CORESTATION
AMULET HOTKEY

CoreStation for AI

AI-Ready Remote Workstation Solutions

Solution Brief

Delivering high performance scalable and flexible AI Ready infrastructure designed and built for delivering GPU accelerated workloads for AI-centric workloads

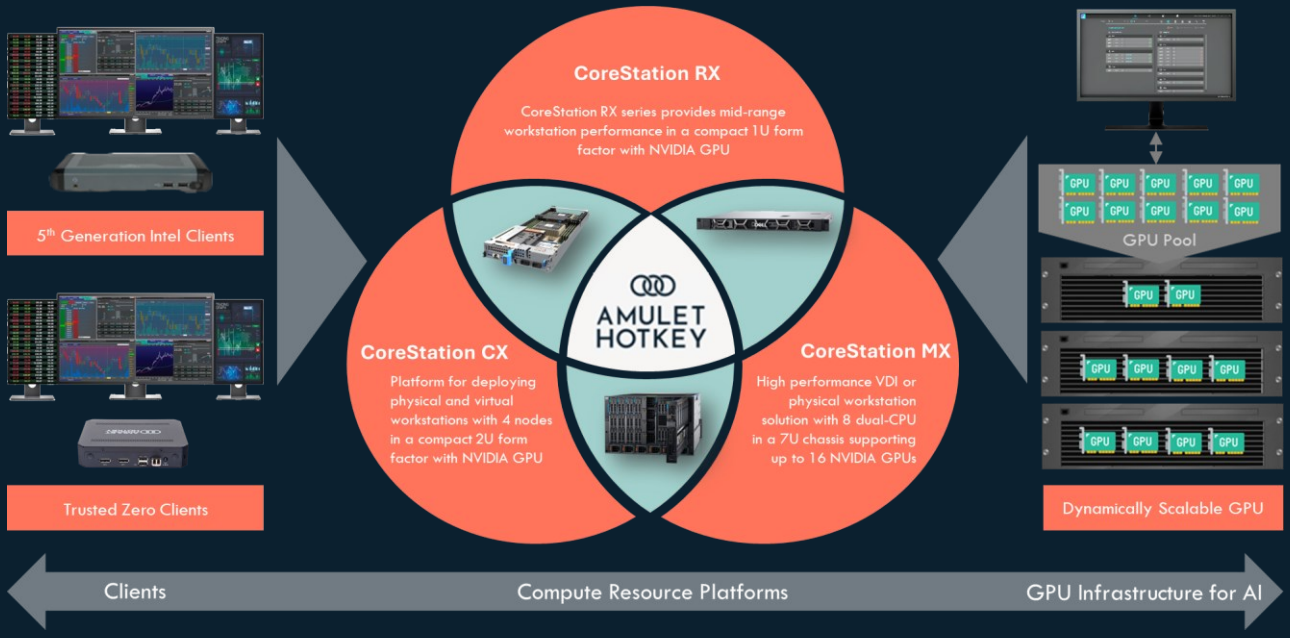
CoreStation for AI: CX Platform

Executive Summary

CoreStation for AI from Amulet Hotkey combines powerful, dedicated remote workstations with class-leading AI-ready capabilities, giving organizations the power to enable employees to work from anywhere, with no compromise on performance, regardless of their role.

By integrating high-performance GPUs into the platform enables customers to utilize workstation-class workloads on a remote platform without sacrificing performance. With this level of scalability workloads such as inferencing, training models, rendering, machine learning can now be cost-effectively deployed.

Unique to Amulet Hotkey is the ability to include the client-side hardware devices to deliver a complete end-to-end remote desktop solution that can fit all sizes regardless of the workload from basic single GPU solutions scaling all the way up to solutions that support multiple GPUs.



At the heart of the Amulet Hotkey CoreStation for AI solution is the CoreStation CX6620. Available as a 2U modular rackmount chassis, each CoreStation CX6620 chassis can house up to four compute nodes (0.5U each) making it the ideal density optimized platform, essential for delivering hardware accelerated graphics.

With the integration of the Liqid SmartStack solution, you can now scale from a single GPU per workstation to an externally configured PCIe chassis that can be configured with up to 30 GPU cards, either directly connected or part of a pool of GPU resource that can be dynamically allocated to workstations on demand.

CoreStation is based on enterprise-class server hardware, certified to run desktop operating systems such as Windows and Linux, as well as being able to support running a hypervisor enabling it to run multiple virtual desktop instances. The flexibility of the CX6620 means you can deliver both physical and virtual all from the same chassis.

