Modern Private Cloud with VMware Cloud Foundation on Lenovo ThinkSystem, ThinkAgile VX/FX Systems and ThinkSystem Storage

Planning / Implementation

The Cloud Operating Model and the Strategic Return to Private Cloud

For many years, enterprise infrastructure strategy involved a compromise between two infrastructure models. Public cloud platforms provided rapid service delivery and operational agility, while on-premises environments provided greater security, data sovereignty, cost predictability, and long-term governance.

Although the public cloud offers lower entry costs, long-term operational costs can become highly variable. Data movement and egress charges, premium storage tiers, and the continuous growth of datasets for analytics and Artificial Intelligence (AI) frequently lead to unpredictable spending and expanded security risk surfaces. As a result, organizations are increasingly adopting a “Private Cloud First” strategy for core enterprise workloads. This shift is further accelerated by the growth of AI workloads, which require high-speed, scalable, and highly reliable infrastructure to maintain predictable performance and low execution latency.

VMware Cloud Foundation (VCF) addresses this long-standing operational challenge by transforming the data center from a collection of isolated infrastructure systems into a unified private cloud platform. By integrating compute, storage, networking, automation, and lifecycle management into a single software-defined stack, VCF 9.0 delivers a cloud operating model within the enterprise data center while maintaining direct control over data location, governance, and security policies.

VCF provides a scalable and high-performance private cloud architecture when paired with:

- ThinkAgile VX V4 and FX V4 hyperconverged platforms and ThinkSystem V4 servers powered by Intel Xeon 6 processors
- ThinkAgile VX V3 and ThinkSystem V3 servers powered by 4th and 5th Gen AMD EPYC processors
- Lenovo ThinkSystem DE/DM/DG/DS Storage systems

This combined disaggregated architecture delivers a storage design leveraging NVMe for high-performance workloads alongside capacity-optimized storage resources, optimized for cost and scaling across modern enterprise workloads including AI, Kubernetes, databases, analytics, and mission-critical applications, while simplifying operations through validated hardware configurations and integrated lifecycle tooling.

Cloud-Scale Infrastructure Foundation for Private Cloud

Lenovo ThinkAgile VX/FX hyperconverged systems, ThinkSystem V3/V4 servers, and ThinkSystem Storage systems are engineered to serve as the infrastructure foundation for modern private cloud environments. These platforms deliver the compute density, high memory bandwidth, storage throughput, and networking performance required to support a wide range of enterprise applications.

Deploying VMware Cloud Foundation 9.0 on Lenovo V3/V4 infrastructure improves performance

consistency, operational efficiency, scalability, and infrastructure resiliency. The combination of VCF software and Lenovo infrastructure enables organizations to establish a unified private cloud operating model capable of supporting traditional enterprise applications, Kubernetes-based services, and AI-driven workloads on a shared physical platform.

The private cloud architecture built on VCF 9.0, Lenovo V3/V4 infrastructure, and ThinkSystem Storage provides flexible deployment models that can be aligned with different operational, storage, and performance requirements.

The following figure shows a typical VCF 9.0 Private Cloud Reference Architecture Models and Deployment Options with HCI, disaggregated and external storage systems.

