



(12) 发明专利

(10) 授权公告号 CN 113689028 B

(45) 授权公告日 2024. 11. 15

(21) 申请号 202110821870.5

(22) 申请日 2021.07.19

(65) 同一申请的已公布的文献号
申请公布号 CN 113689028 A

(43) 申请公布日 2021.11.23

(73) 专利权人 杉数科技(北京)有限公司
地址 100102 北京市朝阳区利泽中园106号
楼2层202A-12

专利权人 上海杉数网络科技有限公司
杉数科技(苏州)有限公司
深圳市杉智科技有限公司

(72) 发明人 王劲松

(74) 专利代理机构 北京众达德权知识产权代理
有限公司 11570
专利代理师 甄伟军

(51) Int.Cl.

G06F 18/10 (2023.01)

G06F 9/50 (2006.01)

(56) 对比文件

CN 112559301 A, 2021.03.26

CN 111369037 A, 2020.07.03

CN 111680108 A, 2020.09.18

审查员 翟燕燕

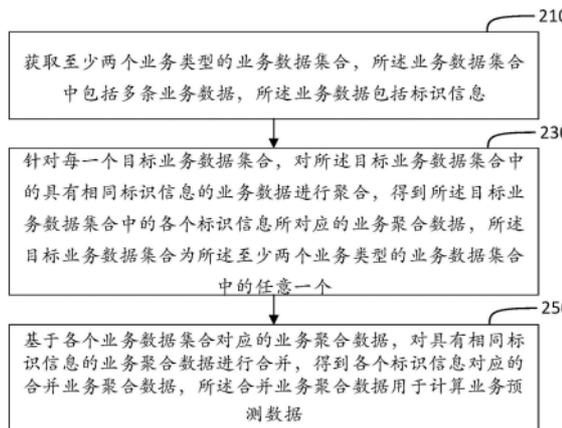
权利要求书2页 说明书12页 附图5页

(54) 发明名称

业务数据处理方法、装置、计算机可读介质及电子设备

(57) 摘要

本申请的实施例提供了一种业务数据处理方法、装置、计算机可读介质及电子设备。该方法包括:获取至少两个业务类型的业务数据集合,该业务数据集合中包括多条业务数据,该业务数据包括标识信息;针对每一个目标业务数据集合,对目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到目标业务数据集合中的各个标识信息所对应的业务聚合数据,目标业务数据集合为至少两个业务类型的业务数据集合中的任意一个;基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的用于计算业务预测数据的合并业务聚合数据。本申请实施例的技术方案可以提高计算业务预测数据的准确性。



1. 一种业务数据处理方法,其特征在于,所述方法包括:

获取至少两个业务类型的业务数据集合,所述业务数据集合包括货物销售数据集合,货物库存数据集合,货物价格数据集合中的至少一种,所述业务数据集合中包括多条业务数据,所述业务数据包括标识信息;

针对每一个目标业务数据集合,对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到所述目标业务数据集合中的各个标识信息所对应的业务聚合数据,所述目标业务数据集合为所述至少两个业务类型的业务数据集合中的任意一个;

还包括:

获取各个标识信息对应的合并业务聚合数据;

根据分布式集群中各个集群节点的负载,将所述各个标识信息对应的合并业务聚合数据分发至所述分布式集群的各个集群节点;由各个集群节点基于被分发的合并业务聚合数据计算对应标识信息的业务预测数据;

基于每一个目标标识信息对应的合并业务聚合数据,通过对应的集群节点计算所述目标标识信息对应的业务预测数据,所述目标标识信息为所述各个标识信息中的任意一个;

基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行拼接,得到各个标识信息对应的合并业务聚合数据,所述合并业务聚合数据用于计算业务预测数据。

2. 根据权利要求1所述的方法,其特征在于,所述对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,包括:

将所述目标业务数据集合中的业务数据发送至分布式集群,并通过所述分布式集群对所述目标业务数据集合中具有相同标识信息的业务数据进行聚合。

3. 根据权利要求2所述的方法,其特征在于,所述通过所述分布式集群对所述目标业务数据集合中具有相同标识信息的业务数据进行聚合,包括:

基于业务数据的标识信息,对所述目标业务数据集合中的业务数据进行分组,得到各个标识信息对应的业务数据组;

通过所述分布式集群中的各个集群节点,并发对得到的各个业务数据组中的业务数据进行聚合。

4. 根据权利要求3所述的方法,其特征在于,所述业务数据还包括业务信息,所述通过所述分布式集群中的各个集群节点,并发对得到的各个业务数据组中的业务数据进行聚合,包括:

根据所述分布式集群中各个集群节点的负载,将所述各个标识信息对应的业务数据组分发至所述各个集群节点;

针对所述分布式集群中的每一个集群节点,通过所述集群节点对所述业务数据组中各条业务数据的业务信息进行序列化聚合。

5. 根据权利要求3所述的方法,其特征在于,所述基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行拼接,得到各个标识信息对应的合并业务聚合数据,包括:

在所述分布式集群的各个集群节点中确定具有相同标识信息的业务聚合数据;

分别对各个标识信息对应的业务聚合数据进行拼接,得到各个标识信息对应的合并业

务聚合数据。

6. 根据权利要求1所述的方法,其特征在于,所述通过对应的集群节点计算所述目标标识信息对应的业务预测数据,包括:

由所述对应的集群节点对所述目标标识信息对应的合并业务聚合数据进行还原处理,得到所述目标标识信息在各个业务类型下对应的业务数据;

由所述对应的集群节点调用与业务需求相匹配的运算代码;

将所述目标标识信息在各个业务类型下对应的业务数据作为输入数据,通过运行所述运算代码计算所述目标标识信息对应的业务预测数据。

7. 一种业务数据处理装置,其特征在于,所述装置包括:

第一获取单元,用于获取至少两个业务类型的业务数据集合,所述业务数据集合包括货物销售数据集合,货物库存数据集合,货物价格数据集合中的至少一种,所述业务数据集合中包括多条业务数据,所述业务数据包括标识信息;

聚合单元,用于针对每一个目标业务数据集合,对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到所述目标业务数据集合中的各个标识信息所对应的业务聚合数据,所述目标业务数据集合为所述至少两个业务类型的业务数据集合中的任意一个;

还包括第二获取单元,被用于获取各个标识信息对应的合并业务聚合数据;分发单元,被用于根据分布式集群中各个集群节点的负载,将所述各个标识信息对应的合并业务聚合数据分发至所述分布式集群的各个集群节点;由各个集群节点基于被分发的合并业务聚合数据计算对应标识信息的业务预测数据;计算单元,被用于基于每一个目标标识信息对应的合并业务聚合数据,通过对应的集群节点计算所述目标标识信息对应的业务预测数据,所述目标标识信息为所述各个标识信息中的任意一个

拼接单元,用于基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行拼接,得到各个标识信息对应的合并业务聚合数据,所述合并业务聚合数据用于计算业务预测数据。

8. 一种电子设备,其特征在于,所述电子设备包括一个或多个处理器和一个或多个存储器,所述一个或多个存储器中存储有至少一条程序代码,所述至少一条程序代码由所述一个或多个处理器加载并执行以实现如权利要求1至6任一项所述的业务数据处理方法所执行的操作。

9. 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中存储有至少一条程序代码,所述至少一条程序代码由处理器加载并执行以实现如权利要求1至6任一项所述的业务数据处理方法所执行的操作。

