KDT 프로젝트 기반 빅데이터 개발자 양성 과정

데이터 엔지니어링 이현수

데이터 엔지니어링 기초

강사 소개

천재교육 AI센터 개발운영팀
데이터 엔지니어

| 프로필

NAME 이현수
E-MAIL hyunsooyein@chunjae.co.kr

| 학력

부산중앙고등학교
연세대학교 기계공학 전공

| 사내 업무

사내 데이터 분석 요구사항 대응
클라우드 환경에서의 데이터 파이프라인 구축/운영
AWS 및 네이버 클라우드 인프라 및 데이터 관리

| 자격증

정보처리기사
빅데이터분석기사
AWS Data Analytics - Specialty
AWS Solutions Architect - Associate
AWS Developer - Associate
Apache Spark Associate Developer

| 개인 사이트

Github > https://github.com/Hyunsoo-Ryan-Lee
Linkedin > https://www.linkedin.com/in/hyunsoo-ryan-lee-824a7917a/
Blog > https://velog.io/@newnew_daddy

데이터 엔지니어링 교육과정 강의 일정

1일차 > 데이터 엔지니어링의 개요 및 실습

- > 데이터 엔지니어링 소개
- > 천재교육 실무에서의 데이터 엔지니어링
- > 실습 범위 안내 및 기초 실습

2일차 > 데이터 파이프라인 구성 실습

- > Sub Module 구성
- > Main Module 구성

3일차 > 데이터 파이프라인 End-to-End 프로젝트

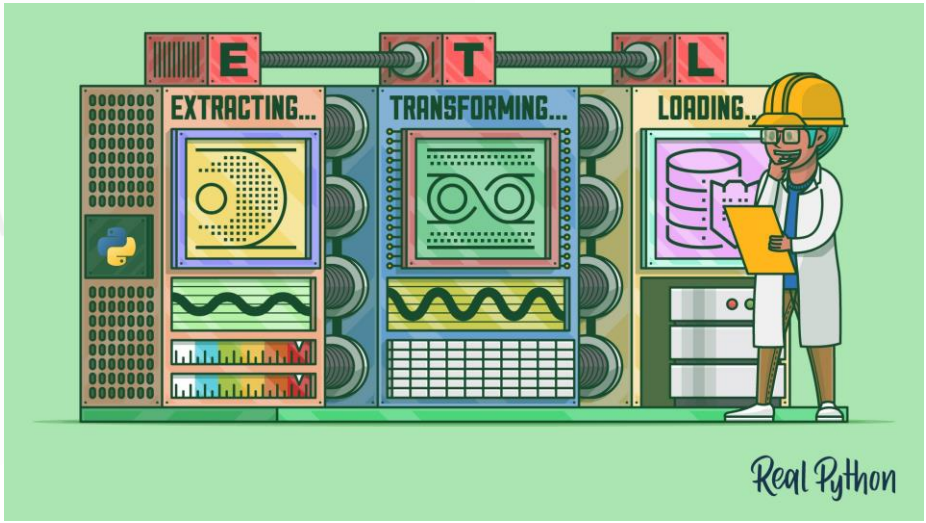
- > 프로젝트 아키텍처 소개
- > 프로젝트 실습

4일차 > Apache Spark의 개요 및 실습

- > 데이터 파이프라인 프로젝트 리뷰
- > Apache Spark 소개
- > Pyspark 환경구성 & 코드 실습
- > 과제 안내

5일차 > Cloud 에서의 데이터 엔지니어링

- > Spark SQL & ML 실습
- > AWS 서비스를 이용한 데이터 처리
- > AWS 서비스 & 관련 자격증 소개



데이터 엔지니어링 소개

01 데이터 엔지니어링이란 무엇인가

02 데이터 엔지니어링이 왜 필요할까?

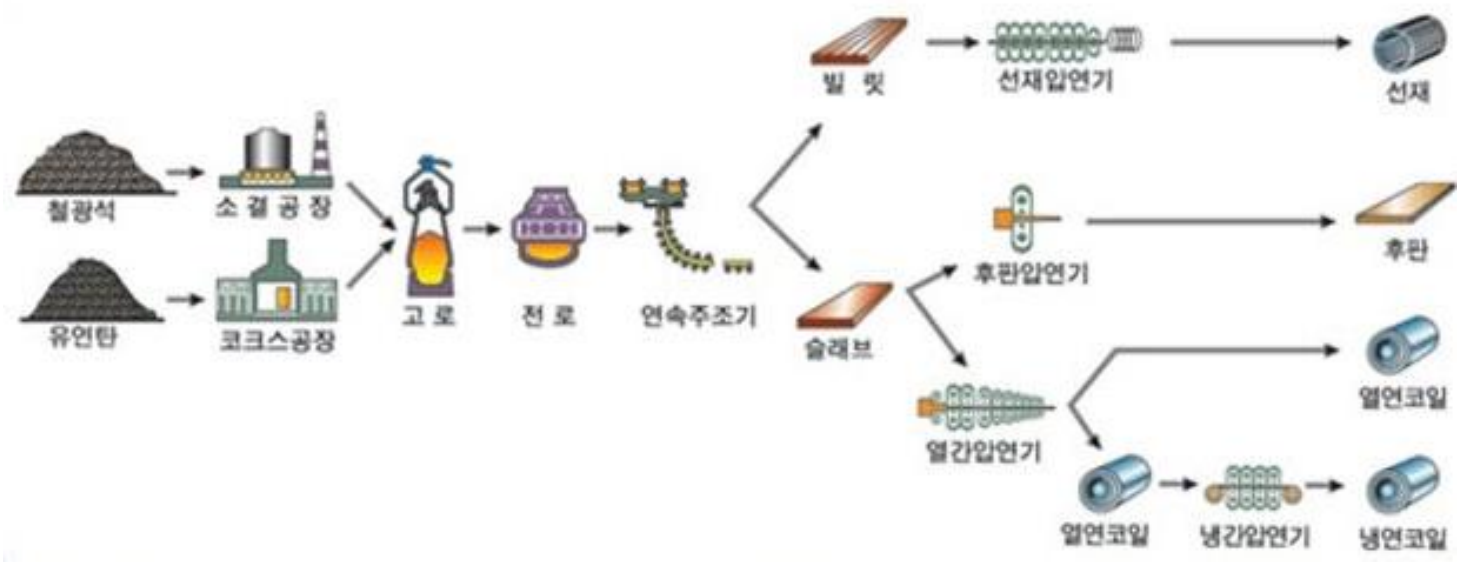
03 데이터 엔지니어가 하는 일은?

04 데이터 엔지니어의 Skill Set

데이터 엔지니어링이란 무엇인가

▶ 엔지니어링의 정의

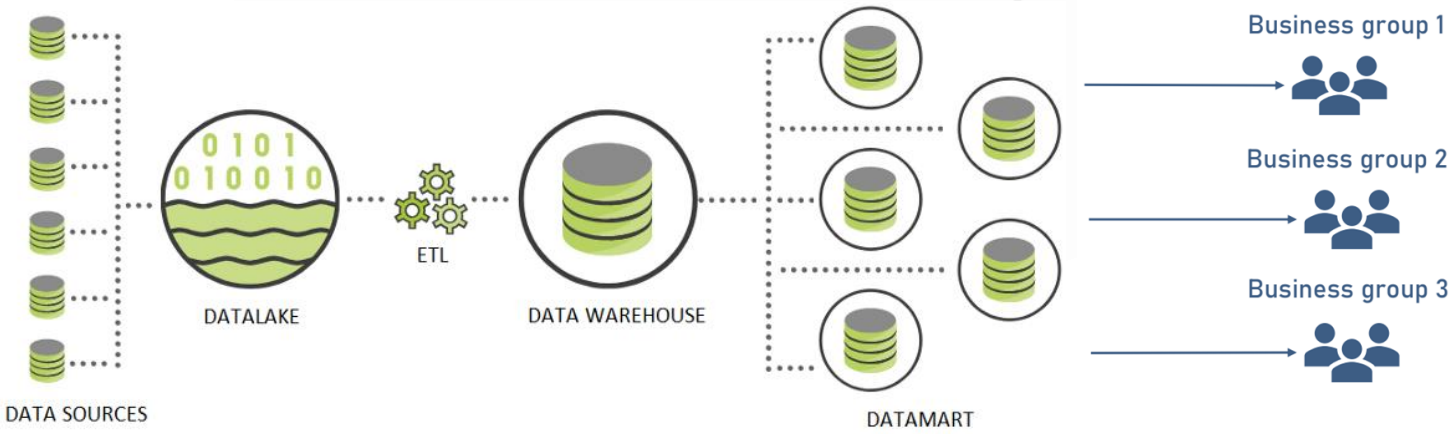
▪ 공업 분야의 응용과학 기술을 연구하는 학문 또는 과학적, 경제학적, 사회적 원리와 실용적 지식을 활용하여 새로운 제품, 도구, 건축물 · 조형물, 시설 등을 만드는 것에 관한 학문.



데이터 엔지니어링이란 무엇인가

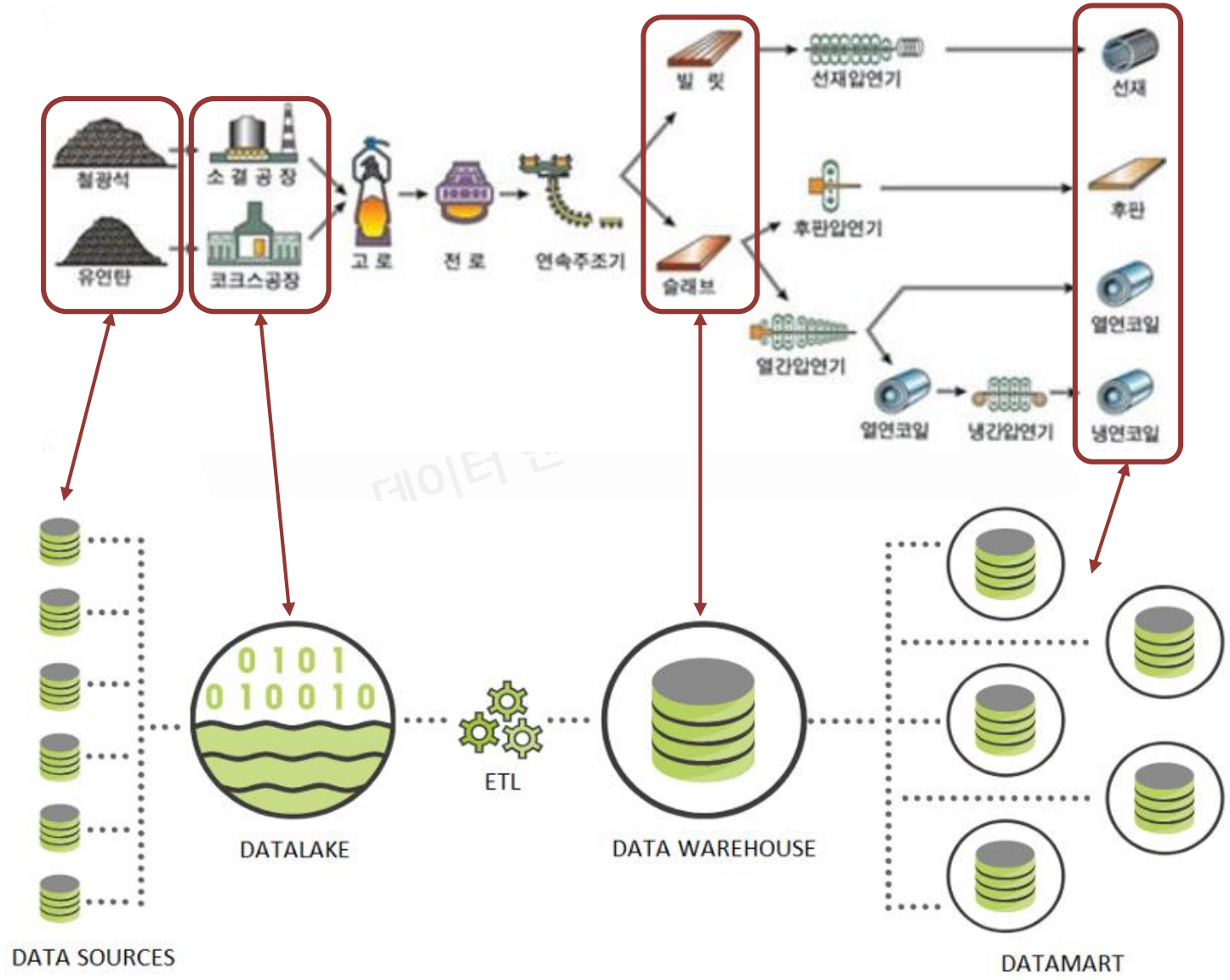
▶ 데이터 엔지니어링이란?

- 데이터 엔지니어링은 데이터를 수집하고 활용할 수 있도록 시스템을 구축하는 것
- 대규모 데이터를 효율적으로 수집, 저장, 처리 및 전송하기 위한 기술과 인프라를 개발하는 작업



데이터 엔지니어링이란 무엇인가

▶ 데이터 엔지니어링이란?



데이터 엔지니어링이란 무엇인가

- Data Lake , Data Warehouse , Data Mart, Data Governance

데이터 엔지니어링 이현수

데이터 엔지니어링이란 무엇인가

▶ Data Lake , Data Warehouse , Data Mart, Data Governance

Data Lake

다양한 형식과 소스에서 대규모의 비정형 및 정형 데이터를 저장하고,
이를 추후 분석이나 처리를 위해 보관하는 저장소

데이터 엔지니어링이란 무엇인가

▶ Data Lake , Data Warehouse , Data Mart, Data Governance

Data Lake

다양한 형식과 소스에서 대규모의 비정형 및 정형 데이터를 저장하고,
이를 추후 분석이나 처리를 위해 보관하는 저장소

데이터 엔지니어링이란 무엇인가

▶ Data Lake , Data Warehouse , Data Mart, Data Governance

Data Lake

다양한 형식과 소스에서 대규모의 비정형 및 정형 데이터를 저장하고,
이를 추후 분석이나 처리를 위해 보관하는 저장소

Data Warehouse

다양한 부서 및 사용자들이 의사결정에 활용할 수 있는
정형화된 데이터를 중앙에서 관리하고 저장하는 데이터 저장소

데이터 엔지니어링이란 무엇인가

▶ Data Lake , Data Warehouse , Data Mart, Data Governance

Data Lake

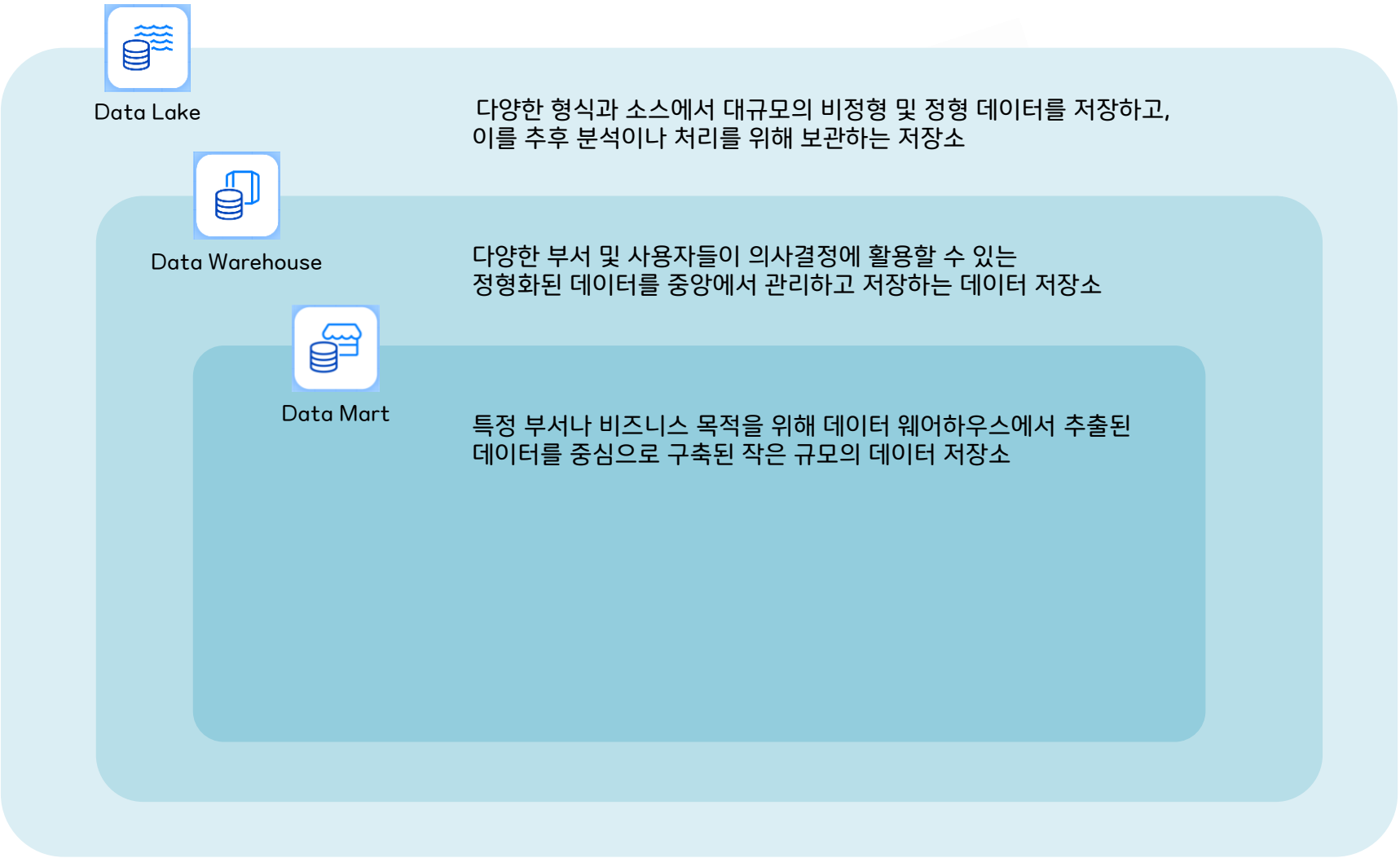
다양한 형식과 소스에서 대규모의 비정형 및 정형 데이터를 저장하고,
이를 추후 분석이나 처리를 위해 보관하는 저장소

Data Warehouse

다양한 부서 및 사용자들이 의사결정에 활용할 수 있는
정형화된 데이터를 중앙에서 관리하고 저장하는 데이터 저장소

데이터 엔지니어링이란 무엇인가

▶ Data Lake , Data Warehouse , Data Mart, Data Governance



데이터 엔지니어링이란 무엇인가

▶ Data Lake , Data Warehouse , Data Mart, Data Governance

Data Lake

다양한 형식과 소스에서 대규모의 비정형 및 정형 데이터를 저장하고,
이를 추후 분석이나 처리를 위해 보관하는 저장소

Data Warehouse

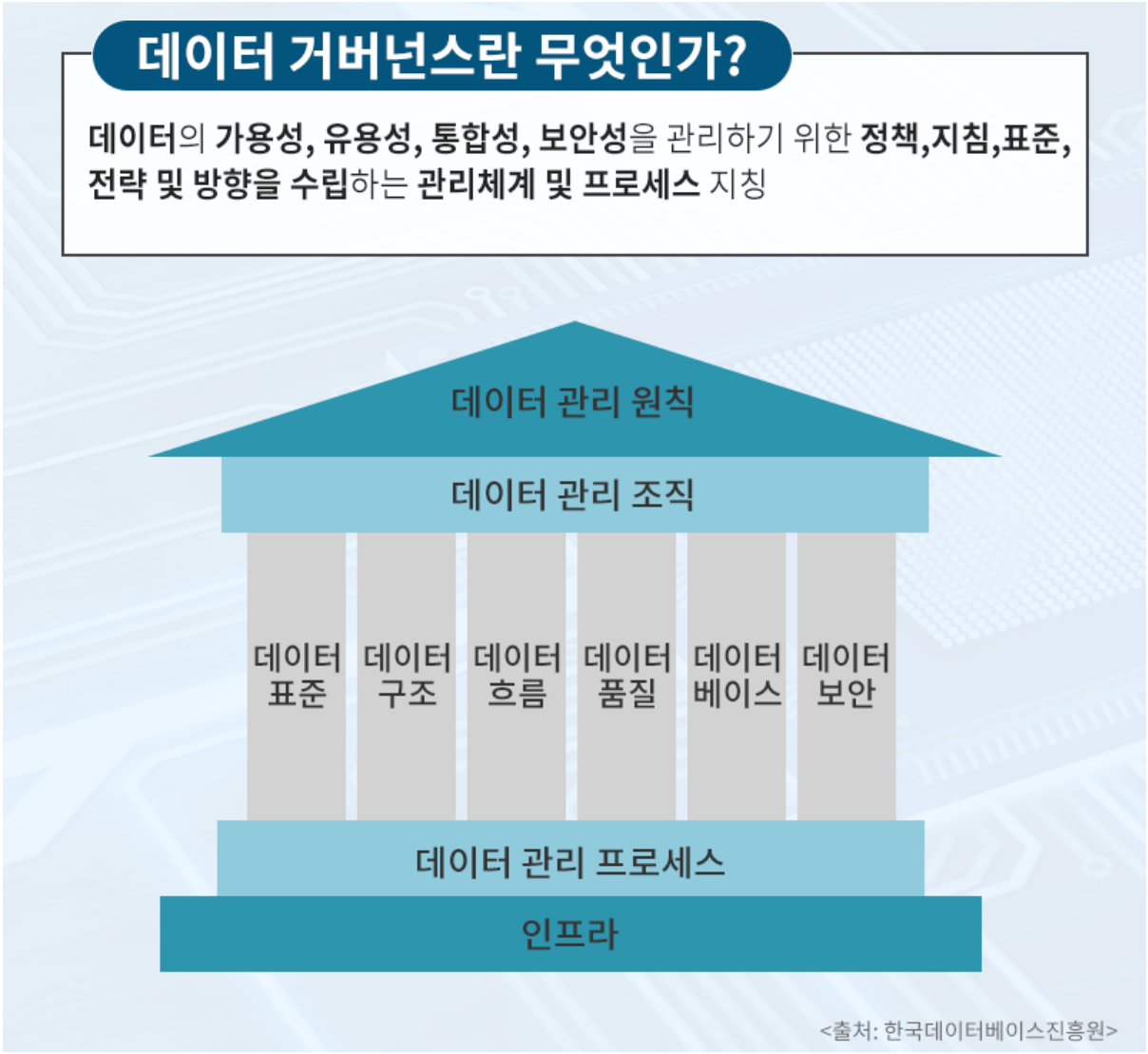
다양한 부서 및 사용자들이 의사결정에 활용할 수 있는
정형화된 데이터를 중앙에서 관리하고 저장하는 데이터 저장소

Data Mart

특정 부서나 비즈니스 목적을 위해 데이터 웨어하우스에서 추출된
데이터를 중심으로 구축된 작은 규모의 데이터 저장소

데이터 엔지니어링이란 무엇인가

▶ Data Lake , Data Warehouse , Data Mart, **Data Governance**



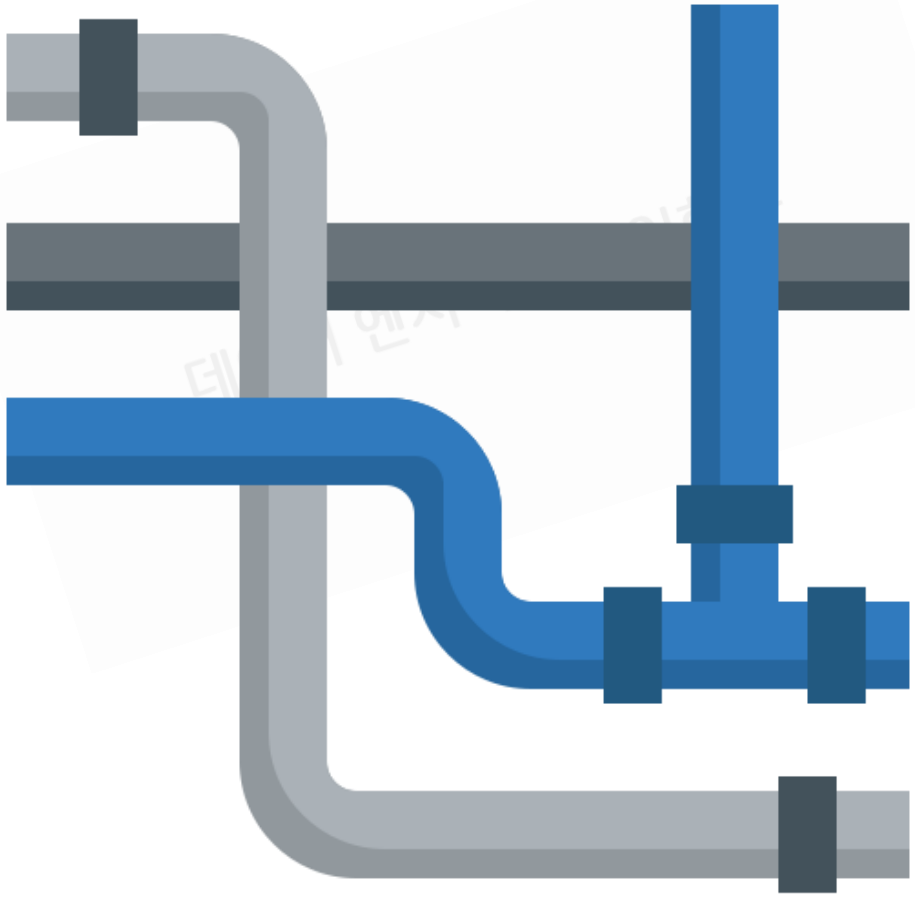
Data Pipeline

- 데이터의 원천부터 시작하여 필요한 데이터를 추출하고, 그 데이터를 정제, 변환, 분석, 저장, 전달하는 일련의 과정

데이터 엔지니어링 이현수

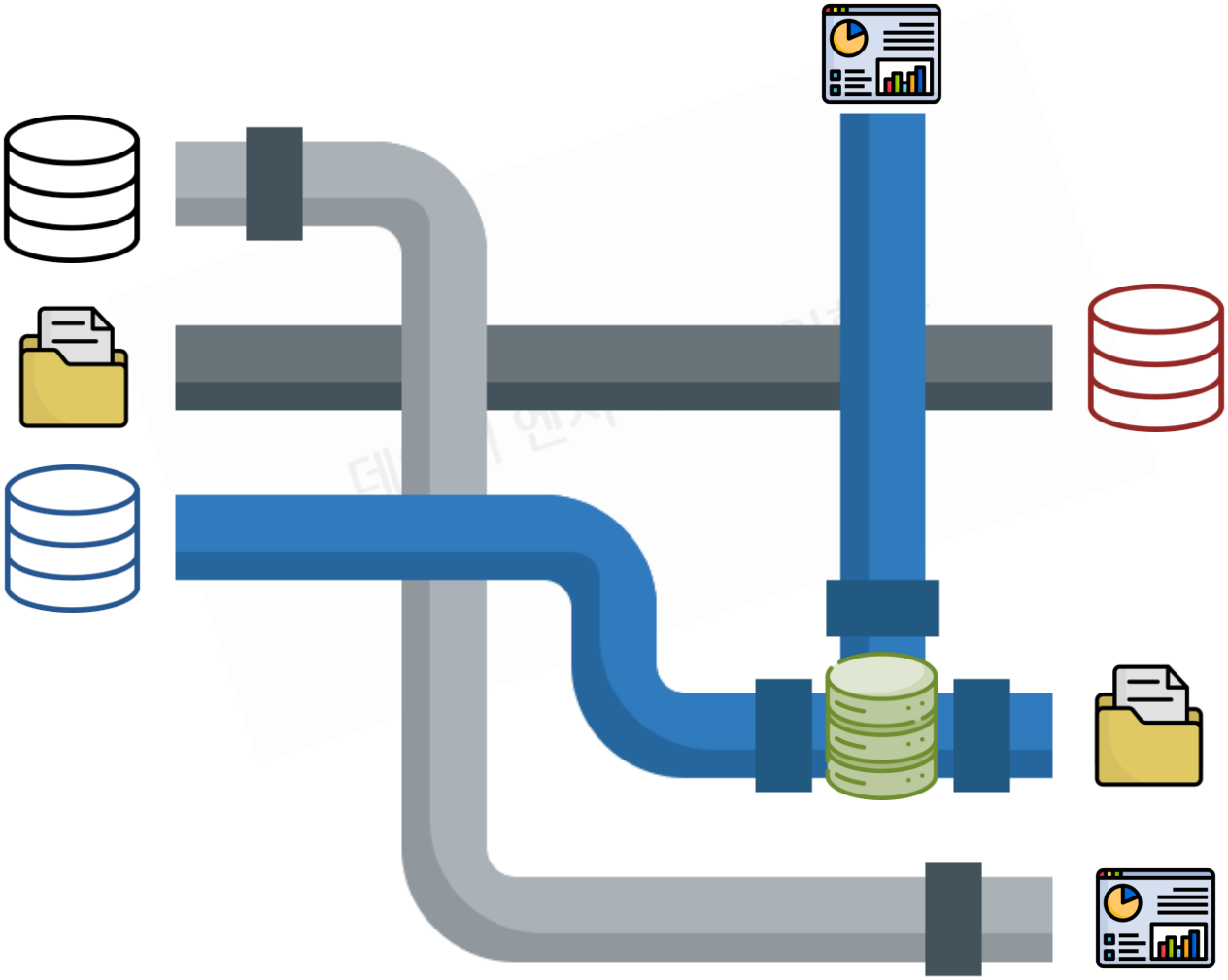
Data Pipeline

- 데이터의 원천부터 시작하여 필요한 데이터를 추출하고, 그 데이터를 정제, 변환, 분석, 저장, 전달하는 일련의 과정



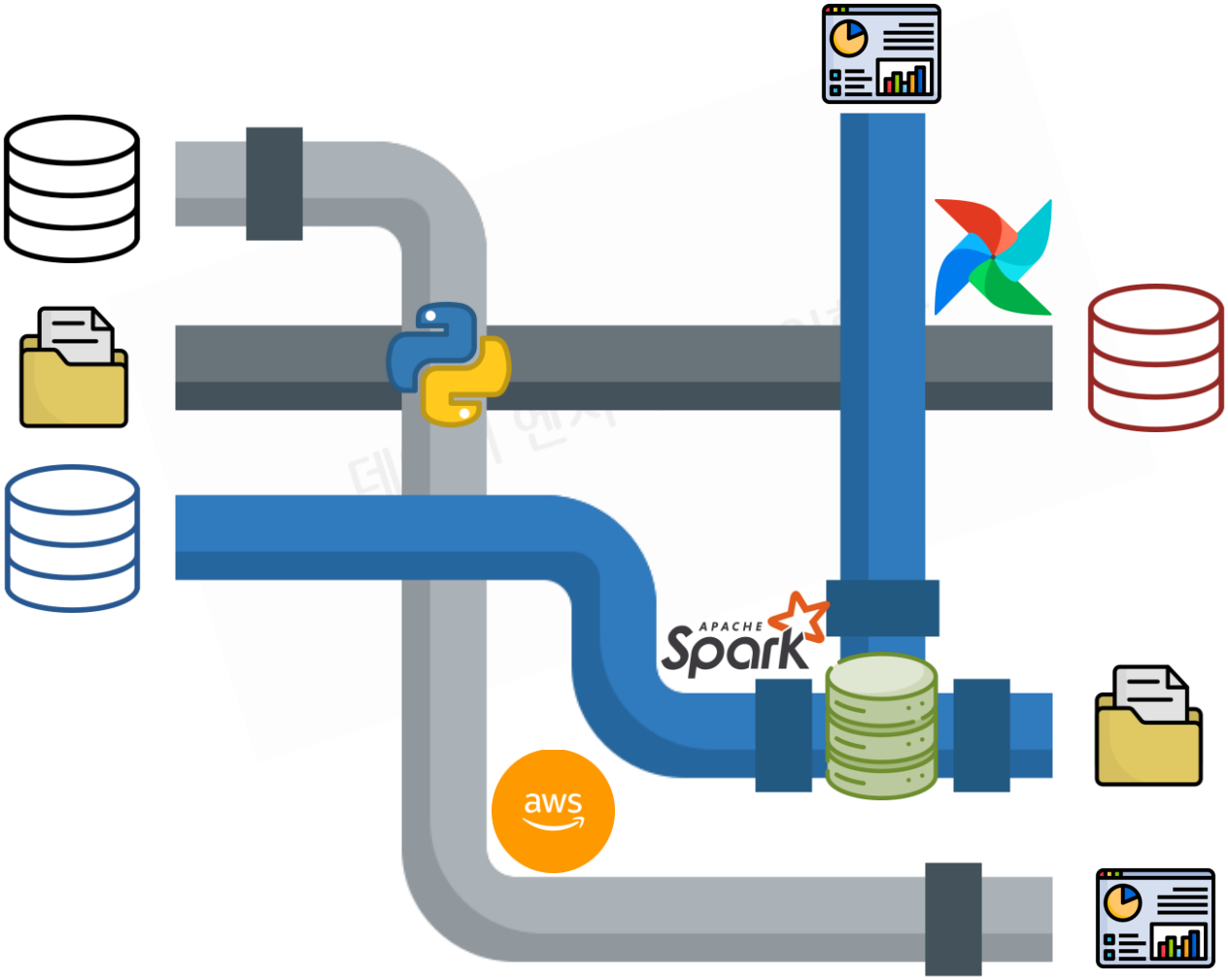
Data Pipeline

- 데이터의 원천부터 시작하여 필요한 데이터를 추출하고, 그 데이터를 정제, 변환, 분석, 저장, 전달하는 일련의 과정



Data Pipeline

- 데이터의 원천부터 시작하여 필요한 데이터를 추출하고, 그 데이터를 정제, 변환, 분석, 저장, 전달하는 일련의 과정



Data vs Metadata

데이터 엔지니어링 이현수

Data vs Metadata

데이터 엔지니어링 이현수

Data vs Metadata

#데이터
#정보의 실체
#사진

데이터 엔지니어링 이현수

▶ Data vs Metadata

Metaata

#데이터
#정보의 실체
#사진

#메타데이터
#데이터의 정보
#사진 정보

Data vs Metadata

f_studytree_id	f_company_id	f_gubun	f_studytree_nm	f_eduprocess_cd	f_area_cd	f_emh_cd	f_goal	f_studymethod	f_ord	f_deleteyn	yyyy	mm
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	6
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	7
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	6
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	6
4	0	01	듣기/말하기/쓰기	01	KO	E0	None	None	4	Y	2021	7
5	0	01	수와 연산	01	MA	E0	None	None	1	Y	2021	7
6	0	01	도형	01	MA	E0	None	None	2	Y	2021	7
7	0	01	측정	01	MA	E0	None	None	3	Y	2021	7

