OVHcloud

Designing for data-intensive workloads

How SaaS and API platforms scale reliably in the cloud

OVHcloud

Scaling data-intensive workloads, in practice

This use case offers a practical, technical view of scaling data-intensive workloads, including:

1. Common challenges encountered as systems grow
2. A technical design checklist to guide infrastructure decisions
3. Deployment examples showing how organisations maintain performance, efficiency and control

By exploring these insights, engineers and architects can better understand the patterns and strategies that enable platforms to scale reliably while managing complexity and cost.

OVHcloud

1. The technical challenges behind scaling data-intensive systems

As data volumes, throughput and concurrency increase, teams commonly encounter challenges including:

Storage performance constraints as datasets grow from gigabytes to terabytes and beyond.

Difficulty predicting costs related to storage growth, I/O usage and data egress.

Maintaining low and consistent latency during traffic bursts across ingestion, query and API layers.

Increasing operational overhead from managing distributed databases and pipelines.

Inability to scale storage, databases and compute independently, leading to inefficiency.

Data residency and compliance requirements adding architectural complexity.

Network bottlenecks caused by data-heavy east-west traffic between services.

Addressing these challenges effectively requires infrastructure that provides visibility and control over how data is stored, moved, processed and scaled. This will facilitate predictable performance and costs as platforms grow.

2. Technical design checklist

Before selecting a cloud platform or committing to an architecture pattern, consider how your workload behaves in production. Large datasets, sustained I/O, and high request volumes place specific demands on storage, networking and distributed compute.

The following checklist is designed to help you clearly identify the technical characteristics to consider when choosing a cloud provider or architecture best suited to your organisational needs.

Storage

- Required storage model: object, block, file, or hybrid
- Peak IOPS requirements
- NVMe requirements for low-latency workloads
- Primary performance driver: throughput vs random read/write
- Data growth trajectory (GB / TB / PB)
- Replication and durability model: single-zone, multi-AZ, or multi-region

Compute & Processing

- Workload profile: CPU-bound, memory-bound, or GPU-assisted
- Data processing mode: batch, streaming, or real-time
- Distributed processing requirements: Spark, Dask, ClickHouse clusters, or others
- Usage pattern: steady-state, burst-driven, or event-based
- Containerisation and autoscaling capability

Networking

- Latency sensitivity level
- Required east-west bandwidth requirements between services or nodes
- Multi-zone or multi-region architecture requirements
- Private networking requirements
- Expected egress volumes

Scaling

- Scaling model: horizontal (stateless), vertical (throughput), or hybrid
- Expected traffic bursts or load spikes: daily, seasonal, or event-driven
- Autoscaling, sharding, or cluster expansion requirements
- Containerisation readiness

OVHcloud

Cost & Resource Behaviour

- Usage predictability: stable vs. highly variable
- Primary cost drivers: storage, compute, egress, or database scaling
- Spend model preference: predictable monthly vs. pay-per-use
- Cost impact of object storage growth and replication

