



RebelRack™

생성형 AI와 에이전틱 워크플로우,
전원만 켜면 끝나는 완성형 턴키 인프라

RebelRack은 대규모 추론 배포의 복잡함을 걷어낸 완성형 AI 팩토리입니다. 고효율·고성능 RebelServer에 엔터프라이즈급 네트워킹과 Rebellions 소프트웨어 스택까지 하나로 묶어, 골치 아픈 통합 과정은 사라지고 곧바로 모델 배포와 추론에 집중할 수 있습니다. 이미 검증을 마치고 실제 양산 환경에서 입증된 구성이기에, DIY 클러스터와는 비교할 수 없이 빠르게 성과를 만들어냅니다.

치솟는 전력 소비와 운영 비용, 추론 배포의 가장 큰 골칫거리를 정면으로 겨냥했습니다. 성능은 끌어올리고 에너지 효율은 극대화하도록 설계된 서버입니다.

이 시스템은 차세대 에이전틱·추론 모델을 위한 토대입니다. MoE(Mixture of Experts) 아키텍처부터 규모를 가리지 않는 모든 모델, 언어·비전·음성을 넘나드는 멀티모달 기능까지 막힘없이 통합합니다.

Key Benefits



폴스택 AI 인프라

프로덕션급 클라우드 오케스트레이션에 맞춰 최적화된, 전원만 켜면 되는 턴키 시스템입니다. Rebel100™ 가속기와 Rebellions 소프트웨어 스택이 오픈소스 엔진과 분산 추론을 지원하고, 사전 검증된 모델까지 갖춰 프로덕션까지 가는 길을 확 단축합니다.



업계 최고 수익성의 추론 플랫폼

대규모 에너지 효율 추론과 에이전틱 워크로드를 위해 태어났습니다. 업계 최고의 와트당 가격 대비 성능으로 전력은 아끼고 처리량은 끌어올려, 운영 비용은 낮추고 AI 투자 수익(ROI)은 극대화합니다.



검증된 안정성, 즉시 배포

DIY 클러스터에 매달리던 수개월간의 구성 작업, 이제 건너뛰세요. 이미 수백 대의 랙이 현장에서 돌아가고 있고, 모든 구성은 실제 AI 워크로드로 혹독하게 검증됐습니다. 도입에서 가동까지, 몇 달이 아니라 며칠이면 충분합니다.



인프라 걱정 없는 소버린 AI

수냉식 설비도, 전력 개조도 필요 없습니다. 지금 쓰는 공냉식 데이터센터에 그대로 들어갑니다. 완전한 온프레미스 운영으로 민감한 정부·기업 데이터의 주권과 보안, 통제권을 온전히 지켜냅니다.

Software Edge: Rebellions Cloud-Native Stack

랙에는 Rebellions 소프트웨어 스택이 이미 통합돼 있습니다. 새로운 기술을 배울 필요 없이 기존 전문성을 그대로 활용할 수 있습니다:

추론 엔진

vLLM, PyTorch와 함께 설계해 호환성을 극대화했습니다.

분산 추론

고속 인터커넥트를 다루는 전용 소프트웨어 라이브러리로 멀티노드 확장이 매끄럽습니다.

벤더 종속 없음

오픈소스 프레임워크와 업계 표준 도구에 자연스럽게 녹아듭니다.

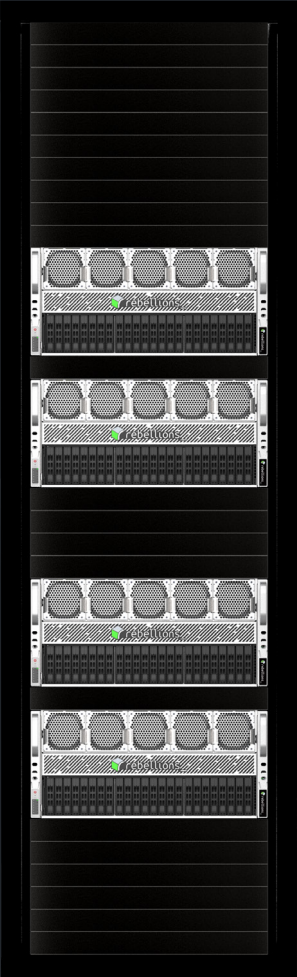
모델 카탈로그

사전 검증·최적화된 다양한 모델을 갖춰, 별도 작업 없이 곧바로 프로덕션에 투입합니다.

RebelRack™ Hardware

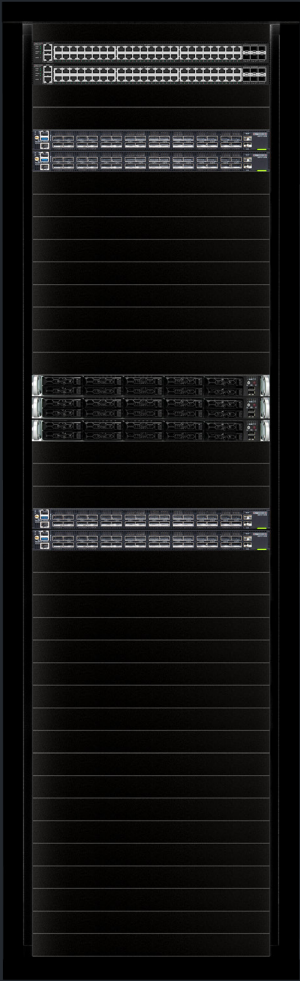
기본구성	1x RebelRack NPU 랙 1x RebelRack 매니지먼트 랙
확장	초기 기본 구성 이후, 네트워크 패브릭을 통해 NPU 랙을 한 대씩 또는 필요에 따라 최대 8대까지 추가할 수 있습니다. NPU 랙이 8대를 초과할 경우 추가 매니지먼트 랙이 필요합니다.

RebelRack™ NPU Rack



서버 모델	4x RebelServer™ (5U)
가속기	32x RebelCard™
NPU 메모리	총 4.61TB, 대역폭 153.6TB/s HBM3e
CPU	8x AMD EPYC™ 9355 Processors (256C/512T)
시스템 메모리	6TB DDR5
네트워크	백엔드 패드릭 : 400G QSFP112-DD 프론트엔드 네트워크 : 25G SFP28 스토리지 네트워크 : 100G (선택 사항)
성능	FP8: 64 PFLOPS FP16: 32 PFLOPS
냉각 방식	공냉식
소비전력	28kW max
랙	높이: 42U, 78.38" (1991mm) 너비: 23.62" (600mm) 깊이: 47.24" (1200mm)
무게	~773 lbs (~351 kg), 미포장 기준
냉각	최대 105,094 BTU/시간

RebelRack™ Management Rack



매니지먼트 서버	3x AMD CPU 서버 (선택 사항)
CPU	3x AMD EPYC 9254/9255 (24 cores)
시스템 메모리	총 384GB (3x 매니지먼트 서버, 각 128GB [4x 32GB DDR5])
네트워크	백엔드 패브릭 : 800G OSFP 프론트엔드 네트워크 : 25G SFP28 + 100G QSFP28
스토리지	검증된 파트너 스토리지 솔루션을 유연하게 지원합니다.
소비 전력	5.2kW max
랙	높이: 42U, 78.38" (1991mm) 너비: 23.62" (600mm) 깊이: 47.24" (1200mm)
무게	~ 558 lbs (~ 253 kg), 비포장 기준
냉각	최대 19,517 BTU/시간

* 기존 데이터센터 환경에 손쉽게 통합할 수 있도록 다양한 랙 구성 옵션을 제공합니다.