k8s 기반의 GPU Job 관리를 위한 Dashboard 개발

2025-2 AJOU SOFTCON

이 름 유수정

지도교수 최재영 교수님 멘 토 강선규(비바리퍼블리카)

개발동기 및 목적

Target

아주대학교 구성원 중 AI 딥러닝 학습을 위해 GPU 리소스가 필요한 사용자

Problem

1. 교내 GPU가 방치되어 있어 자원 사용의 비효율성 발생
2. AI 학습 수요 급증으로 GPU 리소스 수요가 증가했으나, 클라우드 서비스(Colab, AWS 등) 이용 시 일정 사용량 초과 시 과금이 발생하여 학생들에게는 큰 비용 부담

Solution

1. 방치되어 있는 교내 GPU 리소스를 활성화하여, 무료로 사용할 수 있도록 Dashboard 개발
2. Dashboard를 통해 GPU 사용 권한 신청/승인, Job 생성/관리, 실시간 Job 상태 확인 등의 기능을 제공하여 Idle 자원을 사용하고, 비용 부담을 낮춰줌

주요기술

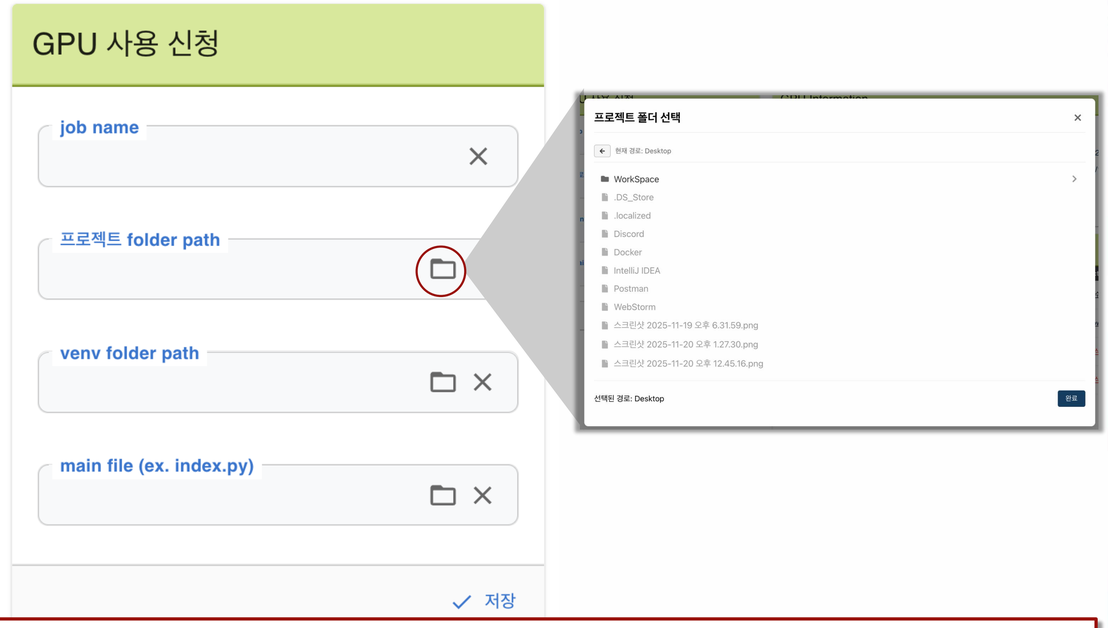
- 통일성과 재활용성을 위해 React.js와 MUI를 사용하여 Dashboard 화면을 구현했습니다.
- 백엔드는 Python 기반의 FastAPI를 사용하여 GPU 리소스 및 Job 관리를 위한 API를 구현하였으며, Asyn 구동을 위해 Uvicorn을 사용합니다.
- 사용자 정보, Job 이력, GPU 상태 등의 데이터는 MongoDB에 저장됩니다.
- Figma를 사용하여 UI/UX를 설계했으며, 코드 관리를 위해 Git/GitHub를 사용했습니다.

개발내용 및 결과

1. GPU Information Card

GPU 상태와 작업 목록을 동시에 가져오는 hook. 컴포넌트 마운트 시 자동으로 데이터 로드.

2. Job 생성 Card



Python os 모듈을 활용해 서버 디렉토리 구조를 스캔하여 반환하고, 프론트엔드에서 파일 트리 UI로 시각화하여 실행 파일 선택.

Job 생성에 성공하면 자동으로 사용 가능한 GPU 할당.

사용 가능한 GPU가 없다면 대기열에서 대기(pending).

3. Job 실행 목록 Card

GPU Job 실행 리스트

Job 요청 시간	Job 이름	Status	Actions
2025-11-19 08:22	test2	pending	수정
2025-11-19 08:22	test	running	show logs
2025-11-19 08:18	second job	failed	삭제
2025-11-19 08:16	first	completed	삭제

프론트엔드에서 Job의 상태에 따라 Action 버튼을 조건부 렌더링
Pending-수정, Running-show log, Failed&Completed-삭제

활용방안 및 기대효과

- 교내 Idle GPU 리소스를 무료로 제공하여, 학습을 위한 학생들의 과금 부담을 해소하고 진입장벽을 낮춘다.
- 아주대학교 코딩 공간인 'AjoupyterHub'와 연동하여 접근성을 높여 학생들의 다양한 AI 학습에 활용한다.
- 복잡한 CLI 환경이나 서버 설정 없이 직관적인 GUI 기반 대시보드를 제공하여 쉽게 GPU 자원을 활용할 수 있다.
- Dashboard를 통해 축적되는 Job 실행 이력과 리소스 점유율 데이터를 분석하여 실제 수요 패턴에 기반한 유의미한 데이터 확보가 가능하다.

오픈소스 URL

Front-end
https://github.com/ytnwjd/gpu_dashboard

Back-end
https://github.com/ytnwjd/gpu_dashboard-server

Front-end Back-end