Deloitte.
Insights
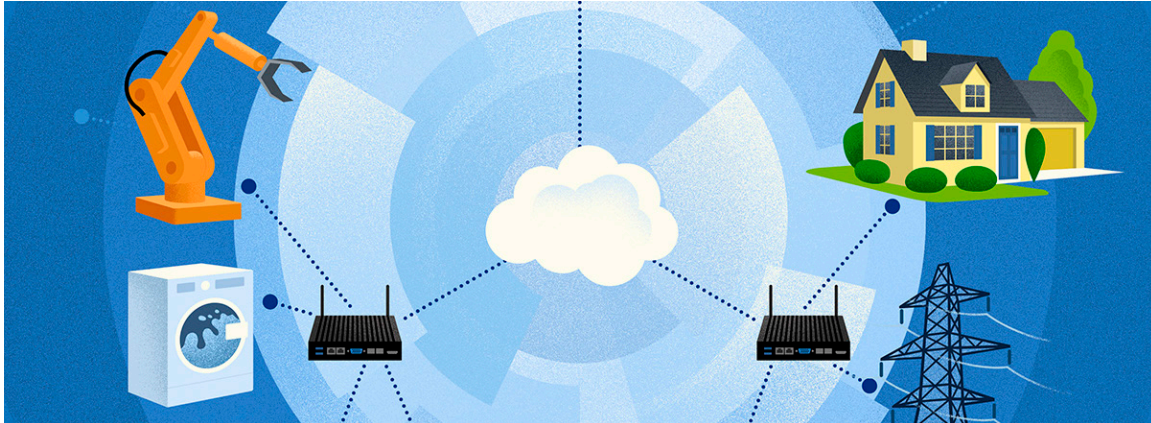
Scaling IoT to meet enterprise needs

Balancing edge and cloud computing

Ken Carroll and Mahesh Chandramouli

Enterprises are increasingly complementing their cloud-based IoT solutions with edge computing to accelerate the pace of data analysis and make better decisions, faster.

J UST A FEW years ago, many expected all the Internet of Things (IoT) to move to the cloud— and much of the consumer-connected IoT indeed lives there—but one of the key basics of designing and building enterprise-scale IoT solutions is to make a *balanced* use of edge and cloud computing.[1] Most IoT solutions now require a mix of cloud and edge computing. Compared to cloud-only solutions, blended solutions that incorporate edge can alleviate latency, increase scalability, and enhance access to information so that better, faster decisions can be made, and enterprises can become more agile as a result.

That being said, complexity introduced by edge computing should justify the objectives at hand, which include scale, speed, and resiliency. A choice that goes too far in one direction typically introduces substantial operational complexities and expenses. Ultimately, the enterprise should take into consideration a full range of factors that reflect its own particular objectives in designing and building an IoT solution in the first place.

In this article, we discuss when and how enterprises can optimally make use of both the edge and the cloud in their IoT solutions. We explain the roles edge and cloud computing play, why the edge may be needed, and how to approach selecting a solution. We also explain some of the complexities with edge computing and provide some use cases.

The cloud explosion and the latency challenge: Enter edge computing

We have experienced a veritable explosion of cloud adoption in the past decade—the IT functionality of many modern companies exists exclusively, or in large part, in the cloud.[2] Among the many benefits of the cloud infrastructure are cost effectiveness, scale, self-service automation, interoperability with traditional back-office systems, and centralized functionality.[3]

At the same time, the amount of sensor-generated data has grown strongly too, and this trend is expected to continue in the years ahead.[4] Because data can become essentially valueless after it is generated, often within milliseconds, the speed at which organizations can convert data into insight

WHAT IS EDGE COMPUTING

Edge computing is a distributed architecture functionality, such as processing and storage, that is located closer to—even *on*—the very source of the data. Examples include cameras with on-device vision processing and wearable medical devices that send data to a mobile phone via Bluetooth. Given these qualities, making a balanced use of both edge and cloud computing is now often considered a key requirement of designing and building enterprise-scale IoT solutions.

The key *representative* functionality often used at the edge can be seen in figure 1.

