

信创架构的基于多 Agent 协作实时数据库查询系统设计与实现

廖化锋¹
LIAO Huafeng

摘要

针对工业级实时数据库数据采集和存储中存在的压缩率增加、延迟高等问题,文章设计了一种基于多 Agent 协作机制的查询系统。系统硬件包括了测量传感器、网络交换机、路由器等,在软件设计方面定义了多 Agent 协作机制,通过对协作机制内数据的分解,解析 SQL 查询语句,从而实现对语句指令的并发控制,完成工业级实时数据库协同查询方法的设计。实际测试的结果表明,所提系统能够提升数据压缩率,且压缩过程中将被采样数据的时延水平控制在 50 ms 以内。由此得出结论,基于多 Agent 协作机制的查询系统实现了工业级数据库对实时数据的高密度存储与快速采集,提升数据查询和处理的效率。

关键词

协同查询;多 Agent 机制;数据分解;SQL 语句

doi: 10.3969/j.issn.1672-9528.2026.03.034

0 引言

工业级实时数据库结合了数据库技术与实时系统的应用优势,专门用于收集、存储、处理和分析来自各种工业设备和传感器的实时数据,在各种环境下都能稳定工作,能够处理大量数据,特别是在数据量急剧增加时可以依然保持稳定的运行状态。采集层负责从各种设备和传感器中收集数据,且被采集到的数据样本可以直接通过无线网络传输至数据库单元之中^[1]。处理层具有实时处理与分析数据的能力,包括进行数据的清洗、聚合和计算等操作,通常采用高性能计算和并行处理技术,以确保数据处理的实时性与高效性。

对于工业级实时数据库的应用,桂智明等^[2]提出了区块链与数据库混合架构下的查询方法,集成空间数据库与 Hy-

perledgerFabric(开源型企业级区块链框架),并将空间操作函数嵌入智能合约中,实现了对查询语句的分布式部署。该查询方法保留了区块链体系原有的分布式特点,能够在查询存储数据的同时,完成对关键信息的实时调用。然而,分布式机制需要的响应时间较长,故该方法并不满足工业级实时数据库快速采集样本信息的需求。刘鹏举等^[3]提出了基于谱聚类的查询方法,根据谱聚类条件生成数据样本的初始分区,再通过对解空间的精细搜索,完成对初始分区的优化。在实际应用过程中,应用该方法不需要完整扫描数据库就可以完成查询任务,故其对于数据访问量的控制较为理想,但非完整扫描意味着数据分区的策略粒度相对较粗,相似数据极易被归于同一分区之中,不利于实现数据样本的高密度压缩。

为解决上述问题,同步实现数据的高密度存储与快速采集,引入信创架构,设计新型的工业级实时数据库协同查询方法。信创架构旨在推动信息技术创新和应用,提高开发效率,

1. 闽西职业技术学院 福建龙岩 364000

[6] 赵丹阳.地市级供电公司会议室调度优化研究[D].天津:天津大学,2019.

[7] 尚玉梅.高校办公室预约管理软件的设计与实现[D].成都:电子科技大学,2014.

【作者简介】

梁雄(2004—),男,福建莆田人,本科,研究方向:计算机应用工程。

陈凌煜(2001—),通信作者(email:chenlingyu@qvtu.edu.cn),男,福建漳州人,本科,研究方向:人工智能与应用系统开发。

吴文辉(1982—),男,福建泉州人,专科,主任,研究方向:高校信息化。

柯燕秋(1998—),女,福建泉州人,本科,研究实习员,研究方向:办公流程管理与信息化。

陈子烜(2004—),男,福建福州人,本科,研究方向:软件工程技术。

黄星涵(2006—),男,江西宜春人,本科,研究方向:软件工程技术。

吴诗淇(2005—),男,福建泉州人,本科,研究方向:软件工程技术。

(收稿日期:2025-08-01 修回日期:2026-03-06)

优化生产过程,提供智能、精准和个性化的服务,并确保信息技术应用的安全性和可靠性。该架构体系在芯片、存储、服务器、云计算等关键组件上实现了自主可控,且在操作系统、数据库、中间件等基础软件上也进行自主研发^[4]。ARM架构作为信创架构的硬件基础,具有功耗低、性能高的优势,其应用有助于实现计算密集型和嵌入式应用的自主可控。

