

《振南 znFAT--嵌入式 FAT32 文件系统设计与实现》一书

【上下册】已正式出版发行

全国各渠道全面发售

关键字"znFAT"即可购买,各地实体书店也有售)

此书是市面上 唯一 一套详细全面而深入讲解嵌入式存储技术、FAT32 文件系统、SD 卡驱动与应用方面的专著。全套书一共 25 章,近 70 万字。从基础、提高、实践、剖析、创新、应用等很多方面进行阐述,力求通俗,振南用十年磨一剑的精神编著此书,希望对广大工程师与爱好者产生参考与积极意义。

此书在各大电子技术论坛均有长期的「抢楼送书活动」,如 211C、elecfans 等等。

振南的【ZN-X 开发板】是市面上唯一全模块化、多元化的开发板,可支持 51、AVR、STM32(M0/M3/M4)

详情请关注 WWW. ZNMCU. CN (振南个人主页!!)

新入渐深,初触核心: 展示 FAT32 绝妙的簇链结构

上一章实现了带 znFAT 雏形的 SD 卡 MP3 数码相框实验,程序已经能够自主进行文件解析,完成文件信息获取与数据定位了,但是最后也提出了一个有关数据连续性的问题。其实在实际读取文件数据的时候,很多情况下数据是不连续的。(为什么会不连续? 这个在第 3 章中其实已经讲过,主要是因为文件删除后遗留的剩余空间被重新利用所造成的。)这一章就来讲解 FAT 表的相关内容,它是 FAT32 中实现数据链式存储的主要手段,是最为核心的部分,整个文件系统都以 FAT 来命名就足以证明这点。

8.1 初识 FAT 表

FAT 表是 FAT32 文件系统的核心功能部分,记录了文件数据之间的链式存储 关系。说白了,我们解析文件目录项得到文件数据的开始簇之后,通过 FAT 表就可 以知道后面的数据存储在哪些簇上了,进而完成数据的读取。

1. 链表思想

链表,基本属于计算机专业数据结构的范畴,电子以及嵌入式相关专业的人可能觉得比较陌生,即便有所了解,应用也不会很多。

讲述链表最好的办法就是用链表图来进行说明,如图 8.1 所示。图中一共有 6 个节点,并用 1~6 进行了编号。每一个节点除了记录自身的编号外,还记录了下一个节点的编号。我们选定一个节点为开始节点,经过一系列的指向,一直到结束节点。这样就产生了形如锁链的结构,各个节点环环相扣,表达了一个链式关系,2-

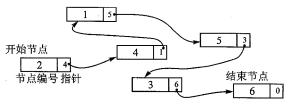


图 8.1 包含 6 个节点的链表

B

嵌入式 FAT32 文件系统设计与实现——基于振南 znFAT(上)

4-1-5-3-6。当然,链表的指向关系是可以重新定义。如图 8.2 所示,它的链式关系为 4-5-1-2-6-3。

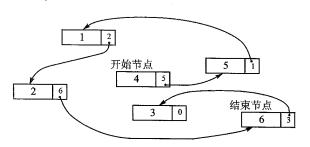


图 8.2 链式关系被重新定义后的链表

讲到这里,读者是不是有所启发呢?链表中节点的位置或者说编号都是固定不变的,通过设置不同的节点指向便可以极为灵活地描述节点之间的链式关系,从而在逻辑上构造了一种节点的次序。如果把节点看成是簇,那它会表达一种什么意义呢?关于链表就说这些,只是让大家有一个基本的思想。

2. FAT 表对簇链的表示

上面振南点了一笔,把节点看成簇。很显然,它可以描述簇间(数据存储单元之间)的链式关系。基于这种机制,一个文件的数据所包含的若干个簇就可以不再连续存储了,有一条"锁链"把它们串了起来。理解了这一点,我们就更加关心 FAT 表是如何具体表示簇链关系的。在上面链表的例子中,每一个节点都记录了自身的一个编号。如果把这些节点按其编号排成一排,此时节点的位置就成了其"天然的编号"。它们便无需再记录自身编号,只需记录下一节点的编号即可,如图 8.3 所示。

