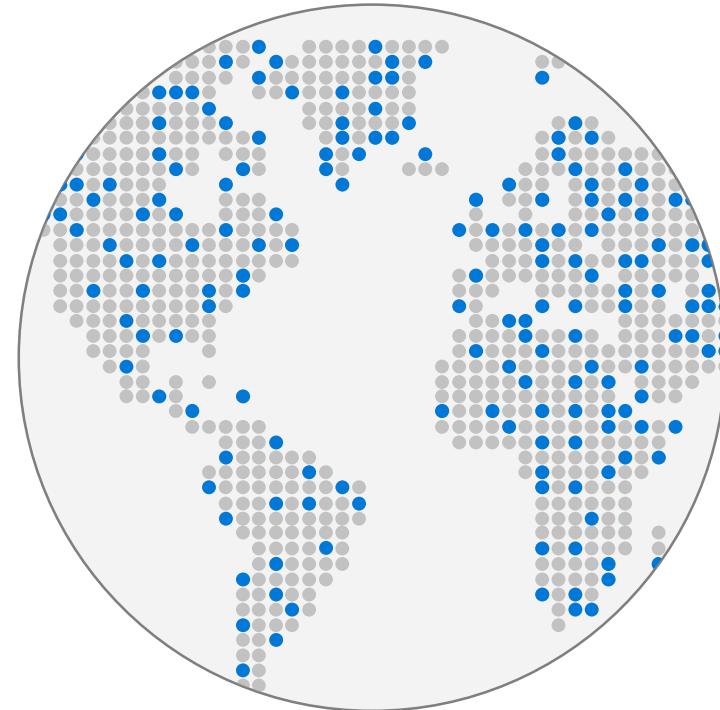
# La llave para la distribución global

¿Qué es Azure Cosmos DB?

Coloque sus datos en donde se encuentren sus usuarios

Los datos se replican de forma automática alrededor del mundo y a través de múltiples regiones.

- Disponible en **todas** las regiones disponibles de Azure
- Failover manual y automático
- Replicación multi región de forma automática y síncrona



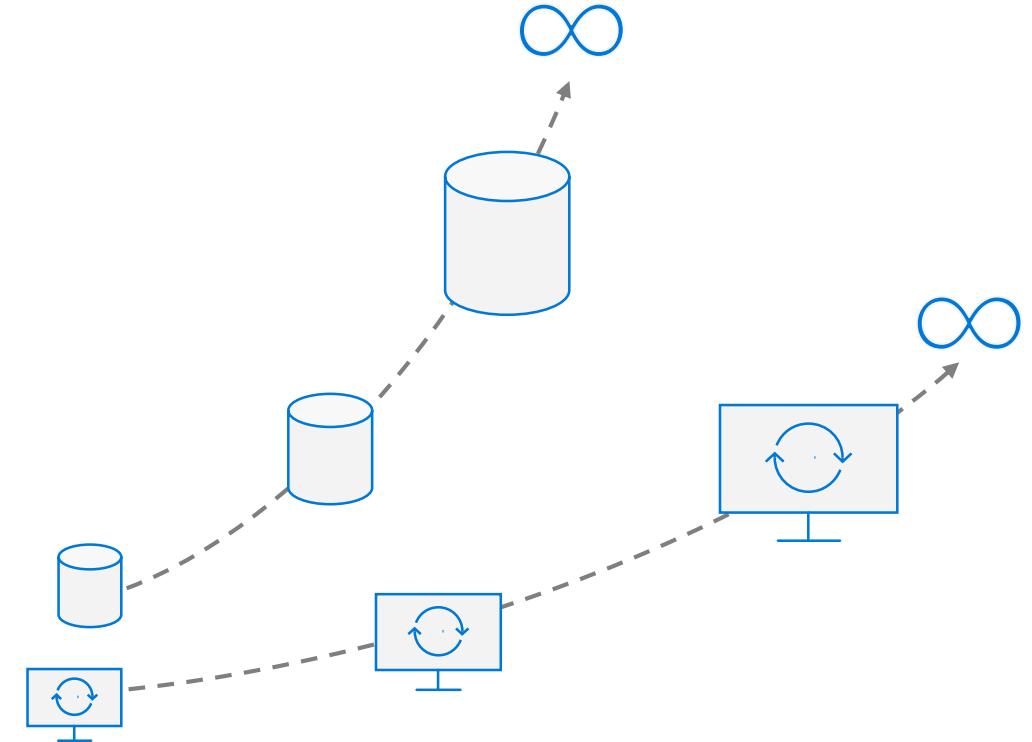
# Scale out elástico para rendimiento y almacenamiento

¿Qué es Azure Cosmos DB?

## Escale en base a las necesidades de sus aplicaciones

Escalamiento de las capacidades de almacenamiento de datos y rendimiento de forma elástica e independiente a través de las diferentes regiones, aún durante cargas de tráfico grandes e impredecibles.

- Escalamiento elástico de 10 -> 100s de millones de solicitudes/seg.
- Solicitudes/seg de diferentes cargas de trabajo.
- Baja latencia, acceso a lecturas en < 10 ms y escrituras de < 15 ms.



# Modelos de consistencia

¿Qué es Azure Cosmos DB?

Proporciona control sobre las compensaciones de la consistencia del rendimiento, respaldadas por las políticas de SLA.

Un modelo de programación intuitivo que ofrece baja latencia y alta disponibilidad para las aplicaciones.



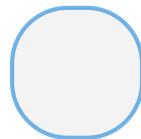
Strong



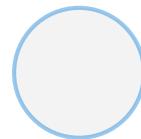
Bounded-stateless



Session



Consistent prefix



Eventual



# Modelos de datos y APIs

¿Qué es Azure Cosmos DB?

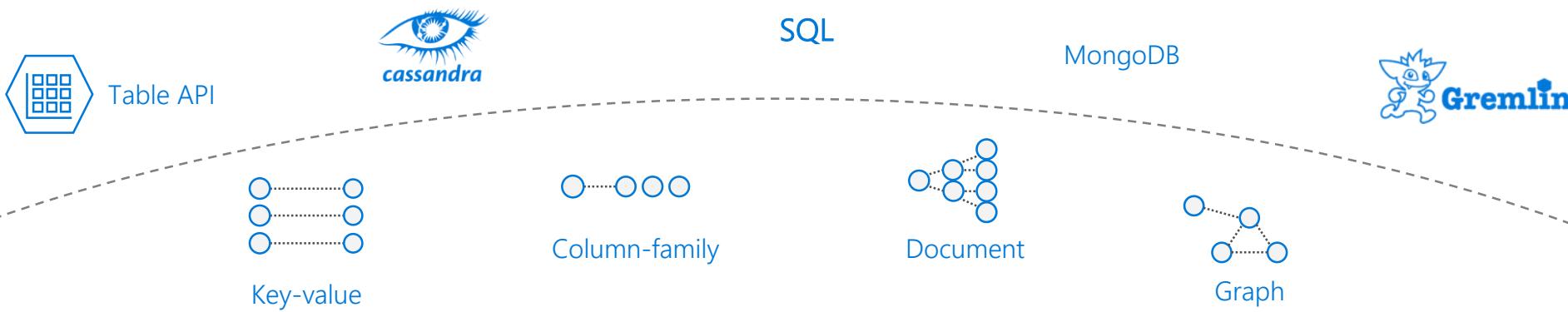
## Adaptable a APIs, herramientas y marcos de trabajo

Acceso a datos como: SQL, OSS APIs, soporte nativo para cargas de trabajo NoSQL.

Uso como: llave-valor, tabular, grafo y documentos.

Los datos son *indexados* de forma automática, sin necesidad de un esquema o índices secundarios.

Búsquedas de datos sin *lag*.



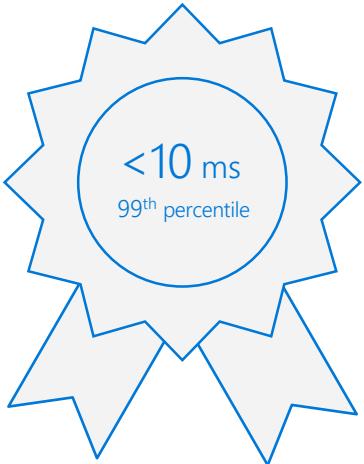
# SLAs

¿Qué es Azure Cosmos DB?

## Infraestructura de clase mundial

Azure Cosmos DB es el único servicio que se encuentra respaldado por una latencia de milisegundos en el percentil 99º, es decir, 99.999% HA bajo la garantía de rendimiento y consistencia.

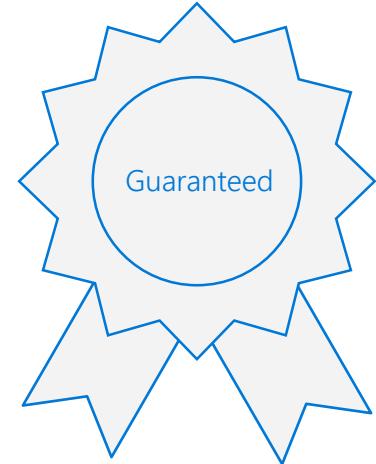
Latency



HA



Throughput



Consistency



# Líder de industria en seguridad y cumplimiento

¿Qué es Azure Cosmos DB?

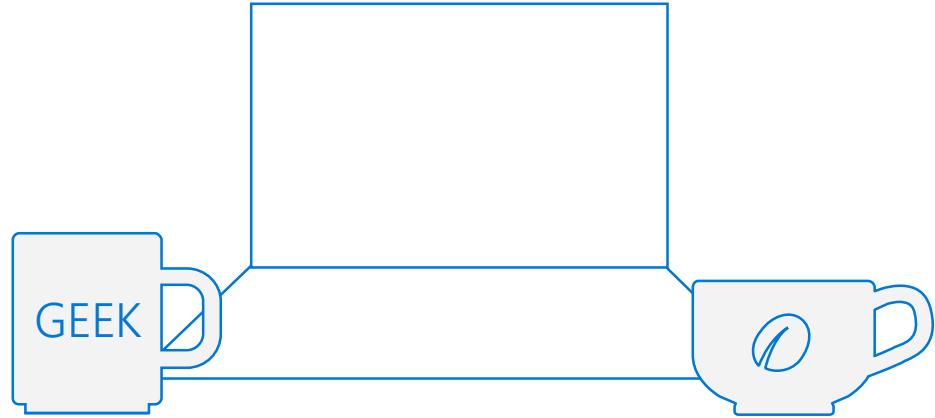
Azure es la plataforma de nube más confiable del mundo, con más certificaciones que cualquier otro proveedor de nube

- Seguridad de grado empresarial
- Cifrado en reposo
- Cifrado de datos habilitado de forma automática por defecto
- Certificación completa de cumplimiento de Azure



# BD sin esquema o indexación

¿Qué es Azure Cosmos DB?



Azure Cosmos DB es un servicio sin esquema de BD que indexa los datos de forma automática sin importar el modelo de datos para entregar resultados de búsquedas con la mejor velocidad.

- Administración de índices automática.
- Auto indexación síncrono
- No requiere esquemas o índices secundarios
- Trabaja sobre cualquier modelo de datos

Item	Color	Microwave safe	Liquid capacity	CPU	Memory	Storage
Geek mug	Graphite	Yes	16oz	???	???	???
Coffee Bean mug	Tan	No	12oz	???	???	???
Surface book	Gray	???	???	3.4 GHz Intel Skylake Core i7-6600U	16GB	1 TB SSD



