



(12)发明专利申请

(10)申请公布号 CN 110830437 A

(43)申请公布日 2020.02.21

(21)申请号 201910908638.8

(22)申请日 2019.09.25

(71)申请人 平安科技(深圳)有限公司

地址 518000 广东省深圳市福田区福田街
道福安社区益田路5033号平安金融中
心23楼

(72)发明人 张鹏 李正洋

(74)专利代理机构 深圳市世联合知识产权代理
有限公司 44385

代理人 谷惠英

(51)Int.Cl.

H04L 29/06(2006.01)

H04L 29/08(2006.01)

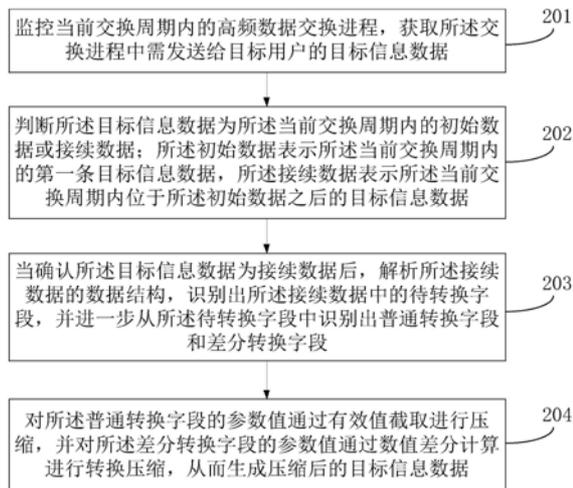
权利要求书2页 说明书12页 附图4页

(54)发明名称

高频业务数据的数据压缩方法、装置、设备
及存储介质

(57)摘要

本申请实施例公开了一种高频业务数据的数据压缩方法、装置、设备及存储介质,涉及大数据处理技术领域。该方法包括:获取目标高频业务数据;判断目标高频业务数据是否为接续数据;若为接续数据,识别出接续数据中的普通转换字段和差分转换字段;对普通转换字段的参数值通过有效值截取进行压缩,对差分转换字段的参数值通过数值差分计算进行转换压缩,生成压缩后的目标高频业务数据。所述方法对高频业务数据中的一部分数据进行截取,摒弃其中无用的数据,并通过应用差分计算的思想,将另一部分符合转换要求的数据的数值进行转换压缩,大幅减小了数据的存储空间和传输时的带宽要求,使服务器压力大幅降低,节省了服务器资源并提升了传输效率。



1. 一种高频业务数据的数据压缩方法,其特征在于,包括:

监控当前交换周期内的高频数据交换进程,获取所述交换进程中需发送给目标用户的目标高频业务数据;

判断所述目标高频业务数据是否为所述当前交换周期内的接续数据;其中,所述接续数据表示所述当前交换周期内位于初始数据之后的目标高频业务数据,所述初始数据表示所述当前交换周期内的第一条目标高频业务数据;

若所述目标高频业务数据为接续数据,解析所述接续数据的数据结构,识别出所述接续数据中的待转换字段,并进一步从所述待转换字段中识别出普通转换字段和差分转换字段;

对所述普通转换字段的参数值通过有效值截取进行压缩,并对所述差分转换字段的参数值通过数值差分计算进行转换压缩,从而生成压缩后的目标高频业务数据。

2. 根据权利要求1所述的高频业务数据的数据压缩方法,其特征在于,所述判断所述目标高频业务数据是否为所述当前交换周期内的接续数据的步骤包括:

识别出所述目标高频业务数据中用于表示数据状态类型的字段,检测该字段下的参数值;

若所述参数值为预设的第一参数判断值,确定所述目标高频业务数据为初始数据;

若所述参数值为预设的第二参数判断值,则确定所述目标高频业务数据为接续数据。

3. 根据权利要求1所述的高频业务数据的数据压缩方法,其特征在于,所述识别出所述接续数据中的待转换字段的步骤之后,所述方法还包括:

对所述接续数据的除所述待转换字段外的其他字段进行识别,以识别出非敏感字段;

将所述非敏感字段从所述接续数据中删除。

4. 根据权利要求1所述的高频业务数据的数据压缩方法,其特征在于,所述对所述普通转换字段的参数值通过有效值截取进行压缩的步骤包括:

截取每个所述普通转换字段的参数值中低位的16比特位数据或24比特位数据,以作为压缩后的目标信息数据中对应的普通转换字段的参数值。

5. 根据权利要求1所述的高频业务数据的数据压缩方法,其特征在于,在所述对所述差分转换字段的参数值通过数值差分计算进行转换压缩的步骤之前,所述方法还包括:

识别所述差分转换字段中各字段的参数值类型,将所述差分转换字段划分为资产数值差分字段和时间差分字段。

6. 根据权利要求5所述的高频业务数据的数据压缩方法,其特征在于,所述识别所述差分转换字段中各字段的参数值类型,将所述差分转换字段划分为资产数值差分字段和时间差分字段的步骤包括:

获取一个历史交换周期内的连续的若干个历史目标信息数据;

统计所述若干个历史目标信息数据中对应的差分转换字段中各字段的参数值,识别所述参数值的变化趋势;

若所述参数值的变化趋势为递增变化,则将该参数值对应的差分转换字段划分为时间差分字段;

若所述参数值的变化趋势为波动变化,则将该参数值对应的差分转换字段划分为资产数值差分字段。

7. 根据权利要求5所述的高频业务数据的数据压缩方法,其特征在于,所述对所述差分转换字段的参数值通过数值差分计算进行转换压缩的步骤包括:

读取最小变动资产数值和高频间隔时间;

计算所述接续数据和所述初始数据之间每个相同的资产数值差分字段下参数值的差值,由每个资产数值差分字段对应的差值除以所述最小变动资产数值后作为压缩后的接续数据对应的资产数值差分字段的参数值;

计算所述接续数据和所述初始数据之间每个相同的时间差分字段下参数值的差值,由每个时间差分字段对应的差值除以所述高频间隔时间后作为压缩后的所述接续数据对应的时间差分字段的参数值。

8. 一种高频业务数据的数据压缩装置,其特征在于,包括:

数据获取模块,用于监控当前交换周期内的高频数据交换进程,获取所述交换进程中需发送给目标用户的目标信息数据;

数据类型判断模块,用于判断所述目标信息数据为所述当前交换周期内的初始数据或接续数据;所述初始数据表示所述当前交换周期内的第一条目标信息数据,所述接续数据表示所述当前交换周期内位于所述初始数据之后的目标信息数据;

字段识别模块,用于当确认所述目标信息数据为接续数据后,解析所述接续数据的数据结构,识别出所述接续数据中的待转换字段,并进一步从所述待转换字段中识别出普通转换字段和差分转换字段;

