



[12] 发明专利说明书

专利号 ZL 200610020462.5

[45] 授权公告日 2007 年 12 月 5 日

[11] 授权公告号 CN 100353331C

[22] 申请日 2006.3.10

[21] 申请号 200610020462.5

[73] 专利权人 四川大学

地址 610065 四川省成都市一环路南一段
24 号

[72] 发明人 李 涛

[56] 参考文献

US6332177B1 2001.12.18

US6081875A 2000.6.27

US6131148A 2000.10.10

CN1352427A 2002.6.5

审查员 赵小宁

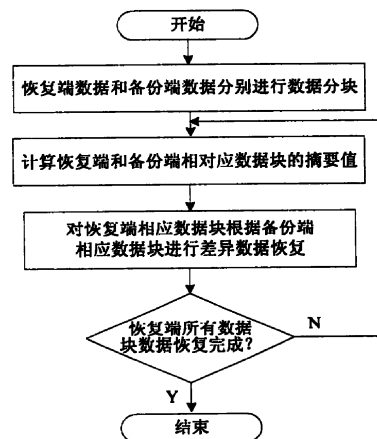
权利要求书 2 页 说明书 5 页 附图 3 页

[54] 发明名称

网络信息系统数据异地快速恢复方法

[57] 摘要

本发明提出了一种网络信息系统数据异地快速恢复方法，属于数据备份与灾难恢复领域。本发明通过对备份端数据与恢复端数据进行差异比较，根据备份端数据对恢复端数据进行差异恢复，其中恢复端数据和备份端数据在物理上的位置不受限，可分处于异地，恢复端数据和备份端数据可以是文件数据、数据库数据、磁盘数据等等。本发明具有数据恢复速度快、对网络的带宽要求低、对应用层透明等特点。本发明可广泛应用于基于 Internet 的异地数据备份与灾难恢复，建设成本低，具有广阔的应用前景。



1. 一种网络信息系统数据异地快速恢复方法，其特征在于包括以下步骤：数据分块的步骤；数据块摘要值计算的步骤；差异数据恢复的步骤；其中
- (1) 数据分块的步骤：采用同样的分块方法对恢复端和备份端进行分块，得到恢复端和备份端各自的数据块集合，恢复端数据块集合中的数据块与备份端数据块集合中的数据块一一对应，具体的数据分块的步骤包括以下步骤：
- 1) 定义数据单元的大小，将数据划分为定长大小的数据单元的步骤；
 - 2) 定义数据块的大小，即定义每个数据块包含的数据单元数的步骤；
 - 3) 计算数据块的数目，即根据数据单元的大小和数据块的大小计算得到数据块数目的步骤；
- (2) 数据块摘要值计算的步骤：对数据分块步骤得到的数据块进行摘要值计算，具体的数据块摘要值计算的步骤包括以下步骤：
- 1) 数据块摘要值计算初始化的步骤；
 - 2) 从数据块的第一个数据单元开始，直到数据块的最后一个数据单元，依次对数据块的每一个数据单元进行以下步骤：
 - ① 计算当前数据单元的摘要值的步骤；
 - ② 累积计算当前数据块摘要值的步骤；
- (3) 差异数据恢复的步骤：根据数据块摘要值的计算结果进行差异数据恢复，具体的差异数据恢复的步骤包括以下步骤：
- 1) 差异数据恢复初始化的步骤；
 - 2) 从恢复端和备份端的第一个数据块开始，直到恢复端和备份端的最后一个数据块，循环进行以下步骤：
 - ① 计算恢复端数据数据块的摘要值的步骤；
 - ② 计算备份端数据相应于所述恢复端数据块的摘要值的步骤；

③ 数据块数据恢复的步骤，包括：

- a) 比较恢复端和备份端数据块的差异的步骤：将恢复端数据块的摘要值与备份端相应于所述恢复端数据块的摘要值进行比较，如果两者相等，则结束本轮循环，否则转 b)；
- b) 获取备份端数据的步骤：获取备份端当前数据块的数据；
- c) 数据恢复的步骤：用备份端当前数据块的数据覆盖恢复端当前数据块的数据。