1 系统框架与硬件设计

信创架构的硬件框架体系应具备高可靠性,确保数据在任何情况下都能被完整、准确地存储和处理,涉及数据采集、数据传输、数据处理和存储、数据备份和恢复四个组成部分。对于工业级实时数据库而言,其所配置的信创架构硬件应支持高效的数据处理和分析,既可以降低系统响应时间,也能够提高传输数据的响应性能^[5]。具体的模型结构如图1所示。

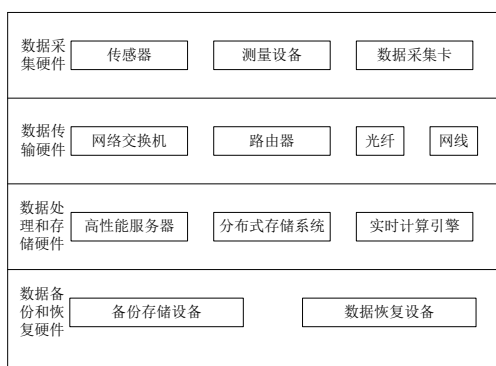


图1 信创架构的硬件框架模型

(1) 数据采集硬件:传感器和测量设备:负责采集工业过程中的各种数据;数据采集卡:可将基层传感器采集到的模拟信号对象转化为数字信号对象,并传输到数据处理中心。

(2) 数据传输硬件:网络交换机和路由器:负责数据的传输和路由,确保数据能够实时、准确的从数据源传输到数据处理中心;光纤和网线:作为数据传输的物理介质,应具备高速、低延迟和高可靠性的特点。

(3) 数据处理和存储硬件:高性能服务器:用于处理和分析实时数据,应具备多核处理器、大容量内存和高速存储设备;分布式存储系统:利用多台服务器和存储设备构建分布式存储系统,提高数据的可靠性和可扩展性;实时计算引擎:用于实现实时数据的处理和分析。

(4) 数据备份和恢复硬件:备份存储设备,用于存储数据的备份副本,以防止数据丢失或损坏;数据恢复设备:在数据丢失或损坏时,用于恢复数据的设备或系统。

2 系统软件设计与实时数据库的多 Agent 协作机制

2.1 基于多 Agent 协作机制的系统软件

数据处理层具有清洗与加工传输数据的能力,能够满足不同的数据处理与分析需求^[6]。该层可以采用大数据处理和分析技术,还可以采用数据压缩和编码技术,以减少存储和

传输成本。须将数据样本存储于不同的节点之中,才能够提供架构体系的可扩展性^[7]。可以选择大数据处理和分析技术,如 Spark、Flink 等,对数据进行高效处理和分析。工业级实时数据库的多 Agent 协作机制是一种分布式智能模型,通过多个具有自主决策和交互能力的智能体 (Agent) 之间的协作,共同完成任务。在信创架构中,每个 Agent 都拥有独特的知识和信息。通过信息交流和共享,Agent 们能够获取更全面的环境信息和状态信息,从而作出更准确的决策^[8]。若 z_1 、 z_2 、 $\dots$ 、 z_n 表示 n 个 Agent 对象,其与各自所对应数据库之间的信息交互关系可表示为:

$$\begin{cases} X_1 = |\omega_1 z_1|^{-\sigma} \\ X_2 = |\omega_2 z_2|^{-\sigma} \\ \vdots \\ X_n = |\omega_n z_n|^{-\sigma} \end{cases} \quad (1)$$

式中: σ 表示信创架构中的信息决策参数; ω_1 、 ω_2 、 $\dots$ 、 ω_n 表示不同的数据交换向量。

信创架构根据每个 Agent 的能力和当前环境状态,将任务合理地分配给各个 Agent,以实现数据库整体性能最优或任务完成时间最短^[9]。对于工业级实时数据库而言,决策制定通常需要通过协调机制来确保各个 Agent 的决策能够协同一致,共同达成系统目标。设 τ_1 、 τ_2 、 $\dots$ 、 τ_n 表示信创架构针对不同 Agent 所设置的任务分配向量,联立式 (1),可将实时数据库的多 Agent 协作机制表达式定义为:

$$C = \frac{|\tau_1 X_1 + \tau_2 X_2 + \dots + \tau_n X_n|^2}{\bar{V} \times (\psi - 1)} \quad (2)$$

