



CLUSTERING FOR AVAILABILITY

HOW TO GET MORE FROM YOUR DATA CENTER

Every day, IT is faced with the challenge of delivering “application dial tone” that carries end-user requests for database, Web, or home-grown applications between the data center and the ultimate user of the applications and services delivered by IT. Just like the dial tone delivered from the phone company, users now expect that their application will be delivered like any other utility—with no downtime or lapse in service acceptable.



To deliver on this promise, IT has had to develop layers of redundancy into the infrastructure—from RAID disk arrays with built-in fault tolerance to multiple network connections and NICs that provide a redundant path between users and their data and programs, and even standby application servers. These layers provide backup computational services should hardware or software become inoperative due to failure or disaster.

LOCALIZED HIGH-AVAILABILITY SOLUTIONS PREDOMINATE

With so many components—including database servers, application servers, networking equipment, and storage of all types—IT has begun to deliver highly available infrastructure for the data center in a limited fashion, with most of these pockets of high availability limited to specific applications in specific locations. Typically, an enterprise will focus first on database availability, allocating a standby database server (which often is standing by just waiting for a failure in the primary server) and a shared storage solution such as a storage area network, or SAN. This type of setup allows a company to separate software and databases from a specific server, and allows multiple servers to access the same physical storage containing the mission-critical applications and data required to keep the enterprise running.

Most clusters to date have played a specific function. Standby servers, for instance, do not perform productive work in a normal everyday environment—they just wait for a failure of a primary server, at which point they take on the workload of the main server through a process known as clustering or failover. Most clusters are highly localized, usually to a single pair of servers, sharing a single disk array,

running a single critical enterprise application. And in most cases, the primary and redundant infrastructure resides in a single physical location. These solutions work well, but new approaches can combine the ability to act as a ‘traffic director’ and move applications among servers and storage devices to fine-tune performance regardless of failure.

“
Since demands on the application infrastructure can shift so rapidly, the problem of providing high availability is compounded by understanding which applications need to be offered in a clustered environment.
”

HIGH AVAILABILITY: A USER’S VIEW

Looking at high availability just from a single back-office database may not be enough to provide continuous application dial tone to end users. A major issue affecting availability is the changing workload that IT is required to manage.

As enterprises move to continuous operation of their IT infrastructure due to the

globalization that Web-based applications and the Internet have enabled, it has become difficult to project exact workloads for applications, Internet connections, servers, and storage resources. For example, Web services-enabled applications now allow employees of many enterprises—large and small alike—to access their HR benefits such as 401(k) information any time of the day or night from their Web browsers. Also, customers on the other side of the world can place orders, check status, and schedule delivery through Web-enabled ERP and CRM systems regardless of the availability of a live customer service representative.

Since demands on the application infrastructure can shift so rapidly, the problem of providing high availability is compounded by understanding which applications need to be offered in a clustered environment. A company must also decide whether or not the cluster can be extended beyond a single room (the data center) to a campus, city, or beyond to enable business continuity and support high availability of applications even if a disaster should bring down an entire data center.

Ideally, the incremental additional infrastructure needed to support high availability of a single application could be shared between multiple applications. So rather than duplicating every application server that requires high availability, a smaller number of standby servers could serve the needs of multiple applications.

WORKLOAD MANAGEMENT CLUSTERS

In such an environment, where many servers are clustered together with one or more standby servers, all attached to a common pool of storage, not only could these standby servers be used in case of

failure of a primary server, but now the additional capacity could also be used to handle peak demands on a server resource, such as a Web server. This would allow the standby server to be used to perform productive work at all times, balancing the workload more evenly and—from the end-users' perspective—enhancing the performance of an application by allowing more compute power to be applied to that application.

There are many options to consider when developing a strategy to support workload management as part of a highly available infrastructure. That's because most vendors of server hardware offer some level of high-availability solution.



VENDOR HARDWARE-BASED SOLUTIONS

When choosing a clustering or high-availability solution, it's important to look at the entire application infrastructure—servers, storage, operating systems, application, and network infrastructure—and ensure that the solution has the flexibility and heterogeneous support for multi-vendor storage environments. Many clustering solutions on the market today require that you use a single vendor's storage or server brand, which is great for that vendor's sales, but may not allow you to leverage your existing resources, while tying you to that vendor's product roadmap and pricing.

Ideally, the solution implemented should be so platform-independent that even if IT is managing multiple clusters of Windows, UNIX, and Linux application servers, a common interface and common tools could be used. Also, a single console should be able to provide a window into the status and availability statistics for each cluster in the enterprise, whether it's local to a single data center or across the Mississippi in another facility.

VERITAS CLUSTERING AND WORKLOAD MANAGEMENT SOLUTIONS

People are IT's most important asset, but every expert can't always be available. If a single tool can provide services for a variety of open-system environments—

including Windows, UNIX, and Linux—deployment of that tool could save precious hours and dollars in overall IT expenditures.

