

## SK네트웍스 Family AI과정 5기

# 모델링 및 평가 시스템 아키텍처

### □ 개요

- 산출물 단계 : 모델링 및 평가
- 평가 산출물 : 시스템 아키텍처
- 제출 일자 : 2025. 02. 11.
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN05-final-3Team>
- 작성 팀원 : 허상호

컴포넌트 다이어그램	<ul style="list-style-type: none"><li>• 구성 요소</li><li>• 설명</li></ul>
시퀀스 다이어그램	<ul style="list-style-type: none"><li>• 참여자 (Actors)</li><li>• 주요 흐름</li></ul>
액티비티 다이어그램	<ul style="list-style-type: none"><li>• 구성</li><li>• 주요 액션 노드</li></ul>

## 1. 컴포넌트 다이어그램

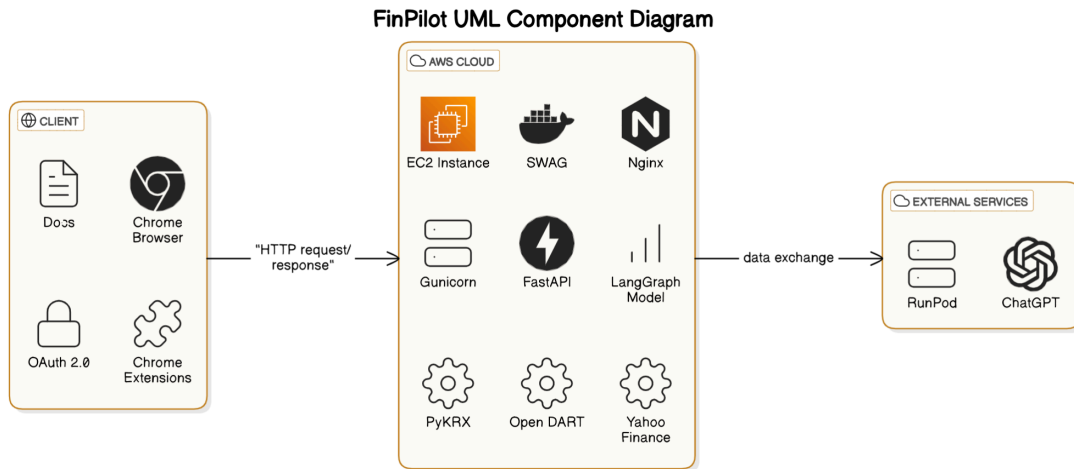


Figure 1. 컴포넌트 다이어그램

### 구성 요소

#### 1. 클라이언트(Client):

##### - 컴포넌트:

- Chrome 브라우저
- Docs
- Chrome Extensions

##### - 인터페이스:

- Google Cloud OAuth 2.0 (인증)
- JSON 데이터 전송 (HTTPS)

#### 2. AWS Cloud:

##### - EC2 인스턴스:

- SWAG (Secure Web Application Gateway)
- Nginx (리버스 프록시 및 정적 파일 제공)
- Gunicorn (WSGI 서버)
- FastAPI (백엔드 애플리케이션)

##### - LangGraph 모델:

- PyKRX, Open DART, Yahoo Finance (외부 데이터 소스)
- Chroma
- Prompt Engineering (AI 모델 최적화)

#### 3. 데이터베이스:

- Amazon RDS (MySQL)

#### 4. 외부 서비스:

- RunPod (sLLM 실행 환경)

## - ChatGPT (OpenAI 언어 모델)

### 설명

- 클라이언트는 Google Cloud OAuth 2.0을 통해 인증을 수행하고, JSON 데이터를 HTTPS를 통해 AWS Cloud로 전송.
- AWS Cloud의 EC2 인스턴스는 요청을 처리하며, Nginx와 Gunicorn을 통해 FastAPI 애플리케이션이 실행.
- LangGraph 모델은 외부 데이터 소스(PyKRX, Open DART 등)와 통합되며, AI 관련 요청은 RunPod(sLLM) 또는 ChatGPT로 전달.
- 최종적으로 처리된 데이터는 클라이언트로 반환.

