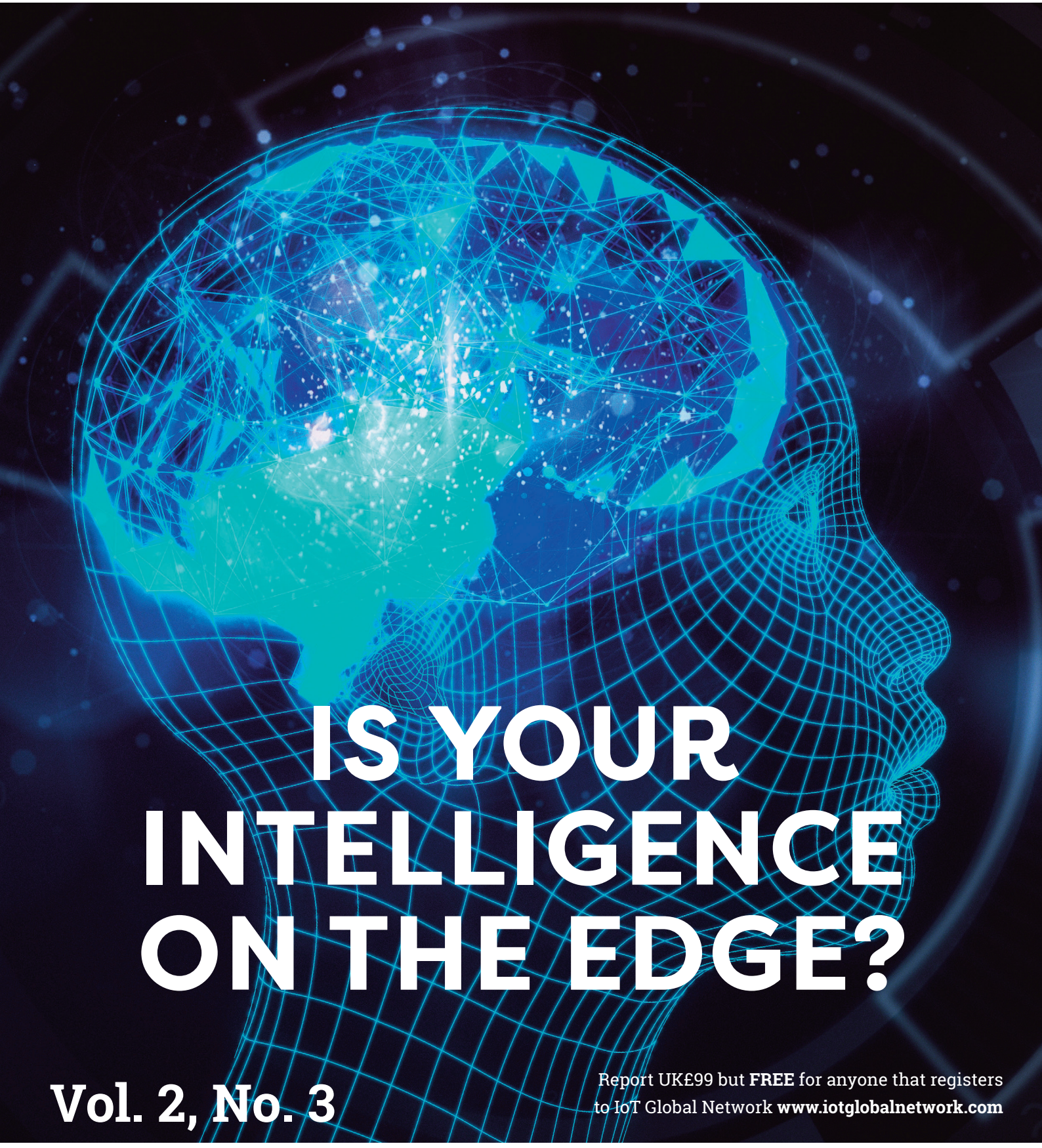
ISSN 2634-9116

VanillaPlus

tech trends

Your inside track on the technology topics that matter



IS YOUR INTELLIGENCE ON THE EDGE?

Vol. 2, No. 3

Report UK£99 but **FREE** for anyone that registers to IoT Global Network **www.iotglobalnetwork.com**

Tech Trends – your new inside track to the topics that **really** matter.

SPONSORED BY
VOLTDB

NEW REPORT

GETTING TO STANDALONE 5G

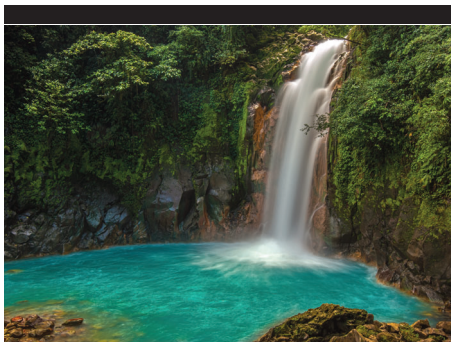
Download the new feature report where T-Mobile, Ericsson, Rakuten Mobile, Vodafone U.K. and VoltDB discuss real-time applications and new service revenue opportunities in the 5G era.

> DOWNLOAD THE REPORT

VOLTDB

For more information, please visit **VOLTDB.COM**

Editor's overview



04

Data lakes and streams give life to intelligence at the edge

Data, communications and computing infrastructure are coming together but intelligence at the edge relies on new tools and processes, writes **George Malim**

Analyst report

06

Edge computing will be critical for digital transformation

Edge computing is a key enabler for solutions that utilise emerging technologies such as AI, 5G, cloud and IoT, and it has become critical to enterprises, writes **Jim Morrish**, a founding partner at Transforma Insights

Sponsored interview

12

5G and MEC encourage edge intelligence to become pervasive

VoltDB's **Dheeraj Remella** tells **George Malim** how the challenges of edge intelligence are being addressed and why there will be no going back to centralised, remote architectures

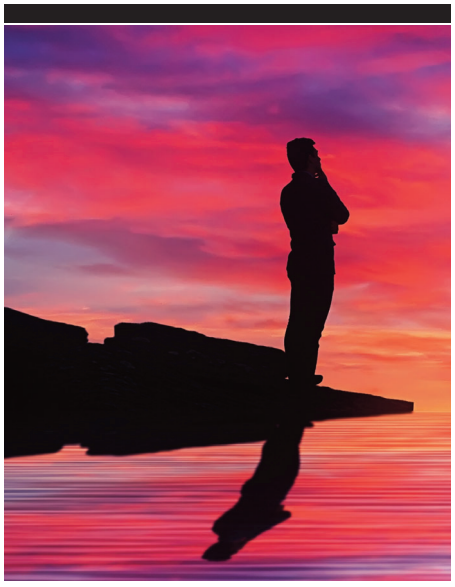
Low latency decisions

16

Why low latency is vital for edge success

5G, edge computing, digital transformation, digital twins, machine learning and AI look like buzzword bingo but there are rational connections between each, writes **Dheeraj Remella**

Securing the edge

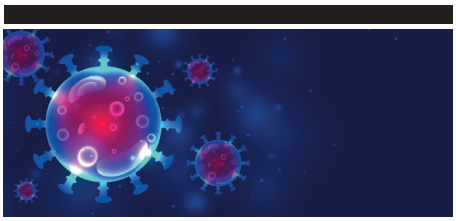


18

Can it be safer on the edge?

Antony Savvas explains why the edge is turning out to be one of the safer locations for hosting processing power and analysing data

Edge analysis



20

Pandemic pressures organisations to push intelligence to the edge

George Malim examines the latest analyst predictions for the increased uptake of MEC and edge intelligence

Edge ROI

22

MEC proves the need for speed

While there are sensible, industrial cost justifications for MEC, **Nick Booth** warns against being distracted by fanciful projections

VOLTDB

About VoltDB

VoltDB enables global organisations to leverage emerging 5G latency standards to power new revenue opportunities, transform their business and operational support system, or develop strategic integrations for their enterprise customers. The platform instantly derives value from anomalous events captured across multiple streams of fast data, delivering precise decisions in less than 10 milliseconds that directly influence in-the-moment monetization, prevent digital fraud, and support digital transformation initiatives. VoltDB is purpose-built to address application-specific scale and latency challenges and augment previous big data and messaging investments to enable businesses to evolve from big data analytics to fast data decisions.

www.voltdb.com

VanillaPlus Tech Trends covers technological & business developments for businesses enabled by the Internet of Things (IoT). © Copyright WKM Ltd. All rights reserved. No part of this publication may be copied, stored, published or in any way reproduced without the prior written consent of the Publisher.

