



Correctly 마무리

플러스 AI 3기
학습메이트 김경호



목차

1. 개요

2. 환경 설치
자동화

3. 데이터 생성
자동화

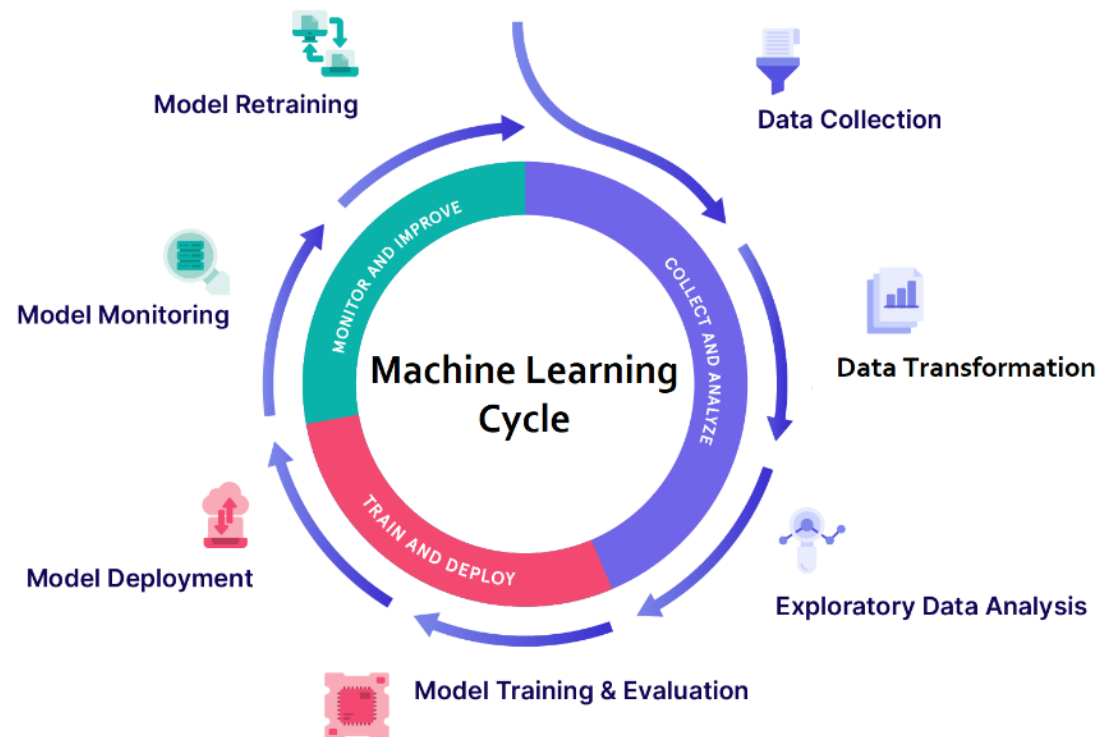
4. 훈련 자동화

5. 난관

6. 결론



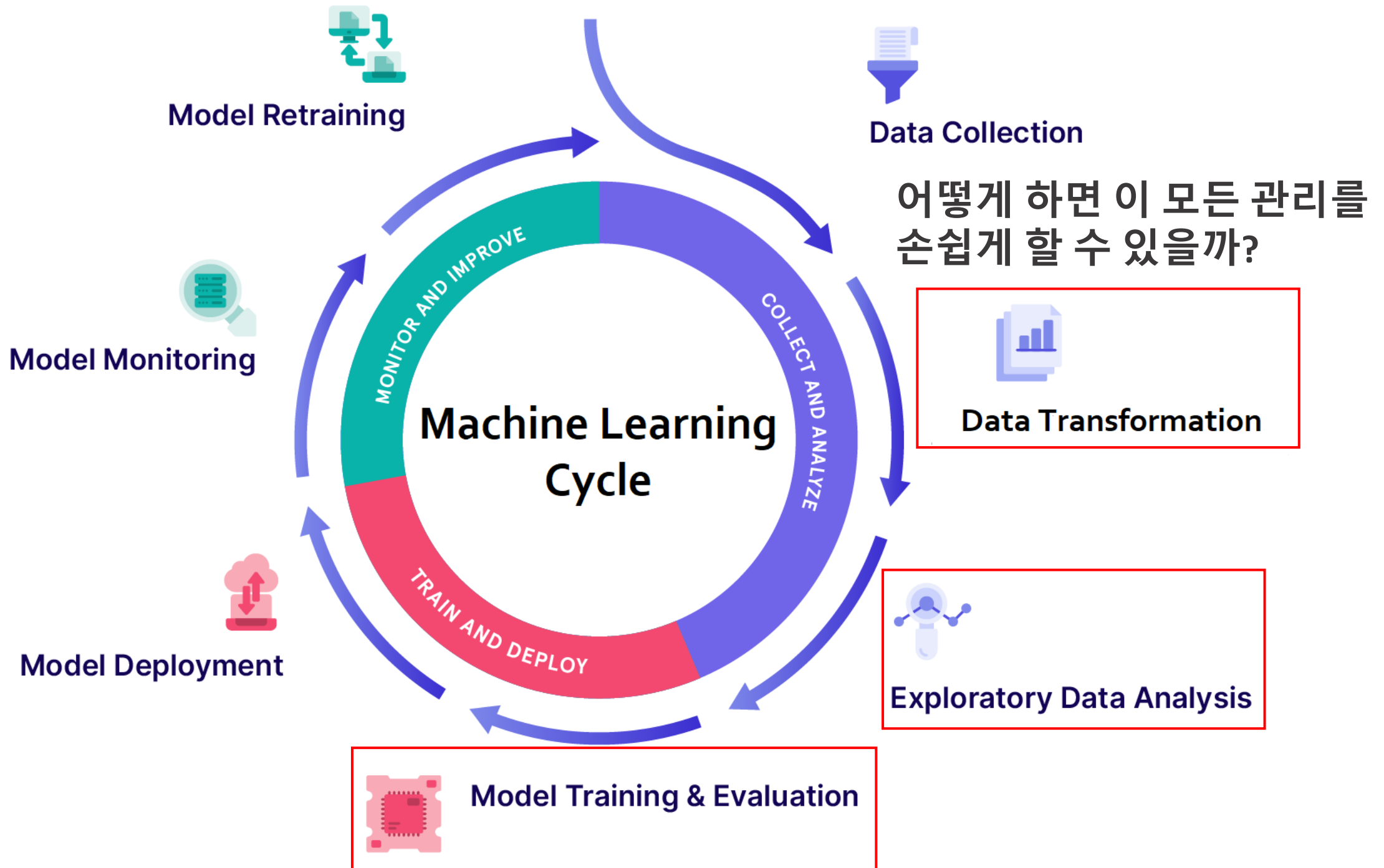
Part 1. 개요



MLcycle: 데이터 수집부터 전처리, 모델 훈련과 배포, 분석에 이르는 하나의 과정

하나의 ML 서비스는 MLcycle을 반복함으로써 개선

정확도와 같은 metric 뿐 아니라 지연 시간과 사용자 경험에 이르는 전체적인 총평을 통하여 ML 기반 서비스를 평가



간단한 데이터 생성 => 모델 훈련
프로젝트

Correctly: 어색한 영어 교정 챗봇



PostgreSQL

mlflow



docker
Compose



Apache
Airflow



Part2.

환경설치 자동화


```
LOG_LEVEL=DEBUG

PROJECT_PATH=/usr/local/project
VENV_PATH=/opt/venv

OPENAI_API_KEY=your_open_api_key

# Autocommit
COMMIT_OPENAI_API_KEY=your_open_api_key
COMMIT_MODEL=gpt-4o-mini

# 포트 설정
AIRFLOW_PORT=8080
MLFLOW_PORT=5000
POSTGRES_PORT=5432

# 네트워크 고정 IP
SUBNET=10.20.0.0/16
POSTGRES_IP=10.20.0.10
AIRFLOW_IP=10.20.0.20
SCHEDULER_IP=10.20.0.21
MLFLOW_IP=10.20.0.30

# Airflow DB 연결
AIRFLOW__CORE__DAGS_FOLDER=/usr/local/project/airflow/dags
AIRFLOW__DATABASE__SQL_ALCHEMY_CONN=postgresql+psycopg2://airflow:airflow@db/airflow
AIRFLOW__CORE__EXECUTOR=LocalExecutor
AIRFLOW__CORE__FERNET_KEY=your_fernet_key
```

모든 정보를 .env에 모아두어 일괄 관리

```
AIRFLOW__CORE__FERNET_KEY=your_fernet_key
AIRFLOW__WEBSERVER__SECRET_KEY=your_webserver_key
AIRFLOW__WEBSERVER__EXPOSE_CONFIG=True
AIRFLOW__CORE__LOAD_EXAMPLES=False
AIRFLOW_ADMIN_USERNAME=admin
AIRFLOW_ADMIN_PASSWORD=admin
AIRFLOW_ADMIN_FIRSTNAME=Air
AIRFLOW_ADMIN_LASTNAME=Flow
AIRFLOW_ADMIN_EMAIL=admin@example.com

# mlflow
MLFLOW_BACKEND_STORE_URI=postgresql://mlflow:mlflow@db/mlflow_db
MLFLOW_DEFAULT_ARTIFACT_ROOT=/mlflow/mlruns

