Unlock the Future: Leveraging Generative AIOPs for Enterprise Innovation and Performance

Sanjit Aiyappa – Cisco, Director Solutions Engineering
Joel Jose – Cisco, Senior Solutions Engineer

AIHUB - 1006

Listening to the market

75%

of large enterprises will rely on AI-infused processes by 2026

GenAI is expected to be the

#1 driver

of **infrastructure investments** over the next 18 months

$300B

global spending on AI by 2026

97%

of companies say the **urgency to deploy AI-powered technologies** has increased

$15.7T

Potential contribution to global economy by 2030

GenAI has potential to generate

$2.6 to $4.4 trillion

in **value across industries**[2]

[1]IDC Artificial Intelligence Systems Spending Guide
[2] McKinsey & Company
[3]Canalys estiates & forecasts, August 2023
Gartner, May 2024

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

Capitalizing on AI for channel partners

AI partner opportunity
$15.4 billion (2023)[3]
$158 billion (2028)

Reselling, co-selling, and upselling AI products with services around those products

AI software development

Advanced data services that use/enable AI

AI services

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

Heeding the voices of our partners

Partner AI readiness index

Biggest challenges supporting customers to deploy AI

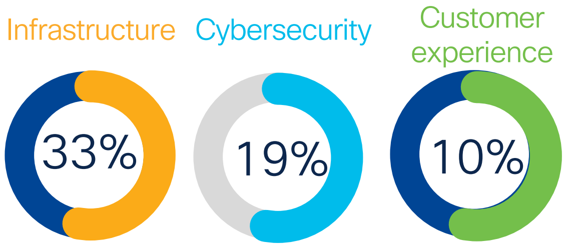
- Lack of experience in new areas of tech deployment 62%
- Lack of knowledge of systems and processes 56%
- Lack of knowledge of available technologies 53%

27% of partners believe

76-100%

of **revenue** will come from **AI technologies** in next 4-5 years

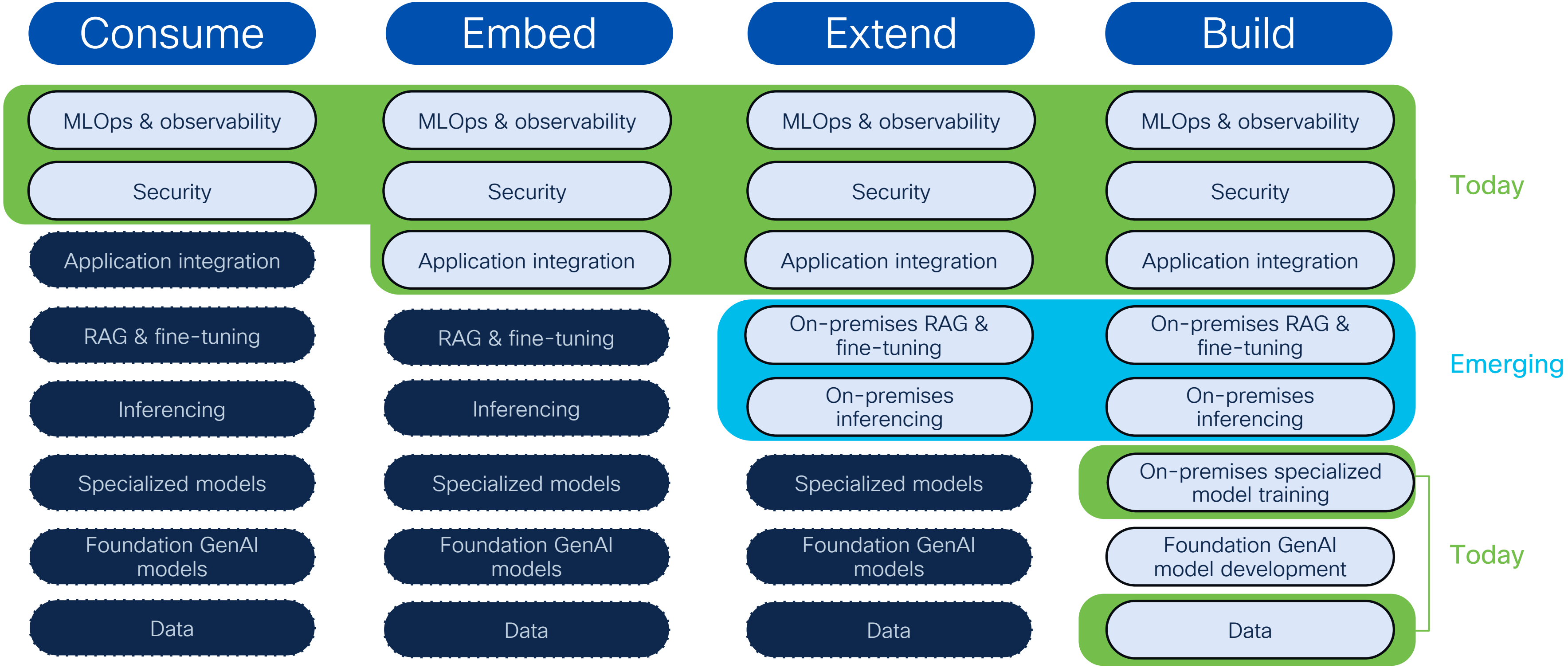
Biggest revenue drivers ALL aligned to Cisco



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

Where is our opportunity?

