OVHcloud

Transform Business Performance with Ultra-Low Latency

Contents

> **Latency is no longer just a technical KPI. It's a competitive lever. In this white paper, learn about the strategic importance of low latency in today's fast-paced digital landscape**

DISCOVER THE OPPORTUNITY FROM LATENCY OPTIMISATION, EDGE INFRASTRUCTURE MODELS, AND HOW TO DELIVER ULTRA-LOW LATENCY TO:

OPTIMISE PERFORMANCE

ENHANCE USER EXPERIENCE

DRIVE BUSINESS SUCCESS

Executive summary

THE SURGE IN DATA FLOWING FROM A GROWING ARRAY OF INDUSTRY ENDPOINTS HAS CREATED SIGNIFICANT IT CHALLENGES FOR ORGANISATIONS.

→ Digital transformation, industrial internet of things (IoT), the impact of consumer expectations and AI processing have made low-latency edge computing a critical factor for success.

→ Three key areas of impact include:

Revenue and conversion: Google found that a half-second delay in page generation could drop traffic by 20%[1].

User expectations and experience: 83% of people expect websites to load in 3 seconds or less, and 73% would try an alternative if a site is too slow[2].

Competitive advantage: Deloitte research suggests that lowering mobile site load time by just 0.1 seconds can boost conversion rates by almost 10%[3].

→ Latency optimisation can be harnessed to improve user experience in high-priority locations and reduce cloud migration risks.

→ Traditional on-premises edge deployments, designed for optimum latency, can cost twice as much as traditional cloud deployments.

→ Cloud service providers are responding with strategically placed cloud data centres close to areas with high user demand.

→ These cloud edge services allow businesses to deliver high-performance applications and services, accelerate cloud infrastructure transformation and enter new markets while maintaining data residency and responsive user experiences.

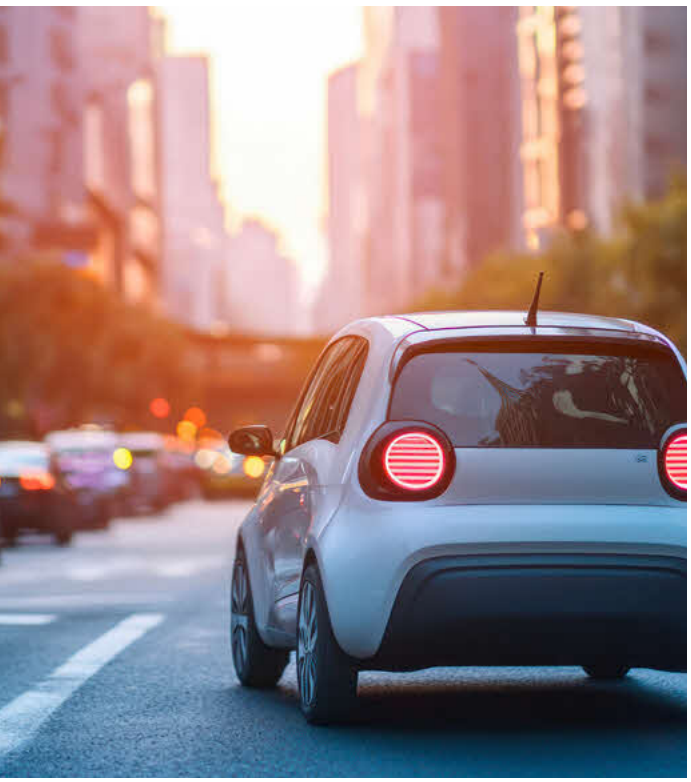
→ Hybrid cloud infrastructures make it possible for any organisation to adopt edge computing at a fraction of the cost of doing it themselves.

Introduction

Cloud migration has underpinned digital transformation and enabled more efficient and secure distributed workforces — solving the scalability, reliability and redundancy challenges of running applications from on-premises data centres.

But modern applications increasingly need more bandwidth and near-real-time feedback for performance, from video chat to autonomous vehicles and telemedicine.

This trend puts network latency in the spotlight. Business leaders increasingly recognise the strategic importance of optimising network latency to drive operational efficiency and business success.



Optimising network latency is increasingly important to drive operational efficiency, and ultimately, business success.

The causes of latency

On a network, latency is a metric for measuring the time it takes data to travel between two points or devices, often referred to as the delay or lag. It's impacted by the CPU, storage, operating system and network, and, ultimately, the distance the data needs to travel.

There are multiple causes for latency as data traverse routers, protocols and switches across the Internet, but the speed of light sets physical limits on the distance a signal can travel in a set timeframe.

Technological drivers of low and ultra-low latency requirements

Several factors are driving low and ultra-low latency requirements:

→ Embedding of real-time applications into critical processes through the digital transformation of operational technology (OT).

