

Sunday, March 30. 2008

Less known Solaris Features: Remote Mirror with AVS - Part 3: Setting up a synchronous mirror with AVS

After this short introduction in the theory of AVS Remote mirror, we will set up such an replication between two hosts. Prerequisites for this tutorial In this tutorial we need to work with two systems. I will use theoden and gandalf again.

10.211.55.200 theoden

10.211.55.201 gandalf Both systems have a boot disk and two disks for data. The data disks have a size of 512 MB. In my example this disks are c1d0 and c1d1. I've partitioned each disk in the same manner: # prtvtoc /dev/rdisk/c1d0s2

. /dev/rdisk/c1d0s2 partition map

[..]

```
.
      First  Sector  Last
. Partition Tag Flags  Sector  Count  Sector  Mount Directory
   0  0  00   6144   65536  71679
   1  0  00   71680  968704  1040383
   2  5  01     0 1042432 1042431
   8  1  01     0   2048   2047
   9  9  00    2048   4096   6143
```

It's important that the primary and secondary partitions and their respective bitmap partitions are equal in size. Furthermore: Don't use cylinder 0 for partitions under the control of AVS. This cylinder may contain administrative information from other components of the systems. Replication of this information may lead to data loss.

In my example i've chosen a 32 MB bitmap partition for a 512 MB data partition. This is vastly too large for the use case. You will find the exact rule for sizing this partition in the documentation

The first partition c1dxs0 is used as the bitmap volume. The second volume c1dxs1 is used as the primary volume on the source of the replication respectively the secondary volume on the target of the replication.

Setting up the replication Okay, at first we have to enable AVS on both hosts. At first we activate it on theoden [root@theoden:~]\$ dscfgadm -e This command may ask for the approval to create the config database, when you run this command for the first time. Answer this question with y. After this we switch to gandalf to do the same. [root@gandalf:~]\$ dscfgadm -e Now we can establish the replication. We login at theoden first, and configure this replication. [root@theoden:~]\$ sndradm -e theoden /dev/rdisk/c1d0s1 /dev/rdisk/c1d0s0 gandalf /dev/rdisk/c1d0s1 /dev/rdisk/c1d0s0 ip sync

Enable Remote Mirror? (Y/N) [N]: y What have we configured? We told AVS to replicate the content of /dev/rdisk/c1d0s1 on theoden to /dev/rdisk/c1d0s1 on gandalf. AVS uses the /dev/rdisk/c1d0s1 volume on each system as the bitmap volume for this application. At the end of this command we configure, that the replication uses IP and it's a synchronous replication.

Okay, but we have to configure it on the targeted system of the replication as well: [root@gandalf:~]\$ sndradm -e theoden /dev/rdisk/c1d0s1 /dev/rdisk/c1d0s0 gandalf /dev/rdisk/c1d0s1 /dev/rdisk/c1d0s0 ip sync

Enable Remote Mirror? (Y/N) [N]: y We repeat the same command we used on theoden on gandalf as well. Forgetting to do this step is one of the most frequent errors in regard of setting up an remote mirror.

An interesting command in regard of AVS remote mirror is the dsstat command. It shows the status and some statistic data about your replication. [root@theoden:~]\$ dsstat -m sndr

```
name      t s  pct role  kps  tps  svt
dev/rdisk/c1d0s1  P L 100.00 net   0   0   0
dev/rdisk/c1d0s0          bmp  0   0  0
```

The 100.00 doesn't stand for "100% of the replication is completed". It standing for "100% of the replication to do". We have to start the replication manually. Okay, more formally the meaning of this column is "percentage of the volume in need of syncing". And as we freshly configured this replication it's obvious, that the complete volume needs synchronisation.

