

Veritas Storage Foundation™ for Oracle® RAC Best Practices Guide



Veritas Storage Foundation for Oracle RAC Best Practices Guide

The software described in this book is furnished under a license agreement and may be used only in accordance with the terms of the agreement.

Document version 1.0

Legal Notice

Copyright © 2009 Symantec Corporation. All rights reserved.

Symantec, the Symantec Logo, Veritas, Veritas Storage Foundation, and FlashSnap are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

The product described in this document is distributed under licenses restricting its use, copying, distribution, and decompilation/reverse engineering. No part of this document may be reproduced in any form by any means without prior written authorization of Symantec Corporation and its licensors, if any.

THE DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID. SYMANTEC CORPORATION SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS DOCUMENTATION. THE INFORMATION CONTAINED IN THIS DOCUMENTATION IS SUBJECT TO CHANGE WITHOUT NOTICE.

The Licensed Software and Documentation are deemed to be commercial computer software as defined in FAR 12.212 and subject to restricted rights as defined in FAR Section 52.227-19 "Commercial Computer Software - Restricted Rights" and DFARS 227.7202, "Rights in Commercial Computer Software or Commercial Computer Software Documentation", as applicable, and any successor regulations. Any use, modification, reproduction release, performance, display or disclosure of the Licensed Software and Documentation by the U.S. Government shall be solely in accordance with the terms of this Agreement.

Symantec Corporation 350 Ellis Street Mountain View, CA 94043

<http://www.symantec.com>

Contents

Chapter 1	Best practices for SF Oracle RAC	7
	Introduction	8
	Intended audience	8
	List of acronymns	8
	About Veritas Storage Foundation for Oracle RAC	10
	Benefits of SF Oracle RAC	10
	Typical SF Oracle RAC configuration	11
	System configuration recommendations	13
	About network configuration	14
	Network configuration recommendations	15
	Storage recommendations	18
	Shared storage options	19
	Disk array requirements	19
	Oracle configuration recommendations	19
	Database storage recommendations	20
	Volume design recommendations	21
	File system recommendations	22
	I/O fencing recommendations	22
	Disaster recovery considerations	23
	Campus cluster	23
	Global cluster	25
	Data replication considerations	25
	Data replication options	25
	References	26
Appendix A	Recommended SF Oracle RAC deployment scenarios	27
	SF Oracle RAC cluster with VCS IPC and PrivNIC agent	27
	SF Oracle RAC cluster with UDP IPC and PrivNIC agent	30
	SF Oracle RAC cluster for multiple databases with UDP IPC and MultiPrivNIC agent	32
	SF Oracle RAC cluster with isolated Oracle traffic and MultiPrivNIC agent	34
	SF Oracle RAC cluster with NIC bonding, VCS IPC and PrivNIC agent	36

SF Oracle RAC cluster with NIC bonding, UDP IPC, and PrivNIC
agent 38

Best practices for SF Oracle RAC

This chapter includes the following topics:

- [Introduction](#)
- [Intended audience](#)
- [List of acronymns](#)
- [About Veritas Storage Foundation for Oracle RAC](#)
- [Typical SF Oracle RAC configuration](#)
- [System configuration recommendations](#)
- [About network configuration](#)
- [Storage recommendations](#)
- [Oracle configuration recommendations](#)
- [Database storage recommendations](#)
- [Volume design recommendations](#)
- [File system recommendations](#)
- [I/O fencing recommendations](#)
- [Disaster recovery considerations](#)
- [References](#)

Introduction

This document is a compilation of the best practices that are recommended for the deployment and administration of Veritas™ Storage Foundation for Oracle® RAC (SF Oracle RAC) in data centers. The various configuration practices discussed in this paper include system configuration, network configuration, storage configuration, Oracle database configuration, disaster recovery configuration, installation, and upgrade.

Intended audience

This document is intended to help system administrators, storage administrators, and database administrators deploy and administer SF Oracle RAC in production environments.

List of acronyms

This section provides the complete list of acronyms used in this document.

Table 1-1 List of acronyms

Acronym	Definition
ASM	Automatic Storage Management
ASL	Array Support Library
CF	Cache Fusion
CFS	Cluster File System
CRS	Cluster Ready Services or Oracle Clusterware
CVM	Cluster Volume Manager
DLM	Distributed Lock Manager
DMP	Dynamic Multi-Pathing
EBN	Enclosure Based Naming
FAN	Fast Application Notification
FC	Fibre Channel
GAB	Group Membership Services/Atomic Broadcast

Table 1-1 List of acronyms (*continued*)

Acronym	Definition
GCO	Global Cluster Option
GCS	Global Cache Service
GES	Global Enqueue Service
HCL	Hardware Compatibility List
IP	Internet Protocol
LLT	Low Latency Transport
LMX	LLT Multiplexer
OCR	Oracle Cluster Registry
OCSSD	Oracle Cluster Synchronization Services Daemon
ODM	Oracle Disk Manager
PGR	Persistent Group Reservation
RAC	Real Application Cluster
SCSI	Small Computer System Interface
SF	Storage Foundation
TCP	Transmission Control Protocol
VCS	Veritas Cluster Server
VCSIPC	VCS Interprocess Communication
VCSMM	Veritas Cluster Server Membership Module
VIP	Virtual Internet Protocol (address)
VXFEN	Veritas I/O Fencing
VVR	Veritas Volume Replicator
VxVM	Veritas Volume Manager

About Veritas Storage Foundation for Oracle RAC

Veritas Storage Foundation™ for Oracle® RAC (SF Oracle RAC) leverages proprietary storage management and high availability technologies to enable robust, manageable, and scalable deployment of Oracle RAC on UNIX platforms. The solution uses cluster file system technology that provides the dual advantage of easy file system management as well as the use of familiar operating system tools and utilities in managing databases.

The solution stack comprises the Veritas Cluster Server (VCS), Veritas Cluster Volume Manager (CVM), Veritas Cluster File System (CFS), and Veritas Storage Foundation, which includes the base Veritas Volume Manager (VxVM) and Veritas File System (VxFS).

Benefits of SF Oracle RAC

SF Oracle RAC provides the following benefits:

- Support for file system-based management. SF Oracle RAC provides a generic clustered file system technology for storing and managing Oracle data files as well as other application data.
- Support for high-availability of cluster interconnects. The combination of LMX/LLT protocols and the PrivNIC/MultiPrivNIC agents provides maximum bandwidth as well as high availability of the cluster interconnects, including switch redundancy.
- Use of clustered file system for placement of Oracle Cluster Registry and voting disks. Clustered file system and volume management technologies provide robust shared block and raw interfaces for placement of Oracle Cluster Registry and voting disks. In the absence of SF Oracle RAC, separate LUNs need to be configured for OCR and voting disks.
- Support for a standardized approach toward application and database management. A single-vendor solution for the complete SF Oracle RAC software stack lets you devise a standardized approach toward application and database management. Further, administrators can apply existing expertise of Veritas technologies toward SF Oracle RAC.
- Increased availability and performance using dynamic multi-pathing (DMP). DMP provides wide storage array support for protection from failures and performance bottlenecks in the HBAs and SAN switches.
- Easy administration and monitoring of SF Oracle RAC clusters from a single web console.
- Support for many types of applications and databases.

