Oracle Fusion Middleware 11g Disaster Recovery Solution Using HP Blade Servers and HP EVA Storage

Oracle Maximum Availability Architecture White Paper
July, 2010

Maximum Availability Architecture

Oracle Best Practices For High Availability

ORACLE

Oracle Fusion Middleware 11g Disaster Recovery Solution Using HP Blade Servers and HP EVA Storage

Executive Overview 3
Introduction 3
Terminology 3
Oracle DR Terminology 3
HP Continuous Access EVA Terminology 5
Oracle Fusion Middleware Disaster Recovery Strategy 7
HP Storage works EVA and HP Continuous Access EVA 8
Introduction 8
HP Continuous Access EVA Features 9
HP Continuous Access EVA Remote Replication Concepts 10
Write Modes 10
DR Groups 10
DR Group Write History Log 13
Managed Sets 13
Failover 13
Use case Objectives 14
Test Cases 14
Reference Architecture 15
System Configuration 16
Hardware Diagram 16
Hardware and Operating System Specifications 17
Planning DR Groups 18
Network Configuration 19
Storage Software Configuration 20
Recommendations During Normal Operations 22
Managing Planned and Unplanned Downtime 22
Site Switchover Procedures 22
Switchback Procedures 26
Site Failover Procedures 26
Failback Procedures 27
Best Practices 27
Choosing Replication Write Modes 27
Choosing the Size of Your Write History Log 29
HP Boot From SAN 29
High Availability for Persistent Stores 30
Data loss and latency requirements 30
HP Virtual Connect for HP BladeSystem Infrastructure Virtualization 30
Local Server Failover for HP Blade Infrastructure using HP Insight Software 31
Oracle Fusion Middleware High Availability Technologies 33
Conclusion 35
References 36

Oracle Fusion Middleware 11g Disaster Recovery Solution Using HP Blade Servers and HP EVA Storage

EXECUTIVE OVERVIEW

The Oracle Fusion Middleware 11g Disaster Recovery solution uses disk replication technology provided by storage vendors for disaster protection of key information on file systems. In addition, Oracle Data Guard is used for replicating database content. This document describes how disaster recovery can be achieved for an Oracle Fusion Middleware 11g environment on HP StorageWorks EVA storage using its replication data functionality, the HP StorageWorks Continuous Access EVA.

INTRODUCTION

Enterprise deployments need protection from unforeseen disasters and natural calamities. One protection solution involves setting up a standby site at a geographically different location than the production site. The standby site may have equal or fewer services and resources compared to the production site. All data including application data, metadata, configuration data, and security data are replicated to the standby site. The standby site is normally in a passive mode; it is started when the production site is not available. This deployment model is sometimes referred to as an active/passive model. This is also applicable to Oracle Fusion Middleware Disaster Recovery solution, which depends upon storage replication techniques to keep middle-tiers across production and standby sites synchronized in addition to using Oracle Data Guard for keeping Oracle databases synchronized.

TERMINOLOGY

Oracle DR Terminology

- Disaster Recovery - The ability to safeguard against natural disasters or unplanned outages at a production site by having a recovery strategy for failing over applications and data to a geographically separate standby site.
- Oracle Fusion Middleware - a collection of best of breed, standards-based software products that spans a range of tools and services: from Java EE and developer tools, to integration services, business intelligence, and

collaboration. Oracle Fusion Middleware offers complete support for development, deployment, and management.

- Oracle SOA Suite: Oracle SOA Suite is a comprehensive, hot-pluggable software suite to build, deploy and manage applications or systems based on Service-Oriented Architectures (SOA). The components of the suite benefit from common capabilities including consistent tooling, a single deployment and management model, end-to-end security and unified metadata management. The key components of the Oracle SOA Suite include BPEL Process Manager, Human Workflow, Adapters, Business Rules, Business Activity Monitoring, Complex Event Processing, Oracle Service Bus, B2B and Oracle Web Services Manager
- Topology - The production site and standby site hardware and software components that comprise an Oracle Fusion Middleware Disaster Recovery solution.
- Site failover - The process of making the current standby site the new production site after the production site becomes unexpectedly unavailable (for example, due to a disaster at the production site).
- Site switchover - The process of reversing the roles of the production site and standby site. Switchovers are planned operations done for periodic validation or to perform planned maintenance on the current production site. During a switchover, the current standby site becomes the new production site, and the current production site becomes the new standby site.
- Site Switchback – The process of reversing the roles of the new production site (old standby) and new standby site (old production). Switchback is applicable after a previous switchover.
- Instantiation – The process of creating a topology at the standby site (after verifying that the primary and standby sites are valid for FMW Disaster Recovery) and synchronizing the standby site with the primary site so that the primary and standby sites are consistent.
- Site synchronization - The process of applying changes made to the production site at the standby site. For example, when a new application is deployed at the production site, you should perform synchronization so that the same application will be deployed at the standby site.
- Oracle home: An Oracle home contains installed files necessary to host a specific product. For example, the SOA Oracle home contains a directory that contains binary and library files for Oracle SOA Suite. An Oracle home resides within the directory structure of the Middleware home. Each Oracle home can be associated with multiple Oracle instances or Oracle WebLogic Server domains.
- WebLogic Server home: A WebLogic Server home contains installed files necessary to host an Oracle WebLogic Server. The WebLogic Server

home directory is a peer of Oracle home directories and resides within the directory structure of the Middleware home.

- Middleware home: A Middleware home consists of the Oracle WebLogic Server home, and, optionally, one or more Oracle homes. A Middleware home can reside on a local file system or on a remote shared disk that is accessible as a SAN or a NAS.
- Oracle instance: An Oracle instance contains one or more active middleware system components, for example Oracle Web Cache, Oracle HTTP Server, or Oracle Internet Directory. You determine which components are parts of an instance, either at install time or by creating and configuring an instance at a later time. An Oracle instance contains files that can be updated, such as configuration files, log files, temporary files.
- Weblogic Server Transaction Logs – Each Weblogic Server instance has a transaction log that captures information about committed XA transactions that may not have completed. The transaction logs enable Weblogic Server to recover transactions that could not be completed before the server failed.
- Recovery Point Objective (RPO) - The maximum age of the data you want to restore in the event of a disaster. For example, if your RPO is 6 hours, you want to be able to restore systems back to the state they were in as of no longer than 6 hours ago.
- Recovery Time Objective (RTO) -The time needed to recover from a disaster—usually determined by how long you could afford to be without your systems.

HP Continuous Access EVA Terminology

- Array - See virtual array and storage system.
- Asynchronous - A descriptive term for computing models that eliminate timing dependencies between sequential processes. In asynchronous write mode, the array controller acknowledges that data has been written at the source before the data is copied at the destination. Asynchronous mode is an optional DR group property. See also synchronous.
- Bandwidth - The transmission capacity of a link or system, usually measured in bits per second.
- Copy Set - A source-destination pair of virtual disks.
- Disk Group - A named group of disks selected from all available disks in an array. One or more virtual disks can be created from a disk group.
- DR Group - Data replication group. A logical group of virtual disks in a remote replication relationship with a corresponding group on another array.
- Destination – The virtual disk, DR group, or virtual array where I/O is stored after replication.

- Enhanced Asynchronous - A write mode in which all host write I/Os are added to write history log. The controller then acknowledges that data has been written at the source before the data is copied at the destination.
- Enterprise Virtual Array (EVA) - An HP StorageWorks product that consists of one or more virtual arrays. See also virtual arrays.
- E-Port - Port used for connection between two Fiber Channel switches. Any port on the switch can be E-port.
- Fabric - A network of Fiber Channel switches or hubs and other devices.
- HP Continuous Access EVA - A storage-based HP StorageWorks product consisting of two or more arrays performing disk-to-disk replication, along with the management user interfaces that facilitate configuring, monitoring, and maintaining the replicating capabilities of the arrays.
- Intersite Link - A connection from an E-port on a local switch to an E-port on a remote switch.
- LUN - Logical unit number. Logical units are the components within SCSI targets that execute I/O commands. Virtual disks that are presented to hosts correspond to logical units and are identified by LUN IDs.
- Managed Set - Selected resources that are grouped for convenient management. For example, you can create a managed set to manage all DR groups whose sources reside in the same rack.
- Management Server - A server, on which HP StorageWorks Enterprise Virtual Array (EVA) management software is installed, including HP StorageWorks Command View EVA and HP StorageWorks Replication Solutions Manager, if used. A dedicated management server runs EVA management software exclusively.
- Mount Point - The file system path and directory where a host volume is accessed.
- Normalization - The initial copy that occurs between source and destination virtual disks or any complete re-synchronization that occurs after the initial copy. (See Instantiation in Oracle DR Terminology section)
- Present a LUN- This is the process in which the Management Console of the storage makes the LUN or virtual disk to be presented (made visible) to the World Wide ID (WWID) of the host (db/middleware) server qlogic hba.
- Remote Copy - A virtual disk on the destination array that is a replica of a virtual disk in the source array.

- Source - The virtual disk, DR group, or virtual array where I/O is stored before replication. See also destination.
- Source-Destination Pair - A copy set.
- Storage Area Network (SAN) - A network of storage devices and the initiators that store and retrieve information on those devices, including the communication infrastructure.
- Storage System - Synonymous with virtual array. The HP StorageWorks Enterprise Virtual Array consists of one or more storage systems. See also virtual array.
- Synchronous - A descriptive term for computing models that perform tasks in chronological order without interruption. In synchronous write mode, the source waits for data to be copied at the destination before acknowledging that it has been written at the source. See also asynchronous.
- Virtual Array - Synonymous with disk array and storage system, a group of disks in one or more disk enclosures combined with control software that presents disk storage capacity as one or more virtual disks. See also virtual disk.
- Virtual Disk - Variable disk capacity that is defined and managed by the array controller and presentable to hosts as a disk.
- XCS - The HP Enterprise Virtual Array software on specific EVA controller models. Controller software manages all aspects of array operation, including communication with HP StorageWorks Command View EVA.

