



DAMIS

中国数据资产管理峰会

CHINA DATA ASSET MANAGEMENT SUMMIT



应对业务部门的数据请求

演讲人：蒋步星



业务部门的数据请求

▼ 业务部门需求数据是常态

- 直接数据请求
- 新增报表查询

▼ 技术部门疲于奔命

- 工作量大
- 响应不及时
- 准确率不高



推给业务部门

▼ 业务部门自助获取数据

- 减少技术部门工作量
- 改善用户体验

▼ 工具引入

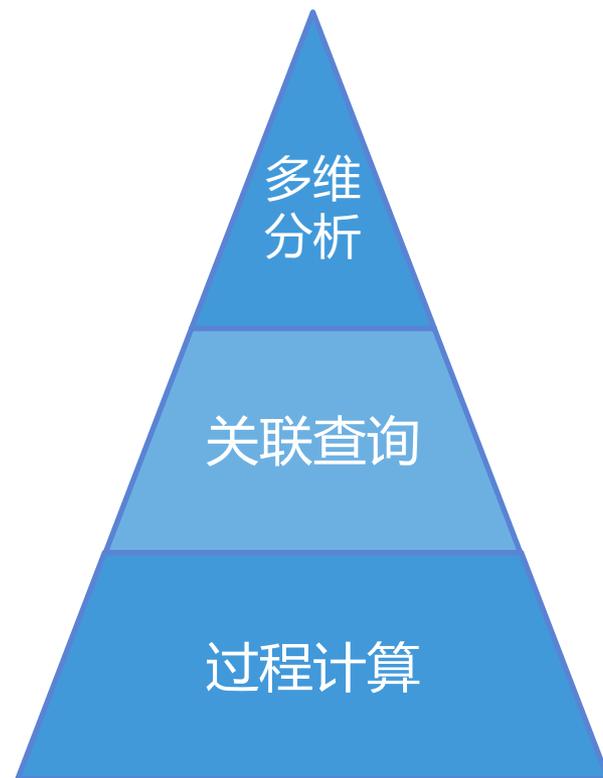
- 自助报表查询
- OLAP与敏捷BI

效果
很有限



数据需求的三个层次

- ▼ 许多工具能解决好多维分析
- ▼ 少量工具能部分处理关联查询
- ▼ 用户的需求点大部分在过程计算



多维分析的功能范围

- ▼ 单数据集上的切片，旋转、钻取
 - 分组交叉表/图、行式表
- ▼ 通用查询和自助报表可理解为多维分析的功能子集及变种

Measures	小家电		白色家电		黑色家电	
	1997年	1998年	1997年	1998年	1997年	1998年
销售量	802		1,590		769	
销售成本	44,932.93		1,055,811.04		421,665.24	
销售费	109,534.60		2,667,380.00		1,051,703.00	
利润	84,601.67		1,611,568.96		830,037.76	
利润环比	0.0%	-100.0%	0.0%	-100.0%	0.0%	-100.0%

关系	括号	字段名	运算符	内容	括号
	{	单据编号	等于	1212	}
并且	{	饰品所在表	等于	现有量表	}
并且	{	创建人	等于	1111	}

多维分析的技术本质

▼ 汇总

```
SELECT D,..., sum(M),... FROM T Group By D,.. Where D=d and ...
```

▼ 界面补充运算

占比、排名、 ...