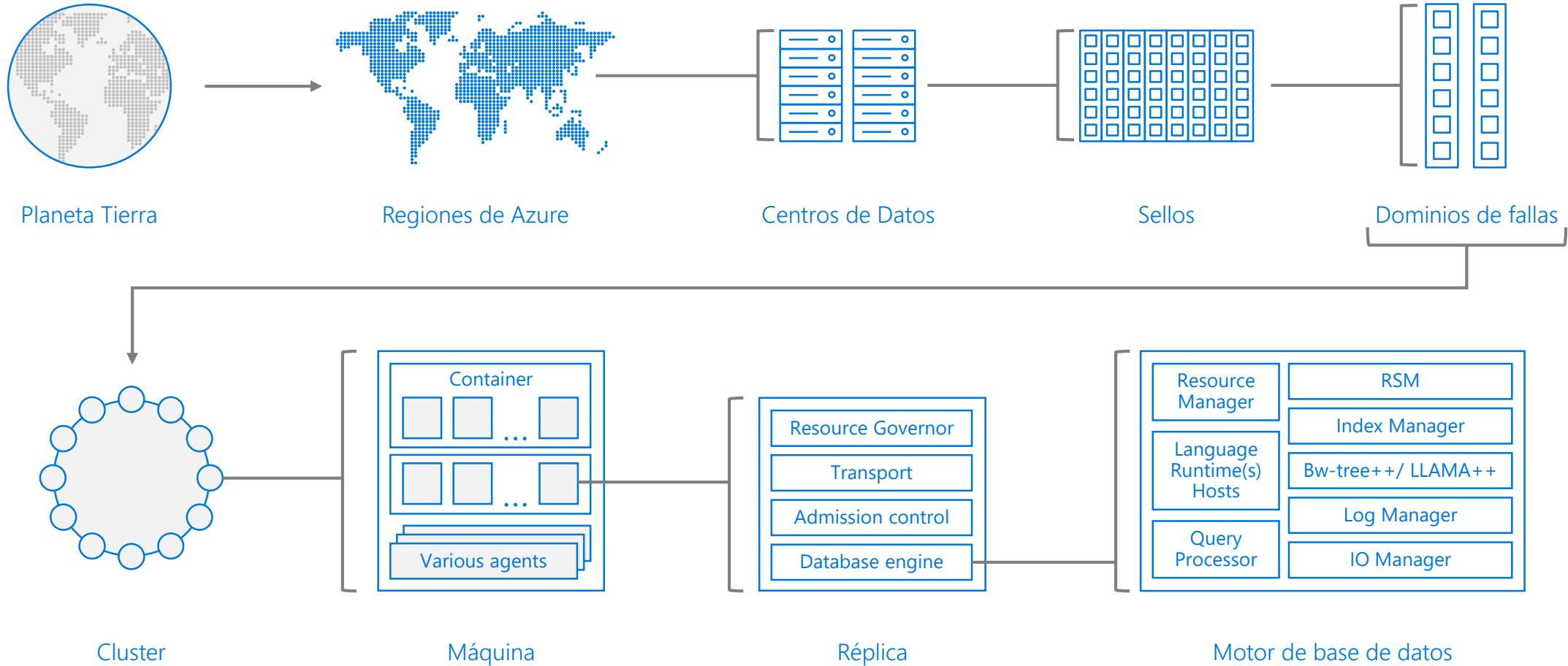
GROW THE BUSINESS TOGETHER

# CosmosDB - Particionamiento



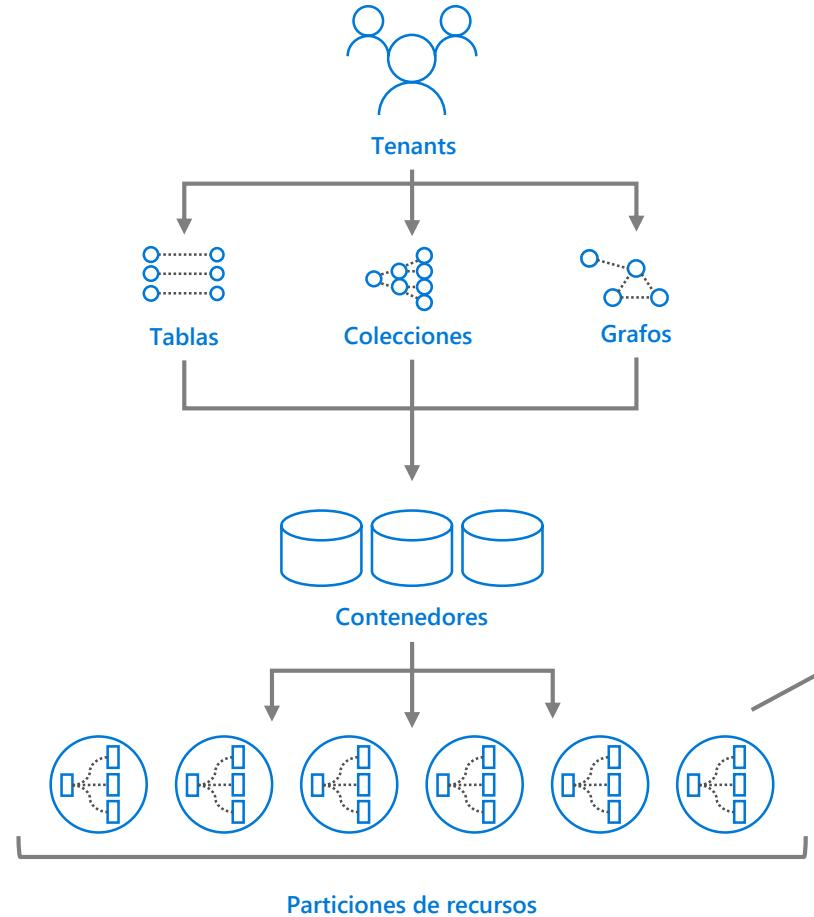
# Topología del sistema

## Partitionamiento



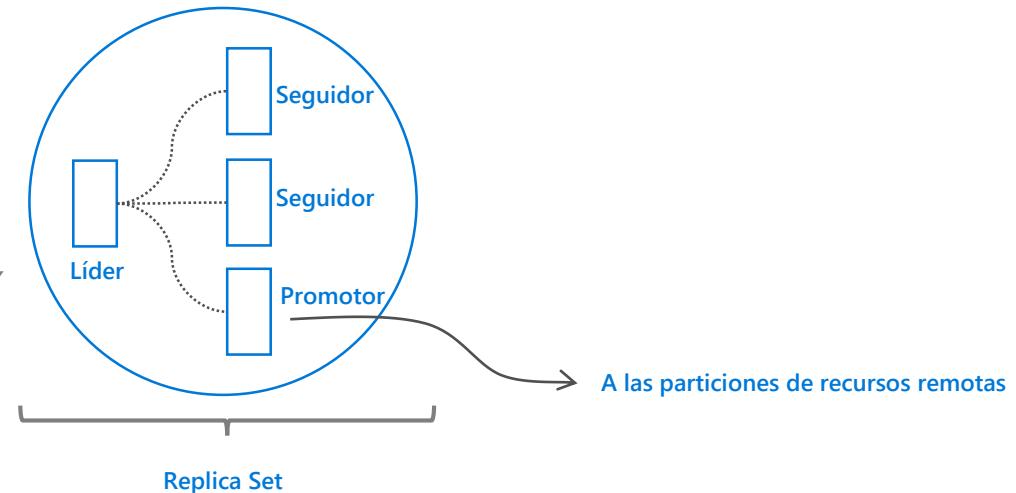
# Jerarquía de recursos

## Partitionamiento



## Contenedores

Los recursos lógicos "emergen" a las API como tablas, colecciones o grafos, que están constituidos por una o más particiones o servidores físicos.



## Particiones de recursos

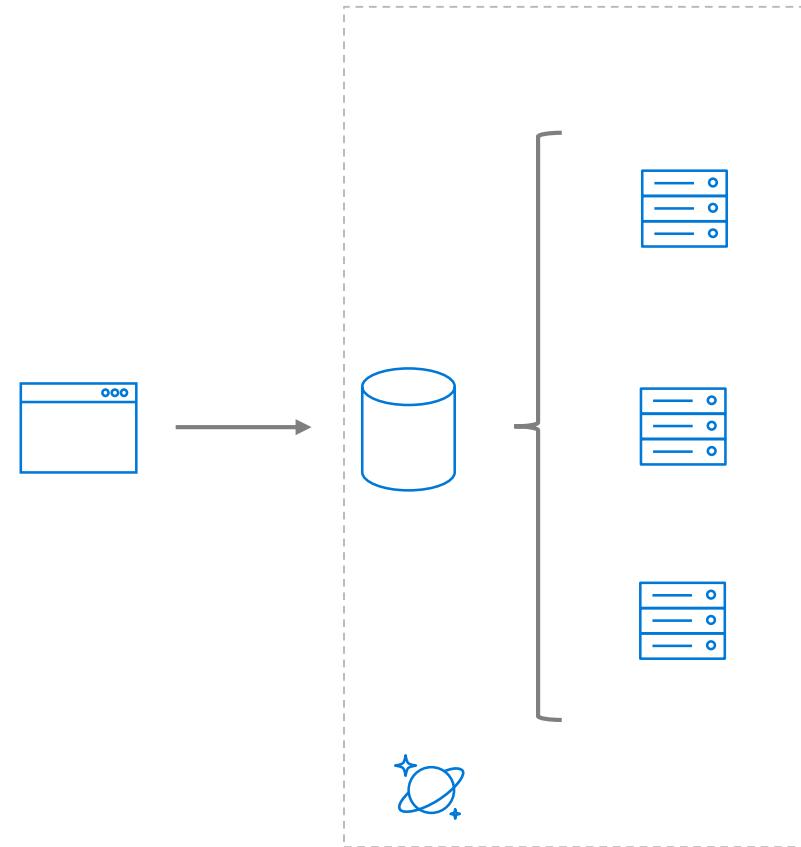
- Primitivas de coordinación coherentes, altamente disponibles y gobernadas por recursos
- Consisten en conjuntos de réplicas, con cada réplica alojando una instancia del motor de base de datos

# Particionamiento

## Particionamiento

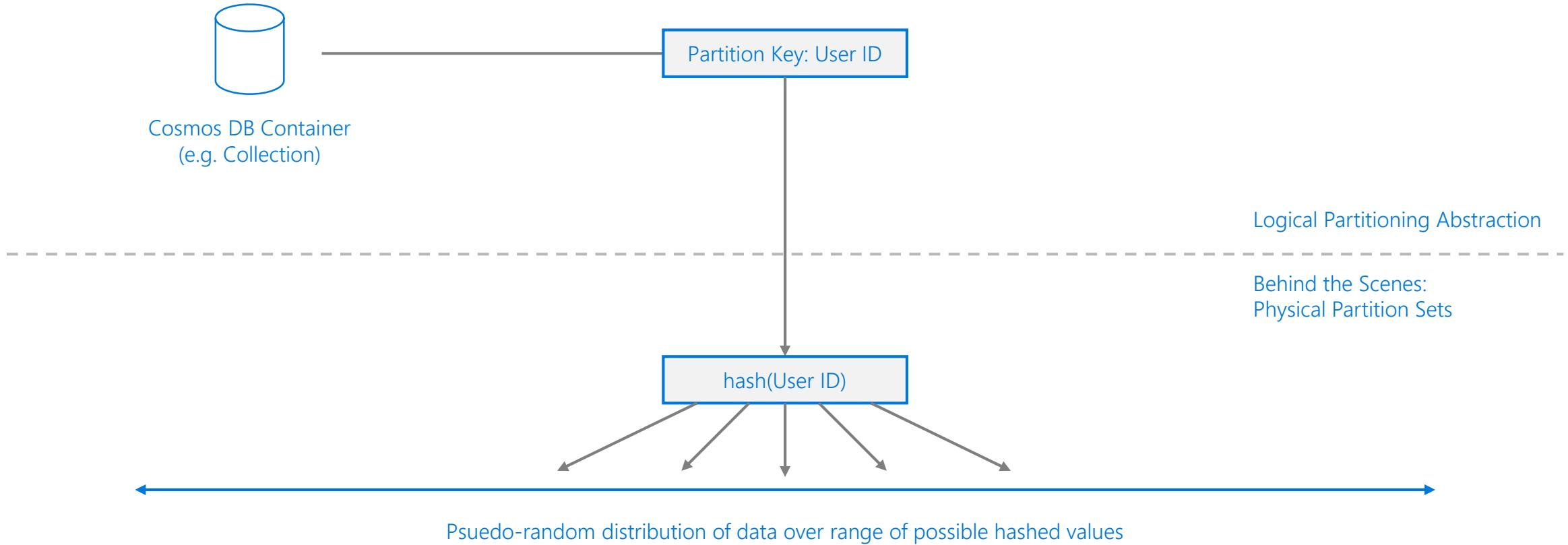
Aprovechando Azure Cosmos DB para escalar automáticamente los datos en todo el mundo

*Este módulo hará referencia a la partición en el contexto de todos los módulos de DB y APIs de Azure cosmos.*



