1. import\_hive.txt  
   将数据导入hive
2. ProcessData.scala  
   进行数据清洗、特征抽取、one hot编码（27个特征值）
3. TrainModel.scala  
   进行模型训练  
   结果:简单测试acc:0.08
4. ProcessDataPro.scala  
   用于替代ProcessData.scala，根据行为数据抽取更多可用特征（73个特征值）
5. ProcessDataToTest.scala  
   用于生成测试集
6. TrainModelPro.scala  
   用于替换TrainModel.scala  
   结果:acc:0.35
7. CheckData.scala  
   检查清洗后的数据有无异常值
8. AnalyseData.scala  
   进行数据的分析，为数据可视化提供支持

执行顺序：  
1. (1)->2->3  
2. (1)->4->5->6  
PS:文件1无需重复运行

其他文件：  
1. webapps.zip  
用户行为数据可视化（下单、关注、删除等）