ISSN 2634-9116

Managing editor:
George Malim
Tel: +44 (0) 1225 319566
g.malim@wkm-global.com

Editorial director & publisher:
Jeremy Cowan
Tel: +44 (0) 1420 588638
j.cowan@wkm-global.com

Digital services director:
Nathalie Millar
Tel: +44 (0) 1732 808690
n.millar@wkm-global.com

Business development:
Cherisse Jameson
Tel: +44 7950 279368
c.jameson@wkm-global.com

Designer:
Jason Appleby
Ark Design Consultancy Ltd
Tel: +44 (0) 1787 881623

Published by:
WeKnow Media Ltd.
Suite 138, 80 Churchill Square,
Kings Hill, West Malling, Kent
ME19 4YU, UK
Tel: +44 (0) 1732 807410

DATA LAKES AND STREAMS GIVE LIFE TO INTELLIGENCE AT THE EDGE

Data, communications and computing infrastructure are all coming together to enable intelligence at the edge so the processes, analytics, management and security are the next step, writes **George Malim.**

Multi-access edge computing (MEC) has existed since 2014 and has often been referred to as mobile edge computing. That tag is fine for wirelessly connected edge devices but increasingly, edge devices can be supported by fixed connectivity hence the change to multi-access. This shift indicates a growing maturity and understanding that it's not just mobile devices that can benefit from intelligence at the edge.

With lower latency networks, enabled by 5G, LTE-Advanced and fixed fibre connectivity, the time it takes to traverse the network from the edge to a centralised server has come down, making various applications practical. However, for many, latency to distant data centres is still an issue and it makes more sense to process or analyse inputs and utilise edge intelligence to provide actionable insights. The edge device therefore becomes not only a sensor collecting information but also something that can take immediate action according to that data. The connectivity is then utilised to update the centralised record and enable functions such as digital twins. ▶

George Malim
Tech Trends

Where is the edge?

Traditionally, the term edge refers to the end of a managed network within 30 miles of an endpoint, such as a connected device or sensor gateway. However, this definition is eroding as the edge is now often closer than that. The edge could be in a vehicle or on the device itself so it's accurate to say the edge is moving further from the centre and closer to the user and the connected devices themselves.

Hyperscale service providers, communications service providers (CSPs) and IT companies are all trying to deliver their definition of edge computing but confining definitions are not helpful. What's really needed is an open understanding that the edge is what an organisation wants it to be. For some, that will be storage and compute resources at the cellular network edge while for others, the device itself will contain processing, storage and analytical power to perform edge intelligence at the device level. In contrast to traditional hub and spoke architectures in which edge data flows up and down a spoke to a centralised hub, MEC utilises a distributed cloud computing model that enables processing and storage of data at the network edge. These typically smaller-scale resources are closer to the devices and the end user and enable a cost-efficient, optimised network architecture. Don't forget that traditional data centres and cloud computing has never been free so often MEC is not an additional but an alternative cost and one that can be offset by the business benefits of low latency or the localised processing of large volumes of data that are critical to high-quality services.

With this edge infrastructure in place, the next step is to extract intelligence from it. The intelligent edge combines connectivity, computing power, artificial intelligence (AI) and data analytics so data can be acted upon more quickly and much closer to the point at which it is captured. As organisations generate ever-greater amounts of data they are engaging in more complex operations and the markets they operate in are transforming at the same time. Real-time data places more demands on businesses to sense more and be responsive to the sensor information they collect. Using data to gain insights quickly and effectively drives operational efficiencies and delivers competitive advantages and this underpins the huge expected growth in MEC and edge intelligence.

Naturally, organisations that invest in edge intelligence are doing so in the expectation of a return and the nature of that return is becoming clearer and clearer as the number of deployments increases. There have been numerous trials and proofs of concepts targeting augmented and virtual reality applications, local content distribution, security and surveillance, manufacturing automation, autonomous vehicles, logistics and many other applications. The deployments are so varied that it looks like there's a use case for virtually any organisation that has multiple sites or a fleet of devices in deployment.

However, the projected benefits will not necessarily come easily. Edge intelligence presents a substantial shift from the cloud computing infrastructure of the previous generation. The speed, while one of the main advantages, is also a massive source of disruption because being able to perform tasks in near real-time transforms business processes which therefore have to adapt to become used to the new rate of progress.

Speed also distorts perceptions and expectations and some of these may become unrealistic or excessively expensive to support. For example, an organisation that relies on very high speed network connectivity to enable its application may find that network latency means it cannot meet its customer's service level or that decisions made at the edge are outdated because the data being utilised is stale.

Edge intelligence is also a complex environment involving multiple technology types that all require integration, security and monitoring. There is substantial risk that this complexity can mushroom making it hard to identify root causes of issues or expensive to monitor effectively. These are the downsides and are common to large-scale immature technology roll-outs and the technology industry is working hard to address them. However, it remains early days and some of the first edge intelligence deployments are hampered by complex architectures, poor integration between different systems and lack of availability of low latency connectivity at various locations.

The upside of edge intelligence on the other hand is only just making itself apparent and many of the potential benefits are not yet widely understood. A system that can ingest, store and analyse data and then make decisions that draw on machine learning and AI information while being located at the edge to ensure speed of response has huge value to add. It feels like today we're on the edge of edge intelligence. Hopefully this issue of Tech Trends helps in pushing further over that edge.

TRANSFORMA INSIGHTS

Jim Morrish
Transforma Insights

VanillaPlus
tech trends REPORT

EDGE COMPUTING WILL BE CRITICAL FOR DIGITAL TRANSFORMATION

Technology enabled solutions are becoming ever more critical to the day-to-day operations of many enterprises. Among the most impactful technologies are artificial intelligence (AI), Internet of Things (IoT), cloud computing, and next generation communication technologies such as 5G. Edge computing may garner fewer headlines, but it is a key enabler for many solutions that utilise the emerging technologies listed above, writes **Jim Morrish**, a founding partner at Transforma Insights.

VOLTDB

SPONSORED REPORT

Effectively, edge computing can allow local devices to operate to some extent autonomously of any cloud infrastructure

Fundamentally, edge computing makes processing and storage resources available in close proximity to edge devices or sensors, complementing centralised cloud resources and allowing for analytics close to those end devices. This results in a number of benefits that can be very relevant in an enterprise context, including:

- **(Near) real-time responsiveness**

Analytics that may have previously been undertaken in offsite cloud locations can potentially be supported locally, avoiding the need for raw data to be transferred to a cloud location and for results of any analyses to follow the same path back to a local device. Accordingly, the time taken for a system to respond to new input information can be reduced to near real-time.

- **Improved device-to-device communications**

Communications and the exchange of data between devices that are co-located together can be routed more directly, and without need to transit cloud infrastructure. In fact, edge intelligence can potentially allow processing resources to be shared between a number of local devices, with certain devices able to call on processing resources residing in other nearby devices in a seamless way.

- **Improved robustness, resilience and reliability**

With more analytics undertaken locally to data sources, systems are not as susceptible to disruption in the case that a connection to a remote cloud location fails. Effectively, edge computing can allow local devices to operate to some extent autonomously of any cloud infrastructure. In some situations, edge devices can operate almost completely autonomously and independently of any connection to cloud infrastructure.

- **Improved security and data protection**

With more data processed locally, many security and privacy issues associated with transmitting data to cloud locations can potentially be mitigated, and it can be easier for enterprises to demonstrate compliance with data privacy and data sovereignty requirements. Alternatively, edge computing can be used to anonymise data locally before onward transmission to cloud infrastructure.

- **Regulatory compliance**

Locally managed information potentially only needs to comply with local regulations, rather than multi-jurisdiction regulations that might apply in a cloud environment.

