

建構AI Agent 時代的高韌性基礎架構



Andy Huang

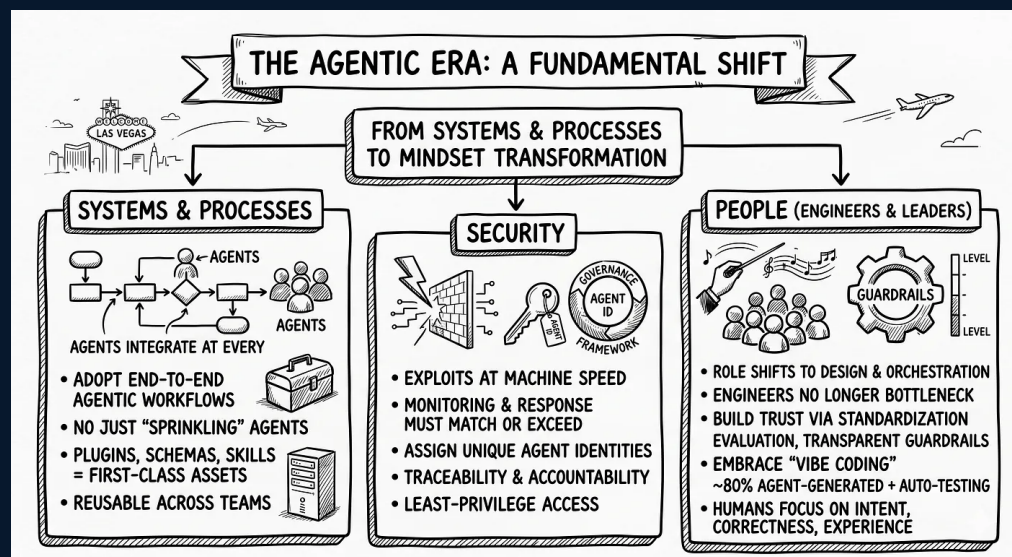
台灣思科AI事業發展總監

Agenda

1. AI Agent對企業帶來的3大衝擊
2. 思科給企業的3個建議:
 - 1) 確實做好Agent 資安
 - 2) 立即開始企業網路轉型
 - 3) 立即採用Cisco Secure AI Factory(SAIF)

我們已身處AI Agent 時代!

from 靜態知識
to 實時知識
to 智慧自主決策



agentic

tool-enabled

chatbots

Demo – AgenticOps Enterprise Use Case with DefenseClaw

– Optimized IT Operation Experience

Rober Feng
台灣思科技術總監



AI Agent 百花齊放

對企業IT帶來的衝擊有哪些？

Automation AI Agents by Monthly Search Volume				
1. Reworkd AgentGPT	2. FATHOM	3. Tactiq	4. Read	5. ADEPT
6. ChatGPT Tasks	7. bardeen	8. botpress	9. cohere	10. HARPA.AI
11. Lindy	12. Orby			
16. Grain	17. LALA			
21. airtop	22. SmythOS			
26. LLMStack	27. text.cortex			
31. Howie	32. Azure AI Agent Service			
36. Ai Agent App	37. fix ai			
41. Digitalfirst.ai	42. QuillCheck			
46. Cimba.ai	47. FL DE			

Conversational AI Agents by Monthly Search Volume				
1. Agentforce	2. D-ID	3. ElevenLabs	4. kustomer	5. Bland
6. Leena AI	7. Voiceflow			
11. Retell AI	12. Synthflow			
16. parloa	17. ada			
21. replicant	22. Recruiter			
26. droxy	27. FlowAI			
31. IrisAgent	32. LiveChatAI			
36. action.ai	37. yuma			
41. SalesAgents AI	42. Formshare			
46. Floatchat	47. B2B ROCKET			

Multi-Agent Platforms by Monthly Search Volume				
1. make	2. crewai	3. LlamaIndex	4. taskade	5. kore.ai
6. Magical	7. Spell	8. Relevance AI	9. aiOps	10. FlowiseAI
11. MindStudio	12. IBM watsonx.ai	13. Breeze	14. mem0	15. Dify
16. Reka	17. DRUID	18. AgentOps.ai	19. lyzr	20. Paradigm
21. Agents.AI	22. AISERA	23. WrenAI	24. AGENTX	25. aixplain
26. appypie ai builder	27. CAMEL-AI	28. OpenDialog	29. Agents Beta	30. Agentive
31. AI AGENT LAYER	32. airkit.ai	33. crackenagi	34. dup.ai	35. SENTIUS
36. INVICTA	37. aomni	38. HubRE.AI	39. Deagent AI Network	40. raia
41. akira AI	42. aixp.pro	43. Eidolon AI	44. giselle	45. X.AGENT
46. NOVUS	47. Neoleads	48. twin	49. amplemarket	50. GenFuse AI

Impact 1: Agent Security 令人擔憂

- Koi Security 對 2,857 個 ClawHub 技能進行檢測，發現其中 341 個為惡意項目，佔比 **11.9%**。
- 一項發表於 arXiv 的研究發現，在所分析的技能中，有 **26.1%** 至少存在一個漏洞。同一研究報告指出，**13.3%** 的技能具有資料外洩模式，**11.8%** 具有權限提升模式。
- 這些數據並不代表每個 OpenClaw 技能都是惡意的。但它們傳達了一個更實際的訊息：**該生態系統中已存在足夠多的風險行為**，因此在沒有前端安全控管的情況下，不應運行 OpenClaw。
- OpenClaw**無障密防護裸奔、代理失控、插件式搶劫、社交平台綁架**等問題層出不窮



RECENT OPENCLAW SECURITY STATS

Published research and recent incidents point to the same problem: without security checks before install and while it runs, OpenClaw can trust malicious skills, unsafe MCP servers, and hidden prompt attacks too easily.

CLAWHUB

11.9%

of audited ClawHub skills were malicious in the marketplace audit

341 malicious skills out of 2,857 audited. Koi report, February 2026.

MOLTBOOK

1.5M

API authentication tokens exposed, plus **35,000** email addresses left exposed in the same data set