## 2. 시퀀스 다이어그램

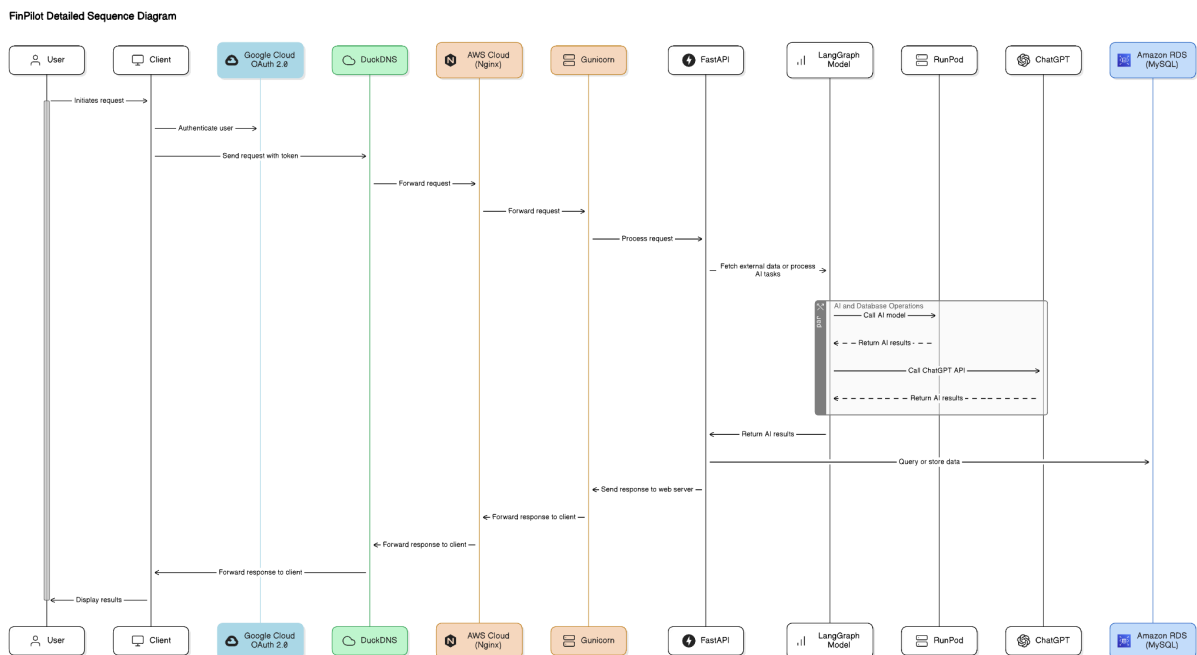


Figure 2. 시퀀스 다이어그램 | 별첨1 사진 참조

### 참여자 (Actors)

1. User: 클라이언트(Chrome 브라우저, Docs, Chrome Extensions)를 통해 요청을 시작하는 사용자.
2. Google Cloud OAuth 2.0: 인증 및 권한 부여를 처리.
3. DuckDNS: 클라이언트 요청을 AWS 서버로 라우팅.
4. AWS Cloud:
  - Nginx: 요청을 수신하고 리버스 프록시 역할 수행.
  - Gunicorn: FastAPI 애플리케이션 실행.

- FastAPI: 요청 처리 및 비즈니스 로직 실행.
  - LangGraph 모델: 외부 데이터 및 AI 모델 호출.
  - Amazon RDS (MySQL): 데이터베이스 작업 수행.
5. RunPod / ChatGPT: AI 모델(sLLM) 또는 ChatGPT API 호출.

