Building High Performance Queues in Rdb Databases

Thomas H. Musson ♦ Jeffrey S. Jalbert ♦ Cheryl P. Jalbert
Keith W. Hare ♦ Jeff Haidet

JCC Consulting, Inc.

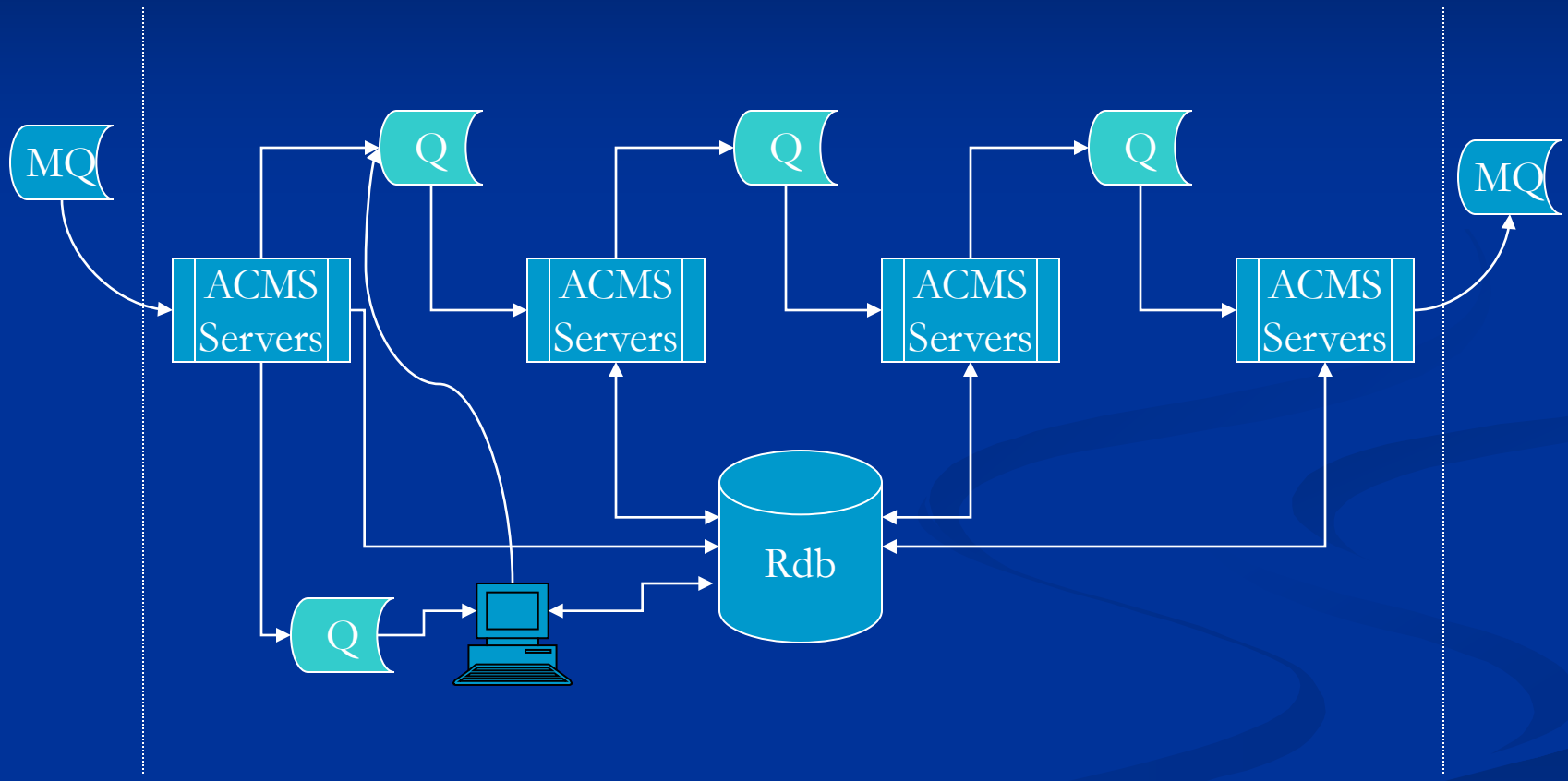
Copyright 2009 JCC Consulting, Inc. All rights reserved.

