

数据处理方法及装置

本申请要求 2020 年 06 月 19 日递交的申请号为 202010567428.X、发明名称为“数据处理方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本说明书实施例涉及数据处理技术领域，特别涉及一种数据处理方法。本说明书一个或者多个实施例同时涉及一种数据处理装置，一种计算设备，以及一种计算机可读存储介质。

背景技术

- 10 在互联网数据库解决方案中，一般会采用读写分离的技术实现数据的读操作和写操作，而传统的读写分离提供一个读写分离连接串，针对非事务读请求按照用户配置的读权重比例将读请求下发到只读数据库，但是事务内的读请求会直接下发至主实例（即主库，提供完整的服务，包括读写链路）上，如果事务中包含了（OLAP, On-Line Analytical Processing）在线分析处理类型的读请求，则会导致 OLAP 的读请求占用主实例过多的资源，
- 15 源，从而影响主实例的性能及稳定性。

因此，急需提供一种可以减少和避免 OLAP 类型的读请求针对主实例的资源抢占，从而保证数据库的稳定性的数据处理方法。

发明内容

- 有鉴于此，本说明书实施例提供了一种数据处理方法。本说明书一个或者多个实施例同时涉及一种数据处理装置，一种计算设备，以及一种计算机可读存储介质，以解决现有技术中存在的技术缺陷。
- 20 根据本说明书实施例的第一方面，提供了一种数据处理方法，包括：

接收携带有查询语句的数据处理请求；

- 基于所述数据处理请求将所述查询语句转换为对应的关系代数树，且基于所述关系代数树确定所述查询语句对应的操作类型；
- 25 在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库，基于所述查询语句在所述第一数据库完成所述数据处理请求。

可选的，所述将所述查询语句转换为对应的关系代数树包括：

- 将所述查询语句解析为对应的抽象语法树，且将所述抽象语法树转换为对应的关系代数树，其中，所述关系代数树由树节点和边组成。
- 30 可选的，所述基于所述关系代数树确定所述查询语句对应的操作类型包括：

基于所述关系代数树确定所述查询语句的结构信息；

在所述查询语句的结构信息满足第一预设规则的情况下，确定所述查询语句对应的操作类型为第一类型。

可选的,所述基于所述关系代数树确定所述查询语句对应的操作类型包括:

基于所述关系代数树确定所述查询语句的逻辑信息;

在所述查询语句的逻辑信息满足第二预设规则的情况下,确定所述查询语句对应的操作类型为第一类型。

5 可选的,所述基于所述关系代数树确定所述查询语句的结构信息之后,还包括:

在所述查询语句的结构信息不满足第一预设规则的情况下,基于所述关系代数树确定所述查询语句的逻辑信息;

在所述查询语句的逻辑信息满足第二预设规则的情况下,确定所述查询语句对应的操作类型为第一类型。

10 可选的,所述基于所述关系代数树确定所述查询语句对应的操作类型包括:

基于所述关系代数树确定所述查询语句的结构信息以及逻辑信息;

在所述查询语句的结构信息不满足第一预设规则,以及所述查询语句的逻辑信息不满足第二预设规则的情况下,基于所述关系代数树确定所述查询语句的执行计划,且根据所述执行计划计算所述查询语句的物理执行代价;

15 在所述物理执行代价小于等于预设代价阈值的情况下,确定所述查询语句对应的操作类型为第一类型。

可选的,所述基于所述关系代数树确定所述查询语句对应的操作类型包括:

基于所述关系代数树确定所述查询语句的执行计划,且根据所述执行计划计算所述查询语句的物理执行代价;

20 在所述物理执行代价小于等于预设代价阈值的情况下,确定所述查询语句对应的操作类型为第一类型。

可选的,所述基于所述关系代数树确定所述查询语句的执行计划,且根据所述执行计划计算所述查询语句的物理执行代价包括:

确定所述关系代数树的树节点,基于所述树节点确定所述查询语句的执行计划;

25 基于所述查询语句的执行计划计算所述查询语句对应的每个树节点的物理执行代价;

将每个树节点的物理执行代价进行加权相加,以获得所述查询语句的物理执行代价。

可选的,所述在所述物理执行代价小于等于预设代价阈值的情况下,确定所述查询语句对应的操作类型为第一类型包括:

30 在所述查询语句对应的任意一个树节点的物理执行代价小于等于预设代价阈值的情况下,确定所述查询语句对应的操作类型为第一类型。

可选的,所述方法,还包括:

在所述物理执行代价大于预设代价阈值的情况下,确定所述查询语句对应的操作类型为第二类型。

可选的，所述确定所述查询语句对应的操作类型为第二类型之后，还包括：

在所述查询语句对应的操作类型为第二类型的情况下，将所述查询语句下发至第二数据库，基于所述查询语句在所述第二数据库完成所述数据处理请求。

可选的，所述第一数据库为处理第一类型的查询语句的分布式读写数据库，所述第
5 二数据库为处理第二类型的查询语句的分布式只读数据库。

可选的，所述第一类型为 OLTP 类型（On-Line Transaction Processing，在线事务处理类型）；

相应的，所述在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库包括：

10 在所述操作类型为 OLTP 类型的情况下，将所述查询语句下发至所述分布式读写数据库。

可选的，所述第二类型为 OLAP 类型（On-Line Analytical Processing，在线分析处理类型）；

相应的，所述在所述查询语句对应的操作类型为第二类型的情况下，将所述查询语
15 句下发至第二数据库包括：

在所述查询语句对应的操作类型为 OLAP 类型的情况下，将所述查询语句下发至所述分布式只读数据库。

根据本说明书实施例的第二方面，提供了一种数据处理装置，包括：

处理请求接收模块，被配置为接收携带有查询语句的数据处理请求；

20 操作类型确定模块，被配置为基于所述数据处理请求将所述查询语句转换为对应的关系代数树，且基于所述关系代数树确定所述查询语句对应的操作类型；

数据库确定模块，被配置为在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库，基于所述查询语句在所述第一数据库完成所述数据处理请求。

