

SPL Base

集算器教案

TopN及变种



CONTENTS

01 最大值/最小值

02 前N个/后N个

03 分组中的使用

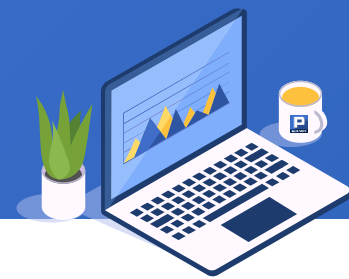


CONTENTS



最大值/最小值

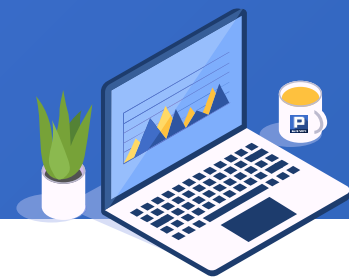
✦ 1. 取值



以成绩表为例，求一班数学最高分。

CLASS	STUDENTID	SUBJECT	SCORE
Class one	1	English	84
Class one	1	Math	77
Class one	1	PE	69
Class one	2	English	81
Class one	2	Math	80
...

✦ 1. 取值



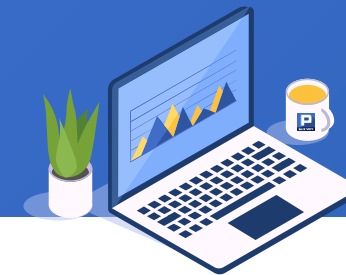
SPL如下，其中用到了max函数来取最大值：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Scores where SUBJECT='Math' and CLASS='Class one'")	/查询一班的数学成绩
3	=A2.max(SCORE)	/取出最高分

同样可以使用min函数来取最小值：

	A	B
3	=A2.min(SCORE)	/取出最低分

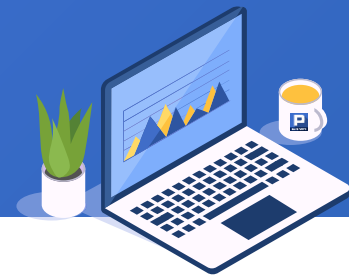
✦ 2. 取对应记录的行号



求2019年股市收盘价最高点那天，相对于前一日的涨幅。

Date	Opening price	Closing price	Amount
2019/12/4	2876.9079	2878.1157	136000000000
2019/12/3	2869.8822	2884.6988	135000000000
2019/12/2	2874.4484	2875.8072	139000000000
2019/11/29	2885.9744	2871.9813	140000000000
2019/11/28	2902.3644	2889.6934	123000000000
...

✦ 2. 取对应记录的行号



我们需要知道股市最高点的记录所在行号，再与上一个交易日比较得出结果。

SPL如下，其中用到了pmax函数来取最大值所在行号：

	A	B
1	=file("000001.csv").import@ct()	/导入数据文件
2	=A1.select(YEAR(Date)==2019)	/选出2019年的记录
3	=A2.pmax('Closing price')	/取出股市最高点所在行号
4	=A2(A3).'Closing price'/A2(A3+1).'Closing price'-1	/使用当天收盘价和前日收盘价计算涨幅

同样可以使用pmin函数来取最小值所在行号：

	A	B
3	=A2.pmin('Closing price')	/取出股市最低点所在行号

✦ 2. 取对应记录的行号



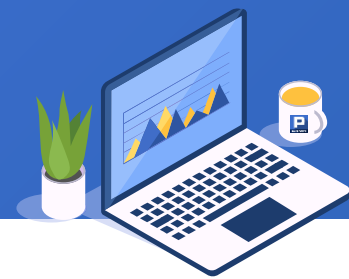
最大值不一定是唯一的，如果想返回所有的行号，可以使用pmax@a选项：

	A	B
3	=A2.pmax@a('Closing price')	/取出所有股市最高点记录所在行号

如果希望从后向前定位，可以使用pmax@z选项：

	A	B
3	=A2.pmax@z('Closing price')	/从后向前取出股市最高点记录所在行号

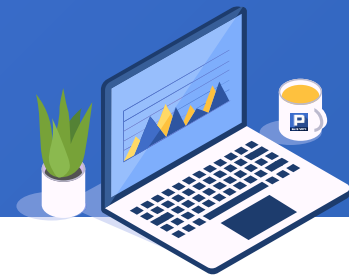
✦ 3. 取对应记录



以成绩表为例，求一班数学最高分的学生ID。

CLASS	STUDENTID	SUBJECT	SCORE
Class one	1	English	84
Class one	1	Math	77
Class one	1	PE	69
Class one	2	English	81
Class one	2	Math	80
...

