DIGITIMES Research: Special Report, 2022

Taiwanese vendor server showcase at Computex 2022

Frank Kung, DIGITIMES Research, June 2022

Introduction

Based on DIGITIMES Research's observation, processor suppliers Intel, AMD and Nvidia as well as server vendors QCT, Ingrasys, Gigabyte, Supermicro and Tyan had demonstrated their new offerings, with an emphasis on cloud, high-performance computing (HPC), AI and 5G edge applications, by hosting events themselves or participating in major international exhibitions including Computex 2022 and ISC 2022 during the first half of 2022.

To address the issues of increasing power consumption and heat generation that come with rising server performance, the vendors have introduced servers with liquid cooling systems to meet energy efficiency demand by cloud datacenters that keep scaling up their server infrastructure.

The server vendors that attended Computex 2022 could be divided into two groups. The first group included subsidiaries of large corporations in charge of server R&D and production or brand positioning, for example, Foxconn's Ingrasys and Quanta's QCT.

The other group included server brands/system integrators (SI) such as Gigabyte and Supermicro. Ingrasys made its debut at Computex in 2022 with an aim to reach more potential customers through its participation at the event as Foxconn's server business becomes a key growth engine for the corporation.

Gigabyte presented the most diverse range of servers featuring different CPU platforms including Whitley (codenamed Ice Lake), third-generation EPYC (codename Milan) and Ampere's next-generation Arm-based Altra Max, targeting HPC, AI and edge computing markets.

Summarizing the specs of new servers introduced by the vendors in first-half 2022, a majority of them feature Ice Lake or Milan processors, PCIe 4.0 and 3200MHz DDR4 DRAM. Going forward, the server vendors will keep exerting efforts toward the cloud computing market while making 5G edge computing servers a key strategic focus.

On top of that, they will emphasize modular design for their servers so that they can more swiftly build a diverse product portfolio to meet customers' small-quantity and large-variety needs. To address supply chain uncertainties that may become a constant pain, the vendors look to strengthen internal logistics control or parts procurement management.

New server platforms and events

The three leading processor suppliers, Intel, AMD and Nvidia, presented their launch schedules for CPU or AI accelerator solutions (GPU) at self-hosted events including Intel Vision 2022 (5/10-5/11), AMD Financial Analyst Day (6/9) and Nvidia GTC 2022 (3/21-3/24).

They also partnered with server vendors to showcase their new chips and server solutions at international events including ISC 2022 (5/29-6/2) and Computex 2022 (5/24-6/6).

At Intel Vision 2022, Intel announced the Aurora supercomputer project in partnership with HPE and Argonne National Labs. The Aurora supercomputer is equipped with Sapphire Rapids CPU and Ponte Vecchio GPU.

Aside from Sapphire Rapids CPUs, Intel also disclosed its next-generation Falcon Shores at ISC 2022, which puts CPU and Xe GPU in a single socket to reduce the latency between the CPU and the AI chip. Falcon Shores is expected to come on the market in 2024.

At ISC 2022, AMD presented the Frontier supercomputer, developed in collaboration with HPE for the US Department of Energy. It features the third-generation EPYC CPU, Instinct MI200 series accelerator and ROCm 5.0 software platform. Frontier is the No. 1 system in the Top 500 supercomputers capable of performing 1,018 floating point arithmetic operations per second (exascale machine).

Nvidia demonstrated its self-developed Arm-based Grace CPU, which, coupled with Nvidia's GPU and NVLink-C2C high-speed interconnect, powers HPC/AI servers targeting digital twin applications including cloud gaming, smart factory and autonomous driving.

At Computex 2022, Nvidia announced partnerships with Supermicro, Gigabyte, QCT, Foxconn and Wiwynn to bring servers featuring Grace CPU on the market in first-half 2023.