根据本说明书实施例的第三方面，提供了一种数据处理系统，包括：

25 路由模块、读写分布式数据节点以及只读分布式数据节点，其中，

所述路由模块，被配置为接收携带查询语句的数据处理请求，且基于所述数据处理请求确定所述查询语句对应的所述操作类型，

在所述操作类型为 OLTP 类型的情况下，将所述查询语句下发至所述读写分布式数据节点，且基于所述查询语句在所述读写分布式数据节点完成所述数据处理请求，

30 在所述操作类型为 OLAP 类型的情况下，将所述查询语句下发至所述只读分布式数据节点，且基于所述查询语句在所述只读分布式数据节点完成所述数据处理请求。

根据本说明书实施例的第四方面，提供了一种计算设备，包括：

存储器和处理器；

所述存储器用于存储计算机可执行指令，其中，所述处理器执行所述计算机可执行

指令时实现所述数据处理方法的步骤。

根据本说明书实施例的第五方面，提供了一种计算机可读存储介质，其存储有计算机可执行指令，该指令被处理器执行时实现所述数据处理方法的步骤。

本说明书一个实施例实现了一种数据处理方法及装置，其中，所述方法包括接收携
5 带有查询语句的数据处理请求；基于所述数据处理请求将所述查询语句转换为对应的关系代数树，且基于所述关系代数树确定所述查询语句对应的操作类型；在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库，基于所述查询语句在所述第一数据库完成所述数据处理请求；所述数据处理方法能够智能识别出客户端发送的 SQL
10 语句对应的操作类型为第一类型，即(OLTP)在线事务处理类型还是第二类型，即(OLAP)在线分析处理类型，且在确定 SQL 语句对应的操作类型为(OLTP)在线事务处理类型的情况下，才会将该 SQL 语句下发至主实例上进行处理，保证了 OLTP 负载流量的低延迟和高吞吐，保证了数据库的稳定性。

附图说明

图 1 是本说明书一个实施例提供的一种数据处理装置的架构示意图；
15 图 2 是本说明书一个实施例提供的一种数据处理方法的流程图；
图 3 是本说明书一个实施例提供的一种数据处理方法在 HTAP 分布式数据库的应用流程图；
图 4 是本说明书一个实施例提供的一种数据处理方法中智能路由对 SQL 语句对应的操作类型的具体分析流程图；
20 图 5 是本说明书一个实施例提供的一种数据处理装置的结构示意图；
图 6 是本说明书一个实施例提供的一种数据处理系统的结构示意图；
图 7 是本说明书一个实施例提供的一种计算设备的结构框图。

具体实施方式

在下面的描述中阐述了很多具体细节以便于充分理解本说明书。但是本说明书能够
25 以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本说明书内涵的情况下做类似推广，因此本说明书不受下面公开的具体实施的限制。

在本说明书一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本说明书一个或多个实施例中使用的术语“和/或”
30 是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

应当理解，尽管在本说明书一个或多个实施例中可能采用术语第一、第二等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书一个或多个实施例范围的情况下，第一也可以被称为第二，

类似地，第二也可以被称为第一。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

首先，对本说明书一个或多个实施例涉及的名词术语进行解释。

5 负载：数据库中的负载主要分为(OLTP)在线事务处理类型负载和(OLAP)在线分析处理类型负载。

OLTP: On-Line Transaction Processing, 在线事务处理类型, 主要包含如涉及少范围数据的增删改查等 SQL。

OLAP: On-Line Analytical Processing, 在线分析处理类型, 主要包含如涉及大范围数据的复杂查询分析型 SQL。

10 HTAP 混合负载: (OLTP)在线事务处理类型负载和(OLAP)在线分析处理类型负载同时共存的负载。

读写实例: 可以进行读和写操作的数据库实例。

只读实例: 只能进行读操作的数据库实例。

15 SQL: 结构化查询语言(Structured Query Language), 简称 SQL, 是一种特殊目的的编程语言, 是一种数据库查询的程序设计语言, 用于存取数据以及查询、更新和管理关系数据库系统。

执行计划: 一条 SQL 会对应多个等价执行计划, 执行计划包含了如何实际执行一条 SQL。

Cost: 执行计划的代价, 包含 CPU, Memory, IO 等信息。

20 RT: Response Time, 事务从请求到系统处理完成后返回结果的时间, 即响应时间。

在本说明书中, 提供了一种数据处理方法。本说明书一个或者多个实施例同时涉及一种数据处理装置, 一种计算设备, 以及一种计算机可读存储介质, 在下面的实施例中逐一进行详细说明。

参见图 1, 图 1 示出了本说明书一个实施例提供的一种数据处理装置的架构示意图, 25 包括: 客户端 102、智能路由 104 (即读写分离链接串), 第一数据库 106, 第二数据库 108, 其中, 第一数据库 106 可以理解为分布式读写数据库, 第二数据库 108 可以理解为分布式只读数据库, 实际应用中, 第一数据库 106 也可以看作分布式读写数据库的读写分布式数据节点, 即 RW 节点, 第二数据库 108 也可以看作分布式只读数据库的只读分布式数据节点, 即 RO 节点。