→ The rise in machine and Internet of Things (IoT) data to improve process efficiency and monitoring.

→ The growing use of AI models to support applications such as real-time analytics on video surveillance at retail checkouts, insights and automation on the factory floor, and telemedicine.

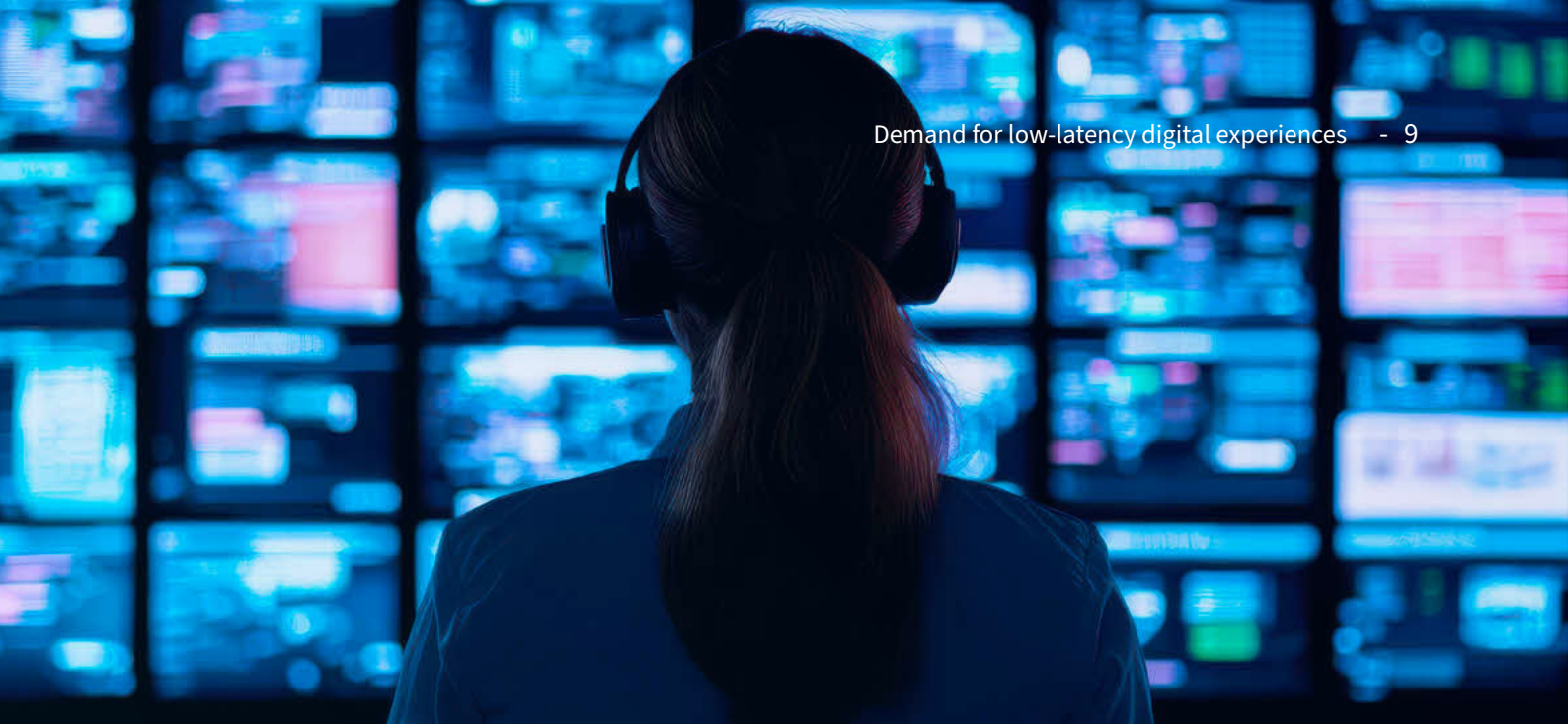
→ An explosion of industry-specific endpoints generating data that must be processed and analysed in real time to meet an application's service level agreement (SLA).

STATISTA, A GERMAN MARKET AND CONSUMER DATA COMPANY, ESTIMATES THERE WERE NEARLY 46 BILLION EDGE-ENABLED IOT DEVICES IN USE GLOBALLY IN 2024, AND THE NUMBER OF EDGE-ENABLED IOT DEVICES IS PROJECTED TO CLIMB TO 77 BILLION BY 2030.[4]

Use cases include sensors for real-time communication in industrial applications, connected and autonomous cars, healthcare monitoring systems and telesurgery, and industrial drones and robotics. These are some of the key emerging applications that are driving digital transformation of OT and its convergence with traditional IT.

Edge computing, cloud and 5G have been critical enablers for the growth in latency-sensitive applications.

5G reduced latency to below 10 ms, compared to 200 ms for 4G[5]. This opened the door to many low-latency applications for IoT, which also needed edge computing for efficiency. And with 6G on the horizon, the opportunities for latency reduction in the future could be vast.



