



(12) 发明专利

(10) 授权公告号 CN 111008521 B

(45) 授权公告日 2023.04.28

(21) 申请号 201911245045.4

(22) 申请日 2019.12.06

(65) 同一申请的已公布的文献号

申请公布号 CN 111008521 A

(43) 申请公布日 2020.04.14

(73) 专利权人 北京三快在线科技有限公司

地址 100080 北京市海淀区北四环西路9号  
2106-030

(72) 发明人 黄晓阳 崔洪振 赵楠

(74) 专利代理机构 北京三高永信知识产权代理

有限责任公司 11138

专利代理师 谢冬寒

(51) Int. Cl.

G06F 40/183 (2020.01)

(56) 对比文件

CN 108228817 A, 2018.06.29

CN 110287192 A, 2019.09.27

CN 110019397 A, 2019.07.16

CN 110399380 A, 2019.11.01

US 2014280374 A1, 2014.09.18

审查员 付苗

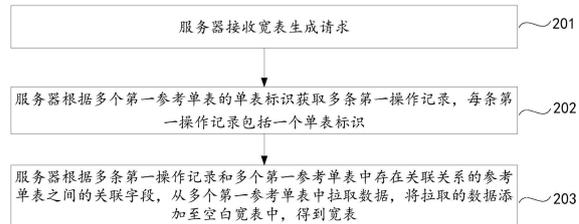
权利要求书4页 说明书15页 附图3页

(54) 发明名称

生成宽表的方法、装置及计算机存储介质

(57) 摘要

本申请公开了一种实时宽表生成方法、装置及计算机存储介质,属于计算机技术领域。该方法包括:接收携带多个第一参考单表的标识以及多个第一参考单表中存在关联关系的参考单表之间的关联字段的宽表生成请求。根据多个第一参考单表的单表标识获取多条第一操作记录。根据多条第一操作记录和多个第一参考单表中存在关联关系的参考单表之间的关联字段,得到宽表。由于该方法是根据各个单表之间的关联字段拉取数据,并将拉取的数据添加至空白宽表中得到的,因此无需预先配置宽表的具体结构,只需用户设置用于构建宽表的参考单表和各个参考单表之间的关联字段,即可通过本申请提供的方法生成宽表,简化了用户操作,从而提高了获取宽表的灵活性。



1. 一种生成宽表的方法,其特征在于,所述方法包括:

接收宽表生成请求,所述宽表生成请求携带多个第一参考单表的标识、以及所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,每个第一参考单表用于存储一种类型的数据;

根据所述多个第一参考单表的单表标识获取多条第一操作记录,每条第一操作记录包括一个单表标识,每条第一操作记录用于指示对包括的单表标识对应的参考单表执行的操作;

根据所述多条第一操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,从所述多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到所述宽表。

2. 如权利要求1所述的方法,其特征在于,所述多条第一操作记录包括数据插入操作记录;

所述根据所述多条第一操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,从所述多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,包括:

根据所述多个第一参考单表构建关联关系树,所述关联关系树中的每个节点指示一个第一参考单表,所述关联关系树中相邻的两个节点用于指示存在关联关系的两个第一参考单表;

从所述多条第一操作记录包括的数据插入操作记录中,确定所述关联关系树中的每个节点对应的数据插入操作记录;

根据所述关联关系树中每个节点对应的数据插入操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,按照所述关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

3. 如权利要求2所述的方法,其特征在于,所述多条第一操作记录还包括数据更新操作记录;

所述依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中之后,还包括:

对于所述多条第一操作记录中的每条数据更新操作记录,基于每条数据更新操作记录,从相应数据更新操作记录包括的单表标识对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

4. 如权利要求2所述的方法,其特征在于,每个第一参考单表包括多个表项;

所述根据所述关联关系树中每个节点对应的数据插入操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,按照所述关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中,包括:

对于所述关联关系树中的第*i*个节点,当所述第*i*个节点为叶子节点时,根据所述第*i*个节点对应的数据插入操作记录,从所述第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中;

当所述第*i*个节点不是叶子节点时,确定所述第*i*个节点对应的第一参考单表与第*i-1*个节点对应的第一参考单表之间的关联字段,所述第*i*个节点为所述第*i-1*个节点的父节点;

确定所述第*i-1*个节点对应的第一参考单表中包括所述关联字段的表项;

根据确定的表项在所述空白宽表中添加空白表项,所述空白表项的表项标识与所述确定的表项的表项标识相同;

根据所述第*i*个节点对应的数据插入操作记录,从与所述第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白表项中。

5.如权利要求1所述的方法,其特征在于,每条操作记录还包括操作执行时间和表项标识,所述表项标识用于指示相应操作记录包括的单表标识对应的参考单表中的一个表项;

所述根据所述多个第一参考单表的单表标识获取多条第一操作记录,包括:

从存储的操作记录中获取包括的单表标识与所述多个第一参考单表中任一个参考单表的标识一致的操作记录;

按照每条操作记录包括的单表标识和表项标识,对获取的操作记录进行分类,得到多个操作记录集合,每个操作记录集合对应一个单表标识和一个表项标识;

获取每个操作记录集合中操作执行时间距离当前时间最近的操作记录,得到所述多条第一操作记录。

6.如权利要求1所述的方法,其特征在于,每条操作记录还包括操作执行时间;

所述将拉取的数据添加至空白宽表中,得到所述宽表之后,还包括:

接收宽表更新请求,所述宽表更新请求携带一个或多个第二参考单表的标识,每个第二参考单表用于存储一种类型的数据;

根据所述一个或多个第二参考单表的标识,获取操作执行时间位于当前时间和当前时间之前最近一次更新所述宽表的时间之间的多条第二操作记录;

根据所述多条第二操作记录从所述一个或多个第二参考单表中拉取数据,将拉取的数据添加至所述宽表中。

7.如权利要求1至6任一所述的方法,其特征在于,所述方法还包括:

接收携带单表标识、表项标识和数据处理操作指示信息的日志消息;

根据所述日志消息中携带的单表标识和表项标识,从存储的多个单表中获取当前待修改的表项,所述多个单表包括所述多个第一参考单表以及一个或多个第二参考单表,每个单表包括多个表项;

根据所述数据处理操作指示信息,修改获取的表项;

根据修改结果生成一条操作记录。

8.一种生成宽表的装置,其特征在于,所述装置包括:

第一接收模块,用于接收宽表生成请求,所述宽表生成请求携带多个第一参考单表的标识、以及所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,每个第一参考单表用于存储一种类型的数据;

第一获取模块,用于根据所述多个第一参考单表的单表标识获取多条第一操作记录,每条第一操作记录包括一个单表标识,每条第一操作记录用于指示对包括的单表标识对应的参考单表执行的操作;

第一拉取模块,用于根据所述多条第一操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,从所述多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到所述宽表。

9. 如权利要求8所述的装置,其特征在于,所述多条第一操作记录包括数据插入操作记录;

所述拉取模块包括:

构建子模块,用于根据所述多个第一参考单表构建关联关系树,所述关联关系树中的每个节点指示一个第一参考单表,所述关联关系树中相邻的两个节点用于指示存在关联关系的两个第一参考单表;

确定子模块,用于从所述多条第一操作记录包括的数据插入操作记录中,确定所述关联关系树中的每个节点对应的数据插入操作记录;

第一拉取子模块,用于根据所述关联关系树中每个节点对应的数据插入操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,按照所述关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

10. 如权利要求9所述的装置,其特征在于,所述多条第一操作记录还包括数据更新操作记录;所述拉取模块还包括:

第二拉取子模块,用于对于所述多条第一操作记录中的每条数据更新操作记录,基于每条数据更新操作记录,从相应数据更新操作记录包括的单表标识对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

11. 如权利要求9所述的装置,其特征在于,每个第一参考单表包括多个表项;

所述第一拉取子模块包括:

第一拉取单元,用于对于所述关联关系树中的第*i*个节点,当所述第*i*个节点为叶子节点时,根据所述第*i*个节点对应的数据插入操作记录,从所述第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中;

