Requirements Specification

Solaris iSCSI Initiator and Target Support

Dan Orbach, Augustus Franklin
September 4th, 2008
Feature Id 2006/1568

Revision History

Version	*Comments*	*Date*	*Author*
1.0	Initial draft	07/16/08	Dan Orbach, Augustus Franklin
1.1	Internal Review	07/18/08	Dan Orbach, Augustus Franklin
1.2	Updated after review from colorado-dev team	07/23/08	Dan Orbach, Augustus Franklin
1.3	Added exceptional behaviour section, ancillary sections and updated scope.	08/11/08	Dan Orbach, Augustus Franklin
1.4	Updated after CLARC review	09/04/08	Augustus Franklin

Always update the revision history after making changes to this document.

1. Project Description

1.1 What it is

iSCSI is a protocol that allows clients (called *initiators*) to send SCSI commands to SCSI storage devices (*targets*) on remote servers. It is a popular Storage Area Network (SAN) protocol, allowing organizations to consolidate storage into data center storage arrays while providing hosts with the illusion of locally-attached disks. Unlike Fibre Channel, which requires special-purpose cabling, iSCSI can be run over long distances using existing network infrastructure. This project would support devices using Solaris iSCSI initiator and target, as valid storage devices, in Sun Cluster.

1.2 Scope

There have been multiple requests from the field to support clustering of x4500 series of Sun Servers, also called as thumpers. Thumpers are four-way x64 based servers that support upto 48

Terrabytes of storage in 4U of rack space. There is interest from customers to be able to use the large amount of storage present in these servers, as storage for running high-available applications. This project will make the necessary changes in Sun Cluster to make this possible.

There is also a general goal to support Sun Cluster with a minimum set of hardware components. Using Solaris iSCSI initiator and target is one way to provide shared storage in Sun Cluster, even though there is no storage that is physically connected to multiple hosts. Project Colorado[2] project is driving this general goal and therefore the changes made in this project will also enable project Colorado to achieve its objectives.

This project will add support for the iSCSI initiator and iSCSI target in Solaris. It is possible for valid iSCSI targets to be implemented on third-party storage arrays. Support for those third-party iSCSI targets is out-of-scope for this project. We expect those storage arrays to be qualified based on customer demand, as follow-on qualification projects.

This project will make Solaris based iSCSI devices available for Sun Cluster use as storage devices. This project will not make iSCSI targets highly available to clients outside the cluster hosts. There are changes needed in the Sun Cluster quorum sub-system to support these iSCSI devices as quorum devices. The support of Solaris iSCSI devices as quorum devices is out-of-scope for this project.

Broadly speaking, this project supports iSCSI based devices the same way as disk devices are used today and will support these iSCSI based disks in all the configuraions that support disks today like ZFS, Solaris Volume manager, global device configuration, shared QFS etc... Veritas volume manager(VxVM) should be supported too, however there could be bugs and other issues beyond our control that could prevent VxVM from working well with Solaris iSCSI devices. This project considers support for Solaris iSCSI devices with VxVM out-of-scope for this release.

The Sun Cluster Geographic edition product should be able to use the iSCSI devices that are used by the Sun Cluster product. We do not expect any additional effort to support these devices on Sun Cluster Geographic edition.

There will not be any network topology restrictions in the supporting Solaris iSCSI targets. These targets could be present in the same subnet as the cluster host, or a different sub-net.

There could be tradeoffs when using iSCSI based devices with respect to, robust availability and performance. However, customers looking for more flexibility in Sun Cluster will find this perfectly acceptable. This project will document any such tradeoffs in its user documentation.

All the existing support present in Sun Cluster for certain iSCSI based devices will be continue to work and this project will not affect existing supported configuraions.

1.2.1 Scope for Project Colorado

For the phase 1 of Project Colorado, this project will support only a sub-set of the total set of requirements. These requirements that are in scope for Project Colorado are prefixed with [CRn].

