Enterprise Strategy Group by TechTarget

SHOWCASE

JULY 2023

How AWS and Its Partner Ecosystem Help Optimize Performance for Low-latency Applications

Paul Nashawaty, Principal Analyst

Overview

Low latency applications are essential for latency-sensitive multiplayer games that need to be closer to end users and for application migrations that are latency-sensitive due to local system/network interdependencies. To perform properly, these latency-sensitive workloads and applications require minimal delays or lagtime to perform optimally. For these applications, milliseconds matter. Not only do today's users expect to interact with low-latency applications with no perceptible delays, but the success of many businesses also depends on the ability to meet specific low-latency requirements. For example, low-latency trading has strict latency service-level agreements, where a millisecond can make the difference between completing a transaction or missing an opportunity. Even a slight latency variance can lead to differences amounting to millions of dollars.

Network latency is the delay in network communication. It shows the time that data takes to transfer across the network. Networks with a longer delay or lag have high latency, while those with fast response times have low latency.

Many organizations with hybrid cloud and edge computing sites in widely dispersed locations are challenged to securely deliver high-performing, low-latency applications to their end users. Other applications are sensitive to latency for performance and have traditionally remained on premises or in data centers to achieve the desired latency profile. As organizations modernize and migrate business-critical legacy applications to the cloud, they may find it challenging due to latency-sensitive system interdependencies between the various components of the application. This paper explores the advantages of working with a trusted cloud provider like Amazon Web Services (AWS) and its partner ecosystem to manage and improve performance for low-latency applications.

