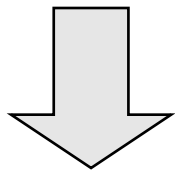
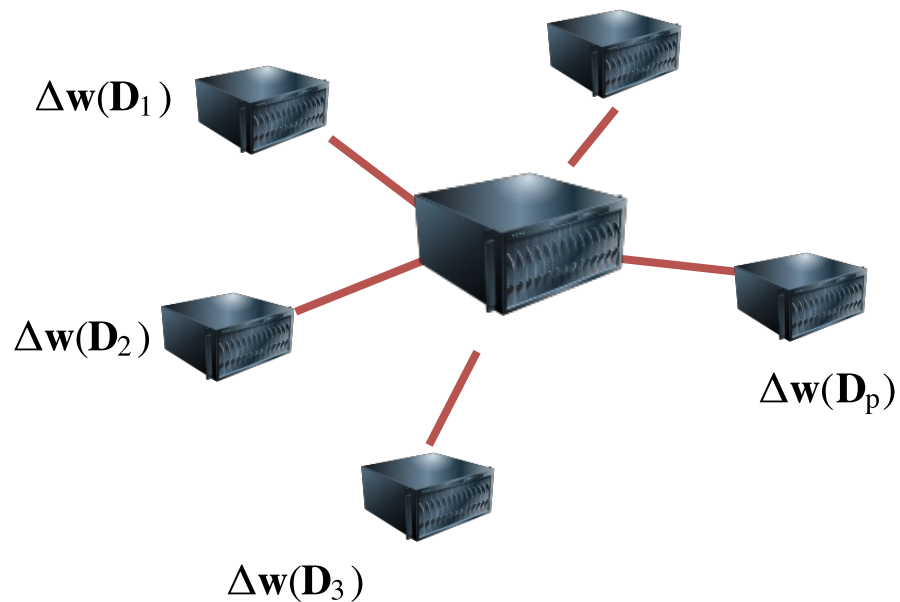


$$\mathbf{w}_{t+1} := \mathbf{w}_t + \Delta \mathbf{w}_t(\mathbf{D})$$

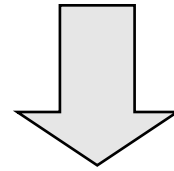
Data Parallelism



$$\mathbf{D} := \{\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_p\}$$



Model Parallelism



$$\mathbf{w} := \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p\}$$

