

# 分布式集群系列

# 基本介绍



BUILDING A BETTER CONNECTED WORLD

Ascend & MindSpore

[www.hiascend.com](http://www.hiascend.com)  
[www.mindspore.cn](http://www.mindspore.cn)

# 关于本内容

## 1. 内容背景

- AI集群+大模型+分布式训练系统

## 2. 具体内容

- 分布式集群
- 分布式算法
- 分布式并行



# 关于本内容

## 1. 内容背景

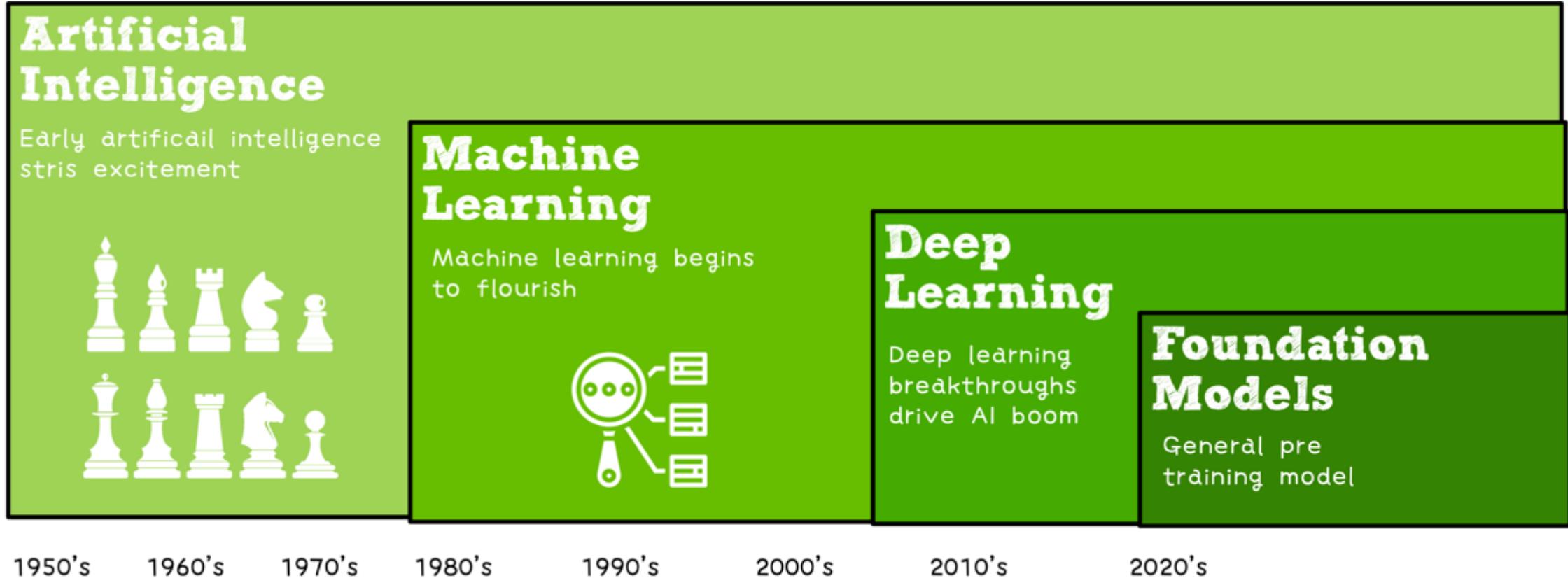
- AI集群+大模型+分布式训练系统

## 2. 具体内容

- **AI集群服务器架构**：参数服务器模式 – 同步与异步并行 - 环同步算法
- **AI集群软硬件通信**：通信软硬件实现 - 通信实现方式
- **分布式通信原语**：通信原语
- **框架分布式功能**：并行处理硬件架构 – AI框架中的分布式训练

