OVHcloud

Designs für datenintensive Workloads

Zuverlässige Skalierung von SaaS- und API-Plattformen in der Cloud

OVHcloud

Skalierung datenintensiver Workloads in der Praxis

Dieser Use Case bietet eine praktische, technische Sicht auf die Skalierung datenintensiver Workloads, einschließlich:

1. Häufig auftretende Herausforderungen, wenn Systeme wachsen
2. Eine Checkliste zum technischen Design als Grundlage für Infrastrukturentscheidungen
3. Bereitstellungsbeispiele, die zeigen, wie Unternehmen Leistung, Effizienz und Kontrolle aufrechterhalten

Anhand dieser Insights können Ingenieur:innen und Architekt:innen sich ein besseres Bild von den Mustern und Strategien machen, die eine zuverlässige Plattformskalierung ermöglichen, ohne dass die Komplexität oder Kosten ausufern.

1. Die technischen Herausforderungen bei der Skalierung datenintensiver Systeme

Wenn Datenvolumen, Durchsatz und Parallelität zunehmen, stoßen Teams häufig auf Herausforderungen, darunter:

- **Leistungsengpässe beim Storage,** wenn Datensätze von Gigabyte auf Terabyte und mehr anwachsen
- **Schwierigkeiten bei der Vorhersage von Kosten** in Zusammenhang mit dem Storage-Wachstum, der I/O-Nutzung und dem Daten-Egress
- **Aufrechterhaltung einer niedrigen und konsistenten Latenz** bei Traffic-Spitzen über die Datenaufnahme-, Datenabfrage- und API-Layer hinweg
- **Zunehmender betrieblicher Overhead** durch die Verwaltung verteilter Datenbanken und Pipelines
- **Keine Möglichkeit, Storage, Datenbanken und Compute-**Ressourcen unabhängig zu skalieren, was zu Ineffizienz führt
- **Datenresidenz- und Compliance-Anforderungen,** die die architektonische Komplexität erhöhen
- **Netzwerkengpässe, die durch datenintensiven** Ost-West-Traffic zwischen Diensten verursacht werden

Zum effektiven Bewältigen dieser Herausforderungen braucht es eine Infrastruktur, die Sichtbarkeit und Kontrolle im Hinblick darauf bietet, wie Daten gespeichert, bewegt, verarbeitet und skaliert werden. Das wird auch eine vorhersehbare Leistung und berechenbare Kosten ermöglichen, wenn die Plattformen wachsen.

2. Checkliste für das technische Design

Bevor Sie eine Cloud-Plattform auswählen oder sich für ein Architekturmuster entscheiden, sollten Sie sich vor Augen führen, wie Ihre Workload sich in der Produktion verhält. Große Datensätze, anhaltende I/O und hohe Request-Volumen führen zu spezifischen Anforderungen in puncto Storage, Netzwerk und verteiltes Computing.

Die folgende Checkliste soll Ihnen helfen, die technischen Merkmale zu bestimmen, die Sie bei der Wahl eines Cloud-Anbieters oder einer Architektur für Ihre Anforderungen berücksichtigen sollten.

Storage

- Erforderliches Speichermodell: Object, Block, File oder hybrid
- Spitzen-IOPS-Anforderungen
- NVMe-Anforderungen für latenzempfindliche Workloads
- Primärer Leistungsfaktor: Durchsatz vs. zufälliges Lesen/Schreiben
- Verlauf des Datenwachstums (GB/TB/PB)
- Modell für die Replikation und Beständigkeit: eine Zone, Mutli-AZ oder mehrere Regionen

Computing und Verarbeitung

- Workload-Profil: CPU-bound, memory-bound oder GPU-gestützt
- Datenverarbeitungsmodus: Batch, Streaming oder Echtzeit
- Anforderungen in Sachen verteilte Verarbeitung: Spark-, Dask-, ClickHouse-Cluster oder andere
- Nutzungsmuster: stabil, Bursting-geprägt oder ereignisbasiert
- Funktionen für die Containerisierung und automatische Skalierung

Netzwerk

- Ausmaß der Latenzempfindlichkeit
- Erforderliche East-West-Bandbreite zwischen Diensten oder Nodes
- Anforderungen in puncto Multi-Zonen- oder Multi-Region-Architekturen
- Anforderungen im Hinblick auf private Netzwerke
- Erwartete Egress-Volumen

