



OVHcloud Products Portfolio Overview

PRODUCT MARKETING

27.01.26



PUBLIC CLOUD PRODUCTS PORTFOLIO

INSTANCES	STORAGE	NETWORK	ENABLER	CONTAINERS	DATABASES	ANALYTICS	AI	QUANTUM	DEVELOPERS EXPERIENCE		
GENERAL PURPOSE	BLOCK	VRACK L2	IAM	KUBERNETES	MYSQL	DATA PLATFORM	NOTEBOOKS	EMULATORS	API		
COMPUTE OPTIMIZED	OBJECT	LOAD BALANCER	LOGS OBSERVABILITY	REGISTRY	POSTGRESQL	KAFKA	TRAINING	QPU (QUANTUM PROCESSING UNITS)	CLI		
MEMORY OPTIMIZED	FILE	GLOBAL ANTI-DDOS	KMS	RANCHER	VALKEY	OPENSEARCH	DEPLOY	CLOUD SHELL	CLOUD SHELL		
STORAGE OPTIMIZED	EDGE FIREWALL	EDGE FIREWALL	SECRET MANAGER	GRAFANA	MONGODB	GRAFANA	ENDPOINTS	SDK	SDK		
DISCOVERY		FLOATING IP	HARDWARE SECURITY MODULE		CLICKHOUSE	CLICKHOUSE	ENDPOINTS				
METAL		PUBLIC GATEWAY	METRICS OBSERVABILITY								
CLOUD GPU		OVH CLOUD CONNECT	CO2 IMPACT								
SERVERLESS		VPC L3	CLOUD EVENTS								
		PRIVATE DNS/DHCP	SEND EMAIL SERVICE								
		VPN / WAF									
		GLOBAL LOAD BALANCER									
DEPLOYMENT MODELS											
					1-AZ REGION	3-AZ REGION	LOCAL ZONE	ON PREM (OPCP)	SNC		
BILLING MODELS											
					HOURLY	MONTHLY	SAVINGS PLANS 1 6 12 24 36 Months				
					RESERVED						
 OVHcloud											
AVAILABLE											
PLANNED											



Compute Instances

Predictable, sovereign and cost-efficient compute to run production workloads at scale.

Key Capabilities

Predictable Performance

Stable and consistent compute performance for production workloads through dedicated vCPU and transparent resource allocation

Cost Control by Design

Lower and predictable TCO compared to hyperscalers

- Simple and transparent pricing
- Savings Plans for long-running workloads

Flexible Compute Portfolio

Right compute for every workload stage.

- General Purpose, Compute Optimized, Memory Optimized, GPU instances
- From sandbox and dev/test to large-scale production

Fast Deployment, Full Automation

Faster time-to-value for engineering teams

- API first, Terraform-ready, full automation
- Native integration with Managed Kubernetes, Network services.

Built-in Sovereignty & Compliance

Full control over data location and infrastructure

- Clear data residency and governance
- Customers meet compliance and regulatory constraints by default through GDPR, ISO 27001, Healthcare certification, ...

Production-Ready by Default

Reliable infrastructure for business-critical workloads

- High availability option through multi-AZ deployment
- Integrated backup, snapshot and distant replication

Top use cases scenarios

Cloud-Native & Containerized Applications



Optimized foundation for Kubernetes workloads, microservices and CI/CD pipelines.

Seamless integration with OVHcloud Managed Kubernetes Services.



Scalable Web & API Platforms

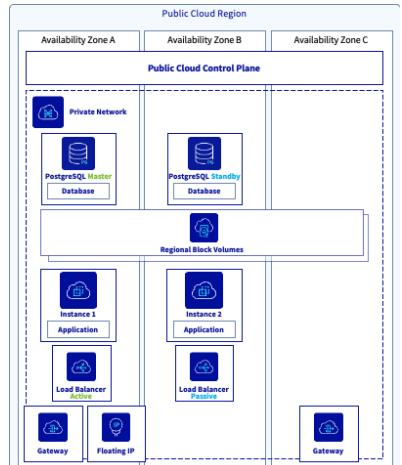
Run customer-facing applications with predictable performance and cost control. Ideal for SaaS, e-commerce, marketplaces and APIs requiring elasticity without vendor lock-in.



Data Processing & AI Inference

Efficient compute and GPU instances for data analytics, ML inference and batch processing, with controlled costs and European data residency.

Architecture - Technical Overview - Offer description



A modular compute layer integrated into the OVHcloud Public Cloud ecosystem

- On-demand virtual machines deployed in seconds
- Integrated networking (Private Network, Load Balancer, Public Gateway), storage (Block, Files & Object Storage) and security
- Designed to support single-AZ, multi-AZ and hybrid architectures

Customers can start small, scale fast and evolve their architecture without replatforming

OVHcloud Public Cloud Instances are designed for customers who want cloud agility, predictable performance and cost control – without compromising on sovereignty or long-term flexibility.

Why OVHcloud ? Our difference

Transparent & Predictable Cloud Economics

No hidden costs, no complex pricing models.
Customers retain long-term cost visibility, even at scale.

European Cloud by Design

A trusted alternative to hyperscalers for regulated industries and sovereignty-driven projects.

Infrastructure Control Without Complexity

High-performance infrastructure with cloud agility, without sacrificing control or predictability.



Cloud GPU

High-performance GPU infrastructure to accelerate AI

Key Capabilities

Production-Grade GPU Performance

NVIDIA GPUs (A10, A100, H100, H200...) with dedicated resources for predictable GPU performance for AI workloads

Cost-Efficient AI Infrastructure

Lower TCO for GPU workloads compared to hyperscalers

- Transparent hourly pricing
- No hidden networking or data egress penalties

Scalable from Experimentation to Production

- Single GPU to multi-GPU instances
- Scale vertically (GPU size) or horizontally (clusters)

Support for the Most Demanding AI Models

- High-memory GPUs and multi-GPU configuration
- NVLink-enabled architectures for large model parallelism

AI-Ready Cloud Ecosystem

- Faster time-to-value for AI and data teams
- Native integration with Kubernetes, Object Storage and CI/CD pipeline

Top use cases scenarios



Use Case #1 - AI Model Training & Fine-Tuning

Train and fine-tune large ML and deep learning models efficiently, with predictable performance and optimized cost.



Use Case #2 - AI Inference

Run real-time or batch inference workloads with low latency and controlled costs, integrated with cloud-native architectures.

Architecture - Technical Overview - Offer description

OVHcloud GPU enables organizations to build, train and deploy AI models faster.

OVHcloud GPU catalogue :

- H200
- H100
- A100
- L40s
- L4
- A10
- V100/V100S
- Quadro RTX 5000

GPU Type	Mem GB	FP8 Support	OVHCloud Avail	Min Nb of GPU for inference (LLM) - FP8					
				Small ~8B / 32k	Medium ~30B / 32k	Large ~70B / 128k	XL ~120B / 128k	Huge ~1TB / 128k	Giant ~671B / 128k
H200	141	Yes	* Coming	1	1	1	2	5	8
RTX 6000	96	Yes	* Coming	1	1	2	2	7	12
A100	80	No	Yes	1	1	3	4	14	14
H100	80	Yes	Yes	1	1	2	3	8	14
L40S	48	Yes	Yes	1	1	3	4	13	23
V100S	32	No	Yes	1	3	7	10	34	34
RTX 5090	16	Yes	No	2	3	9	12	38	67
L4	24	Yes	Yes	1	2	6	8	26	45
A10	24	No	Yes	2	4	9	13	46	45
V100	16	No	Yes	2	5	13	19	68	67

Why OVHcloud ? Our difference

Transparent GPU Pricing

GPU consumption remains understandable and forecastable.

