浅谈ASM

Lunar
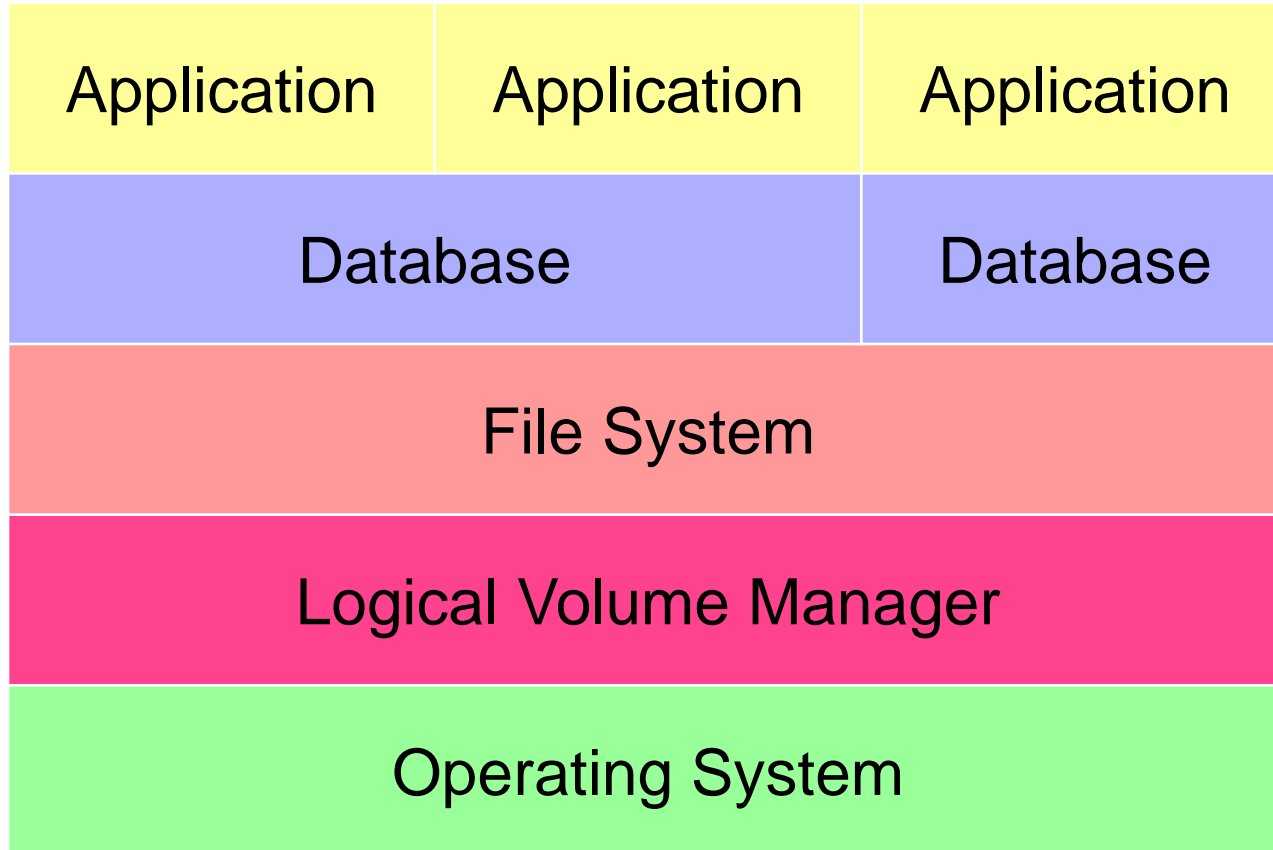
QQ: 5163721

www.lunar2013.com

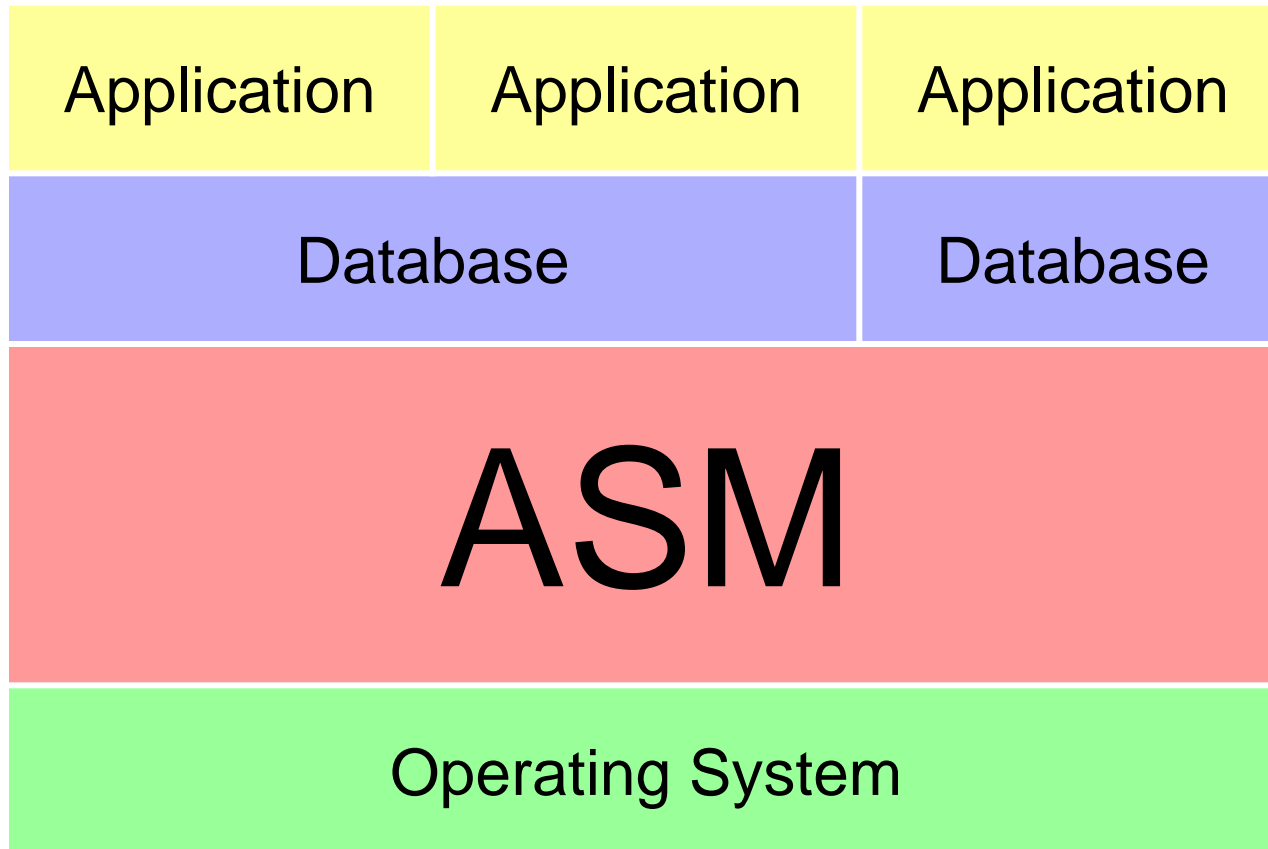# Agend

- **ASM的历史和简介**

- **ASM体系架构**

- **ASM在不同版本的新特性**

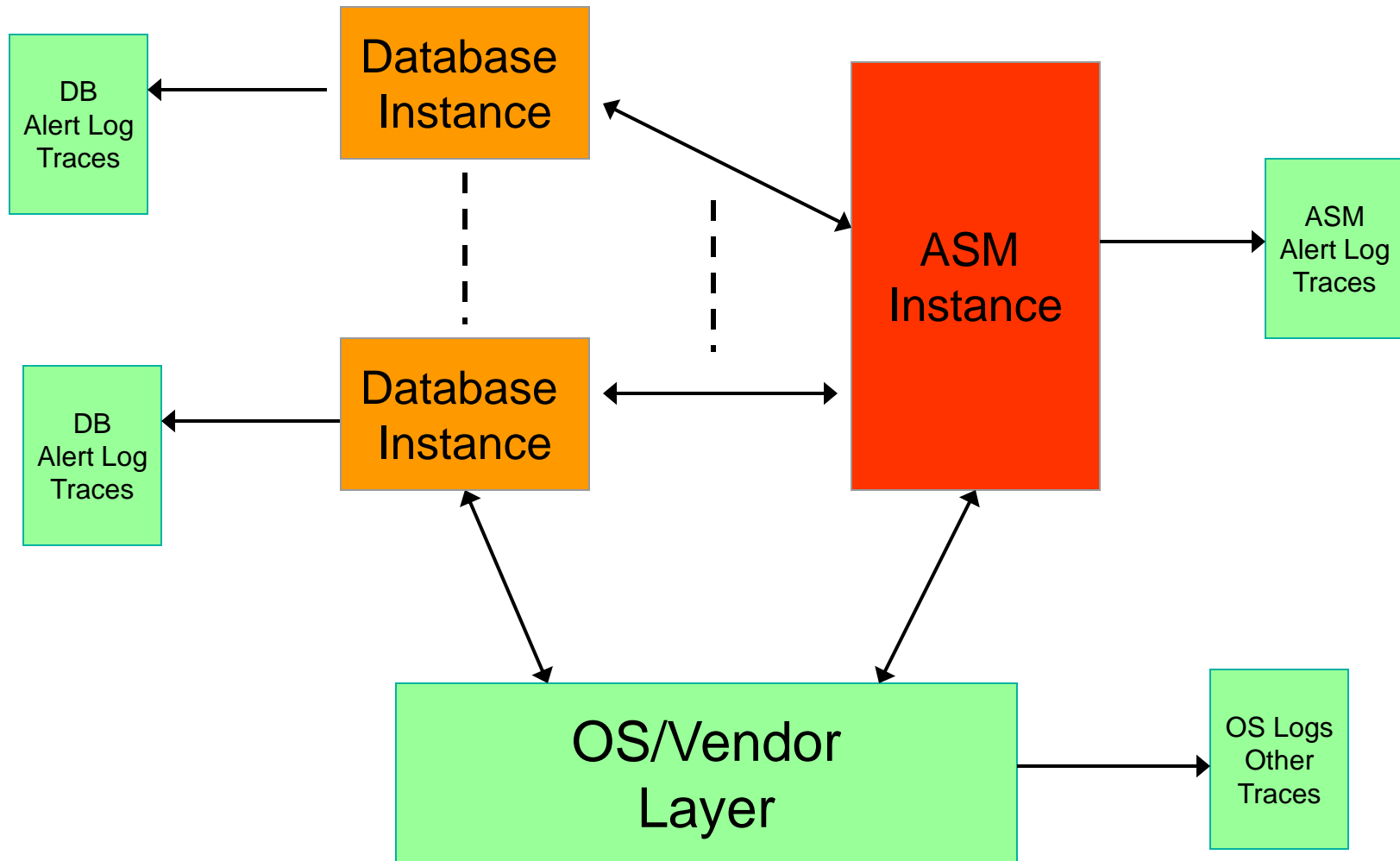- **ASM存储和分配机制**

- **ASM日常使用注意事项**

Oracle 9i Software Stack

Application	Application	Application
Database		Database
File System		
Logical Volume Manager		
Operating System		

Oracle 10g with ASM

Application	Application	Application
Database		Database

ASM

Operating System
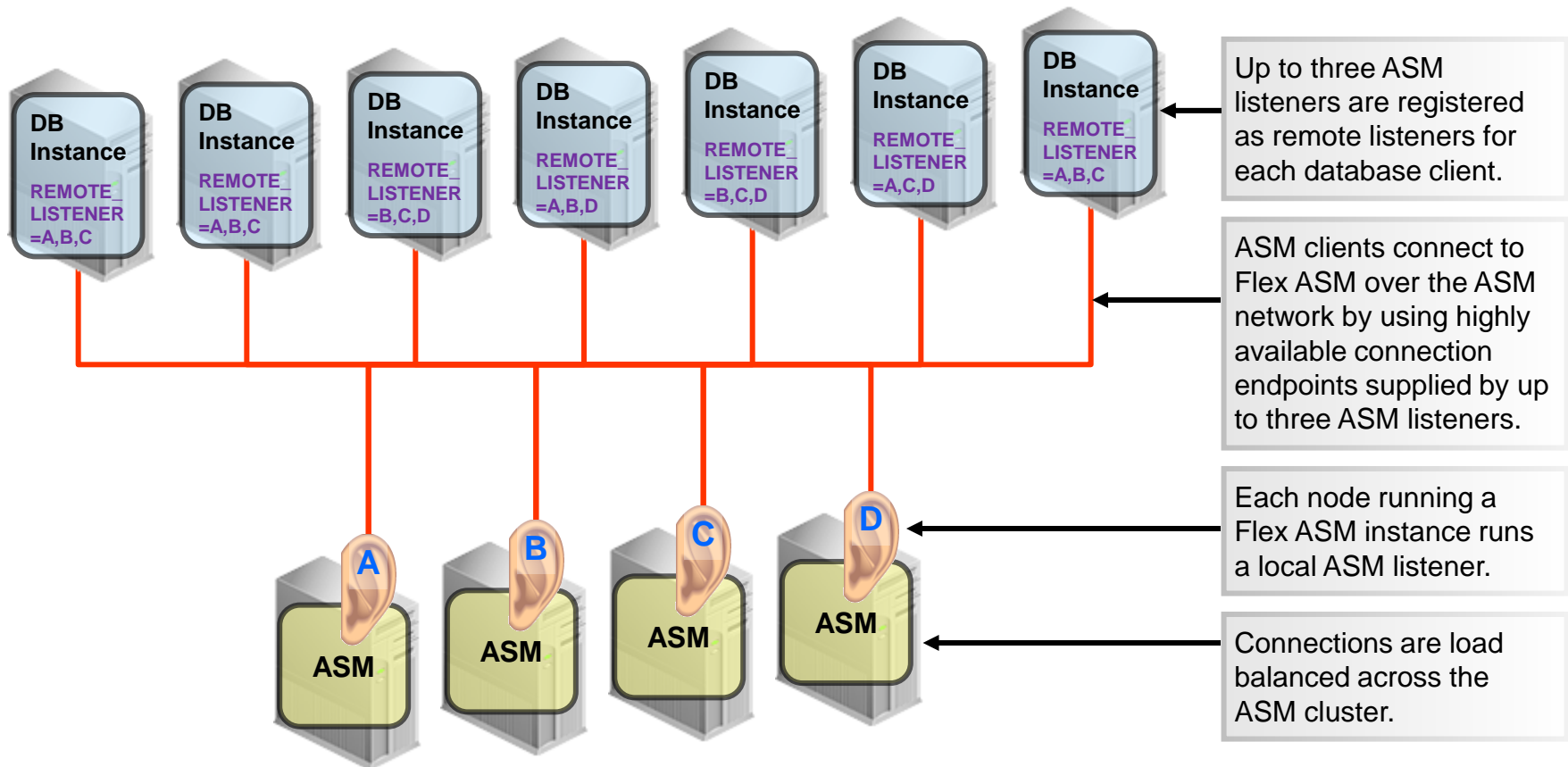
Agend

- **ASM的历史和简介**
- **ASM体系架构**
- **ASM在不同版本的新特性**
- **ASM存储存储和分配机制**
- **ASM日常使用注意事项**
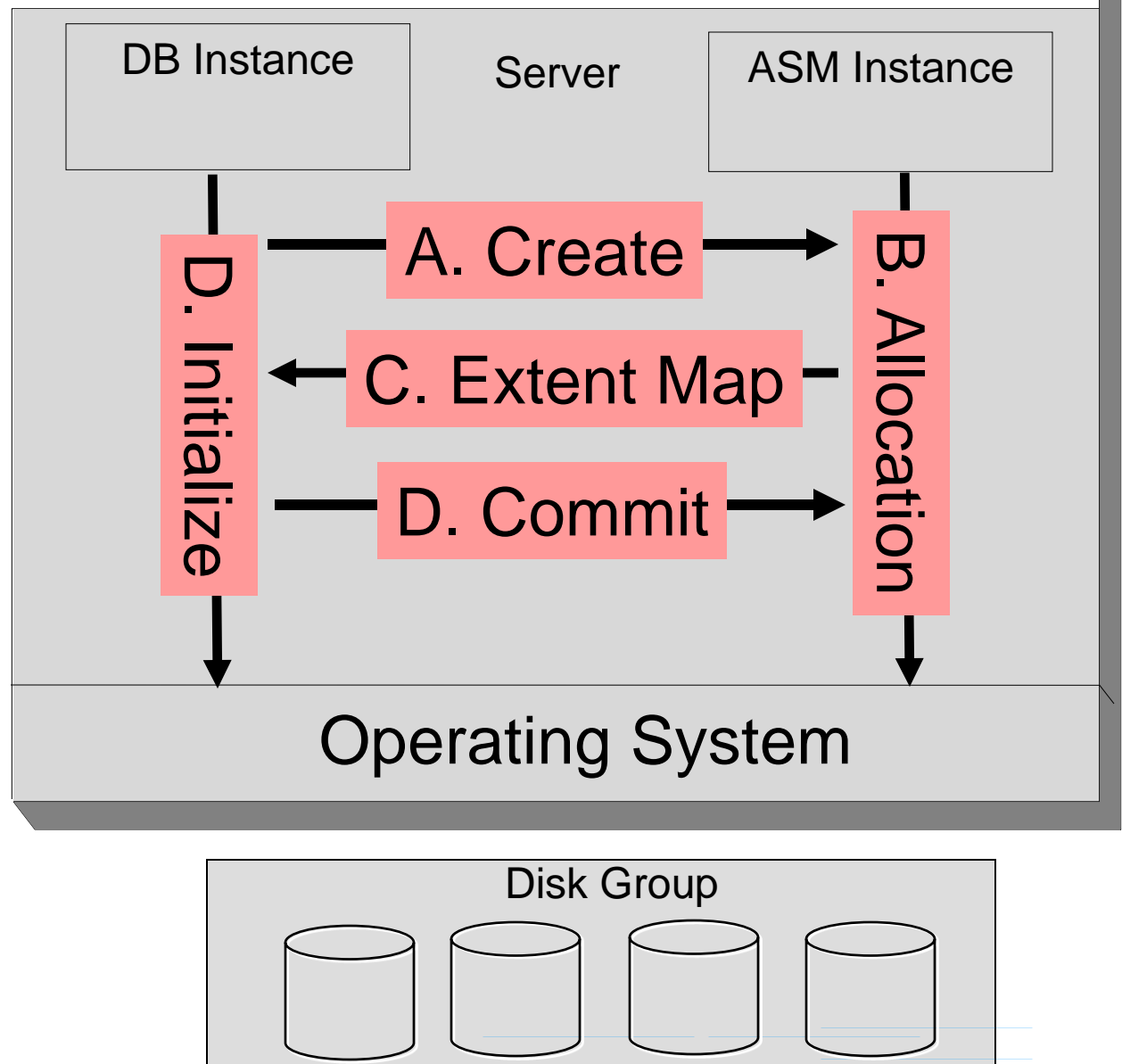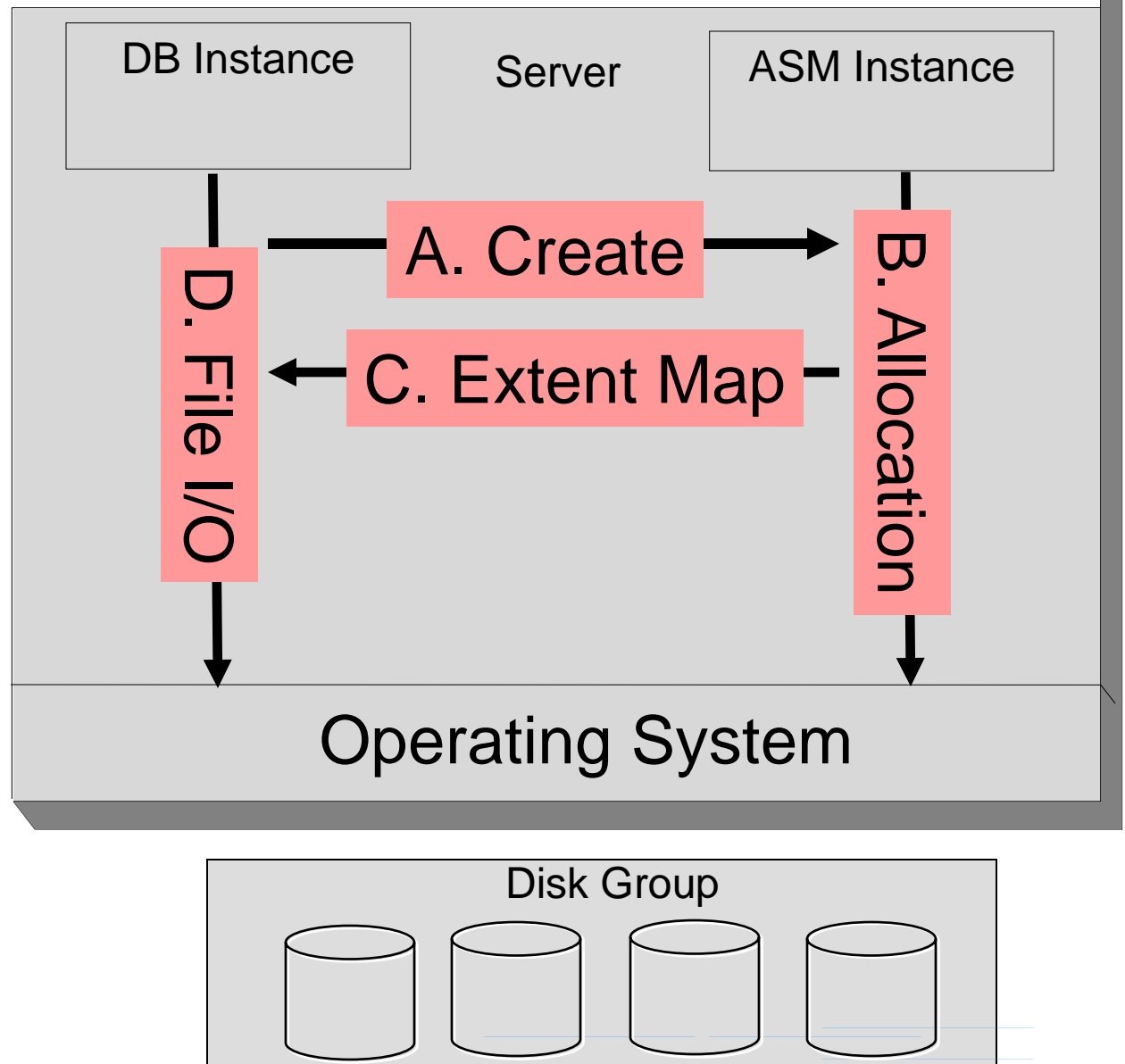
12.1 ASM以前的ASM架构概览

12.1 ASM以后的Flex ASM架构概览

DB Instance REMOTE_LISTENER=A,B,C

DB Instance REMOTE_LISTENER=A,B,C

DB Instance REMOTE_LISTENER=B,C,D

DB Instance REMOTE_LISTENER=A,B,D

DB Instance REMOTE_LISTENER=B,C,D

DB Instance REMOTE_LISTENER=A,C,D

DB Instance REMOTE_LISTENER=A,B,C

A ASM   B ASM   C ASM   D ASM

Up to three ASM listeners are registered as remote listeners for each database client.

ASM clients connect to Flex ASM over the ASM network by using highly available connection endpoints supplied by up to three ASM listeners.

Each node running a Flex ASM instance runs a local ASM listener.

Connections are load balanced across the ASM cluster.

