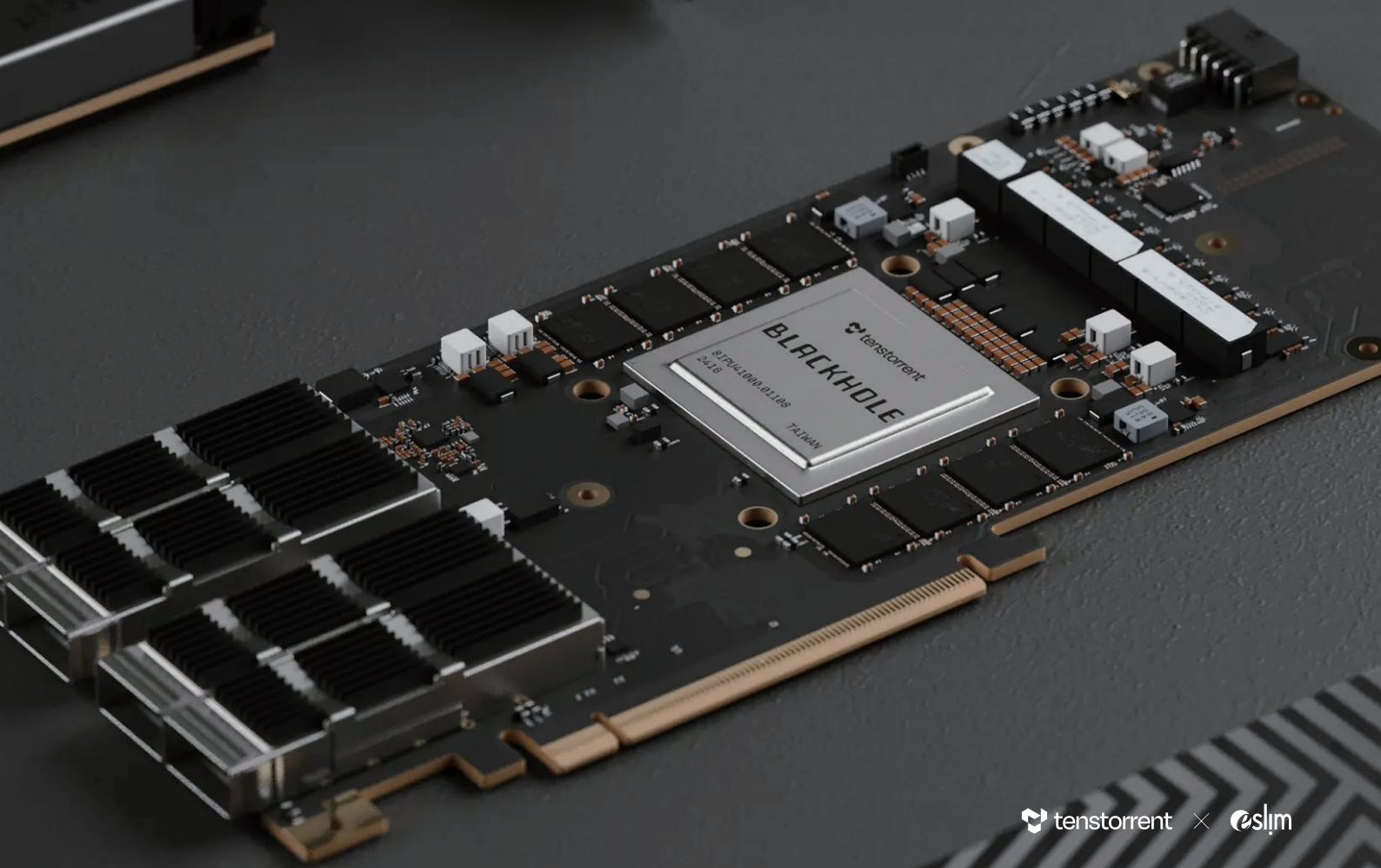


TT-Blackhole™

RISC-V 아키텍처 기반, 고성능-하이엔드 AI 가속기



tenstorrent × eslim

AI 연산의 스케일업, 더 깊은 새로운 차원으로

TT-Blackhole™은 플래그십 AI 가속기로, Tensix Core와 Big RISC-V코어 기반의 SoC 아키텍처를 통해 초고성능 연산과 유연한 시스템 구성을 지원합니다.

TT-Blackhole™은 LLM, 자율주행, 비전 AI 등 고난이도 워크로드에 최적화되어 있으며, TT-Forge™와 TT-Metalium™과 같은 소프트웨어 스택과 연동해 커스터마이징된 연산 경로 설계가 가능합니다.