4.03

average weaknesses per verified malicious skill

3

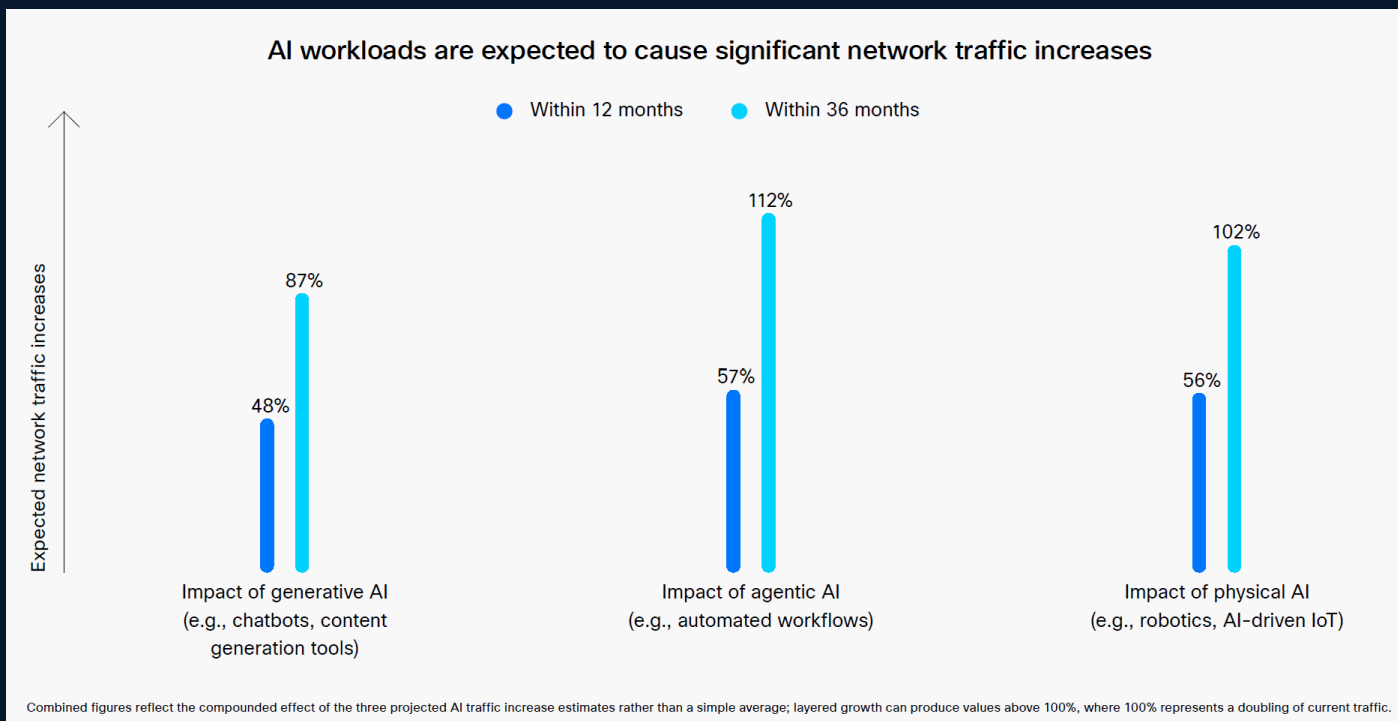
steps in the typical malicious-skill chain

54.1%

of verified malicious cases tied to one actor

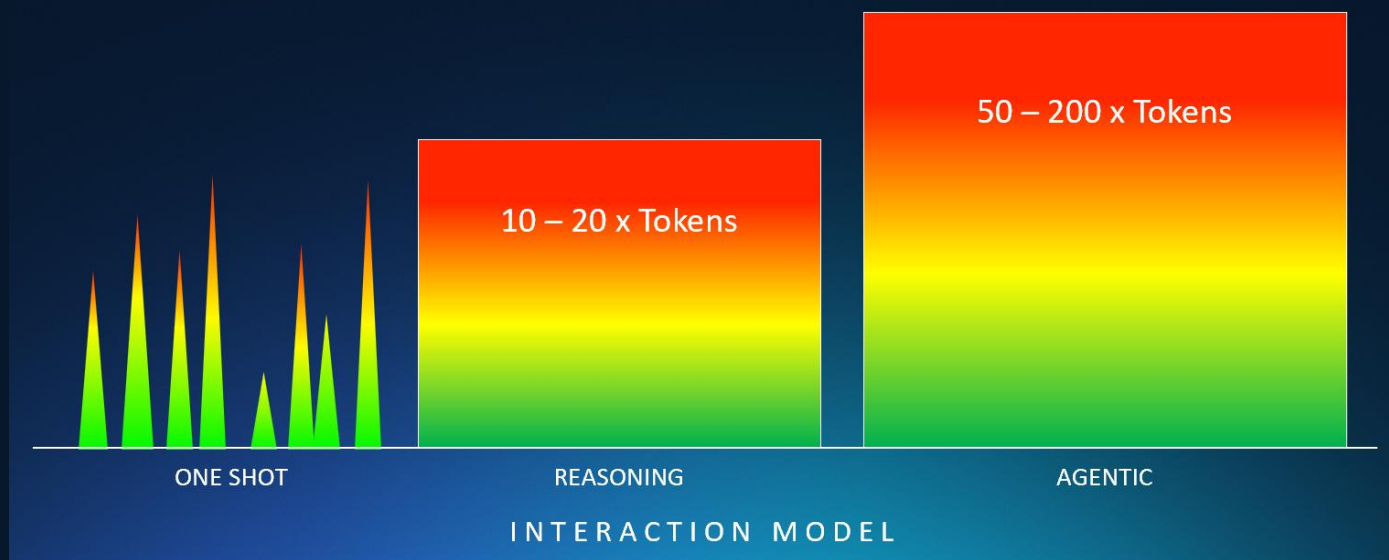
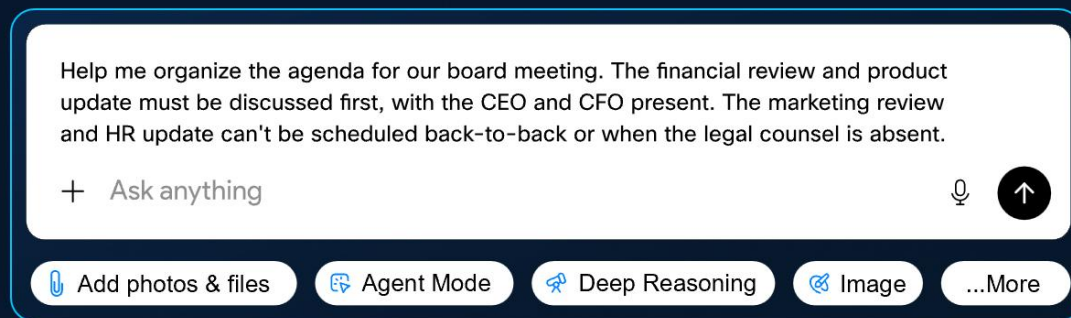
Impact 2: 對基礎架構韌性造成挑戰

AI Workloads 將會帶來網路流量顯著增加



1. 流量規模 (Volume): 3倍成長(未來3年)
2. 傳輸速度 (Velocity): AI 代理需要毫秒級的語境 (context)、嵌入向量 (embeddings) 與 API 狀態交換。
3. 代理間通訊 (Agent-to-Agent, A2A): AI 代理透過協作解決問題, 並深度依賴連續的事件驅動資料流, 而非傳統的靜態儲存。

Impact 3: Token 用量爆炸性的成長



Token = \$

特斯拉 (Tesla) 執行長馬斯克 (Elon Musk) 於 2026 年 7 月 6 日起正式實施新政策，將員工每人每週使用第三方人工智慧 (AI) 工具的支出上限硬性限制在 200 美元。這項被內部稱為對抗「狂燒 Token (Tokenmaxxing)」的緊縮政策，標誌著企業正從盲目擴大 AI 應用轉向嚴格的成本控制與財務治理

企業必須考慮 Model Routing
依 Agent 任務屬性建置地端推論服務

AI Agent 時代Cisco 的建議-1 確實做好Agent 資安



如何做好Agent 資安?

