

游标





目录

CONTENTS

- 1 延迟计算
- 2 并行游标
- 3 管道
- 4 有序游标
- 5 程序游标

01

延迟计算

例：筛选出年龄大于40岁的员工

	A
1	=file("E:/txt/Employees.txt").import@t()
2	=A1.select(age(Birthday)>=40)

A1~A2结果

立即计算

Index	ID	Name	Gender	Post	Birthday	AccountNo	BasePay
1	1	Mike	Female	Sale	1968-12-0...	536936891...	5600.0
2	2	Jake	Male	Vice Presid...	1962-02-1...	964107677...	2500.0
3	3	Lucy	Female	Sale	1973-08-3...	665248245...	10800.0
4	4	Andy	Male	Sales Man...	1968-09-1...	650028860...	7500.0
5	5	Jim	Male	Sales Man...	1965-03-0...	441380247...	4700.0

Index	ID	Name	Gender	Post	Birthday	AccountNo	BasePay
1	1	Mike	Female	Sale	1968-12-0...	536936891...	5600.0
2	2	Jake	Male	Vice Presid...	1962-02-1...	964107677...	2500.0
3	3	Lucy	Female	Sale	1973-08-3...	665248245...	10800.0
4	4	Andy	Male	Sales Man...	1968-09-1...	650028860...	7500.0
5	5	Jim	Male	Sales Man...	1965-03-0...	441380247...	4700.0

	A	B
1	=file("E:/txt/Employees.txt").cursor@t()	/生成外存文件游标
2	=A1.select(age(Birthday)>=40)	/游标中附加select计算
3	=A2.fetch()	/根据条件取数

A1~A3结果

延迟计算

Value	
com.raqsoft.dm.cursor.FileCursor@229db918	

Value	
com.raqsoft.dm.cursor.FileCursor@229db918	

Index	ID	Name	Gender	Post	Birthday	AccountNo	BasePay
1	1	Mike	Female	Sale	1968-12-0...	536936891...	5600.0
2	2	Jake	Male	Vice Presid...	1962-02-1...	964107677...	2500.0
3	3	Lucy	Female	Sale	1973-08-3...	665248245...	10800.0
4	4	Andy	Male	Sales Man...	1968-09-1...	650028860...	7500.0
5	5	Jim	Male	Sales Man...	1965-03-0...	441380247...	4700.0

例：计算公司40岁以上和40岁以下员工的平均工资

	A
1	=file("E:/txt/Employees.txt").import@t()
2	=A1.groups(age(Birthday)>=40;ifgt40;avg(BasePay))

A1~A2结果

立即计算

Index	ID	Name	Gender	Post	Birthday	AccountNo	BasePay
1	1	Mike	Female	Sale	1968-12-0...	536936891...	5600.0
2	2	Jake	Male	Vice Presid...	1962-02-1...	964107677...	2500.0
3	3	Lucy	Female	Sale	1973-08-3...	665248245...	10800.0
4	4	Andy	Male	Sales Man...	1968-09-1...	650028860...	7500.0
5	5	Jim	Male	Sales Man...	1965-03-0...	441380247...	4700.0

Index	ifgt40	avg(BasePay)
1	false	6940.0
2	true	6007.692307692308

	A
1	=file("E:/txt/Employees.txt").cursor@t()
2	=A1.groups(age(Birthday)>=40;ifgt40;avg(BasePay))

A1~A2结果

立即计算

Value
com.raqsoft.dm.cursor.FileCursor@7d632866

Index	ifgt40	avg(BasePay)
1	false	6940.0
2	true	6007.692307692308

物品信息和销售信息存储在两张表中，请计算总的销售额

多级游标

	A	B
1	=file("E:/txt/Products.txt").import@t().primary@i(ID)	/读入商品列表并建立索引
2	=file("E:/txt/Sales.txt").cursor@t()	/建立游标，准备遍历
3	=A2.select(quantity<=10)	/过滤，仍返回游标
4	=A3.switch(productid,A1:ID)	/建立连接指针，仍返回游标
5	=A4.groups(;sum(quantity*productid.Price):total)	/求和汇总