- Improved file system access times using Oracle Disk Manager (ODM).
- Ability to configure ASM disk groups over CVM volumes to take advantage of dynamic multi-pathing (DMP).
- Enhanced scalability and availability with access to multiple Oracle RAC instances per database in a cluster.
- Support for backup and recovery solutions using volume-level and file system-level snapshot technologies. SF Oracle RAC enables full volume-level snapshots for off-host processing and file system-level snapshots for efficient backup and rollback.
- Ability to failover applications without downtime using clustered file system technology.
- Prevention of data corruption in split-brain scenarios with robust SCSI-3 Persistent Reservation (PGR) based I/O fencing.
- Support for sharing all types of files, in addition to Oracle database files, across nodes.
- Fast disaster recovery with minimal downtime and interruption to users. Users can transition from a local high availability site to a wide-area disaster recovery environment with primary and secondary sites. If a node fails, clients that are attached to the failed node can reconnect to a surviving node and resume access to the shared database. Recovery after failure in the SF Oracle RAC environment is far quicker than recovery for a failover database.
- Verification of disaster recovery configuration using fire drill technology without affecting production systems.
- Support for a wide range of hardware replication technologies as well as block-level replication using VVR.

Typical SF Oracle RAC configuration

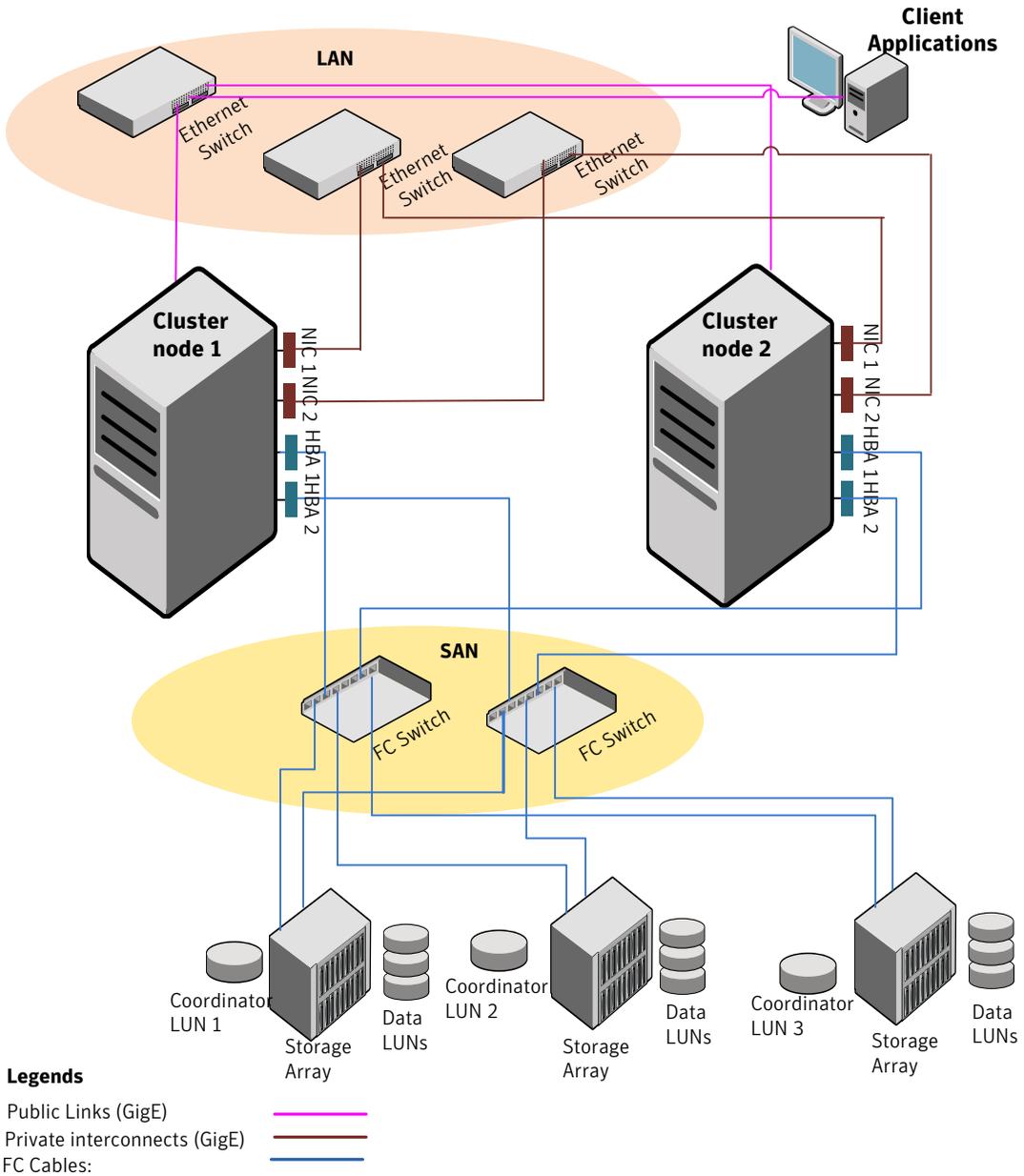
This section illustrates the physical view of a basic two-node SF Oracle RAC cluster.

For other recommended SF Oracle RAC deployment configurations:

See “*Recommended SF Oracle RAC deployment scenarios*” on page 27.

[Figure 1-1](#) illustrates a basic two-node SF Oracle RAC configuration.

Figure 1-1 Basic two-node SF Oracle RAC configuration



System configuration recommendations

The following recommendations provide guidelines for consistent configuration of nodes in an SF Oracle RAC cluster.

Architecture	Oracle RAC requires that all nodes in the cluster have the same architecture. Clusters with mixed architectures, such as x86, SPARC, IA/PA, are not supported.
CPU and memory considerations	The servers in the SF Oracle RAC cluster should have identical CPU speeds and memory.
Swap space requirements	<p>The minimum requirement for Oracle RAC 10g is 4 GB of swap space.</p> <p>The minimum requirement for Oracle RAC 11g is 8 GB of swap space.</p> <p>The operating system requirement for minimum swap is two times the size of RAM for swap space.</p> <p>Between the minimum requirements of Oracle RAC and the operating system, make sure that you meet the minimum requirement that is higher. For example, if the operating system requirement for minimum swap space computes to 5 GB on your Oracle RAC 11g systems, make sure that you meet the minimum swap space requirement of Oracle RAC, that is 8 GB.</p>
OS version and patches	The servers should have identical operating system versions and patches
SF Oracle RAC version	All nodes in the cluster must have the same SF Oracle RAC version.
Oracle version	<p>Use the latest Oracle patchset. The latest information on supported Oracle database versions is available in the following TechNote:</p> <p>http://entsupport.symantec.com/docs/280186</p> <p>SF Oracle RAC supports different versions of the Oracle database in a physical cluster. For example, you can install Oracle 10g R2 and Oracle 11g R1 database in the same cluster with Oracle 11g R1 Clusterware.</p>
System time on cluster nodes	All nodes in the cluster must have the same system time. The Network Time Protocol (NTP) daemon is recommended for synchronizing the time settings on each node.

