— STUDY PACK · POCKET EDITION —

AWS Summit Seoul 2026 — 요약본

Master Script · 6 Areas · v3 Roadmap

대상 Samsung 마케팅 PM (David)

구성 핵심만 — 마스터 스크립트 + 흐름도 + 로드맵 + 핵심 용어

분량 ~30 pages · 휴대용

날짜 2026년 5월

POCKET EDITION

목차 — Contents

핵심 발화·흐름·로드맵·용어 (요약본)

01 마스터 스크립트 (4분) — 임원 보고 풀버전

02 6개 영역 한 페이지 정리

03 통합 흐름도 — 한 요청이 도는 길

04 v3 로드맵 5개 — SmartThings Scenario Agent

05 핵심 용어 25선

06 클라우드 비교 — 요약 행

CH 06 · SYNTHESIS · 마스터

통합편 — 한 장으로

이틀, 18개 트랙, 100여 개 용어, 6개 영역. 이 마지막 챕터가 그 모든 것의 한 줄 요약입니다.

▶ MASTER SPEAKING SCRIPT · FULL　　약 4분 · 임원 보고 풀버전

OPENING · 25초

AWS Summit Seoul 2026, 이틀간 18개 트랙·100여 개 용어 들었습니다. 다 듣고 나서 정리한 한 문장은 이거예요… **"에이전틱 AI 시대로의 풀스택 전환"**. 다른 말로 하면, 클라우드의 모든 부품이 "AI 에이전트를 어떻게 만들고·돌리고·지키고·팔지"라는 한 질문 중심으로 재배치됐다는 뜻입니다.

PART 1 · 6개 영역 · 110초

전체 그림은 6개 영역으로 나뉘었습니다.

첫째, 두뇌. 모델 본체와 에이전트 프레임워크예요. Bedrock·Nova·Claude 같은 파운데이션 모델 위에 AgentCore·Strands SDK·MCP가 얹혀서, 모델이 "도구 쓰는 비서"로 변합니다.

둘째, 공장. AI Native 사고방식 + Kiro·Claude Code 같은 AI 코딩 도구 + CI/CD·IaC 자동화 + MLOps·AIOps·DevSecOps 운영. 코드를 매일 갱신하는 컨베이어 벨트예요.

셋째, 연료. S3에 깔리고 OpenSearch가 검색층이 되고, RAG·GraphRAG로 회사 데이터가 모델에 닿습니다. CDC와 Lakehouse가 실시간 갱신을 담당하고요.

넷째, 엔진. Trainium·Inferentia 자체 칩과 NVIDIA GPU를 얹은 EC2 P/G 위에서, HyperPod 클러스터로 수백 대 GPU를 묶어 굴립니다.

다섯째, 요새. Zero Trust 원칙 + KMS·CloudHSM 키 관리 + AI 전용 위협 방어. 에이전트가 회사 데이터를 진짜로 만지기 시작하면서 보안이 한 겹 더 두꺼워졌어요.

여섯째, 응용. 피지컬 AI—로봇·자동차·공장—와 비즈니스 에이전트—Amazon Connect·Quick Suite·산업별 응용. 가치가 실제로 만들어지는 층입니다.

PART 2 · 6개가 한 흐름으로 · 70초

이 여섯 개가 어떻게 한 흐름으로 묶이느냐… 사용자 질문 한 줄이 도는 길을 따라가면 명확합니다.

사용자가 **비즈니스 앱**에 질문을 던집니다. **요새**가 인증·권한을 통과시키고, **두뇌**의 에이전트가 받아서, **연료**에서 RAG로 회사 데이터를 끌어오고, **엔진** 위에서 추론을 돌립니다. 답이 다시 요새를 거쳐 사용자에게 닿고요. 그리고 이 모든 코드와 운영을 **공장**이 24시간 갱신하고 있어요.

한 줄로 정리하면… 두뇌가 일하고, 연료를 먹고, 엔진에서 돌고, 요새가 지키고, 공장이 갱신하고, 응용이 가치를 만든다.

PART 3 · 우리에게 의미 · 65초

우리에게 의미가 뭐냐… 이미 사내에서 만들고 있는 SmartThings Scenario Agent를 이 6개 그림 위에 올려보면, **응용·두뇌·연료** 세 층까지는 v2로 닿아 있습니다. 5-Agent Pipeline에 BYOK 모델, 27-시나리오 JSON DB가 그 증거예요.

