The Future of ZFS in FreeBSD

Martin Matuška
mm@FreeBSD.org

VX Solutions s. r. o.

bsdday.eu
05.11.2011

About this presentation

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

- What are the old and new features of ZFS?
- What open source operating systems do have ZFS?
- How does the future of ZFS in FreeBSD look like?

Introduction

ZFS feature history

Operating systems

Future of ZFS in FreeBSD

Introduction

- What is ZFS
- ZFS history timeline
- ZFS limits
- Main ZFS objects
- CDDL license

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)
- simplified delegable administration

ZFS history timeline

- 2005/10: OpenSolaris - ZFS introduced in revision 789
- 2006/06: Solaris 10 update 6 - pool v3
- 2008/02: FreeBSD 7.0 - pool v6
- 2009/10: Solaris 10 update 8 - pool v15
- 2009/11: FreeBSD 8.0 - pool v13
- 2010/08: OpenSolaris - closed, last revision 13149 (v28)
- 2010/11: Solaris 11 Express - pool v31
- 2011/01: Linux - native ZFS v28 (KQ Infotech, LLNL)
- 2011/02: FreeBSD 8.2 - pool v15
- 2011/08: Solaris 10 update 10 - pool v29 (w/o dedup)
- 2011/Q4: FreeBSD 9.0 - pool v28

ZFS limits

What are the theoretical 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}

Main ZFS objects

ZFS uses two main objects:

- pool
- dataset

ZFS pool

A ZFS pool is a storage object consisting of virtual devices.
These '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)

CDDL License

ZFS source code is licensed under the
Common Development and Distribution License (CDDL)

- based on Mozilla Public License (MPL) version 1.1
- GPL incompatible
- if binaries are distributed, source code must be distributed
- but only for "Covered Software" = original + modifications
- may be part of a "Larger Work" containing other licenses
- modifications must be CDDL, author ("Contributor") must be disclosed
- terminates if any patent infringements against author or contributors

ZFS feature history

- ZFS pool and filesystem versioning
- ZFS feature history

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 open source ZFS pool version: 28
- latest open source ZFS filesystem version: 5

ZFS feature history 1/3

New features increasing the pool version number:

- hot spares and double-parity raidz (v3)
- gzip compression (v5)
- separate intent log devices (ZIL) (v7)
- delegated administration (v8)
- refquota and refreservation (v9)
- cache devices (L2 cache) (v10)
- user/group space accounting (v15)

ZFS feature history 2/3

New features increasing the pool version number:

- triple parity RAID-Z (v17)
- snapshot user holds (v18)
- intent log device removal (v19)
- deduplication (v21)
- zfs receive properties (v22)
- system attribute support (v24)
- dataset encryption (v30, Solaris only)

ZFS feature history 3/3

Other important 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 operating systems

- OpenSolaris-based distributions
- Other operating systems and distributions

Systems based on OpenSolaris

- OpenSolaris (v28, discontinued)
- Oracle Solaris 10 (U10 v29)
- Nexenta Core (v26)
- OpenIndiana (v28)
- (SchilliX)
- (Belenix)

OpenSolaris

- Project discontinued
- The source of ZFS code for everyone else
- ZFS introduced on 31-Oct-2005 in revision 789
- Last public commit to ZFS on 18-Aug-2010 (rev 13147)
- Bug database not available anymore
- Free successor: Illumos (releases: OpenIndiana)

Oracle Solaris

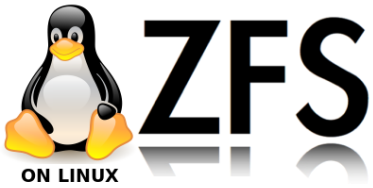
- Commercial OS - Licence Required
- ZFS introduced in Solaris 10 update 6 (Jun-2006)
- Latest: S10 update 10 (Aug-2011) with ZFS v29 (no dedup)
- Solaris 11 Express (Nov-2010) with ZFS v31 (dedup)
- Recommended literature: Oracle® Solaris ZFS Administration Guide

Other Operating Systems

ZFS originates from OpenSolaris - everybody else has to port it.

- FreeBSD (v28)
- Linux (v28, FUSE or kernel module by LLNL)
- Debian GNU/kFreeBSD (uses FreeBSD kernel)
- (MacOS X)(github, development stalled)
- (NetBSD) (development stalled)

Linux



- ZFS-fuse project
  Version 0.7.0 - ZFS pool v23
- ZFS kernel modules by Brian Behlendorf (LLNL)
  Version 0.6 (RC) - ZFS pool v28

FreeBSD

- ZFS introduced in April 2007 (pool version 6)
- Latest release: pool version 15 in 8.2-RELEASE
- Current state: pool version 28 in 10-CURRENT, 9-STABLE and 8-STABLE
- v28 will be part of 9.0-RELEASE and 8.3-RELEASE
- Documentation: wiki, manual pages
- Support: FreeBSD PR's, mailing lists, forums

Future of ZFS in FreeBSD

This section will cover the following topics:

- ZFS at Oracle
- The Illumos Project
- ZFS development at FreeBSD
- Future of ZFS

ZFS at Oracle

A leaked internal Oracle memo from August, 2010 claims the following:

- Oracle will continue to develop ZFS but not in public
- ZFS code will remain CDDL licensed
- CDDL source code will be published with Solaris releases
- development sources will be available only to industry partners via OTN (Oracle Technology Network)

Current pool version: 31 (Solaris Express)

The Illumos Project

- project started by several former OpenSolaris developers
- goal: provide a free ON source (and replace closed parts)
- sponsored and supported primarily by Nexenta and Deplhix
- distributions to build on Illumos: OpenIndiana, Nexenta

ZFS development at FreeBSD 1/2

Main ZFS developers in FreeBSD:

- Paweł Jakub Dawidek (pjd@FreeBSD.org) (maintainer)
- Martin Matuška (mm@FreeBSD.org)
- Andriy Gapon (avg@FreeBSD.org)
- Xin Li (delphij@FreeBSD.org)

ZFS development at FreeBSD 2/2

- Current state: pool version 28
- Import of changes and new features from Illumos
- Bugfixes to common ZFS code - reported to Illumos
- Bugfixes to FreeBSD-specific code (loader, VM, etc.)
- ZFS fault management daemon - zfsd (gibbs@, mav@, delphij@)
- Jail support needs some more work

Future of ZFS

- ZFS will stay at version 28 (until Oracle releases CDDL code)
- There are neither plans nor manpower to maintain a fork
- Important: Future interaction FreeBSD <-> Illumos
- New (private) features at FreeBSD: zfsd
- Bugfixes, bugfixes, bugfixes ...

Thank you for your attention!

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