

An Oracle White Paper
October 2012

SAP with Oracle Real Application Clusters 11g Release 2 and Oracle Automatic Storage Management 11g Release 2

Advanced Configurations & Techniques



Introduction	3
Related SAP Notes.....	3
Storage based Mirroring vs. Host based Mirroring	4
Data Redundancy in Oracle ASM	5
Oracle ASM File Types.....	6
Initial Installation of Oracle Grid Infrastructure	6
Host based Mirroring over multiple Storage Subsystems with Oracle ASM 8	
Configuring Oracle ASM Disks for OCR and Voting.....	10
Preferred Read Failure Groups	15
Zero Downtime Storage Maintenance	16
"Split Mirror" based Online Backups and Database Copies	16
Storage based Online Backups	17
Restore and Recovery from Backup Disks.....	22
Storage based Database Copies	24

Introduction

Oracle Real Application Clusters 11g Release 2 (Oracle RAC) and Oracle Automatic Storage Management 11g Release 2 (Oracle ASM) technologies provide unexcelled availability, scalability and performance for your SAP applications.

This document discusses Oracle Grid Infrastructure and Oracle ASM configurations utilizing 2 or 3 storage systems in different datacenters as well as storage system based methods to create copies or backups from running SAP databases.

Although this whitepaper does not focus on a specific Operating System platform and all configurations and techniques work on each supported platform the examples shown are from a MS Windows based cluster.

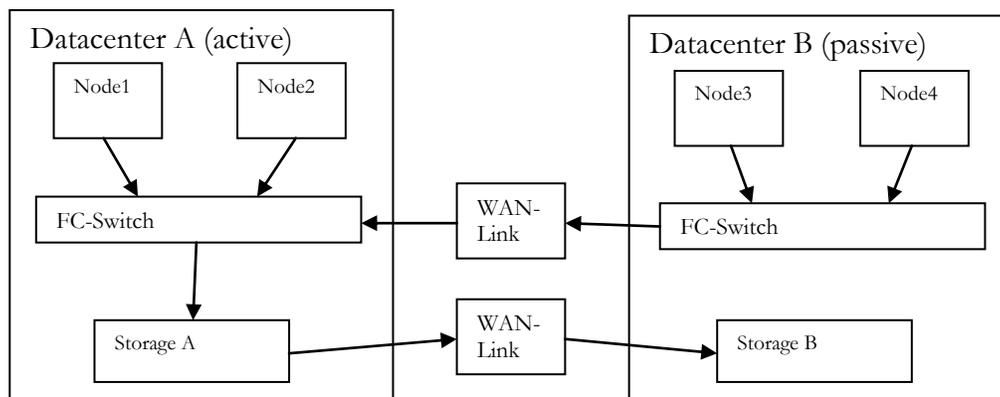
Related SAP Notes

#1570680	Oracle 11g RAC with Oracle ASM in stretched clusters
#1627541	BR*Tools support for Oracle ASM
#1554661	Configuration of environment for user 'oracle'
#1550133	Oracle Automatic Storage Management (ASM)
#1598594	BR*Tools configuration for Oracle installation under 'oracle' user
#1628116	Split-Mirror-Backup of Oracle RAC Databases

Storage based Mirroring vs. Host based Mirroring

With storage based mirroring each write-I/O going to a disk of storage 'A' needs to be mirrored by storage 'A' to storage 'B'. During this time the host waits until both I/Os complete. Usually the host is still connected to both storages but only utilizes one of them actively until the active one fails or the active role is switched between storages. This way of storage based mirroring is called 'active-passive mode'.

The example below shows a write-I/O scenario with storage based mirroring in a stretched 4-node cluster.



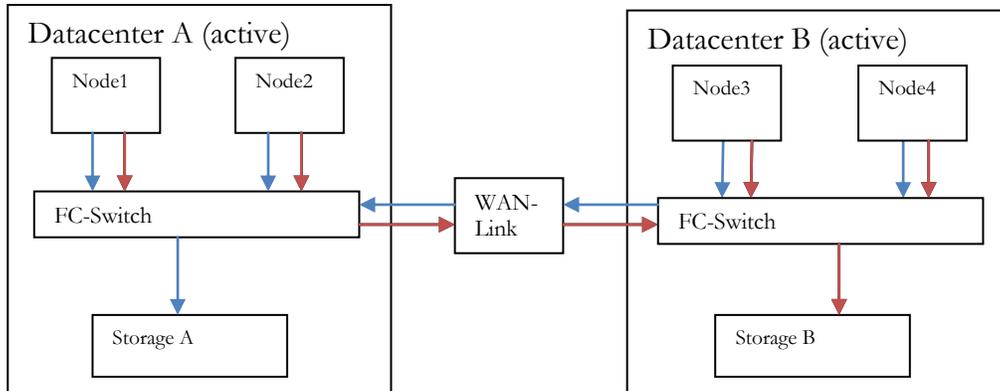
All cluster nodes of datacenter A and datacenter B access storage A.

Storage A communicates with storage B to maintain the mirrored disks.

Host based mirroring does not require the storage subsystem to write to mirror disks. Instead of creating just one write-I/O the host generates two (normal redundancy) or three (high redundancy) write-I/Os by itself. This eliminates the need of a separate high speed link between the storage subsystems. This method is called 'active-active mode'.

The major benefit of host based mirroring in a cluster is that it allows to run storages in active-active mode where the cluster nodes write to both storages but can preferable read from their local storage. In a distributed cluster environment with one storage subsystem per datacenter reading from the local storage is a big advantage as it eliminates most of the I/O load between the different sites.

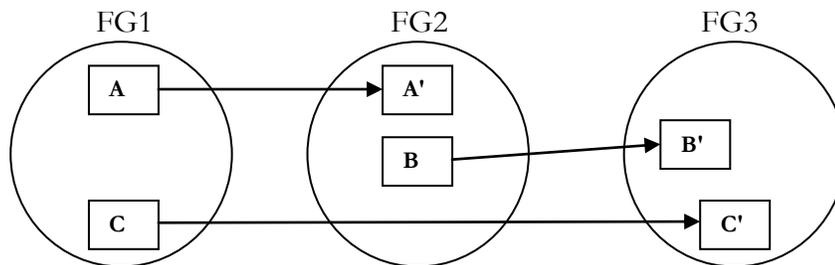
The example below shows a write-I/O scenario with host based mirroring. Each host generates two write-I/O's: One to the local and to the remote storage subsystem.



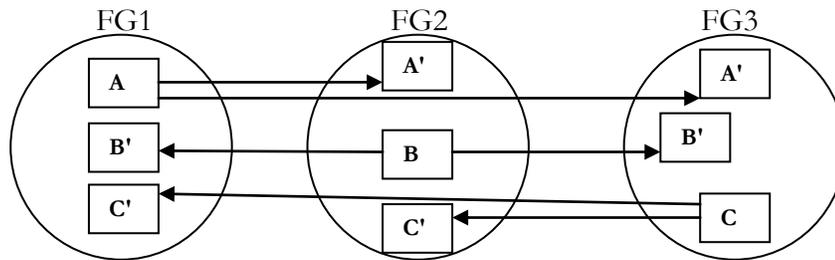
Data Redundancy in Oracle ASM

The most important thing about Oracle ASM redundancy that must be understood is that Oracle ASM does not maintain redundancy like traditional RAID based systems. Oracle ASM does not mirror a whole disk image like RAID1 nor does it use stripes with parity like RAID5. Oracle ASM divides each disk into allocation units (AU). Space on disk is allocated in 'file extents' each consisting of a number of allocation units. Oracle ASM redundancy - normal or high - is achieved through mirroring these file extents to Oracle ASM disks belonging to different failure groups. The Oracle ASM failure group mechanism ensures that the same piece of data will never be stored twice within the same failure group with one exception: If there are no Oracle ASM disks available within the diskgroup that belong to another failure group than the disk to which a file extent belongs to (which means that the required redundancy-level cannot be achieved) ASM will start to mirror file extents within the same failure group but on different ASM disks. By default every Oracle ASM disk that is not explicitly assigned to a failure group will define its own failure group and is therefore a candidate for mirrored file extents.

