

OLAP 가 (OLAP System and Performance Evaluation for Analyzing Web Log Data)

CRM 가 OLAP (On-Line Analysis Processing)
OLAP 가 가
(Explosion)
2,3
가
MOLAP , MS SQL 2000 Analysis
Services, Oracle Express 가
가
: , MOLAP, , 가

Abstract. Nowadays, IT for CRM has been growing and developed rapidly. Typical techniques are statistical analysis tools, on-line multidimensional analytical processing (OLAP) tools, and data mining algorithms (such neural networks, decision trees, and association rules). Among customer data, web log data is very important and to use these data efficiently, applying OLAP technology to analyze multi-dimensionally. To make OLAP cube, we have to precalculate multidimensional summary results in order to get fast response. But as the number of

dimensions and sparse cells increases, *data explosion* occurs seriously and the performance of OLAP decreases.

In this paper, we presented why the web log data sparsity occurs and then what kinds of sparsity patterns generate in the two and the three dimensions for OLAP. Based on this research, we set up the multidimensional data models and query models for benchmark with each sparsity patterns. Finally, we evaluated the performance of three OLAP systems (MS SQL 2000 Analysis Service, Oracle Express and C-MOLAP).

Key Word : Web Log Analysis, MOLAP, Sparsity, Performance Evaluation

1.

CRM(Customer Relationship Management)

. CRM

가

[1]. CRM

가 OLAP(On-Line Analysis

Processing)

가 가

가

가

가

OLAP

OLAP

가 가

(Explosion)

[2,3,4].

2,3

가

MOLAP

MS SQL 2000 Analysis Services, Oracle Express

가

. 2

OLAP

, 2,3

. 3

가

OLAP

C-MOLAP

, 4

가

OLAP

5

2.

OLAP

2.1

OLAP

OLAP

가

[11].

가

가

가

가

가

2.1.1

OLAP

W3C

가

[5]

OLAP

OLAP

4 - , , - 가 ,

가

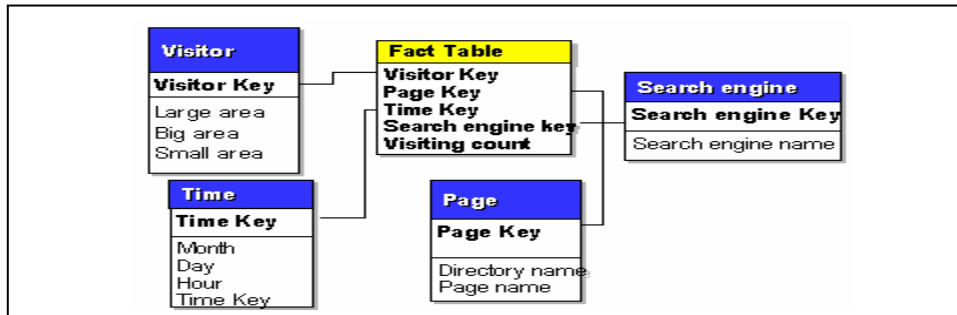
1

1 1

	2000 10 28 ~ 11 4
	-> 2513975 records (163M) -> 35635 records(1.83M)

가

1



1 1

가

15

2

2 2

	2001 02 01 ~ 02 15
	-> 198591 records (120M) -> 72309 records(25.1M)

가

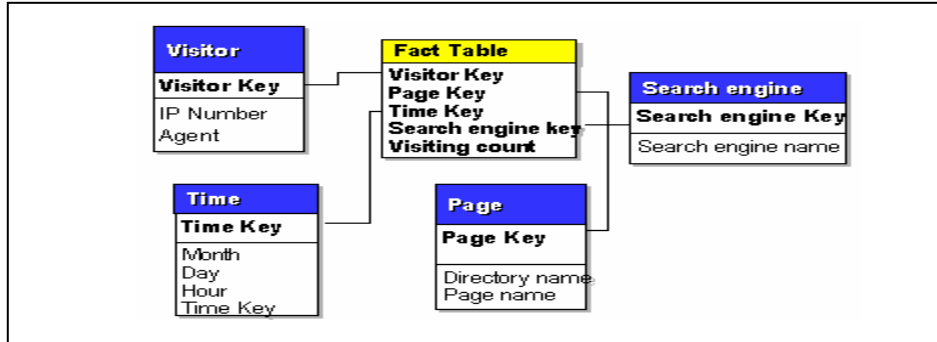
2 가 .