Two other columns are important, too: It's the t and the s column. The t column designates the volume type and the s the status of the volume. In this case we've observed the primary volume and it's in the logging mode. It records changes, but doesn't replicate them right now to the secondary volume.

```
Okay, so let's start the synchronisation:[root@theoden:~]$ sndradm -m gandalf:/dev/rdisk/c1d0s1
Overwrite secondary with primary? (Y/N) [N]: yWe can lookup the progress of the sync with the dsstat command again:
[root@theoden:~]$ dsstat -m sndr
name      t s  pct role  kps tps svt
dev/rdsk/c1d0s1 P SY 97.39 net  Inf  0 -NaN
dev/rdsk/c1d0s0      bmp Inf  0 -NaN
[root@theoden:~]$ dsstat -m sndr
name      t s  pct role  kps tps svt
dev/rdsk/c1d0s1 P SY 94.78 net  Inf  0 -NaN
dev/rdsk/c1d0s0      bmp Inf  0 -NaN
[...]
[root@theoden:~]$ dsstat -m sndr
name      t s  pct role  kps tps svt
dev/rdsk/c1d0s1 P SY 3.33 net  Inf  0 -NaN
dev/rdsk/c1d0s0      bmp Inf  0 -NaN
[root@theoden:~]$ dsstat -m sndr
name      t s  pct role  kps tps svt
dev/rdsk/c1d0s1 P R  0.00 net   0  0  0
dev/rdsk/c1d0s0      bmp  0  0  0
```

When we start the synchronisation the status of the volume switches to SY for synchronizing. After a while the sync is complete. The status switches again, this time to R for replicating. From this moment all changes to the primary volume will be replicated to the secondary one.

Now let's play around with our new replication set by using the primary volume. Create a filesystem on it for example, mount it and play around with it:

```
[root@theoden:~]$ newfs /dev/dsk/c1d0s1
newfs: construct a new file system /dev/rdsk/c1d0s1: (y/n)? y
/dev/rdsk/c1d0s1: 968704 sectors in 473 cylinders of 64 tracks, 32 sectors
473.0MB in 30 cyl groups (16 c/g, 16.00MB/g, 7680 i/g)
super-block backups (for fsck -F ufs -o b=#) at:
32, 32832, 65632, 98432, 131232, 164032, 196832, 229632, 262432, 295232,
656032, 688832, 721632, 754432, 787232, 820032, 852832, 885632, 918432, 951232
[root@theoden:~]$ mount /dev/dsk/c1d0s1 /mnt
[root@theoden:~]$ cd /mnt
[root@theoden:~]$ touch test
[root@theoden:~]$ cp /var/log/* .
[root@theoden:~]$ mkfile 1k test2
```

Okay, in a few seconds i will show you, that all changes really get to the other side.

Posted by Joerg Moellenkamp in English, Solaris at 21:10

I think you make a type in:
What have we configured? We told AVS to replicate the content of /dev/rdsk/c1d0s1 on theoden to /dev/rdsk/c1d0s1 on gandalf.

AVS uses the /dev/rdsk/c1d0s1 volume on each system as the bitmap volume for this application. At the end of this command we configure, that the replication uses IP and it's a synchronous replication.

I would guess that the bitmap volume used on each system would be /dev/rdsk/c1d0s0?
Anonymous on Sep 12 2008, 15:43

"Furthermore: Don't use cylinder 0 for partitions under the control of AVS. This cylinder may contain administrative information from other components of the systems. Replication of this information may lead to data loss."
Was ist mit der o.g. Aussage genau gemeint? Wenn ich eine Partition/Slice anlege, die mit Cylinder 0 beginnt und dort ein UFS-Filesystem mit Anwendungsdaten anlegessollte das doch kein Problem sein, oder?!?
Wenn die Slice für die state database vom SVM verwendet wird macht die Replikation natürlich keinen Sinn. hier habe ich schon mehrere Fälle gesehen, wo die ersten paar Cylinder einer Disks dafür genutzt werden.

Ich hätte noch eine Frage: Ist es mit AVS eigentlich möglich als Quell- und Zielpartition SVM-Metadevices zu nutzen? Interessant wäre es z.B. damit Filesysteme mit verschiedenem SVM-Layout zu replizieren. Extrembeispiel: Filesystem auf Softpartition wird auf Filesystem bestehend aus striped metadevices repliziert. Ist das recht abwegig?

MfG Otmanix
Anonymous on Sep 17 2008, 16:12