ORACLE FUSION MIDDLEWARE DISASTER RECOVERY STRATEGY

Oracle Fusion Middleware has the following artifacts –

- Middleware Product Binaries, Configuration and Metadata Files – Use Disk Replication Technologies offered by storage vendors
- Database content – Use Oracle Data Guard for Oracle Database (and Vendor recommended solutions for non-Oracle databases).

The Oracle Fusion Middleware Disaster Recovery solution requires syncing up these artifacts between two sites.

In case of a failure or planned outage of production site, replication to the standby site will be stopped. The services and applications will subsequently be started on the standby site. As a result, the standby site becomes the new production site. The network traffic should then be routed to the standby site.

For Oracle Database Content, because of its superior level of protection and higher availability, Oracle Data Guard is recommended solution for disaster protection of Oracle Databases. This includes the databases used for Oracle Fusion Middleware Repositories, as well as customer data.

HP STORAGE WORKS EVA AND HP CONTINUOUS ACCESS EVA

Introduction

The Oracle Fusion Middleware Disaster Recovery solution depends upon four specific components from HP StorageWorks:

- The storage array itself, in this case, HP StorageWorks Enterprise Virtual Arrays (EVAs)
- The fabric, the network of Fibre Channel switches that connects the arrays
- The array management software, Replication Solutions Manager (RSM) and Command View EVA
- The replication software, HP Continuous Access EVA software, which is integrated into the management software

HP StorageWorks Continuous Access EVA software is the critical component of Oracle Fusion Middleware Disaster Recovery Solution. HP Continuous Access EVA software provides an array-based application that uses advanced replication technologies to replicate data over distance between HP StorageWorks Enterprise Virtual Arrays. HP StorageWorks Continuous Access EVA utilizes the graphical user interface (GUI) provided by Replication Solutions Manager (RSM) software to create, manage and configure remote replication on the entire HP StorageWorks EVA family of storage arrays. With the combination of unique remote replication technologies and the easy to use RSM Software interface, enterprises can be confident their information is protected in the event of a disaster.

For the Continuous Access EVA Software to work properly, some hardware configuration is required. Specifically, EVA's at two different sites must be connected through intersite links (ISL). The relationship between the hardware and software components is diagramed in Figure 1.

Figure 1 shows a typical remote replication setup with arrays on local and remote sites connected by two linked fabrics. Two intersite links (ISL) connect the local and remote fabrics.

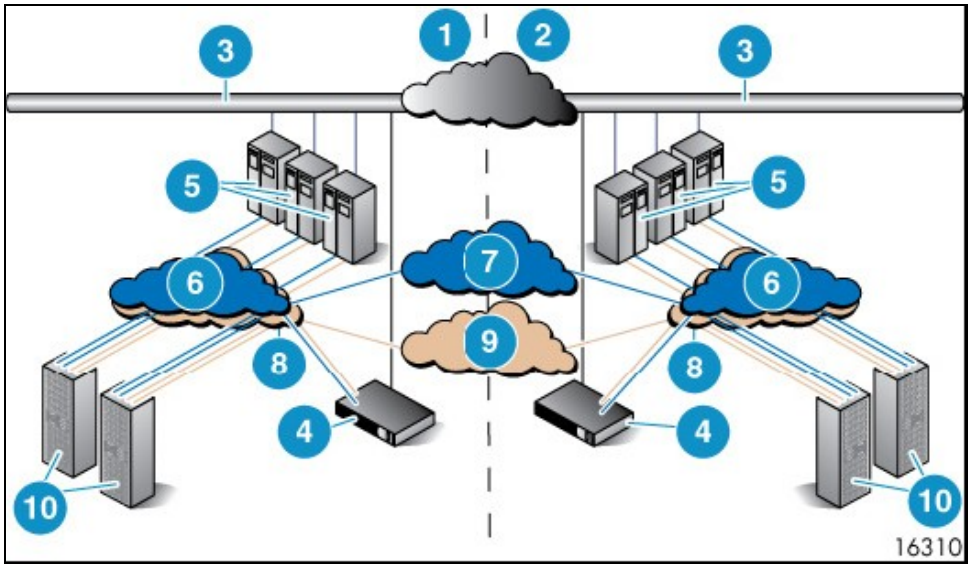


Figure 1 Basic HP Continuous Access EVA Configuration

1. Local site
2. Remote site
3. LAN connection
4. Management server
5. Hosts
6. Local/remote fabric—blue
7. Intersite link—blue
8. Local/remote fabric—gold (for High Availability)
9. Intersite link—gold (for High Availability)
10. Enterprise Virtual Arrays

HP Continuous Access EVA Features

- Continuous replication of local virtual disks to remote virtual disks
- Synchronous and asynchronous replication modes
- Automated failover when used with other solution software
- Failsafe data protection
- Ability to suspend and resume replication
- Bidirectional replication
- Graphical and command line user interfaces
- Automatic suspension of replication if the link between arrays is down
- Support for array-to-array fan-in and fan-out

- Business copy feature facilitates backup of Vdisks, and complements the HP Continuous Access EVA as a best practice. Some of the features of business copy include
    - Snapclone takes physical backup of the disks at a point in time.
    - Vsnap takes a snapshot of a virtual disk.

For more information on any of the features listed above, please see the HP StorageWorks Continuous Access EVA implementation guide (see References section)

HP Continuous Access EVA Remote Replication Concepts

In HP Continuous Access EVA terminology, remote replication is the continuous copying of data from selected virtual disks on a source (local) array to replica virtual disks on a destination (remote) array. Applications continue to run while data is replicated in the background. Remote replication requires a fabric connection (intersite link) between the source and destination arrays and a software connection (DR group) between source virtual disks and destination virtual disks.

Write Modes

The remote replication write modes are:

- Asynchronous—The array acknowledges I/O completion before data is replicated on the destination array. Asynchronous write mode can be standard or enhanced, depending on the software version of the controller.
- Synchronous—The array acknowledges I/O completion only after the data is cached on both the local and destination arrays.

For more information on choosing write modes for Oracle Fusion Middleware Disaster Recovery, see "Choosing Replication Write Modes" in the Best Practices section

DR Groups

A DR group is a logical group of virtual disks in a remote replication relationship between two arrays. The DR group maps to the notion of Consistency Groups as described in the Oracle Disaster Recovery Solution Guide. Hosts write data to the virtual disks in the source array, and the array copies the data to the corresponding virtual disks in the destination array. I/O ordering is maintained across the virtual disks in a DR group, ensuring I/O consistency on the destination array in the

event of a failure of the source array. The virtual disks in a DR group fail over together, share a write history log (DR group log), and preserve write order within the group. A pair of source and destination virtual disks is called a ***copy set***.

Figure 2 illustrates the replication of one DR group between a source array and a destination array. For more information, see "Planning DR groups" under the Best Practices section

