Running MongoDB on Lenovo ThinkAgile HX V4 and FX V4 Servers with Intel Xeon 6 Processors

Planning / Implementation

A Modern Data Platform for Private Cloud, AI/ML, and Analytics

Today's enterprise applications increasingly rely on dynamic, fast changing, and diverse data sources from customer profiles and product catalogs to clickstream events, IoT telemetry, application logs, and operational metrics. These workloads rarely fit neatly into rigid relational schemas. As a result, organizations are turning to document-based data models that align more naturally with the way modern applications evolve. MongoDB has emerged as a leading operational database platform for these scenarios because it enables agility, scalability, and simplicity across the full application lifecycle. MongoDB's document data model is flexible, semi-structured or hierarchical data model to support rapid application development cycle without constant schema redesign and refactoring resulting faster time to market. MongoDB's horizontal scalability across clusters, high availability with built in replication and resilience for mission critical workload makes ideal for ingest-heavy environments such as IoT platforms, streaming analytics, and large-scale operational systems. MongoDB as a vector database combines vector search with full text search to retrieve most relevant results to power AI applications and agentic systems with retrieval augmented generation (RAG).

Lenovo ThinkAgile HX V4 and FX V4 with Intel Xeon 6 Processors

The Lenovo ThinkAgile HX650 V4 and FX650 V4 are 2-socket, 2U systems and The Lenovo ThinkAgile HX630 V4 and FX630 V4 are 2-socket, 1U systems. These hyperconverged systems feature the 6th Generation Intel Xeon Scalable processors (formerly code named "Granite Rapids"). ThinkAgile HX V4 and FX V4 hyperconverged systems are designed for deploying industry-leading hyperconvergence software from Nutanix on Lenovo enterprise platforms. The Nutanix Cloud Platform (NCP) delivers cloud-scale advantages to enterprise applications and databases by combining ThinkAgile HX V4 and FX V4 and public cloud infrastructure. It provides robust enterprise storage, integrated data protection, built-in resilience, unified management, advanced analytics, and end-to-end security and consistent performance for mission-critical workloads.

ThinkAgile FX offers a unique, industry first flexibility for software-defined approach to hyper convergence, leveraging the ability to move between hypervisors of your choice to deliver compute, storage and management in a tightly integrated software stack and future-proof your investment with seamless HCI software transitions.

Combining performance and flexibility, the ThinkAgile HX V4 and FX V4 Hyperconverged systems deliver fully validated and integrated Lenovo hardware and firmware, certified and preloaded with licensed Nutanix software. The systems offer a broad selection of processors, memory and drives, and offers high performance features that industries such as finance, healthcare and telco need. Outstanding reliability, availability, and serviceability (RAS) and high-efficiency design can improve your business environment and can help save operational costs.

