Oracle ASM 12c.

A powerful framework for storage management

CONTENTS

HISTORY

Version	Date	Name	Description
0.1	26.05.2015	Emiliano Fusaglia	Document creation
0.2	28.05.2015	Emiliano Fusaglia	Development of new sections
0.3	08.06.2015	Emiliano Fusaglia	Development of new sections
0.4	12.06.2015	Emiliano Fusaglia	Development of new sections
0.9	15.06.2015	Emiliano Fusaglia	Development of new sections
1.0	19.06.2015	Emiliano Fusaglia	Final Version

REFERENCED DOCUMENTS

Name	Description
	Oracle® Automatic Storage Management Administrator's Guide
	Oracle® Automatic Storage Management New Features White Paper

LIST OF ABBREVIATIONS

Sort the abbreviations in alphabetical order.

| Abbreviation	Description
ASM	Automatic Storage Management
ACFS	Automatic Storage Management Cluster File System
ADVM	ASM Dynamic Volume Manager
ODA	Oracle Database Appliance
HANFS	High Availability Network File System

trivadis

makes **IT** easier.

1   INTRODUCTION

Oracle Automatic Storage Management (ASM) is a well-known, largely used multi-platform volume manager and file system, designed for single-instance and clustered environment. Developed for managing Oracle database files with optimal performance and native data protection, simplifying the storage management; nowadays ASM includes several functionalities for general-purpose files too.

This article focuses on the architecture and characteristics of the version 12c, where great changes and enhancements of pre-existing capabilities have been introduced by Oracle.

Dedicated sections explaining how Oracle has leveraged ASM within the Oracle Engineered Systems complete the paper.
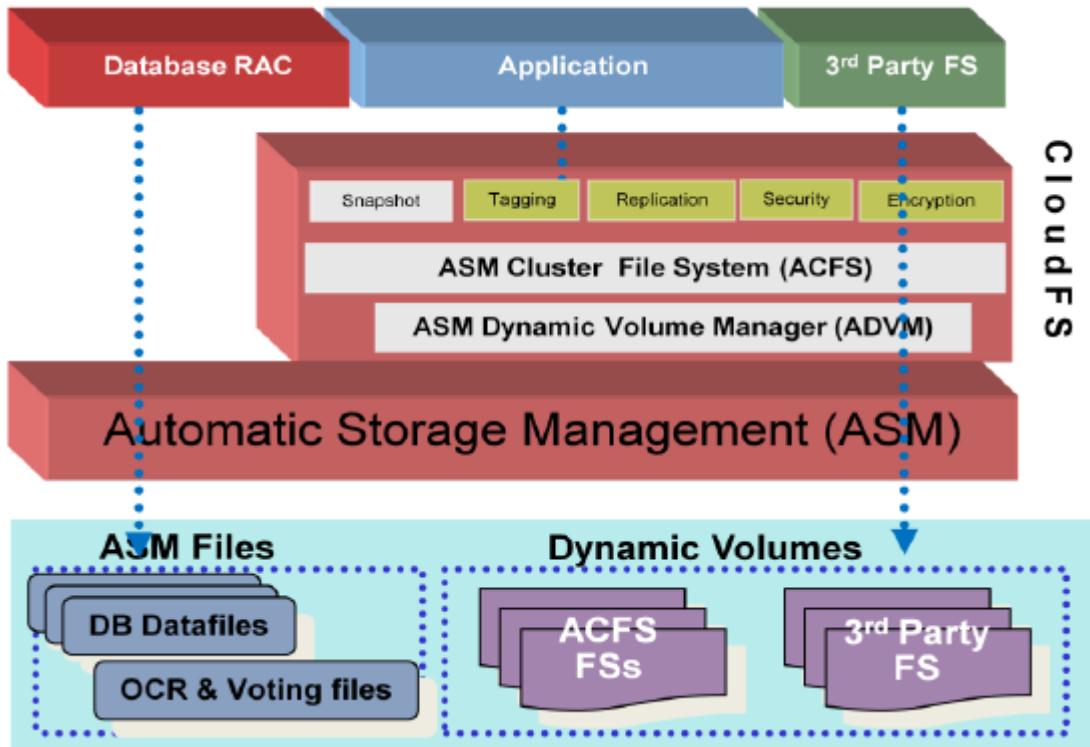
1.1   ASM 12c Instance Architecture Diagram

Below are highlighted the functionalities and the main background components associated to an ASM instance. It is important to notice how starting from Oracle 12c a database can run within ASM Disk Groups or on top of ASM Cluster file systems (ACFS).

Source: Oracle® Automatic Storage Management Administrator's Guide

trivadis

makes **IT** easier.
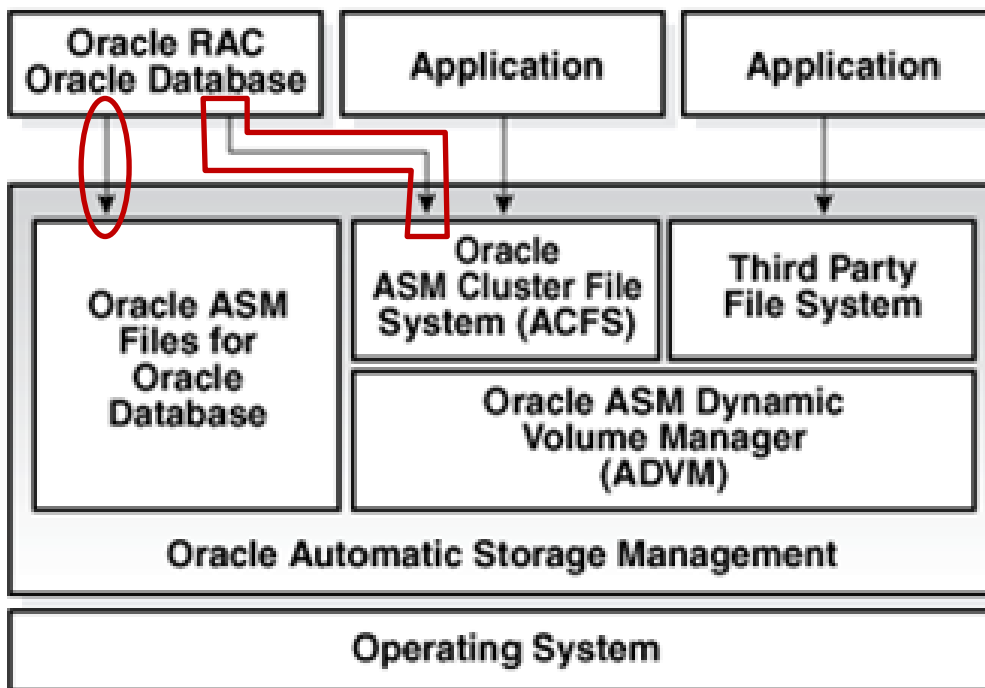
Overview ASM options available in Oracle 12c.