FIGURE 1

The IoT reference architecture with edge-processing tier
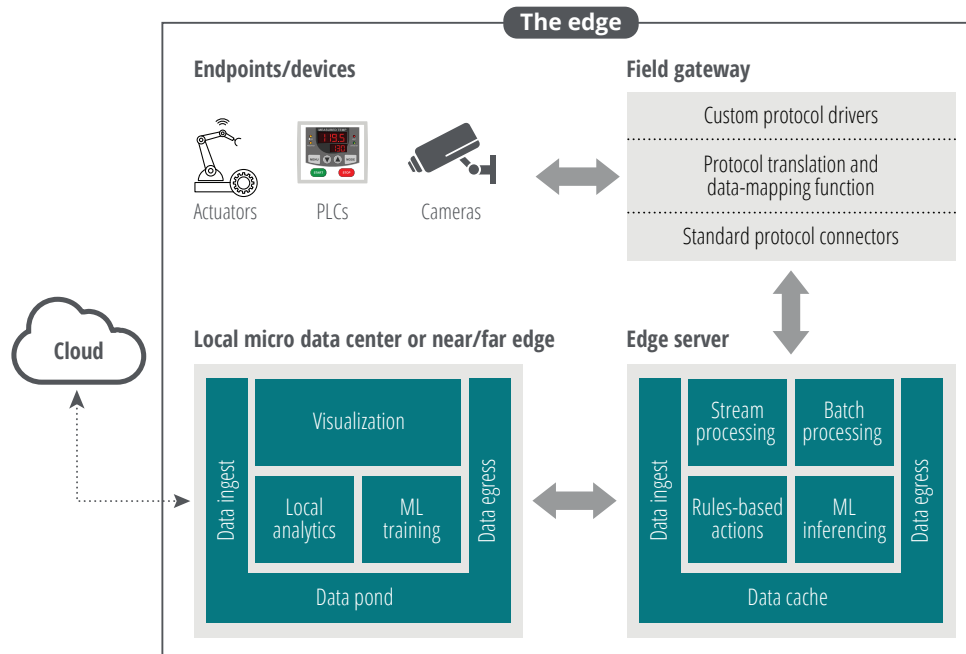
Note: PLC denotes programmable logic controller; ML denotes machine learning.
Source: Deloitte analysis.

and then into action is generally considered mission critical. Therefore, having the smallest possible latency between data generation and the decision or action can be critical to preserve an organization's agility. However, as the speed of data transmission is inviolably bounded by the speed of light, it is only by reducing the distance that data must travel that the latency challenge can be mitigated or avoided altogether. In a cloud-only world the data ends up traveling hundreds or even thousands of miles, so where latency is critical to a solution, edge computing can become key.
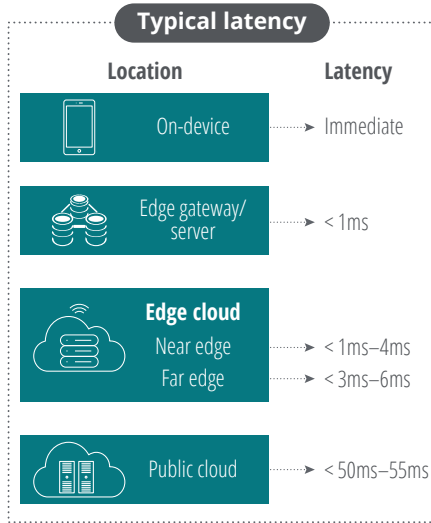
According to one estimate, as much as 55 percent of IoT data could soon be processed near the source, either on the device or through edge computing. Indeed, scale plays a big role in this likely shift—growing data demands will likely put the focus on latency, and decreased latency could dramatically improve the response time, thereby saving both time and money.[5]

Figure 2 shows the relative scale of typical latencies, ranging from on-device to the public cloud.

FIGURE 2

The rise of the edge

Low latency and bandwidth optimization are key drivers in shaping the importance of edge computing

Typical latency

Location	Latency
On-device	Immediate
Edge gateway/server	< 1ms
Edge cloud Near edge	< 1ms–4ms
Far edge	< 3ms–6ms
Public cloud	< 50ms–55ms

Note: Edge gateways/servers are "on-premises," essentially colocated with the devices and can include a micro data center. "Near edge" may be typified by a multiaccess edge computing (or MEC) capability in a nearby cell tower hosted by a telco. "Far edge" is typified by a data center.
Source: Deloitte analysis.