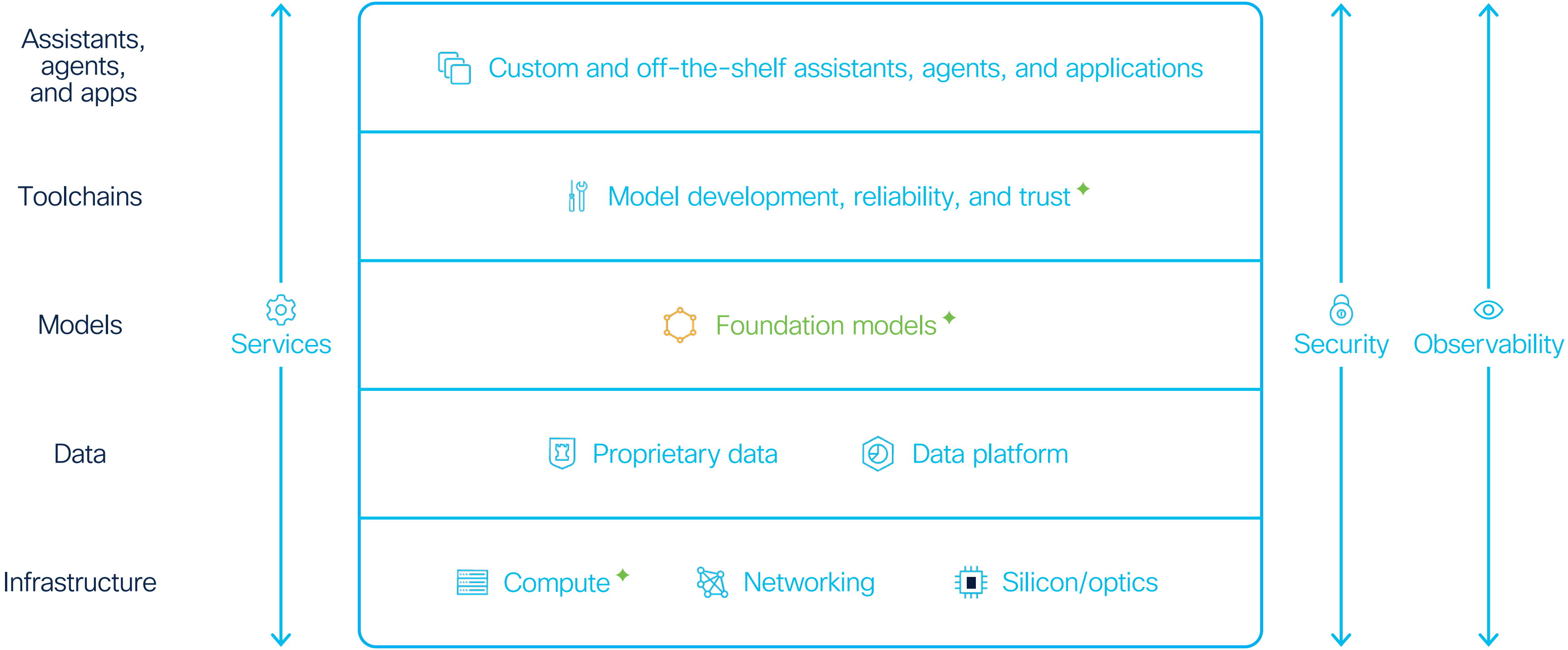
Gartner AI consumption models and customer journey



Consume	Embed	Extend	Build
MLOps & observability	MLOps & observability	MLOps & observability	MLOps & observability
Security	Security	Security	Security
Application integration	Application integration	Application integration	Application integration
RAG & fine-tuning	RAG & fine-tuning	On-premises RAG & fine-tuning	On-premises RAG & fine-tuning
Inferencing	Inferencing	On-premises inferencing	On-premises inferencing
Specialized models	Specialized models	Specialized models	On-premises specialized model training
Foundation GenAI models	Foundation GenAI models	Foundation GenAI models	Foundation GenAI model development
Data	Data	Data	Data

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

Cisco supports customers across the AI stack



Assistants, agents, and apps	Custom and off-the-shelf assistants, agents, and applications
Toolchains	Model development, reliability, and trust ✦
Models	Foundation models ✦
Data	Proprietary data Data platform
Infrastructure	Compute ✦ Networking Silicon/optics

Current Cisco coverage ✦ With partners

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

Map of AI in, on, and @ Cisco

AI in Cisco

- Collaboration
- Security
- Networking
- Observability
- CX

- Summary
- Next step
- Document
- Configure
- Analyze

- Contact center
- Vidcast
- Meetings
- Messaging
- Firewall
- SSE
- XDR
- Duo
- Meraki
- Catalyst Center
- NDB
- ThousandEyes

AI on Cisco

- C8000 | N9300
- UCS-X
- Intersight
- ID intelligence
- User protection
- Panoptica
- FSO
- Motific

- Scalability
- Bandwidth
- Reliability
- Latency

- Large mem
- GPU density
- High TP
- PU lifecycle

- Zero trust
- Data security
- LLM security
- API security

- MLOps
- Governance
- Biz insights
- Trust & safety

- Network
- Compute
- CI & HCI
- Security
- Observability

AI @ Cisco

- Documents
- Convergence
- Next step

- Summarization & search
- Account intelligence
- Territory AP & next steps

- SalesConnect RAG API
- AI teammate
- Beyond ready account prioritization & next-best action

Responsible AI framework

- Controls
- Transparency
- Fairness
- Accountability
- Privacy
- Security
- Reliability