Common across the entire platform is the ability to manage all CoreStation nodes remotely, regardless of form factor, from a single management console that not only makes management simple but also allows service provider partners to deploy Amulet Hotkey solutions to deliver Desktop-as-a-Service solutions.

Solution Overview

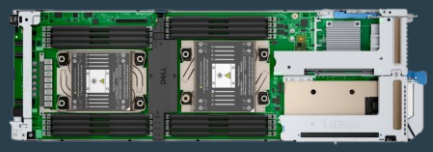
Compute Platform: Amulet Hotkey CoreStation for AI CX

This solutions brief is focused on delivering solutions for large-scale GPU acceleration using the Amulet Hotkey CoreStation for AI CX platform and the CDI ecosystem designed for demanding AI environments.

CoreStation for AI: CX Platform

CoreStation for AI CX platform provides a common platform for deploying both physical and virtual workloads.

Utilizing a 2U enclosure, each node being 0.5U, the CoreStation CX holds up to four compute nodes each with dual processing capabilities, optional GPU cards, and fully redundant power supplies and agent-free management.



It supports native Windows 10, Windows 11, or Linux-based desktop operating systems when deployed as dedicated workstations as well as leading hypervisor platforms for hosting virtual desktop infrastructure that in turn run the AI workloads.

Tech Spec Headlines

Component	Configuration Options
CPU	Up to two 4th or 5th Generation Intel® Xeon® processors with up to 56 cores per CPU
Memory	Up to 5200 MT/s RDIMM memory with 16 x DDR5 DIMM slots supporting up to 4TB
Storage	Up to 16 x 2.5-inch SAS/SATA/NVMe (HDD/SSD) drives max 61 TB
Graphics (GPU)	Support internally for 1 x NVIDIA L4 GPU 24 GB GDDR6 memory, x16 PCIe Gen4
Networking	Intel or Broadcom OCP 3.0 network adapter supporting up to Quad Port 10/25GbE
Expansion Options	Liqid PCIe x16 Gen4 HHHL HBA (up to three EX-4410 PCIe expansion enclosures)

This solutions brief will take advantage of the CoreStation CX6620 flexibility in providing blueprint high-level baseline configurations for different use cases that enable high-end scalable graphically intensive workloads which in turn deliver excellent performance for the applications used in the MEG, Research & Scientific, and Life Sciences industries, as well as 3D modelling applications.

Solution Architecture

Infrastructure platform for delivering GPUs at Scale

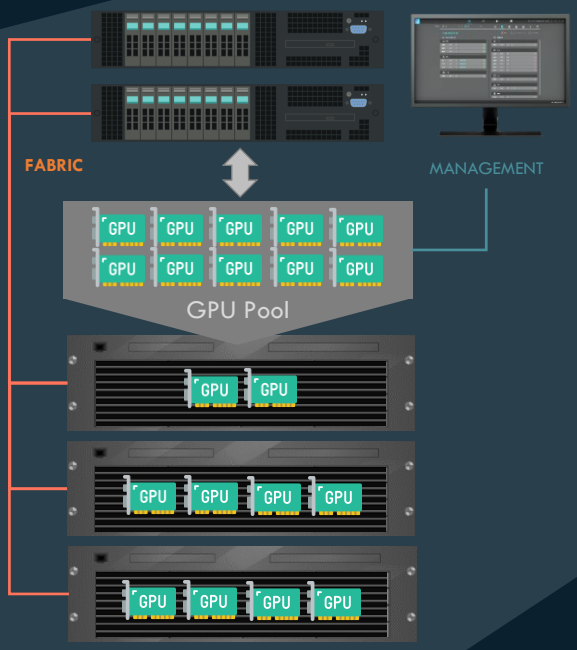
With the CoreStation for AI CX platform, organizations can deliver GPU resources (externally) to density optimized servers that have limited expansion options, yet still deliver high-end CPU and memory capacity.

This ability enables unrivalled GPU density when it comes to running AI-centric workloads, but not only that, it brings with it the flexibility to efficiently manage those GPU resources dynamically across multiple machines.

CoreStation for AI: CX Platform

To deliver this level of GPU scalability and flexibility, CoreStation for AI with the CX workstation platform employs an external enclosure each capable of hosting up to ten GPU cards. This connects to the host machines using a host bus adapter (HBA) that creates a PCIe fabric.