남은 세 층은… **공장(CI/CD 자동화)**, **요새(사내 KMS 연동)**, **엔진(인프라 선택)**… 이게 v3에서 우리가 채워야 할 자리들입니다.

특히 한 가지만 짚자면 — **MCP 서버화**입니다. 시나리오 DB를 MCP라는 표준 도구로 노출하면, 모델이 바뀌어도, 다른 팀 에이전트가 호출해도 그대로 쓸 수 있는 자산이 됩니다. 가장 임팩트 큰 다음 한 수예요.

CLOSING · 30초

정리하면… 100개 용어는 6개 영역의 분담이었고, 6개 영역은 사용자 한 마디 답변을 위한 합주였고, 그 합주는 우리도 이미 일부 하고 있었다, 입니다.

다음 분기에 보일 가장 큰 변화는… "AI 에이전트가 우리 인프라의 디폴트 가정이 된다"는 점입니다. 감사합니다.

6개 영역 한 페이지 정리

영역	역할	핵심 키워드
🧠 **두뇌** Models & Agents	생각하고 도구 쓰는 비서.	Bedrock · Nova · Anthropic · AgentCore · Strands · MCP · A2A · Multi-Agent
🏭 **공장** Dev & Operations	코드·운영을 매일 갱신하는 컨베이어.	AI Native · Spec-driven · AI-DLC · Kiro · Claude Code · CI/CD · IaC · MLOps · AIOps · DevSecOps
⛽ **연료** Data Platform	에이전트가 먹고 일하는 회사 데이터.	S3 · OpenSearch · RAG · GraphRAG · CDC · Lakehouse · AI-Ready Data
⚙️ **엔진** Compute Infra	모델이 실제로 도는 실리콘.	Trainium · Inferentia · EC2 P/G · Nitro · HyperPod · Slurm on EKS · Capacity Blocks
🛡️ **요새** Security	에이전트가 회사 자산을 만지는 걸 지킴.	Zero Trust · KMS · CloudHSM · Guardrails · DevSecOps · Secure AI

영역	역할	핵심 키워드
💼 **응용** Physical & Business	가치가 실제로 만들어지는 사용자 접점.	Amazon Connect · Quick Suite · Physical AI · Robotics FM · Sim-to-Real · AI 컨시어지 · Audience Engine

통합 흐름도 — "오늘 출퇴근 시간 시나리오 추천해줘"의 뒷면

> **USER INPUT:** "오늘 출퇴근 시간 시나리오 추천해줘"

① **응용** — SmartThings 앱이 입력 수신. Amazon Connect·Quick Suite·또는 자체 앱이 1차 접점.

② **요새** — Zero Trust로 요청자 검증. KMS가 호출 키 안전하게 발급. AI 전용 Guardrails가 프롬프트 인젝션 차단.

③ **두뇌** — AgentCore에서 도는 5-Agent가 입력 수신. Curator가 시나리오 후보 분류, Localizer가 한국 맥락 반영. MCP가 외부 도구 호출 표준.

④ **연료** — RAG가 27-시나리오 DB(S3+OpenSearch)에서 관련 시나리오 검색. 운영 DB 변경분은 CDC+Lakehouse 실시간 동기화.

⑤ **엔진** — Bedrock이 Inferentia 또는 NVIDIA GPU(EC2 P/G)에서 답 생성. Nitro 격리. 응답시간 수백 ms.

⑥ **요새** — 출력에 민감 데이터·환각·정책 위반 없는지 가드레일이 한 번 더 검사.

⑦ **응용** — SmartThings 앱에 시나리오 카드 표시. 사용자는 한 번의 응답만 보지만, 뒤에서는 6개 영역이 합주.

> **USER OUTPUT:** "☕ 모닝 루틴: 7:00 자동 기상등 + 커피 머신 ON + 출퇴근 교통 안내"

↑ 그리고 한 가지 더 — 이 모든 흐름의 코드·인프라·모델은 **공장(AI-DLC + CI/CD + AIOps)**이 백그라운드에서 매일 갱신·감시 중입니다.

