

文章编号:2096 - 5389(2020)05 - 0105 - 04

一种保存 Micaps 数据资料的思路和方法

张 华,李细生

(湖南省株洲市气象局,湖南 株洲 412003)

摘 要: Micaps 系统历史数据要素多,占用空间大,不易于长期不间断地保存。该文介绍一种截取适当范围区域、删除不常用要素、改写数据存储格式的思路和方法以压缩数据文件,保存重要的资料。通过最近 2 a 的业务实践应用表明该方法稳定性较高,实用性较好,且操作简便,可解决基层台站海量历史资料的备份难题。

关键词: MICAPS4; 大数据; 历史资料; .NET; 备份

中图分类号: TP311.52 **文献标识码:** B

A New Way to Store MICAPS Data

ZHANG Hua, LI Xisheng

(Zhuzhou Meteorological Bureau of Hunan Province, Zhuzhou 412003, China)

Abstract: There are many elements of MICAPS historical data, which take up a large space and are not easy to keep for a long time. This paper introduces an idea and method of cutting appropriate range area, deleting uncommon elements and rewriting data storage format to compress data files and save important data. The application in the last two years shows that this method is stable, practical and easy to operate, which can solve the problem of massive historical data backup in basic stations.

Key words: MICAPS4; big data; historical data; .NET; backups

0 引言

Micaps4^[1]是目前我国大部分气象台站制作天气预报的操作平台,功能众多,需海量的数据资料支持,国家气象局将这些数据以不同格式(大部分是文本格式)下发,优点是便于直接读取和访问,缺点是空间占用非常大,每天大约有 100 G,且访问效率较低。现在各个省局都建有分布式数据中心,市、县两级再无需维护数据服务器,访问速度和效率较以前大幅度提升,但仍然没有解决基层台站历史资料的长期不间断保存问题,这些海量的历史数据其实也是气象部门的宝贵资源和财富,白白流失实在浪费。基于此,可设计一个资料存储程序,通过一些必要的删减和压缩手段,将每天的资料打包到一个数据文件,既大大节约存储空间,又便于资

料的交流和使用以及日后的分析应用,从而打好气象部门的“大数据”^[2]这张牌。

1 压缩数据文件的主要思路

为了达到数据文件大幅度瘦身的目标,保存资料时必须要有相应的删减和压缩技巧,可通过如下几种方式缩减容量:

1.1 缩小资料覆盖范围

Micaps 资料一般有全球范围,东亚范围等,基层台站如果研究的区域不需要那么大,则可将范围适当缩小。如中央台下发的 EC 细网格资料范围是 73° ~ 135°E、18° ~ 54°N,总共有 $497 \times 289 = 143\ 633$ 个网格点,网格距是 0.125 个经纬度。假如研究目标区域仅仅为江南、华南等地,那么可以将经纬度范围截取为:100° ~ 130°E、20° ~ 40°N,如此一来,网

收稿日期:2019 - 09 - 23

第一作者简介:张华(1975—),女,工程师,主要从事气象服务工作,E-mail:271418415@qq.com。

格点变为 $161 \times 161 = 25\,921$ 个,数据量变为原来的 $1/6$ 。

1.2 删除不常用的要素或产品

Micaps 资料里占用空间较大的有卫星数据产品、雷达产品、数值模式产品等,如果去除这些,或者只选择少数几个精品进行保持,也可减少一半以上存储容量。

1.3 将文本格式的数据改写成二进制格式

Micaps 后台使用的数据很多为文本格式,它是一种典型的顺序文件,其文件的逻辑结构又属于流式文件。可以在 Unix、Macintosh、Microsoft Windows、DOS 和其它操作系统之间自由交互,文本文件编码基于字符定长,译码容易但占用空间较大;二进制文件编码是变长的,所以它灵活,存储利用率高,占用空间较小。因此,可以利用二进制文件占用空间小的优点,通过程序将文本文件中的数据按一定的顺序写入二进制 DAT 文件,从而缩减容量。当然,二进制文件也有可读性差的缺点,所以数据保存好以后一定要用专门的说明文件对数据的格式和顺序进行单独说明,并与数据文件存放在同一路径下。