Demand for low-latency digital experiences

The performance of digital infrastructure is now a direct lever for business growth. Latency — even in the range of milliseconds — has a tangible impact on revenue, customer expectations and market competitiveness.

Three key areas of impact include:

Revenue and conversion:
Google found that a half-second delay in page generation could drop traffic by 20%[6].

User expectations and experience:
83% of people expect websites to load in 3 seconds or less, and 73% would try an alternative if a site is too slow[7].

Competitive advantage:
Deloitte research suggests that decreasing mobile site load time by just 0.1 seconds can boost conversion rates by almost 10%[8].

The financial impact of latency

Measuring the financial impact of network latency is difficult, but milliseconds do matter in numerous use cases.

IDC's Edge Services Thought Leadership Survey found that 46% of healthcare respondents said the latency threshold for their edge initiatives was below 5 ms[9].

In manufacturing, where robotics and automated assembly rely on real-time communication, low-millisecond delays can cause inaccurate movements, high cognitive load and safety hazards for workers[10]. High latency can also undermine the cost benefits of predictive maintenance, remote monitoring of industrial equipment and machine-to-machine communication.

> **A site that loads in 1 second has an ecommerce conversion rate 2.5X higher than a site that loads in 5 seconds.[11]**

In an e-commerce context, reductions in latency can have significant benefits. A 2020 Deloitte study found that a 100-ms speed upgrade on mobile site loads increased retail conversions by 8.4% .

E-COMMERCE

Abandonment rates: 40% of shoppers won't wait more than 3 seconds before abandoning an e-commerce site.

Conversion and revenue: Sites that load faster rank higher in Google search, gaining more visibility and traffic.

Customer loyalty: Almost 70% of consumers say that page speed affects their willingness to buy, and 50% would prefer to skip animations and videos for faster load times[13].

MANUFACTURING

Production efficiency: Edge computing and low-latency networks (such as 5G) enable real-time monitoring and predictive maintenance, reducing downtime and maintenance costs.

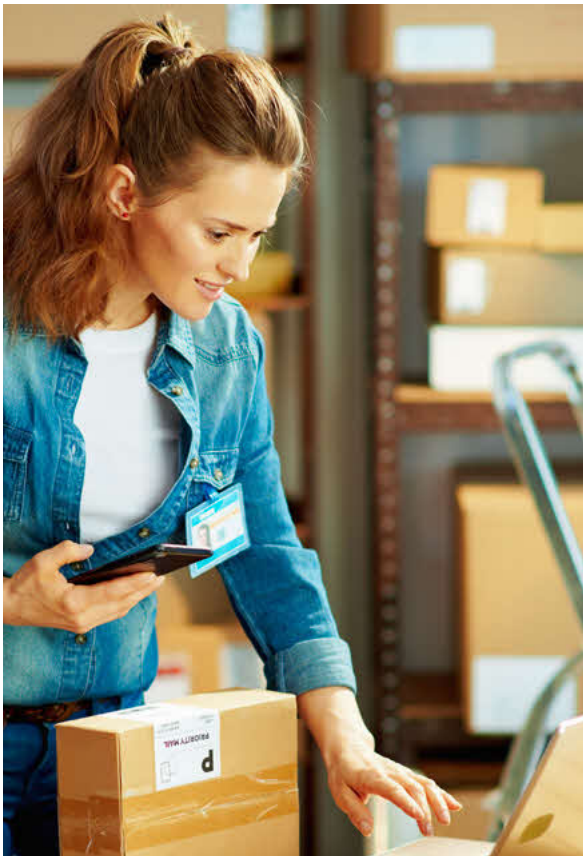
Reliability: Low-latency requirements are often at or below 10 ms for mission-critical manufacturing processes. This enables autonomous guided vehicles (AGVs) and energy management systems to operate safely and efficiently in real time.

Industry value: 5G and edge computing are projected to unlock $740 billion in manufacturing value by 2030, largely through improved connectivity, reliability and low-latency automation.

HEALTHCARE

Real-time applications: Industry trends show that low-latency infrastructure is vital for telemedicine, remote monitoring and AI-assisted diagnostics. Delays can impact patient outcomes, especially in critical care and remote surgery.

Reliability and security: Healthcare organisations prioritise low-latency, secure data transmission for compliance and patient safety, using edge and cloud solutions to support these needs.



Latency thresholds and the optimisation opportunity

How critical low latency is to an application depends on the use case and perspective.

For a video-streaming service, software as a service (SaaS) provider or e-commerce site, latency can be critical for business success, with high latency resulting in lost users, fewer sales, lower feature adoption and missed upsell opportunities.

Regardless of context, all cases demonstrate the benefit of optimising latency with edge computing to achieve a better outcome for the business and users.

