Cisco UCS with VAST Data



© 2026 Cisco and/or its affiliates. All rights reserved.

Contents

Overview 3
Features and benefits 4
Prominent feature 5
Platform support 6
Licensing 7
Product sustainability 8
Product specifications 8
System requirements 11
Ordering information 11
Warranty information 13
Cisco and partner services 14
Cisco Capital 14
Learn more 14

© 2026 Cisco and/or its affiliates. All rights reserved.

The Cisco AI Data Platform with VAST AI OS makes enterprise data continuously usable for AI by unifying files, objects, tables, streams, metadata, vectors, and events in one software platform. Built on Cisco UCS® infrastructure — with Cisco CNode-X (Cisco UCS® C845A M8 Rack Server) for GPU-accelerated data services and Cisco EBox (Cisco UCS C225 M8 Rack Server) for resilient all-flash data management — this solution is a core component of the Cisco Secure AI Factory with NVIDIA, delivering a security-first, enterprise-grade data foundation for agentic and physical AI deployments from core to edge.

Overview

VAST AI OS gives Cisco AI PODs a unified AI data platform that turns enterprise data into live, governed context for models, agents, and modern AI applications. Instead of stitching together storage, pipelines, vector databases, and separate processing tiers, VAST consolidates data services into one software platform that ingests, enriches, indexes, embeds, and serves data in place. The solution brings together Cisco CNode-X systems (Cisco UCS C845A M8 Rack Server) for GPU-accelerated data services and NVIDIA NIM orchestration with Cisco EBox systems (Cisco UCS C225 M8 Rack Server) for resilient all-flash data management. Together, Cisco CNode-X and Cisco EBox form the Cisco® hardware foundation for Cisco AI PODs: EBox delivers the persistent, high-throughput data layer while CNode-X accelerates vector search, real-time pipelines, analytics, and AI services close to the data.

VAST AI OS is built on the VAST Data Platform's Disaggregated Shared-Everything (DASE) architecture, which scales performance and capacity independently across flash and stateless compute to reduce bottlenecks, data copies, stranded capacity, and operational complexity. VAST InsightEngine extends this foundation by continuously classifying content, extracting context, generating embeddings, and turning enterprise knowledge into searchable, AI-ready signals. By partnering with VAST, Cisco was first to market with a unified AI Data Platform, launching VAST InsightEngine on Cisco AI PODs in September 2025.

As a core component of Cisco Secure AI Factory with NVIDIA, the Cisco AI Data Platform combines Cisco's AI infrastructure and networking leadership with VAST's advanced data platform to help enterprises unify, govern, and activate data for AI. Cisco contributes trusted compute, networking, validated designs, and lifecycle support. VAST provides a unified software platform that keeps enterprise data continuously usable for retrieval, inference, analytics, and agents. The result is a faster, simpler path to secure, production-scale AI with performance, scalability, and enterprise readiness customers expect from Cisco. Security is embedded at every layer: Cisco AI Defense protects AI models and applications, Cisco Hybrid Mesh Firewall secures workloads and infrastructure, and Splunk® Observability Cloud delivers end-to-end AI infrastructure and agent monitoring — with Cisco CX providing unified, full-stack support that resolves multi-vendor issues an average of 43% faster.

© 2026 Cisco and/or its affiliates. All rights reserved.

