说明书

发明名称: 数据库性能参数的优化方法、装置以及电子设备

[0001] 本申请要求在2024年12月6日提交中国专利局、申请号为202411786864.0的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

[0002] 本申请涉及分布式数据库技术领域，例如涉及一种数据库性能参数的优化方法、装置以及电子设备。

背景技术

[0003] 分布式数据库在现代业务中被广泛采用，主要通过计算引擎层和数据存储层的分层设计实现高效运行。这种架构中，每一层又由多个节点组成，可以并发处理业务语句，从而在业务量增加时实现近线性性能提升，这种扩展性使得分布式数据库成为大规模、高并发场景下的首选方案。然而，分布式架构的复杂性也带来了显著挑战，尤其是在数据库性能优化方面，由于结构化查询语言（Structured Query Language，SQL）执行涉及多个节点及多个层次，性能表现受限于节点间的协调、资源分配及硬件瓶颈等诸多因素。同时，各层组件之间的交互使得问题排查和优化过程更加复杂，往往需要综合分析包括中央处理器（Central Processing Unit，CPU）、内存、磁盘输入/输出（Input/Output，I/O）和网络带宽在内的多维性能指标。这种多因素耦合关系导致数据库参数调优变得极为困难，不仅需要细化到具体的执行阶段，还需要结合服务器性能进行全面的调试和优化，需要很大的人力分析成本。因此，如何有效收集和分析执行过程中的性能消耗信息，并基于这些数据准确进行参数调优，是当前分布式数据库优化中亟待解决的问题。

发明内容

[0004] 本申请提供了一种数据库性能参数的优化方法、装置以及电子设备，以解决由于分布式数据库架构的复杂性而导致的参数调优难度大的问题。

[0005] 第一方面，本申请提供了一种数据库性能参数的优化方法，包括：将SQL语句集中的每一个SQL语句输入至初始数据库，以使上述初始数据库执行每一个上述SQL语句，并输出每一个上述SQL语句的运行数据，得到第一数据集，其中，上述SQL语句集中包括多类SQL语句，上述运行数据包括上述初始数据库执行上述SQL语句时，上述初始数据库的每一个节点的服务器的性能消耗数据和上述SQL语句在每一个节点的执行耗时数据；对上述第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；根据上述第一数据集，从上述初始数据库的全量参数集当中分析出需要优化的第一参数集；将所有上述综合评价指标值和上述第一参数集输入至预测模型，以使上述预测模型根据所有上述综合评价指标值对上述第一参数集进行初次优化，得到初次优化后的第二参数集；使用上述初始数据库和上述预测模型对上述第二参数集进行多次迭代优化，得到上述初始数据库的最优参数集。

[0006] 第二方面，本申请提供了一种数据库性能参数的优化装置，包括：执行模块，设置为将SQL语句集中的每一个SQL语句输入至初始数据库，以使上述初始数据库执行每一个上述SQL语句，并输出每一个上述SQL语句的运行数据，得到第一数据集，其中，上述SQL语句集中包括多类SQL语句，上述运行数据包括上述初始数据库执行上述SQL语句时，上述初始数据库的每一个节点的服务器的性能消耗数据和上述SQL语句在每一个节点的执行耗时数据；计算模块，设置为对上述第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；分析模块，设置为根据上述第一数据集，从上述初始数据库的全量参数集当中分析出需要优化的第一参数集；第一优化模块，设置为将所有上述综合评价指标值和上述第一参数集输入至预测模型，以使上述预测模型根据所有上述综合评价指标值对上述第一参数集进行初次优化，得到初次优化后的第二参数集；第二优化模块，设置为使用上述初始数据库和上述预测模型对上述第二参数集进行多次迭代优化，得到上述初始数据库的最优参数集。

[0007] 第三方面，本申请提供了一种计算机可读的存储介质，该计算机可读的存储介质中存储有计算机程序，其中，该计算机程序被处理器运行时执行上述的数据库性能参数的优化方法。

[0008] 第四方面，本申请提供了一种电子设备，包括存储器和处理器，上述存储器中存储有计算机程序，上述处理器被设置为通过上述计算机程序执行上述的数据库性能参数的优化方法。

附图说明

[0009] 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。

[0010] 下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍，对于本领域普通技术人员而言，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0011] 一个或多个实施例通过与之对应的附图中的图片进行示例性说明，附图中具有相同参考数字标号的元件表示为类似的元件，除非有特别申明，附图中的图不构成比例限制。