Source: Oracle® Automatic Storage Management New Features White Paper

### 1.2   ASM 12c Multi-Nodes Architecture Diagram

In a Multi-node cluster environment, ASM 12c is now available in two configurations:

- ■ 11gR2 like: with one ASM instance on each Grid Infrastructure node.

- ■ Flex ASM: a new concept, which leverages the architecture availability and performance of the cluster; removing the 1:1 hard dependency between cluster node and local ASM instance. With Flex ASM only few nodes of the cluster run an ASM instance, (the default cardinality is 3) and the database instances communicate with ASM in two possible way: locally or over the ASM Network. In case of failure of one ASM instance, the databases automatically and transparently reconnect to another surviving instance on the cluster. This major architectural change required the introduction of two new cluster resources, ASM-Listener for supporting remote client connections and ADVM-Proxy, which permits the access to the ACFS layer. In case of wide cluster installation, Flex ASM enhances the performance and the scalability of the Grid Infrastructure, reducing the amount of network traffic generated between ASM instances.
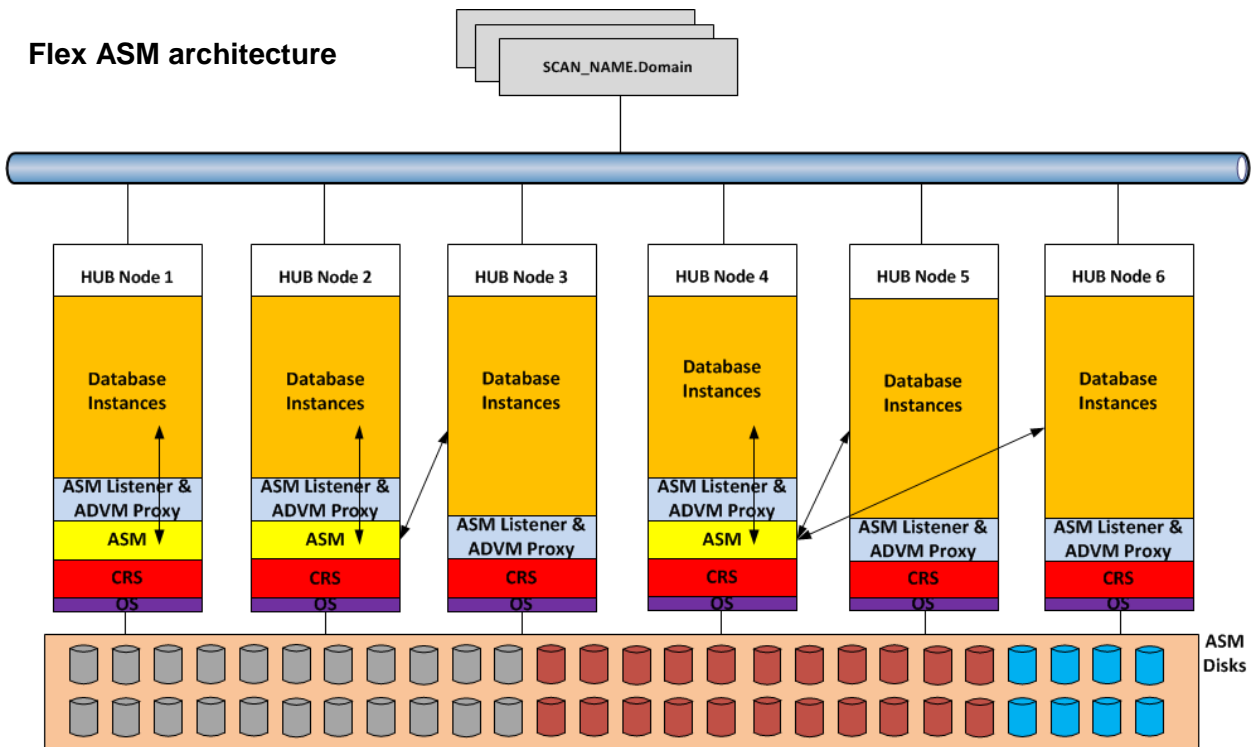
Below two graphical representations of the same Oracle cluster; on the first drawing ASM is configured with pre-12c setup, on the second one Flex ASM is in use.

trivadis
makes **IT** easier.

ASM architecture 11gR2 like

Flex ASM architecture

trivadis
makes IT easier.
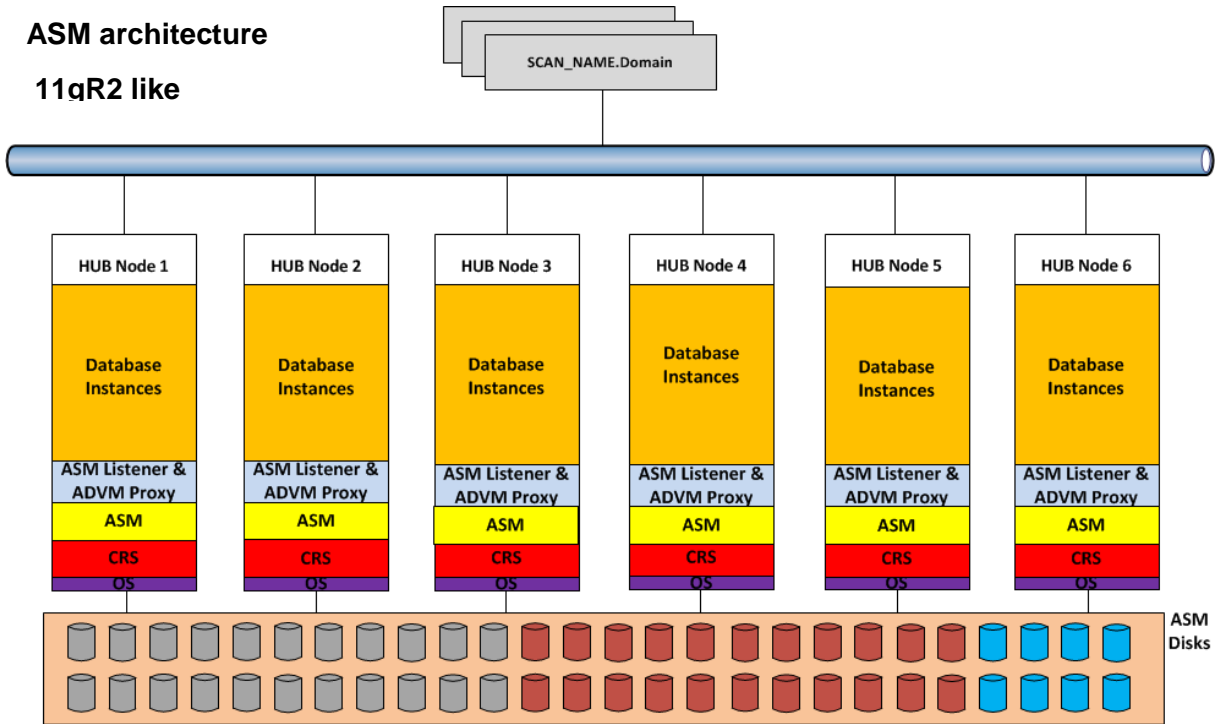
2 ASM 12C NEW FEATURES