SmartThings Scenario Agent — v3 후보 5개

> **현 상태 진단 (v2):**
> ✓ 응용 — SmartThings 사용자 접점 / ✓ 두뇌 — 5-Agent Pipeline + BYOK / △ 연료 — mini-RAG (정통 RAG 아님) / △ 엔진 — Cloudflare 위임 / ✗ 요새 — 개인 단위만 / ✗ 공장 — 수동 배포·수동 운영

01 MCP 서버화 연료 + 두뇌

27-시나리오 JSON DB와 prompt.txt를 MCP 표준 도구로 노출. 모델/팀이 바뀌어도 동일 도구로 호출. **자산화의 가장 큰 한 수.**

IMPACT — 매우 큼

02 Orchestrator 패턴 진화 두뇌

고정 Pipeline → Curator를 Orchestrator로 승격. 시나리오 복잡도에 따라 Localizer 건너뛰기·Expander 두 번 호출 같은 동적 분기. Story Chat을 Sub-agent로 정식 등록.

IMPACT — 큼

03 CI/CD 자동화 공장

git push → wrangler-publish GitHub Action으로 Cloudflare 자동 배포. 수동 콘솔 사라짐. P7-B 같은 잦은 튜닝 사이클에 가장 큰 시간 절약.

IMPACT — 큼

04 정통 RAG 전환 + GraphRAG 검토 연료

현재 facet 기반 retrieval → S3 + OpenSearch 벡터 인덱스로 정식 RAG. 시나리오 간 관계(같은 디바이스·시간대)를 살리려면 GraphRAG. 시나리오 100개 넘으면 임계점.

IMPACT — 중간

05 AIOps 도입 공장

Datadog 또는 Cloudflare Analytics로 P95 지연·LLM 호출 에러율 자동 감지. 너가 직접 로그 보지 않고도 이상 알림 받는 단계. 사용자 수 늘면 필수.

IMPACT — 중간

05 · ESSENTIAL TERMS

핵심 용어 25선

전체 65개 중 가장 자주 부딪힐 25개. 이것부터 외우면 80%는 닿음.

Agent · Agentic AI 에이전트 / 에이전틱 AI

BRAIN

LLM에 도구사용·계획·자율판단을 더한 시스템. 단발 응답이 아니라 'Plan→Tool→Observe→Act' 루프를 도는 다단계 의사결정 AI. Chatbot ≠ Agent.

Amazon Bedrock 베드락

BRAIN

여러 회사 FM(Claude·Llama·Nova 등)을 단일 API로 호출하는 AWS 서버리스 게이트웨이. 가드레일·RAG·에이전트 기능 통합.

Amazon Bedrock AgentCore 에이전트코어

BRAIN

에이전트를 실행하는 AWS 관리형 런타임. 메모리·도구연결·로그·세션을 자동 관리. 코드는 너, 운영은 AWS.

Foundation Model (FM) 파운데이션 모델

BRAIN

대용량 데이터로 self-supervised 사전학습된 거대 신경망. GPT·Claude·Nova·Gemini가 전부 FM. 다양한 task에 재활용 가능해 'Foundation'.

MCP 모델 컨텍스트 프로토콜

BRAIN

모델 ↔ 외부 도구·데이터 표준 연결 규격. Anthropic 주도, OpenAI·Google·AWS 채택. 'AI의 USB-C'. JSON-RPC 기반.

Multi-Agent 멀티 에이전트

BRAIN

에이전트 여러 명을 역할별로 분담시켜 협업. 4가지 패턴 — Single/Pipeline/Orchestrator+Sub-agents/Autonomous. David의 v2 = Pipeline 패턴.

Strands Agent SDK 스트랜즈 SDK

BRAIN

AWS의 오픈소스 에이전트 개발 키트(Python). 가볍고 BYOK 친화적(Bedrock·OpenAI·Anthropic 다 지원). LangChain·LangGraph의 AWS측 대안.

AI Native Development AI Native 개발

FACTORY

AI가 코드 짜는 걸 기본 전제로 두고 워크플로우를 처음부터 재설계. AI-Assisted(추가)와 다름. Cloud Native의 후속 개념.

AI-DLC AI 주도 개발 라이프사이클 FACTORY

기존 SDLC의 AI 시대 버전. Spec→AI 코드생성→AI 테스트→CI/CD→AIOps 전체 루프. AI Native의 사이클화·표준화.

AWS Kiro 키로 · AI 코딩 에이전트 FACTORY

Spec 파일 주면 코드·테스트·PR까지 자동 생성하는 AWS 코딩 에이전트. Spec mode + Vibe mode. Kiro=신규/Claude Code=수정/AWS Transform=현대화.

Claude Code 클로드 코드 FACTORY

터미널 기반 코딩 에이전트(Anthropic). IDE 없이 명령어로 파일을 직접 읽고 수정. Sub-agent 시스템으로 병렬 작업.