▼ 明细

```
SELECT* FROM T Where D=d ...
```

多维分析的局限

▼ 运算模板化

▼ 单一数据集

▼ 事先准备数据集

- 有新需求需要重建模
- 由技术人员实施，失去自助意义

关联查询的必要性

▼ 大多数有意义的数据请求都会涉及多数据集关联

- 北京到上海的电话统计
- 存款10万元以上储户的学历分布

▼ 关联查询是多维分析的前序环节

- 相当于一定程度的业务人员自建模，提高自助能力

现有解决方案

▼ 预先建模

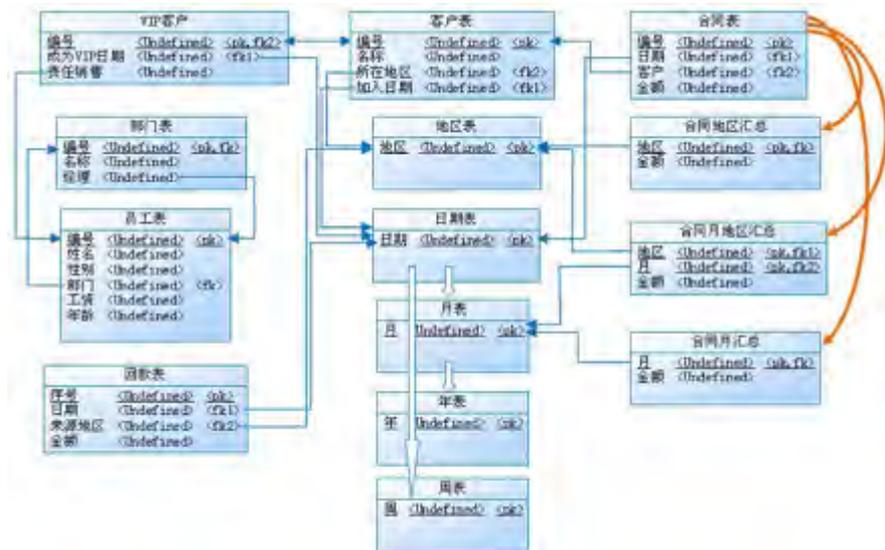
- 相当于不解决

▼ 暴露E-R图

- 数据集较多时超出业务人员的理解能力

▼ 自动同维对齐

- 易于理解的单数据集界面
- 自关联和单表多同维不能识别



关联查询的困难

▼ 关系代数对JOIN运算定义过于简单

- 笛卡尔积再过滤
- 不区分JOIN的各种情况，不能体现运算特征

▼ 检验例子

- 夫妻年龄和大于50岁的家庭
- 女经理的男员工

DQL解决关联查询

▼ 区分JOIN分别处理

1

外键引用

外键属性化

2

同维主子表对齐

同维主子表等同化

3

汇总对齐

按维度自动对齐

DQL书写例子

夫妻年龄和大于50岁的家庭

```
1 SELECT A.姓名, C.姓名 FROM 人员表 A,家庭表 B,人员表 C
2     WHERE A.性别= '男' AND B.关系类型= '配偶'
3     AND A.年龄+C.年龄>50
4     AND A.身份证=B.身份证
5     AND B.关系人身份证=C.身份证
```

```
1 SELECT 身份证.姓名,关系人身份证.姓名 FROM 家庭表
2     WHERE 身份证.性别= '男' AND 关系类型= '配偶'
3     AND 身份证.年龄+关系人身份证.年龄>50
```

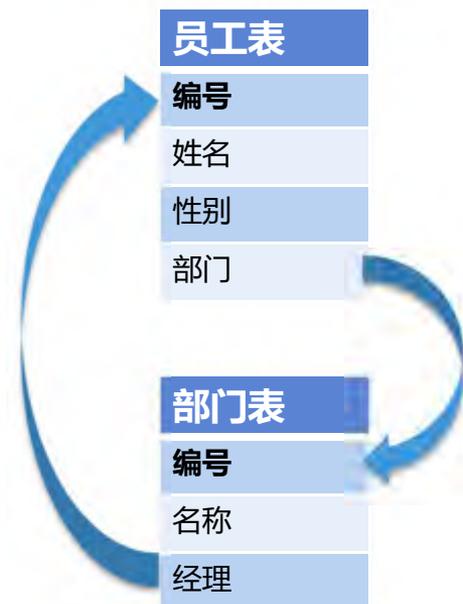


DQL书写例子

女经理的男员工

```
1 SELECT A.* FROM 员工表 A,部门表 B,员工表 C
2     WHERE A.性别= '男'
3     AND C.性别= '女'
4     AND A.部门=B.编号
5     AND B.经理=C.编号
```

```
1 SELECT * FROM 员工表
2     WHERE 性别= '男' AND 部门.经理.性别= '女'
```



DQL查询界面

- ▼ 不涉及到表的概念
- ▼ 直接拖拽数据项
- ▼ 数据项的组织呈树状，体现表之间的外键关系
- ▼ 数据项有子属性



过程计算

▼ 过程计算的普遍性

- 这个月内连涨3天的股票，第4天还继续上涨的比率有多大？
- 哪些半年不出单的客户在更换了销售人员后半年就出单了？
- 语文和数学成绩都在前10名的学生，英语成绩排名是怎样的？