가

3

3

2



[2] 2,3

[3] 3

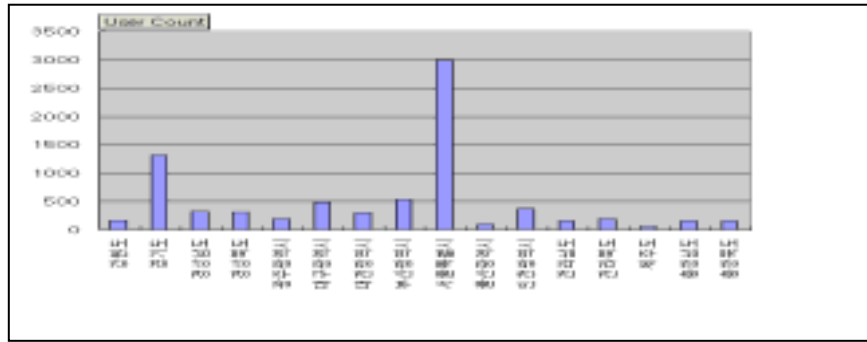
	2001 03 28 ~ 04 03
	-> 21575 records (13.1M) -> 19739 records(4.72M)

2.1.2

1

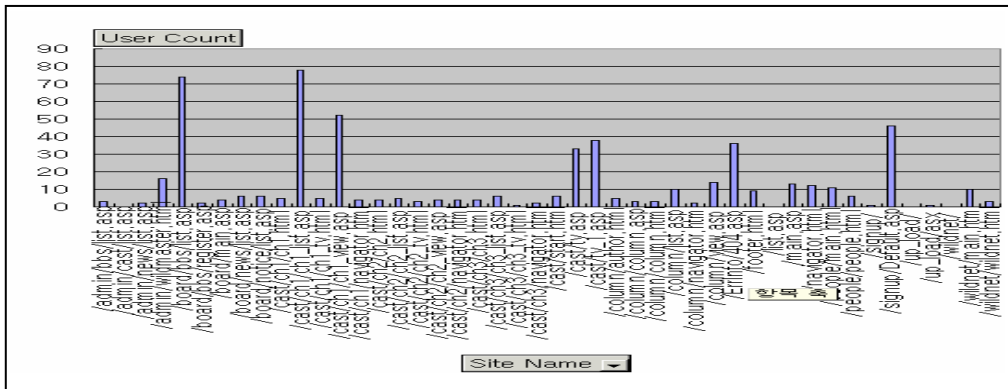
3

가



[3] 1

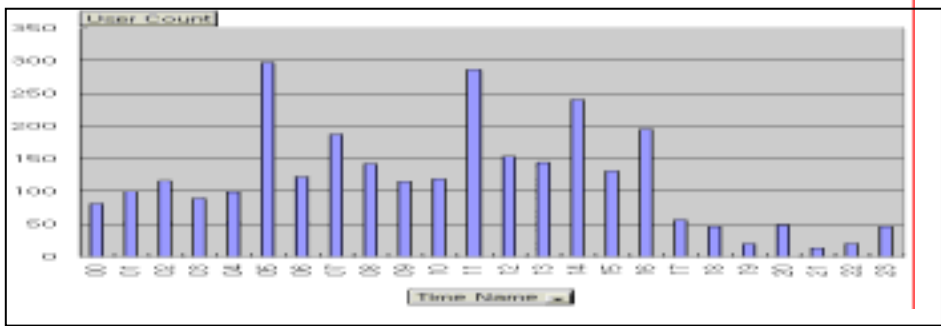
4 3



[4] 3

가 . 5 2

가

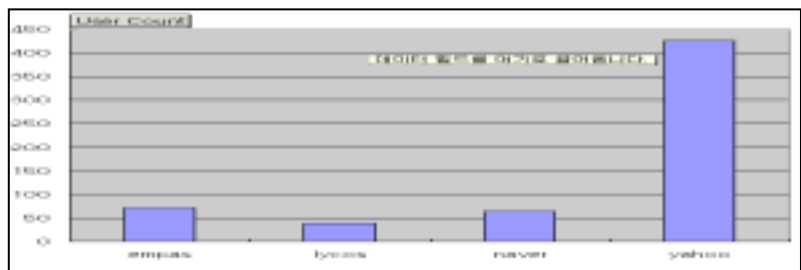


[5] 2

가

가

6 1



[6] 1

가

4가

가

SQL 2000 Analysis Service DBMiner 3D Cube Explore

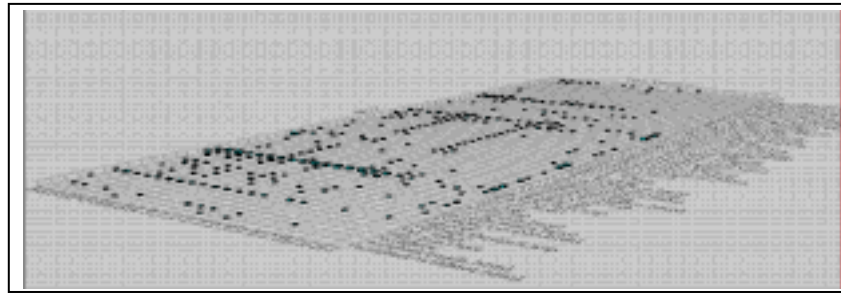
2.2.1 2

3가

OLAP

, 2

7 3



[7] 3 2 : -

(Stripe) 가

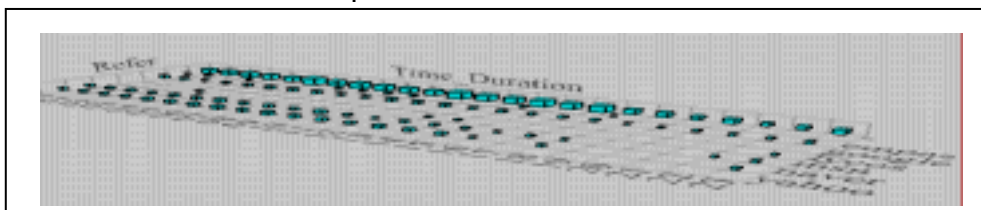
가 (Grid)

(Cluster)

가

8 1

가 (Grid) (Cluster)



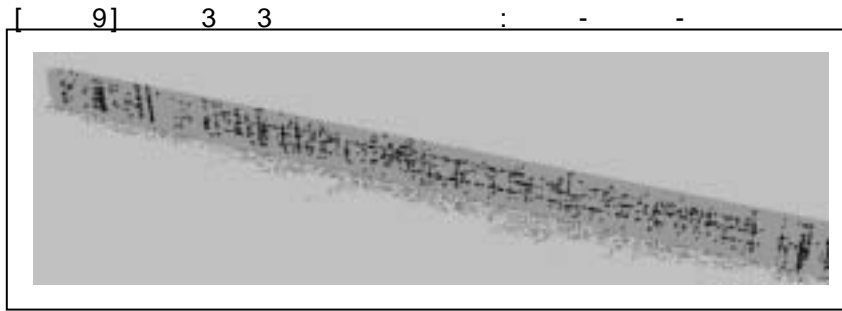
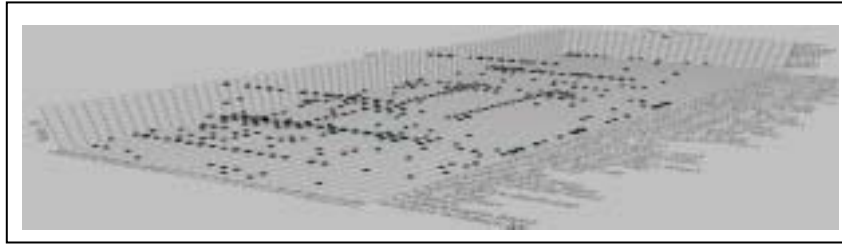
[8] 1 2 : -

2.2.2 3

3

9 10

3 2



[10] 2 3 : - -

2

3

(Grid)

(Cluster)

