

Silicon Graphics, Inc.

XFS Overview & Internals

01 - Background

© Copyright 2006 Silicon Graphics Inc. All rights reserved.

Permission is granted to copy, distribute, and/or modify this document under the terms of the Creative Commons Attribution-Share Alike, Version 3.0 or any later version published by the Creative Commons Corp. A copy of the license is available at <http://creativecommons.org/licenses/by-sa/3.0/us/> .

November 2006



Topics

- Course Objectives
- A Brief History of XFS
- Using XFS
 - Creating XFS Filesystems
 - Mounting XFS Filesystems
- XFS Architecture & Internals
 - File and Directory Operations
 - On-disk Format
- Supporting XFS
 - Repairing XFS Filesystems
 - XFS Triage
- Related Products
 - Backup and Restore
 - DMAPI
 - Volume Managers

Course Objectives

- By the end of this course you will be able to
 - Create and mount XFS filesystems
 - Understand how XFS
 - Creates and manages metadata
 - Allocates extents to files and manages free space
 - Provides extended attributes
 - Tracks filesystem quotas
 - Backup and restore an XFS filesystem
 - Recover and repair XFS filesystems
 - Interpret on-disk and in-core XFS structures

Day 1

- Theory

- Background and History
- XFS Build
- XFS Overview
- Creating Filesystems
- Mounting Filesystems
- Allocators
- Quotas
- Extended Attributes

- Practise

- XFS Build
- Creating and Mounting XFS
- Allocators
- Quotas
- Extended Attributes

Day 2

- Theory
 - XFS Architecture and Internals
 - QA
- Practise
 - XFS On Disk Format
 - QA

Day 3

- Theory & Practise
 - Repairing XFS Filesystems
 - XFS Triage
 - Monitoring

Day 4

- Theory
 - Dump and Restore
 - DMAPI
 - XFS and Volume Managers
- Practise
 - Dump and Restore
 - DMAPI

A Brief History of XFS

- Original design was circulated within SGI in October 1993:
- **xFS: the extension of EFS**
 - "x" for to-be-determined (but the name stuck)
 - Large filesystems: one terabyte, 2^{40} , on 32 bit systems; unlimited on 64 bit systems
 - Large files: one terabyte, 2^{40} , on 32 bit systems; 2^{63} on 64 bit systems
 - Large number of inodes
 - Large directories
 - Large I/O
 - Parallel access to inodes
 - Balanced tree (btree) algorithms for searching large lists
 - Asynchronous metadata transaction logging for quick recover
 - Delayed allocation to improve data contiguity
 - ACL's --Access Control Lists (see `chacl(1)`, `acl(4)`, `acl_get_file(3c)`, `acl_set_file(3c)`)
- First released in IRIX 5.3

XFS on Linux

- Port to Linux began in 1999 against 2.3.40
- Accepted into mainline
 - 2.5 kernel (2002)
 - 2.4 kernel (2004)
 - SLES9 and beyond
- All XFS engineering is now based in SGI's Melbourne, Australia office
 - Coordinate changes to Novell via John Hesterberg's group in Eagan
- A number of contributors in the community
 - Many (but not all) are ex-SGI employees

Who is using XFS

- <http://oss.sgi.com/projects/xfs/users.html>
 - List is out of date but it gives an indication of the spread of users
- XFS is included in a number of distributions
 - Support agreement with Novell for SLES distributions

XFS Distributions – linux/fs/xfs

- Top of tree (tot) XFS on oss.sgi.com
 - Changes appear here shortly after they are checked in internally
 - CVS tree on oss.sgi.com
 - <http://oss.sgi.com/cgi-bin/cvsweb.cgi/linux-2.6-xfs/>
 - <http://oss.sgi.com/cgi-bin/cvsweb.cgi/linux-2.4-xfs/>
 - <ftp://oss.sgi.com/projects/xfs/download>
- Changes routinely pushed to 2.6 mainline kernel updates
 - `git://oss.sgi.com:8090/xfs/xfs-2.6`
 - Maintenance only for 2.4 kernels
- SuSE major releases and service packs
 - As close to top of tree as possible at the time of the code drop
 - Differences between tot, SuSE and mainline kernels highlighted in course notes

XFS Distributions – xfs-cmds

- A large amount of common code between kernel and user-space
 - This is how xfs_repair understands the on-disk format
- User-space commands in a different code base
 - <http://oss.sgi.com/cgi-bin/cvsweb.cgi/xfs-cmds/>
 - This includes test framework under xfstests
- Packaged as
 - xfsprogs
 - xfsprogs-devel
 - xfsdump

XFS Distributions - dmapi

- DMAPI has not been accepted into mainline
 - Unlikely to be accepted without a complete rewrite
- XFS tot distribution on oss.sgi.com includes DMAPI
 - Changes we make to DMAPI will immediately appear in the CVS tree

IRIX vs Linux

- Linux:
 - Does not support V1 directories
 - Filesystem block size \leq PAGE_SIZE only
 - Does not support case insensitive directories
 - Does not support 64 bit inode numbers on 32 bit platforms
 - Will need XFS changes once generic changes in
 - Does not support filestreams allocator
 - Development in progress (Oct 06)
- IRIX:
 - Does not support >512 byte sector sizes
 - MD RAID5
 - Does not support write barriers
 - Does not support “noikeep” functionality

Ongoing Development

- XFS Architecture Team
 - Submits and reviews designs for adding new features, or enhancing existing features, targeting one or more of
 - Interoperability
 - Performance
 - Scalability
 - Reliability
- XFS Triage Roster
 - Engineer rostered for one week to triage incoming bug reports from
 - SGI
 - Novell
 - Community
- Remaining time devoted to developing features and fixing bugs

sgi®