

GIGABYTE 5G and Edge Solutions

Abstract

With growing coverage of 5G network, smart manufacturing, automotive technology and smart cities are driving huge demand of edge data centers. They are usually deployed near data sources and play essential roles in supporting large amount of IoT devices in real time. According to Gartner's estimation, by 2025, there will be 75% of data being processed outside centralized data centers, which indicates the huge growing potential of edge computing.

GIGABYTE's "E-Series edge servers" are the fruition of GIGABYTE's long history of providing 5G and edge computing solutions for key clients. The first model of edge server was applied by NVIDIA to develop NVIDIA Aerial SDK, a framework for building 5G applications. In addition, GIGABYTE's edge server solutions also played important roles in VR concerts hosted by Taipei Music Center and assisted New Taipei City Police Department to maintain public order during huge outdoor events. GIGABYTE's complete edge server solutions are aimed to empower enterprises and institutes to unlock the infinite 5G-related business opportunities.



What are the Attributes of an Ideal Edge Server in the 5G Era?

At its core, an edge server is a solution designed to combine edge computing network architecture with 5th Generation mobile cellular networks, commonly known as 5G. 5G provides greater bandwidth and higher speeds of up to 10 Gbit/s—that is ten times faster than current 4G networks. Edge computing is a way to position computation as close to the end user as possible, further reducing latency and bandwidth consumption. The upshot is that mobile devices can now connect to data centers just as easily as if there were a physical cable between them.

The implications are groundbreaking. Ultra-high-definition video streaming and cloud gaming are just the tip of the iceberg. Radical innovations such as autonomous vehicles may soon become a fact of life, all thanks to 5G communications and edge computing solutions.

The International Telecommunication Union Radiocommunication Sector (ITU-R) defines the three main application areas for 5G communications as Enhanced Mobile Broadband (eMBB), Massive Machine Type Communications (mMTC), and Ultra Reliable Low Latency Communications (URLLC). In layperson's terms, eMBB refers to the massively expanded bandwidth that will enable 4K or even 8K video streaming, as well as truly immersive virtual reality (VR) and augmented reality (AR); mMTC stands for high connection density supporting an enormous number of devices (such as sensors on self-driving cars) that transmit small data packets sporadically; and URLLC offers a target latency of 1 millisecond and 99.999% reliability, which is critical for autonomous vehicles to respond to road conditions, to give one example. An ideal edge computing solution in the 5G era would need to be compatible with these three application areas.

In practice, GIGABYTE E-Series Edge Servers rise to the challenge by meeting multiple requirements. First and foremost, their processors offer superb computing power and contain a high number of cores to service multiple devices simultaneously. Second, an edge server can handle a diverse range of tasks, which is why it features numerous expansion slots to support GPU cards (which are ideal for parallel computing) or networking cards. This provides a degree of flexibility and multifunctionality for the edge server, which may be required to carry out a plethora of tasks in a wide gamut of locations.

The counterweight to these performance requirements is the matter of availability and cost-efficiency. The edge server must be able to physically fit in all kinds of tight spaces, such as in an office cabinet or at the base of a cellular site. It should offer good power usage effectiveness because the power supply may be limited or expensive. Since air conditioning may not be accessible, temperature control should also be considered. Also, the edge server should be easy to maintain and repair to ensure stable operations. On top of all these requirements, edge servers must also offer competitive CAPEX and OPEX, since a lot of such servers will be required to support the 5G communications infrastructure.

GIGABYTE was able to achieve such a succinct and in-depth understanding of how to build ideal edge servers for the 5G era because it has provided edge computing solutions for clients in different industries for many years. Two success cases pivotal to the creation of E-Series Edge Servers are the VR concert viewing system in the Taipei Music Center, which is a shining example of a micro data center that supports edge computing; and the “Private Cell Mobile Command Vehicle” deployed by ITRI and New Taipei City Police at the height of the Pokémon GO craze in Taiwan, which is exemplary of a 5G private cellular network analogue that supports computing, storage, and networking in a compact server system.

Taipei Music Center: Staging VR Concerts with Micro Data Centers on the Edge

Officially opened in 2020, the Taipei Music Center is a futuristic facility for recreation and entertainment. It is also the perfect testing ground for cutting-edge media and communications technologies. One of the eye-popping features of the Music Center is the VR 360 stadium experience enjoyed by guests in the second-floor VIP room. Simply by donning wireless head mounted displays (HMD), spectators can get a 360-degree view of the live concert in 8K resolution. This is better than a ticket to the front row; this is a completely immersive experience, made possible by 5G technology and a built-in micro data center to support edge computing.