30 具体实施时, 客户端 102 发送针对第一数据库 106 或第二数据库 108 的查询语句以进行数据处理, 智能路由 104 接收客户端 102 发送的查询语句后, 对该查询语句进行解析, 确定该查询语句对应的执行计划, 然后基于该执行计划计算出该查询语句的物理执行代价, 例如执行该查询语句时 CPU 要执行的指令的数量, 执行这些数量的指令需要的内存 (Memory), IO 访问次数以及需要多少个网络 (Net) 交互等, 最后根据该查询语

句的物理执行代价确定该查询语句对应的操作类型，在该操作类型为 OLTP 类型的情况下，将查询语句下发至第一数据库 106，而在该操作类型为 OLAP 类型的情况下，将查询语句下发至第二数据库 108，以在第一数据库 106 或第二数据库 108 完成客户端的数据处理请求，其中，所述查询语句为 SQL 语句。

- 5 下面对所述数据处理方法的具体处理过程进行详细说明，具体实施时，所述数据处理方法应用于 HTAP 分布式数据库中，且下述数据处理方法均以智能路由 104 的为执行主体进行描述。

参见图 2，图 2 示出了根据本说明书一个实施例提供的一种数据处理方法的流程图，包括以下步骤：

- 10 步骤 202：接收携带有查询语句的数据处理请求。

其中，数据处理请求包括但不限于数据查询请求、数据更新请求、数据读取请求、以及数据写入请求等；所述查询语句为 SQL 语句。

- 15 而该数据处理请求中携带的查询语句，即为数据处理请求的具体数据要求，例如数据处理请求为数据插入请求，而该数据处理请求中携带的查询语句则可以具体体现出将哪些数据从哪个数据库的表读取出来插入到哪个数据库的哪个表中。

举例说明，客户端发送的数据处理请求为数据插入请求，对应该数据插入请求中携带的查询语句则可以为：insert into T select name from S，即从 S 表中选择 name 字段插入到 T 表中。

- 20 步骤 204：基于所述数据处理请求将所述查询语句转换为对应的关系代数树，且基于所述关系代数树确定所述查询语句对应的操作类型。

其中，基于所述数据处理请求将所述查询语句转换为对应的关系代数树可以理解为，基于所述数据处理请求对所述查询语句进行解析，而具体对所述查询语句的解析过程则是将所述查询语句转换为对应的关系代数树的过程。

具体的，所述将所述查询语句转换为对应的关系代数树包括：

- 25 将所述查询语句解析为对应的抽象语法树，且将所述抽象语法树转换为对应的关系代数树，其中，所述关系代数树由树节点和边组成。

其中，抽象语法树的英文全称为：Abstract Syntax Tree，英文简称为：AST，而将查询语句解析为对应的抽象语法树，则是将查询语句进行解析，以树状的形式表现出其语法结构。

- 30 仍以上述数据插入请求中携带的查询语句：insert into T select name from S 为例，将所述查询语句解析为对应的抽象语法树，则是将查询语句以树状的形式表现出来，例如将“insert”作为树的第一个节点，该节点下面再分为两个输入分支节点，其中，一个节点为“T”，另一个节点为“select name from S”，而“select name from S”节点下又可以有两个分支节点：“S”节点和“name”节点。

而在将查询语句解析为对应的抽象语法树后，再将 ATS 抽象语法树转换为对应的关系代数树，而关系代数树由树节点和边组成，仍以上述数据插入请求中携带的查询语句为例，该关系代数树的树节点至少包括表扫描节点、映射节点以及写入节点，通过该关系代数树的树节点以及连接各树节点的边连接关系，可以确定出该查询语句：需要扫描

5 S 表，然后将 S 表中的 name 字段映射出来，然后将映射出的 name 字段写入到 T 表。

具体的，在将查询语句具体转换为对应的关系代数树之后，即可基于该关系代数树确定查询语句对应的操作类型。

本说明书实施例中，所述数据处理方法通过将查询语句转换为关系代数树结构的优化策略，使得基于关系代数树结构的查询语句针对数据库进行数据处理时，极大的提高了

10 了数据处理效率。

具体实施时，为了快速的确定查询语句的操作类型，可以通过查询语句对应的关系代数树确定的查询语句的结构信息来实现，具体实现方式如下所述：

所述基于所述关系代数树确定所述查询语句对应的操作类型包括：

基于所述关系代数树确定所述查询语句的结构信息；

15 在所述查询语句的结构信息满足第一预设规则的情况下，确定所述查询语句对应的操作类型为第一类型。

其中，第一预设规则为预设的满足第一类型的查询语句对应的结构信息的要求。

具体的，查询语句的关系代数树可以体现出对应的 SQL 的具体结构信息，例如 SQL 中有没有主键、主键有没有索引、索引有没有等值条件等，若第一预设规则为查询语句

20 中要存在主键、主键有对应索引以及索引有等值条件，那么通过查询语句对应的关系代数树中确定到该查询语句的结构信息中包括：主键、主键有对应索引，且索引也存在等值条件，那么则可以确定该查询语句的结构信息满足第一预设规则，此时，则可以确定该查询语句对应的操作类型为第一类型；而第一类型则为上述的 OLTP 类型。

本说明书实施例中，所述数据处理方法通过基于关系代数树结构的查询语句的结构

25 信息与预设规则的匹配，可以快速的确定出查询语句对应的操作类型，后续可以基于准确的操作类型将该查询语句下发到合适的数据库中进行数据处理，极大的提高了数据处理的工作效率。