Agent 資安的三個做法

1. 零信任技術 (Zero Trust)

隨著海量自動化流量的產生，微分段 (Micro-segmentation) 與基於角色的存取控制 (Role Based Access Control) 必須直接內建於網路之中。

2. Agent沙盒

透過在作業系統內運行eBPF 程式沙盒以便在執行時期向作業系統添加額外的功能。作業系統借助即時編譯器和驗證引擎，確保與本機編譯一樣的安全性和執行效率。

3. 工具防護

透過防禦工具提供防護機制 (Guardrails)、工具檢測 (Tool Inspection)、安裝掃描 (Install Scanning)等功能。

Cisco Smart Switch

當資安與網路結合

網路

Nexus

Cisco 9300 Smart Switch



Dansberg

Deschutes

- 整合有狀態服務與網路
- 應用場景：雲端邊緣 (Cloud Edge)、區域型防火牆 (Zone-based Firewall)、數據中心互連 (DCI)、機架頂層 (Top of Rack)

安全

Hypershield

Cisco Security Cloud Control

Autonomous
Segmentation

Distributed
Exploit Protection

Future services

Platform

Workload and network enforcement points

Public Cloud | Private Cloud

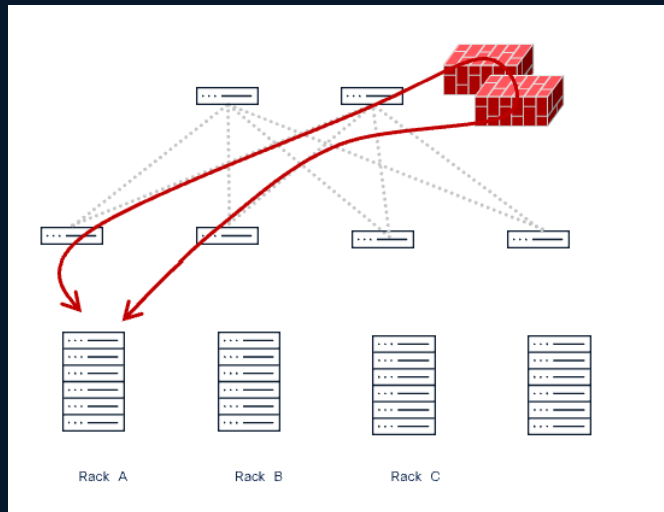
- 智慧安全策略部署
- 自動驗證更新 (雙數據平面)
- 跨工作負載與網路管控的統一架構——適用於公有雲與私有雲

全方位 數據中心安全

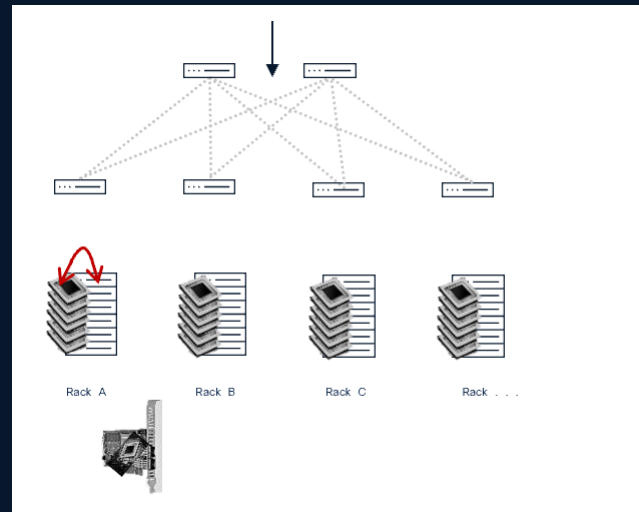
800G 有狀態服務效能與擴展能力

The value of Smart Switch as TOR switch

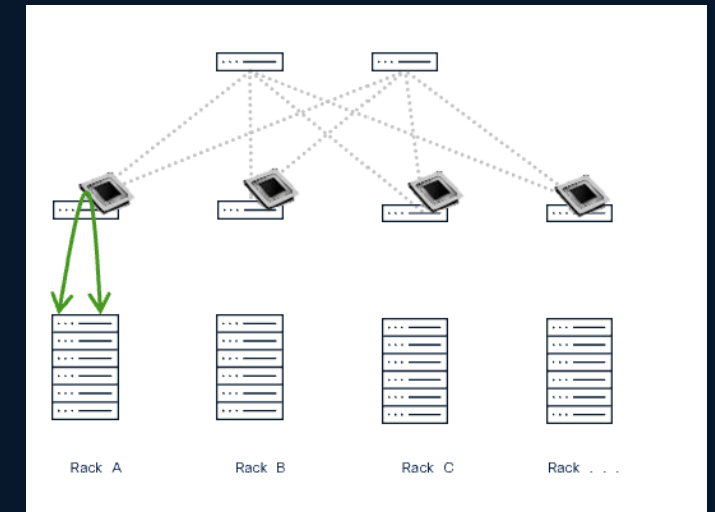
- 專用防火牆
- => multiple hops away from endpoints(距離端點經過多個節點)
- => easy to manage



- DPU SmartNIC
- very close to the endpoint: granular policies高度精確與詳細控管的規則
- many devices to manage



- DUP on Smart Switch
- close to the endpoints
- fewer enforcement points to manage compared to DPUs on the server(相較於伺服器上的 DPU, 需要管理的執行點更少。)



如何做好Agent 資安?

Agent 資安的三個做法

1. 零信任技術 (Zero Trust)

隨著海量自動化流量的產生，微分段 (Micro-segmentation) 與基於角色的存取控制 (Role Based Access Control) 必須直接內建於網路之中。