# postgres
POSTGRES_USER=postgres
POSTGRES_PASSWORD=postgres
POSTGRES_DB=postgres
POSTGRES_HOST=db

# ENG_DB
TEXT_GENERATION_NAME=text_generation
TEXT_GENERATION_USER=text_generation
TEXT_GENERATION_PASSWORD=text_generation
INIT=true
INSERT_SAMPLE=true
```



```

1  # ===== Base =====
2  FROM python:3.10-slim AS base
3  ARG PROJECT_PATH
4  ARG VENV_PATH
5  ENV PROJECT_PATH=${PROJECT_PATH}
6  ENV PYTHONPATH=${PROJECT_PATH}
7  ENV VENV_PATH=${VENV_PATH}
8  ENV PYTHONDONTWRITEBYTECODE=1
9  ENV PYTHONUNBUFFERED=1
10 ENV PATH="${VENV_PATH}/bin:$PATH"
11
12 WORKDIR ${PROJECT_PATH}
13
14 RUN apt-get update && apt-get install -y \
15     gcc libc-dev vim curl \
16     && rm -rf /var/lib/apt/lists/*
17
18 # ===== Builder =====
19 FROM base AS builder
20 WORKDIR ${PROJECT_PATH}
21
22 RUN python -m venv ${VENV_PATH}
23 COPY airflow/requirements.txt ./
24 RUN pip install --upgrade pip && \
25     pip install --no-cache-dir -r requirements.txt && \
26     python -c "import flask_session; print('✅ flask_session import OK)'
27

```

DockerFile

Docker Compose

version: "3.8"

▶ Run All Services

services:

▶ Run Service

postgres:

image: pgvector/pgvector:0.8.0-pg17

container_name: postgres

ports:

- "\${POSTGRES_PORT}:5432"

env_file:

- .env

volumes:

- pgdata:/var/lib/postgresql/data

- ./postgres/init.sql:/docker-entrypoint-initdb.d/01_init.sql

- ./postgres/entrypoint.sh:/docker-entrypoint-initdb.d/02_grant.sh

- ./\${PROJECT_PATH}

networks:

train-server:

aliases:

- db

ipv4_address: \${POSTGRES_IP}

▶ Run Service

airflow:

image: custom-airflow:latest

build:

context: .