Evolving concerns

- Hallucinations
- Harmful Output
- False content
- Bias
- Data theft
- Data leakage
- Data poisoning
- Sustainability

Industry leadership

- AI-enabled ICT workforce consortium

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

Why AIOPs

15-30%
cost savings

30%
increased productivity

≠70%
increase in FTTR

5–10%
revenue growth

-25%
compliance fines

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

AIOPs outcomes

Top use cases

- Anomaly detection
- Root cause analysis
- Predictive analytics
- Change impact analysis
- Capacity planning
- Security monitoring and compliance
- Incident management
- Performance monitoring and optimization
- Change management

Benefits

- Operational efficiency
- Cost reduction
- Faster problem resolution
- Proactive monitoring and predictive analytics
- Enhanced scalability
- Optimized resource utilization
- Reduce alert noise and false positive
- Increase service reliability
- Data-driven insights for decision makers

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

AI-driven portfolio today

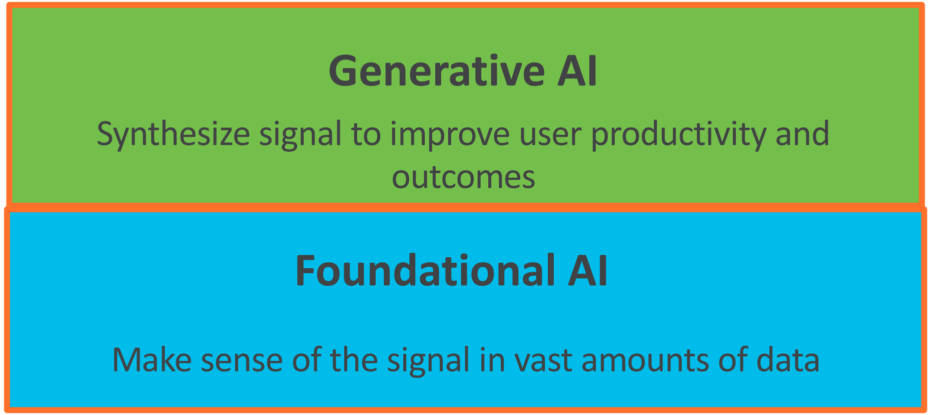
AI with Cisco for MSP

AI ASSISTANT

	Cloud and AI	Networking Cloud	Collaboration	Security	Observability
AI-Driven Portfolio	• Enabled by Silicon One Scheduled Fabric Ethernet Solution • Digital Experience Monitoring & predictive analytics • AI PODs • Nexus with RoCE • Hyperfabric AI/ML • Hyper Shield	• Predictive Path Recommendation (SD-WAN) • Bandwidth Forecast (SD-WAN) • Machine Reasoning Engine (MRE) • AI Network Analytics • AI-Driven Baselining • AI-Enhanced RRM • Anomaly Detection (SD-WAN) • Client Analytics and RCA	Cloud Contact Center • AI Powered Analytics • AI Call Summaries • AI Scripted Agents • Automatic CSAT Scores • Agent Wellness Webex Calling • Catch Up and Summarization Webex Meetings • AI Meeting Summary • AI Codec for Voice and Video	• Extended/Managed Detection & Response • AI Assistant Experience (inc. Firewall, XDR, & cross-portfolio) • SPLUNK (SOC) • Autonomous Actions (incl. incident response, recs, & automation)	• Statistical Modeling with Baselining • Anomaly Detection with dynamic baselining • Intelligent Automation • SPLUNK – Oly & ITSI, AI assistant for SPL • Prompt interface with AI-powered workflows • AI assistant for summarization
Partner led usecases	• Network Management & Deployment • Service Assurance • Anomaly Detection • Change Automation • Predictive Internet Insights • Customized AI Models for Compliance Monitoring • Machine Learning-Enhanced Predictive Analytics for Customer Satisfaction	• Anomaly Detection and Alert Prioritization • Predictive Incident Management • Incident Root Cause Analysis (RCA) with Domain-Specific Insights • Automated, Contextualized Remediation Actions • Dynamic Incident Correlation and Noise Reduction	• NLP Chat bot for ops workflow (knowledge Base Integration) • Custom ML bots integration with CC. • AI model to detect voice/video issues across network path • Proactive notification and auto meeting room booking for employee behalf. • Intelligent workflows to see usage consumption. of meeting rooms • Predictive insights on network usage for Voice and video calls.	• Proactively hunting - searching for hidden threats to prevent the attack from happening • Understanding and Mitigating Advanced Threats • Executing the incident response plan • Define automation and playbooks	• Intelligent root cause analysis with event prioritization • Realtime business health monitoring with data tagging • Data security posture management • Adaptive thresholding for optimized SLAs

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

Types of AI



Generative AI

Synthesize signal to improve user productivity and outcomes

Foundational AI

Make sense of the signal in vast amounts of data

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

Rightsizing GenAI

1. **Understand Workload Needs**
   - Training vs. Inference
   - Model Complexity
   - Batch Size/Dataset

2. **Profile & Benchmark**
   - Tools: NVIDIA Nsight, TensorBoard, PyTorch Profiler.
   - Key Metrics: GPU/CPU utilization, memory, I/O bottlenecks.