ASM 常规体系结构

Database file creation

Database file open

	Server	
DB Instance		ASM Instance

A. Create

B. Allocation

C. Extent Map

D. File I/O

Operating System

Disk Group

ASM的进程介绍--图解ASMB

1. **PMON，SMON，LGWR，ARCH，CKPT，DBWR**

2. **ASM processes in database instance:**

    ASMB - ASM background process used for client extent management

    RBAL - Background process that is used for diskgroup management

    o???? - ASM slave processes

3. **Processes that Run in ASM instance:**

    ASMB - Starts up only during DB startup using spfile

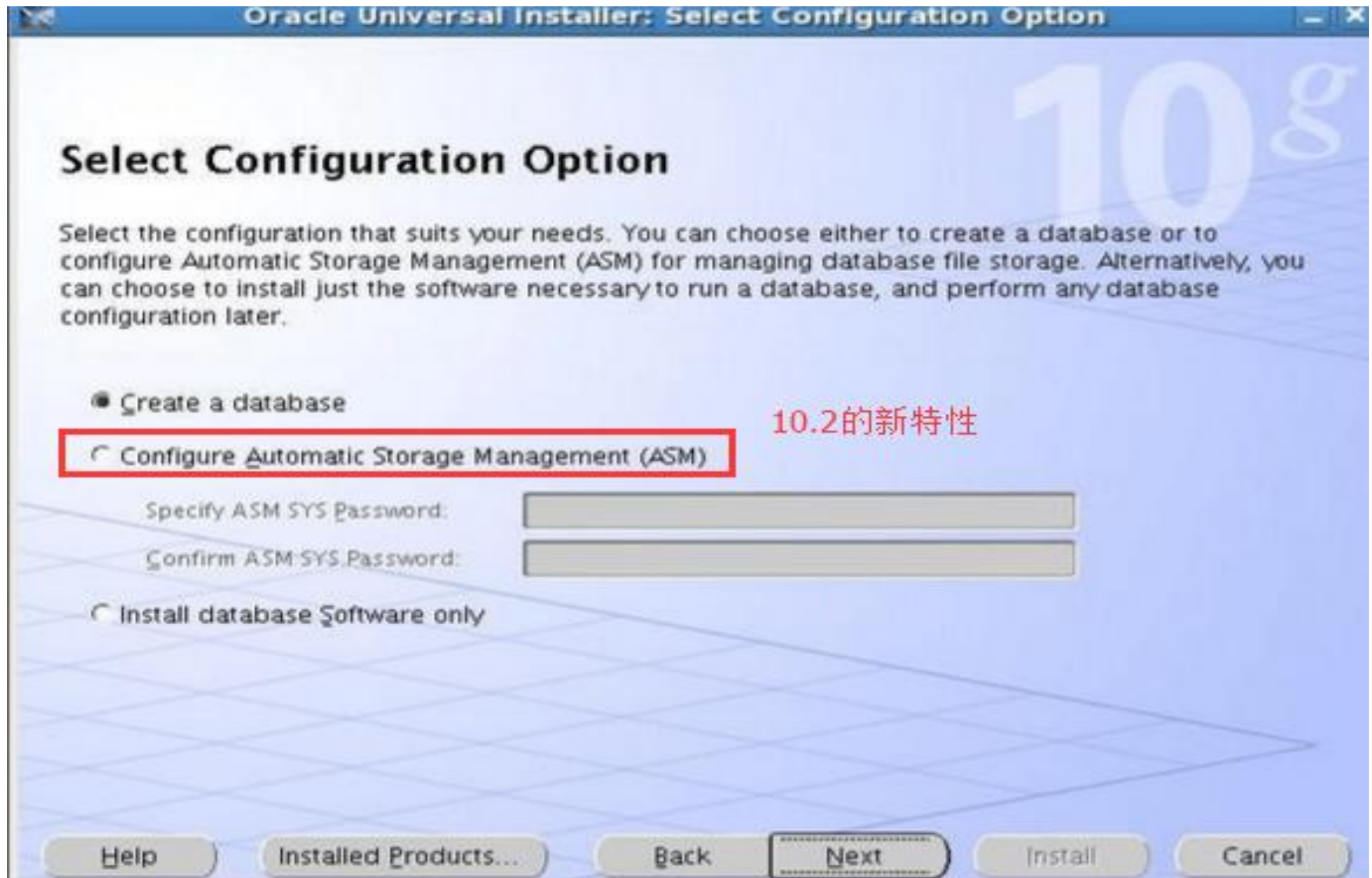   RBAL - Background process that is used for diskgroup management

    ARBn - ASM Rebalance Process.These processes

•   DBW0 - DB writes, same as database DB writer, but deals with ASM cache

    SMON - Recovery process, Same as database SMON, but deals with diskgroup recovery

    CKPT - Checkpoint process, Similar to database CKPT

    PSP0 - Process that Starts other Processes, used to startup other backgrounds

    GMON - Group monitor, used for partner and status table, and node membership

    n00x - GMON Slave processes

    ora_<procname>_<dbsid> - Foregrounds servicing clients commands from client <procname> of database

    <dbsid>

    ora_ASMB_<dbsid>: Special ASM foreground that controls database <dbsid> ASMB

    KATE - Konductor of ASM Temporary Errands, used to process disk online

    VKTM - Process to maintain a fast timer, same as database

    PING - Process to measure network latency, same as database

    DIA? - Diag process, same as database

    DIAG - Diag process, same as database

    LGWR - Log writer, similar to database, but deals with diskgroups

    LMON - Lock monitor, Same as database

    LMS? - Lock monitor slaves, same as database

    MMAN - Autotune SGA process, Same as Database.

    b??? - Slave used to process offline disks.

    x??? - Slave used to expell disks after diskgroup reconfiguration
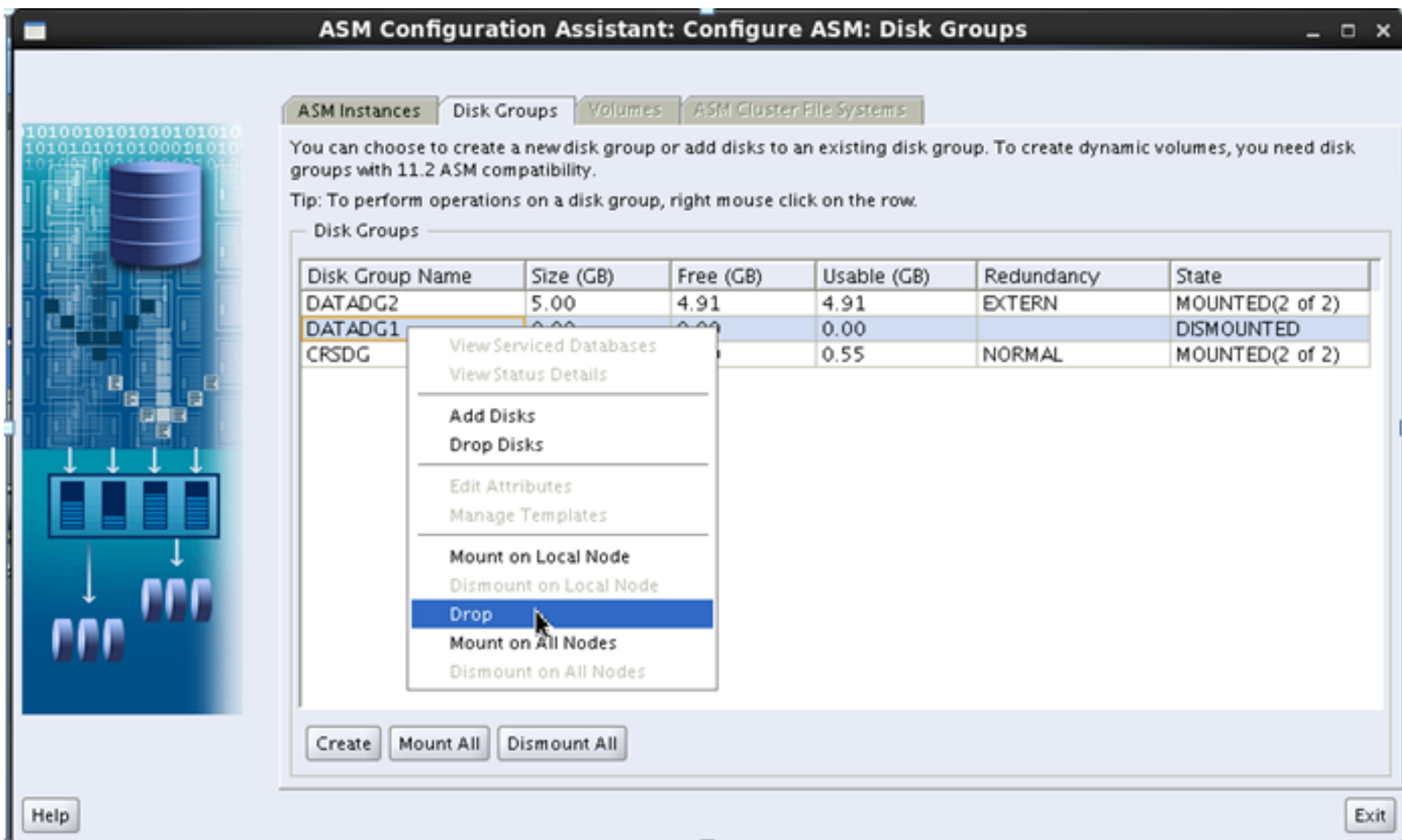
    pz?? - PQ slaves used for global Views

Agend

- **ASM的历史和简介**
- **ASM体系架构**
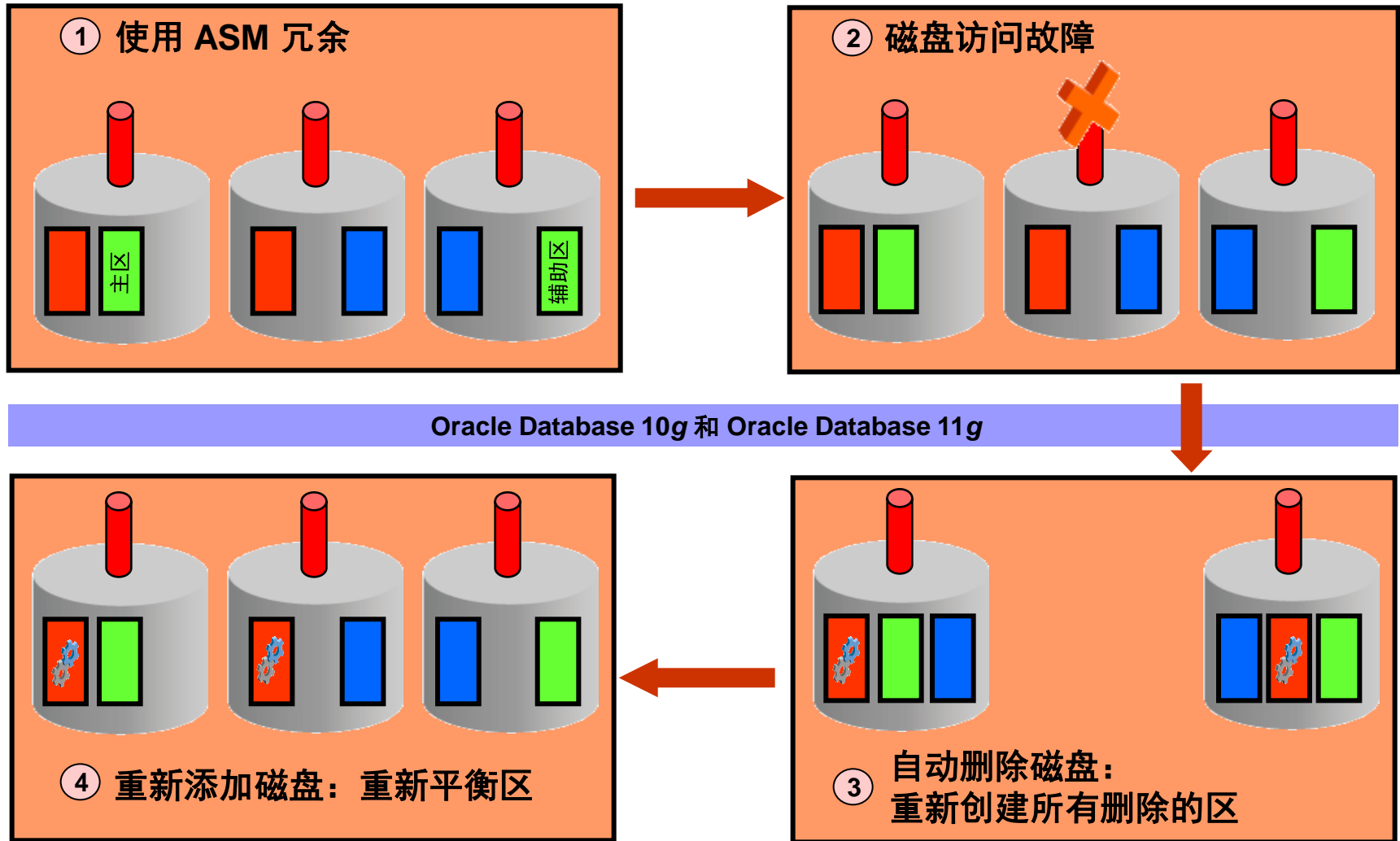- **ASM在不同版本的新特性**
- **ASM存储和分配机制**
- **ASM日常使用注意事项**

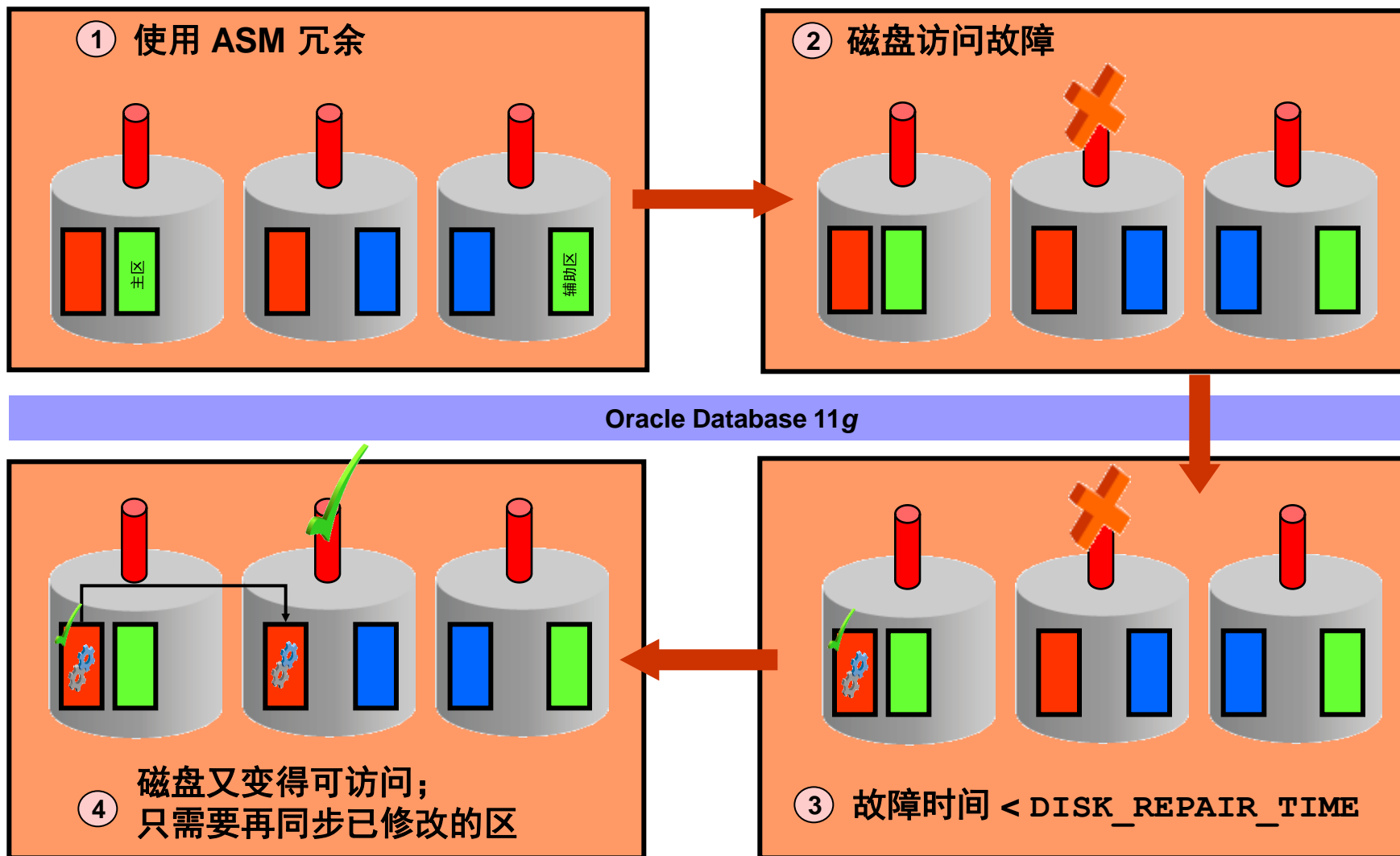# ASM实例在10.2中可以单独配置和管理

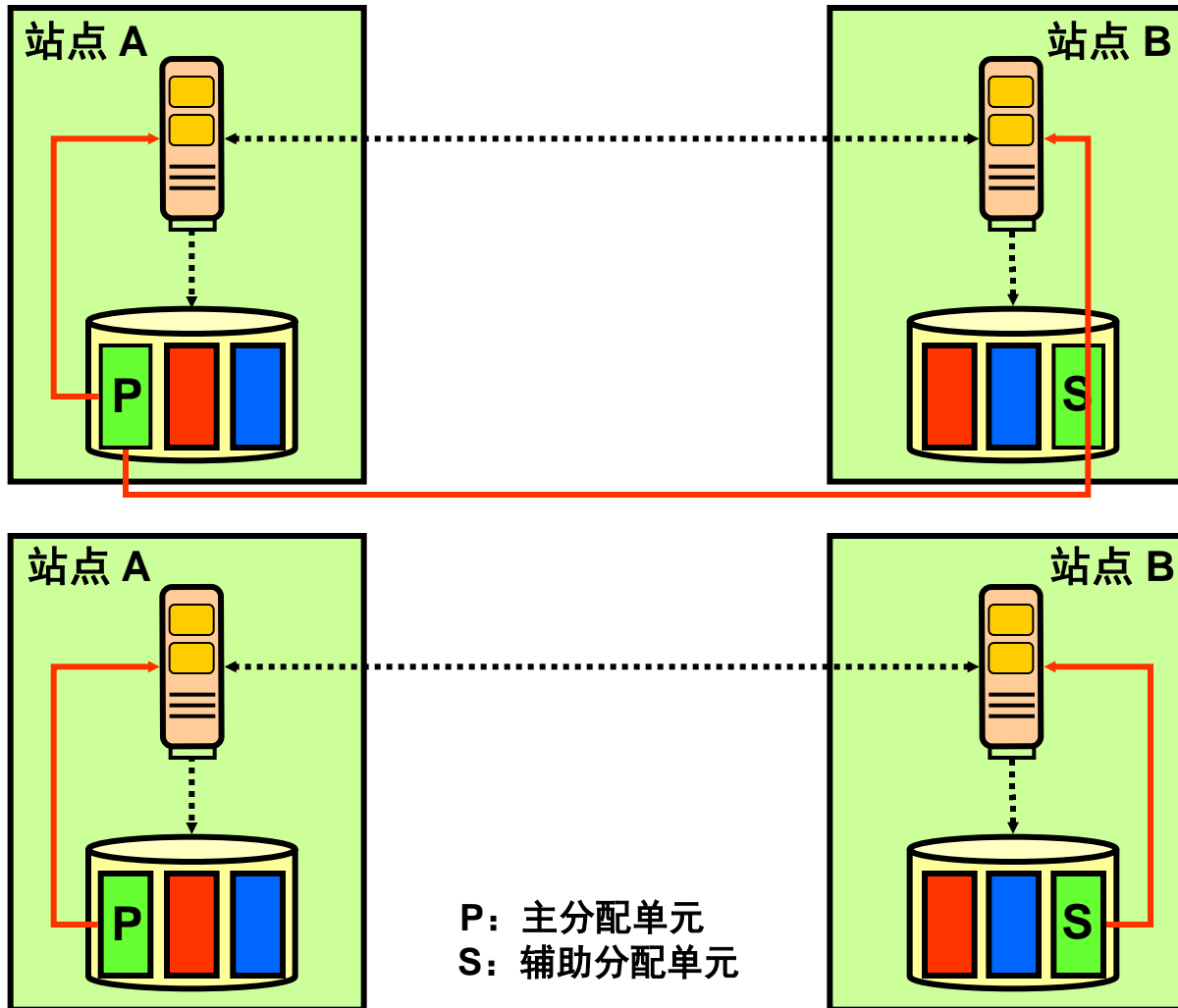# ASM实例在11.2中使用ASMCA配置和管理

无 ASM 快速镜像再同步时

① 使用 ASM 冗余

主区 辅助区

② 磁盘访问故障

Oracle Database 10*g* 和 Oracle Database 11*g*

④ 重新添加磁盘：重新平衡区

③ 自动删除磁盘：重新创建所有删除的区

Oracle 11.1 中的增强
ASM 快速镜像再同步：概览

① 使用 ASM 冗余

主区
辅助区

② 磁盘访问故障

Oracle Database 11*g*

④ 磁盘又变得可访问；
只需要再同步已修改的区

③ 故障时间 < `DISK_REPAIR_TIME`

www.lunar2013.com

ASM 首选镜像读取：概览

镜像IO由谁完成？

P：主分配单元
S：辅助分配单元

ASM 首选镜像读取：设置

设置

```
ASM_PREFERRED_READ_FAILURE_GROUPS=DATA.SITEA
```
在第一个实例上

```
ASM_PREFERRED_READ_FAILURE_GROUPS=DATA.SITEB
```
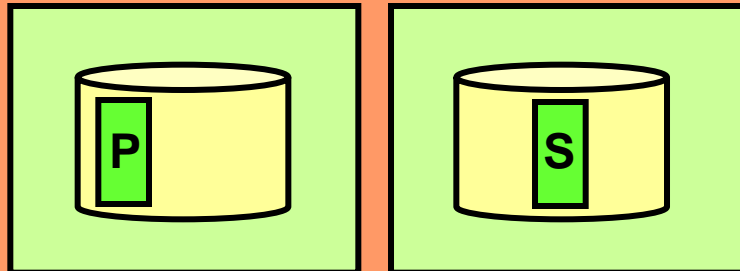在第二个实例上

监视

```
SELECT preferred_read FROM v$asm_disk;
```
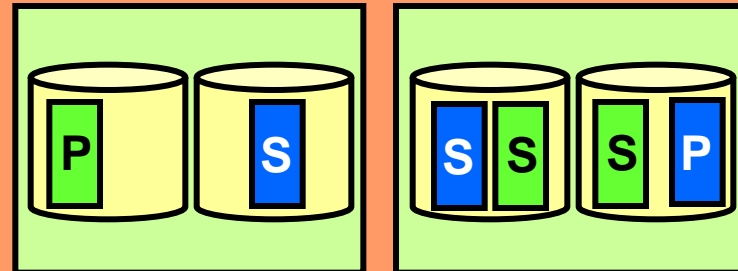
```
SELECT * FROM v$asm_disk_iostat;
```

www.lunar2013.com

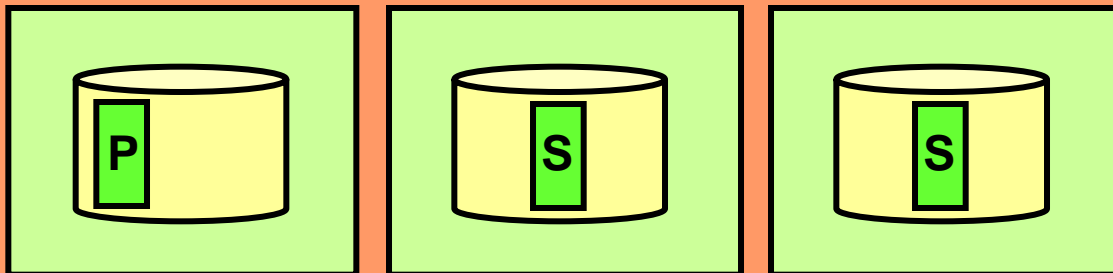9-18

ASM 首选镜像读取：最佳方法

两个站点：正常冗余

P

S

仅两个故障组：每个实例一个

两个站点：高冗余

P S S S S P

最多四个故障组：每个实例二个

三个站点：高冗余

P S S

仅三个故障组：每个实例一个

Oracle Database 11.1 中的 ASM 可伸缩性

ASM 强制实施下列限制：

- **63 个磁盘组**
- **10,000 个 ASM 磁盘**
- **每个 ASM 磁盘 4 PB**
- **40 EB 的存储空间**
- **每个磁盘组 1 百万个文件**
- **最大文件大小：**
  - **外部冗余：140 PB**
  - **正常冗余：42 PB**
  - **高冗余：15 PB**

- **可变extent：**
- **前 20,000 个区 (0–19999) 等于 AU**
- **接下来的 20,000 个区 (20000–39999) 等于 4 个 AU**
- **40,000 个以上的区等于 16 个 AU**

www.lunar2013.com

SYSASM 角色

- 使用 SYSASM 角色管理 ASM 实例可以避免 DBA 与存储管理员之间出现重叠。

```
SQL> CONNECT / AS SYSASM
```

```
SQL> CREATE USER username IDENTIFIED by passwd;
```
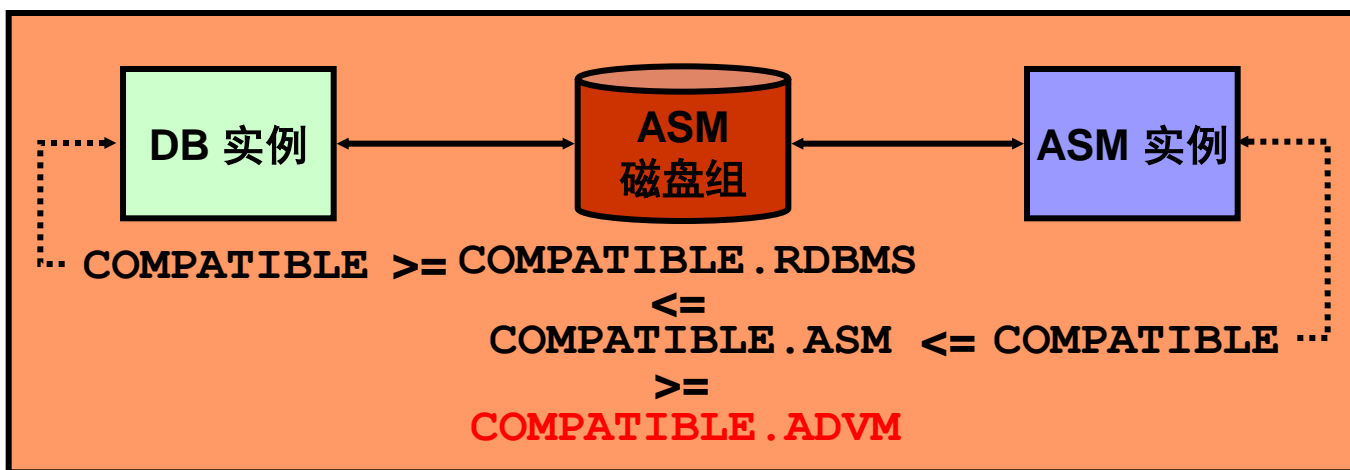
```
SQL> GRANT SYSASM TO username;
```

```
SQL> CONNECT username/passwd AS SYSASM;
```

```
SQL> DROP USER username;
```

SYSDBA 将被废弃：

- Oracle Database 11*g* 版本 1 的行为与 Oracle Database 10*g* 中的相同。
- 在11.2的版本中， ASM 实例中的 SYSDBA 权限将受到限制。

ASM 磁盘组兼容性

- 可以分别控制每个磁盘组的兼容性：
  - ASM 兼容性控制 ASM 元数据的磁盘存储结构。
  - RDBMS 兼容性控制使用者客户机的最低级别。
  - **ADVM 兼容性确定是否可以使用 Oracle ASM 动态卷管理器**。
- 对磁盘组兼容性的设置是不可逆的。

```
DB 实例          ASM          ASM 实例
                磁盘组

COMPATIBLE >= COMPATIBLE.RDBMS
              <=
     COMPATIBLE.ASM <= COMPATIBLE
              >=
     COMPATIBLE.ADVM
```

ASM 磁盘组属性

名称	属性	值	说明
au_size	C	1|2|4|8|16|32|64MB	磁盘组中分配单元的大小
compatible.rdbms	AC	有效的数据库版本	数据库与 ASM 之间交换的消息的格式
compatible.asm	AC	有效的 ASM 实例版本	磁盘上 ASM 元数据结构的格式
disk_repair_time	AC	0 M to 2^{32} D	磁盘脱机之后删除此磁盘之前的时间量
template.*tname*.redundancy	A	UNPROTECT|MIRROR|HIGH	指定模板的冗余
template.*tname*.stripe	A	COARSE|FINE	指定模板的条带化属性

C：CREATE（创建）
A：ALTER（变更）

```
CREATE DISKGROUP DATA NORMAL REDUNDANCY
DISK '/dev/raw/raw1','/dev/raw/raw2'
ATTRIBUTE 'compatible.asm'='11.1';
```

适用于快速重新平衡的受限装载磁盘组

- 磁盘组只能装载在单个实例上。
- 任何数据库客户机或其它 **ASM** 实例都不能获得访问权限。
- 重新平衡可以继续进行，而无锁定开销。

① `ALTER DISKGROUP data DISMOUNT;`

② `ALTER DISKGROUP data MOUNT RESTRICT;`

③   维护任务：添加/删除磁盘...

