NUTANIX

The Future of AI Computing Resides at the Edge

Historians may one day describe 2023 as the tipping point for artificial intelligence (AI). Although consumers have benefitted from AI and machine learning for years — via ridesharing apps, search engines, social media feeds and even Netflix recommendations to name a few — this year's debut of ChatGPT and release of other generative AI tools opened new potential for computing's role in many aspects of daily life.

Business strategists and decision makers are exploring how they can use generative AI and machine learning to improve processes and productivity, and deliver new services. Many believe that organizations that successfully integrate these automated capabilities into their operations will continuously improve their ability to turn big data into transformational business outcomes.

The challenges are as grand as the opportunities. Organizations must acquire or upskill the talent then grow their knowledge and IT infrastructure that are needed to support these new digital capabilities. Without that, it's impossible to collect, store and manage the troves of training data that are required to produce actionable results from modern machine learning models, according to international computer science expert Dr. Douglas Comer, a distinguished professor of computer science at Purdue University.

Comer talked to The Forecast about findings in his 2021 book, The Cloud Computing Book: The Future of Computing Explained.

> AI isn't just a way to be more efficient. Because AI-generated analyses can improve the decision-making process, it's also a way to be more effective

Machine Learning Provides a Competitive Advantage

Thanks to the rise of AI-based technology — including large language models, translation software and voice recognition, just to name a few — machine learning will soon be as fundamental to business as word processing, internet browsing and mobile apps. For enterprise leaders, however, AI isn't just a way to be more efficient. Because AI-generated analyses can improve the decision-making process, it's also a way to be more effective, according to Comer.

"Think of a CFO of a company: They have to make all sorts of important decisions and plan for the company's financial future," he told The Forecast in an exclusive interview. "They have to know about things like bank rates for loans. If it's a multinational corporation, they might have to know about exchange rates or what's going on in other countries. They have to know about stock prices and dividends so they can deliver a fact-based recommendation to the CEO."

Clearly, a lot rides on the decisions of CFOs. With the help of AI and machine learning, they and other C-level executives can make better choices, faster. This ultimately creates a competitive advantage, according to Comer. He said AI can produce real-time insights that human analysts might miss and can do it in a fraction of the time.

Why Cloud Computing is Critical

Comer is author of the seminal computer science series, Internetworking with TCP/IP; a former member of the Internet Architecture Board; a fellow of the Association for Computing Machinery; and an inductee to the Internet Hall of Fame. He believes AI doesn't just represent the future of business. It also represents the future of computing. He sees the widespread adoption of machine learning will have profound consequences for cloud computing, in particular, by dramatically increasing demand for remote data storage.

"Machine learning is based on statistics, which means scientific computation," Comer said.

Large-scale scientific computation requires intensive computing power and large volumes of training data.

"That spells cloud computing," he said. "Cloud data centers have lots of processing power and you can get as much as you want to pay for."

He expects the demand for cloud computing to continue rising as companies lean on data-heavy technologies like machine learning and artificial intelligence.

Edge Computing Fuels Machine Learning

Despite the clear benefits, there are drawbacks associated with using cloud data centers to store and process machine learning datasets.

"We all know that in lots of industries, faster decisions give a competitive advantage," Comer said. "Financial companies that do stock trading will tell you that a millisecond can offer advantages. And we're not talking dollars. It's millions of dollars in a very high-stakes environment."

For organizations that value low latency in machine learning, as well as local control over data, there's a solution that can supplement traditional cloud data centers: edge computing.

"Edge computing means you build multiple small data centers near the sources of data," Comer said. "Then there's another idea called fog, which means you have intermediate data centers."

Collectively, cloud data centers, intermediate fog computing and edge computing form a hierarchy where the lowest form of analysis is performed on-site with immediate results. Data is then aggregated and sent up the hierarchy for additional, more intensive analysis and for training and refining models.

"Additional — especially in aggregate — analysis can be completed in the fog data centers, with results sent to the cloud for the repository and ultimate computation for the overall company," Comer explained.

While he believes adoption will accelerate over the next decade, Comer makes it clear that edge computing will never replace traditional cloud data centers.

"If we can process data simply and quickly put it at the edge, great," he said. "But if we need to have real intensive long-term computation, and we need to keep large amounts of data, that's not going to fit in an edge data center."

Edge Computing and Cybersecurity

Given the risks that digital businesses face, one can't discuss the future of computing without also discussing cybersecurity. Again, edge computing has distinct advantages and disadvantages, according to Comer.

"If you can do processing in the edge center and, in particular, if you can do encryption in the edge center, then all of the data moving upstream to the cloud … is encrypted," he said.

By encrypting data before it leaves the edge computing facility, organizations can prevent sensitive data from being intercepted in transit. This type of approach is especially attractive to organizations that handle data in highly regulated industries like health care, financial services and critical infrastructure sectors, where data breaches come with stiff penalties. And yet, edge computing is not a cybersecurity cure-all.

"More edge data centers actually increase the attack surface," Comer said. "[Because] these are physical locations, someone might be able to break in … These facilities could be sitting out in the middle of nowhere, are automated, with nobody around to see you."

Edge data centers must therefore be physically secured to prevent unauthorized access.

Moving from Cutting Costs to Generating Value

Along with increased speed and security, there's another reason to embrace edge computing: cost.

But that's the wrong reason, said Comer. He worries that cost-conscious leaders might declare, "We can use slower processors that are much cheaper. We can also use less data, which is much cheaper." This approach might work for some things, but it could impact speed and reliability.

Looking at edge computing as a way to replace cloud data centers and save money is a mistake. The real value in edge computing is not cost savings, but the ability to generate more accurate AI results more quickly by using a larger volume of data on-site — a critical advantage for executives who use these insights to guide high-stakes decision-making.

"[Edge computing is] not going to ever overtake cloud data centers, because there are lots of things that you can't do in an edge data center," Comer concluded. "But it is going to be a supplement because you can do things in an edge data center that you can't do in the cloud … It's all about latency."

To learn more about how Nutanix provides AI-ready infrastructure that is easy to get started and run across both your core and edge, visit Nutanix.com.

This article first appeared in The Forecast by Nutanix. For more news about people and trends shaping our future, visit The Forecast by Nutanix.