IBM Fusion HCI

One Platform. Zero Compromises.

Highlights

Integrated by design: Pre-configured hardware and software stack optimized for OpenShift deployment.

Accelerate business growth with parallel data access and scalable data.

Simplify the management of an OpenShift solution.

Data-smart architecture: Transparent access to global data sources for hybrid cloud and AI workloads.

As organizations accelerate their shift to cloud-native architectures, the adoption of containers and Kubernetes continues to surge. A 2024 Red Hat report found that 67% of companies have delayed or slowed down application deployment due to Kubernetes security concerns — underscoring the critical need for robust, secure platforms.[1]

Meanwhile, nearly 48% of enterprises expect their Kubernetes clusters to grow by over 50% in 2025, driven by AI/ML workloads and hybrid cloud expansion.[2]

This momentum reflects a broader business imperative: the need for agility, consistency, and resilience across decentralized environments.

According to IDC, 41% of companies have already deployed containers, with 24% planning production deployments within the next year. Notably, 80% of container adopters are already running them in production — signaling a decisive shift toward cloud-native architectures.

IBM Fusion HCI

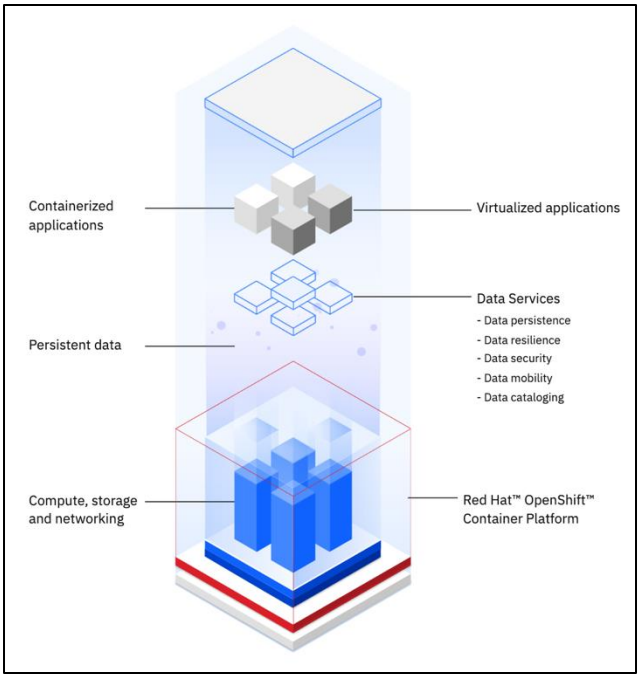
Accelerating Cloud-Native Transformation

As enterprises embrace container-native architectures, Red Hat® OpenShift® stands as a cornerstone for modern application deployment. IBM Fusion elevates this foundation — delivering a unified self-service platform for AI, VMs and Containers running OpenShift applications and watsonx.data in hybrid cloud environments.

Fusion HCI reimagines infrastructure with a hyperconverged approach: compute, storage, and networking unified into composable, software-defined building blocks. This integration simplifies operations, enhances scalability, and unlocks a seamless hybrid cloud experience.

Key features
Fusion HCI isn't just infrastructure — it's a strategic enabler for agility, growth, and innovation. It empowers teams to build, scale, and adapt with confidence in a world where speed and resilience define success.

- Pre-integrated and production-ready: Factory-built to run containerized workloads with speed and precision
- AI-optimized and data-centric: Designed to support demanding applications across edge, core, and cloud
- Unified management: A single dashboard to orchestrate OpenShift resources and global data services
- Accelerated modernization: Rapid deployment of OpenShift environments with enterprise-grade storage and observability
- Scalable architecture: Supports parallel data access and dynamic scaling across environments
- Simplified operations: Hyperconverged design reduces complexity in infrastructure and data layer management
- Enterprise-grade storage: Built-in services for mission-critical applications across edge, core, and cloud



Data architecture for IBM Fusion HCI

90%

Reduce OpenShift subscription costs by up to 90%.[3]

20%

With bare-metal OpenShift licensing get 20% more efficient use of infrastructure vs applications running on hypervisors, and eliminates 100% of hypervisor license costs.[3]

Software-Driven Infrastructure for OpenShift

Fusion HCI is a container-native platform optimized for Red Hat® OpenShift®, with integrated software that simplifies deployment, monitoring, and management across hybrid environments.