} 延迟计算
立即计算

A1~A4结果

Index	ID	Name	Category	Price
1	1	Apple juice	Low-end	18.0
2	2	Mile	Low-end	19.0
3	3	Tomato sa...	Low-end	10.0
4	4	Salt	Low-end	22.0
5	5	Sesame oil	Low-end	21.35

Value
com.raqsoft.dm.cursor.FileCursor@249b5d3b

Value
com.raqsoft.dm.cursor.FileCursor@249b5d3b

Index	total
1	142740.18000000008

注意：游标只遍历一次，不可以重复执行取数操作

02

并行游标

从产品销售记录中找到每年1月份的销售记录

并行游标

	A	单游标
1	=now()	
2	=file("E:/txt/PRODUCT_SALE.txt").cursor@t()	
3	=A2.select(month(Date)=1)	
4	=A3.fetch()	
5	=interval@ms(A1,now())	
		Value 9022

	A	B	多路游标
1	=now()		
2	=file("E:/txt/PRODUCT_SALE.txt").cursor@mt()		
3	fork A2	=A3.select(month(Date)=1)	
4		=B3.fetch()	
5	=A3.conj()	=interval@ms(A1,now())	
		Value 5950	

结果

Index	ID	PID	DATE	QUANTITY	SID
1	1211	10075052	2010-01-01	84	10225
2	2474	10098045	2010-01-01	106	10591
3	10576	10093980	2010-01-01	53	10720
4	12938	10069598	2010-01-01	30	10483
5	15091	10067138	2010-01-01	104	11000

	A	B	非多路游标
1	=now()		
2	=file("E:/txt/PRODUCT_SALE.txt").cursor@mt()		
3	fork A2	=A3.select(month(Date)=1)	
4	=A3.(~.fetch())		
5	=A4.conj()	=interval@ms(A1,now())	
		Value 8659	

注意：如果在fork中没有执行立即计算，则多线程仅仅是用于定义游标，起不到提速的作用，把A3的fork换成for后的运算时间是一样的

可以使用简单的方法实现多路游标并行

	A
1	=now()
2	=file("E:/txt/PRODUCT_SALE.txt").cursor@mt()
3	=A2.select(month(Date)=1)
4	=A3.fetch()
5	=interval@ms(A1,now())
	<div>值 5089</div>

多路游标
读取数据和计算同时并行

	A
1	=now()
2	=file("E:/txt/PRODUCT_SALE.txt").cursor@t().mcursor()
3	=A2.select(month(Date)=1)
4	=A3.fetch()
5	=interval@ms(A1,now())
	<div>值 8183</div>

多路游标
读取数据单线程，计算多线程

Index	ID	PID	DATE	QUANTITY	SID
1	1211	10075052	2010-01-01	84	10225
2	2474	10098045	2010-01-01	106	10591
3	10576	10093980	2010-01-01	53	10720
4	12938	10069598	2010-01-01	30	10483
5	15091	10067138	2010-01-01	104	11000

各部门的员工信息分别存储在一份文件中。

从各部门的员工信息中，找到纽约州的员工（多个文件的游标组成的多路游标）

Administrationinfo.txt
Financeinfo.txt
HRinfo.txt
Marketinginfo.txt
Productioninfo.txt
R&Dinfo.txt
Salesinfo.txt
Technologyinfo.txt

	A	B
1	=directory@p("E:/txt/employee_dept")	/各个部门员工信息的绝对路径
2	=A1.(file(~).cursor@t()).mcursor()	/生成游标序列，并创建多路游标
3	=A2.select(STATE=="New York").fetch()	/多路游标取数

A1~A3结果

Index	Member
1	E:\txt\employee_dept\Administrationinfo.txt
2	E:\txt\employee_dept\Financeinfo.txt
3	E:\txt\employee_dept\HRinfo.txt
4	E:\txt\employee_dept\Marketinginfo.txt
5	E:\txt\employee_dept\Productioninfo.txt
6	E:\txt\employee_dept\R&Dinfo.txt
7	E:\txt\employee_dept\Salesinfo.txt
8	E:\txt\employee_dept\Technologyinfo.txt