Definition of Problem

- Business expected to increase by 100% - 1000%
- Application Description
  - Flow-through design
    - Queues used to migrate data through application
  - Highly scalable
  - Mirrored data centers for primary DR
  - Hot Standby used for secondary DR
- Current Queuing Mechanism
  - ACMS Queues
    - RMS-based
    - Estimated 50% of system I/O
    - a.k.a. "The Bottleneck"

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Application Design



Copyright 2009 JCC Consulting, Inc. All rights reserved.

Application Performance - Before

- Production servers
  - rx8640
- Transaction durations (average for application)
  - Total average 0.128 seconds (128ms)
  - Read only = 0.163 seconds (163ms)
  - Read write = 0.108 seconds (108ms)
- Multiple queue operations per transaction

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Queuing Mechanism

- ACMS Queues
  - Uses RMS journaling
  - 2-phase commit
  - Synchronous I/O
  - Optimal enqueue/dequeue = 0.015 seconds (15ms)
  - Average enqueue/dequeue = 0.025 seconds (25ms)
- Hot Standby DR implications
  - Replicates Rdb changes
  - Queues must be rebuilt on fail-over

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Replacement Requirements

- Many enqueuers
- One dequeuer (per queue)
- FIFO by process
  - Entry is not on Queue until commit
  - Dequeue order can be interleaved among processes
- Resilient to failures

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Solution – Database Queues

- Eliminates
  - RMS
  - 2-phase commit (where possible)
- Rdb performance and reliability
  - Tuning capabilities
  - After Image Journals
  - Replication to standby
- In-memory support structures
- RTV – Real Time Viewer

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Results

- Performance
  - T4
  - RTV – Real Time Viewer
- Integration

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Performance - After

- Development machines
  - rx4640
  - Queues elapsed time ‘too small to measure’
  - Most frequent application transactions
    - Perform other database and application work in addition to queue work
    - Development (new queues) = 0.025 seconds (25ms)
    - Production (legacy queues) = 0.200 seconds (200ms)

