

SPL Base

集算器教案

连接



目录

CONTENTS

**01**

外键对象化

1. 转成对应记录
2. 只保留无匹配的记录

02

常规连接

1. 散列法连接
2. 有序归并法连接
3. 外连接
4. 叉积
5. 直接连接文件

03

按位置连接

1. 序号直接定位
2. 按位置连接
3. 按位置外连接

04

关联连接

1. 关联连接与关联左连接
2. 退化成叉积
3. 复杂用法

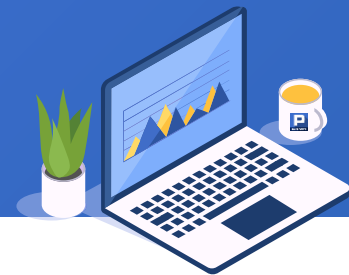
CONTENTS

1. 转成对应记录
2. 只保留无匹配的记录

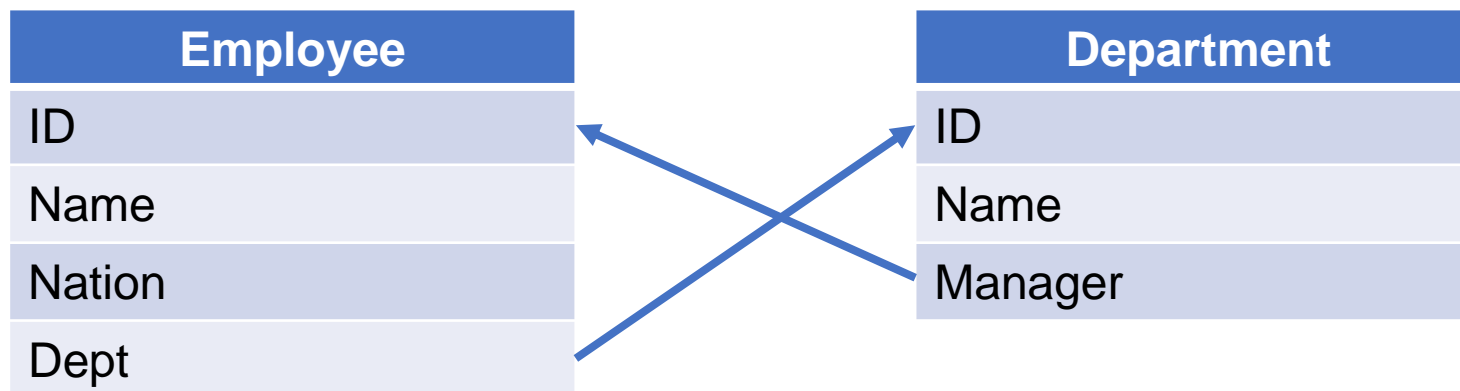


外键对象化

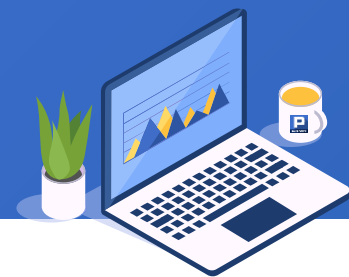
✦ 1. 转成对应记录



哪些美国籍员工有一个中国籍经理？ 员工表和部门表如下：



◆ 1. 转成对应记录

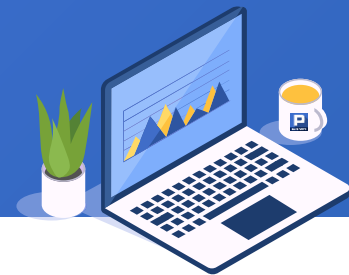


SPL如下，使用了A.switch()函数将外键字段转成外键表中对应记录：

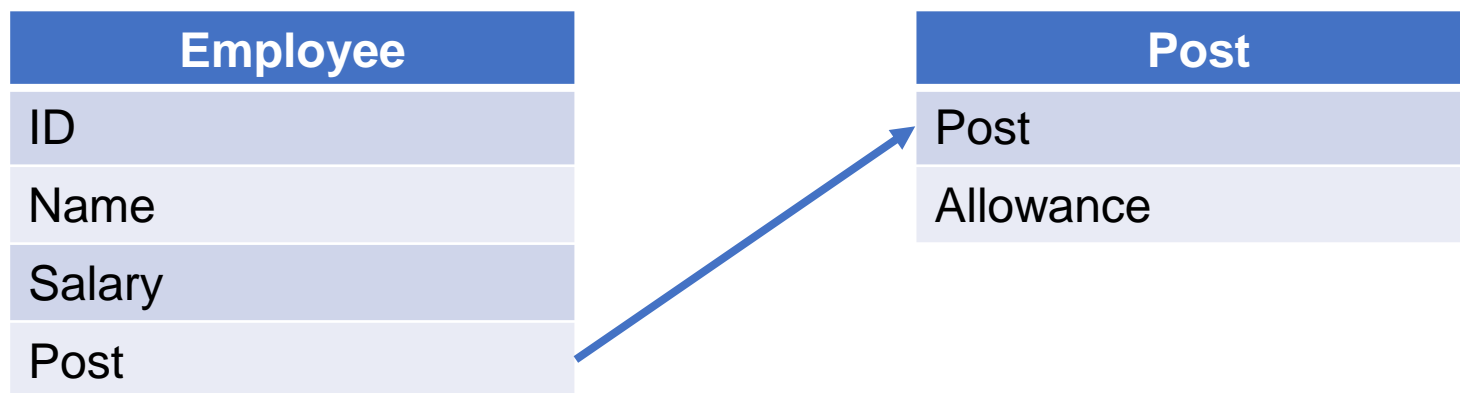
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Employee")	/查询员工表
3	=A1.query("select * from Department")	/查询部门表
4	=A3.switch(Manager, A2:ID)	/使用switch函数将部门表的Manager字段转成对应的员工记录
5	=A2.switch(Dept, A4:ID)	/使用switch函数将员工表的Dept字段转成对应的部门记录
6	=A5.select(Nation=="American" && Dept.Manager.Nation=="Chinese")	/选出国籍是美国，且经理的国籍是中国的员工

A6	ID	Name	Nation	Dept
	11	Simon	American	2
	103	Rudy	American	2

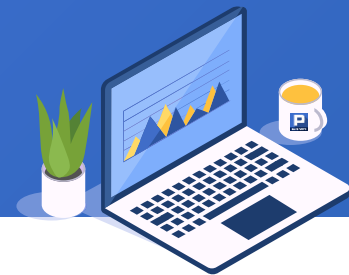
◆ 1. 转成对应记录



除了工资收入，部分员工有岗位津贴，求员工的总收入。员工表和岗位津贴表如下：



◆ 1. 转成对应记录

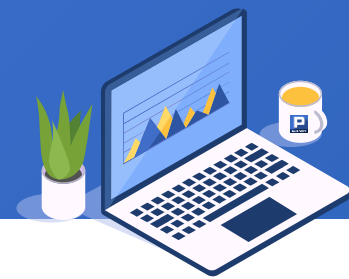


SPL如下，使用了A.switch()函数将外键字段转成外键表中对应记录，不匹配的字
段值将被设置为null:

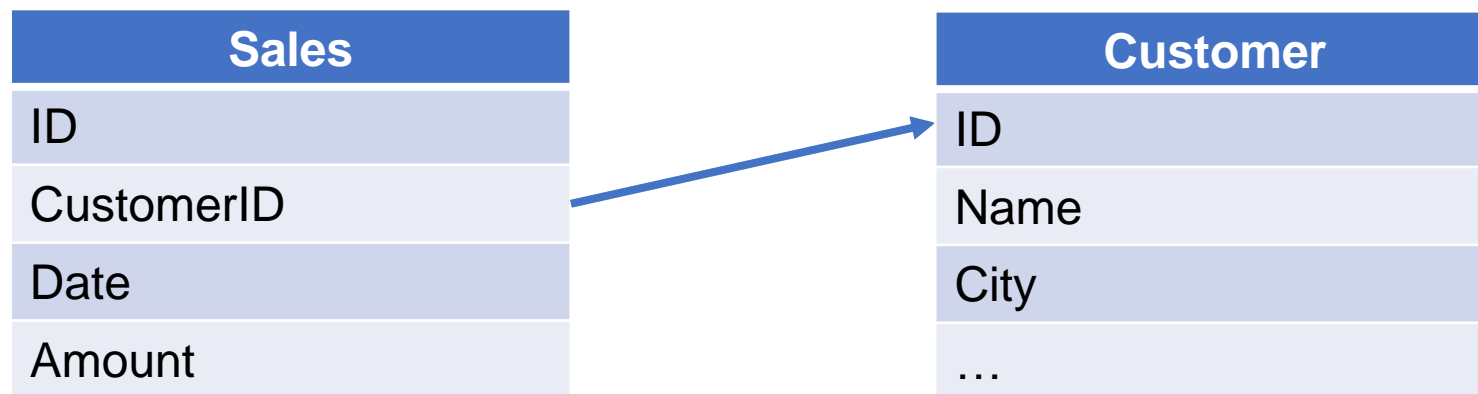
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Employee")	/查询员工表
3	=A1.query("select * from PostAllowance")	/查询岗位津贴表
4	=A2.switch(Post, A3:Post)	/使用switch函数将员工表的岗位字段转成对应的记录，不匹配的岗位会被设置为null
5	=A4.new(ID,Name,Salary+Post.Allowance:Salary)	/创建序表，统计员工收入