④ `ALTER DISKGROUP data DISMOUNT;`

⑤ `ALTER DISKGROUP data MOUNT;`

- **默认情况下，MOUNT 使用 NOFORCE 选项：**
  - 所有磁盘必须可用
- **带有 FORCE 选项的 MOUNT：**
  - 如果存在仲裁磁盘，则会使不可用的磁盘脱机
  - 如果所有磁盘都可用，则操作会失败

```
ALTER DISKGROUP data MOUNT [FORCE|NOFORCE];
```

强制删除磁盘组

- 允许用户删除无法装载的磁盘组
- 如果在任何位置装载了磁盘组，则操作会失败

```
DROP DISKGROUP data FORCE INCLUDING CONTENTS;
```

ASMCMD 扩展

用户创建的目录
模板
磁盘组兼容性
磁盘组名称
磁盘名称和故障组

md_backup

full

cp

$ asmcmd help

md_restore

nodg

lsdsk

newdg
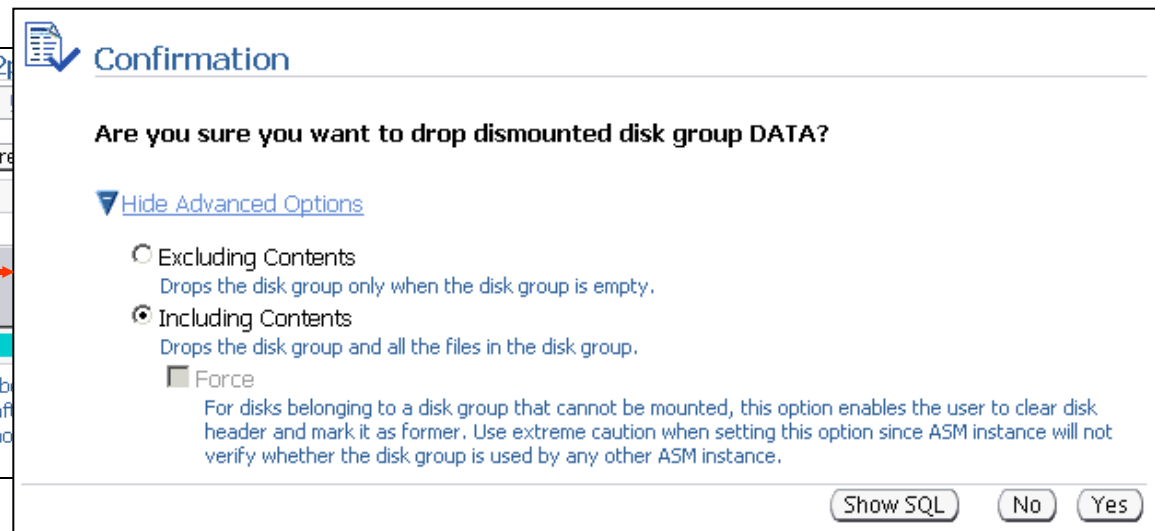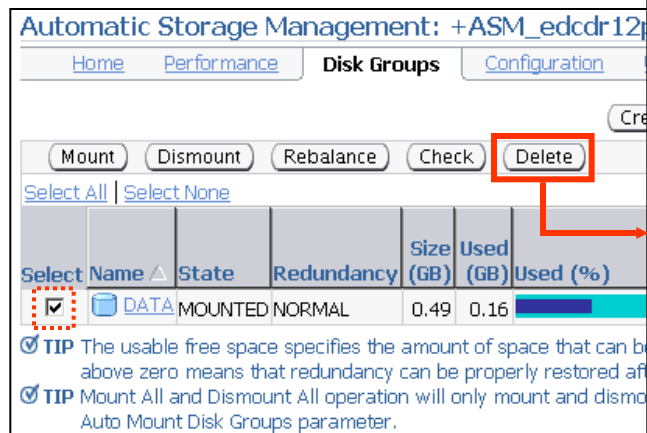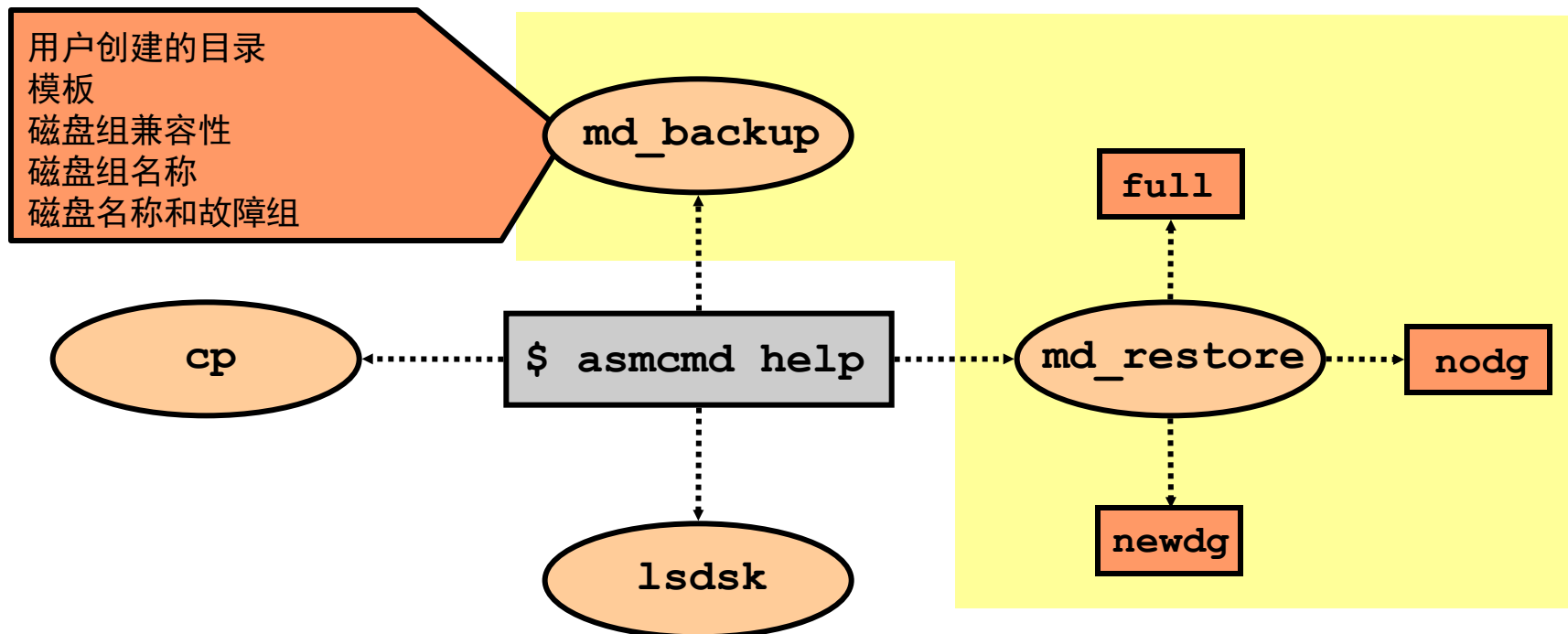
OCR 和Voting Disk放在ASM上

- **OCR 是常规ASM文件**
  - 一个新的ASM文件类型

- **Voting Disk**
  - 存储在选定的ASM磁盘上
  - 使用'crsctl' 为Voting disk指定一个磁盘组
  - 基于 Ext/Normal/High 冗余级别，ASM 自动创建1/3/5 个Voting Disk
  - Quorum failure group

- ASM 基于磁盘组的冗余级别决定OCR和Voting Disk 的冗余度

- ASM SPFILE参数文件也可以被ASM支持

ASM 智能数据放置

COMPATIBLE.ASM>=11.2
COMPATIBLE.RDBMS>=11.2

不使用 IDP 的 ASM

使用 IDP 的 ASM
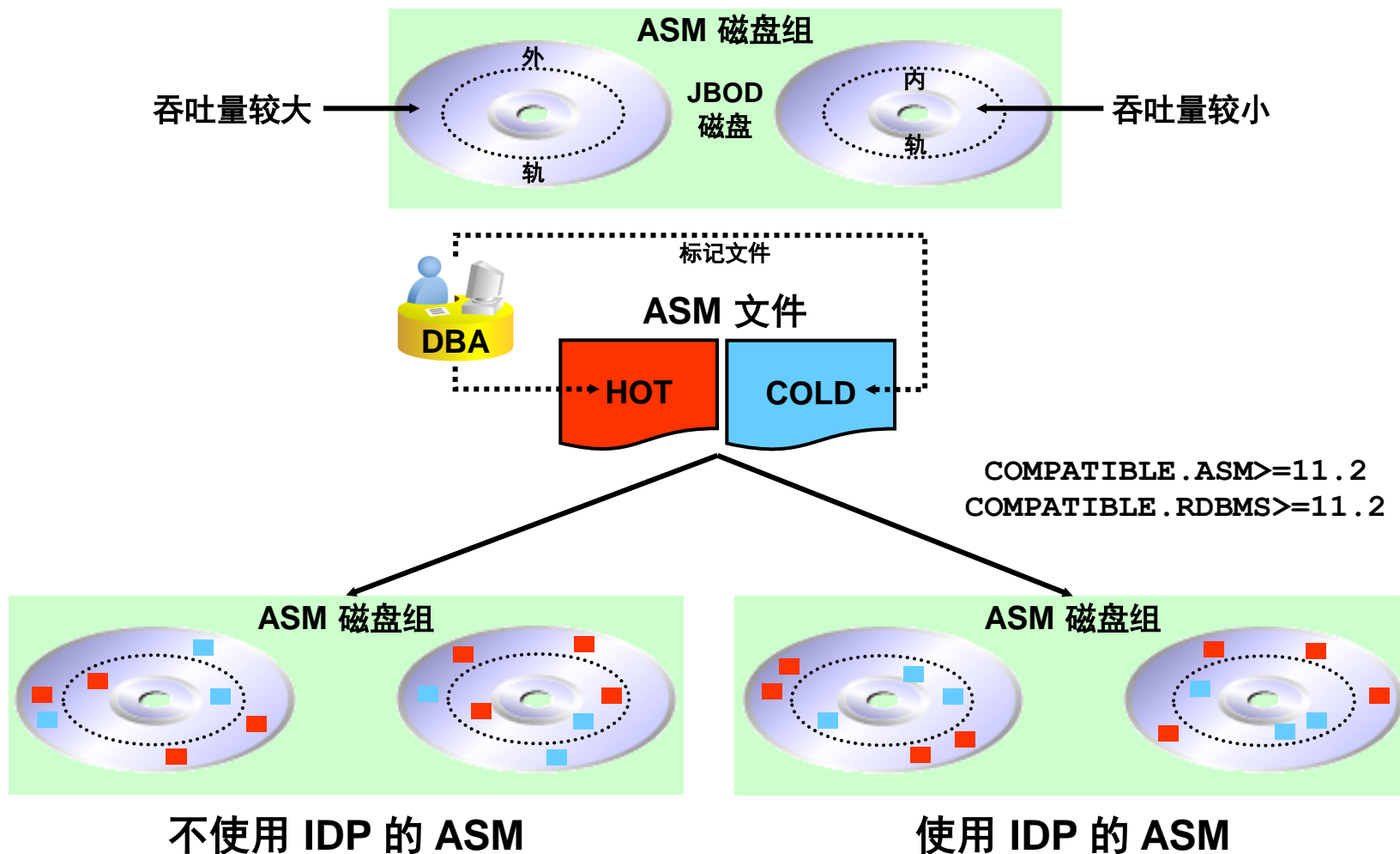
ASM 智能数据放置的最佳实践

读取频率	写入频率	主要区域	辅助区域
高	高	HOT	MIRRORHOT
高	低	HOT	MIRRORCOLD
低	高	COLD	MIRRORHOT
低	低	COLD	MIRRORCOLD

12.1 ASM - 物理元数据冗余

```
NOTE: client cdblunar:cdblunar:ASM mounted group 3 (DATADG1)↓
Sun Nov 08 20:57:42 2015↓
WARNING: cache read a corrupt block: group=3(DATADG1) dsk=0 blk=1 disk=0 (DATADG1_0000) incarn=3916348983 au=0 blk=1 count=1↓
Sun Nov 08 20:57:42 2015↓
Errors in file /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc:↓
ORA-15196: invalid ASM block header [kfc.c:29297] [endian_kfbh] [2147483648] [1] [0 != 1]↓
NOTE: a corrupted block from group DATADG1 was dumped to /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc↓
WARNING: cache read (retry) a corrupt block: group=3(DATADG1) dsk=0 blk=1 disk=0 (DATADG1_0000) incarn=3916348983 au=0 blk=1 count=1↓
Sun Nov 08 20:57:42 2015↓
Errors in file /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc:↓
ORA-15196: invalid ASM block header [kfc.c:29297] [endian_kfbh] [2147483648] [1] [0 != 1]↓
ORA-15196: invalid ASM block header [kfc.c:29297] [endian_kfbh] [2147483648] [1] [0 != 1]↓
WARNING: Failed to verify disk 0 (DATADG1_0000) of group 3 (DATADG1) path /dev/sdb1 reason: endian_kfbh 0 != 1↓
NOTE: corrupt disk header dumped to /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc↓
NOTE: cache repaired a corrupt block: group=3(DATADG1) dsk=0 blk=1 on disk 0 from disk=0 (DATADG1_0000) incarn=3916348983 au=11 blk=1 count=1↓
WARNING: cache read a corrupt block: group=3(DATADG1) dsk=0 blk=3 disk=0 (DATADG1_0000) incarn=3916348983 au=0 blk=3 count=1↓
Sun Nov 08 20:57:42 2015↓
Errors in file /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc:↓
ORA-15196: invalid ASM block header [kfc.c:29297] [endian_kfbh] [2147483648] [3] [0 != 1]↓
NOTE: a corrupted block from group DATADG1 was dumped to /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc↓
WARNING: cache read (retry) a corrupt block: group=3(DATADG1) dsk=0 blk=3 disk=0 (DATADG1_0000) incarn=3916348983 au=0 blk=3 count=1↓
Sun Nov 08 20:57:42 2015↓
Errors in file /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc:↓
ORA-15196: invalid ASM block header [kfc.c:29297] [endian_kfbh] [2147483648] [3] [0 != 1]↓
ORA-15196: invalid ASM block header [kfc.c:29297] [endian_kfbh] [2147483648] [3] [0 != 1]↓
WARNING: Failed to verify disk 0 (DATADG1_0000) of group 3 (DATADG1) path /dev/sdb1 reason: endian_kfbh 0 != 1↓
NOTE: corrupt disk header dumped to /u01/app/grid/diag/asm/+asm/+ASM/trace/+ASM_ora_19789.trc↓
NOTE: cache repaired a corrupt block: group=3(DATADG1) dsk=0 blk=3 on disk 0 from disk=0 (DATADG1_0000) incarn=3916348983 au=11 blk=3 count=1↓
SYS@cdblunar>↓
```
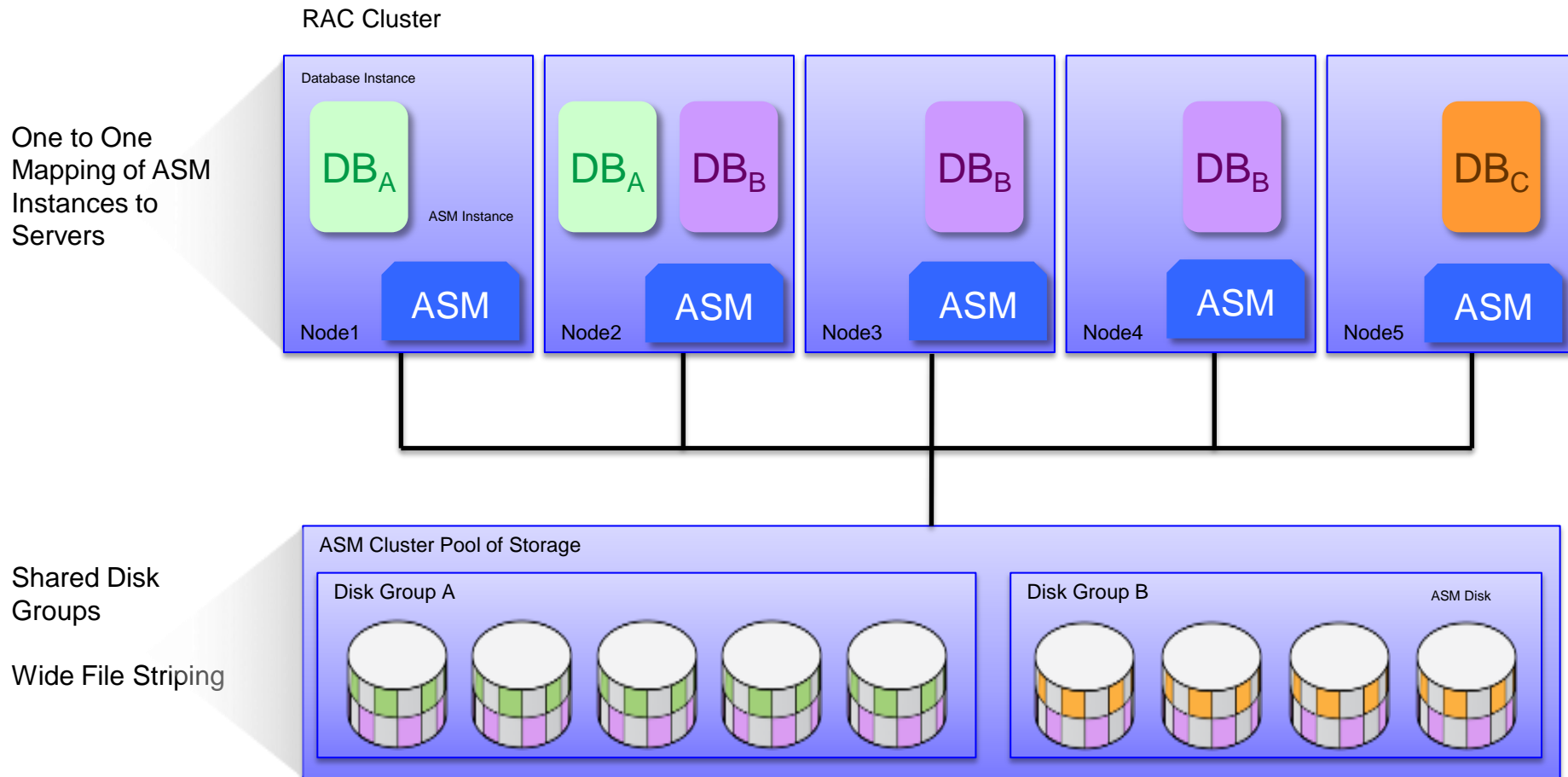
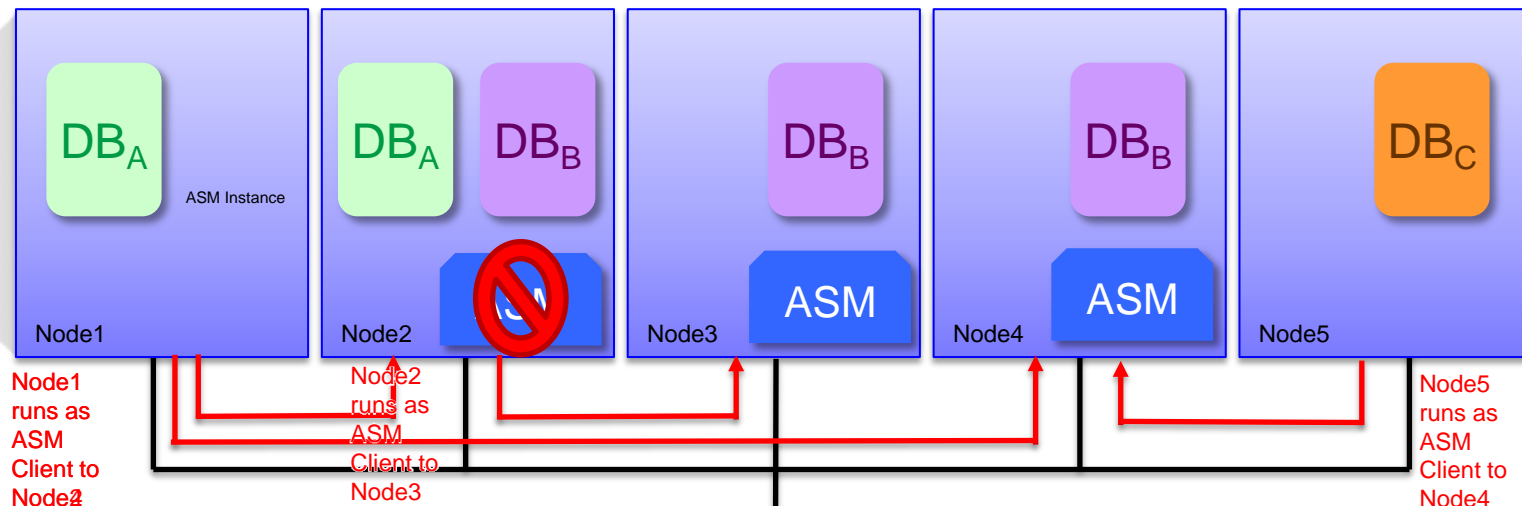# Automatic Storage Management (ASM) Overview

Current State

RAC Cluster

One to One
Mapping of ASM
Instances to
Servers

Database Instance

DB$_A$

ASM Instance

DB$_A$ DB$_B$

DB$_B$

DB$_B$

DB$_C$

ASM   ASM   ASM   ASM   ASM

Node1   Node2   Node3   Node4   Node5

Shared Disk
Groups

Wide File Striping

ASM Cluster Pool of Storage

Disk Group A

Disk Group B

ASM Disk

Flex ASM: Eliminate 1:1 Server Mapping

New: ASM Storage Consolidation in Oracle Database 12c

RAC Cluster

Databases share ASM instances
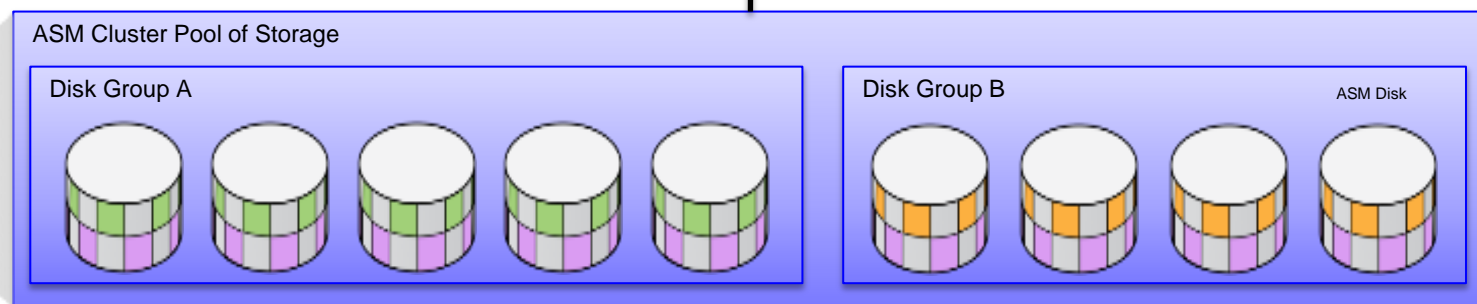
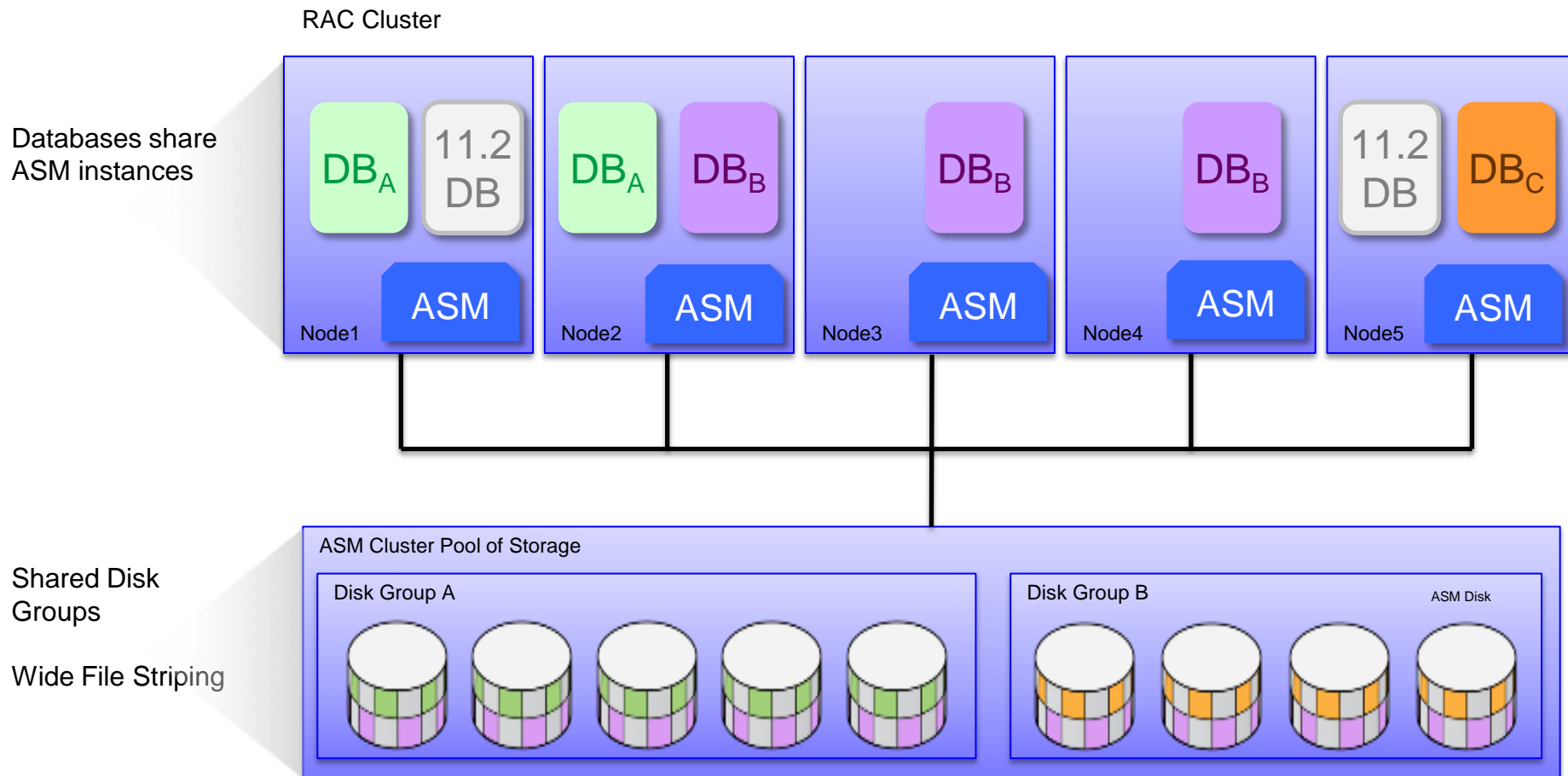DB$_A$ ASM Instance

