



دانشگاه صنعتی شریف
دانشکده کامپیوتر

Main Memory Database

زیر نظر:

استاد سید محمد تقی روحانی رانکوهی

حسن شجاعی مند

۱۳۸۳

AVL

B Tree

T Tree

T

T

T

T

T

T-Tail Tree

DRDB
 MMDB
 DRDB
 object backup
 MMDB

throughput DRDB MMDB

[]:

(

(

(

(

¹ realtime Application

² deadline

³ access time

⁴ volatile

⁵ nonvolatile

⁶ block-oriented

⁷ sequential access

⁸ random access

() (

DRDB

[]:

MMDB

DRDB

()

...

()

DRDB MMDB

DRDB

¹ application interfaces

(B-Tree)

T-Tree

DRDB

MMDB

DRDB

MMDB DRDB

MMDB

[]:

(

¹ index
² UPS

(

(

MMDB

(SM)

(MM)

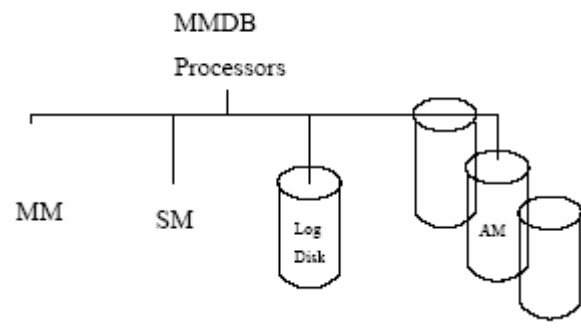
⁴

(SM)

(AM)

.[7]

¹ Main Memory
² Stable Memory
³ Logger Process
⁴ Log Disk
⁵ Archive Memory



MMDB

mmdb

[2]:

•
•
•

[2]:

(

:

:

$O(N)$

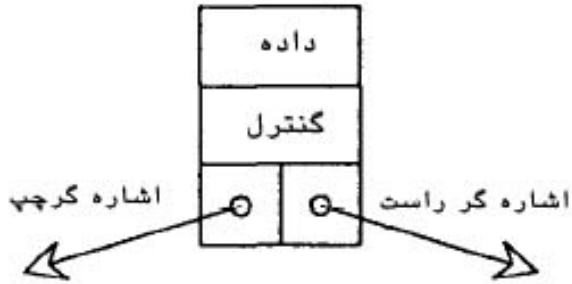
: AVL

¹ index structure
² long field
³ order preserving

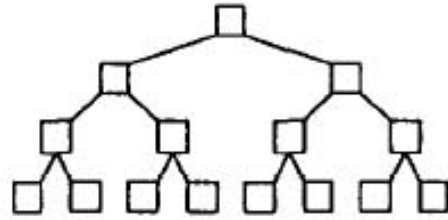
AVL

()

نود درخت AVL



درخت AVL

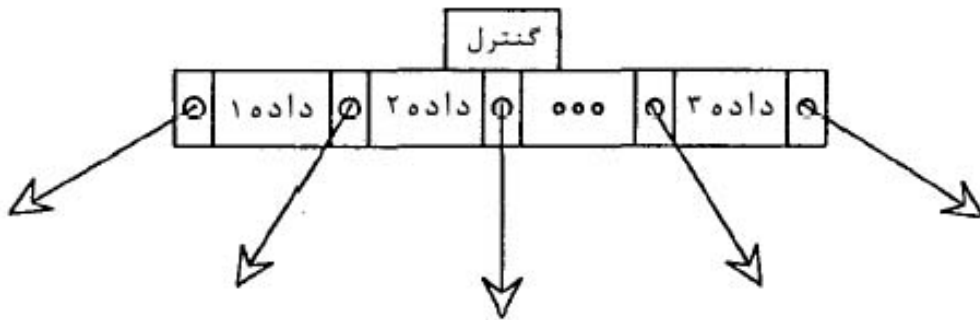


AVL :