3. **Optimize Infrastructure**
   - Compute: Use GPUs (e.g., A100 for training, T4 for inference) or accelerators (TPUs, FPGAs).
   - On-Prem vs. Cloud

4. **Scale Efficiently**
   - Horizontal
   - Vertical
   - Elastic

5. **Enhance Resource Usage**
   - Quantization/Pruning: Optimize model efficiency.
   - Data Pipelines: Caching, sharding for faster I/O.
   - Orchestration: Kubernetes

6. **Monitor Continuously**

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

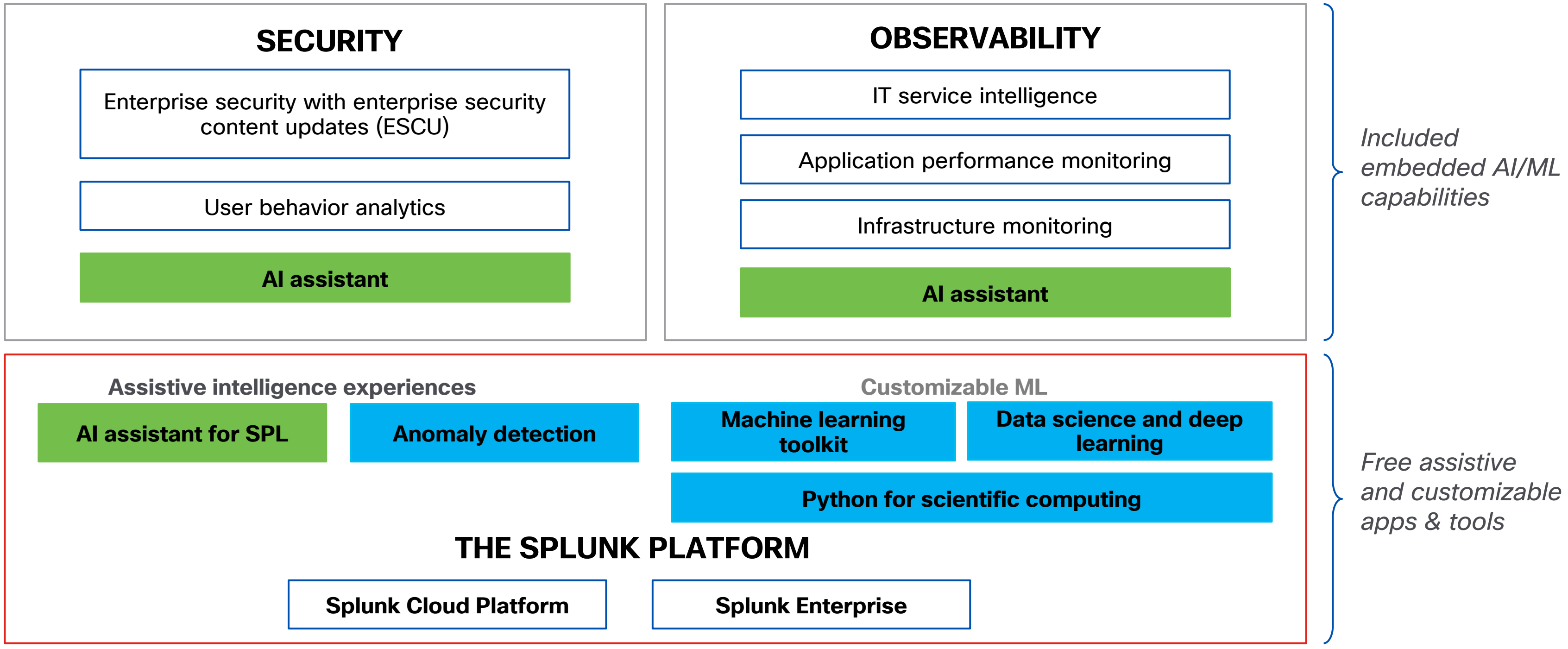
Advanced RAG Implementations

- **Hybrid Retrieval (Dense + Sparse)**: Combines the strengths of dense (e.g., vector-based) and sparse (e.g., BM25) methods for more robust and accurate document retrieval.
- **Iterative RAG (Re-RAG)**: Refines outputs by iteratively retrieving additional context and updating the response for higher accuracy and reduced ambiguity.
- **Memory-Augmented RAG**: Integrates short-term and long-term memory mechanisms to maintain context and personalize interactions over time.
- **Multimodal RAG**: Extends RAG to handle multiple modalities, such as text, images, and videos, enabling applications like visual question answering and multimodal search.
- **RAG with Feedback Loops**: Incorporates human-in-the-loop or automated feedback to iteratively improve retrieval and generation quality, using reinforcement learning or fine-tuning.

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

Powered by Splunk AI

Product overview



SECURITY

Enterprise security with enterprise security content updates (ESCU)

