问题清单  
  
文字问题：  
1、在1.1中的第一句话“测试”后边的逗号应该取消，否则语句不通。  
2、1.1中“本测试文档主要用于测试...”表述不准确。本文档应该是用于介绍相关于测试的一些工作和说明，用于测试的应该是具体的用例，不是文档。

3.3.2.1 "本测试的目的在于测试对爬取数据执行的预处理过程将数据成功的进行了格式化。"这句话不通顺,建议修改

4.4.1.1 "为了测试我们的Map-Reduce函数在修改了代码之后能够以更快的速度运行，所以进行该测试",语句不通顺,略显啰嗦.

1. 描述含糊不清：1）关键数据（如CPU，硬盘等），其中“等”的部分具体指什么？测试应该明确具体包含什么内容，不能省略掉，如果没有其他内容，应该写清楚只有CPU，硬盘两项。此外，CPU的数据包括哪些，硬盘数据包括哪些需要列出，比如CPU包括核心数和主频，硬盘包括转速和容量。2）4.1 "速度明显得到了改善",明显是多明显?应该予以量化。3）2.1.2测试用例中的Basic Flow的“用Shell命令测试环境”与后边Postcondition的“Linux安装成功”关系描述不准确。如果只是测试，那么不能确定安装一定成功。可能测试失败，安装就失败了。后边的几个测试的用例在这部分的描述上都有类似的问题。4）2.1.2测试用例中的Specific Alternative Flows 的“重新安装”和后边Postcondition的“安装失败”的关系不清晰。2.3.2也有类似的问题。

## 格式问题：

1. 目录格式有些错误，例如突然缩进（见2.1.2测试用例等），或者“3.2.1测试用例”写为“2.2.1测试用例”，实现需求模块的“2.1.2”节号有误，应该是3.1.2。
2. 3.2.2  表中第一行和第二行字体大小不一样
3. 中英文混杂：3.2.2 中“PrimaryActor”一栏为“Tester”，而其余都是“测试员”

## 技术问题：

1. 检查Linux是否正确安装，范围过于宽泛。因为Linux操作系统包括的内容很多，而测试用例看起来只是测试jvm有没有正确安装。显然后面的测试中还用到了python环境，如果python没有安装而jvm安装了，本用例会通过测试，但是后面的测试的precondition并不被满足  
   2.测试用例没有给出具体的测试环境，包括操作系统版本，jvm版本，Hadoop版本等，以及测试用的shell命令。  
   3.Hadoop源码编译失败并不能说明Maven环境没有安装成功，也可能是Hadoop源代码的问题，或者Linux系统本身的问题。
2. 数据爬取测试中，没有给出如何判断爬取结果正确完整的方法，是人工对比爬取数据与原始数据，还是计算哈希值，还是通过程序逐字符对比。

测试用例应该给出具体的输入和输出。本用例中的第一步“确定要抓取的网页地址”，应该有具体的哪个或者哪些地址，如果正确执行应该得到什么样的输出，如果不一致，则说明程序存在问题。  
6.课程要求测试重点在扩展功能上，而本项目的扩展是为了在保证准确性的前提下提高计算速度，而“源码改进”部分的测试用例只测试了能否成功编译和执行速度，并不能说明其改进的正确性。  
7.对于性能的测试用例不足。性能的测试应该涵盖多种不同的情况，例如不同数据规模等，只通过一个测试用例不足以说明速度确实得到了提升

1. 未单独列出参考文献或引用文档

9.2.1.2测试用例中Precondition中应为“用Linux的iso光盘”  
10. 3.1.2测试用例中"[size=14.6667px]数据发生了缺省或者出现异常标签"作为爬去结果是否成功的表示，是否过于简单，爬取结果可能出现其他异常，另外爬取的数据的量有无预估。  
11.[size=14.6667px]对你们要改进的"InputSplit"[size=14.6667px]是否应该有专门的描述或讲解  
12.[size=14.6667px]. 文档概述部分貌似与实际内容不匹配

13.本次实验的内容只涉及到测试需求部分，所以标题应该是“测试需求规格说明书”。

14.3.1.1 "抓取的数据应该按照关键词和属性，有序的进行输出". 如何对关键词和属性进行排序呢?  
15.3.4.1 "异常情况下将出现排名错误" 什么是排名错误?请详细写出  
16.2、4.1 运行速度测试,在同样的测试环境下,同样的程序,多次测试可能会出现不同的结果.这个情况应如何处理?