The table below summarizes the list of new functionalities introduced on ASM 12c R1:

Feature	Definition
Filter Driver	*Filter Driver (Oracle ASMFD) is a kernel module that resides in the I/O path of the Oracle ASM disks used to validate write I/O requests to Oracle ASM disks, eliminates accidental overwrites of Oracle ASM disks that would cause corruption. For example, the Oracle ASM Filter Driver filters out all non-Oracle I/Os which could cause accidental overwrites.*
General ASM Enhancements	- *Oracle ASM now replicates physically addressed metadata, such as the disk header and allocation tables, within each disk, offering a better protection against bad block disk sectors and external corruptions.* - *Increased storage limits: ASM can manage up to 511 disk groups and a maximum disk size of 32 PB.* - *New REPLACE clause on the ALTER DISKGROUP statement.*
Disk Scrubbing	*Disk scrubbing checks logical data corruptions and repairs the corruptions automatically in normal and high redundancy disks groups. This process automatically starts during rebalance operations or the administrator can trigger it.*
Disk Resync Enhancements	*It enables fast recovery from instance failure and faster resyncs performance. Multiple disks can be brought online simultaneously. Checkpoint functionality enables to resume from the point where the process was interrupted.*
Even Read For Disk Groups	*If ASM mirroring is in use, each I/O request submitted to the system can be satisfied by more than one disk. With this feature, each request to read is sent to the least loaded of the possible source disks.*
ASM Rebalance Enhancements	*The rebalance operation has been improved in term of scalability, performance, and reliability; supporting concurrent operations on multiple disk groups in a single instance. In this version, it has been enhanced also the support for thin provisioning, user-data validation, and error handling.*
ASM Password File in a Disk Group	*ASM Password file is now stored within the ASM disk group.*
Access Control Enhancements on Windows	*It is now possible to use access control to separate roles in Windows environments. With Oracle Database services running as users rather than Local System, the Oracle ASM access control feature is enabled to support role separation on Windows.*
Rolling Migration Framework for ASM One-off Patches	*This feature enhances the rolling migration framework to apply oneoff patches released for ASM in a rolling manner, without affecting the overall availability of the cluster or the database*

trivadis

makes **IT** easier.

Updated Key Management Framework	*This feature updates Oracle key management commands to unify the key management application programming interface (API) layer. The updated key management framework makes interacting with keys in the wallet easier and adds new key metadata that describes how the keys are being used.*

2.1 ASM 12c Client Cluster

One more ASM functionality explored but still in phase of development and therefore not really documented by Oracle, is ASM Client Cluster

Designed to host applications requiring cluster functionalities (monitoring, restart and failover capabilities), without the need to provision local shared storage.

The ASM Client Cluster installation is available as configuration option of the Grid Infrastructure binaries, starting from version 12.1.0.2.1 with Oct. 2014 GI PSU.

The use of ASM Client Cluster imposes the following pre-requisites and limitations:

- The existence of an ASM Server Cluster version 12.1.0.2.1 with Oct. 2014 GI PSU, configured with the GNS server with or without zone delegation.
- The ASM Server Cluster becomes aware of the ASM Client Cluster by importing an ad hoc XML configuration containing all details.
- The ASM Client Cluster uses the OCR, Voting Files and Password File of the ASM Server Cluster.
- ASM Client Cluster communicates with the ASM Server Cluster over the ASM Network.
- ASM Server Cluster provides remote shared storage to ASM Client Cluster.

As already mentioned, at the time of writing this feature is still under development and without official documentation available, the only possible comment is that the ASM Client Cluster looks similar to another option introduced by Oracle 12c and called Flex Cluster. In fact, Flex Cluster has the concept of HUB and LEAF nodes; the first used to run database workload with direct access to the ASM disks and the second used to host applications in HA configuration but without direct access to the ASM disks.

trivadis
makes **IT** easier.

3   ACFS NEW FEATURES

In Oracle 12c the Automatic Storage Management Cluster File System supports more and more types of files, offering advanced functionalities like snapshot, replication, encryption, ACL and tagging. It is also important to highlight that this cluster file system comply with the POSIX standards of Linux/UNIX and with the Windows standards.

Access to ACFS from outside the Grid Infrastructure cluster is granted by NFS protocol; the NFS export can be registered as clusterware resource becoming available from any of the cluster nodes (HANFS).

Here is an exhaustive list of files supported by ACFS: executables, trace files, logs, application reports, BFILEs, configuration files, video, audio, text, images, engineering drawings, general-purpose and Oracle database files.

The major change, introduced in this version of ACFS, is definitely the capability and support to host Oracle database files; granting access to a set of functionalities that in the past were restricted to customer files only. Among them, the most important is the *snapshot image*, which has been fully integrated with the database Multitenant architecture, allowing cloning entire Pluggable databases in few seconds, independently from the size and in space efficient way using *copy-on-write* technology.

The snapshots are created and immediately available in the "<FS_mount_point>.ASFS/snaps" directory, and can be generated and later converted from read-only to read/write and vice versa. In addition, ACFS supports nested snapshots.

