



(12)发明专利

(10)授权公告号 CN 106933904 B

(45)授权公告日 2019.12.03

(21)申请号 201511032353.0

(22)申请日 2015.12.31

(65)同一申请的已公布的文献号

申请公布号 CN 106933904 A

(43)申请公布日 2017.07.07

(73)专利权人 北京国双科技有限公司

地址 100083 北京市海淀区北四环中路229号海泰大厦4层南401号

(72)发明人 洪超 黄健 何恺铎

(74)专利代理机构 北京康信知识产权代理有限公司 11240

代理人 韩建伟 张永明

(51)Int.Cl.

G06F 16/242(2019.01)

G06F 16/2455(2019.01)

(56)对比文件

CN 104408180 A,2015.03.11,

CN 102567329 A,2012.07.11,

CN 104408169 A,2015.03.11,

CN 104462434 A,2015.03.25,

CN 102567329 B,2013.10.23,

CN 104408183 A,2015.03.11,

CN 103064689 A,2013.04.24,

陈伟杰.“巧用SQL 的外连接”.《电脑知识与技术》.2013,

审查员 王洋洋

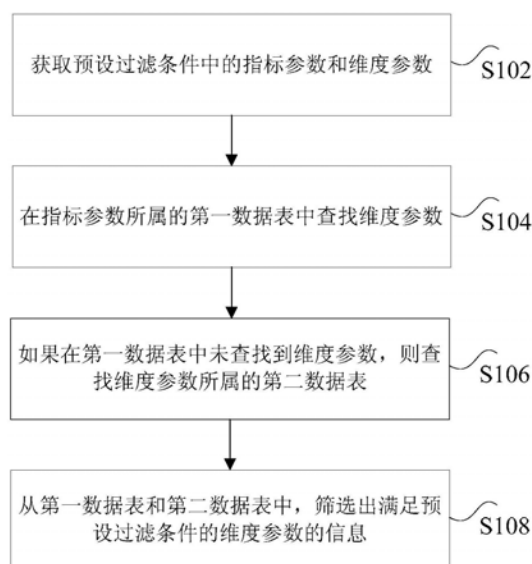
权利要求书2页 说明书10页 附图1页

(54)发明名称

数据的过滤方法和装置

(57)摘要

本申请公开了一种数据的过滤方法和装置。该方法包括:获取预设过滤条件中的指标参数和维度参数,其中,预设过滤条件为基于维度参数确定的对指标参数进行过滤的条件;在指标参数所属的第一数据表中查找维度参数;如果在第一数据表中未查找到维度参数,则查找维度参数所属的第二数据表;从第一数据表和第二数据表中,筛选出满足预设过滤条件的维度参数的信息。本申请解决了现有技术中无法对两个以上的数据表进行指标过滤的技术问题。



1. 一种数据的过滤方法,其特征在于,包括:

获取预设过滤条件中的指标参数和维度参数,其中,所述预设过滤条件为基于所述维度参数确定的对所述指标参数进行过滤的条件;

在所述指标参数所属的第一数据表中查找所述维度参数;

如果在所述第一数据表中未查找到所述维度参数,则查找所述维度参数所属的第二数据表,其中,所述第二数据表包括由多个子表按照逻辑关系进行表连接所得到的子表的集合,每个所述子表对应一种所述维度参数;

从所述第一数据表和所述第二数据表中,筛选出满足所述预设过滤条件的维度参数的信息。

2. 根据权利要求1所述的方法,其特征在于,从所述第一数据表和所述第二数据表中,筛选出满足所述预设过滤条件的维度参数的信息包括:

将所述第一数据表和所述第二数据表进行表连接,得到目标数据表;

从所述预设过滤条件提取所述指标参数的参数阈值;

在所述目标数据表中,筛选出指标参数大于所述参数阈值的所述维度参数的信息。

3. 根据权利要求2所述的方法,其特征在于,将所述第一数据表和所述第二数据表进行表连接包括:

按照预设连接方式,将所述第一数据表和所述第二数据表进行表连接,

其中,所述预设连接方式包括下述至少之一:内连接、左外连接、右外连接以及全外连接。

4. 根据权利要求1所述的方法,其特征在于,如果在所述第一数据表中查找到所述维度参数,所述方法还包括:

从所述预设过滤条件中提取所述指标参数的参数阈值;

在所述第一数据表中,筛选出指标参数大于所述参数阈值的所述维度参数的信息。

5. 根据权利要求1至4中任一项所述的方法,其特征在于,在筛选出满足所述预设过滤条件的维度参数的信息之后,所述方法还包括:

对筛选出的所述维度参数的信息进行聚合,得到信息集合;

