Support of Oracle RAC ASM with SGeRAC

Jan 2008

Introduction

This document discusses the use of the Oracle 10g Release 2 (10g R2) and11g Release 1 (11g R1) database server feature called Automatic Storage Management (ASM) in configurations of HP Serviceguard Extension for Real Application Clusters (SGeRAC).

We begin with a brief review of ASM - its functionality, pros and cons, and method of operation. Then we look in detail at how we configure ASM with SGeRAC (version A.11.17 or later is required).

ASM Background

What is Automatic Storage Management (ASM)?

ASM was introduced by Oracle in 10g as a component of the Oracle database and provides an alternative to platform file systems and volume managers for the management of most file types used to store the Oracle database, including Oracle datafiles, control files, and online and archived redo log files. (The full list of types of files supported by ASM can be found in the section *What Types of Files Does ASM Support?* in Chapter 12, *Using Automatic Storage Management of Oracle® Database Administrator's Guide 10g R2*[1] *or* in the section *About ASM Files Chapter 1, Introduction to Automatic Storage Management (ASM) of Storage Administrator's Guide*[2].) File types not supported by ASM include Oracle database server binaries, trace files, audit files, alert logs, backup files, export files, tar files, core files, as well as the quorum and registry devices used by Oracle Clusterware. Storage for application binaries and data cannot be managed by ASM.

For the supported file types, ASM provides file management services and takes care of mirroring and striping these files. ASM ensures I/O distribution over available disks through its striping capability and it supports online disk addition and deletion.

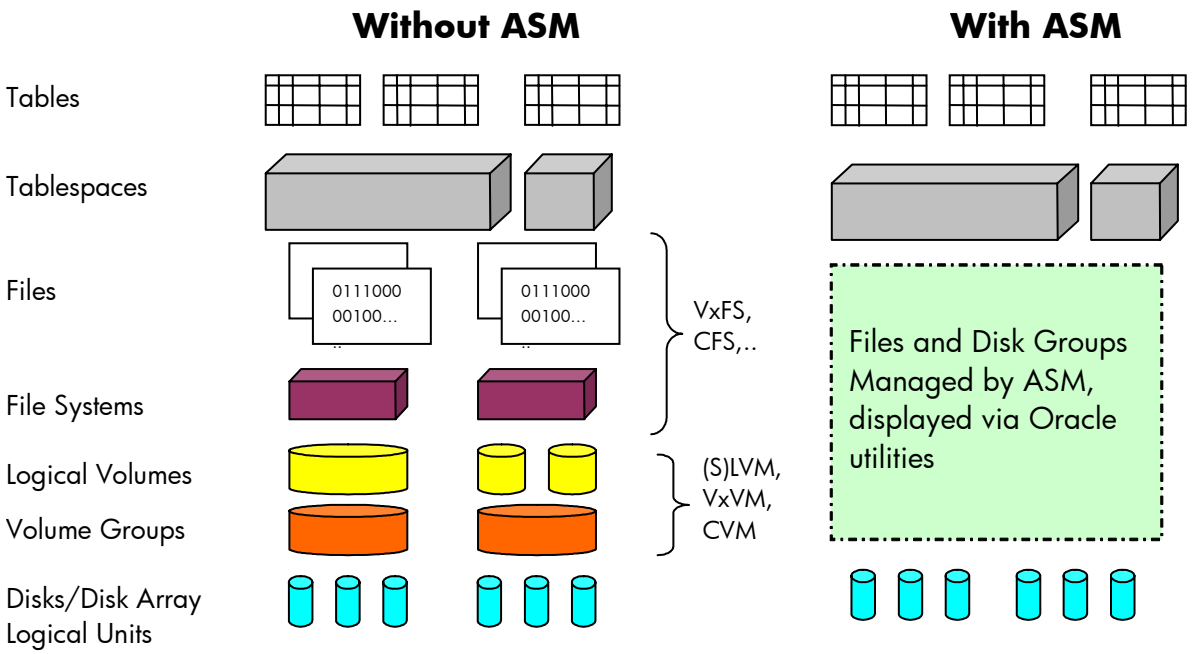
A major advantage of ASM is the ease of management it provides for database files:

- The system administrator has only to specify the set of raw devices to be used in an ASM disk group; the tasks of configuring and administering volume/disk groups and file systems are eliminated. Oracle ASM makes use of the Oracle feature called Oracle-managed files[1] and performs the tasks of creating, deleting and extending files on behalf of database instances; additionally, it manages their mirroring and striping.
- If a device is added to, or deleted from, an ASM disk group, ASM automatically rebalances database file striping based on the new disk layout.

Figure 1 contrasts the Oracle storage hierarchy as it appears when platform or 3rd party volume managers and file systems are used for Oracle data files, compared to when ASM is used. The layers corresponding to file systems and volume managers are absorbed into ASM. The files and directories in the storage hierarchy are not visible using standard operating system commands; to display them the customer must use Oracle utilities.

[1] See http://www.oracle.com/pls/db102/portal.portal_db?selected=4) → Guides → Administrator's Guide
[2] See http://www.oracle.com/pls/db111/portal.portal_db?selected=4&frame= → Storage → Storage Administrator's Guide

Figure 1. Oracle database storage hierarchy without and with ASM



Please note that the 'with ASM' stack shows raw disks/disk array logical units (LUs) being used by ASM. This is consistent with the Oracle perspective of ASM playing the role of the volume manager and file system in managing Oracle data.

Single Node (Non-Clustered) ASM Environment

ASM on HP-UX is a user-space storage manager (although ASM defines a kernel library interface that can provide extended ASM functionality, this library is not provided on HP-UX).

Figure 2 shows the ASM environment within a single non-clustered node. Typically there is only one ASM instance per node (irrespective of whether it is part of a high-availability cluster or not). It provides services to potentially multiple Oracle database instances running on that node and configured to use ASM-managed storage.

