

AI인프라 운영 최적화 : 시스코 AIOps를 활용한 지능형 모니터링

김정훈 상무, 시스코 코리아

시스코 글로벌 Customer Success Specialist팀

Agenda

- AI 인프라 운영 고려사항
- AIOps : Lessons learned
- AI 인프라를 위한 AIOps 방안
- 데모
- Summary

AI 인프라 운영 고려사항

새로운 기술에 대한 학습 필요

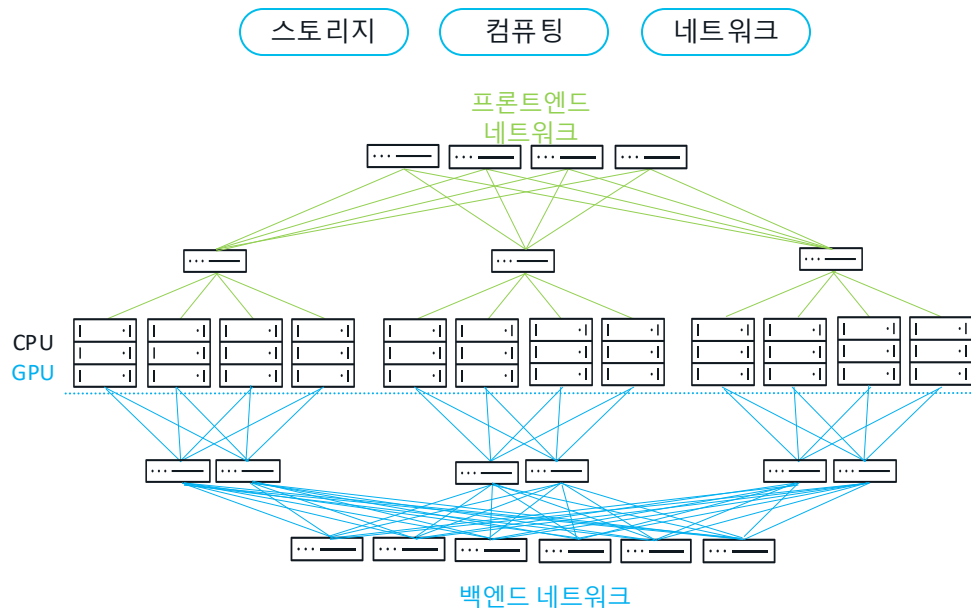
eBGP, EVPN, 부하분산, RoCEv2, GPU 등

기존 인프라에 AI 인프라 추가

운영 CPU+GPU, 트래픽+부하분산, 손실+플로 제어

새로운 방식의 모니터링 필요

텔레메트리 방식, 이상 감지, GPU간 플로 분석



10G | 25G | 50G | 100G | 400G | 800G

Lossless | High-Throughput | Low Jitter | Low-Latency

AI 인프라 운영 고려사항

인터페이스 상태

```

QOS GROUP 2
-----
| Unicast | Multicast
-----
Tx Pkts | 0 | 0
-----
WRED/AFD & Tail Drop
WRED/AFD & Tail Drop
    
```

GPU 자원 사용량

```

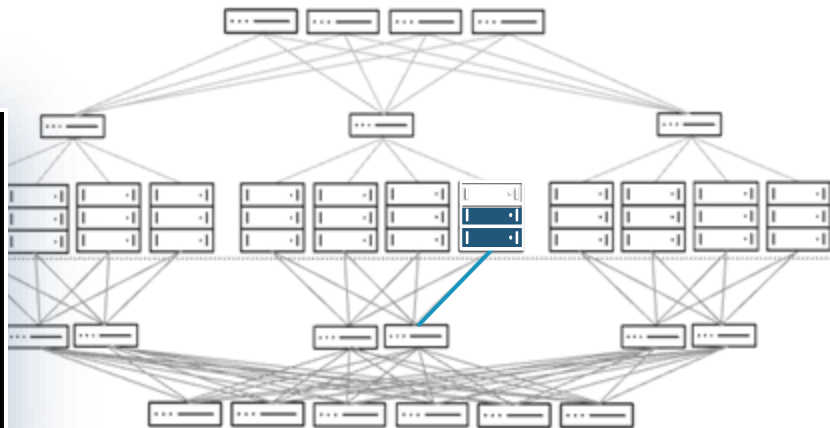
Mon Jan 06 12:26:18 2025 (Press h for help or q to quit)
NVITOP 1.3.2 Driver Version: 535.183.01 CUDA Driver Version: 12.2
-----
GPU Name Persistence-M Bus-Id Memory-Usage Volatile Uncorr. ECC
Fan Temp Perf Pwr:Usage/Cap Memory-Usage GPU-Util Compute M.
-----
0 A40 Off 00000000:5E:00:0 Off MEM: 16.4%
0% 49C P0 69W / 390W 7550M1B / 44,9961B 0% Default 0
-----
Load Average: 0.35 0.63 1.99 GPU MEM: 16.4%
CPU: 0.1%
-----
MEM: 3.39G1B (2.1%)
SWP: 0.00G1B (0.0%)
-----
CAUTION: SUPERUSER LOGGED-IN.
-----
Processes:
GPU PID USER GPU-MEM %SM %CPU %MEM TIME COMMAND
-----
0 10847 C root 7550M1B 0 0.5 0.4 6:05:14 python3 /app/server.py --listen --verbose --api --extensions api --model Phi-3-mini...
    
```

