



(12)发明专利

(10)授权公告号 CN 103473276 B

(45)授权公告日 2017.08.25

(21)申请号 201310376425.8

(22)申请日 2013.08.26

(65)同一申请的已公布的文献号
申请公布号 CN 103473276 A

(43)申请公布日 2013.12.25

(73)专利权人 广东电网公司电力调度控制中心
地址 510699 广东省广州市越秀区梅花路
75号

(72)发明人 林斌 李星南 杨德强 余锦业
包达志 姜绍艳 李溢杰 李伟坚
蒋康明

(74)专利代理机构 广州华进联合专利商标代理
有限公司 44224
代理人 王茹 曾旻辉

(51)Int.Cl.

G06F 17/30(2006.01)

(56)对比文件

US 6092061 A,2000.07.18,
CN 102737132 A,2012.10.17,
CN 102375853 A,2012.03.14,
CN 102112962 A,2011.06.29,

审查员 魏留强

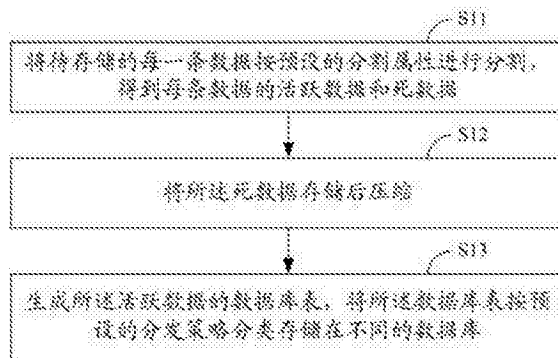
权利要求书2页 说明书5页 附图1页

(54)发明名称

超大型数据存储方法、分布式数据库系统及其检索方法

(57)摘要

本发明提供一种超大型数据存储方法,包括如下步骤:将待存储的每一条数据按预设的分割属性进行分割,得到每条数据的活跃数据和死数据;将所述死数据存储后压缩;生成所述活跃数据的数据库表,将所述数据库表按预设的分发策略分类存储在不同的数据库。本发明还提供一种分布式数据库系统,以及一种分布式数据库系统的检索方法,能解决超大型结构化数据的存储难题,存储资源消耗小,数据的检索速度快。



1. 一种超大型数据存储方法,其特征在于,包括如下步骤:

将待存储的每一条数据按预设的分割属性进行分割,得到每条数据的活跃数据和死数据;其中,所述活跃数据是属性为可能应用于查询条件、统计、关联类的数据,所述死数据是属性为不会应用于查询条件、统计、关联类的数据;

将所述死数据存储后压缩;

生成所述活跃数据的数据库表,将所述数据库表按预设的分发策略分类存储在不同的数据库;其中,所述分发策略包括按照业务和用户的末M位尾号进行分类存储,其中M是1位、2位、...、任意位。

2. 根据权利要求1所述的超大型数据存储方法,其特征在于,还包括步骤:建立每条数据中活跃数据与死数据的映射。

3. 根据权利要求1所述的超大型数据存储方法,其特征在于,将所述死数据存储后压缩的步骤为:将所述死数据按预设的占用单位进行分块存储并压缩。

4. 根据权利要求2所述的超大型数据存储方法,其特征在于,所述建立每条数据中活跃数据与死数据的映射的步骤为:

记录所述死数据压缩后的文件名、存储位置和数据长度;

将所述死数据压缩后的文件名、存储位置和数据长度附加至对应的活跃数据的数据库表中。

5. 一种分布式数据库系统,其特征在于,包括多个数据库,每个所述数据库中存储有不同类别的多个数据库表,所述数据库还以压缩文件包的形式存储有多条死数据;所述数据库表存储多条活跃数据,每条所述活跃数据附加有与对应的死数据的映射;其中,所述死数据和所述活跃数据为根据预设的分割属性从待存储的每一条数据分割而得到,所述活跃数据是属性为可能应用于查询条件、统计、关联类的数据,所述死数据是属性为不会应用于查询条件、统计、关联类的数据;其中,所述数据库表按预设的分发策略分类存储在不同的数据库,所述分发策略包括按照业务和用户的末M位尾号进行分类存储,其中M是1位、2位、...、任意位。

