Technical White Paper
DATA CENTER

www.novell.com

SUSE® Linux Enterprise 10: High Availability Storage Foundation

Enterprise-class Functionality at Open Source Prices

Novell®

Keeping Availability High and Costs Low

Even as data grows, you can lower storage-management costs and still benefit from an easy-to-manage, highly available foundation that scales as needed.

The documentation requirements accompanying recent compliance regulations force companies to continually increase their data storage. This growth not only drives demand for capacity, it also creates a need for storage management that can handle the growing data. Since business continuity relies on uninterrupted access to information and services, the storage management system must ensure both data integrity and availability. High Availability Storage Foundation, an important featured technology in the SUSE® Linux Enterprise 10 platform, satisfies these needs.

Unlike the high costs of proprietary solutions, High Availability Storage Foundation keeps costs low by integrating only open source, enterprise-class components:

- **Heartbeat v2**, *a high-availability resource manager that supports multinode failover*
- **Oracle* Cluster File System 2 (OCFS2)**, *a parallel cluster file system that offers scalability*
- **Enterprise Volume Manager 2 (EVMS2)**, *a cluster-aware volume manager that simplifies operations in a high-availability, scalable environment*

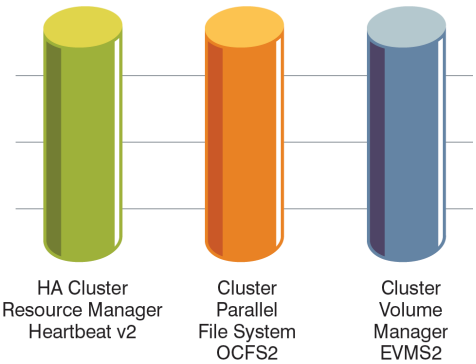

Figure 1. *The Three Pillars of the High Availability Storage Foundation*

Even as data grows, you can lower storage-management costs and still benefit from an easy-to-manage, highly available foundation that scales as needed.

While you could install each of these open source storage technologies separately (and manually), SUSE Linux Enterprise 10 integrates the three. Without this integration, you would have to configure each component separately, and manually prevent conflicting administration operations from affecting shared storage. When delivered as an integrated solution, the High Availability Storage Foundation technology automatically shares cluster configuration, and coordinates cluster-wide activities to ensure deterministic and predictable administration of storage resources for shared disk based clusters. When installing the SUSE Linux Enterprise High Availability pattern and using the YaST based cluster configuration tool, initial set up time is also reduced, with installation and configuration of a multi-node cluster typically taking only a few minutes. Ongoing administration of your shared-disk cluster, for example, adding storage or reconfiguring disks in your production environment, is simplified by the cluster-aware volume manager. Overall, the integration improves the manageability, and therefore reliability of your storage infrastructure.

Due to another open source technology, Internet Small Computer System Interface (iSCSI), High Availability Storage Foundation adds flexibility to your storage area network. This new technology, which combines SCSI, Ethernet and TCP/IP, lowers the cost of your storage area network and makes it easier to manage and deploy. Integration with Multipath I/O ensures uninterrupted access to SAN storage, whether using iSCSI or Fibre Channel-based connectivity.

Heartbeat v2: Keeping Your Business Alive with High-availability Storage

Numerous high-availability features, such as multinode failover, make Heartbeat v2 an enterprise-class resource manager. For example, by monitoring all server nodes within a cluster, Heartbeat v2 eliminates single points of failure and builds in redundancy.

Two-node failover can handle only one machine failure at a time. If one of the nodes goes down, the system no longer has redundancy and remains vulnerable until you repair the failed node. Recent studies show that the average production deployment of high-availability clusters requires six nodes balance resources and to avoid downtime. Heartbeat v2 has completed testing for up to 16 nodes, but theoretically it can manage any number.

Through an Open Cluster Framework (OCF) Resource Agent API, Heartbeat v2 can monitor services and applications as well as nodes. This API allows you to build application and service resource monitors so the cluster manager can detect whether a specific application or service is functioning properly. Even if it appears that the nodes themselves are working, these monitors recognize when the service or application is not responding to health checks. When that happens, Heartbeat v2 can restart the nonfunctioning service immediately or move it to another node in the cluster to get it running again. With its dynamic resource dependency modeling, resource priorities and time-affected rules, Heartbeat v2 even protects multitiered applications, which rely on multiple components. This flexible configuration model monitors whether or not components are available and functioning properly. Heartbeat v2 also works with your existing LSB init scripts, should an OCF compliant resource agent not be available for a given service.

