LV7 · ENTERPRISE

클라우드 AI

AWS Bedrock · Vertex · Azure에서 Claude 운영

75분 · 12 슬라이드 · 아키텍트·CTO · Anthropic + 3대 클라우드

ROADMAP · 75 MINUTES

오늘의 4 블록

개인·팀 규모를 넘어, 회사 전사 규모로 Claude를 운영하려면

BLOCK A · 15 min

왜 클라우드인가

Standard API vs Bedrock/Vertex/Azure / 컴플라이언스·데이터 주권

BLOCK B · 25 min

3대 클라우드 비교

AWS Bedrock · GCP Vertex · Azure AI / 지역·가격·기능 매트릭스

BLOCK C · 25 min

한국 운영 모범

서울 리전·VPC 격리·KISA 준수·KOSPI 사례·정부 클라우드

BLOCK D · 10 min

로드맵

온프레미스 모델·자체 학습·미래 트렌드

핵심 약속: 오늘 끝나면 **"우리 회사 규제·인프라에 맞는 Claude 배포 결정"**이 가능해진다.

PRINCIPLE 01 · WHY CLOUD

왜 클라우드 Claude인가

Beyond Standard API

Anthropic API와 클라우드 Claude는 **동일한 모델** — 다른 점은 **운영·컴플라이언스**

데이터 주권

한국 데이터는 한국에서 처리. **AWS 서울·GCP 서울 리전**에서 Claude 호출 가능.

컴플라이언스

금융·의료·정부 규제 (KISA·금감원·개인정보보호위) 클라우드 인증 활용. Standard API는 별도 평가 필요.

기존 클라우드 통합

이미 AWS·GCP·Azure 쓰는 회사는 같은 계정·인증·결제로 Claude 사용.

엔터프라이즈 기능

VPC 격리, PrivateLink, KMS 암호화, IAM 권한 — Standard API에 없는 엔터프라이즈 기능.

PRINCIPLE 02 · COMPARISON

3대 클라우드 비교

Claude는 AWS Bedrock·GCP Vertex AI·Azure AI 세 곳에서 공식 제공. 모델 자체는 동일.

	AWS Bedrock	GCP Vertex AI	Azure AI
한국 리전	**모델 라인업**	**가격**	**네트워크 격리**
서울 (Sonnet/Haiku)	Opus·Sonnet·Haiku 전체	Standard API와 동일	VPC + PrivateLink
서울 (베타)	Opus·Sonnet·Haiku	동일	VPC-SC
Japan East 권장	Sonnet 주력	동일	Private Endpoint
데이터 잔존	**인증·권한**	**한국 컴플라이언스**	
학습 불사용 · 30일 로그	IAM	KISA CSAP·금감원 인증	
학습 불사용	Cloud IAM	CSAP 진행 중	

PRINCIPLE 03 · CHOOSING

D

선택 가이드

Decision Tree

"어디서 Claude 운영해야 하나?" — 4가지 질문

① 이미 쓰는 클라우드?

AWS 쓰면 Bedrock, GCP면 Vertex, Azure면 Azure AI. **같은 곳이 답**.

② 한국 데이터 주권 필수?

필수면 서울 리전 있는 **AWS Bedrock (현재 가장 안정)**. GCP는 베타 중.

③ 규제 산업 (금융·의료)?

KISA CSAP 인증이 결정적. **Azure CSAP 완료**, AWS·GCP는 단계별.

④ 소규모 시작?

규제 없으면 **Standard API**가 가장 빠르고 저렴. 클라우드는 나중에 마이그레이션 가능.

BLOCK C · KOREAN OPERATIONS

한국 운영 모범

서울 리전 + VPC 격리 + KMS 암호화 + 모니터링 — 4중 안전망.

01 · 서울 리전 사용

한국 데이터 한국에

설정: AWS Bedrock `ap-northeast-2` 서울. 모델 ID는 동일. 지연 시간 ↓ (도쿄 대비).

02 · VPC + PrivateLink

인터넷 안 거치는 호출

구조: 사내 VPC → PrivateLink → Bedrock. 외부 인터넷 노출 0.

03 · KMS 암호화

로그·캐시 암호화

설정: CMK(Customer Managed Key)로 모든 데이터 암호화. 회사가 키 통제.

04 · CloudTrail + Guardrails

전체 호출 감사

로깅: 모든 Claude 호출 CloudTrail 기록. Bedrock Guardrails로 부적절 출력 차단.

한국 사례: KB금융, 우아한형제들, 카카오엔터프라이즈 등이 AWS Bedrock + Claude로 사내 도구 운영 중.

