User exploration

1 Content

Before running the sample codes in this document, establish a connection to the system data source of esProc.

2 Basic Functions

2.1 Sequence

2.1.1 Constant and Sequence

	A	B	C
1	5	esProc	2012-1-1
2	'C:\files		
3	[1,3,5,7,9]	/Define ISeq	
4	[Apple,1.69,LB,1.2]	/Define normal sequence	
5	=[A1,A2,3*3.14]	/Use expression to compute sequence	
6	=[A1:C1]	/ Use expression to define sequence	

Of **A5** and **A6** the computations are as follows:

Member
5
C:\files
9.42

Member
5
esProc
2012-01-01

2.1.2 ISeq and Function to( )

	A	B
1	=to(1,6)	/From 1 to 6
2	=to(5)	/From 1 to 5, the expression can be represented in short
3	=to(7,3)	/From 7 to 3
4	=to@s(7,3)	/The 3 numbers starting from 7 rightward
5	=to@s(7,-3)	/The 3 numbers starting from 7 leftward

Of **A1~A5** the computations are as follows:

Member
1
2
3
4
5
6

Member
1
2
3
4
5

Member
7
6
5
4
3

Member
7
8
9

Member
7
6
5

2.1.3 Set Operation

	A	B
1	[5,3,a]	[b,3,5]
2	=A1|B1	/Sum columns of A1and B1
3	=A1&B1	/Union columns of A1and B1
4	=A1^B1	/Intersect columns of A1 and B1
5	=A1\B1	/Subtract columns of A1and B1
6	=B1*2	/ Complement column of B1 and B2
7	=[1]*4	/Complement column of [1] and 4

Of **A2~A7** the computations are as follows:

Member
5
3
a
b
3
5

Member
5
3
a
b

Member
5
3

Member
a

Member
b
3
5
b
3
5

Member
1
1
1
1

2.1.4 Access to Member

	A	B
1	[1,3,7,null,2,3]	
2	=A1(3)	/Get the 3rd member in the sequence
3	=A1.m(-2)	/Get the next to the last member in the sequence
4	=A1.m@r(9)	/Get the 9th member in the sequence, out-of-range loop

5	=A1.pos(3)	/The first position the 3 appears in the sequence
6	=A1.pos@a(3)	/All positions at which the 3 appears
7	=A1.pos(4)	/The position at which the 4 appears in the sequence. If no member, then return 0

Of **A2~A7** the computations are as follows:

Value	Value	Value	Value	Member	Value
7	2	7	2	2	0
				6	

2.1.5 Subsequence

	A	B
1	[1,3,7,null,2,3]	
2	=A1([1,3,2])	/Get subsequence from sequence according to ISeq
3	=A1.m([1,-2,3])	/Get subsequence from sequence according to ISeq. The ISeq could be negative.
4	=A1.rvs()	/Reverse sequence
5	=A1.dup()	/Duplicate sequence

Of **A2~A5** the computations are as follows:

Member	Member	Member	Member
		3	1
		2	3
			7
1	1	7	
7	2	3	2
3	7	1	3

2.1.6 Converge Computation

	A	B
1	[1,3,7,null,2,3]	
2	=A1.len()	/Length of sequence, that is, all members
3	=A1.count()	/Number of non-null members in the sequence
4	=A1.sum()	/Compute sum of sequence
5	=A1.avg()	/Compute average on sequence
6	=A1.max()	/Maximum value in the sequence
7	=A1.min()	/Minimum value in the sequence
8	=A1.rank()	/Sequence composed of respective ranks of each member

Of **A2~A8** the computations are as follows:

Value	Value	Value	Value	Value	Value	Member
6	5	16	3.2	7	1	5
						2
						1
						6
						4
						2

2.1.7 Sequence and String

	A	B
1	[a,2,[b,4]]	
2	'a,2,[b,4]	/Start with single quotes, and the value is string constant after the single quotes
3	=A1.string(",")	/Convert sequence to string, of which the member of string must be double quoted
4	=A1.string@d(",")	/Convert sequence to string. The double quotes is not required for the member of string
5	=A2.array(",")	/Convert the string to the sequence
6	=A2.array@s(",")	/Convert string to sequence simply, and no further parsing on members.
7	=A2.array@1(",")	/Get the 1st separator, and convert the string to a sequence of 2 members

Of **A3~A7** the computation results are as follows:

Value
"a",2,["b",4]

Value
a,2,[b,4]

Member
a
2
[b,4]

Member
a
2
[b,4]

Member
a
2,[b,4]

Among the results, the member in blue is a sequence for you to double click and view.

2.2 TSeq

- **Generate TSeq**
- **Converge Computation**
- **Index**
- **Filter**
- **Compute and Create New TSeq**
- **Sort**
- **Computed Column and Computed Field**

2.2.1 Generate TSeq

	A	B
1	=file@s("students.txt").import@t()	/Import text file as TSeq, and the 1st row is the column name
2	=demo.query("select * from STUDENTS")	/Execute SQL statement in the esProc database, and get the TSeq
3	=demo.query("select * from STUDENTS where AGE=?",17)	/Execute SQL statement with parameters in database esProc to get the TSeq
4	Apple	1.69
5	Banana	0.69

6	Peach	0.88
7	=create(Name,UnitPrice).record([A4:B6])	/Create TSeq, specify the column name, and populate it with data from cells

Of **A1** the TSeq is as follows:

ID	Name	Gender	Age
1	Emily	F	17
2	Elizabeth	F	17
6	Zachary	M	19
8	Megan	F	16

Of **A2** the TSeq is as follows:

ID	NAME	GENDER	AGE
1	Emily	F	17
2	Elizabeth	F	16
3	Sean	M	17
4	Lauren	F	15
5	Michael	M	16
6	John	M	13
7	Nicholas	M	16

Of **A3** the TSeq is as follows:

ID	NAME	GENDER	AGE
1	Emily	F	17
3	Sean	M	17

Of **A7** the TSeq is as follows:

Name	UnitPrice
Apple	1.69
Banana	0.69
Peach	0.88

2.2.2 Converge Computation

	A	B
1	=demo.query("select * from STATES")	/ get the TSeq from the esProc database
2	=A1.count()	/ Count number of records
3	=A1.sum(POPULATION)	/Sum up the population field values of all records
4	=round(A1.avg(POPULATION))	/Compute the average of population in each state
5	=A1.max(POPULATION/A3)	/Compute the state whose population takes the greatest proportion
6	=A1.min(CAPITAL)	/ Get the first state capital ranked by its name
7	=A1.rank(POPULATION)	/Get the rankings of states sorted by its populations

Of **A2~A7** the computation results are as follows:

Value	Value	Value	Value	Value
51	308745538	6053834.0	0.12066233	Albany

Member
23
47
16
32

2.2.3 Index

	A	B
1	=demo.query("select NAME as STATE, ABBR, CAPITAL, POPULATION from STATES")	/ get the TSeq from the esProc database
2	=A1.pselect(left(STATE,1)=="N")	/The record No of the first state whose initial is N
3	=A1.pselect@a(left(STATE,1)=="N")	/The respective record No of all states whose initial is N
4	=A1.pmax(POPULATION)	/Record no of the largest state
5	=A1.pmin(POPULATION)	/ Record no. of the state with the least population
6	=A1(5)	/ State Record whose serial number of record is 5

Of **A2~A6** the computation results are as follows:

Value
27

Member
27
28
29
30

Value
5

Value
50

STATE	ABBR	CAPITAL	POPULATION
California	CA	Sacramento	37253956

2.2.4 Filter

	A	B
1	=demo.query("select NAME as STATE, ABBR, CAPITAL, POPULATION from STATES")	/ get the TSeq from the esProc database
2	=A1.select(left(STATE,1)=="A")	/All states whose initials are A
3	=A1.select@1(POPULATION>6000000)	/The first state whose population exceeds 6 million
4	[M,N,W]	
5	=A1.select(POPULATION>6000000 && A4.pos(left(STATE,1))>0)	/The states whose population exceed 6 million and initials is M, N, or W respectively

Of **A2** the computation results is as follows:

STATE	ABBR	CAPITAL	POPULATION
Alabama	AL	Montgomery	4779736
Alaska	AK	Juneau	710231
Arizona	AZ	Phoenix	6392017
Arkansas	AR	Little Rock	2915918

Of **A3** the computation results is as follows:

STATE	ABBR	CAPITAL	POPULATION
Arizona	AZ	Phoenix	6392017

Of **A5** the computation result is as follows:

STATE	ABBR	CAPITAL	POPULATION
Massachusetts	MA	Boston	6547629
Michigan	MI	Lansing	9883640
New Jersey	NJ	Trenton	8791894
New York	NY	Albany	19378102
North Carolina	NC	Raleigh	9535483
Washington	WA	Olympia	6724540

2.2.5 Compute and Create New TSeq

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	/ get the TSeq from the esProc database
2	=A1.new(FULLNAME:EName ,if (GENDER =="M", "Male","Female"):Gender)	/Get name and gender to create TSeq. The name field is EName by name, and the abbreviation is not used for the gender.
3	=A1.new(EID, FULLNAME, interval@y(BIRTHDAY,now()): Age)	/Get employee no., name, and compute the employee age, and then create TSeq

Of **A2** and **A3** the computation results are as follows:

EName	Gender
Rebecca Moore	Female
Ashley Wilson	Female
Rachel Johnson	Female
Emily Smith	Female
Ashley Smith	Female

EID	FULLNAME	Age
1	Rebecca Moore	39
2	Ashley Wilson	33
3	Rachel Johnson	43
4	Emily Smith	28
5	Ashley Smith	28

2.2.6 Sort

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ").derive(interval@y(BIRTHDAY,now()): Age)	/ get the TSeq from the esProc database
2	=A1.sort(STATE)	/Sort the employee by its state in ascending order
3	=A1.sort(SALARY:-1)	/Sort the employee by its salary in descending order

4	=A1.sort(SALARY:-1, FULLNAME)	/Sort the employee by its salary in descending order. Then, sort the employee of the same age by its name in ascending order

Of **A2**the computation results are as follows:

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY	Age
102	Christian Smith	M	Alabama	1972-07-2	12000	41
282	Kayla Davis	F	Alabama	1970-09-0	6500	43
419	Sara Davis	F	Alabama	1984-12-3	5000	29
477	Lauren Smith	F	Alabama	1975-10-2	6500	38
100	Danielle Smith	F	Arizona	1986-10-2	7000	27

Of **A3** the computation results are as follows:

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY	Age
5	Ashley Smith	F	Texas	1975-05-13	16000	38
20	Alexis Allen	F	Florida	1977-08-07	16000	36
22	Jacob Davis	M	Texas	1985-05-07	16000	28
47	Elizabeth Brown	F	Pennsylv	1971-08-27	16000	42
40	Madeline Johnson	F	Californi	1971-12-27	15000	42

Of **A4** the computation result is as follows:

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY	Age
20	Alexis Allen	F	Florida	1977-08-07	16000	36
5	Ashley Smith	F	Texas	1975-05-13	16000	38
47	Elizabeth Brown	F	Pennsylvani	1971-08-27	16000	42
22	Jacob Davis	M	Texas	1985-05-07	16000	28
40	Madeline Johnson	F	California	1971-12-27	15000	42

2.2.7 Computed Column and Computed Field

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	=demo.query("select NAME as STATE, ABBR, CAPITAL, POPULATION from STATES")	
3	=A1.derive(A2.select@1(STATE ==A1. STATE):StateInfo)	/In the employee table, add a computed column StateInfo to store the record of state information of respective employee.

After adding computed column and computed field, you can find in **A3** that the original TSeq has been changed:

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY	StateInfo
1	Rebecca Moore	F	Californ	1974-11-20	7000	California
2	Ashley Wilson	F	New Yo	1980-07-19	11000	New York
3	Rachel Johnso	F	New Me	1970-12-17	9000	New Mexi
4	Emily Smith	F	Texas	1985-03-07	7000	Texas
5	Ashley Smith	F	Texas	1975-05-13	16000	Texas

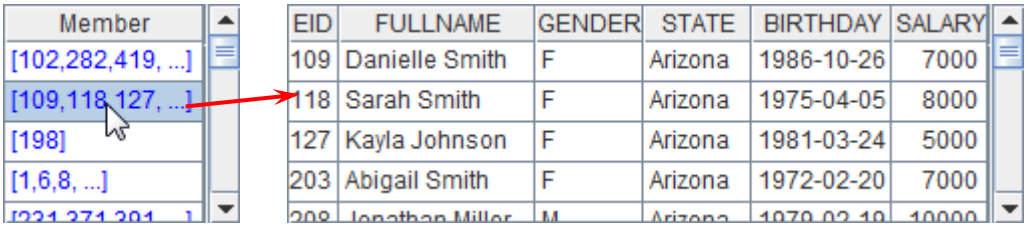
2.3 Group

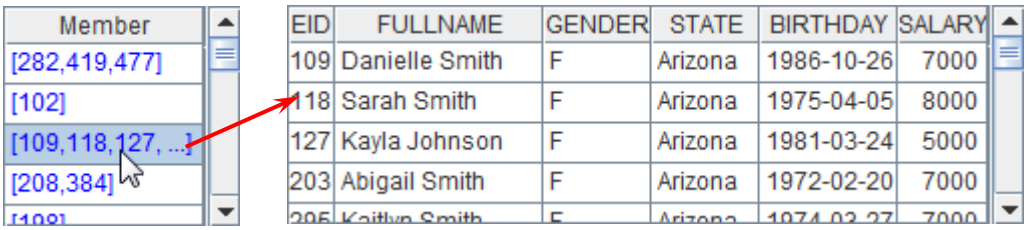
- Equal Group
- Compute Group Summarization Directly
- Equal Group Options
- Align Group
- Enum Group

2.3.1 Equal Group

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	=A1.group(STATE)	/Group the employee record by the state it is located
3	=A1.group(STATE, GENDER)	/Group employee record by its state and gender
4	=A2.new(~.STATE:State,~.count():Count)	/Collect statistics on the total number of employees of each state
5	=A3.new(~.STATE:State,~.GENDER:Gender,~.count():Count)	/Collect statistics on respective total number of female and male employees of each state

Of **A2** and **A3** the computation results are given below. Each group is RSeq. You can double click to view it:





Of **A4** and **A5** the statistics result is as follows:

State	Count
Alabama	4
Arizona	9
Arkansas	1
California	55
Colorado	6

State	Gender	Count
Alabama	F	3
Alabama	M	1
Arizona	F	7
Arizona	M	2
Arkansas	M	1

2.3.2 Compute Group Summarization Directly

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	=A1.groups(STATE:State;count(~):Count)	/Collect statistics the total employees of each state
3	=A1.groups(STATE:State,GENDER:Gender ; count(~):Count)	/Collect statistics the respective total female and male employees of each state

Of **A2** and **A3** the results are the same to that of the **A4** and **A5** in previous example:

State	Count
Alabama	4
Arizona	9
Arkansas	1
California	55
Colorado	6

State	Gender	Count
Alabama	F	3
Alabama	M	1
Arizona	F	7
Arizona	M	2
Arkansas	M	1

2.3.3 Equal Group Options

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	=A1.group@o(STATE)	/Do not sort the employee record, and group the same employees from the neighboring states to one group
3	=A1.group@n(STATE)	/After grouping, only the number employee record is kept in the group
4	=A1.group@1(STATE)	/RSeq consists of the 1st record of each group

Of **A2** the group result is as follows. Each group is an RSeq on which you can double click to view:

Member
[1]
[2]
[3]
[4,5]
[6]

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY
2	Ashley Wilson	F	New York	1980-07-19	11000

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY
4	Emily Smith	F	Texas	1985-03-07	7000
5	Ashley Smith	F	Texas	1975-05-13	16000

Of **A3** the group result is as follows. Each group is a TSeq on which you can double click to view:

Member
[102,282,419, ...]
[109,118,127, ...]
[198]
[1,6,8, ...]
[231,371,391, ...]

