

乾学院 · VIP课堂

商业智能



适合人群

- ✓ 面向有JAVA和SQL基础的程序员
- ✓ 对商业智能感兴趣或业务需要了解这方面的知识

CONTENTS



- 基础报表技术
- 统计图技术
- 呈现与输出技术
- 多维分析
- 复杂报表技术
- DashBoard与大屏
- 应用部署技术

课前准备

环境



Windows 64位 或



Linux64位PC

数据

课堂练习数据

工具

数据库：MySQL5.0以上（需自行准备）

报表工具：润乾报表V2018以上

课前准备（练习数据和工具下载地址）

练习数据

Excel: 销售数据.xlsx, 放到任意目录

MySQL脚本: mybi_v2.sql, 导入数据库

报表工具下载

<https://www.raqsoft.com.cn/download/download-jsbb>

乾学院 · VIP课堂

商业智能 基础报表技术



CONTENTS

本章内容


1. 什么是报表
2. 引入报表工具
3. 三种典型的简单报表
4. 参数
5. 报表的开发技术
6. 练习





什么是报表

报表的概念



报表是具有一定
格式、包含一定
信息的载体！



打印出来的就是
报表！

报表和Excel有什么不同？

- 报表的样式固定但数据并不固定，整个报表会根据数据的变化而变化，行列数都不确定
- Excel则是静态报表，格式和数据都是死的

	A	B	C	D	E	F
1	编号	姓名	生日	性别	学历	奖金
2	1	张三	1974年08月22日	女	中专	14256.00
3	2	李四	1985年10月01日	男	高中	3291.00
4	3	小王	1981年09月29日	女	高中	10851.00
5	4	刘淇	1987年05月03日	男	硕生	8583.00
6	5	小霸	1977年11月17日	女	博士	10936.00
7	6	小究	1979年02月13日	女	博士后	16867.00
8	7	小蒋	1976年06月27日	女	博士	11185.00

计算后

编号	姓名	生日	性别	学历	奖金
1	张三	1974年08月22日	女	中专	14256.00
2	李四	1985年10月01日	男	高中	3291.00
3	小王	1981年09月29日	女	高中	10851.00
4	刘淇	1987年05月03日	男	硕生	8583.00
5	小霸	1977年11月17日	女	博士	10936.00
6	小究	1979年02月13日	女	博士后	16867.00
7	小蒋	1976年06月27日	女	博士	11185.00
8	老贼	1989年04月25日	女	本科	8648.00
9	林郡	1979年11月24日	女	初中	14763.00
10	梁伟	1967年05月02日	女	博士	11643.00
11	赵谦	1974年08月26日	男	本科	13126.00
12	孙翠	1968年05月19日	男	本科	8904.00
13	小周	1988年07月01日	女	中专	1118.00
14	温望	1984年07月02日	男	博士	15158.00
15	孟钗	1988年10月25日	女	小学	3013.00
16	戴玉	1966年06月15日	男	硕生	13461.00
17	小玉	1981年03月05日	女	大专	18265.00
18	喜人	1966年05月18日	男	博士	10830.00

计算前

	A	B	C	D	E	F
1(头)	编号	姓名	生日	性别	学历	奖金
2	=ds1.select(=ds1.EMPN	=ds1.BIRTHDAY	=ds1.GEND	=ds1.DEGR	=ds1.BONUS

报表变化— 举例（列表式情况）

同一个报表模板，根据不同的参数查询出不同数据，展现不同的报表结果

过滤条件为：货主地区 = '华东'

1

报表模板

4(头)	订单ID	客户ID	订购日期	货主地区	运货费
5	=ds1.select(订单ID)	=ds1.客户ID	=ds1.订购日期	=ds1.货主地区	=ds1.运货费

过滤条件为：运货费 >= 80

2

报表结果

订单ID	客户ID	订购日期	货主地区	运货费
10249	东帝望	2012年07月05日	华东	¥11.61
10251	千固	2012年07月08日	华东	¥41.34
10256	凯诚国际顾问公司	2012年07月15日	华东	¥13.97
10257	远东开发	2012年07月16日	华东	¥81.91
10258	正人资源	2012年07月17日	华东	¥140.51
10259	三捷实业	2012年07月18日	华东	¥3.25

报表结果

订单ID	客户ID	订购日期	货主地区	运货费
10255	永大企业	2012年07月12日	华北	¥148.33
10257	远东开发	2012年07月16日	华东	¥81.91
10258	正人资源	2012年07月17日	华东	¥140.51
10263	正人资源	2012年07月23日	华北	¥146.06
10267	友恒信托	2012年07月29日	华东	¥208.58
10270	升格企业	2012年08月01日	华北	¥136.54

报表变化— 举例（交叉表情况）

同一个报表模板，根据不同的参数查询出不同数据，展现不同的报表结果

货主地区包含（东北，华北，西北）
并且
运货商包含（急速快递，统一包裹，联邦货运）

报表模板								
4	<table><tr><th>运货商</th><th>货主地区</th><th>订单数量</th></tr><tr><td colspan="3">=ds1.group(货主地区)</td></tr></table>	运货商	货主地区	订单数量	=ds1.group(货主地区)			
运货商	货主地区	订单数量						
=ds1.group(货主地区)								
5	<table><tr><th>运货商</th><th>货主地区</th><th>订单数量</th></tr><tr><td colspan="2">=ds1.group(运货商;运货商)</td><td>=ds1.sum(订单数量)</td></tr></table>	运货商	货主地区	订单数量	=ds1.group(运货商;运货商)		=ds1.sum(订单数量)	
运货商	货主地区	订单数量						
=ds1.group(运货商;运货商)		=ds1.sum(订单数量)						

1

		报表结果		
运货商	货主地区	东北	华北	西北
	订单数量	20	97	2
	急速快递	25	148	3
	统一包裹	15	108	2

2

		报表结果			
运货商	货主地区	华东	华中	华南	西南
	订单数量	1	0	1	0
	空运	2	0	0	0
	货到付款	0	0	1	0

货主地区包含（华东，华南，西南，华中）
并且
运货商包含（空运，货到付款，普通快递）

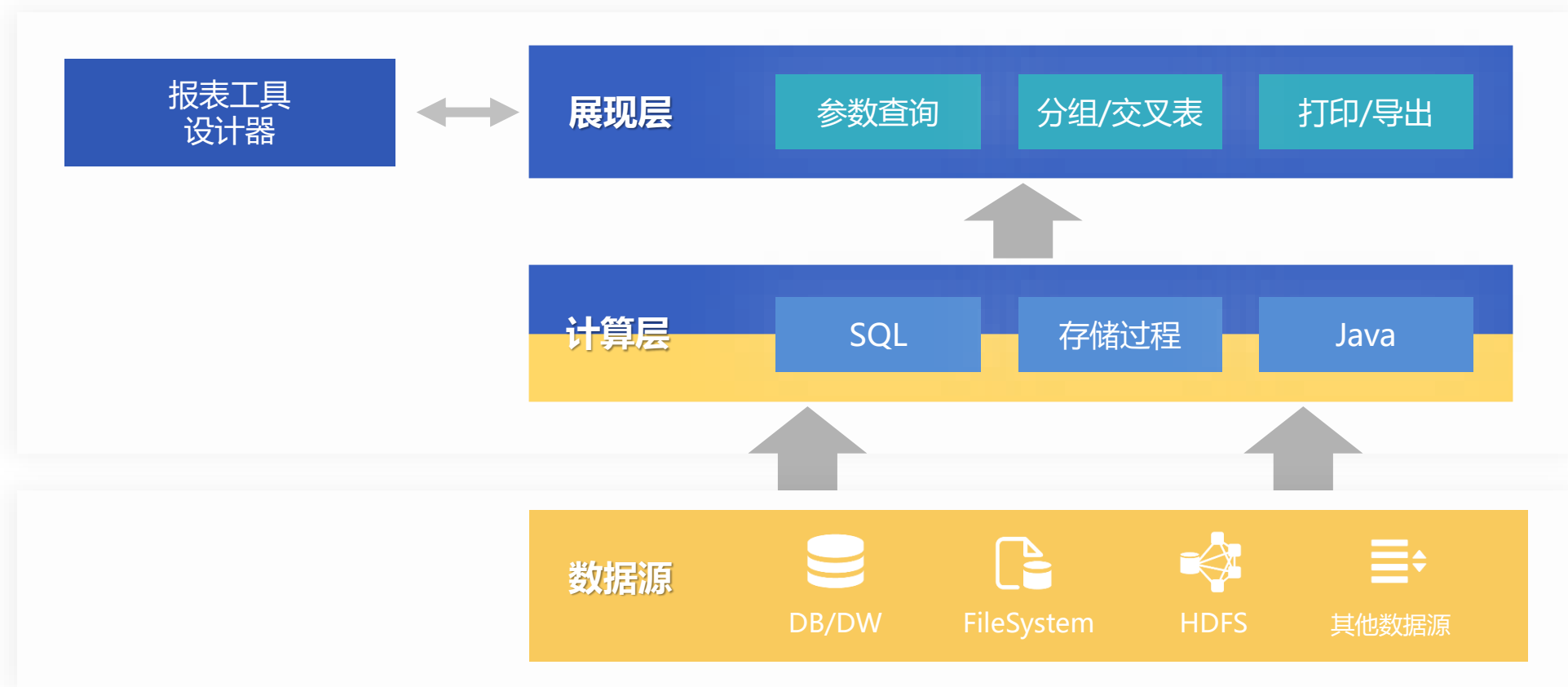


引入报表工具

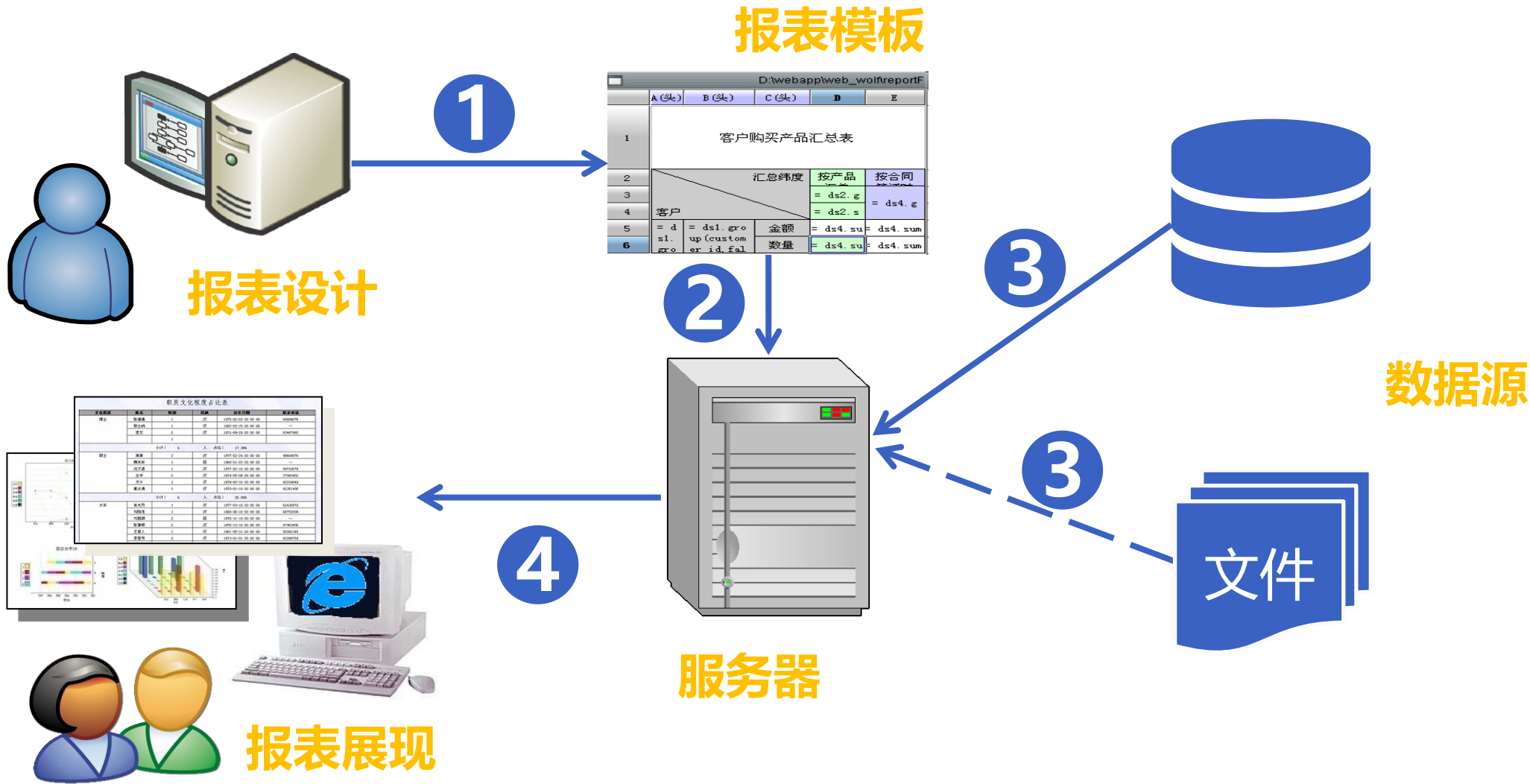
引入报表工具



提高报表制作效率，降低报表开发成本



报表的制作流程



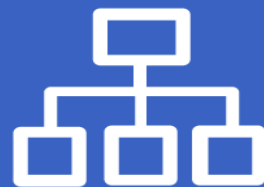


三种典型的简单报表

三种典型的简单报表



列表式报表



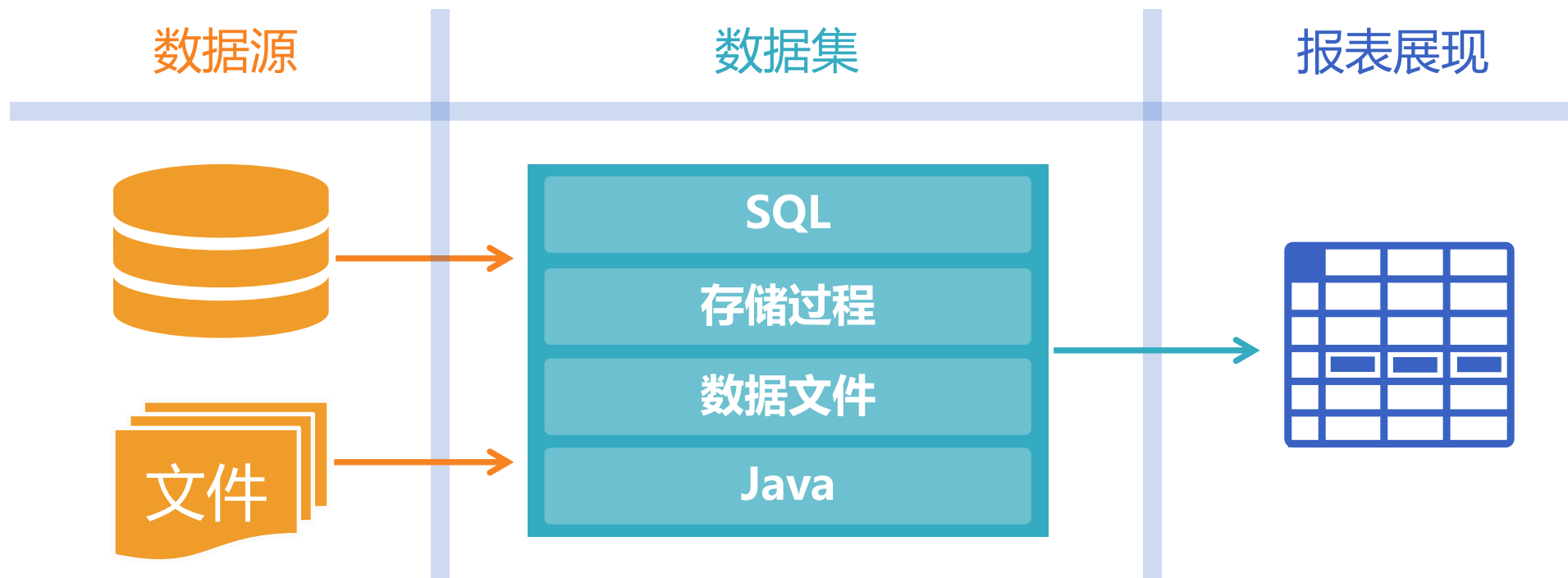
分组式报表



交叉式报表

什么是报表的数据集

- 数据集是报表的数据来源，可以理解为一批结构相同的数据，也就是一句SQL的返回结果
- 数据集可以通过数据库的SQL查询获得，也可以来自其他异构数据源，比如文件，报表不关心数据集的数据来自哪里，只要基于数据集来计算报表



列表式报表——基本特征

列表式报表是将数据以列表的形式展现，可以清楚的看到每条明细数据

订单ID	客户ID	订购日期	货主地区	运货商	运货费
10248	山泰企业	2012年07月04日	华北	3	¥ 32.38
10249	东帝望	2012年07月05日	华东	1	¥ 11.61
10250	实翼	2012年07月08日	华北	2	¥ 65.83
10251	千固	2012年07月08日	华东	1	¥ 41.34
10252	福星制衣厂股份有限公司	2012年07月09日	东北	2	¥ 51.30
10253	实翼	2012年07月10日	华北	2	¥ 58.17
10254	浩天旅行社	2012年07月11日	华中	2	¥ 22.98
10255	永大企业	2012年07月12日	华北	3	¥ 148.33
10256	凯诚国际顾问公司	2012年07月15日	华东	2	¥ 13.97
10257	远东开发	2012年07月16日	华东	3	¥ 81.91
10258	正人资源	2012年07月17日	华东	1	¥ 140.51
10259	三捷实业	2012年07月18日	华东	3	¥ 3.25

列表式报表——数据与报表的关系

数据集（一般指SQL）的行与报表的行是一一对应的关系

```
SELECT EMPID,EMPNAME,BIRTHDAY,GENDER,DEGREE,BONUS FROM 员工
```

SQL结果

EMPID	EMPNAME	BIRTHDAY	GENDER	DEGREE	BONUS
11	赵谦	1974-08-26 ...	1	5	13126.00
14	温望	1984-07-02 ...	1	7	15158.00
16	戴玉	1966-06-15 ...	1	6	13461.00
18	喜人	1966-05-18 ...	1	7	10830.00
19	探春	1991-01-11 ...	1	6	15947.00
4	刘淇	1987-05-03 ...	1	6	8583.00
1	张三	1974-08-22 ...	2	9	14256.00
10	梁炜	1967-05-02 ...	2	7	11643.00
12	孙犁	1968-05-19 ...	1	5	8904.00
2	李四	1985-10-01 ...	1	3	3291.00
25	宝玉	1981-03-05 ...	2	4	18265.00
29	赵究	1979-02-13 ...	2	8	16867.00

报表呈现

编号	姓名	生日	性别	学历	奖金
11	赵谦	1974-08-26	1	5	13126.00
14	温望	1984-07-02	1	7	15158.00
16	戴玉	1966-06-15	1	6	13461.00
18	喜人	1966-05-18	1	7	10830.00
19	探春	1991-01-11	1	6	15947.00
4	刘淇	1987-05-03	1	6	8583.00
1	张三	1974-08-22	2	9	14256.00
10	梁炜	1967-05-02	2	7	11643.00
12	孙犁	1968-05-19	1	5	8904.00
2	李四	1985-10-01	1	3	3291.00
25	宝玉	1981-03-05	2	4	18265.00
29	赵究	1979-02-13	2	8	16867.00

——对应

列表式报表— 取数排序

排序可以在数据集完成，也可以在报表呈现时完成

```
SELECT EMPID,EMPNAME,BIRTHDAY,GENDER,DEGREE,BONUS FROM 员工
ORDER BY EMPID ASC
```

编号	姓名	生日	性别	学历	奖金
11	赵谦	1974-08-26	1	5	13126.00
14	温望	1984-07-02	1	7	15158.00
16	戴玉	1966-06-15	1	6	13461.00
18	喜人	1966-05-18	1	7	10830.00
19	探春	1991-01-11	1	6	15947.00
4	刘淇	1987-05-03	1	6	8583.00
1	张三	1974-08-22	2	9	14256.00
10	梁炜	1967-05-02	2	7	11643.00
12	孙犁	1968-05-19	1	5	8904.00
2	李四	1985-10-01	1	3	3291.00
25	宝玉	1981-03-05	2	4	18265.00
29	赵究	1979-02-13	2	8	16867.00

报表排序前后变化

编号	姓名	生日	性别	学历	奖金
1	张三	1974-08-22	2	9	14256.00
2	李四	1985-10-01	1	3	3291.00
3	小王	1981-09-29	2	3	10851.00
4	刘淇	1987-05-03	1	6	8583.00
5	小霸	1977-11-17	2	7	10936.00
6	小究	1979-02-13	2	8	16867.00
7	小蒋	1976-06-27	2	7	11185.00
8	老贼	1989-04-25	2	5	8648.00
9	林郡	1979-11-24	2	2	14763.00
10	梁炜	1967-05-02	2	7	11643.00
11	赵谦	1974-08-26	1	5	13126.00
12	孙犁	1968-05-19	1	5	8904.00
13	小周	1988-07-01	2	9	1118.00

列表式报表— 显示格式

报表数据可以按照一定格式展现，如金额1234.5678显示为¥ 1,234.56，日期2005/12/25显示为2005-12-25，百分比0.2345显示成23.5%等

编号	姓名	生日	性别	学历	奖金
1	张三	1974-08-22	2	9	14256.00
2	李四	1985-10-01	1	3	3291.00
3	小王	1981-09-29	2	3	10851.00
4	刘淇	1987-05-03	1	6	8583.00
5	小霸	1977-11-17	2	7	10936.00
6	小究	1979-02-13	2	8	16867.00
7	小蒋	1976-06-27	2	7	11185.00
8	老贼	1989-04-25	2	5	8648.00
9	林郡	1979-11-24	2	2	14763.00
10	梁炜	1967-05-02	2	7	11643.00
11	赵谦	1974-08-26	1	5	13126.00
12	孙犁	1968-05-19	1	5	8904.00
13	小周	1988-07-01	2	9	1118.00

显示格式设置前后

编号	姓名	生日	性别	学历	奖金
1	张三	1974年08月22日	女	中专	¥14256.00
2	李四	1985年10月01日	男	高中	¥3291.00
3	小王	1981年09月29日	女	高中	¥10851.00
4	刘淇	1987年05月03日	男	硕士	¥8583.00
5	小霸	1977年11月17日	女	博士	¥10936.00
6	小究	1979年02月13日	女	博士后	¥16867.00
7	小蒋	1976年06月27日	女	博士	¥11185.00
8	老贼	1989年04月25日	女	本科	¥8648.00
9	林郡	1979年11月24日	女	初中	¥14763.00
10	梁炜	1967年05月02日	女	博士	¥11643.00
11	赵谦	1974年08月26日	男	本科	¥13126.00
12	孙犁	1968年05月19日	男	本科	¥8904.00
13	小周	1988年07月01日	女	中专	¥1118.00

显示格式使报表的可读性更强，但并不会改变单元格的真实数据

列表式报表——显示值

因业务需要及设计数据库的考虑，从数据集中取出的很多度量字段的值往往是编码，如员工表的性别，存的是0或1，但报表中希望显示成“男”或“女”

select Name from 学历表 where ID=9

编号	姓名	生日	性别	学历	奖金
1	张三	1974-08-22	2	9	14256.00
2	李四	1985-10-01	1	3	3291.00
3	小王	1981-09-29	2	3	10851.00
4	刘淇	1987-05-03	1	6	8583.00
5	小霸	1977-11-17	2	7	10936.00
6	小究	1979-02-13	2	8	16867.00
7	小蒋	1976-06-27	2	7	11185.00
8	老贼	1989-04-25	2	5	8648.00
9	林郡	1979-11-24	2	2	14763.00
10	梁炜	1967-05-02	2	7	11643.00
11	赵谦	1974-08-26	1	5	13126.00
12	孙犁	1968-05-19	1	5	8904.00
13	小周	1988-07-01	2	9	1118.00

显示值设置前后

编号	姓名	生日	性别	学历	奖金
1	张三	1974年08月22日	女	中专	¥14256.00
2	李四	1985年10月01日	男	高中	¥3291.00
3	小王	1981年09月29日	女	高中	¥10851.00
4	刘淇	1987年05月03日	男	硕士	¥8583.00
5	小霸	1977年11月17日	女	博士	¥10936.00
6	小究	1979年02月13日	女	博士后	¥16867.00
7	小蒋	1976年06月27日	女	博士	¥11185.00
8	老贼	1989年04月25日	女	本科	¥8648.00
9	林郡	1979年11月24日	女	初中	¥14763.00
10	梁炜	1967年05月02日	女	博士	¥11643.00
11	赵谦	1974年08月26日	男	本科	¥13126.00
12	孙犁	1968年05月19日	男	本科	¥8904.00
13	小周	1988年07月01日	女	中专	¥1118.00

虽编码字段在报表中显示成对应的中文，但该单元格的真实编码仍然保留，当别的单元格引用该单元格时，引用的是编码而非中文

列表式报表—— 报表区域

除了数据区外，还可以在报表中设置表头区和标题区，当显示多条数据时，表头被冻结，数据区可以上下拖动

还有分页页眉、页脚等控制打印的选项

		A	B	C	D	E	F
分页页眉	1(眉)	北京润乾软件			创新技术推动应用进步		
头标题	2(标)	员工信息表					
报表头	3(头)	编号	姓名	生日	性别	学历	奖金
数据区	4	=ds1.S	=ds1.EMPNO	=ds1.BIRTHDAY	=ds1.G	=ds1.DEGR	=ds1.BONUS
报表尾	5(尾)					统计时间:	=now()
分页页脚	6(脚)						="第"&pno0+"&

北京润乾软件

创新技术推动应用进步

分页页眉

员工信息表

编号	姓名	生日	性别	学历	奖金
1	张三	1974年08月22日	女	中专	¥14256.00
2	李四	1985年10月01日	男	高中	¥3291.00
3	小王	1981年09月29日	女	高中	¥10851.00
4	刘琪	1987年05月03日	男	硕士	¥8583.00
5	小胡	1977年11月17日	女	博士	¥10936.00
6	小光	1979年02月13日	女	博士后	¥16867.00
7	小陈	1976年06月27日	女	博士	¥11185.00
8	老威	1989年04月25日	女	本科	¥8648.00
9	林娜	1979年11月24日	女	初中	¥14763.00
10	梁炜	1967年05月02日	女	博士	¥11643.00
11	赵谦	1974年08月26日	男	本科	¥13126.00
12	孙翠	1968年05月19日	男	本科	¥8904.00
13	小周	1988年07月01日	女	中专	¥1118.00
14	温望	1984年07月02日	男	博士	¥15158.00
15	宝钗	1988年10月25日	女	小学	¥3013.00
16	戴玉	1966年06月15日	男	硕士	¥13461.00
17	小玉	1981年03月05日	女	大专	¥18265.00
18	喜人	1966年05月18日	男	博士	¥10830.00
19	振春	1991年01月11日	男	硕士	¥15947.00
21	小高	1974年08月26日	男	本科	¥13126.00
22	小武	1988年07月01日	女	中专	¥1118.00
23	小温	1984年07月02日	男	博士	¥15158.00
24	小戴	1966年06月15日	男	硕士	¥13461.00
25	宝玉	1981年03月05日	女	大专	¥18265.00

统计时间: 2020-04-27

第1页

分页页脚

码: 1/5

到第 1 页

最前页

上一页

下一页

最后页

放大

缩小

比例 90%

水平对齐 居中

分栏 1

打印页 全部

打印缩放 无

☐ 镜像打印

☐ 黑白打印

☐ 清除图层

页面设置

打印

关闭

列表式报表—— 条件格式

当报表中的数据可以根据其取值设置不同的显示格式，比如超出范围是高亮，也就是常说的条件预警

编号	姓名	生日	性别	学历	奖金
1	张三	1974年08月22日	女	中专	¥ 14256.00
2	李四	1985年10月01日	男	高中	¥ 3291.00
3	小王	1981年09月29日	女	高中	¥ 10851.00
4	刘淇	1987年05月03日	男	硕士	¥ 8583.00
5	小霸	1977年11月17日	女	博士	¥ 10936.00
6	小究	1979年02月13日	女	博士后	¥ 16867.00
7	小蒋	1976年06月27日	女	博士	¥ 11185.00
8	老贼	1989年04月25日	女	本科	¥ 8648.00
9	林郡	1979年11月24日	女	初中	¥ 14763.00
10	梁炜	1967年05月02日	女	博士	¥ 11643.00
11	赵谦	1974年08月26日	男	本科	¥ 13126.00
12	孙犁	1968年05月19日	男	本科	¥ 8904.00
13	小周	1988年07月01日	女	中专	¥ 1118.00
14	温望	1984年07月02日	男	博士	¥ 15158.00

如右图所示：当员工奖金高于15000时用红色字体显示，这样可以更加清晰的标记出高奖金的员工

列表式报表— 条件格式

相邻的数据行显示不同的背景色，这样不仅美观，也有助于横向查看数据时避免错行，更加清晰准确

编号姓名	生日	性别	学历	奖金
1张三	1974年08月22日	女	中专	¥14256.00
2李四	1985年10月01日	男	高中	¥3291.00
3小王	1981年09月29日	女	高中	¥10851.00
4刘淇	1987年05月03日	男	硕士	¥8583.00
5小霸	1977年11月17日	女	博士	¥10936.00
6小究	1979年02月13日	女	博士后	¥16867.00
7小蒋	1976年06月27日	女	博士	¥11185.00
8老贼	1989年04月25日	女	本科	¥8648.00
9林郡	1979年11月24日	女	初中	¥14763.00
10梁炜	1967年05月02日	女	博士	¥11643.00
11赵谦	1974年08月26日	男	本科	¥13126.00
12孙犁	1968年05月19日	男	本科	¥8904.00
13小周	1988年07月01日	女	中专	¥1118.00
14温望	1984年07月02日	男	博士	¥15158.00
15宝钗	1988年10月25日	女	小学	¥3013.00
16戴玉	1966年06月15日	男	硕士	¥13461.00
17小玉	1981年03月05日	女	大专	¥18265.00
18喜人	1966年05月18日	男	博士	¥10830.00

如右图所示，偶数行背景色显示为浅蓝色，奇数行背景色显示为白色

列表式报表—— 计算列

报表可以基于数据集中已有字段计算出新的字段并呈现，称为计算列

津贴权重系数	
范围	值
<5000	0.4
5000-10000	0.25
>10000	0.1

编号	姓名	生日	性别	学历	奖金	津贴
1	张三	1974年08月22日	女	中专	¥14256.00	¥1425.60
2	李四	1985年10月01日	男	高中	¥3291.00	¥1316.40
3	小王	1981年09月29日	女	高中	¥10851.00	¥1085.10
4	刘琪	1987年05月03日	男	硕士	¥8583.00	¥2145.75
5	小霸	1977年11月17日	女	博士	¥10936.00	¥1093.60
6	小究	1979年02月13日	女	博士后	¥16867.00	¥1686.70
7	小蒋	1976年06月27日	女	博士	¥11185.00	¥1118.50
8	老贼	1989年04月25日	女	本科	¥8648.00	¥2162.00
9	林郡	1979年11月24日	女	初中	¥14763.00	¥1476.30
10	梁炜	1967年05月02日	女	博士	¥11643.00	¥1164.30
11	赵谦	1974年08月26日	男	本科	¥13126.00	¥1312.60
12	孙犁	1968年05月19日	男	本科	¥8904.00	¥2226.00
13	小周	1988年07月01日	女	中专	¥1118.00	¥447.20
14	温望	1984年07月02日	男	博士	¥15158.00	¥1515.80
15	宝钗	1988年10月25日	女	小学	¥3013.00	¥1205.20
16	黛玉	1966年06月15日	男	硕士	¥13461.00	¥1346.10
17	小玉	1981年03月05日	女	大专	¥18265.00	¥1826.50
18	喜人	1966年05月18日	男	博士	¥10830.00	¥1083.00
合计:					¥194898.00	¥25636.65

津贴的算法：奖金*津贴权重系数（按奖金的取值范围定义系数值）

津贴=奖金*0.25

津贴=奖金*0.1

津贴=奖金*0.4

奖金（列）合计

津贴（列）合计

分组式报表— 基本特征

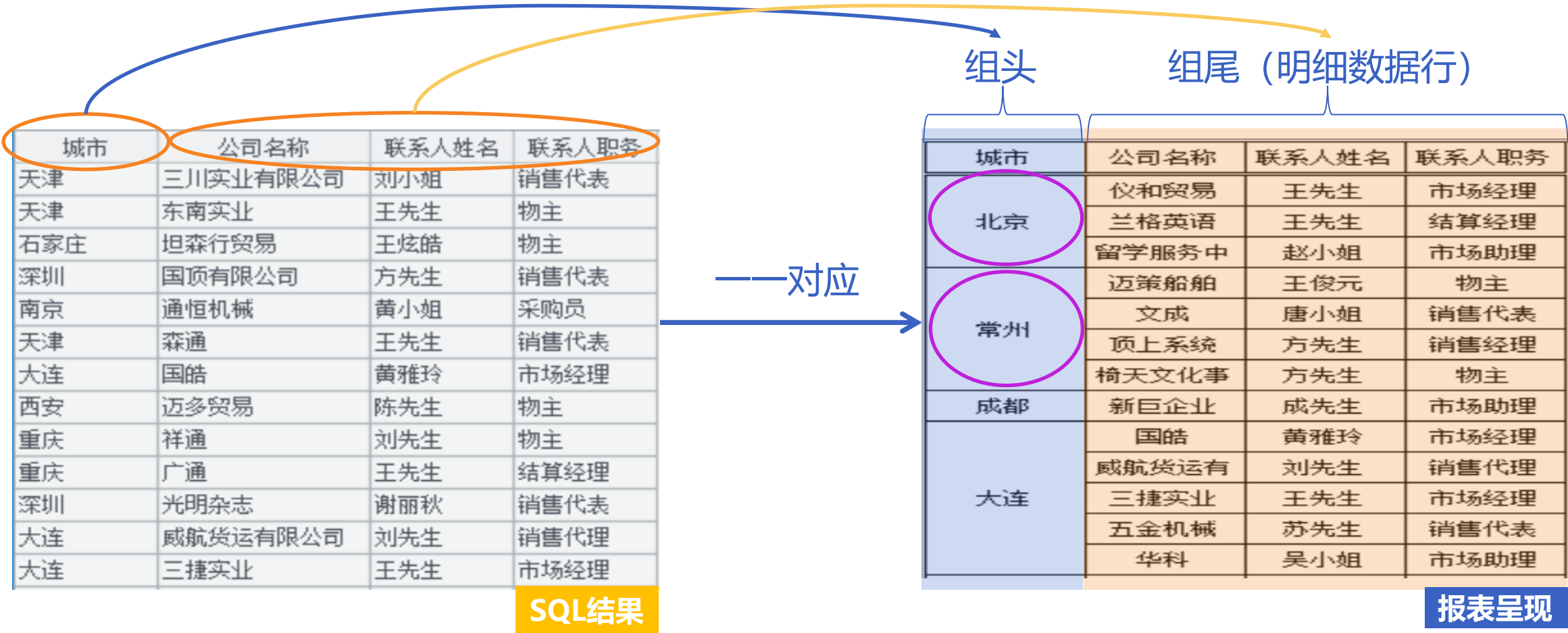
分组式报表是以分组的形式展现报表，使用户清晰快速的了解到组内信息以及组间的信息

供应商	类别	产品	产品名称	库存量	订购量	单价
妙生	调味品	8	胡椒粉	6	0	¥ 40.00
	库存金额合计：					¥ 240.00
为全	肉/家禽	9	鸡	29	0	¥ 97.00
	海鲜	10	蟹	31	0	¥ 31.00
	库存金额合计：					¥ 3774.00
日正	日用品	11	大众奶酪	22	30	¥ 21.00
		12	德国奶酪	86	0	¥ 38.00
	库存金额合计：					¥ 462.00
德昌	海鲜	13	龙虾	24	0	¥ 6.00
	库存金额合计：					¥ 144.00
总库存量：		198		总库存金额：		¥ 4620.00

分组式

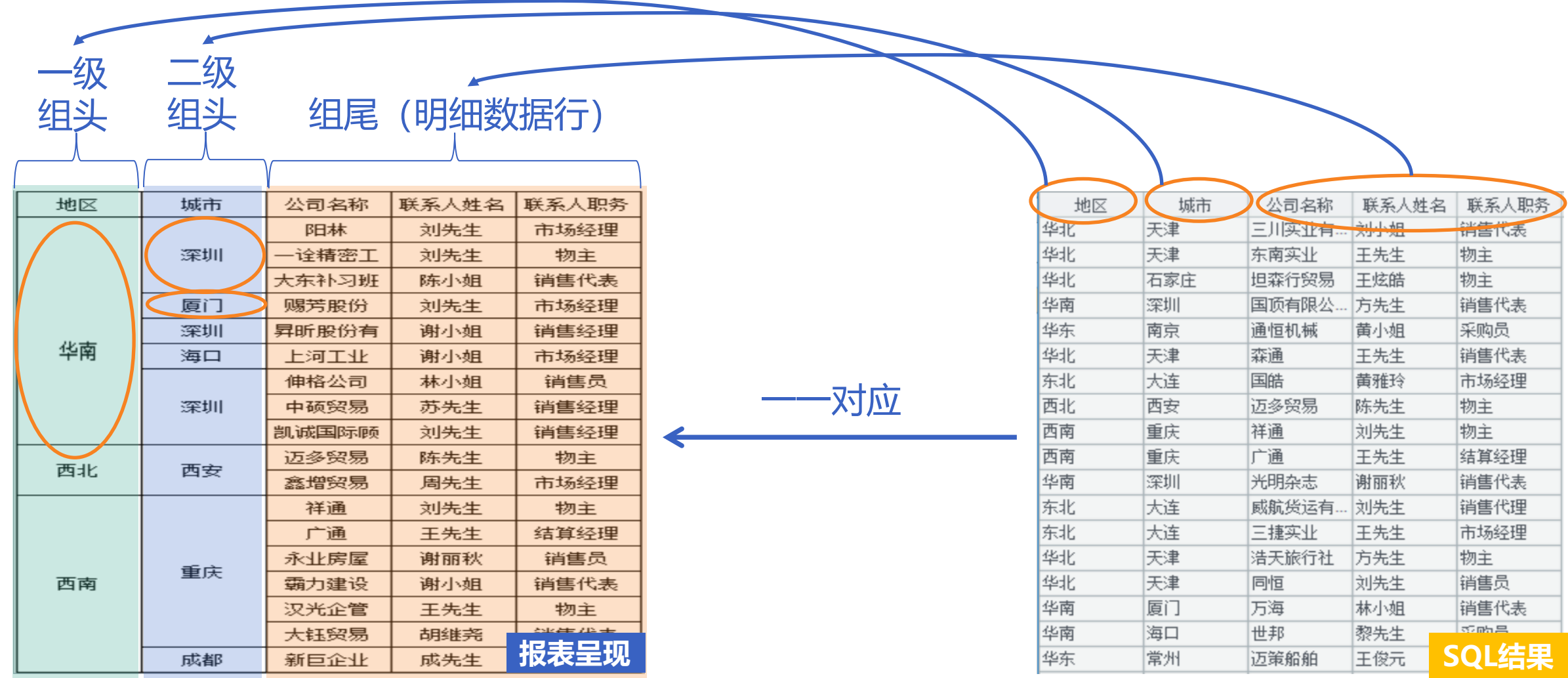
分组式报表——数据与报表的关系

分组式报表有组头和组尾，与数据集（一般指SQL）分出来的组——对应
数据集组内的明细数据和报表内的明细行也是一一对应



分组式报表——数据与报表的关系（多层分组）

多层分组报表（多级组头和组尾），与数据集多层分组和组内明细一一对应
若整个数据集被看成一组，列表式报表本质上是分组式报表的一种特殊情况



分组式报表—— 报表区域

和前面列表式报表同理，除了数据区外，也会有表头、表尾等区域

其中分组表头指当分组跨多页时，每页都能重复显示分组表头信息

		A	B	C	D	E
报表头	1(头)	地区	城市	公司名称	联系人姓名	联系人职务
组表头	2(组)	=ds1.groupk				
	3(组)		=ds1.Groupk			
数据区	4			=ds1.Select	=ds1.联系人	=ds1.联系人
报表尾	5(尾)		公司数	=count(C4{})		

报表结果

第1页				
地区	城市	公司名称	联系人姓名	联系人职务
华南	海口			
		世邦	黎先生	采购员
		保信人寿	方先生	物主
		上河工业	谢小姐	市场经理
	深圳			
		国顶有限公	方先生	销售代表
		光明杂志	谢丽秋	销售代表
		正人资源	谢小姐	销售经理
		红阳事业	王先生	市场助理
		东旗	王先生	市场经理
		远东开发	王先生	销售代表
		春永建设	王先生	市场经理

第2页				
地区	城市	公司名称	联系人姓名	联系人职务
华南	深圳			
		阳林	刘先生	市场经理
		一诺精密工	刘先生	物主
		大东补习班	陈小姐	销售代表
		昇昕股份有	谢小姐	销售经理
		伸格公司	林小姐	销售员
		中硕贸易	苏先生	销售经理
		凯诚国际顾	刘先生	销售经理
	厦门			
		万海	林小姐	销售代表
		兴中保险	方先生	物主
		赐芳股份	刘先生	市场经理
	公司数	20		

分组式报表—— 计算列

分组报表中往往要显示包含明细的多层数据，不能用SQL的GROUP BY来分组
分组动作，以及相关的汇总统计是报表完成的（汇总统计可以看成一种计算列）

```
SELECT 供应商,类别,产品名称,库存量,单价 FROM 产品
```

报表模板

	A	B	C	D	E	F
1(头)	供应商	类别	产品名称	库存量	单价	库存金额
2	=ds1.group(供应	=ds1.group(类别	=ds1.select(产品	=ds1.库存量	=ds1.单价	=D2*E2
3					库存金额合计:	=sum(F2:F)
4		总库存量:	=sum(D2:D)		总库存金额合计:	=sum(F2:F)

报表汇总表达式

报表计算后

报表汇总——

一级组头		二级组头		组尾（来自数据集）		报表汇总
供应商	类别	产品名称	库存量	单价		库存金额
妙生	调味品	胡椒粉	6	¥ 40.00		¥ 240.00
	库存金额合计:					¥ 240.00
为全	肉/家禽	鸡	29	¥ 97.00		¥ 2813.00
	海鲜	蟹	31	¥ 31.00		¥ 961.00
	库存金额合计:					¥ 3774.00
日正	日用品	大众奶酪	22	¥ 21.00		¥ 462.00
		德国奶酪	86	¥ 38.00		¥ 3268.00
	库存金额合计:					¥ 3730.00
德昌	海鲜	龙虾	24	¥ 6.00		¥ 144.00
	库存金额合计:					¥ 144.00
	总库存量:	198			总库存金额合计:	¥ 7888.00

交叉式报表— 基本特征

交叉式报表是行列方向都有分组的报表，帮助用户了解行列各组及相互之间的整体情况

交叉式

运货商	货主地区	东北	华东	华中	华北	华南	西北	西南
	订单数量							
急速快递		20	70	1	97	23	2	35
统一包裹		25	89	1	148	39	3	27
联邦货运		15	79	0	108	29	2	23
空运		0	1	0	1	1	0	0
货到付款		0	2	0	1	0	0	0
普通快运		0	0	0	5	1	0	0

交叉式报表——数据与报表的关系

交叉报表的表头是由上表头和左表头组成，可理解为观察数据的角度，用SQL写出来就是GROUP BY 后面的分组字段

中间区域为横纵两列两个坐标共同决定每个格子的数值，一般为聚合值

```
SELECT TITLE,DEGREE,COUNT(EMPID) FROM
EMPLOYEE GROUP BY TITLE,DEGREE
```

SQL解释数据与交叉表对应关系

报表呈现										
上表头		小学	初中	高中	大专	本科	硕士	博士	博士后	中专
DEGREE	TITLE	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)
初级		0	0	0	0	4	2	2	0	2
中级		2	0	2	0	0	4	0	0	0
高级		0	0	2	2	2	0	4	0	0
特高级		0	2	0	0	0	0	4	2	2

左表头

中间区域

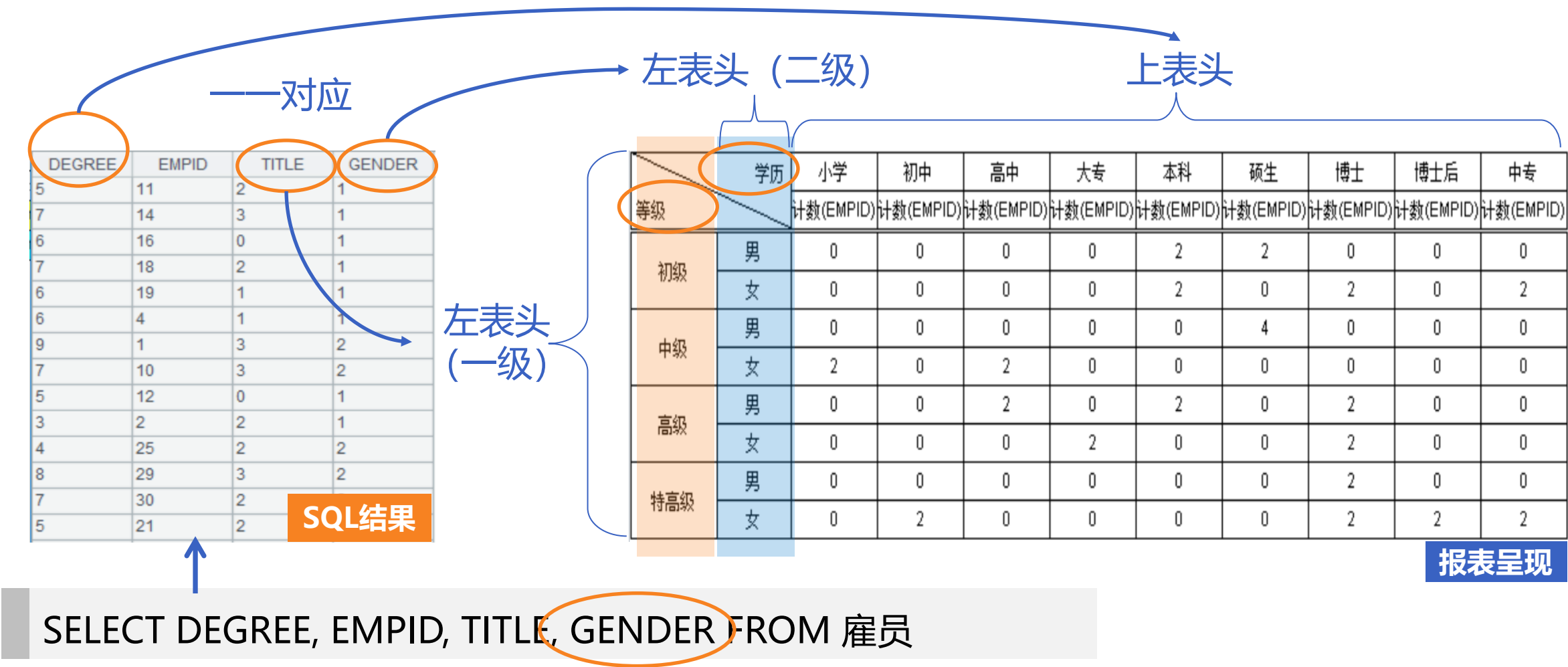
交叉式报表——数据与报表的关系

交叉报表里，中间区域每个格一般都是聚合值，但在某些情况下其实还是单值



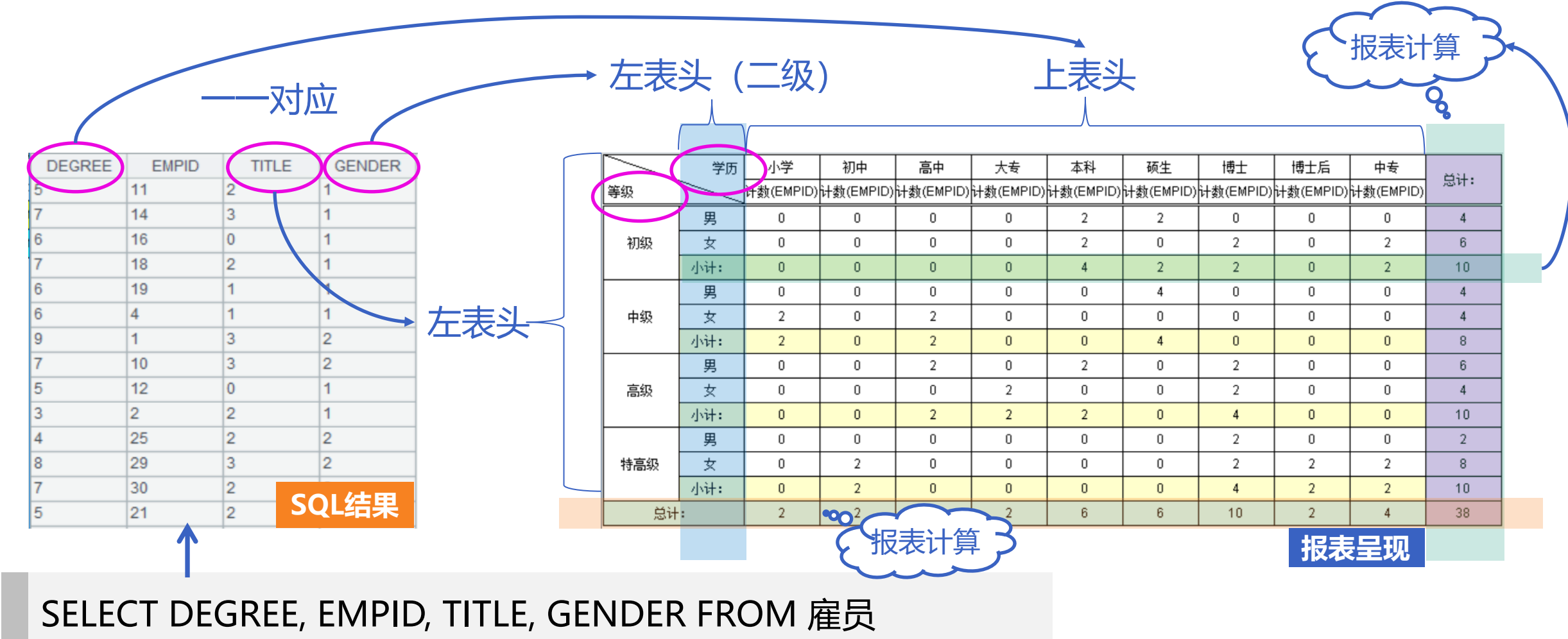
交叉式报表——数据与报表的关系（多层交叉）

在行列方向上通过增加多个分组维度，就形成了多层交叉分组报表



交叉式报表— 计算列

在上一步的基础上，增加行、列上的汇总统计列，形成多层交叉汇总表
汇总统计（计算列）在报表的不同位置上，将会有不同的表现形式



交叉式报表——报表区域

和前面的列表式、分组式报表同理，除了数据区外，也有表头、表尾等区域

		A	B	C	D
头标题	1(标)	员工学历、等级交叉统计表			
	2(标)			报表日期: =now()	
报表头	3(头)		学历	=ds1.Group(总计:
	4(头)		等级	计数(EMPID)	
数据区	5			=ds1.Count(=sum(C5{))
	6	=ds1.Group(小计:	=sum(C5{))	=sum(C5{))
	7	总计:		=sum(C5{))	=sum(C5{))
报表尾	8(尾)				= "第" + pno() +
尾标题	9(注)			审核员: 张三	

<div> 学历 小学 初中 高中 大专 本科 硕士 博士 博士后 中专 总计: </div>		小学	初中	高中	大专	本科	硕士	博士	博士后	中专	总计:
		计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)	计数(EMPID)
初级	男	0	0	0	0	2	2	0	0	0	4
	女	0	0	0	0	2	0	2	0	2	6
	小计:	0	0	0	0	4	2	2	0	2	10
中级	男	0	0	0	0	0	4	0	0	0	4
	女	2	0	2	0	0	0	0	0	0	4
	小计:	2	0	2	0	0	4	0	0	0	8
高级	男	0	0	2	0	2	0	2	0	0	6
	女	0	0	0	2	0	0	2	0	0	4
	小计:	0	0	2	2	2	0	4	0	0	10
特高级	男	0	0	0	0	0	0	2	0	0	2
	女	0	2	0	0	0	0	2	2	2	8
	小计:	0	2	0	0	0	0	4	2	2	10
总计:		2	2	4	2	6	6	10	2	4	38

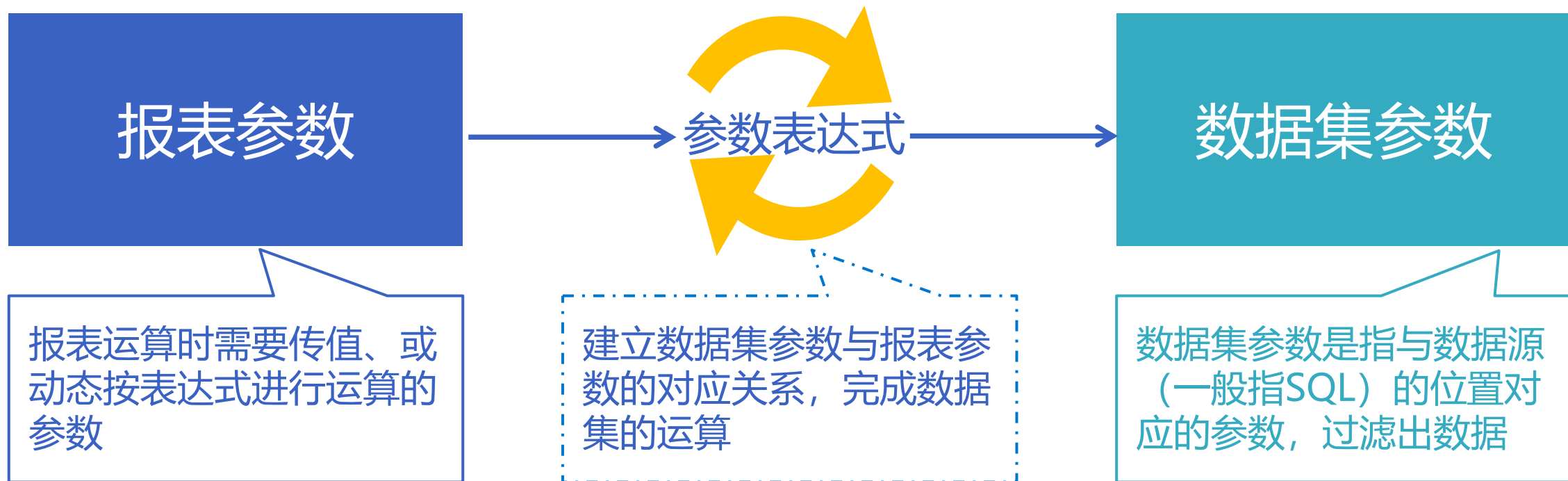


参数

参数

在设计一张报表时，很少有数据是固定不变的，常常需要根据一些条件来过滤出想要的数
据，这就需要通过参数控制报表中的数据

报表中所使用的参数有“报表参数”和“数据集参数”之分,报表参数与数据集参数通过“参
数表达式”建立联系



参数— 简单参数查询

比如：按任意地区查询订单信息；报表的取数SQL写成如下含有参数的形式：

```
SELECT 订单ID,客户ID,订购日期,货主地区,运货商,运货费 FROM 订单 WHERE 货主地区=?
```

华东

华东

华北

华南

东北

西南

华中

西北

查询

建立数据集参数与报表参数的对应关系

订单ID	客户ID	订购日期	货主地区	运货费
10249	东希望	2012年07月05日	华东	¥ 11.61
10251	千国	2012年07月08日	华东	¥ 41.34
10256	凯诚国际顾问公司	2012年07月15日	华东	¥ 13.97
10257	远东开发	2012年07月16日	华东	¥ 81.91
10258	正人资源	2012年07月17日	华东	¥ 140.51
10259	三捷实业	2012年07月18日	华东	¥ 3.25

订单ID	客户ID	订购日期	货主地区	运货费
10248	山泰企业	2012年07月04日	华北	¥ 32.38
10250	实翼	2012年07月08日	华北	¥ 65.83
10253	实翼	2012年07月10日	华北	¥ 58.17
10255	永大企业	2012年07月12日	华北	¥ 148.33
10260	一途精密工业	2012年07月19日	华北	¥ 55.09
10263	正人资源	2012年07月23日	华北	¥ 146.06

参数— 空值参数查询

在前面的基础上，有时第一次查询或参数不输入时，希望报表可以聪明地理解为忽略该参数，显示部分数据，而不是空白页

这种情况下，可以把没有输入的参数认为是取值为空的参数，SQL修改如下：

```
SELECT 订单ID,客户ID,订购日期,货主地区,运货商,运货费 FROM 订单 WHERE 货主地区=? OR ? IS NULL
```

华东

华东

华北

华南

东北

西南

华中

西北

查询

建立数据集参数与报表参数的对应关系

订单ID

客户ID

订购日期

货主地区

运货费

订单ID	客户ID	订购日期	货主地区	运货费
10251	千国	2012年07月08日	华东	¥41.34
10252	福星制衣厂股份有限公司	2012年07月09日	东北	¥51.30
10253	实翼	2012年07月10日	华北	¥58.17
10254	浩天旅行社	2012年07月11日	华中	¥22.98
10255	永大企业	2012年07月12日	华北	¥148.33
10256	凯诚国际顾问公司	2012年07月15日	华东	¥13.97
10257	远东开发	2012年07月16日	华东	¥81.91
10258	正人资源	2012年07月17日	华东	¥140.51

参数—思考

思考：如何查询某段订购日期范围内，
货主地区等于“华东”或为空
和
运货费大于80或为空
和
客户中包含“旅行社”或为空
等多个条件组合出来的订单？

SQL该如何写？

订单ID	客户ID	订购日期	货主地区	运货费
10251	千固	2012年07月08日	华东	¥41.34
10252	福星制衣厂股份有限公司	2012年07月09日	东北	¥51.30
10253	实翼	2012年07月10日	华北	¥58.17
10254	浩天旅行社	2012年07月11日	华中	¥22.98
10255	永大企业	2012年07月12日	华北	¥148.33
10256	凯诚国际顾问公司	2012年07月15日	华东	¥13.97
10257	远东开发	2012年07月16日	华东	¥81.91
10258	正人资源	2012年07月17日	华东	¥140.51

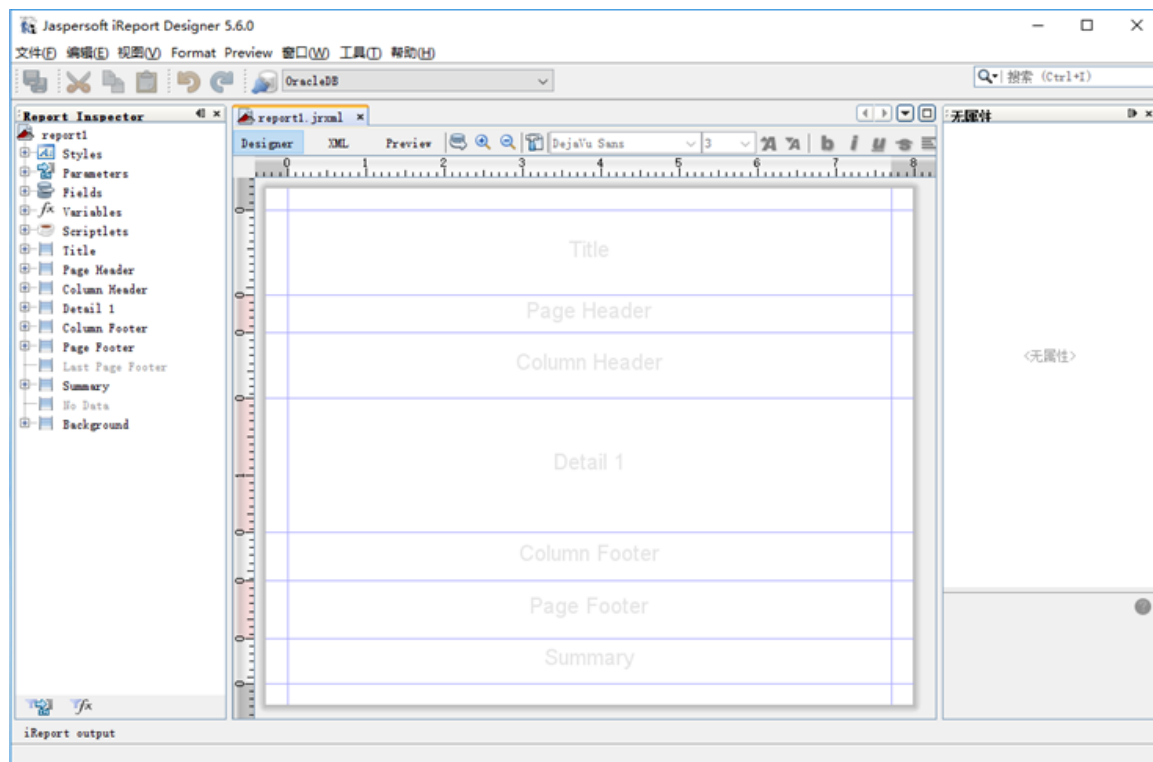


报表的开发技术

报表的开发技术— 分两大类

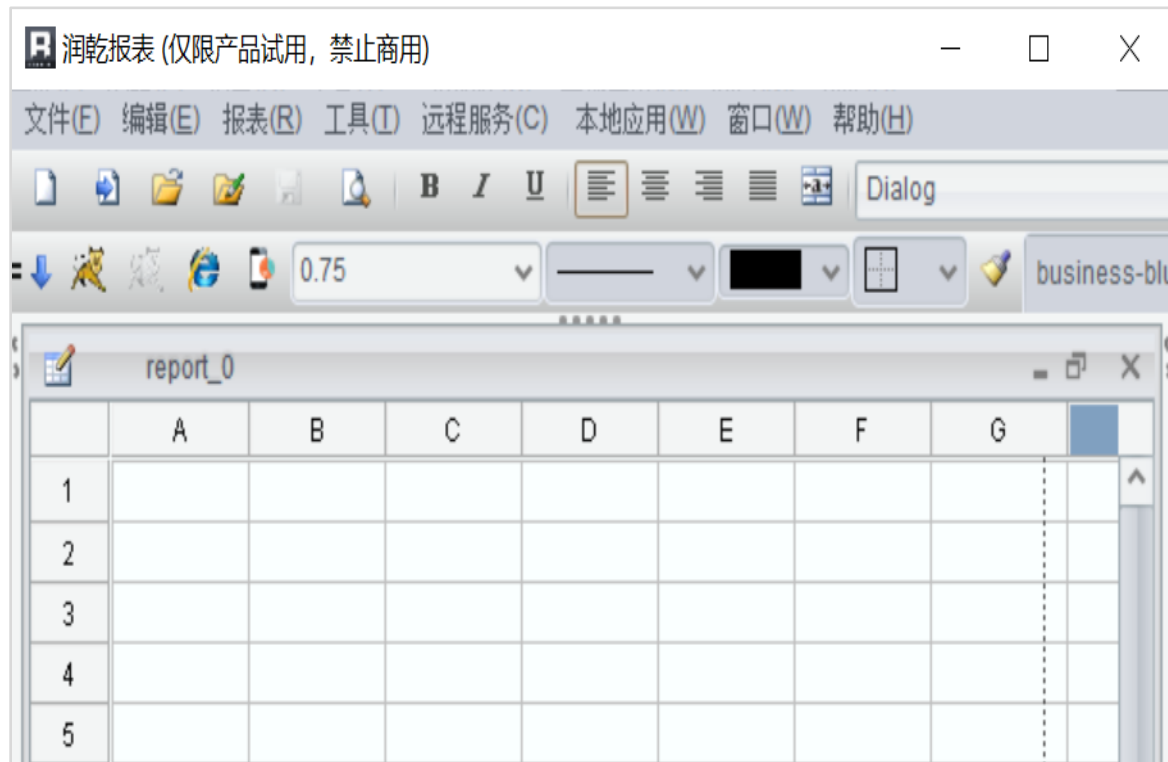
国外传统类

- 条带模型，控件拖拽式布局



新型网格类

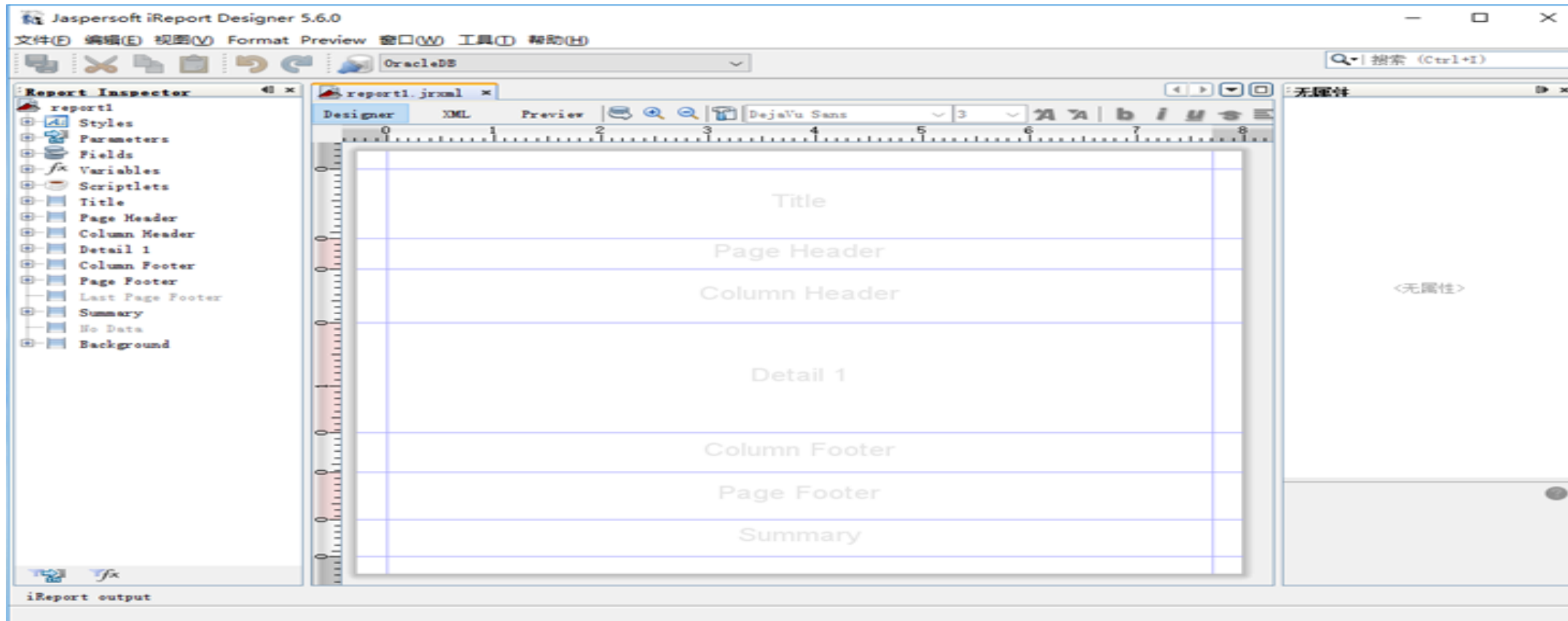
- 类Excel设计模型



报表的开发技术— 模型（国外传统类）

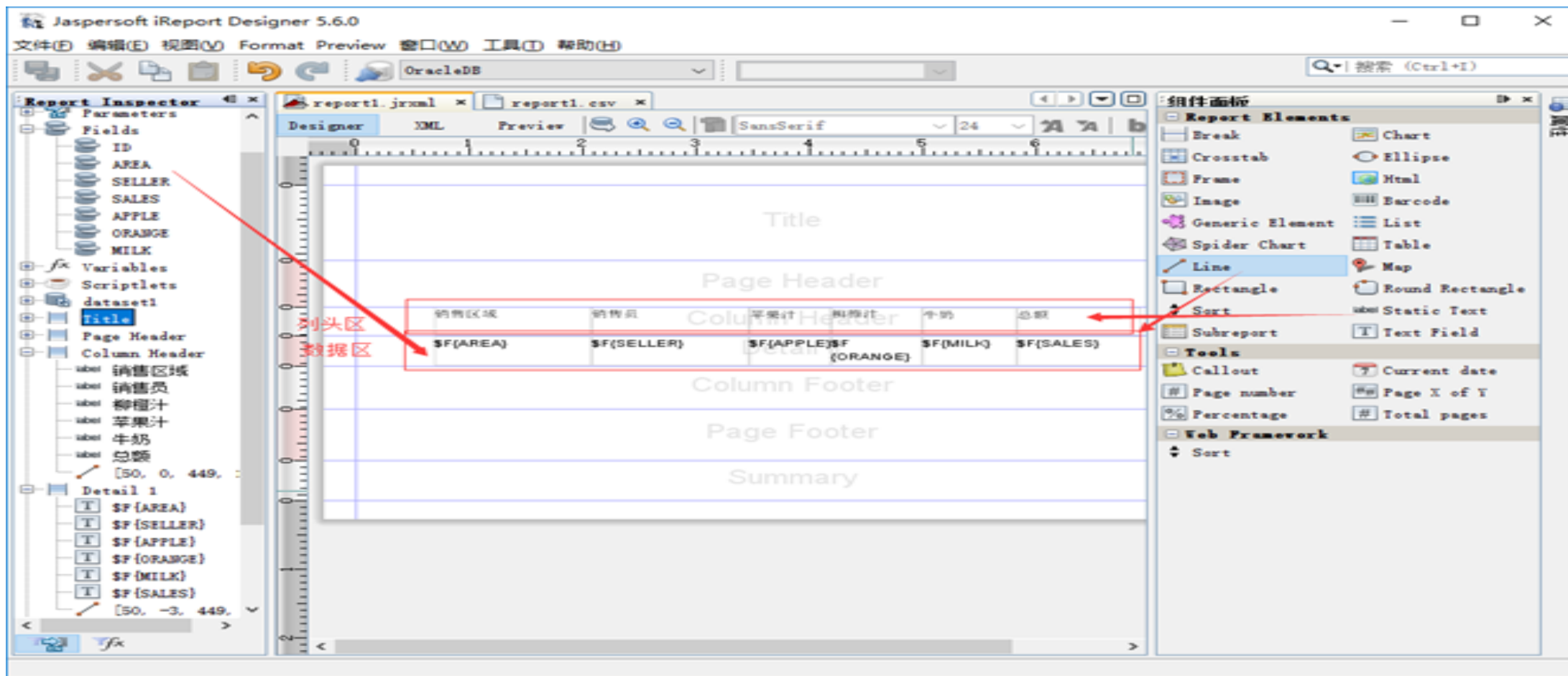
传统条带式 and 前面讲过的分组模型相匹配，每个条带和一个分组表头、表尾、明细区相对应，所以条带上会带有分组信息，从而生成分组报表

而交叉表又是另外一种模型，模型之间互不通用，某种类型只能解决某种问题



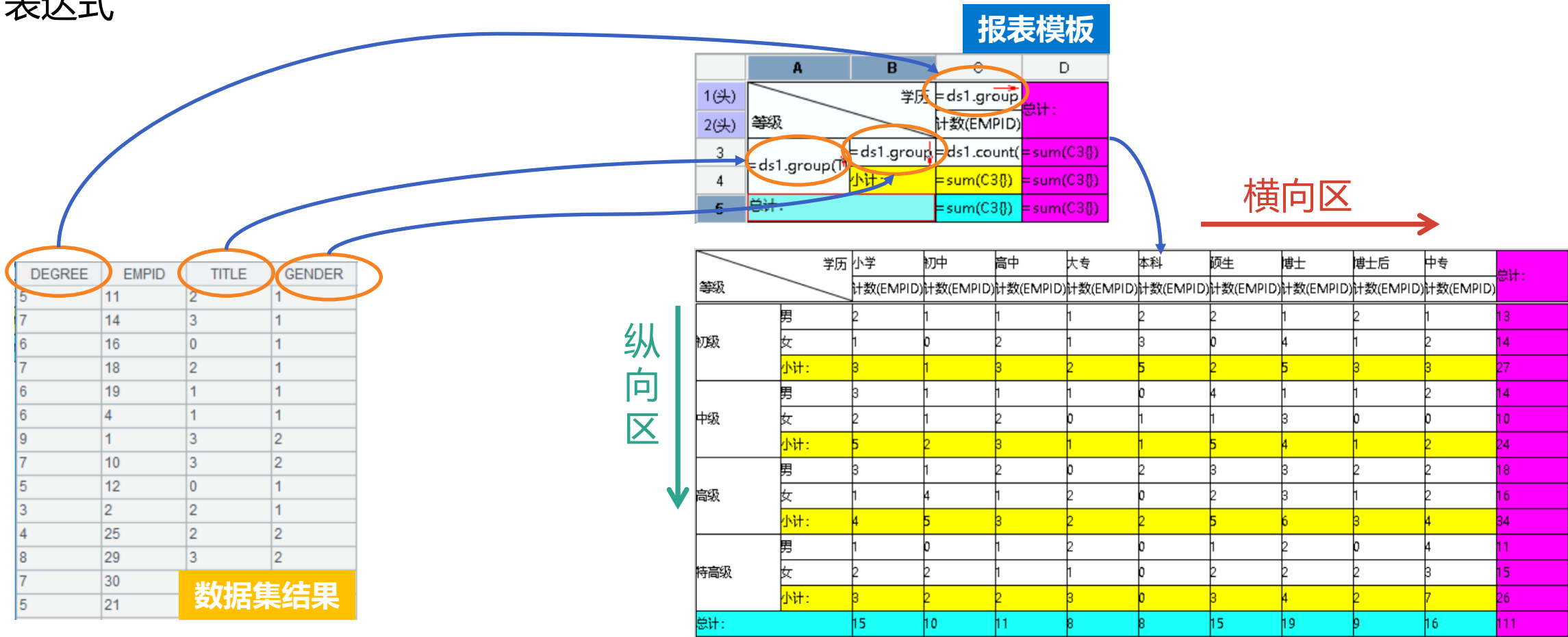
报表的开发技术— 布局（国外传统类）

界面呈条带状，只能通过手动拖拽来制作报表，这就存在位置的摆放精度，边框之间的重合、对齐等问题，制作过程友好度较低



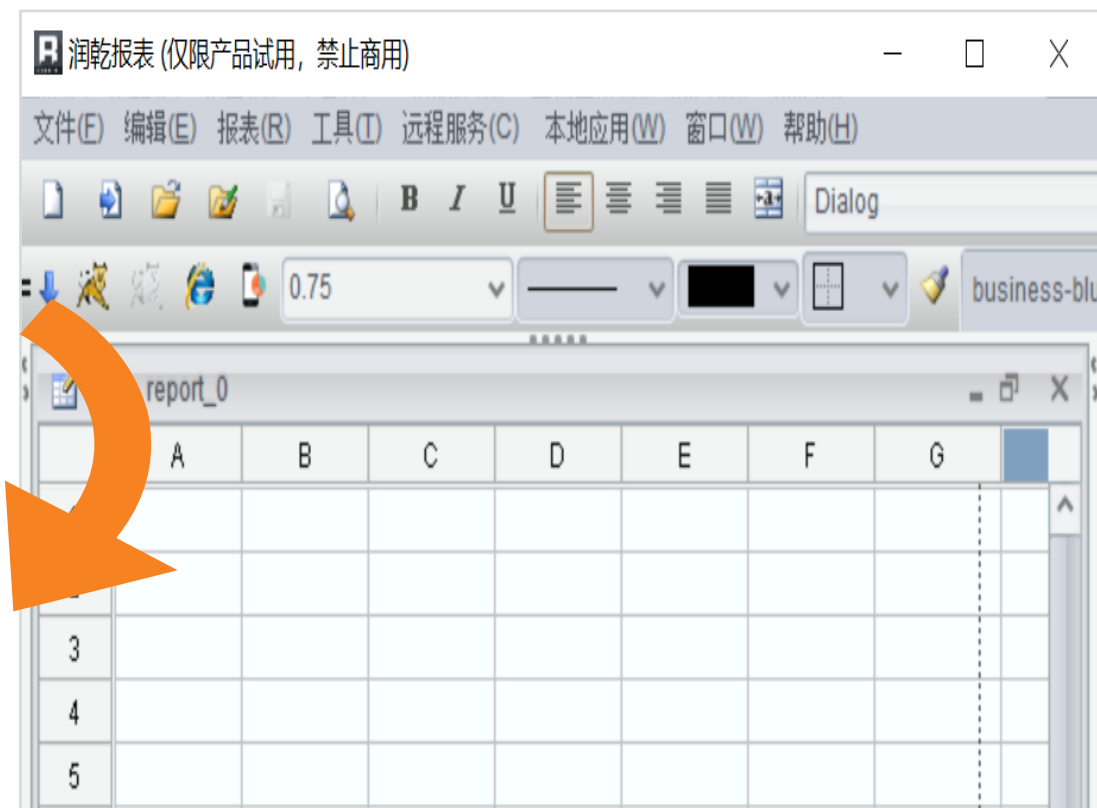
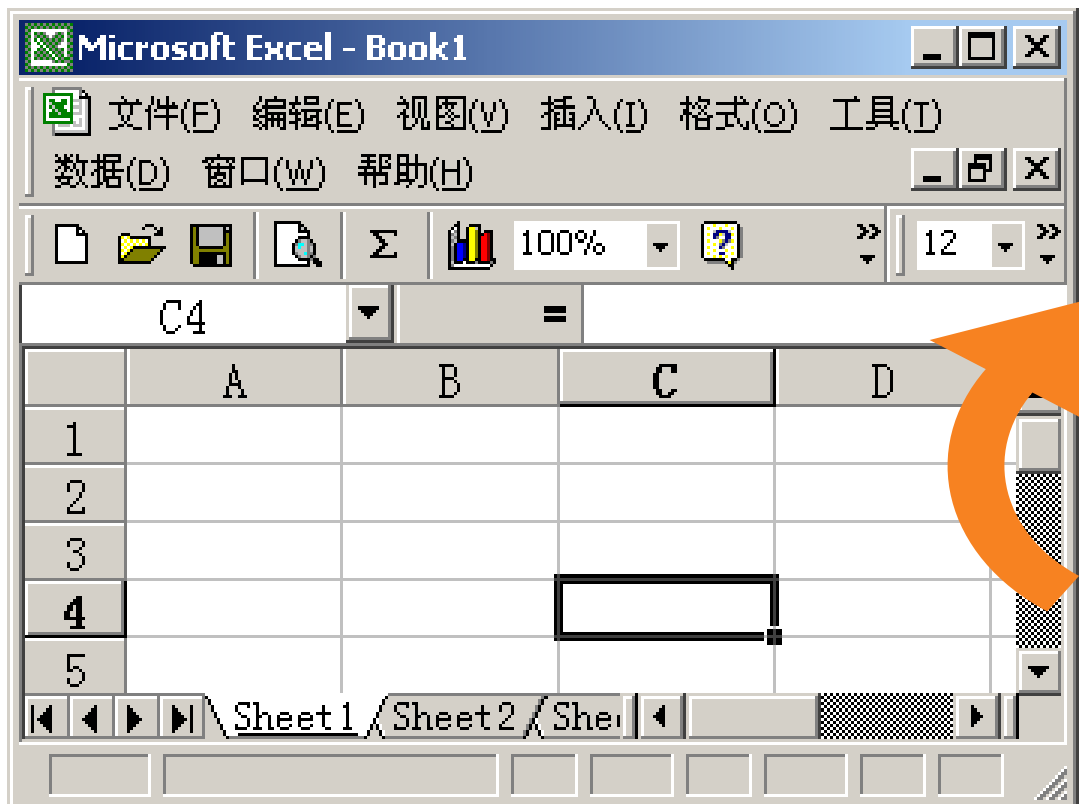
报表的开发技术— 模型（新型网格类）

网格方案采用了行列方向扩展机制，即行方向拥有的数据扩展能力完全实现在列方向上，所有报表均在统一的处理模型下完成，无须专门的交叉表处理方案；在交叉表中也可自由编写汇总表达式



报表的开发技术— 布局（新型网格类）

类Excel报表设计模型充分体现了表格的规律性，解决了传统条带式模型操作不便、难以对齐等问题，报表制作更直观、易上手



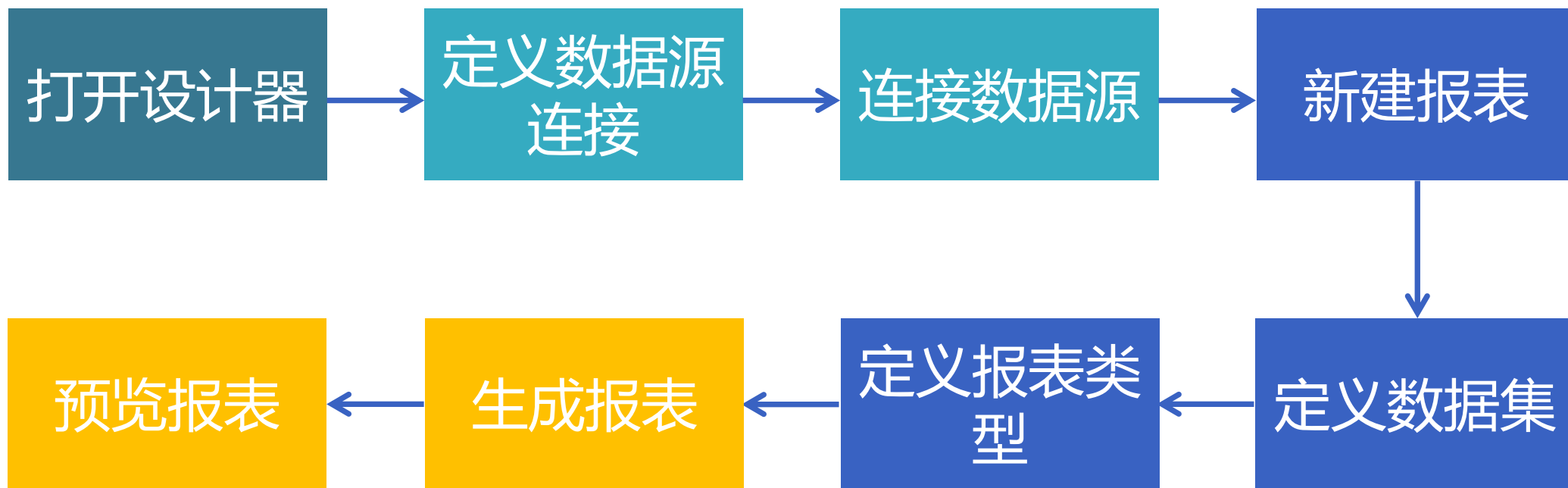


练习

练习



使用新型网格式报表工具，制作几个简单报表，实操感受一下制作报表的基本流程如下：



列表式报表— 练习（知识点）

- 主键升/降排序、显示格式、显示值
- 设置预警条件
- 隔行换色

报表模板

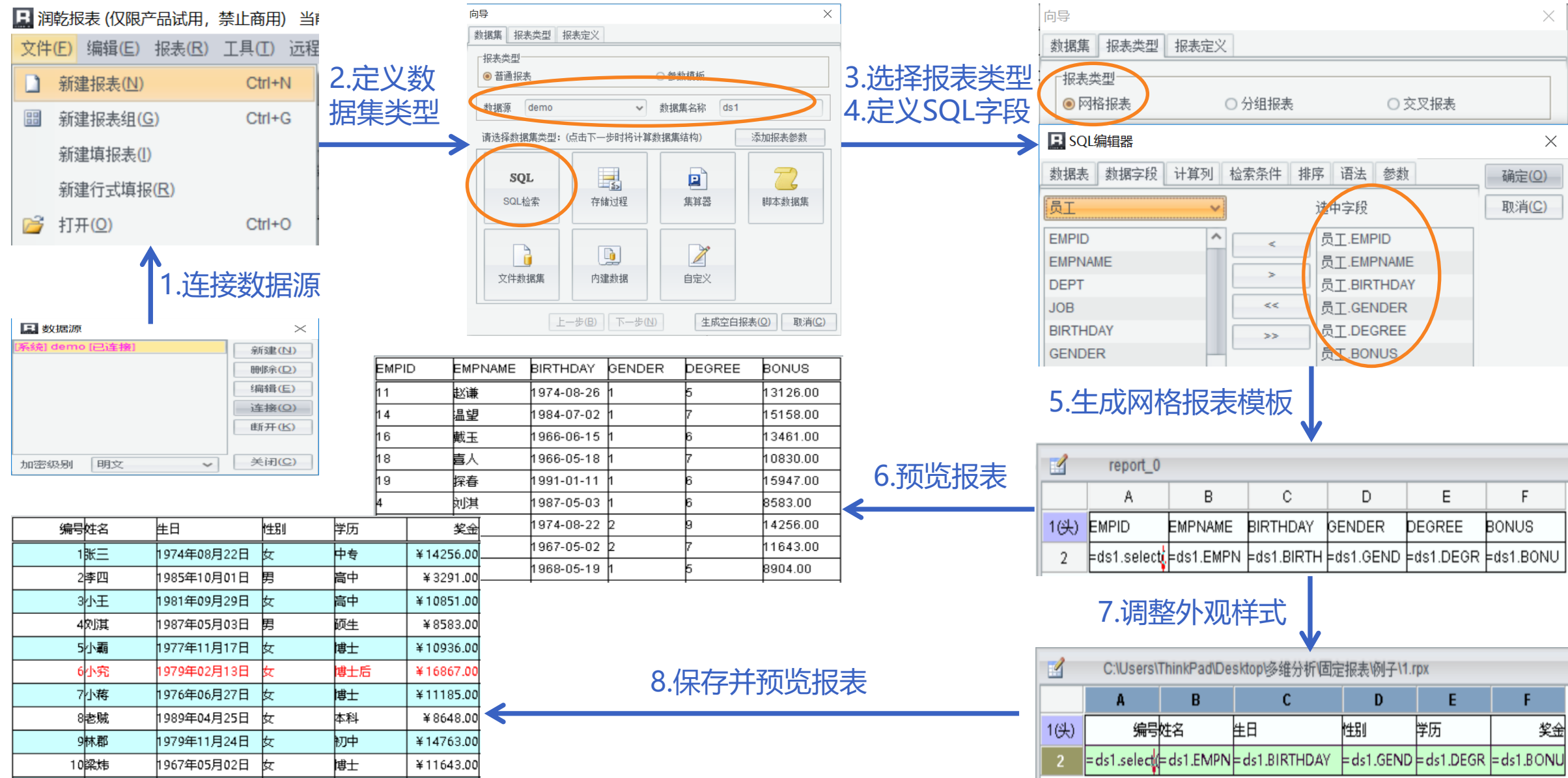
C:\Users\ThinkPad\Desktop\商业智能BI课程_v3\示例1.1.rpx						
	A	B	C	D	E	F
1(头)	编号姓名	生日	性别	学历	奖金	
2	=ds1.select	=ds1.EMPNO	=ds1.BIRTHDAY	=ds1.GENDER	=ds1.DEGREE	=ds1.BONUS

参照例子：1.1.rpx

报表预览

编号	姓名	生日	性别	学历	奖金
1	张三	1974年08月22日	女	中专	14256.00
2	李四	1985年10月01日	男	高中	3291.00
3	小王	1981年09月29日	女	高中	10851.00
4	刘琪	1987年05月03日	男	硕士生	8583.00
5	小霸	1977年11月17日	女	博士	10936.00
6	小究	1979年02月13日	女	博士后	16867.00
7	小蒋	1976年06月27日	女	博士	11185.00
8	老贼	1989年04月25日	女	本科	8648.00
9	林郡	1979年11月24日	女	初中	14763.00
10	梁炜	1967年05月02日	女	博士	11643.00
11	赵谦	1974年08月26日	男	本科	13126.00
12	孙犁	1968年05月19日	男	本科	8904.00
13	小周	1988年07月01日	女	中专	1118.00
14	温望	1984年07月02日	男	博士	15158.00
15	宝钗	1988年10月25日	女	小学	3013.00
16	黛玉	1966年06月15日	男	硕士生	13461.00
17	小玉	1981年03月05日	女	大专	18265.00
18	喜人	1966年05月18日	男	博士	10830.00

列表式报表— 制作流程



分组式报表— 练习（知识点）

- 分组格式
- 小计
- 总计

报表模板

	A	B	C	D
1(头)	城市	公司名称	联系人姓名	联系人职务
2	=ds1.group	=ds1.select	=ds1.联系人	=ds1.联系人
3		小计:	=count(B2:B)	
4	总计:	=count(B2:B)		

参照例子: 1.2.rpx

报表预览

城市	公司名称	联系人姓名	联系人职务
厦 门	万海	林小姐	销售代表
	兴中保险	方先生	物主
	赐芳股份	刘先生	市场经理
	小计：	3	
张 家 口	建资	陈先生	销售经理
	康毅系统	林彩瑜	销售代表
	志远有限公	王小姐	物主/市场助
	小计：	3	
重 庆	祥通	刘先生	物主
	广通	王先生	结算经理
	永业房屋	谢丽秋	销售员
	霸力建设	谢小姐	销售代表
	汉光企管	王先生	物主
	大钰贸易	胡继尧	销售代表
	小计：	6	
总计：	91		

分组式报表— 制作流程

- 1.连接数据源
- 2.定义数据集类型
- 3.选择报表类型
- 4.定义SQL字段

向导

数据表 报表类型 报表定义

报表类型

☐ 网格报表 ☒ 分组报表 ☐ 交叉报表

SQL编辑器

数据表 数据字段 计算列 检索条件 排序 语法 参数

客户

客户ID 公司名称 联系人姓名 联系人职务

选中字段

客户城市 客户公司名称 客户联系人姓名 客户联系人职务

5.定义分组字段

向导

数据集 报表类型 报表定义

分组定义

城市(升)

显示列

☐ 城市 ☒ 公司名称 ☒ 联系人姓名 ☒ 联系人职务

排序类型

☐ 无序 ☒ 升序 ☐ 降序

位置移动

上移 下移

分组汇总

设置分组汇总的函数

计数

上一步 下一步 生成分组报表 取消

6.生成分组报表模板

城市	公司名称	联系人姓名	联系人职务
厦 门	万海	林小姐	销售代表
	兴中保险	方先生	物主
	赐芳股份	刘先生	市场经理
	小计：	3	
张 家 口	建资	陈先生	销售经理
	康毅系统	林彩瑜	销售代表
	志远有限公	王小姐	物主/市场助
	小计：	3	
重 庆	祥通	刘先生	物主
	广通	王先生	结算经理
	永业房屋	谢丽秋	销售员
	丽力建设	谢小姐	销售代表
	汉光企管	王先生	物主
	大钰贸易	胡继尧	销售代表
	小计：	6	
总计：	91		

城市	公司名称	联系人姓名	联系人职务
北京	仪和贸易	王先生	市场经理
	兰格英语	王先生	结算经理
	留学服务中	赵小姐	市场助理
	迈策船舶	王俊元	物主
常州	文成	唐小姐	销售代表
	顶上系统	方先生	销售经理
	椅天文化事	方先生	物主
	新巨企业	成先生	市场助理

7.预览报表

report_3

	A	B	C	D
1(头)	城市	公司名称	联系人姓名	联系人职务
2	=ds1.group	=ds1.select	=ds1.联系人	=ds1.联系人

8.增加汇总, 调整外观

	A	B	C	D
1(头)	城市	公司名称	联系人姓名	联系人职务
2	=ds1.group	=ds1.select	=ds1.联系人	=ds1.联系人
3		小计:	=count(B2:D2)	
4	总计:	=count(B2:D2)		

9.保存并预览报表

交叉式报表— 练习 (知识点)

- 交叉格式
- 设置显示值
- 增加分组列
- 小计、总计

报表模板

	A	B	C	D
1(头)	等级	学历	=ds1.group	总计:
2(头)			计数(EMPID)	
3	=ds1.group(1)		=ds1.group	=sum(C3)
4	小计:		=sum(C3)	=sum(C3)
5	总计:		=sum(C3)	=sum(C3)

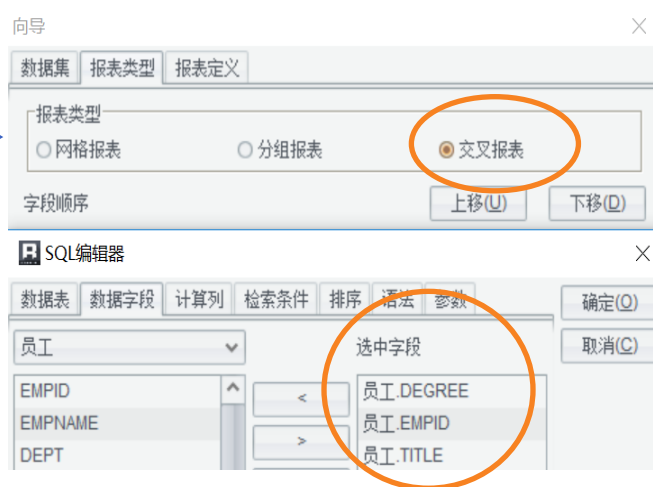
参照例子: 1.3.rpx

报表预览

		小学	初中	高中	大专	本科	硕士	博士	博士后	中专	总计:
等级	男	2	1	1	1	2	2	1	2	1	13
	女	1	0	2	1	3	0	4	1	2	14
小计:		3	1	3	2	5	2	5	3	3	27
中级	男	3	1	1	1	0	4	1	1	2	14
	女	2	1	2	0	1	1	3	0	0	10
	小计:	5	2	3	1	1	5	4	1	2	24
高级	男	3	1	2	0	2	3	3	2	2	18
	女	1	4	1	2	0	2	3	1	2	16
	小计:	4	5	3	2	2	5	6	3	4	34
特高级	男	1	0	1	2	0	1	2	0	4	11
	女	2	2	1	1	0	2	2	2	3	15
	小计:	3	2	2	3	0	3	4	2	7	26
总计:		15	10	11	8	8	15	19	9	16	111

交叉式报表—制作流程

- 1.连接数据源
- 2.定义数据集类型
- 3.选择报表类型
- 4.定义SQL字段



- 5.定义交叉分组



- 6.生成交叉报表模板

	A	B
1(头)	DEGREE	=ds1.group(
2(头)	TITLE	计数(EMPID)
3	=ds1.group(TITLE	=ds1.count()

- 7.预览报表

DEGREE \ TITLE	1	2	3	4	5	6	7	8	9
0	3	1	3	2	5	2	5	3	3
1	5	2	3	1	1	5	4	1	2
2	4	5	3	2	2	5	6	3	4
3	3	2	2	3	0	3	4	2	7

- 8.增加汇总, 调整外观

	A	B	C	D
1(头)	等级	学历	=ds1.group	总计:
2(头)		计数(EMPID)		
3	=ds1.group(TITLE	小计:	=ds1.count(=sum(C3))
4			=sum(C3))	=sum(C3))
5	总计:		=sum(C3))	=sum(C3))

- 9.保存预览

等级 \ 学历		小学	初中	高中	大专	本科	硕士	博士	博士后	中专	总计:
初级	男	2	1	1	1	2	2	1	2	1	13
	女	1	0	2	1	3	0	4	1	2	14
	小计:	3	1	3	2	5	2	5	3	3	27
中级	男	3	1	1	1	0	4	1	1	2	14
	女	2	1	2	0	1	1	3	0	0	10
	小计:	5	2	3	1	1	5	4	1	2	24
高级	男	3	1	2	0	2	3	3	2	2	18
	女	1	4	1	2	0	2	3	1	2	16
	小计:	4	5	3	2	2	5	6	3	4	34
特高级	男	1	0	1	2	0	1	2	0	4	11
	女	2	2	1	1	0	2	2	2	3	15
	小计:	3	2	2	3	0	3	4	2	7	26
总计:		15	10	11	8	8	15	19	9	16	111

