刘教授即将出版的《大数据》和《大数据实验手册》从理论高度展现了大数据的魅力，尤其是教材紧密结合实践，让像我这样的外行，能够在几天的时间内学到了很多关于大数据的知识，不再对大数据望而生畏，而且有了对大数据跃跃欲试的信念。这两本教材对我的帮助，是任何其他教材都无法企及的。

谢谢刘教授，给我们带来如此好的学习大数据的机会。

下面是我读教材过程中，发现的几个小问题和提的小建议。我是外行，不太懂，不对的地方，请包涵：

# 关于《大数据》教材的几点小建议

73页图3-12排版到页底了

# 关于实验教材的几点小建议

建议1：文件所在的包，能一致的就一致，如有的地方用mr,有的地方用cn.store.mr

建议2：代码打包是否指定主类，统一一下：要么指定主类，要么不指定；

# 实验一



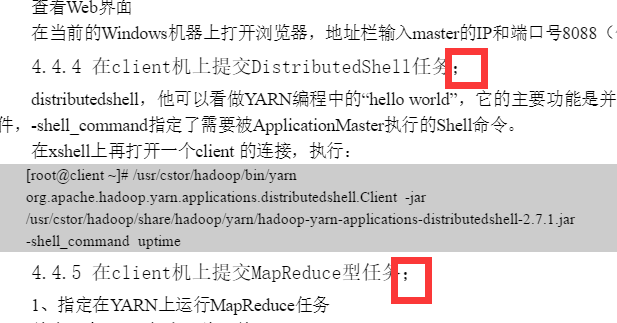


# 实验三

1）Eclips打包操作中，建议给出指定“主类”与否对后面执行jar命令的重要性，因为jar命令指定主类，而打包时也指定主类，代码中args[0][1]等所获取的值是不同的。有些人做实验时会出错。

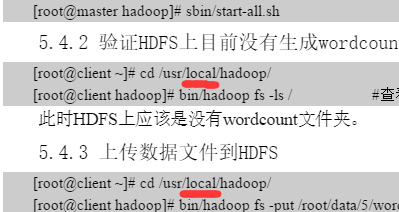
# 实验四

去掉分号



# 实验五

下图应为cstor，而不是local

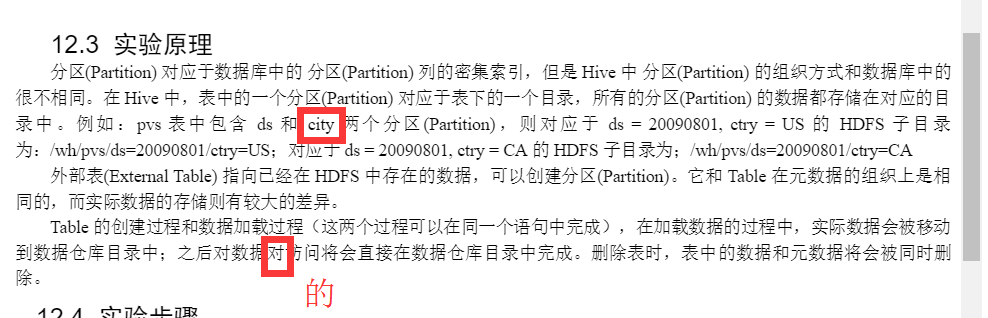


# 实验六

1. 题目改为“MapReduce实验：两关键字排序”是不是好些，即基于主键和从键的排序。
2. 第6.4.2的“打包提交”，如果与3.4.8的打包内容合并一下，是不是会好些。