业务数据处理方法、装置、计算机可读介质及电子设备

技术领域

[0001] 本申请涉及计算机及数据处理技术领域,具体而言,涉及一种业务数据处理方法、装置、计算机可读介质及电子设备。

背景技术

[0002] 在业务数据处理场景中,一般可以通过设计好的自定义函数对数据进行处理,然而,这种情况无法支持对多个类型的业务数据进行处理,在需要通过业务数据计算业务预测数据时,不能保证业务预测数据的准确性。基于此,如何提高计算业务预测数据的准确性是亟待解决的技术问题。

发明内容

[0003] 本申请的实施例提供了一种业务数据处理方法、装置、计算机程序产品或计算机程序、计算机可读介质及电子设备,进而至少在一定程度上可以提高计算业务预测数据的准确性。

[0004] 本申请的其他特性和优点将通过下面的详细描述变得显然,或部分地通过本申请的实践而习得。

[0005] 根据本申请实施例的一个方面,提供了一种业务数据处理方法,所述方法包括:获取至少两个业务类型的业务数据集合,所述业务数据集合中包括多条业务数据,所述业务数据包括标识信息;针对每一个目标业务数据集合,对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到所述目标业务数据集合中的各个标识信息所对应的业务聚合数据,所述目标业务数据集合为所述至少两个业务类型的业务数据集合中的任意一个;基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据,所述合并业务聚合数据用于计算业务预测数据。

[0006] 根据本申请实施例的一个方面,提供了一种业务数据处理装置,所述装置包括:第一获取单元,用于获取至少两个业务类型的业务数据集合,所述业务数据集合中包括多条业务数据,所述业务数据包括标识信息;聚合单元,用于针对每一个目标业务数据集合,对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到所述目标业务数据集合中的各个标识信息所对应的业务聚合数据,所述目标业务数据集合为所述至少两个业务类型的业务数据集合中的任意一个;合并单元,用于基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据,所述合并业务聚合数据用于计算业务预测数据。

[0007] 在本申请的一些实施例中,基于前述方案,所述聚合单元配置为:将所述目标业务数据集合中的业务数据发送至分布式集群,并通过所述分布式集群对所述目标业务数据集合中具有相同标识信息的业务数据进行聚合。

[0008] 在本申请的一些实施例中,基于前述方案,所述聚合单元进一步配置为:基于业务

数据的标识信息,对所述目标业务数据集中的业务数据进行分组,得到各个标识信息对应的业务数据组;通过所述分布式集群中的各个集群节点,并发对得到的各个业务数据组中的业务数据进行聚合。

[0009] 在本申请的一些实施例中,基于前述方案,所述业务数据还包括业务信息,所述聚合单元进一步配置为:根据所述分布式集群中各个集群节点的负载,将所述各个标识信息对应的业务数据组分发至所述各个集群节点;针对所述分布式集群中的每一个集群节点,通过所述集群节点对所述业务数据组中各条业务数据的业务信息进行序列化聚合。

[0010] 在本申请的一些实施例中,基于前述方案,所述合并单元配置为:在所述分布式集群的各个集群节点中确定具有相同标识信息的业务聚合数据;分别对各个标识信息对应的业务聚合数据进行拼接,得到各个标识信息对应的合并业务聚合数据。

[0011] 在本申请的一些实施例中,基于前述方案,所述装置还包括第二获取单元,被用于获取各个标识信息对应的合并业务聚合数据;分发单元,被用于根据分布式集群中各个集群节点的负载,将所述各个标识信息对应的合并业务聚合数据分发至所述分布式集群的各个集群节点;计算单元,被用于基于每一个目标标识信息对应的合并业务聚合数据,通过对应的集群节点计算所述目标标识信息对应的业务预测数据,所述目标标识信息为所述各个标识信息中的任意一个。

[0012] 在本申请的一些实施例中,基于前述方案,所述计算单元配置为:由所述对应的集群节点对所述目标标识信息对应的合并业务聚合数据进行还原处理,得到所述目标标识信息在各个业务类型下对应的业务数据;由所述对应的集群节点调用与业务需求相匹配的运算代码;将所述目标标识信息在各个业务类型下对应的业务数据作为输入数据,通过运行所述运算代码计算所述目标标识信息对应的业务预测数据。

[0013] 根据本申请实施例的一个方面,提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述实施例中所述的业务数据处理方法。

[0014] 根据本申请实施例的一个方面,提供了一种计算机可读介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如上述实施例中所述的业务数据处理方法。

[0015] 根据本申请实施例的一个方面,提供了一种电子设备,包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述一个或多个处理器实现如上述实施例中所述的业务数据处理方法。

[0016] 在本申请的一些实施例所提供的技术方案中,通过对业务数据集中的具有相同标识信息的业务数据进行聚合,得到各个标识信息所对应的业务聚合数据,然后再基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的于计算业务预测数据的合并业务聚合数据。由于通过综合考虑多个业务类型的业务数据集合,得到能够全面表征业务特征的合并业务聚合数据,从而使得能够通过合并业务聚合数据计算出更加准确的业务预测数据。

[0017] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本申请。

附图说明

[0018] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本申请的实施例,并与说明书一起用于解释本申请的原理。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。在附图中:

[0019] 图1示出了可以应用本申请实施例的技术方案的示例性系统架构的示意图;

[0020] 图2示出了根据本申请一个实施例的业务数据处理方法的流程图;

[0021] 图3示出了根据本申请一个实施例的通过所述分布式集群对所述目标业务数据集中具有相同标识信息的业务数据进行聚合的细节流程图;

[0022] 图4示出了根据本申请一个实施例的并发对得到的各个业务数据组中的业务数据进行聚合的细节流程图;

[0023] 图5示出了根据本申请一个实施例的对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据的细节流程图;

[0024] 图6示出了根据本申请一个实施例的业务数据处理方法的进一步流程图;

[0025] 图7示出了根据本申请一个实施例的通过对应的集群节点计算所述目标标识信息对应的业务预测数据的细节流程图;

[0026] 图8示出了根据本申请一个实施例的业务数据流的示意图;

[0027] 图9示出了根据本申请一个实施例的业务数据处理装置的框图;

[0028] 图10示出了适于用来实现本申请实施例的电子设备的计算机系统的结构示意图。

具体实施方式

[0029] 现在将参考附图更全面地描述示例实施方式。然而,示例实施方式能够以多种形式实施,且不应被理解为限于在此阐述的范例;相反,提供这些实施方式使得本申请将更加全面和完整,并将示例实施方式的构思全面地传达给本领域的技术人员。

[0030] 此外,所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施例中。在下面的描述中,提供许多具体细节从而给出对本申请的实施例的充分理解。然而,本领域技术人员将意识到,可以实践本申请的技术方案而没有特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知方法、装置、实现或者操作以避免模糊本申请的各方面。

[0031] 附图中所示的方框图仅仅是功能实体,不一定必须与物理上独立的实体相对应。即,可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0032] 附图中所示的流程图仅是示例性说明,不是必须包括所有的内容和操作/步骤,也不是必须按所描述的顺序执行。例如,有的操作/步骤还可以分解,而有的操作/步骤可以合并或部分合并,因此实际执行的顺序有可能根据实际情况改变。

[0033] 需要说明的是:在本文中提及的“多个”是指两个或两个以上。“和/或”描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[0034] 需要注意的是,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对象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图示或描述的那些以外的顺序实施。

[0035] 图1示出了可以应用本申请实施例的技术方案的示例性系统架构的示意图。

[0036] 如图1所示,系统架构可以包括服务器101(还可以包括智能手机、平板电脑和便携式计算机中的一种或多种)、分布式集群102和网络103。网络103用以在服务器101和分布式集群102之间提供通信链路的介质。网络103可以包括各种连接类型,例如有线通信链路、无线通信链路等等。

