Disaggregated Solutions for VMware: Flexible Infrastructure for AI and Modern Workloads

Solution Brief

Letting Compute and Storage Move Independently

Infrastructure has traditionally been designed with the assumption that compute and storage scale together. Early in a deployment, this works well—workloads are predictable and systems are sized to maintain balance. Over time, however, that balance shifts as data grows faster than expected, new workload types emerge, and performance requirements diverge. AI initiatives further accelerate this trend. Training, retrieval-augmented generation (RAG), and inferencing workloads often require rapid expansion of GPU-accelerated compute resources while storage growth follows a different lifecycle. As these demands evolve independently, traditional scale-together infrastructure models become increasingly inefficient.

When this happens, scaling becomes inefficient. Expanding storage often requires adding full nodes, bringing unnecessary compute, cost, and operational overhead. What began as a simple model becomes increasingly rigid as workloads evolve.

Disaggregation addresses this challenge by allowing compute and storage to scale independently. With ThinkAgile VX, this flexibility is introduced without changing the VMware operational model administrators rely on today.

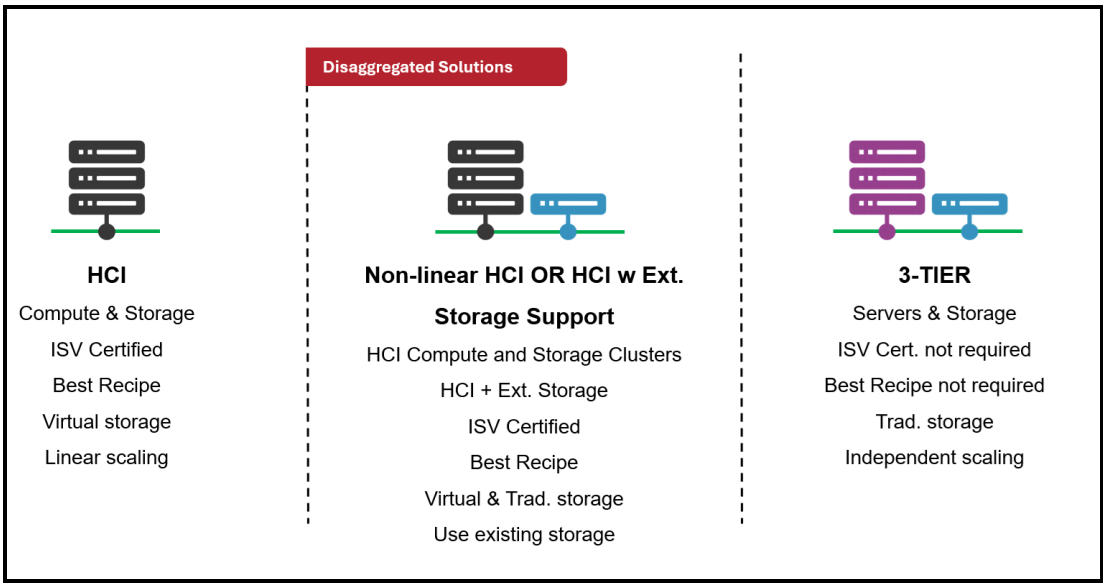

Figure 1. Disaggregated Solutions for VMware bring the best of HCI and 3-Tier Architecture

AI highlights why disaggregation matters. Training environments may require bursts of GPU-enabled compute capacity, while inference platforms, vector databases, and enterprise data repositories often grow on entirely different timelines. Rather than forcing organizations to scale all resources together, disaggregated VMware architectures allow infrastructure investments to align more closely with actual workload demand.

One Architecture, Three Deployment Models

VMware disaggregation is not a new architecture, but a flexible application of the same architecture. ThinkAgile VX supports three deployment models - traditional HCI, vSAN storage clusters, and external storage integration - within a consistent operational framework. These models can coexist, allowing infrastructure to align more closely with workload needs.