第一确定单元,用于当所述第*i*个节点不是叶子节点时,确定所述第*i*个节点对应的第一参考单表与第*i*-1个节点对应的第一参考单表之间的关联字段,所述第*i*个节点为所述第*i*-1个节点的父节点;

第二确定单元,用于确定所述第*i*-1个节点对应的第一参考单表中包括所述关联字段的表项;

添加单元,用于根据确定的表项在所述空白宽表中添加空白表项,所述空白表项的表项标识与所述确定的表项的表项标识相同;

第二拉取单元,用于根据所述第*i*个节点对应的数据插入操作记录,从与所述第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白表项中。

12. 如权利要求8所述的装置,其特征在于,每条操作记录还包括操作执行时间和表项标识,所述表项标识用于指示相应操作记录包括的单表标识对应的参考单表中的一个表项;

所述获取模块包括:

第一获取子模块,用于从存储的操作记录中获取包括的单表标识与所述多个第一参考

单表中任一参考单表的标识一致的操作记录；

分类子模块,用于按照每条操作记录包括的单表标识和表项标识,对获取的操作记录进行分类,得到多个操作记录集合,每个操作记录集合对应一个单表标识和一个表项标识;

第二获取子模块,用于获取每个操作记录集合中操作执行时间距离当前时间最近的操作记录,得到所述多条第一操作记录。

13.如权利要求8所述的装置,其特征在于,每条操作记录还包括操作执行时间;

所述装置还包括:

第二接收模块,用于接收宽表更新请求,所述宽表更新请求携带一个或多个第二参考单表的标识,每个第二参考单表用于存储一种类型的数据;

第二获取模块,用于根据所述一个或多个第二参考单表的标识,获取操作执行时间位于当前时间和当前时间之前最近一次更新所述宽表的时间之间的多条第二操作记录;

第二拉取模块,用于根据所述多条第二操作记录从所述一个或多个第二参考单表中拉取数据,将拉取的数据添加至所述宽表中。

14.如权利要求8至13任一所述的装置,其特征在于,所述装置还包括:

第三接收模块,用于接收携带单表标识、表项标识和数据处理操作指示信息的日志消息;

第三获取模块,用于根据所述日志消息中携带的单表标识和表项标识,从存储的多个单表中获取当前待修改的表项,所述多个单表包括所述多个第一参考单表以及一个或多个第二参考单表,每个单表包括多个表项;

修改模块,用于根据所述数据处理操作指示信息,修改获取的表项;

生成模块,用于根据修改结果生成一条操作记录。

15.一种确定宽表的装置,其特征在于,所述装置包括:

处理器;

用于存储处理器可执行指令的存储器;

其中,所述处理器被配置为执行上述权利要求1至权利要求7中的任一项权利要求所述的方法的步骤。

16.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有指令,所述指令被处理器执行时实现上述权利要求1至权利要求7中的任一项权利要求所述的方法的步骤。

## 生成宽表的方法、装置及计算机存储介质

### 技术领域

[0001] 本申请涉及数据处理技术领域,特别涉及一种生成宽表的方法、装置及计算机存储介质。

### 背景技术

[0002] 目前,在大数据技术领域,通常需要对在线生成的海量数据进行统计,以便根据统计的数据执行后续操作。比如,双十一期间需要统计诸如产品信息、订单信息等数据,以便于根据统计出的数据及时调整后续活动策略。目前在线生成的数据的类型众多,如果将不同类型的数据存储在不同的表中,将不利于用户快速查询不同类型的数据,因此,可以将不同类型的数据存储在一张表中,此时,该表可以称为宽表。

[0003] 相关技术中,宽表中存储的数据的类型是服务器预先配置的,导致后续用户获取的宽表中的数据类型比较单一。

### 发明内容

[0004] 本申请实施例提供了一种生成宽表的方法、装置及计算机存储介质,可以提高获取宽表的灵活性。技术方案如下:

[0005] 一方面,提供了一种生成宽表的方法,方法包括:

[0006] 接收宽表生成请求,所述宽表生成请求携带多个第一参考单表的标识、以及所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,每个第一参考单表用于存储一种类型的数据;

[0007] 根据所述多个第一参考单表的单表标识获取多条第一操作记录,每条第一操作记录包括一个单表标识,每条第一操作记录用于指示对包括的单表标识对应的参考单表执行的操作;

[0008] 根据所述多条第一操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,从所述多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到所述宽表。

[0009] 可选的,所述多条第一操作记录包括数据插入操作记录;

[0010] 所述根据所述多条第一操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,从所述多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,包括:

[0011] 根据所述多个第一参考单表构建关联关系树,所述关联关系树中的每个节点指示一个第一参考单表,所述关联关系树中相邻的两个节点用于指示存在关联关系的两个第一参考单表;

[0012] 从所述多条第一操作记录包括的数据插入操作记录中,确定所述关联关系树中的每个节点对应的数据插入操作记录;

[0013] 根据所述关联关系树中每个节点对应的数据插入操作记录和所述多个第一参考

单表中存在关联关系的参考单表之间的关联字段,按照所述关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

[0014] 可选的,所述多条第一操作记录还包括数据更新操作记录;

[0015] 所述依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中之后,还包括:

[0016] 对于所述多条第一操作记录中的每条数据更新操作记录,基于每条数据更新操作记录,从相应数据更新操作记录包括的单表标识对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

[0017] 可选的,每个第一参考单表包括多个表项;

[0018] 所述根据所述关联关系树中每个节点对应的数据插入操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,按照所述关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中,包括:

[0019] 对于所述关联关系树中的第*i*个节点,当所述第*i*个节点为叶子节点时,根据所述第*i*个节点对应的数据插入操作记录,从所述第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中;

[0020] 当所述第*i*个节点不是叶子节点时,确定所述第*i*个节点对应的第一参考单表与第*i*-1个节点对应的第一参考单表之间的关联字段,所述第*i*个节点为所述第*i*-1个节点的父节点;

[0021] 确定所述第*i*-1个节点对应的第一参考单表中包括所述关联字段的表项;

[0022] 根据确定的表项在所述空白宽表中添加空白表项,所述空白表项的表项标识与所述确定的表项的表项标识相同;

[0023] 根据所述第*i*个节点对应的数据插入操作记录,从与所述第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白表项中。

[0024] 可选的,每条操作记录还包括操作执行时间和表项标识,所述表项标识用于指示相应操作记录包括的单表标识对应的参考单表中的一个表项;

[0025] 所述根据所述多个第一参考单表的单表标识获取多条第一操作记录,包括:

[0026] 从存储的操作记录中获取包括的单表标识与所述多个第一参考单表中任一参考单表的标识一致的操作记录;

[0027] 按照每条操作记录包括的单表标识和表项标识,对获取的操作记录进行分类,得到多个操作记录集合,每个操作记录集合对应一个单表标识和一个表项标识;

[0028] 获取每个操作记录集合中操作执行时间距离当前时间最近的操作记录,得到所述多条第一操作记录。

[0029] 可选的,所述将拉取的数据添加至空白宽表中,得到所述宽表之后,还包括:

[0030] 接收宽表更新请求,所述宽表更新请求携带一个或多个第二参考单表的标识,每个第二参考单表用于存储一种类型的数据;

[0031] 根据所述一个或多个第二参考单表的标识,获取操作执行时间位于当前时间和当前时间之前最近一次更新所述宽表的时间之间的多条第二操作记录;

[0032] 根据所述多条第二操作记录从所述一个或多个第二参考单表中拉取数据,将拉取的数据添加至所述宽表中。

[0033] 可选的,所述方法还包括:

[0034] 接收携带单表标识、表项标识和数据处理操作指示信息的日志消息;

[0035] 根据所述日志消息中携带的单表标识和表项标识,从存储的多个单表中获取当前待修改的表项,所述多个单表包括所述多个第一参考单表以及一个或多个第二参考单表,每个单表包括多个表项;

[0036] 根据所述数据处理操作指示信息,修改获取的表项;

[0037] 根据修改结果生成一条操作记录。

[0038] 另一方面,提供了一种生成宽表的装置,装置包括:

[0039] 第一接收模块,用于接收宽表生成请求,所述宽表生成请求携带多个第一参考单表的标识、以及所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,每个第一参考单表用于存储一种类型的数据;

[0040] 第一获取模块,用于根据所述多个第一参考单表的单表标识获取多条第一操作记录,每条第一操作记录包括一个单表标识,每条第一操作记录用于指示对包括的单表标识对应的参考单表执行的操作;

[0041] 第一拉取模块,用于根据所述多条第一操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,从所述多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到所述宽表。

[0042] 可选的,所述多条第一操作记录包括数据插入操作记录;

[0043] 所述拉取模块包括:

[0044] 构建子模块,用于根据所述多个第一参考单表构建关联关系树,所述关联关系树中的每个节点指示一个第一参考单表,所述关联关系树中相邻的两个节点用于指示存在关联关系的两个第一参考单表;

[0045] 确定子模块,用于从所述多条第一操作记录包括的数据插入操作记录中,确定所述关联关系树中的每个节点对应的数据插入操作记录;

[0046] 第一拉取子模块,用于根据所述关联关系树中每个节点对应的数据插入操作记录和所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,按照所述关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

[0047] 可选的,所述多条第一操作记录还包括数据更新操作记录;所述拉取模块还包括:

[0048] 第二拉取子模块,用于对于所述多条第一操作记录中的每条数据更新操作记录,基于每条数据更新操作记录,从相应数据更新操作记录包括的单表标识对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中。

[0049] 可选的,所述第一拉取子模块包括:

[0050] 第一拉取单元,用于对于所述关联关系树中的第*i*个节点,当所述第*i*个节点为叶子节点时,根据所述第*i*个节点对应的数据插入操作记录,从所述第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白宽表中;

[0051] 第一确定单元,用于当所述第*i*个节点不是叶子节点时,确定所述第*i*个节点对应

的第一参考单表与第i-1个节点对应的第一参考单表之间的关联字段,所述第i个节点为所述第i-1个节点的父节点;

[0052] 第二确定单元,用于确定所述第i-1个节点对应的第一参考单表中包括所述关联字段的表项;

[0053] 添加单元,用于根据确定的表项在所述空白宽表中添加空白表项,所述空白表项的表项标识与所述确定的表项的表项标识相同;

[0054] 第二拉取单元,用于根据所述第i个节点对应的数据插入操作记录,从与所述第i个节点对应的第一参考单表中拉取数据,将拉取的数据添加至所述空白表项中。

[0055] 可选的,所述获取模块包括:

[0056] 第一获取子模块,用于从存储的操作记录中获取包括的单表标识与所述多个第一参考单表中任一个参考单表的标识一致的操作记录;

[0057] 分类子模块,用于按照每条操作记录包括的单表标识和表项标识,对获取的操作记录进行分类,得到多个操作记录集合,每个操作记录集合对应一个单表标识和一个表项标识;

[0058] 第二获取子模块,用于获取每个操作记录集合中操作执行时间距离当前时间最近的操作记录,得到所述多条第一操作记录。

[0059] 可选的,所述装置还包括:

[0060] 第二接收模块,用于接收宽表更新请求,所述宽表更新请求携带一个或多个第二参考单表的标识,每个第二参考单表用于存储一种类型的数据;

[0061] 第二获取模块,用于根据所述一个或多个第二参考单表的标识,获取操作执行时间位于当前时间和当前时间之前最近一次更新所述宽表的时间之间的多条第二操作记录;

[0062] 第二拉取模块,用于根据所述多条第二操作记录从所述一个或多个第二参考单表中拉取数据,将拉取的数据添加至所述宽表中。

[0063] 可选的,所述装置还包括:

[0064] 第三接收模块,用于接收携带单表标识、表项标识和数据处理操作指示信息的日志消息;

[0065] 第三获取模块,用于根据所述日志消息中携带的单表标识和表项标识,从存储的多个单表中获取当前待修改的表项,所述多个单表包括所述多个第一参考单表以及一个或多个第二参考单表,每个单表包括多个表项;

[0066] 修改模块,用于根据所述数据处理操作指示信息,修改获取的表项;

[0067] 生成模块,用于根据修改结果生成一条操作记录。

[0068] 另一方面,提供了一种生成宽表的装置,所述生成宽表的装置包括处理器、通信接口、存储器和通信总线;

[0069] 其中,所述处理器、所述通信接口和所述存储器通过所述通信总线完成相互间的通信;

[0070] 所述存储器用于存放计算机程序;

[0071] 所述处理器用于执行所述存储器上所存放的程序,以实现前述提供的生成宽表的方法。

[0072] 另一方面,提供了一种计算机可读存储介质,所述存储介质内存储有计算机程序,

所述计算机程序被处理器执行时实现前述提供的生成宽表的方法的步骤。

[0073] 本申请实施例提供的技术方案带来的有益效果至少包括：

[0074] 通过接收宽表生成请求,该宽表生成请求携带多个第一参考单表的标识以及多个第一参考单表中存在关联关系的参考单表之间的关联字段。再根据多个第一参考单表的单表标识获取多条第一操作记录。最后根据多条第一操作记录和多个第一参考单表中存在关联关系的参考单表之间的关联字段,从多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到宽表。由于该确定宽表的方法是根据各个单表之间的关联字段拉取数据,并将拉取的数据添加至空白宽表中得到的,因此而无需预先配置宽表的具体结构,只需用户设置用于构建宽表的参考单表和各个参考单表之间的关联字段,即可通过本申请提供的方法生成宽表,简化了用户操作,从而提高了获取宽表的灵活性。

### 附图说明

[0075] 为了更清楚地说明本申请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地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,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的附图。

[0076] 图1是本申请实施例提供的一种确定宽表的系统架构图。

[0077] 图2是本申请实施例提供的一种生成宽表的方法流程图。

[0078] 图3是本申请实施例提供的一种关联关系树示意图。

[0079] 图4是本申请实施例提供的一种生成宽表的装置的结构示意图。

[0080] 图5是本申请实施例提供的一种服务器结构示意图。

### 具体实施方式

[0081] 为使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请实施方式作进一步地详细描述。

[0082] 在对本申请实施例进行详细的解释说明之前,先对本申请实施例涉及的应用场景予以介绍。

[0083] 在大数据技术领域,通常需要对在线生成的海量数据进行统计,以便根据统计的数据执行后续操作。比如,在一种可能的场景下,可以将不同专业的描述信息放置在一个单表中。针对每个专业,将该专业的所有学生的学号信息放置在一个单表。针对每个学生,将该学生的不同课程的成绩信息放置在一张单表中。此时,不同类型的数据是放置在不同的单表中的。目前,为了提高用户查询数据的效率,可以将不同类型的数据存储在一张表中,此时,该表可以称为宽表。比如,可以将放置在前述各个单表中的数据均放置在一张宽表中,此时用户只需要通过一张宽表便可以获取到不同类型的数据。

[0084] 在对本申请实施例进行详细的解释说明之前,先对本申请实施例涉及的系统架构予以介绍。

[0085] 图1是本申请实施例提供的一种确定宽表的系统架构图。如图1所示,该系统中包括部署有生产数据库的数据处理中心101、单表计算引擎102、单表存储中心103、宽表计算引擎104、宽表存储中心105以及缓存中心106。

[0086] 其中,数据处理中心101、单表计算引擎102、单表存储中心103、宽表计算引擎104、宽表存储中心105以及缓存中心106,可以部署于同一服务器中,也可以分别部署于不同的服务器。比如,在一种可能的实现方式中,数据处理中心101部署于第一服务器、单表计算引擎102和单表存储中心103部署于第二服务器、宽表计算引擎104、宽表存储中心105以及缓存中心106三者部署于第三服务器。此时,第一服务器与第二服务器之间通过无线或有线方式连接以进行通信、第二服务器与第三服务器之间通过无线或有线方式连接以进行通信。