[0012] 图1是根据本申请实施例的一种数据库性能参数的优化方法的流程图；

[0013] 图2是根据本申请实施例的一种数据库性能参数的优化方法的架构图；

[0014] 图3是根据本申请实施例的一种数据库性能参数的优化方法的实现流程图；

[0015] 图4是根据本申请实施例的一种数据库性能参数的优化方法的日志分析模块的分析流程图；

[0016] 图5是根据本申请实施例的一种数据库性能参数的优化方法的分析处理模块的分析流程图；

[0017] 图6是根据本申请实施例的一种数据库性能参数的优化装置的结构示意图；

[0018] 图7是根据本申请实施例的一种电子设备的示意图。

具体实施方式

[0019] 下文的公开提供了许多不同的实施例或例子用来实现本申请的不同结构。为了简化本申请的公开，下文中对特定例子的部件和设置进行描述。当然，它们仅

仅为示例。此外，本申请可以在不同例子中重复参考数字和/或字母。这种重复是为了简化和清楚的目的，其本身不指示所讨论各种实施例和/或设置之间的关系。

[0020] 根据本申请实施例的第一方面，提供了一种数据库性能参数的优化方法，可选地，如图1所示，上述方法包括：

[0021] S102，将SQL语句集中的每一个SQL语句输入至初始数据库，以使初始数据库执行每一个SQL语句，并输出每一个SQL语句的运行数据，得到第一数据集，其中，SQL语句集中包括多类SQL语句，运行数据包括初始数据库执行SQL语句时，初始数据库的每一个节点的服务器的性能消耗数据和SQL语句在每一个节点的执行耗时数据；

[0022] S104，对第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；

[0023] S106，根据第一数据集，从初始数据库的全量参数集当中分析出需要优化的第一参数集；

[0024] S108，将所有综合评价指标值和第一参数集输入至预测模型，以使预测模型根据所有综合评价指标值对第一参数集进行初次优化，得到初次优化后的第二参数集；

[0025] S110，使用初始数据库和预测模型对第二参数集进行多次迭代优化，得到初始数据库的最优参数集。

[0026] 可选地，在本实施例中，如图2所示的架构图，包括应用(Application，APP)、数据库、日志分析模块以及分析处理模块，应用用于下发业务语句或SQL语句。数据库负责存储业务数据，并执行业务SQL语句，返回执行结果。分布式数据库一般分为计算引擎层和数据存储层，并且各层分为多个节点，并发处理业务语句，从而实现业务性能的极大近线性提升。日志分析模块接收数据库上传的数据库的每一个节点的服务器的性能消耗数据和SQL语句在每一个节点的详细执行耗时数据，解析SQL语句生成执行计划树，将SQL语句按执行计划树的指纹分类处理，并分析各类SQL语句耗时瓶颈及性能消耗，以此作为对照指标，最终，汇总计算各类SQL语句各执行阶段耗时及性能消耗情况，并直观展示SQL语句耗时

瓶颈处以及性能消耗情况。分析处理模块依据日志分析模块传输的各类SQL语句耗时瓶颈及性能消耗情况，形成调优初步方案和相关参数集，不断调整输入参数集的数值，验证执行结果并收集耗时及性能消耗情况，评估优化效果，更新参数预测模型。多次对比测试后，参数预测模型愈加完善，最终输出最佳参数调优方案及参数集。

[0027] 可选地，在本实施例中，如图3所示的实现流程图，示例性地，应用将SQL语句集中的各类业务SQL语句的每个SQL语句依次输入至已预置好铺底数据的初始数据库，促使其逐一执行，尽可能的覆盖所有业务场景，并且尽量多次下发，统计信息更加准确。这一过程中，数据库记录每个SQL语句的运行数据，包括执行耗时和初始数据库中每个节点的服务器性能消耗(如CPU、内存、磁盘I/O、网络带宽等)以及SQL语句在各个节点的详细执行耗时情况，细化至各执行阶段。这些数据构成了第一数据集，及时上传至日志分析模块。日志分析模块解析各SQL语句生成执行计划树，根据SQL语句的执行计划树的指纹进行分类汇总SQL语句的运行数据，并直观展示SQL语句耗时瓶颈处以及性能消耗情况。分析处理模块接收分类汇总后的各类SQL语句的运行数据，计算每类SQL语句的综合评价指标值，该指标值结合了执行耗时和性能消耗，为性能评估提供量化依据。根据第一数据集，分析出对性能瓶颈有显著影响的数据库参数，从全量参数集当中筛选出需要优化的第一参数集，这些参数通常与SQL执行的关键阶段相关，如索引配置、缓存设置、并行度等。将第一参数集和各类SQL的综合评价指标值输入预测模型，预测模型通过初步分析优化第一参数集，生成初次优化后的第二参数集，即基于当前性能数据得出的改进建议。使用预测模型和初始数据库对第二参数集进行多轮优化，逐步调整参数值并验证优化效果。每次迭代后，预测模型会根据新的性能数据调整优化方向，最终生成最优参数集。

[0028] 可选地，在本实施例中，基于SQL执行数据和节点性能消耗的全面分析，能够准确识别性能瓶颈所在，例如特定节点资源的高消耗或特定SQL类型的执行延迟。通过筛选和优化第一参数集，优化过程聚焦于对性能有最大影响的关键参数，避免全量参数调优的盲目性和低效性。结合预测模型的自动化优化和迭代分析，大幅减少人工调整参数的时间和工作量，提高调优效率。多轮迭代优化

确保数据库的参数配置接近性能最优，适应多样化业务场景需求，提升分布式数据库的整体运行效率。

[0029] 在一些实施例中，对第一数据集进行分类汇总，得到每一类SQL语句的运行数据包括：

[0030] 获取初始数据库在执行每一个SQL语句时生成的执行计划树；

[0031] 根据每一个SQL语句的执行计划树的指纹，对所有SQL语句进行分类，得到多类SQL语句；

[0032] 将目标类SQL语句中的每一个SQL语句的运行数据进行汇总，得到目标类SQL语句的运行数据，其中，目标类SQL语句为多类SQL语句中的任意一类SQL语句。