✦ 3. 取对应记录



SPL如下，其中用到了maxp函数来取最大值所在记录，再从中取学生ID：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Scores where SUBJECT='Math' and CLASS='Class one'")	/查询一班的数学成绩
3	=A2.maxp(SCORE)	/取出最高分所在记录
4	=A3.STUDENTID	/从记录中取学生ID

同样可以使用minp函数来取最小值所在记录：

	A	B
3	=A2.minp(SCORE)	/取出最低分所在记录

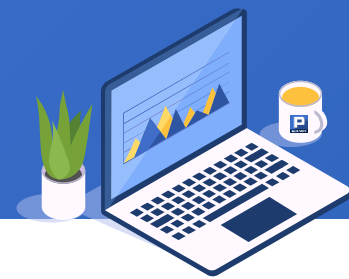
maxp函数和minp函数同样支持@a和@z选项，就不再逐一列举了。

CONTENTS



前N个/后N个

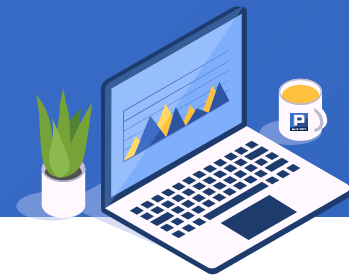
✦ 1. 取值



以成绩表为例，求一班数学前三名的分数。

CLASS	STUDENTID	SUBJECT	SCORE
Class one	1	English	84
Class one	1	Math	77
Class one	1	PE	69
Class one	2	English	81
Class one	2	Math	80
...

✦ 1. 取值



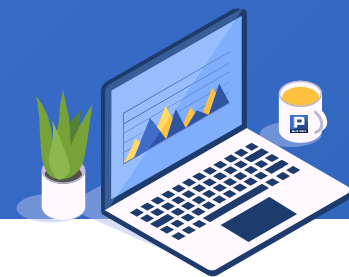
SPL如下，其中用到了top函数，参数N为负数表示从大到小取。

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Scores where SUBJECT='Math' and CLASS='Class one'")	/查询一班的数学成绩
3	=A2.top(-3, SCORE)	/取出前三的分数

这里返回值是得分组成的序列：

Members
97
97
90

✦ 1. 取值



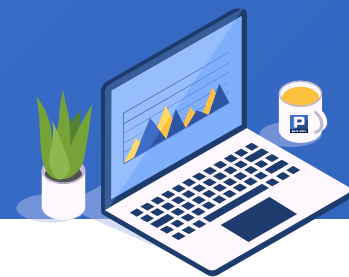
取数学后四名的分数，只要将N设置为4即可，表示从小到大取4名：

	A	B
3	=A2.top(4, SCORE)	/取出后4名的分数

返回值同样是得分组成的序列：

Members
60
60
63
63

✦ 1. 取值



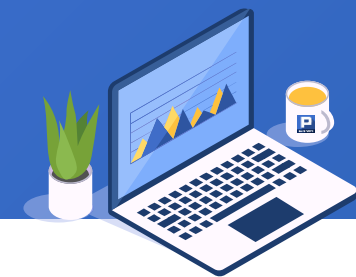
求数学最高分也可以使用top函数，SPL如下：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Scores where SUBJECT='Math' and CLASS='Class one'")	/查询一班的数学成绩
3	=A2.top@1(-1, SCORE)	/取出最高分

当N为±1时，top@1(N, x)函数与max/min函数类似，返回值是数值。

Value
97

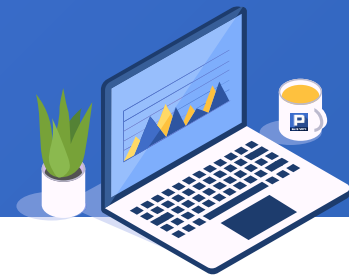
✦ 2. 取对应记录的行号



求2019年股市收盘价最高的三天，交易额相对于前一日的涨幅。

Date	Opening price	Closing price	Amount
2019/12/4	2876.9079	2878.1157	136000000000
2019/12/3	2869.8822	2884.6988	135000000000
2019/12/2	2874.4484	2875.8072	139000000000
2019/11/29	2885.9744	2871.9813	140000000000
2019/11/28	2902.3644	2889.6934	123000000000
...