Note: Before you install or upgrade SF Oracle RAC, run the Veritas Installation Assessment Service (VIAS) utility to verify whether the systems in your environment meet the necessary requirements. For information on downloading and running VIAS:

<https://vias.symantec.com/>

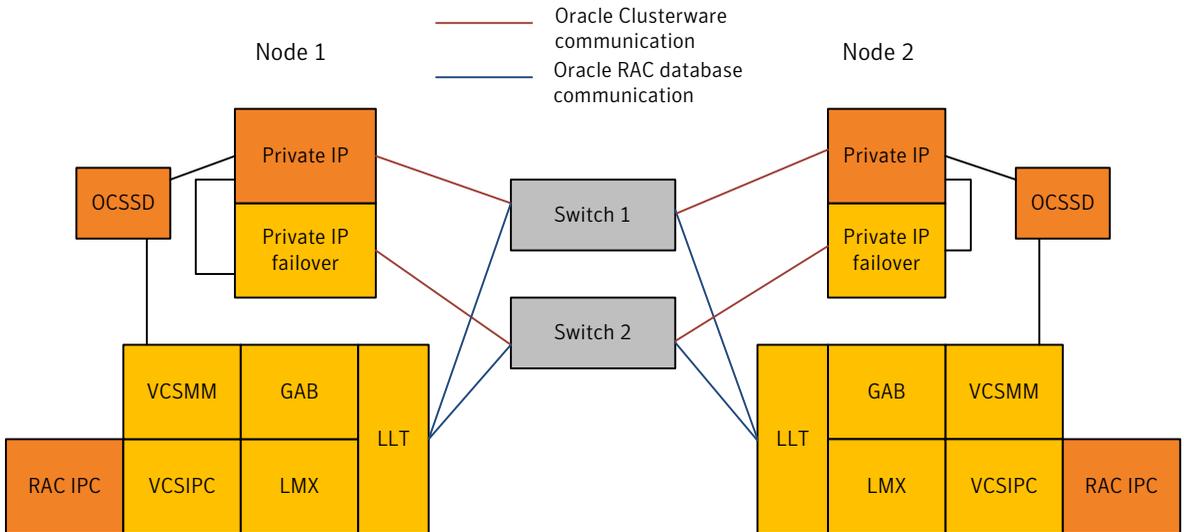
About network configuration

The network connectivity requirements for SF Oracle RAC and Oracle RAC are classified as follows:

- Private network (For communication within the cluster)
- Public network (For communication with the external world)

Figure 1-2 illustrates the communication layers between nodes.

Figure 1-2 Communication layers



The SF Oracle RAC and Oracle RAC components communicate within the cluster over the private network.

The communications can be categorized as follows:

Cluster communication	<p>The cluster communication constitutes the communication between the various components of SF Oracle RAC—Veritas Cluster Server, Cluster File System and Cluster Volume Manager.</p> <p>This communication is multiplexed over multiple LLT links specified in the LLT configuration file.</p>
Oracle Clusterware communication	<p>The Oracle Clusterware communication constitutes the Oracle Clusterware traffic, which includes Cluster Synchronization Service (heartbeat) traffic.</p> <p>The Oracle Clusterware communication uses a single logical connection over a single subnet. This network needs to be protected against link failures to prevent split-brain scenarios.</p>
Oracle RAC database communication (per database)	<p>The Oracle RAC database communication constitutes the RAC instance peer-to-peer communication for Global Cache Service (GCS) and Global Enqueue Service (GES), earlier referred to as Cache Fusion (CF) and Distributed Lock Manager (DLM).</p> <p>The bandwidth requirements for this communication may vary depending on the number of databases configured. The network needs to be protected against link failures to ensure uninterrupted database service.</p>

Network configuration recommendations

The following practices are recommended for a resilient network setup:

- Configure the private cluster interconnect over multiple dedicated gigabit Ethernet links. All single point of failures such as network interface cards (NIC), switches, and interconnects should be eliminated.
- Allow the NICs and switch ports to auto negotiate speed. The NICs used for the private cluster interconnect should have characteristics such as speed, MTU, and duplex on all nodes.
- Configure non-routable IP addresses for private cluster interconnects.
- Enable/disable Jumbo frames for network communication. Jumbo frames for network communications are not recommended when you use the Veritas supplied cache fusion library that employs Low Latency Transport (LLT) protocol for inter-instance communication. Jumbo frames are recommended when you use the User Datagram Protocol (UDP). Extreme inter-instance cache fusion load is necessary to yield any benefit from UDP and jumbo frames.

- In an SF Oracle RAC cluster, Oracle Clusterware heartbeat link **MUST** be configured as an LLT link. If Oracle Clusterware and LLT use different links for their communication, then the membership change between VCS and Oracle Clusterware is not coordinated correctly. For example, if only the Oracle Clusterware links are down, Oracle Clusterware kills one set of nodes after the expiry of the `css-miscount` interval and initiates the Oracle Clusterware and database recovery, even before CVM and CFS detect the node failures. This uncoordinated recovery may cause data corruption.
- Oracle Clusterware interconnects need to be protected against NIC failures and link failures. There are multiple options available to achieve the same. Configuring the Oracle Clusterware interconnects over bonded NIC interfaces or IP failover solutions such as IPMP can provide the first level of protection. In the absence of such mechanisms, the PrivNIC or MultiPrivNIC agent can be used to protect against NIC failures and link failures, if multiple links are available. Even if link aggregation solutions in the form of bonded NICs are implemented, the PrivNIC or MultiPrivNIC agent can be used to provide additional protection against the failure of the aggregated link by failing over to available alternate links. These alternate links can be simple NIC interfaces or bonded NICs.

Using IPMP for Oracle Clusterware

If Oracle Clusterware interconnects are configured over IPMP, all the NICs that are configured under LLT must be configured under the IPMP group. In such a configuration, it is recommended not to manage these links using the PrivNIC/MultiPrivNIC agents.

Using link aggregation/
NIC bonding for Oracle Clusterware

Starting with SF Oracle RAC 5.0 MP3, LLT recognizes aggregated links and treats the aggregated link as a single network interface. Make sure that a link configured under a aggregated link or NIC bond is not configured as a separate LLT link.

When LLT is configured over a bonded interface, a second link needs to be configured to prevent LLT from reporting jeopardy membership. The second link can be either a bonded interface or a simple NIC interface.

Using PrivNIC/MultiPrivNIC agents

The PrivNIC and MultiPrivNIC agents operate on LLT links. To use these agents, both Oracle Clusterware interconnects and Oracle RAC database communication links must be configured as LLT links.

In the event of a NIC failure or link failure, the agent fails over the private IP address from the failed link to the connected or available LLT link.

For more details on PrivNIC and MultiPrivNIC deployments:

See “[Recommended SF Oracle RAC deployment scenarios](#)” on page 27.

- Configure Oracle Cache Fusion traffic to take place through the private network. Oracle database clients use the public network for database services. Whenever there is a node failure or network failure, the client fails over the connection, for both existing and new connections, to the surviving node in the cluster with which it is able to connect. Client failover occurs as a result of Oracle Fast Application Notification, VIP failover and client connection TCP timeout. It is strongly recommended not to send Oracle Cache Fusion traffic through the public network.
- Configure multiple public networks for redundancy so that Oracle can failover virtual IP addresses if there is a public link failure.
- Set the recommended values for LLT peer inactivity timeout and CSS miss-count.

The LLT peer inactivity timeout value indicates the interval after which SF Oracle RAC on one node declares the other node in the cluster dead, if there is no network communication (heartbeat) from that node.

The default value for LLT peer inactivity timeout is 16 seconds. The value should be set based on service availability requirements and the propagation delay between the cluster nodes in case of campus cluster setup.

The CSS miss-count value indicates the amount of time Oracle Clusterware waits before evicting another node from the cluster, when it fails to respond across the interconnect.

The default value for the CSS miss-count in case of SF Oracle RAC is 600 seconds. The value of this parameter is kept much higher than the LLT peer inactivity timeout so that the two clusterwares, VCS and Oracle Clusterware, do not interfere with each other’s decisions on which nodes should remain in the cluster in the event of network split-brain. Veritas I/O fencing is allowed to decide on the surviving nodes first, followed by Oracle Clusterware.