读取内嵌有所述预设过滤条件的目标过滤条件,其中,所述目标过滤条件为基于所述预设过滤条件确定的对筛选维度进行过滤的条件;

从所述信息集合中,筛选出符合所述目标过滤条件的筛选维度的信息。

6. 一种数据的过滤装置,其特征在于,包括:

获取单元,用于获取预设过滤条件中的指标参数和维度参数,其中,所述预设过滤条件为基于所述维度参数确定的对所述指标参数进行过滤的条件;

第一查找单元,用于在所述指标参数所属的第一数据表中查找所述维度参数;

第二查找单元,用于在所述第一数据表中未查找到所述维度参数的情况下,查找所述维度参数所属的第二数据表,其中,所述第二数据表包括由多个子表按照逻辑关系进行表连接所得到的子表的集合,每个所述子表对应一种所述维度参数;

第一筛选单元,用于从所述第一数据表和所述第二数据表中,筛选出满足所述预设过滤条件的维度参数的信息。

7. 根据权利要求6所述的装置,其特征在于,所述第一筛选单元包括:

连接模块,用于将所述第一数据表和所述第二数据表进行表连接,得到目标数据表;

提取模块,用于从所述预设过滤条件提取所述指标参数的参数阈值;

筛选模块,用于在所述目标数据表中,筛选出指标参数大于所述参数阈值的所述维度参数的信息。

8. 根据权利要求7所述的装置,其特征在于,所述连接模块包括:

连接子模块,按照预设连接方式,将所述第一数据表和所述第二数据表进行表连接,

其中,所述预设连接方式包括下述至少之一:内连接、左外连接、右外连接以及全外连接。

9. 根据权利要求6所述的装置,其特征在于,如果在所述第一数据表中查找到所述维度参数,所述装置还包括:

提取单元,用于在所述第一数据表中查找到所述维度参数的情况下,从所述预设过滤条件中提取所述指标参数的参数阈值;

第二筛选单元,用于在所述第一数据表中,筛选出指标参数大于所述参数阈值的所述维度参数的信息。

10. 根据权利要求6至9中任一项所述的装置,其特征在于,所述装置还包括:

聚合单元,用于在筛选出满足所述预设过滤条件的维度参数的信息之后,对筛选出的所述维度参数的信息进行聚合,得到信息集合;

读取单元,用于读取内嵌有所述预设过滤条件的目标过滤条件,其中,所述目标过滤条件为基于所述预设过滤条件确定的对筛选维度进行过滤的条件;

第三筛选单元,用于从所述信息集合中,筛选出符合所述目标过滤条件的筛选维度的信息。

数据的过滤方法和装置

技术领域

[0001] 本申请涉及数据处理领域,具体而言,涉及一种数据的过滤方法和装置。

背景技术

[0002] 在现有技术数据过滤技术中,一般情况下,在一个数据表中按照预定的维度参数筛选满足筛选条件的数据。例如,在订单指标所在的订单表中筛选订单指标大于2的广告,其中,广告即为维度参数。如果订单指标和维度参数(例如,广告)并未存储于同一张数据表中时,通过现有技术便不能实现在订单表中筛选满足过滤条件的订单指标的广告。因此,现有技术在一个数据表中对数据进行过滤的方法,具有一定的局限性,尤其是,当维度参数为多维时,可能会出现维度参数属于不同数据表的现象,此时,需要找到一种合适的方案在上述多个数据表中筛选满足过滤条件的广告。

[0003] 针对现有技术中无法对两个以上的数据表进行指标过滤的问题,目前尚未提出有效的解决方案。

发明内容

[0004] 本申请实施例提供了一种数据的过滤方法和装置,以至少解决现有技术中无法对两个以上的数据表进行指标过滤的技术问题。

[0005] 根据本申请实施例的一个方面,提供了一种数据的过滤方法,该方法包括:获取预设过滤条件中的指标参数和维度参数,其中,预设过滤条件为基于维度参数确定的对指标参数进行过滤的条件;在指标参数所属的第一数据表中查找维度参数;如果在第一数据表中未查找到维度参数,则查找维度参数所属的第二数据表;从第一数据表和第二数据表中,筛选出满足预设过滤条件的维度参数的信息。

[0006] 进一步地,从第一数据表和第二数据表中,筛选出满足预设过滤条件的维度参数的信息包括:将第一数据表和第二数据表进行表连接,得到目标数据表;从预设过滤条件提取指标参数的参数阈值;在目标数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0007] 进一步地,将第一数据表和第二数据表进行表连接包括:按照预设连接方式,将第一数据表和第二数据表进行表连接,其中,预设连接方式包括下述至少之一:内连接、左外连接、右外连接以及全外连接。