- **Reduced operating costs**

Undertaking more analytics locally, supported by edge computing, can reduce the amount of data that needs to be sent to cloud locations for processing, so reducing communications costs associated with data carriage. It also reduces the burden of processing that must be supported by cloud infrastructure and more importantly the amount of data that needs to be stored in the cloud, reducing costs for cloud infrastructure.

Edge locations, data lakes and data streams

Thus far, we've referred to 'the edge' in quite general terms as being characterised by deploying compute power closer to edge devices or sensors. However, there are many different kinds of edge location as illustrated in **Figure 1** overleaf. ▶

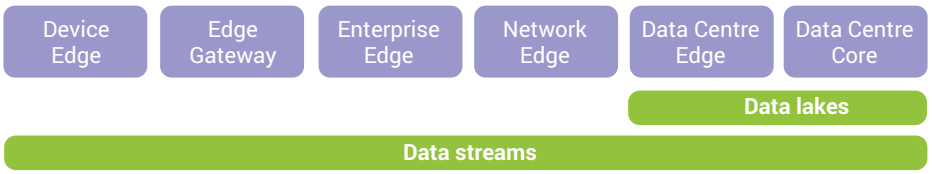


Figure 1: Edge locations, data lakes and data streams

The most local kind of edge computing is where compute power is installed on an actual end device, for example an industrial robot. Next most local would be an edge gateway, located close to an end device; often this would be an industrial computer deployed to connect an operational technology (OT) asset to information technology (IT) systems.

The enterprise edge includes servers and compute power that are local to an enterprise, or on-site at an industrial facility, and sit at the interface between that local network and associated cloud infrastructure. The network edge, specifically multi-access edge computing (MEC), is an evolution of the enterprise edge scenario, where edge processing is provided at the edge of a communications network. This scenario will be increasingly relevant with the advent of 5G communications, and particularly in the case of 5G private networks where the network edge could be located in the same place as the enterprise edge.

Beyond these quintessentially local versions of edge is another definition of edge that has been adopted by providers of cloud infrastructure and which includes the provision of hosting capacity in secondary and tertiary cities: closer to the end devices, from the perspective of a cloud provider. We mention this for completeness only since, from the perspective of a solution designer, such locations are to all intents and purposes still cloud locations. This is entering a grey area though, since the public cloud providers have been partnering with communications service providers (CSPs) to co-locate the cloud edge and network edge together.

In terms of the processing of data, data lakes and large databases of information tend to reside in the cloud, whilst data streams can potentially be processed at any of the edge locations described above.

Dynamics of data at the edge

The next aspect of edge to consider are the dynamics of information, the data, as it flows over edge assets.

The most local kind of edge computing is where compute power is installed on an actual end device

Typically, today most organisations will establish their analytic approaches and frameworks at central, cloud locations. However, as described earlier in this report, there can be significant benefits to undertaking these analyses closer to the end device. A specific example might be the case of AI-enabled analyses of machine status information, where powerful cloud infrastructure can be used to establish rules to support, for example, pre-emptive maintenance alerts, and then any rules identified can be deployed locally to significant benefit and updated dynamically. The first dynamic associated with the advent of edge computing is a trend to push decisions closer to the edge.

Conversely, raw information originating from a device can be redacted or filtered or aggregated as it flows over edge infrastructure, ensuring that only necessary and meaningful information remains for onward transmission. In the example given above, once pre-emptive analytics have been deployed locally, it simply isn't necessary for the same raw data to be transmitted onwards to cloud infrastructure and the digest form could be sent instead to accelerate the central analytics that resides in the cloud. ▶

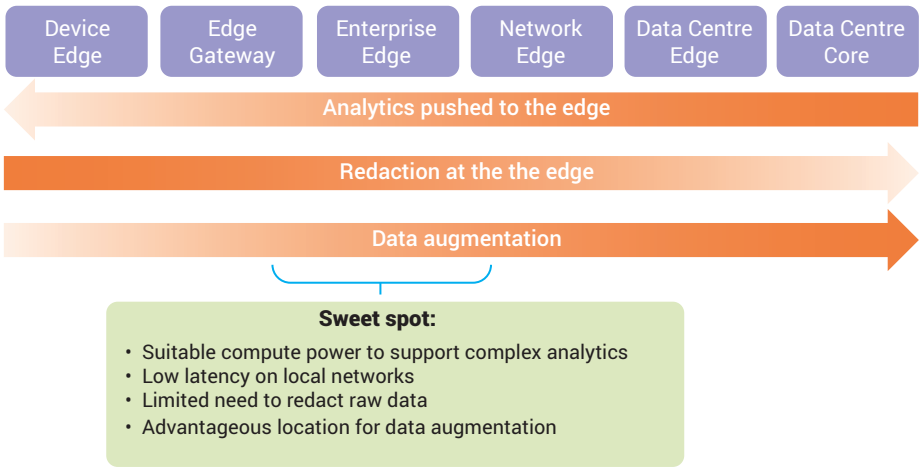


Figure 2: The dynamics of information in the context of edge assets, with sweet-spot

Lastly, information flowing from edge devices can be augmented as it flows across various edge assets by associating contextual and other information, often including information from other local devices. For instance the industrial robot described above may have on-board sensors to monitor operating temperatures and computers located at the device edge or edge gateway level could be well placed to identify when increasing device temperature is a result of some kind of malfunction. However, relevant contextual information might be that a factory air conditioning system has failed, and thus it is likely that increasing device temperatures are simply a product of an increasing ambient temperature. The potential for this kind of data augmentation increases with distance from the end asset. For instance, enterprise edge is a good location from which to analyse and control an entire local facility, including buildings and building condition, the performance of any manufacturing lines, and the organisation of work within the facility.

In general then, volumes of data tend to reduce as they travel away from an end asset, but the information that remains is generally richer and more meaningful.

These dynamics of information in the context of edge assets are illustrated in **Figure 2** above.

Enterprise edge as a sweet spot

Also highlighted in Figure 2, there is a sweet spot for edge computing, identified as the enterprise edge and which combines a number of benefits:

- **Compute power**
  Significant compute power is readily available at the enterprise edge, at a reasonable cost. Compute power deployed at the enterprise edge is also generally scalable, and additional compute power and storage can be added relatively easily.

- **Low latency**
  Analytics at the enterprise edge benefits from high-speed local network connectivity within local facilities and also avoids the need to communicate information to cloud locations.

- **Data redaction**
  Since local connectivity is essentially free within most facilities, there is often limited benefit in redacting data any nearer to end devices than the enterprise edge, while redacting at the enterprise edge can potentially save significant costs associated with processing data using cloud infrastructure.

- **Data augmentation**
  The enterprise edge is an ideal location to draw together all kinds of data streams from within a local facility, ranging from machine performance information to building information and more. It is also an ideal local location to draw in any useful information from remote cloud locations, including information like weather forecasts or supply chain status information.

Overall, these factors make the enterprise edge an ideal location for undertaking all kinds of analyses of the wide range of information generated by a local facility. There is also a particular synergy with the advent of 5G private networks, since 5G networks can support very low latency communications and MEC can provide an efficient bridge ▶

Self-driving cars are a good example of edge computing being used to enable a local system to function autonomously without access to cloud infrastructure.

between the private network and compute power deployed at the enterprise edge.