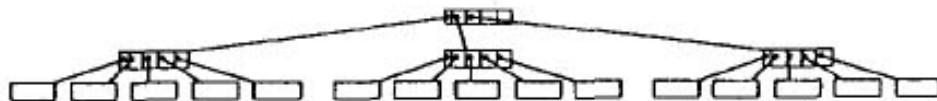
:B-Tree

B-Tree

نود B Tree



B Tree



B :

(

MMDB

(*B*⁺ - *Tree*) B-Tree

T-Tree

B-Tree

update

B Tree AVL

T-Tree

T-Tree

T

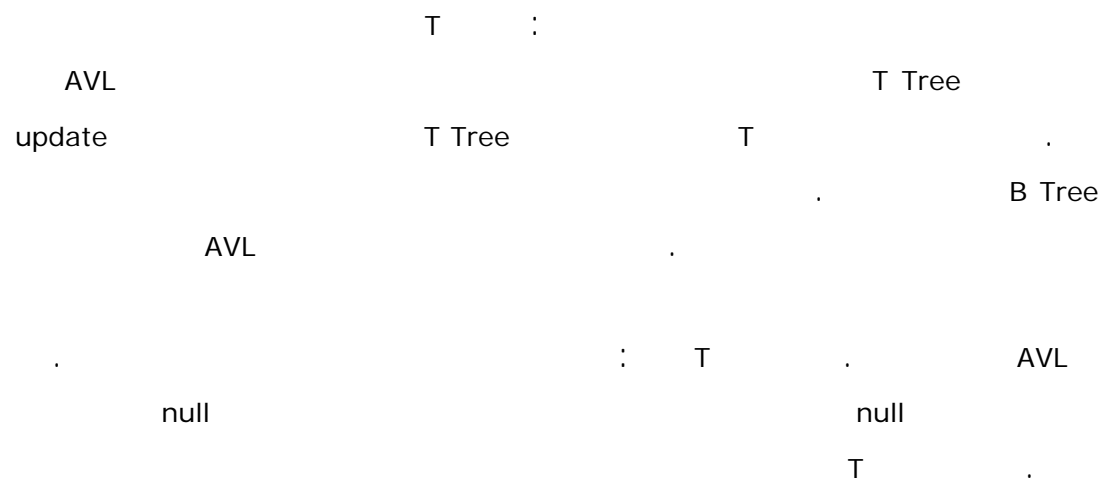
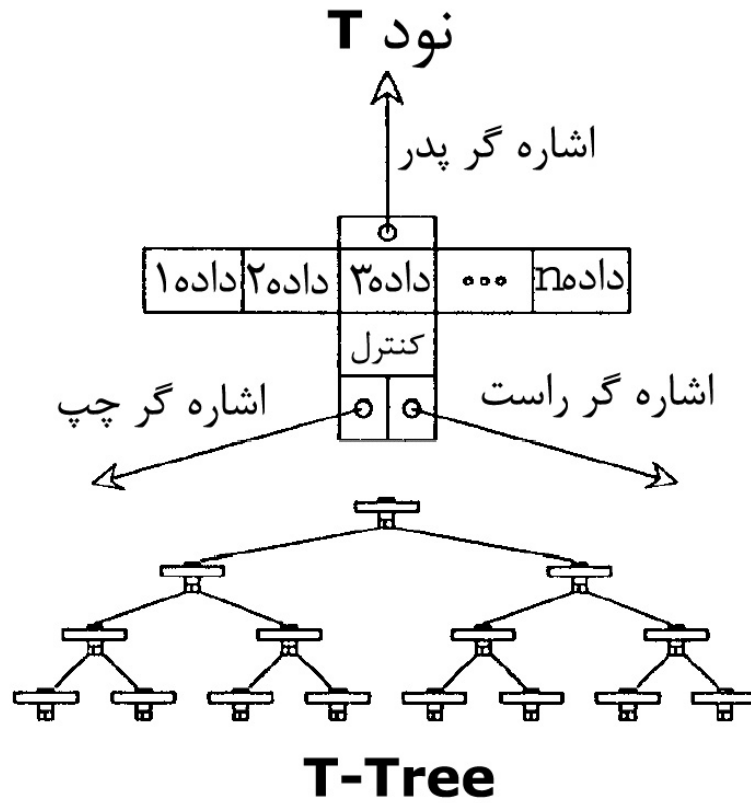
T-Tree