[0008] 进一步地,如果在第一数据表中查找到维度参数,该方法还包括:从预设过滤条件中提取指标参数的参数阈值;在第一数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0009] 进一步地,在筛选出满足预设过滤条件的维度参数的信息之后,该方法还包括:对筛选出的维度参数的信息进行聚合,得到信息集合;读取内嵌有预设过滤条件的目标过滤条件,其中,目标过滤条件为基于预设过滤条件确定的对筛选维度进行过滤的条件;从信息集合中,筛选出符合目标过滤条件的筛选维度的信息。

[0010] 根据本申请实施例的另一方面,还提供了一种数据的过滤装置,该装置包括:获取单元,用于获取预设过滤条件中的指标参数和维度参数,其中,预设过滤条件为基于维度参数确定的对指标参数进行过滤的条件;第一查找单元,用于在指标参数所属的第一数据表中查找维度参数;第二查找单元,用于在第一数据表中未查找到维度参数的情况下,查找维度参数所属的第二数据表;第一筛选单元,用于从第一数据表和第二数据表中,筛选出满足预设过滤条件的维度参数的信息。

[0011] 进一步地,第一筛选单元包括:连接模块,用于将第一数据表和第二数据表进行表连接,得到目标数据表;提取模块,用于从预设过滤条件提取指标参数的参数阈值;筛选模块,用于在目标数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0012] 进一步地,连接模块包括:连接子模块,按照预设连接方式,将第一数据表和第二数据表进行表连接,其中,预设连接方式包括下述至少之一:内连接、左外连接、右外连接以及全外连接。

[0013] 进一步地,如果在第一数据表中查找到维度参数,装置还包括:提取单元,用于在第一数据表中查找到维度参数的情况下,从预设过滤条件中提取指标参数的参数阈值;第二筛选单元,用于在第一数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0014] 进一步地,装置还包括:聚合单元,用于在筛选出满足预设过滤条件的维度参数的信息之后,对筛选出的维度参数的信息进行聚合,得到信息集合;读取单元,用于读取内嵌有预设过滤条件的目标过滤条件,其中,目标过滤条件为基于预设过滤条件确定的对筛选维度进行过滤的条件;第三筛选单元,用于从信息集合中,筛选出符合目标过滤条件的筛选维度的信息。

[0015] 在本申请实施例中,采用获取预设过滤条件中的指标参数和维度参数,其中,所述预设过滤条件为基于所述维度参数确定的对所述指标参数进行过滤的条件;在所述指标参数所属的第一数据表中查找所述维度参数;如果在所述第一数据表中未查找到所述维度参数,则查找所述维度参数所属的第二数据表;从所述第一数据表和所述第二数据表中,筛选出满足所述预设过滤条件的维度参数的信息的方式,通过确定指标参数所属的第一数据表中是否包含预设过滤条件中的维度参数,来实现指标的过滤,若查找到指标参数和维度参数不属于第一数据表,则查找维度参数所属的第二数据表,进而根据第一数据表和 second 数据表筛选满足预设过滤条件的维度参数的信息,相对于现有技术中仅能在一个数据表中对指标进行过滤的方法,达到了在多个数据表中对指标进行过滤的目的,从而实现了在多个数据表中按照维度对指标进行过滤的技术效果,进而解决了现有技术中无法对两个以上的数据表进行指标过滤的技术问题。

附图说明

[0016] 此处所说明的附图用来提供对本申请的进一步理解,构成本申请的一部分,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。在附图中:

[0017] 图1是根据本申请实施例的一种数据的过滤方法的流程图;以及

[0018] 图2是根据本申请实施例的一种数据的过滤装置的示意图。

具体实施方式

[0019] 为了使本技术领域的人员更好地理解本申请方案,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分的实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都应当属于本申请保护的范围。

[0020] 需要说明的是,本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0021] 根据本申请实施例,提供了一种数据的过滤方法的实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0022] 图1是根据本申请实施例的一种数据的过滤方法的流程图,如图1所示,该方法包括如下步骤:

[0023] 步骤S102,获取预设过滤条件中的指标参数和维度参数,其中,预设过滤条件为基于维度参数确定的对指标参数进行过滤的条件。

[0024] 例如,指标参数可以为订单数量,维度参数可以为广告,还可以为电商,预设过滤条件可以为在指标参数中筛选订单数量大于或者等于2的广告,还可以为在指标参数中筛选电商A发布的订单数量大于或者等于2的广告。因此,在本申请实施例中,维度参数(例如,广告或者电商)和指标参数(例如,订单数量)构成本申请实施例中的预设过滤条件。

