Distributed Data Analytics:

STRATEGIC ADVANTAGE AT THE EDGE

database
TRENDS AND APPLICATIONS

actian™
activate your data

Distributed Data Analytics: Strategic Advantage at the Edge

Gaining the advantage in the years to come means going to the edge. Businesses are discovering their future lies in the ability to leverage strategic edge analytics, now possible through the surge in compute intelligence closer to where data is created, leveraging volumes of data being generated through interactions with cameras, sensors, meters, smartphones, wearables, and more. In conjunction, processor, storage and networking capabilities to support local embedded analytics on these devices and across them through peer-to-peer interactions (on local or nearby mezzanine or gateway platforms) is also increasing.

This explosion of intelligence at the networks' edge is often termed the Internet of Things (IoT). The IoT is shaping in many forms—in the consumer space through sensors and systems embedded in products, and in the corporate space as the Industrial IoT, in which data is streamed from production and monitoring systems in factories, offices, healthcare facilities and vehicles. IoT is a clarion call for updating the traditionally centralized, cloud-based model for enterprises to a more distributed and fluid flow of intelligence, and therefore, analytics and supporting data management, out to the edge. The value in the IoT revolution is much more than simply connecting devices and downloading data—it means unfolding systems into global networks connecting companies more intimately with customers, partners, employees, and other constituencies. Strategic edge analytics makes this a reality, without the complexity and latency seen with the centralized hub-and-spoke approaches seen thus far with IoT.

If industry growth projections are any indication, edge computing is on a rapid upward swing. Analysts at Markets and Markets predict 35% annual growth through 2022, while Grand View Research forecasts 41% annual growth through 2025.[1 2] Markets and Markets points to increasing volumes of data coming from IoT sensors, cameras, and beacons that stream into smart applications and are efficiently collected, stored, and processed at the network edge versus a centralized cloud or data center. The fastest-growing segment for edge computing is likely to be retail, says Markets and Markets, while Grand View sees the most growth potential in healthcare and life sciences, due to the storage capabilities and real-time computing offered by edge computing solutions. Still others say manufacturing and additional long-term forecasts often point to smart cities, transportation, and other areas that require significant investment and planning. Taken as a whole, IoT will have a substantial and lasting impact on virtually every industry.

WHAT IT MEANS TO ENTERPRISES TO MOVE TO THE EDGE

There are compelling benefits to be achieved from moving data and processes closer to the edge, where these devices are specified, deployed, and maintained. Today's data analytics—and even artificial intelligence algorithms—can run anywhere along the spectrum, from cloud, to sensors at the edge, and within today's network-oriented world.

There are numerous use cases for edge analytics, such as manufacturing quality control, in which production-floor machine data is analyzed on the spot in real time to identify anomalies and automatically take remedial action. A medical device may require real-time responsiveness to a patient's condition and thus the latency transmitting data back and forth may be a life-or-death issue. Likewise, an autonomous or semi-autonomous vehicle needs to make instantaneous decisions as to whether to yield, stop, or swerve to avoid dangerous situations.

The following trends are shaping the emerging world of strategic edge analytics:

- **The edge is shifting outward:** What was a server is now a cloud service. What were desktops and laptops are now smartphones and specialized PDAs. Applications today are distributed across network devices and out to everything from smartphones to sensors. While cloud platforms are consolidating, cloud services are mushrooming and the need to orchestrate them essential.
- **Business is digital:** The integration of digital technology into all facets of business is profoundly changing how businesses operate and serve customers, from corporate and branch offices, to production facilities, mobile devices, vehicles, and remote job sites. This disruption is changing work cultures, user expectations, and best practices around productivity, efficiency, insights and innovation.
- **"Fog" and "mist" computing mean computing everywhere:** The IoT is eating traditional, centralized cloud architectures. The massive volumes of data generated by connected devices, alongside the growing need for real-time data, is driving the adoption of new cloud architectures that more evenly distribute computing power across the network.[3]
- **Information technology and operational technology converge into one powerful, fully integrated environment:** All systems and applications are increasingly connected into one powerful and responsive network. The standardization of IoT and incorporation of IT architectures, security, and virtual teams into what was formerly siloed operational technology.

WHY CONSIDER MOVING TO THE EDGE?

