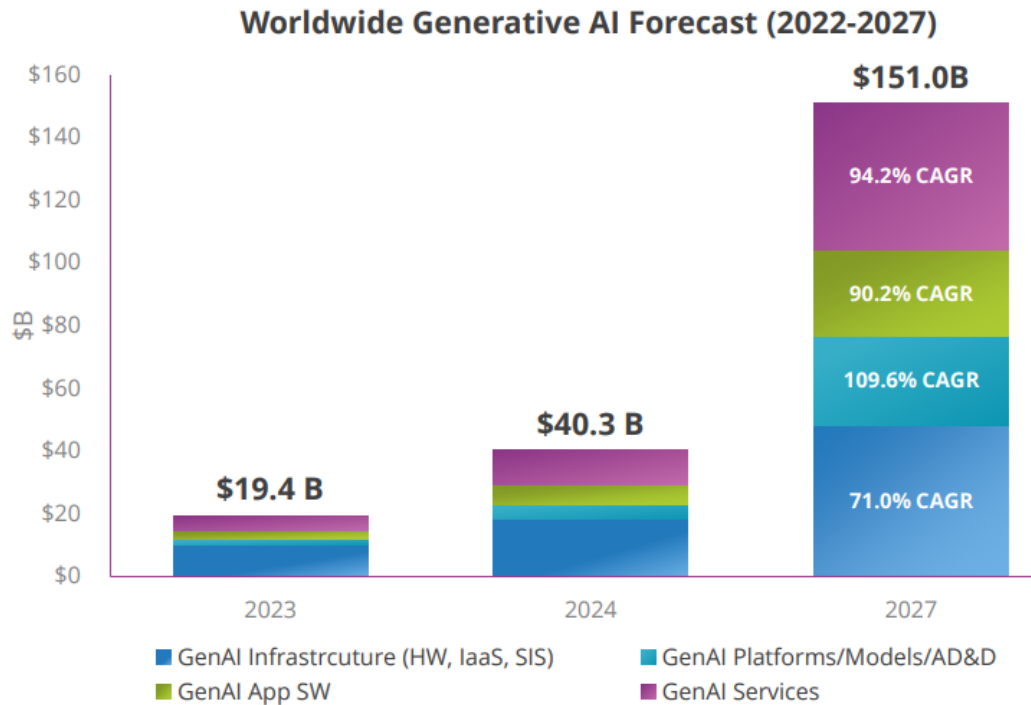


Next Generation Data Platform Hitachi iQ

AI Market Overview



Source: Market Intelligence, IDC

- Purpose-built AI infrastructure is taking over the data center
- Diverse landscape covering cloud, edge, GPUs, CPUs and ASICs as practitioners experiment with new architectures
- Investment in data integrity and data access needed for AI workloads, will spur technology refresh and improvements
- Biggest growth in Platforms/Models and application Software
- Infrastructure and Services remain the largest components of the TAM in 2027

AI Market Overview – Teil 2

- 6-9%
- 8/10
- 33%/8
- 4% - 2023
- Mehr als 80%
- Aller Storage Projekte scheitern
- Aller AI-Projekte scheitern
- Aller AI-Projekte scheitern, weil man die Infrastruktur nicht in den Griff bekommt
- Aller AI-Projekte sind weltweit 2023 bis in Produktion gegangen
- Aller CEOs im deutschen Mittelstand beschäftigen sich mit AI

[GEN]AI | *R/evolution Journey to Data Experienced*

Data Anywhere [GEN]AI Everywhere

Hybrid Cloud



"Unified data Intelligence" from
Data Anywhere and in any format

Infrastructure Solutions [GEN]AI Data

Extreme Performance Systems



An appliance with high
performance and load
optimization for cost efficiency

Simplification Built in [GEN]AI

Smart Companions



"Reactive" to **"Proactive"**
resolution of customer issues for
Hitachi Vantara products and solutions