参数— 练习 (知识点)

以列表式报表举例

输入不同参数值，查询不同报表结果

过滤条件为：姓名为空
和
奖金>5000

姓名:

奖金 > 5000

查询



报表预览

编号	姓名	生日	性别	学历	奖金
1	张三	1974年08月22日	女	中专	¥ 14256.00
3	小王	1981年09月29日	女	高中	¥ 10851.00
4	刘淇	1987年05月03日	男	硕生	¥ 8583.00
5	小霸	1977年11月17日	女	博士	¥ 10936.00
6	小究	1979年02月13日	女	博士后	¥ 16867.00
7	小蒋	1976年06月27日	女	博士	¥ 11185.00
8	老贼	1989年04月25日	女	本科	¥ 8648.00
9	林郡	1979年11月24日	女	初中	¥ 14763.00
10	梁炜	1967年05月02日	女	博士	¥ 11643.00
11	赵谦	1974年08月26日	男	本科	¥ 13126.00
12	孙犁	1968年05月19日	男	本科	¥ 8904.00
14	温望	1984年07月02日	男	博士	¥ 15158.00
16	戴玉	1966年06月15日	男	硕生	¥ 13461.00
17	小玉	1981年03月05日	女	大专	¥ 18265.00
18	喜人	1966年05月18日	男	博士	¥ 10830.00
合计:					¥ 187476.00

过滤条件为：姓名包含 “小”
和
奖金>10000

姓名: 小

奖金 > 10000

查询



参照例子：1.4.rpx
1.4_arg.rpx

参数— 制作流程

参数编辑

名称	描述	数据类型	值表达式	参数类型
name	员工姓名	字符串		普通参数
bonus	奖金	数值		普通参数

确定 取消 增加 删除

1.在列表式报表的基础上，定义报表参数

编号	姓名	生日	性别	学历	奖金
3	小王	1981年09月29日	女	高中	¥10851.00
5	小霸	1977年11月17日	女	博士	¥10936.00
6	小究	1979年02月13日	女	博士后	¥16867.00
7	小蒋	1976年06月27日	女	博士	¥11185.00
17	小玉	1981年03月05日	女	大专	¥18265.00
合计:					¥68104.00

2.定义SQL参数

数据集设置

名称	类型	数据源
ds1	SQL检索	demo
ds2	SQL检索	demo

确定 取消

SQL编辑器

数据表 数据字段 计算列 检索条件 排序 语法 参数

确定 取消

员工

检索条件

EMPID

EMPNAME

DEPT

JOB

and

or

>

(员工.EMPNAME like ? OR ? is null)

AND

(员工.BONUS > ? OR ? is null)

3.定义SQL参数与报表参数的对应关系

SQL编辑器

数据表 数据字段 计算列 检索条件 排序 语法 参数

确定 取消

数据集工厂侦听器类名

数据集工厂过滤器类名

数据集记录行设置

起始行:

结束行:

记录阈值:

0

预警阈值

☐ 无数据时抛出异常

序号 参数表达式 结果类型

1 "%"+@name+"%" 默认

2 "%"+@name+"%" 默认

3 @bonus 默认

4 @bonus 默认

增加 插入 删除 上移

4.输入参数值

设置报表参数值 [双击数值列弹出编辑窗口]

名称 中文说明 类型 数值

name 员工姓名 字符串 小

bonus 奖金 数值 10000

确定 取消

5.预览报表

小结

- ✓ 通过本章的学习，我们了解了报表的概念，三种典型的简单报表（列表、分组、交叉）的基本特征，以及数据源与报表的关系
- ✓ 知道了传统条带式模型和新型网格类报表模型在报表设计上的不同
- ✓ 通过报表设计器能够制作简单的报表，掌握条件属性的概念；学会使用条件属性实现：隔行换色、字体高亮、计算列等
- ✓ 掌握了报表参数的概念、应用场景，能够使用模板参数和数据集参数实现报表查询效果

THANKS

感谢观看



乾学院 · VIP课堂

商业智能 统计图技术



CONTENTS

本章内容

1. 报表中的统计图
2. 【分类-系列】统计图
3. 非【分类-系列】统计图
4. 统计图元素
5. 终端上的统计图技术





报表中的统计图

统计图也是报表中重要的展现形式

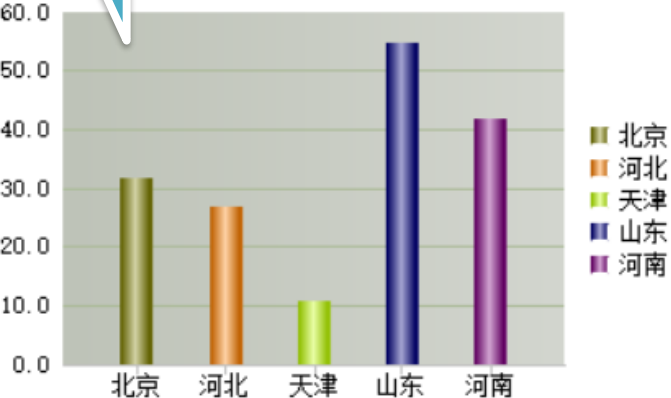


上一章介绍了各种情况的表格。实际的报表中有很多情况是图表结合的，统计图和表格都是报表中重要的展现形式，多图、表共同组成报表。

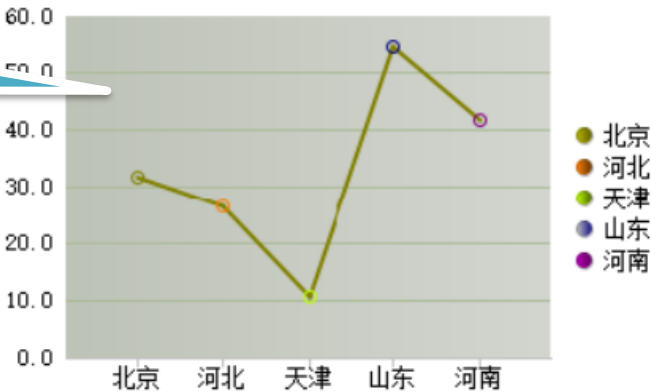


多样的统计图

柱形图



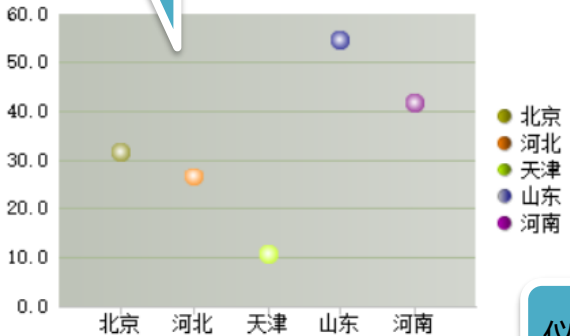
折线图



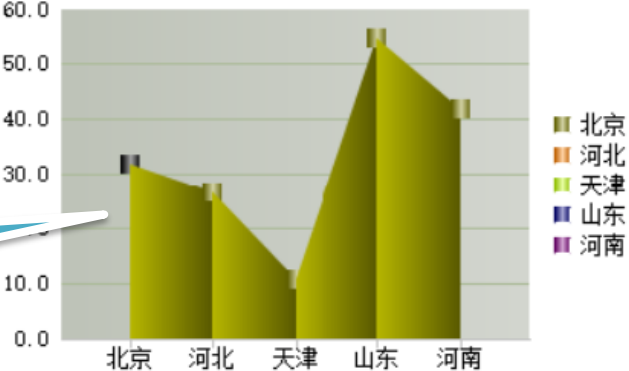
饼图



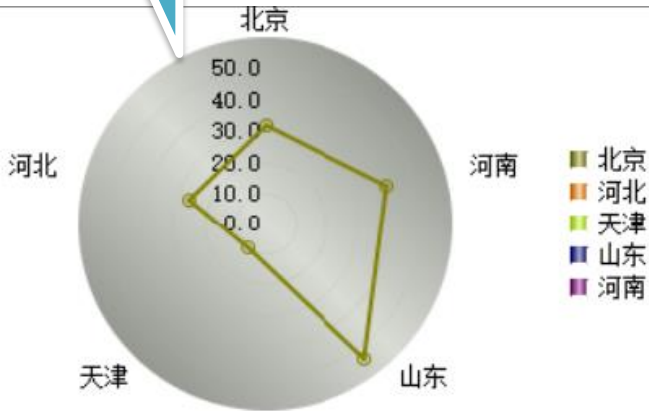
散点图



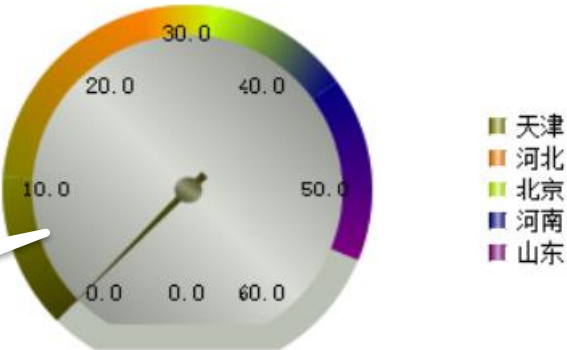
区域图



雷达图



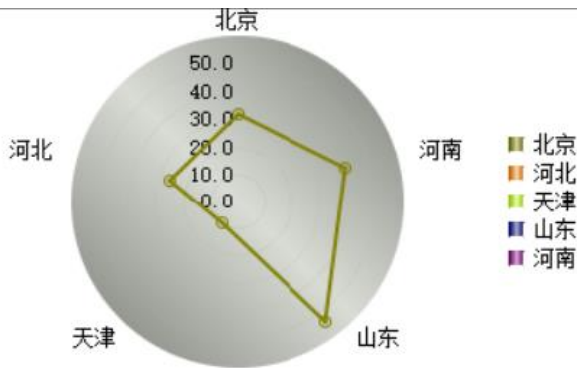
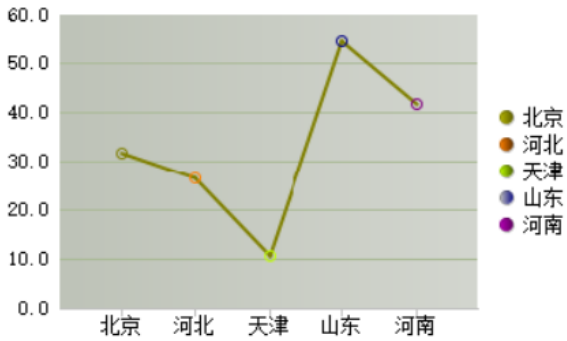
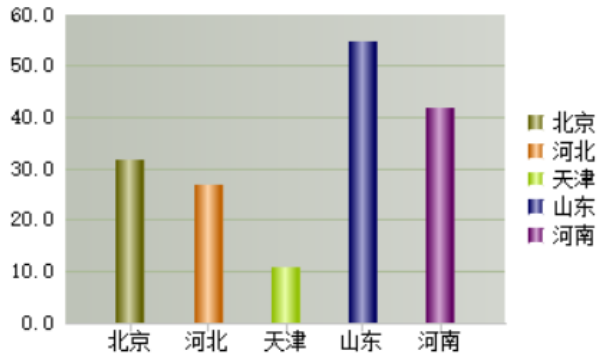
仪表盘





【分类-系列】统计图

知识点-分类、系列



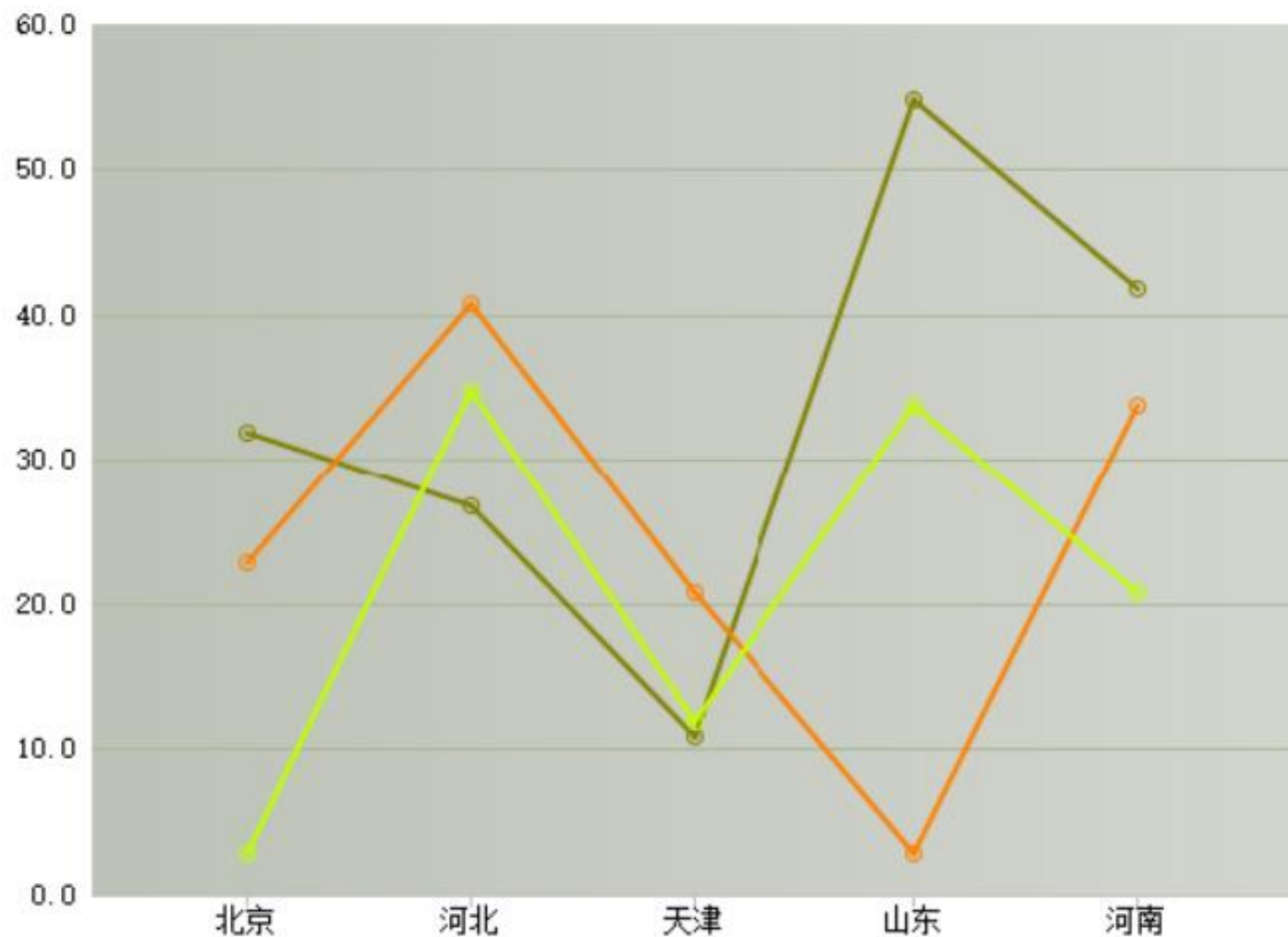
拿三个图我们仔细分析下，这些图虽然样子迥异，但表达的数据却相同，都是有北京、河北、天津、山东、河南五个省份，它们新增病例的数值都是32、27、11、55、42。都能简述成一句话：统计【省份】的【新增病例】。我们把按照【什么】统计【什么】里的第一个什么起一个名字叫【分类】，第二个什么起一个名字叫【系列】。

内建数据集	
字段	数据
省份	新增病例
北京	32
河北	27
天津	11
山东	55
河南	42

分类：是一类事物，来自数据库表的一个字段，字段值是**名称**。

系列：是分类的某种属性的**值**，在数据库表里也是一个字段。

试一试找出下图的分类和系列



分类仍然是省份，但多了两种属性值。这种是一个分类多个系列的情况，后面会详细介绍。

● 新增病例
● 新增疑似
● 治愈出院

数据相应的多了
两个系列的字段

内建数据集

字段 数据

省份	新增病例	新增疑似	治愈出院
北京	32	23	3
河北	27	41	35
天津	11	21	12
山东	55	3	34
河南	42	34	21

最常见的表现形式--图表结合



在应用系统的展示界面里，统计图很少单独出现，经常多图、多表结合的方式组成一个完整的报表。统计图属于报表的组成部分。首先获得一系列相关的数据后，才在在局部设计表或统计图。

下页开始从原始数据出发制作一个最简单的图表结合的报表。

图表结合报表—实现过程1

润乾报表 (仅限润乾公司研发人员专用, 禁止扩散)

文件(E) 工具(T) 远程服务(C) 本地应用(W) 窗口

新建报表(N) Ctrl+N

新建报表组(G) Ctrl+G

新建填报表(I)

新建行式填报(R)

1、打开润乾报表设计器, 新建报表

脚本数据集

缓存文件名

```
return file("d:/dfx/BI/doc/demos/test1.xlsx").xlsimport@tx()
```

3、这里用集算器脚本把原始数据2-1.xlsx加载进来

省份 新增病例 新增疑似 治愈出院

1	北京	32	23	3
2	河北	27	41	35
3	天津	11	21	12
4	山东	55	3	34
5	河南	42	34	21

2、弹出向导, 点击内建数据

向导

数据集 报表类型 报表定义

普通报表

数据源 ds1

请选择数据集类型:

SQL SQL检索

存储过程

集算器

脚本数据集

文件数据集

内建数据

自定义

添加报表参数

上一步(B) 下一步(N) 生成网格报表(O) 取消(C)

4、选择两列数据, 点击生成网格报表, 就创建好新报表了

显示列

☒ 省份

☒ 新增病例

☐ 新增疑似

☐ 治愈出院

report_0

1(头)	A	B
2	=ds1.select	=ds1.新增病

5、生成报表后, 点击预览按钮, 就能看到一个简陋的普通报表了

浏览报表: report_0

省份	新增病例
北京	32
河北	27
天津	11
山东	55
河南	42

图表结合报表—实现过程2

1、选中第二行，右键，追加数据行

2、选中A3/B3，合并两个格子，A3拖拽行列到合适大小

3、右键A3格，选择统计图，弹出“统计图属性定义”窗口

4、选择柱形图；添加分类轴，数据设置A2；添加系列，数据来自B2，系列名“新增病例”

5、设置完毕后，点击预览，就看到有表有图的报表效果了

1(头)	A	B
2	=ds1.select(省份)	=ds1.新增病例

统计图属性定义

图类型及数据 图形特性 显示外观 其他特性

数据集 统计图类型 柱形图 仅用前 条数据记录来画统计图

分类轴

=A2 分类轴定义 添加 删除

系列

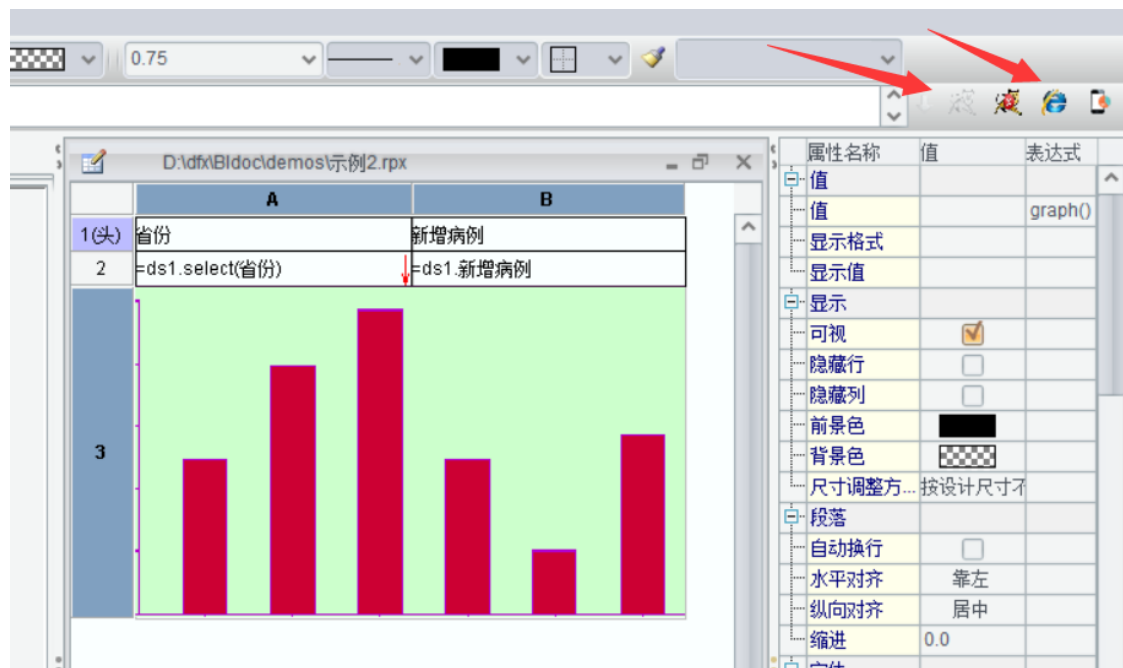
☐ 使用指定轴

系列名称	系列值	标题
=新增病例	=B2	

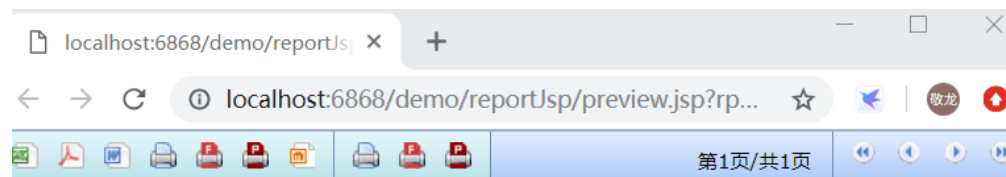
添加 删除

省份	新增病例
北京	32
河北	27
天津	11
山东	55
河南	42

报表作为整体，发布到WEB系统，在浏览器上查看

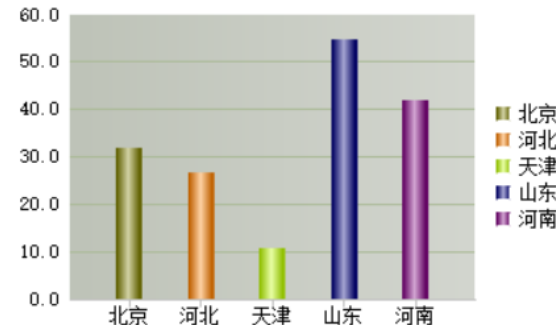


启动报表设计器内置的tomcat服务器，
然后点击浏览器按钮，就会再浏览器里
打开这个报表进行浏览了，如下。



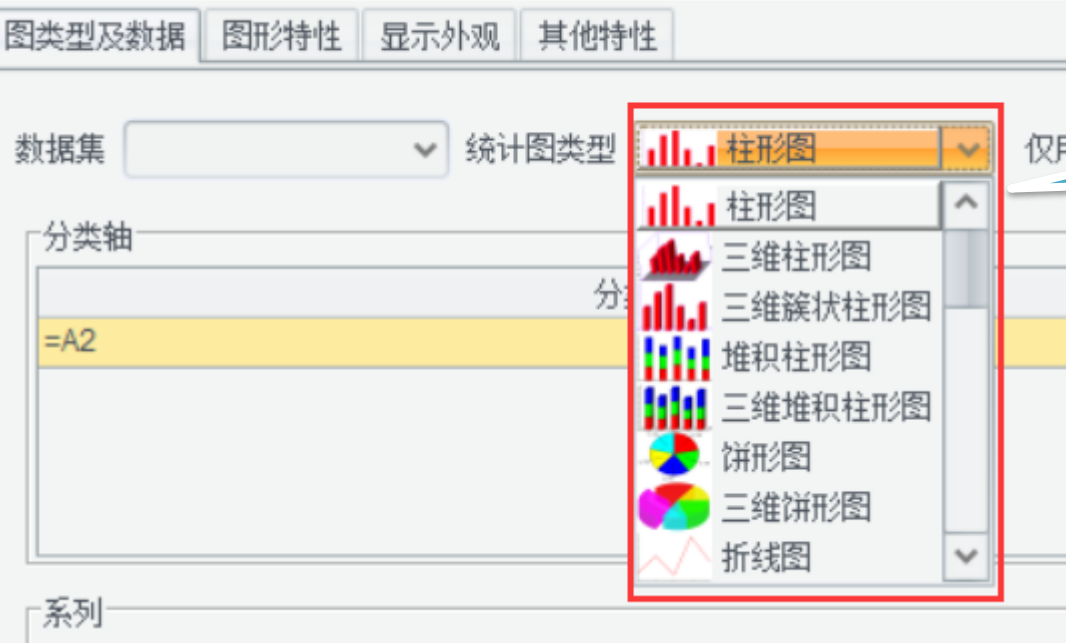
在浏览器里观察统计图部分，通过右键菜单或
开发者工具能知道它是一张图片。在服务器上
生成统计图图片是一种技术实现方式，后面还
会介绍另外一种常见的JS绘图实现方式。

省份	新增病例
北京	32
河北	27
天津	11
山东	55
河南	42



动手试一试一切换统计图类型

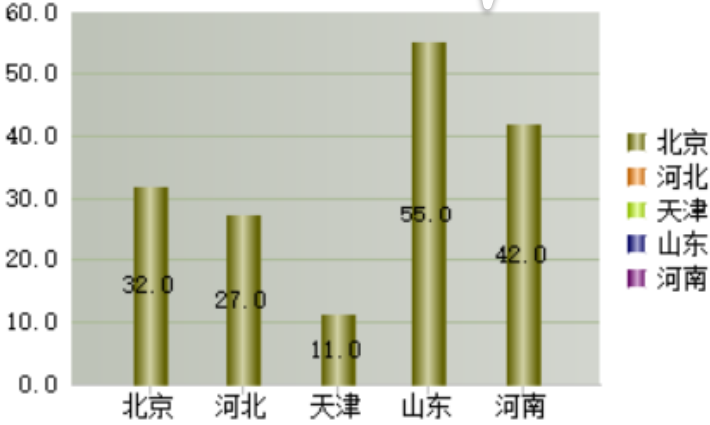
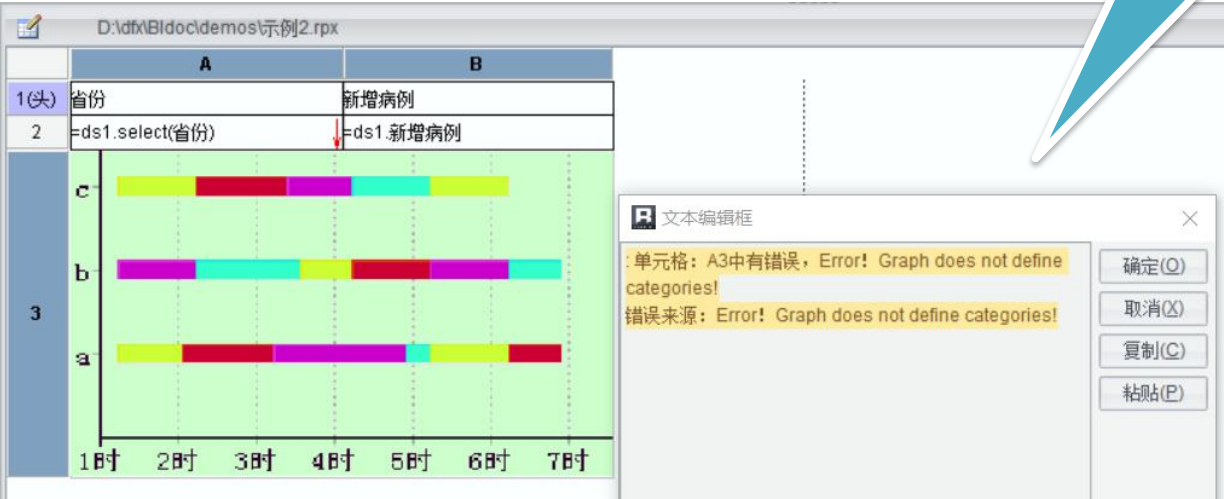
统计图属性定义



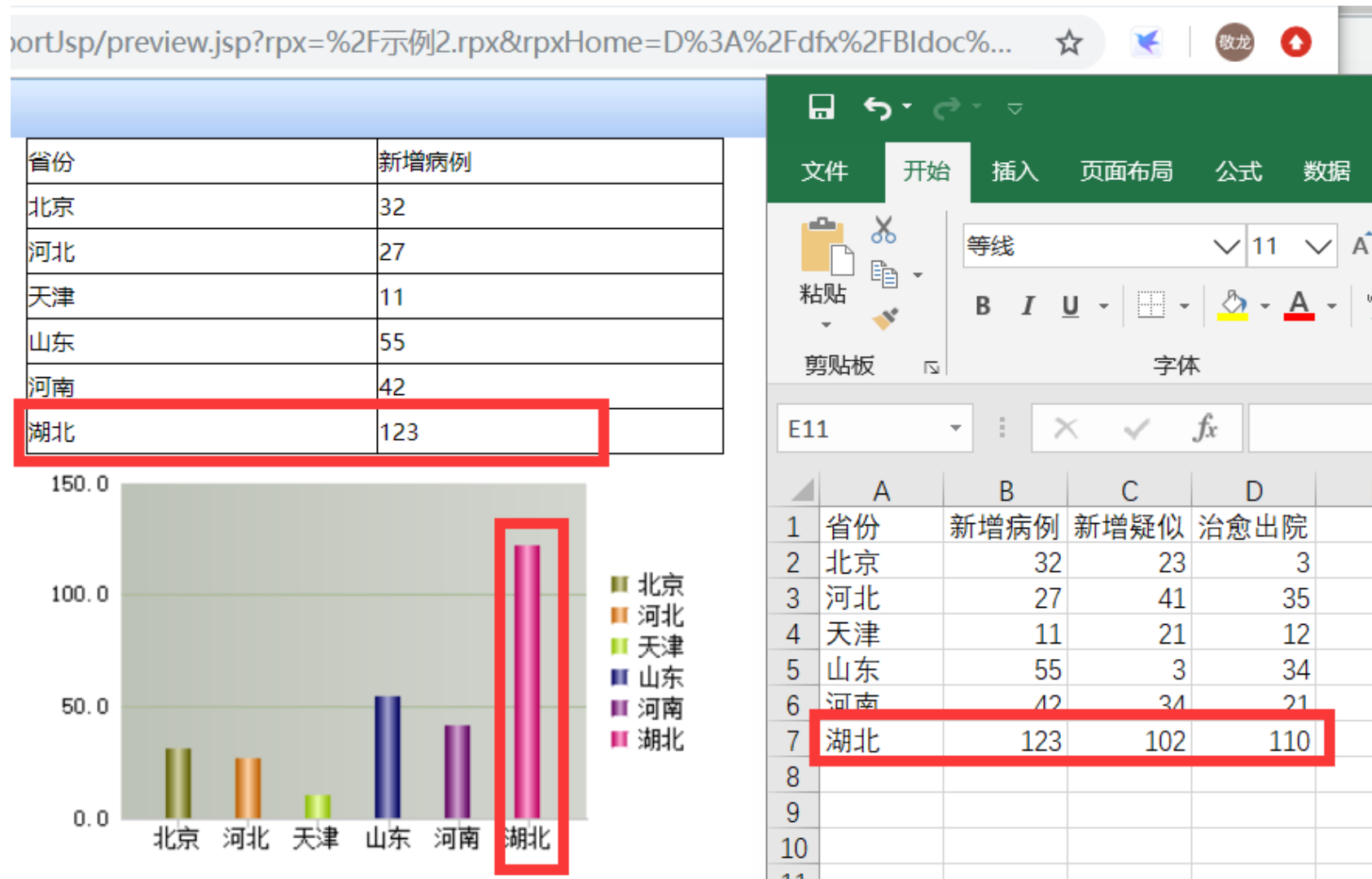
切换不同类型的统计图，预览就能看到上面列举的几种不同统计图。

切换到时序状态图，会遇到报错的情况，这是因为存在不同类型统计图支持的数据存在差异的情况。有通用的，也有不兼容的。

堆积柱形图和柱形图看起来没区别。这两种统计图表现多个系列时才会有现实差异。

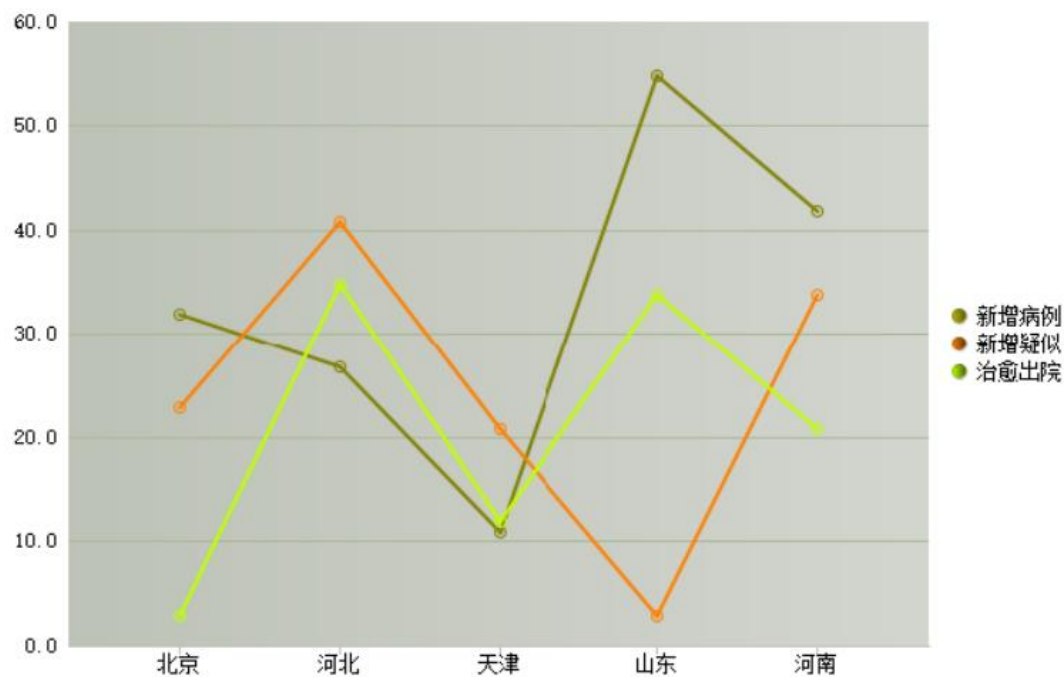
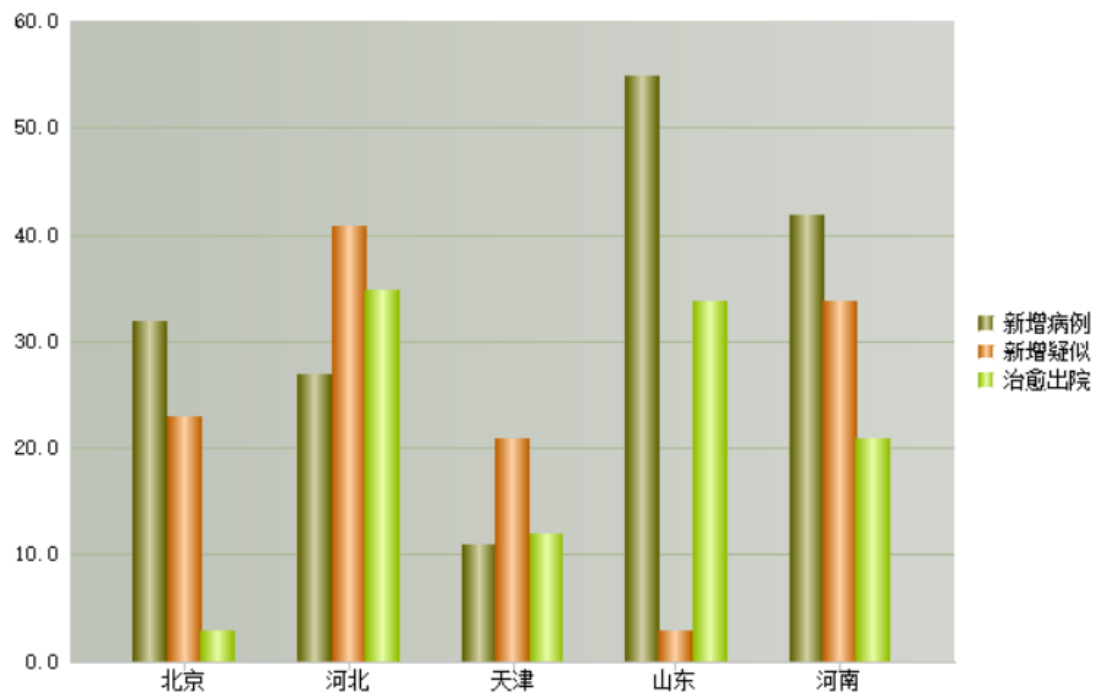


动手试一试—改变原始数据



Excel里的原始数据增加湖北的数据，保存后刷新浏览器页面，表格和统计图也更新好了

一图多系列



多套系列值共用分类。观察数据时，除了可以对比一个系列在不同分类下的值，还可以对比不同系列在一个分类下的值。

实现过程—增加系列


D:\dfx\Bldoc\demos\示例3.rpx

	A	B	C	D
1(头)	省份	新增病例	新增疑似	治愈出院
2	=ds1.select(省份)	=ds1.新增病例	=ds1.新增疑似	=ds1.治愈出院

1、在上一步rpx的基础上，增加C/D两列，加上新增疑似、治愈出院两个新系列

统计图属性定义

图类型及数据 图形特性 显示外观 其他特性

数据集 统计图类型  柱形图 仅用前 ☐ 条数据记录来画统计图

分类轴

分类轴定义

=A2

添加 删除

系列

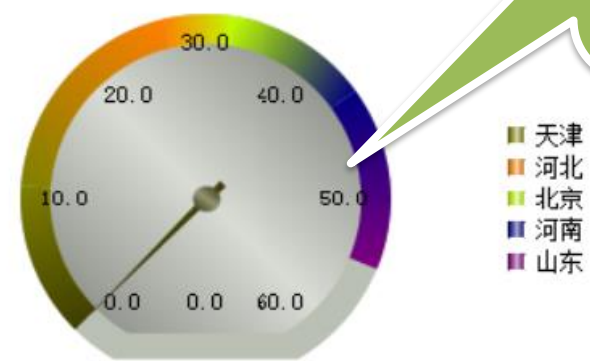
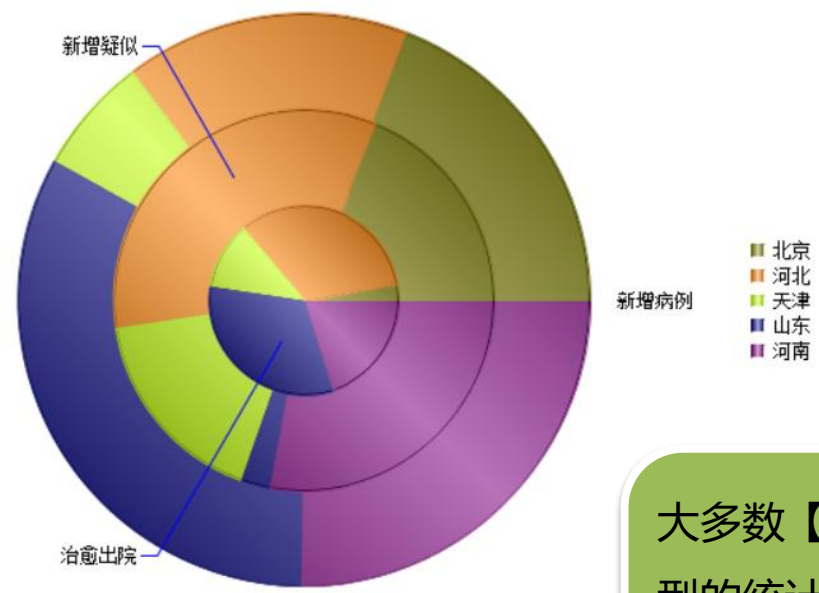
☐ 使用指定轴

系列名称	系列值	标题
=“新增病例”	=B2	
=“新增疑似”	=C2	
=“治愈出院”	=D2	

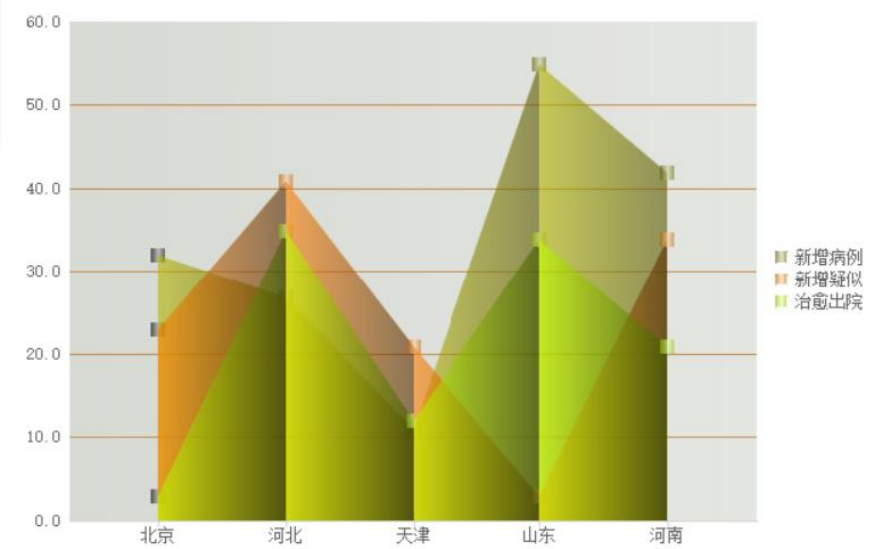
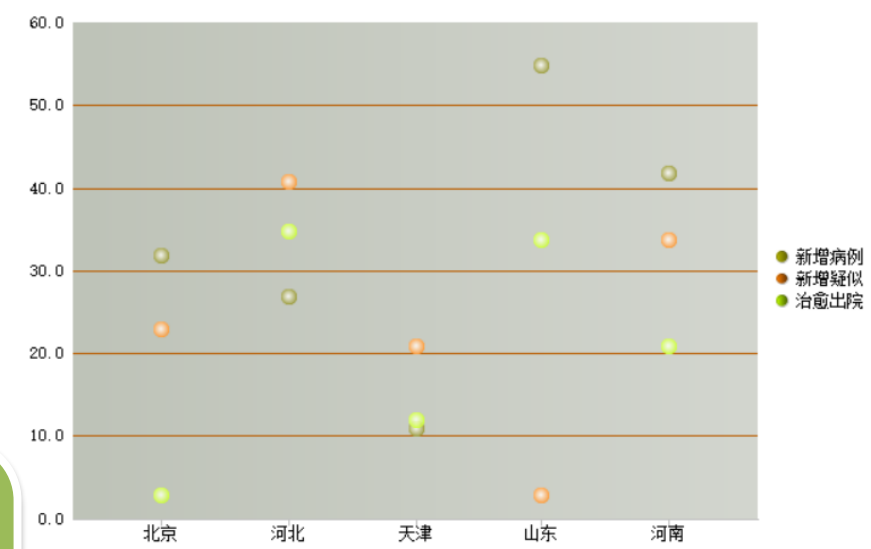
添加 删除

2、选择柱形图；添加分类轴，数据设置A2；添加系列，数据来自B2，系列名“新增病例”

动手试一试—多系列在各种图中都怎么显示



大多数【分类-系列】类型的统计图都支持多系列，也有例外，如仪表盘只能显示一个系列



观察多系列数据1

分类=维度；系列=指标/测度

R 内建数据集

字段	数据		
省份	新增病例	新增疑似	治愈出院
北京	32	23	3
河北	27	41	35
天津	11	21	12
山东	55	3	34
河南	42	34	21

理解方式
其它三种

理解方式1：省份是维度，
其它三种人数为三种指标

两种数据组织方式，展现在报表中时，就是交叉报表行列互换；
在多维分析里，称为旋转；下页用这两种数据制作多系列散点
图做下对比。在理解了数据的情况下，无论是报表、多维分析、
还是统计图场景下，都容易看懂数据。

R 内建数据集

字段	数据				
病例类型	北京	河北	天津	山东	河南
新增病例	32	27	11	55	42
新增疑似	23	41	21	3	34
治愈出院	3	35	12	34	21

理解方式2：病例类型是维
度，其它五个省份是指标

理解方式3：省份、病例类型两
个维度，人数是指标。可以把这
种方式理解成基础数据，然后变
形成前两种容易观察数据的方式。

R 内建数据集

字段	数据	
省份	病例类型	人数
北京	新增病例	32
北京	新增疑似	23
北京	治愈出院	3
河北	新增病例	27
河北	新增疑似	41
河北	治愈出院	35
天津	新增病例	11
天津	新增疑似	21
天津	治愈出院	12
山东	新增病例	55
山东	新增疑似	3
山东	治愈出院	34
河南	新增病例	42
河南	新增疑似	34
河南	治愈出院	21

观察多系列数据2

用上页方式1的数据做成多系列的散点统计图



用上页方式2数据做成多系列散点统计图



两种数据组织方式，做出两种多系列的散点图，都容易从省份、病例类型两个维度观察，这和原始数据里有两个维度也是契合的。

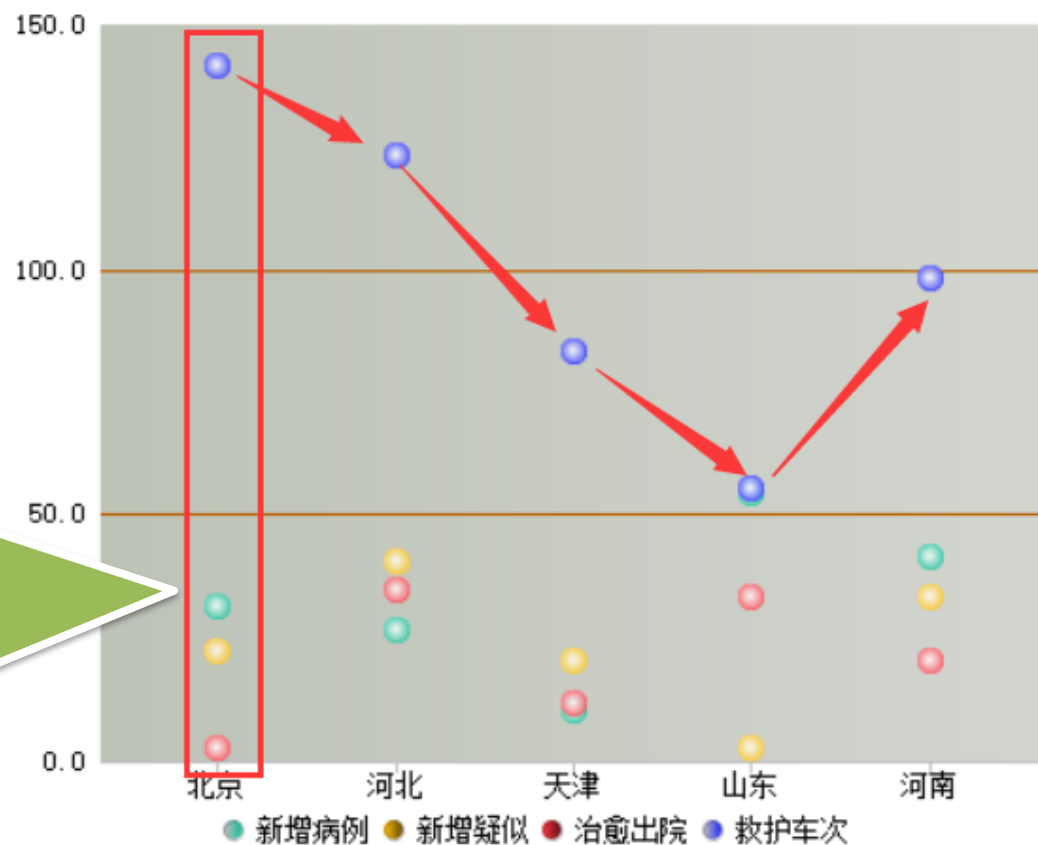
观察多系列数据3

内建数据集

省份	新增病例	新增疑似	治愈出院	救护车次
北京	32	23	3	142
河北	27	41	35	124
天津	11	21	12	84
山东	55	3	34	56
河南	42	34	21	99

在之前数据的基础上，再增加一列“救护车次”。
这个指标是独立的，和人数没关系。

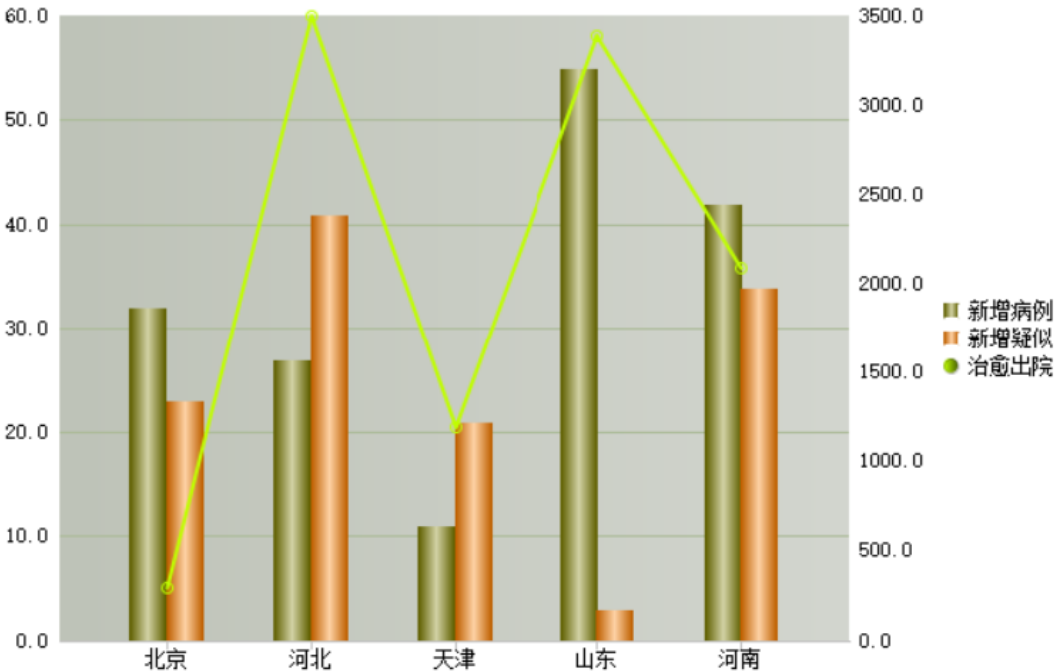
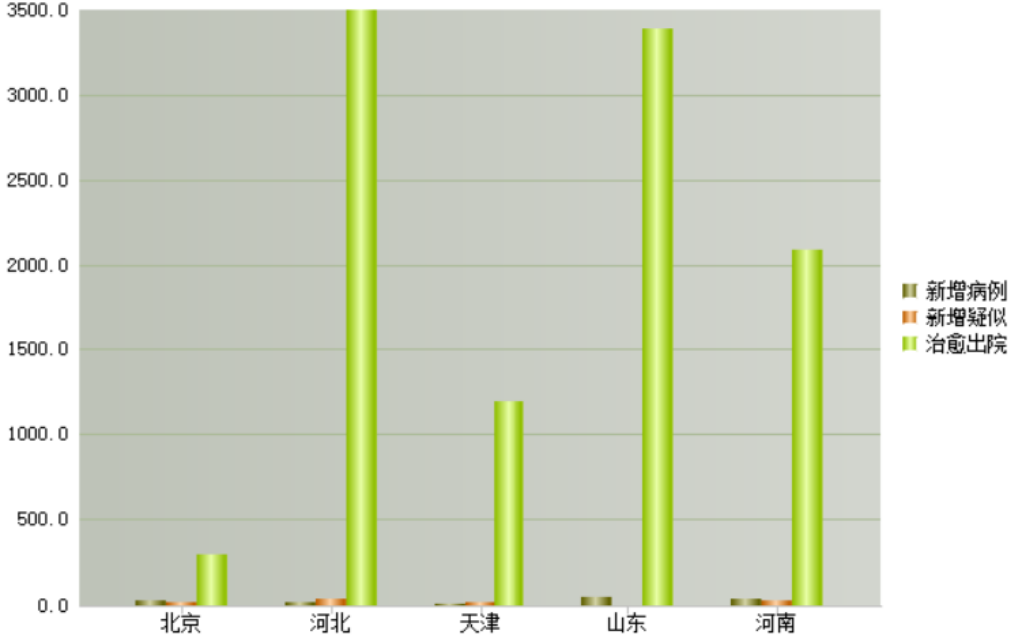
- 1、因为要共用省份维度，这时就只能用省份做分类，如果用病例类型做分类，那就没法展现救护车次这个系列了。
- 2、虽然救护车次和另外三个指标不是一类，但同一分类下的纵向对比，仍然有业务意义，比如了解“救护车次是否和新增疑似的人数增幅相关”。



多个系列时，数值单位差距大时，用双轴统计图

内建数据集

省份	新增病例	新增疑似	治愈出院
北京	32	23	300
河北	27	41	3500
天津	11	21	1200
山东	55	3	3400
河南	42	34	2100



把【治愈出院】系列的数值扩大100倍，再预览柱形图，发现另外两个系列矮到无法辨识了，这是因为纵轴的刻度被撑大了。这时用双轴类的统计图，用增加的右轴显示不同刻度。

实现过程—双轴图

统计图属性定义

图类型及数据

图形特性

显示外观

其他特性

数据集 统计图类型

双轴柱线图

 仅用前 ☐ 条数据记录来画统计图

分类轴

分类轴定义

=A2

添加

删除

系列

☒ 使用指定轴

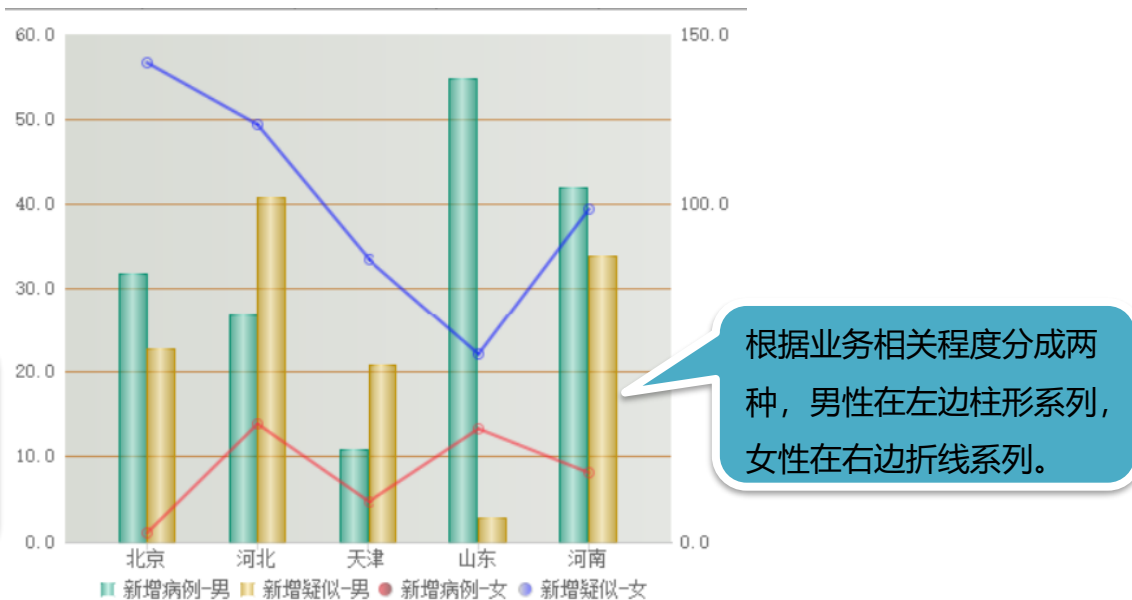
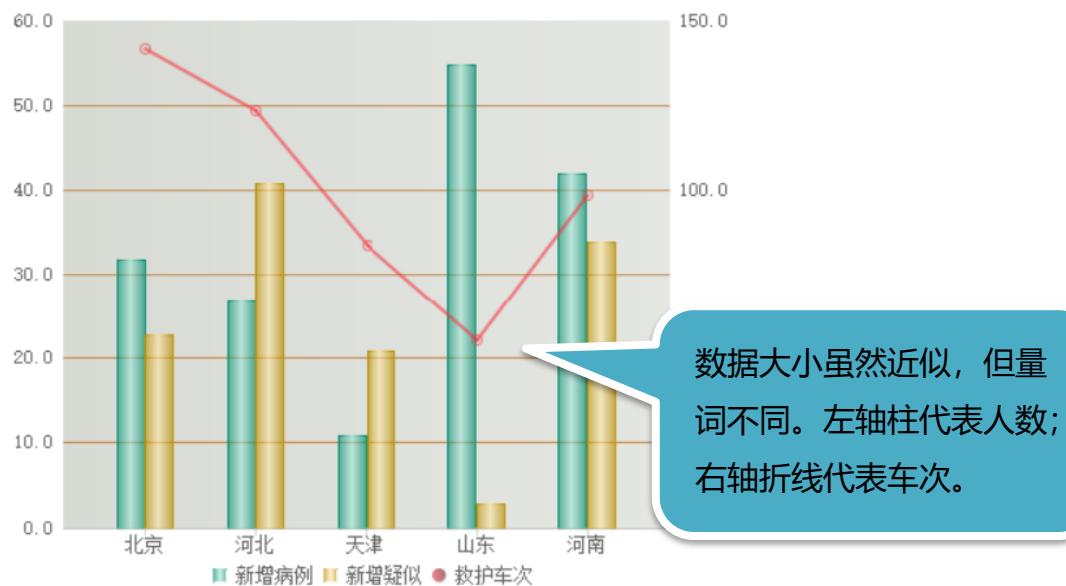
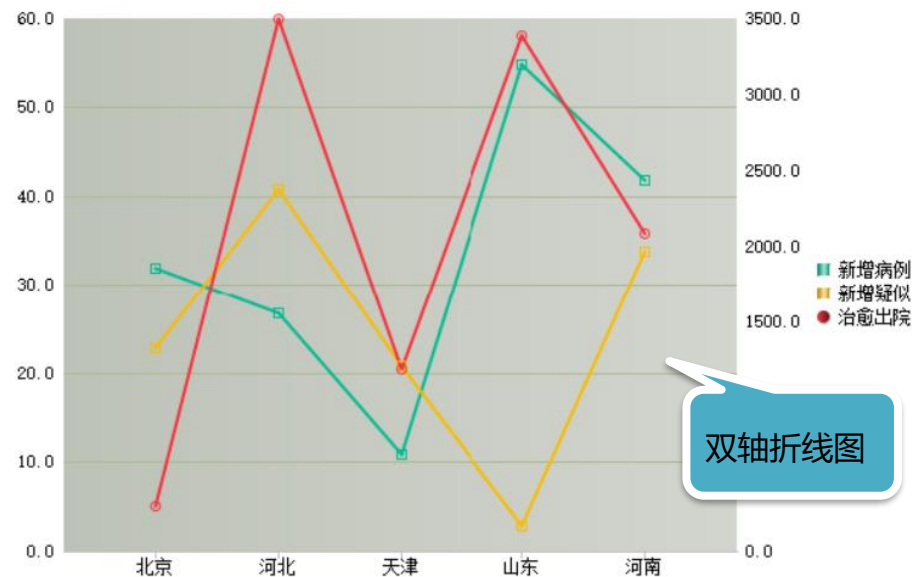
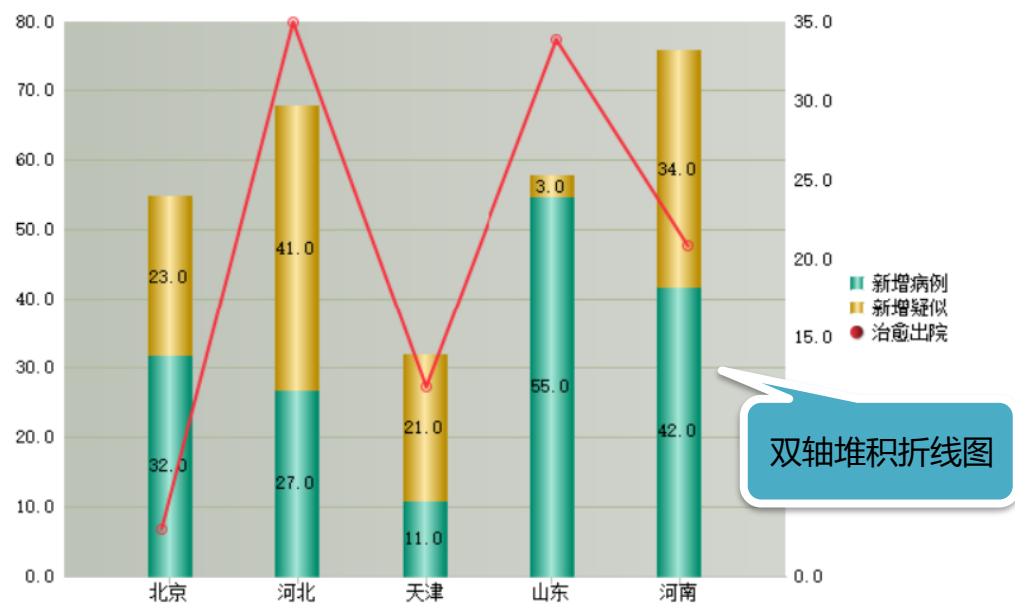
系列名称	系列值	标题	轴
=“新增病例”	=B2		左轴
=“新增疑似”	=C2		左轴
=“治愈出院”	=D2		右轴

添加

删除

选择双轴类的统计图，然后把数据单位大的【治愈出院】设置为右轴（也可以去选“使用指定轴”，让程序根据系列数据自动分配左右轴）

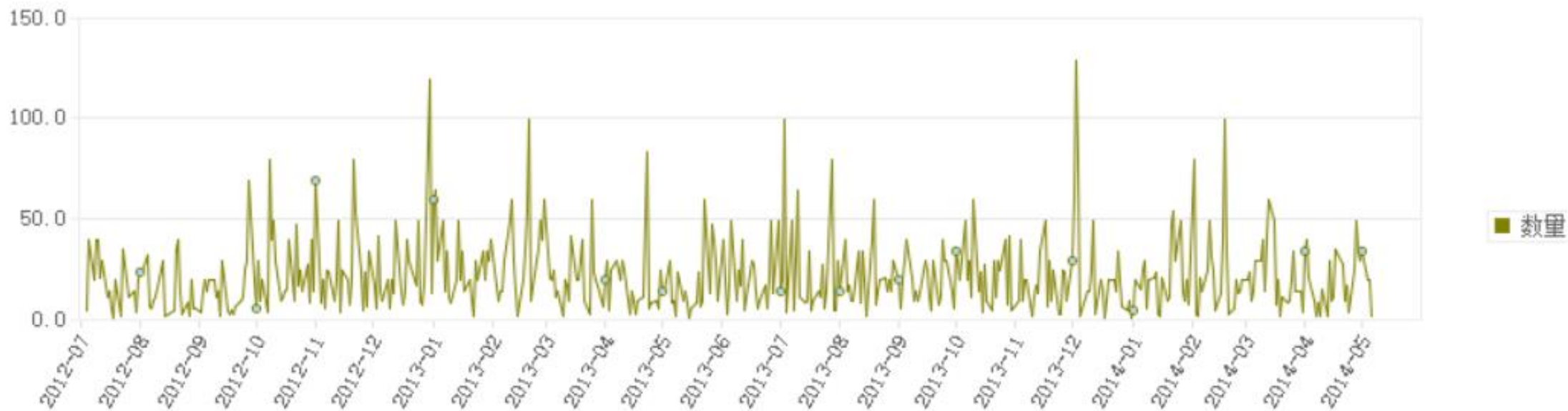
动手试一试：还有哪些双轴图/还有什么情况适用双轴图？



分类轴为指定间隔的时间—时间走势图

在一些业务需求中，经常会遇到随时间变化而变化的大量数据。由于数据量非常大，画数据随时间变化的统计图时，如果每条数据都描点，会使图表非常难以辨认。在润乾报表中，专门针对这种需求提供了时间走势图。

时间走势图



实现过程—时间走势图

D:\dfx\BIdoc\demos\示例5.rpx

	A	B
1	订购日期	数量
2	=ds1.select(订购日期)	=ds1.数量

统计图属性定义

图类型及数据 图形特性 显示外观 其他特性

数据集 统计图类型 时间走势图

分类轴

分类轴定义

=A2

系列

☐ 使用指定轴

系列名称	系列值
=数量	=B2

2、把时间字段设置为分类轴。把数量设置为系列。注意这时分类轴还不明确，成千上万个时间值，没法在图上用点清晰的区分开。

	A	B	C
1	订购日期	数量	
2	2012-07-04	12	
3	2012-07-04	10	
4	2012-07-04	5	
5	2012-07-05	9	
6	2012-07-05	40	
7	2012-07-08	10	
8	2012-07-08	35	
9	2012-07-08	15	

1、三年里有两千多条订购记录，存储在2-2.xlsx里。参照上面的方法制作报表。

3、在其他特性里，通过设置时间分类轴的计算方式，得到最终想看到的效果。

统计图属性定义

图类型及数据 图形特性 显示外观 其他特性

时序状态图或甘特图

状态条高度 时间刻度选择 时

时间走势图横轴坐标取值

起始值	结束值	取样间隔	取样单位	标签间隔	标签单位
2012-07-01	2014-05-31	1	日	1	月

添加 删除

只支持单系列/有特殊值—仪表盘

	A	B
1	分类	温度
2	value	13
3	炎热	40
4	舒适	25
5	凉爽	15
6	寒冷	-3
7	严寒	-15



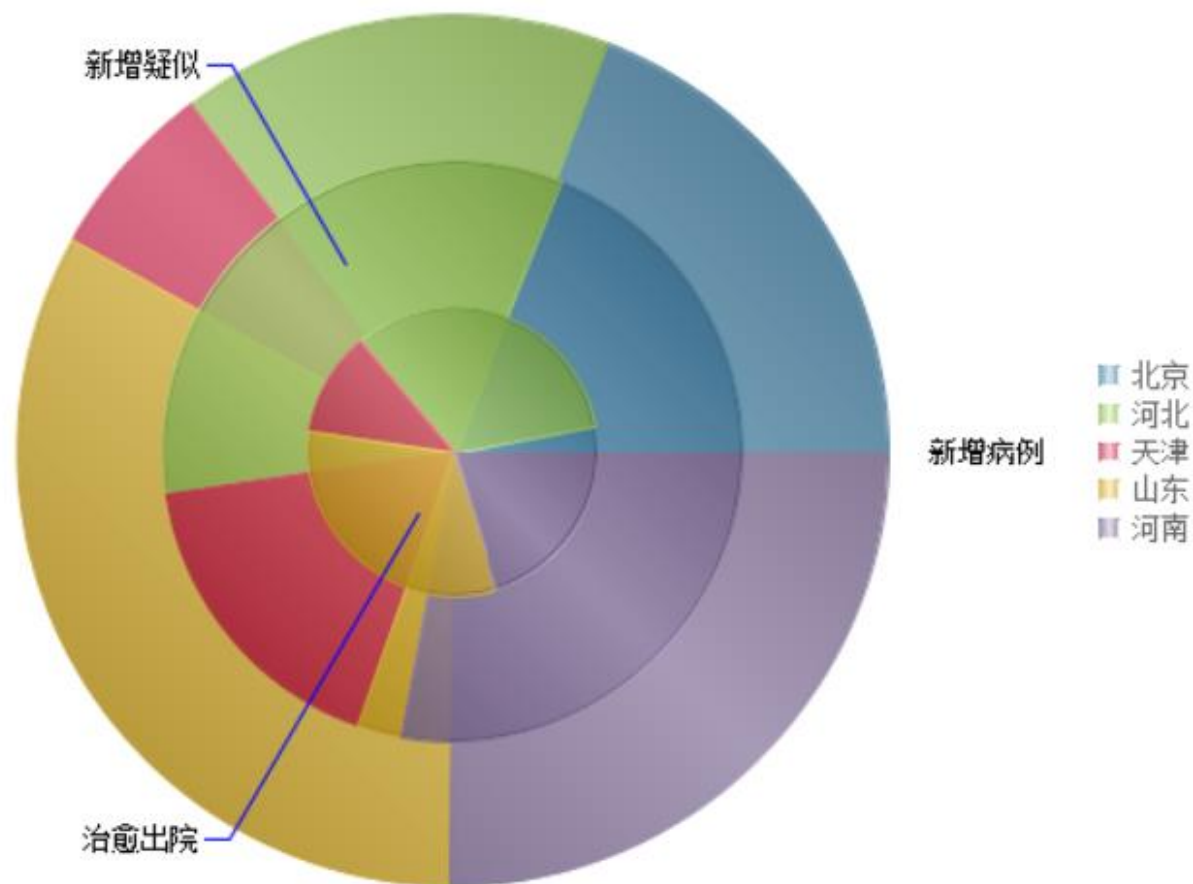
■ 严寒 ■ 寒冷 ■ 凉爽 ■ 舒适 ■ 炎热

观察仪表盘的数据和图，有三个特殊：

- 1、每个分类对应的系列值是一个起始值，实际上分类的含义是一个区间，区间的结束值就下一分类的起始值。
- 2、需要有一个特殊值来定义指针，润乾报表针对这个的设计是：分类字段里始终用“value”为指针值。
- 3、仪表盘不支持多系列。

制作过程没什么特殊的，参考前面步骤设置下分类轴、系列就可以了，示例数据在2-3.xlsx里。

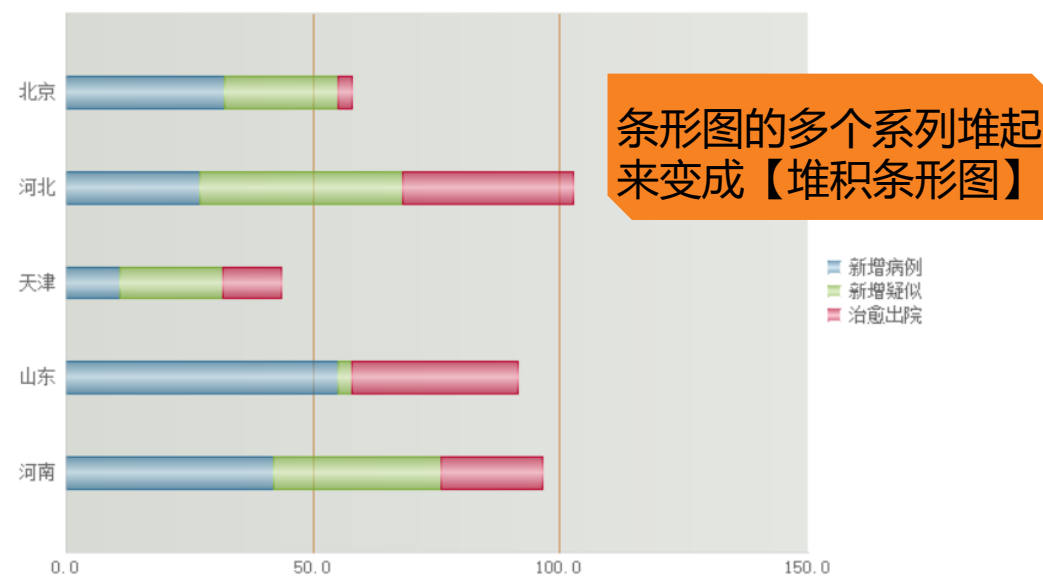
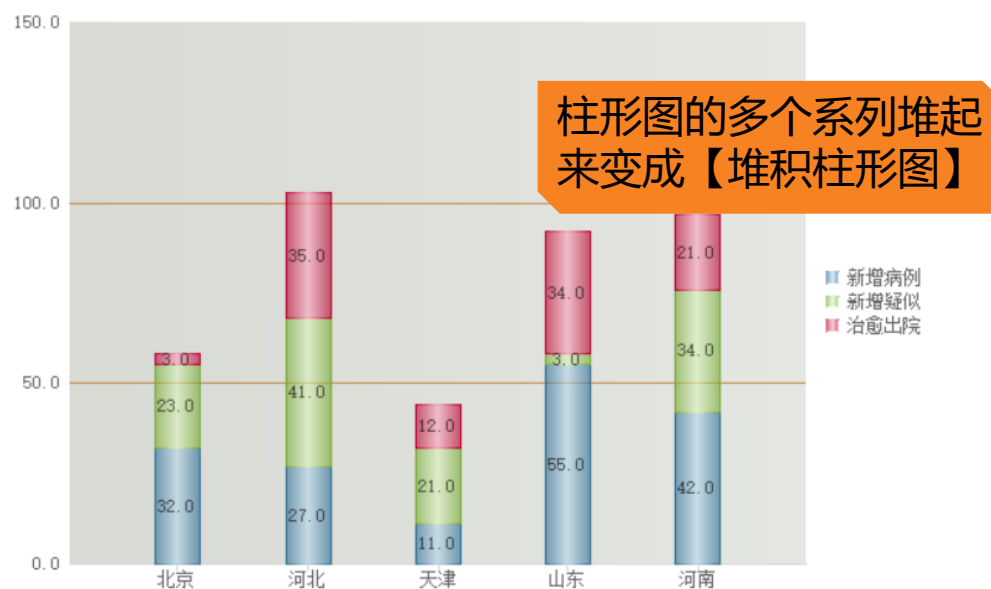
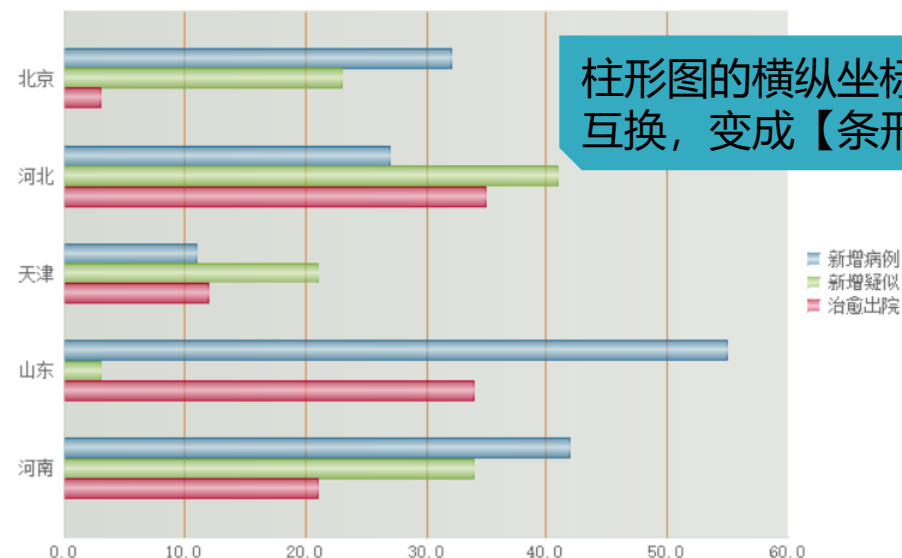
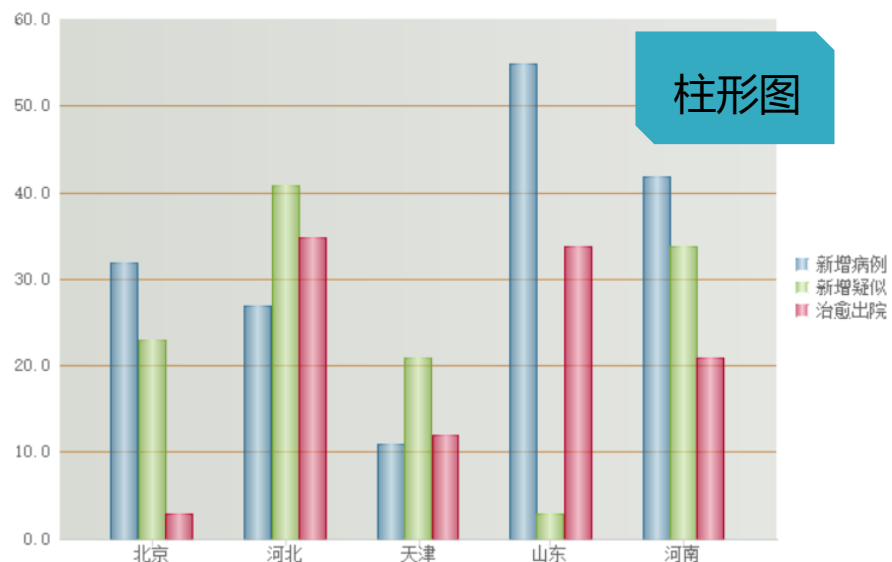
无法表达负数—饼图



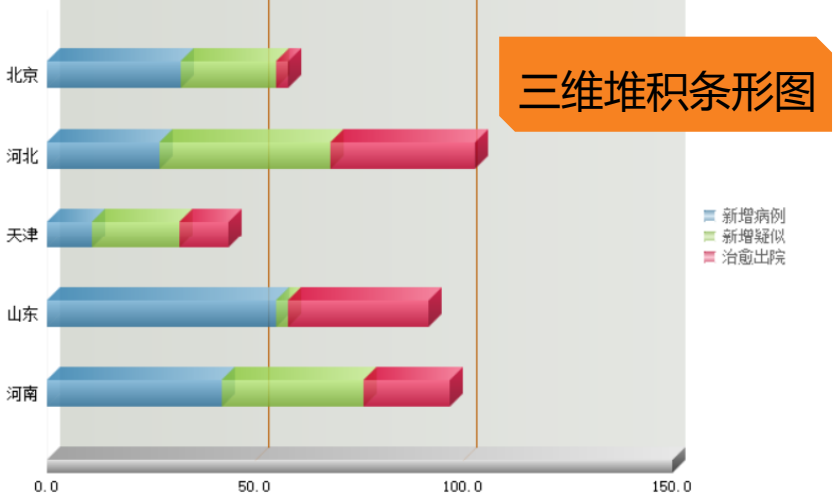
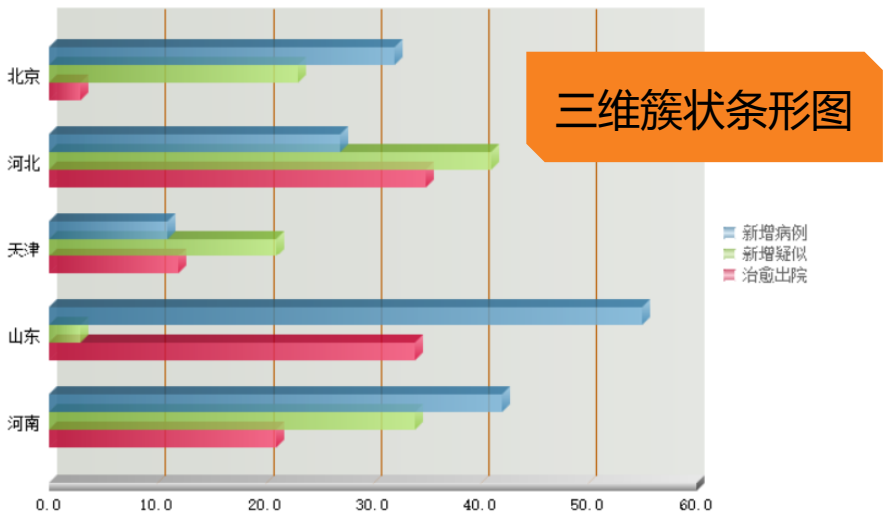
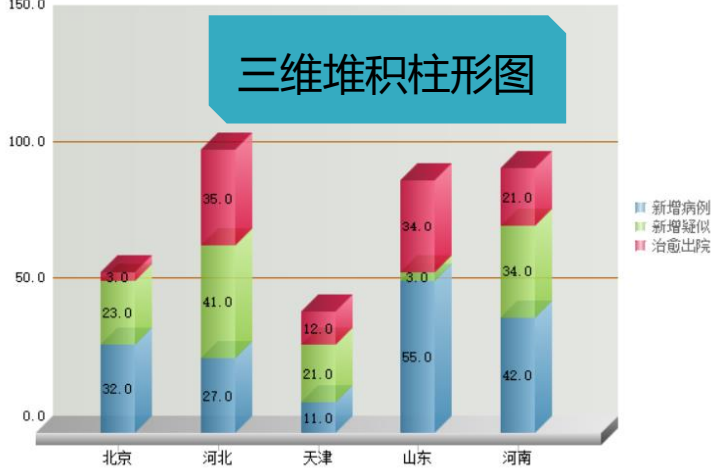
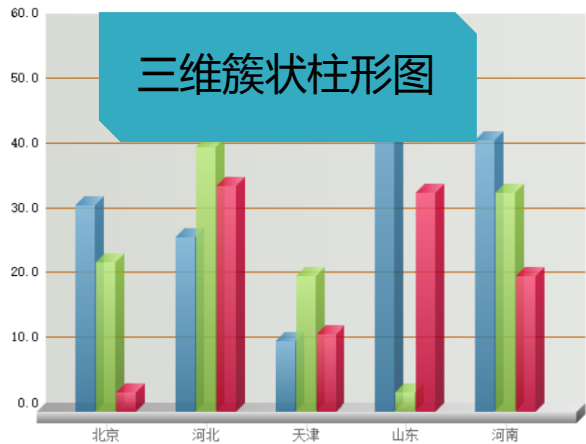
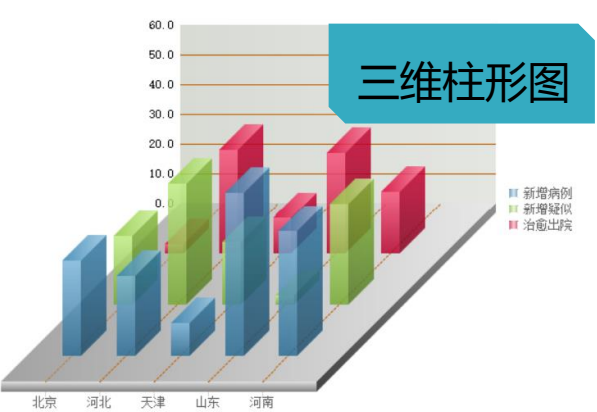
直角坐标系的统计图很多，因为纵轴能表示负数，所以它们也能表示负数。

饼图是用扇形面积表示数值大小，表示的数值必须大于0。

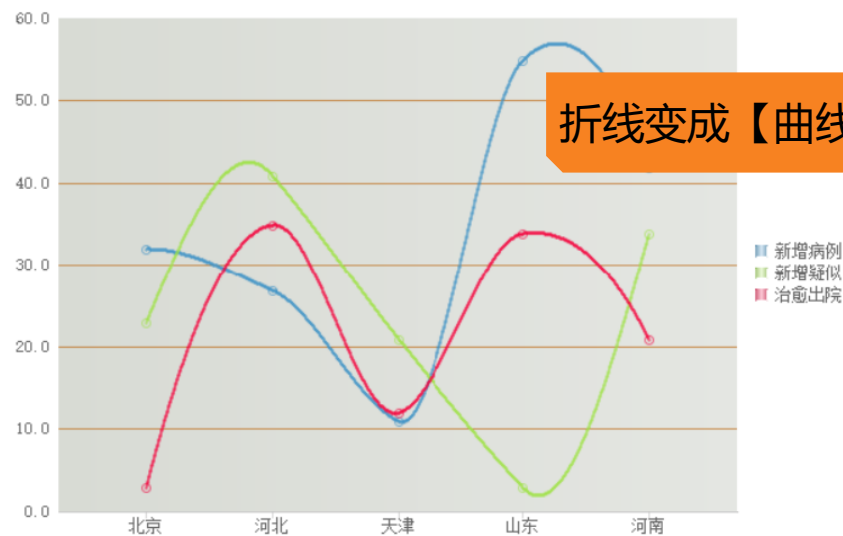
柱形图—平面变种



柱形图—立体变种



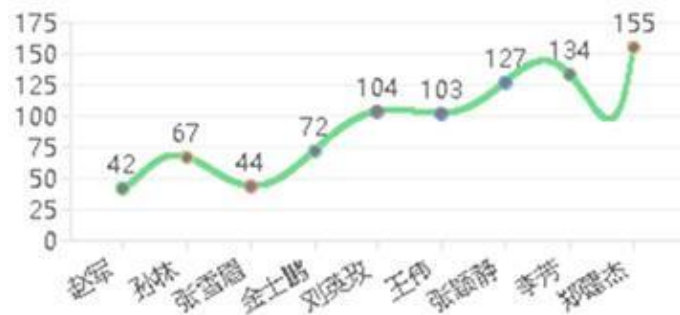
折线图—平面变种



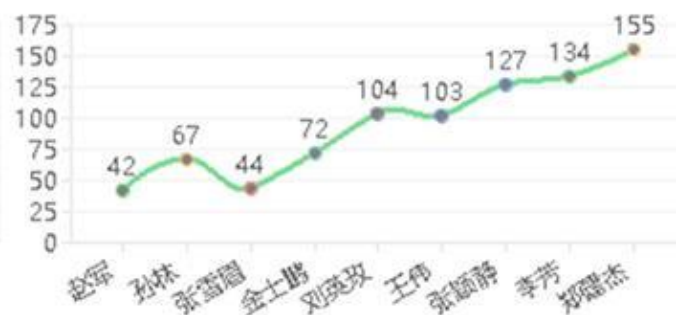
时间走势图



曲线图可以选择曲线类型

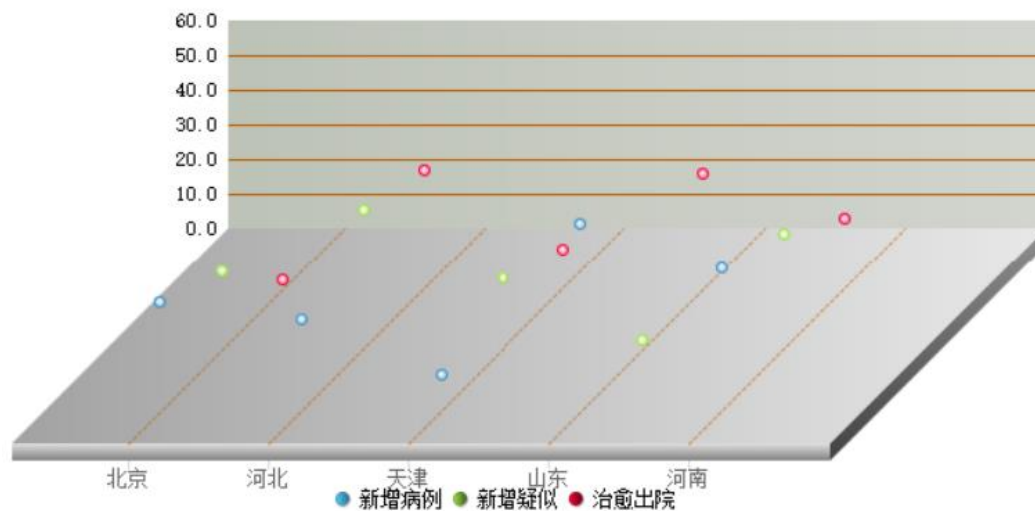
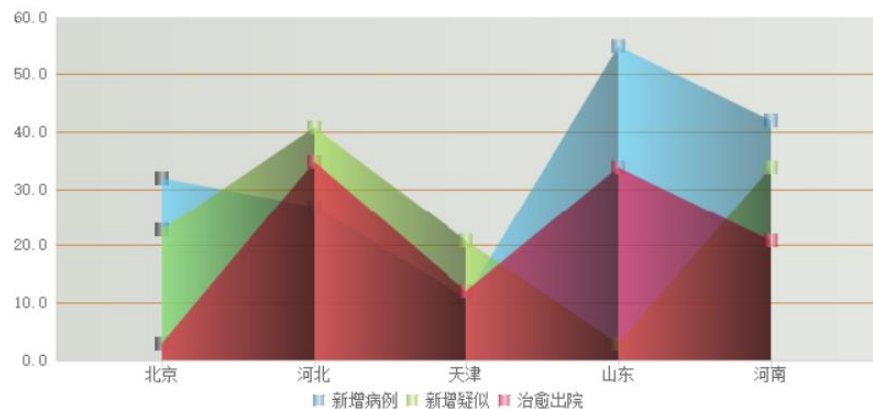
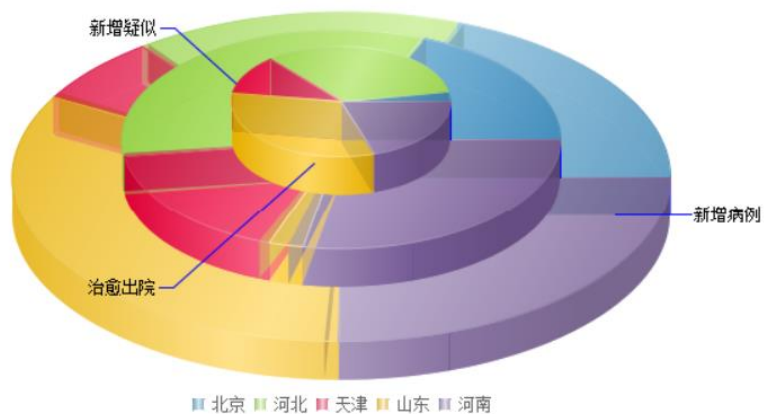
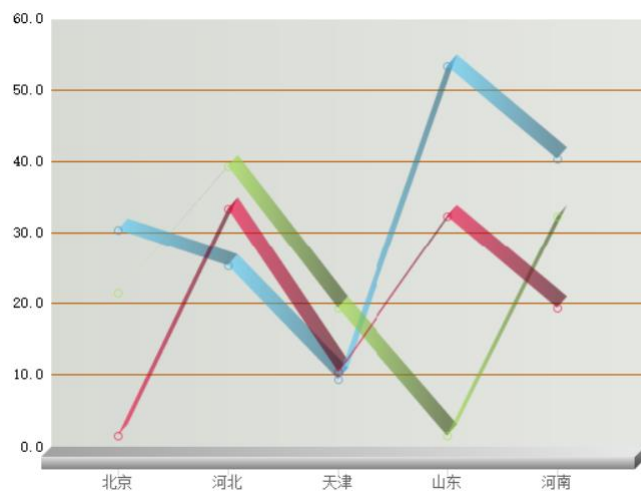


阿格朗日曲线



阿克玛曲线

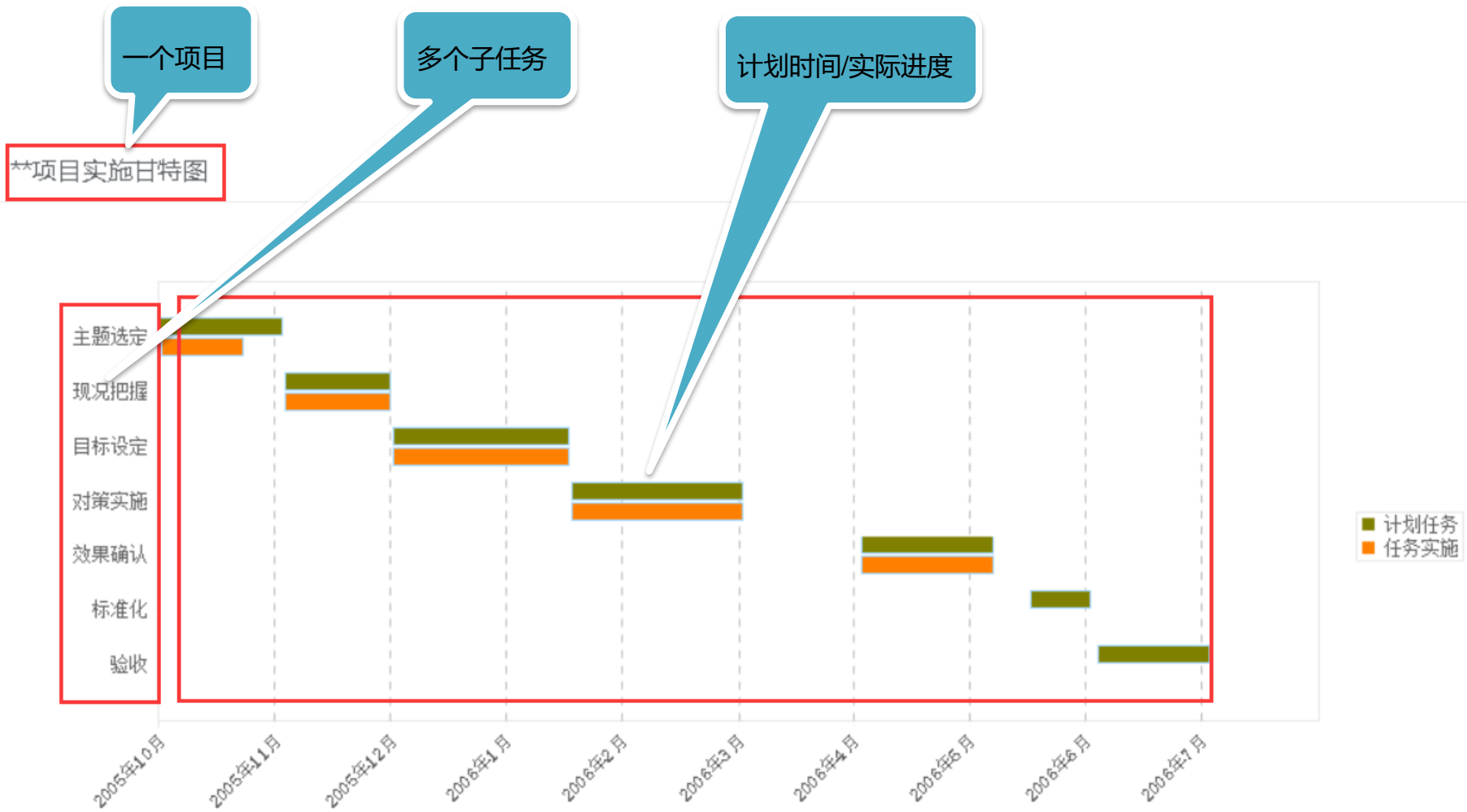
折线图、饼图、区域图、散点图、仪表盘—立体变种





非【分类-系列】统计图

著名的表现项目进度的甘特图



实现过程—甘特图

	A	B	C	D
1	channel	startt	endtt	state
2	现况把握	2005-11-03	2005-11-30	任务实施
3	主题选定	2005-10-01	2005-11-02	计划任务
4	效果确认	2006-04-01	2006-05-05	任务实施
5	标准化	2006-05-15	2006-05-30	计划任务
6	主题选定	2005-10-02	2005-10-23	任务实施
7	目标设定	2005-12-01	2006-01-15	任务实施
8	对策实施	2006-01-16	2006-03-01	计划任务
9	效果确认	2006-04-01	2006-05-05	计划任务
10	验收	2006-06-01	2006-06-30	计划任务
11	对策实施	2006-01-16	2006-03-01	任务实施
12	现况把握	2006-11-03	2006-11-30	计划任务
13	目标设定	2005-12-01	2006-01-15	计划任务

原始数据2-4.xlsx里每条数据
记录一个子任务的一种状态
【计划或实施进度】起止时间。

D:\dfx\Bldocldemos\示例7.rpx

	A	B	C	
1	任务项目	起始时间	结束时间	任务状态
2	=ds1.select(channel)	=ds1.startt	=ds1.endtt	=ds1.state
3				
4				
5				
6				

