AI Dev Kit 상세 가이드

엔터프라이즈 AI 코딩 어시스턴트의 새로운 기준

Databricks Korea

2026년 3월

목차

SECTION 01

Problem

기업 환경에서 AI 코딩 어시스턴트가 부딪히는 벽

AI 코딩 어시스턴트의 현재

- GitHub Copilot, Claude Code, Cursor 등 도구 **폭발적 성장**
- 기업 환경 실증 연구에서 **8~26% 생산성 향상** 확인
- 코드 생성, 리팩토링, 디버깅에서 뛰어난 성과

> ***하지만... 기업 환경에서는 이야기가 달라집니다.***

출처: MIT/GitHub/Accenture 공동 RCT (4,000+ 개발자, 26% 생산성 향상) · DX 리포트 (135K 개발자, 주당 3.6 시간 절감)

엔터프라이즈 GAP

컨텍스트 부재

AI가 우리 회사의 테이블, 스키마, 비즈니스 로직을 모릅니다

거버넌스 사각지대

누가, 어떤 데이터에 접근했는지 추적이 불가능합니다

단절된 워크플로우

코드 작성 → 데이터 탐색 → 배포가 분리된 별개 도구입니다

보안 우려

민감 데이터가 외부 AI 서비스로 유출될 위험이 있습니다

Before vs After

Before: 기존 AI 어시스턴트

- "sales 테이블 조인해줘" → **스키마를 모름**
- Unity Catalog 권한 무시
- Databricks 작업 실행 불가
- 회사 코딩 컨벤션 무시

After: AI Dev Kit

- "sales 테이블 조인해줘" → **스키마 자동 참조**
- Unity Catalog ACL **자동 적용**
- 75개+ 도구로 잡/파이프라인/대시보드 **직접 관리**
- Skills + CLAUDE.md로 **베스트 프랙티스 주입**

"AI 코딩 어시스턴트가
우리 데이터 플랫폼을 네이티브로 이해하면
무엇이 가능해질까?"

이것이 Databricks AI Dev Kit이 답하려는 질문입니다.

SECTION 02

Architecture Overview

4대 구성요소와 Databricks MCP 생태계

AI Dev Kit이란?

- Databricks 플랫폼과 AI 코딩 어시스턴트를 연결하는 **MCP 기반 통합 툴킷**
- Claude Code, Cursor, GitHub Copilot, Gemini CLI 등 **주요 클라이언트 지원**
- **75개 이상의 MCP 도구**로 Databricks 리소스 전체 커버
- Databricks Solutions(Field Engineering) 팀의 **공식 오픈소스** 프로젝트

> *GitHub: **databricks-solutions/ai-dev-kit***

4대 핵심 구성요소

1. databricks-mcp-server

FastMCP 기반 MCP 서버

75개+ 실행 가능한 Databricks 도구를 AI에 노출

2. databricks-skills

25개+ 마크다운 기반 스킬 파일

Databricks 패턴과 베스트 프랙티스를 AI에게 주입

3. databricks-tools-core

공유 Python 라이브러리

LangChain, OpenAI Agents SDK 등에서도 독립 사용 가능

4. databricks-builder-app

Claude Agent SDK 기반 웹 앱

브라우저에서 팀 전체가 AI 코딩 어시스턴트를 사용

MCP란 무엇인가?

Model Context Protocol — AI 어시스턴트를 위한 "USB-C"

핵심 개념

- **Tools**: AI가 호출할 수 있는 함수
- **Resources**: AI가 읽을 수 있는 데이터
- **Prompts**: 재사용 가능한 프롬프트 템플릿

왜 MCP인가?

- Anthropic이 제안한 **오픈 표준**
- 하나의 서버로 **모든 클라이언트 지원**
- JSON-RPC 기반, 구현 간단

MCP를 USB-C에 비유하면 이해가 쉽습니다. 하나의 연결로 모든 도구가 연결됩니다.

