

面向金融场景的下一代数据库测试基准研究

荆一楠^{1,2*}, 张寒冰², 李智鑫², 王晓阳^{1,2}, 吴杰^{1,2}, 柴洪峰^{1,2}

(1. 复旦大学金融科技研究院, 上海 200433; 2. 复旦大学计算机科学技术学院, 上海 200433)

摘要: 银行是我国最主要的金融主体, 对数据库及数据服务解决方案有着更高的性能与安全要求。随着金融数据应用服务的快速发展, 银行数据库所涉及的数据类型、业务场景更加多样化, 用户很难在种类繁多的数据库产品和服务解决方案中做出最优选择。为此, 结合金融行业的数据应用发展需求, 本文通过采用文献调研和理论分析等方法, 全面分析了银行数据库的应用现状, 特别是近几年数据库国产化替代的情况与面临的挑战, 系统调研了国内外主要的数据库测试基准, 展望了构建面向金融场景的下一代数据库测试基准的必要性和重要性。研究发现, 由于金融场景中业务逻辑更复杂、数据模式更多样、安全性要求更高等多方面原因, 现有数据库测试基准在应对金融场景下的数据库测试时存在多处不足, 面临着诸多挑战, 基于此, 本文从工作负载、数据模式、度量指标以及技术架构等方面出发, 对面向金融场景的下一代数据库测试基准的构建提出针对性建议和要求。

关键词: 金融业; 银行; 金融数据; 数据库; 测试基准

中图分类号: TP3 **文献标识码:** A

Next-Generation Database Benchmark for Financial Scenarios

Jing Yinan^{1,2*}, Zhang Hanbing², Li Zhixin², Wang Xiaoyang^{1,2}, Wu Jie^{1,2}, Chai Hongfeng^{1,2}

(1. Institute of Financial Technology, Fudan University, Shanghai 200433, China; 2. School of Computer Science, Fudan University, Shanghai 200433, China)

Abstract: As the major financial entity in China, banks have high performance and security requirements for databases and data service solutions. With the progression of data application services in banking, the data types and business scenarios become more diverse, and it is difficult for users to make optimal choices among a wide diversity of database products and data service solutions. In combination with the data application demands of the financial industry, this study comprehensively analyzes the current status of applications of databases in banking, particularly the status and challenges of database localization in recent years, by using literature research and theoretical analysis. In addition, we systematically investigate the database benchmarks of China and other countries, and further prospect the necessity and importance of constructing next-generation database benchmarks for financial scenarios. We find that current database benchmarks have many deficiencies and face various challenges in dealing with the database testing in financial scenarios owing to the complex business logic, diverse data patterns, and high security requirements. Therefore, to build a next-generation database benchmark that can meet the requirements of financial scenarios, we propose several suggestions to address these challenges, which involve the aspects of workloads, data schemes, metrics, and technical architecture.

Keywords: financial industry; bank; financial data; database; benchmark

收稿日期: 2022-06-19; 修回日期: 2022-07-14

通讯作者: *荆一楠, 复旦大学计算机科学技术学院副教授, 研究方向为数据库; E-mail: jingyn@fudan.edu.cn

资助项目: 中国工程院咨询项目“金融数据安全治理智能化的战略研究”(2022-XY-12); 国家自然科学基金项目(92046024, 92146002)

本刊网址: www.engineering.org.cn/ch/journal/sscae

一、前言

改革开放以来，随着经济稳步增长和经济、金融体制改革的深入，金融业在我国国民经济中的地位愈加重要 [1]。截至 2021 年年底，我国金融业总资产达 381.95 万亿元，同比增长了 8.1%。其中银行业机构总资产为 344.76 万亿元，同比增长 7.8%，在各类金融机构（银行、证券、保险）中占据主要地位 [2]。随着移动互联网时代的到来，手机交易逐渐增多，银行系统内部每时每刻都在产生着大量的数据信息，如何安全有效地管理好这些数据成为银行对自身业务系统的重要诉求 [3]。中国人民银行发布的《金融科技发展规划（2022—2025）》中也明确指出，金融科技整体水平与核心竞争力需要实现跨越式提升，全面加强数据能力建设，建设绿色、高可用的数据中心 [4]。