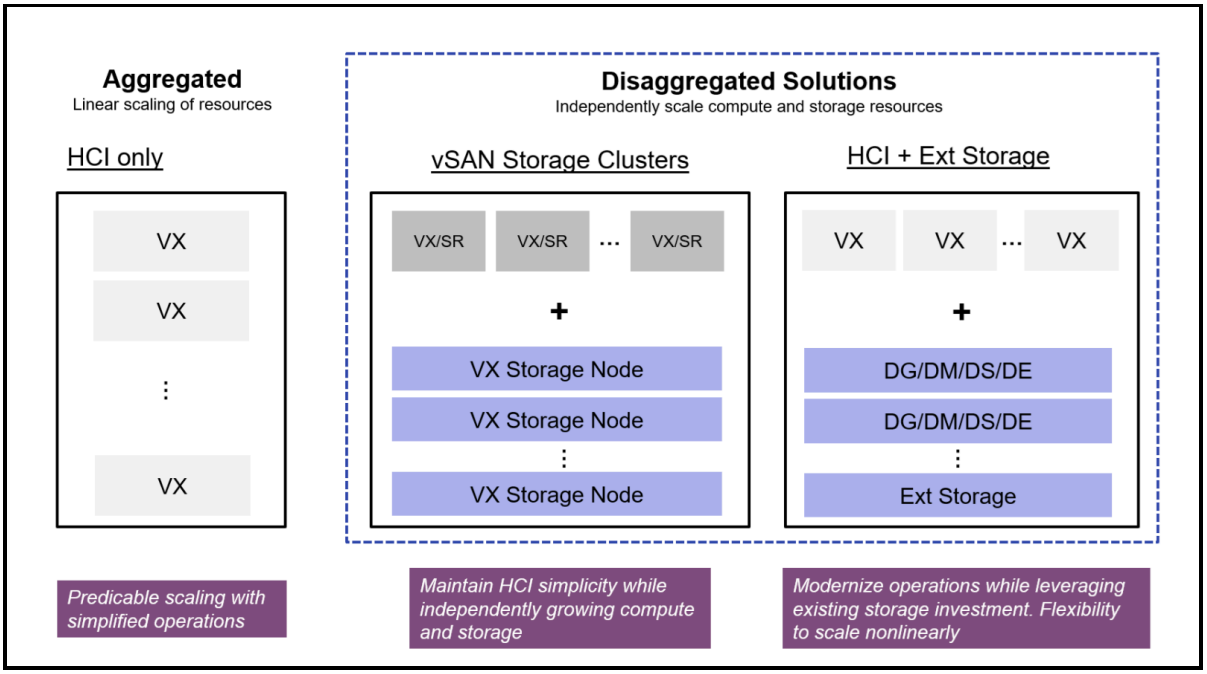

Figure 2. Three Deployment Options for VMware Environments

Scenario 1: Traditional HCI

Traditional HCI remains a foundational model where compute and storage reside on the same nodes and scale together. It is simple, well understood, and effective for predictable workloads. For example, an organization expanding virtual desktop infrastructure (VDI) or adding a limited number of AI inferencing workloads may find traditional HCI remains the most efficient option. In these scenarios, compute and storage requirements often grow in a relatively balanced manner, making a tightly integrated architecture both simple and effective.

Its limitation appears when growth patterns diverge. Scaling one resource requires scaling both, introducing inefficiencies in environments with uneven or rapidly evolving demands.

Scenario 2: vSAN Storage Clusters

vSAN Storage Clusters separate storage from compute while preserving the same operational experience. Storage services run in dedicated clusters and are consumed over the network by compute clusters, all managed through vCenter with consistent policies and lifecycle operations.

This model enables independent scaling, allowing storage to grow based on capacity demand while compute scales and refreshes on its own lifecycle. This is particularly valuable for AI and data-intensive environments where datasets may expand significantly while compute resources are refreshed or scaled independently. Organizations can add storage capacity for growing training datasets, archives, and unstructured data repositories without introducing unnecessary compute resources.

On ThinkAgile VX, these deployments are delivered as validated and engineered configurations, reducing risk and avoiding the complexity typically associated with custom disaggregated designs.

