Hitachi Vantara

Datasheet

Hitachi iQ

Hitachi's End-to-End Solutions Stack for Enterprise AI/ML Workloads

Hitachi iQ delivers extreme performance and resiliency at scale and makes your AI infrastructure simpler and faster to design, deploy, and manage.

Organizations are constantly looking for ways to automate, accelerate time-to-market, and develop new insights, products, or innovations to propel their business forward. At the same time, AI and Generative AI technologies are revolutionizing industries by enhancing existing capabilities and transforming how quickly and creatively business problems can be solved. Whether building specific AI solutions or just starting to identify general-purpose capabilities, Hitachi iQ has the power to automate your business processes and improve your AI experience.

Hitachi iQ allows organizations to automate, expedite, and streamline their business through intelligent, performant, scalable and flexible AI and GenAI solutions. Unlike conventional AI offerings, Hitachi iQ transcends basic integration and storage capabilities by also layering industry-specific AI outcomes within the AI solution. This approach ensures outcomes are finely tuned to each organization's unique needs and objectives.

Hitachi iQ is an accelerated solution that provides unified access to data irrespective of where it resides while ensuring explainability, lineage, data accuracy, security, and traceability at any given point for mission critical solutions. It optimizes AI deployments with an end-to-end software-defined AI and data analytics software to streamline development and deployment of production-grade AI applications, from pilot to production.

Fast and Efficient Results

Harnessing the power of validated and seamlessly integrated industry-leading AI technologies to accelerate your transformation journey and quickly gain insights.

Timely and Accurate Decisions

Ensuring data is relevant and accessible for AI applications and analytics, while prioritizing security and compliance.

Industry Relevant, Meaningful Outcomes

Recognizing not all organizations possess the AI skills for their industry, we bring our expertise to help bridge the gap, creating cutting-edge AI applications, accelerating your outcomes.

- **Personalize Your AI** – Built to meet the rigors of AI and engineered for industry outcomes, but customized to your needs.
- **Achieve Faster Insight** – Accelerated architecture delivers storage performance that improves GPU resource utilization.
- **Simplify to Scale** – Validated reference architecture blueprints provide the flexibility and scale to rapidly develop, test, and deploy modern AI solutions.
- **Lower TCO** – Leverage lower cost, erasure coded, scale-out object storage to safeguard data, catalog for re-use and store data long term.
- **Improve Accuracy** – Increase quantity and quality of data to improve reliability of results. Identify, classify, transform, move, consolidate and prepare data to get the most value out of AI/ML initiatives.

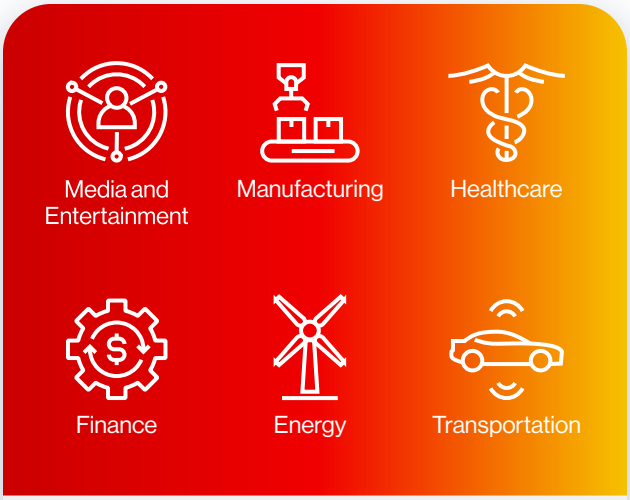
Hitachi iQ is the flagship AI portfolio from Hitachi, powered by NVIDIA® technologies.

When customers try to implement their own AI solution, they are often faced with the challenges of creating complicated systems and performing the necessary integrations on their own. Instead, Hitachi iQ offers a pre-built platform with bespoke customization and delivery. Engineered with Hitachi & NVIDIA solutions, Hitachi iQ is an accelerated integration based on proven foundations. For AI accelerated compute, the NVIDIA DGX™ H100 platform combines the best of NVIDIA software, infrastructure, and expertise in a modern, unified AI development solution. For primary storage, Hitachi Content Software for File provides a distributed parallel file system that delivers the highest performance file services by leveraging NVMe flash, and also includes integrated tiering that seamlessly expands a single namespace to and from hard disk drive (HDD) object storage without the need for special data migration software or complex scripts. Object storage, data tiering, and data protection can be optionally provided by Hitachi Virtual Storage Platform One Object (VSP One Object).