Storage recommendations

The following practices are recommended for efficient functioning of the SF Oracle RAC stack:

- Distribute files between local and shared storage in the following manner for optimal performance benefits:

Local storage

Place the following binaries on local disks:

- SF Oracle RAC binaries
- Oracle Clusterware binaries
- Oracle binaries

Placing the Oracle binaries on local disks enables phased upgrade of the cluster.

Shared storage

Place the following binaries on shared disks:

- SF Oracle RAC fencing coordinator disks
- SF Oracle RAC database storage management repository
- Oracle RAC database
- Oracle Clusterware registry (OCR) and voting disk
- Database datafiles

Store the Oracle database files on CFS rather than on raw device or CVM raw device for easier management. Create separate clustered file systems for each Oracle database. Keeping Oracle database home on separate mount points enables you to unmount the database for maintenance purposes without affecting other databases.

If you plan to store the Oracle database on ASM, configure the ASM disk groups over CVM volumes to take advantage of dynamic multi-pathing.

- Database recovery data (archive, flash recovery)
Place redo and archived logs on CFS rather than on local file systems.
- OCR and vote disks

Place the OCR and vote disks on CVM volumes or CFS file systems to provide high availability to vote disks with dynamic multi-pathing.

Mirror the CVM volumes that contain OCR and vote devices for redundancy.

- CVM provides native naming (cXtXdX) as well as enclosure-based naming (EBN).

Use enclosure-based naming for easy administration of storage. Enclosure-based naming guarantees that the same name is given to a shared LUN on all the nodes, irrespective of the operating system name for the LUN.

Shared storage options

SF Oracle RAC provides the following options for shared storage:

- CVM
- CFS
- Oracle ASM over CVM

It is to be noted that, ASM provides storage only for the data files, control files, online and archive redo log files, and backup files. It does not support Oracle binaries, trace files, alert logs, export files, tar files, core files, Oracle cluster registry devices (OCR) and quorum device (voting disk), application binaries and data.

Disk array requirements

Make sure you meet the following requirements for disk arrays:

- SF Oracle RAC requires that coordinator disks for I/O fencing as well as data disks are SCSI-3 PR compatible.
- Review the current compatibility list to confirm compatibility of your storage array with SF Oracle RAC:
http://www.symantec.com/business/products/otherresources.jsp?pcid=2245&pvid=203_1
- Verify that the correct ASL (Array Support Library) is installed on all the cluster nodes.

Oracle configuration recommendations

For recommended Oracle storage practices:

See “[Storage recommendations](#)” on page 18.

The following Oracle configuration practices are recommended in addition to the storage practices:

Oracle database management

The VCS Oracle agent is recommended for managing Oracle databases. VCS controls the Oracle database in this configuration. The configuration without VCS Oracle agent may be used only in a single database setup.

High availability for Oracle Clusterware and Oracle RAC database communication

Use a native NIC bonding solution to provide redundancy, in case of NIC failures.

Alternatively, use one of the IP failover solutions, PrivNIC or MultiPrivNIC, in the following scenarios:

- Due to operating system limitations, you can not use NIC bonding to provide increased bandwidth using multiple network interfaces.
- The cluster is running multiple databases and you need to isolate the interconnect traffic.

In the event of NIC failure, the PrivNIC/MultiPrivNIC agent provides high availability by failing over the private IP address from the failed NIC to the available NIC.

To configure high availability in such a scenario, you need to specify the private IP address in the `CLUSTER_INTERCONNECTS` initialization parameter.

Additionally, you must remove the `cluster_interconnect` entry from OCR using the following command:

```
$ oifcfg delif
```

Note: When you use PrivNIC or MultiPrivNIC to provide high availability to the private IP addresses used for database cache fusion, do not use the `CLUSTER_INTERCONNECTS` parameter to load balance the interconnect traffic across multiple physical links (by setting multiple IP addresses in the `CLUSTER_INTERCONNECTS` parameter). Using the parameter for load balancing results in loss of high availability to the private interconnects.

ASM

- Use VxVM mirrored volumes with dynamic multi-pathing with external redundancy to ensure high availability.
- Do not use VxVM volumes, which are used by ASM, for any other purpose such as creation of file systems.
- Do not enable ODM when databases are installed on ASM.

Database storage recommendations

The following recommendations ensure better performance and availability of storage.

- Use multiple storage arrays, if possible, to ensure protection against array failures. The minimum recommended configuration is to have two HBAs for each host, two switches and two storage arrays.
- Design the storage layout keeping in mind performance and high availability requirements. Use technologies such as striping and mirroring.
- Use appropriate stripe width and depth to optimize I/O performance.
- Use SCSI-3 PGR compliant storage.
- Provide multiple access paths to disks with HBA/switch combinations to allow DMP to provide high availability against storage link failures and to provide load balancing.
- Use Database Dynamic Storage Tiering (DBDST) to optimize the storage cost. Using DBDST, less frequently used data can be moved to slower, less expensive disks. This also permits frequently accessed data to be stored on faster disks for quicker retrieval.

Volume design recommendations

The following recommendations ensure optimal layout of VxVM/CVM volumes.

- Mirror the volumes across two or more storage arrays, if using VxVM mirrors.
- Separate the Oracle recovery structures from the database files to ensure high availability when you design placement policies.
- Separate redo logs and place them on the fastest storage (for example, RAID 1+ 0) for better performance.
- Use "third-mirror break-off" snapshots for cloning the Oracle log volumes. Do not create Oracle log volumes on a Space-Optimized (SO) snapshot.
- Create as many Cache Objects (CO) as possible when you use Space-Optimized (SO) snapshots for Oracle data volumes.
- Distribute the I/O load uniformly on all Cache Objects when you create multiple Cache Objects.
- If using VxVM mirroring, keep the Fast Mirror Resync region size equal to the database block size to reduce the copy-on-write (COW) overhead. Reducing the region size increases the amount of Cache Object allocations leading to performance overheads.
- Implement zoning to control access to shared storage. Be aware that physical disks may be shared by multiple servers or applications and must therefore be protected from accidental access.

- Choose DMP I/O policy based on the storage network topology and the application I/O pattern.
- Exploit thin provisioning for better return on investment.

File system recommendations

The following recommendations ensure an optimal file system design for databases:

- If using VxVM mirroring, use ODM with CFS for better performance. ODM with SmartSync enables faster recovery of mirrored volumes using Oracle resilvering.
- Create separate file systems for Oracle binaries, data, redo logs, and archive logs. This ensures that recovery data is available if you encounter problems with database data files storage.
- Always place redo and archived logs on CFS file systems rather than local file systems.

I/O fencing recommendations

In an SF Oracle RAC cluster, Veritas fencing handles the split-brain scenarios even though Oracle Clusterware provides I/O fencing with OPROCD. It is therefore mandatory to enable fencing in SF Oracle RAC.

The following practices are recommended:

- Configure three LUNs as coordinator disks. Verify that the LUN is SCSI-3 compliant using the `vxfsntsthdw (1M)` utility.
- Provide high availability to coordinator disks using DMP and multiple physical access paths.
- In the case of SF Oracle RAC in a stretch cluster, place the third coordinator disk at a third location to ensure cluster availability during split-brain scenarios involving site failure.

Starting with SF Oracle RAC 5.0 MP3, iSCSI devices can be used as coordinator disks for I/O fencing.

Note: SF Oracle RAC supports iSCSI devices for I/O fencing only when the disk policy is set to DMP. If you use iSCSI devices, make sure that the `/etc/vxfsenmode` file has the disk policy set to DMP.