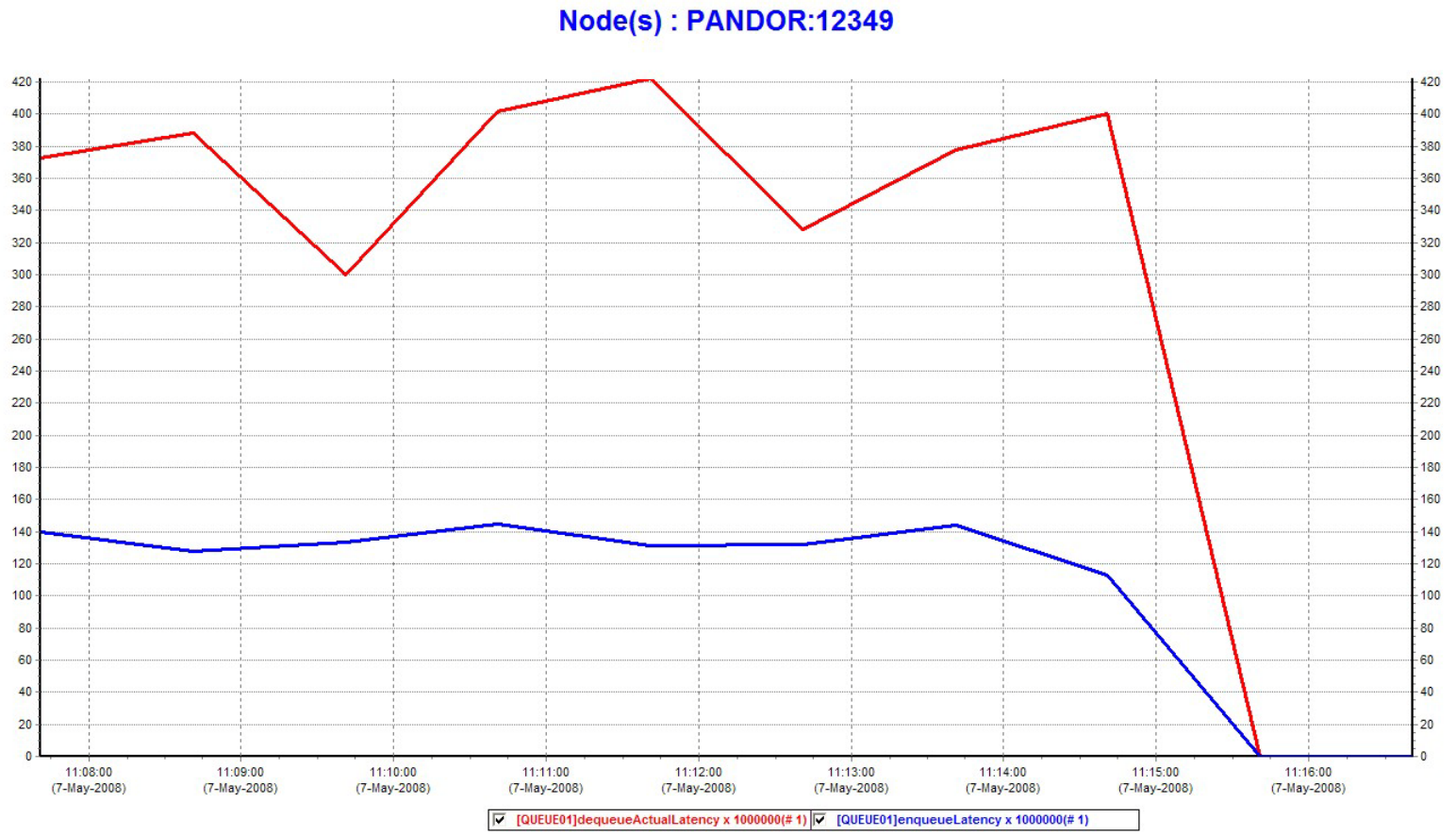
Copyright 2009 JCC Consulting, Inc. All rights reserved.

Performance

- Regression testing measurements demonstrate
  - BL870c
  - Enqueue = 0.000135 seconds (135ns)
  - Dequeue = 0.000370 seconds (370ns)
  - Dequeuing is so *SLOW*… ☺