Data vs Metadata

Data

f_studytree_id	f_company_id	f_gubun	f_studytree_nm	f_eduprocess_cd	f_area_cd	f_emh_cd	f_goal	f_studymethod	f_ord	f_deleteyn	yyyy	mm
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	6
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	7
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	6
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	6
4	0	01	듣기/말하기/쓰기	01	KO	E0	None	None	4	Y	2021	7
5	0	01	수와 연산	01	MA	E0	None	None	1	Y	2021	7
6	0	01	도형	01	MA	E0	None	None	2	Y	2021	7
7	0	01	측정	01	MA	E0	None	None	3	Y	2021	7

▶ Data vs Metadata

Data

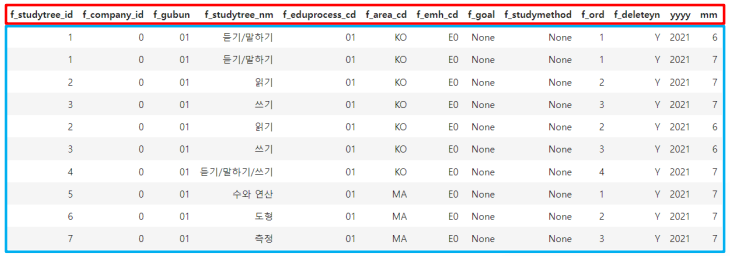
f_studytree_id	f_company_id	f_gubun	f_studytree_nm	f_eduprocess_cd	f_area_cd	f_emh_cd	f_goal	f_studymethod	f_ord	f_deleteyn	yyyy	mm
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	6
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	7
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	6
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	6
4	0	01	듣기/말하기/쓰기	01	KO	E0	None	None	4	Y	2021	7
5	0	01	수와 연산	01	MA	E0	None	None	1	Y	2021	7
6	0	01	도형	01	MA	E0	None	None	2	Y	2021	7
7	0	01	측정	01	MA	E0	None	None	3	Y	2021	7

Metaata

데이터베이스명			ISHERPA_Edubase2	설명	도메인 정보			
테이블명			t_studytree					
			도메인					
	PK	FK	컬럼명	타입	크기	NULL	초기값	설명
1	O		f_studytree_id	int	4	NOT NULL		도메인코드
2			f_company_id	int	4	NULL		회사코드
3			f_gubun	char	2	NULL		도메인구분
4			f_studytree_nm	nvarchar	100	NULL		도메인명
5			f_eduprocess_cd	char	2	NULL		교육과정
6			f_area_cd	char	2	NULL		영역
7			f_emh_cd	char	2	NULL		초중고
8			f_goal	nvarchar	1600	NULL		학습목표
9			f_studymethod	nvarchar	1600	NULL		학습방법
10			f_ord	smallint	2	NULL		순서
11			f_deleteyn	char	1	NULL		삭제여부

데이터 엔지니어링이란 무엇인가

▶ Data vs Metadata



f_studytree_id	f_company_id	f_gubun	f_studytree_nm	f_eduprocess_cd	f_area_cd	f_emh_cd	f_goal	f_studymethod	f_ord	f_deleteyn	yyyy	mm
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	6
1	0	01	듣기/말하기	01	KO	E0	None	None	1	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	7
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	7
2	0	01	읽기	01	KO	E0	None	None	2	Y	2021	6
3	0	01	쓰기	01	KO	E0	None	None	3	Y	2021	6
4	0	01	듣기/말하기/쓰기	01	KO	E0	None	None	4	Y	2021	7
5	0	01	수와 연산	01	MA	E0	None	None	1	Y	2021	7
6	0	01	도형	01	MA	E0	None	None	2	Y	2021	7
7	0	01	측정	01	MA	E0	None	None	3	Y	2021	7

데이터베이스명		ISHERPA_Edubase2		설명	도메인 정보			
테이블명		t_studytree						
		도메인						
	PK	FK	컬럼명	타입	크기	NULL	초기값	설명
1	O		f_studytree_id	int	4	NOT NULL		도메인코드
2			f_company_id	int	4	NULL		회사코드
3			f_gubun	char	2	NULL		도메인구분
4			f_studytree_nm	nvarchar	100	NULL		도메인명
5			f_eduprocess_cd	char	2	NULL		교육과정
6			f_area_cd	char	2	NULL		영역
7			f_emh_cd	char	2	NULL		초중고
8			f_goal	nvarchar	1600	NULL		학습목표
9			f_studymethod	nvarchar	1600	NULL		학습방법
10			f_ord	smallint	2	NULL		순서
11			f_deleteyn	char	1	NULL		삭제여부

구분	데이터 (Data)	메타데이터 (Metadata)
내용	정보의 본문	데이터에 대한 정보
형식	텍스트, 숫자, 이미지, 오디오, 비디오 등	특성, 속성, 구조, 형식 등
의미	정보의 실제 내용	데이터의 특성 및 속성 등
역할	분석, 처리, 저장 등에 사용	데이터 관리, 검색, 이해에 사용
예시	텍스트 문서, 사진, 동영상	파일 크기, 작성자, 생성일 등

▶ OLTP vs OLAP

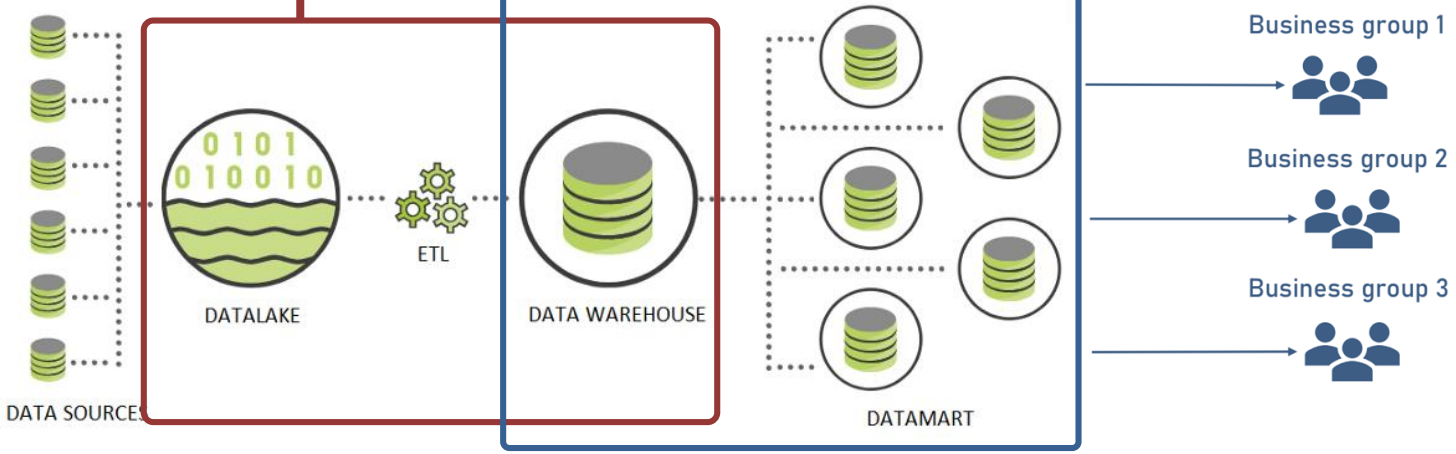
데이터 엔지니어링 이현수

▶ OLTP vs OLAP

구분	OLTP (Online Transaction Processing)	OLAP (Online Analytical Processing)
목적	데이터에 대한 수정 및 추가	데이터에 대한 쿼리
트랜잭션 형태	INSERT, UPDATE, DELETE	SELECT
프로세스 속도	수 초 이내	수 초 이상 수 분 이내
중요점	데이터 정확도, 무결성	결과의 속도, 표현 방식
활용자	데이터 엔지니어, 개발자	분석가, 의사결정자
예시	회원정보 수정, 사용자 로그 기록	1년 간의 주요 인기 트렌드

▶ OLTP vs OLAP

구분	OLTP (Online Transaction Processing)	OLAP (Online Analytical Processing)
목적	데이터에 대한 수정 및 추가	데이터에 대한 쿼리
트랜잭션 형태	INSERT, UPDATE, DELETE	SELECT
프로세스 속도	수 초 이내	수 초 이상 수 분 이내
중요점	데이터 정확도, 무결성	결과의 속도, 표현 방식
활용자	데이터 엔지니어, 개발자	분석가, 의사결정자
예시	회원정보 수정, 사용자 로그 기록	1년 간의 주요 인기 트렌드



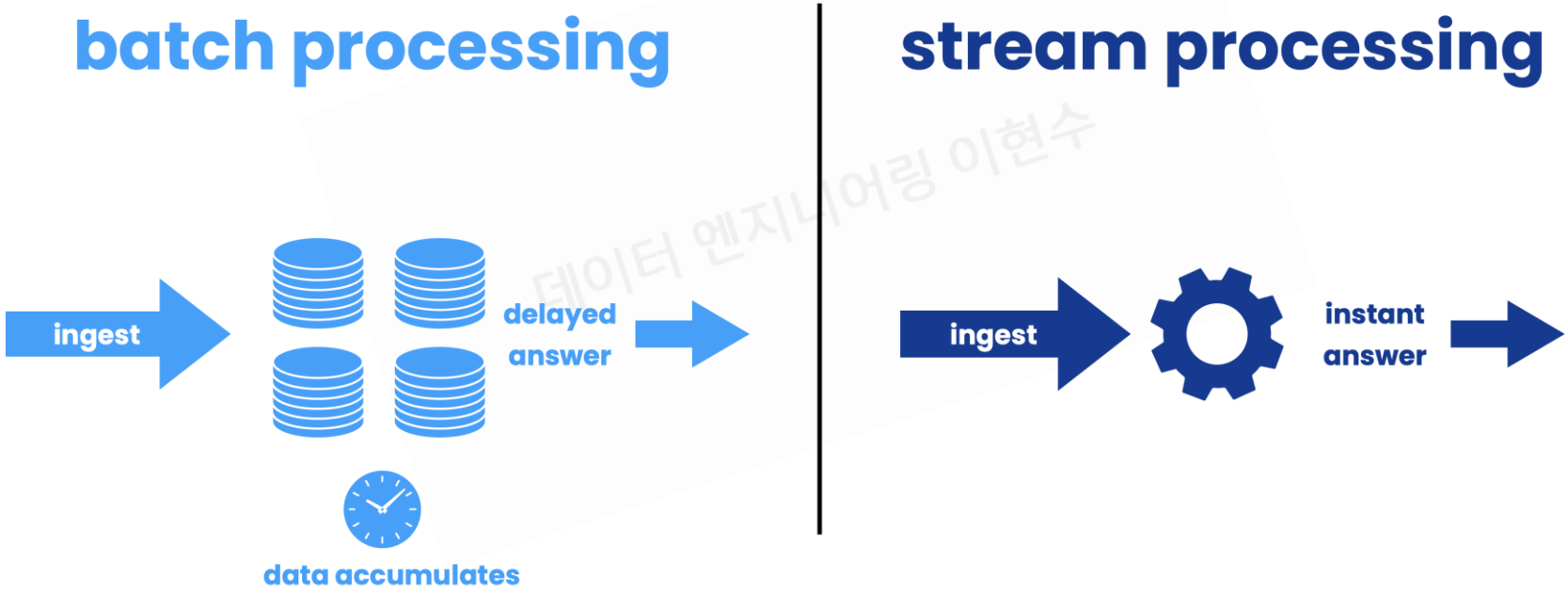
Batch vs Stream

데이터 엔지니어링 이현수

▶ Batch vs Stream

많은 양의 데이터를 정해진 시간에 일괄적으로 처리하는 것

실시간으로 들어오는 데이터를 계속 처리하는 것



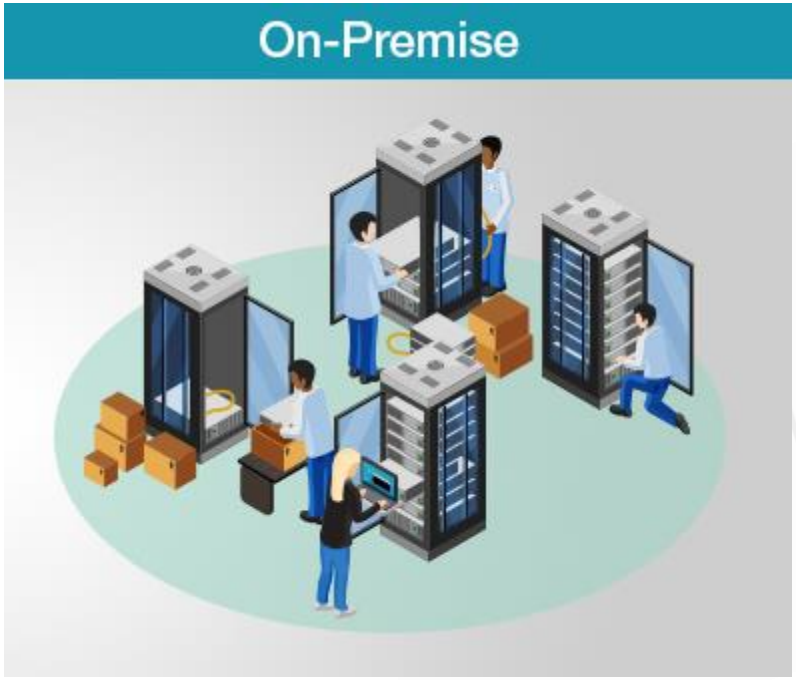
데이터 엔지니어링 이현수

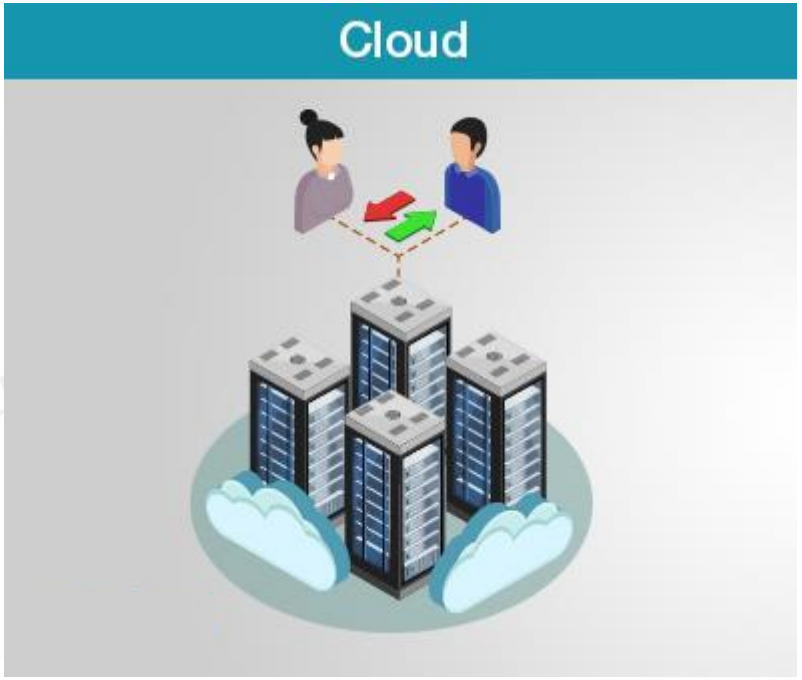
On-Premise vs Cloud

데이터 엔지니어링 이현수

데이터 엔지니어링이란 무엇인가

▶ On-Premise vs Cloud





OVHcloud
Tencent Cloud
Google Cloud
aws
ORACLE CLOUD Infrastructure
Alibaba Cloud
DigitalOcean
IBM Cloud
Akamai
Microsoft Azure

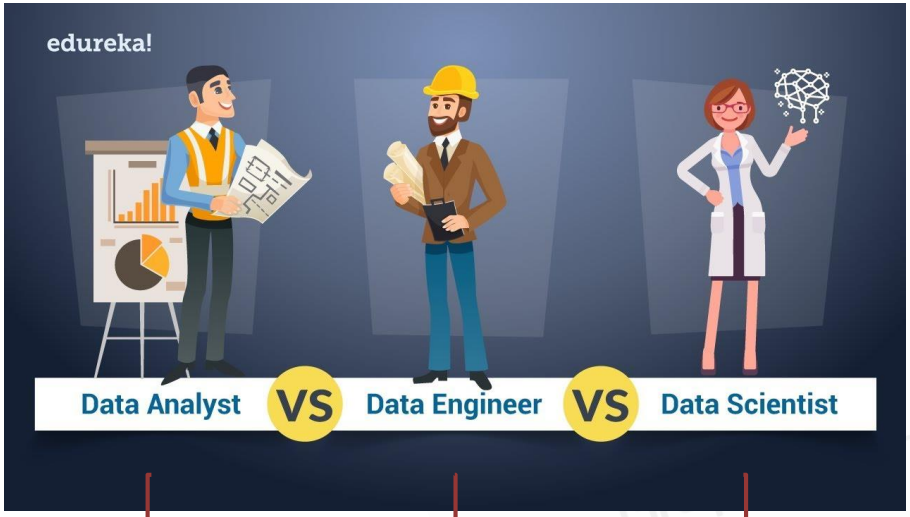
데이터 엔지니어링이란 무엇인가

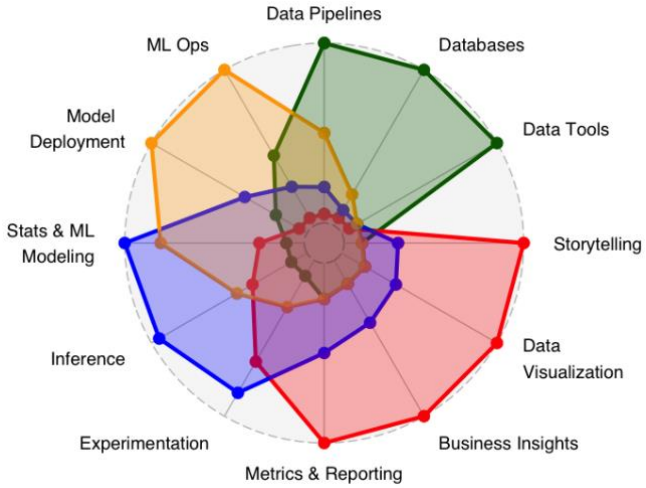
▶ 데이터 엔지니어 vs 데이터 분석가 vs 데이터 사이언티스트

데이터 엔지니어링 이현수

데이터 엔지니어링이란 무엇인가

▶ 데이터 엔지니어 vs 데이터 분석가 vs 데이터 사이언티스트





데이터를 분석하고 예측 모델을 개발하여 비즈니스 문제를 해결

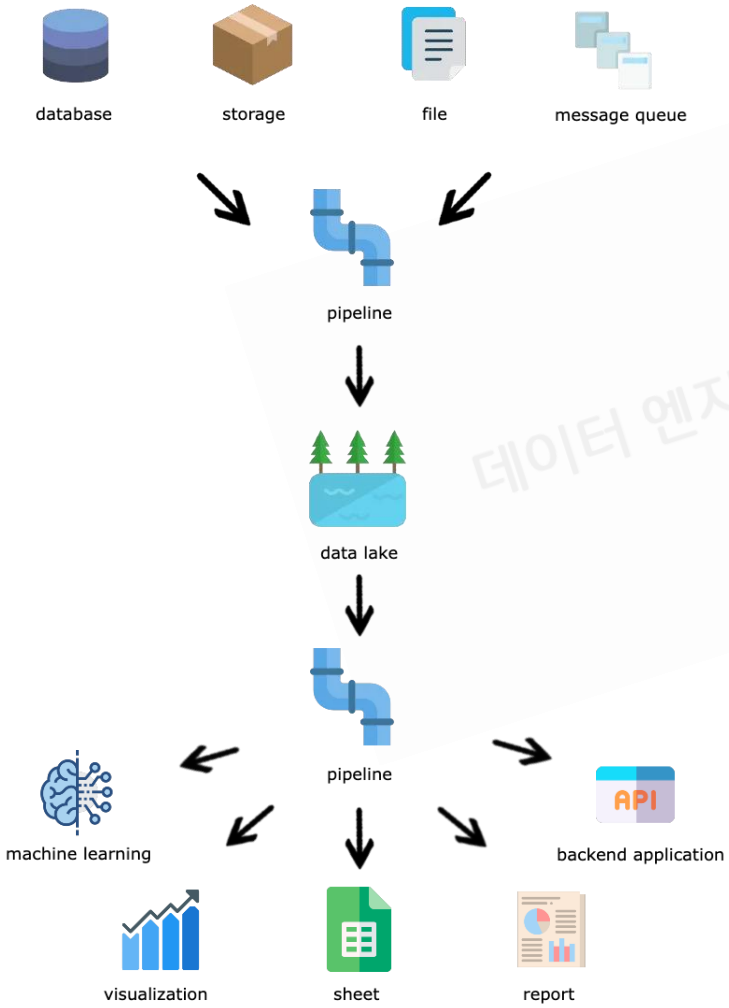
데이터 파이프라인 설계, 구축, 유지 및 관리

데이터를 쿼리하고 분석하여 보고서를 생성하거나 인사이트를 도출

데이터 엔지니어가 하는 일은?

Role & Responsibility

- 데이터가 효율적으로 흘러갈 수 있도록 파이프라인 설계/구축/운영/유지보수



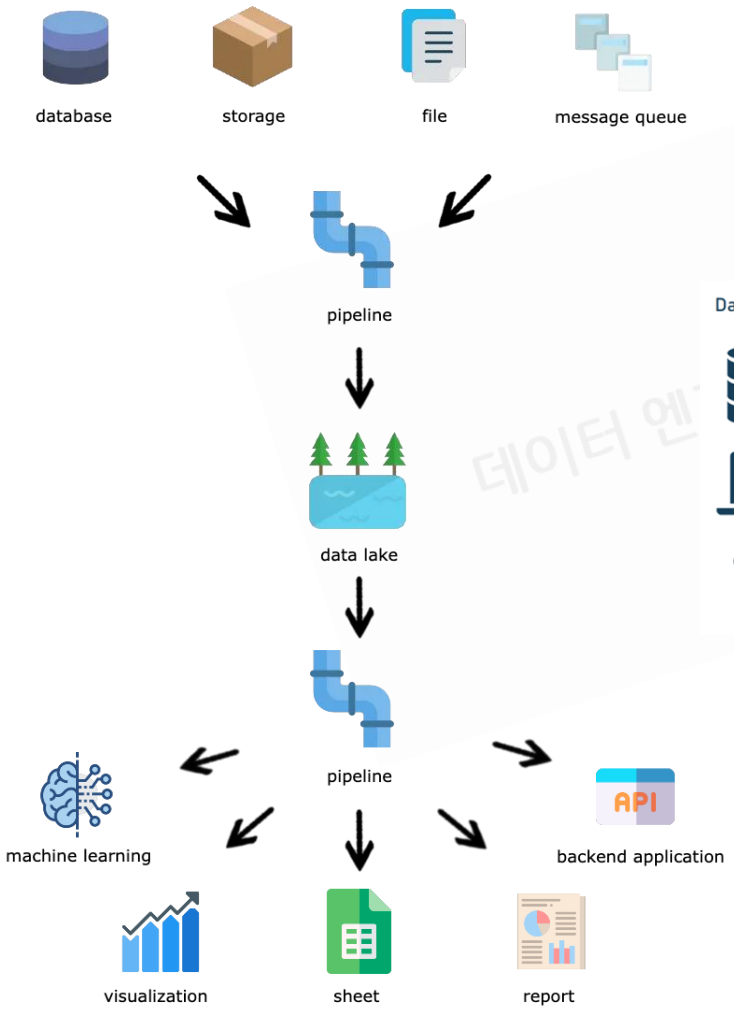
출처 : Socar Tech 블로그

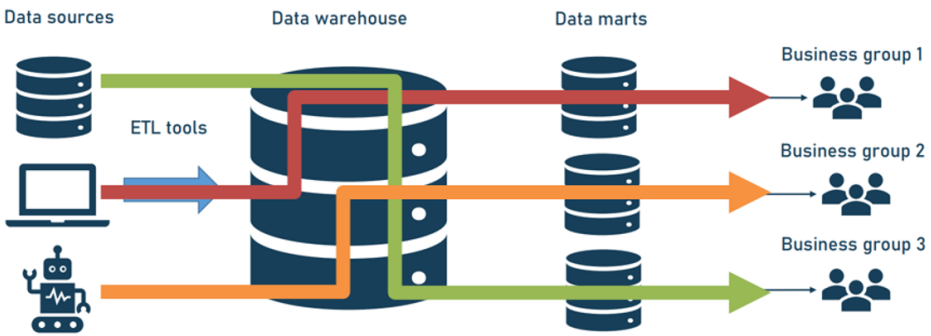
데이터 엔지니어링 이현수

데이터 엔지니어가 하는 일은?

Role & Responsibility

- 데이터가 효율적으로 흘러갈 수 있도록 파이프라인 설계/구축/운영/유지보수





출처 : Socar Tech 블로그

데이터 엔지니어가 하는 일은?

Role & Responsibility

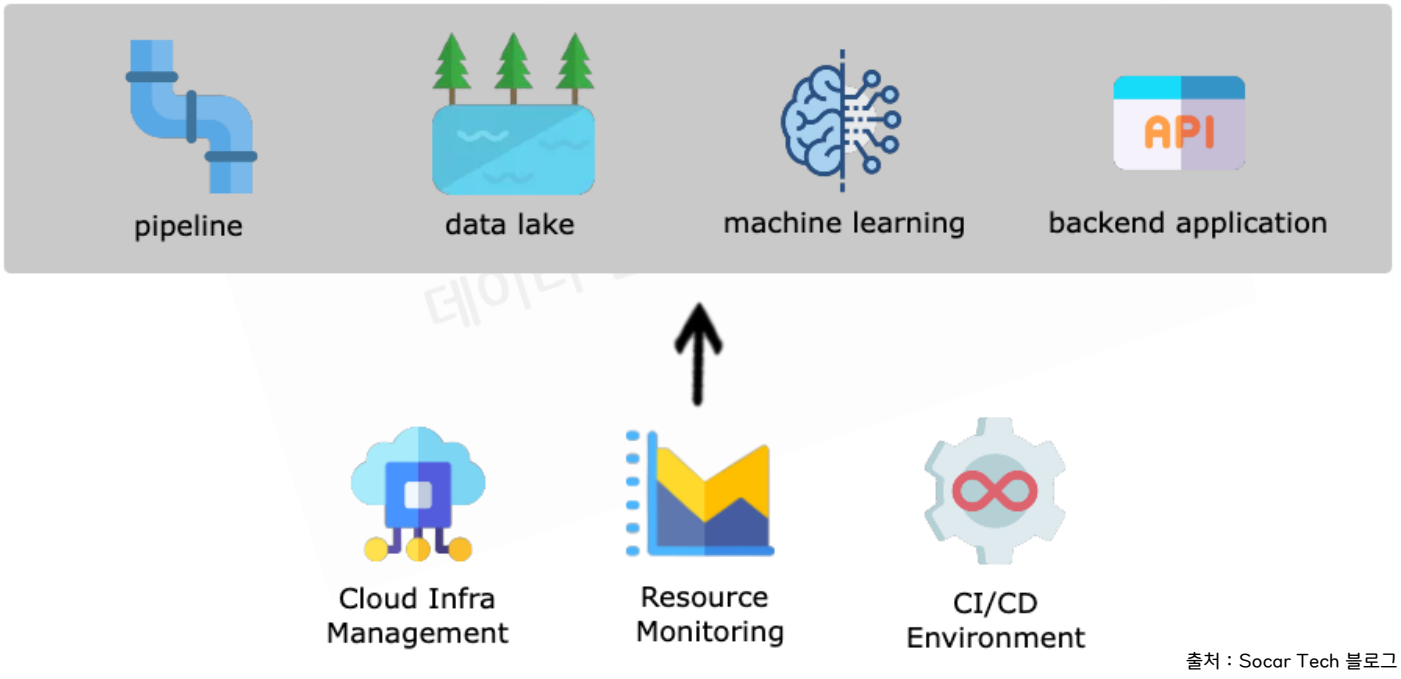
- 데이터가 효율적으로 흘러갈 수 있도록 파이프라인 설계/구축/운영/유지보수
- Endpoint 유저가 원하는 데이터 제공을 위한 ETL 작업 개발

출처 : Socar Tech 블로그

데이터 엔지니어가 하는 일은?