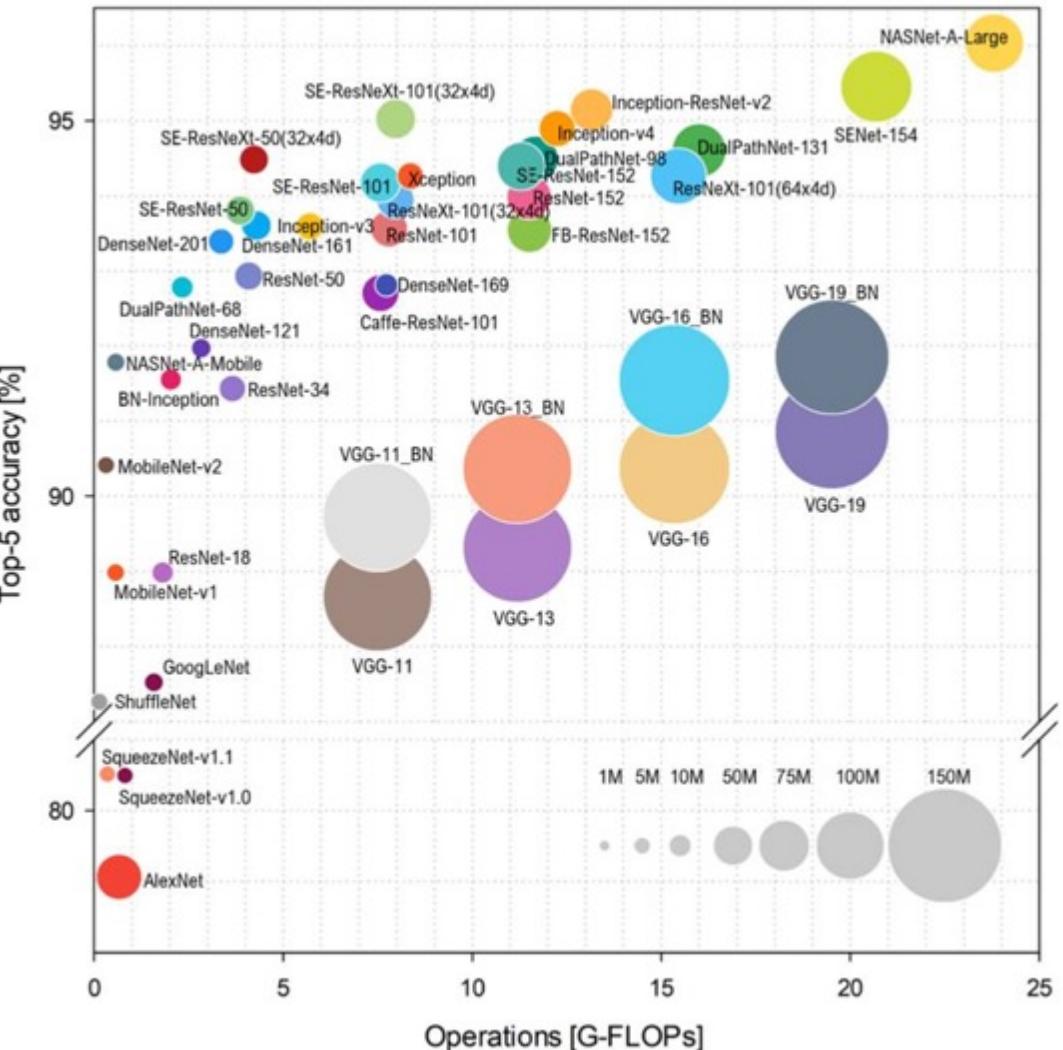
- **大模型算法**：挑战 – 算法结构 – SOTA大模型
- **分布式并行**：数据并行 – 张量并行 – 自动并行 – 多维混合并行

# 人工智能发展与大规模分布式训练关系



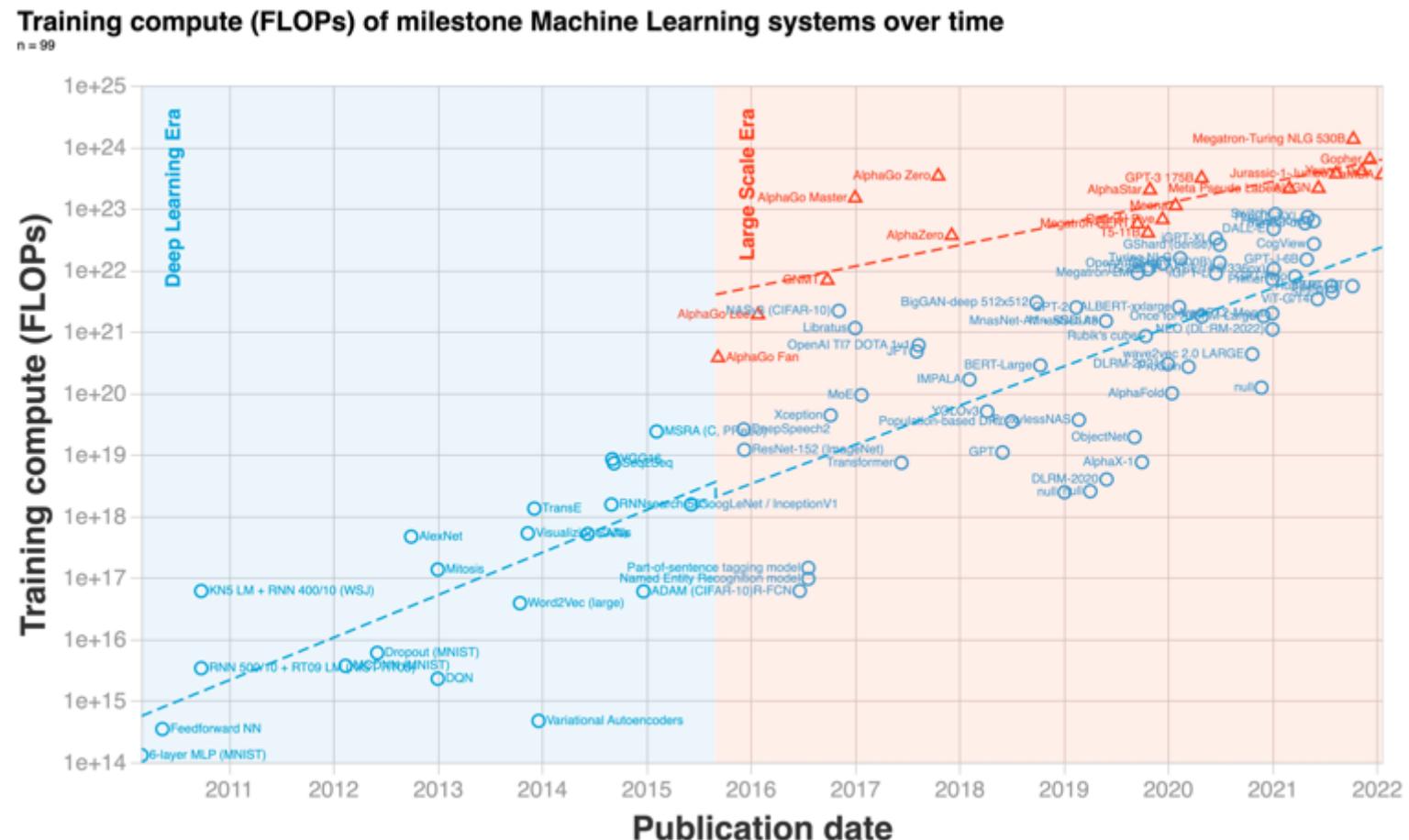
# 从串行计算到并行计算

- 深度学习训练的趋势
  - 更优的模型 -> 更大的计算复杂度
  - 深度学习训练需要巨大的并行能力
- 
- e.g. 语言模型：
    - BERT(Large) 用 V100 GPU 训练需时 >1个月



