Report

Task: (Task 6) Comparison of SIMC Indexing Methods

Scenarios 1

Input: ./create A 10000 5 100

./gendata 10000 5 | ./insert A

./stats A

Output:

Dynamic:

#items: tuples: 10000   tsigs: 10000   psigs: 121   bsigs: 3984

#pages: tuples: 121   tsigs: 15   psigs: 16   bsigs: 16

Static:

tups   #attrs: 5   size: 49 bytes   max/page: 83

sigs   bits/attr: 6

tsigs   size: 48 bits (6 bytes)   max/page: 682

psigs   size: 3984 bits (498 bytes)   max/page: 8

bsigs   size: 128 bits (16 bytes)   max/page: 255

Scenarios 2

Input: ./create B 10000 5 5000

./gendata 10000 5 | ./insert B

./stats B

Output:

Dynamic:

#items: tuples: 10000   tsigs: 10000   psigs: 121   bsigs: 7360

#pages: tuples: 121   tsigs: 27   psigs: 31   bsigs: 29

Static:

tups   #attrs: 5   size: 49 bytes   max/page: 83

sigs   bits/attr: 12

tsigs   size: 88 bits (11 bytes)   max/page: 372

psigs   size: 7360 bits (920 bytes)   max/page: 4

bsigs   size: 128 bits (16 bytes)   max/page: 255

Scenarios 3

Input: ./create C 1000 5 100

./gendata 1000 5 | ./insert C

./stats C

Output:

Dynamic:

#items: tuples: 1000   tsigs: 1000   psigs: 13   bsigs: 3984

#pages: tuples: 13   tsigs: 2   psigs: 2   bsigs: 2

Static:

tups   #attrs: 5   size: 49 bytes   max/page: 83

sigs   bits/attr: 6

tsigs   size: 48 bits (6 bytes)   max/page: 682
psigs   size: 3984 bits (498 bytes)   max/page: 8
bsigs   size: 16 bits (2 bytes)   max/page: 2046

Scenarios 4
Input: ./create D 10000 4 100
         ./gendata 10000 4 | ./insert D
         ./stats D
Ouput:

Dynamic:

#items: tuples: 10000   tsigs: 10000   psigs: 104   bsigs: 3720

#pages: tuples: 104   tsigs: 13   psigs: 13   bsigs: 12

Static:

tups   #attrs: 4   size: 42 bytes   max/page: 97

sigs   bits/attr: 6

tsigs   size: 40 bits (5 bytes)   max/page: 818

psigs   size: 3720 bits (465 bytes)   max/page: 8

bsigs   size: 104 bits (13 bytes)   max/page: 314

1. In not using indexing (use x as the signature type)
Input: ./select A ?,?,a3-039,a4-039,a5-039 x
         ./select B ?,?,a3-039,a4-039,a5-039 x
         ./select C ?,?,a3-039,a4-039,a5-039 x
         ./select D ?,?,a3-039,a4-039 x
Output:

RelName	the total number of page reads	#signature pages	#data pages	Response time(s)	#false match pages
A	121	0	121	0.017	118
B	121	0	121	0.017	118
C	13	0	13	0.005	12
D	104	0	104	0.015	93

Comment:

According to the output table, A and B are the same. It means that pF has a small influence. In the table, C with the minimal number of inserted data has the minimum response time. Meanwhile, the number of false match pages and pages read are minimal as well. For D, the number of attributes is less than A. The number of its page reads and false match is possibly multiple reductions. In addition, the number of signature page is 0 when the query satisfies the condition.

2. Open query (?,?,…,?)
Input: ./select A ?,?,?,?,? t
         ./select B ?,?,?,?,? t

```
./select C ?,?,?,?,? t
./select D ?,?,?,? t
./select D ?,?,?,? p
./select D ?,?,?,? b
```

Output:

RelName	Signature type	the total number of page reads	#signature pages	#data pages	Response time(s)	#false match pages
A	t	136	15	121	0.577	0
B	t	148	27	121	0.348	0
C	t	15	2	13	0.012	0
D	t	117	13	104	0.371	0
D	p	117	13	104	0.105	0
D	b	104	0	104	0.224	0

Comment:

Firstly, the difference between A and B is that B's signature page reads is more than A's because it has smaller pF. Secondly, the number of inserted tuples is a key parameter to affect their performance. Thirdly, for D, response time is simply different between tuple and page signature. Finally, bit-slice signature pages in D have not been read.

3. Query with one solution (100001,?,...,?)

Input:
```
./select A 1000150,?,?,?,? t
./select B 1000150,?,?,?,? t
./select C 1000150,?,?,?,? t
./select D 1000150,?,?,? t
./select D 1000150,?,?,? p
./select D 1000150,?,?,? b
```

Output:

RelName	Signature type	the total number of page reads	#signature pages	#data pages	Response time(s)	#false match pages
A	t	73	15	58	0.011	57
B	t	30	27	3	0.006	2
C	t	7	2	5	0.005	4
D	t	92	13	79	0.014	78
D	p	15	13	2	0.003	1
D	b	8	6	2	0.224	1

Comment:

Compared A with B, the total number of pages reads, data pages and false match pages decreased by reducing pF, while the number of signature pages in A is less than in B. In addition, the preference of tuple signature type in A and D are similar. On the one hand, the number page

signature page reads is twice times as bit-sliced. On the other hand, the number of its data page reads is also far less than tuple signatures'.

4. Query with many solutions (?,?,a3-001,...,?)

Input: ./select A ?,?,a3-021,?,? t
        ./select B ?,?,a3-021,?,? t
        ./select C ?,?,a3-021,?,? t
        ./select D ?,?,a3-021,? t
        ./select A ?,?,a3-021,?,? p
        ./select A ?,?,a3-021,?,? b

Output:

RelNa me	Signature type	the total number of page reads	#signatur e pages	#data pages	Response time(s)	#false match pages
A	t	81	15	66	0.013	25
B	t	68	27	41	0.031	0
C	t	9	2	7	0.005	3
D	t	76	13	63	0.027	22
A	p	57	16	41	0.010	0
A	b	47	6	41	0.009	0

Comment:

Firstly, A, B, D are similar according to the output table. Secondly, page and bit-sliced signature in A have the same number of data page reads and false match pages, while page's number of signature page reads is far more than bit-sliced. Finally, the difference of tuple signature in A is to have a larger number of false match pages.

5. query with multiple values (100001,?,a3-001,...,?)

Input: ./select A 100001,?,a3-001,?,? t
        ./select B 100001,?,a3-001,?,? t
        ./select C 100001,?,a3-001,?,? t
        ./select D 100001,?,a3-001,?,? t
        ./select D 100001,?,a3-001,?,? p
        ./select D 100001,?,a3-001,?,? b

Output:

RelNa me	Signature type	the total number of page reads	#signatur e pages	#data pages	Response time(s)	#false match pages
A	t	17	15	2	0.007	1
B	t	28	27	1	0.006	0
C	t	3	2	1	0.004	0
D	t	14	13	1	0.006	0
D	p	19	13	6	0.004	5

| D | b | 18 | 12 | 6 | 0.004 | 5 |

Comment:

D's three different signatures are similar, but the number of data page reads in p and b type is more than in t type. Meanwhile, A and D in t type are almost the same. However, in the same type, the number of signature page reads in A is a half of B's.