式中: ψ 表示数据库的整体性能参量; $\bar{V}$ 表示信创架构中的数据转存向量。

在工业环境中,多个 Agent 可以协同工作,从各种传感器和设备中实时采集数据,且这些 Agent 对象可以根据实时数据和任务需求,优化生产资源的分配和调度。

2.2 SQL 查询语句

通过数据分解,可以将这些复杂的数据集拆分成更小、更易于管理的部分,从而提高实时数据库的处理和分析效率^[10]。信创架构下工业实时数据库中存储的数据大多属于时序数据,即按照时间顺序记录的数据。时序数据的分解可以根据时间间隔、数据范围条件将数据拆分成多个时间段或数据块,以便进行更高效的数据存储和查询^[11]。例如, v 表示一个随机工业数据样本; $\mathfrak{S}_v$ 、 $\mathfrak{S}_v$ 表示两个包含样本 v 的数据集,且二者不存在包含关系,在数据集 $\mathfrak{S}_v$ 中取一个时间戳参量 b_1 ,在数据集 $\mathfrak{S}_v$ 中取一个时间戳参量 b_2 ,联立公式 (2),可将工业实时数据库中的数据分解表达式定义为:

$$P=\zeta\left(\frac{b_1\times b_2}{|\Delta B|}\right)^2+1+C^2\tilde{m}\tag{3}$$

式中： ΔB 表示单位时间间隔内所存储的工业数据总量； ζ 表示数据范围定义参数； $\tilde{m}$ 表示工业数据样本的拆分向量。

时序分解有助于实现工业数据的局部化和优化处理，提高实时数据库对数据的处理效率和准确性。

工业级实时数据库通常需要对大量数据进行快速读写操作，因此查询语句的性能优化非常重要。在实时环境中，数据一致性是一个关键问题，由于查询结果能够反映最新的数据状态，所以在解析查询语句时，需准确把握当前数据样本在数据库中的分解情况^[12]。结构化查询语言（structured query language, SQL）是用于管理和操作关系数据库的标准语言，该类型查询语句用于从数据库中检索、插入、更新或删除数据^[13]。针对 SQL 查询语句的解析应始终对输入进行验证和清理，以防止 SQL 注入攻击。对于信创架构体系来说，为实现对工业级数据库的协同查询，在定义语句指令时，必须了解数据库的索引和查询优化技术，以提高 SQL 命令的查询性能。

规定 M_p 表示基于式（3）所示数据分解条件所定义的 SQL 数据库查询语句，其计算式为：

$$M_p=\frac{1}{P}\left[\frac{A}{2}+\vartheta\overline{S}-\sqrt{1+\left(\frac{\dot{d}}{f}\right)^2}\right]\tag{4}$$

式中： A 表示工业数据的结构化向量； ϑ 表示数据库更新参数； $\overline{S}$ 表示数据样本的单位存储均值； $\dot{d}$ 表示工业级实时数据库中的实时数据检索特征； f 表示查询语句插入项。

在式（4）的基础上，推导 SQL 查询语句解析表达式为：

$$D=M_p(\theta-\varpi)^2\bigg/\sqrt{\hat{g}-1}\tag{5}$$

式中： θ 表示数据分解系数； ϖ 表示协同标记参数； $\hat{g}$ 表示 SQL 查询语句中目标数据样本的标准解析向量。通过理解和使用 SQL 查询语句，可以有效地管理和操作工业级实时数据库，从而满足各种数据检索、更新和协同处理需求。

2.3 语句并发控制与数据查询的实现

查询语句并发控制是确保多个用户或进程在同时访问数据库时，数据的完整性和一致性得到维护的关键机制。在工业级实时数据库中，数据通常具有高度的动态性和实时性，因此并发控制显得尤为重要。不仅可以防止数据冲突和不一致，还可以确保数据的准确性和可靠性，从而支持实时决策和高效的数据处理。经过 SQL 查询语句的标记后，事务之间的冲突发生概率较低，在提交时数据库只需检查是否发生了冲突，就可以准确定义协同化查询方案。