而具体应用时，每个查询语句对应的关系代数树均有树节点进行表达，即每个查询语句均可以由对应的树节点表达，因此为了更加快速的确定查询语句对应的操作类型，

30 可以通过查询语句对应的树节点进行确定，具体实现方式如下所述：

所述基于所述关系代数树确定所述查询语句的结构信息包括：

基于所述关系代数树确定所述查询语句的结构信息；

相应的，所述在所述查询语句的结构信息满足第一预设规则的情况下，确定所述查询语句对应的操作类型为第一类型包括：

在所述查询语句的结构信息中的树节点满足第一预设规则的情况下，确定所述查询语句对应的操作类型为第一类型。

其中，查询语句对应的关系代数树的树节点可以表达每个查询语句的具体含义，在确定关系代数树的树节点之后，根据该树节点即可以看出对应的查询语句的结构信息，

5 即可以理解为该查询语句的结构信息中包括多个树节点。

仍以上述数据插入请求中携带的查询语句为例，将该查询语句解析后可以确定该查询语句对应的关系代数树包括至少 3 个树节点：表扫描节点、映射节点以及写入节点等，此时则可以确定该查询语句的结构信息中至少包括：表扫描节点、映射节点以及写入节点等；若第一预设规则为查询语句的结构信息中的树节点的数量小于等于预设数量阈值，
10 如 3 个的情况下，确定该查询语句对应的操作类型为第一类型；则根据上述可知该查询语句的结构信息中包括 3 个树节点，因此满足第一预设规则，此时则可以确定该查询语句对应的操作类型为第一类型。

本说明书实施例提供的所述数据处理方法，通过基于关系代数树结构的查询语句的结构信息的树节点与预设规则的匹配方法，可以更快速的确定出查询语句对应的操作类型，
15 后续可以基于准确的操作类型将该查询语句下发到合适的数据库中进行数据处理，极大的提高了数据处理的工作效率。

此外，基于查询语句对应的关系代数树还可以确定查询语句对应的逻辑信息，在另外一种可实现的情况下，还可以基于对查询语句对应的逻辑信息的匹配，对查询语句对应的操作类型进行确定，具体实现方式如下所述：

20 所述基于所述关系代数树确定所述查询语句对应的操作类型包括：

基于所述关系代数树确定所述查询语句的逻辑信息；

在所述查询语句的逻辑信息满足第二预设规则的情况下，确定所述查询语句对应的操作类型为第一类型。

25 其中，第二预设规则为预设的满足第一类型的查询语句对应的逻辑信息的要求，例如通过第二预设规则对表的行数、过滤后的行数目的要求。

具体的，基于关系代数树可以确定出查询语句的逻辑信息，仍以上述数据插入请求中携带的查询语句为例，基于该查询语句的关系代数树可以确定出该查询语句的逻辑信息为：S 表中的总行数为 1000 行，若该查询语句后面有个 where 条件：where age=20，按照该 where 条件过滤后的 S 表中的 name 字段的行数为 300 行。

30 此时，若第二预设规则为：查询语句中表的行数小于等于第一预设行数阈值，例如 1000，过滤后的行数据小于等于第二预设行数阈值，例如 500，那么可以确定上述查询语句的逻辑信息与第二预设规则匹配，此时则可以确定该查询语句对应的操作类型为第一类型。

本说明书实施例提供的所述数据处理方法，通过将查询语句的逻辑信息与预设逻辑

信息进行匹配的方式，快速便捷的确定出查询语句对应的操作类型，提升用户体验。

在另一种可实现的实施例中，为了保证对第一类型的查询语句判断的准确性，即使在查询语句的结构信息不满足第一预设规则的情况下，也不会直接将该查询语句定义为第二类型，而会通过查询语句的逻辑信息再进行一次判断，具体实现方式如下：

5 所述基于所述关系代数树确定所述查询语句的结构信息之后，还包括：

 在所述查询语句的结构信息不满足第一预设规则的情况下，基于所述关系代数树确定所述 查询语句的逻辑信息；

 在所述查询语句的逻辑信息满足第二预设规则的情况下，确定所述查询语句对应的操作类型为第一类型。

10 具体的，在基于关系代数树确定的查询语句的结构信息不满足第一预设规则，但是确定该查询语句的逻辑信息满足第二预设规则的情况下，依然会将该查询语句对应的操作类型定义为第一类型，以便后续可以在该第一类型对应的第一数据库进行该查询语句的数据处理，提高对第一类型的查询语句的低延时和高吞吐的处理。

 实际应用中，即使在查询语句的结构信息不满足第一预设规则，且该查询语句的逻辑信息也不满足第二预设规则的情况下，为了对第一类型的查询语句的无遗漏筛选，还会对该查询语句进行再一次的判断，具体实现方式如下：

 所述基于所述关系代数树确定所述查询语句对应的操作类型包括：

 基于所述关系代数树确定所述查询语句的结构信息以及逻辑信息；

20 在所述查询语句的结构信息不满足第一预设规则，以及所述查询语句的逻辑信息不满足第二预设规则的情况下，基于所述关系代数树确定所述查询语句的执行计划，且根据所述执行计划计算所述查询语句的物理执行代价；

 在所述物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

25 其中，执行计划可以理解为实际执行一条 SQL 语句的计划，例如先执行该 SQL 语句中的表的字段的读取，再执行该 SQL 语句中的读取字段的写入等，而物理执行代价则为 Cost，即执行一条 SQL 语句计划 CPU 要执行的指令数量，执行这些数量的指令需要的 Memory，IO 访问次数以及需要多少个网络（Net）交互等。