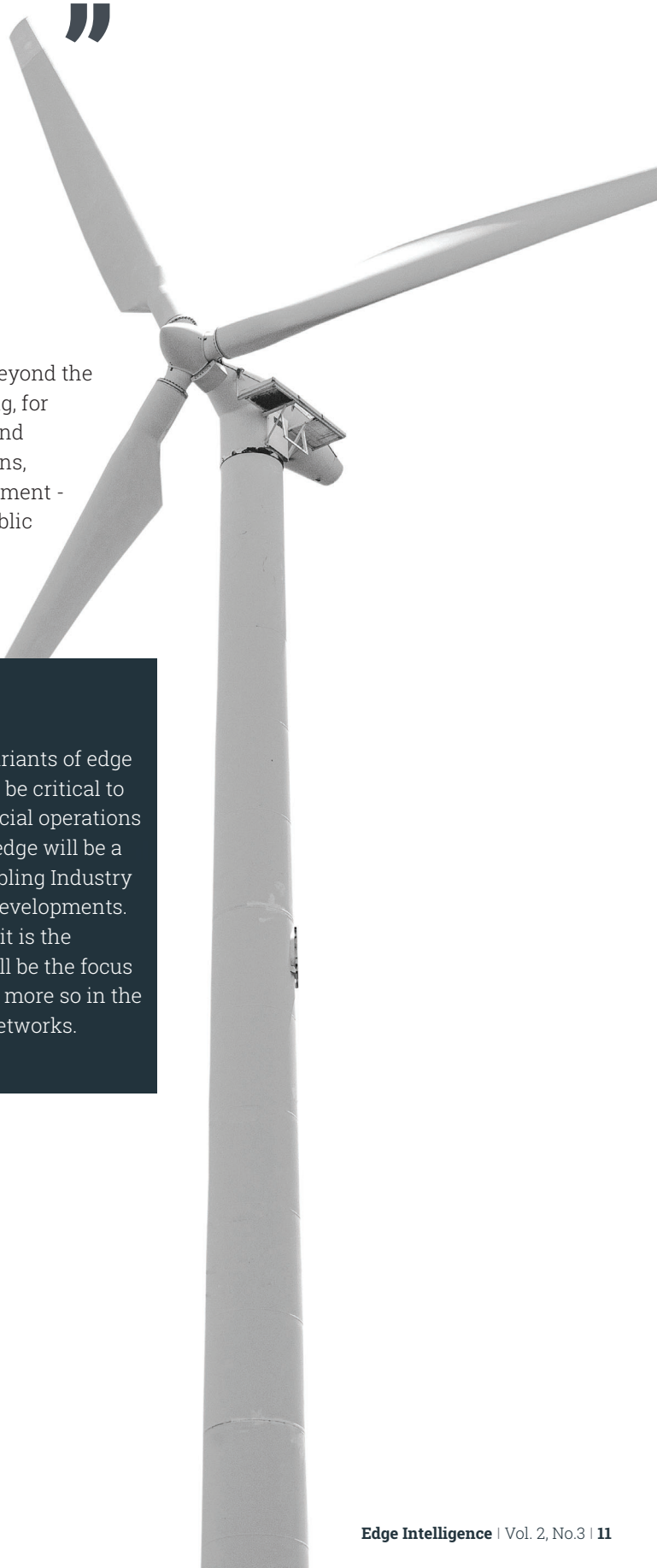
Tangible benefits of edge

The simplest applications of edge technology are probably in building control and facilities management, including heating ventilation and air conditioning (HVAC) control, and security monitoring, access control and alarm systems. In a post-Covid world, such solutions could also extend to include air quality and ventilation monitoring. Clearly such scenarios benefit not only from autonomous operation at the edge, but also the location of analytics at the enterprise edge allows for good oversight of all parameters associated with a particular facility. This kind of application is clearly applicable in almost any vertical sector.

Another potential application is production monitoring and control in an industrial context. For example, an edge-enabled system can be used to monitor a range of production lines and machinery in a manufacturing location to pre-emptively identify any maintenance that is required, and dynamically re-scheduling production given anticipated downtime, availability of parts, and the profile of orders that need to be fulfilled. Information can also be streamed to augmented reality (AR) interfaces such as tablet computers, or video goggles, with minimal latency, Such applications are currently a reality in a production manufacturing environment, but may also find application in smart cities, for example, to control traffic, and also healthcare environments, for example, to manage patient flows. Streaming information to an AR interface can be particularly useful in construction and field support contexts.

The application of AI and machine learning (ML) to closed-circuit television (CCTV) feeds is a key opportunity for edge computing, particularly given the proportion of data communication volumes that it is possible to redact by applying analytic rules at the edge. AI enabled cameras can simply transmit an alert that a certain decision rule has been triggered, averting the need to transmit a full moving image feed to cloud infrastructure for analysis. Such solutions are commonly productised in the form of facial recognition cameras, or security cameras, and ▶

> **The simplest applications of edge technology are probably in building control and facilities management...**



clearly have application in almost all vertical sectors.

Self-driving cars are a good example of edge computing being used to enable a local system to function autonomously without access to cloud infrastructure. The same concepts are relevant in other semi-autonomous vehicles, ranging from factory floor robots to drones, and from automated transportation systems to hospital machinery. Many of the applications associated with these kinds of devices are in fact more complex than autonomous cars, since they must combine local, autonomous, operations with coordination with control applications hosted on cloud infrastructure.

Another application that is particularly suitable for edge computing is the management of an estate of devices, for instance as might be found at a solar farm or a wind farm. Local edge computing devices would be able to compare the performance of individual solar panels or wind turbines, identifying those that need maintenance and ensuring appropriate performance of the overall device estate. Again, such an application concept could have significant relevance beyond the utilities sector including, for instance, agricultural and aquacultural applications, transportation management - particularly rail and public transport, oil and gas extraction and mining operations.

Conclusions

In conclusion, many variants of edge computing are going to be critical to industrial and commercial operations in the near future and edge will be a key technology for enabling Industry 4.0 and many similar developments. Within the edge space, it is the enterprise edge that will be the focus of most attention, even more so in the context of 5G private networks.

VOLTDB

Report sponsor

Multi-access edge computing (MEC) is coming together with the low latency mmWave connectivity enabled by 5G to make intelligence at the edge a reality that truly enables bi-directional digital conversations and gives the potential for technologies such as digital twins to be operated effectively. Here, **Dheeraj Remella**, the chief product officer of VoltDB, tells **George Malim**, the managing editor of VanillaPlus, how the challenges associated with MEC are being addressed and why, for many, once they've looked over the edge, there will be no going back to centralised, remote architectures.

5G AND MEC ENCOURAGE EDGE INTELLIGENCE TO BECOME PERVASIVE

Dheeraj Remella
VoltDB

George Malim: Does the latency offered by mmWave make the goal of linking physical and digital twins in real-time a reality? What are the complex impacts of this on enterprises and can data really be utilised live in the stream of enterprise processes?

Dheeraj Remella: There are a few things that are coming together in this scenario. First, there is the high bandwidth provided by 5G deployments. 5G has three bands of operation: low band, mid-band and high band. High band/mmWave allows massive amounts of data to be transferred over mobile networks. But the caveat is that mmWave doesn't travel too far and is easily interruptible by obstacles. So, an ideal environment for mmWave to be used optimally would be within a building.

This caveat brings us to the second element of the scenario, which is digital transformation. Enterprises often embark on a digital transformation journey but end up trapping themselves into what I call digital transliteration. Digital transliteration is when an enterprise simply digitises its processes status quo using modern technology. I see this as a wasted opportunity to optimise business processes to break free of traditional constraints and evolve to operate in a realm of new possibilities.

The third element here is the need for closed-loop digital twins. Historically, digital twins were state-recording systems whose data was heavily used in analytics to understand the systems, assets, and process behaviour. But given the drive towards automation, digital twins are now the intelligence behind their physical counterparts. Data is not just flowing one way from the physical asset, be it equipment or people or even a business process, to its digital representation. The flow needs to be bidirectional, where data comes from the physical-to-digital path, and the response instructions flow in the digital-to-physical direction. This bidirectional digital conversation completes the picture of the real-time control loop that is going to be a fundamental requirement for successful transformation based on machine-to-machine communication.

The factors in this scenario profoundly impact how enterprises see the value of data. Data's value for analytics is well established by now. But what is hidden is the value of data captured in the first few milliseconds. The low latency event-driven servicing becomes especially significant when we are talking about the digital automation of business processes. Just a few examples of the hidden value from our experience at VoltDB are an 83% reduction in fraudulent transactions completing, 100% prevention of distributed denial of service (DDoS) attacks on the network, and 100% detection of bots before they intrude into a bidding process. All of these have stringent latency service level agreements (SLAs) to make these decisions. Data usage is going to augment post-event analytics by using that intelligence for in-event decision making. Low latency decisions impacting operations require low latency availability of event data. mmWave in private 5G settings will accelerate the shift to tapping into the first ten milliseconds of data's life.

GM: What are the challenges of extracting all the value from the platform?

