



(12) 发明专利

(10) 授权公告号 CN 109408599 B

(45) 授权公告日 2021.09.28

(21) 申请号 201811099754.1

G06F 16/22 (2019.01)

(22) 申请日 2018.09.20

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 109408599 A

CN 102968498 A, 2013.03.13
CN 104794177 A, 2015.07.22
CN 102004761 A, 2011.04.06
CN 103390015 A, 2013.11.13

(43) 申请公布日 2019.03.01

US 2015120745 A1, 2015.04.30

(73) 专利权人 佛山科学技术学院
地址 528000 广东省佛山市南海区狮山镇
仙溪水库西路佛山科学技术学院

张栗粽等. 面向大数据分布式存储的动态负载均衡算法.《计算机科学》.2017,第44卷(第5期),

(72) 发明人 张彩霞 郭静 王向东 王新东
胡绍林

审查员 陈飞

(74) 专利代理机构 广州嘉权专利商标事务所有
限公司 44205
代理人 王国标

(51) Int. Cl.

G06F 16/28 (2019.01)

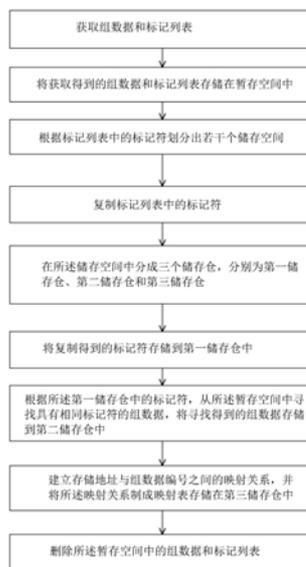
权利要求书1页 说明书3页 附图2页

(54) 发明名称

一种大数据的分布式存储方法

(57) 摘要

本发明公开了一种大数据的分布式存储方法,包括:获取组数据和标记列表;将获取得到的组数据和标记列表存储在暂存空间中;根据标记列表中的标记符划分出若干个存储空间;复制标记列表中的标记符;在所述储存空间中分成三个储存仓,分别为第一储存仓、第二储存仓和第三储存仓;将复制得到的标记符存储到第一储存仓中;根据所述第一储存仓中的标记符,从所述暂存空间中寻找具有相同标记符的组数据,将寻找得到的组数据存储到第二储存仓中;建立存储地址与组数据编号之间的映射关系,并将所述映射关系制成映射表存储在第三储存仓中。本发明存储有序,避免了现有技术中存储混乱的情况。可用于大数据存储领域。



1. 一种大数据的分布式存储方法,其特征在于,包括:
 - 步骤1、获取组数据和标记列表;
 - 步骤2、将获取得到的组数据和标记列表存储在暂存空间中;
 - 步骤3、根据标记列表中的标记符划分出若干个储存空间;
 - 步骤4、复制标记列表中的标记符;
 - 步骤5、在所述储存空间中分成三个储存仓,分别为第一储存仓、第二储存仓和第三储存仓;
 - 步骤6、将复制得到的标记符存储到第一储存仓中;
 - 步骤7、根据所述第一储存仓中的标记符,从所述暂存空间中寻找具有相同标记符的组数据,将寻找得到的组数据存储到第二储存仓中;
 - 步骤8、建立存储地址与组数据编号之间的映射关系,并将所述映射关系制成映射表存储在第三储存仓中。
2. 根据权利要求1所述的一种大数据的分布式存储方法,其特征在于,在步骤1中,所述组数据和标记列表的获取方法包括:
 - 步骤1.1、从工业场景中采集数据,并对采集得到的数据按照采集源头进行分类,分类后的数据记为准类数据;
 - 步骤1.2、设立描述不同采集源头的标记符;
 - 步骤1.3、将标记符对应的加载到所述准类数据中,加载标记符后的准类数据记为类数据;
 - 步骤1.4、将类数据中的标记符分离出来;
 - 步骤1.5、将类数据按照顺序划分出若干个等数据量大小的子数据;
 - 步骤1.6、给所述子数据分配描述子数据之间顺序关系的编号;
 - 步骤1.7、复制两次所述标记符;
 - 步骤1.8、将所述编号与第一次复制得到的标记符对应地附着在子数据中,附着标记符和编号的子数据记为组数据;
 - 步骤1.9、将第二次复制得到的标记符汇集在表格文件中,得到标记列表;
 - 步骤1.10、将所述组数据和标记列表上传至互联网;
 - 步骤1.11、从互联网中获取得到组数据和标记列表。
3. 根据权利要求1所述的一种大数据的分布式存储方法,其特征在于,所述储存空间包括数据库。
4. 根据权利要求1所述的一种大数据的分布式存储方法,其特征在于,在步骤8后面还包括:
 - 步骤9、删除所述暂存空间中的组数据和标记列表。

