

# Fully Sharded Data Parallelism

## Instructor

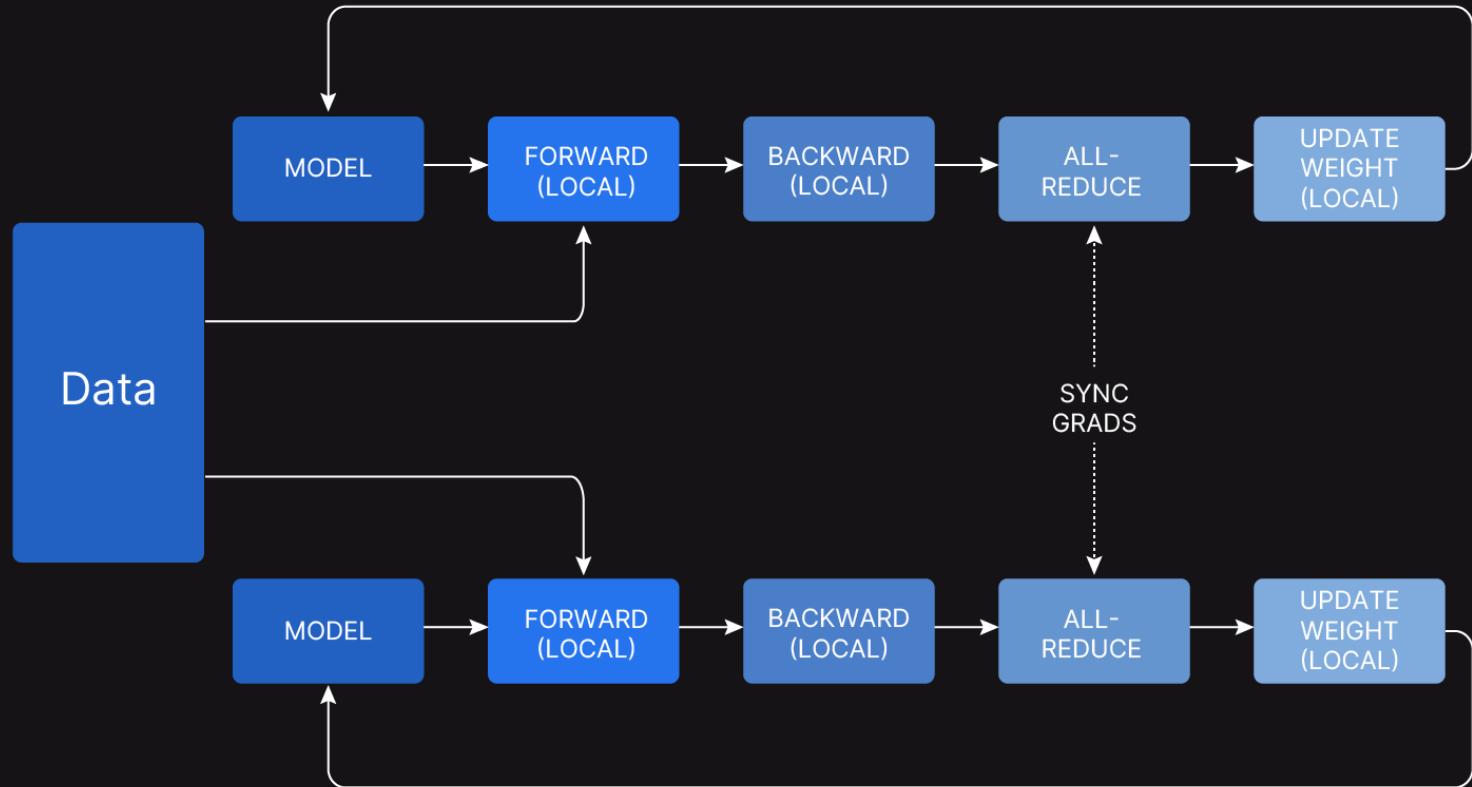
Sourab Mangrulkar

Machine Learning Engineer at

Creator of PEFT



# Recap of Data Parallelism



## Challenges with Data Parallelism

Memory inefficient

# Recap of Pipeline Parallelism



# Challenges with Pipeline Parallelism

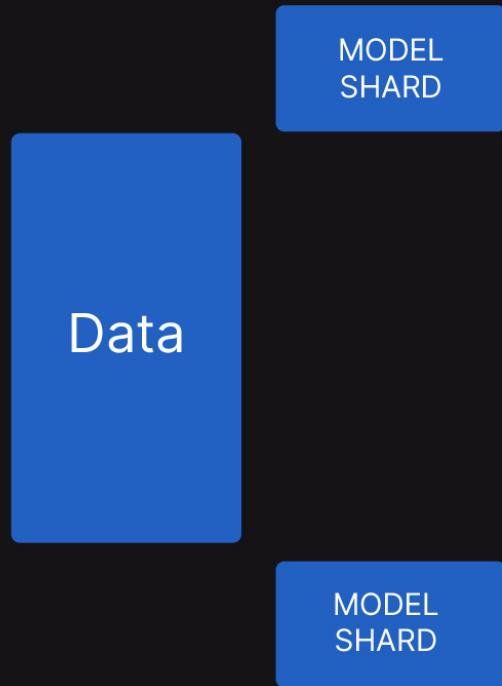