统计图属性定义

图类型及数据 图形特性 显示外观 其他特性

数据集 统计图类型 **甘特图** 仅用前 ☐ 条数据记录来画统计图

分类轴

分类轴定义

系列

☐ 使用指定轴

系列名称 系列值 添加

时序状态图、甘特图及里程碑图数据

起始时间 =B2 项目值 =A2

结束时间 =C2 状态值 =D2

把相应的字段设置给统计图就可以了。

动手试一试—选择合适的时间刻度

统计图属性定义

图类型及数据

图形特性

显示外观

其他特性

时序状态图或甘特图

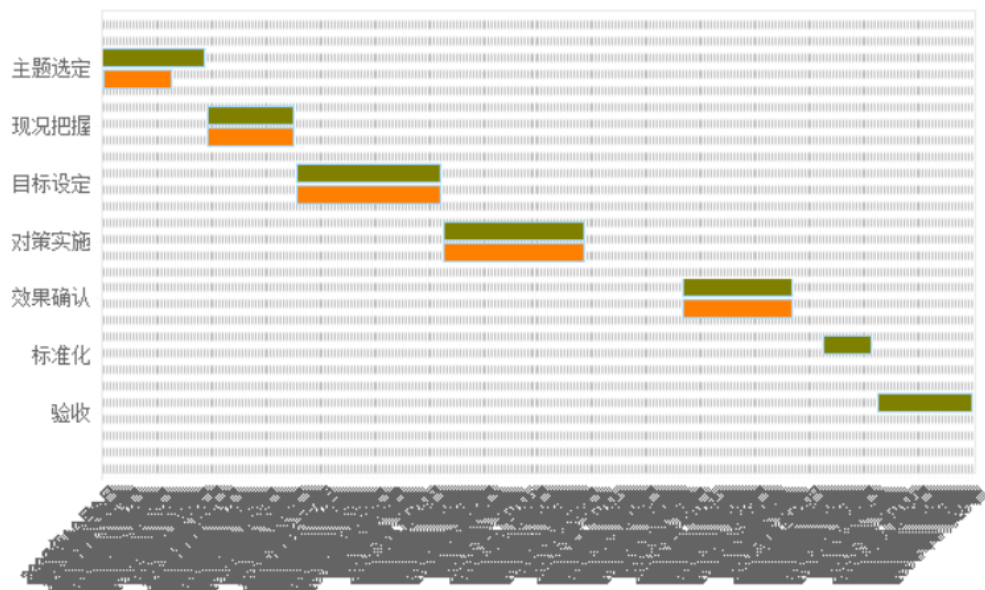
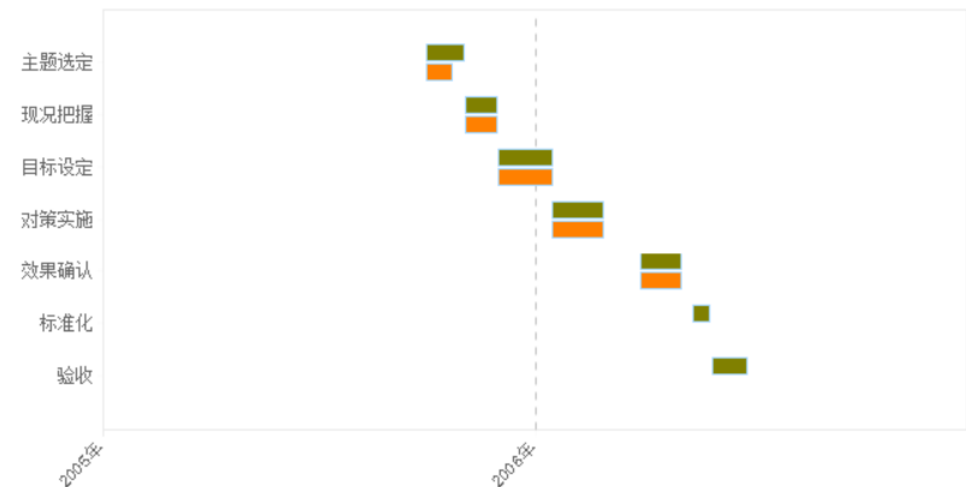
状态条高度

25

时间刻度选择

年

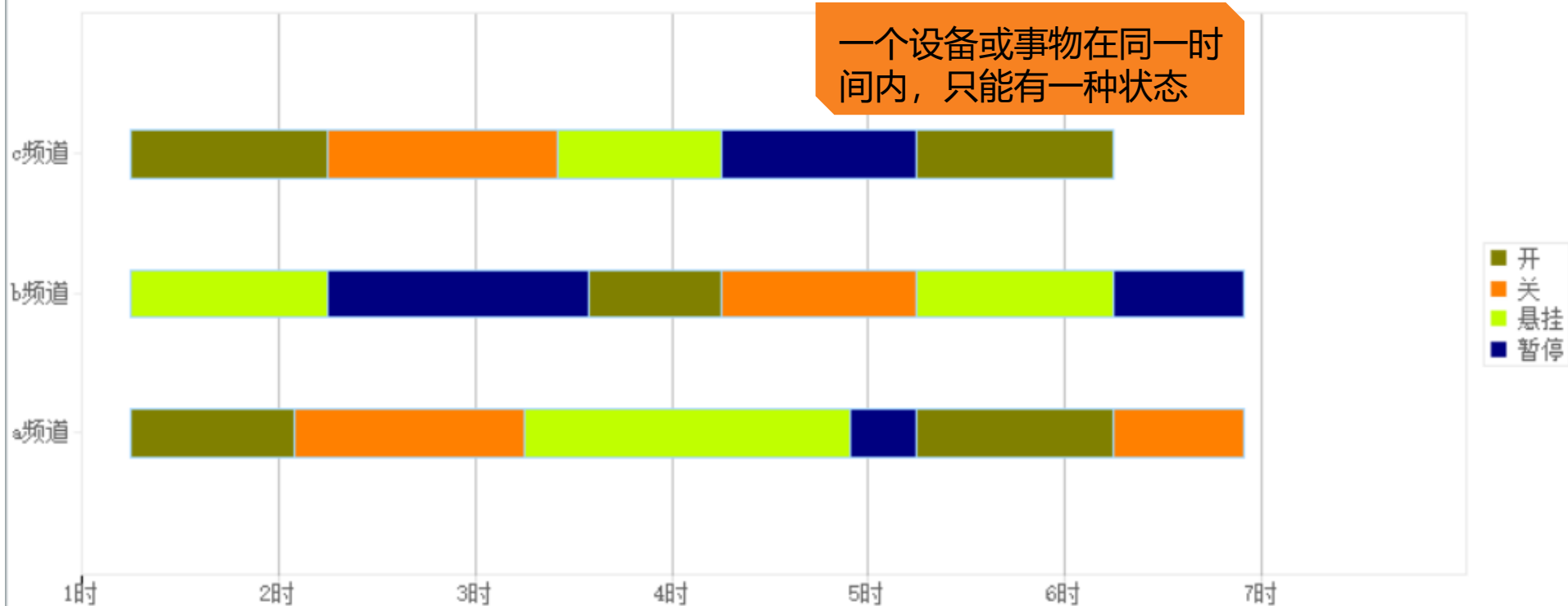
根据子任务实际时间的长短，
选择合适的时间刻度。选不好
就会出现下面的效果。



时序状态图

这种图适合一些多状态的
设备或事物，想要观察它
们在一段时间内的状态

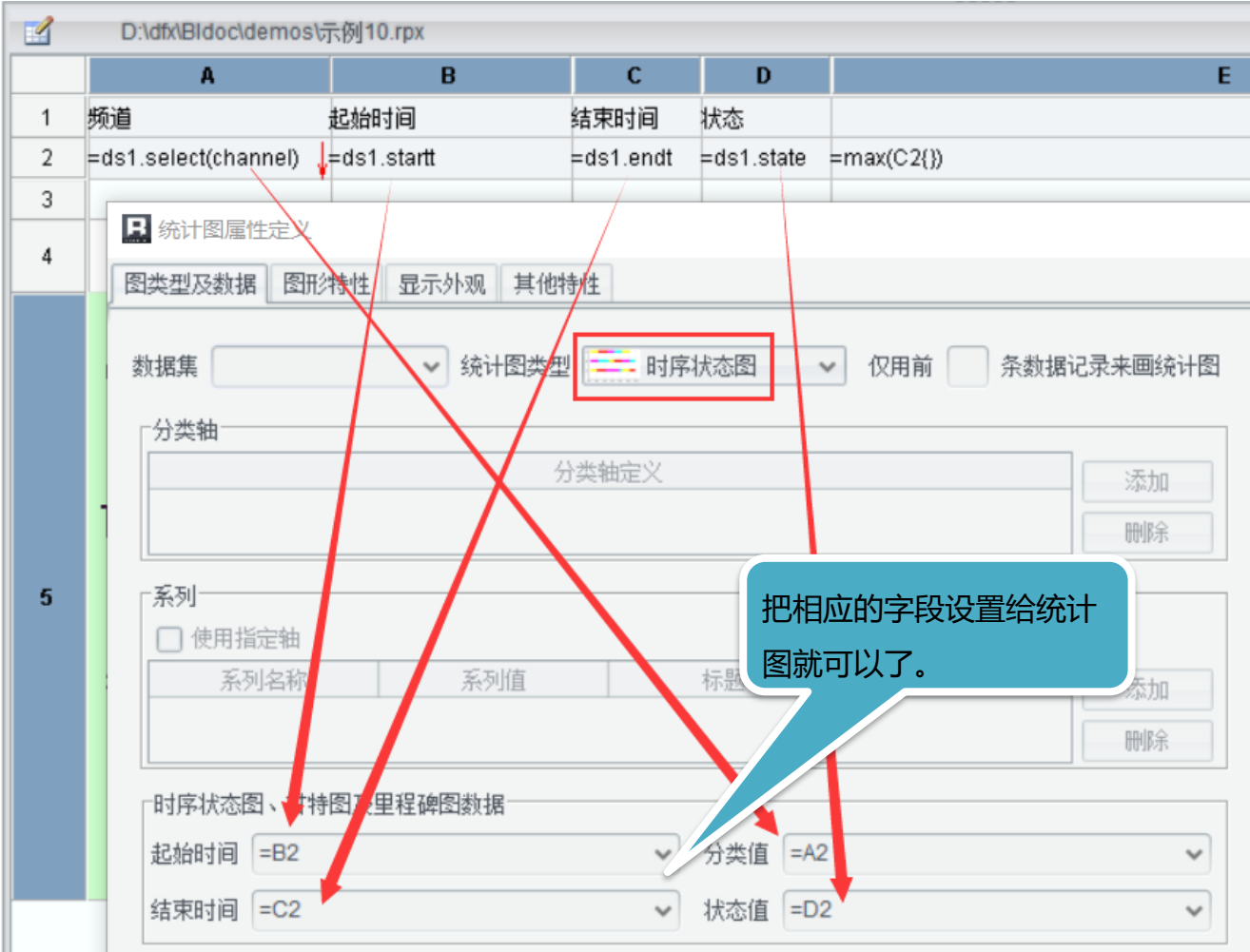
时序状态图



实现过程—时序状态图

	A	B	C	D
1	channel	startt	endt	state
2	a频道	2002-08-08 01:15	2002-08-08 02:05	开
3	a频道	2002-08-08 02:05	2002-08-08 03:15	关
4	a频道	2002-08-08 03:15	2002-08-08 04:55	悬停
5	a频道	2002-08-08 04:55	2002-08-08 05:15	暂停
6	a频道	2002-08-08 05:15	2002-08-08 06:15	开
7	a频道	2002-08-08 06:15	2002-08-08 06:55	关
8	b频道	2002-08-08 01:15	2002-08-08 02:15	悬停
9	b频道	2002-08-08 02:15	2002-08-08 03:35	暂停
10	b频道	2002-08-08 03:35	2002-08-08 04:15	开
11	b频道	2002-08-08 04:15	2002-08-08 05:15	关
12	b频道	2002-08-08 05:15	2002-08-08 06:15	悬停
13	b频道	2002-08-08 06:15	2002-08-08 06:55	暂停
14	c频道	2002-08-08 01:15	2002-08-08 02:15	开
15	c频道	2002-08-08 02:15	2002-08-08 03:15	关
16	c频道	2002-08-08 03:25	2002-08-08 04:15	悬停
17	c频道	2002-08-08 04:15	2002-08-08 05:15	暂停
18	c频道	2002-08-08 05:15	2002-08-08 06:15	开

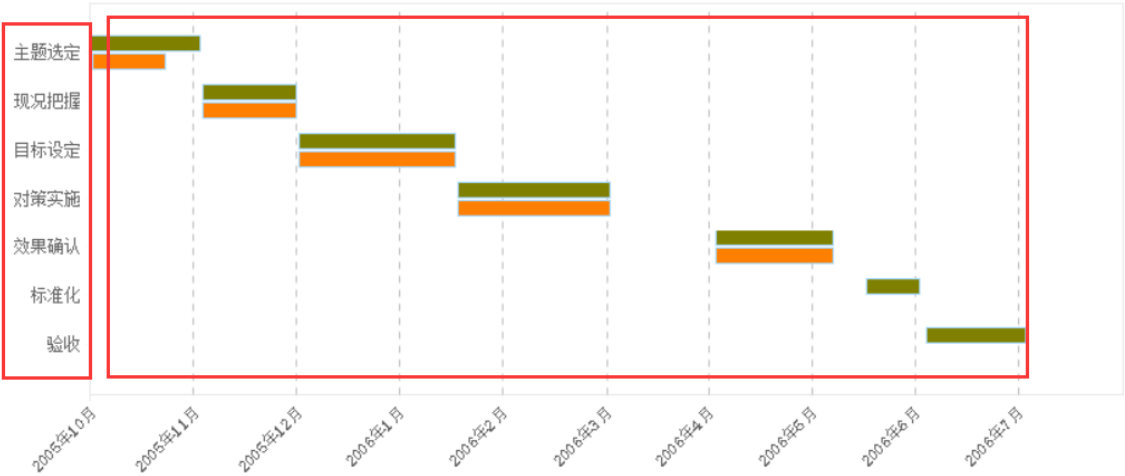
原始数据2-5.xlsx每条数据记录一个设备的一种状态的起止时间。



时序状态图和甘特图，从数据和制作方式上看一模一样，下页详细对比下这两种图

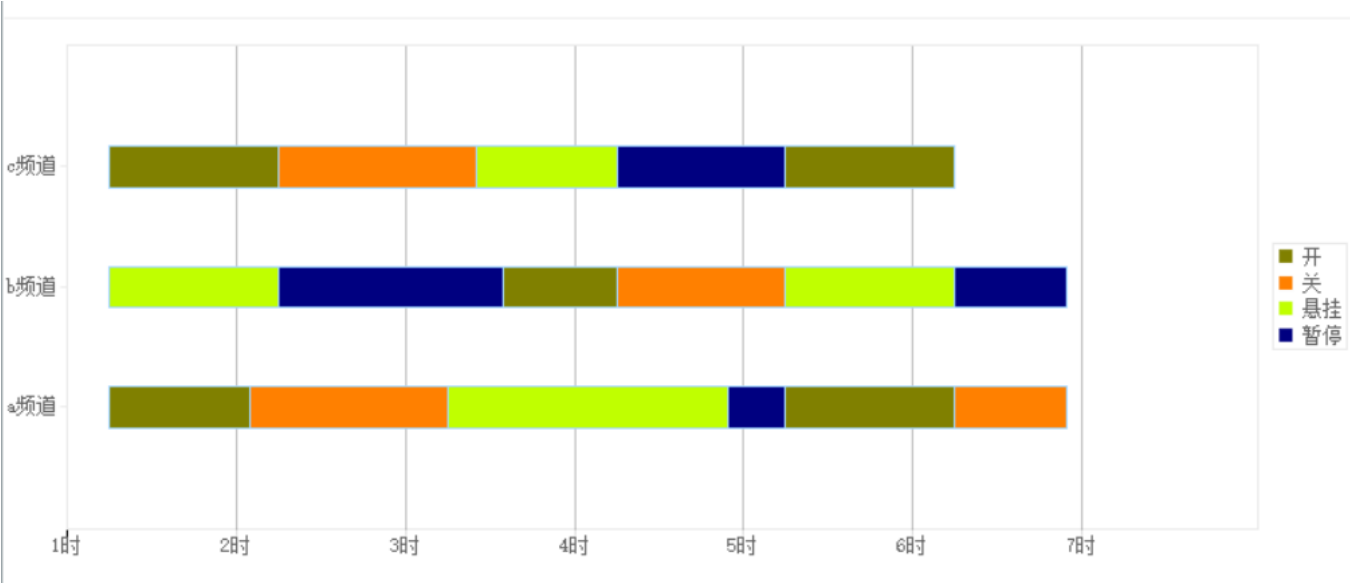
甘特图、时序状态图对比

项目实施甘特图

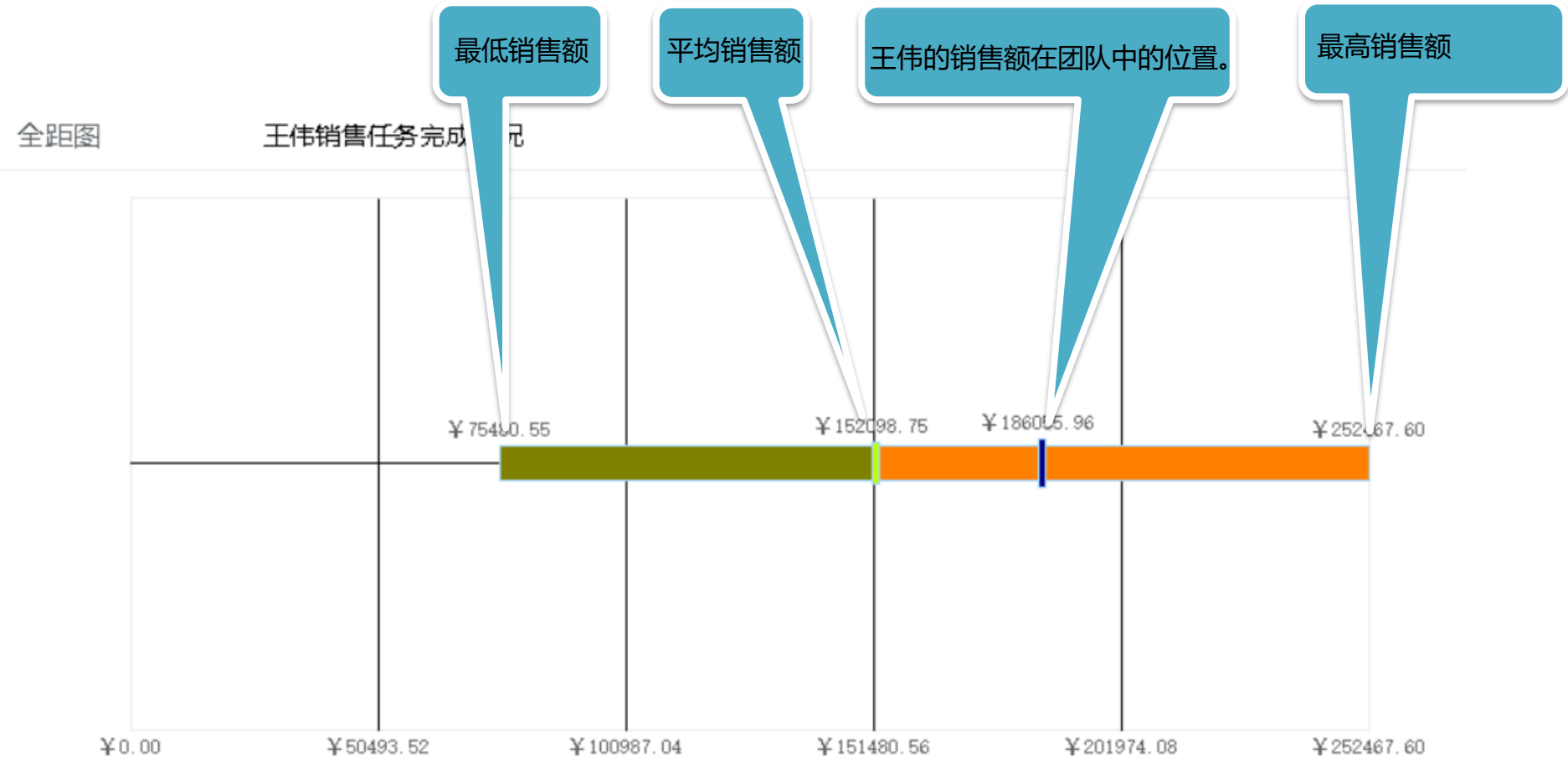


相同：两者都是在观察一些【事物】（子任务、设备）在某些时间的状态；都能在【事物】内部，或多个【事物】间做时间状态对比。比如甘特图里A子任务为什么延期了很久？A子任务结束后，B子任务为什么没有马上开始呢？时序状态图里甲设备最后一次开机是什么时间？甲设备悬停的时候，乙设备为什么没开启？

异同：甘特图里【事物】的状态在时间上有重叠，并且着重对比这些重叠的程度，看看实际时间和计划时间有多大的出入；时序状态图里【事物】的状态在时间上无重叠，一般是连续的，着重对比什么时间进行了状态变换，看看上面时间开机、什么时间又关机了。



对比最大值/最小值/平均值/个体值—全距图



实现过程—全距图

	A	B	
334	张颖静	28.95	
335	张颖静	36	
336	张颖静	31	
337	张颖静	26	
338	王伟	930	
339	王伟	240	
340	王伟	720	
341	王伟	468	
342	王伟	36	
343	王伟	280	

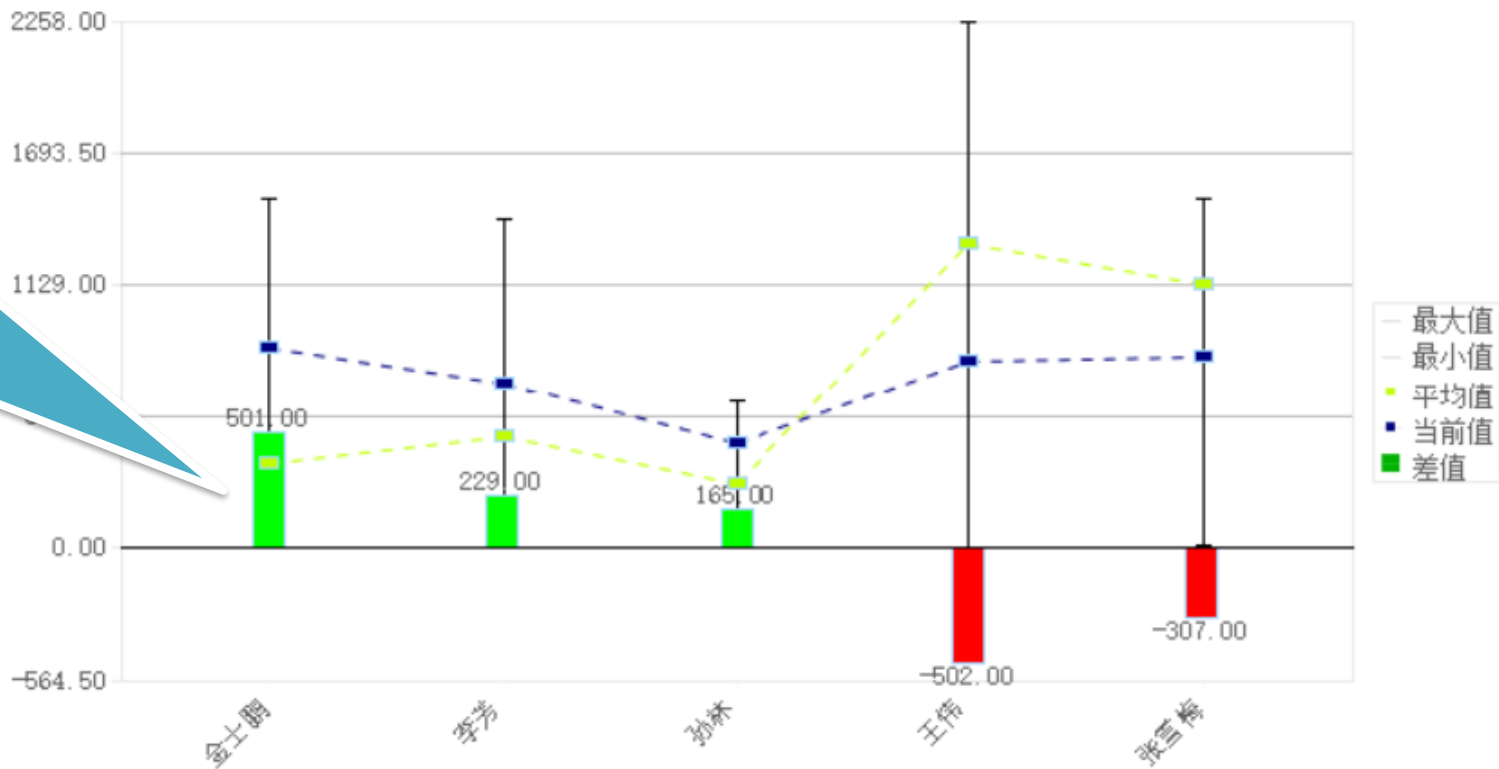
原始数据2-6.xlsx里每个销售有多个订单



对比局部最大值/最小值/平均值/标准值/差值—工字图

工字图

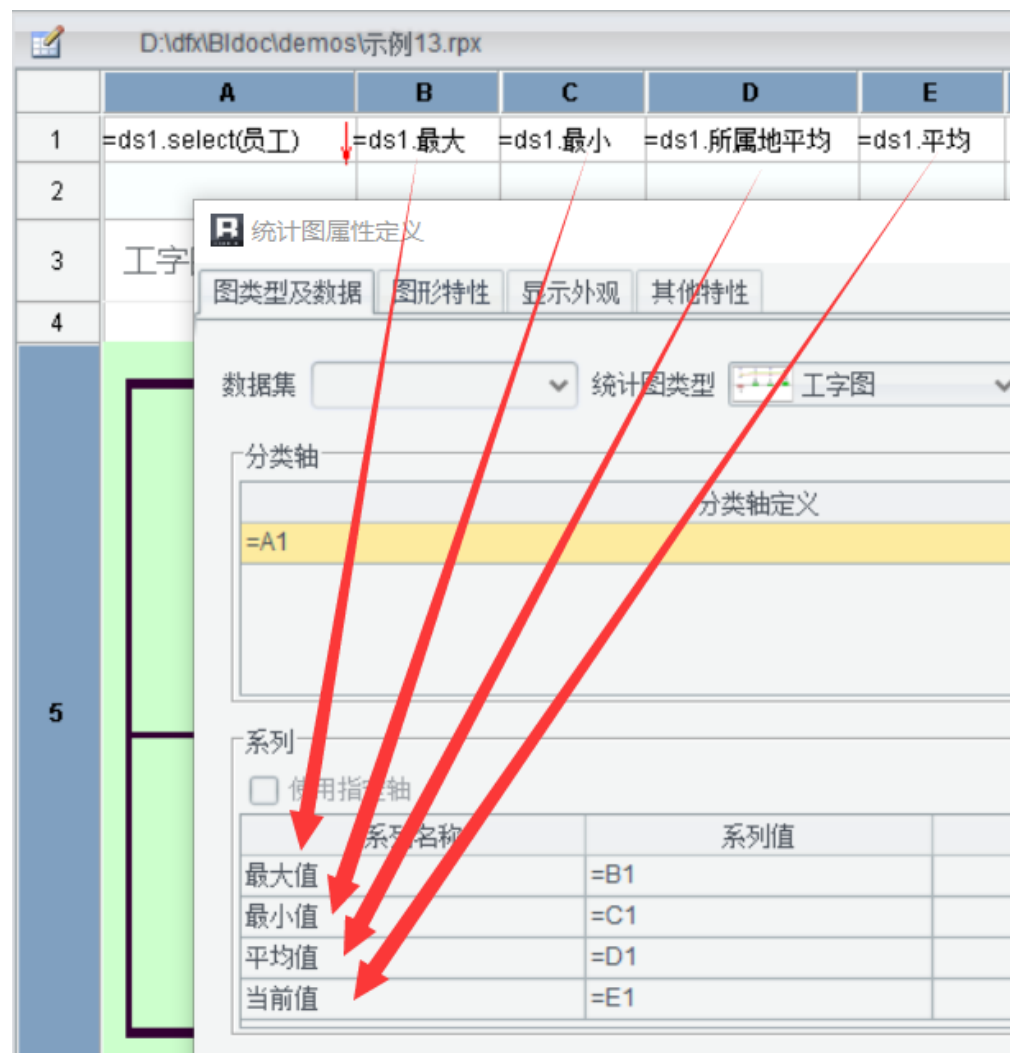
全国各地区每个销售的五个订单金额指标：
最大订单金额、
最小订单金额、
平均订单金额、
销售所在地区的平均订单金额（标准值）、
平均订单金额和标准值的差额（差值）。
绿色差值为正，高于所在地平均水平；
红色差值为负，低于所在地平均水平。



实现过程—工字图

	A	B	C	D	E
1	员工	最大	最小	平均	所属地平均
2	金士鹏	1505	16	865	364
3	李芳	1414	9	711	482
4	孙林	636	20	445	280
5	王伟	2258	4	811	1313
6	张雪梅	1505	12	824	1131

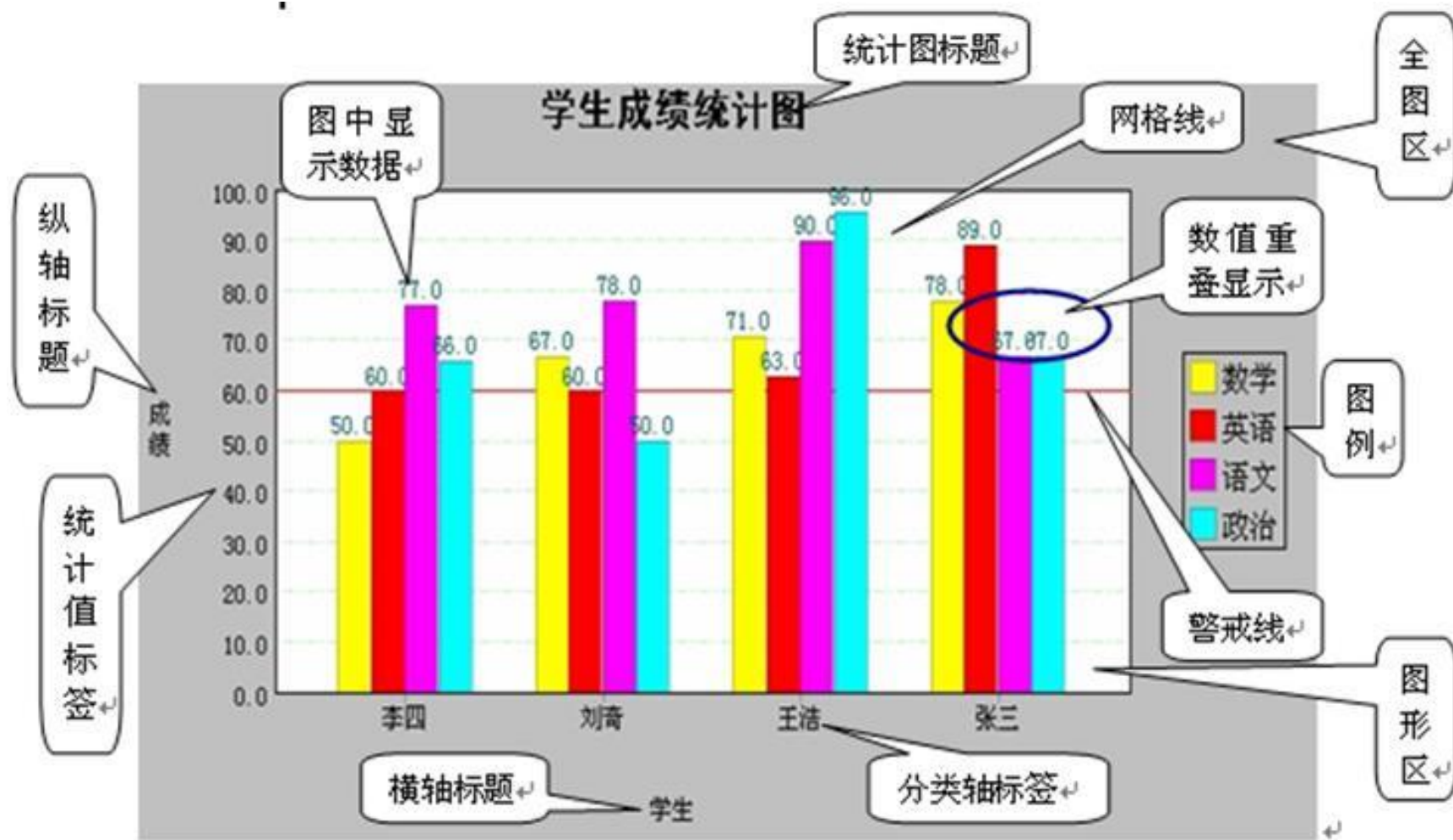
数据来自2-7.xlsx。全距图和工字图
的设置类似，两者都是表达个体在
整体中的位置；全距图表达的是单
个体，而工字图表达是多个体。





统计图元素

图元概览



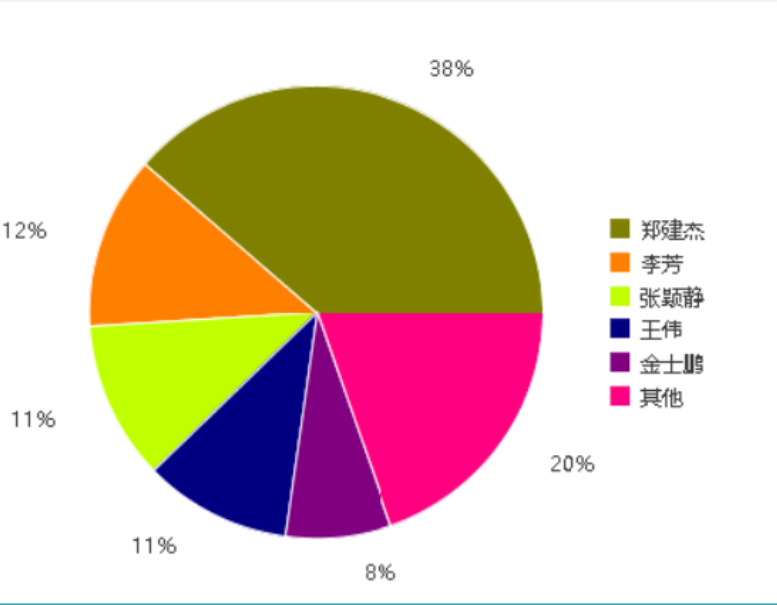
图中尽可能全面的指出了各个统计图的基本元素。最终呈现的统计图是否美观形象，主要就是通过调整这些图元的大小、位置、配色实现的，这个调整过程，最重要的是艺术能力，需要美术人员的参与。

不同类型的统计图，所包含的图元各不相同，即便是一个图中，也可以控制某些辅助图元是否显示。

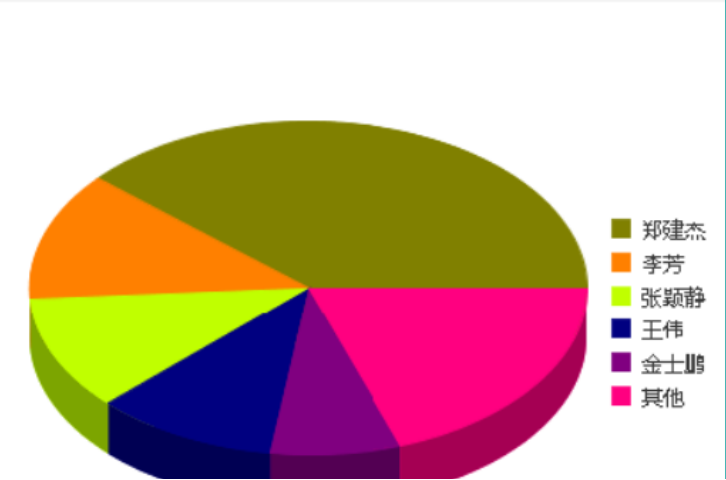
每种制作统计图的产品都会提供对这些图元的外观详细设置功能，设置方式各有差异，但本质上都是针对的这些基础元素。

美工介入效果好

普通饼图

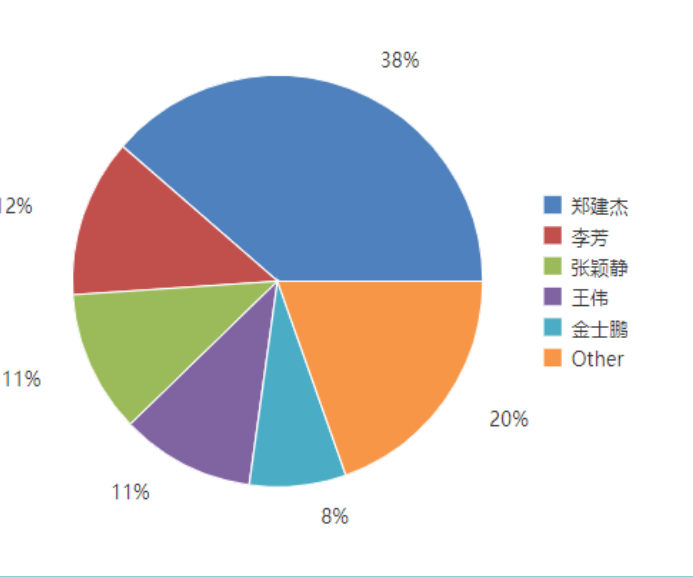


三维饼图

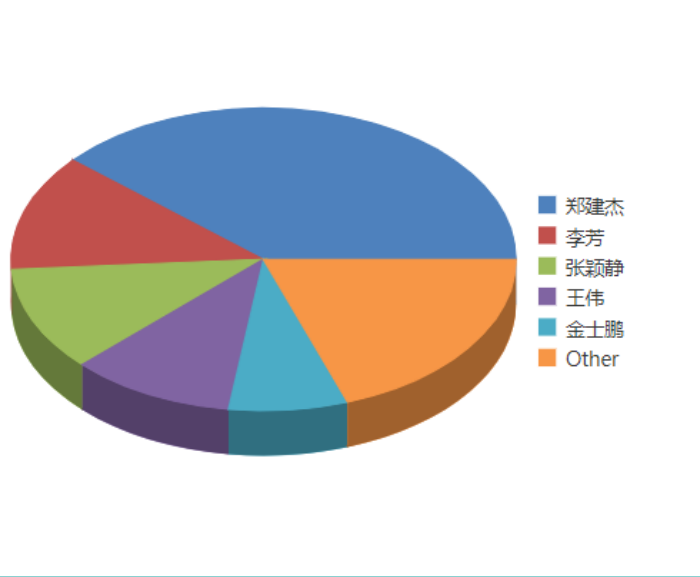


简单调整配色、字体，就能让直观感受不同。这个过程和制作网页一样，非美术的普通人，能感受到正规设计网页与山寨网页的差异，但没能力设计细节。

普通饼图



三维饼图





终端上的统计图技术

显示统计图的主流软件终端-浏览器

HTML5以前，早期的浏览器展现能力不强，报表工具用JAVA等服务器语言实现的统计图，只能以图片形式在浏览器展示，png/svg等格式。

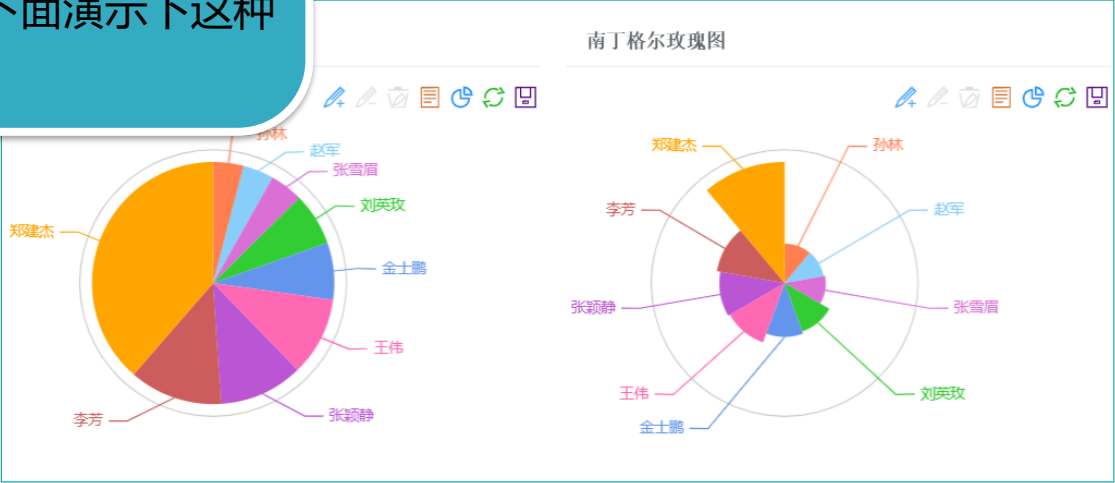
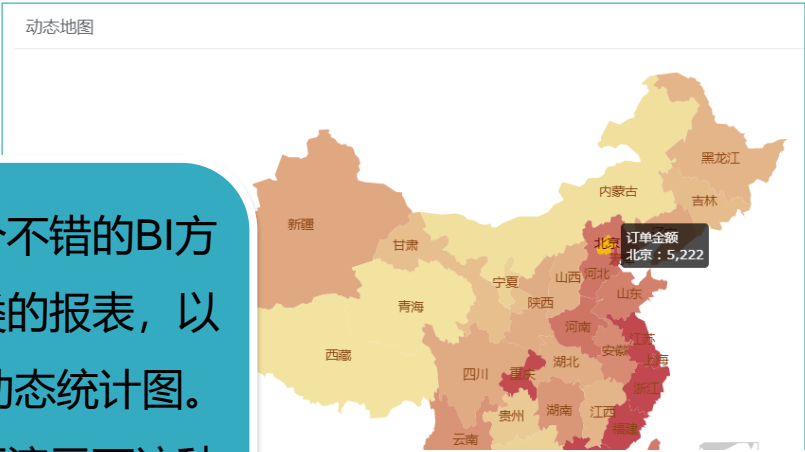
现在浏览器显示技术成熟，特效丰富：诞生了D3、百度echarts等JS图形控件，而且是开源免费的，这些控件擅长于展示、动态互动。

某轮胎		125											
试验报告			LINGLONG TYRE										
申请编号		L16100166		申请人		路波							
申请部门		成品实验室		申请日期		2016-10-27							
试验目的		sdfsdf		测试日期		2016-10-28							
试验结果 Test Result		样品信息		项目信息		谭冬梅002							
						003							
						第1组条件		标准值		结论			
		375ab8				湿度				love		通过	
		eb036a											
		424397											
		7db05											
		140000											
		L161001											
		6601		1									

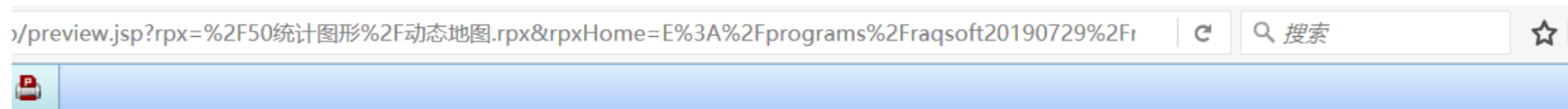
基于这种现状案。以前的报



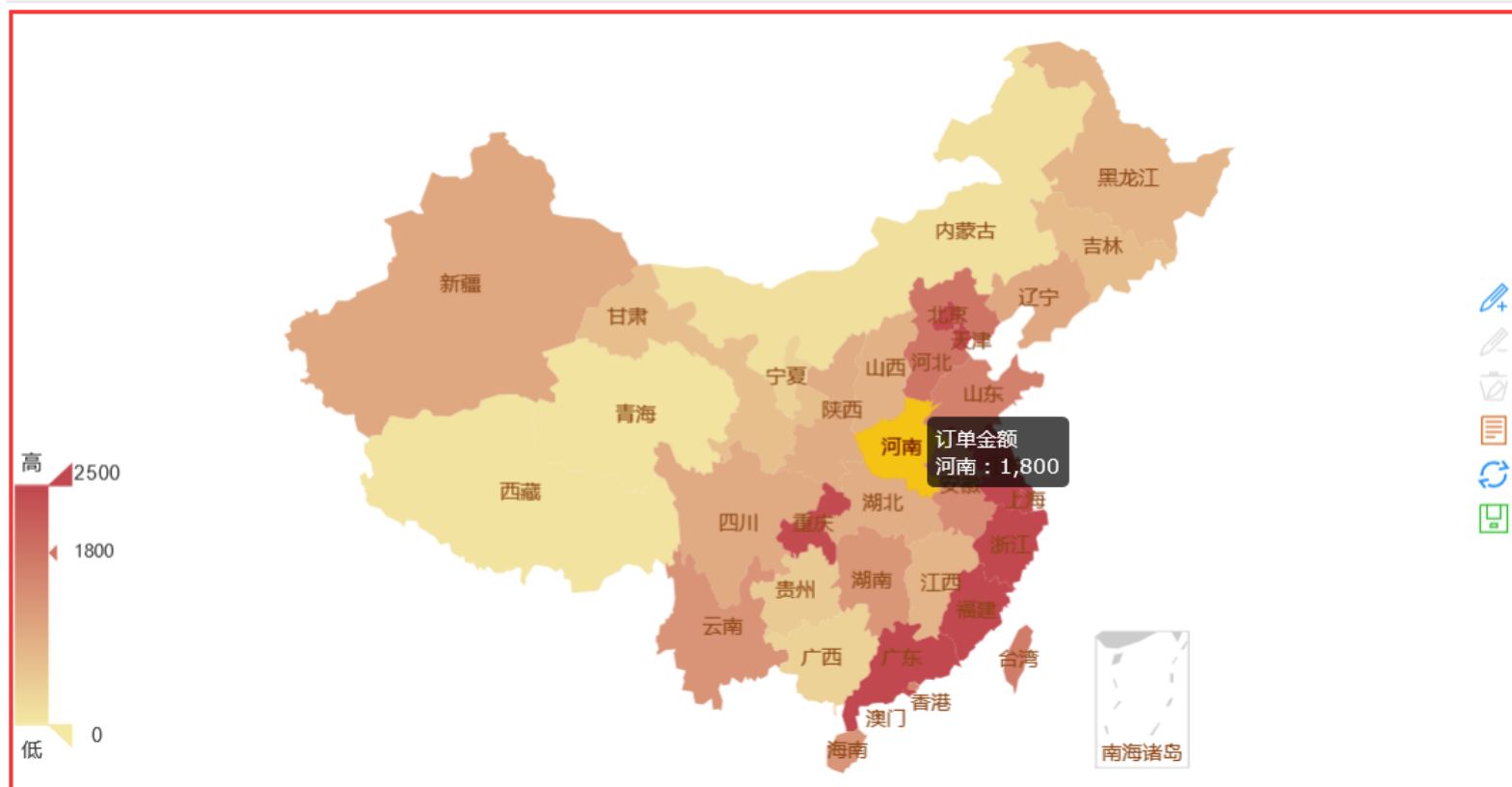
基于这种现状，组合使用也是个不错的BI方案。以前的报表工具负责表格类的报表，以及数据处理；JS控件负责展示动态统计图。本文不讨论方案的性价比，下面演示下这种方案的具体实现。



实现过程1—报表工具—echarts中国地图的订单 (2-8.rpx)



动态地图



先预览一下结果。

这个网页里，红框内是echarts绘制的，之外是报表工具制作。

实现过程2—报表工具+echarts中国地图的订单 (2-8.rpx)

D:\df\Bldoc\demos\示例11.rpx

	A	B	
1	=ds1.select(产品)	=ds1.地区	=ds1.订单金额
2			
3	动态地图		
4			

2、制作报表，A5格子右键“第三方图形”

5

清除编辑风格	
插入行	Ctrl+Enter
插入列	Ctrl+Shift+Insert
追加数据行	Alt+Enter
追加数据列	Alt+Shift+Insert
边框设置	
条形码	
统计图	
第三方图形	
子报表	

内建数据集

地区	订单金额	产品
北京	5222	iphone
天津	4000	iphone
上海	5422	iphone
重庆	2433	iphone

1、准备数据

第三方图形 (双击右边功能面板空白处，对话框最大化或还原)

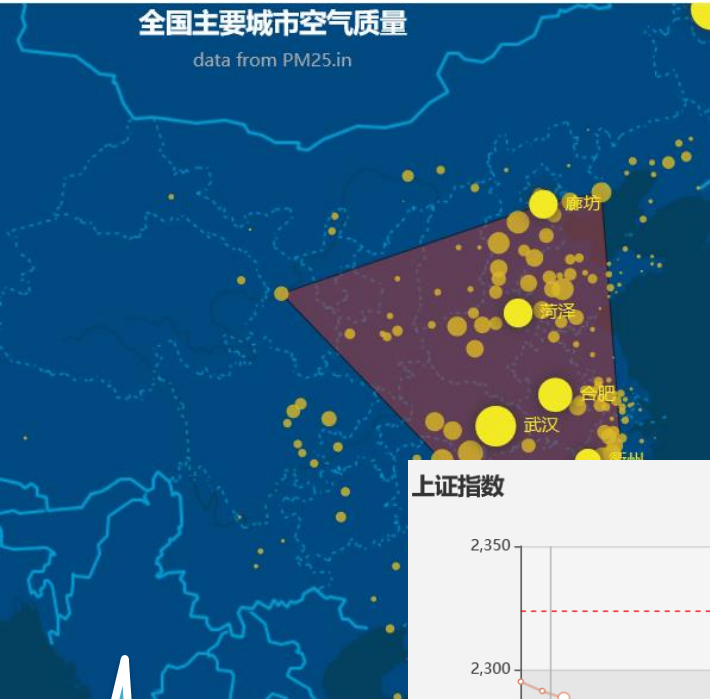
名称	表达式
legend	[订单金额]
seriesName	[订单金额]
dataname	ds1.select@r(地区...地区)
datavalue	C10

JavaScript 代码

```
1 <!-- 为ECharts准备一个具备大小（宽高）的Dom -->
2 <div id="id_4123483372509018" style="width:${width}px;height:${height}
3 <!-- ECharts单文件引入 -->
4 <!-- <script src="http://echarts.baidu.com/build/dist/echarts.js"></script>
5 <script type="text/javascript">
6 var legend = ${legend};
7 var a={
8 toString : function()
9 {
10 var slen=${seriesName}.length;
11 var clen=${dataname}.length;
12 var vsum=[];
13 var value=${datavalue};//[58,345,600,255];
14 var names = ${dataname};//['北京','天津','上海','湖南']
15 var series=${seriesName};
```

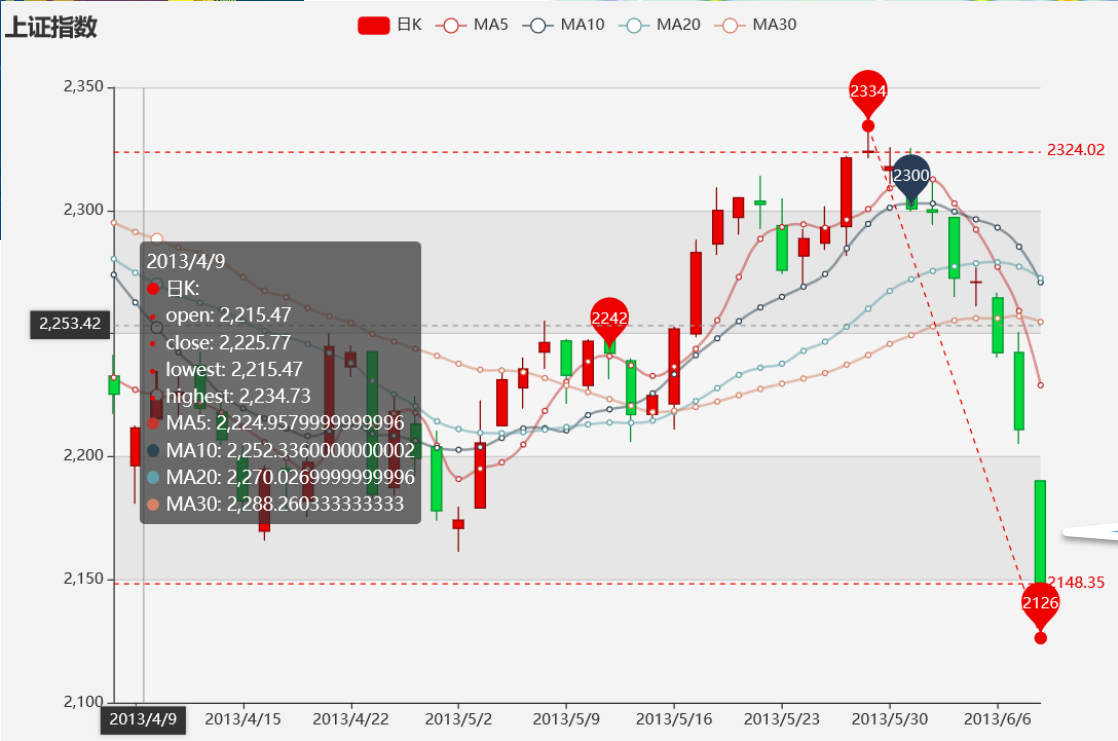
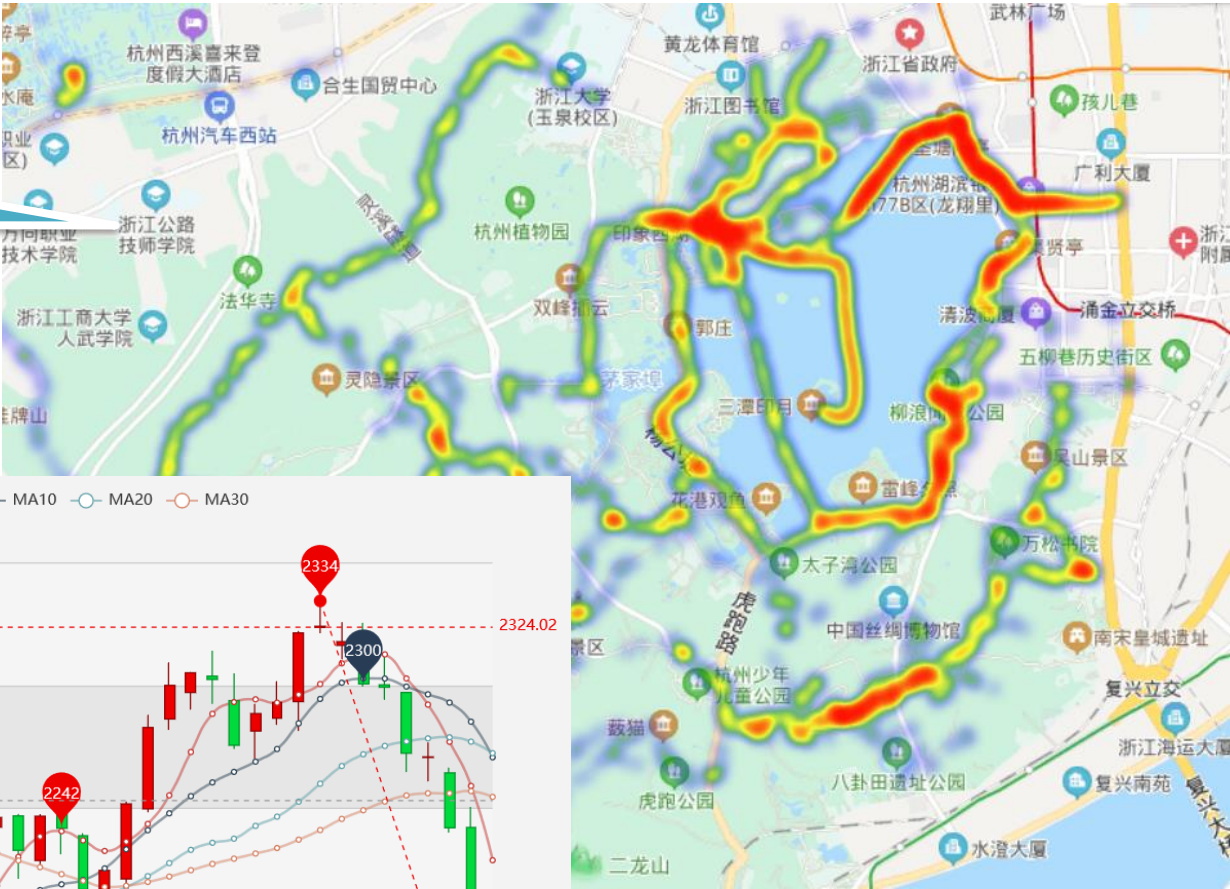
3、红色箭头是报表工具向echarts传送数据的过程。而echarts各种统计图的模板使用方式，去echarts官网学习就可以了，除了包含上面讲到的统计图基本知识外，更多图类型、动态效果的控制组成更丰富、庞大的一系列知识。

JS图形控件优势1—图形更丰富



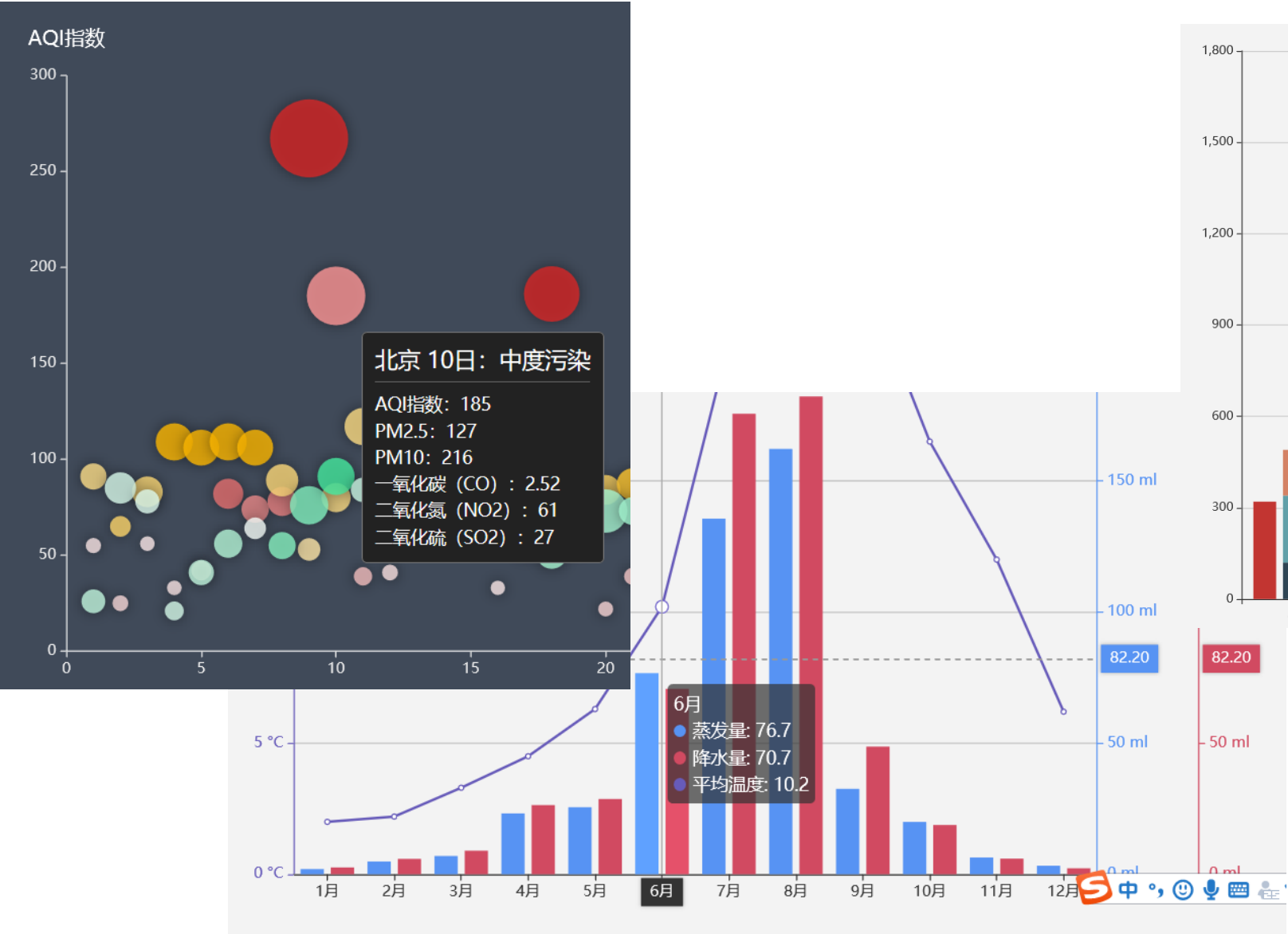
地图

热力图

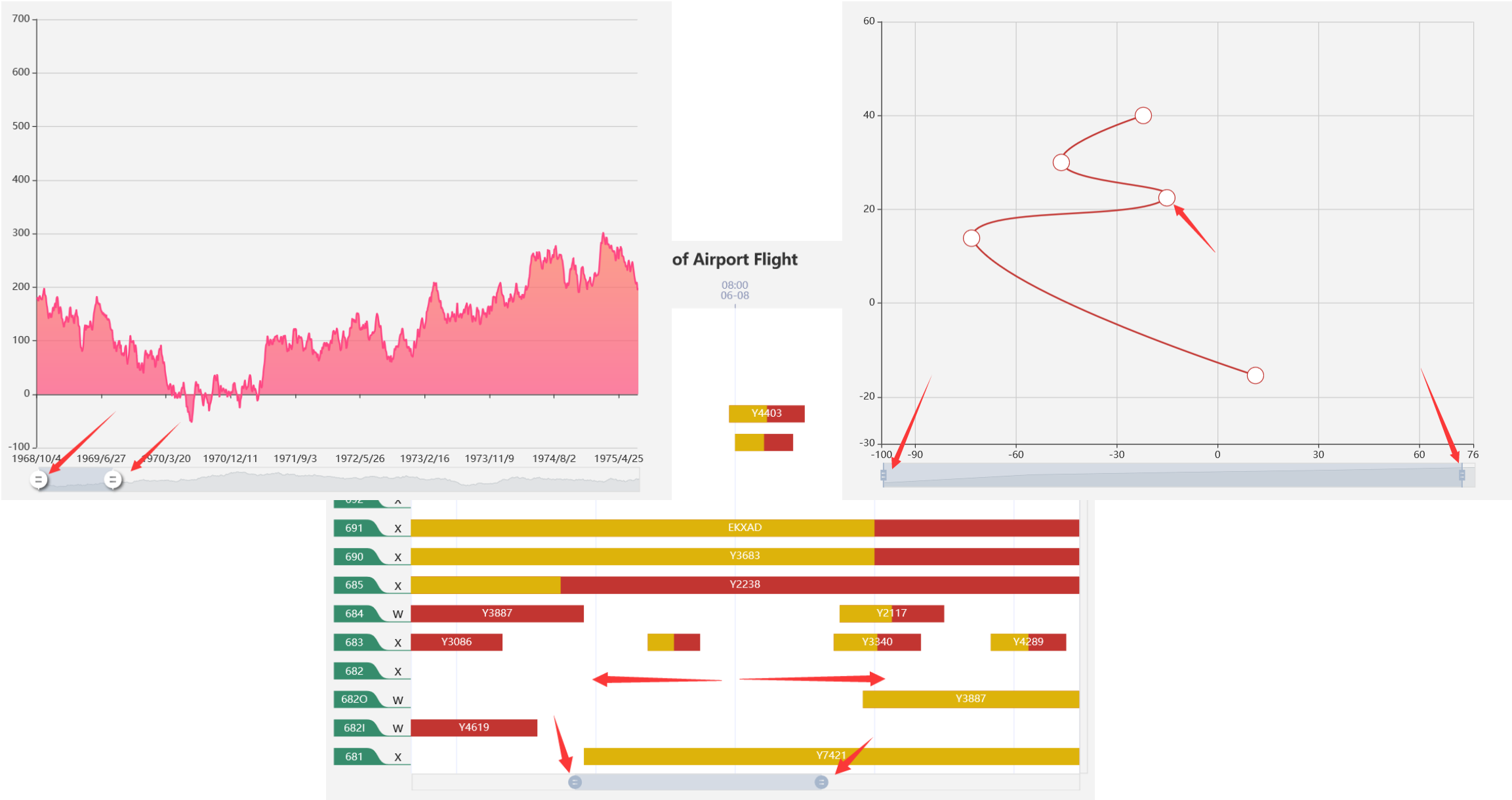


K线图

JS图形控件优势2—鼠标悬浮提示



JS图形控件优势3—拖拽、缩放数据范围



JS图形控件优势4—动态显示/隐藏系列



小结

- ✓ 统计图是报表的重要组成部分
- ✓ 不同统计图类型展示不同的数据特点，【分类-系列】最为常见。统计图组成元素有一定的共性，但也存在较大的个性差异。
- ✓ 近年，开源的JS控件统计图有了很大发展，内置了较强的互动特效、美观样式。

THANKS

感谢观看



乾学院 · VIP课堂

商业智能 呈现与输出技术



CONTENTS

本章内容

1. 打印
2. 导出
3. 宏
4. 前端交互
5. 移动端





打印

报表打印技术



APPLET打印



FLASH打印



PDF打印

打印— APPLET打印

优点

- 功能丰富（相比其他方式）
- 和报表工具都是基于Java开发，所以稳定性和兼容性较好
- 适用于经常进行报表打印且每次打印报表内容较多的情况

缺点

- 客户端必须安装Java运行环境（jre）
- 目前仅IE内核浏览器支持

打印— APPLET打印— 举例



APPLET打印预览画面，一纸多页，可缩放显示，可选择页码打印

试用版 -- 打印预览 (双击右边功能面板空白处，对话框最大化或还原)

本报表主要演示了图片字段的上载下传，演示了卡片式报表的打印分页，能够按数据行数分页，能够在一张打印纸上放置多页报表支持离线填报

员工信息卡

员工编号：1	员工姓名：张君	性别：1
部门：10	岗位：2	专业：jyr
学位：3	职称：0	生日：1972-12-31 00
手机：34554	电话：	
通讯地址：		

本报表主要演示了图片字段的上载下传，演示了卡片式报表的打印分页，能够按数据行数分页，能够在一张打印纸上放置多页报表支持离线填报

员工信息卡

员工编号：11	员工姓名：许先文	性别：0
部门：4	岗位：1	专业：
学位：5	职称：0	生日：1974-12-31 00
手机：	电话：	
通讯地址：		

本报表主要演示了图片字段的上载下传，演示了卡片式报表的打印分页，能够按数据行数分页，能够在一张打印纸上放置多页报表支持离线填报

员

员工编号：10	员工姓名：李
部门：2	岗位：2
学位：7	职称：2
手机：	电话：
通讯地址：	

本报表主要演示了图片字段的上载下传，演示了卡片式报表的打印分页，能够按数据行数分页，能够在一张打印纸上放置多页报表支持离线填报

员

员工编号：2	员工姓名：高
部门：12	岗位：5
学位：6	职称：1
手机：	电话：
通讯地址：	

页码：1/4

到第 1 页

最前页

上一页

下一页

最后页

放 大

缩 小

刷新显示

比例 100%

分栏 1

页面设置

打 印

关 闭

APPLET打印

打印— FLASH打印

优点

- 浏览器几乎都提前装过flash player插件，无需再次下载安装
- 适用打印报表数量和次数较少的客户端使用

缺点

- Flash插件开放接口相对少，不如Applet打印功能丰富
- Flash本身有缺陷，所以也在陆续被限制，没有太多未来

打印— FLASH打印— 举例



FLASH打印预览画面，水平位置对齐和打印缩放等控制面板

localhost:6868/demo/reportServlet?action=41&name=report1&reportFileName=10%24u57FA%24u672C%24u62A5%24u8...

网格式报表

订单ID	客户ID	订购日期	货主地区	运货商	运费
10248	山泰企业	2012年07月04日	华北	3	¥ 32.38
10249	东帝堡	2012年07月05日	华东	1	¥ 11.61
10250	实润	2012年07月08日	华北	2	¥ 65.83
10251	千国	2012年07月08日	华东	1	¥ 41.34
10252	福星制衣厂股份有限公司	2012年07月09日	东北	2	¥ 51.30
10253	实润	2012年07月10日	华北	2	¥ 65.83
10254	海天旅行社	2012年07月11日	华中	1	¥ 11.61
10255	永大企业	2012年07月12日	华北	2	¥ 65.83
10256	凯威国际顾问公司	2012年07月15日	华东	1	¥ 41.34
10257	远东开发	2012年07月16日	华东	1	¥ 41.34
10258	正人资源	2012年07月17日	华东	1	¥ 41.34
10259	三建实业	2012年07月18日	华东	1	¥ 41.34

页码: 1 / 71
到第 1 页

最前页 上一页 下一页 最后页 放大 缩小 比例 100% 水平对齐 居中 分栏 1 打印页 全部 打印缩放 无 黑白打印 清除图层 起始页 1 结束页 71 页面设置 打印 关闭

打印方向: ☒ 纵向 ☐ 横向
页边距毫米(左,右): 24 19
(上,下): 25 25
水平对齐: 居中
分栏: 1
打印缩放: 无
☐ 黑白打印
☐ 清除图层
设置 关闭

FLASH打印

打印方向: ☒ 纵向 ☐ 横向
页边距毫米(左,右): 24 19
(上,下): 25 25
水平对齐: 居中
分栏: 1
打印缩放: 无
☐ 黑白打印
☐ 清除图层
设置 关闭

打印方向: ☒ 纵向 ☐ 横向
页边距毫米(左,右): 24 19
(上,下): 25 25
水平对齐: 居中
分栏: 1
打印缩放: 无
☐ 黑白打印
☐ 清除图层
设置 关闭

打印— PDF打印

优点

- 主流浏览器无限制，兼容性较好
- 应用广泛（国外大部分报表工具都采用的方式）

缺点

- 功能较少

打印— PDF打印— 举例



PDF打印预览画面，功能较少，直接打印

打印

总计：71 张纸

打印

取消

目标打印机

Brother DCP-1618W

页码

全部

例如：1-5、8、11-13

份数

1

更多设置

使用系统对话框进行打印...(Ctrl+Shift+P)

PDF打印

网格式报表

订单ID	客户ID	订购日期	货主地区	运货商
10248	山泰企业	2012年07月04日	华北	3
10249	东帝望	2012年07月05日	华东	1
10250	实翼	2012年07月08日	华北	2
10251	千固	2012年07月08日	华东	1
10252	福星制衣厂股份有限公司	2012年07月09日	东北	2
10253	实翼	2012年07月10日	华北	2
10254	浩天旅行社	2012年07月11日	华中	2
10255	永大企业	2012年07月12日	华北	2
10256	凯诚国际顾问公司	2012年07月15日	华东	2
10257	远东开发	2012年07月16日	华东	1
10258	正人资源	2012年07月17日	华东	1
10259	三捷实业	2012年07月18日	华东	1

打印— 其它打印分页功能— 举例

批量打印

直接打印

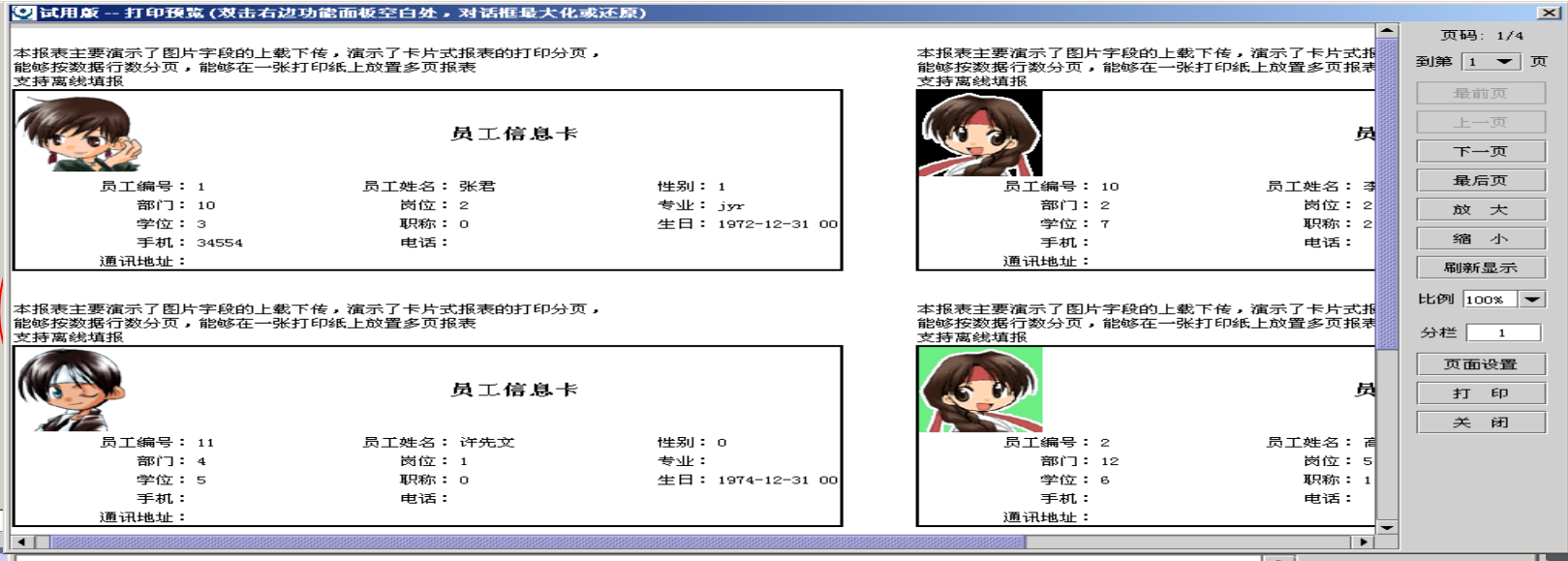
按行分页

按纸分页

强制分页

横向分栏

单元格拆分



打印报表 -- 仅供研发部开发使用 (双击右边功能面板空白处, 对话框最大化或还原)

员工基本信息表

员工编码	员工姓名	生日	性别	部门
1	张君	1972-12-31	女	河南省行
10	李宏浩	1973-12-19	男	石家庄市行
11	许先文	1974-12-31	男	南阳市行
2	高敬东	1975-12-31	男	总行
23	张君	1972-12-31	女	河北省行
24	李宏浩	1973-12-31	男	石家庄市行
25	许先文	1974-12-31	女	南阳市行
26	高敬东	1975-12-31	男	总行
27	尹晋兰	1976-12-30	女	沧州市行
28	范学东	1977-12-30	男	厦门市行
29	张秀兰	1978-12-30	女	福州市行
3	尹晋兰	1976-12-30	女	沧州市行
30	李晶	1979-12-30	男	莆田市行
31	王敏	1980-12-29	女	保定市行

打印—套打绘制—知识点



- 票据打印，即在一张印刷好的空白票据上，准确无误地打印上票据的内容
- 套打的难点在于如何保证绘制出的报表能够精确地套入印刷票据中？
 - ✓ 将票据扫描成图片，作为报表设计的底图，依照图片上格线进行报表描绘
 - ✓ 精确度（以毫米为单位），保证报表的绘制和输出完全与纸张精准一致

打印—套打绘制—举例

底图为票据

套打

	A	B	C	D	E
		发 票 (INVOICE)			
2		致	MINEX METAL CO.,LTD	发 票 编 号	
3		TO:	449-3 SOMAGI OOAZA TSUBAME-CITYNIIGATA-PREF JAPAN	Invoice Number HL22121	
4		Credit No		日 期	
5				DATE DEC. 18,2002	
6		唛 头 , 数 量 及 品 名		单 价	总 价
7		Marks, Quantities&Descriptions		Unit Price	Amount
8		N/M		CNF NIIGATA	
9		PEELER W/WOODEN HANDLE M996070 7200PCS		USD0.21/PC	USD1,512.00
10		S/S YASYOKU-NABE EC E120403 4800PCS		0.17/PC	316.586
11		S/S YASYOKU-NABE EC E120403 4800PCS		0.18/PC	345.685
12		S/S YASYOKU-NABE EC E120403 4800PCS		0.20/PC	564.345
13					
14		MADE IN CHINA			
15		PAY TO:			
16		NINGBO HANGSUN METAL CO.,LTD A/C 8250140010343 BANK OF CHINA,NINGBO BRANCH,ZHEJIANG,CHINA SWIFT: BKCHCNBJ92A			



导出

多种导出类型

多

种导出类型



EXCEL



WORD



PDF



TXT

甚至将**多张报表**导出到一个WORD中**生成报告**

导出— 内容无失真

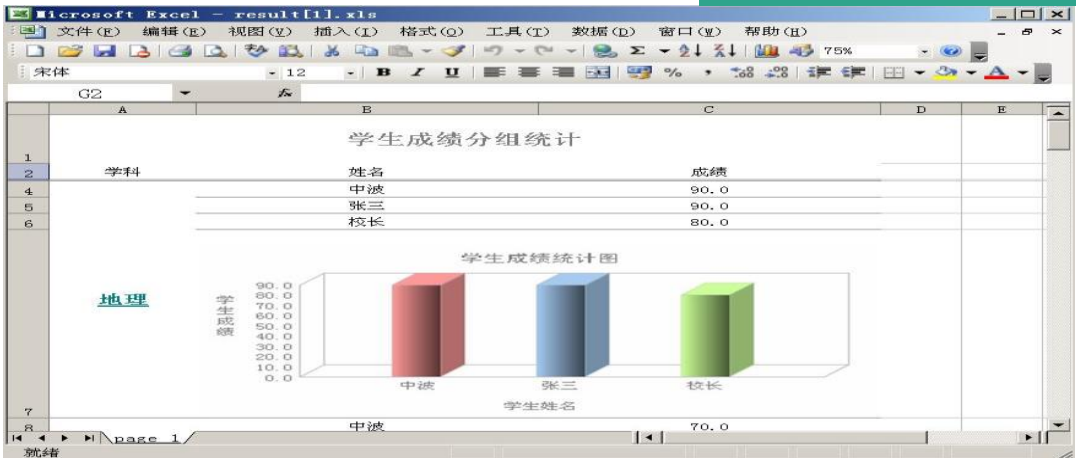


- 在“基础报表技术”中，我们曾讨论过传统条带模型和新型网格类模型的不同
- 采用新型网格类模型的报表编辑方案，避免了传统控件拖拽式编辑方案所带来的位置摆放精度不准，边框之间的重合、对齐等难题
- 也正是因为新型网格类模型能充分体现表格的规律性，所以报表在打印、导出时无失真

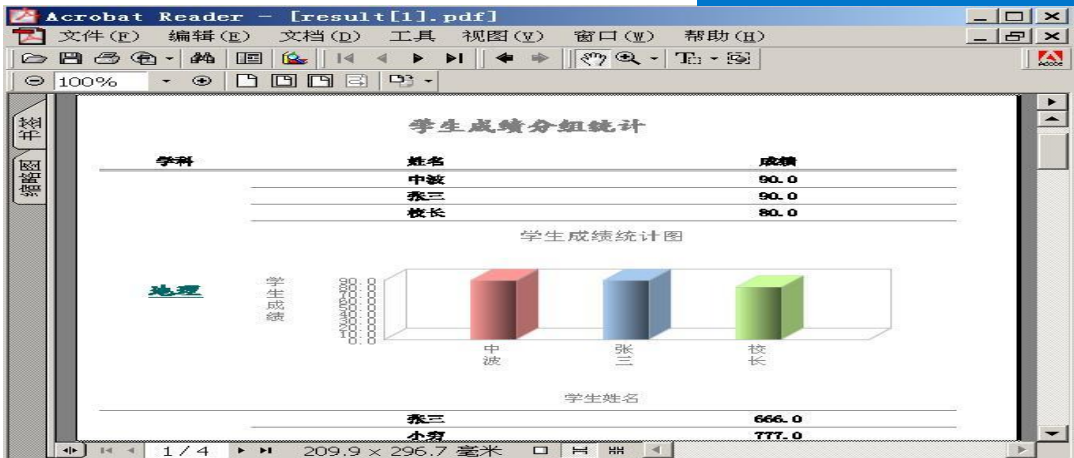
导出无失真

导出—内容无失真—举例

导出EXCEL



导出PDF

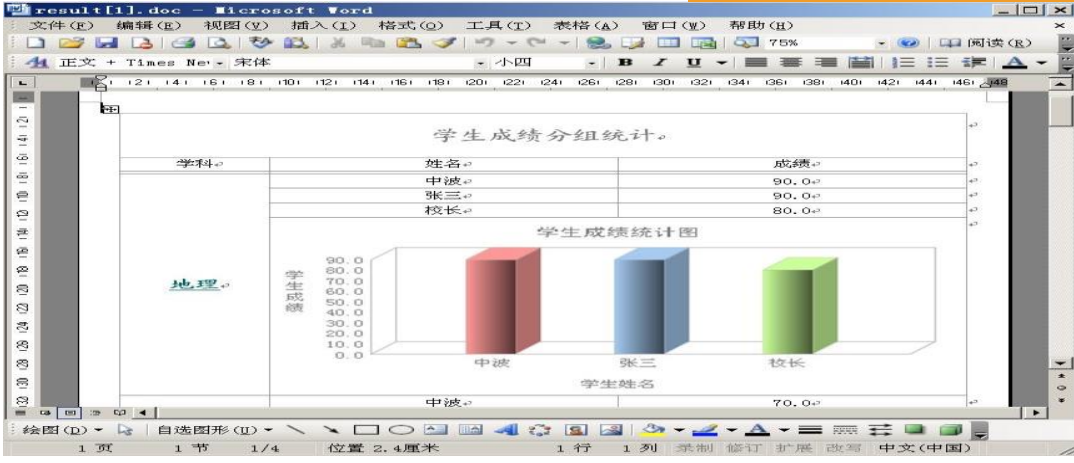


导出TXT

同期比报表


客户名称	年	月	当期合同签订额	比上期	比去年同期
无锡食品加工厂	2002	5	100000.0		
		7	101499.0		
		8	198850.0	195.91%	
		10	153165.0		
	2003	4	147054.0		
5		116659.0	79.33%	116.66%	
6		123719.0	106.85%		
7		399821.0	323.17%	393.92%	
8		144112.0	36.84%	72.47%	
天津长城化工厂	2001	9	168692.0	117.86%	
		10	147715.0	87.56%	96.44%
		4	13200.0		
		6	20640.0		
	2002	7	23876.0	115.68%	
9		20922.0			
2002	4	278833.0		2112.37%	
	6	242453.0		1174.68%	

导出WORD



导出—导入模板（读入EXCEL样式）

- 新型网格模型的报表工具，也能够直接读入Excel绘制好的报表格式，然后在报表设计器中进一步修改，以充分保护用户以往在Excel上的积累，大大减少制表重复工作



Through Put 10月実績(10月10日)v1 - 副本.xls [兼容...]

文件 开始 插入 页面布局 公式 数据 审阅 视图 帮助 操作说明搜索 共享

T5

PH1 Through Put [sheet/1hr] (2011年10月実績) 目標値 82 s/h 以上 Tact 43.9 sec

LOD投入数 75 s/h 以上 Tact 48.0 sec

60 s/h 以上 Tact 60.0 sec

日付	Line	8:00~	9:00~	10:00~	11:00~	12:00~	13:00~	14:00~	15:00~	16:00~	17:00~	18:00~	19:00~	20:00~	21:00~	22:00~	23:00~	24:00~
2011/10/1	BM1	74	77	62	70	56	66	44	32	75	53	77	76	74	65	76	69	66
	RE1	80	75	42	14	81	64	81	66	69	80	81	62	57	16	78	70	82
	BL1	81	40	81	75	45	0	0	22	80	81	71	75	81	74	80	81	34
	GR1	0	0	0	7	7	2	0	0	0	0	0	43	17	0	0	6	24
2011/10/2	PS1	0	3	0	0	3	0	0	20	51	77	19	38	82	40	65	54	81
	BM1	76	76	56	44	77	53	76	26	18	61	50	49	0	0	27	76	76
	RE1	43	23	54	75	29	81	78	44	63	21	0	0	9	81	79	30	46
	BL1	75	71	78	61	0	0	0	0	0	29	81	74	74	76	62	59	81

PH1(DATA) PH2(DATA) 生産量(DATA) 生産量(GRAPH) BM1 ...