T

T Tree

Tree

¹ randomizing
² extensible hashing
³ linear hashing



¹ Rebalancing
² internal node
³ half-leaf node
⁴ leaf node

T

T

(
(

T

(
(

()

(

()

(

T

:

(

(

(

(

(

T

[3]

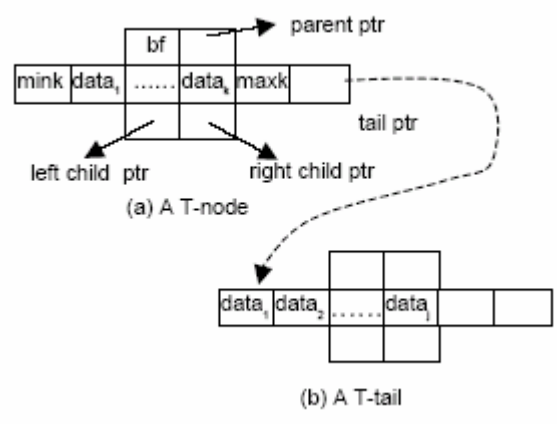
AVL

¹ underflow

[4] T-Tree
 mmdb T-Tree
 T-Tree
 mmdb
 I/O
 T-Tree Carey Lehman
 mmdb
 T-Tree
 B-Tree
 T-Tree

T-Tail Tree

(BTree AVL) T-Tree
 T-Tail-Tree



T-Tail :

T

¹ balanced binary tree

$\text{minK} \leq K \leq \text{maxK}$

T maxK minK
 K T
 K T

T-Tree ()
 -1 0 +1 (bf)
 () A
 A A minK
 A

T-Tail T [4]
 T T-Tail

T

T

¹ balanced factor
² predecessor
³ Successor
⁴ underflow
⁵ overflow
⁶ entry
⁷ completely full

T-Tail

X-Lock SIX-lock S-lock.

:

	S	SIX	X
S	Y	Y	N
SIX	Y	N	N
X	N	N	N

[4] T-Tail

()