Use Case	Target Latency	Business Impact
Retail mobile checkout	100 ms load	+8.4 % conversion per 0.1 s speed (Deloitte)[14]
VoIP calls	150 ms one way	Maintain call quality (IEEE Spectrum)[15]
Multiplayer gaming	40–60 ms	Competitive play and engagement (IEEE Spectrum)[16]
VR applications	7–20 ms	User comfort and immersion (IEEE Spectrum)[17]
Industrial automation	5 ms or below	Prevent errors and hazards (Study: Effects of Signal Latency on Human Performance in Teleoperations[18])
Healthcare applications	5 ms or below	Real-time diagnostics (IDC Edge report)[19]

The Institute of Electrical and Electronic Engineers (IEEE) highlights acceptable latency levels for various technologies today[20]. Acceptable latency for VoIP one-way and video chat is about 150 ms, while the tolerance for latency in multiplayer gaming is sub-30 ms. Augmented reality, haptics and autonomous vehicles demand around 2 ms latency.

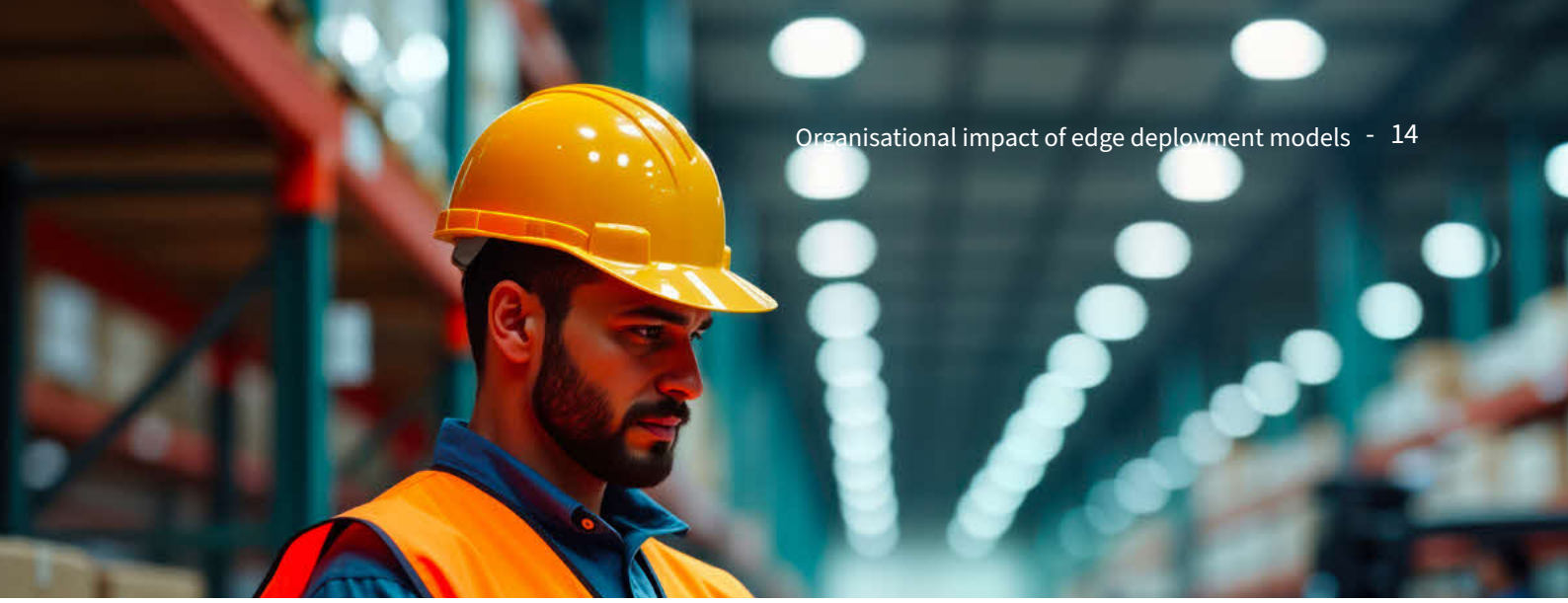
The move to edge computing

The most viable way to reduce latency is to move computing closer to the location where data creation, processing and consumption occurs.

Latency requirements are helping to drive the rise of edge computing, a distributed computing model where processing happens near the source of data generation — the 'edge' of the network — rather than relying on a centralised cloud or data centre. Edge computing delivers several organisational benefits:

- → **Low latency**
- → **Reduced bandwidth requirements**
- → **High performance**
- → **Local data processing**

> **Edge computing delivers several organisational benefits — low latency, reduced bandwidth, high performance and local data processing.**



Organisational impact of edge deployment models

A 2024 survey by S&P Global Market Intelligence commissioned by Red Hat found that:

57%

saw their edge computing investment as “high impact” to their IoT and OT transformation.