Outstanding time-to-value – targeting domain experts without ML experience

Hitachi Ecosystem for Performance Workloads

Data Driven Business Use Cases and Outcomes



Quality
Datasets



360 degrees
Observability



Reporting and
Analysis



Strategic
Analysis



New Business
Revenue



Data
Monetization



AI Driven
Revenue



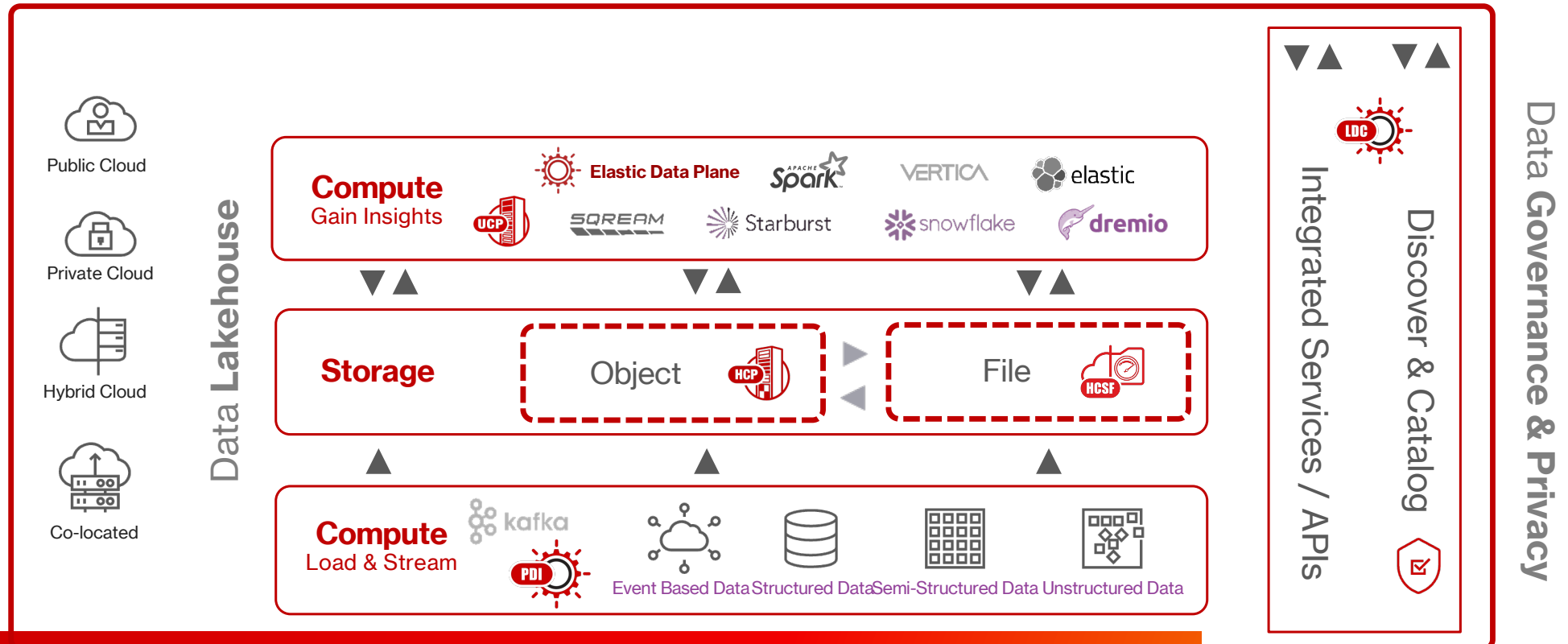
AI Driven
Insights



Improved
Collaboration



Compliance



Industry Leaders Unite to Advance AI

Leader in data
infrastructure
technology &
solution frameworks



Leader in Artificial
Intelligence
Computing



DELIVERING CUSTOMER
VALUE THROUGH
STRATEGIC PARTNERSHIPS

Leader in industrial
and operational
technologies



Accelerating Transformation Across Industrials

Leveraging Hitachi's domain expertise for 100+ years



Finance



- Fraud Detection Systems
- Customer 360
- Algorithmic Trading

Energy



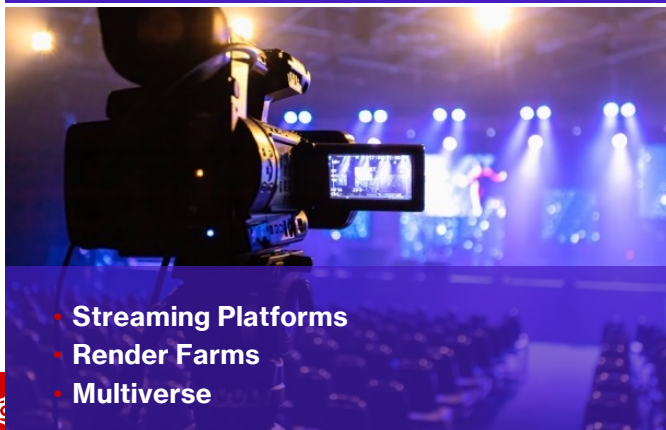
- Smart Power Grid
- Energy Planning
- SCADA Cybersecurity

