

友乾营 · 第十期

如何实现 ●  
报表T+0查询?

# I 关于T+0查询

T+0查询是指全量数据实时查询，即历史数据和当期数据混查

1

准实时同步

2

同库

3

应用层编码

实现T+0的手段

# I 为什么总是T+1

1

只能准实时，而且高频对生产库也有影响

2

业务库放全量数据，对DB压力太大

3

应用层报表工具无法实现，硬编码太复杂

# I T+0的实现方向

数据库层面可操作空间有限

还是要在应用层想办法



数据库 ×

应用端 ✓

# I 应用层实现T+0需要具备哪些能力

- **多数据源支持**

- 很方便从数据库里取数

- **强计算能力**

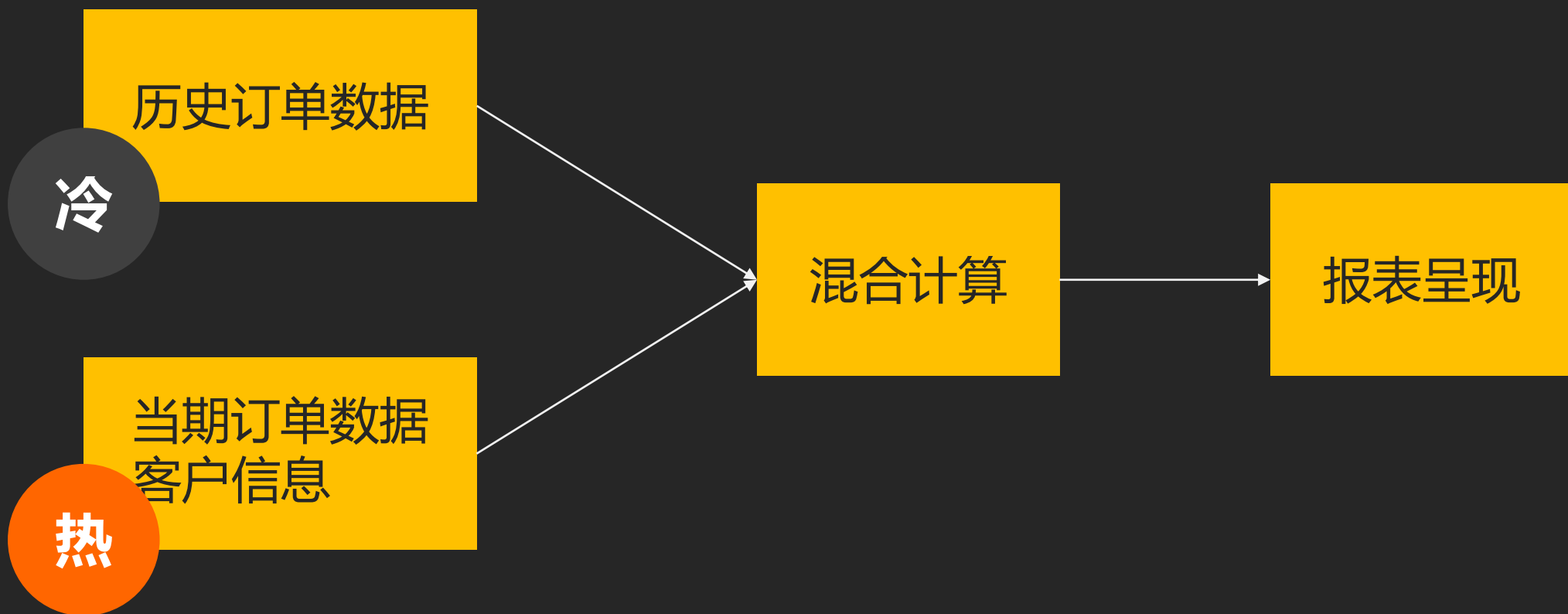
- 可实施复杂计算，计算体系相对完备；实现简单

- **支持多源混合计算**

- 支持异构数据源间的混合计算

# I 举例

近一年内前5大客户的实时交易情况



## I 实现

C:\Users\李松波\Desktop\T+0报表查询.rpx

|   | A              | B               | C          |
|---|----------------|-----------------|------------|
| 1 | 近一年前5大客户实时交易情况 |                 |            |
| 2 | 序号             | 客户              | 订单额        |
| 3 | =&B3           | =ds1.select(客户) | =ds1.金额    |
| 4 |                | 总额:             | =sum(C3{}) |