[0087] 数据处理中心101用于线上实时生成数据,并针对生成的数据生成日志消息(binlog),然后将日志消息通过消息中间件(kafka)传输至单表计算引擎102。单表计算引擎102用于根据日志消息将实时生成的数据添加至单表中,并生成操作记录,将生成的操作记录通过消息中间件(MQ)传输至宽表计算引擎104,此时相当于对日志消息进行了二次分发,通过MQD的方式可以保证当去查找单表时,最新的这条操作记录对应的数据已经落到相应单表中了。宽表计算引擎104用于根据操作记录将添加至单表中的数据也添加至宽表之中。单表存储中心103用于存储每张单表包括的数据,宽表存储中心105用于存储宽表包括的数据。缓存中心106用于存储操作记录。

[0088] 上述消息中间件(kafka)主要存储日志消息,这些日志消息包括来源于线上生产数据库的binlog消息。以生产数据库为mysql为例,binlog消息主要包括以下内容:a、mid这条消息的唯一的id(标识);b、opt操作运算符,主要为INSERT(用于指示插入操作)、UPDATE(用于指示更新操作)、DELETE(用于指示删除操作);c、src,用于指示变更之前的这条数据记录;d、cur,用于指示本次操作引起的哪些字段变更,如INSERT操作时将存这条记录的全量字段,UPDATE操作时将存这条记录的增量变化字段。

[0089] 此外,宽表计算引擎104还用于接收宽表生成请求,该宽表生成请求携带多个第一参考单表的标识、以及所述多个第一参考单表中存在关联关系的参考单表之间的关联字段,每个第一参考单表用于存储一种类型的数据。此时,缓存中心106还用于存储各个参考单表之间的关联字段,以便于宽表计算引擎104通过本申请实施例提供的生成宽表的方法来确定宽表。

[0090] 比如,生产数据库可以为Mysql(瑞典MySQL公司开发的一种关系型数据库)、Oracle(甲骨文公司开发的一种关系数据库)、Sql server(微软公司开发的一种关系数据库),本申请实施例并不限定数据库的具体类型。

[0091] 此外,上述单表存储中心103、宽表存储中心105以及缓存中心106可以为ES(elasticsearch,一种分布式多用户能力的全文搜索引擎)集群。将ES集群作为数据落地的存储引擎,以存储单表或宽表数据,可以借助ES高效的读写性能,来完成实时数据模型的存储和查询功能。

[0092] 接下来对本申请实施例提供的生成宽表的方法进行解释说明。

[0093] 图2是本申请实施例提供的一种生成宽表的方法流程图,该生成宽表的方法可以包括如下几个步骤:

[0094] 步骤201:服务器接收宽表生成请求,宽表生成请求携带多个第一参考单表的标识、以及多个第一参考单表中存在关联关系的参考单表之间的关联字段,每个第一参考单表用于存储一种类型的数据。

[0095] 每个参考单表用于存储一种类型的数据。宽表用于将多个参考单表包括的不同类

型的数据汇总一起并进行存储,也即是,宽表涵盖了不同类型的数据。

[0096] 在一种可能的实现方式中,步骤201可能的实现过程为:管理人员通过终端的预设窗口编辑sql(Structured Query Language,结构化查询语言)语言,该sql语言用于指示多个第一参考单表的标识、以及多个第一参考单表中存在关联关系的参考单表之间的关联字段。终端基于检测到的sql语言生成宽表生成请求,然后将宽表生成请求发送至服务器,服务器接收该宽表生成请求。

[0097] 比如,管理人员通过终端的预设窗口编辑sql语言,该sql语言为:

[0098] select...from A a

[0099] left join B b on a.id=b.id

[0100] left join C c on b.code=c.code

[0101] left join D d on c.no=d.no

[0102] left join E e on a.id=e.id

[0103] 该sql语言用于指示标识分别为A、B、C、D、E的五个第一参考单表、以及这五个第一参考单表中存在关联关系的参考单表之间的关联字段。其中,标识为A的第一参考单表与标识为B的第一参考单表之间的关联字段为id,标识为B的第一参考单表与标识为C的第一参考单表之间的关联字段为code,标识为C的第一参考单表与标识为D的第一参考单表之间的关联字段为no,标识为A的第一参考单表与标识为E的第一参考单表之间的关联字段为id。终端基于检测到的sql语言生成宽表配置请求并发送至服务器之后,服务器接收该宽表配置请求。其中,id、code、no是预先设置的字段,本申请并不限定这些字段的具体含义。

[0104] 步骤202:服务器根据多个第一参考单表的单表标识获取多条第一操作记录,每条第一操作记录包括一个单表标识,每条第一操作记录用于指示对包括的单表标识对应的参考单表执行的操作。

[0105] 在一种可能的实现方式中,步骤202可能的实现过程为:从存储的操作记录中获取包括的单表标识与多个第一参考单表中任一个参考单表的单表标识一致的操作记录;按照每条操作记录包括的单表标识和表项标识,对获取的操作记录进行分类,得到多个操作记录集合,每个操作记录集合对应一个单表标识和一个表项标识;获取每个操作记录集合中操作执行时间距离当前时间最近的操作记录,得到多条第一操作记录。

[0106] 每个操作记录集合包括的所有操作记录对应同一参考单表中的同一表项。当针对同一参考单表中的同一表项的操作记录为多个时,服务器只需要根据操作执行时间距离当前时间最近的操作记录便可确定针对该表项的有效数据。因此,在获取多个第一操作记录时,可以对对应同一参考单表中的同一表项的多条操作记录进行筛选,只筛选出操作执行时间距离当前时间最近的操作记录即可,以减少后续根据操作记录拉取数据的数据处理量,从而提高生成宽表的效率。

[0107] 比如,针对第一参考单表A中的第一表项有3个操作记录,这三个操作记录的操作执行时间分别为T1、T2、T3,距离当前时间最近的执行时间为T3,则直接筛选出操作执行时间为T3的操作记录,将执行时间为T3的操作记录作为第一操作记录。

[0108] 在另一种可能的实现方式中,步骤202可能的实现过程为:从存储的操作记录中获取包括的单表标识与多个第一参考单表中任一个参考单表的单表标识一致的操作记录,将获取的操作记录直接作为多个第一操作记录。此时,无需对操作记录进行筛选,直接将参与参

考单表对应的所有操作记录作为多条第一操作记录。

[0109] 此外,上述任一操作记录用于指示对包括的单表标识对应的参考单表执行的操作。本申请实施例中,生成操作记录实现方式可以为:接收携带单表标识、表项标识和数据处理操作指示信息的日志消息;根据日志消息中携带的单表标识和表项标识,从存储的多个单表中获取当前待修改的表项,多个单表包括多个第一参考单表,每个单表包括多个表项;根据数据处理操作指示信息,修改获取的表项;根据修改结果生成一条操作记录。

[0110] 在一种可能的实现方式中,操作记录可以包括执行的操作的类型、操作执行时间、表项标识、单表标识。其中,执行的操作的类型包括插入操作和更新操作。该插入操作用于在参考单表中插入新的数据。更新操作用于将参考单表中的一条或者多条数据进行更新。其中,进行更新包括用新的数据替换原来的数据,或者将原来的数据删除。为了后续便于说明,将类型为插入操作的操作记录称为数据插入操作记录,将类型为更新操作的操作记录称为数据更新操作记录。

[0111] 另外,表项标识用于指示相应操作记录包括的单表标识对应的参考单表中的一个表项,也即是表项标识用于指示一个表项。

[0112] 步骤203:服务器根据多条第一操作记录和多个第一参考单表中存在关联关系的参考单表之间的关联字段,从多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到宽表。

[0113] 在执行步骤203的拉取数据的操作之前,服务器根据接收到的每个参考单表的标识在数据处理中心中查找出每个参考单表对应的数据的类型,并基于该多个参考单表中每个参考单表对应的数据的类型配置一张包括多种不同类型的数据的空白宽表的结构。服务器再通过步骤203从多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到宽表。

[0114] 基于步骤202可知,操作纪录可以为数据插入操作记录,也可以为数据更新操作记录,在本申请实施例中,针对不同类型的操作记录,拉取数据的实现方式不同,下面针对这两种类型的操作记录分别对步骤203进行解释说明。