The benefits of edge computing in IoT solutions are plentiful: It's not just about latency

Latency is just one of the many reasons to drive the addition of edge functionality to an IoT solution. A fuller list of the potential benefits of edge computing is given in figure 3.

In bringing edge and cloud computing to life, an understanding of *real-world* cases can go a long way. Constant technology evolution, such as the eventual availability of 5G, often affects the cost/latency/balance equation. As such, it is pertinent to consider *current* conditions in making decisions rather than simply defaulting to previous choices. Being mindful of all drivers while designing an IoT solution is important, as multiple drivers may apply in a given situation.

FIGURE 3

Edge computing provides an array of potential benefits

Benefit of edge computing	Description
Bandwidth availability and use	Bandwidth for long-distance communications over the internet is improving, but the amount of data being generated is growing exponentially too, which could soon outpace the available bandwidth.[6] Bringing processing closer to the source of the data could help equalize the equation—local processing (including compression, filtering, and prioritization) can help make effective and efficient use of the available bandwidth.
Network connectivity	Having functionality at the edge can help ensure that applications are not disrupted when network connectivity issues occur. For example, critical applications used at an offshore drilling platform, where the risk of a network connectivity disruption is high, may benefit from edge processing.
Network security	Using a form of hub-and-spoke cloud/edge architecture can help localize or compartmentalize device secrets (tokens, keys, certificates, etc.) into multiple nodes, rather than storing all of them in one place, thereby helping to improve security.
Autonomy	Many IoT solutions require autonomy where the endpoints operate alone or with a gateway and edge server providing sufficient processing and storage capability to monitor, manage, and respond at a local level. Common examples include autonomous vehicles, heart monitors and similar medical devices, and power plants, where the system must continue to work autonomously even if connectivity is lost.

Benefit of edge computing	Description
Data privacy	Each data point comes with significant privacy, security, and regulatory requirements, with many data points needed to be contained close to where they are generated. This is made possible with edge computing. It is especially relevant in certain verticals where privacy is paramount (health care) and/or regulation governing data movement is stringent (Germany).
Data normalization/ homogenization	As data is created, it must be interpreted. There are a large number of protocols that must typically be supported and a wide array of devices.[7] To effectively manage this scale of variety, the sensor data needs to be normalized, using translation to messaging standards.[8] Such homogenization can be more effectively done at the edge, typically in a gateway, rather than in the cloud, where processing resources are far better applied to functions directed at deriving value from the data. Such data normalization or homogenization at the edge is often coupled with filtering and other data "compression" functions, thereby also helping to make more efficient use of available bandwidth.
Data filtering/ prioritization	Only a part of the mammoth data generated by IoT devices is generally relevant, so filtering out redundant or otherwise nonuseful data can save not only data-transfer cost and bandwidth but also cloud-computing and storage cost. In many cases, filtering and prioritization also mean transforming the raw sensor data. Edge computing can save companies from housing and/or transmitting unnecessary data.
Simpler, cheaper devices	Gateways and edge servers enable simpler, cheaper devices as data storage and processing are transferred to the gateway or edge server, instead of embedding them in the sensor or device.
Latency	The speed with which organizations convert data into insight, and then into action, is often critical—especially in cases involving safety. Beyond the theoretical limitation imposed by the speed of light are the much more practical efficiency limitations that network hops and network equipment can impose when massive amounts of data must traverse long distances. Edge computing can help resolve this challenge by moving processing much closer to the source of the data.

Source: Deloitte analysis.

