WHITEPAPER

January 2009

Virtualizing Data Center Memory for Performance and Efficiency

1.0 Introduction

Data centers everywhere face ongoing demands for higher performance and greater efficiency. Cost-efficiency is a primary concern as IT managers seek to enhance performance while reducing overhead, and these managers must also find ways to constantly improve performance for complex applications.

Every aspect of data center infrastructure plays a role in performance, including the speed and number of CPUs, the amount of memory available, the size and performance of storage systems, and the speed and efficiency of the network connecting them. High-demand applications rely on clustered computers to distribute the processing load across multiple CPUs. But while CPU performance, memory, storage, and connectivity all contribute to higher performance across a cluster, the most constrained piece of the infrastructure will always limit total performance.

Unfortunately, not every component of data center infrastructure has experienced the same performance improvements over time. CPU performance has continued to increase according to Moore's Law; and the size and cost of storage have made dramatic improvements with increases in storage density; but access to sufficient memory and the speed of moving bytes across the data center have not kept up. Consequently, the availability of memory and the speed of connectivity among systems remain the true bottlenecks to performance in the data center.

Memory virtualization and high-performance connectivity offer a transparent solution to this problem that delivers order-of-magnitude performance gains while reducing costs and power consumption. This paper discusses how memory virtualization and high-performance connectivity work together to improve data center performance while reducing total cost of ownership.

Data center managers are under constant pressure to improve performance. Increasingly large data sets, increasingly complex applications, and server virtualization are all placing huge new demands on performance. Clustered computing environments are a key strategy for meeting these performance demands.

Using this approach, IT architects can improve performance over time by upgrading to faster multi-core CPUs or adding higher-performance storage systems. Multi-core server performance continues to improve, and storage capacity continues to increase. But achieving the highest performance from clustered computing systems requires shared access to all resources, including memory, and it is gaining access to sufficient memory that is the primary roadblock to cost-effective cluster performance.

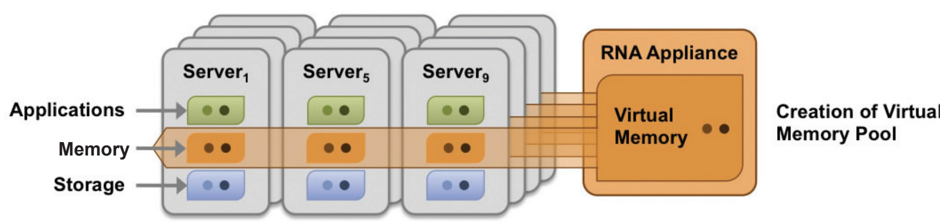
In today's clusters, applications may spread the processing load across multiple CPUs and servers, and servers all have access to a common pool of storage, but each server can access only its own local memory. As a result, servers are limited in the amount of data they can process at a time, and in the time it takes to process that data, regardless of processor or storage system need.

For many applications such as engineering, oil and gas exploration, complex financial modeling, and life sciences research, data sets can run to a terabyte or more, and servers lack sufficient memory to store the entire data set at once. Instead, the server must rely on mechanical disk drives with multi-millisecond access times to retrieve and provide the data, which significantly multiples the processing time.

To date, the typical approach to improving cluster performance is to scale servers or to upgrade CPUs, servers, or storage systems. In many cases, the number of servers or storage nodes is over-provisioned to deliver the needed performance, but the use of legacy connectivity solutions and the continuance of fragmented server memory maintain an insurmountable performance barrier. Over-provisioning servers and storage increases capital and operational costs, but doesn't solve the whole problem.

1.1 Memory Virtualization

RNA Networks' Memory Virtualization Platform™ (MVP) solutions eliminate memory segmentation by creating a pool of shared memory for the entire cluster. MVP combines memory from dedicated appliances with memory from existing servers to create a large memory cache (up to 10 Terabytes) which can be transparently shared among any or all servers or data center in a cluster. It provides a shared network memory resource large enough to accommodate huge data sets, significantly reducing storage access and improving processing performance by orders of magnitude.



Server clusters already rely on remote direct memory access (RDMA) to divide application tasks among CPUs and servers, and RNA's MVP builds on RDMA to provide the fastest possible memory access for all servers in the cluster at any time. In fact, MVP can also scale to very high bandwidth by incorporating portions of each server's memory as part of the virtualized pool. Integrating memory from multiple servers also makes it very easy to scale up the amount of virtual memory available without disrupting operations.

MVP integrates at the file system or block level or the application level. At the file system level, virtualized network memory is mounted like a drive resource, and applications access cached data with no changes whatsoever. But in contrast to data stored on a disk, data in virtualized memory can be accessed in microseconds rather than milliseconds.

Shared memory also saves power costs. Because MVP provides a large amount of memory to any server, it becomes possible to deploy less memory per server across the cluster. MVP also enables fewer servers to accomplish the same level of application performance, so a cluster requires less rack space, less power, and less management labor.

Another advantage to using the MVP is that it uses a copy of the data set from storage to provide full fail-over protection in case the application or cluster fails. The data in memory is a working copy, while the original data remains untouched untouched in the existing storage subsystem.

1.2 High-Performance Connectivity

As an appliance, RNA's MVP is easy to deploy and administer, and adds very little complexity or administrative overhead in the data center. It delivers order-of-magnitude improvements in cluster processing speed without requiring expensive over-provisioning of servers, local server distributed memory, or storage devices.

To optimize the benefits of memory virtualization, data center architects should deploy the fastest possible connectivity between servers, storage, and virtualized memory. Mellanox ConnectX™ InfiniBand-based adapters and InfiniScale™-based switches provide unparalleled support for memory virtualization environments with several key benefits:

- Performance that scales from 10 to 20 to 40 Gb/s with sub-1uSec latency (at 40Gbps)
- Native RDMA support that works transparently with RNA's solution, so no RDMA expertise is required
- Virtualized I/O capabilities, which allow I/O to be scaled and managed independently of server performance. This eliminates the need to scale up to multiple adapters for higher bandwidth and also reduces space, power, and management costs.

With available bandwidth up to four times that of 10-Gigabit Ethernet in addition to lower latency and native RDMA support, Mellanox InfiniBand connectivity solutions provide the highest performance, which maximizes the benefits gained from RNA memory virtualization.

1.3 Applications

Memory virtualization and high-performance connectivity have compelling advantages in applications involving large data sets as well as those that require the least possible latency.

1.4 Large Data Sets

Because it can reduce storage access from hundreds or thousands of reads to just one, virtualized memory delivers huge performance gains in applications using large data sets. For example, NWChem is a computational chemistry application typically deployed in an HPC environment. In a four-node cluster with 4 GB of memory per node running NWChem, a test run required 17 minutes to complete. However, by setting up a small 8GB MVP cache across the cluster, RNA networks improved the test run time by 3X.

1.5 Low Latency

Memory virtualization substantially improves the performance of applications such as trade execution algorithms by allowing them to publish and subscribe to message streams in the shared memory pool. For example, a multibillion dollar hedge fund uses MVP technology to improve trade execution times more than 800 percent.

1.6 Conclusion

Reliance on fragmented local server memory has been a key roadblock to optimizing performance in data center clusters, but memory virtualization eliminates size limits and slashes access times by providing shared memory for all CPUs in a cluster. By combining RNA Networks' Memory Virtualization Platform with Mellanox Technologies' unrivaled connectivity performance, data center architects can achieve new levels of performance with high efficiency and lower costs.

Mellanox TECHNOLOGIES

350 Oakmead Parkway
Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com