A5	ID	Name	Salary
	1	Rebecca	8000
	2	Ashley	12000

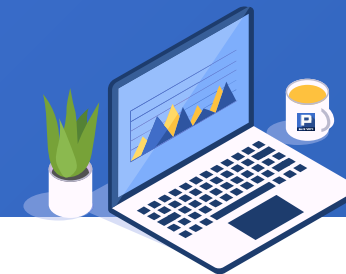
✦ 1. 转成对应记录



查询2014年北京每个客户的销售额，降序排列。销售表和客户表如下：



✦ 1. 转成对应记录

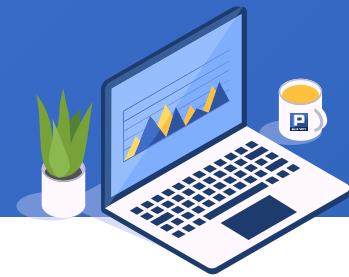


使用了A.switch()函数的@i选项, 找不到外键映射时删除该记录。SPL如下:

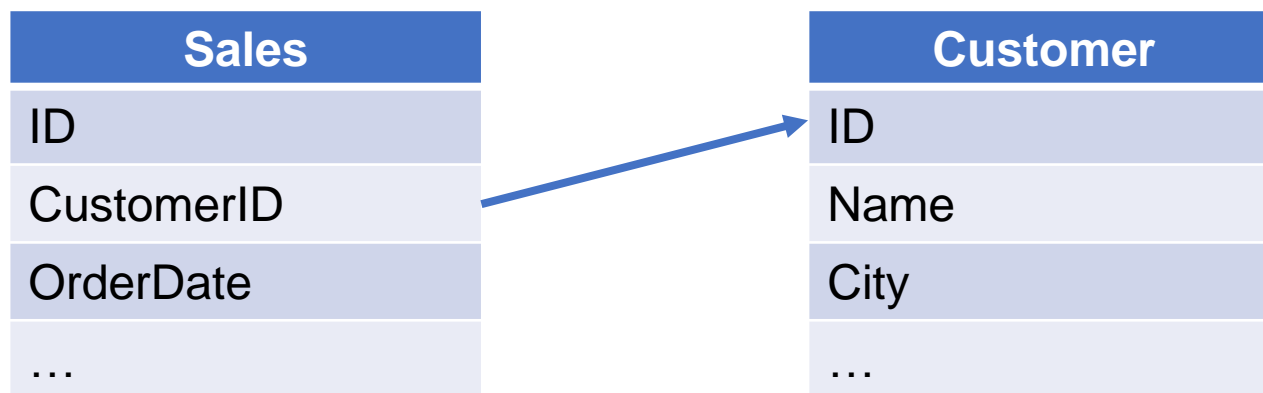
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Sales where year(Date)=2014")	/查询销售表2014年的数据
3	=A1.query("select * from Customer where City='Beijing'")	/查询北京的客户
4	=A2.switch@i(CustomerID, A3:ID)	/使用switch函数的@i选项, 仅保留北京客户的记录
5	=A4.groups(CustomerID.Name:Name; sum(Amount):Amount).sort@z(Amount)	/分组汇总各个客户的销售额, 并按总销售额降序排列

A5	Name	Amount
	SAVEA	130672.64
	HUN	23959.05

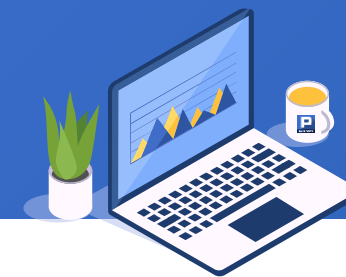
✦ 2. 只保留无匹配的记录



有销售表和客户表，查询2014年的新增客户，即客户ID不在客户表中的记录。



✦ 2. 只保留无匹配的记录



SPL如下，其中用到了A.switch()函数的@d选项，只保留无匹配的记录，此时不会把外键字段设置为null：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Sales where year(OrderDate)=2014")	/查询2014年的销售记录
3	=A1.query("select * from Customer")	/查询客户表
4	=A2.switch@d(CustomerID ,A3:ID)	/使用switch@d()从销售表中选出客户ID在客户表中不存在的记录

A4	ID	CustomerID	OrderDate	...
	10439	MEREP	2014/02/07	...
	10504	WHITC	2014/04/11	...

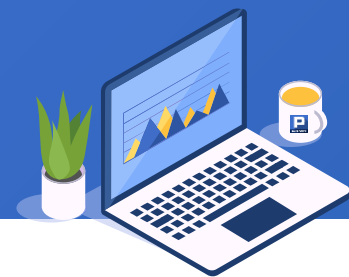
CONTENTS

1. 散列法连接
2. 有序归并法连接
3. 外连接
4. 叉积
5. 直接连接文件

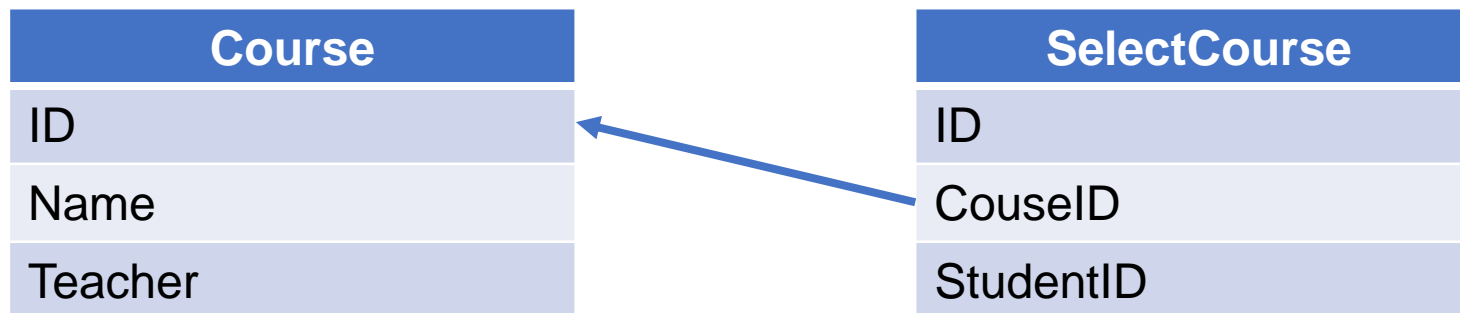


常规连接

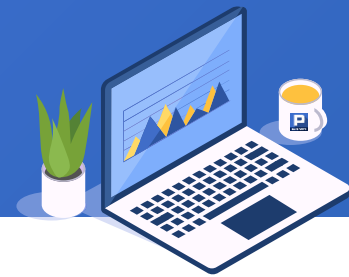
✦ 1. 散列法连接



查询有多少学生选修了“Matlab”课程。课程表和选课表如下：



✦ 1. 散列法连接

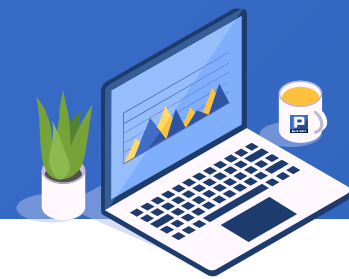


这里用到了A.join()函数的@i选项, 删除不匹配的记录。 SPL如下:

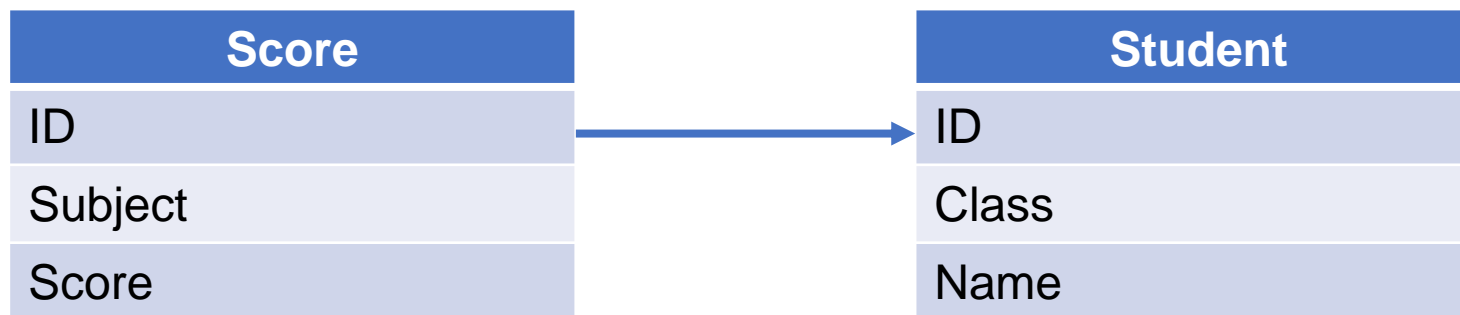
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Course")	/查询课程表
3	=A1.query("select * from SelectCourse")	/查询选课表
4	=A2.select(Name=="Matlab")	/从课程表中选出指定课程
5	=A3.join@i(CourseID,A4:ID).count()	/使用join函数的@i选项进行连接过滤, 并计数