Availability requires more than simply ensuring that the hardware is working correctly. These days it also encompasses the virtual world. Consequently, Heartbeat v2 supports Xen virtualization. If the physical system hosting one or more Xen virtual machines fails, Heartbeat v2 can move the virtual machines to another host system. And since another aspect of availability is ensuring data integrity, Heartbeat v2 includes many methods of fencing or IO isolation. These methods prevent failed nodes from corrupting the data. Ethernet channel bonding and multipath I/O enables highly available LAN and SAN connectivity.

With Heartbeat v2, high availability is only a starting point; the Open Cluster Framework Resource Monitors also improve failure-detection performance. Depending on how you configure the Open Cluster Framework Resource Monitors, the system can detect failures in less than one second and failover immediately. However, even with the potential for configuring sub-second failure detection, the restart time of the application is the gating factor on how quickly you are up and running again. Nonetheless, applications with fast restart capabilities can benefit significantly from this improvement.

Even with its many new features and functions, Heartbeat v2's new graphical user interface simplifies monitoring cluster state and resources as well as managing cluster configuration. Powerful command-line tools and a Common Information Model (CIM) provider allow rich integration with scripts and enterprise-management tools. Furthermore, YaST, the administrative tool for SUSE Linux Enterprise, makes it easy to configure your system for multinode failover.

With its dynamic resource dependency modeling, resource priorities and time-affected rules, Heartbeat v2 even protects multi-tiered applications, which rely on multiple components. This flexible configuration model monitors whether or not components are available and functioning properly.

When you scale out using OCFS2 you add more machines to the cluster. As a result, you increase the size of your clusters, ultimately improving overall reliability.

You can install Oracle RAC directly on SUSE Linux Enterprise Servers without updating the Novell supplied version of OCFS2. Novell makes the management of Oracle RAC installations with OCFS2 even easier.

Maintenance without Downtime

In addition to failing hardware or applications, even scheduled maintenance can become an availability issue. Prior to installing cluster technology, a hospital emergency room required three months of planning for scheduled systems maintenance. Even though they kept the system offline for only an hour, the hospital needed to hire additional staff to maintain operations using a paper-based system. And, by the way, this maintenance was done during off-hours: on Sundays at 2 a.m.

With Heartbeat v2, you can avoid this kind of time-consuming and costly disruption. You simply failover the services to another node and temporarily remove the machine needing maintenance from the cluster. Once the work is done, you bring the machine back into the cluster.

If a physical machine fails unexpectedly, Heartbeat v2 keeps the services up and running with fast detection of the failure and failover. It also notifies administrators of the failure so that they can fix the machine and redeploy it back into the cluster, whenever convenient.

Adding Scalability to High Availability with Oracle Cluster File System 2

As IT organizations continue to evolve from client-server to service-oriented architecture (SOA), parallel cluster file systems like Oracle Cluster File System 2 (OCFS2) become increasingly important. OCFS2 is appropriate for SOA because the applications are stateless—they do not maintain any information about what has previously occurred. Unless specifically constructed so that all of their state information is kept in the file system, client-server applications can use only one node at a time, even though their data is available to all nodes simultaneously. Properly designed SOA applications or services can run and access the same data simultaneously on two or more nodes within a parallel cluster file system. This allows the stateless applications to scale out.

The ability to scale out as opposed to scale up may drive down the total cost of ownership, and it will definitely improve overall reliability. Traditionally, the throughput of client/server applications was increased by adding more processors to a Symmetric Multiprocessing (SMP) machine. Not only can the price of these systems run high, the SMP machine becomes a single point of failure. When you scale out using OCFS2 you add more machines to the cluster. As a result, you increase the size of your clusters, ultimately improving overall reliability.

Oracle retooled its data base technology so it could scale out, creating Oracle Real Application Cluster (RAC), and OCFS2. You can install Oracle RAC directly on SUSE Linux Enterprise Servers without updating the Novell supplied version of OCFS2. That isn't the case with other distributions, where you would have to patch your system before installing Oracle RAC. Novell makes the management of Oracle RAC installations with OCFS2 even easier.

Other workloads that run on OCFS2 include SAP on Oracle RAC, LAMP stacks and Xen virtual machine images. OCFS2 enhances the manageability of Xen virtualization technology. Since OCFS2 was designed to host and perform on larger files in a clustered environment, it works very well hosting virtual machine disk images in a high-availability environment. All nodes can access the Xen images and configuration files. This lets you easily move or migrate virtual machines between clustered servers, Also, you have to deploy a new virtual machine image only

once rather than copying it to each server in the cluster.

In addition to its scalability, OCFS2 can run on the full breadth of architectures supported by SUSE Linux Enterprise 10—from x86 to IBM* zSeries*. OCFS2 also demonstrates the Novell® commitment to community-based projects. It is the only parallel cluster file system to be accepted into the Linux* mainline kernel.

Finally, with Context Dependent Symbolic Links, added in version 2, OCFS maintains node-specific files under a common cluster-wide name, allowing for data that is only relevant to one node in the cluster, to be accessed using the same name on all nodes. The Kernel Journaling Block Device, a common kernel service, provides journaling for file system recovery after node failure.