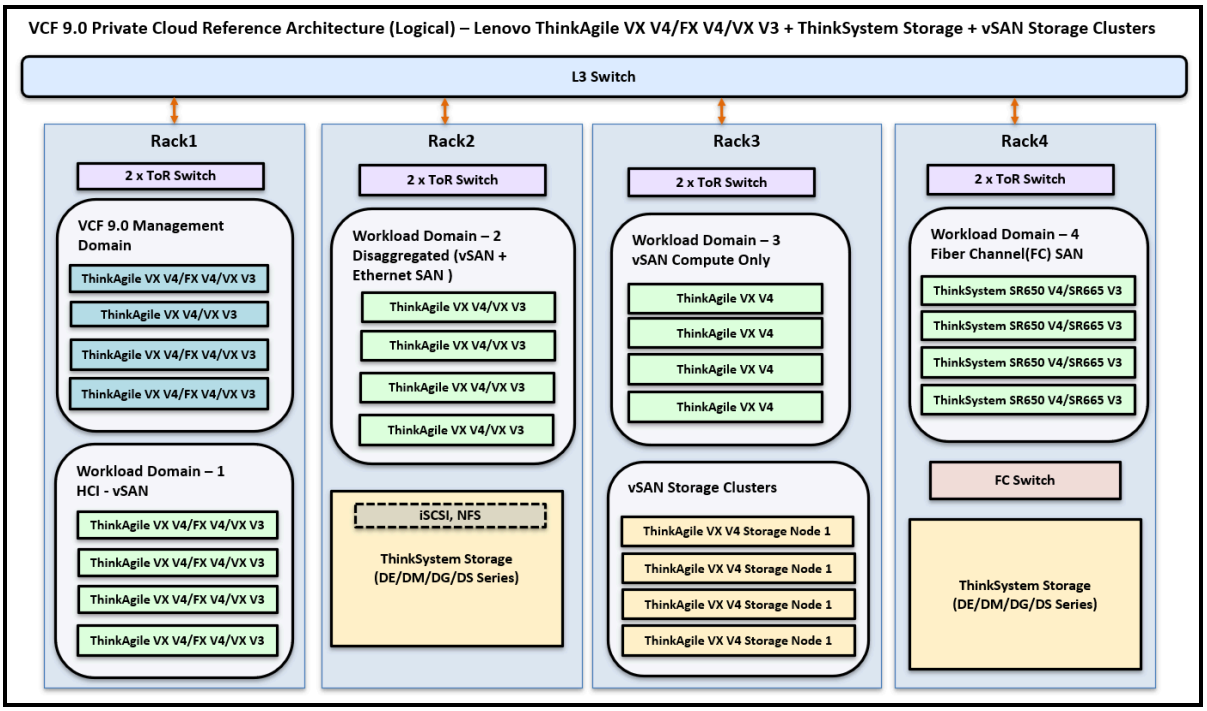

Figure 1. VCF Private Cloud Logical Deployment Architecture with ThinkAgile, ThinkSystem Server and Storage

The design uses distinct workload domains for different workloads based on compute and storage requirements instead of creating a single, massive infrastructure cluster for all enterprise systems.

The workload domains are as follows:

- **Management Domain (Cluster 1 - HCI):** This configuration consists of four Lenovo ThinkAgile VX V4/FX V4/VX V3 dedicated to the VCF management domain. The cluster uses vSAN hyperconverged infrastructure (HCI) to provide highly available management services while maintaining operational separation from production workloads. This cluster can serve compute workloads as well for smaller deployments.
- **Workload Domain (Cluster 2 - HCI):** This deployment consists of four Lenovo ThinkAgile VX V4/FX V4/VX V3 nodes connected through a high-speed Ethernet fabric. The architecture isolates east west vSAN replication and synchronization traffic from application and management traffic to improve operational consistency and storage performance.
- **Workload Domain (Cluster 3 Disaggregated - HCI + Ethernet SAN Storage):** This configuration combines ThinkAgile VX V4/FX V4/VX V3 nodes with external Lenovo ThinkSystem Storage arrays connected through Ethernet fabric. The architecture supports block and file storage services using protocols such as iSCSI and NFS in addition to local vSAN storage. The workloads can leverage both vSAN and external storage based on the performance and operational requirements.

- **Workload Domain (Cluster 4 - ThinkAgile VX V4 Compute-Only + vSAN Storage Cluster):** This architecture separates computing and storage resources using vSAN Storage Cluster. Dedicated compute-only clusters consume storage services from an independent disaggregated vSAN storage cluster, allowing storage and computing capacity to scale independently.
- **Workload Domain (Cluster 5 – ThinkSystem Storage with Fiber Channel):** This model consists of four Lenovo ThinkSystem rack servers connected to external ThinkSystem Storage arrays through a dedicated Fiber Channel SAN infrastructure. Ethernet networking is reserved for VM and management traffic, while storage communication occurs across an isolated FC fabric.

Rack awareness is critical to avoid a single rack failure taking down complete cluster or storage. It ensures high availability by spreading workloads across independent power, cooling, and network domains. This design provides high resilience and faster recovery in multi-cluster and disaggregated deployment.

Principal Storage Support for SAN and NAS

VMware Cloud Foundation 9.0 provides enhanced support for principal storage architectures and converged migration workflows, giving architects greater flexibility when designing disaggregated storage environments.

Native support for Fiber Channel (FC) and NFS v3 as principal storage options allows these storage types to be configured directly during management domain deployment, workload domain creation, or cluster expansion operations. This capability aligns well with Lenovo ThinkSystem DE, DM, DG, and DS storage systems deployed within enterprise SAN and NAS environments.

