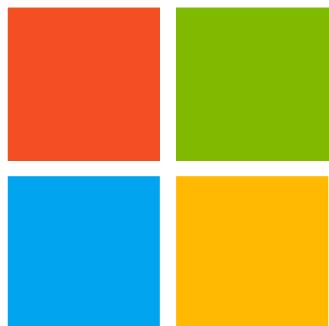






**CLOUDMATE**  
Managed Service Expert



Microsoft



#GlobalAzure

# 애저 PaaS 기반 ML 플랫폼 구성 방법론

배준현

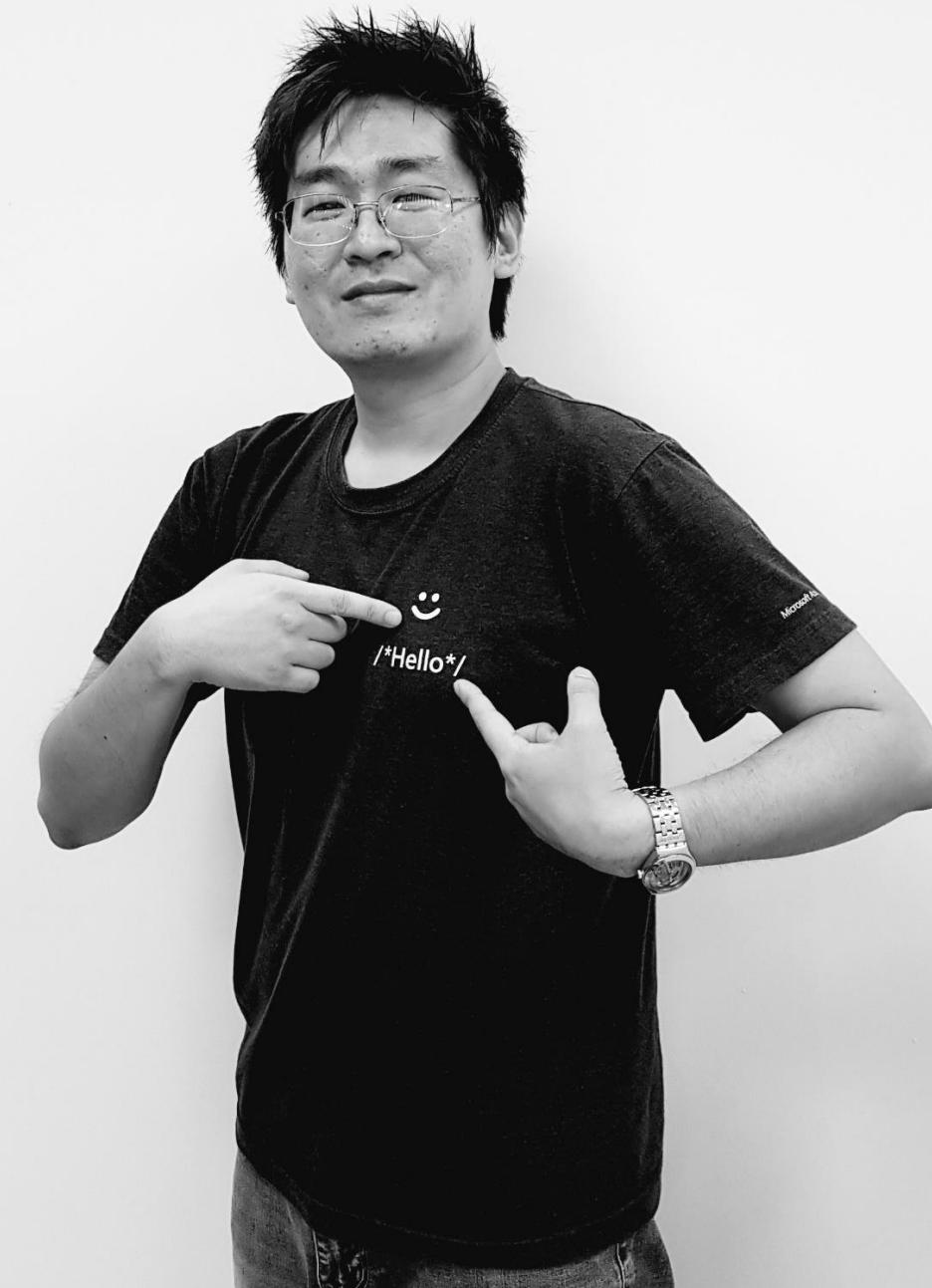


#GlobalAzure

# 배준현

클라우드메이트 Lead Engineering 팀

- Microsoft Azure MVP
- Ubuntu Linux Korea Community Regional Council Leader
- 게임허브 AI팀 팀장
- 유튜브 빅데이터 분석



# Agenda

1. Machine Learning ?
2. Team Data Science Process(TDSP)
3. TDSP 프로젝트 운영



#GlobalAzure

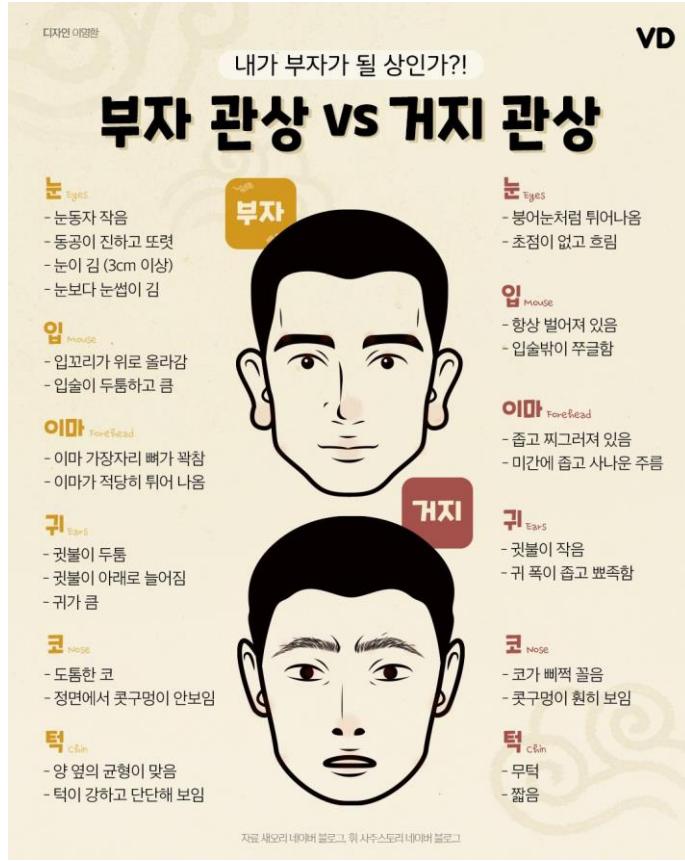
# Machine Learning 어려워요...



#GlobalAzure

Back labradoodle or fried chicken

ML에서 어려운것은 ?



산업 > 바이오&ICT

## "마이너리티 리포트처럼... 범죄예측 AI CCTV 개발"

입력 2020-01-02 17:16:46      수정 2020.01.02 17:16:46      임진혁 기자



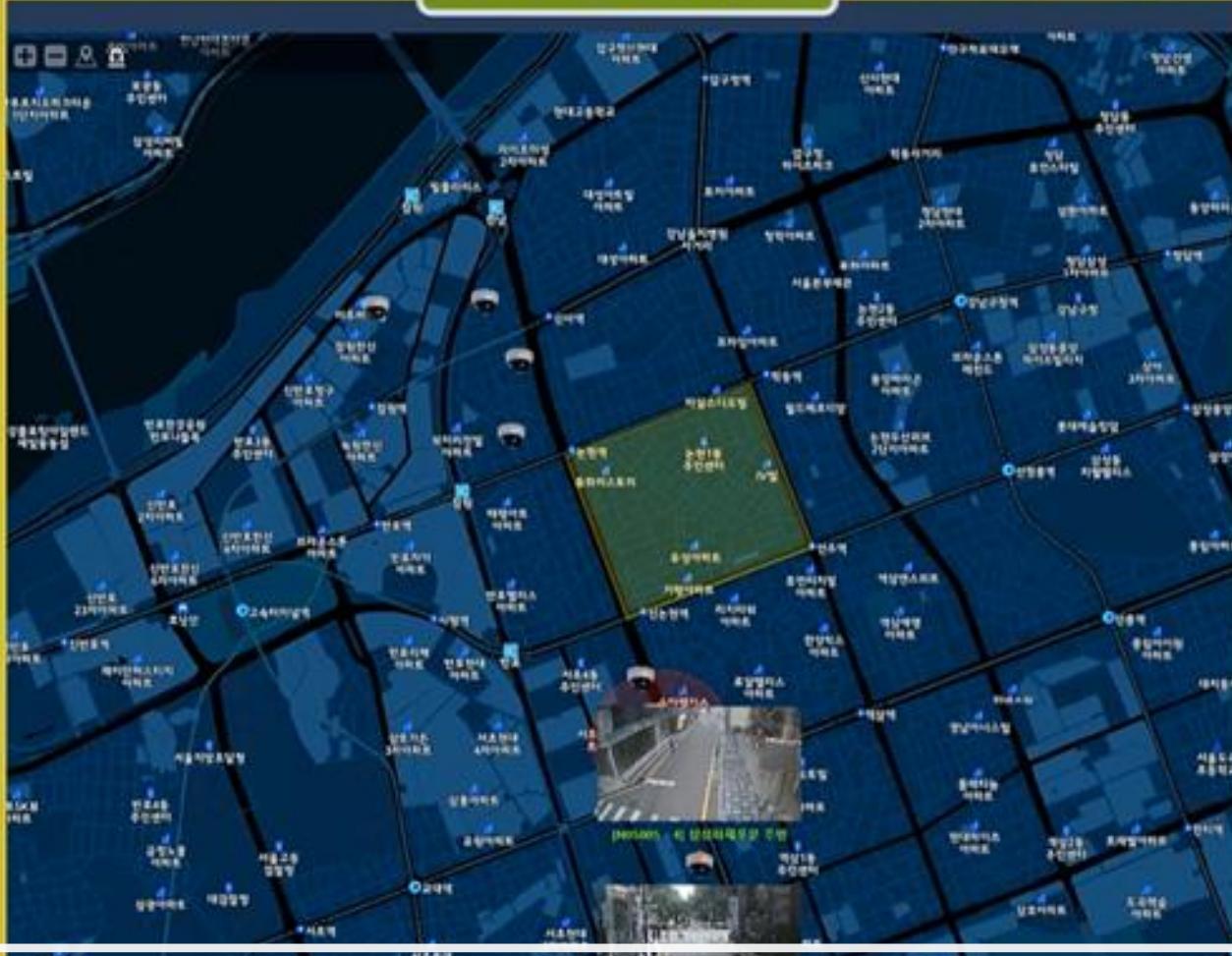
ETRI "2022년 원천기술 확보"  
발걸음 소리 등 AI가 상황 분석  
위험발생 가능성 80%까지 예측



