

# 大数据开发技术

东北林业大学

卢洋



# 第三章

## MapReduce 框架原理



3.3.2

Partition 分区



# Partition分区

## 利用哈希值分区

### 1. 问题引出

要求将统计结果按照条件输出到不同文件中(分区), 比如: 将统计结果按照手机归属地不同省份输出到不同文件(分区)中.

### 2. 默认partition分区

```
public class HashPartitioner<K,V> extends Partitioner<K,V>{  
    public int getPartitions(K key, V value, int NumReduceTasks) {  
        return (key.hashCode() & Integer.Max_VALUE) %  
numReduceTasks;  
    }  
}
```

默认分区是根据key的hashCode对ReduceTask个数取模得到的, 用户无法控制哪个key存储到哪个分区.