A5	Value
	5

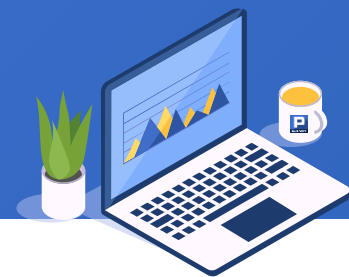
✦ 1. 散列法连接



查询一班每个学生的各科总分。成绩表和学生表如下：



✦ 1. 散列法连接

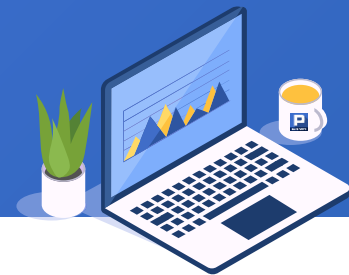


这里用到了A.join()函数的@i选项, 删除不匹配的记录。在连接时把班级指定为常数条件 (一班), 从而实现连接过滤。 SPL如下:

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Score")	/查询成绩表
3	=A1.query("select * from Student")	/查询学生表
4	=A2.join@i(ID:"Class one", A3:ID:Class)	/使用join函数的@i选项, 按ID和Class进行多主键连接过滤
5	=A4.groups(ID; sum(Score):TotalScore)	/分组汇总每个学生的总分

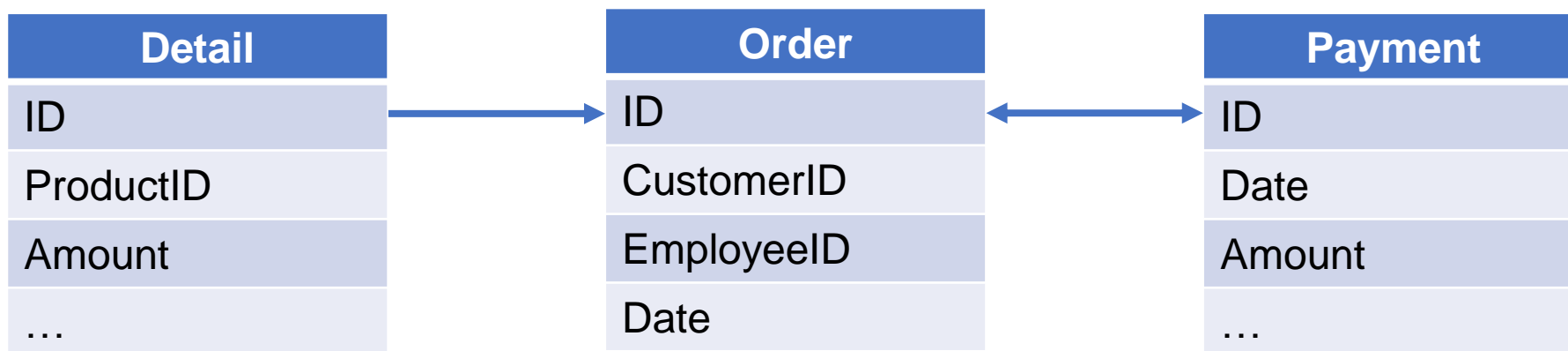
A5	ID	TotalScore
	1	230
	2	258
	3	228

◆ 1. 散列法连接

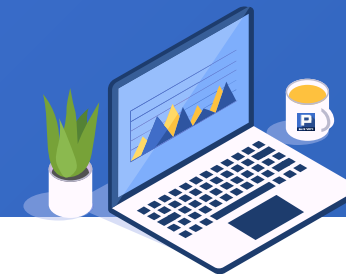


查询哪些订单还没有回款，也就是累计回款金额小于订单总金额的订单。

订单明细表、订单表和订单回款表如下：



◆ 1. 散列法连接

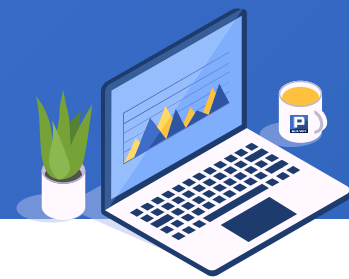


这里用到了join()函数进行连接。 SPL如下：

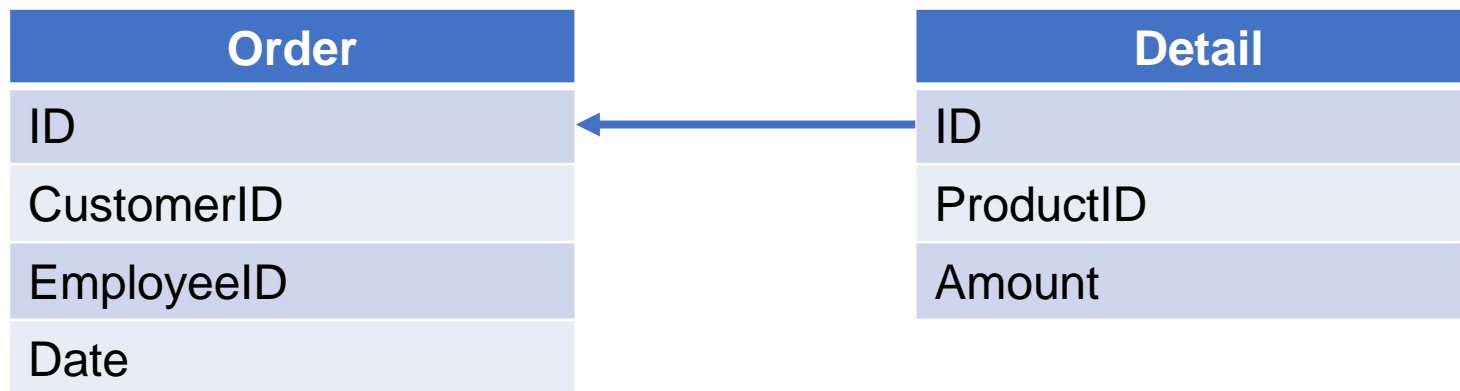
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Order")	/查询订单表
3	=A1.query("select * from Detail")	/查询订单明细表
4	=A1.query("select * from Payment")	/查询订单回款表
5	=A3.group(id)	/订单明细表按订单ID分组
6	=A4.group(id)	/订单回款表按订单ID分组
7	=join(A2:Order,ID; A5:Detail,ID; A6:Payment,ID)	/使用join函数按订单ID连接订单表、订单明细表和订单回款表
8	=A7.new(Order.ID:ID,Detail.sum(Amount):Amount,Payment.sum(Amount):Pay)	/创建序表，汇总每笔订单的订单金额和还款金额
9	=A8.select(Pay<Amount)	/选出还款金额小于订单金额的记录

A9	ID	Amount	Pay
	AROUT	55492.0	35980
	BERGS	3398.55	1080

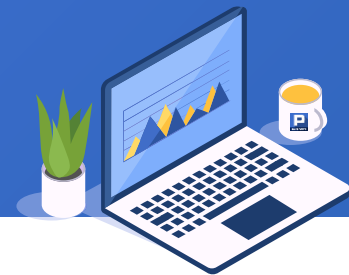
✦ 2. 有序归并法连接



查询2014年每个客户的销售额，订单表和订单明细表如下：



✦ 2. 有序归并法连接

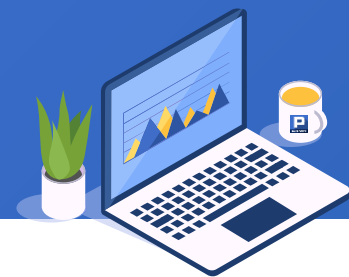


这里用到了join()函数的@m选项，进行有序归并。SPL如下：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Order where year(Date)=2014 order by ID")	/查询2014年的订单，并按订单ID排序
3	=A1.query("select * from Detail order by ID")	/查询订单明细表，并按订单ID排序
4	=join@m(A2:Order,ID;A3:Detail,ID)	/使用join@m函数对订单表和订单明细表有序归并
5	=A4.groups(Order.CustomerID:CustomerID; sum(Detail.Amount):Amount)	/分组汇总每个客户的销售额

