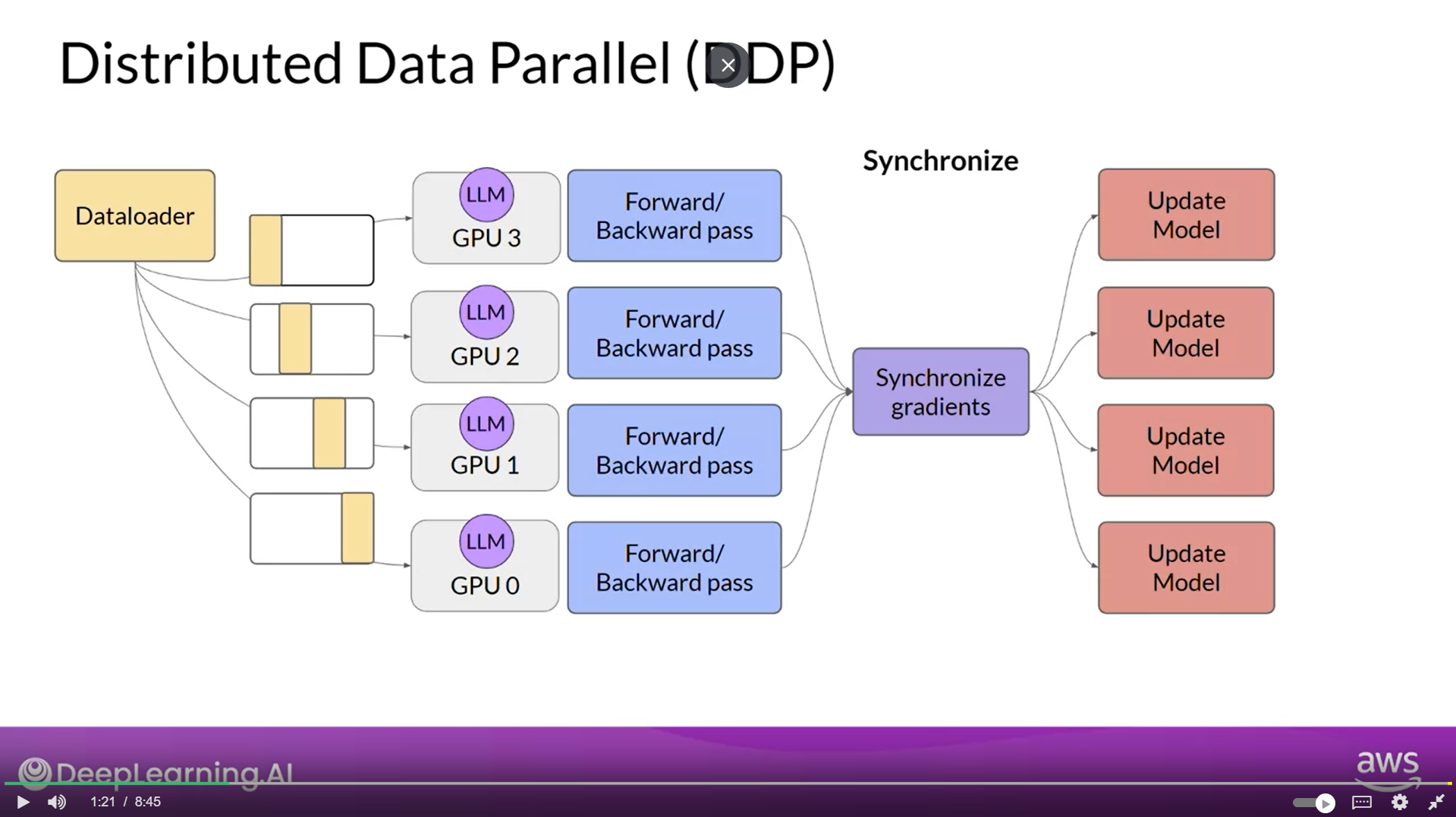
Efficient Multi GPU training Strategies

DDP



Note: DDP assumes that you model is small enough to train on sinlr GPU but you are using multiple GPUs to speed up the training process.

