

在报表中直接使用多样性数据源



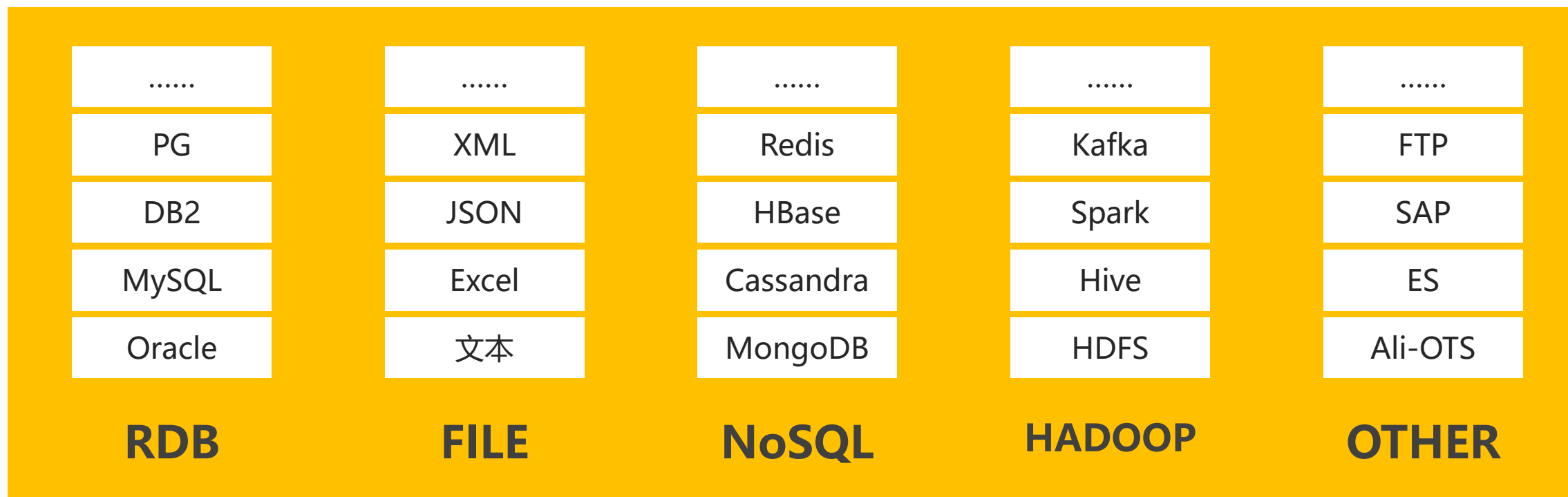
22 期

www.raqsoft.com.cn/yqy

多样性数据源给报表开发带来的新挑战

数据源种类不断增多，给报表开发带来了极大的困难

报表应用



借助RDB中转

将多样性数据源数据统一整理加载到关系库中，再基于关系库出报表
避免报表直接面向多样性数据源

报表应用

RDB



RDB



NoSQL



HDFS



LocalFile



HTTP



rest微服务



IOT数据

应用端硬编码

在应用端通过硬编码处理多样性数据源数据后，交给报表呈现
数据处理环节也可以建设成独立的数据处理层

报表应用

硬编码



RDB



NoSQL



HDFS



LocalFile



HTTP



rest微服务



IOT数据

借助RDB中转

优点

- 单库查询，避免多源问题
- 使用SQL查询相对方便

缺点

- 应用结构复杂，运维成本高
- 数据实时性差
- 与关系库的耦合度过高

应用端硬编码

优点

- 应用结构清晰
- 数据实时性强

缺点

- 硬编码太难，没有SQL简单
- 编译型算法无法热切换
- 应用的耦合度过高

当开发报表时，我们希望：

报表工具具备 直接使用多样性数据源的能力

目标：

RDB+硬编码优点

实现简单，难度与使用SQL相当

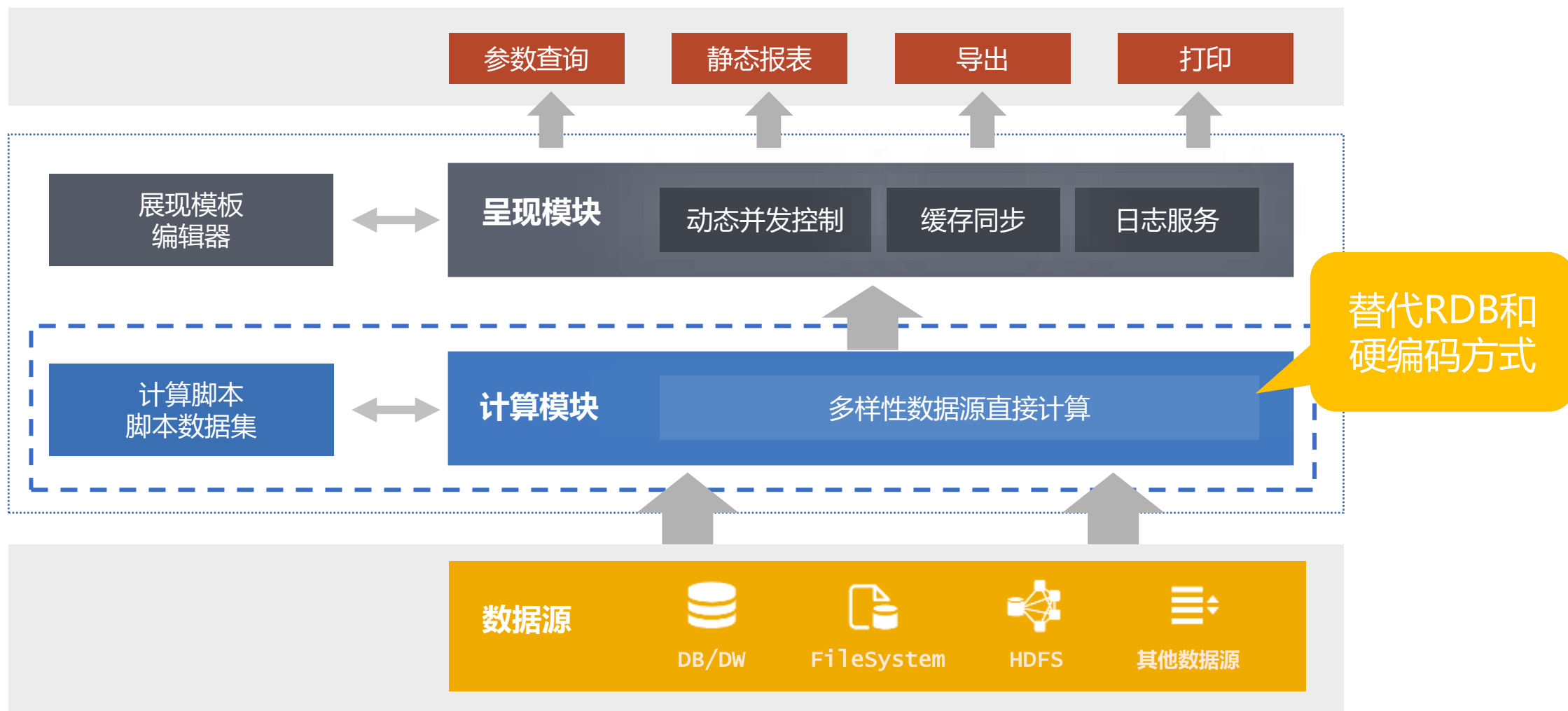
应用结构简单，易于维护

增强数据实时性

有效降低应用耦合性

➤ 润乾报表的方式

润乾报表通过计算模块完成多样性数据源的直接使用，在报表模板的脚本数据集中完成



