

(12) 发明专利申请

(10) 申请公布号 CN 102306114 A

(43) 申请公布日 2012. 01. 04

(21) 申请号 201010292125. 8

(22) 申请日 2010. 09. 25

(71) 申请人 广东电子工业研究院有限公司

地址 523808 广东省东莞市松山湖科技产业  
园区松科苑 10 号楼

(72) 发明人 肖翱 莫展鹏 杨松 岳强  
季统凯

(51) Int. Cl.

G06F 11/14 (2006. 01)

H04L 29/06 (2006. 01)

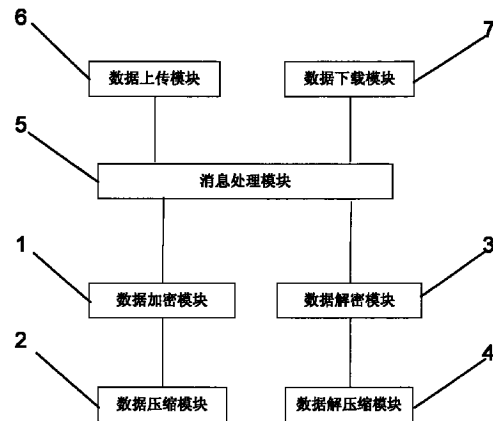
权利要求书 2 页 说明书 4 页 附图 3 页

(54) 发明名称

基于云存储的数据定时备份和恢复方法

(57) 摘要

本发明涉及云计算技术领域, 基于云存储的数据定时备份和恢复方法。包括数据压缩模块、数据解压缩模块、数据加密模块、数据解密模块、消息封装模块、数据上传模块和数据下载模块, 所述方法包括在备份数据时对数据进行加密、压缩、消息处理、上传、校验等操作; 数据恢复时对数据进行下载、消息处理、校验、解压缩、解密等操作。本发明可提升云存储系统数据备份和恢复的速度、效率以及数据传输的安全性; 可应用于升云存储系统的数据备份和恢复。



1. 一种基于云存储的数据定时备份和恢复方法,其特征在于:包括数据加密模块(1)、数据压缩模块(2)、数据解密模块(3)、数据解压缩模块(4)、消息处理模块(5)、数据上传模块(6)和数据下载模块(7);

由所述的数据加密模块(1)通过特定的算法对数据进行加密,此加密算法的描述作为元信息在消息处理模块(5)进行封装,加密之后的数据输出至数据压缩模块(2);

由所述的数据压缩模块(2)使用特定的压缩算法对数据加密模块(1)的输出进行处理,压缩之后的数据作为输入传送到消息处理模块(5),此压缩算法的描述作为元信息也传送到消息处理模块(5);

由所述的数据解密模块(3)从消息处理模块(5)中数据的元信息中获取加密算法的描述,根据此加密算法对数据进行解密,数据解密模块(3)的处理结果作为输入传输至数据解压缩模块(4);

由所述的数据解压缩模块(4)从消息处理模块(5)数据的元信息中获取对压缩算法的描述,根据此压缩算法对数据进行解压缩;

由所述的消息处理模块(5)在数据备份时将数据和元信息进行封装并序列化成语节流,处理结果作为输入传送到数据上传模块(6);在数据恢复时将从数据下载模块(7)接收到的字节流进行反序列化,还原成机器可读的数据及相关信息,处理结果作为输入传送到数据解密模块(3);

由所述的数据上传模块(6)应用预先设置的文件上传策略和多线程方式将消息处理模块(5)输出的字节流通过网络传输到云存储服务器;

由所述的数据下载模块(7)应用预先设置的文件下载策略和多线程方式接收网络传输过来的字节流,输出到消息处理模块(5)。

2. 根据权利要求1所述的数据定时备份和恢复方法,其特征在于:所述的数据上传模块(6)和数据下载模块(7)对应用文件上传策略和文件下载策略步骤之前还包括根据文件上传策略对本地数据和远程数据进行比较。

3. 根据权利要求2所述的数据定时备份和恢复方法,其特征在于:通过以下方式对本地数据和远程数据进行比较:

采用多个线程同时运行的方式分批从云存储服务器上获取数据元信息;