The example below shows a 'normal-redundancy' Oracle ASM diskgroup consisting of 3 Oracle ASM disks. Each file extent is mirrored to only one other failure group where Oracle ASM decides to which one.



The example below shows a 'high-redundancy' Oracle ASM diskgroup consisting of 3 Oracle ASM disks. Each file extent is mirrored to two other failure groups.



Oracle ASM File Types

Oracle ASM knows several file types. A complete list can be found in the *'Oracle Automatic Storage Management Administrators Guide'*. As all file types are striped across Oracle ASM disks of the same failure group the OCRFILE is by default striped across the 3 or 5 Oracle ASM disks of diskgroup 'OCR'. If these Oracle ASM disks are located in different storage subsystems the OCRFILE will be striped across multiple storage subsystems. It will become unusable if one of the storage subsystems fails.

Note that the list of file types does not contain voting files. Voting files are stored on Oracle ASM disks but are not managed by Oracle ASM and are not striped across multiple Oracle ASM disks. Voting files are managed by Oracle Clusterware.

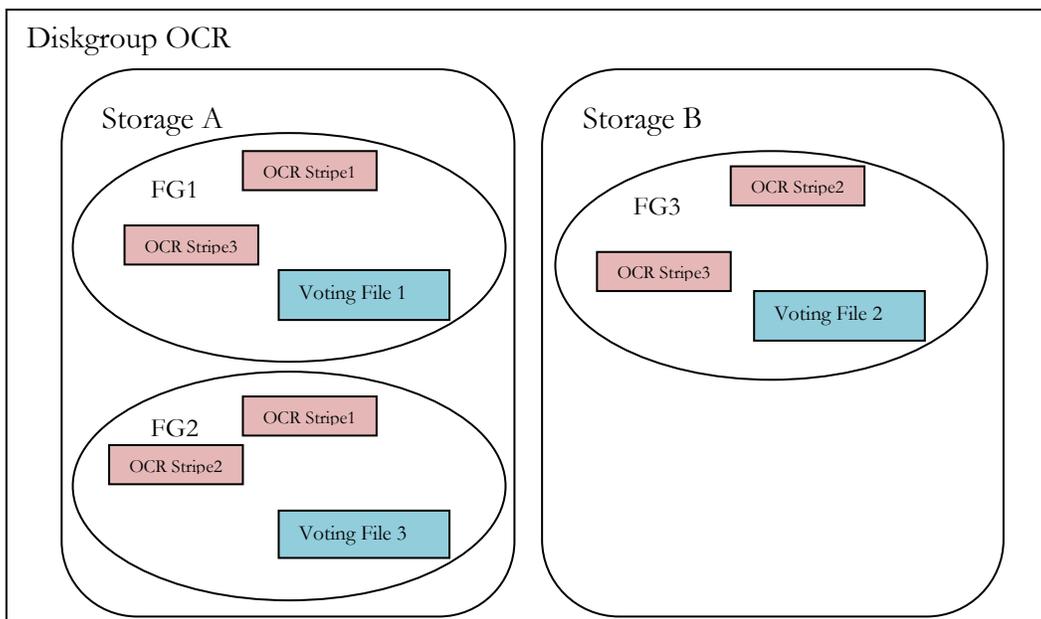
Initial Installation of Oracle Grid Infrastructure

For SAP the initial installation of Oracle Grid Infrastructure recommends to use three Oracle ASM disks for a new 'normal-redundancy' Oracle ASM diskgroup. Three Oracle

ASM disks are required because Oracle Clusterware creates one voting file on each Oracle ASM disk. This configuration is perfect for a cluster that utilizes only **one** fault-tolerant storage system but for clusters that run with 2 or 3 storage subsystems we need to modify the standard installation.

As already mentioned 'OCRFILE' is a predefined Oracle ASM file type that striped across all failure groups within the 'OCR' diskgroup. In a standard installation on a 'normal-redundancy' Oracle ASM diskgroup with 3 Oracle ASM disks this means that Oracle ASM will distribute all the stripes of the OCRFILE across all 3 Oracle ASM disks while voting files are not striped by Oracle ASM.

The example below shows the content of Oracle ASM diskgroup 'OCR' after Grid Infrastructure has been installed with SAP recommended default options and across 2 storage subsystems:



Each file extent is mirrored on another failure group of diskgroup 'OCR'. As Oracle ASM does not know to which storage subsystem an Oracle ASM disk belongs to it distributes the stripes across all available Oracle ASM disks with respect to the failure group. The problem with this configuration is that if one storage subsystem is missing (in this example Storage A) the OCRFILE becomes unusable as 'OCR Stripe 1' would be missing. The same would happen to other striped files on Oracle ASM.

Additionally two voting files would be located on storage subsystem A and one would be located on storage subsystem B. This would cause node evictions on all nodes attached to storage B if a split brain situation occurs.

In order to fix these problems it is recommended to reserve 2 additional Oracle ASM disks on each storage subsystem for a new 'OCR1' diskgroup and a new 'VOTE' diskgroup. After the standard installation has finished we will modify the existing configuration by moving the OCRFILE from diskgroup 'OCR' to diskgroup 'OCR1' and by moving the voting files from diskgroup 'OCR' to diskgroup 'VOTE'. The required steps will be discussed later.

Host based Mirroring over multiple Storage Subsystems with Oracle ASM

Oracle ASM allows organizing disks into failure groups. In order to configure an Oracle ASM diskgroup for normal redundancy with a second storage or for high redundancy with a third storage you just need to assign all disks that are located in the same storage to the same failure group.

The major benefits of such a configuration are zero downtime during maintenance of one out of 2 storage subsystems or two out of 3 storage subsystems or if a storage failure or a complete site failure occurs.

If you want to configure Oracle ASM with 2 storage subsystems for your database you need a quorum disk with a voting file in a third location in order to be able to resolve split brain conditions. This quorum disk can be implemented as a physical disk coming from a third storage, as a file on a NFS share (UNIX and Linux) or as an iSCSI disk (Windows).

For Unix/Linux based systems please refer to the Oracle whitepaper 'Oracle Clusterware 11g Release 2 (11.2) – Using standard NFS to support a third voting file for extended cluster configurations' for a detailed description how to implement an NFS based voting file. (<http://www.oracle.com/technetwork/database/clusterware/overview/grid-infra-thirdvoteonnfs-131158.pdf>).

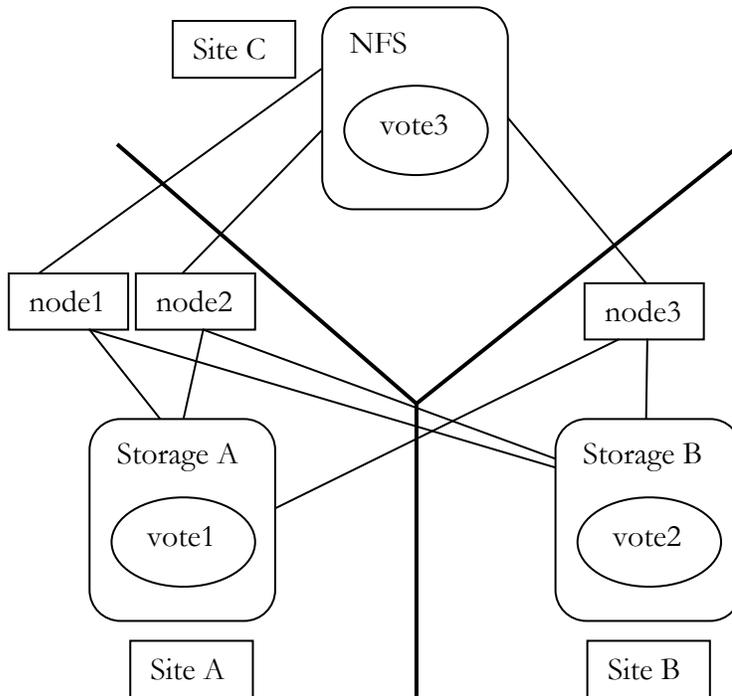
When using Oracle ASM with multiple storage subsystems make sure that all shared disks can be accessed under the same device name across all cluster nodes. On Linux this can be achieved by using WWID's with multipath and/or udev.