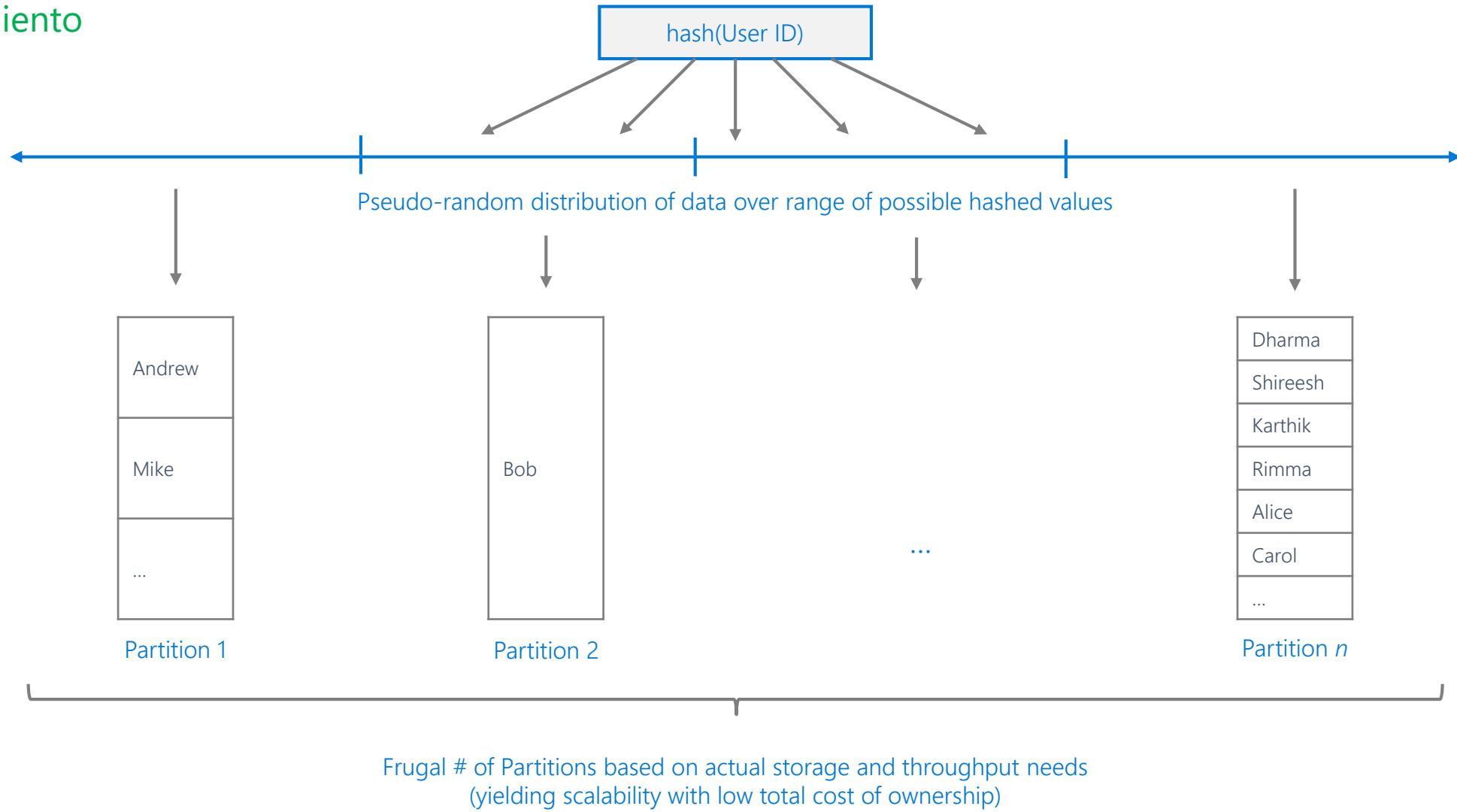
# Particiones

## Partitionamiento



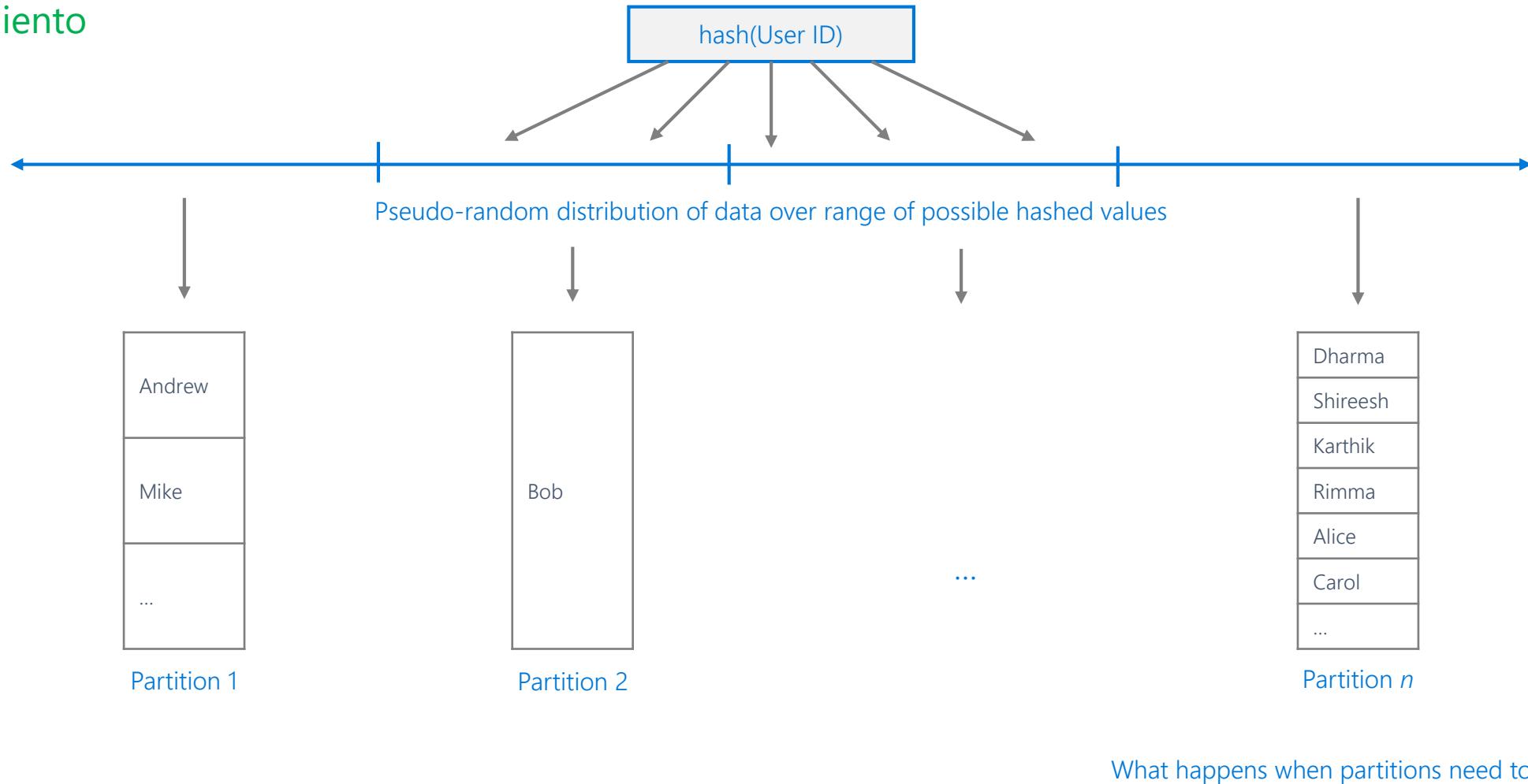
# Particiones

## Partitionamiento



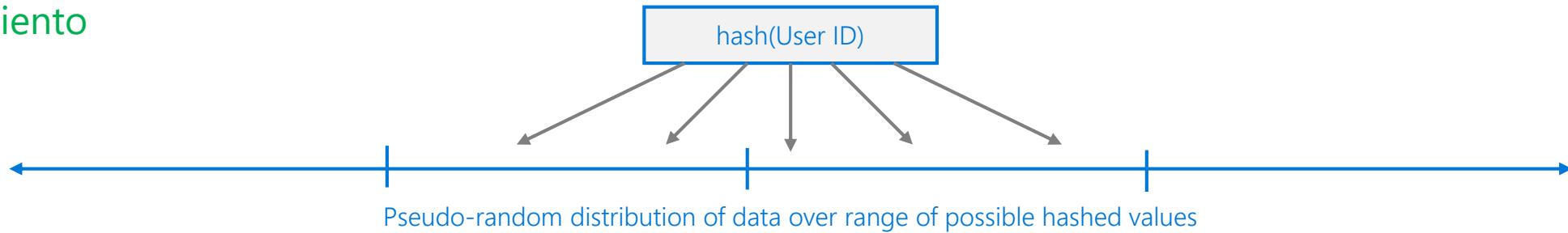
# Particiones

## Partitionamiento



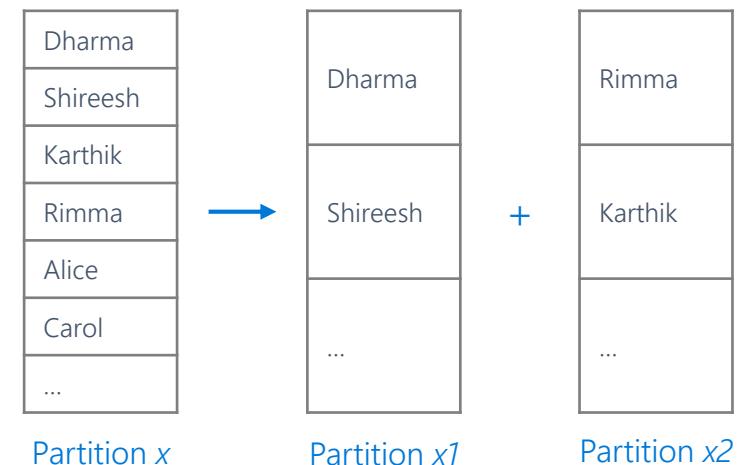
# Particiones

## Partitionamiento



Partition Ranges can be dynamically sub-divided to seamlessly grow database as the application grows while simultaneously maintaining high availability.

**Partition management is fully managed** by Azure Cosmos DB, so you don't have to write code or manage your partitions.



# Diseño de particiones

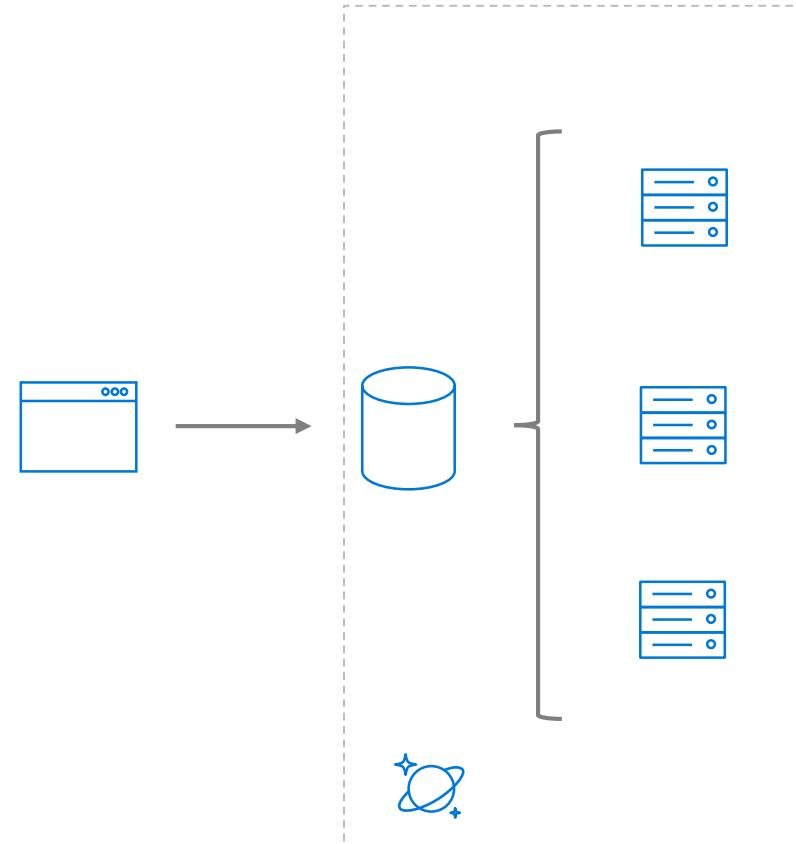
## Particionamiento

Es importante saber seleccionar de forma correcta la partición

Las claves de partición actúan como un medio para enrutar las consultas de forma eficiente y como un límite para transacciones de múltiples registros.

### Motivaciones clave

- Distribuir solicitudes
- Distribuir almacenamiento
- Enrutamiento inteligente de consultas para mejorar la eficiencia



# Diseño de particiones

## Particionamiento

### Escenario

Automotores de México es una empresa de telemática vehicular. Están planeando almacenar datos de telemetría de vehículos de millones de vehículos cada segundo en Azure Cosmos DB para el mantenimiento predictivo de energía, gestión de flotas y análisis de riesgos de controladores.

La clave de partición que seleccionamos será el ámbito para las transacciones de múltiples registros.

### ¿Cuáles son las posibles opciones de llave de partición?

- Modelo del vehículo
- Tiempo actual
- Id del dispositivo
- Llave compuesta – Id del dispositivo + Tiempo actual



# Opciones de llave de partición

## Particionamiento

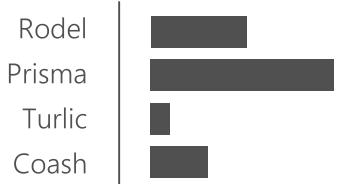
### Modelo del vehículo (p.ej. Modelo A)

La mayoría de los fabricantes de automóviles sólo tienen un par de docenas de modelos. Esto creará un número fijo de valores de llaves de partición lógica; siendo la opción menos granular.

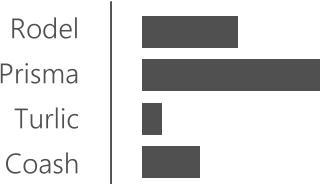
Dependiendo de la uniformidad de las ventas a través de los modelos, se introducen posibilidades para llaves de partición calientes en almacenaje y el rendimiento.



#### Distribución de almacenamiento



#### Distribución del rendimiento

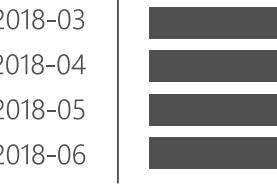


### Mes actual (p.ej. 2018-04)

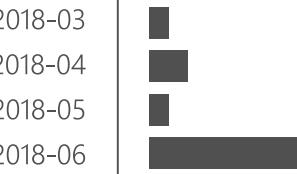
Los fabricantes de automóviles tienen transacciones que ocurren durante todo el año. Esto creará una distribución más equilibrada del almacenamiento de información entre los valores de las claves de partición.

Sin embargo, la mayoría de las transacciones comerciales se producen en datos recientes que crean la posibilidad de una clave de partición caliente para el mes actual sobre el rendimiento.

#### Distribución de almacenamiento



#### Distribución del rendimiento



# Opciones de llave de partición

## Particionamiento

### Id del dispositivo (p.ej. Dispositivo01)

Cada coche tendría un identificador de dispositivo único. Esto crea un gran número llaves de partición y tendría una cantidad significativa de granularidad.

Dependiendo de cuántas transacciones se produzcan por vehículo, es posible una llave de partición específica que alcance el límite de almacenamiento por llave.



#### Distribución de almacenamiento



#### Distribución del rendimiento



### Llave compuesta (p.ej. Id + Tiempo)

Aumenta la granularidad de los valores de las llaves de partición combinando el mes actual y un id de dispositivo. Las llaves específicas -> menos riesgo de alcanzar las limitaciones de almacenamiento de información, ya que sólo se relacionan con un solo mes de datos para un vehículo específico.

El rendimiento se distribuiría más a los valores de la llave de partición lógica para el mes actual.

#### Distribución de almacenamiento



#### Distribución del rendimiento

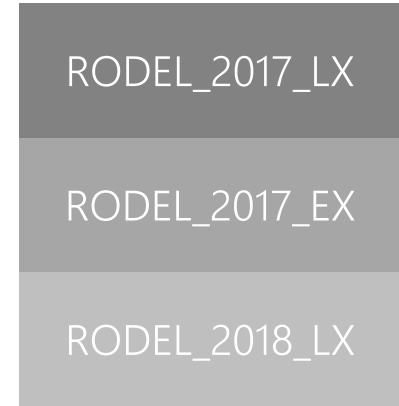


# Granularidad

## Particionamiento

### Selección del nivel adecuado de granularidad para las particiones

Las particiones deben estar basadas en las consultas más comunes y necesidades transaccionales. El objetivo es maximizar la granularidad y minimizar las consultas cruzadas entre particiones.



Más llaves de partición = Mayor escalabilidad

# Granularidad

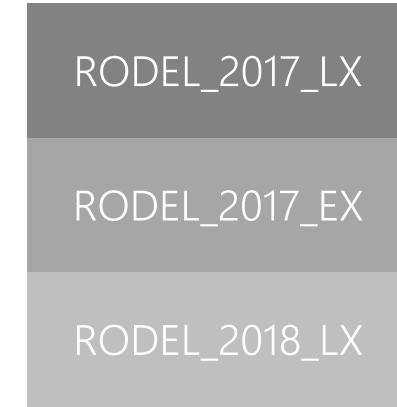
## Particionamiento

### Selección del nivel adecuado de granularidad para las particiones

Considerar los umbrales de almacenamiento y rendimiento



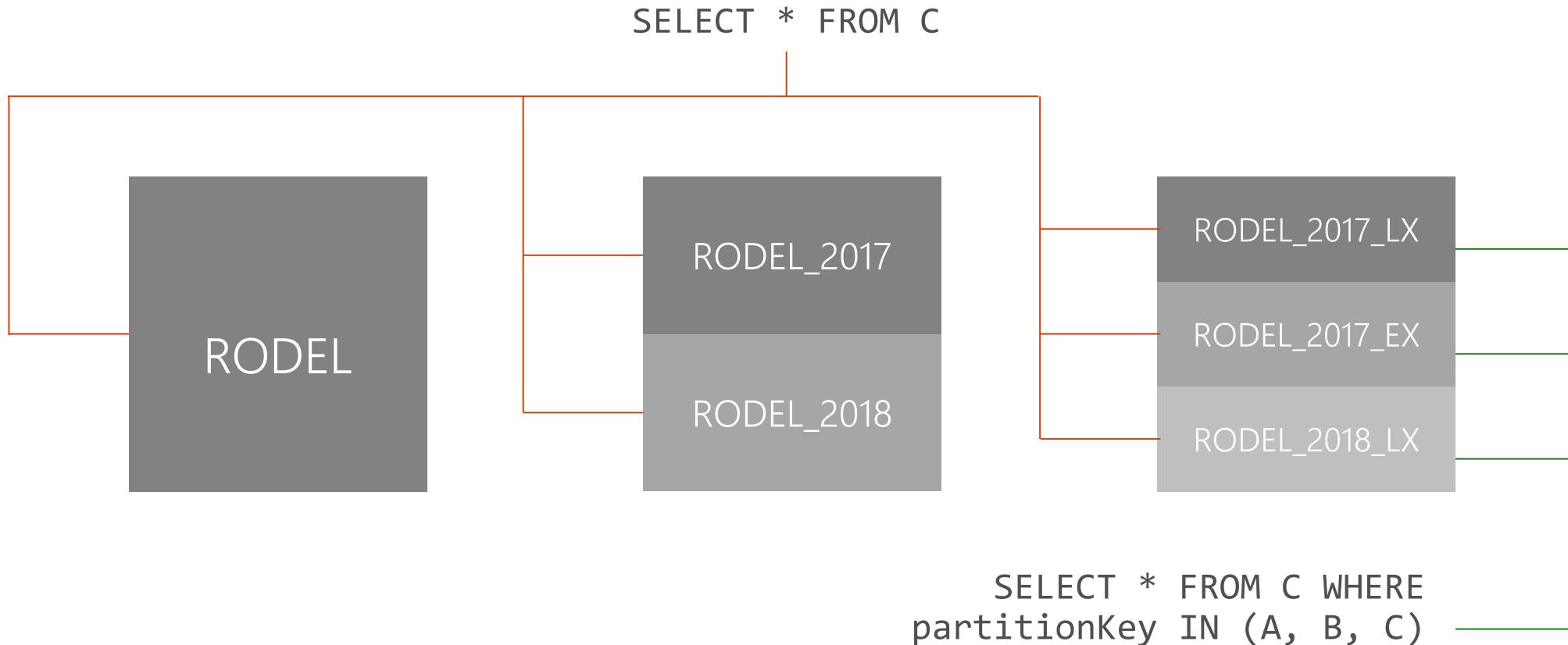
Considerar la probabilidad de consulta con particiones cruzadas



Más llaves de partición = Mayor escalabilidad

# Granularidad

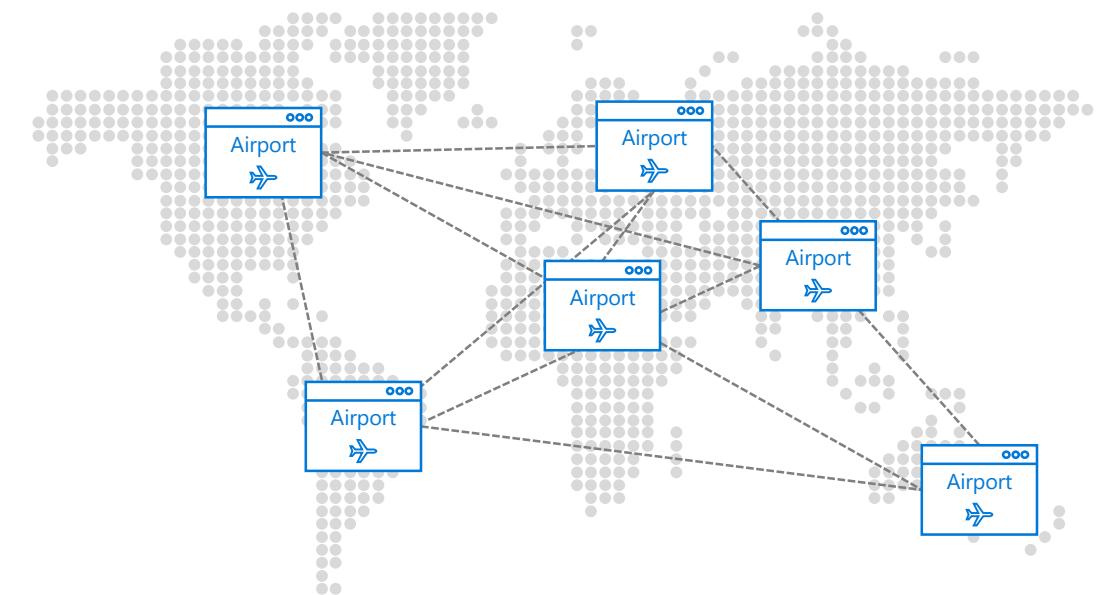
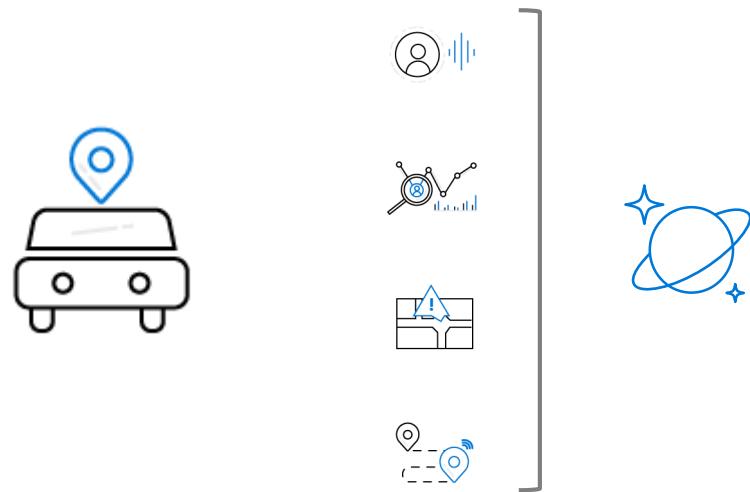
## Particionamiento



# Selección de llave de partición

## Particionamiento

Automotores de México está recolectando y almacenando datos de telemetría vehicular de millones de vehículos. El equipo ha decidido particionar basándose en una clave compuesta consistente en ID de dispositivo + hora actual cuando se produjo la interacción.