European Sovereign AI Infrastructure

A trusted alternative for organizations requiring compliance, data control and strategic independence.

Performance Without Complexity

High-end GPUs delivered as a service, without managing physical infrastructure or vendor lock-in.



Block Storage

Block storage: flexible, secure, and always available for your critical needs.

Key Capabilities

Flexible

Our block storage solution lets you create volumes, ranging from 10 GiB to 12 TiB, to match your exact requirements. Add more storage when needed, without sacrificing strong performance or speed.

Easily interoperable

Cinder CSI driver compatibility enables plug-and-play Kubernetes Persistent Volumes, provision block storage for stateful apps without custom drivers or complex configurations.

Multi-zones

For Classic Block Storage volumes: you can choose between One-Zone, Multi-Zone, or Local Zone configuration with varying SLAs, from 99.9% to 99.99%.

Versatile storage classes

Classic (500+ IOPS, 3-AZ), High Speed (3k IOPS) or High Speed Gen2 (20k IOPS): tailored for analytics or transactional loads.

Budget friendly

Full price transparency, with no extra costs, and optimal performance from the first gigabyte.

Secured

Automatically backup your Block Storage volumes to Object storage. Protect data by encrypting it one the server side with SSE-OMK (OVHcloud Managed Keys).

Top use cases scenarios



Containerised applications

Deploy persistent storage for your containerised environments. Get a secure, resilient infrastructure, supported by triple data replication.



Critical applications

Block Storage offers high fault tolerance, redundancy, and performance for critical apps like high-transaction analytics and data processing.



Hosting relational databases

Relational databases, like MySQL or PostgreSQL, benefit from Block Storage's fast performance and direct access to data blocks.

Architecture - Technical Overview - Offer description

Volume Type	Bandwidth	Performance	Price (ex. VAT/GiB/hour)
Classic	-	From 500 IOPS	€0.000059
High Speed	-	Up to 3000 IOPS	€0.000119
High Speed Gen2	0.5 MB/s/GB, 512 MB/s max	30 IOPS/GB, up to 20k IOPS	€0.000119

Why OVHcloud ? Our difference

Predictable Pricing

Full transparency from GB1 (no minimum fees), vs. hyperscalers' tiered IOPS/throughput surcharges that spike costs.

Sovereignty & compliance

A trusted alternative for organizations requiring compliance, data control and strategic independence. Also, Block Storage meets ISO (27001, 27017, 27018, 27701) standards.

Escape vendor lock-in

OVHcloud Block Storage uses open standards for instant portability e.g Cinder CSI (no proprietary traps like no AWS EBS-specific volumes or Azure Disk formats to re-architect).



File Storage

BETA

Fully managed file storage volumes, built for containers and instances

Key Capabilities

Budget friendly

Instantly adjust your share size at the GiB level instantly, pay only for what you use.

Secured

Encryption with OMK, SLAs of 99.9% and Vrack compatibility. Coming soon: recurrent, automated, distant backups at your fingertips.

Available where you need

Create and manage shares right where your containers or instances are: File Storage is available in all our public cloud regions.

Modular

Linear performance that matches your application's needs (IOPS & MB/s): up to 16 000 IOPS and 128 MB/s.

Integrated to your workflows

Create and manage shares from the OVHcloud API V6, OVHcloud control panel, Manilla CSI Driver, Openstack API or CLI.

Fully managed

No maintenance: File Storage is fully packaged and integrates instantly into your infrastructure.

Top use cases scenarios



Horizontal workload distribution

Distribute workloads across multiple web server instances or containers to balance traffic during peak hours, such as an e-commerce site handling Black Friday sales.



Dynamic resizing

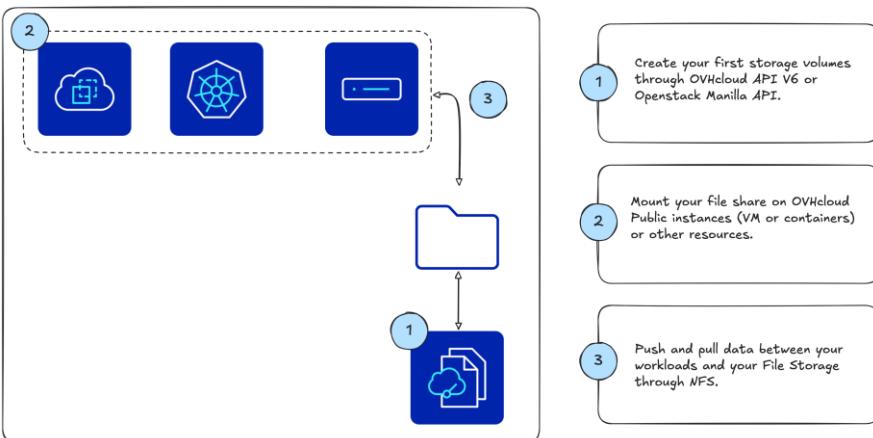
Automatically scale the number of Kubernetes pods or worker nodes in a cluster based on CPU usage, like increasing replicas for a microservices app during a product launch.



Shared asset catalogue

Centralize storage for website assets like product images, promotional videos, and CSS files, enabling a content team to update media across multiple front-end servers simultaneously.

Architecture - Technical Overview - Offer description



Detailed offer (Standard 1AZ)

Pricing	150 GiB to 10 TiB with hourly billing. NB: subject to change after GA (mid-March 2026)
Performance	Up to 16 000 IOPS and 128 MB/s.
Storage capacity	150 GiB to 10 TiB

Why OVHcloud ? Our difference

Competitive performance

Better than most (GCP & Azure for example) between 150GiB and 1iTB

Market leading pricing

5% cheaper than AWS, with a full pay as you go model.

Data sovereignty

No vendor lock-in, no dependency to an external solution (product based on Manilla Openstack). Your data remains yours.



Private Network (vRack)

Secure, layer 2 private networking to connect and scale your cloud architectures without public exposure

Key Capabilities

Private Networks Extended Across Locations

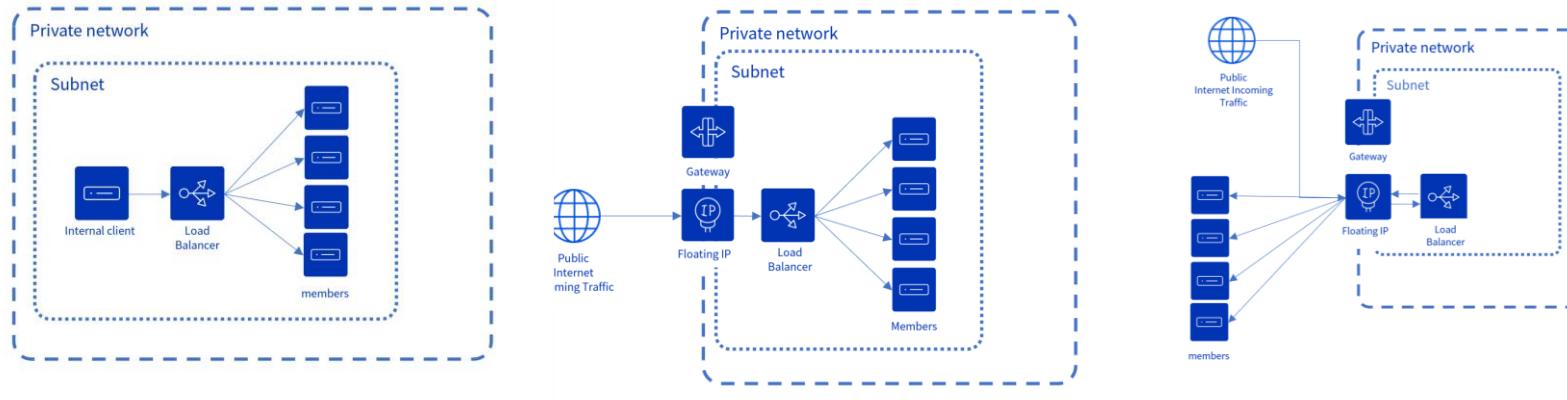
- Seamless private communication between resources in multiple datacenters
- Layer 2 private networking across OVHcloud datacenters worldwide

Private Connectivity Across OVHcloud Services

- Native extension to Dedicated Servers and Private Cloud

Strong Network Isolation by Design

- Secure segmentation of workloads and environments
- Up to 4,000 isolated Layer 2 private networks (VLANs)



Top use cases scenarios



Use Case #1 - Secure Segmentation of Workloads

Isolate control plane, data plane, environments (dev / prod) or tenants to improve security posture and operational resilience.