通过数据标识符的字典顺序与本地文件进行分批比较。

4. 根据权利要求2所述的数据定时备份和恢复方法,其特征在于:文件比较完毕后,将待上传的本地文件放到一个共享的数据上传队列中;上传数据的线程组不断轮询该队列,如果队列为空,上传线程组等待,否则上传线程组传输本地文件到云存储服务器。

5. 根据权利要求3所述的数据定时备份和恢复方法,其特征在于:文件比较完毕后,将待上传的本地文件放到一个共享的数据上传队列中;上传数据的线程组不断轮询该队列,如果队列为空,上传线程组等待,否则上传线程组传输本地文件到云存储服务器。

6. 根据权利要求2所述的数据定时备份和恢复方法,其特征在于:文件比较完毕后,将待下载的远程文件的元信息放到一个共享的数据下载队列中;下载数据的线程组不断轮询该队列,如果队列为空,下载线程组等待,否则下载线程组传输数据到本地文件系统。

7. 根据权利要求3所述的数据定时备份和恢复方法,其特征在于:文件比较完毕后,将待下载的远程文件的元信息放到一个共享的数据下载队列中;下载数据的线程组不断轮询

该队列,如果队列为空,下载线程组等待,否则下载线程组传输数据到本地文件系统。

8. 根据权利要求 1 至 7 任一项所述的数据定时备份和恢复方法,其特征在于:所述数据加密和数据解密还包括对加密算法的选择,如果加密算法没有特别指定,会设置默认的加密算法,同时还设置对应于此加密算法的一个自定义密码。

9. 根据权利要求 1 至 7 任一项所述的数据定时备份和恢复方法,其特征在于:进行数据压缩和数据解压缩的压缩算法是预先定义的。

10. 根据权利要求 1 至 7 任一项所述的数据定时备份和恢复方法,其特征在于:所述对数据进行压缩后再进行传输,同时进行数据分块传输。

11. 根据权利要求 10 所述的数据定时备份和恢复方法,其特征在于:所述的数据压缩使用的算法为标准的 ZLIB 压缩算法。

12. 根据权利要求 1 至 7 任一项所述的数据定时备份和恢复方法,其特征在于:对数据的内容进行加密,同时在上传到云存储服务器端或者下载到本地时都要校验数据的完整性。

13. 根据权利要求 12 所述的数据定时备份和恢复方法,其特征在于:所述的数据加密算法默认为 PBEWITHSHA1ANDDES 算法,但是也可选择其他自定义的加密算法。

## 基于云存储的数据定时备份和恢复方法

### 技术领域

[0001] 本发明涉及云计算技术领域,基于云存储的数据定时备份和恢复方法。

### 背景技术

[0002] 数据备份顾名思义,就是将数据以某种方式加以保留,以便在系统遭受破坏或其他特定情况下,重新加以利用的一个过程。数据恢复就是在系统发生灾难情况下,将数据还原到最近的一个状态,以保障系统的正常运行。

[0003] 传统的数据备份以及数据恢复包括以下几种策略:

[0004] 完全备份:每次都对所有数据都进行备份。例如,星期一用一盘磁带对所有数据进行备份,星期二再用另一盘磁带对所有数据在此进行备份,依此类推。这种备份策略的好处是:当发生数据丢失灾难时,只要用一盘磁带(即灾难发生前一天的备份磁带),就可以恢复丢失的数据。然而它亦有不足之处,首先,由于需每天都对所有数据都进行备份,造成备份的数据大量重复。这些重复的数据占用了大量的磁带空间,这对用户来说就意味着增加成本。其次,由于需要备份的数据量较大,因此备份所需的时间也就较长。

[0005] 增量备份:首次备份所有数据,以后每次都只备份当次新增的数据。比如星期天进行一次完全备份,然后在接下来的六天里只对当天的新数据或被修改过的数据进行备份。这种备份策略的优点是节省磁带空间,缩短了备份时间。但缺点在于,当灾难发生时,数据的恢复比较麻烦。例如,系统在星期三的早晨发生故障,丢失了大量的数据,那么现在就要将系统恢复到星期二晚上时的状态。这时系统管理员就要首先找出星期一的那盘完全备份磁带进行系统恢复,然后找出星期二的磁带来恢复星期二的的数据。很明显,这种方式很繁琐。而且可靠性也很差。在这种备份方式下,各盘磁带间的关系就象链子一样,一环套一环,其中任何一盘磁带出了问题都会导致整条链子脱节。比如在上例中,若星期二的磁带出了故障,那么管理员最多只能将系统恢复到星期一晚上时的状态。

