Meet Cisco Intelligent Packet Flow

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

© 2025 Cisco and/or its affiliates. All rights reserved.

AI and high-performance computing workloads impose unprecedented demands on network infrastructure. They require exceptional bandwidth, ultra-low latency, and intelligent traffic management to maintain GPU efficiency and ensure seamless performance. Traditional methods like static equal-cost multi-path (ECMP) routing cannot keep up with the demanding, bursty, and synchronized east-west traffic patterns typical of AI environments leading to congestion, job delays and resource inefficiencies.

Overview

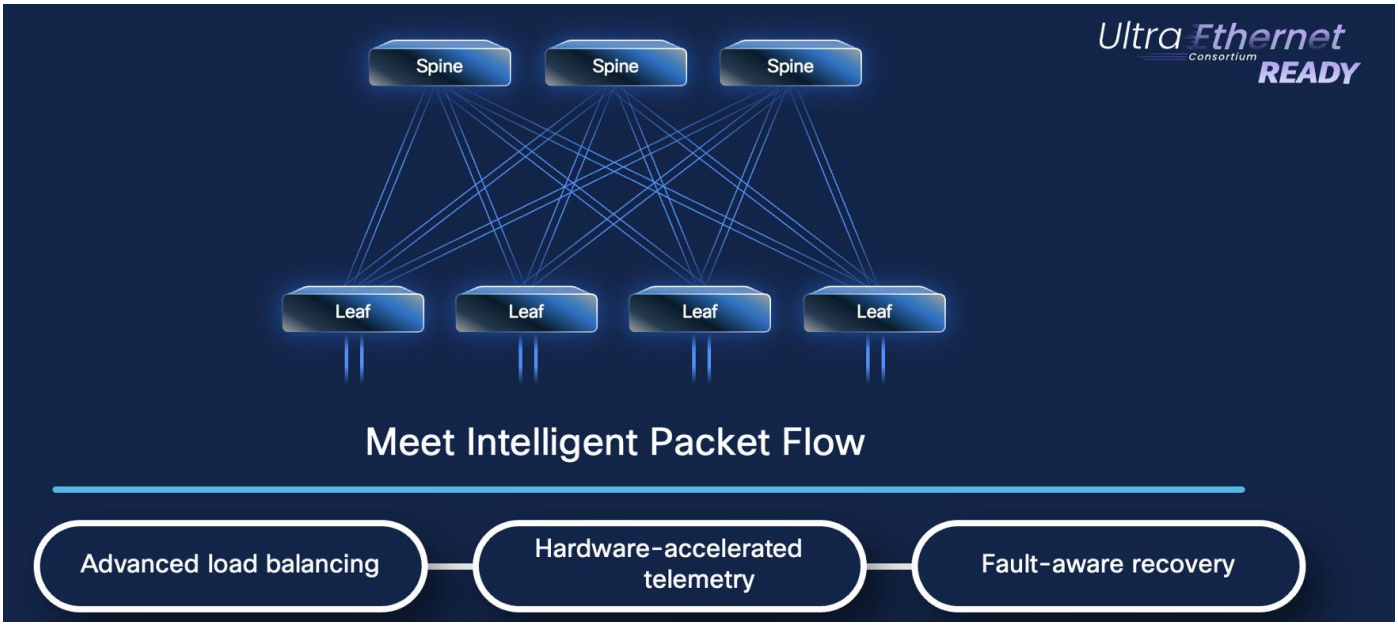

Figure 1. Powering AI fabrics with Intelligent Packet Flow with Cisco Nexus

Feature and Benefits

Table 1. Cisco Intelligent Packet Flow offers a range of benefits to meet the unique demands of AI/ML networks, including:

Feature	Benefit
Dynamic Load Balancing	• Maximizes GPU efficiency by evenly distributing data across available paths.
Adaptive Traffic Steering	• Reduces tail latency by avoiding congestion during peak load.
Real-time fault detection & rerouting	• Maintains uninterrupted AI/ML job performance despite network disruptions.
Embedded telemetry	• Enhanced Observability that deep and actionable visibility into live network behavior.
Policy-based load balancing	• Enables tailored and simultaneous traffic management for training, inference and storage workloads.

© 2025 Cisco and/or its affiliates. All rights reserved.

Trends and Challenges

The rising tide of AI workloads

The rapid adoption of AI and machine learning is reshaping data center architecture. AI training requires extensive GPU clusters and frequently generates synchronized east-west traffic patterns, placing immense pressure on networks. Traditional routing methods fail to keep up, causing latency spikes and inefficient resource utilization.

When AI workloads scale across rows of racks or multi-site clusters, one poorly managed burst of traffic can disrupt the entire fabric, leading to cascading performance issues. This demands a networking solution that delivers both precision traffic engineering and robust congestion control.

Cisco Intelligent Packet Flow addresses these trends, enabling networks to meet the performance demands of AI while maintaining fair resource allocation, optimal throughput and predictable performance at scale.

Key features

Table 2. Cisco Intelligent Packet Flow combines cutting-edge features to deliver precision traffic management in large-scale AI/ML environments.

Key components	Platforms supported
Dynamic load balancing (DLB)	Intelligent flowlet-based traffic distribution ensures efficient utilization of all available paths.
Weighted cost multi-path (WCMP)	Path weighting based on real-time telemetry dynamically allocates workloads to higher-capacity links, ensuring optimal throughput.
Per-packet load balancing	Packet spray maximizes utilization for high-throughput GPU-to-GPU communications.
Policy-based load balancing	Assigns specific traffic-handling strategies to mixed workloads based on ACLs, DSCP markings, or RoCEv2 headers, creating custom-fit efficiency for diverse needs.

Use case

Table 3. Cisco Intelligent Packet Flow enables:

Use case	Application
Real-time insights	Cisco Intelligent Packet Flow offers rich telemetry features, including microburst detection, congestion signaling, and in-band network telemetry (INT), providing deep visibility into network behavior at any scale.
Scale-out training	Adaptive load balancing ensures efficient GPU communication across racks, minimizing congestion and reducing job completion time for large-scale LLM training.
Inference and storage	Policy-based load balancing optimizes mixed workloads, enabling seamless simultaneous training, inference, and data

© 2025 Cisco and/or its affiliates. All rights reserved.

Use case	Application
	ingestion without compromising network performance.
Autonomous recovery	By detecting issues like link failures in real time and rerouting around degraded paths, Intelligent Packet Flow minimizes disruptions, ensuring continuous performance.

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.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **https://www.cisco.com/go/offices**.

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: https://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)

Printed in USA CSM-4796-YT 08/25

© 2025 Cisco and/or its affiliates. All rights reserved.