

IDC TECHNOLOGY SPOTLIGHT

Brought to you by Tech Mahindra, Powered by IDC

A developer-friendly services partner is necessary for successful execution of edge artificial intelligence (AI) initiatives wherein firms embark on gaining unprecedented insights from data generated at the network's edge using AI-enabled applications.

Delivering Edge AI in a Digital Economy

September 2019

Written by Ashish Nadkarni, Group Vice President, Infrastructure Systems, Platforms, and Technologies

Introduction

Digital transformation (DX) — the application of a technology-centric business vision and accompanying strategies — is an essential mandate for businesses to thrive in a digital economy. DX initiatives allow firms to create value and extend their competitive advantage through new products and services, new business relationships, improved customer experiences, and increased operational efficiencies. DX is reaching a macroeconomic scale and is at the heart of business strategies for companies of all sizes. IDC forecasts that:

» By 2020, at least 55% of organizations will be digitally determined (leaders of transformation that have aligned the necessary elements of people, process, and technology), transforming markets and reimagining the future through new business models and digitally enabled products and services.

AT A GLANCE

KEY STAT

By 2020, at least 55% of organizations will be digitally determined, transforming markets and reimagining the future through new business models and digitally enabled products and services. Al at the edge will feature heavily in the strategy of such firms.

KEY TAKEAWAY

Invest in a partnership with an IT services organization that has built deep expertise in edge AI.

» By 2019, 40% of all DX initiatives will be related to artificial intelligence (AI). This is driven by the fact that all data that comes from DX initiatives will have limited value without exploiting the power of AI to extract valuable, accurate, and timely insights from them. AI will be the key technology that will propel organizations through DX.

At the same time, innovation accelerators such as the Internet of Things (IoT), industrial automation, and smart manufacturing continue to fundamentally transform economies worldwide. The era of "connecting the unconnected" is leading to a phenomenon in which compute is moving closer to data, leading to entire industries such as manufacturing, healthcare, retail, and entertainment becoming smarter, more efficient, and more resilient. While a large number of verticals are witnessing the shift from cloud to edge, the trend is being led by those in manufacturing, automotive, aerospace, and healthcare.

In fact, IDC estimates that by 2020, at least 55% of organizations will be digitally determined, transforming markets and reimagining the future through new business models and digitally enabled products and services. IDC anticipates AI at the



edge will feature heavily in the strategy of such firms. In preparation, businesses are increasingly shifting their decisionmaking processes to be distributed, with the edge playing a crucial role. Some concrete examples are as follows:

- » In autonomous driving, vehicles need to detect lane markings, other vehicles, pedestrians, and traffic signs with every incoming camera frame.
- » In manufacturing assembly lines, defects in production output need to be identified within milliseconds to allow that part to be rejected or reworked without holding up the line.
- » In hospital emergency rooms, surgeons operating on injured patients need to make instant decisions aided by Al-based diagnostics regarding the best treatment to administer.

The era of edge computing — which is part of a bigger DX vision — enables firms to obtain valuable insights from data collected via networks of connected devices (beyond computers, laptops, smartphones, and industrial robots). These initiatives enable firms to digitize and better control the "things" they rely on to conduct business. This includes the ability to:

- » Gain unprecedented operational efficiencies through visibility into their product and/or services "supply chain" and exert greater control over it.
- » Better understand the digital "habits" of their customers, specifically their purchasing behaviors, preferences, and tendencies.
- » Change the way they engage with their customers and deliver a consistent, high-quality, and rich customer experience.
- » Conduct lightning-fast business transactions and analyze data gathered from such transactions in real time.
- » Find ways to further improve the quality of their products and services by feeding relevant and actionable information back into the business in a timely manner.

Firms in industries such as transportation logistics, healthcare, hospitality, insurance, manufacturing, and retail are already taking advantage of edge computing initiatives to gain greater efficiency, rapidly go to market with new products and services, develop new customer engagement models, and enhance customer experiences. Examples include:

- » Equipment on factory floors; devices that automate the supply chain; shipping containers; aircraft, trains, and automobiles; and specialty devices such as cranes, elevators, and heavy earth machinery
- » Healthcare measurement devices, home automation devices and appliances, media devices and entertainment consoles, and specialized utility vehicles such as tractors that allow firms to interact with their customers

Businesses are eager to check off the box that declares that they are all in with AI, but this can result in implementing AI without understanding the return on investment. Firms need to identify use cases and applications that can be AI infused, enabled, optimized, or native. This can be a long list, so businesses must prioritize where and how to implement AI in an orderly manner.