[0033] 可选地，在本实施例中，如图4所示的日志分析模块的分析流程图，示例性地，在数据库执行每个SQL语句时，系统会生成相应的执行计划树，执行计划树详细描述了SQL语句的执行流程，包括操作步骤(如索引扫描、表扫描、连接操作等)和资源分配(如CPU、内存等)。这些计划是SQL语句执行过程的关键数据结构，能够反映执行效率和潜在瓶颈。利用每个SQL语句的执行计划树指纹对SQL语句进行分类，执行计划树指纹是一种特征标识，能够唯一表示SQL语句执行逻辑结构的核心模式。通过对指纹的分类，可以将具有相似执行逻辑的SQL语句归为一类，形成多类SQL语句。针对目标类SQL语句(多类SQL中的任意一类)，将该类所有SQL语句的运行数据(如节点性能消耗、执行耗时等)进行汇总，汇总结果包括平均值、峰值、分布统计等，提供目标类SQL语句整体性能表现的全貌。

[0034] 可选地，在本实施例中，基于执行计划树分类，可以精准聚焦不同SQL类型的优化需求，避免“一刀切”的泛化调优，提高针对性。对同类SQL的汇总分析，能发现模式化的性能瓶颈(如特定索引无效、资源争用等)，为优化提供数据支撑。汇总过程覆盖多维性能数据，使结果更具代表性，便于后续综合评价指标计算和性能优化方案设计。分类汇总的数据为后续预测模型提供高质量输入，提升参数优化的准确性和迭代效果。

[0035] 在一些实施例中，在根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值之前，上述方法还包括：

[0036] 通过自适应算法，去除第一数据集中的所有异常值，得到更新后的第一数据集。

[0037] 可选地，在本实施例中，在第一数据集中，可能存在多种异常值，如突发的服务器资源占用(其他任务干扰)、网络波动、硬件故障等引起的毛刺及特殊情况的数据信息，这些异常值可能严重偏离正常数据分布，对分析结果造成误导。因此，日志分析模块在对第一数据集进行分类处理之前，利用自适应算法(如最小二乘法、最小均方算法、维纳滤波等)，动态识别和去除异常值，自适应算法能够根据数据的统计特性(如分布、趋势)自动调整参数，确保去除异常值的同时保留正常数据。去除异常值后，得到清洗后的第一数据集，其数据更加真实和可靠，能够准确反映数据库运行的实际性能。

[0038] 可选地，在本实施例中，去除异常值后，数据分布更加合理，避免了异常点对统计结果的干扰，提高运行数据分类与汇总的准确性。消除了异常值可能引发的错误决策，确保综合评价指标和优化策略基于正常运行数据，提升整体分析效果。数据清洗减少了异常值带来的计算偏差，使后续性能优化的方向更明确，降低无效优化的概率。

[0039] 在一些实施例中，根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值包括：

[0040] 将每一类SQL语句作为当前类SQL语句，对当前类SQL语句执行如下操作：

[0041] 获取当前类SQL语句各个执行阶段的性能消耗数据和执行耗时数据，并将性能消耗数据和执行耗时数据转换为可比单位；

[0042] 为当前类SQL语句各个执行阶段的性能消耗数据和执行耗时数据分配权重因子；

[0043] 将当前类SQL语句的各个执行阶段的性能消耗数据和执行耗时数据分别乘以对应的权重因子，得到当前类SQL语句的各个执行阶段的加权结果；

[0044] 将当前类SQL语句的各个执行阶段的加权结果进行求和，得到当前类SQL语句的综合评价指标值。

[0045] 可选地，在本实施例中，如图5所示的分析处理模块的分析流程图，示例性地，对每一类SQL语句的运行数据，提取其在不同执行阶段(如解析、优化、计

划生成、执行等)产生的性能消耗数据(如CPU、内存、磁盘I/O、网络带宽等)和执行耗时数据，这些数据全面反映了SQL在各阶段的资源使用和性能表现。将性能消耗数据和执行耗时数据转换为可比单位(如转换为百分比或归一化到[0，1]区间)，标准化成相同的量纲(如将内存消耗转换为MB/s，耗时转换为毫秒/ms)，以便后续计算中保持一致性，避免因单位不统一导致结果失真。根据业务需求或实际场景，为SQL语句的各个执行阶段分配不同的权重因子，权重因子表示每个阶段对SQL语句综合评价的重要程度，例如，若业务更关注执行速度，则执行耗时的权重因子可能更高，而对于资源受限场景，性能消耗的权重因子可能更高。将每类SQL语句各个执行阶段的性能消耗数据和执行耗时数据分别乘以对应的权重因子，得到各执行阶段的加权结果，这一过程结合了不同阶段的重要性，使评价结果更加符合业务实际。对每类SQL语句的各个执行阶段的加权结果进行求和，得到该类SQL语句的综合评价指标值，这一指标是对SQL整体性能的量化评价，综合反映了执行效率和资源使用的合理性。

