Infortrend®

Infortrend Enterprise Cloud

Explore IEC

Highlights

High Performance

- One KS provides up to 1M IOPS, 30 GB/s read, and 25 GB/s write throughput
- Selected KS models feature dual Intel Xeon Scalable 64-core/ AMD EPYC™ 9004 Series CPUs and up to 4 Nvidia GPUs

Comprehensive Data Protection

- Multi-node data replication ensures no data loss
- Remote backup of VMs and application-related data to cloud or storage

User-Friendly Management

- EonOne for managing compute, GPU, and storage resources
- EonKube for managing VMs and all deployed applications

Scalability

- Expand KS cluster up to 20 nodes
- Expand storage with EonStor GS and DS solutions

Flexible Hardware Options

- KS 5000U and KS 3000U series offer six models with configurable GPU options for various application needs

Support for Different Application Types

- Run cloud-native applications, traditional applications, and virtual machines

Introduction

Infortrend Enterprise Cloud (IEC) is a private cloud platform that integrates compute, storage, and Kubernetes management into a single solution. It offers cloud capabilities such as scalability and reliability, enabling enterprises to run both cloud-native and legacy applications. IEC supports high-performance CPUs, GPUs, and NVMe SSD drives, making it suitable for AI and other high-performance applications. IEC compute nodes are available in two product series: KS 5000U, designed for enterprise data centers, and KS 3000U, designed for edge and SMB environments. KS 5000U deployments are recommended to start with three nodes, while KS 3000U configurations can begin with one or two nodes, depending on requirements. The platform supports node protection and allows for scale-out to enhance performance. Additionally, IEC provides high-performance storage to meet demanding data storage needs.

High Performance

The first IEC compute node series, KS 5000U, delivers outstanding storage performance, achieving 1M IOPS, 30 GB/s read, and 25 GB/s write throughput with high-speed U.2 NVMe SSDs. Offering flexible and powerful computing options, KS 5000U includes two processor configurations: KSi models with dual Intel Xeon® Scalable processors, capable of scaling up to 2560-core computing power in a cluster, and KSa models with dual AMD EPYC™ 9004 Series processors. Furthermore, the system supports GPU computing, making it ideal for AI, deep learning, HPC, and complex data analysis that demand extensive parallel processing capabilities.

The second series of IEC, KS 3000U, features AMD EPYC™ 8004 Series processors (up to 64 cores) and supports up to two GPU cards. It is designed for low-latency, high-performance workloads at the point of data capture, such as AI inference and edge-based processing.

Comprehensive Data Protection

KS supports multi-node data replication (Replica x2 and x3), ensuring no data loss in the event of node or hard disk failure.

To protect applications, KS can regularly back up data to remote locations, offering flexible backup options, including backup of VMs and application-related data. This ensures fault tolerance and comprehensive disaster recovery capabilities of KS.

User-Friendly Management

KS offers a user-friendly solution for efficient management with EonOne and EonKube graphical user interfaces.

EonOne provides comprehensive hardware management for compute, GPU, and storage resources. Users can effortlessly monitor the resource status of each node and the overall cluster, making it simple to identify hardware issues. This enhances management efficiency and ensures optimal performance across the system.

EonKube is designed for managing software, including VMs and all deployed applications. It provides a simplified Kubernetes application deployment, making it more accessible for enterprises by eliminating the need for complex configuration and coding. The EonKube dashboard offers a clear view of system resources, such as CPU and GPU usage for each application, allowing administrators to monitor and adjust resource allocation as needed. Users can also easily check for the latest versions of built-in applications maintained and verified by Infortrend. Updates can be applied with a single click, streamlining the process and eliminating manual version checks, update package searches, and testing.

BMC (Baseboard Management Controller) support allows IT teams to remotely manage and monitor KS systems in both data center and edge environments.

Scalability

KS features scalability, enabling seamless expansion by non-disruptively adding one node at a time, up to a maximum of 20 nodes for KS 5000U and 5 nodes for KS 3000U. As enterprises' compute and storage requirements increase, they can effortlessly integrate an additional KS node, facilitating the growth of the KS cluster in alignment with business and application needs, thereby optimizing resources for their applications.