一种大数据的分布式存储方法

技术领域

[0001] 本发明创造涉及数据存储技术领域,特别涉及一种大数据的分布式存储方法。

背景技术

[0002] 随着行业应用系统的发展,各种信息系统应用数据快速增长,企业的信息化建设已逐步进入了大数据时代。

[0003] 大数据的存储一般会将数据上传至互联网,通过互联网的储存空间对数据进行存储,现有的存储方法一般会直接将大数据存储在储存空间中,容易造成存储混乱,不利于后续整理。

发明内容

[0004] 本发明的目的是提供一种有序的大数据存储方法。

[0005] 本发明解决其技术问题的解决方案是:一种大数据的分布式存储方法,包括:

[0006] 步骤1、获取组数据和标记列表;

[0007] 步骤2、将获取得到的组数据和标记列表存储在暂存空间中;

[0008] 步骤3、根据标记列表中的标记符划分出若干个储存空间;

[0009] 步骤4、复制标记列表中的标记符;

[0010] 步骤5、在所述储存空间中分成三个储存仓,分别为第一储存仓、第二储存仓和第三储存仓;

[0011] 步骤6、将复制得到的标记符存储到第一储存仓中;

[0012] 步骤7、根据所述第一储存仓中的标记符,从所述暂存空间中寻找具有相同标记符的组数据,将寻找得到的组数据存储到第二储存仓中;

[0013] 步骤8、建立存储地址与组数据编号之间的映射关系,并将所述映射关系制成映射表存储在第三储存仓中。

[0014] 进一步,在步骤1中,所述组数据和标记列表的获取方法包括:

[0015] 步骤1.1、从工业场景中采集数据,并对采集得到的数据按照采集源头进行分类,分类后的数据记为准类数据;

[0016] 步骤1.2、设立描述不同采集源头的标记符;

[0017] 步骤1.3、将标记符对应的加载到所述准类数据中,加载标记符后的准类数据记为类数据;

[0018] 步骤1.4、将类数据中的标记符分离出来;

[0019] 步骤1.5、将类数据按照顺序划分出若干个等数据量大小的子数据;

[0020] 步骤1.6、给所述子数据分配描述子数据之间顺序关系的编号;

[0021] 步骤1.7、复制两次所述标记符;

[0022] 步骤1.8、将所述编号与第一次复制得到的标记符对应地附着在子数据中,附着标记符和编号的子数据记为组数据;

- [0023] 步骤1.9、将第二次复制得到的标记符汇集在表格文件中,得到标记列表;
- [0024] 步骤1.10、将所述组数据和标记列表上传至互联网;
- [0025] 步骤1.11、从互联网中获取得到组数据和标记列表。
- [0026] 进一步,所述储存空间包括数据库。
- [0027] 进一步,在步骤8后面还包括:
- [0028] 步骤9、删除所述暂存空间中的组数据和标记列表。
- [0029] 本发明的有益效果是:本发明通过将储存空间划分为三个储存仓,并从组数据中提取编号和在标记列表中提取标记符,将标记符存储在第一储存仓中,根据所述编号将组数据有序的存储在第二储存仓中,并同时记载组数据的存储地址与编号之间映射关系的映射表存储在第三储存仓中,存储有序,避免了现有技术中存储混乱的情况。