For MS Windows based systems it is recommended to either use an shared SCSI disk from a third SAN or to use an shared iSCSI disk exported by a third storage subsystem for the third voting file.

On MS Windows the storage subsystems must ensure that the disks are always found in the same order across all cluster nodes. By default Windows enumerates the disks in the order they are reported during SCSI inquiry. If multiple storage subsystems are involved it may happen that devices of one storage subsystem are reported faster than devices of another storage subsystem. The order in which the devices are reported could be different

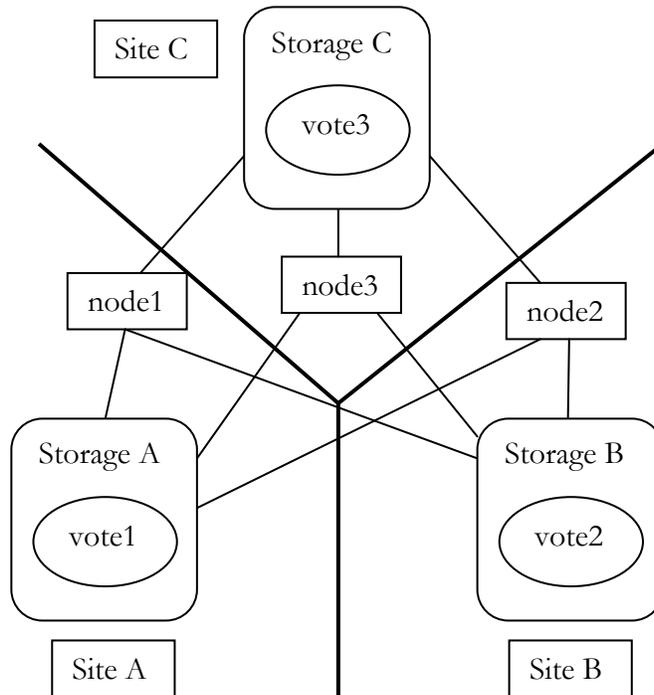
across reboots or across different cluster nodes and cause a lot of trouble for the administrator in identifying the disks and what they were used for.

The example below illustrates the logical connections to each Oracle ASM disk containing a voting file in a 3 node RAC cluster with 2 storage subsystems and an NFS based third voting file acting as quorum:



If the communication between Site A and Site B fails (cluster interconnect and access to remote disks) the cluster will resolve the split brain situation using each site's local voting file together with the quorum voting file of site C.

In a stretched cluster with 3 storage subsystems a quorum disk is not required as the third site will take part in reconfiguration of the cluster.



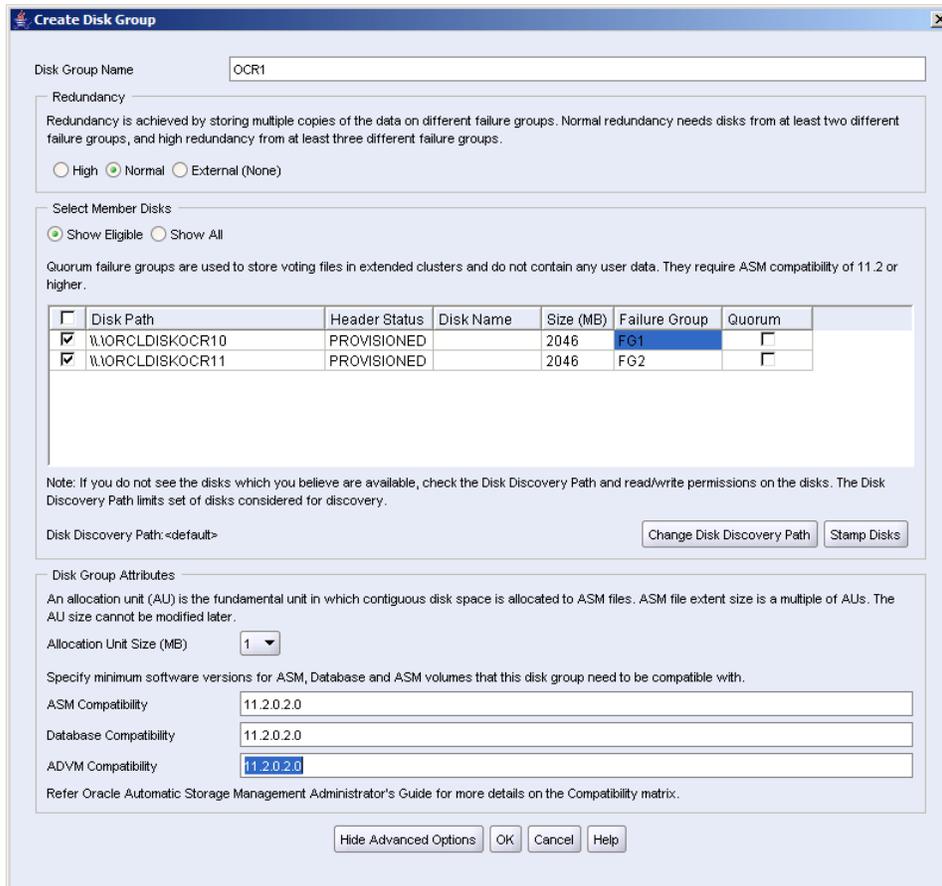
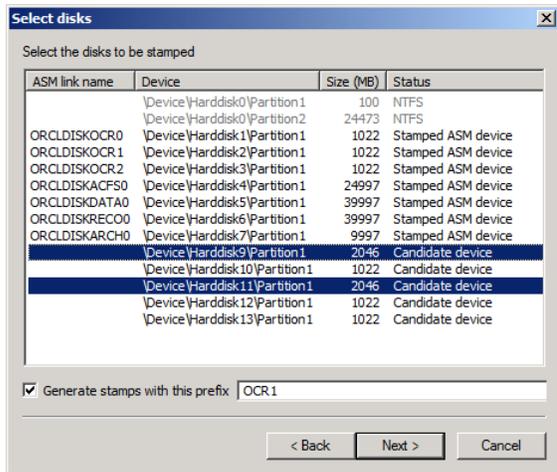
Configuring Oracle ASM Disks for OCR and Voting

The steps in this chapter describe how to transform the standard Grid Installation into a configuration with 2 or 3 storage subsystems. If you plan to run with 2 storage subsystems create a new Oracle ASM diskgroup (e.g. 'OCR1') with 'normal-redundancy' and add one disk from each storage subsystem as shown in the example on the next page.