Use Case #2 - Multi-Datacenter Application Architectures

Deploy applications across several datacenters while keeping all internal communications private, secure and isolated from the public internet.



Use Case #3 - Hybrid Cloud with Dedicated & Private Cloud

Extend Public Cloud networks to Dedicated Servers or Private Cloud, enabling hybrid deployments with private, low-latency connectivity.

Why OVHcloud ? Our difference

True Layer 2 Networking

Unlike traditional cloud networking, vRack enables real Layer 2 extension across infrastructures and locations.

Hybrid by Design

Public Cloud, Dedicated Servers and Private Cloud connected through the same private network fabric.

Secure Networking Without Complexity

No VPN sprawl, no public exposure — private networking delivered as a managed service.



Object Storage

One storage class per data. One deployment per project.

Key Capabilities

Match storage to real data usage

Multiple storage classes, each dataset gets the right performance and cost profile, from frequently accessed data to deep archives.

Deploy where your data belongs

Run Object Storage in 1-AZ, 3-AZ, or Local Zones, and go further with **SecNumCloud-ready** deployments to meet availability, latency, and data residency requirements

Built-in durability and availability

- All data protected with industry-leading 99.99999999% durability,
- Data availability adapts to your deployment model: 99.9% in 1-AZ and Local Zones, and 99.99% in 3-AZ regions.

Control costs with predictable pricing

Pay only for what you store and, depending on the class, what you retrieve. No minimum storage commitments. No API call fees. No ingress or egress charges, just transparent pricing you can forecast.

Automate data lifecycle, reduce waste

Move data seamlessly from frequently to rarely accessed or archived classes using lifecycle policies, within the same region and at no extra cost, while keeping the option to replicate data across regions for colder tiers.

Secure, compliant and sovereign by design

native security features such as Object Lock, replication, encryption at rest (SSE-OMK), and versioning, backed by ISO certifications and deployment options designed for regulatory compliance and strategic autonomy.

Architecture – Technical Overview – Offer description

3-AZ Regions – Availability SLA: 99,99%

	Storage	Retrieval	Minimum billed
Standard	~14-11€/TiB	Included	1 hour
Inf. Access	~9,5€/TiB	0,004€/GiB	30 days
Active Archive	~4,5€/TiB	0,018€/GiB	90 days
Cold Archive	~1,7€/TiB	0,009€/GiB	180 days

1-AZ Regions – Availability SLA: 99,9%

	Storage	Retrieval	Retention
High Performance	~18€/TiB	Included	1 hour
Standard	~7€/TiB	Included	1 hour
Inf. Access	~4€/TiB	0,004€/GiB	30 days

Local Zones – Availability SLA: 99,9%

	Storage	Retrieval	Retention
Standard	~7€/TiB	Included	0 days



Top use cases scenarios



Cloud applications and active data (Hot data)

Modern applications, APIs, and digital services generate data that is accessed continuously, user content, application assets, logs, and transactional data.



Backups and business data retention (Warm data)

As data ages, access frequency drops. Operational backups, historical business records, and compliance datasets are still required, but only occasionally.



Long-term archives and regulatory preservation (Cold data)

Some data must be retained for years or decades, even if it is rarely, or never, accessed. Regulatory archives, historical records, and disaster recovery copies fall into this category.

Why OVHcloud ? Our difference

True choice across classes and deployments

One Object Storage platform with multiple storage classes and multiple deployment models – from 1-AZ to 3-AZ, Local Zones, SecNumCloud, and On-Prem. You choose what fits your data, not what's imposed by the platform.

Pricing you can actually predict

No minimum storage. No API call fees. No ingress or egress charges. You pay for the data you store, and, depending on the class, the data you retrieve. Simple, transparent pricing designed for cost control, not surprises.

Sovereign, secure, and built for compliance

Industry-leading durability, ISO-certified infrastructure, native security features, and deployment options built for data residency and strategic autonomy, including SecNumCloud and On-Prem Cloud Platform.



Public Gateway

Securely expose private cloud workloads to the public internet — without breaking network isolation

Key Capabilities

Controlled Internet Access for Private Networks

Secure outbound and inbound connectivity for private resources

- Public Gateway provides internet access to instances hosted in Private Networks
- No need to assign public IPs to each instance

Reduced Attack Surface by Design

- Improved security posture for cloud workloads
- Private instances remain unreachable directly from the internet
- Centralized internet exposure through a controlled gateway

Centralized Traffic Control & Visibility

- Better control over network flows with a single entry and exit point for private networks

Simplified Network Architecture

- Faster and simpler cloud network designs. Teams deploy secure architectures without networking expertise

Top use cases scenarios



Use Case #1 - Enable internet access to private instances

Allow private instances to access external services (updates, APIs, repositories) while remaining fully isolated from direct internet exposure.



Use Case #2 - Exposing a service on an instance

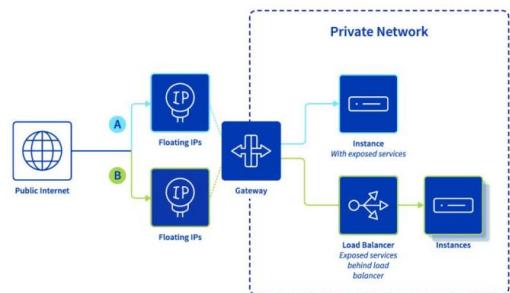
Expose selected services to the public internet in a controlled way, without assigning public IPs to backend resources



Use Case #3 - Exposing services behind Load Balancer

Load Balancer enhances security and enables SSL encryption, and can be seamlessly updated through the Gateway-hosted floating IP

Architecture - Technical Overview - Offer description



Gateway	S	M	L	XL	2XL
Bandwidth capacity	200 Mbps	500 Mbps	2Gbps	4Gbps	8Gbps
Outbound internet connectivity (through SNAT)			Yes		
Inbound internet connectivity (through Floating IP)			Yes		

OVHcloud Public Gateway enables secure and controlled internet connectivity for public cloud architectures, combining simplicity, security and scalability.

Why OVHcloud ? Our difference

Security Networking by Default

Private resources stay private — internet access is centralized and controlled.

Simplicity Over Custom Networking

No complex NAT rules, no self-managed gateways — fully integrated and managed.

Designed for Cloud & Hybrid Architectures

Works seamlessly with Private Network (vRack) and hybrid deployments across OVHcloud.



Load Balancer

Deliver highly available, scalable and resilient applications with managed load balancing.