润乾报表 (仅限产品试用, 禁止商用)

文件(F) 编辑(E) 报表(B) 工具(T) 远程服务(C) 本地应用(W) 窗口(W) 帮助(H)

D4 = 1

C:\Users\ThinkPad\Desktop\维修分析\Through Put 10月実績(10月10日)\2.xls

PH1 Thro 目標値 82 s/h 以上 Tact +3600Msec

LOD投入数 75 s/h 以上 Tact +3600Msec

60 s/h 以上 Tact +3600Msec

日付	Line	8:00~	9:00~	10:00~	11:00~	12:00~	13:00~	14:00~	15:00~	16:00~	17:00~	18:00~	19:00~	20:00~	21:00~	22:00~	23:00~	24:00~
2011-10-01	BM1	74	77	62	70	56	66	44	32	75	53	77	76	74	65	76	69	66
	RE1	80	75	42	14	81	64	81	66	69	80	81	62	57	16	78	70	82
	BL1	81	40	81	75	45	0	0	22	80	81	71	75	81	74	80	81	34
	GR1	0	0	0	7	7	2	0	0	0	0	0	43	17	0	0	6	24
2011-10-02	PS1	0	3	0	0	3	0	0	20	51	77	19	38	82	40	65	54	81
	BM1	76	76	56	44	77	53	76	26	18	61	50	49	0	0	27	76	76
	RE1	43	23	54	75	29	81	78	44	63	21	0	0	9	81	79	30	46
	BL1	75	71	78	61	0	0	0	0	0	29	81	74	74	76	62	59	81

属性名称 值 表达式

值 值 显示格式 显示值 显示 可视 隐藏行 隐藏列 前景色 背景色 尺寸调整方式 按设计尺寸不变 标签 字体 扩展 扩展方式 默认

导出— WORD文档中嵌入报表— 场景

- 在很多业务场景中需要定期生成WORD文档报告。如图所示：



- 这是某大学年度毕业生就业报告，其中表格和统计图的数据来自数据库



- 通过报表工具单独制作表格和统计图部分并不难，但所有内容混合在一起就不容易了

导出— WORD文档中嵌入报表— 常见解决办法



• 解决办法有哪些？

办法一

每次做好表格和统计图后，再导出，然后手工复制粘贴到word报告中



纯手工操作效率低

办法二

把整个报告都做成一个报表模板，然后再一起导出为word



工具排版有限
导致不够美观

导出— WORD文档中嵌入报表— 分析特征

- 这类报告一般都有规定的模板样式（通常很久不会变动），只要定期替换里面动态变化的信息即可

文档编号:
报告时间:

XX 大学 年毕业生就业报告

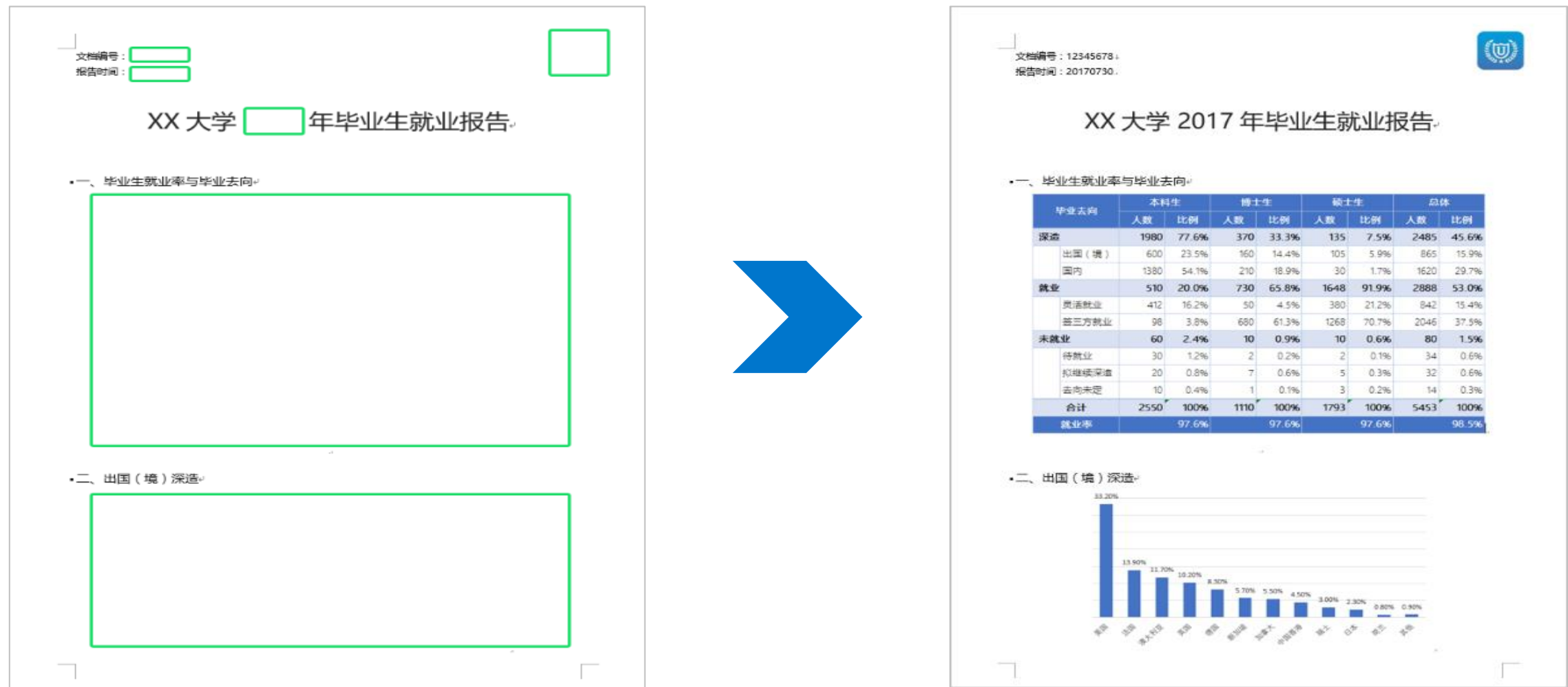
一、毕业生就业率与毕业去向

二、出国（境）深造

绿色线框位置就是需要定期更新的部分，其他文字描述及整体样式固定不变

导出— WORD文档中嵌入报表— 新解决办法

- 预先在word模板的指定位置插入书签，再利用报表引擎将计算好的表格、统计图、文字、图片等内容插入到指定书签位置，定期生成新的word报告





宏

复杂参数

- 我们在“基础报表技术”中，已经知道了报表参数的工作原理，即根据不同参数取出不同数据来呈现。不过，这种方式要求事先把查询条件做死，也就是固定了对应的条件字段，比如按时间段查询：

```
SELECT * FROM 订单 WHERE 订购日期>= ? AND 订购日期<= ?
```

- 基于此，若想用地区查询就不行了，还得再造一个按地区作为查询条件：

```
SELECT * FROM 订单 WHERE 订购日期>= ? AND 订购日期<= ? AND (货主地区= ? OR ? IS NULL)
```

- 如果还需按运货费、客户等条件混合查询，显然，SQL参数写起来非常麻烦
- 如果报表工具只支持这种普通用法，自然不够灵活，于是通用查询就出现了

使用宏表达式

- 报表工具提供一种特殊的字符串型参数，允许将其应用于替换SQL的某一部分，比如WHERE子句
- 前端根据用户输入拼出合法的SQL条件串，作为参数传递给报表替换现有SQL的WHERE子句，这样就可以在同一张报表上实现不同形式的查询条件

```
SELECT 订单ID,客户ID,订购日期,货主地区,运货商,运货费 FROM 订单 WHERE  
${search}
```

- 其中\${search}就是将来会被参数search替换的内容。比如：按地区查询可以把search拼成货主地区=...；无条件时则拼成一个永远为真的条件1=1

使用宏表达式— 举例

- 当\${search}内容为：货主地区= '华东' ， 拼接出的完成SQL如下：

```
SELECT 订单ID,客户ID,订购日期,货主地区,运货商,运货费 FROM 订单 WHERE  
货主地区='华东'
```

- 当\${search}内容为：运货费>=80， 拼接出的完成SQL如下：

```
SELECT 订单ID,客户ID,订购日期,货主地区,运货商,运货费 FROM 订单 WHERE  
运货费>=80
```

- 从上面例子可以看出利用宏可以在报表中轻易替换SQL语句的部分甚至全部
- 同一个报表， 拼接出不同参数值查询不同报表结果， 达到通用查询的目标
- 面对大量类似的报表， 不用单独设计和维护， 能够减轻工作量

参数与宏的对比

思考：既然宏比SQL参数更灵活，那么报表工具还有必要提供SQL参数机制吗？

- 宏只是替换，如何正确的拼接出SQL，保证最后的语句合法，有一定难度
- 数值等简单数据类型可以直接拼到SQL上，日期或字符串常常需要先转义，而不同数据库的规则也有差异，这就会影响报表的兼容性

Oracle宏值写法：

订购日期 = TO_DATE('2014-12-01','yyyy-mm-dd')

Mysql宏值写法：

订购日期 = DATE_FORMAT('2014-12-01','yyyy-mm-dd')



日期类型参数举例

SQL参数则没有这些问题，数据库接口能直接识别各种数据类型的参数，直接传入参数值即可

宏的SQL植入风险

- 报表的参数一般可以通过网页生成并传递到后台，若将恶意的参数值传到后台拼入SQL，欺骗服务器执行恶意SQL命令，就会带来SQL植入风险
- 比如，在前面的宏示例中，若将\${search}值改一下，完整SQL如下：

```
SELECT 订单ID,客户ID,订购日期,货主地区,运货商,运货费 FROM 订单 WHERE  
1=1 UNION SELECT userid,password,null,... FROM user
```

红色为search值

- 上面这是一句可执行的合法SQL，但user表中的信息就被泄露了
- 解决这个问题并不很简单，可以在原SQL的条件上加上括号：

```
SELECT ... FROM 订单 WHERE (${search})
```

- 这看起来是可以挡住上面那种攻击，但仍然可以设计出新的植入串攻击（有兴趣可以思考一下）

小结

- 使用传统的参数方案在有些时候还是相当必要的，这种方案下只能执行固定的SQL，不可能被植入，虽然条件方面不够灵活，但安全性可靠得多

还有个办法是由报表工具提供敏感词检查来规避，当传进来的替换子句包含某些特定词的时候将被拒绝

比如把禁用的敏感词配置到文件中，访问报表服务时，预先判断如果替换子句中包含了这些敏感词，就认为受到攻击并拒绝执行

```
<property name="disallowedParamWordList" value="union, user"/>
```



前端交互

什么是前端交互

- 一般报表都是用来呈现的，但有时也需要简单的交互能力，来表达产品要告诉用户的信息，让用户获得更好、更舒适的体验

表格钻取

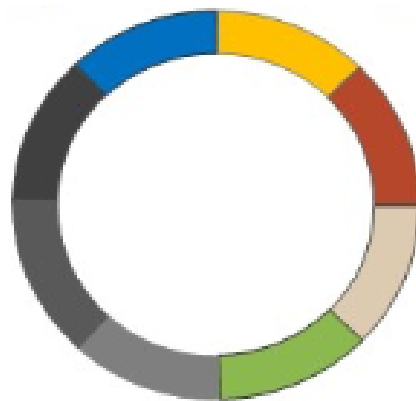
统计图钻取

动态列

点击标题排序

折叠报表

地图联动



前端交互— 钻取

- 钻取是从粗粒度到细粒度观察数据的过程，能够清晰、直观地了解不同层级之间的数据关系，聚焦于感兴趣的内容
- 钻取可以从汇总到明细，从上级到下级等



前端交互— 钻取— 单元格超链接

- 报表中通过对单元格添加超链接方式实现上下级数据钻取的功能，其本质上是利用传递参数的思路，实现对明细报表的过滤和查询

点击任意月份即可钻取出这个月份的订单明细

汇总订单						
年月	订单数量	客户数量	货主数量	运货商数量	雇员数量	订单金额
201508	1	1	1	1	1	¥ 3,748.50
201507	7	4	4	3	3	¥ 11,946.10
201506	1	1	1	1	1	¥ 3,130.00
201505	1	1	1	1	1	¥ 240.00
201503	1	1	1	1	1	¥ 450.00
201502	1	1	1	1	1	¥ 57.00
201412	1	1	1	1	1	¥ 46.70

```
SELECT * FROM 订单
WHERE DATE_FORMAT(订购日期,'%Y%m')=?
```

订单明细						
订购日期	订单编号	客户编号	货主编号	运货商编号	雇员编号	订单金额
2015-07-03	11091	ALFKI	刘小姐	2	3	¥ 450.00
2015-07-09	11105	AROUT	方先生	2	3	¥ 3,116.00
2015-07-17	11104	BLAUS	王先生	2	4	¥ 3,774.05
2015-07-18	11094	BERGS	黄小姐	2	3	¥ 506.05
2015-07-18	11101	AROUT	方先生	4	3	¥ 130.00
2015-07-18	11102	AROUT	方先生	6	3	¥ 240.00
2015-07-18	11103	AROUT	方先生	2	2	¥ 3,730.00

前端交互— 钻取— 多层统计表格（举例）



多张报表通过
参数建立联系
层层钻取，探
索最明细数据

地区销售定额完成统计表

年度：所有历史年度

地区	北京	上海	天津	江苏	浙江	山西	广西	台湾
销售额累计	5963.00	3308.00	1512.00	1701.00	2070.00	2786.00	2394.00	1292.00
销售定额	1665.00	456.00	466.00	1224.00	1577.00	2300.00	1623.76	1811.00
完成比例	358.14%	725.44%	324.46%	138.97%	131.26%	121.13%	147.44%	71.34%

地区：北京

年度：所有历史年度

客户编号	客户名称	销售金额合计
1	蒙牛宏达乳制品公司	4423.00
2	霞光食品添加剂	1540.00
合计：		5963.0

客户：蒙牛宏达乳制品公司

年度：所有历史年度

合同编号	合同日期	合同金额
1	2005-11-09	2855.00
3	2007-11-14	1568.00
合计：		4423.0

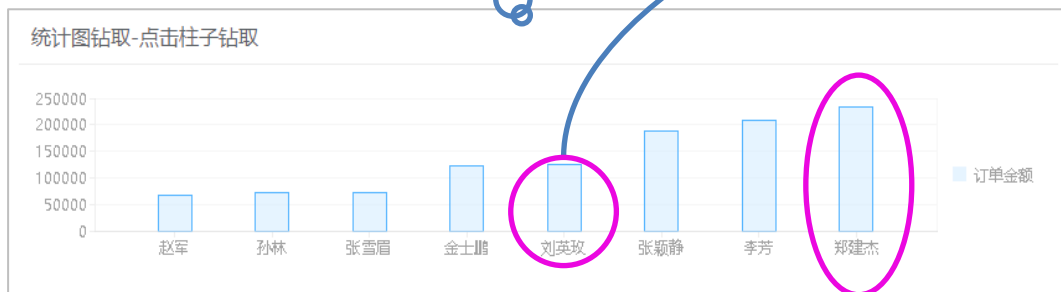
销售订单

合同号	1	交易日期	2005-11-09	
客户	蒙牛宏达乳制品公司			
合同金额	2855.00			
序号	产品	数量	单价	合计
1	河南普通富士苹果	25.00	42.00	1050.00
2	甘肃水晶富士苹果	35.00	31.00	1085.00
3	三月红荔枝	12.00	60.50	720.00

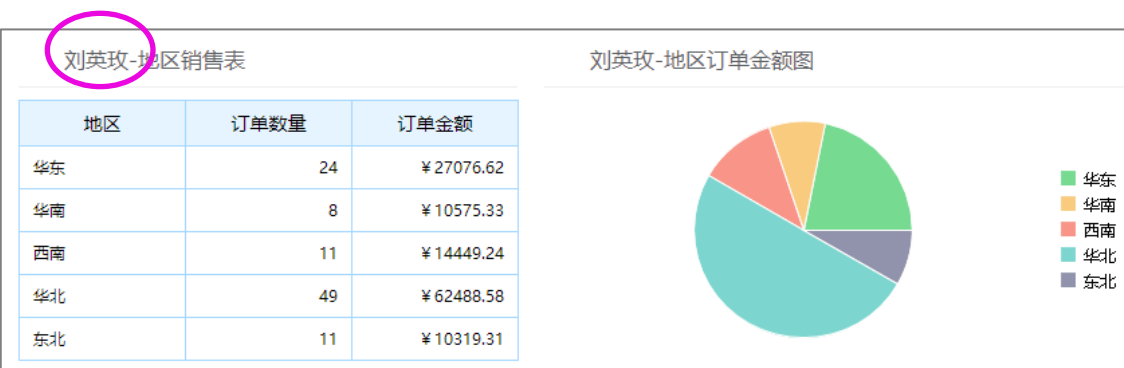
前端交互— 钻取— 统计图超链接

- 统计图超链接和“单元格超链接”的实现原理一样，不同的是统计图需要通过图形、图例的超链接来传递参数，实现对下级报表的过滤和查询

点击柱状图的柱子，页面动态展现该柱子代表的地区销售和订单金额

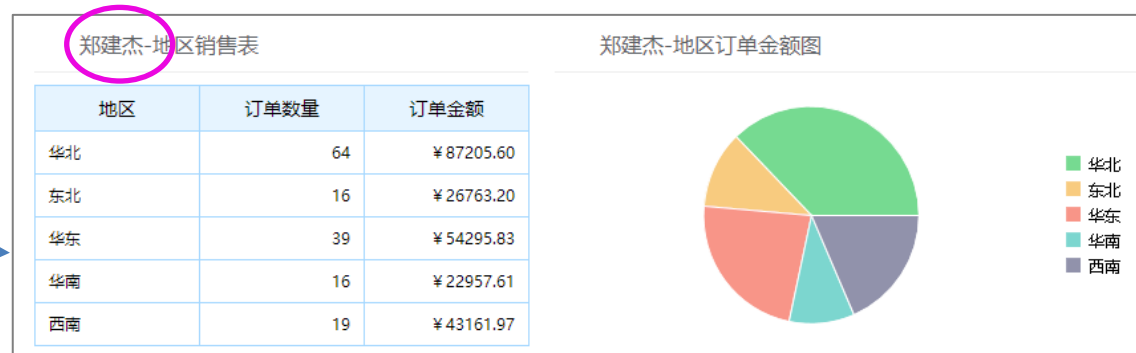


1



- 通过统计图钻取，能够清晰、直观地了解不同层级之间的数据关系

2



前端交互— 点击标题排序

- 超链接结合参数以及单元格表达式来实现，一般步骤：1.通过点击事件确定排序的列.2.判断上一次的排序，决定当次排序取反.3.通过超链接重新请求一次报表并带上1、 2的参数值，传入到报表单元格表达式中完成排序运算并呈现

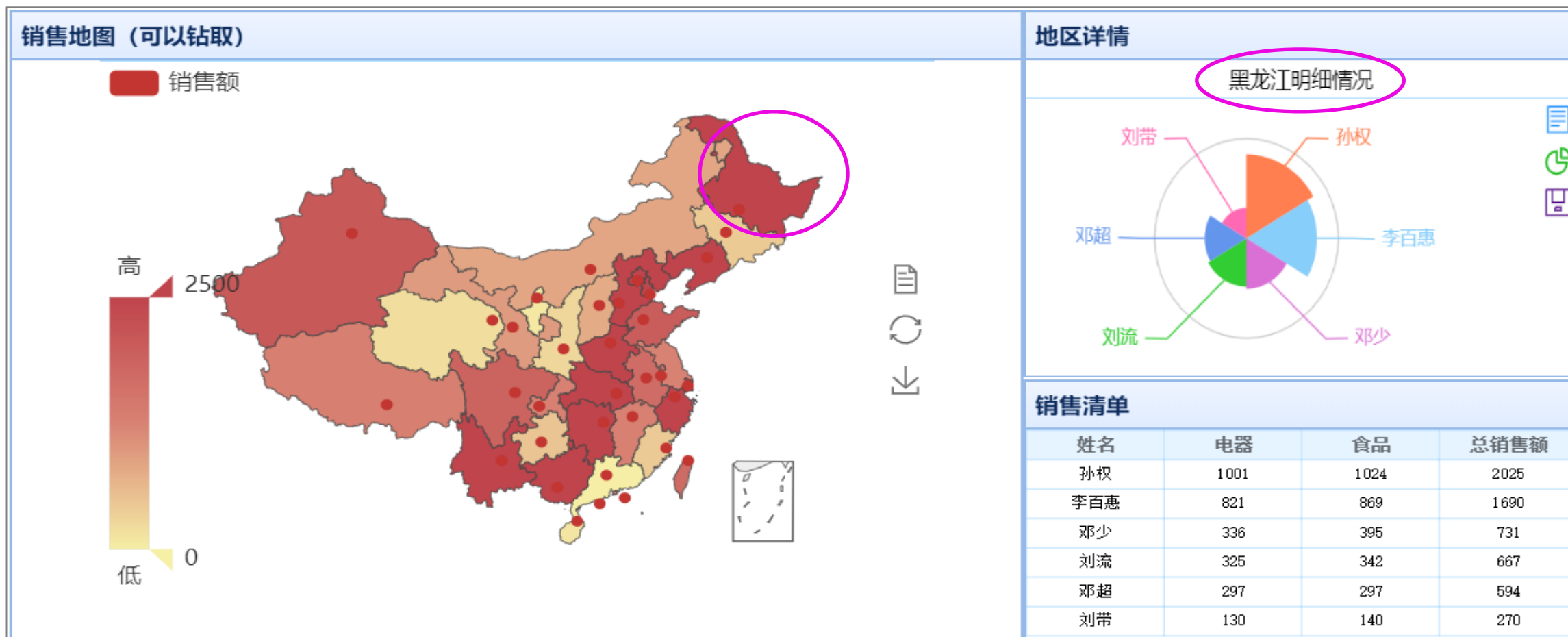
年月	订单数量	客户数量	货主数量	运货商数量	雇员数量	订单金额
201508	1	1	1	1	1	¥ 3,748.50
201507	7					
201506	1					
201505	1					
201503	1					
201502	1					
201412	1					
201411	1					
201409	1					
201408	1					
201407	1					
201406	1					

年月	订单数量	客户数量	货主数量	运货商数量	雇员数量	订单金额
201404	74	47	27	3	9	¥ 123,798.68
201403	73	46				
201401	55	41				
201312	48	36				
201402	54	34				
201308	33	28				
201310	38	27				
201304	31	26				
201301	33	25				
201305	32	25				
201307	33	25				
201309	37	25				

年月	订单数量	客户数量	货主数量	运货商数量	雇员数量	订单金额
201404	74	47	27	3	9	¥ 123,798.68
201403	73	46	22	3	9	¥ 104,932.05
201402	54	34	20	3	9	¥ 99,415.29
201401	55	41	21	3	9	¥ 94,225.31
201312	48	36	17	3	9	¥ 71,398.43
201310	38	27	17	3	9	¥ 66,749.23
201301	33	25	16	3	8	¥ 61,258.07
201309	37	25	16	3	9	¥ 55,629.24
201305	32	25	18	3	9	¥ 53,781.29
201304	31	26	18	3	8	¥ 53,032.95
201307	33	25	14	3	9	¥ 51,020.86
201308	33	28	17	3	9	¥ 47,287.67

前端交互— 地图联动

- 地图联动和“统计图超链接”的实现原理类似，不同的是几张报表都在同一个布局内完成组件间的联动（组件自适应方式、超链接参数传递等）



前端交互— 折叠报表

- 分组统计是常见的数据分析形式，但当分组层级较多时，展现并不直观，希望可以点击各级分组前的按钮，控制组内数据的展开或者收拢
- 成熟的报表工具中，折叠报表制作简单，分组报表结合对应的标签属性即可

地区	城市	最近订购日期	订单个数	订单金额
+ 东北		总计:	16	¥ 26763.20
+ 华北				
+ 华南				

地区	城市	最近订购日期	订单个数	订单金额
- 东北	长春	2013-05-21	5	¥ 10155.16
	大连	2014-03-06	11	¥ 16608.04
		总计:	16	¥ 26763.20
+ 华北				
+ 华南				

地区	城市	最近订购日期	订单个数	订单金额
+ 东北		总计:	16	¥ 26763.20
+ 华北		总计:	64	¥ 87205.60
- 华南	海口	2014-04-02	6	¥ 12203.01
	深圳	2014-04-30	5	¥ 6034.10
	厦门	2014-03-09	5	¥ 4720.50
		总计:	16	¥ 22957.61

前端交互— 动态列



那就是做两张报表分别满足需求就好了？

场景

- 销售人员A：我要随时查看客户表的信息，便于掌握客户最新动态
- 行政人员B：我要随时查看雇员表的信息，便于掌握每位员工的情况

特征

- 格式简单，一般为列表式、交叉式报表

存在的问题

- 若分不同的报表来实现，不仅工作量大而且后期也麻烦
- 利用共同特征，可否一个报表实现动态需求？

前端交互— 动态列

基本原理

- 通过外部参数动态拼接出SQL，作为报表的数据集

```
SELECT ${cols} FROM ${table}
```

这个写法就是我们前面讲过的宏表达式

- 报表中不在写具体字段表达式，而是根据数据集动态获取列名，算出列值

	A	B
1(头)		=to(1,ds1.fcount())
2(头)		=ds1.fname(B1)
3	=ds1.select#0	=ds1.field(B1)

设置B1单元格的扩展方式为横向扩展，A3单元格的扩展方式为纵向扩展

- 该报表相当于在B1中根据数据集ds1的列数进行横向扩展，在A3中根据数据集的行数进行纵向扩展，然后在B2和B3中动态获取列名和列对象，算出列值

前端交互— 动态列— 举例

选择数据表为：产品表
及相关需要显示的字段

报表模板

	A	B
1(头)		=to(1,ds1.fcount())
2(头)		=ds1.fname(B1)
3	=ds1.select(#0	=ds1.field(B1)

选择数据表为：客户表
及相关需要显示的字段

1

2

请选择表：

产品

请选择列：

产品ID,产品名称,单价,单位数量,库存量

查询

产品ID	产品名称	单价	单位数量	库存量
8	胡椒粉	40.0	每箱30盒	6
9	鸡	97.0	每袋500克	29
10	蟹	31.0	每袋500克	31
11	大众奶酪	21.0	每袋6包	22
12	德国奶酪	38.0	每箱12瓶	86
13	龙虾	6.0	每袋500克	24
14	沙茶	23.25	每箱12瓶	35
15	味精	15.5	每箱30盒	39
16	饼干	17.45	每箱30盒	29
17	猪肉	39.0	每袋500克	0
18	墨鱼	62.5	每袋500克	42

请选择表：

客户

请选择列：

客户ID,公司名称,联系人姓名,城市,地区

查询

客户ID	公司名称	联系人姓名	城市	地区
ALFKI	三川实业有限公司	刘小姐	天津	华北
ANATR	东南实业	王先生	天津	华北
ANTON	坦森行贸易	王炫皓	石家庄	华北
AROUT	国顶有限公司	方先生	深圳	华南
BERGS	通恒机械	黄小姐	南京	华东
BLAUS	森通	王先生	天津	华北
BLONP	国皓	黄雅玲	大连	东北
BOLID	迈多贸易	陈先生	西安	西北
BONAP	祥通	刘先生	重庆	西南
BOTTM	广通	王先生	重庆	西南
BSBEV	光明杂志	谢丽秋	深圳	华南
CACTU	威航货运有限公司	刘先生	大连	东北

前端交互— 动态列— 举例（交叉情况）

横向扩展：雇员
纵向扩展：运货商

报表模板		
	A(头)	B
1(头)		=ds1.group(
2	=ds1.group(\${group	=ds1.\${aggr

横向扩展：货主地区
纵向扩展：雇员

1

选上表头： 雇员 选左表头： 运货商 汇总方式： 求和 汇总字段： 订单金额 查询									
	金士鹏	李芳	刘英玫	孙林	王伟	张雪眉	张颖静	赵军	郑建杰
货到付款	0	0	0	0	500	0	98	0	0
急速快递	22,039	50,398	33,959	23,722	63,277	14,806	54,258	24,608	58,060
空运	0	130	0	0	287	0	0	0	0
联邦货运	50,393	64,259	35,396	23,487	41,265	22,589	64,847	13,471	67,725
普通快运	0	3,448	0	0	126	0	137	0	0
统一包裹	52,052	92,043	57,317	26,525	63,185	37,773	72,372	30,629	109,399

2

选上表头： 货主地区 选左表头： 雇员 汇总方式： 求和 汇总字段： 订单金额 查询							
	东北	华北	华东	华南	华中	西北	西南
金士鹏	20,833	57,231	21,793	15,674	0	5,456	3,498
李芳	6,615	110,647	48,987	20,259	0	10,835	12,934
刘英玫	10,319	62,489	27,077	10,575	0	1,763	14,449
孙林	10,813	28,564	18,658	9,856	0	0	5,844
王伟	19,578	66,669	28,481	17,354	1,170	3,749	31,640
张雪眉	224	24,917	18,324	17,517	0	6,750	7,436
张颖静	9,628	90,027	56,713	18,066	0	0	17,280
赵军	1,198	32,944	22,778	6,861	545	0	4,382
郑建杰	26,763	87,206	54,296	22,958	0	800	43,162



移动端

移动端的呈现风格

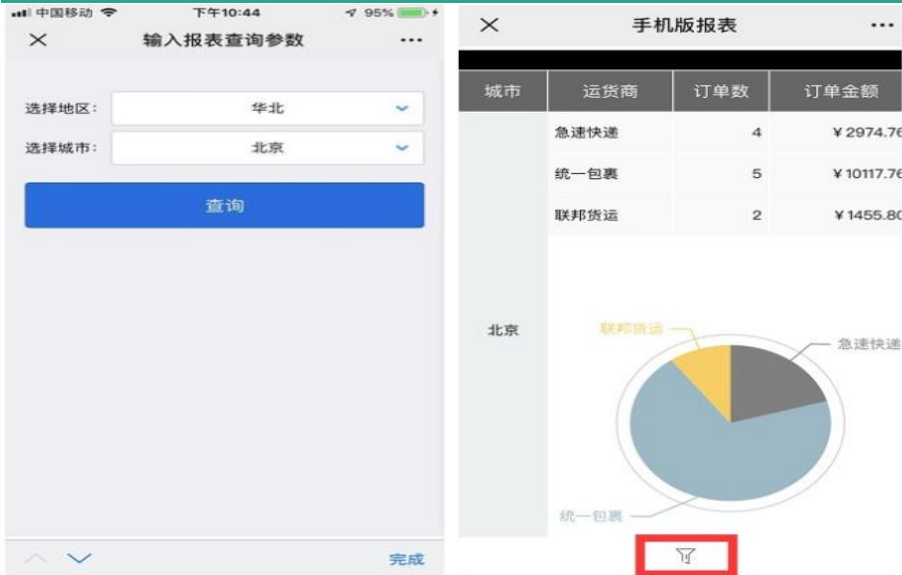
多页签显示



分页显示



条件参数查询



图表联动

长按事件

移动端的报表和 P C 端并没有本质区别，只是在呈现风格上有所不同！



移动端的报表是什么

- 移动端呈现使用HTML5，几乎所有报表工具都能生成HTML网页，所以在移动端呈现并不需要报表工具特殊对待
- 由于移动端在屏幕尺寸变化大，交互方式也所有不同，**自适应屏幕的技术**和**少量特殊事件**（比如长按）需要被关注



移动端自适应技术

- 在报表展现页面中使用JS代码根据屏幕大小及方向对报表尺寸进行自动匹配
- 示例：润乾报表提供自适应模式

取值	含义
0	按设计尺寸展现，不用适应容器
1	始终用报表宽度适应容器宽度，报表高度同比例缩放
2	始终用报表高度适应容器高度，报表宽度同比例缩放
3	竖屏时，相当于 mode=2；横屏时，相当于 mode=1
4	竖屏时，相当于 mode=1；横屏时，相当于 mode=2
5	竖屏时，报表高和宽均分别适应容器的高和宽；横屏时，不需适应容器
6	竖屏时，不需适应容器；横屏时，报表高和宽均分别适应容器的高和宽
7	报表高和宽均分别适应容器的高和宽
8	按宽适应，高度不变

移动端自适应示例— 按报表高宽自适应

- 如果报表想在手机端按高宽自适应的话，那自适应模式应当取值为7，调用该报表的 url 应该为：
- http://ip:6868/demo/matchReport.jsp? rpx=/phone_1.rpx&match=7



手机端浏览器显示的效果

移动端自适应示例— 竖屏按高自适应，横屏按宽自适应

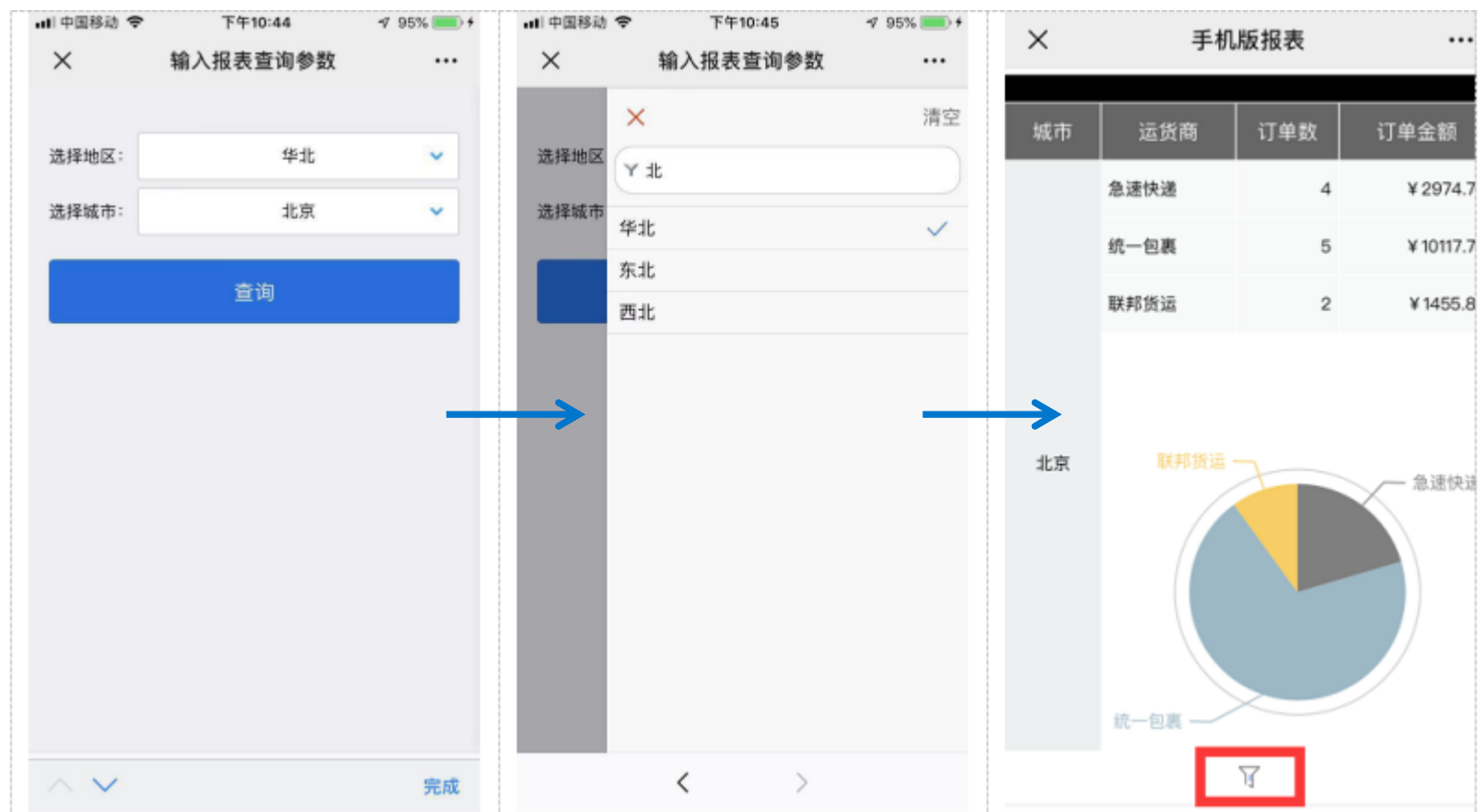
- 如果报表想在手机端竖屏的时候按高自适应，横屏的时候按宽自适应的话，那自适应模式应当取值为3，调用该报表的 url 应该为：
- <http://ip:6868/demo/matchReport.jsp? rpx=/match.rpx&match=3>



手机端浏览器显示的效果

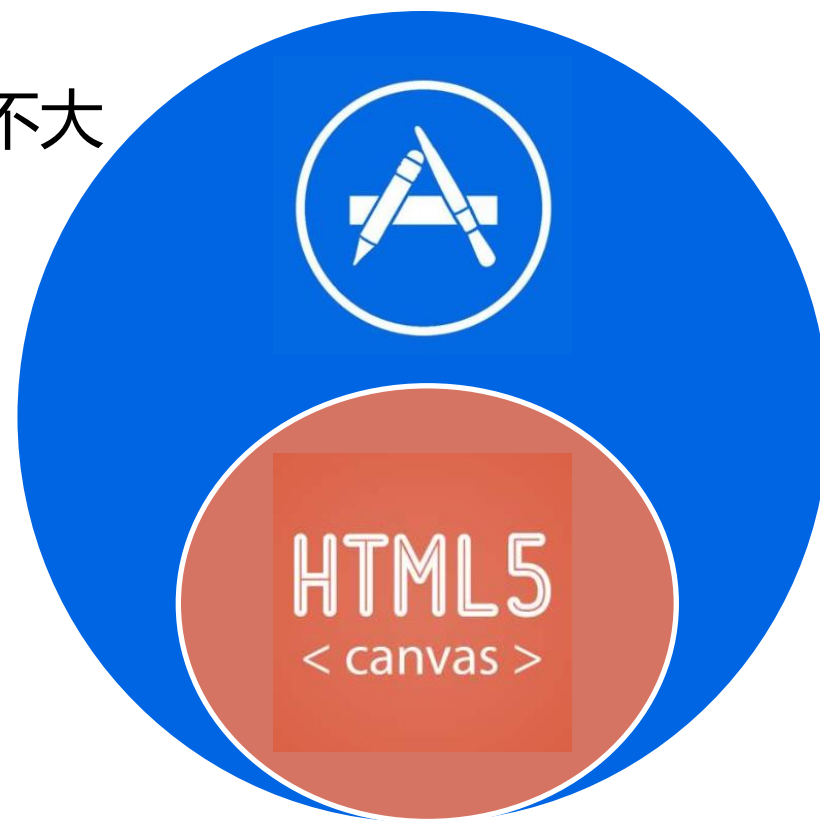
移动端自适应举例— 参数查询

- 参数界面单独显示。其中地区和城市列表做了级联设置，点击页面中的下拉控件，会从右侧推出显示数据项；数据项支持关键字过滤，选择条件后，点击查询，会在另一个页面中显示我们所选条件的数据；点击在页面下方的漏斗图标，返回查询页面



移动端报表APP的开发

- 封装操作系统提供的浏览器控件（所有移动端OS的标配），呈现由后台报表工具生成的HTML
- 外围用APP开发语言实现用户管理等功能
- 微信小程序/公众号等也可以直接呈现HTML
- 报表工具直接提供的APP难以被集成，意义不大



小结

- ✓ 通过本章的学习，我们知道报表除了基本的展现，还有打印、导出、和简单的前端交互
- ✓ 了解了宏的概念、应用场景，以及与参数的区别，还有在报表安全性方面值得注意的方面
- ✓ 掌握前端交互的基本概念，学会使用超级链接实现：表格钻取、统计图钻取、点击标题排序、折叠报表、地图联动等能力
- ✓ 知道了移动端的报表和PC端没有本质区别，只是在呈现风格上有所不同，移动端在自适应屏幕的技术和少量特殊事件需要被关注

THANKS

感谢观看



乾学院 · VIP课堂

商业智能

多维分析技术 (基础分析)



本章目标

- 了解多维分析及其本质
- 掌握多维分析的常见操作
- 学习多维分析中的图表呈现

本章数据说明

- 将“销售数据.xlsx”放置到本地任意目录，如D:\raqsoft
- “销售数据”包括订单、产品、客户、雇员等信息。数据已经加工成一张宽表，为了方便使用方便这里存成Excel文件

订单ID	客户ID	客户名称	雇员ID	姓名	签单日期	签单年	签单月	单价	数量	产品ID	产品名称	城市编码	城市	省直辖市编码	省直辖市
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	14	12	17	猪肉	30201	天津	302	天津
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	9	10	42	糙米	30201	天津	302	天津
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	34	5	72	酸奶酪	30201	天津	302	天津
10249	TOMSP	东帝望	9	张雪眉	2011-02-05	2011	201102	18	9	14	沙茶	20201	青岛	202	山东
10249	TOMSP	东帝望	9	张雪眉	2011-02-05	2011	201102	42	40	51	猪肉干	20201	青岛	202	山东
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	7	10	41	虾子	20401	南昌	204	江西
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	42	35	51	猪肉干	20401	南昌	204	江西
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	16	15	65	海苔酱	20401	南昌	204	江西
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	16	6	22	糯米	30303	秦皇岛	303	河北
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	15	15	57	小米	30303	秦皇岛	303	河北
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	16	20	65	海苔酱	30303	秦皇岛	303	河北
10252	SUPRD	福星制衣厂股份有限公司	2	王伟	2011-02-09	2011	201102	64	40	20	桂花糕	30201	天津	302	天津

订单信息.xlsx

CONTENTS

本章内容

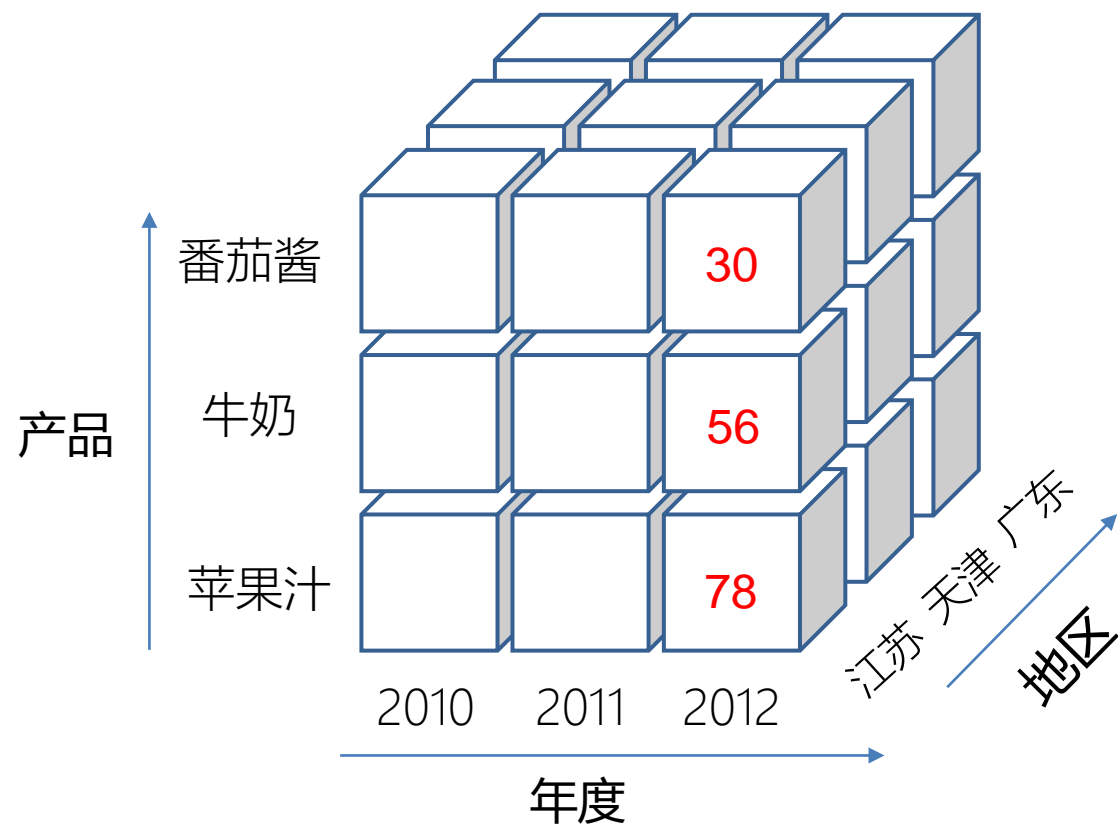
1. 多维分析
2. 汇总方式
3. 维和层
4. 切片
5. 复杂条件
6. 计算列
7. 钻取/旋转
8. 图形





多维分析

多维分析-知识点



常见维度:

日期 (年、季度、月、周)、地区、人员、产品...

常见测度 (指标):

产品价格、销售数量、人数...

基于一个或多个维度汇总指标，如 “查询2012年江苏地区的销售额 ” ($=78+56+30$)

维度和指标

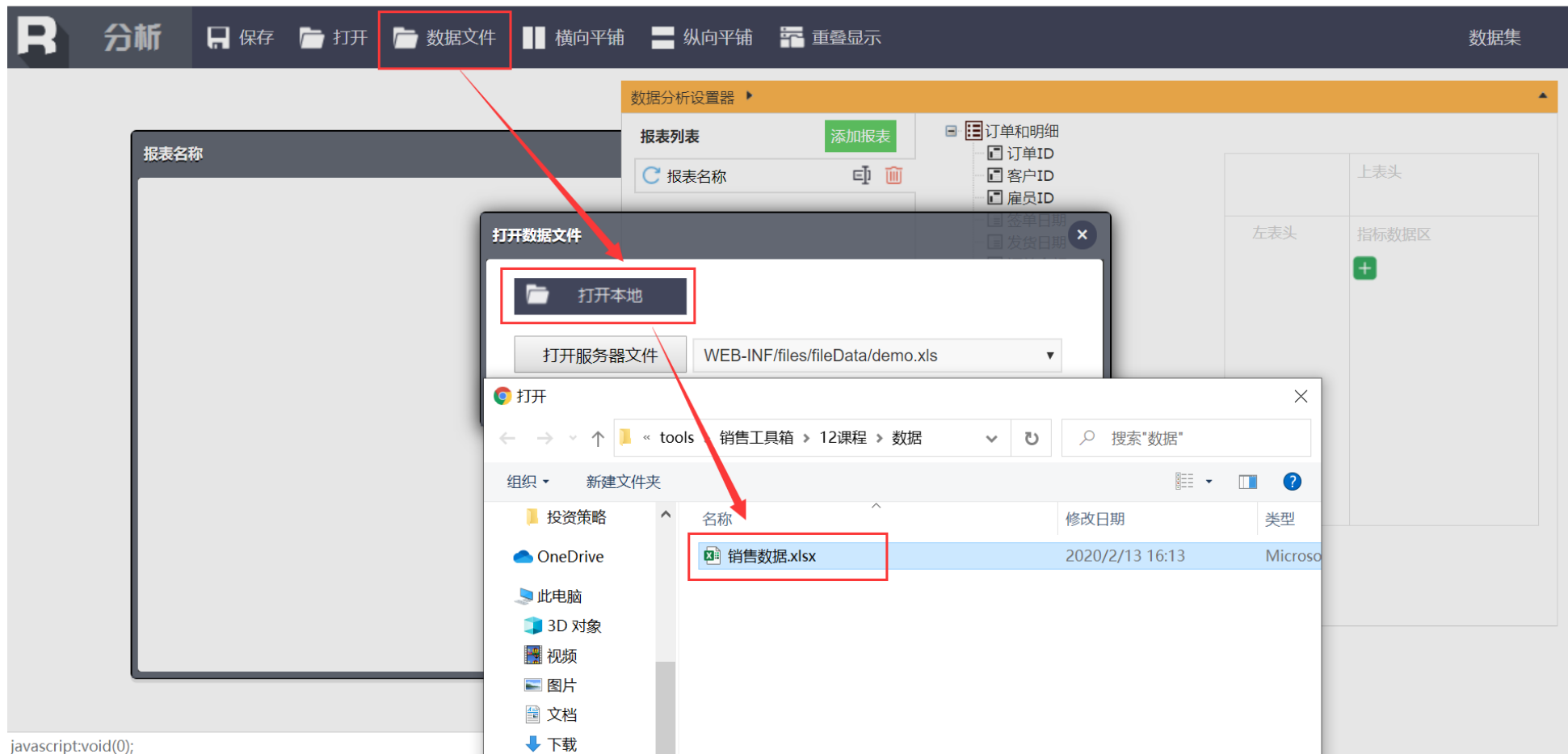
样例数据中，客户、雇员、产品、地区是维度；价格、数量、库存量是指标，也称为测度或度量

维度													测度		
订单ID	客户ID	客户名称	雇员ID	姓名	签单日期	签单年	签单月	产品ID	产品名称	城市编码	城市	省直辖市编码	省直辖市	单价	数量
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	17	猪肉	30201	天津	302	天津	14	12
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	42	糙米	30201	天津	302	天津	9	10
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	72	酸奶酪	30201	天津	302	天津	34	5
10249	TOMSP	东帝望	9	张雪眉	2011-02-05	2011	201102	14	沙茶	20201	青岛	202	山东	18	9
10249	TOMSP	东帝望	9	张雪眉	2011-02-05	2011	201102	51	猪肉干	20201	青岛	202	山东	42	40
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	41	虾子	20401	南昌	204	江西	7	10
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	51	猪肉干	20401	南昌	204	江西	42	35
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	65	海苔酱	20401	南昌	204	江西	16	15
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	22	糯米	30303	秦皇岛	303	河北	16	6
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	57	小米	30303	秦皇岛	303	河北	15	15
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	65	海苔酱	30303	秦皇岛	303	河北	16	20
10252	SUPRD	福星制衣厂股份有限公司	2	王伟	2011-02-09	2011	201102	20	桂花糕	30201	天津	302	天津	64	40
10252	SUPRD	福星制衣厂股份有限公司	2	王伟	2011-02-09	2011	201102	33	浪花奶酪	30201	天津	302	天津	2	25

多维分析-举例

准备数据

打开本地“订单数据.xlsx” Excel，加载订单数据



javascript:void(0);

多维分析-举例

按商品和年份汇总销售数量

报表名称

签单年	2011	2012	
产品名称	数量求和	数量求和	总计
白米	518	746	1264
白奶酪	126	171	297
饼干	545	613	1158
糙米	197	500	697
大众奶酪	293	401	694
蛋糕	137	349	486
德国奶酪	72	272	344
蕃茄酱	160	168	328
干贝	57	68	125
光明奶酪	565	931	1496
桂花糕	160	153	313
海参	175	359	534
海苔酱	348	397	745
海鲜粉	131	632	763

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

签单日期

签单年

签单月

单价

数量

产品ID

产品名称

库存量

城市编码

城市

省直辖市编码

省直辖市

复杂条件

	签单年 ▼
产品名称 ▼	数量求和 ▼
+	

多维分析-练习

练习：按年月汇总每个销售人员（雇员）的销售量（数量）

报表名称

<

思考

1. 上述查询如果用SQL怎么写？（订单信息是一张表）

2. 维度和测度分别在SQL中什么位置？

3. 根据上述SQL，请思考多维分析的本质是什么？

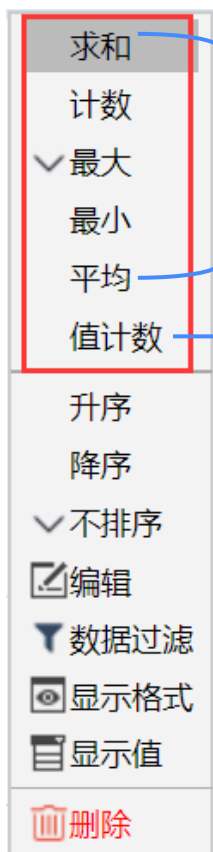
思考



汇总方式

汇总方式-知识点

多维分析中，为了满足不同的统计需要，会有多种汇总方式，包括：求和、最大、最小、求平均、计数、值计数等等



这几个很好理解，不解释

值计数，去重计数，相当于SQL里的去重计数
`COUNT(DISTINCT())`
而计数相当于SQL里的`COUNT()`函数

汇总方式-举例-计数与值计数

查询每年销售过多少种产品

汇总方式-练习与思考

练习：请使用“计数”方式统计上例结果，并比较与值计数的不同

报表名称



签单年	产品名称 计数	总计
2011	823	823
2012	1334	1334
总计	2157	2157

请思考，使用计数统计出来的结果，是什么业务含义？



维和层

维和层-知识点

- **维度**，是我们观察数据的角度，用SQL理解就是group by 后面的分组字段
- 有些维度本身还存在不同的粒度，也就是维的**层次**，比如：时间维包含的不同层次，年、季度、月、周、日等
- 有了层以后就可以方便从同一维度的不同粒度（层）来分析数据

年
季度
月
周
日

时间维的不同层次

国家
地区
省
市
区县

地区维的不同层次

层的使用举例

按签单年、月、日查询产品销量（数量），并逐级汇总

报表名称

签单年:签单月:签单日期

签单年	签单月	签单日期	数量求和	总计
2011	201102	2011-02-04	27	27
		2011-02-05	49	49
		2011-02-08	101	101
		2011-02-09	105	105
		2011-02-10	102	102
		2011-02-11	57	57
		2011-02-12	110	110
		2011-02-15	27	27
		2011-02-16	46	46
		2011-02-17	121	121
		2011-02-18	11	11
		2011-02-19	142	142
		2011-02-22	29	29
		2011-02-23	184	184
		2011-02-24	60	60
		2011-02-25	50	50
		2011-02-26	12	12

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

签单日期

签单年

签单月

签单日

单价

数量

产品ID

产品名称

库存量

城市编码

城市

省直辖市编码

省直辖市

复杂条件

上表头

签单年

签单月

签单日

数量求和

练习与思考

请思考：带有层次的多维分析用SQL语句如何表达？

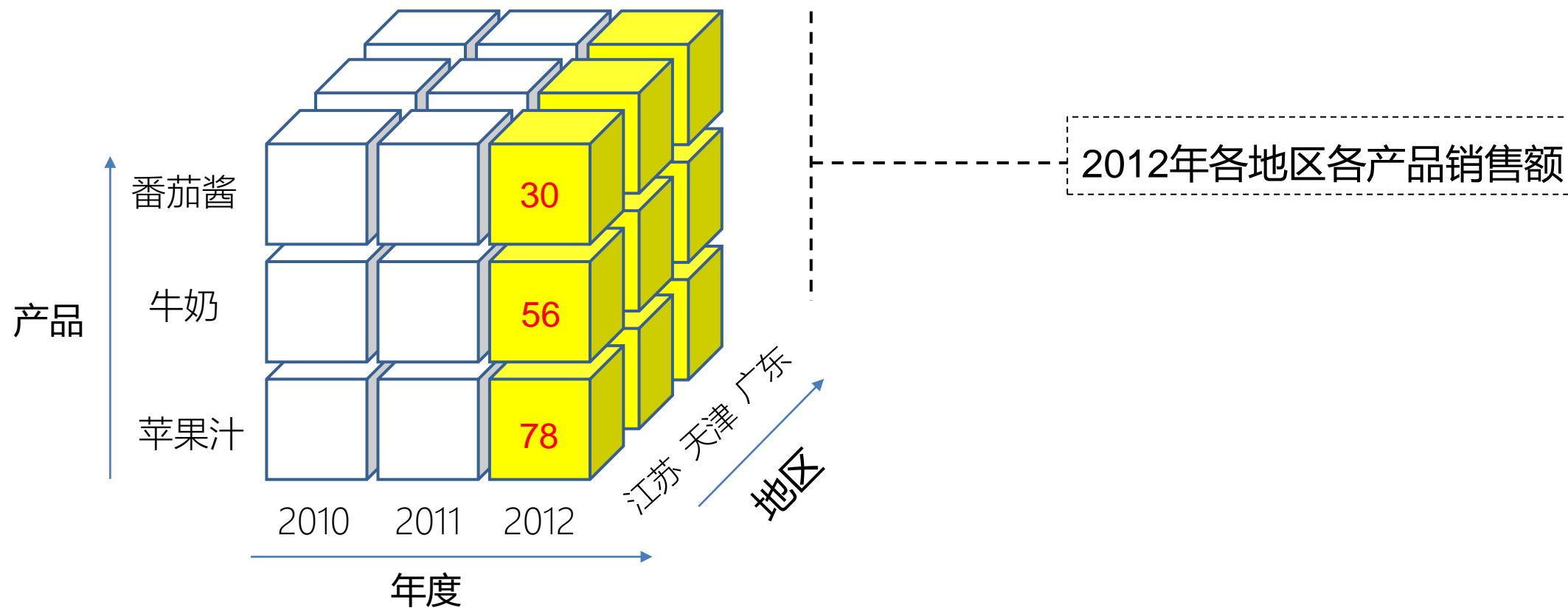
练习：按省市汇总销量，然后调换省市位置，比较前后结果





切片

切片-知识点



指定某些维度值后，剩余2个维度的指标汇总，如 “2012年按地区和产品统计销售额”

切片-举例

使用上述订单数据，查询2012年各类产品在各地区的销量（数量）



切片-举例

使用上述订单数据，查询2012年各类产品在各地区的销量（数量）

报表名称

省直辖市		北京	福建	广东	海
签单年:产品名称		数量求和	数量求和	数量求和	数量
2012	白米	74	21	0	0
	白奶酪	0	0	30	0
	饼干	0	3	71	52
	糙米	0	0	40	0
	大众奶酪	25	10	15	20
	蛋糕	6	0	85	0
	德国奶酪	0	0	86	0
	蕃茄酱	0	0	59	0
	干贝	0	0	0	0
	光明奶酪	30	0	182	30
	桂花糕	0	0	55	0
	海参	15	0	12	30
	海苔酱	0	12	95	21
	海鲜粉	0	0	87	0
	海鲜酱	0	20	216	0

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

签单日期

签单年

签单月

单价

数量

产品ID

产品名称

库存量

城市编码

城市

省直辖市编码

省直辖市

+

复杂条件

	省直辖市 ▼
签单年 ▼	数量求和 ▼
产品名称 ▼	<div>+</div>

切片-练习

练习：请在上例的基础上，查询2012年，福建和江苏，白米和饼干的销售情况

报表名称				
省直辖市		福建	江苏	总计
签单年: 产品名称		数量求和	数量求和	
2012	白米	21	135	156
	饼干	3	175	178
	总计	24	310	334
总计		24	310	334

提示：条件中的多个值使用逗号分隔

思考

如果说多维分析的本质是：

select sum(指标)... from 一个表 group by 维度...

1. 练习中的例子用SQL如何表达？

2. 切片的本质又是什么？



复杂条件

复杂条件-知识点

有时切片切块会带有多个条件，并且条件之间存在不同的优先次序，这就需要在多维分析中将这些复杂条件表达出来

复杂条件在SQL中可以使用多层（括号）嵌套的方式书写，如：

```
(条件1 or 条件2) and 条件3
```

复杂条件-举例

查询2011年四季度和2012年四季度，北京地区的产品销量（数量）

设置查询条件

☐ 简洁显示 (简洁显示会隐藏掉“新增条件、调整条件层次”的功能, 更能凸显条件之间的层次关系)

← → 签单月 等于 201110

← OR | AND

← → 签单月 等于 201111

← OR | AND

← → 签单月 等于 201112

← OR | AND

← → 签单月 等于 201210

← OR | AND

← → 签单月 等于 201211

← OR | AND

← → 签单月 等于 201212

AND | OR

← 省直辖市 等于 北京

保存

清除条件并退出

取消

数据分析设置器

订单ID
客户ID
客户名称
雇员ID
姓名
签单日期
签单年

复杂条件

省直辖市

签单年 数量求和

签单月 +

通过绿色的左右箭头控制条件的层级

复杂条件-练习

练习：沿用上例，请使用月份大于小于设置复杂条件，从而得到上例同样的结果

报表名称			
省直辖市		北京	总计
签单年:签单月		数量求和	
2011	201110	59	59
	201112	20	20
	总计	79	79
2012	201210	155	155
	201211	45	45
	201212	72	72
	总计	272	272
总计		351	351

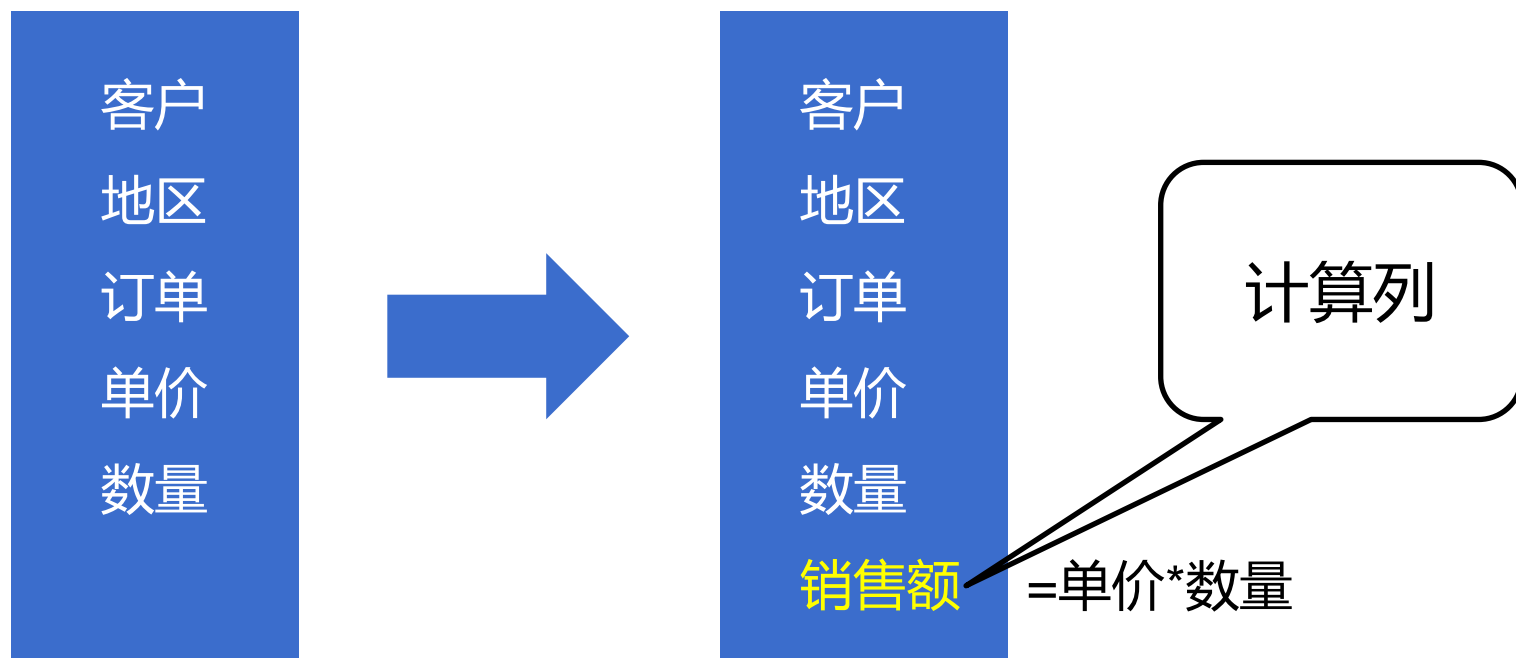
提示：(月份>=201110 and 月份<=201112) or (月份>=201210 and 月份<=201212)



计算列

计算列-知识点

根据已有指标（字段）生成新指标（字段），称新产生的指标为计算列



【注意】计算列通常由指标加工所得，少量可以由维度获得（如根据日期计算工龄、年龄）

计算列-举例

根据销售数量和单价，增加每个产品的**销售额**计算列

编辑计算字段

名称: 销售额

数据类型: 数值

表达式: 单价*数量

订单ID 客户ID 客户名称 雇员ID 姓名 签单日期 签单年 签单月 单价 数量
产品ID 产品名称 库存量 城市编码 城市 省直辖市编码 省直辖市

确定 取消

数据分析设置器

复杂条件

省直辖市

签单年 数量求和
签单月 +

订单ID 客户ID 客户名称 雇员ID 姓名 签单日期 签单年 签单月 单价 数量 产品ID 产品名称 库存量 城市编码 城市 省直辖市编码 省直辖市

+

计算列-举例

按年月和地区汇总销售额

报表名称

省直辖市

北京

福建

广东

海南

签单年:签单月

201102

201103

201104

201105

201106

201107

201108

201109

201110

201111

201112

总计

销售额求和

440

524

0

0

0

13128

3233

0

875

0

385

18585

销售额求和

0

0

0

0

0

0

612

151

0

0

763

销售额求和

7675

822

590

4790

9797

13559

15120

2534

9322

3355

77543

销售额求和

0

0

325

0

440

72

400

0

1750

0

2987

数据分析设置器

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

库存量

销售额

复杂条件

省直辖市

签单年

签单月

销售额求和

计算列-思考与练习

思考：观察本节课程使用的数据，还有哪些指标可以加工成计算列？

练习：根据雇员出生日期，查询销售人员的年龄

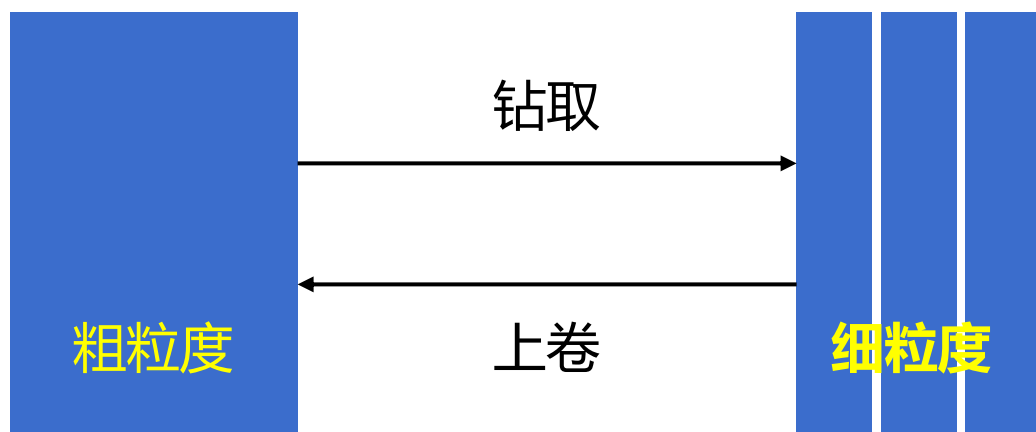
提示：使用出生日期算出年龄



钻取

钻取-知识点

- 钻取是从粗粒度到细粒度观察数据的过程，比如我们发现某个汇总值出现异常，就可以通过**钻取**查看明细数据，找到异常原因。
- 钻取可以从汇总到明细，从上级到下级...



钻取-举例

按年和省份查询销售数量后，查看北京地区2011年的销售明细

报表名称

省直辖市	北京	福建	广东	海南	河北	江苏	浙江
签单年	数量求和	数量求和	数量求和	数量求和	数量求和	数量求和	数量求和
2011	521	54	3771	257	1926	2578	246
2012	1249	262	5858	919	2623	4103	593
总计	1770	316	9629	1176	4549	6681	839

[北京_2011]钻取

省直辖市	签单年	订单ID	客户ID	客户名称	雇员ID	姓名	签单日期	签单
北京	2011	10261	QUEDE	兰格英语	1	张颖	2011-02-19	201102
北京	2011	10261	QUEDE	兰格英语	1	张颖	2011-02-19	201102
北京	2011	10291	QUEDE	兰格英语	1	张颖	2011-03-27	201103
北京	2011	10291	QUEDE	兰格英语	1	张颖	2011-03-27	201103
北京	2011	10291	QUEDE	兰格英语	1	张颖	2011-03-27	201103
北京	2011	10372	QUEEN	留学服务中心	1	张颖	2011-07-04	201107
北京	2011	10372	QUEEN	留学服务中心	1	张颖	2011-07-04	201107
北京	2011	10372	QUEEN	留学服务中心	1	张颖	2011-07-04	201107
北京	2011	10372	QUEEN	留学服务中心	1	张颖	2011-07-04	201107

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

签单日期

签单年

签单月

单价

数量

产品ID

产品名称

库存量

城市编码

城市

省直辖市编码

省直辖市

复杂条件

	省直辖市 ▼
签单年 ▼	数量求和 ▼

从北京地区2011年汇总值查看销售明细记录

思考

我们知道，钻取是从粗粒度到细粒度查看数据的过程，原来的分组汇总现在变成了明细查询

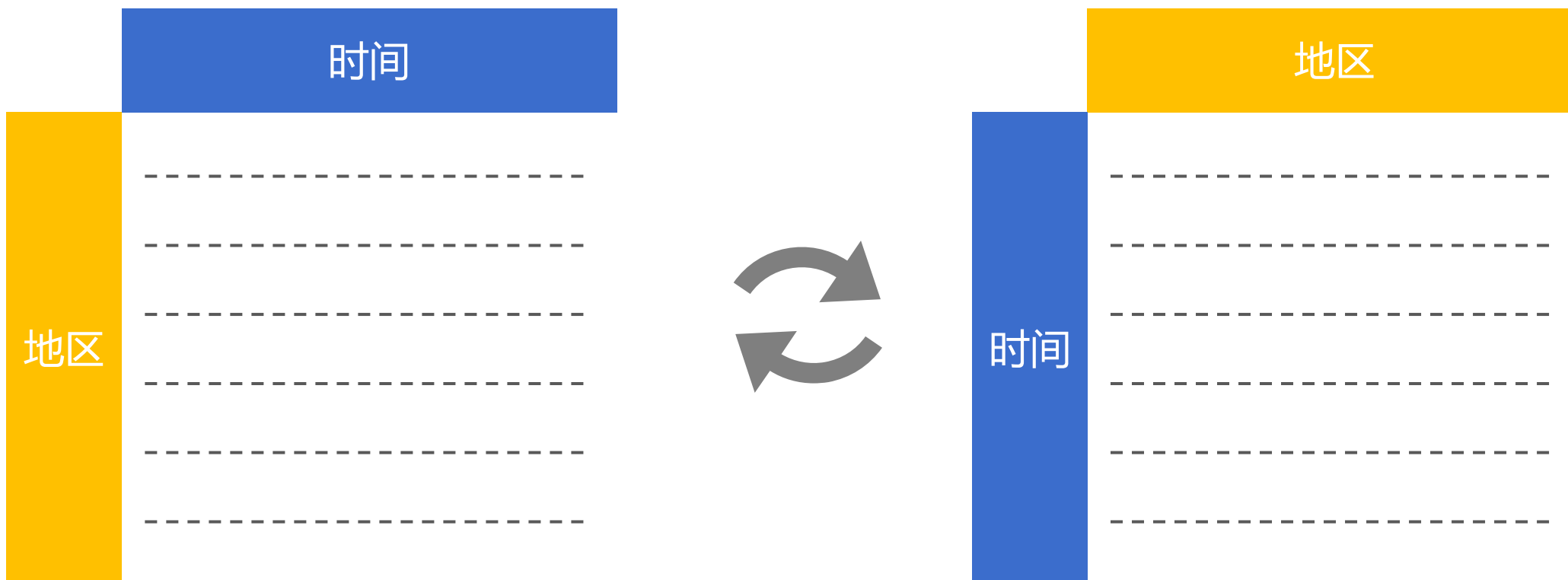
1. 当用户点击了汇总值，后端发生了什么动作？（可以从维度、参数和SQL的角度考虑）
2. 从细粒度到粗粒度查看数据称为上卷，上面的例子如何上卷？



旋转

旋转-知识点

旋转也称行列变换，是指维度从行到列或者从列到行，旋转是为了满足用户观察数据的不同习惯



旋转-举例

按省、年月统计销售数量，并行列互换维度

报表名称

省直辖市		北京	福建	广东	
签单年:签单月		数量求和	数量求和	数量求和	
2011	201102	40	0	480	0
	201103	46	0	81	0
	201104	0	0	50	15
	201105	0	0	50	0
	201106	0	0	428	44
	201107	208	0	735	18
	201108	148	0	751	50
	201109	0	32	236	0
	201110	59	22	431	130
	201111	0	0	339	0
	201112	20	0	190	0
	总计	521	54	3771	
	201201	0	31	350	0
	201202	139	0	462	0

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

签单日期

签单年

签单月

单价

数量

产品ID

产品名称

库存量

城市编码

城市

省直辖市编码

省直辖市

+

复杂条件

省直辖市 ▼

签单年 ▼

签单月 ▼

数量求和 ▼

+

旋转-举例

按省、年月统计销售数量，并行列互换维度

报表名称

单年: 签单月

	201102	201103	201104	201105	201106
省直辖市	数量求和	数量求和	数量求和	数量求和	数量求和
北京	40	46	0	0	0
福建	0	0	0	0	0
广东	480	81	50	50	428
海南	0	0	15	0	44
河北	188	75	85	423	24
江苏	250	202	93	303	197
江西	162	0	0	0	0
辽宁	61	0	80	174	138
山东	49	0	0	0	0
陕西	0	0	44	126	0
上海	0	0	0	0	0
四川	0	0	0	0	0
天津	203	773	582	250	754

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

签单日期

签单年

签单月

单价

数量

产品ID

产品名称

库存量

城市编码

城市

省直辖市编码

省直辖市

复杂条件

签单年 ▼

签单月 ▼

省直辖市 ▼

数量求和 ▼

+

旋转-练习与思考

练习：将上例中的维度任意组合（旋转），并观察结果

思考：解释每次组合后查询结果的业务含义

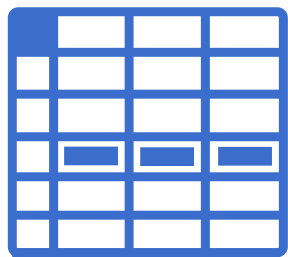


图形

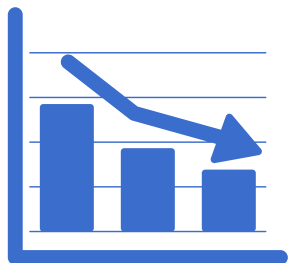
图形-知识点

- 数据呈现形式除了表格还有图形，适当使用图形可以让数据观察更直观
- 而图形的冲击力往往是可视化实现效果的关键（可视化内容在后续课程）

数据
表现
形式



表格



图形

图形-举例

查询2012年各月销量变化趋势（折线图）

数据分析设置器

报表列表

报表名称

添加报表

打印

删除

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

+

复杂条件

上表头

左表头

指标数据区

+

添加报表

fileQuery

月销量走势

☒ 使用报表模板

选择模板...

上子图

折线

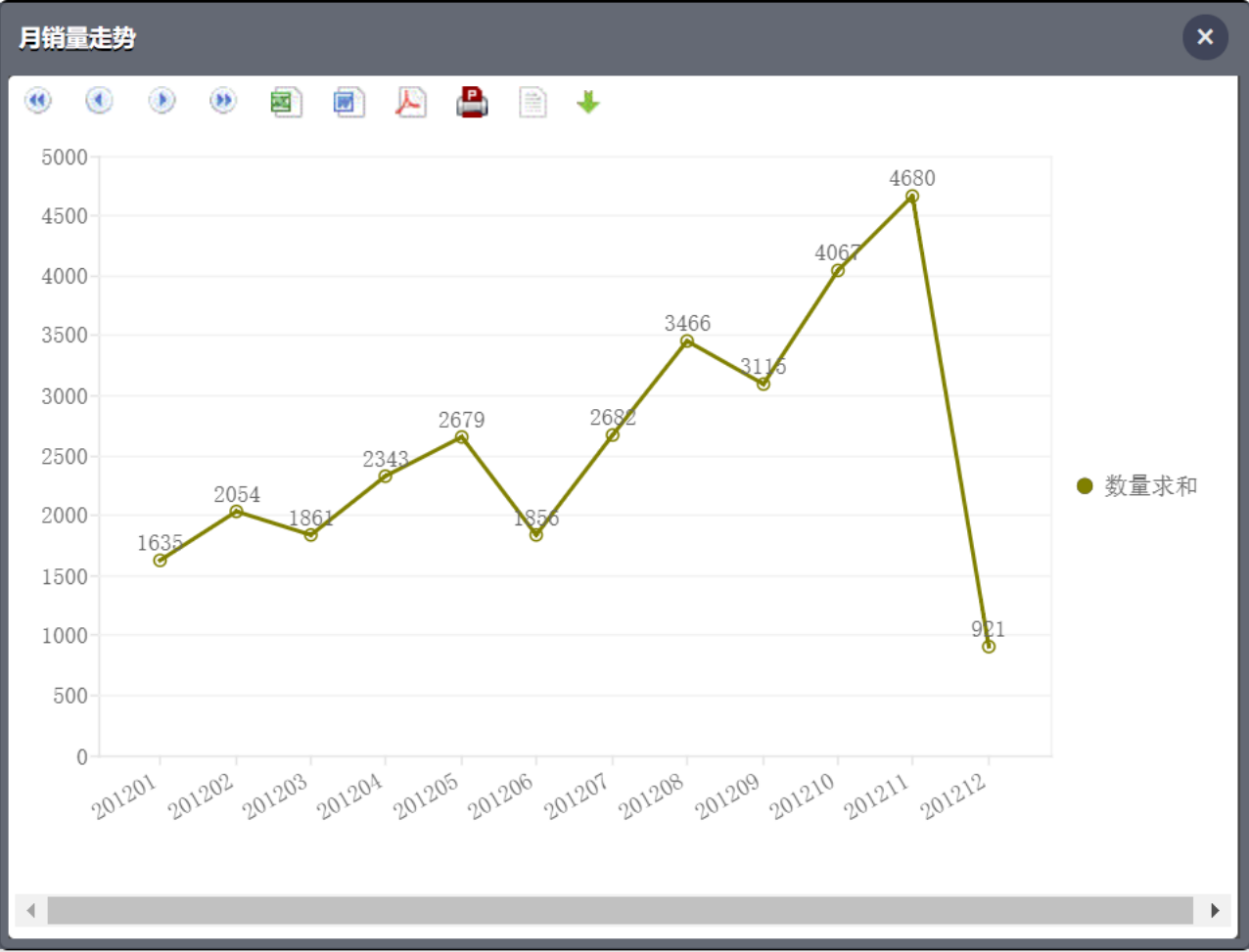
折线图

散列图

时序状态图

图形-举例

查询2012年各月销量变化趋势（折线图）



数据分析设置器

复杂条件

折线图

分类 签单月 ▼

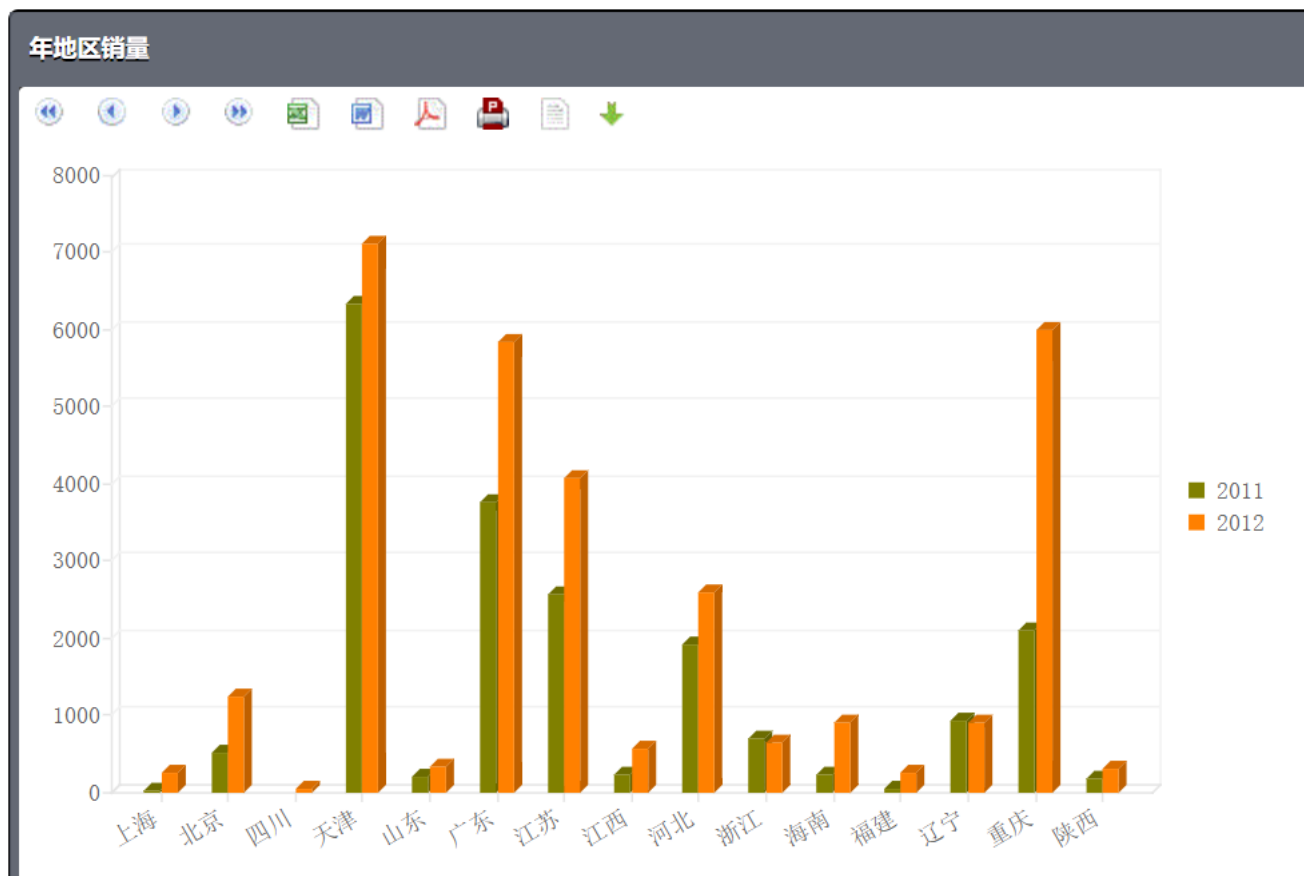
系列 数量求和 ▼

过滤月份为2012年各月
数量汇总方式选择求和

- 订单ID
- 客户ID
- 客户名称
- 雇员ID
- 姓名
- 出生日期
- 签单日期
- 签单年
- 签单月
- 产品ID
- 产品名称
- 城市编码
- 城市
- 省直辖市编码
- 省直辖市
- 单价
- 数量

图形-练习

练习：按年、地区查询销量（用三维簇状柱图表示）



小结

至此，我们已经了解了多维分析及其操作（切片、钻取、旋转等），并认识了多维分析的本质：

针对一个数据集（单表/单文件/单CUBE）的分组汇总

用SQL表达是这样：

```
select sum(m1),... from tbl where dim1=d1,... group by dim1,...
```

上面我们直接使用Excel进行分析，同样的还可以使用TXT和CSV等文本来分析。

多维分析中除了分组汇总、切片、旋转这些操作，像同比环比这些计算如何完成？

我们下节课将一起来看一下这些复杂计算是怎么回事

Summary

THANKS

感谢观看



乾学院 · VIP课堂

商业智能

多维分析技术 (跨行组计算)



上章回顾

- 前面我们基于一份Excel数据讲解了多维分析的概念和相关操作，了解到多维分析的本质是 `<select sum(m1),... from tbl where dim1=d1,... group by dim1,...>` 。
- 通过一些实例掌握了多维分析的基本操作（针对单个或多个维度的分组聚合，求和计数等），这部分计算由于只涉及到汇总值一个层次，实现起来相对简单
- 但当分析涉及更复杂的跨行组运算（如排名、同比、环比、占比等）时，只使用一个汇总层次就不够用了，这时就要指定计算的范围，比如排名要基于组内排名还是组外排名？环比是月环比还是周环比？占比是组内占比还是全部占比？ ...
- 这节课我们来介绍多维分析中的这类复杂运算

本章目标

- 了解多维分析中的复杂计算种类
- 掌握多维分析复杂计算实现
- 明确多维分析能够实现的复杂计算程度

本节数据说明

继续使用尚一章的企业销售信息Excel

维度													测度		
订单ID	客户ID	客户名称	雇员ID	姓名	签单日期	签单年	签单月	产品ID	产品名称	城市编码	城市	省直辖市编码	省直辖市	单价	数量
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	17	猪肉	30201	天津	302	天津	14	12
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	42	糙米	30201	天津	302	天津	9	10
10248	VINET	山泰企业	2	王伟	2011-02-04	2011	201102	72	酸奶酪	30201	天津	302	天津	34	5
10249	TOMSP	东帝望	9	张雪眉	2011-02-05	2011	201102	14	沙茶	20201	青岛	202	山东	18	9
10249	TOMSP	东帝望	9	张雪眉	2011-02-05	2011	201102	51	猪肉干	20201	青岛	202	山东	42	40
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	41	虾子	20401	南昌	204	江西	7	10
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	51	猪肉干	20401	南昌	204	江西	42	35
10250	HANAR	实翼	9	张雪眉	2011-02-08	2011	201102	65	海苔酱	20401	南昌	204	江西	16	15
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	22	糯米	30303	秦皇岛	303	河北	16	6
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	57	小米	30303	秦皇岛	303	河北	15	15
10251	VICTE	千固	9	张雪眉	2011-02-08	2011	201102	65	海苔酱	30303	秦皇岛	303	河北	16	20
10252	SUPRD	福星制衣厂股份有限公司	2	王伟	2011-02-09	2011	201102	20	桂花糕	30201	天津	302	天津	64	40
10252	SUPRD	福星制衣厂股份有限公司	2	王伟	2011-02-09	2011	201102	72	酸奶酪	30201	天津	302	天津	2	25

CONTENTS

本章内容

1. 排名与跨行组运算
2. 占比
3. 同比
4. 环比
5. 累积





排名与跨行组运算

一个例子

根据省市汇总销量，并按销量排名（广东河北江苏天津）

报表名称

数量求和

数量求和

省直辖市:城市		
广东	佛山	3874
	广州	2781
	汕头	2173
	深圳	9629
	珠海	1172
河北	保定	3276
	秦皇岛	2091
	石家庄	1731
	唐山	1975
	张家口	727
江苏	常州	1821
	南京	4860
天津	天津	13494

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

+

复杂条件

	上表头
省直辖市 ▼	数量求和 ▼
城市 ▼	<div><div>+</div></div>

点击添加分析指标

一个例子

根据省市汇总销量，并按销量排名（广东河北江苏天津）

添加分析指标

打开

数据文件

横向平铺

纵向平铺

重叠显示

数据集

数据集[filter]

排名

左表头根层

省直辖市

城市

排名

占比

环比

累积

同比

?1+1

?1

数量求和

查找值

计数

数量求和当前值<数量求和查找值

省直辖市当前值 省直辖市查找值 省直辖市当前序号 省直辖市查找序号 城市当前值 城市查找值 城市当前序号 城市查找序号 数量求和当前值 数量求和查找值 | 等于 大于 小于 并且 或者

+

一个例子

根据省市汇总销量，并按销量排名（广东河北江苏天津）

报表名称

省直辖市:城市

数量求和

排名

广东	佛山	3874	4
	广州	2781	6
	汕头	2173	7
	深圳	9629	2
	珠海	1172	12
河北	保定	3276	5
	秦皇岛	2091	8
	石家庄	1731	11
	唐山	1975	9
	张家口	727	13
江苏	常州	1821	10
	南京	4860	3
天津	天津	13494	1

请观察这个排名有什么问题？

这个排名是针对所有“城市”的排名；而有时我们关心的是同省份各个城市的排名，或者各省的排名

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

签单月

数量

复杂条件

上表头

指标标题

省直辖市 ▼

数量求和 ▼

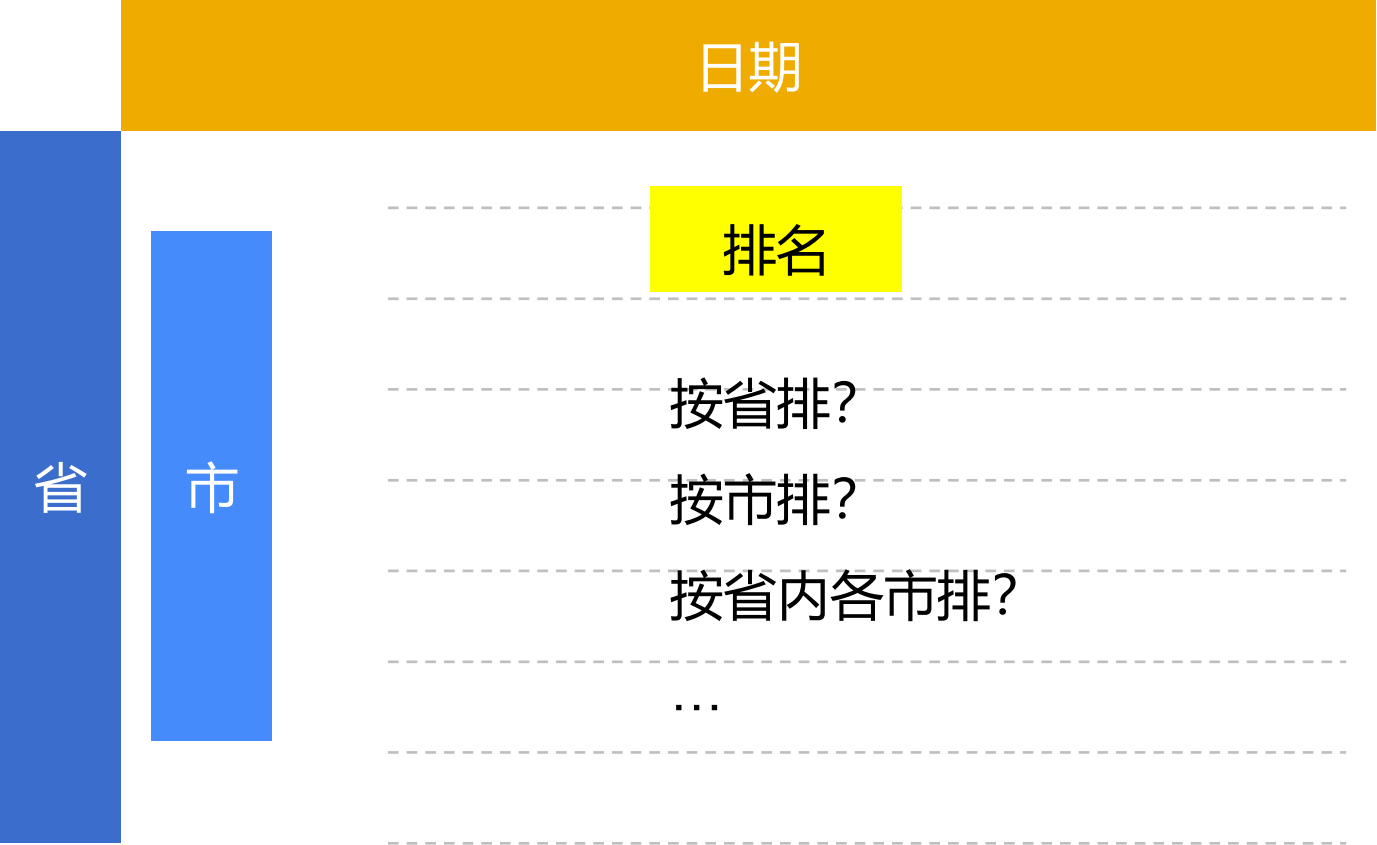
城市 ▼

排名 ▼

+

排名-知识点

排名的难点在于确定排名的层次（维度），即“按照什么排名”
在多维分析中，由于维度会以多层的方式位于行或列上，因此如何指定范围层次至关重要



省内排名

那么如何做省内各市的排名呢？

编辑

打开

数据文件

横向平铺

纵向平铺

重叠显示

数据集

数据集[1] x

省内排名

左表头根层

省直辖市

城市

当前分析设置 排名 占比 环比 累积 同比

?1+1

?1

数量求和

查找值

计数

数量求和当前值<数量求和查找值 && 省直辖市当前值 == 省直辖市查找值

省直辖市当前值 省直辖市查找值 省直辖市当前序号 省直辖市查找序号 城市当前值 城市查找值 城市当前序号 城市查找序号 数量求和当前值 数量求和查找值 | 等于 大于 小于 并且 或者

+

指定排名范围，为同一个省份【重点】

省内排名

报表名称

数据分析设置器

省直辖市:城市		数量求和	排名	省内排名
广东	佛山	3874	4	2
	广州	2781	6	3
	汕头	2173	7	4
	深圳	9629	2	1
	珠海	1172	12	5
河北	保定	3276	5	1
	秦皇岛	2091	8	2
	石家庄	1731	11	4
	唐山	1975	9	3
	张家口	727	13	5
江苏	常州	1821	10	2
	南京	4860	3	1
天津	天津	13494	1	1

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

复杂条件

上表头

指标标题

省直辖市

城市

数量求和

排名

省内排名

这里我们指定范围来做排名，这与前面的“数量求和”有很大不同

按省排名

指定范围可以做最末级（城市）的排名，如何按省排名呢？

报表名称

省直辖市:城市

省直辖市:城市	数量求和	数量求和1
广东	佛山	3874
	广州	2781
	汕头	2173
	深圳	9629
	珠海	1172
河北	保定	3276
	秦皇岛	2091
	石家庄	1731
	唐山	1975
江苏	常州	1821
	南京	4860
天津	天津	13494

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

复杂条件

上表头

指标标题

省直辖市

城市

数量求和

数量求和1

求和

计数

最大

最小

平均

值计数

升序

降序

不排序

编辑

数据过滤

显示格式

显示值表达式

删除

再拖拽一个数量求和，用于计算“省销量”

按省排名

指定范围可以做最末级（城市）的排名，如何按省排名呢？

编辑

报表名称

数据分析设置器

×

省销量

指定分组聚合

不选择任何分组则全部聚合起来；要选择靠前的分组，这样表格展示的数据才整齐易看

省直辖市

城市

选择“指定分组聚合”；
选择聚合层次为“省”

按省排名

指定范围可以做最末级（城市）的排名，如何按省排名呢？

报表名称

省直辖市:城市

数量求和

省销量

广东	佛山	3874	19629
	广州	2781	
	汕头	2173	
	深圳	9629	
	珠海	1172	
河北	保定	3276	9800
	秦皇岛	2091	
	石家庄	1731	
	唐山	1975	
江苏	常州	1821	6681
	南京	4860	
天津	天津	13494	13494

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

+

复杂条件

上表头

指标标题

省直辖市 ▾

城市 ▾

数量求和 ▾

省销量 ▾

+

按省排名

指定范围可以做最末级（城市）的排名，如何按省排名呢？



按省排名

指定范围可以做最末级（城市）的排名，如何按省排名呢？

[illegible]

知识点-跨行组运算

- 通过排名可以看到，在多维分析中排名的计算方式与常规求和计数有很大不同，除了汇总本身还要与其他行/组进行比较，比较时还需要指定范围和层次，我们把这类计算称为**跨行组运算**，类似的计算还有同比、环比、占比、累积等
- 由于报表数据行是根据数据动态扩展出来的，因此进行跨行组运算前无法指定哪些行参与运算，这就要求多维分析产品提供跨行组运算的支持

...	
201901	238
...	
202001	352
...	



与不同行比较计算
“去年同期”

排名-练习

练习：根据省市汇总销量，并添加年省内按市排名（年内&省内排名）

报表名称

省直辖市:城市		2011		2012	
		数量求和	年省内排名	数量求和	年省内排名
广东	佛山	2874	2	3874	2
	广州	781	5	2781	3
	汕头	1173	3	2173	4
	深圳	3771	1	5858	1
	珠海	1072	4	1172	5
河北	保定	2276	2	3276	1
	秦皇岛	890	3	1201	3
	石家庄	787	4	944	4
	唐山	3275	1	1975	2
	张家口	249	5	478	5
江苏	常州	716	2	1105	2
	南京	1862	1	2998	1
天津	天津	6366	1	7128	1

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

+

复杂条件

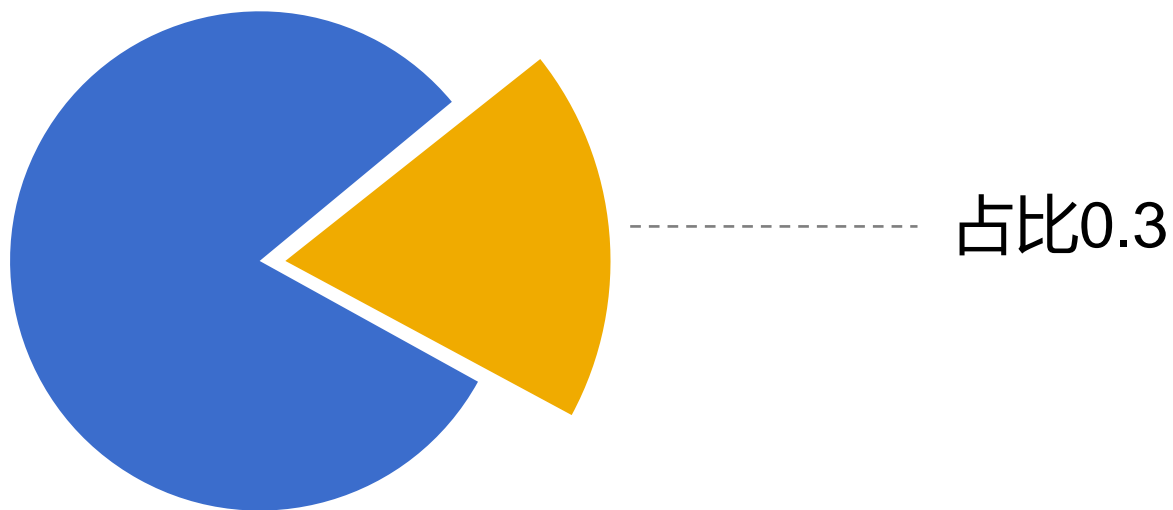
	签单年 ▼
	指标标题
省直辖市 ▼	数量求和 ▼
城市 ▼	年省内排名 ▼
	+



占比

占比-知识点

占比即当前值占总数的比例，在多维分析中计算占比的难点与排名类似，如何确定总数的范围，即“什么的总数”



与排名类似，涉及跨行组的计算都需要采用类似的方式实现，在页面端除了要汇总数据还要指定范围层次

占比

按省市汇总销量，并计算各市销量占比

报表名称

省直辖市:城市

数量求和

总计

广东	佛山	3874	3874
	广州	2781	2781
	汕头	2173	2173
	深圳	9629	9629
	珠海	1172	1172
	总计	19629	19629
河北	保定	3276	3276
	秦皇岛	2091	2091
	石家庄	1731	1731
	唐山	1975	1975
	张家口	727	727
	总计	9800	9800
江苏	常州	1821	1821
	南京	4860	4860
	总计	6681	6681

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单

签单

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

+

复杂条件

上表头

省直辖市 ▼

城市 ▼

数量求和 ▼

+

新增分析指标

占比

按省市汇总销量，并计算各市销量占比

添加分析指标

打开

数据文件

横向平铺

纵向平铺

重叠显示

×

占比

左表头根层 省直辖市 城市

排名 占比 环比 累积 同比

?1/?2

?1 数量求和 当前值

?2 数量求和 查找值 求和

注意这里参数设置，?2参数汇总方式选择求和

查找条件表达式

省直辖市当前值 省直辖市查找值 省直辖市当前序号 省直辖市查找序号 城市当前值 城市查找值 城市当前序号 城市查找序号 数量求和当前值 数量求和查找值 | 等于 大于 小于 并且 或者

+

占比

按省市汇总销量，并计算各市销量占比

报表名称

省直辖市:城市		数量求和	占比	数量求和-总计
广东	佛山	3874	0.08	3874
	广州	2781	0.06	2781
	汕头	2173	0.04	2173
	深圳	9629	0.19	9629
	珠海	1172	0.02	1172
	总计	19629		19629
河北	保定	3276	0.07	3276
	秦皇岛	2091	0.04	2091
	石家庄	1731	0.03	1731
	唐山	1975	0.04	1975
	张家口	727	0.01	727
	总计	9800		9800
江苏	常州	1821	0.04	1821
	南京	4860	0.1	4860
	总计	6681		6681

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

复杂条件

上表头

指标标题

省直辖市

城市

数量求和

占比

+

占比-练习

练习：按省市汇总销量，并计算省内各市销量占比

报表名称

省直辖市:城市		数量求和	占比	省内占比	数量求和-总计
广东	佛山	3874	0.08	0.2	3874
	广州	2781	0.06	0.14	2781
	汕头	2173	0.04	0.11	2173
	深圳	9629	0.19	0.49	9629
	珠海	1172	0.02	0.06	1172
	总计	19629			19629
河北	保定	3276	0.07	0.33	3276
	秦皇岛	2091	0.04	0.21	2091
	石家庄	1731	0.03	0.18	1731
	唐山	1975	0.04	0.2	1975
	张家口	727	0.01	0.07	727
	总计	9800			9800
江苏	常州	1821	0.04	0.27	1821
	南京	4860	0.1	0.73	4860

数据分析设置器

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

+

复杂条件

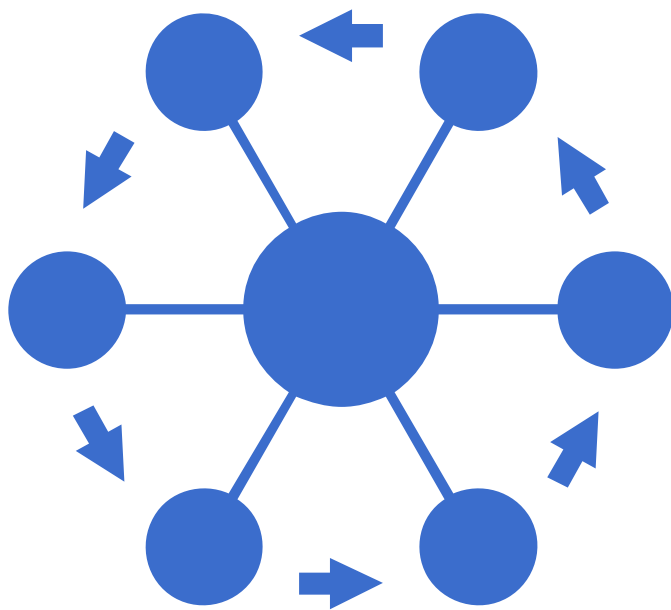
	上表头
	指标标题
省直辖市 ▾	数量求和 ▾
城市 ▾	占比 ▾
	省内占比 ▾
	+



环比

环比-知识点

- 两个周期内值（比如连续两月）的变化比，有时人们更关注环比增长率， $\text{环比增长率} = (\text{本期数} - \text{上期数}) / \text{上期数} \times 100\%$ ，用来反映连续周期内数据的变化幅度
- 在多维分析中，环比跟位置相关，以便确定“连续”，同时周期又与分析时选定的维度层次有关



环比-举例

例1：根据签单年月汇总销量，并添加每月环比

编辑

报表名称

数据分析设置器

月环比

左表头根层 签单年 签单月

当前分析设置 排名 占比 环比 累积 同比

?1/?2

?1 数量求和 当前值

🗑️ ?2 数量求和 查找值 第一个

签单月当前序号==签单月查找序号+1

签单年当前值 签单年查找值 签单年当前序号 签单年查找序号 签单月当前值 签单月查找值 签单月当前序号 签单月查找序号 数量求和当前值 数量求和查找值 | 等于 大于 小于 并且 或者

+

复杂条件

上表头

指标标题

签单年 数量求和

签单月 月环比

+

环比-举例

例1：根据签单年月汇总销量，并添加每月环比

[illegible]

环比-练习

练习：根据签单年月汇总销量，并添加**年内环比**和**环比增长率**

报表名称

签单年:签单月		数量求和	月环比	年内月环比	年内环比增长率	数量求和-总计
2011	201102	1462				1462
	201103	1322	0.9	0.9	-9.58%	1322
	201104	1124	0.85	0.85	-14.98%	1124
	201105	1738	1.55	1.55	54.63%	1738
	201106	1735	1	1	-0.17%	1735
	201107	2200	1.27	1.27	26.80%	2200
	201108	2401	1.09	1.09	9.14%	2401
	201109	2132	0.89	0.89	-11.20%	2132
	201110	1770	0.83	0.83	-16.98%	1770
	201111	1912	1.08	1.08	8.02%	1912
	201112	2164	1.13	1.13	13.18%	2164
	总计	19960				19960
	201201	1635	0.76			1635
	201202	2054	1.26	1.26	25.63%	2054
	201203	1861	0.91	0.91	-9.40%	1861
	201204	2343	1.26	1.26	25.90%	2343

提示：
环比增长率需要修改环比计算公式
指标下拉指定环比增长率显示格式



同比

同比-知识点

- 同比通常是本年和去年同期的比值，因此在多维分析中至少要选出2年数据才能实现同比
- 同比的难点在于如何找到去年同期

情况1：年（2011）、月（2）、日（10）

去年同期值：
年-1，月相等，日相等

其他情况类似，要根据数据特征找同期值

情况2：年（2011）、月（201102）、日（20110210）

按月找去年同期值：
月-100

按日找去年同期值：
日-10000

同比-举例

例1：根据签单年月汇总销量，并计算月同期比

编辑

报表名称

数据分析设置器

×

同比

左表头根层 签单年 签单月

当前分析设置 排名 占比 环比 累积 同比

?1/?2

?1 数量求和 当前值

?2 数量求和 查找值 第一个

签单月当前值 == 签单月查找值 +100

签单年当前值 签单年查找值 签单年当前序号 签单年查找序号 签单月当前值 签单月查找值 签单月当前序号 签单月查找序号 数量求和 当前值 数量求和查找值 | 等于 大于 小于 并且 或者

+

复杂条件

上表头

指标标题

签单年 数量求和

签单月 同比

+

同比-举例

例1：根据签单年月汇总销量，并计算月同期比

报表名称

签单年:签单月

签单年:签单月	数量求和	同比	数量求和-总计
2011	201102	1462	1462
	201103	1322	1322
	201104	1124	1124
	201105	1738	1738
	201106	1735	1735
	201107	2200	2200
	201108	2401	2401
	201109	2132	2132
	201110	1770	1770
	201111	1912	1912
	201112	2164	2164
	总计	19960	
2012	201201	1635	
	201202	2054	1.4
	201203	1861	1.41
	201204	2343	2.08
	201205	2679	1.54
	201206	1856	1.07
	201207	2682	1.22
	201208	3466	1.44
	201209	3115	1.46
	201210	4067	2.3
	201211	4680	2.45
	201212	921	0.43

数据分析设置器

报表列表

拖拽图形

报表名称

添加报表

订单ID

客户ID

客户名称

雇员ID

姓名

出生日期

签单日期

签单年

签单月

产品ID

产品名称

城市编码

城市

省直辖市编码

省直辖市

单价

数量

库存量

复杂条件

	上表头
	指标标题
签单年	数量求和
签单月	同比

与去年同月份比值，其中去年一月没有数据

练习与思考

练习：请使用记录序号（当前序号和查找序号）计算上例中的同期比

报表名称				
签单年:签单月		数量求和	同比	同比-序号
2011	201102	1462		
	201103	1322		
	201104	1124		
	201105	1738		
	201106	1735		
	201107	2200		
	201108	2401		
	201109	2132		
	201110	1770		
	201111	1912		
	201112	2164		
	总计	19960		
2012	201201	1635		
	201202	2054	1.4	1.4
	201203	1861	1.41	1.41
	201204	2343	2.08	2.08
	201205	2679	1.54	1.54
	201206	1856	1.07	1.07
	201207	2682	1.22	1.22
	201208	3466	1.44	1.44
	201209	3115	1.46	1.46
	201210	4067	2.3	2.3
	201211	4680	2.45	2.45
	201212	921	0.43	0.43
	总计	31359		

请思考：

- 1. 使用序号计算同比，需要满足什么条件？
- 2. 序号的实现方式可能存在问题？

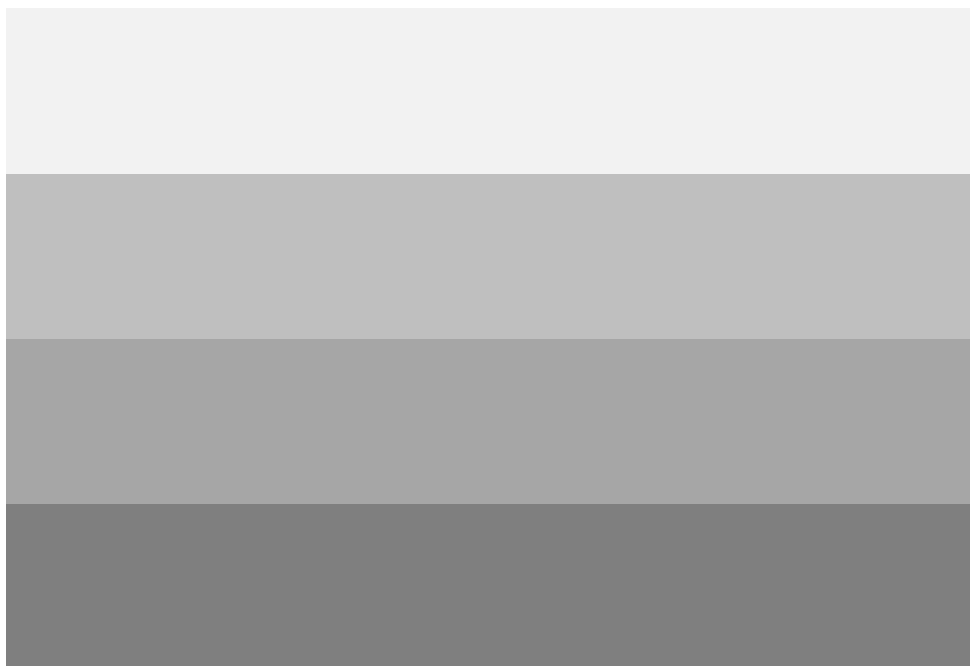


累积

累积-知识点

累积即下一项与上一项的和。

计算累积通常借助位置来完成，因此数据要事先排序，如计算月累积，数据要先按月升序



累积-举例

编辑

报表名称

数据分析设置器

×

累积

左表头根层 签单年 签单月

当前分析设置 排名 占比 环比 累积 同比

?1

?1 数量求和 查找值 求和

签单月当前序号>=签单月查找序号

签单年当前值 签单年查找值 签单年当前序号 签单年查找序号 签单月当前值 签单月查找值 签单月当前序号 签单月查找序号 数量求和

并且 或者

+

注意：比较符是>=

复杂条件

上表头

指标标题

签单年 数量求和

签单月 累积

+

累积-举例

报表名称

↶

↷

↶

↷

📄

📄

📄

📄

📄

⬇️

签单年:签单月		数量求和	累积	数量求和-总计
2011	201102	1462	1462	1462
	201103	1322	2784	1322
	201104	1124	3908	1124
	201105	1738	5646	1738
	201106	1735	7381	1735
	201107	2200	9581	2200
	201108	2401	11982	2401
	201109	2132	14114	2132
	201110	1770	15884	1770
	201111	1912	17796	1912
	201112	2164	19960	2164
总计		19960		19960
2012	201201	1635	21595	1635
	201202	2054	23649	2054
	201203	1861	25510	1861
	201204	2343	27853	2343
	201205	2679	30532	2679
	201206	1856	32388	1856
	201207	2682	35070	2682
	201208	3466	38536	3466
	201209	3115	41651	3115
	201210	4067	45718	4067
	201211	4680	50398	4680
	201212	921	51319	921

数据分析设置器

✚️

订单ID

✚️

客户ID

✚️

客户名称

✚️

雇员ID

✚️

姓名

✚️

出生日期

✚️

签单日期

✚️

✚️

签单月

✚️

✚️

产品名称

✚️

城市编码

✚️

✚️

省直辖市编码

✚️

✚️

单价

✚️

✚️

库存量

复杂条件

	上表头
	指标标题
签单年 ▼	数量求和 ▼
签单月 ▼	累积 ▼
	<div>+</div>

累积-练习

练习：根据例子，求偶数月的累积

报表名称					
签单年:签单月		数量求和	累积	偶数月累积	数量求和-总计
2011	201102	1462	1462	1462	1462
	201103	1322	2784	0	1322
	201104	1124	3908	2586	1124
	201105	1738	5646	0	1738
	201106	1735	7381	4321	1735
	201107	2200	9581	0	2200
	201108	2401	11982	6722	2401
	201109	2132	14114	0	2132
	201110	1770	15884	8492	1770
	201111	1912	17796	0	1912
	201112	2164	19960	10656	2164
	总计	19960			19960
2012	201201	1635	21595	0	1635
	201202	2054	23649	12710	2054
	201203	1861	25510	0	1861
	201204	2343	27853	15053	2343
	201205	2679	30532	0	2679
	201206	1856	32388	16909	1856
	201207	2682	35070	0	2682
	201208	3466	38536	20375	3466
	201209	3115	41651	0	3115
	201210	4067	45718	24442	4067
	201211	4680	50398	0	4680
	201212	921	51319	25363	921

小结

- ✓ 通过本节可以看到，多维分析能够实现一些复杂的跨行组运算，包括排名、同比、环比、占比、累积等等，**跨行组运算与普通的求和计数不同，选择合适的范围层次很关键**
- ✓ 另外也要注意，多维分析实现的“复杂”程度仍然有限，对于表格的格式也有要求（相对规整）；如果遇到更复杂的分析场景可以先用多维分析“试试”能否满足要求
- 前两节以Excel为例（同样适用TXT和CSV等文本）认识了多维分析的基本操作和复杂一些的跨行组运算。而实际业务中多维分析的数据源大多是DB，如何基于DB来分析？后面的课程我们将继续讲解。

Summary

THANKS

感谢观看



乾学院 · VIP课堂

商业智能 多维分析技术 (数据库分析)



上章回顾

- 前两节课我们通过Excel数据源快速认识了多维分析及多维分析的本质
- 上一章学习了在多维分析中如何实现同比环比等跨行组运算，并“感受到”这类运算不同于求和计数，需要指定范围和层次才能正确实现
- 本章来学习如何使用更加广泛的数据库进行多维分析

本章环境准备


1. 启动MySQL数据库
2. 将 “service.xml” 拷贝到 [\[润乾报表安装目录\]\raqsoft\report\services\datalogic](#) 下
3. 启动DQL元数据层 IDE
4. 菜单栏-服务-启动DQL服务器，点击 “启动”

本章数据说明

销售数据存储在sale_data表（与上章使用的Excel数据一致）中，这个表是根据业务表加工出来的一张宽表，数据库中还存储相应的代码表

维度

测度



ORDER_ID	CUSTOMER_ID	EMPLOYEE_ID	BIRTHDAY	ORDER_DATE	ORDER_YEAR	ORDER_MONTH	PRODUCT_ID	CITY_ID	PROVINCE_ID	UNIT_PRICE	AMOUNT
10248	VINET	2	1972-2-19	2011-2-4	2011	201102	17	30201	302	14	12
10248	VINET	2	1972-2-19	2011-2-4	2011	201102	42	30201	302	9	10
10248	VINET	2	1972-2-19	2011-2-4	2011	201102	72	30201	302	34	5
10249	TOMSP	9	1979-7-2	2011-2-5	2011	201102	14	20201	202	18	9
10249	TOMSP	9	1979-7-2	2011-2-5	2011	201102	51	20201	202	42	40
10250	HANAR	9	1979-7-2	2011-2-8	2011	201102	41	20401	204	7	10
10250	HANAR	9	1979-7-2	2011-2-8	2011	201102	51	20401	204	42	35
10250	HANAR	9	1979-7-2	2011-2-8	2011	201102	65	20401	204	16	15
10251	VICTE	9	1979-7-2	2011-2-8	2011	201102	22	30303	303	16	6
10251	VICTE	9	1979-7-2	2011-2-8	2011	201102	57	30303	303	15	15
10251	VICTE	9	1979-7-2	2011-2-8	2011	201102	65	30303	303	16	20
10252	SUPRD	2	1972-2-19	2011-2-9	2011	201102	20	30201	302	64	40
10252	SUPRD	2	1972-2-19	2011-2-9	2011	201102	33	30201	302	2	25
10252	SUPRD	2	1972-2-19	2011-2-9	2011	201102	60	30201	302	27	40
10253	HANAR	9	1979-7-2	2011-2-10	2011	201102	31	20401	204	10	20
10253	HANAR	9	1979-7-2	2011-2-10	2011	201102	39	20401	204	14	42
10253	HANAR	9	1979-7-2	2011-2-10	2011	201102	49	20401	204	16	40
10254	CHOPS	2	1972-2-19	2011-2-11	2011	201102	24	30201	302	3	15
10254	CHOPS	2	1972-2-19	2011-2-11	2011	201102	55	30201	302	19	21

本章目标

- 了解DB数据源时多维分析要解决哪些问题
- 掌握直接基于SQL完成多维分析
- 掌握语义层相关概念与实现

CONTENTS

本章内容

1. SQL分析
2. 语义层





SQL分析

知识点-宽表

- 在OLAP业务中，有时会把多个维度及其属性和测度放到一张表中，形成列数较多的“宽表”
- 宽表往往围绕业务建立，其目标是希望通过一个宽表就能完成一类业务查询分析
- 宽表的优点：查询效率高，查询是不需要再关联其他表
- 宽表的缺点：冗余度太高，会带来各种问题（后面的课程还会详细分析）
- 宽表实际是一种用空间换时间的技术手段

[illegible]

SQL分析

- 基于数据库分析最简单的情况是给一句SQL就能分析
- 这样可以满足临时查询分析的需要，也可以让有SQL基础的业务人员自行组织数据
- 多维分析产品一般都应支持基于SQL的直接分析。
- 下面以润乾报表为例来看一下



给一句SQL
就能做多维分析

数据源配置

打开 [润乾报表安装目录] \raqsoft\report\web\webapps\demo\WEB-INF\raqsoftConfig.xml

在DBList节点下添加mybi数据源供WEB使用

```
<DB name="mybi">
  <property name="url" value="jdbc:mysql://127.0.0.1:3306/mybi?useCursorFetch=true"/>
  <property name="driver" value="com.mysql.jdbc.Driver"/>
  <property name="type" value="10"/>
  <property name="user" value="root"/>
  <property name="password" value="root"/>
  <property name="batchSize" value="1000"/>
  <property name="autoConnect" value="false"/>
  <property name="useSchema" value="false"/>
  <property name="addTilde" value="false"/>
  <property name="dbCharset" value="GBK"/>
  <property name="clientCharset" value="GBK"/>
  <property name="needTransContent" value="false"/>
  <property name="needTransSentence" value="false"/>
  <property name="caseSentence" value="false"/>
</DB>
```

SQL分析-举例

在润乾报表中可以通过页面端数据集编辑SQL查询数据

数据 数据集[dbQuery]已加载完毕, 共 2157 行数据

1. 点击设置数据集

管理数据集 数据分析设置器

fileQuery dbQuery 保存

2. 添加数据集

DQL表 数据库查询 (DQL/SQL) DFX文件 DFX脚本 填报文件

3. 选择新配置的mybi数据源

mybi

4. 输入查询SQL

select * from sale_data

5. 点击将数据以TXT缓存到服务器

q1583803540692.txt 查询数据，缓存入文件

SQL分析-举例

根据dbQuery数据集添加报表进行查询

数据分析设置器

报表列表

请添加报表

添加报表

dbQuery

订单查询

☐ 使用报表模板

选择模板...