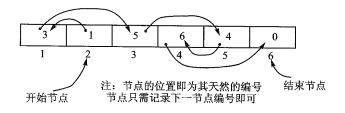


图 8.3 6 个节点顺序排列后所构成的链表

从图 8.3 中可以看到,在给定了开始节点之后,节点的位置与指向关系对链表进行了表达。当节点的数量比较多的时候,这排成一条"长龙"的诸多节点就形成了一个线性的"表"了,如图 8.4 所示。

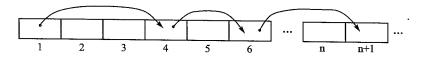


图 8.4 线性表形式的链表

有人说:"这个线性表形式的链表从存储结构上看就是一个一维数组。"没错!其实这已经与实际的FAT表很相似了。FAT表就是由很多顺序排列的簇项(相当于节点)构成的大数组。真正的FAT表的样子如图 8.5 所示。图 8.5 虽然看似二维结构(有行有列),其实只是因为一行放不下那么多的簇项,换了个行而已,实际上就是一个一维结构。

3. FAT 表项的定义

FAT 表的表项(也叫簇项)是簇链表中的节点,用于记录下一个簇的簇号,诸多的簇项顺序地存储在 FAT 表的扇区上。有人问:"FAT 表中一共有多少个簇项?"答:那取决于分区的存储容量。可以

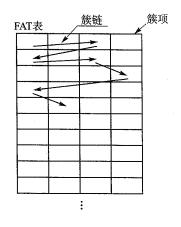


图 8.5 FAT 表表达簇链的形式

说,数据区(DA)有多少个簇,FAT 表中就有多少个簇项,这个道理是不难理解的。 这里比较关心的是 FAT 表项具体是怎样的? 其实非常简单,FAT32 中的簇项就是 一个 32 位无符号整型值(即 4 个字节)。

此时,如果有人问:"簇项为什么是 32 位的? 16 位行不行?"那振南会觉得很高兴,这算是问到"点儿"上了。其实 FAT32 之所以被称为 FAT32,就是因为簇项长度是 32 位,足以看出簇项的定义对 FAT32 文件系统所起的决定性作用。

FAT 表中的簇项与数据区中的簇是一一对应的。假设簇项是一个 16 位的整型值,那么它能表达 65 536 个数值,也就是说,最多只能表达 65 535 个簇链关系。但是如果分区的存储空间比较大,比如说有 90 000 个簇,那簇项将不能表达簇号大于 65 535 的簇的簇链,于是那些无法被表达的簇将无法被利用起来,从而造成存储空间的严重浪费,如图 8.6 所示。所以,簇项的长度直接影响到了文件系统能够支持的最大容量。这是一个非常重要的指标,也是评定一个文件系统是否实用,是否优秀的重要因素之一。

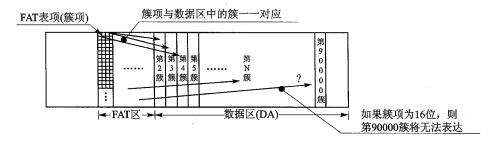


图 8.6 FAT 簇项长度不够造成簇无法利用

有人说:"想要支持更大的容量,簇项如果不够用,那可以把簇定义得大一些,这样簇数不就少了吗?"FAT32 中把最小存储单元重新定义为簇,确实也有这样的用意,用于提高文件系统所能够管理的有效容量,但一味地增加簇的容量,同样也会造

B

嵌入式 FAT32 文件系统设计与实现——基于振南 znFAT(上)

成严重的空间浪费。因此,这不是一个根本的解决方法。

看了上面的这段讲解,你有没有想到一个相反的方面呢?如果簇的容量过小,将使得簇数增加,从而进一步加剧簇项不够用的窘境,唯一的办法就是增加簇项的长度,使其可以表达更大的数值,但是 FAT 表的体积也会随之扩大,从而占用更多的存储空间。所以,在对存储设备进行格式化的时候,文件系统都在进行一种权衡,一种簇容量与存储空间使用率之间的权衡。

很多人在格式化的时候都看到,除了 FAT32 以外还有 FAT(FAT16)和 FAT12 (FAT12 可能现在见得不是很多了),FAT 后面的数字其实就是簇项的位数,这是它们之间最大的不同。其实它们是 FAT 文件系统顺应存储设备容量不断增长而产生的不同版本(即存储设备容量提升了,原来的文件系统用不了了,需要有新的版本来支持),这也体现了存储设备的飞速发展。想想我们现在一个 U 盘、一张存储卡或是一个硬盘,动辄几百 M、几十 G,甚至上 T,但是几年以前提到 M 这个单位就已经觉得是天文数字了。有人问了,FAT 文件系统是不是还在不断更新版本呢,毕竟存储设备依然在不断发展。不错,FAT 系列文件系统的整个演化过程,如图 8.7 所示。