▼ 最重要的自助能力是过程计算

- SQL不提倡分步计算
- 多维分析、关联查询都不能解决过程计算

自助过程计算的三层能力

▼ EXECL

- 基本表格运算

▼ 润乾计算表

- 多层表格与关联运算

业务人员的能力线：抽象思维

▼ 程序脚本

- 循环重复、可复用计算/参数

EXCEL

▼ Excel是“计算器”模式

- 直观具体不抽象
- 分步、每步动作临时决定，不能与无须事先设计

Excel是事实上应用最广泛的桌面BI工具

▼ Excel的四“多”问题

- 多层：多层表格失去运算能力
- 多行：多行记录只能手工操作
- 多表：关联运算描述困难、性能低下
- 多值：不支持集合数据



润乾计算表

▼ 沿用Excel的“计算器”模式

▼ 解决Excel的四“多”问题

- ✓ 扩展多层表格能继续计算
- ✓ 结构化数据模型支持多行记录
- ✓ 扩展关系代数的表间关联运算
- ✓ 集合数据类型



连续上涨股票

▼ 从表中找出连续五交易日上涨的股票

0	1	A	B	C
1-	1	代码	日期	收盘价
1	2	1	2012-03-01	52.8
1	3	2	2012-03-01	43.57
1	4	3	2012-03-01	38.29
1	5	4	2012-03-01	43.55
1	6	5	2012-03-01	66.33
1	7	6	2012-03-01	94.39
1	8	7	2012-03-01	17.33
1	9	8	2012-03-01	107.88
1	10	9	2012-03-01	52.73
1	11	10	2012-03-01	37.86
1	12	11	2012-03-01	93.04
1	13	12	2012-03-01	65.19
1	14	13	2012-03-01	32.25
1	15	14	2012-03-01	37.29
1	16	15	2012-03-01	64.59
1	17	16	2012-03-01	19.68
1	18	17	2012-03-01	100.49
1	19	18	2012-03-01	70.73
1	20	19	2012-03-01	95.42
1	21	20	2012-03-01	101.1
1	22	21	2012-03-01	101.03
1	23	22	2012-03-01	67.21
1	24	23	2012-03-01	16.28
1	25	24	2012-03-01	26.9
1	26	25	2012-03-01	87.87

【思路】

1. 按股票代码分组
2. 在组内按日期排序
3. 添加计算列在组内计算收盘价比上期
4. 添加计算列计算累计上涨天数
5. 在组头上计算连涨天数最大值
6. 按连涨天数从大到排序，收缩显示

连续上涨股票 (续)

1/2) 按代码分组后按日期排序

0	1	2	A	B	C
1-	1		代码	日期	收盘价
1-	2	1			
	3	1		2012-03-01	52.8
	4	1		2012-03-02	78.24
	5	1		2012-03-05	59.58
	6	1		2012-03-06	59.62
	7	1		2012-03-07	61.69
	8	1		2012-03-08	97.37
	9	1		2012-03-09	44.05
	10	1		2012-03-12	56.25
	11	1		2012-03-13	90.26
	12	1		2012-03-14	80.07
	13	1		2012-03-15	24.38
	14	1		2012-03-16	48.85
	15	1		2012-03-19	16.31
	16	1		2012-03-20	68.17
	17	1		2012-03-21	94.46
	18	1		2012-03-22	47.64
	19	1		2012-03-23	35.2
	20	1		2012-03-26	49.89
	21	1		2012-03-27	23.15
	22	1		2012-03-28	16.82
	23	1		2012-03-29	101.36
	24	1		2012-03-30	104.77
1-	25	2			
	26	2		2012-03-01	43.57
	27	2		2012-03-02	68.63

3) 增加D列计算涨幅，公式填入D4以方便引用C3格

0	1	2	A	B	C	D
1-	1		代码	日期	收盘价	
1-	2	1				
	3	1		2012-03-01	52.8	52.8
	4	1		2012-03-02	78.24	=C4-C3/C3
	5	1		2012-03-05	59.58	-18.659999999999997
	6	1		2012-03-06	59.62	0.03999999999999915
	7	1		2012-03-07	61.69	2.0700000000000003
	8	1		2012-03-08	97.37	35.680000000000001
	9	1		2012-03-09	44.05	-53.320000000000001
	10	1		2012-03-12	56.25	12.200000000000003
	11	1		2012-03-13	90.26	34.0
	12	1		2012-03-14	80.07	-10.180000000000007
	13	1		2012-03-15	24.38	-55.69
	14	1		2012-03-16	48.85	24.470000000000002
	15	1		2012-03-19	16.31	-32.540000000000006
	16	1		2012-03-20	68.17	51.86
	17	1		2012-03-21	94.46	26.289999999999992
	18	1		2012-03-22	47.64	-46.819999999999999
	19	1		2012-03-23	35.2	-12.439999999999998
	20	1		2012-03-26	49.89	14.689999999999998
	21	1		2012-03-27	23.15	-26.740000000000002
	22	1		2012-03-28	16.82	-6.329999999999998
	23	1		2012-03-29	101.36	84.539999999999999
	24	1		2012-03-30	104.77	3.4099999999999966
1-	25	2				
	26	2		2012-03-01	43.57	43.57
	27	2		2012-03-02	68.63	25.059999999999995