dockerfile: airflow/Dockerfile

환경 설치 자동화

```
(base) paradeigma@paradeigma:~/workspace/open-webui$ docker compose up -d --build
WARN[0000] /home/paradeigma/workspace/open-webui/compose.yaml: the attribute `version` is obsolete, it will
be ignored, please remove it to avoid potential confusion
[+] Running 9/20
  ⠋ open-webui [#####..] Pulling 39.4s
    ✓ 254e724d7786 Pull complete 7.3s
    ✓ eb0baa05daea Pull complete 7.5s
    ✓ 457229a5b852 Pull complete 8.7s
    ✓ b658f584ba6e Pull complete 8.7s
    ✓ ccb546820a6b Pull complete 8.8s
    ✓ 4f4fb700ef54 Pull complete 8.8s
    ✓ 29444b4738d9 Pull complete 8.8s
    ✓ fe874a226d3a Pull complete 8.9s
    ⠋ eabf08e6753d Downloading 200MB/341.1MB 37.8s
    ✓ 4a19b057bb3f Download complete 4.4s
    ⠋ 0d2e47dd69b3 Downloading 117.3MB/4.803GB 37.8s
    ⠋ 8eb194ba22fd Downloading 68.65MB/90.95MB 37.8s
    ⠋ 3d2fb0291c3c Waiting 37.8s
    ⠋ 93622575a42e Waiting 37.8s
    ⠋ fb7ef2af385e Pulling fs layer 37.8s
  ⠋ watchtower [ ] Pulling 39.3s
    ⠋ 57241801ebfd Waiting 35.7s
    ⠋ 3d4f475b92a2 Waiting 35.7s
    ⠋ 1f05004da6d7 Waiting 35.7s
```

(base) paradeigma@paradeigma:~/workspace/Correctly-mlcycle/src/train\$ docker ps

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS
		NAMES			
4240cbeb47c4	custom-airflow:latest	"airflow/entrypoint..."	3 hours ago	Up 3 hours	8080/tcp
p		airflow_scheduler			
ce089eedfe3d	custom-mlflow:latest	"sh -c 'mlflow serve..."	3 hours ago	Up 3 hours	0.0.0.0
:5000->5000/tcp, [::]:5000->5000/tcp		mlflow			
da29a13ef40a	custom-airflow:latest	"airflow/entrypoint..."	3 hours ago	Up 3 hours	0.0.0.0
:8080->8080/tcp, [::]:8080->8080/tcp		airflow			
0101d4c01e41	pgvector/pgvector:0.8.0-pg17	"docker-entrypoint.s..."	3 hours ago	Up 3 hours	0.0.0.0
:5432->5432/tcp, [::]:5432->5432/tcp		postgres			
1e68d52e6929	moby/buildkit:buildx-stable-1	"buildkitd --allow-i..."	15 hours ago	Up 11 hours	
	buildx buildkit multiplatform-builder0				