Value
com.raqsoft.dm.cursor.MultipathCursors@72b83842

Index	EID	NAME	SURNAME	GENDER	STATE	BIRTHDAY	HIREDATE	DEPT	SALARY
1	2	Ashley	Wilson	F	New York	1980-07-19	2008-03-16	Finance	11000
2	220	Caleb	Smith	M	New York	1976-03-16	2008-01-01	Finance	7000
3	221	Sarah	Davis	F	New York	1982-09-04	2007-03-01	Finance	5000
4	229	Zachary	Taylor	M	New York	1984-09-12	2004-03-01	Finance	7000
5	180	Abigail	Smith	F	New York	1972-09-19	2007-05-01	HR	5000
6	25	Sarah	Davis	F	New York	1976-11-25	2006-11-25	Marketing	12000
7	46	Alexander	Johnson	M	New York	1978-08-20	2008-08-20	Marketing	10000
8	262	Mary	Jackson	F	New York	1987-06-11	2008-03-01	Marketing	6500
9	273	Emma	Williams	F	New York	1984-02-07	2007-11-01	Marketing	10000
10	446	Michael	Johnson	M	New York	1984-07-02	2005-08-01	Marketing	5000

03

管道

游标只能执行一次遍历，但有时我们需要同时计算多个数据，这时就需要将数据压入管道减少遍历次数

例：借贷业务中，2200-01-01定义为逾期，用逾期次数/总次数作为分数，分别为user和listing打分

	A	B
1	=file("E:/txt/user_repay_logs.csv").cursor@tc()	
2	cursor A1	=A2.groups(user_id;count(repay_date==date("2200-01-01"))/count(~):user_score)
3	cursor	=A3.groups(listing_id;count(repay_date==date("2200-01-01"))/count(~):listing_score)
...	...	
n	cursor	可以定义其他管道

A2~A3结果

Index	user_id	user_score
1	1	0.0
2	2	0.0
3	3	0.05

Index	listing_id	listing_score
242110	242110	0.0
242111	242111	0.33333333333333...
242112	242112	0.0

大数据条件下计算中位数

	A	B	
1	=file("E:/txt/CD5-0201FIC10202_2017.csv").cursor@tc()		/创建游标
3	cursor A1	=A2.sortx(Value)	/管道中按照Value排序
4	cursor	=A3.total(count(~))	/另一个管道统计总数
5	=A2.skip((A3-1)\2)		/从排序后的游标中跳过前半
6	=A2.fetch@x(2-A3%2).avg(Value)		/取出中位数

04

有序游标

计算各用户2018年每个月份的在线时长增长率（数据按UID有序）

有序游标

	A	B	C
1	=file("E:/txt/UseLogs.txt").cursor@t()		
2	for A1;UID	=A2.groups(UID,month(LOGIN):m_login;sum(SECOND):total_time)	
3		=to(12)\B2.(m_login)	
4		for B3	> B2.insert(B4,A2.UID,B4,0)
5		=@ B2.derive(if(m_login==1 total_time[-1]==0,null,total_time/total_time[-1]-1):raise_rate)	

循环对UID有序的游标，每次取出相同的用户，对其进行计算，如果用户某个月没登陆则将total_time置0
如果是第一个月或者前一个月未登陆，增长率为null

B5结果

Index	UID	m_login	total_time	raise_rate
1	10000001	1	1312198	(null)
2	10000001	2	984155	-0.24999504647926607
3	10000001	3	1241612	0.2616020850374179
4	10000001	4	1469796	0.18378044026636342
5	10000001	5	1551339	0.055479127715682974
6	10000001	6	1015243	-0.3455698593279741
7	10000001	7	1269041	0.2499874414302783
8	10000001	8	931809	-0.26573767120211245
9	10000001	9	1408993	0.5121049485463223
10	10000001	10	897755	-0.36283927599356425
11	10000001	11	1659301	0.8482782050782229
12	10000001	12	1488339	-0.10303254201618628
13	10000002	1	511107	(null)
14	10000002	2	700521	0.3705955895732205