Member
109
118
127
203
208

Of **A4** the group result is as follows:

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY
102	Christian Smith	M	Alabama	1972-07-25	12000
109	Danielle Smith	F	Arizona	1986-10-26	7000
198	Christopher Robinson	M	Arkansas	1978-10-03	8000
1	Rebecca Moore	F	California	1974-11-20	7000
231	Stephanie Smith	F	Colorado	1986-03-26	6000

2.3.4 Align Group

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	[New York,Texas,California]	
3	=A1.align@a(A2, STATE)	/Retrieve and group the employees from New York, Texas, and California states.
4	=A3.new(A2(#):State,~.count():Count)	/Collect statistics on the grouping results in the A3. The name of state can be obtained from sequence A2 according to its position.

Of **A3** the grouping result is as follows. Each group is a TSeq on which you can click to view:

Member
[2,12,15, ...]
[4,5,9, ...]
[1,6,8, ...]

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY
2	Ashley Wilson	F	New York	1980-07-19	11000
12	Jessica Davis	F	New York	1980-09-11	7000
15	Alexis Smith	F	New York	1983-07-10	8000
25	Sarah Davis	F	New York	1976-11-25	12000
28	Matthew Johnson	M	New York	1972-11-20	6000

Of **A4** the statistics result is as follows:

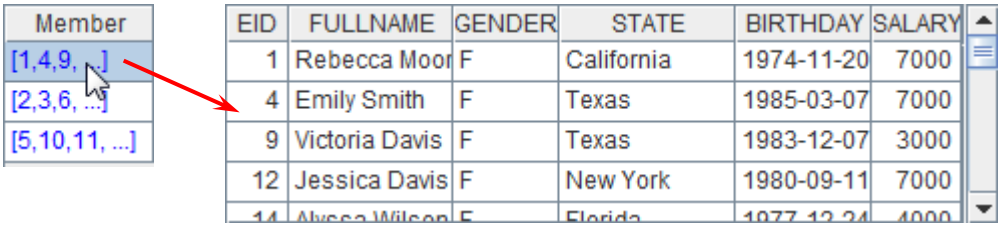
State	Count
New York	45
Texas	54
California	55

2.3.5 Enum Group

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	[?<8000,?<12000,?>=12000]	
3	=A1.enum(A2, SALARY)	/Group by salary range in A2

4	=A3.new(#:Group,~.count():Count)	/Collect statistics on total employees of each range after grouping in the A3
5	[?<8000,?<10000,?<12000,null]	
6	=A1.enum(A5, SALARY)	/Group by salary range in the A5. Those not meet the conditions of the first 3 groups will be grouped in the last group
7	=A6.new(#:Group,~.count():Count)	/Collect statistics on the total employees of each range after grouping in the A6

Of **A3** the group result is as follows. Each group is an RSeq on which you can click to view:



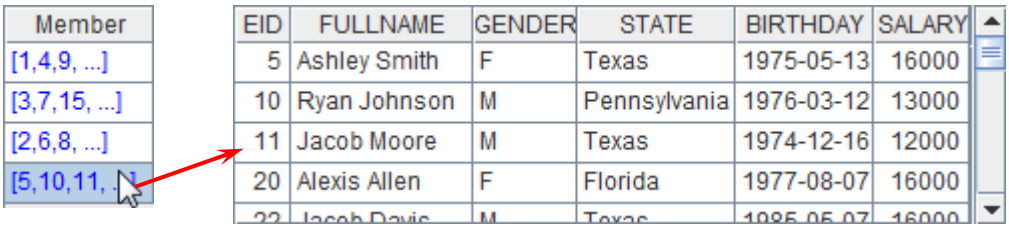
Member
[1,4,9, ...]
[2,3,6, ...]
[5,10,11, ...]

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY
1	Rebecca Moor	F	California	1974-11-20	7000
4	Emily Smith	F	Texas	1985-03-07	7000
9	Victoria Davis	F	Texas	1983-12-07	3000
12	Jessica Davis	F	New York	1980-09-11	7000
14	Alyssa Wilson	F	Florida	1977-12-24	4000

Of **A4** the statistics is as follows:

Group	Count
1	308
2	153
3	39

Of **A6** the group result is as follows. Each group is a RSeq on which you can double click to view:



Member
[1,4,9, ...]
[3,7,15, ...]
[2,6,8, ...]
[5,10,11, ...]

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY
5	Ashley Smith	F	Texas	1975-05-13	16000
10	Ryan Johnson	M	Pennsylvania	1976-03-12	13000
11	Jacob Moore	M	Texas	1974-12-16	12000
20	Alexis Allen	F	Florida	1977-08-07	16000
22	Jacob Davis	M	Texas	1985-05-07	16000

Of **A7** the statistics result is as follows:

Group	Count
1	308
2	75
3	78
4	39

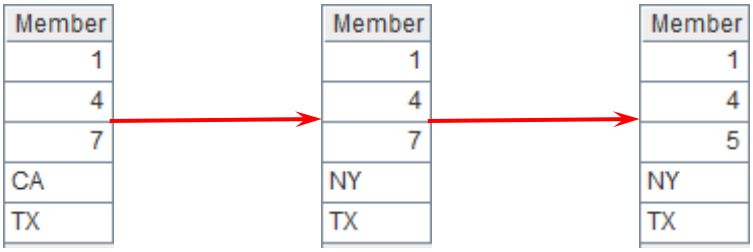
2.4 Modify

- **Modify TSeq Directly**
- **Use Function to Modify Sequence**
- **Modify Record**
- **Modify Data Structure of TSeq**

2.4.1 Modify TSeq Directly

	A	B
1	[1,4,7,CA,TX]	
2	>A1(4)="NY"	/Assign “NY” to the 4th member of the sequence
3	>A1(3)=A1(3)-2	/Subtract 2 from the 3rd member of the sequence

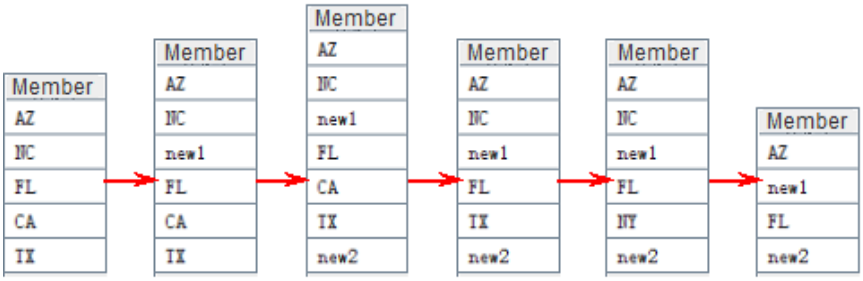
If executing the cellset step by step, you will find that the cell value of **A1** is modified in steps:



2.4.2 Use Function to Modify Sequence

	A	B
1	[AZ,NC,FL,CA,TX]	
2	>A1.insert(3,"new1")	/Insert the member ”new1” at the third position
3	>A1.insert(0,"new2")	/Append the member”new2” to the end of this sequence
4	>A1.delete(5)	/Delete the 5th member
5	>A1.modify(5,"NY")	/Change the 5th member to ”NY”
6	>A1.delete(A1.pselect@a(left(~,1)==”N”))	/Delete all members equal “N” in the sequence

If executing the cellset file step by step, you can find that cell value in the **A1** is changed step by step:



2.4.3 Modify Record

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	>A1.insert(0,2000,"Julia Jones","F","Florida",date("1988-1-10"),10000)	/Append record at the bottom of TSeq
3	>A1.insert(2,20,"M":GENDER,"Brandon Williams":FULLNAME,"California":STATE,"test")	/Insert record at the 2nd position
4	>A1.delete(3)	/Delete the 3rd record
5	>A1.delete([4,20,7])	/Delete the 3 records of the no. 4th, 20th, and 7th
6	>A1.modify(1,1,"Emily	/Modify the 1st record

	Lee","F","California",date("1984-4-2"),8000)	
7	>A1.modify(2,8000:SALARY)	/Modify the salary of the 2nd employee
8	>A1(3).FULLNAME="Helen Brown"	/Modify the name of the 3rd employee
9	>A1.delete(A1.pselect@a(SALARY>15000))	/Delete employee record whose salary is higher than 15000

After modified, TSeq can be viewed in the **A1**:

EID	FULLNAME	GENDER	STATE	BIRTHDAY	SALARY
1	Emily Lee	F	California	1984-04-02	8000
20	Brandon Williams	M	California	test	8000
3	Helen Brown	F	New Mexico	1970-12-17	9000
6	Matthew Johnson	M	California	1984-07-07	11000
8	Megan Wilson	F	California	1970-04-10	11000

2.4.4 Modify Data Structure of TSeq

	A	B
1	=demo.query("select EID, NAME+?+SURNAME as FULLNAME, GENDER, STATE, BIRTHDAY, SALARY from EMPLOYEE"," ")	
2	>A1.alter(ID:EID,Name:FULLNAME, GENDER)	/Change the structure of TSeq so as to keep the serial number, name, and gender fields only, and rename the serial number field to ID

After modifying, the TSeq can be modified and viewed in **A1**:

ID	Name	GENDER
114	Haley Thomas	F
115	Gabriel Johnson	M
116	Alyssa Davis	F
117	Emily Taylor	F
118	Sarah Smith	F

2.5 Flow Control

- Single Line if/else Statement
- If/Else Statement in the Code Block
- Use of Else If
- Non-conditional Loop
- Loop for Specified Times
- Loop Sequence Member
- Record in Loop TSeq

2.5.1 Single Line if/else Statement

	A	B	C	D	E
1	23	123			
2	>C1=A1	if B1>C1	>C1=B1		/Set C1 to the higher one of A1 and B1 values
3	24	21			
4	if A3>B3	>C3=A3	else	>C3=B3	/ Set C3 to the higher one of A3 and B3 values

After executing, the cell values in **C1** and **C3** are as follows:

Value
123

Value
24

2.5.2 If/Else Statement in the Code Block

	A	B	C
1	24	21	
2	if A1>B1		/Set A8 to the higher value of the A1 and the B1. The A9 is the name of this cell
3		>A8=A1	
4		>B8="A1"	
5	else		
6		>A8=B1	
7		>B8="B1"	
8			

Of **A8** and **B8**, the resulting cell values are as follows:

Value
24

Value
A1

2.5.3 Use of Else If

	A	B	C
1	66.8		
2	if A1>=80		/Over 80kg. The level is Heavyweight
3		>A10="Heavyweight"	
4	else if A1>=68		/68~80kg. The level is Middleweight
5		>A10="Middleweight"	
6	else if A1>=58		/58~68kg. The level is Lightweight
7		>A10="Lightweight"	
8	else		/Over 58kg. The level is Flyweight
9		>A10="Flyweight"	
10			

After executing, the cell value of **A10** is as follows:

Value
Lightweight

2.5.4 Non-conditional Loop

	A	B	C	D
1	0			
2	for			/Endless loop
3		>A1=A1+#A2		/Count up in the A1
4		if #A2==100	break	/Quit loop when counting up to 100
5	0			
6	for			/Endless loop
7		>B5=#A6		/Store the count of loops in the B5
8		>A5=A5+B5		/Count up in the A5
9		if A5>=1000	break	/End if count is higher than or equal to 1000

After executing, the result of **A1** is as follows:

Value
5050

Of **A5** the result is as follows:

Value
1035

Of **B5** the result is as follows:

Value
45

2.5.5 Loop for Specified Times

	A	B	C
1	0		
2	for 100		/Loop for 100 times
3		>A1=A1+#A2	/Count in the A1

After executing, the result in **A1** is as follows:

Value
5050

2.5.6 Loop Sequence Member

	A	B	C	D
1	=to(1000)	0		
2	for A1			/Member of loop sequence A1
3		if A2%11!=0	next	/A2 cannot be divided exactly by 11, you can jump to the loop of the next round
4		if A2%2==0	next	/A2 is not odd, so jump to the loop of the next round
5		>B1=B1+1		/Store the number of odds divided by 11 exactly in B1

After executing, the result of **B1** is as follows:

Value
45

2.5.7 Record in Loop TSeq

	A	B	C	D	E
1	=demo.query("select * from STOCKRECORDS where STOCKID=000062")				
2	0	0	0		/The max days of continuous rise is stored in C2
3	for A1				/Loop the stock dealings information
4		if A3.CLOSING >A2	>B2=B2 +1		/Stock rise. Add the days of continuous rise by 1
5		else	if B2>C2	>C2=B2	/End of continuous rise. Set C2
6			>B2=0		/Set the B2 value of days of continuous rise of stock to 0
7		>A2=A3.CLOSING			/Update stock price in A2
8	if B2>C2	>C2=B2			/Set C2 as the last result of B2

In **C2**, compute the days of continuous rise of stock. After executing, the result is as follows:

Value
6

3 Comprehensive Examples

3.1 Common computation

3.1.1 Seek the greatest common divisor of two numbers

In esProc, compute the greatest common divisor of any two numbers.

Solution:

	A	B	C
1	12345678	765432	/Define any two integers
2	>small=A1	>big=B1	/Initialize variable *small* and *big*
3	if small>big	>small=B1	/Set *small* as the smaller integer
4		>big=A1	/Set *big* as the bigger integer
5	for small>0	=big%small	/Loop till *small* becomes 0
6		>big=small	/Set *big* as the original *small*
7		>small=B5	/Set *small* as the remainder of dividing these two numbers
8	=big		/Once loop is completed, the *big* is the greatest common divisor

The result of A8 is shown below:

Value
18

3.1.2 Seek prime number by screening

Seek all prime numbers not greater than 1000000

Solution:

	A	B	C	D
1	1000000			/Set the scope of prime number
2	=to(A1)			/Create an ISeq composed of all integers in the scope, and prepare the result sequence.
3	=int(sqrt(A1))	>A2(1)=0		/The composite number below n must be divisible by an integer below sqrt(n)
4	for A3	if A2(A4)>0	=A1.step(A4,A4*2)	/Loop in proper order. If A2(A4)>0, then the current A4 is the next set prime number
5			>A2(C4)=0	/In A2, the multiples of the current A4 is set to 0
6	=A2.select(~>0)			/All those members that are not set to 0 in A2 are primary numbers

The computation in A6 is shown below:

Member
2
3
5
7
11
13
17
19

3.1.3 Probability test

In the three doors of A, B, and C, there is a hidden prize. When a player selects 1, the judge will choose and open 1 empty door from the 2 remaining gates, then ask the player if he will change his answer to another door. At this time, what choice should the player make to locate the prize more easily?