47%

said their key edge digital enablement goals were to optimise business processes and operations.

45%

wanted to cut costs[21].

Within healthcare specifically, IDC found that 51% of organisations deployed their edge solutions in an on-premises facility.

Although edge computing is sometimes seen as a solution to cloud computing, IDC found that most healthcare organisations were also migrating to a private cloud due to data privacy and compliance requirements, and that they viewed edge as complementary to their cloud strategy rather than as a replacement.

Cloud providers are increasingly recognising the specific needs of the healthcare sector with solutions — for example, OVHcloud’s On Prem Cloud Platform — designed to meet edge computing, data privacy and compliance requirements[22].

The cost of edge infrastructure

Building edge computing on premises comes with capital and operational costs just like those for building data centres. These include costs for servers, data storage, virtualisation software, network, power distribution, cooling, electrical power and maintenance.

> **Estimates suggest it would cost an enterprise twice as much in one year to run edge computing on its own infrastructure compared to traditional cloud.[23]**

Ericsson estimated that it would cost an enterprise twice as much in one year to run edge computing on its own infrastructure compared to traditional cloud infrastructure[24]. This aligns with other estimates that found the total cost of ownership of on-premises edge deployments was 35% higher than traditional cloud after one year and 100% higher over three years[25].

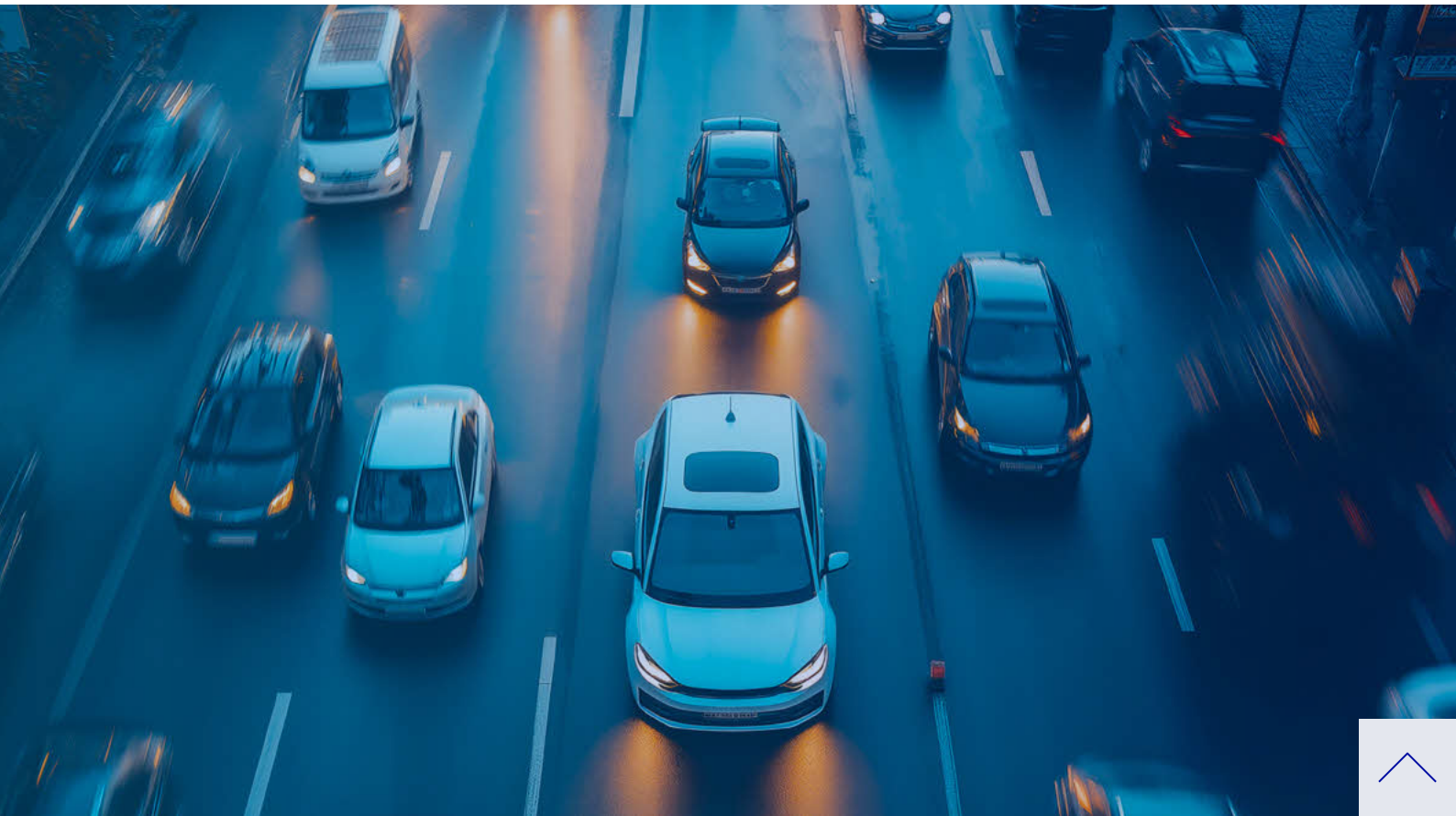
Challenges and costs increase when an organisation considers deploying and staffing a large network of edge sites.