SIX

-1 1

X SIX

T-Tail

T-Tail

X

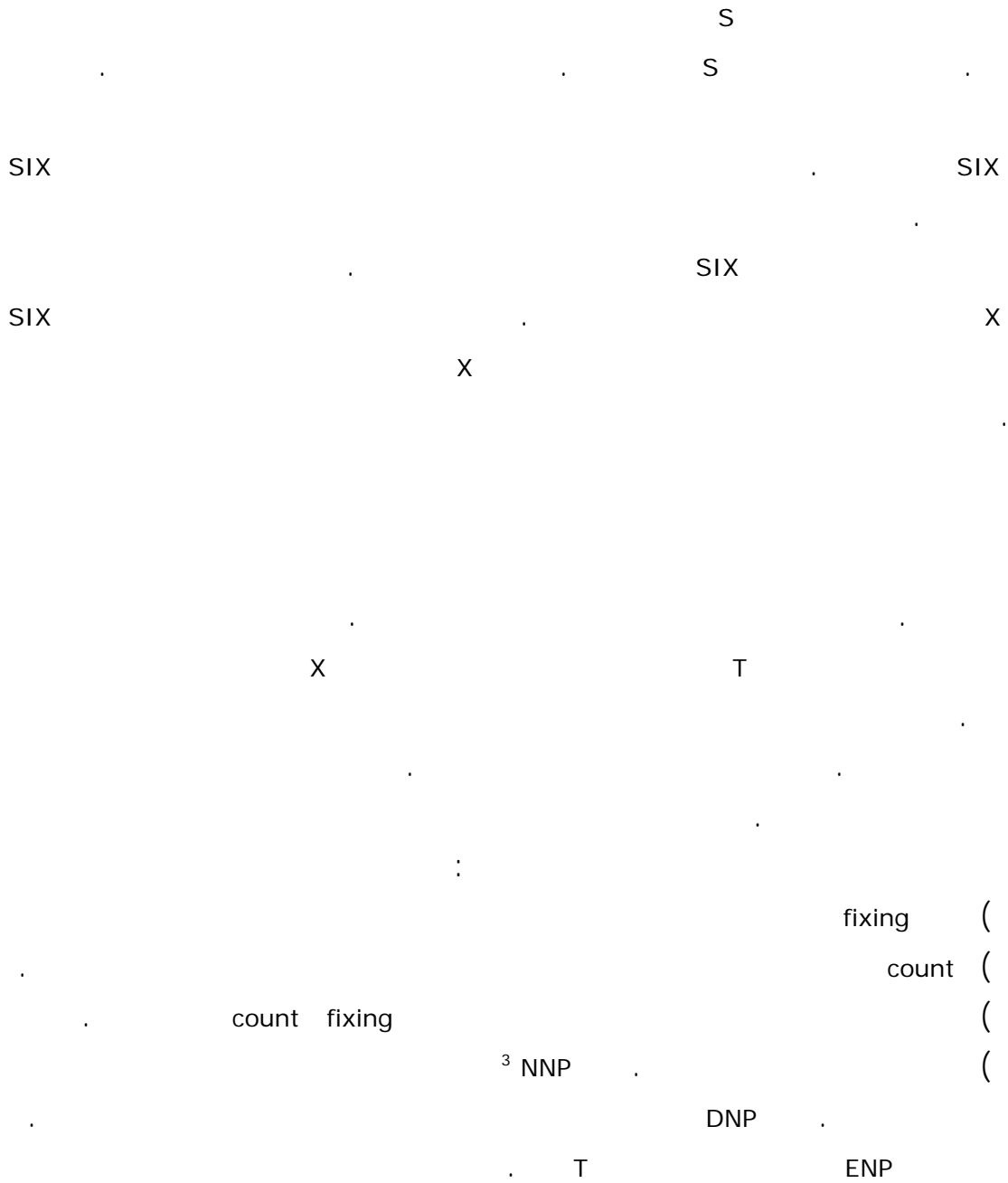
¹ pessimistic

² optimistic

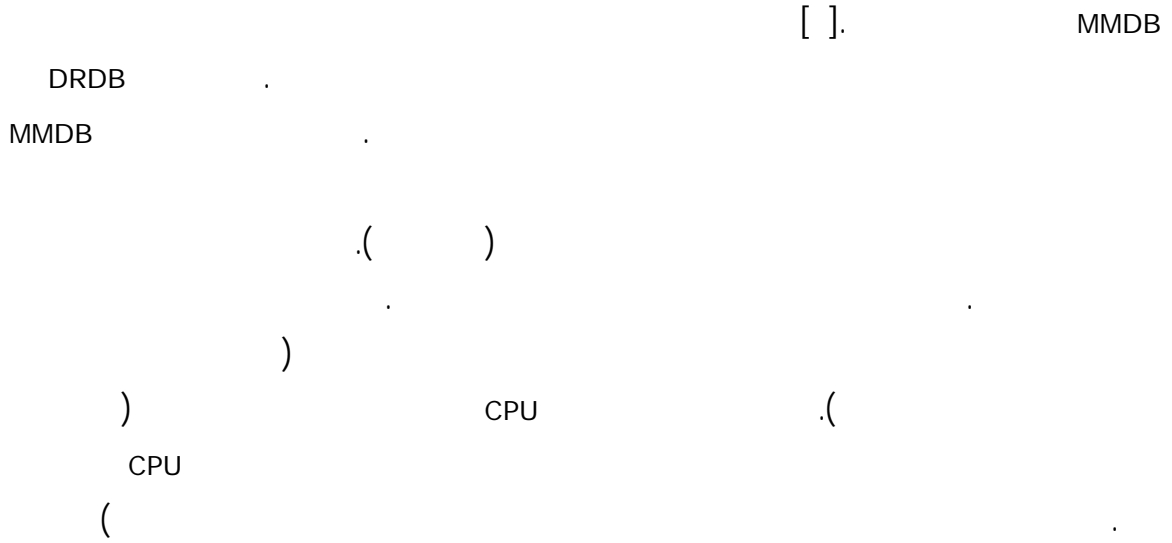
³ lock-coupling

⁴ critical node

⁵ bounding node



¹ updater
² node pointer pool
³ new node pool
⁴ deallocated node pool
⁵ empty node pool



[5]

MMDB

¹ hash table

wake-up

) ()
(

[6]

¹ commit

MMDB

throughput

MMDB

()

¹ bottleneck
² pre-commit
³ group commit

MMDB

. heap

MMDB

sort merge

MMDB

A

S R

A

(R)

a (a)

S

a

MMDB

.[3]

(MM)

(AM)

()

[1,2].

