

(12) 发明专利申请

(10) 申请公布号 CN 103176867 A

(43) 申请公布日 2013.06.26

(21) 申请号 201310112143.7

(22) 申请日 2013.04.02

(71) 申请人 浪潮电子信息产业股份有限公司
地址 250014 山东省济南市高新区舜雅路
1036 号

(72) 发明人 穆裕

(51) Int. Cl.
G06F 11/14 (2006.01)

权利要求书1页 说明书3页 附图2页

(54) 发明名称

一种快速文件差异备份方法

(57) 摘要

本发明提供一种快速文件差异备份方法,该方法是使用备份集 ID 管理每一次备份的数据,通过备份集 ID 定位唯一的数据集,通过 catalog 文件存储每一次备份的文件相关信息,将全量备份集 ID 存储在 catalog 文件前 16 个字节中,这样通过任意 catalog 文件即可定位全量备份的 catalog,从而获取最近一次全量备份时文件的状态,经过与现在文件状态的对比即可知道哪些文件需要备份,该方法通常用于在备份大数据量文件时减少数据传输量,节省数据存储空间。

全备份集ID	
文件1	catalog大小
文件1	类型(文件或是目录)
文件1	文件大小
文件1	文件属性(只读、隐藏)
文件1	文件创建时间
文件1	文件最后访问时间
文件1	文件最后写入时间
文件1	此次文件的备份集ID
文件1	文件路径信息
文件2	catalog大小
.....	

1. 一种快速文件差异备份方法，其特征在于该方法是使用备份集 ID 管理每一次备份的数据，通过备份集 ID 定位唯一的数据集，通过 catalog 文件存储每一次备份的文件相关信息，将全量备份集 ID 存储在 catalog 文件前 16 个字节中，这样通过任意 catalog 文件即可定位全量备份的 catalog，从而获取最近一次全量备份时文件的状态，经过与现在文件状态的对比即可知道哪些文件需要备份，该系统包括：catalog 获取模块(1)、文件比对备份模块(2)、catalog 组织模块(3)，其中：

catalog 获取模块(1)：用于连接存储备份集数据的介质服务器，获取最近一次全量备份的 catalog 文件，首先 catalog 获取模块(1)会向介质服务器获取上一次备份的 catalog 文件，获取成功后读取文件前 16 个字节，获取到全量备份集 ID 信息，然后再次向介质服务器提供全量备份集 ID 以获取全量备份的 catalog 文件；

文件比对备份模块(2)：用于通过文件最后写入时间和文件大小的比对，确定现阶段的文件状态是否与全量备份时的文件状态相同，如果相同则表明该文件不需要进行备份操作，置该文件的备份集 ID 为全量备份集 ID，如果不同则置该文件的备份集 ID 为当前的备份集 ID，最后通过比对每一个文件的备份集 ID 是否等于当前备份集 ID 来确定该文件是否需要备份，即对备份集 ID 为当前 ID 的文件进行备份；

catalog 组织模块(3)：用于组织 catalog 文件，对于每一个文件会有一个结构体去存储该文件的相关信息，包括文件类型包括：文件或是目录、文件大小、文件属性包括：只读，隐藏、文件创建时间、文件最后访问时间、文件最后写入时间、此次文件的备份集 ID、文件路径信息，，不论文件数据是否备份该文件的 catalog 信息都将会写入到 catalog 文件中，所有文件备份完成后将 catalog 获取模块(1)中获取到的全量备份集 ID 信息写入到文件的前 16 个字节中，最后将 catalog 文件备份到介质服务器完成差异备份。

2. 根据权利要求 1 中所述的快速文件差异备份方法，其特征在于通过全量备份集 ID 定位全量备份的 catalog，是由数据备份目的端进行处理。

3. 根据权利要求 1 中所述的快速文件差异备份方法，其特征在于确定文件是否变化，是通过文件的最后写入时间和文件大小来确认。

一种快速文件差异备份方法

技术领域

[0001] 本发明涉及计算机应用领域，具体地说是一种快速文件差异备份方法。

背景技术

[0002] 据 IDC 统计表明，美国在 2000 年前 10 年间发生过数据灾难的公司中，有 55% 当时倒闭，剩下的 45% 中，因为数据丢失，又有 29% 也在 2 年之内倒闭，生存下来的仅占 16%。而拥有有效的备份方法及备份策略的公司则能够将损失降到最低。因此，数据的备份被提上日程。一般传统的本地备份方法如磁带备份等，能抵制硬件故障、计算机病毒等造成的数据损坏，但对大规模的数据灾难却无能为力。而传统的远程备份方法为了完成海量数据备份，需要巨大的代价建造并维护一条独立备份线路。为解决以上问题，本发明采用备份集 ID 标识备份数据，采用文件数据与文件信息分离备份的方式对文件进行快速的差异备份。

