Live On the Edge

How can companies leverage the combination of edge computing & AI to foster sustainability?

By Raju Chellam

Here's a convoluted corporate story: During a company's annual town hall, the CEO proudly discussed their perceived market-leading performance and announced new solutions on the cloud and at the edge. An employee raised a concern: "What steps are we taking to ensure our data isn't compromised? Aren't we worried about cloud security?" The CEO turned to the CSO (Chief Security Officer). "No, we're not worried about cloud security," the CSO blithely replied. "Our data is so disorganized that even we are unable to find anything of value."

If that anecdote made you blink, these statistics should make you think: The global spending on edge computing is set to reach US$378 billion by 2028 – up from roughly US$200 billion in 2023 and US$228 billion in 2024. That's a CAGR (compound annual growth rate) of 12% between 2023 and 2028,

as per IDC estimates. It includes combined enterprise and service provider spending on hardware, software, professional services, and provisioned services for edge solutions.

Edge computing is a distributed model that brings computation and storage closer to the sources of data. Instead of sending data to a centralized data center or cloud for processing, the data is churned and analyzed locally, near where it is generated. This approach helps reduce latency, improve response times, and save bandwidth.

It's akin to a large retail chain which sends the inventory data from each store to a central warehouse (the cloud) for processing. This can be slow, especially during busy shopping seasons. With edge computing, it's like having mini warehouses (edge devices) at each store. These mini warehouses can process inventory data locally, right where it's generated. This means

> ❝ MIGRATE WORKLOADS FROM OWNED FACILITIES TO COLOCATION FACILITIES OR THE EDGE; OR EMBRACE AS-A-SERVICE MODELS FOR PHYSICAL INFRASTRUCTURE AS WELL.

faster restocking, better inventory management, and quicker responses to meet local customer demands.

Why live on the edge now? Because, despite public cloud usage being ubiquitous, many deployments are ad hoc and poorly implemented. "The focus of refactoring cloud infrastructure should be on optimizing costs by eliminating redundant, overbuilt or unused cloud infrastructure," says Gartner. "Build business resilience rather than service-level redundancy. Use cloud infrastructure to mitigate supply chain disruptions. Only then 65% of application workloads will be optimal or ready for cloud delivery by 2027, compared with 45% in 2022."

Moreover, data centers are shrinking and migrating to platform-based colocation providers. Combined with new as-a-service models for physical infrastructure, this can bring cloud-like service-centricity and economic models to on-premises infrastructure. Gartner says 35% of data center infrastructure will be managed from a cloud-based control plane by 2027, from less than 10% in 2022.

"Infrastructure and operations professionals should focus on building cloud-native infrastructure within the data center," Gartner advises. "Migrate workloads from owned facilities to colocation facilities or the edge; or embrace as-a-service models for physical infrastructure as well."

AI ON EDGE

The additional variable these days is AI. "Enterprises are now accelerating their investments in edge and AI to drive real-time analytics, automation, and enhanced customer experiences, particularly in manufacturing, utilities, healthcare, and retail," says Alexandra Rotaru, IDC's analytics manager for Europe. "Technologies like AI-powered devices, edge servers with GPUs, and 5G connectivity are gaining traction, enabling companies to process data closer to the source for higher performance."

The moot benefit is lower latency. "As the focus of AI shifts from training to inference, edge computing will be required to address the need for reduced latency and enhanced privacy," says Dave McCarthy, IDC's research vice president for cloud and edge services. "This trend will not only optimize operational efficiencies but also foster new business

models previously impossible with centralized infrastructure. Distributing applications and data to edge locations can enable faster decision-making with reduced network congestion."

What about the Asia-Pacific region? Enterprise and service provider spending on edge solutions is on track to hit US$81 billion by 2027 – up from US$43.7 billion in 2023. Mobile edge attached to the 5G network is a strategic focus area. "Some of the mobile edge nodes will also be located in 5G private wireless networks with single tenants," says Bill Rojas, IDC's research director for the Asia-Pacific.

IDC has identified more than 400 use cases for edge computing and lists three that will see the largest investments this year: CDN (content delivery networks), virtual network functions, and MEC (multi-access edge computing).

"These are foundational to service providers' edge services offerings," IDC notes. "These three use cases will account for nearly 20% of all edge spending. Service providers will invest more than US$9 billion in total to enable edge offerings in 2023-2024. MEC will see the fastest growth and is critical for supporting low-latency communications required by next-gen apps powered by the spread of 5G networks, IoT, and AI."