连续上涨股票（续）

0	1	2	A	B	C	D	E
1-			代码	日期	收盘价		
	3	1					
	3	1		2012-03-01	52.8	52.8	1
	4	1		2012-03-02	78.24	25.43999999	=IF(D4>0,E3[E2]+1,1)
	5	1		2012-03-05	59.58	-18.65999999	1
	6	1		2012-03-06	59.62	0.0399999999	2
	7	1		2012-03-07	61.69	2.070000000	3
	8	1		2012-03-08	97.37	35.68000000	4
	9	1		2012-03-09	44.05	-53.32000000	1
	10	1		2012-03-12	56.25	12.20000000	2
	11	1		2012-03-13	90.25	34.0	3
	12	1		2012-03-14	80.07	-10.18000000	1
	13	1		2012-03-15	24.38	-55.69	1
	14	1		2012-03-16	48.85	24.47000000	2
	15	1		2012-03-19	16.31	-32.54000000	1
	16	1		2012-03-20	68.17	51.86	2
	17	1		2012-03-21	94.46	26.28999999	3
	18	1		2012-03-22	47.64	-46.81999999	1
	19	1		2012-03-23	35.2	-12.43999999	1
	20	1		2012-03-26	49.89	14.68999999	2
	21	1		2012-03-27	23.15	-26.74000000	1
	22	1		2012-03-28	16.82	-6.329999999	1
	23	1		2012-03-29	101.36	84.53999999	2
	24	1		2012-03-30	104.77	3.409999999	3
	25	2					
	26	2		2012-03-01	43.57	43.57	1
	27	2		2012-03-02	68.63	25.05999999	2

4) 添加E列计算连涨天数，填入E4格以方便引用E3

5) 在E2计算每支股票的最大连涨天数

0	1	2	A	B	C	D	E
1-			代码	日期	收盘价		
	2	1					=E3.max()
	3	1		2012-03-01	52.8	52.8	1
	4	1		2012-03-02	78.24	25.43999999	2
	5	1		2012-03-05	59.58	-18.65999999	1
	6	1		2012-03-06	59.62	0.0399999999	2
	7	1		2012-03-07	61.69	2.070000000	3
	8	1		2012-03-08	97.37	35.68000000	4
	9	1		2012-03-09	44.05	-53.32000000	1
	10	1		2012-03-12	56.25	12.20000000	2
	11	1		2012-03-13	90.25	34.0	3
	12	1		2012-03-14	80.07	-10.18000000	1
	13	1		2012-03-15	24.38	-55.69	1
	14	1		2012-03-16	48.85	24.47000000	2
	15	1		2012-03-19	16.31	-32.54000000	1
	16	1		2012-03-20	68.17	51.86	2
	17	1		2012-03-21	94.46	26.28999999	3
	18	1		2012-03-22	47.64	-46.81999999	1
	19	1		2012-03-23	35.2	-12.43999999	1
	20	1		2012-03-26	49.89	14.68999999	2
	21	1		2012-03-27	23.15	-26.74000000	1
	22	1		2012-03-28	16.82	-6.329999999	1
	23	1		2012-03-29	101.36	84.53999999	2
	24	1		2012-03-30	104.77	3.409999999	3
	25	2					5
	26	2		2012-03-01	43.57	43.57	1
	27	2		2012-03-02	68.63	25.05999999	2

连续上涨股票 (续)

6) 对E2
从大到小
排序

0	1	2	A	B	C	D	E
1-		1	代码	日期	收盘价		
		2	7				8
		3	7	2012-03-01	17.33	17.33	5
		4	7	2012-03-02	61.26	43.93	6
		5	7	2012-03-05	84.26	23.0000	7
		6	7	2012-03-06	93.73	9.46999	8
		7	7	2012-03-07	54.6	-39.13	1
		8	7	2012-03-08	91.88	37.2799	2
		9	7	2012-03-09	109.91	18.03	3
		10	7	2012-03-12	30.37	-79.539	1
		11	7	2012-03-13	25.49	-4.8800	1
		12	7	2012-03-14	51.78	26.2900	2
		13	7	2012-03-15	70.52	18.7399	3
		14	7	2012-03-16	75.84	5.32000	4
		15	7	2012-03-19	39.0	-36.84	1
		16	7	2012-03-20	17.82	-21.18	1
		17	7	2012-03-21	74.48	56.6600	2
		18	7	2012-03-22	91.55	17.0699	3
		19	7	2012-03-23	76.41	-15.14	1
		20	7	2012-03-26	97.36	20.9500	2
		21	7	2012-03-27	101.05	3.68999	3
		22	7	2012-03-28	20.41	-80.64	1
		23	7	2012-03-29	94.79	74.3800	2
		24	7	2012-03-30	89.74	-5.0500	1
		25	14				7
		26	14	2012-03-01	37.29	37.29	4
		27	14	2012-03-02	37.96	0.67000	5