Unified Dashboard
Fusion HCI includes a centralized dashboard that gives administrators full visibility into both hardware and software resources within their OpenShift cluster. From a single interface, users can monitor CPU, memory, storage, and network utilization, track event status, and manage local services such as software upgrades and backup/restore operations. The dashboard also provides direct access to cloud management tools like OpenShift Advanced Cluster Management (ACM), streamlining hybrid operations and reducing context switching. Resource-level insights into server nodes, disks, and switches are readily available, enabling fast diagnostics and performance tuning.

Integrated Management
Fusion HCI's software stack is tightly integrated with its hardware components, allowing for optimized configuration and lifecycle management. The system is intelligently pre-configured on Day 1, and continues to monitor all components to ensure operational integrity. Built-in health checks and telemetry features, including IBM's Call Home service, automatically notify support teams of hardware failures or critical errors. Firmware updates are orchestrated through the same management layer, ensuring consistency and reducing downtime across the infrastructure.

Global Data Mesh
Fusion's architecture supports transparent global data access, allowing containerized applications to interact with remote datasets as if they were local. This abstraction enables seamless integration of S3 object storage (cloud or on-prem), NFS shares from vendors like Dell/EMC and NetApp, and any IBM Storage Scale-compatible system. Applications can access distributed data across geographies without refactoring, which simplifies development and accelerates deployment in hybrid and multi-cloud environments.

Hybrid Cloud Optimization
Fusion HCI helps organizations build a more agile and cost-effective hybrid cloud infrastructure. It reduces data costs through intelligent archiving, simplifies remote access and migration via transparent global visibility, and enables faster application recovery by eliminating the need for data movement. The platform also supports optimized data ingest and collaboration across teams and environments, making it ideal for AI workloads and distributed applications.

Resilient Backup & Recovery
IBM Fusion includes application-aware backup and recovery capabilities powered by IBM Storage Protect Plus. This containerized software provides consistent protection for OpenShift environments, and extends support to VMware and OpenShift VM workloads through a unified interface. By fusing storage services with platform-level orchestration, Fusion ensures data resiliency across critical applications, enabling IT operations to maintain continuity and compliance with minimal overhead.

IBM Fusion and watsonx.data are better together

The ultimate solution for resilient, high-performance AI deployments on OpenShift, tailored to meet your business needs swiftly and securely to accelerate your AI journey. IBM Fusion HCI with local caching capabilities and IBM watsonx.data accelerates remote S3 queries by up to 90X.[4]

watsonx.ai

Watsonx-in-a-Box with Fusion HCI

Fusion HCI provides a robust foundation for deploying watsonx workloads locally, enabling a “Watsonx-in-a-box” architecture that brings AI compute and data services directly to the edge, data center, or private cloud. This configuration supports watsonx.ai for model training and inference, watsonx.data for scalable lakehouse analytics, and watsonx.governance for managing AI lifecycle and compliance — all within a container-native, OpenShift-optimized environment.

Fusion HCI’s integrated hardware stack includes GPU servers configured in redundant pairs, ensuring high availability for accelerated AI workloads. These GPU nodes are tightly coupled with storage and compute servers that form a Scale ECE storage cluster.

The platform’s global data mesh allows watsonx applications to access distributed datasets — including S3 object stores, NFS shares, and IBM Storage Scale volumes — as if they were local. This eliminates the need for data duplication or migration, dramatically reducing latency and simplifying data orchestration across hybrid environments.

Fusion HCI also supports application-aware backup and recovery for watsonx workloads using IBM Storage Protect Plus, ensuring data resiliency and operational continuity. Combined with the unified dashboard and integrated observability, administrators can monitor GPU utilization, storage throughput, and AI job performance from a single pane of glass.

Whether deployed in a secure on-premises environment or extended to the edge, Fusion HCI enables enterprises to run watsonx workloads with full control, scalability, and compliance — making it an ideal platform for AI innovation in regulated or latency-sensitive industries.