Solution:

	A	B	C	D
1	1000000	[A,B,C]		/Set the number of use
2	0	0		/Respectively record the times of locating the prize without/with the original answer changed
3	for A1	=int(rand(3))+1	=int(rand(3))+1	/Generate the location of prize randomly, and let the player choose location.
4		=B1(B3)	=B1(C3)	/Door behind which the prize is hidden; and the initial answer of player.
5		=B1\C4	=B5\B4	/Remove the answer of player; and remove the door behind which the prize is hidden
6		=int(rand(C5.len( )))+1	=C5(B6)	/For the remaining gate, randomly select a door and open.
7		=(B5\C6)(1)		/The left door
8		if B4==C4	>A2+=1	/Prize can be obtained if not changing the answer
9		else if B4==B7	>B2+=1	/Prize can be obtained if changing the answer
10	=A2/A1	=B2/A1		/Compute the probability of locating prize in these two respective cases

In A10, the probability of locating the prize with the answer changed is shown below:

Value
0.333427

In B10, the probability of locating the prize with the answer unchanged is shown below. As can be seen, changing the answer in this game will almost double the probability of locating prize:

Value
0.666573

3.1.4 Estimate π through probability test

Get any two positive integers. The probability of they are relatively prime numbers is $6/\pi^2$. Let's compute the π through the probability tests.

Solution:

	A	B	C	D
1	1000000	0		/Specify the number of tests, and store the times they are relatively prime numbers.
2	for A1	=int(rand()*100000)+1	=int(rand()*100000)+1	/Loop and generate two positive integers randomly every time
3		>small=B2	>big=C2	/Compute the greatest common divisor of these two numbers
4		if small>big	>small=C2	
5			>big=B2	
6		for small>0	=big%small	
7			>big=small	
8			>small=C6	
9		if big==1	>B1+=1	/If the two numbers are relatively prime numbers, then the number in B1 plus 1
10	=sqrt(6.0*A1/B1)			/Compute π

The computation in A10 is shown below:

Value
3.1419753974455222

3.1.5 Referee scoring

In a match, there are 6 referees giving their scores. In computing the final score, the 1 highest score and the 1 lowest score must be removed before averaging the remaining 4 scores. The average score is just the final score. Please compute the final score for a player based on these referees' scores.

Solution:

	A	B	C	D
1	[9.8, 9.4, 9.8, 8.9, 9.2, 9.4]			/Enter the referee score
2	func A4,A1			/Call the subprogram whose master cell is A4 and compute the score
3				
4	func	=A4.count()		/Compute the total scores input to A4
5		if B4!=6	return 0	/If not the scores by 6 referees, return 0 as

				the score
6		=round((A4.sum()-A4.max()-A4.min())/(6-2),3)	return B6	/The scores are up to the rule. So, proceed to compute the final score for player with 3 decimal places.

In A2, the resulting player score is as follows:

Value
9.45

Using subprogram to compute is similar to the model of direct looping and executing. However, the subprogram is easier for code maintenance because the subprogram can be modularized to implement the computation function.

3.1.6 Perpetual calendar

Specify the year and month to export the monthly calendar of the selected month.

Solution:

	A	B	C	D
1	2014	1		/Specify year and month
2	=create(Sun,Mon,Tues,Wed,Thur,Fri,Sat)	=A2.insert(0)		/Create new calendar TSeq, and add the initial record
3	=day@w(date(A1,B1,1))-1	=A2(A2.len())		/In A3, prepare to record the week day of the current date. In B3, it is the record of a week in the monthly calendar.
4	for days(date(A1,B1,1))	>A3+=1		/Loop according to the date of specified month. The A3 will plus 1 each time
5		if A3==8	>A2.insert(0)	/The A3 is 8, which indicates new record is added in the new week.
6			>B3=A2(A2.len())	/Update B3 to the record of a new week
7			>A3=1	/In A3, it is set as the 1st day
8		>B3.field(A3,A4)		/Set the current date to the calendar TSeq

In A2, the resulting calendar is as follows:

Sun	Mon	Tues	Wed	Thur	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

3.1.7 Find out the top 3 integers with the greatest number of occurrences

In the file Integers, there are 100,000 integers and some of them are the same. Please find the top 3 integers with the greatest number of occurrences.

Solution:

	A	B	C	D
1	=file("Integers")	=A1.cursor@b()	>a=0	/ Use cursor to retrieve data from the binary file Integers
2	=B1.groupx(Integer;count(~):Count)	=A2.sortx(-Count)		/Group and count the data. Since the number of groups is quite great, use groupx. In B2, the data

	Count;a)			will be sorted by the number of occurrence descendingly
3	=B2. fetch(3)	>B2. close()		/Find out the top 3 integers with the greatest number of occurrence

When using function groupx to group and summarize, a range of temporary data files can be generated according to the summary result, so as to avoid the memory overflow caused by the big data result set. In A3, the result we've found is as follows:

Integer	Count
37328376	6
340282915	6
8460600	5

3.1.8 Search the median from big data

The file Integers holds 100,000 positive integers. We need to find the middlemost integer - the integer in the 50000th position ranking in descending order.

Solution:

	A	B	C	D
1	=file("Integers")	=A1. cursor@b()	>a=0	/ Use cursor to retrieve data from the binary file Integers
2	=B1. groups@n(~. Integer\1000000+1: ID;count(~): Count)	=A2. sum(Count)		/ In the data traversal for the 1st time, the data will be grouped from the smallest to the greatest and every 1,000,000 integers will be set as one range
3	=A2. select@1((a=a+~. Count,a>B2/2))			/ Select out the range in which the median is located
4	=A1. cursor@b()			/Re-generate the data cursor, and prepare the 2nd traversal
5	=A4. select(~. Integer\1000000+1 ==A3. ID)	=A5. sortx(-Integer)		/ Select out all integers in the range where the median is located, and sort the integers in the cursor in descending order
6	=B5. select@1(#==ceil(a-B2/2))	=A6. fetch()		/Select the median from the range

In the first traversal, compute the range where the median is located and its ranking in this range. Because the number of groups is not great after the range is divided, use the function groups for grouping and summarizing. The median can be computed out in the second traversal. In the computation, you are not required to retrieve all data all at once. This method can be used to find the median for whatever amount of data. In A6, the median is found ultimately, as shown below:

Integer
495188891

3.1.9 Frequent words or phrases

The binary file Words is used to store lots of words or phrases, and many of them may have duplicates. Now, we need to find 100 words or phrase that appears most frequently. Because the memory amount is limited, only 1000 phrases can be computed each time at the most.

Solution:

	A	B	C
1	=file("Words")	=A1. cursor@b()	/Create file object

2	=B1. groupx(Words;count(~):Count;1000)		/Group and summarize the big data, and set each temporary file to store 1000 records
3	=A2. sortx(-Count;1000)		/Sort the big data, and set each temporary file to store 1000 records
4	=A3. fetch(100)		/Retrieve the 100 most frequently-used words.
5	>A3. close()		/Close cursor

In A4, the retrieved result is shown below:

Words	Count
the	3333
and	1724
of	1462
to	1379
he	1293
a	1263
in	898
was	800
his	758

3.2 Structured computing

3.2.1 Query data from database

From the STATES table of database, query the State name, abbreviation, and population. Then, query the State whose initial is T and the State with the largest population.