确定

订单查询

ORDER_YEAR; ORDER_MONTH		AMOUNT求和	总计
2011	201102	1462	1462
	201103	1322	1322
	201104	1124	1124
	201105	1738	1738
	201106	1735	1735
	201107	2200	2200
	201108	2401	2401
	201109	2132	2132
	201110	1770	1770
	201111	1912	1912
	201112	2164	2164
	总计	19960	19960
	201201	1635	1635
	201202	2054	2054

数据分析设置器

- ORDER_ID
- CUSTOMER_ID
- CUSTOMER_NAME
- EMPLOYEE_ID
- EMPLOYEE_NAME
- BIRTHDAY
- ORDER_DATE
- ORDER_YEAR
- ORDER_MONTH
- PRODUCT_ID
- PRODUCT_NAME
- CITY_ID
- CITY_NAME
- PROVINCE_ID
- PROVINCE_NAME
- UNIT_PRICE
- AMOUNT

复杂条件

上表头

ORDER_YEAR ▼

ORDER_MONTH ▼

AMOUNT求和 ▼

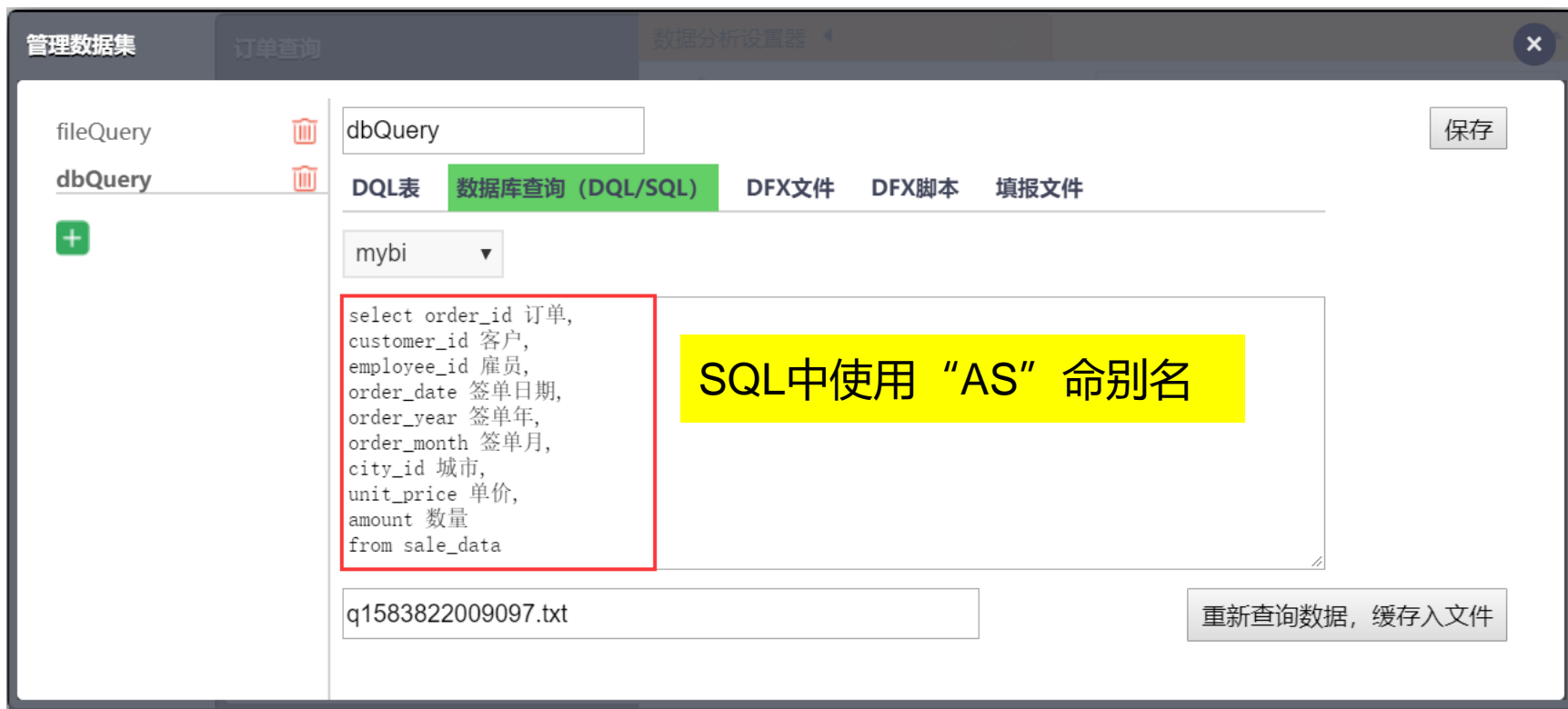
+

拖拽维度指标进行查询，按年月汇总销量

这样通过一句SQL就可以做多维分析了！但维度和指标是英文的，使用起来并不友好，怎么办？

SQL分析-举例

请思考：如何将维度和指标转换成易识别的业务描述？



SQL分析-举例

请思考：如何将维度和指标转换成易识别的业务描述？

订单查询

8

总计	4860	4860
20101	315	315
20201	573	573
20401	83	
20501	13	
30301	72	
30302		
30303	209	
40201	1176	1176
50101	523	523
60101	59	59
总计	9417	9417
总计	51319	51319

9

数据分析设置器

订单

客户

雇员

签单日期

签单年

签单月

城市

单价

数量

复杂条件

上表头

雇员 ▼

城市 ▼

数量求和 ▼

+

人员和地区显示的是编码，怎么显示中文名称呢？

按销售和城市汇总销量

通过SQL “AS 别名”搞定了中文显示的问题；但人员和地区显示的是编码而非中文名称

代码表

数据库存储了城市和雇员代码表，中文名称可以从代码表中读取

城市表(city)	
city_id	
city_name	

city_id	city_name
20501	温州
30101	北京
30201	天津
30301	张家口
30302	石家庄
30303	秦皇岛
40101	深圳

雇员表(employee)	
emp_id	
name	

emp_id	emp_name
1	张颖
2	王伟
3	李芳
4	郑建杰
5	赵军
6	孙林
7	金士鹏
8	刘英玫

SQL分析-举例

请思考：（根据前面给出的数据结构）如何显示人员姓名而非ID？

The screenshot shows a database management interface with a sidebar on the left containing 'fileQuery', 'dbQuery', and 'dbQuery2'. The main area has a tabbed interface with 'DQL表', '数据库查询 (DQL/SQL)', 'DFX文件', 'DFX脚本', and '填报文件'. The '数据库查询 (DQL/SQL)' tab is active, showing a SQL query in a text area. The query is:

```
select order_id 订单, customer_id 客户, sd.employee_id 雇员编号, e.name 雇员姓名, order_date  
签单日期, order_year 签单年, order_month 签单月, city_id 城市编号, c.city_name 城市名称,  
unit_price 单价, amount 数量  
from sale_data sd  
  
left join city c on c.city_id=sd.city_id  
left join employee e on e.emp_id=employee_id
```

The join conditions are highlighted with a red box. A yellow box with the text '关联代码表，查询名称' is overlaid on the right side of the query. Below the query, there is a text input field containing 'q1583832798412.txt' and a button labeled '查询数据，缓存入文件'. A '保存' button is also visible in the top right corner of the main area.

SQL分析-举例

请思考：（根据前面给出的数据结构）如何显示人员姓名而非ID？

订单查询2

王伟	天津	13494	13494
	总计	13494	13494
张雪眉	成都	59	59
	海口	1176	1176
	南昌	839	839
	秦皇岛	2091	2091
	青岛	573	573
	上海	315	315
	石家庄	1731	1731
	温州	1383	1383
	西安	523	523
	张家口	727	727
	总计	9417	9417

数据分析设置器

订单

客户

雇员编号

雇员姓名

签单日期

签单年

签单月

城市编号

城市名称

单价

数量

+

复杂条件

上表头

雇员姓名 ▼

数量求和 ▼

城市名称 ▼

+

通过SQL关联代码表把中文名读出来，使用中文名称进行分组汇总

练习

练习：实现基于页面SQL的多维分析，并按雇员和城市汇总销量

订单查询2

王伟	天津	13494	13494
	总计	13494	13494
张雪眉	成都	59	59
	海口	1176	1176
	南昌	839	839
	秦皇岛	2091	2091
	青岛	573	573
	上海	315	315
	石家庄	1731	1731
	温州	1383	1383
	西安	523	523
	张家口	727	727
	总计	9417	9417

数据分析设置器

订单

客户

雇员编号

雇员姓名

签单日期

签单年

签单月

城市编号

城市名称

单价

数量

+

复杂条件

上表头

雇员姓名 ▼

数量求和 ▼

城市名称 ▼

+

思考

请思考：页面SQL分析存在什么问题？

□ 复用性差

对于经常要使用的表每次都写SQL分析很麻烦，需要复用

□ 准确性和显示问题

使用name做分组可能存在name重名而导致统计错误，比如两个人都叫“王刚”；

使用ID和name同时来分组报表中往往又不需要看到ID，存在显示问题

□ 排序问题

有时需要按照ID排序，只根据name查询就不够用了，比如按城市编码排序显示而非名称

□ 切片太麻烦

基于维度（如城市）切片时需要手动输入条件，十分麻烦。

与之相对应，如果能下拉选择就会好很多...



语义层

知识点-语义层

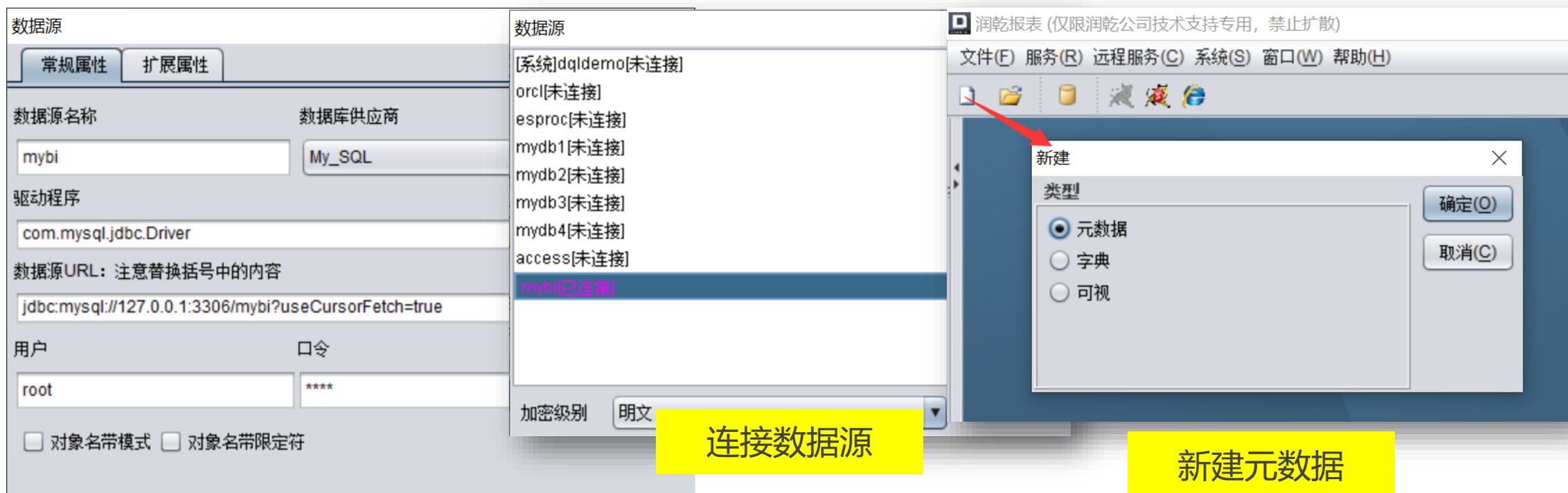
- 为了解决上述问题，可以在多维分析中构建语义层，将数据技术描述转换成业务描述
- 这要求多维分析产品还应具备语义层的能力



语义层-举例

以润乾报表为例，来看一下语义层的做法

首先，使用DQL元数据层IDE（语义层编辑器），建立数据库连接，新建元数据

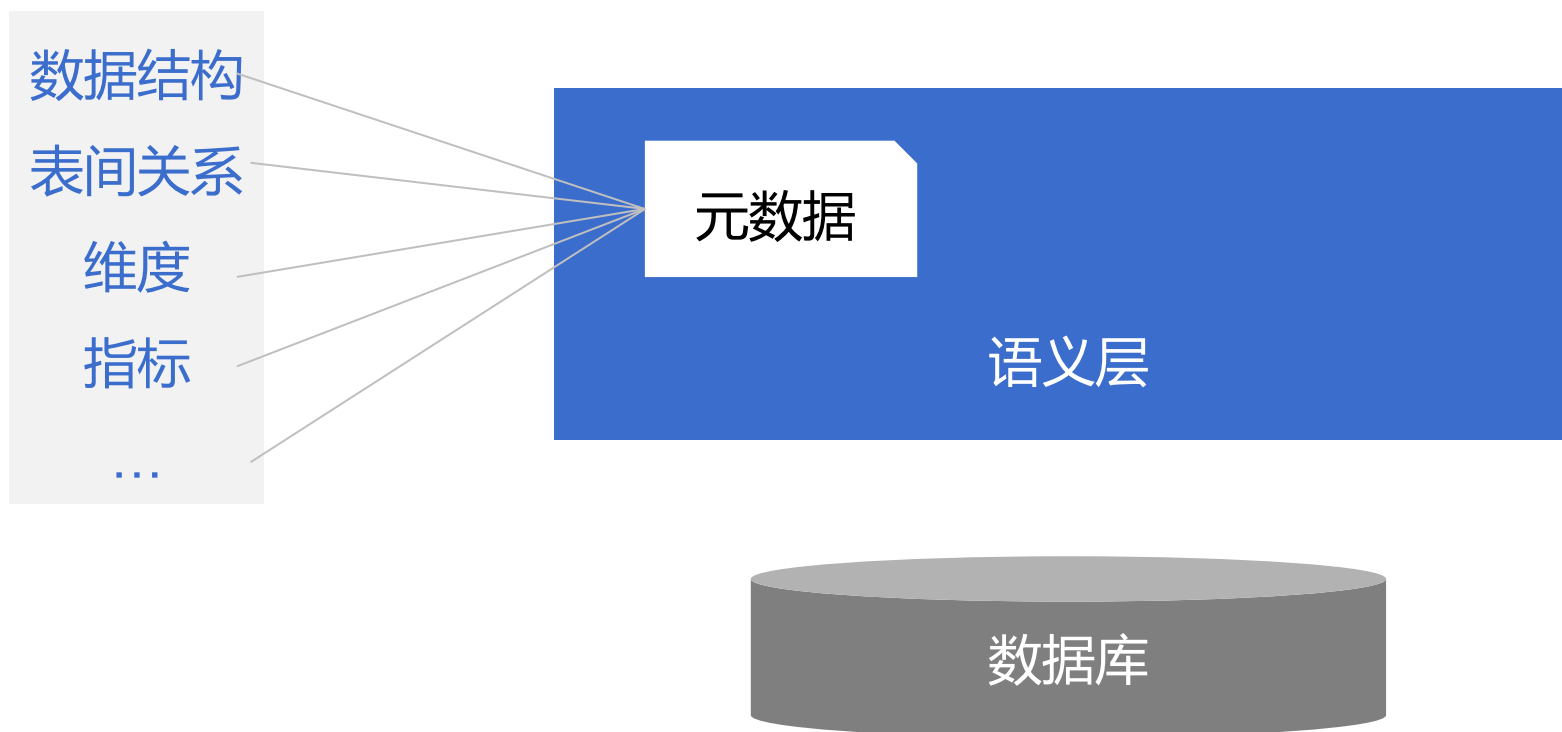


系统-数据源-新建数据源，
建立数据库连接-mybi

元数据参照例子：6.sale_data.lmd
字典参照例子：6.sale_data.dct

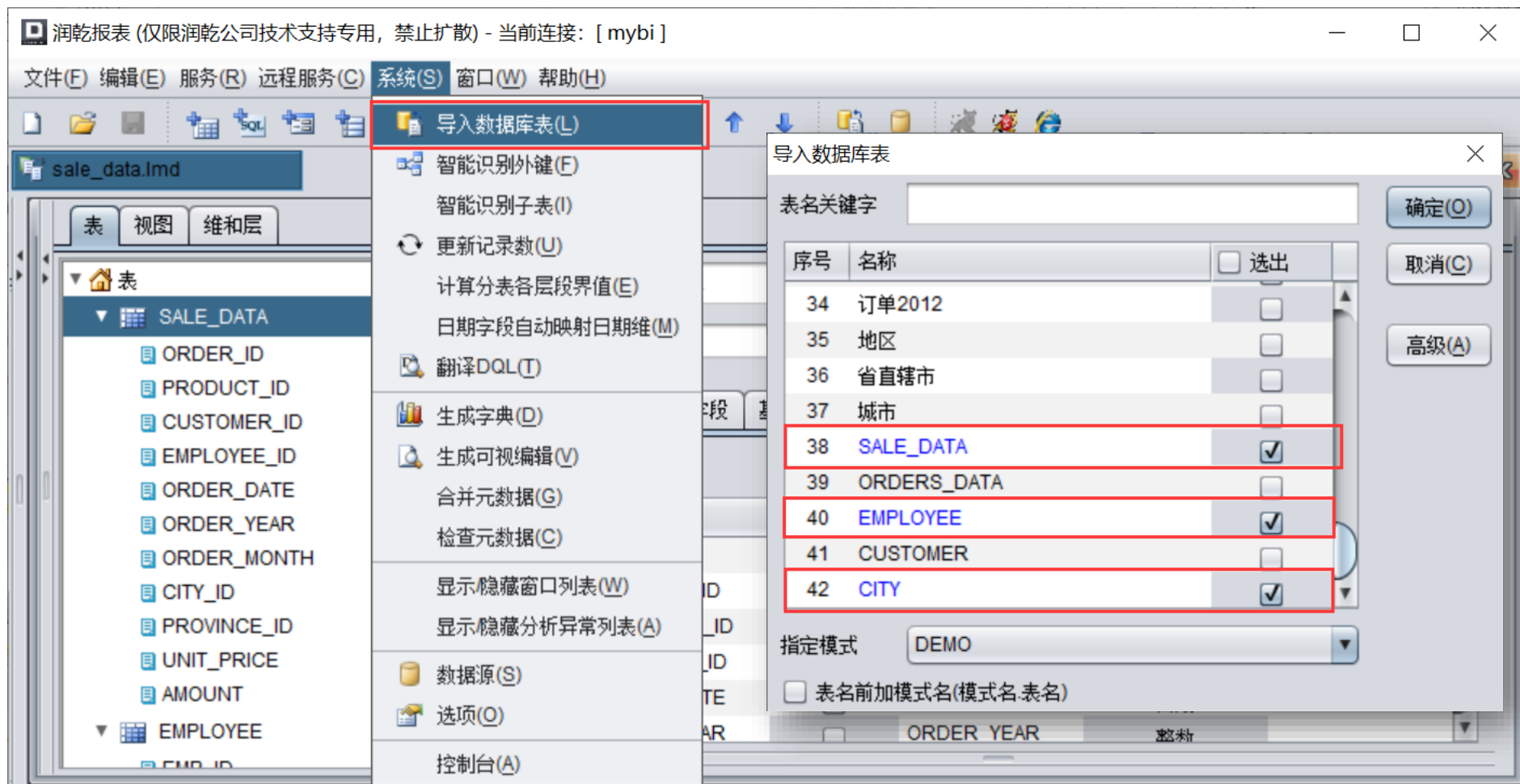
知识点-元数据

- 用来描述数据的数据，我们会在元数据中描述表数据结构、表间关系，指定多维分析使用的维度、层次，预置计算指标等



语义层-元数据

导入销售数据表、地区表和雇员表



语义层-元数据

分别设置三个表的主键

表名: SALE_DATA 物理表名: SALE_DATA
记录数: 0 表描述:

常规 外键 伪字段 基础表 汇总表 子表

字段 输入字段名关键字进行查找:

序号	字段名	主键	数据来源	数据类型	字段描述
1	ORDER_ID	<input checked="" type="checkbox"/>	ORDER_ID	整数	
2	PRODUCT_ID	<input checked="" type="checkbox"/>	PRODUCT_ID	整数	
3	CUSTOMER_ID	<input type="checkbox"/>	CUSTOMER_ID	字符串	
4				整数	
5				日期	
6	ORDER_YEAR	<input type="checkbox"/>	ORDER_YEAR	整数	

勾选主键

表名: EMPLOYEE 物理表名: EMPLOYEE
记录数: 0 表描述:

常规 外键 伪字段 基础表 汇总表 子表

字段 输入字段名关键字进行查找:

序号	字段名	主键	数据来源	数据类型	字段描述
1	EMP_ID	<input checked="" type="checkbox"/>	EMP_ID	整数	
2	NAME	<input type="checkbox"/>	NAME	字符串	

删除多余字段，设置主键

表名: CITY 物理表名: CITY
记录数: 0 表描述:

常规 外键 伪字段 基础表 汇总表 子表

字段 输入字段名关键字进行查找:

序号	字段名	主键	数据来源	数据类型	字段描述
1	CITY_ID	<input checked="" type="checkbox"/>	CITY_ID	整数	
2	CITY_NAME	<input type="checkbox"/>	CITY_NAME	字符串	

指定维度

我们要基于城市和雇员维度进行分析，需要指定这两个维度

润乾报表 (仅限润乾公司技术支持专用, 禁止扩散) [D:\tools\销售工具箱\12课程\程序文件\sale_data.lmd]

文件(F) 编辑(E) 服务(R) 远程服务(C) 系统(S) 窗口(W) 帮助(H)

sale_data.dct sale_data.lmd

添加外键

表 视图 维和层

SALE_DATA

- ORDER_ID
- PRODUCT_ID
- CUSTOMER_ID
- EMPLOYEE_ID
- ORDER_DATE
- ORDER_YEAR
- ORDER_MONTH
- CITY_ID
- PROVINCE_ID
- UNIT_PRICE
- AMOUNT

EMPLOYEE

- EMP_ID

表名: SALE_DATA 物理表名: SALE_DATA

记录数: 0 表描述:

常规 外键 伪字段 基础表 汇总表 子表

外键

序号	外键名称	对应表名
1	fk1	EMPLOYEE
2	fk2	CITY

外键映射

序号	外键字段	引用字段
1	EMPLOYEE_ID	EMP_ID

指定维度

我们要基于城市和雇员维度进行分析，需要指定这两个维度

表名	SALE_DATA	物理表名	SALE_DATA
记录数	0	表描述	
<div>常规 外键 伪字段 基础表 汇总表 子表</div>			
外键			
序号	外键名称	对应表名	
1	fk1	EMPLOYEE	
2	fk2	CITY	
外键映射			
序号	外键字段	引用字段	
1	CITY_ID	CITY_ID	

表 视图 维和层

维

序号	维名	源字段名	时间维类型
1	SALE_DATA_ORDER_ID	SALE_DATA.ORDER_ID	缺省
2	SALE_DATA_PRODUCT_ID	SALE_DATA.PRODUCT_ID	缺省
3	EMPLOYEE_EMP_ID	EMPLOYEE.EMP_ID	缺省
4	CITY_CITY_ID	CITY.CITY_ID	缺省

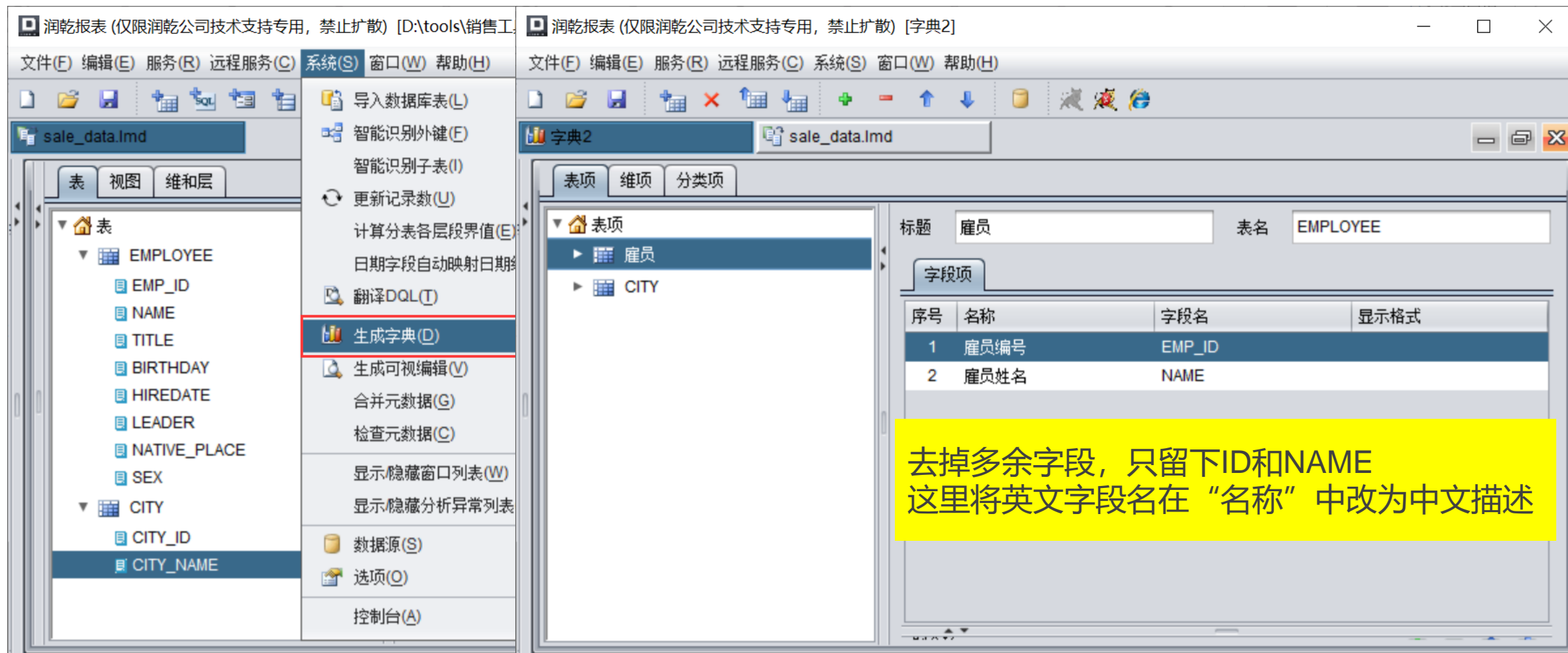
维视图

关于维度

- 在第一节课中，我们提到了维度的概念：观察数据的角度，用SQL理解就是group by 后面的分组字段。常见维度有：地区、日期、人员、产品等等
 - 在语义层中要指定好维度，这样才能正常进行多维分析
-
- 通常多维分析产品仅指定哪几个字段是维度即可，但在润乾报表中，维度需要对应到表（维表），如果不存在对应的表还可以使用虚拟表（假表）。
 - 这样定义的优势会在后面关联分析的课程中讲到！

语义转换

根据元数据生成字典，进行语义转换、显示设置等



The screenshot displays the 'Run Qian Report' software interface, specifically the 'Generate Dictionary' step. The left pane shows a tree view of tables (EMPLOYEE, CITY) and their fields. The right pane shows the 'Dictionary 2' window with a table of fields to be converted.

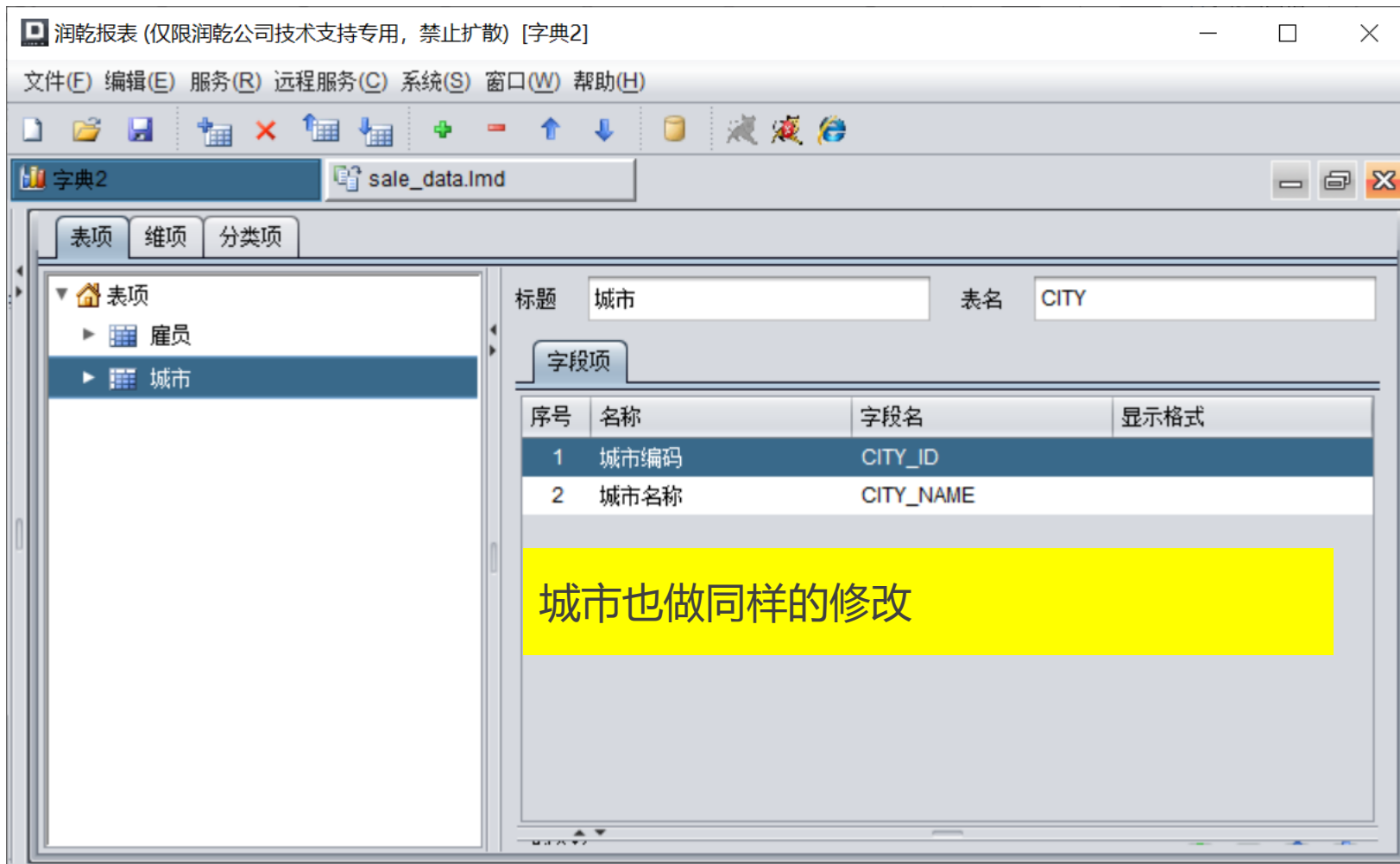
Dictionary 2 Table:

序号	名称	字段名	显示格式
1	雇员编号	EMP_ID	
2	雇员姓名	NAME	

去掉多余字段，只留下ID和NAME
这里将英文字段名在“名称”中改为中文描述

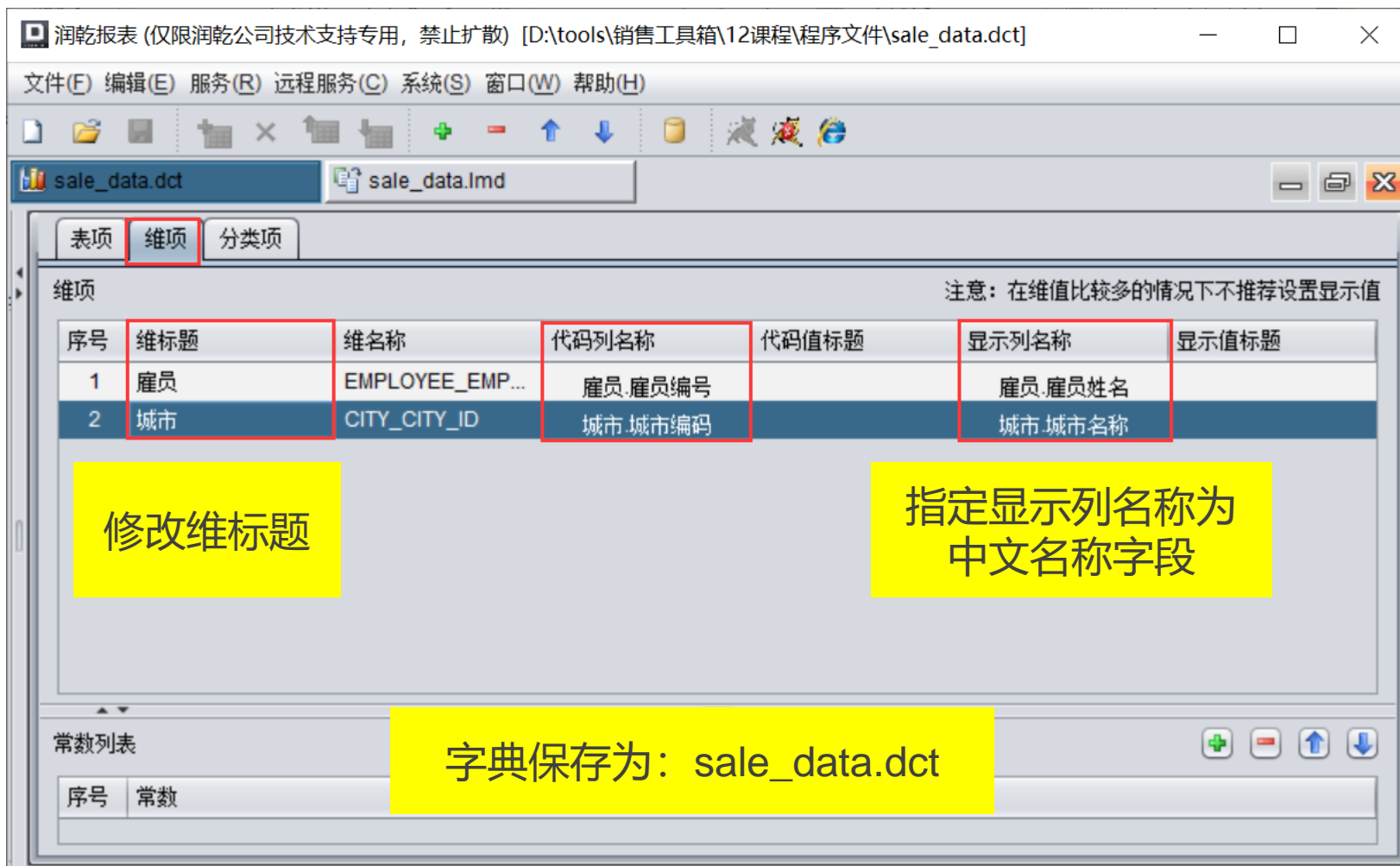
语义转换

根据元数据生成字典，进行语义转换、显示设置等



语义转换

根据元数据生成字典，进行语义转换、显示设置等

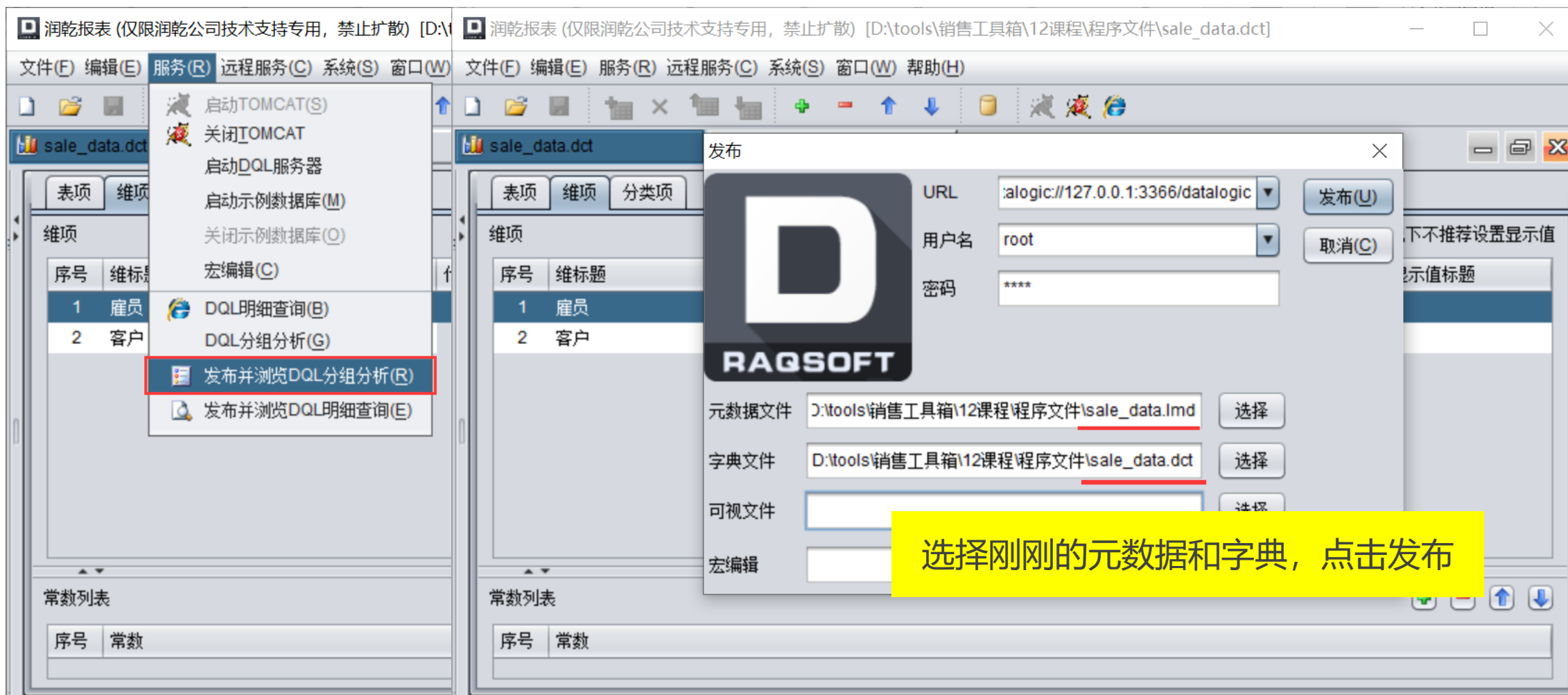


知识点-显示值

- 在多维分析中经常会涉及真实值和显示值，真实值是实际参与计算的数据，而显示值是为了业务需要而显示出来的数据，如最常见的编码和名称，计算的时候使用编码，显示的是名称
- 为什么不用显示值计算？
- 因为真实值（编码）通常是唯一的，使用真实值计算可以保证计算的准确性，而显示值（名称）则可能出现重复，比如两个同名的人，使用名称进行计算（如分组）则会出现计算错误

发布语义层

发布语义层，生成显示值配置



页面分析

多维分析时直接使用我们指定的维度（雇员和城市）进行分析

报表名称

金士鹏

深圳

9629

9629

刘英玫

南京

48

48

张雪眉

上海

31

31

青岛

51

51

南昌

839

839

温州

1383

1383

张家口

727

727

石家庄

1731

1731

秦皇岛

2091

2091

海口

1176

1176

西安

523

523

成都

59

59

总计

9417

9417

总计

51319

51319

编码也直接显示成了“显示值”中文

数据分析设置器

销售数据

订单

产品

客户

雇员

签单日期

签单年

签单月

城市

省份

单价

数量

雇员

雇员

姓名

城市

城市

名称

上表头

雇员 ▼

城市 ▼

数量求和 ▼

+

直接使用语义模型进行分析，同时使用业务汉语描述

页面分析

多维分析时直接使用我们指定的维度（雇员和城市）进行分析

报表名称

雇员:城市

数量求和

总计

张雪眉	上海	315	315
	青岛	573	573
	南昌	839	839
	温州	1383	1383
	张家口	727	727
	石家庄	1731	1731
	秦皇岛	2091	2091
	海口	1176	1176
	西安	523	523
	成都	59	59
	总计	9417	9417
刘英玫	南京	4860	4860
	总计	4860	4860
金士鹏	深圳	9629	9629
	总计	9629	9629

数据分析设置器

销售数据

订单

产品

客户

雇员

签单日期

签单年

签单月

城市

省份

单价

数量

雇员

雇员

姓名

城市

城市

名称

上表头

雇员

城市

数量求和

升序

降序

不排序

编辑

数据过滤

维源字段

显示格式

显示值

显示值表达式

删除

按ID排序,
显示NAME

EMP_ID	NAME
1	张颖
2	王伟
3	李芳
4	郑建杰
5	赵军
6	孙林
7	金士鹏
8	刘英玫
9	张雪眉

切片

基于语义层的多维分析，切片会变得简单

设置查询条件

报表名称

数据分析设置器

×

☐ 简洁显示 (简洁显示会隐藏掉“新增条件、调整条件层次”的功能, 更能凸显条件之间的层次关系)

→ 城市

等于

10101,10102,20101,20201

×

AND | OR

☒ 大连

☒ 沈阳

☒ 上海

☒ 青岛

☐ 济南

☐ 南京

☐ 常州

☐ 南昌

☐ 温州

☐ 北京

☐ 天津

☐ 张家口

☐ 石家庄

☐ 秦皇岛

☐ 深圳

☐ 广州

按城市维度切片
下拉选择

保存

清除条件并退出

取消

切片

基于语义层的多维分析，切片会变得简单

报表名称

雇员:城市

数量求和

总计

张雪眉	上海	315	315
	青岛	573	573
	总计	888	888
郑建杰	大连	1880	1880
	总计	1880	1880
总计		2768	2768

数据分析设置器

销售数据

订单

产品

客户

雇员

签单日期

签单年

签单月

城市

省份

单价

数量

雇员

雇员

姓名

城市

城市

名称

雇员 ▼

城市 ▼

上表头

数量求和 ▼

+

练习

练习：实现上面语义层的开发过程，并查询2012年南京和上海的销量

sale_data

↶

↷

↶

↷

📄

📄

📄

📄

📄

📄

📄

签单年:城市		数量求和	总计
2012	上海	281	281
	南京	2998	2998
	总计	3279	3279
总计		3279	3279

数据分析设置器

⇄ 订单

⇄ 客户

⇄ 雇员

⇄ 性别

⇄ 签单年

⇄ 签单月

⇄ 城市

⇄ 单价

⇄ 数量

+

复杂条件

	上表头
签单年 ▼	数量求和 ▼
城市 ▼	+

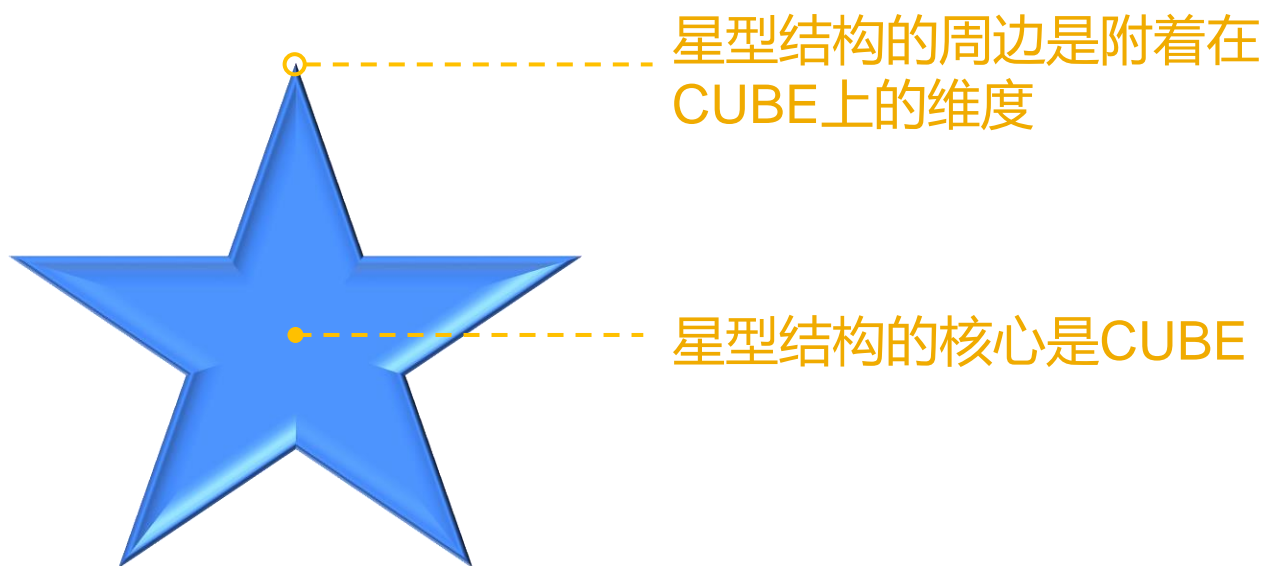
小结

- 通过本节课的学习，我们掌握了通过SQL进行多维分析，并了解了这种方式的优缺点
- 学习了语义层，通过一次建模可以解决查询复用（不必每次都写SQL）问题；解决了显示值方面的显示和排序问题；同时基于语义层的切片会更加方便
- 此外，基于本节（和前两节）使用的数据，还涉及一个概念在总结部分给出——**星型结构**

Summary

知识点-星型结构

- 在前三节课中我们只使用了一张销售数据“宽表（也称为CUBE）”，表里有数量、价格测度，以及和这些指标相关的维度。像这种在一个表中包含了若干测度和与测度直接相关的一组维度的结构我们称为星型结构



□ 那么具备星型结构的宽表如何生成？有哪些注意事项？下节课我们继续探讨。

THANKS

感谢观看



乾学院 · VIP课堂

商业智能

多维分析技术 (宽表使用)



上章回顾

- 上节课我们学习了如何基于数据库（单表）进行多维分析，通过编写SQL可以快速准备数据供多维分析使用
- 如果还需要复用查询、使用显示值，增强操作上的便利，可以通过语义层来实现
- 语义层是多维分析非常重要的组成部分，通过语义层可以很方便地完成语义转换，将数据从技术描述转化为业务描述
- 我们还介绍了星型结构，前面的课程使用的都是具备星型结构的宽表（CUBE），这种结构的表显然不是基础业务表，那是如何生成的呢？本节我们来介绍一下。

本章目标

- 掌握宽表的生成以及注意事项
- 了解ETL的概念
- 熟悉宽表在BI应用中的优缺点

CONTENTS

本章内容

1. 宽表
2. 宽表生成
3. ETL
4. 宽表预汇总
5. 汇总使用
6. 层向透明化

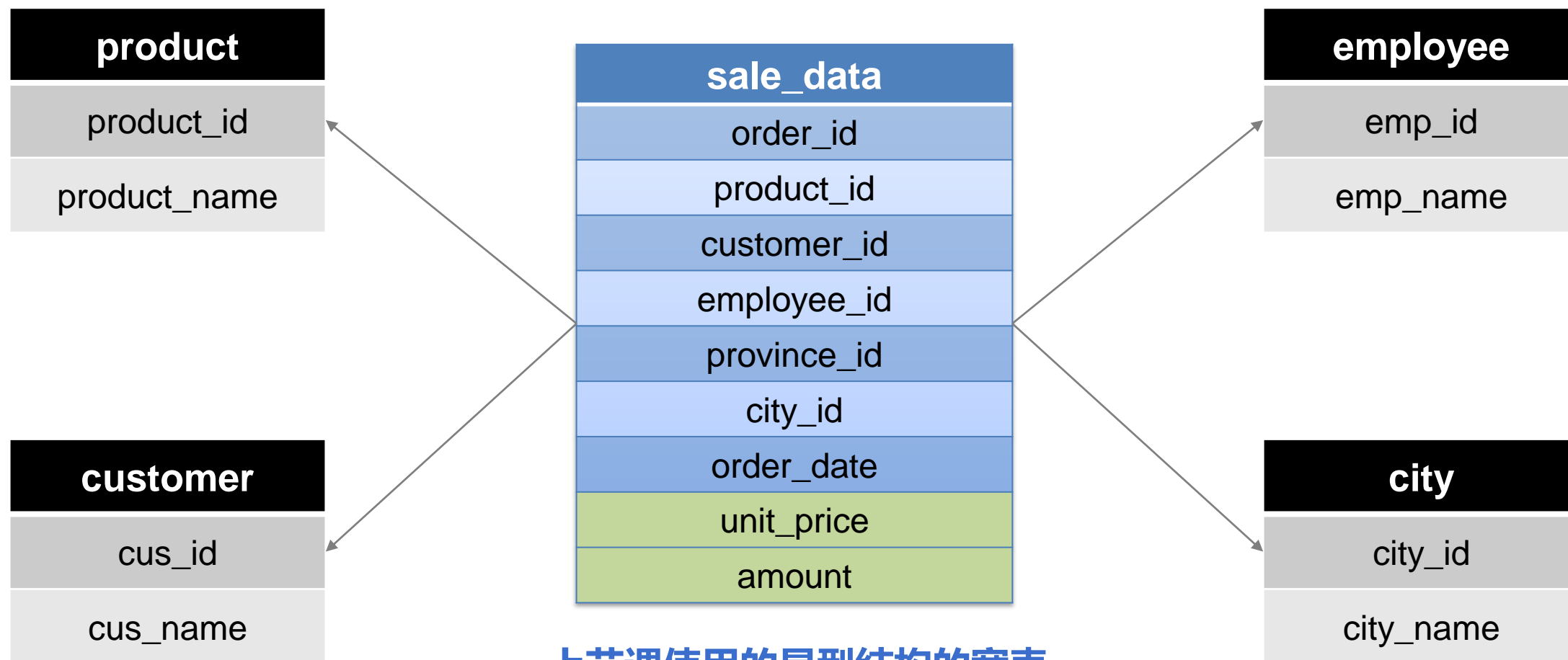




宽表

宽表与CUBE回顾

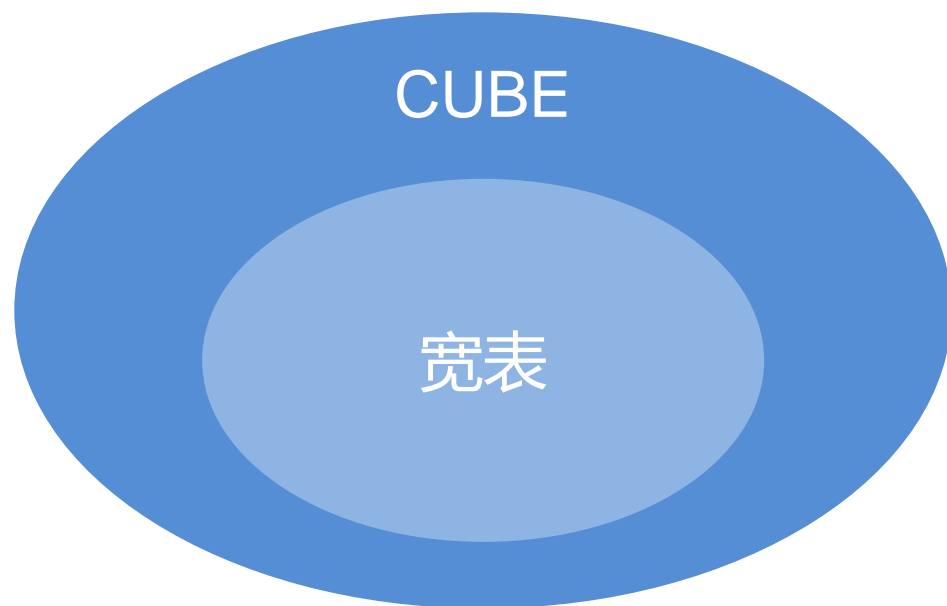
宽表是具有多个维度和指标的数据库表，由于列数较多（来源多个表），因此被称为宽表（CUBE）



上节课使用的星型结构的宽表

知识点-宽表与CUBE的关系

- **宽表就是CUBE!**
- 基于RDB中我们经常会把CUBE称为宽表，DB中更喜欢用“表”的称呼；实际上CUBE的范围更大一些，而“CUBE”的称呼在OLAP体系中也更常用
- 因此请大家注意我们课程中使用的宽表与CUBE是一样的



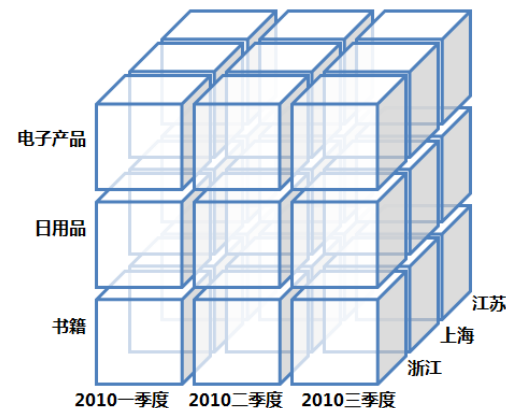
知识点-MOLAP&ROLAP

OLAP技术在发展过程出现过MOLAP和ROLAP两种形式，还有两者的混合HOLAP

MOLAP

- MOLAP表示基于多维数据组织的OLAP实现，多维数据存储是核心
- 其特点是将细节数据和聚合后的数据均保存在CUBE中，所以以空间换效率，查询时效率高，但生成CUBE时需要大量的时间和空间。
- MOLAP服务器/产品包括ESSBASE、Cognos等

CUBE的称呼是从早期MOLAP技术延续下来的



知识点-MOLAP&ROLAP

ROLAP

- ROLAP表示基于关系数据库的OLAP实现，关系数据库是核心，以关系型结构进行多维数据的表示和存储
- 其特点是将细节数据保留在关系型数据库的事实表中，聚合后的数据也保存在关系型的数据库中，ROLAP的灵活性很高，但其应用一直受限于性能不高，但随着数据库的性能不断提升（包括分布式数据库的出现），ROLAP的应用越来越广泛。在新技术的推动下市场上也出现了一批新型的ROLAP产品（本课程中举例使用的润乾多维分析就是其中之一）

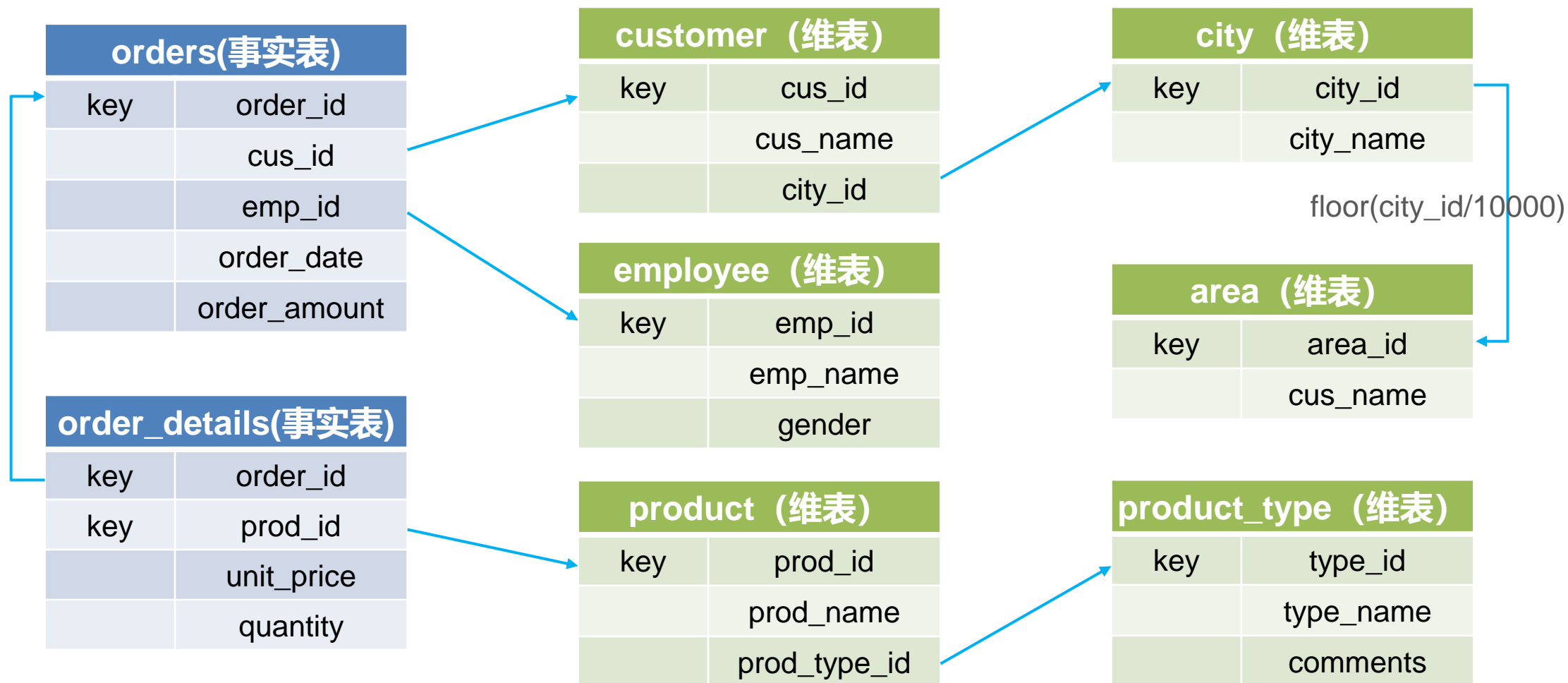
宽表是在ROLAP中CUBE的别称



宽表生成

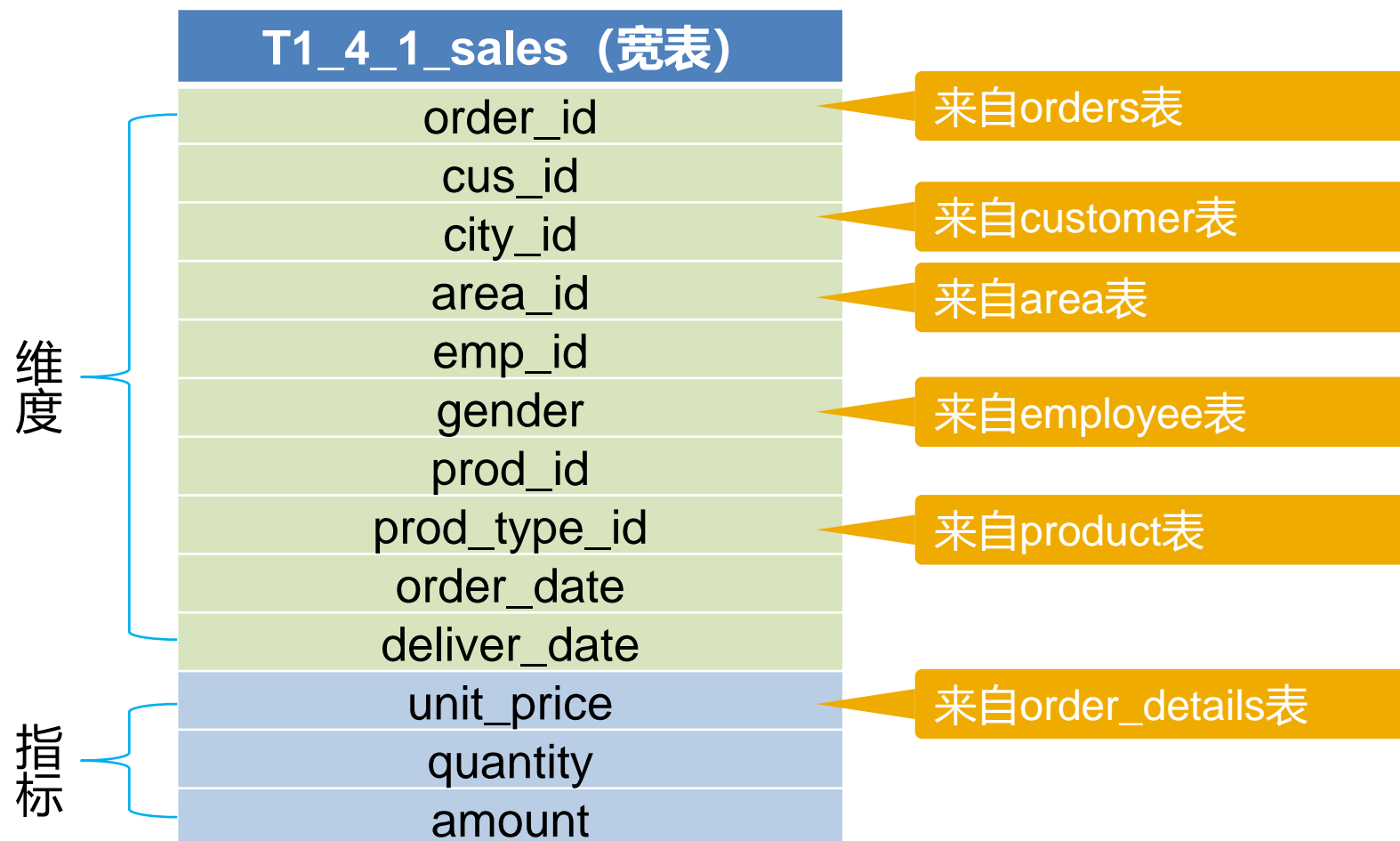
宽表生成

我们继续来看一下宽表应该如何生成？下面是一些业务基础表，我们要基于这些表来生成宽表



宽表生成

现在我们要基于上面的基础表生成如下宽表，用来分析销售情况



宽表生成

数据概况

order_id	cus_id	city_id	area_id	emp_id	gender	prod_id	prod_type_id	order_date	deliver_date	unit_price	quantity	amount
11006	GREAL	30101	3	1	0	1	1	2012-11-07	2012-11-15	18	8	144
10406	QUEEN	30101	3	1	0	1	1	2011-08-07	2011-08-13	14	10	140
10370	CHOPS	30201	3	2	1	1	1	2011-07-03	2011-07-27	14	15	178.5
11047	EASTC	30201	3	2	1	1	1	2012-11-24	2012-12-01	18	25	337.5
10646	HUNGO	30201	3	2	1	1	1	2012-03-27	2012-04-03	18	15	202.5
10413	LAMAI	30201	3	2	1	1	1	2011-08-14	2011-08-16	14	24	336
10317	LONEP	30201	3	2	1	1	1	2011-04-30	2011-05-10	14	20	280
10354	PERIC	30201	3	2	1	1	1	2011-06-14	2011-06-20	14	12	168
10285	QUICK	30201	3	2	1	1	1	2011-03-20	2011-03-26	14	45	504
10691	QUICK	30201	3	2	1	1	1	2012-05-03	2012-05-22	18	30	540
10869	SEVES	30201	3	2	1	1	1	2012-09-04	2012-09-09	18	40	720

T1_4_1_sales (宽表) 数据

宽表生成

在同一个DB内可以使用SQL来生成宽表

```
create table 1_4_1_sales_wide(  
    select  
o.order_id,o.cus_id,c.city_id,area.area_id,o.emp_id,e.gender,od.prod_id,prod.prod_type_id,o  
.order_date,o.deliver_date,od.unit_price,od.quantity,od.unit_price*od.quantity*(1-od.discount)  
as amount  
    from orders o  
    left join order_details od on o.order_id=od.order_id  
    left join customer c on o.cus_id=c.cus_id  
    left join city on c.city_id=city.city_id  
    left join area on floor(city.city_id/10000)=area.area_id  
    left join employee e on o.emp_id=e.emp_id  
    left join product prod on od.prod_id=prod.prod_id  
)
```

生成宽表注意事项

重复数据

业务表通过JOIN (1:N) 生成宽表时要注意是否存在指标重复的情况，指标重复会导致汇总错误

orders (订单表)

order_id	cus_id	amount
10248	VINET	3344.2
10249	TOMSP	1813

order_details (订单明细表)

order_id	prod_id	price	quantity
10248	17	48.5	34
10248	21	32.6	52
10249	13	63.5	70

1:N

使用amount汇总就会出错

order_id	cus_id	amount	prod_id	price	quantity
10248	VINET	3344.2	17	48.5	34
10248	VINET	3344.2	21	32.6	52
10249	TOMSP	1813	13	63.5	70

如果出现这种情况，多维分析产品应该在前端（界面）进行限制，拒绝使用重复的amount汇总

生成宽表注意事项

重复数据

- 出现重复数据是因为在一对多的表间关联中，前一个表（订单表）存在指标数据（订单金额），这样关联以后会被复制出多个重复指标数据，如果基于重复数据汇总就会出错
- 解决的办法除了在前端限制汇总以外，还可以：

1.改用其他指标加工

如我们前面生成的销售数据宽表并没有直接使用订单表中的amount，而是由订单明细中的unit_price*quantity加工而成

2.拆分成多个宽表

如果无法由其他指标加工出来，那应该将订单与其他维度单独建立宽表，而不能再和订单明细关联创建宽表

生成宽表注意事项

2. 拆分多个宽表

如果必须使用这个值

orders(事实表)	
key	order_id
	cus_id
	emp_id
	order_date
	order_amount

用这些表建立一个跟订单相关的宽表

customer (维表)	
key	cus_id
	cus_name
	city_id
employee (维表)	
key	emp_id
	emp_name
	gender

city (维表)	
key	city_id
	city_name
	floor(city_id/10000)
area (维表)	
key	area_id
	cus_name

order_details(事实表)	
key	order_id
key	prod_id
	unit_price
	quantity

用这些表建立一个跟明细相关的宽表

product (维表)	
key	prod_id
	prod_name
	prod_type_id

product_type (维表)	
key	type_id
	type_name
	comments

生成宽表注意事项

增减字段

- 在OLAP业务中，宽表的字段数很多，数据量也会很大，增减字段的开销会非常高
- 比如，增删字段会导致已编译的查询计划失效
- 在建立宽表时要尽量将维度和指标添加完全，避免频繁增删字段

宽表要避免频繁增减字段

*注：关于更多宽表性能方面的内容非本课程重点，请参考我们的性能优化课程

思考

建立宽表时，维度的属性（如名称）是否应该放到宽表中？

加入宽表

- 一种做法是将维度属性加入宽表，这样虽然不满足范式要求，但前端查询页面的实现简单些。
- 不过这种方式会导致宽表冗余度迅速增加（维度属性会很多），从而引起存储空间、查询性能等问题，因此这种方式并不推荐

不加入宽表

- 合理的做法是使用单独的表来存储维度属性，通过语义层设置显示值
- 这种方式带来的影响加入宽表恰好相反，可以获得更小的存储以及更高的性能，但是前端的查询页面实现会复杂些。我们**推荐不把维度属性加入宽表**



ETL

思考

思考1：宽表生成后，如何解决宽表数据更新问题？

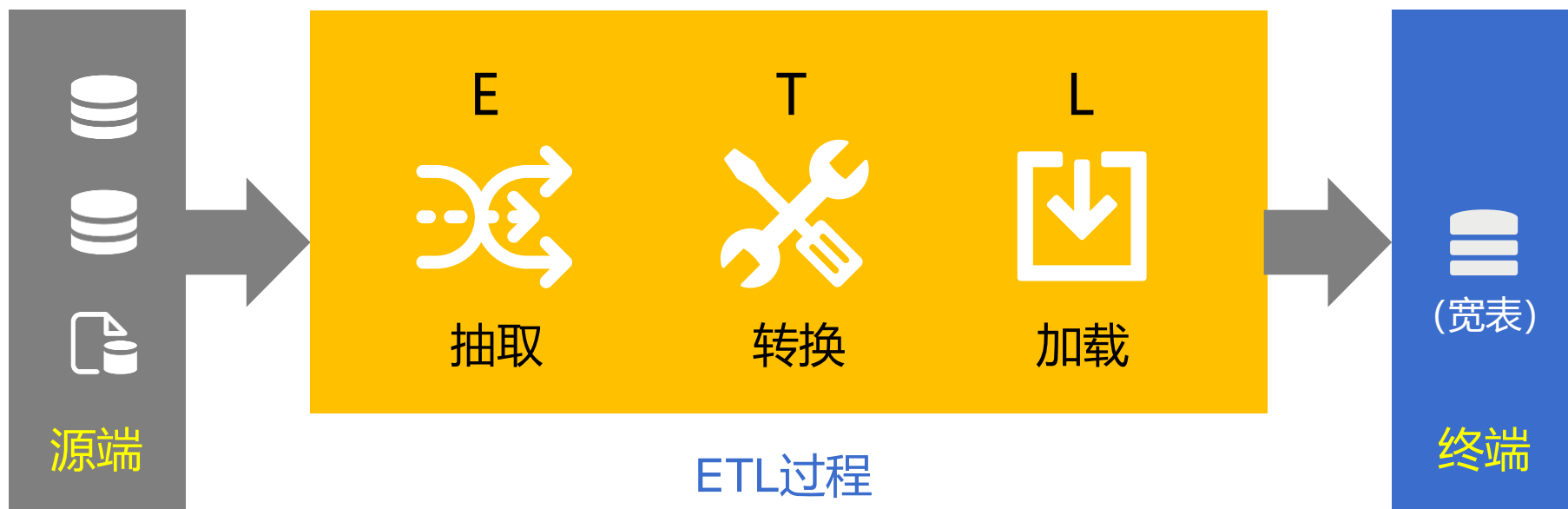
数据库内同步宽表数据可以通过触发器完成，采用适合的CDC策略（如时间戳）就可以完成库内宽表数据同步

思考2：多源数据如何构建宽表？多库、异构源、文件系统等

多源数据无法直接通过SQL搞定，需要选择ETL工具实现。如果同时涉及库内数据和多源数据，可以统一使用ETL工具维护宽表

知识点-ETL

- ETL，是英文Extract-Transform-Load的缩写，用来描述将数据从来源端经过抽取（extract）、转换（transform）、加载（load）至目的端的过程
- ETL不仅可以用于创建宽表，还可用于数据备份、数据仓库等场景，通常涉及数据传输和数据同步都可以尝试使用ETL工具完成
- ETL工具通常都应具备多源的能力，可以将多源数据（DB、File等）处理后加载到目标端



举例-生成宽表与数据同步

将本校学生排名数据（文本）关联学生信息（文本）后写入数据库

ID	SchoolID	ClassName	Score	ClassOrder	SchoolOrder	UnionOrder
7599008	641	03	508	15	15	15
7599011	641	22	499	1	18	18
7599014	641	23	0	66	150	669
7599015	641	22	0	56	150	669
7599022	641	22	0	56	150	669
7599036	641	22	354.5	32	89	143
7599037	641	22	302	43	122	262

成绩排名文本

id	cname	gender	birthday
7599008	张守强	女	1996-02-13
7599011	王思遥	男	1996-04-16
7599014	王浩霖	男	1997-07-26
7599015	王家旭	女	1995-02-11
7599022	国实	女	1995-11-23
7599036	崔海胤	男	1997-05-22
7599037	黄东柱	女	1995-12-31
7599040	孔繁喆	女	1996-06-15

学生代码表

举例-生成宽表与数据同步

将本校学生排名数据（文本）关联学生信息（文本）后写入数据库

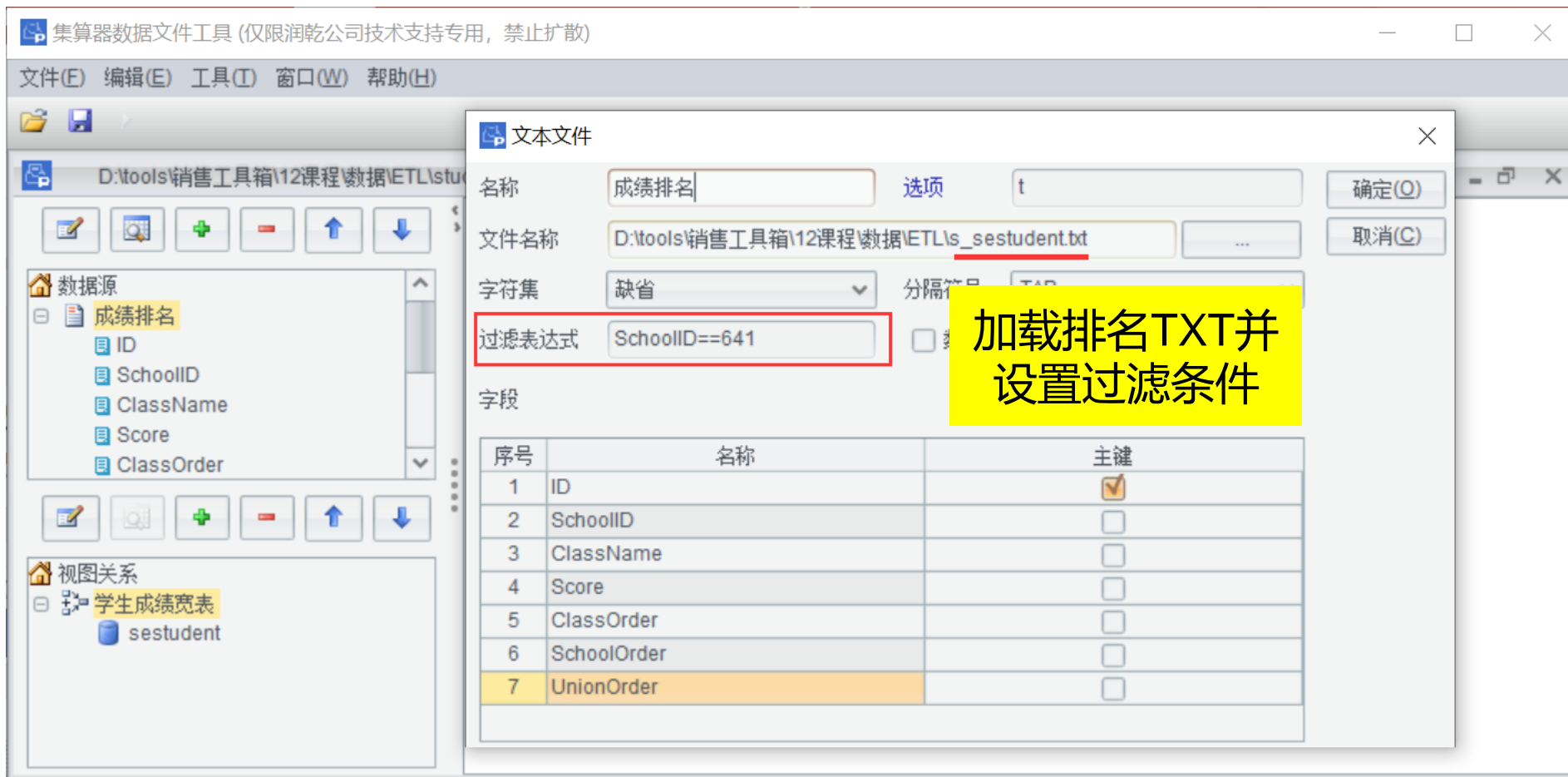
ID	cname	gender	age	SchoolID	ClassName	Score	ClassOrder	SchoolOrder	UnionOrder
7599008	张守强	女	24	641	3	508	15	15	15
7599011	王思遥	男	23	641	22	499	1	18	18
7599014	王浩霖	男	22	641	23	0	66	150	669
7599015	王家旭	女	25	641	22	0	56	150	669
7599022	国实	女	24	641	22	0	56	150	669
7599036	崔海胤	男	22	641	22	354.5	32	89	143
7599037	黄东柱	女	24	641	22	302	43	122	262
7599040	孔繁喆	女	23	641	22	301	44	124	268
7599041	张影	女	23	641	23	430.5	8	44	54
7599045	连婉彤	男	22	641	23	0	66	150	669
7599058	田思雨	男	24	641	3	516.5	13	13	13
7599066	魏春宇	女	23	641	2	485.5	12	22	20
7599067	张络川	男	23						488
7599072	张铭洋	男	24						669
7599075	李金洋	女	23	641	22	410	16	58	75
7599077	康佳馨	男	22	641	23	384	21	71	101

目标表

其中age是根据birthday计算出来的

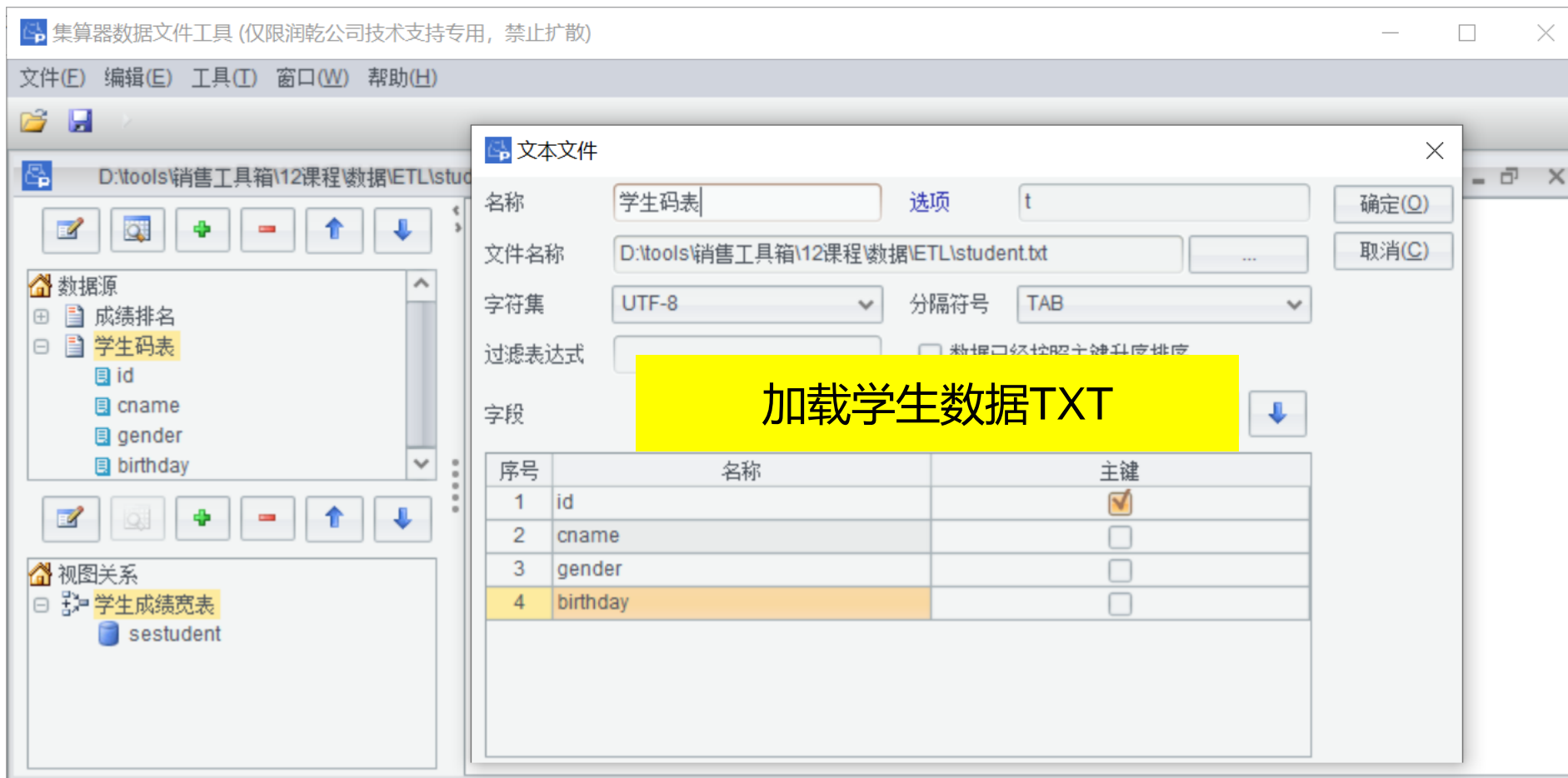
举例-生成宽表与数据同步

以润乾ETL工具为例（例子文件：7.1.student_ETL.ept）



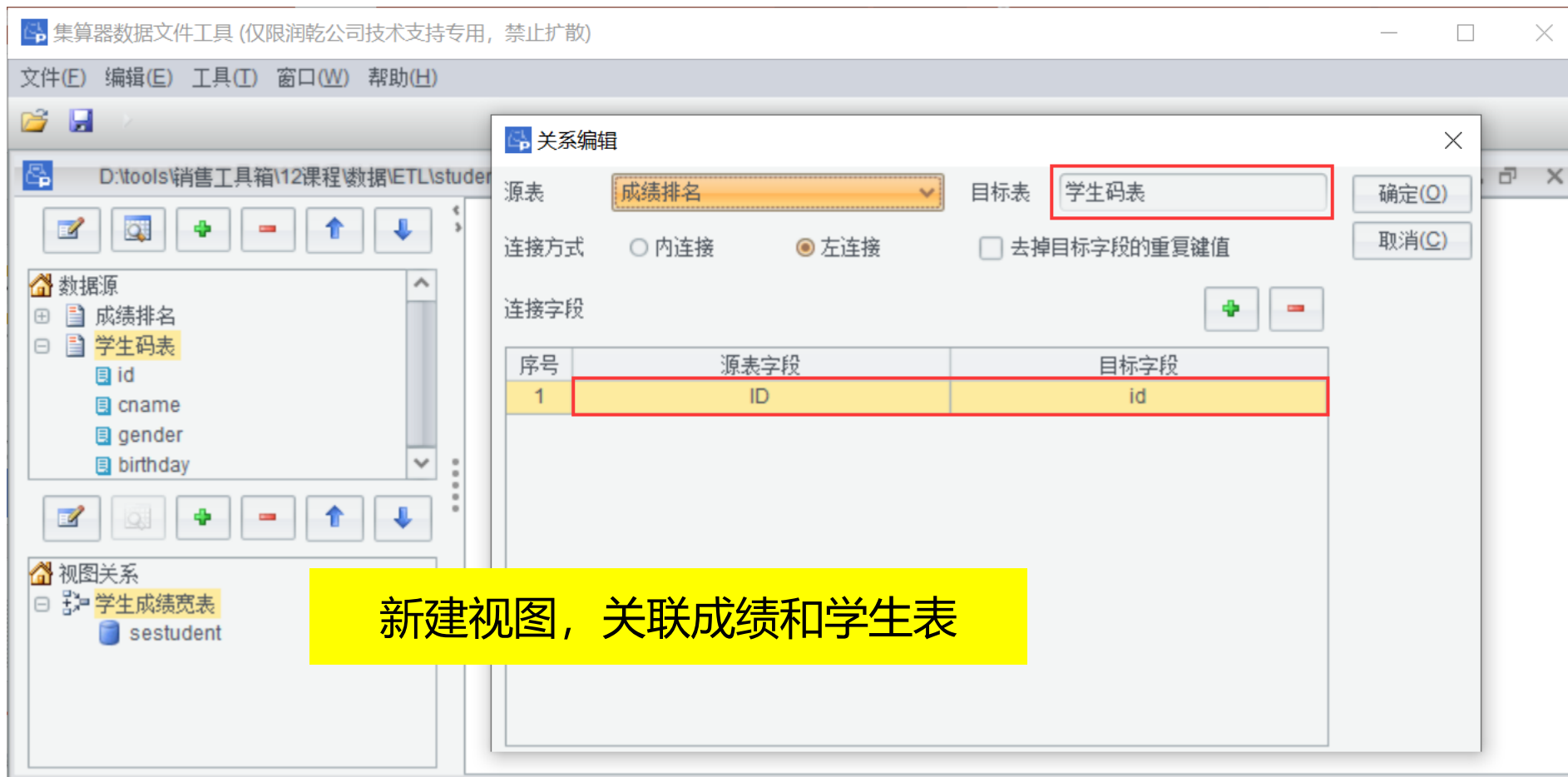
举例-生成宽表与数据同步

以润乾ETL工具为例



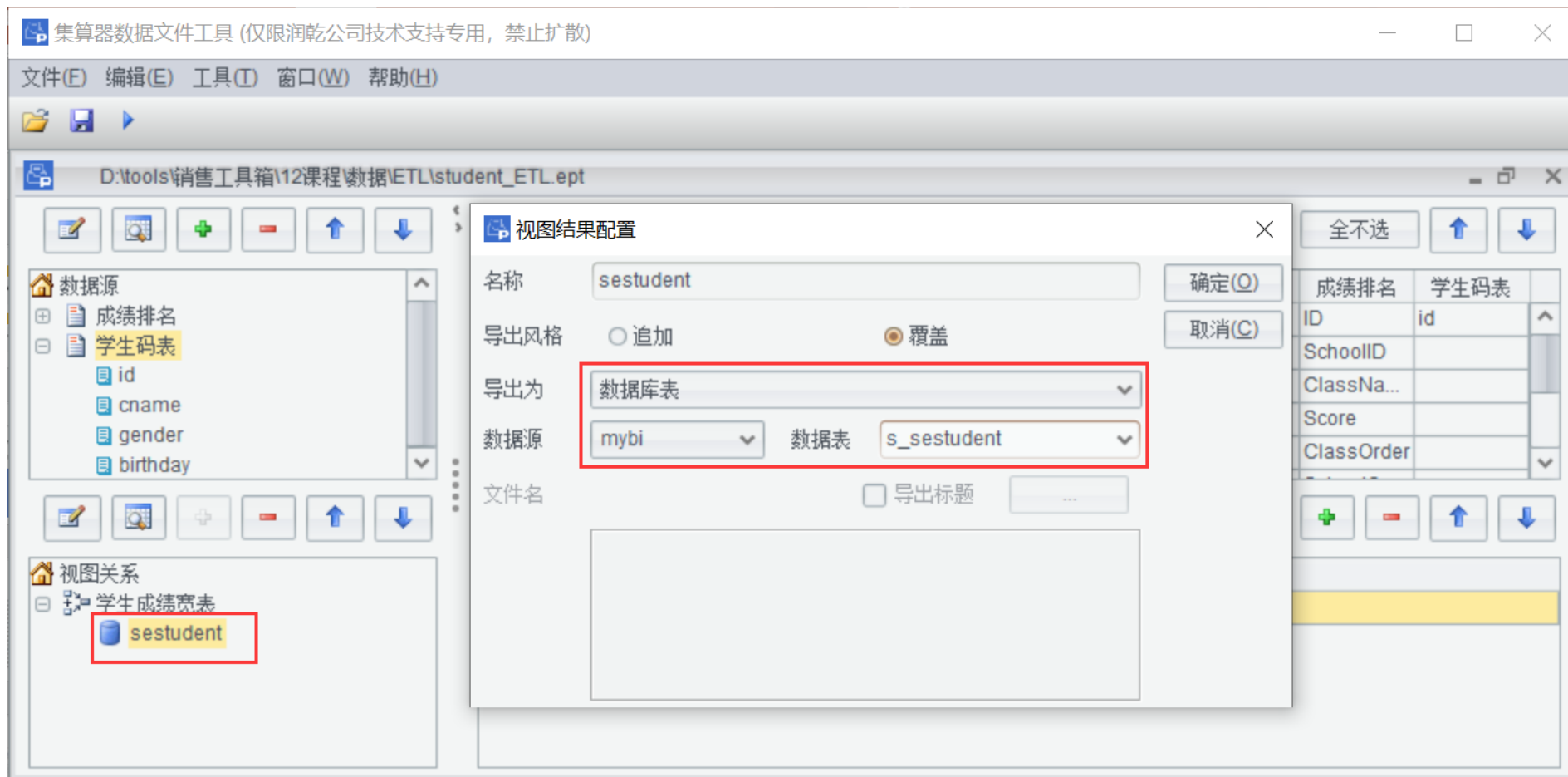
举例-生成宽表与数据同步

以润乾ETL工具为例



举例-生成宽表与数据同步

添加更新，设置选出列（哪些列更新到数据库）



举例-生成宽表与数据同步

添加更新，设置选出列（哪些列更新到数据库）

集算器数据文件工具 (仅限润乾公司技术支持专用, 禁止扩散)

文件(E) 编辑(E) 工具(I) 窗口(W) 帮助(H)

D:\tools\销售工具箱\12课程\数据\ETL\student_E

设置选出字段和别名

序号	别名	选出	主键	引用名字	成绩排名	学生码表
1	ID	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	ID	ID	id
2	SchoolID	<input checked="" type="checkbox"/>	<input type="checkbox"/>	SchoolID	SchoolID	
3	ClassName	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ClassName	ClassNa...	
4	Score	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Score	Score	
5	ClassOrder	<input checked="" type="checkbox"/>	<input type="checkbox"/>	ClassOrder	ClassOrder	

计算列(双击上方引用名字添加到表达式)

序号	别名	计算表达式
1	age	age(birthday)

设置计算列，通过birthday计算age

数据源

- 成绩排名
- 学生码表
 - id
 - cname
 - gender
 - birthday

视图关系

- 学生成绩宽表
 - sestudent

举例-生成宽表与数据同步

手动执行ETL过程

集算器数据文件工具 (仅限润乾公司技术支持专用, 禁止扩散)

文件(E) 编辑(E) 工具(T) 窗口(W) 帮助(H)

D:\tools\销售工具箱\12课程\数据\ETL\student_ETL.ept

数据源

- 成绩排名
- 学生码表
 - id
 - cname
 - gender
 - birthday

视图关系

- 学生成绩宽表
 - sestudent

消息

数据抽取完成, 计算结果已经输出到: s_sestudent。

确定

序号	别名	选出	主键	引用名字	成绩排名	学生码表
				ID	id	
				SchoolID		
				ClassNa...		
				Score		
				ClassOrder		

计算列(双击上方引用名字添加到表达式)

序号	别名	计算表达式
1	age	age(birthday)

举例-生成宽表与数据同步

手动执行ETL过程

mysql

mybi

运行

停止

解释

1 select ID,cname,gender,age,SchoolID,ClassName,Score,ClassOrder,SchoolOrder,UnionOrder from s_sestudent

信息

结果 1

剖析

状态

ID	cname	gender	age	SchoolID	ClassName	Score	ClassOrder	SchoolOrder	UnionOrder
7599008	张守强	女	24	641	3	508	15	15	15
7599011	王思遥	男	23	641	22	499	1	18	18
7599014	王浩霖	男	22	641	23	0	66	150	669
7599015	王家旭	女	25	641	22	0	56	150	669
7599022	国实	女	24	641	22	0	56	150	669
7599036	崔海胤	男	22	641	22	354.5	32	89	143
7599037	黄东柱	女	24	641	22	302	43	122	262
7599040	孔繁喆	女	23	641	22	301	44	124	268
7599041	张影	女	23	641	23	430.5	8	44	54
7599045	连婉彤	男	22	641	23	0	66	150	669
7599058	田思雨	男	24	641	3	516.5	13	13	13
7599066	魏春宇	女	23	641	3	结果表		20	20
7599067	张络川	男	23	641	22			143	488
7599072	张铭洋	男	24	641	23			150	669
7599075	李金洋	女	23	641	22	410	16	58	75
7599077	康佳馨	男	22	641	23	384	21	71	101

举例-生成宽表与数据同步

文件保存为：[ETL_FOR_SALEDATA.ept](#)

手动运行可以点击IDE中的执行按钮 

定时执行可以使用操作系统的定时调度功能，以命令行的形式执行

windows命令行如下：

```
安装目录\bin\esprocx.exe studentETL.ept
```

linux命令行如下：

```
安装目录\bin\esprocx.sh studentETL.ept
```

练习

练习：感兴趣的同学可以自行下载工具练习ETL过程

ETL工具下载地址：

<http://www.raqsoft.com.cn/download/download-jsq>

*需同时下载安装包和开发授权

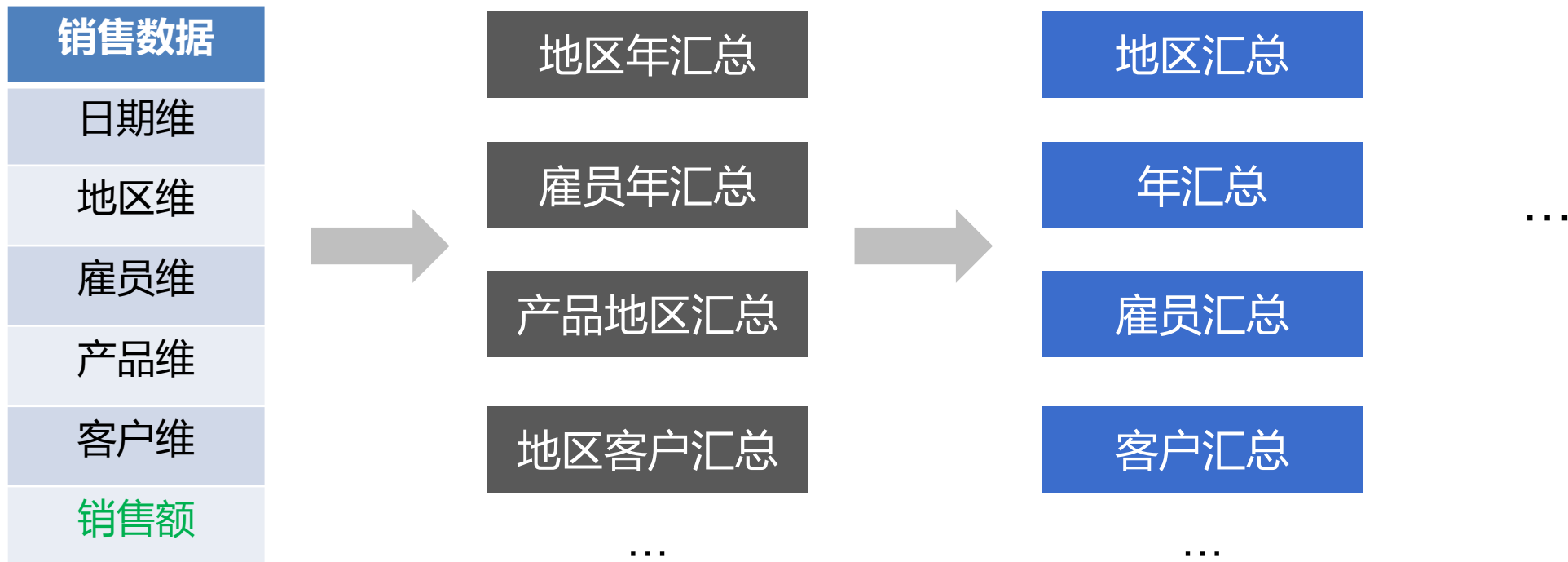
注：除了上述可视化ETL工具，润乾还提供脚本ETL工具，编写脚本可以处理任意复杂的ETL过程，两类ETL工具可以混合使用



宽表预汇总

知识点-宽表预汇总

原始宽表 (sale_data) 包含了多个维度和指标，粒度非常细，如果每次多维分析汇总数据都基于这个表完成性能不高，可以想到的办法是继续使用空间换时间，根据各个维度的组合事先加工汇总，形成不同维度和层级的多个**汇总表**，相当于再做多个（汇总）宽表



*值得注意的是，汇总表不一定由宽表加工而成，使用宽表加工汇总表会比较简单且效率较高

汇总表举例

使用sale_data数据加工雇员销售汇总表和地区销售汇总表

雇员销售汇总表

```
create table t1_4_2_sales_emp(  
    select emp_id,sum(amount) amount from t1_4_1_sales group by emp_id  
)
```

地区销售汇总表

```
create table t1_4_3_sales_city(  
    select city_id,sum(amount) amount from t1_4_1_sales group by city_id  
)
```



汇总表使用

举例-基于宽表和汇总表分析

定义语义层文件，例子参照：[7.2.wide_summary_olap.lmd](#)和[7.2.wide_summary_olap.dct](#)

表名

t1_4_1_sales

物理表名

t1_4_1_sales

记录数

0

表描述

常规

外键

伪字段

基础表

汇总表

子表

外键

序号	外键名称	对应表名
1	fk1	employee
2	fk2	city

外键映射

序号	外键字段	引用字段
1	emp_id	emp_id

表名

t1_4_1_sales

物理表名

t1_4_1_sales

记录数

0

表描述

常规

外键

伪字段

基础表

汇总表

子表

外键

序号	外键名称	对应表名
1	fk1	employee
2	fk2	city

外键映射

序号	外键字段	引用字段
1	city_id	city_id

表名

t1_4_2_sales_emp

物理表名

t1_4_2_sales_emp

记录数

0

表描述

常规

外键

伪字段

基础表

汇总表

子表

外键

序号	外键名称	对应表名
1	fk1	employee

外键映射

序号	外键字段	引用字段
1	emp_id	emp_id

表名

t1_4_3_sales_city

物理表名

t1_4_3_sales_city

记录数

0

表描述

常规

外键

伪字段

基础表

汇总表

子表

外键

序号	外键名称	对应表名
1	fk1	city

外键映射

序号	外键字段	引用字段
1	city_id	city_id

举例-基于宽表和汇总表分析

生成字典，用于语义转化

标题 销售数据 表名 t1_4_1_sales

字段项

序号	名称	字段名	显示格式
1	订单	order_id	
2	产品	prod_id	
3	客户	cus_id	
4	城市	city_id	
5	地区	area_id	
6	雇员	emp_id	
7	性别	gender	
8	产品类型	prod_type_id	
9	订单日期	order_date	
10	发货日期	deliver_date	

标题 雇员销售额汇总 表名 t1_4_2_sales_emp

字段项

序号	名称	字段名	显示格式
1	雇员	emp_id	
2	销售额	amount	
3	fk1	fk1	

标题 雇员 表名 employee

字段项

序号	名称	字段名	显示格式
1	雇员	emp_id	
2	姓名	emp_name	

润乾报表 (仅限润乾公司技术支持专用, 禁止扩散) [D:\tools\销售工具箱\12课程\程序文件\1.4wide_summ

文件(F) 编辑(E) 服务(R) 远程服务(C) 系统(S) 窗口(W) 帮助(H)

1.4wide_summary_olap.dct 1.4wide_summary_olap.l...