가

MOLAP

3. C-MOLAP

OLAP

OLAP

가

[6]

MOLAP

C-MOLAP

C-MOLAP

3.1 C-MOLAP

MOLAP

가

가

C-MOLAP

n

n

Chunk-Offset Compression

[12,13].

OLAP

I/O

$ C_i = i$	$ C_i = D_i * (1/2)^n$	$ D_i = i$	$n:$
-------------	---------------------------	-------------	------

[11]

11

[6] 가

$(1/2)^n$ 가

50K 가

가 50*2160*10000

3.1

7

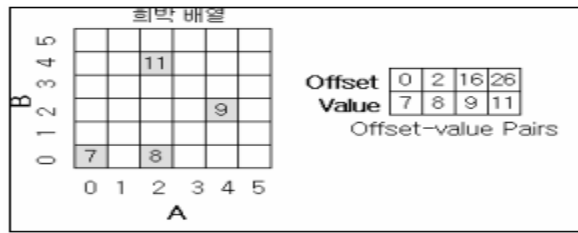
3*16*78

40%

가

“Chunk-Offset Compression”

[12].



[12] Chunk-Offset Compression

offset , offset

가

13

```

typedef struct Chunk_Numbering{
    int start_x; int end_x;
}Chunk_N; //
typedef struct A_Mem_Save{
    int offset; // chunk data offset
    int data_val;
    int chunk_num; }A_Mem;

```

[13]

3.2 C-MOLAP

. N

2^n 가 . 2^n

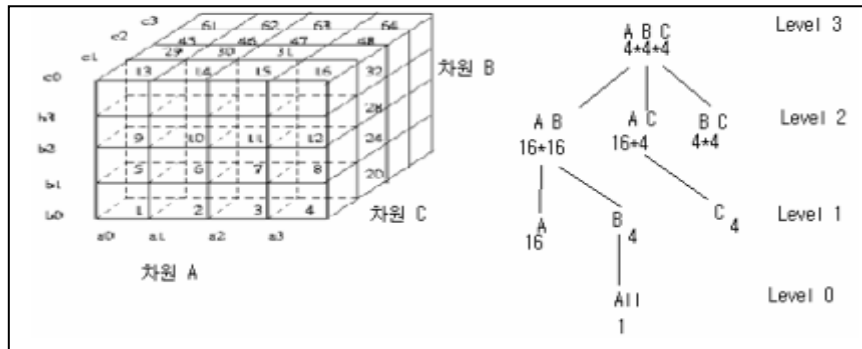
가 .