Role & Responsibility

- 데이터가 효율적으로 흘러갈 수 있도록 파이프라인 설계/구축/운영/유지보수
- Endpoint 유저가 원하는 데이터 제공을 위한 ETL 작업 개발
- 유입 or 유입된 데이터들에 대한 관리 및 모니터링

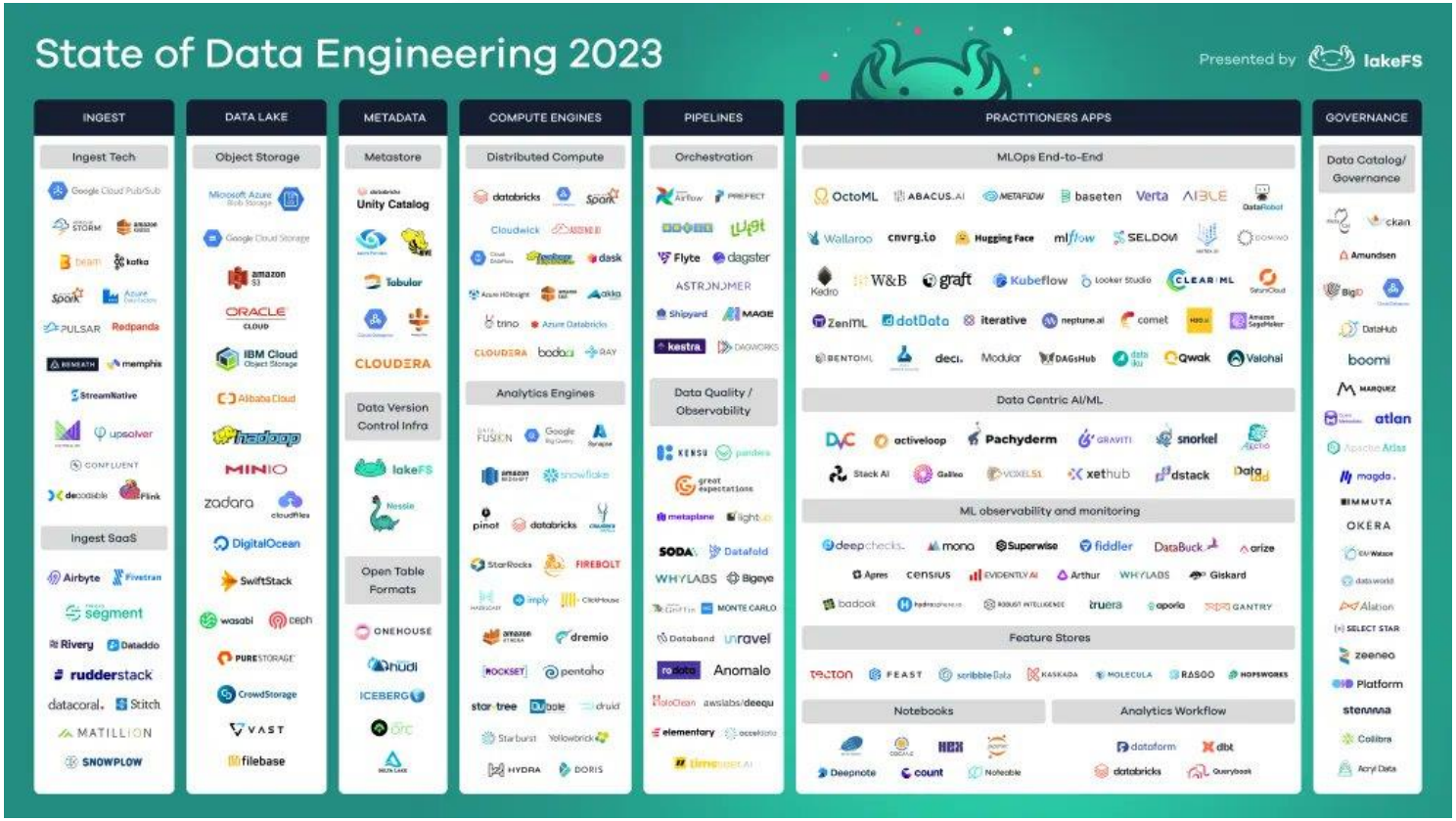


출처 : Socar Tech 블로그

데이터 엔지니어가 하는 일은?

Role & Responsibility

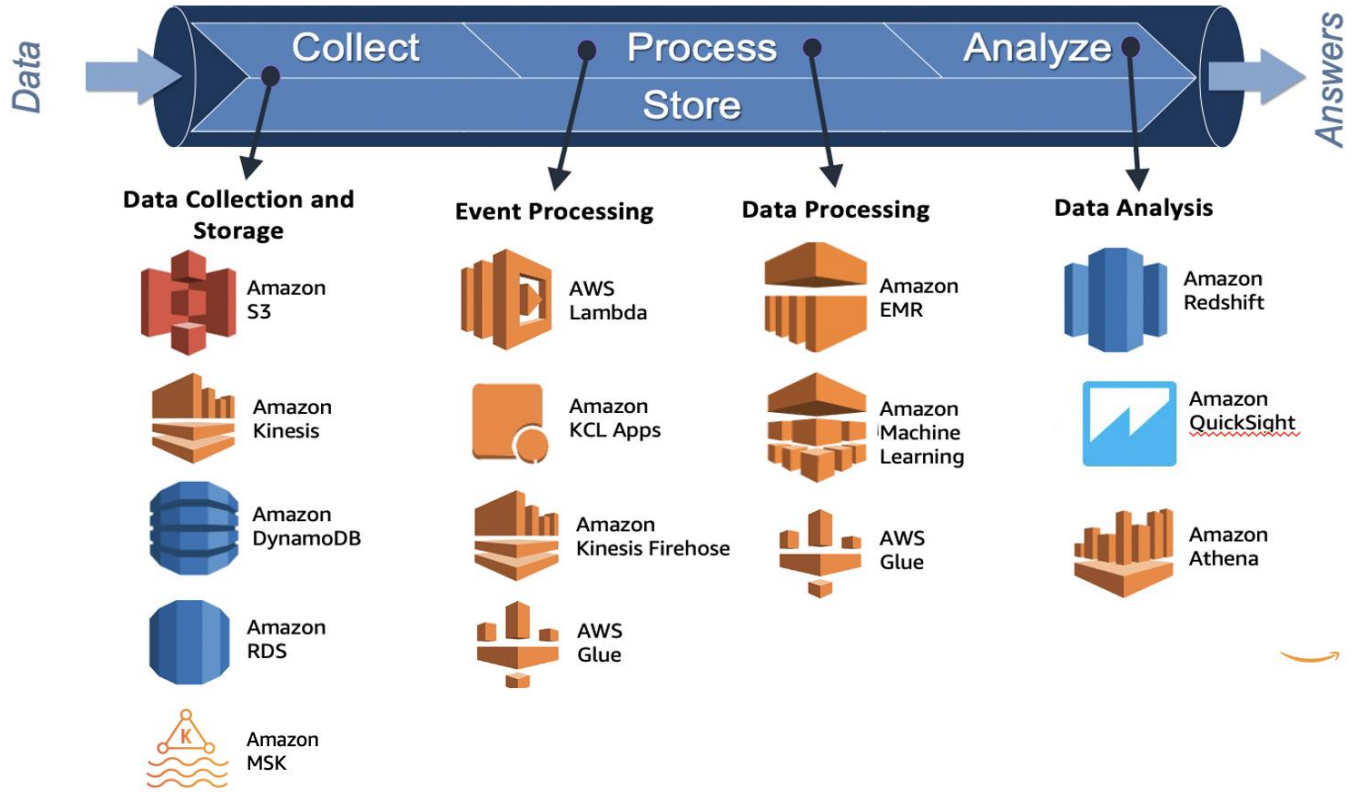
- 데이터가 효율적으로 흘러갈 수 있도록 파이프라인 설계/구축/운영/유지보수
- Endpoint 유저가 원하는 데이터 제공을 위한 ETL 작업 개발
- 유입 or 유입된 데이터들에 대한 관리 및 모니터링
- 여러가지 기술들 중 가장 적합한 Tool을 선택하여 파이프라인 아키텍처 구성



데이터 엔지니어가 하는 일은?

Role & Responsibility

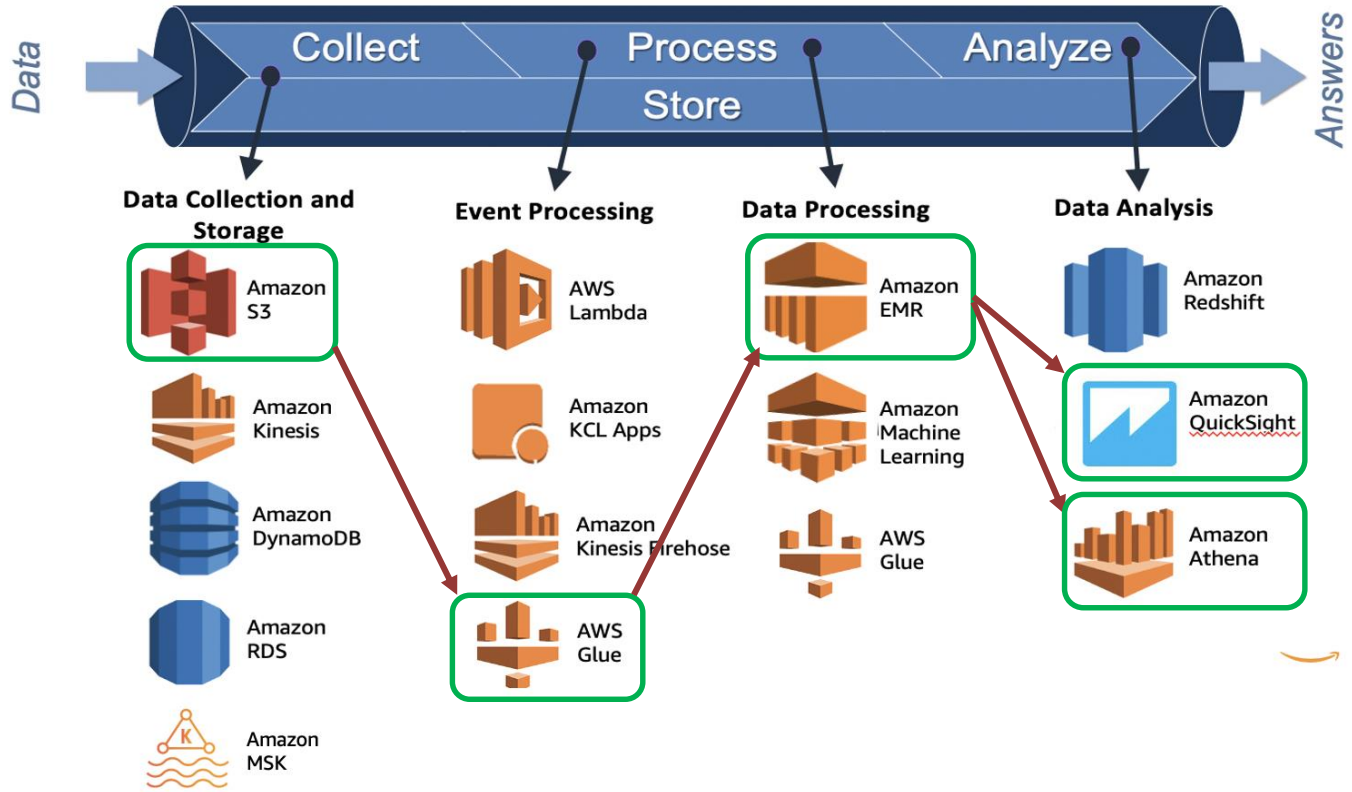
- 데이터가 효율적으로 흘러갈 수 있도록 파이프라인 설계/구축/운영/유지보수
- Endpoint 유저가 원하는 데이터 제공을 위한 ETL 작업 개발
- 유입 or 유입된 데이터들에 대한 관리 및 모니터링
- 여러가지 기술들 중 가장 적합한 Tool을 선택하여 파이프라인 아키텍처 구성



데이터 엔지니어가 하는 일은?

Role & Responsibility

- 데이터가 효율적으로 흘러갈 수 있도록 파이프라인 설계/구축/운영/유지보수
- Endpoint 유저가 원하는 데이터 제공을 위한 ETL 작업 개발
- 유입 or 유입된 데이터들에 대한 관리 및 모니터링
- 여러가지 기술들 중 가장 적합한 Tool을 선택하여 파이프라인 아키텍처 구성



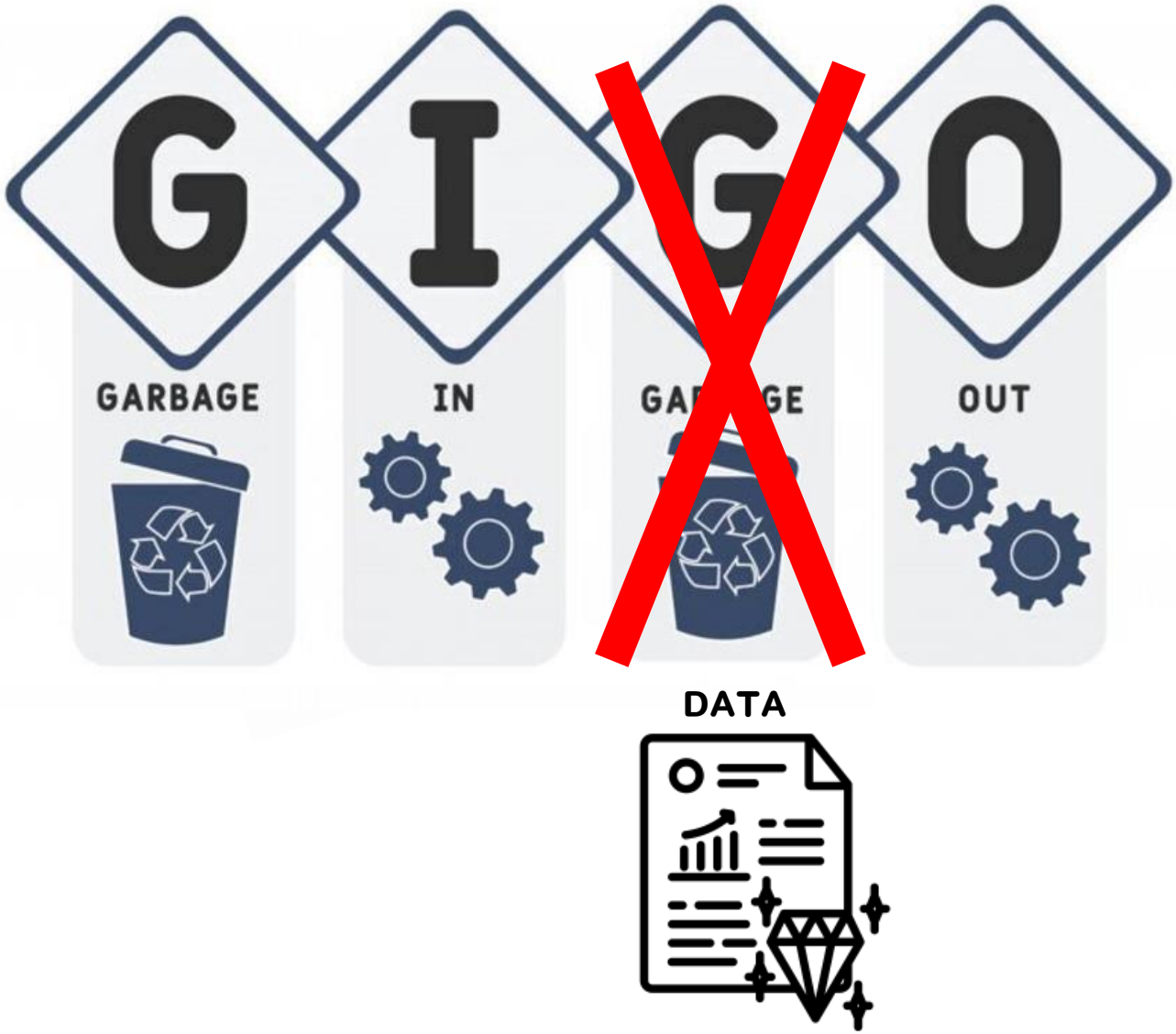
데이터 엔지니어링이 왜 필요할까?

▶ Garbage In Garbage Out

데이터 엔지니어링이 왜 필요할까?

▶ Garbage In ~~Garbage Out~~ Data Out

- 넘쳐나는 데이터들 사이에서 유용한 자료들을 선별하고 가공하여 유의미한 정보로 만드는 것!