Balancing cloud and edge computing in IoT solutions

The IoT can have a dramatic effect on an organization's ability to be more agile. Given below are some ways in which the edge and cloud help aggregate and transmit data on behalf of enterprises connected by the IoT.
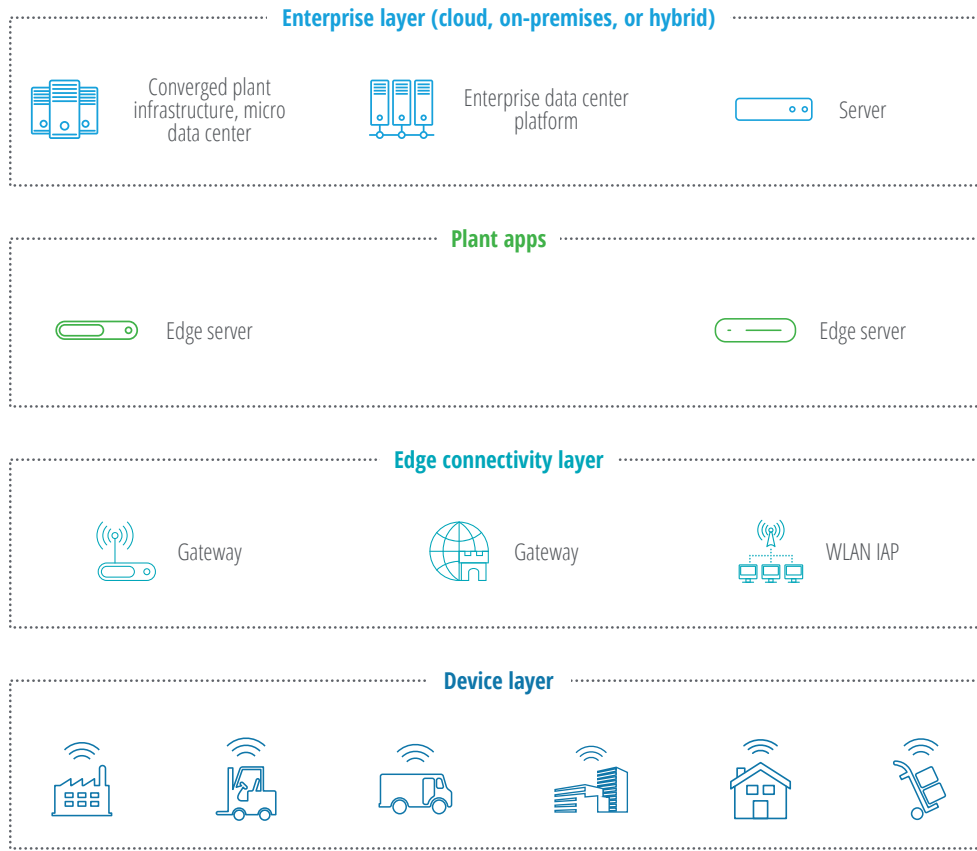
SMART FACTORIES[9]

Enterprises are quickly moving toward an event-driven architecture and real-time automated digital processes.[10] But when you consider that many manufacturers have multiple plants across geographies—each typically with unique character-istics and functional requirements— the challenge to maintain an *exclusively* centralized data-analy-sis capability in the cloud or at a corporate data center becomes apparent.

To be sure, cloud computing offers a number of benefits and it most certainly has a role in smart manufacturing. With data in the cloud, a central-ized operations facility can monitor systems and processes across a large, possibly global, portfolio. It is also possible to undertake comparative analy-sis across the full portfolio that can determine potential for optimizations.

FIGURE 4

A typical cloud and edge computing stack within manufacturing

Source: Deloitte analysis.

Still, we envision that an integrated edge–cloud architecture would provide the kind of speedy and nearly unimpeded connectivity that smart factories require.

Figure 4 illustrates how the edge and cloud typically work with sensors and devices on a manufacturing floor.

- The **device layer** represents individual pieces of equipment that are connected to local operational technology and IoT capabilities for immediate interactions. At this layer, machine learning (ML) scoring or inferencing is done that is based on ML models trained in the cloud. Significant amounts of raw device data are also stored here.

- While the device layer provides visibility and control of individual pieces of equipment, the **plant apps layer** provides visibility and control across all connected pieces of equipment in a plant. The **edge connectivity layer** provides the necessary connectivity between the individual pieces of equipment and the plant apps.

- The **enterprise layer**, which is cloud-hosted, provides primarily visibility, and some control, across multiple plants—a portfolio view. Enterprise analysis and ML algorithms are developed in this layer to predict and provide actionable intelligence—the ML models trained and re-trained here make use of data from across the whole portfolio of plant equipment

and are then "pushed" to the edge and eventually to the IoT software at each piece of equipment, making operations smarter with the information.