Fusion HCI: Integrated, resilient, data-smart infrastructure built to simplify OpenShift and accelerate hybrid cloud.

Hardware Configuration

Fusion HCI is delivered as a fully integrated hardware and software solution. Clients are not required to spend time selecting or procuring hardware components to support the software stack. Instead, the IBM Fusion development team has pre-selected servers, switches, and rack enclosures that are specifically optimized to run the Fusion HCI software stack efficiently and reliably.

Fusion HCI is architected to eliminate single points of failure by incorporating redundant hardware across all critical components. This redundancy not only protects against hardware failures but also enables firmware updates to be performed without system downtime.

Components can be taken offline individually, updated, and restarted while the system remains operational.

Server Types
Fusion HCI includes three distinct server types, each serving a specific function within the system. (Refer to Tables 1 and 2 below for a breakdown of the components and roles associated with each server type.)

Storage and Compute Servers
Storage and compute (S&C) servers form the foundational layer of IBM Fusion HCI. Each deployment includes a minimum of three S&C servers, which are combined to form a Data Foundation storage cluster. Each S&C server is equipped with at least two storage drives, with the option to scale up to a maximum of ten drives per server, depending on workload requirements.

Management Network

The management network is built around a pair of 48-port, 1Gb Ethernet switches. The IMM port of every IBM Fusion HCI server is connected to the first of the management switches using CAT5e cables with RJ45 connectors.
The alternate IMM port, on either the LOM or an OCP adapter, is configured for all Fusion HCI servers and is connected to the second management switch.

These connections are also made using CAT5e cables with RJ45 connectors. This is all done so that there is redundancy to support management functions even if one of the management switches fails or one of the cables becomes disconnected.

Hardware Summary

Table of Specifications – Lenovo

Table 1

LENOVO	Storage & Compute Server	GPU Server	Service Node
CPU	2x 32-core Intel Xeon 4 Gold 6438N 2.0GHz 205W processors or 2x 16-core Intel Xeon 4 Gold 6426Y 2.5GHz 185W processors	2x AMD EPYC 4 9254 24C 200W 2.9GHz Processors	1x 8-core Intel Xeon Bronze 3408U 125W 1.8GHz Processor
Memory	256GB upgradeable to 512GB (32 core) 1024GB upgradeable to 2048GB (64 core)	768GB RAM	128GB RAM
OS Storage	2x 960GB M.2 NVMe OS drives (RAID 1)	2x 960GB M.2 NVMe OS drives (RAID 1)	2x 960GB M.2 NVMe OS drives (RAID 1)
Data Network	1x NVIDIA ConnectX-6 DX dual-port 100GbE PCIe NIC 1x NVIDIA ConnectX-6 LX dual-port 25GbE PCIe NIC	1x NVIDIA ConnectX-6 DX dual-port 100GbE PCIe NIC 1x NVIDIA ConnectX-6 LX dual-port 25GbE PCIe NIC	1x NVIDIA ConnectX-6 LX dual-port 25GbE PCIe NIC
Mgmt Network	Intel I350 1GbE RJ45 4-port OCP Ethernet Adapter	Intel I350 1GbE RJ45 4-port OCP Ethernet Adapter	Intel I350 1GbE RJ45 4-port OCP Ethernet Adapter
Data Storage	0-10x 7.68TB or 3.84TB PCIe Gen 5 NVMe drives		
GPU		1 to 8 NVIDIA H100 NVL or 1 to 8 NVIDIA L40S or 1 to 8 AMD MI210	