The ASM instance is responsible for

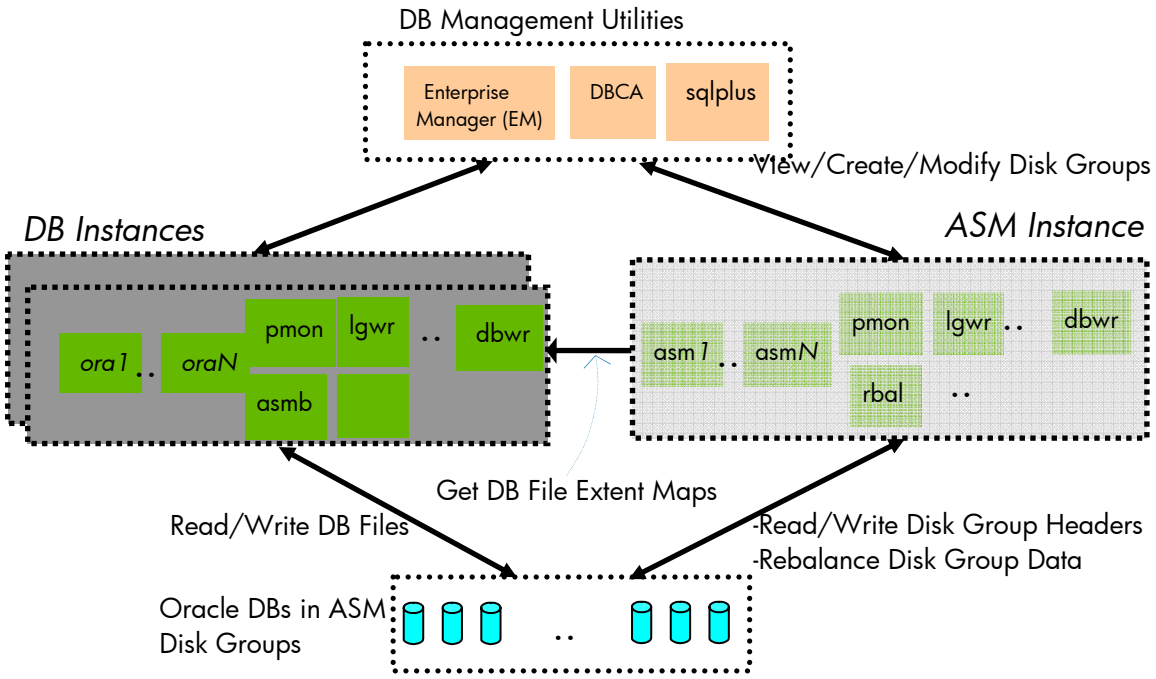
- creating/modifying/destroying ASM disk groups and the database files contained in them,
- managing the mirroring and striping of db files in ASM disk groups, including rebalancing as needed, and
- providing database file layout information to client database instances.

In structure, the ASM instance resembles a database instance.

- It has a set of foreground (asm1.. asmN) and background processes accessing a shared memory segment. The background processes include processes that resemble those of the database instance such as pmon, dbwr, .. as well as some additional ASM-specific processes such as rbal to perform disk group rebalancing.
- It uses its foreground processes to interact with clients such as database instances and db utilities such as sqlplus.

For a database using ASM-managed storage, the database instance has, in addition to the standard foreground (ora1,.. oraN) and background processes (pmon, dbwr, ..), some ASM specific processes such as rbal and asmb for communication with the ASM instance.

Figure 2. Single Node ASM Environment



When a database instance starts up, it opens the database files, and in doing so, the instance acquires file extent maps from the ASM instance for each file. These maps are used by the database instance to translate I/O file addresses to device block addresses.

To provide the extent maps, the ASM instance scans the ASM metadata on the devices that it manages. When the metadata changes, for example due to device addition or deletion, the ASM instance updates the database instance with the new information.

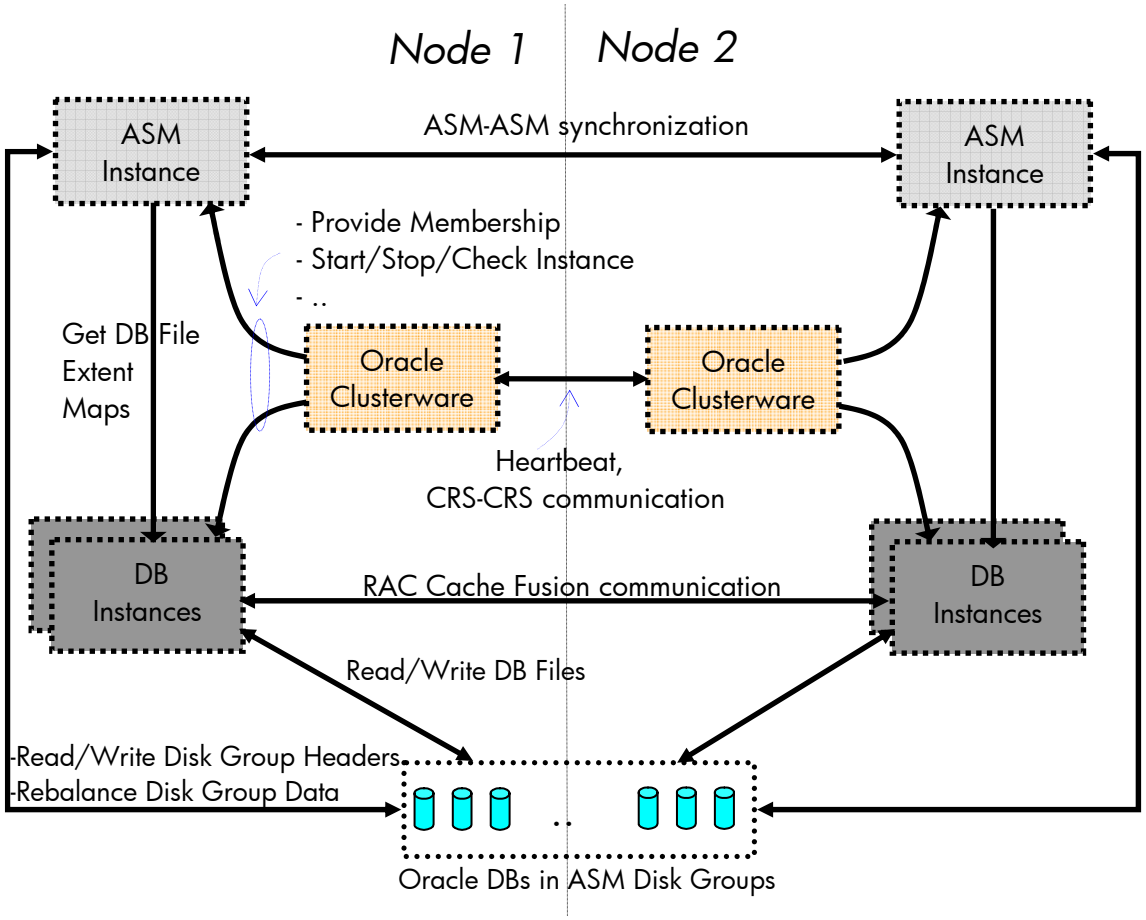
Thus the database instance does not have to communicate with the ASM instance for every I/O read or write; it uses the copy of file extent maps that it has cached to generate one or more corresponding I/O requests directly to the operating system.