[0046] 可选地，在本实施例中，通过单位转换和加权计算，将各类性能数据和耗时指标统一为可比值，使评价结果具有标准化特性，便于SQL类型间的横向对比。权重因子的设置可根据不同业务场景灵活调整，使评价指标更贴合实际需求，例如提高对执行效率或资源消耗的关注度。综合评价指标值能帮助快速发现某类SQL的主要性能瓶颈(如特定阶段资源消耗过高或耗时过长)，为后续优化提供明确方向。评价指标为SQL分类优化和参数调优提供量化依据，简化优化过程，提高调优效率和效果。

[0047] 在一些实施例中，根据第一数据集，从初始数据库的全量参数集当中分析出需要优化的第一参数集包括：

[0048] 根据第一数据集，确定出瓶颈阶段；

[0049] 从全量参数集当中筛选出与瓶颈阶段相关的第三参数集；

[0050] 确定第三参数集中的每一个参数的优化范围；

[0051] 汇总第三参数集及第三参数集中每一个参数的优化范围，得到第一参数集。

[0052] 可选地，在本实施例中，如图5所示的分析处理模块的分析流程图，示例性地，根据第一数据集(包括SQL执行的性能消耗数据和执行耗时数据)，分析数据

库在处理SQL语句时的各执行阶段(如解析、优化、计划生成、执行等)耗时和资源消耗的分布情况。通过对比各阶段的耗时占比和资源消耗比例，定位对整体性能影响较大的瓶颈阶段。例如，若执行阶段的耗时显著高于其他阶段或者设定值，则可以判定其为瓶颈阶段。从初始数据库的全量参数集当中，筛选出与瓶颈阶段直接相关的参数，形成第三参数集。对第三参数集中的每一个参数，结合第一数据集的运行数据和数据库参数设计文档，确定其合理的优化范围。例如，线程池大小可能有一个有效区间(如4−64个线程)，调整过小可能导致并发不足，过大则会消耗额外资源，优化范围可以根据实际瓶颈和数据库负载情况进行缩减。将筛选出的第三参数集及第三参数集中每一个参数的优化范围进行汇总，形成第一参数集。第一参数集是对初始数据库参数的子集，用于参数迭代优化，专注于与瓶颈阶段相关的参数，以缩小优化的范围，提升参数调整的效率。

[0053] 可选地，在本实施例中，瓶颈阶段分析明确了性能问题的重点区域，从而聚焦于最关键的参数，避免全量参数调整带来的复杂性。通过筛选相关参数和限定优化范围，缩小了调整的参数集合，显著降低了调优过程的时间和计算成本。第一参数集以瓶颈阶段为核心，确保优化的针对性和有效性，能够快速取得性能提升。根据不同的瓶颈阶段动态生成第一参数集，使优化过程能够适配不同的SQL类型和运行场景，增强分布式数据库的调优灵活性。

[0054] 在一些实施例中，使用初始数据库和预测模型对第二参数集进行多次迭代优化，得到初始数据库的最优参数集包括：

[0055] 获取第M次优化所用到的第二参数集和数据库，其中，M为大于1的正整数，在M为2时，第M次优化所用到的第二参数集为初次优化后的第二参数集，第M次优化所用到的数据库为初始数据库，在M为大于2时，第M次优化所用到的第二参数集为第M−1次优化后的第二参数集，第M次优化所用到的数据库为第M−1次优化后的数据库；

[0056] 根据第M次优化所用到的第二参数集，优化第M次优化所用到的数据库，得到第M次优化后的数据库；

[0057] 基于SQL语句集和第M次优化后的数据库重新计算每一类SQL语句的综合评价指标值，得到每一类SQL语句第M次优化后的综合评价指标值；

[0058] 将所有第M次优化后的综合评价指标值和第M次优化所用到的第二参数集输入至预测模型，以使预测模型根据所有第M次优化后的综合评价指标值，对第M次优化所用到的第二参数集进行优化，得到第M次优化后的第二参数集，更新M＝M+1，并执行第M+1次优化，直到M等于目标阈值，得到最优参数集。

[0059] 可选地，在本实施例中，第一次优化的输入是初次优化后的第二参数集和初始数据库，随着迭代优化的进行，第M次优化的输入包括第(M−1)次优化生成的第二参数集，和第(M−1)次优化生成的数据库实例。根据第M次优化所用到的第二参数集，对第M次数据库进行调整，包括修改配置参数(如并行度、缓存策略等)并重启或热更新数据库，生成第M次优化后的数据库实例。使用SQL语句集对第M次优化后的数据库进行测试，运行多类SQL语句，统计每一类SQL的性能数据(耗时、CPU、内存、磁盘I/O等)，根据权重因子重新计算第M次优化后的综合评价指标值。将第M次优化后的综合评价指标值与当前的第二参数集作为输入，交由预测模型处理，预测模型通过分析新指标值的变化趋势，进一步调整参数集，生成下一轮优化的第二参数集。每次优化完成后，将迭代次数M增加1，并以新的参数集和数据库实例进入下一轮优化。人工监控并分析当前参数调优的合理性，若部分参数不合理，人工可干预参数进而改变参数预测模型，使其更加适合应用需求。当M达到目标阈值(如性能稳定或未显著提升时)，终止迭代，并输出最终的最优参数集。

