# 实训个人学习报告

-----白文亮 0171121683

## 一丶HDFS

1. hdfs的基本概念：

HDFS：Hadoop Distributed File System ，Hadoop分布式文件系统。主要用来解决海量数据的存储问题。

1. hdfs的写文件过程

将文件存到HDFS集群中

1. hdfs读文件过程

将集群中的数据读出

1. hdfs的优缺点：

优点：

* + 1. 高容错性
    2. 适合批处理
    3. 适合大数据处理
    4. 流式文件访问
    5. 可构建在廉价机器上

缺点：

1. 低延时数据访问
2. 小文件储存
3. 并发写入，文件随机修改

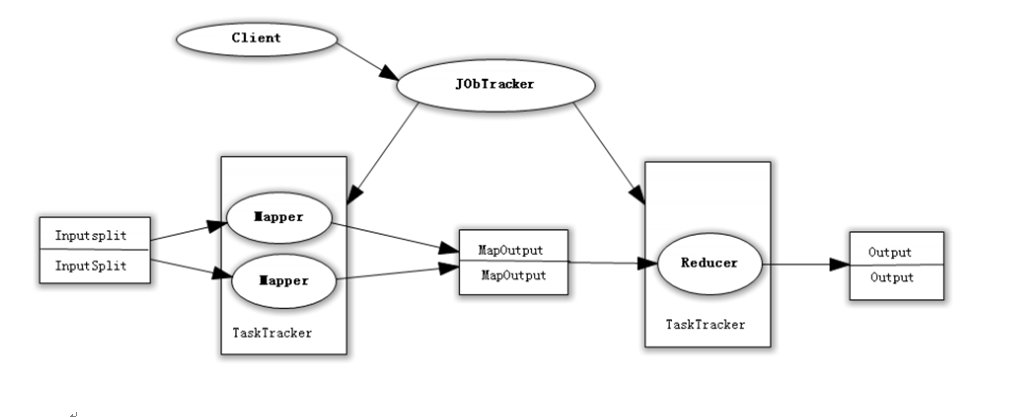
## 二丶MapReduce

1. MapReduce的概念：

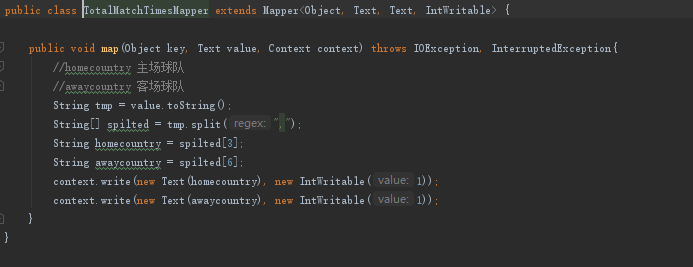
MapReduce是一种分布式计算模型，由Google提出，主要用于搜索领域，解决海量数据的计算问题。MR由两个阶段组成：Map和Reduce，用户只需要实现map()

和reduce()两个函数，即可实现分布式计算。这两个函数的形参是key、value对，表示函数的输入信息。下面是MR的执行流程:

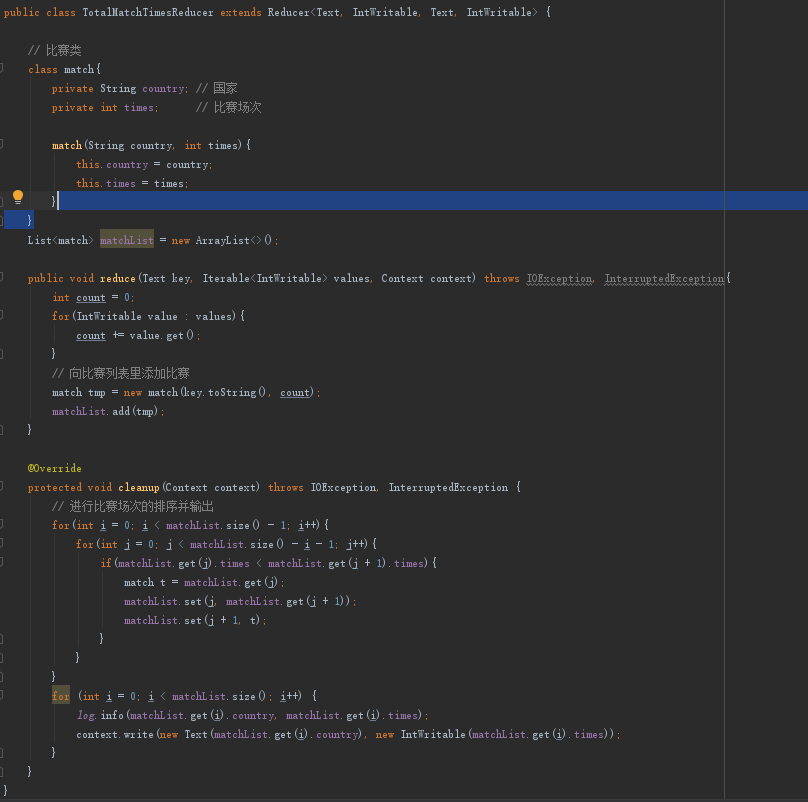
1. 流程图



1. Map任务：
   * 1. 读取从hdfs中读取的内容，解析成，<key,value>对，对输出的每一行数据进行解析，每一个键值对调用一次map函数。
     2. 编写代码实现逻辑，对输入的<key,value>进行处理，转换成新的<key,value>对输出。
     3. 对输出的<key,value>进行分区，同一key值得键值对存放在一个集合中。
     4. 代码示例：



1. Reduce任务：
   * 1. 对多个map任务的输出，按照不同的分区放到不同的reduce节点
     2. 对多个map任务输出的键值对进行合并排序等操作。通过写reduce函数增加自己的逻辑，对输入的<key,value>进行处理，转换成新的<key,value>输出
     3. 讲reduce中处理的数据输出保存到文件中。
     4. 代码示例：



## 三丶Echarts

1. Echarts的概念:

ECharts 是一个使用 JavaScript 实现的开源可视化库，涵盖各行业图表，满足各种需求。

ECharts 遵循 Apache-2.0 开源协议，免费商用。

ECharts 兼容当前绝大部分浏览器（IE8/9/10/11，Chrome，Firefox，Safari等）及兼容多种设备，可随时随地任性展示。

1. 引入Echarts：

<script src="echarts.min.js"></script>

1. Echarts特性
   1. 丰富的可视化类型
   2. 多种数据个数无需转换直接使用
   3. 千万数据的前端展现
   4. 移动端优化
   5. 多渲染方案，跨平台使用
   6. 深度的交互式数据探索
   7. 多维数据的之处以及丰富的视觉编码手段
   8. 动态数据
   9. 绮丽的特效
   10. 通过GL实现更多更强大的三维可视化