Workloads and Use Cases

Whether you are looking for industry-specific AI solutions or just starting to identify general-purpose capabilities, AI has the power to automate your business processes and improve your customer experience. The industry is already using AI for things like creating art and graphic design elements, writing code, and generating marketing tag lines. And new use cases are being identified every day that can provide immediate benefits to any organization. For example:

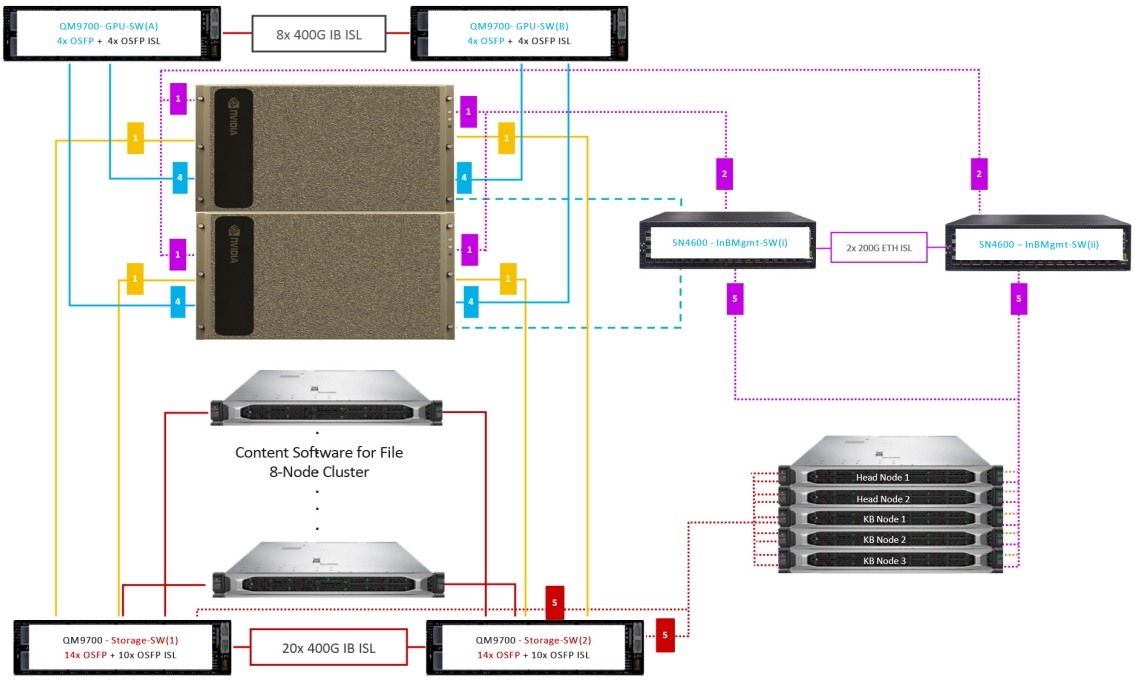
- **Customer Service Voice Assistant:** Revolutionizes customer service with its advanced voice recognition technology, offering real-time responses and personalized assistance.
- **Large Language Model (LLM) Recommender System:** Delivers highly accurate, context-aware suggestions and personalizes content, products, and services, making discovery seamless and engaging.
- **Coding and Development Copilot:** Can act as an invaluable partner in the development process, offering real-time suggestions, debugging assistance, and code optimization, accelerating development cycles and enhancing code quality.
- **Automated Document Processing and Analysis:** Streamlines the processing, analysis, and management of large volumes of documents within enterprise environments, significantly reducing manual workloads, improving accuracy, and enhancing decision-making.
- **Financial Reporting & Accounting:** Can reduce the repetitive tasks that workers and consultants are required to do, helping streamline operations & reducing errors for data entry, transaction categorization, and invoice processing.
- **Edge Inference:** Brings AI computing closer to the data source, minimizing latency and enhancing real-time decision making. It's ideal for applications requiring instant analysis and action, from autonomous vehicles to smart city infrastructure.