스토리지

컴퓨팅

네트워크

프론트엔드 네트워크



백엔드 네트워크

10G | 25G | 50G | 100G | 400G | 800G

전체 AI 인프라를 운영하기 위한 모니터링 방식 필요

...less | High-Throughput | Low Jitter | Low-Latency

AIOps : Lessons Learned

What & Why AIOps?

Gartner가 정의한 AIOps

AIOps는 빅데이터와 머신 러닝을 활용하여 IT 운영을 개선하는 플랫폼

여러 IT 운영 도구 및 데이터를 통합하여 실시간 분석과 자동화를 제공하는 기술

- 이벤트 상관 관계 분석
- 이상징후 감지
- 조치 자동화

CISCO *Connect*

Why AIOps?

AIOps 사용시 운영 비용 15~30% 절감

Enterprise Management Associates. "The Cost-Benefit Analysis of AIOps"

일일업무를 AIOps로 대체하면 생산성 30% 증가

Deloitte "AIOps: A New Era for IT Operations."

AIOps를 적용한 기업에서 MTTR 70% 개선 효과

"State of AIOps" by Gartner

AIOps로 서비스 가용율 개선 시 매출 5%이상 향상

IDC "AIOps and the Future of IT Operations."

컴플라이언스 점검을 통해 과태료 25% 감소

Forrester "The Total Economic Impact of AIOps."

AIOps 도입이 어려웠던 이유

활용 사례가 아닌 기술(AI/ML)로 접근

8년 전 만화



AIOps 도입이 어려웠던 이유

활용 사례가 아닌 기술(AI/ML)로 접근

생성형 AI의 한계

서울에서 집을 사는 방법을 AI에게 물으면?

ChatGPT ▾

공유하기

서울에서 집을 사는 방법 알려줘

서울에서 집을 사는 과정은 여러 단계로 나뉩니다. 아래 절차를 참고하면 도움이 될 거야.

1. 예산 계획 세우기

- 자기 자본 확인: 현재 보유한 자금(예금, 적금, 투자금 등)을 정리.
- 대출 가능 여부 확인: 은행에서 주택담보대출(주담대) 한도와 이자율을 확인.
- 총 예산 결정: 자본 + 대출금 = 구매 가능 금액.

2. 원하는 지역 및 아파트(집) 선정

- 서울 내 선호 지역 선정: 강남, 마포, 송파, 노원 등 원하는 지역 탐색.
- 교통, 학군, 인프라 고려: 직장, 학교, 병원, 편의시설 확인.
- 부동산 시세 조사: 네이버 부동산, KB부동산, 호갱노노 등 활용.

3. 매물 탐색 및 방문

- 부동산 앱 활용: 네이버 부동산, 다방, 직방 등에서 원하는 매물 찾기.
- 현장 방문: 매물의 실제 상태, 주변 환경, 소음, 채광 등 체크.
- 중개업소 방문: 신뢰할 수 있는 공인중개사 통해 매물 상담.

4. 계약 진행

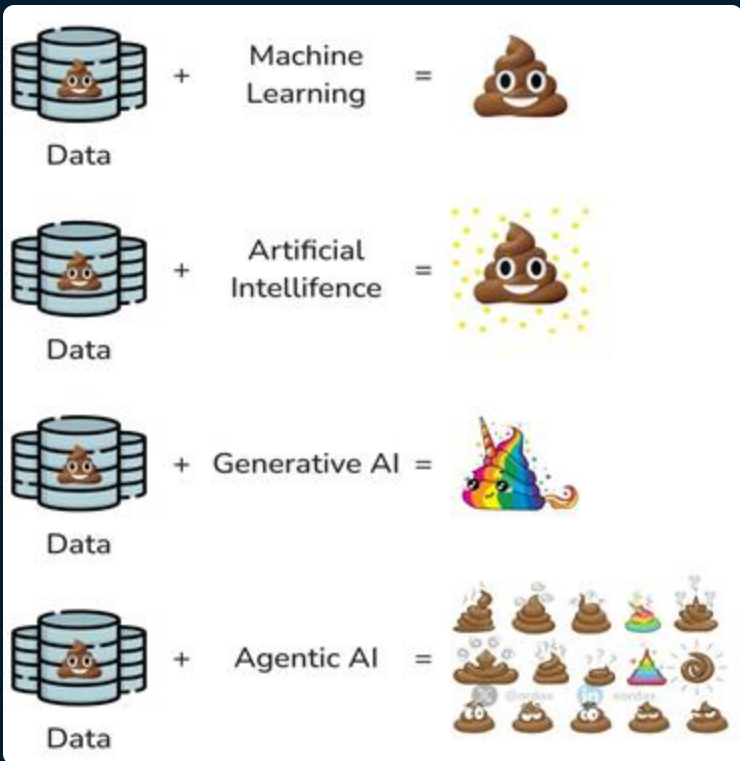
- 계약 전 확인 사항:
 - 등기부등본 (소유권, 근저당 확인)
 - 건축물대장 (불법 건축 여부)