With this unique fabric approach, building a network of GPU resources, you create a Composable Infrastructure. This means that rather than have static GPU resources "locked" to a single machine, GPU resources can now be dynamically allocated to whichever machine requires it. For example, if you have a pool of 10 GPU cards a machine could be allocated a single GPU to start with but as demand increases then all ten could be allocated, all managed simply with the click of a mouse.



Tech Spec Headlines

CoreStation for AI Platform Model	CoreStation for AI CX6620/10	CoreStation for AI CX6620/20	CoreStation for AI CX6620/30
Maximum GPUs	10	20	30
Maximum Workstations Supported	4	8	16
GPUs Per Workstation	Up to 10	Up to 20	Up to 30
Composability Software	Liqid Matrix Software	Liqid Matrix Software	Liqid Matrix Software
PCIe Expansion Chassis	1 x EX-4410 PCIe Chassis	2 x EX-4410 PCIe Chassis	3 x EX-4410 PCIe Chassis
PCIe Fabric Switch	Not Required	1x 48 Port PCIe Switch	1x 48 Port PCIe Switch
PCIe Host Bus Adapters	Up to 4 (1 per host)	Up to 8 (1 per host)	Up to 16 (1 per host)
Rack Units	5U	10U	15U

This solutions brief will take advantage of the CoreStation for AI CX platforms flexibility to provide blueprint high-level baseline configurations coupled with examples of how to increase and dynamically manage and allocate GPU resources using composable infrastructure for delivering AI-centric workloads.

Solution Architecture

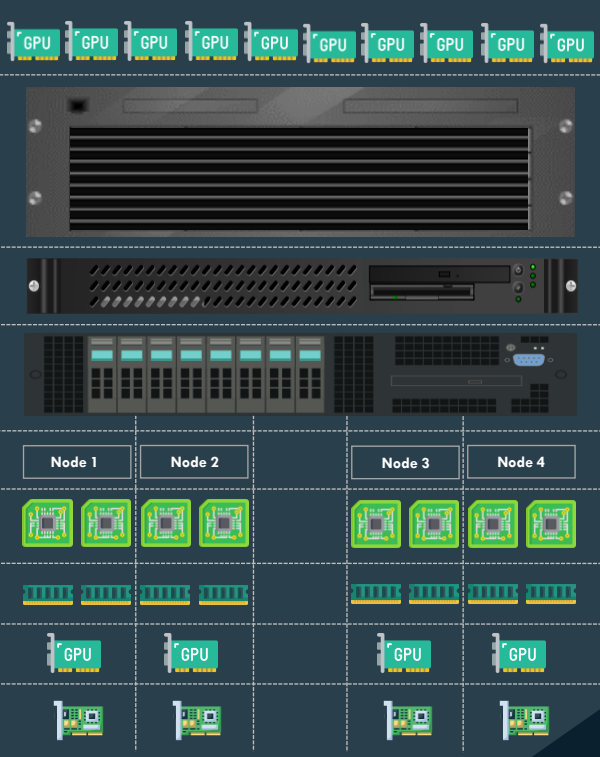
Blueprint for four remote workstations with up to ten GPU cards

This architecture blueprint provides a reference point for delivering up to ten GPU cards, in this example the GPUs in question are NVIDIA L40S cards, however, the external PCIe enclosures support other models of GPU that can also be mixed within the same enclosure. They also support other PCIe cards too.

In terms of workstation nodes, this blueprint is designed to support up to four workstations (a fully-populated CoreStation for AI CX chassis) enabling GPU cards to be dynamically allocated between the four workstations.

CoreStation for AI: CX6620-10

The CoreStation for AI CX6620-10 solution consists of a single EX-4410 enclosure (4U), supporting a maximum of ten GPU cards that can be dynamically allocated to any of the CX6620 workstation nodes (up to four workstation nodes supported in 2U), all managed using the Director management appliance (1U). Each individual workstation node can also be configured with its own local GPU card.



NVIDIA L40S 48GB GPUs (up to 10 supported)
EX-4410 PCIe Expansion Enclosure Up to 10 x PCIe cards supported
Management Appliance: Director (1U)
CoreStation for AI CX Chassis (2U)
Compute Nodes – 4 per CX6620 chassis each
2 x Xeon 4510 CPU, 12C/2.4GHz per node
64/128GB 5200MT/s RAM per node
1 x NVIDIA RTX A4000 ADA
1 x Liqid PCIe x16 Gen4 HHHL HBA

Dynamic Flexibility and Scalability

With the flexibility and density of the CoreStation for AI CX workstations, coupled with the composable infrastructure capabilities of the GPU cards, organizations can not only add more GPU resources to workstations dependent on workload, but they can dynamically allocate those resources on demand.