The evolution of cloud services to deliver edge

Cloud service providers are responding to the growing need for low and ultra-low latency with the introduction of Local Zone cloud services[26]. These are strategically placed cloud data centres close to areas with high user demand.

They allow businesses to deliver high-performance applications and services while maintaining data residency and ensuring responsive user experiences.

Local Zones are increasingly becoming vital within hybrid and multi-cloud infrastructures and can be used in isolation or alongside centralised data centres. This is making it possible for every organisation to adopt edge computing at a fraction of the cost of doing it themselves.



Four powerful uses for cloud-edge Local Zones

1.
IMPROVE USER EXPERIENCE

From real-time online gaming and live video streaming to augmented reality systems, a range of applications are commercially viable only with extremely low latency.

2.
ACCELERATE CLOUD TRANSFORMATION

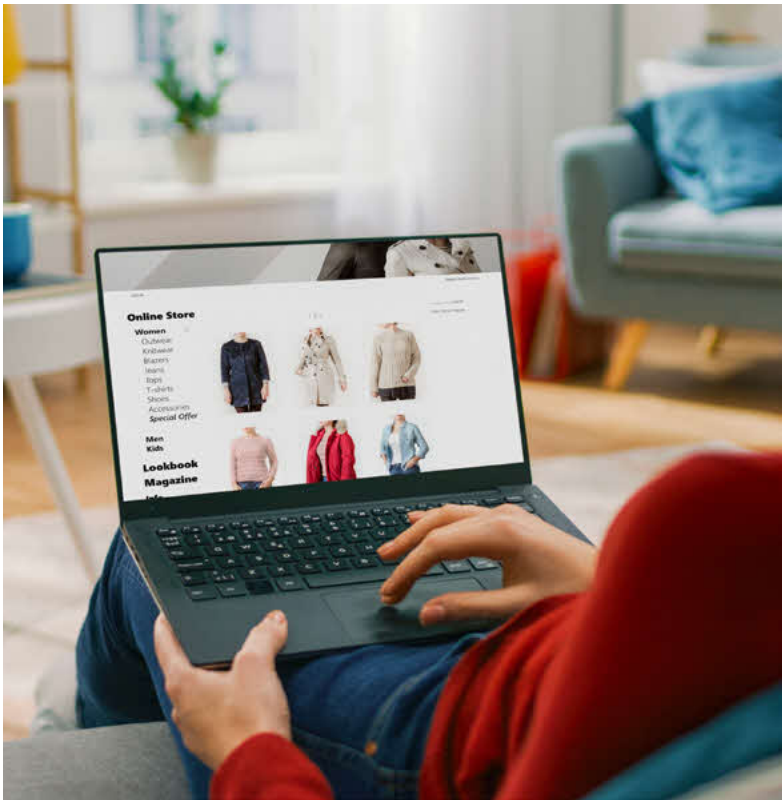
Businesses migrating complex, latency-sensitive applications to the cloud to meet scalability, security, management and cost requirements can benefit from Local Zones. They can use them to deliver a localised subset of services, enabling a smooth migration process, even when the volume of data being exchanged is extremely high.