AIOps 도입이 어려웠던 이유

활용 사례가 아닌 기술(AI/ML)로 접근

생성형 AI의 한계

AIOps로 불필요한 업무 증가



AIOps 도입이 어려웠던 이유

활용 사례가 아닌 기술(AI/ML)로 접근

생성형 AI의 한계

AIOps로 불필요한 업무 증가

운영자 역할이 대체될 것을 우려

" 난 내가 설거지와 빨래를 하기 위해 AI가 내 글쓰기와 창작을 대신 해주기 보단 내가 글쓰기와 창작을 할 수 있도록 AI가 내 빨래와 설거지를 해주었으면 좋겠어요" - 한 소설가



AIOps 도입이 어려웠던 이유

활용 사례가 아닌 기술(AI/ML)로 접근

생성형 AI의 한계

AIOps로 불필요한 업무 증가

운영자 역할이 대체될 것을 우려

이상과 현실의 차이

자율주행 기술 단계

레벨구분	0단계	1단계	2단계	3단계	4단계	5단계
		운전자 보조 기능		자율주행 기능		
명칭	無 자율주행 (No Automation)	운전자 지원 (Driver Assistance)	부분 자동화 (Partial Automation)	조건부 자동화 (Conditional Automation)	고도 자동화 (High Automation)	완전 자동화 (Full automation)
자동화 항목	없음	조향 or 속도	조향 or 속도	조향 or 속도	조향 or 속도	조향 or 속도
운전주시	항시 필수	항시 필수	항시 필수	시스템 요청시	작동구간 내 불필요	전 구간 불필요
자동화구간	-	특정구간	특정구간	특정구간	특정구간	전 구간
예시	사각지대 경고	차선유지 또는 크루즈 기능	차선유지 및 크루즈 기능	혼잡구간 주행지원 시스템	지역 무인택시	운전자 없는 완전 자율주행

레벨별
차량



Q. 어떤 차를 타고 싶으신가요?

AIOps 모델의 진화 방향

“Teaming with AI”

"AIOps is allowing teams to focus on strategic decision-making while automating routine tasks, ultimately leading to faster problem resolution and improved system reliability."



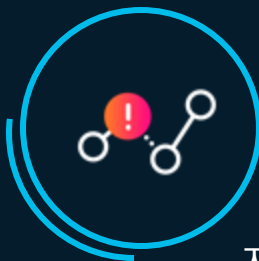
생성형 AI

“사용자를
전문가로”

쉽게 결과를 확인

Unified Observability

AI Assistant



감지 및 예측

실시간 데이터 분석으로 (Telemetry + ML)