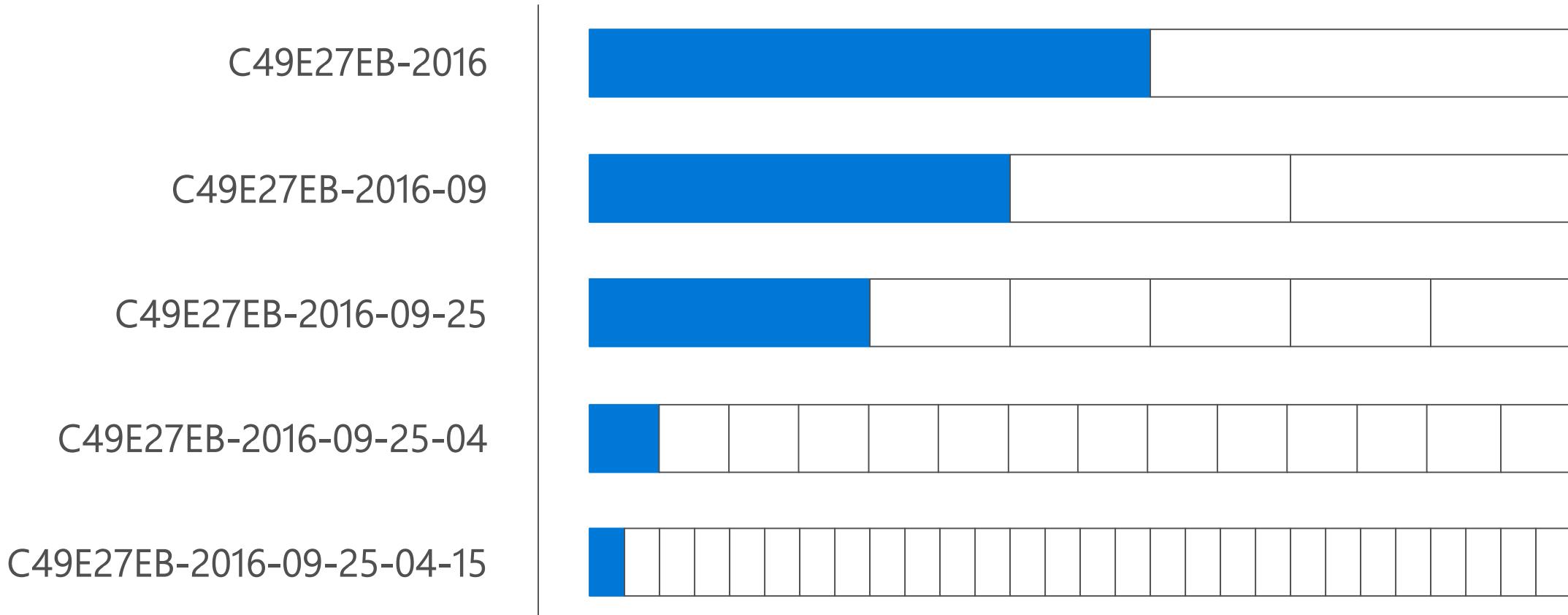


# Escenario

## Particionamiento

Interacción ocurrida en:

Septiembre 25, 2016 a las 4:15 AM UTC

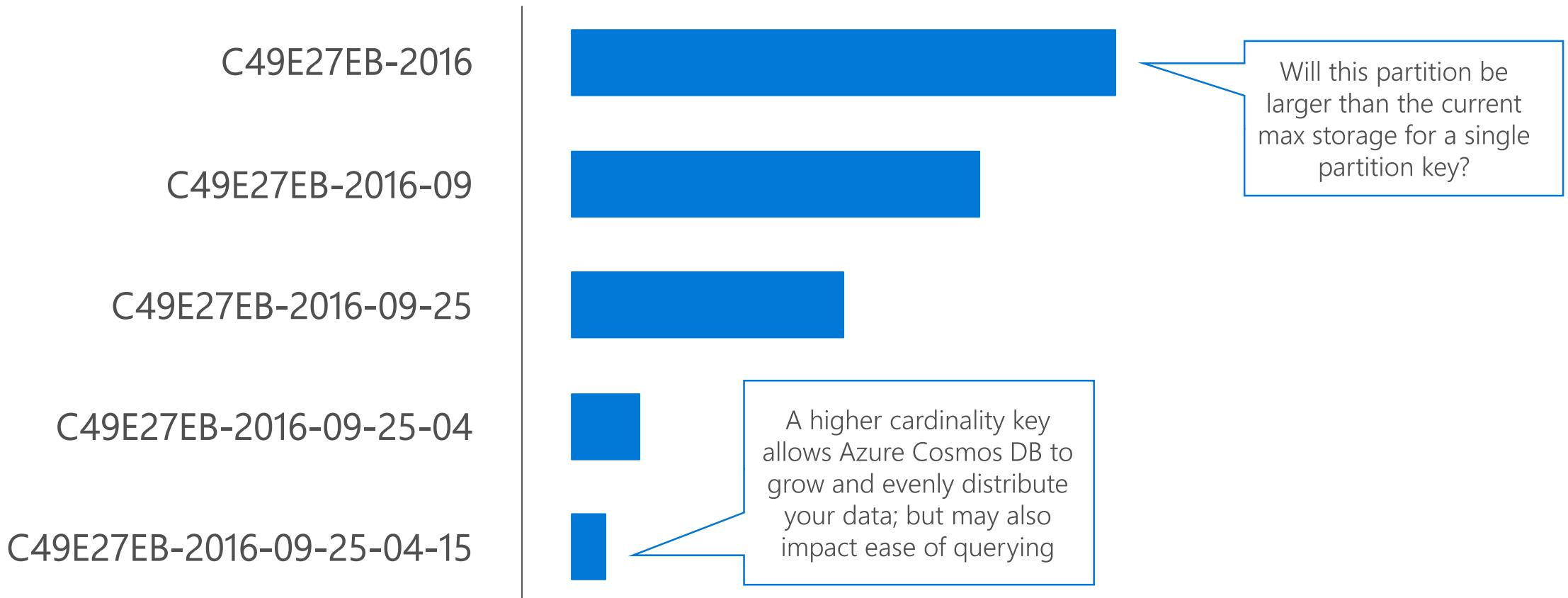


# Escenario

## Particionamiento

Interacción ocurrida en:

Septiembre 25, 2016 a las 4:15 AM UTC



# Mejores prácticas

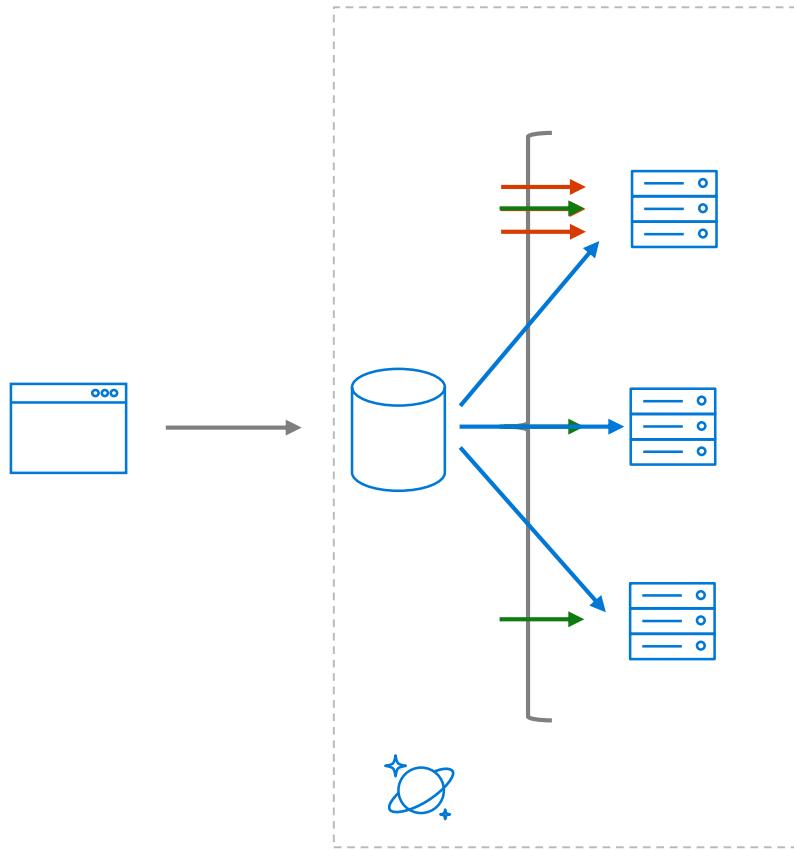
## Particionamiento

### Mejores prácticas: Metas de diseño en la selección

- Distribuir la solicitud global + volumen de almacenamiento
  - Evitar las llaves de partición "Hot"
- La llave de partición es creada para transacciones multi-registro y consultas de enruteamiento
  - Las consultas se pueden enrutar de forma inteligente a través de la llave de partición
  - La omisión de la llave de partición en las consultas requiere de *fan-out*

### Pasos para el éxito

- Considerar los requerimientos de escalación (size/throughput) a largo mediano y largo plazo
- Comprender la carga de trabajo sobre la base de datos
- Número de lecturas contra escrituras por segundo
  - Utilizar el principio de Pareto (80/20 rule) para optimizar la carga de trabajo.
  - Para lecturas – entender las principales 3 – 5 consultas (pensar en los filtros comunes)
  - Para escrituras – comprender las necesidades transaccionales

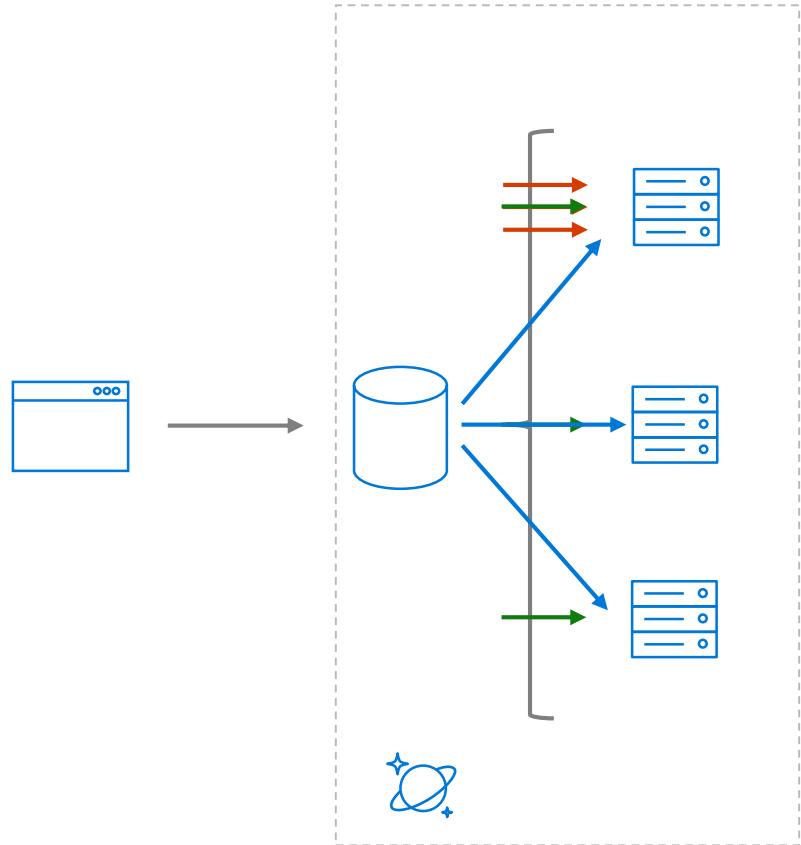


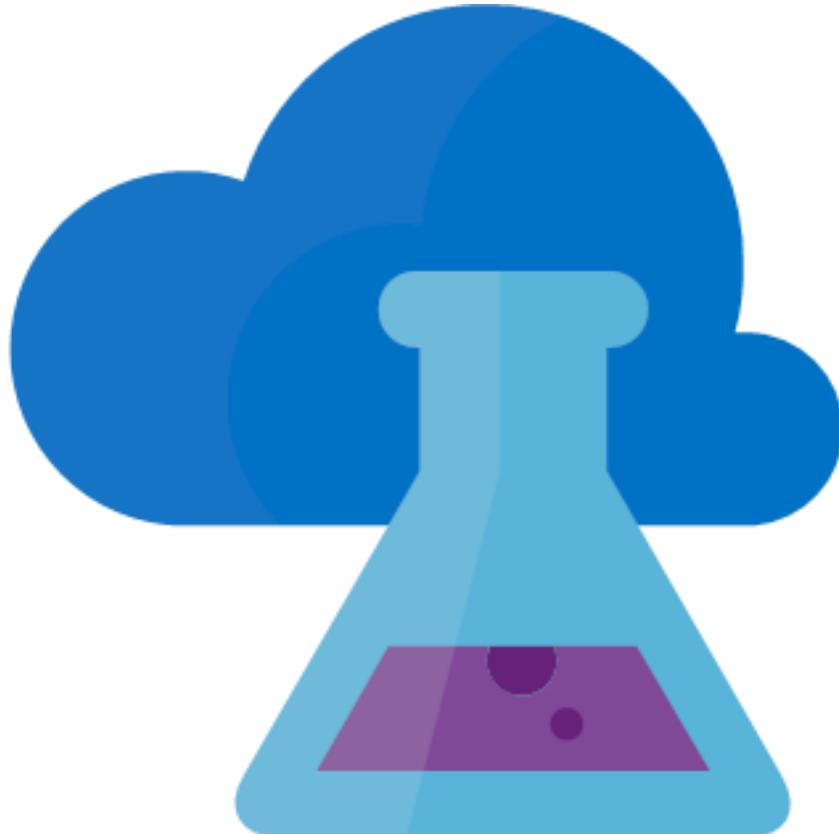
# Tips

## Particionamiento

### Tips

- Construir una prueba de concepto para fortalecer el entendimiento de las cargas de trabajo e iterar (evitando así retrasos por el análisis).
- Considerar que:
  - Las llaves de partición son lógicas
  - Más llaves → Mayor escalabilidad





# Lab 5

## ConsmosDB on Azure Consulta de partición cruzada

**¡Descarga tu material!**

<https://github.com/psmapad/ExpertAcademy>

# Fan Out

## Particionamiento

- Si existen **datos relevantes que devolver**, hacer una consulta de partición cruzada es una carga de trabajo aceptable con un rendimiento predecible.
- En un escenario real, las consultas son filtradas para incluir únicamente aquellas particiones que sean relevantes.
- Existen costos implicados al consultar particiones que no contienen datos relevantes.
- Varias consultas *fan-out* pueden saturar rápidamente las Unidades de Lectura (RU/s) para cada partición.