数据压缩模块,用于对所述普通转换字段的参数值通过有效值截取进行压缩,并对所述差分转换字段的参数值通过数值差分计算进行转换压缩,从而生成压缩后的目标信息数据。

9. 一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器执行所述计算机程序时实现如权利要求1-7中任意一项所述的高频业务数据的数据压缩方法的步骤。

10. 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计算机程序,所述计算机程序被处理器执行时实现如权利要求1-7中任意一项所述的高频业务数据的数据压缩方法的步骤。

高频业务数据的数据压缩方法、装置、设备及存储介质

技术领域

[0001] 本申请涉及大数据处理技术领域,特别是一种高频业务数据的数据压缩方法、装置、设备及存储介质。

背景技术

[0002] 量化业务在处理时以先进的数学模型替代人为的主观判断,利用计算机技术从庞大的历史数据中海选能带来超额收益的多种“大概率”事件以制定策略,极大地减少了参与者情绪波动的影响,避免在市场极度狂热或悲观的情况下作出非理性的业务决策。高频业务则涉及人们无法利用的极为短暂的市场变化中寻求获利的计算机化业务。

[0003] 量化业务具有动态实时、客观、高低频可塑、免疫人主观判断情绪、心理波动等优点,近年来获得了快速发展。特别是高频量化业务由于其具有日内清算、频率高收益回测性能稳健等优点得到特别关注和发展。这里回测是指量化业务平台对设计好的程序化业务策略模拟历史真实的资产走势情况进行业务处理,得到历史真实情景再现后的结果数据以对策略的性能进行评估。

[0004] 由于高频业务数据交换的频率非常高,且策略回测时的样本数据十分庞大,存储数据时所需的存储空间较大,使得数据库压力大,硬件成本较高。

发明内容

[0005] 本申请实施例所要解决的技术问题是,提供一种高频业务数据的数据压缩方法、装置、设备及存储介质,对高频业务数据进行有效压缩,减小高频业务数据所需的存储空间。

[0006] 为了解决上述技术问题,本申请实施例所述的一种高频业务数据的数据压缩方法,采用了如下所述的技术方案:

[0007] 一种高频业务数据的数据压缩方法,包括:

[0008] 监控当前交换周期内的高频数据交换进程,获取所述交换进程中需发送给目标用户的目标高频业务数据;

[0009] 判断所述目标高频业务数据是否为所述当前交换周期内的接续数据;其中,所述接续数据表示所述当前交换周期内位于初始数据之后的目标高频业务数据,所述初始数据表示所述当前交换周期内的第一条目标高频业务数据;

[0010] 若所述目标高频业务数据为接续数据,解析所述接续数据的数据结构,识别出所述接续数据中的待转换字段,并进一步从所述待转换字段中识别出普通转换字段和差分转换字段;

[0011] 对所述普通转换字段的参数值通过有效值截取进行压缩,并对所述差分转换字段的参数值通过数值差分计算进行转换压缩,从而生成压缩后的目标高频业务数据。

[0012] 本申请实施例所述的高频业务数据的数据压缩方法,对高频业务数据中的一部分数据进行截取,摒弃其中无用的数据,并通过应用差分计算的思想,将另一部分符合转换要

求的数据的数值进行转换压缩,从而实现了高频业务数据的高度压缩,大幅减小了高频业务数据的存储空间和数据传输时的带宽要求,使服务器压力大幅降低,节省了服务器资源并提升了传输效率。

[0013] 进一步的,所述的高频业务数据的数据压缩方法,所述判断所述目标高频业务数据是否为所述当前交换周期内的接续数据的步骤包括:

[0014] 识别出所述目标高频业务数据中用于表示数据状态类型的字段,检测该字段下的参数值;

[0015] 若所述参数值为预设的第一参数判断值,确定所述目标高频业务数据为初始数据;

[0016] 若所述参数值为预设的第二参数判断值,则确定所述目标高频业务数据为接续数据。

[0017] 进一步的,所述的高频业务数据的数据压缩方法,所述识别出所述接续数据中的待转换字段的步骤之后,所述方法还包括:

[0018] 对所述接续数据的除所述待转换字段外的其他字段进行识别,以识别出非敏感字段;

[0019] 将所述非敏感字段从所述接续数据中删除。

[0020] 进一步的,所述的高频业务数据的数据压缩方法,所述对所述普通转换字段的参数值通过有效值截取进行压缩的步骤包括:

[0021] 截取每个所述普通转换字段的参数值中低位的16比特位数据或24比特位数据,以作为压缩后的目标信息数据中对应的普通转换字段的参数值。

[0022] 进一步的,所述的高频业务数据的数据压缩方法,在所述对所述差分转换字段的参数值通过数值差分计算进行转换压缩的步骤之前,所述方法还包括:

[0023] 识别所述差分转换字段中各字段的参数值类型,将所述差分转换字段划分为资产数值差分字段和时间差分字段。

[0024] 进一步的,所述的高频业务数据的数据压缩方法,所述识别所述差分转换字段中各字段的参数值类型,将所述差分转换字段划分为资产数值差分字段和时间差分字段的步骤包括:

[0025] 获取一个历史交换周期内的连续的若干个历史目标信息数据;

[0026] 统计所述若干个历史目标信息数据中对应的差分转换字段中各字段的参数值,识别所述参数值的变化趋势;

[0027] 若所述参数值的变化趋势为递增变化,则将该参数值对应的差分转换字段划分为时间差分字段;

[0028] 若所述参数值的变化趋势为波动变化,则将该参数值对应的差分转换字段划分为资产数值差分字段。

[0029] 进一步的,所述的高频业务数据的数据压缩方法,所述对所述差分转换字段的参数值通过数值差分计算进行转换压缩的步骤包括:

[0030] 读取最小变动资产数值和高频间隔时间;

[0031] 计算所述接续数据和所述初始数据之间每个相同的资产数值差分字段下参数值的差值,由每个资产数值差分字段对应的差值除以所述最小变动资产数值后作为压缩后的

接续数据对应的资产数值差分字段的参数值；

[0032] 计算所述接续数据和所述初始数据之间每个相同的时间差分字段下参数值的差值,由每个时间差分字段对应的差值除以所述高频间隔时间后作为压缩后的所述接续数据对应的时间差分字段的参数值。

[0033] 为了解决上述技术问题,本申请实施例还提供一种高频业务数据的数据压缩装置,采用了如下所述的技术方案:

[0034] 一种高频业务数据的数据压缩装置,包括:

[0035] 数据获取模块,用于监控当前交换周期内的高频数据交换进程,获取所述交换进程中需发送给目标用户的目标信息数据;

[0036] 数据类型判断模块,用于判断所述目标信息数据为所述当前交换周期内的初始数据或接续数据;所述初始数据表示所述当前交换周期内的第一条目标信息数据,所述接续数据表示所述当前交换周期内位于所述初始数据之后的目标信息数据;

