

多用户系统下压缩备份的实现

陈元中 (人民银行东莞市分行科技科 257091)

1. 问题的提出

在 XENIX/UNIX 系统环境下支行的业务程序, 数据备份时数据量一般比较大, 动辄就是几张软盘, 由于软盘可靠性差, 直接使用传统的系统命令进行数据备份, 费时、费力、费盘, 容易出错。针对这种情况, 笔者采用 C 语言编写了一段实用程序, 比较好地解决了这个问题。

2. 程序设计思路

分析造成上述问题的原因, 发现问题出在所要备份的数据量上。数据量越大, 问题表现得就越突出。为此我们只能在数据量上作文章, 采取手段对数据进行压缩, 备份时仅备份压缩数据, 恢复时系统自动解压缩, 整个过程对用户来说都是透明的, 业务处理无需任何额外操作。

在 XENIX/UNIX 系统下一般调用系统命令 tar 进行备份, 我们可截获这个系统调用, 掌握系统进行数据备份、恢复操作的控制权。方法很简单, 只需把命令 tar

改名为 systar, 然后再把自己编写的实用程序命名为 tar, 存放在“/usr/bin”目录下。这样一来, 再调用 tar 命令, 实际上调用的是我们自己编写的程序。可以在程序里对命令进行分析, 若是从硬盘到软盘备份数据, 则先压缩数据, 然后再调用 systar 备份压缩后的数据; 若是从软盘到硬盘恢复数据, 则先调用命令 systar 进行数据恢复, 恢复完成后, 还要对数据进行解压缩操作; 其他情况, 直接调用系统命令即可。

考虑到同一系统下可能会有多个注册用户, 有些用户备份时不希望对数据进行压缩。程序中增加了一个模块, 用以判断用户是否是需数据压缩的用户, 如若不是, 则直接调用原系统命令 systar 进行数据备份或恢复。

数据备份时调用 compress 命令压缩数据, 具体操作中若直接对备份数据进行压缩, 将导致一些问题。数据压缩后, 文件名有所改变, 增加了后缀“.Z”。这样一来, 备份完数据后还需对原文件解压还原, 花费一定时间;

更为重要的是在解压结束前,可能会遇到系统掉电一类的问题,系统重新启动后业务程序将不能正常运行,因为有些数据还没有解压还原;另一方面,XENIX系统的文件名长度有限,最多为14个字符,对那些文件名长度多于12个字符的文件,compress命令是无能为力的,影响了数据压缩率。为了保证数据绝对安全可靠,力争取得更好的压缩效果,尽量避免由此引起一些副作用,程序设计时并不直接压缩原文件,而是将原文件A拷贝成另一文件B,压缩的是文件B,备份的是文件“B.Z”。借助这种中间转换机制,很好地解决了上述问题。

3. 程序编码说明

程序主要有data-backup()和data-restore()两个功能模块,分别完成数据备份和数据恢复,其他模块都是为这两个模块服务的。下面对各模块作些扼要说明。

模块chr-in-str(c, str)用来判断字符c是否包含在字符串str之中,若包含则返回1,否则返回0。tar命令是备份还是恢复,取决于该命令的第一个参数。若第一个参数中含有c或r则说明该命令执行备份功能;若含有x则说明要进行数据恢复。本模块主要用以类似的判断。

模块parameter-type(parameter)用来判断参数parameter的类型。tar命令的参数可分为控制参数、块设备文件、其他普通文件、目录等四种类型。程序中正确区分这些参数类型,对构造数据的备份和恢复命令是极为重要的。

模块direct-sys-cmd(argc, argv)用来构造系统备份命令,当不需要压缩时,将直接调用该模块,执行原系统命令systar。参数argc, argv来自主程序main(argc, argv)。

模块test-user(argc, argv)用来判断用户是否需要压缩的用户,若不是则直接执行模块direct-sys-cmd()。需要压缩的用户名事先存放在字符数组user-name[][10]中,程序根据该数组判断用户是否需要压缩。

模块data-backup(argc, argv)首先分析主程序传递来的参数argv,利用“ls-R参数”统计出所有备份文件,并存放于临时文件TMP-DIR-FILE中,该文件是进行压缩、备份的依据。程序依次从该文件里读行并对该行进行分析,结果有三种可能:①该行是空行,下一行肯定是目录名,这是由上述统计命令决定的,此时需要再读一行,记下该目录名,用以构造文件的绝对路径名或相对路径名;②该行是当前目录下的目录名,忽略此行;③该行

是普通文件,这时要看该文件是否已压缩,若已压缩则忽略此行,否则将文件A拷贝到文件B,压缩文件B,并将文件A和文件B.Z的对应关系存到文件DIR-FILE中。这个文件至关重要,数据恢复时它决定着文件B.Z到文件A的转换,因此数据备份时将该文件备份了两次。在进行数据压缩的同时,程序还根据分析结果,重新构造备份命令,完成压缩数据的备份。

模块data-restore(argc, argv)首先分析主程序传递来的参数argv,以此构造数据恢复命令,将软盘上的所有数据拷贝到硬盘。在数据拷贝完成后,程序根据DIR-FILE的内容,将数据压缩还原。

主程序main(argc, argv)调用test-user()判断用户是否需要压缩,若不需要压缩,则直接调用direct-sys-cmd(),执行原系统命令systar;否则将根据第一个命令参数,分别调用不同的功能模块,完成数据备份、恢复以及systar命令的其他功能。

关于程序的安装,大体上可按以下步骤进行:

- ①用vi编辑源程序,生成文件file.c。
- ②用cc编译文件file.c,生成可运行文件file。
- ③在超级用户提示符下,执行以下命令,完成安装:

```
# mv /usr/bin/tar /usr/bin/systar
# mv ./file /usr/bin/tar
# chmod 0711 /usr/bin/tar
# chown bin /usr/bin/tar
# chgrp bin /usr/bin/tar
```

4. 结束语

该程序在XENIX/UNIX系统下已成功运行多年,效果很好,表面上看程序采用以时间换空间策略,其实不然。实践表明,compress命令的压缩效率很高,对业务数据来说,压缩数据仅为原数据的三分之一,甚至更少,原来需三张软盘,现在仅需一张就够了。目前机器档次普遍较高,运算速度快,数据压缩时间比同样数据软盘备份的时间要少得多。因此,该方法不仅减少了软盘备份空间,也大大减少了备份所需时间,在很大程度上提高了备份数据的可靠性。对那些没有磁带机、备份数据又较多的用户,不失为一种很好的备份方法。

(程序清单较长,本刊删略,如读者需要,可与作者联系)

(来稿时间:1997年9月)