[6]

C-MOLAP

● Minimum Memory Spanning Tree(MMST)

MMST group_bys
 가 14 3 가
 , 16*16*16 , 4*4*4 A,B,C 가
 가 가



[14] 3 MMST

가 BC , 1~4
 b_0c_0 BC 4*4 가
 BC AC a_0c_0
 4*16

$$\prod_{i=1}^p |D_i| \times \prod_{i=0}^{k-1} |C_i| \quad |D_i| : \text{차원 } i \text{의 사이즈} \quad |C_i| : \text{차원 } i \text{의 청크 사이즈}$$

[15] group_bys

p 가

최적의 차원의 순서 : |D1| <= |D2| <= <= |Dn|

[16]
 16 가 가
 . 15 16 group_by
 group_by

4. 가
 2 가
 , MS SQL
 2000 Analysis Service, Oracle Express Server, C-MOLAP OLAP

4.1 가 가

, , , .
 4
 [4]

	10,000	50	2,160
	1	4	3

3가 2,3

- 2
- 2 , , 2
- , , 3 , , 2

4

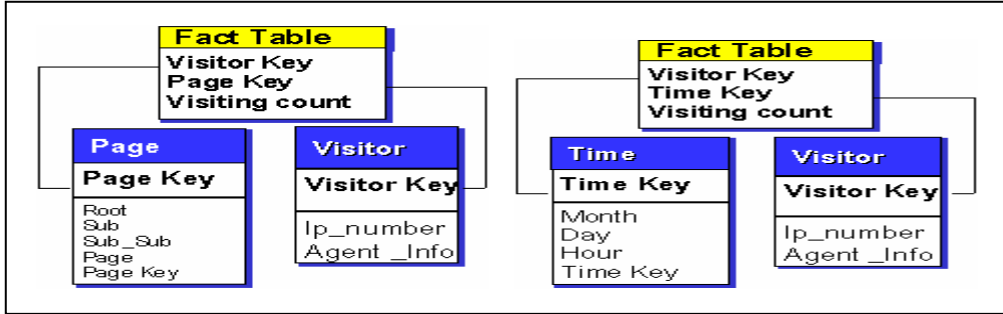
2

A

B 가

A B

17 5



[17] 2

• 3

3

, , 3가

. 3가

2

, 3

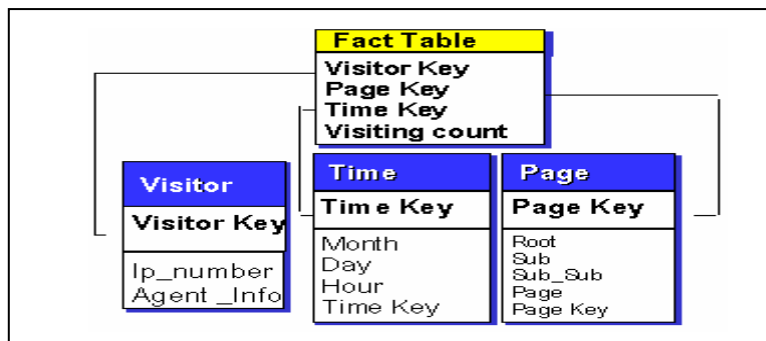
C 가

C

가 18 5

[5] 2,3

			(10%)
A	,	500,000	50,000
B	,	21,600,000	2,160,000
C	,	1,080,000,000	108,000,000



[18] 3

•

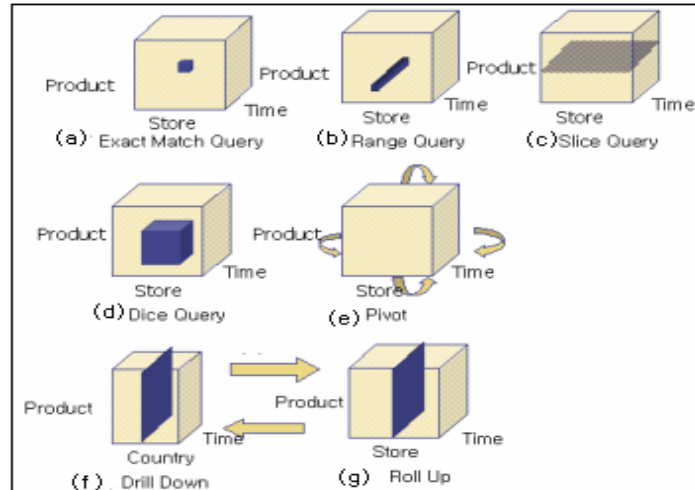
2,3

OLAP

가

OLAP

19 [6, 7].



