LV5 · API & MCP

AI 앱 만들기

Claude API + Model Context Protocol 실전

75분 · 12 슬라이드 · 개발자·기획자 · Anthropic API 기반

ROADMAP · 75 MINUTES

오늘의 4 블록

"AI 챗봇"에서 "AI 앱"으로 — 회사 시스템과 직접 연결

BLOCK A · 15 min

철학 + 차이

API vs 챗봇 / MCP의 발명 의도 / 비용 모델

BLOCK B · 25 min

Claude API 기초

SDK · 메시지 · 시스템 프롬프트 · 비용 통제 · 캐싱

BLOCK C · 25 min

MCP 실전

Tools 호출 / 자체 MCP 서버 / 회사 DB·SaaS 연결 4 시나리오

BLOCK D · 10 min

운영 모범

Rate limit · 토큰 캐싱 · 에러 핸들링 · Lv6 Agents 로드맵

핵심 약속: 오늘 끝나면 **"회사 도구 1개를 Claude와 연결한 작동 가능한 PoC"**를 만들 수 있다.

PRINCIPLE 01 · ANTHROPIC API

A

API ≠ 챗봇

A Programmable Brain

Claude API는 **당신의 앱·서비스에 Claude를 내장**하는 방법

챗봇은 한 사람용

claude.ai는 사람이 직접 대화. **한 사람 한 세션**. UI 고정.

API는 무한 확장

코드 한 줄로 **10명·1만명 동시**. 우리 앱 UI·DB·시스템과 통합.

사용한 만큼 과금

월정액 X. **토큰당 비용**(input/output 다름). 캐싱·배치로 70% 절약 가능.

데이터 보안

Anthropic Standard: **학습에 안 씀**. Enterprise: 데이터 처리 위치 선택. Lv7에서 자세히.

PRINCIPLE 02 · MCP

M

Model Context Protocol

USB-C for AI

MCP = AI가 **외부 도구를 표준 방식으로 호출**하는 프로토콜 (Anthropic, 2024년 11월 공개)

□ 기존 문제

"Claude를 GitHub와 연결" 따로, "Claude를 Slack과 연결" 따로 — 매번 새로 만들기.

□ MCP 해결

표준 인터페이스. **한 번 만든 MCP 서버**는 어떤 AI·앱에서도 호환.

□ 즉시 사용 가능

GitHub, Slack, PostgreSQL, Notion 등 **수십 개 공식 MCP 서버 이미 제공**.

≣ 자체 MCP 서버

회사 ERP·CRM 연결용 자체 MCP 서버는 **Python 100줄 미만**으로 작성 가능.

BLOCK B · CLAUDE SDK

가장 간단한 API 호출

Python·TypeScript SDK 공식 제공. 6줄로 Claude 호출.

```python
# pip install anthropic
import anthropic
client = anthropic.Anthropic()
msg = client.messages.create(
  model="claude-sonnet-4-6",
  max_tokens=1024,
  messages=[{"role":"user","content":"한국어로 인사해줘"}]
)
print(msg.content[0].text)
# → 안녕하세요! 만나서 반갑습니다.
```

> API 키는 ANTHROPIC_API_KEY 환경변수로. 코드에 하드코딩 절대 금지.

BLOCK B · COST CONTROL

비용 70% 절약 — 캐싱

Prompt Caching = 반복되는 시스템 프롬프트·문서를 캐싱. 한 번 보낸 컨텍스트는 다음 호출에 1/10 비용.

```
msg = client.messages.create(
  model="claude-sonnet-4-6",
  system=[{
    "type":"text",
    "text": "우리 회사 매뉴얼 100페이지...",
    "cache_control": {"type":"ephemeral"} #← 캐시 활성화
  }],
  messages=[...]
)
# 첫 호출: 캐시 작성 (조금 비쌈)
# 다음 5분간 호출: 캐시 적중 — 입력 토큰 비용 1/10
```

실무 적용: 사내 매뉴얼·코드베이스·고객 데이터셋을 시스템 프롬프트에 넣고 캐싱 → 일일 비용 70% ↓

SCENARIO 01 · INTERNAL SYSTEMS

사내 시스템 연결

ERP·그룹웨어·CRM에 갇혀 있던 데이터에 AI가 직접 접근. MCP 서버 하나면 전 직원이 자연어로 조회.

01 · ERP 거래내역 조회

"이번 주 매출 어때?"

MCP 서버: PostgreSQL → Claude. 자연어로 SQL 실행, 결과는 표·차트.

02 · 그룹웨어 일정·결재

"내 결재 대기 5건 보고"

MCP 서버: 그룹웨어 API → Claude. 결재 요약 + 우선순위 정리.

