

The REFLINK Operation in OCFS2

Joel Becker

Tao Ma

Oracle

Agenda

- Problems
- Solution – REFLINK in OCFS2
- Status Update

Problems

Virtual Machines want:

- Snapshots
- Clones
- Space Savings

REFLINK operations

- `int reflink(const char *oldpath,`
- `const char *newpath, int preserve);`
-
- `reflink()` has the same calling semantics as `cp(2)`
- but creates a new file that shares all the data extents of the source file

Using REFLINK

- **Snapshots:**

```
# umask 333 && reflink host1.img \  
host1.img.snap
```

- **Restore:**

```
# unlink host1.img  
# reflink host1.img.snap host1.img
```

- **Clones:**

```
# reflink os-base.img host1.img  
# create-vm --name host1 host1.img
```

OCFS2 Implementation

- Use reflink count for a extent-based file.
- CoW when writing to the reflinked file.
- No punishment for read.

OCFS2 Examples

```
ls -l
```

```
-rw-r--r-- 1 root root 4195352576 Oct 28 10:26 el5_u1.img
```

```
df -h .
```

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/sda9	103G	4.2G	99G	5%	/mnt/ocfs2

```
time reflink el5_u1.img el5_u1_1.img
```

```
real 0m0.022s
```

```
user 0m0.000s
```

```
sys 0m0.001s
```

OCFS2 Examples(Contd.)

```
ls -l
```

```
-rw-r--r-- 1 root root 4195352576 Oct 28 10:32 el5_u1_1.img  
-rw-r--r-- 1 root root 4195352576 Oct 28 10:26 el5_u1.img
```

```
df -h .
```

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/sda9	103G	4.2G	99G	5%	/mnt/ocfs2

OCFS2 Examples(Contd.)

```
md5sum el5_u1.img  
6bf239ca47c8d165006047235ae6ba07 el5_u1.img  
real 1m6.425s  
user 0m12.420s  
sys 0m3.200s
```

```
md5sum el5_u1_1.img  
6bf239ca47c8d165006047235ae6ba07 el5_u1_1.img  
real 1m6.837s  
user 0m12.479s  
sys 0m3.222s
```

OCFS2 Examples(Contd.)

```
dd if=/dev/zero of=el5_u1_1.img bs=1M count=1000 seek=500 conv=notrunc  
1000+0 records in  
1000+0 records out  
1048576000 bytes (1.0 GB) copied, 18.4234 seconds, 56.9 MB/s
```

```
ls -l
```

```
-rw-r--r-- 1 root root 4195352576 Oct 28 11:07 el5_u1_1.img  
-rw-r--r-- 1 root root 4195352576 Oct 28 10:26 el5_u1.img
```

```
df -h .
```

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/sda9	103G	5.2G	98G	6%	/mnt/ocfs2

OCFS2 Examples(Contd.)

```
filefrag -sv /mnt/ocfs2/el5_u1_1.img
```

```
Filesystem type is: 7461636f
```

```
File size of /mnt/ocfs2/el5_u1_1.img is 4195352576 (1024256 blocks, blocksize 4096)
```

ext	logical	physical	expected	length	flags
0	0	18688		128000	shared
1	128000	1046016	146687	768	
2	128768	1048832	1046783	255232	
3	384000	404736	1304063	640256	shared,eof

```
/mnt/ocfs2/el5_u1_1.img: 4 extents found
```

Status Update

- REFLINK in OCFS2 has been merged into 2.6.32.
- System call will be integrated into Linux Kernel soon.
- OCFS2 1.6 for enterprise kernel will have REFLINK.

Status Update (Contd.)

Oracle VM 3.0 will have OCFS2 1.6.

OCFS2 will be one type of Oracle Storage Connect Program.

Thanks !

Joel & Tao