[0006] 差异备份:首次备份所有的数据,以后每次备份时都备份与首次相比更新的数据。比如星期天进行一次完全备份,那么在接下来的几天里,只需再将当天所有与星期天不同的数据(新的或修改过的)备份到磁带上。差分备份策略在避免了以上两种策略的缺陷的同时,又具有了它们的所有优点。首先,它无需每天都对系统做完全备份,因此备份所需时间短,并节省了磁带空间,其次,它的灾难恢复也很方便。系统管理员只需两盘磁带,即星期天磁带与灾难发生前一天的磁带,就可以将系统恢复。

[0007] 传统的数据备份和数据恢复依赖的设备也比较昂贵,硬件上一般采用磁盘阵列、光盘塔、光盘库、磁带机以及磁带库等;软件上也许依赖于专业的数据备份和数据恢复软件。

[0008] 云存储是在云计算(Cloud Computing)概念上延伸和发展出来的一个新的概念,它是指通过集群应用、网络技术或分布式文件系统等功能,将网络中大量各种不同类型的存储设备通过应用软件集合起来协同工作,共同对外提供数据存储和业务访问功能的一个系统。

[0009] 云存储可以让用户很容易增加存储容量,而且不需要购买、安装和管理任何存储基础设施,其低成本和易用性等优势对各种规模的企业都很有吸引力,然而基于云存储对数据进行异地的定时备份和恢复的机制却还没有很好的解决方案。

## 发明内容

[0010] 本发明解决的技术问题在于提供一种基于云存储的数据定时备份和恢复方法;可充分利用云存储的即时即地可存储的特征,更有效的保障数据的完备性。

[0011] 本发明解决上述技术问题的技术方案是:

[0012] 包括数据加密模块、数据压缩模块、数据解密模块、数据解压缩模块、消息处理模块、数据上传模块和数据下载模块;

[0013] 由所述的数据加密模块通过特定的算法对数据进行加密,此加密算法的描述作为元信息在消息处理模块进行封装,加密之后的数据输出至数据压缩模块;

[0014] 由所述的数据压缩模块使用特定的压缩算法对数据加密模块的输出进行处理,压缩之后的数据作为输入传送到消息处理模块,此压缩算法的描述作为元信息也传送到消息处理模块;

[0015] 由所述的数据解密模块从消息处理模块中数据的元信息中获取加密算法的描述,根据此加密算法对数据进行解密,数据解密模块的处理结果作为输入传输至数据解压缩模块;

[0016] 由所述的数据解压缩模块从消息处理模块数据的元信息中获取对压缩算法的描述,根据此压缩算法对数据进行解压缩;

[0017] 由所述的消息处理模块在数据备份时将数据和元信息进行封装并序列化成语节流,处理结果作为输入传送到数据上传模块;在数据恢复时将从数据下载模块接收到的字节流进行反序列化,还原成机器可读的数据及相关信息,处理结果作为输入传送到数据解密模块;

[0018] 由所述的数据上传模块应用预先设置的文件上传策略和多线程方式将消息处理模块输出的字节流通过网络传输到云存储服务器;

[0019] 由所述的数据下载模块应用预先设置的文件下载策略和多线程方式接收网络传输过来的字节流,输出到消息处理模块。

[0020] 所述的数据上传模块和数据下载模块对应用文件上传策略和文件下载策略步骤之前还包括根据文件上传策略对本地数据和远程数据进行比较。

[0021] 通过以下方式对本地数据和远程数据进行比较:

[0022] 采用多个线程同时运行的方式分批从云存储服务器上获取数据元信息;

[0023] 通过数据标识符的字典顺序与本地文件进行分批比较。

[0024] 文件比较完毕后,将待上传的本地文件放到一个共享的数据上传队列中;上传数据的线程组不断轮询该队列,如果队列为空,上传线程组等待,否则上传线程组传输本地文件到云存储服务器。