Table 1: CPU makers roadmaps and events for 1H22

CPU maker	1H22 events	Next-generation server CPU development	Status of servers equipping new CPUs
Intel	Hosting Intel Vision 2022 Participating at ISC 2022	Eagle Stream to volume production after 4Q22	Vendors mainly showcase servers equipped with Whitley CPU in 1H22
AMD	Hosting AMD Financial Analyst Day Participating at ISC 2022	4th generation EPYC CPU to volume production in 2H22	New servers equipped with 4th generation EPYC CPUs are still in evaluation. Vendors mainly showcase products with 3rd generation EPYC CPUs
Nvidia	Hosting Nvidia GTC 2022 Participating at Computex 2022 and ISC 2022	In-house developed Arm-based CPU Grace to release in 1H23	Vendors will release servers powered by the Grace CPU in 1H23

Source: DIGITIMES Research, June 2022

Server solutions showcased by vendors

Computex 2022 was a major event in first-half 2022 that most server vendors participated. Among them, QCT, Ingrasys, Gigabyte and Supermicro all showcased their products.

Foxconn had not attended Computex until Ingrasys, in charge of the corporation's server production, made a debut at the 2022 event to increase its visibility as part of the corporation's active efforts to penetrate the datacenter infrastructure market.

Tyan missed Computex 2022 but instead attended ISC 2022. ISC 2022 took place concurrently with Computex 2022. The company is likely to choose ISC over Computex to maximize its exposure to European and American buyers.

Table 2: Server vendor product showcases at events in 1H22

Server vendor	Solutions showcased
QCT	QCT is a subsidiary of Quanta marketing its own brand server solutions. It demonstrated new servers equipped with Ice Lake processors and targeting cloud, high-density and edge applications.
Ingrasys	Ingrasys is a subsidiary of Foxconn targeting the white-box cloud server market. It presented HPC and 5G edge servers as well as software-hardware integrated solutions.
Gigabyte	Gigabyte showcased a diversity of servers equipped with different CPU platforms for cloud, HPC/AI, high-density and edge computing applications.
Supermicro	Supermicro demonstrated HPC, multi-node, high-capacity storage and telecom servers.
Tyan	Tyan is a subsidiary of Mitac targeting integrated brand server solutions. It exhibited new servers powered by Intel and AMD processors for cloud HPC applications.

Source: DIGITIMES Research, June 2022

Product showcase highlights

QCT

At Computex 2022, QCT mainly presented servers equipped with Intel Whitley platforms (codenamed Ice Lake) for datacenters, high-density and 5G telecom network applications.

The QuantaGrid series targets cloud datacenter applications. The D53XQ-2U and D53X-1U models are rackmount servers featuring two Ice Lake processors (dual socket) and PCIe 4.0 support.

The D53XQ-2U comes with a higher storage capacity and supports two dual-width AI accelerators. The D53X-1U supports up to three single-width AI accelerators.

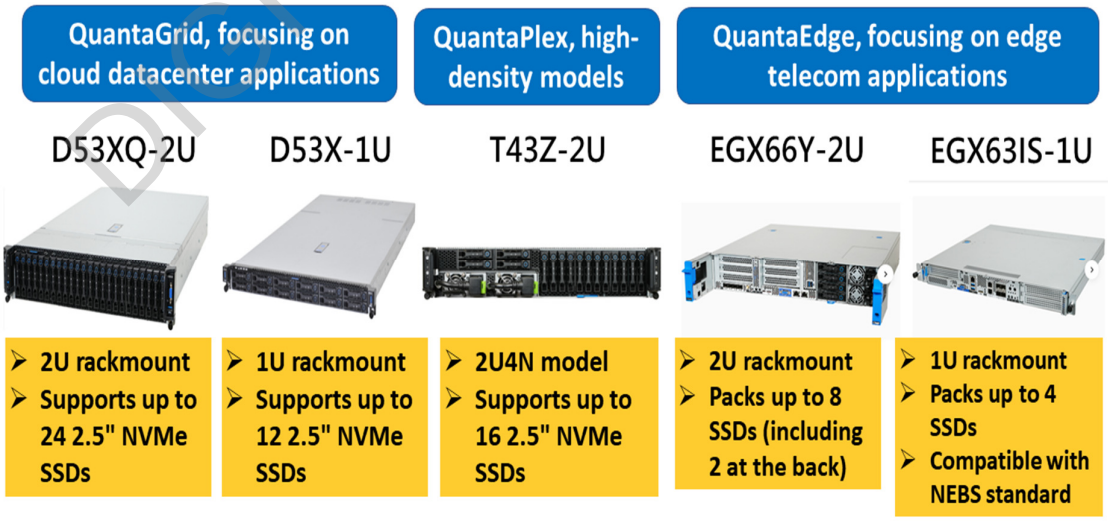
The QuantaPlex is a 2U4N high-density server, targeting HPC applications. Each node can have two Ice Lake processors so the system can have up to eight. It also supports a maximum of 16 3200MHz DDR4 DRAM modules.