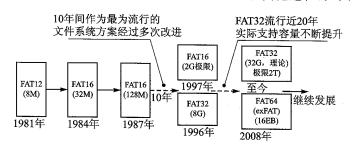


图 8.7 FAT 文件系统的演化与发展

4. 表项的属性

我们已经知道表项是用于记录簇号的,但这只是它最普遍的一种用途。除此之外,它还可以用于标记簇的属性。为了说明这个问题我们来看图 8.8。可以看到, FAT 表项除了有指向的作用以外,还可以用于表达某个簇是否可用、是否完好、是否是簇链的最后一个簇等意义,具体取值如表 8.1 所列。

这就告诉我们,并不是所有的簇项都是有效的。在对簇项进行读/写操作的时候,需要依其属性进行相应的处理,比如在读取的时候,要判断它是不是最后一个簇,写入的时候,要保证簇项是可用的,而不是已被占用或者其他用途。另外,还可以发现簇项的取值只使用了 32 位中的低 24 位,这主要取决于实际现行协议中的定义。本书中介绍的 FAT32 的相关定义遵循微软发布的 FAT32 协议技术手册,与当前使用的 Windows 操作系统中的 FAT32 文件系统相兼容。

另外还有一点,表项也是以小端方式存储的,所以读取它的值与前面几章中解析 参数是一样的。与表项相关的内容就介绍这么多。

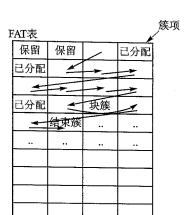


图 8.8 FAT 表项的各种属性

表 8.1 FAT 表项属性取值与意义

属性取值	意义描述
0X0000000	空闲簇
0X00000002~0X0FFFFEF	已被分配
0X0FFFFFF0~0X0FFFFFF6	系统保留
0X0FFFFFF7	坏簇
0X0FFFFFF8~0X0FFFFFF	结束簇

8.2 剖析 FAT 表及簇链

8.2.1 印证文件的连续存储

SD卡的FAT表结构如图 8.9 所示。图中用横线标注的就是FAT表中的表项,可以看到第 0 项和第 1 项的值分别为 0 X0FFFFFF8、0 XFFFFFFFF,这两项我们暂不用关心(数据区的簇号是从 2 开始的,这是 FAT32 的约定)。第 2 项的值为 0 X0FFFFFFF,根据上面的定义,它表示簇链的结束,即再无后继簇。是谁占用了第 2 簇呢?当然是首目录了,它记录了文件目录项。从第 3 项开始,一直到第 40 项(图中用【】进行了标注),似乎这之间的簇号很整齐、很连贯,这也说明了它们所对应数据区中的簇是连续的,没有出现簇的跳跃。占用这块连续存储器空间的就是 PICTURE. BIN 文件。同样的道理,占用 41~440 簇的文件是 TEST. MP3。它们两者的 FAT表项都很连续,这就印证了其数据也是连续存储的(连续存储是 MP3 数码相框实验的前提,只有这样才可以依次按扇区读取数据,而不用顾及簇链)。

有人在这里产生了疑问:"图 8.9 中对于 TEST. MP3 文件,中间的簇项被省略了,我们看不到它所有的簇项,怎么能印证它是连续呢?"嗯,因为篇幅原因这里进行了删减,可以通过 WinHex 软件对簇链的列举功能来看看它到底是不是真的连续,如图 8.10 所示。

图 8.10 中的碎片数量就能告诉我们,TEST. MP3 的簇是连续的。(碎片即磁盘碎片,其实就是指一个连续存储空间的片段。如果 TEST. MP3 文件的数据全都是连续的,那它当然只有一个碎片了。)至于 WinHex 的簇列表功能在哪里调出,请看图 8.11。

此文因版权仅节选一部分,请各位读者见谅!! 完全内容请购买正版书籍!!

感谢对振南及 znFAT 的关注与支持,希望振南在 嵌入式 FAT32 文件系统方面的研究对您有所帮助! 更多内容请关注 振南电子网站

www.znmcu.cn