[0060] 可选地，在本实施例中，通过多次迭代优化，每次以更准确的指标值指导下一轮参数调整，实现对最优参数的渐进逼近。预测模型能基于历史优化数据调整策略，适应不同的SQL负载特征和数据库架构变化。最终输出的最优参数集能够平衡资源利用率和SQL执行效率，显著提升数据库性能，同时确保稳定性。采用自动化迭代优化，降低人工分析参数的复杂性，提高优化效率和精准性。

[0061] 在一些实施例中，基于SQL语句集和第M次优化后的数据库重新计算每一类SQL语句的综合评价指标值，得到每一类SQL语句第M次优化后的综合评价指标值包括：

[0062] 将SQL语句集中的每一个SQL语句输入至第M次优化后的数据库，以使第M次优化后的数据库执行每一个SQL语句，并输出每一个SQL语句的运行数据，得到第M次优化后的第一数据集；

[0063] 对第M次优化后的第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句第M次优化后的综合评价指标值。

[0064] 可选地，在本实施例中，将SQL语句集中的每个SQL语句逐一输入到第M次优化后的数据库执行，数据库在执行SQL语句的过程中，记录运行数据，包括各节点的性能消耗(CPU、内存、磁盘I/O等)，SQL语句在各节点的执行耗时，输出运行数据，形成第M次优化后的第一数据集。根据SQL语句的执行计划树指纹对第M次优化后的第一数据集进行分类，将所有SQL语句划分为多个类别，对每一类SQL语句中的运行数据(耗时和性能消耗)进行汇总，得到每一类SQL语句的运行数据。对每一类SQL语句的运行数据，提取执行阶段的性能消耗数据和执行耗时数据，进行单位统一处理(如转换为毫秒、百分比等)。将每个执行阶段的性能消耗数据和执行耗时数据分别乘以对应的权重因子，得到每个执行阶段的加权结果，对所有执行阶段的加权结果求和，计算出当前类SQL语句的综合评价指标值。重复上述过程，计算所有类别的SQL语句的综合评价指标值，得到第M次优化后的完整评价指标。

[0065] 可选地，本申请的核心在于利用数据库的耗时日志和服务器性能消耗信息进行数据预处理、分类、汇总、计算和迭代优化，以最终获得适用于数据库的最优参数集及参数值。其中，SQL特征分类还可以根据其他规则，如指定的HINT信息（即SQL提示，用于引导数据库生成SQL执行计划的特殊指令）、预编译模板化特征等，灵活调整。预测模型可采用人工智能算法，包括以下两类替代方案：模型类型：监督学习(如回归分析)、非监督学习(如聚类)、半监督学习、强化学习等；训练算法：结合不同算法，如回归算法、决策树、深度学习、关联规则学习等，选择适合的优化方法。

[0066] 对于前述的各方法实施例，为了简单描述，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，本申请并不受所描述的动作顺序的限制，

因为依据本申请，一些步骤可以采用其他顺序或者同时进行。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示例性实施例，所涉及的动作和模块并不一定是本申请所必须的。

[0067] 本申请实施例提供的上述方案与相关技术相比具有如下特点：

[0068] 本申请采用了将SQL语句集中的每一个SQL语句输入至初始数据库，以使初始数据库执行每一个SQL语句，并输出每一个SQL语句的运行数据，得到第一数据集，其中，SQL语句集中包括多类SQL语句，运行数据包括初始数据库执行SQL语句时，初始数据库的每一个节点的服务器的性能消耗数据和SQL语句在每一个节点的执行耗时数据；对第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；根据第一数据集，从初始数据库的全量参数集当中分析出需要优化的第一参数集；将所有综合评价指标值和第一参数集输入至预测模型，以使预测模型根据所有综合评价指标值对第一参数集进行初次优化，得到初次优化后的第二参数集；使用初始数据库和预测模型对第二参数集进行多次迭代优化，得到初始数据库的最优参数集的方法，由于在上述数据库性能参数的优化方法中，通过将SQL语句集中的每个SQL语句输入初始数据库执行，采集包括详细执行耗时、每个节点的服务器性能消耗等运行数据，形成第一数据集。对数据进行分类汇总后，计算每类SQL的综合评价指标值，并从全量参数集当中筛选出需优化的第一参数集。随后，将综合评价指标值和第一参数集输入预测模型，生成初次优化的第二参数集，并通过多轮迭代调整，最终得到最优参数集，从而实现了全面采集性能数据和SQL耗时，精准定位性能瓶颈，聚焦关键参数，避免盲目调优，结合预测模型自动化调整参数，提高调优效率，多轮迭代优化确保数据库性能持续提升，满足复杂业务场景需求，显著降低分布式架构下的调优难度，进而解决了由于分布式数据库架构的复杂性而导致的参数调优难度大的问题。

[0069] 根据本申请实施例的另一方面，还提供了一种数据库性能参数的优化装置，如图6所示，包括：

[0070] 执行模块602，设置为将SQL语句集中的每一个SQL语句输入至初始数据库，以使初始数据库执行每一个SQL语句，并输出每一个SQL语句的运行数据，得到第一数据集，其中，SQL语句集中包括多类SQL语句，运行数据包括初始数据库执行SQL语句时，初始数据库的每一个节点的服务器的性能消耗数据和SQL语句在每一个节点的执行耗时数据；