发明内容

[0003] 本发明的目的是提供一种快速文件差异备份方法。

[0004] 本发明的目的是按以下方式实现的，该方法是使用备份集 ID 管理每一次备份的数据，通过备份集 ID 定位唯一的数据集，通过 catalog 文件存储每一次备份的文件相关信息，将全量备份集 ID 存储在 catalog 文件前 16 个字节中，这样通过任意 catalog 文件即可定位全量备份的 catalog，从而获取最近一次全量备份时文件的状态，经过与现在文件状态的对比即可知道哪些文件需要备份，该系统包括：catalog 获取模块(1)、文件比对备份模块(2)、catalog 组织模块(3)，其中：

catalog 获取模块(1)：用于连接存储备份集数据的介质服务器，获取最近一次全量备份的 catalog 文件，首先 catalog 获取模块(1)会向介质服务器获取上一次备份的 catalog 文件，获取成功后读取文件前 16 个字节，获取到全量备份集 ID 信息，然后再次向介质服务器提供全量备份集 ID 以获取全量备份的 catalog 文件；

文件比对备份模块(2)：用于通过文件最后写入时间和文件大小的比对，确定现阶段的文件状态是否与全量备份时的文件状态相同，如果相同则表明该文件不需要进行备份操作，置该文件的备份集 ID 为全量备份集 ID，如果不同则置该文件的备份集 ID 为当前的备份集 ID，最后通过比对每一个文件的备份集 ID 是否等于当前备份集 ID 来确定该文件是否需要备份，即对备份集 ID 为当前 ID 的文件进行备份；

catalog 组织模块(3)：用于组织 catalog 文件，对于每一个文件会有一个结构体去存储该文件的相关信息，包括文件类型包括：文件或是目录、文件大小、文件属性包括：只读、隐藏、文件创建时间、文件最后访问时间、文件最后写入时间、此次文件的备份集 ID、文件路径信息，，不论文件数据是否备份该文件的 catalog 信息都将会写入到 catalog 文件中，所有文件备份完成后将 catalog 获取模块(1)中获取到的全量备份集 ID 信息写入到文件的前 16 个字节中，最后将 catalog 文件备份到介质服务器完成差异备份。

[0005] 通过全量备份集 ID 定位全量备份的 catalog,是由数据备份目的端进行处理。

[0006] 确定文件是否变化,是通过文件的最后写入时间和文件大小来确认。

[0007] 本发明的有益效果是:用户首先通过 catalog 获取模块(1)获取全量备份集的 catalog 文件,在通过文件比对备份模块(2)确定并且备份文件数据,最后通过 catalog 组织模块(3),将每一个文件的相关信息写入到本次备份的 catalog 文件中,将全量备份集 ID 写入到 catalog 文件前 16 个字节。该方法通常用于在备份大数据量文件时减少数据传输量,节省数据存储空间。

附图说明

[0008] 图 1 是 catalog 文件结构示意图;

图 2 是快速文件差异备份流程示意图。

具体实施方式

[0009] 参照说明书附图对本发明的方法作以下详细地说明。

[0010] 本发明的体系结构主要包括:catalog 获取模块(1)、文件比对备份模块(2)、catalog 组织模块(3),其中:

catalog 获取模块(1)用于连接存储备份集数据的介质服务器,通过提供全量备份集 ID 信息,获取全量备份集 catalog;文件比对备份模块(2)是用于对文件状态信息进行对比,确定文件是否备份并执行备份过程;catalog 组织模块(3)是用于存储每个文件的相关信息,保存最后一次全量备份集 ID。

[0011] 一种快速文件差异备份方法,通常用于在备份大数据量文件时减少数据传输量,节省数据存储空间。用户首先通过 catalog 获取模块(1)获取全量备份集的 catalog 文件,在通过文件比对备份模块(2)确定并且备份文件数据,最后通过 catalog 组织模块(3),将每一个文件的相关信息写入到本次备份的 catalog 文件中,将全量备份集 ID 写入到 catalog 文件前 16 个字节。

[0012] 一种快速文件差异备份方法,通常用于在备份大数据量文件时减少数据传输量,节省数据存储空间。该方法是使用备份集 ID 管理每一次备份的数据,通过备份集 ID 可以定位唯一的数据集,通过 catalog 文件存储每一次备份的文件相关信息,将全量备份集 ID 存储在 catalog 文件前 16 个字节中,这样通过任意 catalog 文件即可定位全量备份的 catalog,从而获取最近一次全量备份时文件的状态,经过与现在文件状态的对比即可知道哪些文件需要备份。该系统体系包括:catalog 获取模块(1)、文件比对备份模块(2)、catalog 组织模块(3)。

[0013] catalog 获取模块(1)用于连接存储备份集数据的介质服务器,获取最近一次全量备份的 catalog 文件。首先 catalog 获取模块(1)会向介质服务器获取上一次备份的 catalog 文件,获取成功后读取文件前 16 个字节,获取到全量备份集 ID 信息,然后再次向介质服务器提供全量备份集 ID 以获取全量备份的 catalog 文件。

