Data Sharing Clusters for Web Serving and Grid Computing

Abstract

Intel-processor-based application servers are becoming more and more common in data centers today. Instead of achieving scalable performance and good response times through larger multi-processor machines (scaling up) IT architects are scaling out using Red Hat Enterprise Linux and low-cost servers and storage.

Version 1 - June 2004

Table of Contents

Scaling out storage and processor resources with shared storage clusters

Data sharing clusters are server farms based on Red Hat Enterprise Linux that share storage devices on a storage area network (SAN), and share data on those storage devices through Red Hat Global File System (GFS), a cluster file system for Linux. In a data sharing cluster, data can be written or read by any server, to or from any file in GFS. As shown in Figure 1, this approach contrasts with the traditional server farms that use only local, direct-attached storage.

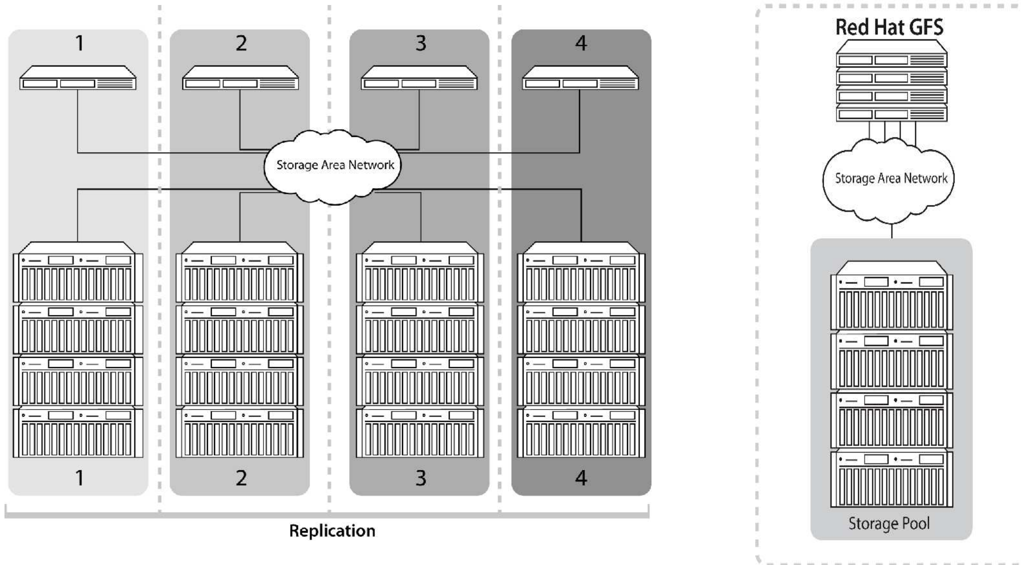

Figure 1: Non-data sharing server farm and a Red Hat data sharing cluster

Red Hat Cluster Suite can be used to increase application availability in a data sharing cluster by allowing applications to be quickly moved in the event of server failure or when server maintenance requires an application move from one server to another. Red Hat Enterprise Linux, GFS, and Cluster Suite are integral parts of the Red Hat Open Source Architecture (OSA), a complete, flexible framework for constructing IT solutions.

Key data sharing advantages with the Open Source Architecture

Data sharing clusters constructed with the Red Hat OSA have several advantages over traditional non-shared storage server farms.

First, good performance can be achieved (both bandwidth and latency) when transferring any file to and from any server. In contrast, in non-data sharing server farms, good performance can be achieved by each server to files on its own local file system, but not to files on other servers due to network and processor overheads. data sharing clusters are ideal for load balancing application work across all servers in the cluster since no application task has an affinity for a particular server due to local file access. Instead, all files are considered local to each server in the data sharing cluster.