이상 감지 및 예측



지능형
모니터링

경보 최적화

상관 관계 분석

중복 제거

시스코 IT AIOps 사례

멀티 도메인 AIOps 플랫폼

Any Platform, Location, Domain

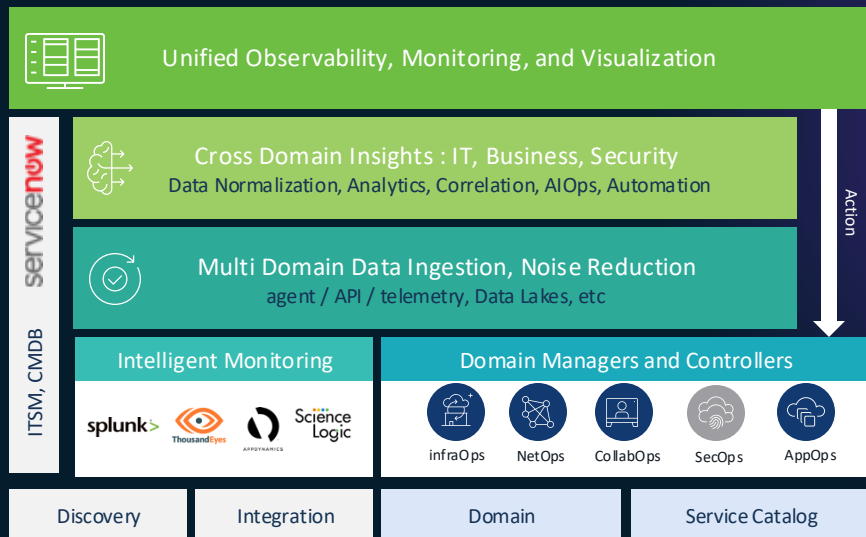
담당자 및 Use Case 별 최적화

도메인간 통합 분석, 통합 대시보드, 자동복구

서비스 품질 및 비즈니스 개선



시스코 IT 멀티도메인 AIOps 아키텍처



퍼블릭 클라우드

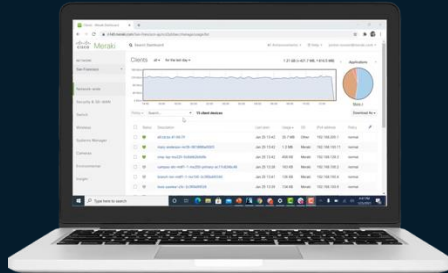
데이터 센터

SaaS

AIOps는 사용자 인터페이스를 진화



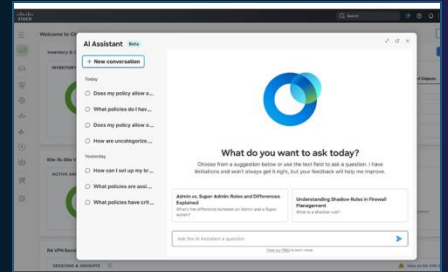