[0025] 步骤S104,在指标参数所属的第一数据表中查找维度参数。

[0026] 需要说明的是,指标参数所属的第一数据表中,可能包含维度参数,可能不包含维度参数。第一数据表中可能包含非预设过滤条件中所包含的维度参数(例如,电商等维度参数)。

[0027] 步骤S106,如果在第一数据表中未查找到维度参数,则查找维度参数所属的第二数据表。

[0028] 若在指标参数所属的第一数据表中不包含维度参数(例如,广告),则查找广告所属的数据表(即,第二数据表),并通过执行下述步骤S108筛选出满足条件的信息。

[0029] 步骤S108,从第一数据表和第二数据表中,筛选出满足预设过滤条件的维度参数的信息。

[0030] 需要说明的是,在本申请实施例中,按照预设过滤条件:订单数量大于2的广告,筛选出的维度参数的信息可以为包含该维度参数的全部信息。例如,订单数量大于或者等于2的广告,以及该广告的下单日期,或者发布该广告的电商等信息。

[0031] 在本申请实施例中,通过确定指标参数所属的第一数据表中是否包含预设过滤条

件中的维度参数,来实现指标的过滤,若查找到指标参数和维度参数不属于第一数据表,则查找维度参数所属的第二数据表,进而根据第一数据表和第二数据表筛选满足预设过滤条件的维度参数的信息,相对于现有技术中仅能在一个数据表中对指标进行过滤的方法,达到了在多个数据表中对指标进行过滤的目的,从而实现了在多个数据表中按照维度对指标进行过滤的技术效果,进而解决了现有技术中无法对两个以上的数据表进行指标过滤的技术问题。

[0032] 在本申请实施例中,第二数据表可以包括多个子表,其中,每个子表对应一个维度,且第二数据表中每个子表之间具有一定的逻辑关系,例如:“逻辑且”、“逻辑或”等逻辑关系。

[0033] 首先,可以按照每个子表之间的逻辑关系(例如,“逻辑且”、“逻辑或”等)将第二数据表中的多个子表进行表连接,连接得到子表的集合之后,将得到的子表的集合与第一数据表进行表连接,得到目标数据表。此时,即可在目标数据表中筛选多维度参数的信息。例如,筛选出订单数量大于2的广告,以及该广告的点击量,其中,广告和点击量为不同的维度,且广告和点击量所属于不同的子表中。

[0034] 可选地,从第一数据表和第二数据表中,筛选出满足预设过滤条件的维度参数的信息包括如下步骤:

[0035] 步骤S1081,将第一数据表和第二数据表进行表连接,得到目标数据表。

[0036] 步骤S1083,从预设过滤条件提取指标参数的参数阈值。

[0037] 步骤S1085,在目标数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0038] 如果在第一数据表中未查找到该维度信息,则查找该维度信息所属的第二数据表,然后,将第一数据表和第二数据表进行表连接,连接之后,得到目标数据表。接下来,在预设过滤条件中提取指标参数的参数阈值,进而,根据该参数阈值在目标数据表中筛选指标参数大于参数阈值的维度参数的信息。

[0039] 需要说明的是,在本申请实施例中,维度参数的种类可以为一种,还可以为多种。

[0040] 若维度参数的种类为一种,且查找到该维度参数所属的数据表为非第一数据表(即,第二数据表),则将第一数据表和第二数据表进行连接,连接之后得到目标数据表,该目标数据表中即包含该维度参数和指标参数,然后,在预设过滤条件中提取指标参数的参数阈值。例如,预设过滤条件为:订单数量大于或者等于2的广告,“订单数量2”即为该指标参数的参数阈值,订单指标所属的第一数据表为订单表,广告所属的第二数据表为会话表。在提取到参数阈值之后,在该目标数据表中,筛选出满足该参数阈值的维度参数的信息,例如,筛选出大于或者等于订单数量为2的维度参数的信息。

[0041] 需要说明的是,在本申请实施例中,第一数据表和第二数据表为相关联的数据表,即第一数据表和第二数据表中包含相同的信息,例如,第一数据表中包含的信息为:用户A购买的订单数量大于2的订单,第二数据表中包含的信息为:用户A根据广告B的购买订单。

[0042] 若维度参数的种类为多种(例如,广告和发布该广告的电商),且查找到该多种维度参数所属的数据表为非第一数据表(即,第二数据表),且广告和发布该广告的电商分属于第二数据表的不同的子表中,则将广告所属的子表和发布该广告的电商所属的子表按照逻辑关系进行表连接,得到子表的集合,然后,将第一数据表与得到的子表的集合按照逻辑关系进行表连接,连接之后得到目标数据表,该目标数据表中即包含该维度参数(例如,广

