ORACLE

# 11 Ways the Cloud Gets Better in 2025

What's new and next for organizations looking to level up? Turns out, a lot—and not just AI

# Table of Contents

# Level Up Your Cloud Strategy: 11 Essential Trends for 2025

Hyperscale cloud service providers that a few years ago envisioned locking enterprises into the 2.0 version of a walled garden are waking up to reality: Business leaders want flexibility, interoperability, and choice.

They also want practical AI capabilities, support for sovereign data requirements, advanced security that lets workers do more while keeping bad actors at bay, and more. Providers are stepping up.

**Here are some top ways your company can benefit from cloud services in 2025.**

# 1 Already excellent security will get better

Enterprises understand that hyperscalers invest millions in security and have a reputational interest in keeping customer data safe. That effort, of course, is an arms race, so providers are often the first to adopt new strategies and technologies. Cloud security innovations coming online now and into 2025 include clouds with zero trust packet routing, use of AI for threat detection, improved identity and access management, and more.

One major advance, the use of a zero trust security model, boils down to a "never trust, always verify" stance that uses granular permissioning to limit the attack surface and assumes the enemy is already past the firewall. Thus, more frequent user authentication and authorization are needed to protect assets, as is continuous monitoring for signs of breaches within the network perimeter.

Organizations working to implement zero trust security can investigate whether their providers can easily define specific access pathways for sensitive data, possibly using a new network security approach called zero trust packet routing, or ZPR. Those not exploring a zero trust architecture may want to reconsider, for a few reasons. First, you have more options for authentication thanks to software, services, and hardware that increasingly support biometrics and other sign-on methods beyond passwords.

**The Path to Passwordless**

Most enterprises understand the need for multifactor authentication for users accessing cloud and on-premises resources. The next frontier—especially for highly regulated industries—is so-called passwordless authentication. This is a method of verifying a user's identity without requiring a password, though passwords may still be used. Passwordless authentication depends on technologies such as short-lived one-time passwords with a mobile authenticator, biometrics, push notifications, one-time passcodes sent via SMS or email, and hardware security tokens such as YubiKeys.

If your organization is looking to reduce exposure to password theft and phishing attacks, ask your cloud provider about its passwordless roadmap. Look for a solution that keeps the login process relatively simple for users without compromising on security.

**And generative AI can now help you enable a zero trust architecture. How?**

- GenAI-powered automated threat detection systems can analyze huge volumes of network traffic and system logs to help identify anomalies and potential threats, and where authorized, stop them in real-time.
- AI can also establish baselines for normal user behavior and analyze patterns to help detect anomalous activity that might indicate, for example, someone trying to log on with stolen credentials.
- AI agents can automate provisioning and deprovisioning, reducing the risk of human error.

Moreover, AI agents can help make sure each system or user is granted only the minimum level of access necessary to get the job done, another tenet of zero trust. These factors combine to make 2025 the year the model gets more realistic for more organizations.

Right now, AI is assisting your cloud providers with anomaly, malware, and intrusion detection by analyzing network traffic, cloud resource use, user behavior, and system logs to spot signs of problems and alert security teams. In 2025, expect providers to make strides in notifying customers promptly of unexpected configuration changes and providing recommendations for responding while implementing strict separation of duties.

> "As companies upgrade encryption, they may also consider implementing zero trust, per Deloitte."[1]
>
> Learn more about zero trust security

[1] *Deloitte Tech Trends 2025*

# 2 GenAI agents will be at your service

Do teams throughout your organization struggle to keep up with their workloads? Large language models (LLMs) are the brains behind new AI agents that, with your staff's oversight, can do a wide variety of jobs.

AI agents can be assigned tasks, examine their environments, and take actions within the parameters of their roles and goals. Agents retain past interactions and refine their decision-making based on positive, neutral, and negative outcomes. Advanced agents can mimic human ingenuity and use tools, including your applications and data sources, APIs, and other agents, to achieve their goals.