[19] OLAP 7가

7가 OLAP OLAP

Top500

A,B,C

7가

OLAP

[6] A OLAP

Exact Match	(PK00003) (VK03018) 가
Range	VK03000~VK03600가 S00001
Slice	(VK03018)가
Dice	S00001
Pivot	,
Drill Down	VK0318 R0001

Roll Up	VK0318
---------	--------

[7] B OLAP

Exact Match	(TK00042) (VK03018) 가
Range	VK03000~VK03600가 8 2
Slice	(VK03018)가
Dice	8 2
Pivot	,
Drill Down	VK3018 8 2
Roll Up	VK3018 8 2 8

[8] C OLAP

Exact Match	(PK00003) (VK00012)가 TK02041
Range	8 1 (VK00001 - VK00100) PK00001 - PK00030
Slice	(VK00001)가
Dice	(VK00300 ~ VK00330) 8 1
Pivot	VK00012
Drill Down	'VK000012' 'PK00008' 8 1
Roll Up	'VK000012' 'PK00008' 8 1

9 3가

Top500

[9] A, B, C Top500

A	가 (PK00003) 가 Top500
B	(10) 가 Top500
C	(10) (PK00003) 가 Top500

4.2

MS SQL 2000 Analysis Service , Oracle Express Server, C-MOLAP 3가

. OLAP 가 10 .

[10]

	Windows NT 4.0 Server
OLAP	Oracle Express, MS SQL 2000 Analysis Service
	Visual Basic 6.0 MDAC (Microsoft Data Access Components) 2.5 SDK (ADO MD: ActiveX Data Objects Multidimensional), MDX (Multidimensional Expressions) OEO (Oracle Express Object) Express Basic, Express Language Visual C++ 6.0

MS SQL 2000 Analysis Service 가

Visual Basic 6.0 . Visual Basic OLAP

ADO MD API , MDX [7].

Oracle Express Server Oracle Express Administrator

, Oracle Express Object Oracle Basic 가

. Oracle Language

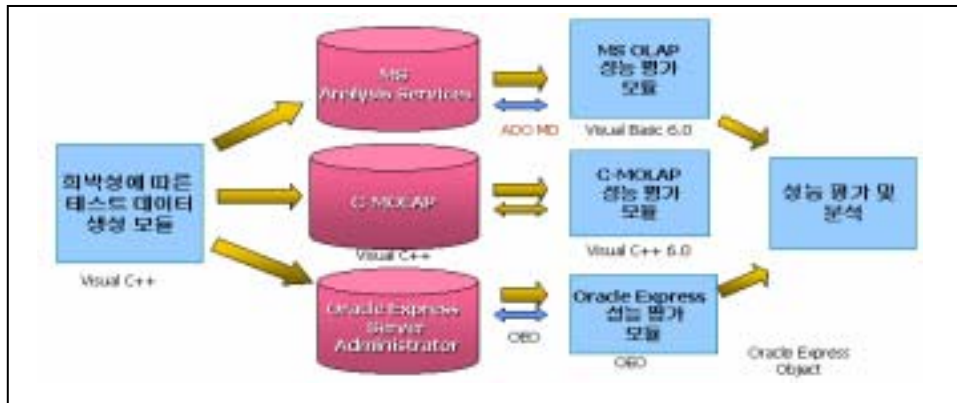
[8,9]. C-MOLAP MOLAP

Visual C++ 6.0

가

가

20



[20]