User behavior analytics

AI assistant

OBSERVABILITY

IT service intelligence

Application performance monitoring

Infrastructure monitoring

AI assistant

Included embedded AI/ML capabilities

Assistive intelligence experiences

AI assistant for SPL

Anomaly detection

Customizable ML

Machine learning toolkit

Data science and deep learning

Python for scientific computing

THE SPLUNK PLATFORM

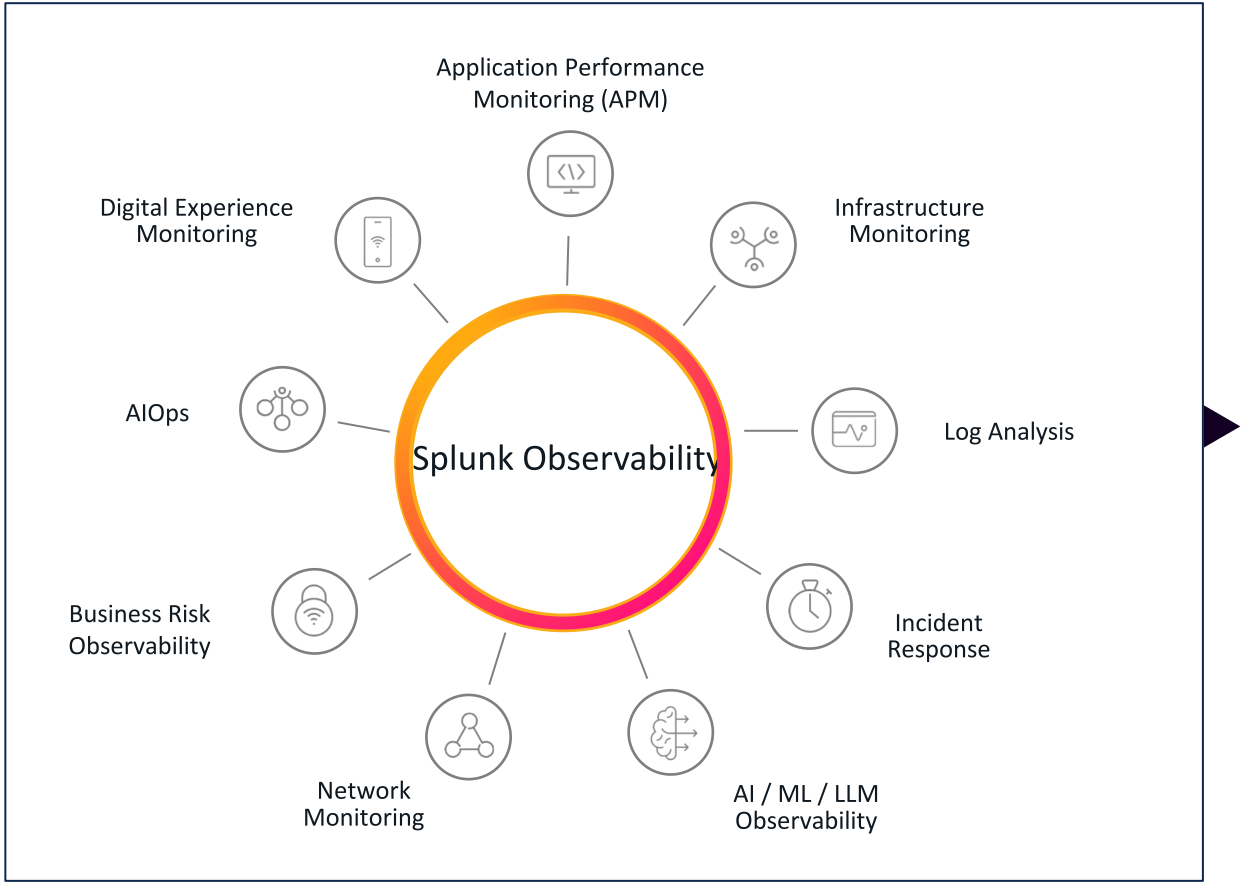
Splunk Cloud Platform

Splunk Enterprise

Free assistive and customizable apps & tools

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

Splunk Observability portfolio

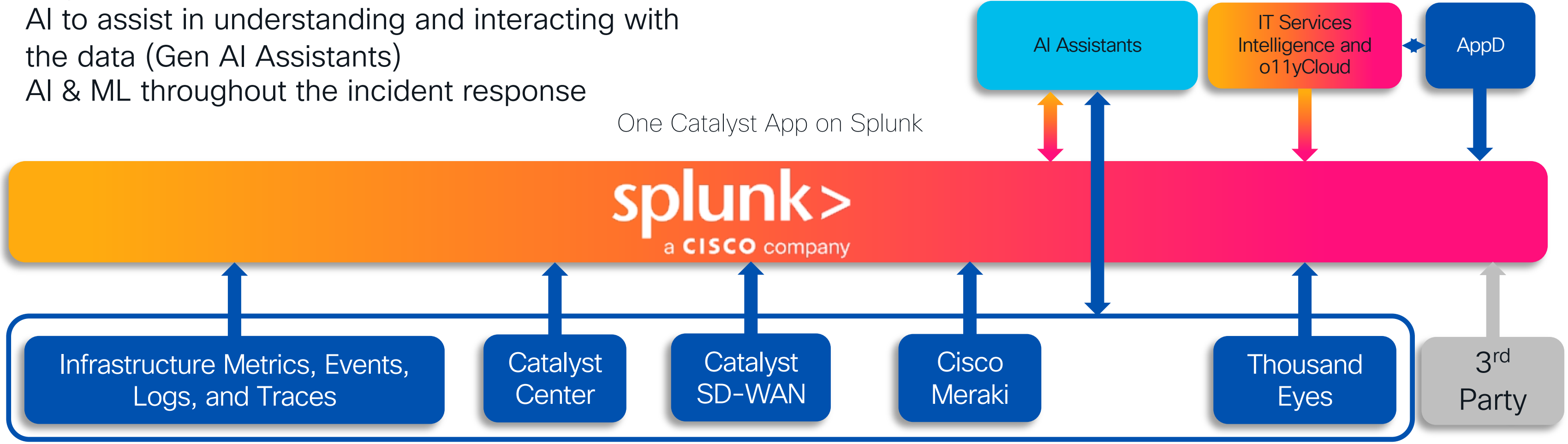

Application Performance Monitoring (APM)
Digital Experience Monitoring
Infrastructure Monitoring
AIOps
Splunk Observability
Log Analysis
Business Risk Observability
Incident Response
Network Monitoring
AI / ML / LLM Observability

