Linux File System

Xie Huagang
xie@chinacluster.com

China Cluster Networks.
Nov 10, 2000

1

Introduction to Filesystem

- ✧ files and directory
  - ✧ normal file
  - ✧ link and fifo
  - ✧ device
- ✧ file attrib
  - ✧ uid,gid
  - ✧ misc
- ✧ file description
- ✧ file I/O
  - ✧ read(),write()

Virutal File System

- ✧ History
  - ✧ s5fs.
  - ✧ vode/vfs from SUN
- ✧ VFS target
  - ✧ support multi-filesystem
  - ✧ sigle-image from varies partions and filesystems
  - ✧ support Network shareing files.
  - ✧ ISV can devlop filesystem via modules.

3

VFS Overview

- ✧ vnode present a file in a filesystem
- ✧ vfs present a filesystem in the kernel.
  - ✧ s5fs, msdos, ext2, nfs.
- ✧ Object Oriented Design
  - ✧ data & function
  - ✧ inherited from base class
- ✧ vnode/vfs abstract
  - ✧ data, struct inode, struct file...
  - ✧ function, struct inode_operations, file_opereations.etc.

Linux VFS Implemetation

- ✧ Data Structure
  - ✧ inode
  - ✧ file
  - ✧ superblock
  - ✧ vfsmount
- ✧ Function member
  - ✧ inode_operations
  - ✧ file_operations
  - ✧ misc

5

---

struct inode

```
struct inode {
        struct list_head i_hash;
        ......
        unsigned long i_ino;
        unsigned int    i_count;
        kdev_t          i_dev;
        ......
        uid_t           i_uid;
        gid_t           i_gid;
        ......
        struct inode_operations *i_op;
        struct super_block  *i_sb;
        ......
        union {
                struct pipe_inode_info pipe_i;
                struct ext2_inode_info ext2_i;
                struct socket          socket_i;
        } u;
}
```

struct file & struct dentry

```
struct file {
        struct  file  *f_next, **f_pprev;
        struct  dentry *f_dentry;
        struct file_operations *f_op;
        ......
        loff_t  f_pos;
        unsigned int  f_uid, f_gid;
        ........
}

struct dentry {
        int d_count;

        struct inode *d_inode;
        struct super_block *d_sb;
        ........
}
```

struct inode_operations

```
include/linux/fs.h

struct inode_operations {
        struct file_operations * default _file_ops;
        int (*create) (struct inode *,struct dentry *,int);
        struct dentry * (*lookup) (struct inode *,struct  dentry *);
        int (*link) (struct dentry *, struct inode *, struct dentry *);
        ...
}

fs/ext2/file.c

struct inode_operations ext2_file_inode_operations = {
        &ext2_file_operations, /* ...*/
        NULL,
        ...
}
```

struct file_operations

```
linux/linux/fs.h
struct file_operations {
        loff_t (*llseek) ( struct file *,loff_t, int);
        ssize_t (*read) (struct file*,char *,size_t, loff_t *);
        ssize_t (*write) (struct file*, const char *..);
        ...............
        int (*ioctl) (struct inode *,...)..
        ............
}

fs/ext2/file.c
struct file_operations ext2_file_operations = {
        ext2_file_lseek,
        generic_file_read,
        ext2_file_write,
        NULL,
        ..............
};
```

FILE operations Implementation

- ✧ sys_open
  - ✧ fd, flip_open
    - ✧ file = flip_open()
      - ✧ dentry = open_namei()
        - ✧ *looup_dentry*
        - ✧
    - ✧ fd_install()
- ✧ lookup_dentry
  - ✧ base = pwd or root
  - ✧ for (;;)
    - ✧ dentry = real_lookup(...);
  - ✧ ..

sys_read()

- ✧ sys_read()
  - ✧ read = file->f_op->read
  - ✧ ret = read(file,buf,count, &file->f_pos);
- ✧ sys_write()
  - ✧ read = file->f_op->write
  - ✧ ret = write(..);

Mount

- ✧ mount a filesystem
- ✧ struct vfsmnt
- ✧ sys_mount
  - ✧ do_mount()
    - ✧ dir_d = namei(dir_name);
    - ✧ sb = read_super();
    - ✧ vfsmnt = add_vfsmnt();

File system

✧ ext2
✧ reiserfs
✧ ext3

13

Distrubuted File System

✧ NFS
✧ GFS
✧ Coda
✧ Inter-mezzo

File System and LIDS

✧ Linux File System Security
  ✧ File is unprotected
  ✧ root is God.
✧ LIDS
  ✧ intercept system call
  ✧ check everytime , sys_open,
    sys_chmod..
  ✧ Reference Monitor
  ✧ Mandatory Access Control
  ✧ Capablity of process
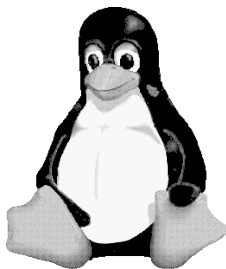
15

LIDS

✧ HOME
  ✧ www.lids.org
✧ Mail List
  ✧ lids-subscribe@egroups.com
✧ Version,
  ✧ 0.9.9 for 2.2.17
  ✧ 1.0.2 for 2.4.0-test9

Thanks

17