6. 根据权利要求5所述的一种分布式数据库系统,其特征在于,所述死数据的映射为所述死数据压缩后的压缩文件名、存储位置和数据长度。

7. 一种分布式数据库系统的检索方法,其特征在于,包括如权利要求6所述的分布式数据库系统,包括如下步骤:

接收数据的检索请求;

若所述检索请求为活跃数据检索请求,则结合数据分发策略根据所述检索请求确定所述分布式数据库系统中需检索的数据库及其数据库表;将检索请求分发到各个所述需检索的数据库;接收各个数据库返回的检索结果后输出;其中,各个数据库根据以下方式进行检索:

先将检索需求进行分析,得到需要在哪些数据库中的哪些表中进行查询,然后将请求分发给每个表,基于每个表进行聚合操作;

每个分布式的表聚合操作完成后,如涉及多个数据库之间的二次聚合,则根据二次聚合特征将数据分发到不同的数据库中,各数据库在收集齐数据后再进行二次聚合操作;

二次聚合操作完成后进行数据统计,将统计后的结果返回;

若所述检索请求为死数据检索请求,则通过对应的活跃数据中的映射,查找到死数据的压缩文件,读取压缩文件中的死数据后输出。

8. 根据权利要求7所述分布式数据库系统的检索方法,其特征在于,所述通过对应的活跃数据中的映射,查找到死数据的压缩文件,读取压缩文件中的死数据后输出的步骤为:

根据所述死数据检索请求,从所述死数据对应的活跃数据查找死数据的压缩文件名,打开对应的所述压缩文件,将读取指针移动到所述存储位置,读取相应数据长度的信息,得到所述死数据。

超大型数据存储方法、分布式数据库系统及其检索方法

技术领域

[0001] 本发明涉及数据库技术领域,特别是涉及一种超大型数据存储方法,一种分布式数据库系统,以及一种分布式数据库系统的检索方法。

背景技术

[0002] 21世纪是一个数据爆炸的时代,特别是对数据定义的逐步对象化、精细化,越来越多的结构化数据会随之产生。特别是当前通信运营商、互联网行业目前对用户行为分析关注度越来越高,伴随着需要存储与查询分析的数据量也越来越大。

[0003] 例如,某省级电信运营商上网清单数据量每天都是在10多亿条记录以上,单表一天数据量达300G以上,而类似的清单数据类型达几十种,由于一般需要保留三个月到半年,因此所需要的存储资源一般可以达到PB级。

[0004] 如此超大数据的存储如果采用传统的关系型数据库来存储,会对主机的性能、存储资源都会有很高的要求,需要有巨大的投资。目前业界比较流行的大数据存储包括hadoop、mongodb、mysql等分布式数据存储与统计技术,但这些技术面对如此大的数据存储与应用,仍然需要较大的硬件开销。

[0005] 大数据存储目前业界一般有二种模式,一种是share nothing架构,另一种是share everything架构,其中share nothing架构主要是数据分散存储,各节点间不进行数据共享访问,share everything架构是数据存储在共享磁阵中,各分布式节点均可以共享访问。

[0006] 其中由于share everything架构需要共享磁阵,导致对磁阵的带宽总线以及网络构造极高,在数据量增长到PB级别时,基本没有实现优势。而share nothing 由于是分布式的数据存储架构,因此节点可以灵活扩展,少到10台左右的集群,多到几万台的集群,均可以满足数据量不断增加的应用需要。

[0007] 目前业界常用的hadoop、mongodb、mysql等分布式架构均是采用share nothing来进行的实现,虽然从数据存储与查询角度都可以满足超大型数据的应用需求,但其对存储资源的消耗都是几倍于原始数据的大小,因此目前迫切需要一种能消耗较少的硬件资源的数据存储方案。

发明内容

[0008] 基于此,本发明提供一种超大型数据存储方法,一种分布式数据库系统,以及一种分布式数据库系统的检索方法,能解决超大型结构化数据的存储难题,存储资源消耗小,数据的检索速度快。

[0009] 一种超大型数据存储方法,包括如下步骤:

[0010] 将待存储的每一条数据按预设的分割属性进行分割,得到每条数据的活跃数据和死数据;

