

并行计算与分布式系统教学 教案

制作人：
时 间：



目录

- 第1章 第1页 : 并行计算与分布式系统的概念
- 第2章 第2页 : 并行计算与分布式系统的区别
- 第3章 第3页 : 并行计算的资源限制
- 第4章 第4页 : 分布式系统的资源利用
- 第5章 第5页 : 并行计算和分布式系统的应用



• 01

第1章 并行计算与分布式系统的概述



并行计算与分布式系统的概念

并行计算指的是在同一时刻，多个处理器或者多个核心同时处理多个任务。分布式系统是指将计算任务分配到多台计算机上进行协同工作的系统。两者的区别在于并行计算更侧重于同一计算机内的多个处理器或核心，而分布式系统则是在不同计算机间进行协作。



并行计算的特点

速度提升

通过同时处理多个任务，大大提高计算速度

提高资源利用率

充分利用多个处理器或核心，提高资源利用率

可扩展性

能够根据需求灵活扩展处理能力



分布式系统的特点

可扩展性

能够根据需求灵活
扩展系统规模

一致性

所有节点在同一时
间看到相同的数据

容错性

即使某台计算机出
现故障，系统仍能
正常运行



并行计算与分布式系统的应用领域

01 科学计算

用于模拟复杂的自然现象和科学问题

02 大数据处理

处理海量数据，提取有用信息

03 人工智能

加速机器学习和深度学习过程



并行计算与分布式系统的应用优势

并行计算

高性能
实时计算
任务划分明确

分布式系统

容错性强
资源共享
灵活可扩展

互补优势

结合并行计算和分布式系统的优势，能够应对更复杂的计算任务



第2章 并行计算模型



workflow 模型

workflow 模型是一种并行计算模型，它将任务分解成一系列的阶段，每个阶段都可以使用并行计算来优化执行时间。这种模型通常用于处理离线批量数据处理任务，如日志分析和数据挖掘。



MapReduce 模型

MapReduce 模型是一种广泛应用于分布式系统的并行计算模型，它将大规模的数据处理任务分解成许多小的子任务，并在多个计算节点上同时执行，最后将结果汇总。这种模型通常用于处理大规模的数据集，如网页搜索和分布式机器学习。



actor 模型

基本概念

Actor、消息、邮
箱

优点

可扩展性、灵活性

缺点

复杂性、调试难度

应用场景

网络服务器、并行
计算



MPI 模型

基本概念

进程、通信、拓扑

优点

高效性、精度

缺点

编程难度、可移植性

应用场景

高性能计算、科学计算



并行计算模型比较

模型

workflow模型
 MapReduce 模型
 actor 模型
 MPI 模型

优点

简单易用、可视化设计
 自适应调度、可扩展性
 灵活性、并发性
 高效性、精度

适用场景

批量数据处理
 大规模数据集处理
 网络服务器、并行计算
 高性能计算、科学计算

缺点

可扩展性有限、调度复杂
 数据倾斜、容错性差
 复杂性、调试难度
 编程难度、可移植性



第3章 分布式数据处理框架



概述

Apache Hadoop

Apache Hadoop 是一个分布式文件系统和分布式计算框架，旨在处理大规模的结构化数据。



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/177126063103006114>