CI/CD 지속적 통합·배포 FACTORY

코드 push → 자동 빌드·테스트·배포 컨베이어. AI 시대엔 PR 리뷰까지 AI가 자동화. GitHub Actions · CodePipeline 등.

MLOps 엠엘옵스 · 모델 운영 FACTORY

모델의 데이터→학습→배포→모니터링→재학습 라이프사이클 자동화. SageMaker AI MLOps가 대표. 모델 버전관리·실험추적·드리프트 감지.

AIOps 에이아이옵스 · 시스템 운영 AI FACTORY

시스템 운영(로그·알람·메트릭)을 AI가 분석해 장애를 자동 진단·복구. MLOps=모델 / AIOps=시스템 / DevSecOps=보안.

DevSecOps 데브섹옵스 · 보안 통합 운영 FACTORY

보안 검사를 출시 직전이 아니라 코딩·빌드·배포 전 단계에 자동 삽입. 'Shift Left Security'. AI 시대엔 프롬프트 인젝션 검사 포함.

RAG 검색증강생성 FUEL

Retrieval-Augmented Generation. LLM이 답하기 전에 회사 DB에서 관련 문서를 먼저 검색해 함께 넘기는 방식. 환각 ↓, 데이터 갱신 즉시 반영. Fine-tuning의 대안.

GraphRAG 그래프 RAG FUEL

문서 사이 관계를 그래프로 두고 그 위에서 검색. 단순 유사도 검색을 넘어 인과·계층 관계 살림. 미래에셋증권 상품DB 사례.

Amazon S3 S3 / 객체 스토리지 FUEL

AWS 객체 스토리지. 모든 종류의 파일(이미지·문서·로그) 적재. AI-Ready 데이터의 기본 저장소. 거의 무한 확장.

Amazon OpenSearch 오픈서치 FUEL

키워드 검색 + 벡터 검색을 모두 지원하는 분산 검색엔진. RAG의 retrieval 층 기본 부품. Elasticsearch 포크에서 출발.

AWS Trainium 트레이니움 · 학습 칩 ENGINE

AWS 자체 AI 칩 — 학습(training) 전용. Inferentia와 페어. Nova가 이 칩 위에 최적화.

AWS Inferentia 인퍼런시아 · 추론 칩 ENGINE

AWS 자체 AI 칩 — 추론(inference) 전용. Trainium의 페어. NVIDIA GPU 의존도 ↓ 비용 ↓.

HyperPod 하이퍼팟 · 관리형 학습 클러스터 ENGINE

수백 대 GPU를 한 묶음으로 굴리는 AWS의 관리형 학습 클러스터. 노드 장애 자동복구·체크포인트 자동저장. 하이퍼커넥트 사례.

Zero Trust 제로 트러스트 FORTRESS

'아무도 기본 신뢰하지 않는다.' 매 요청마다 ID·기기·문맥 동적 검증. 사내·사외 구분 없음. AI 시대엔 에이전트의 API 호출 전부 재확인.

AWS KMS 관리형 키 서비스 FORTRESS

AWS 관리형 암호화 키 서비스. 키 생성·로테이션·접근제어 자동. FIPS 140-2 Level 3.

Physical AI 피지컬 AI APPS

AI가 모니터를 나와 로봇·자동차·공장·가전으로 들어가는 흐름의 우산 개념. Robotics FM + Sim-to-Real이 핵심 부품.

06 · CLOUD COMPARE (요약)

클라우드 비교 — 핵심 12개 역할

가장 자주 등장하는 12개 역할만. 자세한 비교는 full 버전 참조.

역할	AWS	AZURE	GCP
FM 게이트웨이	Bedrock	Azure AI Foundry	Vertex AI
1st-party FM	Nova	Phi	Gemini
AI 코딩	Kiro	GitHub Copilot	Gemini Code Assist
객체 스토리지	S3	Blob Storage	Cloud Storage
벡터 검색	OpenSearch	AI Search	Vertex AI Search
Lakehouse	Lake Formation	Microsoft Fabric	BigLake
학습 칩	Trainium	Maia	TPU
추론 칩	Inferentia	Cobalt	TPU
학습 클러스터	HyperPod	Azure ML Clusters	Vertex AI Training
관리형 키	KMS	Key Vault	Cloud KMS
콜센터 AI	Amazon Connect	Dynamics 365 CC	Contact Center AI
생산성 AI	Quick Suite	M365 Copilot	Gemini Workspace