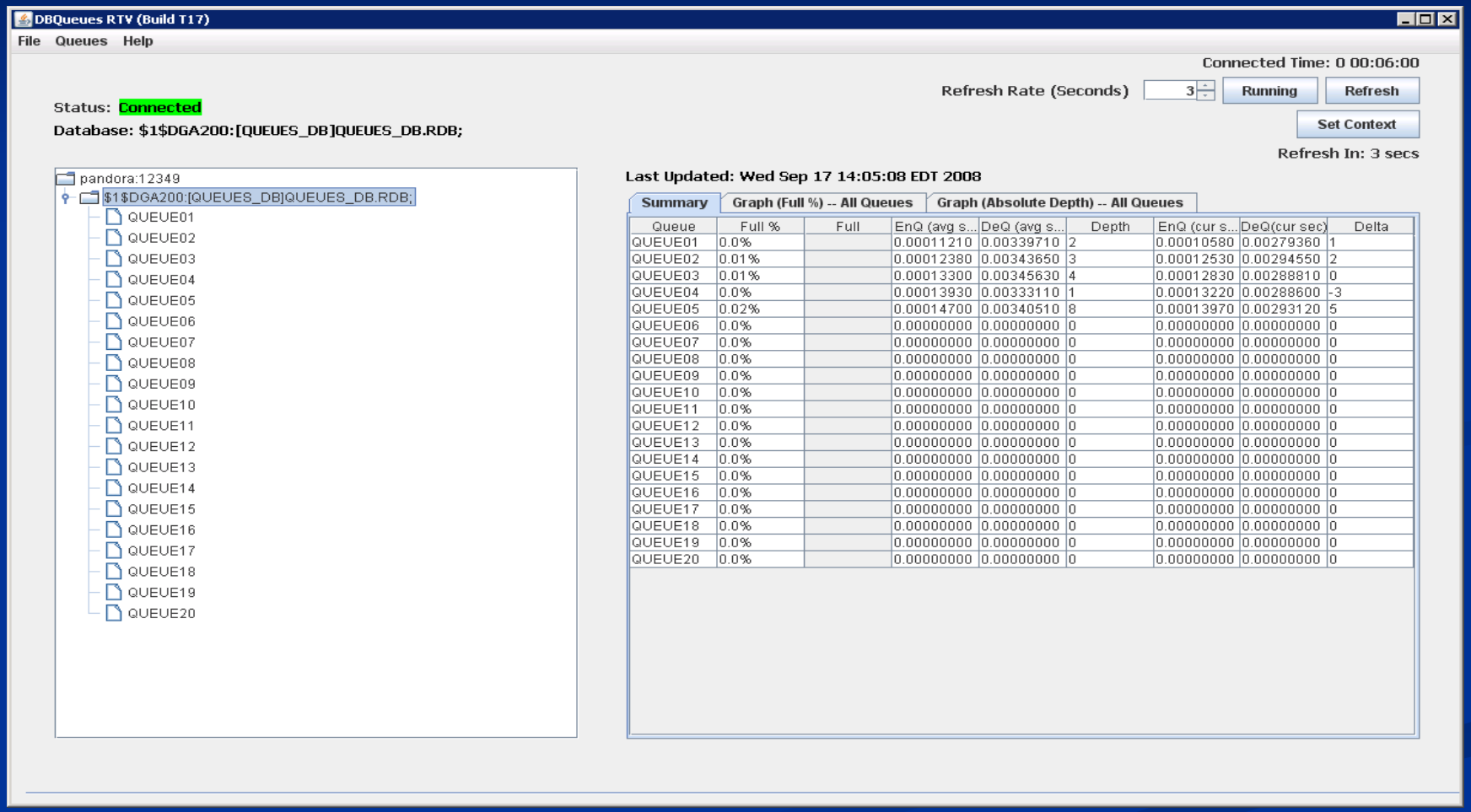
Copyright 2009 JCC Consulting, Inc. All rights reserved.

T4 – Enqueue vs. Dequeue



Copyright 2009 JCC Consulting, Inc. All rights reserved.