CLI 명령어



비기본
인터페이스

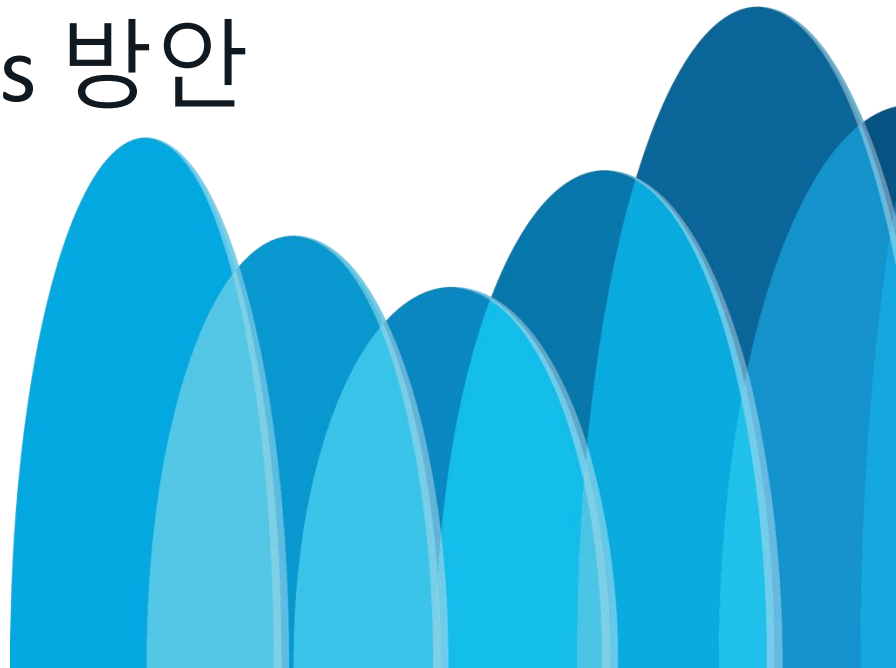


APIs

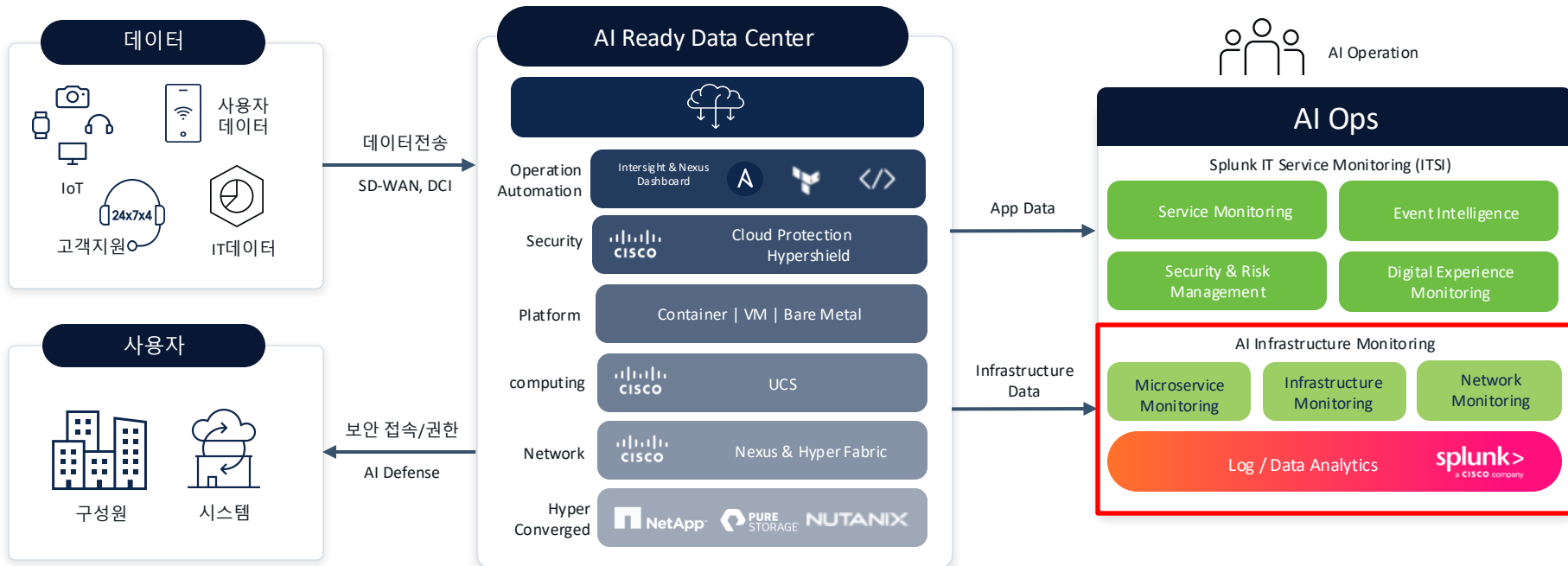


생성형 AI
어시스턴트/에이전트

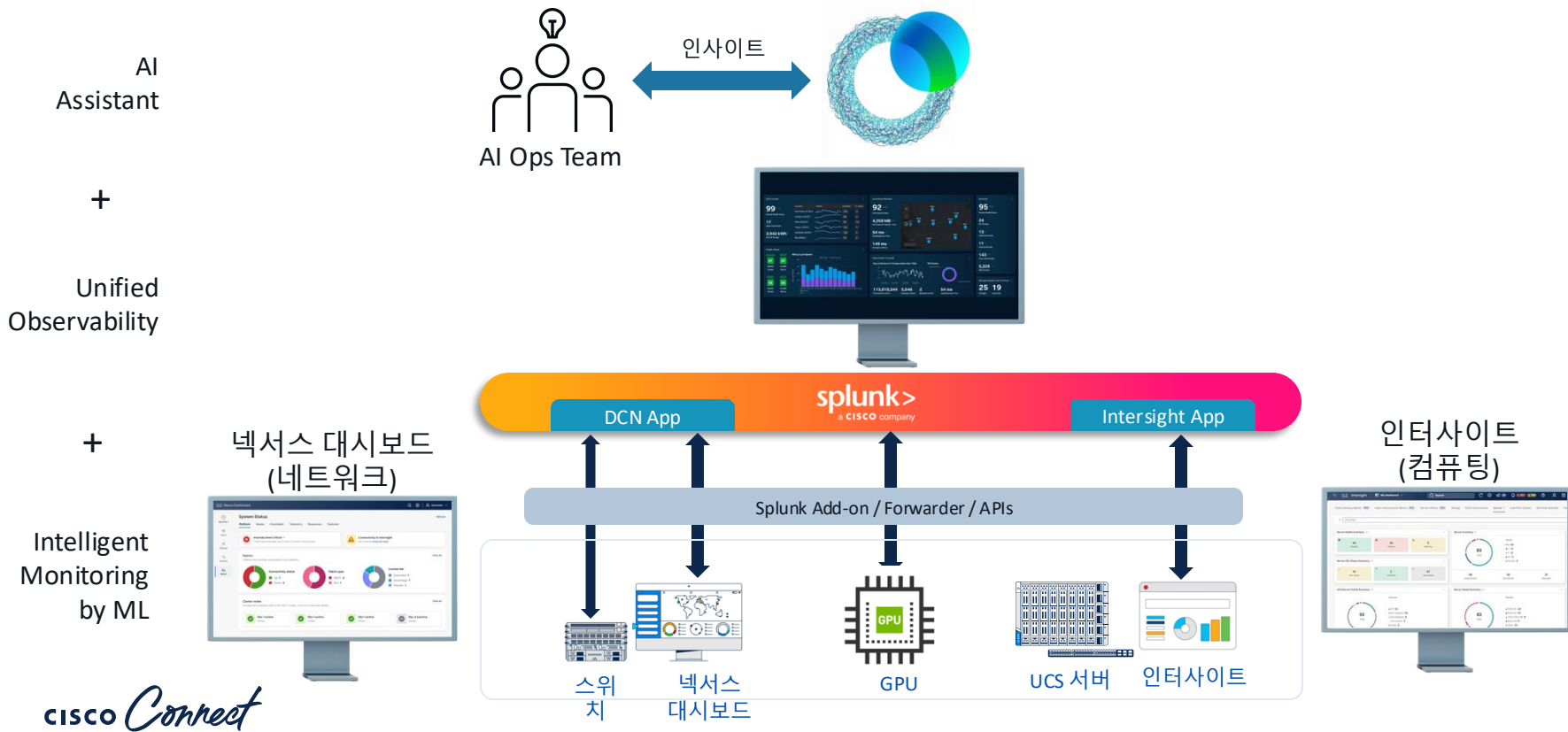
AI 인프라를 위한 AIOps 방안



AI 인프라를 위한 AIOps 모델



AI 인프라 AIOps 구성 예제



AI 인프라를 위한 AIOps 활용 시나리오

AI 인프라 모니터링

eBGP / EVPN, 부하분산, RoCEv2, GPU 등

일반 인프라 모니터링

엔드 포인트, 플로 및 트래픽, H/W 상태 등

이벤트 대응