# 深度学习迎来大模型 ( Foundation Models )

1. 自监督学习方法，可以减少数据标注，降低训练研发成本
  2. 模型参数规模越大，有望进一步突破现有模型结构的精度局限
  3. 解决模型碎片化，提供预训练方案
- e.g. 语言模型 GPT-3
    - 8 张 V100，训练时长 36 年
    - 512 张 V100，训练近 7 个月



## Question?

- 一味让模型变大、参数量爆炸式增长，就是真的智能？
- 大模型能真正让机器迈向智能？

# 分布式深度学习的意义

- 深度学习训练耗时：

$$\text{训练耗时} = \text{训练数据规模} \times \text{单步计算量} / \text{计算速率}$$

模型相关，相对固定                    可变因素

- 计算速率：

$$\text{计算速率} = \text{单设备计算速率} \times \text{设备数} \times \text{多设备并行效率（加速比）}$$

Moore定律+算法优化                    可变因素

# 分布式深度学习的意义

- 深度学习训练耗时：

$$\text{训练耗时} = \text{训练数据规模} \times \text{单步计算量} / \text{计算速率}$$

模型相关，相对固定      可变因素

- 计算速率：

$$\text{计算速率} = \text{单设备计算速率} \times \text{设备数} \times \text{多设备并行效率（加速比）}$$

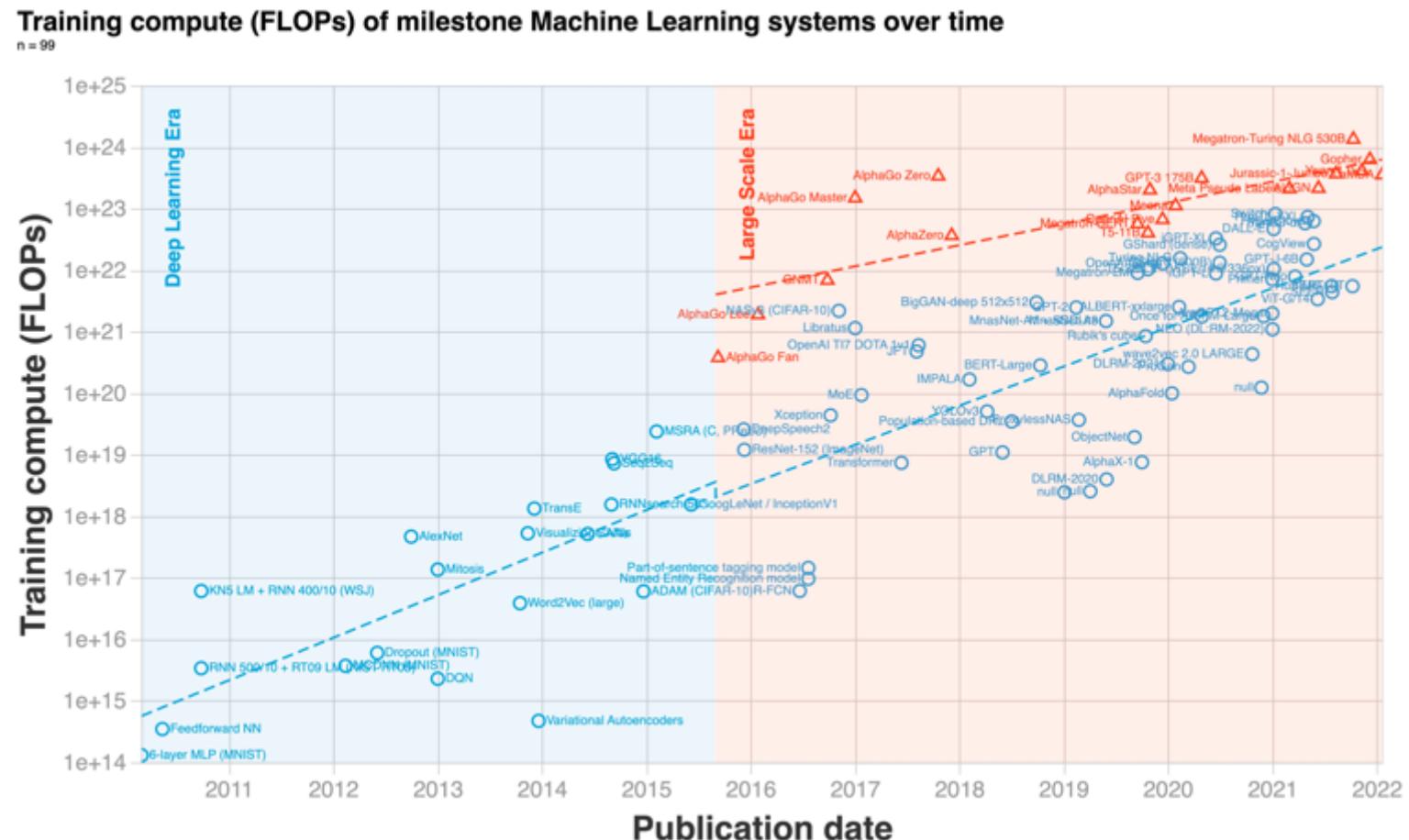
混合精度  
算子融合  
梯度累加

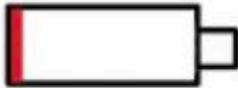
服务器架构  
通信拓扑优化

数据并行  
模型并行  
流水并行

# 深度学习迎来大模型 ( Foundation Models )

1. 自监督学习方法，可以减少数据标注，降低训练研发成本
  2. 解决模型碎片化，提供预训练方案
  3. 模型参数规模越大，有望进一步突破现有模型结构的精度局限
- 
- e.g. 语言模型 GPT-3
    - 8 张 V100，训练时长 36 年
    - 512 张 V100，训练近 7 个月