DR: Most enterprises have already invested in many technologies to fulfil various forms of value extraction from data. When faced with the challenge of tapping into the real-time data streams, these enterprises often resort to figuring out how to make-do with these existing investments. But they fall into the trap of spending inordinate amounts of time, money, and resources, only to end up ▶

making compromises, or even outright failing at the attempt. The low latency expectations for event-driven real-time value extraction from data require many different capabilities to play together in a single unified technology.

When patching together various technologies, enterprises face pitfalls like:

- Communication latency between various layers breaking SLAs
- Infrastructure footprints bloating because each layer requires its resiliency model
- Complex failure handling when aiming to maintain business continuity
- Operations using stale machine learned insights
- Outdated decisions because of not meeting latency SLAs

On the other hand, a unified platform ingests, stores, analyses and makes decisions, which are continuously enriched with machine learning retraining cycles, to address all needs holistically.

GM: Even with the low latency, is the cloud still too far away to enable the round trip to be completed in time to create a meaningful interaction? Does this mean MEC is a must?

DR: This question is an exciting one for me. Cloud computing brings a lot of value to commoditising compute, storage and networking. But, these new low latency expectations and value forces the intelligence closer to the edge. Cloud vendors, for a large part, are centralised in large data centres. This centralisation means that, immaterial of where the event took place, the event data needs to travel quite a bit before realising its value.

The real-time needs of digital transformation are pushing the edge intelligence agenda forward in an accelerated manner. Now the question is, where is edge? You have the edge in devices, gateways, customer premise equipment (CPE), network edge, and then there is the central/cloud data centre. MEC is going through evolution itself. It is transforming from a simple aggregator or local data storage to become more actively involved in edge intelligence. In my opinion, MECs are going to get a lot more capacity allocated to accommodate increasing responsibilities. It might even change the definition of MEC.

MEC on CPE will be the perfect place to land a variety of capabilities:

- Data thinning
- Automated event-driven decision making
- Data preparation for analytics - This becomes even more significant when you consider reducing the infrastructure costs at the central data centre when not all raw data needs to be stored
- Incorporation of machine learning insights into the decision-making process

A plan to succeed in bringing intelligence to the edge must include increasing the role of MEC.

GM: What are the challenges to MEC that need to be addressed? Are costs, security and the definition of the edge itself still barriers?

DR: Currently, MEC is underutilised, although it is in the right place, in the customer premise equipment. There are steps being taken to bring MEC to the network edge, and we've already observed that major communications service providers (CSPs) are partnering with public cloud vendors to get the cloud experience to the CSPs' data centres. In either case, MEC is on single unit boxes and not taking advantage of all the distributed computing architecture innovations. When MEC becomes an integral part of an enterprise's operations, MEC needs to provide the ability to scale and for business continuity by clustering for resilience and performance. The initial thought that comes to mind is 'why invest in near-edge infrastructure when one can have all the hardware necessary procured and managed centrally, especially on top of the central data centre's massive investment?'. Right after that comes the concern of security and intrusion prevention.

My thoughts on this are: if MEC and consequently edge intelligence is implemented well for the right use cases, the returns will easily outweigh the cost. Besides, the intelligence at the edge can even potentially decrease the infrastructure needs in the central datacentre. As for security, it is not something that is an afterthought. Safety is an integral part of everyday operations and weaves into business operations. A no-trust network implementation intertwines checks and prevention into every step and interaction. Now, the biggest challenge is the willingness, or lack thereof, of enterprises to undertake this journey and not shackle themselves with old ways because that is how it is today. ▶

> **The real-time needs of digital transformation are pushing the edge intelligence agenda forward in an accelerated manner**

GM: Is moving intelligence closer to the edge the new standard? Is there no going back?

DR: The bottom line is there is no denying that enterprises are always looking for an edge to take their business to the next level and differentiate themselves from their competitors. With Industry 4.0 getting accelerated by 5G, traditionally non-digital companies are looking to optimise their business processes and utilise communication and computing technology advancements. Innovative leaders that adapt to this fact will leave enterprises that do not behind. Edge will become the most explored area of innovation, bringing better security, optimisations, and user experience. Once the path to the edge is taken, there is no going back.

GM: So where does VoltDB play in all of this?

DR: VoltDB has meaningfully integrated in-memory database and stream processing technologies. This combination brings the best of both worlds, such as data consistency and fast storage with transaction processing from the database world and the ability to integrate with streams for ingestion of data and communication with other applications and systems. Our engineering team has built our technology as a single cohesive product instead of just assembling various open source technologies and calling it a platform. Every step of our design process considers three inextricably linked questions:

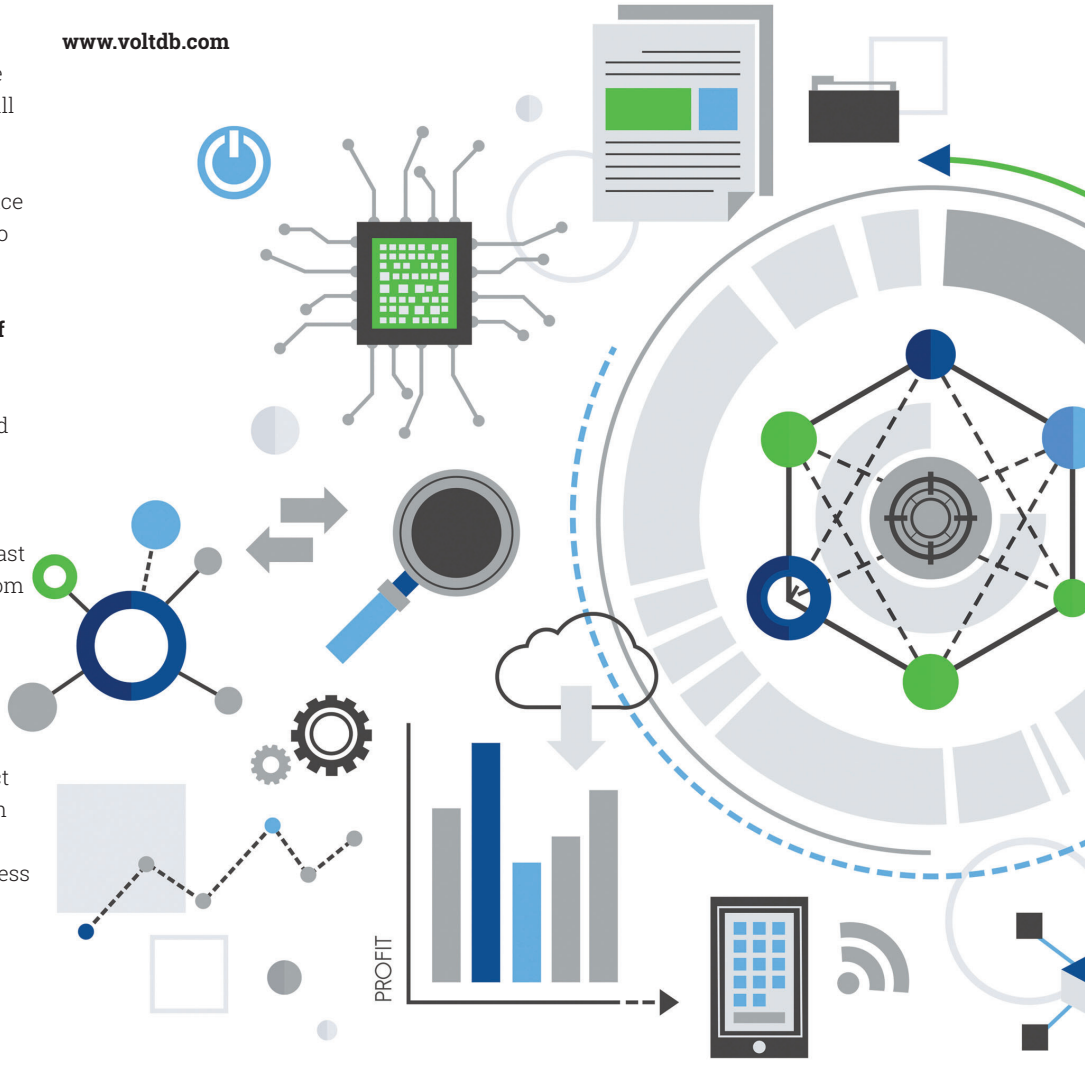
What is the shortest path from an event to a responding action?

