dmraid update
Linux-Kongress Dresden 2009

Heinz Mauelshagen
Consulting Development Engineer
Top

ATAR AID
requirements
Device-Mapper architecture/features
dmraid architecture/features
user Interface
usage examples
summary
URLs
ATARAID

ATARAID is a generic term for a variety of cheap vendor software RAID solutions (e.g., Highpoint, Silicon Image, ...)
typically combines a number of P/SATA adaptors (HBA and/or on MoBo),
a BIOS extension, a host OS driver and a management tool

BIOS extension supports creation of vendor specific metadata describing RAID devices/sets through a management UI and mapping created RAID sets as a single device

OS driver contains the actual RAID logic (i.e., mirroring);
Windows version shipped with the product

RAID driver for a subset of solutions were part of Linux 2.4
Device-mapper in Linux 2.6 avoids having separate RAID drivers for each solution
Requirements

Device-Mapper runtime
Tool to:
discover ATARAIID devices
retrieve RAID properties from vendor metadata
group RAID sets
create/remove mapped devices (ie. logical devices)
display RAID device/set properties
handle device events (eg. disk failure)
avoid hardware dependencies by ignoring interface
types (eg. particular PCI adaptor) completely
Device-Mapper is the generic mapping runtime platform in the Linux kernel (since 2.5):

- can be used by multiple applications which need block device mapping (eg, dmraid, LVM2 or multipathd)
- manages Mapped-Devices (create, remove, …)
- handles text formatted Mapping-Tables (load, unload, reload)
- all table loading actions happen online to reflect changed layouts (eg. an MD size change)
- code to handle mappings factored out into Mapping-Targets, which handle various layouts: linear, striped, mirrored, snapshot, multipath, zero, error, crypt, …
- maps to arbitrary block devices (eg. (i)SCSI)
Device-Mapper architecture/features(2)

Mapping-Targets are dynamically loadable and register with the Device-Mapper core by target name

Mapped-Devices can be stacked in order to build complex mappings (eg. to create a RAID10 set)

more than 2 Terabytes per mapped device in Linux 2.6 (CONFIG_LBD)

comes with a user space library to be interfaced by Device/Volume Management applications and a test tool dmsetup

lib creates nodes to access Mapped Devices in /dev/mapper/
Device-Mapper architecture/features(3)

Examples of Device-Mapper tables which can be activated using the dmsetup tool:

0 1024 linear /dev/hde1 40
   1024 2048 striped 2 64 /dev/hde1 1064 /dev/hdf1 0

0 1024 zero
   1024 1000 error

0 83886080 mirror core 1 64 2 /dev/sda 0 /dev/sdb 0

The device-mapper package provides a tool “dmsetup” for low-level management of mapped devices;

    eg. “dmsetup create device_name mapping_table_file” will create a mapped device named “device_name” with the respective mapping table retrieved from file “mapping_table_file”
Device-Mapper Architecture Overview
dmraid architecture/features(1)

dmraid application falls apart into tool and library with tool exposing the following library functionality:

discovery of block devices
retrieval of device properties (ie. path, size, serial#)
multi RAID metadata format registry (format handlers)
discovery of RAID devices (ie. metadata)
translation of particular metadata format into internal generic RAID device representation

grouping of RAID sets (hierarchical with eg. RAID10) represented via internal generic format

derivation of mapping tables from internal RAID device/set format
dmraid architecture/features(2)

dump RAID metadata to files
erase RAID metadata
creation of mapped devices
activation of (stacked) mappings
logging of block/RAID device and RAID set properties
display RAID device and RAID set properties
list supported metadata formats
dmraid architecture/features(new)

New with Intel IMSM ATARAIMD format (isw):
creation of RAID sets
removal of RAID sets
spare device addition
spare device removal
rebuild of RAID sets utilizing spare device(s)
provides event driven DSO
Architecture Overview

- **tool**
- **libdmraid**
- **libdm**
- **dm**
- **lld**
- **target**

Userspace

Kernelspace
**User Interface(1)**

- **b**: display discovered block devices
- **l**: list all supported metadata formats
- **r**: display discovered RAID devices
- **s**: display grouped RAID sets
- **n**: display native metadata
- **ay**: activate all grouped RAID sets
- **an**: deactivate all grouped RAID sets
- **tay**: display mapping tables
- **p**: avoid activating any partitions on RAID sets
User Interface(2)

-rD : dump vendor metadata into files
-rE : erase vendor metadata on devices
-h : help

-r, -s and -a optionally take -f to select formats
   (identifiers as listed with -l; eg, '-fhpt,sil')

-r and -n take an optional list of device paths
-ay, -an and -s take an optional list of RAID set names
User Interface(new)

-Z: remove partitions on underlying RAID devices when activating set(s)
-C: create RAID set
-R: rebuild RAID set
-S: add hot-spare drive to a RAID set
-x: remove RAID set
Usage Examples(new)

dmraid -f isw -C Raid0 --type 0 --strip 8k --size 20g --disk "'/dev/sd[bc]"
creates an ISW volume with a name of "Raid0", 20 Gigabytes in total, and
8 Kilobytes stripe size on two disks

dmraid -f isw -C Test0 --type 0 --disk "'/dev/sd[de]"
creates an ISW volume with the default size and stripe size

dmraid -f isw -C Test10 --type 01 --strip 128B --disk "'/dev/sd[a-d]"
creates a stacked RAID device, RAID10 (isw format), with a name of "Test10",
128 blocks (512 Bytes) stripe size, and the default volume size on 4 disks

dmraid -f isw -S -M /dev/sde
marks the device /dev/sde as a hot spare for rebuild

dmraid -R isw_djaggchdde_RAID1 /dev/sde
starts rebuild of the RAID volume on device /dev/sde
Summary

dmraid utilizes the generic mapping capabilities of device-mapper in order to make a growing number of ATARAID solutions available on Linux 2.6.

Enhancements to create/remove/rebuild RAID sets and handle spare devices including event driven device monitoring are now supported for the widely available Intel IMSM ATARAID format.
URLs

http://sources.redhat.com/dm (Device-Mapper tool+library, dmraid)
http://www.redhat.com/mailman/listinfo/dm-devel
to subscribe to dm-devel@redhat.com
http://www.redhat.com/mailman/listinfo/ataraid-list
to subscribe to ataraid-list@redhat.com
Q&A