- Communication Overhead
- Underutilisation of GPUs

# Fully Sharded Data Parallelism

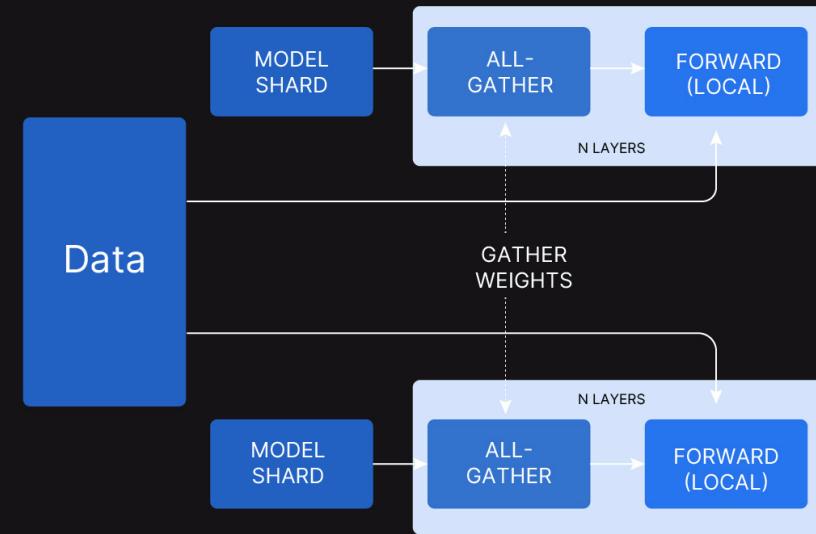
# Fully Sharded Data Parallelism

- Shard Optimiser states (ZeRO Stage I)
- Shard Gradients and Optimiser (ZeRO Stage II)
- Shard Parameters, Gradients and Optimiser (ZeRO Stage III)