A5	CustomerID	Amount
	ALFKI	14848.0
	ANTON	4041.0

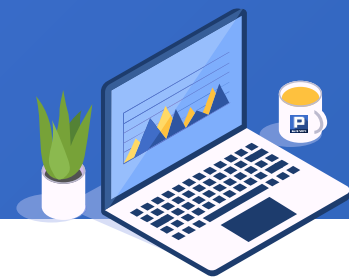
✦ 2. 有序归并法连接



当订单明细和订单表数据量很大，无法加载到内存时，就需要使用游标的有序归并。
要查询总销售额超过一万的客户。



✦ 2. 有序归并法连接

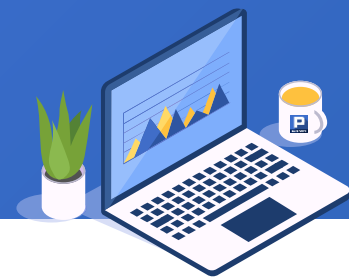


这里用到了joinx()函数进行有序归并。SPL如下：

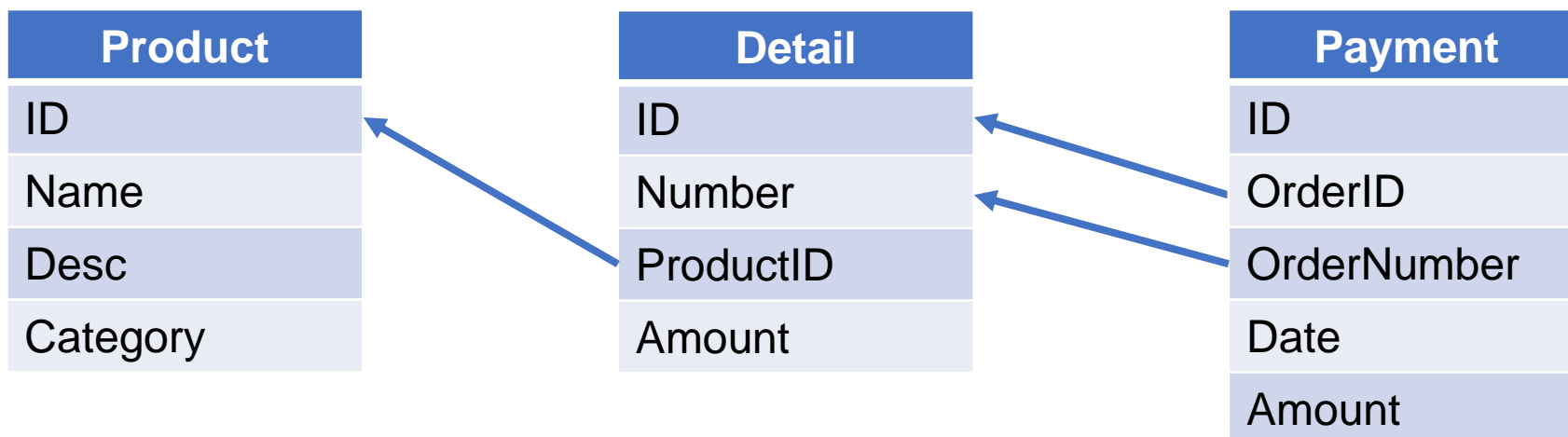
	A	B
1	=connect("db")	/连接数据库
2	=A1.cursor("select * from Order order by ID")	/创建订单表游标
3	=A1.cursor("select * from Detail order by ID")	/创建订单明细表游标
4	=A1.query("select * from Customer")	/查询客户表
5	=A2.switch@i(CustomerID,A4:ID)	/使用switch@i函数，将订单表的客户ID转换成对应记录，并删除不匹配记录
6	=joinx(A5:Order,ID;A3:Detail,ID)	/使用joinx函数对订单表和订单明细表的游标进行有序归并
7	=A6.groups(Order.CustomerID.Name;sum(Detail.Amount):Amount).select(Amount>10000)	/分组汇总每个客户的销售额，并选出销售额大于一万的记录

A7	Name	Amount
	ALFKI	14848.0
	AROUT	55492.0

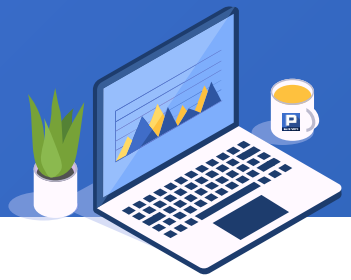
✦ 3. 外连接



查询2014年有回款记录的，单笔订单金额超过500的订购产品名称。订单明细表、订单回款表和产品表如下：



✦ 3. 外连接

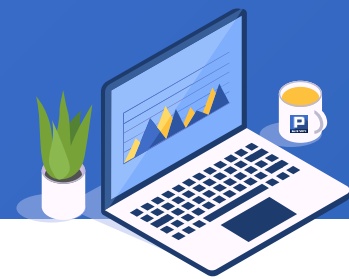


这里用到了A.join()函数进行多主键的外连接。SPL如下：

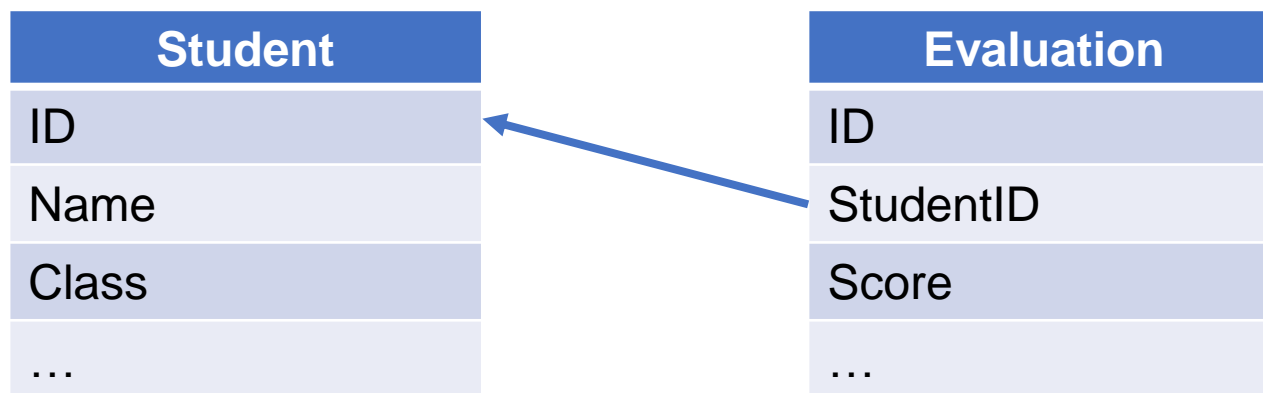
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Detail")	/查询订单明细表
3	=A1.query("select * from Payment")	/查询订单回款表
4	=A1.query("select * from Product")	/查询产品表
5	=A2.switch@i(ProductID,A4:ID)	/使用switch@i函数，将订单明细表的产品ID转换成对应记录
6	=A3.join(OrderID:OrderNumber,A5:ID:Number,~:Detail)	/使用A.join函数连接订单明细表和订单回款表
7	=A6.select(year(Date)==2014 && Detail.Amount>500)	/选出2014年的回款记录，订单金额超过500的。
8	=A7.new(ID,Date,Detail.Product.Name:Name,Detail.Amount:Amount)	/从选出结果创建序表

A8	ID	Date	Name	Amount
	10979	2014/03/26	Soda water	1317
	11011	2014/04/09	Espresso	530

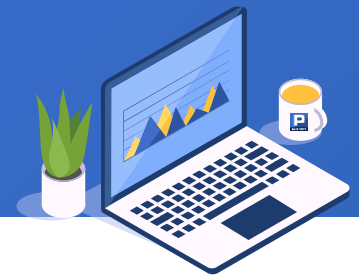
✦ 3. 外连接



有学生表和评价表，学生基础评分70分，查询每个学生的评分。



✦ 3. 外连接

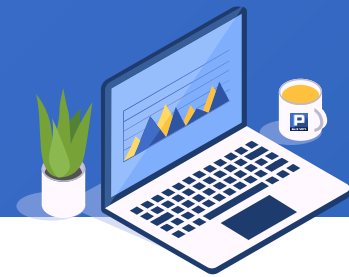


这里用到了join()函数的@1选项, 按照第一个表结构进行连接。 SPL如下:

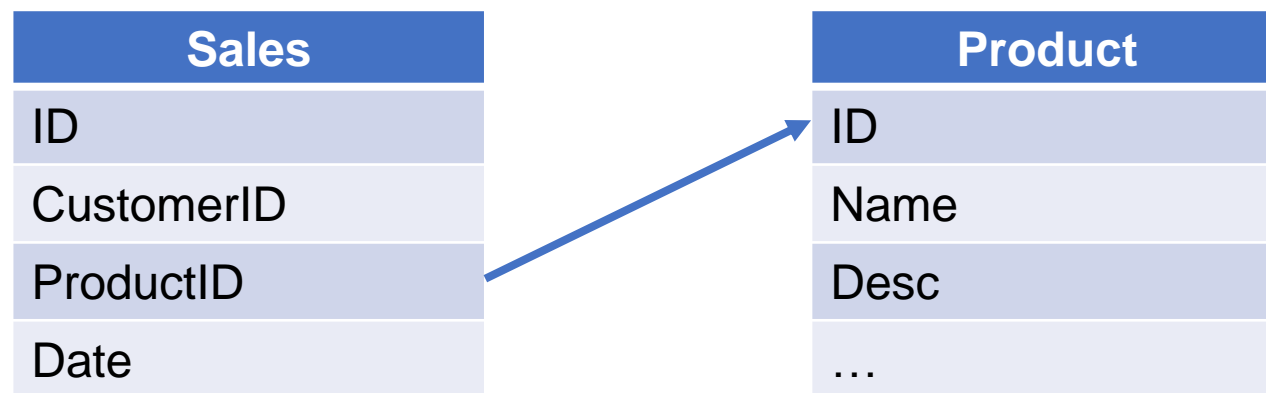
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Students")	/查询学生表
3	=A1.query("select * from Evaluation")	/查询评价表
4	=A3.group(StudentID)	/评价表按照学生编号分组
5	=join@1(A2:Students,ID;A4:Evaluation,StudentID)	/使用join@1按照学生表左连接分组后的评价表
6	=A5.new(Students.ID:ID,Students.Name:Name,70+Evaluation.sum(Score):Score)	/创建序表, 计算每个学生的总评分 (基础分70+评价总分数)

A6	ID	Name	Score
	1	Ashley	85
	2	Rachel	65
	3	Emily	70

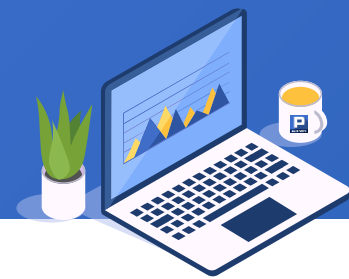
✦ 3. 外连接



销售表和产品表如下，要比较2014年每个月各种产品的销售情况。



✦ 4. 叉积

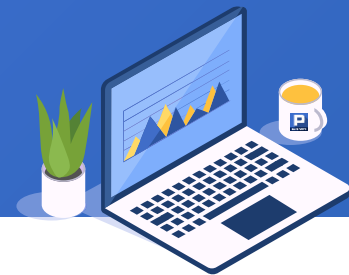


查询哪两种三明治的配料最接近。三明治表和配方表如下：

Sandwich		
ID	Name	Price
1	BLT	5.5
2	Reuben	7.0
3	Grilled Cheese	3.75

Ingredient	
ID	Ingredient
1	bacon
1	lettuce
1	tomato
...	...