数据集设置

| 名称  | 类型  | 数据源 |
|-----|-----|-----|
| ds1 | 集算器 |     |

集算器数据集

DFX文件

p1.dfx

缓存文件名对应的变量

数据管理方式

缓存

参数

| 序号 | 参数名   | 参数值表达式 |
|----|-------|--------|
| 1  | ddate | ddate  |

|   | A                                                                     | B    |
|---|-----------------------------------------------------------------------|------|
| 1 | =connect("demo")                                                      |      |
| 2 | =A1.query("select 客户ID,sum(订单金额) from 订单 group by 客户ID")              | 热数据  |
| 3 | =file("D:\orders.btx").cursor@b(客户ID, 订购日期,订单金额).select(订购日期 >=ddate) | 冷数据  |
| 4 | =A3.groups(客户ID;sum(订单金额):订单金额)                                       |      |
| 5 | =(A2 A4).groups(客户ID;sum(订单金额):订单金额).sort(订单金额-1).to(1,5)             | 冷热归并 |
| 6 | =A1.query@x("select 客户ID,公司名称 from 客户")                               |      |
| 7 | =A5.switch(客户ID,A6.客户ID)                                              | 关联维表 |
| 8 | =A7.new(客户ID.公司名称:客户,订单金额:金额)                                         | 结果集  |

## 报表展现

近一年前5大客户实时交易情况

| 序号  | 客户    | 订单额         |
|-----|-------|-------------|
| 1   | 高上补习班 | ¥ 226340.03 |
| 2   | 大钰贸易  | ¥ 215188.79 |
| 3   | 正人资源  | ¥ 211231.60 |
| 4   | 学仁贸易  | ¥ 100269.36 |
| 5   | 实翼    | ¥ 68482.94  |
| 总额: |       | ¥ 821512.72 |

# I 脚本

|   | A                                                                    | B     |
|---|----------------------------------------------------------------------|-------|
| 1 | =connect("demo")                                                     |       |
| 2 | =A1.query("select 客户ID,sum(订单金额) 订单金额 from 订单 group by 客户ID")        | /热数据  |
| 3 | =file("D:\\orders.btx").cursor@b(客户ID,订购日期,订单金额).select(订购日期>=ddate) | /冷数据  |
| 4 | =A3.groups(客户ID;sum(订单金额):订单金额)                                      |       |
| 5 | =(A2 A4).groups(客户ID;sum(订单金额):订单金额).sort(订单金额:-1).to(1,5)           | /冷热归并 |
| 6 | =A1.query@x("select 客户ID,公司名称 from 客户")                              |       |
| 7 | =A5.switch(客户ID,A6:客户ID)                                             | /关联维表 |
| 8 | =A7.new(客户ID.公司名称:客户,订单金额:金额)                                        | /结果集  |



# I 实现关键

## 计算引擎（集算器）

1. 多数据源支持（可以读文件）
2. 强计算能力（具备完备的计算能力）
3. 提供私有高效数据存储格式
4. 支持并行
5. 支持分布式环境
6. ...