DB$_A$ DB$_B$

DB$_B$

DB$_B$

DB$_C$

ASM   ASM

Node1   Node2   Node3   Node4   Node5

Node1 runs as ASM Client to Node2

Node2 runs as ASM Client to Node3

Node5 runs as ASM Client to Node4

Shared Disk Groups

Wide File Striping

ASM Cluster Pool of Storage

Disk Group A

Disk Group B   ASM Disk

Agend

- **ASM的历史和简介**

- **ASM体系架构**

- **ASM在不同版本的新特性**

- **ASM存储和分配机制**

- **ASM相关产品介绍**

ASM 存储：概念

数据库

表空间

段

区

Oracle 数据块

数据文件

文件系统
文件或裸设备

ASM 磁盘组

ASM 文件
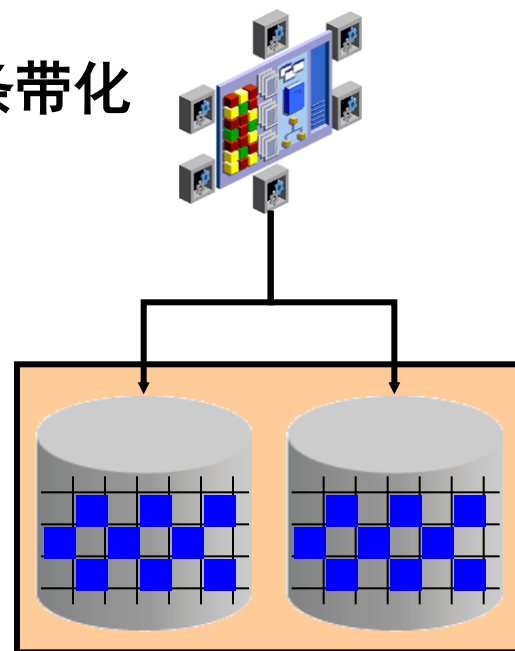**File# 1-255**
File# >=256
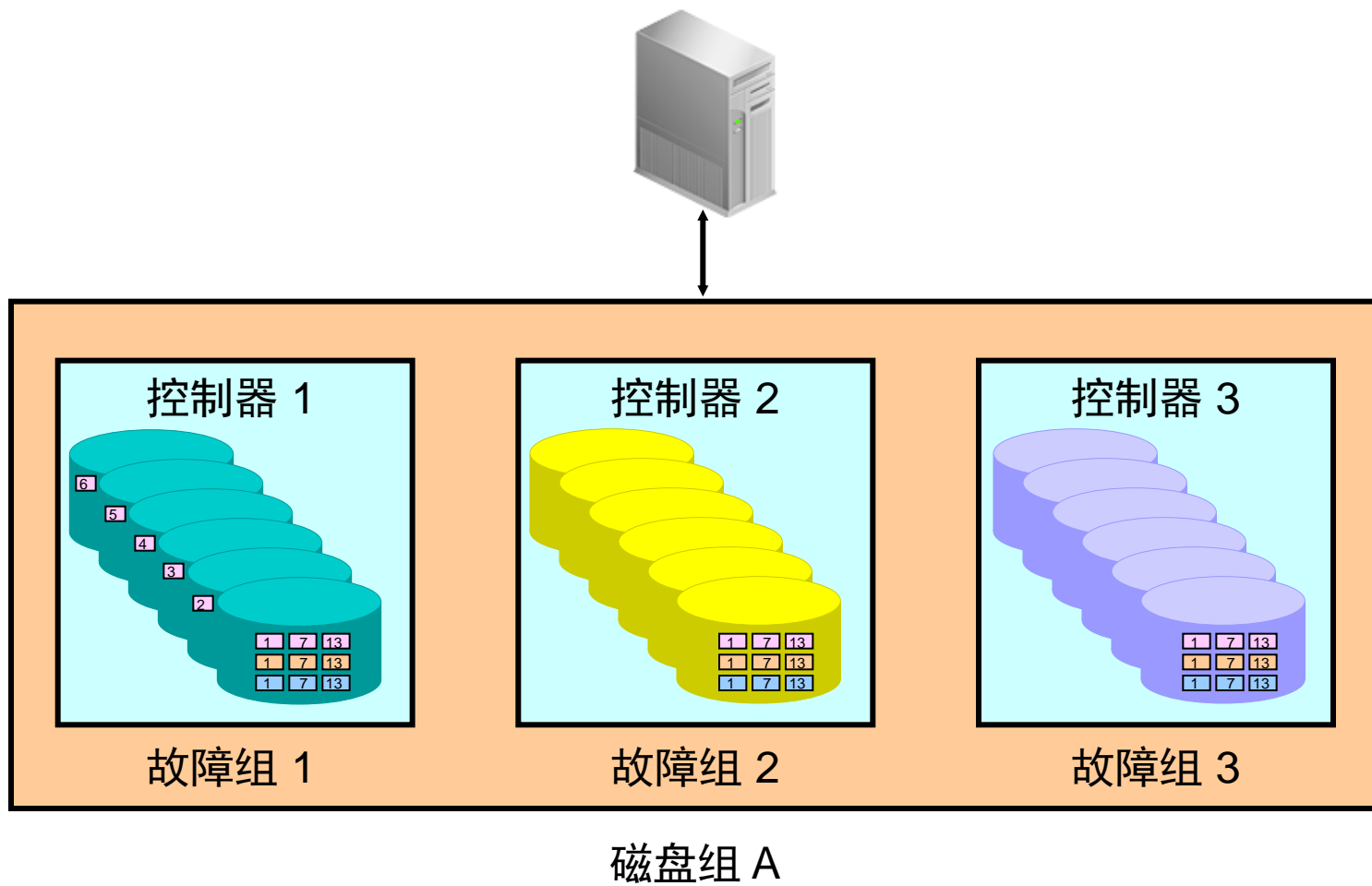
ASM 磁盘

分配单元 (AU)

物理块
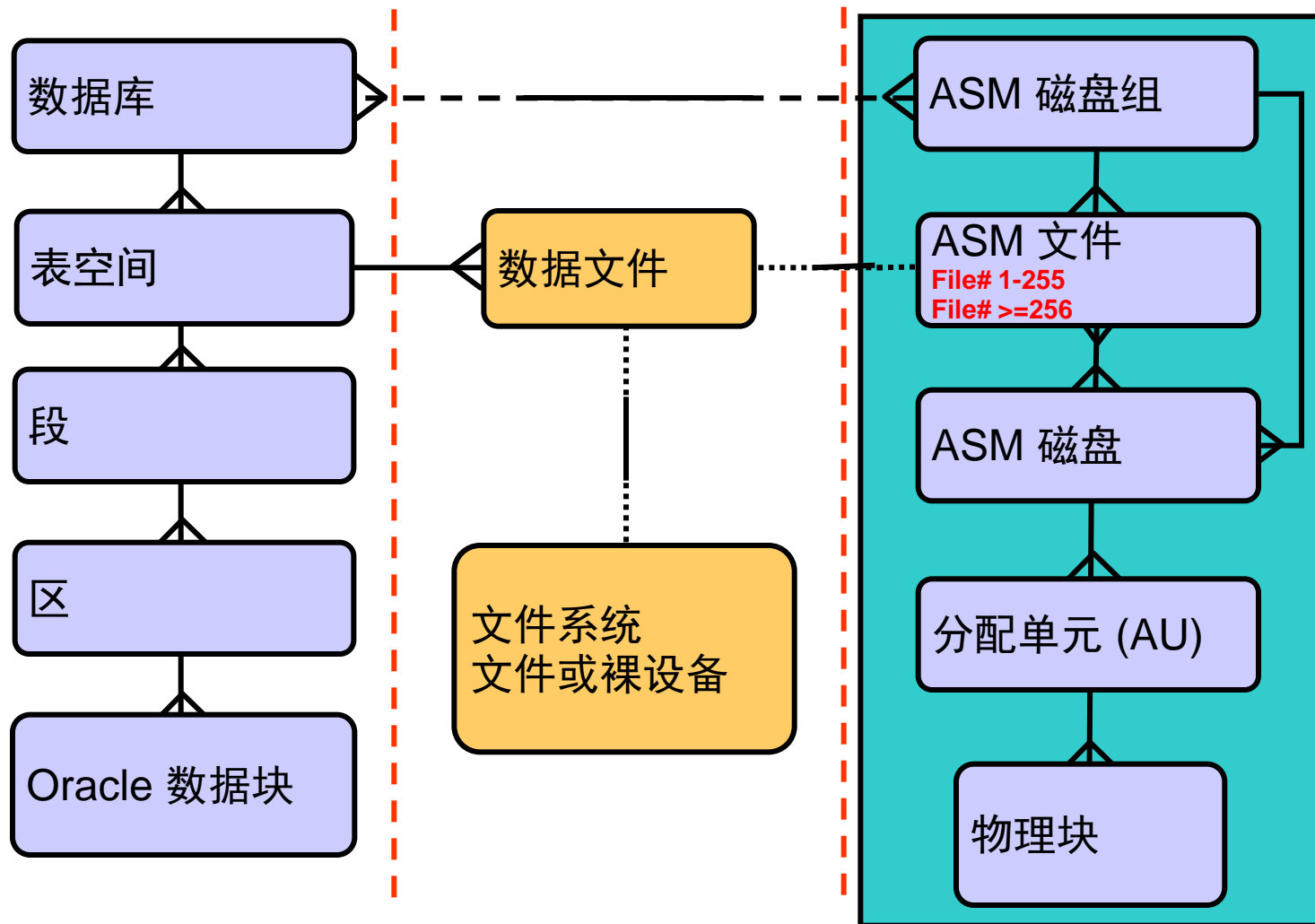
ASM 磁盘组

- 作为逻辑单元管理的磁盘组
- 将磁盘总空间划分为统一大小的单元
- 将各个文件平均分配到所有磁盘中
- 根据文件类型使用粗粒度或细粒度的条带化
- 管理的是磁盘组而非文件

ASM 实例

磁盘组

故障组

磁盘组 A

ASM磁盘和磁盘组（**Disk Groups**）
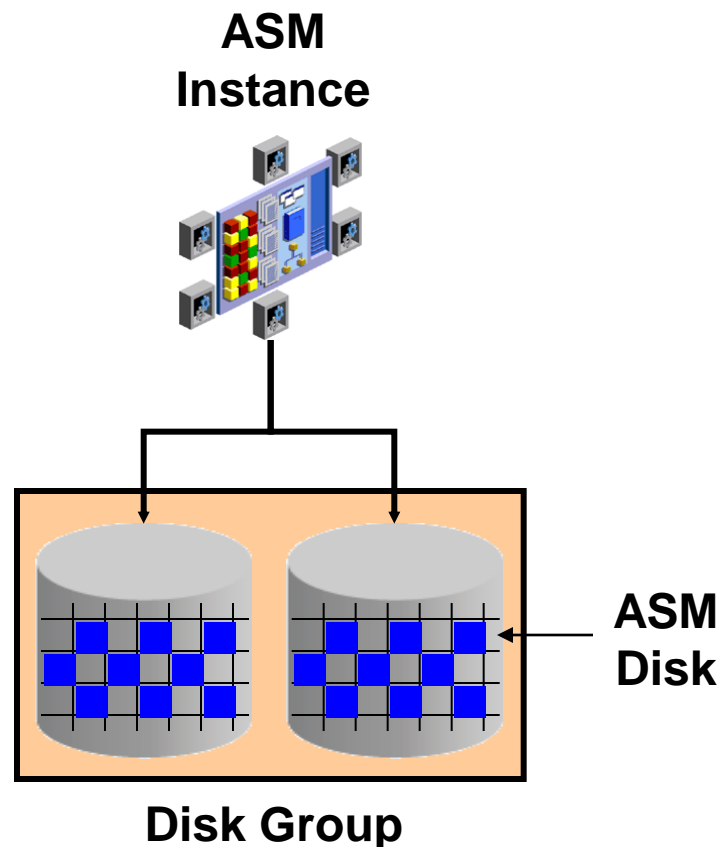
ASM 磁盘 **=** 磁盘分区 或 **LUN**

ASM 磁盘组 **=** 作为一个逻辑单元进行管理的**ASM** 磁盘池

把每一个文件平均散布到跨**ASM**磁盘组的所有**ASM**磁盘

- 使用的分配单元（**AU**）大小为 **1 MB**

条带化

- 基于文件的类型 **(V$ASM_TEMPLATE)**

- **Coarse (1 MB)**

- **Fine (128 KB)**

ASM Instance

ASM Disk

Disk Group

ASM 磁盘组和磁盘

4

9-40

ASM 磁盘组和磁盘

www.lunar2013.com
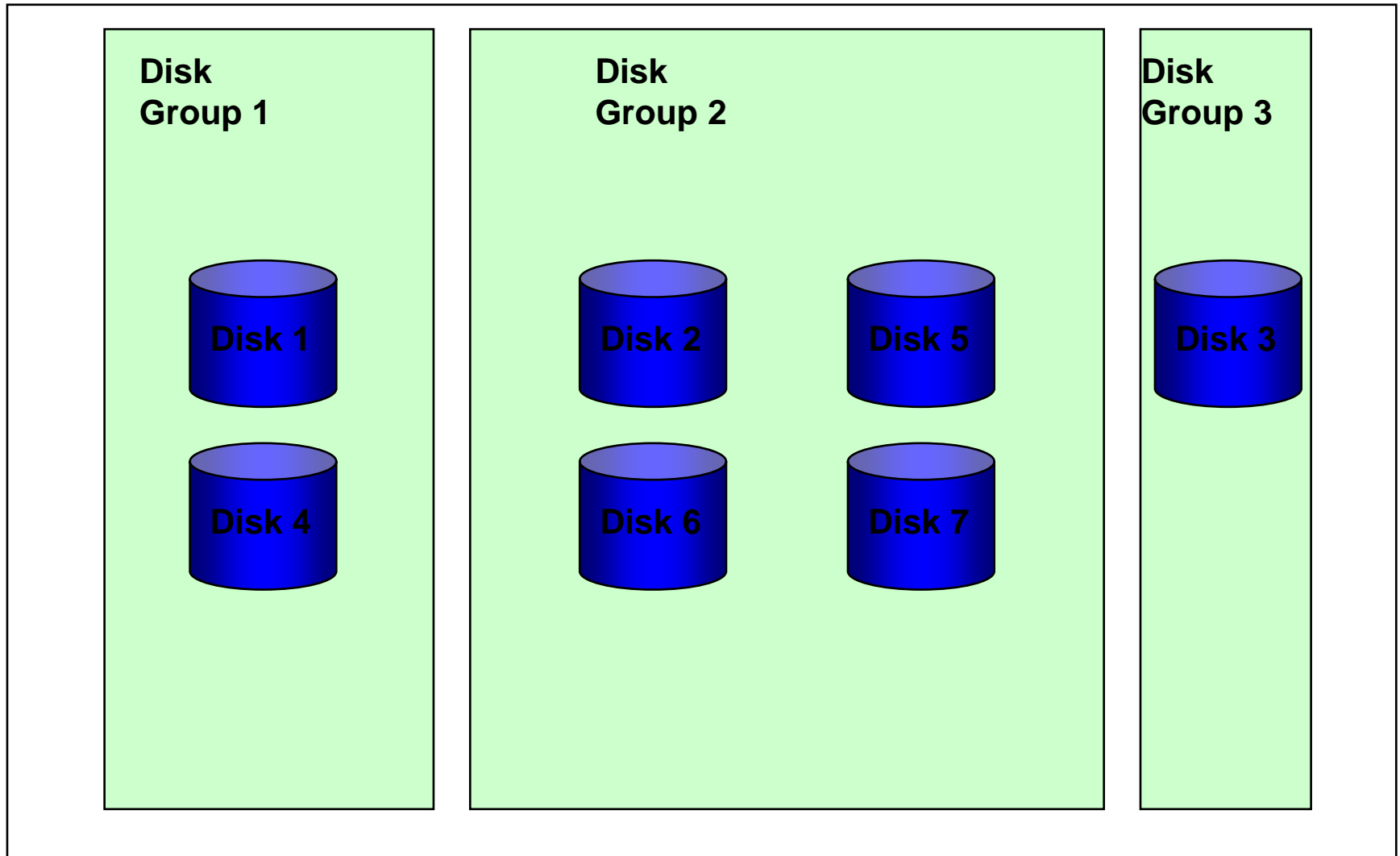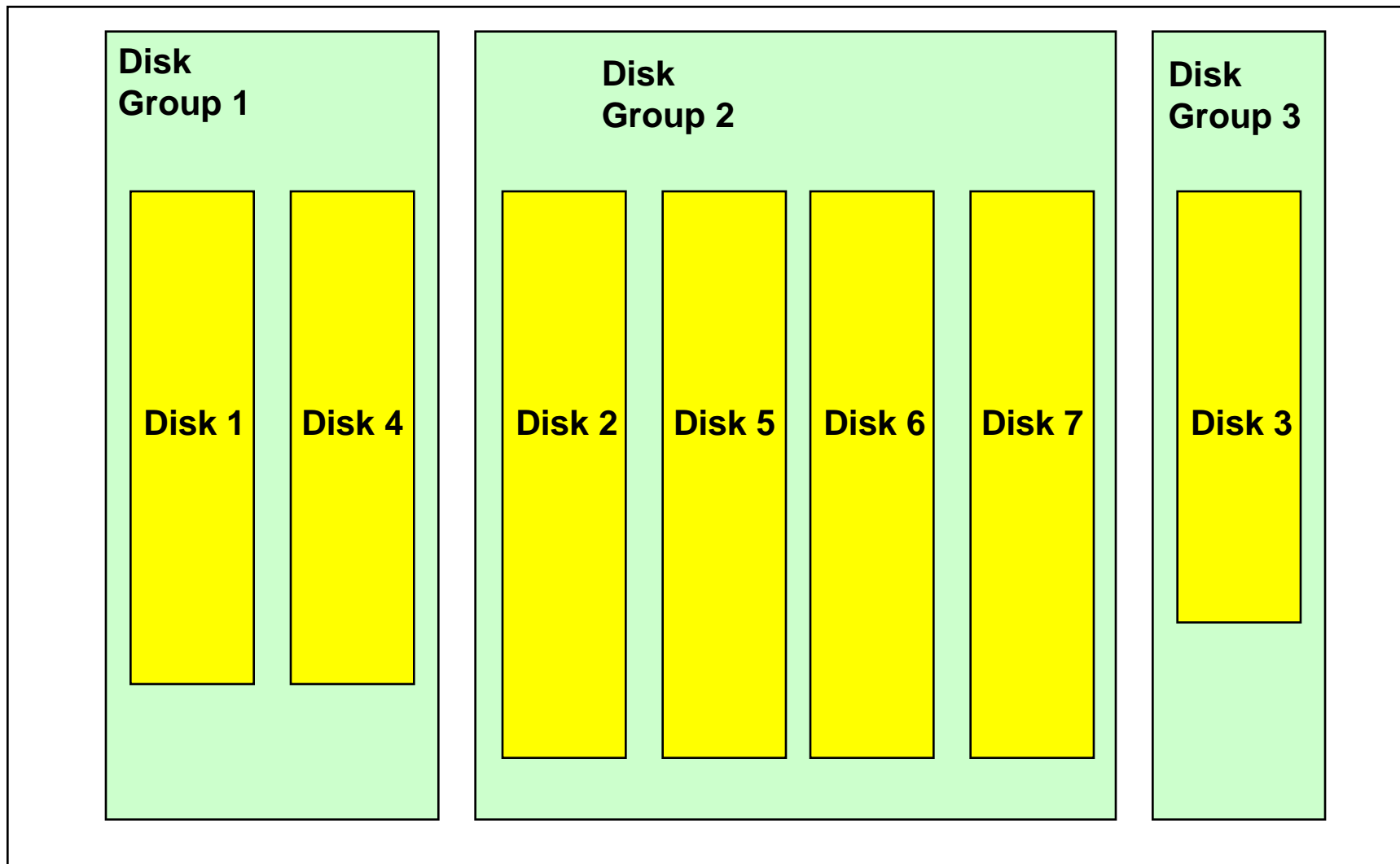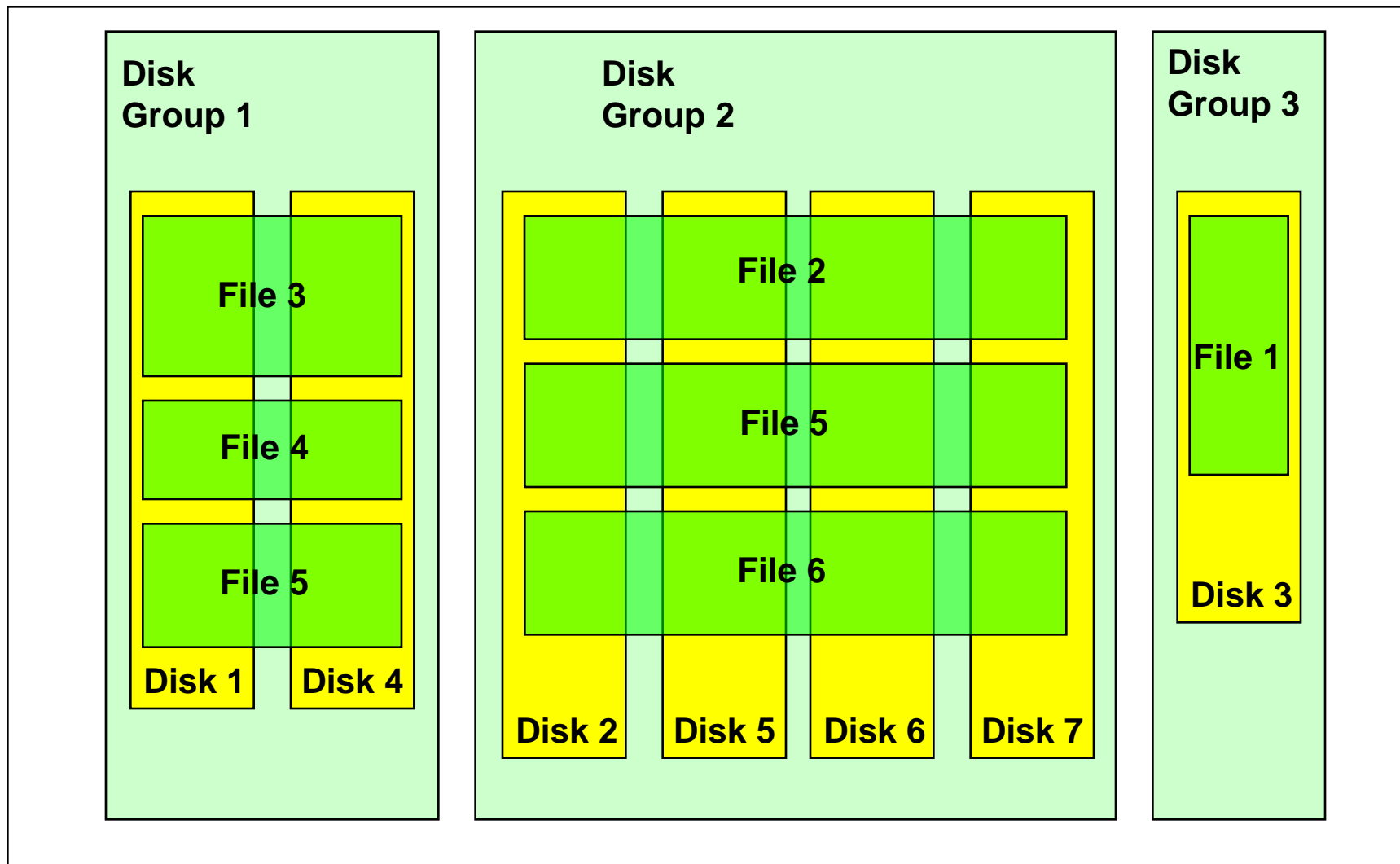
ASM 磁盘组、磁盘和数据库文件

Mirroring & Stripping

- *Mirroring* : Redundant copies of data over multiple disks
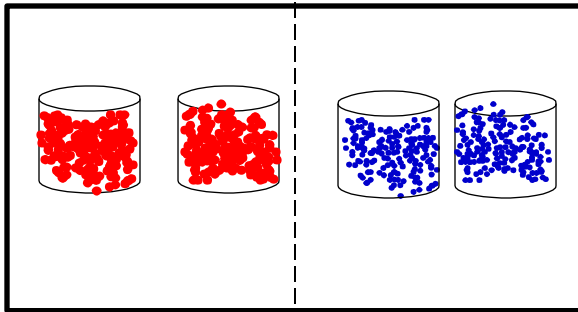
■ *Primary data extents*

■ *Secondary data extents*

- *Stripping* : Spreading data over multiple disks

Unstriped Disks   **Striped Disks**
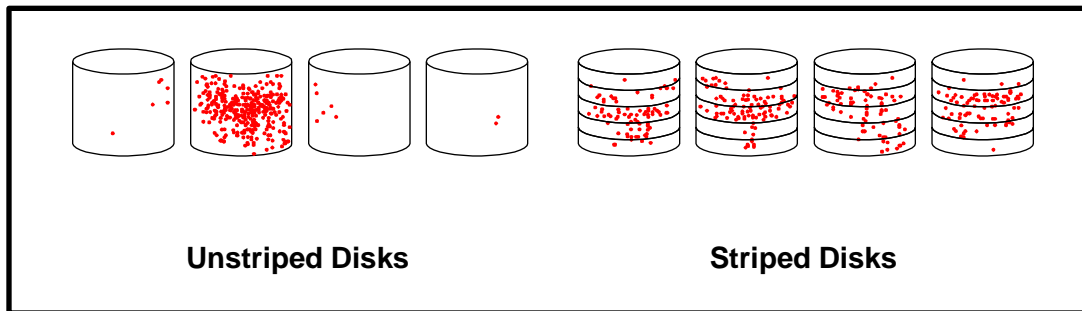
Allocation Unit

ASM Disk
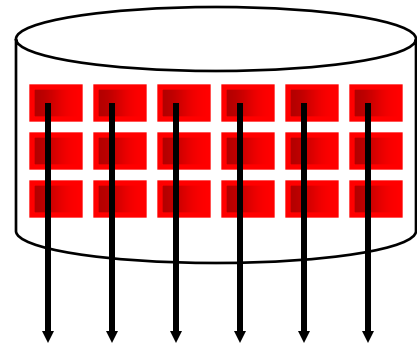
1 MB

ASM disks are divided into allocation units

Unit of allocating storage to a file

One megabyte in size

Small enough not to be hot

Large enough for efficient sequential access

Many allocation units in a single file

_ASM_AUSIZE

Multiple Allocation Unit Size support

- Higher performance for large seq I/O (DW)
- Better leverage of Hardware RAID read-ahead
- Set Oracle MAXIO = AU_size