SMART BUILDINGS

The rise of smart, connected IoT devices and ubiquitous connectivity has created an opportunity to transform buildings—whether they are offices, retail stores, factories, or hospitals—into cost-efficient, responsive environments for delivering exceptional experiences to their occupants.[11] Smart buildings are digitally connected structures that combine optimized building and operational automation with intelligent space management to enhance the user experience, increase productivity, reduce costs, and mitigate physical and cybersecurity risks. Smart, digital buildings span industries and uses, but all of them can provide the same basic capabilities: They connect humans; they provide better control of facilities and operations; they support ways to collaborate digitally; and they enable owners to conserve resources, including space, energy, water, and employees. Each of these four capabilities can form the basis for creating a smart building strategy that can deliver measurable benefits. Figure 5 illustrates some of the different types of sensors and applications that can be utilized in a smart building.

FIGURE 5

Smart buildings use multiple types of sensors and applications
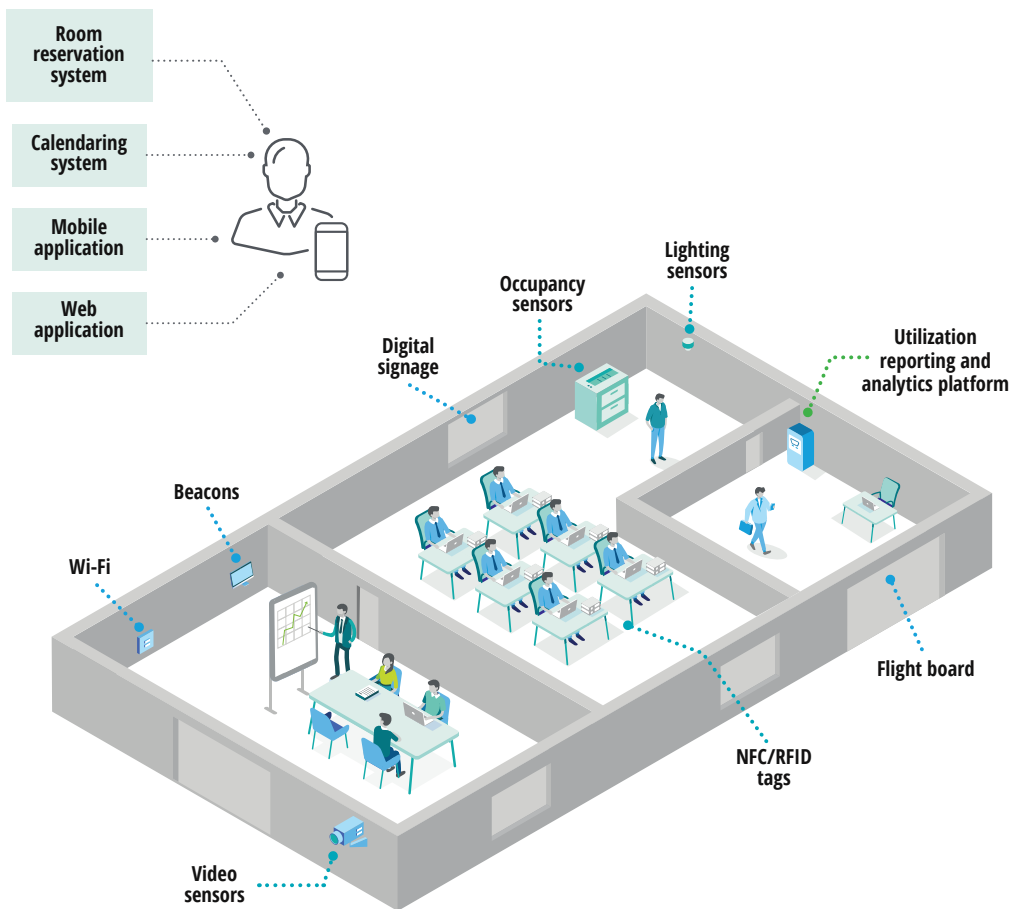
Sources: Deloitte analysis.

For example, 75–80 percent of a building's life cycle costs are related to building operations. All existing commercial and large residential buildings have some form of building automation (or management) system which controls such things as HVAC. In order to introduce smart building features, such as smart lighting with embedded occupancy sensors, and have these interact with the main system, at minimum a gateway is required with some additional capability that would usually come with an edge server.