Databricks의 MCP 생태계

유형	역할	상태
Managed MCP	Databricks 기본 기능 (SQL, Genie, Vector Search, UC Functions)	Public Preview
External MCP	외부 서비스 연결 — GitHub, Google Drive, Slack 등	Public Preview
Custom MCP	조직 자체 도구를 Databricks App으로 호스팅	GA

> *AI Dev Kit은 이 중 **개발자 도구 영역**을 담당합니다.*

AI Dev Kit vs Managed MCP

AI Dev Kit

- **대상**: 개발자 (IDE/브라우저)
- **목적**: 코딩 어시스턴트에 Databricks 능력 부여
- **도구**: 파이프라인/잡/대시보드 생성 등 DevOps
- **실행**: 개발자 로컬 또는 Builder App
- **인증**: Databricks CLI 프로파일

Managed MCP

- **대상**: 프로덕션 AI 에이전트
- **목적**: 에이전트가 기업 데이터에 안전하게 접근
- **도구**: SQL 실행, Genie, Vector Search, UC Functions
- **실행**: Databricks 워크스페이스 내
- **인증**: OAuth + UC ACL

AI Dev Kit 주요 도구 (75개+)

카테고리	주요 도구	설명
Unity Catalog	manage_uc_objects, manage_uc_grants	테이블/스키마/볼륨 탐색 및 권한 관리
SQL	execute_sql, get_best_warehouse	SQL Warehouse에서 쿼리 실행
Genie	ask_genie, create_or_update_genie	자연어 → SQL 변환 및 Genie Space 관리
Jobs/Pipelines	manage_jobs, manage_job_runs	워크플로우 생성, 실행, 모니터링
Dashboards	create_or_update_dashboard	AI/BI 대시보드 생성 및 수정
Model Serving	query_serving_endpoint	모델 엔드포인트 상태 조회 및 쿼리
Vector Search	query_vs_index	벡터 검색 인덱스 질의

SECTION 03

Deep Dive

코드 예시, 보안 모델, 핵심 기능 상세

Unity Catalog 통합

AI가 회사의 데이터 구조를 자동으로 이해합니다

```
# 1. 카탈로그 내 스키마 탐색
manage_uc_objects(object_type="schema", action="list",
                  catalog_name="production")

# 2. 테이블 상세 정보 (스키마 + 통계)
get_table_details(catalog="production", schema="sales",
                  table_names=["orders"])

# 3. 권한 확인
manage_uc_grants(action="get", securable_type="table",
                 full_name="production.sales.orders")
```

> *Unity Catalog의 메타데이터가 AI의 컨텍스트가 됩니다. 테이블 설명, 컬럼 코멘트를 잘 관리할수록 AI가 더 정확해집니다.*

SQL 실행 & Genie 연동

AI가 직접 SQL을 작성하고 실행합니다

```sql
-- AI에게: "지난 달 매출 상위 10개 제품 보여줘"
SELECT p.product_name, SUM(o.amount) as total_revenue
FROM production.sales.orders o
JOIN production.sales.products p ON o.product_id = p.product_id
WHERE o.order_date >= DATE_ADD(CURRENT_DATE(), -30)
GROUP BY p.product_name
ORDER BY total_revenue DESC LIMIT 10;
```

```python
# Genie Space에 자연어 질의
result = ask_genie(space_id="01ef8a...",
                   question="이번 분기 고객 이탈률은?")
# → "이번 분기 고객 이탈률은 4.2%로, 전분기 대비 0.3%p 감소"
```

Skills vs MCP Tools

MCP Tools (75개+)

- AI가 **실행하는 함수**
- execute_sql, manage_jobs 등
- Databricks API를 호출하여 액션 수행
- 비유: AI의 **"손"** — 실제로 작업을 수행

Skills (25개+ 마크다운)

- AI가 **참고하는 지식**
- Databricks 베스트 프랙티스 문서
- 패턴, 안티패턴, 코드 템플릿
- 비유: AI의 **"뇌"** — 어떻게 해야 하는지 앎