데이터 엔지니어의 SKILL SET

Data Engineer Skill Set

초급


Python

SQL


EC2


Linux


RDB


Athena


Quicksight


Git

중급


Glue


EMR

Spark


Lambda


Airflow


Docker

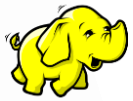
Shell Script


Gitlab

고급


Kubernetes


Kafka


Hadoop Ecosystem

MLOPS


Multi Cloud


Infra Management


Infra as Code

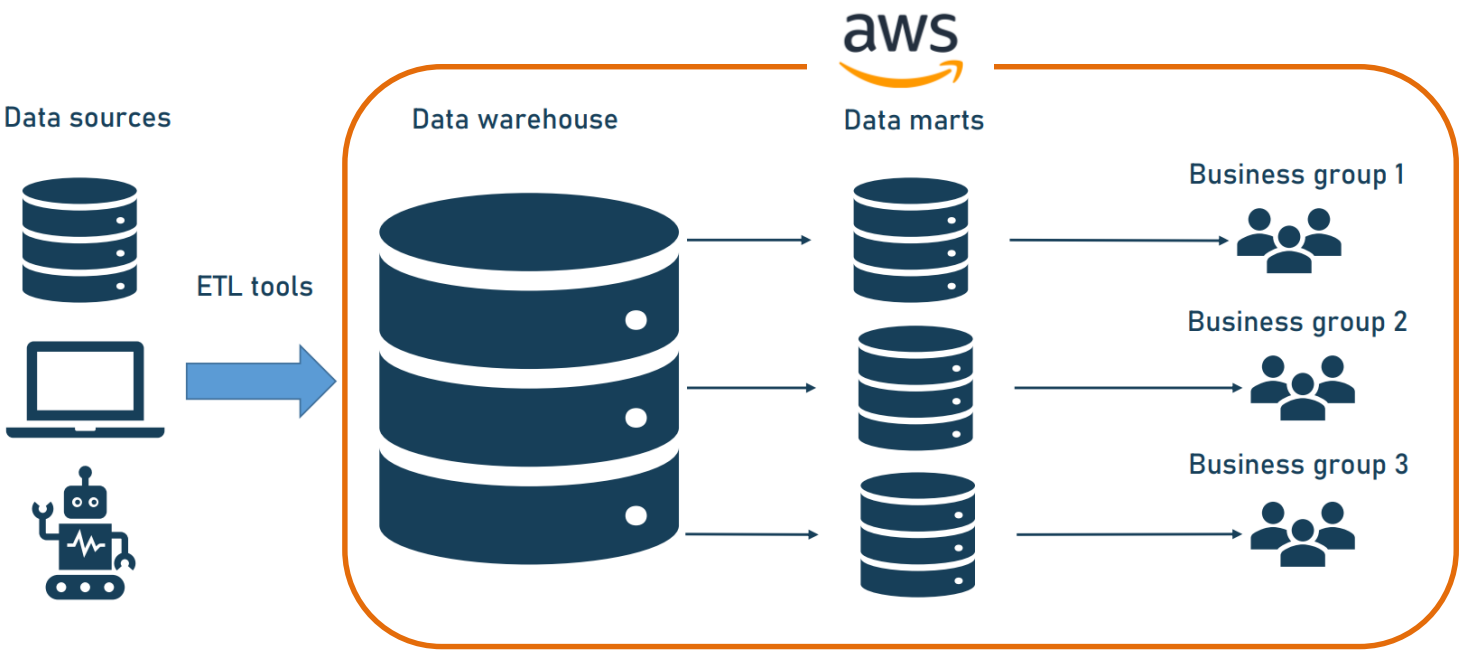
천재교육 실무에서의 데이터 엔지니어링

01 사용되는 Tech Stack 소개

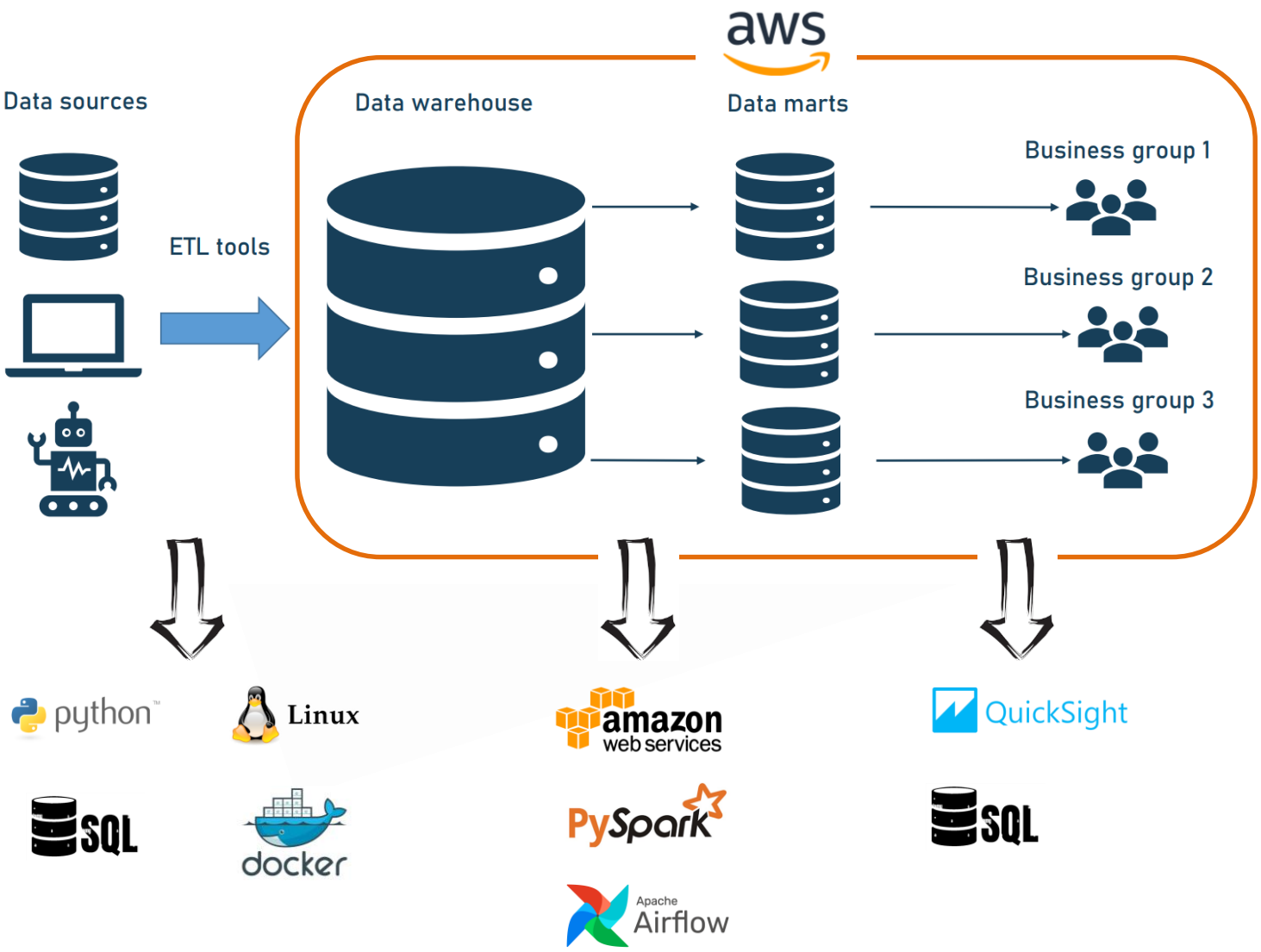
02 데이터 파이프라인 아키텍처

데이터 엔지니어링 이현수

사용되는 Tech Stack 소개

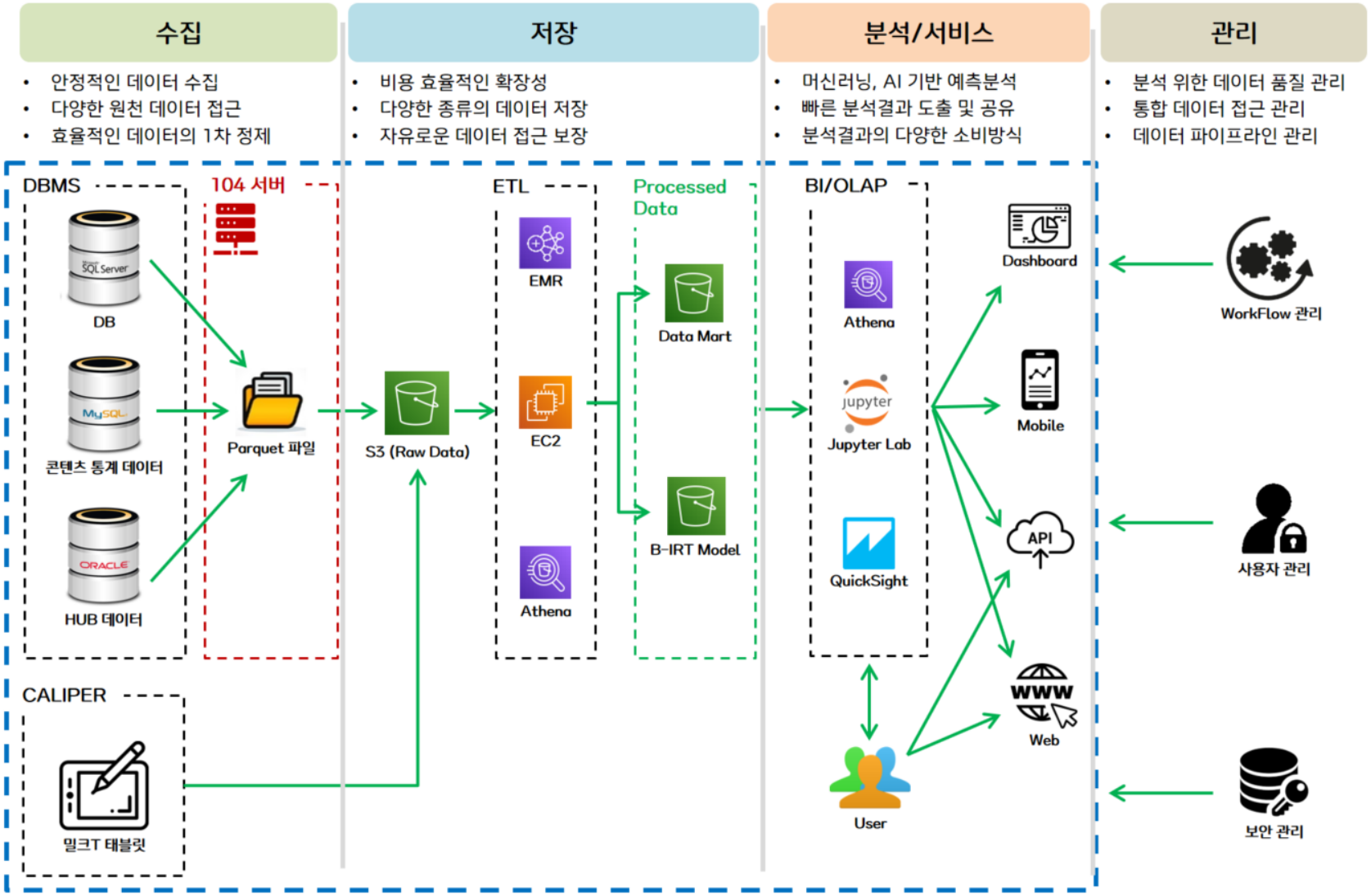


사용되는 Tech Stack 소개



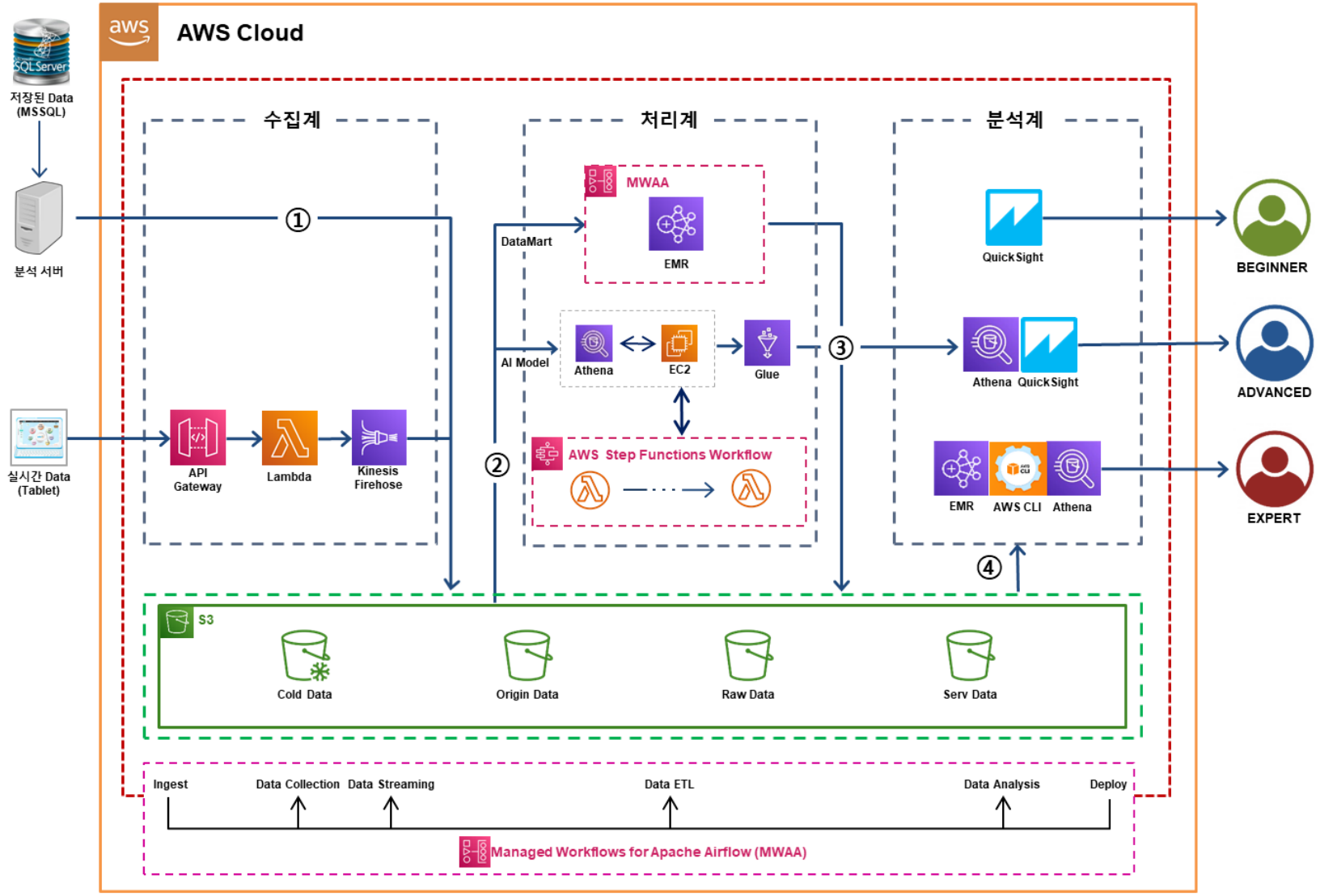
데이터 파이프라인 아키텍처

▶ 밀크T 데이터 분석 파이프라인 아키텍처



데이터 파이프라인 아키텍처

▶ 밀크T 데이터 분석 파이프라인 아키텍처



실습 범위 안내 및 기초 실습

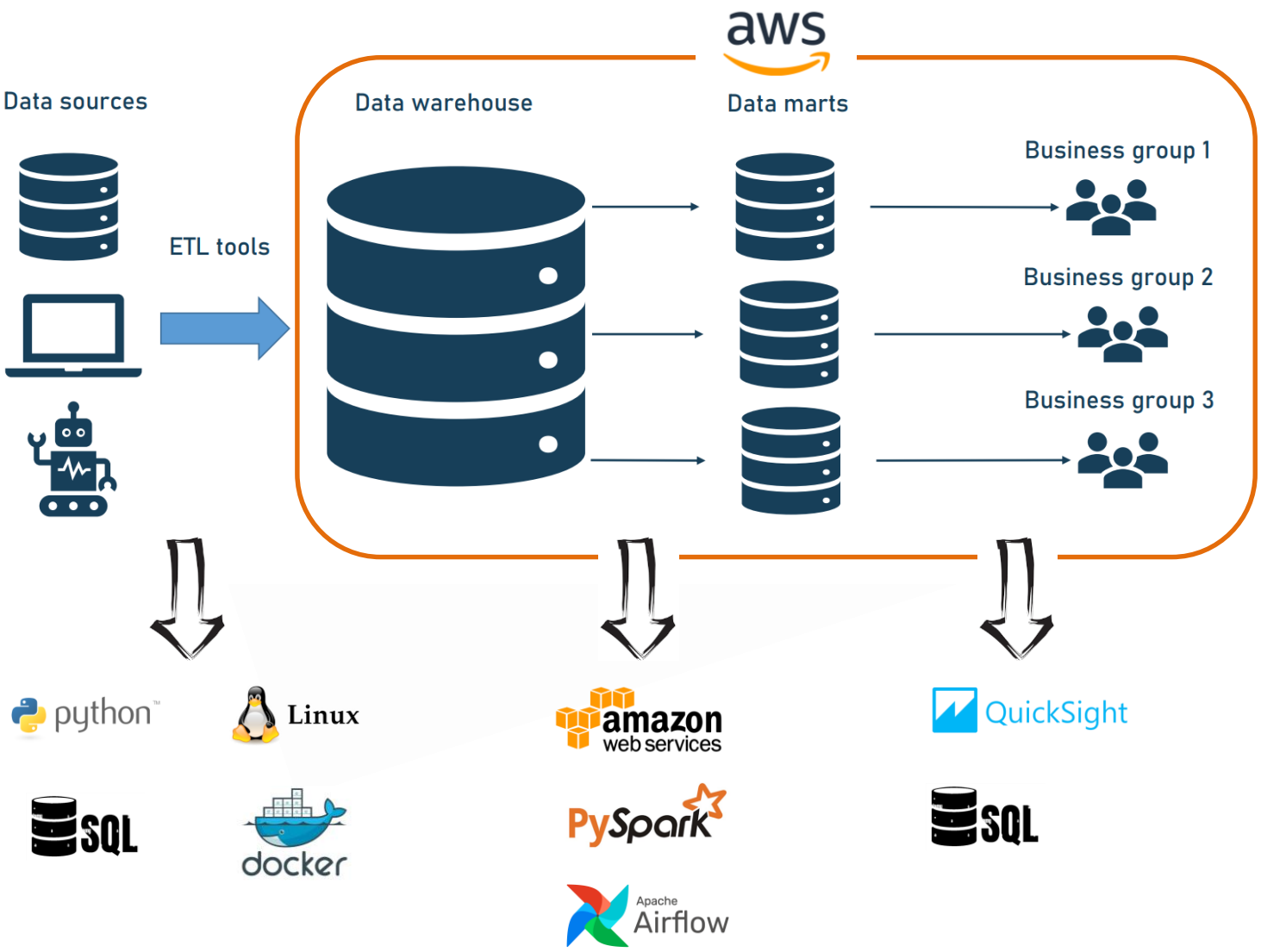
01 실습 범위 소개

02 실습 내용 흐름도

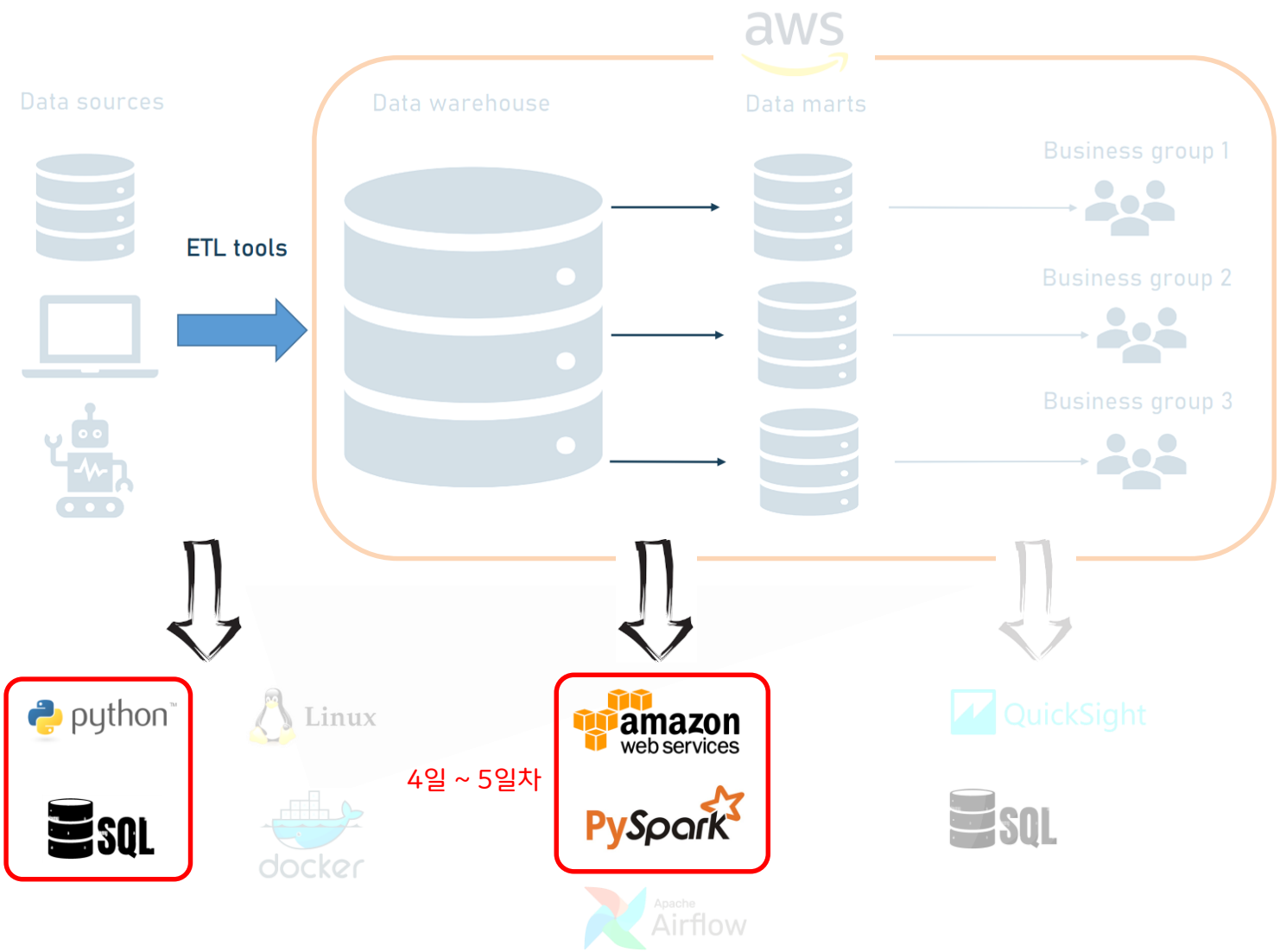
03 환경 구성 및 실습

데이터 엔지니어링 아카데미

실습 범위 소개

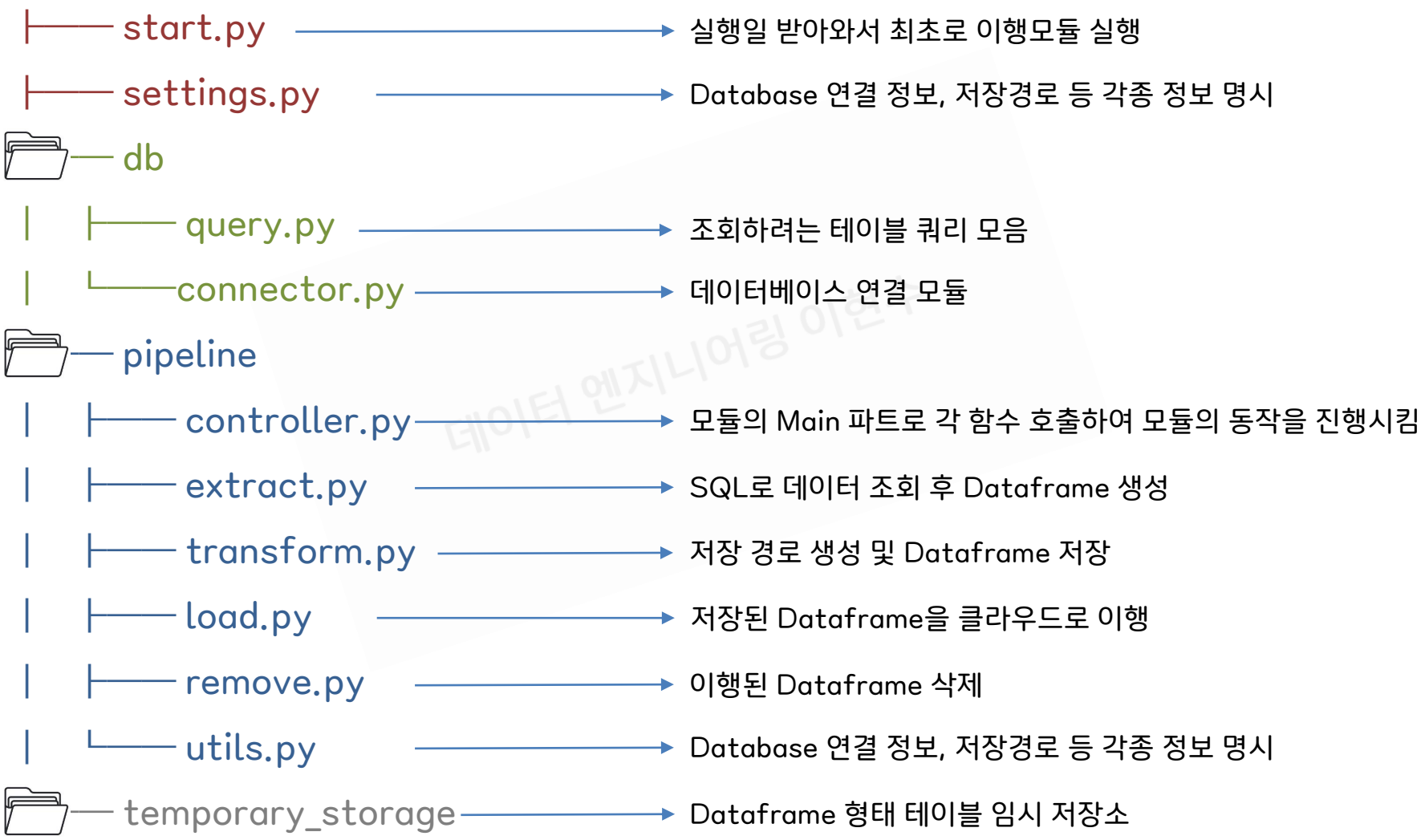

aws
Data sources
Data warehouse
Data marts
ETL tools
Business group 1
Business group 2
Business group 3
python
Linux
amazon web services
QuickSight
SQL
docker
PySpark
SQL
Apache Airflow

실습 범위 소개


aws
Data sources
Data warehouse
Data marts
ETL tools
Business group 1
Business group 2
Business group 3
python
SQL
Linux
docker
amazon web services
PySpark
Apache Airflow
QuickSight
SQL
1일 ~ 3일차
4일 ~ 5일차

실습 내용 흐름도

▶ 데이터 이행 파이프라인 모듈 구성



▶ 데이터 이행 파이프라인 모듈 구성

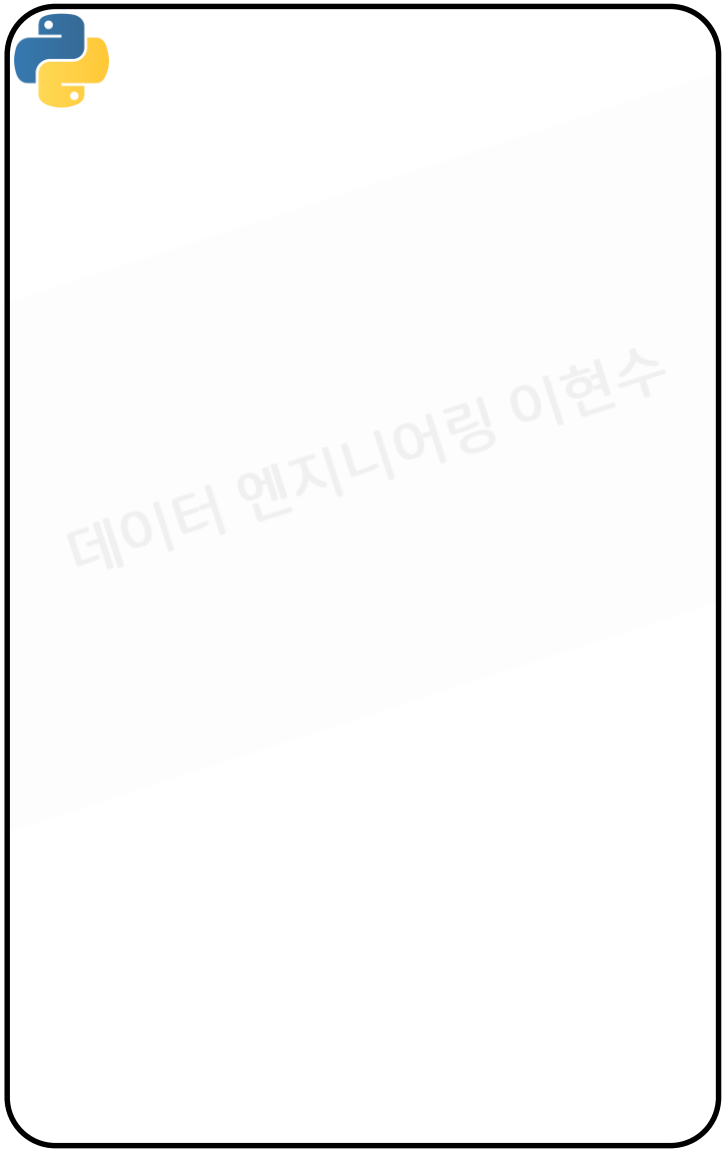
1. 객체 지향적 프로그래밍

2. 하나의 실행 파일은 하나의 기능만 담당

3. 기능과 연관 있는 직관적인 명명

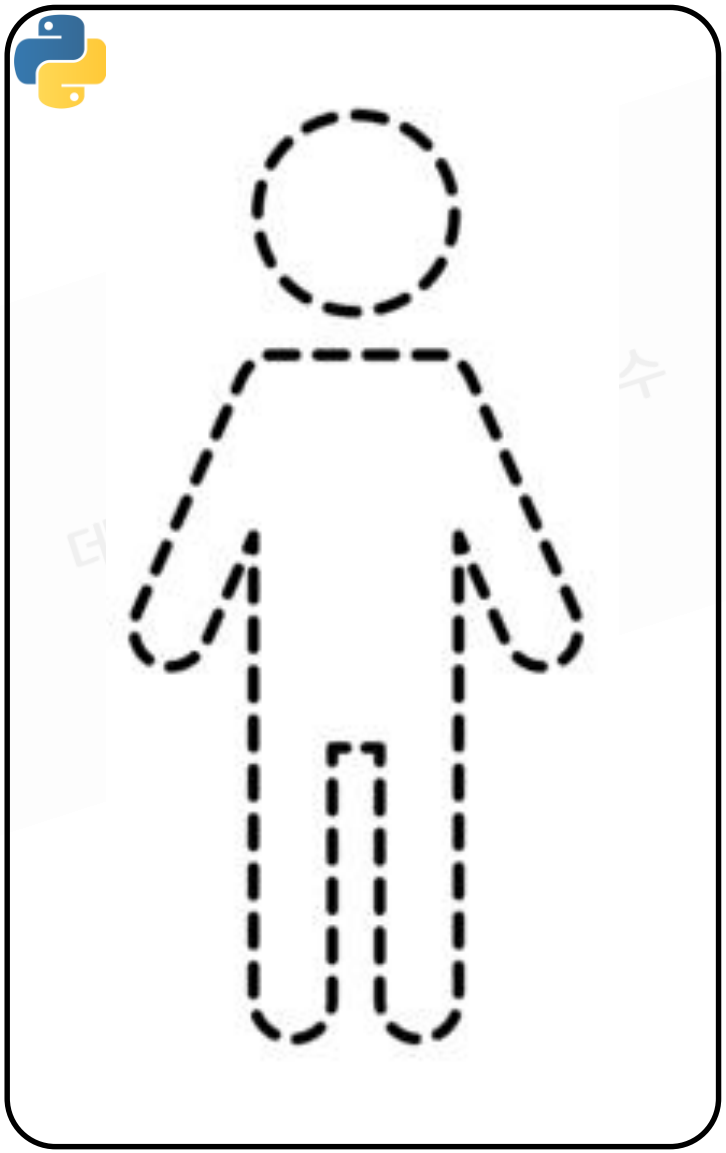
4. 관련성 있는 실행 파일은 같은 폴더 아래에

▶ 절차적 프로그래밍

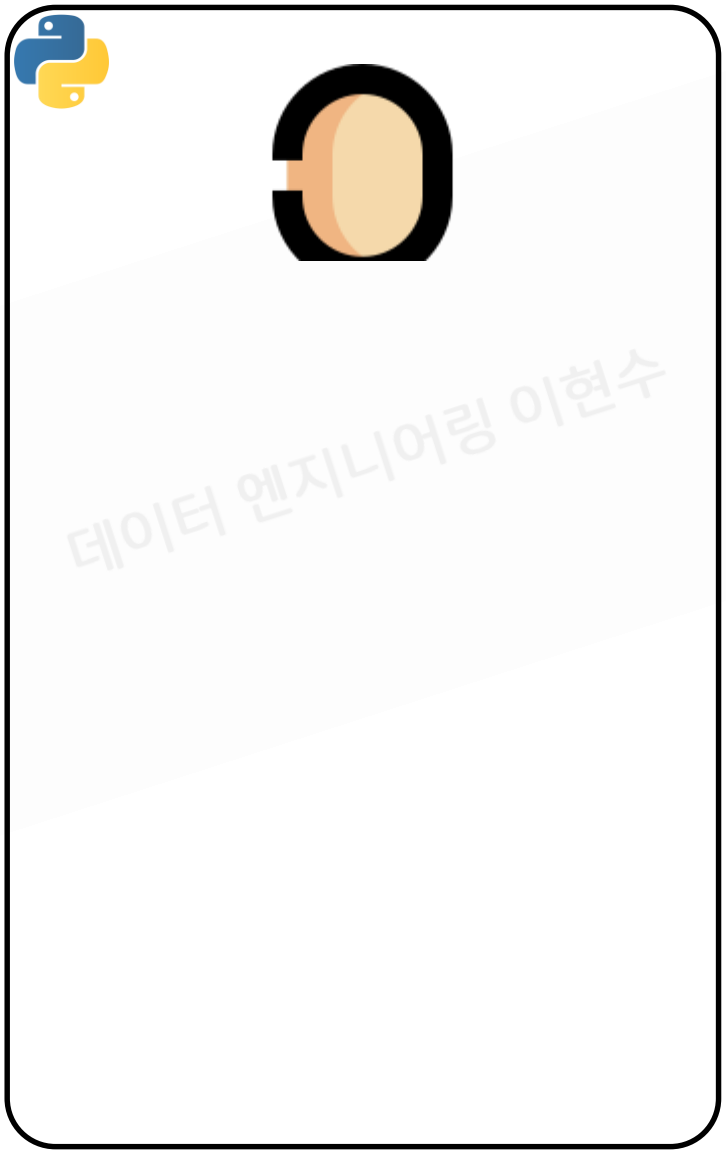


데이터 엔지니어링 이현수

▶ 절차적 프로그래밍

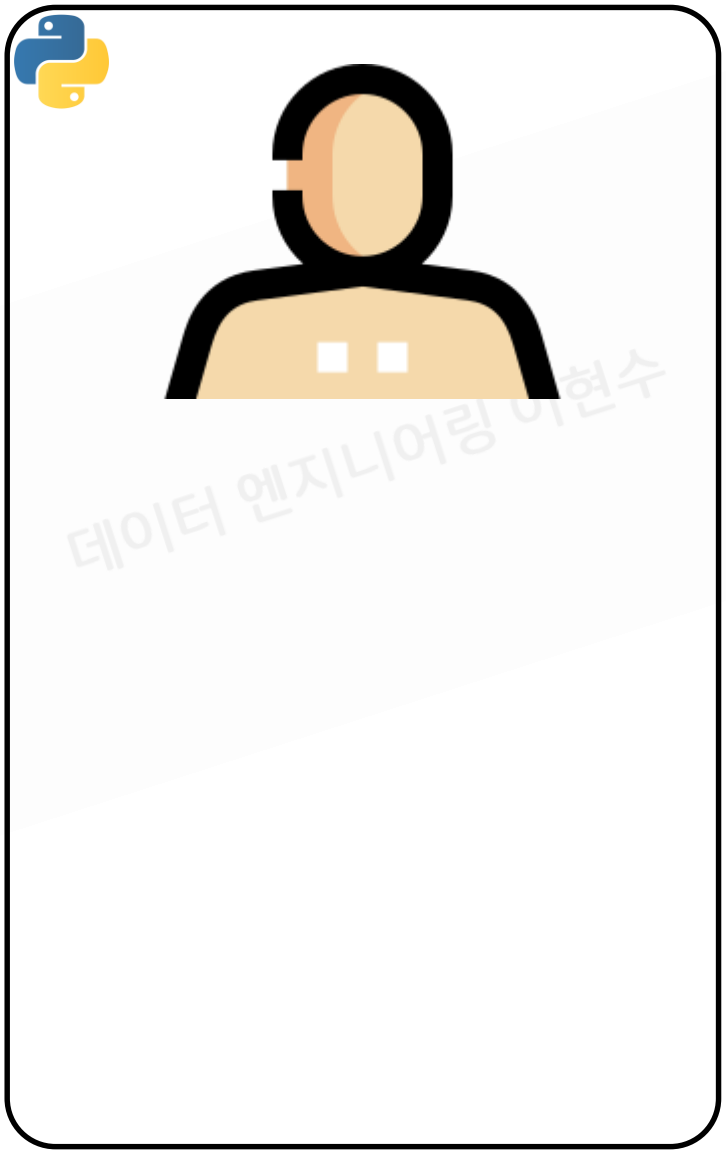


▶ 절차적 프로그래밍



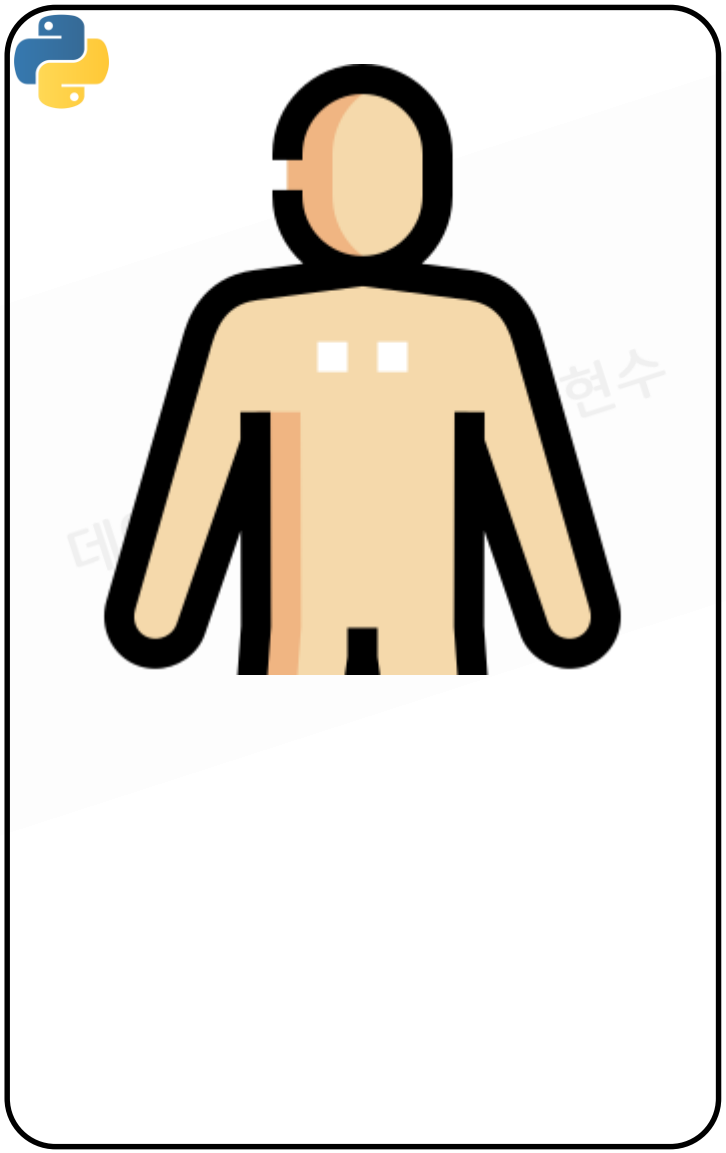
데이터 엔지니어링 이현수

▶ 절차적 프로그래밍

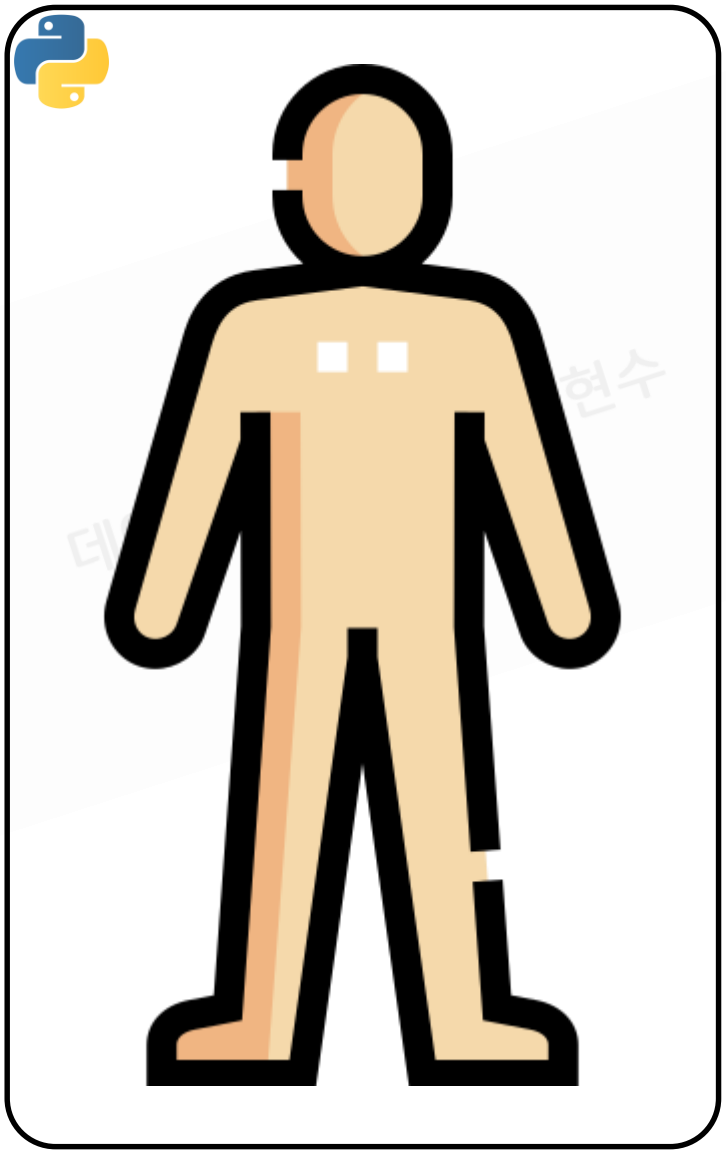


데이터 엔지니어링 이현수

▶ 절차적 프로그래밍



▶ 절차적 프로그래밍

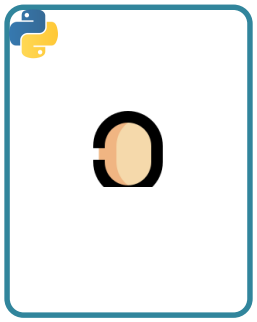


▶ 객체 지향 프로그래밍

데이터 엔지니어링 이현수

▶ 객체 지향 프로그래밍

HEAD.py

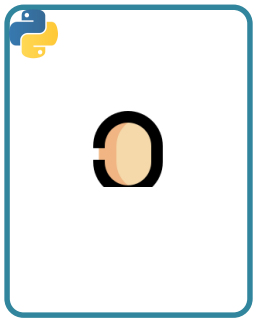


데이터 엔지니어링 이현수

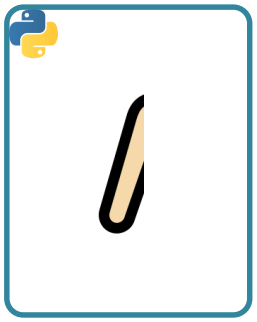
Digital Transformation
All about Education, Chunjae Education.