The edge capability would provide data (occupancy data, for instance) from strategically placed sensors to a cloud-hosted service that performs some specialized analytics. The outcomes from these analytics can be sent back, via the gateway or edge server, to alter the schedules of equipment connected to the main system in order to optimize operations. This configuration is also necessary to obtain a portfoliowide view of building operations and conditions. Edge computing and cloud enable smarter management of resources.

Complexities and challenges in scaling IoT solutions using both edge and cloud computing

While edge computing offers solid benefits, it can also introduce operational and design complexities. Edge processing is highly distributed and often includes far-flung and/or difficult-to-access locations, including sensors/actuators and gateways in offices, plants, at campuses, on pipelines, and in various remote field sites. Any given organization can have thousands of devices and hundreds of associated gateways. All these edge nodes have firmware, operating systems, some form of virtualization and containers, and software installed, some of which are provided by manufacturers and some by solution providers. These need to be properly managed and maintained by the owner/

manager of these edge nodes, mandating an enormous degree of automation (such as for backups, patching, updates, and monitoring).

The number of potential problems is enormous, and troubleshooting can be very challenging and complex in a highly distributed model. In many cases, field service technicians are required to be regularly onsite to address issues that occur as a result of upgrades or even general maintenance. These drivers also tend toward the need for a widespread "software defined everything" approach, as software upgrades are more easily and conveniently achieved than hardware upgrades.

CONTINUAL UPDATING

Despite its challenges, cloud computing obviates concerns with many key IT issues, providing a degree of self-service and automation. Edge processing requires common data center operations (provisioning, updating, change management, and monitoring) in addition to the other higher-level functions (device management, updating machine-learning models, etc.) to be undertaken and replicated to all of the edge nodes and clusters. This is a heavy undertaking, requiring the enterprise to shift the focus from business needs to some extent.

POLICIES AND PRACTICES

Policies and practices used in traditional data centers are often not readily applicable to edge deployments, which are distributed across multiple locations and considerably more dynamic than traditional data centers. Undertaking the operational management of such a system is a complex challenge.

COST

While the cloud offers on-demand scalability and is readily configurable, automated, and resilient, providing for these capabilities at the edge can be costly and complex. Accommodating the expansion of an existing edge deployment to allow for an increased number of devices and edge nodes can

involve significant investment in additional hardware and software, and much complex work.

CYBERSECURITY

Extending the cloud and the data center to the edge with multiple nodes and devices exponentially increases the surface area for cyberattacks. Insecure endpoints, such as devices and edge nodes, can be used as entry points to valuable assets within an enterprise network and for other nefarious purposes, such as distributed denial-of-service attacks. Maintaining the physical and cybersecurity position of all assets in the edge is a complex and critical challenge.

How to determine the need for edge processing in an IoT solution

Including unnecessary complexity in a solution can be costly, risky, and wasteful, so whether to add edge processing to an IoT solution is a decision best taken with care and based on a risk/reward assessment. To that end, figure 6 offers some guidelines that may help.

In many IoT use cases, the edge is simply a necessary or mandatory part of the solution, given the already existing operational technology. Adding a

FIGURE 6

Required capabilities and rationales for considering edge processing within an IoT solution

Capability required	Rationale for choosing edge processing
Autonomy and resiliency	The solution demands autonomy and connection interruptions cannot be tolerated.
Urgency and safety	The speeds at which decisions and actions are needed is important and safety is a key concern or outcome.
Privacy, security, compliance, regulatory adherence	Requirements mandate that some of the data be contained close to where it is generated.
Operational technology and other protocols	These protocols exist due to the simplicity or the nature of the devices and normally mandate that at least an edge gateway be used.
Data volumes and bandwidth	Making effective and economical use of bandwidth and cloud or data center resources, especially when volumes are large or much of the generated data is unnecessary, usually mandates some local processing to filter, compress, or otherwise transform the data prior to transmission to the cloud.

Source: Deloitte analysis.

IOT SCALING CHALLENGES WITH USE OF *ONLY* EDGE OR *ONLY* CLOUD

Designing a large IoT solution that does essentially nothing at the edge and sends all data for action to the cloud often presents scaling challenges in terms of bandwidth usage, possibly necessitating upgrades to networking infrastructure. Additionally, as the solution scales, exclusive use of the cloud could require reconfiguring ingestion engines and load balancing through manual intervention.