你的时间

不看结果  
注重过程

后天上线

明天答辩

梯度检查点

Gradient Checkpointing

梯度累加

Gradient Accumulation

混合精度训练

Mixed Precision

为什么当算法工程师

Go to sleep



洗洗睡吧  
Go to sleep

酷睿i3

V100

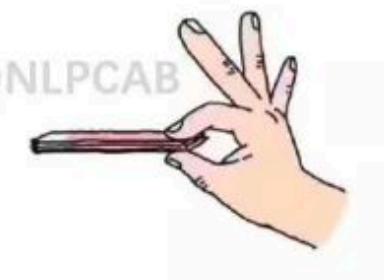
TPU

你的钱

分布式训练  
Distributed Training

并行+加速优化器  
LAMB

@NLPCAB



# AI 集群





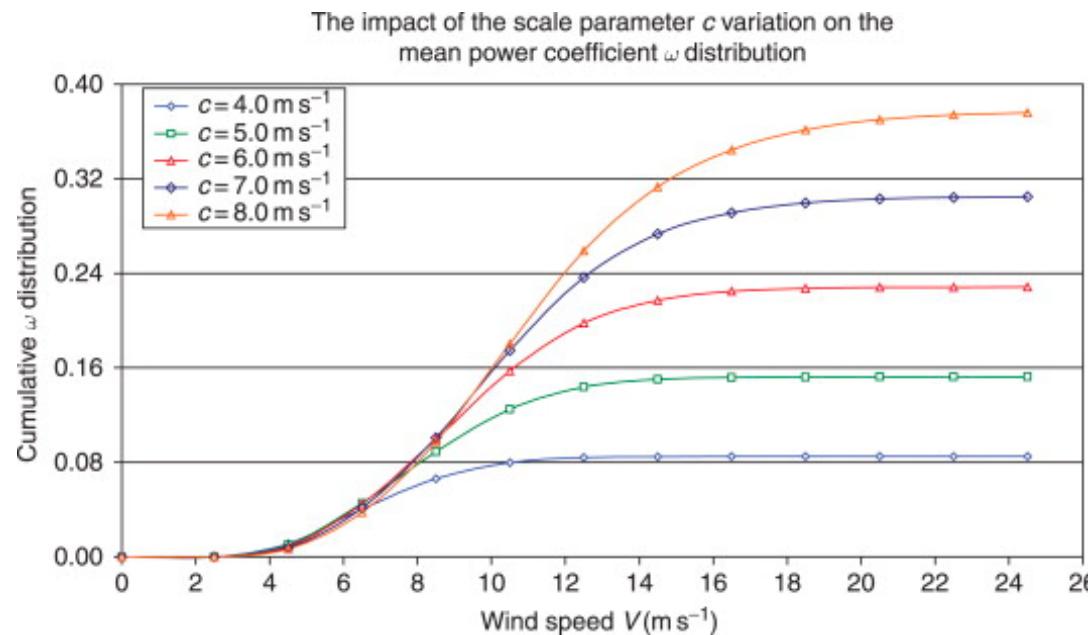




# 加速比

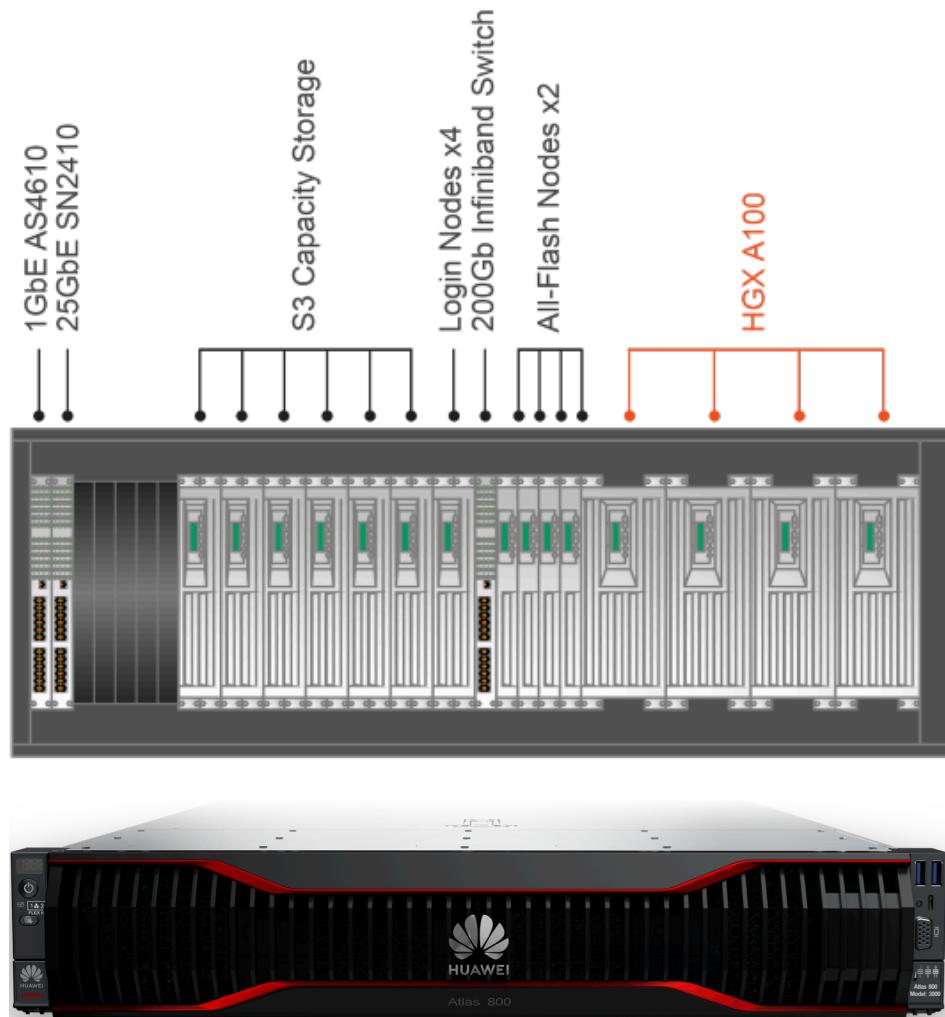
假设单设备吞吐量为 $T$ ， $n$ 个设备系统的吞吐量应为 $nT$ ，系统实际达到吞吐量为 $Tn$ ，则加速比为：