[0037] 字段识别模块,用于当确认所述目标信息数据为接续数据后,解析所述接续数据的数据结构,识别出所述接续数据中的待转换字段,并进一步从所述待转换字段中识别出普通转换字段和差分转换字段;

[0038] 数据压缩模块,用于对所述普通转换字段的参数值通过有效值截取进行压缩,并对所述差分转换字段的参数值通过数值差分计算进行转换压缩,从而生成压缩后的目标信息数据。

[0039] 本申请实施例所述的高频业务数据的数据压缩装置,对高频业务数据中的一部分数据进行截取,摒弃其中无用的数据,并通过应用差分计算的思想,将另一部分符合转换要求的数据的数值进行转换压缩,从而实现了对接续数据的高度压缩,大幅减小了高频业务数据的存储空间和数据传输时的带宽要求,使服务器压力大幅降低,节省了服务器资源并提升了传输效率。

[0040] 为了解决上述技术问题,本申请实施例还提供一种计算机设备,采用了如下所述的技术方案:

[0041] 一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器执行所述计算机程序时实现如上述任意一项技术方案所述的高频业务数据的数据压缩方法的步骤。

[0042] 为了解决上述技术问题,本申请实施例还提供一种计算机可读存储介质,采用了如下所述的技术方案:

[0043] 一种计算机可读存储介质,所述计算机可读存储介质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任意一项技术方案所述的高频业务数据的数据压缩方法的步骤。

[0044] 与现有技术相比,本申请实施例主要有以下有益效果:

[0045] 本申请实施例公开了一种高频业务数据的数据压缩方法、装置、设备及存储介质,本申请实施例所述的高频业务数据的数据压缩方法,通过监控高频数据交换进程,获取需发送给目标用户的目标高频业务数据;再判断目标高频业务数据是否为接续数据;若为接续数据,识别出接续数据中的待转换字段,并进一步从待转换字段中识别出普通转换字段和差分转换字段;最终对普通转换字段的参数值通过有效值截取进行压缩,和对所述差分

转换字段的参数值通过数值差分计算进行转换压缩,从而生成压缩后的目标高频业务数据。所述方法对高频业务数据中的一部分数据进行截取,摒弃其中无用的数据,并通过应用差分计算的思想,将另一部分符合转换要求的数据的数值进行转换压缩,从而实现了对高频业务数据的高度压缩,大幅减小了高频业务数据的存储空间和数据传输时的带宽要求,使服务器压力大幅降低,节省了服务器资源并提升了传输效率。

附图说明

[0046] 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。

[0047] 图1为本申请实施例可以应用于其中的示例性系统架构图;

[0048] 图2为本申请实施例中所述高频业务数据的数据压缩方法的一个实施例的流程图;

[0049] 图3为本申请实施例中所述高频业务数据的数据压缩装置的一个实施例的结构示意图;

[0050] 图4为本申请实施例中计算机设备的一个实施例的结构示意图。

具体实施方式

[0051] 除非另有定义,本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中在本申请的说明书中所使用的术语只是为了描述具体的实施例的目的,不是旨在于限制本申请。

[0052] 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“包括”、“包含”和“具有”以及它们任何变形,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、产品或设备没有限定于已列出的步骤或单元,而是可选地还包括没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。在本申请的权利要求书、说明书以及说明书附图中的术语,诸如“第一”和“第二”等之类的关系术语仅仅用来将一个实体/操作/对象与另一个实体/操作/对象区分开来,而不一定要求或者暗示这些实体/操作/对象之间存在任何这种实际的关系或者顺序。

[0053] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其他实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其他实施例相结合。

[0054] 为了使本技术领域的人员更好地理解本申请的方案,下面将结合本申请实施例中的相关附图,对本申请实施例中的技术方案进行清楚、完整地描述。

[0055] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0056] 用户可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发

送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用,例如网页浏览器应用、购物类应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等。

[0057] 终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、电子书阅读器、MP3播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机和台式计算机等等。

[0058] 服务器105可以是提供各种服务的服务器,例如对终端设备101、102、103上显示的页面提供支持的后台服务器。

[0059] 需要说明的是,本申请实施例所提供的高频业务数据的数据压缩方法一般由服务器/终端设备执行,相应地,高频业务数据的数据压缩装置一般设置于服务器/终端设备中。

[0060] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0061] 继续参考图2,示出了本申请实施例中所述高频业务数据的数据压缩方法的一个实施例的流程图。所述高频业务数据的数据压缩方法,包括以下步骤:

[0062] 步骤201:监控当前交换周期内的高频数据交换进程,获取所述交换进程中需发送给目标用户的目标高频业务数据。

[0063] 高频业务数据交换和策略回测系统由于其自身的特点,对数据源采样率、数据质量有极高要求,大多要求tick级原数据存储。本申请实施例中,发送给目标用户的目标高频业务数据,一般指tick级的原数据。

[0064] Tick数据是一种存有数据交换时的交换明细的数据,高频数据交换的过程中,服务器以每隔n秒的时间截取一个快照数据作为tick数据发送给目标用户。一般而言,高频数据交换中n的取值为0.5,即服务器以每秒两次的频率向各目标用户发送tick级的目标高频业务数据,而tick级的目标高频业务数据存储时即以快照数据的结构存储在服务器的数据库中。

[0065] 高频数据交换时,由于其期间产生的数据交换量十分庞大,因此服务器一般预设一个交换周期,以该交换周期为单位统计该段时间内的交换数据。在本申请实施例中,所述交换周期一般设为一天。

[0066] 在本申请实施例中,所述高频业务数据的数据压缩方法运行于其上的电子设备(例如图1所示的服务器/终端设备)可以通过有线连接方式或者无线连接方式接收需发送给目标用户的目标高频业务数据。需要指出的是,上述无线连接方式可以包括但不限于3G/4G连接、WiFi连接、蓝牙连接、WiMAX连接、Zigbee连接、UWB(ultra wideband)连接、以及其他现在已知或将来开发的无线连接方式。

[0067] 步骤202:判断所述目标高频业务数据是否为所述当前交换周期内的接续数据;其中,所述接续数据表示所述当前交换周期内位于初始数据之后的目标高频业务数据,所述初始数据表示所述当前交换周期内的第一条目标高频业务数据。

[0068] 本申请中压缩所述目标高频业务数据时所采用的一种压缩手段是基于差分计算的思想。在对当前交换周期内的目标高频业务数据通过差分计算进行压缩时,具体是将当前交换周期下的第一条目标高频业务数据与其他目标高频业务数据进行差分计算,用差分

计算的结果代替其他目标高频业务数据中字段的参数值,基于差分计算后的参数值的长度相对于原其他目标高频业务数据的长度会大大减小。

[0069] 于是,首先设定所述当前交换周期内的第一条目标高频业务数据为初始数据,当前交换周期内的其他目标高频业务数据记为接续数据,判断所述目标高频业务数据为初始数据或是接续数据后,进一步执行后续步骤。其中,初始数据保存在数据库中和发送给目标用户时,是以原始的快照数据格式完整地保存和发送的。