Second, data sharing from a single shared pool of storage reduces copying overheads between servers in a cluster. Data copying is often performed in server farms to get the performance advantages of local access. However, data copying introduces multiple copies of a single file and the associated headaches of keeping the copies consistent (i.e., identical across the server farm) in the presence of application changes to the copied file and server failures during the copying process itself. In addition, it wastes precious network and processor cycles during the copying process itself, and for large, fast-changing files, simply does not work.

Third, server farms isolate data on each server. Unfortunately, a server failure renders its own data inaccessible to other servers. For this reason, high availability server farms use active-passive server pairs that share storage devices but not data. The passive server periodically polls the active server to determine if it is still operating. If not, the passive server reserves the shared storage (being careful to ensure the formerly active server can no longer access the same device) and restarts the applications that were being run by the formerly active server. Therefore, at any particular time only one of the servers in the active-passive pair may access the data on shared storage. In contrast, in data sharing clusters based on Red Hat GFS, the loss of a server does not imply lost access to data. These clusters are designed to keep data highly available even when servers fail. In fact, multiple server failures can be tolerated without affecting the surviving servers ability to access shared data. In addition, all servers can simultaneously mount Red Hat GFS, simplifying and speeding up the application failover process. Red Hat Cluster Suite provides a fast, integrated application failover capability that can be used with GFS.

Fourth, additional servers or storage devices can be incrementally added to the data sharing cluster while the system remains on-line. With non-data sharing server farms, an added server requires additional storage, or vice versa. While data sharing clusters allow IT administrators to add more of what they need without the excess overhead and capacity that direct-attach storage often yields.

Fifth, a data sharing cluster based on Red Hat GFS provides simplified management because the cluster shares one or a small number of file systems. In general, each additional file system in a server farm adds significant management overhead, and therefore requires more IT personnel. Costs are avoided in a data sharing cluster because each server does not require its own local file system for shared data.

Finally, unlike common network file protocols like NFS or CIFS, a data sharing cluster can provide true POSIX semantics across the whole cluster. For example, the NFS protocol allows client implementations to cache data aggressively without intervention from the NFS server, so that different NFS clients and the NFS server may have different versions of the same file system at any point in time. Certain distributed applications cannot tolerate this lack of data consistency and therefore cannot use NFS. In contrast, servers in a data sharing cluster have the same view of the shared data across the cluster at all times, maintaining the same POSIX semantics used by nearly all UNIX applications. To make the advantages of the Red Hat Open Source Architecture for data sharing more concrete, the following sections describe how Web serving and grid computing clusters can effectively exploit data sharing.

Using OSA to build a data sharing cluster for grid computing

Grid computing has become a popular technique for achieving scalable performance at a lower cost in technical and scientific computing. Enterprises now employ grid computing for a wide variety of tasks, from genomics research and drug design in the pharmaceutical industry, to risk analysis and trading strategy design in financial services, to the traditional bastions of Computer Aided Design (CAD). The Red Hat OSA--including Red Hat Enterprise Linux, GFS, and Cluster Suite--can deliver data sharing for clustered computing to improve performance, scalability, and efficiency while simplifying management in large computing grids.

Low-cost x86-processor-based server farms have become popular in technical, scientific, and grid computing for several reasons. They include: outstanding price/performance (by exploiting the economies-of-scale of the PC server market), compute-cycle scalability (just add more processors to get more compute cycles), and application affinity (many applications can be and have been made to run in parallel on a cluster). However, without true data sharing, the inherent processing power of these server farms cannot be harnessed because there is a serious imbalance between processor cycles and data transfer (often called IO, for Input-Output) speeds. Processing speed must be balanced with memory bandwidth and sufficient IO capacity so that servers can access and create data at a rate aggregate with this speed. Large imbalances lead to inefficiency as processors must wait for data to be read or written from disk.

For some applications it is possible to use special techniques to balance IO and processor performance. For example, making local copies of shared data that can be read quickly from a local disk during limited phases of computation. But these solutions are generally ad-hoc, complex, and application-specific. In effect, the programmer mapping an application to a non-data sharing server farm must constrict and manipulate the data transfer and disk utilization to adapt to the server farm's limitations, sacrificing simplicity, efficiency, and performance.