경보 알림, 통합 네트워크 분석, 이력 확인 등

상시업무

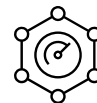
일일점검, 주간 분석, 업그레이드 사전검증 등



통합 대시보드



트러블슈팅



최적화



자동화

AIOps를 위한 어시스턴트

AI 에이전트 모델

스크립트 방식 vs 자율형 방식

정해진 주제, 단기간 구현
저비용

높은 대화 수준, 질문 유추
대량의 지식(KB) 활용

소통 방식 (Conversation)

간단한 텍스트 vs 말하기

CLI명령어 “show”
“플로”, “GPU”

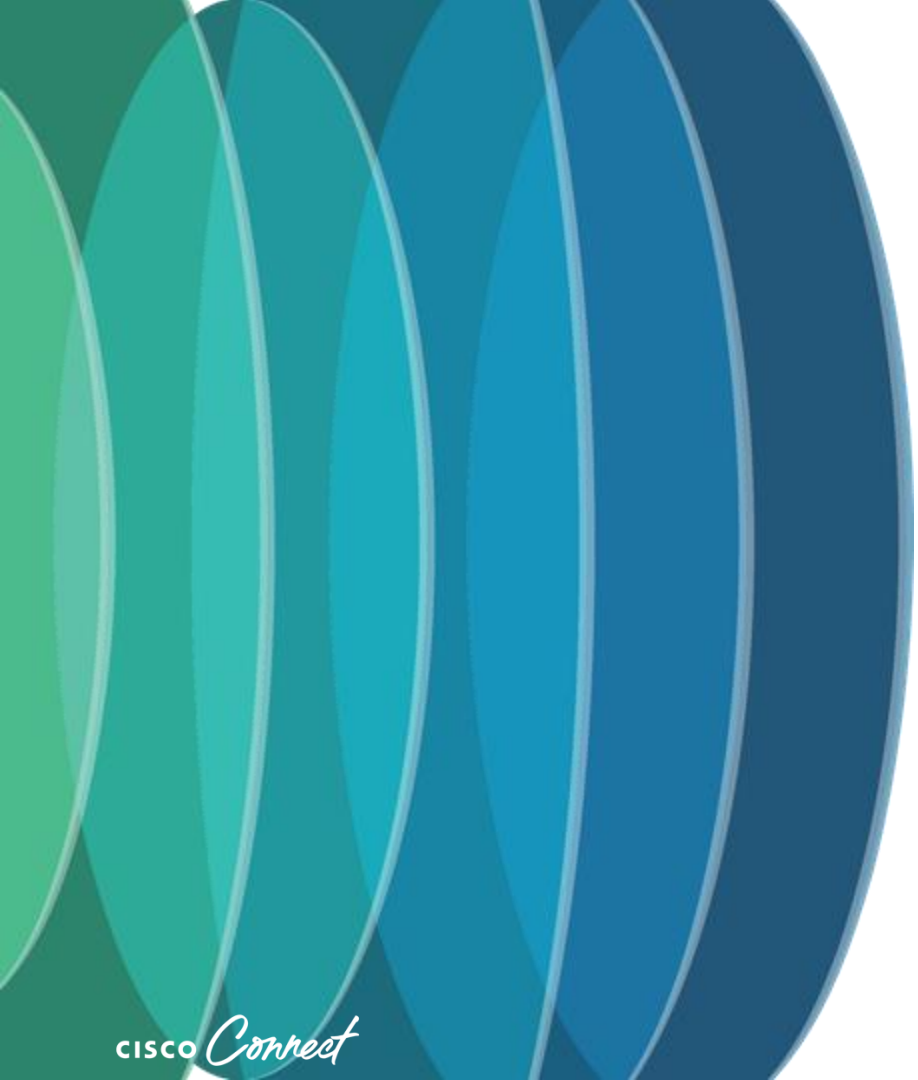
음성 인식, 지속적인 대화
어휘 인식

산출물 (Output)