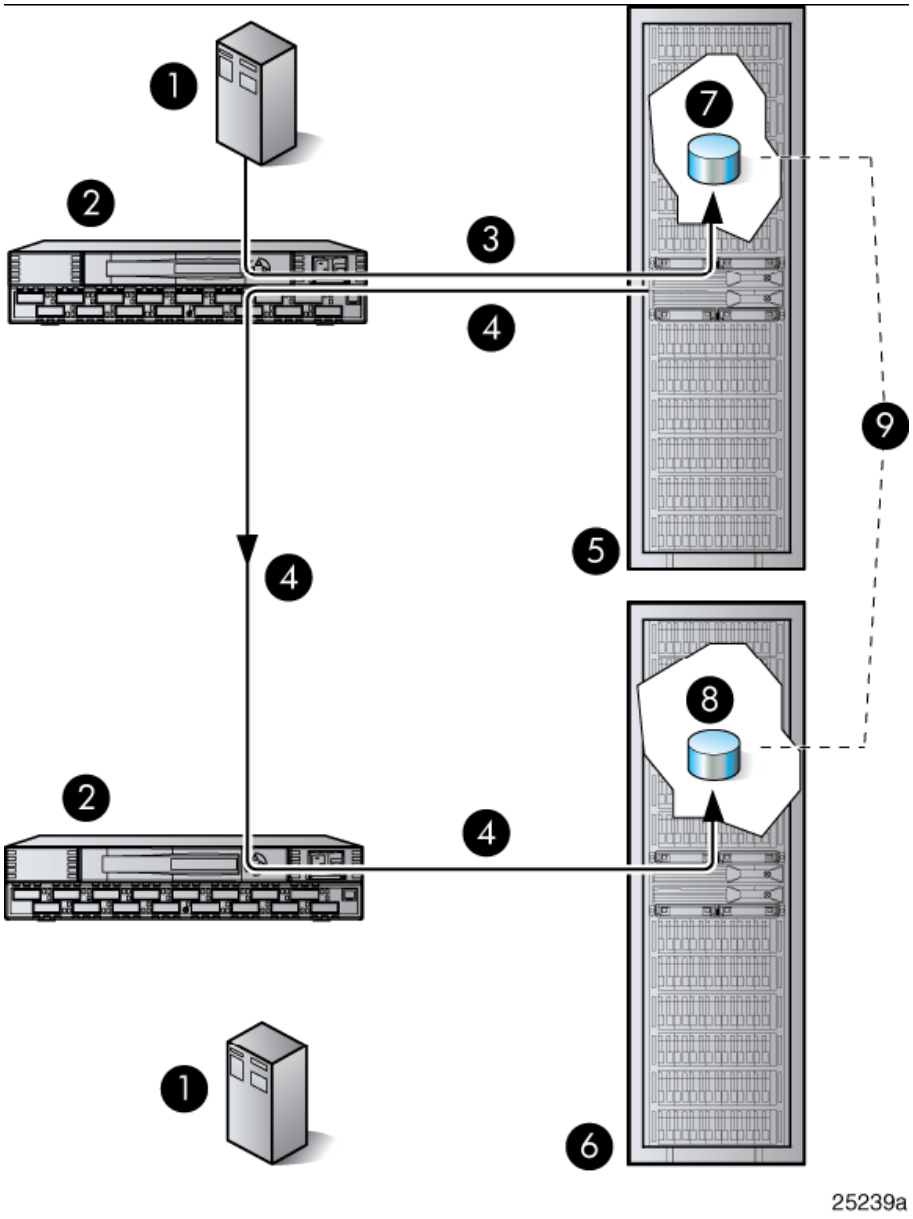


Figure 2 DR Group Replication

1. Host server
2. Fibre Channel switch
3. Host I/O
4. Replication writes
5. Source array
6. Destination array
7. Source virtual disk
8. Destination virtual disk
9. DR group

DR Group Write History Log

The DR group write history log is a virtual disk that stores a DR group's host write data. The log is created when you create the DR group. Once the log is created, it cannot be moved. For more information, see "Choosing Replication Write Modes" in the Best Practices section

Managed Sets

Managed sets are a feature of HP Replication Solutions Manager. A managed set is a named collection of resources banded together for convenient management. Although managed sets can be used for a variety of different types of resources, in context of this project they were used only to aggregate DR groups. Performing an action on a managed set performs the action on all members of the set.

NOTE:

Managed sets are simply a feature that enables you to manage multiple resources easily. They do not contribute to the data consistency of a DR group. Write order consistency is maintained at the DR group level.

In managed sets:

- All resources, or members, in a single managed set must be of the same type (for example, all virtual disks).
- You can add a specific resource to more than one managed set.
- You can add resources on more than one array to a managed set.
- You should create separate managed sets for DR groups so that if a failover occurs, you can perform the actions that correspond to the changed source/destination role of the managed set members.

Failover

In HP Continuous Access EVA replication, failover reverses replication direction for a DR group. The destination array assumes the role of the source, and the source array assumes the role of the destination. For example, if a DR group on array A were replicating to array B, a failover would cause data for the DR group to be replicated from array B to array A. You can failover a single DR group or you can failover multiple DR groups with a single command using a managed set. When you specify a failover action for a specific managed set, the failover occurs for all DR groups contained in the specified managed set. Without managed sets, you must fail over each DR group individually. For more information on failover settings, see "Planning DR groups" under the Best Practices section

USE CASE OBJECTIVES

1. Show a test system architecture deploying key Oracle Fusion Middleware components and HP replication technologies
2. Utilizing HP Continuous Access EVA to protect all non-database content and Oracle Data Guard to protect all Oracle database content
3. Document procedures to handle planned (switchover and switchback) and unplanned outages (failover and failback)
4. Ensure a small RTO and RPO for the Oracle Fusion Middleware environment using HP Continuous Access EVA

TEST CASES

1. Make configuration changes at production site such as, creating a new managed server and deploying a new application and validating them at the standby site to ensure successful configuration replication between production and standby sites
2. Keep XA transactions pending at the production site and making sure they are committed or rolled back at the standby site after switchover/failover to ensure that transaction logs get replicated properly
3. Validate the Oracle Fusion Middleware components on production site, such as, BPEL, Worklist Application, Oracle Fusion Middleware Control, and then revalidate them on standby site after switchover/failover
4. Validate the Oracle Database replication using Oracle Data Guard in context of an Oracle Fusion Middleware SOA application

REFERENCE ARCHITECTURE

This diagram depicts the configuration used for certification of HP storage replication techniques for implementing Oracle Fusion Middleware DR solution. Each site consists of two web hosts, a WLS Admin host, and two clustered application servers with Oracle Fusion Middleware. All non-database objects including the Oracle binaries and Oracle Fusion Middleware configuration are protected using HP Continuous Access EVA. Also, the Oracle RAC database is protected with Oracle Data Guard.