✦ 4. 叉积

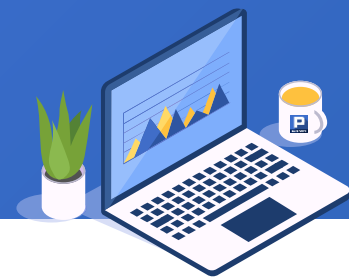


这里用到了xjoin()函数计算叉积。值得注意的是SPL叉积结果的记录是由两个表的记录组成，而不是简单的把所有字段展开。SPL如下：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select i.ID ID, i.Ingredient Ingredient, s.Name Name from Sandwich s, Ingredient i where s.ID=i.ID order by ID")	/查询三明治表和配方表
3	=A2.group@o(ID;Name,~.(Ingredient):Collection)	/使用group@o()按ID归并分组，并将每种三明治配料存储在Collection字段
4	=xjoin(A3:A;A3:B,A.ID<ID)	/使用xjoin函数将查询结果与自身计算叉积，并选出ID不同的组合
5	=A4.new((A.Collection ^ B.Collection).len():Count, A.Name:Name1, B.Name:Name2).sort@z(Count)	/计算两种配方有多少重复的配料，并按照重复数量降序排列。

A5	Count	Name1	Name2
	1	Reuben	Grilled Cheese
	0	BLT	Reuben
	0	BLT	Grilled Cheese

✦ 4. 叉积



矩阵表的数据结构如下，求两个矩阵的乘积：

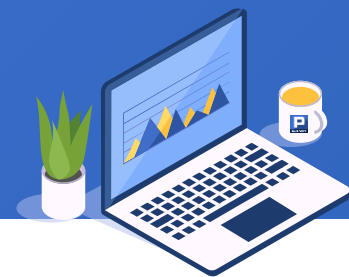
Matrix
row
col
value

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

本例的数学公式如下：

$$C = AB = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 1 \times 1 + 2 \times 2 + 3 \times 3 & 1 \times 4 + 2 \times 5 + 3 \times 6 \\ 4 \times 1 + 5 \times 2 + 6 \times 3 & 4 \times 4 + 5 \times 5 + 6 \times 6 \end{pmatrix} = \begin{pmatrix} 14 & 32 \\ 32 & 77 \end{pmatrix}$$

✦ 4. 叉积

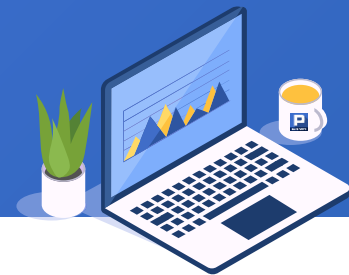


SPL如下，其中用到了xjoin()函数计算叉积，同时可以根据条件过滤：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from MatrixA")	/查询矩阵表A
3	=A1.query("select * from MatrixB")	/查询矩阵表B
4	=xjoin(A2:A; A3:B, A.col==row)	/使用xjoin函数计算两表叉积，同时进行条件过滤
5	=A4.groups(A.row:row,B.col:col;sum(A.value * B.value):value)	/将结果分组汇总，计算每个行列点的值

A5	row	col	value
	1	1	14
	1	2	32
	2	1	32
	2	2	77

✦ 4. 叉积

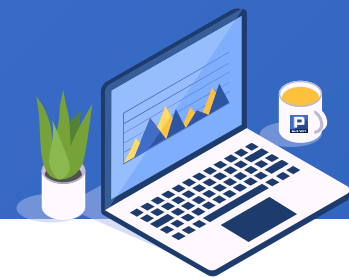


查询社区居民所处的年龄段。社区表和年龄表如下：

Community		
ID	Name	Age
1	David	28
2	Daniel	15
3	Andrew	65
4	Rudy	

Age		
Group	Start	End
Children	0	15
Youth	16	40
Middle	41	60
Old	61	100

✦ 4. 叉积

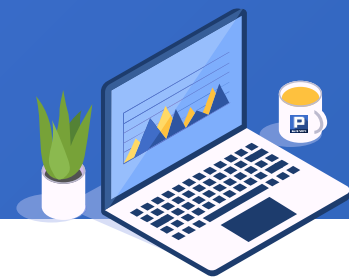


SPL如下，其中用到了xjoin()函数的@1选项计算叉积时使用左连接：

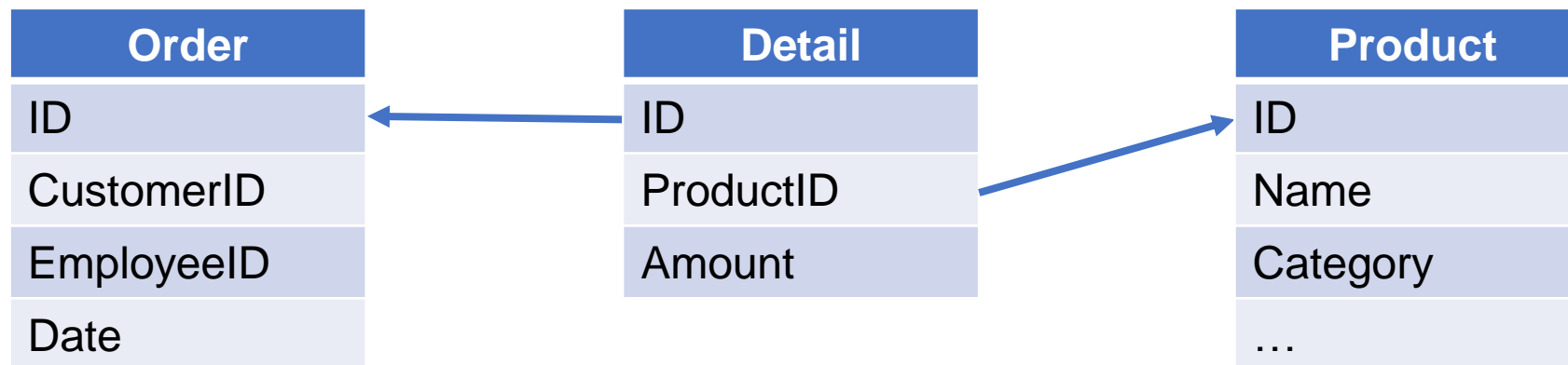
	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Community")	/查询社区表
3	=A1.query("select * from Age")	/查询年龄表
4	=xjoin@1(A2:Person; A3:Age, Start <= Person.Age && End >= Person.Age)	/使用xjoin@1()函数计算叉积时使用左连接，同时选出年龄在相应的年龄区间的记录
5	=A4.new(Person.ID:ID, Person.Name:Name, Person.Age:Age, Age.Group:Group)	/创建序表，返回每个居民所在年龄段

A5	ID	Name	Age	Group
	1	David	28	Youth
	2	Daniel	15	Children
	3	Andrew	65	Old
	4	Rudy	(null)	(null)