Skalierung

- Skalierungsmodell: horizontal (stateless), vertikal (Durchsatz) oder hybrid
- Erwartete Traffic- oder Lastspitzen: täglich, saisonabhängig oder ereignisgesteuert
- Anforderungen rund um Autoskalierung, Sharding oder Cluster-Erweiterung
- Eignung für die Containerisierung

OVHcloud

Kosten und Ressourcenverhalten

- Nutzungsprognose: stabil vs. sehr variabel
- Primäre Kostenfaktoren: Storage-, Compute-, Egress- oder Datenbankskalierung
- Präferenz beim Kostenmodell: monatlich vorhersehbar vs. nutzungsabhängig
- Kostenauswirkungen des Wachstums und der Replikation von Objektspeicher

Compliance und Sicherheit

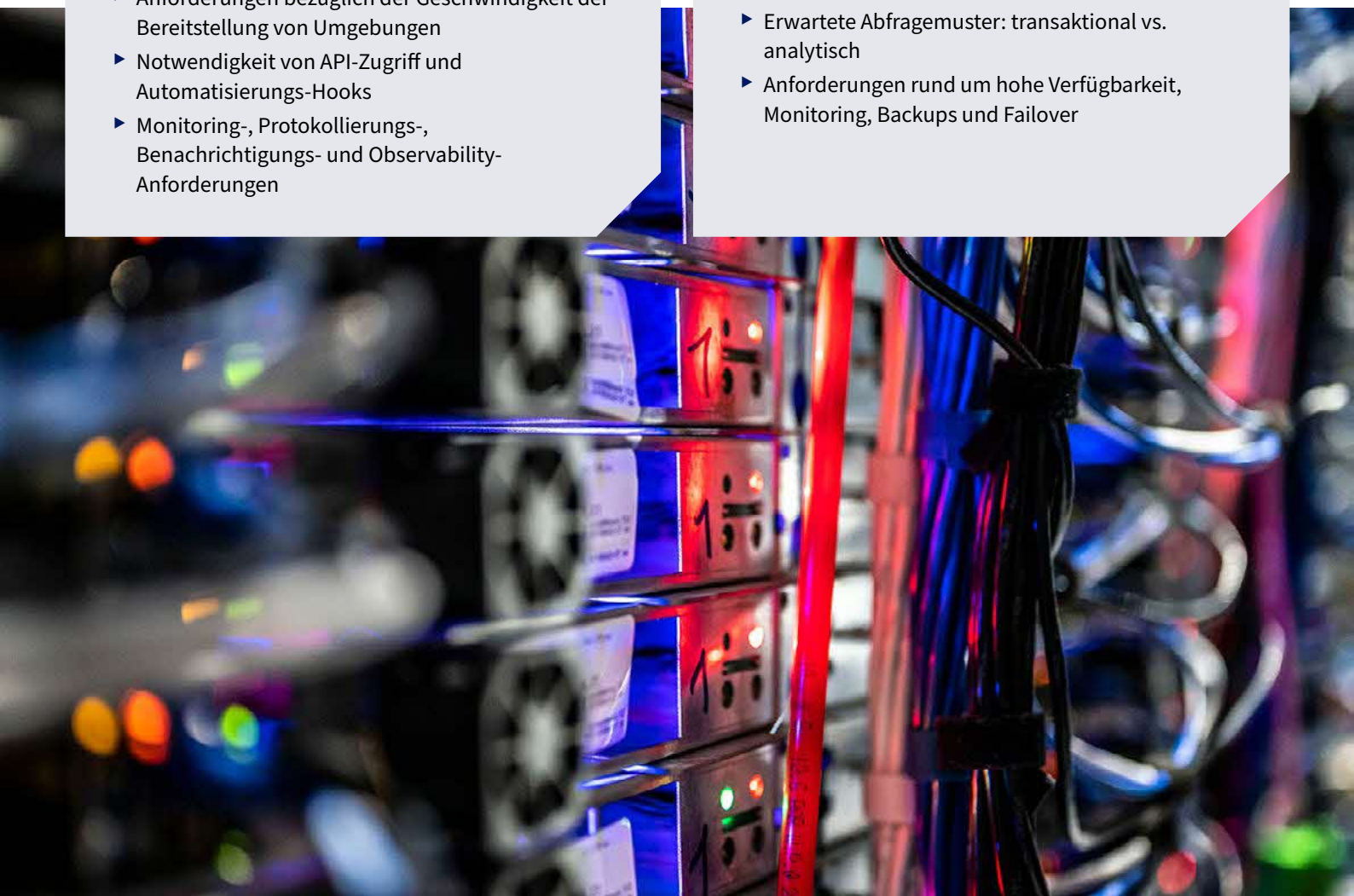
- Datenresidenz- und Souveränitätsanforderungen
- Regulatorischen Auflagen in Bezug auf die Datenlokalisierung
- Erforderliche Zertifizierungen
- Anforderungen in Sachen Verschlüsselung, Netzwerkisolierung und IAM-Integration
- SLA, Kunden-Erwartungen und Governance-Anforderungen in Bezug auf Datenschutz und Compliance