To date, the preferred method for deploying and running data analytics has been within centralized locations, where sophisticated analytic applications coupled with traditional enterprise data warehouses and newly-erected Hadoop data lakes tend to reside. For many organizations, this has delivered great value. Centralized data analytics has been instrumental in environments that are complex, with multiple data sources. There are many situations, however, that require real-time speed and responsiveness.

"There are other situations where IoT data needs to be analyzed in near real time in order to ensure rapid execution and effect change," notes Daniel Kirsch, analyst with Hurwitz Group. "For example, real-time analysis of sensor data on a manufacturing system can detect too much moisture or too high a temperature. This situation will require immediate action to prevent failure."[4]

With IoT, the range of requirements expands beyond many organizations' capabilities. Yet, for others, IoT opens up possibilities to turn far-fetched ideas into practical realities. They need to widen their options with the employment of analytics at the edge. With data analytics running closer to the edge, resulting insights are made immediately available at the point of action rather than in retrospect based on analysis back in the data center.

Centralized data analytics environments face the following challenges with IoT:

- **Security:** To date, sensors and devices have been built with an emphasis on moving data from one location to another—not locking it down and securing it at the point of its inception. Thus, sensors and gateways may serve as entry points to central systems for malicious code or hackers (in other words each device becomes an additional point of attack, massively expanding an enterprise's attack surface).
- **Latency:** Moving data back and forth over the internet or even internal networks may take multiple seconds, or even minutes. For devices requiring real-time responsiveness, for instance, medical equipment, such latency is unacceptable. And, when decisions must be made on a heterogeneous set of devices, and their time series data must undergo temporal or spatial filtering, delays can be debilitating.
- **Lack of standardization:** The billions of devices and endpoints now in today's internet were built according to individual manufacturers' specs and requirements, and may not mesh well within centralized analytical environments.
- **Getting the most out of analytics investments in the data center:** Centralized data analytics applications are significant investments for organizations, and the sizes of these investments are only going to increase with the arrival of newer but more promising techniques such as machine learning and deep learning.

BENEFITS OF ANALYTICS AT THE EDGE

The ability to provide better customer service, connecting more closely with partners, achieving greater scalability, and enabling more reliable delivery of information represents just the beginning of the advantages of strategic edge analytics. Moving and/or extending data analytics out to the edge provides a range of potential benefits:

- **Flexibility:** Data analytics at the edge provides flexibility and agility, as well as the ability to bring decision making based on analysis across multiple disparate systems into a single IoT network closer to the point of action.
- **Adaptability:** While algorithms for machine learning and deep learning must be developed and tuned in the data center, they are self-managing once they are contained within systems or devices at the edge. Once an algorithm is deployed at the edge and set for unsupervised runtime, in many cases all that is needed is lightweight code to run relatively simple analytical tasks, with occasional adjustments and redeployments of the algorithms based on prior outcomes.
- **Security:** Security protocols can be embedded into the edge device or application, particularly with respect to the data management associated with the local analytics. With edge computing, sensitive data is kept within devices. "For example, new retail advertising systems and digital signage are designed to deliver targeted ads and information based on key parameters set on field devices, such as demographic information," says Raj Talluri of *Network World*. "Edge computing in these solutions helps protect user privacy by anonymizing, analyzing, and keeping the data at the source rather than sending identifiable information to the cloud."[5]
- **Skills requirement:** Data analytics at the edge can be supported through embedded algorithms and associated embedded data management that don't require on-site expertise to upgrade or manage.
- **Greater accessibility:** Data and insights are delivered on the spot, or fed directly into the system at the endpoint, providing end users—human and machine—the opportunity to make more accurate decisions.
- **Real-time:** Latency is greatly reduced or eliminated. This is especially important in medical devices or autonomous vehicles, particularly if there are periods of expected—or worse unexpected—lapses in connectivity.
- **Improved end-user experience:** Combining localized analytics of IoT data with location-based data and other formerly siloed data sets can greatly improve customer experience.