[0011] 将所述死数据存储后压缩;

[0012] 生成所述活跃数据的数据库表,将所述数据库表按预设的分发策略分类存储在不同的数据库。

[0013] 一种分布式数据库系统,包括多个数据库,每个所述数据库中存储有不同类别的多个数据库表,所述数据库还以压缩文件包的形式存储有多条死数据;所述数据库表存储多条活跃数据,每条所述活跃数据附加有与对应的死数据的映射;其中,所述死数据和所述活跃数据为根据预设的分割属性从待存储的每一条数据分割而得到。

[0014] 一种分布式数据库系统的检索方法,包括上述分布式数据库系统,检索方法包括如下步骤:

[0015] 接收数据的检索请求;

[0016] 若所述检索请求为活跃数据检索请求,则基于数据分发策略根据所述检索请求确定所述分布式数据库系统中需检索的数据库及其数据库表;将检索请求分发到各个所述需检索的数据库;接收各个数据库返回的检索结果后输出;

[0017] 若所述检索请求为死数据检索请求,则通过对应的活跃数据中的映射,查找到死数据的压缩文件,读取压缩文件中的死数据后输出。

[0018] 上述的超大型数据存储方法中,将数据按预设的分割属性进行分割,对其中的死数据存储后压缩,大大减少了存储资源的消耗;而其他的活跃数据则分类存储在不同的数据库;本发明的超大型数据存储方法能解决超大型数据的存储难题,占用空间较小,大幅提高了数据检索和统计的处理速度。

[0019] 上述的分布式数据库系统中,包括多个数据库,数据库中存储两类数据,一种是不同类别的多个数据库表,对应存储多条业务数据;另一类是按预设的分割属性分割而得到的死数据,对其以压缩文件包的形式存储;通过死数据的压缩存储及活跃数据的分类存储;相对于其他的超大型数据库本发明的分布式数据库具有占用空间小、存储资源消耗小的优点;并且对于该数据库的数据检索和统计的处理速度非常快。

附图说明

[0020] 图1为本发明超大型数据存储方法在一实施例中的流程示意图。

[0021] 图2为本发明分布式数据库系统的检索方法在一实施例中的流程示意图。

具体实施方式

[0022] 下面结合实施例及附图对本发明作进一步详细说明,但本发明的实施方式不限于此。

[0023] 如图1所示,是本发明超大型数据存储方法在一实施例中的流程示意图,包括如下步骤:

[0024] S11、将待存储的每一条数据按预设的分割属性进行分割,得到每条数据的活跃数据和死数据;

[0025] S12、将所述死数据存储后压缩;

[0026] 对于待存储的数据,一般情况下,除了可能应用于查询条件、统计、关联类的属性外,往往还存在大量的非活跃类数据属性(不会应用于查询条件、统计、关联类的属性),即死数据,这些属性数据如果统一采用活跃数据的存储模式,会造成存储空间的浪费,检索效

率的低下。

[0027] 根据实际的业务需要,通过定义不同的数据中哪些属性是死数据,对每一条数据进行分割,将这些死数据直接以压缩文件的方式进行存储,这样对死数据的存储会大大降低存储空间。一般文本文件的压缩程度是在5到100倍左右,因此可带来几倍或是几十倍存储空间的节省,大大减少了存储资源的消耗。

[0028] 其中,所述将所述死数据存储后压缩的步骤可为:将所述死数据按预设的占用单位进行分块存储并压缩;对于死数据除了分离成单独的文件进行存储外,需要同时考虑将文件分成固定大小的小文件块进行存储,如以32M、64M、128M、256M等不同单位进行存储,这样在数据压缩与解压缩时会更加迅速,满足实时用户交互等场合的查询需求。

[0029] S13、生成所述活跃数据的数据库表,将所述数据库表按预设的分发策略分类存储在不同的数据库。

[0030] 一般数据具备多种属性特征,如日期时间、区域、设备、业务、用户等,通过预设的数据分发策略,将数据分类存储在不同的数据库,有利于提高检索和统计的效率。

[0031] 如日期时间一般可以细到天、小时、5分钟、1分钟级别来作为数据分发的策略,将不同时间的数据分发到不同的表中进行存储,在需要相应数据的时候,可以先定位到数据在哪些表中,然后再查这些表来得到相应的数据,避免需要扫描整份数据。