2. Agent沙盒

透過在作業系統內運行eBPF 程式沙盒以便在執行時期向作業系統添加額外的功能。作業系統借助即時編譯器和驗證引擎，確保與本機編譯一樣的安全性和執行效率。

3. 工具防護

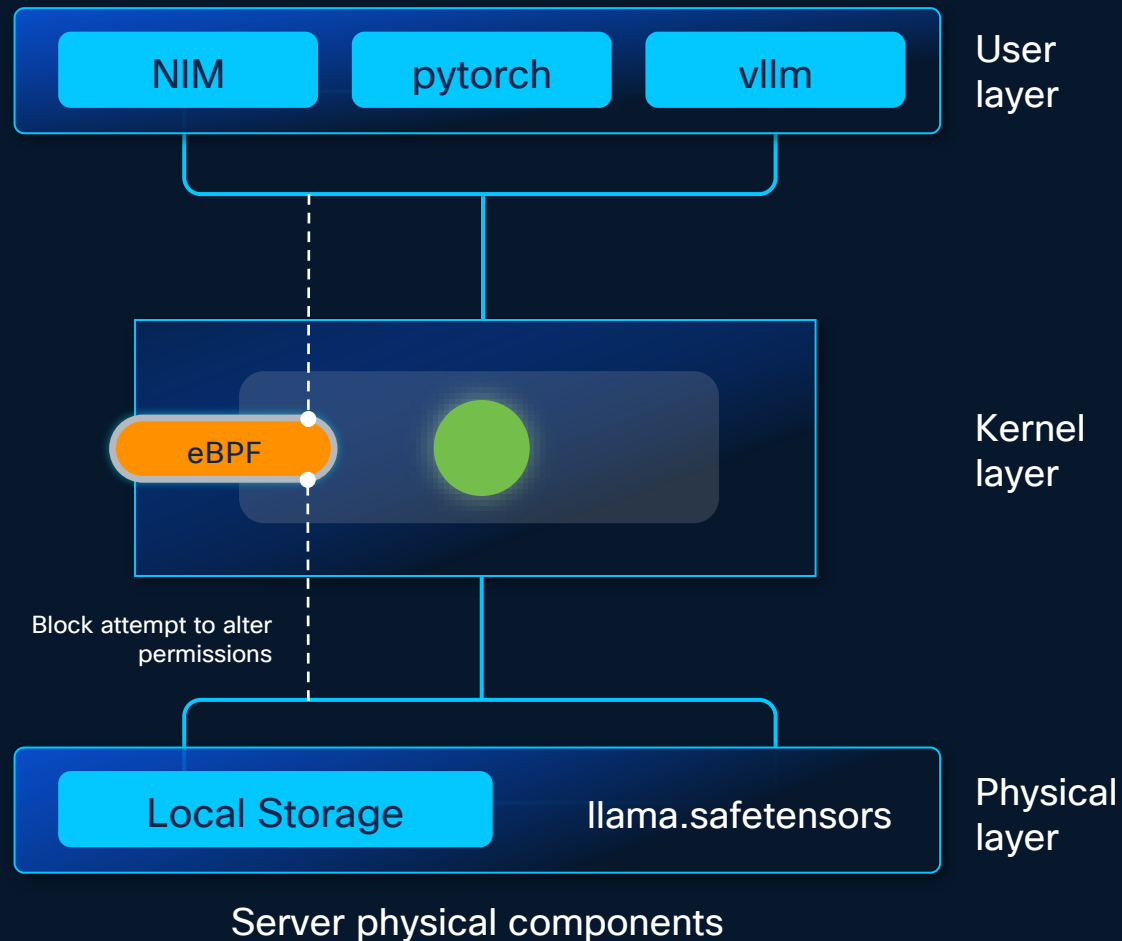
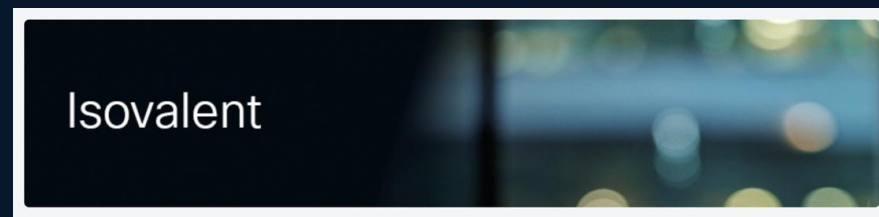
透過防禦工具提供防護機制 (Guardrails)、工具檢測 (Tool Inspection)、安裝掃描 (Install Scanning)等功能。

Securing AI Platform Runtime

安全執行環境與沙盒策略

沙盒策略 (Sandbox policies) 讓平台管理員能夠建立「沙盒」，藉此約束工作負載 (workloads) 所能執行的各項操作。

```
apiVersion: cilium.io/v1alpha1
kind: SandboxPolicyNamespaced
metadata:
  name: "block toxic behavior"
  namespace: "secure-ai-factory-nim-blueprint"
spec:
  syscalls:
    - list:
      - name: "sys_ptrace"
      - name: "sys_execve"
      - name: "sys_chmod"
      - name: "sys_chown"
      op: "In"
    actions:
      - type: "Post" #generate an event
      - type: "Block" #block the operation
      - type: "Signal" #kill the application
```



如何做好Agent 資安?

Agent 資安的三個做法

1. 零信任技術 (Zero Trust)

隨著海量自動化流量的產生，微分段 (Micro-segmentation) 與基於角色的存取控制 (Role Based Access Control) 必須直接內建於網路之中。

2. Agent沙盒

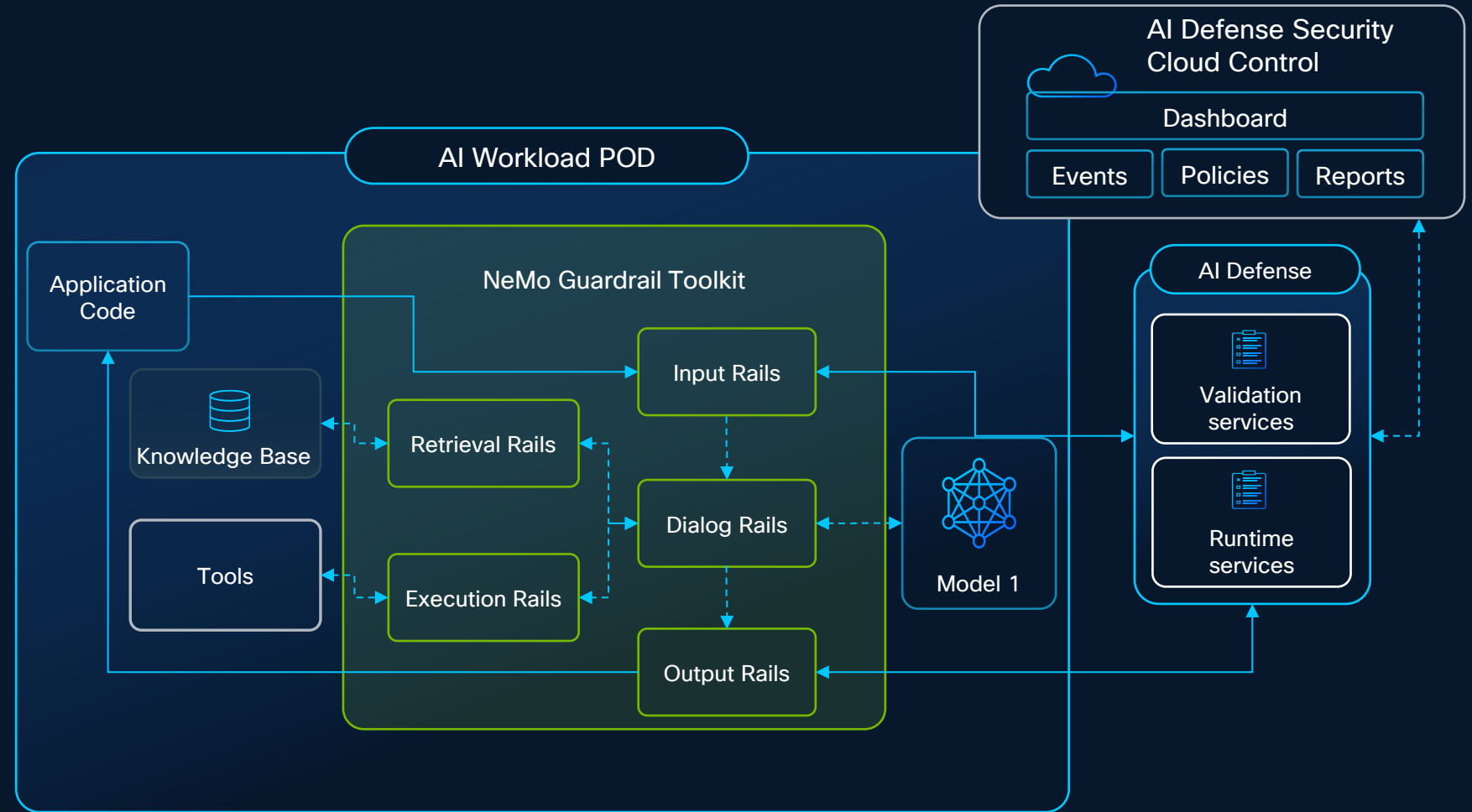
透過在作業系統內運行eBPF 程式沙盒以便在執行時期向作業系統添加額外的功能。作業系統借助即時編譯器和驗證引擎，確保與本機編譯一樣的安全性和執行效率。