数据库系统及数据服务解决方案作为承载数据存储和计算功能的基础软件，向下发挥硬件算力，向上支撑上层应用，是银行业务系统中至关重要的一环。20 世纪 90 年代，我国的银行机构通过使用 Oracle、DB2 等国外数据库产品实现了跨网点的存取款业务功能（通存通兑）。近年来，受银行业务需求和国际形势影响，以银行业为主的众多金融机构逐渐从国外商用数据库向国产分布式数据库过渡 [5]，以在提高业务系统性能的同时避免使用国外商用数据库产品所带来的潜在供应链风险。随着我国对金融数据安全的重视，金融业数据库国产化替代已成为大势所趋。

作为数据库性能测试的依据，数据库测试基准能够在给定的场景下公平、客观地对各类数据库产品和数据服务解决方案进行统一的测试，对用户选择具有重大的参考价值。然而，现有数据库测试基准在应对金融场景下的数据库测试时面临诸多挑战。一方面，以银行为代表的金融应用场景中的业务逻辑更复杂并且同时存在多种模式的数据，现有的测试基准缺乏在这种复杂的环境下对数据库产品及数据服务解决方案进行全面、准确测试的能力，其模拟的应用场景（大多以商业销售为主 [6,7]）与金融场景有着很大的区别。另一方面，金融业对数据安全的要求更高，金融数据的安全关系到国计民生，因此相比于现有测试基准，金融场景下的测试基准需要具备更强大、更全面的可靠性与安全性测

试能力，以辅助实现金融数据的安全治理，为金融数字化转型提供保障。此外，随着金融业数据库国产化替代步伐的不断迈进，分布式改造如火如荼，金融场景下的测试基准需要能够对数据库产品及数据服务解决方案的可移植性、兼容性和分布式架构适配性等方面进行评测，而这正是现有测试基准所欠缺的。因此，亟需构建一个符合金融业务发展需求的数据库测试基准以对数据库产品及数据服务解决方案作出统一的评价和度量，助力金融从业者做出更准确的选择，同时引导数据服务厂商的健康发展。

针对于此，本文首先对银行业数据库的应用发展现状进行阐述，深入剖析新时期银行数据库的新要求与发展趋势，介绍国产数据库在各大银行的替代情况与面临的挑战；其次调研分析国内外主要的数据库测试基准，结合金融行业的数据应用发展需求，分析构建面向金融场景的下一代数据库测试基准的必要性和重要性，以期为金融数据服务领域的从业者和研究者提供参考。

二、我国银行业数据库的应用现状

（一）银行业数据库应用的发展与变迁

在信息时代，数据库系统作为存储和管理数据的基础软件，在银行金融系统中具有重要的作用，直接关系到银行金融系统的稳定。如图 1 所示，从银行信息化 40 多年的发展历程来看，金融行业数据库建设先后经历了手工记账的单机时代、各支行和网点业务数据互通的互联互通时代、数据集中在总行的大集中时代、面向服务的架构（SOA）时代以及目前和未来将长期处于的分布式微服务时代。其中，从互联互通时代开始（20 世纪 90 年代），国外数据库产品已逐渐在我国银行内得到实际应用，打开了我国金融业数据库建设的大门。之后，随着我国数据库行业的不断发展，部分国产数据库于 2017 年左右被投入到包括国有大型银行、股份制银行、城市商业银行在内的众多金融机构中使用并且表现出优异的性能 [5]。目前，国产数据库厂商大致可分为 3 类企业：一是以阿里巴巴集团控股有限公司为代表的互联网企业；二是以北京人大金仓信息技术股份有限公司为代表的传统数据库公司；三是以华为技术有限公司、中兴通讯股份有限公司等