- Refer to the following Symantec Tech Note for various options available to lower the risk of split brain:
<http://support.veritas.com/docs/252635>

Disaster recovery considerations

SF Oracle RAC provides various disaster recovery configurations, such as campus clusters and global clusters, for multi-site clusters. In multi-site clusters, the nodes can be placed in different parts of a building, in separate buildings, or in separate cities. The distance between the nodes depends on the type of disaster from which protection is needed and on the technology used to replicate data. SF Oracle RAC supports various replication technologies for data replication.

To protect clusters against outages caused by disasters, the cluster components must be geographically separated.

Campus cluster

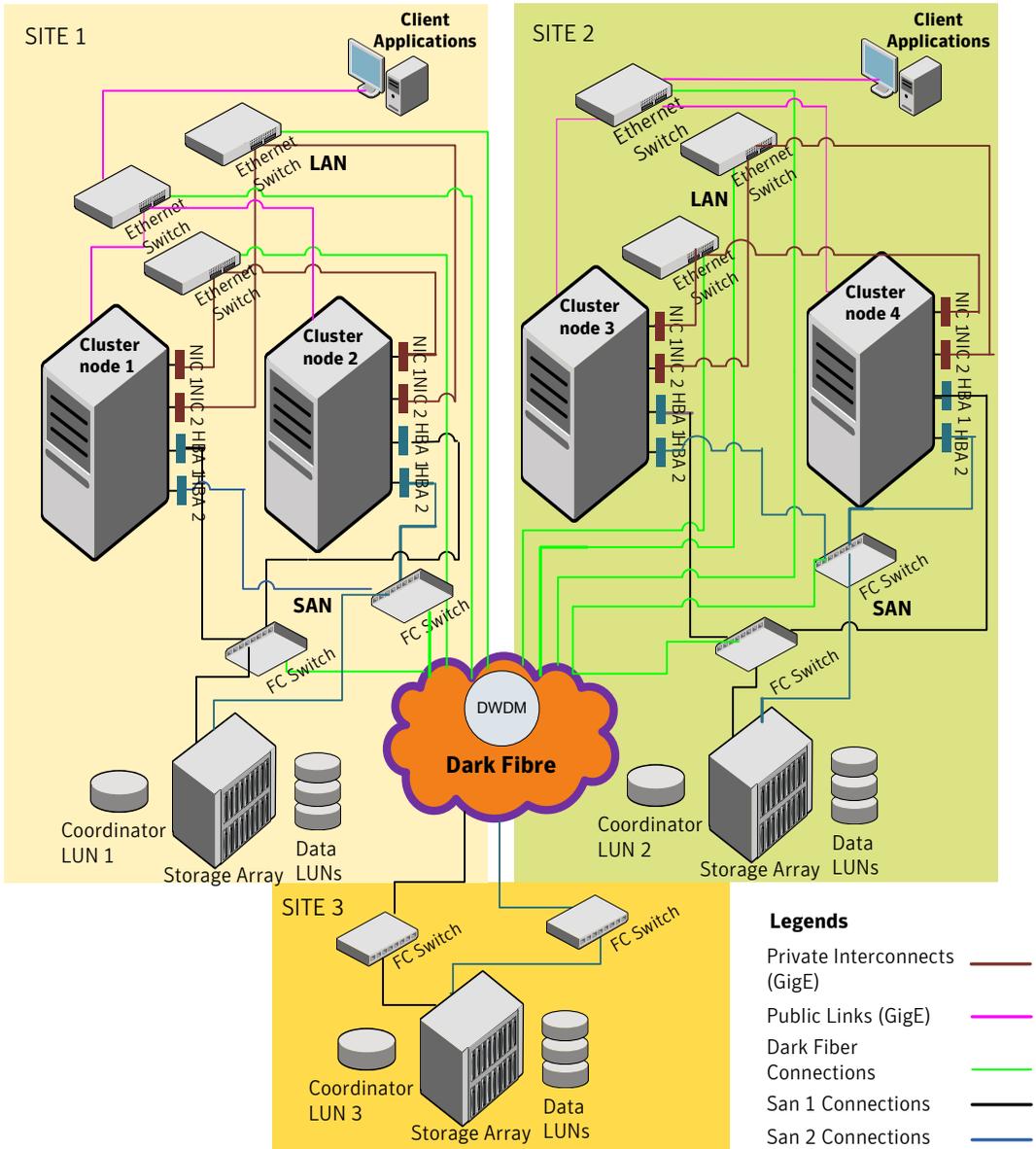
A campus cluster is also known as a stretch cluster or remote mirror configuration. In a campus cluster, the hosts and storage of a cluster span multiple sites separated by a few miles. These sites are typically connected using a redundant high-capacity network that provides access to storage and private link communication between the cluster nodes.

The following best practices may be borne in mind when you configure a SF Oracle RAC campus cluster:

- In the case of a two-node campus cluster, place the third coordinator disk on the third site. You may use iSCSI disk on the third site as an alternative to Dark Fiber connected FC-SAN disk.
- Enable the site consistency feature for site-aware plex detaches. Use site consistency tagging at the enclosure level.
- Set the 'allsites' attribute on the volumes so that the volumes are mirrored automatically across sites.
- Set the global detach policy on the disk groups. The local detach policy is not supported in SF Oracle RAC campus clusters.

Figure 1-3 shows the physical view of a four-node SF Oracle RAC campus cluster.

Figure 1-3 Four-node SF Oracle RAC campus cluster



Global cluster

Global clusters provide the ability to fail over applications between geographically distributed clusters when a disaster occurs.

Global clustering involves two steps:

1. Replication of data between the sites
2. Migration of the application when disaster occurs

The following aspects need to be considered when you design a disaster recovery solution:

- The amount of data lost in the event of a disaster
- The acceptable recovery time after the disaster

Data replication considerations

When you choose a replication solution, one of the important factors that you need to consider is the required level of data throughput. Data throughput is the rate at which the application is expected to write data. The impact of write operations on replication are of more significance than the read operations.

In addition to the business needs discussed earlier, the following factors need to be considered while choosing the replication options:

- Mode of replication
- Network bandwidth
- Network latency between the two sites
- Ability of the remote site to keep up with the data changes at the first site

Data replication options

SF Oracle RAC supports various hardware and software replication technologies.

Hardware replication options

- Hitachi True Copy
- IBM Metro Mirror
- IBM SVC
- EMC Mirror View

Software replication options

- Veritas Volume Replicator (VVR)
- Oracle Data Guard

References

For more information on SF Oracle RAC, refer to the following documents:

Veritas Storage Foundation for Oracle RAC 5.0 MP3 Installation Guide sfrac_install.pdf
Available on the SF Oracle RAC documentation CD

Veritas Storage Foundation for Oracle RAC 5.0 MP3 Installation Guide sfrac_admin.pdf
Available on the SF Oracle RAC documentation CD

Recommended SF Oracle RAC deployment scenarios

This appendix includes the following topics:

- [SF Oracle RAC cluster with VCS IPC and PrivNIC agent](#)
- [SF Oracle RAC cluster with UDP IPC and PrivNIC agent](#)
- [SF Oracle RAC cluster for multiple databases with UDP IPC and MultiPrivNIC agent](#)
- [SF Oracle RAC cluster with isolated Oracle traffic and MultiPrivNIC agent](#)
- [SF Oracle RAC cluster with NIC bonding, VCS IPC and PrivNIC agent](#)
- [SF Oracle RAC cluster with NIC bonding, UDP IPC, and PrivNIC agent](#)

SF Oracle RAC cluster with VCS IPC and PrivNIC agent

This section illustrates the recommended configuration for the following scenario:

Deployment scenario Oracle RAC 10g database cache fusion traffic is distributed over multiple LLT links using VCS IPC over LMX/LLT.