告和发布该广告的电商)和指标参数。然后,在预设过滤条件中提取指标参数的参数阈值,例如,预设过滤条件为:订单数量大于或者等于2的广告,以及发布该广告的电商,“订单数量2”即为该指标参数的参数阈值。在提取到参数阈值之后,在该目标数据表中,筛选出满足该参数阈值的维度参数的信息,例如,筛选出大于或者等于订单数量为2的广告,以及发布该广告的电商等信息。

[0043] 可选地,将第一数据表和第二数据表进行表连接包括如下步骤:

[0044] 步骤S1,按照预设连接方式,将第一数据表和第二数据表进行表连接,其中,预设连接方式包括下述至少之一:内连接、左外连接、右外连接以及全外连接。

[0045] 将第一数据表和第二数据表按照预设连接方式进行连接的方式可以有很多种,在本申请实施例中,可以通过以下至少一种方式进行第一数据表和第二数据表的连接:内连接、左外连接、右外连接以及全外连接。

[0046] 第一种,内连接方式(即Inner Join):该种连接方式是以上述第一数据表和上述第二数据表中的相同属性条件为基准,连接后的目标数据表中包括相同的属性条件对应的数据信息,即为第一数据表和第二数据表的交集;采用该种连接方式得到的目标数据表的数据容量默认预估变为两个连接的数据表中数据容量较小(即第一数据表)的表的预估行数,也可能出现目标数据表的行数小于第一数据表的情况。

[0047] 假设,预设过滤条件为:订单数量大于2的广告。将广告所属的第二数据表和订单数量所属的第一数据表进行内连接,得到目标数据表之后,在目标数据表中筛选大于参数阈值的维度参数的信息可以用下述代码表示:`select ad from session Inner Join order on...group by ad having count(*)>2`。

[0048] 第二种,左外连接方式(即Left Join):将上述第二数据表作为右表,将上述第一数据表作为左表,采用该种连接方式,以左表的过滤条件为基准,过滤右表中记录的数据,即通过左表的过滤条件实现第一数据表和第二数据表的连接,得到上述目标数据表;在该种连接方式下,两表连接后得到的目标数据表的数据容量为左表的行数。

[0049] 假设,预设过滤条件为:订单数量大于2的广告。将广告所属的第二数据表和订单数量所属的第一数据表进行左外连接,得到目标数据表之后,在目标数据表中筛选大于参数阈值的维度参数的信息可以用下述代码表示:`select ad from session Left Join order on...group by ad having count(*)>2`。

[0050] 第三种,右外连接方式(即Right Join):将上述第一数据表作为右表,将上述第二数据表作为左表,采用该种连接方式,以右表的过滤条件为基准,过滤左表中记录的数据,即通过右表的过滤条件实现第一数据表和第二数据表的连接,得到上述目标数据表;在该种连接方式下,两表连接后得到的目标数据表的数据容量为右表的行数。

[0051] 假设,预设过滤条件为:订单数量大于2的广告。将广告所属的第二数据表和订单数量所属的第一数据表进行右外连接,得到目标数据表之后,在目标数据表中筛选大于参数阈值的维度参数的信息可以用下述代码表示:`select ad from session inner join order on...group by ad having count(*)>2`。

[0052] 第四种,全外连接方式(即Full Join):该种连接方式以上述第一数据表和上述第二数据表中的所有属性条件为基准,连接后的上述目标数据表中包括两个表中所有的属性条件对应的数据信息,即为第一数据表和第二数据表的并集;采用该种连接方式得到的目

标数据表的数据容量默认预估变为两个连接的数据表的行数的总和。

[0053] 假设,预设过滤条件为:订单数量大于2的广告。将广告所属的第二数据表和订单数量所属的第一数据表进行全外连接,得到目标数据表之后,在目标数据表中筛选大于参数阈值的维度参数的信息可以用下述代码表示:`select ad from session Full Join order on...group by ad having count(*)>2`。

[0054] 可选地,如果在第一数据表中查找到维度参数,过滤方法还包括如下步骤:

[0055] 步骤S2,从预设过滤条件中提取指标参数的参数阈值。

[0056] 步骤S3,在第一数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0057] 具体地,如果在第一数据表中查找到该维度参数(例如,广告),则在预设过滤条件中提取指标参数的参数阈值,接下来,根据该参数阈值在第一数据表中筛选指标参数大于参数阈值的维度参数的信息。

[0058] 例如,预设过滤条件为:订单数量大于2的广告,且订单数量和广告所属的数据表均为订单表(即,第一数据表)。进而,通过执行下述代码:`select ad from order group by ad having count(*)>2`,在订单表中筛选出订单数据大于2的广告。