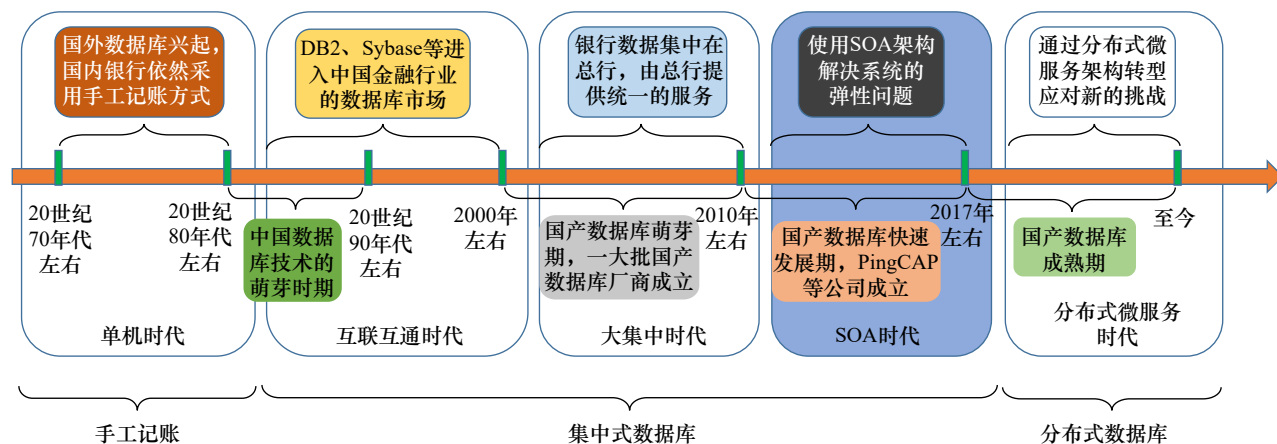


图1 银行业数据库应用的发展历程

为代表的综合型信息技术服务企业。截至2021年6月，我国数据库产品共有135款[8]；但是，在2020年我国高达200亿元的银行数据库软件市场中，Oracle、DB2等国外数据库产品的市场占有率仍超过了80% [9]，这也从侧面表明我国国产数据库拥有广阔的发展空间。

（二）新时期银行数据库的新要求

目前，金融行业在数据库应用方面呈现出了4个新需求[8]。一是随着移动互联网的快速发展，金融业务系统内产生的数据量在不断增长，这对数据库系统的数据存储和管理能力提出了更高的要求；二是随着普惠金融的落地，数据库系统需要具有更强的容灾能力，以保障业务连续性；三是随着电子支付的普及，数据库系统需要具有更强的性能，以应对高并发业务和高用户量所带来的系统压力；四是防止潜在供应链风险，技术层面存在数据库国产化需求，以避免金融数据安全受到威胁。

整体来看，为了适应银行的数字化转型升级，满足业务发展需要，新时代的银行数据库除了满足原子性、一致性、隔离性以及持久性（ACID）等数据库的基本要素之外，还应有区别于传统的特性才能跟上甚至引领时代。这些新要求包括：①可扩展性，数据库系统需要具备扩展数据的存储、访问、计算等方面的能力，特别是横向扩展的能力；②自主性，自主可控是信息安全的前提，是我国金融业务发展的需要，因此金融业特别是银行业务系统中的数据库系统应该是自主可控的；③海量性，随着电子支付的普及，数据库系统需要具备支撑海

量数据存储和计算的能力；④实时性，数据库系统需要具备在高并发环境下实时处理用户业务的能力；⑤高可用性，数据库系统要具备足够的容灾能力，以提供全天候不间断的服务，保障上层业务稳定运行；⑥安全性，数据库系统需要具备足够的安全性，为金融数据安全提供保障；⑦可迁移性，数据库系统需要能够将存储在Oracle、DB2等国外数据库中的业务数据完美地迁移出来，并保证数据的完整性与可用性。

（三）新时期银行数据库发展的新趋势

为了应对手机支付、贷款风险计算、银行卡盗刷研判等需要进行海量数据存储和计算的业务场景所带来的性能压力，保证我国金融信息系统的自主可控，防止其受到国际单边主义和贸易保护主义所带来的负面影响，金融行业特别是银行在数据库应用方面呈现出三大趋势。

1. 分布式数据库改造趋势

随着业务不断增长，金融业务系统中需要处理的数据量急剧增加，而现有的集中式数据库面临着一定的数据处理瓶颈，通过升级硬件来扩展能力成本高且有上限。因此，为了满足金融业务系统中日益增加的性能需求，将现存的集中式数据库改造为能够通过增加存储和计算节点来提升系统性能的分布式数据库，势在必行，也是未来的工作重点。中国人民银行发布的《金融科技（FinTech）发展规划（2019—2021年）》[10]中明确指出，要加强分布式数据库的研发应用，确保分布式数据库在金融领域的稳妥应用，并于2020年发布实施了《分布式数

数据库技术金融应用规范 技术架构》(JR/T 0203—2020)等3项金融行业标准 [11],对分布式数据库在金融业内的应用给出了具体的规范标准。

2. 数据库国产化替代趋势

随着国家政策引导的信息技术应用创新、网络强国、信息安全、大数据等国家战略的推进,我国开发利用数据的需求逐渐增大并且也愈加重视数据的安全问题。数据库作为承载数据存储和计算功能的基础软件,为保障金融数据安全,在金融行业大规模使用国产数据库产品是必然的。此外,受当前国际形势影响,在金融系统中使用国外商业数据库产品存在着诸多风险。同时,多项案例表明,国产数据库在金融业务系统内已呈现出较好的性能水平,这也增强了金融机构选择国产数据库产品的信心。