[0115] 在第一种可能的实现方式中,当多条数据操作记录包括数据插入操作记录时,步骤203可以由步骤(1)~步骤(3)来实现:

[0116] (1)根据多个第一参考单表构建关联关系树。

[0117] 其中,关联关系树用于指示不同第一参考单表之间的关联关系,也即是,关联关系树中相邻的两个节点用于指示存在关联关系的两个第一参考单表。该存在关联关系的两个第一参考单表指示两个第一参考单表包括同一个关联字段。

[0118] 比如,根据多张第一参考单表,该多张第一参考单表如下表所示,分别为表A、表B、表C、表D、表E。根据各个第一参考单表之间的关联字段,构建出关联关系树,该关联关系树如图3所示。其中,节点A对应的第一参考单表为表A,节点B对应的第一参考单表为表B、节点C对应的第一参考单表为表C,节点D对应的第一参考单表为表D、节点E对应的第一参考单表为表E。由于表A和表B为存在关联关系的两个第一参考单表,因此表A对应的节点A与表B对应的节点B为关联关系树中相邻的两个节点。由于表A和表B为存在关联关系的两个第一参考单表,因此表A对应的节点A与表B对应的节点B为关联关系树中相邻的两个节点。由于表B和表C为存在关联关系的两个第一参考单表,因此表B对应的节点B与表C对应的节点C为关

联关系树中相邻的两个节点。由于表C和表D为存在关联关系的两个第一参考单表,因此表C对应的节点C与表D对应的节点D为关联关系树中相邻的两个节点。由于表A和表E为存在关联关系的两个第一参考单表,因此表A对应的节点A与表E对应的节点E为关联关系树中相邻的两个节点。此外,表A与表B的关联字段为专业号0703011。表B与表C的关联字段为学生A。表C与表D的关联字段为课程号A1325。表A与表E的关联字段为专业号0703012。

[0119] 表A

表项	专业 (011)	专业号 (012)	单表标识
A1	物理化学	0703011	001
A2	分析化学	0703012	001
A3	有机化学	0703013	001

[0121] 表B

表项	专业号	学号	学生	单表标识
B1	0703011	201315	学生A	002

[0123] 表C

表项	学生	课程号	成绩	单表标识
C1	学生A	A1325	80	003
C1	学生A	A1326	90	003

[0125] 表D

表项	课程号 (041)	教材 (04)	老师 (04)	单表标识
D1	A1325	大学化学	老师A	004

[0127] 表E

表项	专业号	学号	学生	单表标识
E1	0703012	201917	学生C	005
E2	0703012	201918	学生D	005

[0129] (2) 从多条第一操作记录包括的数据插入操作记录中,确定关联关系树中的每个节点对应的数据插入操作记录。

[0130] 如图3所示,通过步骤(2)将获取的多条第一操作记录中的数据插入操作记录分别映射到关键关系数据树中的每个节点上,从而得到关联关系树中的每个节点对应的数据插入操作记录。

[0131] 在一种可能的实现方式中,对于任一条数据插入操作记录,根据该数据插入操作记录包括的单表标识,确定该该数据插入操作记录对应的节点,从而将该数据插入操作记录作为该节点对应的一条数据插入操作记录。

[0132] 比如,当前有3条数据插入操作记录,分别标记为数据插入操作记录1、数据插入操作记录2、数据插入操作记录3。其中,数据插入操作记录1包括的单表标识为表A、数据插入操作记录2包括的单表标识为表B、数据插入操作记录3包括的单表标识为表A。此时,便可将数据插入操作记录1和数据插入操作记录3确定为图3中节点A对应的数据插入操作记录,将数据插入操作记录2确定为图3中的节点B对应的数据插入操作记录。

[0133] (3) 根据关联关系树中每个节点对应的数据插入操作记录和多个第一参考单表中

存在关联关系的参考单表之间的关联字段,按照关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据,将拉取的数据添加至空白宽表中。

[0134] 在一种可能的场景中,对于关联关系树中的第*i*个节点,当第*i*个节点为叶子节点时,将拉取的数据添加至空白宽表中可能的实现过程为:根据第*i*个节点对应的数据插入操作记录,从第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至空白宽表中。也即是,当操作记录对应节点为关联关系树中的叶子节点时,直接根据操作记录从相应参考单表中拉取数据添加至宽表即可。

[0135] 在另一种可能的场景中,当第*i*个节点不是叶子节点时,将拉取的数据添加至空白宽表中可能的实现过程为:确定第*i*个节点对应的第一参考单表与第*i-1*个节点对应的第一参考单表之间的关联字段,第*i*个节点为第*i-1*个节点的父节点;确定第*i-1*个节点对应的第一参考单表中包括关联字段的表项;根据确定的表项在空白宽表中添加空白表项,空白表项的表项标识与确定的表项的表项标识相同;根据第*i*个节点对应的数据插入操作记录,从与第*i*个节点对应的第一参考单表中拉取数据,将拉取的数据添加至空白表项中。

[0136] 比如,第*i*个节点不是叶子节点时,如图3所示,第*i*个节点为节点B,也即第*i-1*个节点为节点C。节点B为节点C的父节点。如上述表B和表C可知,节点B与节点C之间的关联字段为学生A。确定节点C对应的第一参考单表中包括该关联字段的表项为表项C1和表项C2。根据确定的两个表项在空白宽表中添加两个空白表项,然后基于节点B对应的数据插入操作记录从节点B对应的第一参考单表中拉取数据,将拉取的数据中添加至空白表项。

[0137] 为了后续便于说明,将按照关联关系树中从叶子节点到根节点的顺序,依次从每个节点对应的第一参考单表中拉取数据的过程称为反拉数据的操作。

[0138] 在本申请实施例中,对于所有的数据插入操作记录可以直接执行反拉数据的操作来拉取数据。可选地,也可以在以下场景下才执行反拉数据的操作。该场景为:当第*i*个节点不是叶子节点时,且节点*i*对应的参考单表包括一个表项,节点*i-1*对应的参考单表包括多个表项时,执行反拉数据的操作。

[0139] 比如,节点*i*对应的参考单表为表*i*,节点*i-1*对应的参考单表为表*i-1*,T1时刻之前,表*i*和表*i-1*分别如下图所示:

[0140] 表*i*

单表标识	表项	专业号	学号	学生
002	B1			

[0142] 表*i-1*

单表标识	表项	学生	课程号	成绩
003	C1	学生 A	A1325	80
003	C2	学生 A	A1326	90

003	C3			
-----	----	--	--	--

[0145] 按照关联关系树中从叶子节点到根节点的顺序,先将表*i-1*的数据拉取至宽表中,再根据表*i*和表*i-1*之间的关联字段将表*i*中的数据也拉取至宽表之中。具体的:到达T1时刻之前,表*i*为空表,表*i-1*存储有表项C1对应的数据以及表项C2对应的数据,此时先拉取C中

的数据,再拉取表i中的数据,宽表中将存储有表i-1包括的表项C1对应的数据以及表项C2对应的数据。这种场景下,如果先去表i中拉取数据,由于表i是空白表,此时将拉取不到任何数据,然后无法得到关键字段的具体值,也就无法根据关联字段从表i-1中拉取数据。这时就需要先拉取表i-1中的数据再拉取表i中的数据,才可以获取完整的数据。但是相当于通过3次拉取动作才获取完整的数据。这样耗费更多的计算机资源,因此优先采用反拉方案。这样不仅可以保证宽表中存储的数据是完整的,还可以耗费更少的计算机资源。

[0146] 到达T1时刻时,实时到达数据流为:表i包括的表项B1对应的数据,表i-1包括的表项C3对应的数据。这时,由于单表中的数据会实时更新到宽表中,所以宽表中已经存表i-1的数据,因此可以利用关联字段将表i的相关数据拉取至宽表中,再根据该关联字段将表i-1的相关数据拉取至宽表中,也可以先根据该关联字段将表i-1的相关数据拉取至宽表中,再利用关联字段将表i的相关数据拉取至宽表中。这样通过两次拉取动作就可以将表i和表i-1的与关联字段相关的数据拉取至宽表之中。