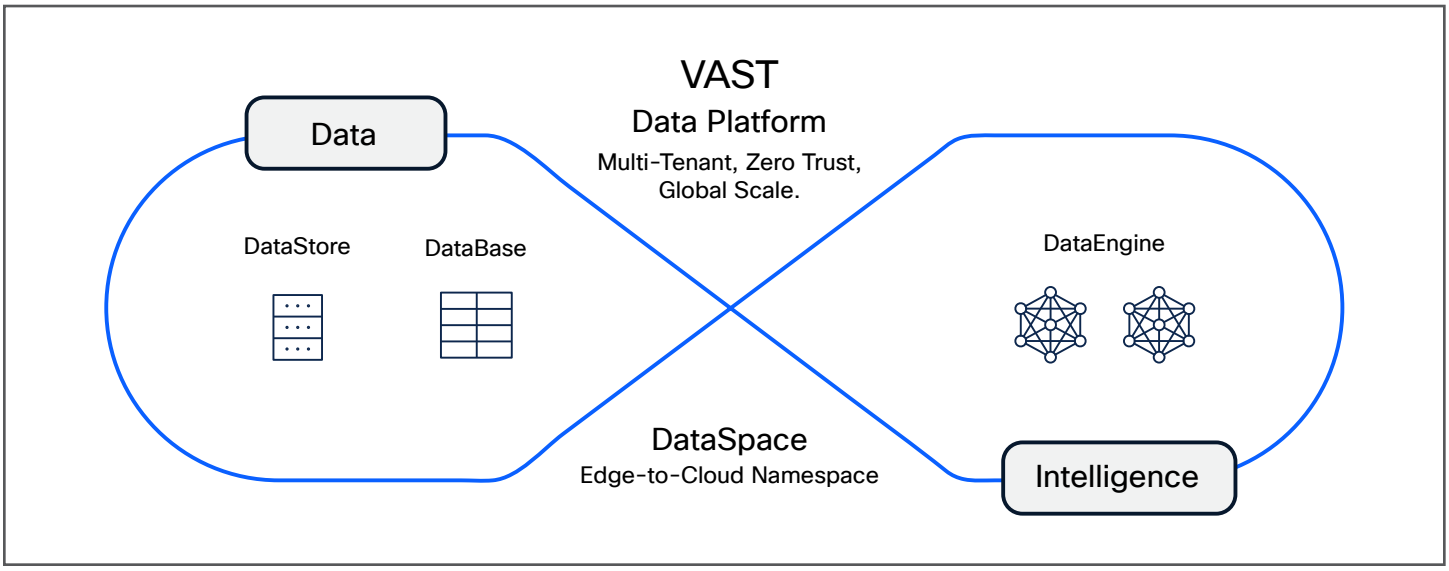


Figure 1. VAST Data Platform high-level solution

Features and benefits

Table 1. VAST InsightEngine and VAST Data Platform features and benefits

Feature	Benefit
Disaggregated, Shared-Everything (DASE) Architecture	Separates storage from compute resources for independent scaling, delivering ultra-low latency and massive bandwidth for real-time analytics and AI workloads.
Unified Global Namespace (DataSpace)	Consolidates workloads onto a single, unified platform, eliminating storage tiering complexity and simplifying management across edge, core, and cloud environments.
Multi-Protocol Data Plane (NFS, SMB, S3, NVMe/TCP)	Connects AI training/inference, MLOps, Business Intelligence (BI), and backup tools without data copying, ensuring seamless integration and data accessibility.
Flash-Optimized Erasure Coding and Similarity-Based Data Reduction	Achieves archive-like economics on all-flash storage, significantly reducing overall infrastructure costs while maintaining high performance.
VAST DataBase (Columnar, Tabular)	Provides low-latency analytics on operational data, supports feature stores, and enables vector/graph storage for real-time RAG data preparation.

© 2026 Cisco and/or its affiliates. All rights reserved.

Feature	Benefit
VAST DataEngine (Programmable, Event-Driven Serverless)	Automates data pipelines (ingest → transform → index) and triggers jobs on data events, accelerating data processing for AI workflows.
InsightEngine (with NVIDIA NIM integration)	Keeps RAG corpora continuously current by embedding new data as it lands, supporting multimodal retrieval and accelerating AI model training.
AgentEngine	Manages and observes agent pipelines at scale, providing an application layer for AI OS that simplifies complex AI deployments.
Enterprise-Grade Reliability and Security	Offers advanced data protection features, including metadata triplication, immutable snapshots, and ransomware-resilient architecture, ensuring data integrity and rapid recovery. Within the Cisco Secure AI Factory, VAST's data protection is complemented by Cisco AI Defense for AI model and application security, an NVIDIA BlueField DPU-based security offload that preserves GPU resources, and Splunk Enterprise Security for threat detection across the AI stack.
GPU-Accelerated Data Services (CNode-X)	Cisco CNode-X (UCS C845A M8) with 8x NVIDIA RTX PRO 6000 GPUs runs AI services where data lives, accelerating vector search, real-time SQL analytics, scalable AI pipelines, and inference for long-context, multimodal, and multi-agent workloads. Removes bottlenecks across ingestion, transformation, indexing, retrieval, analytics, and inference while keeping GPUs saturated.