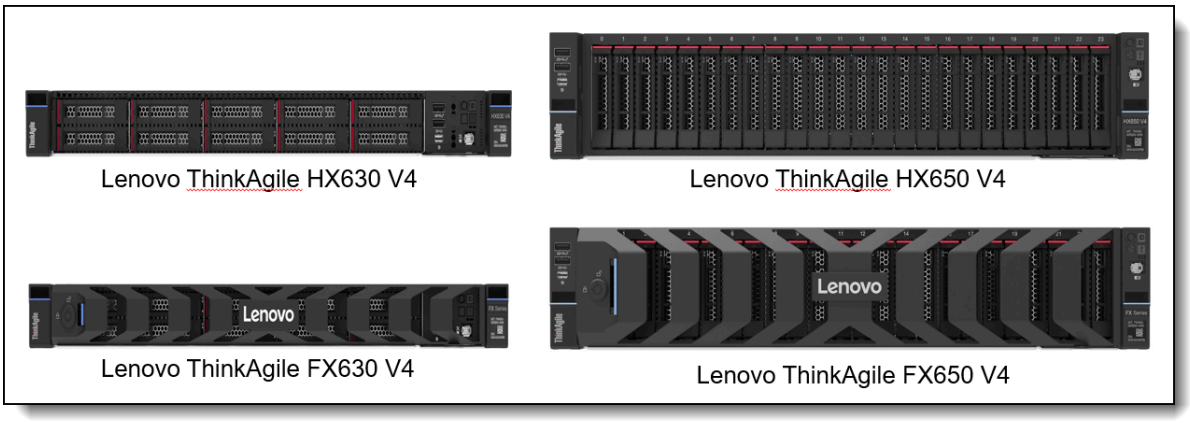

Figure 1. Lenovo ThinkAgile HX and FX V4 Systems

Lenovo ThinkAgile HX V4 and FX V4 Hyperconverged systems running Nutanix software provide a robust foundation for MongoDB by combining compute, storage, and virtualization in a unified hyperconverged platform. This architecture supports simplified operations, efficient scaling, and resilient infrastructure, while allowing MongoDB to take advantage of high core-count processors and large memory capacity for strong performance in insert-heavy and mixed read/write workloads.

MongoDB Features

Transactions, Analytics and Vector database
MongoDB's combination of flexible modeling, distributed architecture, and AI-ready search and vector capabilities makes it a powerful choice for enterprises modernizing their private cloud and data platforms. MongoDB is a Document-based NoSQL database that uses BSON documents to map and store dynamic and semi-structured data which enables to build adaptable applications without rigid relational schemas.

MongoDB architecture consists of:

- **mongod** – primary database engine responsible for storage and querying
- **mongos** – a query router for sharded clusters, routing client requests to appropriate shards
- **Config servers (CSRS)** – store cluster metadata and sharding configuration

Replica Sets ensure high availability and fault tolerance by maintaining multiple copies of data across nodes. Primary node handles writes; secondary nodes replicate via the oplog and supports globally distributed deployments.

Sharding enables horizontal scaling across many machines for large datasets and high-throughput workloads by using shards (Replica set), partitioned data subsets.
MongoDB's aggregation framework supports Time-series optimizations, complex transformations and real time analytics.

MongoDB Ops Manager: The "Command Center" for enterprise deployments. It automates backups, point-in-time recovery, and rolling upgrades.

MongoDB supports AI-native capabilities bringing full-text, semantic, and vector search to self managed and on prem deployment to build RAG (Retrieval Augmented Generation) systems directly where your data lives and eliminate external vector DBs and search engines, reducing complexity and latency.

MongoDB provides 8.2 provide significant performance improvement for unindexed queries and time series bulk inserts. Queryable Encryption enhances secure operations without compromising query expressiveness.

Nutanix Database Service (NDB) enables you to register, provision, clone, and administer all MongoDB databases on one or more Nutanix clusters and it supports single node and multiple node configurations. The MongoDB databases can be deployed manually without NDB and scaled as well.

YCSB overview
Yahoo Cloud Serving Benchmark (YCSB) is a benchmarking framework that is used to evaluate NoSQL and cloud-serving databases. It is designed to simulate application-access patterns and allows for a repeatable means of measuring latency and the throughput of a database under various load conditions.

There are two phases for the YCSB benchmarks:

- **Load** - the dataset used during testing is loaded into the database.
- **Run** - transactional workload (either read, update, insert or scan) defined by the mix of transactions is executed against the database. This makes it possible to test both the pure ingest of data as well as operational workloads randomly using a mixture of transaction types.

In this study, YCSB was used to evaluate MongoDB using only insert workloads and mixed read-write workloads while varying the size of the documents inserted into the database. By changing the size of the documents being added to the database and keeping the client-side concurrency the same, the benchmark demonstrates how the architecture of the database responds to different payload sizes and how throughput and latency react to different patterns of workloads.

Configurations

Hardware Configuration

The Nutanix cluster was built on four Lenovo ThinkAgile HX650 V4 systems. Each host was equipped with two Intel Xeon 6767P processors, with 64 cores per processor, for a total of 128 physical cores per host. Each host also included 1 TB of memory, NVMe-based local storage, and 10/25GbE network connectivity. This provided a dense computing and storage platform for the MongoDB virtual machine and the benchmark client environment.