[0147] 在第二种可能的实现方式中,当多条数据操作记录包括数据更新操作记录时,根据数据插入操作记录生成宽表之后,还需要根据数据更新操作记录重新更新该宽表。具体过程为:对于多条第一操作记录中的每条数据更新操作记录,基于每条数据更新操作记录,从相应数据更新操作记录包括的单表标识对应的第一参考单表中拉取数据,将拉取的数据添加至宽表中。

[0148] 需要说明的是,数据删除操作记录属于特殊类型的数据更新操作记录,当数据更新操作记录用于指示将所有的数据更新为零时,该数据更新操作记录也即是数据删除操作记录。

[0149] 需要注意的是,数据更新操作记录不同于数据插入操作记录拉取数据的方式,不需要依照叶子节点到根节点的顺序拉取数据,只需要同时将叶子节点和根节点的数据同时根据数据更新操作记录进行更新。也即是,当多条第一操作记录中存在多条数据更新操作记录时,可以并行地根据每条数据更新操作记录通过上述实现方式更新宽表。这是由于数据插入操作记录需要根据两个参考单表之间的关联字段依次进行数据的拉取,因此无法同时执行数据插入操作记录。但是,对于多条数据更新操作记录则不存在这个问题。

[0150] 此外,需要说明的是,如果当前的数据操作记录是针对叶子节点的,比如,图3所示的节点D或者节点E,无论是数据更新操作记录还是数据插入操作记录,均直接执行批量更新的操作,不必执行上述先反拉再正拉的操作。

[0151] 此外,由于操作记录通常数量很多,为了实现流式处理这些操作记录,可以设置时间窗口,然后按照操作记录包括的操作执行时间距离当前时间从远到近的顺序将操作记录排序,按照时间窗口对操作记录进行分批,得到多批次的操作记录。对于每批次的操作记录,通过上述步骤203拉取数据更新至宽表。当根据最后一批次的操作记录更新完宽表之后,即可得到本申请实施例最终提供的宽表。

[0152] 此外,当执行完上述的数据插入操作和数据更新操作之后,就保证了当前时间窗口内的所有有关联关系的数据是全的,还有一部分数据与任一参考单表之间都没有任何关联关系,因此这部分的数据也可以插入到宽表之中,当然也可以不插入到宽表中。这时就得到一张完整的宽表。

[0153] 此外,当执行完上述操作之后,需要清理当前的时间窗口的数据,并重新创建时间

窗口,根据新创建的时间窗口,对宽表进行更新。

[0154] 其中,更新宽表可能的实现过程为:在根据用户端首次发起的宽表生成请求之后,服务器还可以接收用户端发起的宽表更新请求,宽表更新请求携带一个或多个第二参考单表的标识,每个第二参考单表用于存储一种类型的数据;根据一个或多个第二参考单表的标识,获取操作执行时间位于当前时间和当前时间之前最近一次更新宽表的时间之间的多条第二操作记录;根据多条第二操作记录从一个或多个第二参考单表中拉取数据,将拉取的数据添加至宽表中。

[0155] 其中,更新宽表的过程和生成宽表的过程基本相同,只是在更新宽表时只需要依据操作执行时间位于当前时间和当前时间之前最近一次更新宽表的时间之间的多条第二操作记录拉取数据即可,而不需要依据全部的操作记录来拉取数据。

[0156] 在本申请实施例中,针对单表的日志消息始终会存一份给宽表复用(以操作记录的方式),宽表反拉正拉的过程就是根据关联键到单表中拉取最新的记录数据,当单表中记录的数据发生了变更时,由于日志消息做了二次分发,宽表同样也会收到一样的操作记录,从而来根据更新宽表,以保证生成的宽表中的数据的时效性。

[0157] 此外,如果已经生成的宽表的外键变化,就是构建宽表的单表之间的关联字段发生变化。由于关联字段发生变化,但是构建宽表的各个单表的主键(也即是单表的表项标识)一定是不会变化的,这个时候需要根据单表主键,找到对应这条单表主键的宽表主键,把对应宽表的表项删了,然后重新根据变化后的关联字段生成宽表即可,在此不再详细说明。

[0158] 本申请实施例中,通过接收宽表生成请求,该宽表生成请求携带多个第一参考单表的标识以及多个第一参考单表中存在关联关系的参考单表之间的关联字段。再根据多个第一参考单表的单表标识获取多条第一操作记录。最后根据多条第一操作记录和多个第一参考单表中存在关联关系的参考单表之间的关联字段,从多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到宽表。由于该确定宽表的方法是根据各个单表之间的关联字段拉取数据,并将拉取的数据添加至空白宽表中得到的,因此而无需预先配置宽表的具体结构,只需用户设置用于构建宽表的参考单表和各个参考单表之间的关联字段,即可通过本申请提供的方法生成宽表,简化了用户操作,从而提高了获取宽表的灵活性。

[0159] 图4是本申请实施例提供的一种生成宽表的装置的结构示意图,该生成宽表的装置可以由软件、硬件或者两者的结合实现。该生成宽表的装置可以包括:

[0160] 第一接收模块401,用于接收宽表生成请求,宽表生成请求携带多个第一参考单表的标识、以及多个第一参考单表中存在关联关系的参考单表之间的关联字段,每个第一参考单表用于存储一种类型的数据;

[0161] 第一获取模块402,用于根据多个第一参考单表的单表标识获取多条第一操作记录,每条第一操作记录包括一个单表标识,每条第一操作记录用于指示对包括的单表标识对应的参考单表执行的操作;

[0162] 第一拉取模块403,用于根据多条第一操作记录和多个第一参考单表中存在关联关系的参考单表之间的关联字段,从多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到宽表。

[0163] 可选的,多条第一操作记录包括数据插入操作记录;

[0164] 拉取模块包括：

[0165] 构建子模块，用于根据多个第一参考单表构建关联关系树，关联关系树中的每个节点指示一个第一参考单表，关联关系树中相邻的两个节点用于指示存在关联关系的两个第一参考单表；

[0166] 确定子模块，用于从多条第一操作记录包括的数据插入操作记录中，确定关联关系树中的每个节点对应的数据插入操作记录；

[0167] 第一拉取子模块，用于根据关联关系树中每个节点对应的数据插入操作记录和多个第一参考单表中存在关联关系的参考单表之间的关联字段，按照关联关系树中从叶子节点到根节点的顺序，依次从每个节点对应的第一参考单表中拉取数据，将拉取的数据添加至空白宽表中。

[0168] 可选的，多条第一操作记录还包括数据更新操作记录；拉取模块还包括：

[0169] 第二拉取子模块，用于对于多条第一操作记录中的每条数据更新操作记录，基于每条数据更新操作记录，从相应数据更新操作记录包括的单表标识对应的第一参考单表中拉取数据，将拉取的数据添加至空白宽表中。

[0170] 可选的，第一拉取子模块包括：

[0171] 第一拉取单元，用于对于关联关系树中的第*i*个节点，当第*i*个节点为叶子节点时，根据第*i*个节点对应的数据插入操作记录，从第*i*个节点对应的第一参考单表中拉取数据，将拉取的数据添加至空白宽表中；

[0172] 第一确定单元，用于当第*i*个节点不是叶子节点时，确定第*i*个节点对应的第一参考单表与第*i*-1个节点对应的第一参考单表之间的关联字段，第*i*个节点为第*i*-1个节点的父节点；

[0173] 第二确定单元，用于确定第*i*-1个节点对应的第一参考单表中包括关联字段的表项；

[0174] 添加单元，用于根据确定的表项在空白宽表中添加空白表项，空白表项的表项标识与确定的表项的表项标识相同；

[0175] 第二拉取单元，用于根据第*i*个节点对应的数据插入操作记录，从与第*i*个节点对应的第一参考单表中拉取数据，将拉取的数据添加至空白表项中。

[0176] 可选的，获取模块包括：

[0177] 第一获取子模块，用于从存储的操作记录中获取包括的单表标识与多个第一参考单表中任一个参考单表的标识一致的操作记录；

[0178] 分类子模块，用于按照每条操作记录包括的单表标识和表项标识，对获取的操作记录进行分类，得到多个操作记录集合，每个操作记录集合对应一个单表标识和一个表项标识；