Scenario 3: vSAN with External Storage

Disaggregation extends further by integrating external storage. ThinkAgile VX compute clusters can connect to ThinkSystem storage arrays over Fibre Channel or iSCSI, bringing additional performance and data service capabilities into the environment.

The ThinkSystem Enterprise Storage portfolio provides tiered options for different workload profiles, including general-purpose storage, high-performance flash, and large-scale capacity solutions. External storage integration can also support AI workflows that rely on large enterprise data repositories. Organizations may choose to keep structured and unstructured business data on enterprise storage platforms while connecting AI-enabled applications and VMware workloads to those datasets. This enables data growth and compute growth to be managed independently while maintaining consistent access to critical information.

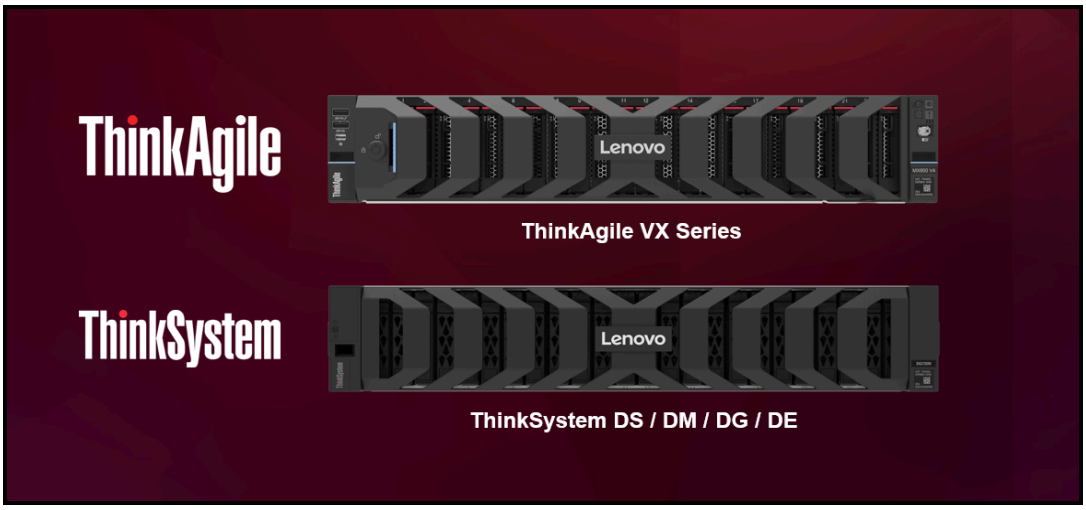

Figure 3. ThinkAgile VX and Supported ThinkSystem Enterprise Storage Arrays

- DE series — cost-effective, high-throughput storage for general-purpose workloads.
- DM series — unified storage with strong data management capabilities.
- DG series — high-performance flash, built for efficiency.
- DS series — scalable capacity for SAN-based deployments.

These can be used alongside vSAN, allowing infrastructure to be tailored to specific application needs. Importantly, this integration maintains operational consistency with ONTAP tools for VMware. Storage provisioning, protection, and monitoring remain visible through familiar VMware tools, avoiding the need for separate management workflows.

Flexibility Without Operational Fragmentation

One of the historical challenges with disaggregated architectures has been operational complexity. This challenge becomes even more significant as organizations introduce AI initiatives. Many IT teams are already managing virtualized infrastructure, containerized applications, and emerging AI workloads. Adding architectural flexibility should not require adding operational complexity.

In many cases, the flexibility gained at the architectural level came at the cost of operational simplicity.

Whether infrastructure is deployed as traditional HCI, vSAN storage clusters, or with external storage, it is managed through a single control plane using VMware vCenter. Policies, lifecycle operations, and monitoring workflows remain consistent across the environment.

In addition, integrations with enterprise storage platforms bring advanced data services into the same management framework using ONTAP tools for VMware vSphere. Capabilities such as snapshots, replication, and storage efficiency are exposed through the same tools used to manage virtual infrastructure, allowing teams to operate more efficiently without introducing additional layers of complexity.