[0014] 文件比对备份模块(2)用于通过文件最后写入时间和文件大小的比对确定现阶段的文件状态是否与全量备份时的文件状态相同。如果相同则表明该文件不需要进行备份操作,置该文件的备份集 ID 为全量备份集 ID,如果不同则置该文件的备份集 ID 为当前的备份

集 ID。最后通过比对每一个文件的备份集 ID 是否等于当前备份集 ID 来确定该文件是否需要备份,即对备份集 ID 为当前 ID 的文件进行备份。

[0015] catalog 组织模块(3)用于组织 catalog 文件,对于每一个文件会有一个结构体去存储该文件的相关信息,包括类型(文件或是目录)、文件大小、文件属性(只读,隐藏)、文件创建时间、文件最后访问时间、文件最后写入时间、此次文件的备份集 ID、文件路径信息等,不论文件数据是否备份该文件的 catalog 信息都将会写入到 catalog 文件中,所有文件备份完成后将 catalog 获取模块(1)中获取到的全量备份集 ID 信息写入到文件的前 16 个字节中,最后将 catalog 文件备份到介质服务器完成差异备份。

[0016] 除说明书所述的技术特征外,均为本专业技术人员的已知技术。



图 1

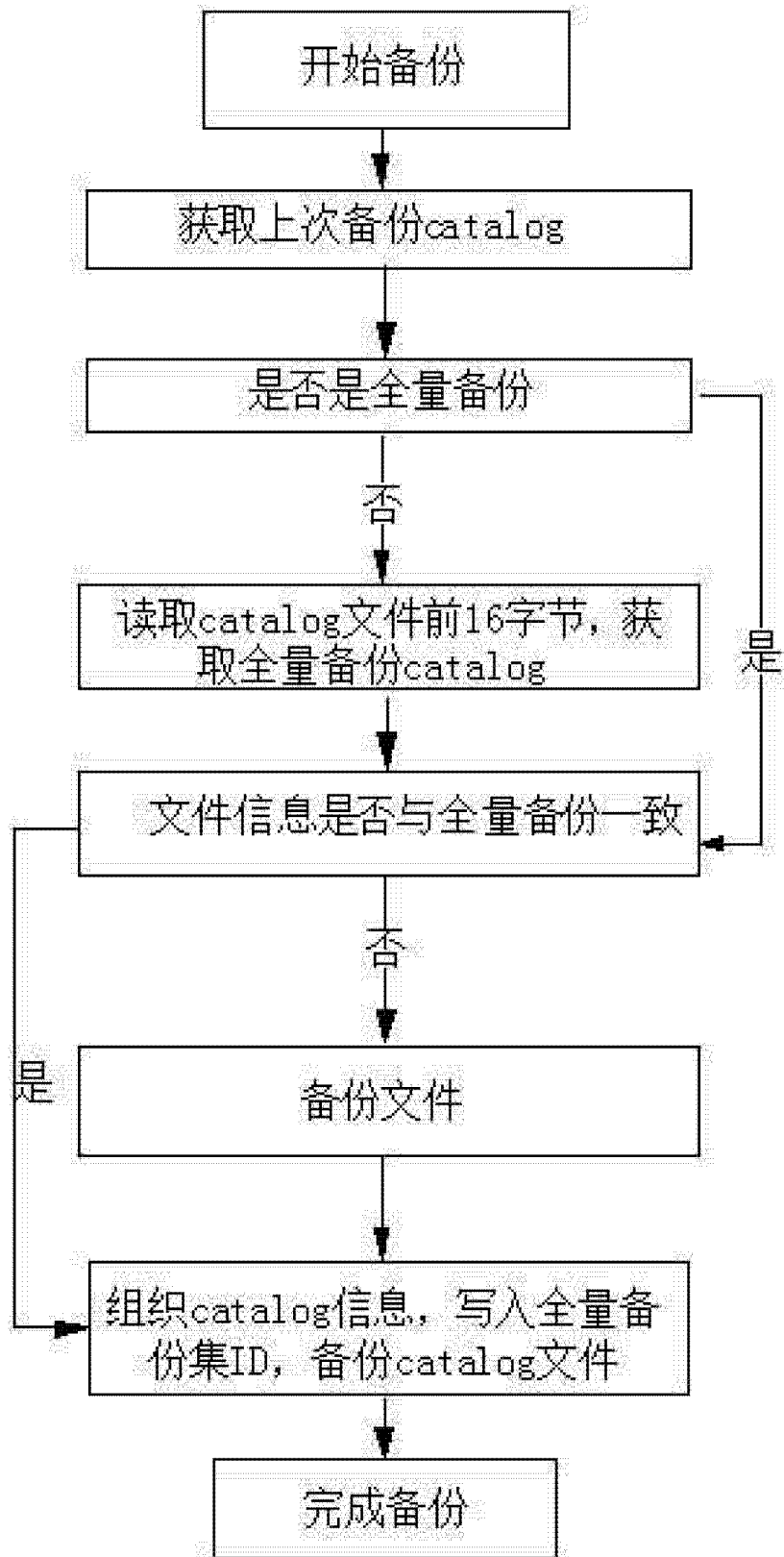


图 2