Example of ACFS snapshot copy:

```
-- Create a read/write Snapshot copy
[grid@oel6srv02 bin]$ acfsutil snap create -w cloudfs snap /cloudfs

-- Display Snapshot Info
[grid@oel6srv02 ~]$ acfsutil snap info cloudfs snap /cloudfs
snapshot name:              cloudfs snap
RO snapshot or RW snapshot:  RW
parent name:                /cloudfs
snapshot creation time:      Wed May 27 16:54:53 2015


-- Display specific file info
[grid@oel6srv02 ~]$ acfsutil info file /cloudfs/scripts/utl env/NEW SESSION.SQL
/cloudfs/scripts/utl env/NEW SESSION.SQL
    flags:        File
    inode:        42
    owner:        oracle
    group:        oinstall
    size:         684
    allocated:    4096
    hardlinks:    1
    device index: 1
    major, minor: 251,91137
    access time:  Wed May 27 10:34:18 2013
    modify time:  Wed May 27 10:34:18 2013
    change time:  Wed May 27 10:34:18 2013
    extents:
        -offset ----length | -dev --------offset
              0       4096 |    1     1496457216
          extent count: 1
```

trivadis

makes **IT** easier.

```
--Convert the snapshot from Read/Write to Read-only
acfsutil snap convert -r cloudfs snap /cloudfs


--Drop the snapshot
[grid@oel6srv02 ~]$ acfsutil snap delete cloudfs_snap /cloudfs
```

Example of Pluggable database cloned using ACFS snapshot copy

List of requirements that must be met to use ACFS SNAPSHOT COPY clause:

- All pluggable database files of the source PDB must be stored on ACFS.

- The source PDB cannot be in a remote CDB.

- The source PDB must be in read-only mode.

- Dropping the parent PDB with the including *datafiles clause,* does not automatically remove the snapshot dependencies, manual intervention is required.

```
SQL> CREATE PLUGGABLE DATABASE pt02 FROM ppq01
  2   FILE NAME CONVERT = ('/u02/oradata/CDB4/PPQ01/',
  3                        '/u02/oradata/CDB4/PT02/')
  4   SNAPSHOT COPY;

Pluggable database created.

Elapsed: 00:00:13.70
```

The PDB *snapshot copy* imposes few restrictions among which the source database opened in read-only. This requirement prevents the implementation on most of the production environments where the database must remain available in read/write 24h/7. For this reason, ACFS for database files is particularly recommended on test and development where flexibility, speed and space efficiency of the clones are key factors for achieving high productive environment.

Graphical representation of how efficiently create and maintain a Test & Development database environment:
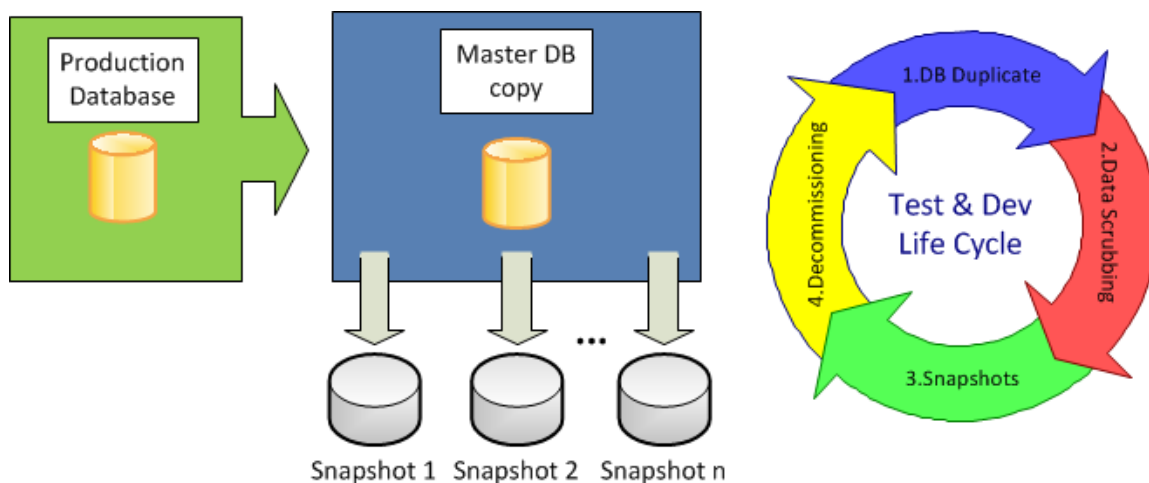
## 4   ASM 12C AND ORACLE ENGINEERED SYSTEMS

Oracle has developed few ASM features to leverage the characteristics of the Engineered Systems. Analyzing the architecture of the Exadata Storage, we see how the unique capabilities of ASM make possible to stripe and mirror data across independent set of disks grouped in different Storage Cells.

The sections below describe the implementation of ASM on the Oracle Database Appliance (ODA) and Exadata systems.

4.1   ASM 12c on Oracle Database Appliance

Oracle Database Appliance is a simple, reliable and affordable system engineered for running database workloads. One of the key characteristics present since the first version is the pay-as-you-grow model; it permits to activate a crescendo number of CPU-cores when needed, optimizing the licensing cost. With the new version of the ODA software bundle, Oracle has introduced the configuration *Solution-in-a-box*; which includes the virtualization layer for hosting Oracle databases and application components on the same appliance, but on separate virtual machines. The next sections highlight how the two configurations are architected and the role played by ASM:

- **ODA Bare metal:** available since version one of the appliance, this is still the default configuration proposed by Oracle. Beyond the automated installation process, it is like any other two-node cluster, with all ASM and ACFS features available.

