



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 01111579.3

[45] 授权公告日 2003 年 11 月 26 日

[11] 授权公告号 CN 1129232C

[22] 申请日 2001.3.22 [21] 申请号 01111579.3
[71] 专利权人 谭伟祥
地址 518067 广东省深圳市南山区蛇口翠竹园 11-202 号
[72] 发明人 谭伟祥
审查员 韩 燕

[74] 专利代理机构 深圳市中知专利代理有限责任公司
代理人 王雄杰

权利要求书 3 页 说明书 7 页

[54] 发明名称 一种对数据库中的数据进行压缩与解压缩的方法

[57] 摘要

本发明提出一种对数据库中的数据进行压缩与解压缩的方法，它是针对数据库中的数值型字段在储存或传输之前，通过制作 $N(N > 10)$ 个不相同的字符串并与数值从 00 到 $N-1$ 按 ASC 码从小到大排列的字符建立一一对应关系，将数字变成字符，然后在读取或接收时进行反变换，还原成原来的数字。由于结构简单，可以直接编译到各个数据库的源程序中，经过本压缩方法压缩的数据库，不用解压缩就可以进行各种数据库操作。采用本发明能大大节省数据库中数据的存储空间、增大信息存储量。

I S S N 1 0 0 8 - 4 2 7 4

1、一种对数据库中的数据进行压缩与解压缩的方法，其特征在于：

压缩方法包括如下步骤：

- a、制作 N 个不相同的字符，其中 $N > 10$ ，
- b、将数值 00 到 $N-1$ 与 N 个按 ASC 码从小到大排列好的字符建立一一对应关系，
- c、若数据库中的数据为负数，设置标识符，若数据为正数，标识符为空，
- d、记录数据库中数据的小数位数 M ，
- e、把数据库中数据的绝对值乘以 10^M 成为整数化数据，
- f、将整数化数据表示成 $\sum X_i * N^i$ ，其中 $0 \leq X_i \leq N-1$ ， i 为零或整数，得到数值 X_i ，
- g、将步骤 f 中各数值 X_i 用步骤 b 中的对应字符代替，得到字符型数据 ΠX_i ，
- h、数据库中的数据记录由标识符、小数位数和字符型数据 ΠX_i 组成；

对上述数据库中由标识符、小数位数和字符型数据 ΠX_i 组成的数据记录的解压缩方法包括如下步骤：

- (1) 根据上述步骤 b 建立的 N 个字符与数值 00 到 $N-1$ 之间的一一对应关系，将字符型数据 ΠX_i 表示成 N 进制数值型数据 $\sum X_i * N^i$ ，其

中 X_I 用对应的数值代替,

(2) 由步骤 (1) 得出 10 进制数值型数据,

(3) 根据记录中的小数位数 M , 把 10 进制数值型数据打上小数点, 成为绝对值数据,

(4) 将标识符放在绝对值数据前恢复成压缩前的带符号数据。

2、根据权利要求 1 所述一种对数据库中的数据进行压缩与解压缩的方法, 其特征在于: $N=100$,

压缩方法包括如下步骤:

a、制作 100 个不相同的字符,

b、将数值从 00 到 99 与 100 个字符建立一一对应关系,

c、若数据库中的数据为负数, 设置标识符, 若数据为正数, 标识符为空,

d、记录数据库中数据的小数位数 M ,

e、把数据库中数据的绝对值乘以 10^M 成为整数化数据,

f、把整数化数据用步骤 b 中的对应字符来表示, 成为字符型数据,

g、数据库中的数据记录由标识符、小数位数和字符型数据组成;

对上述数据库中的数据记录的解压缩方法包括如下步骤:

(1)、根据步骤 b 建立的 100 个字符与数值从 0 到 99 之间的一一对应关系, 将字符型数据表示成数值型数据,

(2)、根据记录中的小数位数 M , 把上述数值型数据打上小数点, 成为绝对值数据,

(3)、将标识符放在绝对值数据前恢复成压缩前的带符号数据。

一种对数据库中的数据进行压缩与解压缩的方法

本发明涉及数据的储存技术和通信传输技术，具体地说，是特指数据库中数值型字段的数据进行储存和传输的压缩与解压缩方法，可广泛用于各种数据库系统。

随着计算机科学的发展，计算机对数据库存储量、存储速度的要求越来越高，特别是在科学计算中，有大量的数值型数据需要进行储存和传输。如何针对数据库中的这种特殊数值型字段的数据进行储存和传输的压缩与解压缩方法，以期提高数据库的存储容量、节省备份空间，提高传输效率，人们一直未能提出有效的解决办法，也未在公开出版物上见到有关这方面的专题报道。