- Real-Time Insights
- AI Powered
- Enterprise Grade
- Open Telemetry Native
- Extensible
- Cross MELT
- Business Context

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

Cisco Networking + Splunk: AI & ML Based

- AI to process the data and find the issues (AI, ML)
- AI to assist in understanding and interacting with the data (Gen AI Assistants)
- AI & ML throughout the incident response



AI Assistants
IT Services Intelligence and o11yCloud
AppD
One Catalyst App on Splunk
splunk>
a cisco company
Infrastructure Metrics, Events, Logs, and Traces
Catalyst Center
Catalyst SD-WAN
Cisco Meraki
Thousand Eyes
3rd Party

Essential for correlation of events across disparate data types and sources

Radio Resource Mgmt
ThousandEyes Agent integrations

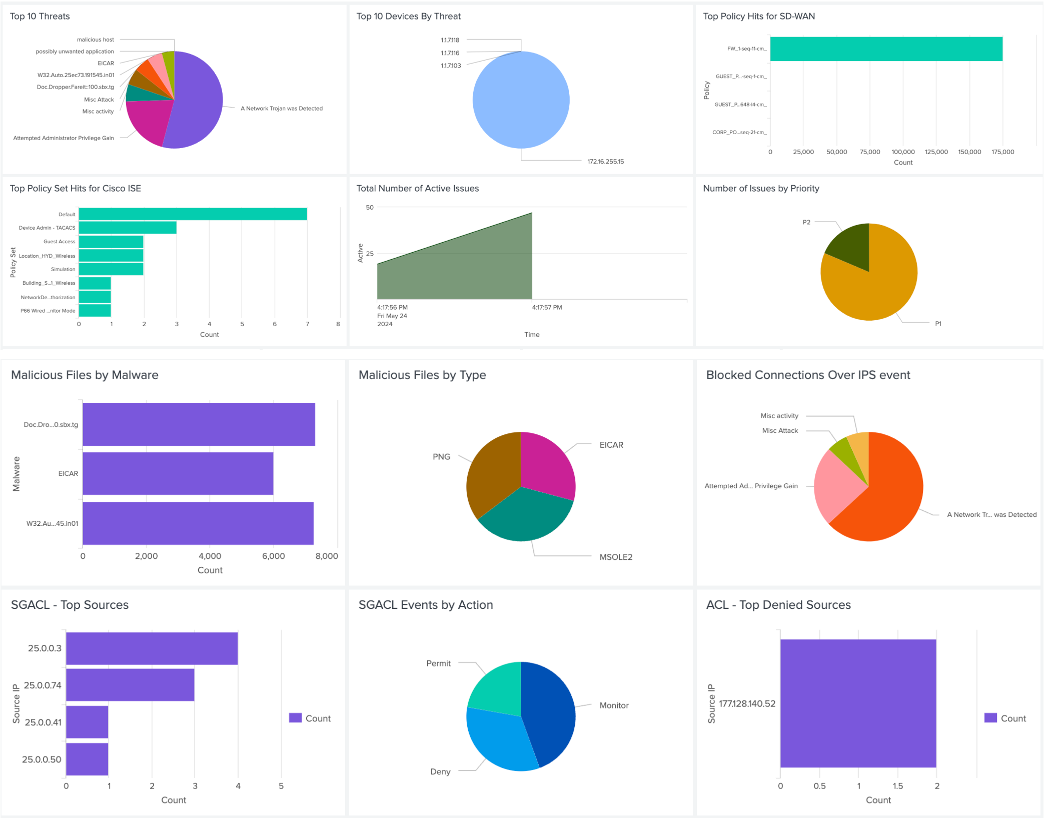
WAN Insights with Predictive Path analysis, Bandwidth Forecasting
Sites, Clients, Circuits, Apps dashboard
Traffic flow and App distribution patterns

Radio Resource Mgmt,
WiFi 3D Analyzer
WiFi 6/6E planning tools
Endpoint Analytics
PoE Analytics
Cisco ecosystem connectivity
Site & Event Analytics

WAN Assurance
Active Application Monitoring
Internet Outage

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

Catalyst Common Splunk Application – Outcomes



Benefits

- Holistic view of Network, Security and Identity events
- Long time data retention for compliance needs
- Charts combine data from multiple sources for ease of visualization (ex Top Threats combine threats from WAN and ISE)

Outcomes

- Improve troubleshooting by tracking device and endpoint issues
- Reduce Mean Time to Resolution (MTTR) by quickly identifying critical events
- Single repository for NGFW events (Malware Protection, Intrusion Prevention, URL-Filtering) and Security Advisories
- Secure grip over the network by tracking Access Control List (ACL) entries across WAN and LAN