$$scale\ factor = \frac{T_n}{nT}$$



边际效应受限

# 通讯硬件

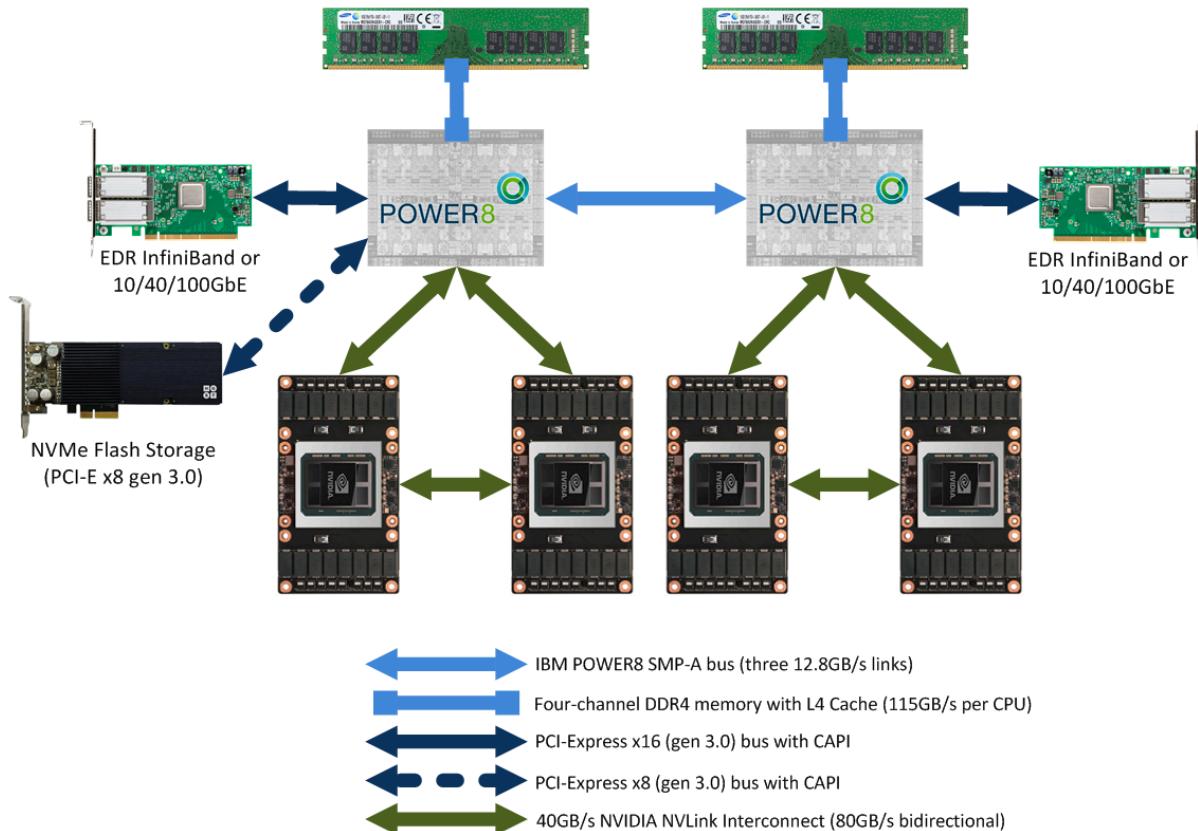


- 机器内通信
  - 共享内存
  - PCIe
  - NVLink (直连模式)
- 机器间通信
  - TCP/IP网络
  - RDMA网络 (直连模式)

# 通讯硬件

## Server Block Diagram

Microway OpenPOWER Server with NVIDIA Tesla P100 NVLink GPUs



### 机器内通信

- 共享内存
- PCIe
- NVLink (直连模式)

### 机器间通信

- TCP/IP网络
- RDMA网络 (直连模式)