 而预设代价阈值可以根据实际应用进行设置，在此不做任何限定，例如设置为 2 万、3 万等等。

30 具体的，首先基于关系代数树确定对应的查询语句的执行计划，然后根据执行计划计算出执行该 SQL 语句时所要付出的物理执行代价，然后在物理执行代价小于等于预设代价阈值，例如 2 万的情况下，确定该查询语句对应的操作类型为第一类型。

 在本说明书另一个可实施的方案中，也可以在接收到查询语句之后，直接基于上述的代价规则实现对查询语句对应的操作类型的判断，以提高数据处理效率，具体实现方

式如下所述：

所述基于所述关系代数树确定所述查询语句对应的操作类型包括：

基于所述关系代数树确定所述查询语句的执行计划，且根据所述执行计划计算所述查询语句的物理执行代价；

- 5 在所述物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

而具体的，所述基于所述关系代数树确定所述查询语句的执行计划，且根据所述执行计划计算所述查询语句的物理执行代价包括：

确定所述关系代数树的树节点，基于所述树节点确定所述查询语句的执行计划；

- 10 基于所述查询语句的执行计划计算所述查询语句对应的每个树节点的物理执行代价；

将每个树节点的物理执行代价进行加权相加，以获得所述查询语句的物理执行代价。

具体实施时，可以根据查询语句的执行计划计算出查询语句中每个树节点的物理执行代价，然后将该查询语句中所有树节点的物理执行代价加权相加，以获得最终的查询语句的物理执行代价。

- 15 实际应用中，每个物理执行代价的权重均不相同，因此在进行查询语句的物理执行代价计算时，每个物理执行代价会乘以对应的权重后相加，以确保最终的查询语句的物理执行代价的精确性。

而具体应用时，物理执行代价和预设代价阈值的比较一般会存在两种情况，第一种：

- 20 所述在所述物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型包括：

在所述查询语句对应的任意一个树节点的物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

- 25 即在计算出查询语句的每个树节点的物理执行代价后，即可基于每个树节点的物理执行代价确定对应的查询语句对应的操作类型，例如每个树节点的物理执行代价均有一个对应的预设代价阈值，如查询语句包括树节点 1 和树节点 2，树节点 1 对应的预设代价阈值为 2000，树节点 2 的预设代价阈值为 3000，那么无论是该查询语句中的树节点 1 或者是树节点 2 小于等于其对应的预设代价阈值，均可以认为该查询语句对应的操作类型为第一类型。

- 30 第二种：所述在所述物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型包括：

在所述查询语句对应的所有树节点的物理执行代价的总和小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

即是在将所有树节点的物理执行代价进行加权相加后，获得查询语句的总的物理执

行代价，在总的物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

本说明书实施例提供的所述数据处理方法，在根据查询语句的执行计划确定出每个树节点的物理执行代价后，在该查询语句对应的任意一个树节点的物理执行代价小于等于预设代价阈值的情况下，就可以确定该查询语句对应的操作类型为第一类型，可以极大的提高第一类型的查询语句的筛选率，保证 OLTP 类型的查询语句后续可以在第一数据库执行，提高其数据处理执行率。

步骤 206：在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库，基于所述查询语句在所述第一数据库完成所述数据处理请求。

具体的，在通过上述多种方式确定了查询语句对应的操作类型为第一类型的情况下，就可以将该查询语句下发至第一数据库，在第一数据库根据该查询语句执行数据处理请求。

具体实施时，所述第一类型为 OLTP 类型，此时，所述第一数据库则为分布式读写数据库；

相应的，所述在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库包括：

在所述操作类型为 OLTP 类型的情况下，将所述查询语句下发至所述分布式读写数据库。

本说明书另一实施例中，在通过上述所有的方式对查询语句进行了判断之后，若该查询语句的物理执行代价大于预设代价阈值，才可以确定该查询语句对应的操作类型为第二类型，具体实现方式如下所述：

所述方法，还包括：

在所述物理执行代价大于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第二类型。

其中，第二类型为 OLAP 类型，即在该查询语句的结构信息不满足第一预设规则，逻辑信息不满足第二预设规则，且物理执行代价大于预设代价阈值的情况下，才可以确定该查询语句对应的操作类型为 OLAP 类型，以实现数据处理的读写分离，保证数据库的稳定性。

而在查询语句对应的操作类型为第二类型的情况下，则是要将查询语句下发至第二数据库，基于该查询语句在第二数据库完成数据处理请求，具体实现方式如下：

所述确定所述查询语句对应的操作类型为第二类型之后，还包括：

在所述查询语句对应的操作类型为第二类型的情况下，将所述查询语句下发至第二数据库，基于所述查询语句在所述第二数据库完成所述数据处理请求。

具体实施时，所述第二类型为 OLAP 类型，此时，第二数据库则为分布式只读数据

库；

相应的，所述在所述查询语句对应的操作类型为第二类型的情况下，将所述查询语句下发至第二数据库包括：

在所述查询语句对应的操作类型为 OLAP 类型的情况下，将所述查询语句下发至所述分布式只读数据库。

具体的，OLAP 类型的查询语句一般为复杂分析的查询语句，若将 OLAP 类型和 OLTP 类型的查询语句均下发至同一个数据库进行数据处理，则会造成资源抢占，而为了保证数据库的稳定性，采用本说明书实施例的所述数据处理方法实现的良好读写分离策略能够减少和避免 OLAP 类型的负载对 OLTP 类型负载的资源抢占，从而使数据库更为稳定性。