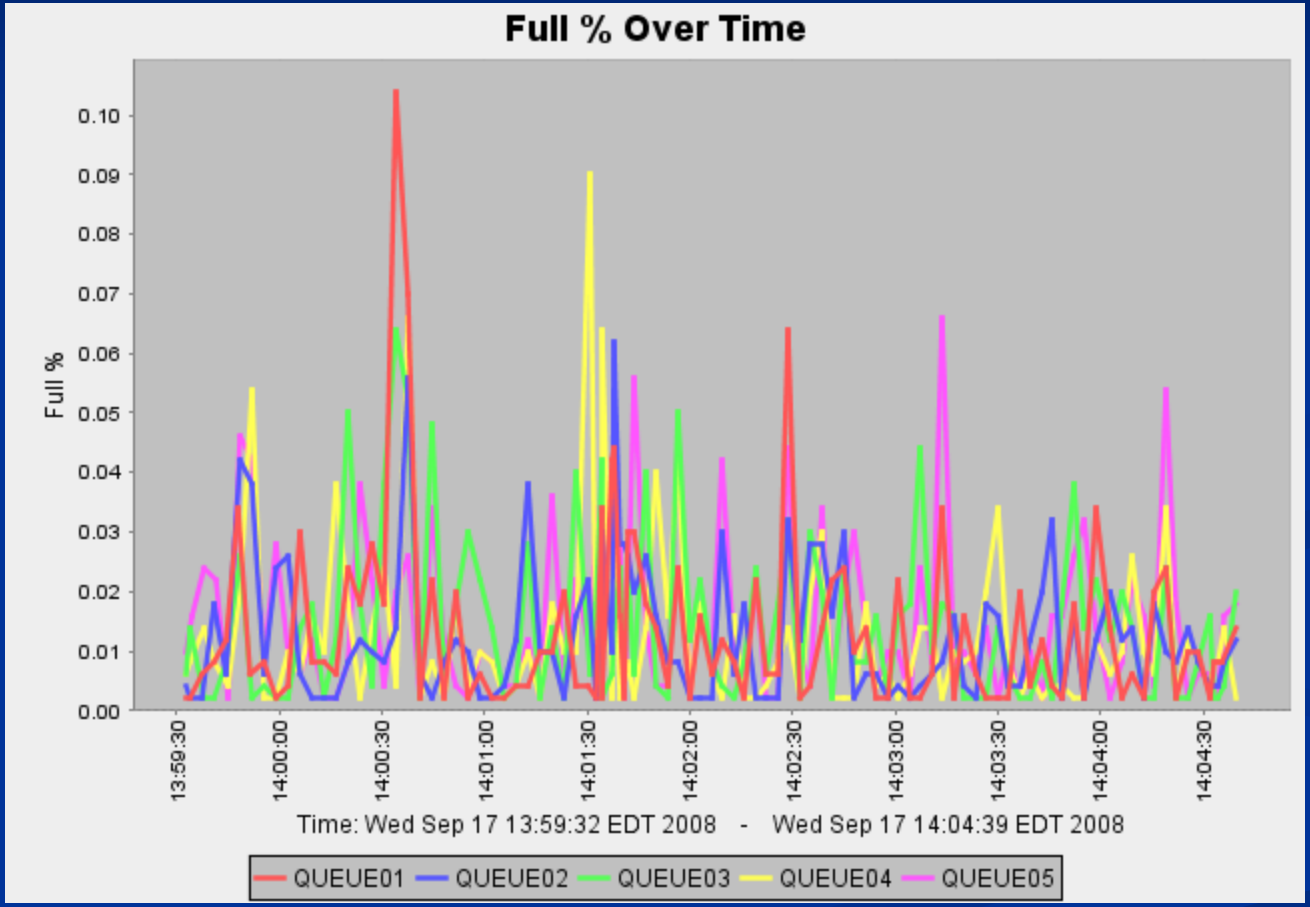
RTV – Real Time Viewer



Queue	Full %	Full	EnQ (avg s...	DeQ (avg s...	Depth	EnQ (cur s...	DeQ(cur sec)	Delta
QUEUE01	0.0%		0.00011210	0.00339710	2	0.00010580	0.00279360	1
QUEUE02	0.01%		0.00012380	0.00343650	3	0.00012530	0.00294550	2
QUEUE03	0.01%		0.00013300	0.00345630	4	0.00012830	0.00288810	0
QUEUE04	0.0%		0.00013930	0.00333110	1	0.00013220	0.00288600	-3
QUEUE05	0.02%		0.00014700	0.00340510	8	0.00013970	0.00293120	5
QUEUE06	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE07	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE08	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE09	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE10	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE11	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE12	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE13	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE14	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE15	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE16	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE17	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE18	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE19	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0
QUEUE20	0.0%		0.00000000	0.00000000	0	0.00000000	0.00000000	0