> ***시너지**: Skills가 "DLT는 이렇게 만들어야 해"라고 가르치고, MCP Tools가 실제 생성을 실행*

CLAUDE.md — 팀 컨텍스트 주입

프로젝트 루트에 CLAUDE.md를 두면 AI가 팀 규칙을 자동으로 따릅니다

```
## 데이터 접근 규칙
- 프로덕션 카탈로그: `prod_catalog`
- 개발 카탈로그: `dev_catalog`
- SQL Warehouse: `shared-warehouse-medium`

## 코딩 컨벤션
- Python: black + ruff 포매팅
- 테스트: pytest, 커버리지 80% 이상

## 금지 사항
- 프로덕션 테이블에 직접 WRITE 금지
- PII 컬럼 SELECT 시 마스킹 필수
```

Cursor는 .cursorrules, Copilot은 .github/copilot-instructions.md 등 유사 기능 제공

보안 & 거버넌스

AI Dev Kit의 보안은 두 레이어로 구성됩니다

레이어	보호 범위	적용 방식
Databricks 접근 제어	누가 어떤 데이터에 접근 가능한지	Unity Catalog ACL 자동 적용, 감사 로그
LLM 데이터 경로	도구 결과가 어떤 LLM으로 전달되는지	사용하는 LLM에 따라 달라짐

외부 LLM (Claude, GPT)

- UC 권한 적용, 결과가 외부 LLM에 전달
- Enterprise 계약 시 학습 미사용 약정

Foundation Model APIs

- 데이터가 Databricks 내부에서만 처리
- 완전한 데이터 주권 확보

SECTION 04

Demo Walkthrough

자연어 요청에서 리포트 생성까지 실제 시연

데모 시나리오

"신규 고객 세그먼트 분석 리포트를 만들어줘"

데이터 탐색 → 쿼리 작성 → 분석 → 리포트 생성까지 AI Dev Kit으로 한 번에 처리

Step 1. 데이터 탐색

UC에서 관련 테이블 자동 검색

customers.profiles (120만 행)

customers.transactions (5,400만 행)

Step 2. 스키마 이해

테이블 구조 자동 분석

컬럼 타입, 설명, 통계 확인

tier, region, ltv, churn_risk

Step 3. SQL 실행 & 분석

세그먼트별 쿼리 자동 작성/실행

인사이트 도출: Bronze 이탈률 41%

Step 4. 리포트 생성

AI/BI 대시보드 자동 생성

4개 위젯 + SQL 쿼리 포함

> *기존 수 시간의 작업이 수 분으로 단축 — 도구 전환 없이 AI 어시스턴트 하나로 완결*

SECTION 05

Use Cases & Impact

팀별 활용 시나리오와 실증 기반 기대 효과

데이터 엔지니어링 — ETL 파이프라인 개발 가속화

- 소스 테이블 스키마 **자동 탐색** → 변환 로직 생성
- 데이터 품질 검증 코드 **자동 작성**
- 기존 Job 파라미터 참조하여 신규 파이프라인 구성

Before

스키마 문서 찾기 → 매핑 정의 → 코드 작성 → 테스트

2-3일 소요

After

AI가 스키마 참조 → 매핑 자동 → 코드 생성 → 검증

2-3시간 소요

데이터 분석 — Ad-hoc 분석 & 리포팅

- Genie Space 활용한 **자연어 데이터 질의**
- 분석 결과 기반 **시각화 코드 자동 생성**
- 정기 리포트 노트북 템플릿 자동 구성

Before

분석가에게 요청 → 대기 → SQL 작성 → 시각화

1-2일 소요

After

자연어 질문 → SQL 자동 → 시각화 자동

10-30분 소요

ML/AI 개발 — 피처 엔지니어링 & 모델 개발

- Unity Catalog 피처 테이블 탐색 및 **재사용**
- 모델 학습 코드 생성 (**MLflow 통합**)
- 모델 서빙 엔드포인트 **설정 자동화**

Before

피처 검색 → 전처리 → 학습 → 등록 → 배포

1-2주 소요

After

AI가 피처 추천 → 코드 생성 → 학습/등록 자동

2-3일 소요

애플리케이션 개발

- Databricks Agent 프레임워크 코드 **자동 생성**
- Streamlit/Gradio 기반 **UI 앱 scaffolding**
- Databricks Apps **배포 설정 자동화**