Automatisierung und Bereitstellung

- Anforderungen rund um Infrastructure-as-Code-Tools
- CI/CD-Integrationsanforderungen
- Anforderungen bezüglich der Geschwindigkeit der Bereitstellung von Umgebungen
- Notwendigkeit von API-Zugriff und Automatisierungs-Hooks
- Monitoring-, Protokollierungs-, Benachrichtigungs- und Observability-Anforderungen

Datenbanküberlegungen

- Datenbankmodell: dokumentorientiert (MongoDB) vs. relational (PostgreSQL)
- Anforderungen in Bezug auf gleichzeitiges Lesen/Schreiben
- Erwartete Abfragemuster: transaktional vs. analytisch
- Anforderungen rund um hohe Verfügbarkeit, Monitoring, Backups und Failover

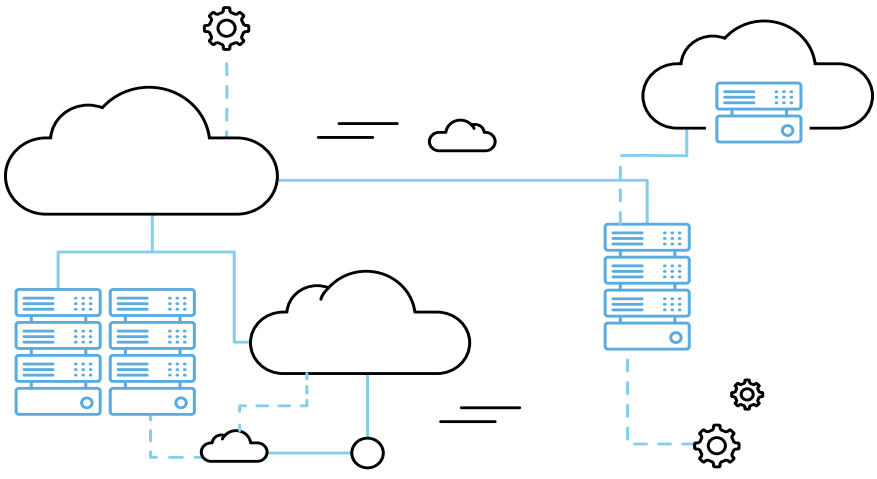


OVHcloud

Engineering-Ergebnisse

Gut gestaltete Plattformen, die mit einem klaren Verständnis dafür entworfen wurden, wie aktuelle und künftige Workloads sich in der Produktion verhalten, ermöglichen es Ingenieur:innen, Verbesserungen in Sachen Zuverlässigkeit, Leistung und Effizienz zu erzielen. Im Wesentlichen kommen sie der Aufrechterhaltung konsistenter Servicelevel, der unabhängigen Skalierung von Ressourcen sowie dem Dienst- und Kostenmanagement zugute.