[0179] 第二获取子模块，用于获取每个操作记录集合中操作执行时间距离当前时间最近的操作记录，得到多条第一操作记录。

[0180] 可选的，该装置还包括：

[0181] 第二接收模块，用于接收宽表更新请求，宽表更新请求携带一个或多个第二参考单表的标识，每个第二参考单表用于存储一种类型的数据；

[0182] 第二获取模块，用于根据一个或多个第二参考单表的标识，获取操作执行时间位

于当前时间和当前时间之前最近一次更新宽表的时间之间的多条第二操作记录；

[0183] 第二拉取模块,用于根据多条第二操作记录从一个或多个第二参考单表中拉取数据,将拉取的数据添加至宽表中。

[0184] 可选的,该装置还包括:

[0185] 第三接收模块,用于接收携带单表标识、表项标识和数据处理操作指示信息的日志消息;

[0186] 第三获取模块,用于根据日志消息中携带的单表标识和表项标识,从存储的多个单表中获取当前待修改的表项,多个单表包括多个第一参考单表以及一个或多个第二参考单表,每个单表包括多个表项;

[0187] 修改模块,用于根据数据处理操作指示信息,修改获取的表项;

[0188] 生成模块,用于根据修改结果生成一条操作记录。

[0189] 在本申请实施例中,通过接收宽表生成请求,该宽表生成请求携带多个第一参考单表的标识以及多个第一参考单表中存在关联关系的参考单表之间的关联字段。再根据多个第一参考单表的单表标识获取多条第一操作记录。最后根据多条第一操作记录和多个第一参考单表中存在关联关系的参考单表之间的关联字段,从多个第一参考单表中拉取数据,将拉取的数据添加至空白宽表中,得到宽表。由于该确定宽表的方法是根据各个单表之间的关联字段拉取数据,并将拉取的数据添加至空白宽表中得到的,因此而无需预先配置宽表的具体结构,只需用户设置用于构建宽表的参考单表和各个参考单表之间的关联字段,即可通过本申请提供的方法生成宽表,简化了用户操作,从而提高了获取宽表的灵活性。

[0190] 需要说明的是:上述实施例提供的生成宽表的装置在生成宽表时,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将设备的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。另外,上述实施例提供的生成宽表的装置与生成宽表的方法实施例属于同一构思,其具体实现过程详见方法实施例,这里不再赘述。

[0191] 图5是本申请实施例提供的一种服务器结构示意图。该服务器可以是后台服务器集群中的服务器。具体来讲:

[0192] 服务器500包括中央处理单元(CPU)501、包括随机存取存储器(RAM)502和只读存储器(ROM)503的系统存储器504,以及连接系统存储器504和中央处理单元501的系统总线505。服务器500还包括帮助计算机内的各个器件之间传输信息的基本输入/输出系统(I/O系统)506,和用于存储操作系统513、应用程序514和其他程序模块515的大容量存储设备507。

[0193] 基本输入/输出系统506包括有用于显示信息的显示器508和用于用户输入信息的诸如鼠标、键盘之类的输入设备509。其中显示器508和输入设备509都通过连接到系统总线505的输入输出控制器510连接到中央处理单元501。基本输入/输出系统506还可以包括输入输出控制器510以用于接收和处理来自键盘、鼠标、或电子触控笔等多个其他设备的输入。类似地,输入输出控制器510还提供输出到显示屏、打印机或其他类型的输出设备。

[0194] 大容量存储设备507通过连接到系统总线505的大容量存储控制器(未示出)连接到中央处理单元501。大容量存储设备507及其相关联的计算机可读介质为服务器500提供

非易失性存储。也就是说,大容量存储设备507可以包括诸如硬盘或者CD-ROM驱动器之类的计算机可读介质(未示出)。

[0195] 不失一般性,计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其他数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括RAM、ROM、EPROM、EEPROM、闪存或其他固态存储其技术,CD-ROM、DVD或其他光学存储、磁带盒、磁带、磁盘存储或其他磁性存储设备。当然,本领域技术人员可知计算机存储介质不局限于上述几种。上述的系统存储器504和大容量存储设备507可以统称为存储器。

[0196] 根据本申请的各种实施例,服务器500还可以通过诸如因特网等网络连接到网络上的远程计算机运行。也即服务器500可以通过连接在系统总线505上的网络接口单元511连接到网络512,或者说,也可以使用网络接口单元511来连接到其他类型的网络或远程计算机系统(未示出)。

[0197] 上述存储器还包括一个或者一个以上的程序,一个或者一个以上程序存储于存储器中,被配置由CPU执行。

[0198] 本申请实施例还提供了一种非临时性计算机可读存储介质,当所述存储介质中的指令由服务器的处理器执行时,使得服务器能够执行上述实施例提供的生成宽表的方法。

[0199] 本申请实施例还提供了一种包含指令的计算机程序产品,当其在服务器上运行时,使得服务器执行上述实施例提供的生成宽表的方法。

[0200] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0201] 以上所述仅为本申请的较佳实施例,并不用以限制本申请,凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