网络信息系统数据异地快速恢复方法

一、技术领域

本发明提出了一种网络信息系统数据异地快速恢复方法，属于数据备份与灾难恢复领域。

二、背景技术

在网络信息系统中，当灾难发生后，能否快速地从异地备份中心恢复数据、减少服务中断时间是保证商业连续的关键。目前灾难恢复的方法有采用 `rsync` 进行数据恢复，利用数据库自己的恢复工具，利用备份端的数据进行完全数据拷贝等方案。这些技术各有其优缺点：`rsync` 在恢复速度和带宽的节约上有一定的优势，但是由于它是基于文件级操作的，针对像数据库这样的大文件无能为力，且其恢复不能满足实时性的要求，只能恢复最近一次备份端的数据。基于数据库的恢复方案的实现与具体的数据库的实现相关，且一般由数据库的厂商提供，通常是非常的昂贵。根据备份端的数据进行完全数据拷贝方案存在的问题是拷贝的数据量大、网络的负载大、恢复的速度低下。

本发明提出了一种通用的网络信息系统数据异地快速恢复方法，具有如下的特点：

- 1) 数据的恢复速度快、对网络的带宽要求低；
- 2) 该方法可基于廉价的 *Internet*，运行成本降低；
- 3) 数据的恢复基于数据块，对应用层透明。

本发明可支持用户在 *Internet* 上以极小的成本构建自己的数据恢复系统，具有广阔的应用前景。

三、发明内容

本发明提出了一种网络信息系统数据异地快速恢复方法，该方法利用备份端数据对恢复端数据进行数据恢复，使恢复端数据和备份端数据处于一致的状态，其中恢复端和备份端在物理上的位置不受限，可分处于异地，恢复端数据和备份端数据可以是文件数据、数据库数据、磁盘数据等。

该方法首先对恢复端数据和备份端数据进行同样的数据分块，分块后恢复端数据和备份端数据分别得到各自的数据块集。在对恢复端数据和备份端数据进行数据分块时，均采用同样的分块方法，因此，恢复端数据分块后得到的数据块和备份端数据分块后得到的数据块一一对应，恢复端数据的各个数据块在备份端数据中有唯一数据块与之对应。其次，对恢复端数据的各数据块和备份端数据的各数据块分别求摘要值；然后，对恢复端数据的各个数据块的摘要值分别和备份端数据相对应的数据块的摘要值进行比较，若恢复端数据的数据块的摘要值和备份端数据与之对应的数据块的摘要值相同，则认为该数据块的数据在恢复端和备份端是一致的，该块数据无需恢复，否则，则认为该数据块的数据在恢复端和备份端不一致，利用备份端的该块数据覆盖恢复端相应的数据块，完成该块数据的恢复工作。对恢复端与备份端的所有数据块均采用同样的差异恢复方法，以完成整个数据的恢复工作。一般地，在实际应用中，当系统发生故障时，系统当前数据与备份数据其间的差异较小，需要恢复的内容不多，因此，本发明能够以高效率的方式恢复系统的数据。

四、附图说明

图1是本发明工作原理图。

图2是数据分块的步骤。

图3是数据块摘要值计算的步骤。

图4是差异数据恢复的步骤。

五、具体实施方式

以下结合附图详细说明本发明的具体方法。

图 1 是本发明工作原理图。

图1本发明的工作原理图，其中恢复端数据和备份端数据在物理上的位置不受限，可分处于异地。由图可知数据恢复主要由数据分块、数据块摘要值计算和差异数据恢复等三个步骤的结合来实现。其中数据分块步骤对恢复端数据和备份端数据进行数据分块，建立各自的数据块集；数据块摘要值计算步骤对恢复端数据和备份端数据的各数据块进行摘要值计算；差异数据恢复步骤对恢复端数据块的摘要值和备份端相应数据块的摘要值进行比较，实现对数据块的差异数据恢复。

