

# 大数据开发技术

东北林业大学

卢洋



# 第一章

## MapReduce概述



1.2

MapReduce优缺点



# 1 优点

## 👁 MapReduce易于编程

它简单的实现一些接口，就可以完成一个分布式程序。这个分布式程序可以分布到大量廉价的PC机器上运行；也就是说，编写一个分布式程序，跟写一个串行程序是一模一样的。就是因为这个特点，使得Hadoop编程变得非常流行。

## 👁 良好的扩展性

当计算资源不能满足的时候，可以通过简单增加机器来扩展集群的计算能力。



# 1 优点

处理大规模  
静态数据

## ● 高容错性

**MapReduce**的设计初衷就是使程序能够部署在廉价机器上，这就要求它具有很高的容错性。比如，其中一个机器挂掉，它可以上面的计算任务转移到另外一个节点上运行，不至于这个任务运行失败，而且这个过程无需人工参与，完全是由**Hadoop**内部完成的。

## ● 适合PB级以上海量数据的离线处理

可以实现上千台服务器集群并发工作，提供数据处理能力。



## 2 缺点

### 1. 不擅长实时计算

**MapReduce**无法像**MySQL**一样，在毫秒或者秒级内返回结果。

### 2. 不擅长流式计算

流式计算的输入数据是动态的，而**MapReduce**的输入数据是静态的，不能动态变化。这是因为**MapReduce**自身的设计特点决定了数据源必须是静态的。

### 3. 不擅长DAG(有向图)计算

多个应用程序存在依赖关系，后一个应用程序的输入为前一个程序的输出。在这种情况下，**MapReduce**并不是不能使用，而是使用后，每个**MapReduce**作业的输出结果都会写入到磁盘，会造成大量的磁盘IO，导致性能的低下。