[0071] 计算模块604，设置为对第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；

[0072] 分析模块606，设置为根据第一数据集，从初始数据库的全量参数集当中分析出需要优化的第一参数集；

[0073] 第一优化模块608，设置为将所有综合评价指标值和第一参数集输入至预测模型，以使预测模型根据所有综合评价指标值对第一参数集进行初次优化，得到初次优化后的第二参数集；

[0074] 第二优化模块610，设置为使用初始数据库和预测模型对第二参数集进行多次迭代优化，得到初始数据库的最优参数集。

[0075] 该实施例中的执行模块602可以设置为执行本申请实施例中的步骤S102，该实施例中的计算模块604可以设置为执行本申请实施例中的步骤S104，该实施例中的分析模块606可以设置为执行本申请实施例中的步骤S106，该实施例中的第一优化模块608可以设置为执行本申请实施例中的步骤S108，该实施例中的第二优化模块610可以设置为执行本申请实施例中的步骤S110。

[0076] 在一些实施例中，计算模块包括：

[0077] 第一获取单元，设置为获取初始数据库在执行每一个SQL语句时生成的执行计划树；

[0078] 分类单元，设置为根据每一个SQL语句的执行计划树的指纹，对所有SQL语句进行分类，得到多类SQL语句；

[0079] 第一汇总单元，设置为将目标类SQL语句中的每一个SQL语句的运行数据进行汇总，得到目标类SQL语句的运行数据，其中，目标类SQL语句为多类SQL语句中的任意一类SQL语句。

[0080] 在一些实施例中，上述装置还包括：

[0081] 去除模块，设置为在根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值之前，通过自适应算法，去除第一数据集中的所有异常值，得到更新后的第一数据集。

[0082] 在一些实施例中，计算模块包括：

[0083] 处理单元，设置为将每一类SQL语句作为当前类SQL语句，对当前类SQL语句执行如下操作：

[0084] 获取当前类SQL语句各个执行阶段的性能消耗数据和执行耗时数据，并将性能消耗数据和执行耗时数据转换为可比单位；

[0085] 为当前类SQL语句各个执行阶段的性能消耗数据和执行耗时数据分配权重因子；

[0086] 将当前类SQL语句的各个执行阶段的性能消耗数据和执行耗时数据分别乘以对应的权重因子，得到当前类SQL语句的各个执行阶段的加权结果；

[0087] 将当前类SQL语句的各个执行阶段的加权结果进行求和，得到当前类SQL语句的综合评价指标值。

[0088] 在一些实施例中，分析模块包括：

[0089] 第一确定单元，设置为根据第一数据集，确定出瓶颈阶段；

[0090] 筛选单元，设置为从全量参数集当中筛选出与瓶颈阶段相关的第三参数集；

[0091] 第二确定单元，设置为确定第三参数集中的每一个参数的优化范围；

[0092] 第二汇总单元，设置为汇总第三参数集及第三参数集中每一个参数的优化范围，得到第一参数集。

[0093] 在一些实施例中，第二优化模块包括：

[0094] 第二获取单元，设置为获取第M次优化所用到的第二参数集和数据库，其中，M为大于1的正整数，在M为2时，第M次优化所用到的第二参数集为初次优化后的第二参数集，第M次优化所用到的数据库为初始数据库，在M为大于2时，第M次优化所用到的第二参数集为第M−1次优化后的第二参数集，第M次优化所用到的数据库为第M−1次优化后的数据库；

[0095] 第一优化单元，设置为根据第M次优化所用到的第二参数集，优化第M次优化所用到的数据库，得到第M次优化后的数据库；

[0096] 计算单元，设置为基于SQL语句集和第M次优化后的数据库重新计算每一类SQL语句的综合评价指标值，得到每一类SQL语句第M次优化后的综合评价指标值；

[0097] 第二优化单元，设置为将所有第M次优化后的综合评价指标值和第M次优化所用到的第二参数集输入至预测模型，以使预测模型根据所有第M次优化后的综合评价指标值，对第M次优化所用到的第二参数集进行优化，得到第M次优化后的第二参数集，更新M＝M+1，并执行第M+1次优化，直到M等于目标阈值，得到最优参数集。

[0098] 在一些实施例中，计算单元包括：

[0099] 执行子单元，设置为将SQL语句集中的每一个SQL语句输入至第M次优化后的数据库，以使第M次优化后的数据库执行每一个SQL语句，并输出每一个SQL语句的运行数据，得到第M次优化后的第一数据集；

[0100] 计算子单元，设置为对第M次优化后的第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句第M次优化后的综合评价指标值。

[0101] 本实施例的其他示例请参见上述示例，在此不在赘述。

[0102] 图7是根据本申请实施例的一种可选的电子设备的示意图，如图7所示，包括处理器702、通信接口704、存储器706和通信总线708，其中，处理器702、通信接口704和存储器706通过通信总线708完成相互间的通信，其中，

