



# Transaction Design for High Performance in Oracle Rdb

By

Magnus Weiman  
Digi-Key Corporation



# Transaction characteristics

- Access mode (aka transaction mode)
- Isolation level
- Wait mode
- Share mode (per table)
- Lock type (per table)
- Horizontal partition specification
- Constraint evaluation specification clause

Example) SET TRANSACTION READ WRITE RESERVING EMPLOYEES  
FOR SHARED READ WAIT 10 ISOLATION LEVEL READ COMMITTED;



# Access Modes

- READ ONLY
- READ WRITE (default)
- BATCH UPDATE



# Isolation levels

Isolation Level	Others allowed to update existing records	Others allowed to insert and delete
<b>SERIALIZABLE</b>	No	No
<b>REPEATABLE READ</b>	No	Yes
<b>READ COMMITTED</b>	Yes	Yes



# Wait modes

- WAIT [seconds] (default)

Waits for other transactions to complete and then proceeds. The default is to wait indefinitely.

- NOWAIT

Immediately returns an error when encountering a locked row



# Share Modes

- SHARED (default)

Others can read and write the table

- PROTECTED

Others can read the table but not write

- EXCLUSIVE

Others have no access to the table



# Lock Types

- READ  
Only reads are allowed
- WRITE  
Reads and writes are allowed
- DATA DEFINITION  
Allows multiple parallel updates to metadata for the same table under certain circumstances



# Horizontal partition specification

- Specifies one or more partitions so that only a subset of a table's partitions are reserved





# Constraint evaluation specification clause

Specifies the point at which the named constraint or constraints are evaluated

- VERB TIME
- COMMIT TIME



## Problem: Users waiting for SEQBLK block and TSN block

```
Node: YEW (1/4/16)      Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 16:30:22.96
Rate: 3.00 Seconds      Stall Messages      Elapsed: 00:08:21.96
Page: 1 of 1           V015:[USER.MAGNUS.TEMP]MF_PERSONNEL.RDB;1      Mode: Online
```

```
-----
Process.ID  Elapsed.... T Stall.reason.....Lock.ID.
23228B44:1  00:00:00.18 - waiting for SEQBLK block 1 (EX)      6105AAD7
23227B5B:1  00:00:00.18 R waiting for SEQBLK block 1 (EX)      2A0569CB
2323434F:1  00:00:00.14 R waiting for SEQBLK block 1 (EX)      57048109
2323554F:1  00:00:00.14 - waiting for TSN block 6 (EX)    1202C1C3
23220598:1  00:00:00.04 - waiting for SEQBLK block 1 (EX)    6A020C45
-----
```



## High I/O rates to the database root device...

OpenVMS Monitor Utility  
DISK I/O STATISTICS  
on node YEW  
21-SEP-2010 16:31:51.34

I/O Operation Rate			CUR	AVE	MIN	MAX
\$1\$DGA1978:	(YEW)	V056	0.00	0.00	0.00	0.00
\$1\$DGA1986:	(YEW)	V057	0.00	0.00	0.00	0.00
\$1\$DGA1994:	(YEW)	V015	467.52	449.45	399.21	505.84
\$1\$DGA2003:	(YEW)	V016	0.00	1.96	0.00	9.99
\$1\$DGA2012:	(YEW)	V017	0.00	0.00	0.00	0.00
\$1\$DGA2015:	(YEW)	V018	0.00	0.00	0.00	0.00
\$1\$DGA2018:	(YEW)	V019	0.00	0.00	0.00	0.00
\$1\$DGA2021:	(YEW)	V095	0.00	0.00	0.00	0.00
\$1\$DGA2031:	(YEW)	V905	0.00	0.00	0.00	0.00
\$1\$DGA2041:	(YEW)	V090	0.00	0.00	0.00	0.00
\$1\$DGA2067:	(YEW)	V5000	0.00	0.00	0.00	0.00
YEW\$DKA0:		YEW0	0.00	0.00	0.00	0.00



## ... and I/O queue

OpenVMS Monitor Utility  
DISK I/O STATISTICS  
on node YEW  
21-SEP-2010 16:32:57.36

I/O Request Queue Length	CUR	AVE	MIN	MAX
\$1\$DGA1978: (YEW) V056	0.00	0.00	0.00	0.00
\$1\$DGA1986: (YEW) V057	0.00	0.00	0.00	0.00
\$1\$DGA1994: (YEW) V015	0.33	0.33	0.00	0.66
\$1\$DGA2003: (YEW) V016	0.00	0.00	0.00	0.00
\$1\$DGA2012: (YEW) V017	0.00	0.00	0.00	0.00
\$1\$DGA2015: (YEW) V018	0.00	0.00	0.00	0.00
\$1\$DGA2018: (YEW) V019	0.00	0.00	0.00	0.00
\$1\$DGA2021: (YEW) V095	0.00	0.00	0.00	0.00
\$1\$DGA2031: (YEW) V905	0.00	0.00	0.00	0.00
\$1\$DGA2041: (YEW) V090	0.00	0.00	0.00	0.00
\$1\$DGA2067: (YEW) V5000	0.00	0.00	0.00	0.00
YEW\$DKA0: YEW0	0.00	0.00	0.00	0.00



## Average sustained transaction rate 900-1000 tps for 11 processes doing short READ ONLY transactions

Node: YEW (1/4/4) Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 18:26:05.10  
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:00:28.02  
Page: 1 of 1 V015:[USER.MAGNUS.TEMP]MF\_PERSONNEL.RDB;1 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	991	985	965.6	27057	1.0
verb successes	4944	4917	4822.5	135127	4.9
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	516	472	482.6	13525	0.4