✦ 2. 取对应记录的行号



我们需要知道股市收盘价最高的三天记录所在行号，再与上一个交易日比较得出结果。SPL如下，其中用到了ptop函数来取最高三天所在行号：

	A	B
1	=file("000001.csv").import@ct()	/导入数据文件
2	=A1.select(year(Date)==2019)	/选出2019年的记录
3	=A2.ptop(-3, 'Closing price')	/取出股市收盘价最高三天所在行号
4	=A3.run(~=A2(~).Amount/A2(~+1).Amount-1)	/循环用每天交易量和前日交易量计算涨幅

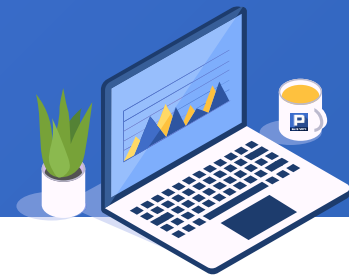
A3结果：

VALUE
154
156
157

A4结果：

VALUE
-0.027777777777777779
-0.013888888888888884
0.011235955056179803

✦ 2. 取对应记录的行号



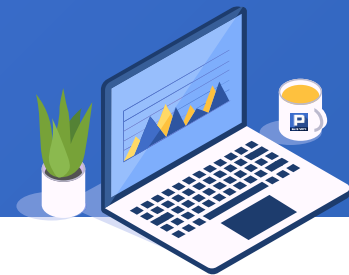
求股市收盘价最低点的两天记录所在行号。

	A	B
1	=file("000001.csv").import@ct()	/导入数据文件
2	=A1.select(YEAR(Date)==2019)	/选出2019年的记录
3	=A2.ptop(2, 'Closing price')	/取出股市收盘价最低两天所在行号

ptop函数默认返回的是行号组成的序列。A3结果如下：

VALUE
224
225

✦ 2. 取对应记录的行号



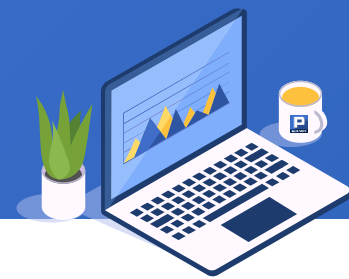
求股市收盘价最高点所在行号也可以使用ptop函数，SPL如下：

	A	B
1	=file("000001.csv").import@ct()	/导入数据文件
2	=A1.select(YEAR(Date)==2019)	/选出2019年的记录
3	=A2.ptop(-1, 'Closing price')	/取出股市收盘价最高点所在行号

当N为±1时，ptop@1(N, x)函数与pmax/pmin函数类似，返回值是数值。

Value
154

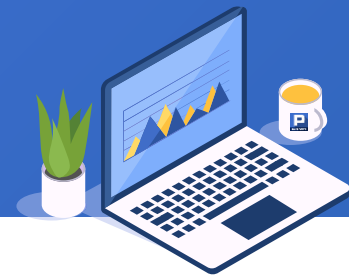
✦ 3. 取对应记录



以成绩表为例，求一班数学前三名的学生信息。

CLASS	STUDENTID	SUBJECT	SCORE
Class one	1	English	84
Class one	1	Math	77
Class one	1	PE	69
Class one	2	English	81
Class one	2	Math	80
...

✦ 3. 取对应记录



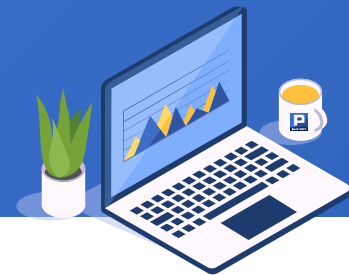
SPL的top(n; x)函数可以取前n条记录。SPL如下：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Scores where SUBJECT='Math' and CLASS='Class one'")	/查询一班的数学成绩
3	=A2.top(-3; SCORE)	/取出前三名的记录

A3结果：

CLASS	STUDENTID	SUBJECT	SCORE
Class one	13	Math	97
Class one	10	Math	97
Class one	7	Math	90

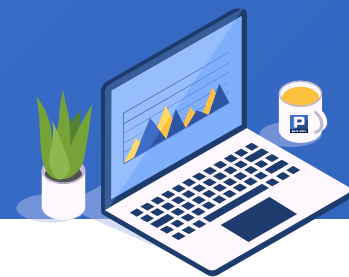
✦ 3. 取对应记录



以登录日志表为例，求某个用户第一次登录的IP地址。

ID	USERID	IP	LOGINTIME
1	1	37.17.184.11	2012/05/09 09:01:10
2	3	61.134.201.1	2012/05/09 09:02:43
3	7	124.114.171.101	2012/05/09 09:03:18
4	2	183.202.48.25	2012/05/09 09:05:15
5	15	1.24.216.5	2012/05/09 09:05:55
...