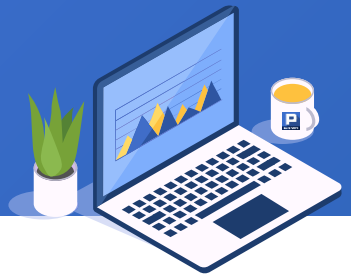
✦ 5. 直接连接文件



查询2014年1月各种产品的销售数量。订单表、订单明细表和产品表如下，其中产品表集文件对ID有序：



✦ 5. 直接连接文件

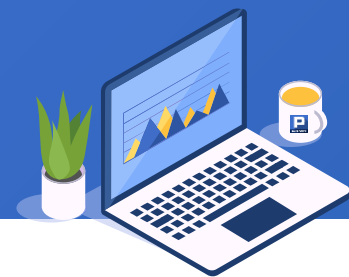


本例用到了cs.joinx()函数连接文件，集文件对于连接字段必须有序。SPL如下：

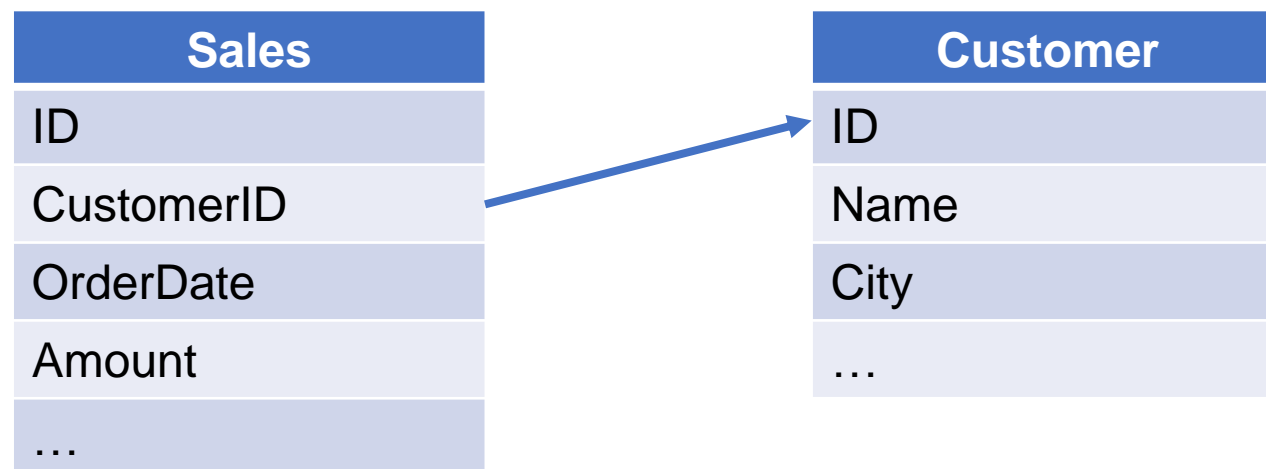
	A	B
1	=file("Detail.ctx").create().cursor()	/创建订单明细表游标
2	=file("Order.ctx").create().cursor(;year(Date)=2014 && month(Date)=1)	/创建2014年1月的订单游标
3	=file("Product.btx")	/创建产品表集文件对象
4	=A1.joinx@i(ID,A2:ID)	/使用cs.joinx函数的@i选项进行连接过滤
5	=A4.joinx(ProductID,A3:ID,Name:ProductName)	/使用cs.joinx函数按产品ID连接订单明细表和产品表
6	=A5.groups(ProductName; count(~):Count)	/分组汇总各产品的销售数量

A6	ProductName	Count
	Milk	32
	Coffee	60

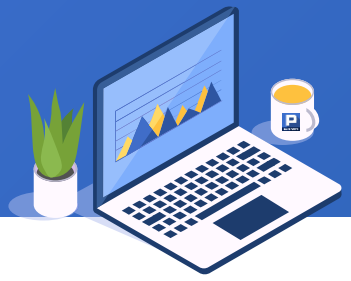
✦ 5. 直接连接文件



查询2014年销售额前三的客户名称。销售表和客户表如下，其中客户表集文件对ID有序：



✦ 5. 直接连接文件



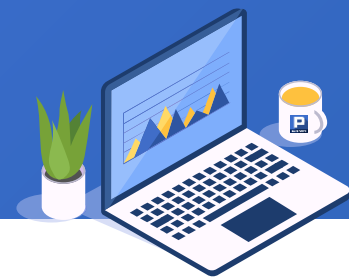
使用cs.joinx()函数与可分段集文件进行连接，当数据量不大时可以使用@q选项提速。

SPL如下：

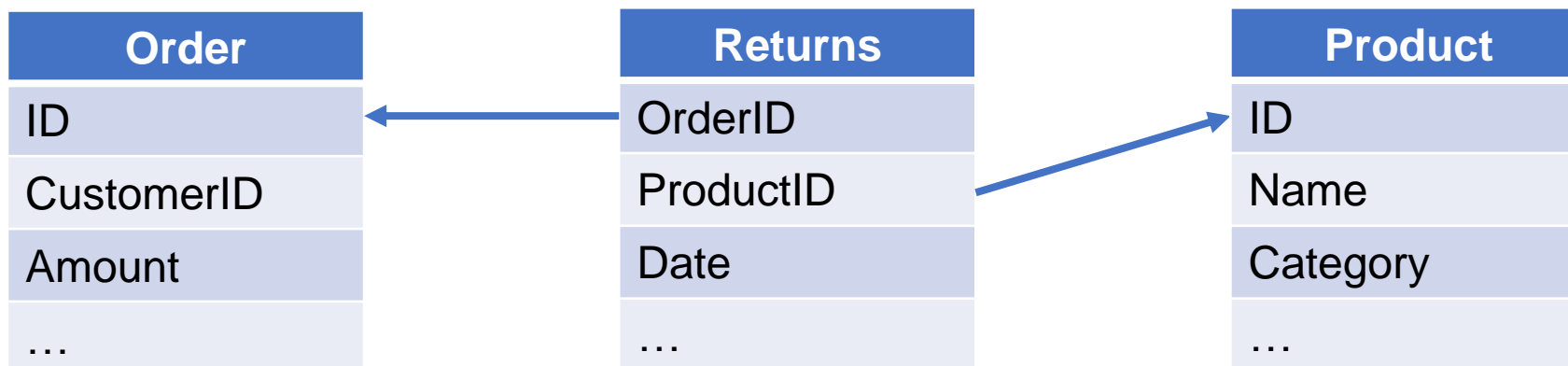
	A	B
1	=file("Sales.btx").cursor@b().select(year(Date)==2014)	/从集文件创建游标，选出2014年的销售记录
2	=file("Customer.btx")	/创建客户表集文件对象，该文件对于客户ID有序
3	=A1.groups(CustomerID;sum(Amount):Amount)	/分组汇总客户的销售额
4	=A3.top(-3;Amount)	/选出销售额前三名
5	=A4.joinx@q(CustomerID,A2:ID,Name:CustomerName).fetch()	/使用cs.joinx函数按客户ID连接销售表和客户表，数据量不大时使用了@q选项提速。

A5	CustomerID	Amount	CustomerName
	71	130672.64	SAVEA
	63	64238.0	QUICK
	20	53467.38	ERNSH

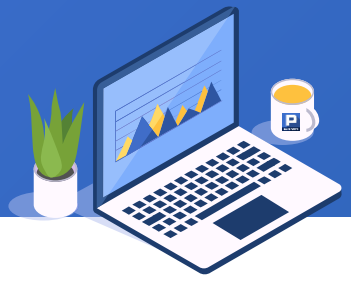
✦ 5. 直接连接文件



查询2015年各类产品的退货金额。订单表、退货表和产品表如下：



✦ 5. 直接连接文件



本例用到了`cs.joinx()`函数连接文件，当游标对第一个连接字段有序时可以使用`@c`选项提速，`@c`选项可以和`@q`选项同时使用。SPL如下：

	A	B
1	<code>=file("Returns.btx").cursor@b().select(year(Date)==2015)</code>	/创建订单明细表游标
2	<code>=file("Order.btx")</code>	/创建订单表集文件对象
3	<code>=file("Product.btx")</code>	/创建产品表集文件对象
4	<code>=A1.joinx@qc(OrderID,A2:ID,Amount;ProductID,A3:ID,Category)</code>	/使用 <code>cs.joinx</code> 函数按订单ID连接订单表，按产品ID连接产品表， <code>@qc</code> 选项同时使用进行提速
5	<code>=A4.groups(Category; sum(Amount))</code>	/分组汇总各类产品的退货金额

A5	Category	Amount
	Electric appliance	1854.5
	Fruits	251.5

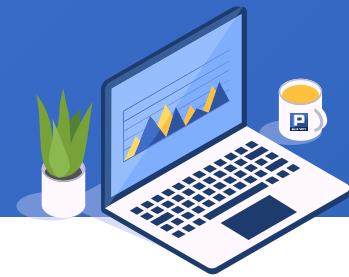
CONTENTS

1. 序号直接定位
2. 按位置连接
3. 按位置外连接

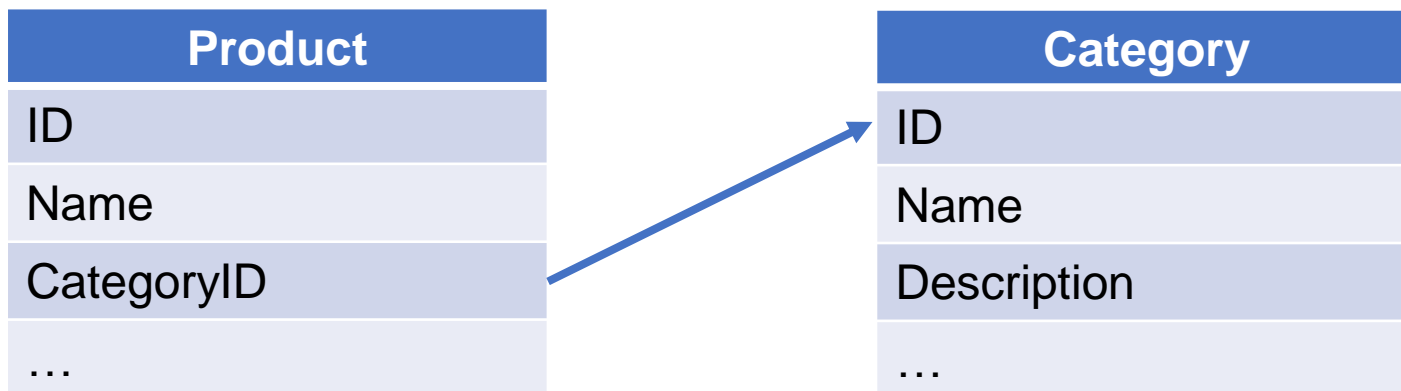


按位置连接

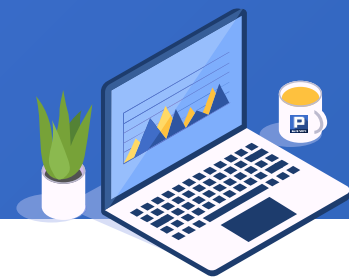
✦ 1. 序号直接定位



有产品表和类别表，要查询产品类别名称包含"drink"的产品，如下图：



✦ 1. 序号直接定位

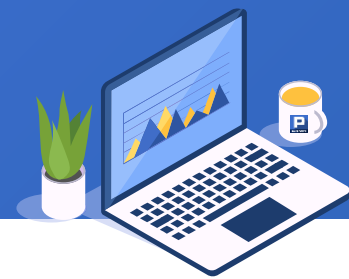


这里用到了A.join()函数进行连接，使用#按序号定位，#1表示第一个字段。SPL如下：

	A	B
1	=connect("demo")	/连接数据源
2	=A1.query("select * from Product")	/查询产品表
3	=A1.query("select * from Category")	/查询类别表
4	=A3.select(like@c(Name, "*drink*"))	/选出类别名称中包含drink的记录，不区分大小写
5	=A2.join@i(CategoryID,A4:#1)	/产品表的类别ID字段与类别表的第一个字段 (ID) 字段进行连接过滤