# ML에서 어려운 것은 ?

데이터

## PCM 전자지도



## VMS 연동 영상관제



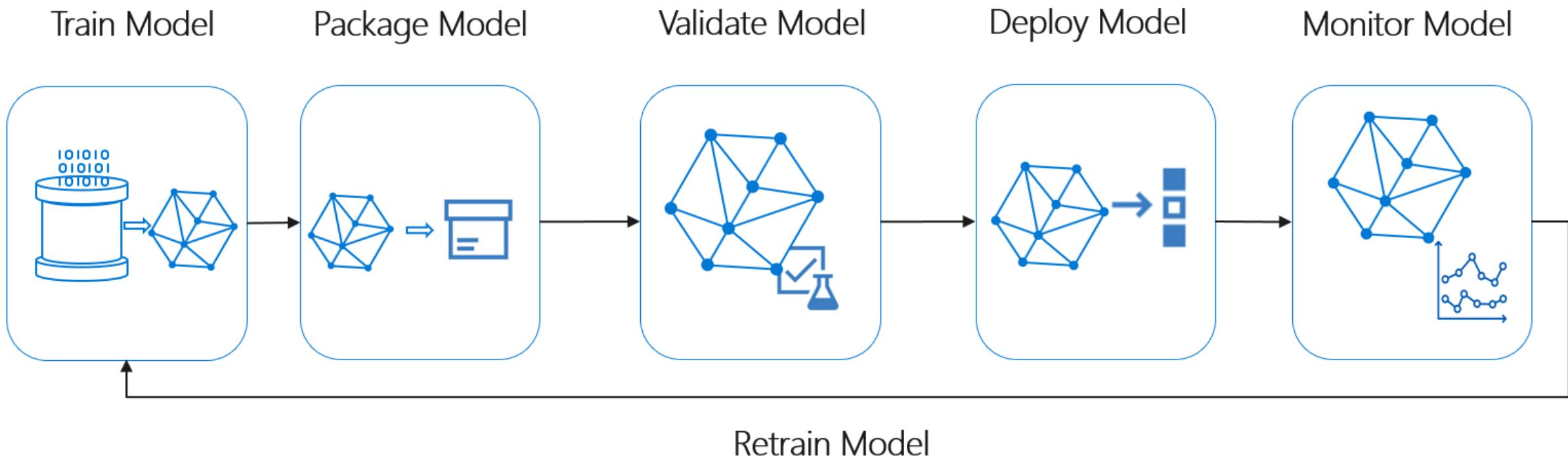
## 휴면속성 분석

Analytical Information	
Name	History
Person ID: 1	성별: Male 나이: 30 인종: CR 모자: CR 가방: CR
Person ID: 2	성별: Male 나이: 40 인종: CR 모자: CR 가방: CR
Person ID: 3	성별: Female 나이: 20 인종: CR 모자: CR 가방: CR
Person ID: 4	성별: Male 나이: 50 인종: CR 모자: CR 가방: CR
Person ID: 5	성별: Male 나이: 35 인종: CR 모자: CR 가방: CR
Person ID: 6	성별: Female 나이: 25 인종: CR 모자: CR 가방: CR
Person ID: 7	성별: Male 나이: 45 인종: CR 모자: CR 가방: CR
Person ID: 8	성별: Female 나이: 30 인종: CR 모자: CR 가방: CR
Person ID: 9	성별: Male 나이: 55 인종: CR 모자: CR 가방: CR
Person ID: 10	성별: Female 나이: 28 인종: CR 모자: CR 가방: CR

ML에서 어려운 것은 ?