举例

EXAMPLE



友乾营

专注数据技术的社群

1

文本/Excel数据源

2

JSON数据源

3

WEB HTTP数据源

4

MongoDB数据源

5

混合数据源



1-文本/Excel数据源

历史订单数据归档到文本中，现需要查询历史订单信息

【报表需求】基于历史的文本数据，查询2012年各地区、各销售人员销售情况

1	订单ID,客户ID,雇员ID,订购日期,到货日期,发
2	货日期,运货商,运货费,货主名称,订单金额
10248,VINET,5,2012-07-04,2012-08-01,2012-0	
7-16,3,32.38,余小姐,428.0	
10249,TOMSP,6,2012-07-05,2012-08-16,2012-0	
7-10,1,11.61,谢小姐,1842.0	
10250,HANAR,4,2012-07-08,2012-08-05,2012-0	
7-12,2,65.83,谢小姐,1523.4999898076057	
10251,VICTE,3,2012-07-08,2012-08-05,2012-0	
7-15,1,41.34,陈先生,624.9499997608364	
10252,SUPRD,4,2012-07-09,2012-08-06,2012-0	
7-11,2,51.3000000000000004,刘先生,3559.4999	
980553985	
10253,HANAR,3,2012-07-10,2012-07-24,2012-0	
7-16,2,58.17,谢小姐,1428.0	
10254,CHOPS,5,2012-07-11,2012-08-08,2012-0	