Compliance & Security

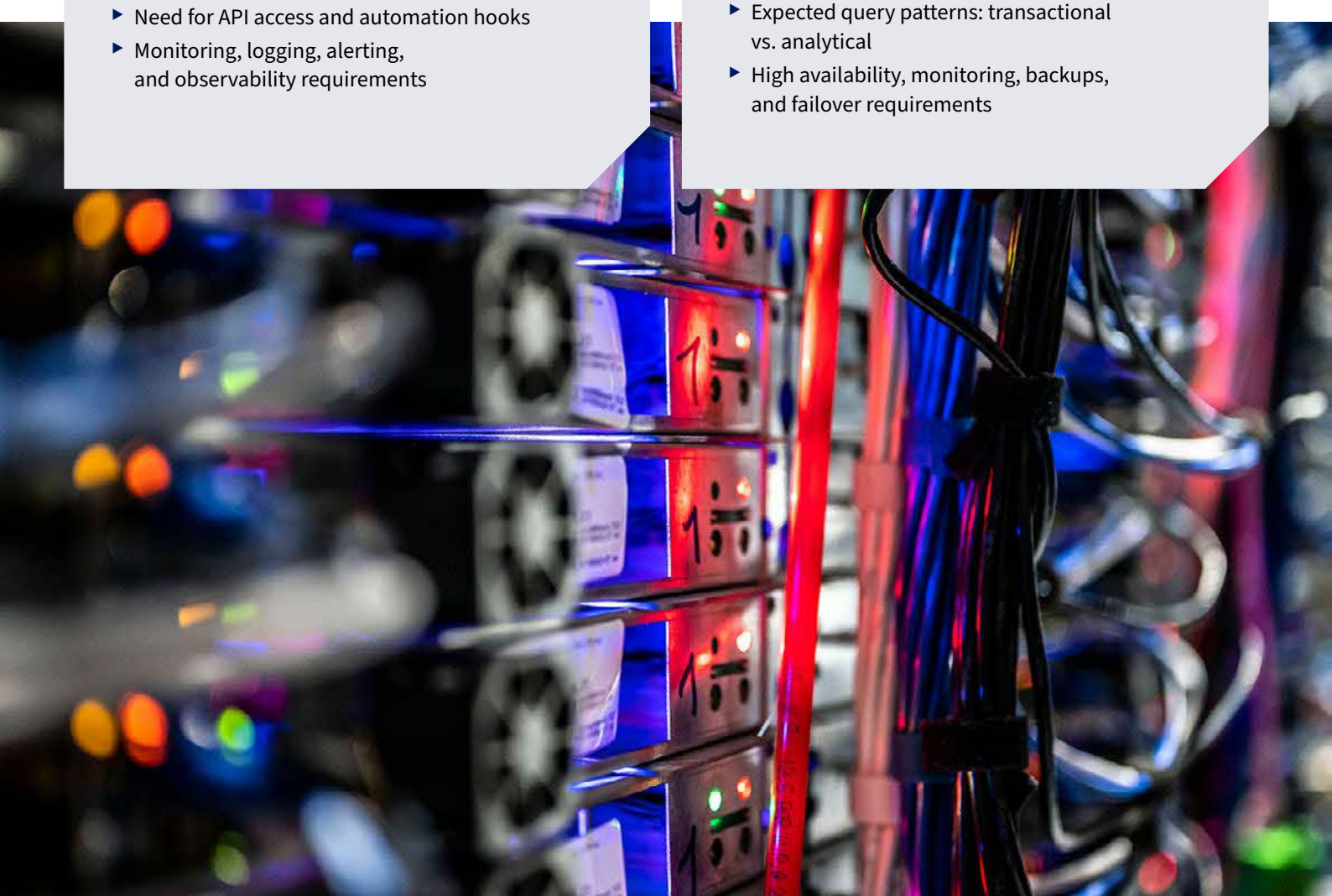
- Data residency and sovereignty requirements
- Regulatory data locality constraints
- Required certifications
- Encryption, network isolation, and IAM integration requirements
- SLA, customer and governance expectations around privacy and compliance

Automation & Deployment

- Infrastructure-as-Code tooling requirements
- CI/CD integration requirements
- Environment provisioning speed requirements
- Need for API access and automation hooks
- Monitoring, logging, alerting, and observability requirements

Database Considerations

- Database model: document (MongoDB) vs. relational (PostgreSQL)
- Concurrent read/write requirements
- Expected query patterns: transactional vs. analytical
- High availability, monitoring, backups, and failover requirements