而具体的，所述第一数据库为处理第一类型的查询语句的分布式读写数据库，所述第二数据库为处理第二类型的查询语句的分布式只读数据库，即可以理解为主库，第二数据库为备库，其中，主库中的数据会同步到备库中，但是备库中的数据会有一定的延时，由于 OLTP 类型的查询语句的数据处理一般会对数据库中的数据进行变化，而 OLAP 类型的查询语句的数据处理一般是对数据库中数据的查询、探索等，为了保证数据的完整性以及数据处理的效率，即将 OLTP 类型的查询语句的数据处理分发到第一数据库执行，以保证数据的读写完整性，将 OLAP 类型的查询语句的数据处理分发到第二数据库执行，以避免 OLAP 类型的负载对 OLTP 类型负载的资源抢占，从而使数据库更为稳定性，以及提高数据处理的工作效率。

参见图 3，以本说明书提供的数据处理方法在 HTAP 分布式数据库的应用为例，对所述数据处理方法中的 SQL 语句对应的操作类型为事务处理类型的情况进行进一步说明，具体包括以下步骤：

步骤 302：客户端（Client）发送携带有 SQL 语句的数据处理请求至 HTAP 分布式数据库。

步骤 304：分布式数据库中的智能路由对 SQL 语句进行解析，确定 SQL 语句对应的操作类型。

步骤 306：分布式数据库中的智能路由确定 SQL 语句对应的操作类型为 OLTP 类型的情况下，将 SQL 语句下发至对应的 RDS1 数据节点。

其中，RDS 数据节点即为读写数据库的分布式数据节点，例如图 3 中的 RDS1 和 RDS2 均为读写数据库的分布式数据节点，区别仅在于 RDS1 与 RDS2 中存放的数据是不同的，例如 RDS1 对应存储的是字段 name、age 的数据，RDS2 中对应存储的是字段 duty、address 的数据，那么对 SQL 语句进行解析后，将需要针对 RDS1 中的字段进行数据处理的 SQL 语句下发至 RDS1，将需要针对 RDS2 中的字段进行数据处理的 SQL 语句下发至 RDS2。

步骤 308：RDS1 接收 SQL 语句，并在其数据节点上执行该 SQL 语句。

步骤 310: RDS1 将对 SQL 语句的执行结果返回至分布式数据库。

步骤 312: 分布式数据库将 SQL 语句下发至对应的 RD2 数据节点。

步骤 314: RDS2 接收 SQL 语句, 并在其数据节点上执行该 SQL 语句。

步骤 316: RDS2 将对 SQL 语句的执行结果返回至分布式数据库。

5 步骤 318: 分布式数据库对 RDS1 和 RDS2 返回的结果进行整合计算。

步骤 320: 分布式数据库将整合计算的结果返回至 Client。

具体的, 步骤 306、步骤 308、步骤 310 与步骤 312、步骤 314、步骤 316 的执行无先后顺序要求, 实际使用中可以并行执行, 根据实际应用设置执行顺序即可。

此外, 一个 SQL 语句可能会同时下发至分布式数据库的多个 RDS 数据节点进行执行, 具体需要多少数量的 RDS 数据节点执行该 SQL 语句, 则是要根据解析后的 SQL 语句的具体待执行内容进行确定, 在此不做任何限定, 例如该 SQL 语句中待执行的数据分别存放在 5 个 RDS 数据节点, 则执行该 SQL 语句需要同时将该 SQL 语句下发至对应的 5 个 RDS 数据节点进行数据处理。

15 本说明书实施例提供的所述数据处理方法, 在智能路由模块接收到用户的 SQL 语句的数据处理请求后, 对 SQL 语句进行负载类型分析, 根据分析结果确定 SQL 语句的负载类型为 OLTP 类型负载的情况下, 在读写数据库的分布式数据节点执行该 SQL 语句, 以提高数据处理的低延迟和高吞吐。

参见图 4, 图 4 示出了本说明书一个实施例提供的一种数据处理方法中智能路由对 SQL 语句对应的操作类型的具体分析流程图, 具体包括以下步骤:

20 步骤 402: SQL 解析为 AST 抽象语法树。

具体的, 在接收客户端发送的 SQL 语句之后, 将 SQL 语句解析为 AST 抽象语法树。

步骤 404: AST 抽象语法树转化为关系代数树。

步骤 406: 前置规则根据结构信息及逻辑信息判断路由。

其中, 前置规则可以理解为上述实施例的第一预设规则以及第二预设规则。

25 具体的, 前置规则根据结构信息及逻辑信息判断路由, 即可以理解为, 判断 SQL 语句的结构信息及逻辑信息是否与前置规则匹配, 若是, 则执行步骤 412, 若否, 则执行步骤 408。

步骤 408: 通过基于代价的优化器优化出物理执行计划。

30 具体的, 基于代价的优化器优化出物理执行计划, 即基于代价的优化器为待执行的 SQL 语句制定的物理执行计划, 该物理执行计划包含了如何实际执行该 SQL 语句。

步骤 410: 代价规则根据代价信息判断路由。

具体的, 根据上述 SQL 语句制定的物理执行计划计算出执行该 SQL 语句的物理执行代价, 其中, 物理执行代价可以参见上述实施例, 在此不再赘述; 然后根据计算得到的物理执行代价判断该 SQL 语句需要路由到哪个数据库。

步骤 412: 路由结果。

具体的, 在 SQL 语句的结构信息及逻辑信息与前置规则匹配的情况下, 可以确定该 SQL 语句需要路由到读写数据库的分布式数据节点上, 而在 SQL 语句的结构信息及逻辑信息与前置规则不匹配, 然后可以根据 SQL 语句的物理执行代价确定该 SQL 语句需要