FEATURE	BETRIEBLICHE VORTEILE
✓ Stabiler Durchsatz bei der Datenaufnahme	✓ Weniger Verluste, konsistente SLAs
✓ Kürzere Abfrage- und Verarbeitungszeiten	✓ Verbesserte API-Latenz, Reaktionsfähigkeit bei Analysen
✓ Unabhängige Skalierung und Speicherung	✓ Separates Skalieren von Storage, Datenbanken und Compute-Ressourcen
✓ Weniger betrieblicher Overhead	✓ Geringerer Zeitaufwand für Patching, Hochverfügbarkeit und Backups
✓ Berechenbare Cloud-Kosten	✓ Keine überraschenden Egress- oder IOPS-Gebühren
✓ Datenkontrolle und Compliance	✓ Klare Datenplatzierung und -residenz
✓ Offene Architekturen	✓ Lock-in vermeiden, einfachere Portabilität



OVHcloud

Designs für AI-Workloads

AI beschleunigt das Wachstum datenintensiver Workloads, die hohe Anforderungen in puncto Compute-Ressourcen, Speicherdurchsatz und Effizienz verursachen. Beim Skalieren dieser Workloads benötigen Teams eine Infrastruktur, die eine vorhersehbare Leistung, flexible Bereitstellungsmodelle und operative Konsistenz ermöglicht, ohne dass es zu unnötiger Komplexität kommt.

Ein praktischer Ansatz besteht in der Standardisierung mit einem weitreichend kompatiblen Technologiestack über hybride Umgebungen hinweg.

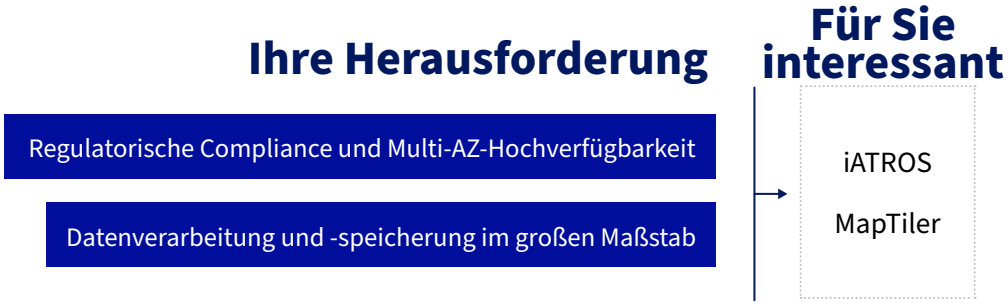
Plattformanbieter wie OVHcloud unterstützen dies mit einer Kombination aus sofort bereitstellbaren On-Premise-Servern (On-Prem Cloud Platform), Bare-Metal-Servern und klassischen Cloud-Instanzen, die oft auf leistungsstarken Architekturen wie AMD EPYC basieren.

So sind Engineering-Teams in der Lage, die richtige Infrastruktur für jede Workload zu wählen – und profitieren gleichzeitig von vertrauten Tools, vorhersehbarem Verhalten und Effizienz im großen Maßstab.

3. Bereitstellungsbeispiele: Wie Unternehmen datenintensive SaaS und APIs skalieren

Bei datenintensiven SaaS- und API-Plattformen sind Sie je nach Traffic, Datensatzgröße und Workload-Typ mit unterschiedlichen Infrastrukturherausforderungen konfrontiert.

Die folgenden Beispiele verdeutlichen, wie wachsende Unternehmen für durchweg vorhersehbare Leistung sorgen, effizient skalieren und die Kosten im Griff behalten.



iATROS: Eine sichere und konforme digitale Gesundheitsplattform im großen Maßstab

iATROS bietet eine digitale Gesundheitsplattform, die sensible Patientendaten für Hunderttausende User erfasst, analysiert und bereitstellt. Um strenge regulatorische Anforderungen – einschließlich DSGVO und branchenspezifische Standards – zu erfüllen und die Latenz für geografisch verteilte User zu reduzieren, hat das Team seinen Stack mit einer Multi-Cluster-Cloud-Infrastruktur mit hoher Verfügbarkeit und starken Kontrollen als Herzstück neu gestaltet.