EDGE OR CENTER? DECIDING WHERE STRATEGIC EDGE DATA ANALYTICS FITS BEST

Of course, there are many different scenarios unfolding for IoT, and many different ways the business case for strategic edge analytics needs to be made. The following are some of the deciding factors for the extent to which analytics at the edge should be deployed:

- **Latency requirements, need for speed:** Sending data back and forth between devices and centralized systems can create problems, especially in situations that call for rapid response is required. The need for edge analytics may be seen in critical medical devices, factory automation, aerospace and defense applications, and onboard vehicle sensors and engine monitoring equipment, to name just a few.
- **Performance requirements:** Related to latency are the performance issues that may be introduced to systems depending on distributed networks.
- **Improved systems of engagement:** Moving intelligence closer to the end-user—whether they be knowledge workers or customers—may be facilitated through edge analytics.
- **Data reliability and quality:** It may not be cost-effective to flow data into centralized systems if much of that data has relatively little value because it either doesn't change or has a very low signal-to-noise ratio.
- **Cost of data movement and storage:** Sending all data instead of performing preprocessing consumes significant bandwidth, putting pressure on expensive upgrades to networking systems. Once data is moved to the cloud or private data centers there is always a cost of storage and, in some regulated industries, once the data is placed there it must be kept for many years inclusive of security safeguards that have evolved from periodic and cursory to continuous and complex.
- **Available processing power:** Grid computing based on single-CPU rack-mounted systems was instrumental in shifting data center computing from complex SMP systems with large monolithic applications to an open standard microservices architecture for applications on Linux and Windows platforms. Hardware at the edge is increasingly capable of supporting robust applications running on full-fledged operating environments with embedded data management and analytics at roughly the same level of capability equivalent to an advanced data center 10 to 15 years ago—but now out at the edge.
- **Data types and usage:** Data at the edge will be no different than any other data in that it will be operational and transactional. In the case of operational data much of it will be time series data that must be synchronized and reconciled against more than one time series vector, time, GPS location, or other customized parameters either directly in the data or bolted on as metadata. When taken from disparate devices or gateways, there will be a need to often reorder data, process, and enhance it and its metadata. Similarly, with transactional data, local processing may include temporary storage in case of disconnection, encryption locally, elements of fraud detection or cybersecurity algorithms that should be run based on local data.

TECHNOLOGY CONSIDERATIONS IN MOVING TO THE EDGE

Of course, an organization's infrastructure has to be ready to accommodate and support the move to strategic edge computing. The old ways of doing things—managing data, developing applications and deploying applications—need to be overhauled:

- **Edge computing sees information technology (IT) and operational technology (OT) as one.** IT and OT, considered two separate disciplines, are equal parts of the edge computing and IoT revolution. While IT focuses on data processing and storage, which often require large data environments, OT is seen on production floors, with many smaller-footprint or embedded databases. Edge analytics is rooted in both IT and OT, seeing both as a single converged system.
- **Reorient applications to be distributed and intelligent.** Edge applications are configured to leverage the IoT, and have different requirements. Enterprise architects, data scientists and their line-of-business counterparts can no longer assume that making everything at the edge a thin client or, at the opposite extreme, siloed applications, will satisfy their business outcome objectives. They can no longer run only localized applications with limited intelligence. Instead, they will take advantage of the richer hardware and software resources to build more balanced approaches, determining the best locations for improved situational awareness and decision support at the point of action by humans and machines at the edge. Manual processing will be automated, automated processing will be revised, all expressed in distributed applications with embedded analytics that better support streamlined processing, exception handling, and closed feedback loops that review and adjust the decisions and actions of humans and machines alike. Distributed applications will need to communicate with each other, reacting to how each piece of the application and a series of instances of that component and application across multiple devices are operating. Context, analysis of operations, and responses between devices will generate a need for yet more embedded analytics.
- **The new role of databases** as localized persistent data management systems capable of shared and distributed data analytics support infrastructure. As the intelligence and capabilities of strategic edge analytic applications expands, the need for embedded local and persistent data will also grow on devices, be they sensors, smartphones, or gateways at the cloud's edge. Furthermore, process across applications, governance of them and the underlying software and hardware environments, storage for temporary data that must be processed and then thrown away or stored for some limited time will be required as well. This will promote data sharing across environments in a way that ensures the integrity of the data, management of various clients requesting withdrawal, deposit, or deletion of the data. In other words, this enables a full-fledged data management system but tailored to the edge , paving the way for iterative improvements to embedded analytics and ultimately business insights.
- **Moving data to where it is needed.** Of course, organizations have the option to retain the data generated in more centralized locations as well. Data lakes and similar environments enable the storage of data streaming from edge environments to be made available for future applications or innovations.
- **Standardization and reuse.** Edge applications and analytics are configured and reconfigured as business needs require. The need for local data is increasing, even where applications are still monolithic and extend to mobile devices only through a thin-client web browser interface. If not by architectural requirements then by hard experience, data must not be lost on edge devices during points of network disconnection. Without an off-the-shelf data management system tailored to the edge, developers must code create, read, update, and delete (CRUD) logic to store or retrieve data with care, and coding may need to vary across APIs and file systems to ensure data consistency and avoid corruption to ensure flat file data consistency and avoid corruption.

