MongoDB on Lenovo ThinkAgile HX650 V4 and FX650 V4

Solution Brief

Modern applications increasingly generate dynamic, semi-structured data: customer profiles, product catalogs, clickstream events, IoT telemetry, application logs, and operational metrics. These data sources rarely fit neatly into rigid relational schemas, and organizations are turning to document-based databases that evolve alongside the applications they support. MongoDB has become a leading operational database for these scenarios, combining a flexible document model with horizontal scalability, built-in replication, and native support for vector and full-text search that powers retrieval-augmented generation (RAG) and other AI applications.

At the same time, infrastructure teams are consolidating compute, storage, and virtualization onto hyperconverged platforms to simplify operations and reduce the number of systems they manage. The arrival of Intel Xeon 6 processors (code named "Granite Rapids") in the Lenovo ThinkAgile HX650 V4 and FX650 V4 gives that consolidated infrastructure a substantial increase in core count and memory capacity, precisely the resources MongoDB needs to handle insert-heavy and mixed read/write workloads at scale.

Figure 1. ThinkAgile HX650 V4 (top) and HX630 V4 (bottom) for Nutanix hyperconverged infrastructure

Business Challenge

Choosing infrastructure for MongoDB is rarely just a sizing exercise. IT teams evaluating a platform for document databases face several intersecting pressures:

- **Unpredictable performance at scale:** as document sizes and concurrent client counts grow, throughput and latency do not scale linearly, making it hard to size infrastructure with confidence for insert-heavy ingestion or mixed transactional workloads.
- **Fragmented AI infrastructure:** adding retrieval-augmented generation and vector search often means standing up a separate vector database and search engine, increasing operational complexity and latency between systems.
- **Siloed compute and storage:** traditional three-tier architectures separate compute, storage, and virtualization, adding management overhead and slowing the ability to scale a MongoDB cluster on demand.
- **Locked-in hypervisor choices:** many hyperconverged platforms tie an organization to a single hypervisor, limiting flexibility as licensing, cost, and operational requirements change over time.
- **Lifecycle and data protection burden:** provisioning, cloning, backing up, and patching database instances consistently across a growing MongoDB estate consumes scarce administrator time.

These pressures push organizations to look for infrastructure that is validated for MongoDB, flexible in its software stack, and simple to operate at scale.

Solution

Lenovo addresses these challenges with the ThinkAgile HX650 V4 and FX650 V4 hyperconverged systems, 2-socket, 2U platforms (also available as 1U ThinkAgile HX630 V4 and FX630 V4 systems) built on Intel Xeon 6 processors. Both systems are fully validated and integrated by Lenovo, certified and preloaded with licensed Nutanix software, and designed to run the Nutanix Cloud Platform (NCP). NCP unifies compute, storage, and virtualization into a single hyperconverged layer that provides enterprise storage, integrated data protection, built-in resilience, unified management, advanced analytics, and end-to-end security for mission-critical workloads such as MongoDB.

Figure 2. ThinkAgile FX650 V4 (top) and FX630 V4 (bottom) designed for flexible hyperconverged infrastructure

ThinkAgile FX650 V4 adds a further layer of flexibility: organizations can move between hypervisors of their choice while keeping the same integrated compute, storage, and management stack, future-proofing their investment against hypervisor licensing or strategy changes.

MongoDB runs as a virtual machine on the Nutanix AHV hypervisor, using the WiredTiger storage engine, and can take full advantage of the high core counts and large memory capacity that Intel Xeon 6 processors and dense NVMe storage provide. For teams that want to standardize database operations, Nutanix Database Service (NDB) registers, provisions, clones, and administers MongoDB databases across one or more Nutanix clusters in single node or multi-node configurations; MongoDB can also be deployed and scaled manually outside of NDB. Lenovo XClarity One provides ongoing systems management across the cluster, backed by Lenovo's broad selection of processor, memory, and drive options to match the deployment to the workload.

Use Cases

Lenovo ThinkAgile HX650 V4 and FX650 V4 support MongoDB across a range of enterprise scenarios, from consolidating private cloud infrastructure to running the ingest-heavy and AI-driven workloads that modern applications generate. The following use cases show how organizations put the platform's performance, flexibility, and validated benchmark results to work.

- **Private cloud consolidation and hypervisor flexibility**

  Enterprises consolidating MongoDB alongside other virtualized workloads onto a single private cloud platform deploy ThinkAgile HX650 V4 or FX650 V4 running the Nutanix Cloud Platform. The hyperconverged architecture combines compute, storage, and virtualization into one manageable system, reducing the number of silos administrators must maintain. Organizations that want the option to change hypervisors over time choose ThinkAgile FX650 V4, which lets them move between hypervisors of their choice without re-architecting the underlying compute, storage, or management stack, protecting their MongoDB investment as strategy or licensing needs evolve.