For 5G edge telecom applications, QCT introduced the QuantaEdge series, featuring a short depth (< 500mm). The EGX66Y-2U model is designed for RAN edge central units (CU), distributed units (DU) and mobile edge computing (MEC) applications while the EGX63IS-1U model is designed with Telco NEBS Level 3 compliance.

The EGX63IS-1U comes with one Ice Lake processor, eight 3200MHz DDR4 DRAM modules and four PCIe 4.0 slots for external storage devices or AI accelerators.

The EGX66Y-2U features a dual-socket design and supports up to 16 3200MHz DDR4 DRAM modules and six PCIe 4.0 slots.

Table 3: QCT servers showcased at Computex 2022



Source: QCT; compiled by DIGITIMES Research, June 2022

Ingrasys

Ingrasys, fully owned by FII under Foxconn, joined online exhibitions at Computex for the first time this year. It demonstrated solutions for cloud datacenters as well as various applications.

Its cloud HPC servers are mainly based on AMD third-generation EPYC CPU (codenamed Milan), including the 2U4N dual-socket server SV2080A, 2U dual-socket rackmount server SV2020A and single-socket tower server WS5010A.

The high-density SV2080A supports two AMD CPUs and 16 3200MHz DDR4 DRAM modules per node.

The 2U rackmount server SV2020 comes with two Milan CPUs and an additional storage capacity of up to 26 2.5-inch NVMe solid-state drives (SSD) to meet the needs of HPC/AI and software-defined storage applications at large cloud datacenters.

The 5U tower server WS5010A is built with one AMD CPU and 16 3200MHz DDR4 DRAM modules to target cloud media streaming, 8K video editing and AI accelerator applications.

Featuring a barebone 1U form factor, the SV1020i packs in two Intel Ice Lake processors and up to 32 3200MHz DDR4 DRAM modules and supports Intel Optane Persistent Memory.

The edge servers Ingrasys presented focused on CU and DU telecom solutions, including the CU-CI5520 and DU-DO4020, which are based on Intel Xeon Scalable processors. Designed for operation in the edge environment, the servers feature a wide operating temperature range (-40°C to 65°C), ideal for 5G smart city applications, rural use scenarios and autonomous driving.

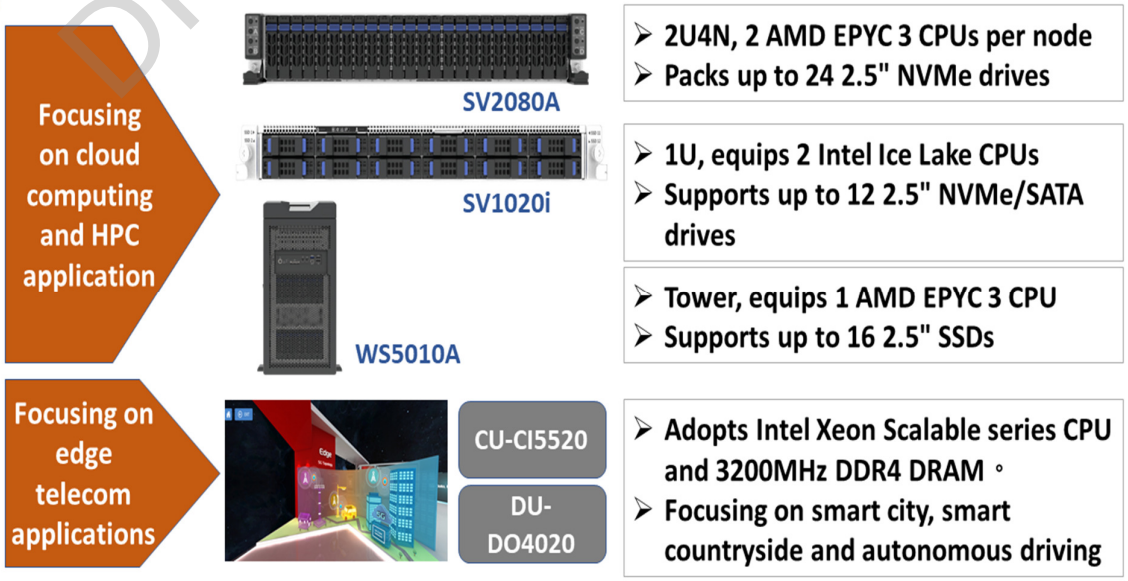
The CU- CI5520 is equipped with one Intel Xeon Scalable processor and 16 3200MHz DDR4 DRAM modules.