[0059] 可选地,在筛选出满足预设过滤条件的维度参数的信息之后,过滤方法还包括如下步骤:

[0060] 步骤S4,对筛选出的维度参数的信息进行聚合,得到信息集合。

[0061] 步骤S5,读取内嵌有预设过滤条件的目标过滤条件,其中,目标过滤条件为基于预设过滤条件确定的对筛选维度进行过滤的条件。

[0062] 步骤S6,从信息集合中,筛选出符合目标过滤条件的筛选维度的信息。

[0063] 在本申请实施例中,在按照其他过滤条件(例如,目标过滤条件)对指标参数进行进一步筛选时,目标过滤条件中可能内嵌有预设过滤条件。因此,在按照预设过滤条件对指标参数进行过滤之后,可以将过滤出的维度参数的信息进行聚合,得到信息集合。如果确定出目标过滤条件中内嵌有预设过滤条件时,可以在信息集合中筛选满足目标过滤条件的筛选维度的信息。

[0064] 在本申请实施例中,通过将按照预设过滤条件过滤出的维度参数的信息进行聚合,得到信息集合,再按照内嵌有预设过滤条件的目标过滤条件对指标参数进行过滤时,能够快速并准确地过滤出满足目标过滤条件的维度参数的信息。

[0065] 例如,预设过滤条件为订单数量大于2的广告,目标过滤条件为电商A发布的订单数量大于2的广告,其中,“订单数量大于2的广告”内嵌于“电商A发布的订单数量大于2的广告”中。在按照预设过滤条件在第一数据表或者目标数据表中筛选出订单数量大于2的广告的全部信息之后,对筛选得到的广告的全部信息进行聚合,得到信息集合。然后,通过判断可知,目标过滤条件:电商A发布的订单数量大于2的广告中内嵌有预设过滤条件“订单数量大于2的广告”,此时,根据该目标过滤条件在信息集合中筛选电商A发布的订单数量大于2的广告。

[0066] 本申请实施例还提供了一种数据的过滤装置,该装置可以用于执行本申请实施例的数据的过滤方法。

[0067] 图2是根据本申请实施例的一种数据的过滤装置的示意图,如图2所示,该装置包括获取单元21、第一查找单元22、第二查找单元23和第一筛选单元24,其中:

[0068] 获取单元21,用于获取预设过滤条件中的指标参数和维度参数,其中,预设过滤条件为基于维度参数确定的对指标参数进行过滤的条件。

[0069] 例如,指标参数可以为订单数量,维度参数可以为广告,还可以为电商,预设过滤条件可以为在指标参数中筛选订单数量大于或者等于2的广告,还可以为在指标参数中筛选电商A发布的订单数量大于或者等于2的广告。因此,在本申请实施例中,维度参数(例如,广告或者电商)和指标参数(例如,订单数量)构成本申请实施例中的预设过滤条件。

[0070] 第一查找单元22,用于在指标参数所属的第一数据表中查找维度参数。

[0071] 需要说明的是,指标参数所属的第一数据表中,可能包含维度参数,可能不包含维度参数。第一数据表中可能包含非预设过滤条件中所包含的维度参数(例如,电商等维度参数)。

[0072] 第二查找单元23,用于在第一数据表中未查找到维度参数的情况下,查找维度参数所属的第二数据表。

[0073] 若在指标参数所属的第一数据表中不包含维度参数(例如,广告),则查找广告所属的数据表(即,第二数据表)。并通过调用第一筛选单元筛选出满足条件的信息。

[0074] 第一筛选单元24,用于从第一数据表和第二数据表中,筛选出满足预设过滤条件的维度参数的信息。

[0075] 需要说明的是,在本申请实施例中,按照预设过滤条件:订单数量大于2的广告,筛选出的维度参数的信息可以为包含该维度参数的全部信息。例如,订单数量大于或者等于2的广告,以及该广告的下单日期,或者发布该广告的电商等信息。

[0076] 在本申请实施例中,通过确定指标参数所属的第一数据表中是否包含预设过滤条件中的维度参数,来实现指标的过滤,若查找到指标参数和维度参数不属于第一数据表,则查找维度参数所属的第二数据表,进而根据第一数据表和第二数据表筛选满足预设过滤条件的维度参数的信息,相对于现有技术中仅能在一个数据表中对指标进行过滤的方法,达到了在多个数据表中对指标进行过滤的目的,从而实现了在多个数据表中按照维度对指标进行过滤的技术效果,进而解决了现有技术中无法对两个以上的数据表进行指标过滤的技术问题。

[0077] 在本申请实施例中,第二数据表可以包括多个子表,其中,每个子表对应一个维度,且第二数据表中每个子表之间具有一定的逻辑关系,例如:“逻辑且”、“逻辑或”等逻辑关系。