A5	ID	Name	CategoryID
	24	Soda	1
	34	Beer	1
	35	Orange Juice	1

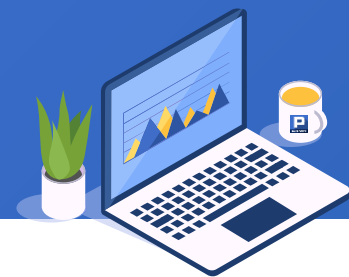
✦ 2. 按位置连接



想要将数据库某列的值打乱，并将新值对应写回数据库中。引用值表如下：

ID	ORIGINAL_VALUE	SHUFFLED_VALUE
1	D	N
2	U	n
3	j	K
4	N	D
...

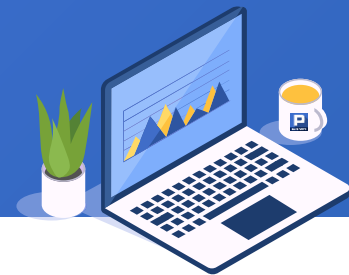
✦ 2. 按位置连接



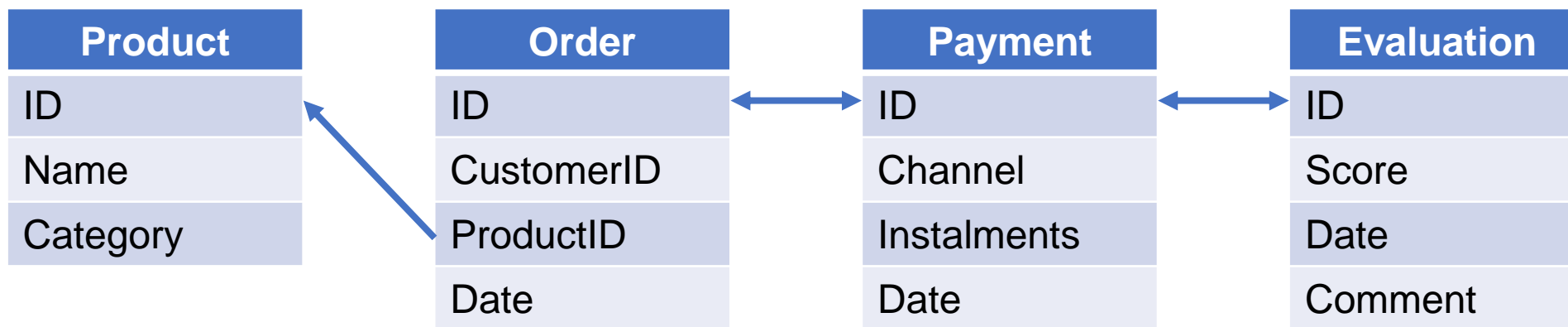
这里用到了join()函数的@p选项，按位置进行连接。SPL如下：

	A	B
1	=connect("demo")	/连接数据库
2	=A1.query("select ID,ORIGINAL_VALUE from REF_VALUES")	/查询引用值表
3	=A2.sort(rand())	/按随机种子排序，用来扰乱序表
4	=join@p(A2.(ID);A3.(ORIGINAL_VALUE))	/使用join@p按位置连接原始ID和扰乱后的值
5	=A1.update@u(A4, REF_VALUES, ID:_1, SHUFFLED_VALUE:_2; ID)	/根据主键ID，将A4的结果更新到数据库表REF_VALUES

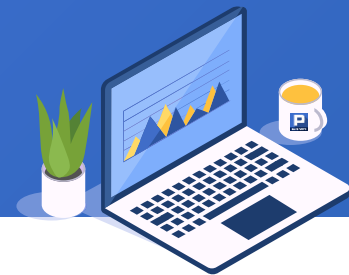
✦ 2. 按位置连接



查询2014年未使用分期付款支付的订单，各类产品的平均评分是多少。其中订单表、支付表和评价表的订单编号相同，关系如下：



✦ 2. 按位置连接

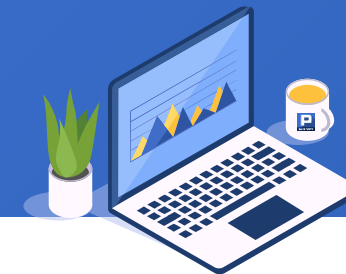


这里用到了join()函数的@p选项，按位置进行连接。SPL如下：

	A	B
1	=connect("demo")	/连接数据库
2	=A1.query("select * from Order order by ID")	/查询订单表
3	=A1.query("select * from Payment order by ID")	/查询支付表
4	=A1.query("select * from Evaluation order by ID")	/查询评价表
5	=A1.query("select * from Product")	/查询产品表
6	=A2.switch(ProductID, A5:ID)	/使用switch函数将产品ID转换为对应的产品记录
7	=join@p(A6:Order;A3:Payment;A4:Evaluation)	/使用join@p函数将三个表按位置进行连接
8	=A7.select(year(Order.Date)==2014 && !Payment.Instalments)	/选出2014年未使用分期付款的订单
9	=A8.groups(Order.ProductID.Category; avg(Evaluation.Score):Score)	/分组汇总不同类别产品的平均评分

A9	Category	Score
	Electric appliance	3.98
	Fruits	3.86

✦ 3. 按位置外连接



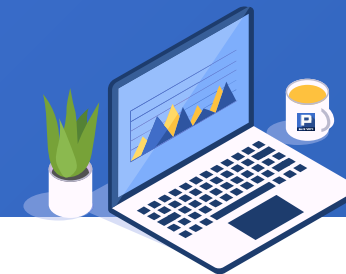
世界城市人口表，如下图：

Continent	Country	City	Population
Africa	Egypt	Cairo	6789479
Asia	China	Shanghai	24240000
Europe	Britain	London	7285000

分栏列出欧洲和非洲人口超 200 万的城市名称及人口（每栏按从多到少排序），期望结果如下图：

Europe City	Population	Africa City	Population
Moscow	8389200	Cairo	6789479
London	7285000	Kinshasa	5064000
St Petersburg	4694000	Alexandria	3328196

✦ 3. 按位置外连接



使用A.paste()函数按位置外连接欧洲和非洲数据，这里省略了字段名，会按照字段顺序进行修改。SPL如下：

	A	B
1	=connect("db").query("select * from World where Continent in('Europe', 'Africa') and Population >= 2000000")	/连接数据库，取出欧洲和非洲超过200万人口的记录
2	=A1.select(Continent:"Europe")	/使用select()函数取出欧洲数据
3	=A1.select(Continent:"Africa")	/使用select()函数取出非洲数据
4	=A2.new(City:'Europe City',Population:'Europe Population','Africa City','Africa Population')	/通过欧洲数据创建序表
5	=A4.paste(A3.(City):#3,A3.(Population):#4;1)	/使用paste()函数按位置外连接欧洲数据和非洲数据，将值序列依次填充到第三列至第四列

A5	Europe City	Population	Africa City	Population
	Moscow	8389200	Cairo	6789479
	London	7285000	Kinshasa	5064000
	St Petersburg	4694000	Alexandria	3328196

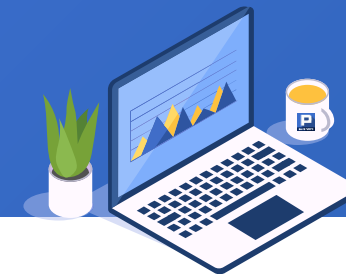
CONTENTS

1. 关联连接与关联左连接
2. 退化成交积
3. 复杂用法



关联连接

◆ 1. 关联连接与关联左连接

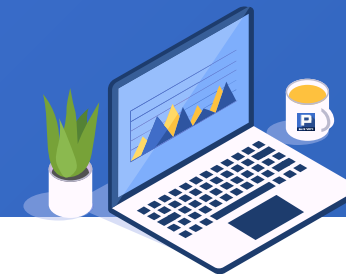


某小学调查学生在线教学可以使用的终端，现在要统计各终端占比。各班问卷及汇总目录如下：

▼	Primary School
>	Grade1
>	Grade2
▼	Grade3
	Class1
	Class2
	Class3
	Class4
	Class5
	Class6
>	Grade4
>	Grade5
>	Grade6

ID	STUDENT_NAME	TERMINAL
1	Rebecca Moore	Phone
2	Ashley Wilson	Phone,PC,Pad
3	Rachel Johnson	Phone,PC,Pad
4	Emily Smith	Phone,Pad
5	Ashley Smith	Phone,PC
6	Matthew Johnson	Phone
7	Alexis Smith	Phone,PC
8	Megan Wilson	Phone,PC,Pad
...

