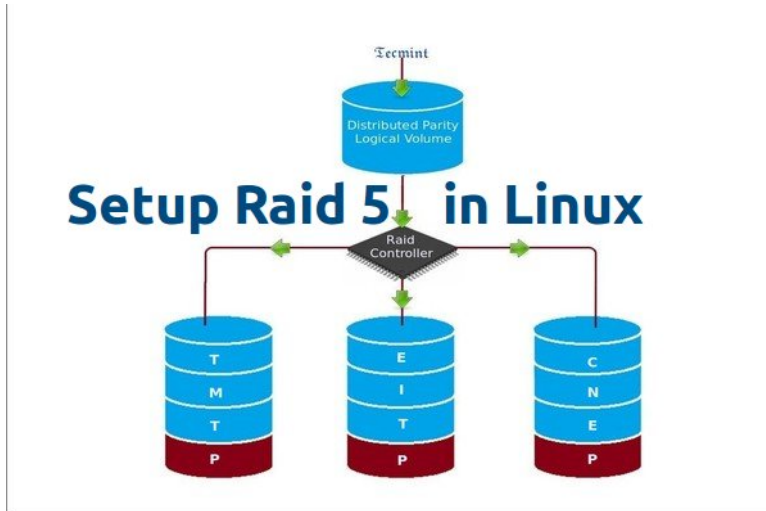


4 在 Linux 下使用 RAID (四) : 创建 RAID 5 (条带化与分布式奇偶校验)

<http://www.linuxidc.com/Linux/2015-08/122343.htm>

在 RAID 5 中, 数据条带化后存储在分布式奇偶校验的多个磁盘上。分布式奇偶校验的条带化意味着它将奇偶校验信息和条带化数据分布在多个磁盘上, 这样会有很好的数据冗余。



在 Linux 中配置 RAID 5

对于此 RAID 级别它至少应该有三个或更多个磁盘。RAID 5 通常被用于大规模生产环境中, 以花费更多的成本来提供更好的数据冗余性能。

什么是奇偶校验?

奇偶校验是在数据存储中检测错误最简单的常见方式。奇偶校验信息存储在每个磁盘中, 比如说, 我们有4个磁盘, 其中相当于一个磁盘大小的空间被分割去存储所有磁盘的奇偶校验信息。如果任何一个磁盘出现故障, 我们可以通过更换故障磁盘后, 从奇偶校验信息重建得到原来的数据。

RAID 5 的优点和缺点

- 提供更好的性能。
- 支持冗余和容错。
- 支持热备份。
- 将用掉一个磁盘的容量存储奇偶校验信息。
- 单个磁盘发生故障后不会丢失数据。我们可以更换故障硬盘后从奇偶校验信息中重建数据。
- 适合于面向事务处理的环境, 读操作会更快。
- 由于奇偶校验占用资源, 写操作会慢一些。
- 重建需要很长的时间。

要求

创建 RAID 5 最少需要3个磁盘, 你也可以添加更多的磁盘, 前提是你要有多个端口的专用硬件 RAID 控制器。在这里, 我们使用“mdadm”包来创建软件 RAID。mdadm 是一个允许我们在 Linux 下配置和管理 RAID 设备的包。默认情况下没有 RAID 的配置文件, 我们在创建和配置 RAID 后必须将配置文件保存在一个单独的文件 mdadm.conf 中。

在进一步学习之前, 我建议你通过下面的文章去了解 Linux 中 RAID 的基础知识。

- 介绍 RAID 的级别和概念 <http://www.linuxidc.com/Linux/2015-08/122191.htm>
- 使用 mdadm 工具创建软件 RAID 0 (条带化) <http://www.linuxidc.com/Linux/2015-08/122236.htm>
- 用两块磁盘创建 RAID 1 (镜像) <http://www.linuxidc.com/Linux/2015-08/122289.htm>
- 如何在Linux上构建 RAID 10阵列 <http://www.linuxidc.com/Linux/2014-12/110318.htm>
- Debian软RAID安装笔记 - 使用mdadm安装RAID1 <http://www.linuxidc.com/Linux/2013-06/86487.htm>
- 常用RAID技术介绍以及示例演示(多图) <http://www.linuxidc.com/Linux/2013-03/81481.htm>
- Linux实现最常用的磁盘阵列-- RAID5 <http://www.linuxidc.com/Linux/2013-01/77880.htm>
- RAID0+1和RAID5的性能测试结果 <http://www.linuxidc.com/Linux/2012-07/65567.htm>
- Linux入门教程: 磁盘阵列 (RAID) <http://www.linuxidc.com/Linux/2014-07/104444.htm>