GIGABYTE G481-HA0, a 4U GPU Server with 10 slots for GPGPU cards, is used in the Taipei Music Center to recreate live concerts in virtual reality with 8K resolution. This is an example of an on-location micro data center that provides high bandwidth and low latency media streaming to fill in the gaps of edge computing.

This is how it works. Data from a multitude of cameras and recording devices in the concert hall are streamed to the GIGABYTE **G481-HA0**, a 4U GPU Server from the G-Series with slots for up to ten GPGPU cards. The server processes the data in real time to recreate the concert in virtual reality. By storing and routing the footage locally instead of through a faraway cloud computing center, latency caused by network backhaul is eliminated. Thus, the criteria of high bandwidth and low latency are fulfilled, so viewers can truly experience a feast for the senses.

While the G481-HA0 does not match the definition of an edge server in the strictest sense—for one thing, the 4U form factor will not fit in smaller spaces—it does set an example for the type of processing power and multifunctionality necessary in an ideal edge server. Based on 2nd Gen Intel® Xeon® Scalable Processors, the G481-HA0 supports up to 28 cores and 56 threads per socket, with support for Intel® Optane™ Persistent Memory and increased DDR4 memory speed and capacity. These features can be counted upon to match the growing needs of different edge computing scenarios. The G481-HA0 also features industry-leading GPU density, with slots for up to ten GPGPU cards such as NVIDIA® Tesla® V100, which can be used to accelerate processing speed until it reaches the level of high performance computing (HPC). This is especially pertinent in the case of the Taipei Music Center, which strives to astound visitors by pairing multimedia content with the latest communications technologies.

ITRI's Private Cell Mobile Command Vehicle: 5G Network on the Go

The mobile game Pokémon GO has been the pride and joy of Pokémon fanatics everywhere—as well as a source of constant headache for public safety officials. Every time there is a special in-game event, mobs of enthusiasts congregate at real-world locations, overwhelming the local cellular network. If the repercussion was just that a few digital critters escaped the grasp of eager players, there would be no cause to fret. Unfortunately, police officers monitoring the event and maintaining order rely on the same web connection. Network congestion leads to low image resolution and sporadic transmission problems, making it harder for the police to do their job.



Designed by ITRI and deployed by New Taipei Police, the “Private Cell Mobile Command Vehicle” provided internet connectivity analogous to a private 5G network for officers maintaining public order. At the heart of the solution lies GIGABYTE’s H281-PE0, a precursor to future Edge Servers compatible with 5G networks. (Image courtesy of ITRI)

In preparation for a large-scale Pokémon GO event, the New Taipei City Police Department worked with Taiwan's Industrial Technology Research Institute (ITRI) to deploy a "Private Cell Mobile Command Vehicle". In effect, this was a 5G private cellular network analogue, delivering high bandwidth, low latency, and ultra-reliable communications for police use. Officers were able to monitor high-resolution video footage in real time from a variety of sources, including surveillance cameras, body-worn cameras, and aerial drones. The command vehicle featured an AI-enhanced machine vision system that could identify crowded hot spots, so police resources could be allocated to areas where they were needed the most.

At the heart of the "Private Cell Mobile Command Vehicle" is the GIGABYTE **H281-PE0**, a 2U 4-node High Density Server from the H-Series. The H281-PE0 meets the hardware interface requirements of the Multi-Access Edge Computing (MEC) virtualization platform employed to enable low-latency transmissions, and it features multiple expansion slots to support adequate networking ports. Its four nodes are used to utilize Network Functions Virtualization (NFV) technology to deliver cellular core network services and to support edge computing or application services. Like the G481-HA0 used in the Taipei Music Center, the H281-PE0 offers unrivaled flexibility and multifunctionality that can support AI-based image recognition for video analytics software used in police work.

The H281-PE0 is also a big step in the advancement of 5G-capable edge servers in the sense that its 2U form factor takes up only half the space compared to a 4U form factor, without compromising on computing power or component density. Because everything is packed closely together, heat dissipation becomes very essential. The dual-fan setup was installed in the center of the server to generate strong airflow from the front to the rear of the chassis. Dynamic fan speed control was included to adjust fan speed according to the ambient temperature inside the van, which may shift as the vehicle changed locations or as personnel entered or exited the vehicle. The server's baseboard management controller (BMC) monitors key components such as the CPU heatsink and automatically adjusts fan speed to maintain temperature control.