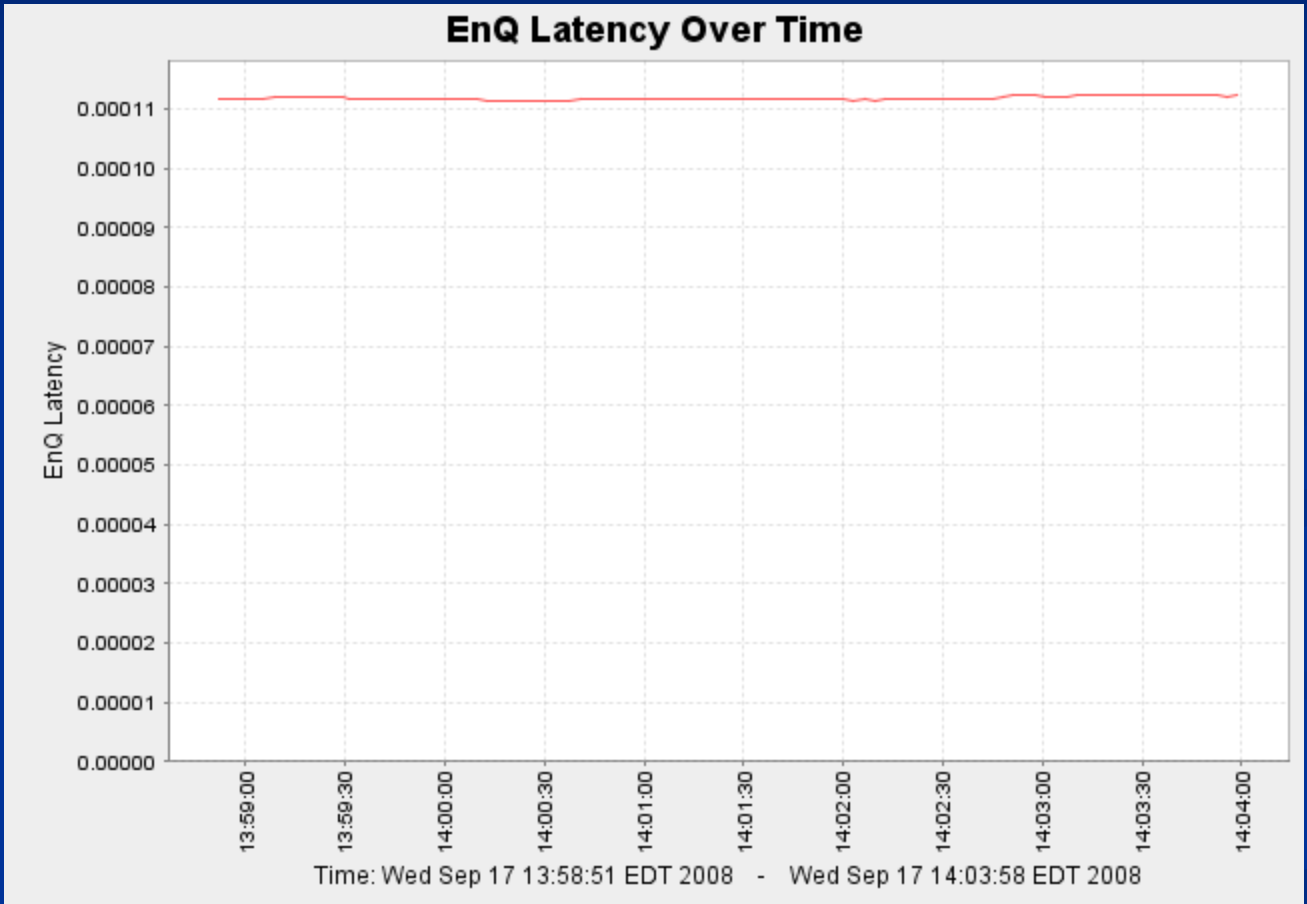
Copyright 2009 JCC Consulting, Inc. All rights reserved.

RTV – Real Time Viewer



Copyright 2009 JCC Consulting, Inc. All rights reserved.

RTV – Real Time Viewer



Copyright 2009 JCC Consulting, Inc. All rights reserved.