Key Capabilities

Ensure application continuity and resilience

- Traffic distributed across multiple backend instances
- Automatic failover in case of instance or service failure

Secure Entry Point for Public & Private Workloads

- Acts as a secure gateway between public internet and private networks
- Backend nodes remain isolated within Private Network (vRack)

Flexible Deployment Models

- Deployable per region, close to end users
- Public or private Load Balancer configurations

Simplified Traffic Management & Operations

- Managed service with no infrastructure to maintain
- Multiple health check protocols (TCP, HTTP, etc.)

Cloud-Native & Ecosystem Integration

- Native support for Octavia API
- Compatible with Terraform and automation tools
- Works with Public Cloud instances, Kubernetes, Bare Metal and Private Cloud backends

Top use cases scenarios



Use Case #1 - Highly Available Web & API Applications

Distribute traffic across multiple backend instances to ensure uptime, performance and resilience for customer-facing services.



Use Case #2 - Secure Frontend for Private Architectures

Expose applications through a single, secured entry point while keeping backend services isolated in Private Networks.



Use Case #3 - Hybrid & Multi-Platform Load Balancing

Use a single Load Balancer to distribute traffic across Public Cloud instances, Kubernetes workloads, Hosted Private Cloud or Bare Metal servers.

Why OVHcloud ? Our difference

Offer Description - schema

Load Balancer Size	Bandwidth	Applicable to the following listeners:					
		All	HTTP/TCP/HTTPS*		HTTP/ HTTPS	TERMINATED_HTTPS*	
Concurrent active session	Session created per second	Requests per second	SSL/TLS session created per second	Requests per second	Packet per second		
Size S	200 Mbs (up/down)	10k	8k	10k	250	5k	10k
Size M	500 Mbs (up/down)	20k	10k	20k	500	10k	20k
Size L	2 Gbps (up/down)	20k	10k	40k	1,000	20k	40k
Size XL	4 Gbps (up/down)	20k	10k	80k	2,000	40k	80k

Security & Isolation Built In

Backend workloads stay private; exposure is centralized and controlled.

Open, Cloud-Native Load Balancing

Based on open technologies (Octavia), fully integrated into the Public Cloud ecosystem.

One Load Balancer, Multiple Backends

Public Cloud instances, Kubernetes, Bare Metal and Private Cloud — all behind the same entry point.



Floating IP

Ensure application continuity with a static, reassignable public entry point for cloud workloads.

Key Capabilities

High Availability Through IP Reassignment

Maintain service availability during failures or maintenance

- Static public IP that can be reassigned dynamically
- Instant failover between instances or services

Predictable & Reserved Public Entry Point

Stable network identity for critical applications

- Reserve public IPs early in the project lifecycle

Customers plan architectures and integrations without network uncertainty

Secure Exposure of Private Workloads

Expose services while preserving network isolation

- Floating IP attached via Gateway
- Instances remain inside Private Networks

Customers expose only what is needed, without compromising security

Zero-Configuration Failover

Simplified operations and maintenance

- Reassignment handled at the network controller level
- No OS or application reconfiguration required

Teams perform upgrades, replacements and incident recovery faster

Top use cases scenarios



Use Case #1 - Transparent Maintenance & Server Replacement

Move a public service from one instance to another instantly, without changing IP address or impacting users.



Use Case #2 - Secure Exposure of a Private Instance

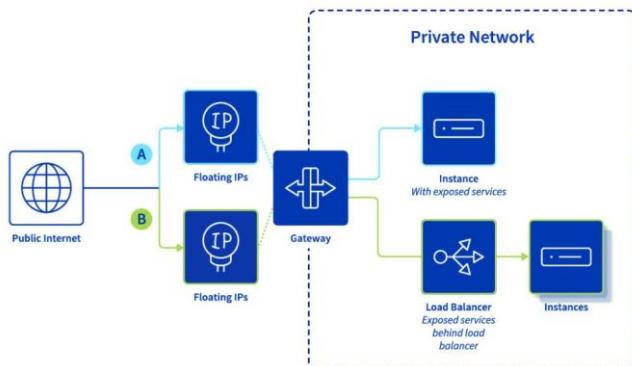
Expose a service hosted in a Private Network to the internet using a Floating IP, while keeping the instance isolated from direct access.



Use Case #3 - Highly Available Application Frontend

Attach a Floating IP to a Load Balancer to create a single, secure and stable entry point for multi-instance applications.

Why OVHcloud ? Our difference



Combine a Floating IP, a Gateway and a Load Balancer in your architecture to achieve suitable rules for network accessibility and the level of security that you require.

True Network-Level Mobility

IP reassignment happens instantly, without touching operating systems or applications.

Designed for Secure Cloud Architectures

Floating IPs integrate natively with Gateway and Private Network for controlled exposure.

Automation-Ready by Design

Full API and IaC support for scalable, repeatable operations.



Managed Kubernetes Service

Benefits of a fully Managed Kubernetes Service platform , highly available , secure and resilient for every containerized workloads size.

Key Capabilities

Highly available and resilient

- Cross zonal control plane to guarantee continuity of service
- Dedicated ETCD to improve performance, stability and scalability of cluster's data plane.

Storage orchestration

- Automatically mount the storage system of your choice, whether from local
- Storage block storage or share storage such like FSaaS or NFS.

Automated rollout & rollback

- Managed progressive changes rolls our to your application or its configuration, for a full continuity of service.

Self healing

- Restarts containers that fail, replaces and reschedules containers when nodes die.

Service discovery

- No need to modify your application to use an unfamiliar service discovery mechanism.

Top use cases scenarios

Simplify and accelerate application deployment

Forget about infrastructure dependencies and scale micro-services at high speed. Embrace the modern application world by stacking containers and accelerate time to market innovation.

Optimize infrastructure cost ownership

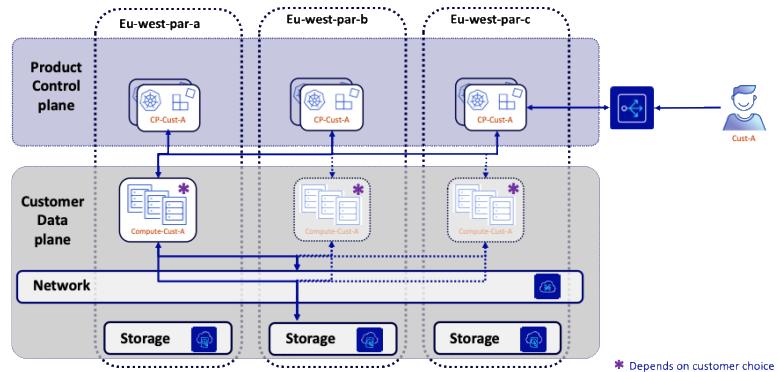
Densify your data plane utilization on clusters to have the best return on cloud investment and avoiding to handle multiple kubernetes cluster to manage.

Failure-proof your cloud native application

Deploy highly resilient cloud native application on our multi zonal region. Kubernetes control plane is deployed regionally and data plane is multi zonal.

Why OVHcloud ? Our difference

MKS Standard – Multi zonal deployment model



	MKS Free	MKS Standard new
Control plane	Managed	Managed & Cross-AZ resilient
Availability	99,5% SLO	99,99% SLA
ETCD	400MB	Dedicated - 8GB
Max cluster size	Up to 100 nodes	Up to 500 nodes
Regional disponibility	Single-regions	Paris & Milan (3AZ) Mumbai, Roubaix (1AZ)

Highly versatile

- Market standards certified Kubernetes based, with all features
- Cloud Native Computing Foundation (CNCF) certified
- Flexible infrastructure , works with any product or system without restrictions.

Scalability

Enables on demand dynamic scaling . Only use the needed resources. Improves the resilience of your applications.

Open ecosystem

Large catalog of instances for all use cases: intensive computing, AI (GPU), big data.

As well as our Private Registry, load balancing, storage (block, object) solutions...