订单.txt

1	客户ID,公司名称,联系人姓名,联系人职务,地址
2	城市,地区,邮政编码,国家,电话,传真
ALFKI,三川实业有限公司,刘小姐,销售代表,大	
崇明路 50 号,天津,华北,343567,中国,(030)	
30074321,(030) 30765452	
ANATR,东南实业,王先生,物主,承德西路 80	
号,天津,华北,234575,中国,(030)	
35554729,(030) 35553744	
AROUT,国顶有限公司,方先生,销售代表,天府东	
街 30 号,深圳,华南,890879,中国,(0571)	
45557788,(0571) 45556750	
BERGS,通恒机械,黄小姐,采购员,东园西甲	
30 号,南京,华东,798089,中国,(0921)	
9123465,(0921) 55123467	
BLAUS,森通,王先生,销售代表,常保阁东 80	
号,天津,华北,787045,中国,(030)	

客户.txt

1	雇员ID,姓名,职务,尊称,出生日期,雇用日期,地
2	址,城市,地区,邮政编码,国家,家庭电话,分机,
3	备注,上级,ISMODITY
1,张颖静,销售代表,女士,1968-12-08	
00:00:00,1992-05-01 00:00:00,复兴门 245	
号,北京,华北,100098,中国,(010)	
65559857,5467,获北京大学心理学学士学位。她	
同时完成了“冷食的艺术”。张颖是国际美食协会	
的会员。2,N	
2,王伟,副总裁(销售),博士,1962-02-19	
00:00:00,1992-08-14 00:00:00,罗马花园	
890 号,北京,华北,109801,中国,(010)	
65559482,3457,王伟获南京大学商业学士学位,	
获该校国际营销博士学位。他能说流利的法语和	
意大利语并能阅读德语。他加入公司时是销售代	
表,被提拔为销售经理并升任销售副总裁。王伟	

员工.txt

➤ 润乾报表实现

新建报表，使用【脚本数据集】准备数据，实现文本数据查询



【其中】y_date为报表参数，查询年份

	A	B
1	=file("订单.txt"). cursor@t (雇员ID,客户ID,订购日期,订单金额;","")	/由于历史订单数据量较大，这里使用文件游标流式读取，还可以增加@m选项进行并行取数
2	=A1.select(year(订购日期)==y_date)	/根据报表参数y_date（年份）过滤数据
3	=file("客户.txt"). import@t (客户ID,地区;","").keys(客户ID)	/读取客户数据
4	=file("员工.txt"). import@t (雇员ID,姓名;","").keys(雇员ID)	/读取员工数据
5	=A2.join(客户ID,A3,地区;雇员ID,A4,姓名)	/关联运算，join三个结果集
6	=A5.groups(地区,姓名;sum(订单金额):订单金额)	/按照地区和人员分组汇总订单金额，统计销售情况
7	return A6	/为报表返回计算后的结果集

报表模板设计

根据已准备的数据集设计报表模板，直接进行
分组呈现

	A	B	C
1	=@y_date+"年销售情况"		
2	地区	人员	订单金额（万元）
3	=ds1.group(地区;地区:1)	=ds1.select(姓名)	=ds1.订单金额
4	总计		=sum(C3{})

报表模板设计

发布报表，报表呈现效果如下：

2012年销售情况		
地区	人员	订单金额（万元）
东北	孙林	2625.00
	张颖静	841.00
	李芳	2124.88
	王伟	4524.00
	赵军	1380.00
	郑建杰	7490.20
华北	刘英玫	10167.17
	孙林	5148.66
	张颖静	14074.08
	李芳	8222.75
	王伟	8512.88
	郑建杰	16189.30
		16861.64

报表展现效果

还能怎么做？

还能用SQL查文本！！

方式二

润乾报表提供了基于SQL的文本查询方式，前面的脚本数据集可以改成：

	A	B
1	=connect()	/连接文件系统，接下来就可以进行文件的SQL查询了
2	\$(A1) select t2.地区 地区,t3.姓名 姓名,sum(t1.订单金额) as 订单金额 from 订单2.txt t1 join 客户2.txt t2 on t1.客户ID=t2.客户ID join 员工2.txt t3 on t1.雇员ID=t3.雇员ID where year(t1.订购日期)=2012 group by t2.地区,t3.姓名	/基于A1的连接，使用标准SQL查询文本，关联三个文本内容并进行分组汇总 SQL是大家都熟悉的方式，没有学习成本，可以直接上手
3	return A2	

还能怎么做?

还可以用SQL数据集!!!