# ML 파이프라인

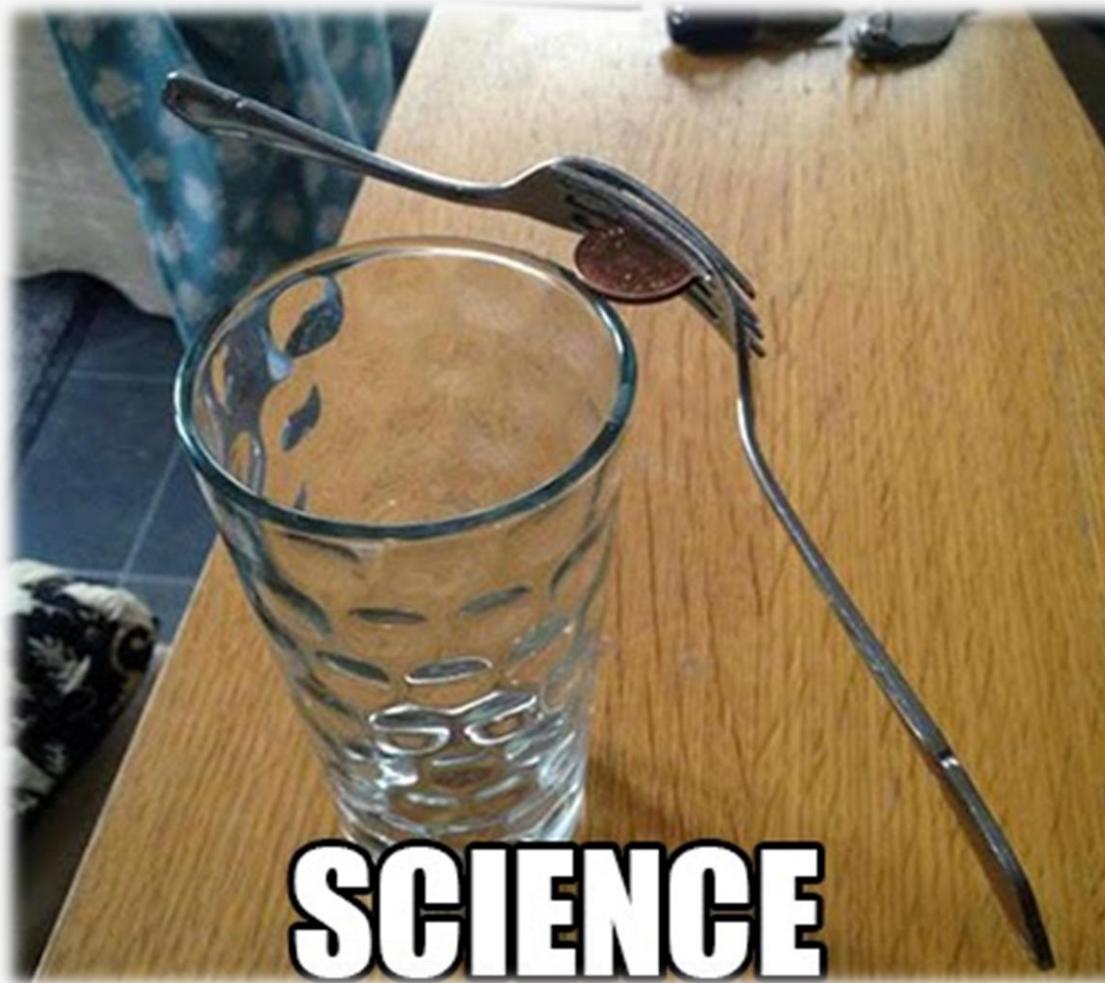
모델을 구성하고 >> 실험 >> 실험이 잘되면 >> 알아서 작동되게



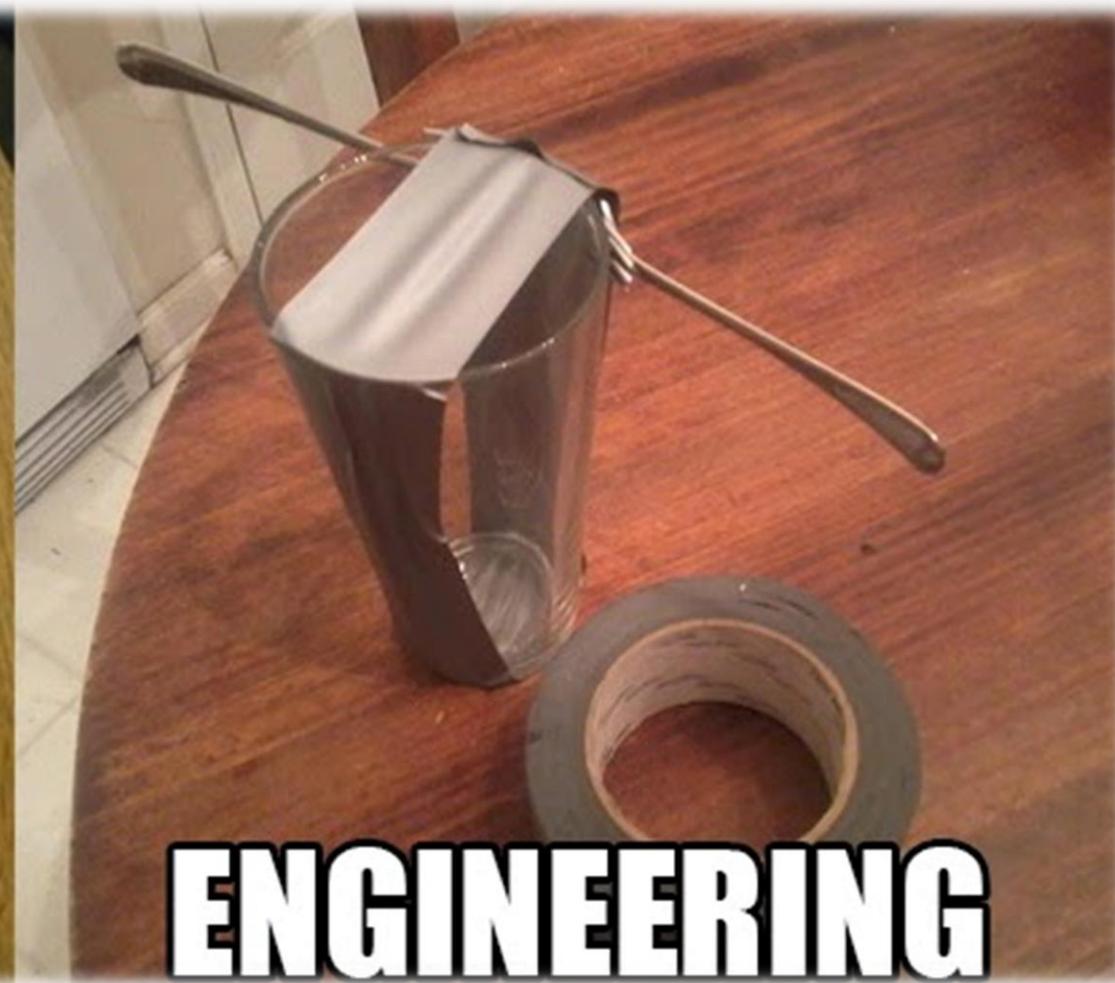


알아서...?

# Scientist / Engineer 그들을 이해 해야 한다



**SCIENCE**



**ENGINEERING**

# 요구사항을 정의

1. 클라우드 인프라 사용
2. 로컬 ML 실험 연동
3. 클라우드에서의 작동 보장
4. 이 모든 것을 자동화
5. 보안에 대한 우려
6. 기왕이면 DevOps 문화 도입도...



#GlobalAzure

# TDSP (Team Data Science Process)



#GlobalAzure

## 제목으로 필터링

## Team Data Science Process 설명서

## 개요

- > 수명 주기
- > 역할 및 작업
- 프로젝트 계획
- > 개발
- > 운영화
- > 연습 예제
- > 학습
- > 방법
- > 관련 항목
- > 리소스

# Team Data Science Process란 무엇인가요?

2020. 11. 17. • 읽는 데 8분 걸림 • +3

TDSP(Team Data Science Process)는 예측 분석 솔루션 및 지능형 애플리케이션을 효율적으로 제공하는 기민한 반복 데이터 과학 방법론입니다. TDSP는 팀 역할이 가장 잘 작동하는 방법을 제안하여 팀 협업 및 학습을 개선하는 데 도움이 됩니다. TDSP는 데이터 과학 이니셔티브를 성공적으로 구현하는 데 도움이 되는 Microsoft 및 기타 산업 선두 업체의 모범 사례 및 구조를 포함합니다. 목표는 회사가 해당 분석 프로그램의 이점을 완전히 이해하도록 돋는 것입니다.

이 문서에서는 TDSP 및 주요 구성 요소의 개요를 제공합니다. 다양한 종류의 도구로 구현될 수 있는 프로세스의 일반적인 설명을 제공합니다. 프로젝트 작업 및 프로세스의 수명 주기와 관련된 역할의 더 자세한 설명은 추가 연결된 항목에서 제공됩니다. 팀에서 TDSP를 구현하는 데 사용하는 Microsoft 도구 및 인프라의 특정 집합을 사용하여 TDSP를 구현하는 방법에 대한 지침도 제공됩니다.

## TDSP의 주요 구성 요소

PDF 다운로드

TDSP에는 다음과 같은 주요 구성 요소가 있습니다.

이 페이지가 도움  
이 되었나요?

Yes    No

### 이 문서의 내용

#### TDSP의 주요 구성 요소

[데이터 과학 수명  
주기](#)

[표준화된 프로젝트  
구조](#)

[데이터 과학 프로젝  
트에 대한 인프라  
및 리소스](#)

[프로젝트 실행을 위  
한 도구 및 유틸리  
티](#)

[다음 단계](#)

# TDSP (Team Data Science Process)

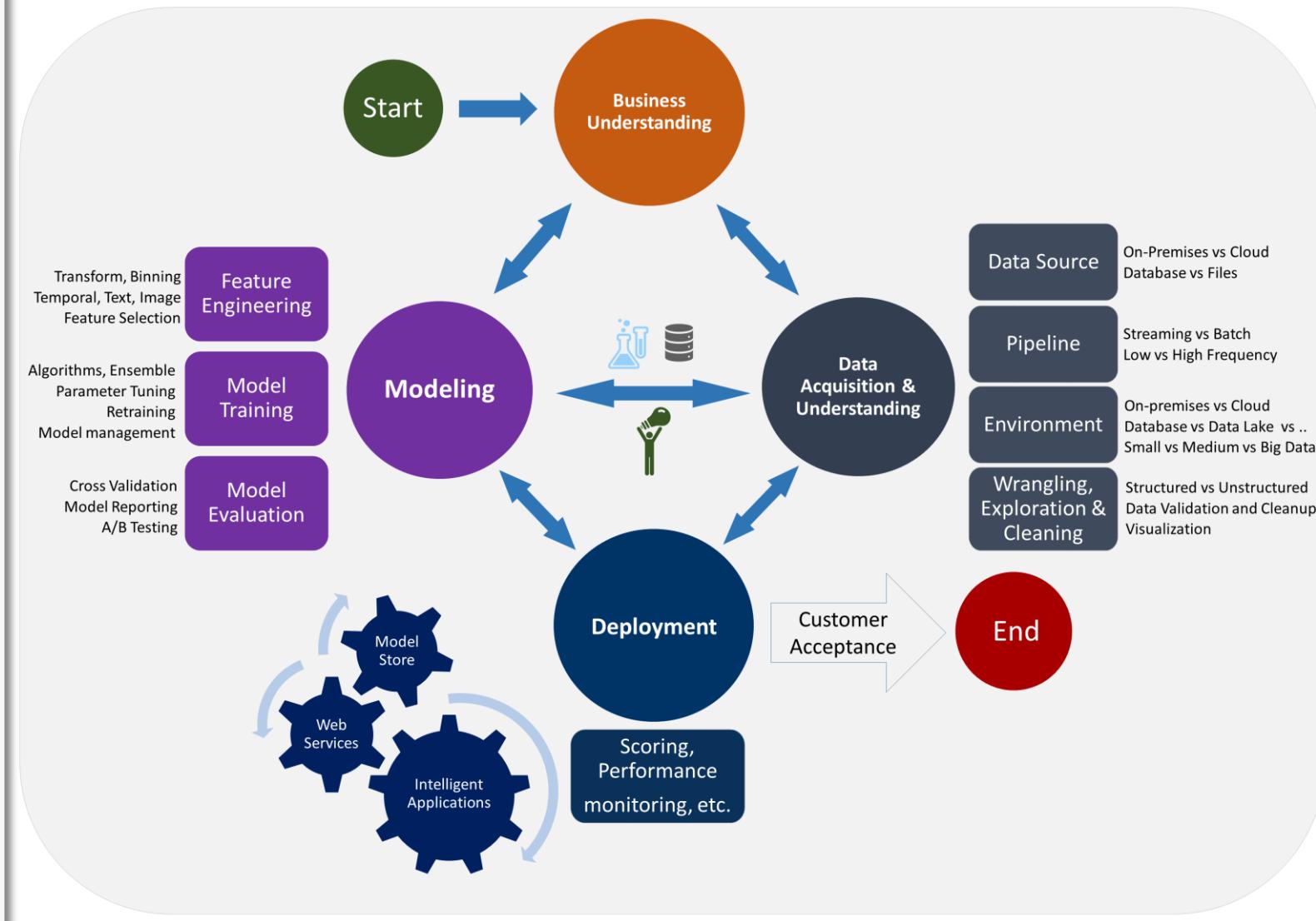
TDSP(Team Data Science Process)는 예측 분석 솔루션 및 지능형 애플리케이션을 효율적으로 제공하는 데이터 과학 방법론입니다.

TDSP는 팀 역할이 가장 잘 작동하는 방법을 제안하여 팀 협업 및 학습을 개선하는데 도움이 됩니다. TDSP는 데이터 과학 이니셔티브를 성공적으로 구현하는데 도움이 되는 Microsoft 및 기타 산업 선두 업체의 모범 사례 및 구조를 포함합니다.