[0078] 首先,可以按照每个子表之间的逻辑关系(例如,“逻辑且”、“逻辑或”等)将第二数据表中的多个子表进行表连接,连接得到子表的集合之后,将得到的子表的集合与第一数据表进行表连接,得到目标数据表。此时,即可在目标数据表中筛选多维度参数的信息。例如,筛选出订单数量大于2的广告,以及该广告的点击量,其中,广告和点击量为不同的维度,且广告和点击量所属于不同的子表中。

[0079] 可选地,第一筛选单元包括:连接模块,用于将第一数据表和第二数据表进行表连接,得到目标数据表;提取模块,用于从预设过滤条件提取指标参数的参数阈值;筛选模块,用于在目标数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0080] 具体地,如果在第一数据表中未查找到该维度信息,则查找该维度信息所属的第二数据表,然后,通过调用连接模块将第一数据表和第二数据表进行表连接,连接之后,得

到目标数据表。接下来,通过调用提取单元在预设过滤条件中提取指标参数的参数阈值,进而,根据该参数阈值在目标数据表中筛选指标参数大于参数阈值的维度参数的信息。

[0081] 可选地,连接模块包括:连接子模块,按照预设连接方式,将第一数据表和第二数据表进行表连接,其中,预设连接方式包括下述至少之一:内连接、左外连接、右外连接以及全外连接。

[0082] 具体地,将第一数据表和第二数据表可以通过调用连接子模块按照预设连接方式进行连接第一数据表和第二数据表,其中,预设连接方式可以有很多种,在本申请实施例中,可以通过以下至少一种方式进行第一数据表和第二数据表的连接:内连接、左外连接、右外连接以及全外连接。

[0083] 第一种,内连接方式(即Inner Join):该种连接方式是以上述第一数据表和上述第二数据表中的相同属性条件为基准,连接后的目标数据表中包括相同的属性条件对应的数据信息,即为第一数据表和第二数据表的交集;采用该种连接方式得到的目标数据表的数据容量默认预估变为两个连接的数据表中数据容量较小(即第一数据表)的表的预估行数,也可能出现目标数据表的行数小于第一数据表的情况。

[0084] 第二种,左外连接方式(即Left Join):将上述第二数据表作为右表,将上述第一数据表作为左表,采用该种连接方式,以左表的过滤条件为基准,过滤右表中记录的数据,即通过左表的过滤条件实现第一数据表和第二数据表的连接,得到上述目标数据表;在该种连接方式下,两表连接后得到的目标数据表的数据容量为左表的行数。

[0085] 第三种,右外连接方式(即Right Join):将上述第一数据表作为右表,将上述第二数据表作为左表,采用该种连接方式,以右表的过滤条件为基准,过滤左表中记录的数据,即通过右表的过滤条件实现第一数据表和第二数据表的连接,得到上述目标数据表;在该种连接方式下,两表连接后得到的目标数据表的数据容量为右表的行数。

[0086] 第四种,全外连接方式(即Full Join):该种连接方式以上述第一数据表和上述第二数据表中的所有属性条件为基准,连接后的目标数据表中包括两个表中所有的属性条件对应的数据信息,即为第一数据表和第二数据表的并集;采用该种连接方式得到的目标数据表的数据容量默认预估变为两个连接的数据表的行数的总和。

[0087] 可选地,如果在第一数据表中查找到维度参数,过滤装置还包括:提取单元,用于在第一数据表中查找到维度参数的情况下,从预设过滤条件中提取指标参数的参数阈值;第二筛选单元,用于在第一数据表中,筛选出指标参数大于参数阈值的维度参数的信息。

[0088] 具体地,如果在第一数据表中查找到该维度参数(例如,广告),则通过调用提取单元在预设过滤条件中提取指标参数的参数阈值,接下来,通过调用第二筛选单元根据该参数阈值在第一数据表中筛选指标参数大于参数阈值的维度参数的信息。

[0089] 可选地,过滤装置还包括:聚合单元,用于在筛选出满足预设过滤条件的维度参数的信息之后,对筛选出的维度参数的信息进行聚合,得到信息集合;读取单元,用于读取内嵌有预设过滤条件的目标过滤条件,其中,目标过滤条件为基于预设过滤条件确定的对筛选维度进行过滤的条件;第三筛选单元,用于从信息集合中,筛选出符合目标过滤条件的筛选维度的信息。

[0090] 在本申请实施例中,在按照其他过滤条件(例如,目标过滤条件)对指标参数进行进一步筛选时,目标过滤条件中可能内嵌有预设过滤条件。因此,在按照预设过滤条件对指