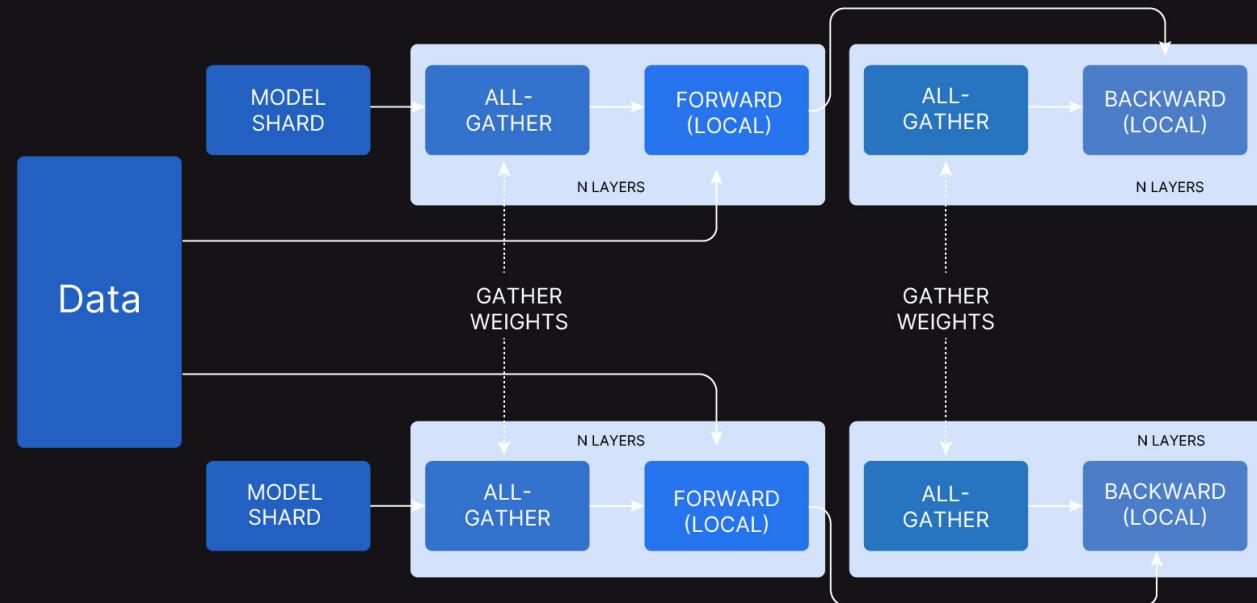
# Fully Sharded Data Parallelism (FSDP)