计算所有股票最高收盘价当天的增长率(数据按stockid有序)

有序游标

	A	B
1	=file("E:/txt/Stock_Price.txt").cursor@t()	
2	=create(stockid,max_price_raise_rate)	
3	for A1;stockid	=A3.pmax(CL)
4		=A3.calc(B3,if(day(DT)==1,A3.CL,A3.CL/A3.CL[-1]-1))
5		=A2.insert(0,A3.stockid,B4)

循环对stockid有序的游标，每次取出一组stockid，从中找到最高收盘价的位置，利用此位置定位计算，当天的增长率，并把结果填入到A2创建的序表中

A3结果

Index	stockid	max_price_raise_rate
1	1001	0.09691629955947145
2	1026	0.04958677685950419
3	1028	0.015592077538980176
4	1070	0.07007203667321549
5	1107	0.09358288770053469
6	1134	0.028019925280199365
7	1137	0.05385810460901075
8	1147	0.0014598540145984717
9	1206	14.01
10	1213	40.94

邮件信息如下：每一份邮件都以RECIPIENT开头，分为收件人，发件人和内容三部分，
请将邮件整理成结构化数据

RECIPIENT:730284595@xx.xx

SENDADDRESS:106383734@xx.xx

CONTENT: Harry Potter and the Sorcerer's Stone

CHAPTER ONE

THE BOY WHO LIVED

Mr. and Mrs. Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you very much. They were the last people you'd expect to be involved in anything strange or mysterious, because they just didn't hold with such nonsense.

Mr. Dursley was the director of a firm called Grunnings, which made drills. He was a big, beefy man with hardly any neck, although he did have a very large mustache. Mrs. Dursley was thin and blonde and had nearly twice the usual amount of neck, which came in very useful as she spent so much of her time craning over garden fences, spying on the neighbors. The Dursleys had a small son called Dudley and in their opinion there was no finer boy anywhere.

The Dursleys had everything they wanted, but they also had a secret, and their greatest fear was that somebody would discover it. They didn't think they could bear it if anyone found out about the Potters Potter was Mrs Dursley's sister, but they hadn't met for several years; in fact, Mrs Dursley pretended she didn't have a sister, because her sister and her good-for-nothing husband were as unDursleyish as it was possible to be. The Dursleys shuddered to think what the neighbors would say if the Potters arrived in the street. The Dursleys knew that the Potters had a small son, too, but they had never even seen him. This boy was another good reason for keeping the Potters away; they didn't want Dudley mixing with a child like that

When Mr

RECIPIENT:717308235@xx.xx

SENDADDRESS:853474825@xx.xx

CONTENT:

When Mr Dursley woke up on the dull, gray Tuesday our story starts, there was nothing about the cloudy sky outside to suggest that strange and mysterious things would soon be happening all over the country Dursley hummed as he picked out his most boring tie for work, and Mrs Dursley gossiped away happily as she wrestled a screaming Dudley into his high chair

None of them noticed a large, tawny owl flutter past the window

At half past eight, Mr

RECIPIENT:206817811@xx.xx

SENDADDRESS:839801895@xx.xx

CONTENT:

	A	B	
1	=file("E:/txt/email.txt").cursor()		
2	=create(RECIPIENT,SENDADDRESS,CONTENT)		
3	for A1;_1.split(":")(1)=="RECIPIENT"	=A3._1(1).split(":")	
4		=A3._1(2).split(":")	
5		=A3._1(3).split(":")	
6		=B5.m(2:).concat(":")+A3._1.m(4:).concat()	
7		=A2.insert(0,B3(2),B4(2),B6)	
8		if A2.len()==500	=file("E:/txt/email_pre.txt").export@at(A2)
9			>A2.reset()
10	=file("E:/txt/email_pre.txt").export@at(A2)		
11	=file("E:/txt/email_pre.txt").cursor@t().fetch@x(10)	/读入查看前10行结果	