表项 维项 分类项

维项

序号	维标题	维名称	代码列名称	代码值标题
1	雇员	employee_emp_id	employee.emp_id	
2	城市	city_city_id	city.city_id	

页面多维分析

发布语义层到web，进行多维分析

保存

打开

数据文件

横向平铺

纵向平铺

重叠显示

数据集

报表名称

数据分析设置器

报表列表

添加报表

报表名称

打印

删除

销售数据

订单

产品

客户

城市

地区

雇员

性别

产品类型

订单日期

发货日期

单价

数量

销售额

雇员销售额汇总, 雇

雇员

销售额

雇员

姓名

城市销售额汇总, 城

城市

销售额

销售数据宽表

上表头

左表头

指标数据区

+

雇员销售额汇总表

城市销售额汇总表

举例-基于宽表和汇总表分析

查询各地区的销售情况/各个销售人员的销售额

这时我们有两种选择，从销售数据宽表中汇总，或者从汇总表中直接查询

报表名称

城市	销售额求和	总计
大连	55787	55787
上海	8662	8662
青岛	15224	15224
南京	126007	126007
常州	57949	57949
南昌	33920	33920
温州	51823	51823
北京	56387	56387
天津	361406	361406
张家口	20129	20129
石家庄	43263	43263
秦皇岛	41525	41525
深圳	220864	220864
海口	24730	24730

数据分析设置器

报表列表

添加报表

报表名称

销售数据

- 订单
- 产品
- 客户
- 雇员
- 签单日期
- 签单年
- 签单月
- 城市**
- 省份
- 单价
- 数量
- 销售额**

雇员销售汇总, 雇员

- 雇员
- 销售额
- 雇员
- 姓名

城市销售汇总, 城市

- 城市
- 销售额
- 城市

上表头

城市

销售额求和

基于明细表汇总

举例-基于宽表和汇总表分析

查询各地区的销售情况/各个销售人员的销售额

这时我们有两种选择，从销售数据宽表中汇总，或者从汇总表中直接查询

The screenshot shows a BI tool interface with the following components:

- 报表名称 (Report Name):** A table with columns '城市' (City) and '销售额' (Sales Amount). The data is as follows:

城市	销售额
大连	55787
上海	8662
青岛	15224
南京	126007
常州	57949
南昌	33920
温州	51823
北京	56387
天津	361406
张家口	20129
石家庄	43263
秦皇岛	41525
深圳	220864
海口	24730
厦门	7522

- 数据分析设置器 (Data Analysis Settings):** A sidebar with a tree view of data sources. The '城市销售汇总, 城市' (City Sales Summary, City) source is selected, and its fields '城市' (City) and '销售额' (Sales Amount) are highlighted with red boxes. Red arrows point from these boxes to the '城市' and '销售额' fields in the summary table.
- 报表列表 (Report List):** A list of reports with a '添加报表' (Add Report) button.
- 报表预览 (Report Preview):** A table with columns '左表头' (Left Header) and '右表头' (Right Header). The '右表头' column has a sub-header '指标标题' (Indicator Title). The '城市' and '销售额' fields are listed under '右表头' with a green plus sign next to '销售额'.
- 汇总表直接查询 (Direct Query from Summary Table):** A yellow box with text indicating that data can be queried directly from the summary table.

汇总表的使用体验问题

汇总表只有看到才能被使用，即只有现实在页面上业务人员才能查询。但汇总表的数量可能很多（取决于性能要求），都显示在页面上是不可接受的。如何解决这个问题呢？

报表名称

城市

销售额

大连

55787

上海

8662

青岛

15224

南京

126007

常州

57949

南昌

33920

温州

51823

北京

56387

天津

361406

张家口

20129

石家庄

43263

秦皇岛

41525

深圳

220864

海口

24730

厦门

7522

数据分析设置器

报表列表

添加报表

报表名称

产品

客户

雇员

签单日期

签单年

签单月

城市

省份

单价

数量

销售额

雇员销售汇总, 雇员

雇员

销售额

雇员

姓名

城市销售汇总, 城市

城市

销售额

城市

名称

上表头

指标标题

左表头

城市

销售额

+

汇总表
看到才能使用

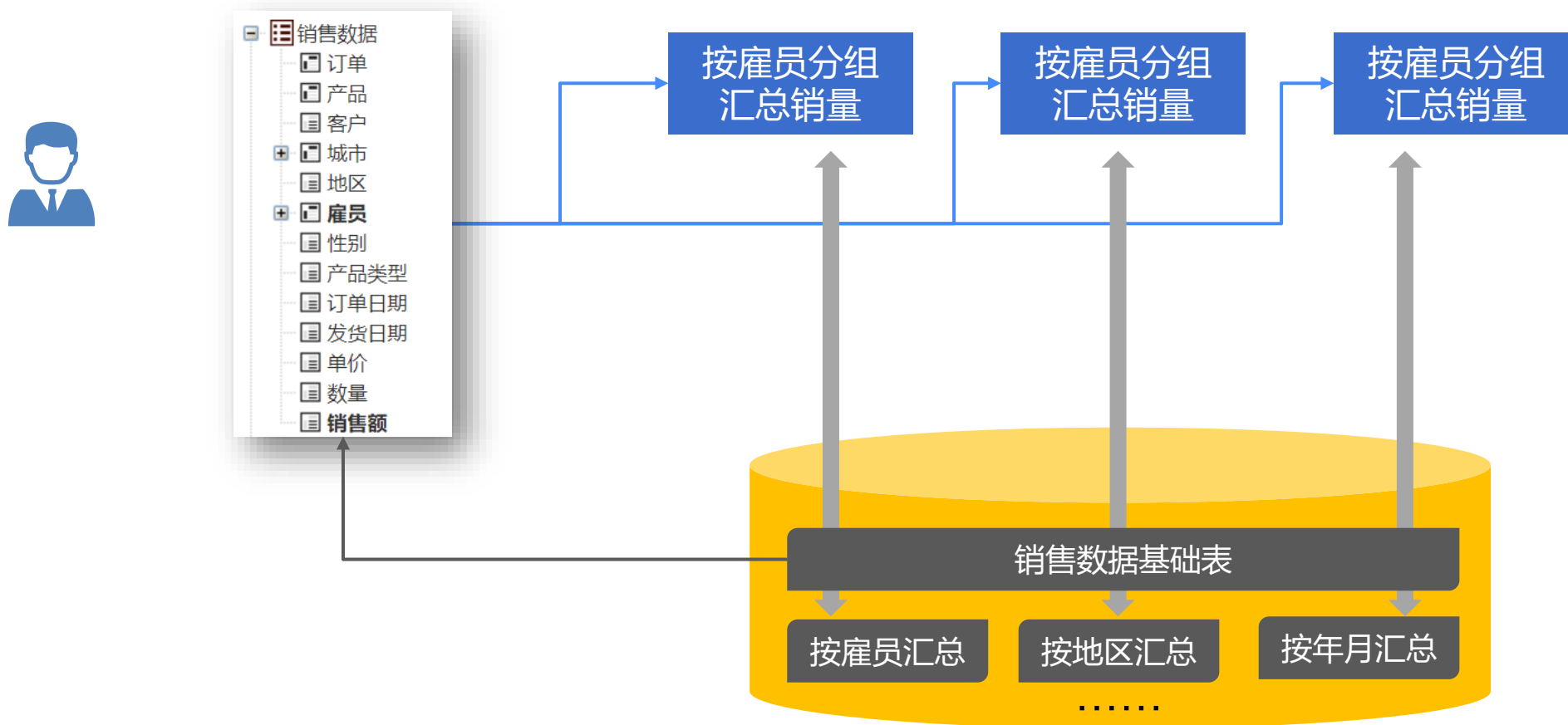
前面的例子只使用了2
个汇总表，如果有20
个，200个会怎么样？



层向透明化

知识点-透明化

透明化是指使用多维分析的业务人员**无需关心汇总表**是否存在，直接针对基础表汇总查询，引擎自动将针对基础表的汇总转化成针对汇总表的查询



知识点-透明化

- 很多OLAP工具会支持自动使用汇总数据，只要事先建立关联
- 专门做CUBE的MOLAP产品会把汇总表和基础表放在一个CUBE里，物理上看不到汇总表，只能看到基础表
- 但ROLAP产品使用关系数据库，那汇总表自然就能看到了
- 而随着ROLAP应用越来越广泛，透明化也就成为ROLAP产品面临的一个问题了

举例-自动使用汇总表

要解决透明化的问题，可以在元数据中建立汇总表和基础表的关系，从而做到自动使用汇总表

润乾报表 (仅限润乾公司技术支持专用, 禁止扩散) [D:\tools\销售工具箱\12课程\程序文件\1.4wide_summary_olap.lmd]

文件(F) 编辑(E) 服务(R) 远程服务(C) 系统(S) 窗口(W) 帮助(H)

1.4wide_summary_olap.dct 1.4wide_summary_olap.l...

表 视图 维和层

表名: t1_4_2_sales_emp

记录数:

在汇总表上添加基础表, 基础表就是我们的宽表

基础表

序号	基础表名
1	t1_4_1_sales

与汇总表相关的维度和指标

汇总表字段

序号	字段名	广义外键或表达式
1	emp_id	emp_id
2	amount	sum(amount)

这样定义以后，多维分析页面就只需要显示基础表就可以了，使用基础表针对employee汇总销售额时：

```
select emp,sum(amount)
from sales (基础表)
group by emp
```

会被转换成：

```
select emp, amount
from sales_emp (汇总表)
```


举例-自动使用汇总表

城市销售额汇总表也做同样的设置

禁止扩散) [D:\tools\销售工具箱\12课程\程序文件\1.4wide_summary_olap.lmd]

统(S) 窗口(W) 帮助(H)

te_summary_olap.i...

表名: t1_4_3_sales_city 物理表名: t1_4_3_sales_city

记录数: 0 表描述:

常规 外键 伪字段 基础表 汇总表 子表

基础表

序号	基础表名
1	t1_4_1_sales

字段

序号	字段名	广义外键或表达式
1	city_id	city_id
2	amount	sum(amount)

禁止扩散) [D:\tools\销售工具箱\12课程\程序文件\1.4wide_summary_olap.lmd]

统(S) 窗口(W) 帮助(H)

te_summary_olap.dct

表名: t1_4_1_sales 物理表名: t1_4_1_sales

记录数: 0 表描述:

常规 外键 伪字段 基础表 汇总表 子表

汇总表名称

序号	汇总表名称
1	t1_4_3_sales_city
2	t1_4_2_sales_emp

汇总表物理字段

序号	汇总表物理字段
1	city_id
2	amount

切换到基础表的“汇总表”标签
可以看到有两个汇总表

基础表和汇总表是相互的关系，
一个表是另一个的汇总表，另一个就是这个表的基础表

举例-自动使用汇总表

发布语义层到WEB，可以看到“汇总表”都已经没有了，我们按雇员汇总销售额

报表名称

雇员

销售额求和

总计

张颖	50424.45	50424.45
王伟	338618.6	338618.6
李芳	53154.65	53154.65
郑建杰	51926.65	51926.65
赵军	185070.8	185070.8
孙林	7116.15	7116.15
金士鹏	207699.6	207699.6
刘英玫	116513.55	116513.55
张雪眉	242008.52	242008.52
总计	1252532.9 7	1252532.9 7

数据分析设置器

销售数据

订单

产品

客户

城市

地区

雇员

性别

产品类型

订单日期

发货日期

单价

数量

销售额

雇员

雇员

姓名

城市

城市

名称

上表头

雇员

销售额求和

页面上汇总表不见了，我们直接针对基础表汇总

举例-自动使用汇总表

发布语义层到WEB，可以看到“汇总表”都已经没有了，我们按雇员汇总销售额

报表名称

雇员	销售额求和	总计
张颖	50424.45	50424.45
王伟	338618.6	338618.6
李芳	53154.65	53154.65
郑建杰	51926.65	51926.65
赵军	185070.8	185070.8
孙林	7116.15	7116.15
金士鹏	207699.6	207699.6
刘英玫	116513.55	116513.55
张雪眉	242008.52	242008.52
总计	1252532.9	1252532.9

数据分析设置器

销售数据

订单

产品

客户

城市

地区

雇员

性别

产品类型

订单日期

发货日期

单价

数量

销售额

雇员

雇员

姓名

上表头

雇员

销售额求和

页面上汇总表不见了，我们直接针对基础表汇总

DQL服务器 127.0.0.1 : 3366

```
2020/03/25 13:13:34:[DQL]:SELECT T1.sum(amount) 销售额求和 ON employee_emp_id 雇员 FROM t1_4_1_sales T1 BY T1.emp_id
2020/03/25 13:13:34:[SQL]:SELECT T_1_1.emp_id "雇员",T_1_1.amount "销售额求和" FROM t1_4_2_sales_emp T_1_1
```

启动(S)

停止(Q)

退出(Q)

复制(C)

后台可以看到针对基础表sales的汇总查询最终被转化成针对汇总表sales_emp的查询了

基础表、宽表和汇总表比较

上面我们用到了基础表、宽表和汇总表，这里总结一下。宽表由基础业务表加工而成；汇总表往往由宽表加工而成，我们从查询的灵活性和性能两方面比较一下三种类型的表

	基础表	宽表	汇总表
灵活性	强	一般	差
性能	低	一般	高

- 宽表和汇总表都是采用空间换时间的手段，前者主要为了避免关联查询（大多数多维分析产品都不支持关联查询，只能针对一个宽表/CUBE分析），后者则采用更高层次的聚合来获得高性能；而多维分析的灵活性则随着数据的不断加工越来越差，没建立在宽表中的维度是不能查询的（汇总表也一样）
- 那么有没有办法基于DB多维分析时性能高且灵活性又强呢？后面的课程我们继续探讨

小结

- 本章了解了宽表的生成与注意事项
 - 使用ETL工具可以完成同库或异构源的宽表建立，可以通过适合的CDC策略更新宽表数据
 - 对查询性能有要求时还可以通过预汇总数据表来提升查询性能，事实上宽表数据规模较大时汇总表往往必不可少
 - 由于绝大多数多维分析产品不支持关联查询，只能使用一个个单独的宽表/CUBE分析，使用适当的透明化机制可以减轻业务人员的使用成本
- 下节课我们将沿着本节的遗留问题，来看看关联分析的效力！

THANKS

感谢观看



乾学院 · VIP课堂

商业智能

多维分析技术 (关联分析)



上章回顾

- 上章我们了解了宽表的生成，知道如果在库内可以通过SQL来完成，如果涉及多源则需要通过ETL工具实现
- 宽表生成时的注意事项，注意指标重复会引起汇总错误，不要对宽表频繁增减列，我们建议不要将维度属性加入宽表中
- 了解了MOLAP和ROLAP的概念，现在主流大家都在使用ROLAP，而ROLAP会面临透明化的问题，通过适当的机制可以解决从基础表汇总到汇总表查询的转换

本章目标

1. 了解星型模型的缺点
2. 深入了解雪花型模型
3. 基于雪花模型实现的关联分析及其意义

CONTENTS

本章内容

1. 星型结构回顾
2. 雪花型结构与关联分析
3. 事实星座





星型结构回顾

星型结构回顾

- 为了多维分析（从一张表分组汇总）方便，将测度和相关维度放到一张宽表中，该宽表具备星型结构
- 星型结构存在如下问题：

需ETL数据
实时性差

数据冗余
容量大

固定维度
灵活性受限

字段太多
可用性受限

可能导致
数据不一致

其中，可用性和数据错误影响比较大
如何理解这些问题？下面具体来看一下

星型结构的缺点

星型结构的前三个缺点比较容易理解：

数据实时性差

物理宽表因为需要ETL，所以数据往往不是实时数据；通过逻辑宽表可以缓解这个问题

数据冗余容量大

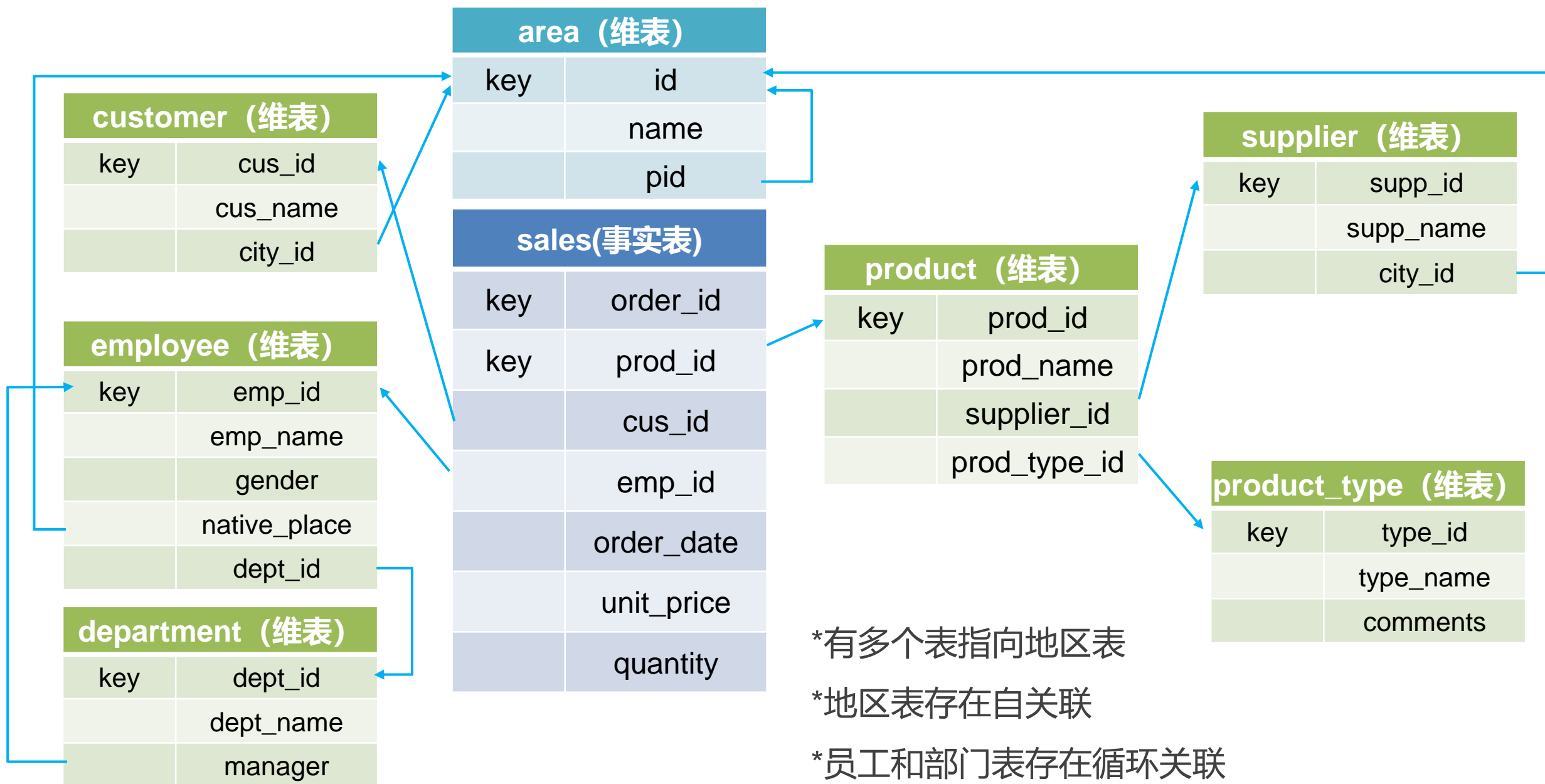
宽表不满足范式，存在大量冗余，占用空间较大；本质上星型结构是一种空间换时间的手段

灵活性受限

宽表采用的是按需建模手段，将可能用到的维度组织到宽表中，如果事先没预置则无法查询

对于后两个问题，我们首先看一下下面的数据结构，也是本节课要用到的表！

数据结构



星型结构的缺点

关于可用性的问题，我们通过例子来看一下：

分别按客户所在地区、供应商所在地区、雇员籍贯汇总销售额

如果要基于宽表实现上面的查询，那宽表该如何构建呢？

订单	产品	客户	客户地区	供应商	供应商地区	销售	销售籍贯	...	销售额
			北京		青岛		沈阳		

地区维被重复了3次，字段多了3个

- 如果三个维表还指向另外一个维表，那宽表的维字段就又多出3个
- 这样在多层维表的加持下，宽表字段的数量会急剧增加，有时可以达到上千个，多维分析页面几乎无法选择和使用

星型结构的缺点

再看一个例子：

根据员工和部门表，查询中国经理的美国员工

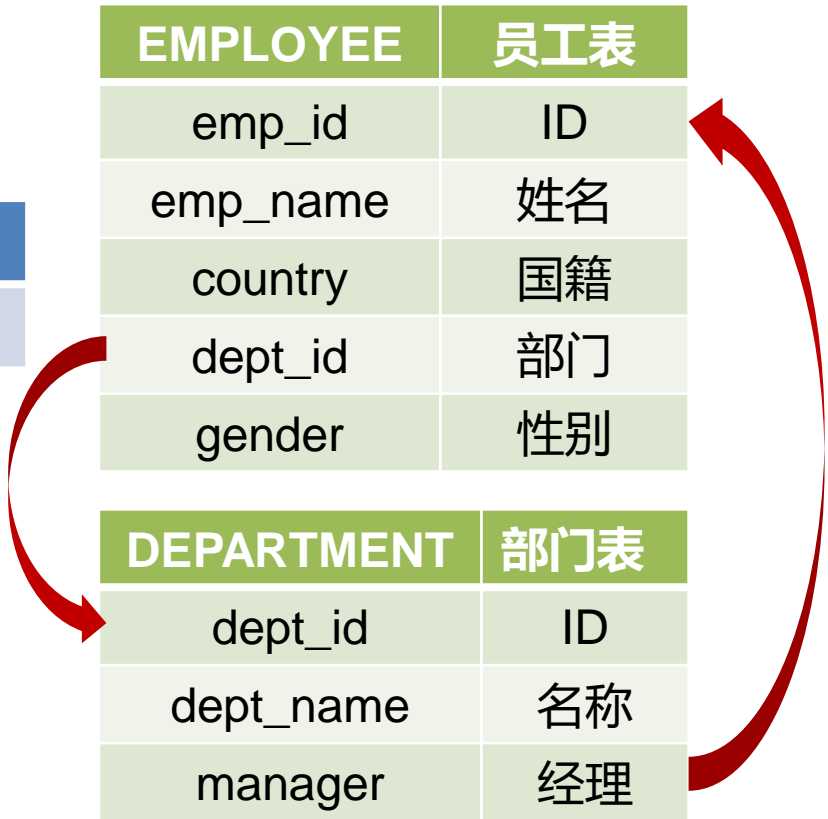
在上面的表结构中，员工表和部门表存在循环关联，这时要完成上面的查询需要如何构建宽表呢？

员工	姓名	...	国籍	经理	经理姓名	经理国籍	...
				北京		青岛	

员工信息（员工表）

经理信息（员工表）

循环关联会导致员工表字段被复制多份，宽表字段数增加；如果还叠加多层维表的情况，宽表的字段数又会成倍增加



星型结构的缺点

- 在多层维表和循环关联（自关联）的情况下，宽表的字段数量会非常多
- 这个问题无论物理宽表还是逻辑宽表都存在
- 用户从数以千计的字段中拖拽分析是不可接受的现实



星型结构的缺点

星型结构不满足范式，可能导致数据错误（不一致）

订单	产品	客户	客户地区	销售	性别	籍贯	供应商	供应商地区	...	单价	数量
10619	22	MEREP	10101	4	1	30101	9	10102		21	40
10566	18	BLONP	10101	4	0	30101	9	10102		62	18
10332	18	MEREP	10101	4	1	30101	9	10102		50	40
11001	22	FOLKO	20301	8	0	30101	9	20102		21	25

同一个销售员出现了不同的性别

同一个供应商出现了不同的所在地

- 宽表不满足范式，无论是ETL程序错误，还是传输异常导致的错误数据都可以顺利地写入宽表
- 错误数据会直接导致多维分析结果错误，而异常值往往很难追踪到



雪花型结构

如何解决星型结构存在的问题？

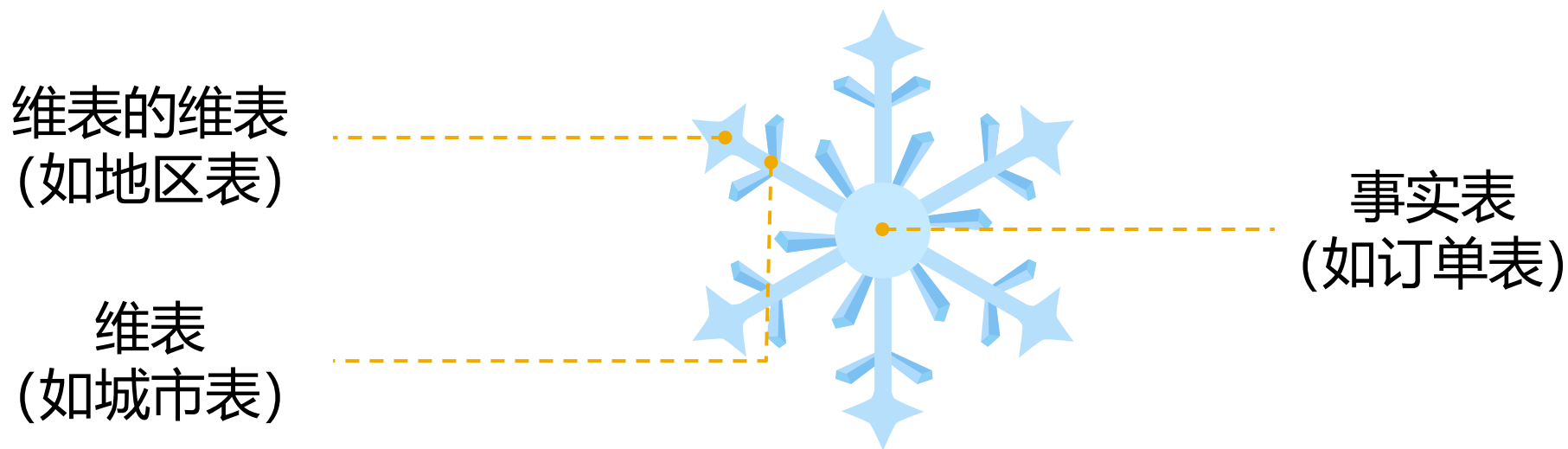
规范化！

- 规范化数据，让数据结构符合范式，消除冗余
- 数据规范以后，原来的单个宽表就会变成多个具有关联关系的表
- 多维分析中，如何很好地描述表间关联，从而让业务人员能顺利使用，是我们接下来要探讨的内容
- 下面首先了解一下，规范化以后的数据形成的雪花型结构

规范化

知识点-雪花型结构

- 在雪花型结构中，将维度进行规范化，消除数据冗余，存在多层维表与事实表相连
- 由于其结构与雪花形似，故称为雪花型结构
- 雪花结构满足范式，可以解决星型结构存在的那些问题



知识点-维表和事实表

- 雪花型结构涉及多个表，其中保存测度值或事实的表称为事实表，如订单表
- 另一类仅包含维度（如产品）及其属性（如产品颜色）称为维表，维表为用户提供了分析数据的窗口
- 维度还包含层次（如日期维分年、月、季度、周等诸多层次），在多维分析中可以方便用户从不同粒度观察数据

order_id	cus_id	emp_id	order_date	deliver_date	order_amount
10248	VINET	2	2011-02-04	2011-02-16	440
10249	TOMSP	9	2011-02-05	2011-02-10	1863
10250	HANAR	9	2011-02-08	2011-02-12	1813
10251	VICTE	9	2011-02-08	2011-02-15	670
10252	SUPRD	2	2011-02-09	2011-02-11	3730
10253	HANAR	9	2011-02-10	2011-02-16	1444
10254	CHOPS	2	2011-02-11	2011-02-23	625
10255	RICSU	8	2011-02-12	2011-02-15	2490

事实表（订单表）

cus_id	cus_name	contact_name	Contact_title	city_id
ALFKI	三川实业有限公司	刘小姐	销售代表	30201
ANATR	东南实业	王先生	物主	30201
ANTON	坦森行贸易	王炫皓	物主	30302
AROUT	国顶有限公司	方先生	销售代表	40101
BERGS	通恒机械	黄小姐	采购员	20301
BLAUS	森通	王先生	销售代表	30201
BLONP	国皓	黄雅玲	市场经理	10101
BOLID	迈多贸易	陈先生	物主	50101

维表（客户表）

业界常见做法

数据规范化以后，会面临多表关联查询的问题，业界在基于雪花型结构时通常的做法有三种：

1.使用逻辑宽表、2.页面自动关联、3.由用户指定关联

1. 逻辑宽表

采用物理宽表的思路，面对雪花结构时通过逻辑宽表事先完成表关联，再交给业务人员使用

缺点

逻辑宽表无法解决多层维表和循环关联（自关联）产生的多字段问题，**太多字段**对于业务人员仍然不可用

无法解决上千个字段的可用性问题的

业界常见做法

2. 页面自动关联

- 既然事先关联存在可用性问题，那能不能用的时候自动关联呢？
- 能想到的办法是在页面上，由用户选择的表，然后根据字段类型、长度、甚至值匹配程度等信息自动完成关联

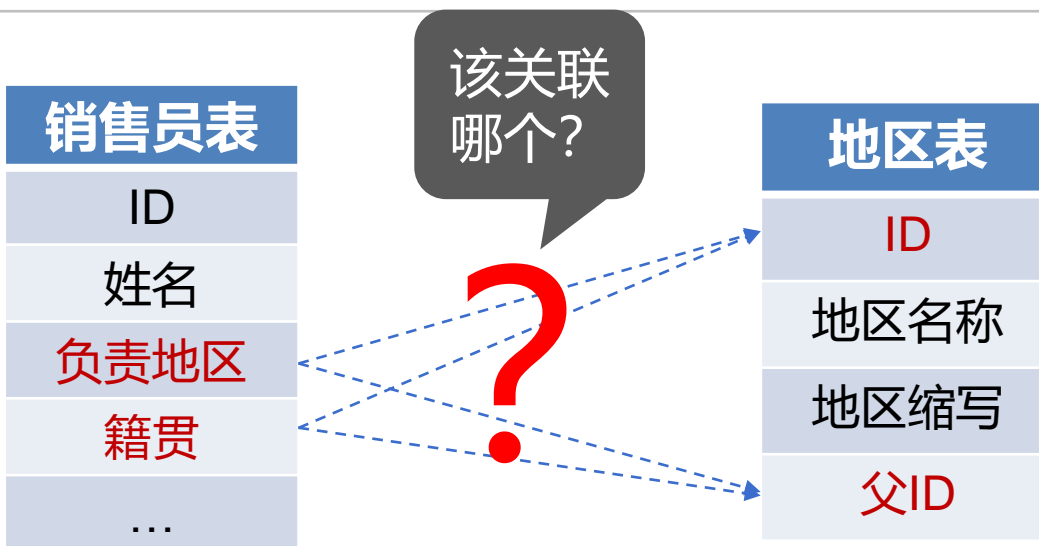
缺点

在页面上自动关联表只适用简单的情况，当遇到同维字段（如同一个表有2个以上地区字段）时就无法匹配了；自关联的情况也没法处理

页面自动关联的缺点

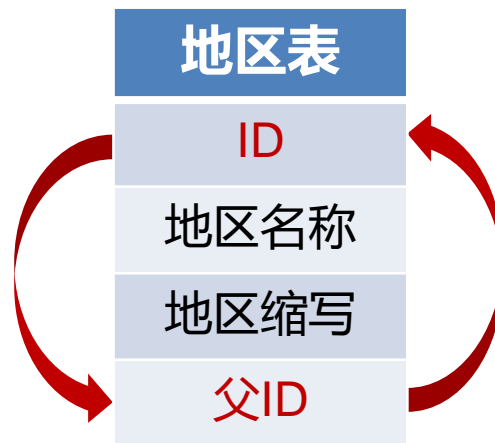
同维字段无法关联

销售表和地区表各有一对同维字段，无论是字段类型、长度、取值都相同，这时页面自动关联就不知所措了



自关联无法处理

地区表存在自关联的情况，如果想按层级汇总需要关联多次，这种情况页面自动关联是无法处理的



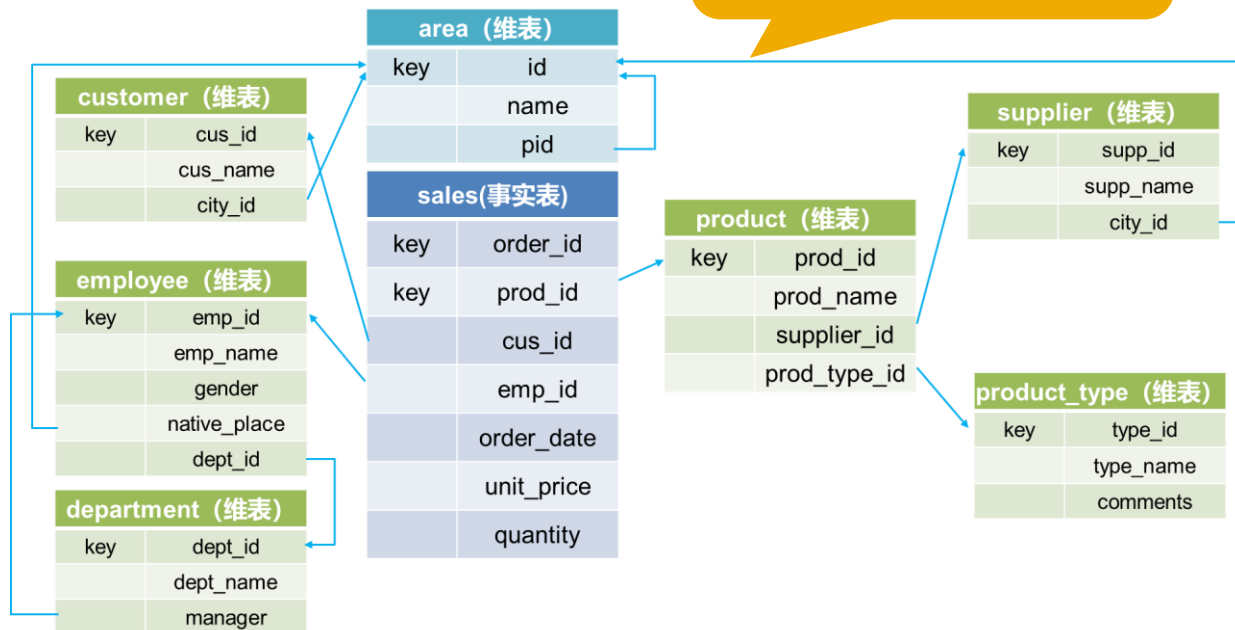
业界常见做法

3. 由用户指定关联

无法事先关联，也不能自动实现，自然想到把关联操作交给用户来完成，将网状结构的表间关联关系暴露给用户，让用户自行关联

缺点

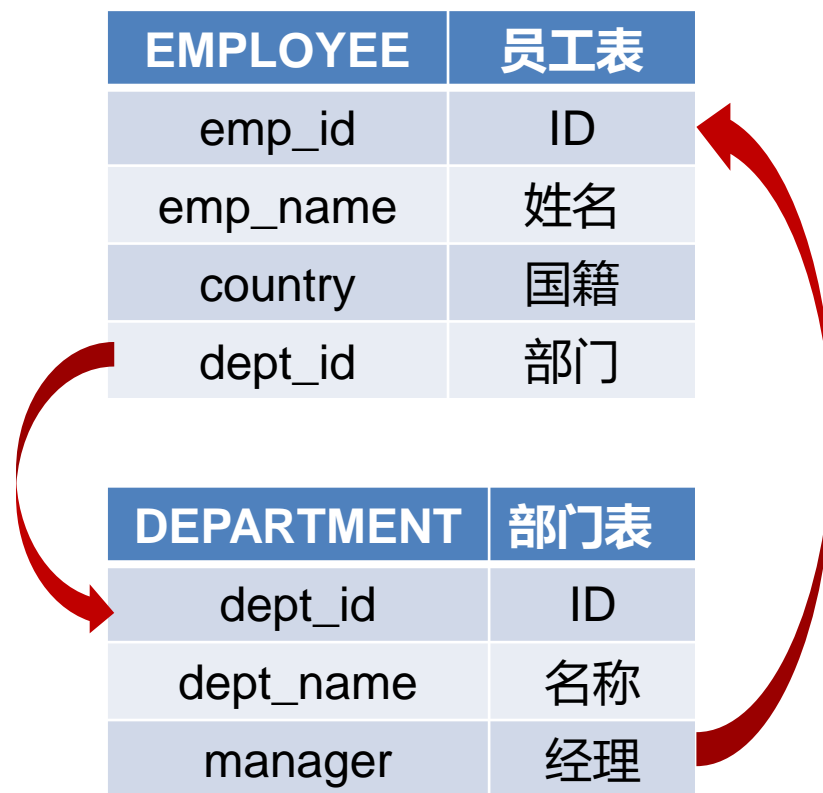
用户面对复杂的表间关系往往会比较“晕”，实在难以理解，即使把表和字段都翻译成业务汉语也很难理解表间关系，关联也就很难操作了



如何看待数据关联?

- 那造成“难以表达表间关联”的原因是什么呢?
- 我们一直沿用SQL看待数据关联的方式, 在SQL体系内不区分维表和事实表, 表和表是等同的。来看一下前面例子的SQL写法

```
-- 中国经理的美国员工
SELECT  e2.*
FROM    department d,
        employee e,
        employee e2
WHERE   d.manager = e.emp_id
        AND e.country = '中国'
        AND d.dept_id = e2.dept_id
        AND e2.country = '美国'
```



*这里用了2次employee (员工表) 关联查询得到了查询结果

换个思路看待表间关联

- 根据现在的新思路，将表进行显示地区分，将外键指向的表（字段）看做本表的一个属性
- 这样就可以把前面的例子写成这样：

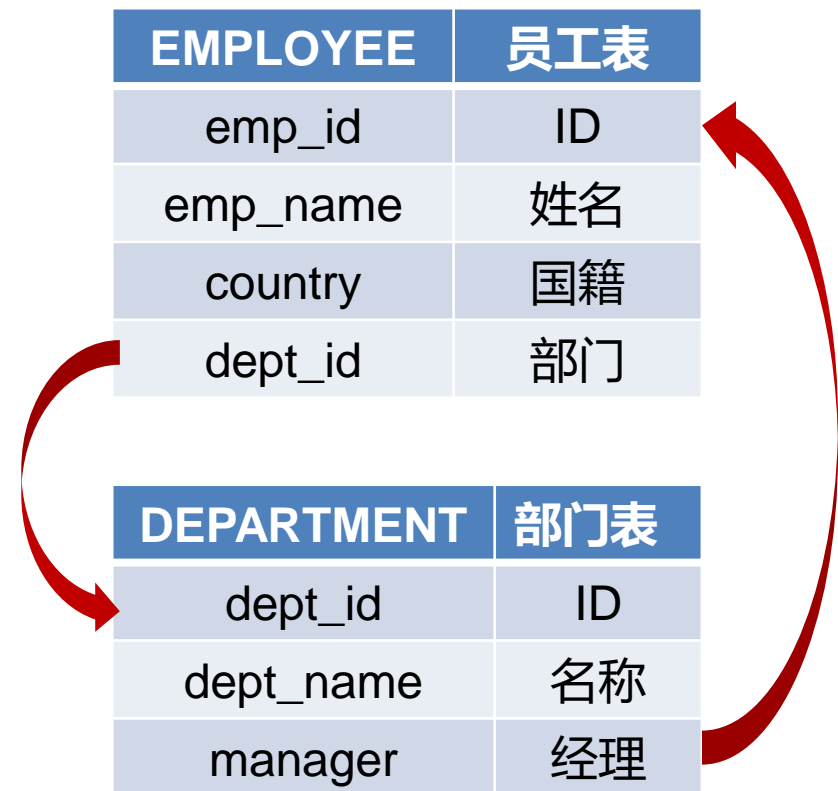
```
-- 中国经理的美国员工  
SELECT * FROM 员工表  
WHERE 国籍='美国' AND 部门.经理.国籍='中国'
```

DQL语法

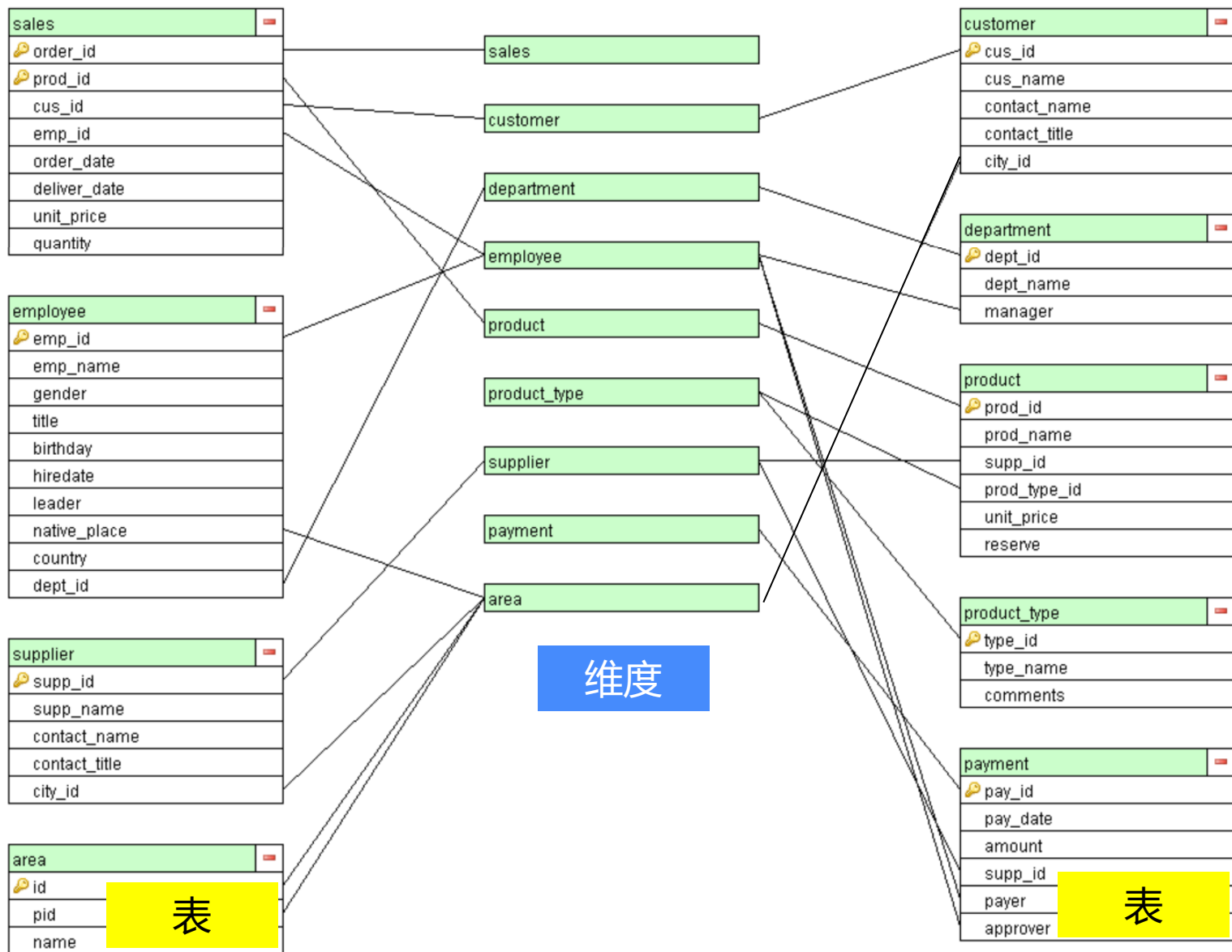
在这样的语法体系中，外键被看成了属性：

- 外键指向表的字段可直接用子属性的方式引用
- 允许多层和递归引用

【注】DQL是润乾多维分析的查询语言，全称是Dimensional Query Language，
是这种看待关联新思路的一种实现



以维度为中心的结构



- 将所有的表都定义成维度, 所有表都只和维度相连;
- 同一个维度可以作用到多个表上
- 从而形成以维度为中心的总线式结构 (不同于E-R结构)
- 这样结合外键属性化就很容易描述雪花结构中多层外键和自关联的情况

以这种结构开发出的页面

当我们通过将表看做维度、外键指向表的字段看做维度的属性后，页面就可以通过树状目录的方式逐级展开，直到最后一个层级（自关联递归引用）

pid	id	name
303	30302	石家庄
303	30303	秦皇岛
3	303	河北
401	40101	深圳
401	40102	广州
4	401	广东
402	40201	海口
4	402	海南
403	40301	厦门
403	40302	福州
4	403	福建
501	50101	西安
5	501	陕西
601	60101	成都
6	601	四川
602	60201	重庆
6	602	重庆
(Null)	1	华北
(Null)	2	华东
(Null)	3	华北
(Null)	4	华南
(Null)	5	西北
(Null)	6	西南

DB中的
地区表

比如根据销售记录表和自关联的地区表按地区、省、市三级汇总销售额

报表名称

地区:省:市

地区:省:市			销售额求和	总计
东北	辽宁	大连	55787	55787
		总计	55787	55787
	总计		55787	55787
华东	上海	上海	8662	8662
		总计	8662	8662
	山东	青岛	15224	15224
		总计	15224	15224
	江苏	南京	126007	126007
		常州	57949	57949
		总计	183956	183956
	江西	南昌	33920	33920
		总计	33920	33920
	总计		241762	241762
北京	北京	北京	56387	56387
		总计	56387	56387

数据分析设置器

销售记录

订单

产品

客户

客户名称

联系人姓名

联系人职位

所在城市

上级地区

上级地区

地区名称

地区名称

雇员

订单日期

发货日期

单价

数量

销售额

客户

客户编号

客户名称

联系人姓名

上表头

地区

省

市

销售额求和

+

润乾报表关联分析实现

通过以表作为维度、外键指向的表字段作为属性（外键属性化）以后，就比较容易实现基于雪花结构的关联分析了。在润乾报表中，要实现关联分析大概需要如下步骤：



举例-语义层建模

导入数据库表，根据表间关系，设置外键关联；定义字典

表名 物理表名

记录数 表描述

外键

序号	外键名称	对应表名
1	fk1	product
2	fk2	customer
3	fk3	employee

元数据

标题 表名

序号	名称	字段名	显示格式
1	订单	order_id	
2	产品	prod_id	
3	客户	cus_id	
4	雇员	emp_id	
5	订单日期	order_date	
6	发货日期	deliver_date	
7	单价	unit_price	
8	数量	quantity	
9	销售额	amount	
10	fk1	fk1	

字典

例子文件参考：
[8.relation_olap.lmd](#)
[8.relation_olap.dct](#)

举例-关联查询

发布语义层模型，在页面上查询前面的例子

例子：分别按客户所在地区、供应商所在地区、雇员籍贯汇总销售额

报表名称

地区

销售额求和

总计

华北	55787	55787
华东	241762	241762
华北	522710	522710
华南	304939	304939
西北	12412	12412
西南	202685	202685
总计	1340295	1340295

数据分析设置器

产品

产品名称

+

供应商

+

产品类别

单价

库存

客户

客户名称

联系人姓名

联系人职位

所在城市

上级地区

+

上级地区

地区名称

地区名称

雇员

订单日期

发货日期

单价

数量

销售额

客户

地区

上表头

销售额求和

+

举例-关联查询

发布语义层模型，在页面上查询前面的例子

例子：分别按客户所在地区、供应商所在地区、雇员籍贯汇总销售额

报表名称

供应商所在

销售额求和

总计

华北	55787	55787
华东	241762	241762
华北	522710	522710
华南	304939	304939
西北	12412	12412
西南	202685	202685
总计	1340295	1340295

数据分析设置器

供应商

供应商名称

联系人姓名

联系人职务

所在城市

上级地区

上级地区

地区名称

地区名称

产品类别

单价

库存

客户

客户名称

联系人姓名

联系人职位

所在城市

雇员

订单日期

发货日期

单价

数量

销售额

上表头

供应商所在地区

销售额求和

举例-关联查询

发布语义层模型，在页面上查询前面的例子

例子：分别按客户所在地区、供应商所在地区、雇员籍贯汇总销售额

报表名称

雇员地区

销售额求和

总计

华北	55787	55787
华东	241762	241762
华北	522710	522710
华南	304939	304939
西北	12412	12412
西南	202685	202685
总计	1340295	1340295

数据分析设置器

所在城市

雇员

籍贯

上级地区

上级地区

上级地区

地区名称

地区名称

地区名称

国籍

部门

订单日期

发货日期

单价

数量

销售额

雇员地区

销售额求和

上表头
雇员地区
销售额求和

举例-关联查询

例子：查询中国经理的美国员工

报表名称

姓名 国籍 经理 经理国籍

杜邦	美国	李强	中国
詹姆斯	美国	李强	中国
罗伯特	美国	李强	中国
乔治	美国	李强	中国

数据分析设置器

雇员

雇员编号
姓名
性别
职位
出生日期
入职日期
直属领导
籍贯
国籍
部门
部门名称
经理
姓名
性别
职位
出生日期
入职日期
直属领导
籍贯
国籍
部门

左表头
上表头
指标标题
姓名
国籍
经理
经理国籍
+

举例-关联查询

例子：查询中国经理的美国员工

设置查询条件 名称

数据分析设置器

☐ 简洁显示 (简洁显示会隐藏掉“新增条件、调整条件层次”的功能, 更能凸显条件之间的层次关系)

→

雇员.部门.经理.国籍

▼

等于

中国

✕

AND | OR

→

雇员.国籍

▼

等于

美国

✕

AND | OR

保存

清除条件并退出

取消

上表头

指标标题

姓名 ▼

国籍 ▼

经理 ▼

经理国籍 ▼

+



事实星座

知识点-事实星座

- 除了星型结构和雪花型结构，多维分析中还包括一种事实星座结构
- 当多个事实表共用维表时就形成了事实星座结构，从广义上来讲，事实星座是星型结构和雪花型结构的延伸
- 事实星座可以继续通过上述手段来完成关联分析（同一个维度被多个表指向），此处不再赘述感兴趣的同学可以尝试一下



小结

- 本节课我们学习了基于雪花型结构的多维分析，并了解了实现雪花型多维分析的意义
- 相比星型结构，雪花型结构更合理
- 在多维分析中应用星型结构主要因为很难让业务人员描述多表关联关系
- 通过润乾多维分析举例，实现雪花型结构的多维分析（关联分析）

Summary

THANKS

感谢观看



乾学院 · VIP课堂

商业智能 复杂报表技术



CONTENTS

本章内容

1. 什么是复杂报表
2. 复杂报表的特征
3. 复杂报表开发技术对比
4. 多源分片报表设计（练习）
5. 数据源开发技术
6. 性能与容量





什么是复杂报表

什么是复杂报表

- 复杂报表在应用中很常见，几个典型特征：格式复杂、信息量大、表头乱等
- 报表内容根据数据变化而变化，分片区呈现（静态、动态区域任意混合编排）



复杂报表若干例子



中国通信服务用工分类月报

指标名称	序号	小计	主业划转	社会招聘	其他从业	派遣制	指标名称	序号	小计	主业划转	社会招聘	其他从业	派遣制
甲	乙	1	2	3	4	5	甲	乙	1	2	3	4	5
一.按文化程度分类	1	18521	6386	10218	333	15844	4.助理级	29	1092	472	614	1	5
1.研究生及以上	2	574	122	435	13	45	5.员级	30	270	166	102	0	2
其中:博士	3	24	2	22	0	0	五.按业务版块分类	31	9786	3268	5232	387	899
硕士	4	550	120	413	13	4	(一) 综合管理	32	2399	1111	1206	6	76
2.本科(含双学士)	5	5487	1063	4003	98	323	1.企业经营管理	33	438	331	106	1	0
3.大专	6	3988	1337	2110	28	513	2.行政综合	34	599	296	266	5	32
4.中专及以下	7	7898	3742	3235	181	740	3.人力资源	35	190	69	116	0	5
二.按年龄分类	8	10174	3387	5430	387	970	4.财务	36	279	80	186	0	13
1.24岁及以下	9	52	0	26	3	23	5.审计	37	25	9	16	0	0
2.25-29岁	10	819	0	588	89	142	6.IT管理	38	53	11	38	0	4
3.30-34岁	11	1865	0	1610	6	249	7.党群	39	92	60	32	0	0
4.35-39岁	12	1918	30	1682	6	200	8.市场	40	691	232	437	0	22
5.40-44岁	13	1317	322	847	53	95	9.战略与发展	41	32	23	9	0	0
6.45-49岁	14	1268	720	388	98	62	(二) 网络建设	42	3221	942	1888	0	391
7.50-54岁	15	1051	773	167	70	41	1.工程施工	43	1834	795	795	0	244
8.55岁及以上	16	1884	1542	122	62	158	2.工程设计	44	760	84	592	0	84
三.按合同期限分类	17	18954	4783	11874	146	2151	3.工程监理	45	627	63	501	0	63
1.1年以内(不含1年)	18	3118	207	2682	0	229	(三) 外包服务	46	1446	615	522	55	254
2.1到5年(含5年)	19	9664	278	7464	2	1920	1.维护	47	590	236	233	55	66
3.5-10年(含10年)	20	512	201	309	0	2	2.设施管理	48	565	345	137	0	83
4.10年以上	21	100	100	0	0	0	3.渠道	49	291	34	152	0	105
5.无固定期限	22	5456	3997	1418	41	0	(四) 内容应用及其他	50	2143	393	1335	326	89
6.未签订合同	23	0	0	0	0	0	1.IT应用	51	828	7	717	98	6
7.完成一定工作期限	24	104	0	1	103	0	2.语音增值	52	343	31	287	1	24
四.专业技术职务分类	25	3096	1537	1470	638	641	3.互联网服务	53	83	12	69	0	2
1.正高级	26	25	22	3	0	0	4.培训业务	54	94	32	40	0	22
2.副高级	27	363	241	122	0	0	5.其他	55	795	311	222	227	35
3.中级	28	1346	636	707	0	3	(五) 辅助岗位	56	36	0	5	0	31
							(六) 综合业务	57	541	207	276	0	58

复杂报表若干例子

零售

National -- FYAug'11 counter Offtake summary						Reporting Date: 1st~ 31st Aug.2011		Completion Rate		Reporting Date: 1st~ 31st Aug.2011	Workweeks this Month	
						South Completion Rate		7780654%	0%	Available Weeks	4	100%
						North Completion Rate		19697688%	0%	Elapsed Weeks	4	100%
						National Completion Rate		13739171%	0%	Remaining Weeks	0	0%
City	Sales	2011-07					2011-08					
		美发—July'11	须刀—July'11	Braun—July'11	—July'11	TTL—July'11	美发—Aug'11	须刀—Aug'11	Braun—Aug'11	—Aug'11	TTL—Aug'11	
National Summary	Target销售目标	84.00	46.00	130.00	0.00	130.00	84.00	46.00	130.00	0.00	130.00	
(Only Counter)	Sales实际销售金额	13,342,598.59	6,671,299.29	20,013,897.88	0.00	20,013,897.88	11,907,281.74	5,953,640.87	17,860,922.60	0.00	17,860,922.60	
	Ach.%=实际销售金额/目标	15884045.94%	14502824.55%	15395306.06%	0.00%	15395306.06%	14175335.40%	12942697.54%	13739171.23%	0.00%	13739171.23%	
	YA去年销售金额	39,245,259.26	39,245,259.26	78,490,518.51	39,245,259.26	117,735,777.77	34,199,432.70	34,199,432.70	68,398,865.41	34,199,432.70	102,598,298.11	
	YA%同期金额对比=今年/去年	34.00%	17.00%	25.50%	0.00%	17.00%	34.82%	17.41%	26.11%	0.00%	17.41%	
	Target	42.00	23.00	65.00	0.00	65.00	42.00	23.00	65.00	0.00	65.00	
	Sales	10,334,730.43	5,167,365.22	15,502,095.65	0.00	15,502,095.65	8,535,664.97	4,267,832.48	12,803,497.45	0.00	12,803,497.45	
	Ach.%	24606501.03%	22466805.29%	23849377.92%	0.00%	23849377.92%	20323011.82%	18555793.40%	19697688.38%	0.00%	19697688.38%	
	YA	19,622,629.63	19,622,629.63	39,245,259.26	19,622,629.63	58,867,888.88	17,099,716.35	17,099,716.35	34,199,432.70	17,099,716.35	51,299,149.06	
北区	YA%	52.67%	26.33%	39.50%	0.00%	26.33%	49.92%	24.96%	37.44%	0.00%	24.96%	
	Target	42.00	23.00	65.00	0.00	65.00	42.00	23.00	65.00	0.00	65.00	
	Sales	3,007,868.16	1,503,934.08	4,511,802.23	0.00	4,511,802.23	3,371,616.77	1,685,808.39	5,057,425.16	0.00	5,057,425.16	
	Ach.%	7161590.85%	6538843.82%	6941234.21%	0.00%	6941234.21%	8027658.98%	7329601.68%	7780654.09%	0.00%	7780654.09%	
南区	YA	19,622,629.63	19,622,629.63	39,245,259.26	19,622,629.63	58,867,888.88	17,099,716.35	17,099,716.35	34,199,432.70	17,099,716.35	51,299,149.06	
	YA%	15.33%	7.66%	11.50%	0.00%	7.66%	19.72%	9.86%	14.79%	0.00%	9.86%	
	Target	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Sales	6,413,961.31	3,206,980.66	9,620,941.97	0.00	9,620,941.97	649,138.73	324,569.37	973,708.10	0.00	973,708.10	
长春	Ach.%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	YA	19,622,629.63	19,622,629.63	39,245,259.26	19,622,629.63	58,867,888.88	17,099,716.35	17,099,716.35	34,199,432.70	17,099,716.35	51,299,149.06	
	YA%	32.69%	16.34%	24.51%	0.00%	16.34%	3.80%	1.90%	2.85%	0.00%	1.90%	
	Target	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Sales	3,007,868.16	1,503,934.08	4,511,802.23	0.00	4,511,802.23	3,371,616.77	1,685,808.39	5,057,425.16	0.00	5,057,425.16	
	Ach.%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	YA	19,622,629.63	19,622,629.63	39,245,259.26	19,622,629.63	58,867,888.88	17,099,716.35	17,099,716.35	34,199,432.70	17,099,716.35	51,299,149.06	
	YA%	15.33%	7.66%	11.50%	0.00%	7.66%	19.72%	9.86%	14.79%	0.00%	9.86%	
长沙	Target	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Sales	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Ach.%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	YA	19,622,629.63	19,622,629.63	39,245,259.26	19,622,629.63	58,867,888.88	17,099,716.35	17,099,716.35	34,199,432.70	17,099,716.35	51,299,149.06	
成都	YA%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	Target	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Sales	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Ach.%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
	YA	19,622,629.63	19,622,629.63	39,245,259.26	19,622,629.63	58,867,888.88	17,099,716.35	17,099,716.35	34,199,432.70	17,099,716.35	51,299,149.06	
	YA%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	

复杂报表若干例子



PH1 Through Put [sheet/1hr] (2011年6月実績)

82s/h 以 Tact 42.0 sec
75s/h 以 Tact 48.0 sec
60s/h 以 Tact 60.0 sec

LOD投入数

日付	Line	昼勤													夜勤										投入数			Through Put [sheet/1hr]								
		8:00-	9:00-	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	24:00	1:00-	2:00-	3:00-	4:00-	5:00-	6:00-	7:00-	昼勤	夜勤	合計	1位	2位	3位	4位	昼勤Average	夜勤Average	Average	
2011-06-01	BM1	0	0	0	6	0	0	0	3	0	0	15	0	0	18	47	31	78	41	62	64	80	53	38	0	24	512	536	80	78	64	62	2	43	22	
	RE1	41	56	50	71	62	38	73	67	71	62	51	59	68	69	67	71	54	74	47	67	75	61	74	26	701	753	1454	75	74	73	71	58	63	61	
	BL1	41	76	58	83	18	0	0	32	15	0	58	58	84	74	54	37	59	56	47	73	49	83	72	26	439	714	1153	84	83	76	74	37	60	48	
	GR1	0	0	16	0	24	54	67	66	25	30	61	57	85	53	76	66	57	68	77	46	79	57	80	12	400	756	1156	85	80	79	77	33	63	48	
	PS1	74	72	50	66	53	68	83	63	75	82	68	82	83	30	0	0	9	82	85	82	54	44	24	76	836	569	1405	85	83	82	76	70	47	59	
2011-06-02	BM1	0	0	17	77	76	40	46	61	76	78	49	60	2	0	0	3	0	0	23	0	0	0	17	79	580	124	704	79	78	77	76	48	10	29	
	RE1	40	0	0	49	11	0	6	0	0	0	7	0	22	0	0	36	18	70	32	46	77	4	61	77	113	443	556	77	70	61	49	9	37	23	
	BL1	55	56	67	82	53	83	59	77	83	69	12	0	0	0	0	0	79	47	58	49	79	67	70	76	696	525	1221	83	82	79	77	58	44	51	
	GR1	0	36	78	70	61	78	75	67	70	44	60	71	73	48	38	62	83	70	56	60	64	58	58	28	710	698	1408	83	78	75	73	59	58	59	
	PS1	84	56	54	82	81	83	83	52	82	55	27	0	1	63	85	82	0	1	52	46	42	50	83	85	739	590	1329	85	84	83	82	62	49	55	
2011-06-03	BM1	14	45	43	78	77	31	78	67	77	79	69	69	69	0	0	0	0	43	53	59	70	51	7	60	727	412	1139	79	78	77	70	61	34	47	
	RE1	0	14	0	0	41	82	70	63	57	83	68	72	76	54	82	69	46	0	0	0	0	0	0	0	550	327	877	83	82	76	72	46	27	37	
	BL1	0	13	0	7	0	0	46	50	0	0	0	0	0	0	0	21	0	0	5	0	0	0	0	0	116	26	142	50	46	21	13	10	2	6	
	GR1	5	0	4	0	0	0	29	45	29	0	0	1	19	0	1	0	58	70	82	65	77	73	8	0	113	453	566	82	77	73	70	9	38	24	
	PS1	52	82	81	56	9	0	0	0	0	0	0	0	0	0	0	0	82	84	85	56	66	85	84	83	280	625	905	85	84	83	82	23	52	38	

小结

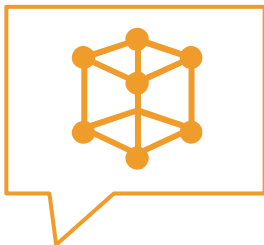


• 格式复杂、信息量大意味着什么？



格式复杂

一般表头比较乱，三五层的大格套小格很常见(还有著名的斜线)，没有格线容易看不清



格式复杂意味着要有好的工具可以画出来

信息量大

人们总是期望在一张报表上获得更多的信息，往往把各种各样的数据都集中在一张纸上



信息量大意味着大量的、复杂的统计运算



业务人员无法用自助报表做出来



复杂报表的特征

复杂报表的特征



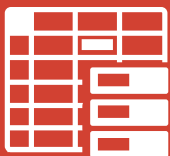
多源分片

数据来源多个数据表或数据库
在报表中以多分片形式呈现



不规则分组

支持按段分组/重叠分组/
归并分组等复杂分组形式



动态格间计算

根据格子坐标引用其他格
方便计算同比/环比/累积/排名



行列对称

行列属性一致
纵向和横向同时扩展

多源分片—举例

多源

著名的斜线

产品表

客户表

多片区

客户金额

已签合同

已发货

年度合同金额

已付款

合同表

客户			汇总纬度	化工				食品	
				苯乙酮	苯甲酸钠	苯磺酸	对氯二苯	大豆	色拉油
按客户汇总	国内客户	无锡食品加工厂	金额	400,000	500,000	800,000	500,000	900,000	700,000
			数量	1,540	1,530	4,630	2,631	3,125	2,902
		天津长城化工厂	金额	816,000	1,020,000	1,632,000	1,020,000	1,836,000	1,428,000
			数量	3,080	3,060	9,260	5,262	6,250	5,804
		中国化工原料厂	金额	94,400	118,000	188,800	118,000	212,400	165,200
			数量	1,540	1,530	4,630	2,631	3,125	2,902
	国外客户	Crystal	金额	3,142,400	3,928,000	6,284,800	3,928,000	7,070,400	5,499,200
			数量	4,620	4,590	13,890	7,893	9,375	8,706
		Chemist	金额	481,600	602,000	963,200	602,000	1,083,600	842,800
			数量	3,080	3,060	9,260	5,262	6,250	5,804
		Murli	金额	164,400	205,500	328,800	205,500	369,900	287,700
			数量	3,080	3,060	9,260	5,262	6,250	5,804
已签合同			金额	5,314,600	6,643,250	10,629,200	6,643,250	11,957,850	9,300,550
			数量	20,020	19,890	60,190	34,203	40,625	37,726
已发货			金额	816,000	1,020,000	1,632,000	1,020,000	1,836,000	1,428,000
			数量	1,540	1,530	4,630	2,631	3,125	2,902
按合同年度汇总	2001	金额	174,000	217,500	348,000	217,500	391,500	304,500	
		数量	3,080	3,060	9,260	5,262	6,250	5,804	
	2002	金额	441,800	552,250	883,600	552,250	994,050	773,150	
		数量	3,080	3,060	9,260	5,262	6,250	5,804	
	2003	金额	4,012,000	5,015,000	8,024,000	5,015,000	9,027,000	7,021,000	
		数量	9,240	9,180	27,780	15,786	18,750	17,412	
	2004	金额	686,800	858,500	1,373,600	858,500	1,545,300	1,201,900	
		数量	4,620	4,590	13,890	7,893	9,375	8,706	
已付款			金额	0	0	0	0	0	0
			数量	0	0	0	0	0	0

多源分片—知识点

多源

报表的数据来源来自多个数据表或视图，甚至是多个数据库

分片

报表的纵向或横向或双向同时被分成了多个区域，每个区域规则不同，而又可能相互运算



多源往往带来分片，正是由于分片的特征，使得报表设计必须直接基于多源进行，而不能简单的将多源转成单源！

不规则分组—不完全分组—举例



只关心重点区域、城市的销售额，非重要城市归为其它

地区	城市	销售额
东北	大连	¥ 86056.70
华北	北京	¥ 83795.99
	秦皇岛	¥ 60369.21
	石家庄	¥ 59649.52
	天津	¥ 554732.33
	张家口	¥ 33434.70
华东	常州	¥ 102282.55
	南昌	¥ 57244.20
	南京	¥ 182536.39
	青岛	¥ 20965.26
	上海	¥ 10219.23
	温州	¥ 76812.50
华南	海口	¥ 35604.91
	深圳	¥ 353504.06
	厦门	¥ 12241.67
西北	西安	¥ 21538.90
西南	成都	¥ 2273.24
	重庆	¥ 312385.39



地区	城市	销售额
华北	北京	¥ 83795.99
	秦皇岛	¥ 60369.21
	石家庄	¥ 59649.52
	天津	¥ 554732.33
	张家口	¥ 33434.70
华东	常州	¥ 102282.55
	南昌	¥ 57244.20
	南京	¥ 182536.39
	青岛	¥ 20965.26
	上海	¥ 10219.23
	温州	¥ 76812.50
其它	成都	¥ 2273.24
	大连	¥ 86056.70
	海口	¥ 35604.91
	深圳	¥ 353504.06
	西安	¥ 21538.90
	厦门	¥ 12241.67
	重庆	¥ 312385.39



不规则分组—枚举分组—举例

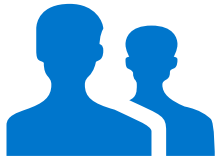


人为地把某几组合并成一组，且合并规则并无规律，需要逐组枚举的报表，我们称为枚举分组

地区	城市	销售额
东北	大连	¥ 86056.70
华北	北京	¥ 83795.99
	秦皇岛	¥ 60369.21
	石家庄	¥ 59649.52
	天津	¥ 554732.33
	张家口	¥ 33434.70
华东	常州	¥ 102282.55
	南昌	¥ 57244.20
	南京	¥ 182536.39
	青岛	¥ 20965.26
	上海	¥ 10219.23
	温州	¥ 76812.50
华南	海口	¥ 35604.91
	深圳	¥ 353504.06
	厦门	¥ 12241.67
西北	西安	¥ 21538.90
西南	成都	¥ 2273.24
	重庆	¥ 312385.39

地区	城市	销售额
北方地区	北京	¥ 83795.99
	大连	¥ 86056.70
	秦皇岛	¥ 60369.21
	石家庄	¥ 59649.52
	天津	¥ 554732.33
	张家口	¥ 33434.70
南部地区	成都	¥ 2273.24
	海口	¥ 35604.91
	深圳	¥ 353504.06
	厦门	¥ 12241.67
	重庆	¥ 312385.39
	常州	¥ 102282.55
其他	南昌	¥ 57244.20
	南京	¥ 182536.39
	青岛	¥ 20965.26
	上海	¥ 10219.23
	温州	¥ 76812.50
	西安	¥ 21538.90

不规则分组—重叠分组—举例



同一部分记录在不同组中被重复统计的分组报表称为重叠分组

不良贷款情况分析表

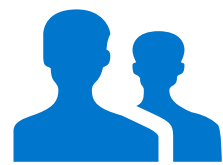
固定列分组

制表日期 2006年03月12日

地区	贷款余额	不良贷款 余额	不良贷款内部原因				
			合计	“三查” 不落实	决策失误	人为缩短贷款 期限	其他
北京	1,970,955.00	1,372,169.00	1,372,169.00	234,455.00	345,645.00	134,534.00	657,535.00
天津	28,721,313.00	21,067,924.00	21,067,924.00	7,544,534.00	875,646.00	3,890,998.00	8,756,746.00
河北	98,767,579.00	98,767,579.00	98,767,579.00	98,767,579.00	0.00	0.00	0.00
山东	1,345,654.00	1,345,654.00	1,345,654.00	1,345,654.00	0.00	0.00	0.00
其中 青岛	465,465.00	465,465.00	465,465.00	465,465.00	0.00	0.00	0.00
山西	89,676,786.00	89,676,786.00	89,676,786.00	89,676,786.00	0.00	0.00	0.00

山东和青岛
有重复部分

不规则分组—按段分组—举例



按值段进行分组的分组报表被称为按段分组

按部门两级分组

年龄区间只能枚举

部门		年龄段						性别	
		0-8岁	8-15岁	15-27岁	27-33岁	33-49岁	49岁以上	男	女
化工产品部	化工业务一部	0	0	0	2	0	0	1	1
	化工业务二部	0	0	0	2	0	0	0	2
	化工业务三部	0	0	0	2	0	0	2	0
	其他	0	0	0	0	1	0	0	1
	合计	0	0	0	6	1	0	3	4
食品销售部	食品一部	0	0	0	2	0	0	0	2
	食品二部	0	0	2	0	0	0	2	0
	食品三部	0	0	2	0	0	0	0	2
	其他	0	0	0	2	0	0	2	0
	合计	0	0	4	4	0	0	4	4
后勤部	财务部	0	0	2	0	1	0	3	0
	行政部	0	0	2	0	0	0	0	2
	其他	0	0	0	0	0	0	0	0
	合计	0	0	4	0	1	0	3	2

不规则分组—知识点



- 不规则分组是相对于规则分组而言，传统工具中的数据分组是模拟了SQL中的GROUP BY 运算，即分组标准一致且有规则（一般都按某个字段或表达式）
- 不规则分组是指分组标准看不出规律（往往是用户最关注的数据角度，只能穷举），所有事实不一定全部出现在分组结果中、个别事实还可能重复出现，次序也与原数据记录无关

自由格间计算—跨行组运算—举例



比上期（跨行）和比去年同期（跨组）就是典型的跨行组运算；同期比的难点在于如何准确的定位到去年同期

行

历年签约对比表

组

年	季	当期合同额	比上期	比去年同期
2001	1	0.0		
	2	120895.00	∞%	
	3	161081.00	133.24%	
	4	0.0	0.00%	
2002	1	0.0	∞%	∞%
	2	889282.00	∞%	735.58%
	3	1001285.00	112.59%	621.60%
	4	169584.00	16.94%	∞%
2003	1	0.0	0.00%	∞%
	2	2041479.00	∞%	229.56%
	3	3492677.00	171.09%	348.82%
	4	1147466.00	32.85%	676.64%

自由格间计算—独立格运算—举例



独立格可以自由引用表格中任意数据进行运算，比如带了某种条件（前三名），还可能会引用到报表外的数据（北京市的全部销售额不在报表中）

跨行运算计算排名

城八区电信销售统计表								
地区	合计		固定电话		小灵通		移动电话	
	销售额	排名	销售额	排名	销售额	排名	销售额	排名
朝阳	972,102,181.00	1	9,879,866.00	2	897,868,567.00	1	64,353,748.00	4
崇文	194,828,511.00	2	8,567,857.00	4	98,686,090.00	2	87,574,564.00	2
东城	12,516,819.00	8	3,452,345.00	8	8,967,578.00	5	96,896.00	8
丰台	72,399,706.00	5	7,646,346.00	5	78,575.00	8	64,674,785.00	3
海淀	101,782,940.00	4	6,345,768.00	6	89,768,585.00	3	5,668,587.00	7
石景山	161,115,195.00	3	9,879,868.00	1	63,547,869.00	4	87,687,458.00	1
西城	13,003,779.00	7	5,668,746.00	7	867,575.00	7	6,467,458.00	6
宣武	22,898,007.00	6	9,876,757.00	3	6,485,896.00	6	6,535,354.00	5

前三名城销售总额和：1,328,045,887.00
城八区占北京市的销售比例：79.49%

表内的跨行
独立格运算

占总额比例：85.64%

自由格间计算—序号运算—举例



- 按照城市分组后，对每个城市里所有公司进行局部编号
- 按照城市分组后，对所有公司进行全局编号

序号	城市	公司名称	联系人姓名	联系人职务
1	北京	仪和贸易	王先生	市场经理
2		兰格英语	王先生	结算经理
3		留学服务中	赵小姐	市场助理
1	常州	迈策船舶	王俊元	物主
2		文成	唐小姐	销售代表
3		顶		市场经理
4		物主		物主
1	成都		王	市场助理
1	大连	国皓	黄雅玲	市场经理
2		威航货运有	刘先生	销售代理
3		三捷实业	王先生	市场经理
4		五金机械	苏先生	销售代表
5		华科	吴小姐	市场助理



序号	城市	公司名称	联系人姓名	联系人职务
1	北京	仪和贸易	王先生	市场经理
2		兰格英语	王先生	结算经理
3		留学服务中	赵小姐	市场助理
4	常州	迈策船舶	王俊元	物主
5		文成	唐小姐	销售代表
6		顶上系统		市场经理
7		椅天文格		市场助理
8	成都	新巨企业	王	市场助理
9	大连	国皓	黄雅玲	市场经理
10		威航货运有	刘先生	销售代理
11		三捷实业	王先生	市场经理
12		五金机械	苏先生	销售代表
13		华科	吴小姐	市场助理



自由格间计算—知识点



自由格间运算，**特指浮动行列**的情况，即行列均为动态扩展出来的，此时行间、列间、甚至组间的运算。基本特征如下：

- 行列浮动
- 运算可能跨行、跨列，还可能跨组
- 独立格运算，即位于扩展区域之外的某些独立格，需要引用扩展区域内的单元格的值
- 序号，为扩展出来的格子编号，例如组内编号、统一编号等

行列对称—知识点

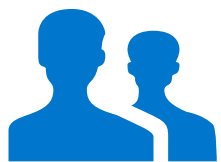


- 所有以上的讨论中还贯穿了一点，即**行列对称**
- 纵方向上拥有的自动复制扩展的能力需要完全实现在横方向上
- 报表可以横向分组、横向分片，对于横向变列的报表可以定义跨列组运算等



复杂报表开发技术对比

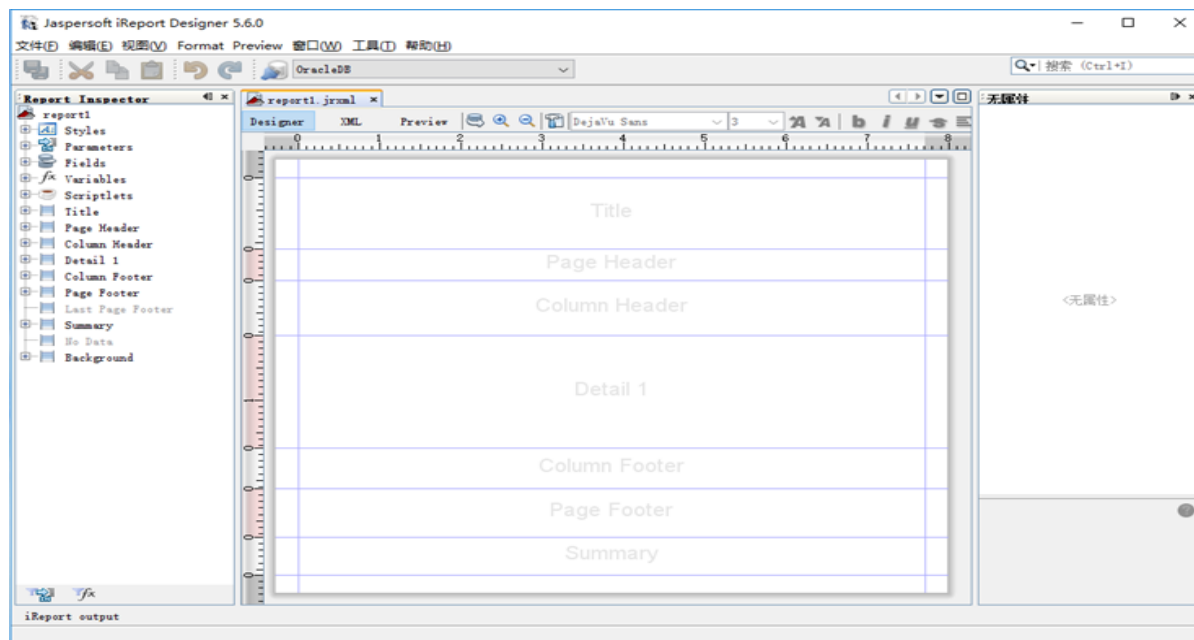
复杂报表开发技术对比



- 在“基础报表技术”中，我们讨论过两类报表模型与布局上的不同，接下来我们看一下，在复杂报表开发技术方面的对比

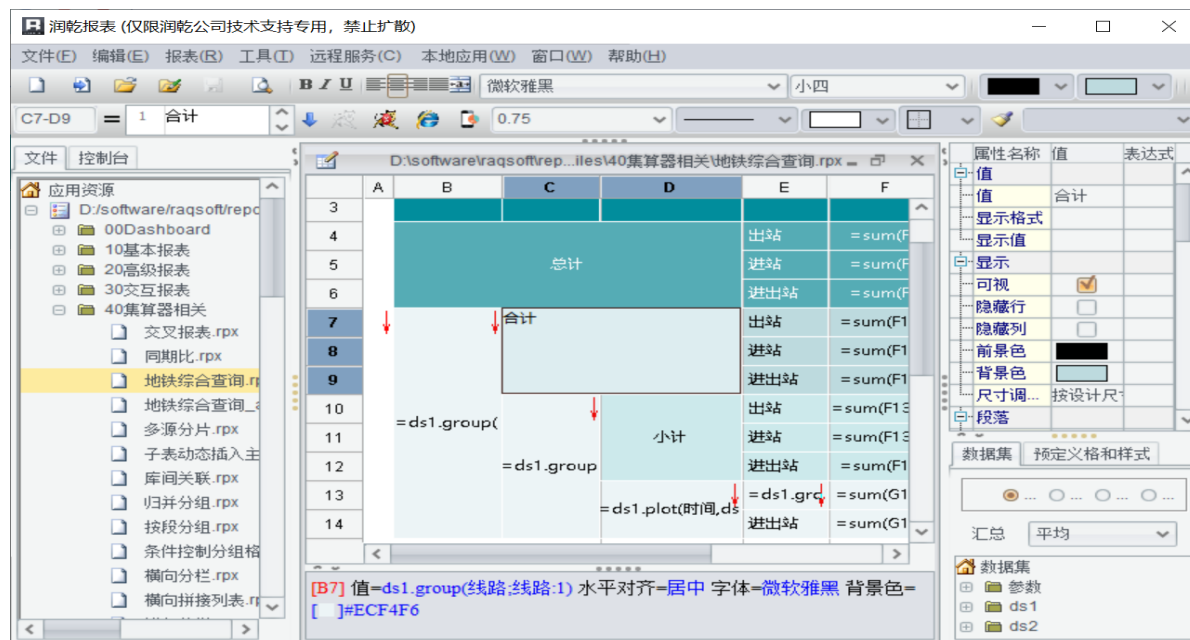
国外传统类

- 条带模型，控件拖拽式布局



新型网格类

- 类Excel设计模型

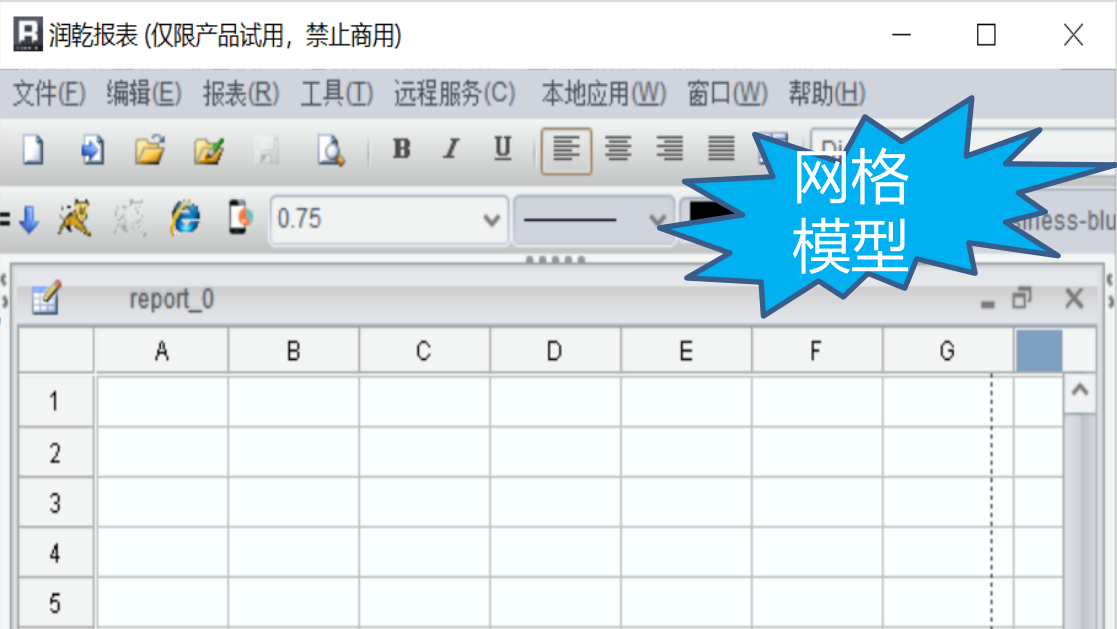
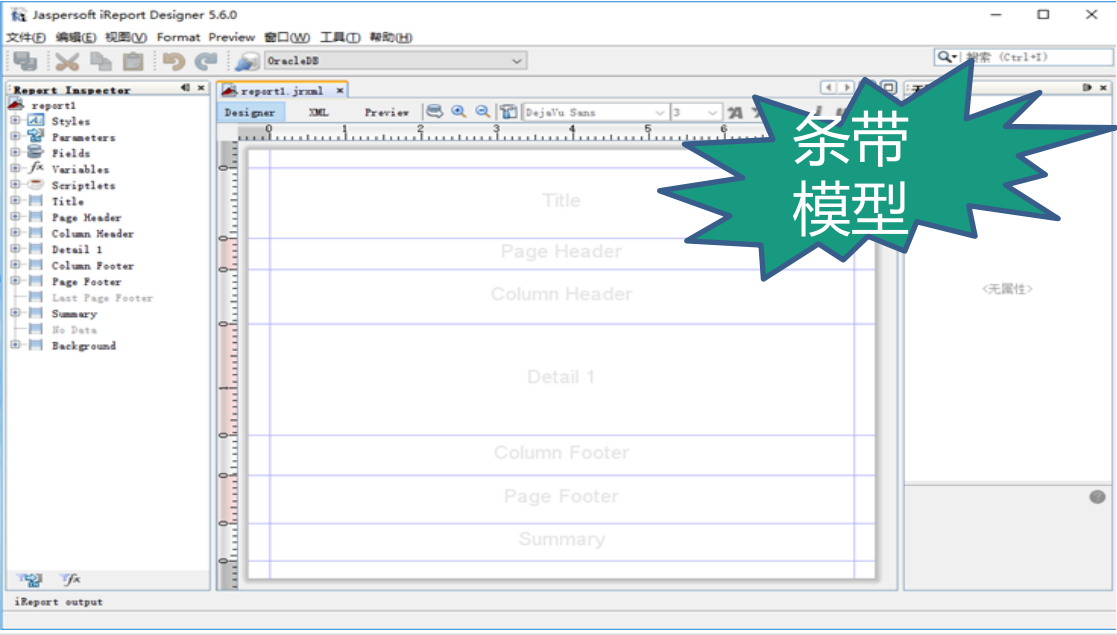