[0070] 在本申请实施例中,所述步骤202包括:

[0071] 识别出所述目标高频业务数据中用于表示数据状态类型的字段,检测该字段下的参数值;

[0072] 若所述参数值为预设的第一参数判断值,确定所述目标高频业务数据为初始数据;

[0073] 若所述参数值为预设的第二参数判断值,则确定所述目标高频业务数据为接续数据。

[0074] 为了判断目标高频业务数据的数据状态类型为初始数据或是接续数据,可以预先在目标高频业务数据的数据结构下的参数中选择一个用于表示数据状态类型的参数,规定该参数下各具体参数值分别表示的不同数据状态类型,以通过设定该参数下的具体参数值的方式对其所属的数据状态类型进行标记。

[0075] 通过举例进一步理解,在目标高频业务数据的快照数据结构下,可以通过设定如以order_type这一具体字段表示该目标高频业务数据代表的数据状态类型,具体地,order_type字段下取2比特位的具体数值表示不同的数据状态类型,如在一种具体实施方式中,可以设定最高位的2比特位的数值为11时表示初始数据,数值为00时表示接续数据。

[0076] 如此通过识别order_type这一字段下的具体参数值,即可识别所述目标高频业务数据的数据状态类型,判断其为初始数据或是接续数据。

[0077] 步骤203:若所述目标高频业务数据为接续数据,解析所述接续数据的数据结构,识别出所述接续数据中的待转换字段,并进一步从所述待转换字段中识别出普通转换字段和差分转换字段。

[0078] 当识别出所述目标高频业务数据为当前交换周期内的初始数据时,会以原始的快照数据格式完整地保存和发送,不会对其数据大小进行压缩,因此本申请中所述方法的后续实施步骤只在目标高频业务数据为接续数据时进一步执行。

[0079] 在本申请的一些实施例中,对目标高频业务数据下的字段类型进行划分时,首先将部分字段划分为无需进行压缩处理的字段,另一部分便为需要进行压缩处理的字段,需要进行压缩处理的字段即为所述待转换字段。而对同一目标高频业务数据的有些不同字段需要采用不同的压缩手段,因此在确定所述目标高频业务为接续数据,并进一步识别出需要进行压缩处理的待转换字段后,还需要区分待转换字段的类型,以实施不同的压缩手段。

[0080] 本申请实施例中,将待转换字段中需通过差分计算的方式进行压缩处理的字段视为差分转换字段,除差分转换字段外的字段即可视为普通转换字段。

[0081] 目标用户会预先基于目标高频业务数据的快照数据格式进行判断,根据其用户需求设定好快照数据格式中的哪些字段为需要进行转换的字段,以及需进行转换的字段下哪些为普通转换字段哪些为差分转换字段。

[0082] 在一种具体实施方式中,目标用户可以配置一个包含其转换需求的转换指令,在识别待转换字段和从待转换字段中识别出普通转换字段和差分转换字段时,通过接收所述转换指令,并解析出该转换指令中的目标用户的转换需求,便能确认接续数据中的字段哪些属于待转换字段,而哪些又属于待转换字段下的普通转换字段和差分转换字段。

[0083] 在本申请的一些实施方式中,在步骤203之后,所述高频业务数据的数据压缩方法还包括:

[0084] 对所述接续数据的除所述待转换字段外的其他字段进行识别,以识别出非敏感字段;

[0085] 将所述非敏感字段从所述接续数据中删除。

[0086] 在接续数据的数据结构中的各个字段,除所述待转换字段外,剩下的即视为无需进行压缩处理的字段。而无需进行压缩处理的字段中,存在部分字段为目标用户所不感兴趣或不需要的数据,由于目标用户对这些数据不敏感,因此发送给目标用户并无意义,将所述非敏感字段代表的从接续数据中删除,能够有效地压缩数据,减小数据大小。

[0087] 步骤204:对所述普通转换字段的参数值通过有效值截取进行压缩,并对所述差分转换字段的参数值通过数值差分计算进行转换压缩,从而生成压缩后的目标高频业务数据。

[0088] 在本申请实施例中,快照数据结构中各个字段的数据类型均为双精度浮点型,即长度为8字节(64位)。但对于当前交换周期内的快照数据,其各字段下的参数值,若表示无符号的整形数据时,一般2字节便完全能够满足当前的需要了,而若需要表示带符号的整形数据,则以3字节的长度表示相应的数据即可。

[0089] 所述接续数据的待转换字段中的普通转换字段,一般而言,其数值大小无需用到参数值的全部64比特位进行表示,因此对于普通转换字段,通过截取其比特位中的有效数值,去掉其中以0表示的无效比特位,便能对普通转换字段的参数值进行有效压缩。

[0090] 而所述接续数据的待转换字段中的差分转换字段,一般而言,其数值大小同样无需用到参数值的全部64比特位,通过与初始数据中相同字段的参数值进行差分计算,计算出相对数值,以该相对数值代替差分转换字段的原参数值,可以有效减小差分转换字段的长度。

[0091] 在本申请实施例的一种具体实施方式中,有效值截取和差分计算的具体操作方式便可以预先设置在压缩规则中,通过调用压缩规则,压缩目标高频业务数据时基于该压缩规则即可。

[0092] 在一种具体的实施方式中,进一步理解上述步骤中的压缩过程,如order_type字段可以在步骤203中划分至所述普通转换字段,在步骤204中对该字段进行压缩时,可以只需截取其最高位的2比特位的数值,如此进行压缩后的order_type字段的长度便为1字节。而为了在实施所述高频业务数据的数据压缩方法后,方便目标用户定位至该字段以判断接收到的目标高频业务数据为初始数据或接续数据,可以移动order_type字段将其作为目标高频业务数据的第一个字段。

[0093] 在本申请的一些实施例中,步骤204中所述对所述普通转换字段的参数值通过有效值截取进行压缩的步骤包括:

[0094] 截取每个所述普通转换字段的参数值中低位的16比特位数据或24比特位数据,以

作为压缩后的目标信息数据中对应的普通转换字段的参数值。

[0095] 所述普通转换字段下的参数值,用于表示无符号的整形数据时,其2字节(64比特)中高位的48比特位的数值均为0,截掉上述无效数据,保留低位的16比特位下的数值作为普通转换字段的参数值,从而能不影响数据正常使用的同时有效对普通转换字段的数据长度进行压缩。

[0096] 所述普通转换字段下的参数值,用于表示带符号的整形数据时,最多只需截取低位3字节(即24比特位)的数值作为普通转换字段的参数值,以对普通转换字段进行压缩。