Managed Private Registry

Store, manage and access your container images and Helm charts by offering enterprise-grade features and predictable pricing.

Ecosystem integration

Full interoperability, facilitating streamlined CI/CD pipelines and enhancing development workflows.

Image & AI artefact security enforcement

Robust security features, with vulnerability scanning and access control, container images are stored securely, protecting against unauthorized access and potential threats

Ease of use

Simplifies operations, freeing up internal resources and eliminating the need for in-house expertise.

Key Capabilities

Cost efficient

Cost efficiency and predictable pricing model, with multiple pay as you go plans

Highly available and resilient

M or L plans come with a cross zonal private registry to guarantee continuity of service and improved performance, stability and scalability of images' catalog.

Top use cases scenarios

Easily set up your CI/CD

Facilitate the entire integration with your CI/CD via API. Accelerate your software releases, from development to production, while OVHcloud manages your private container registry, ensuring the availability of images and a smooth, stress-free application deployment process.

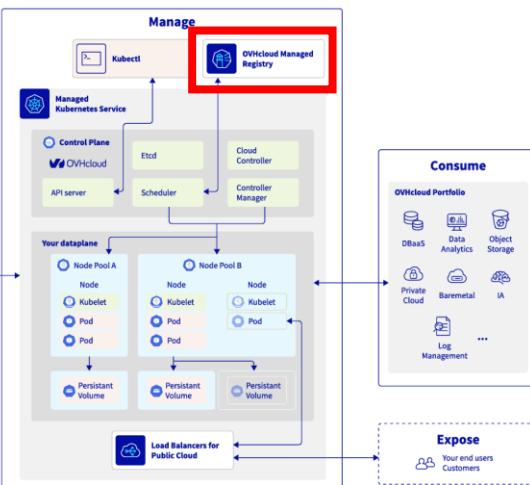
Improve team collaboration

Enable the efficient and secure delivery of images, everywhere you operate. Depending on their rights, developers access approved images, reducing the risk of obsolescence. Managed Private Registry also enables multiple connections to your container registry to take place simultaneously, further enhancing the agility of your teams.

Streamline management of image or AI artefacts

Manage your private storage configuration via a single, dedicated Harbor interface, or through the Harbor API. This allows you to ensure the integrity of your images, by synchronising multiple container registries and automating the removal of unused images.

Architecture - Technical Overview - Offer description



	S	M	L			
Price	Mono AZ	Multi AZ	Mono AZ	Multi AZ	Mono AZ	Multi AZ
Storage	200 GB		600 GB		5TB	
HTTP concurrent	15 (3x5)*		45 (9x5)*		90 (18x5)*	
Durability	100 %		100 %		100 %	
Features set	Basics		basic + Vulnerability scanning			
SLA	99,9% core & registry – 1AZ & 3AZ		99,95% core & registry – 99,99% in 3AZ 99,9% on other components – 99,99% in 3AZ			
Price (/m)	16€	19,053€	40€	43,216€	190€	208,78€

Why OVHcloud ? Our difference

Full interoperability

MPR is built on open-source solutions, such as Docker and the CNCF Harbor project, to guarantee interoperability. It supports all OCI (Open Container Initiative) images and Helm charts and can work with any containerisation platform or orchestration tool, such as the Kubernetes ecosystem.

Maximum security

Secure access for your teams with Role-Based Access Control (RBAC) and ensure the integrity of your images' sources with Content Trust.

Predictable pricing

MPR offers a range of pay-as-you-go plans to suit your specific needs. These include the amount of stored data, concurrent connections to your private container registry, and SLAs.



Managed Rancher Service

Managed Rancher Service simplifies Kubernetes clusters management to ensure consistent operations, workload management, and enterprise-grade security.

Key Capabilities

Multi cluster management

- Eliminates the complexity of managing disparate containerized applications.
- Provides a single, unified view of container ecosystem with centralized control
- Improved monitoring, and streamlined operations.

Platform elasticity

- Eliminates the complexity of predicting and provisioning underlying infrastructure.
- Define the desired state of your application, and the platform will automatically manage and scale the compute resources

Portability. 100% opensource. 0% lock-in

- Freedom to deploy workloads regardless of chosen, cloud provider
- Access to the vibrant Rancher community and its extensive ecosystem of applications

Workflow automation with gitops tools

Leverage the combined power of DevOps methodologies, Agile development practices, and leading GitOps tools for streamlined continuous delivery, ensuring your application remains current and adaptable to changing requirements.

Top use cases scenarios



Modernizing and scaling infrastructure efficiently

Maintaining and deploying new services without handling infrastructure and dependencies complexity.



Multi-cluster and multi-cloud management

Address enterprise level Kubernetes cluster provisioning, upgrades, user and policy management anywhere (on premise, on virtualized environment or in baremetal) in any cloud providers.



Edge computing with K3S

Run containerized apps on edge (K3S) and keep the same level of administration, monitoring, scaling and updating if it was on traditional cloud infrastructure

Why OVHcloud ? Our difference

Versatility

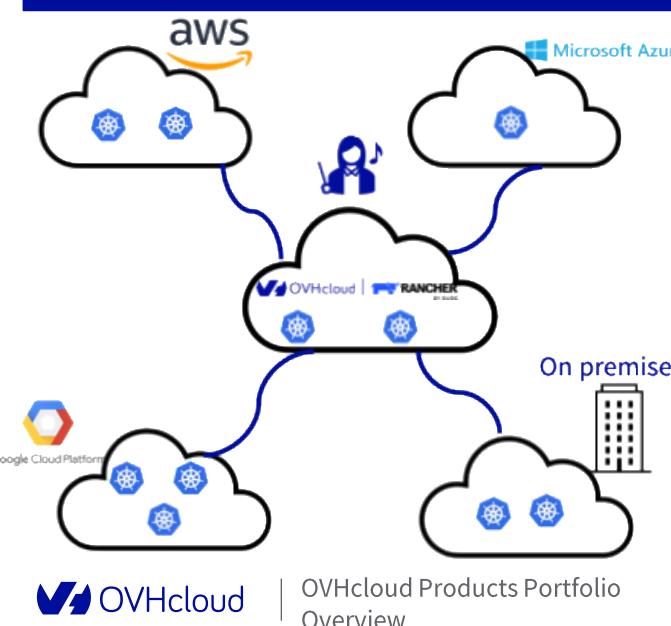
No application lockin. Any kubernetes and containers distro are working. Compatible with any CSP even on vendor managed kubernetes (EKS,GKE,ACK..) You can instantly move your workloads anywhere, anytime. Capacity to self manage cluster on top of any IaaS. Fully opensource and un-opinionated

Scalability

Until 1000 worked node allowed. Fast provisioning of cluster. Integrated with OVHcloud 80+ services across different deployment models (public and private). Rich catalogue of integrated opensource and industry-leading Kubernetes services.

Ease of use and maintain

Fully managed control plane and friendly UI to manage your cluster. Designed with Suse and powered by OVHcloud. With both teams committed to support your cloud native projects. committed to support your cloud native projects.



Limitations

- Not compatible with vRack
- Not compatible MKS driver for cluster import
- OVHcloud drivers are not open sourced
- Not possible to import existing Rancher
- No custom DNS configuration
- Not available for the US (FY26)
- Service based in France region, but with a global reach. It can deploy Kubernetes clusters on any infrastructure everywhere (OVHcloud or not).



Managed Databases

Databases as a Service (DBaaS) is a fully managed cloud solution that allows organisations to deploy, manage, and scale relational and non-relational databases without the complexities of infrastructure management

Key Capabilities

Automated Management

Our customer no longer needs to worry about database setup, maintenance, backups, and updates.