가 OLAP

. MS SQL 2000 Analysis Service

MS SQL

2000 DTS

RDB

Analysis Service 2,3

[10]. Oracle Express

Oracle Express Administrator Database

. 가 OLAP

2,3

가

3가

가

4.3 가

MS SQL 2000 Analysis Service

Visual Basic

Analysis Service

Microsoft MDAC SDK 2.5

ADO MD API

21 MS SQL 2000 Analysis Service

가

C

7가

OLAP

Slice

Oracle Express Server

Express Language

가

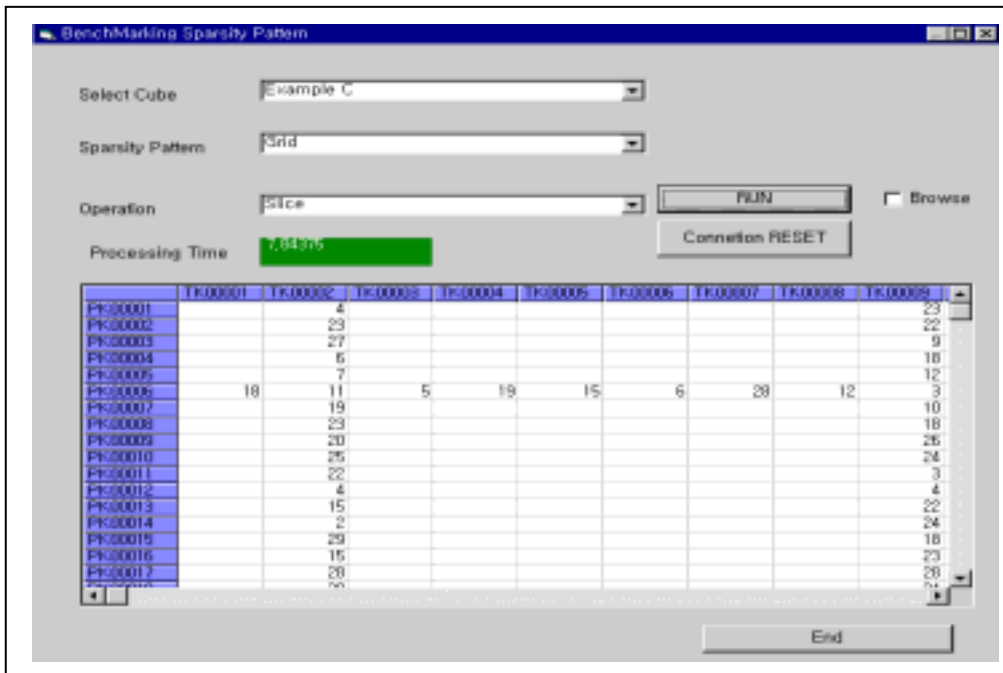
Oracle Express Object

Express Basic

22

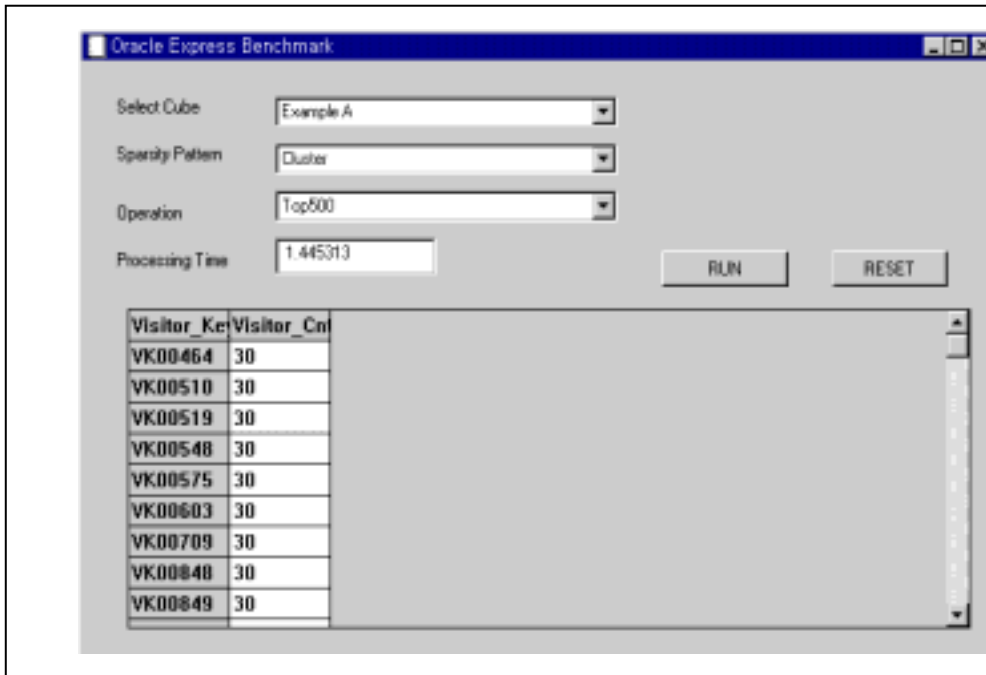
A

Top500



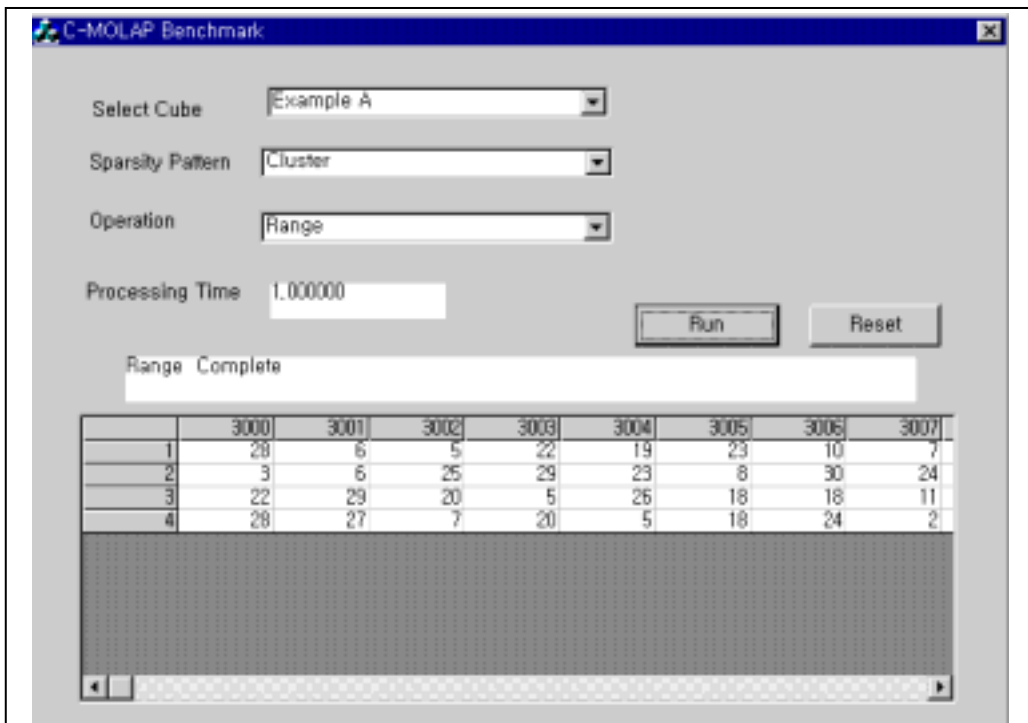
21. MS SQL 2000 Analysis Service

가



22. Oracle Express 가

23 C-MOLAP 가 A Range



23. C-MOLAP 가

4.4 가

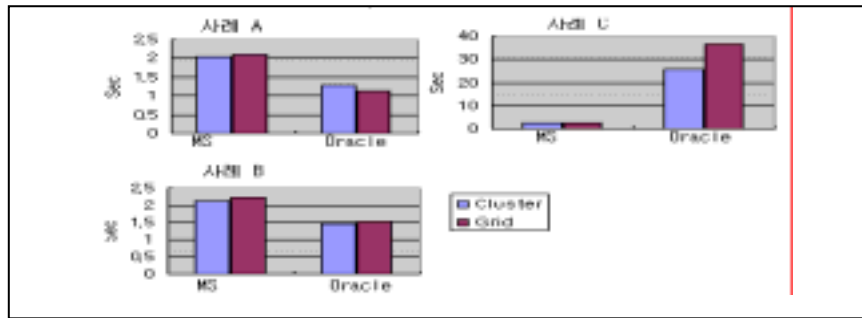
3가 OLAP

가

(Sec)