[0097] 在本申请的一些实施例中,在步骤204中所述对所述差分转换字段的参数值通过数值差分计算进行转换压缩的步骤之前,所述高频业务数据的数据压缩方法还包括:

[0098] 识别所述差分转换字段中各字段的参数值类型,将所述差分转换字段划分为资产数值差分字段和时间差分字段。

[0099] 对于差分转换字段参数值的压缩处理,本申请实施例中采用两种基于差分计算的压缩方式,其中一种是对差分转换字段的参数值表示的数据是资产数值(如价格)时的情况,另一种是差分转换字段的参数值表示的数据是时间时的情况,因此首先需在识别出的差分转换字段中,进一步确认哪些是资产数值差分字段,哪些是时间差分字段。

[0100] 在进一步的一种具体实施方式中,所述识别所述差分转换字段中各字段的参数值类型,将所述差分转换字段划分为资产数值差分字段和时间差分字段的步骤包括:

[0101] 获取一个历史交换周期内的连续的若干个历史目标信息数据;

[0102] 统计所述若干个历史目标信息数据中对应的差分转换字段中各字段的参数值,识别所述参数值的变化趋势;

[0103] 若所述参数值的变化趋势为递增变化,则将该参数值对应的差分转换字段划分为时间差分字段;

[0104] 若所述参数值的变化趋势为波动变化,则将该参数值对应的差分转换字段划分为资产数值差分字段。

[0105] Tick数据中各字段下的参数值,若是表示时间的参数值,在高频数据交换进程中,其变化的方式会是连续不断趋于递增的,但表示资产数值的参数值,其会在一个数值上限和一个数值下限之间不断无规律的上下波动。因此通过这种方式可以较有效地判断差分转换字段的类型,尤其是参考的数据量较大时,判断的准确率会非常高。

[0106] 在进一步的一种具体实施方式中,所述对所述差分转换字段的参数值通过数值差分计算进行转换压缩的步骤包括:

[0107] 读取最小变动资产数值和高频间隔时间;

[0108] 计算所述接续数据和所述初始数据之间每个相同的资产数值差分字段下参数值的差值,由每个资产数值差分字段对应的差值除以所述最小变动资产数值后作为压缩后的接续数据对应的资产数值差分字段的参数值;

[0109] 计算所述接续数据和所述初始数据之间每个相同的时间差分字段下参数值的差值,由每个时间差分字段对应的差值除以所述高频间隔时间后作为压缩后的所述接续数据对应的时间差分字段的参数值。

[0110] 高频业务数据交换中,资产数值波动都有一个最小的计算波动单位,以这个单位的整数倍进行资产数值涨跌波动,本提案中以预设的最小变动资产数值表示所述资产数值

波动时的最小的计算波动单位。而高频业务数据交换中资产数值每天有涨跌停的限制,所以有限的所述交换周期内资产数值波动范围是有限的离散的资产数值区间。这个区间或者这个能波动的范围比较小,所述目标高频业务数据中的差分转换字段的参数值是以8字节的双精度浮点型来表示资产数值数据。

[0111] 通过将接续数据同初始数据进行差分,计算所述资产数值差分字段的参数值与初始数据中相同的资产数值差分字段的参数值的差值,再除以所述最小变动资产数值得到散点数值,以散点数值作为接续数据下对应的资产数值差分字段的参数值,从而对所述接续数据进行压缩。而基于该散点数值和已知的初始数据,能较容易地推断出压缩前的接续数据中资产数值差分字段下的参数值,如此显然无需用到8字节的长度存储相关资产数值数据,能大大减小数据长度,节省存储空间。

[0112] 前述的内容中,服务器以每隔n秒的时间截取一个快照数据作为tick数据发送给目标用户,对于资产数值差分字段,本申请中所述高频间隔时间即指由该n秒表示的将数据发送给目标用户的间隔时间,一般而言,高频业务数据交换中n的取值为0.5。

[0113] 同样将接续数据同初始数据进行差分,计算所述时间差分字段的参数值与初始数据中相同的时间差分字段的参数值的差值,得到一个相对时间后以该相对时间除以所述高频间隔时间,得到最终的时间偏移量作为接续数据下对应的时间差分字段下的参数值,从而对所述接续数据进行压缩。而基于该时间偏移量和已知的初始数据,能较容易地推断出压缩前的接续数据中价格差分字段下的参数值,如此同样无需用到8字节的长度存储相关时间数据,能大大减小数据长度,节省存储空间。

[0114] 本申请的具体实施方式中,一般压缩后的资产数值差分字段和时间差分字段的长度均为2字节,2字节对应的十进制的最大数据量为65536,根据高频业务数据交换的数据统计经验,2字节表示的数据量已足够满足大部分高频业务数据交换的场景。但对于一些高频业务数据交换场景,资产数值差分字段中的参数值可能超过2字节的最大数据量,此时会以3字节的数据长度表示相应的数据。

[0115] 基于本申请中的tick数据压缩方法,对上述实施例中将字段长度压缩为3字节的高频业务数据交换场景,数据量可由一个tick数据的120字节压缩到了14字节,压缩程度约为8.5倍;而对于上述实施例中将字段长度压缩为2字节的高频业务数据交换场景,数据量可由一个tick数据的120字节压缩到了11字节,压缩程度约为10.9倍。

[0116] 本申请实施例所述的高频业务数据的数据压缩方法,对高频业务数据中的一部分数据进行截取,摒弃其中无用的数据,并通过应用差分计算的思想,将另一部分符合转换要求的数据的数值进行转换压缩,从而实现了对高频业务数据的高度压缩,大幅减小了高频业务数据的存储空间和数据传输时的带宽要求,使服务器压力大幅降低,节省了服务器资源并提升了传输效率。

[0117] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,该计算机程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,前述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory,ROM)等非易失性存储介质,或随机存储记忆体(Random Access Memory,RAM)等。

[0118] 应该理解的是,虽然附图的流程图中的各个步骤按照箭头的指示依次显示,但是

这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,其可以以其他的顺序执行。而且,附图的流程图中的至少一部分步骤可以包括多个子步骤或者多个阶段,这些子步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,其执行顺序也不必然是依次进行,而是可以与其他步骤或者其他步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

[0119] 进一步参考图3,图3示出了为本申请实施例中所述高频业务数据的数据压缩装置的一个实施例的结构示意图。作为对上述图2所示方法的实现,本申请提供了一种高频业务数据的数据压缩装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0120] 如图3所示,本实施例所述的高频业务数据的数据压缩装置包括:

[0121] 数据获取模块301;用于监控当前交换周期内的高频数据交换进程,获取所述交换进程中需发送给目标用户的目标信息数据。

[0122] 数据类型判断模块302;用于判断所述目标信息数据为所述当前交换周期内的初始数据或接续数据;所述初始数据表示所述当前交换周期内的第一条目标信息数据,所述接续数据表示所述当前交换周期内位于所述初始数据之后的目标信息数据。