What is necessary to drive those actions intelligently?

How can we do this while using the least amount of hardware and resources?

I would highly encourage our audience to check out our paper on Intelligence at the Edge.

www.voltdb.com



WHY LOW LATENCY DECISIONS ARE KEY FOR EDGE SUCCESS

5G, edge computing, digital transformation, digital twins, machine learning and AI appearing together looks like a buzzword bingo. But there is a rational connection between these topics and technologies, writes **Dheeraj Remella**, the chief product officer at VoltDB.

To frame these topics in simple terms, they can be defined as follows:

5G - Provides high bandwidth low latency connectivity to the mobile network, which can be private or public depending on the use case

Edge computing - Brings intelligence meaningfully closer to the source of events

Digital transformation - An exercise that an enterprise undertakes to break free of archaic processes and redefine operations to make use of modern technology

Digital twins - The next step in the digital representation of organisational assets, such as people, assets and processes, moving from simple state recording to full sense control feedback loop

Machine learning - Employing computers and algorithms to understand what the data is telling us

Artificial intelligence - Operationalising the machine learning insights to facilitate ever-evolving 'do-learn-do better' cycles ▶

SPONSORED ARTICLE

Now, with this backdrop of understanding of the terms, let's examine the fundamental thread that connects these together. The singular intent of each one of these technologies is to bring more self-learning automation to business processes. Data generated by enterprise assets and processes should get consumed as close to the source of the events as possible. This ensures decisions are not made - and actions are not taken - on stale information. This would render those decisions and actions obsolete as the universe has since moved on.

The decisions are not made based on a static set of rules. Instead, they have a characteristic of dynamism due to the speed at which events get generated and decisions are made. The overarching business process is fluid to adapt to current conditions. These dynamic rules are generated by training and retraining learning models.

When dealing with real-life processes and industrial applications, 'good enough' is not enough. Every decision and action needs to be correct based on the current situation and business rules. This is where bringing intelligence beyond simple aggregation and to the edge becomes important. Now, what is edge though? Is it in the device or gateway or some kind of on-premise datacentre or network, or is it in the cloud? It depends on the application. But in most cases, one thing becomes concretely evident. The edge on the devices is too narrow to make any meaningful contextual decisions.

Having the edge layer in the cloud is too far away to be able to act on events within a reasonable amount of time. Gateways are slowly becoming irrelevant with narrowband IoT and Cat-M devices being able to connect directly to the network through embedded (eSIM). So that leaves the on-premises data centre and the network edge as the best possible candidates for intelligent interactions. Applications that are more industrial by nature will best benefit by taking the on-premises approach while applications that are more consumer-oriented will benefit from the network edge. But in either case, while the network ping roundtrip latency is important, the service latency in the middle of the communication roundtrip is also essential.

To keep the service latency low, a data platform that serves multiple purposes together is necessary to ensure the data value is extracted without an artificially complicated technology stack. The event data needs to be ingested, stored and aggregated, either as a single business event or a set of events - think complex events, and compared to the aggregated data to measure some key performance metrics. Any deviation in behaviour needs to be acted upon for either monetisation or prevention of some form of threat.

Monetisation opportunities can range from the personalisation of user experience to determining the most profitable end product - at this moment - from refined crude oil. Threats can range from potential machine downtime to robotic network intrusion. Given the core objective of digital transformation is to automate business processes by shifting to a machine-to-machine communication paradigm, the latency of these decisions to invoke appropriate actions needs to be in single-digit milliseconds and, in a more stringent environment, less than a millisecond. If this window is missed, a cascade of inefficiencies is put into play. The decisions are stale, which translates into wrong learnings. This in turn creates insight that is not congruent with the reality of the enterprise.

Depending on the use case, VoltDB's customers typically allocate anywhere between 0.25 milliseconds to less than 10 milliseconds for making the decisions, and those decisions are made with 100% accuracy and uncompromising guarantees for resiliency. Completing intelligent decisions and taking action within this timeframe is no longer a nice-to-have. It is a must-have to uncover the latent value of the data in its infancy. Our customers are able to gain unprecedented advantages ranging from being able to prevent 100% of bot intrusion to being able to take the best next action by adhering to stringent low latency service level agreements (SLAs).

www.voltdb.com

CAN IT BE SAFER ON THE EDGE?

Children are warned to stay away from the edge of cliffs, bridges, roads and rivers, but in computing, the edge is turning out to be one of the safer locations for hosting processing power and analysing data. **Antony Savvas** explores why this is the case, while also addressing the security concerns that do exist around edge computing and assessing how these are being resolved.

The edge computing market includes hardware, edge nodes or gateways and servers, sensors and routers, software including databases and analytics, services and edge-managed platforms. The global edge computing market size covering all these areas is anticipated to reach US$43.4bn by 2027, seeing a CAGR of 37% over the forecast period, according to a March 2020 report by **Grand View Research**.

Covering just hardware, platforms and services, rival research house **MarketsandMarkets** projects the global edge computing market will grow from US$3.6bn in 2020 to US$15.7bn by 2025, at a CAGR of 34% during the forecast period.

Some key players in the edge computing market are ADLINK, Altran, Amazon Web Services (AWS), Axellio, Belden, Cisco Systems, Clearblade, Dell Technologies, Digi International, EdgeConneX, Edge Intelligence, Edgeworx, FogHorn Systems, GE Digital, Google, Hewlett Packard Enterprise, IBM, Intel, Juniper Networks, Litmus Automation, MachineShop, Microsoft, Moxa, Nokia, Sierra Wireless, SixSq, Vapor IO, VMware and VoltDB, among many others.

The main drivers

5G is expected to act as a big catalyst for market growth. Applications using 5G are expected to change traffic demand patterns, "enabling technology growth avenues" for the communications service providers (CSPs), said Grand View. The cloud market leaders see this as a threat and have started investing in the edge ecosystem themselves by engaging in partnerships with CSPs.

CSPs are expected to embrace new opportunities in the multi-access edge computing (MEC) market place, said Grand View. MEC allows providers to mitigate network congestion and ensure higher application performance by bringing processing tasks and running applications closer to the cellular customer. The implementation of MEC at mobile base stations or edge nodes is expected to facilitate the rapid and flexible deployment of new services and applications for customers, which "promises healthy market growth", Grand View said.

Furthermore, there is an anticipated wave of micro edge data centre (EDC) capacity that differs from large centralised data centres. This new capacity is expected to range from small clusters of edge cloud resources located on street-lights to a few racks located in a shelter at the base of a cell tower or inside buildings. ▶

Antony Savvas

In addition, 5G networks can use EDC facilities to provide efficient local data services, redirecting edge traffic away from the carrier networks to local public internet networks. Various start-ups, such as **EdgeMicro**, are in the process of deploying commercial mini data centres with IT computing stacks, redundant cooling, fire suspension and biometric security.

MarketsandMarkets warns however that the costs of moving to the edge can be significant. It says: "Edge computing might reduce data transmission and storage costs through localised processing, but investing in edge infrastructure still adds to the capex of companies, including heavy investment in edge nodes, other edge devices and edge data centres."

They would also be required to spend more on making the devices and the entire network secure. But, MarketsandMarkets added: "The edge infrastructure cost is a restraining factor, though, with advancement and continuous R&D, the cost of edge technology is expected to reduce soon." So despite some obstacles, there is plenty of interest in processing data at the edge. But why is it safer, overall, for receiving, processing and exchanging data - instead of relying on cloud data centres to do everything?

Edge safety and performance

Dheeraj Remella, the chief product officer at edge database and analytics provider **VoltDB**, says: "It is pure and simple. Getting closer to the event source decreases the time elapsed before the event data becomes stale. You can apply this to a variety of value extraction principles such as personalisation, operational automation, preventative maintenance and most importantly securing assets, processes and even employees and customers."

"While the central cloud data centres offer larger infrastructure capacity, just the travel time for the data to get to the data centre robs the enterprise of the opportunity to respond to the event in a timely manner," he adds. "You can observe this emphasis on getting close to the event source in efforts by CSPs partnering with cloud vendors for the deployment of edge data centres. And enterprise customers stand to benefit by bringing the intelligence even closer to their premises to support lower latency applications and deliver improved security."