✦ 1. 关联连接与关联左连接

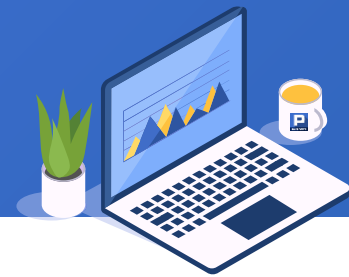


使用了A.news()函数将序表与终端拆分的序列进行连接。SPL如下：

	A	B	
1	=directory@ps("D:/Primary School")		/递归遍历目录，列出所有文件
2	for A1	=file(A2).xlsimport@t()	/循环导入各年级各班的问卷excel文件
3		=@ += B2.len()	/统计总行数，即总人数
4		=B2.news(B2.TERMINAL.split@c(); ID, STUDENT_NAME, ~:TERMINAL)	/使用news函数将问卷表与拆分后的终端连接
5		=B4.groups(TERMINAL; count(~):Count) @	/分组汇总各终端的数量，结果合并到本格。如果合并B4的结果可能导致内存不足。
6	=B5.groups(TERMINAL;string(sum(Count)/B3, "#.##%"):PERCENTAGE)		/对每个班的统计结果再分组汇总

A6	
TERMINAL	PERCENTAGE
PC	70%
Pad	56.67%
Phone	93.33%

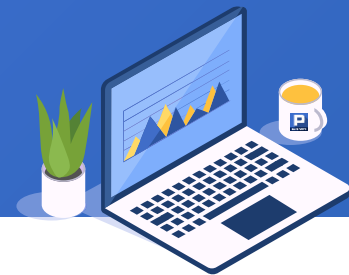
◆ 1. 关联连接与关联左连接



下面是发帖记录表，统计每位作者最常用的标签。

ID	TITLE	Author	Label
1	Easy analysis of Excel	Ashley	Excel,ETL,Import,Export
2	Early commute: Easy to pivot excel	Rachel	Excel,Pivot,Python
3	Initial experience of SPL	Rebecca	
4	Talking about set and reference	Emily	Set,Reference,Dispersed,SQL
5	Early commute: Better weapon than Python	Emily	Python,Contrast,Install
...

✦ 1. 关联连接与关联左连接

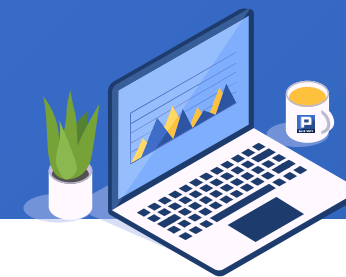


SPL如下，使用A.news()函数的@1选项进行左连接：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from PostRecord")	/查询发帖记录表
3	=A2.news@1(A2.Label.split@c(); ID,Title,Author,~:Label)	/使用A.news函数的@1选项左连接帖子的标签，当标签不存在时仍保留发帖记录
4	=A3.groups(Author,Label;count(~):Count)	/分组汇总每位作者各标签的数量
5	=A4.group(Author).conj(~.maxp@a(Count))	/按作者分组选出数量最多的标签，并列的会同时选出。

A5	Author	Label	Count
	Rebecca	(null)	1
	Ashley	Excel	3
	Ashley	SPL	3
	Rachel	Python	4

✦ 2. 退化交叉积

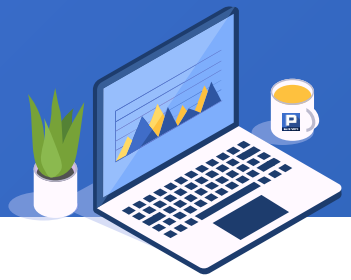


下面是任课教师表和课程表。要查询每个课程可以安排的教师。

Teachers		
Teacher	Branch	Courses
Petitti	Matematica	28,33,30,35
Canales	Apesca	11,16,12,17,13,18,14,19
Lucero	NavegacionI	6,11,16,21,7,12,17,22,...
Bergamaschi	TecPesc	1,26,2,27,3,28,4,29,5,30
...

Courses	
ID	Name
1	lua
2	maa
3	mia
4	jua
...	...

✦ 2. 退化交叉积

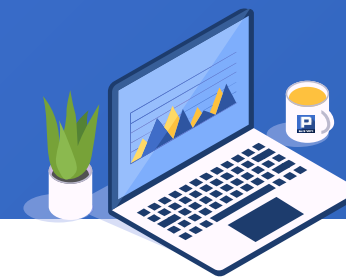


SPL如下，其中用到了A.news()函数与其他序列计算交叉积：

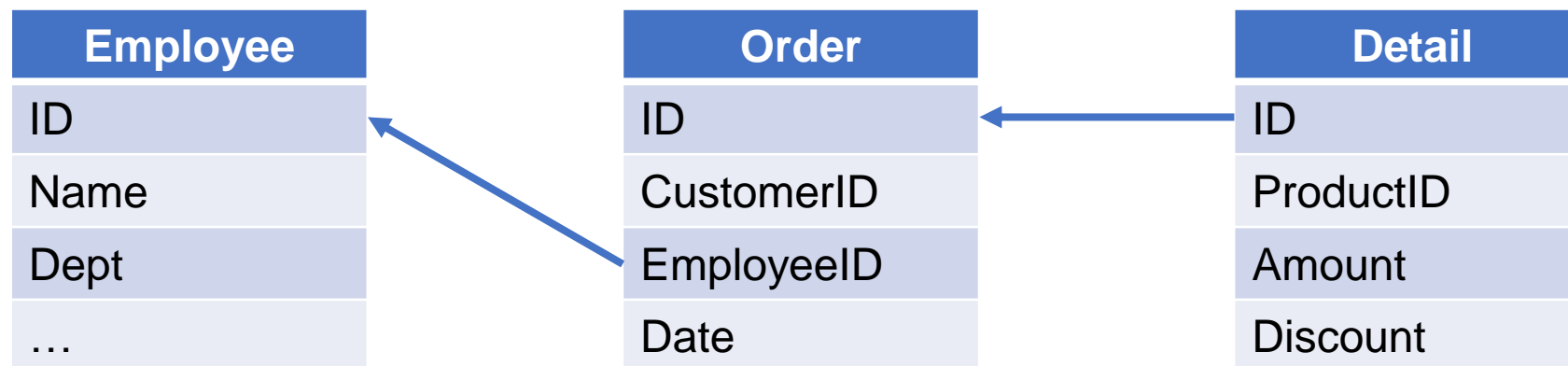
	A	B
1	<code>=file("Teachers.txt").import@t().run(Courses=Courses.split@cp())</code>	/导入教师表，并将课程按逗号拆分成序列
2	<code>=file("Courses.txt").import@t()</code>	/导入课程表
3	<code>=A2.news(A1;ID,Name:Course,Teacher,Courses)</code>	/使用A.news函数计算课程表和教师表的交叉积
4	<code>=A3.select(Courses.contain(ID))</code>	/选出课程序列包含当前课程ID的记录
5	<code>=A4.group(Course).new(Course,~.(Teacher).concat@c():Teachers)</code>	/按课程分组后创建序列，将教师序列用逗号拼在一起组成教师字段

A5	Course	Teachers
	jua	Bergamaschi,Puebla,Jimenez
	jub	Lucero,Mazza,Puebla,Chiatti,Jimenez,Luceroo
	juc	Canales,Lucero,Mazza,Puebla,Chiatti,Luceroo

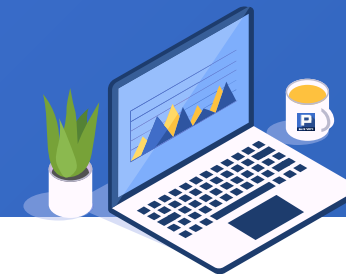
✦ 3. 复杂用法



对单笔订单的实际金额超过1000的销售人员，再给予 5% 的业绩奖励。
求2014年每位员工的实际销售额。



✦ 3. 复杂用法



这里用到了A.news()函数进行连接和计算。SPL如下：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Order where year(Date)=2014")	/查询订单表2014年数据
3	=A1.query("select * from Detail")	/查询订单明细表
4	=A1.query("select * from Employee")	/查询员工表
5	=A2.switch(EmployeeID,A4:ID)	/使用switch函数，将订单表的员工ID转换成对应记录
6	=A3.group(ID)	/订单明细表对订单ID分组
7	=A6.news(A2.select(ID:A6.~.ID); EmployeeID,(s=sum(Amount*(1-Discount)), if(s>1000, s*1.05, s)):Amount)	/使用news函数，将订单明细表按照订单ID与订单表进行连接，并计算出每笔订单的实际业绩
8	=A7.groups(EmployeeID.Name:Name; sum(Amount):Amount)	/分组汇总每位员工的总销售业绩

A8	Name	Amount
	Alexis	358882.02
	Emily	432435.85

THANKS

感谢观看