The DU- DO4020 comes with one Intel Xeon Scalable processor and eight 3200MHz DDR4 DRAM modules.

Targeting MEC applications, the MEC-MI4010 with an Intel Xeon D CPU and four 2933MHz DDR4 DRAM modules are designed with a lower spec compared to the CU and DU servers.

Ingrasys also exhibited liquid cooling solutions to address the issue of rising power consumption by next-generation datacenters. The solutions achieve a thermal dissipation capacity up to 450W (as opposed to 350W by conventional cooling systems) and reduce datacenter power usage effectiveness (PUE) to 1.03 (as opposed to 1.4-1.6 by conventional cooling systems).

Table 4: Ingrasys servers showcased at Computex 2022



Source: Ingrasys; compiled by DIGITIMES Research, June 2022

Gigabyte

At Computex 2022, Gigabyte showcased a slew of edge and cloud servers including its R and G series for cloud and AI/HPC applications, H series featuring high density and E series for 5G and edge computing applications.

Among the server vendors, Gigabyte presented the most diverse range of servers featuring different CPU platforms based on Intel x86, AMD and Ampere Arm architectures.

The R282-Z93 in the R series comes with a 2U form factor, two AMD third-generation EPYC processors and up to 32 3200MHz DDR4 DRAM modules. Certified as an Nvidia NGC Ready server, the R282-Z93 is ideal for cloud AI, video processing or storage-intensive applications.

The G series comprises the G262 and G492 families.

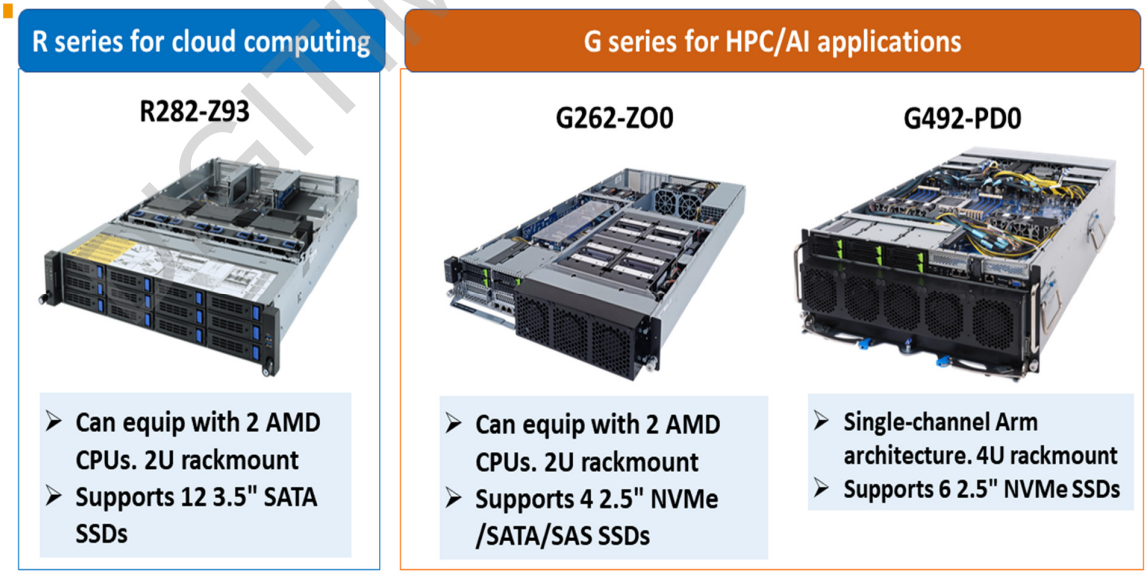
The G262-ZO0 is built to be a total AMD-based solution that comes with two AMD third-generation EPYC CPUs, four AMD Instinct MI250 GPUs and AMD Infinity Fabric interconnect technology to support the use scenarios of high-performance AI computing.