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

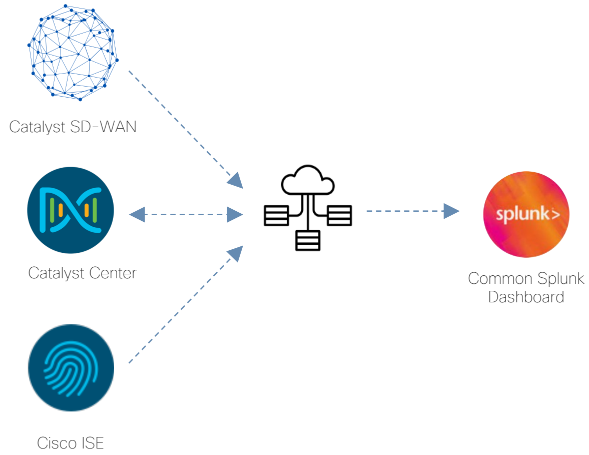
Catalyst Common Splunk Application – Use Cases

Consolidated Visibility

- Consume data from different solutions (DNA/SD-WAN/ISE) using standard data sets
- Common dashboard that visualizes data from all solutions

- Outcomes : Realtime Monitoring, Historical Insight, Network and Client Insight, Security Insight, Compliance & Security advisory

MVP Focus



Playbook Driven Response

- Playbooks that look for certain event triggers and generate API calls back to the appropriate domain (SD-WAN/DNA/ISE)

- Outcome: Automated Event Response

Analytics Dashboard

- Correlate information from the various solutions and establish baselines

- Outcome: Detect and report on anomalies based on deviation from the baselines

Splunk Ecosystem Partner

- Trigger Notifications to 3rd Party Platforms
- Splunk Already Integrates with (1,000+) existing apps

- Outcome: Integration with third party event management systems

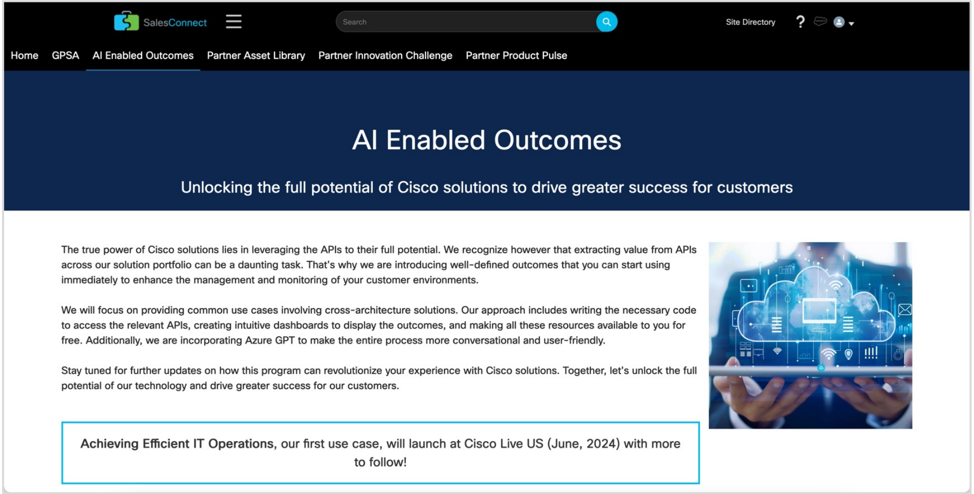
© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Demo

- MSP BOT with troubleshooting using GenAI
- Advanced Cost-effective RAG for Enterprise
- Interactive API documentation using GenAI
- Alert reduction and troubleshooting using GenAI

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

Cisco supports to build outcomes



AI Enabled Outcomes

Unlocking the full potential of Cisco solutions to drive greater success for customers

The true power of Cisco solutions lies in leveraging the APIs to their full potential. We recognize however that extracting value from APIs across our solution portfolio can be a daunting task. That's why we are introducing well-defined outcomes that you can start using immediately to enhance the management and monitoring of your customer environments.

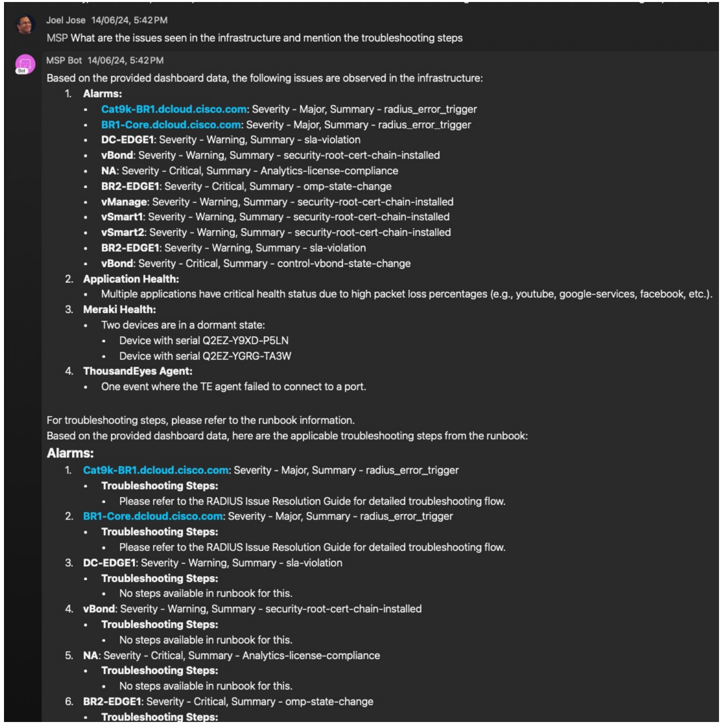
We will focus on providing common use cases involving cross-architecture solutions. Our approach includes writing the necessary code to access the relevant APIs, creating intuitive dashboards to display the outcomes, and making all these resources available to you for free. Additionally, we are incorporating Azure GPT to make the entire process more conversational and user-friendly.