환경 설치 자동화

AIRFLOW

Task 자동화 및 관리 툴

AIRFLOW SCHEDULER

AIRFLOW에서의 작업 스케줄링을 담당

MLflow

모델의 레지스트리를 저장 및 성능을 분석

POSTGRESQL
+ pgvector

훈련 데이터 및 추론 데이터 저장

Pgvector를 쓰면 vector 자료형 사용 가능

물리 서버

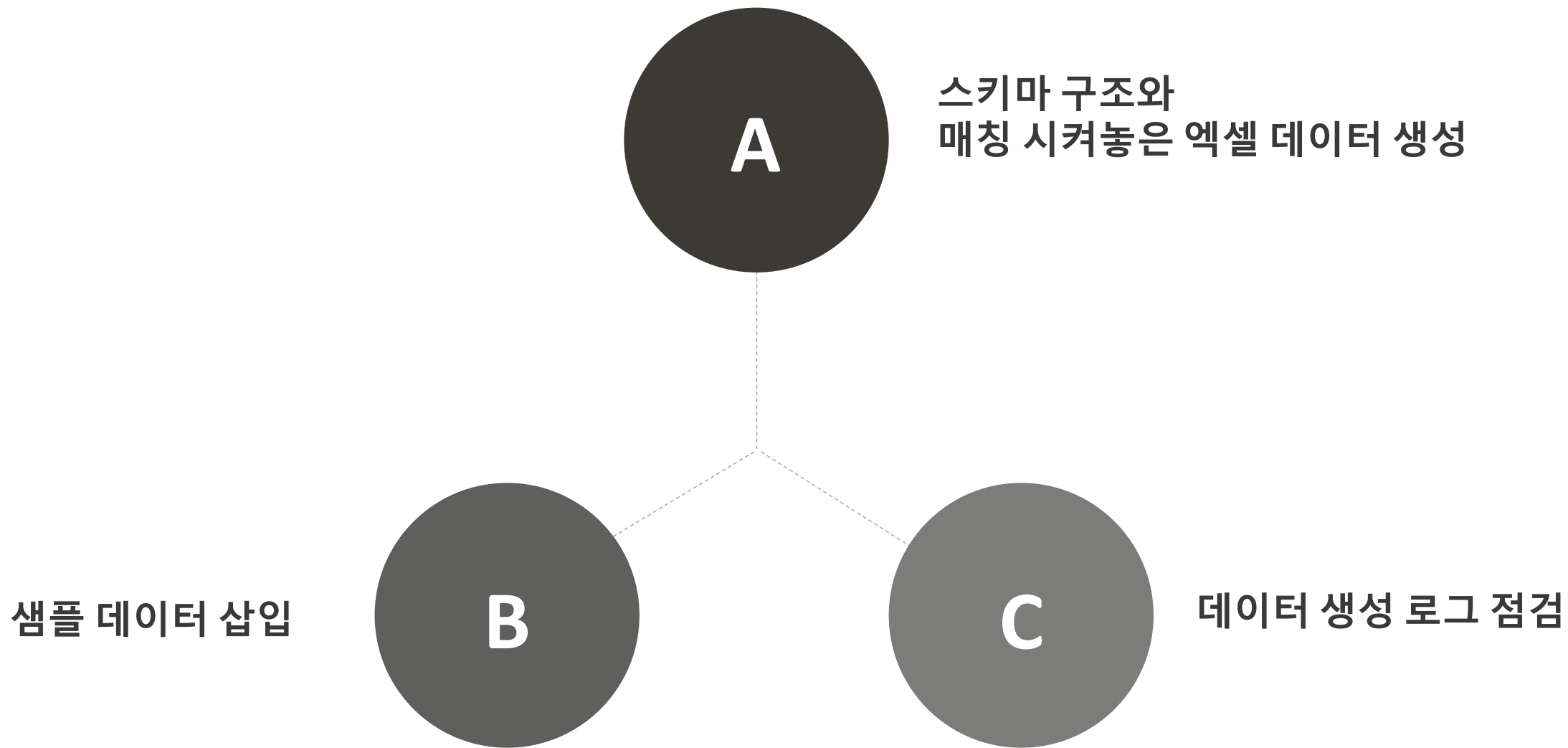
Train Container

훈련 전용 컨테이너(ssh로 연결)



Part 3.

데이터 생성 자동화



교정 전

당신은 영어 문장을 교정하는 선생님입니다.
다음의 어색한 영어 문장을 원어민이 작성한 것처럼 자연스럽게 고급스럽게 교정해주세요.

교정 전

I have a friend who likes a girl but he don't know how to talk to her. He say he is very nervous because he think she is pretty. He always say "Hi" to her but she never looks at him. He ask me for advice but I don't know what to tell him. 그래서 나는 그냥, "Be yourself!" 라고 했어요. 그러나 그는 대답해요, "But what if she don't like me?"

해석

나는 한 친구가 있는데, 그 친구는 한 여자를 좋아하지만 그녀에게 어떻게 말을 걸어야 할지 모른다. 그는 그녀가 예쁘다고 생각하기 때문에 매우 긴장한다고 말한다. 그는 항상 "안녕"이라고 그녀에게 말하지만 그녀는 결코 그를 쳐다보지 않는다. 그는 나에게 조언을 구하지만, 나는 그에게 무엇을 말해야 할지 모르겠다. 그래서 나는 그냥 "너 자신이 되어라!"라고 말했다. 그러나 그는 "그런데 만약 그녀가 나를 좋아하지 않으면 어찌지?"라고 대답한다.

해석

교정 후

교정 후

I have a friend who likes a girl, but he doesn't know how to talk to her. He says he is very nervous because he thinks she is pretty. He always says "Hi" to her, but she never looks at him. He asks me for advice, but I don't know what to tell him. So, I simply said, "Be yourself!" However, he replies, "But what if she doesn't like me?"

교정 지점

1. "but he don't know" → "but he doesn't know": 주어에 맞는 동사 형태로 수정하여 문법적으로 정확하게 만듦.
2. "He say" → "He says": 주어에 맞는 동사 형태로 변경하여 일관성을 유지.
3. "he think" → "he thinks": 주어에 맞는 동사 형태로 수정하여 문법적 오류 제거.
4. "He always say" → "He always says": 주어에 맞는 동사 형태로 변경하여 문법적으로 정확하게 만듦.
5. "never looks at him" → "never looks at him": 동일하나, 문장 연결을 부드럽게 하기 위해 쉼표를 추가.
6. "He ask me for advice" → "He asks me for advice": 주어에 맞는 동사 형태로 수정하여 문법적 오류 제거.
7. "하지만 나는..." → "but I don't know...": 자연스러운 영어 표현으로 변경하여 문맥을 원활하게 만듦.
8. "그래서 나는 그냥, 'Be yourself!'" → "So, I simply said, 'Be yourself!'" : 자연스러운 영어 표현으로 변경하여 문맥을 부드럽게 만듦.
9. "그는 대답해요" → "However, he replies": 자연스러운 표현으로 변경하여 일관성을 유지.
10. "But what if she don't like me?" → "But what if she doesn't like me?": 주어에 맞는 동사 형태로 수정하여 문법적 오류 제거.

교정 포인트

correction_script_type | speaker | style | theme | key_expression

Database Navigator

Filter connections by name

> airflow_db localhost:5432
> mlflow_db localhost:5432
v text_generation localhost:5432

Databases

text_generation

Schemas

correction_script

Tables

> correction_script_type 32K
> example 288K
> generated_text 208K
> instruction_prompt 32K
> key_expression 16K
> speaker 32K
> style 32K
> theme 32K
> train 256K

Foreign Tables

Views

Materialized Views

Indexes

Project - General

Name

DataSource

> Bookmarks
> Dashboards
> Diagrams
> Scripts

generated_text

correction_...

public

instruction...

dag

text_generation

train

Properties

Data

Diagram

Enter a SQL expression to filter results (use Ctrl+Space)