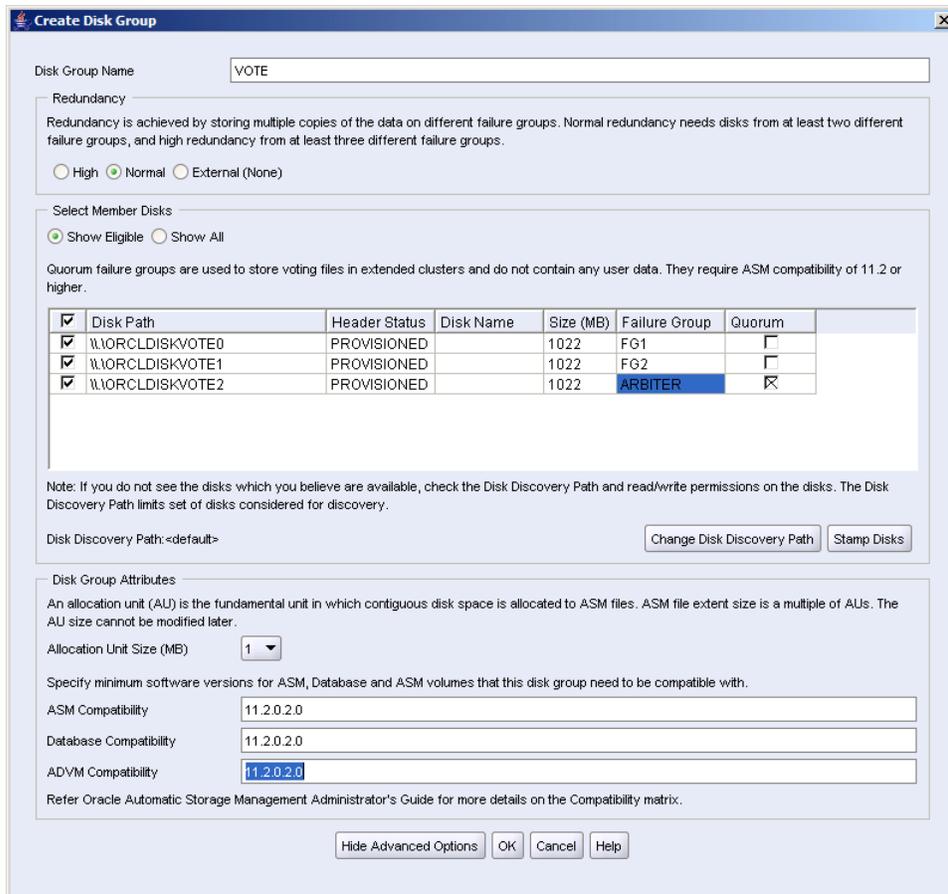
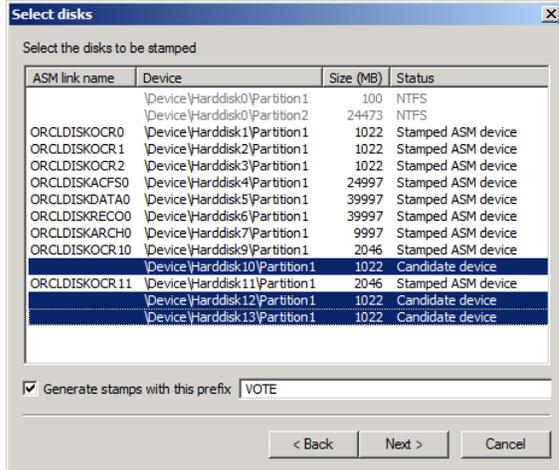
For configurations with 3 storage subsystems choose 'high-redundancy' and add one disk from each storage subsystem.

Make sure to set ASM, database and ADVM compatible values to at least 11.2.0.2.0 as this enables important functionality like the Oracle ASM command 'alter diskgroup <dg> online all'. For **new** SAP installations with SAPINST versions that do not yet allow specifying the Oracle Database initialization parameter 'compatible' you have to keep 'compatible.rdbms' on 11.2.0 until the installation has finished.

On MS Windows 'asmca' is used to stamp the new disks and to create the new diskgroups:

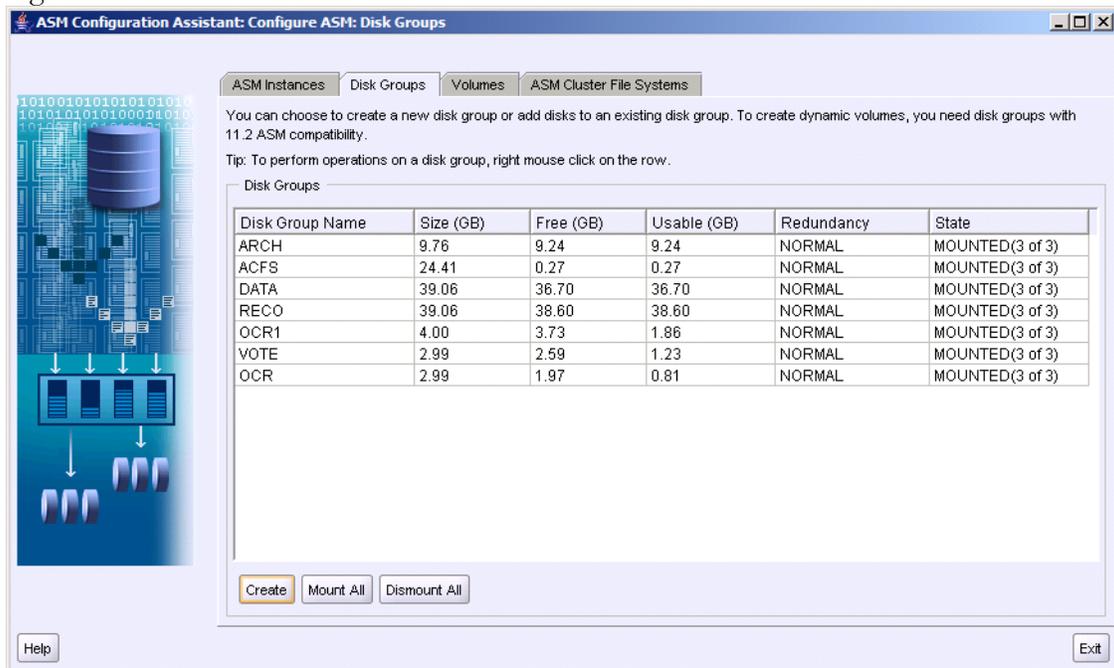


Create a new Oracle ASM diskgroup (e.g. 'VOTE') with 'normal-redundancy' and add an Oracle ASM disk from each storage subsystem plus the NFS or iSCSI based quorum disk if you run with just 2 storage subsystems.



At this point you have an Oracle ASM diskgroup 'OCR' that contains the OCR file and the voting files and 2 empty diskgroups e.g. 'OCR1' and 'VOTE'.

E.g.



The next step is to move the OCR from diskgroup 'OCR' to the new diskgroup 'OCR1' and to replace the voting files on diskgroup 'OCR' with voting files on diskgroup 'VOTE'.

- Check the consistency and location(s) of OCR files.

```
C:\Users\oracle>ocrcheck
Status of Oracle Cluster Registry is as follows :
Version                :                3
Total space (kbytes)   :           262120
Used space (kbytes)    :             3812
Available space (kbytes) :         258308
ID                     : 1526919991
Device/File Name       :      +OCR
                       Device/File integrity check succeeded
                       Device/File not configured
                       Device/File not configured
                       Device/File not configured
                       Device/File not configured
Cluster registry integrity check succeeded
Logical corruption check succeeded
```

- Add a mirror OCR file and check again.

```
C:\Users\oracle>ocrconfig -add +OCR1
C:\Users\oracle>ocrcheck
Status of Oracle Cluster Registry is as follows :
    Version                :                3
    Total space (kbytes)    :           262120
    Used space (kbytes)     :             3812
    Available space (kbytes):           258308
    ID                      :    1526919991
    Device/File Name        :             +OCR
                                Device/File integrity check succeeded
    Device/File Name        :             +OCR1
                                Device/File integrity check succeeded
                                Device/File not configured
                                Device/File not configured
                                Device/File not configured
    Cluster registry integrity check succeeded
    Logical corruption check succeeded
```

- Remove the OCR mirror on diskgroup 'OCR' and check again.

```
C:\Users\oracle>ocrconfig -delete +OCR
C:\Users\oracle>ocrcheck
Status of Oracle Cluster Registry is as follows :
    Version                :                3
    Total space (kbytes)    :           262120
    Used space (kbytes)     :             3812
    Available space (kbytes):           258308
    ID                      :    1526919991
    Device/File Name        :             +OCR1
                                Device/File integrity check succeeded
                                Device/File not configured
                                Device/File not configured
                                Device/File not configured
    Cluster registry integrity check succeeded
    Logical corruption check succeeded
```

- Query the voting files and location) that are currently in use.

```
C:\Users\oracle>crsctl query css votedisk
##  STATE  File Universal Id                File Name Disk group
--  -
  1.  ONLINE  4a11f6d099f64f04bf2615f54323fd6d  (\.\ORCLDISKOCR0) [OCR]
  2.  ONLINE  345700482a084f0cbf6aa951b46805bd  (\.\ORCLDISKOCR1) [OCR]
  3.  ONLINE  9b96fe635ebd4f07bfb68edc347acfb3  (\.\ORCLDISKOCR2) [OCR]
```

