



Whitepaper / Benchmark

Oracle RAC performance quadruples on HP cluster with Dolphin Express

Pierre Orzechowski, Hewlett-Packard ESS
Wiekus Beukes, Hewlett-Packard ESS
Einar Rustad, Dolphin Interconnect Solutions
Friedrich Siefert, Dolphin Interconnect Solutions

Dolphin Interconnect Solutions

February 27th, 2008

1 EXECUTIVE OVERVIEW

Dolphin Express with SuperSockets provides a scalable high performance interconnect solution for a horizontal Scale-Out approach for enterprises using Oracle 10g and Oracle Real Application Cluster (RAC). This paper discusses the advantages of using SuperSockets with Dolphin Express to build scalable, high performance and highly cost-effective Oracle RAC clusters with industry standard servers and Dolphin components.

Large database systems have traditionally been run on high-cost mainframe systems. With modern hardware and software components built on standard high volume technologies, commodity servers can be interconnected to form clusters and grids that provide huge cost savings compared with mainframe systems. Coupled with a high performance, low latency interconnect, clusters and grids can provide similar performance levels to that of mainframes at a fraction of the cost. The features of Oracle RAC coupled with the fault-tolerant features built into the Dolphin Express hardware and the SuperSockets software provide highly available shared database solutions on clusters built from industry standard servers.

The benchmark results in this paper show up to a **quadrupling of performance** when using Dolphin Express compared to gigabit Ethernet, both with regards to the number of transactions per second and the response time experienced by the user.

2 WHY ORACLE WILL RUN FASTER WITH DOLPHIN EXPRESS

Oracle RAC performance can be heavily influenced by the cluster interconnect. If the workloads are interconnect intensive, the performance of Oracle RAC systems with Gigabit Ethernet will suffer from bottlenecks in latency and overhead in the cluster messaging and the Oracle cache fusion traffic. This traffic consists in nature of relatively short individual transfers and those are inherently limited in performance by interconnect latency and protocol overhead. Dolphin Express with SuperSockets address this through unique hardware architecture with features that minimize latency and remove protocol overhead. Dolphin Express with SuperSockets reduces basic process-to-process latency in the order of a factor of ten compared with Gigabit Ethernet for socket based communication with a ping-pong half latency as low as $2.2\mu s$. This is done without compromising standard socket functionality; SuperSockets is compatible with applications using TCP, UDP or RDS, so Oracle RAC can be run right out of the box with the standard UDP protocol or with RDS through the RDS interface developed by Oracle.

Dolphin Express features include both RMA (Remote Memory Access) and RDMA (Remote Direct Memory Access). The RMA is used for short data transfers and has proven extremely efficient for reducing latency. The data to be transferred are normally located in the L1 cache and the RMA functionality allows the CPU to store 8 bytes to the remote memory in a single CPU cycle. With a 3 GHz processor, this is just a fraction

above 300 picoseconds. Since Dolphin Express has built-in end-to-end data integrity features that ensure that transfers complete successfully without corruption or data loss, no additional transaction protocol layer is necessary spending correspondingly fewer processor cycles for error checking.

In the unlikely event of a severe hardware malfunction, the error checking ensures that fault-tolerance software is invoked to manage retransmission of data across an alternative connection, either through an alternative path through the interconnect fabric, or through a secondary fabric and eventually through an Ethernet connection. All this is done automatically and fully transparent to the application.

3 THE BENCHMARK SETUP

To document the performance improvements with Dolphin Express, HP provided a set of ProLiant DL385 G2 servers that were equipped with the Dolphin Express host bus adapter cards and Dolphin's SuperSockets software library.

The benchmarks were run at the Hewlett-Packard Solution Alliance Engineering lab in Cupertino, California for the purpose of showing users of Oracle RAC environments how much performance improvement can be expected in their own systems by using industry standard Hewlett-Packard servers and Dolphin Express interconnects.

The benchmark consists of a block-transfer test, provided by the Oracle RAC Development team, which is used by Oracle RAC development for the purpose of determining interconnect performance. The database is an Oracle 10g RAC database running on a two-node cluster.