设 $\tilde{j}$ 表示实时数据库针对 SQL 查询语句所定义的数据并发处理向量， K_1 表示非冲突查询参数； K_2 表示冲突查询参数；

ϑ 表示工业数据的协同化传输系数； μ 表示尝试查询过程中的数据信息连接项，联立式（5），推导查询语句并发控制表达式为：

$$Q=\tilde{j}\cdot\frac{\left(\vartheta\frac{K_1}{K_2}\right)-1}{\mu D^2}\tag{6}$$

3 实验结果与分析

设置工业级实时数据库的目的是实现工业数据的高密度存储与快速采集，因此在查询目标信息时，应确保数据样本具有较低的时延水平和较高的压缩率。选择信创架构下的工业级实时数据库协同查询方法（第 1 组）、区块链与数据库混合架构下的查询方法（第 2 组）、基于谱聚类的查询方法（第 3 组）三组不同的技术手段进行验证，首先按照表 1 所示标准配置实验所需的测试机，然后记录实验过程中数据库机制中传输数据的压缩率与采样时延，最后分析所得结果并总结实验规律。

选择 MySQL 环境作为工业级实时数据库的运行背景，但由于 MySQL 数据库在运行过程中不会记录所存储数据的传输路径。DevOps 场景是 MySQL 数据库中最基本的数据存储场景，在该场景下激活 cpu_only 集群，在确保输入数据能够完整写入数据库的情况下开始实验。

表 1 测试机配置

序号	1	2	3
设备名称	服务器集群	单机服务器	测试机
型号	Huawei2288H V5	Huawei2288H V5	LNTL-PC009
硬件配置	处理器：Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10 GHz×10	处理器：Intel(R) Xeon(R) Platinum 8160 CPU @ 2.10 GHz×8	处理器：Intel(R) core(TM) i5-4590 CPU@3.3 GHz
	内存容量：16 GB	内存容量：16 GB	内存：8 GB
	硬盘容量：1 TB	硬盘容量：1 TB	硬盘：1 TB
软件配置	操作系统：openEuler-22.03-LTS	操作系统：openEuler-22.03-LTS	操作系统：Win10 64 位操作系统
	被测试软件：信创版工业实时数据库系统 RTDB V1.0	被测试软件：信创版工业实时数据库系统 RTDB V1.0	测试工具：Time Series Benchmark Suite

图 2 反映了不同查询方法下，工业数据压缩率的具体实验情况。分析图 2 可知，第 1 组方法的查询过程中，工业数据压缩率始终保持高水平状态，数据输入量对压缩情况的影响不明显；第 2 组方法的查询过程中，在数据输入量达到 3.0×10^{13} MB 之前，其压缩率的数值水平较高，与第 1 组数值的出入不大，但在数据输入量达到 4.0×10^{13} MB 后，其压缩率水平则明显下降，后续实验过程中，最大值仅为 63.14%，

远小于第1组；第3组方法的查询过程中，工业数据压缩率始终明显小于第1组，特别是在数据输入量达到 6.0×10^{13} MB后，其压缩率已经低至 66.71%。

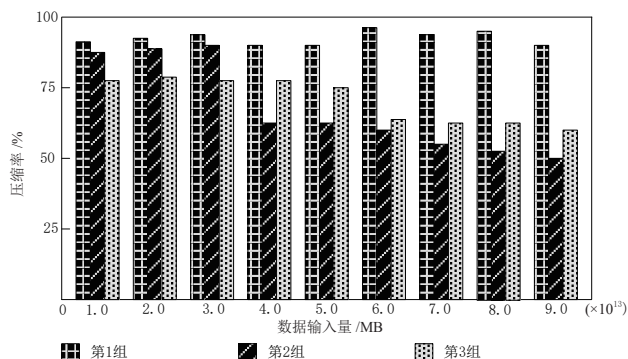


图2 工业数据压缩率

图3反映了压缩过程中被采样数据的详细时延情况。数据输入量达到 5.0×10^{13} MB之前，第1组采样延时始终不断增大，但达到最大值 47.38 ms后，则开始趋于平稳，整个实验过程中，其均值水平相对较低；数据输入量达到 4.0×10^{13} MB之前；第2组采样延时低于第1组，但从该节点开始，其数值水平明显增大，故整体均值水平依然高于第1组；数据输入量达到 3.0×10^{13} MB之前；第3组采样延时也低于第1组，但其后续实验曲线的走势与第2组较为相似，故整体均值水平也高于第1组。