- Replace the old voting files with voting files on diskgroup 'VOTE'.

```
C:\Users\oracle>crsctl replace votedisk +VOTE
Successful addition of voting disk 776d91a6b2714f0fbf3fe7d464719174.
Successful addition of voting disk 58960399706e4f57bf9d42b4ee0051e4.
Successful addition of voting disk 66fce74f02bf4fe3bff723f93af3325f.
Successful deletion of voting disk 4a11f6d099f64f04bf2615f54323fd6d.
Successful deletion of voting disk 345700482a084f0cbf6aa951b46805bd.
Successful deletion of voting disk 9b96fe635ebd4f07bfb68edc347acfb3.
Successfully replaced voting disk group with +VOTE.
CRS-4266: Voting file(s) successfully replaced
```

- Use sqlplus to create a new spfile file for the Oracle ASM instance(s) on diskgroup 'OCR1'.

```
C:\Users\oracle>sqlplus

SQL*Plus: Release 11.2.0.2.0 Production on Wed Aug 24 15:41:38 2011
Copyright (c) 1982, 2010, Oracle. All rights reserved.

Enter user-name: / as sysasm

Connected to:
Oracle Database 11g Enterprise Edition Release 11.2.0.2.0 - 64bit
Production
With the Real Application Clusters and Automatic Storage Management
options

SQL> create pfile='c:\asmpfile.ora' from spfile;

File created.

SQL> create spfile='+OCR1' from pfile='c:\asmpfile.ora';

File created.
```

- As the Oracle ASM instances have been started with a spfile from the old 'OCR' diskgroup this file is still in use. So restart the Clusterware or reboot the cluster nodes and drop diskgroup 'OCR' if no longer needed.

Preferred Read Failure Groups

In a stretched cluster environment with more than one storage system it is recommended to configure the ASM instances to prefer reads from their 'local' storage subsystem over reads from a 'remote' storage subsystem.

This is achieved by setting the initialization parameter <asminstancename>.asm_preferred_read_failure_groups to a list of failure groups that are located in the 'local' storage.

The format of the parameter is
<asminstancename>.asm_preferred_read_failure_groups =
<dname>.<failgroupname>,<dname>.<failgroupname>,...

E.g.

```
+ASM1.ASM_PREFERRED_READ_FAILURE_GROUPS = C11_DATA.FG1
+ASM2.ASM_PREFERRED_READ_FAILURE_GROUPS = C11_DATA.FG2
```

Zero Downtime Storage Maintenance

Configuring Oracle ASM with 2 or 3 storage subsystems allows taking all storage subsystems except one offline without the need to stop your SAP system.

Before you shutdown a storage system the disks in that storage subsystem should be taken offline in Oracle ASM. When the storage maintenance is finished the Oracle ASM disks can be taken online again and Oracle ASM will resynchronize the disks.

Before the maintenance window:

```
alter diskgroup DATA offline disks in failgroup FG2;  
alter diskgroup RECO offline disks in failgroup FG2;  
alter diskgroup ARCH offline disks in failgroup FG2;  
alter diskgroup ACFS offline disks in failgroup FG2;  
alter diskgroup VOTE offline disks in failgroup FG2;  
alter diskgroup OCR1 offline disks in failgroup FG2;
```

After the maintenance window:

```
alter diskgroup DATA online all;  
alter diskgroup RECO online all;  
alter diskgroup ARCH online all;  
alter diskgroup ACFS online all;  
alter diskgroup VOTE online all;  
alter diskgroup OCR1 online all;
```

Note that if the contents of FG2 have been overwritten or deleted, or if the ASM disks have been offline longer than 'disk_repair_time' the ASM disks of the specific failure group – FG2 in our example – cannot be onlined with 'alter diskgroup <dg> online all'. Instead the ASM disks must be wiped and added to the ASM diskgroup (as a member of FG2) like a new disk.

"Split Mirror" based Online Backups and Database Copies

Oracle ASM allows the implementation of very powerful strategies for database backup/restore and database copies on storage side. Using these features requires an intelligent storage subsystem that allows to **atomically** clone all disks of an Oracle ASM failure group and the ability to set all disks of a specific failure group to a **'not ready'** state where the disks can be seen but not accessed by the cluster nodes. The later condition is important to prevent cluster nodes from mounting or accessing the affected disks while an image copy is in progress e.g. during restore from a set of backup disks and to ensure that Oracle ASM uses the correct mirror during first mount of a newly restored diskgroup.

Atomically cloning or splitting disks means that all disks are cloned/split at the same point in time.

You cannot use the storage based techniques described below if your storage subsystem does not offer these functionalities.

It is strictly recommended to configure 4(5) Oracle ASM diskgroups (DATA, ARCH, OLOG, MLOG, (RECO)) per database as described in the whitepaper '*SAP Databases on Oracle Automatic Storage Management 11g Release 2 - Configuration Guidelines for UNIX and Linux Platforms*' under '*Variant 3 – very large data and data change volumes, restore time crucial*'. This configuration guarantees that all diskgroups that belong to a specific database are independent from diskgroups that belong to other databases.

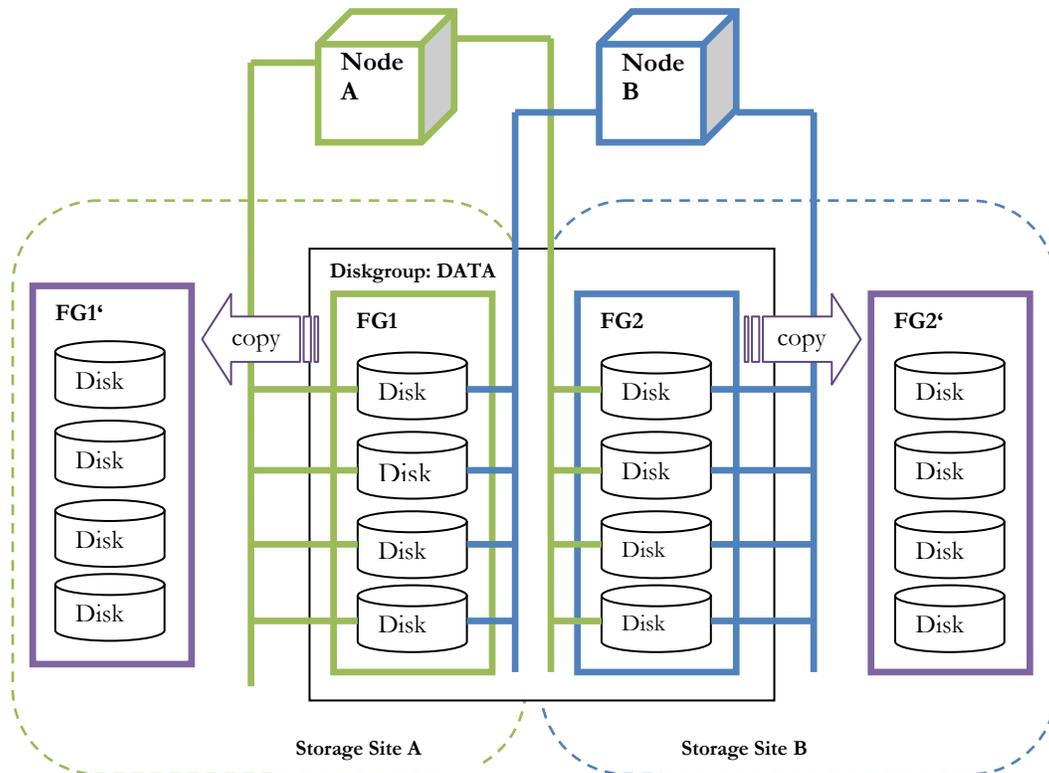
Note that the " " behind the name of a Oracle ASM diskgroup or failure group is used to distinguish Oracle ASM diskgroups and failure groups from their clones and is not part of the diskgroup's or failure group's name. E.g. +DATA and +DATA' or FG1 and FG1'.