trivadis
makes **IT** easier.
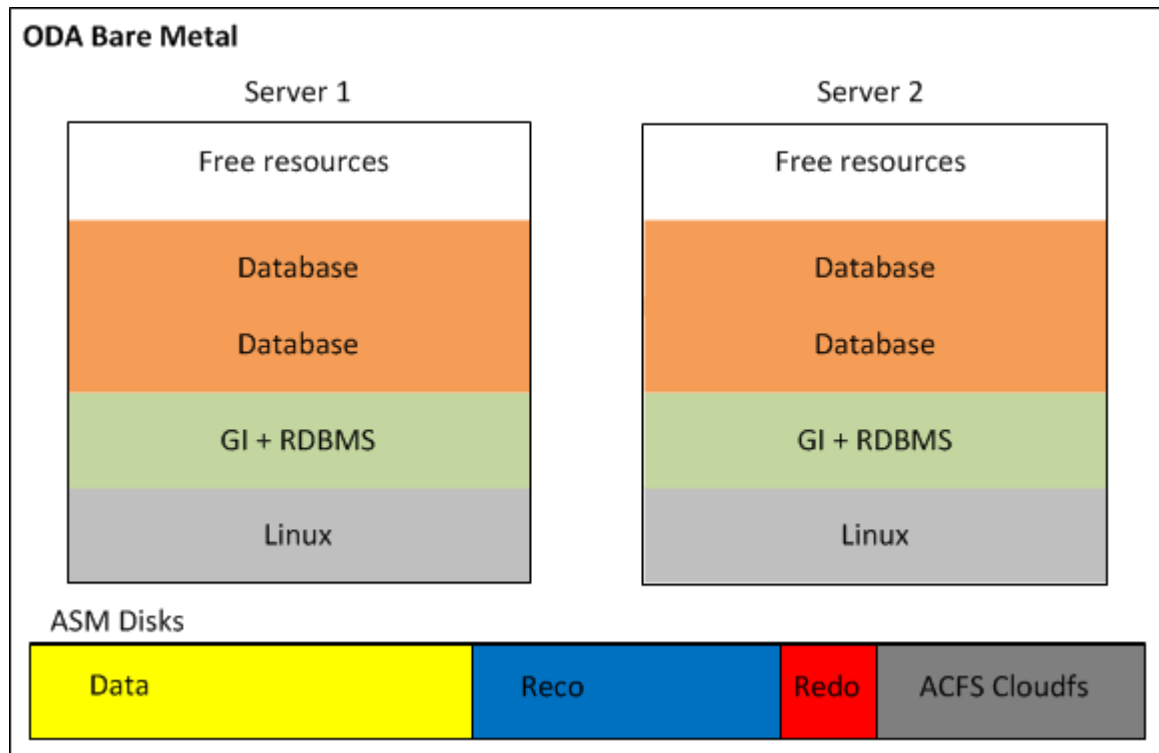
■ **ODA Virtualized:** on both ODA servers runs the Oracle VM Server software, also called Dom0. Each Dom0 hosts the ODA Base (or Dom Base), a privileged virtual machine where it is installed the Appliance Manager, Grid Infrastructure and RDBMS binaries. The ODA Base takes advantage of the Xen PCI Pass-through technology to provide direct access to the ODA shared disks presented and managed by ASM. This configuration reduces the VM flexibility; in fact, no VM migration is allowed, but it guarantees almost no I/O penalty in term of performance. After the Dom Base creation, it is possible to add Virtual Machine where running application components. Those optional application virtual machines are also identified with the name of Domain U.

By default, all VMs and templates are stored on a local Oracle VM Server repository, but in order to be able to migrate application virtual machines between the two Oracle VM Servers a shared repository on the ACFS file system should be created.

The implementation of the *Solution-in-a-box* guarantees the maximum Return on Investment of the ODA, because while licensing only the virtual CPUs allocated to Dom Base, the remaining resources are assigned to the application components as showed on the picture below.
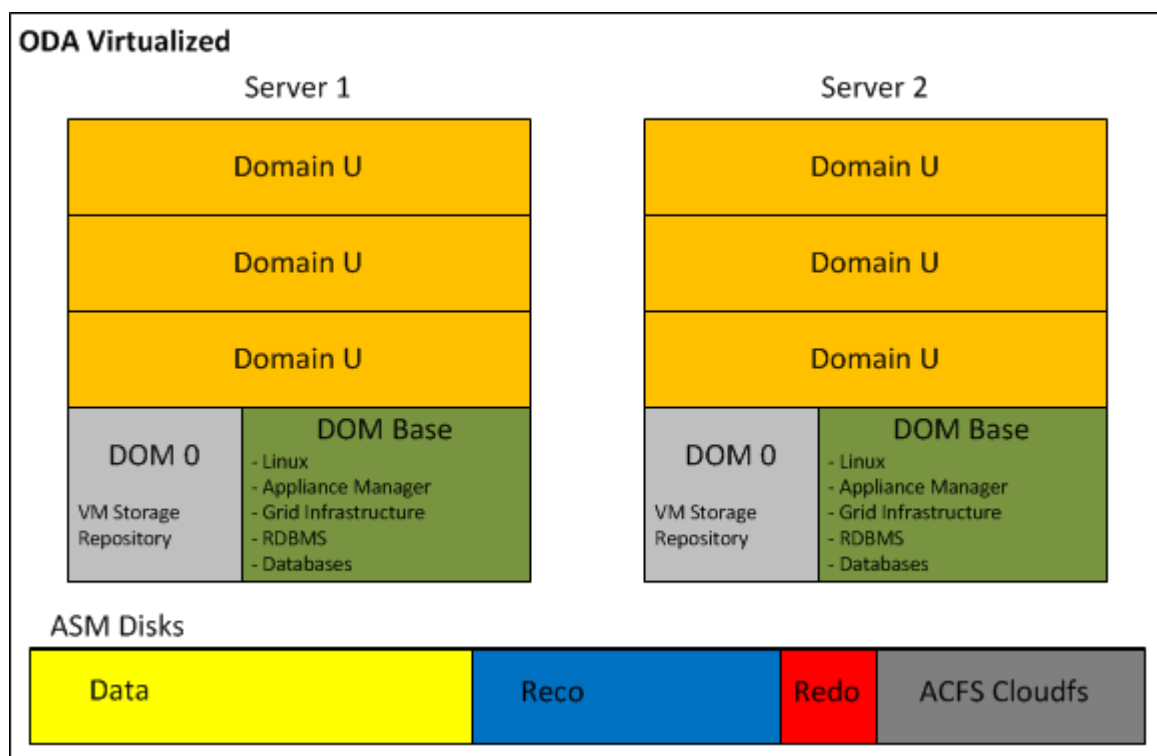
4.2 ACFS becomes the default database storage of ODA

Starting from Version 12.1.0.2, a fresh installation of the Oracle Database Appliance adopts ACFS as primary cluster file system to store database files and general-purpose data. Three file systems are

trivadis
makes **IT** easier.

created in the ASM disk groups (DATA, RECO, and REDO) and the new databases are stored in these three ACFS file systems instead of in the ASM disk groups.

In case of ODA upgrade from previous release to 12.1.0.2, all pre-existing databases are not automatically migrated to ACFS; but can coexist with the new databases created on ACFS.
At any time, the databases can be migrated from ASM to ACFS as post upgrade step.

Oracle has decided to promote ACFS as default database storage on ODA environment for the following reasons:

■ ACFS provides almost equivalent performance than Oracle ASM disk groups.