So is it just 5G and the Internet of Things (IoT) that have driven the edge data processing and analytics market? Or has this been an evolving migration driven by the needs of communication service providers and enterprises, and if so, what else do they want out of the edge?

"5G and IoT are accelerants to this awakening to the real-time needs of digital transformation," Remella explains. "Customers already in the space of tapping into event-driven real-time decisions and automation have benefited from revenue increases and greater security. These organisations were pioneers because they saw the value of what is being ignored in the first ten milliseconds or less."

"What 5G brings to the table, especially when combined with narrowband IoT (NB-IoT) and Cat-M, is the ability to do away with hops and interim aggregators to get to the network directly," he says. "Gateway-less IoT is going to become mainstream and this will allow much richer intelligence near the edge, be it machine learning or event-driven real-time decisions."

It's time to get sassy

With more data moving to the edge, last December analyst house Gartner promoted the Secure Access Service Edge (SASE) framework. Pronounced sassy, SASE enables increasingly distributed and mobile workforces to remotely and securely access corporate networks and clouds. Interest in SASE has grown substantially due to Covid-19 as enterprises recognise its potential as a business continuity solution.

SASE combines network security functions - such as secure web gateway (SWG), cloud access security broker (CASB), firewall as-a-service (FWaaS) and zero trust network access (ZTNA) - with software-defined wide area networks (SD-WAN) to support the dynamic secure access needs of organisations. These capabilities, says Gartner, are delivered primarily as-a-service with the ability to identify sensitive data or malware and the capability to decrypt content at line speed.

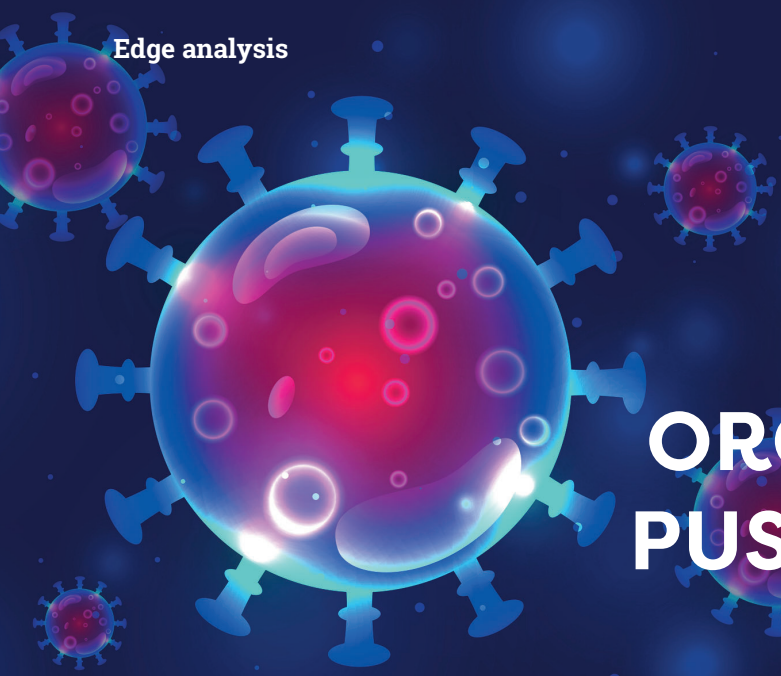
"Although SASE is relatively new, the ongoing pandemic has fostered the need for business continuity plans that include flexible, anywhere, anytime, secure remote access at scale, even from untrusted devices," said Gartner analyst Joe Skorupa in August 2020. "Mobile workforce, contractor access and edge computing applications that are latency sensitive are three market opportunities. Over the last three months, SASE has been adopted by more than 40% of global remote workers."

Many suppliers have already launched SASE-based products onto the market, with **VMware** one of the latest to do so in September 2020. The VMware SASE Platform converges cloud networking, cloud security and zero trust network access with web security, to deliver flexibility, agility and scalability for enterprises of all sizes, said the vendor. With the edge data processing/analytics market expanding as billions more things are added to the edge, we are inevitably going see more edge data leakages and security mishaps hitting the headlines.

To mitigate matters VoltDB's Remella says edge intelligence is going to be a central theme in data-driven enterprises. He says real-time dashboards at operations centres are going to give way to automated processes that will decrease the burden of manual intervention and resolution.

"Machine learning is going to slowly start moving elements from the central data centre to the edge data centre," he explains. "What if we can do away with all the unnecessary raw data and instead digest it locally near the edge? And the central data centres are intelligence aggregation points where only the learnings from the edges need to be mixed together to create higher level models, which can then be sent back to individual edge centres to incorporate into local decisions?"

This would reduce the amount of data being stored in cloud data centres, reducing energy consumption, with organisations only having to store human transactional data covering financial or medical or insurance matters, for instance. "I still encounter enterprises talking about storing several petabytes of data and the need for massive high-performance compute clusters to churn through that data," says Remella. "People might call me utopian, but I strongly feel we can get there."



PANDEMIC PUTS PRESSURE ON ORGANISATIONS TO PUSH INTELLIGENCE TO THE EDGE

Multi-access edge computing (MEC) has turned the corner from being seen as an interesting alternative to the hub and spoke architecture of cloud computing to being recognised as a vital technological enabler of low latency, connected intelligence. **George Malim** examines the latest analyst predictions for the technology's increased uptake.

It comes as no surprise given the drivers that are affecting the technology market that substantial growth in MEC is expected. The arrival of 5G, with its low latency communications capability, the dispersal of more and more people and endpoints to the edges of networks, and the increased familiarity with applying intelligence at the edge have all come together to provide important stimulae for growth. This is borne out by research from **Dell'Oro Group** which projected in its July 2020 forecast that the MEC market will grow at a compound annual growth rate (CAGR) of 169% between 2019 and 2024.

Coronavirus-related lockdowns and new ways of working are demonstrating that more needs to be done from remote locations and, with office work now moving away from large, centralised locations, so too are computing resources. Dell'Oro's latest estimate is a revision of its January 2020 forecast and reflects the industry collaboration and momentum being achieved. The firm says that this traction has caused it to double its predicted CAGR. Nevertheless, the firm does not project steep immediate growth. It continues to expect the market to start slowly before accelerating significantly in the second half of the 2019-2024 forecast period. The firm expects China to be the leading market in terms of the scale of MEC deployments and the country will represent the largest regional market.

Frost & Sullivan's recent analysis, *5G and Edge Computing—Cloud Workloads Shifting to the Edge, Forecast to 2024*, also finds that edge computing is a foundational technology for industrial enterprises because of the lower latencies, robust security, responsive data collection and reduced costs it can offer. The firm reports that, in spite of being at a nascent stage, the MEC market is estimated to grow at a CAGR of 157.4%, garnering a revenue of US$7.23bn by 2024, up from US$64.1m in 2019.

"The recent launch of the 5G technology coupled with MEC brings computing power close to customers and also allows the emergence of new applications and experiences for them," said ▶

> **"...many believe lockdowns and quarantines have only accelerated digital transformation in general"**

> **5G and MEC are an opportunity for telecoms operators to launch innovative offerings**

Renato Pasquini, the Information & Communication Technologies research director at Frost & Sullivan. "5G and MEC are an opportunity for telecoms operators to launch innovative offerings and also enable an ecosystem to flourish in the business-to-business (B2B) segment of telecoms service providers using the platform."

A recent **Heavy Reading** survey has also uncovered that communications service providers (CSPs) see MEC as an opportunity. Almost 85% of network operator professionals polled believed that edge computing would be critical or important to their network evolution strategy, the firm says.

For Pasquini, the CSP opportunity comes second to the edge applications and software market. "From the perspective of the MEC ecosystem, software - edge applications and solutions - promises the highest CAGR followed by services," he said. "[These are composed of] telecoms operators' services, cloud providers' infrastructure-as-a-service, and edge data centre colocation services."