方式三

通过配置专用数据源后，可以直接使用SQL数据集，编写SQL查询文件数据

1. 配置esproc数据源

数据源

常规属性 扩展属性

数据源名称 数据库类型

esproc UNKNOWN

驱动程序

com.esproc.jdbc.InternalDriver

数据源URL: 注意替换括号中的内容

jdbc:esproc:local://

用户 口令

☐ 使用带模式的表名称 ☐ 使用带引号的SQL ☒ 自动提交

确定(O) 取消(C)

➤ 使用SQL数据集

2. 新建SQL数据集，编写SQL直接查询文本，实现关联、过滤、分组汇总等运算



这种方式更适合熟悉SQL技术人员的习惯，可以直接使用

润乾报表（计算模块）让文本拥有了**计算能力**



写脚本、写SQL都可以查文件

报表可以直接查询文本，无需入库



更进一步

某些情况可以使用文本存储数据，无需建设RDW

➤ 2-JSON数据源

基于JSON格式数据进行查询统计

【报表需求】基于JSON文件数据统计，查询每种服装的数量

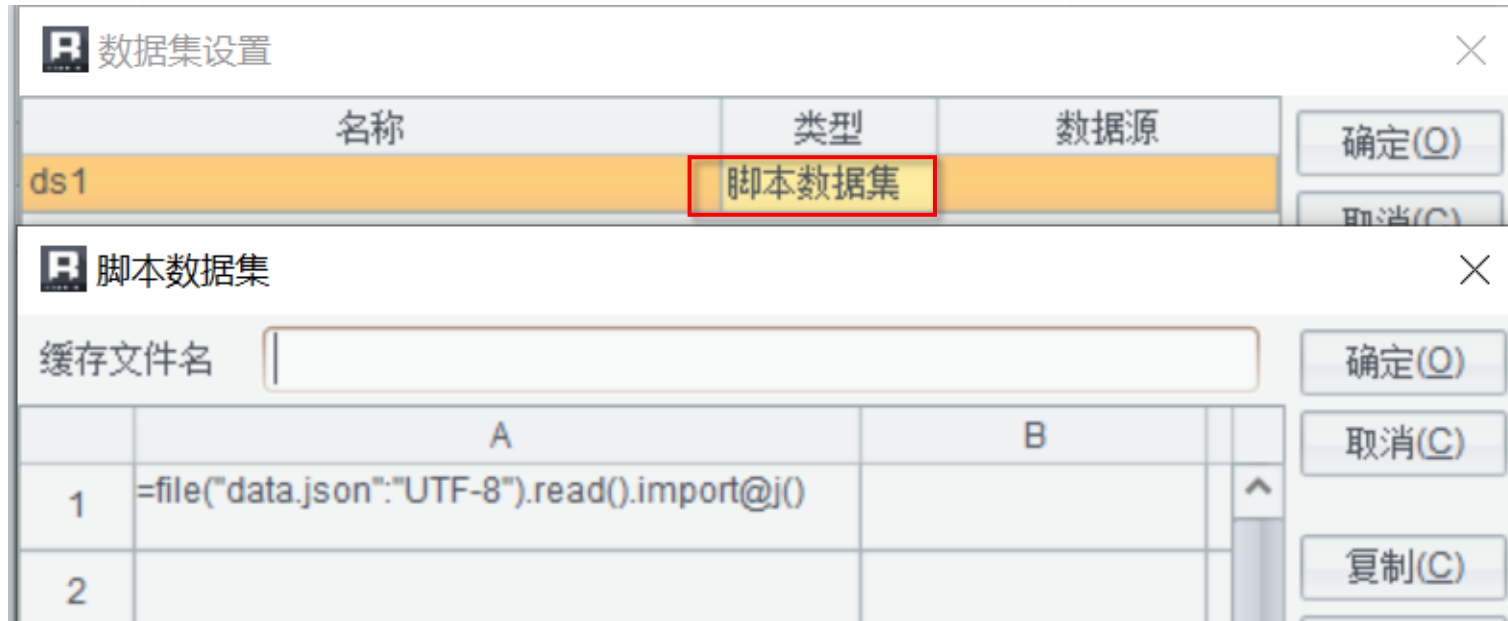


```
1  [
2  {
3    "id": 1,
4    "name": "男士",
5    "productTypeDtos": [
6      {
7        "id": 23,
8        "name": "上装",
9        "productTypeDtos": [
10         {
11           "id": 33,
12           "name": "毛衣"
13         },
14         {
15           "id": 34,
16           "name": "T恤"
17         },
18         {
19           "id": 35,
20           "name": "衬衫"
21         },
22         {
23           "id": 36,
24           "name": "卫衣"
25         },
26         {
27           "id": 37,
28           "name": "背心"
29         },
30         {
31           "id": 38,
32           "name": "针织衫"
33         }
34       ]
35     }
36   ]
37 }
```

 data.json

➤ 润乾报表实现

新建报表，使用【脚本数据集】准备数据，实现json数据查询



A1: =file("data.json":"UTF-8").read().import@j()

