

El diseño pensado para cargas de trabajo intensivas en datos

Cómo las plataformas SaaS y API escalan de forma fiable en el cloud



Escalar cargas de trabajo intensivas en datos: un enfoque práctico

Este caso de uso ofrece una visión práctica y técnica de la escalabilidad de cargas de trabajo intensivas en datos, e incluye:

1

Desafíos comunes que surgen a medida que los sistemas crecen

2

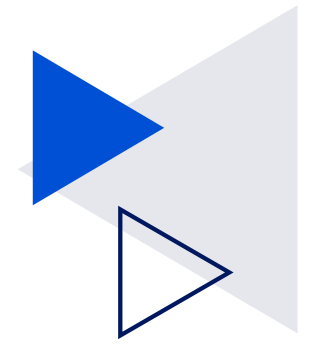
Una checklist de diseño técnico para apoyar la toma de decisiones de infraestructura

3

Ejemplos de despliegue que muestran cómo las empresas mantienen el rendimiento, la eficiencia y el control

Al explorar estas ideas, los ingenieros y arquitectos pueden comprender mejor los patrones y estrategias que permiten a las plataformas escalar de manera fiable, sin perder el control sobre la complejidad y los costes.





1. Los desafíos técnicos de la escalabilidad de los sistemas intensivos en datos

A medida que aumentan los volúmenes de datos, el rendimiento y la concurrencia, los equipos suelen encontrar los siguientes desafíos:



Restricciones de rendimiento en el almacenamiento a medida que los datasets escalan de gigabytes a terabytes o más



Dificultad para predecir costes relacionados con el crecimiento del almacenamiento, el uso de I/O y la salida de datos



Gestión de una latencia baja y consistente durante picos de tráfico en las capas de ingesta, consulta y API



Aumento de la sobrecarga operativa por la gestión de bases de datos y pipelines distribuidos



Imposibilidad de escalado independiente entre almacenamiento, bases de datos y computación, lo cual reduce la eficiencia



Exigencias de cumplimiento y residencia de datos que añaden complejidad arquitectónica



Cuellos de botella de red provocados por tráfico este-oeste intensivo en datos entre servicios

Abordar estos desafíos de manera efectiva requiere una infraestructura que proporcione visibilidad y control sobre cómo se almacenan, mueven, procesan y escalan los datos. Esto facilitará un mejor rendimiento y costes predecibles a medida que las plataformas crezcan.

2. Checklist de diseño técnico

Antes de adoptar una plataforma cloud o una arquitectura concreta, es importante evaluar el comportamiento real de la carga de trabajo en producción. En este sentido, el crecimiento de los datasets, el I/O continuo y los altos niveles de tráfico pueden plantear demandas específicas en almacenamiento, redes y cálculo distribuido.

Esta checklist está diseñada para ayudarte a identificar claramente las características técnicas que debes considerar al elegir un proveedor cloud o una arquitectura para que se ajuste mejor a las necesidades de tu empresa.



Almacenamiento

- ▶ Modelo de almacenamiento requerido: objeto, bloque, archivo o híbrido
- ▶ Requisitos de IOPS en picos de carga
- ▶ Requisitos de NVMe para cargas de trabajo de baja latencia
- ▶ Principal impulsor del rendimiento: rendimiento frente a lectura/escritura aleatoria
- ▶ Trayectoria de crecimiento de datos (GB / TB / PB)
- ▶ Modelo de replicación y durabilidad: zona única, multi-AZ o multiregión



Computación y procesamiento

- ▶ Perfil de carga de trabajo: orientada a CPU, memoria o asistida por GPU
- ▶ Modo de procesamiento de datos: batch, streaming o en tiempo real
- ▶ Requisitos de procesamiento distribuido: clústeres de Spark, Dask, ClickHouse u otros
- ▶ Patrón de uso: carga estable, picos de demanda o basado en eventos
- ▶ Capacidad de contenerización y autoescalado



Redes

- ▶ Nivel de sensibilidad a la latencia
- ▶ Requisitos de ancho de banda este-oeste entre servicios o nodos
- ▶ Requisitos de arquitectura de multizona o multiregión
- ▶ Requisitos de redes privadas
- ▶ Volúmenes de salida esperados