Hitachi iQ leverages its deep domain experience with 110+ years of experience in the Operational Technology (OT) market in industries such as finance, transportation, energy, media & entertainment, manufacturing, and healthcare. With Hitachi's OT and IT legacy, the Hitachi iQ portfolio integrates infrastructure and capabilities to deliver industry-specific AI solutions tailored to our customers' needs.

Solution Architecture

Hitachi iQ with Hitachi Content Software for File is a certified NVIDIA DGX BasePOD™ solution , creating a robust offering for high-performance computing (HPC) and advanced AI workloads. This certification validates the seamless integration and interoperability between NVIDIA's accelerated infrastructure and Hitachi's scalable file storage system, ensuring optimal performance and reliability. For more information and detail about the Hitachi iQ reference architecture and testing for NVIDIA DGX BasePOD, please refer to "Hitachi iQ: Hitachi Content Software for File with NVIDIA DGX BasePOD" on the Hitachi Vantara website.



Compute Fabric in NDR IB
GPU-GPU bandwidth: 6.4 Tb/s

Storage Nodes Fabric in NDR IB
8x Content Software for File storage bandwidth: 6.4 Tb/s

Compute Storage Fabric in NDR200 IB
2x DGX H100 data network bandwidth: 0.8 Tb/s

Management node storage fabric 200G IB
5x Management nodes storage/compute bandwidth: 2 Tb/s

DGX/Management node SN4600 200G ETH
2x DGX nodes SN4600 bandwidth: 0.8 Tb/s
5x Management nodes SN4600 bandwidth: 2 Tb/s

DGX Management connection 100G ETH

Table 1: Hitachi iQ Components and Performance

NVIDIA Architecture	**NVIDIA DGX H100 System:** Powered by eight NVIDIA H100 Tensor Core GPUs per node, the NVIDIA DGX H100 is engineered to maximize AI throughput. **NVIDIA AI Enterprise:** Comprehensive, cloud-native software platform designed to accelerate data science pipelines and streamline the development and deployment of production-grade co-pilots and generative AI applications. **NVIDIA Base Command™:** Enterprise grade cluster management, libraries that accelerate compute, storage, network infrastrsucture, and system software optimized for running AI workloads. **NVIDIA GPU Direct Storage:** Enables memory access and coherency features that bypass system CPUs and/or memory and ensures data consistency and integrity between GPUs and storage devices.		
Storage	**Hitachi Content Software for File 36116 Storage Nodes:** Primary storage		
Networking	**NVIDIA QM9700 400Gb/s InfiniBand Switches:** Compute and storage fabric **NVIDIA SN4600 200GbE Ethernet Switches:** Storage fabric for management servers and in-band management network **NVIDIA SN2201 1GbE Ethernet Switch:** Out-of-band management network		
Management Servers	**Hitachi Vantara HA810 G3 Servers**		
Hitachi iQ Solution Performance*	**Performance Characteristic**	**NVIDIA Recommendation (Best)**	**Hitachi iQ Performance**
	Single-node read	40 GB/s	60 GB/s
	Single-node write	20 GB/s	60 GB/s
Hitachi Content Software for File Performance*	**Workload Type**	**Peak Performance**	
	Sequential read performance	745 GB/s	
	Sequential write performance	257 GB/s	
	Random 4kB read performance	26.2 MIOPS	
	Random 4kB write performance	6.16 MIOPS	
	4kB read latency	112µs	
	4kB write latency	78µs	

About Hitachi Vantara

Hitachi Vantara, a wholly owned subsidiary of Hitachi Ltd., delivers intelligent data platforms, infrastructure systems and digital expertise that supports more than 80% of the Fortune 100. To learn how Hitachi Vantara turns businesses from data-rich to data-driven through agile digital processes, products, and experiences, visit hitachivantara.com.

Hitachi Vantara
A Hitachi Group Company

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
hitachivantara.com | community.hitachivantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
hitachivantara.com/contact

HV-BTD-DS-Hitachi iQ-DS-16July24-A