Stay tuned for further updates on how this program can revolutionize your experience with Cisco solutions. Together, let's unlock the full potential of our technology and drive greater success for our customers.

Achieving Efficient IT Operations, our first use case, will launch at Cisco Live US (June, 2024) with more to follow!

AIOPS hackathon workshops for partners

Cisco Black Belt for AI



Joel Jose 14/06/24, 5:42 PM

MSP What are the issues seen in the infrastructure and mention the troubleshooting steps

MSP Bot 14/06/24, 5:42 PM

Based on the provided dashboard data, the following issues are observed in the infrastructure:

1. **Alarms:**
   - **Cat9k-BR1.dcloud.cisco.com**: Severity - Major, Summary - radius_error_trigger
   - **BR1-Core.dcloud.cisco.com**: Severity - Major, Summary - radius_error_trigger
   - **DC-EDGE1**: Severity - Warning, Summary - sla-violation
   - **vBond**: Severity - Warning, Summary - security-root-cert-chain-installed
   - **NA**: Severity - Critical, Summary - Analytics-license-compliance
   - **BR2-EDGE1**: Severity - Critical, Summary - omp-state-change
   - **vManage**: Severity - Warning, Summary - security-root-cert-chain-installed
   - **vSmart1**: Severity - Warning, Summary - security-root-cert-chain-installed
   - **vSmart2**: Severity - Warning, Summary - security-root-cert-chain-installed
   - **BR2-EDGE1**: Severity - Warning, Summary - sla-violation
   - **vBond**: Severity - Critical, Summary - control-vbond-state-change
2. **Application Health:**
   - Multiple applications have critical health status due to high packet loss percentages (e.g., youtube, google-services, facebook, etc.).
3. **Meraki Health:**
   - Two devices are in a dormant state:
     - Device with serial Q2EZ-Y9XD-P5LN
     - Device with serial Q2EZ-YGRG-TA3W
4. **ThousandEyes Agent:**
   - One event where the TE agent failed to connect to a port.

For troubleshooting steps, please refer to the runbook information.
Based on the provided dashboard data, here are the applicable troubleshooting steps from the runbook:

Alarms:

1. **Cat9k-BR1.dcloud.cisco.com**: Severity - Major, Summary - radius_error_trigger
   - **Troubleshooting Steps:**
     - Please refer to the RADIUS Issue Resolution Guide for detailed troubleshooting flow.
2. **BR1-Core.dcloud.cisco.com**: Severity - Major, Summary - radius_error_trigger
   - **Troubleshooting Steps:**
     - Please refer to the RADIUS Issue Resolution Guide for detailed troubleshooting flow.
3. **DC-EDGE1**: Severity - Warning, Summary - sla-violation
   - **Troubleshooting Steps:**
     - No steps available in runbook for this.
4. **vBond**: Severity - Warning, Summary - security-root-cert-chain-installed
   - **Troubleshooting Steps:**
     - No steps available in runbook for this.
5. **NA**: Severity - Critical, Summary - Analytics-license-compliance
   - **Troubleshooting Steps:**
     - No steps available in runbook for this.
6. **BR2-EDGE1**: Severity - Critical, Summary - omp-state-change
   - **Troubleshooting Steps:**

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

Webex App

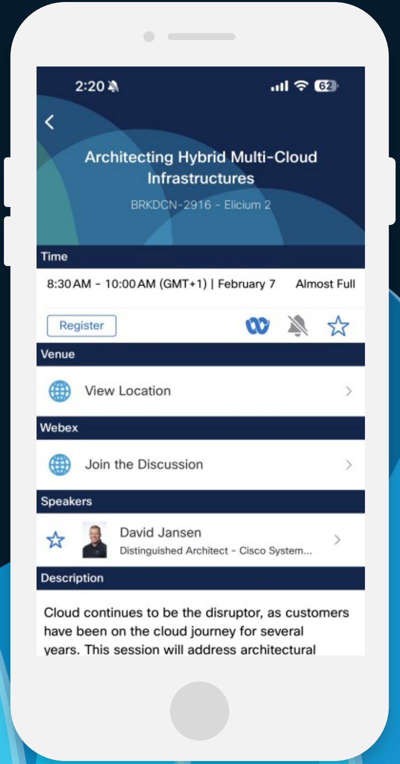
Questions?

Use the Webex app to chat with the speaker after the session

How

1. Find this session in the Cisco Events mobile app
2. Click “Join the Discussion”
3. Install the Webex app or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until February 28, 2025.



Architecting Hybrid Multi-Cloud Infrastructures

BRKDCN-2916 - Elicium 2

Time

8:30 AM - 10:00 AM (GMT+1) | February 7 Almost Full

Venue

View Location

Webex

Join the Discussion

Speakers

David Jansen
Distinguished Architect - Cisco System...

Description

Cloud continues to be the disruptor, as customers have been on the cloud journey for several years. This session will address architectural

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

Fill Out Your Session Surveys

Participants who fill out a minimum of 4 session surveys and the overall event survey will get a unique Cisco Live t-shirt.

(from 11:30 on Thursday, while supplies last)

All surveys can be taken in the Cisco Events mobile app or by logging in to the Session Catalog and clicking the 'Participant Dashboard'

Content Catalog

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

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at ciscolive.com/on-demand. Sessions from this event will be available from March 3.

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

Thank you

cisco Live!

GO BEYOND