Transportation



- Autonomous Drive
- ADAS
- Predictive Maintenance

Media & Entertainment



- Streaming Platforms
- Render Farms
- Multiverse

Manufacturing



- Process Automation
- Smart Manufacturing
- Predictive Maintenance

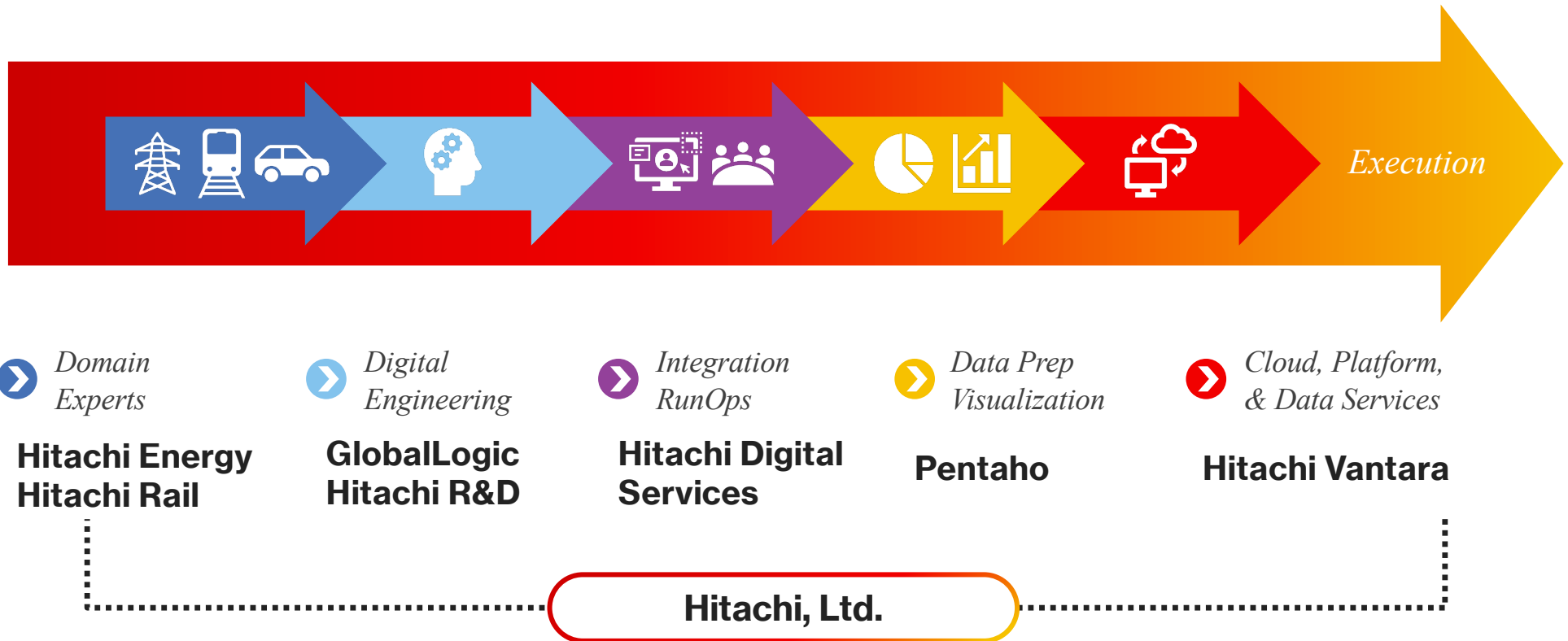
Healthcare



- Gene Research
- Genome Mapping
- Precision, Remote Medicine

The Sum Of All Parts

Hitachi's End-to-End capabilities enabling our clients AI journey



Hitachi iQ | With

ENTRY



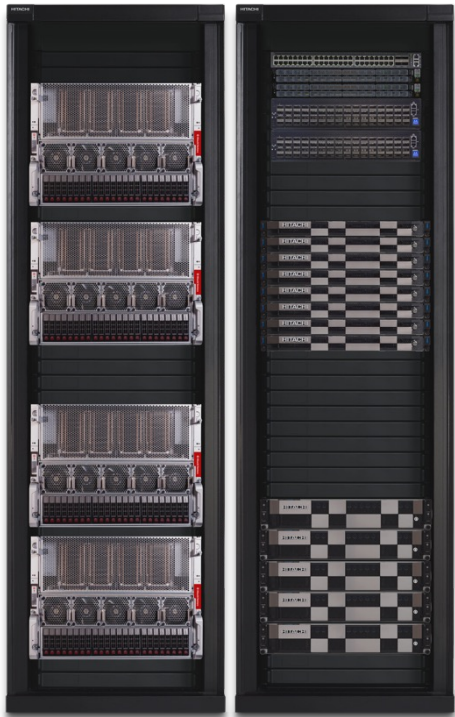
Hitachi HA820

MIDRANGE



Hitachi L40S/H100

HIGHEND



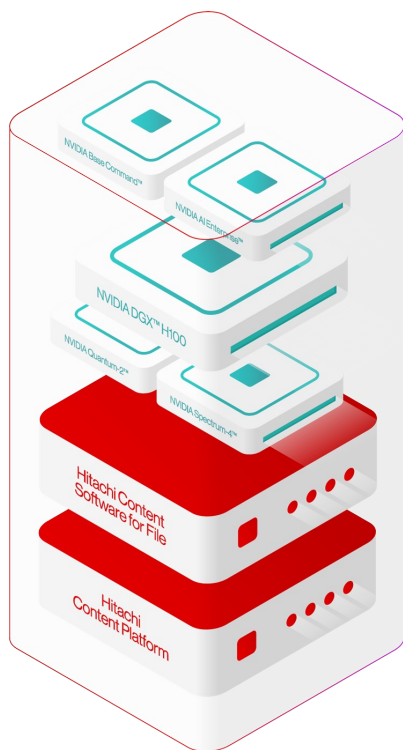
Hitachi HGX

*DGX**



NVIDIA DGX

AI-Ready Infrastructure

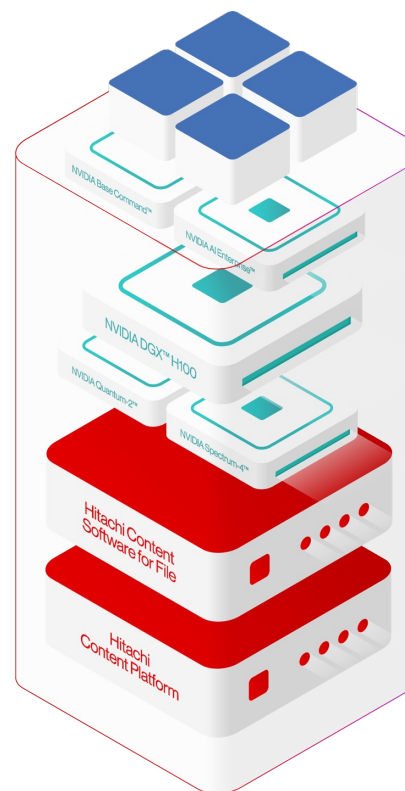


Benefits

Meet performance, scale, efficiency, integration, and reliability demands.

- Single vendor approach
 - Reliable infrastructure
 - Reference architectures
- NVIDIA DGX BasePOD™ certified solution meets the criteria of the "Best" category.
- Scale compute and storage independently
- Improve economics and sustainability