Solution:

	A	B
1	=demo.query("select NAME,ABBR,POPULATION from STATES where left(NAME,1)='T'")	/Query with the set SQL conditions
2	=demo.query("select NAME,ABBR,POPULATION from STATES where left(NAME,1)=?","T")	/Query with SQL clause with parameters
3	=demo.query("select NAME,ABBR,POPULATION from STATES where POPULATION = (select max(POPULATION) from STATES)")	/Use the nested SQL statement to query directly
4	=demo.query("select NAME,ABBR,POPULATION from STATES")	/Query all data in table with SQL statement
5	=A4.select(left(NAME,1)=="T")	/Filter in esProc. In esProc, == is used to determine the equality.
6	=A4.maxp(POPULATION)	/Query the record with the max population in esProc

The query result in A1, A2, and A5 is as follows:

NAME	ABBR	POPULATION
Tennessee	TN	6346105
Texas	TX	25145561

The query results in A3 and A6 are the same, as shown below:

NAME	ABBR	POPULATION
California	CA	37253956

3.2.2 Cross-row computation and proportion

From the STATES table in the database, query the data of State name, abbreviation, and

population. After sorting by population, compute the population difference between each state and the other state ranking ahead of it, and compute the population proportion of each State.

Solution:

	A	B
1	=demo.query("select NAME,ABBR,POPULATION from STATES")	/Retrieve data of all States from database
2	=A1.sort@z(POPULATION)	/Sort by population descendingly
3	=A2.sum(POPULATION)	/Compute total population
4	=A2.derive(Difference,Proportion)	/Add fields in RSeq and return
5	>A4.(Difference=if(#==1,0,POPULATION[-1]-POPULATION))	/Compute population difference
6	>A4.(Proportion=round(POPULATION/A3,4))	/Compute the population proportion, and keep 4 decimal places

The computation in A4 is shown below:

NAME	ABBR	POPULATION	Difference	Proportion
California	CA	37253956	0	0.1207
Texas	TX	25145561	12108395	0.0814
New York	NY	19378102	5767459	0.0628
Florida	FL	18801310	576792	0.0609
Illinois	IL	12830632	5970678	0.0416
Pennsylvania	PA	12702379	128253	0.0411
Ohio	OH	11536504	1165875	0.0374
Michigan	MI	9883640	1652864	0.032

3.2.3 Grouping statistics

From the text file temperature.txt, get the Fahrenheit average temperature data of a certain place. Compute the average temperatures of each quarter and the whole year.

Solution:

	A	B
1	=file@s("temperature.txt").import@t()	/Get the temperature data
2	=A1.derive(int((Month+2)/3):Quarter)	/Add the quarter field to the temperature data, and compute by month
3	=A2.group(Quarter)	/Group by quarter
4	=A3.new(Quarter,round(~.avg(Temperature),2):Average)	/Based on the result of grouping, compute the average temperature of each quarter
5	=round(A1.avg(Temperature),2)	/Compute the mean annual temperature
6	>A4.insert(0,"annual",A5)	/To facilitate viewing, the average temperature of the whole year is added to the result.

The computation in A4 is shown below:

Quarter	Average
1	61.03
2	77.37
3	82.03
4	66.33
annual	71.69

3.2.4 Top n clients which account for half of the total sales

Get the sales data of 2010 from SALES table in a database, and pick out the big client in this

year. The big client refers to the top n clients which account for half of the total sales.

Solution:

	A	B
1	=demo.query("select * from SALES where year(ORDERDATE)=2010")	/Retrieve the sales data of 2010
2	=A1.groups(CLIENT;sum(AMOUNT):Amount)	/Group by client then summarize, compute the sales of each client
3	=A2.sort@z(Amount)	/Sort the sales descendingly
4	=A3.sum(Amount)/2	/Compute half of the total sales for each client in each year
5	0	/Store the accumulated sales
6	=A3.select((A5=A5+Amount,A5<=A4))	/Loop through every client, accumulate the sales, and select the data of clients accounting for first half of the total sales.
7	=A6.(CLIENT)	/List the big client

The computation in A7 is shown below:

Member
HANAR
EGU
HP
JAYB
DILRT
JOPO
BTMMU
DNEDL
ERNSH

3.2.5 Clients whose sales are among the top 15 over the years

Get the sales data from the SALES database table, and find the clients whose sales in each year always are among the top 15

Solution:

	A	B
1	=demo.query("select * from SALES")	/Get all sales data
2	=A1.group(year(ORDERDATE))	/Group by year
3	=A2.(~.group(CLIENT))	/Group the sales data of each year by client
4	=A3.(~.new(CLIENT,~.sum(AMOUNT):Amount))	/Compute the total sales of each client in each year
5	=A4.(~.sort@z(Amount).select(#<=15))	/Sort by annual total sales descendingly, and keep the top 15 clients in each year
6	=A5.(~.(CLIENT))	/Get the clients whose annual sales are among the top 15
7	=A6.isect()	/Compute the intersection set, and find the clients whose sales in each year are always among the top 15

The computation in A7 is shown below:

Member
JAYB
DNEDL

3.2.6 Link relative ratio and year-over-year computation

From the text file quartersale.txt, retrieve the quarterly sales data. Compute the growth ratios of sales of each quarter compared with that of its previous quarter and the same quarter in the previous year respectively.

Solution:

	A	B
1	=file@s("quartersale.txt").import@t()	/Retrieve the sales data
2	=A1.derive('Link-relative','Year-on-year')	/Add blank field for recording the link relative ratio and the year-over-year growth ratio
3	=A2.sort(Year,Quarter)	/Sort the sales data by year and quarter
4	>A3.run('Link-relative'=if(#==1,0,round((Amount-Amount[-1])/Amount[-1],4)))	/Compute the year-over-year growth ratio. The Amount[-1] represents the Amount field of the previous record in A3, that is, sales of the previous month
5	=A3.group(Year)	/Group sales records by year
6	>A5.(~.run('Year-on-year'=if(A5.#==1,0,round((Amount-A5[-1](#).Amount)/A5[-1](#).Amount,4))))	/Compute the link relative growth ratio. A5[-1] represents the sales data of previous year, and A5[-1](#) represents the sales data of the same quarter in the previous year.

The computing result in A3 is shown below:

Year	Quarter	Amount	Link-relative	Year-on-year
2011	1	117700	0	0
2011	2	121900	0.0357	0
2011	3	127200	0.0435	0
2011	4	135000	0.0613	0
2012	1	138600	0.0267	0.1776
2012	2	137600	-0.0072	0.1288
2012	3	138200	0.0044	0.0865
2012	4	145400	0.0521	0.077

Note: The field name with dash is used in the computation, for example, the Link-relative must be quoted with single quotes when being referenced.

3.2.7 Data computing involving multiple homogeneous datasheets

There are 3 text files Finance.txt , Sales.txt, and Marketing.txt. They are of the same structure and respectively stores the employee data of Finance Department, Sales Department, and Marketing Department. Please compute the total number of male and female employees of these 3 departments and the average age of all employees.