3. 工具防護

透過防禦工具提供防護機制 (Guardrails)、工具檢測 (Tool Inspection)、安裝掃描 (Install Scanning)等功能。

AI Defense & NeMo Guardrails Integration

- AI Defense provides input & output guardrails via API disposition
- Common guardrail policy for on-premises and cloud deployments.
- Supports additional guardrail types included in the NeMo Guardrail Toolkit
- [Nvidia documentation link](#)

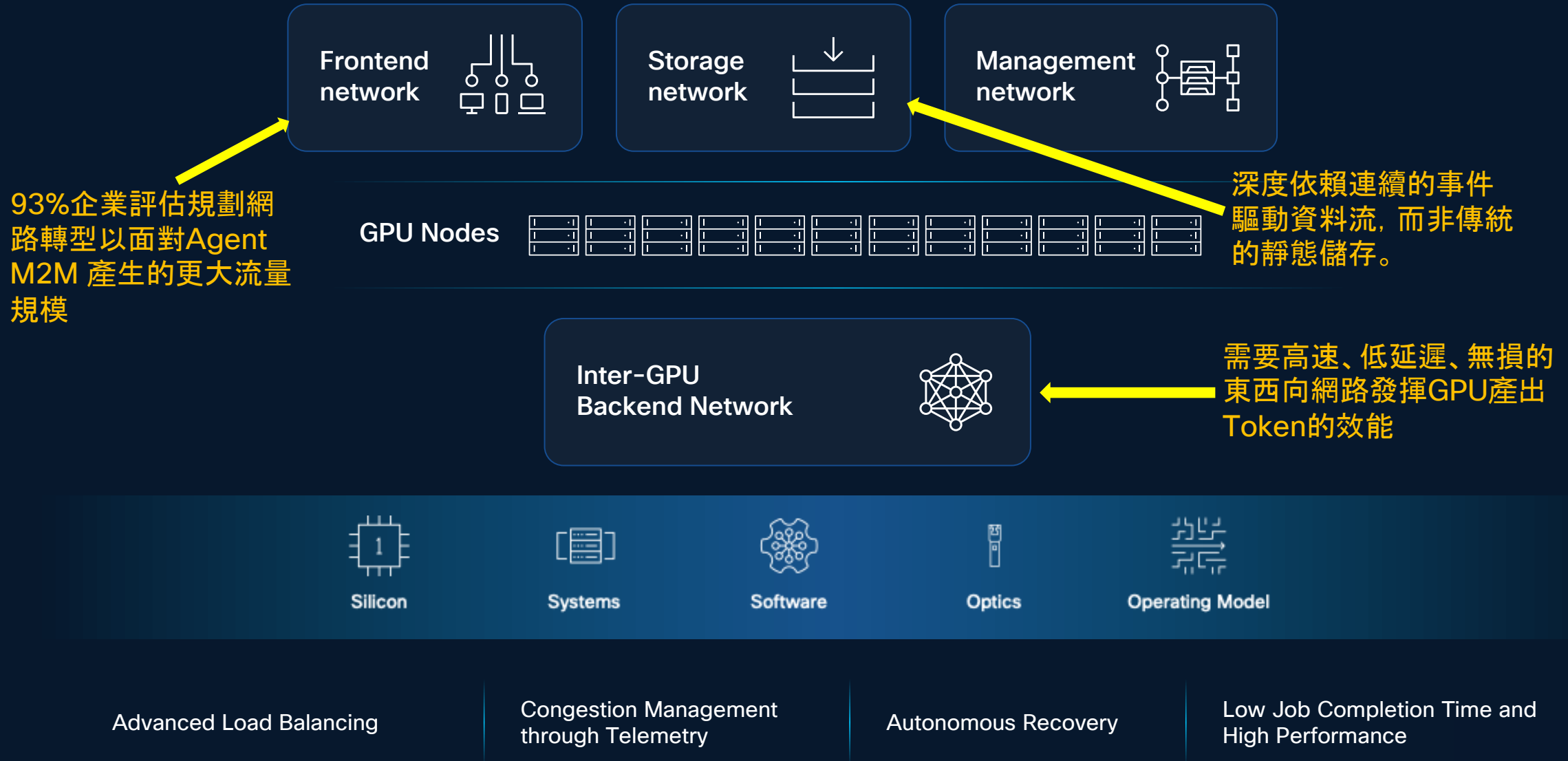


AI Agent 時代Cisco 的建議-2 立即開始企業網路轉型

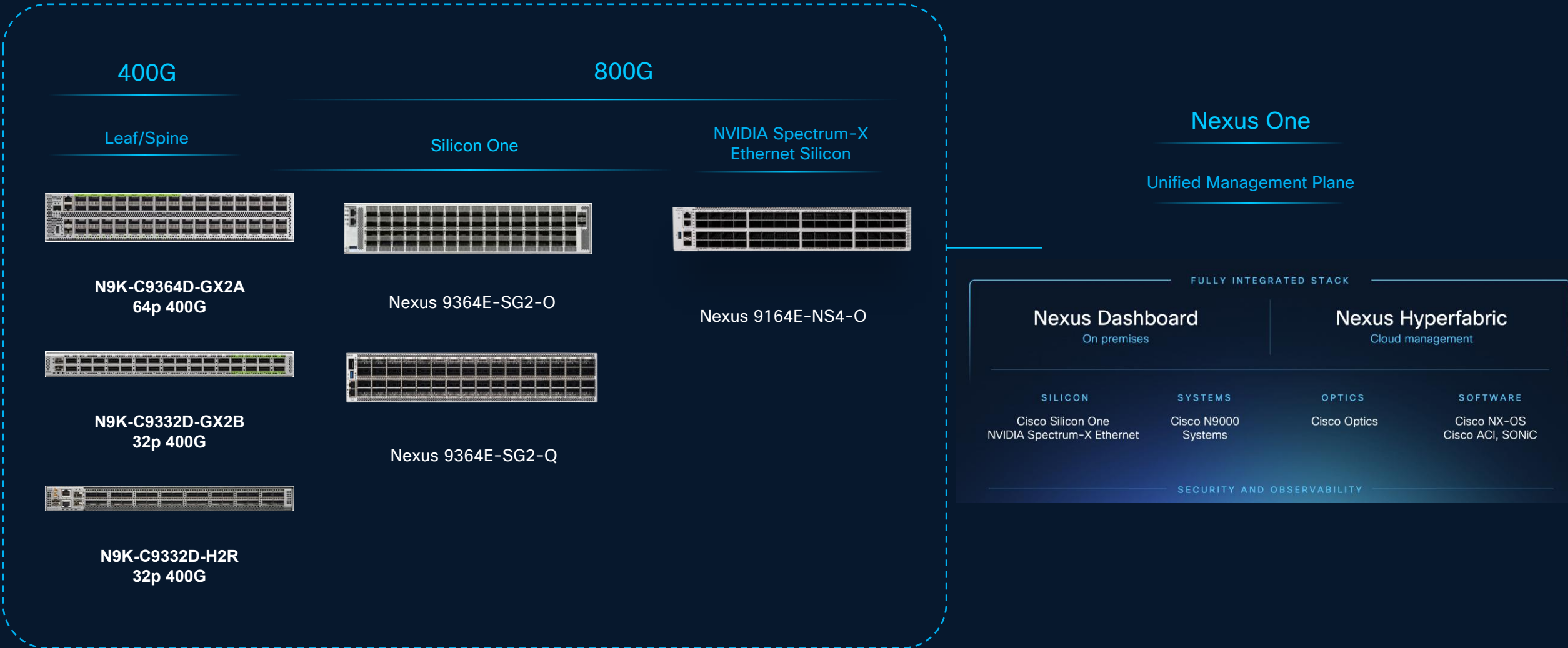


推進網路轉型 以提升AI 效能!