Prominent feature

CNode-X + EBox: GPU-Accelerated AI Services Where Data Lives

Cisco CNode-X with VAST AI OS extends the Cisco Secure AI Factory with NVIDIA with a unified, GPU-accelerated platform for Cisco AI PODs. Built on the Cisco UCS C845A M8 with 8x NVIDIA RTX PRO 6000 GPUs (5th Gen Tensor, 2nd Gen Transformer Engine, FP4, 96GB GDDR7), CNode-X runs AI services where data lives, accelerating the full path from data ingest to inference as a single, integrated system. It accelerates vector search and retrieval, real-time SQL analytics, scalable AI pipelines, and inference for long-context, multimodal, and multi-agent workloads. By bringing GPU acceleration into the VAST AI OS, CNode-X removes bottlenecks across ingestion, transformation, indexing, retrieval, analytics, and inference while simplifying the AI data stack. Cisco EBox, built on the UCS C225 M8, delivers the persistent, high-throughput, all-flash data layer that keeps GPUs fed for training, RAG, inference, and agentic AI workloads. Together, CNode-X and EBox form the hardware foundation for the Cisco AI Data Platform.

© 2026 Cisco and/or its affiliates. All rights reserved.

Revolutionary Disaggregated, Shared-Everything (DASE) Architecture

At the core of the VAST Data Platform is its unique Disaggregated, Shared-Everything (DASE) architecture. This innovative approach separates stateless front-end compute nodes (CNodes) from flash enclosures (DNodes), which provide persistence over NVMe over Fabrics (NVMe-oF). This disaggregation allows for independent scaling of compute and capacity resources, ensuring that any compute node can read or write any data without burdensome east-west chatter. The DASE architecture also features a global, transactional element store with Atomicity, Consistency, Isolation, Durability (ACID) semantics, eliminating per-node caches and metadata bottlenecks. Combined with flash-optimized erasure coding and Similarity-Based Data Reduction, DASE delivers breakthrough performance and efficiency, offering archive-like economics on all-flash storage.

Unified Data Platform for All Workloads

The VAST Data Platform consolidates all data types—files, objects, and tables—into a single, unified platform. Its global namespace (DataSpace) provides a consistent read/write namespace spanning edge, core, and cloud environments, complete with global read/write leases and cross-site Global Snapshot Clones for instant writable copies. This eliminates data silos and tiering complexity, allowing IT teams to manage all workloads, from traditional enterprise NAS consolidation to cutting-edge AI/ML pipelines, on one infrastructure. The platform's multi-protocol data plane (NFSv3/4.1, SMB 2.1/3.1, S3 object, Non-Volatile Memory Express [NVMe]/Transmission Control Protocol [TCP], GPUDirect/Remote Direct Memory Access [RDMA]) ensures seamless connectivity and high-throughput access for diverse applications.

Platform support

Table 2. Platform support

Product Family	Platforms Supported
Cisco UCS Servers	Cisco UCS C845A M8 (CNode-X), Cisco UCS C225 M8 (EBox)
Cisco Networking	Cisco N9300 Series Switches (for example, N9K-C9332D-GX2B, N9K-C9364D-GX2A)
Management	Cisco Intersight® SaaS
Networking Connectivity	Mellanox ConnectX-7 (CX7), NVIDIA BlueField-3 DPU

© 2026 Cisco and/or its affiliates. All rights reserved.

Licensing

VAST Data software is licensed per usable capacity and CPU cores, and all licenses are subscription-based with terms ranging from 36 to 60 months. There is no perpetual license offering. VAST Data software licenses are portable, allowing customers to transfer existing licenses to new equipment upon retirement or upgrade, ensuring continued use without additional licensing costs. Pricing is monthly but billed upfront for the full term. The same term must be selected across all related SKUs within a given quote or configuration to ensure contract alignment.