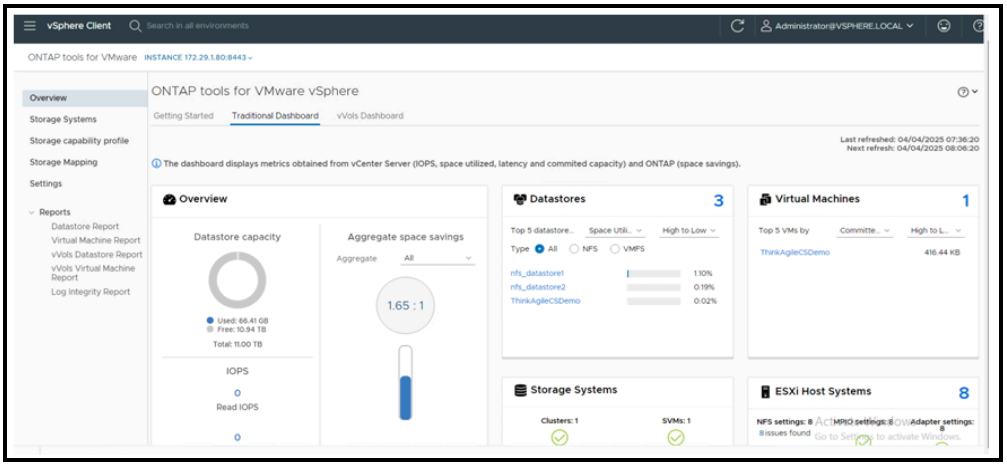

Figure 4. ONTAP tools for VMware vSphere to provision ThinkSystem Enterprise Storage

In addition to operational consistency, enterprise storage platforms provide advanced data protection capabilities. Features such as encryption, replication, secure authentication, WORM storage, and AI-powered Autonomous Ransomware Protection help safeguard critical business data while supporting compliance requirements.

The Lenovo Advantage

The value of disaggregation depends on how it is implemented. Delivering it as a collection of independent components can introduce risk and complexity. Lenovo addresses this by delivering disaggregated solutions as validated, lifecycle-managed systems supported by engineered configurations, reference architectures, and unified support. This approach reduces deployment risk while accelerating time to value.

ThinkAgile VX integrates traditional HCI, vSAN storage clusters, and external storage into a single engineered framework. This allows organizations to support a broad range of workloads—from traditional business applications and databases to modern analytics and emerging AI initiatives—while maintaining a consistent VMware operational model.

Combined with the ThinkSystem storage portfolio, this approach enables organizations to adapt infrastructure to changing workload demands while maintaining operational simplicity.

In closing, disaggregated infrastructure on ThinkAgile VX provides multiple ways to deploy and scale VMware environments, including traditional HCI, dedicated storage clusters, and external storage integration. These approaches can be combined within a single environment, allowing infrastructure to evolve alongside workload demands. Whether supporting traditional enterprise applications, cloud-native services, analytics platforms, or emerging AI initiatives, organizations increasingly need the ability to scale compute and storage independently without increasing operational complexity. Through validated VMware-native architectures, ThinkAgile VX delivers that flexibility while preserving the operational simplicity organizations depend on.

For More Information

- For more details on ThinkAgile VX Storage Clusters, check out this recent article: Modern HCI needs Options - Rethinking Storage at Scale with VMware Storage Clusters
- Learn more in this reference architecture for Enterprise Storage support: Disaggregated Solutions for VMware with ThinkAgile VX and ThinkSystem DE/DM/DG/DS Storage

Author

Catherine Maina is a Senior Product Manager for ThinkAgile VX at Lenovo, where she leads portfolio strategy and market direction for VMware based hyperconverged and software defined infrastructure platforms. She works across engineering, sales, marketing, and ecosystem partners to shape co engineered solutions that support modern hybrid cloud architectures, balancing scalability and operational simplicity for enterprise customers.

Related product families

Product families related to this document are the following:

- ThinkAgile VX Series for VMware

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, LP2464, was created or updated on July 1, 2026.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2464
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2464.

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®

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