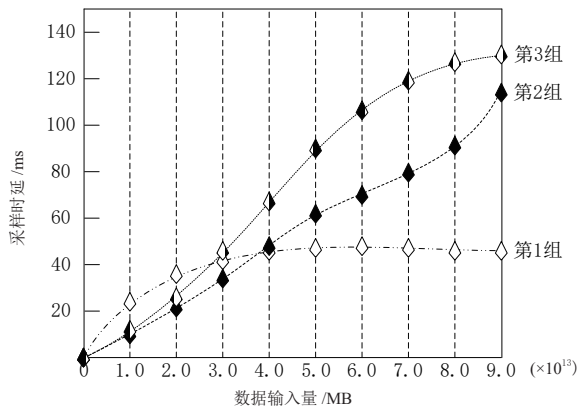


图3 数据采样时延

4 结语

在协同查询技术的研究过程中，深入分析了实时数据库的特点和需求，结合信创架构的特性和要求，设计并实现了高效的协同查询算法和机制。这些算法和机制在多个实际场景中得到了验证和应用，取得了显著的效果。此外，还对实时数据库的协作机制、查询语句等方面进行了深入研究，进一步提升了实时数据库的性能和效率。未来，相关组织机构可以继续深化信创架构下工业级实时数据库协同查询技术，一方面优化和完善现有的协同查询算法，提高其实用性与效

率；另一方面探索新的应用场景和领域，推动实时数据库在更多领域的应用和发展。

参考文献：

- [1] 吴昊亮,刘永宏,危建朋,等.异构体系结构下的混合负载数据库自适应并行技术[J].计算机应用,2024,44(S1):89-94.
- [2] 桂智明,孙忠祥,黄舟.区块链与数据库混合架构下的空间查询优化方法[J].北京大学学报(自然科学版),2023,59(2):261-270.
- [3] 刘鹏举,李好洋,王天一,等.基于谱聚类的在线数据库垂直分区多阶段生成方法[J].软件学报,2023,34(6):2804-2832.
- [4] 陈子勇,朱德剑.基于信创平台的新一代中波安播监控系统研究[J].电视技术,2024,48(11):210-213.
- [5] 李龙海,谢鹏,何列松,等.面向非关系型数据库的时空数据信息量平衡索引方法[J].测绘科学,2022,47(4):160-168.
- [6] 陆海,薛显斌,朱岩,等.可搜索加密及其驱动的SQL隐私数据库设计[J].工程科学学报,2024,46(11):2085-2098.
- [7] 于晓昕,何东,叶子铭,等.一种利用词典扩展数据库模式信息的Text2SQL方法[J].四川大学学报(自然科学版),2024,61(1):84-94.
- [8] 屠要峰,陈河堆,王涵毅,等.面向GoldenX软硬协同优化的异构加速列式存储引擎研究[J].计算机学报,2022,45(1):207-223.
- [9] 汪静,钱晓东.区块链环境中基于局部敏感哈希的协同过滤推荐研究[J].计算机工程与科学,2022,44(3):436-446.
- [10] 陈昆伦,李佳姪,李传文,等.面向高维数据的Skyline查询处理技术研究[J].小型微型计算机系统,2023,44(12):2819-2824.
- [11] 张啸剑,徐雅鑫,孟小峰.基于本地化差分隐私的空间数据近似k-近邻查询[J].计算机研究与发展,2022,59(7):1610-1624.
- [12] 王肇康,潘佳辉,周璐.支持亿级数据的高效密文范围查询完整性验证[J].模式识别与人工智能,2024,37(1):27-46.
- [13] 原旭,黄笠煌,陈志奎,等.基于状态树的链上数据高效可信查询索引模型及方法[J].重庆大学学报,2023,46(7):9-22.

【作者简介】

廖化锋(1981—),男,福建龙岩人,本科,研究方向:计算机应用。

(收稿日期:2025-08-27 修回日期:2026-03-05)