我的服务器设置

1. 操作系统: CentOS 6.5 Final
2. IP 地址: 192.168.0.227
3. 主机名: rd5.tecmintlocal.com
4. 磁盘1 [20GB]: /dev/sdb
5. 磁盘2 [20GB]: /dev/sdc
6. 磁盘3 [20GB]: /dev/sdd

这是9篇系列教程的第4部分, 在这里我们要在 Linux 系统或服务器上使用三个20GB (名为/dev/sdb, /dev/sdc 和 /dev/sdd) 的磁盘建立带有分布式奇偶校验的软件 RAID 5。

第1步: 安装 mdadm 并检验磁盘

1、正如我们前面所说, 我们使用 CentOS 6.5 Final 版本来创建 RAID 设置, 但同样的做法也适用于其他 Linux 发行版。

1. # lsblk_release -a
2. # ifconfig | grep inet

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# # lsb_release -a  
LSB Version: :base-4.0-amd64:base-4.0-noarch:core-4.0-amd64:core-4.0-noarch  
Distributor ID: CentOS  
Description: CentOS release 6.5 (Final)  
Release: 6.5  
Codename: Final  
[root@rd5 ~]#  
[root@rd5 ~]#  
[root@rd5 ~]# # ifconfig | grep inet  
inet addr:192.168.0.227 Bcast:192.168.0.255 Mask:255.255.255.0  
inet6 addr: fe80::5054:ff:feb8:73fc/64 Scope:Link  
inet addr:127.0.0.1 Mask:255.0.0.0  
inet6 addr: ::1/128 Scope:Host  
[root@rd5 ~]#  
[root@rd5 ~]#  
[root@rd5 ~]# # hostname  
rd5.tecmintlocal.com  
[root@rd5 ~]# http://www.tecmint.com
```

CentOS 6.5 摘要

2、如果你按照我们的 RAID 系列去配置的，我们假设你已经安装了“mdadm”包，如果没有，根据你的 Linux 发行版使用下面的命令安装。

1. # yum install mdadm [在 RedHat 系统]
2. # apt-get install mdadm [在 Debain 系统]

3、“mdadm”包安装后，先使用fdisk命令列出我们在系统上增加三个20GB的硬盘。

1. # fdisk -l | grep sd

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# # fdisk -l | grep sd  
Disk /dev/sda: 19.3 GB, 19327352832 bytes  
/dev/sda1 * 1 64 512000 83 Linux  
/dev/sda2 64 2350 18361344 8e Linux LVM  
Disk /dev/sdb: 19.3 GB, 19327352832 bytes  
Disk /dev/sdc: 19.3 GB, 19327352832 bytes  
Disk /dev/sdd: 19.3 GB, 19327352832 bytes  
[root@rd5 ~]#  
[root@rd5 ~]# # ls -l /dev | grep sd  
brw-rw----. 1 root disk 8, 0 Oct 12 2014 sda  
brw-rw----. 1 root disk 8, 1 Oct 12 05:41 sda1  
brw-rw----. 1 root disk 8, 2 Oct 12 2014 sda2  
brw-rw----. 1 root disk 8, 16 Oct 12 2014 sdb  
brw-rw----. 1 root disk 8, 32 Oct 12 2014 sdc  
brw-rw----. 1 root disk 8, 48 Oct 12 05:41 sdd  
[root@rd5 ~]# http://www.tecmint.com
```