Solution Architecture

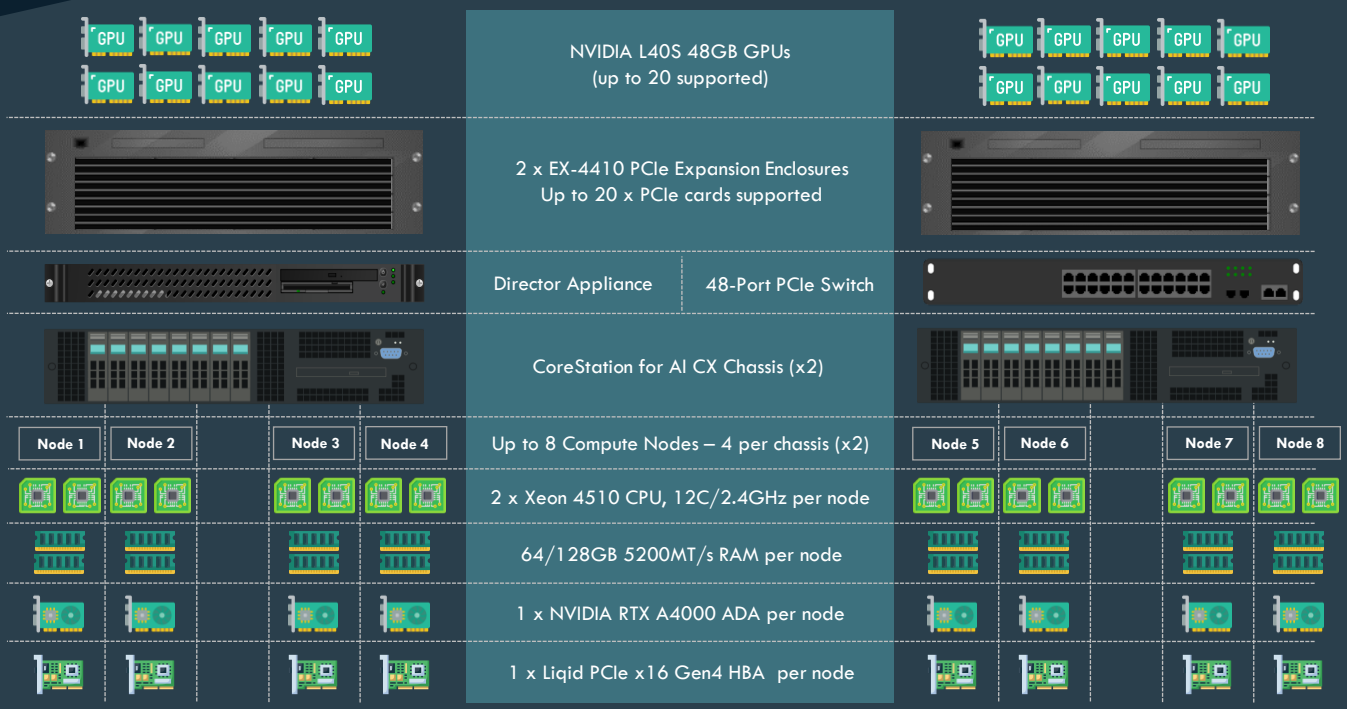
Blueprint for eight remote workstations with up to twenty GPU cards

This architecture blueprint provides a reference point for delivering up to twenty GPU cards, in this example the GPUs in question are NVIDIA L40S cards, however, the external PCIe enclosures support other models of GPU that can also be mixed within the same enclosure. They also support other PCIe cards too.

In terms of workstation nodes, this blueprint supports up to eight workstations (two fully-populated CoreStation for AI CX chassis) enabling GPU cards to be dynamically allocated between all workstations.

CoreStation for AI: CX6620-20

The CoreStation for AI CX6620-20 solution consists of a two EX-4410 enclosures (8U), supporting a maximum of twenty GPU cards that can be dynamically allocated to any of the CX workstation nodes (up to eight workstation nodes supported in 4U), all managed using the Director management appliance (1U), and all connected via a PCIe fabric switch (1U).