# Fan Out

## Particionamiento

### Concurrencia y consultas *Fan-Out*

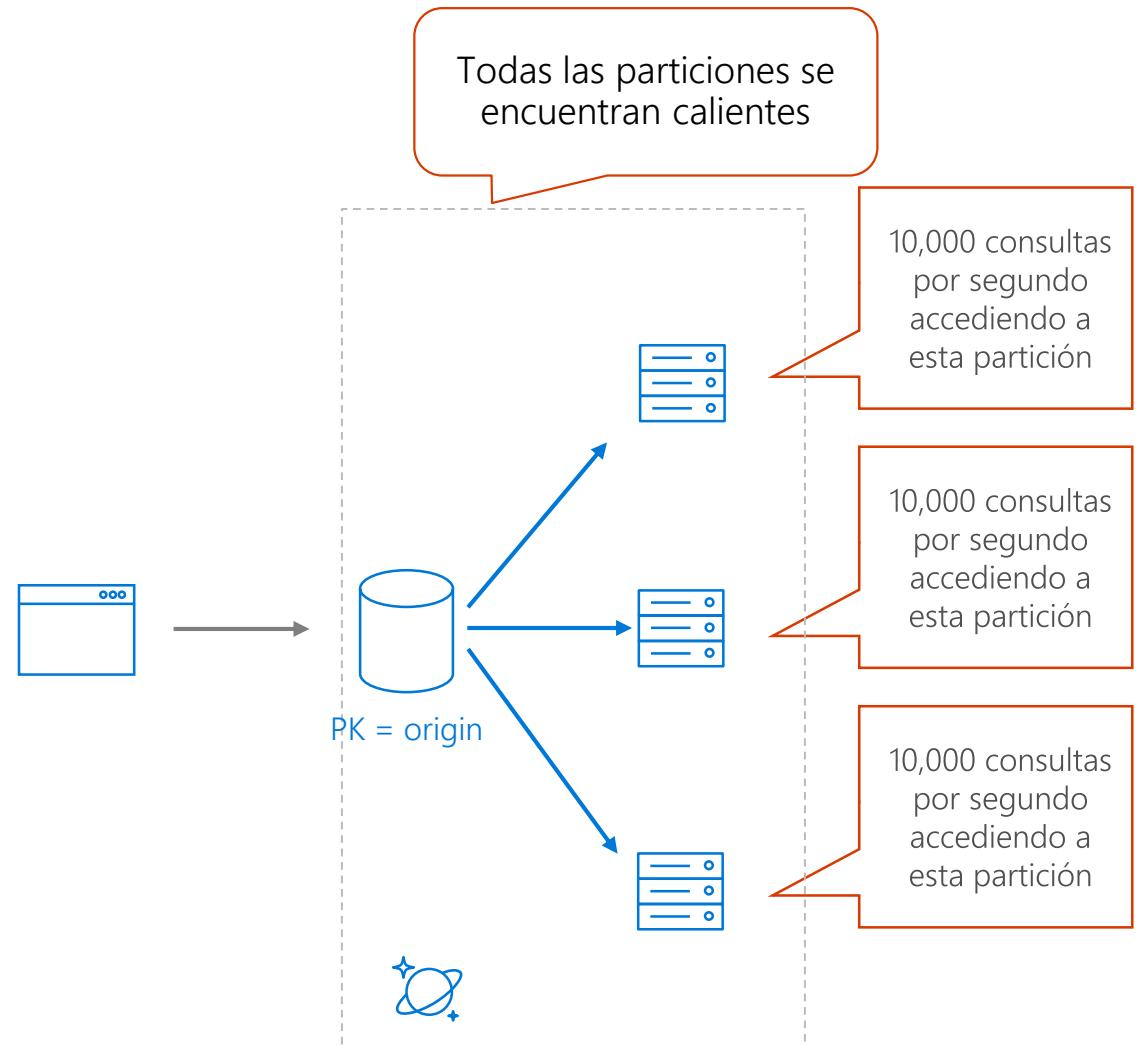
>10,000 consultas fan-out pueden generar calentamiento de todas las particiones.

**Ejemplo:** Consulta en la base de datos vehicular, particionada por modelo de vehículo, en donde la consulta es filtrada por año sin filtrar para incluir únicamente las particiones relevantes.

```
SELECT * FROM car a WHERE a.year = "2015"
```

↑  
|  
>10,000 más consultas  
por segundo  
|  
↓

```
SELECT * FROM car a WHERE a.year = "2016"
```



# Fan Out

## Particionamiento

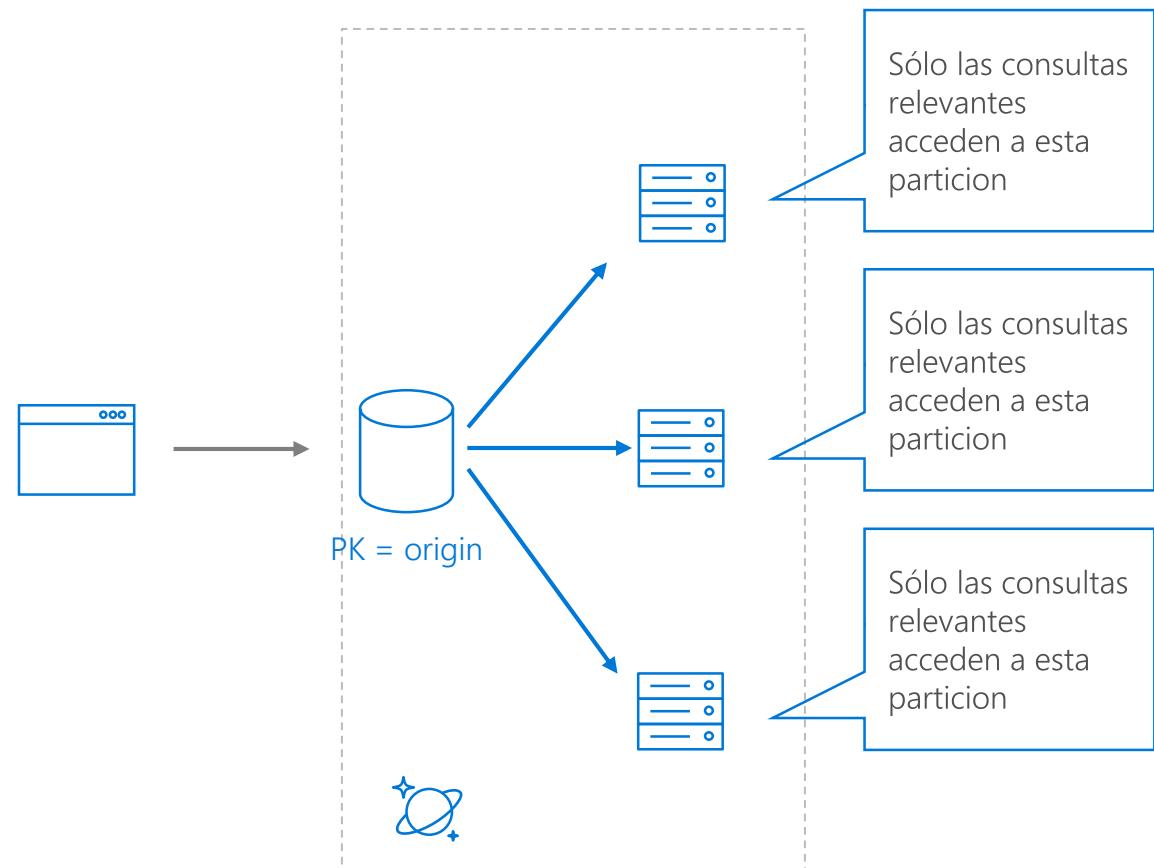
### Concurrencia y consultas *Fan-Out*

Filtrar las consultas para incluir únicamente valores de llaves de partición relevantes, reduciendo así el esfuerzo mal invertido y permitiendo concentrar las búsquedas sobre las particiones relevantes.

```
SELECT * FROM car a WHERE a.year = "2015"
```

>10,000 más consultas  
por segundo

```
SELECT * FROM car a WHERE a.year = "2016"
```

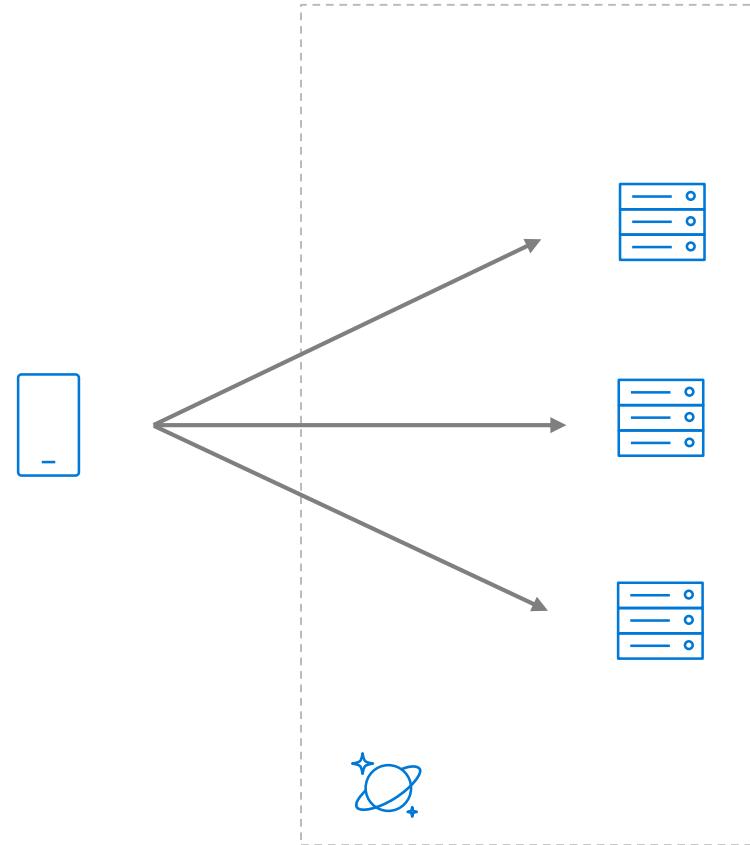


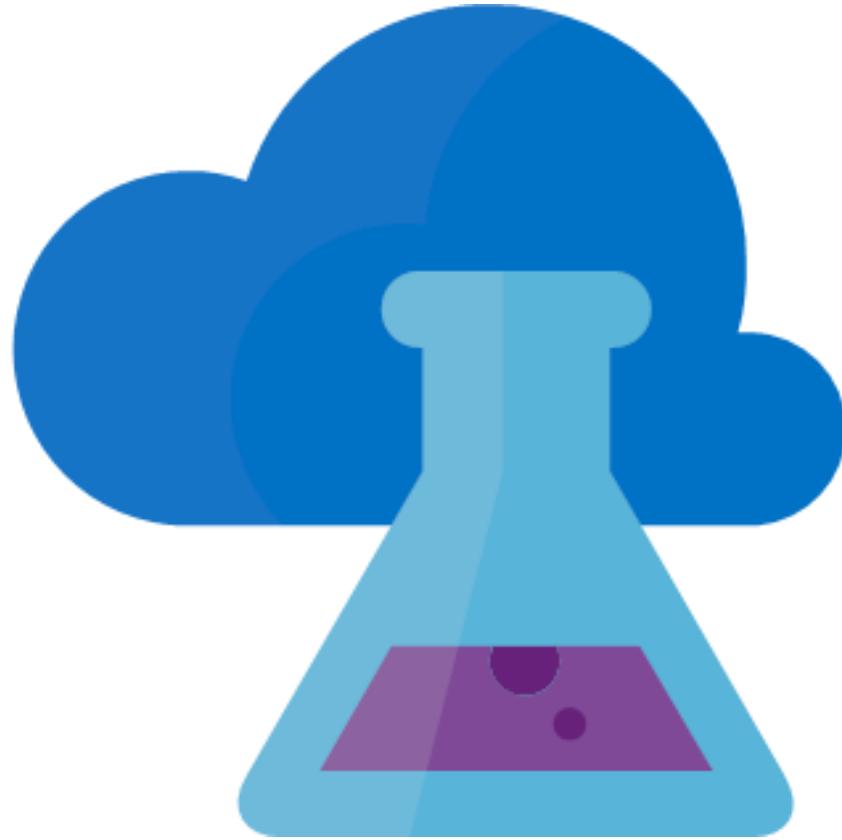
# Consultas Fan Out de lado del cliente

## Particionamiento

Es factible hilos y paralelismo del lado del cliente para atacar el *fan-out* de las consultas entre particiones.

- Aprovechar las capacidades de los dispositivos del lado del cliente.
- Realizar las consultas en paralelo a través de las unidades.
- Potenciar el procesamiento de resultados tempranos antes que se completen las solicitudes ejecutadas en paralelo.





# Lab 6

ConsmosDB on Azure  
Crear una solución multi-partición

¡Descarga tu material!

<https://github.com/psmapad/ExpertAcademy>

# Lab: Crear una solución multi-partición

Particionamiento

1. Crear un contenedor ilimitado
2. Ejecutar consultas de partición cruzadas





# CosmosDB - Consultas

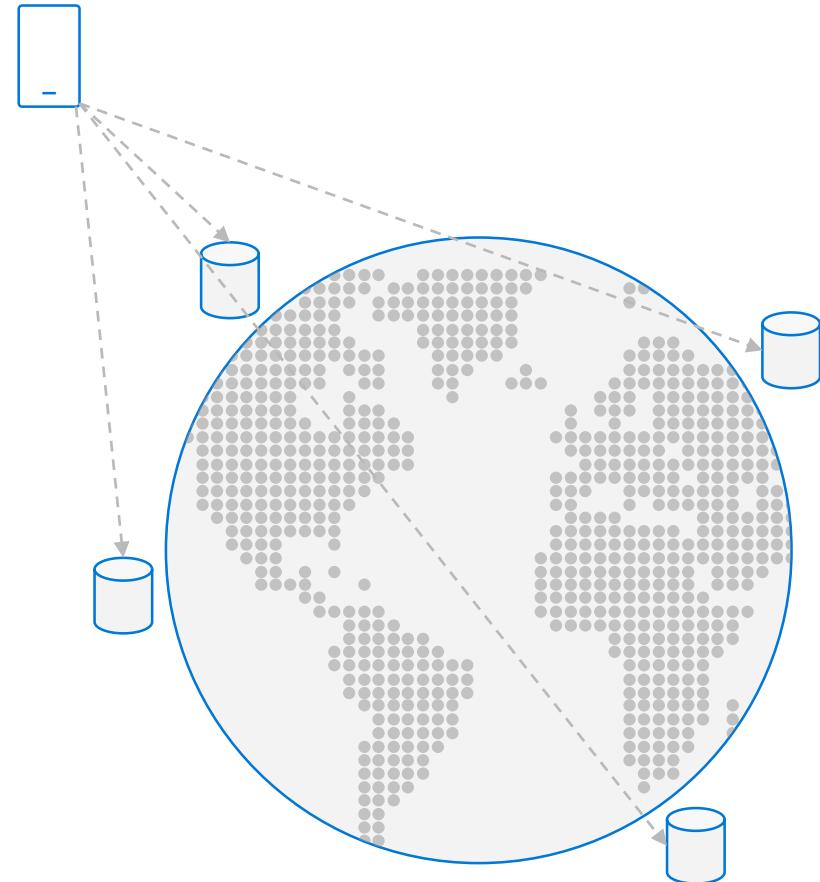


# Consultas Fan Out de lado del cliente

## Consultas

La mejora en la forma de consultar, las técnicas y parámetros para hacer más eficiente el uso de un servicio de base de datos distribuido.