(BT)

(ET)

¹ Logging

² Checkpointing

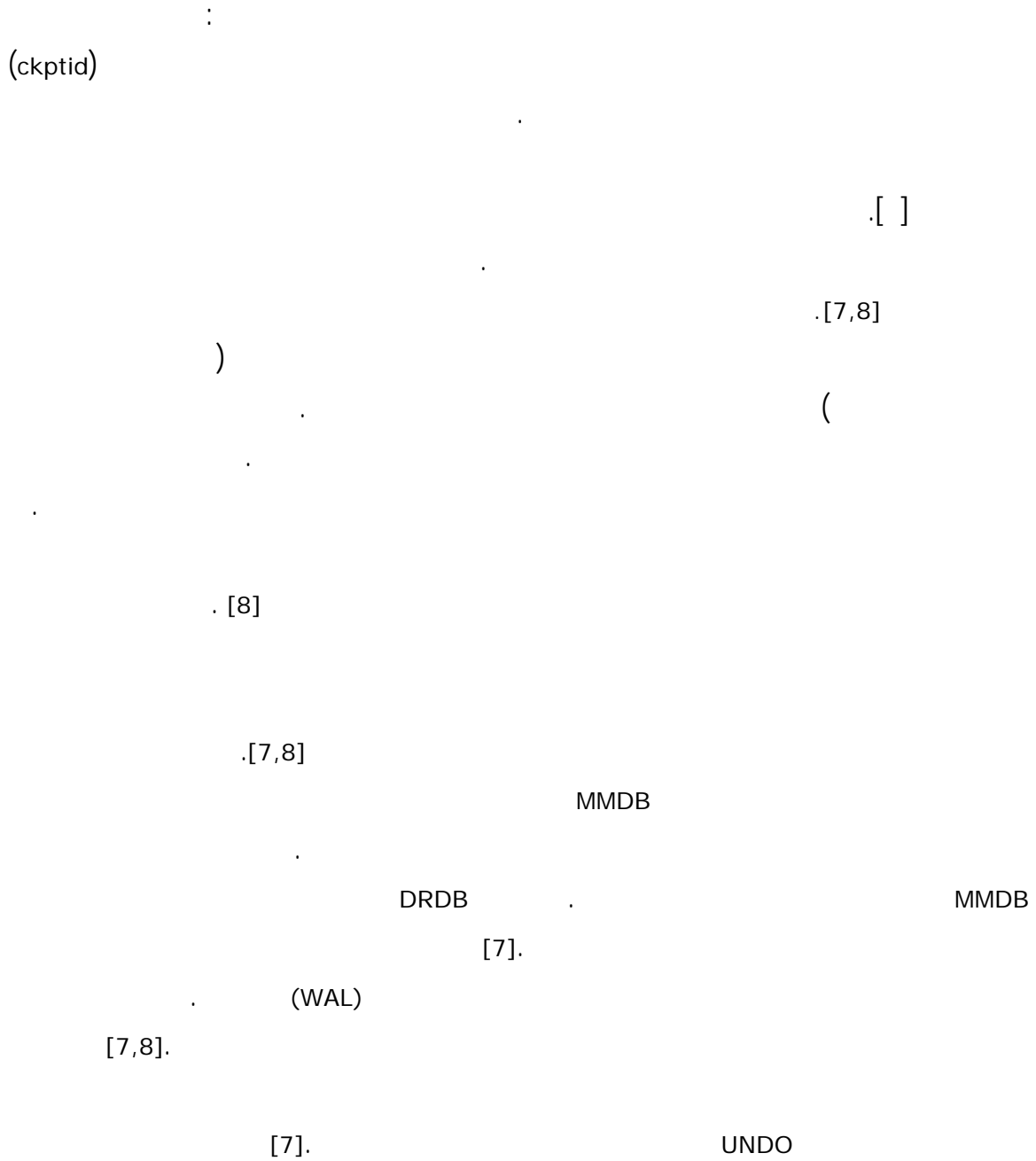