Storage based Online Backups

Storage based online backups require at least one additional set of disks where the storage subsystem will copy the contents of your production system's disks to (e.g. EMC BCV's). The storage based backup procedure will then set the database into backup mode and atomically copy all the contents from the production system's disks to the backup disks. The way how this is achieved depends on the type of the storage subsystem. Some storage subsystems copy the contents one time and then just refresh the backup disks on all subsequent backups. This kind of 'storage subsystem based online backup' also allows restoring and recovering a production system very quickly even if your database is several terabyte in size. In addition of just copying the data to some kind of backup disks this technology allows to mount the backup disks on another server (e.g. a dedicated single node backup server) where you could backup the database to tape. This reduces the impact of database backups on your production system to nearly zero.

It is important to mention that only one side of an Oracle ASM mirror needs to be copied. Therefore all disks of a mirror side should belong to the same Oracle ASM failure group (e.g. 'FG1'). Usually only disks that belong to diskgroup '+DATA' need to be split as archivelog files will be backed up by your traditional backup solution (e.g. BR*Backup / Backint).

The example below shows a system with 2 RAC nodes and 2 storage subsystems at different sites. Each storage subsystem has a second set of disks (FG1' and FG2') that are configured as backup disks (e.g. EMC BCV's). One day the disks of FG1' are overwritten, the next day the disks of FG2' are overwritten.



Remote Online Backup to Backup Disks

In order to perform an online backup to backup disks it is necessary to:

- Backup the controlfile.
- Set the database into backup mode.
- Rebuild / Refresh the backup disks (e.g. copy FG1 to FG1').
- End database backup mode.
- Archive all current online redolog files and backup the archivelog files.

Note that this kind of backup is supported by SAP BR*Tools in combination with a dedicated backup server where customers have to use their own scripts that implement the necessary actions on Oracle ASM and storage subsystem side while BR*Tools will call those scripts when required. This way SAP BR*Tools (BRBACKUP) controls the flow of the backup but the specific steps to actually perform the required tasks (e.g. refreshing the clone) have to be implemented by the customer.

Refer to SAP notes #1627541 and #1628116 for a detailed description of how to configure BR*Tools for remote online split mirror backups.

The backup server (the server where the backup disks are mounted by Oracle ASM and where the backup to tape is run) is usually a single server (no cluster) with much less RAM and CPU than the production system. Therefore the customer needs to prepare a pfile and/or spfile that allows to mount the database on the backup server in order to be able to run the backup. A good approach is to copy the spfile from the production system (on Oracle ASM) to the backup server's local Oracle RDBMS Home (e.g. \$ORACLE_HOME/dbs or %ORACLE_HOME%\database) and to make the following adjustments:

- Reduce all memory specific parameters (SGA, PGA,...)
- cluster_database=FALSE
- Remove all controlfile locations except the one on your DATA diskgroup
- Remove or adjust all other entries that refer to the production system (e.g. remote_listener, dataguard specific entries,...)

As most customers will not create all scripts to control and run this kind of backups on their own we will focus on BR*Tools to control the flow of the backup and some customer specific scripts required for Oracle ASM and storage subsystem.

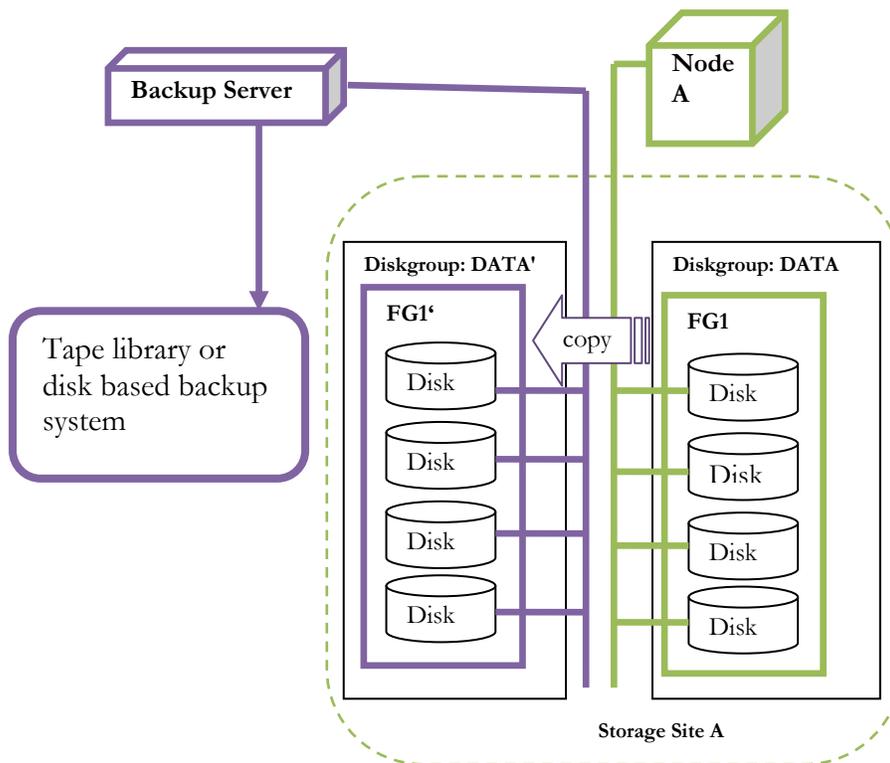
Implementation example

Assumptions

- You want to run the backup at site A and utilize storage subsystem A that contains all disks of failgroup FG1 and the backup disks FG1' of diskgroup +DATA.
- You have a standalone single instance Oracle ASM installation with just 2 Oracle ASM diskgroups (+OCR and +DATA) on your backup server where Oracle Grid Infrastructure is installed under the 'oracle' user and in the same path than on your production servers.
- You have installed an Oracle RDBMS Home in the same path as on your production servers but on a local filesystem.

- You have an Oracle ASM diskgroup +DATA with 2 failure groups FG1 + FG2 on production side and a set of backup disks FG1' for copies of FG1 that are attached to the backup server.
- Diskgroup +DATA on production side is containing all datafiles of the database you want to backup.
- The database instance on the backup server has been shutdown

The figure below shows storage subsystem A with the backup server and cluster node A attached.



- Prepare a database instance with the same name than on your production system and an spfile that allows you to startup your database on backup side into mount state.
 - Reduce all memory specific parameters (SGA, PGA,...)
 - cluster_database=FALSE
 - Remove all controlfile locations except the one on your DATA diskgroup
 - Remove or adjust all other entries that refer to the production system (e.g. remote_listener, dataguard specific entries,...)
- Modify tnsnames.ora (under your Oracle RDBMS Home!) so that the backup server can connect to the local database instance and to one of the RAC instances

on the production system.

```
E.g.
#backupserver
C111 =
  (DESCRIPTION =
    (ADDRESS = (PROTOCOL = TCP) (HOST =
oracx3v4.oracledev.wdf.sap.corp) (PORT = 1521))
    (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = C111)
    )
  )
)
#production system node A
C111_PRIMARY =
  (DESCRIPTION =
    (ADDRESS = (PROTOCOL = TCP) (HOST =
oracx3v1.oracledev.wdf.sap.corp) (PORT = 1521))
    (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = C111)
    )
  )
)
```

- Create a SQL script that dismounts diskgroup '+DATA'. Note that the script must be run on the Oracle ASM instance of your backup server. Therefore ORACLE_HOME and ORACLE_SID must be set accordingly.

```
E.g.
connect / as sysasm
alter diskgroup DATA dismount force;
exit;
```

- Create a script that atomically clones the disks of diskgroup DATA failgroup FG1 to the disks of diskgroup DATA' failgroup FG1' in your storage subsystem.
- Create a SQL script that mounts diskgroup '+DATA' after cloning. Note that the script must run on the Oracle ASM instance of your backup server. Therefore ORACLE_HOME and ORACLE_SID must be set accordingly.

```
E.g.
connect / as sysasm
alter diskgroup DATA mount force;
exit;
```

The 'mount force' option is required to mount the diskgroup with just the disks in failure group FG1' as the disks of FG2 are missing.

- Prepare an init<SID>.sap for remote online split backup with BRBACKUP.
- Run BRBACKUP to perform the backup.

