数据仓库课程练习：

1. 数据来源：

<http://snap.stanford.edu/data/web-Movies.html>

1. 需要存储的信息

电影ID，评论用户ID，评论用户ProfileName，评论用户评价Helpfulness，评论用户Score，评论时间Time，评论结论Summary，评论结论Text，电影演员，电影上映时间，电影风格，电影导演，电影主演，电影演员，电影版本等信息。

1. 环境：

构建关系型数据仓库系统，自行设计数据存储模式

数据建模工具请参考使用ERWin，ETL工具请参考使用Pentaho DI，均使用Community的版本。

1. 常见查询及统计（占总查询数目）=80%）：
   1. 按照时间进行查询及统计（例如XX年有多少电影，XX年XX月有多少电影，XX年XX季度有多少电影，周二新增多少电影等）
   2. 按照电影名称进行查询及统计（例如 XX电影共有多少版本等）
   3. 按照导演进行查询及统计（例如 XX导演共有多少电影等）
   4. 按照演员进行查询及统计（例如 XX演员主演多少电影，XX演员参演多少电影等）
   5. 按照电影类别进行查询及统计（例如 Action电影共有多少，Adventure电影共有多少等）
   6. 按照上述条件的组合查询和统计
2. 提交内容
   1. ETL脚本
   2. E-R图
   3. 物理存储模型（Schema）
   4. 数据仓库导出文件（请按照每张表分别导出csv格式文件并压缩。文件名同表名）
   5. 每组准备一个答辩ppt，说明你的存储方案和思考过程。