But they're not typically DIY. Like all AI, agents demand significant computational power and storage and often work best in the context of business applications such as ERP, inventory, HR, or supply chain management systems. When agents are delivered in the cloud by your technology partners—especially those that you trust with your data—it's up to the provider to have trained the agents and given them proper access to the data they'll need within the limits you set.

2025 will see a flood of AI agents that can help with tasks as varied as contract creation, project management, accounting, and customer support. The good news is that providers are very focused on ease of setup, use, and monitoring.

> Advanced agents can mimic human ingenuity and use tools to achieve their goals.

Learn more about AI agents

**Agentic AI case study: A recommendation engine**

A prime cloud-supported use case: customer purchasing history analysis agents that integrate with your CRM and ERP suites to provide insights into buyer behavior and help improve marketing and sales efforts.

An AI agent ingests data that could include customer demographics and purchasing behavior as well as product pricing and inventory projections and uses machine learning tools to identify patterns and trends. After segmenting customers based on factors your team has identified as relevant, the agent can apply predictive models to forecast behavior and predict which of your products a cadre is likely to purchase.

Other AI agents could then craft personalized marketing emails. And, by collecting data like customer lifetime value, agents can help identify your most valuable customers and alert sales to reach out and discuss new opportunities.

# 3 Intercloud partnerships will make multicloud a bigger win for businesses

That cloud migration you've been wanting to do if only your strategic data could come along? It's time.

> "We're entering a new phase where services on different clouds work gracefully together. The clouds are becoming open. They're no longer walled gardens; customers have choices and will use multiple clouds together."
>
> **Larry Ellison**
> Chairman and CTO of Oracle

Multicloud in 2025 means the ability to manage and use your primary database to inform applications in the hyperscale clouds of your choosing. That helps IT leaders deliver a diverse yet unified cloud portfolio. AWS, Google Cloud, Microsoft Azure, Oracle Cloud Infrastructure—hyperscalers are forging partnerships that allow you to have full access to your data within the facilities of any cloud provider, or several, and without massive data egress fees. Customers can now natively run Oracle Database as a managed service within AWS, Google Cloud, and Microsoft Azure, reducing latency and helping simplify budgeting and data governance.

89% of respondents employ a multicloud strategy now, according to Flexera's 2024 State of the Cloud Report.[2]

> Clouds are becoming open.
> They're no longer walled gardens.

Learn more about multicloud

[2] *Flexera 2024 State of the Cloud Report*

# 4 GenAI will assist with cloud native application development

In our 2024 trends report, we called out low-code and no-code platforms that empower people without programming knowledge to create their own lightweight applications to solve business problems. In 2025, look for low-code platforms that can be easily integrated with cloud services, such as your database and analytics system, and that use GenAI to generate code based on natural language prompts: People describe what to create and let the system decide how.

But GenAI can do even more. A 2025 advance is generative development, where GenAI tools augment enterprise development teams by handling increasingly complex code creation. For example, Oracle is working to help dev teams move beyond code assistants to having AI generate application modules that humans can link together, review, and edit. In particular, AI can help architect cloud native applications that are built to take advantage of the cloud's inherent scalability and other benefits by using technologies such as microservices and containers.

For companies looking to refactor applications to run in the cloud, GenAI also offers assistance. LLMs can examine legacy code and help describe how it functions so teams can begin to translate monolithic applications into more robust cloud native apps.

67% of AI decision-makers say their organization plan to increase investment in generative AI in the coming year, per Forrester's May 2024 Artificial Intelligence Pulse Survey.[3]

> Use GenAI to generate code based on natural language prompts: People describe what to create and let the system decide how.

Learn more about cloud native

[3] *Forrester, "Generative AI Trends For Business: Why, When, And Where To Begin" May 2024*

# 5 Data lakes will flow into the cloud and become smarter

Data silos are your enemy—you can't get full value from isolated information—so consider making 2025 the year you commit to gathering diverse structured and unstructured data from all your sources into a central, accessible repository.