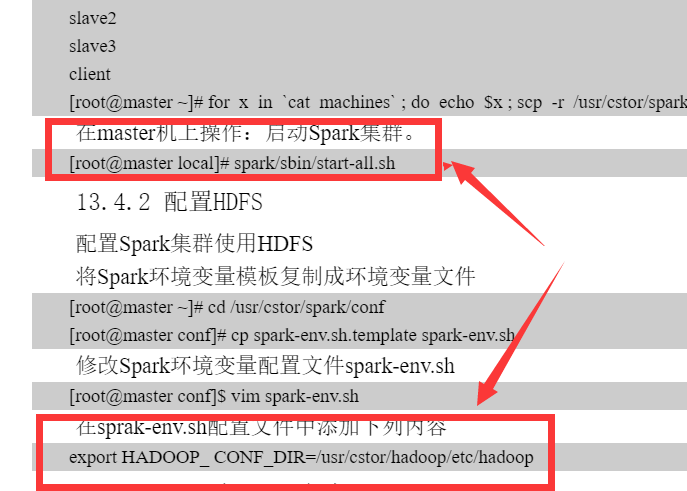
# 实验十二

下图中city是不是应该为ctry，这样就与后面的语句一致了。



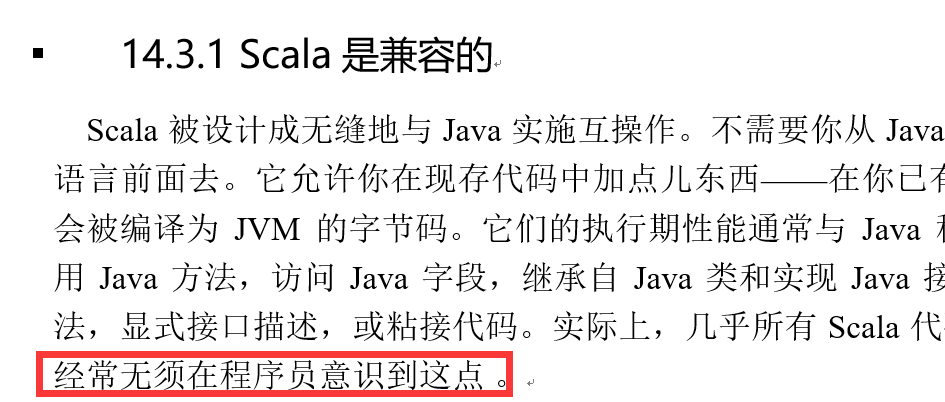
# 实验十三

下图中配置HDFS与启动Spark是否要调换一下顺序？



# 实验十四

安装Java时，安装路径不要带空格，如”program files”中的空格，否则intellij编译代码时，会出现line broken的错误。

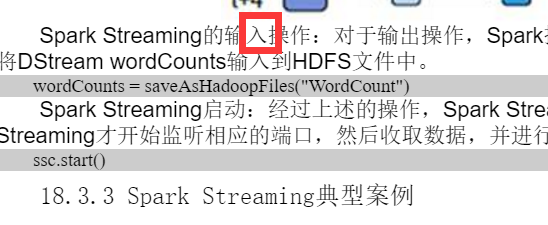
1. 下面这句话不通顺2）14.3.2 节最后一段倒数第3句话“而这些有可以被灵活地混合在一起。”中的“有”字是不是应该是“又”3）14.3.3节最后一段倒数第3句话“Scala里的函数式文本真地很轻量”，“真地”是否应该是“真的”