A chart with text on it

Description automatically generated with medium confidence

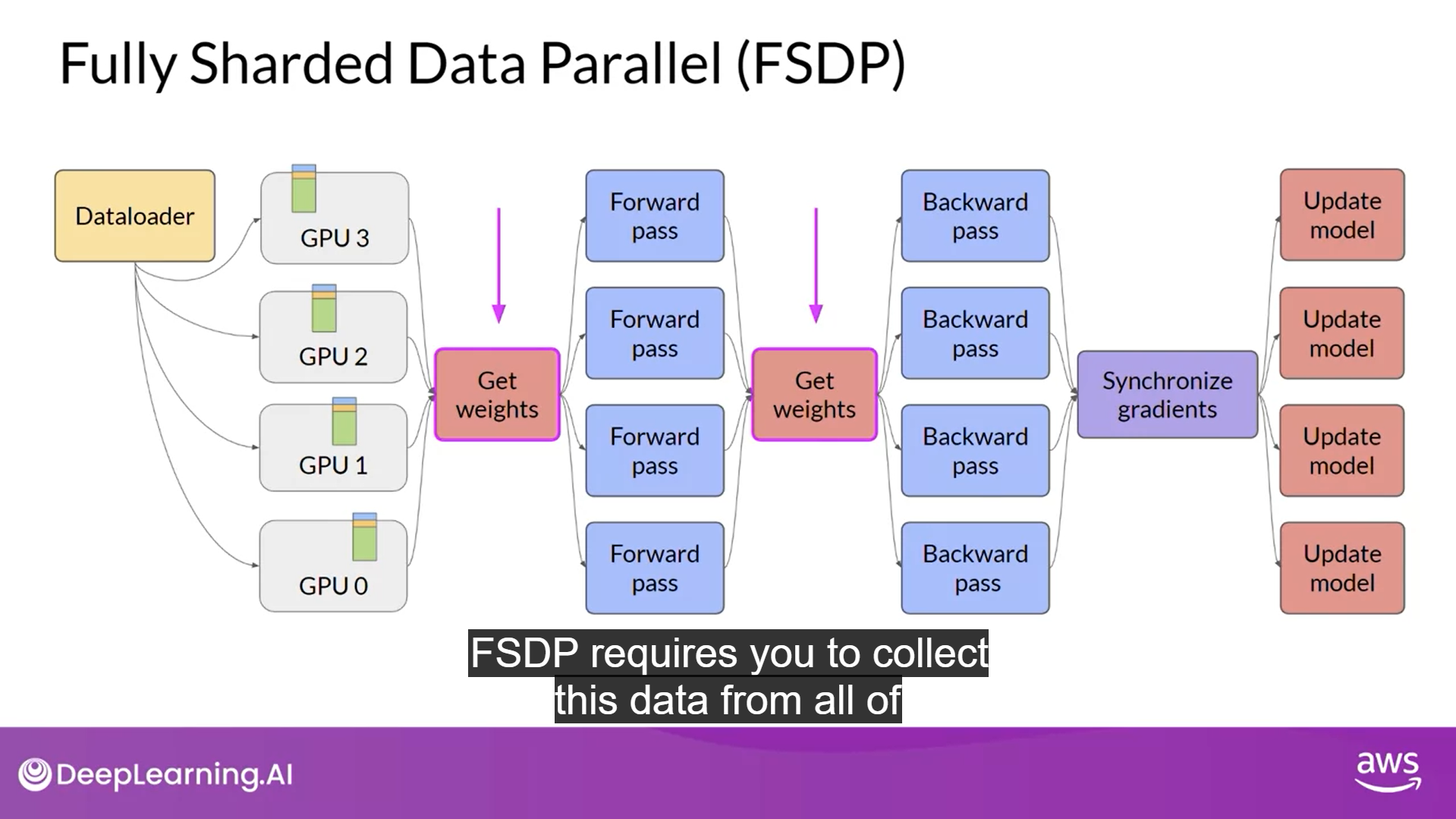
Fully Shared Data Parallel (FSDP) is a type of Modal Sharding

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated



A screenshot of a computer

Description automatically generated

FSDP vs DDP Performance

A graph with numbers and a bar chart

Description automatically generated with medium confidence