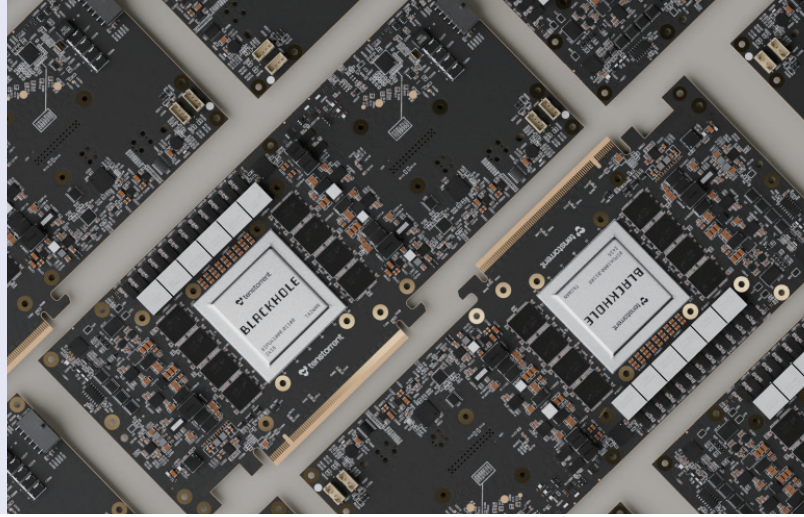
텐스토렌트
한국공식총판

이슬림코리아
www.eslim.co.kr



Blackhole™ Tensix 프로세서

고성능 • 하드웨어 유연성 • 확장 및 연결 최적화



Blackhole™은 Tenstorrent의 차세대 AI 가속기로, 최신 세대 Tensix 아키텍처를 기반으로 고성능 AI 연산과 고도화된 모델 처리를 위해 설계되었습니다. 이 가속기는 최대 140개의 Tensix 코어와 16개의 Big RISC-V 코어를 탑재하여, FP8 기준 최대 745 TFLOPs, FP16 기준 372 TFLOPs의 연산 성능을 제공합니다. 특히 대규모 언어 모델(LLM), 컴퓨터 비전, 자율주행, 과학적 시뮬레이션 등 복잡한 워크로드를 안정적으로 처리할 수 있도록 설계되었습니다. Blackhole의 p150 시리즈는 32GB GDDR6 메모리와 512GB/s 대역폭, PCIe 5.0 x16 인터페이스를 갖추었으며, 4개의 QSFP-DD 800G 포트를 통해 최대 3.2Tbps의 내부 연결 대역폭을 지원합니다. 이를 통해 다수의 카드를 유기적으로 연결해 클러스터를 구성하거나 메모리를 풀링하여 확장성을 극대화할 수 있습니다. 또한 TT-Metalium™(로우 레벨), TT-Forge™(컴파일러)와 같은 오픈소스 SDK를 통해 다양한 정밀도 포맷(FP8, FP16, BF16, TF32 등)의 유연한 연산 설정과 모델 최적화가 가능하며, 개발자는 하드웨어 수준까지 제어하며 시스템을 커스터마이징할 수 있습니다. Blackhole은 고성능, 확장성, 유연성을 통합한 Tenstorrent의 플래그십 AI 가속 솔루션으로, 고도화된 AI 인프라 환경에 최적의 선택지를 제공합니다.

고도화된 설계

강력함 :

Blackhole 칩은 300W의 전력으로 동작하며, 최대 745 TFLOPs(FP8), 372 TFLOPs(FP16)의 연산 성능을 제공하는 고성능 AI 가속기로 설계되었습니다.

확장성 :

p150 시리즈는 QSFP-DD 800G 포트를 통해 다수의 카드를 직접 연결할 수 있어, 대규모 클러스터 구성과 메모리 풀링을 유연하게 지원합니다.

오픈 소스 :

TT-Metalium™(로우 레벨)과 TT-Forge™(중간 레벨 컴파일러) SDK를 통해 모델 실행 최적화와 하드웨어 자원 제어가 자유롭게 가능합니다.

커스터마이징 :

정밀도, 연산 경로, 메모리 처리 방식 등 다양한 파라미터를 개발자가 직접 조정할 수 있어, 특정 워크로드에 맞춤 대응이 가능합니다.

유연성 :

FP8, FP16, BF16, TF32, VFP32 등 다양한 정밀도 연산 포맷을 자동 최적화 방식으로 지원하여, 처리 효율과 연산 정확도를 균형 있게 유지합니다.

멀티 유저 친화적 :

SoC 구조 기반의 아키텍처는 여러 사용자가 동시에 병렬 연산 및 OS 기반 작업을 수행할 수 있도록 설계되어, 고도화된 서버 환경에 적합합니다.

Blackhole™ Tensix 프로세서

* 텐스토렌트 Blackhole™ 기반 카드에만 연결할 수 있습니다.

Card	p100a	p150a	p150b
Tensix Cores	120	140	140
AI Clock	1.35 GHz	1.35 GHz	1.35 GHz
“Big RISC-V” Cores	16	16	16
SRAM	180MB	210MB	210MB
Memory	28GB GDDR6	32GB GDDR6	32GB GDDR6
Memory Speed	16 GT/sec	16 GT/sec	16 GT/sec
Memory Bandwidth	488 GB/sec	512 GB/sec	512 GB/sec
TFLOPS (FP8)	638	745	745
TFLOPS (FP16)	319	372	372
TBP(Total Board Power)	300W	300W	300W
Connectivity	N/A	4x QSFP-DD 800G (Passive)*	4x QSFP-DD 800G (Passive)*
System Interface	PCI Express 5.0 x16	PCI Express 5.0 x16	PCI Express 5.0 x16
Cooling	Active	Active	Passive
FormFactor	Dual Slot, FHFL	Dual Slot, FHFL	Dual Slot, FHFL