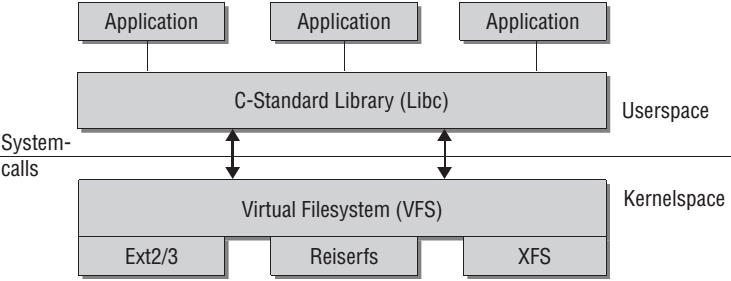
Hadoop Fuse

# 一、简介

## 1.1 VFS

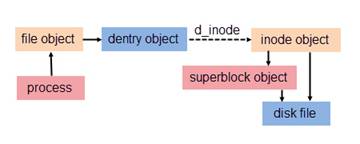
VFS是Linux核心的概念，Linux大部分的操作都要用到VFS相关功能。VFS（Virtual Filesystem Swith）是Linux中的一个虚拟文件文件系统，称为虚拟文件系统交换层。它为应用程序提供一层抽象，屏蔽底层各种文件系统的差异，如下图所示：



Linux文件系统架构

不同的文件系统，如Ext2/3,XFS,FAT32等，具有不同的结构，用户对文件相同的操作，不同的文件系统的实现是不同的。为了屏蔽这种差异，Linux引入了VFS的概念，相当于Linux构建了新的贮存在内存中的文件系统，所有其他文件系统都需要先转换成VFS的结构才能会用户所调用。

Linux VFS存在四个基本对象，超级块对象（superblock object），索引节点对象（inode object），目录项对象（dentry object）及文件对象（file object），VFS对象之间的处理如下图所示：



VFS对象之间的处理

超级块对象代表一个已安装的文件系统，索引节点对象代表一个文件，目录项对象代表一个目录项（如设备文件event5在路径/dev/input/event5中），文件对象代表由进程打开的文件。在图中，d\_inode为硬链接，为文件路径的快速解析，Linux VFS设计了目录项（Dentry cache）。

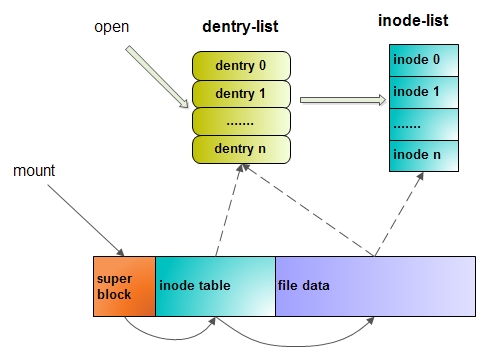
VFS主要由dentry与inode构成。Dentry用于维护VFS的目录结构，每个dentry项代表ls时显示的目录和文件。Inode为文件节点，与文件一一对应。在Linux中，目录也是一种文件，所以dentry也会对应一个inode节点。下图是目录模型在VFS结构：

|  |  |
| --- | --- |
| C:\Users\yshe\Desktop\0S132C39-0.jpg | C:\Users\yshe\Desktop\0S132A24-3.jpg |

目录结构在VFS中的结构

## 1.2 VFS的构建

VFS的构建就是加载实际文件系统的过程，也就是mount被调用的过程，以ext2文件系统加载为例，如下图：



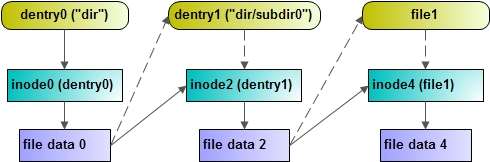
文件系统加载

mount时，linux首先找到磁盘分区super block，然后通过解析磁盘的inode table与file data，构建自己的dentry列表与inode列表。Ext的inode节点在磁盘上，VFS的inode节点在内存中。

当用户用ls命令查看某一个目录或用open命令打开一个文件时，VFS会为用到的每个目录项与文件建立dentry项与inode(按需创建)，采用dentry cache的设计，然后维护一个LRU列表，当linux认为VFS占用太多资源时，VFS会释放掉没有被使用的dentry项与inode项。

## 1.3 定位文件

在VFS中并不能随时都有dentry项与inode项可用，当无dentry项与inode项时，定位文件的过程如下图：



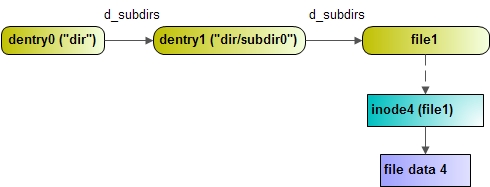
无dentry项时定位文件

1）首先通过dir对应的dentry0，找到inode0节点，从inode0节点读取目录中的信息（file data），包含该目录包含的下一级目录与文件的文件列表，包括name与inode号（用ls命令查看的就是这些信息）。

2）通过subdir0对应的inode号重建inode2，并通过文件数据与inode2重建subdir0的dentry节点：dentry1

3）根据file1对应的inode号重建inode4，并通过文件数据与inode4重建dentry节点，最后通过inode4节点访问文件。

一旦在Dentry cache中建立了dentry项，下次访问就可以快速定位文件，如下图所示：



有dentry项时定位文件

1）在代表dir目录的dentry0的d\_subdirs中查找名字为"subdir0"的dentry项，找到dentry1.