The G492-PD0 is a single-socket system based on Ampere Altra Max or Altra Arm CPU with up to 128 cores. It also supports Nvidia A100 GPU and NVlink interconnect protocol.

The G492-ID0 is an Intel Ice Lake-based server with support for Intel Optane Persistent Memory 200 series as well as Nvidia A100 GPU and NVlink.

Gigabyte also showcased its G492-ZL2, featuring a liquid cooling solution, two AMD third-generation EPYC processors as well as Nvidia A100 GPU and NVlink support.

Table 5: Gigabyte servers showcased at Computex 2022 (part 1)



Source: Gigabyte; compiled by DIGITIMES Research, June 2022

The H series products Gigabyte demonstrated were of the 2U4N high-density family H262, including the H262-ZL0 and H262-Z6B.

The H262-ZL0 comes with two new-generation Milan CPUs and 16 3200MHz DDR4 DRAM modules per node. In particular, it features a liquid cooling system with cold and hot tubes deployed over each node to enhance the thermal dissipation efficacy.

The H262-Z6B comes with two AMD Rome CPUs and also 16 3200MHz DDR4 DRAM modules per node to satisfy enterprise users' needs for hybrid cloud, private cloud or hyper-converged infrastructure (HCI) applications.

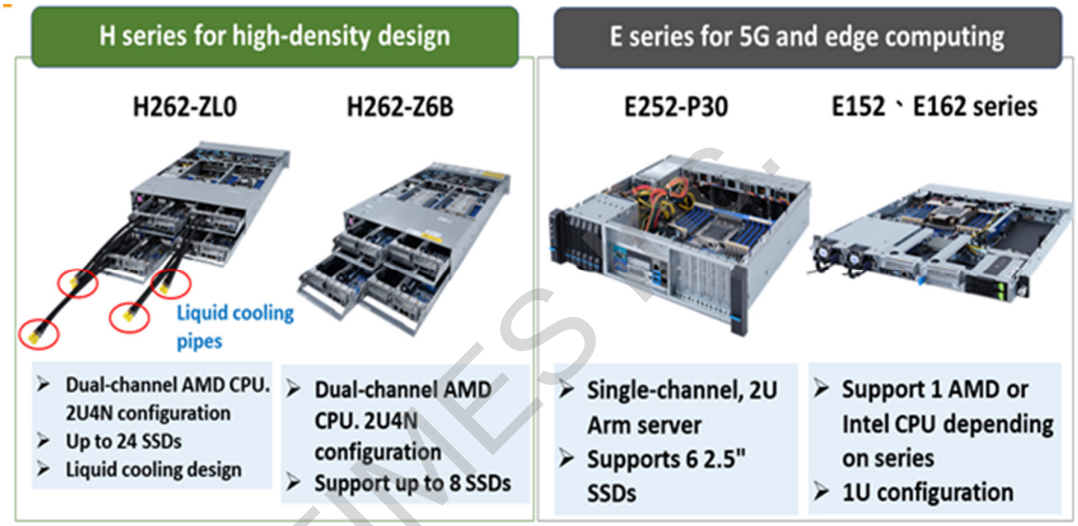
Designed for 5G and edge computing applications, the E series includes the E252 family with Arm-based processors as well as the E152 and E162 families with x86-based AMD and Intel

processors, all featuring a short depth (400-500mm) that is half of that of conventional servers.

The E252-P30 is equipped with one Arm-based Ampere Altra Max or Altra CPU and 16 3200MHz DDR4 DRAM modules for edge computing applications.

Both designed with a short depth, the E152-ZE0 and E162-220 respectively come with one AMD Milan CPU and Intel Ice Lake CPU, as well as eight and 16 3200MHz DDR4 DRAM modules. Furthermore, they support one double-width GPU for the acceleration of edge AI inference.

Table 6: Gigabyte servers showcased at Computex 2022 (part 2)



Source: Gigabyte; compiled by DIGITIMES Research, June 2022

Supermicro

At Computex 2022, Supermicro exhibited cloud HPC, high-density, high storage capacity and edge 5G servers to address a variety of application scenarios.

The servers for large cloud datacenters and HPC applications Supermicro put on exhibit included the A+ and SYS series.