Table of Specifications – Dell

Table 2

DELL	Storage & Compute Server	GPU Server	Service Node
CPU	2x 32-core Intel Xeon 4 Gold 6438N 2.0GHz 205W processors or 2x 16-core Intel Xeon 4 Gold 6426Y 2.5GHz 185W processors	2x 64-core AMD EPYC 5 9575F 3.3GHz 400W processors with 256M cache	1x 8-core Intel Xeon Bronze 3408U 125W 1.8GHz Processor
Memory	256GB upgradeable to 512GB (32 core) 1024GB upgradeable to 2048GB (64 core)	1536GB RAM 2304GB RAM (8x H200 NVL)	128GB RAM
OS Storage	2x 960GB M.2 NVMe OS drives (RAID 1)	2x 960GB M.2 NVMe OS drives (RAID 1)	2x 960GB M.2 NVMe OS drives (RAID 1)
Data Network	1x NVIDIA ConnectX-6 DX dual-port 100GbE PCIe NIC	4x NVIDIA ConnectX-6 DX dual-port 100GbE PCIe NIC	1x NVIDIA ConnectX-6 LX dual-port 25GbE PCIe NIC
Mgmt Network	Intel I350 1GbE RJ45 4-port OCP NIC	1x Broadcom 5719 quad-port 1GbE PCIe NIC	Intel I350 1GbE RJ45 4-port OCP NIC
Data Storage	0-10x 7.68TB or 3.84TB PCIe Gen 5 NVMe drives		
GPU		4 or 8 NVIDIA H200 NVL or 4 or 8 NVIDIA L40S	

Usable Storage Summary

Storage disks
In this configuration, the failure of any two storage and compute (S&C) servers can be tolerated without losing any data.

Base configuration

The Base configuration includes *
- 42U rack
- Two Ethernet ToR switches (200GbE)
- Two Ethernet management switches (1Gb)
- Three 1U x86 storage/compute nodes
- Service Node server

**Additional configuration options available*

It is possible to expand Fusion HCI beyond the minimum of three S&C servers. Servers can be added to a maximum of 16. Each of the S&C servers added to Fusion HCI is also added to the storage cluster, increasing the total storage capacity.

Table of Specifications – Usable Storage Capacity
Table 3

Usable Storage Capacity in TB (Using DF 3-Way Copy)						
		# Storage / Compute Servers				
2	30*	40	50	60	70	81
4	60	81	101	121	141	162
6	91	121	152	182	212	243
8	121	162	202	243	283	324
10	152	202	253	304	354	405
Drive Capacity (TB)	7.68					

**Aproximate capacity values rounded down to the nearest TB.*

High speed data network

Fusion HCI has two physical networks defined within it: a high-speed network for use by the storage cluster and applications, and a management network that is used for controlling the servers and monitoring the health of the servers.

The high-speed network is built around a pair of 32-port, 200Gb Ethernet switches. The switches are configured together using MLAG to create a redundant pair. All of the S&C servers and the GPU servers have a 2-port, 100Gb Ethernet adapter.

One port on the adapter is connected to the first high-speed switch and the second port is connected to the second high-speed switch. This 100GbE connections are reserved for use by the Data Foundation storage cluster.

Table of Specifications – Ethernet Switches
Table 4

High-Speed Ethernet Switch	Management Ethernet Switch
2x NVIDIA/Mellanox SN3700V	2x NVIDIA SN2201
Configured as a redundant pair using MLAG	Configured as a redundant pair using MLAG
32x QSFP56 200GbE ports in each switch	48x 1GbE RJ-45 ports
All servers have 1x 100GbE connection and 1x25GbE connection to each switch	All IMMs connected to one switch, the alternate IMM to the other switch
Total of 4x 200GbE uplink customer data ports	Total of 4x QSFP28 100GbE ports
Switch throughput of 6.4Tb/s	Up to 5 PDUs connect to each switch
Cumulus Network OS	Cumulus Network OS
1U height	1U height

Why IBM?

IBM Fusion is a container-native global data platform designed to simplify hybrid cloud operations, accelerate AI readiness, and unify infrastructure management. Built to optimize Red Hat® OpenShift® environments, Fusion delivers integrated storage, observability, and data services that scale from edge to core to cloud. By fusing infrastructure and data into a single, intelligent platform, IBM Fusion enables organizations to build, deploy, and adapt with speed, confidence, and strategic clarity.

For more information

To learn more about IBM Fusion, contact your IBM representative or IBM Business Partner, or visit ibm.com/products/storage-fusion

1. "Kubernetes adoption, security, and market trends report 2024," June, 2024, red
2. "Kubernetes and Containerization Trends in 2025", March, 2025.
3. "Fusion HCI Hardware Component Overview", December, 2024
4. Redbbok, "Accelerating IBM watsonx.data with IBM Fusion HCI," March 2024

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
October 2025