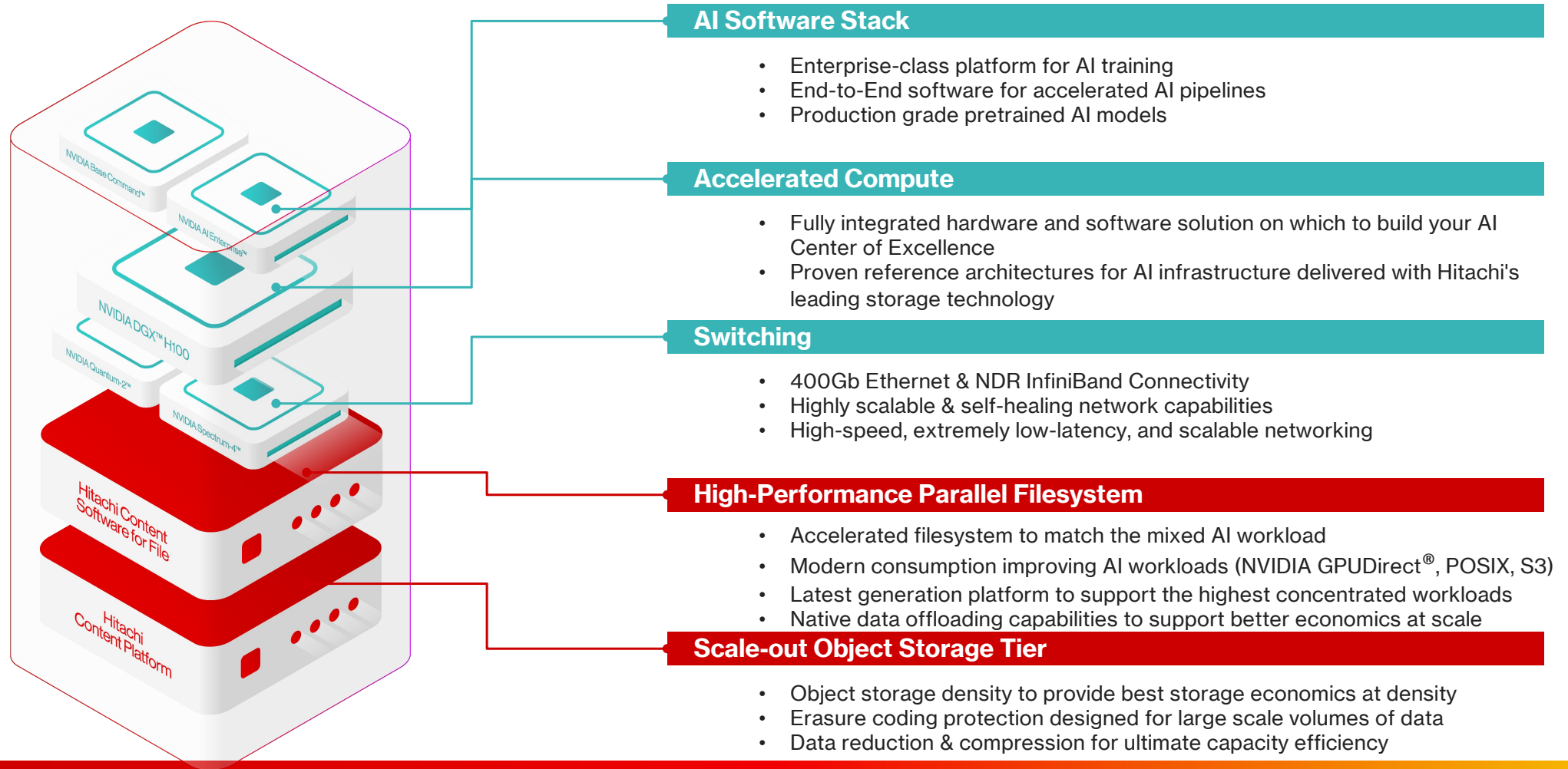
AI-Ready Solutions



Benefits

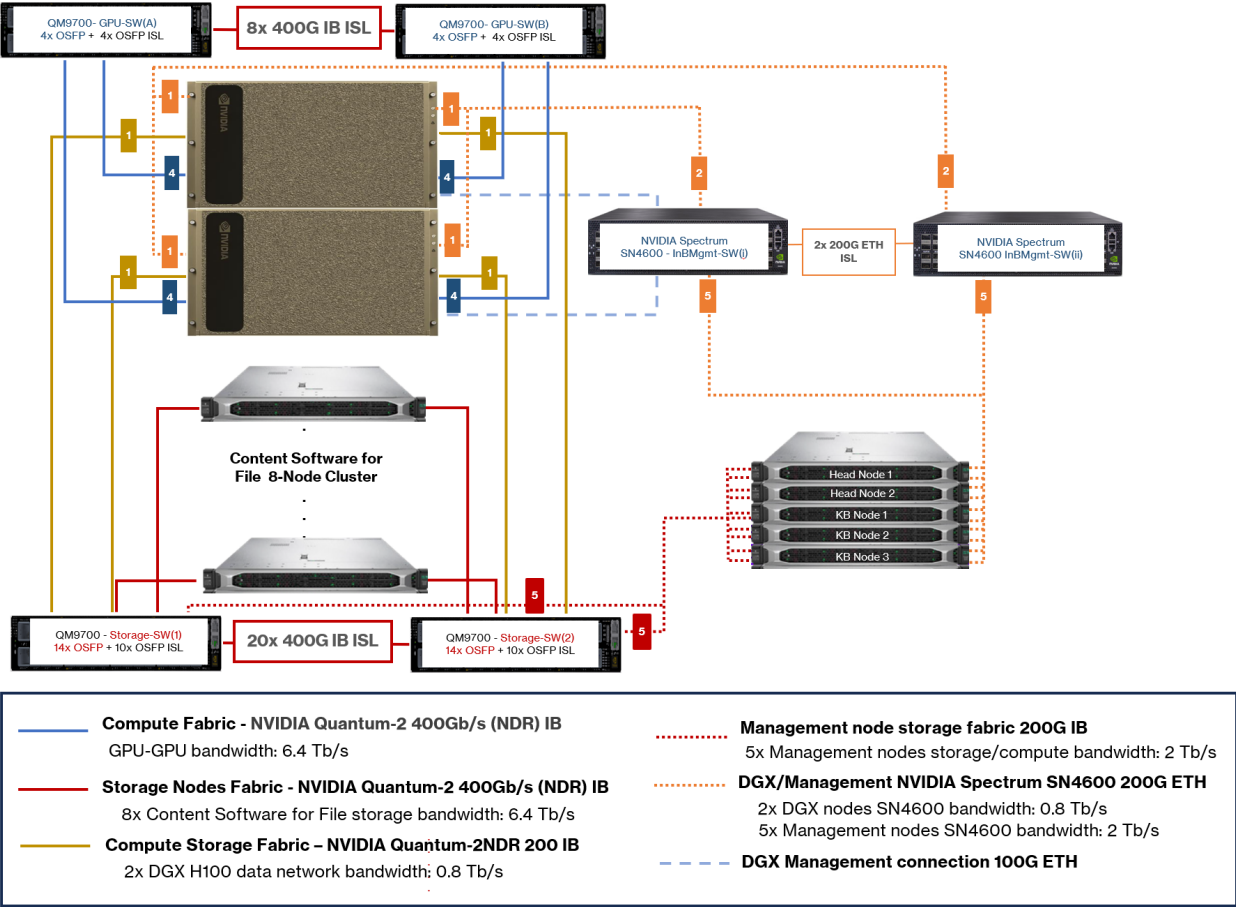
- Realize faster outcomes with Hitachi led solutions across multiple BU's (Hitachi Digital & Hitachi Ltd.)
- Outcome-led design
- Engineered with Hitachi & NVIDIA Solutions
 - Long history of domain expertise in industries such as rail, energy, manufacturing
- Accelerated integration based on a pre-built foundation
- Bespoke customization & delivery

Hitachi iQ AI-Ready Infrastructure with NVIDIA DGX



AI Reference Architecture

Topology and Components



NVIDIA Architecture	NVIDIA DGX H100 System: Powered by eight NVIDIA H100 Tensor Core GPUs per node. NVIDIA AI Enterprise NVIDIA Base Command Manager NVIDIA GPU Direct Storage
Storage	Hitachi Content Software for File 36116 Storage Nodes
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	Hitachi Vantara HA810 G3 Servers