Fast, Secure & High Availability

Our customer can rely on a secure setup, with SLAs, fast connections and automated backups or failovers in the event of an outage. Safeguarding data integrity and sovereignty is a key topic.

Scalability for Large Volumes of Data

Our customer can set up a database in a few clicks and then dynamically scale its resources vertically and horizontally based on workload demands without downtime or manual intervention

Total Cost of Ownership (TCO) reduction

Administration of DBs is difficult and DB administrators are very specific and costly jobs. With DBaaS, our customers can focus on their core product and business

Top use cases scenarios



Managing High Volumes of Structured Data

SQL engines provide high-performance, reliable transactional data storage with ACID compliance, ensuring data integrity for all transactions.



Multimedia data organization & analysis

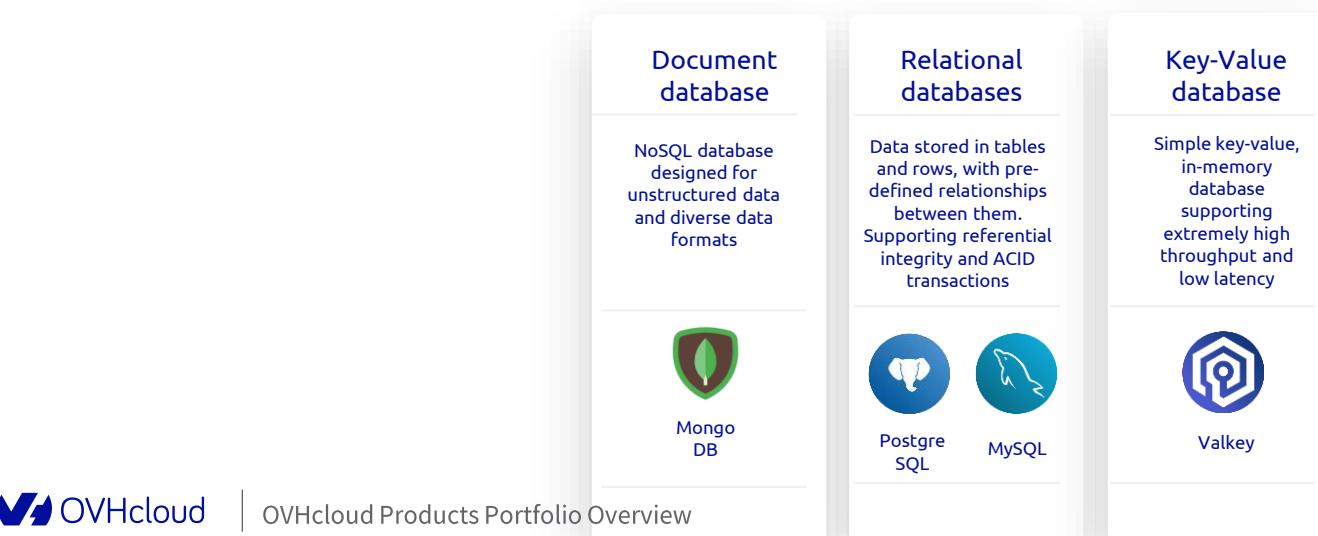
When it is crucial to organize multiple types of data and deliver it fast to end users in a personalized way, NoSQL engines will provide higher performance and scalability.



Managing Real-Time Data and Caching

In-memory databases (IMDBs) offer ultra-fast data access by storing data directly in RAM instead of traditional disk-based storage. This architecture enables low-latency real-time processing and immediate data retrieval.

Architecture - Technical Overview - Offer description



Why OVHcloud ? Our difference

OVHcloud Managed Databases are resilient & hassle-free

OVHcloud provides a full and clear portfolio of managed databases which will reduce your time-to-market and ensure your service is always up and running and adapts to evolving business needs.

OVHcloud Managed Databases provide high value for a predictable price tag

No hidden fees and costs are affordable and predictable, including IOPS, traffic and backups

OVHcloud Managed Databases safeguard your data

Not only is the offering based on open-source technology which allows freedom to move and innovate, but it is also hosted in infrastructure compliant and protected from extra-territorial laws.



Managed Analytics

Cloud Analytics is a fully managed cloud solution that allows businesses to collect, process, and analyze large volumes of data without having to manage the underlying infrastructure.

Key Capabilities

Save Money

Managed Analytics are a cost-effective way to compress, store large datasets and analyze them as you go along. You can start small and expand, reducing costs and required skillset.

Fast, Secure & High Availability

Our customers can rely on a secure setup, with SLAs, fast connections and automated backups or failovers in the event of an outage. Safeguarding data integrity and sovereignty is a key topic.

Effortless Scalability & Resilience

Launched within minutes, Managed Analytics are adapted to peaks, to unexpected demands as well as outages as they scale with usage. Backups and failover options provide added security.

Fully Managed Experience

OVHcloud handles provisioning, scaling, patching, and backups, allowing teams to focus on rapid deployment of observability stacks without the complexity of managing clusters manually.

Top use cases scenarios



Managing High Volumes of Data in Real Time

ClickHouse enables real-time analytics at a petabyte scale with exceptional query speed and storage efficiency.



Managing Real-Time Data Pipelines

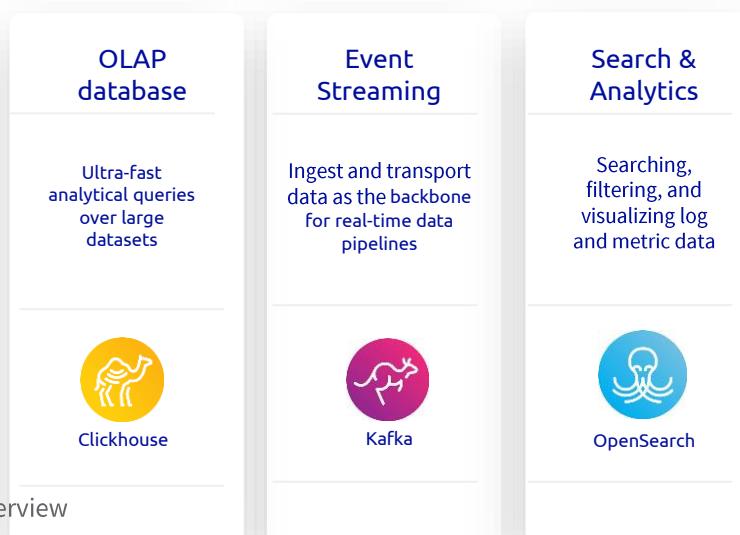
Kafka provides a reliable, scalable, and high-performance backbone for event-driven architectures, enabling companies to build resilient, decoupled systems.



Managing Logs, Metrics, and Search at Scale

OpenSearch enables fast, scalable, and flexible analysis of both structured and unstructured data, empowering operations and security teams to troubleshoot issues and monitor system health.

Architecture - Technical Overview - Offer description



Why OVHcloud ? Our difference

OVHcloud Managed Analytics are built for scale

OVHcloud provides a full and clear portfolio of managed analytics engines which will reduce your time-to-market and ensure your service is always up and running and adapts to evolving business needs.

OVHcloud Managed Analytics are managed so you can fully focus

OVHcloud handles provisioning, upgrades, scaling, failover, and monitoring—so your teams can focus on product, data, and customer value instead of system babysitting.

OVHcloud Managed Analytics as a guarantor of transparency

Not only is the offering based on open-source technology which allows freedom to move and innovate, but it is also hosted in infrastructure compliant and protected from extra-territorial laws.



Data Platform

Speed-up and simplify data & analytics projects with a unified platform

Key Capabilities

Simplify data ops

Thanks to a fully managed solution, your data organisation and operations become much easier regardless of the volume.