For storage protocols that are not directly integrated into greenfield VCF workflows—such as iSCSI, NFS v4.1, FCoE, and NVMe over Fabrics (including NVMe/FC and NVMe/TCP)—VCF 9.0 supports converged deployment workflows and the ESXi hosts, and their associated storage configurations are prepared before domain ingestion, enabling organizations to incorporate existing enterprise storage investments into the VCF operating model during migrations from VMware vSphere Foundation (VVF), vSphere environments, or vCloud Suite deployments.

vSAN ESA 9.0 – Optimized and Resilient Hyperconverged Storage

The Express Storage Architecture (ESA) in vSAN 9.0 is a complete redesign of the software-defined storage data path, engineered from the ground up to maximize the efficiency of high-speed NVMe flash storage media.

Performance Without Compromise

vSAN ESA 9.0 implements a new log-structured file system that facilitates Zero-Loss Compression and Deduplication to achieve substantial storage space savings without suffering the performance penalties historically associated with data reduction algorithms. This is achieved because data reduction tasks are entirely offloaded and parallelized across modern, multi-threaded CPU cores, keeping the active I/O path unencumbered.

Native Snapshot Performance

To defend against modern security threats like ransomware, snapshot capabilities must be both frequent and performant. vSAN ESA 9.0 is entirely redesigned native snapshot engine that operates up to 100x faster than previous generations. Because these snapshots introduce minimal performance impact to active virtual disks, architects can deploy aggressive protection intervals (e.g., every 15 minutes) without causing operational degradation to high-transaction production databases.

Advanced Traffic Separation

In a disaggregated or hyperconverged vSAN storage cluster deployment, each host can configure dedicated, isolated networking paths. Architects can tag one VMkernel port specifically for vSAN storage cluster client traffic and a separate VMkernel port for back-end vSAN storage activities.

The client port handles guest virtual machine I/O to and from the vSphere clusters mounting the datastore, while the core vSAN port is used exclusively by the storage cluster for background activities like data resilience syncing, cluster rebalancing, and drive repairs.

The VMkernel port assigned to vSAN client traffic can run on an entirely separate VLAN and network subnet from the back-end vSAN storage traffic port. This network isolation ensures that heavy background storage maintenance tasks never induce latency on front-end user applications. This capability gives deployment teams the flexibility to design strict traffic isolation within Top-of-Rack (ToR) switches or distribute traffic across dedicated physical switches using 10, 25, or 100GbE bandwidth paths.

Global Deduplication and TCO Optimization

vSAN 9.0 improves the performance of its global deduplication engine to deliver enhanced cost savings for the enterprise. When combined with the VCF licensing mode which includes a 1 TiB entitlement of raw vSAN storage for every licensed VCF core, this global deduplication functionality yields a significantly lower Total Cost of Ownership (TCO) without diminishing application performance.

The system leverages the distributed scale-out architecture of vSAN, utilizing the entire cluster as a single, contiguous deduplication domain. As new nodes are added to the cluster, the deduplication domain automatically scales out. This expansion increases the mathematical probability of identifying duplicate data blocks, delivering a progressively higher deduplication ratio over time.

File Services and Site Resilience

- **Enterprise File Services Scale:** The scalability of native vSAN File Services increases to support up to 500 file shares per cluster within VCF 9.0. This allows organizations to execute broad file-based workload consolidation strategies, eliminating the cost and complexity of deploying external, third-party file systems for enterprise file storage.
- **Site-Level Resilience Design Options:** VCF 9.0 provides architectural support for stretched compute-only clusters connected to stretched storage cluster deployments (or vSAN Max configurations). This gives architects advanced multi-site availability options to protect critical applications and data volumes against localized data center failures.

Memory Tiering over NVMe

This feature allows the hypervisor to utilize high-speed local NVMe storage as an extension of physical memory capacity. ESXi monitors memory access behavior and dynamically relocates colder or less frequently accessed memory pages to NVMe storage while preserving active operational data within DDR5 or MRDIMM memory.

By optimizing memory utilization in this manner, organizations can increase VM density and improve infrastructure efficiency without requiring maximum DRAM population across every physical server. Memory Tiering over NVMe can benefit memory-intensive enterprise applications such as databases, analytics platforms and workloads which require moderate active memory. Also software mirroring adds redundancy to NVMe tier layer to protect workloads from single device failure. These features are enabled through vSphere Configuration profile to apply consistently across nodes in the cluster.