Escalabilidad

- ▶ Modelo de escalado: horizontal (stateless), vertical (throughput) o híbrido
- ▶ Picos de tráfico o cargas esperadas: diarias, estacionales o impulsadas por eventos
- ▶ Requisitos de escalado automático, fragmentación o expansión de clústeres
- ▶ Preparación para la contenerización



Gestión de costes y de recursos

- ▶ Previsibilidad de uso: estable vs. altamente variable
- ▶ Costes principales: almacenamiento, cálculo, salida de datos o escalado de bases de datos
- ▶ Preferencia del modelo de gasto: mensual predecible vs. pago por uso
- ▶ Impacto del coste del crecimiento y replicación del almacenamiento de objetos



Cumplimiento normativo y seguridad

- ▶ Requisitos de residencia y soberanía de datos
- ▶ Restricciones de localización de datos regulatorios
- ▶ Certificaciones requeridas
- ▶ Requisitos de cifrado, aislamiento de red e integración de IAM
- ▶ SLA, expectativas de clientes y gobernanza en torno a la privacidad y el cumplimiento normativo



Automatización y despliegue

- ▶ Herramientas de infraestructura como código
- ▶ Integración CI/CD
- ▶ Requisitos de velocidad de aprovisionamiento del entorno
- ▶ Acceso a la API y hooks de automatización
- ▶ Monitorización, registro, alertas y observabilidad



Requisitos de bases de datos

- ▶ Modelo de base de datos: documento (MongoDB) vs. relacional (PostgreSQL)
- ▶ Necesidades de lectura/escritura concurrentes
- ▶ Patrones de consulta esperados: transaccional vs. analítico
- ▶ Alta disponibilidad, monitorización, copias de seguridad y failover



Resultados de ingeniería

Una arquitectura bien definida, basada en la comprensión del comportamiento de las cargas de trabajo presentes y futuras en producción, permite a los equipos de ingeniería lograr mejoras significativas en fiabilidad, rendimiento y

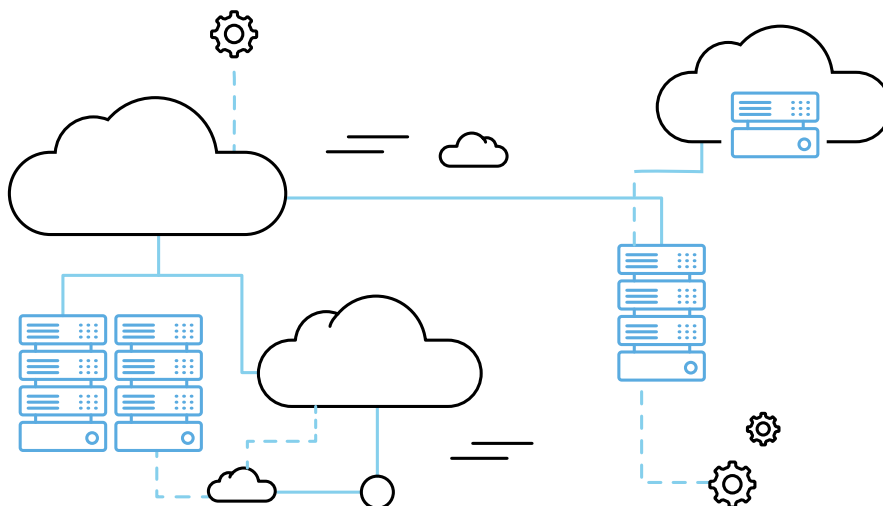
eficiencia. Además, hacen posible mantener niveles de servicio consistentes, escalar los recursos de forma independiente y gestionar tanto los servicios como los costes.