Stick to open standards

Reversible and interoperable with open technologies: Apache Parquet, Iceberg, Spark, SuperSet, Jupyter Notebooks & Trino.

Deploy rich dashboards, faster than ever

Configure and feed precise dashboards in minutes, enabling proper visualisation and monitoring of data for better decisions

Benefit from a flexible, predictable pricing model

Allying pay-as-you-go with the cloud's agility and full transparency

Top use cases scenarios



Retail & Ecommerce

Securing proper Customer Segmentation based on purchasing behaviour or launching Personalized Recommendations based on Analyzing customer browsing analysis.



Financial Services & Banking

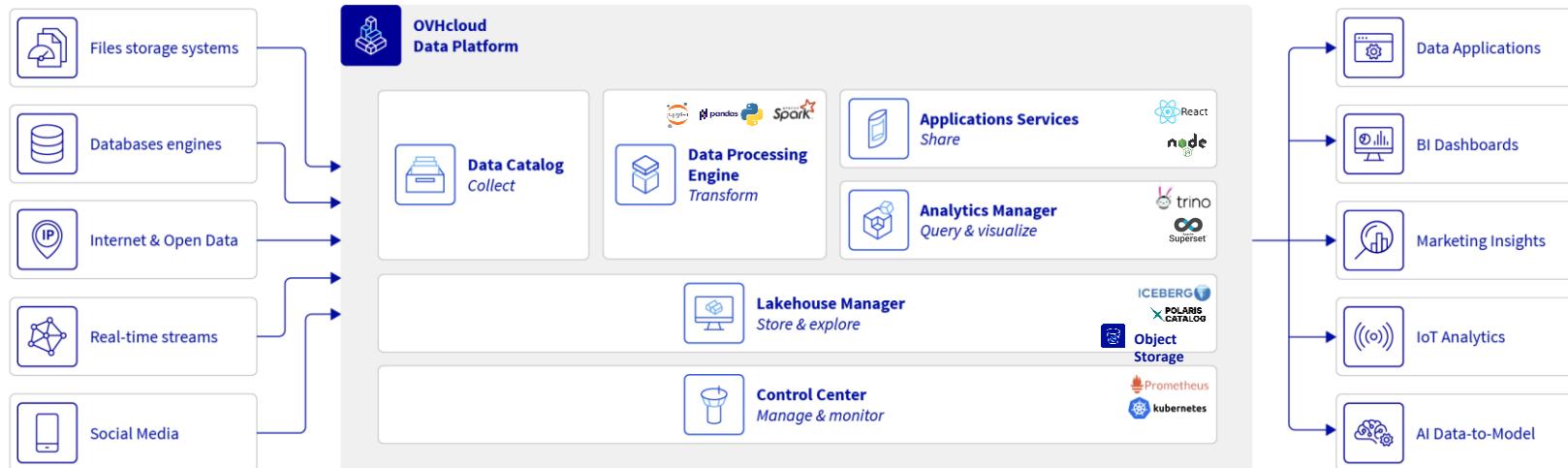
Enabling Fraud Detection by using machine learning to identify suspicious transactions or pinpoint Credit Scoring by evaluating transactional history.



Healthcare & Life Science

Identifying patients at high risk for diseases using AI-driven models. Or analyzing patient data to accelerate drug discovery and trial effectiveness.

Architecture - Technical Overview - Offer description



Why OVHcloud ? Our difference

A complete data & analytics platform

Unlike hyperscalers and BI software data stacks, Data Platform offers a unified solution with all features natively available via a single low-code capable interface.

Simplicity, fully automated PaaS

Unlike hyperscalers, Data Platform offers a fully integrated, managed and secure service from physical hardware all the way to the software. No infra expertise needed.

Transparent & simple pricing

Unlike Snowflake or MS Fabric, Data platform offers a free tier and a pay-as-you-go, including a serverless lakehouse, or alternatively a transparent reserved capacity business model.



AI Solutions – AI Notebooks

Developers & Data scientists, accelerate your project and model releases with fully managed notebooks in seconds

Key Capabilities

Most popular preconfigured AI frameworks

Ready to use widely-known frameworks : TensorFlow, PyTorch, Scikit-learn, MXNet and Hugging Face.

Managed Notebook Environments

Launch Jupyter-based notebooks instantly without installing or maintaining local tooling or infrastructure.

Seamless Data access & management

Connect directly to OVHcloud Object Storage, work efficiently with large dataset.

Collaboration & reproducibility

Share notebooks, environments, and experiments across teams to improve collaboration and reproducibility.

On demand Compute resources

Select CPU or GPU-backed resources and adjust capacity as workloads evolve from exploration to experimentation.

Integrated security & IAM

Protect model endpoints with identity management, access controls, and network isolation aligned with enterprise security standards.

Top use cases scenarios



Data exploration, analysis and visualisation

Explore, clean, and analyze datasets interactively to prepare data for machine learning and AI workloads.



Model Prototyping and Experimentation

Rapidly test algorithms, architectures, and parameters before scaling training jobs.



AI Proof of Concept

Validate AI ideas quickly and collaboratively before committing to large-scale training or deployment.

Architecture – Technical Overview – Offer description

Node	Memory	vCore	Ephemeral local storage	Public Network	Price Ex. VAT/hour
ai1-1-CPU	4 Go	1	40 Go	500 Mbit/s	0,03 €

Node	Memory	vCore	GPU	Ephemeral local storage	Public Network	Price Ex. VAT/hour
a10-1-gpu	40 Go	28	NVIDIA Ampere A10 24 Go 31.2 TFLOPS	1536 Go	5 Gbit/s	0,90 €
a100-1-gpu	160 Go	13	NVIDIA Ampere A100 80 Go 19.5 TFLOPS	80 Go	1.5 Gbit/s	3 €
ai1-1-GPU	40 Go	13	NVIDIA Tesla V100S 32 Go 16.4 TFLOPS	750 Go	1.5 Gbit/s	1,93 €
ai1-le-1-GPU	40 Go	13	NVIDIA Tesla V100S 32 Go 16.4 TFLOPS	500 Go	1.5 Gbit/s	0,90 €
h100-1-gpu	350 Go	28	NVIDIA Hopper H100 80 Go 51 TFLOPS	3072 Go	5 Gbit/s	3,10 €
l4-1-gpu	80 Go	20	NVIDIA Ada Lovelace L4 24 Go 30.3 TFLOPS	500 Go	5 Gbit/s	0,83 €
l40s-1-gpu	80 Go	13	NVIDIA Ada Lovelace L40S 48 Go 91.6 TFLOPS	1536 Go	5 Gbit/s	1,55 €

Why OVHcloud ? Our difference

Part of an End-to-End Platform

AI Notebooks integrates natively with AI Training and AI Deploy, enabling a smooth transition from experimentation to production.

Open and Flexible Environments

Use standard notebooks, frameworks, and libraries without proprietary constraints or vendor lock-in.

Enterprise-Grade Cloud Foundations

Run notebooks on secure, compliant cloud infrastructure with strong data control and governance.



AI Solutions – AI Training

Train faster. Scale confidently. Control your costs

Key Capabilities

GPU accelerated training

Access to high-performance GPUs optimized for AI and deep learning workloads.

Resources usage optimization

Managed by our solution, with transparent minute billing.

Cost Control

Transparent pricing and the ability to scale resources up or down on demand.

Distributed training

Native support for multi-node and multi-GPU training to reduce training time.

Integrated security & IAM

Protect model endpoints with identity management, access controls, and network isolation aligned with enterprise security standards

Top use cases scenarios



Large scale Model Training

Train foundation models, large language models, or computer vision models on GPU-enabled infrastructure.