[0123] 字段识别模块303;用于当确认所述目标信息数据为接续数据后,解析所述接续数据的数据结构,识别出所述接续数据中的待转换字段,并进一步从所述待转换字段中识别出普通转换字段和差分转换字段。

[0124] 数据压缩模块304;用于对所述普通转换字段的参数值通过有效值截取进行压缩,并对所述差分转换字段的参数值通过数值差分计算进行转换压缩,从而生成压缩后的目标信息数据。

[0125] 在本申请实施例中,所述数据类型判断模块202用于识别出所述目标高频业务数据中用于表示数据状态类型的字段,检测该字段下的参数值;若所述参数值为预设的第一参数判断值,确定所述目标高频业务数据为初始数据;若所述参数值为预设的第二参数判断值,则确定所述目标高频业务数据为接续数据。

[0126] 在本申请的一些实施方式中,所述高频业务数据的数据压缩装置还包括:非敏感字段识别模块。在所述字段识别模块203识别出所述接续数据中的待转换字段之后,所述非敏感字段识别模块用于对所述接续数据的除所述待转换字段外的其他字段进行识别,以识别出非敏感字段;将所述非敏感字段从所述接续数据中删除。

[0127] 在本申请的一些实施例中,所述数据压缩模块204包括:截取压缩子模块。所述截取压缩子模块用于截取每个所述普通转换字段的参数值中低位的16比特位数据或24比特位数据,以作为压缩后的目标信息数据中对应的普通转换字段的参数值。

[0128] 在本申请的一些实施例中,所述数据压缩模块204包括:差分字段分类子模块。在所述数据压缩模块204对所述差分转换字段的参数值通过数值差分计算进行转换压缩之前,所述差分字段分类子模块用于识别所述差分转换字段中各字段的参数值类型,将所述差分转换字段划分为资产数值差分字段和时间差分字段。

[0129] 在进一步的一种具体实施方式中,所述差分字段分类子模块用于获取一个历史交换周期内的连续的若干个历史目标信息数据;统计所述若干个历史目标信息数据中对应的差分转换字段中各字段的参数值,识别所述参数值的变化趋势;若所述参数值的变化趋势

为递增变化,则将该参数值对应的差分转换字段划分为时间差分字段;若所述参数值的变化趋势为波动变化,则将该参数值对应的差分转换字段划分为资产数值差分字段。

[0130] 在进一步的一种具体实施方式中,所述数据压缩模块204还包括:差分转换子模块。所述差分转换子模块用于读取最小变动资产数值和低频间隔时间;计算所述接续数据和所述初始数据之间每个相同的资产数值差分字段下参数值的差值,由每个资产数值差分字段对应的差值除以所述最小变动资产数值后作为压缩后的接续数据对应的资产数值差分字段的参数值;计算所述接续数据和所述初始数据之间每个相同的时间差分字段下参数值的差值,由每个时间差分字段对应的差值除以所述低频间隔时间后作为压缩后的所述接续数据对应的时间差分字段的参数值。

[0131] 本申请实施例所述的高频业务数据的数据压缩装置,对高频业务数据中的一部分数据进行截取,摒弃其中无用的数据,并通过应用差分计算的思想,将另一部分符合转换要求的数据的数值进行转换压缩,从而实现了对高频业务数据的高度压缩,大幅减小了高频业务数据的存储空间和数据传输时的带宽要求,使服务器压力大幅降低,节省了服务器资源并提升了传输效率。

[0132] 为解决上述技术问题,本申请实施例还提供计算机设备。具体请参阅图4,图4为本实施例计算机设备基本结构框图。

[0133] 所述计算机设备6包括通过系统总线相互通信连接存储器61、处理器62、网络接口63。需要指出的是,图中仅示出了具有组件61-63的计算机设备6,但是应理解的是,并不要求实施所有示出的组件,可以替代的实施更多或者更少的组件。其中,本技术领域技术人员可以理解,这里的计算机设备是一种能够按照事先设定或存储的指令,自动进行数值计算和/或信息处理的设备,其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit,ASIC)、可编程门阵列(Field-Programmable Gate Array,FPGA)、数字处理器(Digital Signal Processor,DSP)、嵌入式设备等。

[0134] 所述计算机设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机设备可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

[0135] 所述存储器61至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如,SD或DX存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中,所述存储器61可以是所述计算机设备6的内部存储单元,例如该计算机设备6的硬盘或内存。在另一些实施例中,所述存储器61也可以是所述计算机设备6的外部存储设备,例如该计算机设备6上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。当然,所述存储器61还可以既包括所述计算机设备6的内部存储单元也包括其外部存储设备。本实施例中,所述存储器61通常用于存储安装于所述计算机设备6的操作系统和各类应用软件,例如高频业务数据的数据压缩方法的程序代码等。此外,所述存储器61还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0136] 所述处理器62在一些实施例中可以是中央处理器(Central Processing Unit,CPU)、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器62通常用于控制所述计

计算机设备6的总体操作。本实施例中,所述处理器62用于运行所述存储器61中存储的程序代码或者处理数据,例如运行所述高频业务数据的数据压缩方法的程序代码。

[0137] 所述网络接口63可包括无线网络接口或有线网络接口,该网络接口63通常用于在所述计算机设备6与其他电子设备之间建立通信连接。

[0138] 本申请还提供了另一种实施方式,即提供一种计算机可读存储介质,所述计算机可读存储介质存储有高频业务数据的数据压缩程序,所述高频业务数据的数据压缩程序可被至少一个处理器执行,以使所述至少一个处理器执行如上述的高频业务数据的数据压缩方法的步骤。

[0139] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。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(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本申请各个实施例所述的方法。

[0140] 在本申请所提供的上述实施例中,应该理解到,所揭露的装置和方法,可以通过其他的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅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,多个模块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。

[0141] 所述模块或组件可以是或者也可以不是物理上分开的,作为模块或组件显示的部件可以是或者也可以不是物理模块,既可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块或组件来实现本实施例方案的目的。

[0142] 本申请不限于上述实施方式,以上所述是本申请的优选实施方式,该实施例仅用于说明本申请而并不用于限制本申请的范围,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,其依然可以对前述各具体实施方式所记载的技术方案进行若干改进和修饰,或者对其中部分技术特征进行等效替换。凡是利用本申请说明书及附图内容所做的等效结构,直接或间接运用在其他相关的技术领域,均同理应视为包括在本申请的保护范围之内。

[0143] 显然,以上所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例,附图中给出了本申请的较佳实施例,但并不限制本申请的专利范围。本申请可以以许多不同的形式来实现,相反地,提供这些实施例的目的是使对本申请的公开内容的理解更加透彻全面。尽管参照前述实施例对本申请进行了详细的说明,对于本领域的技术人员来而言,其依然可以对前述各具体实施方式所记载的技术方案进行修改,或者对其中部分技术特征进行等效替换。基于本申请中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,以及凡是利用本申请说明书及附图内容所做的等效结构,直接或间接运用在其他相关的技术领域,均同理在本申请专利保护范围之内。