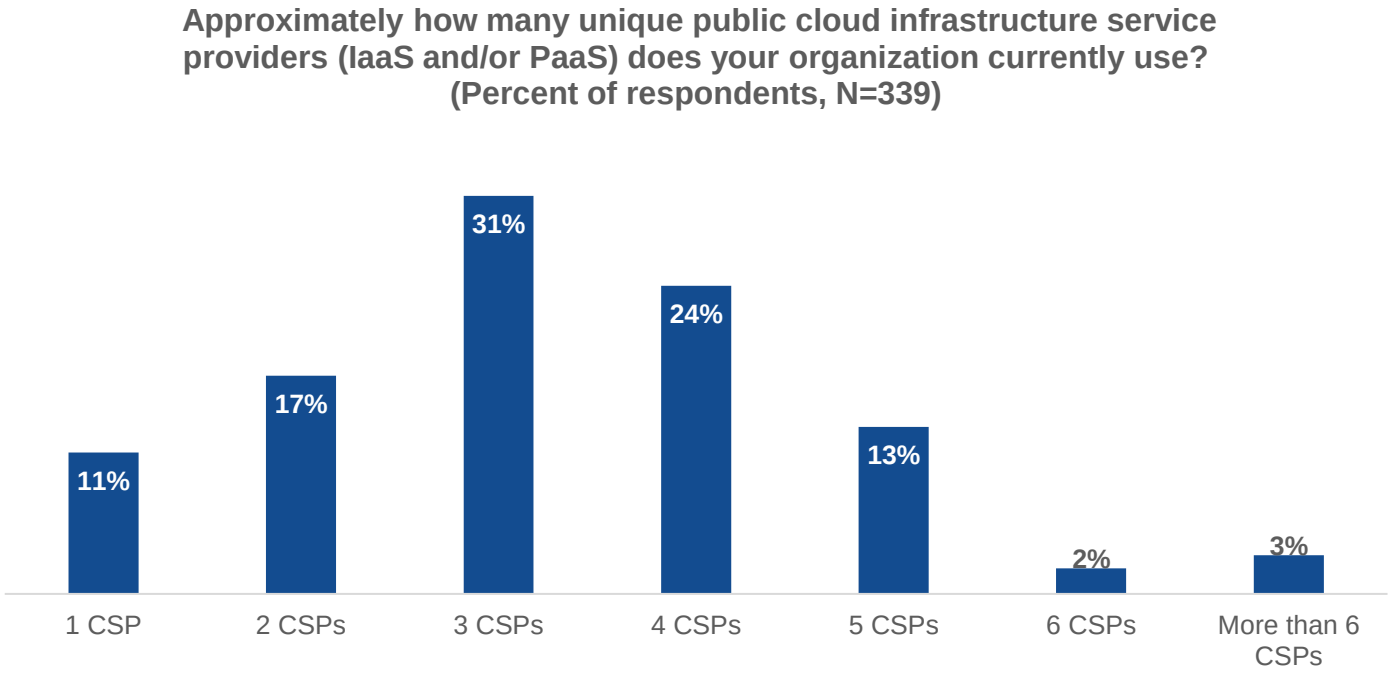
Introduction

The popularity of cloud-native applications is on the rise, according to research from TechTarget's Enterprise Strategy Group, which shows that spending is expected to significantly increase by 43%.[1] In addition, a separate Enterprise Strategy Group research survey revealed that 37% of respondents reported they have a shortage of skills in cloud architecture and planning.[2]

Companies delivering low-latency applications are reexamining their strategy to consider working with a cloud service provider to leverage best-in-class infrastructure services for improved user experiences. In fact, Enterprise Strategy Group research shows that 55% of organizations use 3 or 4 cloud service providers for their applications and workloads (see Figure 1).

[1] Source: Enterprise Strategy Group Research Report, *Cloud Native Applications*, May 2022.
[2] Source: Enterprise Strategy Group Research Report, *2023 Technology Spending Intentions Survey*, November 2022.

Figure 1. Number of Unique Public Cloud Infrastructure Service Providers Currently Used



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Let's look at why this approach for low-latency applications can be advantageous. Businesses that want to take advantage of a hybrid approach often wind up with on-premises services and infrastructure, alongside the cloud, which can leave them with various tools and platforms that don't communicate with each other, as well as require that they train both IT teams and developers. Not only is this solution difficult to scale, but it's also unnecessarily complex and costly. After seeing this traditional approach to a hybrid environment insufficiently solve an organization's challenged, businesses find that it's key to have a seamless hybrid approach.

Achieve Optimal Performance at Any Location

Latency is a function of the physical distance between application servers and end users. The closer an application server is physically located to end users, the lower the latency and the faster the performance. To deliver highly performant applications to a global and distributed audience, organizations require a vast network of global data centers close to their users; however, managing this kind of network in-house is expensive and inefficient.

Location impacts an application's performance because of:

- **Internet service providers (ISPs):** The performance of low-latency applications can be impacted by the quality and reliability of the ISPs used by end users. For example, if the ISP is congested or experiencing issues, this can increase latency.
- **Geopolitical factors**: Some industries are subject to regulatory requirements put in place to protect data and user privacy. These regulations can place limitations on the flow of data across borders and can impact the performance of low-latency applications.

With AWS, organizations can:

- Accelerate their innovation with the broadest and deepest set of capabilities.
- Create a truly consistent and secure experience across their environments.
- Optimize their applications with experienced partners and the best support.

Use Cases

Low-latency applications are especially critical in certain domains, including:

- Manufacturing for enhanced efficiency and effectiveness of automation across various industrial processes.
- Content creation and more immersive visuals for media production, live streaming, virtual reality/augmented reality (VR/AR), gaming, and real-time collaboration platforms.
- Financial trading, especially for high-frequency trading platforms and algorithmic-based trading strategies.
- Autonomous vehicles, to enable real-time decision-making, vehicle-to-everything communication, sensor fusion and object tracking, rapid emergency response, cloud-based services and updates, and local decision-making through edge computing.
- Inference at the edge, performing data processing, analysis, and decision-making tasks directly on edge devices or local computing resources.

With increasing globalization and digitalization, organizations need easier ways to deliver low-latency applications while complying with ever-changing data protection and privacy rules. A cloud provider such as AWS has data centers located in different regions around the world, in AWS Regions and AWS Local Zones (LZs). Choosing a Region or an LZ closest to end users or sources of data can achieve the desired latency profile and improve application performance. In addition, AWS offers services that solve for low latency like AWS Outposts , AWS Wavelength, and AWS Snow Family.

