

Parallel and Distributed Paradigm

Instructor

Sourab Mangrulkar

Machine Learning Engineer at

Creator of PEFT



Introduction

- Huge number of Parameters
- Trained on Massive Datasets
- Massive Computing GPU

Compute Power for Training LLMs



GPT4: 25000 GPUs

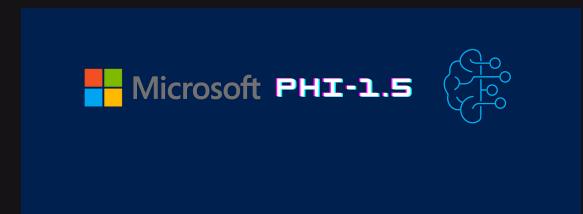


Llama 2

Meta AI

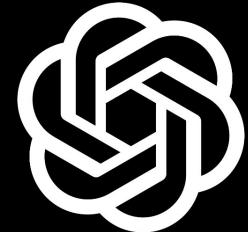


Falcon 180B: 4096 GPUs



Phi 1.5: 16 GPUs

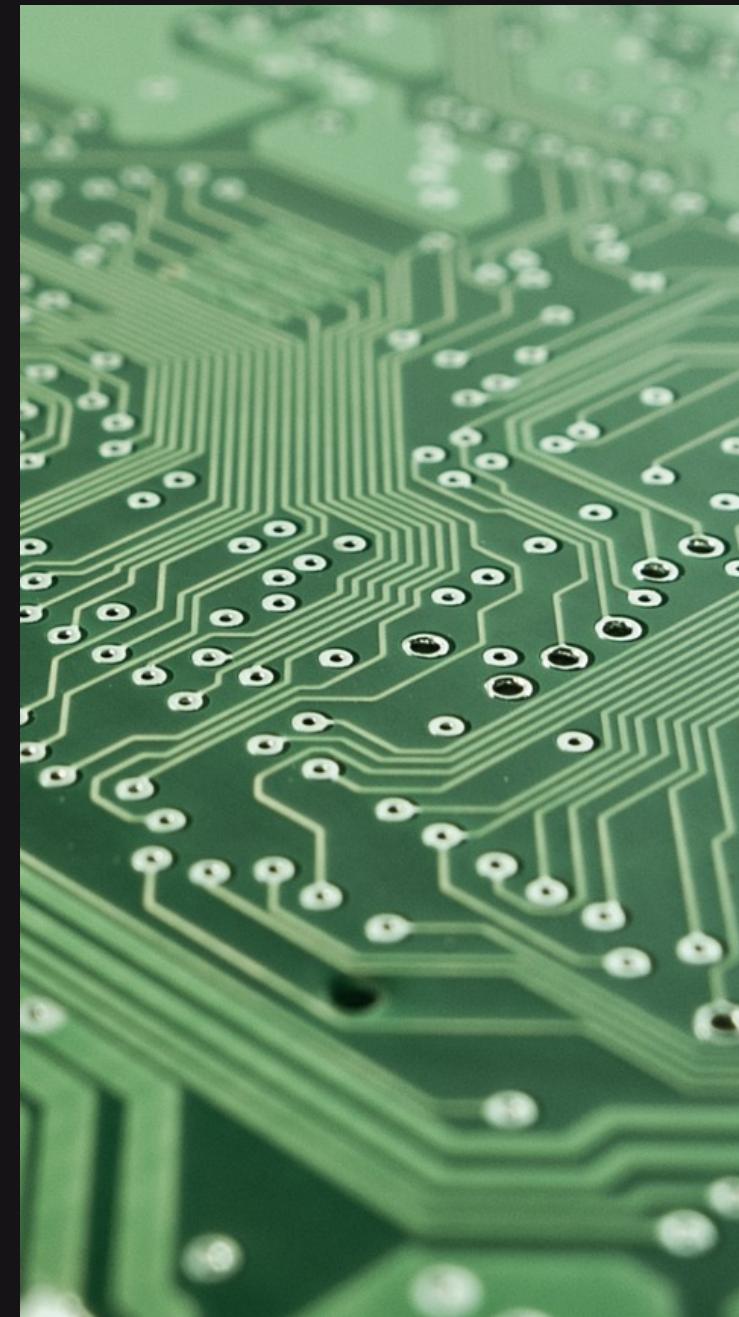
Training GPT-3 on Single GPU