Restoring a database or parts of a database from tape to Oracle ASM can be performed by BR*Tools.

Restore and Recovery from Backup Disks

Using remote online backups with backup disks does not only lower the impact of online backups to the production system. It also allows to quickly restore and recover your production database in case of an error by copying the disk images from the backup disks to the production disks using fast storage subsystem based techniques.

Complete Restore and Recovery

If the complete diskgroup DATA is damaged or a larger number of datafiles is lost or corrupt it can be faster to restore the complete database from backup disks instead of restoring only the corrupt or lost parts from tape.

The steps below outline the path to restore and recover your database using this technique.

- Shutdown all database instances that access diskgroup DATA and check the status.
E.g.
Set ORACLE_HOME and ORACLE_SID to your ASM environment.

```
srvctl stop database -d C11 -o immediate
srvctl status resource -t
```
- Copy your controlfile from a diskgroup that is still intact to a local directory.
E.g.
Set ORACLE_HOME and ORACLE_SID to your RDBMS environment and start RMAN.

```
RMAN> connect target /
RMAN> startup nomount;
RMAN> restore controlfile to 'C:\temp\cntrlc11.dbf' from
'+ARCH/C11/cntrlc11.dbf';
RMAN> shutdown;
```
- Set 'asm_power_limit=0' to prevent Oracle ASM from rebalancing the contents of diskgroup DATA failgroup FG1 until we have finished restoring the complete failure group FG1 from FG1'.
E.g.
Set ORACLE_HOME and ORACLE_SID to your ASM environment and start SQLPLUS.

```
sqlplus / as sysasm
SQL> alter system set asm_power_limit=0 scope=memory;
```
- Dismount the diskgroup **on all** cluster nodes.
E.g.
Set ORACLE_HOME and ORACLE_SID to your ASM environment and start SQLPLUS.

```
sqlplus / as sysasm
alter diskgroup DATA dismount force;
```
- Set **all** disks of diskgroup DATA to 'not ready' state (all disks of FG1,FG2,...).

- Clone all disks of diskgroup DATA failure group FG1' to diskgroup DATA failure group FG1.
- Set all disks of diskgroup DATA failure group FG1 to 'ready' state (NOT the disks of FG2,...).
- Mount diskgroup DATA on all cluster nodes using the 'force' option.
E.g.
Set ORACLE_HOME and ORACLE_SID to your ASM environment and start SQLPLUS.

```
sqlplus / as sysasm  
alter diskgroup DATA mount force;
```
- Startup one database instance into 'nomount' state and restore the controlfile from your local directory.
E.g.
Copy your controlfile from a diskgroup that is still intact to a local directory.
E.g.
Set ORACLE_HOME and ORACLE_SID to your RDBMS environment and start RMAN.

```
RMAN> connect target /  
RMAN> startup nomount;  
RMAN> restore controlfile from 'C:\temp\cntrlc11.dbf';  
RMAN> shutdown;
```
- Startup one database instance into 'mount' state and start recovery (auto mode). This assumes that every archive log file needed is still on diskgroup ARCH. If this is not the case the missing archives must be restored first.
E.g.
Set ORACLE_HOME and ORACLE_SID to your RDBMS environment and start SQLPLUS.

```
SQL> startup mount;  
SQL> recover database;
```
- Open the database after successful complete recovery and shut it down again.
E.g.
Set ORACLE_HOME and ORACLE_SID to your RDBMS environment and start SQLPLUS.

```
SQL> alter database open;  
SQL> shutdown immediate;
```
- Start all database services with the instances they depend on.
E.g.
Set ORACLE_HOME and ORACLE_SID to your ASM environment.

```
srvctl start service -d C11
```
- Check if all database instances and services are up and running. Your system is available again but with reduced redundancy.
- Set all the remaining disks of diskgroup DATA to 'ready' state (FG2,...).
- Set 'asm_power_limit' to a value greater than zero (e.g. 250) to allow Oracle ASM to resynchronize the contents of diskgroup DATA failgroup FG2 with the contents of FG1 when the disks of FG2 are taken online.
E.g.

```
Set ORACLE_HOME and ORACLE_SID to your ASM environment and
start SQLPLUS.
```

```
sqlplus / as sysasm
```

```
SQL> alter system set asm_power_limit=250 scope=memory;
```

- Start resynchronization by setting the remaining disks (FG2,...) online.

E.g.

```
Set ORACLE_HOME and ORACLE_SID to your ASM environment and
start SQLPLUS.
```

```
sqlplus / as sysasm
```

```
alter diskgroup DATA ONLINE ALL;
```

- Check V\$ASM_OPERATION and V\$ASM_DISK for information about the progress of resynchronization. If V\$ASM_OPERATION is empty Oracle ASM has finished resynchronization.

Storage based Database Copies

Customers that separated their databases to different diskgroup sets (e.g. <DBNAME>_DATA, <DBNAME>_ARCH, ...) can use storage based cloning or image copying facilities to quickly create copies of those databases. This method usually works much faster than other strategies especially for large databases. The technique is basically the same than with remote online split backups but is not fully supported by BR*Tools. In this case BRBACKUP just initiates the split by issuing 'begin and end backup' to the source database and by calling the clone or split scripts provided by the customer. The system change number (SCN) of each thread is written to the logfile and required to find out the archive logfiles needed to perform a manual recovery before the copied database can be opened.

Note that in this mode brbackup does not backup any files. This mode is also called 'remote online split backup in query mode' which means that brbackup will execute all the steps that belong to a 'remote online split backup' but omit backing up any files to tape. This type of operation is usually run on the production side.

In the example below the steps needed to copy a production database from a 3 node RAC cluster (PROD) to a 2 node RAC quality assurance cluster (QAS) and to rename the production database 'XWP' to 'XWK' are outlined.

Assumptions:

You already have installed Oracle Grid Infrastructure for RAC on XWK and you have configured all the Oracle ASM diskgroups required for the database you are going to copy. Although the names of the systems and the names of the database will be different (XWP, XWK) the name of the diskgroups will be the same. In other words: You should use the same names for your Oracle ASM diskgroup on XWK as you have on XWP because storage based cloning will also take over the names of the Oracle ASM diskgroups. To simplify the example we have omitted the 3rd voting file and we will use diskgroup name 'DATA' as synonym for diskgroup name 'XWP_DATA'.

E.g.

SAP Systemname: XWP	SAP Systemname: XWK
Database name: XWP	Database name: XWK
Diskgroup names: XWP_DATA, XWP_ARCH, XWP_OLOG, XWP_MLOG, (XWP_RECO)	Diskgroup names: XWP_DATA, XWP_ARCH, XWP_OLOG, XWP_MLOG, (XWP_RECO)
OHRDBMS: /oracle/XWP/11202 or C:\ORACLE\XWP\11202 on MS Windows.	OHRDBMS: /oracle/XWK/11202 or C:\ORACLE\XWK\11202 on MS Windows.