▶ 객체 지향 프로그래밍

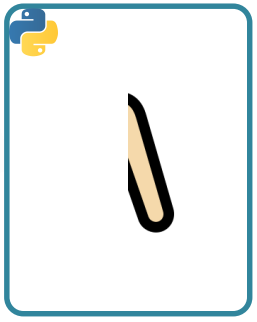
HEAD.py



ARM_R.py



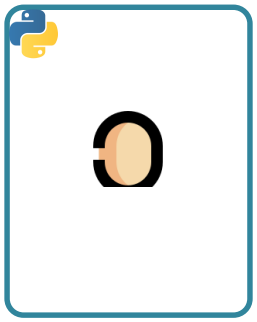
ARM_L.py



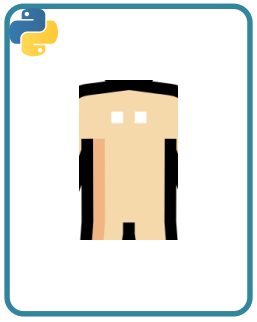
데이터 엔지니어링 이현수

▶ 객체 지향 프로그래밍

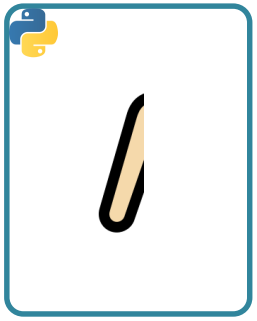
HEAD.py



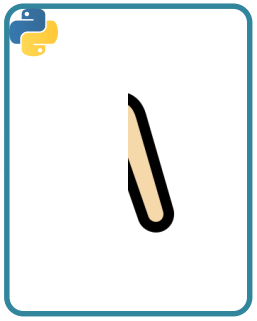
BODY.py



ARM_R.py



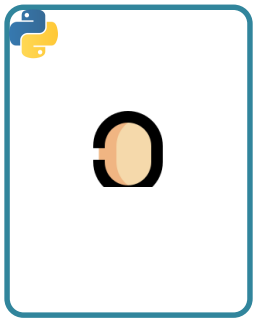
ARM_L.py



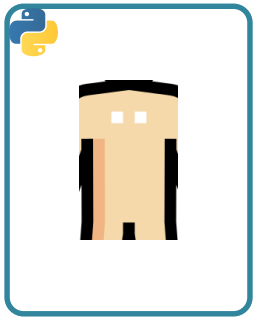
데이터 엔지니어링 이현수

▶ 객체 지향 프로그래밍

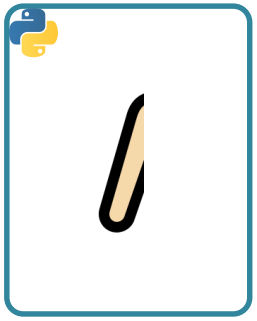
HEAD.py



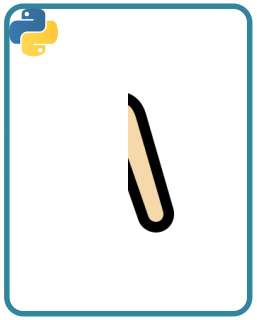
BODY.py



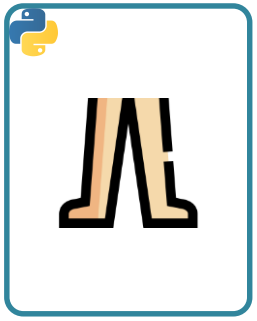
ARM_R.py



ARM_L.py

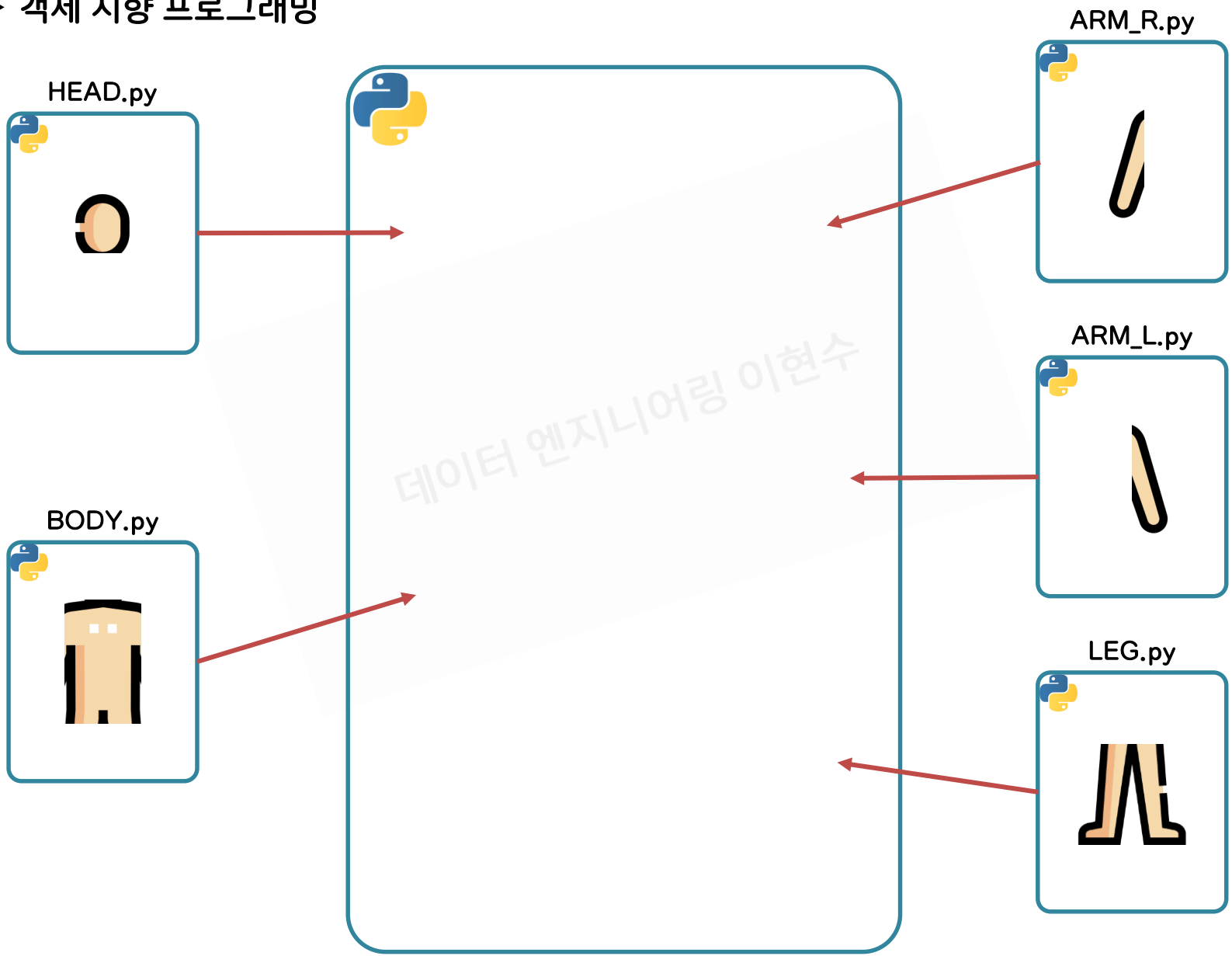


LEG.py



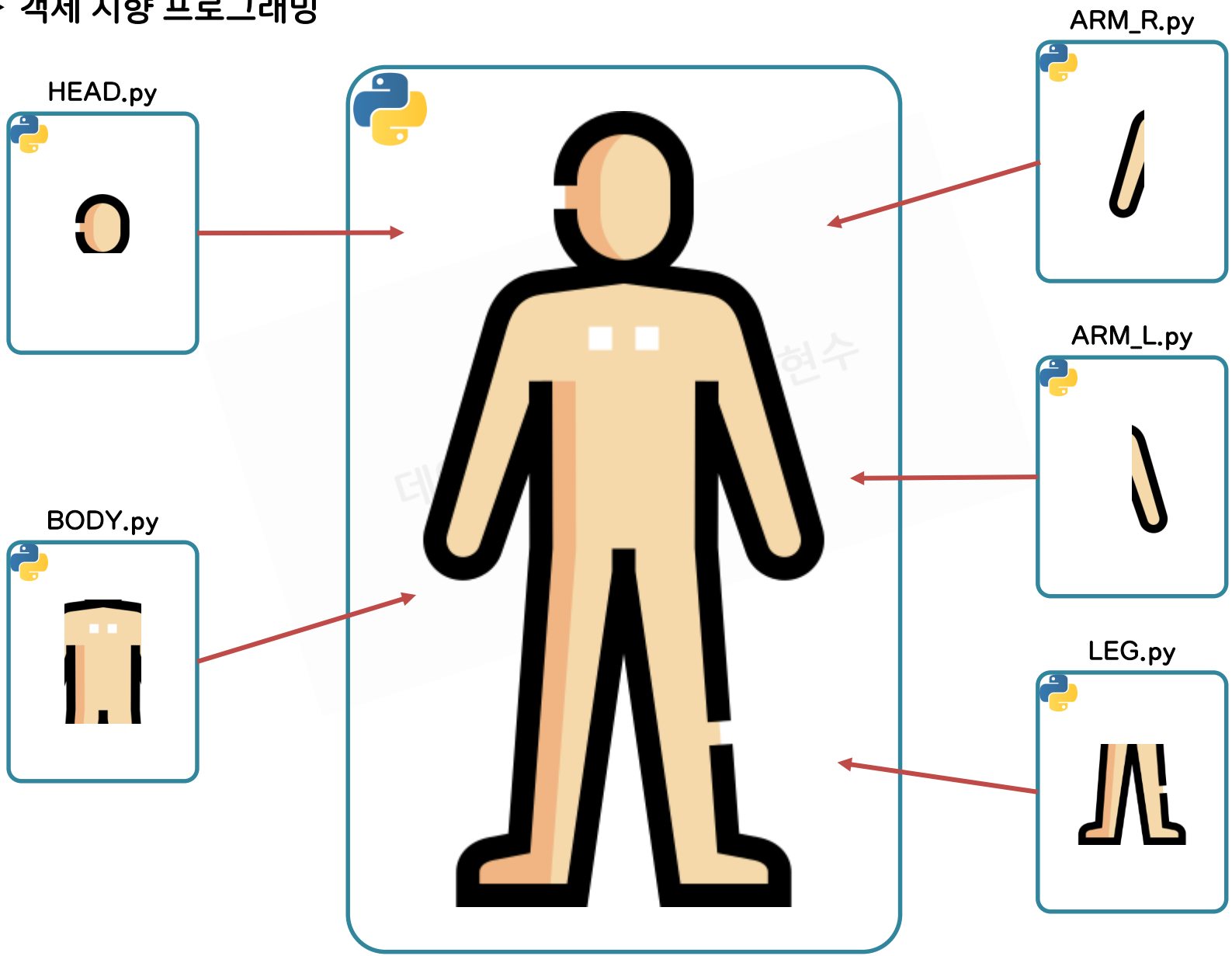
데이터 엔지니어링 이현수

▶ 객체 지향 프로그래밍

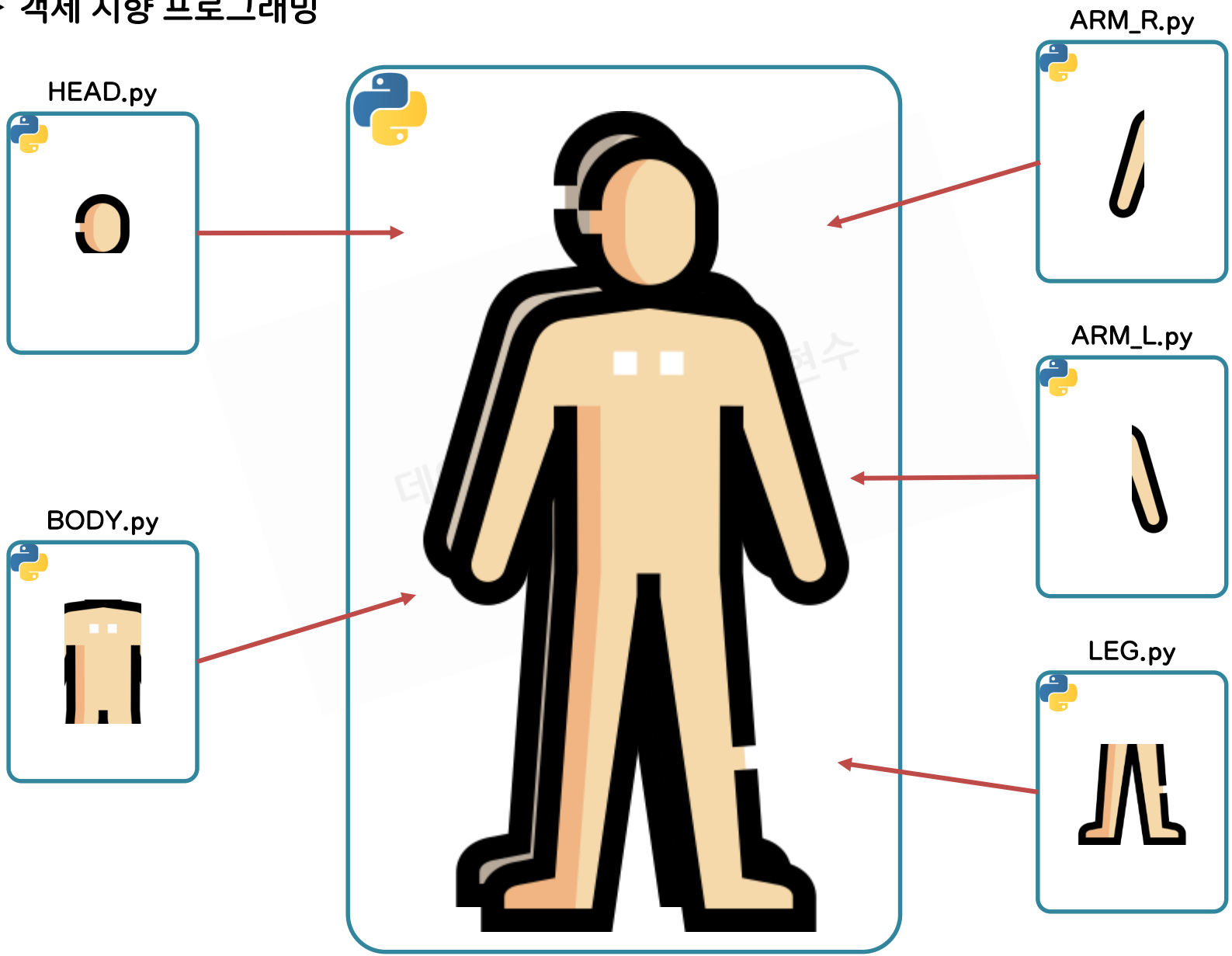


데이터 엔지니어링 이현수

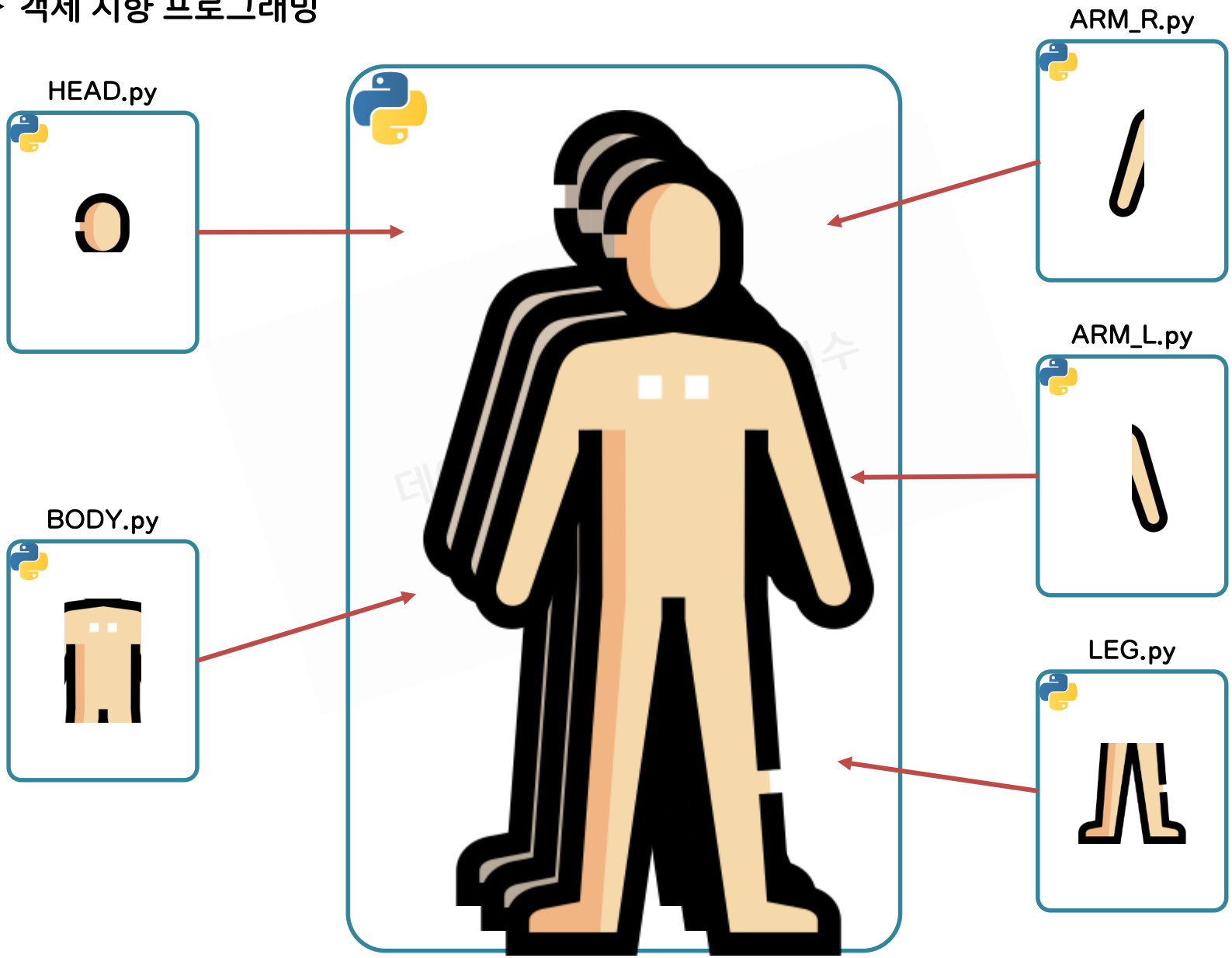
▶ 객체 지향 프로그래밍



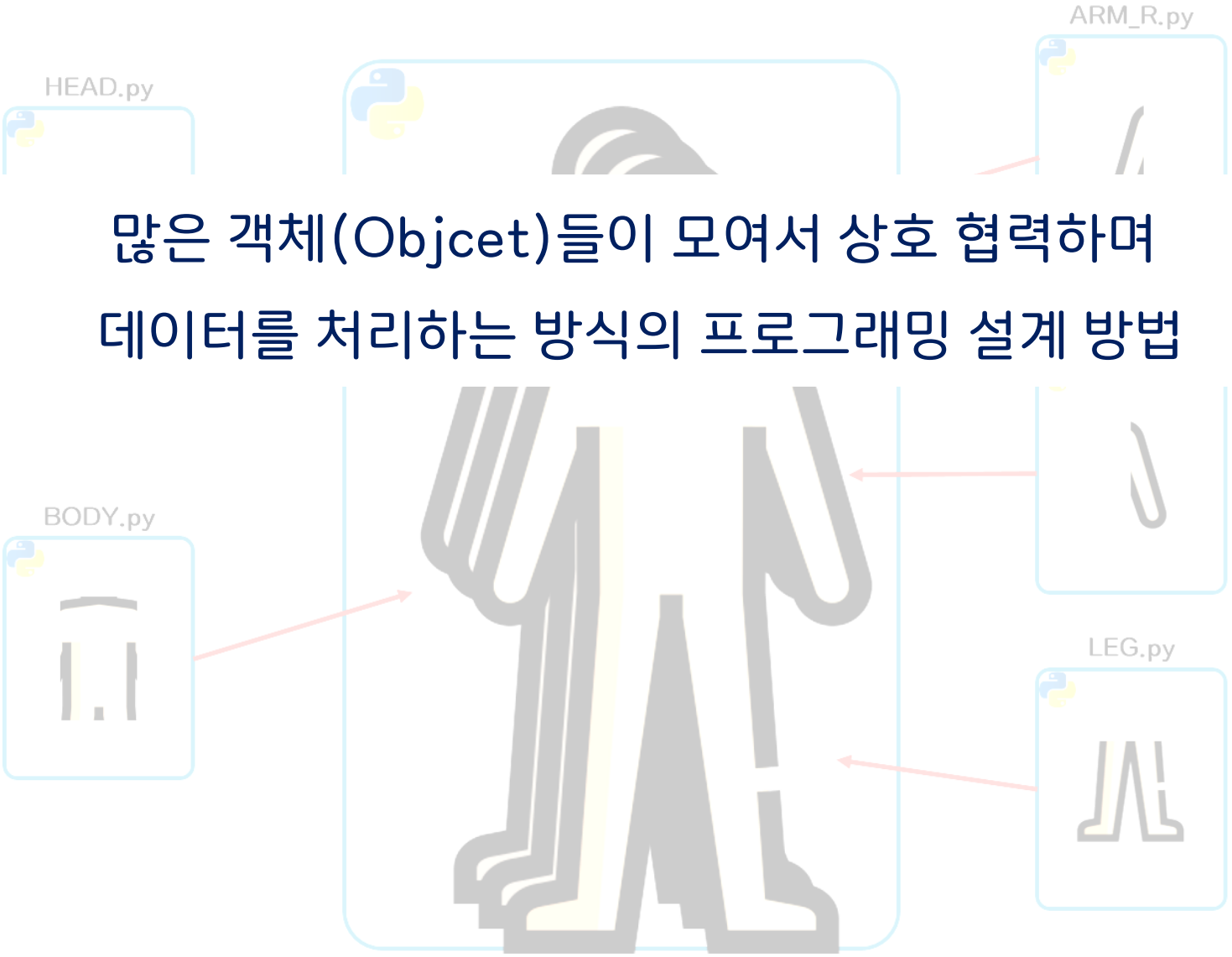
▶ 객체 지향 프로그래밍



▶ 객체 지향 프로그래밍



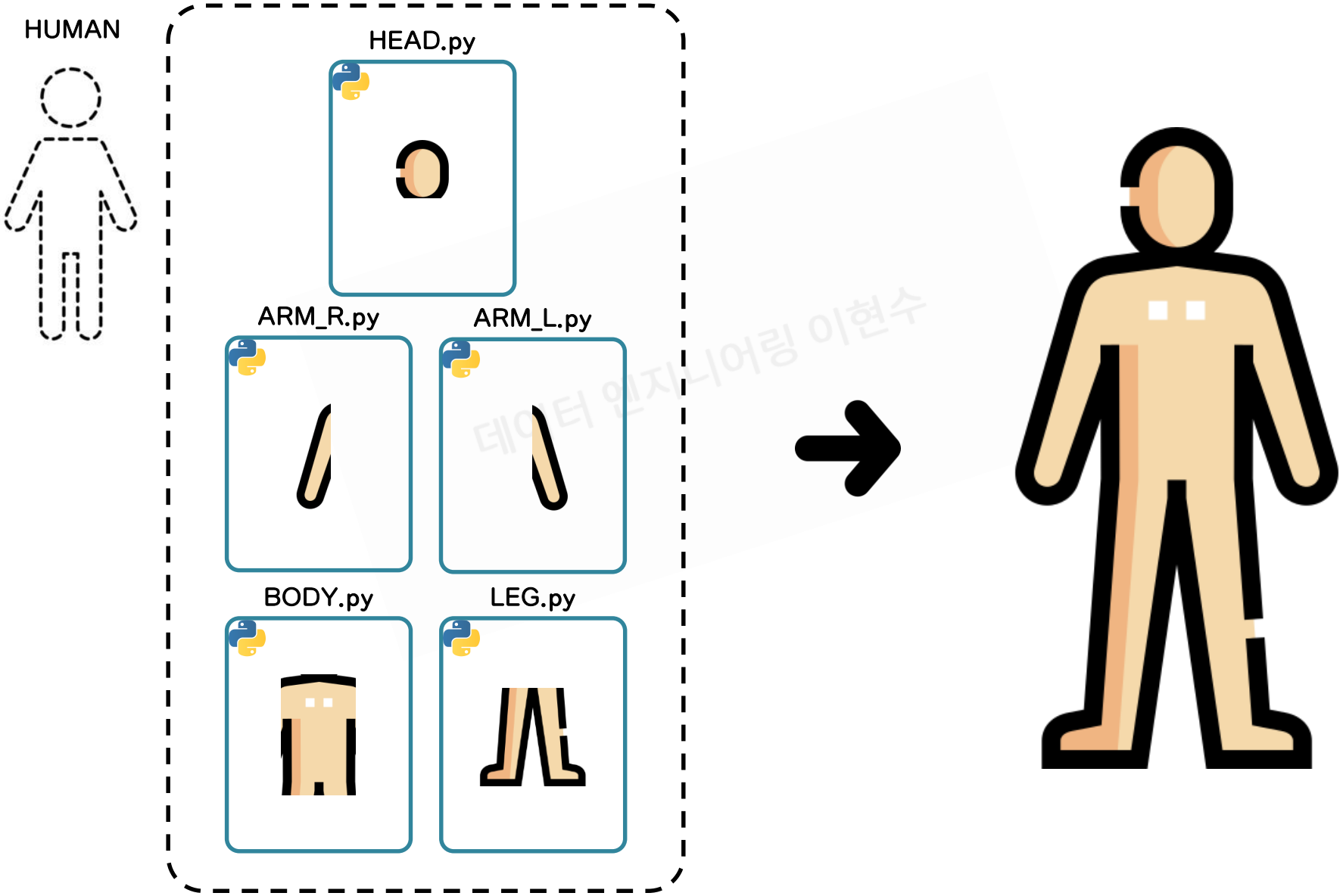
▶ 객체 지향 프로그래밍



많은 객체(Objcet)들이 모여서 상호 협력하며

데이터를 처리하는 방식의 프로그래밍 설계 방법

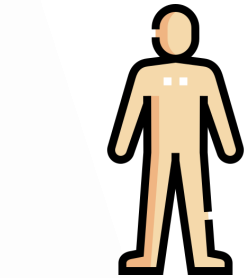
▶ 객체 지향 프로그래밍

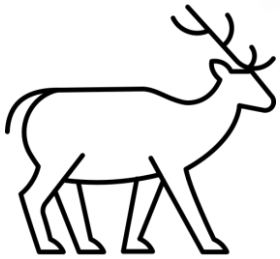


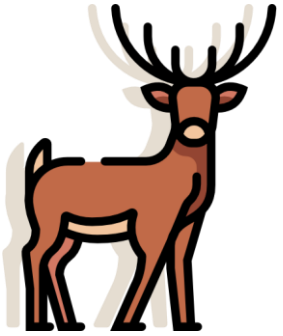
데이터 엔지니어링 이현수

▶ 객체 지향 프로그래밍

클래스 객체

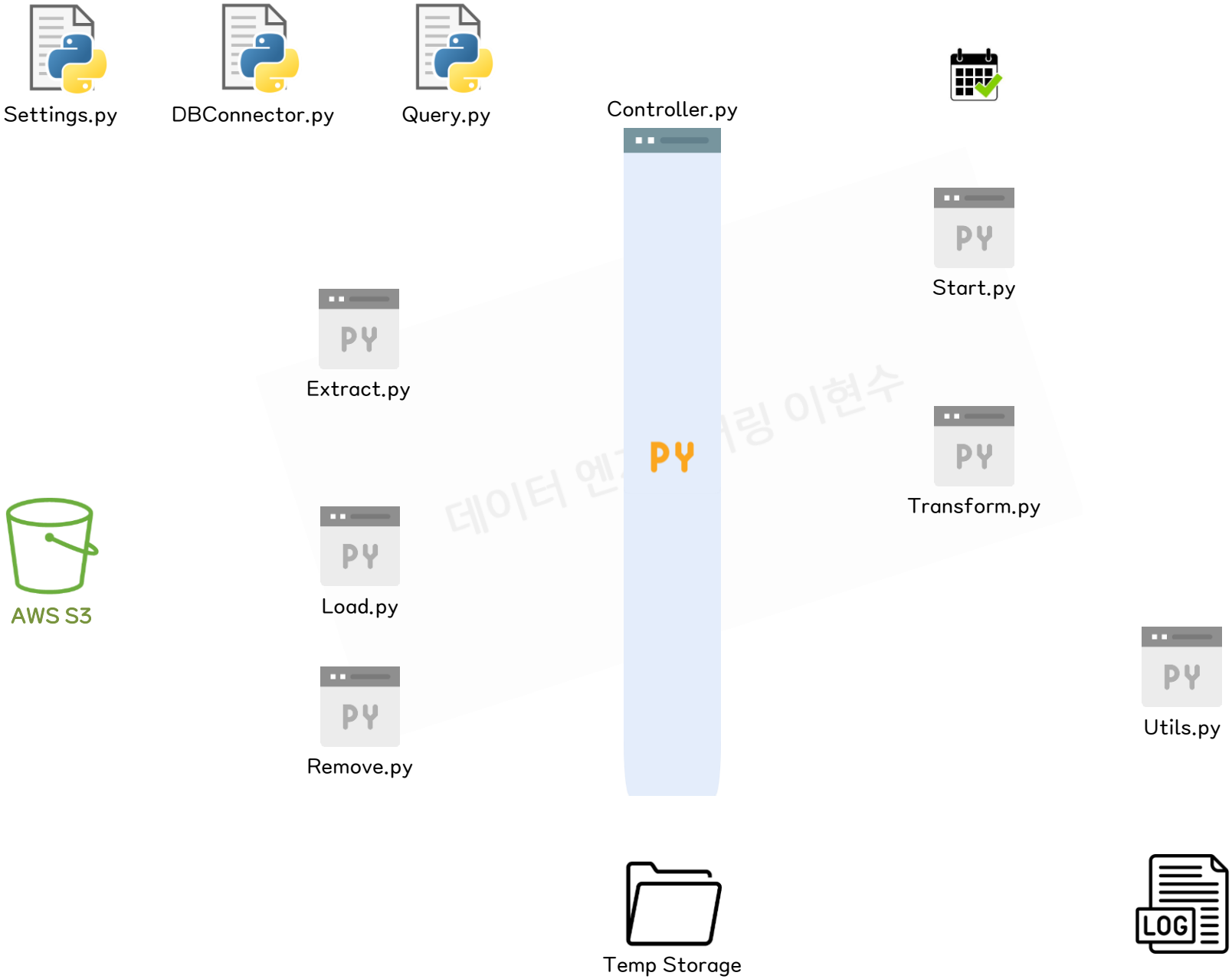




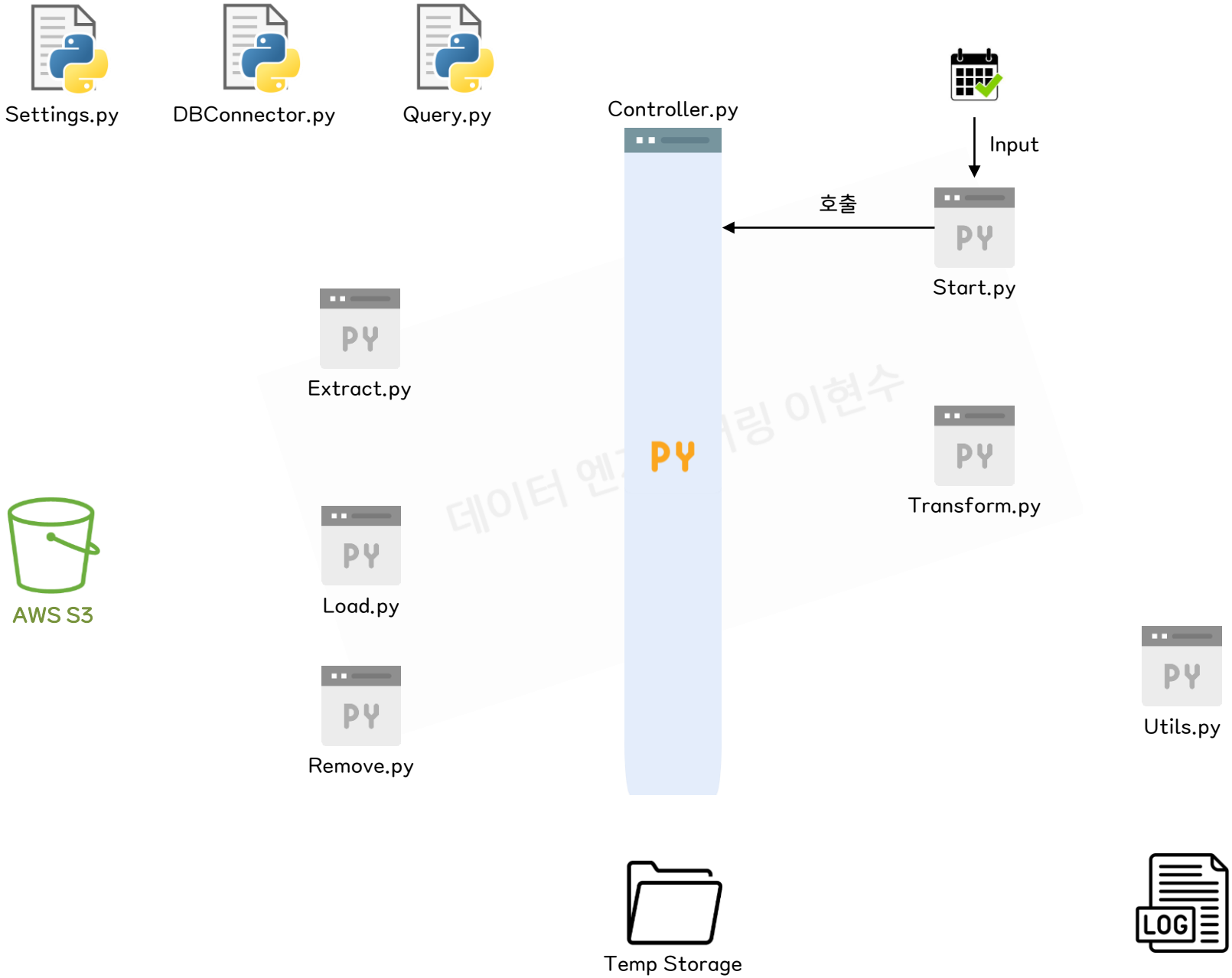


데이터 엔지니어링 이현수

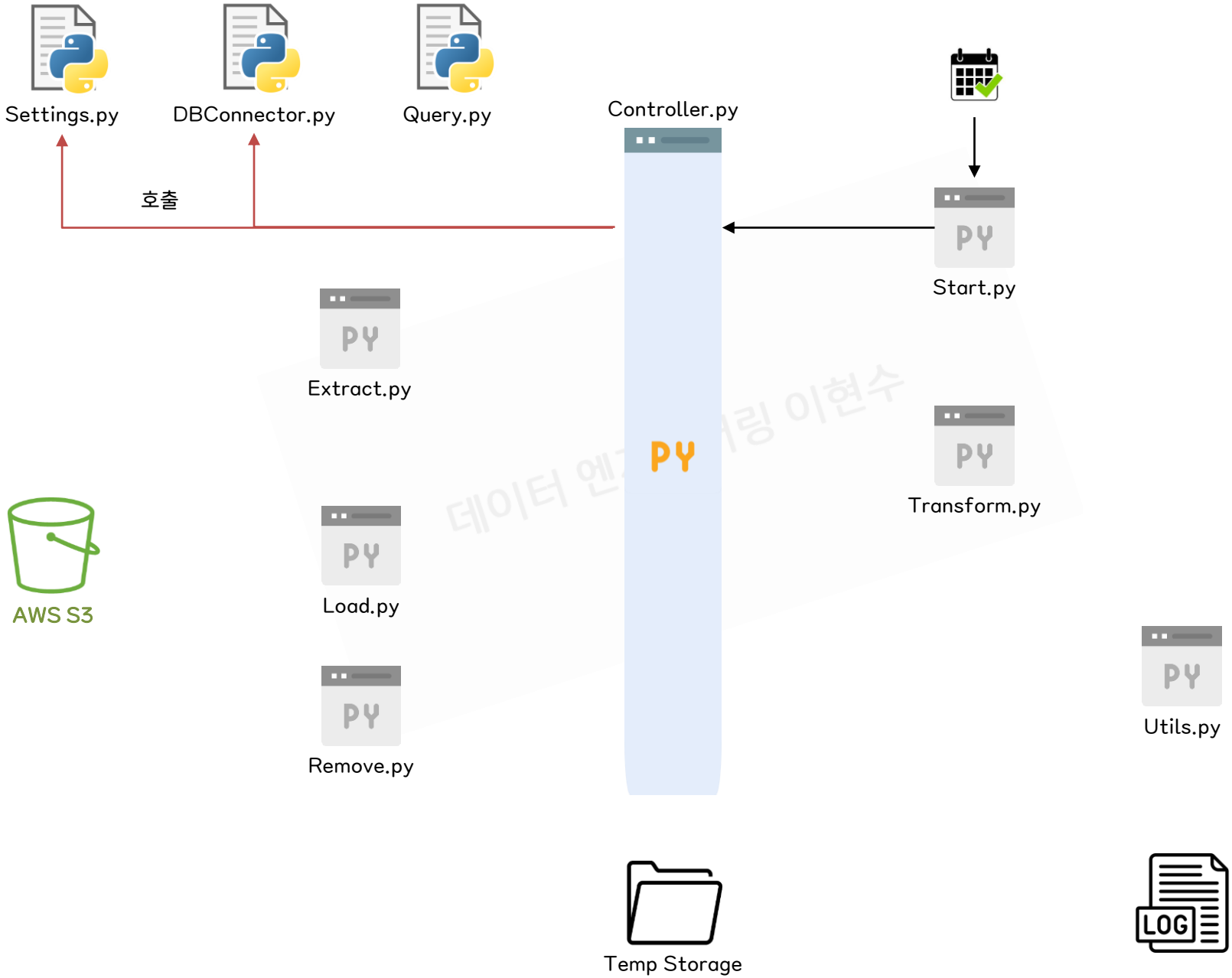
실습 내용 흐름도


Settings.py
DBConnector.py
Query.py
Controller.py
Start.py
Extract.py
Transform.py
AWS S3
Load.py
Utils.py
Remove.py
Temp Storage
LOG