Transforming AI Networks for Performance and Scale



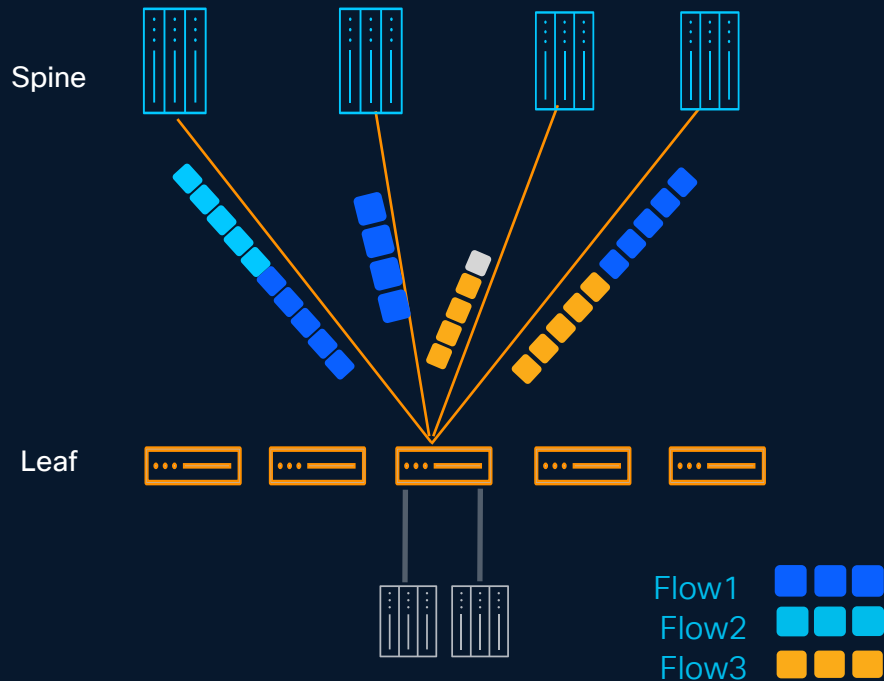
Cisco AI Networking 完整的產品線



What Makes Cisco AI Networking Unique? – Intelligent Networking

Maximize GPU utilization, optimize KV cache performance

Dynamic Load Balancing



Intelligent Packet Flow

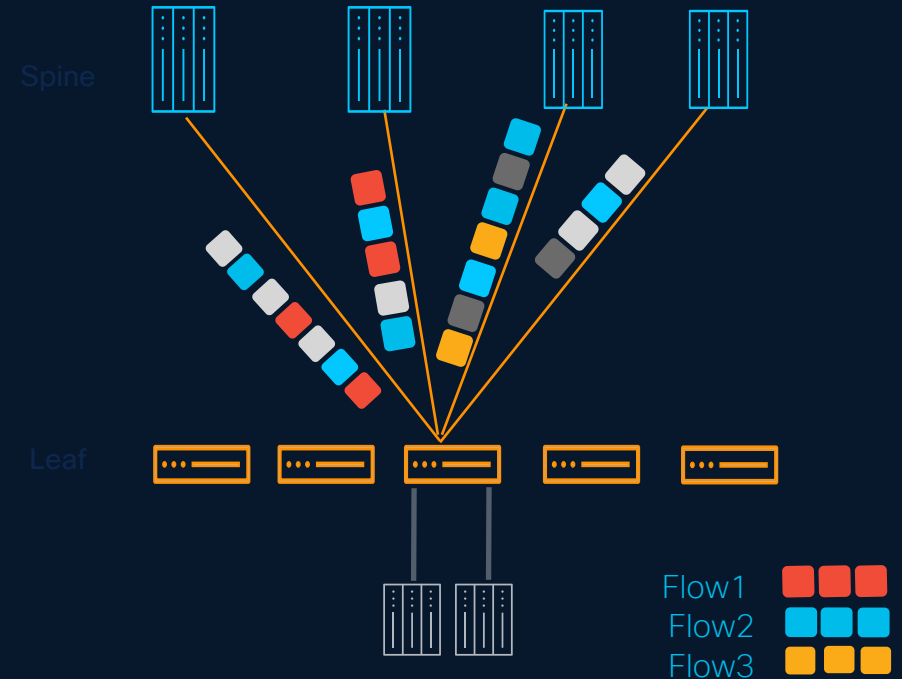
Cisco innovation,
works with NVIDIA
Spectrum-X



Spectrum-X

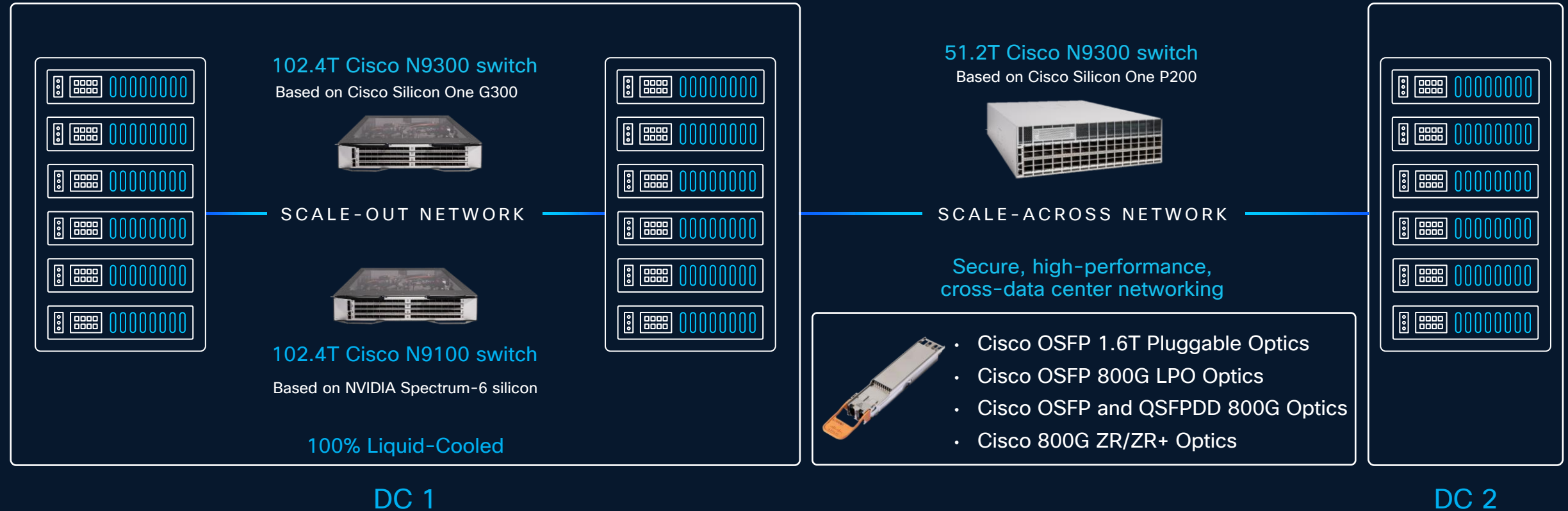
NVIDIA innovation,
available with Cisco
switches