附图说明

[0030] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单说明。显然,所描述的附图只是本发明的一部分实施例,而不是全部实施例,本领域的技术人员在不付出创造性劳动的前提下,还可以根据这些附图获得其他设计方案和附图。

[0031] 图1是本发明方法的步骤流程图;

[0032] 图2是组数据和标记列表的获取方法的步骤流程图。

具体实施方式

[0033] 以下将结合实施例和附图对本发明的构思、具体结构及产生的技术效果进行清楚、完整地描述,以充分地理解本发明的目的、特征和效果。显然,所描述的实施例只是本发明的一部分实施例,而不是全部实施例,基于本发明的实施例,本领域的技术人员在不付出创造性劳动的前提下所获得的其他实施例,均属于本发明保护的范围。另外,文中所提到的所有联接/连接关系,并非单指构件直接相接,而是指可根据具体实施情况,通过添加或减少联接辅件,来组成更优的联接结构。本发明创造中的各个技术特征,在不互相矛盾冲突的前提下可以交互组合。

[0034] 实施例1,参考图1和图2,一种大数据的分布式存储方法,包括:

[0035] 步骤1、获取组数据和标记列表;

[0036] 步骤2、将获取得到的组数据和标记列表存储在暂存空间中;利用暂存空间作为存储的缓冲。

[0037] 步骤3、根据标记列表中的标记符划分出若干个储存空间;

[0038] 步骤4、复制标记列表中的标记符;

[0039] 步骤5、在所述储存空间中分成三个储存仓,分别为第一储存仓、第二储存仓和第三储存仓;

[0040] 步骤6、将复制得到的标记符存储到第一储存仓中;

[0041] 步骤7、根据所述第一储存仓中的标记符,从所述暂存空间中寻找具有相同标记符的组数据,将寻找得到的组数据存储到第二储存仓中;

[0042] 步骤8、建立存储地址与组数据编号之间的映射关系,并将所述映射关系制成映射

表存储在第三储存仓中；

[0043] 步骤9、删除所述暂存空间中的组数据和标记列表。

[0044] 本发明通过将储存空间划分为三个储存仓，并从组数据中提取编号和在标记列表中提取标记符，将标记符存储在所述第一储存仓中，根据所述编号将组数据有序的存储在第二储存仓中，并同时记载组数据的存储地址与编号之间映射关系的映射表存储在第三储存仓中，存储有序，避免了现有技术中存储混乱的情况。

[0045] 其中，在步骤1中，所述组数据和标记列表的获取方法包括：

[0046] 步骤1.1、从工业场景中采集数据，并对采集得到的数据按照采集源头进行分类，分类后的数据记为准类数据；

[0047] 其中，所述采集源头可以为不同的设备，可以对采集得到的数据按照不同设备进行分类，比如，将从A设备采集的数据为一类，从B设备采集的数据为另一类。

[0048] 步骤1.2、设立描述不同采集源头的标记符；例如，从A设备采集的数据，则可设立标记符a，从B设备采集的数据则可设立标记符b。

[0049] 步骤1.3、将标记符对应的加载到所述准类数据中，加载标记符后的准类数据记为类数据；

[0050] 步骤1.4、将类数据中的标记符分离出来；

[0051] 步骤1.5、将类数据按照顺序划分出若干个等数据量大小的子数据；具体的，将类数据按照数据量的大小划分出若干个等数据量大小的子数据，例如类数据的数据量为100KB，那么可以划分出5个子数据，每一个子数据的数据量为20KB。

[0052] 步骤1.6、给所述子数据分配描述子数据之间顺序关系的编号；

[0053] 步骤1.7、复制两次所述标记符；

[0054] 步骤1.8、将所述编号与第一次复制得到的标记符对应地附着在子数据中，附着标记符和编号的子数据记为组数据；

[0055] 步骤1.9、将第二次复制得到的标记符汇集在表格文件中，得到标记列表；

[0056] 步骤1.10、将所述组数据和标记列表上传至互联网；

[0057] 步骤1.11、从互联网中获取得到组数据和标记列表。

[0058] 作为优化，所述储存空间包括数据库。

[0059] 以上对本发明的较佳实施方式进行了具体说明，但本发明创造并不限于所述实施例，熟悉本领域的技术人员在不违背本发明精神的前提下还可做出种种的等同变型或替换，这些等同的变型或替换均包含在本申请权利要求所限定的范围内。