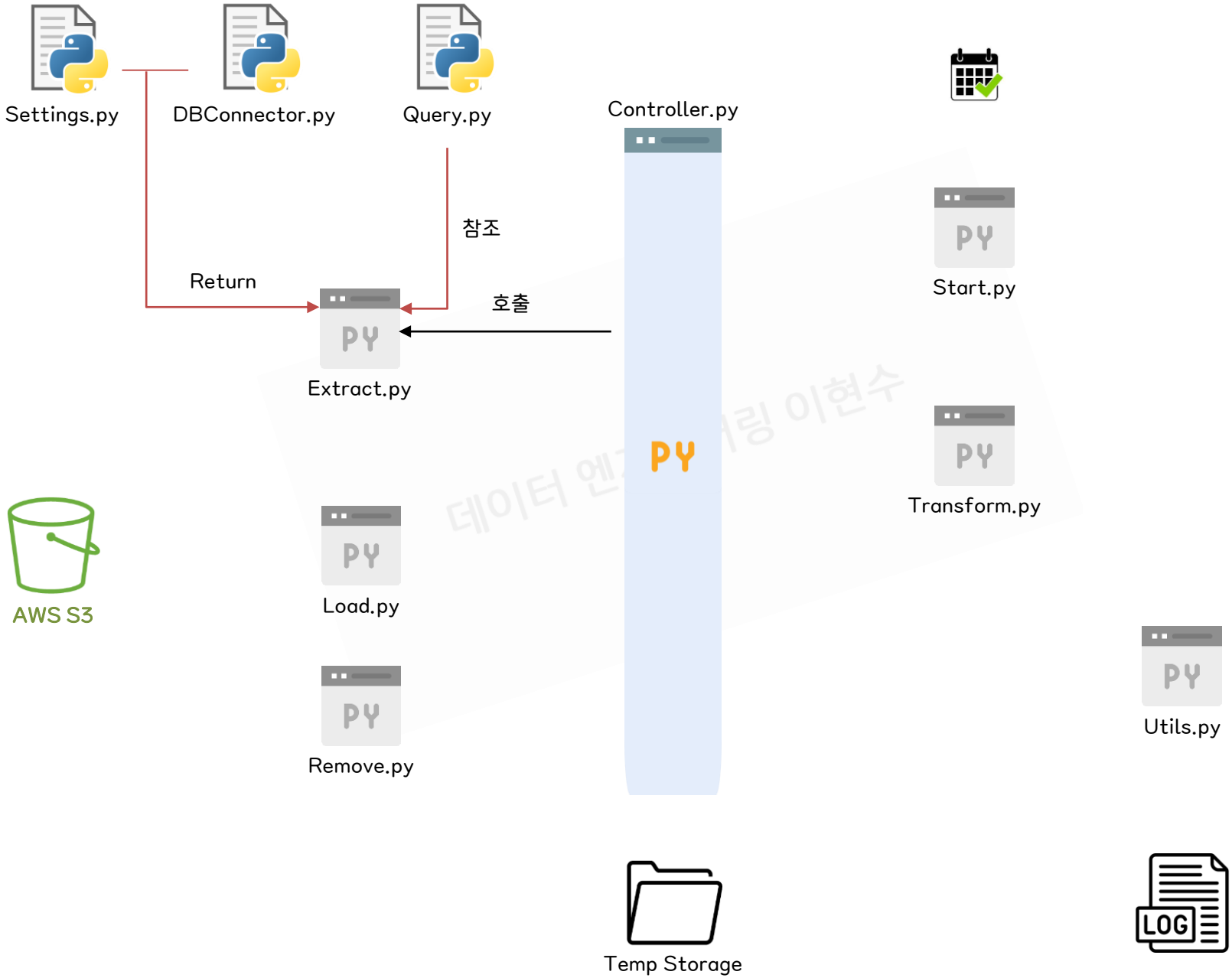
실습 내용 흐름도 - Initial Input



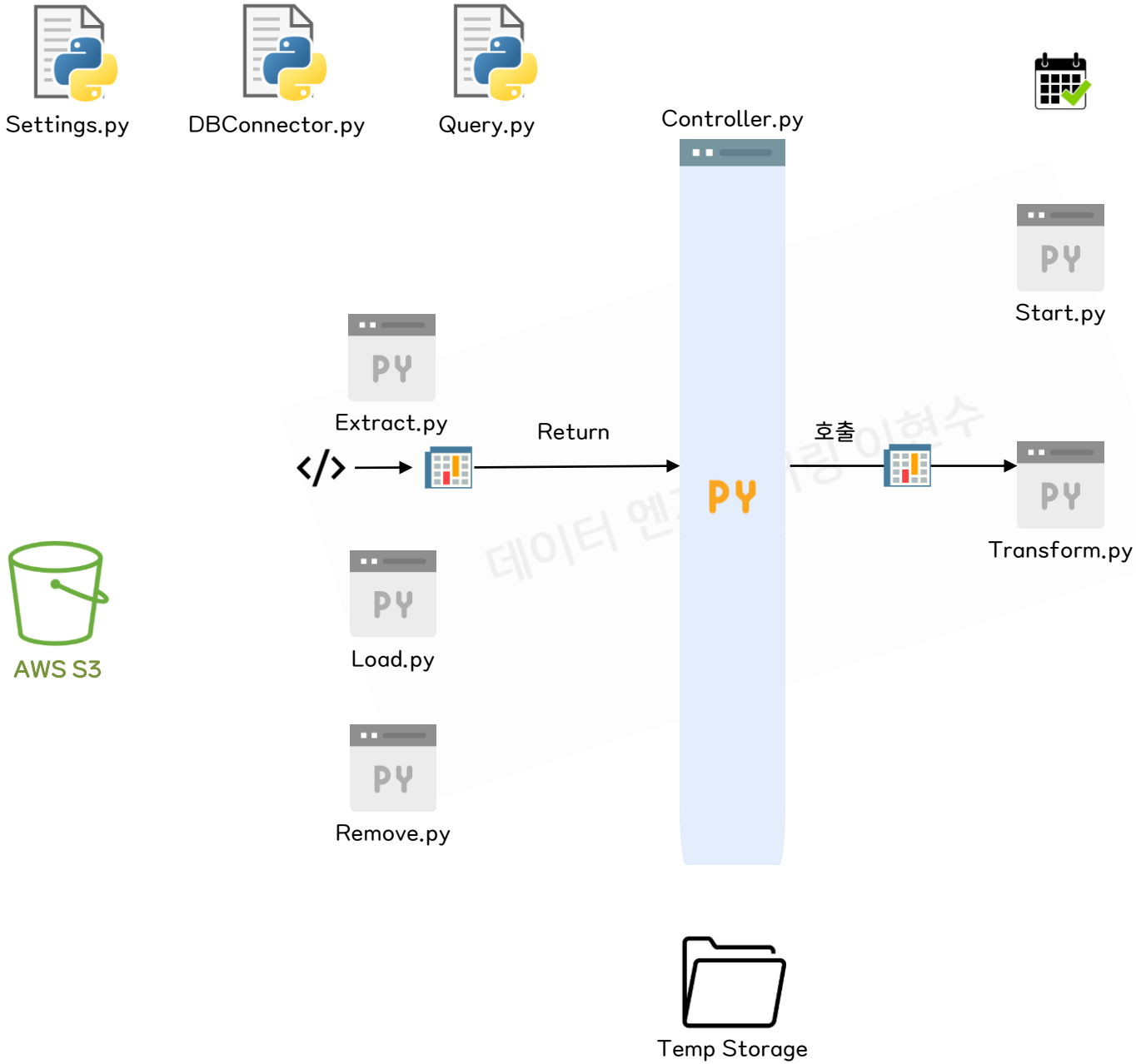
실습 내용 흐름도 - DB Connector 생성



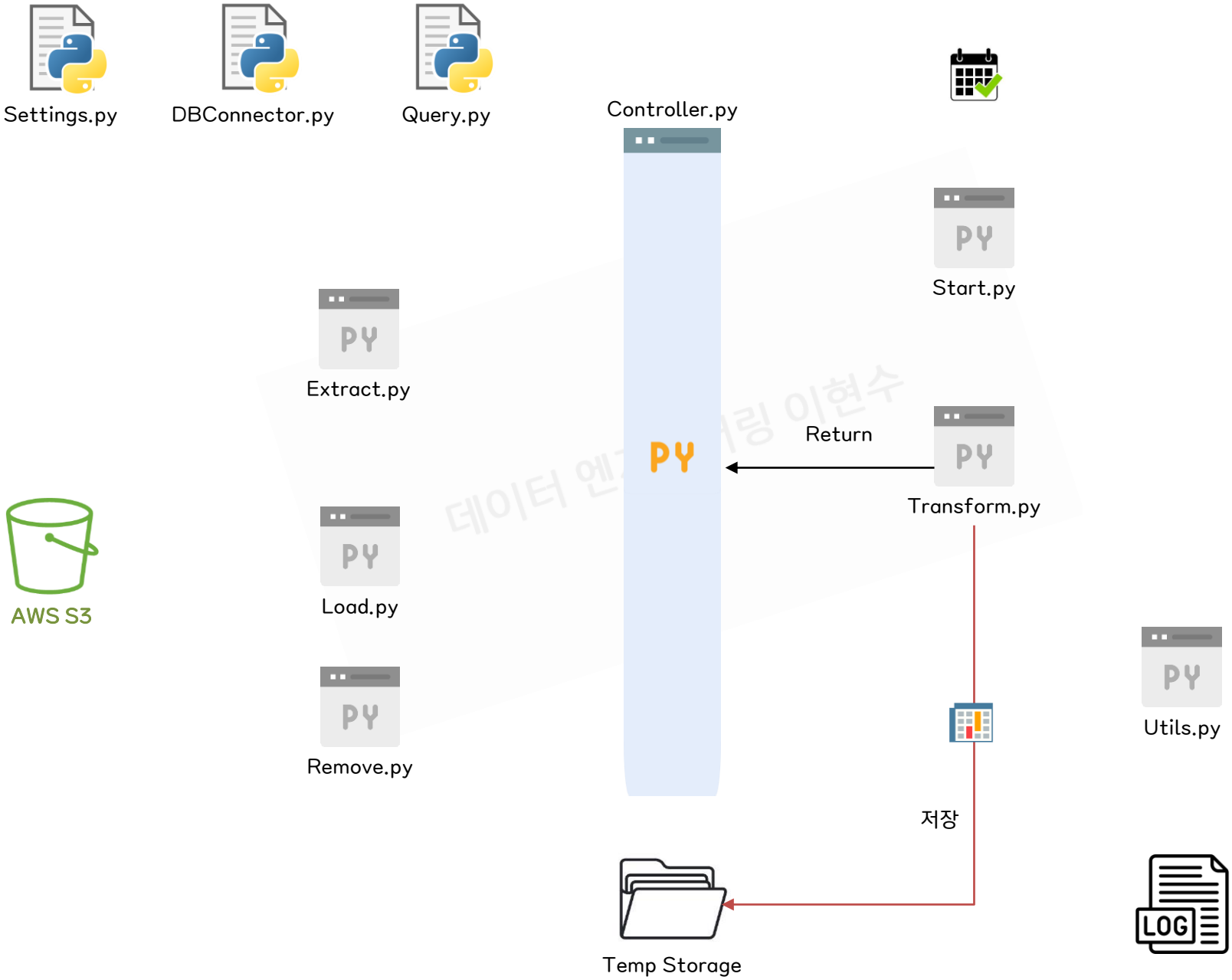
실습 내용 흐름도 - Database 조회



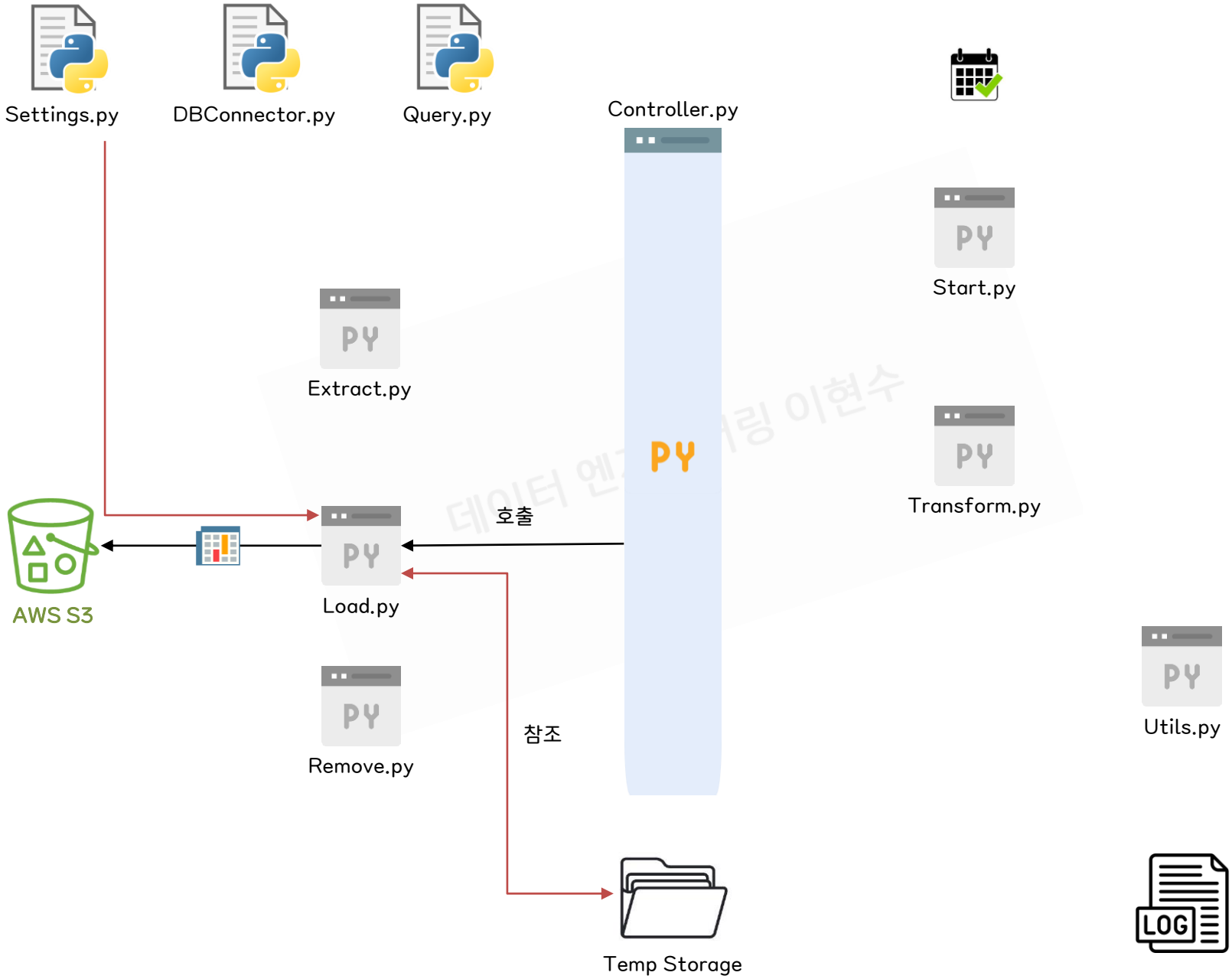
실습 내용 흐름도 - Dataframe 생성



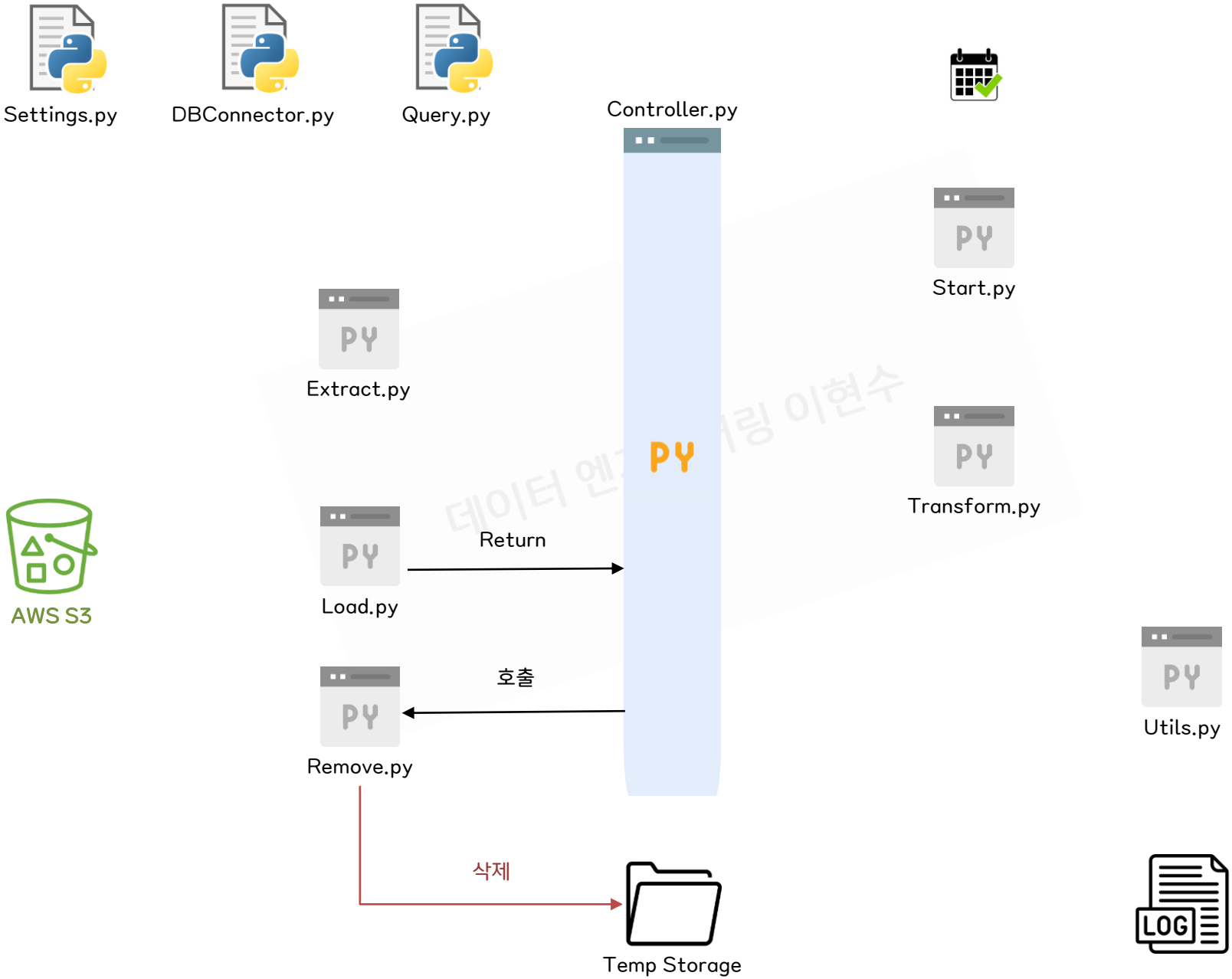
실습 내용 흐름도 - Dataframe 임시 저장



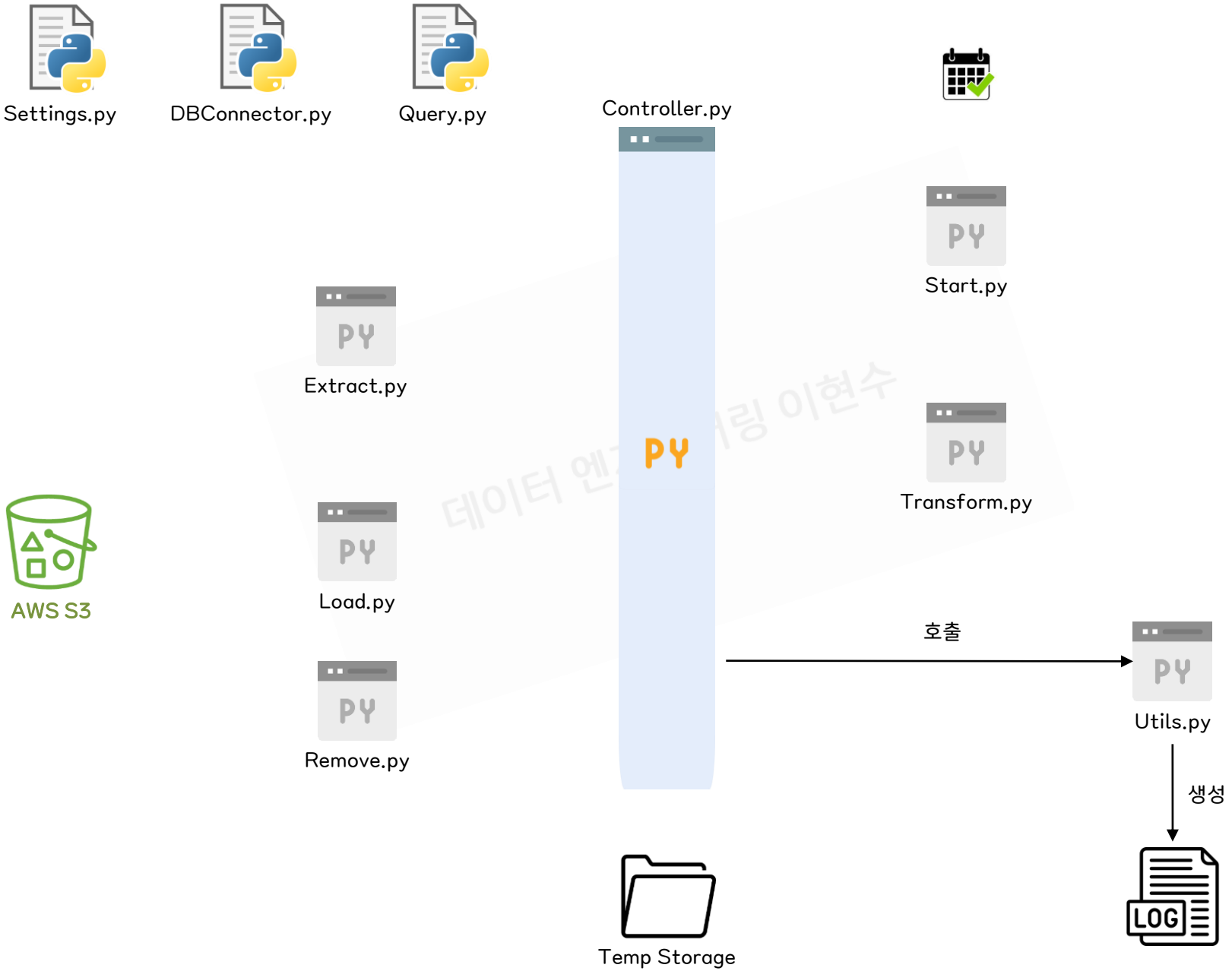
실습 내용 흐름도 - Dataframe 업로드



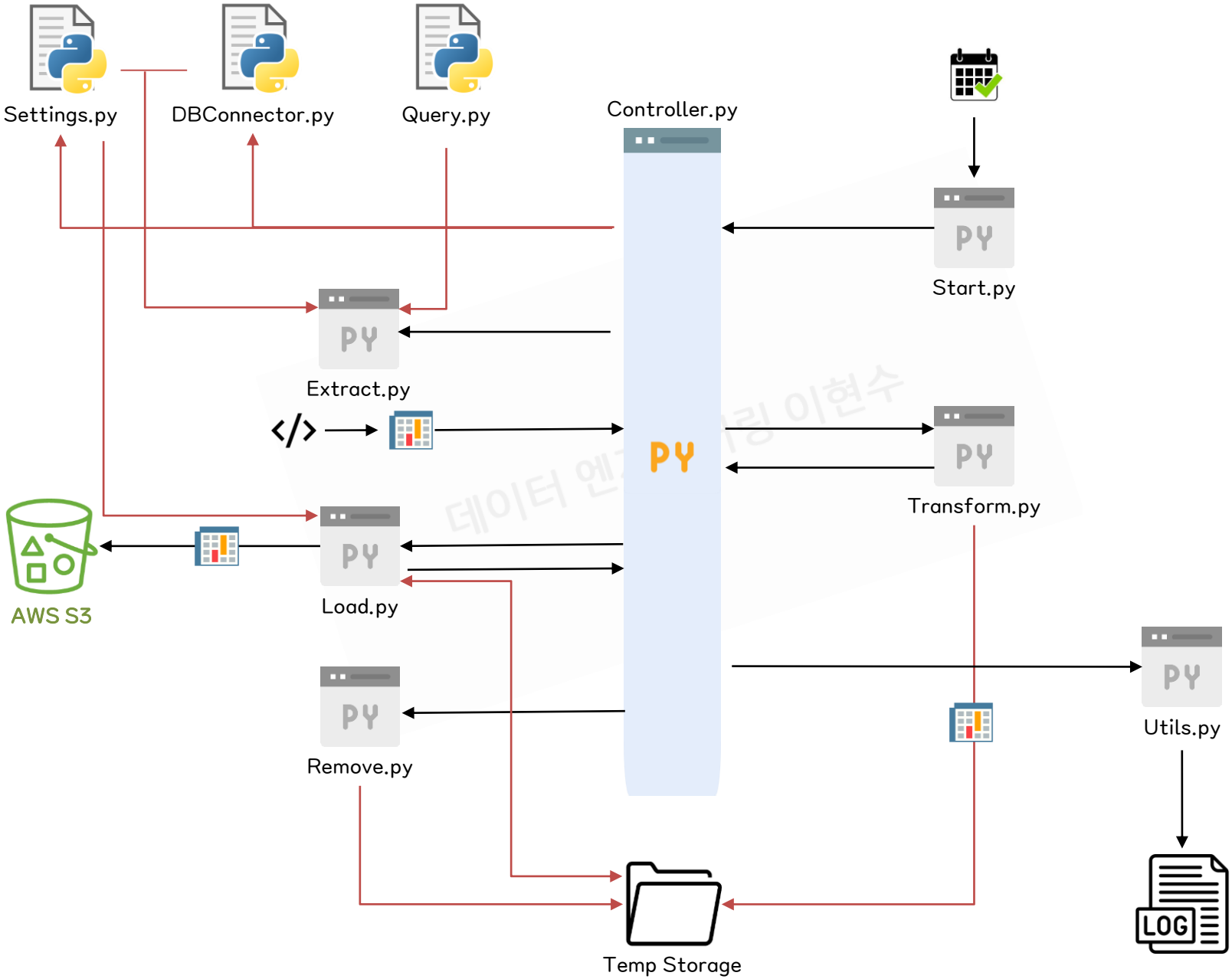
실습 내용 흐름도 - Dataframe 삭제



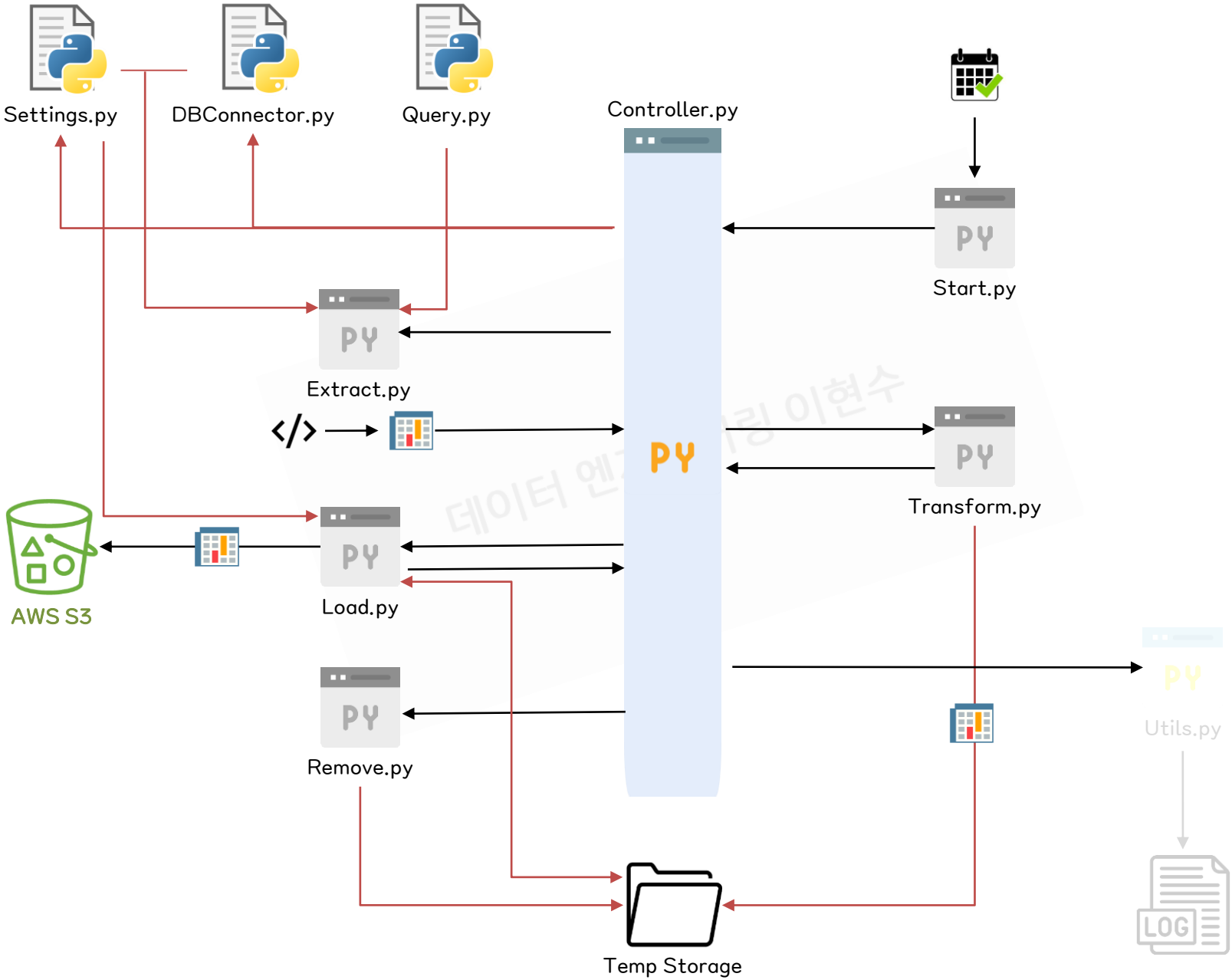
실습 내용 흐름도 - 로그파일 생성



실습 내용 흐름도


Settings.py
DBConnector.py
Query.py
Controller.py
Start.py
Extract.py
Transform.py
AWS S3
Load.py
Remove.py
Utils.py
Temp Storage
LOG

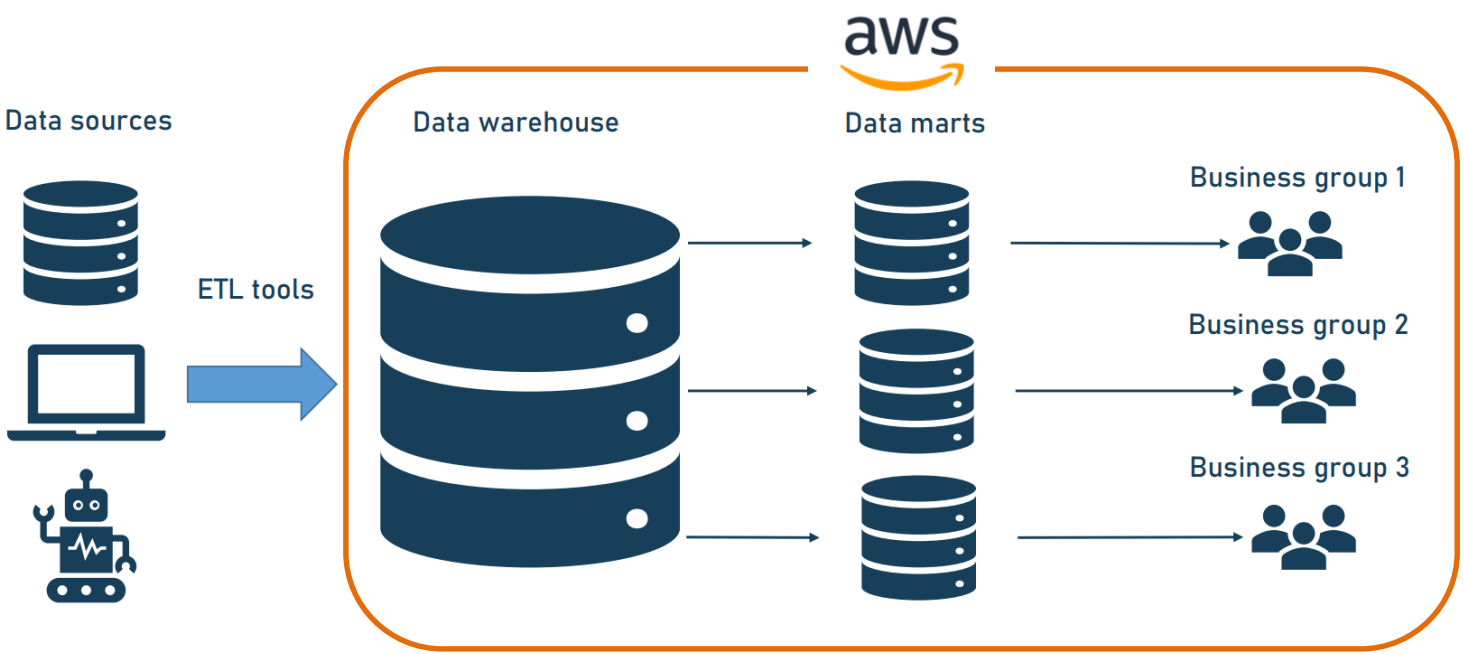
실습 내용 흐름도


Settings.py
DBConnector.py
Query.py
Controller.py
Start.py
Extract.py
Transform.py
AWS S3
Load.py
Remove.py
Utils.py
Temp Storage
LOG

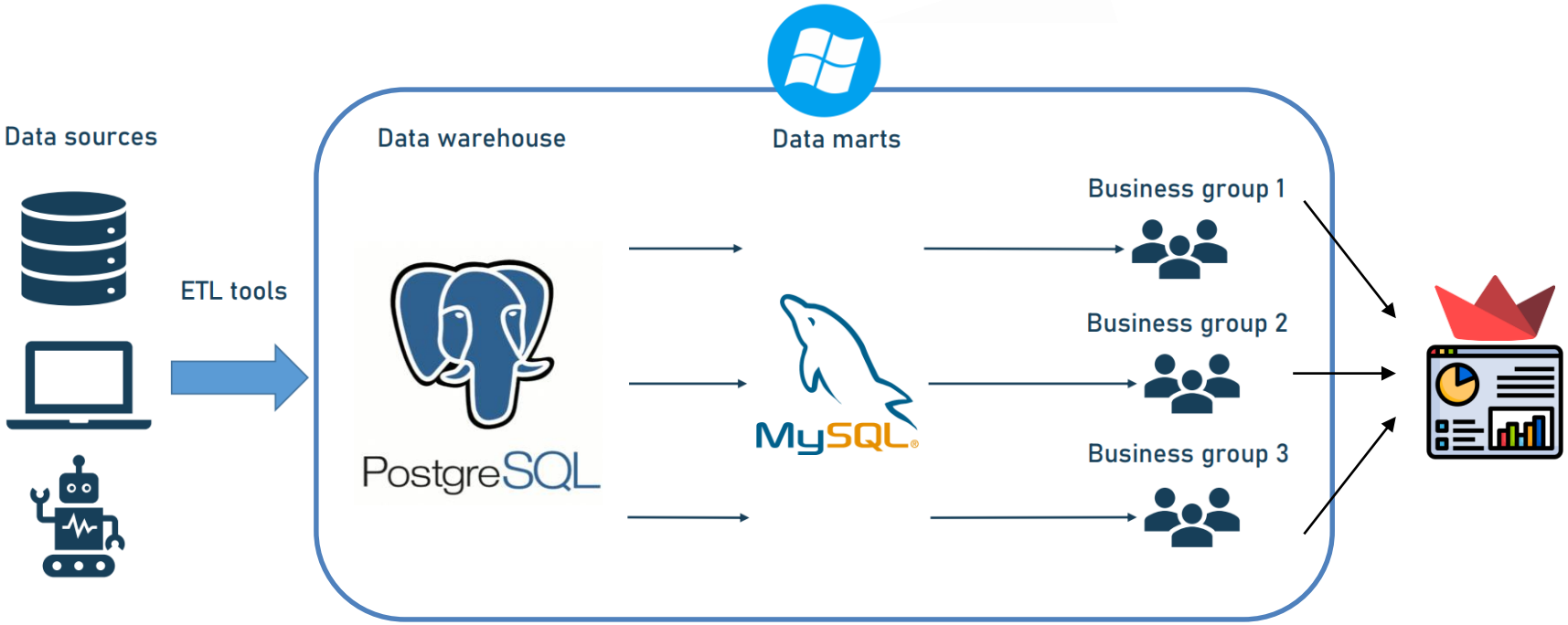
데이터 파이프라인 End-to-End 프로젝트

데이터 엔지니어링 이현승

▶ 프로젝트 아키텍처


aws
Data sources
ETL tools
Data warehouse
Data marts
Business group 1
Business group 2
Business group 3

▶ 프로젝트 아키텍처


Data sources
ETL tools
Data warehouse
Data marts
PostgreSQL
MySQL
Business group 1
Business group 2
Business group 3