[0037] 在本申请的一个实施例中,所述业务数据处理方法可以由服务器101执行,进一步的,可以由分布式集群102配合服务器101执行。

[0038] 具体的,可以是由服务器101获取至少两个业务类型的业务数据集合,所述业务数据集合中包括多条业务数据,所述业务数据包括标识信息,针对每一个目标业务数据集合,可以通过分布式集群102对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到所述目标业务数据集合中的各个标识信息所对应的业务聚合数据,所述目标业务数据集合为所述至少两个业务类型的业务数据集合中的任意一个,然后,服务器101可以基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据,所述合并业务聚合数据用于计算业务预测数据。

[0039] 本申请通过综合考虑多个业务类型的业务数据集合,得到能够全面表征业务特征的合并业务聚合数据,从而使得能够通过合并业务聚合数据计算出更加准确的业务预测数据。

[0040] 以下对本申请实施例的技术方案的实现细节进行详细阐述:

[0041] 图2示出了根据本申请一个实施例的业务数据处理方法的流程图,该业务数据处理方法可以由具有计算处理功能的设备来执行,比如可以由图1中所示的服务器和分布式集群来配合执行。参照图2所示,该业务数据处理方法至少包括步骤210至步骤230,详细介绍如下:

[0042] 在步骤210中,获取至少两个业务类型的业务数据集合,所述业务数据集合中包括多条业务数据,所述业务数据包括标识信息。

[0043] 在本申请中,所述业务数据处理方法可以应用在处理货物管理数据的场景中,其中,货物管理数据包括货物销售数据,货物库存数据,货物价格数据等多种业务类型的数据。

[0044] 进一步的,在处理货物管理数据的场景中,获取至少两个业务类型的业务数据集合,可以是指获取货物销售数据集合,货物库存数据集合,货物价格数据集合。

[0045] 每一个业务类型的业务数据集合中包括有多条业务数据,每一条业务数据包括一个标识信息,例如,在货物销售数据集合中包括有多条货物销售数据,其中,每一条货物销售数据对应包括一个标识信息,所述标识信息可以是货物种类,也可以是货物型号,还可以是货物名称或者货物编号。

[0046] 需要说明的是,在本申请中,所述业务数据处理方法除了应用在处理货物管理数据的场景中,也可以应用在其他场景中,例如处理金融数据的场景。

[0047] 还需要说明的是,业务数据集中的业务数据为历史业务数据。

[0048] 继续参照图2,在步骤230中,针对每一个目标业务数据集合,对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到所述目标业务数据集合中的各个标识信息所对应的业务聚合数据,所述目标业务数据集合为所述至少两个业务类型的业务数据集合中的任意一个。

[0049] 在本申请的一个实施例中,对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,可以是将所述目标业务数据集合中的业务数据发送至分布式集群,并通过所述分布式集群对所述目标业务数据集合中具有相同标识信息的业务数据进行聚合。

[0050] 应该理解的是,在本申请中,需要对每一个业务数据集合中的业务数据进行聚合处理。

[0051] 在本实施例中,分布式集群可以是Spark集群,其中,Spark是基于内存计算的大数据并行计算框架,通过master+slave的管理模式,可以以集群的形式处理大规模的数据。

[0052] 在本实施中,通过所述分布式集群对所述目标业务数据集合中具有相同标识信息的业务数据进行聚合,可以按照如图3所示的步骤执行。

[0053] 参见图3,示出了根据本申请一个实施例的通过所述分布式集群对所述目标业务数据集合中具有相同标识信息的业务数据进行聚合的细节流程图。具体包括步骤231至步骤232:

[0054] 在步骤231中,基于业务数据的标识信息,对所述目标业务数据集合中的业务数据进行分组,得到各个标识信息对应的业务数据组。

[0055] 具体的,由于业务数据集合中的每一条业务数据都对应有一个标识信息,因此可以将具有相同标识信息的业务数据分为一组。例如,在处理货物管理数据的场景中,货物销售数据集合中包括如下表1所示的6条销售数据,在该货物销售数据集合中,可以将“(标识信息,销售信息1,销售信息2,销售信息3) = (a,a1,a2,a3)”,“(标识信息,销售信息1,销售信息2,销售信息3) = (a,a4,a5,a6)”,“(标识信息,销售信息1,销售信息2,销售信息3) = (a,a7,a8,a9)”分为一组;将“(标识信息,销售信息1,销售信息2,销售信息3) = (b,b1,b2,b3)”,“(标识信息,销售信息1,销售信息2,销售信息3) = (b,b4,b5,b6)”分为一组;将“(标识信息,销售信息1,销售信息2,销售信息3) = (c,c1,c2,c3)”分为一组。

标识信息	销售信息1	销售信息2	销售信息3
型号a	a1	a2	a3
型号c	c1	c2	c3
型号b	b1	b2	b3
型号a	a4	a5	a6
型号c	c4	c5	c6
型号a	a7	a8	a9

[0056] 表1

[0057] 在步骤232中,通过所述分布式集群中的各个集群节点,并发对得到的各个业务数据组中的业务数据进行聚合。

[0058] 在本申请中,所述业务数据还包括业务信息,例如,在表1中,“销售信息1,销售信息2,销售信息3”即为所述业务数据的业务信息。

[0060] 在本申请的一个实施例中,通过所述分布式集群中的各个集群节点,并发对得到的各个业务数据组中的业务数据进行聚合,可以按照如图4所示的步骤执行。

[0061] 参见图4,示出了根据本申请一个实施例的并发对得到的各个业务数据组中的业务数据进行聚合的细节流程图。具体包括步骤2321至步骤2322:

[0062] 在步骤2321中,根据所述分布式集群中各个集群节点的负载,将所述各个标识信息对应的业务数据组分发至所述各个集群节点。

[0063] 在步骤2322中,针对所述分布式集群中的每一个集群节点,通过所述集群节点对所述业务数据组中各条业务数据的业务信息进行序列化聚合。

[0064] 具体的,例如,在表1中,标识信息为“型号a”的业务数据组包括“(标识信息,销售信息1,销售信息2,销售信息3) = (a,a1,a2,a3)”,“(标识信息,销售信息1,销售信息2,销售信息3) = (a,a4,a5,a6)”,“(标识信息,销售信息1,销售信息2,销售信息3) = (a,a7,a8,a9)”3条销售数据,通过集群节点对3条销售数据中的业务信息进行序列化聚合,得到标识信息为“型号a”所对应的业务聚合数据“(销售信息1销售信息2销售信息3) = (a1 a2 a3,a4 a5 a6,a7 a8 a9)”。

[0065] 在本申请中,通过所述分布式集群中的各个集群节点,并发对得到的各个业务数据组中的业务数据进行聚合,其好处在于,可以增强对业务数据集合中业务数据进行聚合的速度,提高聚合效率。

[0066] 继续参照图2,在步骤250中,基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据,所述合并业务聚合数据用于计算业务预测数据。

[0067] 通过在步骤230中对每一个业务数据集合中的业务数据进行聚合,使得每一个业务数据集合对应包括有各个标识信息的业务聚合数据。由于一个业务数据集合中各个标识信息的业务数据由不同集群节点聚合,因此,可以理解的是,每一个业务数据集合中各个标识信息的业务聚合数据分布在不同的集群节点。

[0068] 基于此,在本申请的一个实施例中,基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据可以按照如图5所示的步骤执行。

[0069] 参照图5,示出了根据本申请一个实施例的对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据的细节流程图。具体包括步骤251至步骤252:

[0070] 在步骤251中,在所述分布式集群的各个集群节点中确定具有相同标识信息的业务聚合数据。

[0071] 在步骤252中,分别对各个标识信息对应的业务聚合数据进行拼接,得到各个标识信息对应的合并业务聚合数据。

[0072] 具体的,例如,在处理货物管理数据的场景中,在销售数据集合中,标识信息“型号a”所对应的业务聚合数据为“(销售信息1销售信息2销售信息3) = (a1 a2 a3,a4 a5 a6,a7 a8 a9)”;在库存数据集合中,标识信息“型号a”所对应的业务聚合数据为“(库存信息1库存信息2库存信息3) = (aa1 aa2 aa3,aa4 aa5 aa6,aa7 aa8 aa9)”;在价格数据集合中,标识信息“型号a”所对应的业务聚合数据为“(价格信息1价格信息2价格信息3) = (aaa1 aaa2

aaa3,aaa4 aaa5 aaa6,aaa7 aaa8 aaa9)”。

[0073] 通过标识信息“型号a”所对应的各个业务聚合数据进行合并,得到标识信息“型号a”所对应的合并业务聚合数据“(销售信息1销售信息2销售信息3) = (a1 a2 a3,a4 a5 a6, a7 a8 a9);(库存信息1库存信息2库存信息3) = (aa1 aa2 aa3,aa4 aa5 aa6,aa7 aa8 aaa9);(价格信息1价格信息2价格信息3) = (aaa1 aaa2 aaa3,aaa4 aaa5 aaa6,aaa7 aaa8 aaa9)”。

[0074] 在本申请中,首先将标识信息对应的业务数据聚合为业务聚合数据,然后将标识信息对应的业务聚合数据合并为合并业务聚合数据,可以将零散的业务数据合为能够全面表征业务特征的一条数据,从而使得能够通过合并业务聚合数据计算出更加准确的业务预测数据。

[0075] 在本申请中,在得到各个标识信息对应的合并业务聚合数据之后,还可以执行如图6所示的步骤。

[0076] 参见图6,示出了根据本申请一个实施例的业务数据处理方法的进一步流程图。具体包括步骤260至步骤280:

[0077] 在步骤260中,获取各个标识信息对应的合并业务聚合数据。

[0078] 在步骤270中,根据分布式集群中各个集群节点的负载,将所述各个标识信息对应的合并业务聚合数据分发至所述分布式集群的各个集群节点。

[0079] 在步骤280中,基于每一个目标标识信息对应的合并业务聚合数据,通过对应的集群节点计算所述目标标识信息对应的业务预测数据,所述目标标识信息为所述各个标识信息中的任意一个。

[0080] 在本申请中,将各个合并业务聚合数据分发至各个集群节点,由各个集群节点基于被分发的合并业务聚合数据计算对应标识信息的业务预测数据,其好处在于,可以增强对业务预测数据进行计算的速度,提高计算效率。

[0081] 而根据各个集群节点的负载将各个合并业务聚合数据分发至各个集群节点,可以均衡各个集群节点的计算压力,同时充分利用每个集群节点计算资源,进一步提高对业务预测数据进行计算的效率。

[0082] 在步骤280的一个实施例中,通过对应的集群节点计算所述目标标识信息对应的业务预测数据,可以按照如图7所示的步骤执行。

[0083] 参见图7,示出了根据本申请一个实施例的通过对应的集群节点计算所述目标标识信息对应的业务预测数据的细节流程图。具体包括步骤281至步骤283:

[0084] 在步骤281中,由所述对应的集群节点对所述目标标识信息对应的合并业务聚合数据进行还原处理,得到所述目标标识信息在各个业务类型下对应的业务数据。

[0085] 具体的,如表2,则为在处理货物管理数据的场景中,标识信息“型号a”在各个业务类型下对应的业务数据。

[0086]	型号 a	销售(类型)数据	销售信息 1	销售信息 2	销售信息 3		
			a1	a2	a3		
			a4	a5	a6		
					a7	a8	a9
		库存(类型)数据	库存信息 1	库存信息 2	库存信息 3		
			aa1	aa2	aa3		
			aa4	aa5	aa6		
					aa7	aa8	aa9
		价格(类型)数据	价格信息 1	价格信息 2	价格信息 3		
	aaa1		aaa2	aaa3			
	aaa4		aaa5	aaa6			
				aaa7	aaa8	aaa9	

[0087] 表2

[0088] 在步骤282中,由所述对应的集群节点调用与业务需求相匹配的运算代码。

[0089] 在步骤283中,将所述目标标识信息在各个业务类型下对应的业务数据作为输入数据,通过运行所述运算代码计算所述目标标识信息对应的业务预测数据。

[0090] 具体的,例如,在处理货物管理数据的场景中,业务预测数据可以包括销售预测数据,也可以包括库存预测数据,还可以包括价格预测数据。

[0091] 为了使本领域技术人员更好的理解本申请,下面将结合图8对本申请提出的业务数据处理方法进行说明。

[0092] 参见图8,示出了根据本申请一个实施例的业务数据流的示意图。

[0093] 如图8所示,在步骤801中,获取第一业务类型的业务数据集合和第二业务类型的业务数据集合。

[0094] 在步骤802中,在第一业务类型的业务数据集合和第二业务类型的业务数据集合中,分别对标识信息1和标识信息2的业务数据进行聚合,得到业务聚合数据。

[0095] 在步骤803中,按照标识信息1对第一业务类型对应的业务聚合数据和第二业务类型对应的业务聚合数据进行合并,且按照标识信息2对第一业务类型对应的业务聚合数据和第二业务类型对应的业务聚合数据进行合并。

[0096] 在步骤804中,基于步骤803中的得到数据计算标识信息1和标识信息2对应的业务预测数据。

[0097] 在本申请的一个实施例中,所提出的业务数据处理方法适用于基于Py Spark(即Spark套件中的Python编程接口)中Pandas UDF的编程方法,具体的,可以通过Spark将多个算法输入源(即多个类型的业务数据集合)读取成Dataframe后,并按照最小并行单元Key(即标识信息)进行分组折叠和序列化(即聚合),再按Key去做Join合并所有Dataframe得到全数据Dataframe,最后对全数据Dataframe执行Group by key apply的方式执行Pandas UDF来得到计算结果。

[0098] 具体的,在本实施例中,包括如下6个步骤:

[0099] 步骤1,确定最小并行单元Key(关键字),定义最小并行单元为算法设计中跑完一次算法流程需要最小数据单元。同时也是Pandas UDF做Group by的关键字。例如一个预测算法是按单个商品建模的,那最小并行单元是商品ID或者商品名称,如果一个预测算法是品类建模的,那最小并行单元即为品类ID或者品类名称。

[0100] 步骤2,通过Spark从文件或者数据库将多个输入数据读取为多个Dataframe。

[0101] 步骤3,通过设计的Pandas UDF函数对步骤2中读取的Dataframe进行折叠,这个函数中,会按最小并行单元将数据进行分组,并对分组后的数据进行序列化。每个经过折叠的Dataframe经过折叠函数都会变成一个2列n行的新Dataframe,一列并行单元key,一列是按并行单元序列化后的json字符串,这个json字符串中包含这个Dataframe中关于这个并行单元的所有信息;n为这个Dataframe中包含的并行单元key的个数。

[0102] 步骤4,所有的Dataframe都经过步骤3的合折叠后再以最小并行单元Key以join的方式合并成一大表,这一步会得到一个m+1列,n行的全数据Dataframe,m为步骤2中读取的Dataframe个数,另外一列为并行单元key;n包含的并行单元key的个数。

[0103] 步骤5,经过以上操作,全数据Dataframe中的一个单元就包含一个最小并行单元下某一类输入的所有数据,接下来只需要写一个全流程的,在数据获取部分使用pd.read_json.,函数去读取之前单元中序列化的json字符串即还原这个最小并行单元的数据,其格式为Pandas Dataframe,所有的算法主流程都可以写到这个Pandas UDF中,完全由python实现。