ASM disk1   ASM disk2   ASM disk3   ASM disk4

1		2		3		
5		6		7		4
						8

20,001   20,002   20,003   20,004

20,005   20,006   20,007
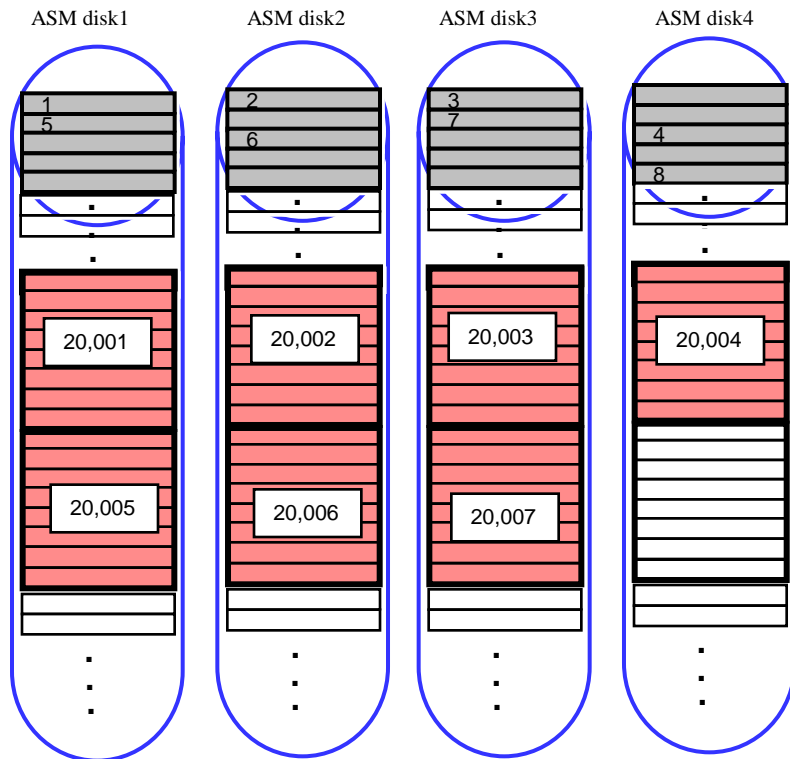
Allocation Unit (AU)
Selected at disk group creation time
and may be 1,2,4,8,16,32,64 MB

Variable size ASM file extents
Extent size = AU size up to 20,000 extents
Extent size = 8*AU up to 40,000 extents
Extent size = 64*AU beyond 40,001 extents

Striping
Coarse Stripe size always = one AU
Fine Stripe size always = 128 KB

ASM中Extent 的分布-1块磁盘成1个磁盘组

http://juliandyke.com/

Metadata

0	
1	
2	
3	
4	
5	
6	
7	

Disk 0

Metadata   Data
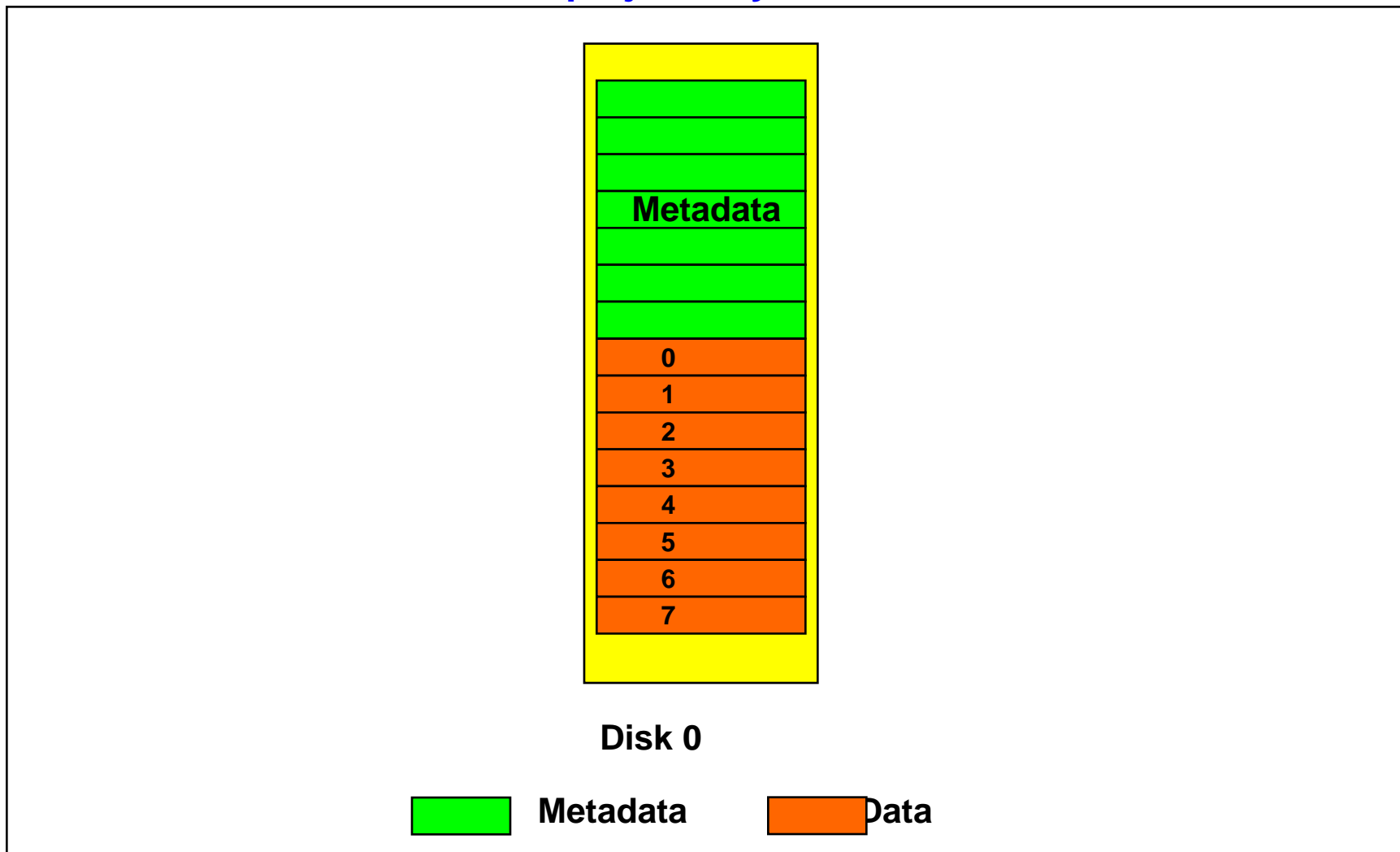
ASM中Extent 的分布-2块磁盘成1个磁盘组
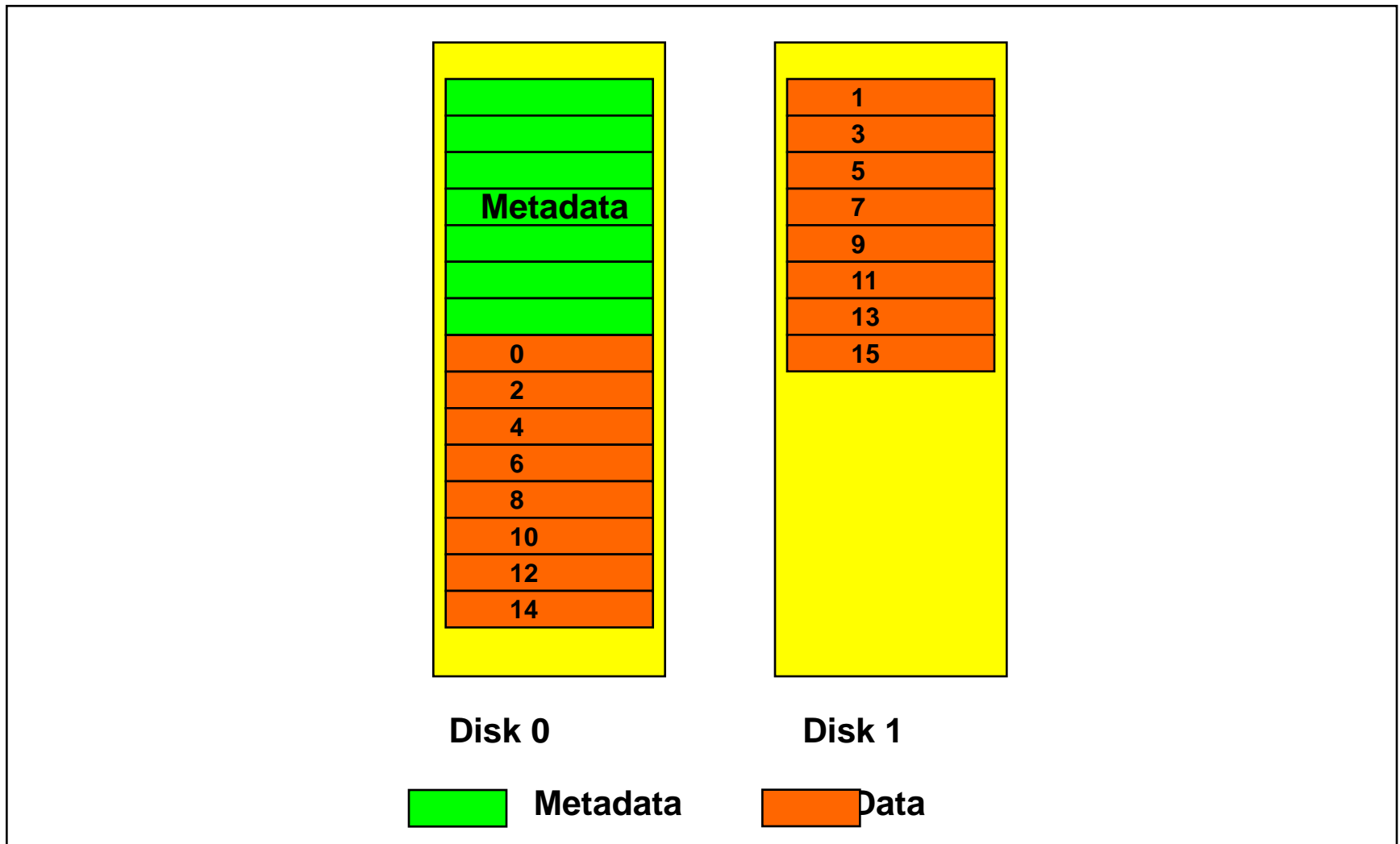
ASM中Extent 的分布-4块磁盘成1个磁盘组

Metadata	2	1	3
	6	5	7
	10	9	11
	14	13	15
	18	17	19
	22	21	23
	26	25	27
0	30	29	31
4			
8			
12			
16			
20			
24			
28			

Disk 0　　　**Disk 1**　　　**Disk 2**　　　**Disk 3**

Metadata　　　Data
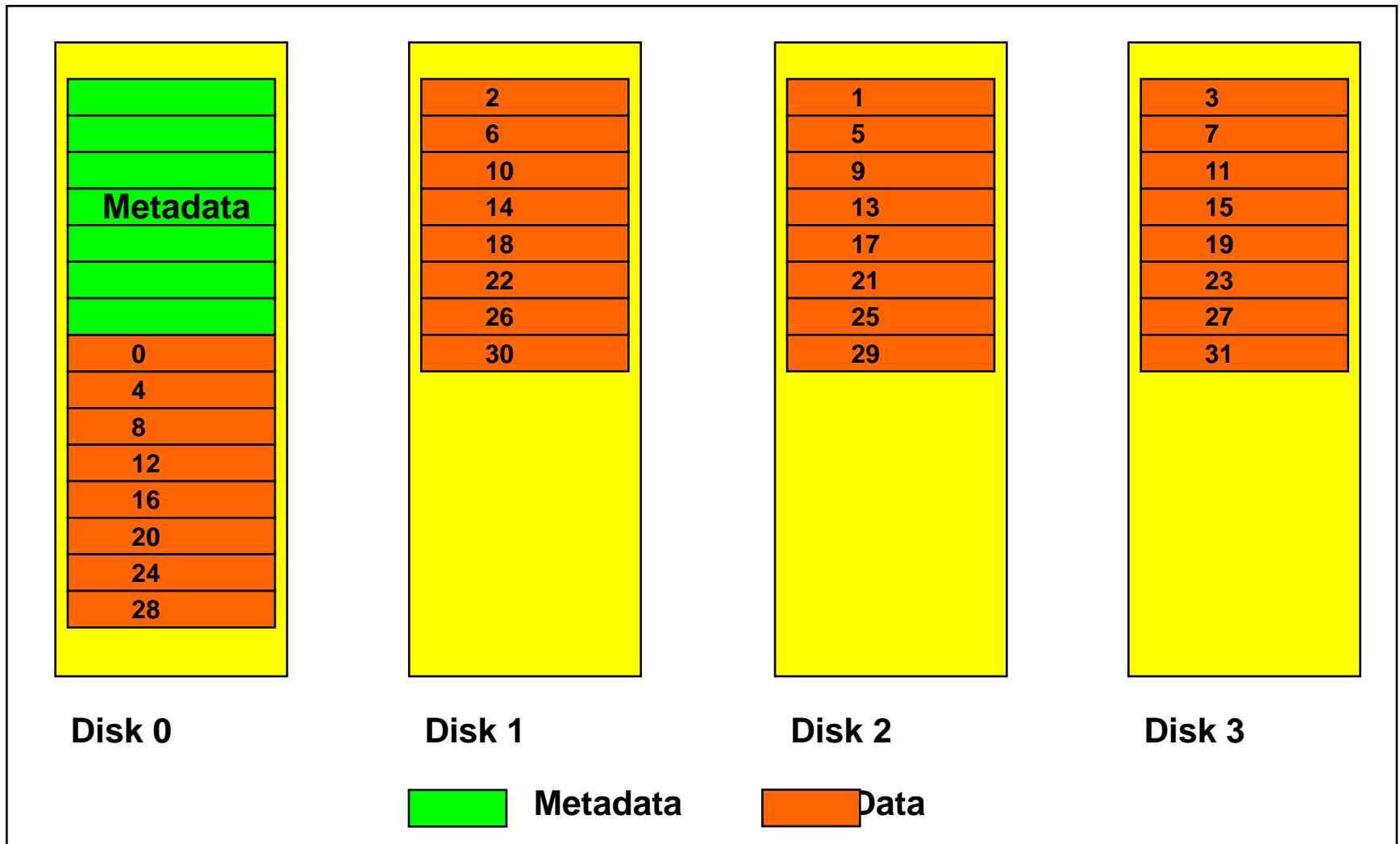
ASM中Extent 的分布-1块大磁和1块小盘组成一个磁盘组
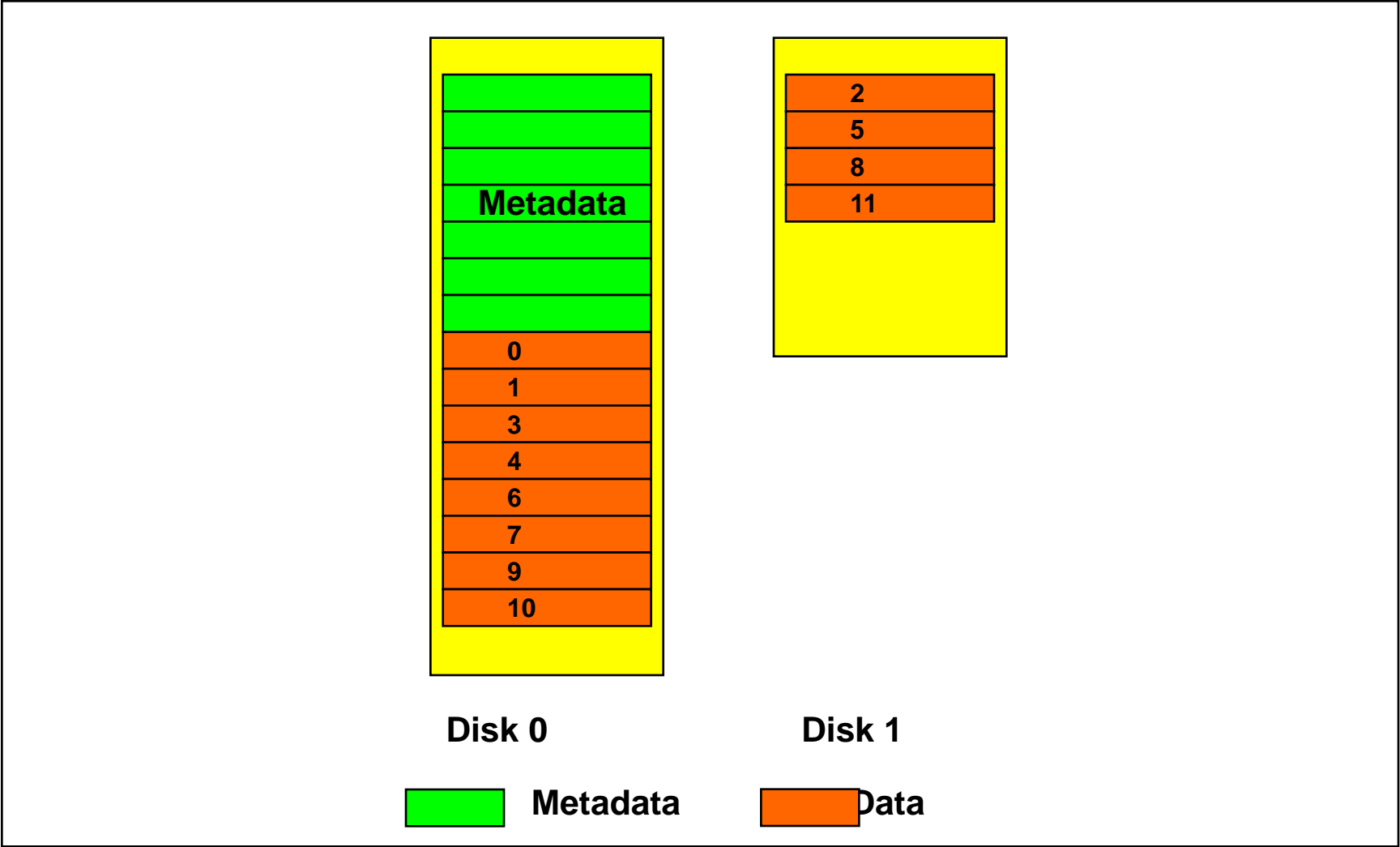
ASM中Extent 的分布-1块大磁和3块小盘组成一个磁盘组

Disk 0　　　　　　Disk 1　　　　　Disk 2　　　　　Disk 4

■ Metadata　　　■ Data
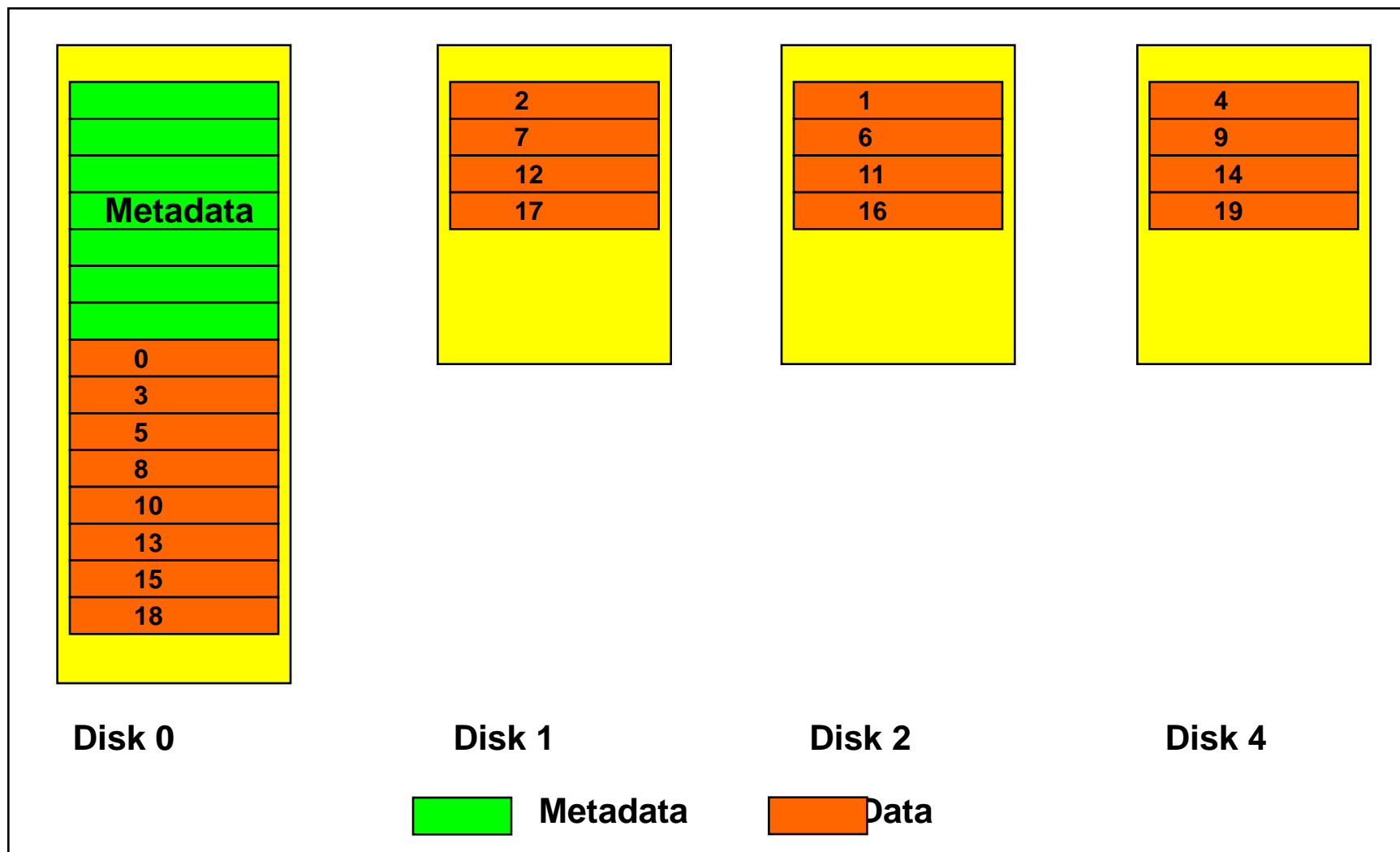
ASM中Extent 的分布-2块大磁和2块小盘组成一个磁盘组

Disk 0	Disk 1	Disk 2	Disk 4

Metadata | Data
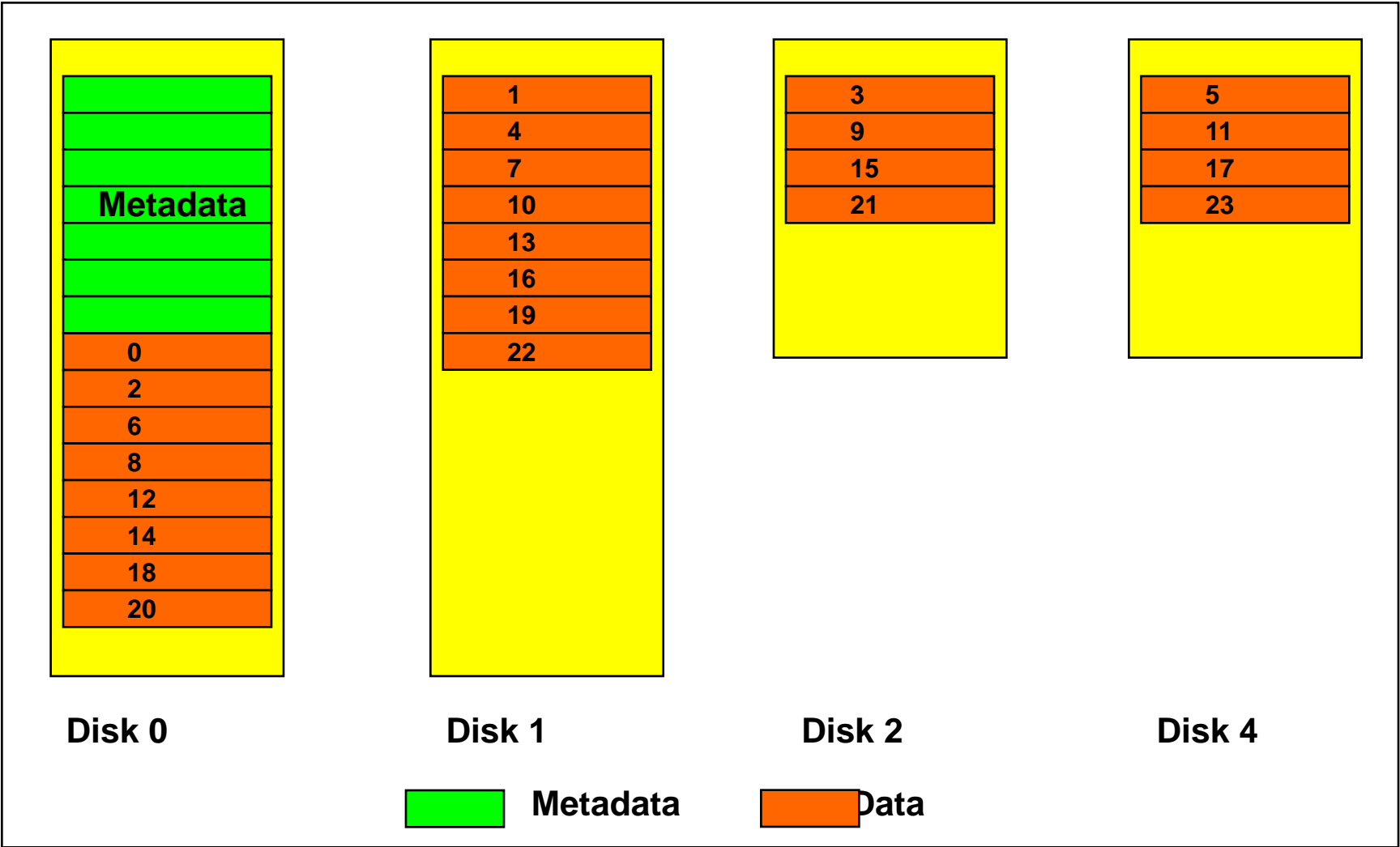
磁盘组的镜像

- 在**extent**级别镜像

- **Primary 和 mirror extent** 不在相同的**ASM**磁盘

- 失败组的概念（**Failure group**）
  - **Default failure group = Disk**

- **External redundancy**
  - 由底层硬件实现**mirroring**

- **Normal redundancy**
  - **2-way mirroring**
- **High redundancy**
  - **3-way mirroring**
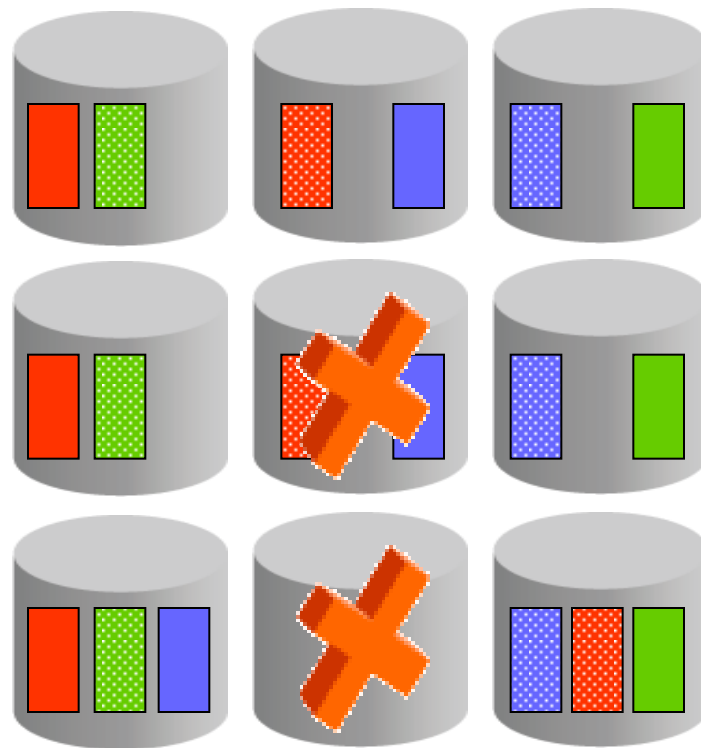
Normal Redundancy
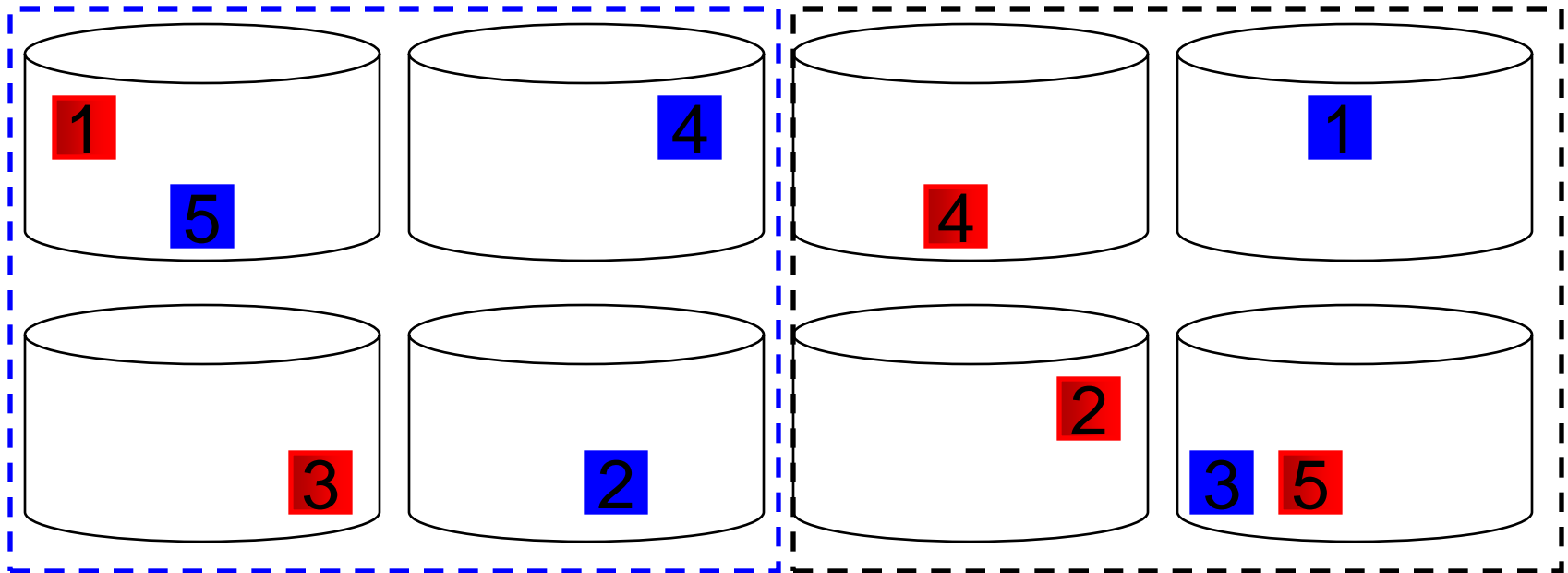
Failure group 1   Failure group 2