The challenges of implementation are exacerbated when businesses conflate AI with edge computing. The need for extracting or delivering timely insights in a time-critical and localized manner will drive the intersection of AI and edge computing. IDC defines "edge AI" as a strategy that organizations pursue to deliver AI at the edge of a network. A modern infrastructure supporting a firm's digital journey must support edge AI.



There is only one problem: If tying AI initiatives to the firm's DX strategy introduces a complex task of measuring business outcomes before and after the AI endeavor, then the task of introducing edge AI makes this problem extremely complex. IDC forecasts that by 2024, a 7% rise in AI-based IT implementation project automation will drive a new wave of business process redesign, requiring services from firms with deep industry and functional expertise.

Fortunately, by partnering with IT services organizations that have built deep expertise in edge computing, artificial intelligence, and edge AI services, firms can build coherency and consistency into their edge AI initiatives. This in turn enables firms to focus on their "big picture" business outcomes versus getting mired in tactical implementation issues.

Benefits

There are several short-term and long-term benefits to firms partnering with IT services organizations with deep edge AI expertise. At first glance, edge computing deployments appear straightforward. Firms assume that deploying an edge AI infrastructure and pushing out applications to it should not be any different from deploying their "core" infrastructure, which is typically the primary, corporatewide IT infrastructure. In reality, on the infrastructure side, edge computing is a multitiered mix of assets arranged in a use-case and workload-centric fashion.

An "intelligent" edge tier is a crucial link between the core and endpoints that not only provides a distributed compute, data persistence, and network aggregation layer but also serves as the intermediary analytics of collected data. For operational technology (OT) use cases, the edge tier hosts actuation and control equipment, while in communications technology (CT) use cases, it can even interface with communications equipment such as booster and relay equipment. The lack of industrywide standards and architectural approaches for edge computing means that today most of the edge deployments are highly custom in nature. When an additional layer of complexity is added in the form of AI-enabled applications, the situation can quickly get out of control. Consider the need to maintain the following key practices — and the implication on a firm's ability to conduct its business should these practices fail:

- » **Infrastructure architecture** involves the selection of an appropriate computing platform, along with connectivity and data persistence. The platform must support bare-metal and virtualized workloads. Organizations must treat the edge as an extension of their core from an architecture perspective.
- » Asset and application management requires control of asset sprawl, which is alleviated by a cataloged deployment and management IT, OT, and CT applications. Deploy a software-defined infrastructure solution that treats the edge assets like a cloud.
- » Data management and governance involves the life-cycle management of data generated or collected at the edge. Ephemeral data has to be analyzed and then disposed of, and what is persistent has to be protected. Organizations must define a core-edge-endpoint data management paradigm that ensures that data is managed according to its value.
- Infrastructure security requires managing device, user, application, and data security. It requires that edge devices be managed according to corporate governance, risk, and compliance policies. Here, organizations can take a multipronged "always on" approach to security.

These are not trivial issues — they require a partner with deep expertise in infrastructure management, application development, security, and automation.



Considering Tech Mahindra for Edge AI

Tech Mahindra is a multinational provider of IT, networking technology solutions, integrated engineering solutions (IES), and business process outsourcing (BPO) to various industry verticals and horizontals. With operations in 90 countries and over 122,000 employees worldwide, Tech Mahindra is one of the fastest-growing strategic IT services firms.

Tech Mahindra has amassed deep expertise in helping its clients build and operationalize AI applications. It has also helped clients deploy edge computing infrastructure as part of their DX initiatives. It has now combined these domains to develop industry-leading edge AI services expertise. Its AI services are developed at an enterprise-grade lab in Fremont, California. Key differentiators of Tech Mahindra's approach to edge AI services are as follows:

- » An experienced team of IT practitioners, advisors, and developers tasked with building innovative artificial intelligence and machine learning (AI/ML) solutions and services utilizing the latest algorithms and state-of-the-art toolkits and simulators
- » Deep expertise in a variety of use cases and industry verticals that include manufacturing, automotive, aerospace, healthcare, semiconductor, oil and gas, banking and financial services, retail, telecom, media and entertainment sector
- » Integrated hardware and software stack deployment capabilities that ensure that edge AI deployments are consistent with edge computing and cloud infrastructure strategies
- » A cultural approach to techno-functional teamwork and passion that addresses client requirements from a unique perspective
- » Investment in Tech Mahindra–owned IP and asset development to accelerate AI/ML deployments supplemented by strategic investments in start-ups that are creating industry-leading IP