³ Reloading

⁴ Logging

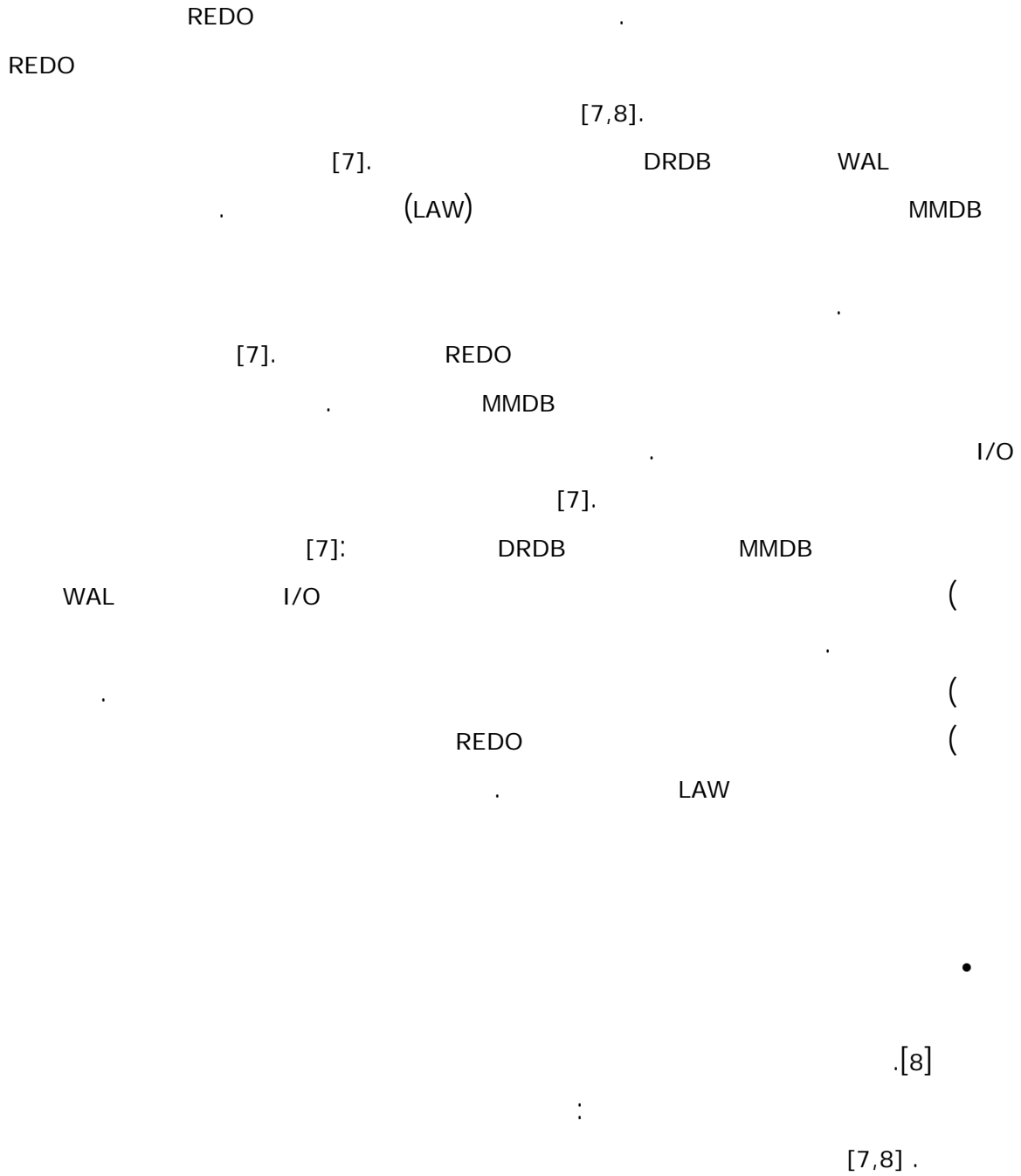
⁵ BFIM(Before Image)