CAPACIDAD

- ✓ Rendimiento de ingesta estable
- ✓ Tiempos de consulta y procesamiento más rápidos
- ✓ Escalabilidad y almacenamiento independientes
- ✓ Reducción de costes operativos
- ✓ Costes cloud impredecibles
- ✓ Control de datos y cumplimiento normativo
- ✓ Arquitecturas abiertas

GANANCIAS OPERATIVAS

- ✓ Menos eventos perdidos, SLA consistente
- ✓ Mejor latencia de API, capacidad de respuesta de análisis
- ✓ Escalado independiente el almacenamiento, las bases de datos y los recursos de cálculo
- ✓ Menos tiempo dedicado a parches, alta disponibilidad y copias de seguridad
- ✓ Sin sorpresas en los costes de salida ni de IOPS
- ✓ Ubicación y residencia de los datos claramente definidas
- ✓ Sin lock-in, mayor portabilidad



Diseño para cargas de trabajo de IA

La IA está acelerando el crecimiento de cargas de trabajo intensivas en datos que combinan alta demanda de computación, rendimiento de almacenamiento y estrictos requisitos de eficiencia. A medida que estas cargas escalan, los equipos necesitan infraestructuras que ofrezcan un rendimiento predecible, modelos de despliegue flexibles y consistencia operativa, sin añadir complejidad innecesaria.

Un enfoque práctico es estandarizar un stack tecnológico con amplia compatibilidad en entornos híbridos.

Plataformas como OVHcloud son un buen ejemplo, ya que ofrecen una combinación de servidores on-premise listos para desplegar (On-Prem Cloud Platform), servidores bare metal e instancias cloud tradicionales, a menudo basadas en arquitecturas de alto rendimiento como AMD EPYC.

Esto permite a los equipos de ingeniería adaptar la infraestructura a los requisitos de cada carga de trabajo, manteniendo el uso de herramientas conocidas, un rendimiento predecible y eficiencia operativa a gran escala.

3. Ejemplos de despliegue: cómo las empresas escalan SaaS y API intensivas en datos

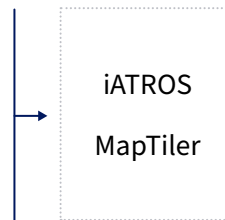
Las plataformas SaaS y API intensivas en datos enfrentan diferentes desafíos de infraestructura dependiendo del tráfico, el tamaño del conjunto de datos y el tipo de carga de trabajo.

Los siguientes ejemplos muestran cómo las empresas en crecimiento mantienen un rendimiento predecible, escalan de manera eficiente y controlan sus costes.

Si tu desafío es... Podría interesarte...

Cumplimiento normativo y alta disponibilidad multi-AZ

Procesamiento y almacenamiento de datos a gran escala



iATROS: una plataforma a gran escala de salud digital segura y conforme a la normativa

La empresa iATROS ofrece una plataforma de salud digital que recopila, analiza y gestiona datos sensibles de pacientes para cientos de miles de usuarios. Para cumplir con los requisitos regulatorios y mejorar el rendimiento para usuarios en

distintas regiones, el equipo rediseñó su stack apoyándose en una infraestructura cloud multiclúster con alta disponibilidad y controles avanzados.

EL RETO CLAVE

- ▶ Migrar desde una nube que no cumplía completamente con los requisitos europeos de protección de datos
- ▶ Garantizar alta disponibilidad y baja latencia para usuarios distribuidos geográficamente
- ▶ Cumplir estrictos estándares de cumplimiento y gobernanza (RGPD, ISO, sector salud)
- ▶ Mantener una infraestructura resiliente a través de distintos dominios de fallo