1.3 Terminology and Definitions

Initiator - An initiator is an iSCSI client. An initiator typically serves the same purpose to a computer as a SCSI bus adapter would, except that instead of physically cabling SCSI devices (like hard drives and tape changers), an iSCSI initiator sends SCSI commands over an IP network. This project only addresses the initiator that is implemented in Solaris. All references to initiator refers to the Solaris based iSCSI initiator in this docment.

Target - A target is a storage resource located on an iSCSI server (more generally, a target is one of potentially many *instances* of iSCSI running on that server). An iSCSI target usually represents hard disk storage. As with initiators, software to provide an iSCSI target is available for most mainstream operating systems. This project only addresses the target that is implemented in Solaris. All references to target refers to the Solaris based iSCSI target in this docment.

Logical Unit Number (LUN) - LUN is a *logical unit number*, and represents an individual SCSI device. In an iSCSI environment, LUNs are essentially numbered disk drives. An initiator negotiates with a target to establish connectivity to a LUN; the result is an iSCSI session that emulates a SCSI hard disk. Initiators treat iSCSI LUNs the same way as they would a raw SCSI or IDE hard drive; for instance, rather than mounting remote directories as would be done in NFS or CIFS environments, iSCSI systems format and directly manage filesystems on iSCSI LUNs.

In enterprise deployments, LUNs usually represent slices of large disk arrays, often allocated one per client. iSCSI imposes no rules or restrictions on multiple computers sharing individual LUNs; shared access to a single underlying filesystem is instead left as a task for the operating system.

Internet Storage Name Service (iSNS) - iSNS, a proposed standard at IETF, is a protocol allows automated discovery, management and configuration of iSCSI and Fibre Channel devices on a TCP/IP network.

2. Functional Requirements

2.1 Normal Case Behavior

R1: Solaris iSCSI initator and target support depends on the stability of iSCSI software in Solaris. The iSCSI software was introduced to Solaris in version 10. This project will support iSCSI based devices in Sun Cluster in Solaris 10 Update 7 and in all its subsequent versions.

The current plan of record is that fix for CR 6725957 will be present in Solaris 10 Update 7. If this plan changes, then this project will be supported on the update release of Solaris 10 in which CR 6725957 is fixed.

R2: [CR1] OpenSolaris 2009.04 and all its subsequent versions will also be supported.

R3: [CR2] All hosts that can run the supported version of Solaris as mentioned in R1 and R2 will be supported as cluster nodes to use this feature.

There will be certain hosts on which Sun Cluster is not supported due to other factors, then those servers will remain unsupported.

R4: Sun Cluster will support iSCSI targets on storage that is physically attached to a different non-cluster host from a cluster host.

This is a scenario where a non-cluster host is exporting it's local storage as "shared" (via Solaris iSCSI target) to a cluster.

R5: [CR3] Sun Cluster will support iSCSI targets on storage that is local on-board and locally attached to the cluster host.

iSCSI targets could be present in a remote host or on a local host. This project will support both of these configurations. For Project Colorado, R5 is a necessary requirement, but R4 is a nice to have requirement.

R6: Sun Cluster will not support iSCSI targets on storage that is already shared between the cluster nodes.

Storage that is already shared between the nodes of the cluster can be used in cluster configuraions to provide high availability. The extra complexity of supporting this configuration using iSCSI does not provide corresponding benefit.

iSCSI based devices will supported in Sun Cluster in the following configurations.

R7: [CR4] Highly available ZFS filesystem will be supported using iSCSI devices

iSCSI targets will be imported to the host as a LUN using the iSCSI initiator. Zpools can be created using these LUNs. These Zpools will then be made highly available.

R8: Raw iSCSI devices will be supported as a global device in Sun Cluster.

iSCSI devices will have entry in the global device namespace like other devices and it will be managed by the device services in Sun Cluster. This project supports iSCSI based devices the same way as disk devices are used today and will support these iSCSI based disks in all the configuraions that support disks today like ZFS, Solaris Volume manager, global device configuration, shared QFS and UFS.

