

CPU vs. GPU - Architectural Comparison

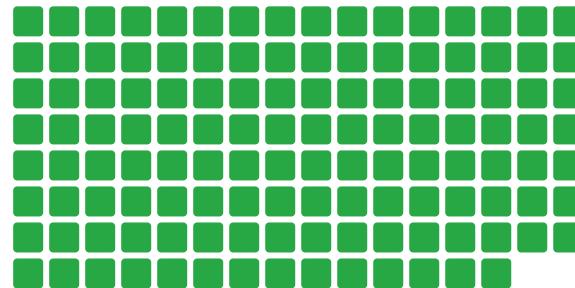
CPU



Few powerful cores (4-64)

- Optimized for sequential tasks
- Better for control flow & branching
- General computing excellence

GPU



Thousands of simple cores (1000s-10000s)

- Massive parallel processing
- 10-100x faster for matrix ops
- Ideal for ML workloads

GPU Performance: **10-100x faster for matrix operations**

Modern ML Workflow

GPU: Training

CPU: Preprocessing / Serving