3. 非关系型数据库及多模数据库应用趋势

随着金融业务系统中存储的数据量急剧增长,面对贷款风险计算、大额交易判断、银行卡盗刷预警等需要进行数据分析的业务场景,使用传统的关系型数据库来分析系统中所存储的海量数据已难以满足这些场景中对处理速度的要求,金融业务系统中需要使用一些非关系型数据库及多模数据库来分类存储和管理金融数据,以提高数据分析效率,保障人民财产安全。例如,针对需要分析多层交易记录路径的贷款风险计算场景,将账户及交易记录等数据转换为图数据并使用图数据库(如Neo4j)来对这些数据进行存储和分析,能够比关系型数据库更加快速地获得任务结果。此外,在一部分业务系统中,需要使用能够同时支持对多种类型的数据进行集中存储、查询和处理的多模数据库,以满足系统对结构化、半结构化和非结构化数据的统一管理需求。因此,为了应对由海量数据所带来的性能压力,满足各类业务场景对数据分析速度的要求,未来会有大量的非关系型数据库及多模数据库应用在金融业务系统中。

(四) 银行数据库国产化替代现状与面临的挑战

受银行业务需求和国际形势影响,银行在广泛使用Oracle、DB2等国外数据库产品时,存在诸多潜在风险。因此,包括国有大型银行、股份制银行、城市商业银行在内的众多金融机构,都在逐步从国外商用数据库向国产数据库过渡 [5]。具体而言,各大有商业银行因尚不能在保证数据完整性

和业务不间断的前提下,将数据从正在使用的数据库迁移到新的数据库中,而没有在主要核心业务系统中进行数据库国产化替代,但是在一些需要进行海量数据处理的新兴业务系统以及一些不涉及核心业务的业务系统中,已经逐步引入性能表现良好的国产数据库,如北京平凯星辰科技发展有限公司的TiDB、天津南大通用数据技术股份有限公司的GBase、北京奥星贝斯科技有限公司的OceanBase等;股份制商业银行不仅仅是在数据分析和模型开发等领域采纳国产数据库,同时对于传统的事务型系统也有部分银行采用国产数据库,如中信银行在事务型业务(OLTP)领域采用中兴通讯股份有限公司开发的分布式数据库GoldenDB [12],交通银行基于OceanBase数据库自主研发了分布式数据库CBase并将全行的借记卡数据从DB2数据库下移至CBase数据库,使性能取得了显著提升 [13];城市商业银行或者规模更小的银行由于客户量较少,涉及到的数据量远远小于大型银行,因此除在外围业务和新兴业务上实现数据库国产化替代外,在部分核心系统上的替换也屡见不鲜,如南京银行使用北京奥星贝斯科技有限公司的Oceanbase数据库打造了完整的银行业务核心系统。整体来看,国产数据库具备自主可控、高可扩展、高性能、高可用等特性,可以很好地满足线上化、高频、多维度、高并发的场景需求,帮助金融机构解决技术瓶颈。分布式、云计算、混合部署等架构在国产数据库中的广泛应用还可以大幅提升成本控制和优化水平以达到“降本增效”的目的,有利于进一步实现金融数字化转型。

目前,在银行数据库国产化替代的过程中,也面临着一些挑战。一是架构转型挑战。金融行业中大量使用集中式数据库并具备了丰富的运维经验,如何快速有效地实现从集中式数据库到分布式数据库的转型,成为金融机构必须面对的挑战。此外,金融行业特别是银行对数据库有着极为严苛的要求,需要分布式数据库具备高度的稳定性和安全标准 [14]。二是数据迁移挑战。由于金融业长期使用Oracle、DB2等国外数据库产品,这些数据库中存储了大量的业务数据,如何高效率、低成本地将数据从正在使用的数据库迁移到新的数据库中,以保证数据完整性和业务系统不间断,成为金融机构当前面对的一大挑战。三是数据库产品选择挑战。现