[0032] 对于区域,可以细到省、地市、县区、营服、网格粒度来作为数据分发策略;对于设备,可以细到具体的设备。

[0033] 对于业务、用户可以按照业务、用户的末M位尾号(其中M可以是1位、2位、...任意位)来定义数据分发,除了可以达到时间、区域细分的效果外,在应用计算用户数、业务数的应用场合可以避免数据重叠原因造成的频繁交互比较。

[0034] 通过组合不同的分发属性策略可以达到更优的数据分发效果,让数据检索与统计效率达到最优。

[0035] 在一较佳实施例中,还可包括步骤:建立每条数据中活跃数据与死数据的映射;

[0036] 其中,所述建立每条数据中活跃数据与死数据的映射的步骤可为:

[0037] 记录所述死数据压缩后的文件名、存储位置和数据长度;

[0038] 将所述死数据压缩后的文件名、存储位置和数据长度附加至对应的活跃数据的数据库表中;

[0039] 本实施例中通过建立映射保持每条数据中活跃数据和死数据的关联,记录死数据被压缩后的压缩文件名称,该条死数据在压缩文件中的存储位置,以及该条死数据的数据长度;当需要查询死数据时,可先通过活跃数据找到文件ID,然后打开这个文件,并将文件指针移动到文件存储位置的首地址,然后读取到相应数据长度的信息后返回,满足死数据的查询需求。

[0040] 本发明的超大型数据存储方法中,将数据按预设的分割属性进行分割,对其中的死数据存储后压缩,大大减少了存储资源的消耗;而其他的活跃数据则分类存储在不同的数据库;本发明的超大型数据存储方法能解决超大型数据的存储难题,占用空间较小,大幅提高了数据检索和统计的处理速度。

[0041] 本发明还提供一种分布式数据库系统,包括多个数据库,每个所述数据库中存储有不同类别的多个数据库表,所述数据库还以压缩文件包的形式存储有多条死数据;所述

数据库表存储多条活跃数据,每条所述活跃数据附加有与对应的死数据的映射;其中,所述死数据和所述活跃数据为根据预设的分割属性从待存储的每一条数据分割而得到。

[0042] 其中,所述死数据的映射为所述死数据压缩后的文件名、存储位置和数据长度。

[0043] 对于待存储的数据,一般情况下,除了可能应用于查询条件、统计、关联类的属性外,往往还存在大量的非活跃类数据属性(不会应用于查询条件、统计、关联类的属性),即死数据,这些属性数据如果统一采用活跃数据的存储模式,会造成存储空间的浪费,检索效率的低下。

[0044] 根据实际的业务需要,通过定义不同的数据中哪些属性是死数据,对每一条数据进行分割,将这些死数据直接以压缩文件的方式进行存储,这样对死数据的存储会大大降低存储空间。一般文本文件的压缩程度是在5到100倍左右,因此可带来几倍或是几十倍存储空间的节省,大大减少了存储资源的消耗。

[0045] 其中,可将所述死数据按预设的占用单位进行分块存储并压缩;因为对于死数据除了分离成单独的文件进行存储外,需要同时考虑将文件分成固定大小的小文件块进行存储,如以32M、64M、128M、256M等不同单位进行存储,这样在数据压缩与解压缩时会更加迅速,满足实时用户交互等场合的查询需求。

[0046] 而一般的活跃数据具备多种属性特征,如日期时间、区域、设备、业务、用户等,通过预设的数据分发策略,将数据分类存储在不同的数据库,有利于提高检索和统计的效率。

[0047] 本发明的分布式数据库系统中,包括多个数据库,数据库中存储两类数据,一种是不同类别的多个数据库表,对应存储多条业务数据;另一类是按预设的分割属性分割而得到的死数据,对其以压缩文件包的形式存储;通过死数据的压缩存储及活跃数据的分类存储;相对于其他的超大型数据库本发明的分布式数据库具有占用空间小、存储资源消耗小的优点,并且数据库的数据检索和统计的处理速度非常快。