Specifically, an intelligent, cloud-based data lake enables you to store, organize, and analyze diverse data types, both structured and unstructured. While many companies established data lakes on-premises, cloud data lakes offer many advantages, including scalability, cost-effectiveness, easier management, and, notably, having all data readily accessible to applications no matter where they run. And the cloud supports another key benefit of data lakes—namely that they ingest data quickly and prepare it on the fly as people request it. When the lake is in the cloud, that access is eased no matter where users are located, and there's plentiful processing power available for data preparation.

Data lakes built on open data standards are better able to handle diverse inputs. You'll get real-time connections, quicker transformations, AI-assisted data prep, and strong role-based access controls—all essential features. By significantly reducing data silos in 2025, you'll set your organization up for more accurate analytics that can drive better business decisions.

> Data lakes built on open data standards are better able to handle diverse inputs.
>
> Learn more about Oracle's data services

# 6 Digital sovereignty will grow as a concern, and cloud providers can help

You're familiar with data sovereignty, which refers to the ability to control and regulate the collection, storage, and use of data within a geographic border. Digital sovereignty is a broader concept, focusing not just on data but on the ability to control the IT infrastructure, policies, standards, and cybersecurity methodologies. If you're not concerned with digital sovereignty, it's likely you have a customer or partner that is, so it's a trend worth exploring.

For organizations that want cloud benefits but need strict localization and control, providers are responding with fully managed dedicated regions that deliver essentially the same experience as public clouds, including scalability and access to a range of services and applications. While you can create dedicated, private connections to the public cloud, a dedicated region is a separate realm managed using its own console, and if desired, by staffers who work for you.

And today's offerings are accessible to smaller enterprises—Oracle Cloud Infrastructure Dedicated Region requires a minimum of just three racks of gear and scales up from there, and Oracle Exadata Cloud@Customer and Oracle Compute Cloud@Customer both have a low cost of entry. By maintaining complete control of your data, you can more easily address data residency, security, and connectivity concerns related to both data and digital sovereignty.

> If you're not concerned with digital sovereignty, it's likely you have a customer or partner that is
>
> Learn more about Oracle Cloud@Customer

# 7 The cloud will be your AI's classroom

GenAI training is notoriously compute intensive and increasingly the realm of specialist vendors that create foundation LLMs, which companies can then fine-tune for specific tasks.

The cloud is the place to take those generic LLMs and make them your own. Cloud platforms provide foundation LLMs and specialized tools for vector database management, retrieval-augmented generation (RAG), and more, simplifying the development process.

## Here are three advances to know for 2025:

Vector databases natively store and manage vector embeddings and handle the unstructured source data vectors describe, such as documents, images, video, and audio. By using AI to create numerical representations that describe data's content—aka vectors—unstructured data becomes highly searchable. In addition, having vectors and the original source data reside in the same database delivers advantages in contextual understanding and data provenance and gives you more flexibility in how you use and analyze information. Searching vectors makes it easier to identify similar content in unstructured or complex business data, such as images of two plants that resemble one another or anomalous content that needs to be examined further.

Retrieval-augmented generation, or RAG, helps optimize the output of an LLM without modifying the underlying model by providing it with up-to-date, organization-specific information. As a result, the GenAI system can provide more contextually appropriate answers to prompts by basing those answers on your data. An LLM on its own can't tell you how different products are selling. An LLM that's provided information via RAG can.

Fine-tuning a general-purpose GenAI model involves providing additional rounds of training on a smaller, domain-specific data set and adjusting the model's parameters based on this training. This helps the model perform better on tasks specific to your organization or industry. A generic LLM by itself can't help much with accounting, for example, but one that's been fine-tuned using FASB rules and other specialized data can supplement finance teams.