*Este módulo hará referencia al proceso de consulta en el contexto de SQL API para Azure Cosmos DB*



# Mejora en las consultas

## Consultas

Hay múltiples factores que pueden impactar en el rendimiento de una consulta:

- Rendimiento provisionado
- Particionamiento y llaves de particionamiento
- SDK y opciones de consulta
- Latencia de la red
- Políticas de indexación
- Complejidad de la consulta
- Métricas de ejecución de la consulta

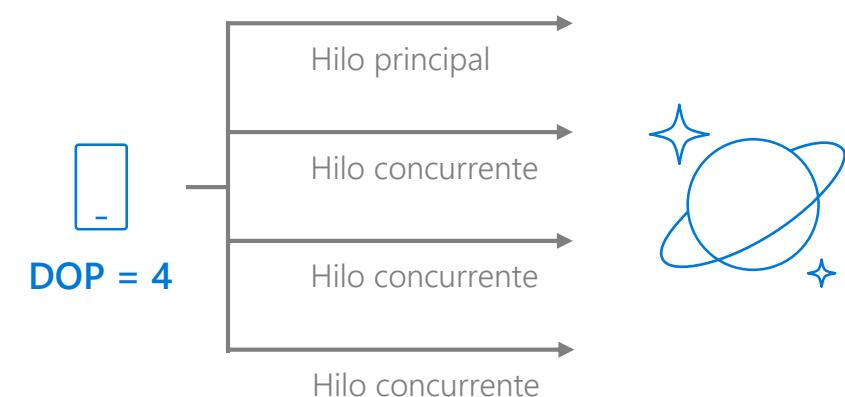
# Paralelismo en las consultas

## Consultas

Las consultas de particiones cruzadas se pueden paralelizar para usar tantos hilos como sea posible

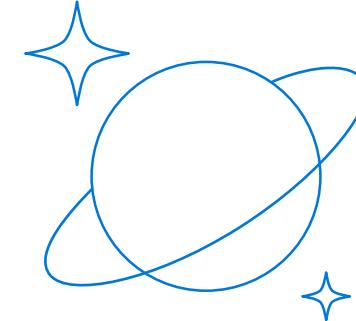
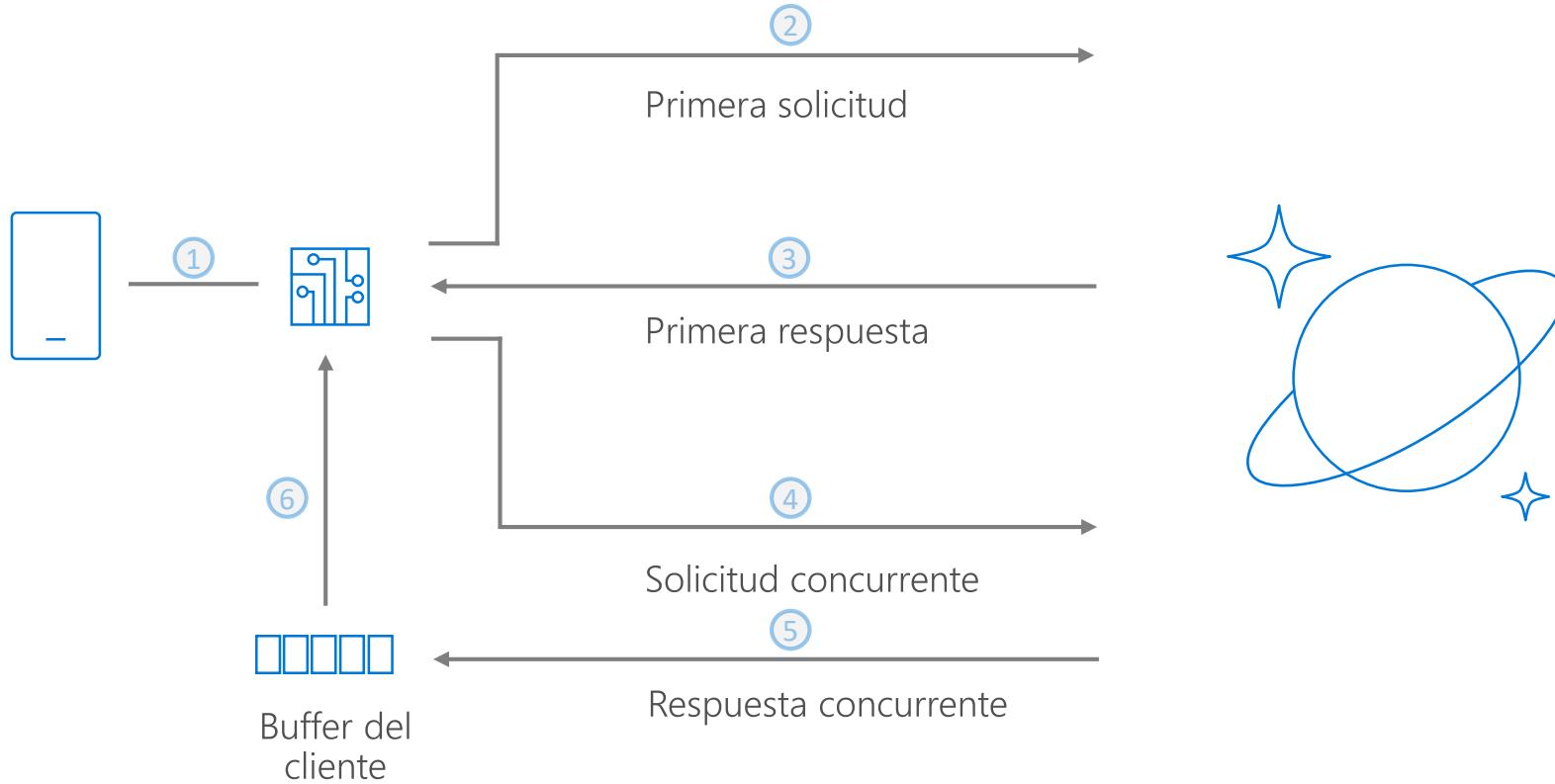
Los procesadores modernos tienen *hyper-threading* en sus núcleos tanto físicos como virtuales. Para una consulta de partición cruzada, el SDK puede utilizar hilos concurrentes para solventar la consulta a través de las particiones subyacentes.

Por defecto, el SDK utiliza un **algoritmo de arranque lento** para consultas sobre particiones cruzadas. El incremento es exponencial sobre cualquier limitación física o de conexión de red.



# Buffer de respuesta

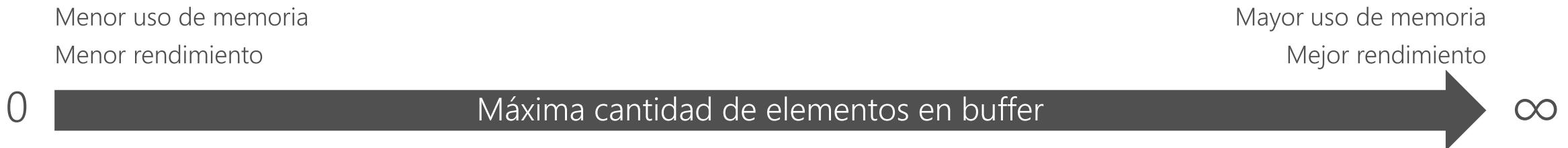
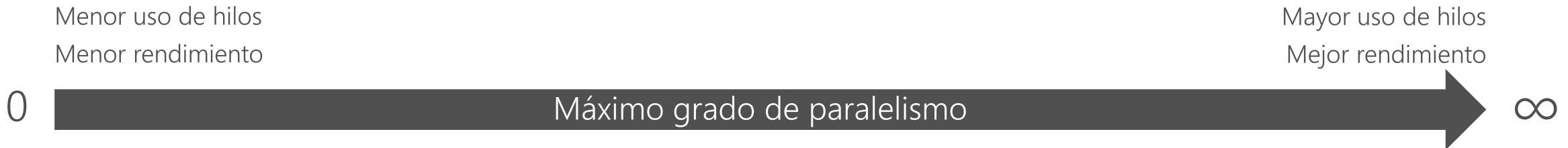
Consultas



# Opciones de consulta del SDK

## Consultas

El equilibrio en el rendimiento esta dado por estas dos propiedades



# Opciones de consulta del SDK

## Consultas

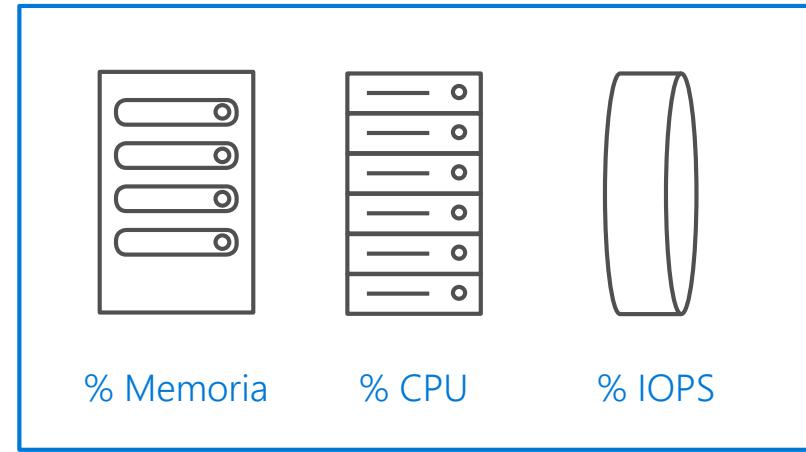
Setting	Value	Effect
<b>MaxDegreeofParallelism</b>	<b>-1</b>	El sistema decidirá de forma automática el número de elementos en el buffer
	<b>0</b>	No agregar hilos concurrentes adicionales
	<b>&gt;= 1</b>	Agregar el número específico de hilos concurrentes adicionales
<b>MaxBufferedItemCount</b>	<b>-1</b>	El sistema decidirá de forma automática el número de operaciones concurrentes que es capaz de ejecutar
	<b>0</b>	No mantiene un buffer del lado del cliente
	<b>&gt;= 1</b>	Especifica el tamaño máximo de los elementos que se encontrarán en el buffer del lado del cliente.

# Opciones de consulta del SDK

## Consultas

Las unidades de solicitud o Request Units (RUs) es una divisa basada en tarifas

- Abstracción de los recursos físicos para ejecutar las solicitudes
- Clave para que sea multi-tenant, SLAs y la eficiencia de los COGS.
- Actividades de primer plano y actividades de fondo.



# Unidades de solicitud

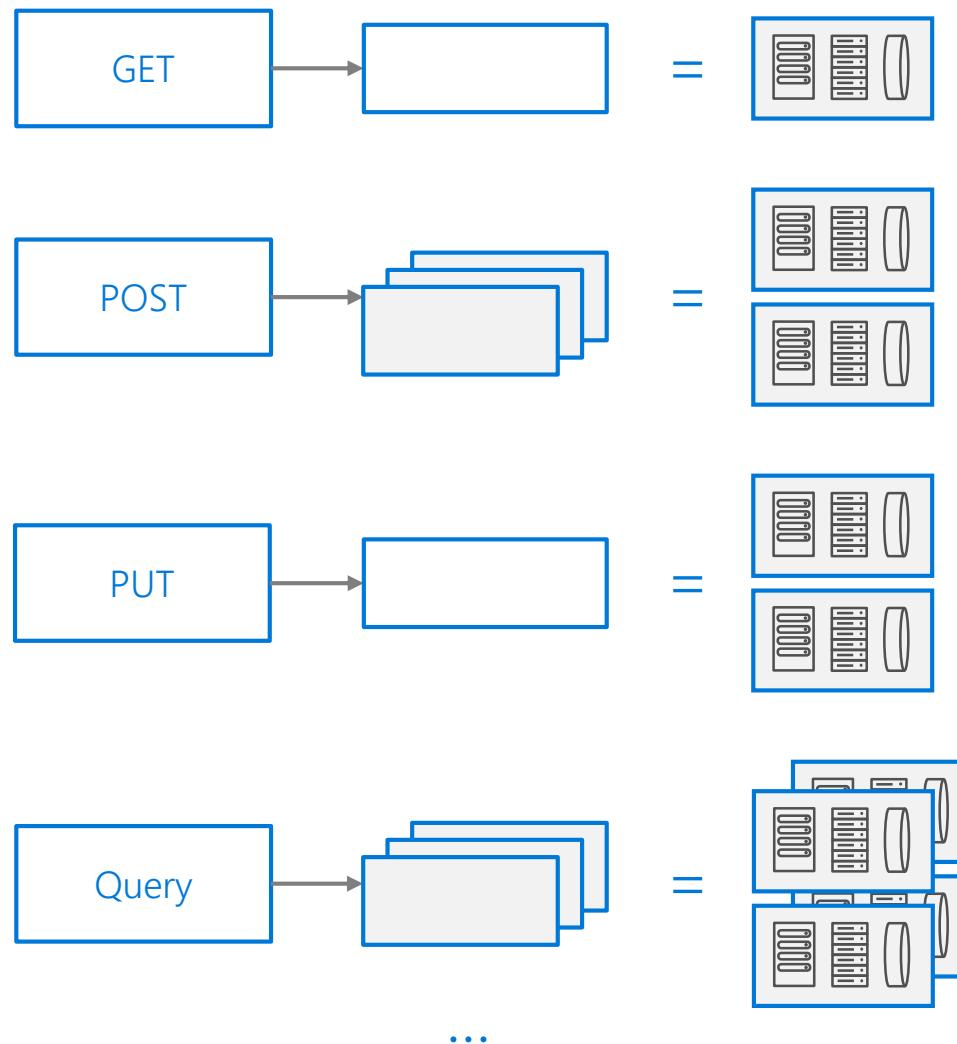
## Consultas

Normalizada a través de varios métodos de acceso

- 1 RU = 1 lectura de 1 KB de documento

Cada solicitud consume RUs fijas

- Aplica a lecturas, escrituras, consultas, y stored procedures



# Unidades de solicitud

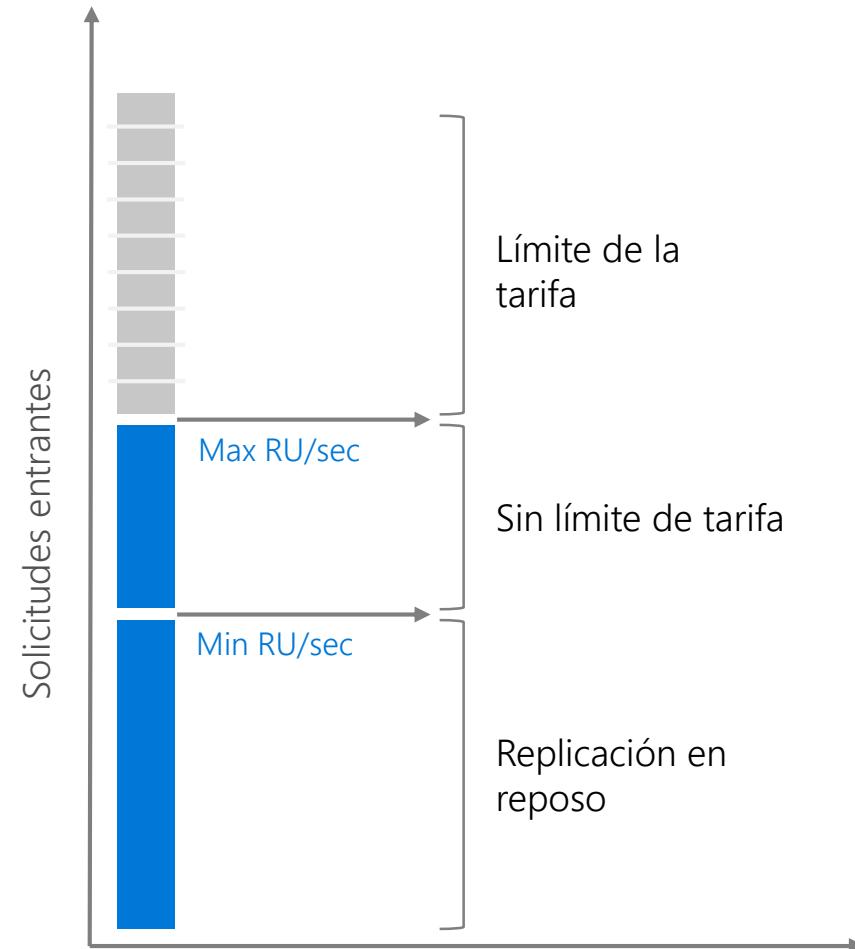
## Consultas

### Provisionada en términos de RU/seg

- Limitación de la tasa basada en la cantidad de rendimiento provisionado
- Puede ser incrementado o decrementado de manera instantánea