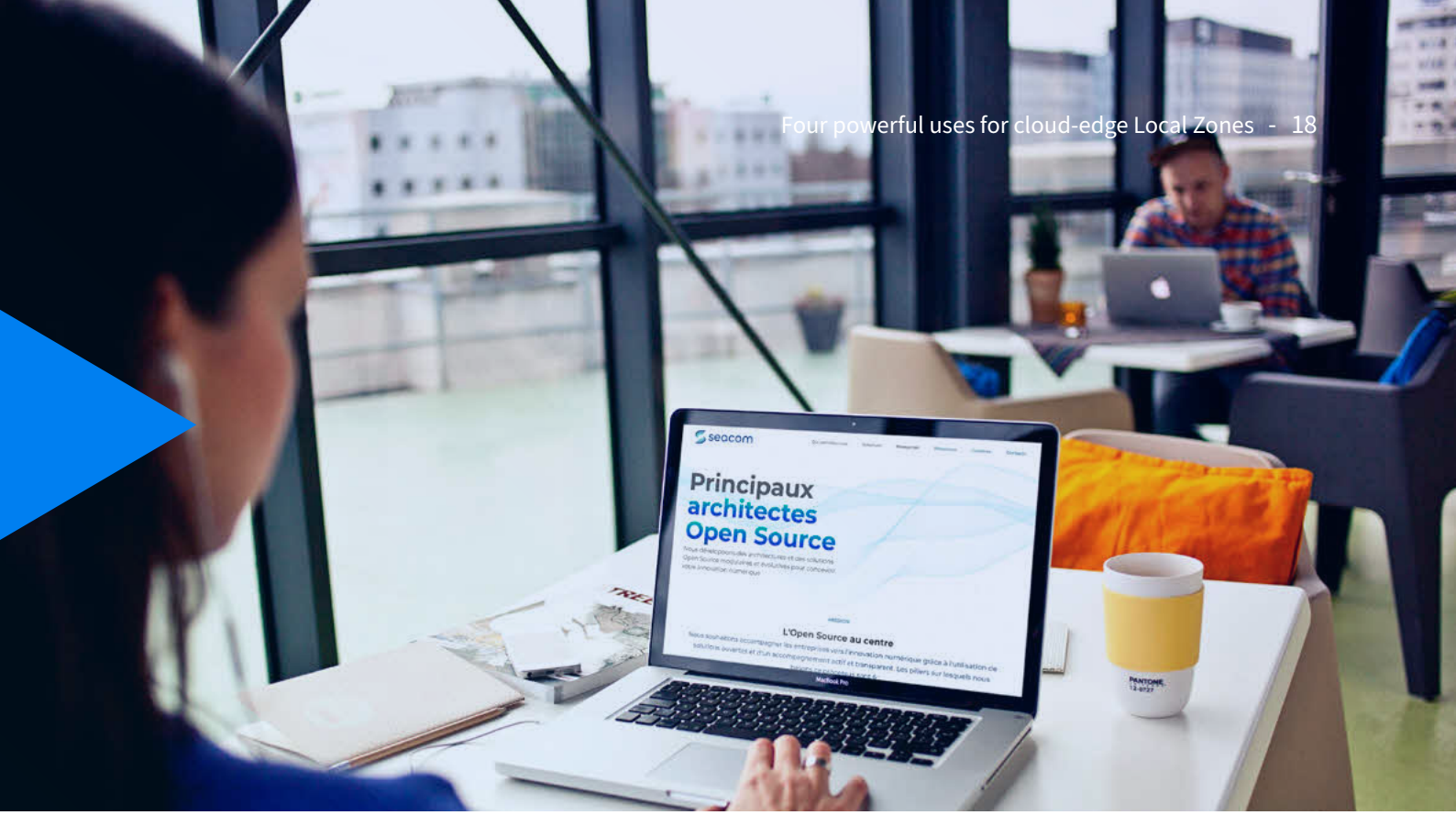
3.
ENTER NEW MARKETS

Local Zones enable organisations to maintain high performance and consistent infrastructure, regardless of location, while delivering single-millisecond latency on computing and storage instances across a wide range of new territories. Because capacity is scalable based on need, the sunk cost of entering new markets is reduced.

4.
DEPLOY A LARGE VPN

Demand for virtual private networks (VPNs) has skyrocketed in recent years. 41% of people in the UK report using a VPN once per week, while nearly a quarter of people in France, Germany and Sweden uses one weekly[27]. Deploying a large VPN network requires reliable ultra-low latency infrastructure.





Case study:

CLOUD-EDGE LOCAL ZONE – SEACOM

Seacom, an open source architecture and solutions consultancy, needed to rapidly scale its infrastructure to meet the needs of its growing business. Seacom chose OVHcloud Local Zones to meet both scalability and data compliance requirements.

> **The OVHcloud Local Zone in Milan has empowered Seacom to scale its infrastructure while ensuring full regulatory compliance and local data residency. By leveraging a range of flexible solutions — from storage to public cloud instances — and centralising through a single-entry point, we achieve streamlined, secure and transparent management with greater agility.**
>
> **Alessandro Fortunati**, CTO at Seacom

Conclusion

Latency optimisation in IT infrastructure delivers measurable business value: higher revenue, improved conversion, greater operational efficiency and more customer satisfaction. E-commerce, manufacturing and healthcare sectors all report significant gains from faster, more reliable digital experiences.

Although the cost and management time of edge computing infrastructure can be prohibitive, Local Zone cloud services give businesses the infrastructure required to deliver high-performance applications and services while maintaining data residency, ensuring responsive user experience and controlling costs.

OVHcloud Local Zones

DELIVERING HYBRID-CLOUD INFRASTRUCTURE

Local Zones are an extension of OVHcloud regions that bring OVHcloud services closer to specific locations, offering reduced latency and improved performances for applications[28].