- **Mixed read/write transactional applications**

  Applications such as order processing, customer profile updates, and catalog management generate a steady mix of reads and writes against MongoDB. Benchmark testing on ThinkAgile HX650 V4 showed the platform sustaining strong throughput across both balanced 50/50 and read-heavy 80/20 read/write workloads, with the 80/20 profile remaining especially resilient even as document size grew to 8,192 bytes. For teams running transactional, customer-facing MongoDB applications, this means predictable performance across the document sizes and concurrency levels typical of production traffic.

- **High-volume operational data**

  IoT telemetry, clickstream events, application logs, and operational metrics arrive continuously and at high volume, making insert throughput the critical metric. Testing showed ThinkAgile HX650 V4 sustaining over 100,000 load operations per second for small and medium-size documents (256 bytes and 2,000 bytes) across 64 to 128 concurrent client threads. The platform's high core-count Intel Xeon 6 processors and NVMe-based local storage let MongoDB absorb ingest-heavy traffic from streaming analytics, IoT platforms, and other large-scale operational systems without sacrificing responsiveness.

- **AI-ready search and retrieval-augmented generation**

  MongoDB's native support for vector search combined with full-text search lets organizations retrieve the most relevant results to power AI applications and agentic systems that use retrieval-augmented generation (RAG). Running this capability on ThinkAgile HX650 V4 or FX650 V4 keeps vector, text, and operational data together in the same self-managed, on-premises MongoDB deployment, eliminating the need for a separate external vector database or search engine and reducing the complexity and latency of building RAG pipelines directly where enterprise data already lives.

Workloads

ThinkAgile HX650 V4 and FX650 V4 are validated for the workload types that make up most production MongoDB deployments, from high-volume data ingestion to AI-driven search. Rather than testing a single synthetic benchmark point, Lenovo varied document size, client concurrency, and read/write mix to mirror the range of traffic patterns these workloads actually generate. The table below summarizes the representative workload profiles and how each one performed.

Table 1. Representative MongoDB workloads on ThinkAgile HX650 V4 and FX650 V4

Workload Profile	Represents	Validated Performance
High-volume data ingestion	IoT telemetry, event streaming, clickstream and log data, initial bulk loads (100% write)	Up to 113,868 operations per second, low and predictable latency
Balanced transactional applications	Order processing, customer profile and catalog updates (50/50 read/write)	Up to 107,890 operations per second; sub-1 ms latency for small and medium documents
Read-heavy operational and reporting workloads	Customer-facing lookups, dashboards, reporting (80/20 read/write)	Peak of 116,149 operations per second, the strongest result of any workload tested
Rich, content-heavy documents	Order details, IoT payloads, and nested records up to 8,192 bytes	Solid sustained throughput, with a predictable latency trade-off versus smaller documents
AI-native search and retrieval-augmented generation	Vector and full-text search combined for RAG and agentic applications	Runs in the same self-managed MongoDB deployment; no separate vector database required
Private cloud consolidation	Multiple MongoDB clusters and other virtualized workloads sharing infrastructure	Nutanix Cloud Platform HCI, with FX650 V4 adding hypervisor flexibility

Across every workload profile tested, ThinkAgile HX650 V4 sustained strong throughput with consistently low latency, giving IT teams a proven performance envelope to size against rather than a single best-case number. Full lab configuration and measured results follow below.

Test configuration

Lenovo validated this configuration on a four-node Nutanix cluster built from Lenovo ThinkAgile HX650 V4 systems, with MongoDB deployed as a single, standalone instance, not a replica set or shared cluster, inside one large virtual machine sized to take advantage of the platform's core count and memory capacity. Across every workload tested, CPU utilization stayed below 50% while memory was the more constrained resource, so production deployments should size CPU and memory independently to match their own workload rather than treating this lab configuration as a fixed template.

The following table lists the configuration of each ThinkAgile HX650 V4 host.

Table 2. Lenovo ThinkAgile HX650 V4 hardware configuration (per host, four hosts total)

Item	Description
Server platform	4 x Lenovo ThinkAgile HX650 V4
Processor per host	2 x Intel Xeon 6767P, 64 cores, 350 W, 2.4 GHz (128 physical cores per host)
Memory	1 TB per host (16 x ThinkSystem 64 GB TruDDR5 6400 MHz RDIMM)
Storage	2 x ThinkSystem M.2 7450 PRO 480 GB NVMe SSD; 8 x ThinkSystem 2.5" U.2 PM9D3a 1.92 TB NVMe SSD
Network	2 x Mellanox ConnectX-6 Lx 10/25GbE SFP28 2-port adapters
Nutanix software	AHV 10.3.0.1, AOS 7.3.0.5
CVM configuration	16 vCPUs, 64 GB RAM

The following table lists the MongoDB virtual machine configuration.