In contrast, data sharing clusters are ideal for grid computing because the architecture can be incrementally scaled and therefore is inherently balanced. As more processors are added, more shared storage can be added to balance processor and IO speed. In addition, instead of separate islands of partitioned data to which only one server has fast access, a data sharing cluster provides all servers fast access to all shared data. This allows any processor task to be executed on any server, without an artificial affinity between certain servers and certain data. This greatly increases performance and efficiency, and simplifies load balancing across the cluster.

In the past, specialized supercomputers generally had very fast access to their local storage, but servers for pre- and post-processing data had much slower access to the data stored on the supercomputer. Yet again, the inherent problem was the lack of data sharing between the supercomputer and the pre- and post-processing server. Data sharing clusters based on the OSA, using Red Hat GFS, can efficiently share data between a computing grid and desktop machines, allowing users fast access to computed data across the storage area network.

Incremental scalability to hundreds of servers and many terabytes of storage, seamless application load balancing, failover in the event of server failures, high data transfer rates, and avoidance of needless data copying make data sharing clusters using Red Hat technology the right approach for technical and scientific computing.

Using the OSA to build a data sharing cluster for Web serving with Apache

Web services are now an integral part of enterprise computing, providing customers and employees with a critical path to an enterprise's key information resources. The popular open source Apache Web server is widely used as a component in Web serving infrastructures.

As shown in Figure 2, data sharing clusters can provide scalable performance, simplified management, and highly available Web serving with Apache. Notice that Apache client http requests are balanced across all Web servers in the cluster. The servers access shared Apache data including log files, configuration files, and static and dynamic data files.

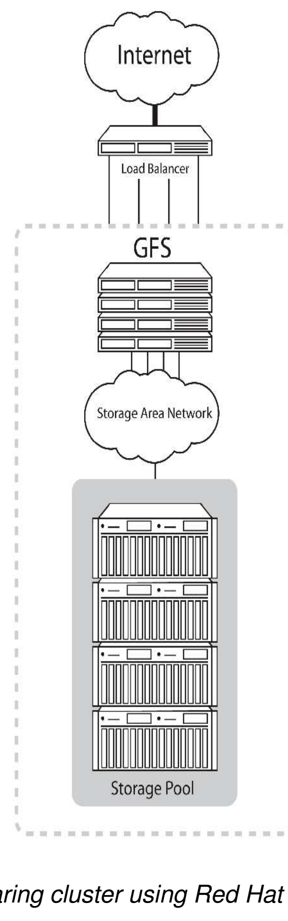

Figure 2: A data sharing cluster using Red Hat GFS for Web serving with Apache.

Load balancing is easy because all files are accessible to all servers, including rapidly changing dynamic data files. No file copying or complex scripting is necessary as data files change or new ones are created because these files are instantly visible to all Web servers in the cluster. This means the network, processor, and disk space overheads of this copying are avoided, as are the inherent difficulty of keeping all these changing files consistent in the context of server, storage, or network failures.

System administrators can also perform log file analysis on the fly across all log files (each specific to a particular server but located on GFS), even log files from servers that have crashed or been disabled by some kind of external attack. In contrast, access to log files from servers in non-data sharing clusters is lost when the server goes down, hampering efforts to analyze server or network problems. Server failures do not significantly impact performance as the failed server's load can be transferred to other servers by a load-balancing switch.

Dynamic analysis of shared log files, seamless load balancing and failover in the event of server failures, and avoidance of needless data copying and synchronization scripts make data sharing clusters based on Red Hat Enterprise Linux, GFS, and Cluster Suite the right approach for reducing management complexity in Apache Web serving. Combine this with incrementally scalable performance and storage capacity, and the overwhelming advantages of data sharing clusters for Web serving are clear.

www.redhat.com