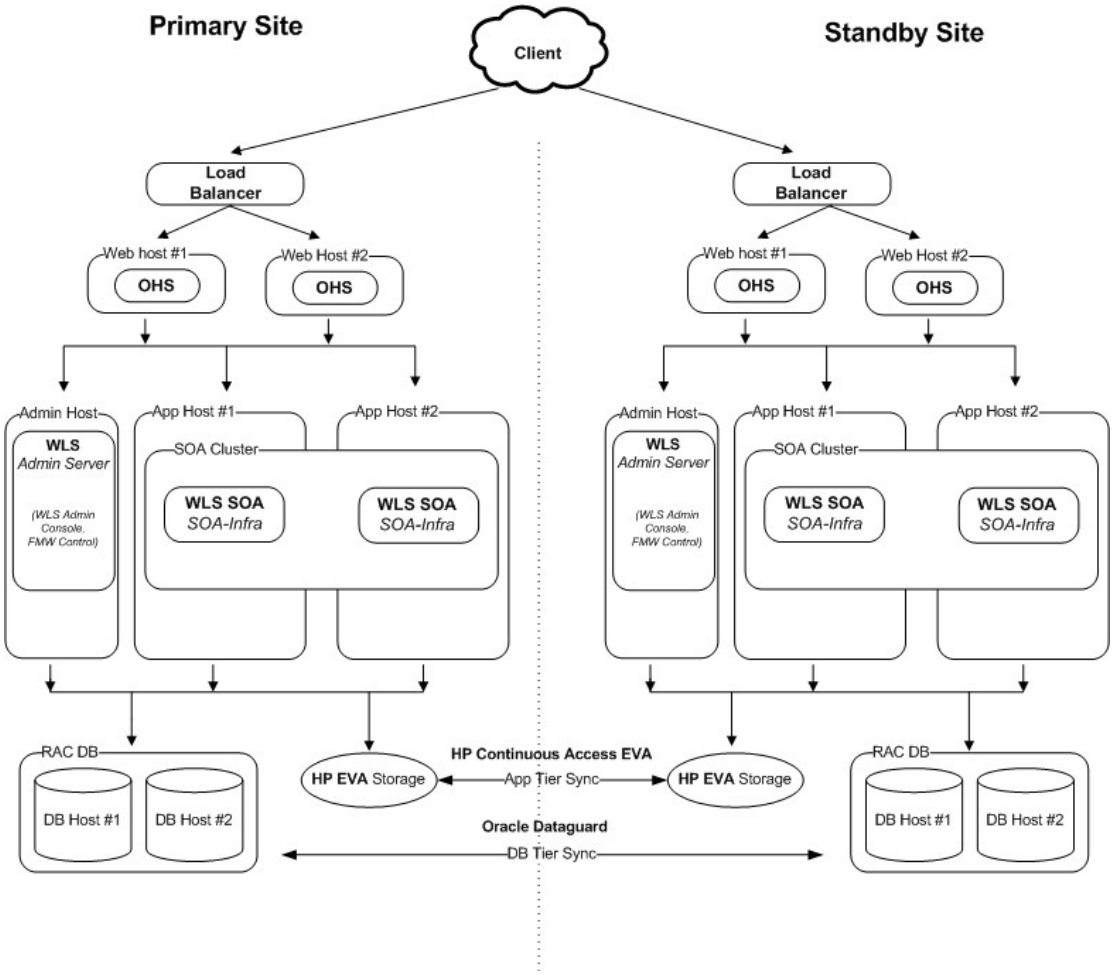


Figure 3 Reference Topology Diagram

SYSTEM CONFIGURATION

Hardware Diagram

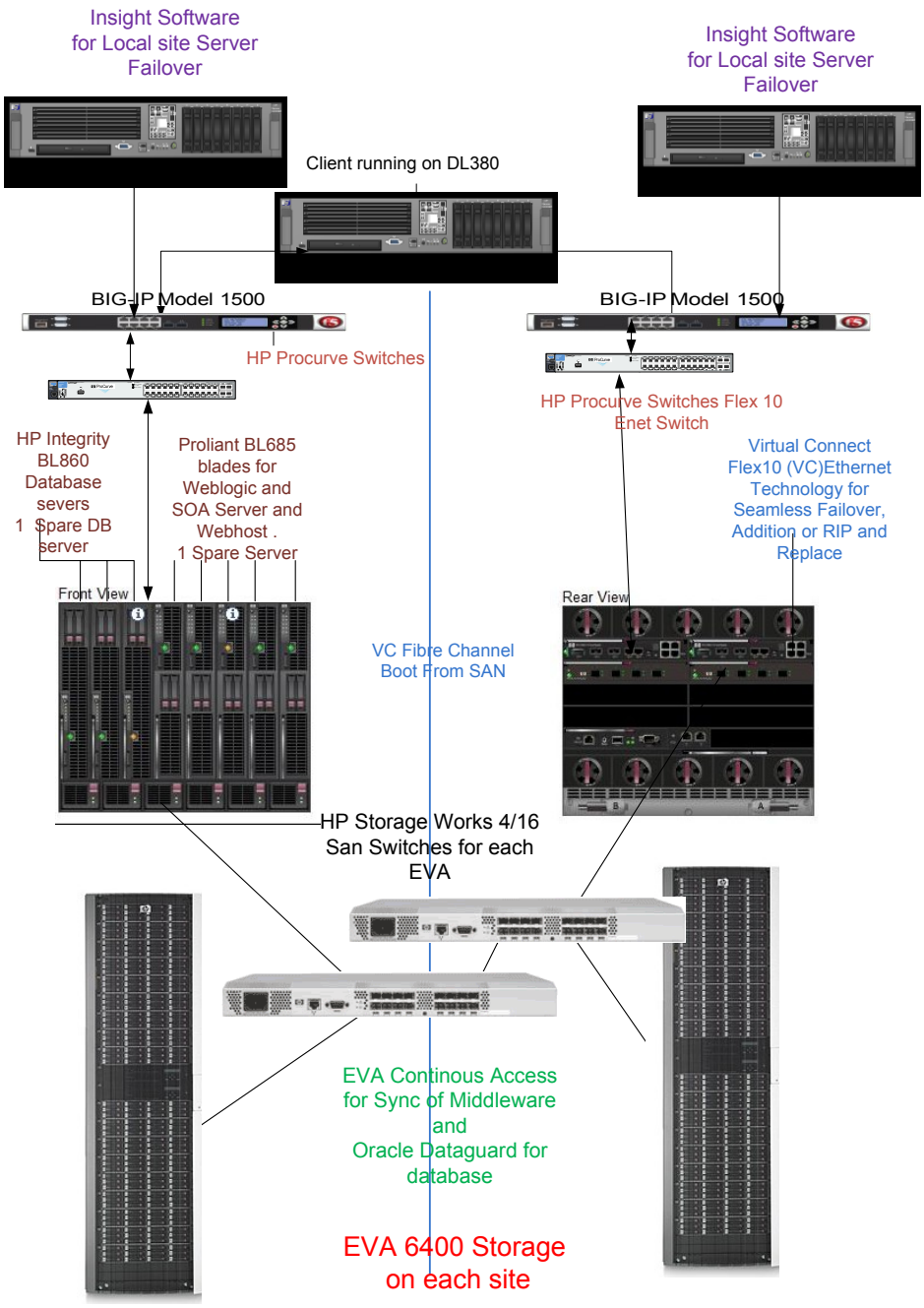


Figure 4 Hardware Diagram

Hardware and Operating System Specifications

- Client Host Configuration (Used for Validation Testing)

  HP Proliant DL380 (8 core, 2.8 GHz x86, 4 GB RAM)

  Windows 2003 Version 5.2

  Selenium Remote Control v1.0

- Load Balancer (1 per site)

  Site 1: BigIP Model 1500

  Site 2: BigIP Model 3400

  SW Version: BIG-IP 9.4.6 Build 401.0 Final

- Administration Server Host Configuration (1 per site)

  HP Proliant DL380 (8 core, 2.8 GHz x86, 4 GB RAM)

  Red Hat Enterprise Linux 5.3 (64 bit)

  Oracle Web Logic Server 10.1.3.1

- Web Host Configuration (2 per site)

  HP Proliant BL685c (8 core, 2.6 GHz x86, 8 GB RAM)

  Red Hat Enterprise Linux 5.3 (64 bit)

  Oracle HTTP Server (11.1.1.1)

- App Host Configuration (2 per site)

  Proliant BL685c (8 core, 2.6 GHz AMD, 16 GB RAM)

  Red Hat Enterprise Linux 5.3 (64 bit)

  Oracle Web Logic Server 10.1.3.1

  Oracle SOA Suite 11g (11.1.1.1)

- Database Server Host (2 per site)

  HP Integrity BL860 (4 core, 1.66 GHz Itanium 2, 64 GB RAM)

  HPUX 11iV3

  Oracle Database 11g RAC (11.1.0.7)

  HP Service guard for network redundancy.

- Storage Management Host

  Windows 2003 Version 5.2

  HP Command View EVA 9.0

HP Replication Solutions Manager 5.0.118

- Storage Array (1 per site)

  HP Storage Works EVA 6400

  HP 2 x HSV Controllers

  2.25 TB raw capacity

- HP StorageWorks SAN switch 4/16 (2 per site)
- HP Virtual Connect Flex10 Ethernet Switches (2 per site)
- HP VC FC SAN switches (2per site)
- HP Procurve Switches (2per site)

Planning DR Groups

The general rule for assigning virtual disks to DR groups is that virtual disks associated with the same application should be in the same DR groups. The term "application" really means a distinct software installation. In the Oracle Fusion Middleware environment, there are three distinct software installations, the Oracle HTTP Server installation on the Web Hosts, the Oracle Weblogic Server and Oracle SOA Suite 11g on managed server hosts. The Oracle Weblogic admin server is targeted to run on the admin server hosts.

A DR group consisting of one or more virtual disk has to be created for any replication to take place between two sites. In this case a DR group is created for each of the Oracle home's and additionally for TLog and JMS files to ensure all the middleware data is replicated. The following table shows the different directories that are part of DR group. For details on creating the required volumes refer to the Oracle Fusion Middleware Disaster Recovery Guide and Oracle Fusion Middleware Enterprise Deployment Guide for SOA Suite.

For details on configuring DR groups for other product suites besides Oracle SOA Suite, that are part of Oracle Fusion Middleware 11g, refer to the Oracle Fusion Middleware Disaster Recovery Guide. Below is the summary on how to configure DR Groups for Oracle Fusion Middleware 11g.

Oracle Fusion Middleware DR Groups			
DR Group Name	Host(s)	File systems	# of virtual disks
Admin	Administration Server	/oracle/home, /oracle/config	2

Web Host	WebHost1, WebHost2	/u01/app/oracle/product/fmw/web/ohs	2
Managed	AppHost1, AppHost2	/u01/app/oracle/product/fmw	2
JMS and Tlog Data	AppHost1, AppHost2	/u01/app/oracle/admin/soa_domain/SOA_Cluster/jms /u01/app/oracle/admin/soa_domain/SOA_Cluster/Tlogs	4 (2per host)

Mounting File Systems

Format and mount these partitions for jms and Tlog directories. For example on Apphost1:

mkfs –t ext3 /dev/sdb

mkfs –t ext3 /dev/sdc

/dev/sdc on /u01/app/oracle/admin/soa_domain/SOA_Cluster/jms type ext3 (rw)
/dev/sdb on /u01/app/oracle/admin/soa_domain/SOA_Cluster/tlogs type ext3 (rw)

The Apphost2 will have luns presented example sdd and sde and similarly the log files can be mounted.

Network Configuration

Because the sites in a disaster recovery scenario are generally separated from each other by a considerable distance, they are also likely to be on separate networks. This was the case in our test environment as well.

When a site switchover or site failover occurs, the hostnames and ip addresses associated with the primary site will continue to work on the secondary, because the DNS server on the secondary site has aliases for all of the hosts on the primary site. For example, the primary site might have a Web Host called webhost1_1 with an IP address of 10.10.11.101, while the secondary site has a Web Host called webhost1_2 with an IP address of 10.10.11.102. To enable the same name to be used on the secondary site, the DNS server on the secondary site would contain an alias (canonical name) for webhost1_1 associating it with webhost1_2. The BigIP load balancer would use the actual names and IP addresses of the hosts to do its load balancing, but the Oracle Fusion Middleware configuration would still be able to make use of the primary site names.

In our test environment, the client host can see both sites because the primary site's DNS server was its primary DNS server and the secondary site's DNS server

was its alternate DNS server. However like any client out on the web, the client software only knows how to reach the services provided by Oracle Fusion Middleware through a single URL:

http://bigiplb.mycompany.com:<port number>/console

This works because both the primary and secondary site DNS servers have aliases for bigiplb, even though the aliases point to different BigIP load balancers. When a failover occurs, the client host is able to detect that its primary DNS server was down, and begin to use it alternate DNS server (on the secondary) site, yet the client software is unaware that anything has changed.

Storage Software Configuration

Basic management of an HP Storage Works is accomplished through HP Storage Works Command View utility. This utility runs a Management Server host that is connected via Fiber Channel to the SAN associated with the EVA. The following picture shows typical EVA command view screen –

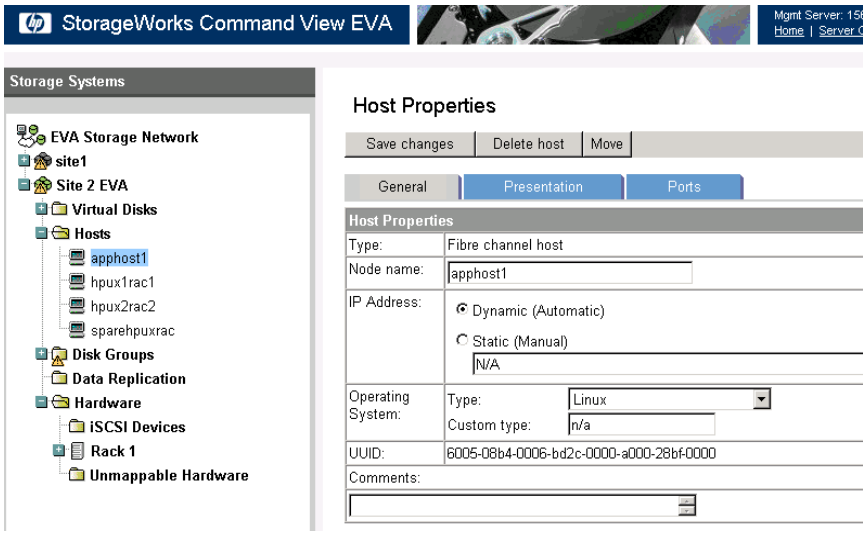


Figure 5 HP StorageWorks Command View EVA

Through the Command View GUI, we performed all of the basic array management tasks including:

- Installing Licenses

Ensure that the EVA licenses are updated, includes the EVA command view, Continuous Access. The command view license needs to be at least version 9.0 required for async failover.