Grid	input	A-Z output	A-Z rejected	123 generated_text_id	create
1	교정 전 ###Young people these da	### 교정 전 ###Young people these da	[NULL]	81	2025-05
2	교정 전 ###Travel is very good for	### 교정 전 ###Travel is very good for	[NULL]	80	2025-05
3	교정 전 ###Travel is very fun, and r	### 교정 전 ###Travel is very fun, and r	[NULL]	79	2025-05
4	교정 전 ###Reading book is very g	### 교정 전 ###Reading book is very g	[NULL]	78	2025-05
5	교정 전 ###Many people is using t	### 교정 전 ###Many people is using t	[NULL]	77	2025-05
6	교정 전 ###Reading habit is very in	### 교정 전 ###Reading habit is very in	[NULL]	76	2025-05
7	교정 전 ###I have a friend who like	### 교정 전 ###I have a friend who like	[NULL]	75	2025-05
8	교정 전 ###Many people are using	### 교정 전 ###Many people are using	[NULL]	74	2025-05
9	교정 전 ###Today I wake up early k	### 교정 전 ###Today I wake up early k	[NULL]	73	2025-05
10	교정 전 ###I have a friend who like	### 교정 전 ###I have a friend who like	[NULL]	72	2025-05
11	교정 전 ###Yesterday I go to the p	### 교정 전 ###Yesterday I go to the p	[NULL]	71	2025-05
12	교정 전 ###My parents always say	### 교정 전 ###My parents always say	[NULL]	70	2025-05
13	교정 전 ###My parents always say	### 교정 전 ###My parents always say	[NULL]	69	2025-05
14	교정 전 ###Young people these da	### 교정 전 ###Young people these da	[NULL]	68	2025-05
15	교정 전 ###Young people these da	### 교정 전 ###Young people these da	[NULL]	67	2025-05
16	교정 전 ###Travel is very fun and r	### 교정 전 ###Travel is very fun and r	[NULL]	66	2025-05
17	교정 전 ###Travel is very exciting, a	### 교정 전 ###Travel is very exciting, a	[NULL]	65	2025-05
18	교정 전 ###Reading book is very ir	### 교정 전 ###Reading book is very ir	[NULL]	64	2025-05
19	교정 전 ###Many people using the	### 교정 전 ###Many people using the	[NULL]	63	2025-05
20	교정 전 ###Reading habit is very in	### 교정 전 ###Reading habit is very in	[NULL]	62	2025-05
21	교정 전 ###I have a friend who like	### 교정 전 ###I have a friend who like	[NULL]	61	2025-05
22	교정 전 ###Many people are using	### 교정 전 ###Many people are using	[NULL]	60	2025-05

Refresh

Save

Cancel

Export data

200

81

DAGs

All 3Active 3Paused 0

Running 1Failed 0

Filter DAGs by tag

Search DAGs

Auto-refresh

<div><div></div></div> DAG <div></div>	Owner <div></div>	Runs <div></div>	Schedule	Last Run <div></div> <div></div>	Next Run <div></div> <div></div>	Recent Tasks <div></div>
<div><div></div></div> <div>correction_script_example</div> <div>llmtestyaml</div>	airflow	<div><div></div><div>2</div><div>3</div><div></div></div>	None <div></div>	2025-05-23, 15:35:54 <div></div>		<div><div></div><div></div><div></div><div></div><div>3</div><div>6</div><div></div><div></div></div>
<div><div></div></div> <div>correction_script_generate_text</div> <div>llmtestyaml</div>	airflow	<div><div></div><div>3</div><div></div><div>3</div></div>	None <div></div>	2025-05-23, 12:50:36 <div></div>		<div><div></div><div></div><div></div><div></div><div></div><div>3</div><div></div><div></div></div>
<div><div></div></div> <div>correction_script_make_train_data</div> <div>llmtestyaml</div>	airflow	<div><div></div><div>1</div><div></div><div></div></div>	None <div></div>	2025-05-23, 12:56:34 <div></div>		<div><div></div><div></div><div></div><div></div><div></div><div>2</div><div></div><div></div></div>

Part 3 >>

데이터 생성 자동화

DAG: correction_script_example

1) YAML 로드 → 2) LLM 테스트 → 3) 예시 생성

Schedule: None

Next Run ID: None

05/23/2025

03:35:55 PM

All Run Types

All Run States

Clear Filters

Auto-refresh

25

Press shift + / for Shortcuts

deferred

failed

queued

removed

restarting

running

scheduled

shutdown

skipped

success

up_for_reschedule

up_for_retry

upstream_failed

no_status

Duration

00:19:52

00:09:56

00:00:00

May 23, 12:50

prepare_generation_data

test_llm_generation

create_example

DAG

Run

correction_script_example / ▶ 2025-05-23, 15:35:54 UTC

Clear

Mark state as...

Details

Graph

Gantt

Code

Event Log

DAG Run Notes

Add Note

Dag Run Details

Status

running

```
[2025-05-23, 15:36:15 UTC] {_client.py:1025} INFO - HTTP Request: POST https://api.openai.com/v1/chat/completions "HTTP/1.1 200 OK"
[2025-05-23, 15:36:15 UTC] {logging_mixin.py:190} INFO - [2025-05-23 15:36:15] DEBUG | [Apply Request] Raw Output: ChatCompletion(i
[2025-05-23, 15:36:15 UTC] {logging_mixin.py:190} INFO - [2025-05-23 15:36:15] INFO | [Apply Request] Assistant: ### 교정 전 ###
Today, I go to school and I very happy. My friend, he is always make joke, and we laugh a lot. 수업에서, the teacher tell us about hi
#####
### 교정 후 ###
Today, I went to school and I was very happy. My friend always makes jokes, and we laugh a lot. In class, the teacher told us about
### 해석 ###
오늘 나는 학교에 갔고 매우 행복했다. 내 친구는 항상 농담을 하고, 우리는 많이 웃었다. 수업에서는 선생님이 역사에 대해 말씀해 주셨다. 가끔 지루할 수 있지만
### 교정 지점 ###
1. **시제 일치**: "go" → "went", "is" → "was"로 변경하여 과거의 일상을 표현했습니다.
2. **주어-동사 일치**: "he is always make joke" → "always makes jokes"로 수정하여 주어와 동사의 일치를 맞췄습니다.
3. **단수/복수 형태**: "joke" → "jokes"로 변경하여 복수형을 사용했습니다.
4. **어순 수정**: "the teacher tell us about history" → "the teacher told us about history"로 시제와 어순을 올바르게 조정했습니다.
5. **불규칙 동사 수정**: "try" → "tried"로 수정하여 과거 시제를 유지했습니다.
6. **관사 사용**: "with my 친구들" → "with my friends"로 한국어를 영어로 바꾸어 자연스럽게 표현했습니다.
7. **과거 시제 사용**: "is very fun" → "was very fun"으로 과거의 즐거운 경험을 나타냈습니다.
8. **부사 사용**: "I am tired" → "I was tired"로 과거 시제에 맞춰 수정했습니다.
9. **구어체 표현 수정**: "집에 가면" → "When I got home"으로 자연스러운 영어 표현으로 바꾸었습니다.
```




