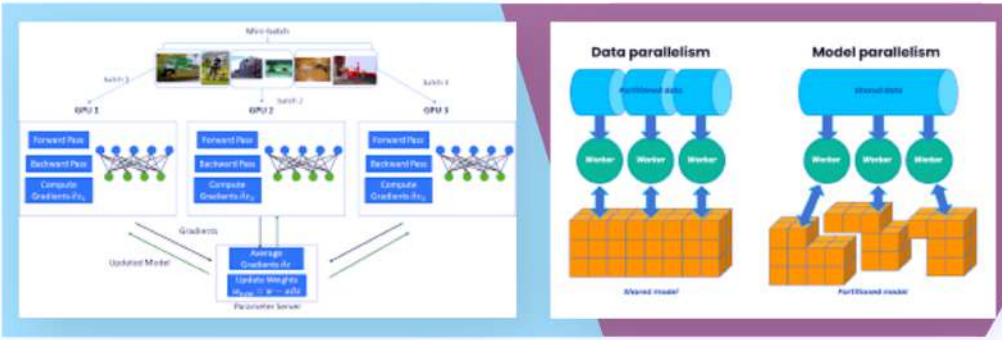
Parallel and Distributed Machine Learning

Parallel and Distributed Learning

parallel machine learning algorithms can be executed simultaneously on several computers or processors

CPU VS GPU

Distributed machine learning is a technique that splits the data and/or the model across multiple machines or nodes, and coordinates the communication and synchronization among them.

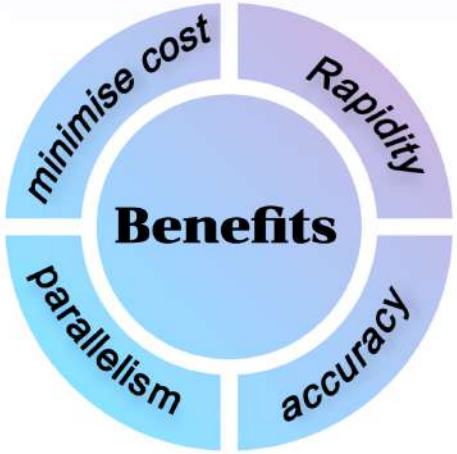


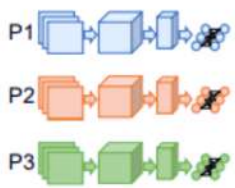
CHALLENGES

I. Partitioning
II. Communication bottlenecks
III. Scalability
IV. Extra Hardware

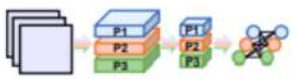
POINTS OF COMPARE

- Speed -> Parallel: Fast || Distributed: Scalable
- Scalability -> Limited || Extensible
- Fault Tolerance -> Vulnerable || Robust
- Communication -> Local || Networked
- Resource Utilization -> Localized || Distributed
- Complexity -> Simple || Comprehensive
- Cost -> Affordable || Expensive

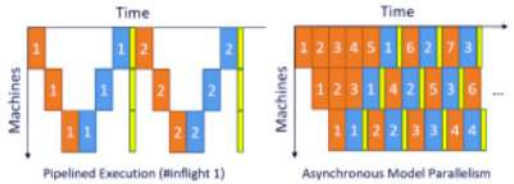




(a) Data Parallelism



(b) Model Parallelism



(a) Pipelined Asynchronous Execution [76]

Illustration

The publication "Parallel stochastic gradient descent for deep learning: A survey" by J. Dean et al. (2012)

- provides an overview of optimisation technique for deep neural network training: stochastic gradient descent (SGD).
- The authors analyse the effectiveness of several ways to parallelizing SGD