The complexities of an exclusively edge-based distributed architecture (especially one that may involve distributed processing) are no less, and they increase with scale. Systems and application management are highly complex, and tools needed for these have yet to mature. In many cases, edge deployments do not adequately consider expandability, complicating the support of more devices and more data.

cloud-hosted component of an IoT solution requires some degree of edge-computing presence, even if primarily a gateway. Similarly, the desire to add smart capabilities to existing building management system infrastructure and provide for a cloud-based real estate portfolio view necessitates use of some edge-processing capability.

Where to start

The first step is to assess whether edge computing is needed at all. It may well be that the best solution is a purely cloud solution. The next step is to determine the capabilities needed at the edge, and followed by determining the most appropriate deployment model, given that edge processing can be on devices, gateways, edge servers, possibly in multiple tiers, or micro data centers. There can be large variations in compute capabilities, responsiveness, and placement.

In some cases, preconfigured solutions, packaged into a single, integrated product, can offer simplicity, but at the possible expense of flexibility. The capability provided by the flexibility of self-constructing from best-of-breed components is attractive, though it comes at the expense of software/product development and stabilization, which lengthens the time taken to deliver a solution and brings a number of inherent risks.

Paying attention to the changes in the edge compute landscape is important, as is conducting a proof of concept with the relevant deployment design in order to finalize the best fit choice for the use case involved.

There is another variable worth noting—the edge-computing vendor landscape. It is undergoing rapid change. Most IoT infrastructure or platform vendors recognize that edge computing is an important part of many IoT solutions and delivery hardware, such as gateways or servers, with some

data processing, analytics, and local storage capabilities. These hardware vendors tend to rely on others for device management, protocol handling, and conversion, in addition to other capabilities. Significant consolidation is likely to occur in this space as vendors seek to offer end-to-end solutions.[12]

Takeaways

IoT devices and the data they can provide are changing the world and how we interact. Much of the connected-consumer IoT world resides primarily on the cloud largely because of its copious benefits. However, in most cases an IoT solution will involve some mix of the edge and the cloud. Bringing it to the edge can alleviate latency, boost scalability, and increase access to information so that better, faster decisions can be made, and organizations can become more agile as a result.

While deciding on the right balance of edge and cloud functionality in an IoT solution, it is helpful to keep in mind that edge computing has various configurations, and all can bring their own benefits, but they also can present unique challenges. Substantial operational complexities and expenses can emerge in no time, so enterprises should take into consideration a full range of factors while designing and building any given IoT solution.

Even then, an IoT solution should be only as simple as it needs to be, and no simpler. Conversely, it should be only as complex as it needs to be, and no more complex. These seemingly straightforward, yet essential, points can make a difference in the success of a solution.

Clearly, there is no single correct answer in the cloud vs. edge assessment in an IoT context. Every situation is unique. What is clear, however, is that a balance between cloud and edge computing will likely make up tomorrow's IoT architecture.

Endnotes

1. For more information, read Deloitte Insights' suite of articles on Internet of Things.

2. Barb Renner, Curt Fedder, and Jagadish Upadhyaya, *The adoption of disruptive technologies in the consumer products industry: Spotlight on the cloud*, Deloitte Insights, April 4, 2019.

3. Ibid.

4. Muhammad Syafrudin, "Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing," *Sensors* 18, no. 9 (2018): 2964, DOI: 10.3390/s18092946.

5. Michael Kanellos. "Edge or cloud? The five factors that determine where to put workloads," CloudTech, June 20, 2018.

6. Jeff Hecht, "The bandwidth bottleneck that is throttling the internet," *Scientific American*, August 10, 2016.

7. Examples of such protocols include Modbus, OPC-UA, Zigbee, BACnet, and LWM2M.

8. Examples of such standards include AMQP and MQTT.

9. For more on smart factories, read: Rick Burke et al., *The smart factory: Responsive, adaptive, connected manufacturing*, Deloitte University Press, 2017.