### Medida por hora

- Los procesos de fondo como expiración del TTL, transformaciones de índice programadas cuando están en reposo





# CosmosDB – Sintaxis de SQL

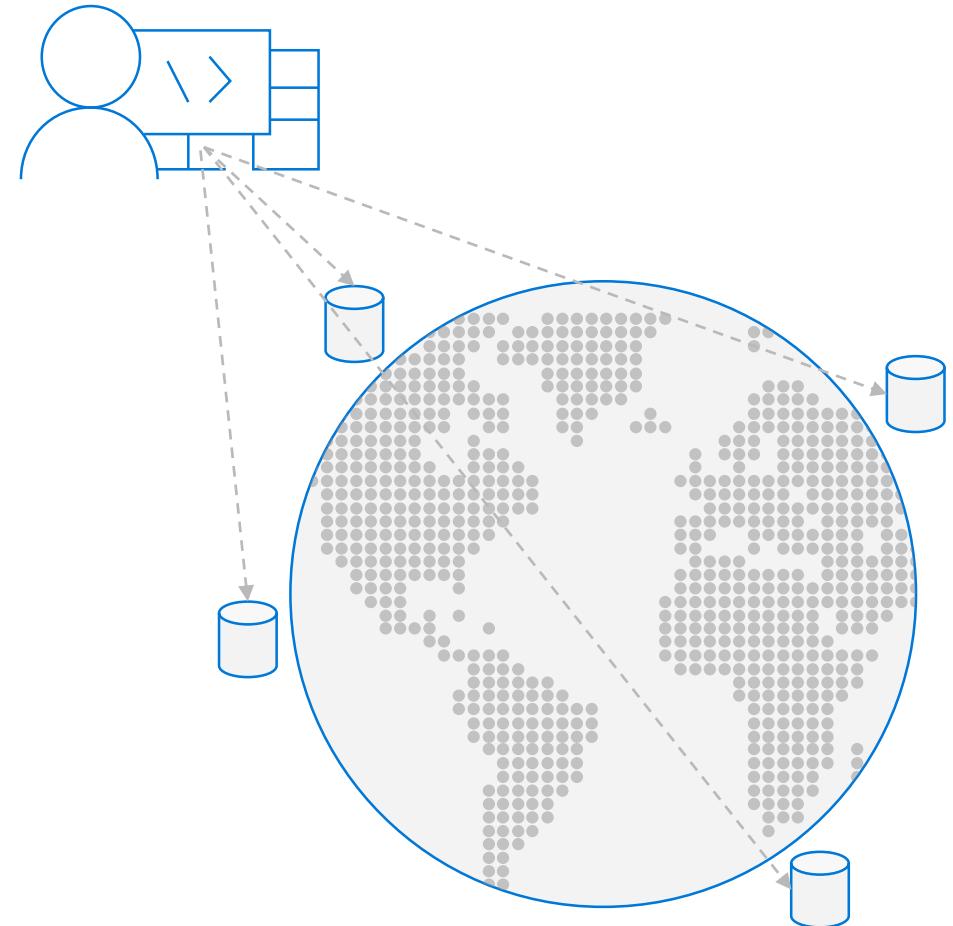


# Sintaxis de SQL

## Sintaxis de SQL

Se puede utilizar la sintaxis de SQL para realizar consultas sobre datos semi-estructurados de tipo JSON.

*Este modulo hará referencia a las consultas en el contexto del API de SQL para Cosmos DB*



# Sintaxis de las consultas

## Sintaxis de SQL

### Sintaxis básica de la consulta

**SELECT & FROM** siguen siendo los componentes básicos de las consultas.

```
SELECT
```

```
    tickets.id,  
    tickets.pricePaid  
FROM tickets
```

```
SELECT
```

```
    t.id,  
    t.pricePaid  
FROM tickets t
```

# Sintaxis de las consultas

## Sintaxis de SQL

### Filtrado

**WHERE** soporta expresiones escalares complejas incluyendo aritmética, comparación y operadores lógicos.

SELECT

```
    tickets.id,  
    tickets.pricePaid
```

FROM tickets

WHERE

```
    tickets.pricePaid > 500.00 AND  
    tickets.pricePaid <= 1000.00
```

# Sintaxis de las consultas -

## Sintaxis de SQL

### Proyección JSON

**WHERE** soporta expresiones escalares complejas incluyendo aritmética, comparación y operadores lógicos.

SELECT

```
    tickets.id,  
    tickets.pricePaid  
FROM tickets  
WHERE  
    tickets.pricePaid > 500.00 AND  
    tickets.pricePaid <= 1000.00
```

# Sintaxis de las consultas – Proyección

## Sintaxis de SQL

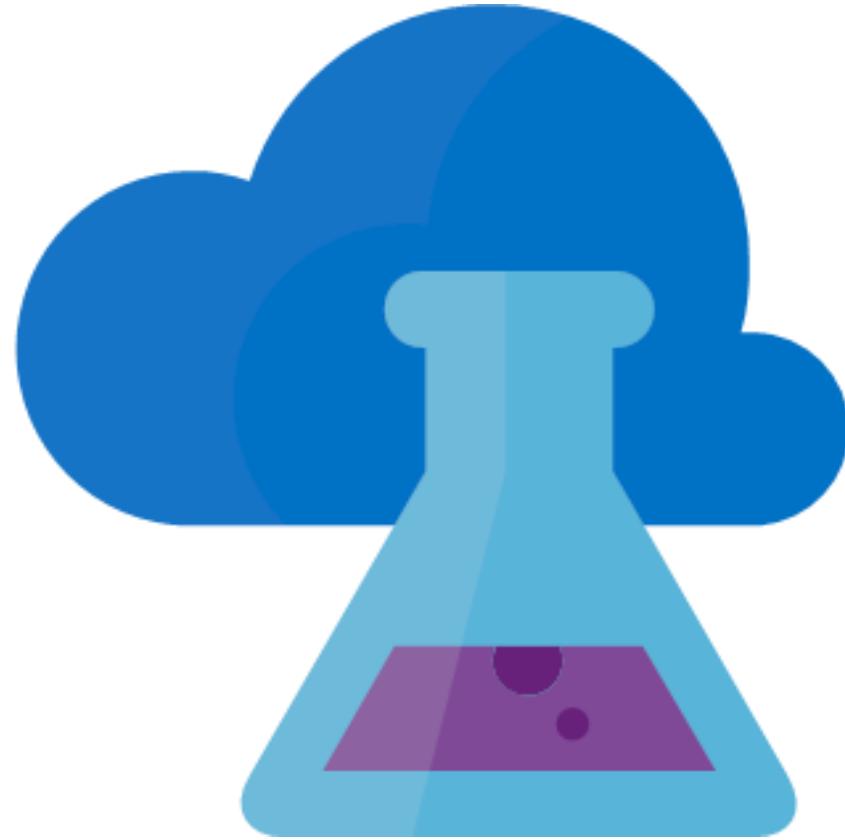
### SELECT VALUE

La palabra clave **VALUE** puede acoplar la colección de resultados si es necesario para una carga de trabajo de aplicación específica.

```
SELECT VALUE {
    "id": tickets.id,
    "flightNumber": tickets.assignedFlight.flightNumber,
    "purchase": {
        "cost": tickets.pricePaid
    },
    "stops": [
        tickets.assignedFlight.origin,
        tickets.assignedFlight.destination
    ]
}
FROM tickets
```



```
[ {
    "id": "6ebe1165836a",
    "purchase": {
        "cost": 575.5
    },
    "stops": [
        "SEA",
        "JFK"
    ]
}]
```



# Lab 7

ConsmosDB on Azure  
Consultas a Cosmos DB  
empleando SQL

**¡Descarga tu material!**  
<https://github.com/psmapad/ExpertAcademy>

# Lab: Crear una solución multi-partición

## Particionamiento

1. Ejecutar consultas
2. Ejecución de consultas dentro de documentos
3. Proyectos de los resultados de consultas





GROW THE BUSINESS TOGETHER

# CosmosDB – Programación

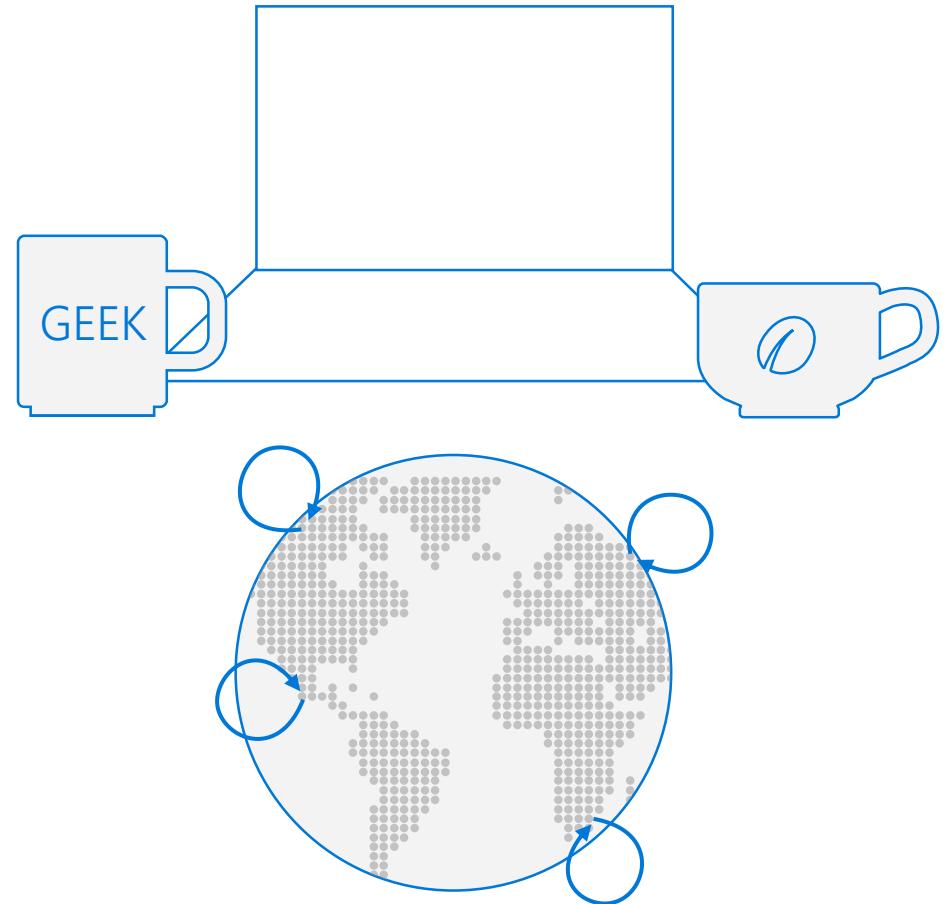


# Programación

## Programación

Ejecución de la lógica de programación del lado de servidor de forma nativa con JavaScript para realizar transacciones multi registro.

*Este módulo hace referencia a la programación en el contexto del API de SQL.*

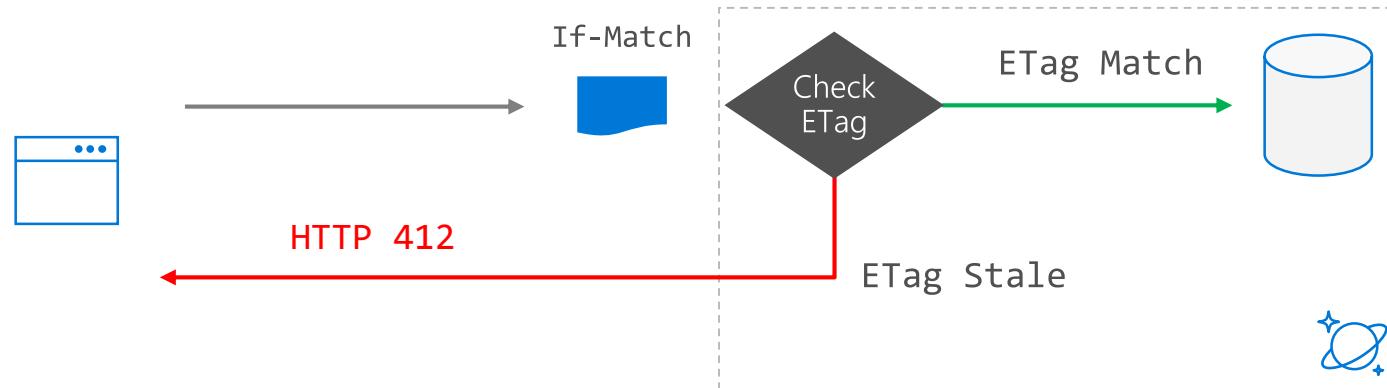


# Control de la concurrencia con ETags

## Programación

### Concurrencia optimista

- El API de SQL soporta control de la concurrencia optimista (OCC) a través de tags de entidades en HTTP o Etags.
- Cada recurso del API de SQL tiene un sistema de propiedades Etag, el valor de Etag es generado en el servidor cada vez que un documento se actualiza.
- Si el valor del Etag se mantiene constante, eso significa que ningún otro proceso ha actualizado el documento. Si el valor del Etag muta de forma inesperada, entonces otro proceso concurrente ha realizado la actualización del documento.
- Los ETags pueden ser usados con el encabezado de la solicitud de If-Match de HTTP para permitir al servidor decidir si un recurso debe ser actualizado:



# Control de la concurrencia con ETags

## Programación

```
try
{
    var ac = new AccessCondition { Condition = readDoc.ETag, Type = AccessConditionType.IfMatch };

    updatedDoc = await client.ReplaceDocumentAsync(readDoc, new RequestOptions { AccessCondition = ac });

}
catch (DocumentClientException dce)
{
    if (dce.StatusCode == HttpStatusCode.PreconditionFailed)
    {
        Console.WriteLine("Another concurrent process has updated the record");
    }
}
```