Building Flexibility on Top of Availability and Scalability with Enterprise Volume Management System

The third component of High Availability Storage Foundation, the (disk) volume manager, allows you to virtualize storage into logical groupings. Between cluster awareness, a layered plug-in model and a single mechanism for all tasks, Enterprise Volume Management System 2 (EVMS2) takes logical volume management for Linux to a completely new level of flexibility and ease of use. This new approach provides a flexible and extensible framework so that you can easily expand and customize various levels of volume management.

EVMS2 configures storage through containers, which you create by aggregating multiple physical disks. Once you have created a container, EVMS2 gives you the option of partitioning it or attaching it to servers, even in a cluster configuration *(see Figure 2: EVMS2 Architectural Overview)*. You can form private cluster containers (which are owned or used by only one node at a time) or shared ones (which all nodes can use simultaneously). This versatility allows you to design storage based on your application or service needs, as well as to maintain the security and integrity of your application data.

Between cluster awareness, a layered plug-in model and a single mechanism for all tasks, Enterprise Volume Management System 2 (EVMS2) takes logical volume management for Linux to a completely new level of flexibility and ease of use.

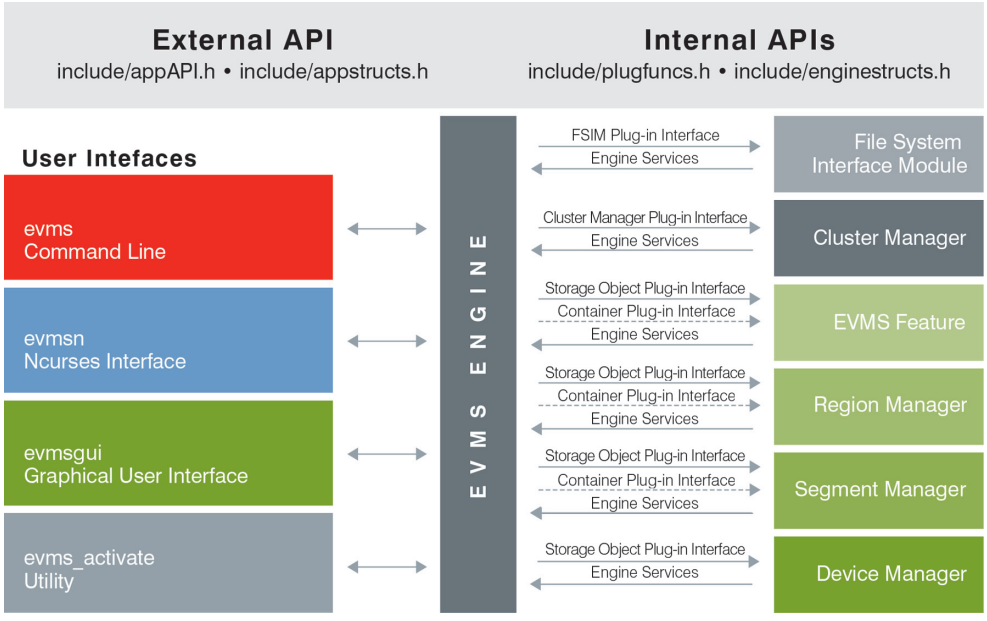

Figure 2. *EVMS2 Architectural Overview (Source: http://evms/sourceforge.net/architecture/)*

The integration of Heartbeat v2, OCFS2 and EVMS2 within SUSE Linux Enterprise 10 gets you there—and with a low total cost of ownership.

EVMS2 achieves much of its versatility through the plug-in model. With a file system interface module (FSIM), EVMS2 integrates various file systems. It also becomes cluster-aware through a different plug-in called the EVMS2 cluster engine (ECE).

Cluster awareness simplifies volume management. When you modify a cluster configuration in any way, such as increasing the size of a file system on a logical volume, or mounting a new volume, EVMS2 informs all nodes about the change. With a volume-manager that's not aware of the cluster, the administrator has to ensure all nodes are refreshed of the configuration changes made to shared storage. This leaves the storage vulnerable because different nodes may retain conflicting views of shared storage configuration. For file systems that can mount only one node at a time, such as EXT3 or Reiser, cluster-aware EVMS2 prevents two nodes from mounting the same volume simultaneously.

In addition to being flexible, EVMS2 adds to the overall robustness of High Availability Storage Foundation with a rich assortment of maintenance and storage-management tasks. It can perform bad block relocation when it detects a write failure. It supports snapshots, which are useful for backups on a live system. It also consolidates storage management because you no longer need separate utilities. EVMS2 provides a single mechanism for each task.