Dynamic Flexibility and Scalability

Key to the CoreStation for AI CX platform solution is its scalability and density. This blueprint demonstrates a solution consisting of up to twenty GPU cards that can be shared with up to eight remote workstation nodes. Including the locally configured GPU cards, in total this solution can deliver up to 28 GPUs in just 14U of rack space. With the added flexibility of composable infrastructure, external GPU cards can be dynamically allocated on demand to the workstation nodes as workloads dictate.

Solution Architecture

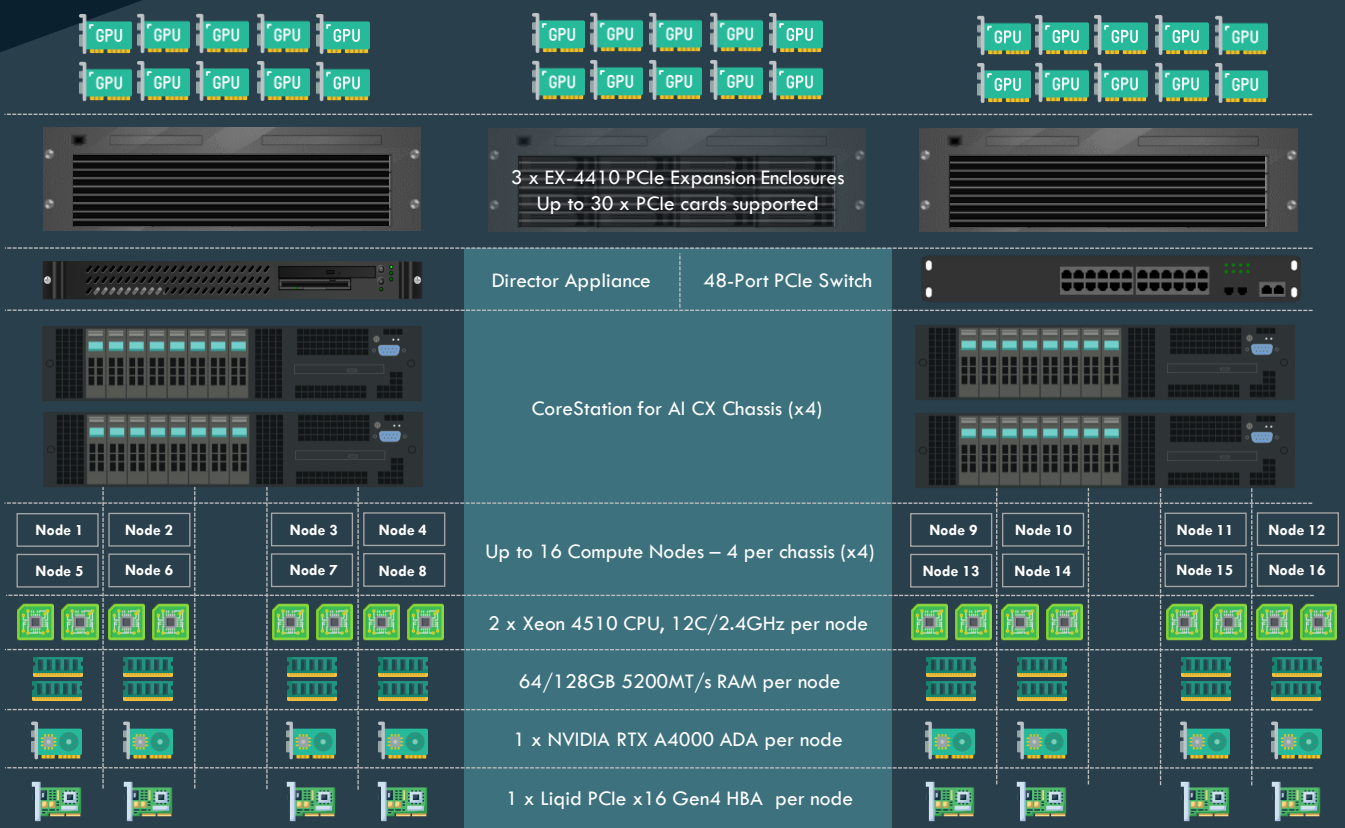
Blueprint for sixteen remote workstations with up to thirty GPU cards

This architecture blueprint provides a reference point for delivering up to thirty GPU cards, in this example the GPUs in question are NVIDIA L40S cards, however, the external PCIe enclosures support other models of GPU that can also be mixed within the same enclosure. They also support other PCIe cards too.

In terms of workstation nodes, this blueprint supports up to sixteen workstations (four fully-populated CoreStation for AI CX chassis) enabling GPU cards to be dynamically allocated between all workstations.