[0025] 文件比较完毕后,将待下载的远程文件的元信息放到一个共享的数据下载队列中;下载数据的线程组不断轮询该队列,如果队列为空,下载线程组等待,

[0026] 所述数据加密和数据解密还包括对加密算法的选择,如果加密算法没有特别指

定,会设置默认的加密算法,同时还设置对应于此加密算法的一个自定义密码。

[0027] 进行数据压缩和数据解压缩的压缩算法是预先定义的。

[0028] 所述对数据进行压缩后再进行传输,同时进行数据分块传输。

[0029] 所述的数据压缩使用的算法为标准的 ZLIB 压缩算法。

[0030] 对数据的内容进行加密,同时在上传到云存储服务器端或者下载到本地时都要校验数据的完整性。

[0031] 所述的数据加密算法默认为 PBEWITHSHA1ANDDES 算法,但是也可选择其他自定义的加密算法。

[0032] 本发明对数据进行加密、压缩后,分块上传、备份;对数据进行压缩、分块下载后,进行解密;可有效提升云存储系统数据备份和恢复的速度、效率以及数据传输的安全性。

### 附图说明

[0033] 下面结合附图对本发明进一步说明:

[0034] 图 1 是本发明数据备份和恢复方法采用系统结构框图;

[0035] 图 2 是本发明进行数据备份流程示意图;

[0036] 图 3 是本发明进行数据恢复流程示意图。

### 具体实施方式

[0037] 见图 1 所示,本发明数据备份和恢复方法采用系统包括数据加密模块 1、数据压缩模块 2、数据解密模块 3、数据解压缩模块 4、消息处理模块 5、数据上传模块 6 和数据下载模块 7;的数据加密模块 1 通过特定的算法对数据进行加密,此加密算法的描述作为元信息在消息处理模块 5 进行封装,加密之后的数据输出至数据压缩模块 2。数据压缩模块 2 使用特定的压缩算法对数据加密模块 1 的输出进行处理,压缩之后的数据作为输入传送到消息处理模块 5,此压缩算法的描述作为元信息也传送到消息处理模块 5。数据解密模块 3 从消息处理模块 5 中数据的元信息中获取加密算法的描述,根据此加密算法对数据进行解密,数据解密模块 3 的处理结果作为输入传输至数据解压缩模块 4。数据解压缩模块 4 从消息处理模块 5 数据的元信息中获取对压缩算法的描述,根据此压缩算法对数据进行解压缩。消息处理模块 5 在数据备份时将数据和元信息进行封装并序列化成字节流,处理结果作为输入传送到数据上传模块 6;在数据恢复时将从数据下载模块 7 接收到的字节流进行反序列化,还原成机器可读的数据及相关信息,处理结果作为输入传送到数据解密模块 (3)。数据上传模块 6 应用预先设置的文件上传策略和多线程方式将消息处理模块 5 输出的字节流通过网络传输到云存储服务器。数据下载模块 7 应用预先设置的文件下载策略和多线程方式接收网络传输过来的字节流,输出到消息处理模块 5。

[0038] 在前述数据上传模块 6 和数据下载模块 7 对应用文件上传策略和文件下载策略步骤之前还包括根据文件上传策略对本地数据和远程数据进行比较;比较的方法是:

[0039] 采用多个线程同时运行的方式分批从云存储服务器上获取数据元信息;

[0040] 通过数据标识符的字典顺序与本地文件进行分批比较。

[0041] 文件比较完毕后,将待上传的本地文件放到一个共享的数据上传队列中;上传数据的线程组不断轮询该队列,如果队列为空,上传线程组等待,否则上传线程组传输本地文

件到云存储服务器。

[0042] 文件比较完毕后,将待下载的远程文件的元信息放到一个共享的数据下载队列中;下载数据的线程组不断轮询该队列,如果队列为空,下载线程组等待,否则下载线程组传输数据到本地文件系统。

[0043] 前述数据加密和数据解密还包括对加密算法的选择,如果加密算法没有特别指定,会设置默认的加密算法,同时还设置对应于此加密算法的一个自定义密码。

[0044] 前述对数据进行压缩后再进行传输,同时进行数据分块传输。数据压缩使用的算法为标准的 ZLIB 压缩算法。

[0045] 前述机制中数据的内容都是经过加密,同时在上传到云存储服务器端或者下载到本地时都要校验数据的完整性。数据加密算法默认为 PBEWITHSHA1ANDDES 算法,但是也可选择其他自定义的加密算法。

