Model Parallelism: Building and Deploying Large Neural Networks

ID MPBDLNN **Prix** sur demande **Durée** 1 jour

Pré-requis

Familiarity with:

- Good understanding of PyTorch
- Good understanding of deep learning and data parallel training concepts
- Practice with deep learning and data parallel are useful, but optional

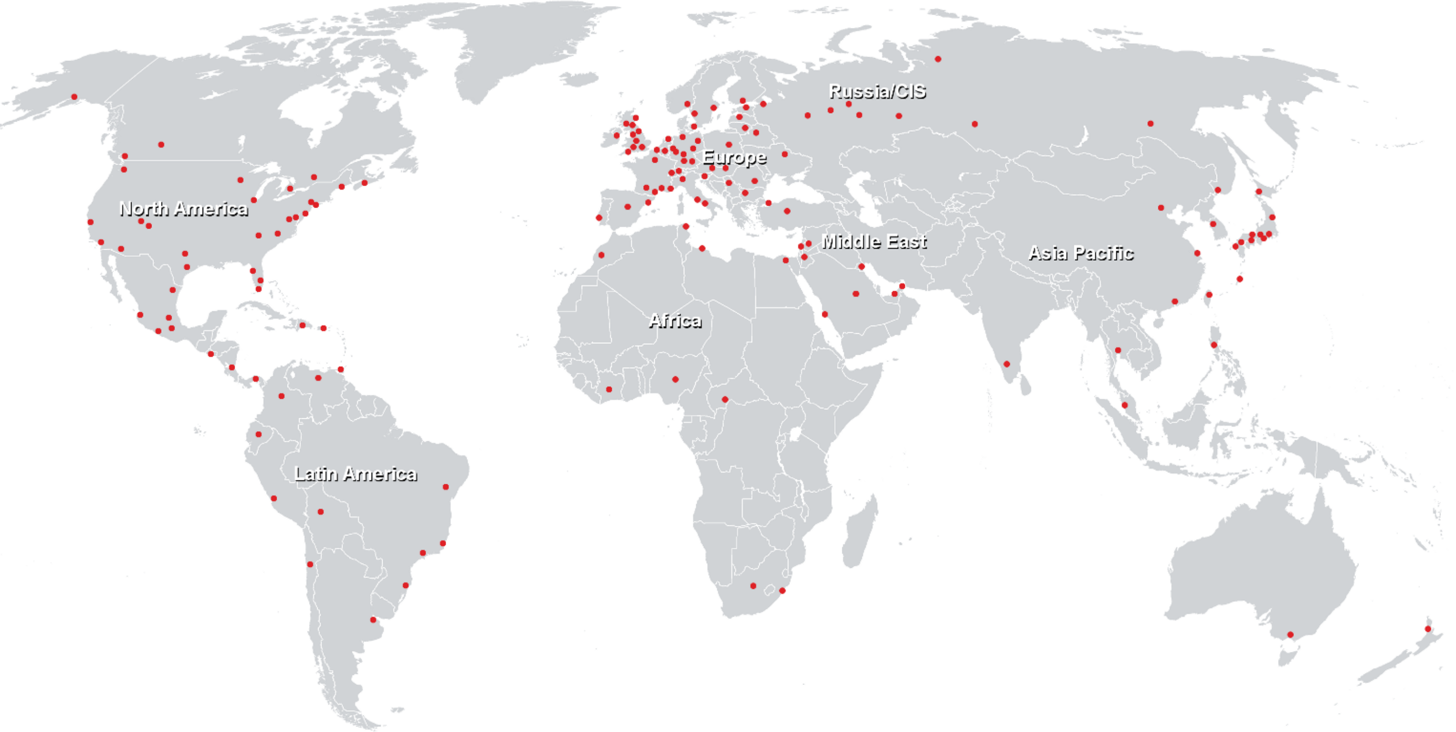
Objectifs

In this workshop, participants will learn how to:

- Train neural networks across multiple servers
- Use techniques such as activation checkpointing, gradient accumulation, and various forms of model parallelism to overcome the challenges associated with large-model memory footprint
- Capture and understand training performance characteristics to optimize model architecture
- Deploy very large multi-GPU models to production using NVIDIA Triton™ Inference Server

Model Parallelism: Building and Deploying Large Neural Networks (MPBDLNN)

Centres de formation dans le monde entier



Fast Lane Institute for Knowledge Transfer (Switzerland) AG

Husacherstrasse 3
CH-8304 Wallisellen
Tel. +41 44 832 50 80

info@flane.ch, https://www.flane.ch