03 · CRM 고객 응대 히스토리

"이 고객 지난 6개월"

MCP 서버: CRM REST API → Claude. 응대 패턴 분석 + 다음 액션 제안.

04 · 사내 위키·문서 검색

RAG 패턴

MCP 서버: Notion/Confluence/사내 폴더 검색 → 답변 + 출처.

핵심: MCP 서버 1개 만들면 전 직원 도구로. 개별 통합 X.

SCENARIO 02 · CUSTOMER-FACING

고객 응대 자동화

웹사이트 챗봇, 이메일 자동 응대, 음성 상담 1차 처리 — API + 우리 데이터로 맞춤형.

01 · 사이트 상담 챗봇

제품 Q&A + 견적

흐름: 사용자 질문 → API + RAG로 우리 제품 매뉴얼 검색 → 답변 + 견적 폼 유도.

02 · 이메일 1차 응대

받자마자 답장

흐름: 받은 메일 → API 분류 → 일반 문의는 자동 답장, 복잡한 건 사람 알림.

03 · 음성 상담

전화 → 텍스트 → 답변

흐름: Whisper → Claude → TTS. 전화 첫 30초 응대 자동화.

04 · 사용자 도움말

제품 안내·튜토리얼

흐름: 우리 앱 안에서 "?"버튼 클릭 → 현재 화면 컨텍스트 + Claude 도움말.

전제: 외부 노출 챗봇은 system prompt에 회사 규칙·금기 명시 + 사용량 모니터링 필수.

SCENARIO 03 · CONTENT & BATCH

콘텐츠 대량 생산

반복 생산 콘텐츠 (상품 설명·이미지 캡션·번역·SEO 메타)를 API + 배치로 자동.

01 · 쇼핑몰 상품 설명

1만 개 제품 자동 작성

패턴: 제품 스펙 CSV → API 배치 → 카피 + SEO 메타 자동 생성.

02 · 다국어 번역

한 → 영·일·중 동시

패턴: 한국어 콘텐츠 → 3개 언어 + 톤별 (비즈니스·친근) 옵션.

03 · 이미지 캡션·alt text

Vision 모델 + 자동 SEO

패턴: 제품 이미지 → Claude Vision으로 캡션 → SEO 친화 alt text.

04 · Batch API

대량 작업 50% 절약

Batch API: 100만 건 작업을 24시간 안에 비동기 처리. 비용 50% ↓.

운영 팁: 실시간이 필요 없으면 무조건 Batch API — 비용 절반.

SCENARIO 04 · CUSTOM MCP

자체 MCP 서버 만들기

우리 시스템 전용 MCP 서버는 Python 100줄. 한 번 만들면 전사 표준.

01 · 매출 조회 MCP

get_sales(month, region)

구조: ERP DB 쿼리 → JSON. Claude가 "이번 달 서울 매출?"이라 물으면 자동 호출.

02 · 고객 정보 MCP

find_customer(query)

보안: 권한 체크 + PII 마스킹 후 응답. 평직원은 익명 통계만.

03 · 사내 위키 MCP

search_docs(query)

RAG: Vector DB로 의미 검색 → 출처 첨부. 환각 ↓.

04 · 작업 자동화 MCP

create_invoice(data)

주의: 쓰기 작업은 사람 확인 필수. Claude가 자동 실행 X.

공유: 한 번 만든 MCP는 사내 Claude Desktop, API 앱, 다른 LLM에서도 동일하게 사용.

PRINCIPLE 03 · OPERATIONS

0

운영 모범

Production Best Practices

PoC는 쉽지만 운영은 다르다 — 4가지 필수 체크

Rate Limit 대응

지수 백오프: 실패 시 1s → 2s → 4s 재시도. 라이브러리에 내장됨.

Prompt Caching 활용

시스템 프롬프트·자주 쓰는 문서 캐싱 → 비용 70% ↓ (Slide 6 참고).

Observability

모든 호출에 **request_id 로깅**, 비용·지연 시간 모니터링. Anthropic Console 무료 제공.

Fallback

API 다운 시 **임시 정적 답변** 또는 **다른 모델로 전환**. 단일 의존 위험 회피.

NEXT STEPS

다음은 Agents

LV6 · AGENTS

Agents & Skills

여러 단계 작업을 자율 처리하는 에이전트 설계. Anthropic Building Effective Agents 백서 기반.

LV7 · CLOUD

클라우드 AI 엔터프라이즈

AWS Bedrock · Vertex · Azure에서 Claude 대규모 운영.

LV8 · EDU

교사·강사를 위한

교육 현장 적용 + 학습 효과 분

 오늘의 약속: 일주일 안에 **회사 도구 1개를 API로 자동화한 PoC**를 만든다.