- Upgrading Firmware (XCS)
  Ensure that the latest firmware is installed on the EVA controllers.
- Creating and Managing Disk Groups
- Defining Hosts
- Creating Virtual Disks
- Presenting Virtual Disks to Hosts
- Creating DR groups and
- Adding and Removing members from DR groups

HP also provides a special utility that is designed specifically for data replication, Replication Solutions Manager. This tool cannot perform the basic functions that Command View can such as installing licenses or upgrading firmware, but it does provide a much more robust set of capabilities for managing DR groups and other features associated with data replication including creating and failing over Managed Sets. You can use the replication manager to:

- Automatically discover array, virtual disk, host, and application resources
- Present virtual disks for host access
- Copy virtual disks and host volumes using snapshot, snapclone, and mirror clone technology
- Remotely replicate and fail over virtual disks
- Group and replicate resources as a unit called a managed set
- Dynamically mount virtual disks on enabled hosts
- Automate replication tasks using replication manager jobs and schedule jobs for future purposes.
- Monitor replication status by array : Verification of whether data is being replicated to other site can be monitored using the EVA command view status tool.
- Backup and restore replication manager configuration and jobs
- Replicate application resources on enabled hosts (for example, Oracle tablespaces)
- Visually manage replication resources with the topology viewer
- Configure security using operating system-based authentication and user administration with audit capabilities
- Integrate with a variety of backup and recovery solutions

- Create round robin snapshots and snap clones of host volumes using the host volume replication wizard. The oldest replica is deleted automatically, allowing the space it occupied to be reused.
- Create virtual disks, containers and DR groups
- Manually or automatically perform dynamic capacity management

RECOMMENDATIONS DURING NORMAL OPERATIONS

In a normal Oracle Fusion Middleware Disaster Recovery configuration, the following are true:

1. Disk replication is used to copy Oracle Fusion Middleware file systems and data from the production site shared storage to the standby site shared storage. During normal operation, the production site is active and the standby site is passive. When the production site is active, the only write operations made to the standby site shared storage are the disk replication operations from the production site shared storage to the standby site shared storage
2. Oracle Data Guard is used to copy database data for the production site Oracle databases to the standby databases at standby site. By default, the production site databases are active and the standby databases at the standby site are passive. The standby databases at the standby site are in managed recovery mode while the standby site is in the standby role (is passive). When the production site is active, the only write operations made to the standby databases are the database synchronization operations performed by Oracle Data Guard.
3. When the production site becomes unavailable, the standby site is enabled to take over the production role. If the current production site becomes unavailable unexpectedly, then a site failover operation is performed to enable the standby site to assume the production role. Or, if the current production site is taken down intentionally (for example, for planned maintenance), then a site switchover operation is performed to enable the standby site to assume the production role.

MANAGING PLANNED AND UNPLANNED DOWNTIME

Site Switchover Procedures

Switchovers are planned operations done for periodic validation or to perform planned maintenance on the current production site. During a switchover, the current standby site becomes the new production site, and the current production site becomes the new standby site. The following steps need to be performed in sequence –

1. Shutdown all the Oracle Fusion Middleware components on production site either manually or using relevant management software
2. If the replication write mode is not already set to synchronous, then change the write mode to synchronous and wait for any pending writes to complete
3. Unmount the file systems associated with the DR groups on production site
4. Switchover database using Oracle Data Guard to the standby site and ensure it comes up successfully
5. Failover the DR groups associated with Oracle middleware. The following screen shots demonstrate the failover process for one DR group; this process needs to be repeated for all the DR groups.