# 实验十六

实验十四也介绍了Scala,这里有部分内容与前面重复。

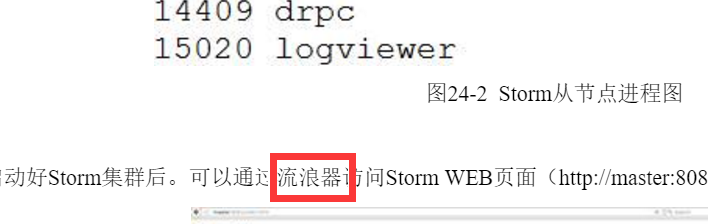
# 实验十八

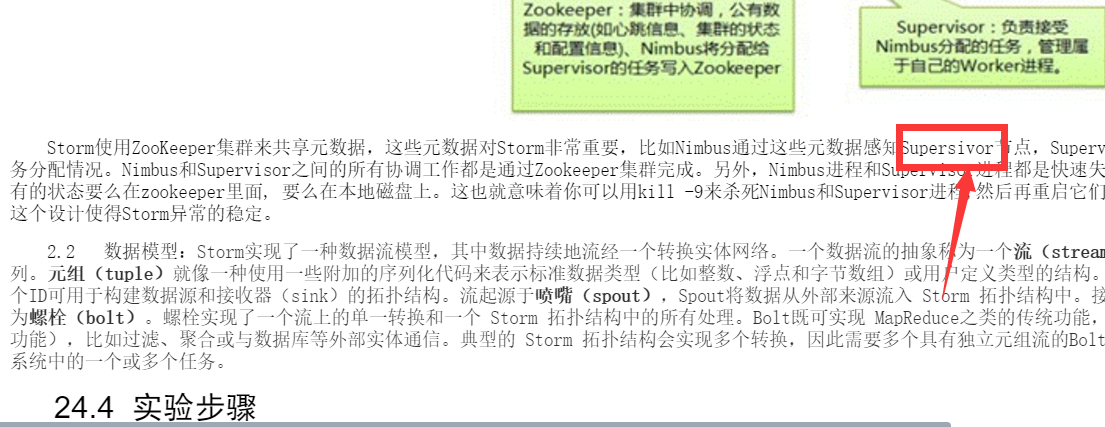
下面这个地方似乎应该是“输出”



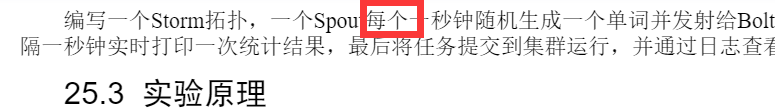
# 实验二十四

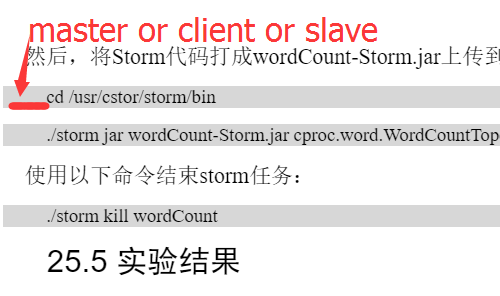
1. 少一个主类代码
2. 代码缺少import部分，建议加上，便于学习者比对。

