Intelligence at the Edge

5G has been steadily bringing the benefits of speed and scale to the forefront, and is clearly going to change the world of communication in ways never before realised. But who is going to benefit from all these investments? When one reads about 5G in mainstream media, it is often misleading, says Dheeraj Remella, chief product officer at VoltDB. There is a lot of talk about 5G and how it provides blazingly fast internet on mobile devices. At times, as with many new technologies, facts give way to hype and sensationalism. To help clear that up, let's start by clarifying a few things about what 5G is.

5G has many components that will be different from 4G including the radio, network core and the various bands. Within the context of this paper, we will address the bands of 5G.

5G operates across three bands: low band (< 1GHz), mid-band (3 GHz to 24 GHz) and high band (> 24 GHz, also known as mmWave).

- While the low band provides higher coverage, the speeds are less than expected from LTE (Long Term Evolution to 4G).
- The mid-band is where most operators are operating today, between 3.3 GHz and 3.8 GHz. More frequencies are getting opened to the 5G usage in the mid-band. The term mmWave refers to the spectrum above 24 GHz, which opens up much higher speeds and lower latencies.
- The higher frequencies have an inherent constraint on the distance the signal can travel and are susceptible to interference. These constraints restrict the applicability of the technology across vast areas.

Enterprises are the primary beneficiaries

When you put together the facts that mmWave offers - the fastest connectivity at the lowest latency, and the 5G objectives of low latency, connection density, and support for machine-type communications, it's evident that enterprises will be the primary beneficiaries of 5G advancements.
Connection latency of 1 millisecond, while supporting a connection density of 1 million devices per km2, is necessary to accelerate business process digitalisation in many industries. 4G connectivity of 2000 devices/km^2 is not going to be sufficient to achieve this objective. ▶

SPONSORED FEATURE

Business process transformation

Let's take a more in-depth look into what it takes to accomplish business process transformation. A high-level process overview could look like this:

- Create digital twins of the physical assets. The initial stage of creating digital twins is to create a state machine that stores the state changes of an asset.
- Use this state change information to analyze and predict future changes.
- Determine what actions can address these predictions, i.e., prescriptions.
- Apply the predictions rules and prescriptions rules to the state change events to decide the best action.
- Refine these predictions and prescriptions with updated information on an on-going basis.

These steps map the evolution of the usage of data as well. First came databases to store data for querying. Then came big data for analytics: predictive and prescriptive. Today's demands for speed and volume require a data platform that makes decisions — leveraging the state database and intelligence from big data - to drive actions.

Predictions can be mainly of two types:

1. Capitalisation opportunities
2. Threat prevention opportunities

Prescriptions can range from presenting the best offer to a customer, to picking the right next step in an industrial automation setting or even preventing a
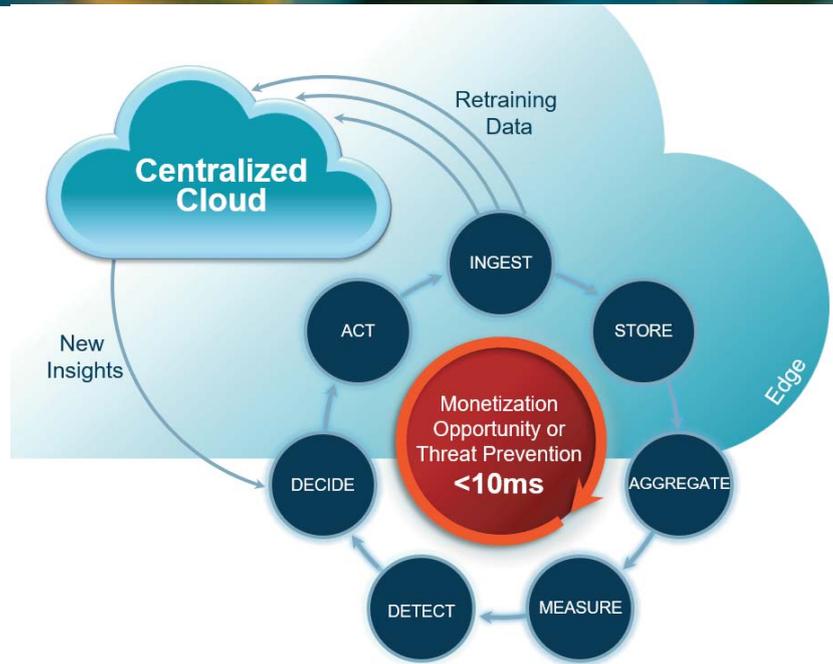
network intrusion. There is real differentiating value in implementing these prescriptions within milliseconds of detecting the prediction.

5G connectivity has ushered in a redefinition of real-time and data-value. Real-time is now in milliseconds. But, more importantly, the value of data has moved from just state recording and analytics to in-event decision making. This shift in the data-value definition requires a data platform that addresses the value extraction process's entirety. ▶

"The goal remains the same; to deliver the lowest latency journey from an event to a responding action."