[0046] 见图 2 所示,本发明中基于云存储进行数据备份的方法包括以下实施步骤:

[0047] 第 1 步,将待备份的数据块分别进行压缩,记录所使用压缩算法;

[0048] 第 2 步,将压缩后的数据进行加密,该加密算法是可自定义的,加密算法也会被记录;

[0049] 第 3 步,对加密后的数据以及期间所使用的算法和 hash 检验值等进行封装;

[0050] 第 4 步,通过网络传输封装后的消息数据,

[0051] 第 5 步,云存储服务端接收到消息后,根据消息中的元信息对数据进行校验;

[0052] 如果通过,那么说明数据在传输过程中未被篡改,数据将被写入文件系统;

[0053] 如果检验失败,那么会丢弃该数据块,并返回数据备份失败响应。

[0054] 如图 3 所示,本发明中基于云存储进行数据恢复的方法包括以下实施步骤:

[0055] 第 1 步,接收云存储服务端的消息,并验证;

[0056] 如果数据块通过一致性校验,那么将生成一个新的对象,该对象包括数据内容和对应的元信息;

[0057] 如果数据块的一致性校验失败,那么说明该数据块在传输过程中被篡改,该数据块被丢弃;

[0058] 第 2 步,从消息中获取数据块的解密算法,对数据块进行解密操作;

[0059] 第 3 步,从消息中获取数据块的解压缩算法,对数据库进行解压缩操作;