Clustered ASM Environment

In an Oracle Clusterware environment, the individual ASM instances, one per node, cooperate to offer shared disk group access across the cluster. One or more RAC databases may be configured in the cluster, using the shared disk groups. Single instance databases may also be configured to use the shared disk groups but Oracle Clusterware does not provide built-in management and monitoring for them.

The ASM instances are under the control of Oracle Clusterware. Oracle Clusterware manages the start, stop and health check functions for the ASM instances just as it does for RAC database instances. Oracle Clusterware also ensures that the RAC database instances and the ASM instance start and shutdown in the proper order (that is, the ASM instance starts up before the database instances and shuts down after the database instances). Figure 3 shows the clustered ASM environment.

Figure 3. RAC ASM Environment



The ASM instances communicate with each other to maintain the consistency of the metadata of the disk groups they are managing.

Oracle Clusterware maintains a resource dependency on the ASM instance running on a node from every RAC database instance on that node that uses ASM services; this ensures that the RAC database instances using the services of a failed ASM instance will be halted. The other ASM instances in the cluster are responsible to address the issue of operations only partially completed by the failed ASM instance and to preserve the integrity of the ASM disk group metadata.

SG/SGeRAC Support for ASM on HP-UX 11i v2

Overview

Support for ASM with SG/SGeRAC commences with SGeRAC A.11.17.
However, please note that it has been possible to set up Oracle databases using ASM on HP-UX from the time Oracle 10g became available on HP-UX, using the following configurations that do not require SG/SGeRAC:

- non-clustered single instance Oracle
- Oracle single instance and RAC databases running in a pure Oracle clusterware environment

The following requirements/restrictions apply to SG/SGeRAC (A.11.17 or later) support of ASM (and are summarized in Table 1):

Table 1. ASM support with SG/SGeRAC A.11.17 or later

SUPPORTED	**NOT SUPPORTED**
Oracle RAC database	Oracle single instance database, with or without Oracle clusterware
SLVM Logical Volumes as ASM Disk Group Members	Raw disks/LUs or CVM Logical Volumes as ASM Disk Group Members
ASM configurations on SGeRAC using SGeRAC A.11.17 or later, and Oracle 10g R2 RAC	ASM configurations on SGeRAC using SGeRAC 11.16, Oracle 10g R1 RAC or HP-UX 11i v1
ASM configurations on SGeRAC using SGeRAC A.11.17 or later, and Oracle 11g R1	Single instance Oracle database with ASM managed storage running in Serviceguard cluster.
Maximum SGeRAC cluster size for ASM configurations is 16	Single instance Oracle database with ASM-managed storage in SGeRAC configuration wherein SGeRAC provides node membership to Oracle Clusterware.
SGeRAC Toolkit[3] framework based on Multi-Node Package (MNP) /Simple Package Dependency to manage ASM configurations on SGeRAC up to a maximum cluster	

[3] See the whitepaper "*Use of Serviceguard Extension for RAC Toolkit*" at http://www.docs.hp.com

size of 16	

Why ASM over SLVM?

As mentioned above, we require ASM disk group members in SGeRAC A.11.17 (or later) configurations to be raw logical volumes managed by SLVM.

The main reason for this requirement is to expedite ASM support with SGeRAC. We leverage existing HP-UX capabilities to provide multipathing for SLVM logical volumes, using either the PV Links feature, or separate products such as HP StorageWorks Secure Path that provide multipathing for specific types of disk arrays. Other advantages of the "ASM-over-SLVM" configuration are as follows:

- ASM-over-SLVM ensures that the HP-UX devices used for disk group members will have the same names (the names of logical volumes in SLVM volume groups) on all nodes, easing ASM configuration.
- ASM-over-SLVM protects ASM data against inadvertent overwrites from nodes inside/outside the cluster. If the ASM disk group members are raw disks, there is no protection currently preventing these disks from being incorporated into LVM or VxVM volume/disk groups.