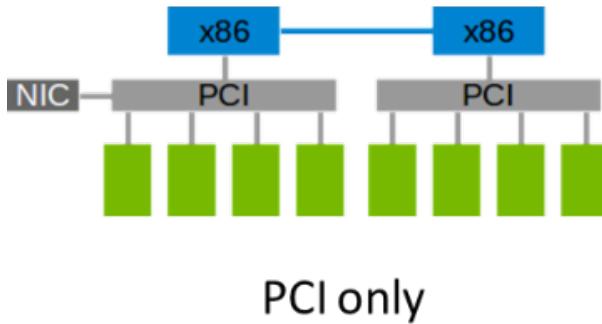
# 通信软件：提供集合通信

- **MPI**

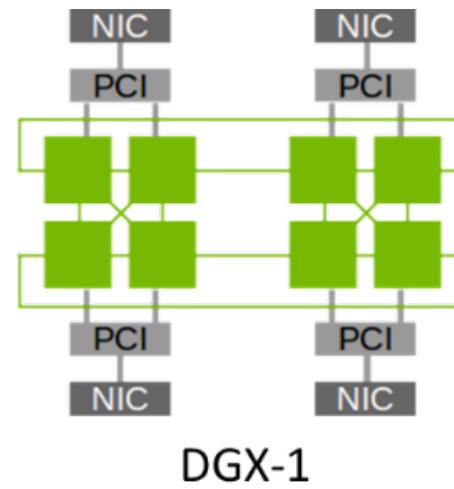
- 通用接口，可调用 Open-MPI, MVAPICH2, Intel MPI, etc.