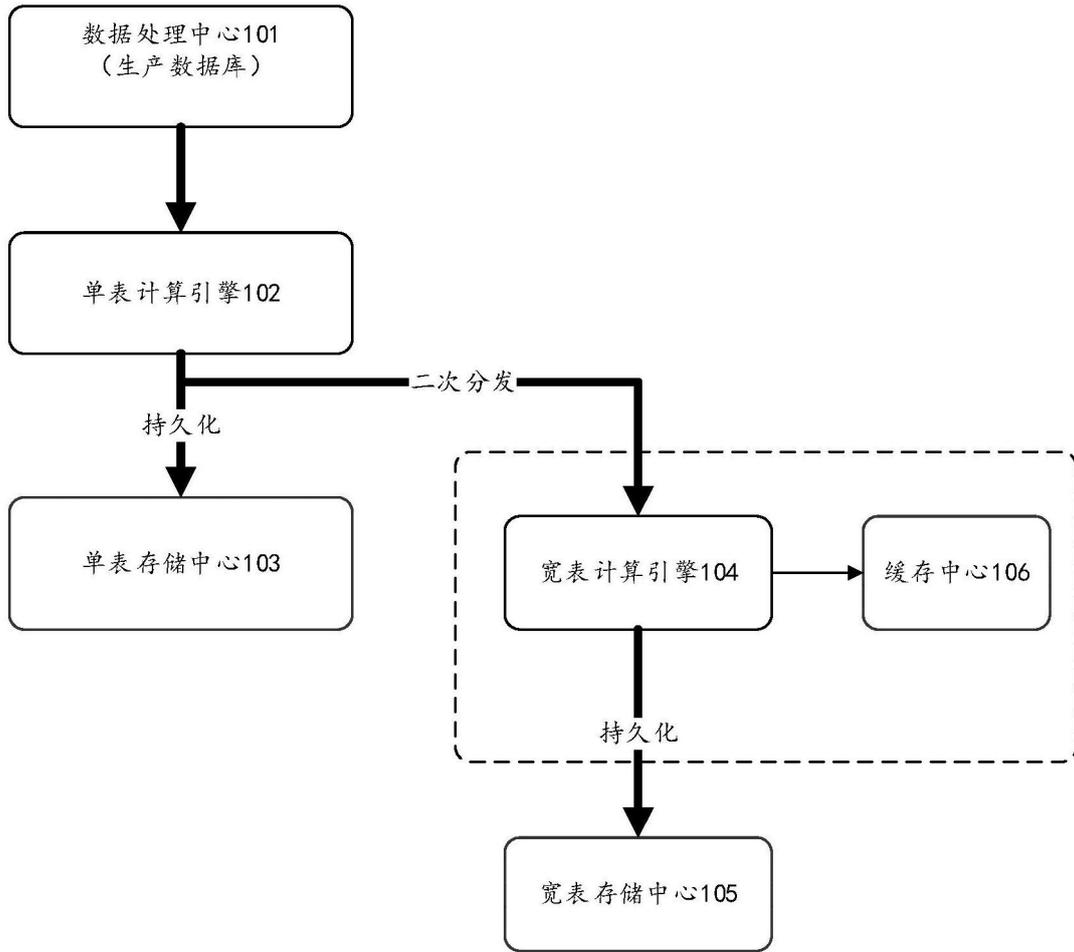


图1

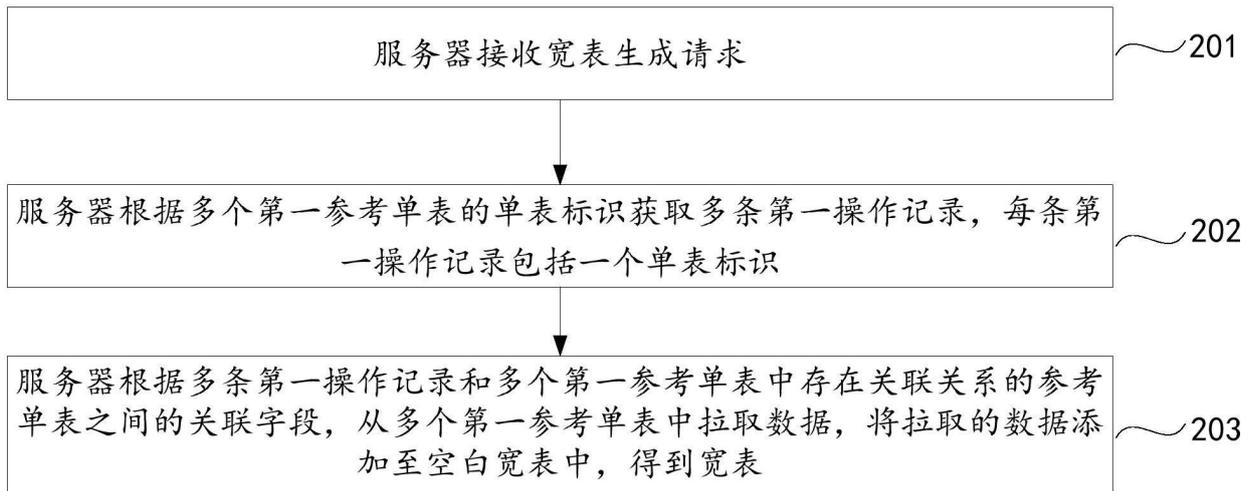


图2

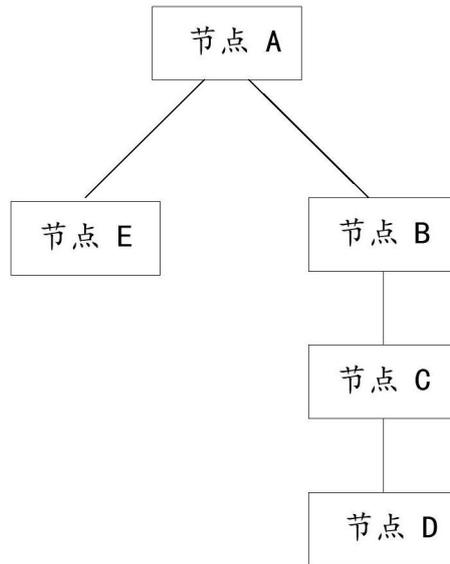


图3

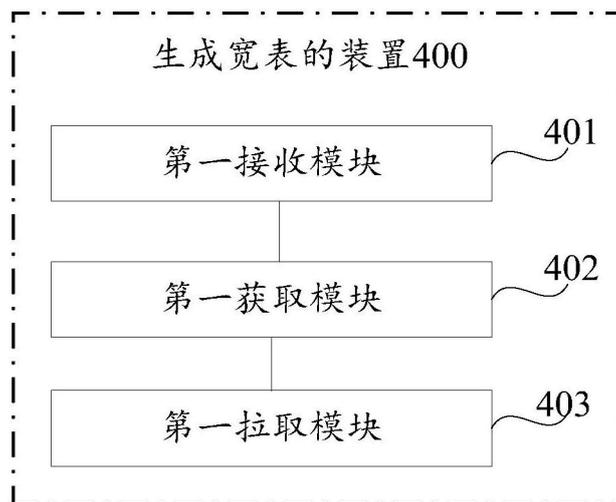


图4

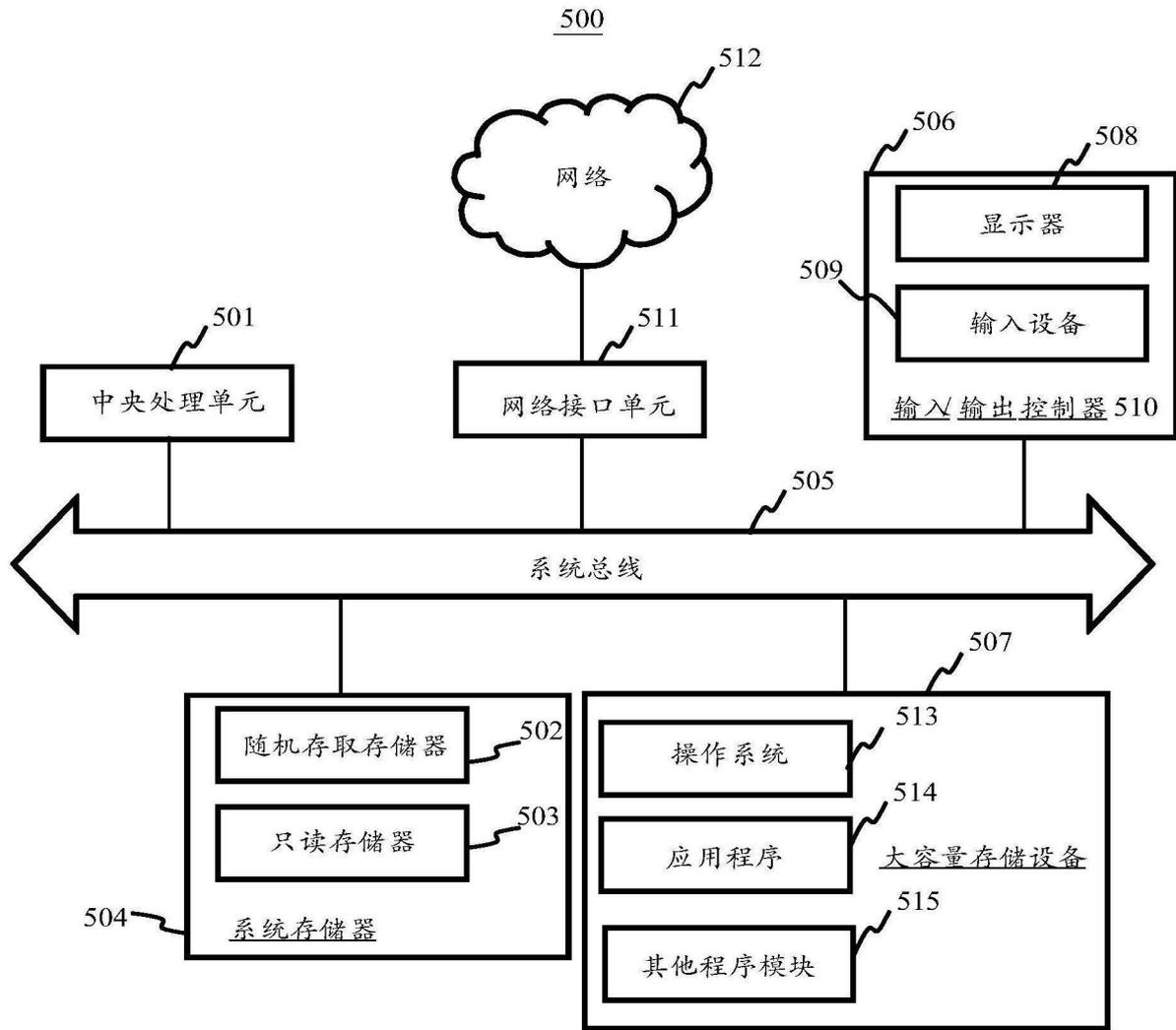


图5