Tech Mahindra is actively investing in edge AI in the following verticals:

- » Automotive. The automotive industry is undergoing one of the most profound changes in its history with the advent of electric vehicles, connected cockpits, autonomous driving, and monetizable recommendations. The two main drivers of edge AI in automotive are the need to make decisions at millisecond-level granularity and the scarcity of wireless bandwidth, typically bounded by available data plans. The former limits the volume and rate of data that can continually be streamed from the vehicle to the cloud while the latter requires a high level of local intelligence at the vehicle.
- » **Manufacturing.** The manufacturing sector is undergoing profound change in several areas, including predictive maintenance, cognitive defect inspection, robotics, assembly line efficiencies, and inventory management. While the technologies for these applications are quite diverse, there are two common threads: They all benefit from the use of AI and ML, and most manufacturers have very strict policies restricting the transfer of data outside their enterprises. This places a strong requirement on edge AI for these manufacturers that is driven by both corporate policy and applications that need real-time local intelligence.



» Healthcare. As with the automotive sector, healthcare is undergoing profound change enabled in a major way by AI. The healthcare sector has an added requirement, which is the confidentiality of patient data. Indeed, HIPAA regulations are very stringent in this regard, and physicians, hospitals, and insurance companies have to comply with strict privacy guidelines. The area of automated diagnostics is undergoing very fast adoption (see Figure 1). In many cases, physicians performing a procedure need real-time AI-based assessments that factor in a multitude of variables to enable them to make the right decisions. In other scenarios, the diagnoses need to be conducted locally (at the edge) because the patients' data cannot be transferred outside the hospital premises. In addition, the healthcare industry is moving gradually in the direction of in-home AI-assisted self-diagnoses by patients, which require home-based intelligent medical devices. In short, edge AI is becoming increasingly pervasive in healthcare.



FIGURE 1: AI-Based Remote Diagnosis

Source: Tech Mahindra, 2019



- » Aerospace/unmanned aerial vehicles (UAVs). In the aerospace industry, there are multiple growing trends, including streaming services and UAV-based services. With the latter, there is a growing market for autonomous drones that can perform useful operations such as perimeter security, defect inspection, warehouse inventory management, and short-range delivery. In all these applications, the UAV has cameras and other sensors that feed the AI algorithms for assessment of the situation. In most cases, both the assessment and the response must be executed swiftly, often within milliseconds. This requires edge AI.
- Supply chain management. In supply chain, there are three broad categories: inbound inventory operations, outbound inventory operations, and distribution to outlets. It is critical for all three to execute optimally to minimize inefficiencies and to maximize profits. Although cloud-based AI approaches were used initially for supply chain management, there are several benefits to AI models that are a hybrid of edge and cloud. A major focus of supply chain management is process optimization, which in turn benefits from rapid assessment and response. Although edge AI is not critical, an edge-cloud hybrid architecture will typically result in the best solution.
- » **Retail management.** The in-store retail industry is progressing in the following directions: better understanding customers' in-store behavioral patterns, experimenting with in-store marketing to see what resonates with customers and the characteristics of those customers, and cash register opt-ins to receive targeted marketing material that is aligned with the customers' preferences. Edge AI plays a major role in the first two previously mentioned areas in being able to tag customers as they enter a store, identify their demographic characteristics anonymously (age bracket, gender, ethnicity, etc.), track them as they move around the store, and identify the items they pick up for checkout.
- Banking and financial services. Security has always been of utmost importance in the banking and financial services sector, and it is critical to detect and respond to threats instantly. Malicious attacks can come from two sources: bad actors that aim to hack into enterprises' databases to steal data and malware that's injected into enterprises with the goal of crippling operations. In either scenario, there is a need to detect the attack quickly and take preventive action. Such real-time responses can be effective only if there is local intelligence at the edge that detects the attack and immediately generates an alert. A more comprehensive investigation can be conducted in the cloud once initial preventive measures are taken rapidly (see Figure 2).