收缩显示，
观察E列 ≥ 5
的代码即是
结果

0	1	2	A	B	C	D	E
1-		1	代码	日期	收盘价		
		2	7				8
		25	14				7
		48	44				7
		71	8				6
		94	22				6
		117	47				6
		140	48				6
		163	49				6
		186	79				6
		209	2				5
		232	10				5
		255	12				5
		278	18				5
		301	21				5
		324	28				5
		347	30				5
		370	31				5
		393	39				5
		416	42				5
		439	51				5
		462	56				5
		485	58				5
		508	62				5
		531	64				5
		554	73				5
		577	76				5

各科成绩前10名

▼ 从成绩表中找出各科成绩均在前10名的学生

0	1	A	B	C
1-	1	姓名	科目	成绩
1	2	赵一伤	语文	89
1	3	钱二败	语文	77
1	4	孙三败	语文	31
1	5	李四崖	语文	37
1	6	郭峰	语文	92
1	7	杨康	语文	47
1	8	周五输	语文	76
1	9	吴六破	语文	35
1	10	郑七灭	语文	54
1	11	王八衰	语文	88
1	12	杨过	语文	57
1	13	令狐冲	语文	94
1	14	王重阳	语文	93
1	15	黄药师	语文	49
1	16	洪七公	语文	81
1	17	欧阳峰	语文	99
1	18	周伯通	语文	57
1	19	东方不败	语文	72
1	20	任我行	语文	56
1	21	萧峰	语文	41
1	22	虚竹	语文	62
1	23	段誉	语文	35
1	24	慕容复	语文	98

【思路】

1. 按科目分组
2. 在科目内排序
3. 计算科目内名次
4. 删除名次>10的行
5. 拆除分组
6. 按学生分组
7. 统计组下科目数量，该数与科目总数相同者即是结果

各科成绩前10名（续）

按科目
分组后
按日成绩
排序

0	1	2	A	B	C
1-		1	姓名	科目	成绩
1-		2		语文	
		3	慕容复	语文	98
		4	令狐冲	语文	94
		5	王重阳	语文	93
		6	郭靖	语文	92
		7	赵一伤	语文	89
		8	欧阳锋	语文	89
		9	王八衰	语文	88
		10	洪七公	语文	81
		11	钱二败	语文	77
		12	周五输	语文	76
		13	东方不败	语文	72
		14	虚竹	语文	62
		15	鸠摩智	语文	62
		16	杨过	语文	57
		17	周伯通	语文	57
		18	任我行	语文	56
		19	郑七灭	语文	54
		20	黄药师	语文	49
		21	杨康	语文	47
		22	萧峰	语文	41
		23	李四摧	语文	37
		24	吴六破	语文	35
		25	段誉	语文	35
		26	孙三毁	语文	31

增加D列
计算各科
前10名

0	1	2	A	B	C	D
1-		1	姓名	科目	成绩	
1-		2		语文		
		3	慕容复	语文	98	=#<=10
		4	令狐冲	语文	94	true
		5	王重阳	语文	93	true
		6	郭靖	语文	92	true
		7	赵一伤	语文	89	true
		8	欧阳锋	语文	89	true
		9	王八衰	语文	88	true
		10	洪七公	语文	81	true
		11	钱二败	语文	77	true
		12	周五输	语文	76	true
		13	东方不败	语文	72	false
		14	虚竹	语文	62	false
		15	鸠摩智	语文	62	false
		16	杨过	语文	57	false
		17	周伯通	语文	57	false
		18	任我行	语文	56	false
		19	郑七灭	语文	54	false
		20	黄药师	语文	49	false
		21	杨康	语文	47	false
		22	萧峰	语文	41	false
		23	李四摧	语文	37	false
		24	吴六破	语文	35	false
		25	段誉	语文	35	false
		26	孙三毁	语文	31	false
0-		27		数学		