VCF Operations tracks DRAM and NVMe tier performance and project potential cost savings and capacity gains to do capacity planning and hardware expansion efficiently.

Built-in Kubernetes for Modern Applications

VMware Cloud Foundation is designed to support both traditional enterprise applications and modern Kubernetes-based workloads within a unified operational model. Rather than maintaining separate infrastructure silos for containers and virtual machines, organizations can run both application types on shared infrastructure while maintaining common operational workflows for networking, security, monitoring, and lifecycle management.

VCF Operations provides integrated observability and monitoring for both the vSphere Supervisor Cluster and VMware Kubernetes Service (VKS). Through integrated telemetry collection services, infrastructure and Kubernetes operational metrics are continuously streamed into the VCF Operations platform.

Built-in dashboards and operational views provide visibility into multiple layers of the Kubernetes infrastructure stack:

- **Compute and Capacity Monitoring:** Real-time visibility into CPU, memory, and storage utilization across Kubernetes infrastructure and associated virtualized resources.
- **Kubernetes Construct Visibility:** Health monitoring and operational status tracking for clusters, namespaces, worker nodes, and Kubernetes services.
- **Container-Level Observability**: Detailed visibility into container operations and performance behavior to assist operations teams in identifying bottlenecks and application performance issues.

Virtual Private Cloud Ready Experience

The networking enhancements in VCF 9.0 are designed to simplify private cloud networking consumption through automation and software-defined networking services. Virtual Private Cloud (VPC)-style constructs allow infrastructure and application teams to provision isolated network environments using policy-based automation.

Application teams can provision Network connectivity, Segmentation policies, Firewall rules, NAT services and Load balancing within isolated logical network environments while maintaining centralized governance and security oversight. To simplify communication between VPC environments and external enterprise services, the workload domain Transit Gateway architecture streamlines inter-VPC routing and provides structured connectivity to shared enterprise services and external networks.

The NSX networking layer is also integrated closely with VCF lifecycle management workflows, enabling automated certificate management, credential rotation, and coordinated patching operations across the networking stack.

Unified Architecture and Eliminating Operational Silos

VMware Cloud Foundation unifies compute, storage, networking, automation, and lifecycle management into a centralized operational framework. This approach reduces operational fragmentation by enabling infrastructure teams to manage distributed infrastructure environments through standardized deployment workflows, centralized visibility, and coordinated lifecycle operations.

Through VCF Automation, enterprise teams can provision virtual machines, Kubernetes clusters, networking services, and application environments using self-service catalogs, APIs, and automated deployment blueprints. This allows organizations to reduce manual provisioning activities while improving deployment consistency across environments.

The Unified Operations Console and Fleet Scale Management

VCF is built with Unified Operations Console, which consolidates private cloud management into a centralized operational interface. VCF Operations provides integrated monitoring, lifecycle management,

capacity visibility, and operational diagnostics across all active workload domains and infrastructure clusters.

VCF Fleet construct allows multiple distributed VCF environments to be managed collectively through centralized VCF Operations and VCF Automation services. This architecture allows organizations to manage vCenter Servers, NSX Managers, Workload Domains, vSphere Clusters, VCF Automation instances and VCF Operations services through a unified operational framework.

Key operational capabilities include:

- **Streamlined Expansion:** Unified deployment workflows simplify infrastructure expansion and improve deployment consistency.
- **Fleet Lifecycle Operations:** Software patching, upgrade orchestration, and compliance operations are coordinated across the infrastructure fleet.
- **Financial and Capacity Visibility:** Native operational dashboards provide visibility into infrastructure utilization, growth trends, and capacity planning.
- **Centralized Licensing:** Licensing administration and core utilization visibility are consolidated through VCF Operations.
- **End-to-End Diagnostics:** Infrastructure diagnostics provide visibility into storage, compute, networking, and service-level dependencies.
- **Integrated Network Insights:** Application discovery and traffic flow analysis improve operational understanding of multi-tier applications.
- **Unified Observability Plane:** Logs, events, metrics, and infrastructure telemetry are consolidated into a unified operational analytics platform.
- **Unified System Health Dashboard:** Critical alerts and operational insights are aggregated into a centralized health monitoring interface.
- **Integrated cost visibility and capacity analytics:** Helps infrastructure teams understand compute, storage, and networking consumption patterns across business units and application environments.

Platform Security, Compliance, and Governance

VCF security operational model is aligned with modern enterprise security and compliance requirements. It is centered on Zero Trust architecture to protect data and production workloads.