10. Gartner, "Gartner says CIOs must define an event-centric digital business strategy," press release, July 10, 2017.

11. Surabhi Kejriwal and Saurabh Mahajan, *Smart buildings: How IoT technology aims to add value for real estate companies*, Deloitte University Press, 2016; Paul Wellener et al., *Smart buildings: Four considerations for creating people-centered smart, digital workplaces*, Deloitte Insights, December 13, 2018.

12. Param Singh and Brian Buntz, "IoT technology: A platform for innovation, but not a market," IoT World Today, September 28, 2018.

Acknowledgments

About the authors

Ken Carroll | kencarroll@deloitte.com

Ken Carroll is a leader in Deloitte's Smart Buildings, IoT, and Cloud practices. With over 36 years of experience in software product architecture, creation, and delivery, he has worked extensively in IoT, engineering, energy management, smart buildings, financial services, and educational publishing domains. In addition, he has led IoT practices and driven key IT transformations at many leading organizations. He is based in Boston.

Mahesh Chandramouli | mchandramouli@deloitte.com

Mahesh Chandramouli is a senior manager in Deloitte Digital's IoT practice, Deloitte Consulting LLP, and has extensive experience in delivering IoT strategy, solutions, and delivery engagements in retail, manufacturing, life sciences, health care, telecom, media, and technology industries. He brings more than 15 years of experience in helping companies create unique digital experiences for their customers. Chandramouli is based in Dallas.

Contact us

Our insights can help you take advantage of change. If you're looking for fresh ideas to address your challenges, we should talk.

Industry/practice leadership

Andy Daecher
Principal and Internet of Things practice leader | Deloitte Consulting LLP
+1 415 783 6525 | adaecher@deloitte.com

Andy Daecher leads Deloitte Digital's Internet of Things practice, advising executives on the practical applications of emerging technologies, effective management of IT organizations, and execution of complex, transformational, technology-enabled projects. Daecher is based in San Francisco.

Ken Carroll
Senior manager | Deloitte Consulting LLP
+1 781 223 8863 | kencarroll@deloitte.com

Ken Carroll is a leader in Deloitte's Smart Buildings, IoT, and Cloud practices. With over 36 years of experience in software product architecture, creation, and delivery, he has worked extensively in IoT, engineering, energy management, smart buildings, financial services, and educational publishing domains. Carroll is based in Boston.

Industry center leadership

Jonathan Holdowsky
Senior manager | Deloitte Services LP
+1 617 437 3198 | jholdowsky@deloitte.com

Jonathan Holdowsky is a senior manager with Deloitte Services LP and part of Deloitte's Center for Integrated Research, leading thought leadership initiatives that explore the promise of emerging and disruptive technologies. Holdowsky is based in Boston.

About the Deloitte Center for Integrated Research

Deloitte's Center for Integrated Research focuses on developing fresh perspectives on critical business issues that cut across industries and functions, from the rapid change of emerging technologies to the consistent factor of human behavior. We look at transformative topics in new ways, delivering new thinking in a variety of formats, such as research articles, short videos, in-person workshops, and online courses. We are at the forefront of many Deloitte Big Idea campaigns, which focus on topics we can own in the marketplace, with a direct line of sight to commercialization and multiple intersections with sectors and offerings.

Connect

To learn more about the vision of the Center for Integrated Research, its solutions, thought leadership, and events, please visit
https://www2.deloitte.com/us/en/pages/about-deloitte/solutions/center-integrated-research.html

Subscribe

Sign up for updates at www.deloitte.com/insights

Engage

Follow us on Twitter at: @DeloitteInsight

Engage with us on LinkedIn via Deloitte Insights

POWERING TOMORROW'S DIGITAL NETWORK WITH THE IOT
Deloitte understands the power of the Internet of Things—not just for DSNs, but for an even broader set of applications from smart cities to digital utilities to tomorrow's workplace to next-generation transportation. With a dedicated IoT practice focused on the connected enterprise, we stand ready to help you deploy new capabilities around the Internet of Things—whether you want to build a digital supply network or another innovative digital network enabled by IoT. The goal? To help you create real business value, not just connect things for the sake of connecting them. **Let's chat**.

Deloitte.
Insights

Sign up for Deloitte Insights updates at www.deloitte.com/insights.

Follow @DeloitteInsight

About Deloitte Insights

Deloitte Insights publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte Insights is an imprint of Deloitte Development LLC.

About this publication

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.