安装 mdadm 工具

4、现在检查这三个磁盘是否存在 RAID 块，使用下面的命令来检查。

1. # mdadm -E /dev/sd[b-d]
2. # mdadm --examine /dev/sdb /dev/sdc /dev/sdd # 或

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# # mdadm -E /dev/sd[b-d]  
mdadm: No md superblock detected on /dev/sdb.  
mdadm: No md superblock detected on /dev/sdc.  
mdadm: No md superblock detected on /dev/sdd.  
[root@rd5 ~]#  
[root@rd5 ~]#  
[root@rd5 ~]# # mdadm --examine /dev/sdb /dev/sdc /dev/sdd  
mdadm: No md superblock detected on /dev/sdb.  
mdadm: No md superblock detected on /dev/sdc.  
mdadm: No md superblock detected on /dev/sdd.  
[root@rd5 ~]#  
[root@rd5 ~]# http://www.tecmint.com
```

检查 Raid 磁盘

注意: 上面的图片说明，没有检测到任何超级块。所以，这三个磁盘中没有定义 RAID。让我们现在开始创建一个吧！

第2步：为磁盘创建 RAID 分区

5、首先，在创建 RAID 前磁盘 (/dev/sdb, /dev/sdc 和 /dev/sdd) 必须有分区，因此，在进行下一步之前，先使用fdisk命令进行分区。

1. # fdisk /dev/sdb
2. # fdisk /dev/sdc
3. # fdisk /dev/sdd

创建 /dev/sdb 分区

请按照下面的说明在 /dev/sdb 硬盘上创建分区。

- 按 n 创建新的分区。
- 然后按 p 选择主分区。选择主分区是因为还没有定义过分区。
- 接下来选择分区号为1。默认就是1。
- 这里是选择分区大小，我们没必要选择指定的大小，因为我们需要为 RAID 使用整个分区，所以只需按两次 Enter 键默认将整个容量分配给它。
- 然后，按 p 来打印创建好的分区。
- 改变分区类型，按 t 可以列出所有可用的类型。

- 按 `t` 修改分区类型。
- 这里使用 `fd` 设置为 RAID 的类型。
- 然后再次使用 `p` 查看我们所做的更改。
- 使用 `w` 保存更改。

```

root@rd5:~
[root@rd5 ~]# fdisk /dev/sdb

WARNING: DOS-compatible mode is deprecated. It's strongly recommended to
switch off the mode (command 'c') and change display units to
sectors (command 'u').

Command (m for help): n
Command action
  e   extended
  p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-2349, default 1):
Using default value 1
Last cylinder, +cylinders or +size{K,M,G} (1-2349, default 2349):
Using default value 2349

Command (m for help): t
Selected partition 1
Hex code (type L to list codes): fd
Changed system type of partition 1 to fd (Linux raid autodetect)

Command (m for help): p

Disk /dev/sdb: 19.3 GB, 19327352832 bytes
255 heads, 63 sectors/track, 2349 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0xed21efaf

   Device Boot      Start         End      Blocks   Id  System
/dev/sdb1             1         2349     18868311    fd  Linux raid autodetect

Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
[root@rd5 ~]#

```

<http://www.tecmint.com>

创建 `sdb` 分区

注意: 我们仍要按照上面的步骤来创建 `sd` 和 `sdd` 的分区。

创建 `/dev/sdc` 分区

现在, 通过下面的截图给出创建 `sd` 和 `sdd` 磁盘分区的方法, 或者你可以按照上面的步骤。

1. # `fdisk /dev/sdc`

```
root@rd5:~  
[root@rd5 ~]# fdisk /dev/sdc  
WARNING: DOS-compatible mode is deprecated. It's strongly recommended to  
switch off the mode (command 'c') and change display units to  
sectors (command 'u').  
Command (m for help): n  
Command action  
  e   extended  
  p   primary partition (1-4)  
p  
Partition number (1-4): 1  
First cylinder (1-2349, default 1):  
Using default value 1  
Last cylinder, +cylinders or +size{K,M,G} (1-2349, default 2349):  
Using default value 2349  
.  
Command (m for help): t  
Selected partition 1  
Hex code (type L to list codes): fd  
Changed system type of partition 1 to fd (Linux raid autodetect)  
Command (m for help): p  
Disk /dev/sdc: 19.3 GB, 19327352832 bytes  
255 heads, 63 sectors/track, 2349 cylinders  
Units = cylinders of 16065 * 512 = 8225280 bytes  
Sector size (logical/physical): 512 bytes / 512 bytes  
I/O size (minimum/optimal): 512 bytes / 512 bytes  
Disk identifier: 0x26ad0e05  


| Device    | Boot | Start | End  | Blocks   | Id | System                |
|-----------|------|-------|------|----------|----|-----------------------|
| /dev/sdc1 |      | 1     | 2349 | 18868311 | fd | Linux raid autodetect |