To expand storage capacity, KS supports various storage expansion options, allowing connection to EonStor GS unified storage or EonStor DS SAN storage.

Flexible Hardware Options

KS offers six compute node models across the KS 5000U and 3000U series with U.2 NVMe SSDs and configurable options for GPUs, allowing enterprises to select the model that best fits their application needs. All KS models come in a space-efficient 2U form factor and support 25GbE and 100GbE network interfaces.

Designed for standard enterprise data centers and high-performance workloads, the **KS 5000U** series is ideal for AI training, deep learning, HPC, and other compute-intensive applications. It includes the following models:

- **KSi 5008U / KSa 5008U** is designed as a GPU-intensive computing node, supporting up to four high-performance Nvidia GPU cards. This model is ideal for AI deep learning and high-performance computing tasks.
- **KSi 5016U** is a general-purpose computing node with two Nvidia GPU cards, providing efficient and scalable performance suitable for various enterprise data services, such as database management systems, data search engines, and other containerized workloads.
- **KSi 5024U** is a storage-oriented computing node supporting 24 SSDs. It is perfect for email servers, big data processing, and other storage-intensive applications.

Designed for edge deployments at the source of data generation, the **KS 3000U** series offers a short-depth 2U (50 cm) design and is suitable for time-sensitive, low-latency workloads requiring real-time decision-making, such as AI inference and video analytics. It includes the following models:

- **KSa 3004U** is best suited as a rackmount edge compute node, supporting up to two full-height 450W GPU cards, providing powerful edge AI computing for edge data centers.
- **KSa 3004UE** is a low-noise model operating at 55dB or less, primarily intended for desktop edge deployments in people-occupied environments such as offices and retail stores. It features low-power operation (down to 80W) and supports one GPU card, making it a flexible and efficient edge AI compute server.

Support for Cloud-Native and Traditional Applications

KS features a built-in Kubernetes platform for running cloud-native containerized applications. With a built-in hypervisor, KS provides support for traditional applications by running VMs. This facilitates enterprises to run all workloads on a single platform.

Furthermore, KS offers a built-in Application Marketplace with a diverse selection of applications for multiple domains and industries. The Marketplace includes commonly used enterprise applications, such as database systems, big data analytics platforms, and software development tools like CI/CD, as well as domain-specific applications for HPC, AI, and machine learning (ML), along with tools for managing related tasks. To minimize maintenance efforts, KS supports seamless one-click deployment and rolling updates, ensuring that applications stay up to date.

High Availability

The high-availability design of KS features the advanced failover function. The system can automatically detect node failures and quickly perform a failover, transferring applications to healthy nodes. Faulty node can be completely removed from and rejoined to the cluster system after repair, without requiring system downtime. This greatly enhances availability and ensures the continuity of enterprise applications.

Automatic Scaling

KS can dynamically scale cloud-native applications across nodes according to real-time node load conditions. It achieves this by continuously monitoring the hardware load of computing nodes, including metrics such as CPU, GPU, memory, storage space, and network usage. In response to heightened computing demands on an application, the system reallocates resources by automatically scaling the application to underutilized computing nodes. Conversely, when applications no longer require additional system resources, KS seamlessly downscales the allocated resources to optimize efficiency. By dynamically adjusting to changing demands, the automated scaling mechanism helps to ensure consistent application performance and optimized resource utilization.

Easy Deployment

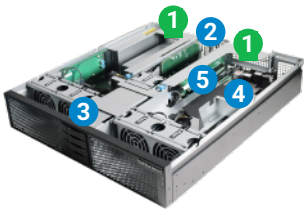
KS combines computing, storage, and Kubernetes platform within a single node appliance. This greatly simplifies deployment and management as KS eliminates the need for users to purchase these components separately and navigate through complex configurations, which can otherwise require substantial time, finances, and technical expertise.