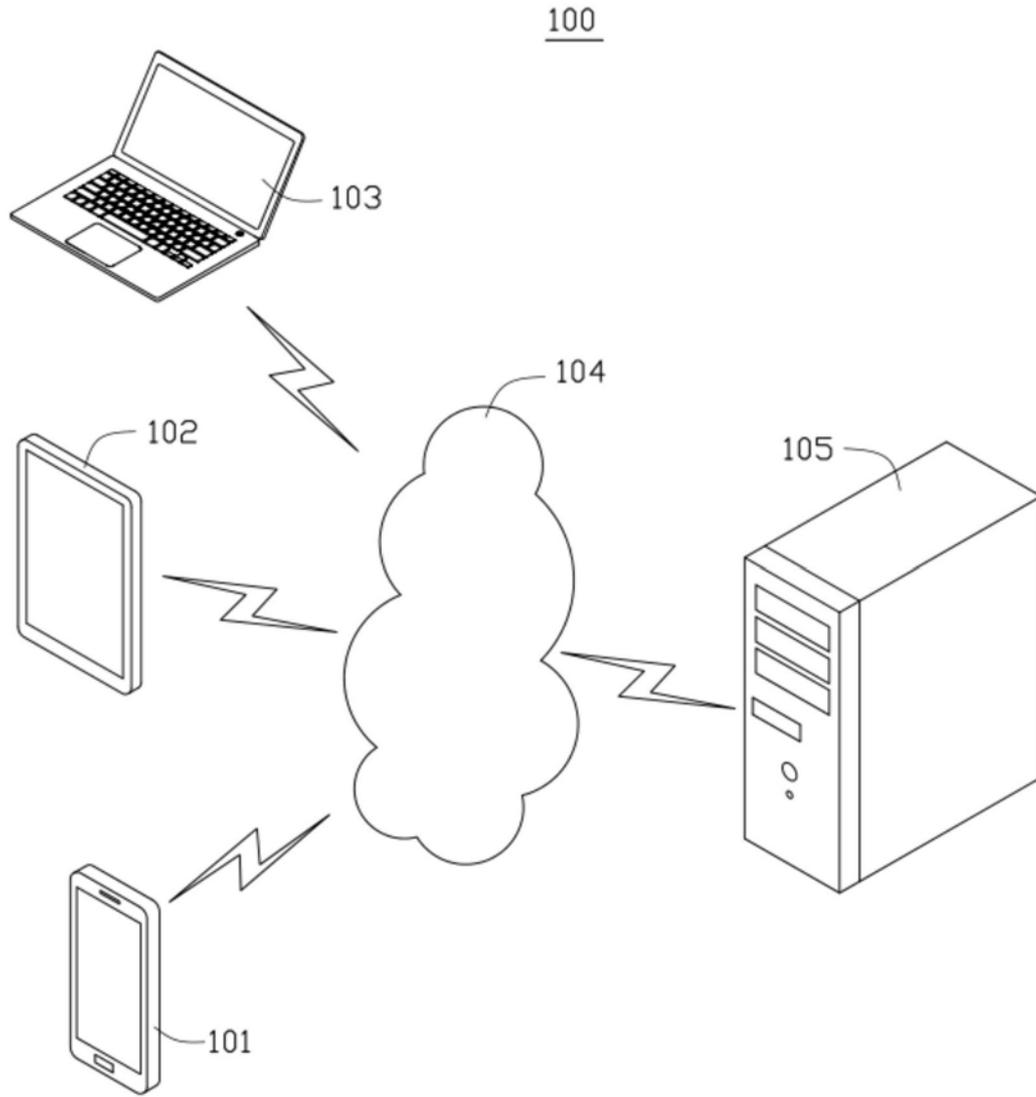


图1

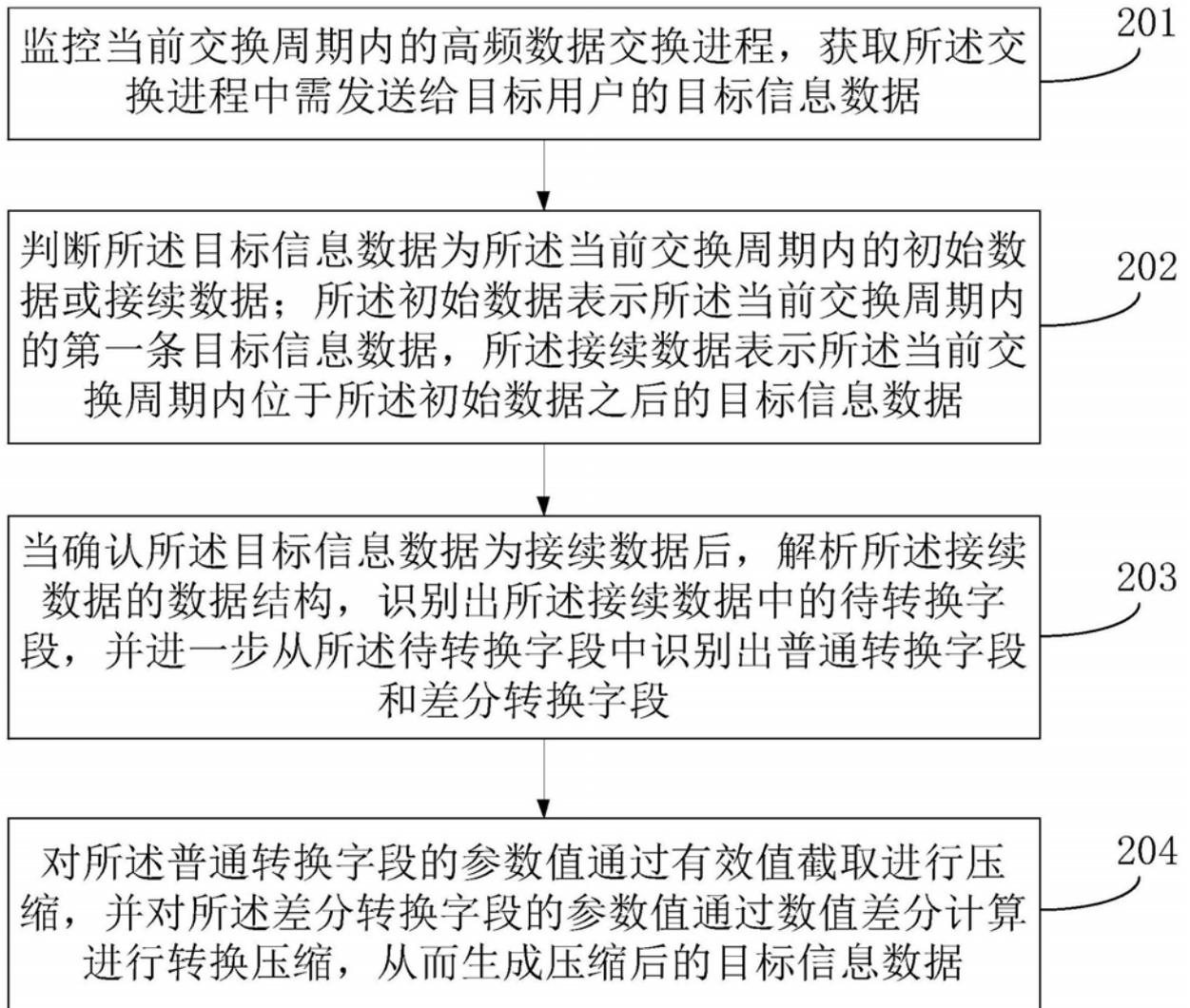


图2