本发明的目的意在克服上述现有技术的不足，提出一种能大大节省数据库中数据的存储空间、增大信息存储量的压缩与解压缩方法。

本发明的次一目的是通过这种压缩，提高数据的传输效率。

实现上述目的的技术方案：一种对数据库中的数据进行压缩与解压缩的方法，其特征在于：

压缩方法包括如下步骤：

- a、制作 N ($N > 10$) 个不相同的字符，
- b、将数值 00 到 $N-1$ 与 N 个按 ASC 码从小到大排列好的字符建立一一对应关系，

- c、若数据库中的数据为负数，设置标识符，若数据为正数，标识符为空，
- d、记录数据库中数据的小数位数 M ，
- e、把数据库中数据的绝对值乘以 10^M 成为整数化数据，
- f、将整数化数据表示成 $\sum X_i * N^i$ ($0 \leq X_i \leq N-1$, i 为零或整数)，得到数值 X_I ，
- g、将步骤 f 中各数值 X_i 用步骤 b 中的对应字符代替，得到字符型数据 ΠX_i (即：表示成 $X_{n-1} X_{n-2} \dots X_1 X_0$ 的形式)，
- h、数据库中的数据记录由标识符、小数位数和字符型数据 ΠX_i 组成；

对上述数据库中由标识符、小数位数和字符型数据 ΠX_i 组成的数据记录的解压缩方法包括如下步骤：

(1) 根据上述步骤 b 建立的 N 个字符与数值 00 到 $N-1$ 之间的一一对应关系，将字符型数据 ΠX_i 表示成 N 进制数值型数据 $\sum X_i * N^i$ ，其中 X_i 用对应的数值代替，

(2) 由步骤 (1) 得出 10 进制数值型数据，

(3) 根据记录中的小数位数 M ，把 10 进制数值型数据打上小数点，成为绝对值数据，

(4) 将标识符放在绝对值数据前面恢复成压缩前的带符号数据。

一种对数据库中的数据进行压缩与解压缩的方法，其特征在于：
 $N=100$ ，

压缩方法包括如下步骤：

- a、制作 100 个不相同的字符，
 - b、将数值从 00 到 99 与 100 个字符建立一一对应关系，
 - c、若数据库中的数据为负数，设置标识符，若数据为正数，标识符为空，
 - d、记录数据库中数据的小数位数 M ，
 - e、把数据库中数据的绝对值乘以 10^M 成为整数化数据，
 - f、把整数化数据用步骤 b 中的对应字符来表示，成为字符型数据，
 - g、数据库中的数据记录由标识符、小数位数和字符型数据组成；
- 对上述数据库中的数据记录的解压缩方法包括如下步骤：

(1)、根据步骤 b 建立的 100 个字符与数值从 0 到 99 之间的一一对应关系，将字符型数据表示成数值型数据，

(2)、根据记录中的小数位数 M ，把上述数值型数据打上小数点，成为绝对值数据，

(3)、将标识符放在绝对值数据前面恢复成压缩前的带符号数据。

采用上述技术方案，本发明突出的技术进步在于：1、它的物理意义是将数据字符化，在字符长度相同的情况下，可以表达更多的信息量。如：在一个字符的长度上，由十进制数据的 10 种变化，根据需要可扩展成 100 个变化甚至于 1000 个变化，信息量大大增加。2、通过对数据库中数值型字段数据的压缩和解压缩，能够大大节省存储

空间、节省备份的空间和时间、节省数据的传输时间，适合于各种数据库系统。例如：按照上述方法制作 100 个不相同的字符，对 3 位、4 位整数数字进行压缩，可以使字符数减少大约一半；对于带有小数点的长度为 5 的数 XX.XX 或 XXX.X，正常情况下要 5 位，经过变换后只需 2 位，字符数只有原来的 40%；而对于 0.XX 的情况，变换后只有一位，压缩到原来的 25%。将其用于数据库的存储方面，大大节省了存储空间和备份的空间。将其用于 FM 调频广播的股票数据传输方面，由于 FM 调频广播单位时间所传输的信息量有限，成为提高传输速度的瓶颈，数字经过压缩后再进行 FM 传输，由于有大量带两位小数点的 5 位数，有希望成倍以上地提高其传输效率，从而大大加强与视频传输的竞争力。3、按照本发明方法编写的程序简单（在无负数或小数位数一定的情况下，程序还可以更简单），所以可以方便地为其它程序和函数随时调用，也可以直接出现在计算式中，可以直接编译到各个数据库的源程序中，使其成为数据库中数值型数据的一种新的记录方式。因此，经过本压缩方法压缩的数据库，不用解压缩就可以进行各种数据库操作。