FIGURE 2: AI-Based Fraud Detection



Source: Tech Mahindra, 2019

Examples of Recent Tech Mahindra Edge AI Deployments

Tech Mahindra provides a turnkey solution for edge AI deployments. A differentiator of Tech Mahindra's edge AI solution is that it focuses on use cases that require heavy customization and optimization as part of the solution. Its expertise becomes evident when such customization and optimizations transcend software and hardware stacks.

The following are examples of actual projects involving edge AI that Tech Mahindra implemented for its customers worldwide:

» AI-Based Legal Assistance. As a value-added technology consulting company, Tech Mahindra engages with hundreds of clients and partners. Tech Mahindra's legal department expertly handles nondisclosure agreements (NDAs), partnership agreements, master services agreements (MSAs), statements of work (SOWs), and trademarks/patents with all these organizations, covering all aspects of the relationships, including revenue sharing, technology



implementation, work detail, timelines for delivery, payment terms, and forfeiture terms. This paperwork used to be a very laborious process at Tech Mahindra, with endless variations from one agreement to the next. Tech Mahindra implemented a smart automation solution that learned from a large trove of historical contracts to identify critical clauses and identify important decision parameters in these documents. The final product reduced legal document processing times from three hours to a few seconds, with a significant improvement in accuracy. This solution ison track to being universally deployed within Tech Mahindra for its contracts.

- » AI-Based Telecom Network Operations. Tech Mahindra AI for Telco Networks offers an open platform for network fault predictions. It is designed to increase network availability using industry-standard open source—based AI tools and frameworks. The solution provides auto—fault remediation capabilities and can monitor key performance indicator (KPI) status for potential breaches. It uses a data lake for providing a single, centralized, and trusted source of information for access, core, and transport domains. Views can be customized to different audiences — from an executive-level view to deep technical views for operators.
- » Cognitive Signature Detection and Verification Verification of physical signatures was a huge problem for a Fortune 500 multinational bank and financial services company with over 55,000 employees because of the large number of transactions it executed every day, including those at point-of-sale terminals. Tech Mahindra implemented an edge AI solution that would cognitively detect a human signature and match it up with its reference with superior low false positive and false negative rates. This solution is leading to a huge savings in processing time, in addition to a reduction in cost and an improvement in accuracy of signature matching.
- » Drone-Enabled Cognitive Inspection (DECI). A multinational aircraft manufacturer headquartered in North America was facing a problem in the inspection of aircraft for dents in the fuselage, cracks, corrosion, and damage. Commercial aircraft generally have hard-to-reach surfaces for comprehensive inspection, and many problems were going unnoticed for too long. Tech Mahindra implemented a deep learning—based solution with high-definition cameras mounted on drones to visually inspect aircraft, capture high-resolution images, and detect damage or defects in real time. This is leading to great savings for the customer by reducing around 50% of defects that would otherwise slip through, an 80% reduction in inspection time, and a 40–70% reduction in manual inspection efforts.
- » **Digital Freight Matching.** Representatives of a leading logistics provider and load brokerage firm based in the United States were spending a lot of time and manual effort on matching given cargo loads with transportation carriers. Tech Mahindra implemented an AI-based recommendation engine that was trained on historical data and learned to match various types of loads with the appropriate carriers, optimizing various constraints such as cost, delivery time, cargo type, and weather. This led to an improvement in the quality of matches and reduced the manual effort of the agents. Furthermore, the AI recommendation engine that was built was designed to learn continually so that it keeps getting better as increasing amounts of data become available to it.
- Cognitive Assembly Inspection. A United States-based medical company that develops and provides blood component and cellular technologies with a combination of apheresis collections, cell therapy systems, and pathogen reduction was facing a challenge in the frequently incorrect assembly of apheresis equipment that led to unreliable diagnostic results. Tech Mahindra implemented a cognitive assembly inspection system that visually inspects the equipment and its connections and provides an instant automated verification of correct assembly. Technicians are alerted upon incorrect assembly and can fix it before the physician proceeds with the diagnosis. This has resulted in improved process quality and increased control, leading to significantly better diagnoses.