2 写入二进制文件的源码和注意事项

下面以 ECMWF_THIN 某个时次的低温预报数据为例来说明改写程序的编写。编程语言为 VB.NET,源程序用 Microsoft Visual Studio 2010 编译^[3]通过,代码的功能是将某一时间点 EC 细网格 2 m 最低气温预报(间隔 6 h、共 38 个预报时次)通过范围截取后写入二进制 DAT 文件,设计思路是将原文本数据文件去除文件头,选取适当范围数据再改变顺序写入二进制文件 *.dat。源代码如下:

```
Public Function DuECTmn ( ByVal Parth As String)
...
Dim Jd(2), Jdt(2) As Single 原经度范围,细网格 73 < jd < 135
Dim Wd(2), Wdt(2) As Single 原纬度范围,细网格 18 < wd < 54
Dim T1(17), Tm1(38) As Integer
Dim T2(17), Tm2(38) As String
Dim myFileTmn As New IO. FileStream( "d:\tftx\dat\" & ComboBox4. Text & "Tmn6. dat", FileMode. Create, FileAccess. Write) '创建新的二进制文件。
Dim bwTmn As New BinaryWriter( myFileTmn, System. Text. Encoding. UTF32) '定义二进制文件的数据压缩格式为 UTF32。
```

For K = 1 To 38 按顺序循环读取低温文件 38 个(6h 间隔)

txtFmn6 = TxtTmn6 ° ComboBox4. Text ° Tm2 (K) EC6 h 低温全路径文件名。

If My. Computer. FileSystem. FileExists(txtFmn6) = False Then ‘如果原始资料缺失则提示并退出。

MsgBox(txtFmn6 & " 文件缺失!")

bwTmn. Flush()

bwTmn. Close()

myFileTmn. Close()

Exit Function

End If

Dim strRead As String() = IO. File. ReadAllLines(txtFmn6)

Dim Temp(strRead. Length - 1)() As String

For i = 0 To strRead. Length - 1

strRead(i) = Trim(strRead(i))

Do While InStr(strRead(i), " ") 如果有连续的两个空格就替换成一个空格,用循环方式去除文本文件中多余的空格。

strRead(i) = Replace(strRead(i), " ", " ")

Loop

If i < 6 Then

Temp(i) = strRead(i). Split(" ") '用回车符分割数据

Else

Temp(i) = strRead(i). Split(" ") '用空格分割数据

End If

Next

Jd = {107, 117} 缩小经度范围

Wd = {23, 31} 缩小纬度范围

Dim Ds(Int((Wd(1) - Wd(0))/0.125), Int((Jd(1) - Jd(0))/0.125)) As Single

x = 0

For i = 0 To Int((Wd(1) - Wd(0))/0.125) - 1

For j = 0 To Int((Jd(1) - Jd(0))/0.125) - 1

Ds(i, j) = Temp(2 * (54 - Wd(0))/0.125 - 2 * i + 6) ((Jd(0) - 73)/0.125 + j) '将文本文件数据读入数组,并改变顺序写入二进制文件,+6 是为了去除文本文件的前 6 行(文件头)。

Gf = Val(Ds(i, j))

bwTmn. Write(Gf) '写入二进制文件

x = x + 1

```
Next
Next
Next
bwTmn. Flush( )
bwTmn. Close( )
myFileTmn. Close( )
End Function
```

上述子程序将读取数据范围设置成 107° ~ 117°E、23° ~ 31°N ,网格点变为 101 × 81 = 8 181 个, 6 h 一次、38 个预报时次的文件(TXT 格式)本来总共占 33 M 空间,改写成二进制 dat 数据文件后缩小为 680 K,约为原来的 1/50,文件大幅度变小,访问效率提升。

这里需要注意两个问题:一是写 DAT 文件时的编码格式,“System. Text. Encoding. UTF32”必须是 UTF32 编码,其它编码格式用 GRADS 软件读取时为错误数据;二是写数据的顺序,Micaps 的数据顺序是从高纬度到低纬度,写数据时必须改成从低纬度到高纬度。

Micaps 的文本数据一般是一个文件保存二个维度数据,写成二进制文件时可以根据需要将一天的所有资料保存到一个文件里,那么就有 5 个维度(空间三维,时间一维,要素一维)的数据,如此处理之后,每天 100 G 的原始数据除去不常用的,全部写入一个 dat,大小变为 100 ~ 500 M,压缩比例为 200 ~ 1 000 倍,保存几年的数据毫无压力。