5 路由到读写数据库还是只读数据库中, 实现 SQL 语句的读写分离, 使得最终 OLTP 类型的请求会被路由至 RW (读写) 节点, 而 OLAP 类型的请求会被路由至 RO (只读) 节点, 从而保证了高并发的 OLTP 类型的 SQL 有稳定的 RT 和吞吐, 且不会被消耗大量资源的 OLAP 类型的 SQL 请求影响, 保证数据库的稳定性。

本说明书实施例提供的所述数据处理方法设计了一种能够智能识别出 OLTP 和

10 OLAP 类型的负载的读写分离路由算法, 用于读写分离, 能够做到 OLAP 负载流量不影响 OLTP 负载流量, 保证 OLTP 负载流量低延迟和高吞吐, 无需人工干预, 就可以很好地解决 OLAP 负载与 OLTP 负载相互影响的问题。

与上述方法实施例相对应, 本说明书还提供了数据处理装置实施例, 图 5 示出了本说明书一个实施例提供的一种数据处理装置的结构示意图。如图 5 所示, 该装置包括:

15 处理请求接收模块 502, 被配置为接收携带有查询语句的数据处理请求;

操作类型确定模块 504, 被配置为基于所述数据处理请求将所述查询语句转换为对应的关系代数树, 且基于所述关系代数树确定所述查询语句对应的操作类型;

数据库确定模块 506, 被配置为在所述操作类型为第一类型的情况下, 将所述查询语句下发至第一数据库, 基于所述查询语句在所述第一数据库完成所述数据处理请求。

20 可选的, 所述操作类型确定模块 504, 进一步被配置为:

将所述查询语句解析为对应的抽象语法树, 且将所述抽象语法树转换为对应的关系代数树, 其中, 所述关系代数树由树节点和边组成。

可选的, 所述操作类型确定模块 504, 进一步被配置为:

基于所述关系代数树确定所述查询语句的结构信息;

25 在所述查询语句的结构信息满足第一预设规则的情况下, 确定所述查询语句对应的操作类型为第一类型。

可选的, 所述操作类型确定模块 504, 进一步被配置为:

基于所述关系代数树确定所述查询语句的逻辑信息;

30 在所述查询语句的逻辑信息满足第二预设规则的情况下, 确定所述查询语句对应的操作类型为第一类型。

可选的, 所述装置, 还包括:

逻辑信息确定模块, 被配置为在所述查询语句的结构信息不满足第一预设规则的情况下, 基于所述关系代数树确定所述查询语句的逻辑信息;

事务类型确定模块, 被配置为在所述查询语句的逻辑信息满足第二预设规则的情况

下，确定所述查询语句对应的操作类型为第一类型。

可选的，所述操作类型确定模块 504，进一步被配置为：

基于所述关系代数树确定所述查询语句的结构信息以及逻辑信息；

- 5 在所述查询语句的结构信息不满足第一预设规则，以及所述查询语句的逻辑信息不满足第二预设规则的情况下，基于所述关系代数树确定所述查询语句的执行计划，且根据所述执行计划计算所述查询语句的物理执行代价；

在所述物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

可选的，所述操作类型确定模块 504，进一步被配置为：

- 10 基于所述关系代数树确定所述查询语句的执行计划，且根据所述执行计划计算所述查询语句的物理执行代价；

在所述物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

可选的，所述操作类型确定模块 504，进一步被配置为：

- 15 确定所述关系代数树的树节点，基于所述树节点确定所述查询语句的执行计划；
基于所述查询语句的执行计划计算所述查询语句对应的每个树节点的物理执行代价；

将每个树节点的物理执行代价进行加权相加，以获得所述查询语句的物理执行代价。

可选的，所述操作类型确定模块 504，进一步被配置为：

- 20 在所述查询语句对应的任意一个树节点的物理执行代价小于等于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第一类型。

可选的，所述装置，还包括：

第二类型确定模块，被配置为在所述物理执行代价大于预设代价阈值的情况下，确定所述查询语句对应的操作类型为第二类型。

- 25 可选的，所述装置，还包括：

数据处理模块，被配置为在所述查询语句对应的操作类型为第二类型的情况下，将所述查询语句下发至第二数据库，基于所述查询语句在所述第二数据库完成所述数据处理请求。

- 30 可选的，所述第一数据库为处理第一类型的查询语句的分布式读写数据库，所述第二数据库为处理第二类型的查询语句的分布式只读数据库。

可选的，所述第一类型为 OLTP 类型；

相应的，所述数据库确定模块，进一步被配置为：

在所述操作类型为 OLTP 类型的情况下，将所述查询语句下发至所述分布式读写数据库。

可选的，所述第二类型为 OLAP 类型；

相应的，所述数据处理模块，进一步被配置为：

在所述查询语句对应的操作类型为 OLAP 类型的情况下，将所述查询语句下发至所述分布式只读数据库。

5 本说明书一个实施例实现了一种数据处理装置，能够智能识别出客户端发送的 SQL 语句对应的操作类型为（OLTP）在线事务处理类型还是（OLAP）在线分析处理类型，且在确定 SQL 语句对应的操作类型为（OLTP）在线事务处理类型的情况下，才会将该 SQL 语句下发至主实例（即读写数据库）上进行处理，保证了 OLTP 负载流量的低延迟和高吞吐，保证了数据库的稳定性。

10 上述为本实施例的一种数据处理装置的示意性方案。需要说明的是，该数据处理装置的技术方案与上述的数据处理方法的技术方案属于同一构思，数据处理装置的技术方案未详细描述的细节内容，均可以参见上述数据处理方法的技术方案的描述。