## Database open on one node – no difference

Node: YEW (1/1/4) Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 18:29:32.28  
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:00:25.79  
Page: 1 of 1 V015:[USER.MAGNUS.TEMP]MF\_PERSONNEL.RDB;1 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	988	932	947.8	24454	1.0
verb successes	4919	4653	4730.8	122057	4.9
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	444	429	410.7	10597	0.4



## NUMBER OF CLUSTER NODES 1 >5000 tps, no root file I/O!

Node: YEW (1/1/1) Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 18:37:15.02  
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:00:44.33  
Page: 1 of 1 V015:[USER.MAGNUS.TEMP]MF\_PERSONNEL.RDB;1 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	6569	5743	5539.3	245614	1.0
verb successes	32834	28704	27686.9	1227641	4.9
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	0	0	0.0	0	0.0



## Problem – NUMBER CLUSTER NODES 1 prevents database from being open on more than one node

There are at least three different things you can do here:

- Open the database on one node and use remote access from other nodes
- Use locally attached storage for the database root
- Replace short READ ONLY transactions with READ WRITE SHARED READ (isolation level READ COMMITTED can be used to minimize locking impact)





## 11 processes doing short READ ONLY transactions via remote access

Node: YEW (1/1/1) ← Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 18:45:48.70  
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:00:33.60  
Page: 1 of 1 V015:[USER.MAGNUS.TEMP]MF\_PERSONNEL.RDB;1 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	6379	5386	5634.2	189367	1.0
verb successes	31890	26926	28160.7	946482	4.9
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	0	0	0.0	0	0.0



## 11 processes doing short READ ONLY transactions with database root on locally attached storage

Node: YEW (1/1/4) ← Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 18:53:58.63  
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:00:13.73  
Page: 1 of 1 YEW\$DKA0:[TEMP]MF\_PERSONNEL.RDB;1 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	3031	2606	2761.6	37917	1.0
verb successes	15149	13023	13801.4	189494	4.9
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	4687	3772	4115.8	56510	1.4



## 11 processes doing transactions READ WRITE SHARED READ ISOLATION LEVEL READ COMMITTED

Node: YEW (1/1/4) ← Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 19:09:13.06  
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:00:10.75  
Page: 1 of 1 V015:[USER.MAGNUS.TEMP]MF\_PERSONNEL.RDB;1 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	101759	101188	100157.5	1077695	1.0
verb successes	503904	501339	495856.1	5335412	4.9
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	4	0	1.1	12	0.0



## 11 processes doing transactions READ WRITE SHARED READ ISOLATION LEVEL SERIALIZABLE

Node: YEW (1/1/4) Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 19:22:19.86  
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:00:11.14  
Page: 1 of 1 V015:[USER.MAGNUS.TEMP]MF\_PERSONNEL.RDB;1 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	534410	534296	524624.5	5844318	1.0
verb successes	2122359	2122359	2069214.0	23051045	3.9
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	2	0	0.5	6	0.0



Here we do the same thing but actually reading a row per transaction

Node: YEW (1/1/4)      Oracle Rdb V7.2-410 Perf. Monitor 21-SEP-2010 19:28:29.23  
Rate: 3.00 Seconds      Summary IO Statistics      Elapsed: 00:00:11.11  
Page: 1 of 1      V015:[USER.MAGNUS.TEMP]MF\_PERSONNEL.RDB;1      Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	161569	161259	160647.2	1784791	1.0
verb successes	798041	777261	782381.9	8692263	4.8
verb failures	0	0	0.0	0	0.0
synch data reads	0	0	0.0	0	0.0
synch data writes	0	0	0.0	0	0.0
asynch data reads	0	0	0.0	0	0.0
asynch data writes	0	0	0.0	0	0.0
RUJ file reads	0	0	0.0	0	0.0
RUJ file writes	0	0	0.0	0	0.0
AIJ file reads	0	0	0.0	0	0.0
AIJ file writes	0	0	0.0	0	0.0
root file reads	0	0	0.0	0	0.0
root file writes	0	0	0.0	0	0.0



# Conclusions

- Setting the number of cluster nodes to 1 increased max TPS for READ ONLY transactions by a factor of 5
- Using local storage for the database root increased max TPS 3-10 times
- Using READ WRITE SHARED READ transactions increased max TPS 100 times or more!
- YMMV



## Summary: sustained number of transactions per second for 11 simultaneous database users

READ ONLY transactions NUMBER OF CLUSTER NODES > 1 Root file on SAN device Local database access	970
READ ONLY transactions NUMBER OF CLUSTER NODES = 1 Root file on SAN device Local database access	5,500
READ ONLY transactions NUMBER OF CLUSTER NODES = 1 Root file on SAN device Remote database access	5,600
READ ONLY transactions NUMBER OF CLUSTER NODES > 1 Root file on local device Local database access	2,700
READ WRITE SHARED READ transactions NUMBER OF CLUSTER NODES >1 Local database access	>100,000



# What are the downsides?

- Database open on only one cluster node can reduce availability and increase downtime after a node failure
- Using local storage for the database root might mean less redundancy
- READ WRITE SHARED READ might cause locking issues if transactions are not kept short. Requires changes in the application code.





# Additional findings

- If SAN disaster recover replication is disabled for the database root device then the maximum sustained DIO rate increases from around 1000 DIO/s to almost 4000.
- This virtually eliminates the I/O bottleneck for our database root.



# Questions?

[magnus.weiman@digkey.com](mailto:magnus.weiman@digkey.com)