Another critical lesson from the deployment of the "Private Cell Mobile Command Vehicle" is the effects of a harsh and rugged user environment. Even though the H281-PE0 was installed inside the van, it was still subject to external stress factors such as temperature variations, jostling and vibrations, and exposure to dusts. These elements could not be allowed to interfere with the system. GIGABYTE was able to overcome these challenges with mechanical engineering ingenuities, such as using rubber pads to reduce the vibration caused by fans, which could affect the read and write speed of hard disks. The H281-PE0 also passed stringent vibration, sweep frequency, and pressure testing before being shipped to the client, to ensure stable operations in the field.

The Birth of E-Series Edge Servers: E251-U70 Chosen by NVIDIA

In September 2020, GIGABYTE announced the E251-U70, the first product in a new line of E-Series Edge Servers created especially for edge applications.



Built especially for edge computing, the E251-U70 takes design cues from its predecessors, such as a scalable processor with a high number of cores, and multiple expansion slots for flexibility and multifunctionality. It also features brand new optimizations, such as decreased depth and a front-access server chassis.

The E251-U70 and other E-Series products represent the best of GIGABYTE's edge computing solutions, as demonstrated by the aforementioned success stories. A quick rundown of the highlighted features:

- A scalable CPU with a high number of cores and threads, to enable mMTC—supporting many devices simultaneously.
- A multitude of expansion card slots to provide flexibility and multifunctionality, so an edge server can handle computing, networking, and storage. Depending on the types of expansion cards used, the server may fulfill the requirements of eMBB (high bandwidth) and URLLC (low latency).
- A high-density design in a compact chassis, supported by GIGABYTE's innovative thermal solution to ensure stable operations.
- Good total cost of ownership (TCO) and power efficiency to provide clients with competitive CAPEX and OPEX.

What's more, the E251-U70 improves on the dense and compact design of a 2U form factor by further trimming the dimensions down to 439 x 86 x 449mm (WxHxD). In other words, the depth has been decreased from a standard 800 to 860mm to just 449mm, while maintaining the structural integrity and thermal control of a top-quality GIGABYTE server. A front-access server chassis has been incorporated for easy access and maintenance in confined spaces. A tool-less rail kit design is also included for convenient installation and maintenance.

As for software, the E251-U70 features GIGABYTE's patented SCMP (Smart Crises Management/ Protection) function, which automatically forces the CPU to enter ultra-low frequency mode (ULFM) to prevent system shutdown when the BMC is alerted to a critical fault or error, such as overheating or a power surge. The server will return to normal power mode once the issue has been resolved. The GIGABYTE Server Management (GSM), a proprietary multiple server remote management application, can also be used by administrators to remotely manage and update the edge server through their own computers or mobile devices, which is imperative considering how many edge servers are needed to complement a 5G network.

It was not long before GIGABYTE's new Edge Servers caught the eye of some of the biggest tech companies in the world. In October, NVIDIA announced the NVIDIA Aerial SDK, a software development kit for building and deploying software-defined, cloud-native, and GPU-accelerated 5G Virtual Radio Access Networks (vRAN) or other services to address rising consumer demand. The hardware platform they chose was the GIGABYTE E251-U70.

The E251-U70 was chosen for NVIDIA Aerial SDK due to its outstanding attributes, namely the compact spatial design, scalable CPU, high-redundancy features, low power consumption requirements, ease of maintenance, and multifunctionality thanks to its numerous accelerator slots.

GIGABYTE Expands 5G Business Opportunities with Industries Worldwide

As applications of 5G and IoT keeps growing, the market demand of edge servers is also on the rise. With long-standing expertise, GIGABYTE has developed edge servers able to operate under rigorous environments. This allows people around the world to experience a more convenient and smarter life brought by 5G technology.

Take E251-U71 for example, this server model is NEBS (Network Equipment-Building System) Level 3 design-ready, which indicates the highest industry standard of stability and durability. E251-U71 can operate under environments with high humidity, high altitudes, and various levels of temperature. It is suitable for lightweight edge data centers deployed outdoor or in remote areas to expand 5G-related services.

With the corporate vision "Upgrade Your Life", GIGABYTE aimed to bring a better life prospect to all mankind. GIGABYTE continues to drive innovation and cooperate with global partners to connect the world with 5G advancements and infinite opportunities.

GIGABYTE™

GIGABYTE TECHNOLOGY CO., LTD.

- All intellectual property rights, including without limitation to copyright and trademark of this work and its derivative works are the property of, or are licensed to, GIGA-BYTE TECHNOLOGY CO., LTD. Any unauthorized use is strictly prohibited.
- The entire materials provided herein are for reference only. GIGABYTE reserves the right to modify or revise the content at anytime without prior notice.
- All other brands, logos and names are property of their respective owners.