“

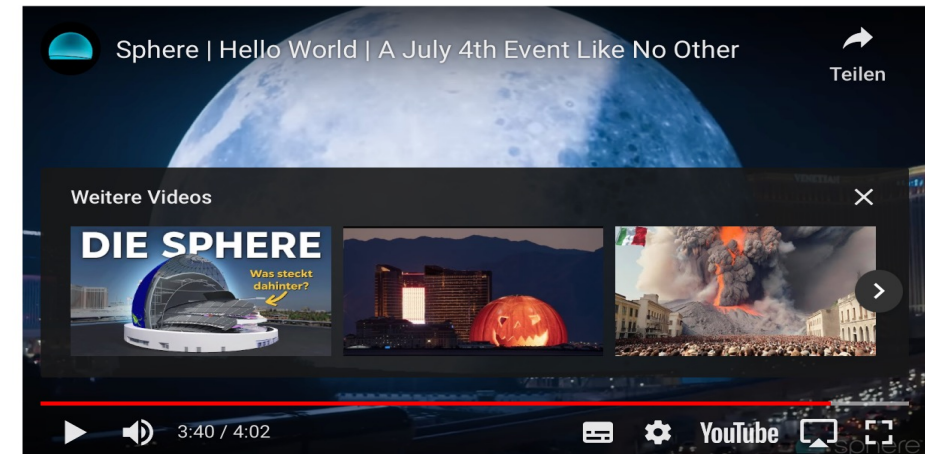
Hitachi Vantara worked with our team to develop a solution that's fast, reliable, and efficient.

Through their expertise and technology, Sphere's displays deliver content that is captivating audiences not only in the venue in Las Vegas, but also around the world on social media."

Alex Luthwaite, SVP,
Show Systems Technology,
Sphere Entertainment



The Sphere measures 580,000 square feet, while the inner side measures 160,000 square feet. The Sphere is capable of some truly outstanding visuals. From planets and moons to trippy art pieces, the outside of the Sphere is more than well-equipped for its high-res projection purposes. Since it was activated for the first time on July 4, 2023, with a high-res view of Earth, it has consistently been praised for its novelty—though one day it may become a Times Square-esque advertiser's surface, which is much less exciting.



Reportedly, both the internal and external screens support **16K resolution** video output. They each rely on a collection of 27 nodes, each streaming at 4K through Hitachi Vantara's software, with a whopping 4 Petabytes (1 Petabyte = 1000 TB) of flash memory capable of 400 GB/s speeds. Full 4:4:4 chroma subsampling is also used, and reportedly, the displays can achieve a latency of around five milliseconds or less.

Customer Profile: Autonomous EV Company

❑ Deep Learning for Autonomous Vehicle

- Autopilot

❑ Problem

- Epoch runs were taking over two weeks to complete
- Data management on local drives induced overhead
- NFS performance was limited
- Hot and cold data were stored on disparate systems

❑ HCSF Solution

- Three training clusters overall 6PB flash, and 27PB object storage
- In production since late 2017

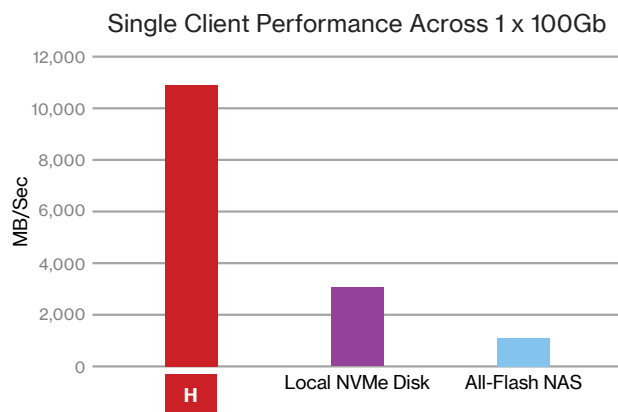
❑ With HCSF

- Epoch time was reduced from 2 weeks to 4 hours
- Single namespace for all data



Why HCSF: Powered Performance

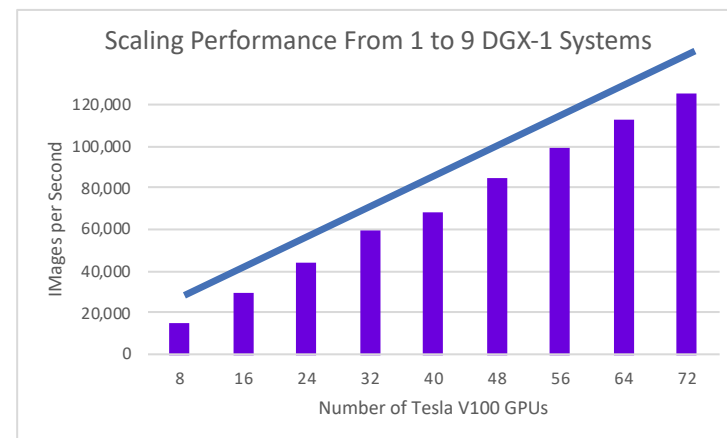
GPU-accelerated applications are data hungry