Table 3. MongoDB virtual machine configuration

Item	Description
Database	MongoDB 7.0.3
Deployment topology	Standalone instance (not a replica set or sharded cluster)
Operating system	Ubuntu 22.04.5 LTS
VM sizing	240 vCPUs, 384 GB memory, 600 GB disk, NUMA enabled (4 NUMA nodes, about 60 CPUs and 96 GB memory each)
Storage engine	WiredTiger, 64 GB cache
Data path	/data/mongodb, ext4 filesystem
Primary virtual disk	/dev/sda, 600 GB
Port	27017
Observed resource utilization	CPU below 50% throughout testing; workloads were memory-bound

Test results

Testing used the Yahoo Cloud Serving Benchmark (YCSB) against a single, standalone MongoDB instance, using a 10 million record, 10 million operation dataset across three document sizes and three client concurrency levels.

The following table shows the measured throughput for every combination tested.

Table 4. Throughput results

Document size	Threads	Load (ops/sec)	50/50 Read/Write (ops/sec)	80/20 Read/Write (ops/sec)
256 bytes	64	113,868	101,690	116,149
256 bytes	96	112,165	107,890	108,226
256 bytes	128	109,902	102,529	108,980
2,000 bytes	64	105,069	96,462	102,298
2,000 bytes	96	99,693	97,011	110,568
2,000 bytes	128	94,767	80,146	80,499
8,192 bytes	64	70,898	45,777	87,659
8,192 bytes	96	73,254	53,671	95,160
8,192 bytes	128	59,079	58,305	61,658

The following table shows the average latency for every combination tested.

Table 5. Latency results

Document size	Threads	Load (ms)	50/50 Read (ms)	50/50 Update (ms)	80/20 Read (ms)	80/20 Update (ms)
256 bytes	64	0.551	0.611	0.630	0.535	0.570
256 bytes	96	0.842	0.870	0.887	0.871	0.891
256 bytes	128	1.150	1.222	1.238	1.157	1.167
2,000 bytes	64	0.602	0.613	0.693	0.602	0.667
2,000 bytes	96	0.952	0.909	1.043	0.845	0.908
2,000 bytes	128	1.336	1.548	1.612	1.555	1.626
8,192 bytes	64	0.894	0.863	1.910	0.696	0.821
8,192 bytes	96	1.299	1.062	2.484	0.949	1.193
8,192 bytes	128	2.148	1.970	2.382	2.031	2.159

Business Outcome

Across all three workload profiles, ThinkAgile HX650 V4 delivered strong throughput and stable latency. Peak throughput reached 116,149 operations per second on the 80/20 read/write workload with 256-byte documents at 64 threads, and load throughput exceeded 100,000 operations per second for both 256-byte and 2,000-byte documents. Latency stayed low and predictable for small and medium documents, typically under 1 millisecond at 64 and 96 threads.

Larger, 8,192-byte documents showed the expected trade-off: the balanced 50/50 workload saw average update latency rise to 2.484 ms at 96 threads, while the read-heavy 80/20 workload stayed far more resilient, with both read and update latency noticeably lower at the same document size and thread count. That resilience means organizations running read-dominant MongoDB applications can expect the platform to hold its performance even as payloads grow, while write-heavy applications benefit most from keeping documents compact.

Testing also showed these workloads to be memory-bound rather than CPU-bound: CPU utilization stayed below 50% throughout, even at the highest concurrency levels, on a standalone MongoDB instance rather than a replica set or shared cluster. Organizations planning a production deployment should size CPUs and memory independently to match their own workload profile, rather than treating this lab configuration as a fixed template.

Hardware Bill of Materials

The feature codes below are the exact configuration used for this validation, for one Lenovo ThinkAgile HX650 V4 server node; multiply quantities by four for the full cluster. Use them as a starting point when configuring an order with your Lenovo representative or Lenovo Business Partner.