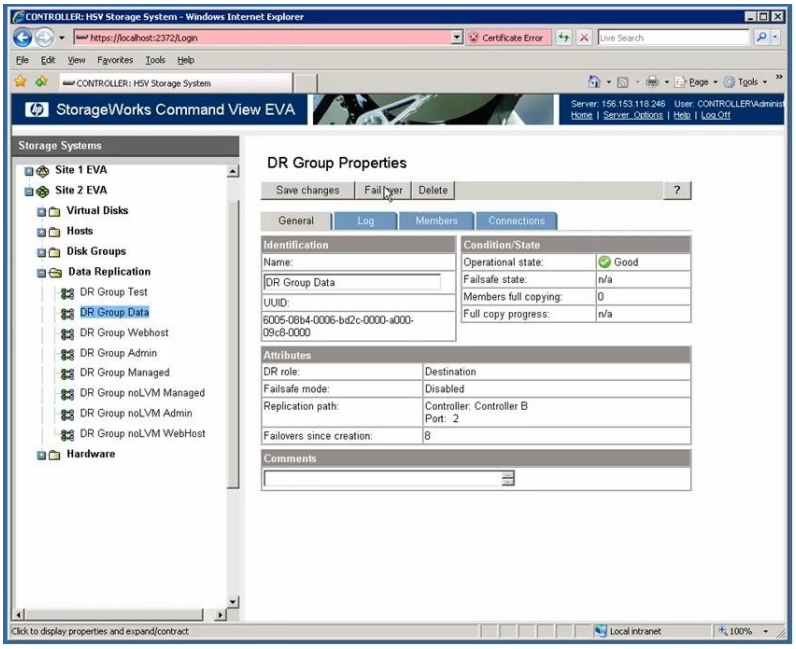


Figure 6 Select Failover for a DR Group

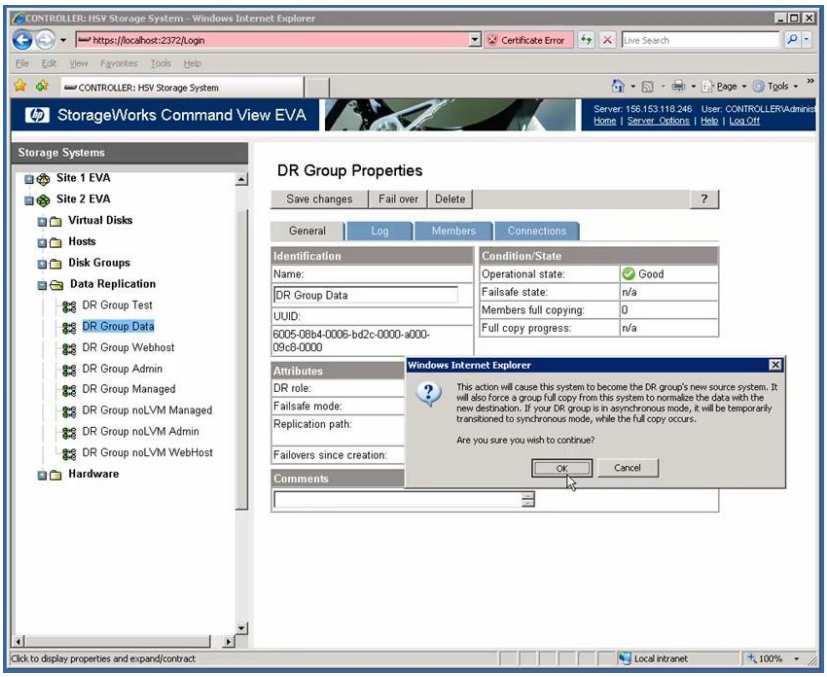


Figure 7 DR Group Failover Warning

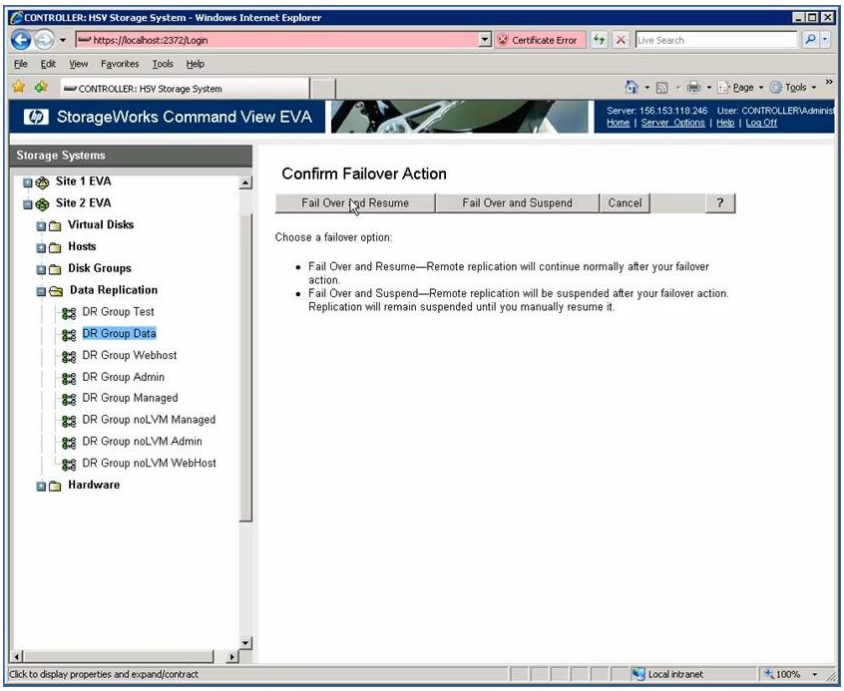


Figure 8 Confirm Failover for DR Group

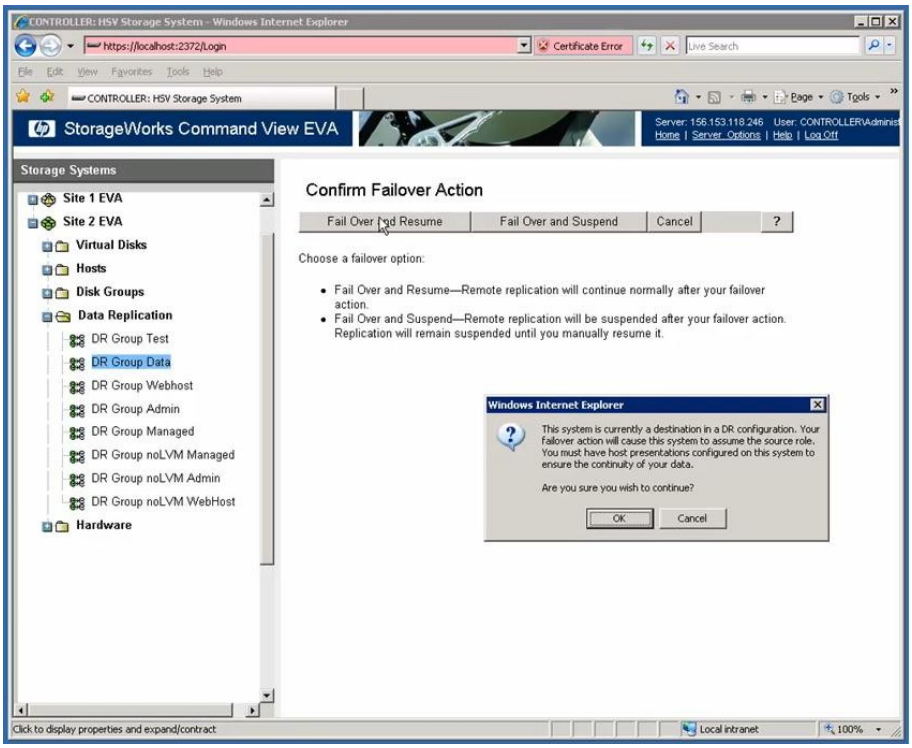


Figure 9 Confirm Failover Action for DR Group

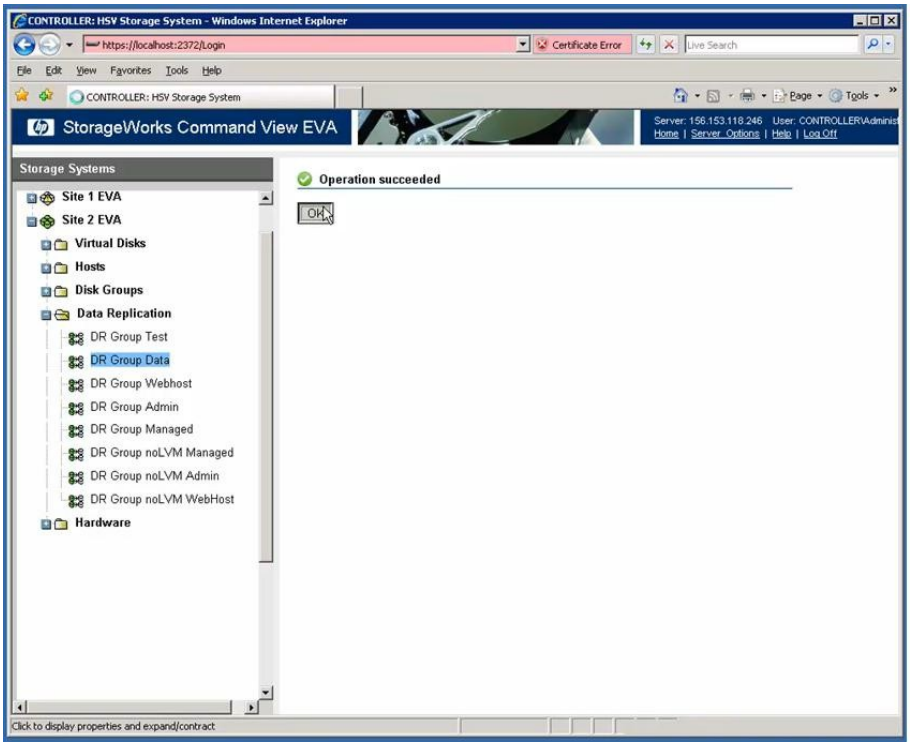


Figure 10 DR Group Failover Completion

6. Reconfigure the replication write mode to its original asynchronous setting, if any of the DR groups settings were modified in Step 2
7. Run device scan utility on each of the standby site's middle-tier hosts to ensure that failed over storage is now visible. This is process in which host runs a scan on all the devices presented to it. It may be required to be done on each server node to make sure all the LUNs are visible
8. Mount the file systems associated with the DR groups on standby site
9. Start all the Oracle Fusion Middleware components on standby site and ensure they are started successfully
10. Ensure that all user requests are routed to the standby (new production) site. This can be achieved through a global DNS push or something similar
11. The standby site has assumed the role of production site and vice versa
12. Use a browser client to perform post-switchover testing to confirm that requests are being resolved and redirected to the new production site

Switchback Procedures

Repeat all the steps in the switchover section to perform switchback to the original production site.

Site Failover Procedures

The process of making the current standby site the new production site after the production site becomes unexpectedly unavailable (for example, due to a disaster at the production site). The following steps need to be performed in sequence –

1. Detect that the production site no longer available
2. Determine the actual status of production site - has there been a real disaster or it something simpler like a network node failure? Can it be corrected without a site failover?
3. If it is determined that a site failover is required, failover the DR groups associated Oracle middleware to the standby site using HP Command View EVA as shown in 'Site Switchover Procedures' section
4. Run device scan utility on each of the standby site's middle-tier hosts to ensure that the failed over storage is now visible. This is process in which host runs a scan on all the devices presented to it. It may be required to be done on each server node to make sure all the LUNs are visible
5. Mount the file systems on standby site hosts

6. Failover the database using Oracle Data Guard and ensure it is started successfully
7. Start Oracle Fusion Middleware components on the standby site and ensure they are started successfully
8. Ensure that all user requests are routed to the standby site by performing a global DNS push
9. At this point, the standby site has assumed the role of production site
10. Use a browser client to perform post-failover testing to confirm that requests are being resolved and redirected to the new production site
11. When the original production site is back up, resynchronize the DR groups on the new production site (old standby site) with the DR groups on the original productions site
12. If the DR groups on original production site are not recoverable, then recreate them and then resynchronize the DR groups on the new production site (old standby site) with the DR groups on the original productions site

Failback Procedures

Failback involves re-instantiating both database and middle tiers.

BEST PRACTICES

Choosing Replication Write Modes

For Oracle Fusion Middleware Disaster Recovery, Enhanced Asynchronous and Synchronous replication write modes are recommended. The replication write mode can be set on specific DR group(s) depending upon frequency of data change. For example, Synchronous mode should be set for DR group pertaining to Tlogs due to its dynamic nature.

The base Asynchronous replication write mode is not recommended

Synchronous replication prevents any loss of data, however it also requires each write I/O to be completed on the destination array before it is considered completed for the local array. Therefore in an environment where there are a lot of write I/O's, synchronous replication is a potential drag on performance.

Fortunately, there is a solution that is nearly as robust as synchronous replication and that is enhanced asynchronous replication. In enhanced asynchronous replication, write I/Os don't have to be completed on the destination array before they are marked as completed locally.

At the same time, there is protection against data loss, because each write I/O is written both to the local array and to the DR group write history log before it is considered to be complete. The write history log is written in the same order that the write I/O's are written to the local array. As the write I/O's are propagated to

the destination array, they are removed from the write history log, so the write history log is a sequential record of all write I/Os written to the local array that have not yet been acknowledged to be completed on the destination array.

In the event of a failure while enhanced asynchronous write mode is being used, all pending write I/O's are preserved in the write history log. In this scenario, one option is to simply wait until the source array can be brought back up. If the failure is only temporary and can be corrected in a short period of time, this is probably the best option, because it ensures that no data will be lost.

In the case where the failure is not temporary or the production environment needs to be brought back online quickly, the customer will have to fail over the production site to the standby site. In enhanced asynchronous write mode, this means that all pending write I/O's in the write history log will be lost. The number of writes lost can be minimized if the writes are being processed quickly, and the therefore the number of pending writes is low. The rate of write processing should be estimated by customers when they are setting their RPO. The RPO is dependent on the bandwidth of the inter site link, which is in turn dependent on the distance between the arrays, the type of interconnect, and other factors. Careful analysis of the application's write profile and the replication link speed can determine what the worst case RPO will be for the solution. For complete details on RPO's, bandwidth, and inter site links, see the HP StorageWorks Continuous Access EVA implementation guide (see References section).