Index	RECIPIENT	SENDADDRESS	CONTENT
1	730284595@xx.xx	106383734@xx.xx	Harry Potter an...
2	717308235@xx.xx	853474825@xx.xx	When MrDursl...
3	206817811@xx.xx	839801895@xx.xx	At half past eig...
4	983837696@xx.xx	1034188652@xx.xx	He got into his car ...
5	860688317@xx.xx	973356725@xx.xx	They were whisperi...
6	239022673@xx.xx	579899833@xx.xx	He didn't know why...
7	746677348@xx.xx	1034632680@xx.xx	He looked back at t...
8	264687557@xx.xx	351534290@xx.xx	It was a few secon...
9	352747334@xx.xx	348118929@xx.xx	Going to be any mo...
10	533754054@xx.xx	1057123562@xx.xx	"Owls... shooting st...

A11结果

每封邮件都是以RECIPIENT开头的，循环游标，当以 “: ” 分割后的第一部分为 “RECIPIENT” 时开始读取游标直到下次又遇到 “RECIPIENT” 。这样每循环一次代表了一封邮件，在for循环中处理这一封邮件即可，当等于500封时，导出数据，并将序表清空。

05

程序游标

前面介绍的多级游标很方便，但有时会遇到某一步计算时较复杂，单一函数无法完成计算，这时使用程序游标，返回中间结果，更加容易理解和日后维护。

例：整理用户信息并筛选出纽约州的用户，用户信息如下：

1	10308583	F	2007-02-07	
2018-03-11	Michigan	10308583_***@mail.xxx	acappella	
2	10902344	F	2011-02-17	
2018-02-25	Ohio	10902344_***@mail.xxx	big-beat	
3	10550284	F	2010-12-26	
2018-10-23	Illinois	10550284_***@mail.xxx	acappella	
4	10719361	M	2003-01-15	
2018-07-01	Missouri	10719361_***@mail.xxx	new age	
5	10329553	F	2015-09-21	
2018-08-08	Texas	10329553_***@mail.xxx	ambient	
6	10321518	F	2015-11-28	
2018-07-07	Pennsylvania	10321518_***@mail.xxx	trip-hop	

用户信息分两行，需要先将用户信息整理成结构化数据

先整理用户信息，然后导出成文件，再用游标读取找出纽约州的用户

程序游标

	A	B
1	=file("E:/txt/user_info.txt").cursor()	
2	for A1,5000*2	=A2.group((#-1)\2)
3		=B2.(~.(~.array()).conj())
4		=B3.new(~(1):ID,~(2):UID,~(3):GENDER,~(4):RDATE,~(5):LDATE,~(6):STATE,~(7):EMAIL,~(8):STYLE)
5		=file("E:/txt/user_info_pre.txt").export@at(B4)
6	=file("E:/txt/user_info_pre.txt").cursor@t()	
7	=A6.select(STATE=="New York")	
8	=A7.fetch()	

循环游标将数据整理成结构化数据，分批导出追加到同一个文件。 创建游标，附加select运算，得到最终结果

最后一次循环结束后， B2~B4结果

B2

Index	Member			
1	[[95001,10085776,M,...],[2018-10-25,Florida,10085776_***@mail.xxx,...]]			
2	[[95002,10522585,F,...],[2018-11-16,Texas,10522585_***@mail.xxx,...]]			
3	[[95003,10073788,M,...],[2018-07-31,Florida,10073788_***@mail.xxx,...]]			

Index	_1	_2	_3	_4
1	95001	10085776	M	2006-01-18
2	2018-10-25	Florida	10085776_***...	big-beat

B3

Index	Member	
1	[95001,10085776,M,...]	
2	[95002,10522585,F,...]	
3	[95003,10073788,M,...]	