The platform features an intuitive installation wizard that guides users through the entire setup process, which is completed within 30 minutes. After setup, enterprises can immediately start installing and running applications. The initial cluster setup can start with three KS or more computing nodes, and expansion is easy – users can simply add KS nodes one by one, with each node requiring just a few simple steps through the management interface. Alternatively, users can start with one KS node in a setup mode that does not support expansion.

KS 3000U Series

KSa 3004UE

KSa 3004U



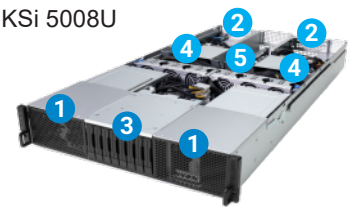
1 **KSa 3004UE**: FHHL single- or dual-width GPU slot
1 **KSa 3004U**: FHFL dual-width GPU slot
2 Network adapter slots
3 NVMe SSD slots
4 AMD EPYC™ 8004 Series Processor
5 DDR5 DIMM slots

PHYSICAL SPECIFICATIONS

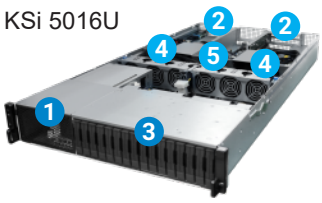
Product Series		KS 3000U	
Model		KSa 3004UE	KSa 3004U
Form Factor		2U 4-bay	2U 4-bay
Mounting Options		Desktop / Rackmount	Rackmount
Processor		Single AMD EPYC™ 8004 Series Processors	
Processor Core Available		16 to 24 cores	16 to 64 cores
Cache Memory		• 6 x DDR5 DIMM slots • Default 128GB, up to 576GB DDR5 RDIMM	
Supported Drives		2.5" U.2 NVMe SSD	
		Note: 1. A minimum of four drives is required. 2. For the latest Compatibility Guide, refer to our official website.	
PCIe Expansion Slots		• 1 x FHHL PCIe Gen5 x16 slot for GPU • 1 x HHHL PCIe Gen5 x16 slot • 2 x HHHL PCIe Gen5 x8 slots	• 2 x FHFL PCIe Gen5 x16 slots for GPUs • 1 x HHHL PCIe Gen5 x16 slot • 2 x HHHL PCIe Gen5 x8 slots
		Note: The number of supported network adapters per node via PCIe expansion slots depends on the deployment mode (standalone or cluster mode).	
Network Adapter Options		• Dual-port 25GbE SFP28 adapter • Dual-port 100GbE QSFP56 adapter	
		Note: Network adapters must be purchased from Infortrend.	
Max. FHHL Single- or Dual-width GPUs		1	0
Max. FHFL Dual-Width GPUs		0	2
System Fans		5	
Rear IO Ports		• 1 x RJ45 dedicated BMC/IPMI LAN port • 1 x RJ45 1GbE management port (shared with IPMI access) • 1 x RS232 serial port (COM) • 1 x VGA port • 4 x USB 3.2 Gen1 ports • 1 x RJ45 2.5GbE LAN port • 2 x RJ45 10GbE LAN ports	
Dimensions (W x H x D) (Without Chassis Ears and Protrusions)		445 x 86.8 x 498 mm	
Package Dimensions (W x H x D)		600 x 215 x 780 mm	
Power Supply	Power Supplies	2 x 1300W 80 PLUS Titanium redundant hot-plug	
	AC Input	100-127VAC/12A 200-240VAC/8.5A, 50/60Hz	
		Note: Please use 200-240VAC for the KS 3004U model	
Environmental (Without GPU)	Temperature	Operating: 10 to 35°C; non-operating: -40 to 60°C	
	Altitude	Operating: Sea level to 3,660m (12,000ft); non-operating: sea level to 12,192m (40,000ft)	
	Relative Humidity	5 to 95% non-condensing, operating and non-operating	
Safety Standards		• Electromagnetic compatibility: CE, BSMI, FCC	• Safety: UL, BSMI, CB