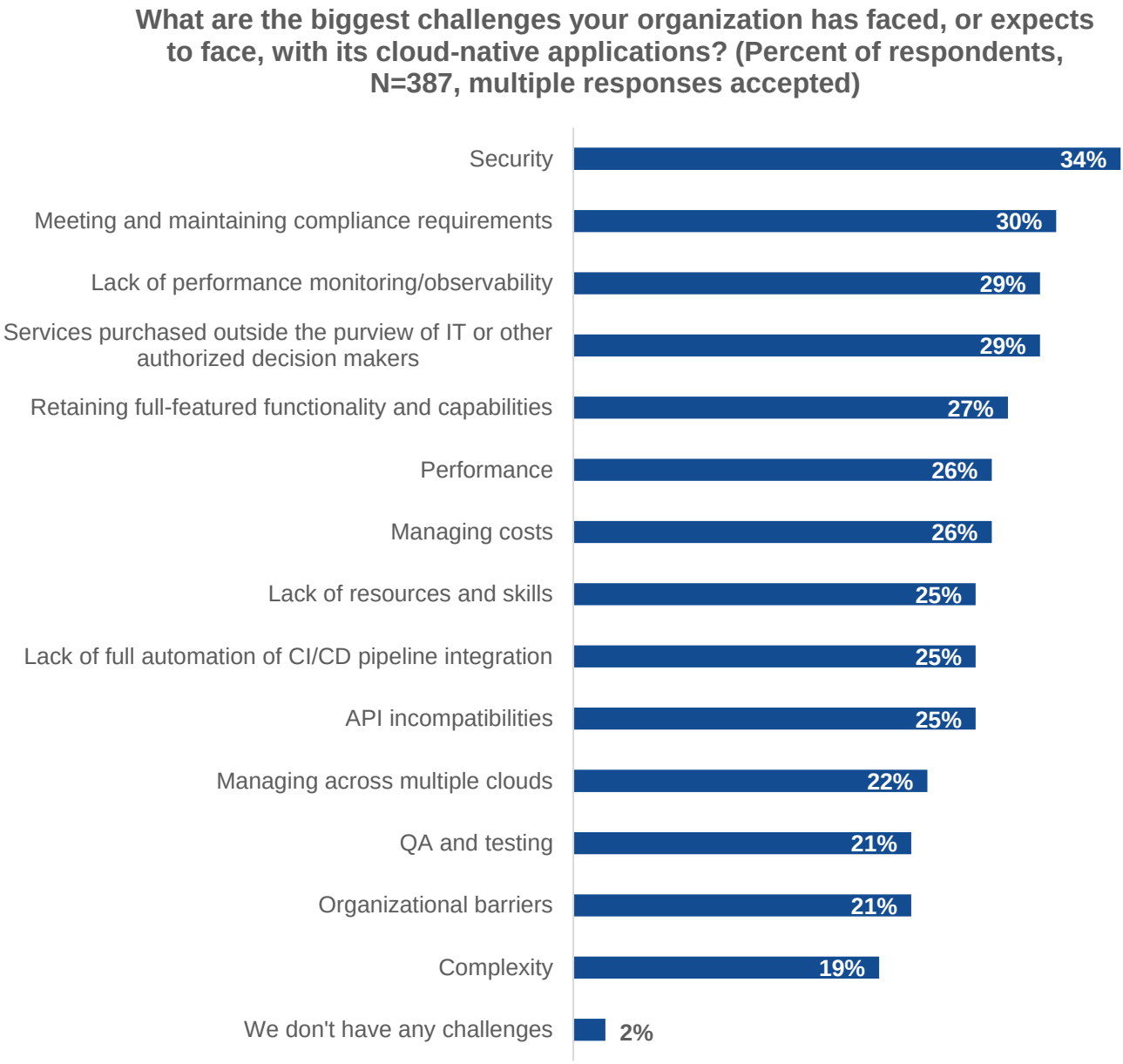
Scalability and Security for Low-latency Apps

As the number of users of these applications grows, so does the volume of data and processing speed and power required—making it difficult to maintain low latency while scaling the application. These applications require fast and reliable access to data and real-time processing.