Table 3. VAST Data Software Licensing

License Type	Description	Application Details
VAST-SW-1TB	VAST Data software subscription (1TB)	VAST Data Subscription (quoted per 1TB), 36-60 months term. It includes DataStore, DataBase, DataEngine, and DataSpace)
VAST-SW-1CPU-CORE	VAST CPU Core License, quoted per core beyond included entitlement	License for CPU cores within the servers comprising the VAST cluster and CPU cores orchestrated by VAST software. Licensed per Core. 36 - 60 months term
VAST-SW-1CPU-ACC	VAST CPU Core Accelerated License	Refer to Ordering guide for details.
VAST-PS-COPILOT	VAST CoPilot service for proactive monitoring, upgrades, and system health management	Charged per cluster (not per node). Only one CoPilot service is required per deployment, regardless of node count. (36-60 months term)
VAST-PS	VAST Installation Services for VAST software	Apply 1x VAST-PS per cluster for up to 12 EBox nodes. For clusters larger than 12 nodes, add more quantities accordingly.

© 2026 Cisco and/or its affiliates. All rights reserved.

Product sustainability

Information about Cisco's Environmental, Social and Governance (ESG) initiatives and performance is provided in Cisco's CSR and sustainability reporting.

Table 4. Cisco Environmental Sustainability Information

Sustainability Topic		Reference
General	Information on product-material-content laws and regulations	Materials
	Information on electronic waste laws and regulations, including our products, batteries and packaging	WEEE Compliance
	Information on product takeback and reuse program	Cisco Takeback and Reuse Program
	Sustainability Inquiries	Contact: csr_inquiries@cisco.com
Power	Typical Power	650-800W/Max Power 850-1000W (per UCS C225M8 EBox)
Material	Product packaging weight and materials	Contact: environment@cisco.com

Product specifications

Table 5. Cisco UCS C845A M8 Rack Server CNode-X configuration

Component	Specification
Form factor	4RU GPU-accelerated compute node
CPU	Dual AMD EPYC 5th Generation (Turin) processors
RAM	Up to 4TB DDR5 memory (24 DIMM slots)
GPU	8x NVIDIA RTX PRO 6000 (96GB GDDR7 each, 768GB total GPU memory)
GPU interconnect	NVLink Bridge for GPU-to-GPU communication
Data connectivity	NVIDIA ConnectX-7 or BlueField-3 DPU (400GbE)
Boot	2x M.2 NVMe boot SSDs
Management network	2x 10 GbE mgmt
PSU	4x 3000W Platinum power supplies
Management	Cisco Intersight® (cloud or connected virtual appliance)

© 2026 Cisco and/or its affiliates. All rights reserved.

Table 6. Cisco UCS C225 M8 Rack Server EBox configuration

Component	Specification
Form factor	1U single-node rack server
CPU	AMD Epyc 9454P (48-core, 2.75GHz) AMD Turin 9555P (64 Cores, 3.2GHz)
RAM	384GB (12x 32GB) DDR5-4800 (Epyc) 384GB (12X32GB) DDR5-6400 (Turin)
Storage (drive layout options)	2x 960 GB SCMs, 8x 15TB (122TB usable per node)
2x 1.9TB SCMs, 8x 30TB (240TB usable per node)*	
3x 1.9TB SCMs, 7x 60 TB*	
Connectivity	2x CX-7
Boot	2x 1TB on-board M.2 boot SSDs
Management network	2x 10 GbE mgmt
PSU	Typical power 650-800W / max power 850-1000W

* Please check with the Cisco Compute BU for the latest drive capacity qualifications.

Table 7. Cisco EBox performance (per EBox)

Metric	Value
Read throughput	21GB/s
Max Write throughput*	10GB/s
Sustained Write throughput	2.6GB/s

* Maximum write performance is delivered automatically and can absorb write bursts of up to 30 minutes.

Table 8. Typical Data Reduction Rates (DRR)

Workload	DRR Rate
AI	2:1
Media	1.5:1

© 2026 Cisco and/or its affiliates. All rights reserved.