具体地，本发明提出的网络信息系统数据异地快速恢复方法的步骤包括以下步骤：

- 1) 数据分块的步骤；
- 2) 数据块摘要值计算的步骤；
- 3) 差异数据恢复的步骤。

图 2 是数据分块的步骤。

图 2 给出了对一个数据进行数据分块的示例。图中数据是由数据单元组成，各数据单元大小相同，若数据的最后一个数据单元的大小小于其它数据单元的大小，则用空白补齐最后一个数据单元使它的大小和其它数据单元的大小一致。数据块由逻辑上连续的一个或多个数据单元组成，各数据块的大小可以不同。这样，数据可以看成是由数据单元组成，也可以看成是由数据块组成。

数据分块步骤如下：

- 1) 恢复端数据分块的步骤：对恢复端数据进行分块，具体步骤如下：
 - ① 定义数据单元的大小：将数据划分为定长大小的数据单元，若数据的最后一个数据单元的大小小于定义数据单元的大小，则用空白补齐最后一个数据单元使它的大小为定义数据单元的大小；

- ② 定义数据块的大小：规定数据块的大小，即包含的数据单元数，各数据块的大小可以不同。
- ③ 计算数据块的数目：计算构成整个数据的数据块的数目。
- 2) 备份端数据分块的步骤：对恢复端数据进行分块，具体步骤如下：
- ① 定义数据单元的大小：将数据划分为定长大小的数据单元，若数据的最后一个数据单元的大小小于定义数据单元的大小，则用空白补齐最后一个数据单元使它的大小为定义数据单元的大小；
- ② 定义数据块的大小：规定数据块的大小，即包含的数据单元数，各数据块的大小可以不同。
- ③ 计算数据块的数目：计算构成整个数据的数据块的数目。

通过以上的步骤，对恢复端数据和备份端数据进行一样的数据分块，即恢复端数据和备份端数据在分块时数据单元的大小、数据块的数目、各数据块的大小都相同，分块后得到恢复端数据和备份端数据各自的数据块集合。

图 3 数据块摘要值计算的步骤。

图 3 给出了数据块摘要值计算的示例。图中 H 为摘要计算函数，如 $MD5$ 、 $SHA-1$ 等，图示的数据块由 m 个数据单元组成，每个数据单元有一个摘要值，图中通过累积计算各个数据单元的摘要值得出数据块的摘要值，图中 h 为图示数据块的摘要值。

数据块摘要值计算的步骤如下：

- 1) 数据块摘要计算初始化：初始化有关变量，令数据块的摘要值 $h = \text{空}$ ，数据单元的编号 $i = 1$ ；
- 2) 计算当前数据单元的摘要值 h' ： $h' = H(d_i)$ ，其中 H 为摘要函数，如 $MD5$ 、 $SHA-1$ 等， d_i 是当前数据块的第 i 块数据单元；
- 3) 累积计算当前数据块摘要值： $h = H(h \odot h')$ ，其中 \odot 表示字符串连接运算； $i = i + 1$ ；若 i 不大于当前数据块的大小，则转 2)，否则 h 即为当前数据块的摘要值。

图 4 是差异数据恢复的步骤。

图 4 给出了差异数据恢复的步骤，如图所示，首先进行差异数据恢复初始化工作；其次分别计算恢复端和备份端数据相对应数据块的摘要值；然后比较数据块的摘要值，若数据块的摘要值相同，则认为恢复端数据的该数据块的数据和备份端数据的相应数据块的数据处于一致的状态，若数据块的摘要值不同，则认为恢复端数据的该数据块的数据和备份端数据的相应数据块的数据处于不一致的状态，此时利用备份端的该块数据覆盖恢复端相应的数据块，完成该块数据的恢复工作；对恢复端与备份端的所有数据块均采用同样的差异恢复方法，以完成恢复端整个数据的恢复工作。