To fully utilise data, there are two cycles of processing:

1. The maintenance of state and decision-making processes together
2. Big data analytics for predictive and prescriptive insight generation

As the diagram shows, the fast cycle feeds the slow cycle with new retraining data, and in turn, the slow cycle feeds the decision-making process with new insights. This separation of the slow cycle and the fast cycle brings us to this paper's main point. Why does one need to move intelligent decision-making to the

edge? Simple: To deliver the lowest latency journey from an event - single or complex - to a responding action.

Decisions close the source

A system's decisions need to be close to the source of the events. This proximity enables the system to complete the round trip required to create a meaningful interaction. Decisions in the cloud are too far away from the source and actors. The round trip to the data centre, alone, will make the decision moot. On the other hand, the decision cannot be made inside devices such as gateways since this makes the decision contextually too narrow and will render it inconsequential.

There are many examples of the monetary value hidden in the first few milliseconds from an event taking place. For example:

- A credit card company was able to reduce fraud by 83% by applying 1,500+ rules on every transaction within 30 milliseconds.
- A telecom operator increased their offer acceptance rate by 253% by presenting the best offer at the moment of engagement - deciding what to show and then showing it in under 10 milliseconds.

But in the end, the benefits depend on the application and the organisation's data usage maturity. An Internet Service Provider (ISP) can detect and thwart 100% of the distributed denial of service (DDoS) attacks on their customers' sites. An AdTech security company can identify bot publishers in less than 10 ms and prevent them from stealing campaign funds by not bidding on their real estate.

Early forms of transformation

These examples are some of the early forms of digital transformation. Several industries are still misperceiving digital transformation as just translating existing processes into their digital versions. These are the cases where, although the original intent was to make it a little better, one ends up with a far worse situation (Look up Verschlimmbesserung - It is one of these German words that captures a complex scenario in a single word). Effective digital transformation requires a rethinking of the existing business processes to do the right thing in the right way instead of a business-as-usual approach.

Industry 4.0, with its reimagined business processes due to better connectivity and fast communication, is a seemingly obvious recipient of 5G benefits. But a low latency communication is not that beneficial when the service that is sitting in the middle of this communication round-trip is situated far away or ▶

The author is **Dheeraj Remella**, chief product officer, VoltDB

takes too long to make decisions. Multi-access edge computing (MEC) is a ripe space to address the real-time needs for business process transformation and exploit the connectivity improvements.

Multi-access edge computing

Historically, organisations used MEC for data thinning and aggregation but not much for process automation. It is becoming apparent now that with modern hardware, next-generation MEC can drive automation onsite or at least at the network edge. Edge computing is going through the same transformation as physical computing went through to cloud computing. In telecom, Advanced Telephony Computing Architecture (ATCA) gave way to Virtualised Network Functions (VNFs), which in turn gave way to Containerised Network Functions (CNFs).

Intelligence is moving from firmware within devices to software for commodity hardware. The need for agility and an ever-increasing pressure to shrink the time to market for new solutions is driving this shift to software-enabled everything.

Next-generation MEC brings the scalability of distributed computing and low latency of gateway-less communication over private 5G. The software transformation brings the agility of applications (or microservices) and data running in containers. But there is a caveat here. There are conflicting forces in action. Microservices' dictum is to separate business logic and data. Teams that take that message verbatim end up with applications that depend on pure data storage technologies. But for this combination of applications and storage technologies to work together, data needs to be moved to the processing layer, i.e., the application microservices. This need to transfer data to a separate processing layer is in direct contrast to one of container-based architecture's constraints: low network bandwidth allowance.

To address this misaligned architecture, one needs to rethink the bifurcation point away from orthodox business logic and data standpoint.

The data is of two types:
1. Business-related data
2. Application state-related data

Similarly, the business logic is of two types:
1. Application state control logic
2. Data-driven decision-making logic

Once these second-level details come in place, it becomes clear that to utilise the network bandwidth optimally, one will need to process data within the data storage platform. This finer-grained separation of concerns will eliminate expensive failure handling and resilience management in catastrophic failure, sub-optimal network usage.

VOLTDB

Conclusion

In conclusion, moving intelligence closer to the event source will become the de facto pattern of fast data solution building. But done naively, the efforts will result in a Verschlimmbesserung situation! The definition of the edge itself will, of course, vary from application to application. Communication service providers are actively involved in the network edge by colocating their data centres with public cloud vendors' edge data centres; think of Verizon's partnership with AWS Wavelength or AT&T's collaboration with Azure Edge Zones. In industrial automation, the edge is at a plant level while in autonomous cars, it is in the vehicle.

Regardless of the edge definition, though, the goal remains the same; to deliver the lowest latency journey from an event to a responding action and, in doing so, reap the tangible business benefits of intelligent decision-making at the edge.

VoltDB empowers global organisations to leverage emerging 5G latency standards to power new revenue opportunities, transform their business and operational support systems, or develop strategic integrations for their enterprise customers. The platform instantly derives value from anomalous events captured across multiple streams of fast data, directly influencing in-the-moment monetization, preventing digital fraud, and supporting 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. For more information, please visit **voltdb.com/**. ⊘

"Moving intelligence closer to the event source will become the de facto pattern of fast data solution building."