Recommendation Use the PrivNIC agent.

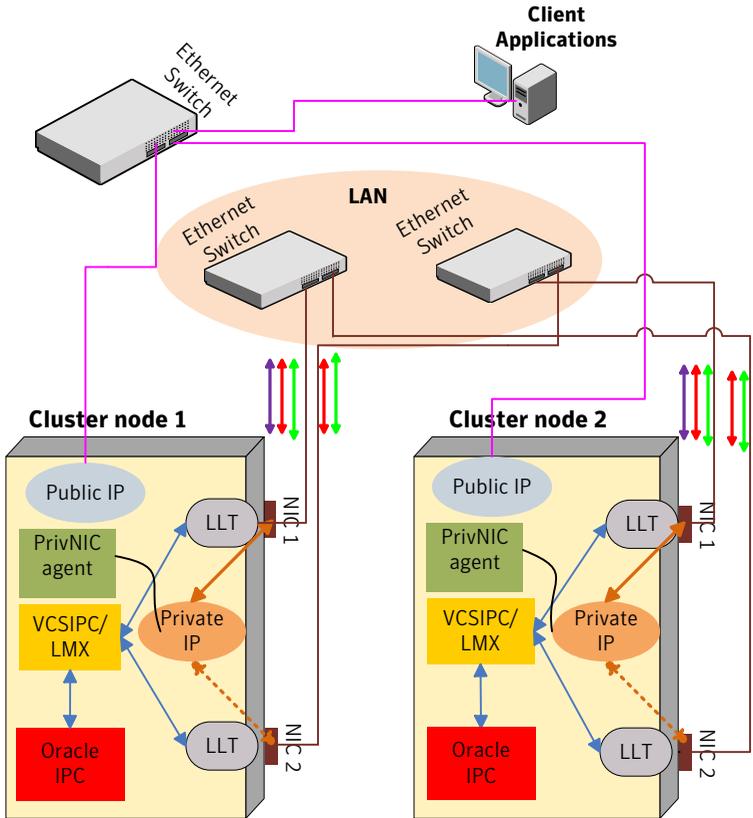
In the illustrated configuration:

- Oracle uses VCS IPC over LMX/LLT for cache fusion.
- One private network IP address is used for Oracle Clusterware communication that takes place over one of the LLT private interconnect links.

- The CFS/CVM/VCS metadata and the Oracle database cache fusion traffic travels through LLT over the two physical links that are configured as LLT links.
- In the event of a NIC failure or link failure, the PrivNIC agent fails over the private network IP address from the failed link to the available LLT link. When the link is restored, the IP address is failed back to the original link.

[Figure A-1](#) illustrates the logical view of a two-node SF Oracle RAC cluster with VCS IPC and PrivNIC agent (Oracle RAC 10g).

Figure A-1 SF Oracle RAC cluster with VCS IPC and PrivNIC agent (Oracle RAC 10g)



Legends

- | | | | |
|------------------------------|--|--|--|
| Public link (GigE) | | Oracle inter-process communication | |
| Private Interconnect (GigE) | | Active connection | |
| Oracle Clusterware Heartbeat | | Failover connection for CFS/CVM/VCS Metadata | |
| Oracle DB Cache Fusion | | PrivNIC Agent | |
| CFS/CVM/VCS Metadata | | | |

SF Oracle RAC cluster with UDP IPC and PrivNIC agent

This section illustrates the recommended configuration for the following scenario:

Deployment scenario	Oracle RAC 10g or Oracle RAC 11g configured with UDP IPC for database cache fusion.
Recommendation	Use the PrivNIC agent.

In the illustrated configuration:

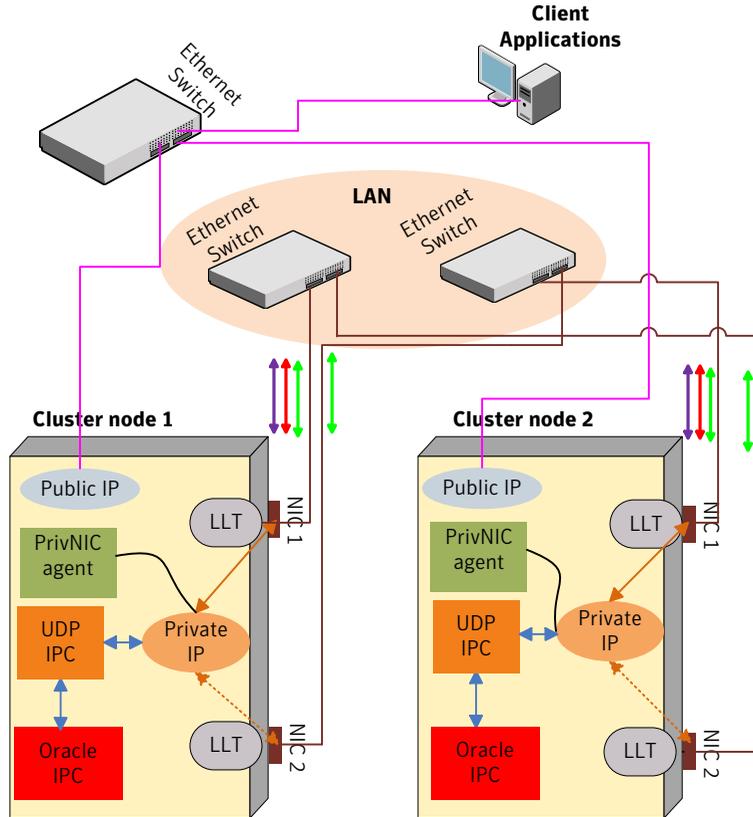
- A common IP address is used for Oracle Clusterware communication and Oracle database cache fusion.

Note: You need to specify the private network IP address used for Oracle database cache fusion in the `CLUSTER_INTERCONNECTS` initialization parameter. Remove the `cluster_interconnect` entry, if it exists in the OCR, using the following command: `$ oifcfg delif.`

- Oracle Clusterware communication and Oracle database cache fusion traffic flows over one of the LLT private interconnect links.
- The CFS/CVM/VCS metadata travels through LLT over the two physical links that are configured as LLT links.
- In the event of a link failure, the PrivNIC agent fails over the private network IP address from the failed link to the available LLT link. When the link is restored, the IP address is failed back to the original link.

[Figure A-2](#) illustrates the logical view of a two-node SF Oracle RAC cluster with UDP IPC and PrivNIC agent (Oracle RAC 10g/Oracle RAC 11g).

Figure A-2 SF Oracle RAC cluster with UDP IPC and PrivNIC agent (Oracle RAC 10g/Oracle RAC 11g)



Legends

- | | | | |
|------------------------------|--|------------------------------------|--|
| Public link (GigE) | | Oracle inter-process communication | |
| Private Interconnect (GigE) | | Active connection | |
| Oracle Clusterware Heartbeat | | Failover connection for | |
| Oracle DB Cache Fusion | | PrivNIC Agent | |
| CFS /CVM / VCS Metadata | | | |

SF Oracle RAC cluster for multiple databases with UDP IPC and MultiPrivNIC agent

This section illustrates the recommended configuration for the following scenario:

Deployment scenario	Oracle RAC 10g or Oracle RAC 11g is configured with UDP IPC for database cache fusion.
Recommendation	Use the MultiPrivNIC agent.