Another emerging trend? RAFT, or retrieval-augmented fine-tuning, combines RAG and fine-tuning to help you get even more relevant results in a specific domain. It provides a way to efficiently access external data sources, including your proprietary data. It takes some expertise from your cloud provider, but the results are highly accurate, relevant, context-aware—dare we say, super insightful—outputs.

> An LLM on its own can't tell you how different products are selling. An LLM that's provided information via RAG can.
>
> Learn more about RAG and fine-tuning

# 8 Cloud-based AI analytics adds predictive power

A cloud-based AI analytics platform uses machine learning to uncover hidden patterns, predict future trends, and automate even very complex data analysis tasks. The technology is poised to transform the way businesses make decisions in 2025 by providing faster, timelier, and clearer insights. Predictive analytics has long been used in traditional data analytics and is now benefiting from the speed and compute power of the cloud infrastructures that enable AI analytics. The applications of AI analytics are vast and varied, ranging from business automation and data security to logistics and product design.

Because cloud platforms can store and process massive amounts of data, your tuned and trained AI models can consider many more factors when making predictions. Cloud-based AI tools often take advantage of more sophisticated algorithms and techniques to uncover complex patterns and relationships in data. A cloud native service is scalable, can integrate with enterprise systems, and lets businesspeople and data experts collaborate on the AI analytics process.

## Top AI Use Cases based on 5 Year CAGR (2023-2028) (Value (Constant))[4]

| Use Case | CAGR (5 years) |
|---|---|
| Augmented Claims Processing | +35.8% |
| Digital Commerce | +33.2% |
| Augmented Sales Planning and Prospecting | +32.8% |
| Smart Factory Floor | +32.5% |
| Augmented Product Requirements, Design & Collaboration | +32.2% |
| Others | +28.6% |

> The top AI use cases cited above represent massive opportunities for businesses of all types and sizes.
>
> Learn more about AI and analytics

[4] *IDC's Worldwide AI and Generative AI Spending Guide, V2 2024, August 2024*
Data: The IDC Spending Guide quantifies the AI opportunity by providing data for 42 use cases across 27 industries in nine regions and 32 countries

# 9 Specialized cloud providers (perhaps like you) will fill regional and market needs

Do you have a great idea for a new service or application or a particular need that other providers aren't filling? New cloud infrastructure platforms from hyperscalers let companies become cloud providers themselves. These tools enable the creation of customized cloud environments inside your data centers so you can maximize your control and address data and digital sovereignty requirements.

If you're on the provider side, look for a system that lets you market your service under your own brand, with full control over where data is located and your customer relationships and the flexibility to set your own pricing, rate cards, account types, and discount schedules.

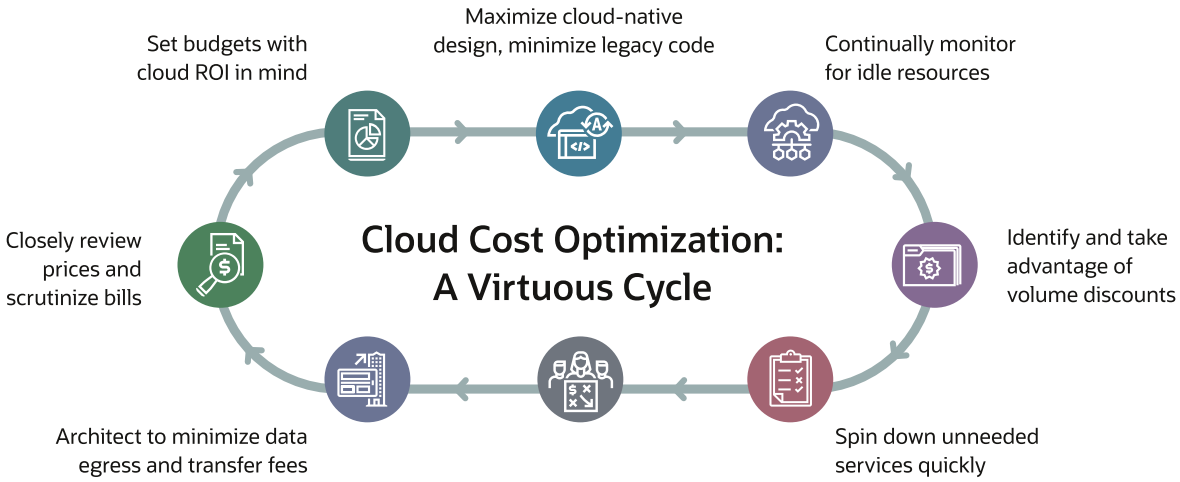
If you're in the market for a highly specialized or localized application, consider unconventional providers that are spinning up services on these cloud platforms. As an example, users of the Oracle Alloy platform have access to the same development and security tools Oracle uses to build its own services, which they can pair with a deep understanding of a particular market, region, or industry. It's a win-win, with customers getting bespoke services from a specialist with the assurance of a hyperscale cloud infrastructure.