CoreStation for AI: CX6620-30

The CoreStation for AI CX6620-30 solution consists of three EX-4410 enclosures (12U), supporting a maximum of thirty GPU cards that can be dynamically allocated to any of the CX workstation nodes (up to sixteen workstation nodes supported in 8U), all managed using the Director management appliance (1U), and all connected via a PCIe fabric switch (1U).



Dynamic Flexibility and Scalability

This blueprint demonstrates a solution capable of hosting up to thirty GPU cards that can be shared with up to sixteen remote workstation nodes. Including locally configured GPU cards, this solution can deliver up to 38 GPUs. With the added flexibility of composable infrastructure, external GPU cards can be dynamically allocated on demand to the workstation nodes as workloads dictate.

Connecting Remotely from the Edge

Client Solutions for Remote Connectivity

The final piece of the solution, now that the remote workstation platform is in place, and just as important to get right, is the device from which the end user is going to connect to their virtual desktop machine. If the device is underpowered then the end user could experience poor performance regardless of the hosting infrastructure. Therefore, it is critical to use the correct device to match the end users use case.

The Amulet Hotkey 5th generation DX range of thin clients provide not only a range of hardware specifications to meet the end user requirements but also are agnostic when it comes to what operating system they run. This means you can also match the most suitable operating system to the use case.

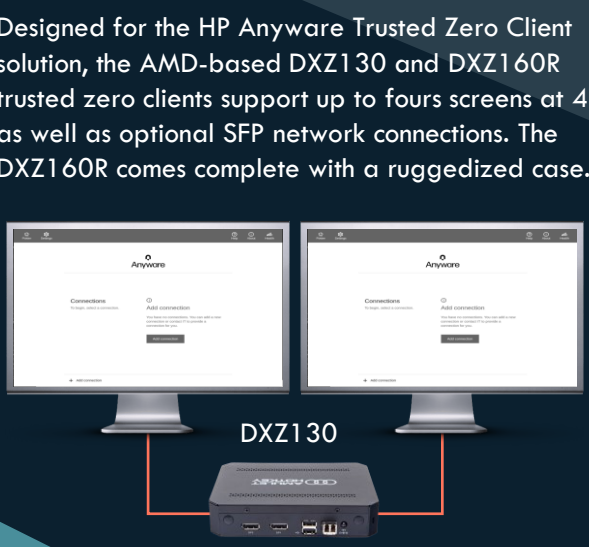
High Performance Secure Thin Clients

The DX1500 and DX1700 series of thin clients are designed up to support four screens each running at 4k (3840 x 2160) and are available with up to 16GB RAM and Intel i3, i5, or Intel U300 CPUs.

Trusted Zero Clients

Anyware

Designed for the HP Anyware Trusted Zero Client solution, the AMD-based DXZ130 and DXZ160R trusted zero clients support up to fours screens at 4k as well as optional SFP network connections. The DXZ160R comes complete with a ruggedized case.



In Summary

This solutions brief has provided a high-level blueprint that serves as an example of how to approach the challenge that AI-centric workloads present to organizations when multiple GPU resources are required to fulfil these high workload demands, solving the inflexibility and non-scalability that organizations face today when adopting AI models.

In this example blueprint, the compute requirements are delivered using the Amulet Hotkey CoreStation for AI CX platform solution as the compute resource, optimized for density and scalability, with the ability to deliver four high-spec workstation nodes in a single 2U chassis form factor, including an internal GPU card.

When it comes to delivering multiple GPU resources for AI-centric workloads, CoreStation for AI allows static barriers to be removed by abstracting GPU resources from individual workstations. It creates a datacenter-based centralized pool of GPU resource that can be dynamically allocated to individual workstations as workloads and demands dictate.

Find out more

To find out more about the Amulet Hotkey CoreStation for AI CX solution, Amulet Hotkey 5th Gen DX thin clients, or Amulet Hotkey Trusted Zero Clients for delivering a complete end-to-end solution for delivering remote workstation capabilities designed to deliver AI-centric workloads, click the icons below:

CoreStation for AI CX

5th Gen DX Thin Client

Trusted Zero Client

Composable Infrastructure

Contact Us

EMEA Sales
+44 (0) 20 7960 2400
emeasales@amulethotkey.com

N America Sales
+1 212 269 9300
ussales@amulethotkey.com

APJ Sales
+61 409 930 884
apsales@amulethotkey.com