Table 6. Hardware bill of materials, per ThinkAgile HX650 V4 node

Feature code	Description	Quantity
7DG4CTO1WW	Server: Lenovo ThinkAgile HX650 V4 Hyperconverged System	1
C6TQ	ThinkAgile HX650 V4 Base	1
B15S	Nutanix Software Stack on Nutanix AHV	1
BVKV	Nutanix Cloud Platform (NCP) Pro Software License with Mission Critical Support	1
C5QY	Intel Xeon 6767P 64C 350W 2.4GHz Processor	2
B15S	ThinkSystem 64GB TruDDR5 6400MHz (2Rx4) RDIMM	16
C26V	ThinkSystem M.2 RAID B545i-2i SATA/NVMe Adapter	1
C46P	ThinkSystem 2U V4 8x2.5" NVMe Backplane	2
B0SW	Nutanix Flash Node Config	1
C2BR	ThinkSystem 2.5" U.3 7500 PRO 1.92TB Read Intensive NVMe PCIe 4.0 x4 HS SSD	6
BKSR	ThinkSystem M.2 7450 PRO 960GB Read Intensive NVMe PCIe 4.0 x4 NHS SSD	2
BN2T	ThinkSystem Broadcom 57414 10/25GbE SFP28 2-Port OCP Ethernet Adapter	1
C0U3	ThinkSystem 2000W 230V Titanium CRPS Premium Hot-Swap Power Supply	2
6400	2.8m, 13A/100-250V, C13 to C14 Jumper Cord	2
C2DJ	ThinkSystem Advanced Toolless Slide Rail Kit V4	1
C3RG	ThinkSystem SR650 V4 Left Rack Latch with USB/MiniDP	1
C3RD	ThinkSystem 2U 6056 20K Performance Fan Module	6
SCJC	XClarity One - Managed Device, Per Endpoint w/1 Yr SW S&S	1

Conclusion

MongoDB running on Lenovo ThinkAgile HX650 V4 and FX650 V4 with Nutanix software delivers strong throughput and consistently low latency across the document sizes, concurrency levels, and workload mixes that enterprise applications actually generate. The platform combines validated, integrated hardware and firmware with licensed Nutanix Cloud Platform software, giving organizations a single hyperconverged foundation for private cloud, AI/ML, and analytics workloads built on MongoDB. Whether consolidating infrastructure, scaling operational data ingestion, running mixed transactional applications, or building retrieval-augmented generation pipelines, ThinkAgile HX650 V4 and FX650 V4 give MongoDB the scalable, predictable performance modern data platforms require.

For More Information

To learn more about MongoDB on Lenovo ThinkAgile HX650 V4 / FX650 V4, contact your Lenovo representative or Lenovo Business Partner, or visit the resources below.

References:

- Running MongoDB on Lenovo ThinkAgile HX V4 and FX V4 Servers with Intel Xeon 6 Processors (LP2412):
  https://lenovopress.lenovo.com/LP2412
- Lenovo ThinkAgile HX Series:
  https://www.lenovo.com/us/en/servers-storage/thinkagile/hx-series/
- Lenovo ThinkAgile FX Series:
  https://www.lenovo.com/us/en/servers-storage/thinkagile/fx-series/
- MongoDB:
  https://www.mongodb.com/
- Nutanix Cloud Platform:
  https://www.nutanix.com/products/cloud-platform
- Multi-Vendor Hyperconverged Infrastructure with Lenovo ThinkAgile FX V4 (LP2520):
  https://lenovopress.lenovo.com/lp2520-multi-vendor-hyperconverged-infrastructure-with-lenovo-thinkagile-fx-v4

Authors

Chandrakandh Mouleeswaran is a Solution Architect with 18+ years of experience in software development, performance testing and engineering, having worked on designing and architecting many scalable enterprise applications. He has spent a decade in technical enablement and partner solution development for VMware, Nutanix, Oracle and other ISVs across industries and technologies. He specializes in architecting infrastructure solutions for virtualization, VDI, database, cloud, data science, AI/ML solutions and various enterprise workloads.

Cristian Ghetau is an Advisory Engineer for Lenovo in Romania and has experience in Cloud Infrastructure technologies. He has had more than 13 years of experience working with virtual environments from VMware, Microsoft, Oracle, and Linux.

Chris Honoré is a Solutions Product Manager at Lenovo with deep expertise in datacenter products and solution offerings. He has a strong background in consulting and solution development, helping customers design and support on-premises and hybrid environments. Chris has spent the past 15 years with IBM and Lenovo, specializing in x86 server and data center solutions. Prior to that, he built two decades of experience in the telecommunications industry, serving in both technical and business leadership roles.

Related product families

Product families related to this document are the following:

- ThinkAgile FX Series
- ThinkAgile HX Series for Nutanix

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION ”AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document, LP2524, was created or updated on September 10, 2026.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2524
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2524.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:
Lenovo®
ThinkAgile®
ThinkSystem®
XClarity®

The following terms are trademarks of other companies:

Intel®, the Intel logo and Xeon® are trademarks of Intel Corporation or its subsidiaries.

Other company, product, or service names may be trademarks or service marks of others.