Per-Packet/Packet Spray



Scale-out and scale-across networking

Built on high performance Silicon, Systems and Optics



Benchmark proven AI/ML performance: [Cisco Enterprise & Cloud Reference Architecture](#)

Compliant with NVIDIA ERA and NCP RA [with Spectrum-X technology](#)

CISCO NEXUS ONE
UNIFIED MANAGEMENT CONTROL PLANE



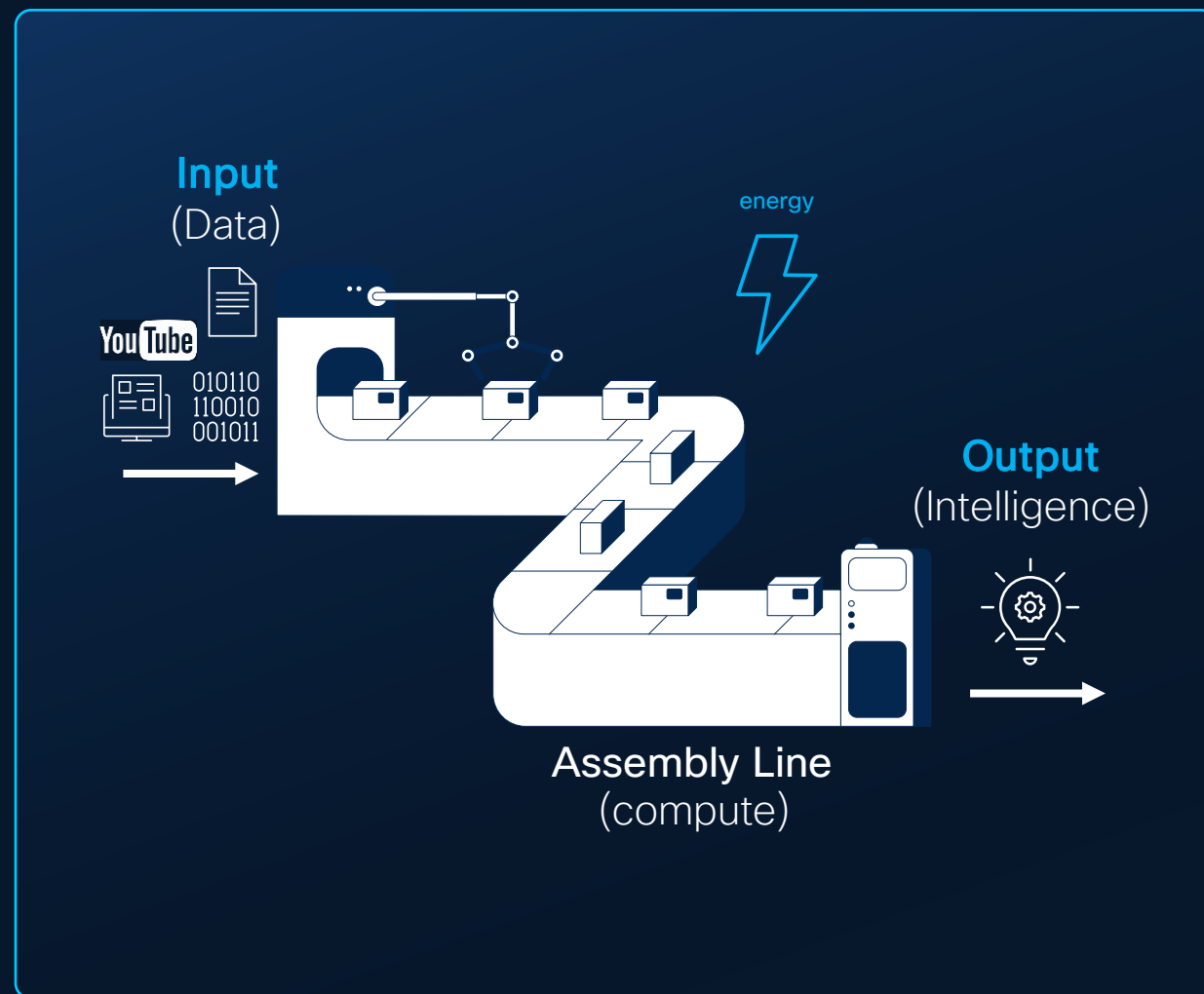
AI Agent 時代Cisco 的建議-1 立即採用Cisco Secure AI Factory



What is an AI Factory?