[0048] 本发明还提供一种分布式数据库系统的检索方法,包括上述的分布式数据库系统,其检索方法包括如下步骤:

[0049] S21、接收数据的检索请求;

[0050] S22、若所述检索请求为活跃数据检索请求,则结合数据分发策略根据所述检索请求确定所述分布式数据库系统中需检索的数据库及其数据库表;将检索请求分发到各个所述需检索的数据库;接收各个数据库返回的检索结果后输出;

[0051] S23、若所述检索请求为死数据检索请求,则通过对应的活跃数据中的映射,查找到死数据的压缩文件,读取压缩文件中的死数据后输出;

[0052] 在一较佳实施例中,所述通过对应的活跃数据中的映射,查找到死数据的压缩文件,读取压缩文件中的死数据后输出的步骤为:

[0053] 根据所述死数据请求,从所述死数据对应的活跃数据查找死数据的压缩文件名,打开对应的所述压缩文件,将读取指针移动到所述存储位置,读取相应数据长度的信息,得到所述死数据。

[0054] 基于上述分布式数据库系统中的数据存储特点,对于活跃数据的检索可采用查询及合并的模式进行实现;

[0055] 可先将活跃数据的检索请求结合数据分发策略进行分析,得到需要在哪个数据库的哪些表中进行查询,然后将请求分发给各个数据库中对应的数据库表,基于每个表进行

查询操作;每个数据库的每个表查询完成后就可以进行数据合并,将合并后的数据返回。

[0056] 举例来说,对于存储用户在每个小区下的http请求响应清单来说,如果清单采用按用户号码尾号进行分发的策略进行存储,对于查询某些用户的访问清单需求来说就需要执行如下步骤:

[0057] 先在各个数据库的各个表中执行特定用户访问清单查询;

[0058] 在每个数据库均查询完成后,将各节点数据进行合并返回,达到本次检索需要。

[0059] 对于活跃数据中,需进行统计分析的检索请求,可采用聚合+分发+聚合+合并模式进行实现:

[0060] 先将检索需求进行分析,得到需要在哪些数据库中的哪些表中进行查询,然后将请求分发给每个表,基于每个表进行聚合操作;

[0061] 每个分布式的表聚合操作完成后,如涉及多个数据库之间的二次聚合,则需要根据二次聚合特征将数据分发到不同的数据库中,各数据库在收集齐数据后再进行二次聚合操作;

[0062] 二次聚合操作完成后就可以进行数据统计,将统计后的结果返回。

[0063] 举例来说,对于存储用户在每个小区下的http请求响应清单来说,如果清单采用按用户号码尾号进行分发的策略进行存储。对于统计各个小区下的访问流量需求来说就需要执行如下步骤:

[0064] 先在各个数据库的各个表中执行小区访问流量汇聚;

[0065] 将小区流量汇聚结果按小区进行分发给各处理数据库(不同的处理数据库处理固定的某些小区,可以按小区的hash码来分发);

[0066] 各处理数据库在收齐完小区流量汇聚记录后重新执行二次汇聚操作;

[0067] 所有数据库均汇聚完成后,将各数据库数据进行合并后返回,达到本次统计需要。

[0068] 而对于如果需要统计每个用户的访问流量的需求来说,因为清单数据是采用用户号码的尾号进行分发存储,则可以只执行上述第一和第四步就可以完成。

[0069] 对于死数据的检索请求,上述的分布式数据库系统中,在死数据存储时,需要在活跃数据中附加存储死数据存储在哪个文件(四个字节足够使用)、压缩文件位置(通过设置起始指针即可实现,四个字节足够使用)、数据长度(二个字节足够使用)这三类信息。

[0070] 在需要检索死数据时,先通过活跃数据找到文件ID,然后打开这个文件,并将文件指针移动到“文件位置起始指针”,然后读取相应“数据长度”的信息,即可读取到需要的死数据,返回结果。

[0071] 采用此种方式因为是直接定位到数据位置,没有数据的比较操作,因此定位实现会非常快,具体数据获取速度基本可以接近于硬盘的I/O速度。而虽然在活跃数据存储时至少需要增加10个字节的存储空间,但相对于死数据压缩带来的空间节省,这个浪费基本可以忽略不计,实践证明一般情况下最大浪费开销不会超过5%。