⁶ AFIM(After Image)

⁷ Begin Transaction



¹ End Transaction
² Commit
³ Abort
⁴ Logical Logging
⁵ Physical logging
⁶ Fuzzy Checkpointing
⁷ Write Ahead Logging



¹ Commite Rule
² Logging After Writing
³ After Image
⁴ non-fuzzy checkpoint
⁵ log-driven checkpoint

(

[8].

[]

[7]. UNDO

[7].

UNDO

WAL

[]

Dali

REDO

(

MMDB

¹ Transation Oriented

² Action Oriented

³ Lehman

⁴ Carey

⁵ Salem

⁶ Garica-Molina

⁷ Update Transaction

⁸ Jagadish et al

⁹ Dirty Pages

¹⁰ Hagmann

¹¹ Checkpointer

[].

[7,8].

[7].

[1,3] .

().

[7].

(

[7].

LAW

(

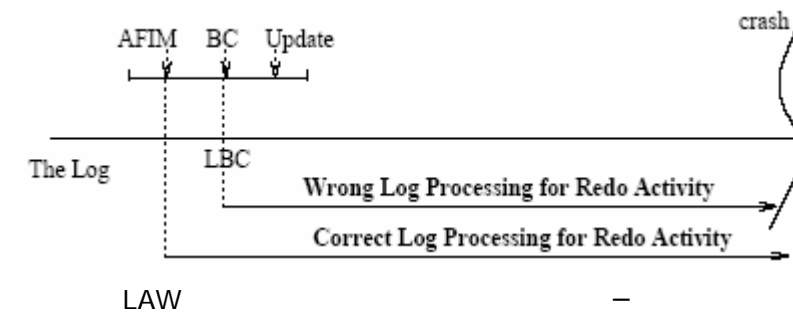
LAW

¹ Segments
² Ping-Pong Scheme
³ Dirty bit
⁴ Li et al

REDO

().

REDO



MMDB

[1].

[7]:

[8].

(ORP)

:

[7].(FR)

(SR)

:

commit

ORP

SR

¹ Page fault
² Simple Reloading
³ Concurrent Reloading
⁴ Gruenwald
⁵ Ordered Reload with Prioritization
⁶ Smart Reload
⁷ Frequency Reload
⁸ Waiting Transaction
⁹ Executing Transaction

AM

ORP

FR

throughput

FR

[7].

[1] Hector Garcia-Molina and Kenneth Salem , "Main Memory Database Systems: An Overview" , IEEE Transactions on Knowledge and Data Engineering, Vol. 4 ,No. 6 ,Page 509 , December 1992

[2] T. J. Lehman and M. J. Carey, "Query processing in main memory database management systems," in *Proc. ACM-SIGMOD Conf.*, Washington, DC, 1986, pp. 239-250.

[3] A. Aho, J. Hopcroft, and J. Ullman. *The Design and Analysis of Computer Algorithms*. Addison Wesley, Reading, MA, 1974.

[4] H. Lu, Y. Y. Ng, and Z. Tian, "T-tree or b-tree: Main memory database index structure revisited," in *Proceedings of the 11th Australian Database Conference*, 2000, pp. 65--73.

"() " **[5]**

[6] R. Rastogi, S. Seshadri, P. Bohannon, D. W. Leinbaugh, A. Silberschatz, and S. Sudarshan. Logical and physical versioning in main memory databases. In *The VLDB Journal*, pages 86--95, 1997

[7] Margaret H. Dunham , "Recovery In Main Memory Databases" Submitted to *International Journal of Engineering Intelligent Systems*, July 1996

[8] ANURAG GUPTA and HAN-YIN CHEN, "Recovery System for Main Memory Databases" *Computer Sciences Department ,University of Wisconsin, Madison – 53706, WI, May 14,1999*