Table 1. Lenovo ThinkAgile HX650 V4 configuration

Item	Description
Server platform	4 x Lenovo ThinkAgile HX650 V4
Processor per host	2 × Intel Xeon 6767P 64C 350W 2.4GHz Processor
Memory	1 TB (16 x ThinkSystem 64GB TruDDR5 6400MHz (2Rx4) RDIMM)
Storage	• 2 x ThinkSystem M.2 7450 PRO 480GB Read Intensive NVMe PCIe 4.0 x4 NHS SSD • 8 x ThinkSystem 2.5" U.2 PM9D3a 1.92TB Read Intensive NVMe PCIe 5.0 x4 HS SSD
Network	2 × Mellanox ConnectX-6 Lx 10/25GbE SFP28 2-port PCIe Ethernet Adapter
Nutanix Software	AHV 10.3.0.1, AOS 7.3.0.5
CVM configuration	16 VCPUs, 64 GB RAM

MongoDB virtual machine configuration

MongoDB was deployed in a virtual machine running Ubuntu 22.04.5 LTS. The guest operating system exposed the VM as a four-NUMA-node system. Each NUMA node contained 60 CPUs and approximately 96 GB of memory, for a total guest size of 240 vCPUs and about 384 GB of RAM.

Table 2. MongoDB configuration

Item	Description
Database	MongoDB 7.0.3
OS	Ubuntu 22.04.5 LTS
VM Configuration	240 vCPUs, 384 GM Memory, 600GB disk, NUMA Enabled
Storage engine	WiredTiger
WiredTiger cache size	64 GB
Data path	/data/mongodb
Port	27017
MongoDB data filesystem	ext4
Primary virtual disk	/dev/sda, 600 GB

Testing

Testing Methodology

The goal of testing was to evaluate how changing document sizes, and the number of concurrent clients would affect throughput and latency.
For this testing, the YCSB dataset contained a record count of 10,000,000, operation count of 10,000,000 and a field count of 1. The document sizes were chosen to represent three practical application profiles, respectively, rather than a single synthetic point.

- **256 bytes** represents small records, such as lightweight metadata, compact session objects, short key-value style application data, or simple event entries.
- **2000 bytes** represents a medium-size operational document, which is a realistic size for many business applications using MongoDB for customer, catalog, profile, or transactional data.
- **8192 bytes** represents a larger document profile, useful for understanding how the platform behaves when records contain richer payloads, embedded structures, or more application context.

The testing performed with 64, 96, and 128 concurrent threads to stress the database and observe not only the effects of document size, but also how well the platform performed as client concurrency is increased.

The following workload profiles tested, as follows:

Table 3. Workload profiles

Insert	**100% write**
50/50 READ/WRITE	Reasonably balanced mix of transactional activity
80/20 READ/WRITE	Read intensive workloads

The data was loaded into the database at the selected document sizes at each client thread level for insert testing. The workload data was loaded first into the database and then tested with concurrent threads for the mixed workloads.

Test Results

The MongoDB YCSB results collected on the Nutanix-based Lenovo ThinkAgile HX650 V4 environment are presented below.

Throughput Performance

The following table summarizes throughput results for the three workload profiles tested. This dataset gives a very clear look into how MongoDB performs with increasing concurrency and IO sizes.

Nutanix platform performs strongest overall with the 256 B document profile. The highest values in the table are all in that range, including the peak result of 116,149 ops/sec for the 80/20 workload at 64 threads. That suggests the platform handles smaller MongoDB records very efficiently.

The best thread count is not the same for every workload. For load, the highest throughput appears at 64 threads for 256 B and 2000 B, while 8192 B peaks at 96 threads. For 50/50 read/write, 96 threads workload is generally the strongest point. For 80/20 read/write, the best point varies by document size. This indicates the platform does not scale linearly with thread count and that the optimal operating point depends on both payload size and workload mix.