> If you're in the market for a highly specialized or localized application, consider smaller and unconventional providers
>
> Learn more about Oracle Alloy

# 10 Cost transparency and optimization will improve

In 2025, your IT and finance teams will gain more timely and detailed insights into your cloud spending and find new help identifying areas for optimization thanks to advanced cloud cost analytics tools powered by GenAI. These systems continually monitor and analyze cloud resource use and recommend ways to reduce spending without hurting performance. They also make it possible to allocate costs to specific teams, projects, or applications and automate reporting. That way, each department knows exactly what costs it's incurring for compute resources. This sort of insight is critical for calculating metrics like cost of goods sold and for any company considering chargebacks or showbacks.

GenAI cost optimization tools can also get prescriptive. In 2025, it will be possible to, for example, receive frequent rightsizing recommendations on optimal configurations for your specific mix of services and alerts on discounts for long-term commitments to reserved resources.




> Minimizing egress fees is all about positioning your data where it's needed. New intercloud partnerships make that much easier.
>
> Learn more about cloud cost optimization

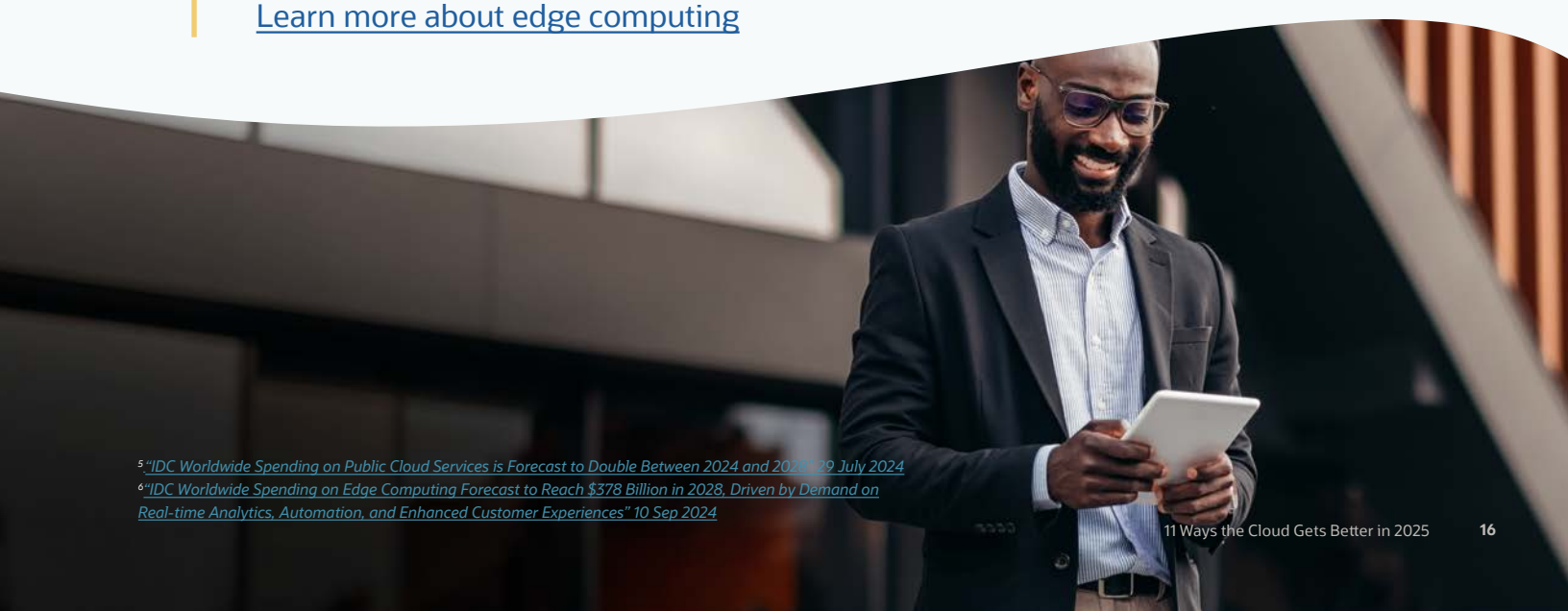
# 11 “Cloud first” will accelerate, especially at the network edge