Solution:

	A	B
1	=file@s("Finance.txt").import@t()	/Get the employee data of Finance Department
2	=file@s("Sales.txt").import@t()	/Get the employee data of Sales Department
3	=file@s("Marketing.txt").import@t()	/Get the employee data of Marketing Department
4	=A1|A2|A3	/Combine records from the 3 homogeneous TSeq
5	=A4.count@b(GENDER=="M")	/Compute the total number of male employees
6	=A4.count@b(GENDER=="F")	/Compute the total number of female employees
7	=round(A4.avg(age(BIRTHDAY)),2)	/Compute the average age of all employees

The computations in A5, A6 and A7 are shown below:

Value	Value	Value
143	167	35.37

3.2.8 Statistics by age group

A company needs to make statistics on the salary level of employees in each age bracket. Please retrieve the employee data from EMPLOYEE database table, and compute the number and average salaries of employees below 30, between 30 -40, and over 40.

Solution:

	A	B
1	=demo.query("select * from EMPLOYEE")	/Get the employee data
2	[?<30,?<=40,null]	/Set the condition for grouping. For the last group, null is used to represent the remaining record
3	=A1.enum(A2,age(BIRTHDAY))	/Group by employee age on the conditions in A2
4	=A3.new(A2(#):Group,~.count():Count,round(~.avg(SALARY),2):AvgSalary)	/Compute over the grouping data in A3 for each age group

The computation in A4 is shown below:

Group	Count	AvgSalary
?<30	94	7324.47
?<=40	302	7410.6
	104	7413.46

3.2.9 Make statistics on the best player

From the text file Basketball.txt, get the individual score data of a team in a few games, and select the players whose scores are all among the top 3 in each game.

Solution:

	A	B
1	=file("Basketball.txt").import@t()	/Get the individual sore data
2	=A1.derive(Rank)	/Add blank field Rank, and prepare the record score ranking
3	=A2.group(Game)	/Group by games
4	=A3.((ScoreAll=~.(PTS),~.run(Rank=ScoreAll.ranki(PTS))))	/Loop each game, and compute the score ranking of each player
5	=A4.(~.select(Rank<=3).(PLAYER))	/Select the top 3 athletes whose scores are among the top 3 in each game
6	=A5.isect()	/Compute the intersection set, and find the best player

The computation in A6 is shown below:

Member
Kerron Diaz
Ryan Williams

Because the scores may be the same, simply sorting and retrieving the top 3 records cannot get the desired rankings.

3.2.10 Prepare test data

Prepare 1000 rows of test data: generate the serial number of item in proper order, randomly generate the unit price and purchase quantity, and compute the total price. The unit price is between 1 yuan – 100 yuan with two decimal places. The purchase quantity is an integer between 1 and 1000.

Solution:

	A	B	C
1	=create(ID,UnitPrice,Quantity)		/Create blank result TSeq
2	for 1000	=round(int(rand(9900))/100+1,2)	/Loop, and generate unit price of item randomly
3		=int(rand(1000))+1	/Generate purchase quantity randomly
4		>A1.insert(0,A2,B2,B3)	/Insert the resulting record in the result TSeq
5	=A1.derive(round(UnitPrice*Quantity,2):Total)		/Add the field of total price in the result TSeq

Ultimately, the result in A5 is shown below:

ID	UnitPrice	Quantity	Total
1	8.72	66	575.52
2	99.3	94	9334.2
3	27.64	690	19071.6
4	77.8	245	19061.0
5	90.15	795	71669.25
6	50.72	219	11107.68
7	19.1	544	10390.4
8	12.56	344	4320.64

3.2.11 StockRise for 10 Consecutive Day

From the closing prices of several stocks in this month, find the stocks that are rising for 10 consecutive days

Solution:

	A	B
1	=file@s("StockRecords.txt").import@t()	/ Retrieve data from StockRecords.txt
2	=A1.derive(UP)	/ Add a calculation column to store the number of consecutive days to that day during which the stock keeps rising
3	=A2.sort(CODE,DT)	/ Sort the data by the stock code and dealing date
4	=A3.group@o(CODE)	/ Group the data by the stock code
5	=A4.run(~.run(UP=if(CL<=CL[-1],1,UP[-1]+1)))	/ Calculate the number of consecutive days of rising to that day
6	=A5.select(~.max(UP)>10).(CODE)	/ Find out the stocks rising for 10 consecutive days

The computation in A6 is shown below:

成员
120123
129401
711953
900698

3.2.12 Average Interval between Buying Jetta and Magotan

Based on the automobile sales data, calculate the average interval in days between buying the Jetta first and then the Magotan car.

Please note that a customer may buy the car of the same model for several times, and thus

there may be several intervals. In this case, we will take the shortest interval, that is, the number of days from the date buying Jetta for the last time to the date buying Magotan for the first time.

Solution:

	A	B
1	=file@s("Volkswagen.txt").import@t()	/ Retrieve the raw data
2	=A1.sort(Customer,Date)	/ Sort by the customer code and date
3	=A2.select(Model=="Jetta"||Model=="MAGOTAN").dup@t()	/ Filter out the records of Jetta and Magotan
4	=A3.derive(interval(Date[-1],Date):Interval)	/ Calculate the interval of neighboring records
5	=A4.select(Model[-1]=="Jetta" && Model=="MAGOTAN" && Customer==Customer[-1])	/ Filter out the actual interval in days
6	=round(A5.avg(Interval))	/ Calculate the average interval in days

The computation in A6 is shown below:

值
1233.0

3.2.13 Calculate Payroll according to Absenteeism and Performance

Calculate the payroll table of this month on the basis of the basic information of employees, the absenteeism data of this month, and the performance data of this month. The basic algorithm is: **Actual Pay = Base Pay*(1 – Absenteeism Factor + Performance Factor).**

Solution:

	A	B
1	=file@s("Employees.txt").import@t()	/ Retrieve data from Employees.txt
2	=file@s("Absenteeism.txt").import@t()	/ Retrieve data from Absenteeism.txt
3	=file@s("Performance.txt").import@t()	/ Retrieve data from Performance.txt
4	=join@1(A1:E, Name;A2:A, Name;A3:P, Name)	/ Align the data in the respective table
5	=A4.new(E.Name:Name,round(E.BasePay*(1-A.Absenteeism+P.Performance)):SalaryPayable)	/ Calculate the actual pay according to the expression

The computation in A5 is shown below:

Name	SalaryPayable
Mike	4928.0
Jake	2200.0
Lucy	12312.0
Andy	8625.0
Jim	4653.0
David	8649.0

3.2.14 Locate records in big data file

The text file PersonnelInfo.txt is used to store a great deal of employee data. Please find the data of first 50 male employees in the Texas.

Solution:

	A	B

1	=file("PersonnelInfo.txt")	/Create a file object
2	=A1.cursor@t()	/Create a cursor
3	=A2.select(Gender=="M"&&State=="TX")	/Select the male employee of Texas state, and return a cursor
4	=A3.fetch(50)	/Get the data of top 50 employees
5	>A3.close()	/Close cursor

When handling the big data text, the data neither can be all read into the memory, nor partly selected with SQL as we usually do for database. Under this circumstance, the cursor can be of great help. In A4, retrieve the required data of the first 50 employees. The result is shown below:

ID	Name	Gender	Birthday	City	State
6	Benjamin Powell	M	1960-07-11	Lubbock	TX
22	Timothy Brown	M	1991-01-13	Fort Worth	TX
43	Sean Gray	M	1981-12-06	Laredo	TX
46	Willie Sanders	M	1993-04-30	Dallas	TX
62	Shawn Ross	M	1972-02-21	Fort Worth	TX
88	Jimmy Stewart	M	1992-09-08	Dallas	TX
118	Alan Morris	M	1991-09-08	Lubbock	TX

3.2.15 Compute proportion for big data

Suppose a telecommunications company stores the data of customer phone charges in a binary file TelBill. Please compute the proportion of monthly call charge in the total phone charges for each customer.