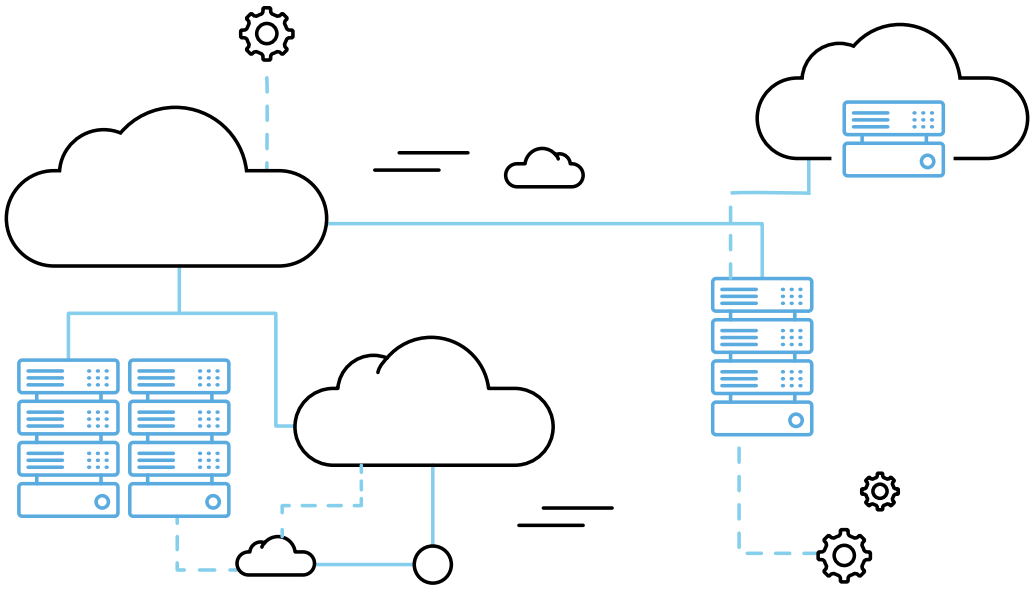


Engineering outcomes

Well-architected platforms, designed with a clear understanding of how current and future workloads behave in production enable engineers to achieve improvements in reliability, performance and efficiency.

Critically, they allow maintenance of consistent service levels, independent scaling of resources, and service and cost management.

Capability	Operational Gains
✓ Stable ingestion throughput	✓ Fewer dropped events, consistent SLAs
✓ Faster query & processing times	✓ Improved API latency, analytics responsiveness
✓ Independent scaling & storage	✓ Scale storage, databases and compute separately
✓ Reduced operational overhead	✓ Less time spent on patching, HA, backups
✓ Predictable cloud costs	✓ No surprise egress or IOPS charges
✓ Data control & compliance	✓ Clear data placement and residency
✓ Open architectures	✓ Avoid lock-in, easier portability



OVHcloud

Designing for AI workloads

AI is accelerating the growth of data-intensive workloads that combine **high compute demand, storage throughput and strict efficiency requirements.**

As these workloads scale, teams need infrastructure that supports predictable performance, flexible deployment models and operational consistency – without adding unnecessary complexity.

One practical approach is to standardise on a broadly compatible technology stack across hybrid environments.

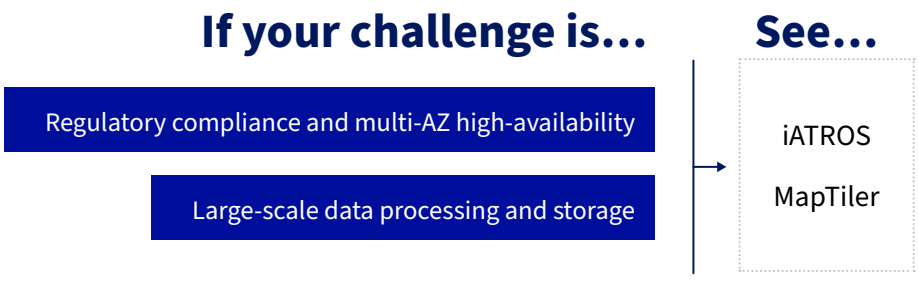
Platforms such as OVHcloud illustrate this by offering a mix of on-premise ready-to-deploy servers (On-Prem Cloud Platform), bare-metal servers and traditional cloud instances, often built on high performance architecture such as AMD EPYC™.

This allows engineering teams to select **the right infrastructure for each workload** while benefiting from familiar tooling, predictable behaviour and efficiency at scale.

3. Deployment examples: How organisations scale data -intensive SaaS and APIs

Data-intensive SaaS and API platforms face different infrastructure challenges depending on traffic, dataset size and workload type.

The following examples highlight how scaling businesses maintain predictable performance, scale efficiently and control costs.



iATROS: A secure and compliant digital health platform at scale

iATROS delivers a digital health platform that collects, analyses and serves sensitive patient data for hundreds of thousands of users. To satisfy stringent regulatory requirements – including GDPR and sector-specific standards – and reduce latency for geographically distributed users, the team rearchitected its stack around multi-cluster cloud infrastructure with high availability and strong controls.