HIGH-SPEED NETWORK SATURATION

162 GB/sec & 1.5M IOPs
Performance to a *Single GPU client*

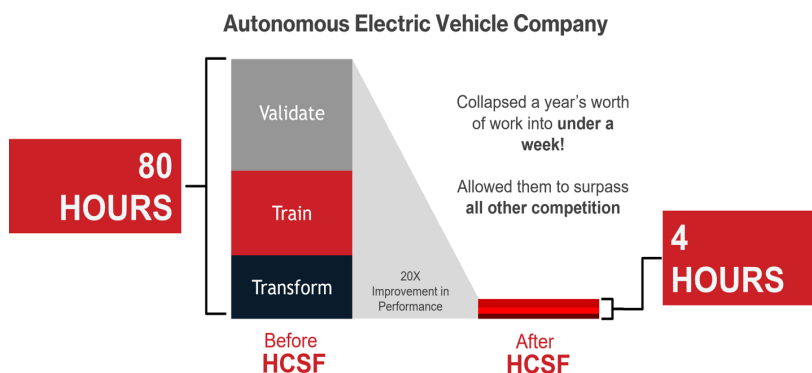
MASSIVE SINGLE CLIENT PERFORMANCE



PERFORMANCE SCALES LINEARLY TO MULTIPLE CLIENTS

AI/ML, DEEP LEARNING
DATA PIPELINE FOR
AUTONOMOUS DRIVING

- REDUCED DATA EPOCH RUNTIME FROM 2 WEEKS TO 4 HOURS VIA MODERNISATION OF NETWORKING & STORAGE WITH HCSF



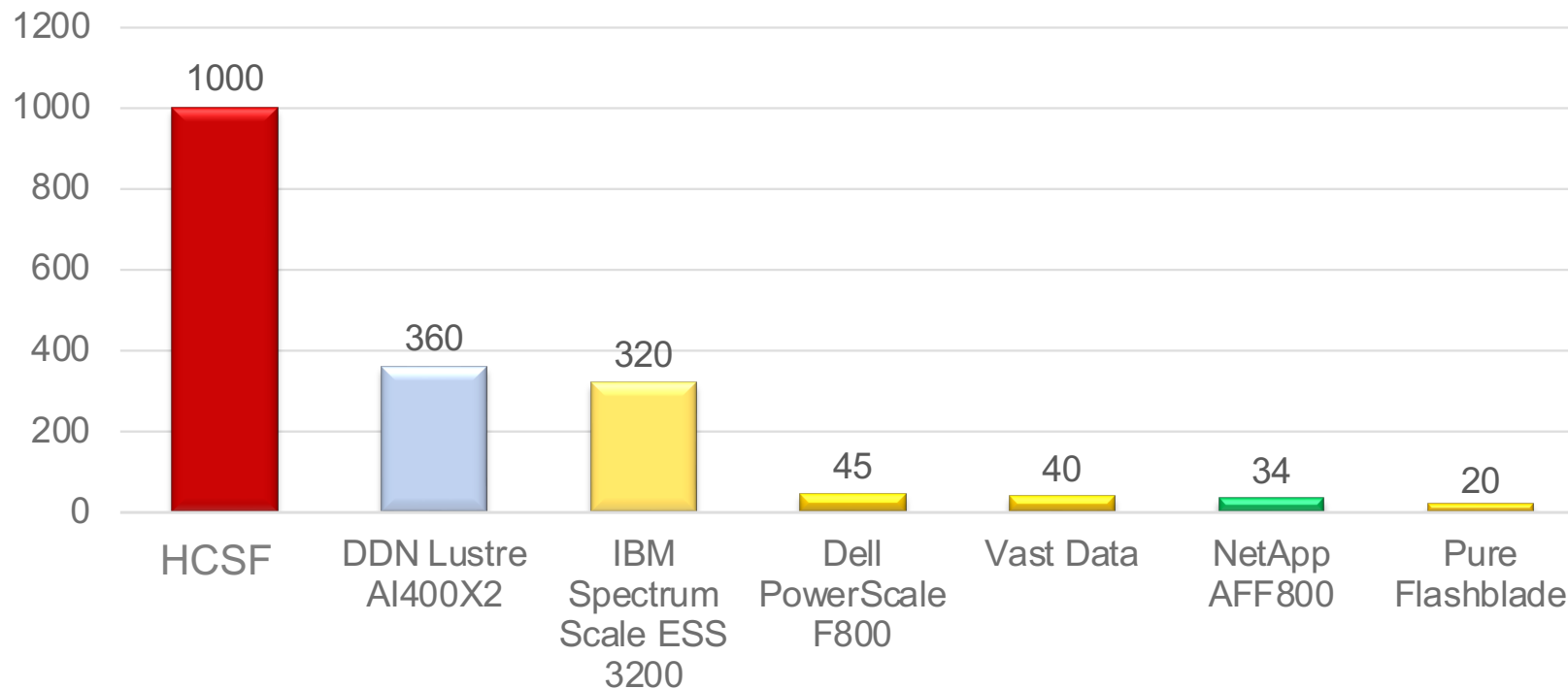
Benchmark	DAS (Optane SSD Server)	SAN (NVMe-oF)	NAS (All-Flash)	WekaFS (HPE NVMe Servers)
100T.YR1VWAB-12D-HO	15633	1886	4183	1028
100T.YR2VWAB-12D-HO	18114	1418	3294	892
100T.YR3VWAB-12D-HO	20730	1910	4773	1141
100T.YR4VWAB-12D-HO	24741	3317	7037	1550
100T.YR5VWAB-12D-HO	36888	22389	11376	4808
10T.YR2-MKTSNAP	176	355	6898	655
10T.YR3-MKTSNAP	176	358	7855	675
10T.YR4-MKTSNAP	149	375	8531	711
10T.YR5-MKTSNAP	155	393	8684	726
1T.2YRHIBID	645	374	1419	309
1T.3YRHIBID	1129	630	2737	480
1T.4YRHIBID	1957	1082	4881	804
1T.5YRHIBID	3234	1804	8589	1234
1T.OLDYRHIBID	61	46	129	48
1T.YR1VWAB-12D-HO	334	226	545	294
1T.YR2VWAB-12D-HO	394	268	632	355
1T.YR3VWAB-12D-HO	462	347	750	430
1T.YR4VWAB-12D-HO	553	517	928	547
1T.YR5VWAB-12D-HO	841	769	1298	732
50T.YR1VWAB-12D-HO	1089	1748	4302	2300
50T.YR2VWAB-12D-HO	1988	1774	4798	1971
50T.YR3VWAB-12D-HO	2865	2278	6253	2409
50T.YR4VWAB-12D-HO	4195	3118	8840	3077
50T.YR5VWAB-12D-HO	6731	4625	13597	4111
Average Result (lower is better)	5968.33	2166.958	5097.041	1303.625