[0103] 存储器706，设置为存储计算机程序；

[0104] 处理器702，设置为执行存储器706上所存放的计算机程序时，实现如下步骤：

[0105] 将SQL语句集中的每一个SQL语句输入至初始数据库，以使初始数据库执行每一个SQL语句，并输出每一个SQL语句的运行数据，得到第一数据集，其中，SQL语句集中包括多类SQL语句，运行数据包括初始数据库执行SQL语句时，初始数据库的每一个节点的服务器的性能消耗数据和SQL语句在每一个节点的执行耗时数据；

[0106] 对第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；

[0107] 根据第一数据集，从初始数据库的全量参数集当中分析出需要优化的第一参数集；

[0108] 将所有综合评价指标值和第一参数集输入至预测模型，以使预测模型根据所有综合评价指标值对第一参数集进行初次优化，得到初次优化后的第二参数集；

[0109] 使用初始数据库和预测模型对第二参数集进行多次迭代优化，得到初始数据库的最优参数集。

[0110] 可选地，在本实施例中，上述的通信总线可以是外设部件互连标准(Peripheral Component Interconnect，PCI)总线、或扩展工业标准结构(Extended Industry Standard Architecture，EISA)总线等。该通信总线可以分为地址总线、数据总线、控制总线等。为便于表示，图7中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。通信接口用于上述电子设备与其他设备之间的通信。

[0111] 存储器可以包括随机存取器(Random Access Memory，RAM)，也可以包括非易失性存储器(non-volatile memory)，例如，至少一个磁盘存储器。可选地，存储器还可以是至少一个位于远离前述处理器的存储装置。

[0112] 作为一些示例，上述存储器706中可以包括上述数据库性能参数的优化装置中的执行模块602、计算模块604、分析模块606、第一优化模块608以及第二优化模块610。此外，还可以包括上述数据库性能参数的优化装置中的其他模块单元。

[0113] 上述处理器可以是通用处理器，例如可以包括：中央处理器(Central Processing Unit，CPU)、网络处理器(Network Processor，NP)等；还可以是数字信号处理器(Digital Signal Processing，DSP)、专用集成电路(Application Specific Integrated Circuit，ASIC)、现场可编程门阵列(Field-Programmable Gate Array，FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0114] 可选地，本实施例中的示例可以参考上述实施例中所描述的示例，本实施例在此不再赘述。

[0115] 本领域普通技术人员可以理解，图7所示的结构仅为示意，实施上述数据库性能参数的优化方法的设备可以是终端设备，该终端设备可以是智能手机(如安卓

（Android）手机、苹果操作系统（iPhone Operating System，iOS）手机等)、平板电脑、掌上电脑以及移动互联网设备(Mobile Internet Devices，MID)、便携式平板设备（Portable Android Device，PAD）等终端设备。图7并不对上述电子设备的结构造成限定。例如，电子设备还可包括比图7中所示更多或者更少的组件(如网络接口、显示装置等)，或者具有与图7所示的不同的配置。

[0116] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令终端设备相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括：闪存盘、只读存储器(Read-Only Memory，ROM)、RAM、磁盘或光盘等。

[0117] 根据本申请的实施例的又一方面，还提供了一种计算机可读的存储介质，该计算机可读的存储介质中存储有计算机程序，其中，该计算机程序被处理器运行时执行上述数据库性能参数的优化方法中的步骤。

[0118] 可选地，在本实施例中，本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令终端设备相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，存储介质可以包括：闪存盘、ROM、RAM、磁盘或光盘等。

[0119] 上述实施例中的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，可以存储在上述计算机可读取的存储介质中。基于这样的理解，本申请的技术方案本质上或者说对相关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存储介质中，包括若干指令用以使得一台或多台计算机设备(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。

[0120] 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，一些实施例中没有详述的部分，可以参见其他实施例的相关描述。

[0121] 在本申请所提供的几个实施例中，应该理解到，所揭露的客户端，可通过其它的方式实现。其中，以上所描述的装置实施例是示意性的，例如所述单元的划分，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。

[0122] 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单元来实现本实施例方案的目的。

[0123] 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。

权利要求书

[权利要求 1] 一种数据库性能参数的优化方法，包括：

将结构化查询语言SQL语句集中的每一个SQL语句输入至初始数据库，以使所述初始数据库执行每一个所述SQL语句，并输出每一个所述SQL语句的运行数据，得到第一数据集；其中，所述SQL语句集中包括多类SQL语句，所述运行数据包括所述初始数据库执行所述SQL语句时，所述初始数据库的每一个节点的服务器的性能消耗数据和所述SQL语句在每一个节点的执行耗时数据；

对所述第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；

根据所述第一数据集，从所述初始数据库的全量参数集当中分析出需要优化的第一参数集；

将所有所述综合评价指标值和所述第一参数集输入至预测模型，以使所述预测模型根据所有所述综合评价指标值对所述第一参数集进行初次优化，得到初次优化后的第二参数集；

使用所述初始数据库和所述预测模型对所述第二参数集进行多次迭代优化，得到所述初始数据库的最优参数集。

[权利要求 2] 根据权利要求1所述的方法，其中，所述对所述第一数据集进行分类汇总，得到每一类SQL语句的运行数据，包括：

获取所述初始数据库在执行每一个所述SQL语句时生成的执行计划树；

根据每一个所述SQL语句的执行计划树的指纹，对所有所述SQL语句进行分类，得到多类SQL语句；

将目标类SQL语句中的每一个所述SQL语句的运行数据进行汇总，得到所述目标类SQL语句的运行数据，其中，所述目标类SQL语句为多类SQL语句中的任意一类SQL语句。