[0104] 步骤6,对全数据Dataframe按Key做Group by再apply步骤5中的pandas UDF,即可得到格式为Spark Dataframe的最终算法结果。

[0105] 通过这样的调度方式,Spark只承担了数据的读取处理和分发过程,具体的算法流程在Pandas UDF中,只需要付出很小的开发代价,就可以实现既使用Python中各种算法包,也利用了Spark的并行模式,加速算法,提高计算效率。

[0106] 在本申请的一些实施例所提供的技术方案中,通过对业务数据集中的具有相同标识信息的业务数据进行聚合,得到各个标识信息所对应的业务聚合数据,,然后再基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的于计算业务预测数据的合并业务聚合数据。由于通过综合考虑多个业务类型的业务数据集合,得到能够全面表征业务特征的合并业务聚合数据,从而使得能够通过合并业务聚合数据计算出更加准确的业务预测数据。

[0107] 以下介绍本申请的装置实施例,可以用于执行本申请上述实施例中的业务数据处理方法。对于本申请装置实施例中未披露的细节,请参照本申请上述的业务数据处理方法的实施例。

[0108] 图9示出了根据本申请一个实施例的业务数据处理装置的框图。

[0109] 参照图9所示,根据本申请一个实施例的业务数据处理装置900,包括:第一获取单元901、聚合单元902和合并单元903。

[0110] 其中,第一获取单元901,用于获取至少两个业务类型的业务数据集合,所述业务数据集合中包括多条业务数据,所述业务数据包括标识信息;聚合单元902,用于针对每一个目标业务数据集合,对所述目标业务数据集合中的具有相同标识信息的业务数据进行聚合,得到所述目标业务数据集合中的各个标识信息所对应的业务聚合数据,所述目标业务

数据集合为所述至少两个业务类型的业务数据集合中的任意一个;合并单元903,用于基于各个业务数据集合对应的业务聚合数据,对具有相同标识信息的业务聚合数据进行合并,得到各个标识信息对应的合并业务聚合数据,所述合并业务聚合数据用于计算业务预测数据。

[0111] 在本申请的一些实施例中,基于前述方案,所述聚合单元902配置为:将所述目标业务数据集合中的业务数据发送至分布式集群,并通过所述分布式集群对所述目标业务数据集合中具有相同标识信息的业务数据进行聚合。

[0112] 在本申请的一些实施例中,基于前述方案,所述聚合单元902进一步配置为:基于业务数据的标识信息,对所述目标业务数据集合中的业务数据进行分组,得到各个标识信息对应的业务数据组;通过所述分布式集群中的各个集群节点,并发对得到的各个业务数据组中的业务数据进行聚合。

[0113] 在本申请的一些实施例中,基于前述方案,所述业务数据还包括业务信息,所述聚合单元902进一步配置为:根据所述分布式集群中各个集群节点的负载,将所述各个标识信息对应的业务数据组分发至所述各个集群节点;针对所述分布式集群中的每一个集群节点,通过所述集群节点对所述业务数据组中各条业务数据的业务信息进行序列化聚合。

[0114] 在本申请的一些实施例中,基于前述方案,所述合并单元903配置为:在所述分布式集群的各个集群节点中确定具有相同标识信息的业务聚合数据;分别对各个标识信息对应的业务聚合数据进行拼接,得到各个标识信息对应的合并业务聚合数据。

[0115] 在本申请的一些实施例中,基于前述方案,所述装置还包括第二获取单元,被用于获取各个标识信息对应的合并业务聚合数据;分发单元,被用于根据分布式集群中各个集群节点的负载,将所述各个标识信息对应的合并业务聚合数据分发至所述分布式集群的各个集群节点;计算单元,被用于基于每一个目标标识信息对应的合并业务聚合数据,通过对应的集群节点计算所述目标标识信息对应的业务预测数据,所述目标标识信息为所述各个标识信息中的任意一个。

[0116] 在本申请的一些实施例中,基于前述方案,所述计算单元配置为:由所述对应的集群节点对所述目标标识信息对应的合并业务聚合数据进行还原处理,得到所述目标标识信息在各个业务类型下对应的业务数据;由所述对应的集群节点调用与业务需求相匹配的运算代码;将所述目标标识信息在各个业务类型下对应的业务数据作为输入数据,通过运行所述运算代码计算所述目标标识信息对应的业务预测数据。

[0117] 图10示出了适于用来实现本申请实施例的电子设备的计算机系统的结构示意图。

[0118] 需要说明的是,图10示出的电子设备的计算机系统1000仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0119] 如图10所示,计算机系统1000包括中央处理单元(Central Processing Unit, CPU) 1001,其可以根据存储在只读存储器(Read-Only Memory, ROM) 1002中的程序或者从缓存部分1008加载到随机访问存储器(Random Access Memory, RAM) 1003中的程序而执行各种适当的动作和处理,例如执行上述实施例中所述的方法。在RAM 1003中,还存储有系统操作所需的各种程序和数据。CPU 1001、ROM 1002以及RAM 1003通过总线1004彼此相连。输入/输出(Input/Output, I/O) 接口1005也连接至总线1004。

[0120] 以下部件连接至I/O接口1005:包括键盘、鼠标等的输入部分1006;包括诸如阴极

射线管 (Cathode Ray Tube, CRT)、液晶显示器 (Liquid Crystal Display, LCD) 等以及扬声器等的输出部分 1007; 包括硬盘等的储存部分 1008; 以及包括诸如 LAN (Local Area Network, 局域网) 卡、调制解调器等网络接口卡的通信部分 1009。通信部分 1009 经由诸如因特网的网络执行通信处理。驱动器 1010 也根据需要连接至 I/O 接口 1005。可拆卸介质 1011, 诸如磁盘、光盘、磁光盘、半导体存储器等等, 根据需要安装在驱动器 1010 上, 以便于从其上读出的计算机程序根据需要被安装入储存部分 1008。

[0121] 特别地, 根据本申请的实施例, 上文参考流程图描述的过程可以被实现为计算机软件程序。例如, 本申请的实施例包括一种计算机程序产品, 其包括承载在计算机可读介质上的计算机程序, 该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中, 该计算机程序可以通过通信部分 1009 从网络上被下载和安装, 和/或从可拆卸介质 1011 被安装。在该计算机程序被中央处理单元 (CPU) 1001 执行时, 执行本申请的系统中限定的各种功能。

[0122] 需要说明的是, 本申请实施例所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一一但不限于一一电、磁、光、电磁、红外线、或半导体的系统、装置或器件, 或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于: 具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器 (RAM)、只读存储器 (ROM)、可擦式可编程只读存储器 (Erasable Programmable Read Only Memory, EPROM)、闪存、光纤、便携式紧凑磁盘只读存储器 (Compact Disc Read-Only Memory, CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中, 计算机可读存储介质可以是任何包含或存储程序的有形介质, 该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中, 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号, 其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式, 包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质, 该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输, 包括但不限于: 无线、有线等等, 或者上述的任意合适的组合。

[0123] 附图中的流程图和框图, 图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。其中, 流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分, 上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意, 在有些作为替换的实现中, 方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如, 两个接连地表示的方框实际上可以基本并行地执行, 它们有时也可以按相反的顺序执行, 这依所涉及的功能而定。也要注意, 框图或流程图中的每个方框、以及框图或流程图中的方框的组合, 可以用执行规定的功能或操作的专用的基于硬件的系统来实现, 或者可以用专用硬件与计算机指令的组合来实现。