데이터 파이프라인 프로젝트 실습 흐름도

데이터 파이프라인 프로젝트 상세 흐름도


Settings.py

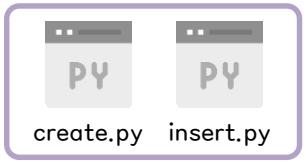

DBConnector.py


Query.py

Controller.py


호출

FAKE DATA




Start.py


Extract.py


Transform.py

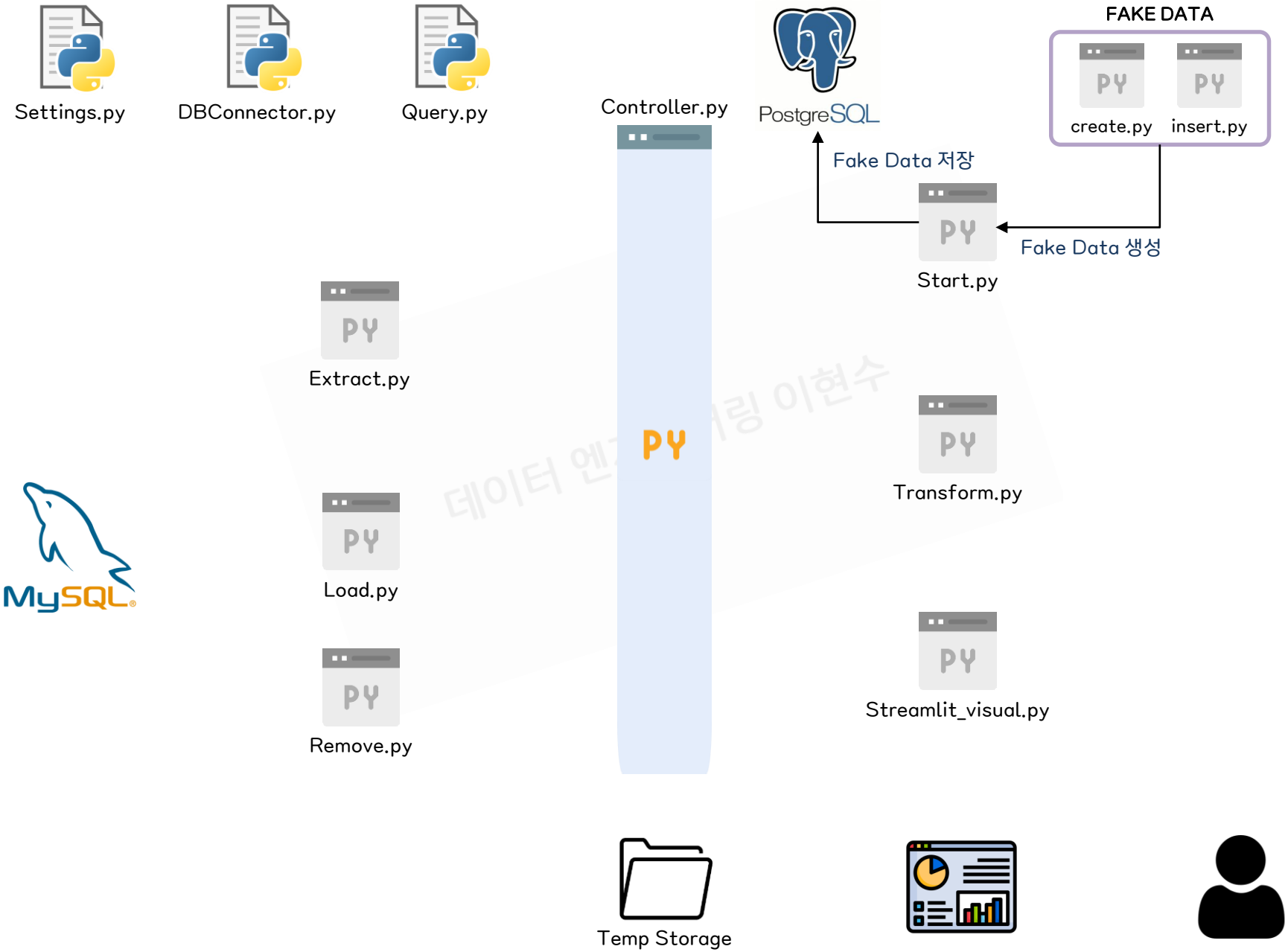

Load.py


Streamlit_visual.py

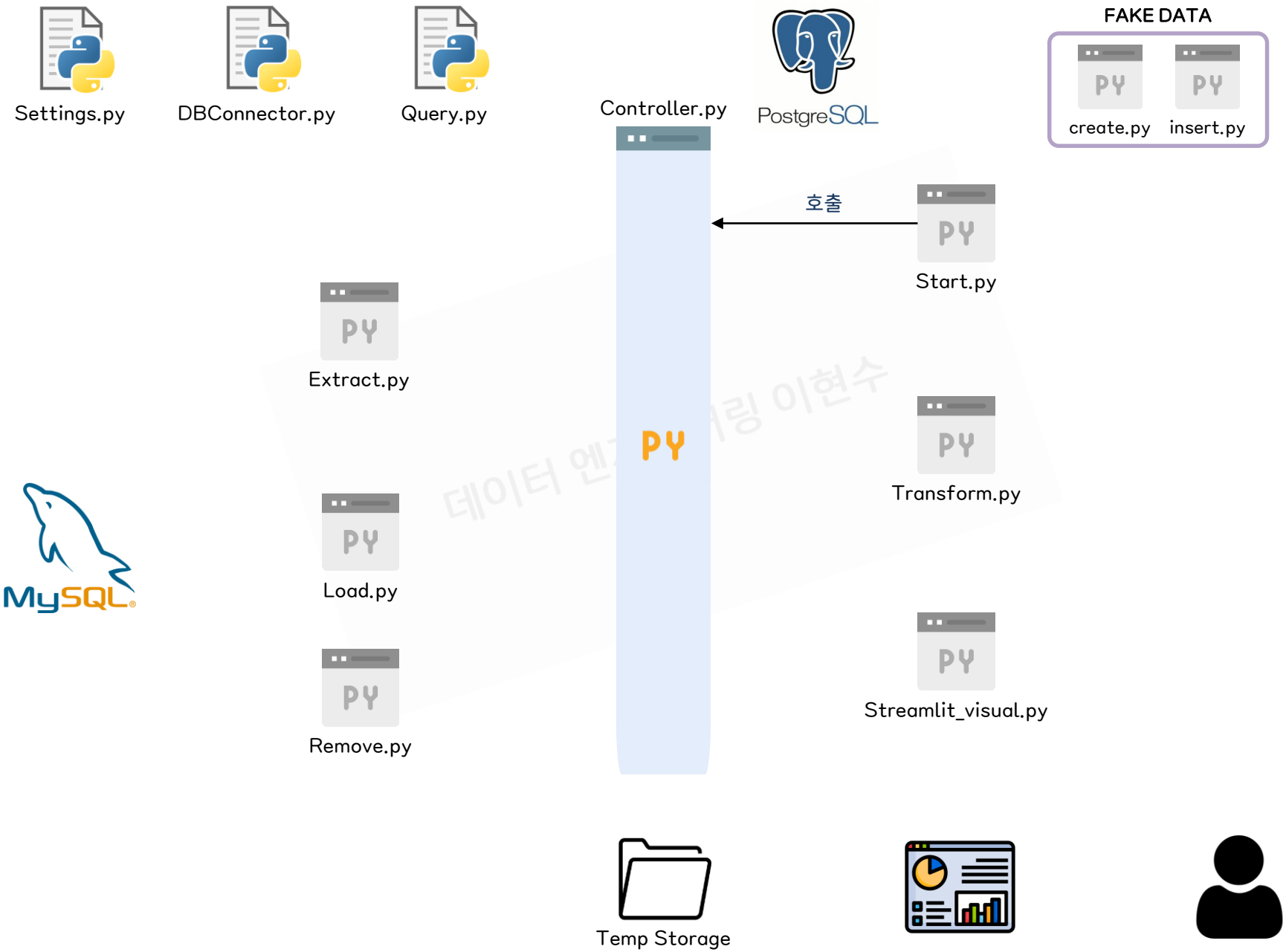

Remove.py


Temp Storage

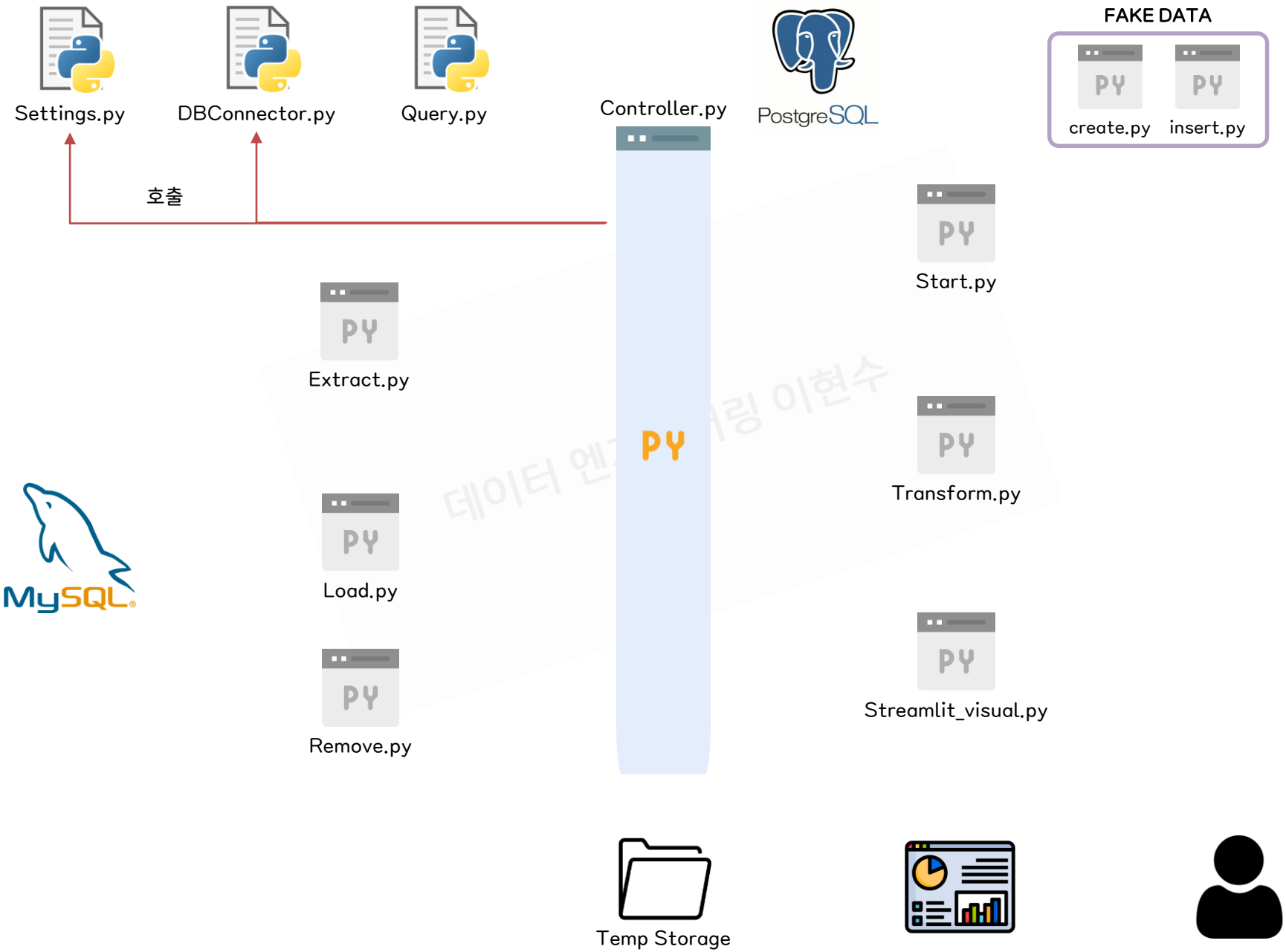
데이터 파이프라인 프로젝트 상세 흐름도



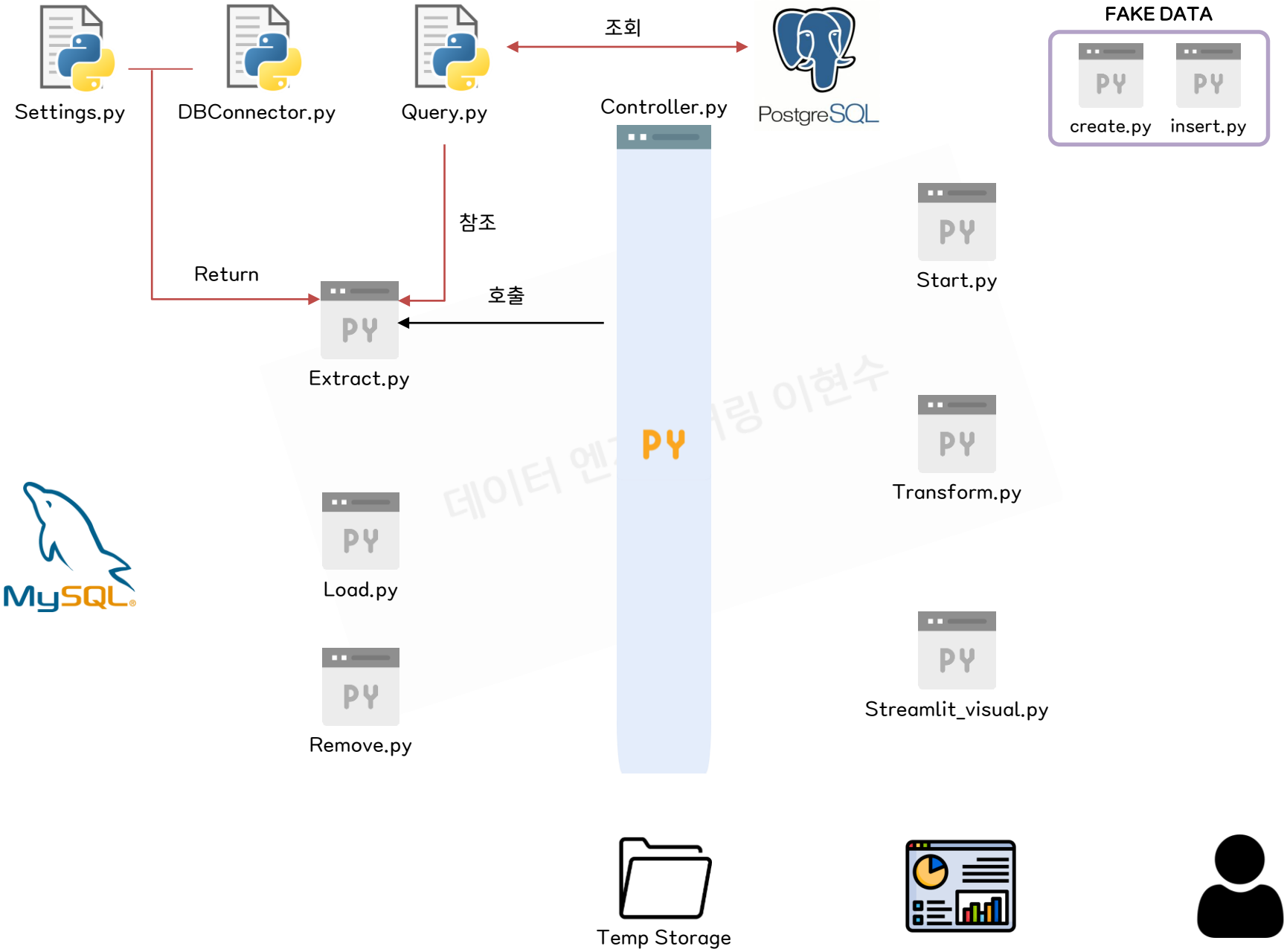
데이터 파이프라인 프로젝트 상세 흐름도



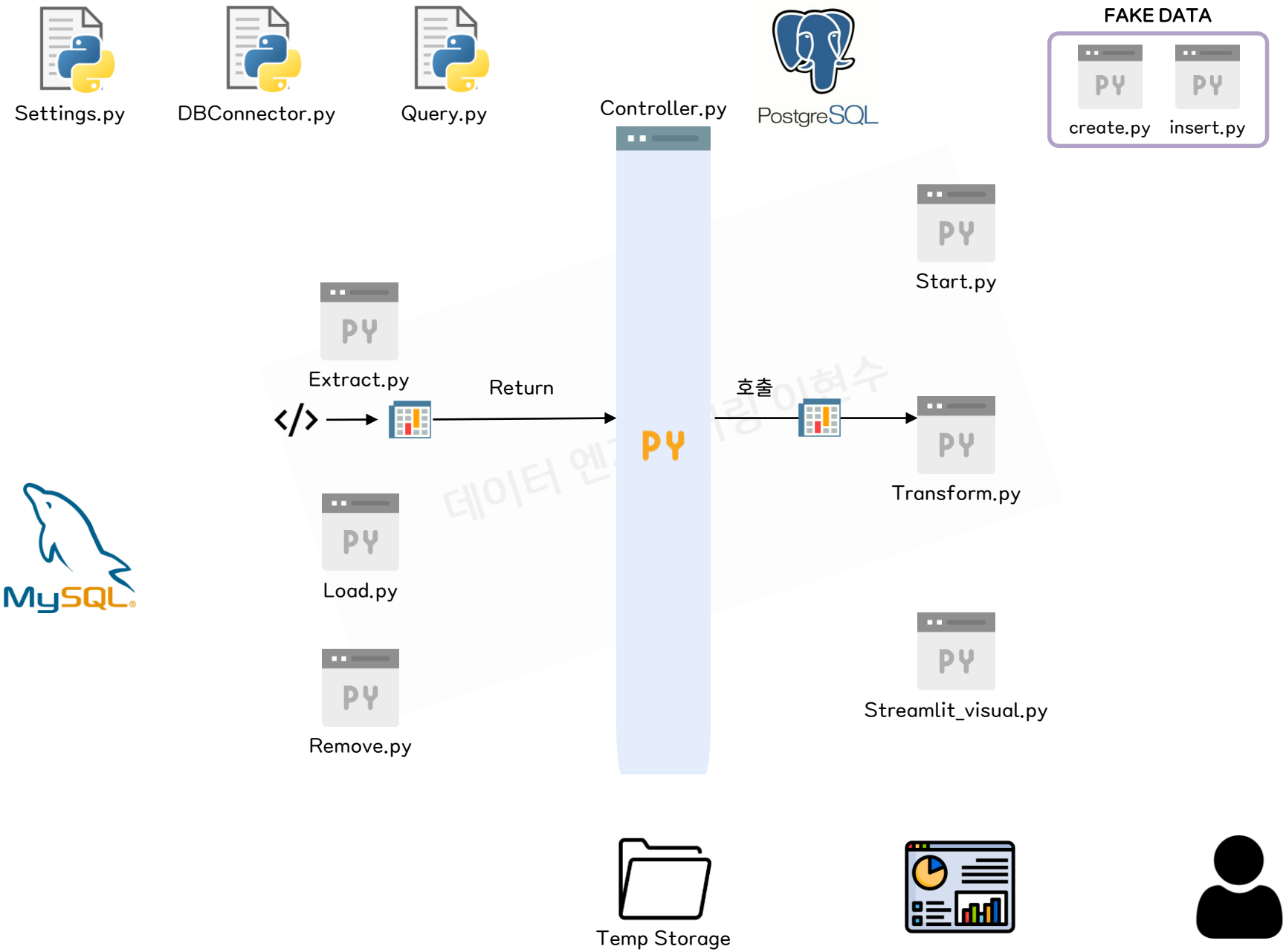
데이터 파이프라인 프로젝트 상세 흐름도



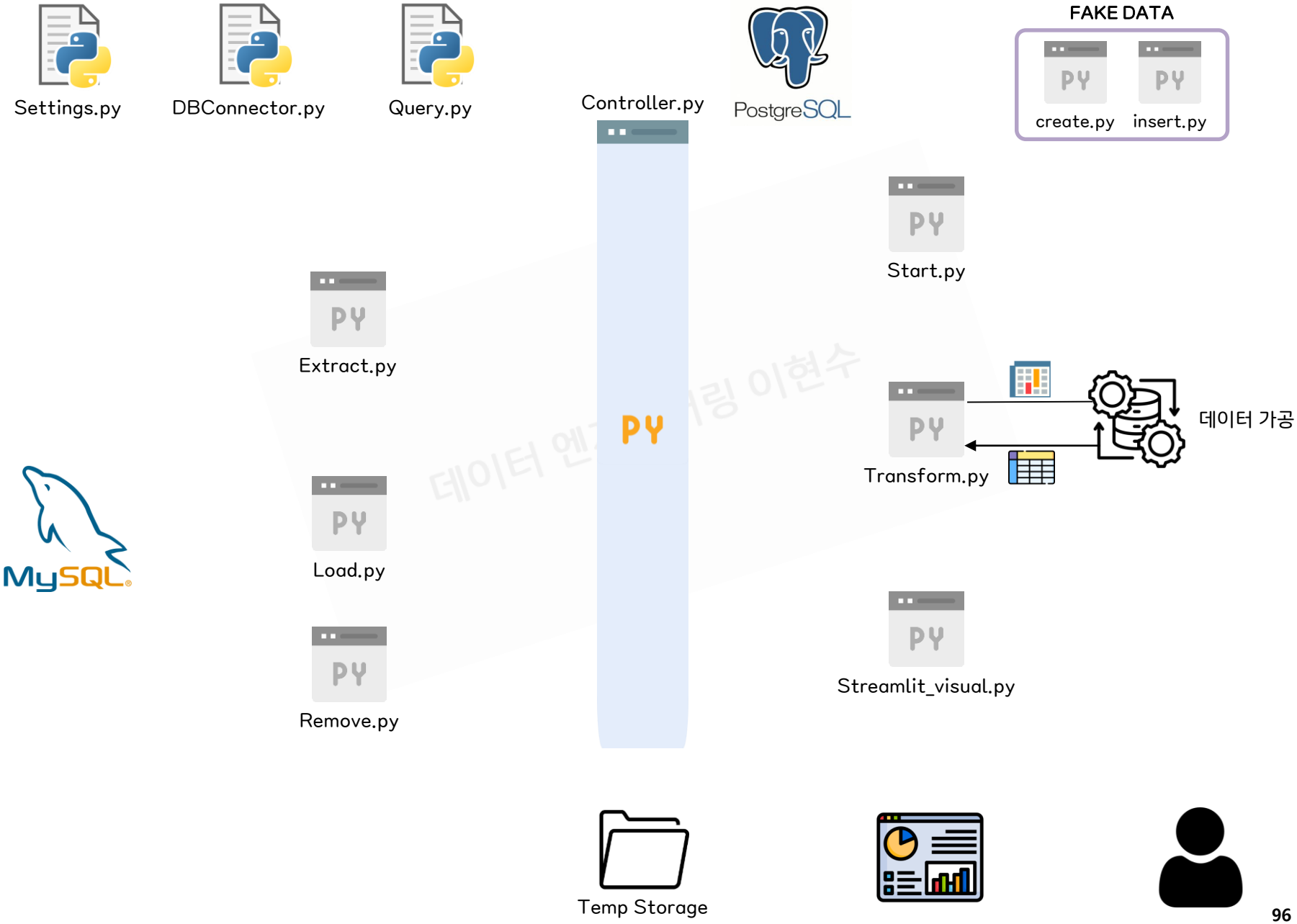
데이터 파이프라인 프로젝트 상세 흐름도



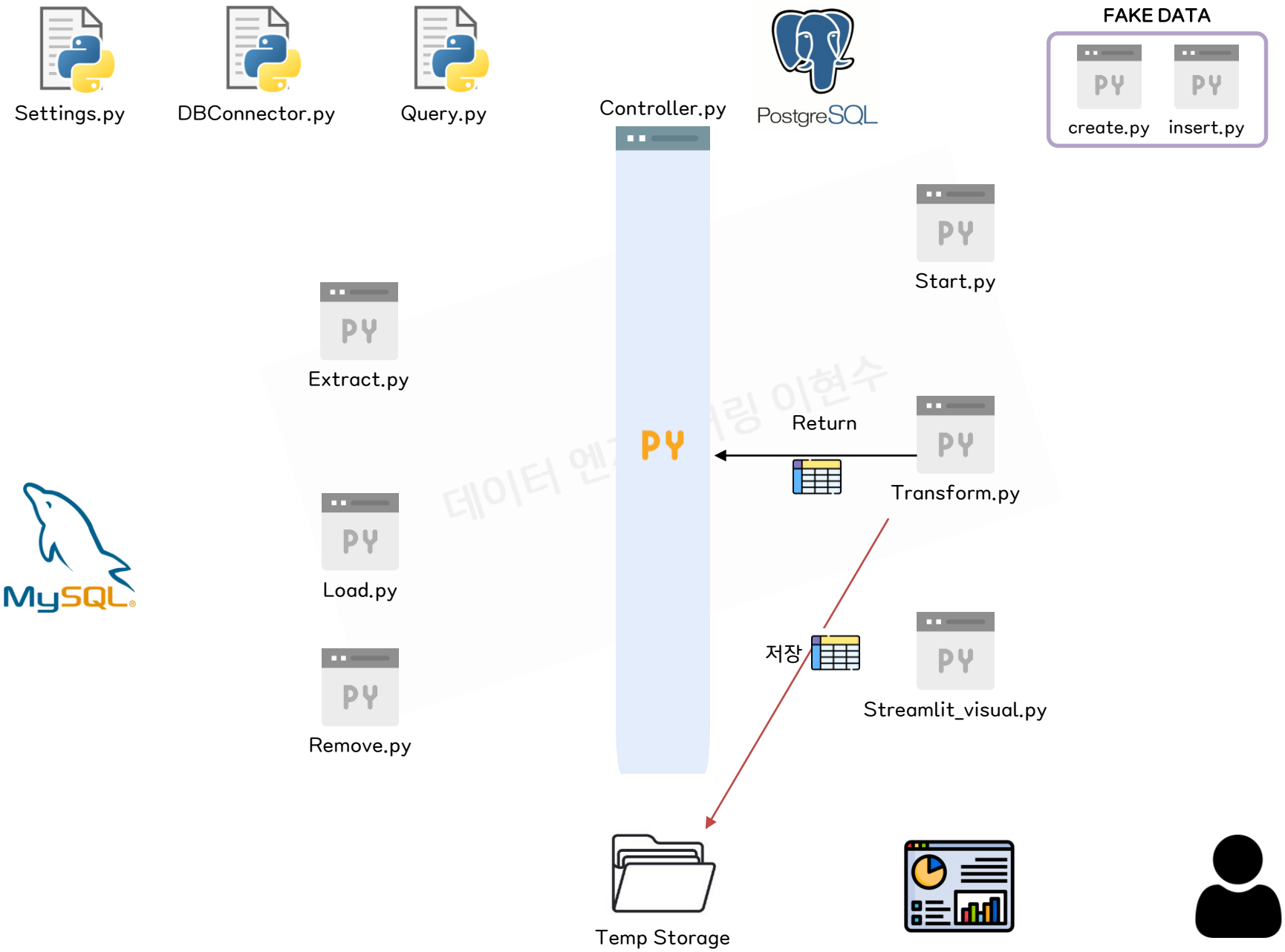
