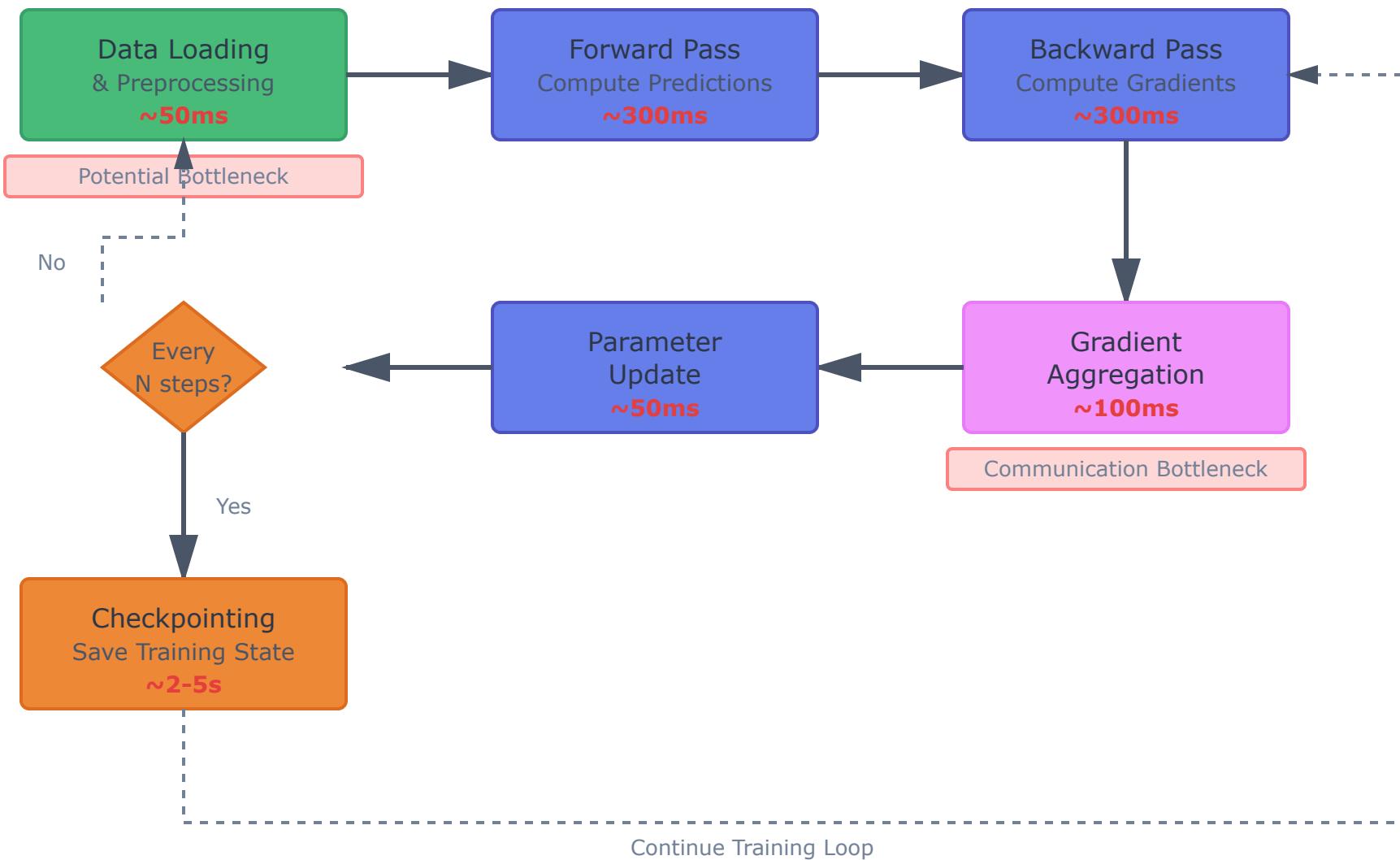


Training Pipeline Architecture



Total Time per Step: ~800ms (without checkpoint) or ~3-5s (with checkpoint)

Key Bottlenecks:

1. Data Loading: Can limit GPU utilization if not optimized
2. Gradient Aggregation: Communication overhead scales with GPU count