仅通过一行脚本即可读取JSON数据，此时为报表返回的是带层次的结果集

➤ 润乾报表层次数据集

脚本数据集中返回的结果集：

序号	id	name	productTypeDtos
1	1	男士	[[23,上装],[33,毛衣], ...]
2	2	女士	[[26,上装],[56,T恤], ...]
3	3	童装	[[31,男童],[85,T恤], ...]

序号	id	name	productTypeDtos
1	23	上装	[[33,毛衣],[34,T恤], ...]
2	24	裤装	[[40,背带裤],[41,休...]
3	25	外套	[[49,皮衣],[50,风衣], ...]

序号	id	name
1	33	毛衣
2	34	T恤
3	35	衬衫
4	36	卫衣
5	37	背心
6	38	针织衫
7	39	POLO衫

报表数据集呈现的结构：



报表模板设计

使用层次数据集设计报表模板，进行分组汇总

	A	B	C
1	服装种类统计		
2	着装群体	服装类型	种类数量
3	=ds1.select(name)	=ds1.productTypeDtos.select	=ds1.productTypeDtos.pro

A3: =ds1.select(name)
B3: =ds1.productTypeDtos.select(name)
C3:
=ds1.productTypeDtos.productTypeDtos.c
ount()

报表模板设计

发布报表，报表呈现效果如下：

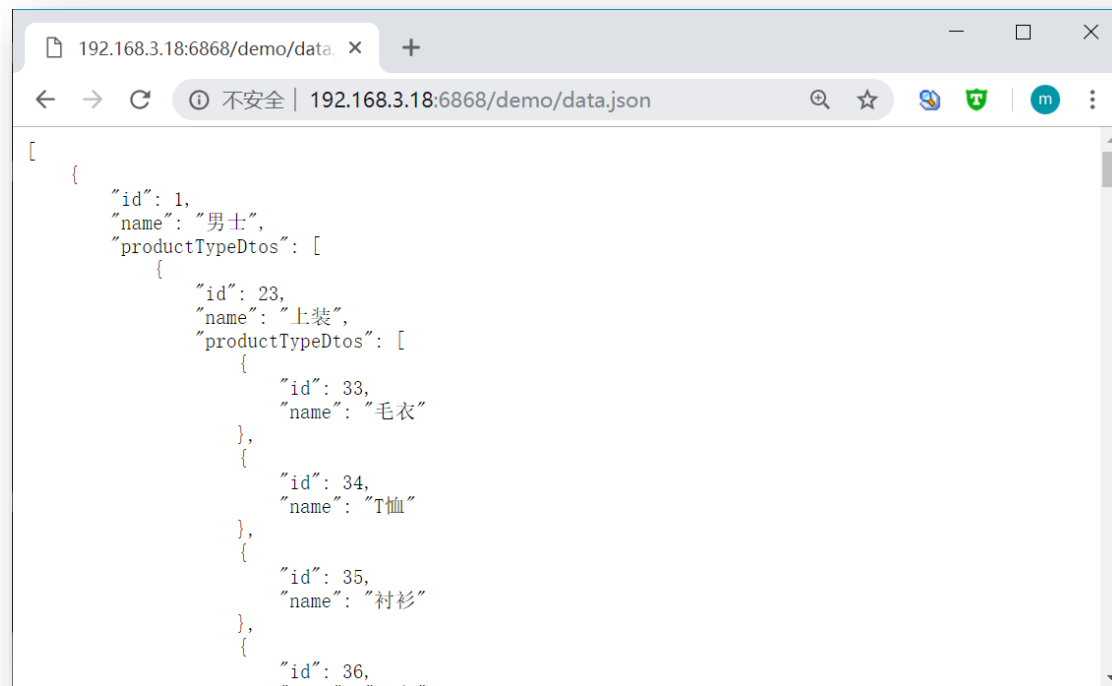
服装种类统计		
着装群体	服装类型	种类数量
男士	上装	7
	裤装	9
	外套	7
女士	上装	7
	裤装	8
	裙装	5
	外套	8
	内衣	1
童装	男童	5
	女童	10