Server type	HP ProLiant DL385 G2
CPU	2x dual-core AMD Opteron 2218, 2.6 GHz
Interconnect	Dolphin Express D352
Operating system	RedHat Enterprise 4 Update 4, x86_64

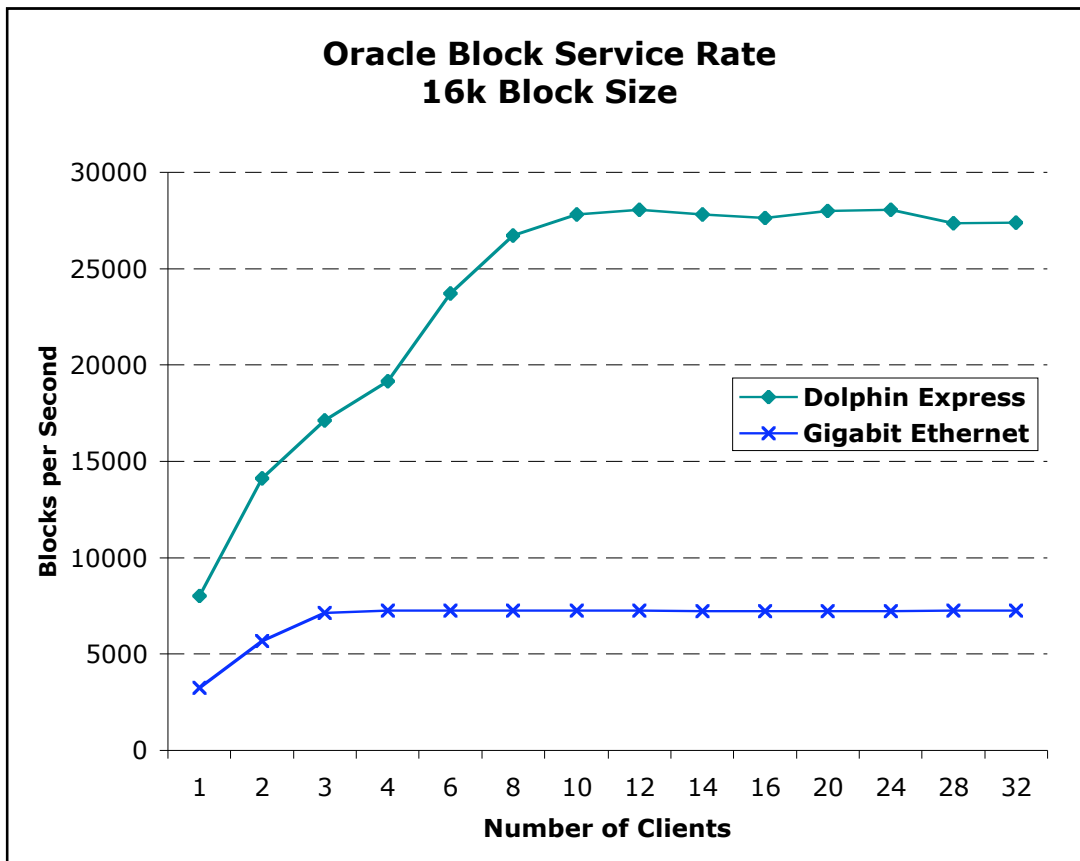
The transparency provided by the Dolphin SuperSockets library allowed for running the Oracle developed tests without any modifications to any of the Oracle software modules.

4 BENCHMARK RESULTS

The results are presented in the below two graphs comparing performance to Gigabit Ethernet. The results show significant performance improvements when running Dolphin Express with RDS.

