

# All-Reduce data parallel

[arxiv.org/abs/1706.02677](https://arxiv.org/abs/1706.02677)

**Idea:** get rid of the host, each gpu runs its own computation

**Q:** why will weights be equal after such step?