MEC is not simply a CSP opportunity. Frost & Sullivan predicts that approximately 90% of industrial enterprises will utilise edge computing by 2022, presenting immense growth prospects for MEC market participants. Consulting firm, **Deloitte**, has reported that more than 80% of executives surveyed in a recent report believe that advanced connectivity is very or extremely important to their ability to capitalise on advanced technologies such as artificial intelligence (AI), edge computing and data analytics.

In fact, the importance of edge is being recognised to the extent that **ABI Research** expects AI chipset revenues from edge deployments will dethrone cloud as the leading market by 2025. At that point edge AI chipsets will generate revenues of US$12bn per year, outpacing the cloud AI chipset market, which will reach US$11.9 billion in 2025.

Cloud remains at the centre of AI today with most workloads served in public and private clouds. However, the industry is beginning to shift away from centralised cloud resources driven by the need for privacy, cybersecurity and low latency and AI training and inference workloads on gateways, devices and sensors is starting to happen. Recent advancements in key domains, including connectivity to cloud computing, new AI learning architecture and high-performance computational chipsets have played a critical role in this shift, the firm says.

"As enterprises start to look for AI solutions in the areas of image and object recognition, autonomous material handling, predictive maintenance and human-machine interface for end devices, they need to resolve concerns around data privacy, power efficiency, low latency and strong on-device computing performance," explains Lian Jye Su, a principal analyst at ABI Research. "Edge AI will be the answer to this. By integrating an AI chipset designed to perform high-speed inference and quantized federated learning or collaborative learning models, edge AI brings task automation and augmentation to device and sensor levels across various sectors. So much that it will grow and surpass the cloud AI chipset market in 2025."

The pandemic has disrupted demand for many smart consumer devices, notably smartphones, smart home, and wearables. These would have helped further stimulate deployment of AI accelerating technologies at the edge and, at the same time, implementation of AI in industrial manufacturing, retail and other verticals has been postponed or put on hold. This is likely to be only a temporary state of affairs.

"ABI Research expects the market to rebound in 2022. It is important to note that the impact on the chipset supply chain has been relatively minimal since fabrication factories in Singapore and Taiwan remained operational during the outbreak," Su points out, adding that vendors of key connectivity technologies such as 5G, Wi-Fi 6, and autonomous solutions such as autonomous vehicles see minimal impact to their product roadmaps. "Catalysing many other emerging technologies, edge AI will pave the way for a variety of new business opportunities in the consumer and enterprise segments."

Covid-19 may have caused a slight pause in production of technologies that pave the way towards intelligence at the edge but it has also made people look in more detail at the edge as they assess what the new world of work will look like. In addition, many believe lockdowns and quarantines have only accelerated digital transformation in general and that can only further accelerate the march towards greater intelligence at the multi-access edge.

MEC PROVES THE NEED FOR SPEED

There are some very sensible, industrial strength cost justifications for installing multi-access edge computing (MEC) but sometimes attention can be diverted by more fanciful projections, writes **Nick Booth.**

Forget about augmenting and virtualising realities, there is plenty of work to do in the factories, according to Dean Bubley, founder of **Disruptive Analysis**. "A lot of what gets discussed in 5G and edge-computing conferences is either hyped or undeliverable," he says. "Many of the use-cases can be adequately serviced with 4G mobile or Wi-Fi - or a person on a bicycle delivering a USB memory stick."

However, teleprotection systems for high-voltage utility grids will demand latency of between 6-10ms. Bubley has identified a range of other emerging practical use cases in manufacturing. Sensitive industrial process-control systems will ruin their finely-calibrated machinery if they wait ten milliseconds to respond. Image sensors and network sync mechanisms work in nanoseconds and a picosecond is a long time to a Photon sensor. Ultra-fast laser pulses for machining glass or polymers need feedback in femtoseconds.

Kiva Allgood, the head of IoT at **Ericsson**, agrees that manufacturing is the most realistic target market for MEC, and identifies fabrication and 3D printing as avid consumers. ▶

Meanwhile, software vendor **VoltDB** reports areas in which clients face problems now. "It's hard to bill accurately under the present circumstances and a local presence would help," says Dheeraj Remella, the company's chief product officer.

Security, too, is becoming unfeasible without a rapid response. Nobody can defend against a distributed denial of service (DDoS) attack if their intelligence has to be sent to a distant data centre to be processed. All reporting and reconciliation of intrusions must be instant. Similarly, automated processes that are fine-tuned by artificial intelligence (AI) will falter if the machine learning is delayed, says Remella.

Communications service providers (CSPs) can't work out hyper-personalised offers to each subscriber and capitalise on the mass marketing campaigns unless they can live in the moment. Which calls for some heavy lifting at the edge.

There are exceptions. No credit card company can run a full fraud inference routine on a server at the edge because there is too much information and context to process. In this case, not every event is significant, so the client can thin the data out before sending it to a central data centre.

No company should digitise any process if they can't guarantee its safety, says Remella, and MEC can be a reassuring presence. Security probes and the cross-checking of events are now inextricable parts of the new process definition. The meta-data describing these jobs can triple the amount of event information generated.

The cases for MEC needs clarity and that is provided by definition, says Remella. Edge is a contentious term and it's possible to get lost in the semantics of defining in-devices, gateways, customer premises and network edges. It's better to ignore these philosophical topics – the question that matters is: How close can I get to the event before it peters out?

Thomas Neren, Ericsson's head of dedicated networks, is another who decries the lack of focus and warns that people are getting confused over 5G's potential to meet the challenge.

"We need to standardise our definitions and make sure we are all talking about the same thing," says Neren. "In 5G terms, one millisecond is the latency of the radio interface only, whereas a measure of total, end-to-end latency is the sum of response times across the transport layer, the core, cloud storage and computing systems."

Creating ultra-low latency across a high capacity wide area, would take a massive investment and the natural starting point for confined areas might be dedicated networks, says Neren.

Can a 5G network deliver the same benefit as MEC?

Eventually, yes, says Ericsson's Allgood. In theory Releases 15 and 16 of 5G will deliver the low latency and high frequency connectivity needed for IoT.

Anyone who waits for 5G will be disappointed says Joshua Norrid, senior technical director at **DataStax**, because many 5G standards don't specify how data is stored, managed or replicated. "Enterprises want a consistent approach to managing data from the edge to centre and in-between and they won't find this model flexible enough," he says. "If it's hard to use your data, then developers will struggle to build the IoT applications that will need it."

Dr Paul Carter, the chief executive of mobile network benchmarker **Global Wireless Solutions**, urges caution over your choice of MEC. "Think about the connectivity charges: is it cheaper to connect via Wi-Fi or a non-cellular network," says Carter.

Still, smartphones and tablets will eventually bring AR apps to the masses at which point CSPs will have to compete for subscribers on their ability to match the digital picture to the physical, predicts Kevin Riley, the CTO at **Ribbon Communications**. Another increasingly popular security app, face recognition for site access, will need MEC as it processes data locally. The systems will fall down if they need to send buckets of raw data back to HQ.

Neither will be realistic unless the computer comes back with an answer in under 12 milliseconds, and "Ideally less than seven," says Riley.

Latency is important, for application developers, enterprises and many classes of IoT device and solution. But we have been spectacularly vague at defining what low-latency actually means, and where it's needed, says Disruptive Analyst Bubley. What is likely is that average latencies will fall with 5G. An app developer that currently expects a 30 to 70ms latency on 4G, and probably lower on Wi-Fi, will gradually adapt to 20-40ms on mostly-5G networks and eventually 10-30ms. If it's a smartphone app, they likely won't use ultra-reliable low latency comms (URLLC) anyway. Specialised IoT developers in industrial settings will work with specialist providers, like mobile network operators, fully-private networks and integrators, to hit more challenging targets. "Returns on investment and safety constraints will justify the cost," says Bubley. They may even get down to one millisecond at some point in the medium term, but it's far from clear they will be contributing massively to the edge-providers' bottom line.

Joshua Norrid
DataStax

Dr Paul Carter
Global Wireless Solutions

NEW eBOOK

THE HIDDEN INFLECTION POINT IN 5G

WIll the changing definition of real-time break your existing tech stack? Dive into our in-depth analysis, including 5 points to help you evaluate your 5G readiness.

VOLTDB

> GRAB YOUR COPY

For more information, please visit **VOLTDB.COM**