各科成绩前10名（续）

删除各科的非前10名

0	1	2	A	B	C	D
1-		1	姓名	科目	成绩	
1-		2		语文		
		3	慕容复	语文	98	true
		4	令狐冲	语文	94	true
		5	王重阳	语文	93	true
		6	郭靖	语文	92	true
		7	赵一伤	语文	89	true
		8	欧阳锋	语文	89	true
		9	王八衰	语文	88	true
		10	洪七公	语文	91	true
		11	钱二败	语文	77	true
		12	周五输	语文	76	true
1-		13		数学	10	
		14	黄药师	数学	96	true
		15	郑七灭	数学	91	true
		16	钱二败	数学	90	true
		17	令狐冲	数学	88	true
		18	王八衰	数学	82	true
		19	洪七公	数学	82	true
		20	周五输	数学	81	true
		21	东方不败	数学	78	true
		22	王重阳	数学	77	true
		23	赵一伤	数学	76	true
1-		24		英语	10	
		25	鸠摩智	英语	99	true
		26	洪七公	英语	94	true
		27	段誉	英语	94	true

拆除分组

0	1	A	B	C	D	
1-		1	姓名	科目	成绩	
			慕容复	语文	98	true
		3	令狐冲	语文	94	true
		4	王重阳	语文	93	true
		5	郭靖	语文	92	true
		6	赵一伤	语文	89	true
		7	欧阳锋	语文	89	true
		8	王八衰	语文	88	true
		9	洪七公	语文	81	true
		10	钱二败	语文	77	true
		11	周五输	语文	76	true
		12	黄药师	数学	96	true
		13	郑七灭	数学	91	true
		14	钱二败	数学	90	true
		15	令狐冲	数学	88	true
		16	王八衰	数学	82	true
		17	洪七公	数学	82	true
		18	周五输	数学	81	true
		19	东方不败	数学	78	true
		20	王重阳	数学	77	true
		21	赵一伤	数学	76	true
		22	鸠摩智	英语	99	true
		23	洪七公	英语	94	true
		24	段誉	英语	94	true
		25	周五输	英语	89	true
		26	虚竹	英语	89	true
		27	王重阳	英语	88	true

各科成绩前10名（续）

再按学生
分组并计
算科目数

0	1	2	A	B	C	D
1-		1	姓名	科目	成绩	
1-		2	东方不败	=E(B3).count()		
1-	1	3	东方不败	数学	78	true
1-	1	4	东方不败	物理	75	true
1-	1	5	东方不败	化学	95	true
1-		6	令狐冲	3		
1-	1	7	令狐冲	语文	94	true
1-	1	8	令狐冲	数学	88	true
1-	1	9	令狐冲	历史	84	true
1-		10	吴六破	1		
1-	1	11	吴六破	物理	94	true
1-		12	周五输	5		
1-	1	13	周五输	语文	76	true
1-	1	14	周五输	数学	81	true
1-	1	15	周五输	英语	89	true
1-	1	16	周五输	化学	79	true
1-	1	17	周五输	历史	97	true
1-		18	周伯通	1		
1-	1	19	周伯通	物理	94	true
1-		20	孙三殿	1		
1-	1	21	孙三殿	历史	83	true
1-		22	慕容复	4		
1-	1	23	慕容复	语文	98	true
1-	1	24	慕容复	英语	82	true
1-	1	25	慕容复	物理	95	true
1-	1	26	慕容复	化学	95	true

排序后收缩
显示，得到
每名学生前
10名科目数

0	1	2	A	B	C
1-		1	姓名	科目	成绩
1-		2	周五输	5	
1-		8	洪七公	5	
1-		14	赵一伤	5	
1-		20	钱二败	5	
1-		26	慕容复	4	
1-		31	东方不败	3	
1-		35	令狐冲	3	
1-		39	李四摧	3	
1-		43	欧阳锋	3	
1-		47	王重阳	3	
1-		51	郭靖	3	
1-		55	黄药师	3	
1-		59	段誉	2	
1-		62	王八衰	2	
1-		65	虚竹	2	
1-		68	郑七灭	2	
1-		71	鸠摩智	2	
1-		74	吴六破	1	
1-		76	周伯通	1	
1-		78	孙三殿	1	
1-		80	杨康	1	
1-		82	萧峰	1	

计算表的理论模型类比

▼ 从整数到有理数

- 整数范围内可以做自由连续加、减、乘运算，但不能随意做除法，结果可能超出整数范围
- 将数域扩大到有理数后，除法运算封闭，能自由连续地进行四则运算
- 在有理数域内要重新定义加、减、乘、除运算规则

▼ 从单层数据模型到多层数据模型

- 单层多行数据范围内可以自由进行过滤、排序、计算列等运算，但分组后的数据不能再用单层模型表示，单层模型对分组运算（在关系代数中恰好被称为除法）不封闭
- 扩展到多层数据模型使之对分组运算封闭，从而使各项运算得以连续进行，在新体系下再重新定义原有各项运算