Five megabyte normal redundancy file

Primary Extents   # Secondary Extents

Virtual Extents (Extent Sets)
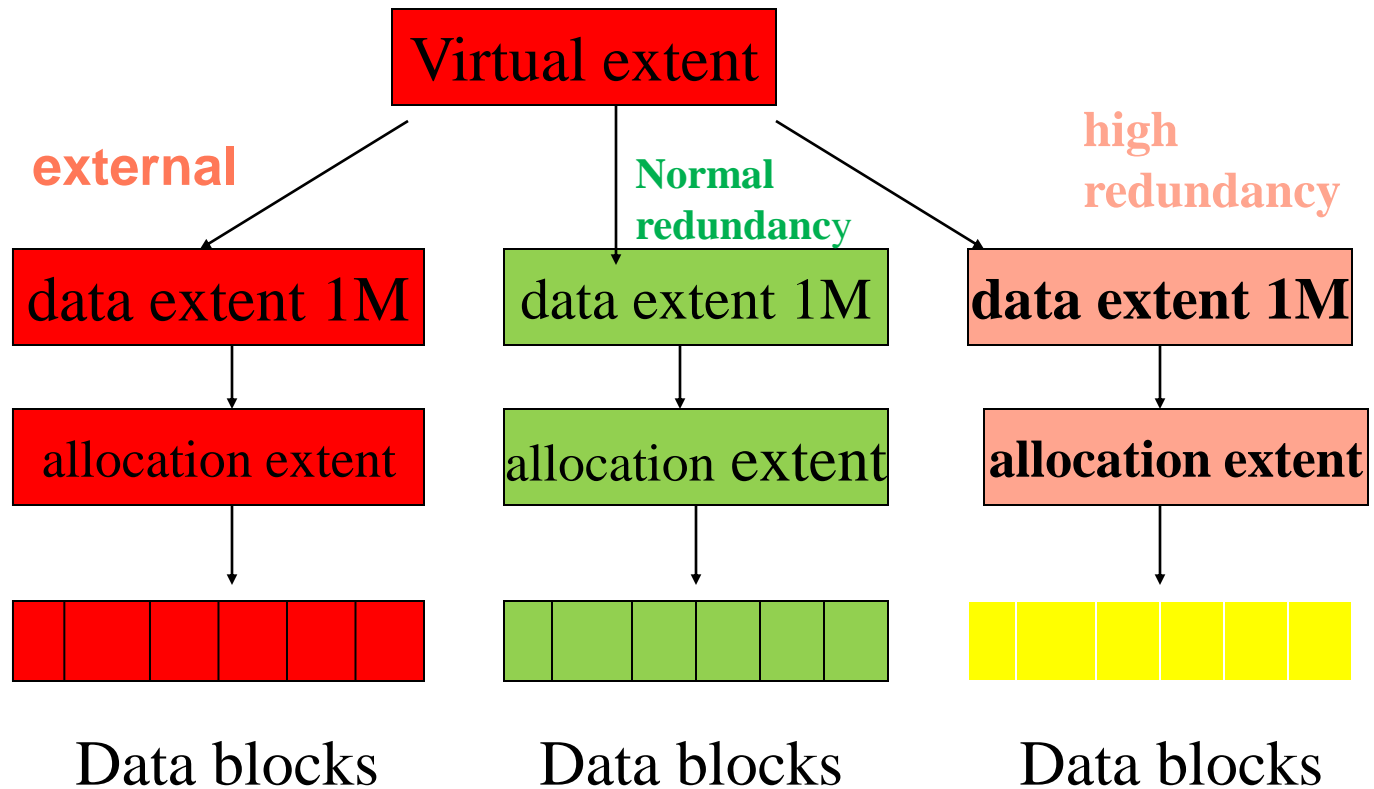
ASM故障组-Normal Redundancy
每个故障组包含1块磁盘

ASM故障组-Normal Redundancy
每个故障组包含2块磁盘

Failure Group 1		Failure Group 2	
0	**1**	**0**	**1**
3	**2**	**3**	**2**
4	**5**	**4**	**5**
7	**6**	**7**	**6**
0	**1**	**0**	**1**
3	**2**	**3**	**2**
4	**5**	**4**	**5**
7	**6**	**7**	**6**
8	**9**	**8**	**9**
11	**10**	**11**	**10**
12	**13**	**12**	**13**
15	**14**	**15**	**14**
16	**17**	**16**	**17**
19	**18**	**19**	**18**
20	**21**	**20**	**21**
23	**22**	**23**	**22**
Disk 0	Disk 1	Disk 2	Disk 3

Metadata	Primary	Secondary
Data	Primary	Secondary
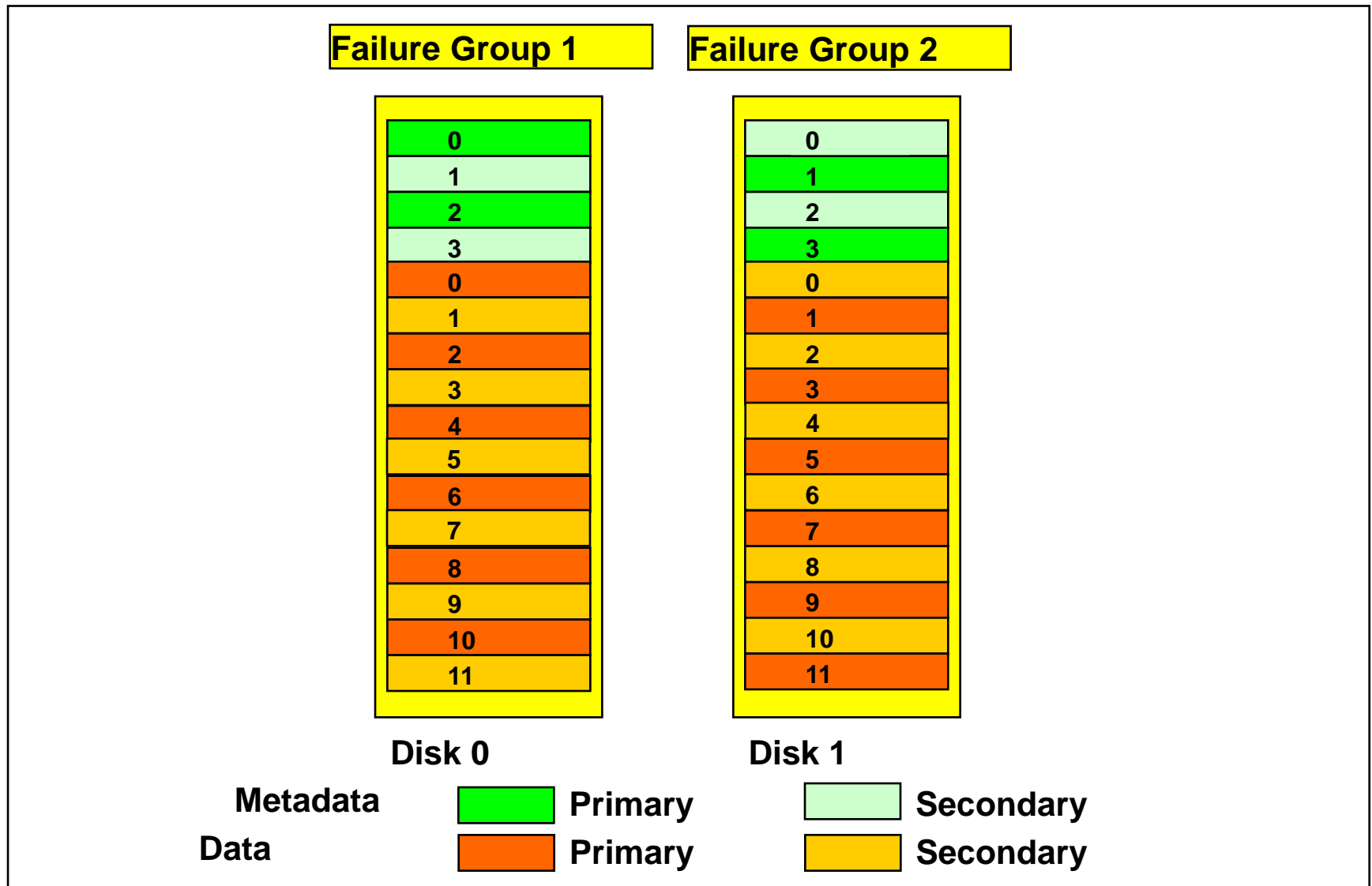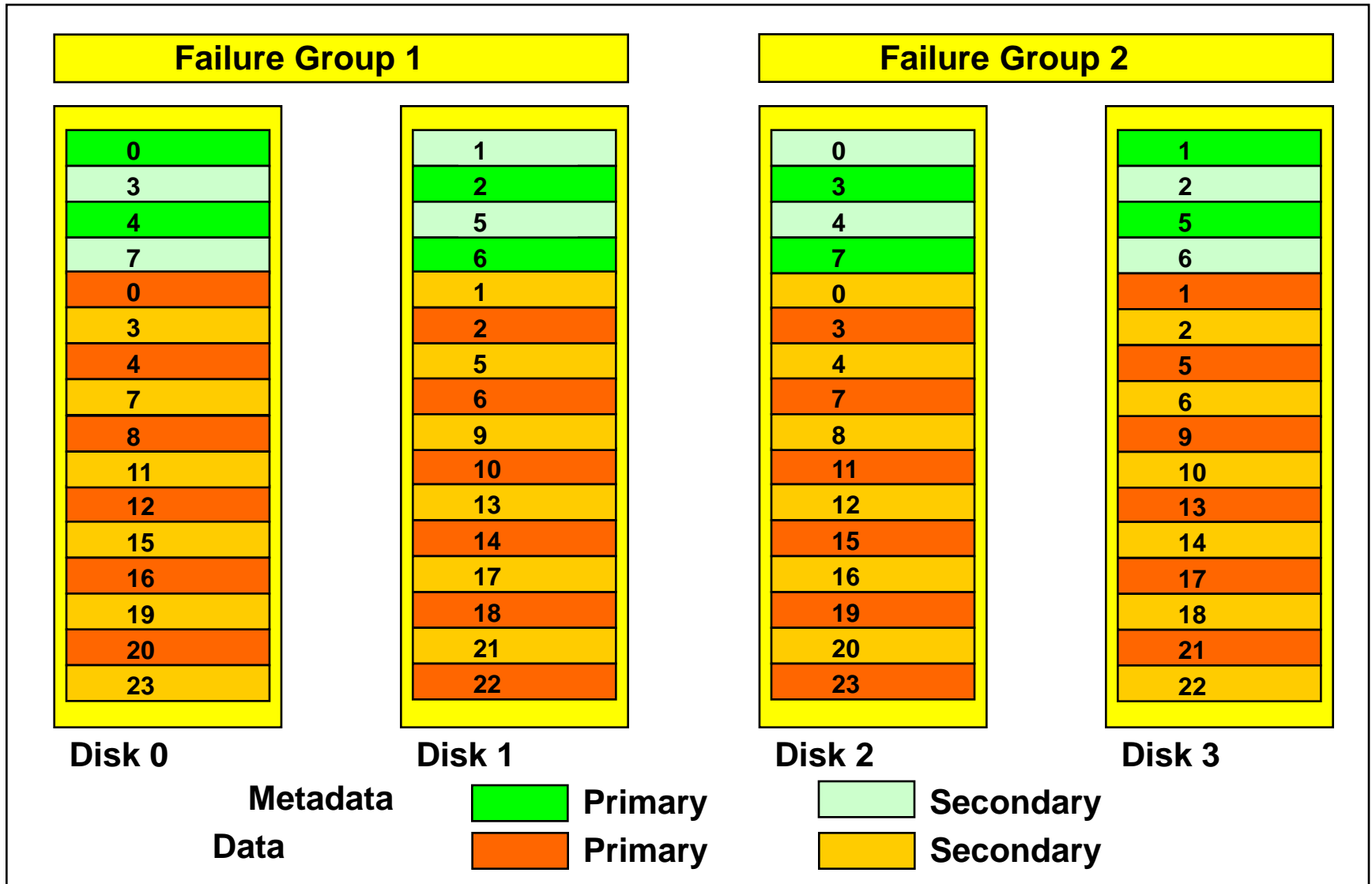
ASM故障组-High Redundancy
每个故障组包含1块磁盘

伙伴的概念（**Partnering**）

Failure Group 1

Failure Group 2

Stripes

Allocation Unit

Failure Group 3

Failure Group 4

9-59
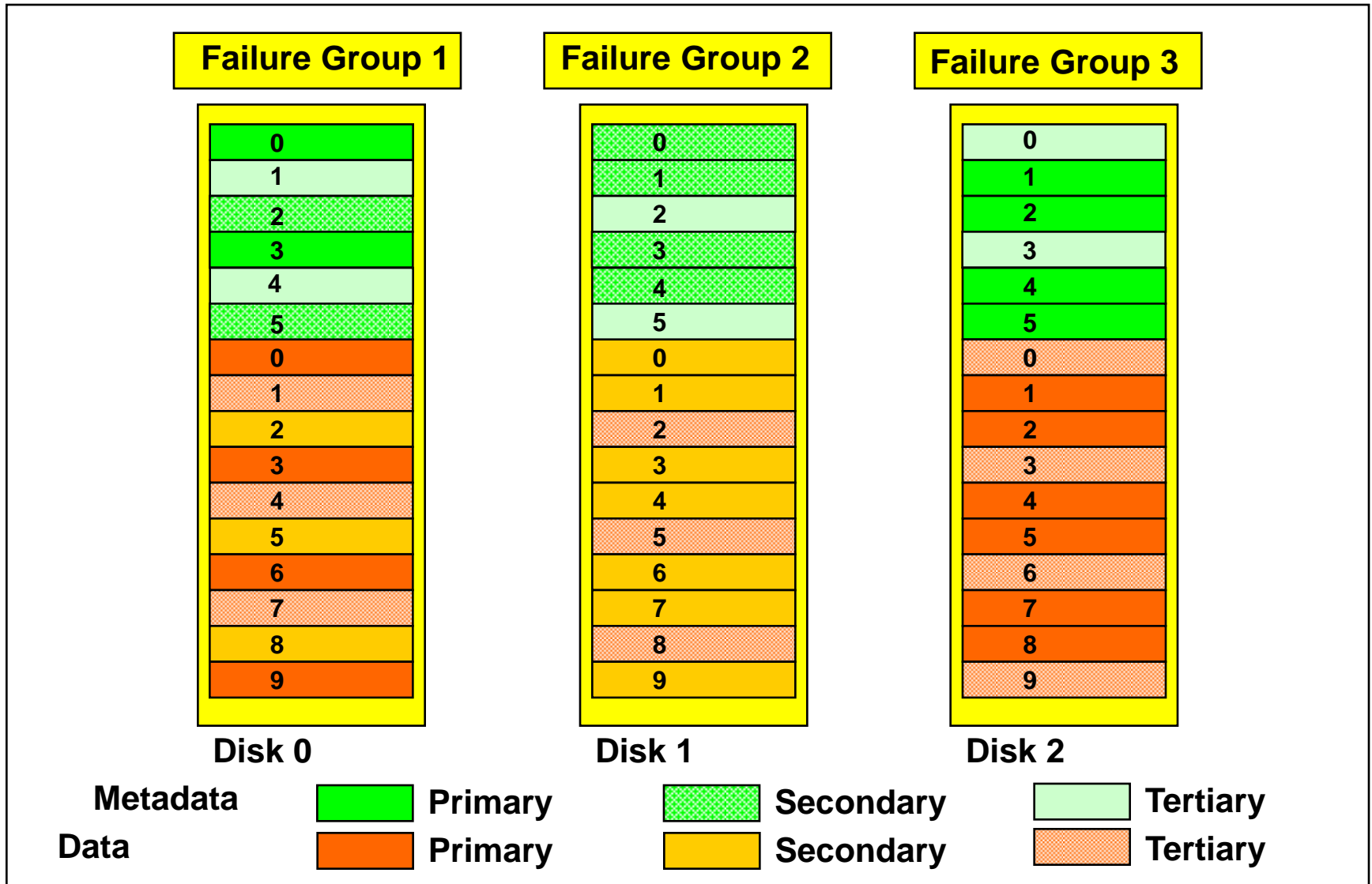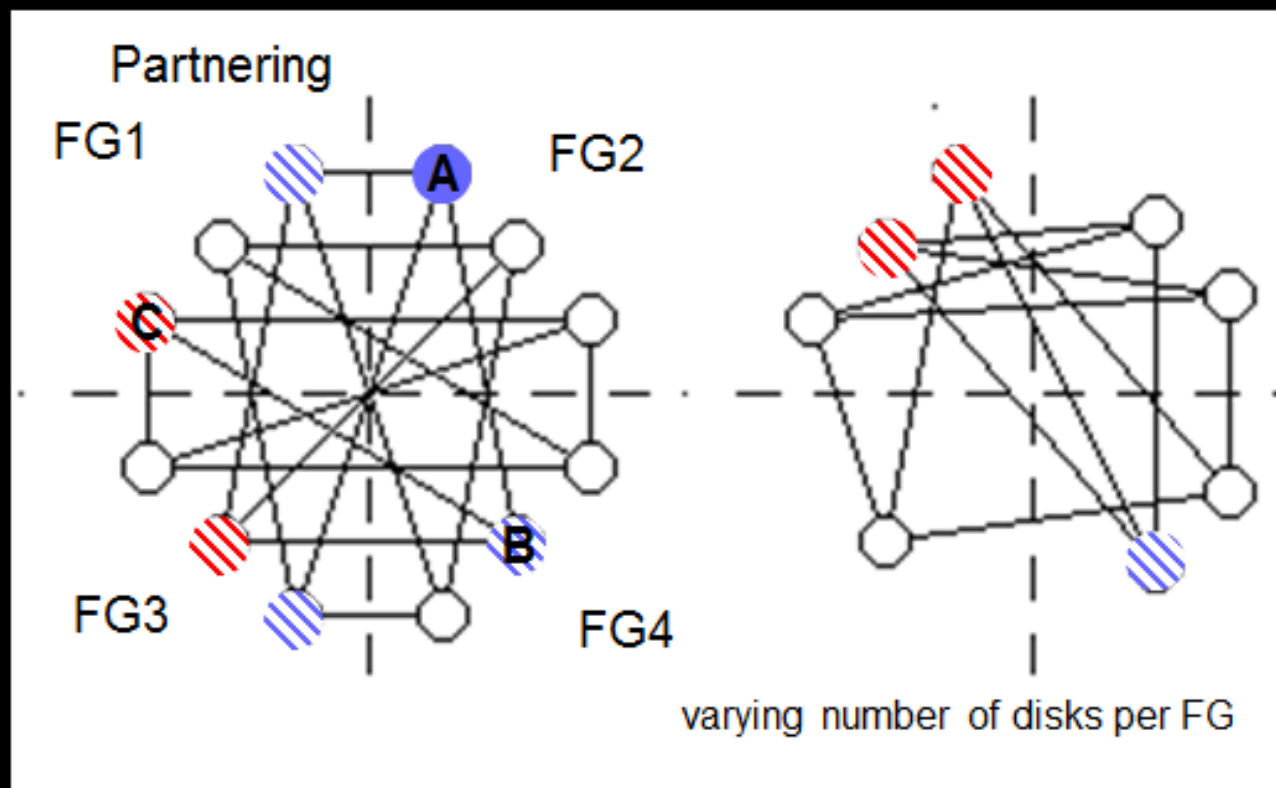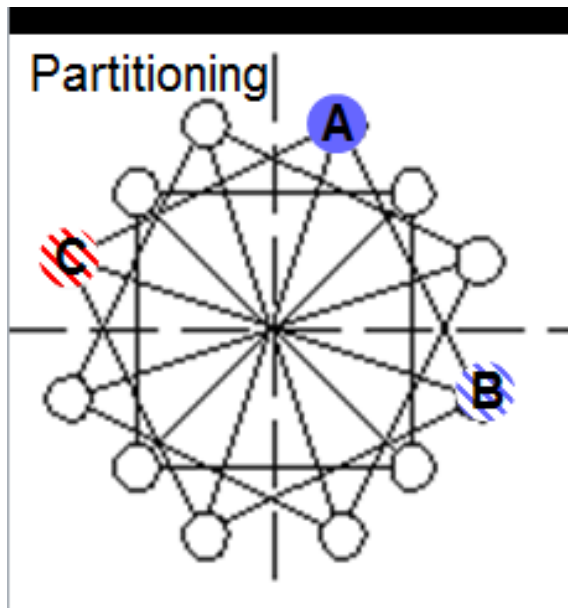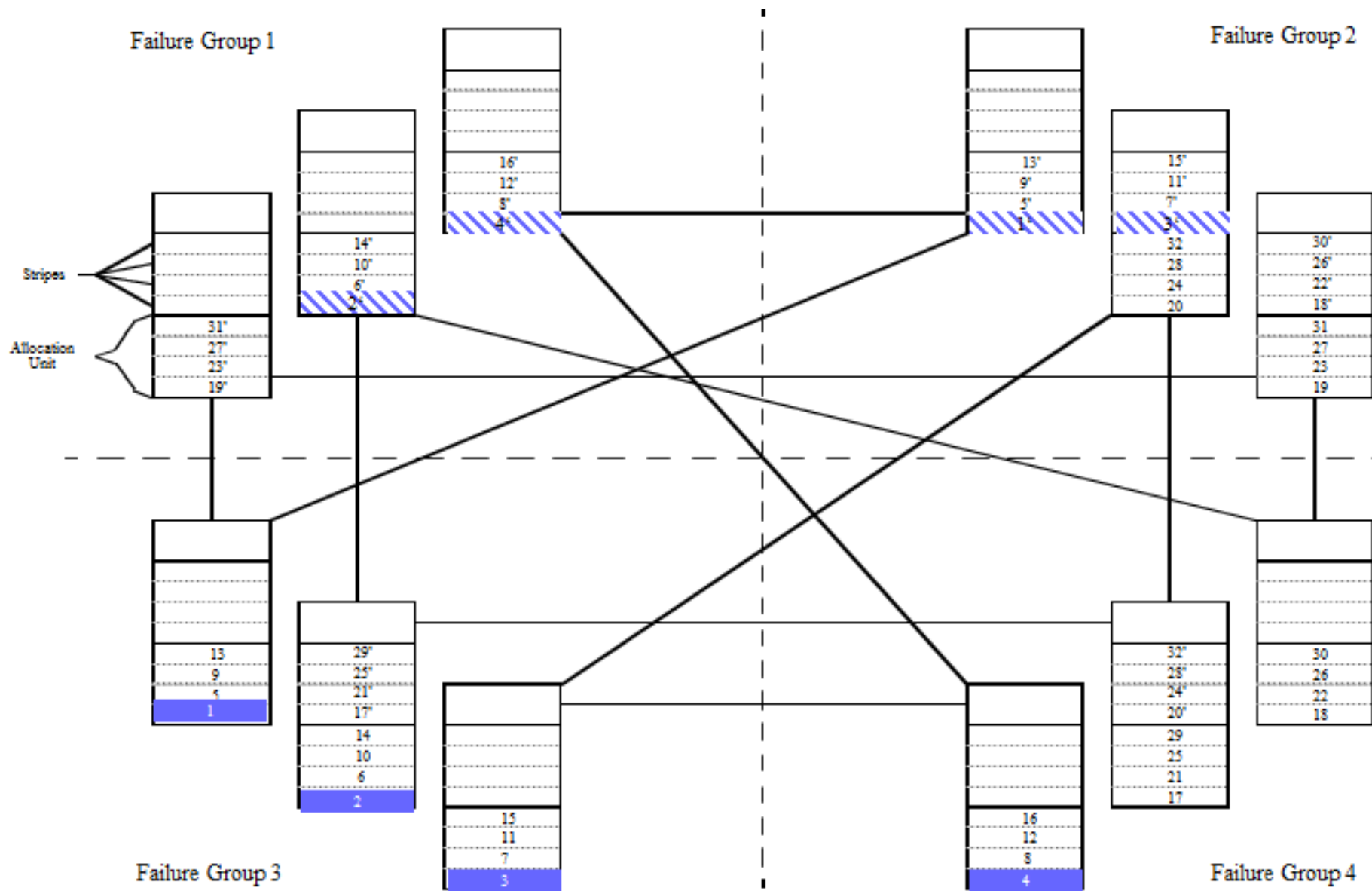
How ASM allocates redundant copies?