报表展现效果

3-WEB HTTP数据源

有时数据通过WEB方式传递后进行报表呈现，沿用上述例子，json数据通过web传递过来

【报表需求】基于WEB传递的数据，进行报表查询统计



```
[
  {
    "id": 1,
    "name": "男士",
    "productTypeDtos": [
      {
        "id": 23,
        "name": "上装",
        "productTypeDtos": [
          {
            "id": 33,
            "name": "毛衣"
          },
          {
            "id": 34,
            "name": "T恤"
          },
          {
            "id": 35,
            "name": "衬衫"
          },
          {
            "id": 36,
            "name": "卫衣"
          }
        ]
      }
    ]
  }
]
```

➤ 润乾报表实现

新建报表，使用【脚本数据集】准备数据，接收web数据进行呈现



A1: =httpfile("http://192.168.3.18:6868/demo/data.json":"UTF-8").read().import@j()

使用httpfile函数读取远程HTTP数据源，这里也只用一行脚本实现
模板设计与上例相同，此处省略

4-MongoDB数据源

读取MongoDB数据，开发查询统计报表

【报表需求】读取MongoDB数据，完成查询统计报表

```
>
  _id: ObjectId("5dc0d196e793caae78ac41d0")
  id: 3
  name: "童装"
  productTypeDtos: Array
    0: Object
      id: 31
      name: "男童"
      productTypeDtos: Array
        0: Object
          id: 85
          name: "T恤"
        1: Object
        2: Object
        3: Object
        4: Object
    1: Object
```

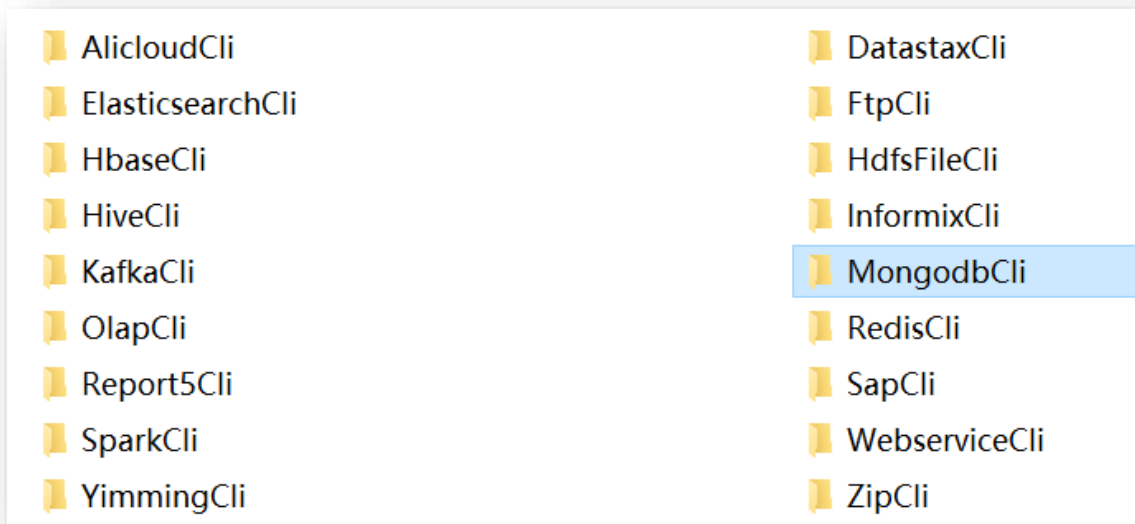
```
_id: ObjectId("5dc0d196e793caae78ac41d1")
id: 2
name: "女士"
productTypeDtos: Array
```

数据内容与JSON例子一致

外部库配置

在润乾报报表中，对于没有JDBC或ODBC的数据源可以通过外部库的方式连接、取数、计算，本例即通过外部库方式连接MongoDB，完成数据查询统计。以下为配置步骤：

1. 到【乾学院】下载外部库所需jar包 (<http://c.raqsoft.com.cn/article/1569805469486>)，其中MongoDB的jar为mongocli.jar



