



# 서던 캘리포니아 대학교 (USC) 어드밴스드 리서치 컴퓨팅 센터가 Backend.AI로 GPU 클러스터를 관리하는 방법

Backend.AI를 통해 쉽게 풀어낸  
복잡한 고성능 컴퓨팅 인프라 관리 방법

# 서던 캘리포니아 대학교 (USC) 어드밴스드 리서치 컴퓨팅 센터가 Backend.AI로 GPU 클러스터를 관리하는 방법

“대학이 보유한 자원을 학생들이 제대로 활용할 수 있도록 돕는 플랫폼의 존재는 매우 중요합니다.  
Backend.AI는 AI 인프라를 쉽게 관리하고, 자원에 쉽게 접근할 수 있게 만드는 운영 플랫폼입니다.”

서던 캘리포니아 대학교 (USC)  
김병도 Associate Chief Research Information Officer

미국 서던 캘리포니아 대학교(University of Southern California)는 컴퓨터공학과 인공지능 분야에 강점을 가진 대학으로, 미국 내 상위권 대학으로 꾸준히 이름을 올리고 있습니다. 특히 소프트웨어 엔지니어링, 게임 개발, 로보틱스 등 다양한 컴퓨터 과학 세부 분야에서 세계 최상위권 수준의 연구 역량을 인정받고 있으며, 산업계와의 긴밀한 협력을 통해 연구 성과를 실제 컴퓨팅 기술과 서비스 혁신으로 이어가고 있습니다.

## 학생들을 위한 GPU 접근성 문제 해결

USC Center for Advanced Research Computing (이하 USC CARC)은 CPU 기반 슈퍼컴퓨팅 자원을 효율적으로 관리하기 위한 인프라를 구축했고, 해당 인프라는 학내 구성원들에게 좋은 반응을 얻었습니다. USC CARC는 슈퍼컴퓨팅 자원 풀을 GPU로 확장하는 과정에서 GPU 역시 비슷한 방식으로 관리, 유연하게 인프라를 할당하고 추적, 회수할 수 있는 시스템을 구축하고자 했습니다. 이를 위해서는 여러 장의 GPU를 온전히 통제하면서도 각 부서 혹은 학과가 필요할 때 기다림 없이 자원을 할당받아 사용할 수 있도록 허용하는 시스템이 필요했습니다. 이에 더해 지속가능한 인프라 관리를 위해서는 정확한 사용량 추적과 사용자별 비용 처리를 가능하게 하는 통합 과금체계를 구축, HPC 클러스터 관리 시스템에 연동해야 했습니다. 하지만 미국 내 연구기관 중에서 이 정도로 통합된 자원 관리 체계를 구축한 사례를 찾기 어려웠고, USC CARC는 플랫폼 구축을 위한 기술적 어려움을 극복해야 하는 상황이었습니다.

## 캠퍼스 고성능 컴퓨팅(HPC) 인프라의 어려움을 Backend.AI를 통해 풀어낸다

래블업은 다양한 글로벌 기업과 연구기관의 GPU 인프라 효율화를 지원해 온 경험과 기술력을 바탕으로 USC CARC의 HPC 인프라 관리 솔루션으로 Backend.AI를 제공하게 되었습니다. USC CARC는 Backend.AI의 컨테이너 수준 GPU 분할가상화 기술과 Sokovan 스케줄러를 활용하여 연구실·학과·강의별 실시간 수요에 맞춰 자원을 유연하게 배분했고, 사용자별로 완전히 분리된 스토리지 그룹을 제공해 동시다발적인 학술 연구 환경에서도 안정적인 운영을 가능하게 하였습니다. Backend.AI를 도입함으로써 USC CARC는 보유한 GPU 하드웨어 자원의 활용률을 극대화했으며, 기관의 운영 정책을 반영한 통합 과금 시스템을 Backend.AI에 통합함으로써 학생·교수·행정 전반에 걸쳐 투명하고 공정한 자원 관리 체계를 확립했습니다.