## 주요 흐름

1. 사용자가 클라이언트(브라우저 또는 확장 프로그램)를 통해 요청을 시작.
2. Google Cloud OAuth 2.0에서 사용자를 인증하고 액세스 토큰을 반환.
3. DuckDNS는 클라이언트의 요청을 AWS 서버로 전달.
4. AWS 클라우드 내에서 Nginx가 요청을 수신하고 Gunicorn 및 FastAPI로 전달.
5. FastAPI는 LangGraph 모델과 통신하여 필요한 데이터를 조회하거나 AI 작업 수행.
6. LangGraph 모델은 외부 데이터 소스(PvKRX, Open DART, Yahoo Finance 등) 또는 Amazon RDS(MySQL)와 상호작용.
7. AI 관련 요청은 RunPod(sLLM) 또는 ChatGPT로 전달.
8. 결과가 반환되고, FastAPI를 통해 응답이 클라이언트로 전달.
9. 최종적으로 사용자는 결과를 확인.

## 3. 액티비티 다이어그램

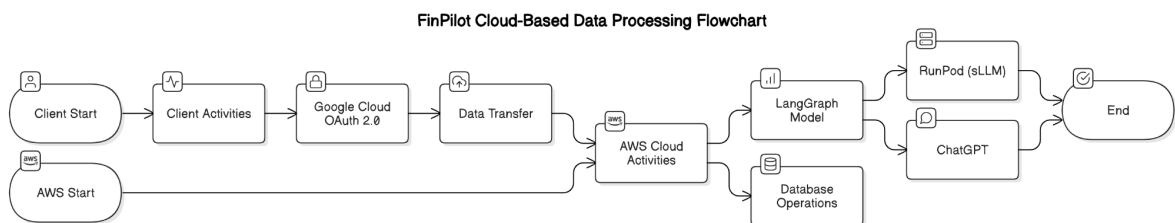


Figure 3. 액티비티 다이어그램 | 별첨2 사진 참조

## 구성

1. 시작 지점: 사용자가 클라이언트에서 작업을 시작합니다.
2. 클라이언트 활동:
  - Chrome 브라우저, Google Docs, 또는 Chrome 확장 프로그램을 통해 요청 생성.
  - Google Cloud OAuth 2.0을 통해 인증이 수행.
3. 데이터 전송:
  - DuckDNS를 통해 클라이언트 요청이 AWS 클라우드로 전달.
  - JSON 형식의 데이터가 HTTPS를 통해 안전하게 전송.

4. AWS 클라우드 내부 활동:

- Nginx가 요청을 수신하고 리버스 프록시 역할을 수행.
- Gunicorn이 FastAPI 애플리케이션에 요청을 전달하여 처리.

5. LangGraph 모델 호출:

- LangGraph 모델은 PyKRX, Open DART, Yahoo Finance 등 외부 데이터를 호출하거나 AI 프롬프트 최적화(Chroma, Prompt Engineering)를 수행.

6. 데이터베이스 작업:

- Amazon RDS(MySQL)를 통해 필요한 데이터를 조회하거나 저장.

7. 외부 서비스 호출:

- RunPod(sLLM) 또는 ChatGPT로 AI 관련 요청 전달.

8. 결과 반환:

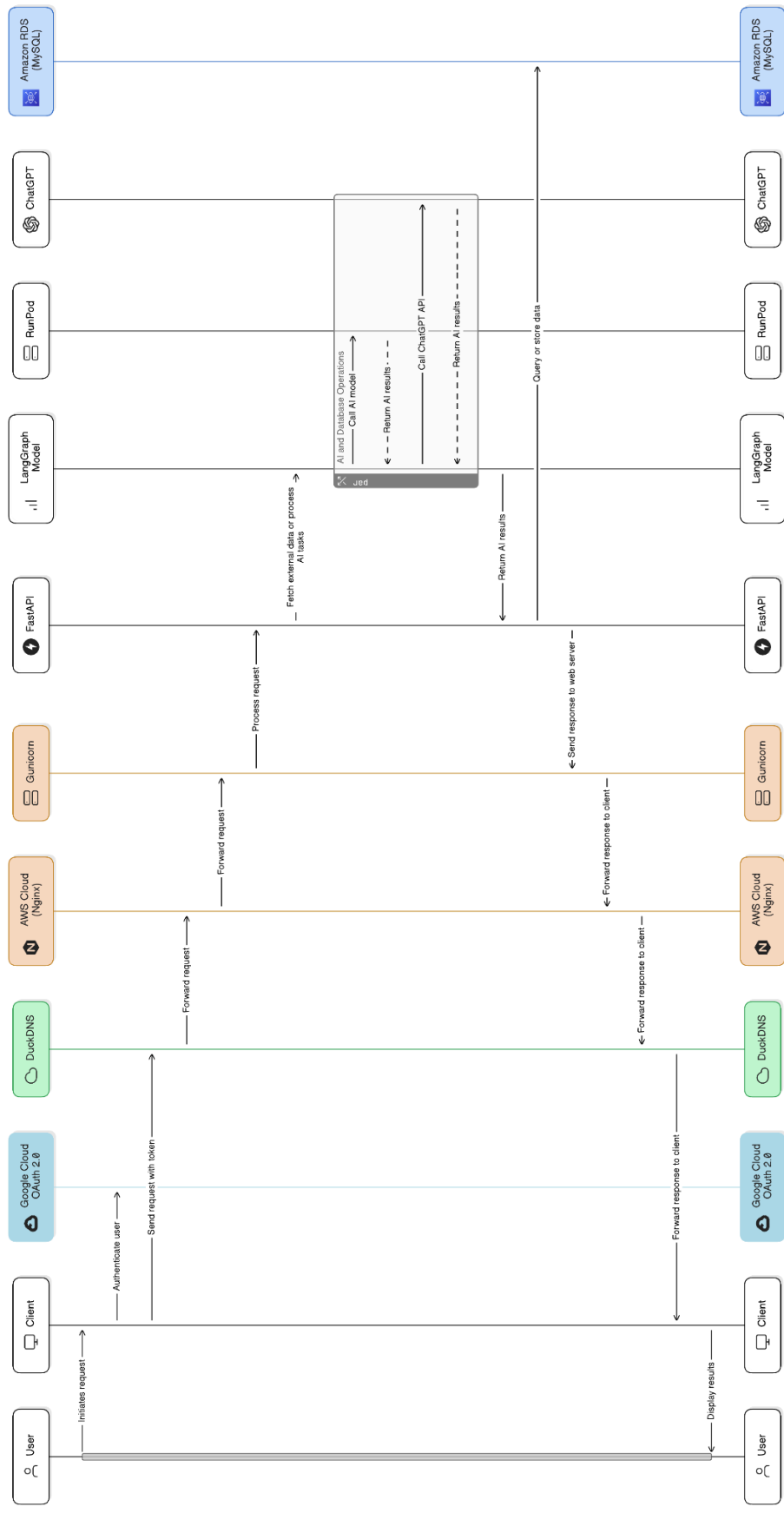
- 처리된 결과는 클라이언트로 반환.

9. 종료 지정: 작업 완료.

**주요 액션 노트**

- 클라이언트 활동: 사용자가 Chrome 브라우저, Docs, 또는 확장 프로그램에서 작업을 시작.
- 데이터 전송: DuckDNS와 HTTPS를 통해 요청이 서버로 전달.
- AWS 클라우드 활동: 서버 측에서 요청을 처리하고 필요한 데이터를 조회하거나 외부 서비스를 호출.
- 외부 서비스 호출: RunPod 또는 ChatGPT와 같은 AI 서비스를 사용하여 추가 처리 수행.

FinPilot Detailed Sequence Diagram



## FinPilot Cloud-Based Data Processing Flowchart