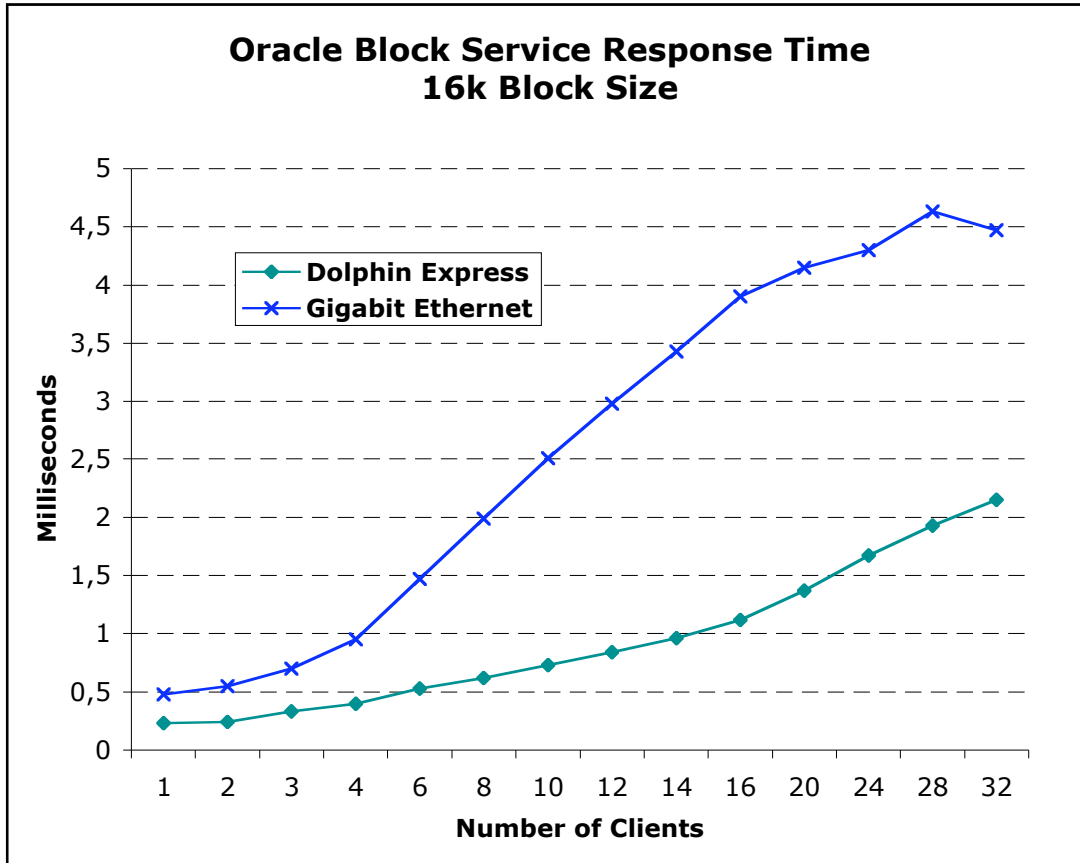
The test results are comparing multi-block messaging properties of Dolphin Express with Gigabit Ethernet. The charts show how the database performs when a client database instance asks other database instances in a cluster to process data-blocks. More specifically, the test uses an internal database process for passing data called Lock Manager Server (LMS). LMS is performed by the database's Global Cache Service (GCS) and is the service responsible for processing GCS requests from remote Oracle instances, and for sending requested data-blocks back and forth between nodes in the RAC cluster. The benchmark uses 8 LMS server-processes.

The first chart below shows throughput measured in blocks-per-second vs. number of client processes for 16k size blocks.



This proves that the Dolphin Express solution is almost quadrupling the number of 16k size blocks being transferred between the nodes in the RAC database.

The chart below shows response times for different number of client processes for 16k size blocks.



Again we are seeing that Dolphin Express provides a close to 400% improvement in response time over Gigabit Ethernet on the Oracle RAC database.

5 CONCLUSION

The benchmark in this paper has shown a close to 400% performance improvement (both throughput and response time) over gigabit Ethernet when deploying Oracle RAC on an HP ProLiant cluster with Dolphin Express interconnect.

These results are explained by the big increase in transfer rate for the relevant data block sizes used by Oracle when comparing Dolphin Express with gigabit Ethernet. The Dolphin technology utilizes Remote Memory access to deposit data directly from the sending process into remote memory through CPU store operations, effectively avoiding overhead and reducing latency.