In the illustrated configuration:

- One private IP address is used for each database for database cache fusion. One of the private IP addresses used for the database cache fusion is shared by Oracle Clusterware communication.

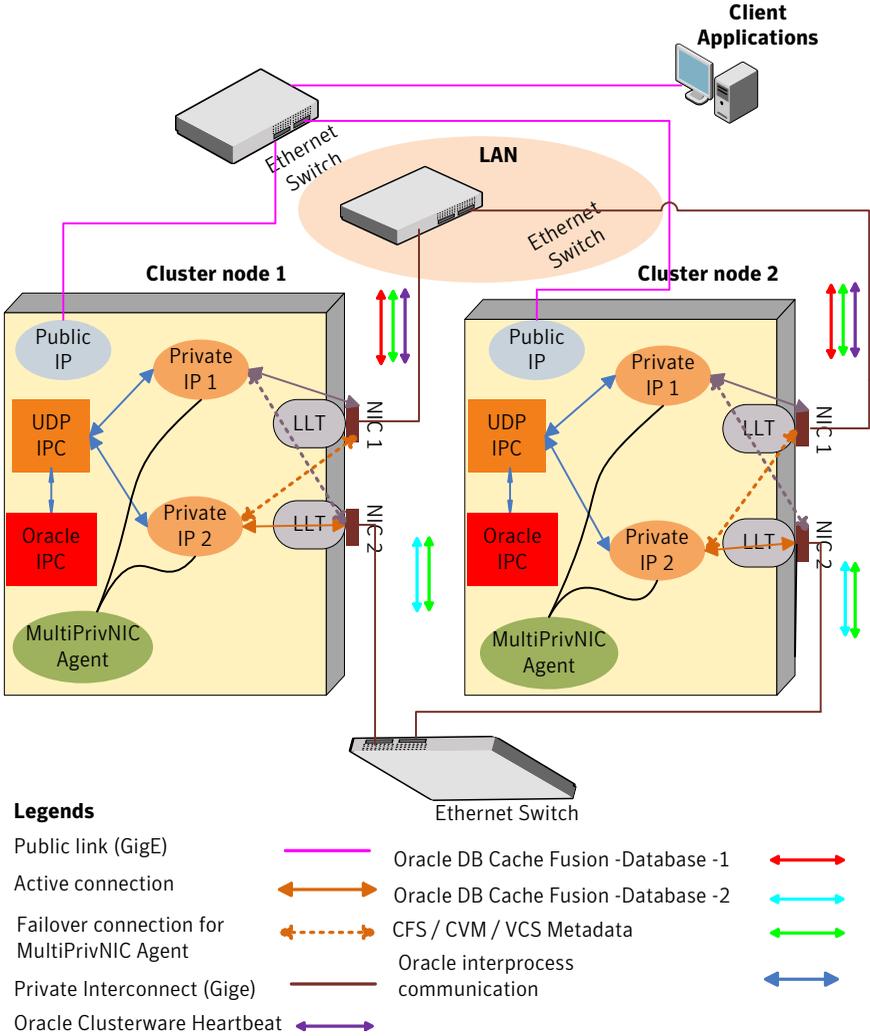
Note: You need to specify the private network IP address used for Oracle database cache fusion in the `CLUSTER_INTERCONNECTS` initialization parameter for each database. Remove the `cluster_interconnect` entry, if it exists in the OCR, using the following command: `$ oifcfg delif.`

The CFS/CVM/VCS metadata also travels through these links.

- In the event of a link failure, the MultiPrivNIC agent fails over the private network IP address from the failed link to the available LLT link. When the link is restored, the IP address is failed back to the original link.

[Figure A-3](#) illustrates the logical view of a two-node SF Oracle RAC cluster for multiple databases with UDP IPC and MultiPrivNIC agent (Oracle RAC 10g/Oracle RAC 11g).

Figure A-3 SF Oracle RAC cluster for multiple databases with UDP IPC and MultiPrivNIC agent (Oracle RAC 10g/Oracle RAC 11g)



SF Oracle RAC cluster with isolated Oracle traffic and MultiPrivNIC agent

This section illustrates the recommended configuration for the following scenario:

Deployment scenario	Oracle RAC 10g or Oracle RAC 11g database cache fusion traffic is isolated from the CFS/CVM/VCS metadata.
Recommendation	Use the MultiPrivNIC agent.

In the illustrated configuration:

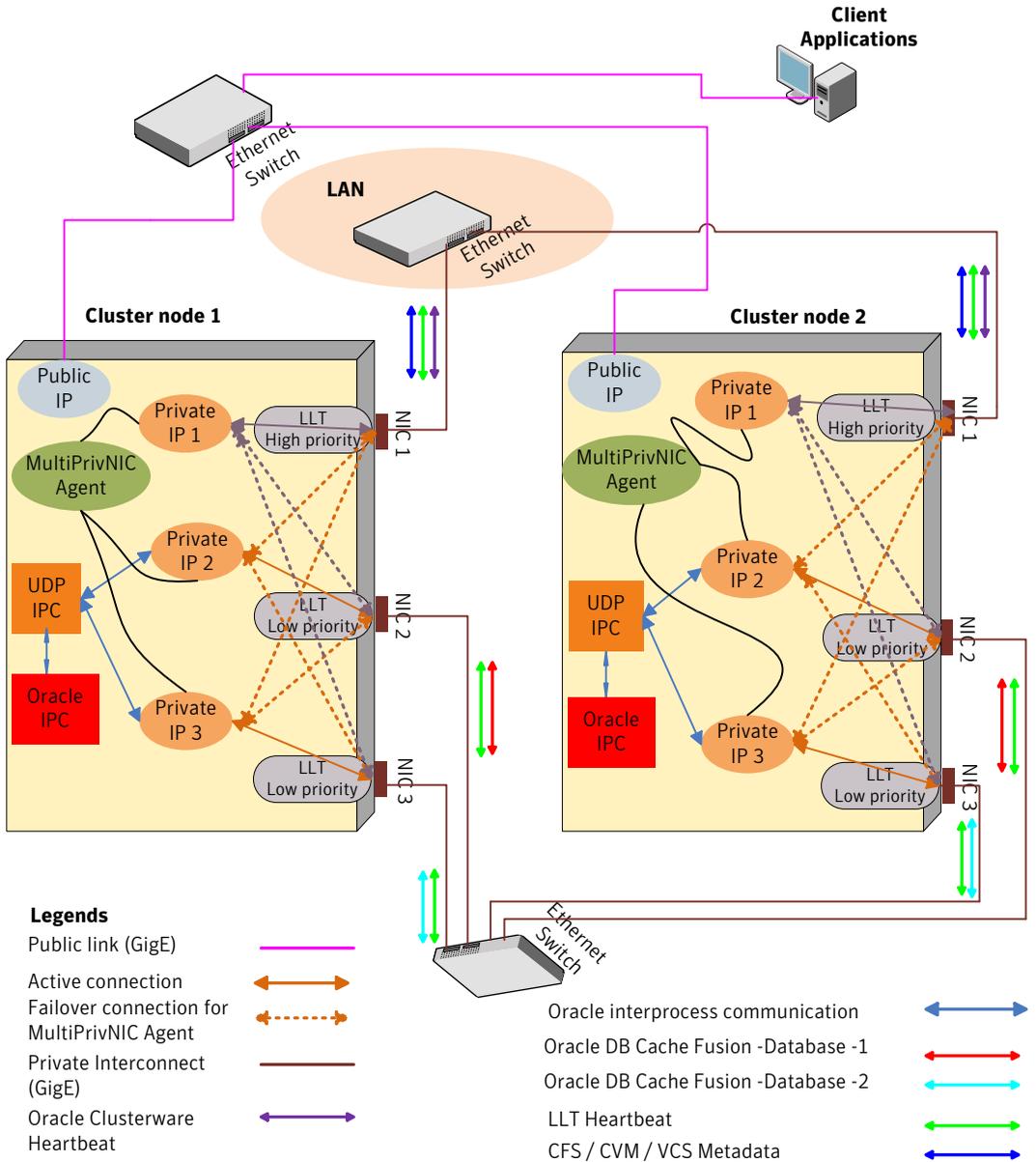
- The private network IP address used for database cache fusion is configured over a dedicated link for each database. These links are configured as low-priority LLT links.