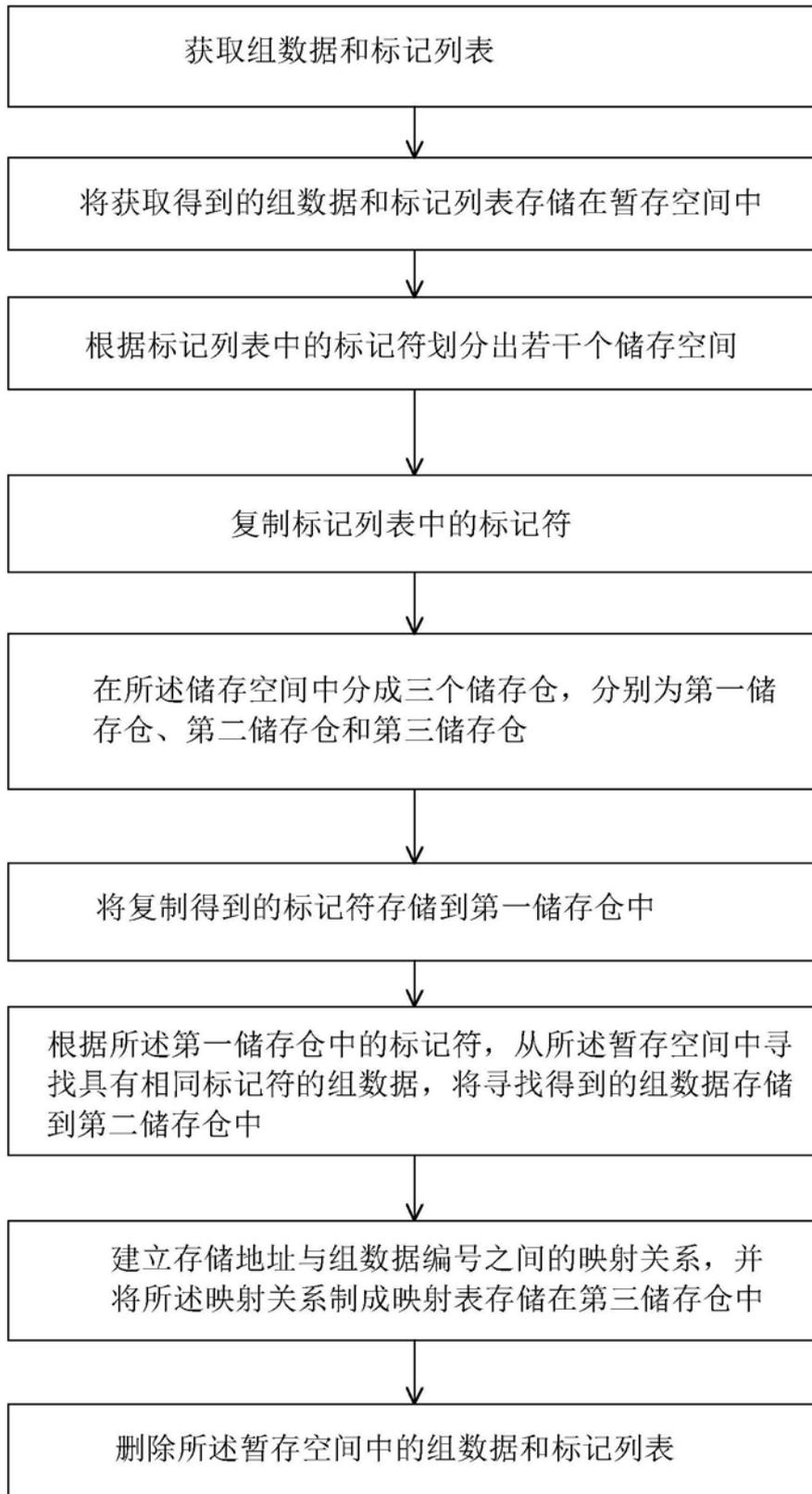


图1