- **NCCL / HCCL**

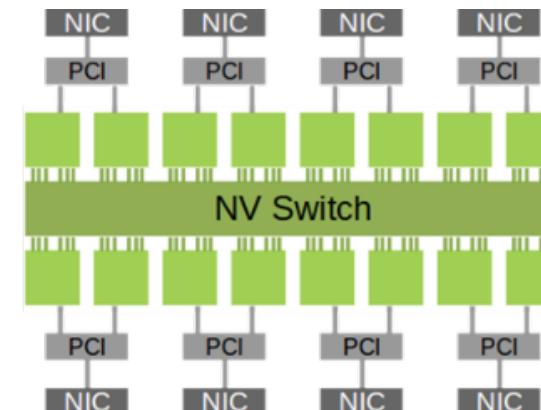
- GPU通信优化，仅支持集中式通信



PCI only



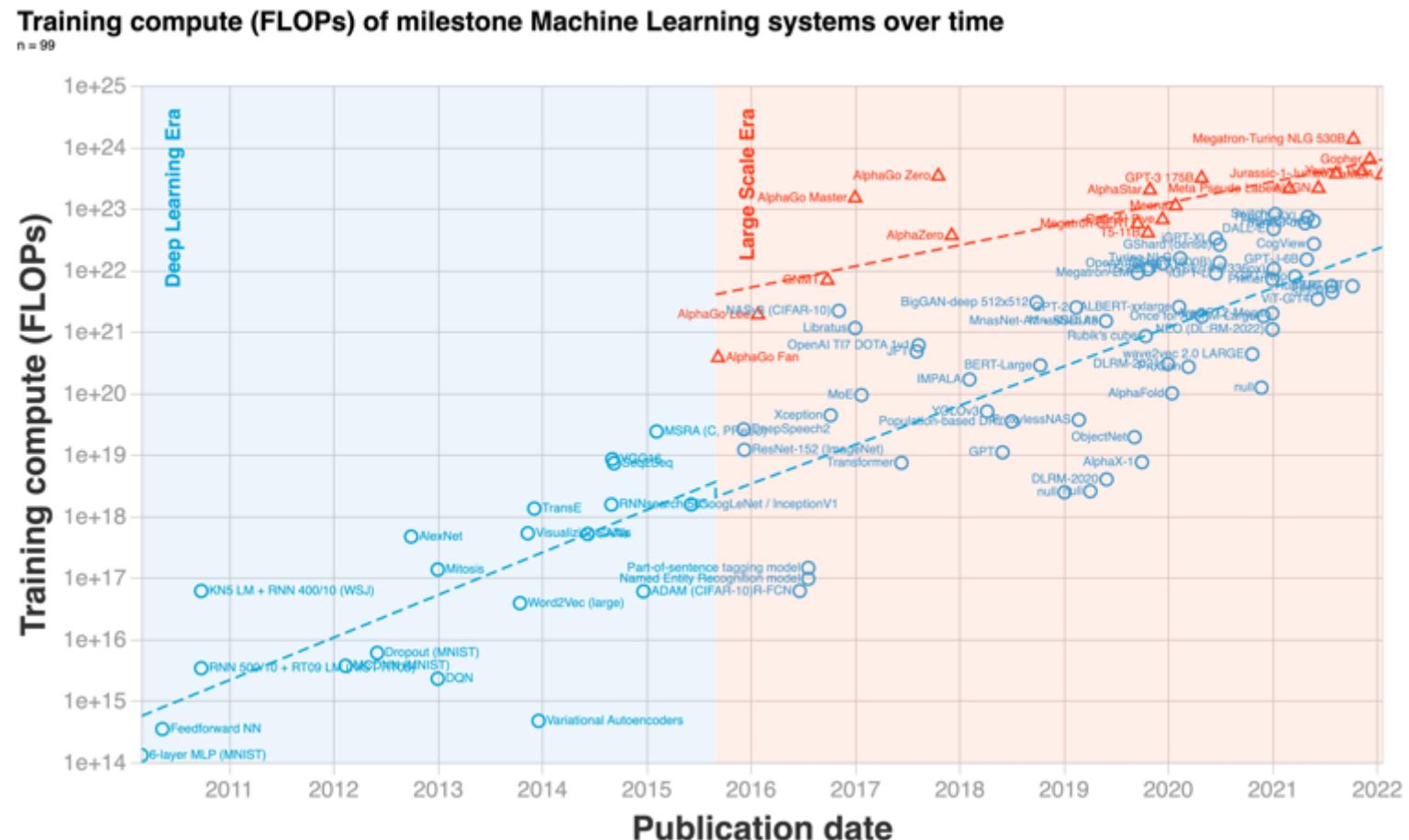
DGX-1



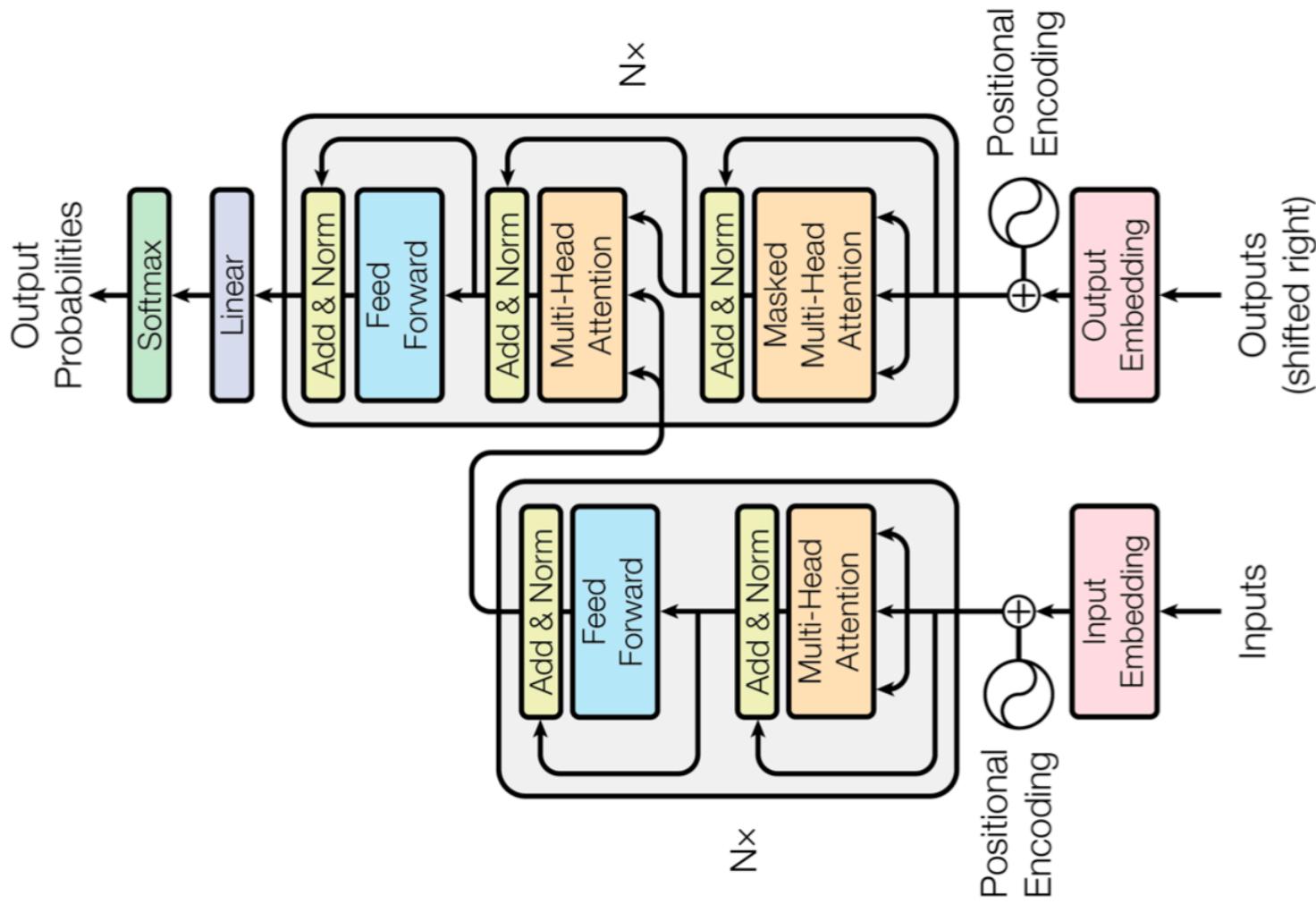
DGX-2

# 深度学习迎来大模型 ( Foundation Models )

1. 自监督学习方法，可以减少数据标注，降低训练研发成本
  2. 解决模型碎片化，提供预训练方案
  3. 模型参数规模越大，有望进一步突破现有模型结构的精度局限
- 
- e.g. 语言模型 GPT-3
    - 8 张 V100，训练时长 36 年
    - 512 张 V100，训练近 7 个月