Core Challenge

- Migrating from a cloud that didn't fully meet European data-protection requirements
- Delivering high availability and low latency for geographically distributed users
- Meeting stringent compliance and governance standards (GDPR, ISO, health sector)
- Maintaining resilient infrastructure across fault domains

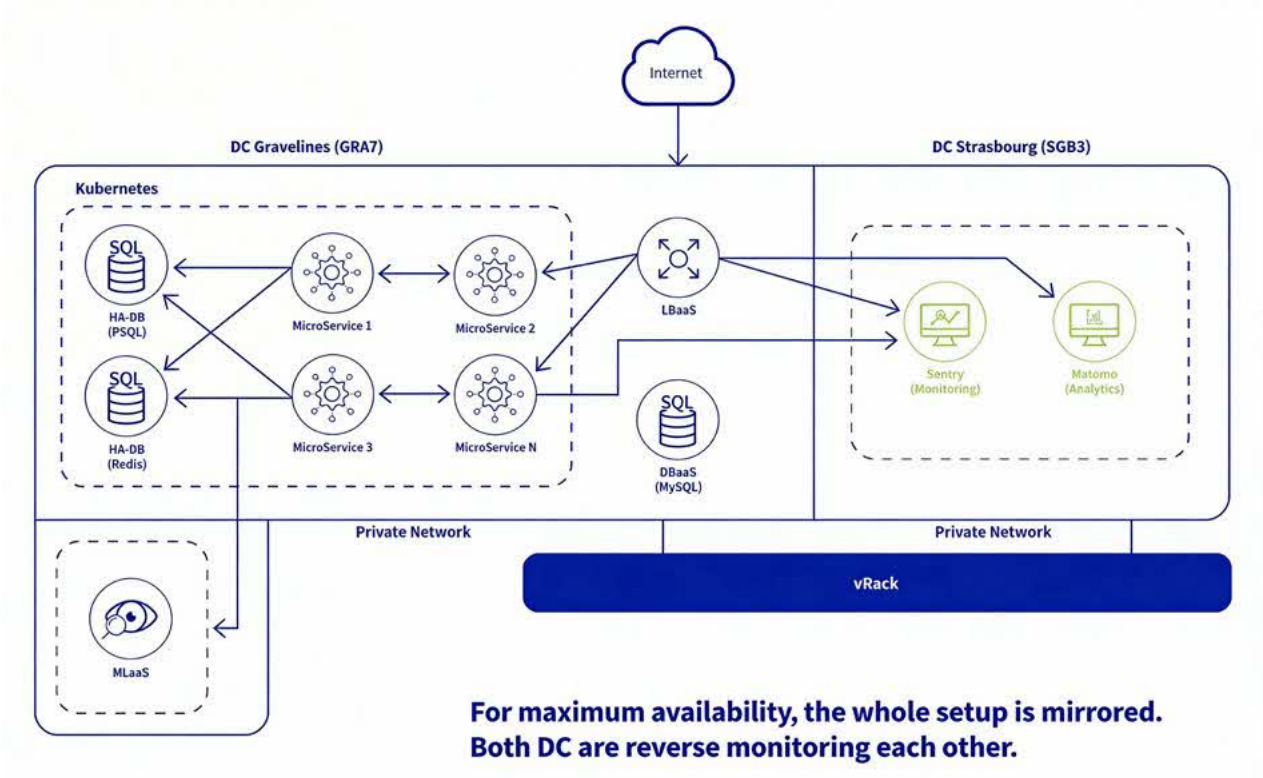
The Solution

- Migrated workloads to OVHcloud EU-based data centres for GDPR compliance
- Deployed high-availability PostgreSQL clusters across multiple AZs
- Introduced vRack private networking for low-latency east-west traffic
- Used Managed Databases and scalable compute instances for predictable performance

The Results

- ~20% reduction in resource needs and cost compared with prior setup
- Significant latency improvements regardless of user location
- Fully secure and GDPR-compliant data hosting with robust governance and HA pattern

Read more



MapTiler: Scalable satellite map generation with unlimited cloud instances

MapTiler, a Swiss geospatial scale-up, builds and delivers high-performance basemaps and custom map data used by applications across industries like logistics, real-estate, defence, and tourism.