The disadvantages of the ASM-over-SLVM configuration are as follows:

- Additional configuration and management tasks are imposed by the extra layer of volume management (administration of volume groups, logical volumes, physical volumes)
- There is a small performance impact from the extra layer of volume management
- SLVM has some restrictions in the area of online reconfiguration, the impact of which will be examined later in this document

Configuring SLVM Volume Groups for ASM Disk Groups

Ideally, ASM disk group members should be raw disks or array logical units, but due to the lack of built-in multipathing we require them to be SLVM raw logical volumes in SGeRAC configurations. But we would still like these logical volumes presented to ASM to resemble raw disks, as far as possible.

Hence, each SLVM logical volume (LV) used as a member of an ASM disk group is required to be laid out to occupy the usable space, in contiguous fashion, of exactly one single physical volume (PV). This implies that the LV

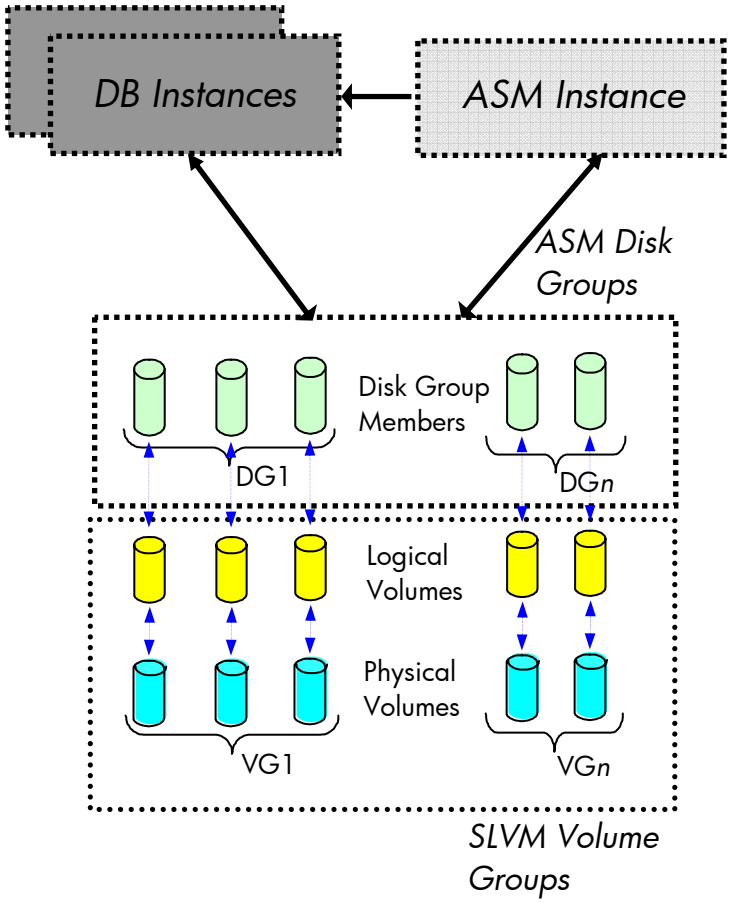
- should be contiguous,
- should not be striped or mirrored,
- should not span multiple PVs
- and should not share a PV with other LVs.

The idea is that ASM provides the mirroring, striping, slicing and dicing functionality as needed and SLVM supplies the multipathing functionality not provided by ASM. Figure 4 indicates this 1-1 mapping between SLVM PVs and LVs used as ASM disk group members.

Further, the default retry behavior of SLVM could result in an I/O operation on an SLVM LV taking an indefinitely long period of time. This behavior could impede ASM retry and rebalance capabilities; hence a finite timeout must be configured for each SLVM LV. For example, the timeout could be configured to the value (total number of physical paths to the PV * PV timeout), providing enough time for SLVM to try all available paths, if needed.[4]

The PVs used in an ASM disk group can be organized into SLVM volume groups as desired by the customer. In the example shown in Figure 4, for each ASM disk group, the PVs corresponding to its members are organized into a separate SLVM volume group.

Figure 4. 1-1 mapping between SLVM logical and physical volumes for ASM configuration



[4] If the LVM patch PHKL_36745 (or equivalent) is installed in the cluster, a timeout equal to (*2* PV timeout*) will suffice to try all paths.

The SLVM volume groups are marked as shared volume groups and exported across the SGeRAC cluster using standard SGeRAC procedures. As noted above, multiple physical paths to each physical volume should be configured using the LVM PV Links feature or a separate multipathing product such as HP StorageWorks Secure Path.

Please note that, for the case in which the SLVM PVs being used by ASM are disk array LUs, the requirements in this section do not place any constraints on the configuration of the LUs. The LUs may be configured with striping, mirroring and other characteristics at the array level, following guidelines provided by Oracle and the array provider for use by ASM.

Sample Command Sequence for Configuring SLVM Volume Groups

In this section, we provide an example of a command sequence that can be used to prepare SLVM Logical Volumes for use by ASM to meet the requirements specified above.

The scenario for our example is that we are preparing a new volume group named vgora_asm with two PVs, each with two physical paths. The physical paths for the first PV are /dev/dsk/c9t0d1 and /dev/dsk/c10t0d1 and those for the second PV are /dev/dsk/c9t0d2 and /dev/dsk/c10t0d2.

1. Create the volume group with the two PVs, incorporating the two physical paths for each (choosing *hh* to be the next hexadecimal number that is available on the system, after the volume groups that are already configured) :

```
# pvcreate -f /dev/dsk/c9t0d1
# pvcreate -f /dev/dsk/c9t0d2

# mkdir /dev/vgora_asm
# mknod /dev/vgora_asm/group c 64 0xhh0000

# vgcreate /dev/vgora_asm /dev/dsk/c9t0d1
# vgextend /dev/vgora_asm /dev/dsk/c9t0d2
# vgextend /dev/vgora_asm /dev/dsk/c10t0d1
# vgextend /dev/vgora_asm /dev/dsk/c10t0d2
```

2. For each of the two PVs, create a corresponding LV
   - Create an LV of zero length
   - Mark the LV as contiguous
   - Extend each LV to the maximum size possible on that PV (the number of extents available in a PV can be determined via vgdisplay -v <vgname>)
   - Configure LV timeouts, based on the PV timeout and number of physical paths, as described in the previous section. If a PV timeout has been explicitly set, its value can be displayed via pvdisplay -v. If not, pvdisplay will show a value of default, indicating that the timeout is determined by the underlying disk driver. For SCSI, in HP-UX 11i v2, the default timeout is 30 seconds.
   - Null out the initial part of each LV to ensure ASM accepts the LV as an ASM disk group member.[5] Note that we are zeroing out the LV data area, not its metadata. It is the ASM metadata that is being cleared.

