

第八章

第8章：8.1节、8.2节

数据密集型计算存储系统（高性能分布式文件系统和存储云、NoSQL）、编程平台（MapReduce编程模型、MapReduce变形和扩展）

8.1 什么是数据密集型计算

p 181

- 数据密集型计算的定义
- 元数据定义
- 数据密集型计算应用领域

8.1.1 数据密集型计算特性

p182

8.1.2 未来的挑战

p182

8.1.3 历史背景

8.2 数据密集型计算技术

8.2.1 存储系统

p186

- 导致数据库、数据管理行业研究的转折点的因素
- 数据密集型计算的主要方向构成

1. 高性能分布式文件系统和存储云

p187~p188

- 常见的高性能分布式文件系统

2. NoSQL系统

P189

- 什么是NoSQL系统？
- 决定NoSQL运动发展的两个因素
- NoSQL实现分类
- 一些常见的支持数据密集型应用程序的显著成就

8.2.2 编程平台

1. MapReduce编程模型

p193

- MapReduce模型的输入输出是什么？
- MapReduce计算的参考工作流程

p194~195

- MapReduce灵活性的例子
- 能够以MapReduce计算型是表达的两个阶段
- MapReduce如何给开发者使用
- 两个运行在分布式基础设施上的进程
 - 主进程负责的工作
 - 工作进程负责的工作

P196

- 进程输入输出缓冲区的作用
- 主进程、工作进程节点处理方式

2. MapReduce的变形和扩展

P196

- 变形和扩展的目的

p197

- 和MapReduce相近的模型以及和MapReducue不同的点