■ Additional functionalities on industry standard POSIX file system.

■ Database snapshot copy of PDBs, and NON-CDB version 11.2.0.4 of greater.

■ Advanced functionality for general-purpose files such as replication, tagging, encryption, security, and auditing.

Database created on ACFS follows the same Oracle Managed Files (OMF) standard used by ASM.

4.3 ASM 12c on Exadata

Oracle Exadata Database machine is now at the fifth hardware generation; the latest software update has embraced the possibility to run virtual environments, but differently from the ODA or other Engineered System like Oracle Virtual Appliance, the VMs are not intended to host application components. ASM plays a key role on the success of the Exadata, because it orchestrates all Storage Cells in a way that appear as a single entity, while in reality, they do not know and they do not talk to each other.

The Exadata, available in a wide range of hardware configurations from 1/8 to multi-racks, offers a great flexibility on the storage setup too. The sections below illustrate what is possible to achieve in term of storage configuration when the Exadata is exploited bare metal and virtualized:
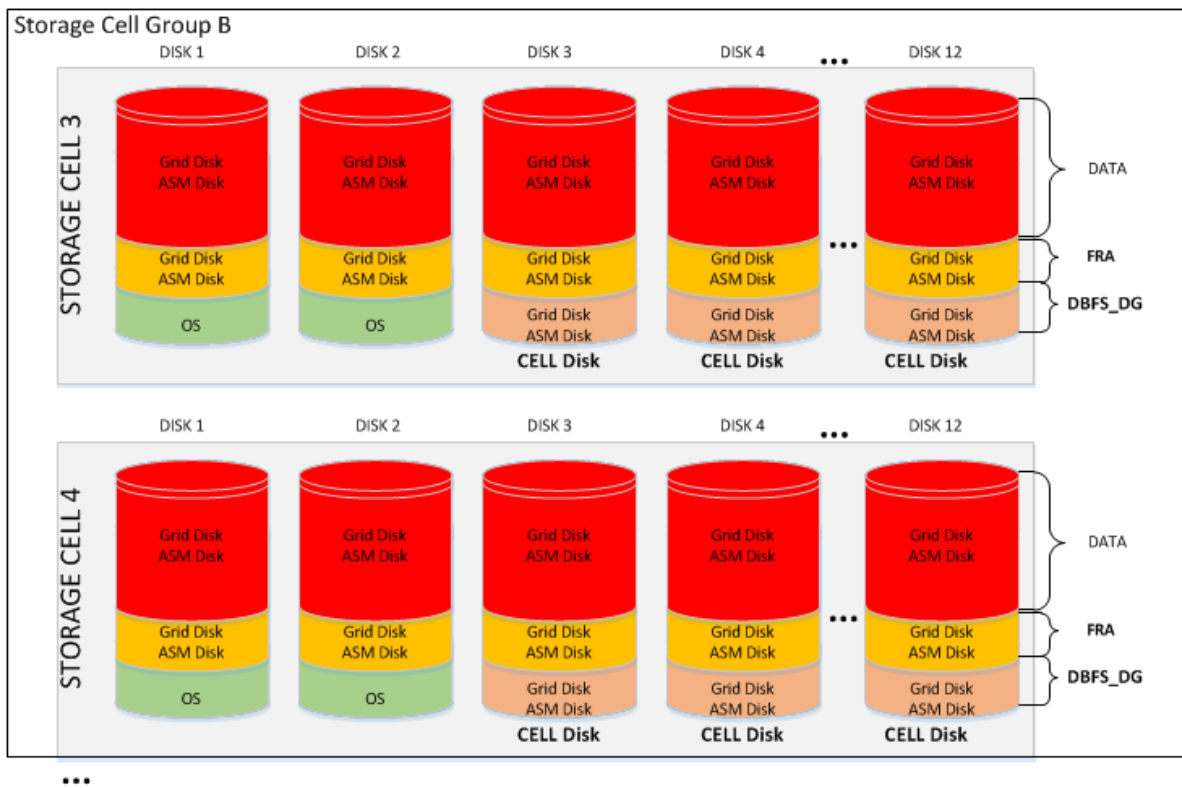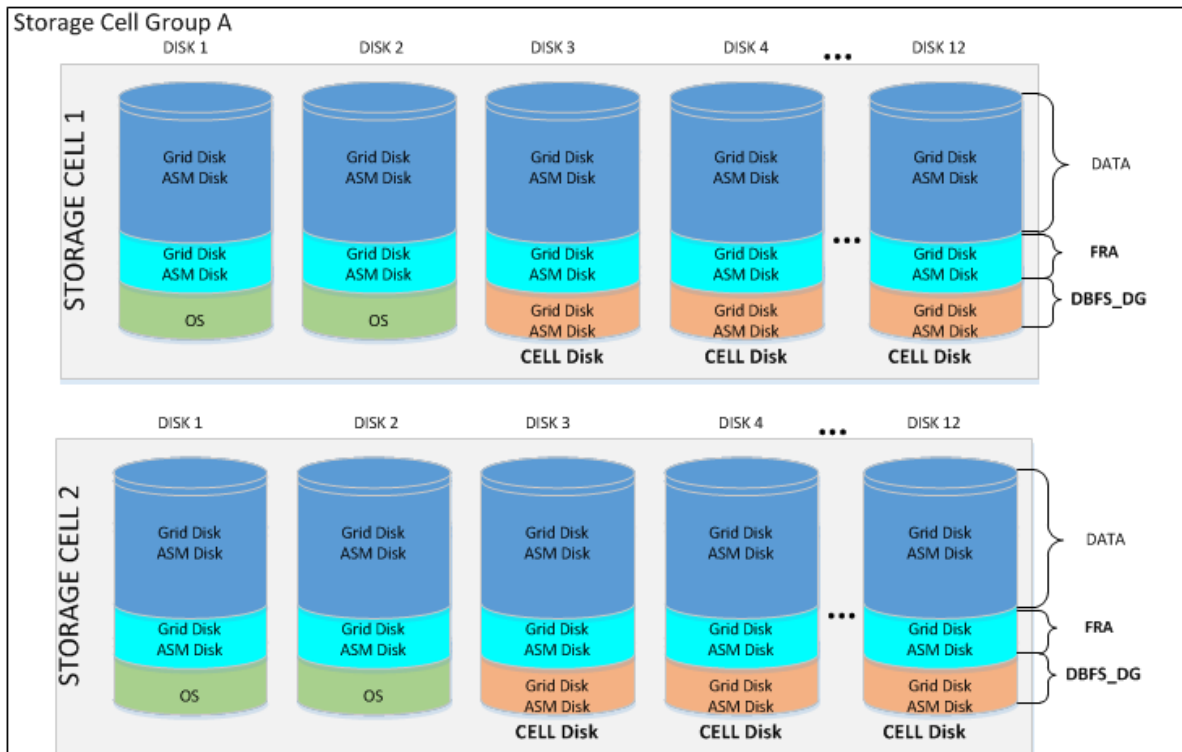
■ **Exadata Bare Metal:** despite the default storage configuration, which foresees three disk groups striped across all Storage Cells, guaranteeing the best I/O performance; as post-installation step, it is possible to deploy a different configuration. Before changing the storage setup, it is vital to understand and evaluate all associated consequences. In fact, even though in specific cases can be a meaningful decision, any storage configuration different from the default one, has as result a shift from optimal performance to flexibility and workload isolation.