Larger documents affect the workloads differently. In the load phase, throughput drops as document size grows, which is expected. In the 50/50 read/write case, the penalty is much stronger, especially for 8192 B documents. In contrast, the 80/20 workload remains much more resilient, even with 8192 B records, where throughput stays relatively high compared with the 50/50 case. That suggests the Nutanix-based environment is particularly comfortable with read-dominant MongoDB activity, even when document size increases.

Table 4. Throughput results

Document size	Threads #	Load (ops/sec)	50 / 50 Read/Write (ops/sec)	80 / 20 Read/Write (ops/sec)
256 B	64	113,868	101,690	116,149
256 B	96	112,165	107,890	108,226
256 B	128	109,902	102,529	108,980
2000 B	64	105,069	96,462	102,298
2000 B	96	99,693	97,011	110,568
2000 B	128	94,767	80,146	80,499
8192 B	64	70,898	45,777	87,659
8192 B	96	73,254	53,671	95,160
8192 B	128	59,079	58,305	61,658

Latency Performance

The latencies generally increase with both document size and thread count. That trend is especially visible in the load phase, where average latency rises steadily from 256 B to 8192 B and from 64 to 128 threads.

For small and medium-size documents the latencies are stable for most workloads, especially at 64 and 96 threads. That suggests the Nutanix-based environment handles small and medium MongoDB documents efficiently without a large latency penalty.

For large-size documents we observe an increase in latencies, especially in the 50/50 read/write workload. For example, at 96 threads, 8192 B gives about 1.062 ms average read latency but 2.484 ms average update latency. This indicates that larger documents impose a heavier penalty on the write/update path than on reads.

Finally, the 80/20 read/write workload is generally more latency-friendly than the 50/50 workload, especially for larger documents. For 8192 B at 64 and 96 threads, both read and update latency are noticeably lower in 80/20 than in 50/50. That aligns with the throughput results and suggests the platform is more comfortable with read-dominant access patterns than with evenly balanced mixed activity.

The following table shows the latency results.

Table 5. Latency results

Document size	Threads #	Load (ms)	50 / 50 Read/Write (ms)	50 / 50 Read/Write (ms)	80 / 20 Read/Write (ms)	80 / 20 Read/Write (ms)
256 B	64	0.551	0.611	0.630	0.535	0.570
256 B	96	0.842	0.870	0.887	0.871	0.891
256 B	128	1.150	1.222	1.238	1.157	1.167
2000 B	64	0.602	0.613	0.693	0.602	0.667
2000 B	96	0.952	0.909	1.043	0.845	0.908
2000 B	128	1.336	1.548	1.612	1.555	1.626
8192 B	64	0.894	0.863	1.910	0.696	0.821
8192 B	96	1.299	1.062	2.484	0.949	1.193
8192 B	128	2.148	1.970	2.382	2.031	2.159

Conclusion

The results indicate that MongoDB running on Lenovo ThinkAgile HX V4 systems with Nutanix software delivers strong throughput and consistently low latency across a range of realistic workload profiles. Peak throughput was highest with smaller documents, while medium-size and larger documents also maintained solid performance, showing that the platform can support both lightweight transactional records and richer application data effectively.

Overall, Lenovo ThinkAgile HX V4 with Nutanix is a strong platform for MongoDB deployments that need scalable performance, predictable responsiveness, and support for mixed read/write application behavior. The combination of high throughput, stable latency, and broad workload coverage makes it well suited for modern enterprise database environments .

Bill of Materials

The following table lists the feature codes for the lab configuration.

Table 6. ThinkAgile VX650 V4

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
BVKV	Nutanix Cloud Platform (NCP) Pro Software License with Mission Critical Support	1
SCJC	XClarity One - Managed Device, Per Endpoint w/1 Yr SW S&S	1

Author

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, Linux.

Related product families

Product families related to this document are the following:

- ThinkAgile FX Series
- ThinkAgile VX Series for VMware
- VMware vSphere

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, LP2412, was created or updated on March 30, 2026.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2412
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2412.

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.