  
Command (m for help): w  
The partition table has been altered!  
Calling ioctl() to re-read partition table.  
Syncing disks.  
[root@rd5 ~]#
```

<http://www.tecmint.com>

创建 sdc 分区

创建 /dev/sdd 分区

1. # fdisk /dev/sdd

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# fdisk /dev/sdd  
WARNING: DOS-compatible mode is deprecated. It's strongly recommended to  
switch off the mode (command 'c') and change display units to  
sectors (command 'u').  
Command (m for help): n  
Command action  
  e   extended  
  p   primary partition (1-4)  
p  
Partition number (1-4): 1  
First cylinder (1-2349, default 1):  
Using default value 1  
Last cylinder, +cylinders or +size{K,M,G} (1-2349, default 2349):  
Using default value 2349  
.  
Command (m for help): t  
Selected partition 1  
Hex code (type L to list codes): fd  
Changed system type of partition 1 to fd (Linux raid autodetect)  
Command (m for help): p  
Disk /dev/sdd: 19.3 GB, 19327352832 bytes  
255 heads, 63 sectors/track, 2349 cylinders  
Units = cylinders of 16065 * 512 = 8225280 bytes  
Sector size (logical/physical): 512 bytes / 512 bytes  
I/O size (minimum/optimal): 512 bytes / 512 bytes  
Disk identifier: 0x3fbe8772  


| Device    | Boot | Start | End  | Blocks   | Id | System                |
|-----------|------|-------|------|----------|----|-----------------------|
| /dev/sdd1 |      | 1     | 2349 | 18868311 | fd | Linux raid autodetect |

  
Command (m for help): w  
The partition table has been altered!  
Calling ioctl() to re-read partition table.  
Syncing disks.  
[root@rd5 ~]#
```

<http://www.tecmint.com>

创建 sdd 分区

6. 创建分区后，检查三个磁盘 sdb, sdc, sdd 的变化。

1. # mdadm --examine /dev/sdb /dev/sdc /dev/sdd

2. # mdadm -E /dev/sd[b-c] # 或

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# mdadm --examine /dev/sdb /dev/sdc /dev/sdd  
/dev/sdb:  
  MBR Magic : aa55  
Partition[0] : 37736622 sectors at 63 (type fd)  
/dev/sdc:  
  MBR Magic : aa55  
Partition[0] : 37736622 sectors at 63 (type fd)  
/dev/sdd:  
  MBR Magic : aa55  
Partition[0] : 37736622 sectors at 63 (type fd)  
[root@rd5 ~]# http://www.tecmint.com
```

检查磁盘变化

注意: 在上面的图片中, 磁盘的类型是 fd。

7、现在在新创建的分区检查 RAID 块。如果没有检测到超级块, 我们就能够继续下一步, 在这些磁盘中创建一个新的 RAID 5 配置。

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# mdadm --examine /dev/sdb1 /dev/sdc1 /dev/sdd1  
mdadm: No md superblock detected on /dev/sdb1.  
mdadm: No md superblock detected on /dev/sdc1.  
mdadm: No md superblock detected on /dev/sdd1.  
[root@rd5 ~]# http://www.tecmint.com
```

Check Raid on Partition

*在分区中检查 RAID *

第3步: 创建 md 设备 md0

8、现在使用所有新创建的分区(sdb1, sdc1 和 sdd1)创建一个 RAID 设备 “md0” (即 /dev/md0), 使用以下命令。

1. # mdadm --create /dev/md0 --level=5 --raid-devices=3 /dev/sdb1 /dev/sdc1 /dev/sdd1
2. # mdadm -C /dev/md0 -l=5 -n=3 /dev/sd[b-d]1 # 或