广义OLAP概念

▼ 狭义化的OLAP特指多维分析

▼ 广义的OLAP概念

- 猜测、证实/证伪的循环过程
- 不需要也不可能事先建模，临时决定动作
- 面向业务人员，不需要技术人员参与

关联查询+过程计算：回归OLAP本意

自助之外：技术人员协助

▼ 技术人员协助的必要性

- 抽象算法
- 数据量大，无法导出
- 数据源杂，不一定是数据库，难以做关联查询
- 数据集太多，都导出太繁琐

技术协助的努力方向

▼ 让业务人员更高效便捷地获得技术资源

- 流畅获取需求和提交答案
- 准确理解需求
- 快速实现需求
- 低成本建设技术队伍

代码管理

▼ 历史代码库

- 直接搜索后执行
- 协助技术人员准确理解业务需求

▼ 脚本工具选择

- 可管理
- 易于集成
- 开发效率高
- 环境配置简单

脚本选择

	可管理性	开发效率	环境配置
▼ SQL	好	低	易
▼ 存储过程	差	低	难
▼ Python/perl	好	中	中
▼ Java	差	低	难
▼ 集算器	好	高	易

集算器

编辑调试功能，执行、单步执行、设置断点、执行到光标等

The screenshot displays the Jisuanqi 3.1 software interface. The main window is titled "集算器3.1 - 内部测试版 [D:\Program Files\raqsoft\esProc\demo\zh\Structural\db08.dfx]". The menu bar includes "文件(F)", "编辑(E)", "程序(P)", "工具(T)", "窗口(W)", and "帮助(H)". The toolbar contains icons for file operations and execution. The code editor shows a script with the following content:

```
A4 = 1 =A2.group(Employee;~:g,0:AccWorkDays_NYC)
1 =file("..\demo\zh\bf\Employees.bt").import@t0
2 =file("..\demo\zh\bf\EmpTransfer.bt").import@t0.sort(Date)
3 >A1.primary(ID), A2.switch(Employee,A1)
4 =A2.group(Employee;~:g,0:AccWorkDays_NYC)
5 for A4
6   =A5.g
7   =B5.new(~:Date:StartDate, if(#==B5.len(),now(), B5(#+1).Date):EndDate, Area:WorkArea)
8   =B6.select(WorkArea=="New York")
9   >A5.AccWorkDays_NYC=B7.sum(interval
10  =A4.select(AccWorkDays_NYC>=5*365)
11  =A9.new(g.Employee.Name:Employee, int(AccWorkDays_NYC/365):AccWorkYears_NYC)
```

The results panel on the right shows a table with columns "Employee", "g", and "AccWorkDay...". The data is as follows:

Employee	g	AccWorkDay...
16	[16,16,16, ...]	1485
5	[5,5,5, ...]	2594
11	[11,11]	
3	[3,3,3]	
9	[9]	
21	[21,21,21, ...]	51
6	[6]	5203
22	[22,22]	
17	[17,17]	
8	[8,8,8]	
1	[1,1]	

At the bottom of the interface, there is a table for variables:

序号	名称	值
----	----	---

结果面板，实时显示格子计算结果

网格式编程，语法完善，编写简单，比R、Python更简单

集算器与SQL对比

▼ 销售额占前一半的大客户

SQL

```
1 SELECT CUSTOMER, AMOUNT, SUM_AMOUNT
2 FROM (SELECT CUSTOMER, AMOUNT,
3       SUM(AMOUNT) OVER(ORDER BY AMOUNT DESC) SUM_AMOUNT
4       FROM (SELECT CUSTOMER, SUM(AMOUNT) AMOUNT
5             FROM ORDERS GROUP BY CUSTOMER))
6       WHERE 2 * SUM_AMOUNT < (SELECT
7       SUM(AMOUNT) TOTAL FROM ORDERS)
```

集算脚本

	A
1	=orders.groups(CUSTOMER;sum(AMOUNT):AMOUNT).sort(AMOUNT:-1)
2	=A1.derive(SUM_AMOUNT).run(SUM_AMOUNT=A1.AMOUNT+SUM_AMOUNT[-1])
3	=A2.select(SUM_AMOUNT<=A1.sum(AMOUNT)/2)

与SQL这类本身就支持结构化计算的语言相比，集算器的语法完善了对分步计算、集合化、有序计算和对象引用等几方面的支持；对于日期和字符串等运算，集算器也比大部分SQL提供了更丰富的方法。