Model Fine-Tuning and Optimization

Adapt pre-trained or open-source models to proprietary data and specific business use cases.



AI Research and Experimentation

Run controlled experiments to evaluate architectures, parameters, and datasets before production deployment.

Architecture – Technical Overview – Offer description

Node	Memory	vCore	Ephemeral local storage	Public Network	Price ex. VAT/hour
ai1-1-CPU	4 Go	1	40 Go	500 Mbit/s	0,03 €

Node	Memory	vCore	GPU	Ephemeral local storage	Public Network	Price Ex. VAT/hour
a10-1-gpu	40 Go	28	NVIDIA Ampere A10 24 Go 31.2 TFLOPS	1536 Go	5 Gbit/s	0,90 €
a100-1-gpu	160 Go	13	NVIDIA Ampere A100 80 Go 19.5 TFLOPS	80 Go	1.5 Gbit/s	3 €
ai1-1-GPU	40 Go	13	NVIDIA Tesla V100S 32 Go 16.4 TFLOPS	750 Go	1.5 Gbit/s	1,93 €
ai1-le-1-GPU	40 Go	13	NVIDIA Tesla V100S 32 Go 16.4 TFLOPS	500 Go	1.5 Gbit/s	0,90 €
h100-1-gpu	350 Go	28	NVIDIA Hopper H100 80 Go 51 TFLOPS	3072 Go	5 Gbit/s	3,10 €
l4-1-gpu	80 Go	20	NVIDIA Ada Lovelace L4 24 Go 30.3 TFLOPS	500 Go	5 Gbit/s	0,83 €
l40s-1-gpu	80 Go	13	NVIDIA Ada Lovelace L40S 48 Go 91.6 TFLOPS	1536 Go	5 Gbit/s	1,55 €

Why OVHcloud ? Our difference

Part of an End-to-End Platform

AI Training integrates natively with AI Notebooks and AI Deploy, enabling a smooth transition from experimentation to production.

Data Sovereignty by Design

Train and fine-tune AI models on cloud infrastructure with full control over data location and compliance

Open, Framework-First Platform

Run industry-standard AI frameworks and your own containers with no proprietary lock-in, enabling full portability across cloud, on-prem, and hybrid environments.



AI Solutions – AI Deploy

From model to product, without friction

Key Capabilities

Managed model deployment solution

Serverless solution that deploys AI models as production-ready services without managing servers, orchestration, or runtime infrastructure.

Scalable inference

Automatically scale inference workloads up or down based on demand to ensure consistent performance and optimized resource usage.

Cost Control

Transparent pricing and the ability to scale resources up or down on demand.

Integrated security & IAM

Protect model endpoints with identity management, access controls, and network isolation aligned with enterprise security standards

Top use cases scenarios



Application AI Inference

Embed AI capabilities such as text generation, image analysis, or recommendations directly into customer-facing or internal applications.



Enterprise AI APIs

Expose models as internal or external APIs to standardize AI consumption across teams and business units.



Operational Automation

Deploy models that automate decision-making, classification, or prediction tasks within business processes.

Architecture – Technical Overview – Offer description

Node	Memory	vCore	Ephemeral local storage	Public Network	Price ex. VAT/hour
ai1-1-CPU	4 Go	1	40 Go	500 Mbit/s	0,03 €

Node	Memory	vCore	GPU	Ephemeral local storage	Public Network	Price Ex. VAT/hour
a10-1-gpu	40 Go	28	NVIDIA Ampere A10 24 Go 31.2 TFLOPS	1536 Go	5 Gbit/s	0,90 €
a100-1-gpu	160 Go	13	NVIDIA Ampere A100 80 Go 19.5 TFLOPS	80 Go	1.5 Gbit/s	3 €
ai1-1-GPU	40 Go	13	NVIDIA Tesla V100S 32 Go 16.4 TFLOPS	750 Go	1.5 Gbit/s	1,93 €
ai1-le-1-GPU	40 Go	13	NVIDIA Tesla V100S 32 Go 16.4 TFLOPS	500 Go	1.5 Gbit/s	0,90 €
h100-1-gpu	350 Go	28	NVIDIA Hopper H100 80 Go 51 TFLOPS	3072 Go	5 Gbit/s	3,10 €
l4-1-gpu	80 Go	20	NVIDIA Ada Lovelace L4 24 Go 30.3 TFLOPS	500 Go	5 Gbit/s	0,83 €
l40s-1-gpu	80 Go	13	NVIDIA Ada Lovelace L40S 48 Go 91.6 TFLOPS	1536 Go	5 Gbit/s	1,55 €

Why OVHcloud ? Our difference

Data Control and Sovereignty

Run AI inference on infrastructure that guarantees data residency, access control, and compliance with regulatory requirements.

Open and Portable Deployment

Deploy your own models using open standards and containers, ensuring portability across environments and avoiding vendor lock-in.

Enterprise-Grade Reliability

Benefit from a production-ready platform designed for high availability, resilience, and predictable performance.



AI Endpoints

Serverless API access to powerful AI models without the need to manage infrastructure or model hosting

Base API – Balanced Performance

Standard synchronous API with consistent performance and throughput for typical interactive workloads. Ideal for chat features, document Q&A, and semantic search.

Fast API – Ultra low latency

soon

High-performance synchronous API with minimal time-to-first-token and predictable behavior for latency-sensitive applications (coming soon).

Batch API - Cost-Efficient Bulk Processing

soon

Architecture – Technical Overview – Offer description

	Usual Flow	Specific Needs	
	Base API	Fast API	Batch API
Models available	All catalog	Subset of Base API models	LLMs & Embeddings
Use cases	Ideal for interactive workloads with standard throughput , where users experience usual interactions	Ideal for private workloads requiring high throughput and low TTFT (Time To First Token)	Optimized for large-volume or delayed workloads . Ideal for all your requests that can wait
Prompts per API Call	1 (max 2 Mo) / 1 - 25 for Embeddings	1 (Max 2 Mo)	Max 50 000
Response time	Standard response time experience	Minimal TTFT (Time To First Token) and TPOT (Time Per Output Token)	Up to 24h
Rate Limit	400 Requests Per Minute per project. (higher for embeddings)	Rate limit will depend on models	n.a.
Pricing model <small>Based on consumption unit (per token or per image or per second)</small>	With standard price, no commitment required, enjoy a pay-as-you-go consumption mode	You commit with minimum consumption, You are rewarded with lightening fast response time.	Cost-optimized, with batch requests billed at a significant discounted price compared to the Base API rate.
API Type	Synchronous	Synchronous	Asynchronous
SLA	99,8% uptime	99,8% uptime, guaranteed TTFT & TPOT	99,8% uptime, max 24h response time

Top use cases scenarios



AI-Enhanced Applications

Embed interactive intelligence (chatbots, semantic search, knowledge assistants) with consistent API performance using the Base API.



Predictable, High Throughput Services

Support enterprise-grade user-facing services and internal platforms that demand ultra-low latency or guaranteed API responsiveness with Fast API.



Bulk AI Workloads

Process large datasets (mass classification, embeddings generation, content tagging) in asynchronous batches to optimize cost and throughput.

Why OVHcloud ? Our difference

Performance-Tiered API Portfolio

Choose between Base, Fast, and Batch APIs that align with your application's performance and cost profile, from interactive responsiveness to large-scale asynchronous processing.

Serverless, API-First Simplicity

No infrastructure to manage. Developers focus on building features while OVHcloud handles scalability, availability, and model hosting.

Sovereign, Secure Cloud

Run AI inference on cloud infrastructure with transparent data governance and privacy guarantees.