Note: You need to specify the private network IP addresses used for Oracle database cache fusion in the `CLUSTER_INTERCONNECTS` initialization parameter. Remove the `cluster_interconnect` entry, if it exists in the OCR, using the following command: `$ oifcfg delif`

- The private network IP address used for Oracle Clusterware communication is configured over a high-priority LLT link. This link is also used for the CFS/CVM/VCS metadata transfer.
- In the event of a link failure, the MultiPrivNIC agent fails over the private network IP address from the failed link to the available LLT link. When the link is restored, the IP address is failed back to the original link.

[Figure A-4](#) illustrates the logical view of a two-node SF Oracle RAC cluster with Oracle traffic isolated from VCS / CVM / CFS traffic (Oracle RAC 10g/Oracle RAC 11g).

Figure A-4 SF Oracle RAC cluster with isolated Oracle traffic and MultiPrivNIC agent (Oracle RAC 10g/Oracle RAC 11g)



SF Oracle RAC cluster with NIC bonding, VCS IPC and PrivNIC agent

This section illustrates the recommended configuration for the following scenario:

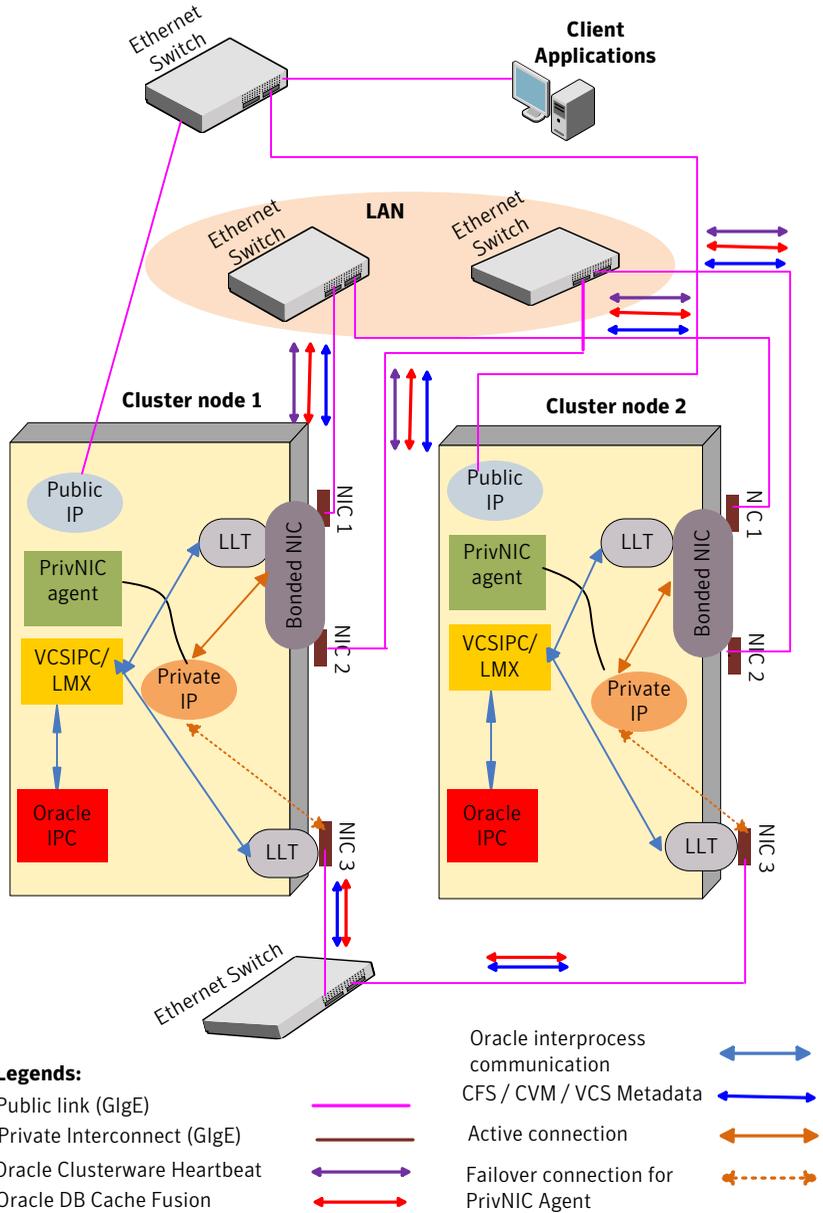
Deployment scenario	A bonded NIC interface is used along with another NIC interface to distribute Oracle RAC 10g database cache fusion traffic using VCS IPC over LMX/LLT.
Recommendation	Use the PrivNIC agent.

In the illustrated configuration:

- Oracle uses VCS IPC over LMX/LLT for cache fusion.
- One common private network IP address is used for Oracle database cache fusion and Oracle Clusterware communication that takes place over the bonded NIC interface, configured as an LLT link.
- The Oracle Clusterware communication takes place over the other LLT private interconnect link.
- The Oracle database cache fusion as well as the CFS/CVM/VCS metadata travels through LLT over the three physical links.
- In the event of a bonded NIC interface failure, the PrivNIC agent fails over the private network IP address from the failed link to the other available LLT link. When the link is restored, the IP address is failed back to the original link.

[Figure A-5](#) illustrates the logical view of a two-node SF Oracle RAC cluster with NIC bonding, VCS IPC, and PrivNIC agent (Oracle RAC 10g).

Figure A-5 SF Oracle RAC cluster with NIC bonding, VCS IPC and PrivNIC agent (Oracle RAC 10g)



SF Oracle RAC cluster with NIC bonding, UDP IPC, and PrivNIC agent

This section illustrates the recommended configuration for the following scenario:

Deployment scenario	Oracle RAC 10g or Oracle RAC 11g with UDP IPC is configured to use a bonded NIC interface for distribution of Oracle database cache fusion traffic. A second link is configured as a standby link.
Recommendation	Use the PrivNIC agent.

In the illustrated configuration:

- A common IP address is used for Oracle Clusterware communication and Oracle database cache fusion that is distributed over two underlying physical links of the bonded NIC interface. The bonded NIC interface is configured as a single LLT link.

Note: You need to specify the private network IP addresses used for Oracle database cache fusion in the `CLUSTER_INTERCONNECTS` initialization parameter. Remove the `cluster_interconnect` entry, if it exists in the OCR, using the following command: `$ oifcfg delif`

- The CFS/CVM/VCS metadata also travels through the bonded link.
- In the event of a bonded link failure, the PrivNIC agent fails over the private network IP address from the failed link to the available standby LLT link. When the link is restored, the IP address is failed back to the original link.

[Figure A-6](#) illustrates the logical view of a two-node SF Oracle RAC cluster with NIC bonding, UDP IPC, and PrivNIC agent.

Figure A-6 SF Oracle RAC cluster with NIC bonding, UDP IPC, and PrivNIC agent