Shown below a graphical representation of the default Exadata storage setup, compared to a custom configuration, where the Storage Cells have been divided in multiple groups, segmenting the I/O workloads and avoiding disruption between environments.

trivadis
makes **IT** easier.

Default Storage Cell setup

trivadis
makes IT easier.
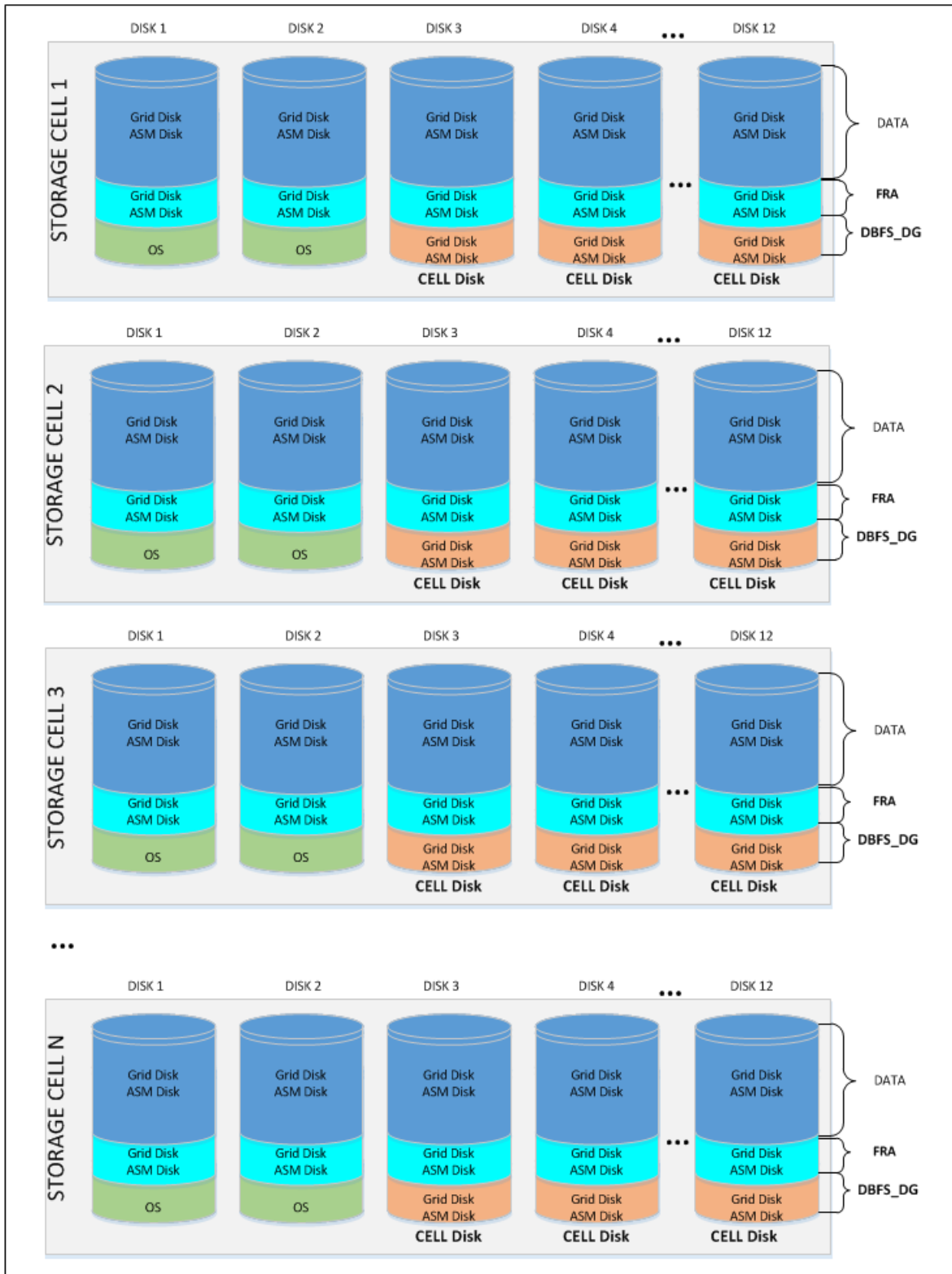
Segmented Storage Cell setup

trivadis
makes **IT** easier.

■ **Exadata Virtualized:** the installation of the Exadata with the virtualization option foresees a first step of meticulous capacity planning, defining the resources to allocate to the virtual machines (CPU and memory) and the size of each ASM disk group (DBFS, Data, Reco) of the clusters. This last step is particularly important, because unlike the VM resources, the characteristics of the ASM disk groups cannot be changed.

The new version of the Exadata Deployment Assistant, which generates the configuration file to submit to the Exadata installation process, now in conjunction with the use of Oracle Virtual Machines, permits to enter the information related to multiple Grid Infrastructure clusters.

The hardware-based I/O virtualization (so called Xen SR-IOV Virtualization), implemented on the Oracle VMs running on the Exadata Database servers, guarantees almost native I/O and Networking performance over InfiniBand; with lower CPU consumption when compared to a Xen Software I/O virtualization. Unfortunately, this performance advantage comes at the detriment of other virtualization features like Load Balancing, Live Migration and VM Save/Restore operations.

If the Exadata combined with the virtualization open new horizon in term of database consolidation and licensing optimization, do not leave any option to the storage configuration. In fact, the only possible user definition is the amount of space to allocate to each disk group; with this information, the installation procedure defines the size of the Grid Disks on all available Storage Cells.