[5] See Oracle Metalink Doc ID: Note:268481.1 *RE-CREATING ASM INSTANCES AND DISKGROUPS* at https://metalink.oracle.com/ (Oracle MetaLink account required)

```
# lvcreate -n lvol1 vgora_asm
# lvcreate -n lvol2 vgora_asm

# lvchange -C y /dev/vgora_asm/lvol1
# lvchange -C y /dev/vgora_asm/lvol2

# Assume vgdisplay shows each PV has 2900 extents in our example
# lvextend -l 2900 /dev/vgora_asm/lvol1 /dev/dsk/c9t0d1
# lvextend -l 2900 /dev/vgora_asm/lvol2 /dev/dsk/c9t0d2

# Assume a PV timeout of 30 seconds.
# There are 2 paths to each PV, so the LV timeout value is 60 seconds
# lvchange -t 60 /dev/vgora_asm/lvol1
# lvchange -t 60 /dev/vgora_asm/lvol2

# dd if=/dev/zero of=/dev/vgora_asm/rlvol1 bs=8192 count=12800
# dd if=/dev/zero of=/dev/vgora_asm/rlvol1 bs=8192 count=12800
```

3. Export the volume group across the SGeRAC cluster and mark it as shared, as specified by SGeRAC documentation.[6] Assign the right set of ownerships and access rights to the raw logical volumes on each node as required by Oracle (oracle:dba and 0660, respectively).

We can now use the raw logical volume device names as disk group members when configuring ASM disk groups using the Oracle database management utilities. There are a number of ways of doing this described in Oracle ASM documentation, including the dbca database creation wizard as well as sqlplus.

The same command sequence, with some modifications, can be used for adding new disks to an already existing volume group that is being used by ASM to store one or more RAC databases. If the database(s) should be up and running during the operation, we use the Single Node Online volume Reconfiguration (SNOR) feature of SLVM.[6]

Step 1 of the above sequence is modified as follows:

- First, deactivate the volume group vgora_asm on all nodes but one, say node A. This requires prior shutdown of the database(s) using ASM-managed storage and ASM itself, on all nodes but node A. See the section ASM Halt is needed to ensure disconnect of ASM from SLVM Volume Groups to understand why it is not adequate to shut down only the database(s) using the volume group to be reconfigured, and why we must shut down ASM itself and therefore all database(s) using ASM-managed storage, on all nodes but node A.

- Next, on node A, switch the volume group to exclusive mode, using SNOR.

- Initialize the disks to be added with pvcreate, and then extend the volume group with vgextend.

Step 2 remains the same. Logical volumes are prepared for the new disks in the same way.

[6] See *Using Serviceguard Extension for RAC, June 2007* at http://docs.hp.com/en/ha.html → Serviceguard Extension for Real Application Cluster (Serviceguard OPS Edition).

In step 3, switch the volume group back to shared mode, using SNOR, and export the VG across the cluster, ensuring that the right ownership and access rights are assigned to the raw logical volumes. Activate the volume group, and restart ASM and the database(s) using ASM-managed storage on all nodes (they are already active on node A).

SG/SGeRAC Support for ASM on HP-UX 11i v3

Overview

Support for ASM with SG/SGeRAC commences with SGeRAC A.11.17.01 on HP-UX 11i v3. SGeRAC support ASM where the member disk groups are either logical volumes managed by HP Shared Logical Volume Manager (SLVM) or raw disks/disk array logical units(LUs).

A new I/O infrastructure that enables the native built-in multipathing functionality is introduced in HP-UX 11i v3. This feature offers users a continuous I/O access to a LUN or disk if any of the paths fails. This feature is enabled in the operating system by default. In addition, new DSF(device special file) format is introduced in this operating system. An example of new DSF is /dev/disk/disk1, compare to the legacy DSF, /dev/rdsk/cxtydz.

Please note that it has been possible to set up Oracle databases using ASM on HP-UX from the time Oracle 10g became available on HP-UX, using the following configurations that do not require SG/SGeRAC:

- non-clustered single instance Oracle
- Oracle single instance and RAC databases running in a pure Oracle clusterware environment

The following requirements/restrictions apply to SG/SGeRAC (A.11.17.01 or later) support of ASM (summarized in Table 2):

Table 2 : ASM support with SG/SGeRAC A.11.17.01 or later

SUPPORTED	**NOT SUPPORTED**
Oracle RAC database	Oracle single instance database, with or without Oracle clusterware
SLVM Logical Volumes as ASM Disk Group Members.	CVM Logical Volumes as ASM Disk Group Members
ASM configurations on SGeRAC using SGeRAC A.11.17.01 or later, and Oracle 10g R2 RAC	Single instance Oracle database with ASM managed storage running in Serviceguard cluster.
ASM configurations on SGeRAC using SGeRAC A.11.17.01 or later, and 11g R1 RAC	Single instance Oracle database with ASM-managed storage in SGeRAC configuration wherein SGeRAC provides node membership to Oracle Clusterware.

Maximum SGeRAC cluster size for ASM configurations is 16	
SGeRAC Toolkit[3] framework based on Multi-Node Package (MNP) /Simple Package Dependency to manage ASM configurations on SGeRAC up to a maximum cluster size of 16	
Raw disks/disk array LUs as ASM Disk Group Members.	