本发明将惠及各个数据库系统，有着广泛的用途。

下面通过实施例对本发明作进一步详细的说明：

实施例：一种对数据库中的数据进行压缩与解压缩的方法，包括压缩方法和解压缩方法。

压缩方法包括如下步骤：

a、首先，制作字符串 N ($N > 10$)，字符串中的每一个字符都不相

同，字符串中的所有字符按 ASC 码从小到大排列好（这样就可以在索引 INDEX 或排序 SORT 等操作中直接对字段名进行而不用变换）。

字符串一旦选定，就变成了存取数字的密码；

b、将数值 00 到 N-1 与 N 个按 ASC 码从小到大排列好的字符建立一一对应关系；

c、判断数据库中的十进制数据正负数，若为负数，设置标识符“-”，若为正数，标识符为空字符串；

d、记录上述数据库中数据的小数位数 M；

e、把数据库中数据的绝对值乘以 10^M 成为整数化数据；

f、将整数化数据表示成 $\sum X_i * N^i$ ($0 \leq X_i \leq N-1$, i 为零或整数)，得到数值 X_i ；

g、将步骤 f 中各数值 X_i 用步骤 b 中的对应字符代替，得到字符型数据 ΠX_i ；

h、数据库中的数据记录由标识符、小数位数和字符型数据 ΠX_i 组成。

对上述数据库中由标识符、小数位数和字符型数据 ΠX_i 组成的数据记录的解压缩方法包括如下步骤：

(1) 根据上述步骤 b 建立的 N 个字符与数值 00 到 N-1 之间的一一对应关系，将字符型数据 ΠX_i 表示成 N 进制数值型数据 $\sum X_i * N^i$ ，其中 X_i 用对应的数值代替；

(2) 由步骤 (1) 得出 10 进制数值型数据；

(3) 根据记录中的小数位数 M，把 10 进制数值型数据打上小数

点，成为绝对值数据；

(4) 将标识符放在绝对值数据前面恢复成压缩前的带符号数据。

作为特例，取 $N=100$ ，对数据库中的数据“-780036.26”进行压缩的方法包括如下步骤：

- a、首先制作 100 个不相同的字符；
- b、将数值从 00 到 99 与 100 个字符建立一一对应关系；
- c、将标识符设置为“-”；
- d、取 $M=2$ ；
- e、将数据 780036.26 乘以 10^2 成为整数化数据 78003626；

f、根据数值从 00 到 99 与 100 个字符建立的一一对应关系，假定其中 00 对应 A、26 对应 Z、36 对应 α 、78 对应 ϕ ，由于 $78003626=78*100^3+00*100^2+36*100^1+26*100^0$ ，因此，整数化数据 78003626 用对应字符来表示时的字符型数据为 $\phi A \alpha Z$ ，压缩后的数据记录由标识符为“-”、小数位数 $M=2$ 的字符型数据 $\phi A \alpha Z$ 组成。

对上述 $N=100$ 时数据库记录中标识符为“-”、小数位数 $M=2$ 的字符型数据 $\phi A \alpha Z$ 的解压缩方法，包括如下步骤：

(1) 根据预先建立的 N 个字符与数值从 0 到 $N-1$ 之间的一一对应关系，将字符型数据 $\phi A \alpha Z$ 表示成数值型数据 78003626；

(2) 根据记录中的小数位数 $M=2$ ，把 10 进制数值型数据打上小数点，成为绝对值数据 780036.26；

d、将标识符“-”放在绝对值数据 780036.26 前面恢复成压缩前的带符号数据-780036.26。

采用上述压缩与解压缩方法，若用于数据库存储器中，其物理意义在于：若不进行压缩，需要 10 个字节才能存储数据“-780036.26”，经本发明压缩后只需要 6 个字节就能存储（包括 4 个字节的字符型数据 $\Phi A \alpha Z$ 、各 1 个字节的标识符“-”和小数位数 2），节省存储空间 40%（压缩比为 50%）。若为无符号整数，则节省存储空间 50%！

不失一般性，根据商业需要，字符数 N 的设置也可以是 100 以上、甚至 1000，用于进一步增加提高压缩比；字符数 N 的设置也可以是 100 以下，此时，压缩比有不同程度的降低。不过，这两种做法虽然能增强保密性，然而其数据缺乏流通性。