报表问题

▼ 报表是获取数据的重要手段

- 许多业务用户只会看报表

▼ 报表业务的不稳定是常态

- 要建立长期应对机制

▼ 报表的困难点在于数据源

- 占用大量开发时间



报表模块/平台的目标

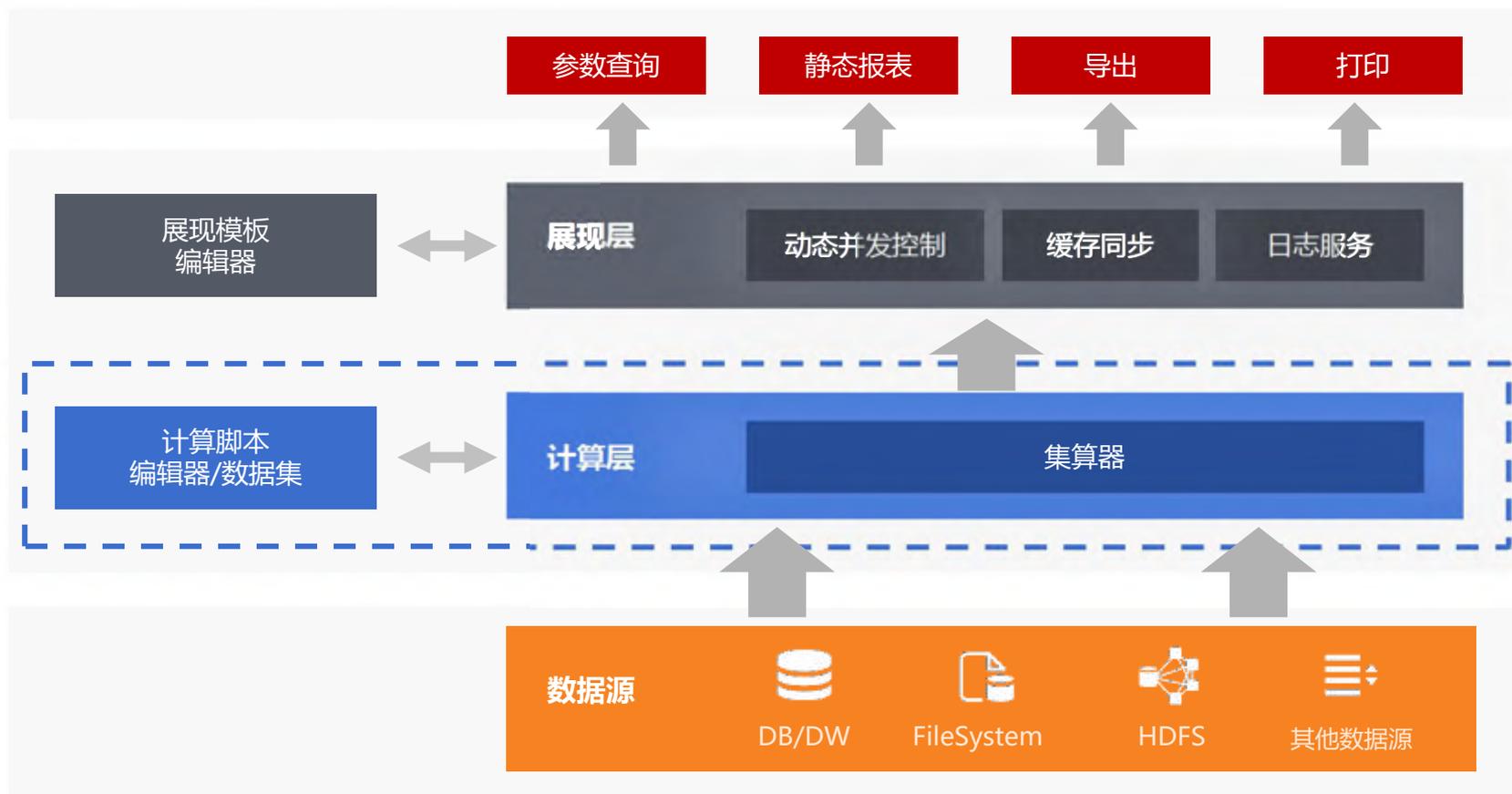
▼ 应对报表的业务不稳定性

- 与应用程序脱耦
- 提高开发效率
- 降低人员要求



报表应用结构

引入数据计算层



集算器实现数据计算层



高效率

更丰富的语法和类库



低耦合

报表模板与算法一体化



热切换

解释执行无须重启应用



简单开发

无须环境配置与应用层代码引用

扫码关注 《数据蒋堂》



- ✓ 技术干货分享
- ✓ 专注数据，每周一期
- ✓ 微信/直播交流



DAMS

中国数据资产管理峰会

CHINA DATA ASSET MANAGEMENT SUMMIT

THANK YOU !