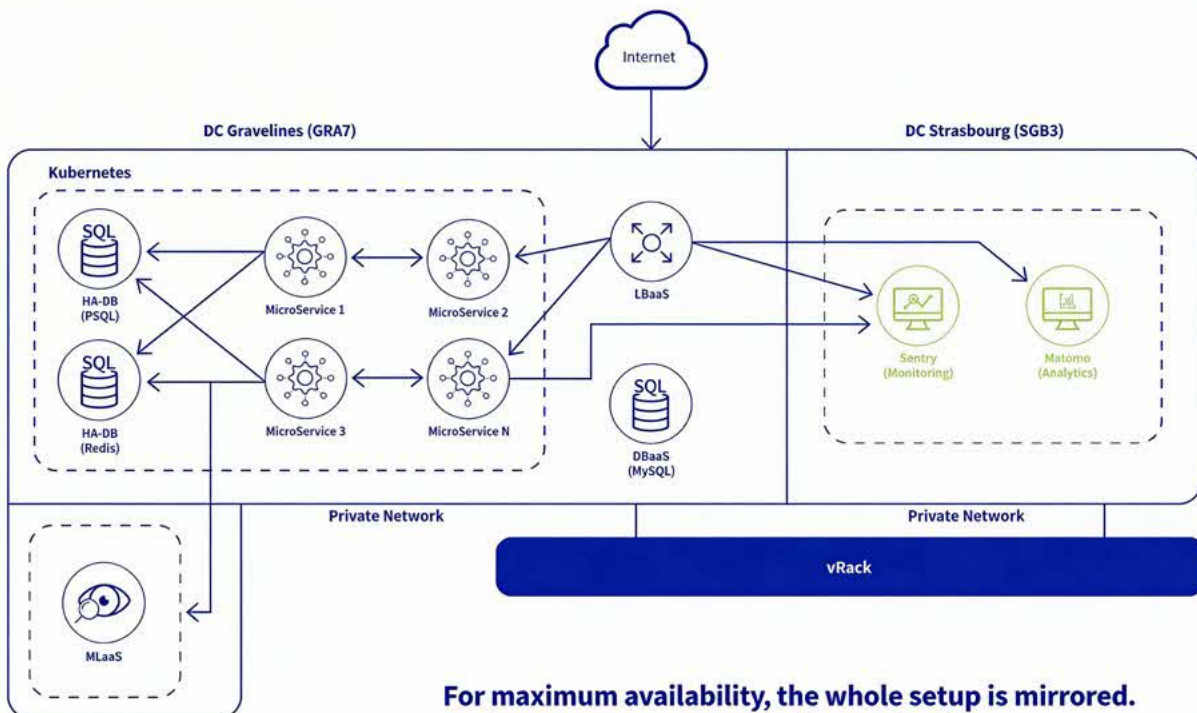
LA SOLUCIÓN

- ▶ Migración de cargas de trabajo a datacenters de OVHcloud en la UE para cumplir con el RGPD
- ▶ Despliegue de clústeres de PostgreSQL de alta disponibilidad en múltiples AZ
- ▶ Introducción de redes privadas vRack para tráfico este-oeste de baja latencia
- ▶ Uso de bases de datos gestionadas e instancias de computación escalables para un rendimiento predecible

LOS RESULTADOS

- ▶ Reducción de aproximadamente un 20 % en los requerimientos de recursos y costes en comparación con la configuración anterior
- ▶ Mejoras significativas en la latencia, independientemente de la ubicación del usuario
- ▶ Alojamiento de datos totalmente seguro y conforme al RGPD, con gobernanza sólida y patrón de alta disponibilidad

[Leer más](#)



For maximum availability, the whole setup is mirrored. Both DC are reverse monitoring each other.

MapTiler: generación de mapas satelitales escalables con instancias cloud ilimitadas

MapTiler, una scaleup suiza del sector geoespacial, crea y proporciona mapas base de alto rendimiento y datos cartográficos personalizados utilizados por aplicaciones en industrias como logística, bienes raíces, defensa y turismo. Para mantenerse competitiva con imágenes satelitales

actualizadas y atender cientos de millones de visualizaciones de mapas cada mes, MapTiler necesitaba un proveedor cloud capaz de escalar de manera rentable y eliminar los cuellos de botella en la infraestructura.



