### 6.23

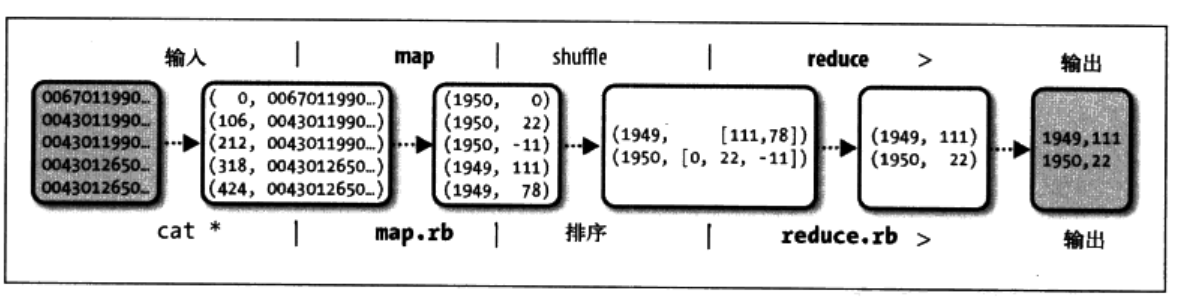
1813091 邢钦云

1. MapReduce
2. 使用unix tools分析数据
3. 使用hadoop分析数据

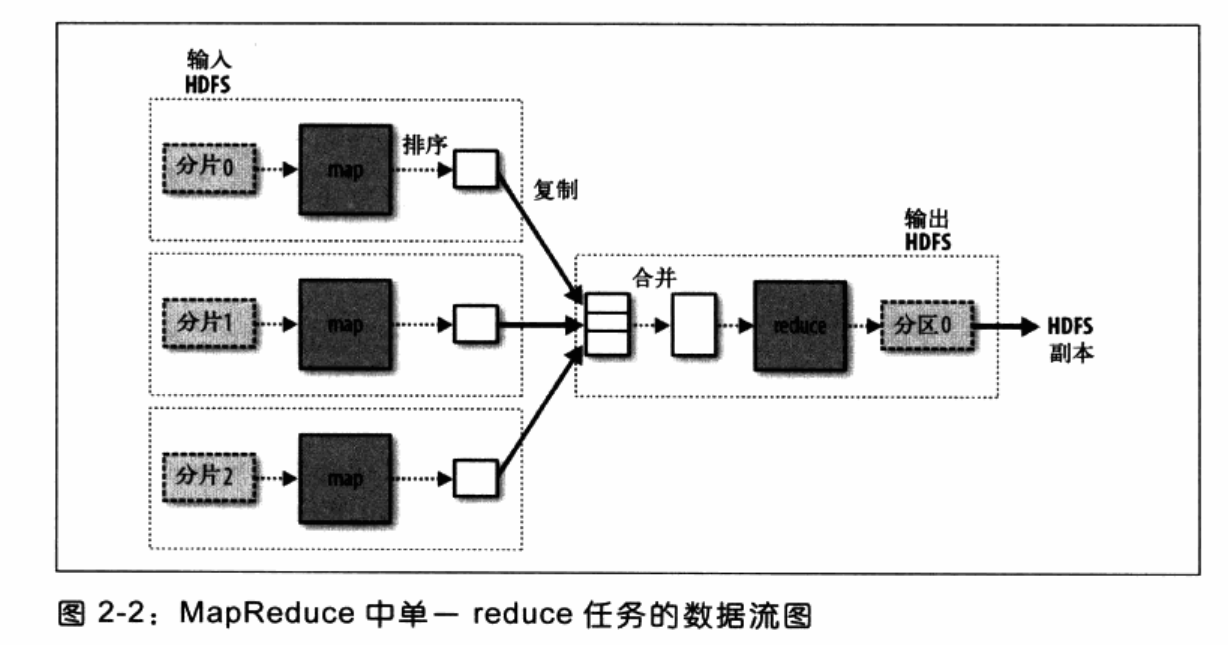
Map函数：数据准备

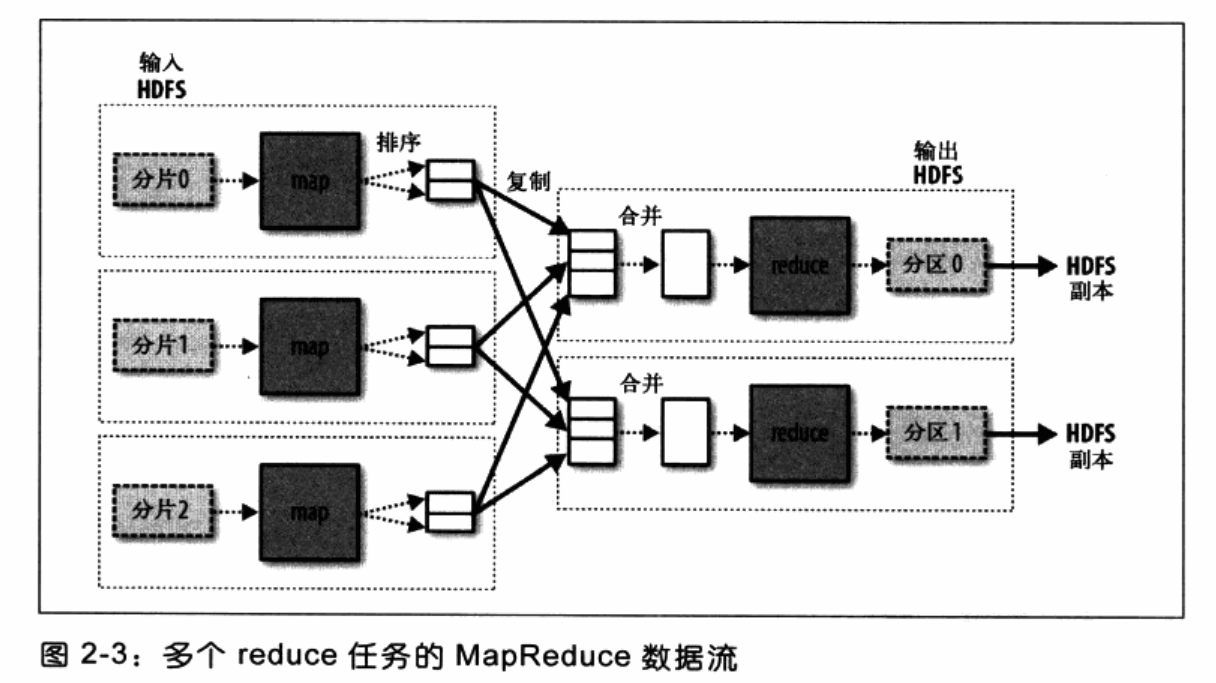
Reduce函数：进一步处理数据

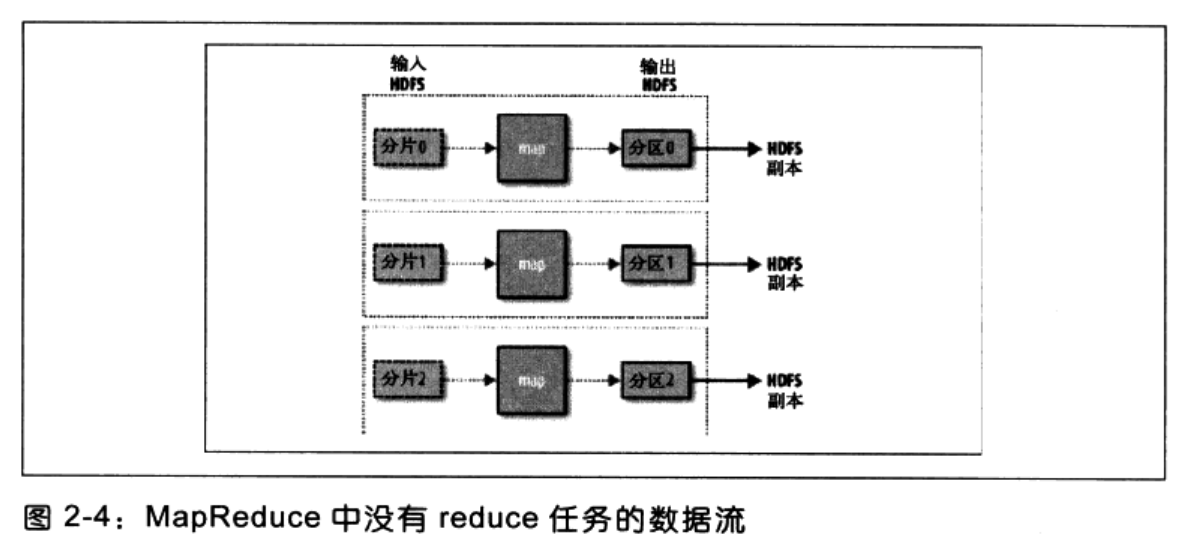
Mapreduce的逻辑数据流



1. 分布化
2. 数据流