KS 5000U Series

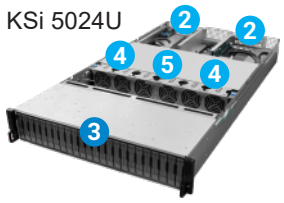
KSi 5008U



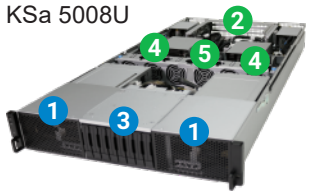
KSi 5016U



KSi 5024U



KSa 5008U



1. FHFL dual-width GPU slots
2. KSi: Network adapter slots
2. KSa: Network adapter slots
3. NVMe SSD slots
4. KSi: Intel® Xeon® Scalable Processor (4th or 5th Gen)
4. KSa: AMD EPYC™ 9004 Series Processor
5. KSi: DDR5 DIMM slots
5. KSa: DDR5 DIMM slots

PHYSICAL SPECIFICATIONS

Product Series	KS 5000U				
Model	KSi 5008**U**		KSi 5016**U**	KSi 5024**U**	KSa 5008**U**
Form Factor	2U 8-bay		2U 16-bay	2U 24-bay	2U 8-bay
Mounting Options	Rackmount		Rackmount	Rackmount	Rackmount
Processor	Dual 4th Gen Intel® Xeon® Scalable Processors	Dual 5th Gen Intel® Xeon® Scalable Processors	Dual 4th Gen Intel® Xeon® Scalable Processors		Dual AMD EPYC™ 9004 Series Processors
Processor Core Available	12 to 32 cores	64 cores	12 to 32 cores	12 to 32 cores	32 cores
Cache Memory	• 32 x DDR5 DIMM slots • Default 128GB, up to 2TB DDR5 RDIMM				• 24 x DDR5 DIMM slots • Default 192GB, up to 1.5TB DDR5 RDIMM
Supported Drives	2.5” U.2 NVMe SSD				
	Note: 1. A minimum of four drives is required. 2. For the latest Compatibility Guide, refer to our official website.				
PCIe Expansion Slots	• 4 x FHFL PCIe Gen5 x16 slots for GPUs • 2 x FHHL PCIe Gen5 x16 slots • 2 x OCP NIC 3.0 PCIe Gen5 x16 slots		• 2 x FHFL PCIe Gen5 x16 slots for GPUs • 2 x FHHL PCIe Gen5 x16 slots • 2 x OCP NIC 3.0 PCIe Gen5 x16 slots	• 2 x FHHL PCIe Gen5 x16 slots • 2 x OCP NIC 3.0 PCIe Gen5 x16 slots	• 4 x FHFL PCIe Gen5 x16 slots for GPUs • 3 x FHHL PCIe Gen5 x16 slots • 1 x OCP NIC 3.0 PCIe Gen5 x16 slots
	Note: The number of supported network adapters per node via PCIe expansion slots depends on the deployment mode (standalone or cluster mode).				
Network Adapter Options	• Dual-port 25GbE SFP28 OCP NIC 3.0 adapter • Dual-port 100GbE QSFP56 adapter				
	Note: Network adapters must be purchased from Infortrend.				
Max. FHFL Dual-Width GPUs	4	4	2	0	4
System Fans	6				
Rear IO Ports	• 1 x RJ45 dedicated BMC/IPMI LAN port • 1 x RJ45 1GbE management port • 1 x RJ45 serial port (COM) • 1 x Mini-DP port (supports VGA output) • 2 x USB 3.2 Gen1 ports (Type A)				• 1 x RJ45 dedicated BMC/IPMI LAN port • 1 x RJ45 1GbE management port • 1 x RJ45 serial port (COM) • 1 x DB15 VGA port • 2 x USB 3.2 Gen1 ports (Type A)
Dimensions (W x H x D) (Without Chassis Ears and Protrusions)	449 x 88 x 922 mm			449 x 88 x 852 mm	449 x 88 x 922 mm
Package Dimensions (W x H x D)	596 x 262 x 1188 mm			596 x 262 x 1118 mm	596 x 262 x 1188 mm
Power Supply – Power Supplies	2 x 2700W 80 PLUS Titanium redundant hot-plug	2 x 3200W 80 PLUS Titanium redundant hot-plug	2 x 2700W 80 PLUS Titanium redundant hot-plug	2 x 1600W 80 PLUS Titanium redundant hot-plug	2 x 2700W 80 PLUS Titanium redundant hot-plug
Power Supply – AC Input	200-240VAC/16-13.5A, 50/60Hz	100-127VAC/18A, 200-240VAC/16A, 50/60Hz	200-240VAC/16-13.5A, 50/60Hz	100-127VAC/12A, 200-240VAC/10A, 50/60Hz	200-240VAC/16-13.5A, 50/60Hz
	Note: Please use 200-240VAC for all KS model				
Environmental (Without GPU) – Temperature	Operating: 10 to 35°C; non-operating: -40 to 60°C				
Environmental (Without GPU) – Altitude	Operating: Sea level to 3,660m (12,000ft); non-operating: sea level to 12,192m (40,000ft)				
Environmental (Without GPU) – Relative Humidity	5 to 95% non-condensing, operating and non-operating				
Safety Standards	• Electromagnetic compatibility: CE, BSMI, FCC • Safety: UL, BSMI, CB				