[0124] 描述于本申请实施例中所涉及到的单元可以通过软件的方式实现, 也可以通过硬件的方式来实现, 所描述的单元也可以设置在处理器中。其中, 这些单元的名称在某种情况

下并不构成对该单元本身的限定。

[0125] 作为另一方面,本申请还提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上述实施例中所述的业务数据处理方法。

[0126] 作为另一方面,本申请还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该电子设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该电子设备执行时,使得该电子设备实现上述实施例中所述的业务数据处理方法。

[0127] 应当注意,尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元,但是这种划分并非强制性的。实际上,根据本申请的实施方式,上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之,上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[0128] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本申请实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、触控终端、或者网络设备等)执行根据本申请实施方式的方法。

[0129] 本领域技术人员在考虑说明书及实践这里公开的实施方式后,将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。

[0130] 应当理解的是,本申请并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本申请的范围仅由所附的权利要求来限制。

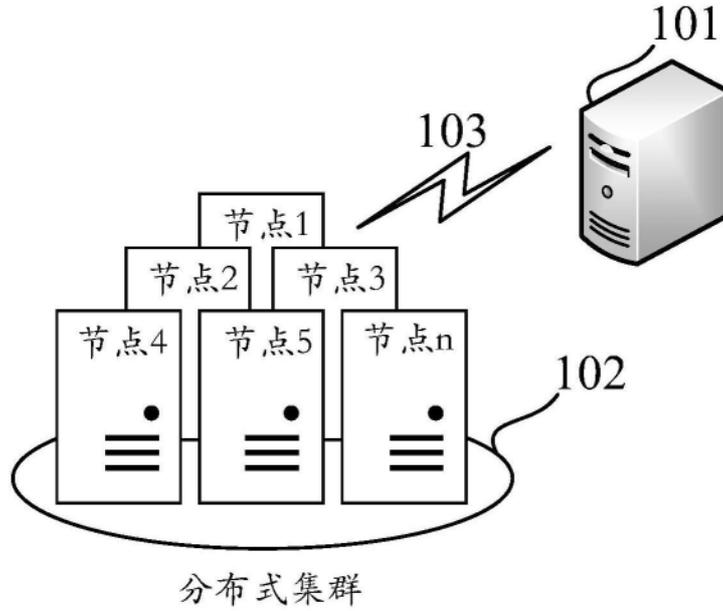


图1

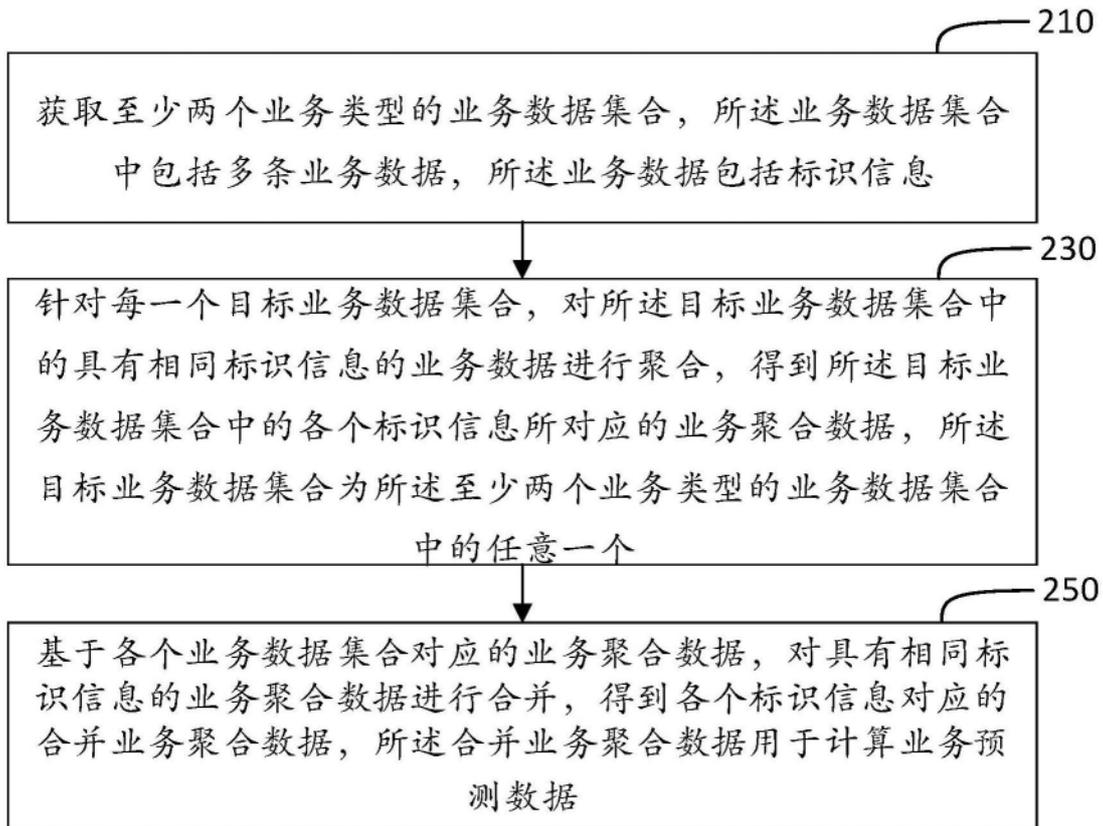


图2

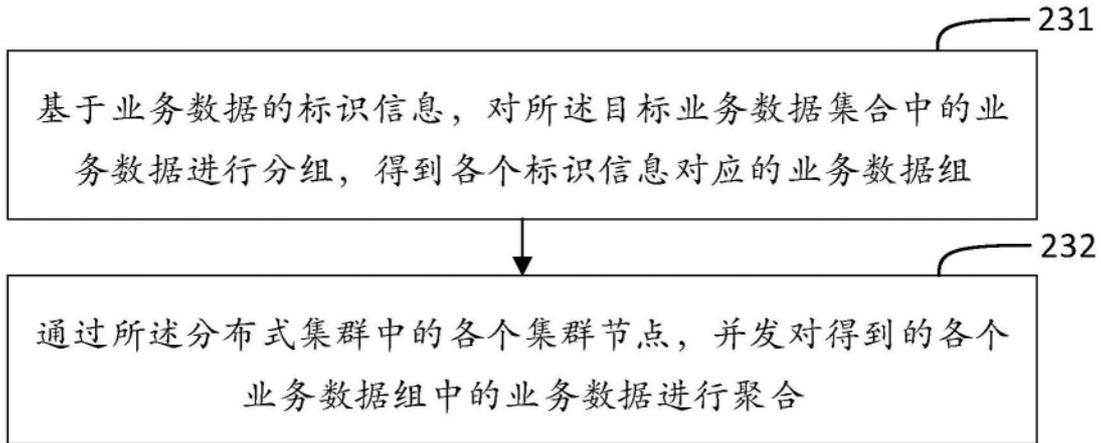


图3

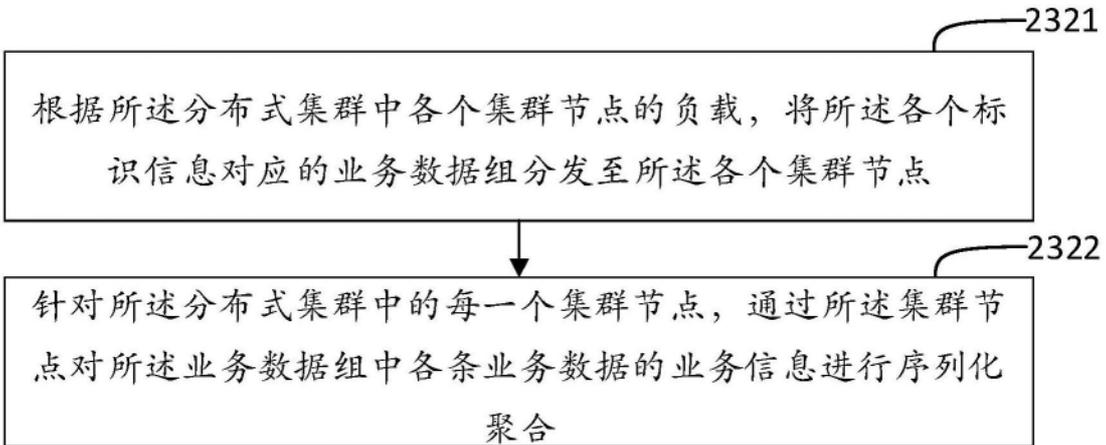


图4

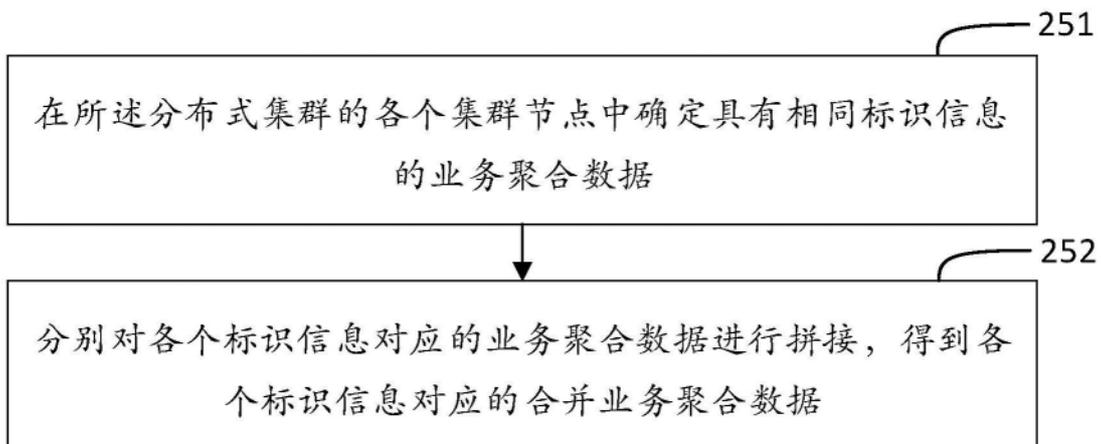


图5

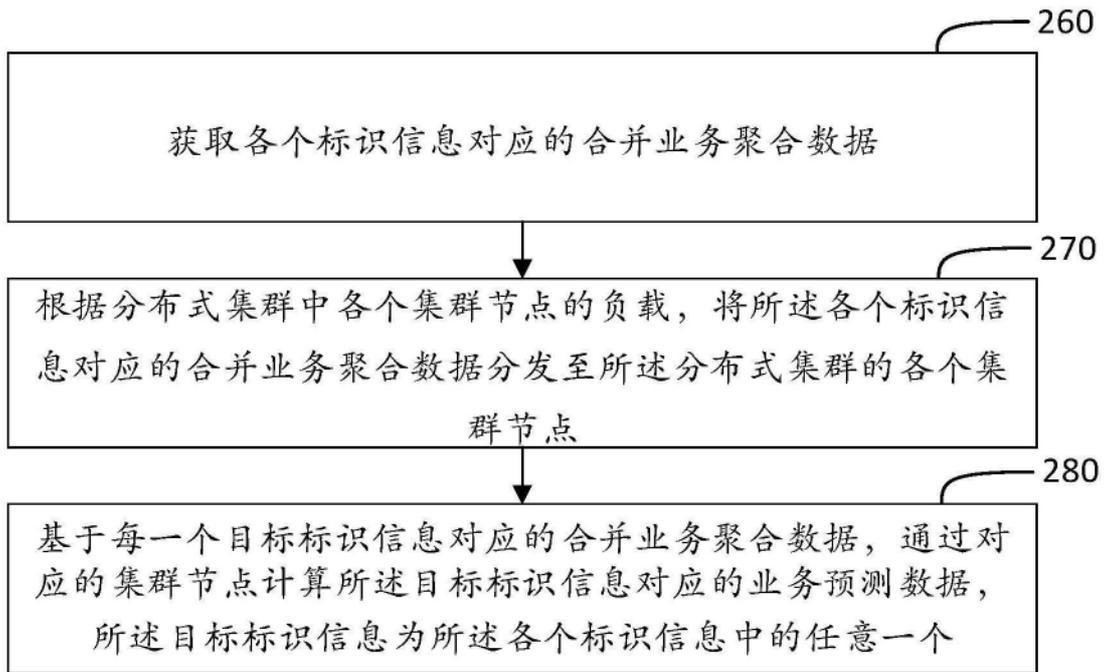


图6

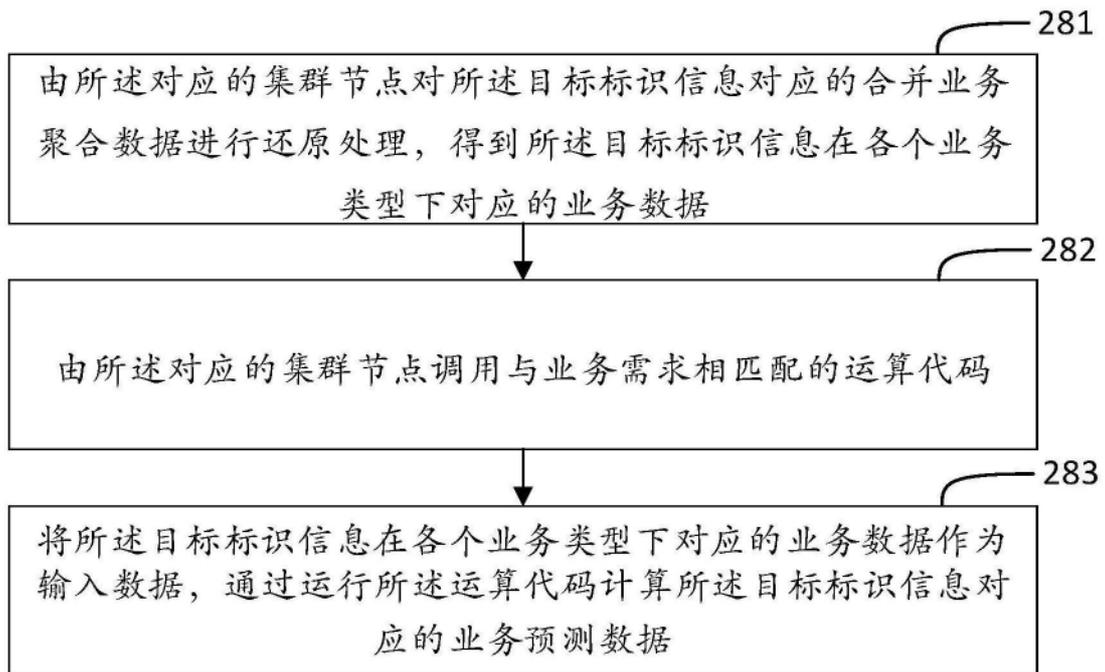


图7

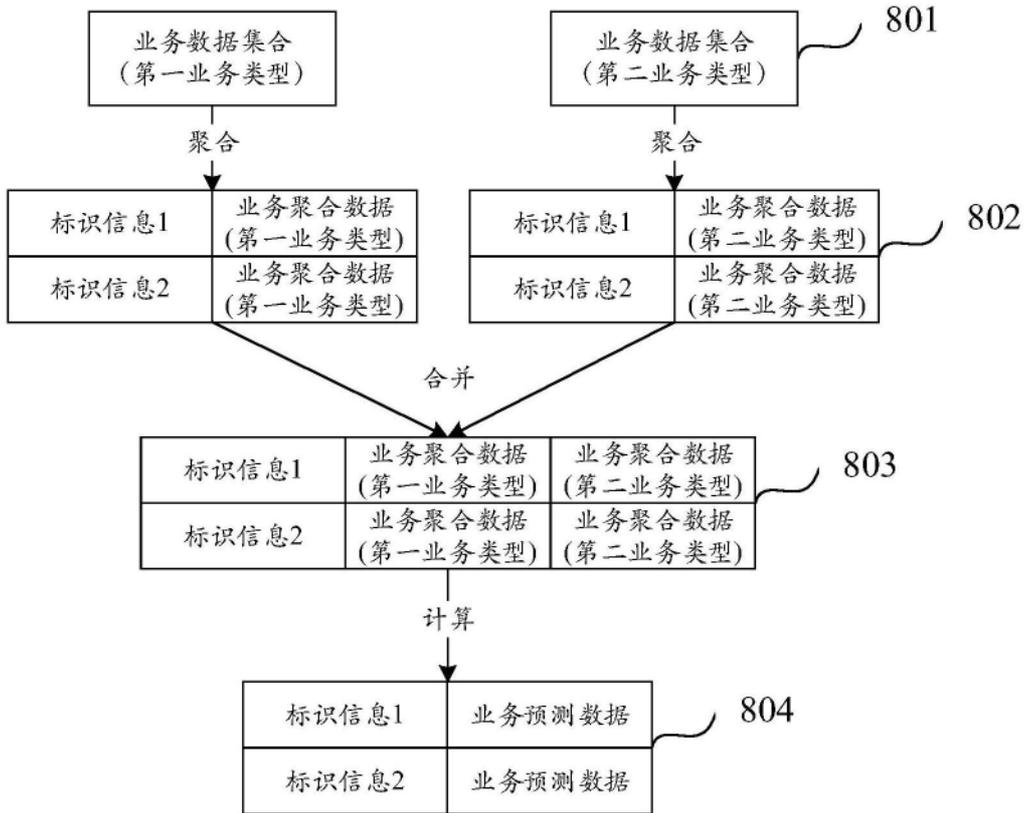


图8

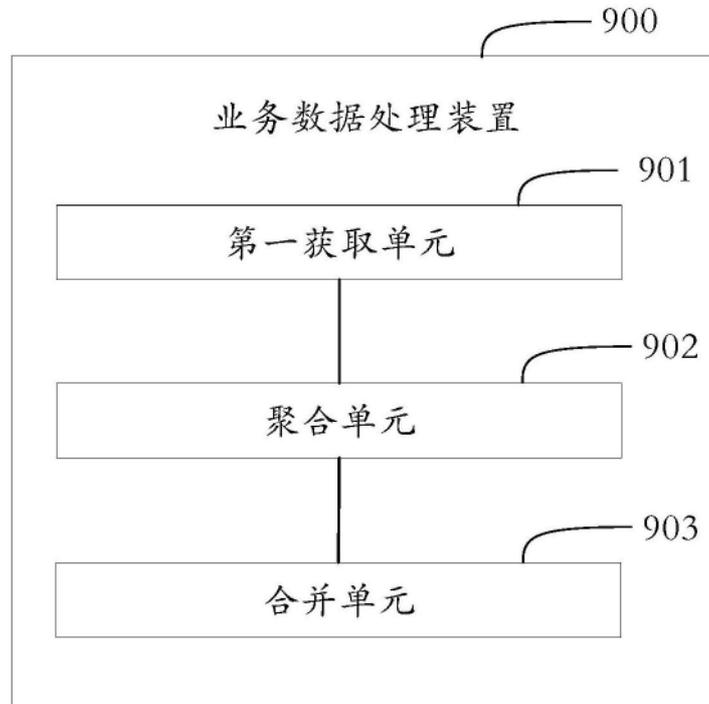


图9

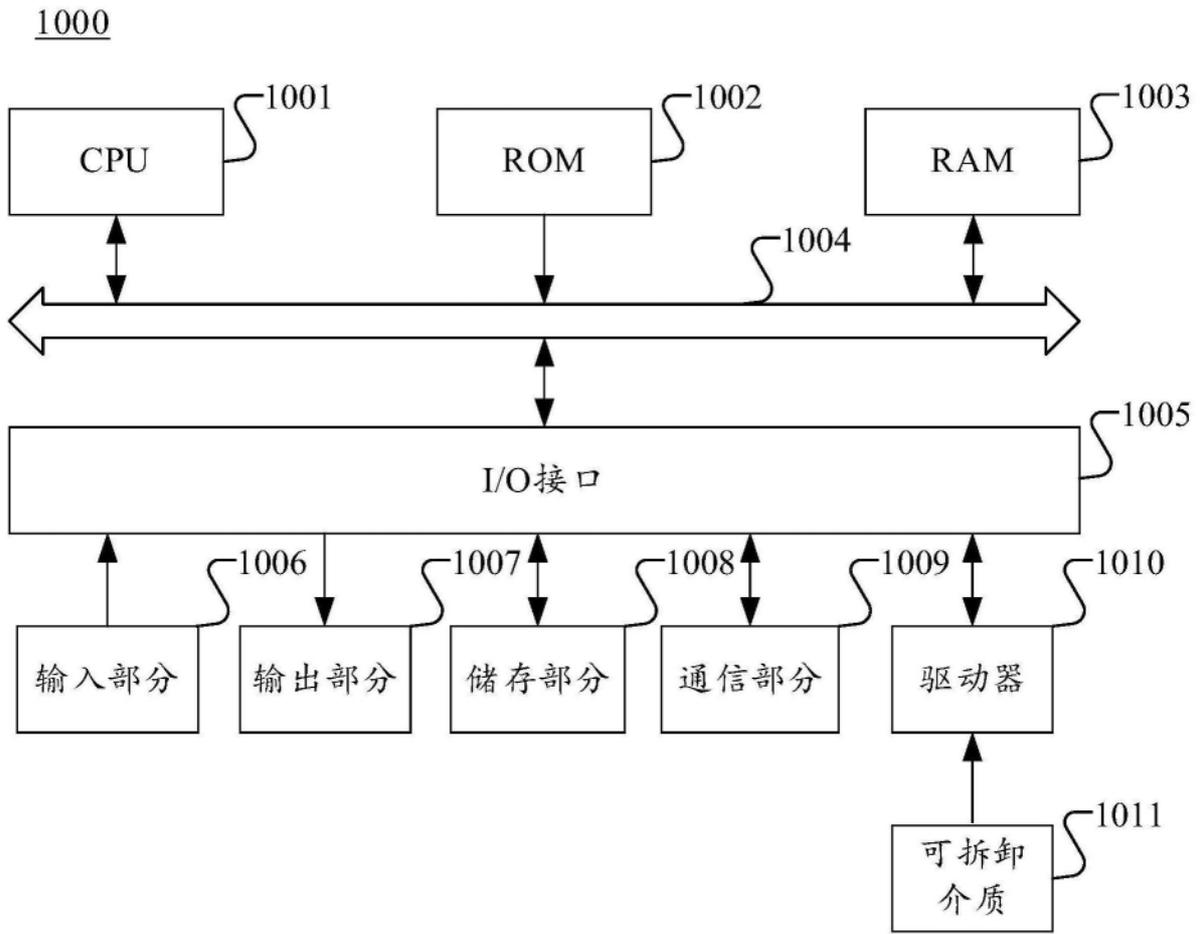


图10