EL RETO CLAVE

- ▶ Gestionar enormes y crecientes necesidades de almacenamiento a medida que se acumulan los datos satelitales
- ▶ Atender de manera fiable alrededor de 400 millones de solicitudes diarias de teselas de mapas
- ▶ Eliminar costes mensuales impredecibles vinculados a volúmenes de datos fluctuantes
- ▶ Reducir el tiempo de procesamiento de décadas a semanas con capacidad de computación ilimitada



LA SOLUCIÓN

- ▶ Migración a OVHcloud Public Cloud con almacenamiento de objetos escalable compatible con S3
- ▶ Aprovechamiento de instancias cloud ilimitadas para el procesamiento de datos en paralelo
- ▶ Adopción de estructuras de costes predecibles para almacenamiento y computación
- ▶ Simplificación de la generación de mapas satelitales mediante infraestructura rentable

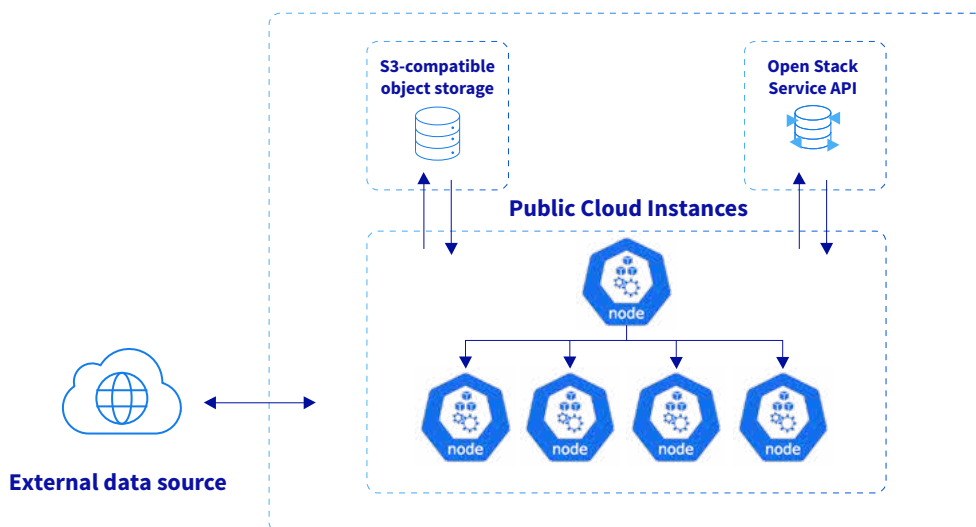


LOS RESULTADOS

- ▶ El tiempo de procesamiento de datos satelitales se redujo de más de 18 años estimados a solo unas pocas semanas
- ▶ Costes mensuales predecibles, independientemente de las cargas de trabajo de datos fluctuantes
- ▶ Instancias de computación ilimitadas que previenen techos de capacidad y aceleran la producción de mapas

[Leer más](#)

MapTiler's cloud infrastructure at OVHcloud



*S3 is a registered trademark of Amazon Technologies, Inc. OVHcloud services are not sponsored or approved by, nor affiliated with Amazon Technologies, Inc. in any way.

Escalar cargas de trabajo con gran volumen de datos con confianza

A medida que las cargas de trabajo intensivas en datos crecen, la complejidad operativa aumenta junto con el volumen de datos, los usuarios y los servicios. Sin un diseño estratégico, pueden aparecer cuellos de botella en el rendimiento, costes crecientes y sobrecarga operativa, especialmente cuando los sistemas se encuentran bajo presión.

Los equipos exitosos se aseguran de que el almacenamiento, el cálculo, las bases de datos y las redes escalen de manera independiente. Cuando esto se combina con servicios administrados para tareas como almacenamiento, orquestación y monitorización, los ingenieros pueden

concentrarse en las cargas de trabajo principales mientras mantienen la flexibilidad operativa.

La escalabilidad a largo plazo permite a los equipos mantener el control incluso cuando:

- ▶ El volumen de datos aumenta, ya sea por 10 o por 100
- ▶ Los servicios o pipelines se multiplican
- ▶ Los fallos se vuelven más frecuentes

Una infraestructura cloud flexible minimiza los obstáculos e impulsa un crecimiento continuo sin complicaciones.

Escala SaaS intensivos en datos y API de alto volumen con confianza.

Descubre un cloud que crece contigo.

¿Quieres saber más?

Programar una llamada con un arquitecto de soluciones:

[Solicitar una llamada](#)

Descubre el cloud escalable para empresas en crecimiento:

[Saber más](#)