SCENARIO · ENTERPRISE PATTERNS

엔터프라이즈 패턴

대규모·다부서 운영의 4가지 핵심 패턴.

01 · Multi-Tenant

부서별 격리

구조: 부서별 별도 IAM 역할 + 별도 한도 + 별도 비용 청구. 비용 사고 격리.

02 · Cross-Region Failover

장애 대비

구조: 서울 다운 시 도쿄·싱가포르 자동 전환. RTO < 1분.

03 · Hybrid

온프레미스 + 클라우드

구조: 민감 데이터 온프레미스 (Llama 등) + 일반 데이터 Bedrock Claude. 라우터로 분기.

04 · Cost Allocation

부서별 비용 가시화

태그: 모든 호출에 부서·프로젝트 태그 → 월별 비용 청구·예산 통제.

비용 통제: Bedrock Provisioned Throughput으로 고정 비용 옵션 가능 (사용량 예측 가능 시).

SCENARIO · KR COMPLIANCE

한국 규제 대응

금융·의료·정부 — 한국 특수 컴플라이언스 사항.

01 · 금감원 클라우드 가이드

금융 ISMS-P · 망분리

필요: 망분리 환경에서 Bedrock 호출 → VPC + Transit Gateway 구조. 별도 컨설팅 권장.

02 · 개인정보보호법

PII 처리 가명화

설계: Claude 호출 전 PII 마스킹 (이름·연락처). 출력 후 재식별 위험 검증.

03 · 의료 (HIPAA·국내)

환자 정보 격리

설계: 환자 식별 정보는 Claude에 보내지 않음. 의료 영상·통계만 분석.

04 · 정부·공공

국가정보자원관리원·G-Cloud

현황: CSAP 인증 클라우드 (Azure 완료, AWS 일부)에서 Claude 사용. 정보자원원 사용 가이드 준수.

실전 팁: 규제 검토는 법무·정보보안팀과 사전 협의. "Standard API 학습 불사용" 조항 NDA·계약서에 명시.

PRINCIPLE 04 · MONITORING

M

모니터링·운영

Observability at Scale

전사 규모 운영의 4 필수

사용량 대시보드

부서별·모델별·시간대별. Bedrock + CloudWatch 또는 Vertex + Looker.

비용 알람

부서별 월 예산 + 80% 도달 시 알람. AWS Budgets / GCP Billing alerts.

이상 패턴 탐지

갑작스러운 호출 증가·실패율 ↑ → 자동 알림. 사고·악용·버그 조기 발견.

감사 로그

3년 보관(개인정보보호법 기준). CloudTrail·Audit Logs 자동 S3 백업.

PRINCIPLE 05 · GOVERNANCE

G

거버넌스

AI Governance

대기업·중견기업이 반드시 갖춰야 할 4가지

AI 사용 정책

"어떤 데이터를 AI에 보낼 수 있나·없나" 사내 정책 문서화.

직원 교육

Lv1·Lv2·Lv3 같은 사내 교육 의무화. 이해 못한 사용 = 사고 위험.

⚖ 책임 라인

AI 사용으로 인한 사고 시 책임자 명확. 보통 IT 보안팀 + 법무.

정기 감사

분기 1회 사용 패턴·비용·사고 점검. ISMS-P 인증 회사는 필수.

PRINCIPLE 06 · FUTURE

F

미래 트렌드

What's Next

2026년 기준 — 향후 12개월 흐름

온프레미스 모델 부상

Llama·DeepSeek·Qwen 등 오픈웨이트 모델이 사내 H100 서버에서 운영 가능. 민감 데이터 완전 격리.

한국 자체 모델

NCSOFT VARCO·KT 믿음·SKT A.X·Kakao kanana 등 한국어 특화 모델 부상. Claude와 병행 운영 트렌드.

⚡ Edge AI

노트북·스마트폰에서 도는 작은 모델 (Claude Haiku 4.5 Local). 지연·비용 0.

Agentic 인프라

Lv6 Agent를 클라우드 네이티브로 — Bedrock Agents, Vertex Agents가 점차 표준화.

CONCLUSION

Lv1~7 완주

LV8 · EDU

교사·강사를 위한 AI

교육 현장 적용 + 학습 효과 분석 + Claude for Education.

DEPT-AI

부서별 AI 도입

15강 사내 출강 — 영업·인사·기획·재무·운영 부서별 실전.

BUILD-AI

사내 AI 구축

30강 사내 출강 — IT 인프라·[illegible]가이드.

오늘의 약속: 우리 회사 규제·인프라에 맞는 **Claude 배포 결정안**을 한 페이지로 정리한다.