The A+ series comes with AMD third-generation EPYC processors and a choice between AMD Instinct MI250 GPU and Nvidia A100 GPU as the accelerator. The SYS series is Whitley based.

Both the A+ and SYS series support up to 8TB of 3200MHz DDR4 DRAM. The SYS series further supports Intel Optane Persistent Memory.

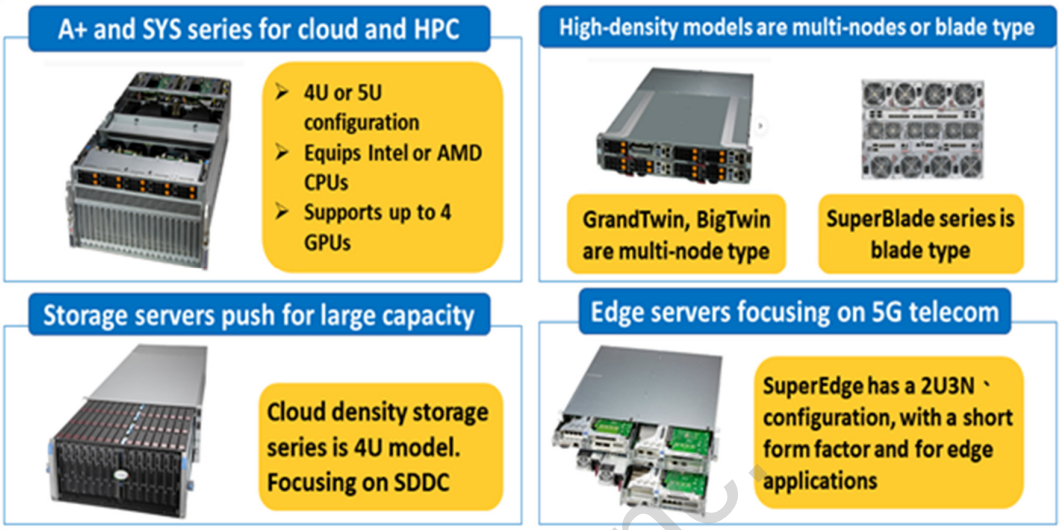
Supermicro's AMD-based high-density offerings include the 2U4N GrandTwin and BigTwin, as well as the 8U20N SuperBlade, built for large datacenter operators' content delivery network (CDN), cloud gaming and HPC, needs as well as semiconductor vendors' EDA applications.

Targeting high storage capacity and software-defined datacenter (SDDC) needs, Supermicro's Cloud-Density Storage servers in a 4U form factor accommodate up to 90 hot-swap drives and feature a tool-less design for easy service and maintenance at the customer's site.

The SuperEdge series is designed with a 2U3N form factor and a short depth (430mm). Each node comes with one Intel Ice Lake processor and eight 3200MHz DDR4 DRAM modules, ideal as Open RAN telecom servers and enterprise edge servers.

Supermicro further showcased the barebone 2U short-depth SYS-210P with one Ice Lake CPU, eight 3200MHz DDR4 DRAM modules and Telco NEBS Level 3 compliance.

Table 7: Supermicro servers showcased at Computex 2022



Source: Supermicro; compiled by DIGITIMES Research, June 2022

Tyan

Tyan is a member of the Mitac group. It is the server brand under Mitac Computing Technology. It showcased servers equipped with Intel Ice Lake and AMD third-generation EPYC CPUs at ISC 2022.

Tyan's Thunder HX series features Ice Lake processors and comes in 2U and 4U form factors, both with PCIe 4.0 support.