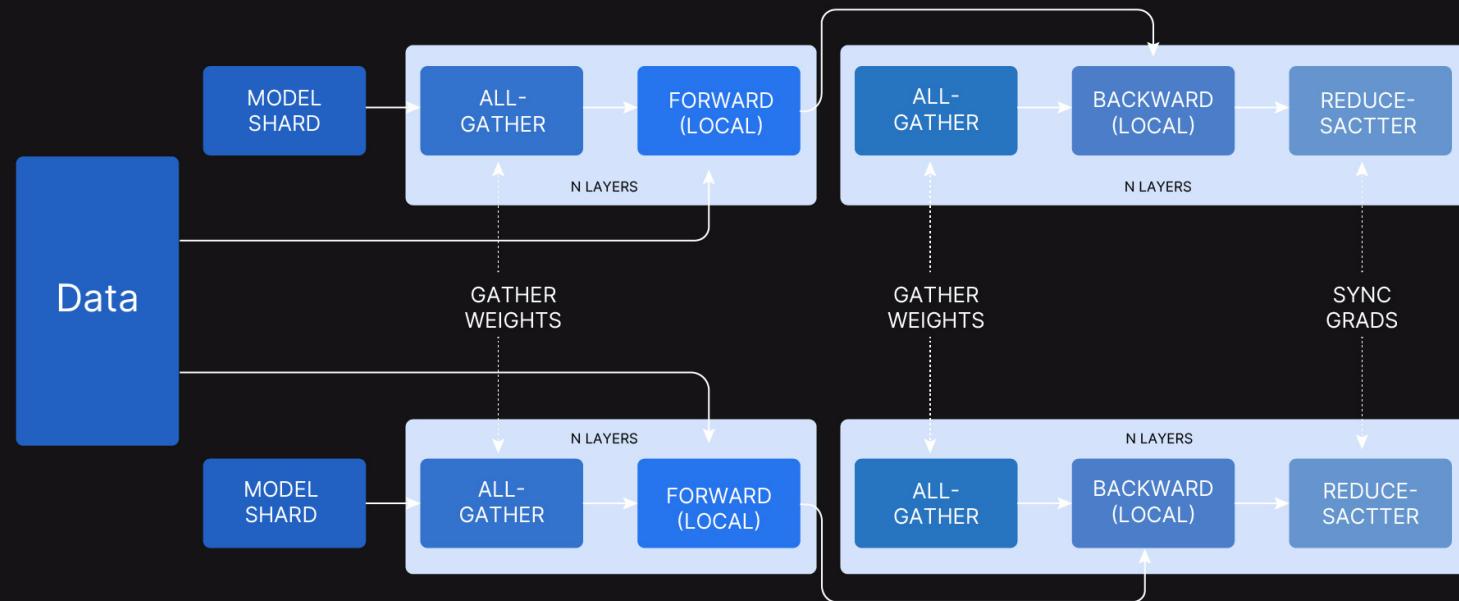
# FSDP



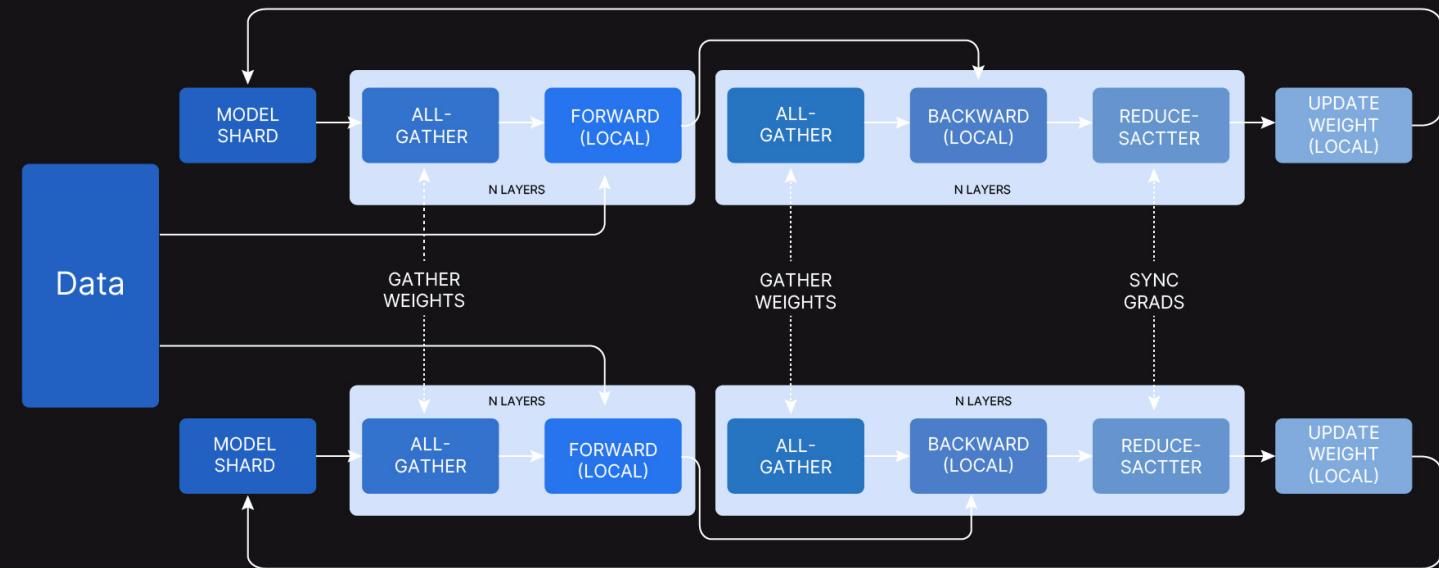
# FSDP



# FSDP



# FSDP



## Pros

- ✓ Memory efficient
- ✓ Compute efficient

# Thank You

---