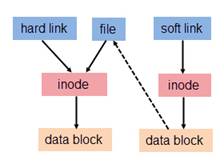
2）在dentry1中查找名字为file1的dentry项，然后找到file1对应的dentry

3）通过file1对应的dentry项获得file1对应的Inode4

在dentry cache中有dentry项时，操作简单很多

## 1.4 Symbolic link与hard link

为了解决文件的共享使用，Linux系统引入了两种链接：硬链接（hard link）和软链接（符号链接，soft link或symbolic link），链接还能带来隐藏文件路径、增加权限安全及节省存储等好处，硬链接与软链接的概念如下图所示：



硬链接与软连接

若一个inode号对应多个文件名，则称这些文件为硬链接(同一个文件使用多个别名)，如上图，硬链接具有以上几点特性：

文件由相同的inode和data block

只能对已存在的文件进行创建

不能交叉文件系统进行硬链接的创建

不能对目录进行创建，只可对文件创建

删除一个硬链接并不影响其他相同inode号文件

可以通过link或ln创建硬链接，命令如下：

1. $link old.file hard.link
2. //通过以下的命令,查看文件的具体信息
3. $stat old.file
4. $stat hard.link

文件old.file与hard.link有相同的inode号及文件权限。

软链接与硬链接不同，软链接指向的用户数据块内存放的内容是另一文件的路径名的指向。软链接是一个普通文件，只是数据内容有点特殊，软链接有着自己的inode号以及用户数据块。因此软链接的创建与使用没有类似硬链接的诸多限制

软链接有自己的文件属性及权限

可对不存在的文件或目录创建软权限

软链接可交叉文件系统

软链接可对文件或目录创建

创建软链接时，链接计数i\_nlink不会增加

删除软链接并不影响被指向的文件，但若被指向的原文件被删除，则相关软链接被称为死链接（dangling link）

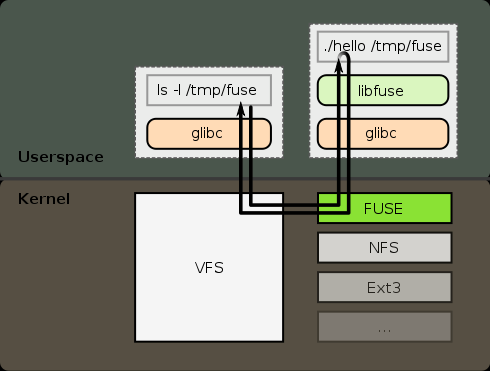
通过ln命令创建软链接：

$ln -s old.file soft.link

# 二、Fuse

Fuse（FileSystem in Userspace），用户空间文件系统，是类Unix系统平台上可加载的内核模块，允许非特权用户创建功能完备的文件系统，而不需要重新编译内核。Fuse模块仅仅提供kernel模块的接入口，而本身的主要代码主要位于用户空间中。对于读写虚拟文件系统来说，Fuse是不错的选择。但是在用户态文件系统必然会引入额外的内核态/用户态切换带来的开销，对性能会产生一定影响。

Fuse的框架如下图所示：



Fuse 运行机制

Fuse包括三个模块：内核模块，用户空间库以及mount工具，其中内核模块实现了一个完整文件系统的框架，但具体操作没有实现（由程序员在用户空间实现）。用户空间库给程序员提供编程接口，程序员通过实现fuse提供的fuse\_operation或fuse\_ll\_operation的所有或者部分操作来实现文件系统。mount工具fusermount用于挂载用fuse实现的文件系统。

用户空间进程在实行操作文件的系统调用的时候，在内核空间，VFS会调用各文件系统定义的对应操作函数。Fuse内核模块中被定义的操作函数，把和它对应的请求送到实现文件系统的用户空间进程（Fuse文件系统，后台程序），并等待回应。因为fuse涉及到多次内核态到用户态的切换，性能有一定的降低。

# 三、Hadoop Fuse

Hadoop源码中自带了contrib/fuse-dfs模块，用于实现通过libhdfs和fuse将HDFS mount到linux本地。在linux 内核2.6.14以上版本添加了Fuse模块，不需要再安装fuse组件。

使用的Hadoop版本是hadoop-2.3.0-cdh5.0.0，cloudera的开源Hadoop，在cdh中，hadoop-fuse被编译成rpm安装包（需要在src中添加hadoop-fuse 子项目），链接如下：

http://archive.cloudera.com/cdh5/redhat/6/x86\_64/cdh/5.0.0/RPMS/x86\_64/。

1）编译libhdfs组件

首先安装ant

*$yum -y install ant*

通过yum安装ant，环境变量会自动配置，略过

然后打开$HADOOP\_HOME，编译libhdfs

2）fuse编译

fuse-dfs所处的目录为：$HADOOP\_HOME/src/hadoop-hdfs-project/hadoop-hdfs/src

/main/native/fuse-dfs

3）ambari fuse的安装

# 四、Ambari的安装

1）下载repo

2）安装Ambari-server

*yum -y install ambari-server*

配置jdk及数据库:ambari-server setup

启动：ambari-server start

3）安装Ambari-agent

*yum -y install ambari-agent*

配置agent

/etc/ambari-agent/conf/ambari-agent.ini中配置hostname

agent启动：ambari-agent start

4）安装服务

安装HDFS

安装Hdfs后，已经安装后了fuse

使用命令：

hadoop-fuse-dfs dfs://ambarimaster:8020 testfuse/

将hdfs挂载到testfule