- **데이터 과학 수명 주기 정의**
- **표준화된 프로젝트 구조**
- **데이터 과학 프로젝트에 권장되는 인프라 및 리소스**
- **프로젝트 실행에 권장되는 도구 및 유틸리티**

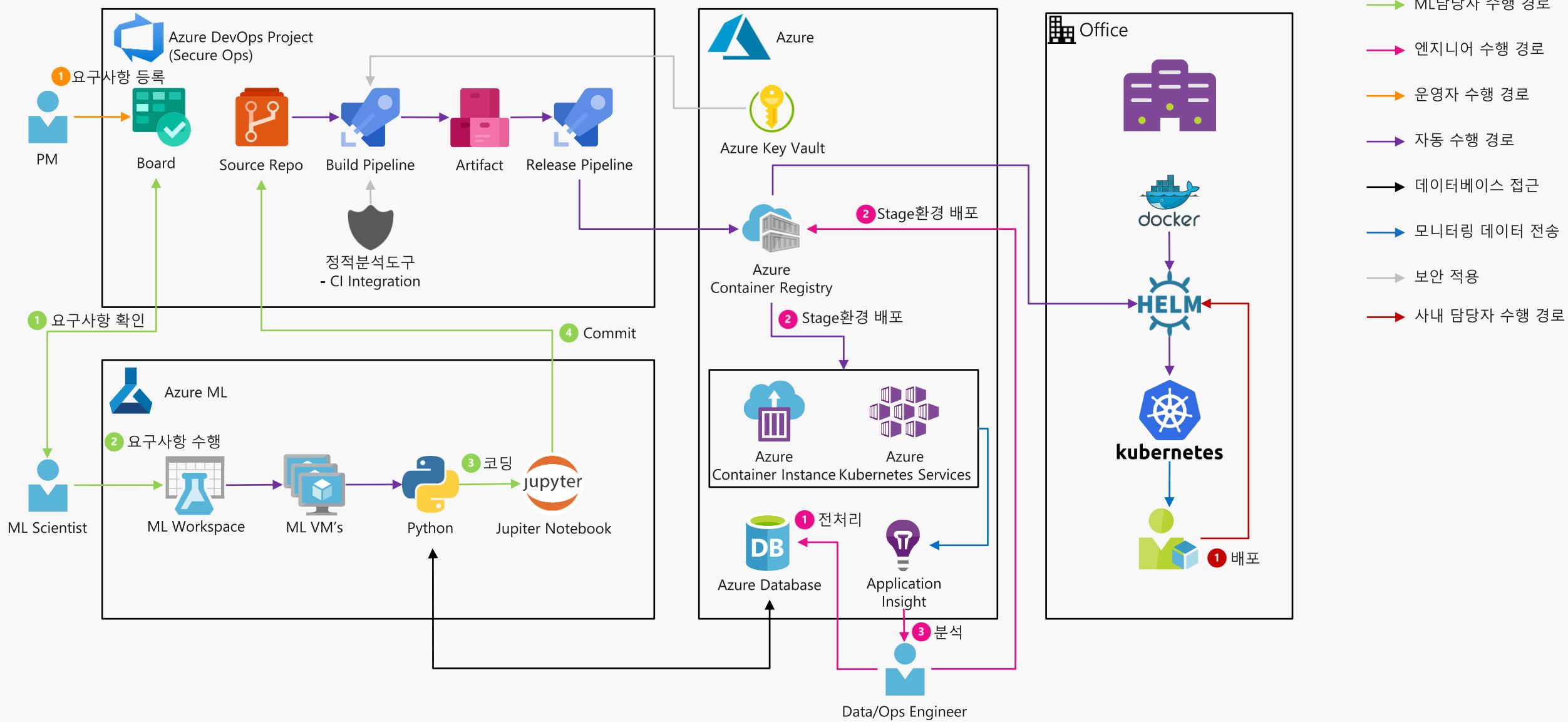
# 1. 데이터 과학 수명 주기 정의

## Data Science Lifecycle

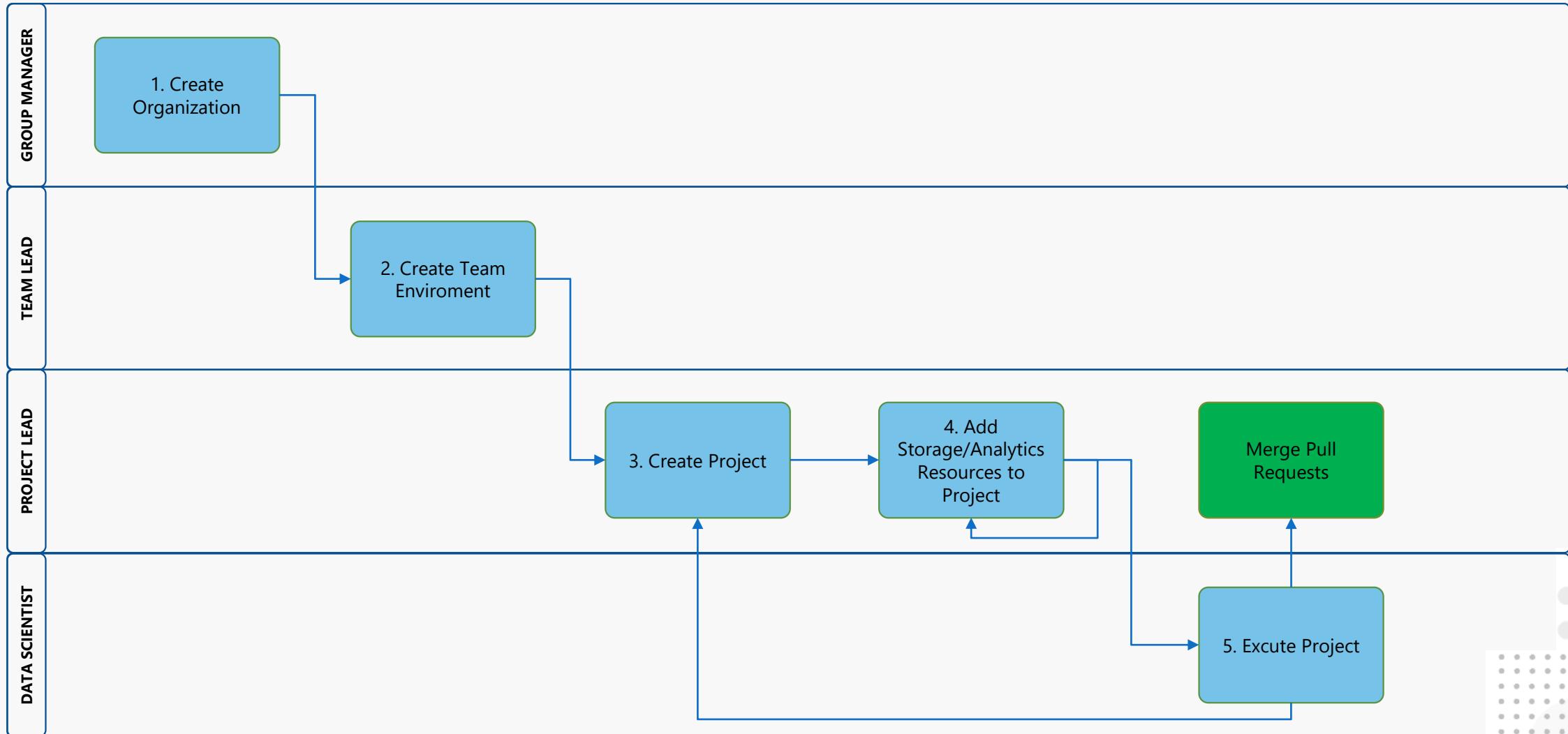


시나리오에 따른  
라이프사이클이  
달라질 수 있음.