B4

Index	ID	UID	GENDER	RDATE	LDATE	STATE	EMAIL	STYLE
1	95001	10085776	M	2006-01-18	2018-10-25	Florida	10085776...	big-beat
2	95002	10522585	F	2007-03-03	2018-11-16	Texas	10522585...	brit-hop
3	95003	10073788	M	2005-07-03	2018-07-31	Florida	10073788...	jungle

	A	B
1	=file("E:/txt/user_info.txt").cursor()	
2	for A1,5000*2	=A2.group((#-1)\2)
3		=B2.(~.(~.array()).conj())
4		=B3.new(~(1):ID,~(2):UID,~(3):GENDER,~(4):RDATE,~(5):LDATE,~(6):STATE,~(7):EMAIL,~(8):STYLE)
5		return B4

B4结果

Index	ID	UID	GENDER	RDATE	LDATE	STATE	EMAIL	STYLE
1	95001	10085776	M	2006-01-18	2018-10-25	Florida	10085776...	big-beat
2	95002	10522585	F	2007-03-03	2018-11-16	Texas	10522585...	brit-hop
3	95003	10073788	M	2005-07-03	2018-07-31	Florida	10073788...	jungle

	A	B
1	=cursor("E:/esproc_test/cursor/prepare_user_data.dfx")	/调用数据处理程序
2	=A1.select(STATE=="New York")	/筛选纽约州的用户
3	=A2.fetch()	/取出数据

A3结果

Index	ID	UID	GENDER	RDATE	LDATE	STATE	EMAIL	STYLE
1	1	10182061	M	2013-09-26	2018-03-11	New York	10182061...	big-beat
2	20	10325095	M	2010-03-27	2018-09-26	New York	10325095...	acappella
3	29	10533676	M	2011-02-01	2018-03-28	New York	10533676...	trip-hop
4	45	10023684	F	2005-05-12	2018-08-23	New York	10023684...	brit-hop
5	54	10052401	M	2011-08-25	2018-08-14	New York	10052401...	ambient

也可以使用func的方式在同一个脚本调用

	A	B	C
1	=file("E:/txt/user_info.txt").cursor()		
2	func	for A2,5000*2	=B2.group((#-1)\2)
3			=C2.(~.(~.array()).conj())
4			=C3.new(~(1):ID,~(2):UID,~(3):GENDER,~(4):RDATE,~(5):LDATE,~(6):STATE,~(7):EMAIL,~(8):STYLE)
5			return C4
6	=cursor@c(A2,A1)		
7	=A7.select(STATE=="New York")		
8	=A8.fetch()		

A9结果

Index	ID	UID	GENDER	RDATE	LDATE	STATE	EMAIL	STYLE
1	1	10182061	<u>M</u>	2013-09-26	2018-03-11	<u>New York</u>	10182061...	<u>big-beat</u>
2	20	10325095	<u>M</u>	2010-03-27	2018-09-26	<u>New York</u>	10325095...	<u>acappella</u>
3	29	10533676	<u>M</u>	2011-02-01	2018-03-28	<u>New York</u>	10533676...	<u>trip-hop</u>
4	45	10023684	<u>F</u>	2005-05-12	2018-08-23	<u>New York</u>	10023684...	<u>brit-hop</u>
5	54	10052401	M	2011-08-25	2018-08-14	New York	10052401...	ambient

例：订单表原来已按时间排序，需要将数据按日期、产品去除重复，再统计记录条数（groupx）

	A	B
1	=now()	
2	=file("E:/txt/PRODUCT_SALE.txt").cursor@t()	/创建游标
3	=A2.groupx(date(DATE),PID)	/去重
4	=A3.skip()	/统计记录条数
5	=interval@ms(A1,now())	

A4、A5结果

Value
9849397

Value
55647

数据按DATE有序，否则不可以这么做

	A	B	C
1	=now()		
2	=file("E:/txt/PRODUCT_SALE.txt").cursor@t()		
3	func	for A3;DATE	=B3.id(PID)
4			return C3
5	=cursor@c(A3,A2)		
6	=A5.skip()		
7	=interval@ms(A1,now())		

A3、A4结果

Value
9849397

Value
12767

THANKS

感谢聆听 批评指导