FINANCE, TICK DATA
ANALYTICS APPLICATION
PERFORMANCE

- 4X FASTER THAN ALL-FLASH NAS
- 4.5X FASTER THAN DAS WITH OPTANE
- 1.6X FASTER THAN ALL-FLASH ARRAY

HCSF vs Market – 1 Petabyte Bandwidth

**Bandwidth Performance GB/s
(1PB Usable Capacity)**



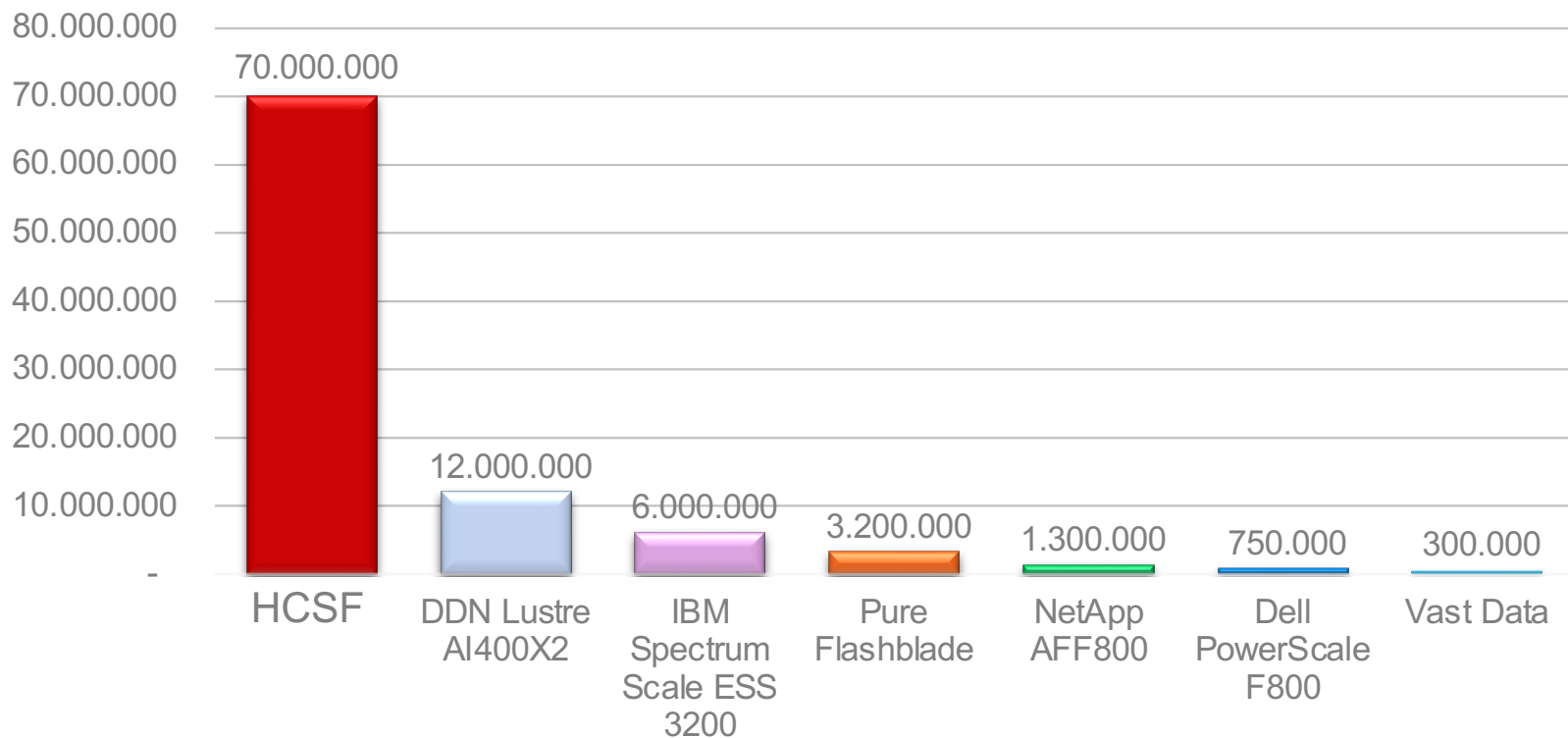
50x
More BW than **Pure**

25x
More BW than **Netapp**

3x
More BW than **IBM**

HCSF vs Market – 1 Petabyte IOPS

IOPS Performance (1PB Usable Capacity)



53x

More IOPS than **Netapp**

22x

More IOPS than **Pure**

12x

More IOPS than **IBM**

Getting Started with AI

AI DISCOVERY SERVICE FOR HITACHI IQ

Discovery Sprint ~3 Weeks

Value	✓
Use Case	✓
Data Assess	✓
Roadmap	✓

Week 1

Opportunity Analysis

Immersion in company priorities, inventory of internal data sources, risk identification, and opportunity prioritization.

Week 2

Use Case Selection

Identify priority use cases to optimize the business and identify a candidate for initial pilot based on value creation opportunity, risk, and data availability.

Week 3

Planning

Plan a proof-of-concept vision, approach, and timeline, aligned to value creation goals, hypotheses to validate, and success criteria. Recommend high-level roadmap for next actions.

Results

Deliverables

Research report on user insights, prioritized AI use cases roadmap, and AI concept documentation and recommendations for roadmap and proof-of-concept

AI Discovery for Hitachi iQ

	Discovery Sprint ~3 Weeks	AI Advisory ~5 Weeks	AI Jumpstart ~8-12 Weeks
Value	✓	✓	✓
Use Case	✓	✓	✓
Data Assess	✓	✓	✓
Roadmap	✓	✓	✓
Tech Assess		✓	✓
Model Selection		✓	✓
POC			✓
Prod Planning			✓

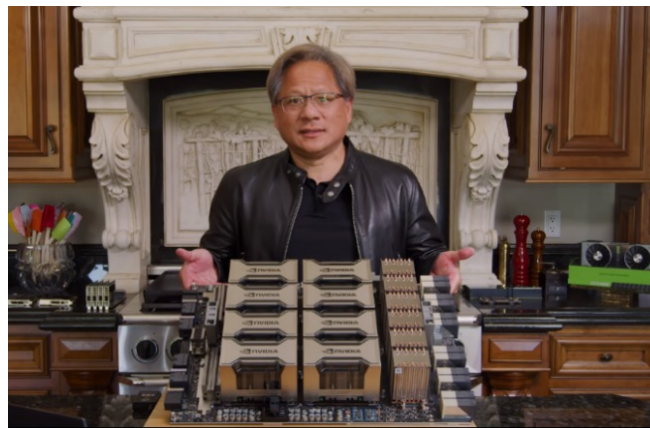
AI Discovery Accelerator

Prioritize a Use Case

Build a Roadmap

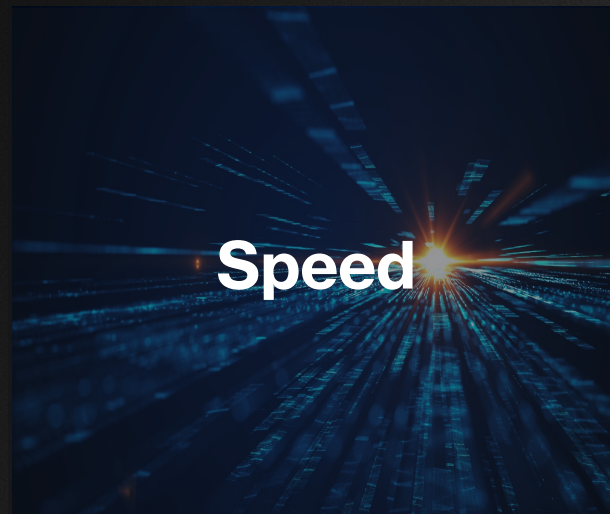
Develop a POC

Many Companies *Already* Bought into the AI/GPU Craze



Hitachi Content Software for File & SQream

Big Data Analytics has a Glass Ceiling Problem....