R9: Solaris Volume Manager metasets, both single master and shared metasets (OBAN), will be supported using iSCSI devices.

Currently, there are open bugs in Solaris that prevents this feature from being supported. The

requirement on Sun Cluster will be dependent on these Solaris bugs being fixed. See CR 6691027

R10: [CR5] Sun Cluster DID driver loading must be hardened to support iSCSI devices.

iSCSI initator attaches devices to the host only when there is access to the devices. This causes problems when the Sun Cluster DID driver is loaded before the iSCSI devices are attached to the host. This problem will be addressed by this project.

R11: [CR6] Sun Cluster will only support iSCSI initiators that are connected to the iSCSI target using the static configuration and using the SendTargets discovery.

iSCSI initiator and targets can be connected using three approaches in Solaris. They are static configuration, discovery of iSCSI targets using SendTargets, and using a external iSNS server. Many deployment goals for iSCSI can be met via the static configuration and SendTargets discovery. Configuration is also a lot easier to use and administer in the static configuration. Therefore, this project will recommend static configuration.

R12: [CR7] Sun Cluster will support only iSCSI targets that are configured in emulation mode.

When iSCSI targets are configured in emulation mode, the iSCSI target software responds to SCSI commands. When iSCSI targets are used in pass through mode, the underlying device must support SCSI commands to be able to respond to them. The fencing feature of Sun Cluster depends on SCSI reservation commands to work. The Solaris iSCSI target team recommends that using iSCSI targets in emulation mode provides better performance than using them in pass-thru mode. Therefore, this project will support only iSCSI targets are configured in emulation mode.

R13: [CR8] Sun Cluster will support only iSCSI targets that are backed by a slice of a disk.

The iSCSI targets could be backed up by a slice of a disk, or a file or zvols. This project will only support iSCSI targets that are backed up by a slice of a disk in Sun Cluster.

R14: [CR9] Multi-pathing of iSCSI devices will be supported using MPxIO and IPMP

Using iSCSI multi-pathing is described in the blue print, "Using iSCSI Multipathing in the Solaris 10 Operation System"[1]. Sun Cluster will be supported in the configurations that are recommended for Solaris.

2.2 Exceptional Behavior

2.2.1 Errors Arising from User Interactions

There will certain iSCSI configuration that are supported in Sun Cluster and some configuraions

will not be supported in Sun Cluster. The supported configurations will be documented in the user documentation. Errors arising from a configuration that is not supported, should not prevent Sun Cluster from functioning normally, except with respect to the Solaris iSCSI device.

2.2.2 Errors Arising from System Failures

The error handling in existing devices code will work for iSCSI based devices as well. The connectivity to iSCSI based devices can be less reliable than the connectivity to standard SCSI disks. We expect errors arising from any loss of connectivity to be handled by the existing code that handles loss of connectivity for SAN based devices today.

2.3 scchecks/eRAS checks

The project does not introduce any new checks.

2.4 Accessibility (Section 508) Requirements

This project does not modify any existing CLIs or GUIs, therefore this section does not apply to this project.

2.5 Internationalization (I18N) and Localization (L10N) Requirements

This project does not modify any existing CLIs or GUIs, therefore this section does not apply to this project. There will be new log messages introduced by this project, but Sun Cluster does not translate any of its log messages.

2.6 Cluster Events

This project does not introduce any new cluster events.

3. Efficiency Requirements

3.1 Time and Space Requirements (Performance)

Configurations including Solaris iSCSI target devices will very liklely be less perfornant than raw SCSI devices. There will be no performance requirements for Solaris iSCSI devices from this project. The other advantages of using iSCSI devices is a trade-off to the difference in performance.

4. References

(1) Using iSCSI Multipathing in the Solaris 10 Operating System - www.sun.com/blueprints/1205/819-3730.pdf