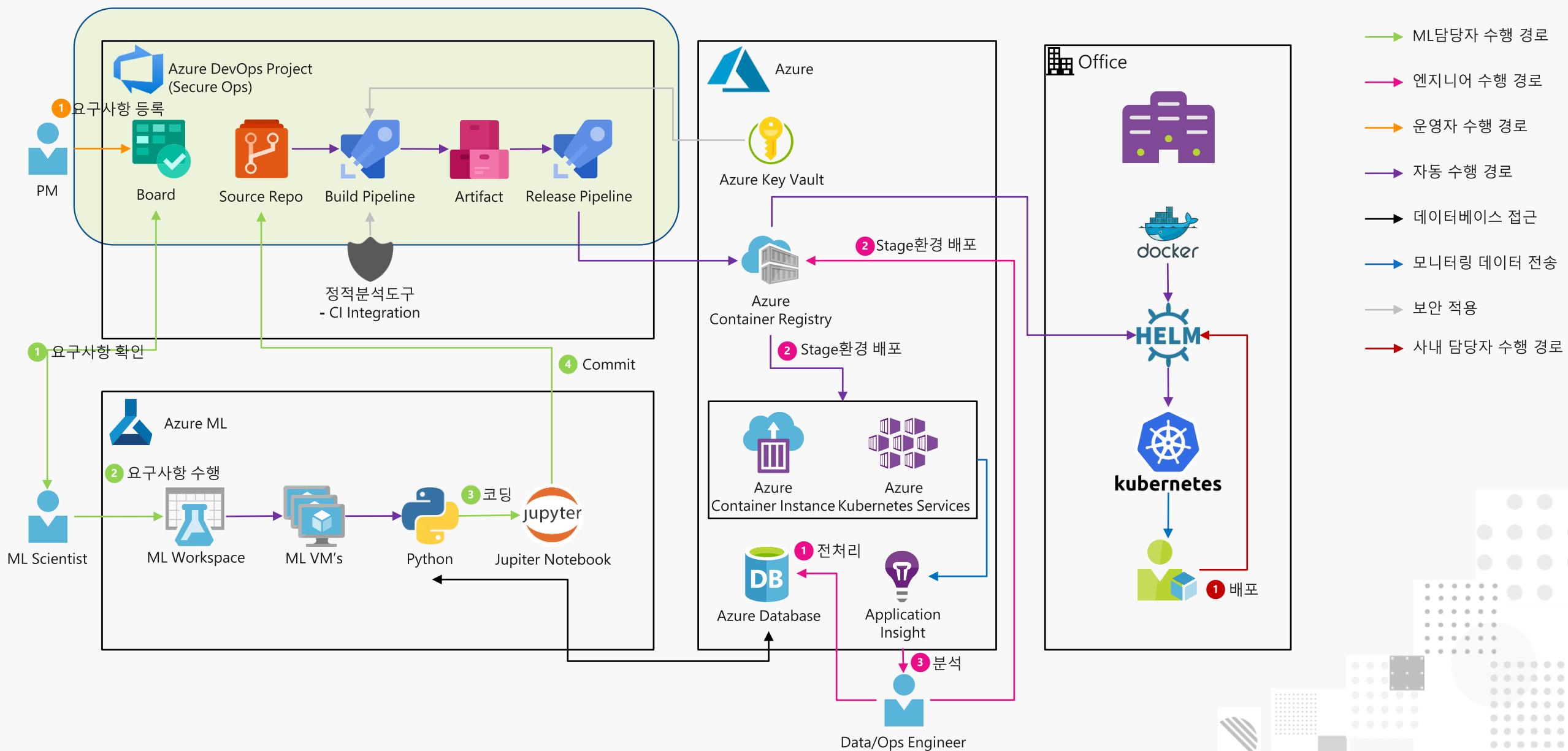
# 1. 데이터 과학 수명 주기 정의



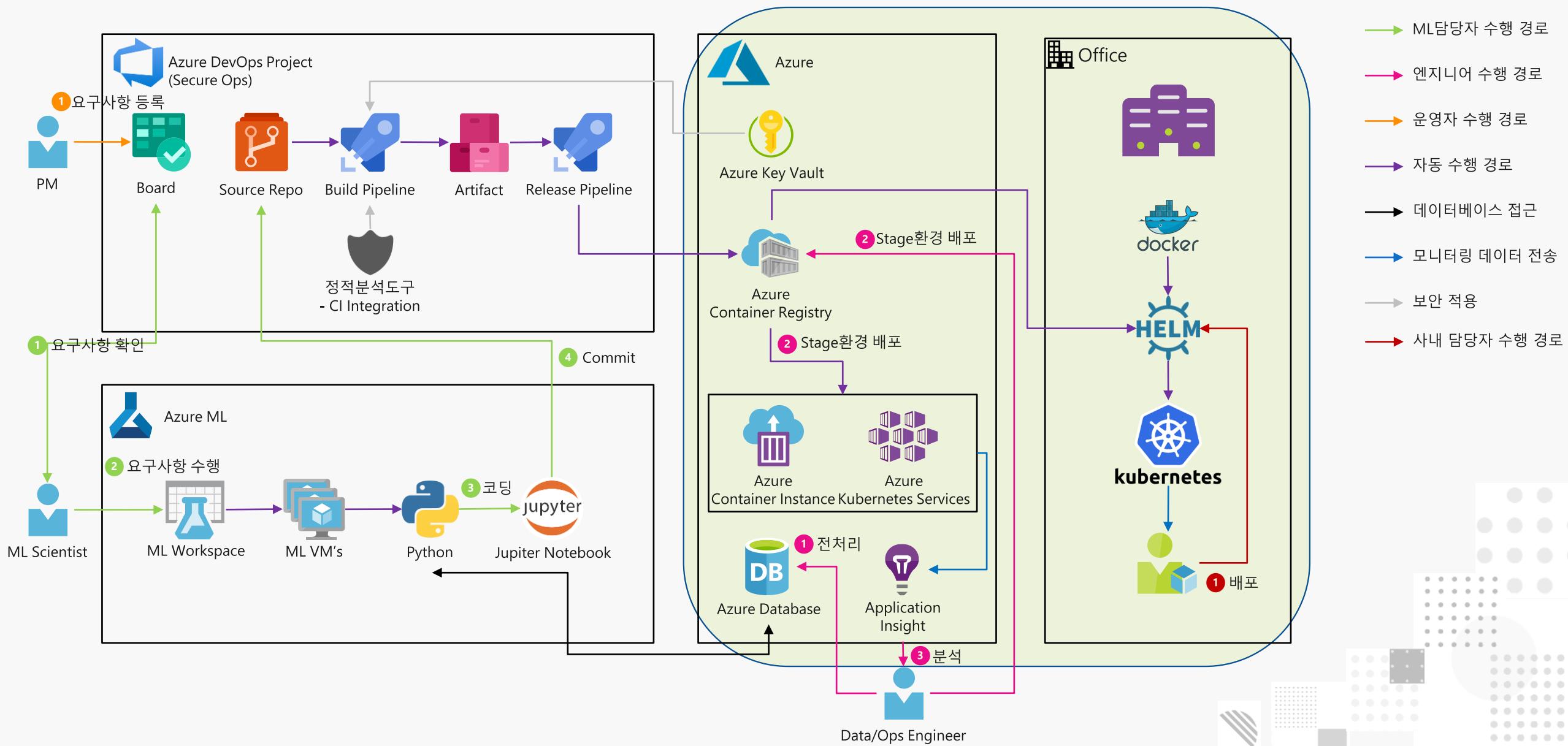
## 2. 표준화된 프로젝트 구조



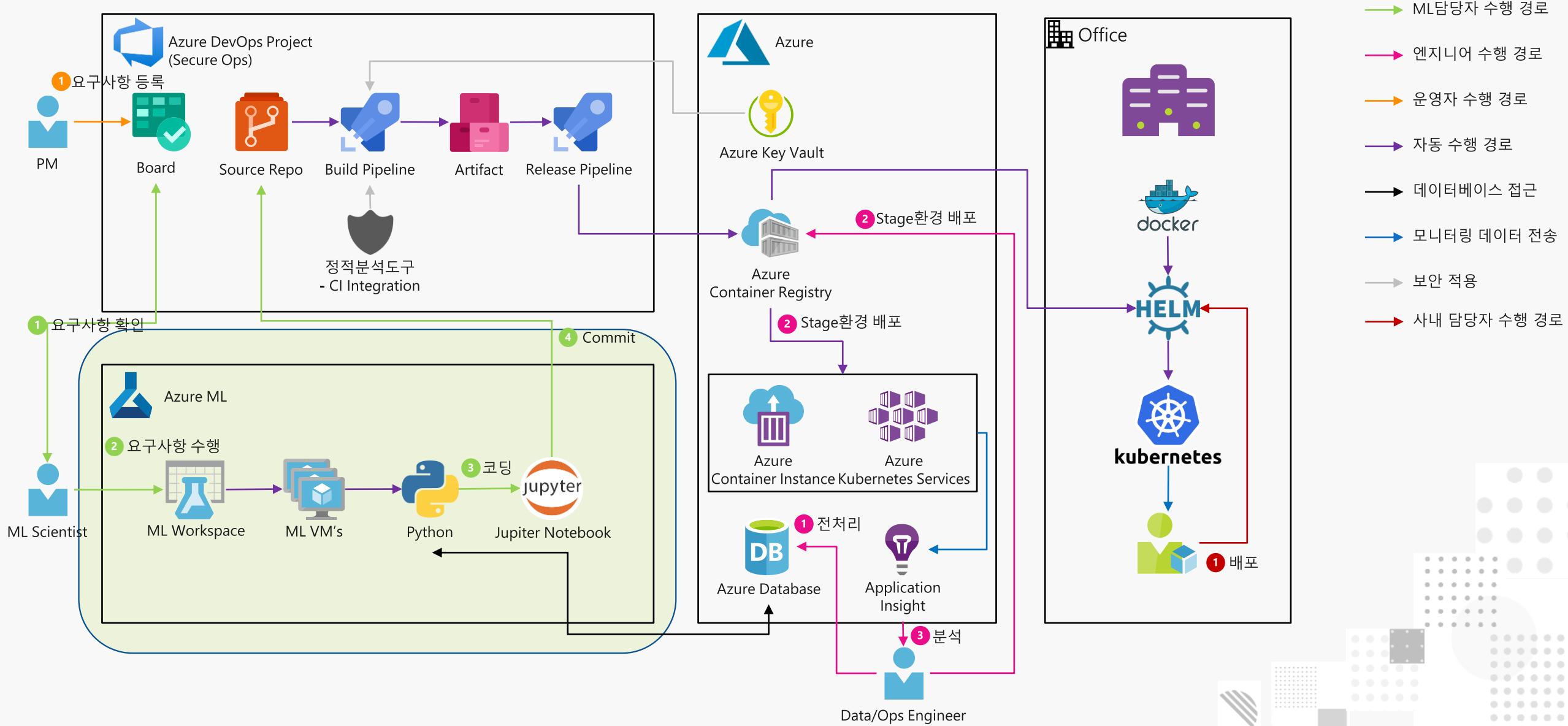
# 2. 표준화된 프로젝트 구조



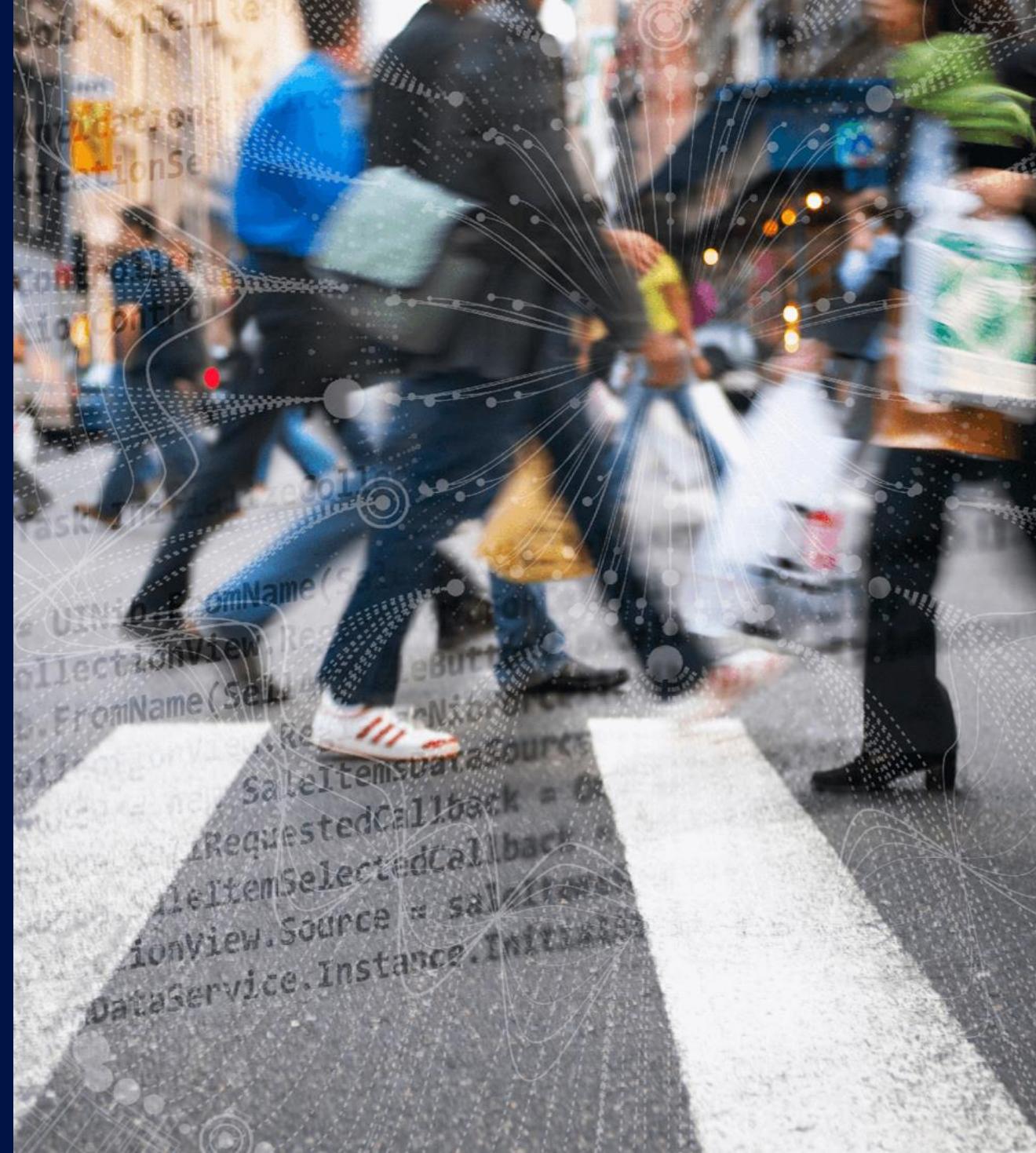
### 3. 데이터 과학 프로젝트에 권장되는 인프라 및 리소스



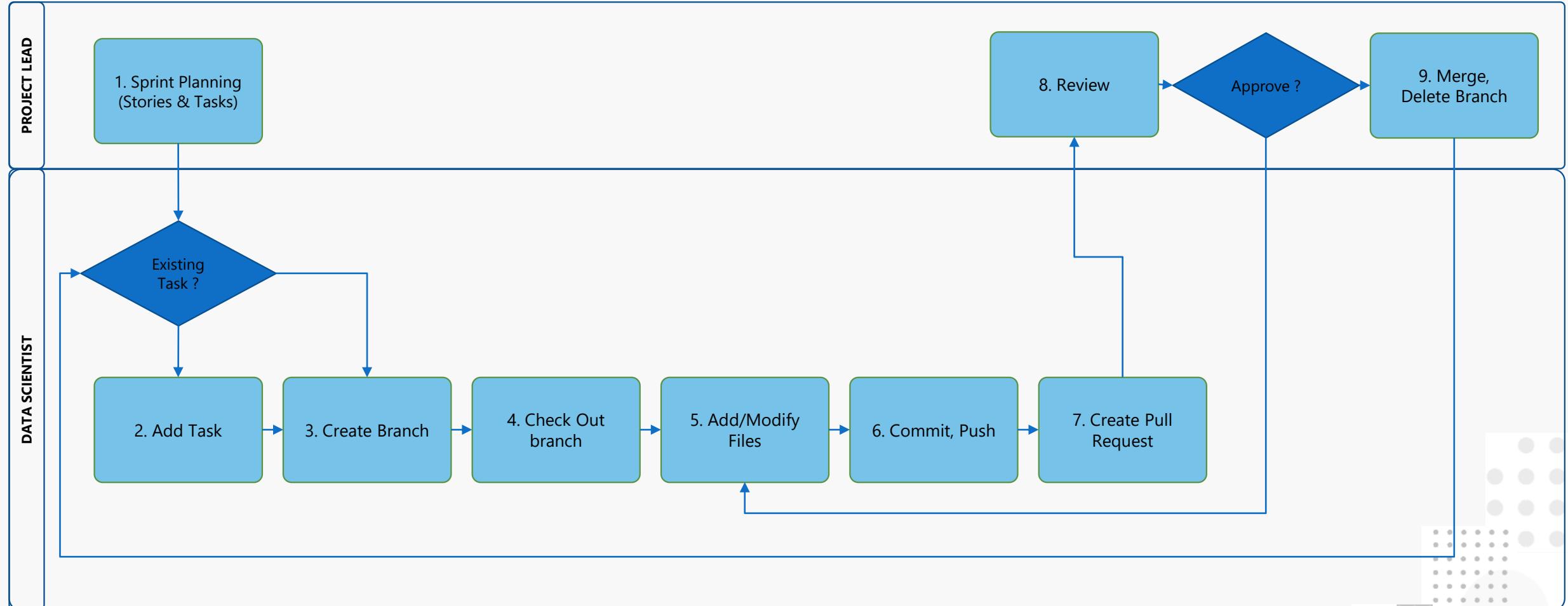
# 4. 프로젝트 실행에 권장되는 도구 및 유틸리티



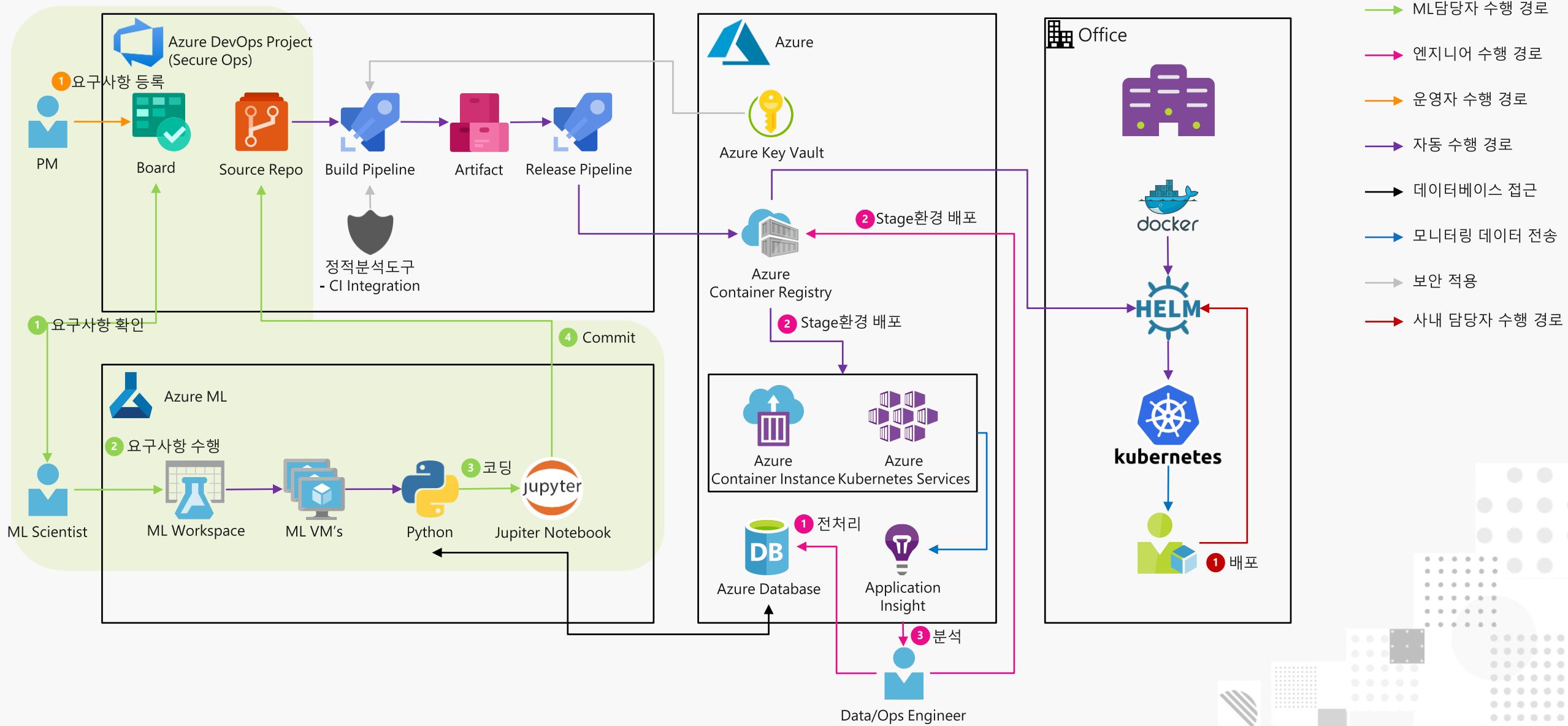