9、创建 RAID 设备后, 检查并确认 RAID, ♦♦♦ mdstat 中输出中可以看到包括的设备的 RAID 级别。

1. # cat /proc/mdstat

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# mdadm --create /dev/md0 --level=5 --raid-devices=3 /dev/sdb1 /dev/sdc1 /dev/sdd1  
mdadm: Defaulting to version 1.2 metadata  
mdadm: array /dev/md0 started.  
[root@rd5 ~]#  
[root@rd5 ~]# cat /proc/mdstat  
Personalities : [raid6] [raid5] [raid4]  
md0 : active raid5 sdd1[3] sdc1[1] sdb1[0]  
      37703680 blocks super 1.2 level 5, 512k chunk, algorithm 2 [3/2] [UU]  
      [=>.....] recovery = 0.3% (73088/18851840) finish=47.0min  
      speed=6644K/sec  
  
unused devices: <none>  
[root@rd5 ~]# http://www.tecmint.com
```

验证 Raid 设备

如果你想监视当前的创建过程, 你可以使用 watch 命令, 将 cat /proc/mdstat 传递给它, 它会在屏幕上显示且每隔1秒刷新一次。

1. # watch -n1 cat /proc/mdstat

```
root@rd5:~  
[root@rd5 ~]#  
[root@rd5 ~]# watch -n1 cat /proc/mdstat  
http://www.tecmint.com
```

监控 RAID 5 构建过程

```
root@rd5:~  
Every 1.0s: cat /proc/mdstat Sun Oct 12 07:48:09 2014  
  
Personalities : [raid6] [raid5] [raid4]  
md0 : active raid5 sdd1[3] sdc1[1] sdb1[0]  
      37703680 blocks super 1.2 level 5, 512k chunk, algorithm 2 [3/2] [UU]  
      [=>.....] recovery = 7.4% (1411840/18851840) finish=35.1min  
      speed=8276K/sec  
  
unused devices: <none>  
http://www.tecmint.com
```

Raid 5 过程概要

10、创建 RAID 后, 使用以下命令验证 RAID 设备

1. # mdadm -E /dev/sd[b-d]1

```

root@rd5:~
[root@rd5 ~]#
[root@rd5 ~]# mdadm -E /dev/sd[b-d]1
/dev/sdb1:
    Magic : a92b4efc
    Version : 1.2
    Feature Map : 0x0
    Array UUID : 0185fe22:d0e31edf:c313404e:f25172e3
    Name : rd5.tecmintlocal.com:0 (local to host rd5.tecmintlocal.com)
    Creation Time : Sun Oct 12 07:44:47 2014
    Raid Level : raid5
    Raid Devices : 3

    Avail Dev Size : 37703854 (17.98 GiB 19.30 GB)
    Array Size : 37703680 (35.96 GiB 38.61 GB)
    Used Dev Size : 37703680 (17.98 GiB 19.30 GB)
    Data Offset : 32768 sectors
    Super Offset : 8 sectors
    State : clean
    Device UUID : 4c3fdef8:2b9bcfb0:231ba9bd:a4f96798

    Update Time : Sun Oct 12 08:29:23 2014
    Checksum : 6b5133a3 - correct
    Events : 18

    Layout : left-symmetric
    Chunk Size : 512K

    Device Role : Active device 0
    Array State : AAA ('A' == active, '.' == missing)
/dev/sdc1:

```

<http://www.tecmint.com>

验证 Raid 级别

注意: 因为它显示三个磁盘的信息, 上述命令的输出会有点长。

11、接下来, 验证 RAID 阵列, 假定包含 RAID 的设备正在运行并已经开始了重新同步。

1. # mdadm --detail /dev/md0

```

root@rd5:~
[root@rd5 ~]# mdadm --detail /dev/md0
/dev/md0:
    Version : 1.2
    Creation Time : Sun Oct 12 07:44:47 2014
    Raid Level : raid5
    Array Size : 37703680 (35.96 GiB 38.61 GB)
    Used Dev Size : 18851840 (17.98 GiB 19.30 GB)
    Raid Devices : 3
    Total Devices : 3
    Persistence : Superblock is persistent

    Update Time : Sun Oct 12 08:29:23 2014
    State : clean
    Active Devices : 3
    Working Devices : 3
    Failed Devices : 0
    Spare Devices : 0

    Layout : left-symmetric
    Chunk Size : 512K

    Name : rd5.tecmintlocal.com:0 (local to host rd5.tecmintlocal.com)
    UUID : 0185fe22:d0e31edf:c313404e:f25172e3
    Events : 18

    Number Major Minor RaidDevice State
    0      8      17        0      active sync  /dev/sdb1
    1      8      33        1      active sync  /dev/sdc1
    3      8      49        2      active sync  /dev/sdd1

```

<http://www.tecmint.com>

验证 RAID 阵列