VERITAS Software, the number-one clustering vendor according to IDC, has developed a suite of tools designed to enhance high availability as well as provide a new level of adaptive workload management to an organization's entire infrastructure. The core of this suite is VERITAS Cluster Server (VCS). Here is a look at some of the key benefits to IT that can be realized by deploying VERITAS Software clustering solutions, and the products that enable them.

VERITAS CLUSTER SERVER (VCS) is a robust application availability architecture that keeps applications online in many sophisticated ways. VCS's improved failover technology supports up to 32 nodes in a cluster, all nodes communicating with each other to ensure that clients maintain continual access to applications and data. Since many systems in the cluster are monitoring each other simultaneously, there is no single point of failure, even in the control of the cluster.

Further, VCS assures that planned server outages for upgrades or scheduled maintenance can occur with minimal performance impact and without additional migration of storage or software before a server in the cluster is taken offline for maintenance. In a VERITAS Software cluster environment, new levels of high availability are attained. Failover options inherent in VCS allow configuration of 1:1, any:1, or any:any failover environments, so any server in the cluster can handle failover of any of the active servers in that cluster. No longer is IT forced to double the number of application servers in the data center just to guarantee availability.

In choosing application servers, heterogeneity is a fact of life in today's open-systems environment. Combinations of Windows, UNIX variants, and Linux are commonplace. Multiple clusters are required to support multiple operating system platforms and applications, since no single cluster can support failover between different platforms. However, with VCS, even if multiple clusters are deployed, they can all be managed and controlled from a single, Web-based console, regardless of the OS or platform that the cluster application servers support. This allows IT to leverage its most important asset—its people—eliminating the need to train in multiple clustering tools to support multiple cluster platforms.

VERITAS CLUSTER SERVER TRAFFIC DIRECTOR

provides workload balancing and high availability to user traffic. By configuring the allocation of tasks via round-robin or least-connection attributes, VCS Traffic Director ensures that all servers in the cluster are used in the most efficient manner. It also ensures that a failure in any server in the cluster will have minimal impact on the end-user experience. An extension of VCS, called VCS Traffic Director, offers traffic management capabilities which allow incoming requests for application services to be distributed among multiple nodes running a Web server, application server, or content management software. With VCS Traffic Director, each server is able to handle requests from user traffic, and if a server or component fails, user requests can be redistributed among remaining servers.

VERITAS Software's tradition of supporting all major third-party storage products is extended with VCS in both SAN and client-server storage products. With VCS you can utilize your existing storage solutions, preserving and leveraging your investment in installed storage devices.

Tight integration with VERITAS Volume Manager, the leading virtualization and volume-management software solution, enables clusters to see all the storage on the SAN by grouping physical devices into logical volumes. It also allows VERITAS' software to remap failed server storage to the new server and provide continuous data access.

Since today's IT environments often span multiple data centers and locations, VERITAS has provided tools to extend VCS beyond the confines of the data center, over SAN or standard IP connectivity.



VERITAS VOLUME REPLICATOR (VVR) is an enhancement of VERITAS Volume Manager that provides the ability to extend Volume Manager's mirroring capabilities over any IP connection. This allows IT to replicate mission-critical data to another location without any distance limitations, and with the same familiar interface that Volume Manager provides for local storage. With VVR, the ability to replicate your data and applications to another location seamlessly between any storage array, through either dedicated or shared IP connectivity enables a key advantage: the ability to extend clustering technologies over a wide-area link, enabling a level of availability and disaster recovery previously unattainable.

TOOLS FOR ONE-CLICK DISASTER RECOVERY

When an enterprise's clustering requirements

demand a multi-cluster, multi-site cluster configuration, the ability to manage all the applications, replication, and traffic from a single point of control becomes critical. To enable IT to manage such a cluster or group of clusters and replication jobs, VERITAS Software offers Global Cluster Manager (GCM). Through GCM, the entire organization's clusters and replication tasks can be managed and supported from one central console, so enterprise-wide clustering activities and status are easily displayed and managed.

This tight integration between the elements of VERITAS Software's clustering and high-availability solutions pays off in time savings—time to deploy, time to successful failover, and time to resumption of business activities after a failure or disaster.

VERITAS Software has consultants who can work with you to design a cluster configuration that can be deployed in a fraction of the time that alternative solutions would demand. When failure or disaster occurs, applications protected by VERITAS Cluster Server and its associated tools allow IT to return failed applications to service in minutes, rather than hours or days that might otherwise be required if VCS was not in place. ▾

RESOURCES:

Find out why more users trust VERITAS Software for maintaining high-availability environments than any other vendor.

For more information on high availability, clustering, and workload management, be sure to register at www.veritas.com/ha for VERITAS Software's upcoming eSeminar or call us at 1-800-729-7894 (reference code 8091516).



Copyright 2002 VERITAS Software Corporation. All rights reserved. VERITAS, the VERITAS Logo and all other VERITAS product names and slogans are trademarks or registered trademarks of VERITAS Software Corporation. VERITAS and the VERITAS Logo Reg. U.S. Pat. & Tm. Off. Other product names and/or slogans mentioned herein may be trademarks or registered trademarks of their respective companies.