复杂报表开发技术对比—多源分片



这种多源分片的报表怎么样做？

客户			汇总纬度	化工				食品	
				苯乙酮	苯甲酸钠	苯磺酸	对氯二甲	大豆	色拉油
按客户汇总	国内客户	无锡食品加工厂	金额	400,000	500,000	600,000	500,000	900,000	700,000
			数量	1,540	1,530	4,630	2,631	3,125	2,902
		天津长城化工厂	金额	816,000	1,020,000	1,632,000	1,020,000	1,836,000	1,428,000
			数量	3,080	3,060	9,260	5,262	6,250	5,804
	国外客户	中国化工原料厂	金额	94,400	118,000	188,800	118,000	212,400	165,200
			数量	1,540	1,530	4,630	2,631	3,125	2,902
		Crystal	金额	3,142,400	3,928,000	6,284,800	3,928,000	7,070,400	5,499,200
			数量	4,620	4,590	13,890	7,893	9,375	8,708
		Chemist	金额	481,600	602,000	963,200	602,000	1,083,600	842,800
			数量	3,080	3,060	9,260	5,262	6,250	5,804
按合同年度汇总	已签合同	Murli	金额	164,400	205,500	328,800	205,500	369,900	287,700
			数量	3,080	3,060	9,260	5,262	6,250	5,804
	已发货		金额	5,314,600	6,643,250	10,629,200	6,643,250	11,957,850	9,300,550
			数量	20,020	19,890	60,190	34,203	40,625	37,726
	2001		金额	174,000	217,500	348,000	217,500	391,500	304,500
			数量	3,080	3,060	9,260	5,262	6,250	5,804
		2002	金额	441,800	552,250	883,600	552,250	994,050	773,150
			数量	3,080	3,060	9,260	5,262	6,250	5,804
		2003	金额	4,012,000	5,015,000	8,024,000	5,015,000	9,027,000	7,021,000
			数量	9,240	9,180	27,780	15,786	18,750	17,412
		2004	金额	686,800	858,500	1,373,600	858,500	1,545,300	1,201,900
			数量	4,620	4,590	13,890	7,893	9,375	8,708
已付款			金额	0	0	0	0	0	0
			数量	0	0	0	0	0	0



复杂报表开发技术对比—多源分片—实现思路（国外传统类）

多源转化成单源

- 编写存储过程或其它代码来准备数据
- 多库时，无法直接写出复杂SQL或存储过程，需要架构专门的数据库桥后才可以完成

子表拼接

- 每个分片做成一个子表，再在主表中拼接起来
- 各个子表各自独立，很难将格线统一对齐
- 各子表之间无法沟通数据，引用计算困难

- 由于模型原因，行列严重不对称，列方向的扩展能力远远弱于行方向
- 理论上是通过子表拼接完成，但实际操作上非常困难，有待验证

复杂报表开发技术对比—多源分片—实现思路（新型网格类）

- 直接支持多源（集）设计，在同一张报表中可同时引用多个数据集
- 单元格根据数据表达式进行扩展，扩展部分与固定部分可自由结合

报表模板				
	A	B	C	D
1	客户			汇总纬度
2	按客户	= ds1.group(customer_id, false)	金额	= ds2.group
3	已签合同		数量	= ds2.select
4			金额	= ds4.sum(amo
5			数量	= ds4.sum(qua
6			金额	= ds4.sum(amo
7			数量	= ds4.sum(qua
8	按合同	= ds4.group(year(sign	金额	= ds4.sum(amo
9			数量	= ds4.sum(qua
10			金额	= ds4.sum(amo
11			数量	= ds4.sum(qua
12			金额	= ds4.sum(amo
			数量	= ds4.sum(qua

客户			汇总纬度	化工				食品	
				苯乙酮	苯甲酸钠	苯磺酸	对氯二苯	大豆	色拉油
按客户汇总	国内客户	无锡食品加工厂	金额	400,000	500,000	800,000	500,000	900,000	700,000
			数量	1,540	1,530	4,630	2,631	3,125	2,902
		天津长城化工厂	金额	816,000	1,020,000	1,632,000	1,020,000	1,836,000	1,428,000
			数量	3,080	3,060	9,260	5,262	6,250	5,804
		中国化工原料厂	金额	94,400	118,000	188,800	118,000	212,400	165,200
			数量	1,540	1,530	4,630	2,631	3,125	2,902
	国外客户	Crystal	金额	3,142,400	3,928,000	6,284,800	3,928,000	7,000,000	499,200
			数量	4,620	4,590	13,890	7,893	9,375	8,706
		Chemist	金额	481,600	602,000	963,200	602,000	1,020,000	816,000
			数量	3,080	3,060	9,260	5,262	6,250	5,804
		Murli	金额	164,400	205,500	328,800	205,500	312,500	247,700
			数量	3,080	3,060	9,260	5,262	6,250	5,804
已签合同			金额	5,314,600	6,643,250	10,629,200	6,643,250	11,957,850	9,300,550
			数量	20,020	19,890	60,190	34,203	40,625	37,726
已发货			金额	816,000	1,020,000	1,632,000	1,020,000	1,836,000	1,428,000
			数量	1,540	1,530	4,630	2,631	3,125	2,902
按合同年度汇总	2001	金额	174,000	217,500	348,000	217,500	391,500	304,500	
		数量	3,080	3,060	9,260	5,262	6,250	5,804	
	2002	金额	441,800	552,250	883,600	552,250	994,050	773,150	
		数量	3,080	3,060	9,260	5,262	6,250	5,804	
	2003	金额	4,012,000	5,015,000	8,024,000	5,015,000	9,027,000	7,021,000	
		数量	9,240	9,180	27,780	15,786	18,750	17,412	
	2004	金额	688,800	858,500	1,373,600	858,500	1,545,300	1,201,900	
		数量	4,620	4,590	13,890	7,893	9,375	8,706	
已付款			金额	0	0	0	0	0	0
			数量	0	0	0	0	0	0

复杂报表开发技术对比——跨行组排名

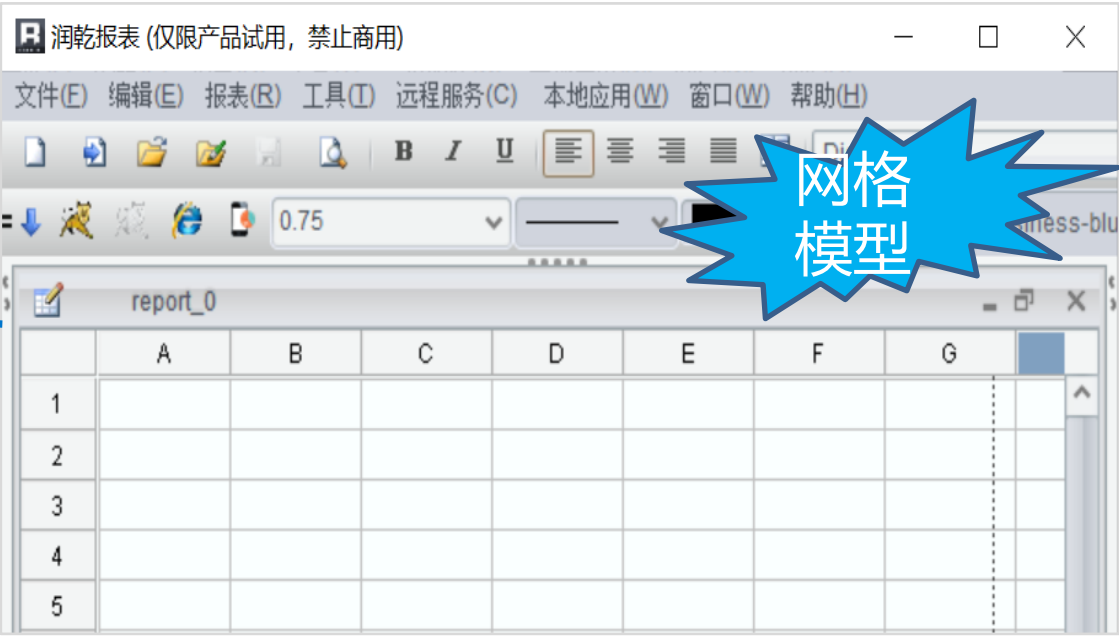
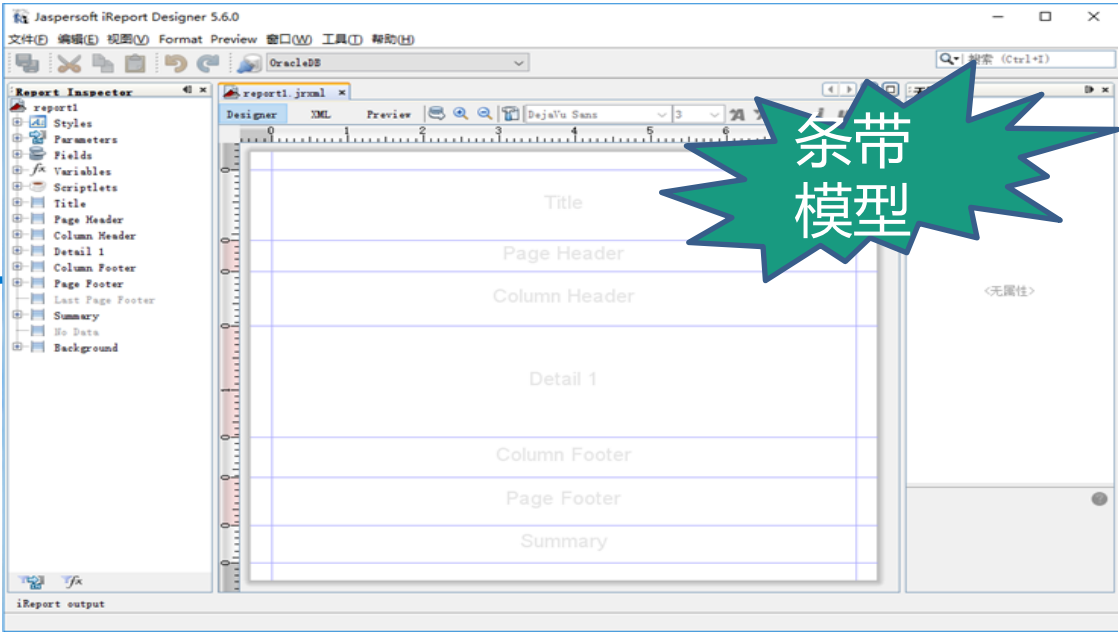


跨行组排名的报表怎么样做？

城八区电信销售统计表								
地区	合计		固定电话		小灵通		移动电话	
	销售额	排名	销售额	排名	销售额	排名	销售额	排名
朝阳	972,102,181.00	1	9,879,866.00	2	897,868,567.00	1	64,353,748.00	4
崇文	194,828,511.00	2	8,567,857.00	4	98,686,090.00	2	87,574,564.00	2
东城	12,516,819.00	8	3,452,345.00	8	8,967,578.00	5	96,896.00	8
丰台	72,399,706.00	5	7,646,346.00	5	78,575.00	8	64,674,785.00	3
海淀	101,782,940.00	4	6,345,768.00	6	89,768,585.00	3	5,668,587.00	7
石景山	161,115,195.00	3	9,879,868.00	1	63,547,869.00	4	87,687,458.00	1
西城	13,003,779.00	7	5,668,746.00	7	867,575.00	7	6,467,458.00	6
宣武	22,896,007.00	6	9,876,757.00	3	6,485,896.00	6	6,535,354.00	5

前三名城区销售额总和：1,328,045,887.00
城八区占北京市的销售比例：79.49%

占总额比例：85.64%



复杂报表开发技术对比—跨行组排名—实现思路（国外传统类）

- 由于模型原因，只有简单计算列功能（比如同行内统计值等），没有很好的运算后数据项命名机制（一般用列名命名设计阶段的数据单元），无法精准引用

```
SQL脚本.txt
--生成省内排序、全国排序、全国同级别城市排序数据加工的 存储过程脚本:
CREATE PROCEDURE Order_P
AS
    DECLARE @ls_kc as varchar(20)
    drop table temptable
    Select no=Identity(int,1,1),* Into temptable From gy_amount order by amount DESC
    update a set a.COrder=b.no from gy_amount a,temptable b where a.province=b.province
    and a.city=b.city and a.level=b.level
--生成省内排序
DECLARE C1 Cursor For
SELECT DISTINCT province FROM gy_amount where province is not null
OPEN C1
FETCH NEXT FROM C1 INTO @ls_kc
WHILE (@@Fetch_Status = 0)
BEGIN
    drop table temptable
    Select no=Identity(int,1,1),* Into temptable From gy_amount where province=@ls_kc order by province ,amount DESC
    update a set a.POrder=b.no from gy_amount a,temptable b where a.province=b.province
    and a.city=b.city and a.level=b.level
    FETCH NEXT FROM C1 INTO @ls_kc
END
CLOSE C1
--生成全国同级别排序
DECLARE C2 Cursor For
SELECT DISTINCT level FROM gy_amount where level is not null
OPEN C2
FETCH NEXT FROM C2 INTO @ls_kc
WHILE (@@Fetch_Status = 0)
BEGIN
    drop table temptable
    Select no=Identity(int,1,1),* Into temptable From gy_amount where level=@ls_kc order by level ,amount DESC
    update a set a.LOrder=b.no from gy_amount a,temptable b where a.province=b.province
    and a.city=b.city and a.level=b.level
    FETCH NEXT FROM C2 INTO @ls_kc
END
CLOSE C2
DEALLOCATE C2
drop table temptable
```

存储过程

事先把数据计算好，再往报表格子里填

复杂报表开发技术对比—跨行组排名—实现思路（新型网格类）

- 对扩展后的每个单元格进行精确命名，即可在设计时自由引用任何一个扩展后单元格进行关联运算（像Excel表达式一样）
- 单元格还可以方便地嵌入SQL或其它函数单独引用外部信息

报表模板					
	A	B	C	D	E
1	城八区电信销售统计表				
2	地区	合计		= ds1.group(类别; false)	
3		销售额	排名	销售额	排名
4	= ds1.group(=sum(b4{}	=count(B4[~0;~0]{B4>\$b4))+1	= ds1.sum(销售
5					=count(B4[~0
6(尾)	前三名城区销售额总和:		=sum(b4{c4<=3}))		占总额比例: =c6/sum(b4{}
7(尾)	城八区占北京市的销售比		=sum(b4{}/ ds2.select_one(总销售额)		

城八区电信销售统计表								
地区	合计		固定电话		小灵通		移动电话	
	销售额	排名	销售额	排名	销售额	排名	销售额	排名
朝阳	972,102,181.00	1	9,879,866.00	2	897,868,567.00	1	64,674,185.00	4
崇文	194,828,511.00	2	8,567,857.00	4	98,686,090.00	2	6,467,458.00	2
东城	12,516,819.00	8	3,452,345.00	8	8,967,578.00	8	6,467,458.00	2
丰台	72,399,706.00	5	7,646,346.00	5	78,575.00	8	64,674,185.00	3
海淀	101,782,940.00	4	6,345,768.00	6	89,768,585.00	3	5,668,587.00	7
石景山	161,115,195.00	3	9,879,868.00	1	63,547,869.00	4	87,687,458.00	1
西城	13,003,779.00	7	5,668,746.00	7	867,575.00	7	6,467,458.00	6
宣武	22,898,007.00	6	9,876,757.00	3	6,485,896.00	6	6,535,354.00	5

前三名城区销售额总和：1,328,045,887.00

城八区占北京市的销售比例：79.49%

占总额比例：85.64%

思考与小结

思考1：报表工具最重要的指标是什么？

数据统计模型，即在不编代码的情况下能够制作出何种报表，它决定了报表工具最基本的制表能力

思考2：市场上存量软件对复杂报表处理的如何？

事实上，市场上众多国外或开源产品以及BI类产品，在处理复杂报表时都存在明显缺陷，大量报表仍需编写代码准备数据

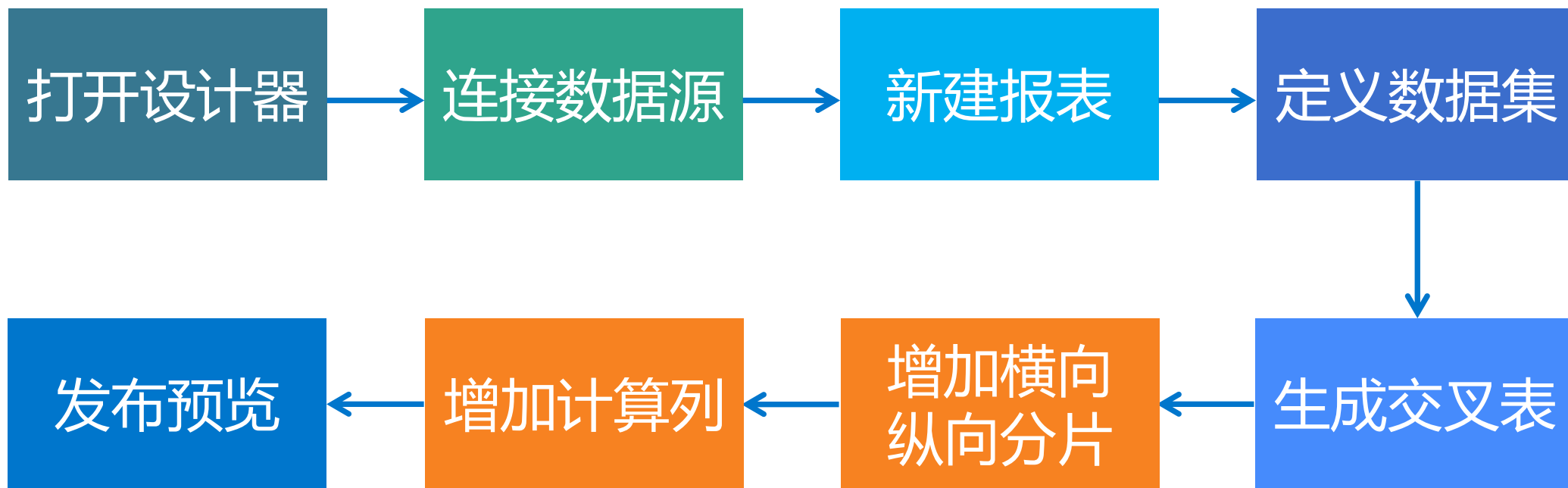


多源分片报表设计

多源分片报表设计



- 使用新型网格式报表工具，制作多源分片报表，实操感受一下
- 制作报表的基本流程如下：



多源分片报表——举例

- 纵向分片
- 横向分片
- 扩展区域与静态区域混合

报表模板

	A	B	C	D	E	F	G
1	统计维度	产品信息		=ds1.grou	销售额大于10000的总	产品类别	总计
2		销售额		=ds1.grou		=ds4.grou	
3	=ds1.group(year(订购日期))		=ds1.g	=ds1.sum	=sum(D3{D	=ds1.sum	=sum(F3{F
4	销售额大于10000的总计			=sum(D3{		=sum(F3{F	
5	销售员业绩	=ds3.group	=ds3.g	=ds1.sum	=sum(D5{D	=ds1.sum	=sum(F5{F
6	销售员业绩总计			=sum(D5{F		=sum(F5{F	

参照例子：9.rpx

扩展区域与静态区域/混合

↓

报表预览

产品信息 统计维度			华南			西南		销售额大于10000的总计	产品类别								总计
			海口	深圳	厦门	成都	重庆		饮料	调味品	点心	日用品	谷类/麦片	肉/家禽	特制品	海鲜	
2014年	1月		¥4731.26	¥12024.10	¥931.50	¥0.00	¥22924.14	¥34948.24	¥13786.00	¥1130.50	¥4761.30	¥5545.00	¥2859.50	¥1242.20	¥1068.00	¥10218.50	¥40611.00
	2月		¥0.00	¥9881.60	¥1559.10	¥0.00	¥15535.71	¥15535.71	¥3815.50	¥3144.65	¥2656.85	¥6037.00	¥2701.75	¥4073.36	¥360.00	¥4187.30	¥26976.41
	3月		¥3024.90	¥23011.05	¥2052.50	¥0.00	¥22474.60	¥45485.65	¥6342.00	¥5651.60	¥15993.95	¥9472.80	¥1923.50	¥564.00	¥4412.00	¥6203.20	¥50563.05
	4月		¥5181.80	¥23585.00	¥210.00	¥326.00	¥28198.05	¥51783.05	¥8015.00	¥4132.50	¥3660.30	¥17698.10	¥2007.00	¥10787.40	¥7351.00	¥3849.55	¥57500.85
	5月		¥86.85	¥5218.00	¥0.00	¥0.00	¥5779.30	¥0.00	¥152.00	¥500.00	¥1137.00	¥0.00	¥4322.50	¥3919.00	¥465.00	¥588.65	¥11084.15
2015年	7月		¥0.00	¥7216.00	¥0.00	¥0.00	¥0.00	¥0.00	¥0.00	¥610.00	¥230.00	¥4192.00	¥2184.00	¥0.00	¥0.00	¥0.00	¥7216.00
销售额大于10000的总计			¥0.00	¥58620.15	¥0.00	¥0.00	¥89132.50		¥13786.00	¥0.00	¥15993.95	¥17698.10	¥0.00	¥10787.40	¥0.00	¥10218.50	
销售员业绩	副总裁(销售)	王伟	¥1393.00	¥11847.00	¥0.00	¥0.00	¥11952.05	¥23799.05	¥2118.00	¥2415.00	¥5430.00	¥7506.00	¥292.50	¥2634.00	¥750.00	¥4046.55	¥25192.05
	内部销售协调员	刘英玫	¥2544.00	¥6646.20	¥0.00	¥0.00	¥2285.50	¥0.00	¥1235.00	¥1536.50	¥2738.70	¥3266.80	¥428.00	¥0.00	¥780.00	¥1490.70	¥11475.70
	销售代表	张颖静	¥4277.21	¥9425.70	¥0.00	¥0.00	¥12830.85	¥12830.85	¥4725.00	¥3026.00	¥4896.26	¥3201.60	¥1069.00	¥4575.00	¥2622.00	¥2418.90	¥26533.76
		李芳	¥2584.50	¥16653.25	¥2052.50	¥326.00	¥17407.10	¥34060.35	¥14652.50	¥1049.00	¥2876.35	¥5874.00	¥2254.00	¥96.00	¥2120.00	¥10101.50	¥39023.35
		郑建杰	¥264.00	¥8754.50	¥1450.60	¥0.00	¥11163.80	¥11163.80	¥3422.00	¥752.60	¥2238.90	¥3375.00	¥6354.50	¥1432.40	¥465.00	¥3592.50	¥21632.90
		孙林	¥0.00	¥2582.60	¥108.50	¥0.00	¥5133.65	¥0.00	¥1180.50	¥623.10	¥344.05	¥599.00	¥1349.00	¥295.20	¥2667.00	¥766.90	¥7824.75
		金士鹏	¥60.00	¥12309.50	¥931.50	¥0.00	¥22532.90	¥34842.40	¥3395.50	¥2135.00	¥3214.00	¥10921.50	¥3752.50	¥7718.40	¥4252.00	¥445.00	¥35833.90
		张雪眉	¥0.00	¥12717.00	¥0.00	¥0.00	¥9129.15	¥12717.00	¥1382.00	¥3190.00	¥5226.54	¥6862.50	¥0.00	¥3594.96	¥0.00	¥1590.15	¥21846.15
	销售经理	赵军	¥1902.10	¥0.00	¥210.00	¥0.00	¥2476.80	¥0.00	¥0.00	¥442.05	¥1474.60	¥1338.50	¥498.75	¥240.00	¥0.00	¥595.00	¥4588.90
销售员业绩总计			¥13024.81	¥80935.75	¥4753.10	¥326.00	¥94911.80		¥32110.50	¥15169.25	¥28439.40	¥42944.90	¥15998.25	¥20585.96	¥13656.00	¥25047.20	

多源分片报表—制作过程 (1)

- 第一步：打开设计器
- 第二步：连接DEMO数据源
- 第三步：通过“报表” -> “数据集” 菜单，首先定义数据集：

ds1: SELECT 客户.地区,客户.城市,订单明细.数量,订单明细.折扣,订单明细.单价,订单.雇员ID,订单.订购日期,订单明细.产品ID FROM 订单明细,订单,客户 WHERE 客户.客户ID = 订单.客户ID AND 订单.订单ID = 订单明细.订单ID and 订单.订购日期 is not null and 客户.地区 in ('华南','西南') and year(订单.订购日期)>=2014

- 第四步：写入单元格表达式

- (1)在A3单元格录入表达式: =ds1.group(year(订购日期);year(订购日期):1)+"年"
- (2)在C3单元格录入表达式: =ds1.group(month(订购日期);month(订购日期):1)+"月"
- (3)在D1单元格录入表达式: =ds1.group(地区;地区:1), 扩展方向: 横向
- (4)在D2单元格录入表达式: =ds1.group(城市;城市:1), 扩展方向: 横向
- (5)在D3单元格录入表达式: =ds1.sum(数量*单价)

统计维度	产品信息 销售额	华南			西南	
		海口	深圳	厦门	成都	重庆
2014年	1月	¥ 4731.26	¥ 12024.10	¥ 931.50	¥ 0.00	¥ 22924.14
	2月	¥ 0.00	¥ 9881.60	¥ 1559.10	¥ 0.00	¥ 15535.71
	3月	¥ 3024.90	¥ 23011.05	¥ 2052.50	¥ 0.00	¥ 22474.60
	4月	¥ 5181.80	¥ 23585.00	¥ 210.00	¥ 326.00	¥ 28198.05
	5月	¥ 86.85	¥ 5218.00	¥ 0.00	¥ 0.00	¥ 5779.30
2015年	7月	¥ 0.00	¥ 7216.00	¥ 0.00	¥ 0.00	¥ 0.00

	A	B	C	D
1	统计维度	产品信息		=ds1.group(
2		销售额		=ds1.group(
3		=ds1.group(year(订购日期);year(订购日期):1)+"年"	=ds1.group(month(订购日期);month(订购日期):1)+"月"	=ds1.sum(数量*单价)

多源分片报表—制作过程（2）

- 在上一步的基础上，增加数据集ds3：

SELECT 雇员.雇员ID,订单明细.产品ID,雇员.职务,雇员.姓氏||雇员.名字 as 姓名 FROM 雇员,订单,订单明细 WHERE 雇员.

雇员ID = 订单.雇员ID AND 订单.订单ID = 订单明细.订单ID ORDER BY 订单.订购日期 ASC

- 在报表下方追加一行，设定各单元格表达式

(1) 在B4单元格输入表达式：= ds3.group(职务;职务:1)

(2) 在C4单元格输入表达式：= ds3.group(雇员ID;雇员ID:1),

设置显示值为：ds3.姓名

(3) 在D4单元格输入表达式：=ds1.sum(数量*单价,雇员ID==C4),

设置显示格式为：¥#0.00

	A	B	C	D
1			产品信息	=ds1.group(
2	统计维度		销售 额	=ds1.group(
3	=ds1.group(year(订购日期)	=ds1.group(=ds1.sum(
4	销售员业绩	=ds3.group(=ds3.group(=ds1.sum(

统计维度		产品信息	华南			西南	
		销售 额	海口	深圳	厦门	成都	重庆
2014年		1月	¥ 4731.26	¥ 12024.10	¥ 931.50	¥ 0.00	¥ 22924.14
		2月	¥ 0.00	¥ 9881.60	¥ 1559.10	¥ 0.00	¥ 15535.71
		3月	¥ 3024.90	¥ 23011.05	¥ 2052.50	¥ 0.00	¥ 22474.60
		4月	¥ 5181.80	¥ 23585.00	¥ 210.00	¥ 326.00	¥ 28198.05
		5月	¥ 86.85	¥ 5218.00	¥ 0.00	¥ 0.00	¥ 5779.30
2015年		7月	¥ 0.00	¥ 7216.00	¥ 0.00	¥ 0.00	¥ 0.00
销售员业绩	副总裁(销售)	王伟	¥ 1393.00	¥ 11847.00	¥ 0.00	¥ 0.00	¥ 11952.05
	内部销售协	刘英玫	¥ 2544.00	¥ 6646.20	¥ 0.00	¥ 0.00	¥ 2285.50
	销售代表	张颖静	¥ 4277.21	¥ 9425.70	¥ 0.00	¥ 0.00	¥ 12830.85
		李芳	¥ 2584.50	¥ 16653.25	¥ 2052.50	¥ 326.00	¥ 17407.10
		郑建杰	¥ 264.00	¥ 8754.50	¥ 1450.60	¥ 0.00	¥ 11163.80
		孙林	¥ 0.00	¥ 2582.60	¥ 108.50	¥ 0.00	¥ 5133.65
		金士鹏	¥ 60.00	¥ 12309.50	¥ 931.50	¥ 0.00	¥ 22532.90
		张雪眉	¥ 0.00	¥ 12717.00	¥ 0.00	¥ 0.00	¥ 9129.15
	销售经理	赵军	¥ 1902.10	¥ 0.00	¥ 210.00	¥ 0.00	¥ 2476.80

多源分片报表—制作过程 (3)

统计维度	产品信息 销售额	华南			西南		产品类别							
		海口	深圳	厦门	成都	重庆	饮料	调味品	点心	日用品	谷类/麦片	肉/家禽	特制品	海鲜
2014年	1月	¥ 4731.26	¥ 12024.10	¥ 931.50	¥ 0.00	¥ 22924.14	¥ 13786.00	¥ 1130.50	¥ 4761.30	¥ 5545.00	¥ 2859.50	¥ 1242.20	¥ 1068.00	¥ 10218.50
	2月	¥ 0.00	¥ 9881.60	¥ 1559.10	¥ 0.00	¥ 15535.71	¥ 3815.50	¥ 3144.65	¥ 2656.85	¥ 6037.00	¥ 2701.75	¥ 4073.36	¥ 360.00	¥ 4187.30
	3月	¥ 3024.90	¥ 23011.05	¥ 2052.50	¥ 0.00	¥ 22474.60	¥ 6342.00	¥ 5651.60	¥ 15993.95	¥ 9472.80	¥ 1923.50	¥ 564.00	¥ 4412.00	¥ 6203.20
	4月	¥ 5181.80	¥ 23585.00	¥ 210.00	¥ 326.00	¥ 28198.05	¥ 8015.00	¥ 4132.50	¥ 3660.30	¥ 17698.10	¥ 2007.00	¥ 10787.40	¥ 7351.00	¥ 3849.55
	5月	¥ 86.85	¥ 5218.00	¥ 0.00	¥ 0.00	¥ 5779.30	¥ 152.00	¥ 500.00	¥ 1137.00	¥ 0.00	¥ 4322.50	¥ 3919.00	¥ 465.00	¥ 588.65
2015年	7月	¥ 0.00	¥ 7216.00	¥ 0.00	¥ 0.00	¥ 0.00	¥ 0.00	¥ 610.00	¥ 230.00	¥ 4192.00	¥ 2184.00	¥ 0.00	¥ 0.00	¥ 0.00
销售员业绩	副总裁(销售)	王伟	¥ 1393.00	¥ 11847.00	¥ 0.00	¥ 0.00	¥ 11952.05	¥ 2118.00	¥ 2415.00	¥ 5430.00	¥ 7506.00	¥ 292.50	¥ 2634.00	¥ 750.00
	内部销售协	刘英玖	¥ 2544.00	¥ 6646.20	¥ 0.00	¥ 0.00	¥ 2285.50	¥ 1235.00	¥ 1536.50	¥ 2738.70	¥ 3266.80	¥ 428.00	¥ 0.00	¥ 780.00
	销售代表	张颖静	¥ 4277.21	¥ 9425.70	¥ 0.00	¥ 0.00	¥ 12830.85	¥ 4725.00	¥ 3026.00	¥ 4896.26	¥ 3201.60	¥ 1069.00	¥ 4575.00	¥ 2622.00
		李芳	¥ 2584.50	¥ 16653.25	¥ 2052.50	¥ 326.00	¥ 17407.10	¥ 14652.50	¥ 1049.00	¥ 2876.35	¥ 5874.00	¥ 2254.00	¥ 96.00	¥ 2120.00
		郑建杰	¥ 264.00	¥ 8754.50	¥ 1450.60	¥ 0.00	¥ 11163.80	¥ 3422.00	¥ 752.60	¥ 2238.90	¥ 3375.00	¥ 6354.50	¥ 1432.40	¥ 465.00
		孙林	¥ 0.00	¥ 2582.60	¥ 108.50	¥ 0.00	¥ 5133.65	¥ 1180.50	¥ 623.10	¥ 344.05	¥ 599.00	¥ 1349.00	¥ 295.20	¥ 2667.00
		金士鹏	¥ 60.00	¥ 12309.50	¥ 931.50	¥ 0.00	¥ 22532.90	¥ 3395.50	¥ 2135.00	¥ 3214.00	¥ 10921.50	¥ 3752.50	¥ 7718.40	¥ 4252.00
		张雪眉	¥ 0.00	¥ 12717.00	¥ 0.00	¥ 0.00	¥ 9129.15	¥ 1382.00	¥ 3190.00	¥ 5226.54	¥ 6862.50	¥ 0.00	¥ 3594.96	¥ 0.00
	销售经理	赵军	¥ 1902.10	¥ 0.00	¥ 210.00	¥ 0.00	¥ 2476.80	¥ 0.00	¥ 442.05	¥ 1474.60	¥ 1338.50	¥ 498.75	¥ 240.00	¥ 595.00

在在上一步的基础上，增加如下数据集：

ds2: SELECT 类别.类别ID,类别.类别名称 FROM 类别

ds4: SELECT 产品.类别ID,产品.产品ID FROM 产品

在报表右侧追加一列，设定各单元格表达式

(1) 在E2单元格输入表达式：=ds4.group(类别ID,类别ID!=null;类别ID:1)

设置显示值为：ds2.select (类别名称,类别ID==value(),1)

扩展方向：横向扩展

(2) 在E3单元格输入表达式：=ds1.sum(数量*单价,产品ID in ds4.select(产品ID))

设置显示格式为：¥ #0.00

(3) 在E4单元格输入表达式：=ds1.sum(数量*单价,产品ID in ds4.select(产品ID) and 雇员ID==C4)

设置显示格式为：¥ #0.00

	A	B	C	D	E
1	统计维度	产品信息			产品类别
2		销售额			=ds1.group(
3		=ds1.group(year(订购日期),=ds1.group(=ds1.sum(
4		销售员业绩	=ds3.group(=ds3.group(=ds1.sum(

多源分片报表—制作过程（4）

- 在上一步基础上，再分别增加两行两列，在新加的行列中增加一些表达式，对不同扩展区域

的数据分别进行统计：

- (1) 在E3格输入：=sum(D3{D3>10000})
- (2) 在E5格输入：=sum(D5{D5>10000})
- (3) 在D4格输入：=sum(D3{D3>10000})
- (4) 在D6格输入：=sum(D5{})
- (5) 在G3格输入：=sum(F3{})
- (6) 在G5格输入：=sum(F5{})
- (7) 在F4格输入：=sum(F3{F3>10000})
- (8) 在F6格输入：=sum(F5{})
- (9) 以上单元格设置显示格式为：¥#0.00

统计维度\产品信息			华南			西南		销售额大于10000的总计	产品类别								总计
			销售 额	海口	深圳	厦门	成都		重庆	饮料	调味品	点心	日用品	谷类/麦片	肉/家禽	特制品	
2014年		1月	¥ 4731.26	¥ 12024.10	¥ 931.50	¥ 0.00	¥ 22924.14	¥ 34948.24	¥ 13786.00	¥ 1130.50	¥ 4761.30	¥ 5545.00	¥ 2859.50	¥ 1242.20	¥ 1068.00	¥ 10218.50	¥ 40611.00
		2月	¥ 0.00	¥ 9881.60	¥ 1559.10	¥ 0.00	¥ 15535.71	¥ 15535.71	¥ 3815.50	¥ 3144.65	¥ 2656.85	¥ 6037.00	¥ 2701.75	¥ 4073.36	¥ 360.00	¥ 4187.30	¥ 26976.41
		3月	¥ 3024.90	¥ 23011.05	¥ 2052.50	¥ 0.00	¥ 22474.60	¥ 45485.65	¥ 6342.00	¥ 5651.60	¥ 15993.95	¥ 9472.80	¥ 1923.50	¥ 564.00	¥ 4412.00	¥ 6203.20	¥ 50563.05
		4月	¥ 5181.80	¥ 23585.00	¥ 210.00	¥ 326.00	¥ 28198.05	¥ 51783.05	¥ 8015.00	¥ 4132.50	¥ 3660.30	¥ 17698.10	¥ 2007.00	¥ 10787.40	¥ 7351.00	¥ 3849.55	¥ 57500.85
		5月	¥ 86.85	¥ 5218.00	¥ 0.00	¥ 0.00	¥ 5779.30	¥ 0.00	¥ 152.00	¥ 500.00	¥ 1137.00	¥ 0.00	¥ 4322.50	¥ 3919.00	¥ 465.00	¥ 588.65	¥ 11084.15
2015年		7月	¥ 0.00	¥ 7216.00	¥ 0.00	¥ 0.00	¥ 0.00	¥ 0.00	¥ 0.00	¥ 610.00	¥ 230.00	¥ 4192.00	¥ 2184.00	¥ 0.00	¥ 0.00	¥ 0.00	¥ 7216.00
销售额大于10000的总计			¥ 0.00	¥ 58620.15	¥ 0.00	¥ 0.00	¥ 89132.50		¥ 13786.00	¥ 0.00	¥ 15993.95	¥ 17698.10	¥ 0.00	¥ 10787.40	¥ 0.00	¥ 10218.50	
销售员业绩	副总裁(销售)	王伟	¥ 1393.00	¥ 11847.00	¥ 0.00	¥ 0.00	¥ 11952.05	¥ 23799.05	¥ 2118.00	¥ 2415.00	¥ 5430.00	¥ 7506.00	¥ 292.50	¥ 2634.00	¥ 750.00	¥ 4046.55	¥ 25192.05
	内部销售协调员	刘英玖	¥ 2544.00	¥ 6646.20	¥ 0.00	¥ 0.00	¥ 2285.50	¥ 0.00	¥ 1235.00	¥ 1536.50	¥ 2738.70	¥ 3266.80	¥ 428.00	¥ 0.00	¥ 780.00	¥ 1490.70	¥ 11475.70
	销售代表	张颖静	¥ 4277.21	¥ 9425.70	¥ 0.00	¥ 0.00	¥ 12830.85	¥ 12830.85	¥ 4725.00	¥ 3026.00	¥ 4896.26	¥ 3201.60	¥ 1069.00	¥ 4575.00	¥ 2622.00	¥ 2418.90	¥ 26533.76
		李芳	¥ 2584.50	¥ 16653.25	¥ 2052.50	¥ 326.00	¥ 17407.10	¥ 34060.35	¥ 14652.50	¥ 1049.00	¥ 2876.35	¥ 5874.00	¥ 2254.00	¥ 96.00	¥ 2120.00	¥ 10101.50	¥ 39023.35
		郑建杰	¥ 264.00	¥ 8754.50	¥ 1450.60	¥ 0.00	¥ 11163.80	¥ 11163.80	¥ 3422.00	¥ 752.60	¥ 2238.90	¥ 3375.00	¥ 6354.50	¥ 1432.40	¥ 465.00	¥ 3592.50	¥ 21632.90
		孙林	¥ 0.00	¥ 2582.60	¥ 108.50	¥ 0.00	¥ 5133.65	¥ 0.00	¥ 1180.50	¥ 623.10	¥ 344.05	¥ 599.00	¥ 1349.00	¥ 295.20	¥ 2667.00	¥ 766.90	¥ 7824.75
		金士鹏	¥ 60.00	¥ 12309.50	¥ 931.50	¥ 0.00	¥ 22532.90	¥ 34842.40	¥ 3395.50	¥ 2135.00	¥ 3214.00	¥ 10921.50	¥ 3752.50	¥ 7718.40	¥ 4252.00	¥ 445.00	¥ 35833.90
		张雪眉	¥ 0.00	¥ 12717.00	¥ 0.00	¥ 0.00	¥ 9129.15	¥ 12717.00	¥ 1382.00	¥ 3190.00	¥ 5226.54	¥ 6862.50	¥ 0.00	¥ 3594.96	¥ 0.00	¥ 1590.15	¥ 21846.15
	销售经理	赵军	¥ 1902.10	¥ 0.00	¥ 210.00	¥ 0.00	¥ 2476.80	¥ 0.00	¥ 0.00	¥ 442.05	¥ 1474.60	¥ 1338.50	¥ 498.75	¥ 240.00	¥ 0.00	¥ 595.00	¥ 4588.90
销售员业绩总计			¥ 13024.81	¥ 80935.75	¥ 4753.10	¥ 326.00	¥ 94911.80		¥ 32110.50	¥ 15169.25	¥ 28439.40	¥ 42944.90	¥ 15998.25	¥ 20585.96	¥ 13656.00	¥ 25047.20	

	A	B	C	D	E	F	G
1	产品信息		=ds1.group		销售额大于10000的总计	产品类别	总计
2	统计维度		=ds1.group				
3	=ds1.group(year(订购日期))		=ds1.group	=ds1.sum	=sum(D3{D3>10000})	=ds1.sum	=sum(F3{})
4	销售额大于10000的总计			=sum(D3{D3>10000})		=sum(F3{F3>10000})	
5	销售员业绩		=ds3.group(销售员)	=ds1.sum	=sum(D5{D5>10000})	=ds1.sum	=sum(F5{})
6	销售员业绩总计			=sum(D5{})		=sum(F5{})	

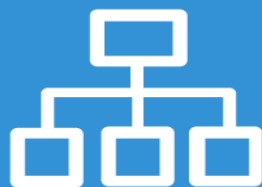


数据源开发技术

常见数据源开发技术



存储过程



中间表辅助



JAVA自定义数据源

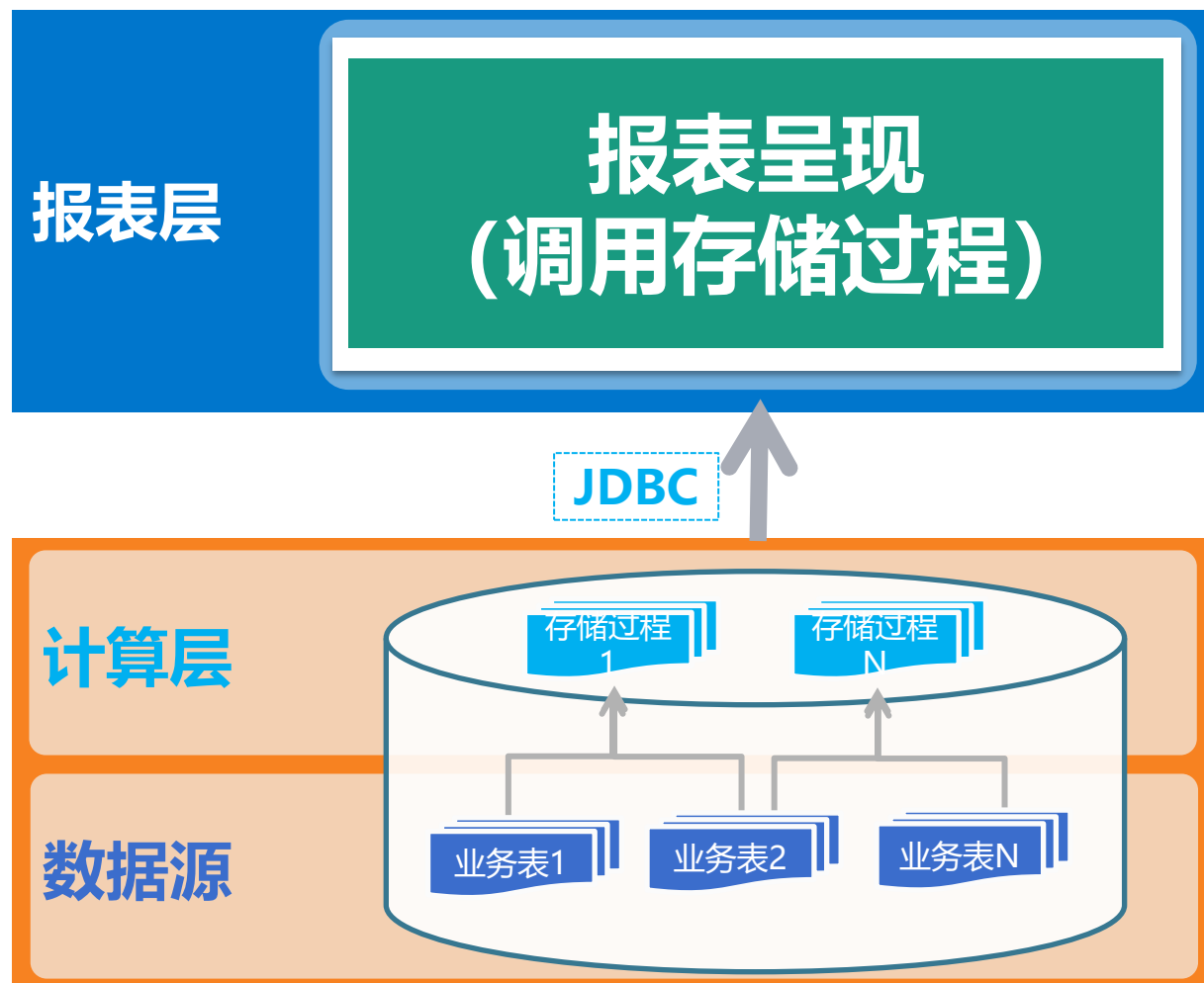
数据源开发技术—存储过程



采用存储过程的原因

- 支持多步运算
 - ✓ 在存储过程中可以通过多步运算，实现复杂数据处理逻辑
- 多利于界面与逻辑分离
 - ✓ 计算逻辑放在数据库中完成，利于计算存储层和前端应用层分离
- 计算性能高
 - ✓ 存储过程采用预编译机制运算性能更高

数据源开发技术—存储过程



存储过程存在的问题:

管理与安全

存储过程和报表的存放位置不同，导致对应难度很大；存储过程修改需要分配相应的数据库权限，存在安全隐患

开发调试困难

难以扩展和移植

分库后无法跨库查询；更换数据库无法移植

应用高耦合

容易被其他应用调用，造成多应用间的耦合

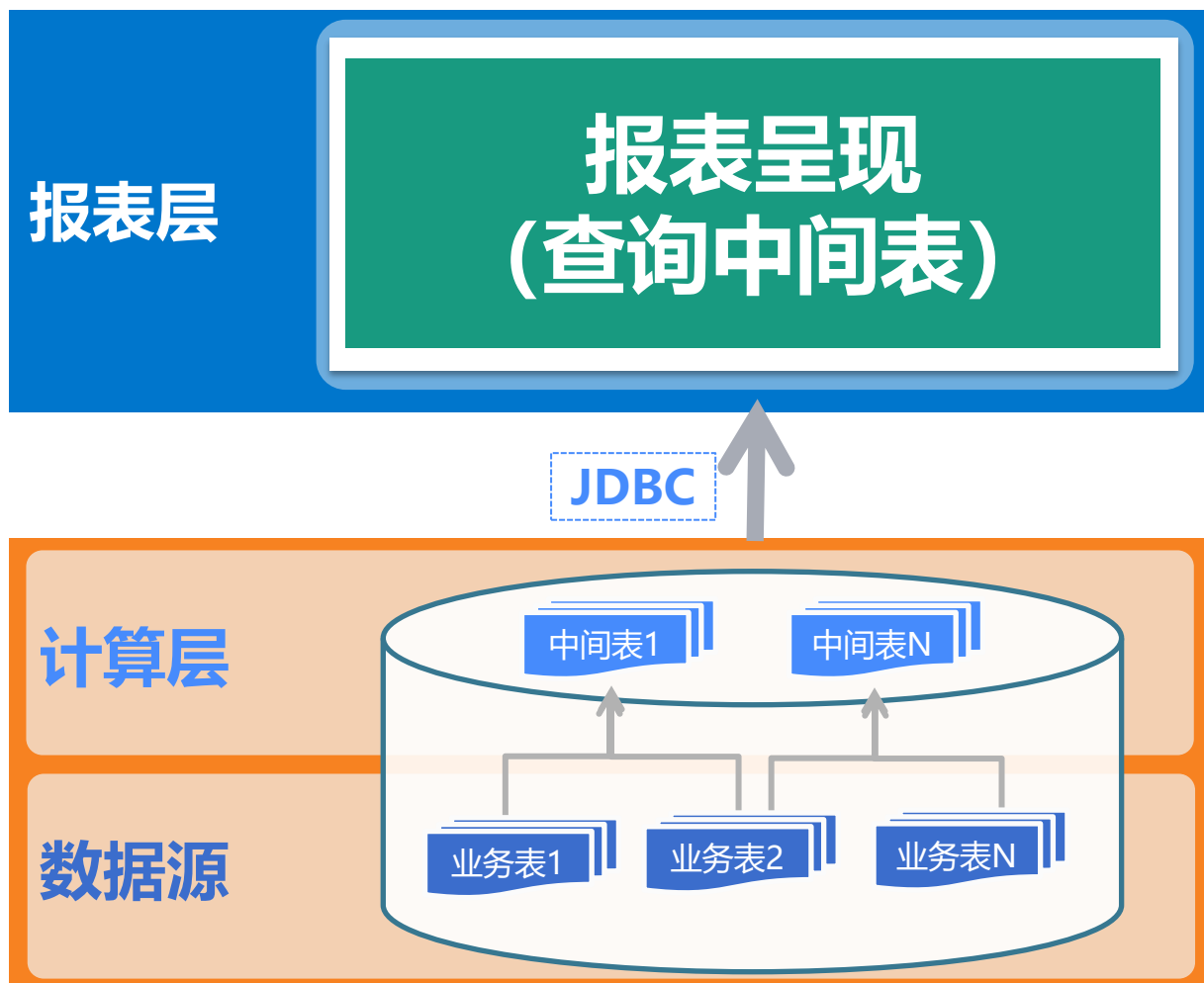
数据源开发技术—中间表辅助



中间表产生的原因

- 预处理跑批
 - ✓ 数据量大或逻辑复杂，避免报表等待时间过长，预先计算好中间数据，存放中间表里
- 一步算不出来
 - ✓ 原始数据需要经过多步骤的复杂计算才能得出有意义的数​​据，供报表使用
- 多样性数据源参与计算
 - ✓ 来自NoSQL/文本/XML/JSON等外部数据，需与库内数据进行混合计算，必须先入库

数据源开发技术—中间表辅助



中间表存在的问题:

管理混乱

各个应用的累积大量中间表存储在线性结构的数据库中造成数据库管理混乱，不易清理

资源浪费

不用的中间表，仍然需要相应ETL过程向其更新数据，浪费数据库资源，性能不断下降

扩容压力

存储过多的中间数据，加速数据库扩容，用户不得不经常升级

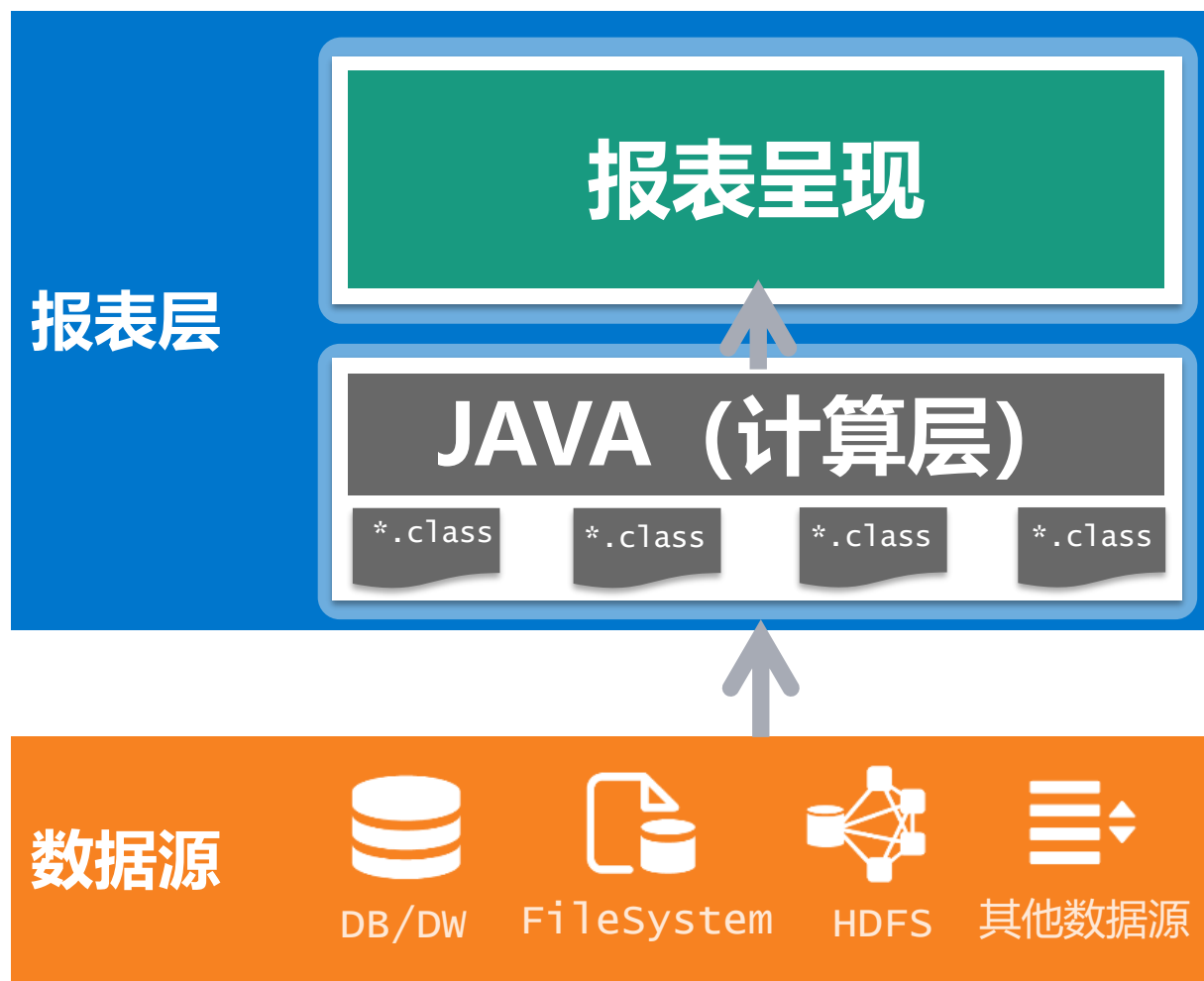
数据源开发技术—JAVA



采用JAVA的原因

- 复杂逻辑控制
 - ✓ 为报表准备数据时涉及复杂计算或逻辑控制，直接SQL不易实现，需Java辅助
- 多样性数据源
 - ✓ 涉及NoSQL/文本/XML/JSON等，报表工具无法实现混合计算，需用Java实现
- 方便应用扩展
 - ✓ 业务增长需对应用进行水平扩展，使用SQL不利于未来分库

数据源开发技术—JAVA



JAVA计算存在的问题:

模块化困难

Java程序必须和主应用**一起编译打包**，耦合度高

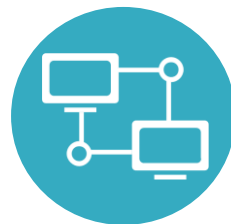
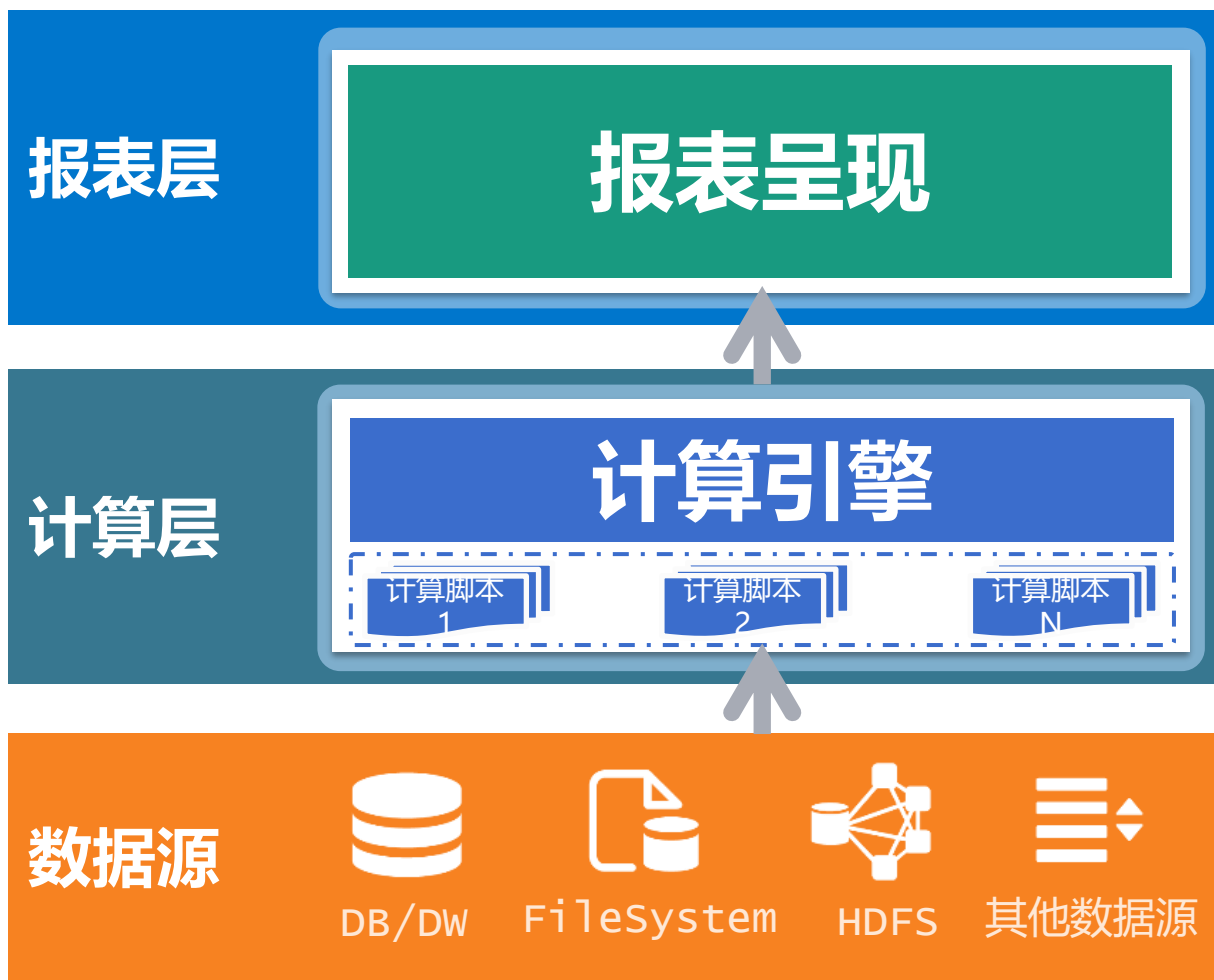
难以热切换

Java编写的报表数据准备算法有修改后会导致整个应用**重新编译部署**，很难做到热切换

类库少

Java程序在结构化计算方面的类库较少，写个简单分组，动辄数百行的冗长代码

数据源开发技术— 数据计算层



独立的计算层可以极大的降低报表与应用、应用与数据库的耦合



当需求变更时修改此计算模块不影响系统其他部分，完全模块化



计算脚本是解释执行语言，很容易做到热切换，无须重启应用

数据源开发技术— 数据计算层— 跨库关联举例

报表层

城市	订单数	订单金额
Beijing	33	¥ 600083.00
Changzhou	34	¥ 655112.00
Chengdu	3	¥ 25458.00
Chongqing	88	¥ 2165427.00
Dalian	36	¥ 468148.00
Haikou	22	¥ 275773.00
Nanchang	14	¥ 138293.00

计算层

数据集设置	
名称	类型
ds1	脚本数据集
脚本数据集	
	A
1	=Mysql.query("SELECT O_ORDERKEY,O_CUSTKEY,O_ORDERDATE,O_TOTALPRICE FROM ORDERS")
2	=ORA.query("SELECT C_CUSTKEY,C_NAME,C_CITY FROM CUSTOMER")
3	>A1.switch(O_CUSTKEY,A2:C_CUSTKEY)
4	=A1.groups(O_CUSTKEY.C_CITY:CITY;count(O_ORDERKEY):COUNT,sum(O_TOTALPRICE):AMOUNT)

数据源

Mysql: ORDERS				
Index	O_ORDERKEY	O_CUSTKEY	O_ORDERDATE	O_TOTALPRICE
1	10262	RASTC	1996-07-22	14487.0
2	10263	ERNSH	1996-07-23	43818.0
3	10264	FOLKO	2007-12-18	1101.0
4	10265	BLONP	1996-07-25	5528.0
5	10266	WARTH	1996-07-26	7719.0
6	10267	FRANK	1996-07-29	20858.0
7	10268	GROSR	1996-07-30	19887.0
8	10269	WHITC	1996-07-31	456.0

ORACLE: CUSTOMER			
Index	C_CUSTKEY	C_NAME	C_CITY
1	BLONP	The hao	Dalian
2	CACTU	Wayair freight co. LTD	Dalian
3	CENTC	Three jie industrial	Dalian
4	HUNGC	Hardware mechanical	Dalian
5	MEREP	huake	Dalian
6	ALFKI	Sanchuan industrial ...	Tianjin



性能与容量

性能与容量— 什么是性能问题

报表展现慢

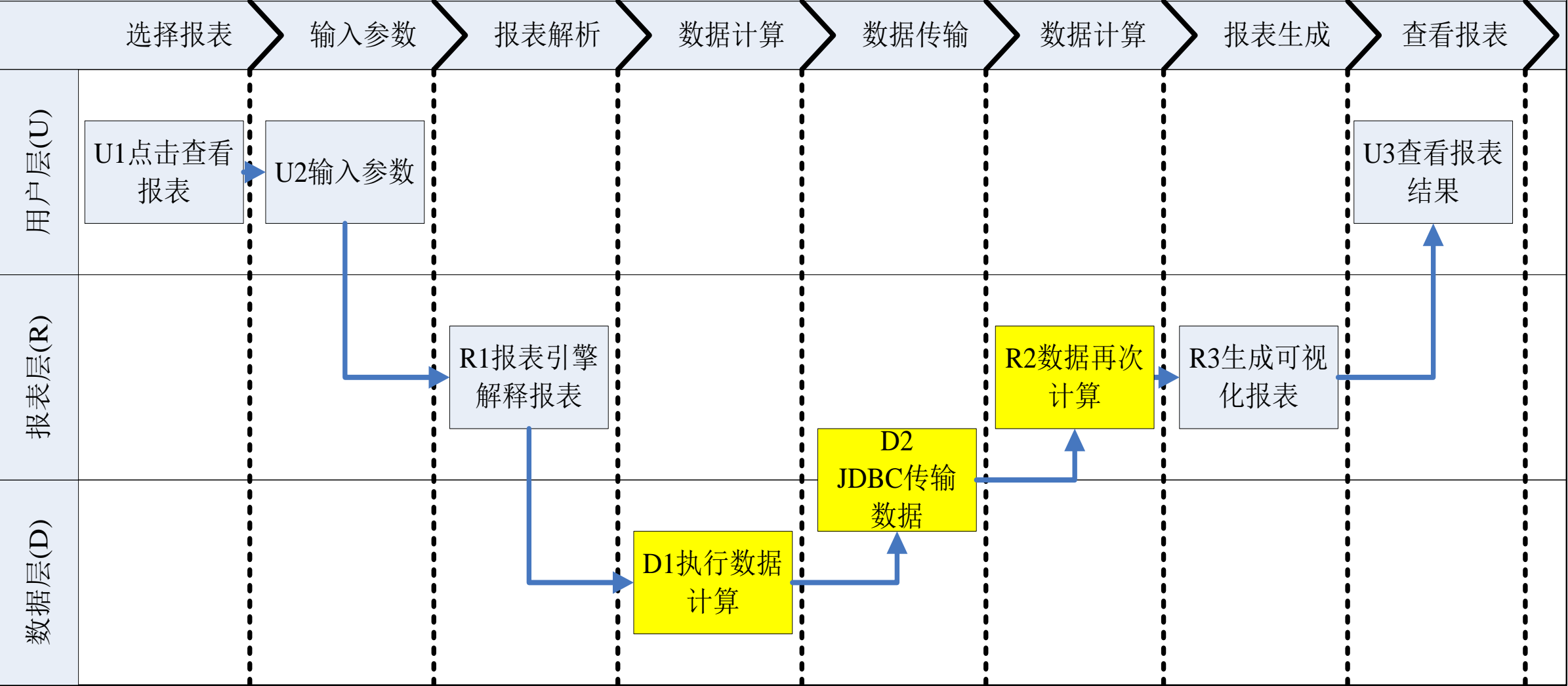


这是表象

内存溢出

性能与容量— 影响性能的主要因素

报表处理过程示意图



性能与容量—数据准备慢（原因分析）

数据量大

报表呈现的数据量一般很小，但涉及的原始数据量可能巨大，把大数据汇总和过滤成小数据需要很长时间

复杂计算

复杂的业务逻辑计算也是类似，主要时间消耗在数据准备阶段

- 这个环节主要是数据库来计算，其性能与报表工具没有什么关系！

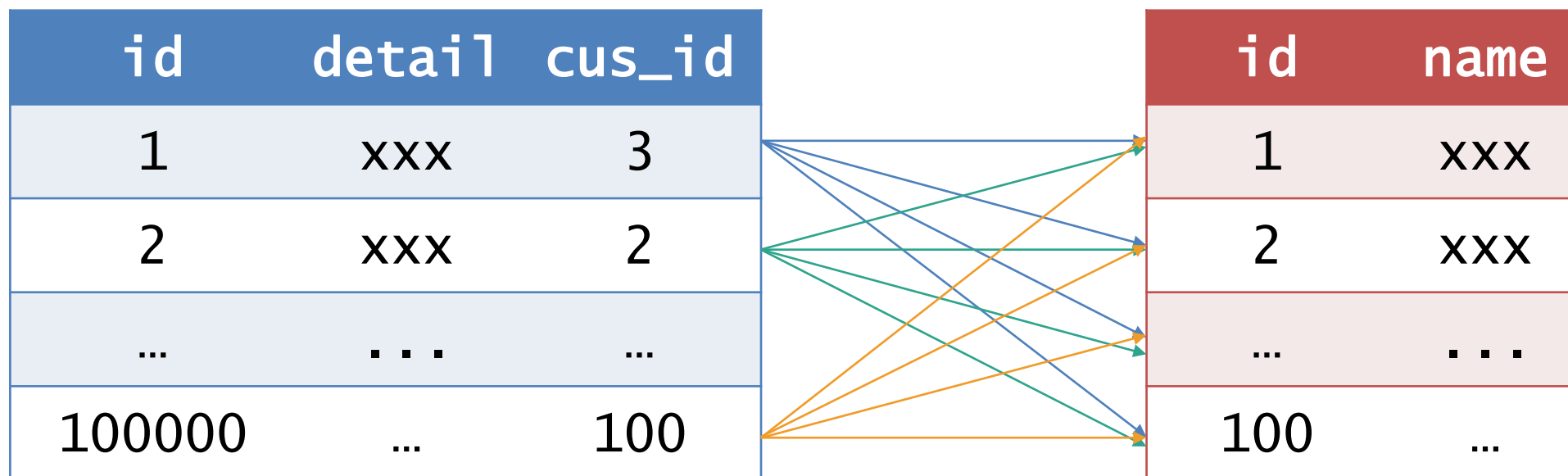
性能与容量—数据传输慢（原因分析）

主流数据库提供的JDBC驱动在将数据流转换成Java对象时的效率很低，但又不可避免，在大数据量时取数就会非常慢

- 不同数据库JDBC-IO性能有所不同： MySQL<Oracle<MSSQL<DB2
- JDBC驱动是由数据库厂家提供的，所以，这方面的性能和报表工具也没有太大关系！

性能与容量—报表计算慢（原因分析）

- 报表计算性能差的情况相对少，比如典型的场景是多源关联报表，在报表示单元格中用数据集过滤来描述本格和其它数据集的关联，当数据量不大时也无所谓，但数据量稍大（几千行）且涉及数据集较多时，性能会急剧下降



单元格表达式类似：`ds2.select(name,id==ds1.cus_id)`

性能与容量一小结



通过上面的分析可以知道，看起来是报表环节表现出来的问题，但绝大多数是数据源环节的问题

重点关注

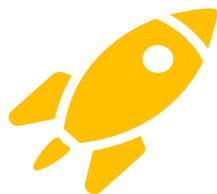


大部分性能问题出现在数据源层或需要由数据源层解决



少部分出现在报表层的问题也可以通过数据源层解决

性能与容量—数据准备慢—优化思路



- 数据准备的优化是报表提速的关键！

优化数据准备代码

- 一般是SQL（或存储过程），某些时候是应用程序的代码（涉及非数据库或多数据库时）

数据库扩容

- 数据量大，代码不能再优化时，还可以扩容数据库，比如采用集群方案

采用高性能计算引擎

- 数据库在实现某些运算时性能较差或成本太高，可以更换为其它计算机制

性能与容量—数据传输慢—优化思路



- 无论应用开发商还是报表工具厂商都没办法改变数据库的JDBC驱动，只能在数据准备阶段想办法

多线程并行取数

- 采用多线程并行取数，同时建立多个连接从数据库分段取数（DB连接数足够），但目前报表工具并没有直接提供并行取数的功能

可控缓存

- 报表自带的缓存功能可以在一定程度有效地改善响应体验，但机制过于死板
- 通外部程序实现灵活的缓存机制，比如部分缓存、多个报表之间缓存复用、以及不同缓存的不同生存周期，采用可控的缓存策略，有效减少重复运算

性能与容量—报表计算慢—优化思路



- 比如多源关联报表场景，如果把运算移到报表外，在数据准备阶段时处理，就可以大幅度提升性能

同库HASH JOIN算法

- 如果数据来自同一个数据库，那么用SQL写JOIN语句即可

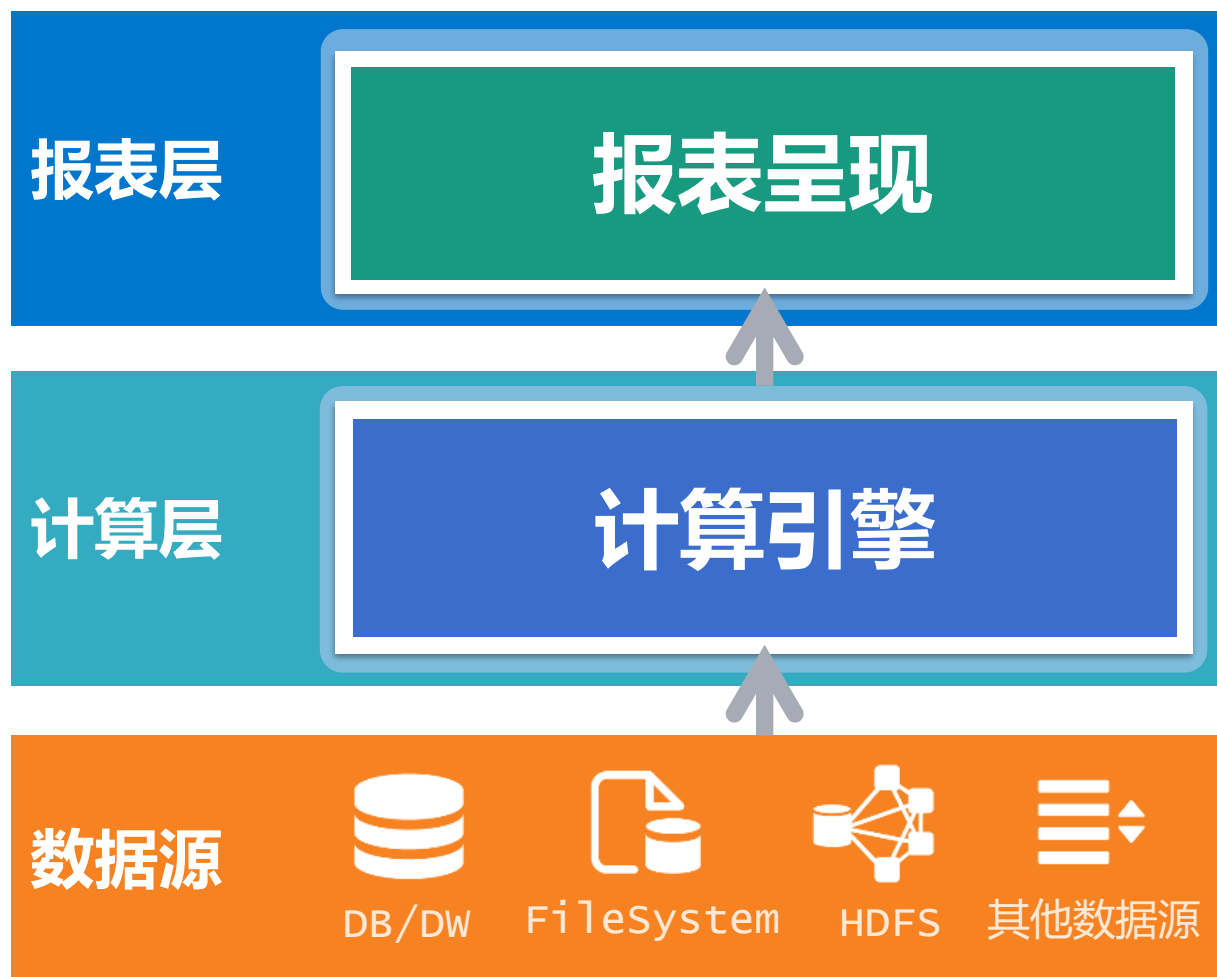
外部HASH JOIN算法

- 如果数据集来自多库或者希望减轻数据库的计算压力，也可以通过外部程序实现HASH JOIN算法
- HASH JOIN算法可以整体地看待几个数据集，效率比报表工具采用的过滤式关联要高得多

性能与容量—数据计算层



- 前面讨论的优化思路都需在数据源阶段才能解决，那么引入计算中间层希望更便捷地实现数据准备也是比较自然的想法



解决有些不能或不合适使用数据库的计算（涉及非数据库或多数据库或数据库负担重，及存储过程导致的性能或管理问题等）

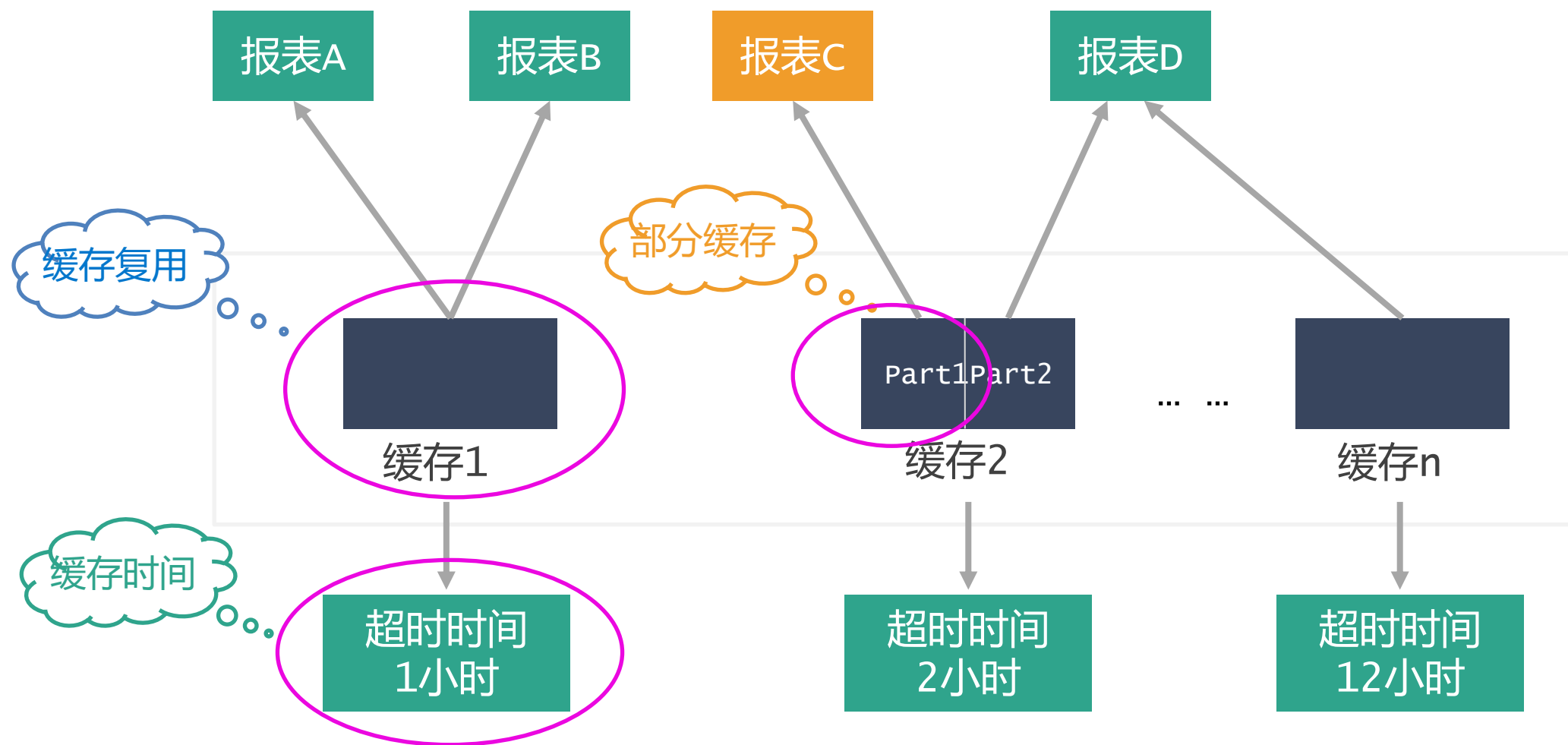


解决不适合在报表中的计算或报表工具不直接具备的能力（多线程并行取数、灵活可控缓存机制、高并发内存共享机制等）

性能与容量—数据计算层—多线程取数举例

	A	B	C
1	fork 4	=connect(db)	/分4线程，要分别建立连接
2		=B1.query@x(“select * from T where part=?”,A1)	/分别取每一段
3	=A1.conj()		/合并结果

性能与容量—数据计算层—可控缓存举例



练习—多源关联—需求场景

- 三个数据集，分别是订单数据、用户数据和销售人员数据。数据结构与关联关系如下：



【报表需求】 按客户所在地区和销售人员统计订单情况

练习—多源关联—表样

地区销售订单情况汇总

地区	销售编号	销售姓名	订单数量	订单金额	运货费
东北	2	王伟	2002	12413501.10	1162981.82
	1	张颖静	10010	20354734.38	1238257.02
	2	王伟	6006	10012002.00	181581.40
	3	李芳	16016	19886166.28	1141720.58
	4	郑建杰	16016	26105679.60	806105.30
	5	赵军	2002	2762760.00	11491.48
	6	孙林	8008	9916726.82	375234.86
	7	金士鹏	6006	20824723.85	840179.34
	8	刘英玫	12012	8348139.80	443803.36
	9	张雪眉	4004	3550546.98	179179.00
华北			4004	456856.40	1207206.00
	1	张颖静	100100	117743659.93	5764178.42
	2	王伟	92092	152269016.76	7206779.58
	3	李芳	88088	110062622.59	7661513.86
	4	郑建杰	128128	198836452.63	9942112.18
	5	赵军	46046	81573141.56	5644939.30
	6	孙林	56056	51163691.54	1742460.72
	7	金士鹏	58058	61082146.04	3311468.16
	8	刘英玫	90090	122059147.11	7030843.82
	9	张雪眉	16016	19702367.64	527707.18

来源
客户数据

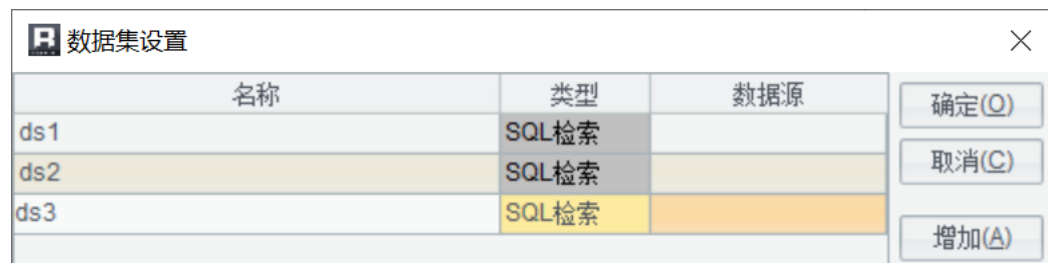
来源
订单数据

来源
销售数据

由于数据可能分别来自不同库；

练习—多源关联—在报表现单元格中完成数据集关联

- 新建报表，使用3个SQL数据集分别准备订单数据、客户数据和销售人员数据



- 如前所述，3个数据集可能来源多个库，或因为实际业务中每个数据集SQL过于复杂无法在数据库中完成关联

练习—多源关联—在报表现单元格中完成数据集关联

- 根据关联关系，定义报表表达式：

	A	B	C	D	E	F
1	地区销售订单情况汇总					
2	地区	销售编号	销售姓名	订单数量	订单金额	运货费
3	=ds2.group(地区)	=ds1.group(雇员)	=ds3.select(姓)	=ds1.count()	=ds1.sum(订单金额)	=ds1.sum(运货
4	总计：			=sum(D3{})	=sum(E3{})	=sum(F3{})

A3: 使用ds2按照客户地区分组，表达式： ds2.group(地区;地区:1)

B3: 使用ds1， 根据ds2中的客户ID查找该客户地区下的销售人员， 表达式：
ds1.group(雇员ID,客户ID in ds2.select(客户ID);雇员ID:1)

C3: 使用ds3， 根据ds1中的销售ID查找对应的姓名， 表达式：
ds3.select(姓氏+名字,雇员ID==B3,1)

练习—多源关联—在报表单元格中完成数据集关联

- 浏览报表，记录时间：

	时间（秒）
准备数据集（取数）	6
报表计算（关联计算）	32
总时间	38

运行时间

练习—多源关联—在计算层完成数据集关联

- 新建报表，使用脚本数据集准备数据

数据集设置			×
名称	类型	数据源	确定(Q)
ds1	脚本数据集		取消(C)

脚本数据集		
	A	B
1	=connect("demo")	
2	=A1.query("select 订单ID,雇员ID,客户ID,订单金额,运货费 from 订单")	
3	=A1.query("select * from 客户").keys(客户ID)	
4	=A1.query@x("select * from 雇员").keys(雇员ID)	
5	=A2.join(客户ID,A3,地区;雇员ID,A4,姓氏+名字:姓名)	
6	=A5.groups(地区,雇员ID,姓名;count(订单ID):订单数量,sum(订单金额):订单金额,sum(运货费):运货费)	

- A2-A4读取三部分数据，其中SQL可以任意复杂
- 在A5中使用：A2.join(客户ID,A3,地区;雇员ID,A4,姓氏+名字:姓名)
- 订单数据通过两个外键分别与客户、雇员关联
- A6按照地区和人员分组求得汇总值，结果集返回报表

直接使用准备好的结果集：

	A	B	C	D	E	F
1	地区销售订单情况汇总					
2	地区	销售编号	销售姓名	订单数量	订单金额	运货费
3	=ds1.group(地区)	=ds1.group(雇员)	=ds1.姓名	=ds1.订单数量	=ds1.订单金额	=ds1.运货费
4	总计：			=sum(D3{})	=sum(E3{})	=sum(F3{})

报表中只使用ds1进行分组展现

练习—多源关联—在计算层完成数据集关联

- 浏览报表，记录时间：

	时间（秒）
准备数据集 (取数+关联计算)	12
报表计算	0
总时间	12

运行时间

练习—多源关联—结果分析

(全部时间)

38 -> 12



32 -> 6

(计算时间)



5.3倍提升



计算性能获得

性能与容量—什么是大报表

- 报表要展现的数据可达百万、甚至千万级别，通常以清单列表方式呈现，少数情况以分组报表形式呈现

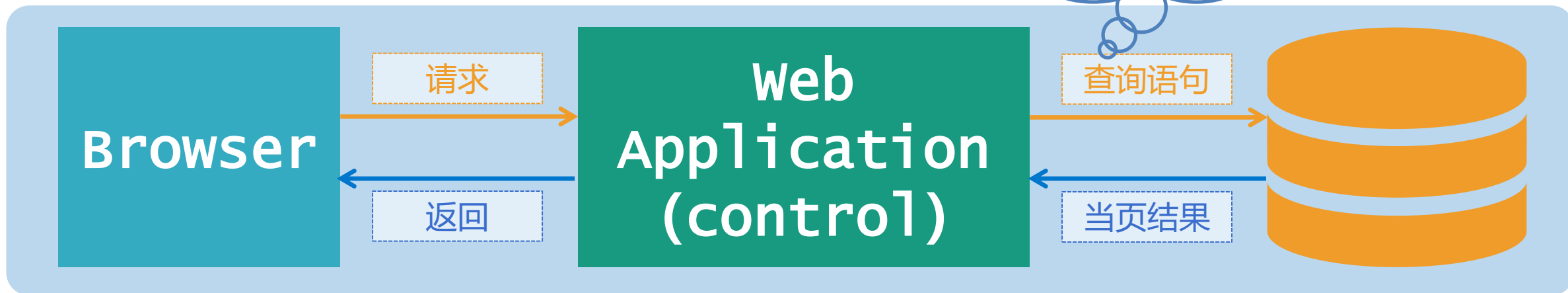
订单ID	客户ID	订购日期	运货费	货主名称	货主地区
10248	VINET	2012-07-04	¥ 32.38	余小姐	东北
10249	TOMSP	2012-07-05	¥ 11.61	谢小姐	
10250	HANAR	2012-07-08	¥ 65.83	谢小姐	
10251	VICTE	2012-07-08	¥ 41.34	陈先生	
10252	SUPRD	2012-07-09	¥ 51.30	刘先生	
10253	HANAR	2012-07-10	¥ 58.17	谢小姐	
10254	CHOPS	2012-07-11	¥ 22.98	林小姐	
10255	RICSU	2012-07-12	¥ 148.33	方先生	
10256	WELLI	2012-07-15	¥ 13.97	何先生	
10257	HILAA	2012-07-16	¥ 81.91	王先生	
10258	ERNSH	2012-07-17	¥ 140.51	王先生	华东
10259	CENTC	2012-07-18	¥ 3.25	林小姐	
10260	OTTIK			徐文彬	
10261	QUEDE			刘先生	
10262	RATTC	2012-07-22	¥ 48.29	干先生	
货主地区	订单ID	客户ID	订购日期	运货费	订单金额
东北	10783	HANAR	2013-12-18	¥ 124.98	¥ 1442.50
	10932	BONAP	2014-03-06	¥ 134.64	¥ 1788.63
	10723	WHITC	2013-10-30	¥ 21.72	¥ 468.45
	10641	HILAA	2013-08-22	¥ 179.61	¥ 2054.00
	10700	SAVEA	2013-10-10	¥ 65.10	¥ 1638.40
	10701	SAVEA	2013-10-13	¥ 220.31	¥ 2864.50
	10730	BONAP	2013-11-05	¥ 20.12	¥ 484.26
	10760	MAISD	2013-12-01	¥ 155.64	¥ 2917.00
	10670	FRANK	2013-09-16	¥ 203.48	¥ 2301.75
	11068	QUEEN	2014-05-04	¥ 81.75	¥ 2027.08
				订单金额小计:	¥ 102410.19
华东	10737	VINET	2013-11-11	¥ 7.79	¥ 139.80
	10894	SAVEA	2014-02-18	¥ 116.13	¥ 2753.10
	10502	PERIC	2013-04-10	¥ 69.32	¥ 816.30
	10681			¥ 1.13	¥ 1287.40
	10503			¥ 0.74	¥ 2048.50
	10689	BERGS	2013-10-01	¥ 15.42	¥ 472.50
	10885	SUPRD	2014-02-12	¥ 5.64	¥ 1209.00

性能与容量——大报表——为什么要采用数据库分页

- 如果等着把全部记录取出后再呈现，那需要很长时间，用户体验恶劣
- 报表一般采用内存运算机制，所以一次装不下太多数据，使用分页呈现也是很自然的想法，这样每次只显示一页数据，就不会造成内存溢出

数据库分页的大致做法：

```
1 SELECT * FROM  
2 (  
3   SELECT A.*, ROWNUM RN  
4   FROM (SELECT * FROM TABLE_NAME) A  
5   WHERE ROWNUM <= 40  
6 )  
7 WHERE RN >= 21
```



性能与容量—大报表—数据库分页带来的问题

翻页效率差

页码小时，感觉不明显；
页码较大时，翻页会有明显的等待感

可能出现汇总错误

每页的SQL都是单独发送的，如果期间数据库发生增删操作，可能导致数据汇总错误

无法实现分组效果

由于每页取数无法保证取出整组数据，因而也无法实现分组报表

无法使用非RDB数据源

数据库分页基于RDB，其他数据源（如NoSQL或文件则无法使用）

性能与容量—大报表—更高级的解决办法

游标

向数据库发出取数SQL生成游标，从中取出一页后呈现，但并不终止这个游标，要取下一页的时候再继续取数

只能向后翻页

混用

向后翻页时用游标，一旦发生向前翻页时，则重新执行取数SQL

解决问题不彻底

性能与容量—大报表—另一个很大的问题

数据源耦合性太强

每种数据库的实现方式不同，更换数据源都要重新实现

强耦合

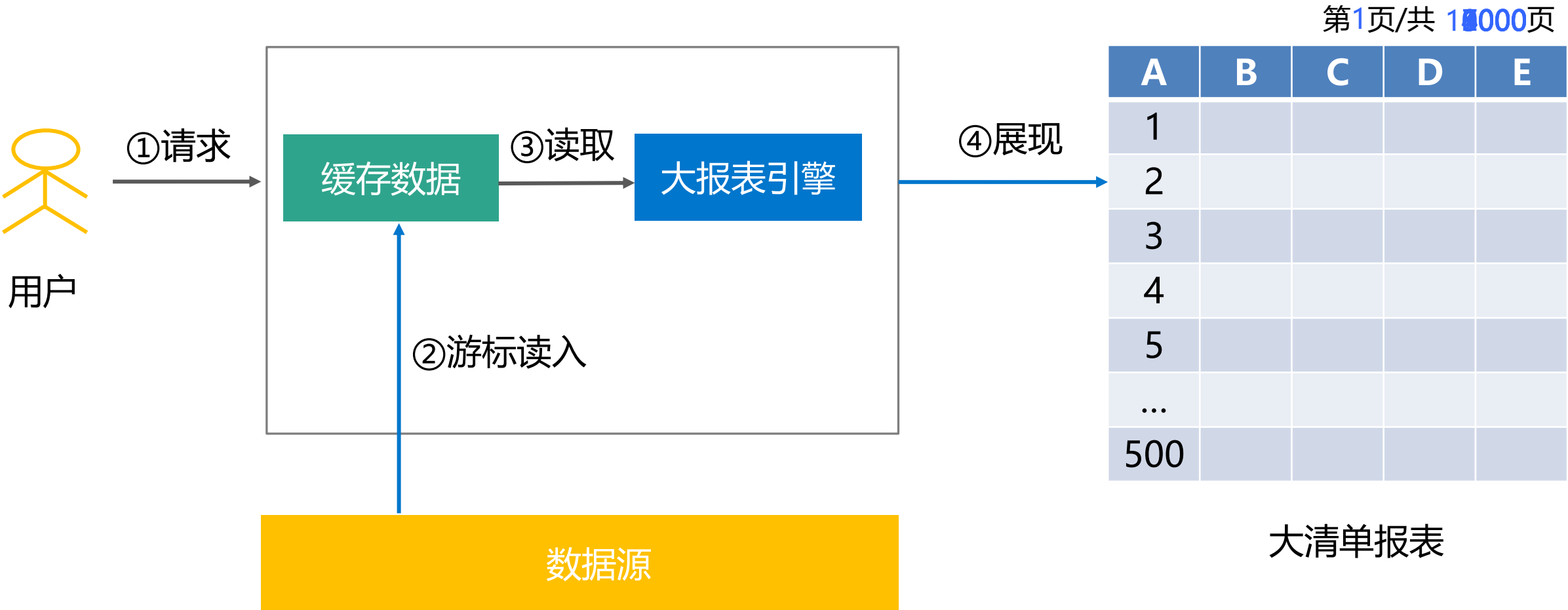
性能与容量—大报表—如何从根本上解决

取数和呈现做成两个异步线程

- 取数线程发出SQL后就不断取出数据并缓存到本地存储中，呈现线程根据页数计算出行数到本地缓存中去获取数据显示
- 只要已经取过的数据就能快速呈现，不会有等待感，还没取到的数据需要等待一下也是正常可理解的
- 取数线程只涉及一句SQL，在数据库中是同一个事务，也不会有不一致的问题

应用层面

性能与容量—大报表—异步双线程工作原理



小结

- ✓ 通过本章的学习，我们知道复杂报表一般都有哪些基本特征：多源分片、不规则分组、动态格间计算、行列对称；以及在应对复杂报表时，传统条带式模型所表现出来的不足，而新型网格类报表模型就可以轻松应对
- ✓ 了解常见数据源开发技术的优缺点和新的解决思路
- ✓ 知道影响报表性能的主要因素有哪些，它们会发生在哪些环节以及优化思路
- ✓ 了解传统大报表实现思路（数据库分页）和带来的问题，以及如何从根本上解决

THANKS

感谢观看



乾学院 · VIP课堂

商业智能 DashBoard 与大屏技术



CONTENTS

本章内容

1. 数据可视化
2. 初识BI仪表盘
3. 用报表组布局功能快速实现示例
4. WEB技术开发方式制作DBD
5. BI工具制作DBD
6. 手机端DBD
7. 结论





数据可视化

报表、统计图属于可视化

序号	所属板块	项目名称	商品房销售				非住宅(门面、车库)				住宅											
			套数	成交面积	成交金额	均价	套数	面积	金额	均价	别墅				多层				小高层、高层			
											套数	面积	金额	均价	套数	面积	金额	均价	套数	面积	金额	均价
1	行政	兆基君城	544	64292.08	227999108	3546													544	64292.1	227999108	3546
2	行政	中源·凝香华都	450	52810.59	240533950	4555	16	3630	41759277	11503												
3	行政	绿地·中央花园	329	47932.03	219803369	4586	28	10636	72127940	6782												
4	行政	长沙富豪山庄盛	260	42074.9	180798178	4297	3	332.6	5823800	17500												
5	行政	龙湾一品	248	35112.81	145612818	4147	11	6238	39115077	6270												
6	行政	华都丽景	105	18756.83	117763600	6278	39	11832	95914933	8107												
7	行政	玉虹·蓝庭国际城	135	18415.88	80169413	4353	5	3937	28203746	7165												
8	行政	钱隆公馆	138	17042.34	59797982	3509																
9	行政	铂金汉宫	125	14243.62	56192501	3945	3	304.3	4938540	16232												
10	行政	瑞景春天	79	11897.14	66454719	5586	28	5323	45000329	8454												
11	行政	福临门·名郡	92	11849.45	41793417	3527																
12	行政	一品江山	220	11780.37	86652181	7356	216	11135	84929582	7627												
13	行政	翰林华府	109	11755.6	39321377	3345																
14	行政	中央领御	39	6820.73	31432625	4608	5	2316	12173011	5256												
15	行政	和泰家园	18	3709.65	13329738	3593	8	2537	11223818	4424												
16	行政	嘉诚花园	74	3602.85	17119304	4752	48	1661	9945409	5987												
17	行政	世纪花园	18	1850.44	5946918	3214																
18	行政	春城佳苑	24	1719.27	6730226	3915																
19	行政	水晶城	17	975.08	7908997	8111	6	376.7	5685602	15092												
20	行政	玉龙国际花园	6	809.61	2651038	3274																
21	行政	同兴花园	4	376.58	1672016	4440	4	376.6	1672016	4440												
22	行政	兴旺佳园	3	309.93	384120	1239																
23	行政	维也纳花城	4	306.79	2185187	7123	3	151	1858007	12305												
24	行政	水晶郦城	1	137	411000	3000																
25	行政	时代和府	1	59.76	370000	6191	1	59.76	370000	6191												



比起用户界面中其它形式，报表呈现的数据更多，统计图呈现数据更形象，它们是可视化主要的呈现形式。

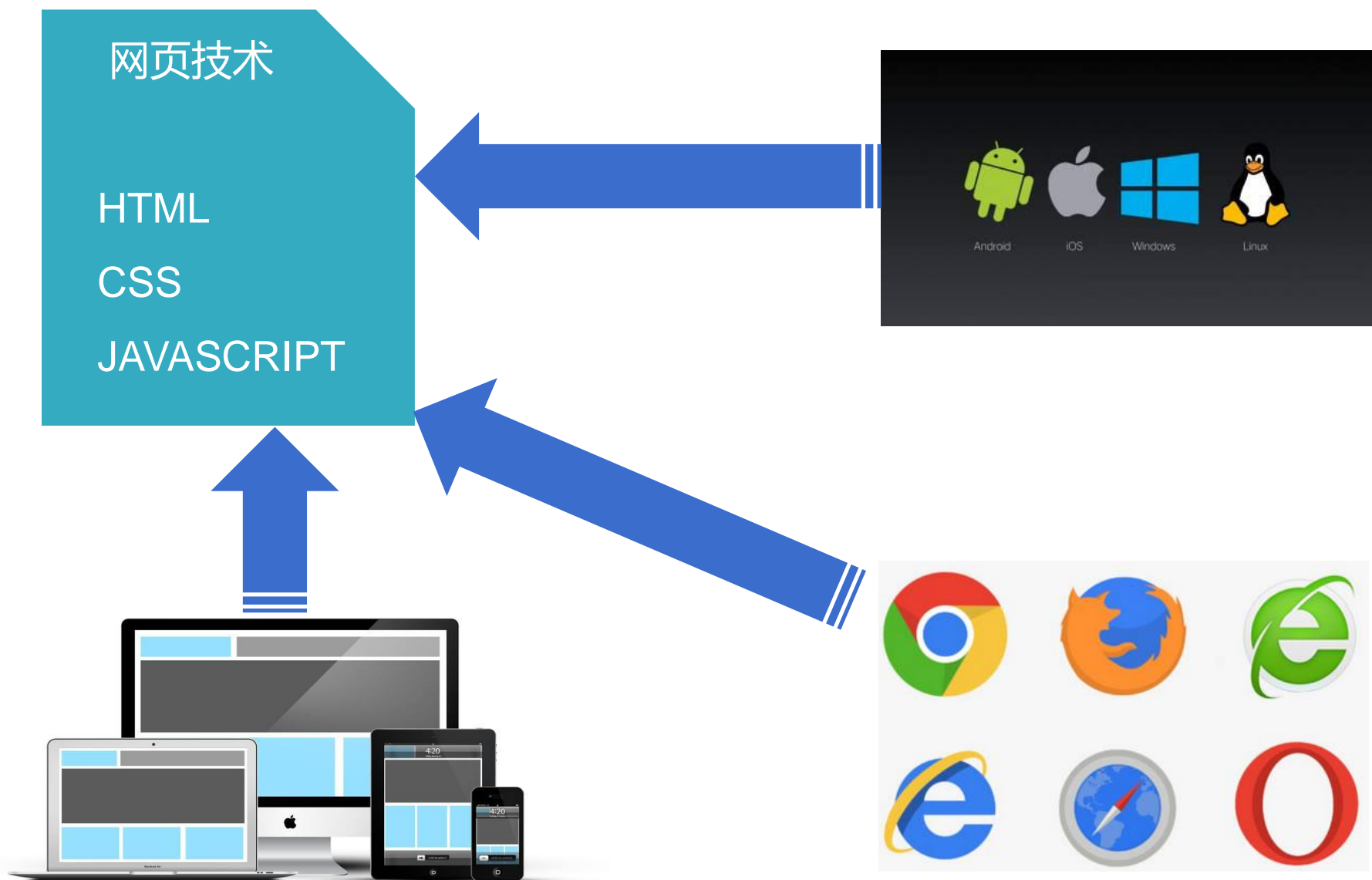
我们看到的所有网页、软件界面都属于数据可视化



其它界面元素, 无论数据多少、显示是否美观、操作是否酷炫, 同样是把数据库里的数据呈现给人看, 仍然属于可视化。



各种语言的界面中，最广泛，最通用的是网页



报表和统计图发布到网页上展现时，仍然是这些技术

什么是Dashboard



有时出于多角度观察相关数据的目的，会把相关的几个报表放到一个网页里，而且几个报表之间的数据存在共享参数、互相联动，但不管多复杂，本质仍然是网页技术，只是内容多，操作复杂了。

报表或BI工具是否支持大屏显示/是否支持手机?

