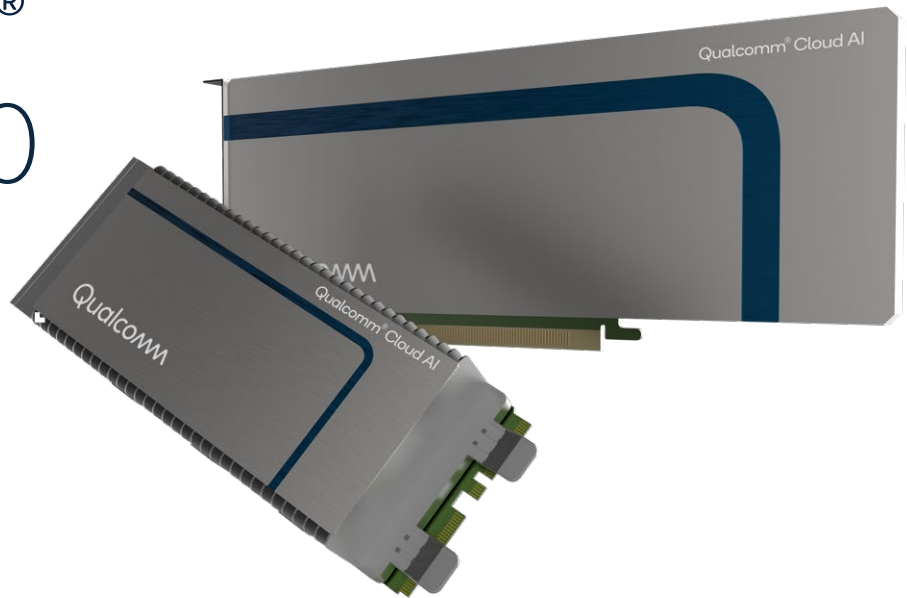




Qualcomm® Cloud AI 100

**Purpose-built for
high performance,
low-power AI
processing in the cloud.**



Schedule	Sampling Now - Commercial launch 1H 21
Process Node	7nm
Card Types	Dual M.2 (edge): 15W TDP Dual M.2: 25W TDP PCIe (HHHL): 75W TDP
Card Performance (Raw TOPS)	Dual M.2 (edge): 70 TOPS Dual M.2: 200 TOPS PCIe: 400 TOPS
AI Cores	Up to 16 cores
On Die SRAM	144MB (9MB Each AI Core)
On Card DRAM	Up to 32GB w/ 4x64 LPDDR4x @ 2.1GHz
PCIe (connection to the host)	8 lane Gen3/4 (PCIe) or 4 lane Gen3/4 (Dual M.2)
Data Types	INT8, INT16, FP16, FP32