3 二进制数据文件的访问和使用

通过前面方法处理保存的数据文件的说明文件通常在同路径下可以找到,如果这些数据需要在

Micaps 里使用,用简单的少数几行代码恢复成一个文本格式文件,再加上相应的文件头即可。

如需要用 GRADS 等绘图软件查看数据,则另写数据描述文件。前文 EC 低温预报的数据描述文件如下:

```
dset d:\tfxt\dat\19091008Tmn6. dat
TITLE EC dw
UNDEF -9.99E+33
options little_endian
XDEF 80 LINEAR 107 0.125
YDEF 64 LINEAR 23 0.125
ZDEF 1 LEVELS 999
TDEF 38 LINEAR 08Z10Sep2019 6hr
vars 1
dw 0 99 dw
ENDVARS
```

4 备份程序的业务使用情况

株洲市气象台是湖南省的地区级台站,备份数据需要从湖南省气象局数据服务器远程拷贝下载,每天的数据量很大,如果是直接下载再保存于本地,受限于网络带宽和硬盘容量,基本无法实现,用前文所述方法加以处理后,如果只选取常规地面、高空资料,EC - Thin 资料等,一天的数据压缩后大约只有 500 M,这样一来,1TB 的硬盘可以轻松保存 5 a 以上资料。基于此思路,我们用 VB. NET 编写了可执行程序,该程序的主要功能是定时备份远程服务器上的 M4 后台数据,定时功能由 WIN 系统的计划任务实现,将该程序放入系统计划任务栏每日 02—03 时自行启动(如图 1),备份完毕后自动退出。



图 1 WIN7 系统计划任务中的 M4 资料备份程序

Fig. 1 M4 data backup program in win7 system planning task

具体需要备份哪些资料由设置文件设置(如图 2),该设置文件列举需要备份的资料及其目录地址,备份主程序启动后搜索该目录下一天之内的所有新文件并逐一写入 DAT 文件,图 2 所示的设置文件备份高空、地面和 EC 细网格资料,日储存数据量

大约为原数据量的 25%,占用空间 150 M,一般 1 h 内即可完成备份。通过最近 2 a 的实践运行,此程序稳定可靠,操作简单,设置好参数文件后无需人工干预即可每日定时完成资料备份工作。

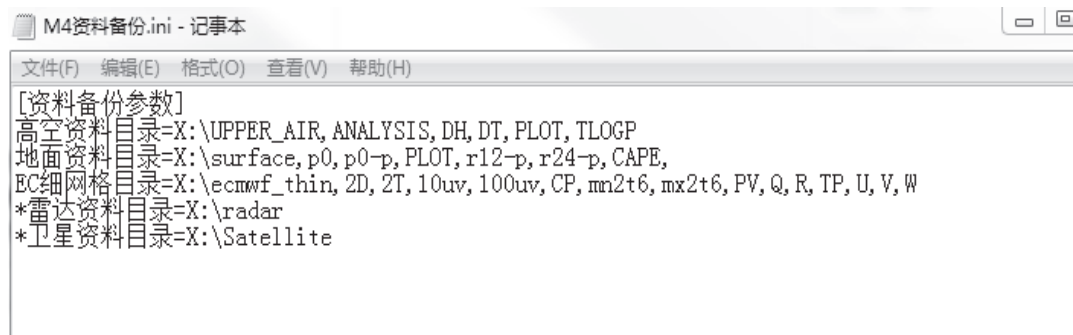


图 2 备份程序的设置文件
Fig. 2 The setup file of backup program

5 结语

资料备份工作非常重要,前文所述的思路和方法可以概括为:删除不常用数据,保留必需的核心数据,并用二进制编码进一步压缩数据文件容量,后期数据使用可借助 GRADS 软件或者恢复成 M4 文本文件。2 年的实践使用证明该方法稳定、实用且操作简便,保存和使用资料的成本降低,效率提高,可为

基层台站解决海量历史资料备份问题提供借鉴。

参考文献

[1] 气象信息综合分析处理系统第四版(MICAPS4.0)客户端使用指南[M]. 北京:气象出版社,2017-01-01.
[2] 气象大数据[M]. 上海:上海科学技术出版社,2018:20-25.
[3] VB.NET 程序设计[Computer Science and Technology][M]. 北京:清华大学出版社,2011.