# I 详细的例子

友乾营

乾学院

## 《实时报表 T+0 的实现方案》

<http://c.raqsoft.com.cn/article/1541494770016>

# I 数据预处理的方式

文件系统具备更高的I/O性能、低廉的成本，常用于存储历史数据



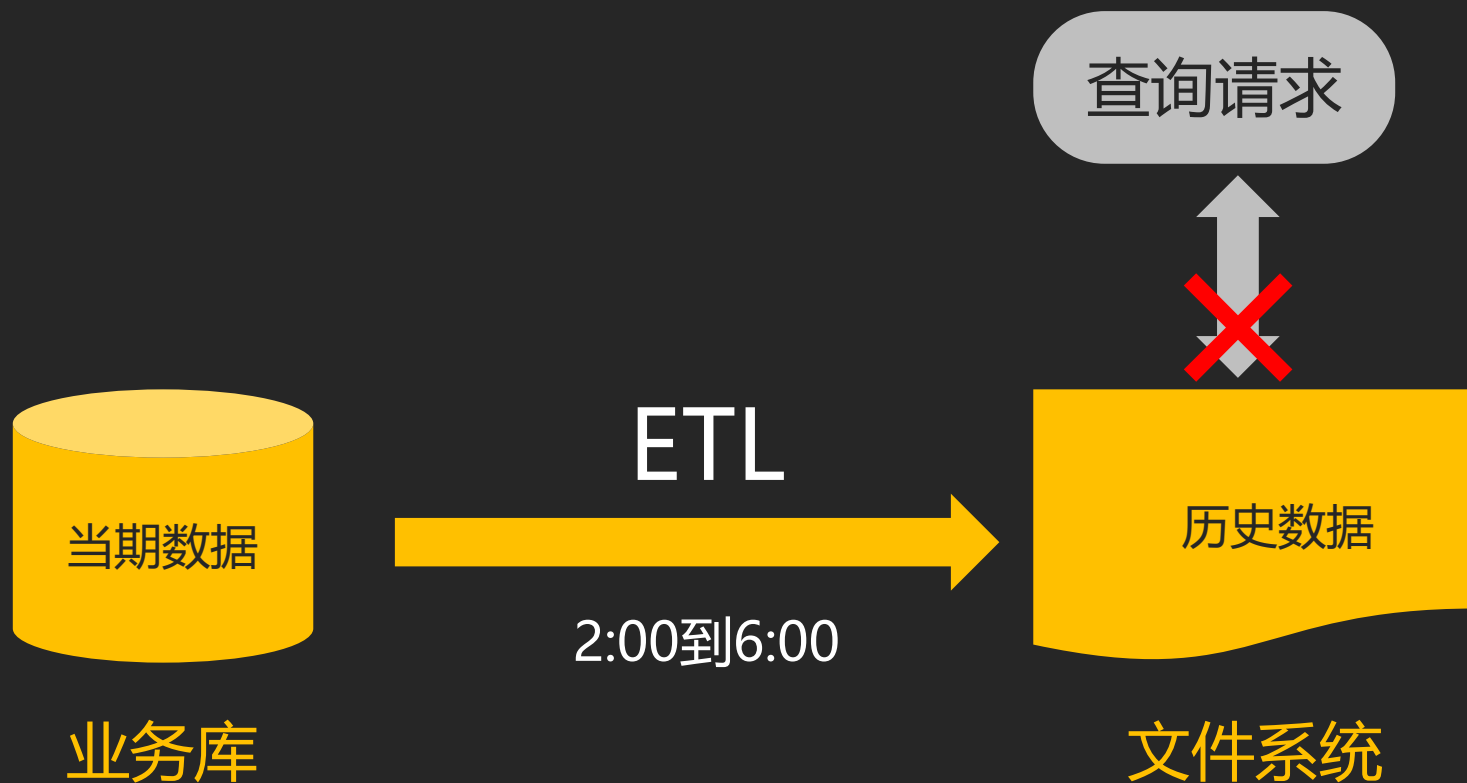
冷导出

温导出

热导出

数据导出的三种方式

# I 冷导出

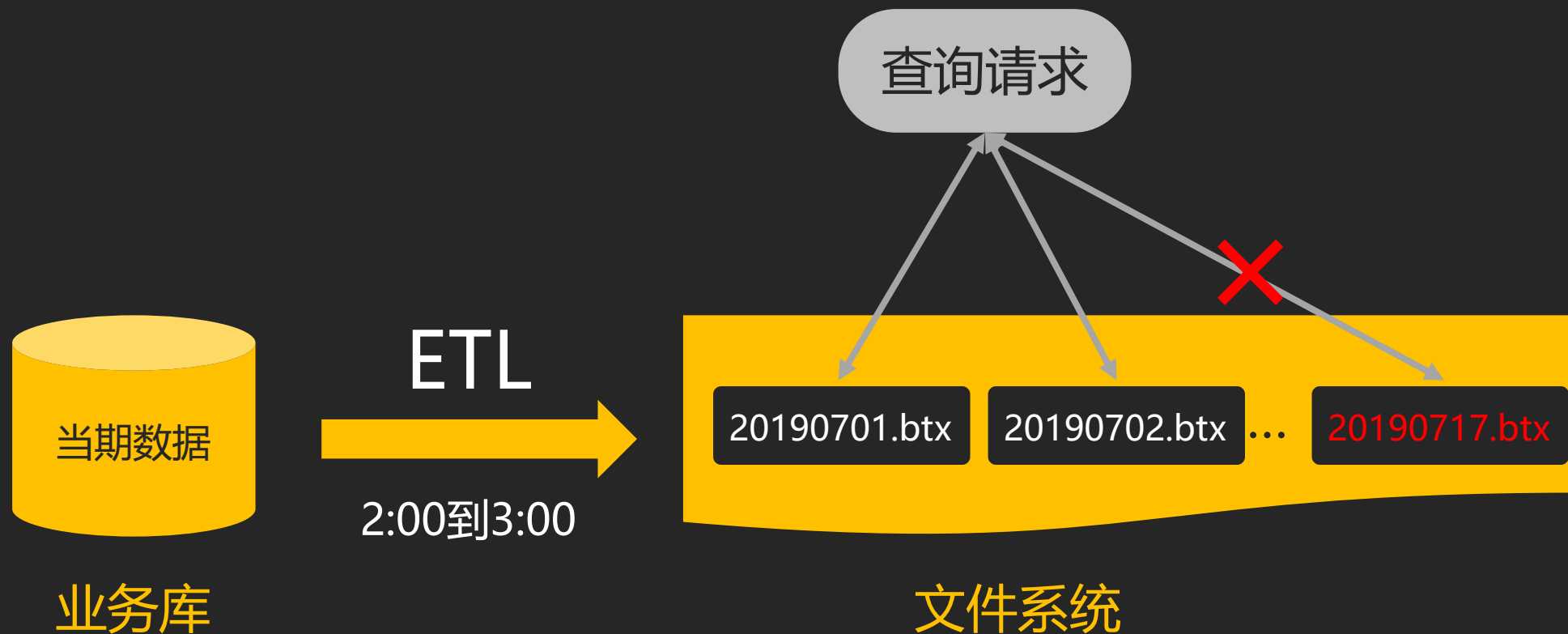


**冷导出：**固定时间从生产库导出数据到文件系统中，期间分析系统暂停使用，导出完成后恢复

**缺点：**分析系统需要停机

