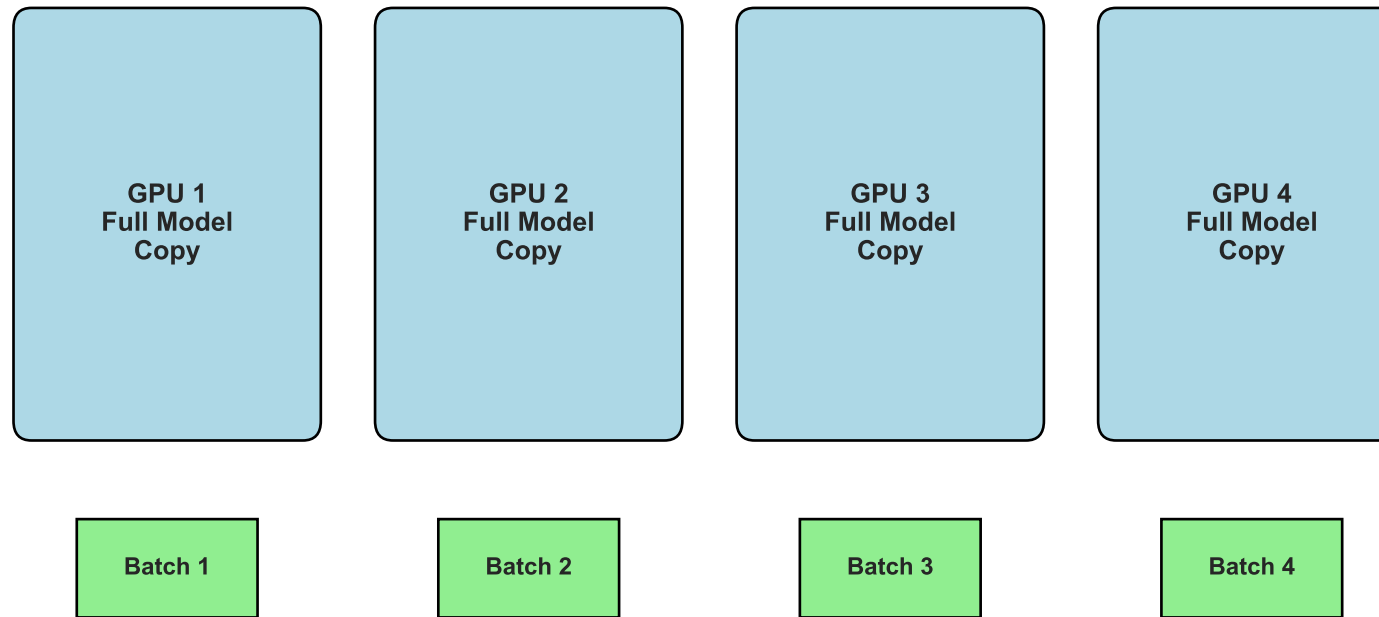


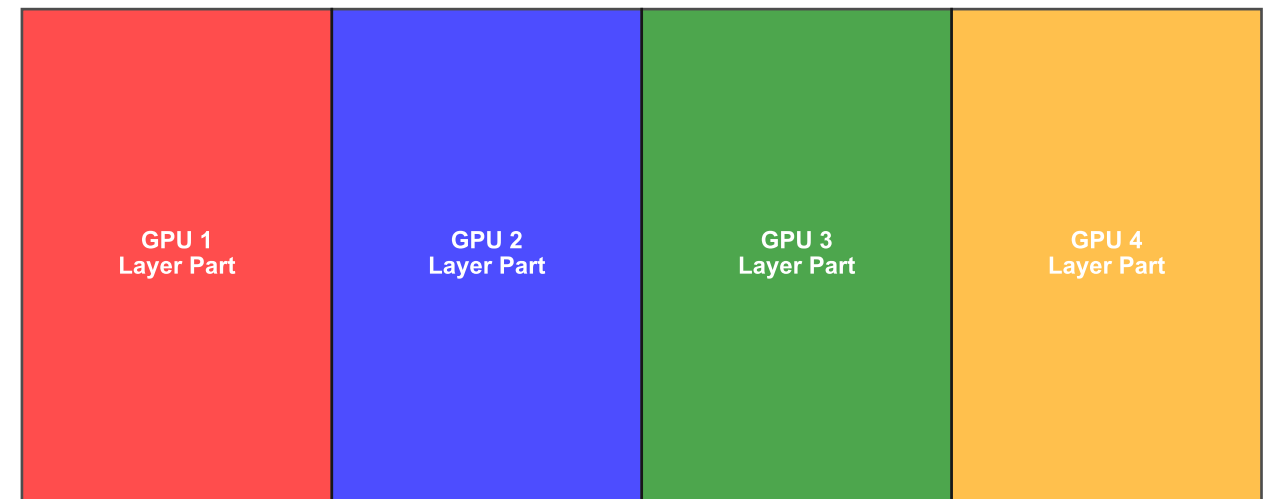
Model Parallelism: Training at Scale

Data Parallelism



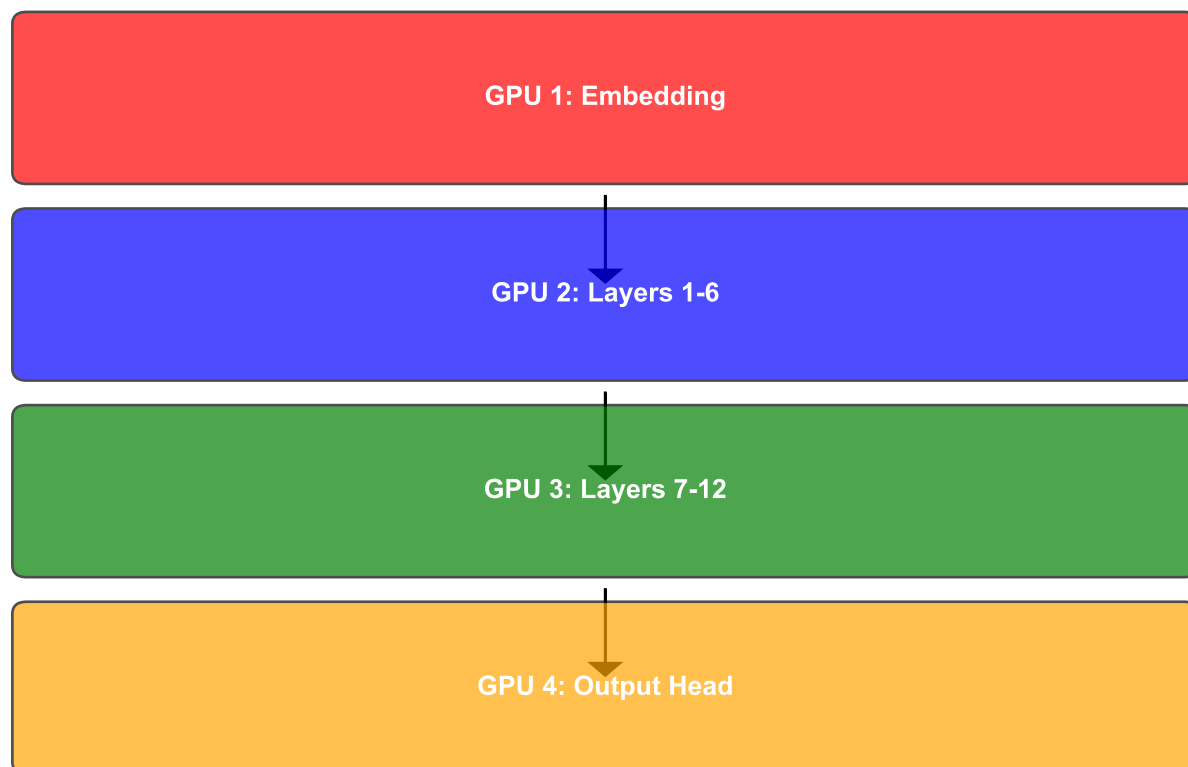
Gradients aggregated across GPUs

Tensor Parallelism



Single layer split across GPUs

Pipeline Parallelism



Parallelism Strategy Comparison

