ZFS and FreeBSD

Martin Matuška
mm@FreeBSD.org

VX Solutions s. r. o.

BSDDay 2010
Eötvös Loránd University
Budapest, Hungary, 20.11.2010

About this presentation

This presentation will give a very brief introduction into ZFS and try to answer the following questions:

- ▶ What are the newest features in ZFS?
- ▶ What is the state of ZFS in FreeBSD?

Introduction

- What is ZFS
- ZFS history
- Main ZFS objects
- ZFS limits

What is ZFS?

ZFS is the "Zettabyte filesystem"

Original ZFS features by design:

- ▶ pooled storage (integrated volume manager)
- ▶ transactional semantics (copy on write)
- ▶ checksums and self-healing (scrub, resilver)
- ▶ scalability
- ▶ instant snapshots and clones
- ▶ dataset compression (lzjb, gzip)
- ▶ simplified delegable administration

ZFS history

- 2005/10: OpenSolaris - ZFS introduced in revision 789
- 2006/06: Solaris 10 update 6 - pool version 3
- 2007/04: FreeBSD (CURRENT) - pool version 6
- 2008/11: FreeBSD (CURRENT) - pool version 13
- 2009/10: Solaris 10 update 8 - pool version 15
- 2010/07: FreeBSD (CURRENT) - pool version 15
- 2010/08: OpenSolaris - closed, last revision 13149 (v28)
- 2010/09: Solaris 10 update 9 - pool version 22 (no dedup)
- 2010/11: Solaris 11 Express - pool version 31

Main ZFS objects

The two main ZFS objects are:

- pool
- dataset

ZFS pool

A ZFS pool is a storage object consisting of virtual devices.
'vdevs' can be:

- disk (partition, GEOM object, ...)
- file (experimental purposes)
- mirror (groups two or more vdevs)
- raidz, raidz2, raidz3 (single to triple parity RAIDZ)
- spare (pseudo-vdev for hot spares)
- log (separate ZIL device, may not be raidz)
- cache (L2 cache, may not be mirror or raidz)

ZFS dataset

Each ZFS pool contains ZFS datasets.
ZFS dataset is a generic name for:

- ▶ file system (posix layer)
- ▶ volume (virtual block device)
- ▶ snapshot (read-only copy of filesystem or volume)
- ▶ clone (filesystem with initial contents of a snapshot)

ZFS limits

What are the limits of ZFS?

- ▶ ZFS is a 128-bit filesystem
- ▶ Maximum pool size: 256 quadrillion zettabytes
  ($= 256 * 10^{36}$ bytes)
- ▶ Maximum filesystem/file/attribute size: 16 exabytes
- ▶ Maximum pools/filesystems/snapshots: 2^{64}

New features

- ZFS pool and filesystem versioning
- New ZFS pool versions (15-22)
- New ZFS pool versions (23-31)
- Other new user-visible features

ZFS pool and filesystem versioning

- ZFS pools and filesystems have a version number
- incompatible structural changes lead to a version increase
- backwards compatibility is provided
- forward compatibility is NOT provided
- version downgrade is NOT possible
- latest ZFS pool version: 31
- latest ZFS filesystem version: 5

New ZFS pool versions (15-22)

- pool version 15: user/group space accounting
- pool version 16: STMF property support
- pool version 17: triple parity RAID-Z
- pool version 18: snapshot user holds
- pool version 19: log device removal
- pool version 20: ZLE compression (zero-length encoding)
- pool version 21: deduplication
- pool version 22: zfs receive properties

New ZFS pool versions (23-31)

- pool version 23: slim ZIL
- pool version 24: system attribute support
- pool version 25: improved scrub statistics
- pool version 26: improved snapshot deletion performance
- pool version 27: improved snapshot creation performance
- pool version 28: multiple vdev replacements
- pool version 29: RAID-Z/mirror hybrid allocator
- pool version 30: encryption
- pool version 31: improved 'zfs list' performance

Other new user-visible features

Other important new features not touching pool versions:

- device autoexpansion (post-v16)
- ZFS pool recovery (post-v19)
- deduplication of zfs send streams (post-v21)
- splitting mirrors into separate pools (post-v22)
- ZIL synchronicity setting for datasets (post-v24)
- diff between snapshots (post-v28)

ZFS in FreeBSD

- Overview of ZFS in FreeBSD
- Other operating systems / distributions

ZFS in FreeBSD - Overview

- ZFS introduced in Apr-2007 (pool version 6)
- Latest release: pool version 14 in 8.1-RELEASE
- Current state: pool version 15 in 9-CURRENT and 8-STABLE + some backported improvements (L2ARC, Metaslabs, ACL cache, ...)
- Developer testing: version v28
- Documentation: wiki, manual pages
- Support: mailing lists, forums
- Future: cooperation with Illumos?

FreeBSD ZFS developers

- Paweł Jakub Dawidek (pjd@FreeBSD.org) (maintainer)
- Andriy Gapon (avg@FreeBSD.org)
- Xin Li (delphij@FreeBSD.org)
- Martin Matuska (mm@FreeBSD.org)
- External developers - patches for p4

Ongoing ZFS work at FreeBSD

- ▶ Backported improvements from higher versions:
  L2ARC, Metaslabs, ACL caching, ...
- ▶ Improving ARC and VM paging interaction, sendfile(2)
  (avg@FreeBSD.org)
- ▶ TBD: public testing of pool version 28

Thank you for your attention!

http://blog.vx.sk
http://www.vx.sk