✦ 3. 取对应记录



oracle的keep函数可以实现这个需求:

```
select
    min(IP) keep (dense_rank first order by LOGINTIME) IP
from
    LoginLog
where
    USERID=8
```

执行结果:

IP
223.223.118.1

✦ 3. 取对应记录



其实我们关心的就是用户第一次登录的那条记录，SPL的top(n; x)函数可以更清晰的实现这个需求。SPL如下：

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from LoginLog where USERID=8")	/查询用户ID是8的登录记录
3	=A2.top(1; LOGINTIME)	/取出第一次登录的记录
4	=A3.IP	/从记录中取IP字段值

A3结果：

ID	USERID	IP	LOGINTIME
8	8	223.223.118.1	2012/05/09 09:07:17

A4结果：

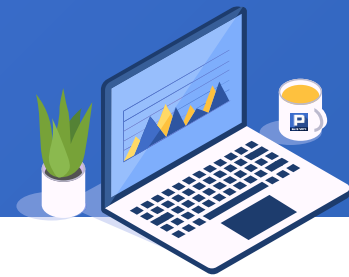
IP
223.223.118.1

CONTENTS



分组中的使用

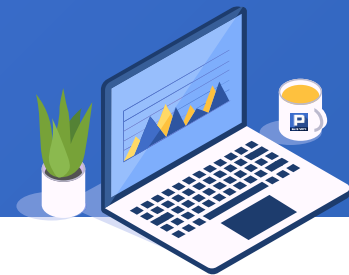
✦ 1. 子集进行top类运算



以成绩表为例，求各班每科前两名。

CLASS	STUDENTID	SUBJECT	SCORE
Class one	1	English	84
Class one	1	Math	77
Class one	1	PE	69
Class one	2	English	81
Class one	2	Math	80
...

✦ 1. 子集进行top类运算



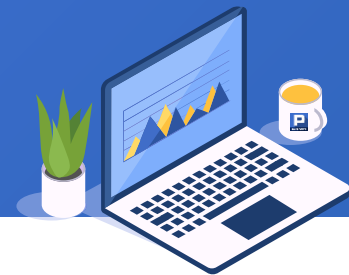
SPL如下:

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from Scores")	/查询学生成绩
3	=A2.group(CLASS,SUBJECT;~.top(-2;SCORE):TOP2)	/按班级和学科分组并取出每组分数前两名
4	=A3.(TOP2).conj()	/将所有班级各科的前两名记录合并

A4结果:

CLASS	STUDENTID	SUBJECT	SCORE
Class one	4	English	96
Class one	9	English	93
Class one	13	Math	97
Class one	10	Math	97
...

◆ 1. 子集进行top类运算



A3: 使用group函数按照班级分组，分组时把topN看成是一种聚合运算，对每个子集取前两名。

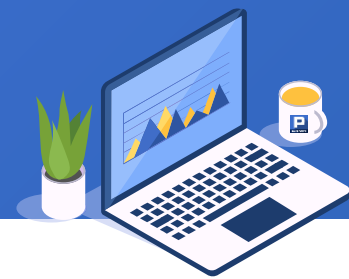
	A	B
3	=A2.group(CLASS;~.top(-2;SCORE):TOP2)	/按班级分组并取出每组分数前两名

A3结果:

CLASS	SUBJECT	Members
Class one	English	[[Class one,4,English,96],[Class one,9,English,93]]
Class one	Math	[[Class one,13,Math,97],[Class one,10,Math,97]]
...

CLASS	STUDENTID	SUBJECT	SCORE
Class one	13	Math	97
Class one	10	Math	97

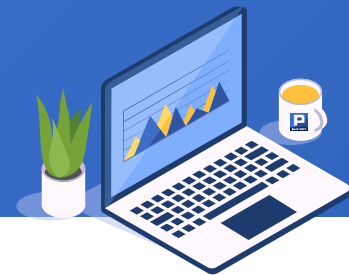
✦ 2. 以累计方式进行top类运算



以雇员表为例，求每个部门薪水前三高的员工信息。

EID	NAME	DEPT	SALARY
1	Rebecca	R&D	7000
2	Ashley	Finance	11000
3	Rachel	Sales	9000
4	Emily	HR	7000
5	Ryan	R&D	13000
...

✦ 2. 以累计方式进行top类运算



SPL如下:

	A	B
1	=connect("db")	/连接数据库
2	=A1.query("select * from EMPLOYEE")	/查询雇员表
3	=A2.groups(DEPT; top(-3;SALARY):TopSalary)	/按部门分组并取出每组薪水前三的记录
4	=A3.(TopSalary).conj()	/把各部门薪水前三的记录合并

A4结果:

EID	NAME	DEPT	SALARY
20	Alexis	Administration	16000
42	Michael	Administration	12000
18	Jonathan	Administration	7000
2	Ashley	Finance	11000
32	Andrew	Finance	11000
...

A.groups()函数用累计方式计算，不产生分组子集。

THANKS

感谢观看