> **Through OVHcloud Local Zones, customers can reduce cloud migration risks through hybrid edge-core deployments and achieve Total Cost of Ownership (TCO) savings of 41% compared to hyperscaler edge solutions.**

→ **Sub-10ms latency by design:** Engineered for latency-critical use cases to deliver edge-native application performance.

→ **Fully cost optimised:** The benefits of edge computing without the project and maintenance costs of on-premises infrastructure. Egress-fee-free, on-demand scalability and transparent pricing.

→ **More than 30 locations:** Through EMEA and America, expanding to 150 cities across the globe in 3 years.

→ **Seamless DevOps integration:** Provision Local Zones instances within minutes with Terraform and existing automation pipelines.

→ **Edge expertise on demand:** Certified partners and experts available to support complex deployments.

The new Local Zones are an interesting product to further lower the latency of connections to distinct regions. We're always keen to improve our services and grow together with OVHcloud.

Fabian Girsch, Head of Backend Operations, AnyDesk

Get support with qualifying use cases, building a business case or help to prepare for scaling.

Contact our team to find out more: https://www.ovhcloud.com/nl/contact/

MENTIONS

[1] 2 6 7 13 91 powerful statistics that show why your ecommerce site speed matters in 2025
► https://queue-it.com/blog/ecommerce-website-speed-statistics/

3 Latency is Having a Huge Negative Impact on eCommerce Companies
► https://www.globaldots.com/resources/blog/latency-is-having-a-huge-negative-impact-on-ecommerce-companies/

4 15 edge computing trends to watch in 2025 and beyond
► https://www.techtarget.com/searchcio/tip/Top-edge-computing-trends-to-watch-in-2020

5 5G Will Accelerate a New Wave of IoT Applications
► https://newsroom.ibm.com/5G-accelerate-IOT

8 12 14 Milliseconds Make Millions
► https://www.deloitte.com/ie/en/services/consulting/research/milliseconds-make-millions.html

9 Edge Computing: Transforming Healthcare by Increasing Resilience
► https://www.multivu.com/players/English/85243514-lumen-healthcare-systems-idc-survey/docs/InfoBriefPDF_1624646596597-663284142.pdf

10 Effects of Signal Latency on Human Performance in Teleoperations
► https://www.researchgate.net/publication/354509211_Effects_of_Signal_Latency_on_Human_Performance_in_Teleoperations

[1]1 Site Speed is (Still) Impacting Your Conversion Rate
► https://portent.com/blog/analytics/research-site-speed-hurting-everyones-revenue.htm

[1]5 16 17 20 Breaking the Latency Barrier
► https://spectrum.ieee.org/breaking-the-latency-barrier

[1]8 Effects of Signal Latency Human Performance in Teleoperations
► https://www.researchgate.net/publication/354509211_Effects_of_Signal_Latency_on_Human_Performance_in_Teleoperations

[1]9 Edge Computing: Transforming Healthcare by Increasing Resilience
► https://www.multivu.com/players/English/85243514-lumen-healthcare-systems-idc-survey/docs/InfoBriefPDF_1624646596597-663284142.pdf

2[1] Industry-specific needs call for differentiated edge deployments
► https://www.redhat.com/rhdc/managed-files/cl-451-research-on-edge-computing-analyst-material-1323765-202408-en.pdf

22 OVHcloud On-Prem Cloud Platform
► https://www.ovhcloud.com/en-gb/dc-as-a-service/

23 24 Edge computing – significant growth potential for service providers
► https://www.ericsson.com/en/reports-and-papers/mobility-report/articles/enabling-demanding-use-cases-with-csp-edge-computing

25 Analysis: The economics of edge computing
► https://www.edgecomputing-news.com/news/analysis-economics-of-edge-computing/

26 Accelerate performance where milliseconds matter
► https://www.ovhcloud.com/en-gb/about-us/global-infrastructure/local-zone/

27 How often do you use a VPN?
► https://www.statista.com/statistics/1219770/virtual-private-network-use-frequency-us-uk/

28 OVHcloud: Worldwide Public Cloud presence
► vhttps://www.ovhcloud.com/en/public-cloud/regions-availability/

Transform Business Performance with Ultra-Low Latency

OVHcloud

About OVHcloud

OVHcloud is a global cloud player and Europe's leading cloud services provider. The company operates more than 450,000 servers in 40 datacentres across 4 continents, serving 1.6 million customers in more than 140 countries. As the driving force behind creating a trusted cloud and pioneering a sustainable cloud with the best value for performance, for over 20 years the Group has employed an integrated model that gives it complete control over its value chain. From designing its own servers to building and managing its datacentres to orchestrating its fibre-optic network, this unique approach allows OVHcloud to independently cover all use cases by enabling its clients to benefit from its environmentally-sound model, with a careful use of resources and one of the best carbon footprints in the industry. OVHcloud now offers latest-generation solutions that combine high performance, predictable pricing and full data sovereignty to support their clients' unfettered growth.