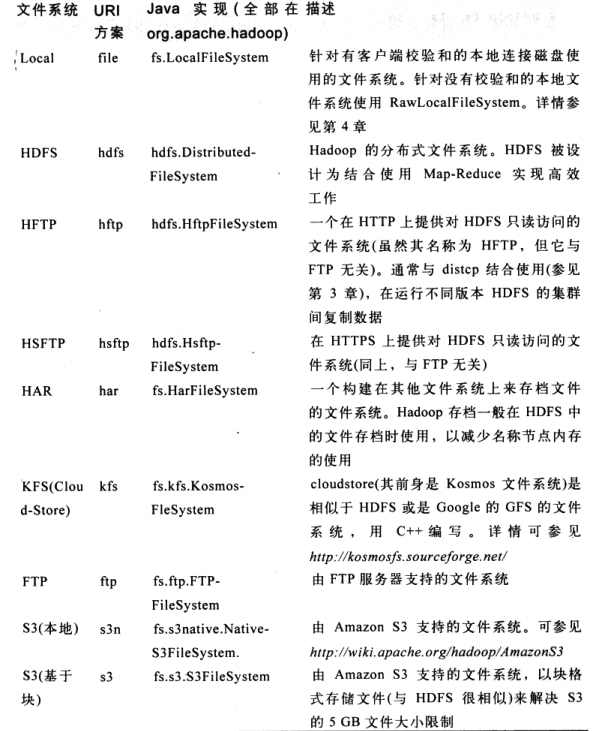


1. 定义combiner
2. Hadoop流

Hadoop提供了一个API来运行mapreduce。Hadoop流使用unix标准流作为hadoop和程序之间的接口，所以可以使用任何语言。

流适用于文字处理

1. Hadoop管道（c++接口）
2. Hadoop分布式文件系统（HDFS）
3. HDFS的设计
4. HDFS的概念
5. 块
6. 名称节点和数据节点
7. 命令行接口
8. 文件系统



1. Java接口
2. 数据流
3. 通过distcp进行复制
4. Hadoop归档文件