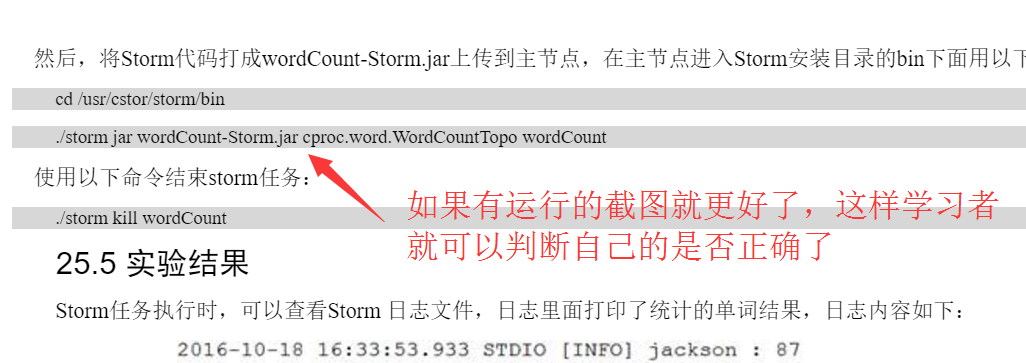


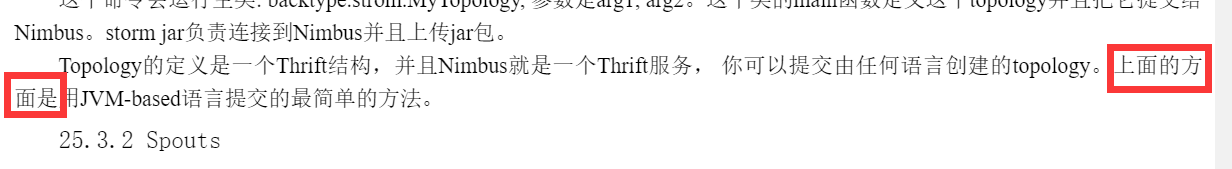


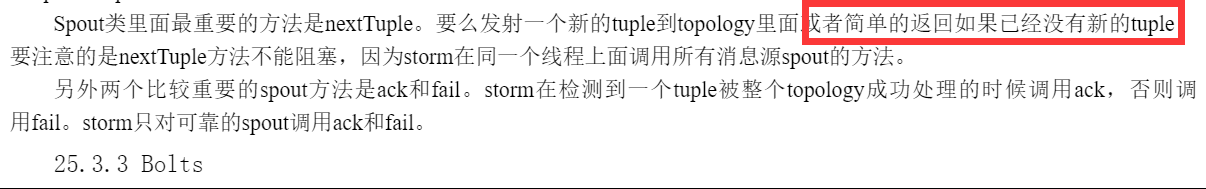
# 实验二十五



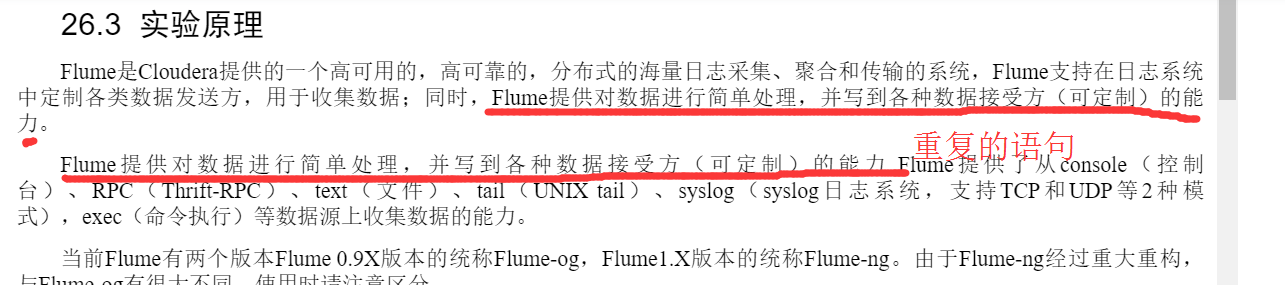


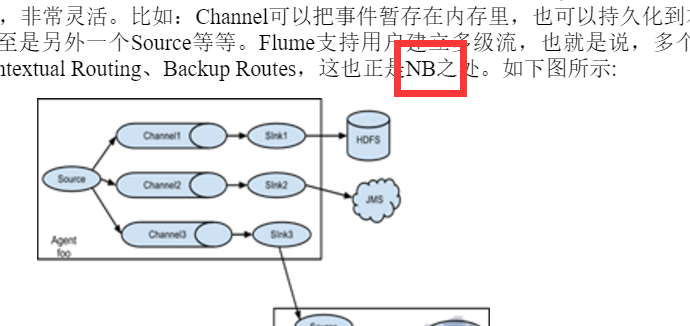




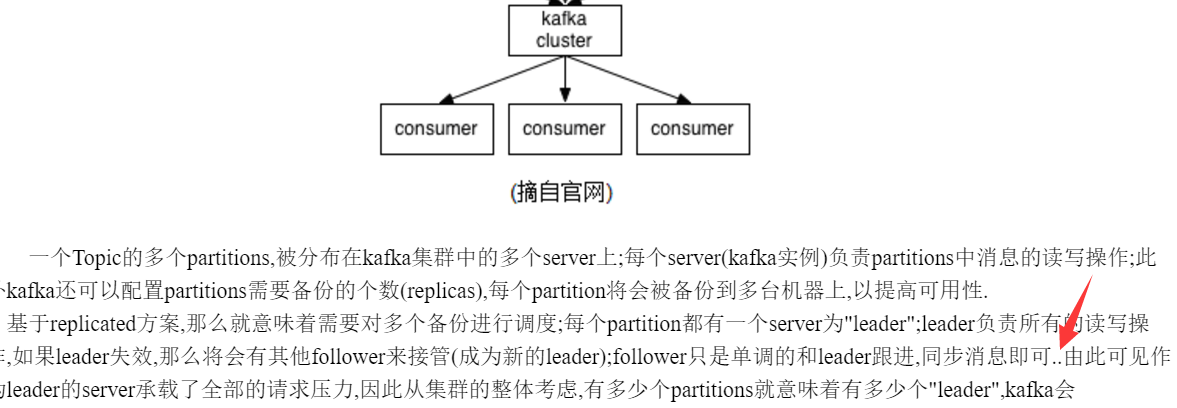