Workload	DRR Rate
HPC	2:1
General file and object	3:1

Table 9. Cisco UCS Ebox Raw Capacity Scaling (Usable TB)

Number of EBoxes	Cisco UCS 122 (8x 15.3TB)	Cisco UCS 244 (8x 30TB)	Cisco UCS 430 (8x 61.4TB)
8x EBoxes (min)*	638 TB	1276 TB	2233 TB
9x EBoxes	754 TB	1507 TB	2638 TB
10x Eboxes	868 TB	1737 TB	3040 TB
11x Eboxes	982 TB	1,965 TB	3439 TB
12x EBoxes	1,096 TB	2,192 TB	3837 TB
13x EBoxes	1,209 TB	2,419 TB	4234 TB
14x EBoxes	1,322 TB	2,645 TB	4630 TB
15x EBoxes	1,435 TB	2,871 TB	5025 TB
Each additional EBox adds at least**	113 TB	225 TB	393 TB
25x EBoxes	2,558 TB	5,117 TB	8956 TB
50x EBoxes	5,355 TB	10,710 TB	18743 TB
100x EBoxes	10,868 TB	21,736 TB	38039 TB

*** When you select fewer than 12 EBoxes for a cluster, this reduces the maximum write performance because the system sets the write buffer stripe width at initial installation, and it cannot be changed later.**

- Design with 8–10 EBoxes is resulting in approximately 15% less write bandwidth.
- Design with 11 EBoxes is resulting in approximately 6% less write bandwidth.
- 12 EBoxes or more deliver full write bandwidth.

For optimal performance, start with 12 or more EBoxes. If you're unsure how this impacts your cluster design, please reach out to the Cisco or your VAST sales contacts for guidance.

**** Capacity scaling is slightly better than linear.**

© 2026 Cisco and/or its affiliates. All rights reserved.

System requirements

Table 10. System requirements

Feature	Description
Minimum Cluster Size	Minimum of 8 Cisco UCS C225 M8 EBox storage nodes (12 recommended); CNode-X GPU compute nodes added based on AI workload requirements (minimum 1 CNode-X per cluster for accelerated services)
Networking	Minimum 2 Cisco Nexus® switches (N9332D-GX2B or N9364D-GX2A as leaf switches)
Management Network	1G/10G management switches (for example, 93108TC-FX3)
Cabling	Copper transceivers (QDD-2Q200-CU3M) for lengths up to 3m. Fiber cables (SFP-200G-SR4 or QDD-400G-SR8 with breakout) for longer lengths
Software Version	VAST AI OS (includes DataStore, DataBase, DataSpace, InsightEngine, DataEngine, and AgentEngine)

Ordering information

The Cisco AI Data Platform with VAST AI OS, including Cisco UCS CNode-X and EBox hardware plus VAST Data software, can be ordered through Cisco Commerce.

Ordering VAST Data Software Only (SolutionsPlus):

To order VAST Data software only, search for VAST-DATA-SPLUS in CCW. This ATO (Approved To Order) contains the VAST Data software licenses (VAST-SW-1TB, VAST-SW-1CPU-CORE, VAST-SW-1CPU-ACC) and services (VAST-PS-COPILOT, VAST-PS). Ensure to select the appropriate quantities and a consistent subscription term (36-60 months).

Ordering Cisco UCS Hardware and VAST Software (MLB):

For a single integrated bundle of Cisco UCS infrastructure and VAST Data software, use the Multi-Line Bill of Materials (MLB): VAST-DATA-MLB.

1. In CCW, search for and add VAST-DATA-MLB to your configuration.
2. Click “Select Options” to customize the MLB.
3. Select Cisco UCS C-Series hardware: UCSC-C845A-M8 for CNode-X GPU compute nodes and UCSC-C225M8N-EBOX for EBox storage nodes (use UCSC-C225M8N-EXP for expansion/RMA).

© 2026 Cisco and/or its affiliates. All rights reserved.

4. Configure PCIe NIC (Mellanox ConnectX-7 or NVIDIA BlueField-3 DPU) and country-specific selections for each EBox.
5. Add corresponding VAST Data software product IDs (PIDs): VAST-SW-1TB and VAST-SW-1CPU-CORE for EBox storage, VAST-SW-1CPU-ACC for CNode-X accelerated services, plus VAST-PS-COPILOT and VAST-PS for support. Configure quantities and subscription terms.
6. Add networking components: Select Cisco Nexus® Switches (for example, N9K-C9332D-GX2B, N9K-C9364D-GX2A) and necessary transceivers/cables. Ensure to include SKU: DCN-AI for AI intent.
7. Add Cisco Intersight for cloud-based management: Select DC-MGT-IS-SAAS-AD and specify quantity (typically 1 per UCS server).