可以看到外部库支持的数据源种类很多，这些数据源都可以在润乾报表中直接使用

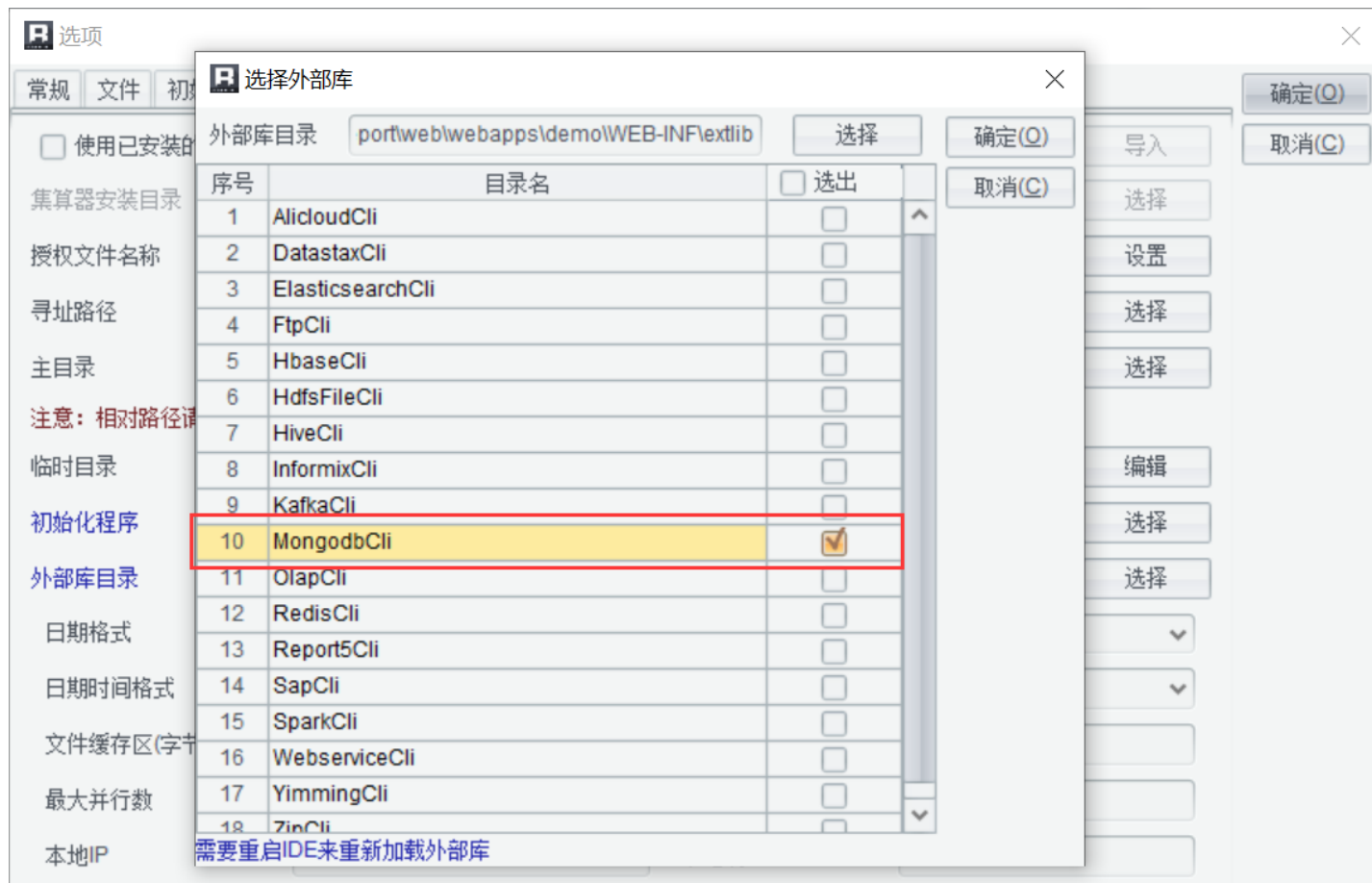
外部库配置

2. 下载MongoDB所需驱动jar，选择相应版本的驱动即可，这里我使用bson-3.11.1.jar、mongodb-driver-3.11.1.jar、mongodb-driver-core-3.11.1.jar
3. 将前两步的jar拷贝到任意目录（如：
D:\software\raqsoft\report\web\webapps\demo\WEB-INF）
4. 在润乾报表IDE中注册外部库，“工具菜单-选项-集算器选项”，选择外部库目录



外部库配置

选择要加载的外部库，这里选择MongoDbCli；注意勾选后要**重启IDE**才能生效



5. 报表在WEB端发布，还需要在应用配置外部库

打开报表应用 WEB-INF 下的 raqsoftConfig.xml, 找到<Esporc>节点，在这个节点内增加：

```
<extLibsPath>D:\software\raqsoft\report\web\webapps\demo\WEB-INF\extlib</extLibsPath>
  <importLibs>
    <lib>MongoDbCli</lib>
  </importLibs>
```

注：extLibsPath 标签设置外部库的根目录

importLibs 设置外部库中对应数据源的 jar 存放目录

➤ 润乾报表实现

新建报表，使用【脚本数据集】准备数据，读取MongoDB数据



A1: =mongo_open("mongodb://localhost:27017/test")