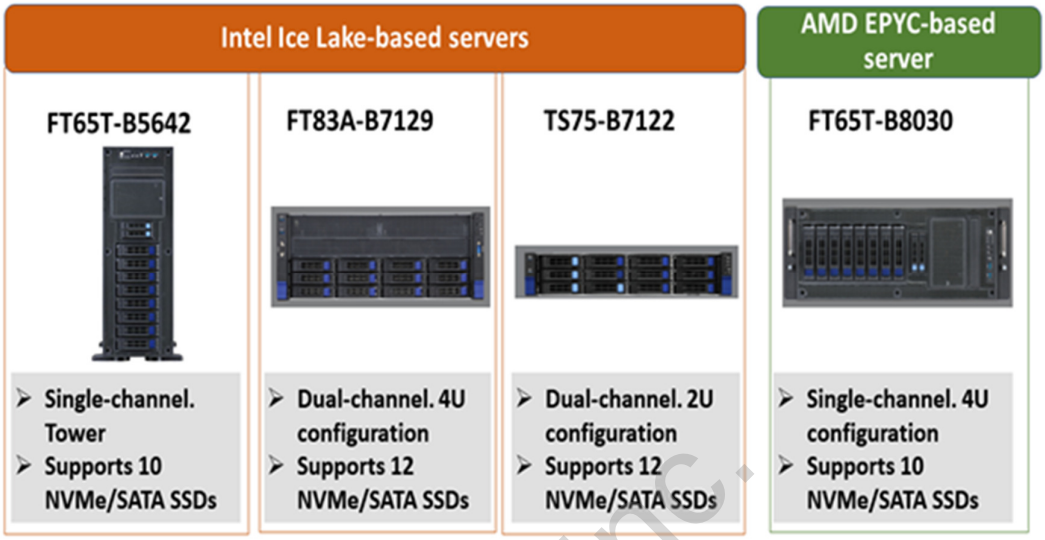
The FT65T-B5642 is a single-socket tower server with eight 3200MHz DDR4 DRAM modules, targeting enterprise datacenters.

The FT83A-B7129 is a dual-socket 4U server designed for use scenarios that require higher computing performance. It supports two CPUs, 32 3200MHz DDR4 DRAM modules and up to ten GPU accelerators.

The TS75-B7122 with two Ice Lake processors, a 2U form factor, 32 3200MHz DDR4 DRAM modules and five PCIe 4.0 slots for two double-width GPU accelerators or one OCP 3.0 LAN mezzanine card is ideal for high-density cloud HPC and virtualization applications.

Also on exhibit was Tyan's AMD-based FT65T-B8030 with one third-generation EPYC processor, eight 3200MHz DDR4 DRAM modules and five PCIe 4.0 slots.

Table 8: Tyan servers showcased at Computex 2022



Source: Tyan; compiled by DIGITIMES Research, June 2022

Server vendor development in 1H22

DIGITIMES Research observed that the servers presented by the vendors in first-half 2022 mainly come with the Intel Whitley platforms or AMD third-generation EPYC CPUs in addition to PCIe 4.0 support.

Intel Eagle Stream and AMD fourth-generation EPYC processors are expected to begin volume production in third-quarter or fourth-quarter 2022 and to be adopted by server vendors starting first-half 2023.

A majority of the new servers developed by the vendors target HPC or AI model training uses by cloud datacenters as well as edge 5G telecom or AI inference applications. They feature NEBS compliance or a short-depth design.

Additional key development trends include modular designs, next-generation liquid cooling solutions and software-hardware integrated total cloud server solutions.

Ingrasys and Supermicro offered OCP reference designs for cloud datacenter operators or developed commonly used modules for key elements of the motherboard such as BMC, network, storage and AI accelerators so that they can more swiftly put together a diverse server product portfolio.

Another advantage of modular design is that it allows server vendors to more conveniently control supply chain operations or parts procurement and accelerate the production of custom-made cloud servers.

Thermal dissipation for datacenters is also a key focus for the vendors. Ingrasys and Gigabyte both exhibited such solutions at Computex 2022, including servers with liquid cooling solutions or liquid cooling systems for datacenter racks.

Cloud and HPC/AI applications remain the main growth drivers for the server market. In view of this, aside from next-generation server hardware development, the server vendors are also exerting efforts toward corresponding software system solutions.

At Computex 2022, Ingrasys showcased its HCI solution featuring its self-developed software Composable Disaggregated Infrastructure (CDI), which enables customers to make more flexible use of server computing and storage resources to integrate virtualization and software-defined anything (SDx) implementations.

Table 9: Server vendor development in 1H22

Server vendor development
Most server vendors presented servers with PCIe 4.0 support in 1H22.
Their exhibitions highlighted cloud, HPC/AI and edge 5G applications.
They showcased servers featuring modular design for cloud datacenters.
They also emphasized liquid cooling solutions to address requirements by large datacenters.
They made effort toward server software system R&D, delivering software-hardware integrated solutions.

Source: DIGITIMES Research, June 2022