DIE WESENTLICHEN *HERAUSFORDERUNGEN*

- Migration weg von einer Cloud, die die europäischen Datenschutzanforderungen nicht vollständig erfüllte
- Hohe Verfügbarkeit und niedrige Latenz für geografisch verteilte User gewährleisten
- Strenge Compliance- und Governance-Standards erfüllen (DSGVO, ISO, Gesundheitssektor)
- Resiliente Infrastruktur über Fault Domains hinweg gewährleisten

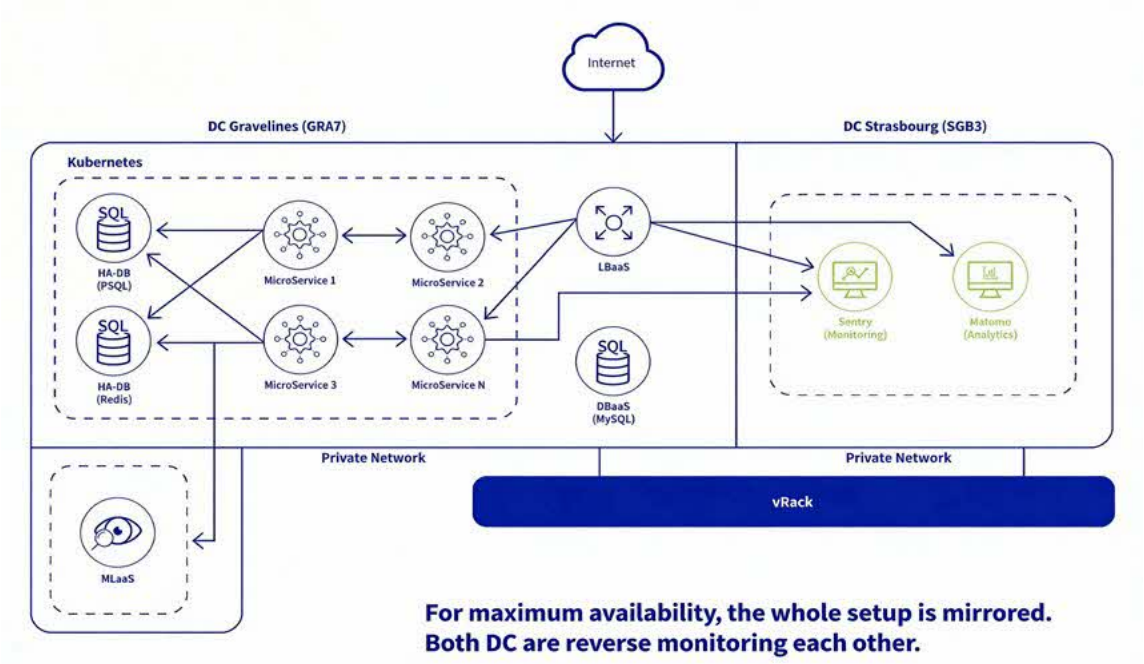
DIE LÖSUNG

- Migration von Workloads zu den EU-basierten Rechenzentren von OVHcloud für die Einhaltung der DSGVO
- Bereitstellung hochverfügbarer PostgreSQL-Cluster über mehrere AZs hinweg
- Einführung des privaten vRack Netzwerks für Ost-West-Traffic mit niedriger Latenz
- Nutzung von Managed Databases und skalierbaren Compute-Instanzen für vorhersehbare Leistung

DIE ERGEBNISSE

- Reduzierung des Ressourcenbedarfs und der Kosten um ~20 % im Vergleich zum vorherigen Setup
- Bedeutende Latenzverbesserungen unabhängig vom Standort des Users
- Komplett sicheres und DSGVO-konformes Datenhosting mit robuster Governance und Hochverfügbarkeit

Mehr erfahren



MapTiler: Skalierbare Generierung von Satellitenkarten mit unbegrenzten Cloud-Instanzen

Das im Bereich Geodaten tätige Scale-up MapTiler aus der Schweiz generiert Basemaps und individuelle Kartendaten, die von Anwendungen in verschiedensten Branchen wie Logistik, Immobilien, Verteidigung und Tourismus genutzt werden. Um mit aktuellen Satellitenbildern wettbewerbsfähig zu bleiben und jeden Monat Hunderte Millionen Kartenansichten zu bedienen, benötigte MapTiler einen Cloud-Anbieter, der kosteneffizient skalieren und Infrastrukturengpässe beseitigen kann.

