

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

自定义Partition分区

1. 自定义类，继承Partitioner，重写getPartition()方法

```
public class CustomPartitioner extends Partitioner<Text , FlowBean>{  
    @override  
    public int getPartition(Text key, FlowBean value, int numPartitions){  
        // 控制分区代码逻辑  
        ...  
        return partition;  
    }  
}
```

2. 在Job驱动类中，设置自定义Partitioner

```
job.setPartitionClass(CustomPartitioner.class);
```

3. 自定义Partitioner类之后，要根据自定义的Partitioner的逻辑设置相应数量的ReduceTask

```
job.setNumReduceTasks(5);
```