Part4. 후련 자동화

훈련 자동화

AIRFLOW 컨테이너

외부 서버(A100)

데이터
생성

데이터
검증

훈련 데
이터 전
처리

SSH 통신
명령

언어모델
훈련 전용
외부 GPU

```

t1 = PythonOperator(
    task_id="prepare_generation_data",
    python_callable=prepare_generation_data,
    provide_context=True
)

t2 = PythonOperator(
    task_id="write_train_data",
    python_callable=write_train_data,
    provide_context=True
)

```

```

t2 = SSHOperator(
    task_id="model_train",
    ssh_hook=ssh_hook,
    command="""
bash -c '
export PYTHONPATH={{ params.remote_workdir }} && \
cd {{ params.remote_workdir }} && \
nohup /venv/main/bin/python train/train_sft.py \
--train_json data/{{ params.id }}/{{ params.train_data }} \
--test_json data/{{ params.id }}/{{ params.val_data }} \
--report_json data/{{ params.id }}/report.json \
--model_name "{{ params.model }}" \
--output_dir data/{{ params.id }}/model \
--batch_size {{ params.batch_size }} \
--num_epochs {{ params.num_epoch }} \
--save_total_limit {{ params.save_total_limit }} \
> data/{{ params.id }}/train.log 2>&1 &

echo "✅ Training started in background"
'
""",
    do_xcom_push=False,
)

```

PythonOperator와
SSHOperator 차이?

항목	PythonOperator	SSHOperator
실행 환경	현재 Airflow 환경 (컨테이너 or 호스트 내 Python)	외부 서버의 Python 환경 (보통 SSH를 통한 리눅스 서버)
사용 목적	간단한 Python 함수 실행	외부 서버에서 script 실행 (GPU 훈련, 리소스 분리 등)
의존성 관리	Airflow Python env 와 공유	원격 서버에 사전 구성된 환경 사용 필요
예시	전처리, 간단한 DB 작업	모델 학습, 훈련 스크립트 실행, 대규모 연산
구현 복잡도	쉬움	SSH 키/포트 설정, 보안 설정 필요 등 상대적으로 복잡
성능/리소스 분리	Airflow Worker에 의존	외부 서버 리소스 사용 (Airflow 자원 과 완전 분리, GPU 사용 가능)

KK

Kim Kyungh...
Individual

Search

Templates

Instances

1

- Account
- Billing
- Earnings
- Members
- Audit Logs
- Keys
- Settings

GPU Name

Verification Status

Sort Options

All

All

CLEAR FILTERS

All

20498687

m:30416

datacenter:125728

Czechia, CZ

80.188.223.202:14041-14917

verified

vast.ai

1x A100 SXM4

15.6 TFLOPS

Max CUDA: 12.6

0WGY16

PCIe 4.0, 16x 13.5 GB/s

AMD EPYC 7513 ...

32.0/128 cpu ?/0 GB

3961.5 Mbps

12038.5 Mbps

SAMSUNG MZW...

2766 MB/s 9.0/130.0 GB

Age: 1h 7m

Remaining: 4 mon.

OPEN

\$1.239/hr

GPU: 0% 27C , CPU: 0%

Status: success, running vastai/base-image_cuda-12.6.3-auto/jupyter

©Saebyeol Yu. Saebyeol's PowerPoint

An aerial, high-angle photograph of a busy pedestrian crossing. Numerous people of various ages and ethnicities are walking across the street, which is marked with white zebra stripes. Long, dark shadows are cast across the pavement, indicating it is either early morning or late afternoon. The overall scene is one of a crowded, active urban environment.

Part 5.

관

한계(1): 어떤 Operator를 쓸 것인가

가정: 하나의 물리 머신

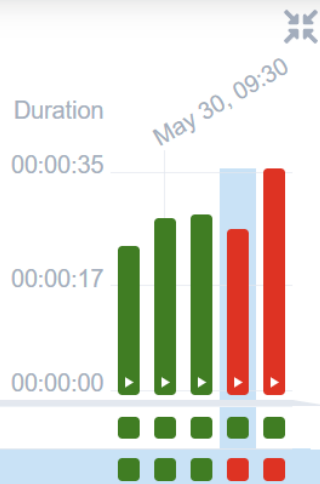
문제1: Airflow Container에서 GPU?

문제2: SSH operator로 우회

불안정한 통신 환경

Airflow가 컨테이너 환경에서 작동하다 보니, 훈련에 필요한 GPU 자원을 직접 활용하는 데 한계가 있었습니다. 특히 PyTorch, CUDA 등 무거운 라이브러리를 컨테이너에 통합하는 작업은 비효율적이었고, 로컬 머신의 8GB VRAM으로는 대규모 언어 모델 훈련을 감당하기 어려웠습니다.

이를 해결하기 위해 외부 GPU 서버를 임대하고, Airflow에서 SSHOperator를 통해 훈련을 원격 실행하는 방식으로 우회했습니다. 그러나 이 방식은 SSH 연결이 불안정하거나 세션이 끊길 경우 훈련 자체가 실패하는 등 통신 안정성 문제에 취약했습니다.



DAG

Run

Task

correction_script_model_train / 2025-05-30, 09:37:35 KST / model_train

Details

Graph

Gantt

Code

Event Log

Logs

XCom

Task Duration

All Levels

All File Sources

Log message source details

[2025-05-30, 09:37:58 KST] {local_task_job_runner.py:123} Pre task execution logs

[2025-05-30, 09:37:58 KST] {ssh.py:285} WARNING - No Host Key Verification. This won't protect against Man-In-The-Middle attacks

[2025-05-30, 09:37:59 KST] {transport.py:1944} INFO - Connected (version 2.0, client OpenSSH_9.6p1)

[2025-05-30, 09:38:00 KST] {transport.py:1944} INFO - Auth banner: b'Welcome to vast.ai. If authentication fails, try again after 30 seconds'

[2025-05-30, 09:38:00 KST] {transport.py:1944} INFO - Authentication (publickey) successful!

