**GradientFlow: Optimizing Network Performance for Large-Scale Distributed DNN Training (**[**DOI: 10.1109/TBDATA.2019.2957478**](https://doi.org/10.1109/TBDATA.2019.2957478)**)**

This paper goes in to introduce deep neural net training for those who don’t know. They then discuss how they distribute the data from the data into the model. They discus both low level and high level aspects of the process. They then go into showing data they collected from their model with varying numbers of GPUs from 16 to 128. They discuss different cluster orientations and different architectures to speed up communication between the components.