

SK네트웍스 Family AI 과정 15기

모델배포 시스템 구성도

산출물 단계	모델배포
평가 산출물	시스템 구성도
제출 일자	2025년 10월 02일
깃허브 경로	<a href="https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN15-FINAL-3TEAM">https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN15-FINAL-3TEAM</a>
작성 팀원	이세진

\*아키텍처, 구성도 등 도식화 흐름 이미지 삽입 및 부가설명

개요

- AWS 시스템 구상도

The diagram illustrates the AWS system architecture for model deployment. It is organized into several functional layers and components:

- Frontend (Amplify, Route53):** Shows an Actor interacting with Route53 (DNS) and Amplify.
- Web Tier (EC2 + Docker + Nginx + Gunicorn + Django):** The core application layer. It includes Nginx, Gunicorn, and Django running on EC2 instances, all managed by Docker.
- Inference (RunPod + FastAPI + LangChain + Qwen3 + FAISS):** The AI inference layer. It features FastAPI, LangChain, Qwen3, and FAISS running on RunPod, with a Hugging Face logo indicating the model source.
- Data (RDS MySQL, S3):** The data storage layer, consisting of Amazon RDS for MySQL and Amazon S3.
- Ops/Security (IAM, Parameter Store/Secrets, CloudWatch):** The operations and security layer, including IAM for access management, Parameter Store/Secrets for configuration, and CloudWatch for monitoring.
- CICD (GitHub -> ECR):** The CI/CD pipeline, showing the workflow from GitHub to Amazon ECR (Elastic Container Registry).

Arrows indicate the flow of traffic and data between these components, showing a clear separation of concerns and a scalable, cloud-native architecture.

## Backend with FastAPI

### Backend CI/CD Pipeline

- 애플리케이션 코드를 자동으로 빌드(Build) 및 배포(Deploy)

### ECR (Elastic Container Registry)

- 컨테이너 실행을 위한 Docker 이미지 저장소

### ECS (Elastic Container Service) with Fargate

- 서버리스 방식의 컨테이너 실행 및 관리

### RDS (PostgreSQL, MySQL)

- 관계형 데이터베이스 관리 서비스 (PostgreSQL, MySQL 엔진 지원)

### Amazon S3

- PDF 등 대용량 파일 저장을 위한 오브젝트 스토리지

### Parameter Store (SSM)

- API 키, DB 비밀번호 등 민감 데이터 안전 저장 및 관리

---

## Frontend with Vue.js

### Frontend CI/CD Pipeline

- Vue.js 기반 애플리케이션 자동 빌드 및 배포

### AWS Amplify

- 정적 웹사이트 빌드 및 호스팅, 글로벌 CDN 제공

### CodeCommit

- Git 기반 소스 코드 저장소

### CloudFormation

- 인프라 자동화 및 관리 (Infrastructure as Code)

---

## Authentication

IAM

- 팀원 별 계정 생성 및 권한 부여

---

## DNS & HTTPS

Route 53

- DNS 기반 도메인 관리

## 개발자 작업

1. GitHub 또는 CodeCommit에 코드 푸시
2. Backend CI/CD Pipeline을 통해 Docker 이미지를 빌드 후 **ECR**에 푸시
3. Frontend CI/CD Pipeline으로 정적 파일을 **AWS Amplify**에 배포

## 백엔드

1. **FastAPI** 기반 애플리케이션이 **ECS (Fargate)**에서 실행
2. **Amazon S3**에 PDF 파일 저장 및 제공
3. **RDS (PostgreSQL, MySQL)**와 통신

## 프론트엔드

1. **Amplify**에서 Vue.js 기반 정적 웹사이트 제공

## 사용자 요청

1. **Route 53 (DNS)** 를 통해 사용자 요청이 **Amplify (프론트엔드)** 및 **ALB (백엔드)** 로 전달