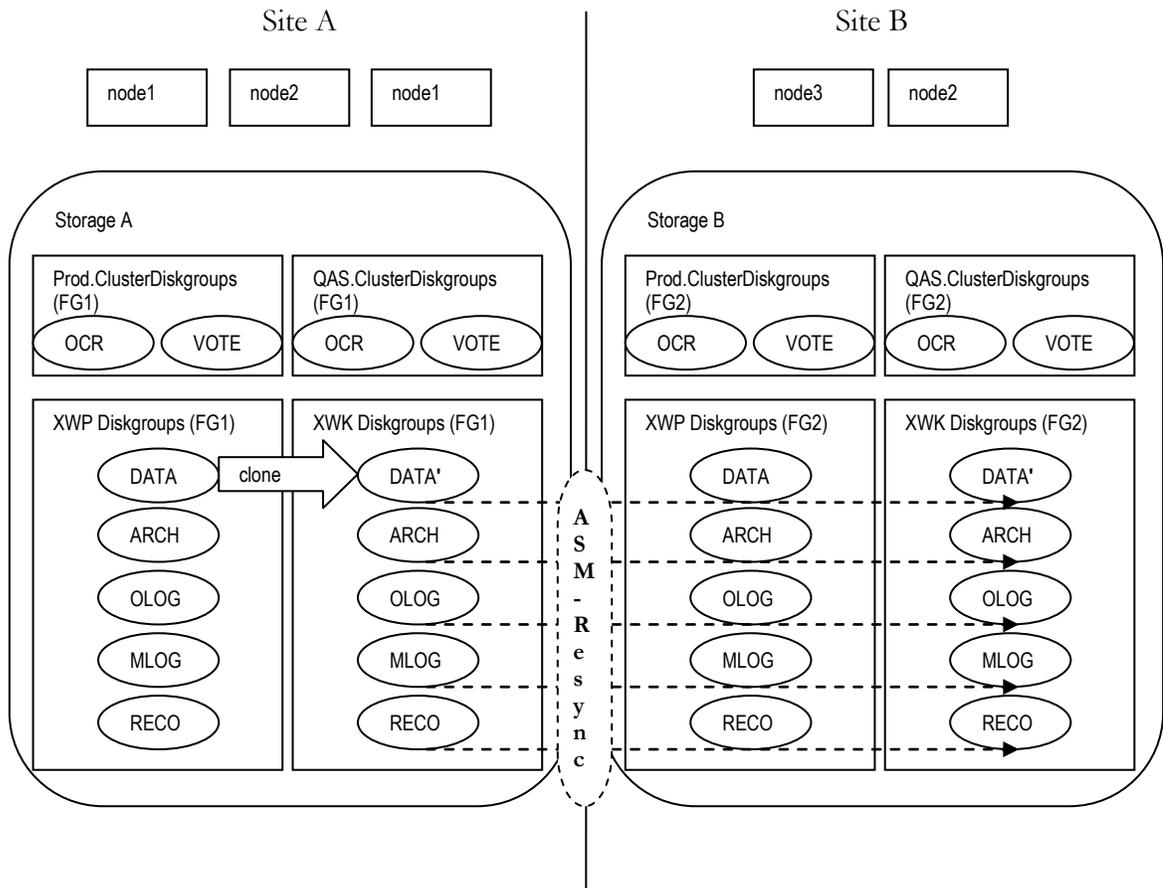
Note that only diskgroup DATA failure group FG1 of XWP has to be cloned to diskgroup DATA' failure group FG1 of XWK. The archive logfiles that are required to bring the database XWK to a consistent state must be copied and applied manually.

Also note that having all the diskgroups we have on PROD on QAS is not a mandatory and fully depends customer's requirements. Instead of creating the same diskgroups you could decide to create just diskgroup DATA' failure group FG1 with the same number of disks and size than diskgroup DATA and to place a new controlfile and spfile on diskgroup DATA' too after cloning.

In our example the XWK diskgroup's failure group's FG2 are completely optional. If you decide to run your QAS system with just one failure group FG1 per diskgroup then you just need to set 'asm_power_limit' to zero and keep that value as your default by storing it in the spfile of your Oracle ASM instances. This allows forcibly mount diskgroups even if failure groups FG2 are missing and prevents Oracle ASM from trying to restore redundancy within failure groups FG1 as this would quickly exhaust the free space.

If you decide that XWK diskgroups should have two failure groups FG1 and FG2 like XWP on PROD note that only DATA FG1 needs to be cloned to DATA' FG1 by the storage subsystem. DATA' FG2 will be rebuild by Oracle ASM when all the disks belonging to FG2 will be set online **after** the diskgroup was forcibly mounted just with FG1.

Never clone other diskgroups than DATA in online mode.



- On QAS
 - Make sure that no database instance on QAS is accessing the diskgroups belonging to XWK.
 - If there are already any files on the Oracle ASM diskgroups that belong to XWK remove them.
 - Forcibly dismount diskgroup DATA' on all cluster nodes (QAS)
E.g.
Set ORACLE_HOME and ORACLE_SID to your ASM environment and start SQLPLUS.
sqlplus / as sysasm
alter diskgroup XWP_DATA dismount force;
 - Set all Oracle ASM disks of diskgroup DATA' to 'not ready' state.
- On PROD
 - Create a 'create controlfile script' by backing up the controlfile of XWP to trace. This is required later when we create new controlfiles for XWK on QAS.
E.g.

```
Alter database backup controlfile to trace as
'c:\temp\xwk_cntrl.sql';
```

- Make sure that all current online redologs have been archived.

E.g.

```
alter system archive log current;
```

- Run brbackup to clone DATA FG1 to DATA' FG1

E.g.

```
brbackup.exe -p initclonexwptoxwk.sap -d tape -t
online_split -m full -q split -k no -e 0 -l E -U
```

- Check log of brbackup. It contains the last SCN of each thread. Write down the lowest SCN of all thread. This SCN is required to find out the names of all archive logs required to recover the database.
- Get a list of all the archive logs and copy them to a local temporary directory. The script below prompts for the SCN from the previous step and creates a batch script that copies the files from Oracle ASM to '/TEMP'.

E.g.

```
set linesize 200
set heading off
set pagesize 100
set echo off
set feedback off
set verify off
define scn=&&firstarchivescn
define sid=&&sid
spool ksysteme_copyarchives.cmd
select 'call env_asm.cmd' from dual;
select 'call asmcmd --privilege sysasm -p cp '||name||'
/temp/'||'&sid'||'_'||substr(name,instr(name,'/',1,4)+1)
from v$archived_log where FIRST_CHANGE# >= &scn;
spool off
```

- Run the generated script to copy the archived redologs to '/TEMP'.

- On QAS

- Set all Oracle ASM disks belonging to DATA' FG1 to 'ready' state. Do **not** include Oracle ASM disks of DATA' FG2 at this point.

E.g.

```
alter diskgroup XWP_DATA mount force;
```

Note: If you do **not** want a second failure group FG2 on your QAS system you would drop all Oracle ASM disks belonging to FG2 at this point. This does not change the Oracle ASM diskgroup's redundancy level but removes the missing disks from the diskgroup. Also turn off rebalancing for that diskgroup.

E.g.

```
alter diskgroup XWP_DATA drop disks in failgroup FG2
force;
```

```
alter diskgroup XWP_DATA rebalance power 0;
```

- Copy the 'create controlfile script' together with the required archive logfiles from PROD ('/temp') to QAS.
- Prepare an pfile that allows you startup a database instance of XWK into 'nomount' state (and later into 'mount' state) and modify your 'create controlfile script' as needed to start the database instance using the pfile and to create a new controlfile that matches for XWK.

E.g. Change the following lines as shown in the example below and remove everything that is not required to create the new controlfile.

```
STARTUP NOMOUNT
pfile='c:\oracle\XWK\11202\database\initXWK_fordup.ora'
CREATE CONTROLFILE REUSE SET DATABASE "XWK" RESETLOGS
ARCHIVELOG
```

- If the controlfile creation was successful shutdown the database instance, bring it into 'mount' state and start recovery.

E.g.

```
recover database using backup controlfile until cancel;
```

Apply all the archive logfiles in the appropriate order (seq# and thread#) and open the database when finished using the 'resetlogs' option.

- Disable threads that are not required anymore and remove the associated logfile groups and undo tablespaces.
- Add the tempfiles that are missing after controlfile creation.
- Check the database for invalid objects and recompile them.
- If have configured Oracle ASM diskgroup DATA' FG2 set the disks to 'ready' state and take them online in Oracle ASM.

E.g.

```
alter diskgroup XWP_DATA online all;
```

This will start resynchronization of FG2. You can monitor the progress in V\$ASM_OPERATION.



SAP with Oracle Real Application Clusters 11g
Release 2 and Oracle Automatic Storage
Management 11g Release 2
Advanced Configurations & Techniques

October 2012

Author: Markus Breunig

Contributing Authors: Andreas Becker, Jan
Klokkers, Christoph Kurucz

Oracle Corporation
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



Oracle is committed to developing practices and products that help protect the environment

Copyright © 2011, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark licensed through X/Open Company, Ltd. 0611

Hardware and Software, Engineered to Work Together