[2025-05-30, 09:38:00 KST] {ssh.py:431} INFO - Running command:

export PYTHONPATH=/workspace/Correctly-mlcycle && cd /workspace/Correctly-mlcycle && /venv/main/bin/python

[2025-05-30, 09:38:06 KST] {taskinstance.py:3313} ERROR - Task failed with exception

Traceback (most recent call last):

File "/opt/venv/lib/python3.10/site-packages/airflow/models/taskinstance.py", line 763, in _execute_task

result = _execute_callable(context=context, **execute_callable_kwargs)

File "/opt/venv/lib/python3.10/site-packages/airflow/models/taskinstance.py", line 734, in _execute_callable

return ExecutionCallableRunner(

File "/opt/venv/lib/python3.10/site-packages/airflow/utils/operator_helpers.py", line 252, in run

return self.func(*args, **kwargs)

File "/opt/venv/lib/python3.10/site-packages/airflow/models/baseoperator.py", line 424, in wrapper

return func(self, *args, **kwargs)

File "/opt/venv/lib/python3.10/site-packages/airflow/providers/ssh/operators/ssh.py", line 173, in execute

result = self.run_ssh_client_command(ssh_client, self.command, context=context)

File "/opt/venv/lib/python3.10/site-packages/airflow/providers/ssh/operators/ssh.py", line 161, in run_ssh_client_command

self.raise_for_status(exit_status, agg_stderr, context=context)

File "/opt/venv/lib/python3.10/site-packages/airflow/providers/ssh/operators/ssh.py", line 155, in raise_for_status

raise AirflowException(f"SSH operator error: exit status = {exit_status}")

airflow.exceptions.AirflowException: SSH operator error: exit status = 1

[2025-05-30, 09:38:06 KST] {taskinstance.py:1226} INFO - Marking task as FAILED. dag_id=correction_script_model_train, task_id=model_train

[2025-05-30, 09:38:06 KST] {taskinstance.py:341} Post task execution logs

한계(2): SSH 통신의 불안정성

통신 방식	특징	주요 용도	실시간 성	보안성	주석
SSH	명령어 기반 원격 접속, 터미널 환경에 적합	서버 제어, 원격 코드 실행, 파일 관리	낮음	높음	전통적인 CLI 제어
HTTP	요청-응답 기반, 단방향 전송	REST API, 웹 서비스	낮음	중간	브라우저/앱 통신
WebSoc ket	양방향 통신 가능, 지속 연결 유지	채팅, 실시간 데이터 스트리밍	높음	중간	연결 유지 비용 있음
gRPC	HTTP/2 기반 바이너리 통신, 빠른 직렬화	마이크로서비스 간 통신, 고성능 API	높음	높음	양쪽 모두 protobuf 필요
FTP/SFTP	파일 전송 전용 (SFTP는 SSH 기반)	대용량 파일 업로드/다운로드	낮음	SFTP는 높음	FTP는 보안 취약

1. SSH는 단방향 비저장성 통신 방식입니다.

2. 그러나 훈련이 잘 되고 있는지 수시로 점검해야하고 통신이 끊겼을 때 다시 연결을 하는등의 로직이 필요합니다.

3. 이러한 특성은 ssh 통신 방식에 부적합합니다.

SSH의 부적합성

한계(3): 제한된 서버 환경과 비용

항목	완전한 IaaS (예: AWS EC2, GCP VM)	제한된 IaaS (예: Vast.ai, RunPod)
루트 권한	✓ 가능 (sudo 포함)	⚠ 일부 환경에서는 제한됨 또는 가짜 루트
Docker 설치	✓ 가능	⚠ 이미 설치되어 있고, 제어 불가능한 경우 많음
Docker socket 공유	✓ /var/run/docker.sock 사용 가능	✗ 기본적으로 불가능 또는 보안상 차단됨
DinD (Docker in Docker)	✓ 지원	✗ 대부분 불가능
커널 모듈 제어	✓ 가능 (커스텀 드라이버 가능)	✗ 제한됨
GPU 독점 여부	✓ 가능 (VM 단독 사용)	⚠ 공유일 가능성 있음
적합 용도	K8s 노드, CI/CD, ML 플랫폼 구축	빠른 훈련 실행, 저렴한 추론 작업

한계(3): 제한된 서버 환경과 비용

머신 구성

이름 *
instance-20250530-013531

리전 *
us-central1 (아이오와)

영역 *
모두

리전은 영구적입니다.

Google에서 VM 확보 가능성을 극대화하는 영역을 대신 선택합니다. 영역은 영구적입니다.

신규: 범용 C4D 머신 시리즈(프리뷰)

✓ 까다로운 범용 워크로드를 최적화하기 위해 뛰어난 가격 대비 성능과 고급 기능을 갖춘 [지금 사용해 보기](#)

- 범용
- 컴퓨팅 최적화
- 메모리 최적화
- 스토리지 최적화
- ✓ GPU

그래픽 처리 장치(GPU)는 인스턴스에서 머신러닝 및 데이터 처리와 같은 특정 워크로드를 가속화합니다. [Learn More](#)

GPU 유형
NVIDIA A100 80GB

GPU 수
1

☐ 가상 워크스테이션 사용 설정(NVIDIA GRID)

us-central1 할당량	현재 사용량/한도	한도 필요
NVIDIA A100 80GB GPUs	0/0	1

[Request quota adjustment](#)