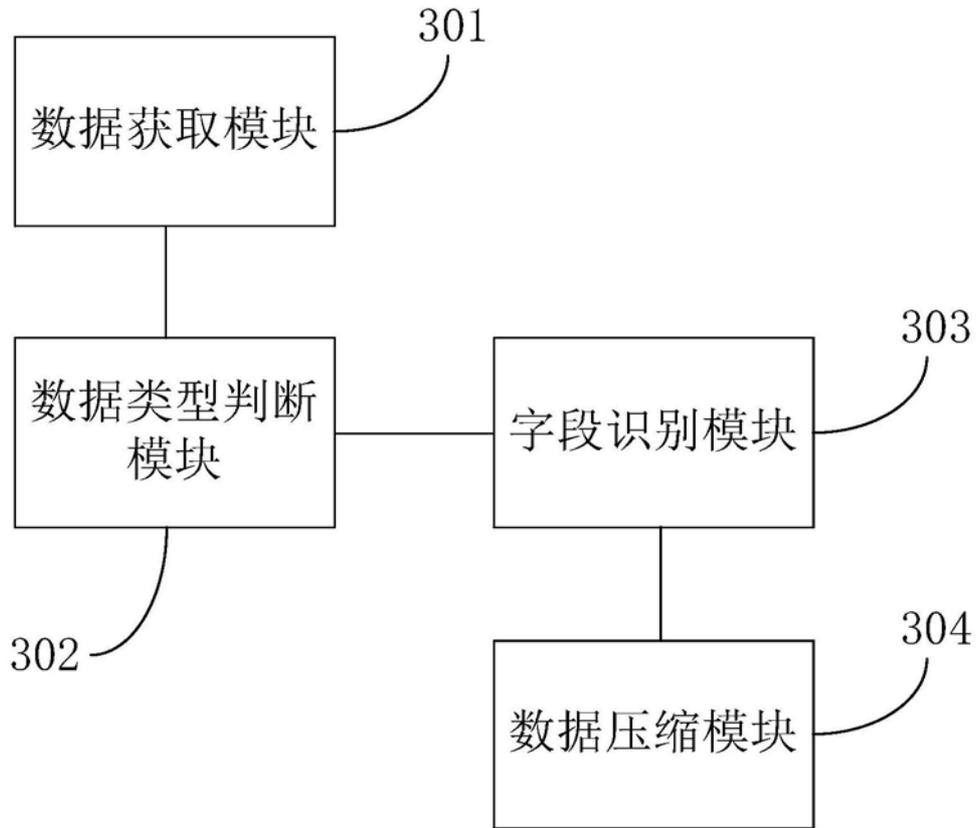


图3

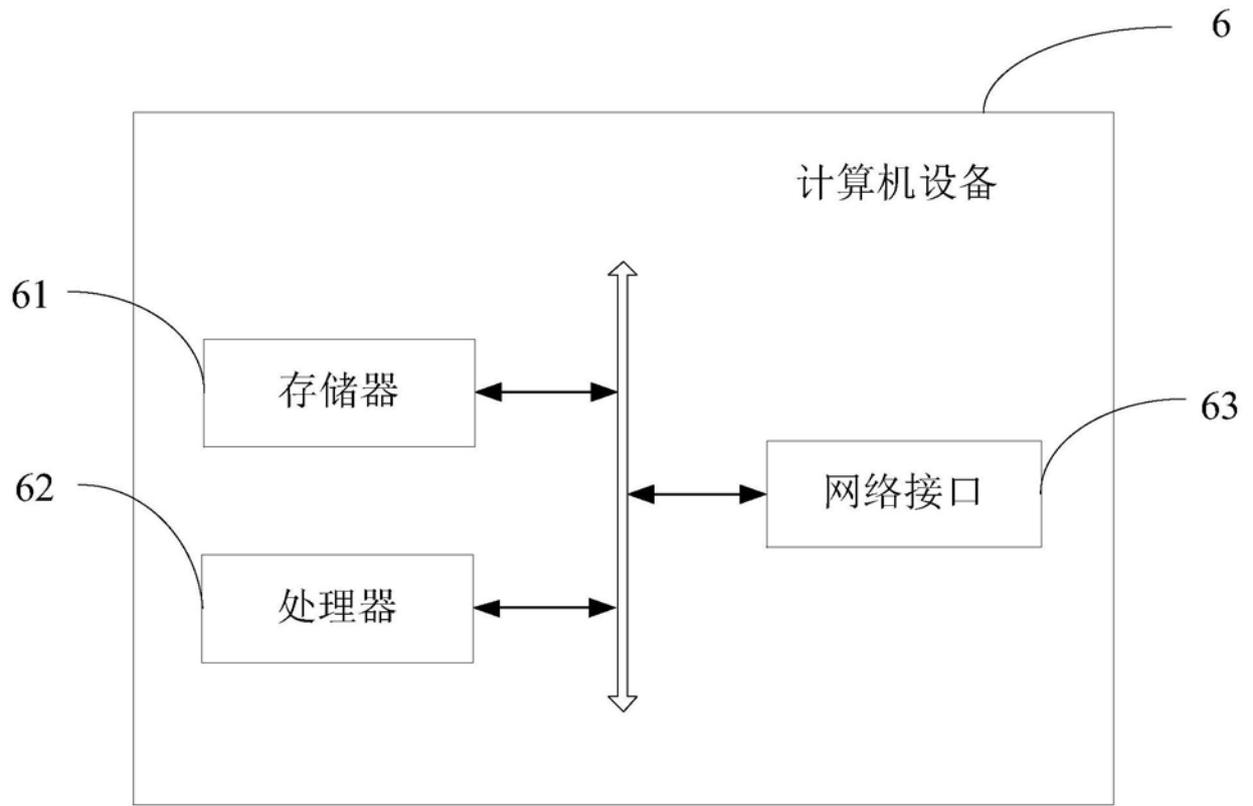


图4