> ***이 발표 자료 자체가 Claude Code로 만들어져 Databricks App으로 배포되었습니다!***

기업 MCP 활용 시나리오

데이터 거버넌스 팀

AI Agent가 UC 권한을 자동 준수

거버넌스 오버헤드 제로

Managed MCP + UC Functions

플랫폼 팀

조직 자체 도구를 MCP 서버로 구축

Databricks App으로 중앙 호스팅

Custom MCP

데이터 엔지니어링 팀

파이프라인 개발 가속

UC 메타데이터 직접 참조

AI Dev Kit

MLOps 팀

모델 배포 워크플로우 자동화

MLflow + 서빙 엔드포인트

AI Dev Kit + Managed MCP

실증 데이터 기반 기대 효과

26%
개발 생산성 향상
(MIT/GitHub RCT, 4,000+ 개발자)

3.6h
주당 절감 시간
(DX, 135K 개발자 자가보고)

UC
거버넌스 내장
(도구 레벨 ACL + 데이터 주권)

0
별도 인프라 구축
(기존 워크스페이스 위에서 동작)

수치 출처: MIT Economics working paper, GitHub/Microsoft/Accenture 공동 3개 RCT · DX AI-Assisted Engineering Q4 2025 Impact Report

왜 AI Dev Kit인가?

구분	일반 AI 코딩 도구	AI Dev Kit
데이터 컨텍스트	코드만 이해	UC 테이블/스키마/메타데이터 직접 참조
실행 능력	코드 생성만	SQL 실행, 잡/파이프라인 관리, 대시보드 생성
거버넌스	별도 설정 필요	UC ACL 자동 적용 + 데이터 주권 확보
베스트 프랙티스	일반적 코딩 패턴	Databricks 네이티브 패턴 자동 적용
배포	수동	Asset Bundles 등 자동화

SECTION 06

Getting Started

3분 설치 가이드와 PoC 제안

사전 준비

Databricks Workspace

Unity Catalog가 활성화된
Databricks 워크스페이스

Databricks CLI

```
brew tap databricks/tap
brew install databricks
```

AI 코딩 어시스턴트

Claude Code, Cursor, Copilot,
Gemini CLI, OpenAI Codex 등

Python & uv

Python 3.11+ 및 uv
AI Dev Kit 설치 시 venv 관리용

3분 퀵 스타트

Step 1. Databricks CLI 인증

```
# 워크스페이스 인증 (OAuth 권장)
databricks auth login --host https://your-workspace.cloud.databricks.com
```

Step 2. AI Dev Kit 설치 (원커맨드)

```
# 설치 스크립트가 모든 설정을 자동 처리
bash <(curl -sL https://raw.githubusercontent.com/\
databricks-solutions/ai-dev-kit/main/install.sh)
```

Step 3. 바로 사용!

```
claude   # Claude Code 시작 (MCP 설정 자동 완료)
# 첫 번째 질문: "prod 카탈로그에 어떤 테이블이 있어?"
```

다음 단계 제안

1. PoC 진행

팀 내 2-3명으로 1주일 파일럿
일상 업무에 적용해보기

2. CLAUDE.md 작성

팀 컨벤션, 데이터 접근 규칙,
프로젝트 컨텍스트 정리

3. Unity Catalog 정비

테이블/컬럼 설명 추가
AI가 이해할 수 있는 메타데이터

4. 팀 확대

PoC 결과 공유 및
팀 전체 도입 검토

참고 리소스

자료	링크
GitHub (메인)	github.com/databricks-solutions/ai-dev-kit
MCP on Databricks	docs.databricks.com/.../generative-ai/mcp
MCP 표준	modelcontextprotocol.io
Databricks CLI	docs.databricks.com/dev-tools/cli

AI Dev Kit은 Databricks Solutions(Field Engineering) 팀의 오픈소스 프로젝트입니다.

감사합니다

"AI 코딩 어시스턴트에 Databricks의 데이터 인텔리전스를 연결하세요"

Databricks