Integration

- Issues
  - Database handle
  - 100% Dynamic SQL
- Must be linked with application

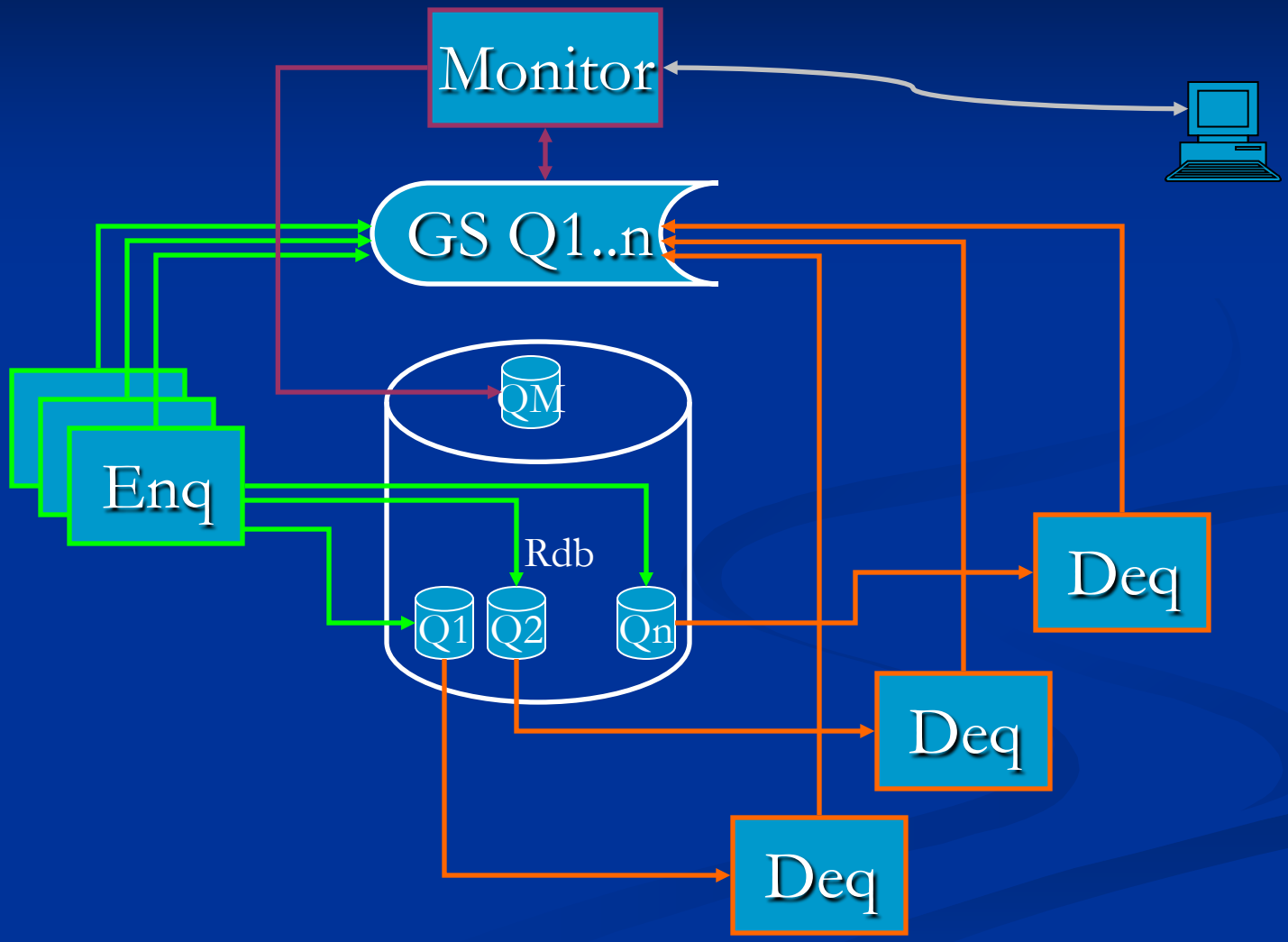
Copyright 2009 JCC Consulting, Inc. All rights reserved.

How We Did It

- Technical details
- Processes
- Regression test

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Technical Details



Copyright 2009 JCC Consulting, Inc. All rights reserved.