[0060] 第 4 步,将数据保存至本地文件系统。

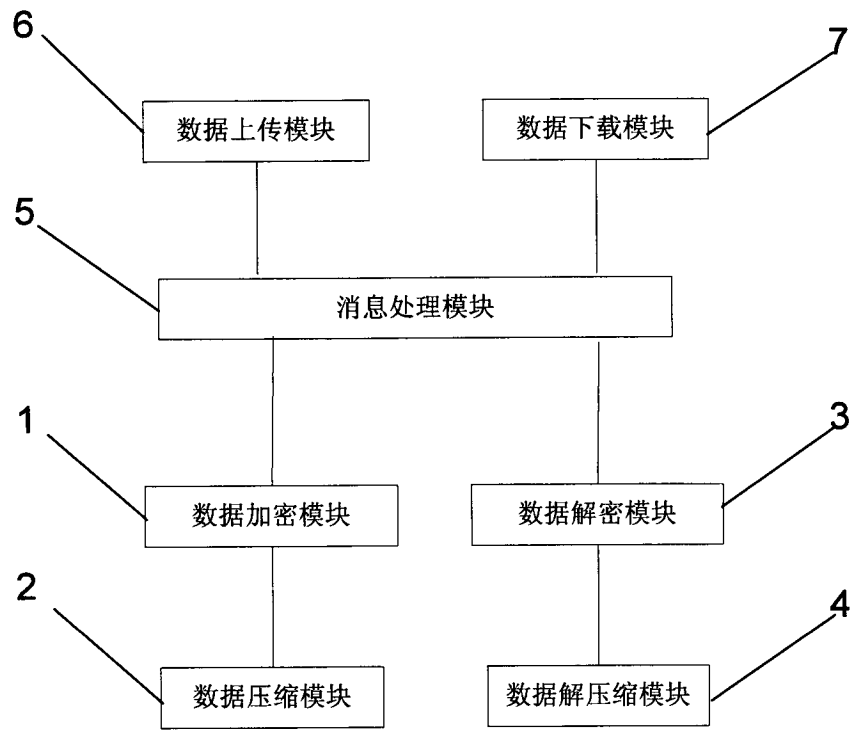


图 1



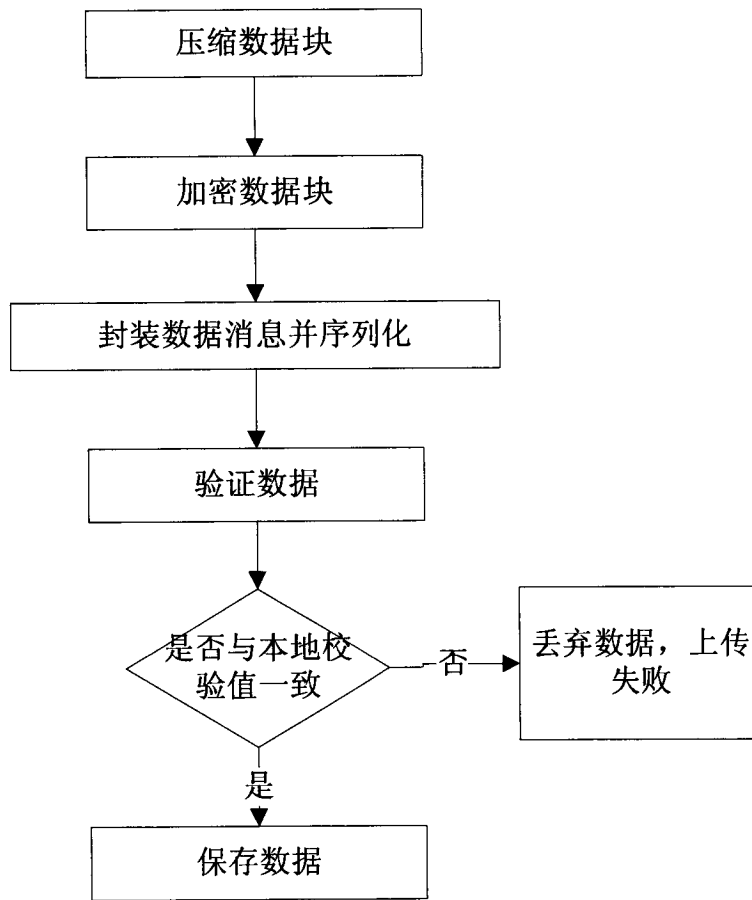


图 2

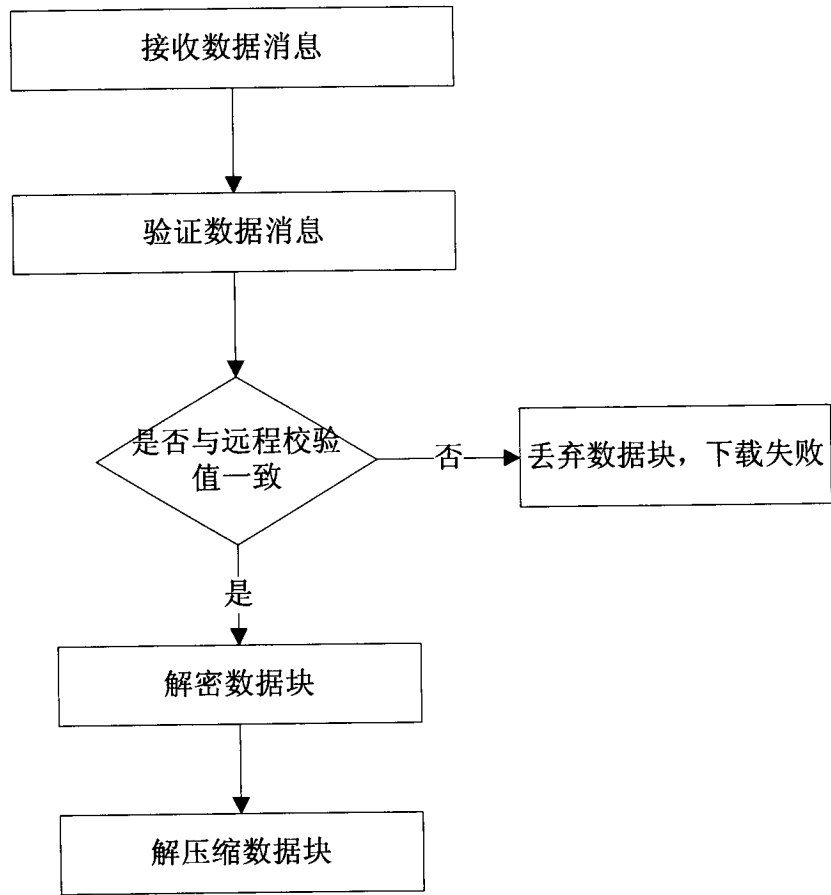


图 3