结课报告——数据库部分

组49

# 数据源：

[basketball-references.com](http://www.basketball-references.com)

82games.com

espn.com

# 获取数据方式：

用java实现的爬虫

# 数据内容：

NBA整个历史上的全部球员资料

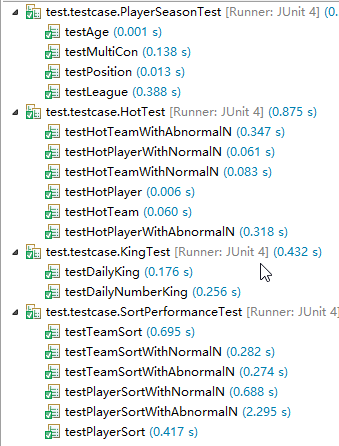
有记载以来的完整的比赛数据，包括比赛文字实录、比分、每一场每一个球员的表现

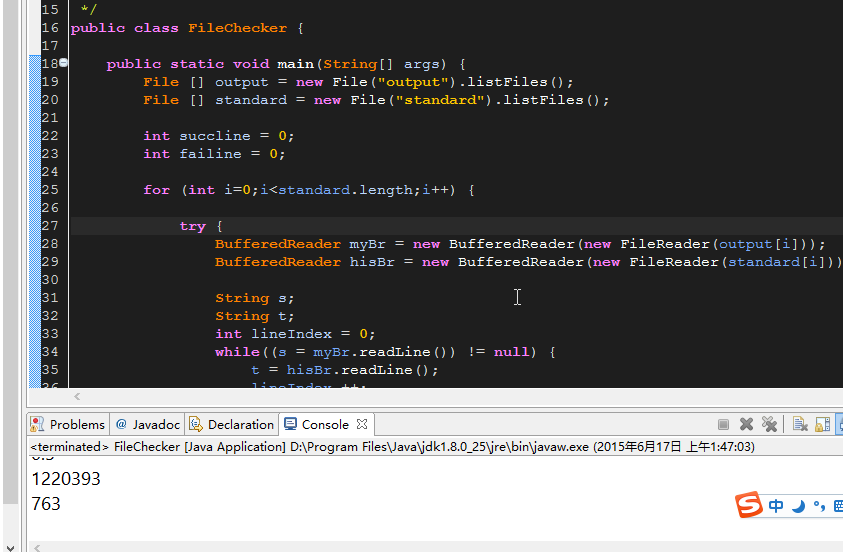
球队资料，包括每个赛季的数据汇总、比赛记录、阵容

赛季数据，包括基本数据（如球员和球队的命中率、得分等）和高端数据（如球员的WAR值、球员在Clutch Time的得分、球队的防守效率等）

# 如何保证数据的正确性：

1. 使用已知的计算公式，自己编写计算逻辑，使用抓取的比赛数据，计算出赛季数据和高端数据，与网站上核对，保证计算逻辑的正确性。使用正确的逻辑，生成可以保证正确性的数据，作为标准。
2. 每次改变逻辑、更换数据源等，均使用之前的测试用例和正确数据进行回归测试，保证新数据的正确性。





如图，在一次回归测试过程中，输出的122万行语句有763行与标准结果不匹配。考虑到浮点误差，这说明数据正确性没有问题。

# 数据库设计：

# 