ASM over SLVM

SGeRAC support for ASM-over-SLVM continues in HPUX 11i v3. SGeRAC 11.17.01 or later is required for ASM-over-SLVM configurations on HP-UX 11i v3.

The advantages of the "ASM-over-SLVM" configuration are as follows:

- ASM-over-SLVM ensures that the HP-UX devices used for disk group members will have the same names (the names of logical volumes in SLVM volume groups) on all nodes, easing ASM configuration.
- ASM-over-SLVM protects ASM data against inadvertent overwrites from nodes inside/outside the cluster. If the ASM disk group members are raw disks, there is no protection currently preventing these disks from being incorporated into VxVM volume/disk groups.

The disadvantages of the ASM-over-SLVM configuration are as follows:

- Additional configuration and management tasks are imposed by the extra layer of volume management (administration of volume groups, logical volumes, physical volumes)
- There is a small performance impact from the extra layer of volume management
- SLVM has some restrictions in the area of online reconfiguration, the impact of which will be examined later in this document

Configuring SLVM Volume Groups for ASM Disk Groups

We would like the logical volumes presented to ASM to resemble raw disks, as far as possible. Hence, each SLVM logical volume (LV) used as a member of an ASM disk group is required to be laid out to occupy the usable space, in contiguous fashion, of exactly one single physical volume (PV). This implies that the LV

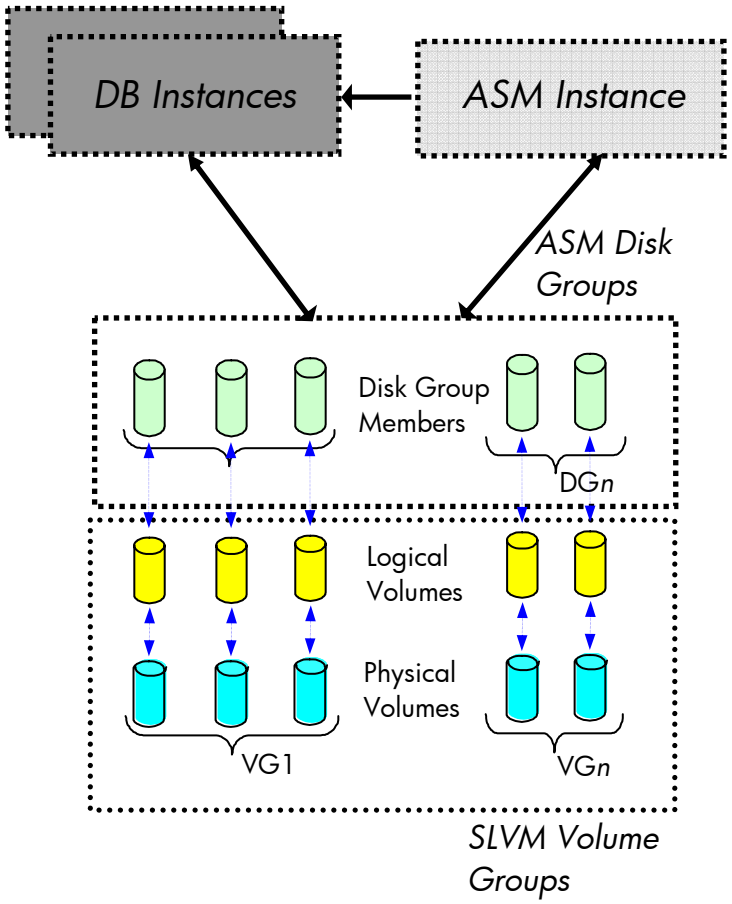
- should be contiguous,
- should not be striped or mirrored,
- should not span multiple PVs
- and should not share a PV with other LVs.

The idea is that ASM provides the mirroring, striping, slicing and dicing functionality as needed and SLVM supplies the multipathing functionality not provided by ASM. Figure 5 indicates this 1-1 mapping between SLVM PVs and LVs used as ASM disk group members.

Further, the default retry behavior of SLVM could result in an I/O operation on an SLVM LV taking an indefinitely long period of time. This behavior could impede ASM retry and rebalance capabilities; hence a finite timeout must be configured for each SLVM LV. For example, the timeout could be configured to the value (*total number of physical paths to the PV * PV timeout*), providing enough time for SLVM to try all available paths, if needed.

The PVs used in an ASM disk group can be organized into SLVM volume groups as desired by the customer. In the example shown in Figure 5, for each ASM disk group, the PVs corresponding to its members are organized into a separate SLVM volume group.

Figure 5. 1-1 mapping between SLVM logical and physical volumes for ASM configuration



The SLVM volume groups are marked as shared volume groups and exported across the SGeRAC cluster using standard SGeRAC procedures.

Please note that, for the case in which the SLVM PVs being used by ASM are disk array LUs, the requirements in this section do not place any constraints on the configuration of the LUs. The LUs may be configured with striping, mirroring and other characteristics at the array level, following guidelines provided by Oracle and the array provider for use by ASM.

Sample Command Sequence for Configuring SLVM Volume Groups

In this section, we provide an example of a command sequence that can be used to prepare SLVM Logical Volumes for use by ASM. The example below demonstrates how volume group is created using new DSF format. HP-UX will automatically use the redundant path for the volume group in the background.

The scenario for our example is that we are preparing a new volume group named *vgora_asm*. The physical path for the first PV is */dev/rdisk/disk1* and the second PV is */dev/rdisk/disk2*.