具体地，差异数据恢复的步骤如下：

- 1) 差异数据恢复初始化：初始化有关变量，令数据块的编号 $i=1$ ；
- 2) 计算恢复端数据相应数据块的摘要值：调用数据块摘要值计算的步骤计算恢复端数据的第 i 块数据块的摘要值 h ；
- 3) 计算备份端数据相应数据块的摘要值：调用数据块摘要值计算的步骤计算备份端数据的第 i 块数据块的摘要值 h' ；
- 4) 数据块数据恢复：通过比较两端摘要值的差异，采取差异恢复方法进行数据块恢复，具体步骤如下：
 - ① 比较两端数据块的差异：若 $h = h'$ ，则两端数据一致，该块数据不需要恢复，否则，转②；
 - ② 获取备份端数据：在备份端提取提取第 i 块数据 d_i ；
 - ③ 数据恢复：利用 d_i 的数据覆盖恢复端第 i 块数据；
- 5) 循环恢复下一数据块： $i = i + 1$ ；若 i 不大于数据块的总数，则转 2)，否则结束数据恢复工作。

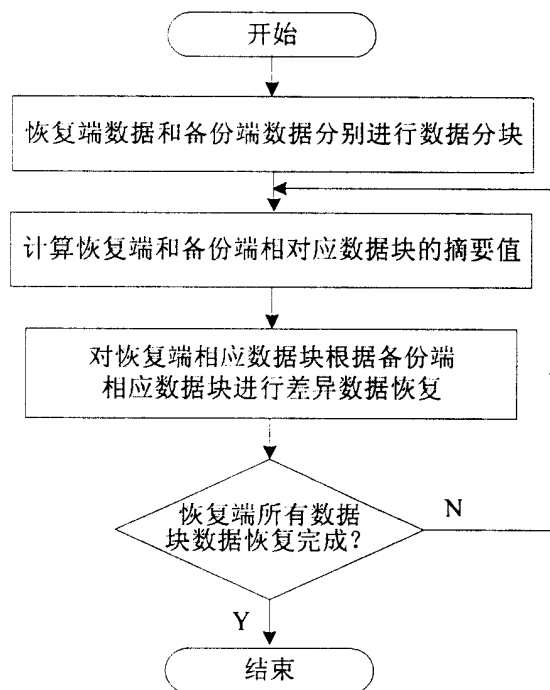


图 1

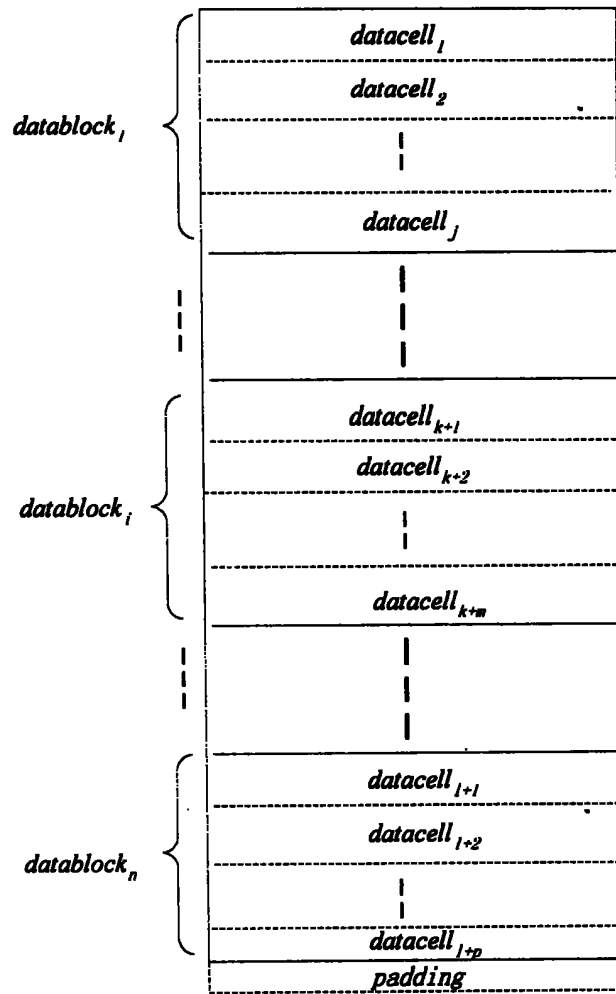


图 2

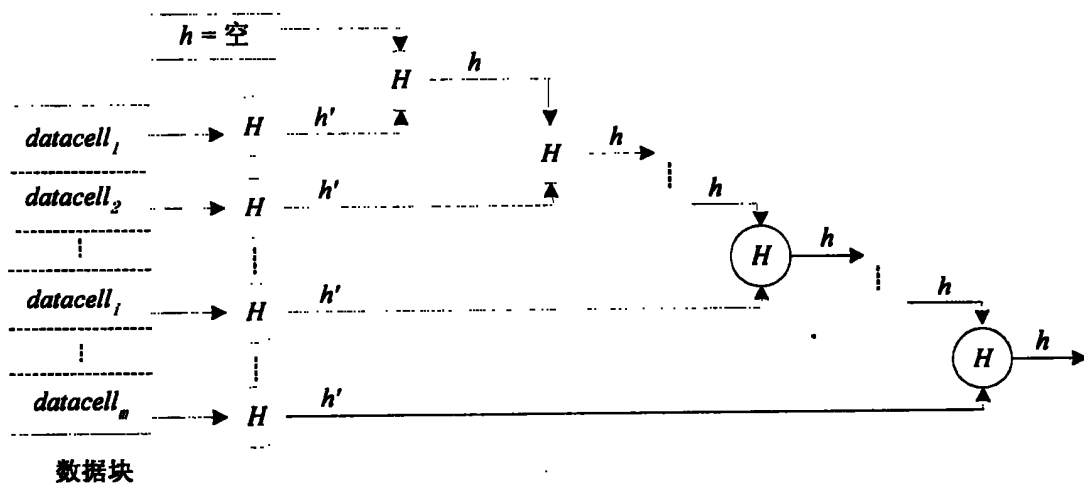


图 3

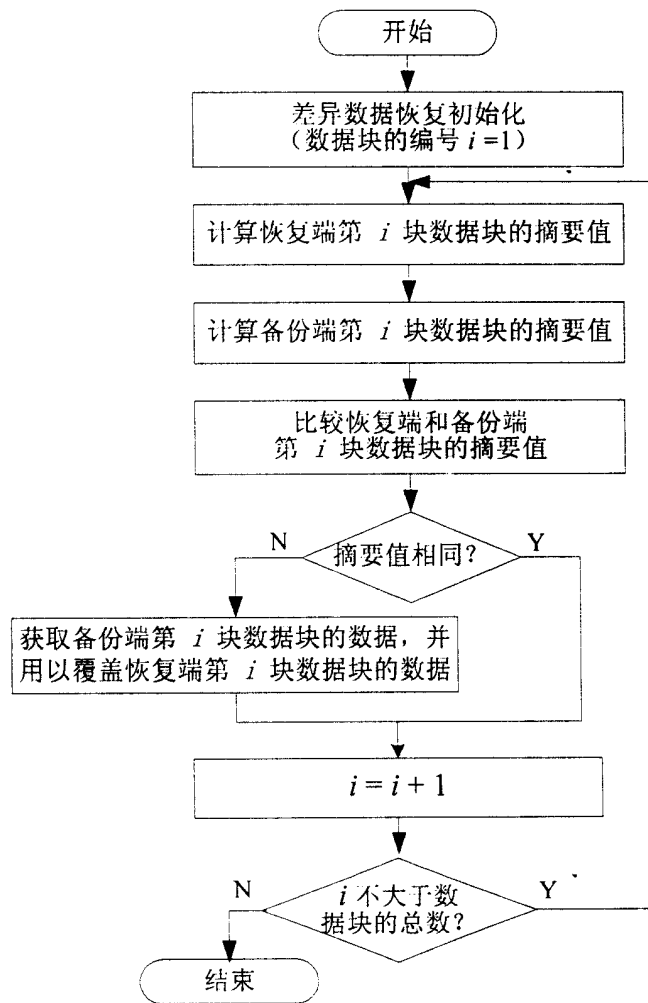


图 4