[0072] 以上所述实施例仅表达了本发明的几种实施方式,其描述较为具体和详细,但不能因此而理解为对本发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变形和改进,这些都属于本发明的保护范围。因此,本发明专利的保护范围应以所附权利要求为准。

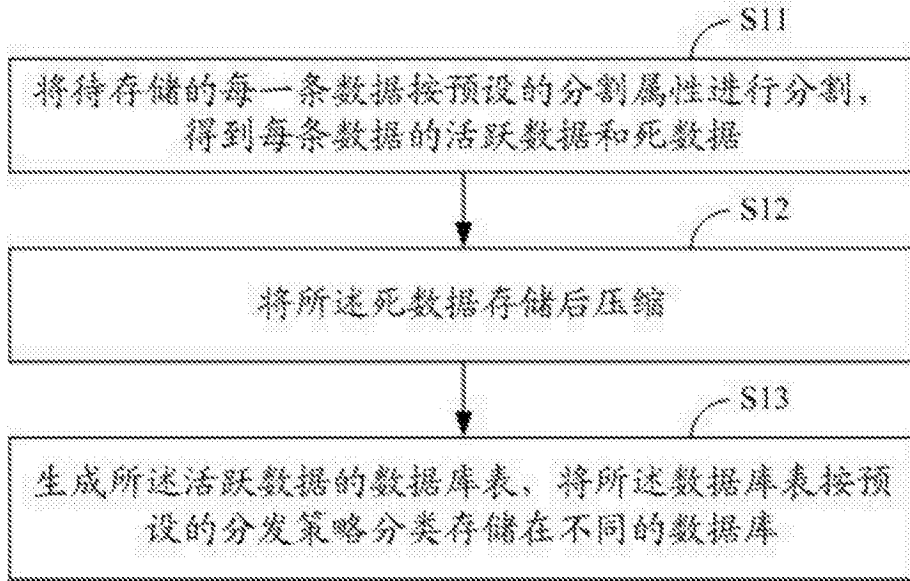


图1

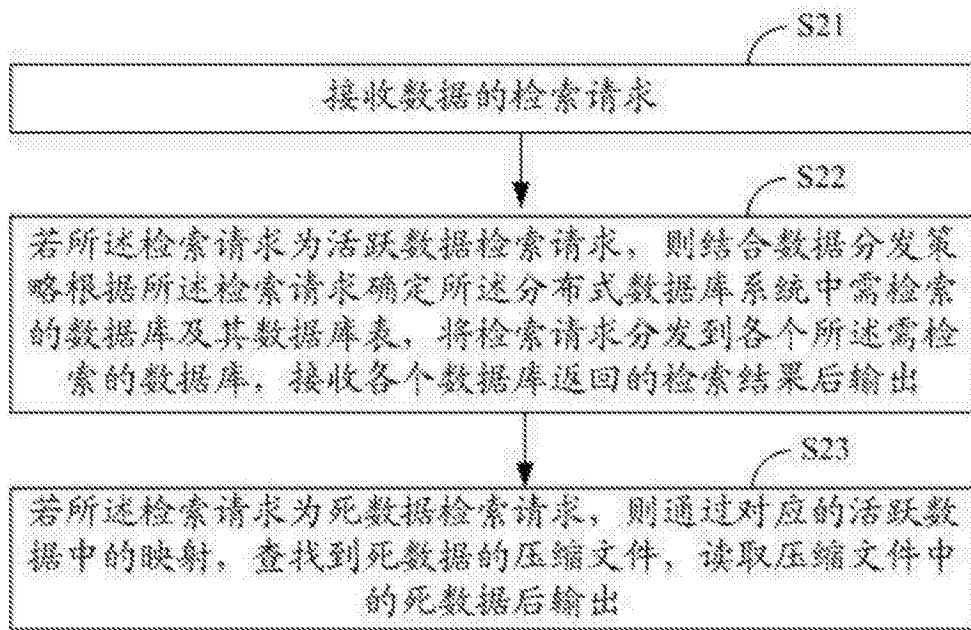


图2