参见图 6，图 6 示出了根据本说明书一个实施例提供的一种数据处理系统的结构示意图。

15 所述数据处理系统包括：路由模块 602、读写分布式数据节点 604 以及只读分布式数据节点 606，其中，

所述路由模块 602，被配置为接收携带查询语句的数据处理请求，且基于所述数据处理请求确定所述查询语句对应的所述操作类型，

20 在所述操作类型为 OLTP 类型的情况下，将所述查询语句下发至所述读写分布式数据节点 604，且基于所述查询语句在所述读写分布式数据节点 604 完成所述数据处理请求，

在所述操作类型为 OLAP 类型的情况下，将所述查询语句下发至所述只读分布式数据节点 606，且基于所述查询语句在所述只读分布式数据节点 606 完成所述数据处理请求。

25 本说明书实施例提供的所述数据处理系统中，路由模块 602 在接收客户端发送的携带有查询语句的数据处理请求后，基于所述数据处理请求将所述查询语句转换为对应的关系代数树，且基于所述关系代数树确定所述查询语句对应的操作类型；在所述操作类型为 OLTP 类型的情况下，将所述查询语句下发至所述读写分布式数据节点 604，且基于所述查询语句在所述读写分布式数据节点 604 完成所述数据处理请求；或者在所述操作类型为 OLAP 类型的情况下，将所述查询语句下发至所述只读分布式数据节点 606，且基于所述查询语句在所述只读分布式数据节点 606 完成所述数据处理请求，所述路由模块 602 能够智能识别出客户端发送的查询语句对应的操作类型为 OLTP 类型还是 OLAP 类型，且在确定查询语句对应的操作类型为 OLTP 类型的情况下，才会将该查询语句下发至主实例上进行处理，保证了 OLTP 负载流量的低延迟和高吞吐，保证了数据

30

库的稳定性。

图 7 示出了根据本说明书一个实施例提供的一种计算设备 700 的结构框图。该计算设备 700 的部件包括但不限于存储器 710 和处理器 720。处理器 720 与存储器 710 通过总线 730 相连接，数据库 750 用于保存数据。

5 计算设备 700 还包括接入设备 740，接入设备 740 使得计算设备 700 能够经由一个或多个网络 760 通信。这些网络的示例包括公用交换电话网（PSTN）、局域网（LAN）、广域网（WAN）、个域网（PAN）或诸如因特网的通信网络的组合。接入设备 740 可以包括有线或无线的任何类型的网络接口（例如，网络接口卡（NIC））中的一个或多个，诸如 IEEE802.11 无线局域网（WLAN）无线接口、全球微波互联接入（Wi-MAX）接口、
10 以太网接口、通用串行总线（USB）接口、蜂窝网络接口、蓝牙接口、近场通信（NFC）接口，等等。

在本说明书的一个实施例中，计算设备 700 的上述部件以及图 7 中未示出的其他部件也可以彼此相连接，例如通过总线。应当理解，图 7 所示的计算设备结构框图仅仅是出于示例的目的，而不是对本说明书范围的限制。本领域技术人员可以根据需要，增添
15 或替换其他部件。

计算设备 700 可以是任何类型的静止或移动计算设备，包括移动计算机或移动计算设备（例如，平板计算机、个人数字助理、膝上型计算机、笔记本计算机、上网本等）、移动电话（例如，智能手机）、可佩戴的计算设备（例如，智能手表、智能眼镜等）或其他类型的移动设备，或者诸如台式计算机或 PC 的静止计算设备。计算设备 700 还可以是移动式或静止式的服务器。
20 以是移动式或静止式的服务器。

其中，处理器 720 用于执行如下计算机可执行指令：

接收携带有查询语句的数据处理请求；

基于所述数据处理请求将所述查询语句转换为对应的关系代数树，且基于所述关系代数树确定所述查询语句对应的操作类型；

25 在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库，基于所述查询语句在所述第一数据库完成所述数据处理请求。

上述为本实施例的一种计算设备的示意性方案。需要说明的是，该计算设备的技术方案与上述的数据处理方法的技术方案属于同一构思，计算设备的技术方案未详细描述的细节内容，均可以参见上述数据处理方法的技术方案的描述。

30 本说明书一实施例还提供一种计算机可读存储介质，其存储有计算机指令，该指令被处理器执行时以用于：

接收客户端发送的携带有查询语句的数据处理请求；

基于所述数据处理请求将所述查询语句转换为对应的关系代数树，且基于所述关系代数树确定所述查询语句对应的操作类型；

在所述操作类型为第一类型的情况下，将所述查询语句下发至第一数据库，基于所述查询语句在所述第一数据库完成所述数据处理请求。

上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是，该存储介质的技术方案与上述的数据处理方法的技术方案属于同一构思，存储介质的技术方案未详细描述的细节内容，均可以参见上述数据处理方法的技术方案的描述。

上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

所述计算机指令包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、电载波信号、电信信号以及软件分发介质等。需要说明的是，所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减，例如在某些司法管辖区，根据立法和专利实践，计算机可读介质不包括电载波信号和电信信号。

需要说明的是，对于前述的各方法实施例，为了简便描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本说明书实施例并不受所描述的动作顺序的限制，因为依据本说明书实施例，某些步骤可以采用其它顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定都是本说明书实施例所必须的。

在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中未详述的部分，可以参见其它实施例的相关描述。

以上公开的本说明书优选实施例只是用于帮助阐述本说明书。可选实施例并没有详尽叙述所有的细节，也不限制该发明仅为所述的具体实施方式。显然，根据本说明书实施例的内容，可作很多的修改和变化。本说明书选取并具体描述这些实施例，是为了更好地解释本说明书实施例的原理和实际应用，从而使所属技术领域技术人员能很好地理解和利用本说明书。本说明书仅受权利要求书及其全部范围和等效物的限制。