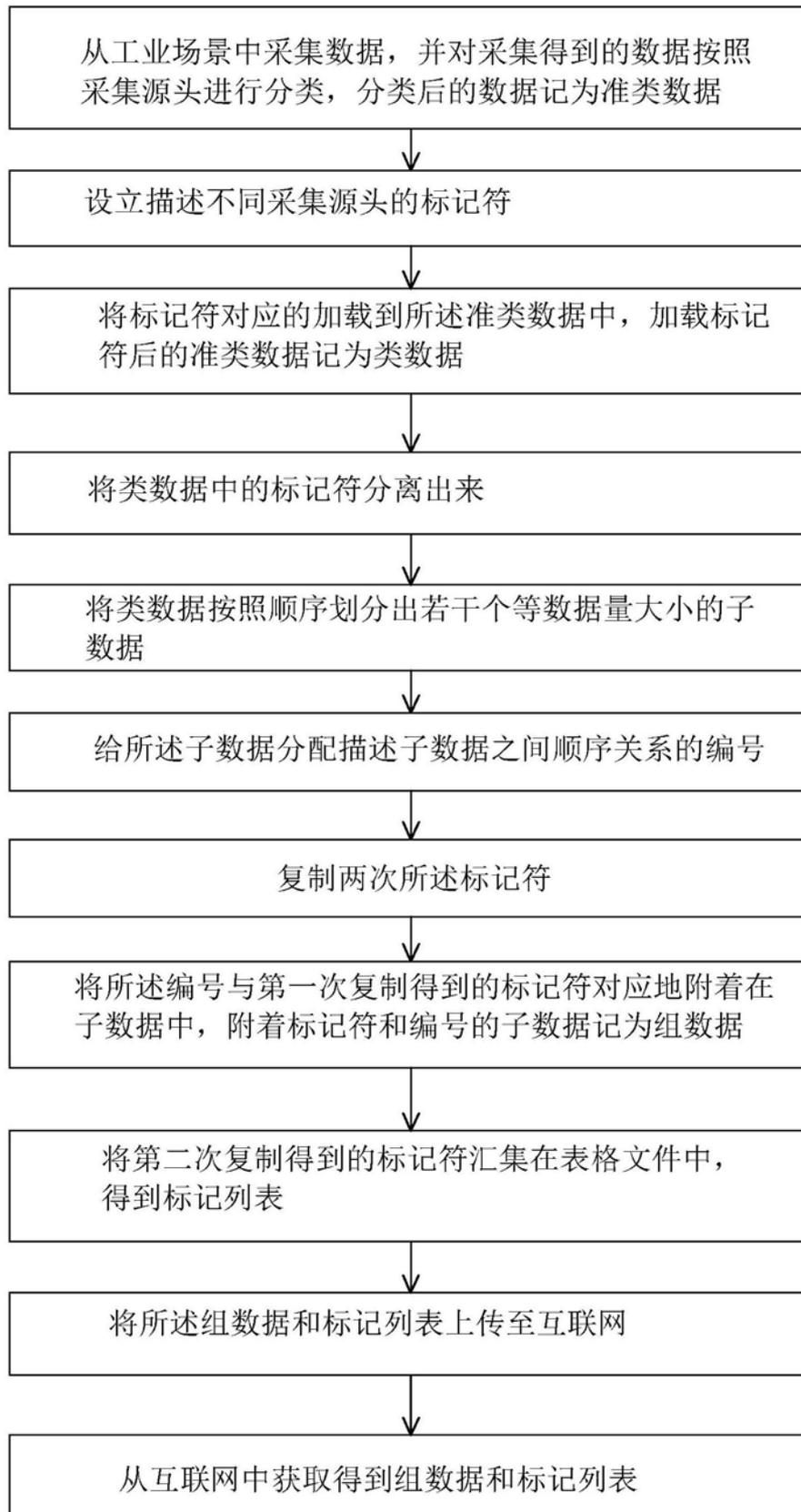


图2