간략히

월별 예상 가격

US\$3,701.22
시간당 약 US\$5.07

사용한 만큼만 비용 지불: 선불 비용 없이 초당 청구

항목	월별 예상 가격
12 vCPU + 170 GB memory	US\$802.72
1 NVIDIA A100 80GB	US\$2,867.50
375GiB 로컬 SSD 디스크	US\$30.00
10GB 분산된 영구 디스크	US\$1.00
스냅샷 일정	다양한 비용
Total	US\$3,701.22

한계(4): MLFLOW와의 연결

네트워크 드라이버	동작 방식	특성	사용 시점	주의사항
bridge (기본)	컨테이너 간 가상 네트워크	격리된 환경, 포트 매핑 필요	단일 머신에서 내부 서비스 간 연결	외부 접속은 직접 불가
host	호스트 네트워크 직접 사용	빠른 통신, 포트 충돌 주의	퍼포먼스 중시, Linux 전용	Windows/WSL2 불가
macvlan	고유 IP 부여	외부 장비와 직접 통신	NAS, IoT 기기 등 연동 시	구성 복잡, 라우팅 주의
overlay	멀티 노드 통신	Docker Swarm 기반	클러스터 네트워크 구성	Compose 단독 사용 불가
none	네트워크 없음	완전 격리	보안 실험, 디버깅	네트워크 기능 사용 불가

1. MLFLOW를 외부 GPU 서버에서 사용하려면 컨테이너가 외부에서 직접 접근 가능해야 한다.

2. 그러나 bridge 네트워크는 노드에서 직접 접근이 불가능해 별도 포트 매핑이 필요합니다.

Bridge 네트워크의 부적합성

An aerial, high-angle photograph of a very busy pedestrian crossing. The crossing is marked with wide white stripes on a dark asphalt surface. Hundreds of people of various ages and ethnicities are walking across the street in different directions. Long, dark shadows are cast by the pedestrians, indicating it is either early morning or late afternoon. The overall scene conveys a sense of a high-traffic, urban environment.

Part6.

기록

결론(1): 스키마 전략을 단순하게 MongoDB로

1. RDB로 하려니 관계 설정과 스키마 전략을 구성하는데 너무 오래걸렸습니다.
2. 또한 스키마가 자주 변경될 가능성이 높기 때문에 MongoDB를 사용하고 벡터는 Chorma와 같은 Vector DB를 따로 쓰는 게 좋을 것입니다.

Postgresql+Pgvector=>MongoDB+VectorDB



결론(2): 자동화를 만드는 것은 Airflow가 아니라 스크립트 단위 관리

1. Airflow를 구성하려면 불가분 모듈화가 잘 된 스크립트가 생명이었다.
2. 따라서 스크립트 단위에서 모듈화를 못하면 Airflow에서 자동화가 될 것이라는 것도 허상이었다.
3. 스크립트 관리 만으로도 어플리케이션 레벨에서 자동화를 잘 할 수있으므로 Airflow 자체도 사용 여부 자체를 재고

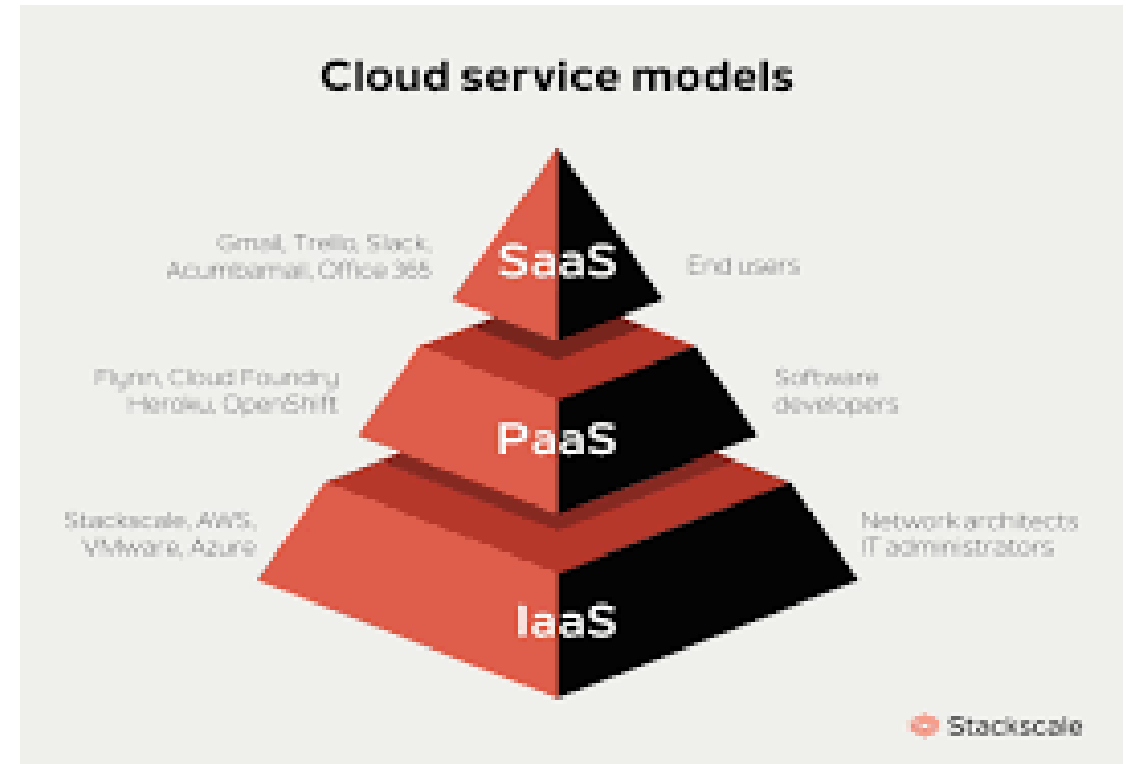
스크립트관리우선!



결론(3): Vast.ai 환경에서 훈련 자동화는 구조적으로 한계

1. Airflow를 구성한다 하더라도 외부 GPU 환경에서 훈련하는 것까지 자동화하긴 어렵다.

훈련은 반자동화



Q&A