- ★ 大屏显示仅仅是可视化内容变多了，无论从制作方式，还是涉及的软件技术，没有任何新东西。手机显示也是一样。
- ★ 在做界面设计时，要根据不同硬件上的操作特点、界面大小进行合理布局。比如早期800*600和近期1920*1280的分辨率的PC，设计网页时布局会做适应性调整，甚至完全重新设计。
- ★ 工具制作的网页展现到哪里都可以，区别仅发生在制作过程，布局和操作要适应目标场景。





初识BI仪表盘

BI仪表盘是什么

一张数据图表就是一个报表。但观察数据时，通常不是单独的一个报表，而是图文结合，多个图表混合的形式。就类似驾驶室里，很多个仪表盘分别表达不同的数据。驾驶车辆需要仪表盘，那驾驶企业，就需要通过企业数据制作出来的表示企业运营状况的BI仪表盘。

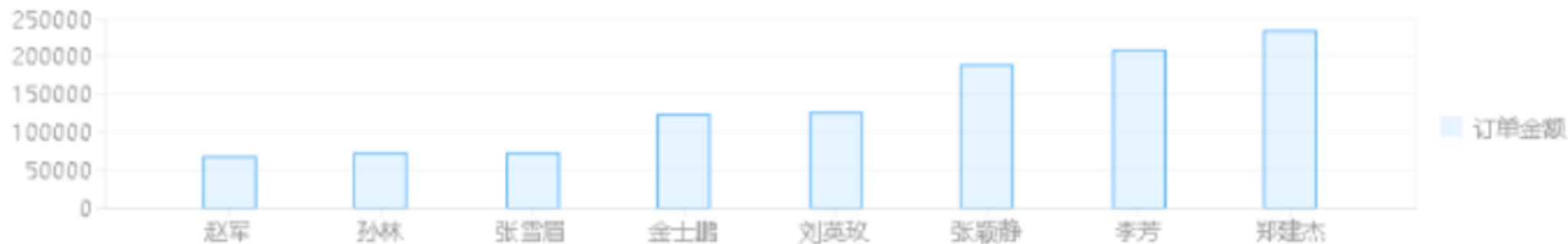


BI仪表盘，名称也可能是：**商业智能仪表盘（BI Dashboard、DBD）、企业驾驶仓。**

以下的课程中采用**DBD**这个简称来表达BI仪表盘。

初识DBD-示例1

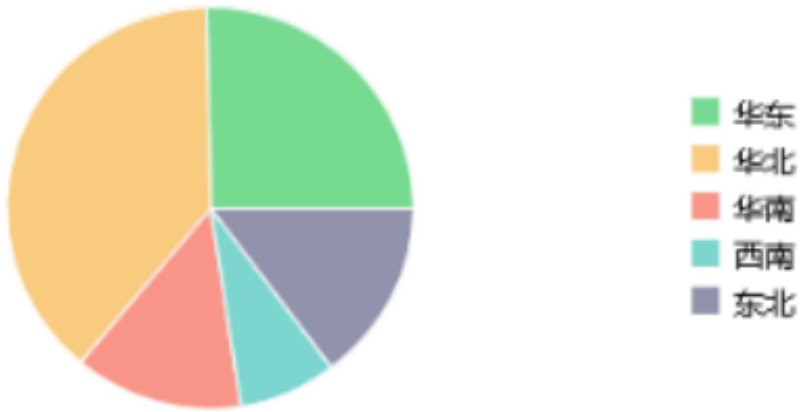
统计图钻取-点击柱子钻取



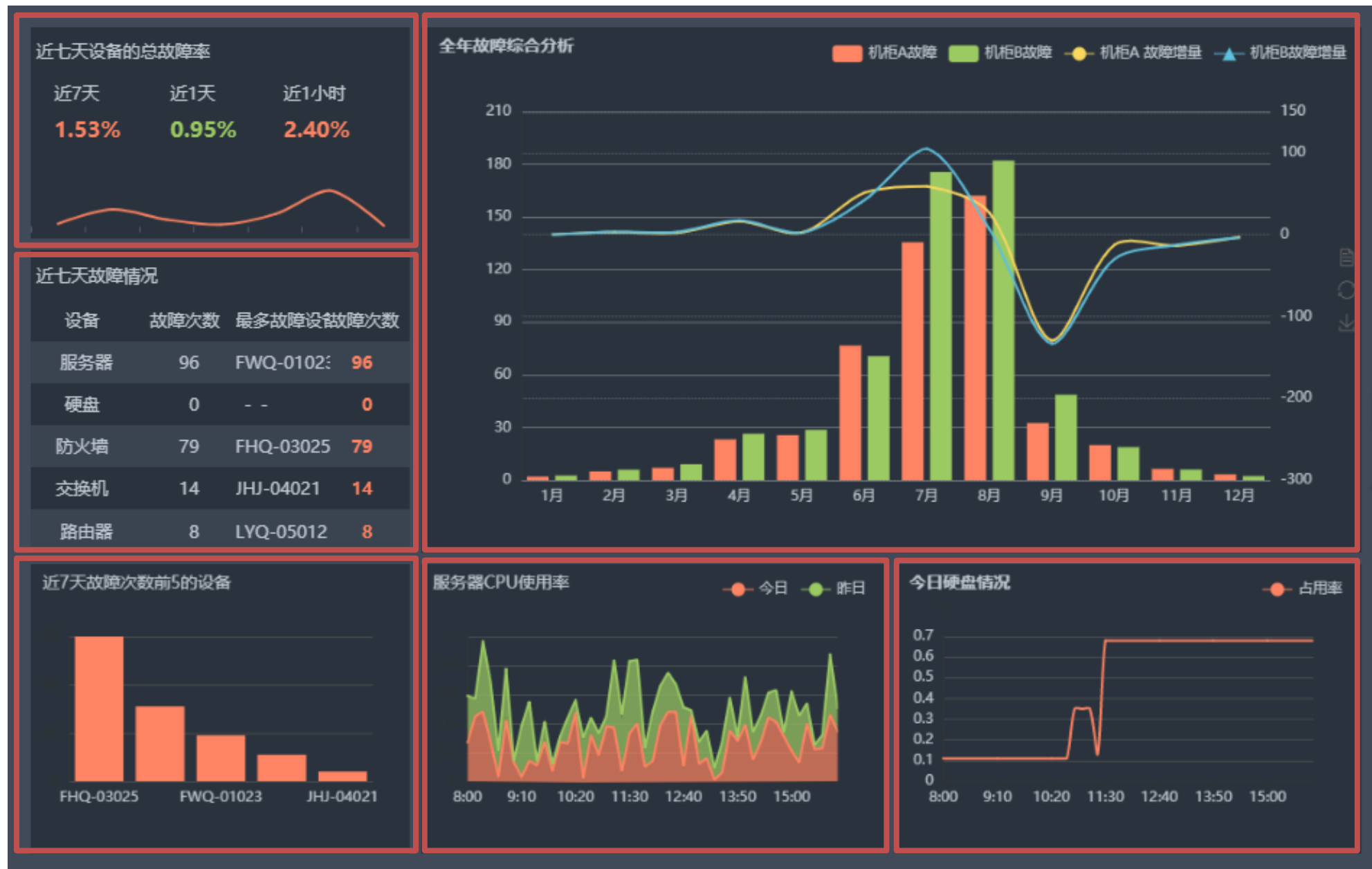
孙林-地区销售表

地区	订单数量	订单金额
华东	20	¥ 18657.62
华北	25	¥ 28563.55
华南	9	¥ 9855.56
西南	9	¥ 5844.44
东北	4	¥ 10812.56

孙林-地区订单金额图



初识DBD-示例2



统一的背景色

初识DBD-示例3

2018电网监控及全国用电情况分析

● 用电总量统计、对比



● 2017年全年用电量与2018年上半年及下半年用电预测量



● 2018年上半年发电情况分布



● 2018年上半年全国各省市用电量



● 全国全社会用电量情况

第一、二、三产业及城乡居民生活上半年用电对比 (亿千瓦时)



● 用电大省的经济发展情况 (前五省份)

用电量 (单位: 亿千瓦时)



● 2018年上半年发电汇总 (含各月情况)

总发电量 (亿千瓦时)

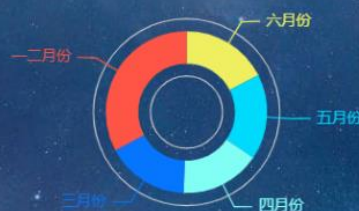
31945.3

同比增长率 (%)

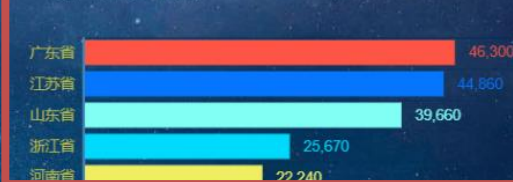
6.7

累计增长率 (%)

8.3



GDP (单位: 亿元)



初识DBD-示例4

一些功能按钮



观察：DBD由复杂布局的多个报表组成



观察--独立于图表外的参数控件控制多表联动

置于三个图
表外的独立
参数控件

选择地区：

华东

选择城市：

上海

开始日期：

2013-01-01

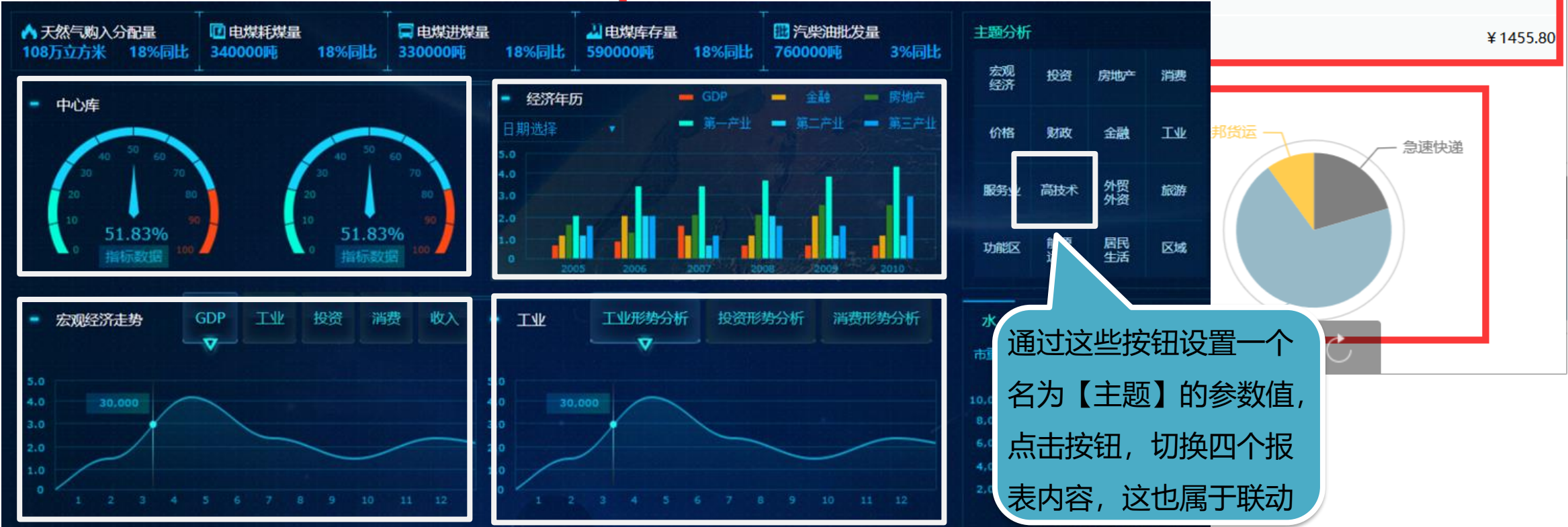
结束日期：

2015-12-31

查询

统计周期：自2013-01-01到2015-12-31

运货商	订单个数	订单金额
急速快递	4	¥ 2974.76
统一包裹	5	¥ 10117.76
		¥ 1455.80



观察：另外一种联动形式--报表之间联动

统计图钻取-点击柱子钻取



观察示例1，三个图表都和“孙林”有关。期望点击第一个图里的“孙林”柱子时，下面两个图也变成“孙林”的数据。这是报表之间的联动。

孙林-地区销售表

地区	订单数量	订单金额
华东	20	¥18657.62
华北	25	¥28563.55
华南	9	¥9855.56
西南	9	¥5844.44
东北	4	¥10812.56

孙林-地区订单金额图



无论哪种方式联动，都可以理解成通过**共享参数**实现，通过相同的参数值让相关的报表达达到表现一致。

DBD总结

组成：主体是多个报表；独立的共享参数控件、标签页等控制控件及一些装饰性区域为点缀。

独立的共享参数控件产生的**联动**、报表之间的**联动**是DBD界面中主要的交互方式。

DBD怎么实现

1、制作局部的每个报表，这在以前的课程中详细过。

2、针对多个报表、各种控件进行**布局**，这是制作DBD的重要本职工作。

3、布局同时，考虑共享参数如何实现联动。

报表工具主要的功能是制作报表，有时也会提供多个报表组合的功能。下来我们用报表工具的**报表组布局**功能试试实现下示例1。

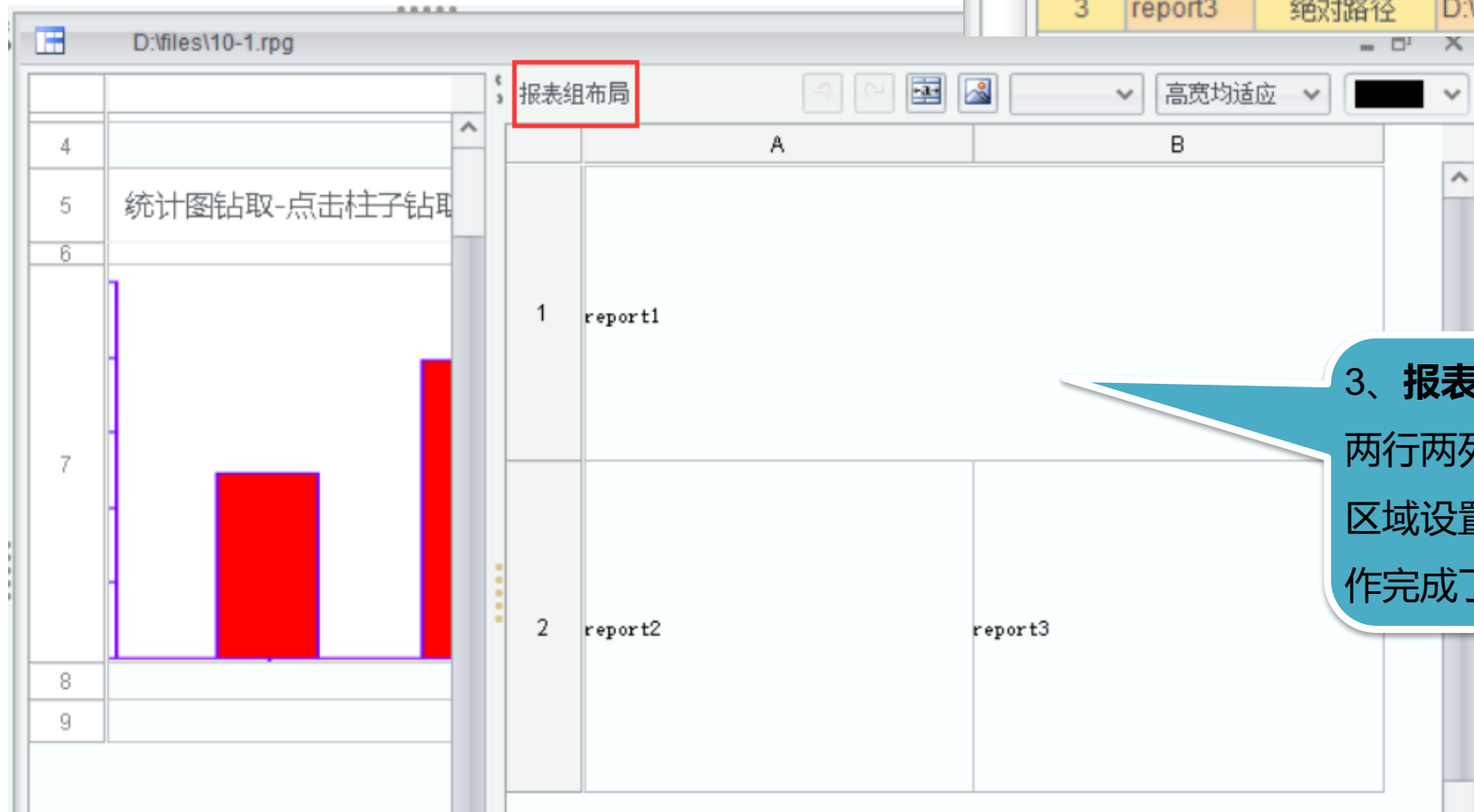


用报表组布局功能 快速实现示例

报表组布局功能实现布局

1、参照以前的课程单独制作出三个独立报表10-2.rpx、10-3.rpx、10-4.rpx。

2、用这三个报表创建报表组10-1.rpg。



3、**报表组布局**里删除多余行列，剩余两行两列；A1,B1设置合并，并对每个区域设置上相应报表，这样基本上就制作完成了，web上预览就能看到效果。

报表工具的参数模板制作独立的参数控件



	A	B	C	D	E
1					
2		姓名:		奖金>	
3		性别:		出生日期>	1970-01-01

编辑风格 下拉列表框 设置(S) 确定(O) 取消(C)

☒ 是否可写

编辑风格 编辑框 设置(S) 确定(O) 取消(C)

☒ 是否可写

姓名: 奖金> 查询

性别: 出生日期> 1970-01-01

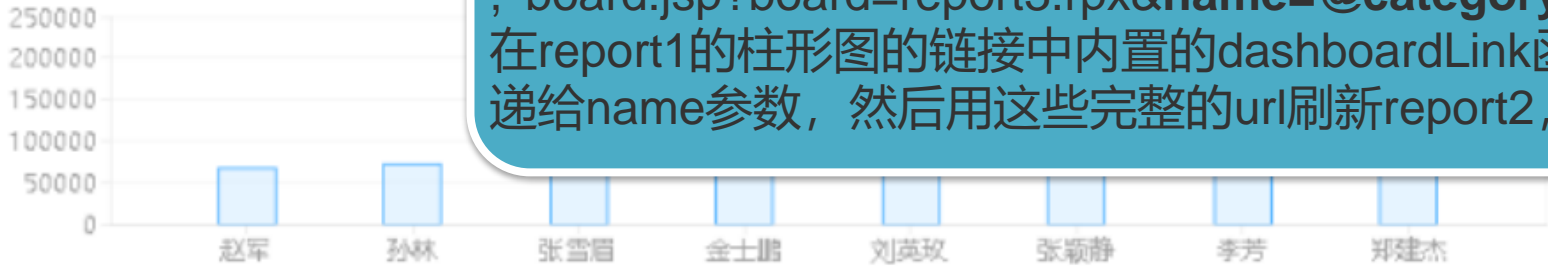
通过简单的设置，发布后，WEB上生成独立控件给报表和DBD提供参数。

四个参数固定参数名设置好参数值后，通知相关报表按照参数值刷新，可以灵活决定哪个报表被哪些参数影响。

实现图表之间联动

= "javascript:dashboardLink(
'board.jsp?board=report2.rpx&name=@category', 'report2'
, 'board.jsp?board=report3.rpx&name=@category', 'report3')"
在report1的柱形图的链接中内置的dashboardLink函数，把当前姓名这个分类值传递给name参数，然后用这些完整的url刷新report2，report3区域，实现联动效果。

统计图钻取-点击柱子钻取



孙林-地区销售表

地区	订单数量	订单金额
华东	20	¥ 18657.62
华北	25	¥ 28563.55
华南	9	¥ 9855.56
西南		
东北		

孙林-地区订单金额图



下面这两个报表制作的时候要定义name参数，根据name动态生成图标。

参数编辑

名称	描述	数据类型	值表达式	参数类型	
name	姓名	字符串	孙林	普通参数	<div>确定(O)</div> <div>取消(C)</div>

用报表工具做DBD，看起来把功能基本实现了，但实际上这种方式只能制作出前三个示例。

WEB技术开发方式

制作DBD

观察1、DBD里一些非标准报表、控制组件



一些非标准的文字报表，还有滚动轮播消息这类动态效果

异动指标

2017年乌鲁木齐高新区服务业7-12月经济运行报告

2017年乌鲁木齐高新区软件和信息服务业1-6月经济运行报告

2017年乌鲁木齐高新区软件和信息服务业7-12月经济运行报告

2017年乌鲁木齐高新区软件和信息服务业1-6月经济运行报告



一些非标准的统计图报表



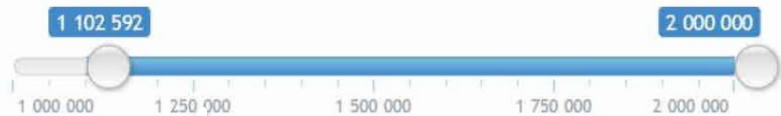
漂亮效果的标签页、分页



这些用报表工具很难做出来，只能HTML+JS开发实现。

观察2、DBD中一些输入、布局控件

这些用报表工具做不出来，需要基于针对开源JS控件做二次开发。



滑动方式的数值范围选择器

2017-04-20

2017-04-20 周四

1	17
2	18
3	19
2017 年	4 月 20 日
2018	5
2019	6
2020	7

多种样式的日历控件

加多宝凉茶, 饮料, 成功图像...

搜索

产品

- ☒ 加多宝凉茶
- ☒ 饮料
- ☐ 太太乐鸡精
- ☐ 嘉士伯啤酒

字典

- ☒ 成功图像类型
- ☐ 门店类型

多选下拉

选择城市时，按下 Ctrl 或 Cmd 键可以进行多选

城市: 北京, 重庆, 保定 选择

- 北京
- 天津
- 上海
- 重庆
- 河北省
 - 石家庄
 - 保定
 - 邯郸
 - 承德
- 广东省
 - 广州
 - 深圳

多选下拉树

北京市

省份

北京市

山西省

带搜索功能的多级下拉选择

A1

B1

C1

C2

C3

B2

嵌套的布局

拖拽区域

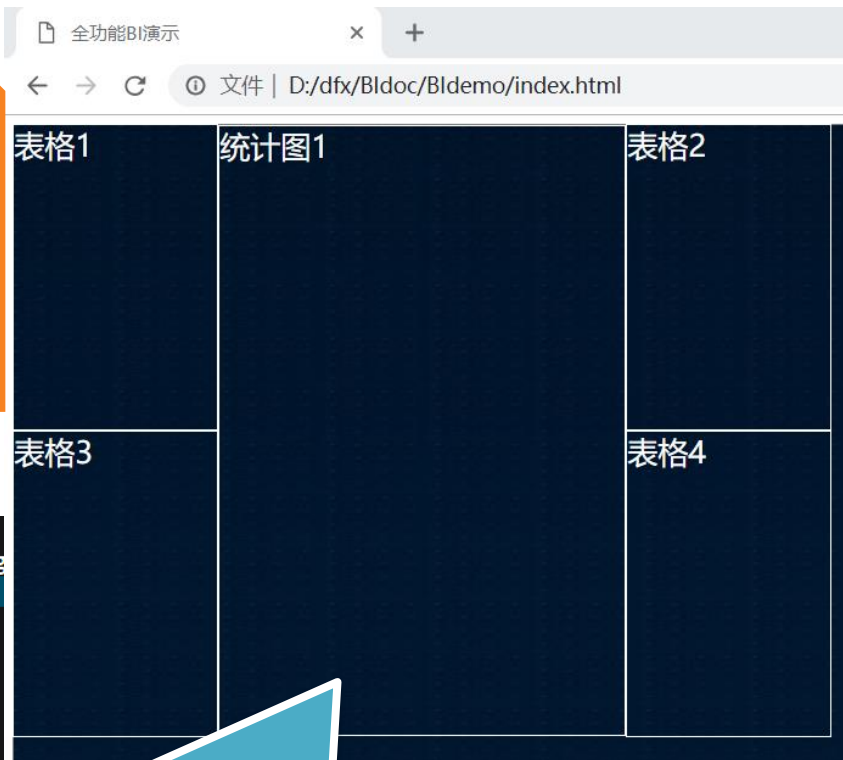
我是内容我是内容我是内容我是内容我是内容我是内容我是内容

可拖拽改变大小的浮动框

结论：报表工具对复杂的DBD帮助不大

最近几年新热的“大屏显示”是典型的布局复杂，要求精美的DBD；这种DBD经常用上两页这些局部，只能用HTML+JS硬写出来。用报表工具做DBD实现的简单布局，手写HTML也容易实现。所以，复杂精美DBD只能用WEB技术开发方式；报表工具帮不上忙。

```
17 <body>
18 <div style='color:#FFF;font-size:30px;width:5000px;height:2000px;border:1px solid g
19 <table border=0 style="border:0;border-collapse:collapse;border:0px;margin:0px 0
20 <tr>
21 <td>
22 <div style='width:200px;height:300px;border:1px solid #FFF;'>表格1</div>
23 </td>
24 <td rowspan=2>
25 <div style='width:400px;height:600px;border:1px solid #FFF;'>统计图1</div>
26 </td>
27 <td>
28 <div style='width:200px;height:300px;border:1px solid #FFF;'>表格2</div>
29 </td>
30 </tr>
31 <tr>
32 <td>
33 <div style='width:200px;height:300px;border:1px solid #FFF;'>表格3</div>
34 </td>
35 <td>
36 <div style='width:200px;height:300px;border:1px solid #FFF;'>表格4</div>
37 </td>
38 </tr>
39 </table>
40 </div>
```



用html里的table布局一个两行三列的图表组合，中间一列合并，每个格子里用iframe展示独立的报表。

简单HTML+JS代码功能演示

```
<script>
$(document).ready(function() {
  $('#dateinput').datepicker({
    changeMonth: true,
    changeYear: true,
    dateFormat : 'yy-mm-dd'
  });
})
</script>
```

```

<script>
$(document).ready(function() {
    $('#td1').find("span").css('padding','0 10px').click(function(){
        $('#td1').find("span").css({'background-color':''});
        $(this).css({'background-color':'#AD1345'});
        $('#td1').find('div[idx]').css({'display':'none'});
        $('#td1').find('div[idx='+$(this).attr('idx')+']').css({'display':'block'});
    });
    $('#td1').find("span")[0].click();
})
</script>
<td id='td1' style='border:1px solid #FFF;'>
<div style='width:600px;border:1px solid #FFF;'>
    <span idx=1>标签11</span><span idx=2>标签12</span>
</div>
<div idx=1 style='width:600px;height:550px;border-top:1px solid #FFF;'>
<div idx=2 style='width:600px;height:550px;border-top:1px solid #FFF;'>
</td>

```

图表外表单参数:

[illegible]

图表外表单参数: 分类1 分类2 分类3 | 2020-04-21 | ☒选项1 ☐选项2 ☒选项3

标签11 标签12

表格11

各式各样的参数输入方式，比如按钮选择、日期选择、多选等。

file:///D:/dfx/BI doc/BI demo/index.html

标签11 标签12

统计图1

表格1

演示标签页功能，同一个区域切换或轮播报表。

表格3

§11</div>
§12</div>

这里演示只用了最基础的JS实现，而实际上WEB开发时，经常会引入一些功能齐全的JS控件，提高WEB开发效率，但这些控件只提供接口，需要技术人员学习才会用起来。

思考DBD的技术本质

通过拖拽设计
DBD图表、布局
及共享参数控件。

把拖拽生成的内置配置翻译成WEB需要的HTML、CSS、JS；同时需要引入内置的、各种功能的JS控件库。

WEB端展现、操作；这时的浏览器区分不出来你到底是普通网页、还是多维分析图表、甚至DBD。

既然浏览器区分不出来普通网页和DBD，那制作DBD和开发普通网页有什么区别？

技术上没区别，但DBD更偏重呈现，对艺术要求高。

那制作DBD为什么不能做到完全工具化实现？

艺术占比大，怎么搭配展现更美、更和谐没有标准答案；

功能繁杂，控件众多，用户需要控制的细节太庞大。

工具帮助制作DBD，本质是**对WEB开发，进行有限功能的可视化。**

WEB技术方式制作DBD不能应对需求频繁变化怎么办？

WEB技术方式开发DBD，周期太长，那就让业务人员直接制作DBD，BI工具就是做这个事情，但各方面程度都会受限。

BI工具
制作
DBD受
限程度

采用模板的方式快速得到一些通用布局。复杂布局和丰富样式的参数控件搞不定。

只能生成简单些的多维分析报表。报表工具制作的复杂样式报表、WEB方式制作的个性化报表都做不到。

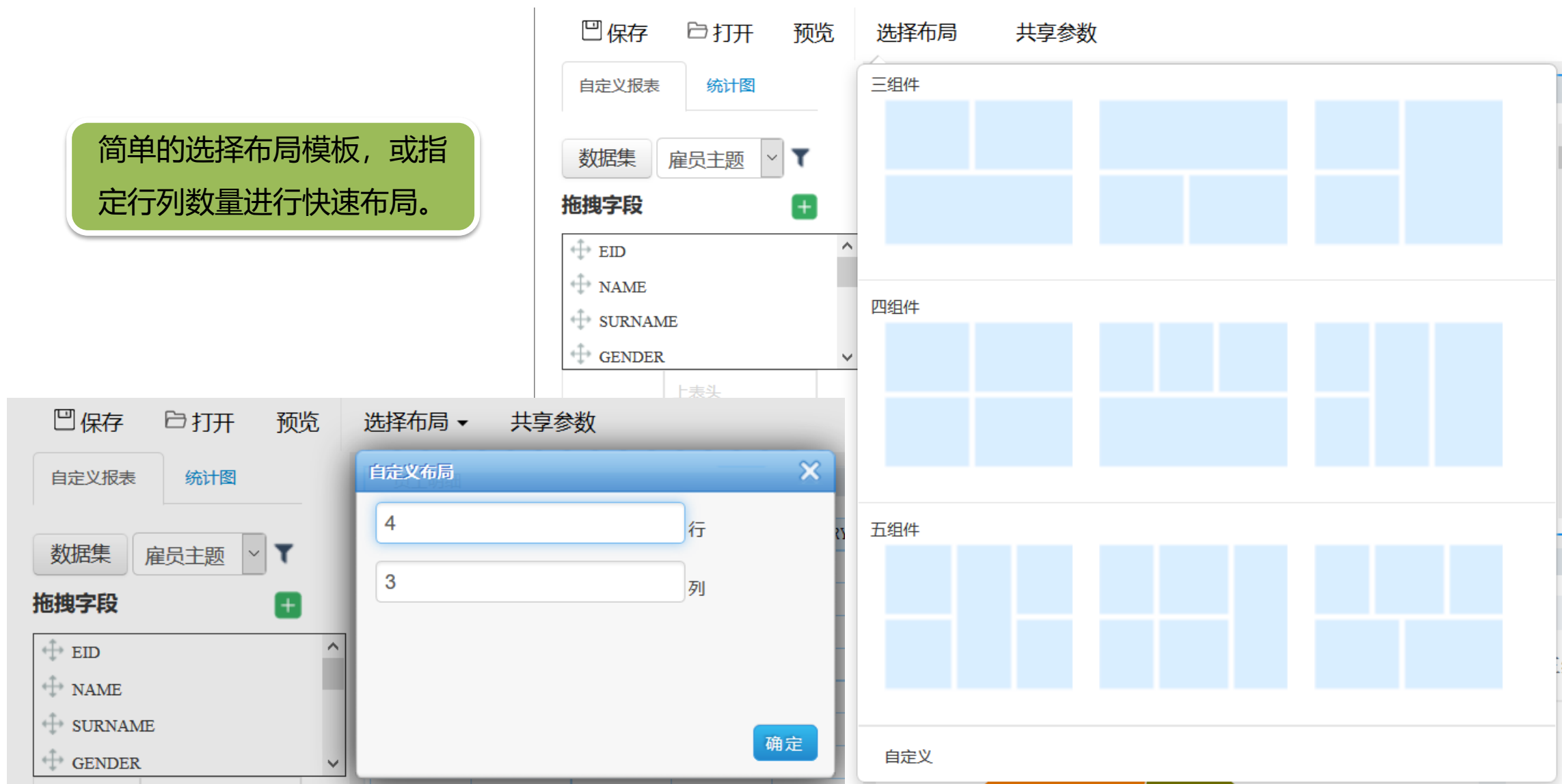
通过给数据集设置共享参数能实现较低程度的联动。图表之间的联动不太容易实现。



BI工具制作DBD

布局模板快速布局

简单的选择布局模板，或指定行列数量进行快速布局。



拖拉拽方式的多维分析生成每一个报表

保存 打开 预览 选择布局 共享参数

自定义报表 统计图

数据集 雇员主题

拖拽字段

EID
NAME
SURNAME
GENDER

上表头
指标标题

左表头

管理数据集

雇员主题
订单主题

雇员主题

DQL表 数据库查询 (DQL/SQL) DFX文件 DFX脚本 填报文件

demo

select * from employee

每个多维分析报表是针对一个数据集做的，但DBD里的不同的报表可能针对不同数据集，所以有个数据集管理，来维护多个数据集。

DBD每块区域能被选中，总是针对选中的区域制作多维分析报表。

保存 打开 预览 选择布局

自定义报表 统计图

数据集 SQL数据集

拖拽字段

SURNAME
GENDER
STATE
BIRTHDAY

STATE
指标标题
EID计数
DEPT
总薪资

员工明细					
EID	STATE	BIRTHDAY	DEPT	SALARY	AGE
1	California	1974-11-20	R&D	7000	45
2	New York	1980-07-19	Finance	11000	39
3	New Mexico	1970-12-17	Sales	9000	49
4	Texas	1985-03-07	HR	7000	35
5	Texas	1975-05-13	R&D	16000	44
6	California	1984-07-07	Sales	11000	35
7	Illinois	1972-08-16	Sales	9000	47
8	California	1979-04-19	Marketing	11000	40

图表名称1							
DEPT	STATE	Alabama		Arizona		Arkansas	
		EID计数	总薪资	EID计数	总薪资	EID计数	总薪资
Administration		0	0	0	0	0	0
Finance		0	0	0	0	0	0
HR		0	0	0	0	0	0
Marketing		1	6500	0	0	0	0
Production		0	0	2	13500	0	0
R&D		0	0	0	0	1	8000
Sales		3	23500	5	40500	0	0
Technology		0	0	2	15000	0	0
总计		4	30000	9	69000	1	8000

共享参数实现简单联动

保存 打开 预览 选择布局 共享参数

自定义报表 统计图

数据集 雇员主题

拖拽字段

- EID
- NAME
- SURNAME
- GENDER

上表头

指标标题

左表头

- EID
- STATE
- BIRTHDAY
- DEPT
- SALARY

共享参数

名称	值	类型
员工姓名	张静	字符串
订单日期	2018-03-01	日期
年份	2018	字符串

添加

保存 退出

DBD里增加的共享参数，用在数据集的查询条件里。给共享参数设置的值，会作用到所有使用这些数据集的报表。

自定义报表 统计图

数据集 雇员主题

设置查询条件

☐ 简洁显示 (简洁显示会隐藏掉“新增条件、调整条件层次”的功能，更能凸显条件之间的层次关系)

NAME 等于 张静

AND | OR

员工姓名

EID	STATE	BIRTHDAY	DEPT	SALARY	AGE
1	California	1974-11-20	R&D	7000	45
2	New York	1980-07-19	Finance	11000	39
3	New Mexico	1979-12-17	Sales	9000	40

STATE
DEPT
Administration



手机端DBD

手机端DBD特点

★一般自上而下的纵向排列

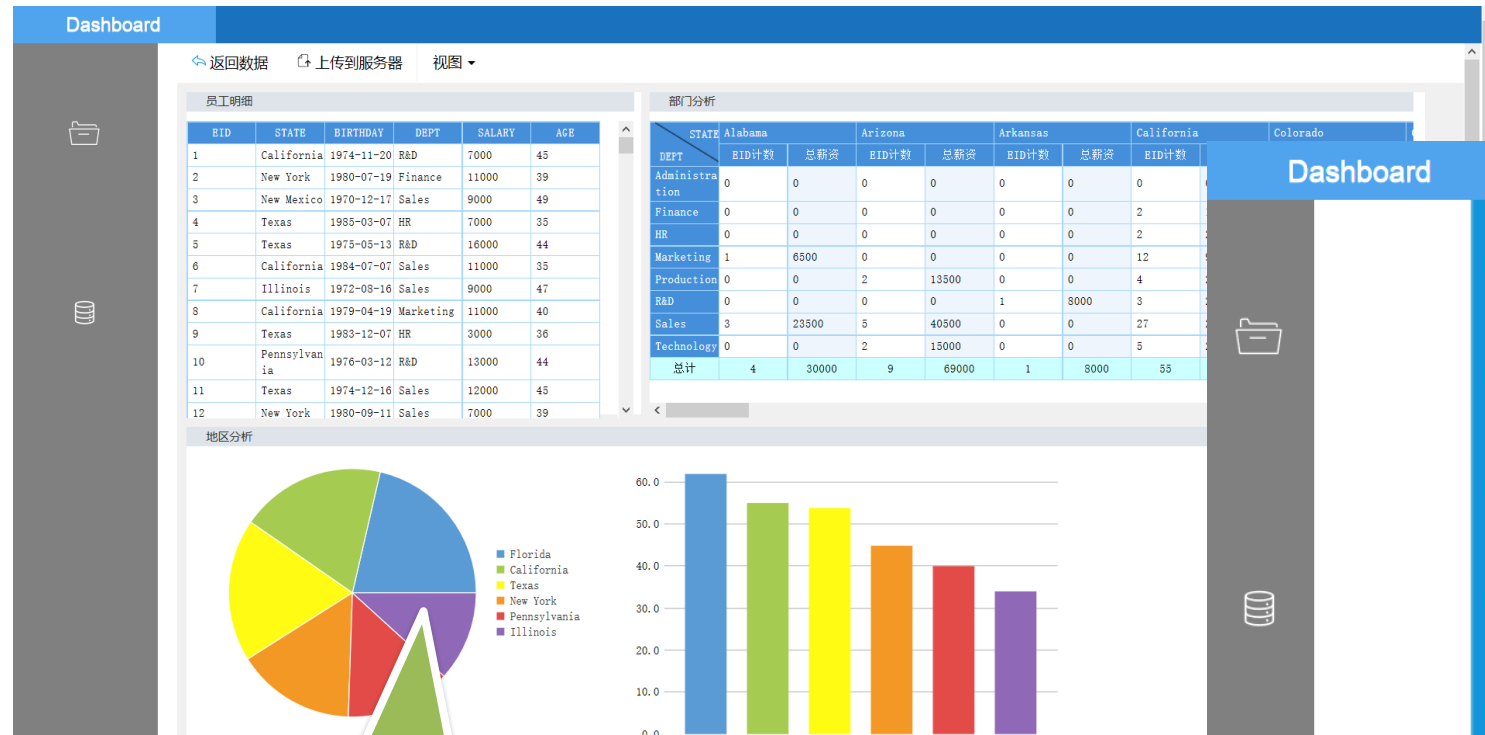
★手机型号不同，分辨率各异

★交互方式和电脑端差异大

针对这些特点，手机端DBD比较重要的一个功能是样式布局**自适应**。H5里有根据屏幕宽度自适应的设置。实际所做的局部报表还要具体的观察自适应后是否和谐。

```
<meta name="viewport" content="width=device-width, user-scalable=no, initial-scale=1.0, maximum-scale=1.0, minimum-scale=1.0"/>
```

BI工具做的DBD在手机端自动布局



在电脑上浏览DBD，会看到和设计布局一样的效果

切换成手机上浏览，布局自动变换成了自上而下

这种简单的逐个从上到下排列局部，效果不会太好。如单个报表太宽、共享参数的交互不适合手机等问题。需要效果好同样需要WEB技术方式开发。



WEB技术方式做效果好的DBD，电脑端和手机端分别开发



两种终端在显示、交互差异比较大，要做细致时，纠缠在一起耦合性很大。分开开发更有效率。

即便分开开发了，手机端因为分辨率复杂，自适应能力仍然需要较大工作量。



结论

DBD制作方式的对比

上面介绍了三种实现DBD的方式，我们把DBD拆成**制作单个报表**和**制作整体页面**两部分，对比下三种方式的特点。

	制作每个报表			制作整体页面		
	要求技术能力	制作方式	效果	要求技术能力	制作方式	效果
报表组布局	高	报表工具	复杂度高的报表	低	报表组布局	互动简单，没有标签页、丰富样式的共享参数控件等
WEB技术开发	高	报表工具+WEB手工方式	复杂度高、个性化度高的报表	高	HTML,JS直接开发	任意布局，任意界面互动，美术人员可以重度参与，但丰富的JS控件开发工作量很大
BI工具制作	低	多维分析	简单的多维分析报表	低	选模板或简单自定义行列	互动简单，没有标签页、丰富样式的参数控件等

DBD制作结论

★DBD最终的展现效果由每个报表和整体布局决定。要效果都好（如大屏），就只能投入更多的技术进行开发。

★对DBD要求不高时，报表工具/BI工具能提供一些帮助。

★不同制作方式不兼容，工具是个封闭的系统，很难对它二次开发，来加快WEB技术开发DBD。

不论是常规DBD、大屏DBD，还是手机端DBD，这些结论都适用

THANKS

感谢观看



乾学院 · VIP课堂

商业智能 应用部署技术



CONTENTS

本章内容

1. 集成部署
2. 独立部署
3. 应用结构
4. 报表没完没了
5. 按需建模



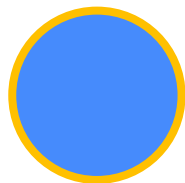
部署方式

BI应用在实际使用中集成部署和独立部署两种方式，可以根据实际情况选择



集成部署

将BI应用模块嵌入已有系统中，BI模块负责数据分析呈现，用户管理、权限控制等管理功能则借助已有系统实现



独立部署

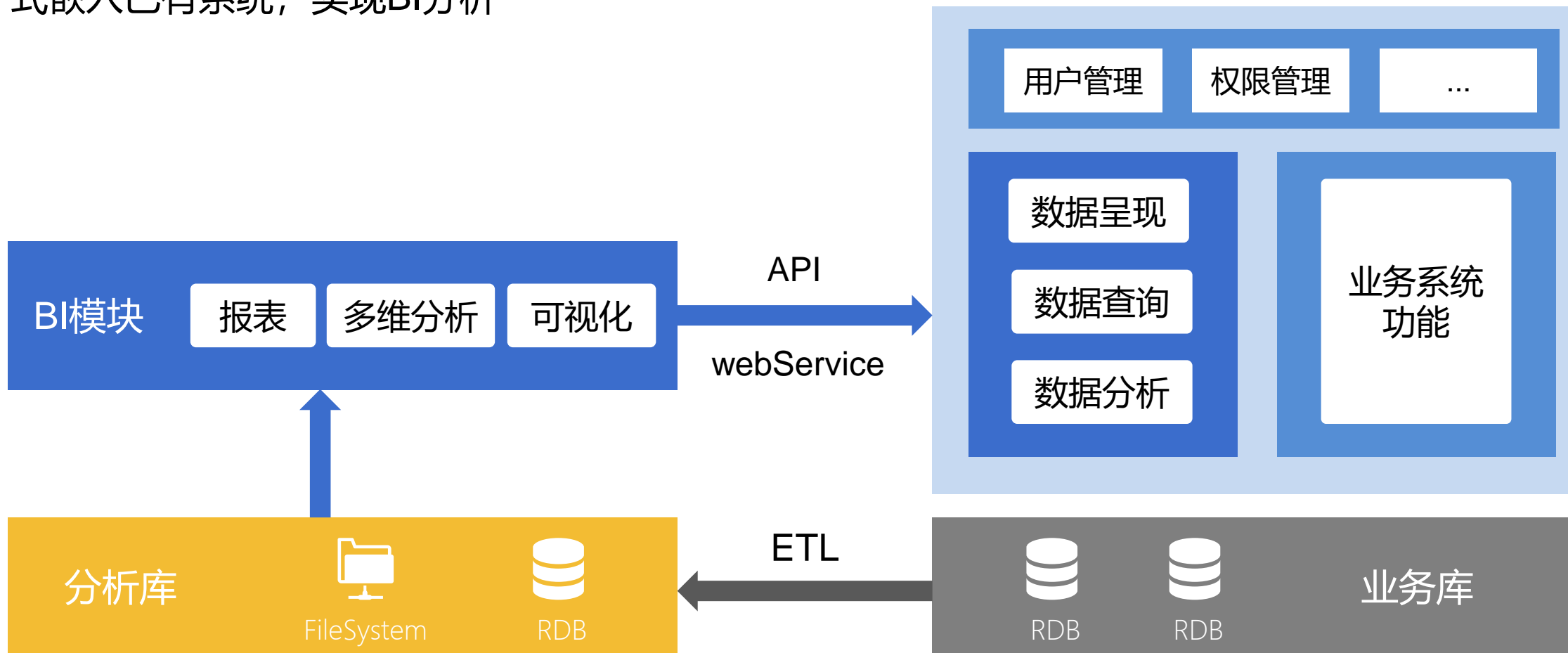
BI应用作为完整系统部署，应用本身除了BI分析功能外，还包括用户管理、权限控制等管理功能



集成部署

集成部署

将多维分析、报表等BI模块通过API或服务形式嵌入已有系统，实现BI分析



两种集成方式

嵌入式紧集成

将BI模块嵌入已有系统中，
通过API方式进行调用，物理
上仅部署在一处

服务式松集成

将BI模块发布成独立的服务，
已有系统通过服务方式访问
BI模块，物理上可以部署在
两处或多处

两种集成方式无优劣之分，可根据实际情况选择；但两种方式注意事项有所不同

嵌入式集成注意事项

1.API是否丰富以满足需要

嵌入式集成时，BI产品要提供丰富的API供已有系统调用，包括：

1. 报表定义接口
2. 参数控制接口
3. 报表计算接口
4. ...
5. 元数据获取接口
6. 权限控制接口
7. 数据源访问接口
8. ...

API丰富程度

嵌入式集成注意事项

2.JAR包冲突

- 嵌入式集成时，工程中已引入的第三方JAR与BI模块中使用的第三方JAR存在冲突
- 此时可以选择使用最高版本
- 事实上，BI产品应提供适当机制尽量避免冲突的出现

JAR冲突

嵌入式集成注意事项

3.配置文件冲突

- 嵌入式集成时，工程中已有的配置文件与BI模块中的配置文件冲突
- 如常见的web.xml
- 出现这种情况时，需要将配置文件的内容合并来解决冲突问题

配置文件冲突

服务式集成注意事项

1. 权限控制

系统调用BI模块功能时需要提供足够的参数以完成数据查询、权限控制等目标

可能需要传递的参数包括：

1. 用户角色信息
2. 机构信息
3. 数据过滤参数信息
4. ...

权限控制

服务式集成注意事项

2.安全性

- 采用服务式调用BI模块功能时，需避免URL被复制进行非法访问、参数被修改、以及参数过长被截断等问题
- 此时需对服务请求增加加密机制以保证安全性

安全控制

服务式集成注意事项

3.跨域问题

- 采用服务式调用BI模块功能时，当使用导出打印功能，或BI模块回调业务系统功能时可能存在跨域访问问题
- 此时需要BI产品提供相应的解决机制

跨域访问

集成部署举例

- 润乾报表支持嵌入式和服务式集成部署
- 其中服务式部署很简单，将产品自带的web应用直接发布后，通过web服务访问即可
- 这里我们以集成式部署为例，简单说明集成过程（也非常简单）

1.拷贝相应目录

将raqsoft目录拷贝到项目根目录（名称不可修改）

2.引入jar包

将dm.jar和raqsoftReport.jar等必须jar引入到项目中

3.添加/合并配置 文件

添加raqsoftConfig.xml，合并web.xml

详细部署内容参考：<http://doc.raqsoft.com.cn/report/preference/wyym184.html>

集成部署举例

集成部署后，BI模块仅提供数据查询、分析、呈现功能，不涉及用户与系统管理



The screenshot displays the BI module interface. At the top is a dark navigation bar with icons for file operations (save, open, data files) and visualization settings (horizontal/vertical layout, overlapping display). The main area is divided into three sections: a report title bar, a data table, and a data analysis settings panel. The data table shows sales data categorized by region (Northeast, East, North China, South China) and province, with columns for sales amount and total. The settings panel on the right allows for configuring the data source (Sales Record) and the dimensions to be analyzed (Region, Product, Customer).

上级地区: 上级地区1		销售额求和	总计
东北	辽宁	55787	55787
	总计	55787	55787
华东	上海	8662	8662
	山东	15224	15224
	江苏	183956	183956
	江西	33920	33920
	总计	241762	241762
华北	北京	56387	56387
	天津	361406	361406
	河北	104917	104917
	总计	522710	522710
华南	浙江	51823	51823
	广东	220864	220864
	海南	24730	24730

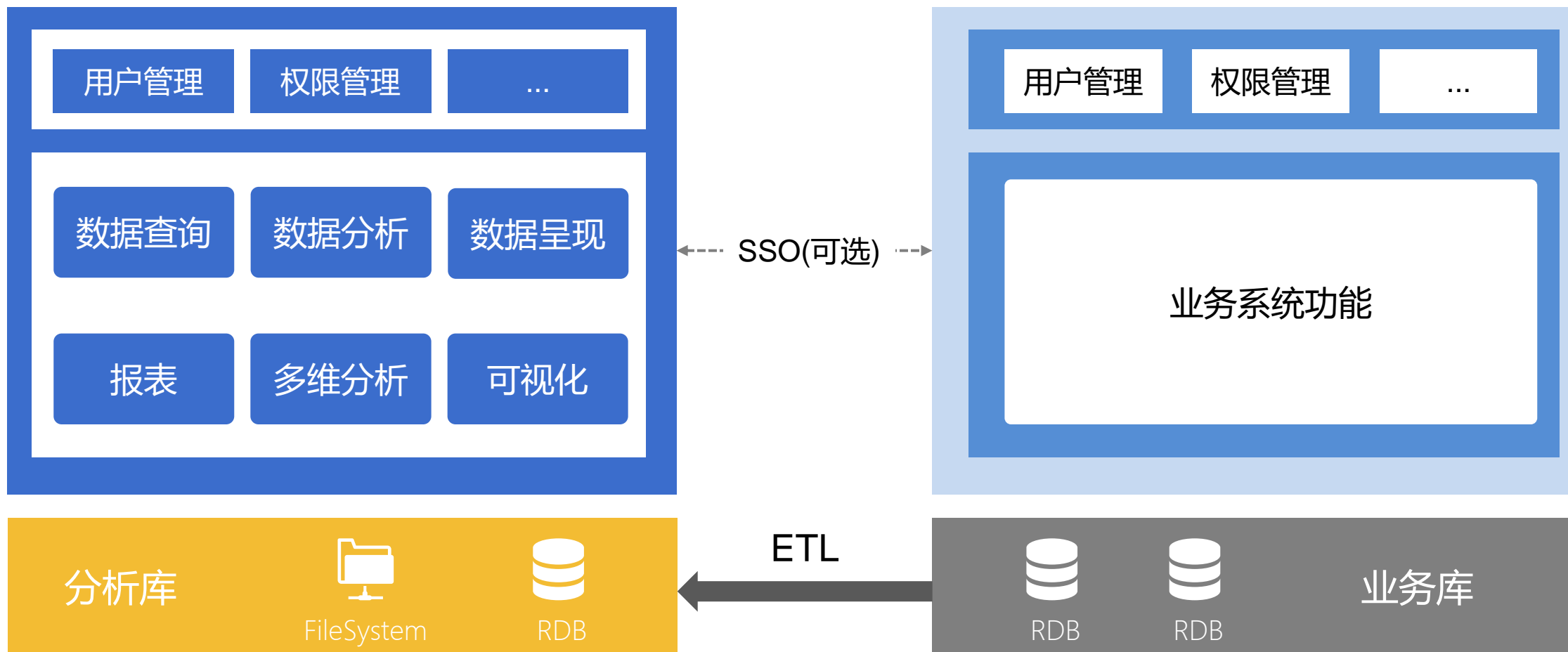
集成后BI模块提供多维分析功能



独立部署

独立部署

将多维分析、报表等BI功能作为一个完整应用系统，包含用户、组织机构、权限设置等系统管理功能



独立部署注意事项

1.需具备强维护能力

- BI应用作为独立应用部署后，与已有业务系统形成完全独立的两个系统，此时BI应用除了要满足日常的查询统计分析需求
- 还要维护一套完整的用户、机构、权限体系，往往这些内容还要与业务系统一致
- 用户、机构等信息发生变化时，要同步维护两套系统
- 因此独立部署的BI应用需要使用者具备较强的维护能力

强维护能力

独立部署注意事项

2.单点登录 (SSO) 问题

- BI应用作为独立应用部署后，为了满足单系统登录，多系统使用的需求，可能要实现单点登录 (SSO)
- 这时需要借助第三方工具或自行实现

单点登录

独立部署举例

将润乾报表以独立应用部署后，报表中心除了可以查询数据外，还可以独立管理用户、机构

报表中心

← → ↺ localhost:6868/d

润乾报表中心

报

润乾报表DEMO

- Dashboard
- 基本报表
- 高级报表
- 交互报表
- 集算器相关
 - 交叉报表
 - 同期比
 - 地铁综合查询
 - 多源分片
 - 子表动态插入主表
 - 库间关联
 - 归并分组
 - 按段分组
 - 条件控制分组格式
 - 横向分栏
 - 横向拼接列表
 - 错行分栏

润乾报表中心

显示节点 机构管理 用户管理 报表管理 查询分析管理 管理员退出 当前用户: 管理员: admin 密码管理

+ 添加

删除

设置

<input type="checkbox"/>	用户名称	机构名称	邮箱	电话	操作	dql宏
<input type="checkbox"/>	raq_visitor	访客				编辑
<input type="checkbox"/>	王伟	副总裁			保存	编辑
<input type="checkbox"/>	郑建杰	销售经理			保存	编辑
<input type="checkbox"/>	张颖	销售代表			保存	编辑
<input type="checkbox"/>	李芳	销售代表			保存	编辑
<input type="checkbox"/>	李松波				保存	编辑

用户管理



应用结构

应用结构导致的运维困难

采用合理的应用结构可以降低BI系统运维复杂度，通常BI系统在运维中会面临两个主要问题，导致运维成本居高不下

报表没完没了

报表的业务稳定性天生很差，业务开展会不断刺激出新的查询统计需求，报表没完没了，专业程序员做报表成本高，缺乏低成本应对方案

多维分析频繁建模

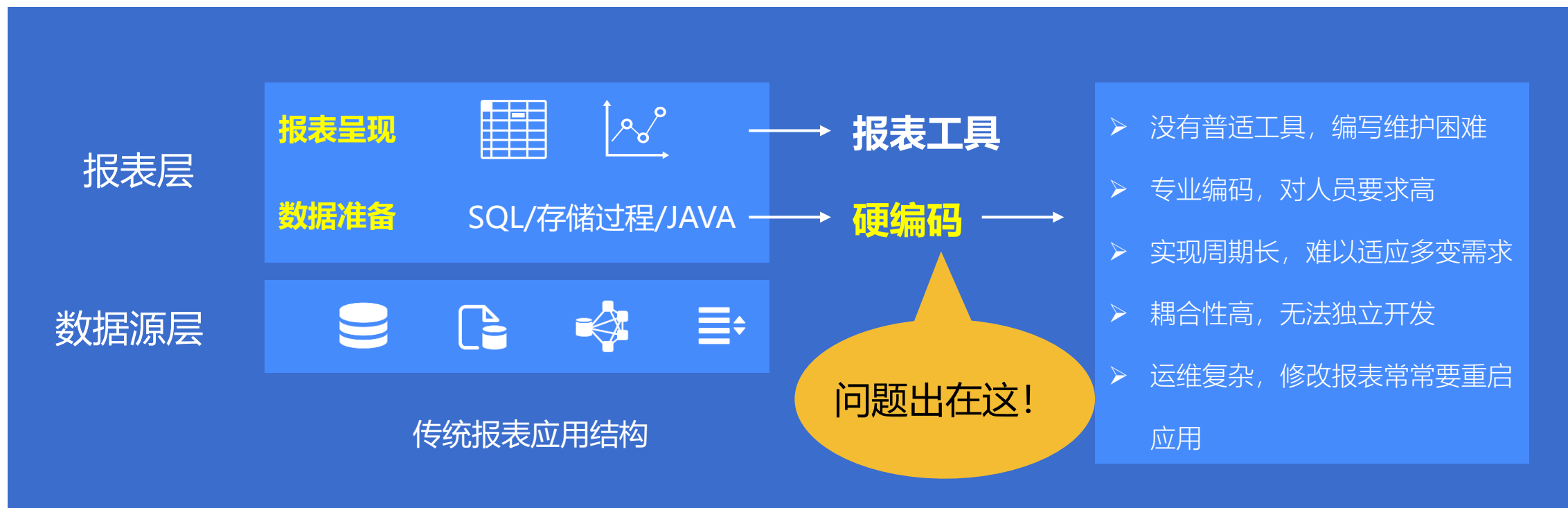
多维分析/自助报表会随着需求增加或变化由技术人员不断修改模型（宽表），导致自助分析“不自助”，技术人员频繁参与成本高



报表没完没了

报表没完没了带来的运维成本

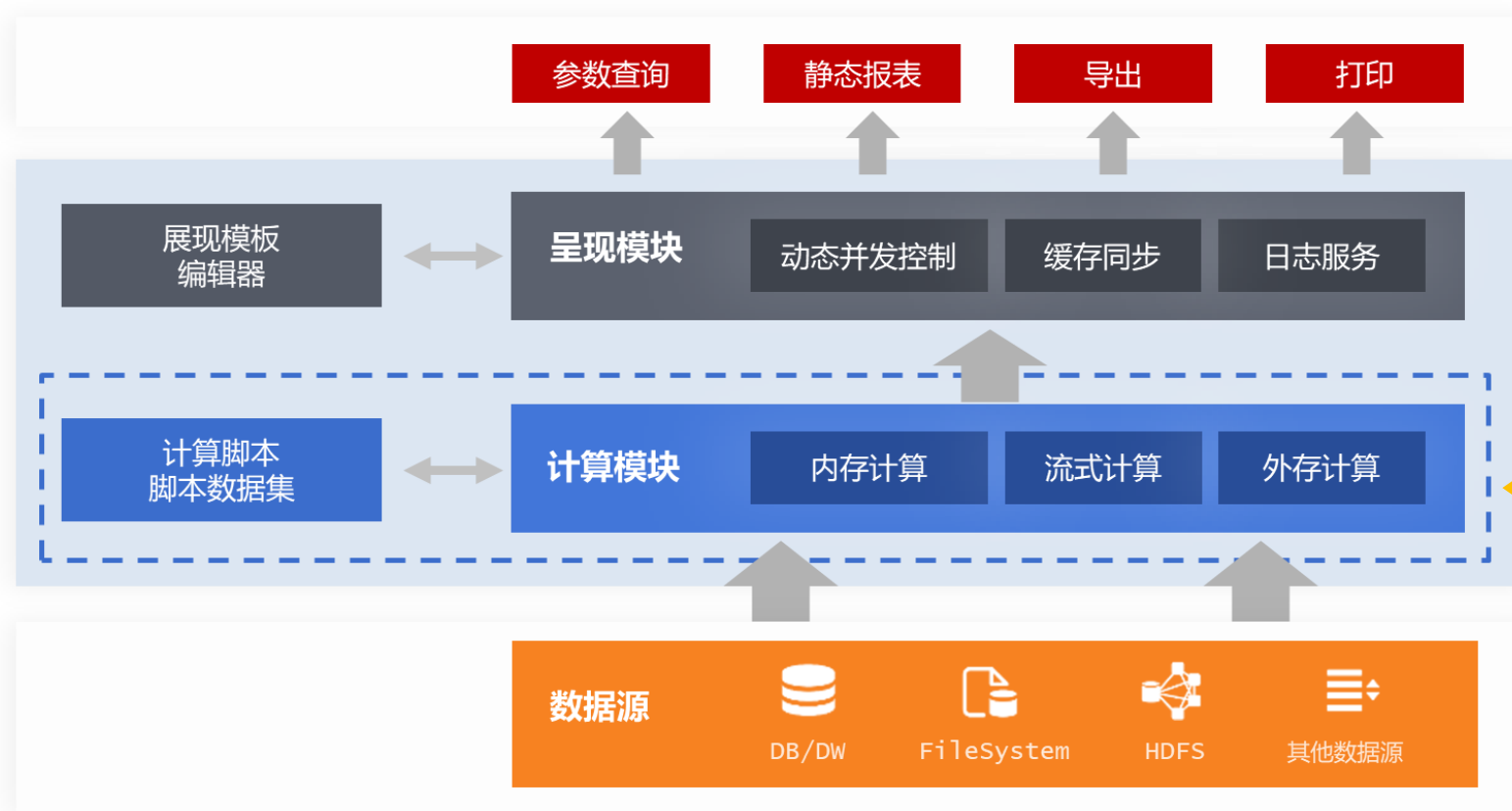
- 报表业务不稳定，新需求导致报表没完没了，这是常态，无法消除只能适应
- 以往报表工具只能解决呈现端问题，无法解决数据源准备（要靠硬编码），引起报表运维成本过高问题



——• 报表做不完，主要数据源 •——

改进后报表应用结构

在报表工具中**独立**用于数据准备的**计算模块**，将原来使用SQL/存储过程/JAVA实现的数据准备算法，全部通过计算模块完成，使得报表开发彻底工具化，简化开发，降低成本



计算模块具备的特性

- 解释执行，可插拔支持热切换
- 可理解为内置的计算中台

报表模板（计算和呈现）置于应用系统之外，报表实施过程全面工具化

独立运维报表模块

报表开发全面工具化以后，就可以着手梳理应用结构，独立报表模块，从而将报表模块从业务系统中解耦出来，独立维护

独立报表模块可以按照以下步骤着手：

1. 梳理数据源

2. 剥离报表业务

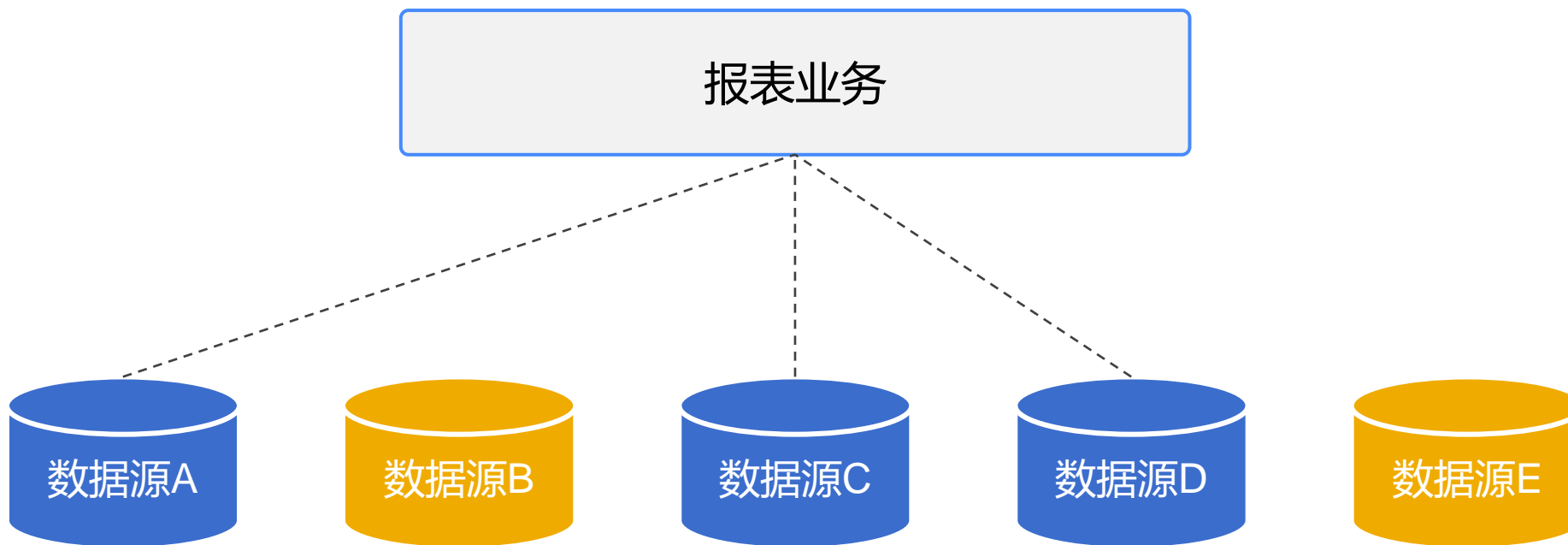
3. 独立报表模块

4. 调整应用结构



1. 梳理数据源

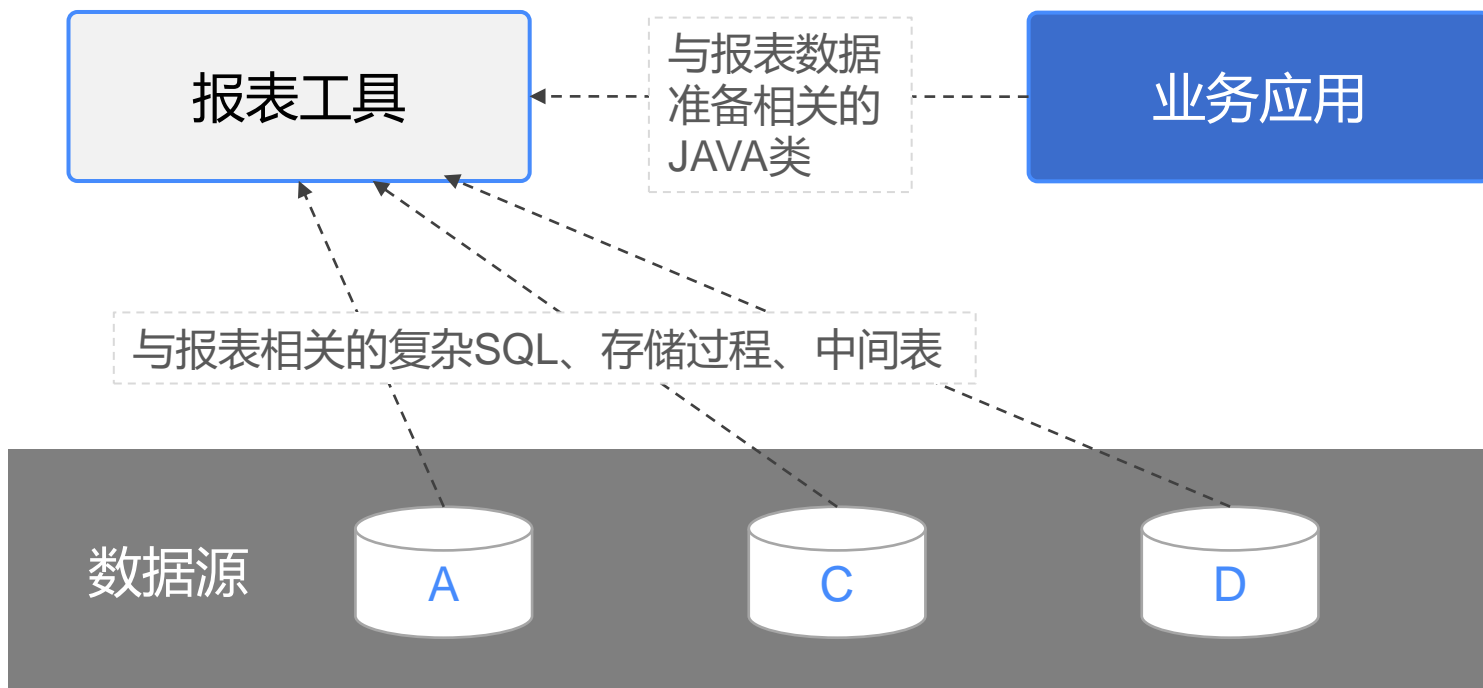
首先梳理数据源，将报表业务相关的数据源都整理出来，以后的报表开发只需要和这些数据源打交道；同时为下一步剥离报表业务做准备



数据源A C D都与报表业务相关，将其单独整理出来

2. 剥离报表业务

将数据源和业务应用中与报表相关的数据准备（复杂SQL、存储过程、中间汇总表、自定义JAVA类）全部转移到报表工具中实现；同时将这些内容从数据库和业务应用中清除，完成报表业务剥离



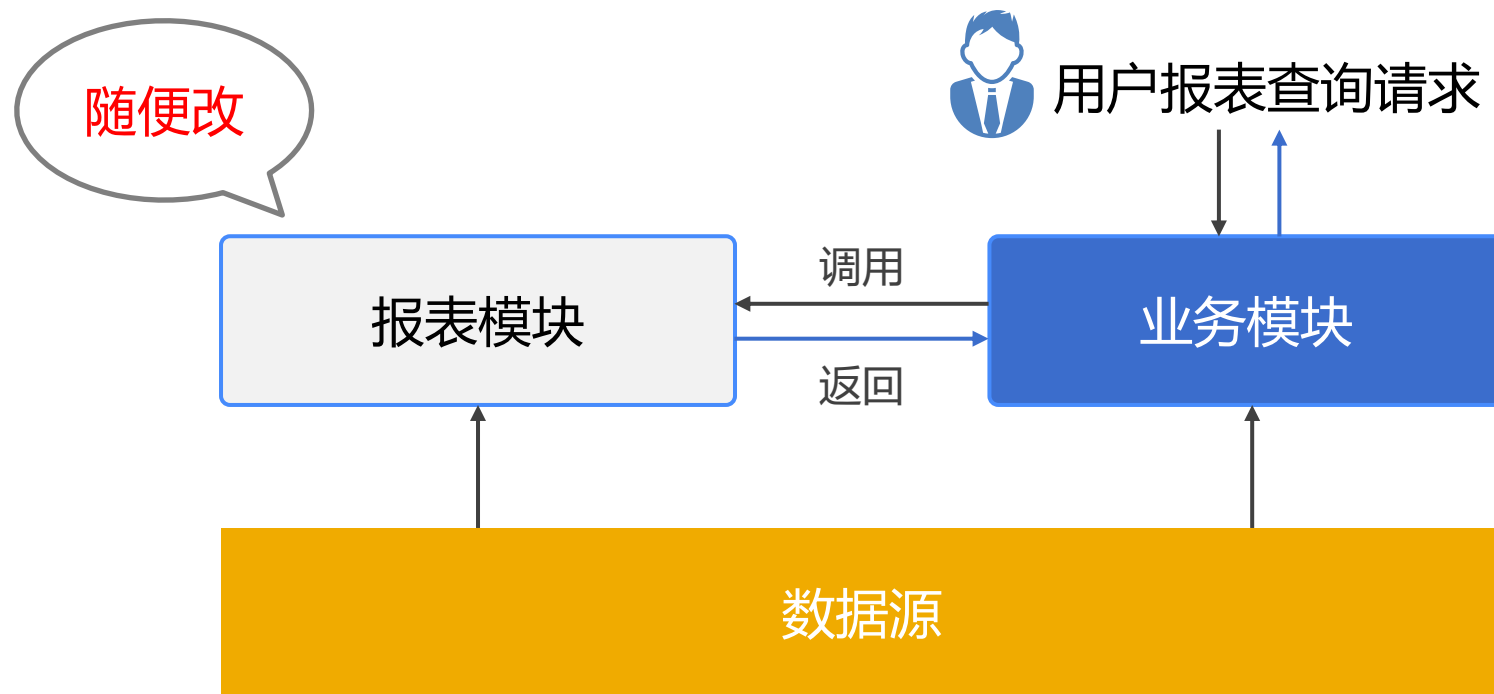
3. 独立报表模块

报表业务完全剥离后，在报表工具中实现的报表在物理上就可以独立于数据源和业务模块存储，不再与业务系统和数据源紧密耦合，完全独立报表模块



4. 调整应用结构

独立报表模块后，报表模块与业务模块共享数据存储；应用系统调用报表模块为用户输出报表结果；同时解释执行报表模块支持热切换，即改即用，无需重启应用；
报表模块应对多变的报表需求，而不会影响相对稳定的业务模块



报表模块采用集成部署，通过API或服务方式与已有交互

举例-润乾报表嵌入式脚本计算

根据订单数据，查询某年销售额占前一半的大客户及订单金额

订单表

序号	订单ID	客户ID	订购日期	货主地区	订单金额
1	10248	VINET	2012-07-04	华北	428.0
2	10249	TOMSP	2012-07-05	华东	1842.0
3	10250	HANAR	2012-07-08	华北	1523.4999898...
4	10251	VICTE	2012-07-08	华东	624.94999976...
5	10252	SUPRD	2012-07-09	东北	3559.4999980...
6	10253	HANAR	2012-07-10	华北	1428.0
7	10254	CHOPS	2012-07-11	华中	545.39999735...
8	10255	RICSU	2012-07-12	华北	2450.0
9	10256	WELLI	2012-07-15	华东	510.0
10	10257	HILAA	2012-07-16	华东	1109.0
11	10258	ERNSH	2012-07-17	华东	1603.9999940...
12	10259	CENTC	2012-07-18	华东	100.0
13	10260	OTTIK	2012-07-19	华北	1461.75
14	10261	QUEDE	2012-07-19	华东	440.0

源数据

销售额占前一半的大客户

统计年份：2013

序号	客户	订单金额 (万元)
1	QUICK	64174.26
2	SAVEA	61811.55
3	ERNSH	48320.26
4	MEREP	23332.31
5	RATTC	19383.75
6	BERGS	16344.36
7	SIMOB	16232.41
8	OTTIK	14016.27
9	HILAA	13482.74
10	FOLKO	13314.67
11	BLONP	13111.93
12	LEHMS	13076.13

报表表样

举例-润乾报表独立数据准备算法

将原来在SQL/JAVA中实现的数据准备工作转移到报表模板内实现

R 数据集设置

名称	类型	数据源
ds1	脚本数据集	

确定(O)

取消(C)

R 脚本数据集

缓存文件名

确定(O)

取消(C)

复制(C)

粘贴(P)

	A
1	=connect("demo")
2	=A1.query@x("select 订单ID,客户ID,订购日期,货主地区,订单金额 from 订单 where year(订购日期)=?,data_year)
3	=A2.group(客户ID;~.sum(订单金额):订单金额)
4	=A3.sort(-订单金额)
5	=A3.sum(订单金额)/2
6	=0
7	=A4.pselect((A6=A6+订单金额,A6>=A5))
8	=A4(to(A7))

数据准备
(脚本数据集)

	A	B	C
1	销售额占前一半的大客户		
2			=统计年份: "+@data_year
3	序号	客户	订单金额(万元)
4	=ds1.select(#0)	=ds1.客户ID	=ds1.订单金额

呈现模板

将数据准备在脚本中实现后，计算更加简单；
且整个报表不涉及存储过程/JAVA，完全独立维护，降低报表运维成本

进一步降低报表运维成本

引入报表工具、计算模块、独立报表模块后，还可以通过建立报表团队和完善沟通机制进一步降低报表运维成本

建立报表团队

报表模块独立后，建设专门的报表开发团队，不再需要高成本应用程序员或DBA参与报表开发，进一步降低报表开发运维成本

完善沟通机制

通过建设报表知识库，降低业务部门与技术部门沟通成本，减少沟通中的误解，实现报表复用，促进知识传承

建设报表团队

建设专门的报表团队应对频繁多变的报表需求，报表开发人员无需具备专业的程序员能力

需具备

- ✓ 熟悉业务逻辑
- ✓ 了解数据结构
- ✓ 熟练使用报表工具

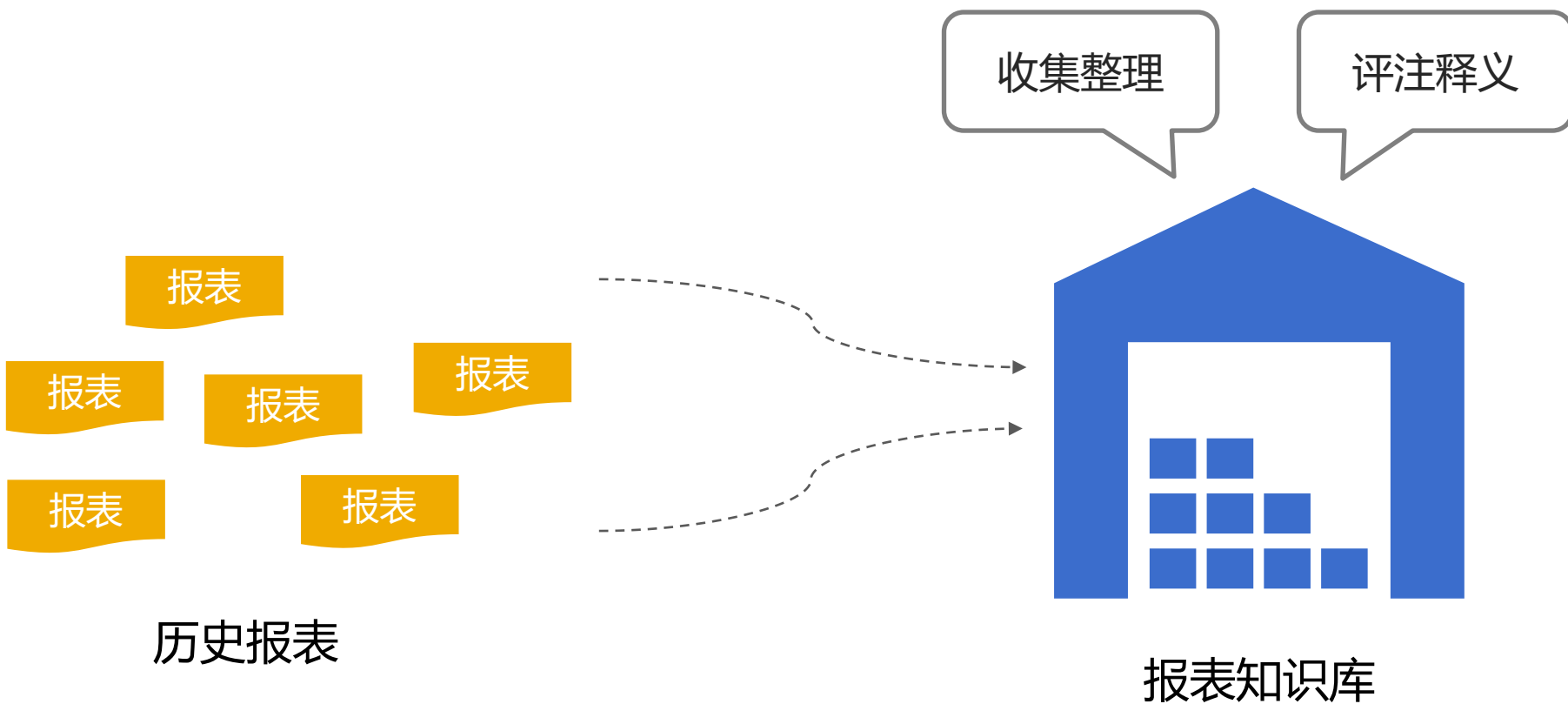
无需具备

- 复杂SQL/存储过程编码能力
- JAVA编码能力
- 应用环境部署及维护能力

报表模块独立后，报表开发人员就无需应对纷繁复杂的应用环境，更无需关系应用架构，只需根据已有架构开发报表就可以了

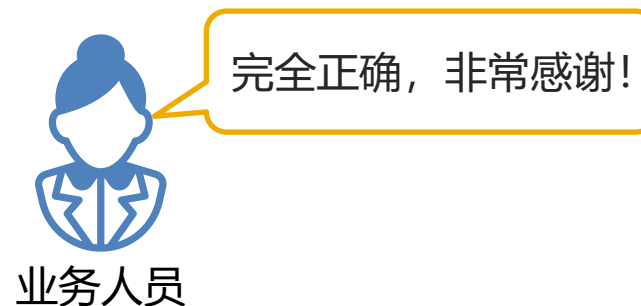
完善沟通机制

通过**建设报表知识库**，降低业务部门与技术部门沟通成本，减少沟通中的误解
实现报表复用，促进知识传承



知识库搜索

当有了新的报表需求时，可以搜索历史库中是否曾经做过类似的报表（出现过相同的业务术语）



促进知识传承

新人接手报表开发后，可以利用知识库，复用历史报表，促进知识传承



领导

小王，财务提了一个需求，你做一下这个报表



新人小王

这个应该怎么做呢？

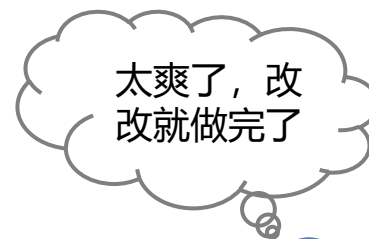


领导

到报表知识库搜一下



财务xx报表



新人小王

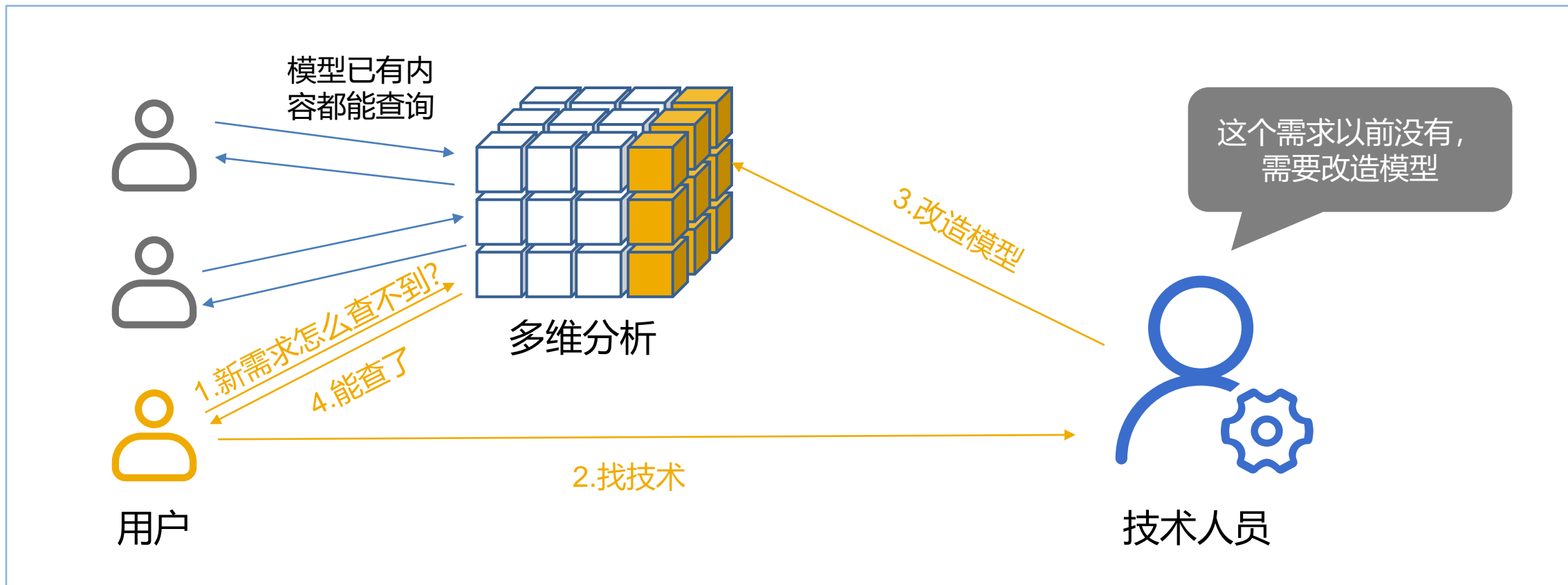
报告老板，我做完了！



按需建模

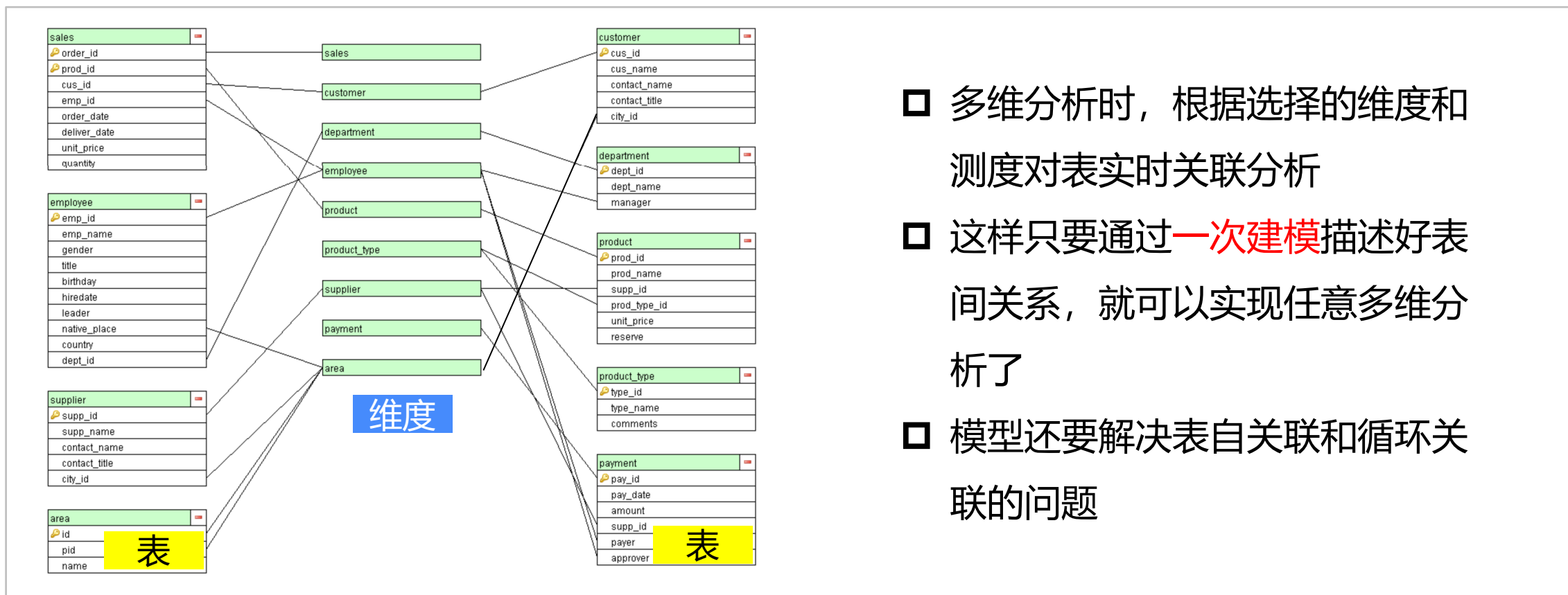
多维分析建模

通过物理/逻辑宽表实现多维分析，当需求变化时往往需要技术人员改造模型来满足查询需求，导致自助查询不灵活的同时，由于模型不断修改造成运维成本居高不下



通过关联分析减少频繁建模

- 基于星型结构的宽表会引起频繁修改模型导致运维成本高的问题
- 要解决这个问题需要基于雪花结构的多表实现，也就是要实现多表关联分析



- ❑ 多维分析时，根据选择的维度和测度对表实时关联分析
- ❑ 这样只要通过一次建模描述好表间关系，就可以实现任意多维分析了
- ❑ 模型还要解决表自关联和循环关联的问题

关联分析的意义

我们在多维分析课程的《关联分析》章节中详细讲解了实现关联分析的意义与重要性，这里再简单回顾一下，更详细的内容可参考对应课程

关联分析意义回顾

1. 一次性建模，避免频繁修改模型，降低运维成本
2. 提升数据实时性
3. 减少数据冗余
4. 增强查询的灵活性
5. 解决宽表字段多带来的可用性问题
6. 满足范式，避免数据出错

润乾报表关联分析举例

在润乾报表中通过一次定义的语义模型，在页面上通过拖拽实时关联表进行分析
下面是一个自关联维表（地区）与事实表（销售）的关联示例

pid	id	name
303	30302	石家庄
303	30303	秦皇岛
3	303	河北
401	40101	深圳
401	40102	广州
4	401	广东
402	40201	海口
4	402	海南
403	40301	厦门
403	40302	福州
4	403	福建
501	50101	西安
5	501	陕西
601	60101	成都
6	601	四川
602	60201	重庆
6	602	重庆
(Null)	1	华北
(Null)	2	华东
(Null)	3	华北
(Null)	4	华南
(Null)	5	西北
(Null)	6	西南

DB中的
地区表

根据销售记录表和白关联的地区表按地区、省、市三级汇总销售额

报表名称

地区:省:市

			销售额求和	总计
东北	辽宁	大连	55787	55787
		总计	55787	55787
	总计		55787	55787
华东	上海	上海	8662	8662
		总计	8662	8662
	山东	青岛	15224	15224
		总计	15224	15224
	江苏	南京	126007	126007
		常州	57949	57949
		总计	183956	183956
		江西	南昌	33920
	总计		33920	33920
	总计		241762	241762
	北京	北京	56387	56387

数据分析设置器

销售记录

订单

产品

客户

客户名称

联系人姓名

联系人职位

所在城市

上级地区

上级地区

地区名称

地区名称

雇员

订单日期

发货日期

单价

数量

销售额

客户

客户编号

客户名称

联系人姓名

上表头

地区

省

市

销售额求和

+

THANKS

感谢观看