DIE WESENTLICHEN HERAUSFORDERUNGEN

- Bewältigung enormer, schnell zunehmender Storage-Anforderungen im Zuge der Erfassung von Satellitendaten
- Zuverlässige Unterstützung von ~400 Millionen täglichen Anfragen für Kartenkacheln
- Beseitigung unvorhersehbarer monatlicher Kosten, die an schwankende Datenvolumen geknüpft sind
- Verkürzung der Verarbeitungszeit von Jahrzehnten auf Wochen mit unbegrenzter Rechenkapazität

DIE LÖSUNG

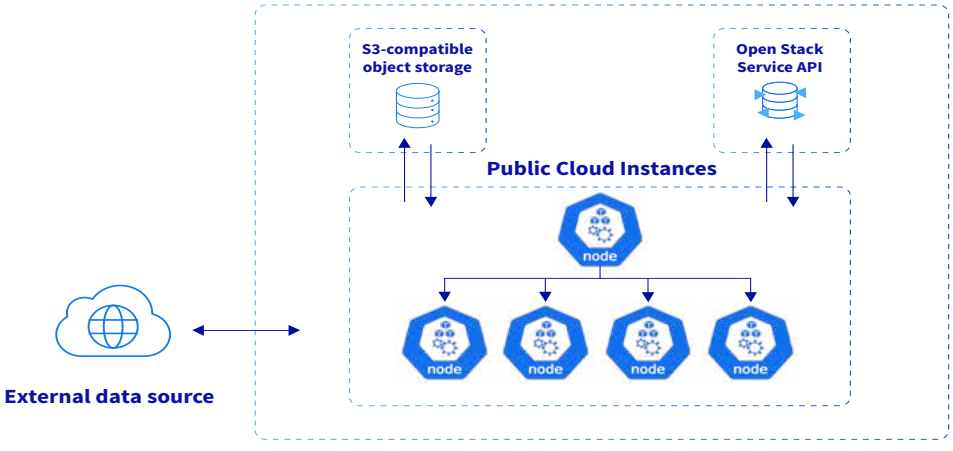
- Migration zur OVHcloud Public Cloud mit skalierbarem S3-kompatiblem Object Storage
- Nutzung unbegrenzter Cloud-Instanzen für die parallele Datenverarbeitung
- Einführung vorhersehbarer Kostenstrukturen für Storage und Compute
- Vereinfachte Satellitenkartenerstellung mit kosteneffizienter Infrastruktur

DIE ERGEBNISSE

- Verkürzung der Verarbeitungszeit für Satellitendaten von geschätzten >18 Jahren auf nur wenige Wochen
- Vorhersehbare monatliche Kosten unabhängig von schwankenden Daten-Workloads
- Keine Kapazitätseinschränkungen und Beschleunigung der Kartenproduktion dank unbegrenzter Compute-Instanzen

Mehr erfahren

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.*

OVHcloud

Souveränes Skalieren datenintensiver Workloads

Mit dem Wachstum datenintensiver Workloads steigt die betriebliche Komplexität in Zusammenhang mit Datenvolumen, Usern und Diensten. Ohne sorgfältiges Design können Leistungsengpässe, steigende Kosten und betrieblicher Overhead entstehen.

Erfolgreiche Teams stellen sicher, dass Storage, Compute-Ressourcen, Datenbanken und Netzwerke unabhängig skalieren können. Durch die Kombination mit Managed Services für Aufgaben wie die Speicherung, Orchestrierung und Überwachung hat das Engineering-Team mehr Zeit für die wesentlichen Workloads – bei gleichzeitiger Wahrung der betrieblichen Flexibilität.

Die langfristige Skalierbarkeit ermöglicht es Teams, die Kontrolle zu behalten, selbst wenn:

- Das Datenvolumen steigt, ob um das 10-Fache oder um das 100-Fache
- Dienste oder Pipelines sich vervielfachen
- Fehler häufiger auftreten

Eine flexible Cloud-Grundlage reduziert die Probleme und unterstützt nachhaltiges Wachstum ohne Kompromisse.

Skalieren Sie datenintensive SaaS und hochvolumige APIs souverän. OVHcloud macht den Unterschied.

Sie wollen mehr erfahren?

Vereinbaren Sie einen Anruf mit einem Solution Architect:

Rückruf anfordern

Entdecken Sie die skalierbare Cloud für wachsende Unternehmen:

Mehr erfahren