A2: =mongo_shell(A1,"goods.find()").fetch()

通过封装好的函数连接MongoDB，完成数据读取，后面可以继续计算，这里直接返回给报表。报表模板设计与JSON例子一致，不再赘述

➤ 还支持哪些数据源

再罗列一下除了MongoDB，还支持哪些数据源

- Elasticsearch
- HDFS
- Hbase
- Hive
- Spark
- Kafka
- Redis
- 多维数据库
- SAP
- Ali-OTS
- FTP

在润乾报表中
这些数据源都可以直接使用

5-混合数据源

报表经常涉及多源，多种数据源混合的报表该如何开发？这里沿用文本例子的数据，三部分数据分别来源于文本（历史数据）、RDB（业务系统）和Excel（财务系统导出）

【报表需求】查询2012年各地区、各销售人员销售情况

1	订单ID,客户ID,雇员ID,订购日期,到货日期,发
2	货日期,运货商,运货费,货主名称,订单金额
10248,VINET,5,2012-07-04,2012-08-01,2012-0	7-16,3,32.38,余小姐,428.0
10249,TOMSP,6,2012-07-05,2012-08-16,2012-0	7-10,1,11.61,谢小姐,1842.0
10250,HANAR,4,2012-07-08,2012-08-05,2012-0	7-12,2,65.83,谢小姐,1523.4999898076057
10251,VICTE,3,2012-07-08,2012-08-05,2012-0	7-15,1,41.34,陈先生,624.9499997608364
10252,SUPRD,4,2012-07-09,2012-08-06,2012-0	7-11,2,51.300000000000004,刘先生,3559.4999
980553985	
10253,HANAR,3,2012-07-10,2012-07-24,2012-0	7-16,2,58.17,谢小姐,1428.0
10254,CHOPS,5,2012-07-11,2012-08-08,2012-0	

订单.txt

序号	客户ID	公司名称	联系人姓名	联系
1	ALFKI	三川实业有...	刘小姐	销售
2	ANATR	东南实业	王先生	物主
3	AROUT	国顶有限公...	方先生	销售
4	BERGS	通恒机械	黄小姐	采购
5	BLAUS	森通	王先生	销售
6	BLONP	国皓	黄雅玲	市场
7	BOLID	迈多贸易	陈先生	物主
8	BONAP	祥通	刘先生	物主
9	BOTTM	广通	王先生	销售

客户数据库表

雇员ID	姓名	职务	尊称	出生日期	雇用日期	地址	城市
1	张颖静	销售代表	女士	#####	#####	复兴门 24	北京
2	王伟	副总裁(销	博士	#####	#####	罗马花园 8	北京
3	李芳	销售代表	女士	#####	#####	芍药园小区	北京
4	郑建杰	销售代表	先生	#####	#####	前门大街 7	北京
5	赵军	销售经理	先生	#####	#####	学院路 78	北京
6	孙林	销售代表	先生	#####	#####	阜外大街 1	北京
7	金士鹏	销售代表	先生	#####	#####	成府路 11	北京
8	刘英玫	内部销售	女士	#####	#####	建国门 76	北京

员工.xlsx

➤ 润乾报表实现

新建报表，使用【脚本数据集】准备数据，分别读取三个数据源数据，混合计算

数据集设置

名称	类型	数据源
ds1	脚本数据集	

确定(O)
取消(C)

脚本数据集

缓存文件名

	A	B
1	=file("订单.txt").cursor@t(雇员ID,客户ID,订购日期,订单金额;","")	建立文件游标，读取文件数据
2	=A1.select(year(订购日期)==2012)	过滤数据
3	=demo.query("select 客户ID,地区 from 客户").keys(客户ID)	读取数据库数据
4	=file("员工.xlsx").xlsimport@t(雇员ID,姓名).keys(雇员ID)	读取Excel数据
5	=A2.join(客户ID,A3,地区;雇员ID,A4,姓名)	关联计算
6	=A5.groups(地区,姓名,sum(订单金额):订单金额)	分组汇总
7	return A6	

确定(O)
取消(C)
复制(C)
粘贴(P)

报表模板设计有文本数据源例子一致，此处不再赘述

润乾报表处理多样性数据源时的特点：

- 直接使用，无需借助RDB，不需要硬编码，数据实时性强
- 脚本实现简单，甚至比SQL更简单
- 脚本与报表统一管理，应用结构简单清晰，易于维护
- 脚本解释执行，支持热切换
- 降低与数据源的耦合性
- 降低报表与应用的耦合性

好多乾

润乾线上直销系统



玩转好多乾

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