DATA SERVICES

EonSight		Optional	EonSight enables AI-driven system monitoring and analytics, providing anomaly detection, performance insights, and capacity forecasting to help optimize cluster stability and operational efficiency.
Cluster Upgrade	Single-to-Cluster Upgrade	Optional	Cluster Upgrade enables single-node systems to scale into a multi-node cluster, supporting distributed workloads, high availability, and centralized management.
		Note: 1. This license is applicable only to single-node systems and enables scale-out to a cluster configuration. 2. After upgrading to cluster mode, each node requires two network adapters.	
Application Services	AI	Optional	AI License enables one-click deployment of built-in AI applications via the App Marketplace, accelerating AI environment setup and supporting advanced data analytics and machine learning workloads.
	HPC	Optional	HPC License enables deployment and management of built-in HPC applications through the App Marketplace, supporting efficient job scheduling, workload monitoring, and centralized management of HPC data and tasks.
	Manufacture	Optional	Manufacturing License enables one-click deployment of built-in manufacturing applications via the App Marketplace, providing a streamlined way to support and manage manufacturing operations and workflows.

WARRANTY AND SERVICE

Service and Support	Standard Service	3-year limited hardware warranty and 8 x 5 phone, web, and email support (batteries are covered under warranty for 2 years)
	Upgrade or Extension Options	Warranty extension: Standard service can be extended up to 5 years. The following services can be upgraded to 5 years. • Upgrade: Replacement part dispatch on the next business day • Advanced service: Phone, web, and email support + onsite diagnostics on the next business day • Premium service: Phone, web, and email support + onsite diagnostics within 4 hours
		Note: Options may vary by region. For more details, please contact our sales representatives.
	Technical Support	Get information on system installation and maintenance, download technical documents and software, or issue a support ticket
	Product Services	Register products, download firmware, apply for licensing services, create product repair tickets, or check product repair status

Asia Pacific (New Taipei, Taiwan)
Infortrend Technology, Inc.
Tel : +886-2-2226-0126
E-mail : sales.ap@infortrend.com

China (Beijing, China)
Infortrend Technology, Ltd.
Tel : +86-10-6310-6168
E-mail : sales.cn@infortrend.com

Japan (Tokyo, Japan)
Infortrend Japan, Inc.
Tel : +81-3-5730-6551
E-mail : sales.jp@infortrend.com

Americas (Sunnyvale, CA, USA)
Infortrend Corporation
Tel : +1-408-988-5088
E-mail : sales.us@infortrend.com

EMEA (Düsseldorf, Germany)
Infortrend Technology, Inc.
E-mail: sales.de@infortrend.com

Contact Sales

Visit Our Website