1. Create the volume group with the two PVs (choosing *hh* to be the next hexadecimal number that is available on the system, after the volume groups that are already configured):

```
# pvcreate -f /dev/rdisk/disk1
# pvcreate -f /dev/rdisk/disk2

# mkdir /dev/vgora_asm
# mknod /dev/vgora_asm/group c 64 0xhh0000

# vgcreate /dev/vgora_asm /dev/disk/disk1
# vgextend /dev/vgora_asm /dev/disk/disk2
```

2. For each of the two PVs, create a corresponding LV
   - Create an LV of zero length
   - Mark the LV as contiguous
   - Extend each LV to the maximum size possible on that PV (the number of extents available in a PV can be determined via *vgdisplay -v <vgname>*)
   - Configure LV timeouts, based on the PV timeout and number of physical paths, as described in the previous section. If a PV timeout has been explicitly set, its value can be displayed via *pvdisplay -v*. If not, *pvdisplay* will show a value of *default*, indicating that the timeout is determined by the underlying disk driver. For SCSI, in HP-UX 11i v3, the default timeout is 30 seconds.
   - Null out the initial part of each LV to ensure ASM accepts the LV as an ASM disk group member.[5] Note that we are zeroing out the LV data area, not its metadata. It is the ASM metadata that is being cleared.

```
# lvcreate -n lvol1 vgora_asm
# lvcreate -n lvol2 vgora_asm

# lvchange -C y /dev/vgora_asm/lvol1
# lvchange -C y /dev/vgora_asm/lvol2

# Assume vgdisplay shows each PV has 2900 extents in our example
# lvextend -l 2900 /dev/vgora_asm/lvol1 /dev/disk/disk1
# lvextend -l 2900 /dev/vgora_asm/lvol2 /dev/disk/disk2

# Assume a PV timeout of 30 seconds.
# There are 2 paths to each PV, so the LV timeout value is 60 seconds
# lvchange -t 60 /dev/vgora_asm/lvol1
# lvchange -t 60 /dev/vgora_asm/lvol2

# dd if=/dev/zero of=/dev/vgora_asm/rlvol1 bs=8192 count=12800
# dd if=/dev/zero of=/dev/vgora_asm/rlvol1 bs=8192 count=12800
```

1. Export the volume group across the SGeRAC cluster and mark it as shared, as specified by SGeRAC documentation.[6] Assign the right set of ownerships and access rights to the raw logical volumes on each node as required by Oracle (*oracle:dba* and *0660*, respectively).

We can now use the raw logical volume device names as disk group members when configuring ASM disk groups using the Oracle database management utilities. There are a number of ways of doing this described in Oracle ASM documentation, including the *dbca* database creation wizard as well as *sqlplus*.

The same command sequence, with some modifications, can be used for adding new disks to an already existing volume group that is being used by ASM to store one or more RAC databases. If the database(s) should be up and running during the operation, we use the Single Node Online volume Reconfiguration (SNOR) feature of SLVM.

Step 1 of the above sequence is modified as follows:

- First, deactivate the volume group *vgora_asm* on all nodes but one, say node *A*. This requires prior shutdown of the database(s) using ASM-managed storage and ASM itself, on all nodes but node *A*. See the section to understand why it is not adequate to shut down only the database(s) using the volume group to be reconfigured, and why we must shut down ASM itself and therefore all database(s) using ASM-managed storage, on all nodes but node *A*.
- Next, on node *A*, switch the volume group to exclusive mode, using SNOR.
- Initialize the disks to be added with *pvcreate*, and then extend the volume group with *vgextend*.

Step 2 remains the same. Logical volumes are prepared for the new disks in the same way.

In step 3, switch the volume group back to shared mode, using SNOR, and export the VG across the cluster, ensuring that the right ownership and access rights are assigned to the raw logical volumes. Activate the volume group, and restart ASM and the database(s) using ASM-managed storage on all nodes (they are already active on node *A*).

ASM over Raw disk

As mentioned above, a new I/O infrastructure that enables the native built-in multipathing functionality is included in HP-UX 11i v3. This feature offers users a continuous I/O access to a LUN or disk if any of the paths fails. This newly added functionality enables SGeRAC (11.17.01 or later) to support ASM on raw disks/disk array LUs. In HP-UX 11i v3, new DSF is introduced. SGeRAC will support the DSF format that ASM support with the restriction that native multipathing feature is enabled.

The advantages for "ASM-over-raw" are as follow:

- There is a small performance improvement from one less layer of volume management.
- Online disk management (adding disks, deleting disks) is supported with ASM-over-raw.

The disadvantages for "ASM-over-raw" are as follow:

- Might not see the HP-UX devices (raw disks/disk array LUs) used for disk group members as the same names on all nodes.
- There is no protection to prevent the raw disks from being incorporated into VxVM volume/disk groups.

Configure Raw Disks/Disk Array Logical Units for ASM Disk Group

Oracle provides instructions on how to configure disks for ASM where the member disks are raw logical volume. The instructions to configure raw disks/disk LUs are:

> For Oracle 10g R2, please refer to *Oracle Database Installation Guide 10g Release 2 for hp-ux Itanium* , Chapter 2, Preinstallation Tasks, section "Preparing Disk Group for an Automatic Storage Management Installation".
>
> For 11g R1, please refers to *Oracle Clusterware Installation Guide 11g Release 1 (11.) for HP-UX*, Chapter 5, Configuring Oracle Real Application Clusters Storage, section "Configuring Disks for Automatic Storage Management".

Then, these raw devices can be used as disk group members to configure ASM disk group members using Oracle database management utilities.

Additional Hints on ASM Integration with SGeRAC

In this section, we provide some pointers that may be useful when deploying ASM in a SGeRAC environment.

Consider using the MNP/Simple Dependency-based SGeRAC Toolkit's Framework

The SGeRAC Toolkit[3] which provides a framework to integrate Oracle 10g R2 RAC or 11g R1 RAC with SGeRAC and is based on the SGeRAC A.11.17 multi-node package and simple package dependency features provides a uniform, intuitive and easy-to-manage method to co-ordinate between SGeRAC and Oracle Clusterware and manage all the storage options supported by SGeRAC, including ASM-over-SLVM and ASM-over-raw devices.