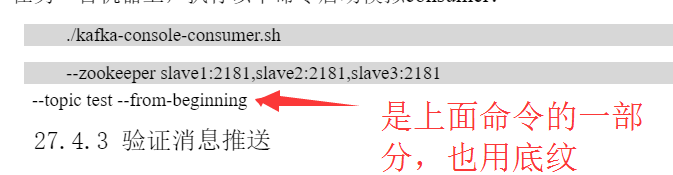
# 实验二十六



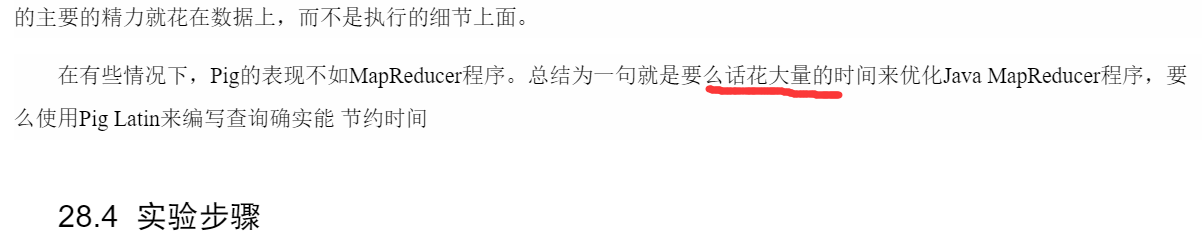


# 实验二十七





# 实验二十八



# 实验二十九



# 实验三十三

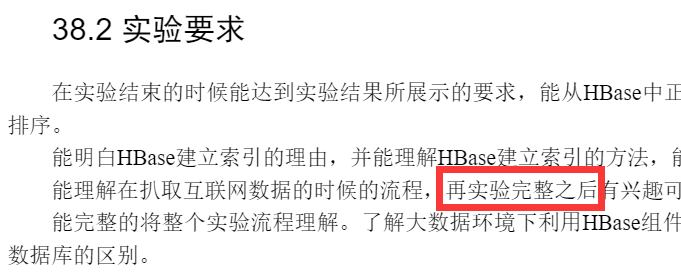




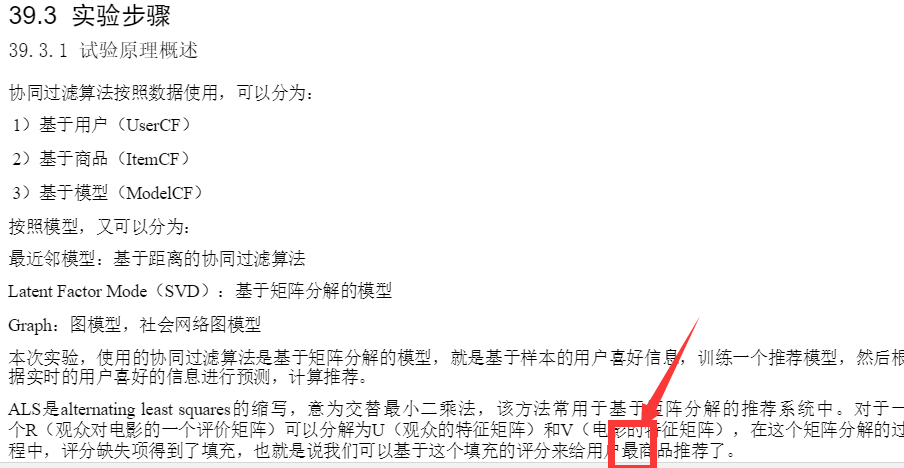
# 实验三十四

对K-Means的介绍与实验三十三有重复

# 实验三十八



# 实验三十九



# 实验二十三