For 2024, IDC forecast [5] worldwide spending on public cloud services of US$805 billion, with spending expected to double by 2028. What that means is a sharp upward trajectory in workloads moving to the cloud, starting in 2025. Much of this growth is driven by the trends we’ve discussed, especially AI. But a need for better data management and an increase in edge computing will also play a role, says IDC.

While we didn’t include edge computing as a standalone trend, innovation is picking up there too. Consider improvements in portable hardware platforms providing cloud-integrated compute and storage at the edge of networks and in disconnected locations. With enough power to drive data-intensive and AI workloads, these ruggedized devices will enable use cases like predictive maintenance on oil rigs, real-time intelligence analysis in the field for first responders, and active monitoring of patient data in remote hospitals. It’s all powered by the cloud.

**$378B** Forecast global spending on edge computing by 2028, per IDC.[6]

> One key to success for applications at the network edge is low latency, and that means direct links to the cloud.
>
> Learn more about edge computing



[5] *“IDC Worldwide Spending on Public Cloud Services is Forecast to Double Between 2024 and 2028,” 9 July 2024*
[6] *“IDC Worldwide Spending on Edge Computing Forecast to Reach $378 Billion in 2028, Driven by Demand on Real-time Analytics, Automation, and Enhanced Customer Experiences” 10 Sep 2024*

# How Oracle helps

In case you hadn't noticed, a theme for 2025 is the power of the cloud plus AI. With Oracle Cloud Infrastructure (OCI), Oracle delivers a comprehensive AI portfolio to help you take advantage of AI in the way that makes the most sense for you. Additionally, using its distributed cloud, OCI provides a broad set of deployment options for AI that are unmatched by other cloud service providers. For example, OCI makes it easy to bring AI-generated insights to your key business functions by embedding AI in Oracle Fusion Applications. To help you build AI into your own applications, OCI features a broad array of AI services with models that can be customized using your own business data.

AI wouldn't be where it is today without cloud computing, and many of the cloud advances we're looking forward to for 2025 will come courtesy of AI. Cloud providers such as Oracle offer the compute architectures needed to train a thriving mix of AI models, and they're building on-ramps that allow you to take advantage of AI's growing capabilities.

**Learn more about OCI**

Join us at CloudWorld Tour 2025

## Connect with us

Call +1.800.ORACLE1 or visit oracle.com

Outside North America, find your local office at oracle.com/contact