While it is possible that a failover using enhanced asynchronous write mode could result in zero data loss if the write log is empty, enhanced asynchronous replication is can never be guaranteed to achieve that objective. Synchronous replication is the only way to guarantee zero data loss. Oracle Fusion Middleware environment does not generally require synchronous replication, the Java Message Services (JMS) and Transaction Logs (TLogs) files can be put on enhance asynchronous replication so that the performance is not affected. If these logs are to be replicated synchronously when the rest of environment is using enhanced asynchronous replication, they must be in a separate file system and a different DR group.

The third write mode option, asynchronous without the write history log, is not recommended. It is the only asynchronous option available for older versions of firmware (pre XCS 6.xxx versions). To use enhanced asynchronous mode, both local and remote arrays must be running XCS 6.xxx firmware or greater. For arrays that are not capable for running XCS 6.xxx or greater, we would recommend either upgrading to new storage arrays or running in synchronous mode.

In event of failure in the middle of a large block transfer, the HP Continuous Access EVA synchronous replication guarantees the IO write order during replication to other site. So this becomes a case of how Oracle handles a local site

failure due to power or any other reason when a database goes down. The synchronous replication also ensures that any message is written to the disk is written on both sites. If the replication is enhanced asynchronous the replication is again ensured to write IO order protected due to the history log.

In event of failure in the middle of a large block transfer, the HP Continuous Access EVA synchronous replication, guarantees the IO write order during replication to the other site. The synchronous replication also ensures that any message is written to the disk is written on both sites. If the replication is enhanced asynchronous, the replication is again ensured to write IO order protected due to the history log.

In addition to the HP Continuous Access EVA replication, it is recommended that the user maintain a daily backup of the sites using business copy on each site. Additional features like snap clone takes point in time physical copy of the virtual disk, whereas the Vsnap features takes a snapshot of the virtual disk. The HP Continuous Access EVA is a highly reliable and highly available configuration. The user needs to select the proper procedure to complement the HP Continuous Access EVA replication with other features of backup based on the risk profile of the business needs.

Choosing the Size of Your Write History Log

As noted, enhanced asynchronous replication is the preferred write mode for most parts of the Oracle Fusion Middleware environment. For enhanced asynchronous write mode to work properly, the write history log must be large enough to hold the entire write I/Os for a system that is under peak load. This is extremely important, since a full write log results in a process called normalization, which will force a synchronization of the source and destination arrays. Under peak load, a forced normalization would have a very negative impact on performance. Another reason to set the size of the write history log correctly from the point when the DR group is created is that changing the size requires you to switch to synchronous, drain the write log, and then switch back to enhanced asynchronous mode.

HP Boot From SAN

Recommend addition to a Disaster Recovery Solution

Traditionally, HP Integrity servers have booted their operating systems (OSs) from local disks or Direct-Attached Storage (DAS). Today's rapidly evolving technologies now enable servers to boot from a Redundant Array of Independent Disks (RAID) unit located on a Storage Area Network. An HP StorageWorks EVA storage array offers support for this server boot model.

Boot from SAN capability allows any server on the network to boot from a Fiber Channel (FC) RAID unit located somewhere on the SAN. As a result, when a server needs to be taken offline or replaced, the system administrator need only deploy a new server with the same configuration, direct the storage volume to this server, and boot the server from the RAID unit. Booting in this way from an external device can decrease downtime in the event of server failure as there is no

need to re-install the OS and application software to make the replacement server functional.

Separating the boot image from the server helps fully leverage an organization's investment in high availability, data integrity and storage management.

In this setup, each of the OS luns were carved out on the EVA 6400 storage system. The hpux 11iv3 and Red Hat Linux 5.3 were installed. This was done for both primary and standby sites for each of the servers. This provides additional local failover capability within each setup using logical server migration and as well as ease of management.

High Availability for Persistent Stores

The Oracle WebLogic application servers are usually clustered for high-availability. For high-availability of the Oracle SOA Suite within a site, a persistent file-based store is used for the Java Message Services (JMS) and Transaction Logs (TLogs). This file store needs to reside on shared disk, which is accessible by all members of the cluster.

Data loss and latency requirements

The Recovery Point Objective (RPO) and Recovery Time Objective (RTO) define the time taken to recovery and the point until which data is recovered. Failover distances at metro level (typically less than 150 miles) can be on synchronous replication. There is minimal data loss of less than 5 minutes, and recovery time within a few hours (2-5) to a few days depending on the criticality of the application and SLA requirements.

For long distance recovery (greater than 150miles) when it is required to protect from major area wide disasters, asynchronous replication is used. The recovery time can be anywhere between 2 hours to 20 days depending on SLA requirements and data loss can be as low as 10 minutes to up to 2 days depending on the criticality and SLA. Careful analysis of SLA and appropriate procedures for disaster recovery need to be adopted to arrive at suitable solution for the business thereof.

HP Virtual Connect for HP BladeSystem Infrastructure Virtualization

HP Virtual Connect Flex 10 (VC) is an industry standards-based implementation of server-edge input-output (I/O) virtualization. It puts an abstraction layer between servers and the external networks so that the Local Area Network (LAN) and SAN see a fixed pool of servers rather than individual servers.

The ability to support Virtual Connect is built in to each HP BladeSystem c-Class component, including the Onboard Administrator (OA), FC Host Bus Adapters (HBAs), Network Interface Controllers (NICs), and the HP Integrated Lights-Out (iLO) communication channels. VC modules work with the standard Ethernet NICs and FC HBAs available for HP BladeSystem c-Class server blades; no other specialized mezzanine cards are required.

VC Ethernet and FC modules simplify the connection of server NICs and HBAs to the data center environment and extend the capabilities of these standard server devices by supporting the secure administration of their Ethernet MAC addresses

and FC WWNs. The effect of the VC modules make it appear that there are no virtual devices but that all the bay assigned addresses are the real and only identifiers seen by the system, the OS, and the network.

The Virtual Connect Manager allows you to create I/O profiles, defining unique Media Access Control (MAC) addresses and World Wide Names (WWN). A VC profile is assigned to a specific bay in the blade enclosure, thus the addresses assigned to the bay override the factory defaults of the LAN and FC hardware on the blade server installed in the bay. The bay assigned MAC addresses and WWNs are the only addresses seen by the upper level operating system and other external references.

This approach allows you to upgrade or replace a blade server without the need to change any of the external resources used by that server. This approach allows you to upgrade or replace a blade server without the need to change any of the external resources used by that server.

Application and Database landscapes commonly need dense stacks of servers with hundreds of Ethernet and Fiber Channel (FC) connections to make the infrastructure work. Virtual Connect eliminates point to point connections and configuration complexity.

Local Server Failover for HP Blade Infrastructure using HP Insight Software

To have a robust and complete solution for any architecture, the configuration must account for individual server failure. The user must be able to quickly and easily restore balance when one of the servers fail. This can be accomplished using the HP Logical server migration feature of the Insight Software.

HP Insight Software is a new class of management software that allows you to continuously analyze and optimize your Converged Infrastructure. It's the world's first integrated solution that lets you visualize, plan and change physical and virtual resources in exactly the same way. It combines the best of the industry-leading HP infrastructure management portfolio – including HP Systems Insight Manager, HP Insight Control and HP Virtual Server Environment – into one integrated offering for HP BladeSystem servers. It makes your infrastructure change-ready, with the freedom and flexibility of virtualization delivered across your physical infrastructure.

Logical servers bring the freedom and flexibility of virtualization to physical servers. The logical server is a server profile that is easily created and freely moved across physical and virtual machines. By detaching the logical identity from the physical resource, you can create or move logical servers on any suitable virtual or physical machine—on demand. For example, a logical server profile would include entitlements such as power allocation, processor and memory requirements, network connections, and storage—everything that the OS and application stack requires to operate.

With a logical server approach, you can even create templates for your frequently used applications with specific configurations. These templates can be stored and reactivated in minutes, when needed. The capability to move server profiles across physical servers is available on HP Proliant and HP Integrity blades. The insight software is resident on a HP Blade Server.

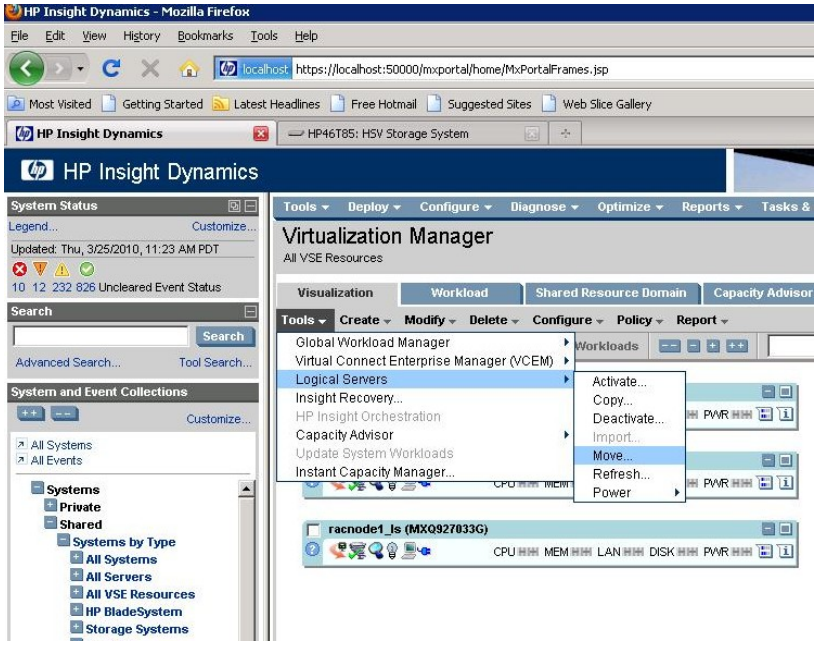
Since the OS is resident on the SAN or EVA on a virtual disk or LUN as indicated in the boot from SAN section. When one of the blade server fails, all we need is to boot up another blade server and point it to the earlier LUN or virtual disk.