FG1
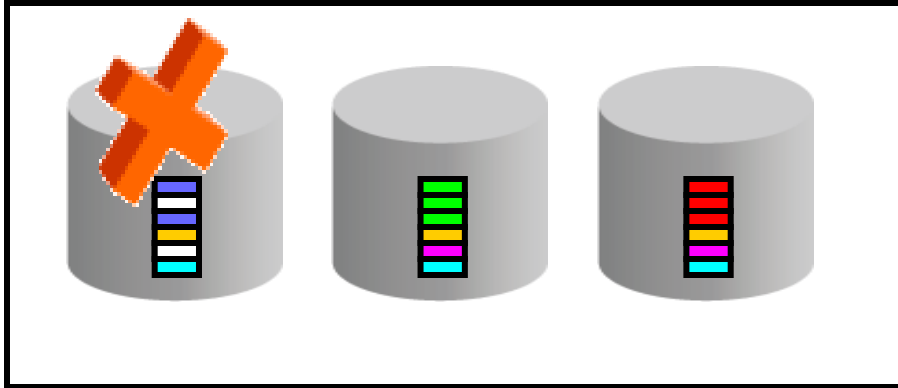
FG2
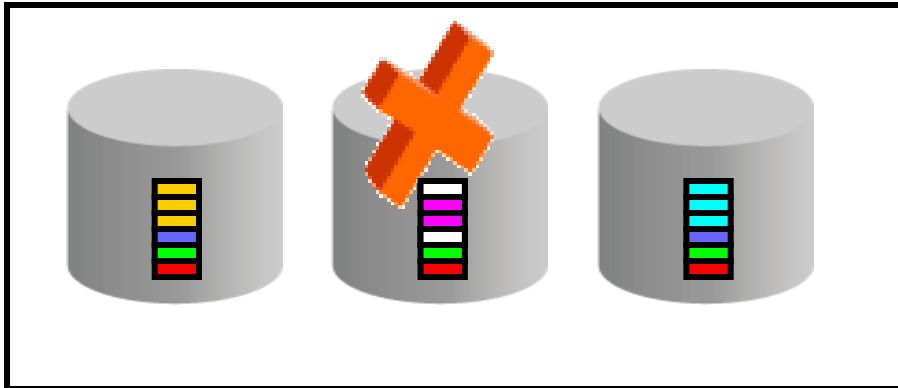
ASM partnering concept

FG1

FG2
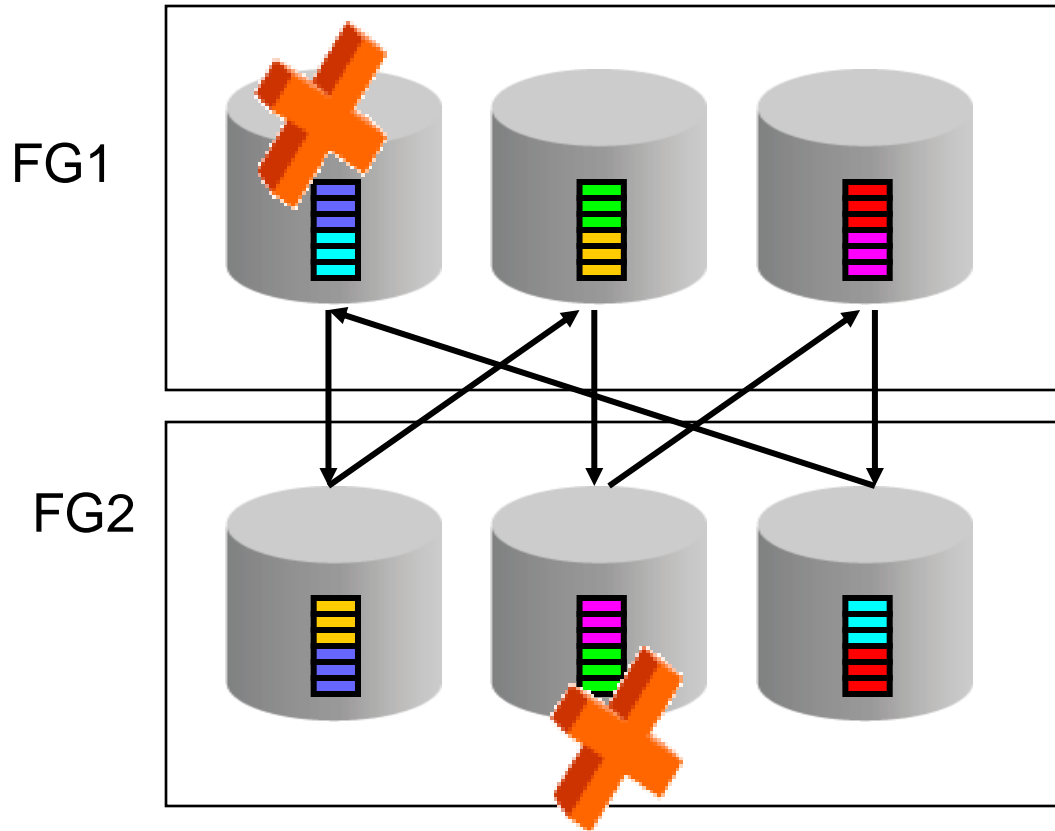
ASM重平衡（**Rebalancing**）
Adding disks - 1 disks to 2 disks

Disk 0   Disk 1

Metadata   Data
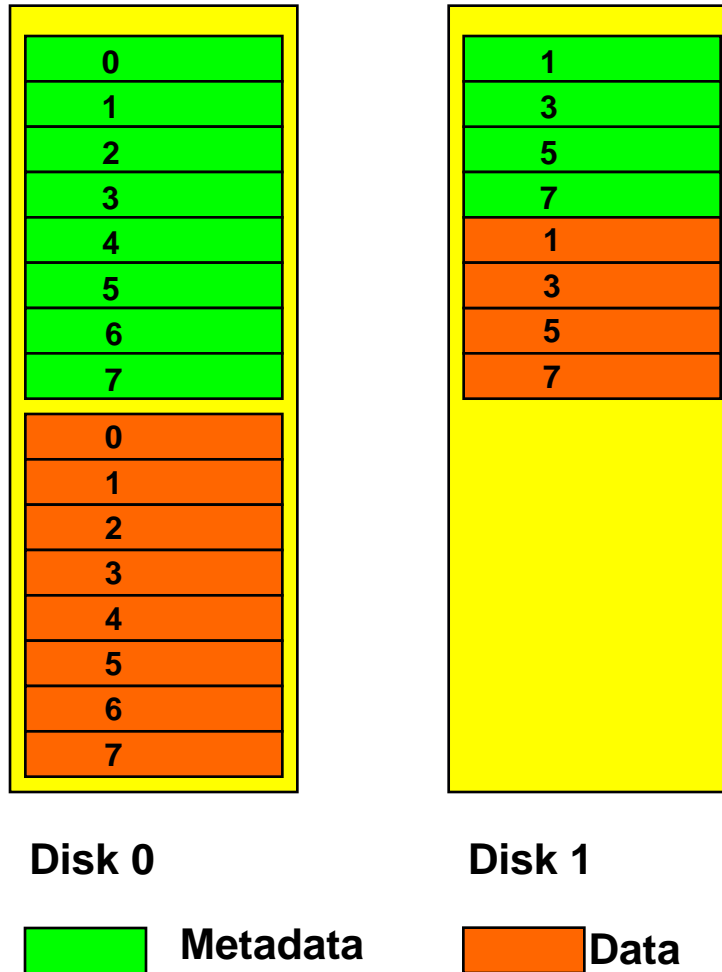
ASM重平衡（Rebalancing）
Adding disks - 1 disks to 4 disks

| Disk 0 | Disk 1 | Disk 2 | Disk 3 |

Disk 0 (Metadata): 0, 1, 2, 3, 4, 5, 6, 7
Disk 0 (Data): 0, 1, 2, 3, 4, 5, 6, 7

Disk 1 (Metadata): 2, 6
Disk 1 (Data): 2, 6

Disk 2 (Metadata): 1, 5
Disk 2 (Data): 1, 5

Disk 3 (Metadata): 0, 4
Disk 3 (Data): 0, 4

Metadata   **Data**
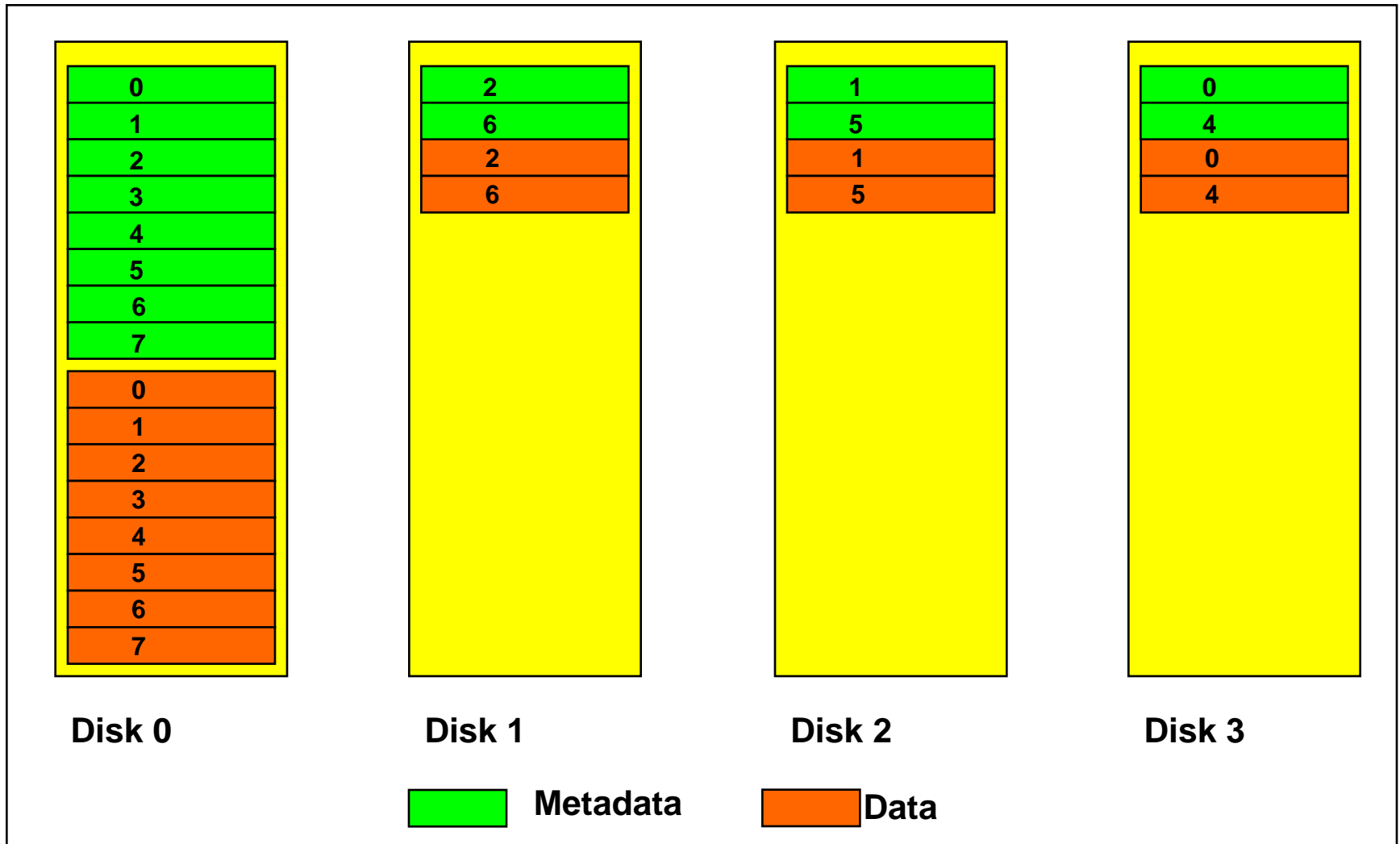
ASM重平衡（Rebalancing）
Adding disks - 2 disks to 3 disks
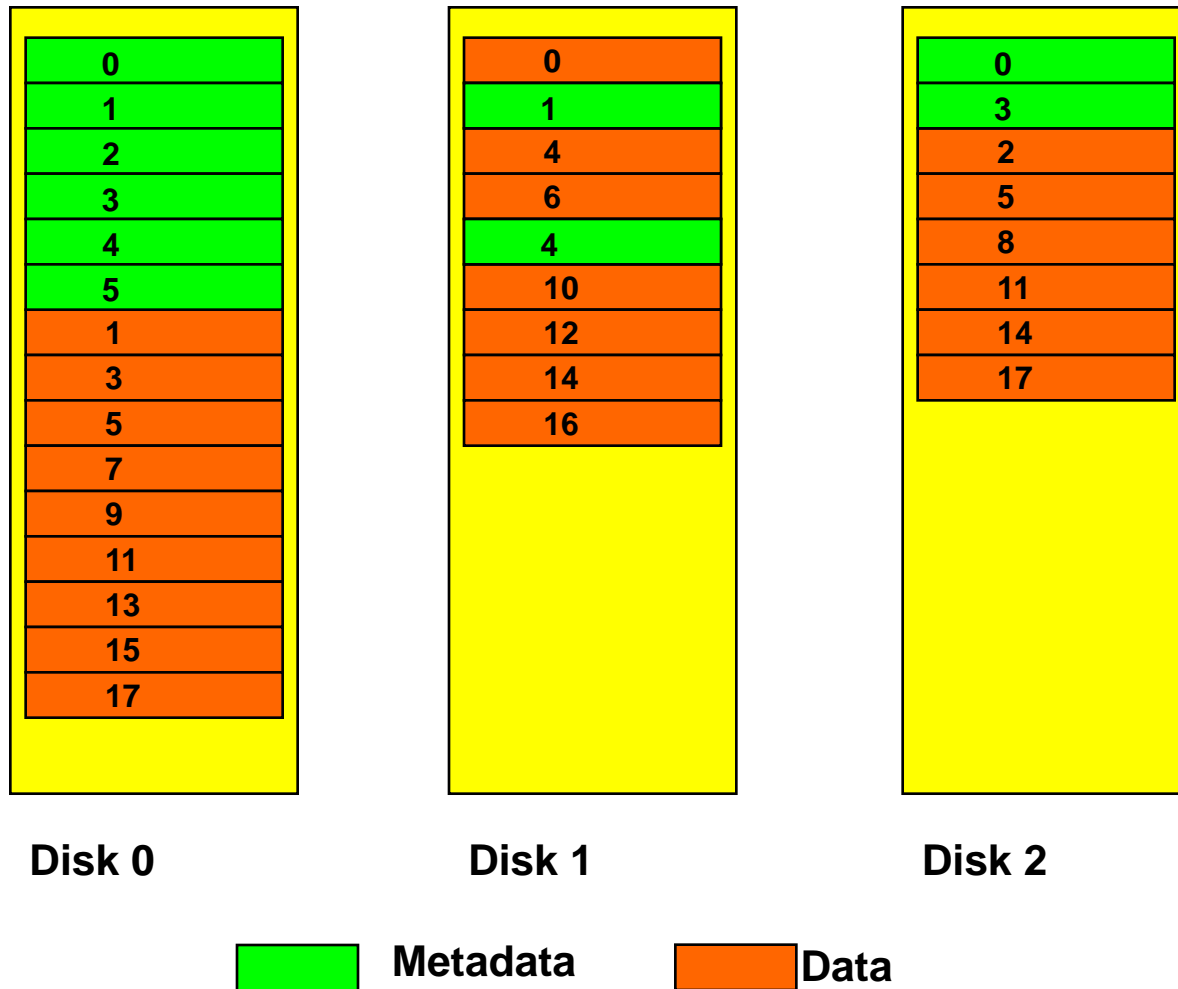
ASM重平衡（**Rebalancing**）
Adding disks - 2 disks to 4 disks

Disk 0	Disk 1	Disk 2	Disk 3
0	2	1	0
1	2	5	4
2	6	1	0
3	6	5	4
4	8	9	8
5	10	13	12
6	12		
7	14		
1			
3			
5			
7			
9			
11			
13			
15			

Metadata **Data**
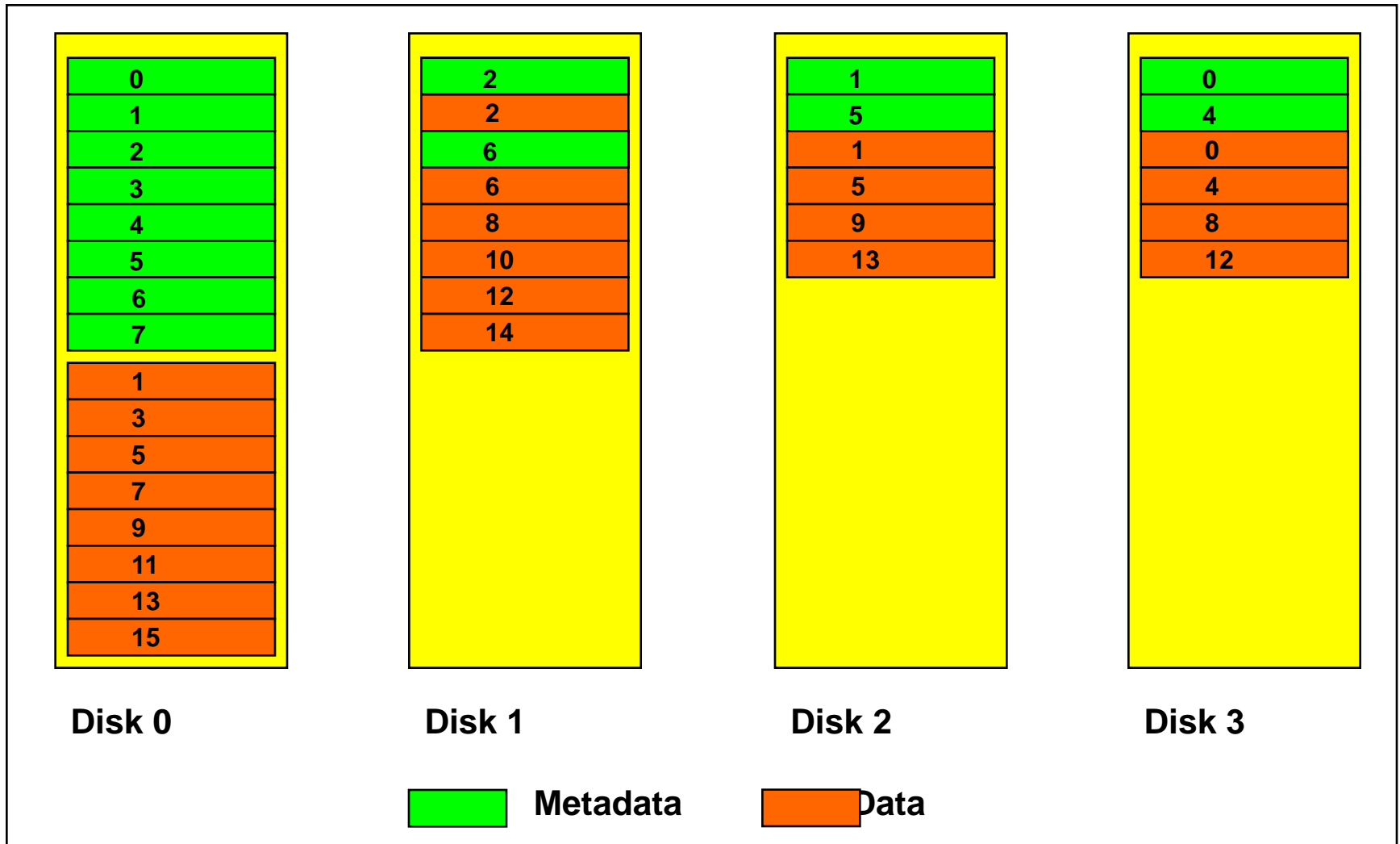
ASM重平衡（**Rebalancing**）
Dropping disks - 3 disks to 1 disk

Disk 0

Disk 1

Disk 2

Metadata **Data**
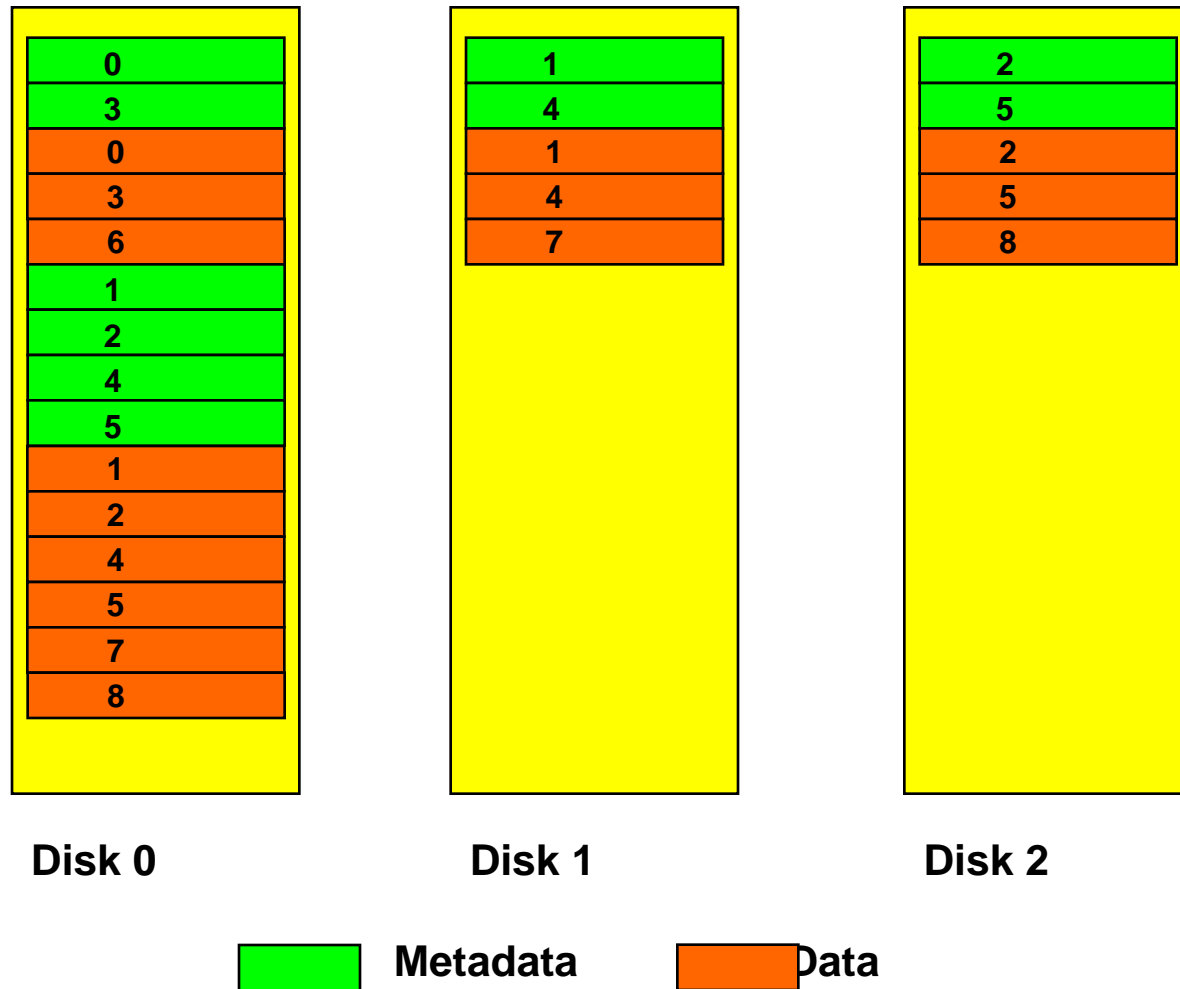
ASM重平衡（Rebalancing）
Moving disks - 2 disks to 2 disks

Disk 0	Disk 1	Disk 2
0	0	0
1	2	2
2	4	4
3	6	6
4	8	0
5	10	2
6	12	4
7	14	6
1		8
3		10
5		12
7		14
9		
11		
13		
15		