EVMS2 can leverage multipath I/O, which accesses storage devices through multiple channels to create redundancy for your storage area network. Although EVMS2 is not integrated with multipath I/O, it can still use its tools to achieve greater load balancing and fault tolerance for improved uptime and protection.

Lowering Your Storage Costs with iSCSI

Using a commodity server and Ethernet connections, iSCSI lets you create a fully functional disk storage server. iSCSI storage area networks enable remote storage accessibility over the global IP Network, which eliminates any distance limitations. iSCSI also increases interoperability by reducing disparate networks and cabling, allowing you to standardize on standard Ethernet equipment.

The use of Ethernet switches compared to the more complex Fibre Channel infrastructure eliminates the need for special skills and training. Commodity networking hardware also helps lower cost of ownership. Yet, with High Availability Storage Foundation, this low-cost alternative becomes easy to manage through an EVMS2 plug-in that allows you to create and manage iSCSI targets.

High Availability Storage Foundation: Designed and Delivered for the Enterprise

The five characteristics required to claim an enterprise-class solution are:

- *High Availability*
- *Robustness*
- *Manageability*
- *Flexibility*
- *Scalability*

A high price tag often accompanies this difficult-to-attain list. Therefore, Novell looked for a complementary set of open source modules that, when integrated, would supply the features and functions necessary to operate Linux in your Open Enterprise. The integration of Heartbeat v2, OCFS2 and EVMS2 *(see Figure 3)* within SUSE Linux Enterprise 10 gets you there—and with a low total cost of ownership.

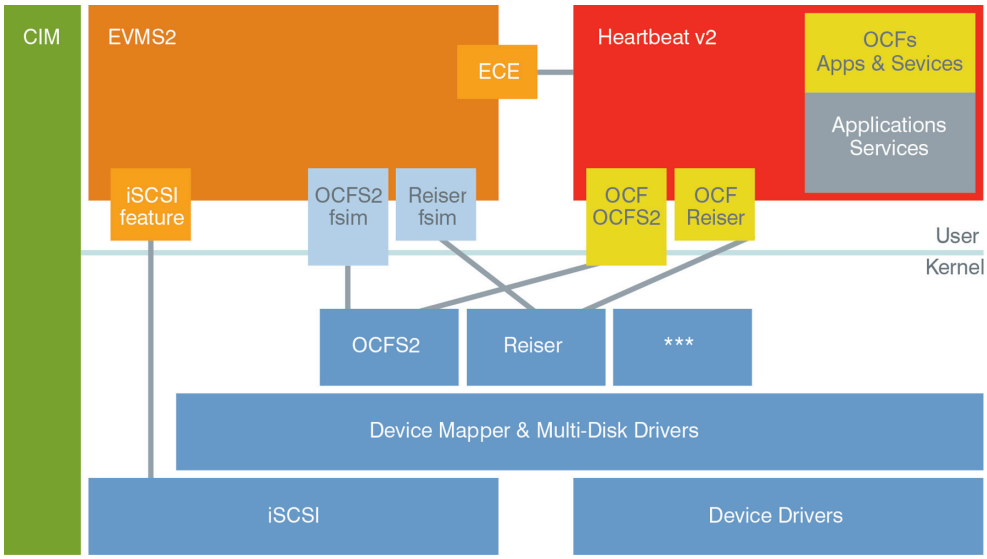

Figure 3. *High Availability Storage Foundation in SUSE Linux Enterprise—Integration of OCFS2, Heartbeat v2 and EVMS2*

High Availability Storage Foundation offers file systems beyond OCFS2, including EXT3, ReiserFS and XFS. Together, these file systems support a broad range of application types—from databases to client-server to Web services—and storage configurations, scaling from small file systems to millions of files requiring terabytes of storage. Like OCFS2, these other file systems are integrated with EVMS2 and Heartbeat v2. Even when the amount of data is enormous and you are managing many volumes, this storage infrastructure maintains its performance.

The multinode failover support in Heartbeat v2, the improved node and journaling recovery in OCFS2, and the snapshots in EVMS2 represent a small sampling of the high-availability features in the storage foundation. In addition, other features such as the cluster awareness and ready-to-run support of Oracle RAC enrich the environment, simplifying administrative tasks or eliminating them completely. And iSCSI gives you the flexibility you need for low-cost storage area networks.

Overall, High Availability Storage Foundation protects your data in a way that lowers costs, simplifies storage management and, most importantly, keeps your enterprise running.

High Availability Storage Foundation protects your data in a way that lowers costs, simplifies storage management and, most importantly, keeps your enterprise running.

www.novell.com

Contact your local Novell Solutions Provider, or call Novell at:

1 888 321 4272 U.S./Canada
1 801 861 4272 Worldwide
1 801 861 8473 Facsimile

Novell, Inc.
404 Wyman Street
Waltham, MA 02451 USA

Novell®