TEN TIPS

So how can organizations leverage the combination of edge computing & AI to foster sustainability? Here's my list of ten potential applications – in alphabetical order – for you to consider:

- **Agriculture:** Implement edge computing in smart farming to monitor soil conditions, weather, and crop health. Optimize resource usage to increase yields.
- **B2C:** Deploy AI-powered edge devices in retail stores to analyze customer behavior and preferences, thereby enhancing personalized shopping experiences.
- **Cars:** Use AI at the edge to process data from sensors and cameras locally for safer driving in autonomous vehicles – or use AI for real-time diagnostics locally in non-autonomous vehicles.
- **Deliveries:** Factor AI at the edge to track and manage logistics inventories in real-time and

> ❝ EDGE COMPUTING CAN HELP WILDLIFE OFFICIALS AND PARK RANGERS IDENTIFY AND STOP POACHING ACTIVITIES, SOMETIMES BEFORE THESE OFFENSES EVEN OCCUR.

space, optimizing supply chain operations and reducing delays.

- **Energy Management:** Implement AI at the edge to monitor and manage energy consumption in high-power equipment, leading to more efficient energy use and cost savings.
- **Financial Services:** Use AI at the edge for real-time AML (anti-money laundering), fraud detection, and risk management, boosting security and customer trust.
- **Government:** Deploy AI-powered edge devices for surveillance and emergency response, improving public safety and reducing response times. Implement AI-driven edge computing in urban infrastructure to manage traffic, utilities, and public services more efficiently.
- **Healthcare:** Utilize AI at the edge for remote patient monitoring and diagnostics, providing close to real-time health data analysis to improve patient care in hospital and home-care settings.
- **InstaCom:** Enhance network performance and reduce latency by processing data at the edge with AI, improving telecom QoS. Enhance content delivery networks with AI at the edge for faster and more reliable entertainment and media streaming services.
- **JIT:** Deploy AI-driven edge devices to monitor and optimize production lines for JIT (just-in-time) ops, thereby reducing downtime and improving efficiency in smart manufacturing settings.

SUSTAINABILITY ON EDGE

Can edge computing also benefit the environment? Yes, according to a study by IBM Corp. "One direct way is by using edge computing to monitor protected species of wildlife inhabiting remote places," IBM says. "Edge computing can help wildlife officials and park rangers identify and stop poaching activities, sometimes before these offenses even occur."

Another relates to energy management. "Edge computing supports the use of smart grids, which can deliver energy more efficiently and help businesses leave a smaller carbon footprint," IBM notes. "Grid or distributed computing is where a group of machines and networks work together for a common computing purpose. Resources are utilized in an optimized manner, thus reducing the amount of waste that can occur when large quantities of power are consumed."

More significantly, edge computing can also support the remote monitoring of oil and gas assets. That's no small feat given that some locales where oil is drilled can be rugged or dangerous, such as on ocean floors. Edge computing can help with real-time analytics and do that closer to the specific asset, limiting the need for cloud connectivity as well.

Regulators are increasingly forcing tough sustainability standards on urban planning and design. Civil engineers can leverage AI and edge computing to craft urban designs for smart cities and drive civic innovation and sustainable use of resources. "Edge devices can be used to process usage data in the field," IBM advises. "For instance, municipalities can use edge devices to capture data from public infrastructure, power grids and other sources where urgent action is needed – and avert a crisis."

The bottom line: By harnessing the combined power of AI and edge computing, enterprises and governments can drive sustainability by enabling real-time, intelligent decision-making at the source, reducing environmental impact, conserving resources, and fostering a resilient, eco-friendly future.

Since we started with a corny corporate story, let's end with another: At the company's quarterly meeting, the CEO proudly announced new AI-driven solutions that will run on the edge. "How are we ensuring that our data is secure with these new technologies?" An employee asked. "Is our data encryption strong enough?" The CEO turned to the Chief Information Officer. "Don't worry about that," the CIO said. "Our data is so powerfully encrypted that even our own AI can't figure it out." 🄳🄠

Raju Chellam is a former Editor of Dataquest and is currently based in Singapore, where he is the Editor-in-Chief of the AI Ethics & Governance Body of Knowledge, and Chair of Cloud & Data Standards.

maildqindia@cybermedia.co.in