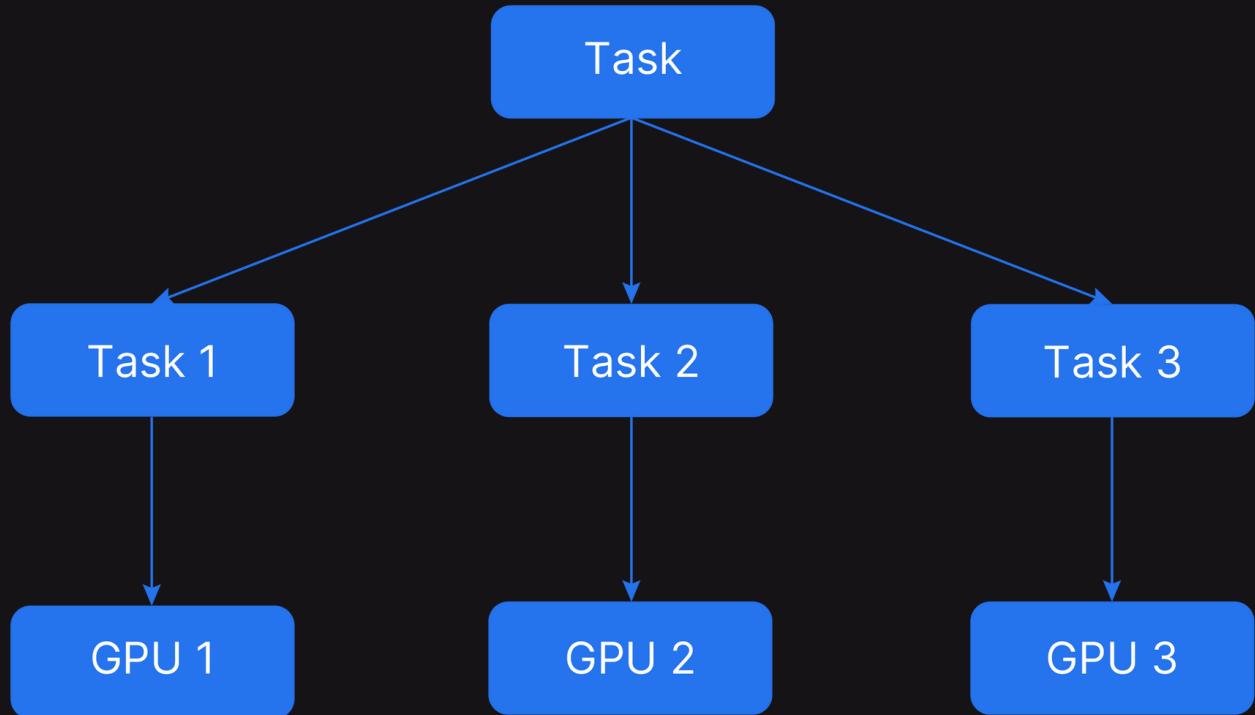
GPT-3

Training GPT-3 on a single
GPU would take **355 years**

Parallel and Distributed System



What is a Parallel and Distributed System?



Parallel and Distributed Systems



Accelerates the Training
Process



Memory Efficient

Different Parallel and Distributed Techniques

- Data parallelism
- Model parallelism
- Pipeline parallelism
- Fully Sharded Data Parallelism/ZeRO III
- Tensor parallelism
- 2D and 3D parallelism

Thank You