Metadata   **Data**
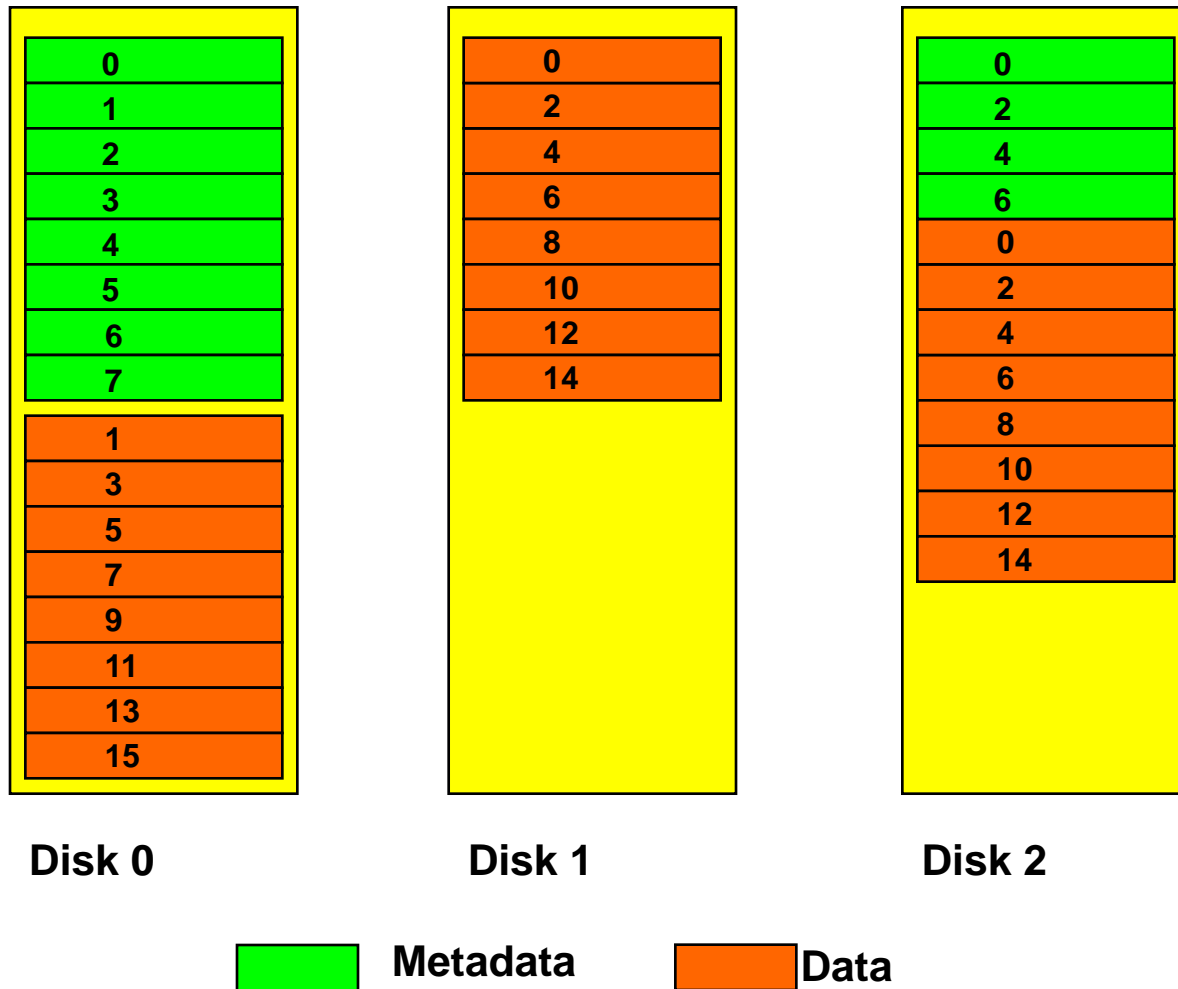
ASM 文件

```
CREATE TABLESPACE sample DATAFILE '+dgroupA';
```

数据库文件

RMAN

自动创建
ASM 文件

ASM 文件在磁盘组 dgroupA 中自动分布
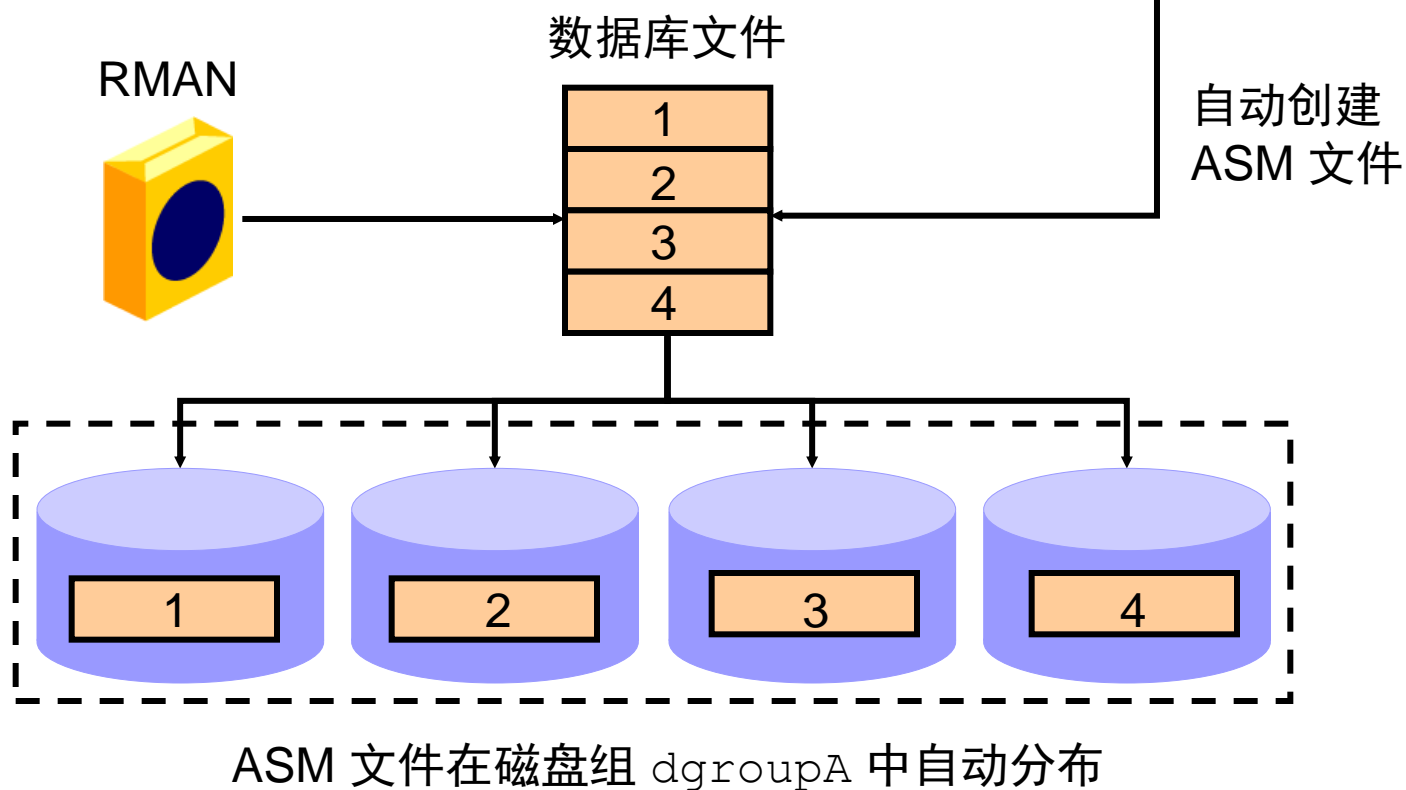
ASM Disk directory tree
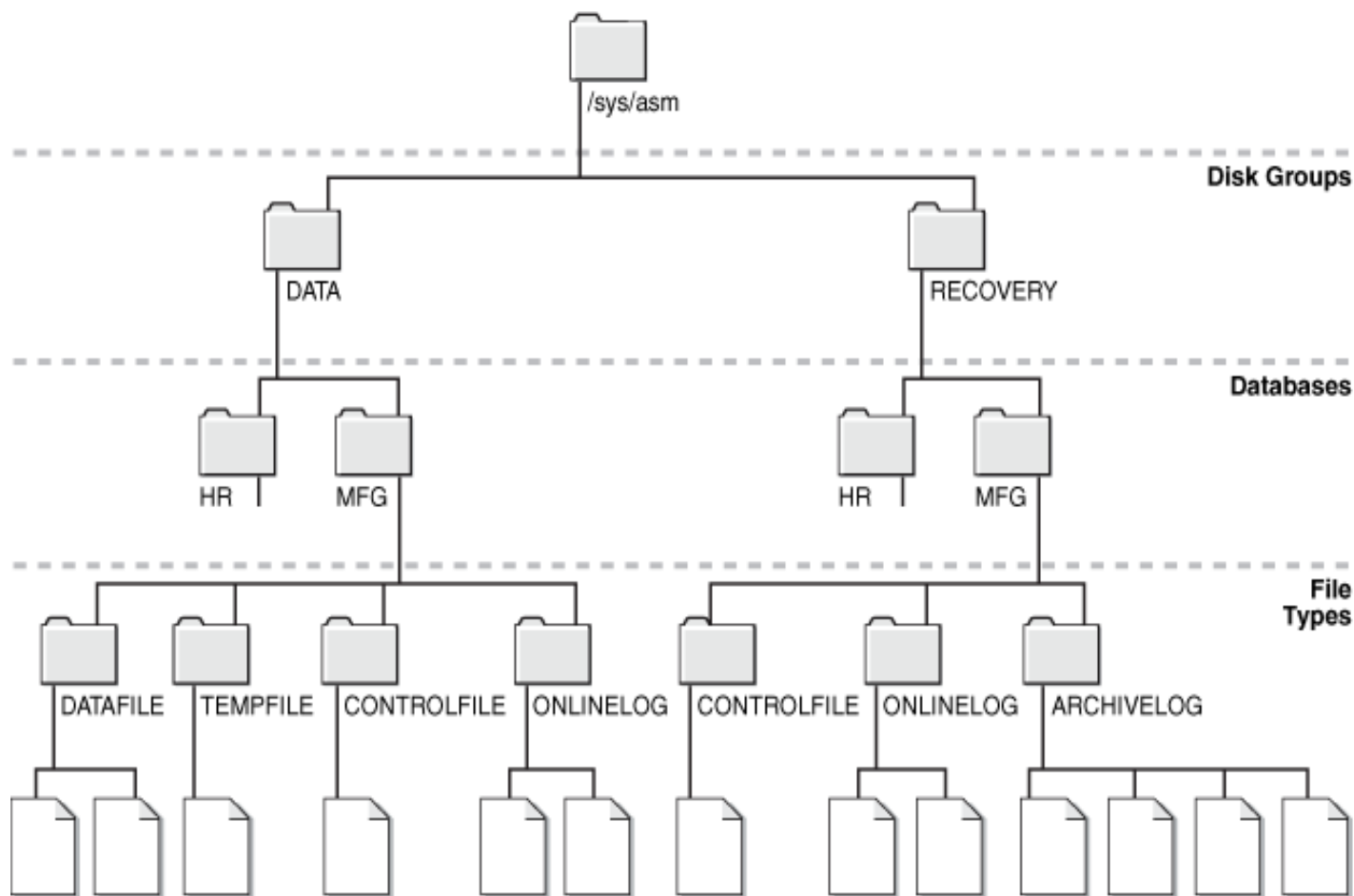
X$KFFXP

Maps file extents to allocation units

Only populated in ASM instance

Columns include

Column Name	Description
GROUP_KFFXP	Disk Group Number
NUMBER_KFFXP	File Number
COMPOUND_KFFXP	Disk Group Number || File Number
INCARN_KFFXP	Incarnation Number
PXN_KFFXP	Physical Extent Number (within file)
XNUM_KFFXP	Logical Extent Number (within file)
LXN_KFFXP	0=primary, 1=first mirror, 2=second mirror
DISK_KFFXP	Disk Number
AU_KFFXP	Allocation Unit Number (within disk)
SIZE_KFFXP	Size (# allocation units)

ASM Metadata

Metadata Header

Partner Status Table

File Directory

Disk Directory

Active Change Directory

Continuing Operations Directory

Template Directory

Alias Directory

Disk Header

Free Space Table

Allocation Table

Metadata Header
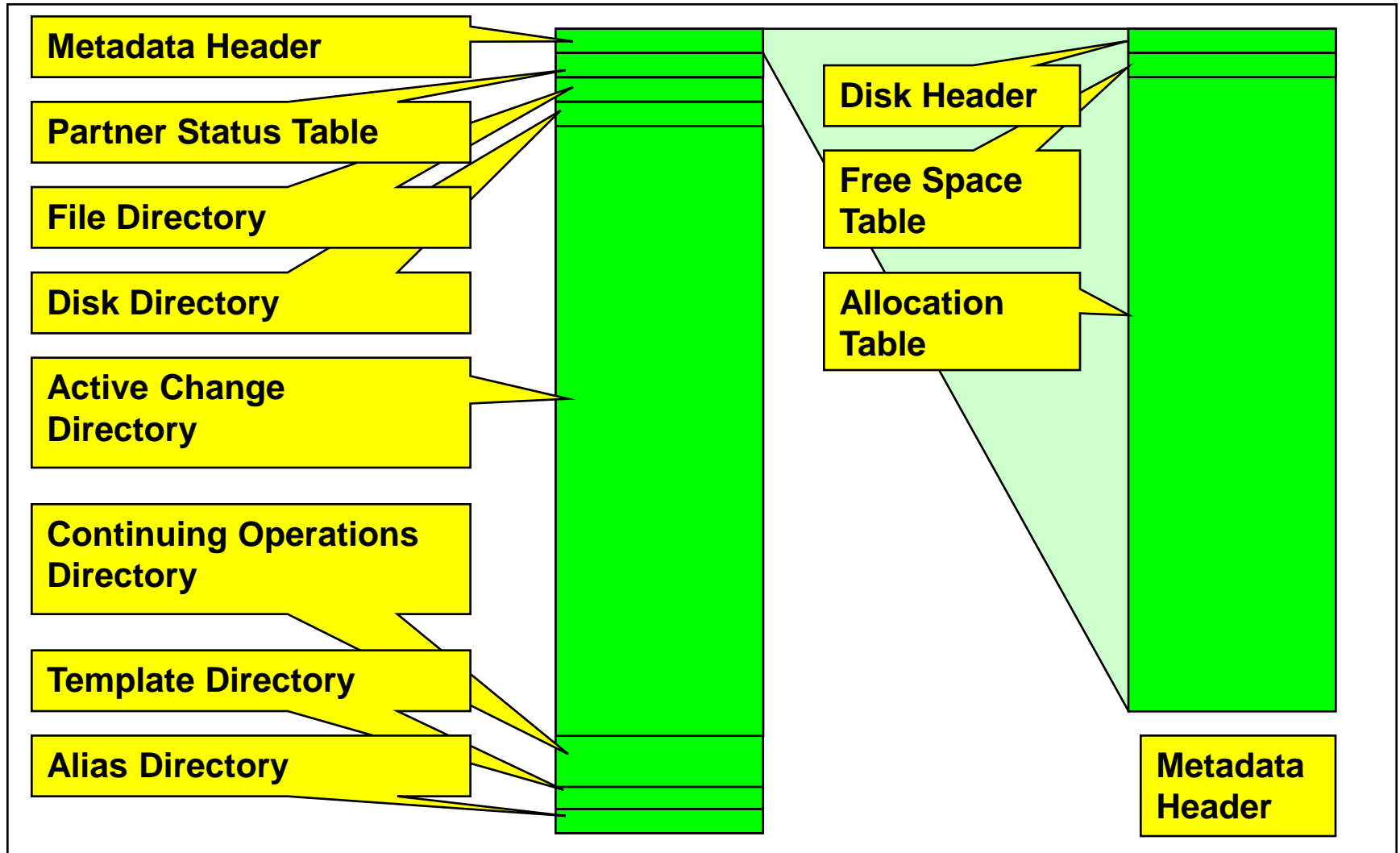
ASM Metadata

Metadata is stored in first 256 files in ASM disk group

- **Space is initially allocated when disk group is created**
- **Can be subsequently extended**

Metadata allocation units are divided into blocks

- **Each block is 4096 bytes**
- **Block size specifed using _asm_blksize**

Metadata files include

File#	Description
0	**Metadata Header**
1	**File Directory**
2	**Disk Directory**
3	**Active Change Directory**
4	**Continuing Operations Directory**
5	**Template Directory**
6	**Alias Directory**
9	**Attribute directory (optional)**
12	**Staleness registry (optional)**

kfed - **K**ernel **F**ile metadata **Ed**itor

11.1以前：

```
cd $ORACLE_HOME/rdbms/lib

modify ins_rdbms.mk

make  -f  ins_rdbms.mk ikfed
```

```
$ORACLE_HOME/bin/kfed -h
as/mlib                ASM Library [asmlib='lib']
aun/um                 AU number to examine or update [AUNUM=number]
aus/z                  Allocation Unit size in bytes [AUSZ=number]
blkn/um                Block number to examine or update [BLKNUM=number]
blks/z                 Metadata block size in bytes [BLKSZ=number]
ch/ksum                Update checksum before each write [CHKSUM=YES/NO]
cn/t                   Count of AUs to process [CNT=number]
d/ev                   ASM device to examine or update [DEV=string]
o/p                    KFED operation type
                       [OP=READ/WRITE/MERGE/NEW/FORM/FIND/STRUCT]
p/rovnm                Name for provisioning purposes [PROVNM=string]
s/eek                  AU number to seek to [SEEK=number]
te/xt                  File name for translated block text [TEXT=string]
ty/pe                  ASM metadata block type number [TYPE=number]
```

```
for (( f = 12544 ; f < 12544 + 256 ; f++ ))
do
      kfed op=read blkn=$f dev='/dev/oracleasm/disks/VOL1' > blk${f}
done
```

ASM File Allocation

- ASM files分为多个小片，每一片叫做一个DE(Data extents)
- 每个ASM files都跨越在同一个磁盘组的多个磁盘上
- 将extents存储到磁盘上的位置叫做AU(Allocation Unit)

www.lunar2013.com
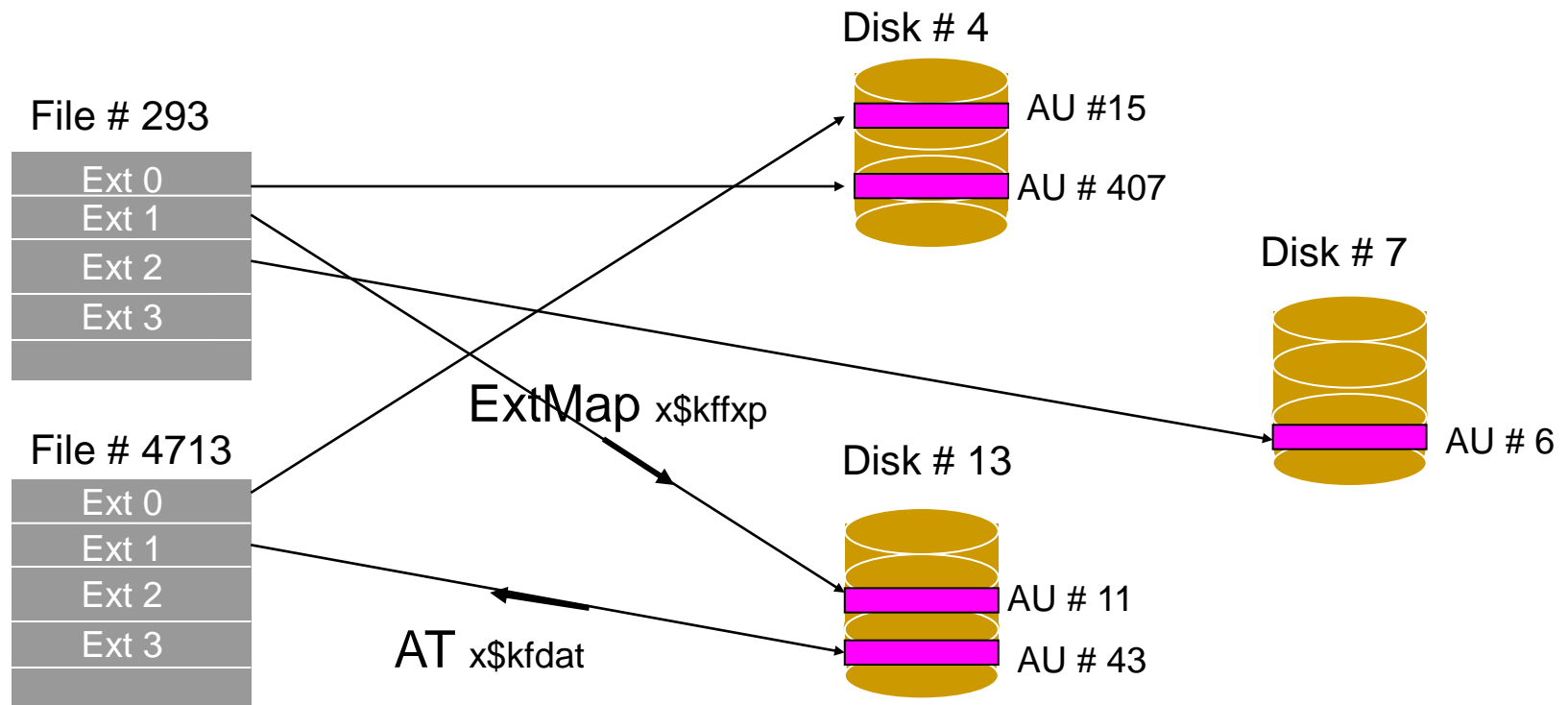
X$KFFXP Extent Maps
ASMFILE# 1

one row per extent over all files

GROUP_KFFXP   **diskgroup number (1 - 63)**

NUMBER_KFFXP   **file number for the extent**

COMPOUND_KFFXP **(group_kffxp << 24) + file #**

INCARN_KFFXP   **file incarnation number**

PXN_KFFXP   **physical extent number**

XNUM_KFFXP   **extent number**

- **bit 31 set if indirect**

LXN_KFFXP   **logical extent number**

DISK_KFFXP   **disk on which AU is located**

AU_KFFXP   **AU number on disk of AU**

Datafile size 3MB => 4 AU's (4MB)

X$KFDAT – Allocation Tables

One row per AU over all disks

`GROUP_KFDAT`   **diskgroup # (1 - 63)**

`NUMBER_KFDAT`   **disk # for the AU**

`COMPOUND_KFDAT` **(**`group_kfdat` **<< 24) +** `number_kfdat`

`AUNUM_KFDAT`   **AU # within disk**

`V_KFDAT`   **Valid: Y if allocated, N if unused**

`FNUM_KFDAT`   **Meaningless unless** `V_KFDAT` **set**
- **File number using this AU**
- **0 for physical/PST AUs**
- **-1 if past end of disk**

`I_KFDAT`   **Indirect: Meaningless unless** `V_KFDAT` **set**
- **Y if indirect extent using the AU**
- **N if direct extent using this AU**

`XNUM_KFDAT`   **Physical extent number using AU**

`RAW_KFDAT`   **Raw 8-byte contents of extent pointer**
- **useful when** `V_KFDAT` **not set to check for corruptions**

Metadata Block Layout

e.g. Diskheader

e.g. PST

kfbh.endian:	1 ; 0x000: 0x01	
kfbh.hard:	130 ; 0x001: 0x82	**32 byte**
kfbh.type:	**1 ; 0x002: KFBTYP_DISKHEAD**	**BlockHeader**
kfbh.datfmt:	1 ; 0x003: 0x01	
kfbh.block.blk:	0 ; 0x004: T=0 NUMB=0x0	
kfbh.block.obj:	2147483648 ; 0x008: TYPE=0x8 NUMB=0x0	**kfbh**
kfbh.check:	3768521839 ; 0x00c: 0xe09f146f	
kfbh.fcn.base:	1696 ; 0x010: 0x000006a0	
kfbh.fcn.wrap:	0 ; 0x014: 0x00000000	
kfbh.spare1:	0 ; 0x018: 0x00000000	**kfbh**
kfbh.spare2:	0 ; 0x01c: 0x00000000	

kfbh.endian:	1 ; 0x000: 0x01
kfbh.hard:	130 ; 0x001: 0x82
kfbh.type:	**17 ; 0x002: KFBTYP_PST_META**
kfbh.datfmt:	1 ; 0x003: 0x01
kfbh.block.blk:	256 ; 0x004: T=0 NUMB=0x100
kfbh.block.obj:	2147483648 ; 0x008: TYPE=0x8 NUMI
kfbh.check:	2396453971 ; 0x00c: 0x8ed6fc53
kfbh.fcn.base:	0 ; 0x010: 0x00000000
kfbh.fcn.wrap:	0 ; 0x014: 0x00000000
kfbh.spare1:	0 ; 0x018: 0x00000000
kfbh.spare2:	0 ; 0x01c: 0x00000000

kfdhdb.driver.provstr:	ORCLDISK ; 0x000: length=8	
kfdhdb.driver.reserved[0]:	0 ; 0x008: 0x00000000	
kfdhdb.driver.reserved[1]:	0 ; 0x00c: 0x00000000	
kfdhdb.driver.reserved[2]:	0 ; 0x010: 0x00000000	
kfdhdb.driver.reserved[3]:	0 ; 0x014: 0x00000000	
kfdhdb.driver.reserved[4]:	0 ; 0x018: 0x00000000	**kfdhdb**
kfdhdb.driver.reserved[5]:	0 ; 0x01c: 0x00000000	
kfdhdb.compat:	168820736 ; 0x020: 0x0a100000	
kfdhdb.dsknum:	0 ; 0x024: 0x0000	
kfdhdb.grptyp:	2 ; 0x026: KFDGTP_NORMAL	
kfdhdb.hdrsts:	3 ; 0x027: KFDHDR_MEMBER	
kfdhdb.dskname:	ASMDG3_0000 ; 0x028: length=11	
kfdhdb.grpname:	ASMDG3 ; 0x048: length=6	
kfdhdb.fgname:	ASMDG3_0000 ; 0x068: length=11	
kfdhdb.capname:	; 0x088: length=0	

kfdpHdrB.time.hi:	32870224 ; 0x000: HOUR=0x10 DA
kfdpHdrB.time.lo:	238088192 ; 0x004: USEC=0x0 MSE
kfdpHdrB.last:	5 ; 0x008: 0x00000005
kfdpHdrB.next:	5 ; 0x00c: 0x00000005
kfdpHdrB.copyCnt:	2 ; 0x010: 0x02
kfdpHdrB.ub1spare:	0 ; 0x011: 0x00
kfdpHdrB.ub2spare:	0 ; 0x012: 0x0000
kfdpHdrB.incarn:	0 ; 0x014: 0x00000000
kfdpHdrB.copy[0]:	0 ; 0x018: 0x0000
kfdpHdrB.copy[1]:	1 ; 0x01a: 0x0001
kfdpHdrB.copy[2]:	2 ; 0x01c: 0x0002
kfdpHdrB.copy[3]:	0 ; 0x01e: 0x0000
kfdpHdrB.copy[4]:	0 ; 0x020: 0x0000
kfdpHdrB.dtaSz:	3 ; 0x022: 0x0003
ub1[0]:	2 ; 0x024: 0x02

Agend

- **ASM的历史和简介**
- **ASM体系架构**
- **ASM在不同版本的新特性**
- **ASM存储和分配机制**
- **ASM日常使用注意事项**

ASM实例管理的注意事项

- 在10.2的环境中，该ASM实例使用ASMM的方式管理，设置为100M足够了

- **从11.2开始，Oracle强烈推荐使用AMM方式管理ASM实例**，并且最小值是 **MEMORY_TARGET (256 MB）**，**最佳实践是1.5GB**。大部分平台的缺省值是 MEMORY_TARGET (272 MB）。也可以采用传统模式：ALTER SYSTEM SET SGA_TARGET=0

- 12c（12.1）的文档中除了11.2的上述描述外，增加了一个内容：
  In an Oracle Exadata environment, the recommended settings for managing memory are SGA_TARGET = 1250MB, PGA_AGGREGATE_TARGET = 400MB, MEMORY_TARGET = 0, and MEMORY_MAX_TARGET = 0.

其他注意事项

1，存储上的条带大小

2，ASM磁盘大小

3，同一磁盘组中ASM磁盘大小和数目

4，在ASM中添加ASMDISK（LUN或者其他）

5，数据库维护（RMAN备份，mv数据库文件等）时，如何避免ASM文件或者目录自动被删除的情况

6，双存储使用ASM镜像

Q & A