오브젝트 vs 문장

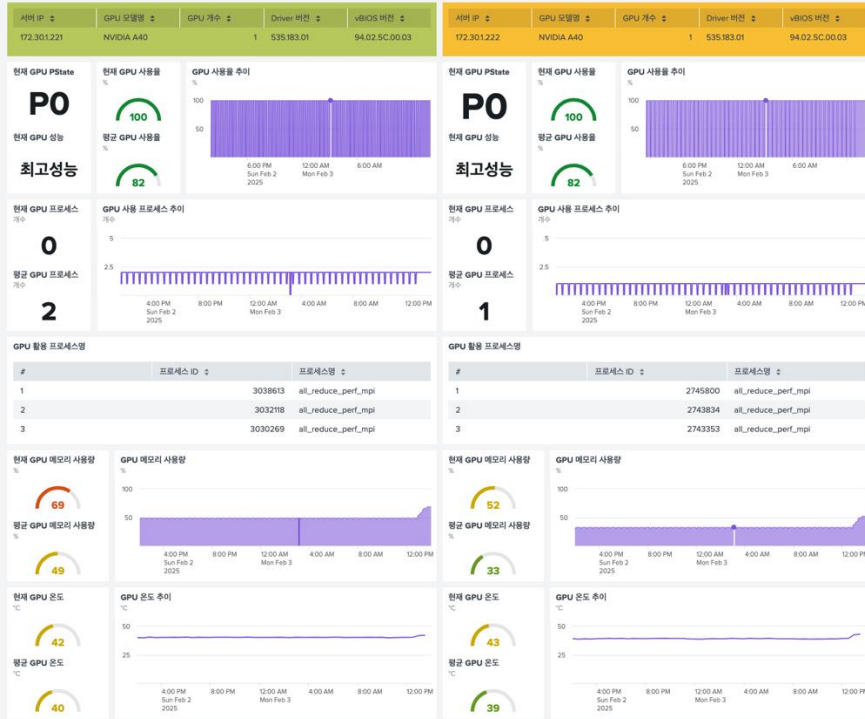
텍스트, 이미지, 파일
접속 링크

인사이트를 포함한 문장
추가 질문

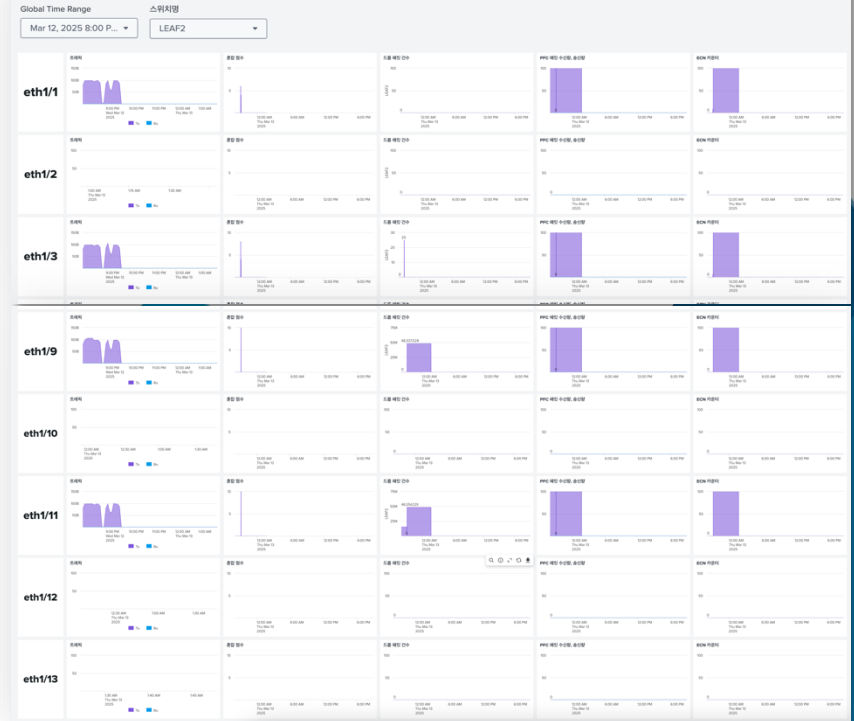


AIOps 데모

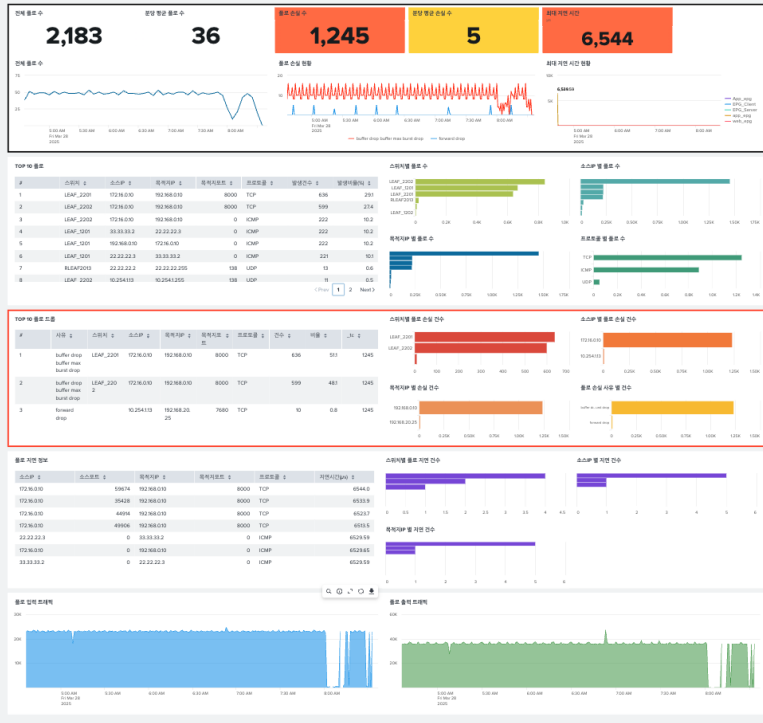
#1. AI 인프라 모니터링



AI Infra Monitoring - Interfaces QoS



#2. 플로 및 호스트 분석



#3. 이상감지 및 OS 검증

시스코 커넥트 AI Ops팀 ☆

메시지 사용자 (3) 콘텐츠 미팅 + 앱

저는 당신의 AI서비스입니다!

나의AI서비스 오전 11:40
안녕하세요! 무엇을 도와드릴까요?

본인 오전 11:40
나의AI서비스 anomaly

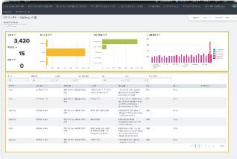
나의AI서비스 오전 11:40
#SHOW_ANOMALY 확인해드릴게요, 조금만 기다려주세요~♡

나의AI서비스 오전 11:40

ndi_anomaly_...01_024008.csv
41 KB 안전

최근 하루 동안 발생한 네트워크 이상 경보내역입니다! 📊

나의AI서비스 오전 11:40



http://20.196.90.138:8000/en-US/aop/cisco-cx-demo/anomaly?form.global_time_earliest=-24h%40h&form.global_time_latest=now&form.tk_severity=&form.tk_category=&form.tk_node_name=&form.tk_detail=&form.tk_comment=&form.tk_ack=&form.tk_clear=&tab=layout_1

본인 오전 11:41
고마워~

Shift + Enter를 눌러 새로운 줄 생성
필 물어볼까?

시스코 커넥트 AI Ops팀 ☆

메시지 사용자 (3) 콘텐츠 미팅 + 앱

본인 오전 11:43

나의AI서비스 APIC OS 검증해줘~

나의AI서비스 오전 11:44
#SHOW_APIC_OS_VERSIONS 확인해드릴게요, 조금만 기다려주세요~♡

나의AI서비스 오전 11:44

result_20250401024417.txt