If concurrent reads or writes must be supported across multiple threads, file systems do not ensure consistency, so a database with full ACID support is a must-have, particularly if the file system acts as a temporary cache yet has a spotty network connection. ACID capability is mandatory if the application and associated data are exposed to machine crashes—otherwise partial writes can leave data in an inconsistent state.

- **Modularization and multi-model support.** The overall computer, storage, and network intelligence will exponentially increase at the edge. However, the actual device, gateway, and complex equipment resources will vary from implementation to implementation as will the use cases associated with each entity and their interactions with others in the LAN, WAN, and global, cloud-connected environments in which they will operate. Product line and service practice managers, in conjunction with their enterprise architects, developers on both the IT and OT sides of the house will want to identify data management systems based on a common architecture that can support a range of programming languages, operating environments, and underlying hardware platforms capable of handling transactional and operational data. Developers can then build reusable API knowledge and even reuse code built for one device or use case on successive projects. Data scientists and business analysts can likewise remotely extract data to meet their modelling and reporting needs.

CONCLUSION AND CALL TO ACTION

Strategic edge computing promises to strengthen the connections between businesses, customers, partners, employees, and other constituencies.

The following steps will help make the most of this emerging opportunity:

- **Put the business first.** Map out the business opportunity and requirements. It's important that the business identify critical requirements, what data would be needed to satisfy these requirements, and then the appropriate technology be implemented to meet those requirements. Not every edge device needs to be built up with analytical capabilities, as this could be a cost-prohibitive exercise when tens of thousands of devices are involved, with very low return for the business. But don't be scared away by thinking all analytic capabilities fit into the nascent category of artificial or machine intelligence, oftentimes, simple mature analytics at the edge is beneficial and easily achievable within relatively constrained resources.
- **Work with customers and partners.** IoT inherently depends on collaboration, co-ownership, and cooperation across your organizational siloes as well as with networks of customers or partners, who likely own and maintain the devices and systems that supply the data. The deployment of analytics within these systems will benefit them just as much as it will benefit the host company. Build a strategic roadmap but start small, find reasonable projects with low risk to get going.
- **Create and analytics model.** Developing an analytics model will provide a roadmap and consistency with the architecture. This involves such areas as collecting data, preparing data, selecting algorithms, training the algorithms on a continuous basis, and finally, deploying the models, states Ramesh Dontha of Digital Transformation Pro. "The processing and storage capacity at the edge also plays a key role. Some of the merging deployment models include decentralized and peer-to-peer deployment models with pros and cons for each."[6]
- **Select the data.** In a typical IoT scenario with multiple streams of data from a variety of sources. The type of data employed is critical. In addition, organizations need to ensure the quality and viability of data employed. A digital, AI-driven enterprise is built on trust of the data being deployed and analyzed. With data coming from an ever-changing array of multiple sources, there is risk of data that is overlapping, duplicative, missing elements, or inconsistent.

The growth of strategic edge analytics promises to rearrange corporate priorities. There are a wide array of benefits of moving data and processes closer to the edge, including enhanced flexibility, greater adaptability, expanded security, skills mitigation, and real-time capabilities.

1 Edge Computing Market worth 6.72 Billion USD by 2022, Markets and Markets, October 2017.
2 Edge Computing Market Worth $3.24 Billion By 2025, Grand View Research, March 2018.
3 Fog Computing Conceptual Model, Michaela Iorga, Nedim Goren, Larry Feldman, Robert Barton, Michael Martin, Charif Mahmoudi, National Institute for Standards and Technology, March 2018.
4 The Value of Bringing Analytics to the Edge, Daniel Kirsch, Hurwitz & Associates, 2015.
5 Why edge computing is critical for the IoT, Raj Talluri, Network World, October 2017.
6 Edge Analytics – What, Why, When, Who, Where, How?, Ramesh Dontha, KDNuggets, Oct. 2017.