# I 温导出

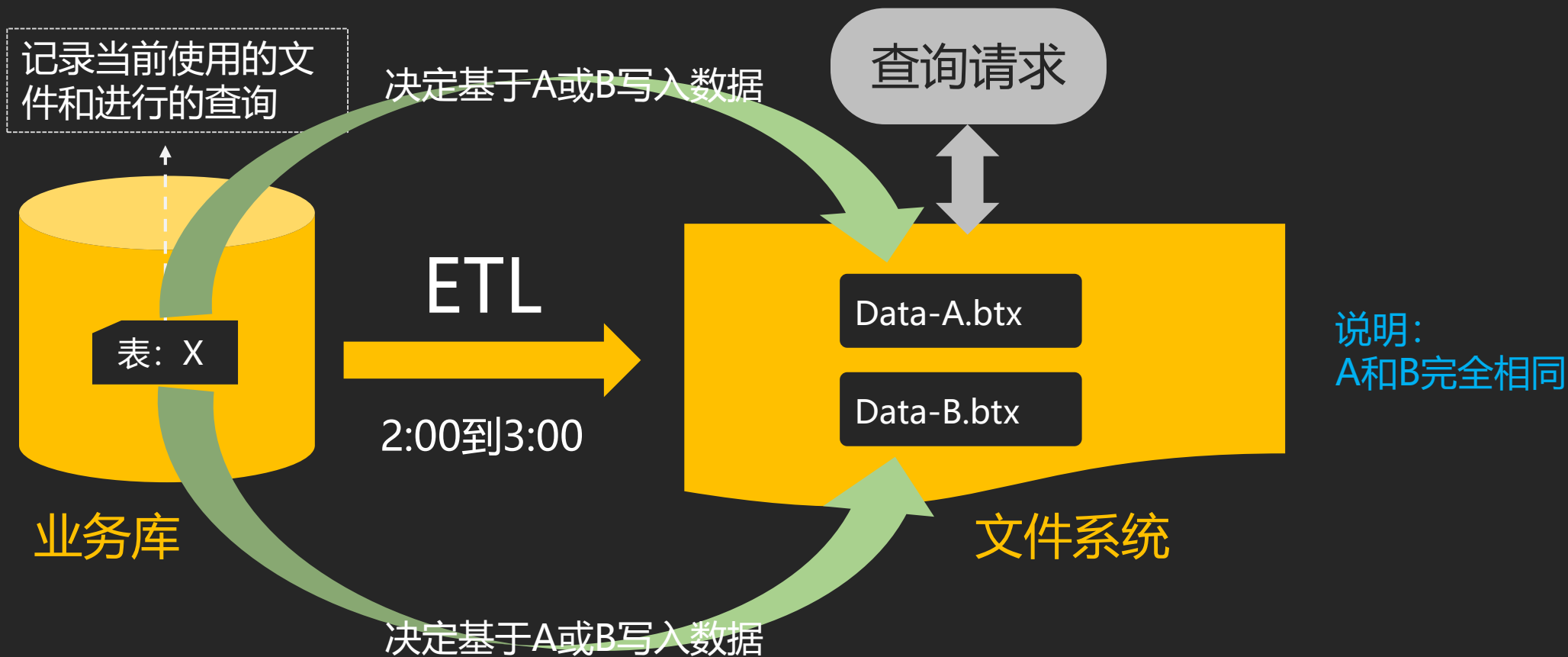
友乾营



**温导出:** 折中办法，把数据拆细，如按日存储，当天数据导出时不允许查询，其他日数据可以正常使用

**缺点:** 文件拆分过多，管理有难度；基于多个文件的计算代码实现复杂，也可能引起性能问题

# I 热导出



**热导出:** 借助关系数据库的一致性, 实现热导出, 分析系统永不停机, 24小时在线

**缺点:** 实现相对复杂

# I 热导出注意事项

热导出基本原理是在导出数据过程中使用另一个文件，完成导出后再换回来去维护备份文件。期间要考虑到查询的并发性，借助数据库的一致性确保不会发生写入和查询在同一个时刻针对同一文件进行。并发读写导致的错误，不是刻意或大规模使用时很难测试出来，在设计时要特别小心。

要使用具备一致性能力的数据库（X表）确保多并发下不会出现读写错误

乾学院

## 《实时报表 T+0 的实现方案》

<http://c.raqsoft.com.cn/article/1541494770016>

### 【产品支持】

集算器仓库版已经支持这一机制，直接向组表追加数据就可以了，集算器会自动处理热导出中的问题。集算器不依赖数据库，可以延续文件系统的好处



# 好多乾

## 注册润乾线上直销系统



玩转好多乾

<http://www.raqsoft.com.cn/wx/hdq-strategy.html>