标参数进行过滤之后,可以通过调用聚合单元将过滤出的维度参数的信息进行聚合,得到信息集合。如果通过读取单元读取到目标过滤条件中内嵌有预设过滤条件时,通过调用第三筛选单元在信息集合中筛选满足目标过滤条件的筛选维度的信息。通过将按照预设过滤条件过滤出的维度参数的信息进行聚合,得到信息集合,再按照内嵌有预设过滤条件的目标过滤条件对指标参数进行过滤时,能够快速并准确地过滤出满足目标过滤条件的维度参数的信息。

[0091] 所述数据的过滤装置包括处理器和存储器,上述获取单元、第一查找单元、第二查找单元和第一筛选单元等均作为程序单元存储在存储器中,由处理器执行存储在存储器中的上述程序单元。

[0092] 处理器中包含内核,由内核去存储器中调取相应的程序单元。内核可以设置一个或以上,通过调整内核参数在多个数据表中对指标进行过滤的目的,进而解决了现有技术中无法对两个以上的数据表进行指标过滤的技术问题。

[0093] 存储器可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM),存储器包括至少一个存储芯片。

[0094] 本申请还提供了一种计算机程序产品的实施例,当在数据处理设备上执行时,适于执行初始化有如下方法步骤的程序代码:获取预设过滤条件中的指标参数和维度参数,其中,所述预设过滤条件为基于所述维度参数确定的对所述指标参数进行过滤的条件;在所述指标参数所属的第一数据表中查找所述维度参数;如果在所述第一数据表中未查找到所述维度参数,则查找所述维度参数所属的第二数据表;从所述第一数据表和所述第二数据表中,筛选出满足所述预设过滤条件的维度参数的信息。

[0095] 上述本申请实施例序号仅仅为了描述,不代表实施例的优劣。

[0096] 在本申请的上述实施例中,对各个实施例的描述都各有侧重,某个实施例中沒有详述的部分,可以参见其他实施例的相关描述。

[0097] 在本申请所提供的几个实施例中,应该理解到,所揭露的技术内容,可通过其它的方式实现。其中,以上所描述的装置实施例仅仅是示意性的,例如所述单元的划分,可以为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,单元或模块的间接耦合或通信连接,可以是电性或其它的形式。

[0098] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0099] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0100] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式

体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可为个人计算机、服务器或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0101] 以上所述仅是本申请的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本申请原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本申请的保护范围。

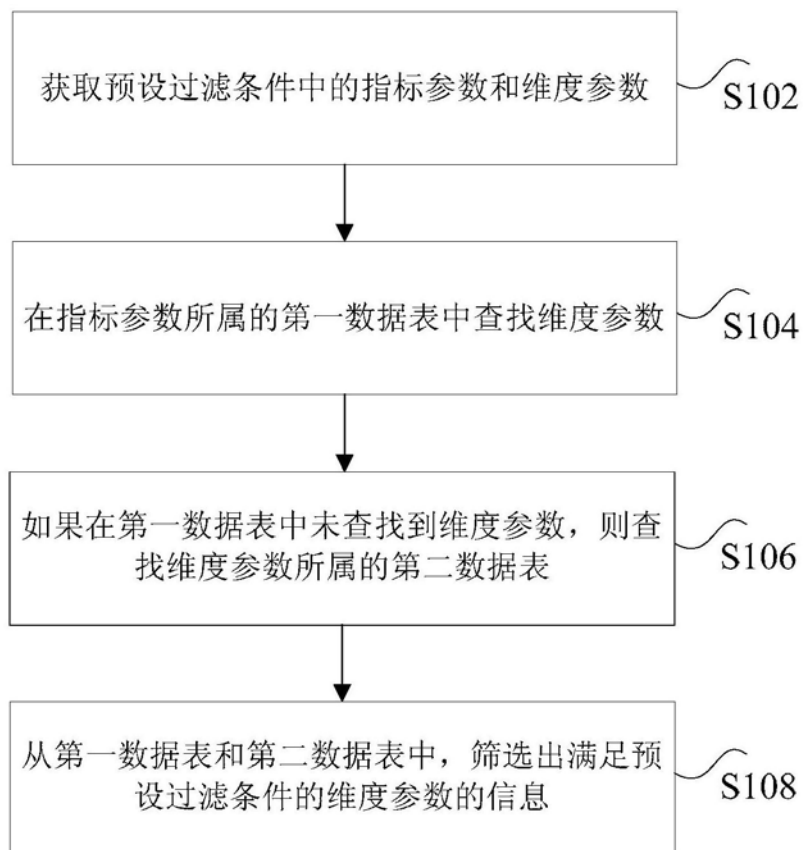


图1



图2