# 小组第一次讨论

时间：2020/11/16

成员: 1.徐莘杰（组长）2018081065

2.邓棋 2018081062

3.孟令迪 2018081050

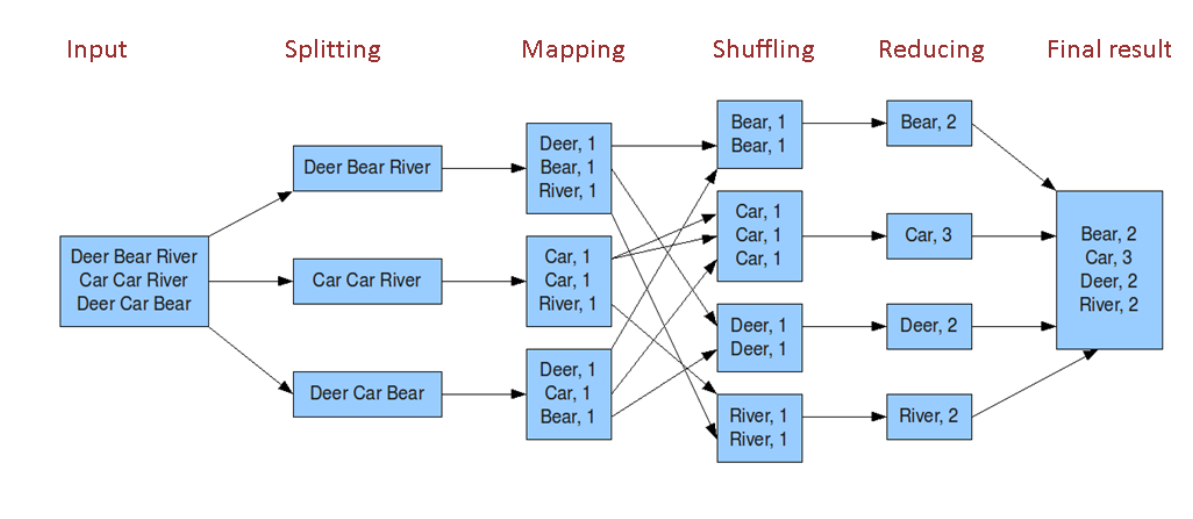
地点：寝室

讨论主题：Mapreduce计算原理或流程

讨论结果：

1. Mapreduce执行流程…

如果一切都按部就班的进行，那么整个作业的计算流程，应该是作业的提交 -> Map任务的分配和执行 -> Reduce任务的分配和执行 -> 作业的完成。而在每个任务的执行中，又包含输入的准备 -> 算法的执行 -> 输出的生成，三个子步骤。

1. Mapreduce计算原理  
   
2. Mapreduce优缺点

（1）优点  
1>易于编程：以普通程序的编程方法加上使用MapReduce提供的接口，可以快速完成分布式程序的编写。  
2>良好的扩展性：计算资源得不到满足时，可以通过简单的增加计算机器来扩展计算能力  
3>高容错性：如果一个任务所在计算节点挂了，上面的计算任务可以自动转移到另外的节点上执行，即故障自动转移，这个过程是内部完成的，无需人工干预  
4>适合PB级别以上数据的离线处理

（2）缺点  
1>实时计算：无法像mysql一样在毫秒级或者秒级返回计算结果  
2>流式计算：流式计算的输入数据是动态的，而MapReduce要求输入数据是静态的，已经持久化在存储上的。  
3>DAG（有向无环图）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的输出，这种情况下，MapReduce的性能很低。因为MapReduce的每个阶段的输出结果都会先写入到磁盘中，大量的磁盘IO会造成性能的急剧下降。

讨论相关照片