19 KB 안전

현재 APIC 버전[5.2(4d)]에서 권장 APIC 버전[5.3(1d)]으로 업그레이드 사전 검증 결과입니다!
상세 검증 내용은 첨부파일을 확인해주세요~♡

```
1 === Summary Result ===
2
3 PASS                               : 50
4 FAIL - OUTAGE WARNING!!           : 6
5 FAIL - UPGRADE FAILURE!!          : 3
6 MANUAL CHECK REQUIRED              : 2
7 POST UPGRADE CHECK REQUIRED        : 1
8 N/A                                : 7
9 ERROR !!                           : 0
10 TOTAL                             : 69
```

확인함: 🟢

Shift + Enter를 눌러 새로운 줄 생성
나의AI서비스 아이쿠... 큰일났네 했네..!

Summary



Summary

- AI 인프라 위한 새로운 운영 방식과 모니터링 환경이 필요합니다.
- AIOps는 운영자가 빠른 의사결정을 할 수 있도록 돕는 역할로 진화하고 있습니다.
- 시스코의 다양한 솔루션을 활용하면 AI인프라를 위한 AIOps를 구성할 수 있습니다.

CX 데모 부스 안내

캠퍼스와 데이터센터 한번에 알아보기

- 국내 운영환경을 위한 Splunk 통합 플랫폼 소개
- 운영 업무 별 대시보드
데이터 센터, AI인프라, 유무선 네트워크

데이터센터 네트워크 분석 꿀팁

- ACI 검증 자동화
- 건강검진 2.0 소개
- AI 어시스턴트를 통한 AIOps 예제

데모부스
데워치



김인섭 이사
Customer Success Specialist
AI어시스턴트, Splunk



최건 프로
Customer Success Specialist
Splunk 통합 플랫폼, 건강검진



김봉원 프로
Customer Success Specialist
건강검진, ACI OS 사전 검증



The bridge to possible

감사합니다!

CISCO *Connect*