ASM Halt is needed to ensure disconnect of ASM from SLVM Volume Groups

This section is specific to ASM-over-SLVM only.

When an ASM disk group is dismounted on a node in the SGeRAC cluster, there is no guarantee that processes in the ASM instance on that node and client processes of the ASM instance will close their open file descriptors for the raw volumes underlying the members of that ASM disk group.

Consider a configuration in which there are multiple RAC databases using ASM to manage their storage in a SGeRAC cluster. Assume each database stores its data in its own exclusive set of ASM disk groups.

If we shut down the database instance for a specific RAC database on a node, and then dismount its ASM disk groups on that node, some Oracle processes may still hold open file descriptors to the underlying raw logical volumes. Hence an attempt at this point to deactivate the corresponding SLVM volume group(s) on the node may fail. The only way to ensure success of the deactivation of the volume groups is to first shut down the ASM instance and its clients (including all databases that use ASM based storage) on that node.

The major implications of this behavior include the following:

Many SGeRAC customers use SGeRAC packages to start and shut down Oracle RAC instances. In the startup and shutdown sequences, the package scripts activate and deactivate the SLVM volume groups used by the instance.

For the ASM environment, it is not appropriate to include SLVM volume group activation and deactivation in the database instance package control script, since the deactivation may fail. In the SGeRAC Toolkit that is MNP/Simple Dependency-based, the SLVM volume groups underlying ASM disk groups are managed instead from the package control script for Oracle Clusterware[3].

When there are multiple RAC databases using ASM to manage their storage in the cluster, online reconfiguration of SLVM volume groups for one database will impact the others, even if each database is configured with its own set of ASM disk groups and underlying SLVM volume groups.

An SLVM volume group is reconfigured online, using the SLVM SNOR feature. This procedure requires the volume group to be deactivated on all nodes but one. This in turn requires shutting down the ASM instance and all client database instances on all nodes but one.

However, note that many storage reconfiguration operations can be confined to the ASM layer. For example, if a new file has to be created in a disk group, there will be no SLVM operation required, if the disk group has adequate free space. Adding disks to, and deleting disks from, ASM disk groups may require SLVM reconfiguration (online disk addition is discussed in the section).

It is recommended that, at the time a physical volume is added to a volume group for use by ASM, that the corresponding logical volume be created at the same time, to avoid the potential impact of a future online SLVM reconfiguration operation to create the logical volume. Note that, when adding one or more physical volumes in an SLVM configuration, one can avoid an online SLVM reconfiguration operation by creating a new volume group for the physical volume(s).

ASM may require Modified Backup/Restore Procedures

Backup/restore procedures for databases stored in ASM must make use of Oracle Recovery Manager (RMAN), per Oracle documentation. Customers wishing to deploy ASM should ensure that their backup/restore tools/scripts/procedures are compatible with this ASM requirement.

Installation, Configuration, Support and Troubleshooting

Oracle ASM is part of the Oracle database server installation (both 10g R2 and 11g R1). No additional software from HP is required to operate ASM in SGeRAC environment. Information on how to use the optional MNP/Simple Dependency based SGeRAC Toolkit is separately documented.[3]

Oracle ASM and ASM disk groups may be configured at the time of creating a database or as a separate step. These tasks are carried out using Oracle database management tools, for example, the database creation wizard *dbca*. In either case, ensure that SLVM volume groups and raw disks/LUs have been prepared and activated prior to ASM disk group configuration or reconfiguration, as discussed in this document. For SLVM, use the names of raw LVs contained in the SLVM volume groups when Oracle commands and utilities require the specification of the names of raw HP-UX devices for storing ASM disk groups.

For troubleshooting, it may be useful to look at subdirectories for database and ASM instances under *$ORACLE_BASE/admin/*, where log and trace files for these instances are deposited. Oracle clusterware deposits log and trace information regarding its management of ASM and database instances under *$ORACLE_HOME/log/<hostname>/racg*.

Related Documentation and Scripts

- *Using Serviceguard Extension for RAC, June 2007* at http://docs.hp.com/en/ha.html → Serviceguard Extension for Real Application Cluster (Serviceguard OPS Edition)
- Whitepaper "*Use of Serviceguard Extension for RAC Toolkit with Oracle 10g RAC*" http://docs.hp.com under High Availability, Serviceguard Extension for RAC December 2006
- Serviceguard Extension for RAC Toolkit README is available when SGeRAC 11.18 is installed or download from http://software.hp.com under High Availability, Serviceguard Extension for RAC Toolkit
- Oracle Clusterware Installation Guide 11g Release 1 (11.1) for HP-UX at http://www.oracle.com/pls/db111/portal.portal_db?selected=11&frame= → HP-UX Installation Guides → Clusterware Installation Guide for HP-UX
- ASM related sections in Oracle Manuals
    - Oracle® Database Administrator's Guide 10g R2 (10.2) at http://www.oracle.com/pls/db102/portal.portal_db?selected=4 → Guides → Administrator's Guide
    - Oracle® Database Oracle Clusterware and Oracle Real Application Clusters Administration and Deployment Guide10g R2 (10.2) at http://www.oracle.com/pls/db102/portal.portal_db?selected=4 → High Availability → Oracle Clusterware and Oracle Real Application Clusters Administration and Deployment Guide
    - Oracle® Database Administrator's Guide 11g R1 at http://www.oracle.com/pls/db111/portal.portal_db?selected=4&frame= → Supporting Documentation → Administrator's Guide
    - http://www.oracle.com/technology/products/database/asm/index.html
    - http://www.oracle.com/pls/db111/portal.portal_db?selected=4&frame= → Storage → Storage Administrator's Guide

XXXX-XXXXEN, 1/2008