ZFS on FreeBSD: A Quick Tutorial

Tai-hwa Liang
<avatar@FreeBSD.org>

History/Availability

- Designed by Sun Microsystems (Sep-14-2004).

- First OpenSolaris inclusion: Nov-16-2005.

- Solaris 10 inclusion: Jun-2006.

- Became part of FreeBSD on Apr-06-2007.
  - Ported by Pawel Jakub Dawidek <pjd@FreeBSD.org>.

- Mac OS X 10.5 Leopard (?)

- Linux (via FUSE)
  - CDDL is not 'compatible' with GPL

ZFS Features

- ☐ Pooled Storage
  - ○ Build-in volume manager (LVM-like) and RAID

- ☐ Transactional Semantics
  - ○ Think about databases...
  - ○ No more fsck, journaling

- ☐ Lightweight filesystem creation
  - ○ Snapshot (rollback to "undo" anything bad)
  - ○ Clones (writable snapshots)

- ☐ Mind-boggling Scalability
  - ○ 128 bits(256 quadrillion zettabytes)
    - ▷ giga - 10^9
    - ▷ zetta - 10^21

ZFS Features (cont.)

☐ Data Integrity
- ○ Checksumming
- ○ Self healing

☐ Platform Independent Endianness

☐ Built-in Compression
- ○ lzjb
- ○ gzip-N

☐ Simplified Administration
- ○ No more LVM
- ○ Delegated to non-root users

ZFS Limitations

- Per-user or per-group quota is not supported.
  - Workaround: create user-owned filesystem and set its own size limit with 'zfs set quota=XXG'

- Transparent encryption is not supported, yet.
  - WIP: http://www.opensolaris.org/os/project/zfs-crypto
  - Workaround: geli(8)

- It's not possible to change the number of disk in a RAID-Z set.

- It's not possible to reduce the number of vdevs in a zpool.
  - WIP: http://www.opensolaris.org/jive/thread.jspa?messageID=118280

What About The Performance?

- Testing environment:
  - 7.0-CURRENT-200706(snapshot) GENERIC kernel
  - Intel Core 2 T5500 @ 1.66GHz
  - 1GB memory

- Extract a 447MB tarball(-CURRENT source tree as of Aug-21-2007):
  - UFS + SoftUpdate: 2:45.86
  - ZFS: 1:06.00

- 'make buildworld buildkernel':
  - UFS + SoftUpdate: 1:48:11.96
  - ZFS: 2:21:16.75

Host with 'Smaller' Memory

☐ 512MB memory(it's strongly discouraged to use ZFS without at least 1 GB of memory.)

- ○ /boot/loader.conf:
  - ▷ vm.kmem_size="256M"
  - ▷ vfs.zfs.prefetch_disable="1"

- ○ /etc/sysctl.conf:
  - ▷ kern.maxvnodes=22500

☐ 1024MB memory

- ○ /boot/loader.conf:
  - ▷ vfs.zfs.arc_max="100M"
  - ▷ vm.kmem_size_max="600M"
  - ▷ vfs.zfs.prefetch_disable="1"

- ○ /etc/sysctl.conf:
  - ▷ kern.maxvnodes=50000

Troubleshooting

- For ZFS-on-root setup can't mount ZFS root:
  - /boot/zfs/zpool.cache must be up-to-date on the boot filesystem(zfs import).

- Ordinary users can not create snapshot?
  - Make sure that they are in the 'operator' group.

- panic: kmem_alloc(131072): kmem_map too small: xxxxx total allocated
  - Not recommended for host < 1GB of memory
    - http://people.freebsd.org/~pjd/patches/vm_kern.c.2.patch
  - Set vm.kmem_size and vm.kmem_size_max

References

- Live Demo
  - http://people.freebsd.org/~pjd/misc/zfs

- Guides
  - Solaris ZFS Administration Guide
    - http://opensolaris.org/os/community/zfs/docs/zfsadmin.pdf
  - ZFS - FreeBSD Wiki
    - http://wiki.freebsd.org/ZFS

- NFS and ZFS, a fine combination(read before you want to set zil_disable=1)
  - http://blogs.sun.com/roch/entry/nfs_and_zfs_a_fine

- FreeBSD/ZFS - last word in operating/file systems
  - http://people.freebsd.org/~pjd/pubs/eurobsdcon07_zfs.pdf