Following a graphical representation of the Exadata Storage Cells, partitioned for holding three virtualized clusters. For each cluster, ASM access is automatically restricted to the associated Grid Disks.

trivadis

makes **IT** easier.

Storage Cell Virtualized setup

trivadis
makes **IT** easier.
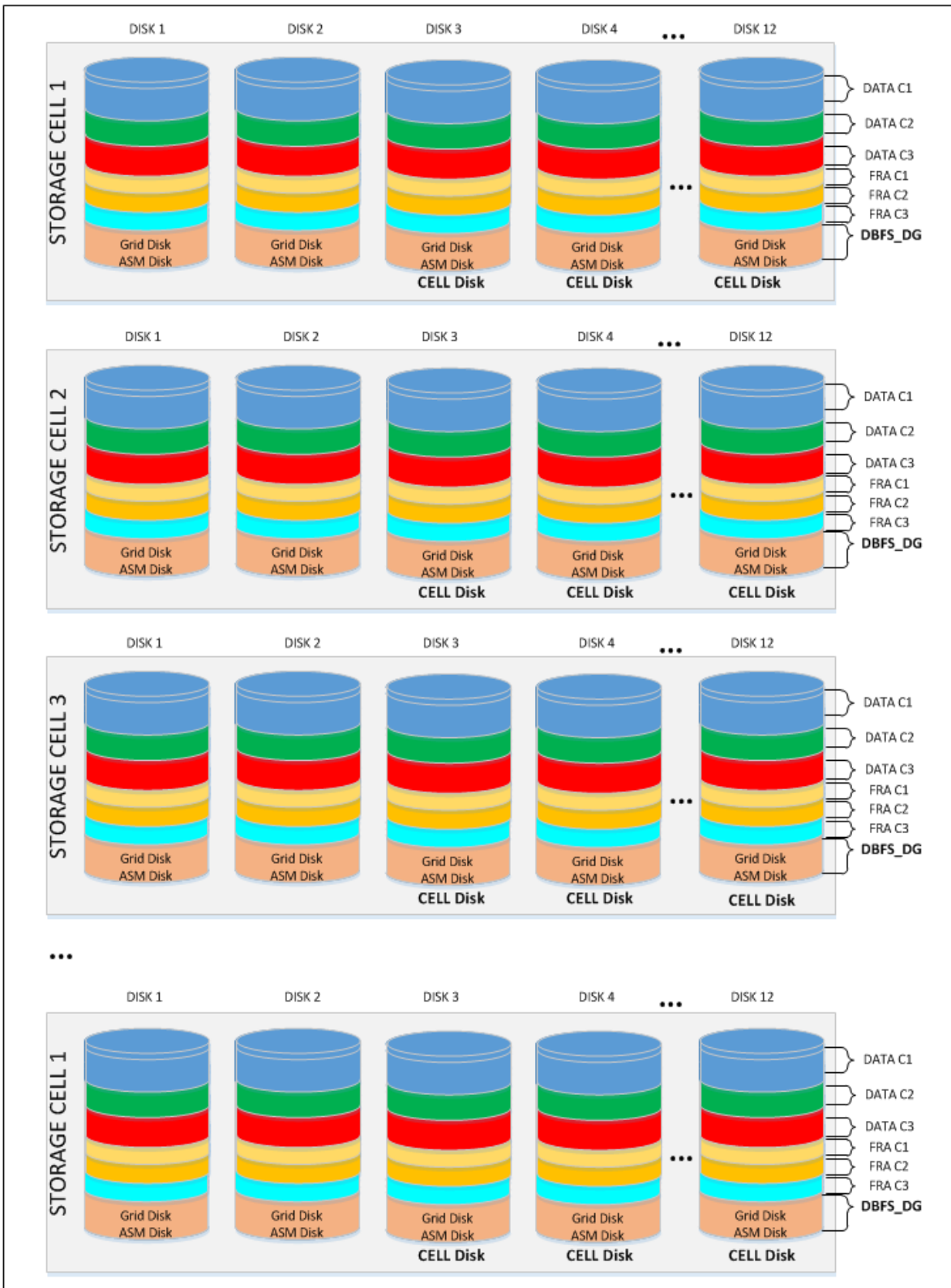
4.4   ACFS on Linux Exadata Database Machines

Starting from version 12.1.0.2, the Exadata Database Machine running Oracle Linux, supports ACFS for database file and general-purpose, with no functional restriction.

This makes ACFS an attractive storage alternative for holding: external tables, data loads, scripts and general-purpose files.

In addition, Oracle ACFS on Exadata Database Machines supports database files for the following database versions:

- Oracle Database 10g Rel. 2 (10.2.0.4 and 10.2.0.5)

- Oracle Database 11g (11.2.0.4 and higher)

- Oracle Database 12c (12.1.0.1 and higher)

Since Exadata Storage Cell does not support database version 10g, ACFS becomes an important storage option for customers wishing to host older databases on their Exadata system.

However, those new configuration options and flexibility come with one major performance restriction. When ACFS for database files is in use, the Exadata is still not supporting the Smart Scan operations and is not able to push database operations directly to the storage. Hence, for a best performance result, it is recommended to store database files on the Exadata Storage using ASM disk groups.

As per any other system, when implementing ACFS on Exadata Database Machine, snapshots and tagging are supported for database and general-purpose files, while replication, security, encryption, audit and high availability NFS functionalities are only supported with general-purpose files.

trivadis
makes **IT** easier.

5 CONCLUSION

Oracle Automatic Storage Management 12c is a single integrated solution, designed to manage database files and general-purpose data under different hardware and software configurations. The adoption of ASM and ACFS not only eliminates the need for third party volume managers and file systems, but also simplifies the storage management offering the best I/O performance, enforcing Oracle best practices. In addition, ASM 12c with the Flex ASM setup removes previous important architecture limitations:

- Availability: the hard dependency between the local ASM and database instance, was a single point of failure. In fact, without Flex ASM, the failure of the ASM instance causes the crash of all local database instances.

- Performance: Flex ASM reduces the network traffic generated among the ASM instances, leveraging the architecture scalability; and it is easier and faster to keep the ASM metadata synchronized across large clusters. Finally yet importantly, only few nodes of the cluster have to support the burden of an ASM instance, leaving additional resources to application processing.

Oracle ASM offers a large set of configurations and options; it is now our duty to understand case-by-case, when it is relevant to use one setup or another, with the aim to maximize performance, availability and flexibility of the infrastructure.

trivadis

makes **IT** easier.