- » Al-Driven Intelligence Business Operations. A leading telecom provider in Germany was facing a problem with the quality of its customer support operations. Issues were logged with insufficient detail, and the mean time to resolution (MTTR) was far higher than acceptable. Tech Mahindra implemented an edge Al-based solution that was able to learn from the customer's past customer relationship management (CRM) records and improve customer support efficiency. It uses natural language processing (NLP) to identify the types of incidents, classifies the incidents into one of several categories, and provides recommendations to address the problem. This has improved the quality of customer support and reduced the MTTR and average handle time (AHT) by a significant margin.
- » Virtual Agent for Intelligent Patient Engagement. A United Kingdom–based multinational pharmaceutical company faced problems related to patients' inappropriate use of medicines contributing to anti-microbial resistance, lack of effective reporting of adverse effects, and long lead times for providing advice to patients. Tech Mahindra implemented a conversational AI assistant that was able to understand the patients' queries, track the use of medicines and associated compliance, and provide automated advice that factored in the patients' condition combined with the characteristics and side effects of the medicines. This led to superior administration of medicine, improved patient compliance, and on overall improvement in the satisfaction of the patients as well as physicians.

Challenges and Opportunities

CIOs and other IT leaders must prepare their wider organization and key stakeholders for the necessary steps to successfully deploy and manage an edge computing infrastructure, complete with AI and ML capabilities as necessary. Firms that acknowledge edge AI as a long-term investment generally fall into the following categories:

- » Firms that have embraced edge computing early on. They have invested in custom or semi-custom approaches and are well on their way to realizing the business benefits of moving compute closer to data. Many of the early adopters of the Internet of Things fall into this category.
- » Firms that are still on the fence with edge computing. They are evaluating the way in which they can deploy edge computing in production, the changes they have to make their IT processes, and how they can reap long-term business benefits. These firms will generally take an approach that involves an industry-standard infrastructure.

(Note: This classification does not include "traditional" edge deployments such as remote and branch offices. Rather, it examines edge deployments for newer use cases such as the Internet of Things and by telecom service providers, oil and energy providers, and retail merchants.)

For firms in the second category, IDC recommends rigorous due diligence on a build versus partner strategy as a part of their edge AI investments. Avoiding this rigor could result in wasted spend on an incomplete and suboptimal solution or lost revenue and additional costs to address crises that arise from an incomplete or improperly implemented infrastructure. Organizations must evaluate partnerships with IT services firms especially when heavy hardware and/or software optimizations are involved.

Edge computing forms an important tier in an organization's next-generation infrastructure. It can offer enormous benefits to a firm when deployed in the right context, for the right use case, with the appropriate architecture, and with the right investments in technologies.



Conclusion

Firms that thrive in the digital economy are those that can bridge the divide between their core and edge strategies. They are able to effectively combine artificial intelligence and edge computing to create a business outcomes–based edge AI strategy. They can formulate and execute on an edge computing strategy that can combine differentiating AI capabilities with data sources for unprecedented time to value. As the previously mentioned examples show, such firms exist: They are firms that have consciously taken the risk to invest in next-generation technologies. As a result, they will be rewarded handsomely in the form of reduced cost of operations as well as effective scaling of the business and in turn will be well positioned to compete digitally. Firms that thrive in the digital economy are those that can effectively combine artificial intelligence and edge computing to create a business outcomes—based edge AI strategy.



MESSAGE FROM THE SPONSOR

Tech Mahindra

Tech Mahindra is a multinational provider of Information Technology (IT), networking technology solutions, Integrated Engineering Solutions (IES) and Business Process Outsourcing (BPO) to various industry verticals and horizontals. With operations in 90 countries and over 122,000 employees worldwide, Tech Mahindra is one of the fastest growing strategic IT services firms. Tech Mahindra has amassed deep expertise in helping its clients build and operationalize AI applications. It has also helped clients deploy edge computing infrastructure as a part of their DX initiatives. It has now combined these domains to develop industry-leading edge AI services expertise. Its AI services are developed at an enterprise-grade lab in Fremont, California. Contact <u>techmahindra.com</u> to learn how we can help you with accelerate your edge AI initiatives.

About the analyst:



Ashish Nadkarni, Group Vice President, Infrastructure Systems, Platforms, and Technologies

Ashish Nadkarni is Group Vice President within IDC's worldwide infrastructure practice. Mr. Nadkarni leads a team of analysts who engage in delivering qualitative and quantitative research on computing, storage, and data management infrastructure platforms and technologies via syndicated research programs (subscription services), data products (IDC Trackers), and custom engagements.

O IDC Custom Solutions

IDC Corporate USA

5 Speen Street Framingham, MA 01701, USA T 508.872.8200 F 508.935.4015 Twitter @IDC idc-insights-community.com www.idc.com This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason. Copyright 2019 IDC. Reproduction without written permission is completely forbidden.

