Agenda

- Existing performance implications
  - Database access in cluster
  - SMP scaling
- Recent performance enhancements
- Rdb’s Galaxy implementation
Rdb at the High End

- Database servers pushing technology limits
  - Applications utilizing more disk space, more I/O, more memory and more CPU
  - Configurations limited by accessible CPU power and I/O capability
Cluster Performance Cost

- Application performance suffers vs. running on single node
- Remote locking overhead
- Inability to take advantage of
  - Internal optimizations
  - Global buffer effectiveness
  - Rdb Performance Features
    - Row Cache
    - Page Transfer Via Memory
Distributed Locking Costs
GS140 - CI Interconnect (estimated)

Network
Remote CPU
Local CPU
Cluster Access with Database Open on Single Node

Application Front-end Servers…

Database access via Rdb remote server from Standby to Master

AlphaServer

AlphaServer

AlphaServer

Database

Hot Standby

Standby Database
SMP System Limitations

- Some Rdb database servers are effectively limited in SMP system
  - Additional CPUs yield no performance improvement

- “MPSYNC” / Spinlock contention
  - CPU time spent waiting for another CPU
  - Lost CPU resources - wasted potential
  - Significant OpenVMS V7.3 improvements in this area (dedicated lock manager, SCSI/Fibre fast path, etc)

- Primary CPU handles cluster, remote lock & I/O traffic
  - Primary processor saturation
  - Fast Path provides help in some cases
Application OLTP Simulator

- Replicate behavior of customer application
  - Gathered and analyzed RMU /SHOW STATISTICS and MONITOR binary data
    - >90% R/O transactions
    - Row Cache; little actual database I/O
  - GS140 / 8 CPU / EV6 / 8gb / VMS V7.2-1

- Program simulating actual production system database access patterns
  - Lookups / Modifies per transaction
  - Averaged think times
  - Variable workload

- Nearly duplicates actual production statistics
Result 1
Root Object Sharing

- Control structure object sharing optimizations when “NUMBER OF CLUSTER NODES IS 1”
- Significant performance improvement for high read-only transaction rates
- Reduced locking activity and root file I/O for frequently accessed objects
  - SEQBLK
  - TSNBLK
- Introduced with Rdb V7.0.4
Rdb 7.0.4

Throughput Increase...
...Idle Time Doubled

[Graph showing the relationship between Users and Transactions Per Second for different versions of Oracle, with a note indicating that the idle time has doubled.]
MPSYNC Halved

60% MPSYNC = 5 CPUs wasted!
Result 2
Read-Only Txn Start/Commit

- Read-write start/commit optimizations for read-only
- Reduced TSNBLK I/O and locking for TSNBLKs in high contention situations
- Lab measured > 8x improvement transaction start/commit throughput
- Customer measured 5x transaction duration improvement in production
- Introduced with Rdb V7.0.5
However...

- Need more than CPUs & Memory
- Need more primary CPUs
  - Network I/O
  - Non-fast path I/O
- Nicer to be able to run in cluster
  - Upgrades
  - Isolation
  - Node failure protection
What is OpenVMS Galaxy?

- Software architecture
- Multiple OpenVMS instances execute cooperatively in single computer
- Dynamic resource allocation between instances
- Memory shared between instances
- *Capable* of “Cluster in a box”
  - Move from existing single system environment
Each OpenVMS Instance...

- 1 or more CPUs
- 1 or more I/O adapters
- Private memory
- Access to “Galactic” shared memory
- Boot/Shutdown independently
Rdb 7.1 & Galaxy

- Natural growth of existing Rdb cluster sharing
  - Galactic shared memory for
    - Global Buffers
    - Row Caches
    - Shared database objects

- Leverage Galaxy technology for performance
  - Greater effective scaling
  - Reduced locking and I/O
  - More Primary CPUs

- High-end database server
Rdb Sharing in Galaxy

- Each Instance has
  - Rdb Monitor (RDMMON)
  - Database Recovery Server (RDMDBR)
  - AIJ Buffers and AIJ Log Server (RDMALS)
  - “NODGBL” global sections
    - Statistics

- Shared
  - “TROOT” global sections
  - Database root objects
  - Global buffers
  - Row caches & Row Cache Server (RDMRCS)
Setting up Rdb and Galaxy

- No application changes required
- Configure Galaxy environment
  - Enough shared memory for existing database global sections
- Set database parameters
  - NUMBER OF CLUSTER NODES
  - GALAXY
- Open database on each instance
Operational Considerations

- All instances must use *identical* Rdb software
- Many RMU operations are per-instance
  - SHOW STATISTICS
  - SHOW USERS
  - Monitor/Server Start & Stop
- Database Open / Close on each instance
  - RCS started on first instance to open database
  - Must close database last on instance with RCS
  - Use manual open mode
- DBR or RCS failure shuts down database
Rdb Galaxy Advantages vs. Traditional Clustering

- Significantly less root I/O, locking & blocking ASTs
  - Shared TROOT objects
  - Much lower overhead for R/O transaction start/stop
  - Reduced spinlock (MPSYNC) contention

- Reduced database I/O & page locking
  - Row caches shared among instances
  - Global buffers shared among instances

- Most rapid cluster interconnect
  - VMS uses shared memory for node-node SCS traffic

- Row cache runs in cluster!
RMU/SHOW SYSTEM Example

CLICK$ RMU/SHOW SYSTEM

database DGA5:[DB]RND.RDB;1
- opened 18-APR-2001 14:59:32.38 (elapsed 0 01:01:47)
- current after-image journal file is DGA5:[DB]RND.AIJ;1
- global buffer count is 6911; 5105 global buffers free
- maximum global buffer count per user is 86
- global section resides in OpenVMS Galaxy shared memory
- all database users use the same copy of shared memory
- 11 active database users on this node
- database is also open on the following node:
  - BOING as monitor ID 2 - Galaxy
More availability, Performance and Scalability with Rdb & Galaxy

FibreChannel/HSG80 Disk Interconnect

Hot Standby

GS320

Shared Database Structures

Global Buffer Pool

Row Caches

48gb shared memory

WLDFR1
8 CPU
4gb private

WLDFR2
8 CPU
4gb private

WLDFR3
8 CPU
4gb private

WLDFR4
8 CPU
4gb private
Mixing Galaxy and non-Galaxy

- Both Galaxy and non-Galaxy nodes in cluster can access database
- Not all features available in this environment
  - Row Cache
  - Page Transfer Via Memory
  - Some TROOT optimizations
  - All require direct access to common memory
Availability

- **Oracle Rdb**
  - Oracle Rdb V7.1 ~ Q3 CY2001
  - Oracle Rdb V7.1.0.1 ~ Q4 CY2001
  - Oracle Rdb V7.1.0.2 ~ Q1 CY2002
  - Rdb’s Phase I of Galaxy support

- **OpenVMS Galaxy**
  - OpenVMS V7.2 ~ Q1 CY1999
  - OpenVMS V7.3 ~ Q2 CY2001
Questions? Comments?

norman.lastovica@oracle.com