Solution:

	A	B
1	=file("TelBill")	/Create a file object
2	=A1.cursor@b()	/Create a cursor
3	=A2.groupx(ID;sum(Amount):Total;200)	/Summarize by group. Compute the total phone charges of each customer
4	=A1.cursor@b()	/Re-create the file object cursor
5	=A4.sortx(ID)	/Sort by customer
6	=join@x(A5:TelBill,ID;A3:Sum,ID)	/Join the phone charge data of each customer to the result of grouping and summarizing
7	=A6.new(TelBill.Date:Date,TelBill.Name:Name,TelBill.Amount:Amount,Sum.Total:Total,round(Amount/Total,4):Ratio)	/Compute the proportion of monthly call charge for each customer
8	=A7.fetch(200)	/For example, only retrieve the first 200 records
9	>A7.close()	/Close cursor

To compute the proportion regarding the big data, firstly, compute the total call charge of each customer. If there is too much data to be held in the TSeq, use the cursor, then align and join the data in the cursors. In A8, only the first 200 results are retrieved, as shown below:

Date	Name	Amount	Total	Ratio
201310	Tina Diaz	92.88	1689.18	0.055
201311	Tina Diaz	75.02	1689.18	0.0444
201312	Tina Diaz	38.26	1689.18	0.0227
201401	Tina Diaz	126.63	1689.18	0.075
201402	Tina Diaz	21.98	1689.18	0.013
201403	Tina Diaz	171.59	1689.18	0.1016
201301	Todd Martin	46.17	1405.24	0.0329

3.2.16 Find possible customer churn

Suppose a telecommunications company stores the customer phone charge data in a binary file TelBill. Please find the 50 customers who are the most likely to lose out for retaining them. To achieve this, compare the phone charge in March of 2014 with the average monthly phone charge in the recent 1 year. If the charge drops sharply, then this customer can be taken as the one who are very likely to lose out.

Solution:

	A	B	C
1	=file("TelBill")	=A1.cursor@b()	/Establish the file object cursor
2	=B1.select(Date>201303 && Date<=201403)		/Select the phone charge data of the recent 1 year, and generate the cursor
3	=A2.new(ID,Date,Name,Amount,if(Date==201403,Amount,0):Last)		/Add a field for the call charge of the last month, so as to prepare the grouping and summarizing, and create the cursor
4	=A3.groupx(ID,Name;sum(Amount):Total,max(Last):Last;100)		/ Group by the user to summarize. Generate a new cursor. Collect statistics on the total sum of phone charges in the last year and the last month
5	=A4.new(ID,Name,Last,round(Total/12,2):Avg,round(Last/Avg,4):LastRatio)		/Compute the proportion of the charge in the last month to the average monthly phone charges. Then, prepare the customer data
6	=A5.sortx(LastRatio)		/Sort by proportion with the sorting function sortx for big data
7	=A6.fetch(50)		/Retrieve the 50 customers whose proportions of phone charges in the last month are among the lowest
8	>A6.close()		/Close cursor

In A7, find the 50 customers whose proportions of phone charges in the last month are among the lowest. Their data is shown below:

ID	Name	Last	Avg	LastRatio
000549	Alan Jackson	0.22	98.48	0.0022
000597	Benjamin Gutierrez	0.21	93.76	0.0022
000135	Jose Baker	0.24	92.25	0.0026
000581	Christina Jackson	0.3	112.66	0.0027
000183	Betty King	0.38	90.97	0.0042
000539	Melissa Nelson	0.55	75.29	0.0073
000439	Kathy Lewis	1.02	75.78	0.0135

3.2.17 Count and index the word occurrences in a given text

In the binary file Articl, there are some texts. Count the number of appearance and create a index for each word in these texts. Save the result as a binary file Article_index, and find out the texts in which both *will* and *about* appear.

Solution:

	A	B	C	D	E
1	=file("Article")	=create(Word,Count, Index)			/Read the data and establish the index table
2	=A1. cursor@b()				/Use the file cursor
3	for A2	=A3. Text	=len(B3)		/Loop the cursor and retrieve the text
4		=C3. (mid(B3,#,1))	=B4. (if(isalpha(~),lower(~)," "))		/Convert the text to the single-character sequence, and convert to lower case characters and white spaces
5		=C4. select(~! =" "||~[-1]! =" ")	=B5. conj@s(). array(" ")		/Combine the consecutive white spaces into 1 white space, convert to a character string and delimit the characters with white space
6		=C5. groups(string(~): Word;count(~): Count)	=B6. len()		/ Count the number of occurrence for each words in the given text
7		=B1. len()	>i1=min(1,B7)	>i2=1	/Loop through the index table
8		for i2<=C6	=B1. m(i1)	=B6(i2)	/Loop and summarize each word in the result
9			=cmp(C8. Word,D8. Word)		/Compare each word in the index table
10			if i1==0||C9>0	>B1. insert(i1,D8. Word,D8. Count,[#A3])	/If a certain word does not exist in the index table, then add it to the index table
11				>i2=i2+1	
12				>i1=if(i1>0&&i1<B7,i1+1,0)	
13			else if C9<0	>i1=if(i1>0&&i1<B7,i1+1,0)	/The word from index table does not appear. Go ahead to search the next word from the index table.
14			else	>B1. modify(i1,Count+D8. Count: Count,Index|#A3: Index)	/The word from index table appears in this given text. Combine the result for processing.
15				>i2=i2+1	
16				>i1=if(i1>0&&i1<B7,i1+1,0)	

17	=file("Article_index")	>A17. export@b(B1)			/Save the index as binary file

According to the problem statement, the memory is not limited. Generally speaking, the number of English words is limited. So, to summarize, use the TSeq to store the index table directly. With the binary index file Article_index, we can find the related texts with one or more key words:

	A	B	C	D
1	=file("Article")	=file("Article_index")		/Create file object
2	=B1. import@b()	[about,will]		/The amount of words is limited, and you can retrieve the index table directly
3	=A2. (Word). pos(B2)	=A2(A3). (Index). isect()		/Select out the serial number of text holding the specified words
4	=A1. cursor@b()	[]	0	/Prepare cursor to retrieve the text data from Article
5	for B3	>A4. skip(A5-C4-1)	>C4=A5	/Loop according to the serial number in B3
6		>B4=B4|A4. fetch(1)		/Store every retrieved result in the RSeq B4
7	>A4. close()			/Close cursor

The sorted index TSeq is as follows:

Word	Count	Index
angry	27	[33, 482, 494, ...]
anguish	10	[1890, 2712, 4798, ...]
anhonest	1	[6388]
anhour	8	[576, 1909, 2513, ...]
anice	1	[6955]
anill	2	[877, 2560]
animal	26	[152, 157, 172, ...]
animalcould	1	[195]
animals	19	[141, 170, 196, ...]

The final result is as follows:

Text
Cooling yourselves! retorted Monks. Not all the rain thatever fell, or ever will fall, will put as
You must have some kind gentleman about you that will hear it asa secret, and advise you what to do
Brownlow, interrupting Rose as she was about to speak. Thepromise shall be kept. I dont think it w
Yes, yes, replied the other. His dog has been seen lurkingabout some old haunt, and there seems li
I havent the slightest doubt you will make a success. You mustcome to the park to-morrow morning s
At last the Chicago firm answered. It was by Mr. Moysdictation. He was astonished that Hurstwood
Dont worry about it. Maybe the grocer will wait. He can dothat. Weve traded there long enough to
Now, continued Mr. Withers, swaying his derby hat softly andbeating one of his polished shoes upon
I have a room-mate, she added, who will have to go wherever Ido. I forgot about that.