Virtual connect switches enable this logical server migration by virtualizing the MAC and WW ID. The OS resides on the virtual disk on EVA. In event of local blade server failure, another blade server can take the identity of the failed server and boot up with the LUN on the EVA. In the same way the blade server characteristics known as logical server can be migrated to another blade.

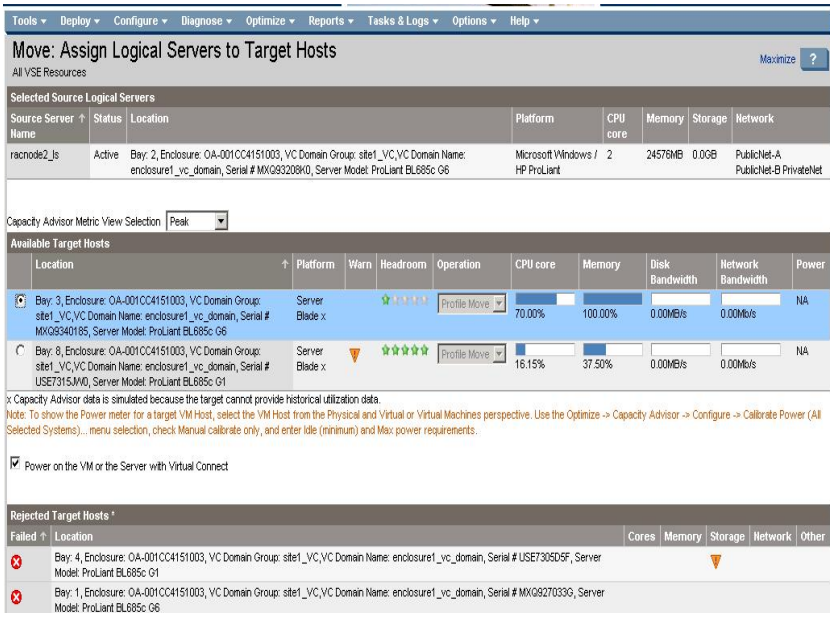
A logical server profile would include entitlements such as power allocation, processor and memory requirements, network connections, and storage connection and everything that the OS and application stack requires to operate.

The following figures illustrate how to quickly move a profile from one blade server to another in an event of local failure and ease of use using the HP insight manager.

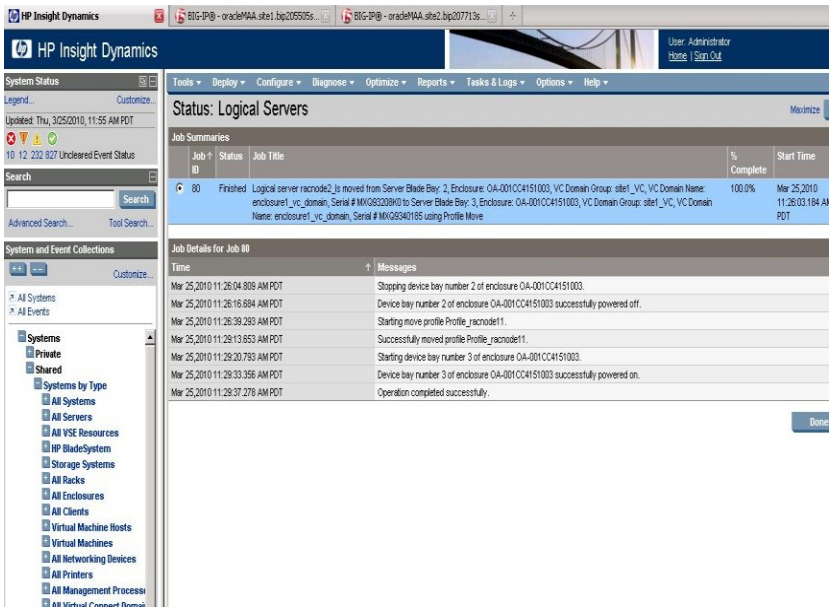
1. Select the logical server profile which needs to move



2. Once a logical server to move is picked, the next screen as shown below lists all the potential target servers which have the same cpu, memory and other features required for suitable logical server. Select the appropriate slot or bay of the enclosure where you want to move the profile to.



The job to do the logical server migration runs and the task is completed.



Oracle Fusion Middleware High Availability Technologies

Some of the high availability features provided by the Oracle Fusion Middleware infrastructure are as follows (Refer to the Oracle Fusion Middleware High Availability Guide for additional details):

- Process death detection and automatic restart

  For Java EE components running on WebLogic Server, Node Manager monitors the Managed Servers. If a Managed Server goes down, it attempts to restart the Managed Server for a configured number of times.

  For system components, OPMN monitors the processes. If a system component process goes down, OPMN attempts to restart the process for a configurable number of times.

- Clustering

  Oracle Fusion Middleware Java EE components leverage underlying powerful WebLogic Server clustering capabilities to provide clustering. Oracle Fusion Middleware uses WebLogic clustering capabilities, such as redundancy, failover, session state replication, cluster-wide JNDI services, Whole Server Migration, and cluster wide configuration. These capabilities provide for seamless failover of all Java EE Oracle Fusion Middleware system components transparent to the client preserving session and transaction data, as well as ensuring data consistency. System components can also be deployed in a run time cluster. They are typically front-ended by a load balancer to route traffic.

- State replication and routing

  Oracle WebLogic Server can be configured for replicating the state of stateful applications. It does so by maintaining a replica of the state information on a different Managed Server, which is a cluster member. Oracle Fusion Middleware components, which are stateful, leverage this feature to ensure seamless failover to other members of the cluster.

- Failover

  Typically, a Managed Server running Oracle Fusion Middleware Java EE components has a Web server, such as Oracle HTTP Server, clustered in front of it. The Web server proxy plug-in (mod_wl_ohs) is aware of the run time availability of the different Managed Servers, as well as the location of the Managed Server on which the state replica is maintained. If the primary Managed Server becomes unavailable, the plug-in routes the request to the server where the application is available. For stateful applications, the location of the replica is also taken into account while routing to the new Managed Server.

- Server Migration

  Oracle Fusion Middleware components, such as SOA, which uses pinned services, such as JMS and JTA, leverage WebLogic Server capabilities to provide failover and automatic restart on a different cluster member.

- Integrated High Availability

  Oracle Fusion Middleware has a comprehensive feature set around load balancing and failover to leverage availability and scalability of Oracle

RAC databases. All Oracle Fusion Middleware components have built-in protection against loss of service, data or transactions as a result of Oracle RAC instance unavailability due to planned or unplanned downtime. This is achieved by using Oracle WebLogic Server multi data sources. Additionally, components have proper exception handling and configurable retry logic for seamless failover of in-flight transactions at the time of failure.

- Rolling Patching

  Oracle WebLogic Server allows for rolling patching where a minor maintenance patch can be applied to the product binaries in a rolling fashion without having to shut down the entire cluster.

- Backup and Recovery

  Oracle Fusion Middleware backup and recovery is a simple solution based on file system copy for Middle-tier components. RMAN is used for Oracle databases. There is also support for online backups. With Oracle Fusion Middleware, you can integrate with existing backup and recovery tools, or use scheduled backup tasks through oracle Fusion Middleware Enterprise Manager or cron jobs.

CONCLUSION

Storage replication is a key requirement in providing disaster recovery protection for Oracle Fusion Middleware environments. HP Continuous Access EVA provides comprehensive features to handle all the unique requirements for replicating Oracle Fusion Middleware components in conjunction with Oracle Data Guard for Oracle database replication. HP Continuous Access EVA provides different modes of replication techniques that could be used depending on product requirements for making sure that customer environments are protected against any unforeseen disasters.

Using HP Continuous Access EVA with Oracle Data Guard provides users with the maximum benefit out of their investment to protect their entire Oracle environment.

REFERENCES

Oracle's Middleware Disaster Recovery Guide and Disaster Recovery Terminology

HP StorageWorks Command View EVA user guide

HP StorageWorks Continuous Access EVA implementation guide

HP StorageWorks Continuous Access EVA Software – Overview &Features

Configure DR Solution using HP EVA Continuous Access

HP StorageWorks Enterprise Virtual Array compatibility reference

HP StorageWorks Enterprise Virtual Array configuration best practices white paper

HP StorageWorks Replication Solutions Manager help and user guide

HP licenses installation

HP Reference Architectures for Application Deployment in Virtualized Environments

http://h71028.www7.hp.com/enterprise/us/en/technologies/virtualization-vsereferencearchitectures-whitepapers.html

HP BladeSystem http://h18000.www1.hp.com/products/blades/bladesystem/

HP BladeSystem Virtual Connect http://h18006.www1.hp.com/products/blades/components/ethernet/vc/index.html

Oracle Fusion Middleware 11g Disaster Recovery Solution Using HP EVA Storage
July, 2010
Author: Prasad Kona from Oracle, Sathya Krishnaswamy from HP
Contributors: Kathir Radhakrishnan, Milton Wan, Aalok Muley, Sunita Sharma, Anuj Sahni from Oracle
Stan Kellet, Jeff Smith from HP

Oracle USA, Inc.
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com