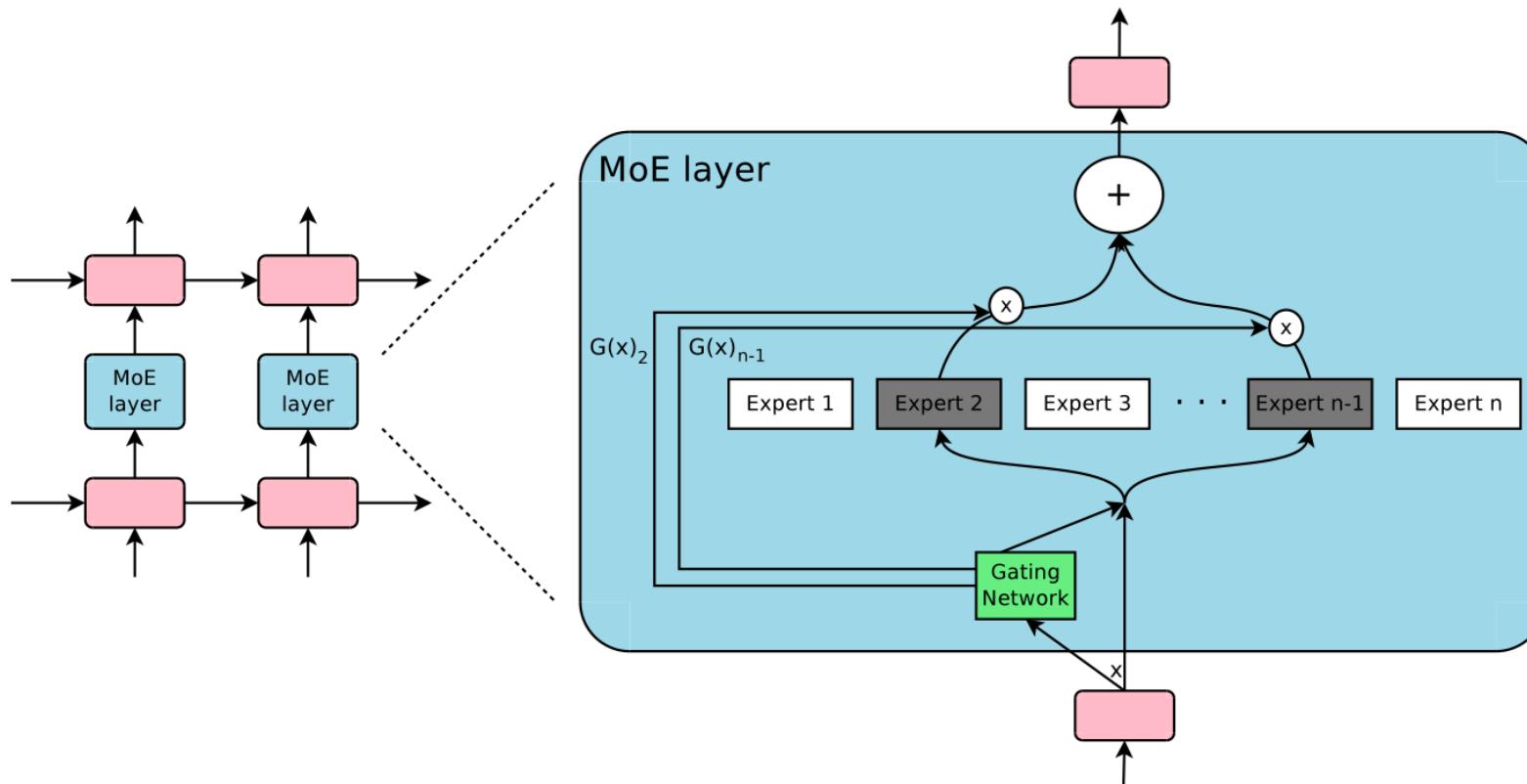


# Transformer 取代RNN、CNN进入大模型时代



# MoE 稀疏混合专家结构模型参数量进一步突破

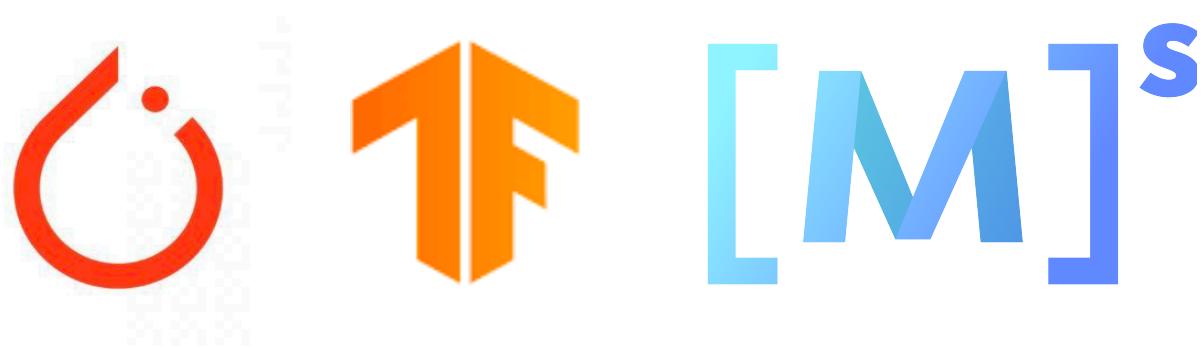
稀疏门控专家混合模型（ Sparsely-Gated MoE ）：旨在实现条件计算，即神经网络的某些部分以每个样本为基础进行激活，作为一种显著增加模型容量和能力而不必成比例增加计算量的方法。



# 分布式训练系统

定义：能够分布式地执行深度学习的训练的系统

- 分布式用户接口
  - 用户通过接口，实现模型的分布化
- 执行单节点训练
  - 产生本地执行的逻辑
- 通信协调
  - 实现多节点之间的通信协调



意义：提供易于使用，高效率的分布式训练

# 分布式训练系统



# Summary

1. 了解分布式训练的总体内容
2. 了解了深度学习为什么需要大规模分布式训练





BUILDING A BETTER CONNECTED WORLD

THANK YOU

Copyright©2014 Huawei Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

# 引用

1. [Introduction to Parallel Computing Tutorial](#)
2. [Seppo Linnainmaa, Algoritmin kumulatiivinen pyoristysvirhe yksittäisten pyoristysvirheiden taylor-kehitelmana](#)
3. [Benchmark Analysis of Representative Deep Neural Network Architectures](#)
6. [NVIDIA Tensor Core GPUs Train BERT in Less Than An Hour](#)
7. [Large Batch Optimization for Deep Learning: Training BERT in 76 minutes](#)
8. [Joseph E. Gonzalez AI-Systems Distributed Training](#)
4. [阿姆达尔定律](#)
5. [Gustafson定律](#)