Technical Details

- Database structures
  - Queue Management table
    - One row per Queue
    - Queue configuration information
  - Queue data table
    - One table per Queue
    - One row per available queue entry
    - Primary Key index

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Technical Details

- Queue data table
  - Pre-populate rows
  - Rows utilized as circular queue
  - Represents the backing store for the queue
  - Queued data
  - Flag to indicate queued data or not
  - Timestamps, Usernames for Enqueue, Dequeue

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Technical Details

- Global Section
  - All queues represented in a single section
  - Each queue has:
    - Configuration information
    - Current enqueue/dequeue/reclaim locations
    - Performance data
    - Enqueued bitmap
    - Dequeued (not really) bitmap
  - Locks used to coordinate update access

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Processes

- Monitor
- Enqueue/Dequeue Clients
- RTV
- Queue Scripting

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Monitor

- Responsibilities
  - Populate Global Section from database
  - Reclaim dequeued sections of bitmaps
  - Collect performance data
  - Provide data to Real Time Viewer
    - TCPIP interface
    - Can be disabled if necessary
  - Write T4 data files
  - Client exception/debug logging

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Enqueue/Dequeue Clients

- Enqueue API
  - Get ENQ lock
  - Get entry lock on GS next available
  - Increment GS next available entry
  - Release ENQ bitmap lock
  - Update Queue table entry

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Enqueue/Dequeue Clients (cont.)

- Dequeue API
  - Get DEQ lock
  - Get GS entry starting position
  - Compare ENQ and DEQ bitmaps for potential entries
    - ENQ set, DEQ unset
  - Attempt to get entry lock
    - Bypass if lock not available (locked for ENQ or DEQ)
    - Read Queue row if lock available
      - Set DEQ bitmap entry if not Enqueued
  - Continue to end of Queued data (ENQ bit set)

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Enqueue/Dequeue Clients (cont.)

- Commit and Rollback API
  - Execute Commit/Rollback
    - Or sysend_transw/sysabort_transw if 2PC context
  - Release resources
    - Enq entry locks
    - Deq entry locks

Copyright 2009 JCC Consulting, Inc. All rights reserved.

RTV

- Request real-time data from the Monitor process
  - TCPIP requests
  - XML returned
- Display returned data
- Plot data over time

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Queue Scripting

- Builds SQL scripts
  - Add Queue Management table, if necessary
  - Add Queue Management row
  - Add Queue table
  - Create Queue index
  - Create Row Caches
  - Create Storage Areas
  - Populate Queue rows

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Regression Test

- Uses 5 queues with radically different payload sizes
- Workload is randomized
  - Writes to random queues
  - May write to multiple queues per transaction
  - Number of entries to each queue is random
  - Uses random choice of distributed or simple transactions
- Runs randomly either wait or do not wait between transactions
  - Tests conditions when queues are both full and empty since dequeue is more costly than enqueue
- Performs only queue work

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Regression Test

- 5 separate processes make entries randomly to the 5 queues
- Each queue has a max of 50,000 entries
- Each session enqueues about ½ million entries to each queue
- One process per queue dequeue entries and write to dedicated Rdb tables
- Payload includes information to allow proper analysis of dequeue ordering and completeness after a session

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Regression Test

- Post-processing analyzes all results and ensures that all entries that were enqueued were also dequeued in the correct order
- Upon success, the regression cycle repeats

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Regression Test

- With Fast Commit, commit to journal and OPT the regression test was so fast that ALS had difficulties keeping up
  - Rdb engineering repaired some bugs for us

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Recent Changes

- Enqueue support in SQL Stored Procedures
  - Allows for enqueues from remote SQL and JDBC clients

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Documentation

- Describes
  - Goals
  - Philosiphy
  - Use
- Complete documentation is available on request
- It is the usual JCC style and completeness

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Directions

- Generalize to shareable image
- VMS authentication for RTV
- Other ideas?

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Acknowledgements

- Thanks to Rdb engineering for their support and counsel
- Thanks to our Customers for making this possible

Copyright 2009 JCC Consulting, Inc. All rights reserved.

Questions

http://www.jcc.com/

Join the worldwide Rdb community. Send mail to

OracleRdb-request@JCC.com with "SUBSCRIBE" in the body of the message.

For more information send mail to info@jcc.com

Copyright 2009 JCC Consulting, Inc. All rights reserved.