[权利要求 3] 根据权利要求1所述的方法，在根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值之前，所述方法还包括：

通过自适应算法，去除所述第一数据集中的所有异常值，得到更新后的第一数据集。

[权利要求 4] 根据权利要求1所述的方法，其中，所述根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值，包括：

将每一类SQL语句作为当前类SQL语句，对所述当前类SQL语句执行如下操作：

获取所述当前类SQL语句各个执行阶段的性能消耗数据和执行耗时数据，并将所述性能消耗数据和执行耗时数据转换为可比单位；

为所述当前类SQL语句各个执行阶段的性能消耗数据和执行耗时数据分配权重因子；

将所述当前类SQL语句的各个执行阶段的性能消耗数据和执行耗时数据分别乘以对应的权重因子，得到所述当前类SQL语句的各个执行阶段的加权结果；

将所述当前类SQL语句的各个执行阶段的加权结果进行求和，得到所述当前类SQL语句的综合评价指标值。

[权利要求 5] 根据权利要求1所述的方法，其中，所述根据所述第一数据集，从所述初始数据库的全量参数集当中分析出需要优化的第一参数集，包括：

根据所述第一数据集，确定出瓶颈阶段；

从所述全量参数集当中筛选出与所述瓶颈阶段相关的第三参数集；

确定所述第三参数集中的每一个参数的优化范围；

汇总所述第三参数集及所述第三参数集中每一个参数的优化范围，得到所述第一参数集。

[权利要求 6] 根据权利要求1所述的方法，其中，所述使用所述初始数据库和所述预测模型对所述第二参数集进行多次迭代优化，得到所述初始数据库的最优参数集，包括：

获取第M次优化所用到的第二参数集和数据库，其中，所述M为大于1的正整数，在所述M为2时，第M次优化所用到的第二参数集为所述初次优化后的第二参数集，第M次优化所用到的数据库为初始数据库，在所述M为大于2时，第M次优化所用到的第二参数集为第M−1次优化后的第二参数集，第M次优化所用到的数据库为第M−1次优化后的数据库；

根据所述第M次优化所用到的第二参数集，优化所述第M次优化所用到的数据库，得到第M次优化后的数据库；

基于所述SQL语句集和所述第M次优化后的数据库重新计算每一类SQL语句的综合评价指标值，得到每一类SQL语句第M次优化后的综合评价指标值；

将所有所述第M次优化后的综合评价指标值和所述第M次优化所用到的第二参数集输入至所述预测模型，以使所述预测模型根据所有所述第M次优化后的综合评价指标值，对所述第M次优化所用到的第二参数集进行优化，得到第M次优化后的第二参数集，更新M＝M+1，并执行第M+1次优化，直到所述M等于目标阈值，得到所述最优参数集。

[权利要求 7] 根据权利要求6所述的方法，其中，所述基于所述SQL语句集和所述第M次优化后的数据库重新计算每一类SQL语句的综合评价指标值，得到每一类SQL语句第M次优化后的综合评价指标值，包括：

将所述SQL语句集中的每一个SQL语句输入至所述第M次优化后的数据库，以使所述第M次优化后的数据库执行每一个所述SQL语句，并输出每一个所述SQL语句的运行数据，得到第M次优化后的第一数据集；

对所述第M次优化后的第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句第M次优化后的综合评价指标值。

[权利要求 8] 一种数据库性能参数的优化装置，包括：

执行模块，设置为将SQL语句集中的每一个SQL语句输入至初始数据库，以使所述初始数据库执行每一个所述SQL语句，并输出每一个所述SQL语句的运行数据，得到第一数据集；其中，所述SQL语句集中包括多类SQL语句，所述运行数据包括所述初始数据库执行所述SQL语句时，所述初始数据库的每一个节点的服务器的性能消耗数据和所述SQL语句在每一个节点的执行耗时数据；

计算模块，设置为对所述第一数据集进行分类汇总，得到每一类SQL语句的运行数据，并根据每一类SQL语句的运行数据，计算得到每一类SQL语句的综合评价指标值；

分析模块，设置为根据所述第一数据集，从所述初始数据库的全量参数集当中分析出需要优化的第一参数集；

第一优化模块，设置为将所有所述综合评价指标值和所述第一参数集输入至预测模型，以使所述预测模型根据所有所述综合评价指标值对所述第一参数集进行初次优化，得到初次优化后的第二参数集；

第二优化模块，设置为使用所述初始数据库和所述预测模型对所述第二参数集进行多次迭代优化，得到所述初始数据库的最优参数集。

[权利要求 9] 一种计算机可读的存储介质，所述计算机可读的存储介质存储有计算机程序，所述计算机程序被处理器运行时执行所述权利要求1至7任一项中所述的数据库性能参数的优化方法。

[权利要求 10] 一种电子设备，包括存储器和处理器，所述存储器中存储有计算机程序，所述处理器被设置为通过所述计算机程序执行所述权利要求1至7任一项中所述的数据库性能参数的优化方法。