Table 11. Ordering information

Cisco PID	PID Description
VAST-SW-1TB	VAST Data Subscription (quoted per 100TB), 36–60 months term. Includes DataStore, DataBase, DataEngine, DataSpace
VAST-SW-1CPU-CORE	VAST Data CPU CORE (quoted per core), 36–60 months term
VAST-SW-1CPU-ACC	VAST Data CPU CORE Accelerated (quoted per core), 36–60 months term
VAST-PS-COPILOT	VAST Data CoPilot for system monitoring, upgrades, expansions, and overall system health management
VAST-PS	VAST Data Installation Services for VAST software (per cluster up to 12 EBoxes)

Table 12. Core Components of Cisco + VAST Data Solution Offering

MLB	Category Name	ATO	PID Description
VAST-DATA-MLB	VAST Data Software and Cisco Hardware MLB		
	Cisco UCS C845A M8	UCSC-C845A-M8	CNode-X GPU-accelerated compute node (8x NVIDIA RTX PRO 6000)

© 2026 Cisco and/or its affiliates. All rights reserved.

MLB	Category Name	ATO	PID Description
	CISCO UCS C225 M8	UCSC-C225M8N-EBOX	UCS C225 M8 1U Rack Server for VAST with 15.3 TB drives
		UCSC-C225M8N-EXP	UCSC-C225-M8N-1P VAST expansion/RMA Server, 15.3 TB drives
	VAST DATA SOFTWARE	VAST-DATA-SPLUS	VAST Data Solution Plus Offer
	CISCO INTERSIGHT	DC-MGT-SAAS	Cisco Intersight SaaS
		CNDL-DESELECT-D	Conditional Deselect
	NEXUS SWITCHES	N9K-C9332D-GX2B	Nexus 9300 Series, 32p 400G Switch
		N9K-C9364D-GX2A	Nexus 9300 Series, 64p 400G Switch

Deal Registration Process: All opportunities must be registered via VAST Data's partner portal (partners.vastdata.com). Use your cisco.com email address and associate yourself with "Cisco Systems Inc." VAST registered deals are valid for 180 days from approval if submitted through Cisco and can be extended upon demonstration of active engagement. VAST will review and respond to submissions within two business days.

Warranty information

Cisco hardware components included in the VAST Data solution carry standard Cisco hardware warranties. For VAST Data software, customers with a current (non-expired) VAST Data license subscription will have access to the latest software updates and patches. Specific warranty details for VAST Data software should be confirmed with VAST Data or Cisco sales representatives.

© 2026 Cisco and/or its affiliates. All rights reserved.

Cisco and partner services

Cisco and its certified partners offer comprehensive services to support the deployment, management, and optimization of the VAST InsightEngine and VAST Data Platform. Cisco Solution Support is available for VAST Data software, with Cisco Technical Assistance Center (TAC) providing Level 1 support and engaging VAST Data engineering for software-related issues. Customers also have the option to contact VAST Data directly for support. Services include planning and design, implementation, and ongoing optimization to ensure peak performance and accelerate business innovation. Renewals of VAST software originally purchased through Cisco's Global Price List (GPL) shall be transacted through Cisco's GPL.

Cisco Capital

Flexible payment solutions to help you achieve your objectives

Cisco Capital® makes it easier to get the right technology to achieve your objectives, enable business transformation and help you stay competitive. We can help you reduce the total cost of ownership, conserve capital, and accelerate growth. In more than 100 countries, our flexible payment solutions can help you acquire hardware, software, services and complementary third-party equipment in easy, predictable payments. Learn more.

Learn more

Unlock the Full Potential of Your Data for the AI Era.

Are you ready to transform your data center into an AI-ready powerhouse? The Cisco + VAST Data solution provides the simplicity, scalability, and performance required to accelerate your AI, machine learning, and data-intensive workloads. Eliminate complexity, reduce costs, and gain actionable insights faster than ever before. As part of the Cisco Secure AI Factory with NVIDIA, the VAST Data Platform benefits from a security-first architecture with end-to-end observability and flexible deployment options — from modular Cisco AI PODs to turnkey Cisco Nexus Hyperfabric™-managed stacks — so you can operationalize AI at your own pace. For more information, please contact your Cisco sales representative or visit: https://www.cisco.com/site/us/en/solutions/artificial-intelligence/infrastructure/ai-pods.html.

© 2026 Cisco and/or its affiliates. All rights reserved. Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)