FIPS-compliant cryptographic standards are integrated across core infrastructure services, supporting encrypted communication and secure data handling. Integrated security operations dashboards provide visibility into Host encryption status, vSAN encryption status, Certificate lifecycle monitoring, Security vulnerability visibility and Compliance drift detection.

Automated compliance monitoring assists organizations in identifying deviations from operational baselines and regulatory requirements, including frameworks such as PCI-DSS, HIPAA, and GDPR.

VMware Advanced Cyber Compliance provides comprehensive cyber recovery capabilities for on-premise and ThinkAgile VX V3/V4 servers can be used to build dedicated vSAN storage cluster on the recovery sites to build Isolated Recovery Environment (IRE) and perform cyber recovery without a dependency on the public cloud.

Policy as Code and Advanced Automation

VCF Automation provides expanded Infrastructure as Code (IaC) and policy-driven automation capabilities. Administrators can define infrastructure blueprints, Kubernetes clusters, networking policies, and application environments using declarative configuration models integrated with source code repositories such as GitHub, GitLab, and Bitbucket. YAML-Based Policy as Code enables organizations to standardize governance and automate compliance enforcement across Kubernetes and virtualization environments.

Private AI: The Next Frontier of the Private Cloud

VMware Cloud Foundation combined with Lenovo ThinkAgile V3/V4 infrastructure provides a validated foundation for VMware Private AI deployments.

This architecture enables organizations to deploy, operate, and manage AI services within enterprise-controlled infrastructure environments while maintaining governance, operational visibility, and data protection.

- **GPU Resource Allocation**: VCF 9.0 provides mechanisms for reserving and assigning GPU resources to AI and high-performance computing workloads. This improves workload isolation and ensures predictable access to accelerator resources for critical AI services.
- **Optimized Storage for Retrieval-Augmented Generation (RAG):** Modern AI environments depend heavily on rapid access to enterprise datasets and vector databases. vSAN ESA 9.0 provides the low-latency NVMe storage infrastructure required to support RAG pipelines, inference services, and data-intensive AI operations.
- **AI Blueprints:** Organizations can package AI development environments and deploy frameworks into reusable automation blueprints that can be provisioned through self-service workflows.
- **Private AI Services:** VCF-based AI infrastructure also supports shared AI model services, allowing multiple business units or tenants to consume centralized AI models while maintaining data isolation and governance boundaries.

Conclusion

VMware Cloud Foundation provides advanced capabilities for enterprise private cloud architecture and addresses traditional private cloud challenges by integrating compute virtualization, software-defined storage, networking, automation, lifecycle management, and Kubernetes services into a unified cloud operating framework. vSAN ESA is optimized for dense flash storage, high-throughput networking, and modern multi-core processors, allowing organizations to support increasingly demanding enterprise workloads more efficiently.

Lenovo ThinkAgile VX V3/V4, FX V4, ThinkSystem V4 platforms and ThinkSystem Storage provide the infrastructure foundation required to deploy scalable private clouds to address modern workloads performance and enterprise operational requirements. The value of the combined solution provides a scalable and repeatable framework for organizations modernizing private cloud operations, standardizing Kubernetes infrastructure, maintaining data sovereignty, and preparing enterprise environments for AI-enabled applications.

For more information

For more information, see these resources:

- ThinkAgile VX Series
  https://lenovopress.lenovo.com/datasheet/ds0104
- ThinkAgile FX Series
  https://lenovopress.lenovo.com/datasheet/ds0203
- ThinkAgile VX for VMware vSAN Storage Clusters
  https://lenovopress.lenovo.com/lp2422
- Reference Architecture: Disaggregated Solutions for VMware with ThinkAgile VX and ThinkSystem Storage
  https://lenovopress.lenovo.com/lp2201
- Deployment Guide: VMware Cloud Foundation on Lenovo ThinkAgile VX and FX
  https://lenovopress.lenovo.com/lp1581
- VMware Cloud Foundation
  https://www.vmware.com/products/cloud-infrastructure/vmware-cloud-foundation

Author

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.

Related product families

Product families related to this document are the following:

- ThinkAgile FX Series
- ThinkAgile VX Series for VMware
- VMware Alliance

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, LP2462, was created or updated on August 25, 2026.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2462
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2462.

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®

The following terms are trademarks of other companies:

AMD and AMD EPYC™ are trademarks of Advanced Micro Devices, Inc.

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.