# Stored procedures

## Programación

```
function createSampleDocument(documentToCreate) {  
    var context = getContext();  
    var collection = context.getCollection();  
    var accepted = collection.createDocument(  
        collection.getSelfLink(),  
        documentToCreate,  
        function (error, documentCreated) {  
            context.getResponse().setBody(documentCreated.id)  
        }  
    );  
    if (!accepted) return;  
}
```

# Transacciones multi documento

## Programación

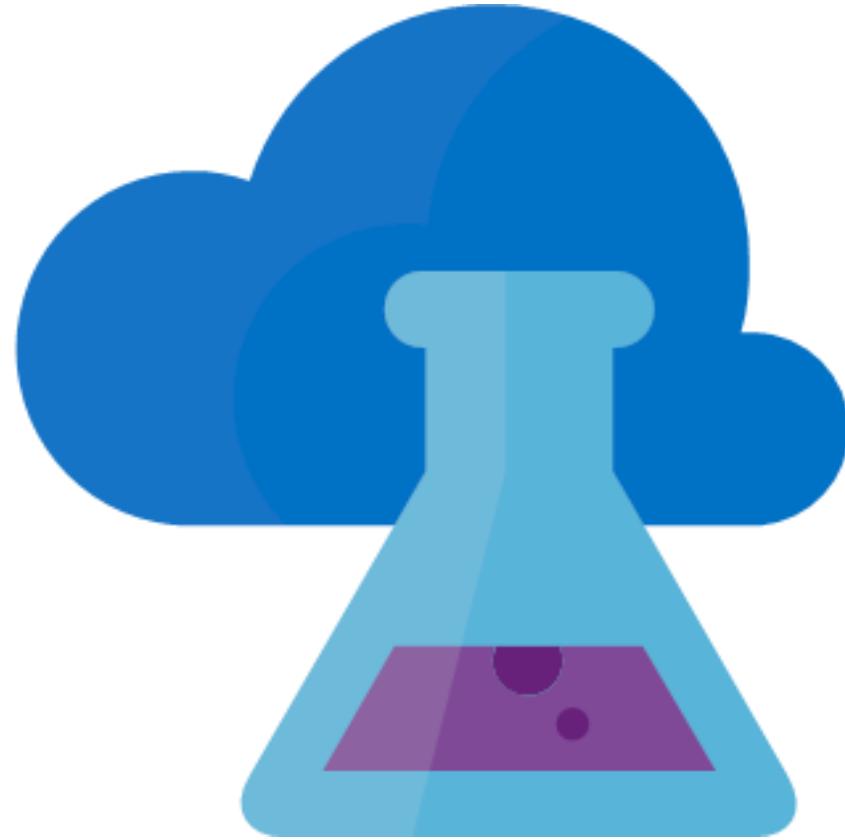
### Transacciones de base de datos

En una base de datos típicamente una transacción puede ser definida como una secuencia de operaciones realizadas como una unidad de trabajo única. Cada transacción provee de garantías ACID.

En Cosmos DB, JavaScript esta hospedado en el mismo espacio de memoria que la base de datos. Por lo tanto, las solicitudes realizadas dentro de los stored procedures y triggers se ejecutan en el mismo ámbito de una sesión de base datos.

Los stored procedures utilizan snapshots aislados para garantizar que todas las lecturas dentro de la transacción tengan un snapshot consistente de los datos.





# Lab 8

## ConsmosDB on Azure Creación de Stored Procedures

¡Descarga tu material!

<https://github.com/psmapad/ExpertAcademy>

# Funciones definidas por el usuario

## Programación

### UDF

- Las funciones definidas por el usuario (UDFs) se utilizan para ampliar la gramática del lenguaje de consulta del API SQL de Cosmos DB y para implementar la lógica empresarial personalizada. Las UDFs sólo se pueden llamar desde el interior de la consulta.
- No tienen acceso al contexto del objeto y están destinadas a utilizarse como código de una sola operación.

# Funciones definidas – Definición

## Programación

```
var taxUdf = {  
    id: "tax",  
    serverScript: function tax(income) {  
        if (income == undefined)  
            throw 'no input';  
        if (income < 1000)  
            return income * 0.1;  
        else if (income < 10000)  
            return income * 0.2;  
        else  
            return income * 0.4;  
    }  
}
```

# Funciones definidas – Uso en consultas

Programación

SELECT

\*

FROM

TaxPayers t

WHERE

udf.tax(t.income) > 20000



# CosmosDB – Casos de Uso



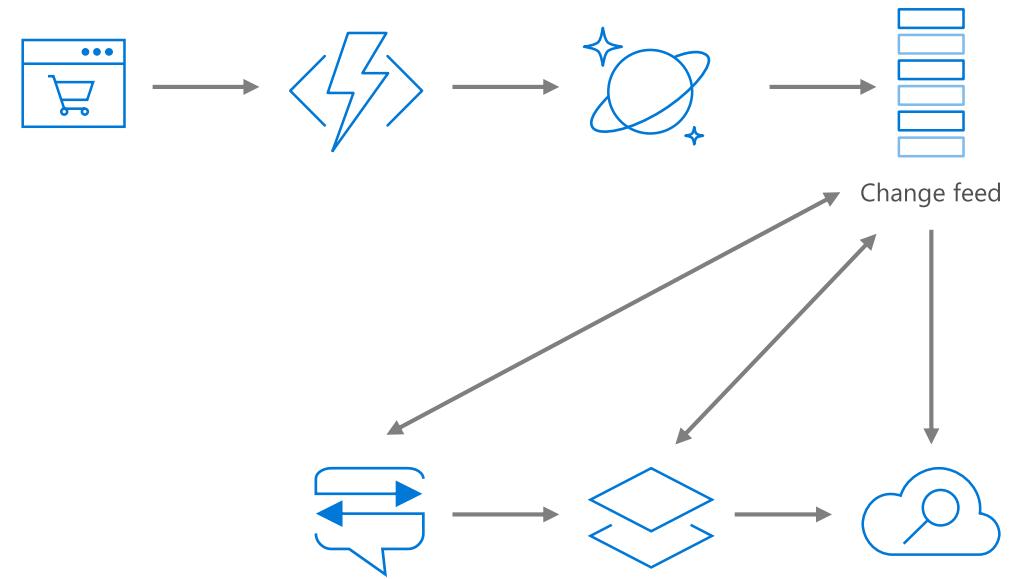
# Aplicaciones modernas reactivas

## Casos de uso

Las aplicaciones de lote, juego, Retail y registro operacional necesitan rastrear y responder a una tremenda cantidad de datos que son ingeridos, modificados o eliminados de una base de datos globalmente escalada.

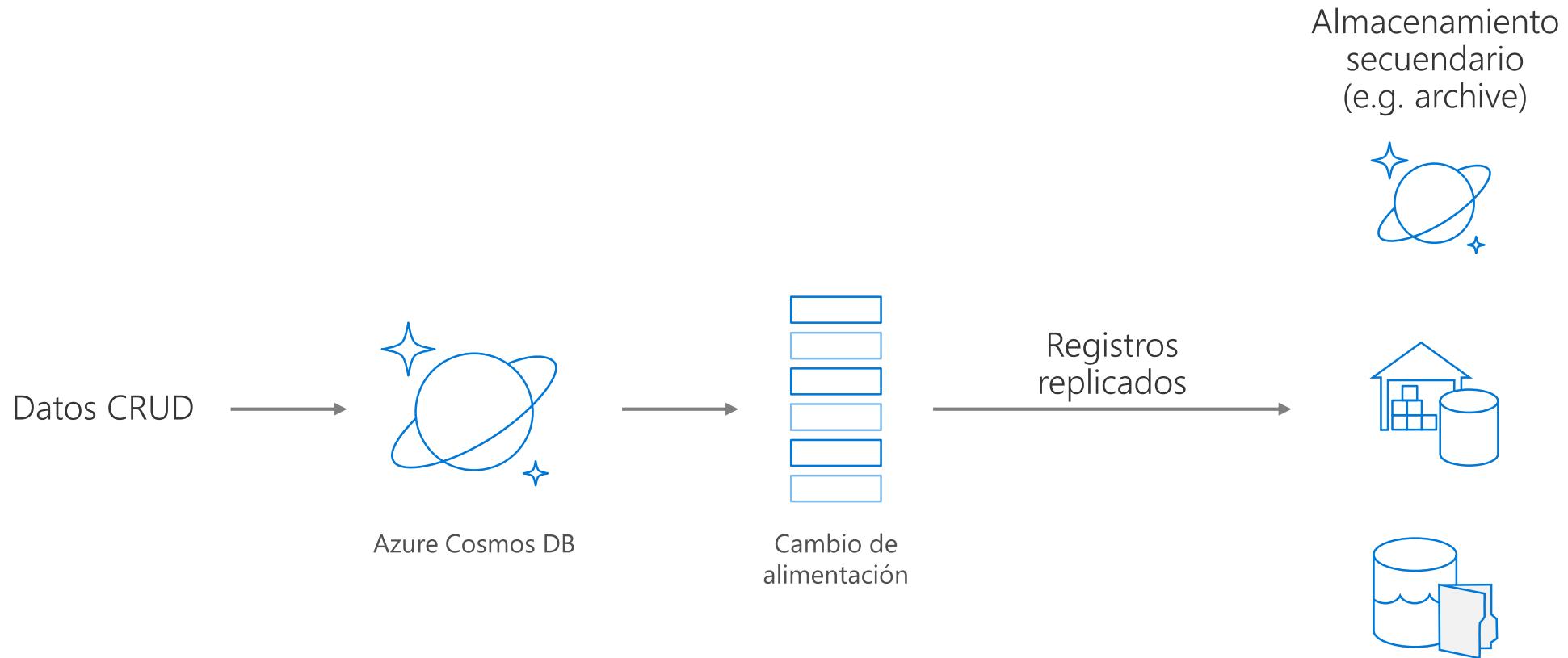
### Escenarios comunes

- Notificación de triggers para elementos nuevos
- Realizar análisis en tiempo real en datos transmitidos
- Sincronización de los datos con un caché, un motor de búsqueda o un almacén de datos.



# Replicación de datos

## Casos de uso



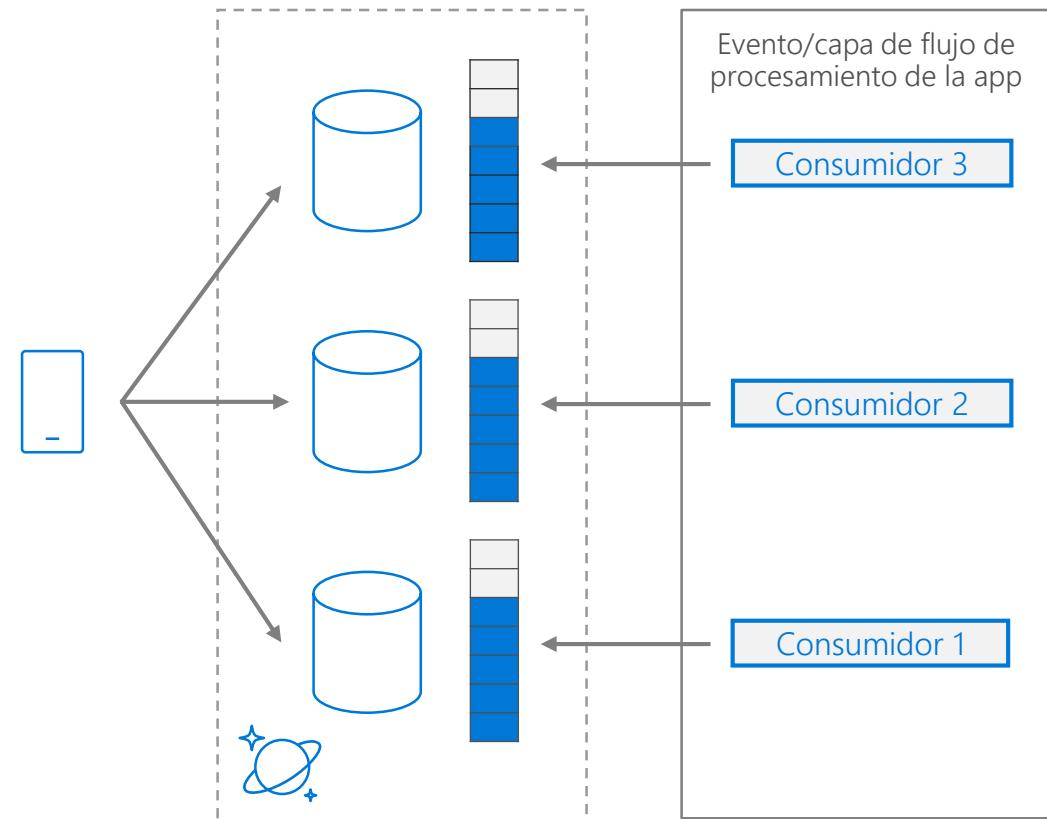
# Cambio de la alimentación con particiones

## Casos de uso

### Paralelización de consumidores

Cambio de alimentación escucha sobre cualquier cambio en la colección Cosmos DB. Después, se emite la lista ordenada de documentos que se cambiaron en el orden en que se modificaron.

Los cambios se persisten, se pueden procesar de forma asíncrona e incremental y la salida se puede distribuir a través de uno o varios consumidores para el procesamiento en paralelo. El cambio de alimentación está disponible para cada intervalo de llaves de partición dentro de la colección de documentos y, por lo tanto, se puede distribuir en uno o varios consumidores para el procesamiento en paralelo.



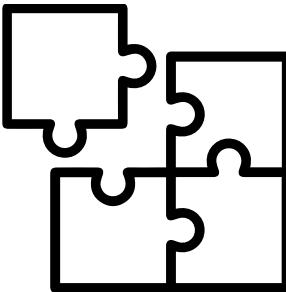
# Beneficios Azure en OSS

Open Source en Azure



## Ahorro:

Solo pagas lo que usas y utilizas todas las bondades que proporciona Azure.



## Compatible:

Se cuentan con los sabores mas utilizados de Linux y tecnologías reconocidas en el mundo Open Source Software.



## Disminuye recursos:

Automatiza y optimiza tus tecnologías y soluciones  
**Open Source Software en Azure.**

# Conclusiones

## Open Source en Azure

- **Azure** está listo para recibir sus servicios Open Source de manera nativa con todas las bondades y características que esta nube representa.
- De la mano con bSide y nuestro equipo de especialistas de Open Source generaremos la **estrategia ideal** para su organización en temas de Open Source en Azure.
- Le acompañamos en este nuevo desafío para potenciar la **transformación digital** de su compañía.

 **powerful**

# Nuestros servicios de OSS

## Open Source en Azure

El mundo de **Open Source** Software es tan extenso, tanto en **Azure** como en el mundo real, que entender y/o **implementar** nuevas **tecnologías** en su organización en Azure puede ser un gran **reto**. En **bSide** nos unimos a su causa y lo **acompañamos** paso a paso en el camino para que sus **iniciativas** de Open Source Software sean **exitosas**

¡Cuento con especialistas altamente calificados en Open Source Software!





# PREGUNTAS

¿Tienes alguna duda?



Para **bSide** ¡Su opinión es de vital importancia! Agradecemos su participación en la siguiente encuesta, cuyo objetivo consiste en la recopilación de información que aportará de forma valiosa a nuestro proceso de mejora continua.

**Ingresar a encuesta ➔**

En caso de cualquier duda contactar al correo:  
[oss@bside.com.mx](mailto:oss@bside.com.mx)

<https://bit.ly/2SwSMTw>



all you can b | [bside.com.mx](http://bside.com.mx) | [f](#) | [Linkedin](#).



## NUESTRA RAZÓN DE SER

"Sumarnos a nuestros clientes innovando con lo último en  
tecnología Microsoft para potenciar confiablemente su negocio."

---

Gracias

**Contacto:**  
[oss@bside.com.mx](mailto:oss@bside.com.mx)