# TDSP 프로젝트 운영



# 1. 기민한 개발을 위한 방법론

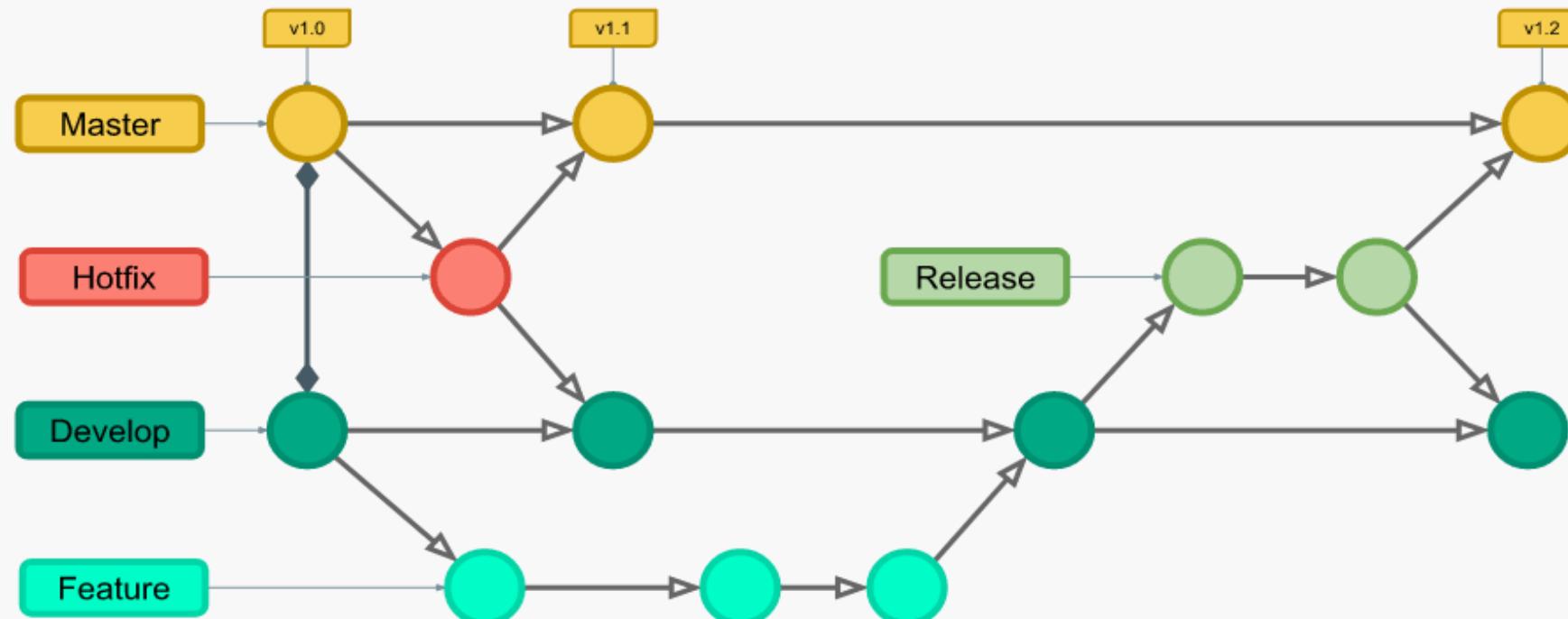


# 1. 기민한 개발을 위한 방법론

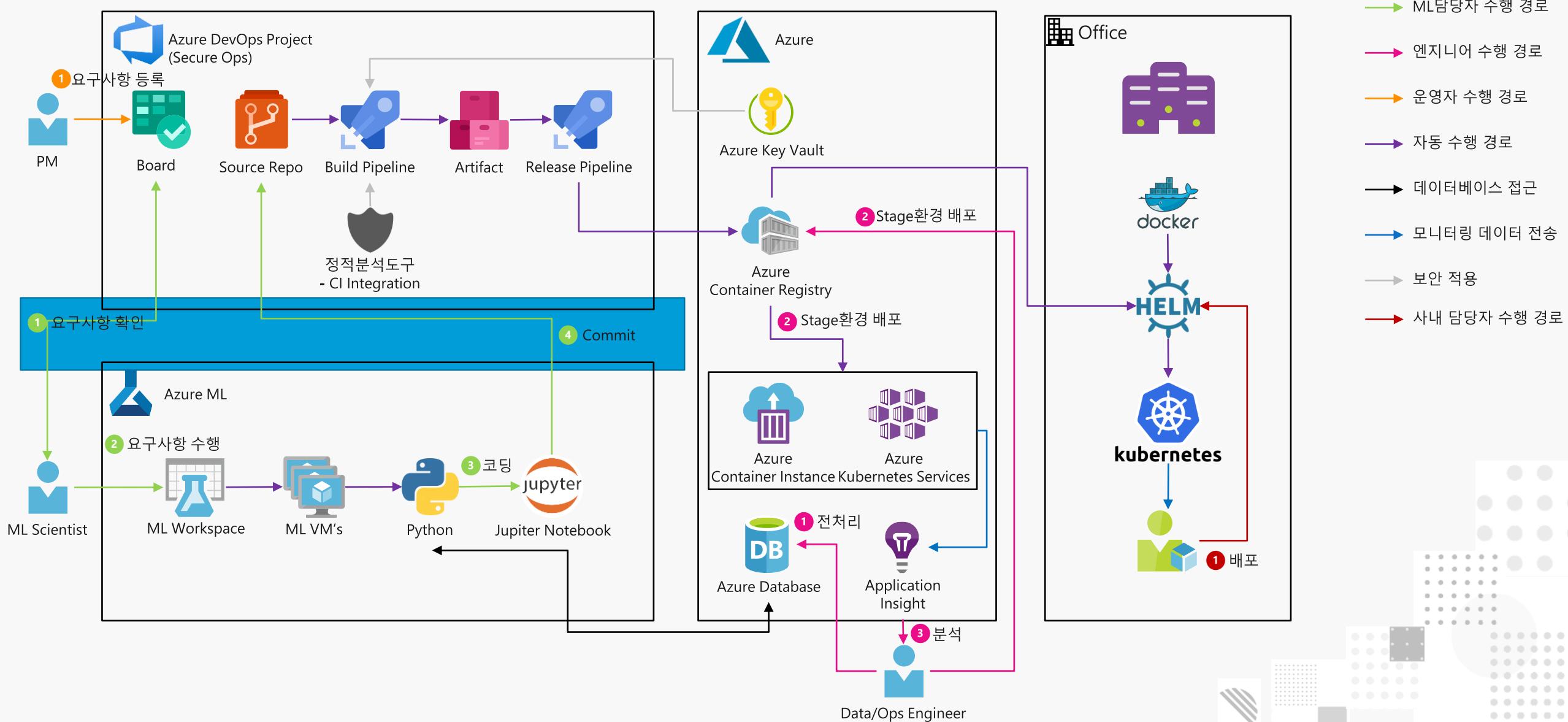


## 2. GIT을 사용한 공동 코딩

Vincent Driessen의 Branching Model은 개발자가 더 큰 소프트웨어 프로젝트에서 feature, hotfixes 및 releases를 추적하는 데 도움이 되는 git branch 및 release management 전략입니다. git 하위 명령의 git-flow 라이브러리가 있어 흐름의 일부분을 자동화하여 작업을 훨씬 쉽게 할 수 있습니다.



## 2. GIT을 사용한 공동 코딩



# 3. 데이터 과학 작업

Microsoft Azure Machine Learning

새로 만들기

홈

작성자

Notebooks

자동화된 ML

디자이너

자산

데이터 세트

실험

파이프라인

모델

엔드포인트

관리

컴퓨팅

데이터 저장소

데이터 레이블 지정

연결된 서비스

+

새로 만들기

Notebooks

Python SDK를 사용하여 코딩하고 샘플 실험을 실행합니다.

지금 시작

자동화된 ML

대상 메트릭을 사용하여 모델을 자동으로 학습하고 튜닝합니다.

지금 시작

디자이너

데이터 준비에서 모델 배포로 끌어서 놓기 인터페이스입니다.