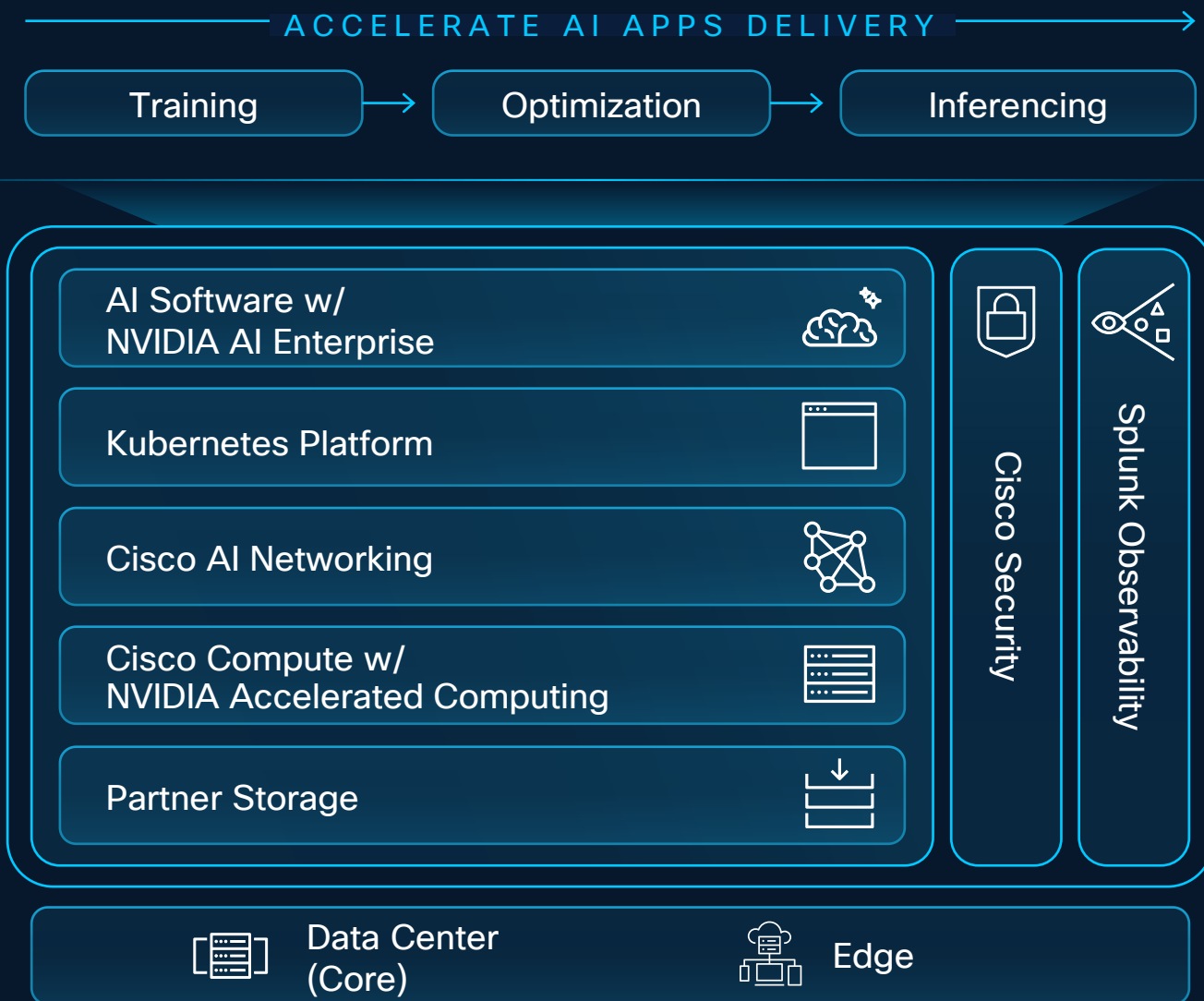
負責執行大型語言模型（LLM）推論、處理並生成 **Token** 的基礎設施

全球各地的組織都在思考
如何以最快速、安全且具
成本效益的方式生成
Token。



Cisco Secure AI Factory with NVIDIA

結合高效能基礎架構、全方位安全防護與可觀測性的**模組化**參考設計

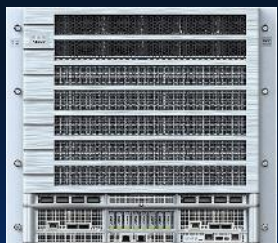


Sample Bill of Materials (C880A w/ B300)

128 GPU Secure AI Factory with NVIDIA - Including Backend Network & Optics Scalable to 256 GPUs

128 GPU Secure AI Factory with NVIDIA - Including Backend Network & Optics Scalable to 256 GPUs

Compute - AI Nodes



16 - Cisco UCS C880A
NVIDIA HGX 2-8-10-800 Design
2x B3220 2x200G
8x CX-8 (1x 800G OSFP -> 2x400G cables)

128 - NVIDIA B300 HGX GPU
(SXM Form Factor, 8 per node)

Security - AI Defense Nodes



3 - Cisco UCS C845A
(NVIDIA MGX 2-8-5-200G Design)

24 - NVIDIA L40S PCIe GPU
(PCIe Form Factor, 8 per node)

Network



12 - Cisco Nexus 9364E 800G
(8 Leaf Switches, 4 Spine Switches)



6 - Cisco Nexus 9364D 400G
(4 Leaf Switches, 2 Spine Switches)



2 - Cisco Nexus 9332D
(VAST Backend)



4 - Cisco 93108TC-FX3
(Management Switches)



884x 200 to 800G Optics | 698x Fiber Cables
(counts will vary with final design)

Management Nodes



5 - Cisco UCS C225 M8
(Control Cluster K8s & SLURM)



3 - Cisco UCS C225 M8
(Nexus Dashboard Cluster)



3 - Cisco UCS C220 M7
(AI Defense Mgmt Cluster)

Software

152 - NVIDIA Enterprise

128 - Run:ai

20 - AI Defense

39 - Intersight

24 - DCN Subscription

12 - Spectrum-X Add-on

261 - Cloud Protection

19 OpenShift Container Platform

152 - Red Hat AI Accelerator

11 - VAST Capacity

348 - VAST Cores

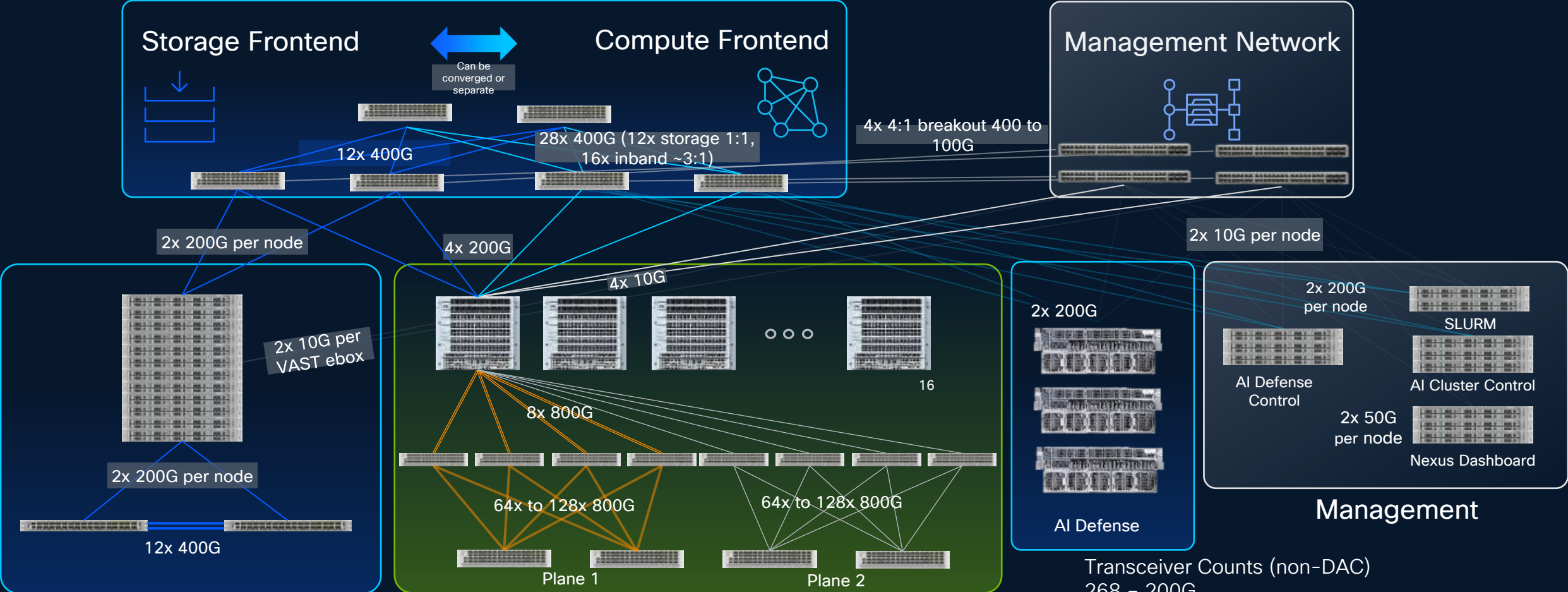
Storage



12 - Cisco UCS C225 M8
(VAST EBox Cluster)

Sample Bill of Materials

128 GPU (16 node) Secure AI Factory with NVIDIA – High Level Hardware Architecture



Transceiver Counts (non-DAC)
268 – 200G
104 – 400G
512 – 800G
698 – Fiber Cables
*counts will vary with final design
*Nexus dashboard has 50G VIC

結論: AI Agent 時代企業最需在意的3件事

1



Security & Observation



Cisco 提供Agent 資安
完整解決方案!

2



Infra resilience



立即開始企業網路轉型!

3



Token Cost



立即採用Cisco SAIF

Call to action

1

到 Cisco Connect Demo Booths 了解更多!

2

造訪 Cisco “**Securely scale AI**” Website.

High-performance networking, security, observability, and AI-native tools to power AI and operate the enterprise.

<https://www.cisco.com/site/us/en/solutions/artificial-intelligence/index.html>

3

想獲得更多細節，請立即聯繫您的思科業務窗口!

Thank you!