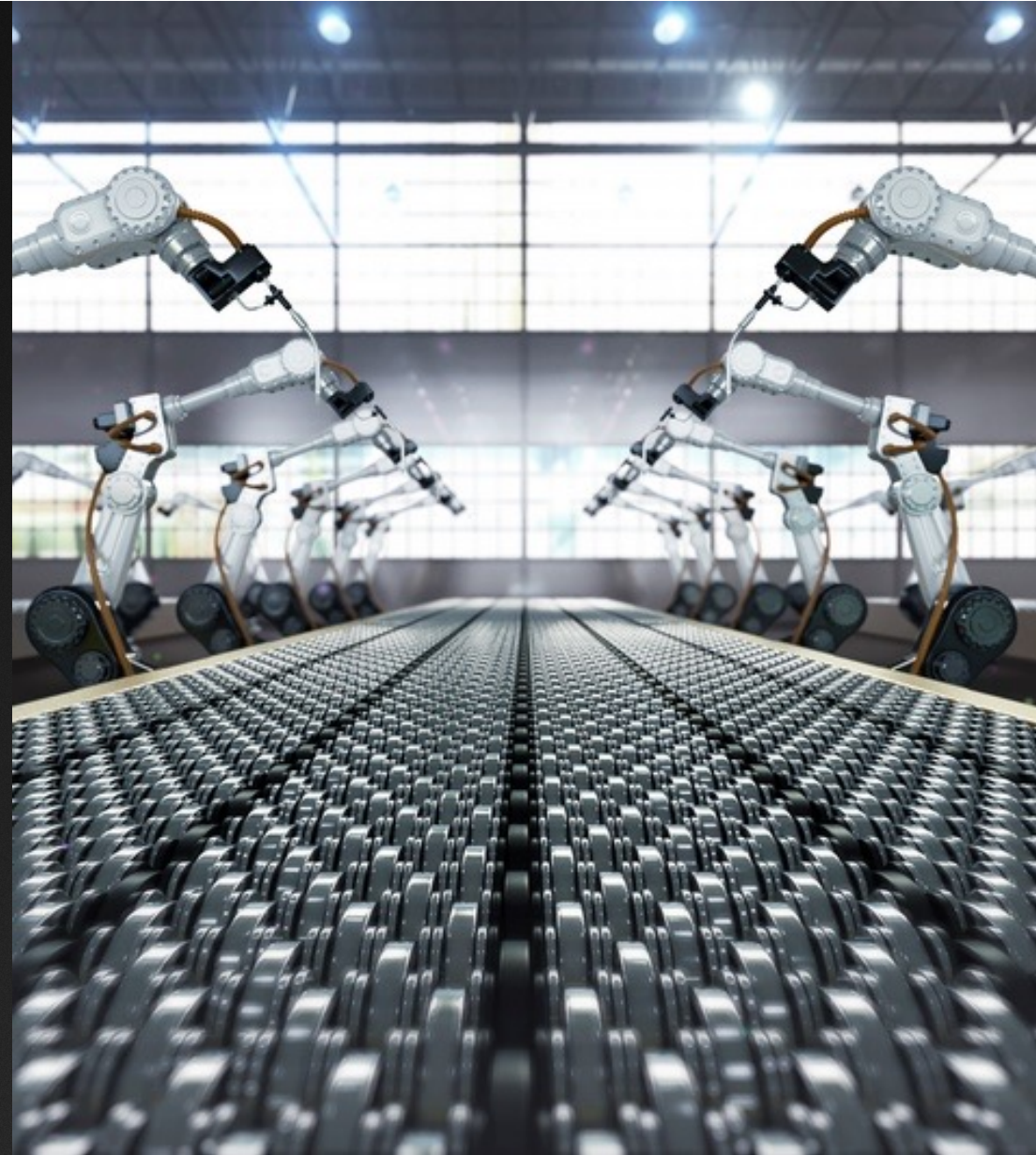


Who , what, and why SQream?

- **SQream** is a patented GPU-based Data & Analytics Acceleration Platform that enables complex analytics at unprecedented speed and in the most efficient footprint.
- **SQream** is software that integrates with Hitachi Content Software for File
- **SQream** provides seamless integration between GPU, File Management Software, and Object Storage
- **SQream Use Cases**
 - Complex Analytics
 - Data Ingest Engine
 - Prepare & Transform Accelerator
 - Query Engine Optimizer
 - Backup/Disaster Recovery
 - Industry Specific Use Cases

The logo for SQREAM, featuring the word "SQREAM" in a bold, italicized, sans-serif font. Below the text are five horizontal bars of increasing length, creating a sense of motion or a stylized underline.

Industry Case Studies



Global Electronics Manufacturer



Prominent Global Electronics manufacturer struggling with data disparity, unable to implement an AI platform

Limitation: Current infrastructure can't cope with scale; Outdated insights

Objective: Improve machines health

Results: Full data consolidation; 80% Reduced hardware maintenance cost; 50% to 90% improved factory yield

Legacy Infrastructure



Failed after 2 weeks

Filtered only 300M

Failed to achieve KPI

10 nodes

*Running on 300B records

Ingest 500B records

Filter 1B records

Ad-hoc analytical queries

10x footprint

SQREAM

Hitachi Vantara

39 minutes

3 minutes

58 minutes

1 server, 2 GPUs



Enterprise Semi-Conductor Manufacturer

US manufacturer had to support new factories; Failed to consolidate enterprise and FAB data

Limitation: Existing solution doesn't scale to PBs; Can't deal with outburst of loading jobs

Objective: Fulfill 'Factory of the Future' vision for all existing and new sites

Results: Better performance at scale; Consolidated data platform for all sites; Initiated ML platform project

Legacy Infrastructure



263 minutes

8% workloads faster

\$6.2M

26 nodes

263x Daily ETL task time

Spotfire workloads time

4x Licensing Cost

26x footprint

SQREAM

1 minute

38% workloads faster

\$1.5M

1 server, 2 GPUs



Thank you



Hitachi Vantara



@HitachiVantara



Hitachi Va

Thank You