To stay competitive with up-to-date satellite imagery and serve hundreds of millions of map views every month, MapTiler needed a cloud provider that could scale cost-efficiently and eliminate infrastructure bottlenecks.

Core Challenge

- Handling enormous, rapidly growing storage needs as satellite data accumulates
- Supporting ~400 million daily requests for map tiles reliably
- Eliminating unpredictable monthly costs tied to fluctuating data volumes
- Reducing processing time from decades to weeks with unlimited compute capacity

The Solution

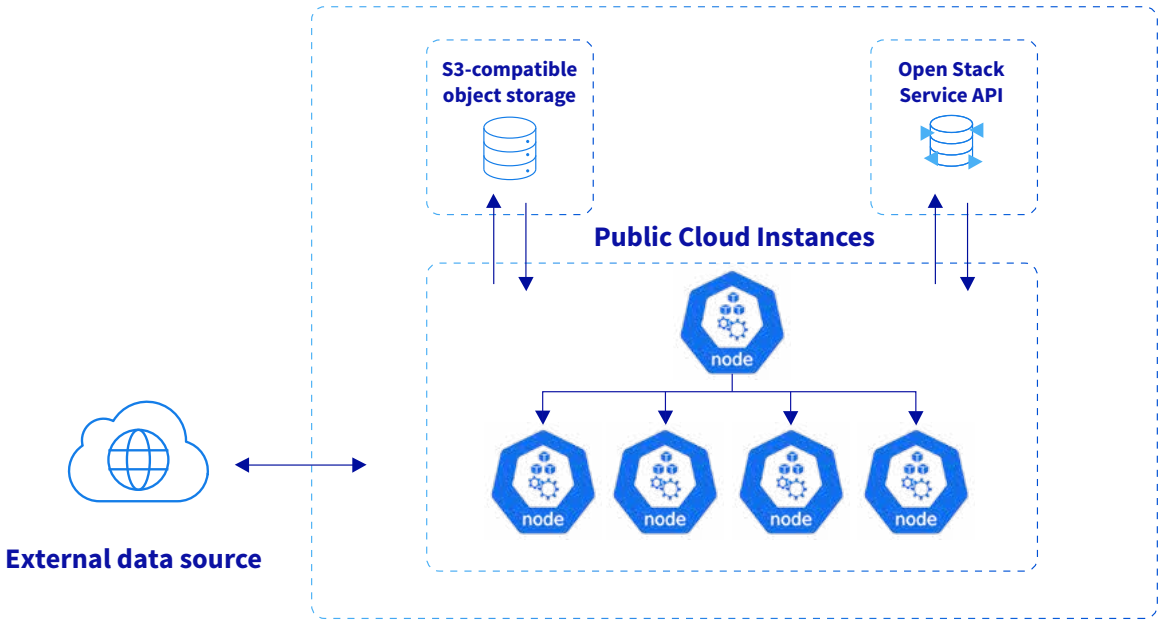
- Migrated to OVHcloud Public Cloud with S3*-compatible object storage
- Leveraged unlimited cloud instances for parallel data processing
- Adopted predictable cost structures for storage and compute
- Simplified satellite map generation with cost-effective infrastructure

The Results

- Processing time for satellite data shrank from an estimated >18 years to just a few weeks
- Predictable monthly costs regardless of fluctuating data workloads
- Unlimited compute instances prevent capacity ceilings and accelerate map production

Read more

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

Scaling data-intensive workloads with confidence

As data-intensive workloads grow, operational complexity escalates with data volume, users and services. Without careful design, performance bottlenecks, increasing costs and operational overhead may arise – often when systems are under strain.

Successful teams ensure storage, compute, databases and networking scale independently. When used alongside managed services for tasks like storage, orchestration and monitoring, engineers are free to focus on core workloads while maintaining operational flexibility.

Long-term scalability lets teams maintain control even when:

- Data volume increases, whether by 10x or 100x
- Services or pipelines multiply
- Failures become more frequent

A flexible cloud foundation reduces friction and supports sustained growth without compromise.

Scale data-intensive SaaS and high-volume APIs with confidence. Discover the OVHcloud difference.

Want to find out more?

Arrange a call
with a Solution Architect:

Request a call

Explore scalable cloud
for growing businesses:

Learn more