데이터 파이프라인 프로젝트 상세 흐름도



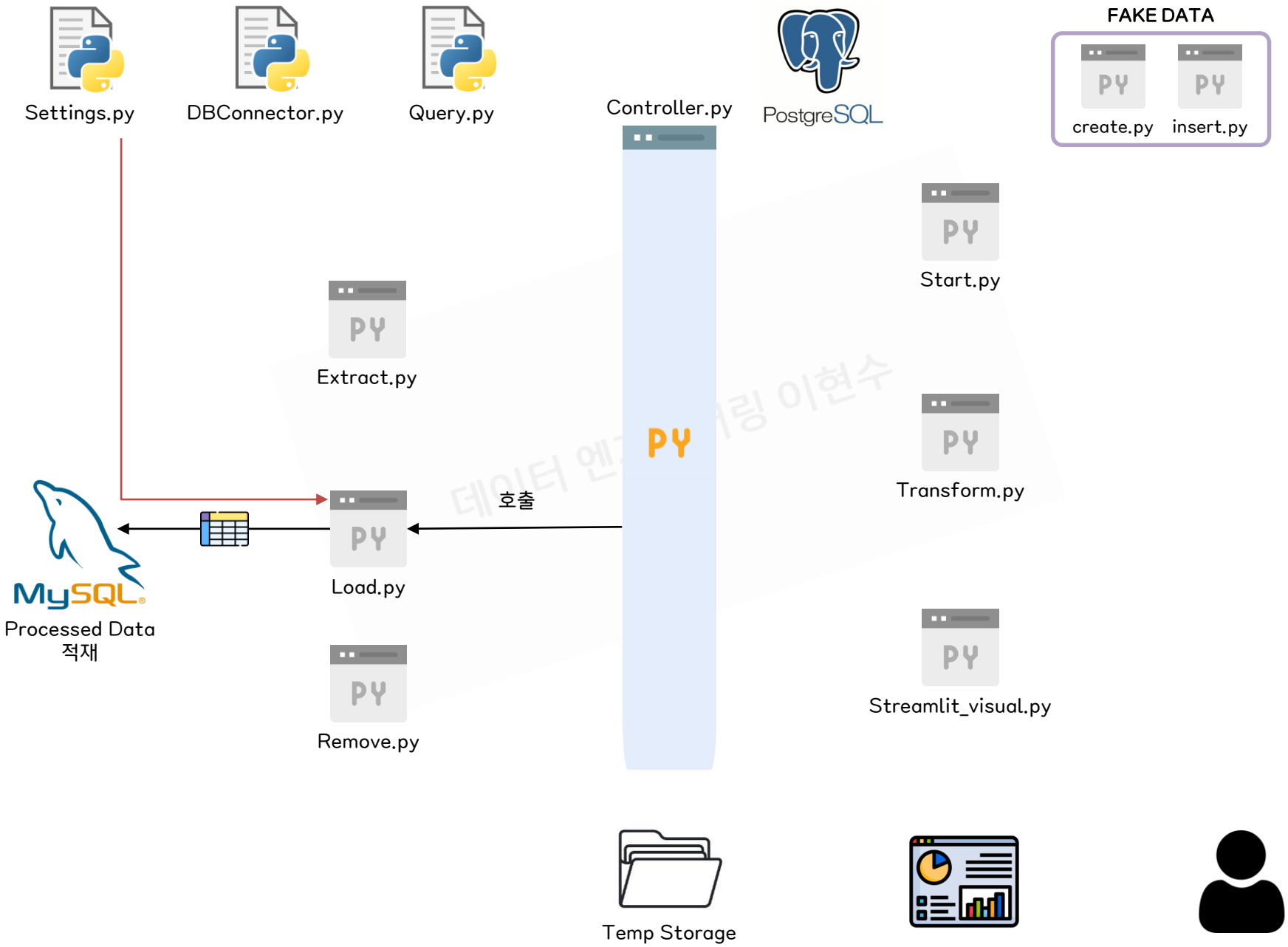
데이터 파이프라인 프로젝트 상세 흐름도



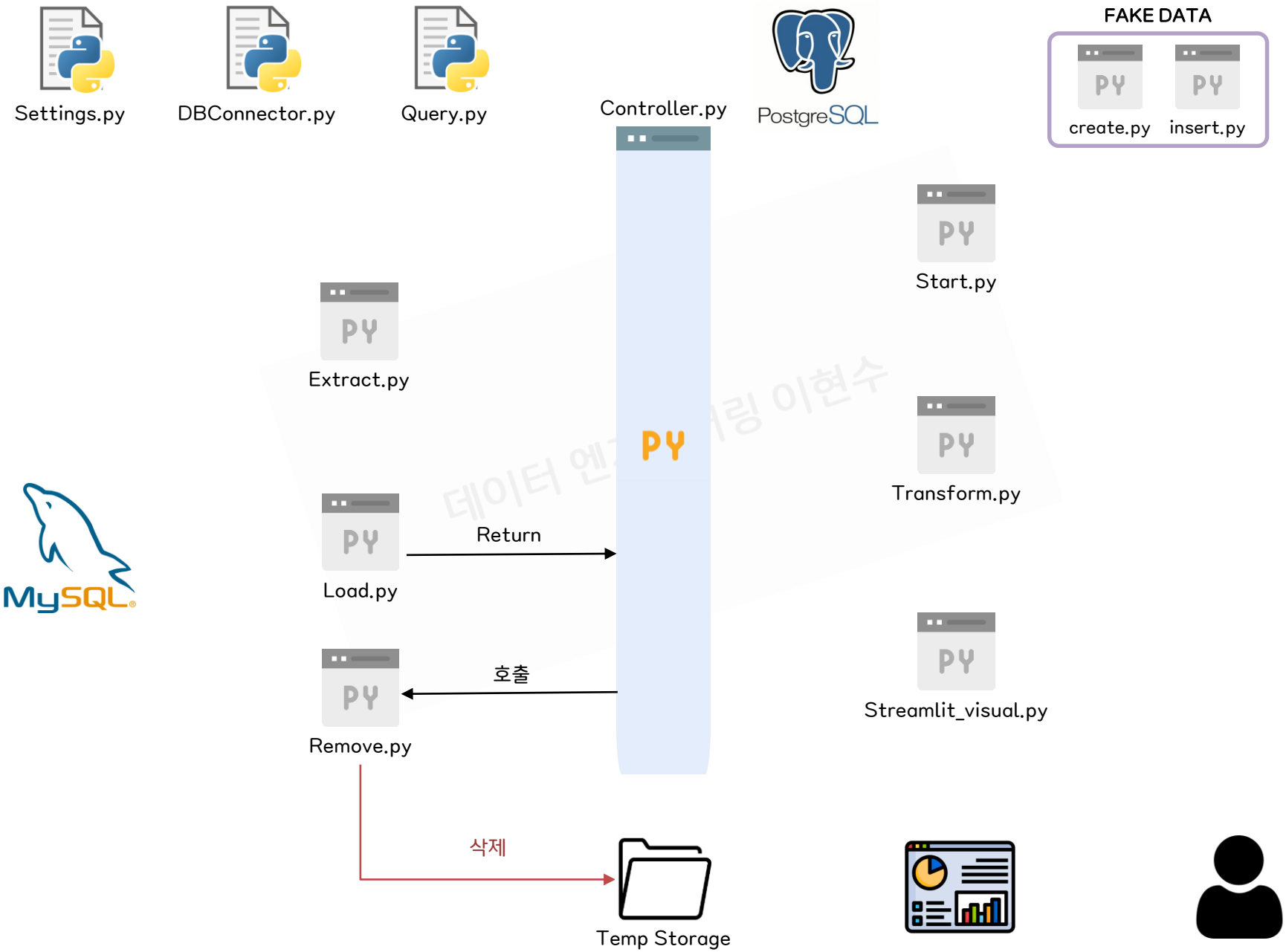
데이터 파이프라인 프로젝트 상세 흐름도



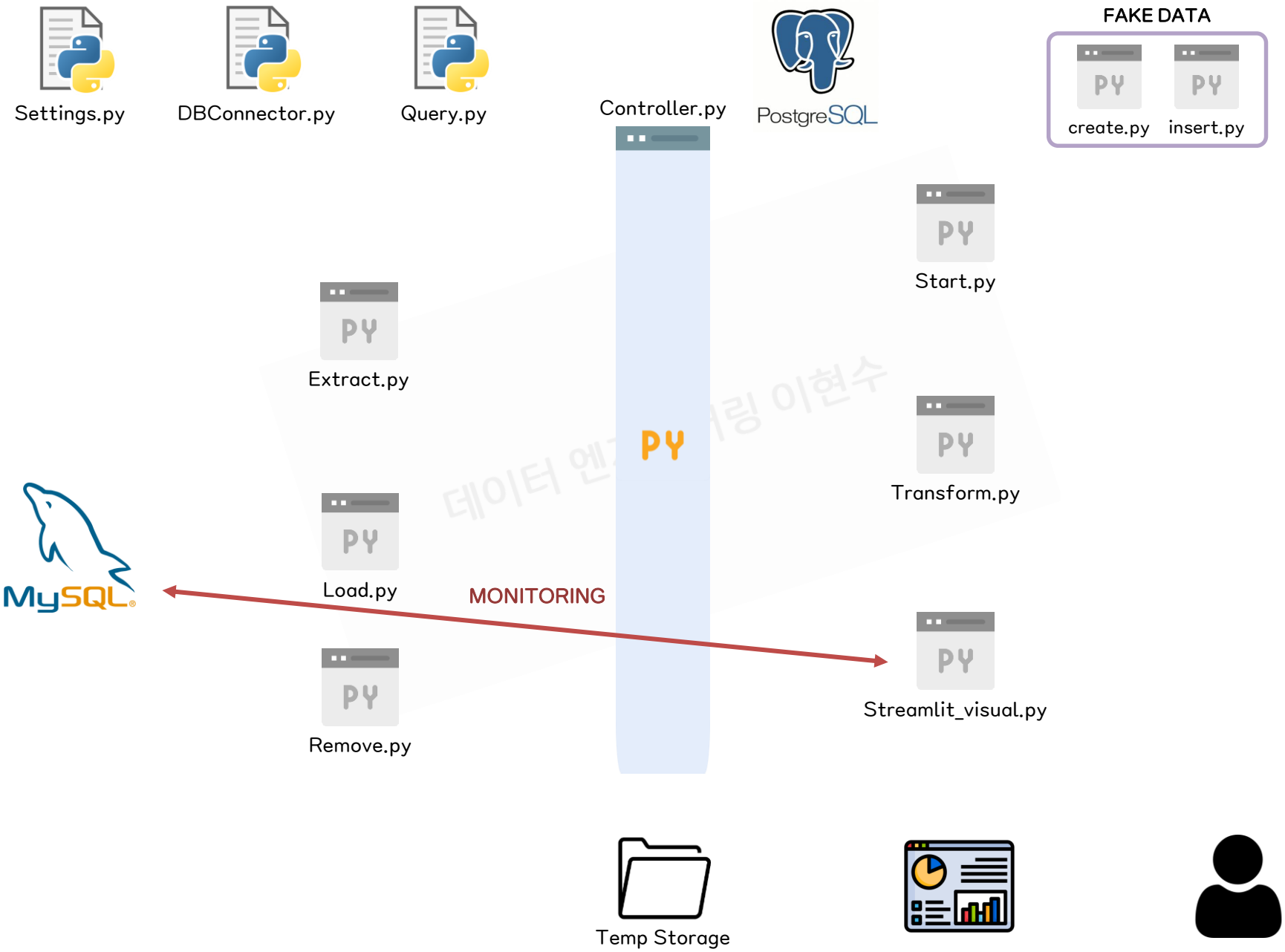
데이터 파이프라인 프로젝트 상세 흐름도



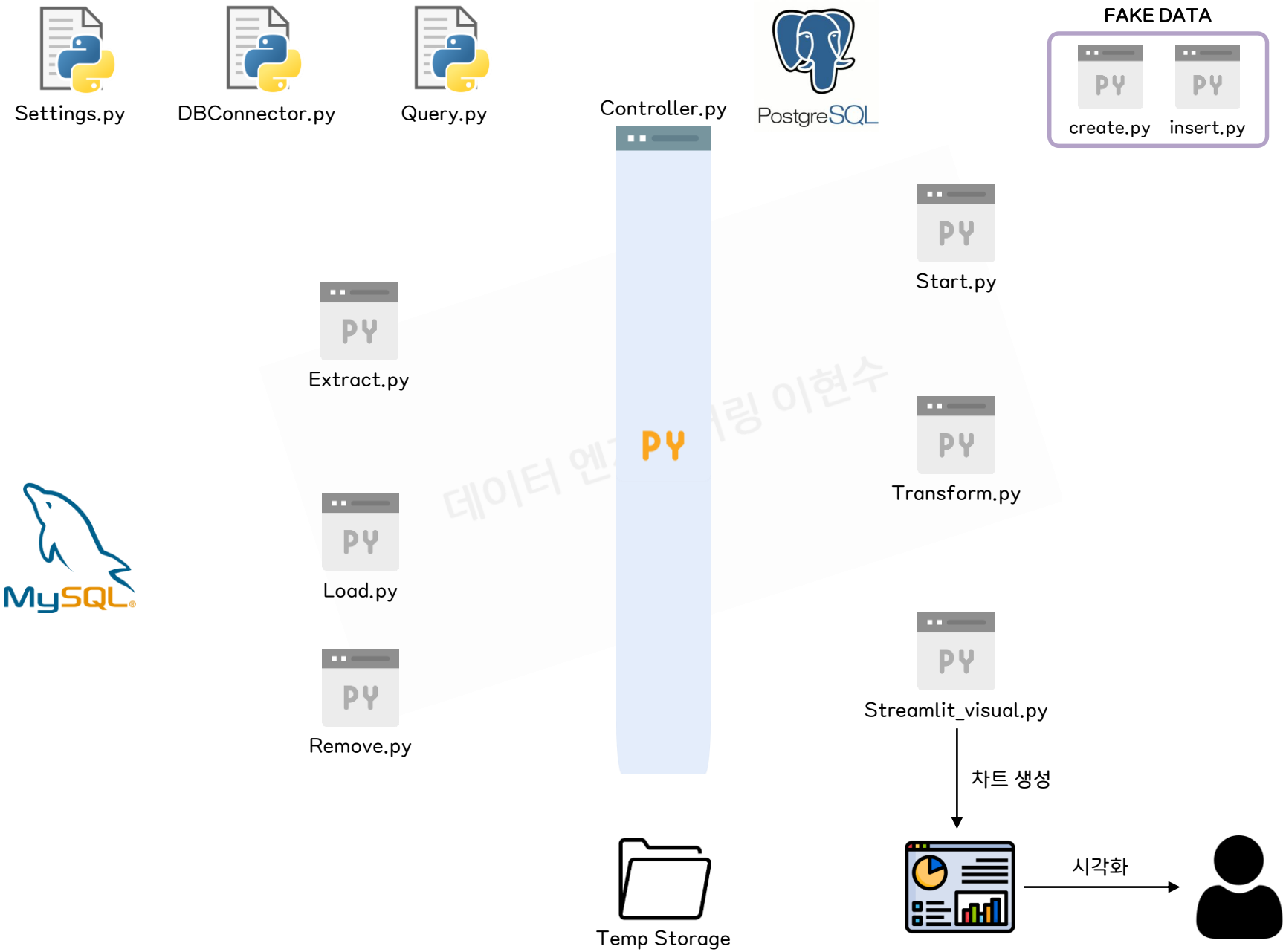
데이터 파이프라인 프로젝트 상세 흐름도



데이터 파이프라인 프로젝트 상세 흐름도

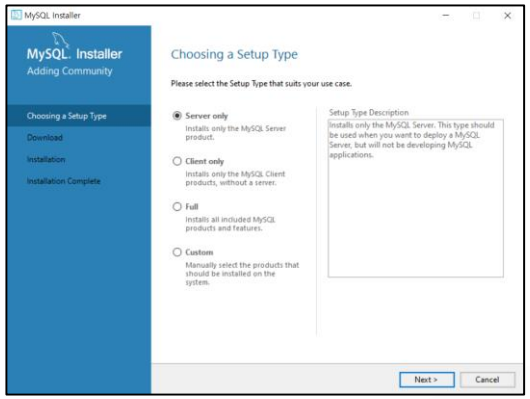


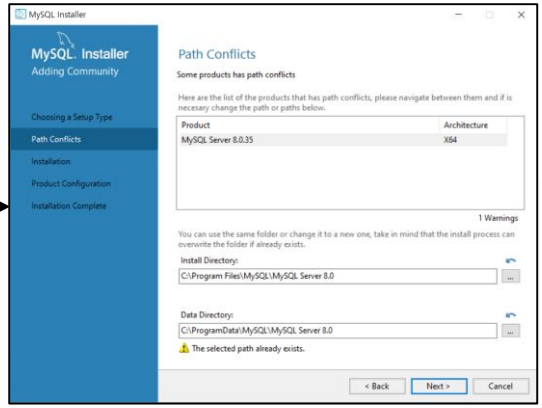
데이터 파이프라인 프로젝트 상세 흐름도

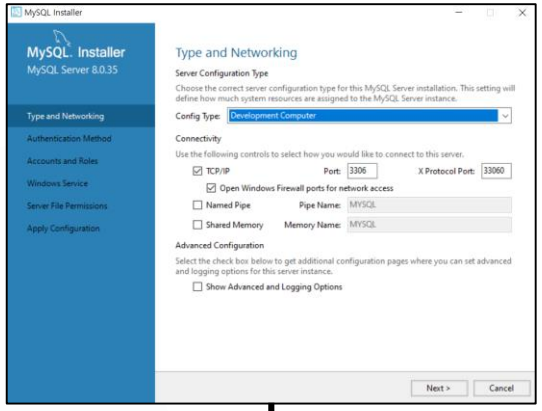


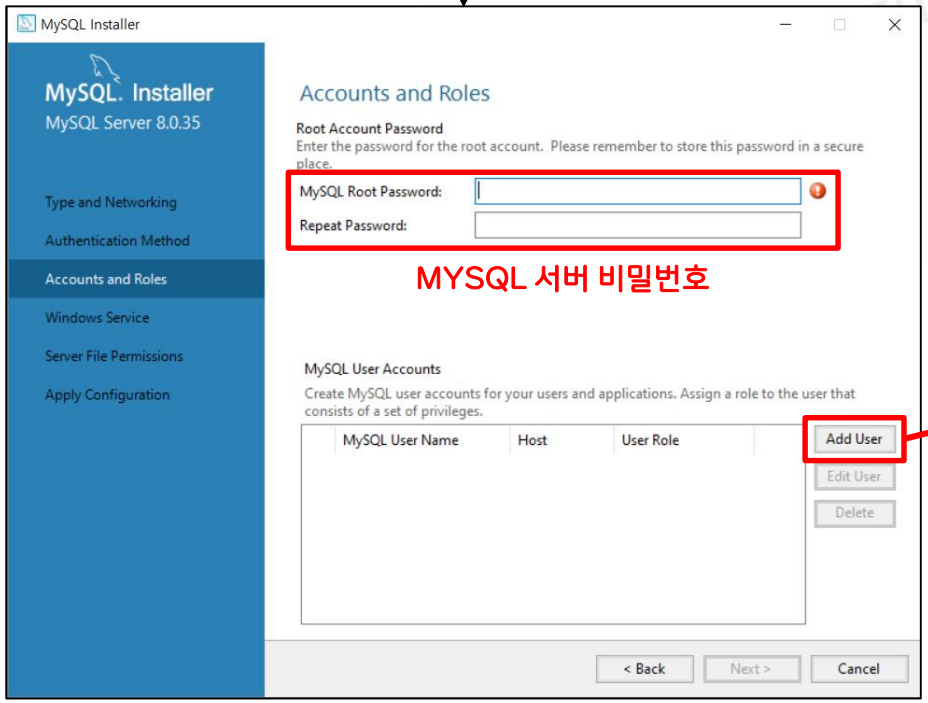
데이터 파이프라인 전체 실습 흐름도

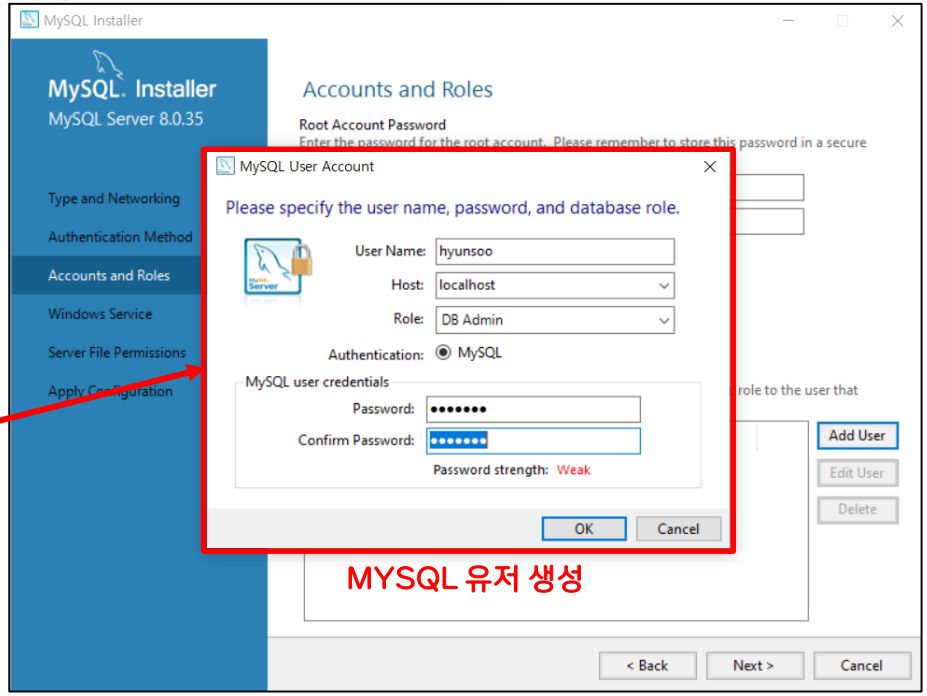
기본 세팅을 유지한 채 계속 NEXT 클릭











데이터 파이프라인 전체 실습 흐름도