

# 我了个去，还能这么测数据？

报表类型数据项目测试分享

演讲人：灵石

2012年11月24日

@一淘测试 @淘了啊

- 报表类型数据的定义和测试挑战
- 开始测试之封装
- 开始测试之数据来源
- 报表类型数据测试的分层

# 什么是报表类型的数据计算

- **上游：各个环节收集的日志和数据**
- **产出：各类报表**
- **用户：淘宝卖家，内部小二**
- **被测试对象使用：**
  - **Hive**
  - **Hadoop**
  - **Storm**
- **过程和流程简单，不涉及UI**

- 计算量庞大，需要分布式支持
- 开发节奏快，数据挖掘类型
- 数据上游不掌握在我们手中
- 输入和产出都是各种格式的数据
- 数据正确是头等大事

# 如何测试？

- **第一步：封装所有无关操作**
- **第二步：测试数据准备**

- **让测试人员关注真正应该关注的事！**
- **Don't Repeat Yourself**
- **如何做？**
  - **面向数据的测试框架SmartTest**
  - **测试用例：input => output**



# 伪造数据VS真实数据

问题	伪造数据	真实数据
是否能覆盖到输入数据的所有情况?	不能	几乎可以
数据表之间的关联信息是否可以模拟?	不容易	轻松
发现问题便于调试?	容易调试	不易调试
期望结果如何获取?	造数时	进行计算
测试思维和开发思维的关系?	反向	正向
可重复运行?	可以	不容易
对于数据格式变更敏感?	不敏感	敏感

# 伪造数据VS真实数据

- 伪造数据更容易测试和自动化
- 但是无法替代真实数据的真实地位
- 我们两种数据都使用，并且将测试分层



- 格式和数据内容分离
- 从头开始做数据
- 对数据完全控制（不做随机，除非需要）
- 使用默认数据



一致性测试

黑盒测试

单元测试

- 使用自己制造的数据
- 开发自测
- 测试提供工具和平台支持

# 如何对HIVE进行单元测试

- HIVE目前没有很好的单元测试支持
- HIVE的测试其实是数据的测试
- 测试提供平台降低单元测试门槛
  - 管理所有的HIVE表信息
  - 在平台界面上造数
  - 平台从机器池中取机器运行、收集结果
  - 生成测试用例
  - 持续集成

- 使用自己的数据
- 将整个项目作为一个黑盒
  - 忽略其中的过程
  - 按照输入输出进行测试
- 使用造数框架对关联性进行重点造数

# 黑盒测试：造数框架实例

访问日志	
用户ID	URL

成交日志	
用户ID	宝贝ID

店铺信息	
PK	<u>店铺ID</u>
	所属类目

宝贝信息	
PK	<u>宝贝ID</u>
	店铺ID

被测对象

店铺维度汇总	
PK	<u>店铺ID</u>
	成交 PV .....

宝贝维度汇总	
PK	<u>宝贝ID</u>
	成交 PV .....



# 黑盒测试：造数框架详解

用户1 SV: 用户ID

首页

宝贝页面

购买

用户2 SV: 用户ID

搜索

宝贝页面

另一个宝贝

# 黑盒测试：造数框架详解

报表

店铺 ScopeVar: 店铺ID

宝贝 ScopeVar: 宝贝ID

用户1 SV: 用户ID

首页



宝贝页面



购买

用户2 SV: 用户ID

搜索



宝贝页面



另一个宝贝

# 黑盒测试：造数框架详解

报表

店铺 ScopeVar: 店铺ID

宝贝1 ScopeVar: 宝贝ID

用户1 SV: 用户ID

用户2 SV: 用户ID

宝贝2 ScopeVar: 宝贝ID

用户1 SV: 用户ID

用户2 SV: 用户ID

- 使用线上数据的结果进行测试
- 建立一个完备的一致性验证库
  - 跳失率大于等于0小于等于1
  - $PV \geq IP$
  - 店内宝贝PV  $\leq$  整店PV
  - 大店的某小时PV 在全天pv的3%到7%之间
  - 和修改之前的经验值做对比
  - 和某店的经验值做对比

- 封装所有测试不需要关心的事情
- 伪造数据VS真实数据
- 数据测试如何划分层次
- HIVE单元测试平台
- 如何伪造大量的关联数据
- 如何使用真实数据进行测试

# 谢谢大家&QA

- **关于一淘测试**
  - 网站：[testing.etao.com](http://testing.etao.com)
  - 微博：[@一淘测试](#)
  - 加入我们：[testing.etao.com/jobs](http://testing.etao.com/jobs)
- **关于我**
  - 花名：灵石
  - 博客：[www.colblog.net](http://www.colblog.net)
  - 微博：[@淘了啊](#)