지금 시작

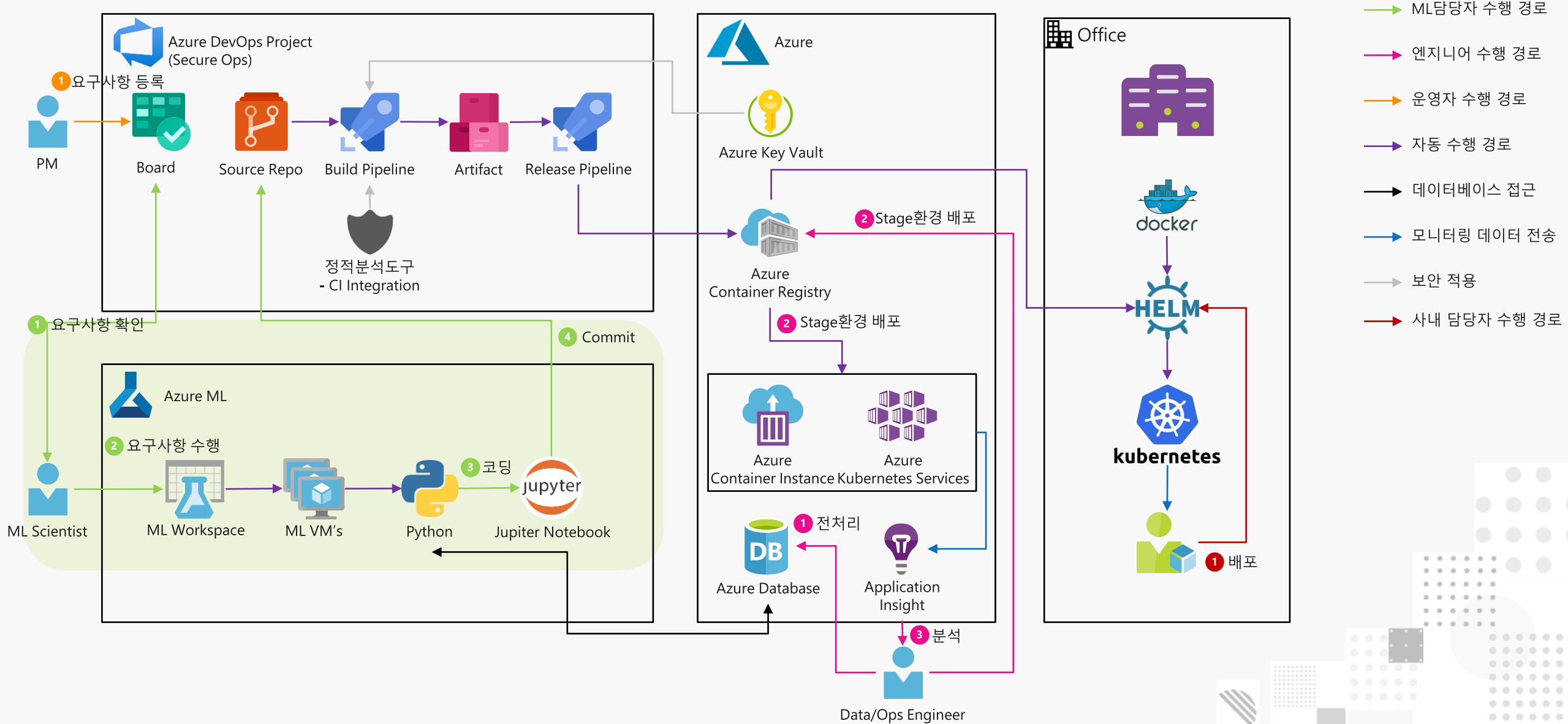
## Azure Machine Learning Studio

### 내 최근 리소스

#### 실행

실행	실행 ID	실험	상태	제출된 시간	제출한 사람	실행 형식
실행 10	16861e26-f551-4b1...	diabetes-...	완료됨	2021-01-05T06:47:5...	Junhyun Bae	
실행 9	281bcc54-f9ba-48fe...	diabetes-...	완료됨	2021-01-05T06:47:4...	Junhyun Bae	
실행 8	c8f03e30-4bb6-498...	diabetes-...	완료됨	2021-01-05T06:47:4...	Junhyun Bae	
실행 7	4581b3fb-6293-4d7...	diabetes-...	완료됨	2021-01-05T06:47:3...	Junhyun Bae	
실행 6	2386f7db-b38b-481...	diabetes-...	완료됨	2021-01-05T06:47:2...	Junhyun Bae	
실행 5	3ff91a73-b4b4-4c6c...	diabetes-...	완료됨	2021-01-05T06:47:1...	Junhyun Bae	

### 3. 데이터 과학 작업



# 4. 작업 테스트

diabetes-experiment

열 편집 새고침 차트 추가 현재 보기: 기본값 보기 저장 보기 편집 보기 공유

검색 필터 추가 자식 실행 포함 내 실행만 보기

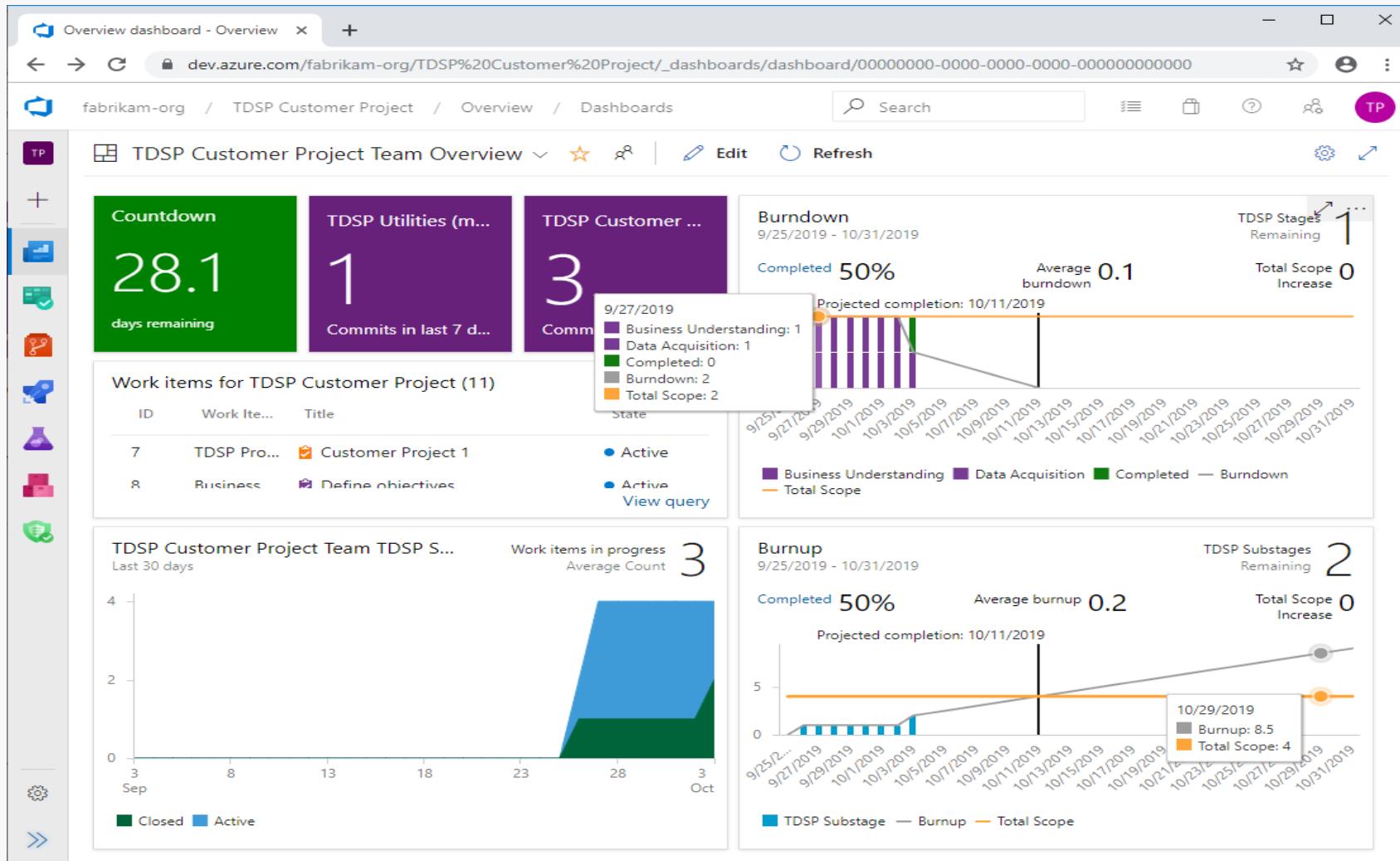
실행 상태  
● 실행 중 0 ● 실패 0 ● 대기 0 선택한 행만 표시 (5개 선택됨)

alpha\_value

rmse

실행	실행 ID	상태	제출된 시간 ↓	기간	제출한 사람	컴퓨팅 대상	실행 형식	마지막(alpha...)	마지막(rmse)	태그
실행 10	16861e26-f551-4b10-b36c-e7cf4e04...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		1	56.661	
실행 9	281bcc54-f9ba-48fe-80a8-2d3b4bee0...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		0.9	56.654	
실행 8	c8f03e30-4bb6-4980-8748-9f49846e...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		0.8	56.647	
실행 7	4581b3fb-6293-4d7b-8798-03929ed...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		0.7	56.641	
실행 6	2386f7db-b38b-4818-823c-3440307...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		0.6	56.634	
실행 5	3ff91a73-b4b4-4c6c-beb6-8dcb52261...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		0.5	56.628	
실행 4	dd702787-293d-48d9-ace3-8f162f33...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		0.4	56.622	
실행 3	80720bd5-3dfe-47ab-979e-d232f7c3...	완료됨	Jan 5, 2021 3:47 PM	8초	Junhyun Bae	local		0.3	56.616	

# 5. 진행률 추적



# TDSP 기반의 애저 PaaS 기반 ML 플랫폼 구성 방법론의 결론

# 클라우드 PaaS / SaaS 복합 구성

1. 클라우드를 통해 SLA 보장
2. PaaS 도입으로 구축의 편의성
3. SaaS 도입으로 사용성 증대



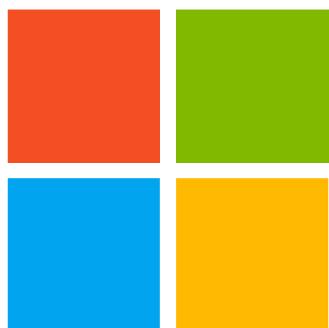
포기하면  
그 순간이 바로  
**개발완료**  
예요…

감사합니다.





**CLOUDMATE**  
Managed Service Expert



Microsoft