3

OLAP



[24] Top500 가

[24] Top500

. Top500

OLAP

C-

MOLAP

가

A B

가

Oracle

Express가 1

C

가

MS SQL 2000 Analysis Service가 10

OLAP

7가

가

3가

가 2

가

25, 26, 27

A, B, C

A B

Dice Range

MS SQL 2000 Analysis Service가

가

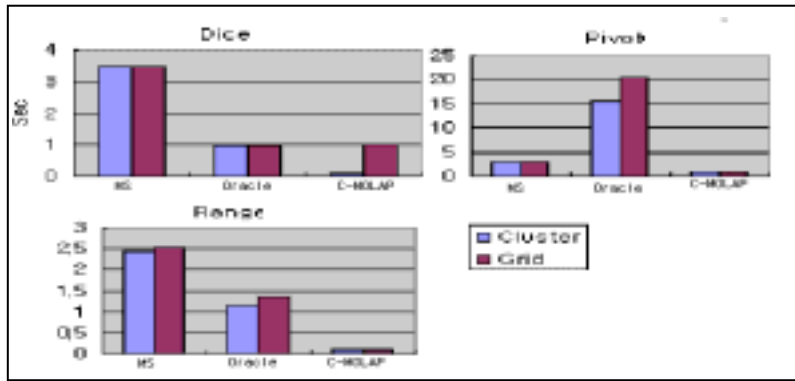
1~2

가

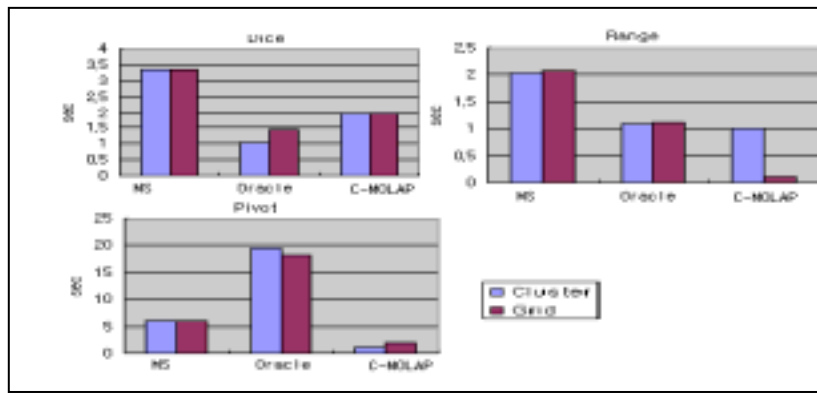
C

Oracle Express

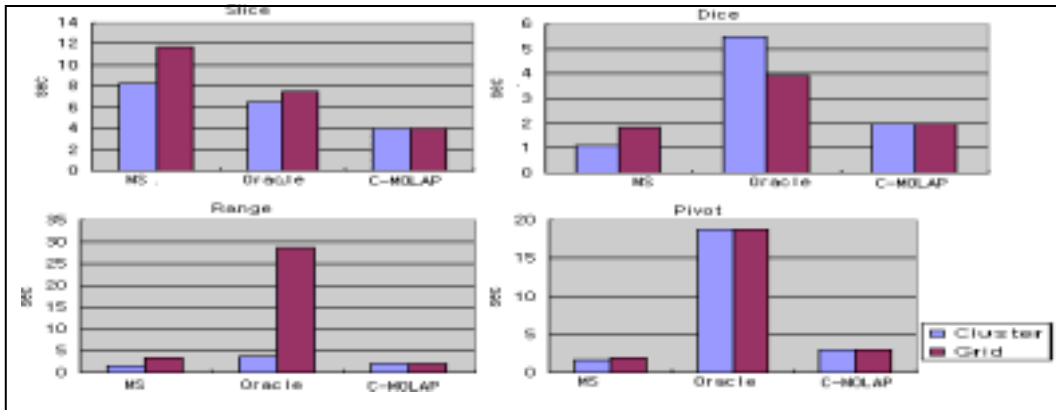
2~25



[25] A 가



[26] B 가



[27] C 가

Range Oracle Express
 Pivot Oracle Express가
 가
 , 가

Dice Range, Pivot Oracle Express 가

5.

CRM

가

. CRM

OLAP

, 가 OLAP , (Data Explosion)

OLAP

2,3

(Grid)

(Cluster)

가

. 7가 OLAP

Top500

3가 OLAP

(MS SQL 2000 Analysis Service, Oracle Express, C-MOLAP) 가

가

가

Top500

, Dice

, Range

MS SQL 2000 Analysis Service

가 가

가

3가

Oracle

Express

. Pivot

Oracle

Express

가

-

가

Data Mining

CRM

가


- C-MOLAP


가

가 I/O , 가
가 . 가
가

- [1] MaxScan Corp, White Paper. “An Accelerator for Click Stream Data Analysis Applications”
- [2] *White Paper*, <http://www.olapreport.com/DatabaseExplosion.htm>
- [3] Michael L. Gonzales, “Estimating the Explosion of Derived Cells”, DB2 magazine, pp 14-17, 2000.
- [4] Sanjay Goil and Alok Choudhary, “ Sparse Data Storage of Multi-Dimensional Data for OLAP and Data Mining”, Technical Report CPDC-TR-9801-005, Center for Parallel and Distributed Computing, Northwestern University, 1997.
- [5] Robert Cooley, Bamshad Mobasher, Jaideep Srivastava, “Data Preparation for mining world wide web browsing patterns”, the Journal of Knowledge and Information System, Vol. 1, No. 1, 1999
- [6] Y. Zhao, P. M. Deshpande, and J. F. Naughton, “An Array-Based Algorithm for Simultaneous Multidimensional Aggregates”, In Proc. of ACM SIGMOD, pp 159-170, 1997
- [7] MDAC(Microsoft Data Access Component)Documentation, <http://www.microsoft.com/data/whatcom.htm>
- [8] Oracle Corp, “Oracle Express Objects User’s Guide Release 6.3”, Aug 1999
- [9] Oracle Corp, “ Oracle Express Programmer’s Guide to the Express Language Release 6.3”, Sep 1999

- [10] Microsoft Corp, "Microsoft Corp. SQL Server 2000 Analysis Services", Nov 2000
- [11] Ju-young Kang, "Classification of sparsity patterns and performance evaluation in OLAP system", the Master's Thesis of Department of Computer Science and Engineering, Ewha Institute of Science and Technology, 2000
- [12] Sanjay Goil and Alok Choudhary, " Sparse Data Storage of Multi-Dimensional Data for OLAP and Data Mining, *Technical Report CPDC-TR-9801-005*, Center for Parallel and Distributed Computing, Northwestern University, 1997.
- [13] Robert J. Earle. Arbor Software corporation, "Method and Apparatus for Storing and Retrieving Multi-dimensional Data in Computer Memory", *U.S. patent #5359724*, Oct. 1994.

	<p>1999</p> <p>2002</p> <p>2002~ Systems</p> <p>< > OLAP, CRM</p>
--	---

	<p>1983</p> <p>1985</p> <p>1985 ~ 1989</p> <p>1994</p> <p>2002 IBM T.J. Watson Research Center Visiting Researcher</p> <p>1995 ~ :</p> <p>< > , ,</p> <p style="text-align: center;">XML</p>
---	---