Another challenge with low-latency applications is keeping them secure and compliant when running on company-owned infrastructure. According to Enterprise Strategy Group research, security and meeting and maintaining compliance requirements were cited as the most common cloud-native application challenges (see Figure 2).[3] Reliable network infrastructure and observability tools that can handle the increased traffic and data volume need to be in place.

[3] Source: Enterprise Strategy Group Research Report, *Cloud-native Applications*, May 2022.

Figure 2. Security and Compliance Are the Most Common Cloud-native Application Challenges



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Another challenge related to security for low-latency applications is ensuring that the data is encrypted in transit and at rest to prevent data leaks and unauthorized access. Additionally, companies need to implement strict access controls and authentication mechanisms to prevent unauthorized access to data.

Working with a cloud service provider like AWS gives organizations access to tools like Amazon CloudFront, which securely delivers content with low latency and high transfers speeds. It offers improved security with traffic encryption and access controls and uses AWS tools like AWS Shield Standard to defend against DDoS attacks.

The Right Approach

Ensuring that low-latency applications are secure and performing optimally requires organizations to be able to address the ever-evolving technical challenges that come with global users in myriad locations, including network congestion, bandwidth, data storage and analytics, and processing power. AWS provides a robust platform with a wide range of features and services and a vast partner ecosystem that can help companies optimize the performance of their low-latency applications while reducing the operational and cost burden and improving security and compliance.

Organizations can bring AWS wherever they need it for low latency from the cloud to data centers, large metro areas and 5G networks, smart factories, or buildings, and beyond. AWS is the world's most comprehensive and broadly adopted cloud, offering the most advanced services and capabilities, including a global infrastructure. AWS has a vast global network of data centers, enabling organizations to deploy their applications closer to their end users, reducing latency and improving performance. The AWS global infrastructure includes:

- **AWS Regions.** An AWS Region is a physical geographic location that has a cluster of AWS data centers within it. These data centers are known as Availability Zones (AZs), LZs, and edge networking and are all used by grouping in a minimum of three data centers per Region to provide redundancy and enable organizations to build resilient applications and services.
- **On-premises and hybrid solutions**. For companies using a hybrid cloud, AWS Outposts is a fully managed service that offers AWS infrastructure, services, APIs, and tools to virtually any data center, colocation space, or on-premises facility for a truly consistent hybrid experience. AWS Outposts is ideal for workloads that require low-latency access to on-premises systems, local data processing, and data residency.
- **Security and compliance**. Security continues to be a serious concern for companies hosting low-latency applications on their own infrastructure because of the sensitive and critical data involved and the need to maintain a secure and reliable network infrastructure.

Using a cloud services provider like AWS for low-latency applications can help companies ensure not only secure application performance but also, with the economies of scale, a more cost-effective pay-as-you-go pricing model.

AWS Partners

Organizations of every size around the world, including large enterprises, government organizations, and startups, trust AWS and its partner community to run their most essential applications—including mission-critical, enterprise, and cloud-native applications—and accelerate their digital transformation. Customers can bring AWS wherever it's needed for low latency—from the cloud to data centers, large metro areas and 5G networks, smart factories or buildings, and beyond. And with services and solutions from AWS Partners, customers can further optimize, manage, and improve performance for low-latency applications. Emerging applications, including VR/AR, connected vehicles, immersive events and experiences, and more can be supported by the innovative offerings available from AWS Partners with expertise in low-latency scenarios.

Need help finding a partner? AWS Partner specialists can help you find and connect with an experienced provider with the expertise to meet your low-latency, compliance, and data residency needs.

Find an AWS Outposts Ready Partner.

Conclusion

AWS is reinventing hybrid cloud and edge computing by bringing AWS wherever companies need it—from the cloud to data centers, to large metro areas and 5G networks, to smart factories and buildings, and more.

Working with AWS and its partners allows an organization to build, deploy, and manage low-latency applications on a common platform that uses the same hardware, services, security model, and tools across their environment for consistency, allowing them to improve end-user satisfaction and IT and developer productivity.

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com