有的数据库产品及数据服务解决方案种类繁多,彼此之间在金融业务场景中的优劣程度并没有标准的评价指标,如何选择适合自身业务需求并具有高性价比的数据库产品或数据服务解决方案,是金融机构当前面临的挑战。现有典型的数据库测试基准如 TPC-C [14,15]、TPC-H [5]、TPC-DS [6]等针对不同的场景模式提供了相应的测试工具和说明文档来对数据库产品进行评测,但是由于金融应用场景中的业务逻辑与现有测试基准所模拟的商务应用场景有所不同,因此在金融场景下直接使用现有的测试基准对数据库产品进行评测存在许多不足之处,亟需一个金融数据库测试基准来统一地评价数据库产品及数据服务解决方案在金融场景下的性能优劣。

三、数据库测试基准的研究现状

(一) 数据库测试基准的发展历程

半个多世纪以来,数据库技术经过不断发展和完善,已经从最初简单的层次模型、网络模型发展为关系型数据库、时间序列数据库、图数据库等更加适合现有应用场景的技术方法 [16],为数据库领域注入了新的活力。然而,面对各种各样的数据库产品,如何选择一个符合用户需求的解决方案就变得困难起来。例如,针对社交网络中的数据分析任务,使用以 Neo4j 为代表的图数据库能够比传统关系型数据库更好地表征数据关系;而针对海量数据分析需求,选用分布式数据库更能有效应对大数据和高并发的请求处理。正是因为应用场景复杂、产品选择多样,设计一定的测试基准就变得非常必要。在数据库领域,一个可靠的测试基准能够公平、客观地反映不同数据库系统在同一测试标准下的性能水平,为用户选择数据库产品及数据服务解决方案提供参考。

一个有效的数据库测试基准需包括数据模式、工作负载和度量指标 3 个部分,共同构成数据库测试基准。具体来看,数据模式在很大程度上决定了测试基准的应用场景,根据目标应用场景描述数据组成和结构。目前大多数测试基准都是通过设置与真实数据相似的数据规模、数据结构、数据分布和数据相关性等生成尽可能贴近真实数据的测试数据集 [17]。工作负载用于评测数据库系统在给定查询场景中的具体性能,是测试基准的核心。根据应用

领域、计算范式、密集类型和延迟要求的不同,不同测试基准中的工作负载需要具备不同的特性 [18]。度量指标通常用于描述测评对象的性能。由于业务场景、目标需求不同,因此测试基准中度量指标的选取需要尽可能公平、客观,常用的指标包括缓存命中率、吞吐率、查询执行时间、资源利用率以及系统可扩展性等。

图 2 描述了数据库测试基准的主要发展历程。数据库测试基准始于 1983 年,为解决当时大量数据库系统孰优孰劣的争执,研究者提出了威斯康星基准 [19],由 3 张表和 32 个结构化查询语言 (SQL) 语句组成,以总的运行时间为性能指标。虽然该基准较为简单,但为各关系型数据库管理系统厂家提供了很好的评测方式,也打开了测试基准的先河。威斯康星基准的发布也促进了事务处理委员会 (TPC) 的成立,目前应用较多的数据库测试基准主要由 TPC 发布,包括 TPC-C、TPC-H、TPC-DS 等。这些测试基准针对不同的场景模式提供了相应的测试工具和说明文档,对不同的数据库产品进行评测,具有较高的权威性,如近期关于 Databricks 和 Snowflake 之间的性能优劣争论,就是通过使用 TPC-DS 对其性能进行评测的 [20~22]。此外,随着数据量的不断增长以及新型应用场景的不断出现,产生了大量的新型测试基准,如面向大数据应用的测试基准 TPCx-BB [23]面向时间序列数据库的 TS-Benchmark [24]以及面向图数据库的 Graph 500 [25]等。综上,数据库技术的发展带来了新的测试基准需求,数据库测试基准促使数据库厂商进一步完善已有问题,不断提升系统的可用性,二者相互促进,共同发展。

(二) 面向关系型数据库的测试基准

按照数据库的作用进行划分,关系型数据库可以分为 OLTP 数据库、分析型 (OLAP) 数据库、混合负载 (HTAP) 数据库。针对这 3 种类型的数据库分别有相应的测试基准对其进行性能评估,具体如下。

1. 面向事务型数据库的测试基准

OLTP 数据库主要用于事务处理,具备 ACID 特性,能够在写入或更新数据的过程中保证事务是正确可靠的。其中,事务是指由一系列数据库操作组成的一个完整的逻辑过程。

TPC-C: TPC-C 是 TPC 组织提出的专门针对事

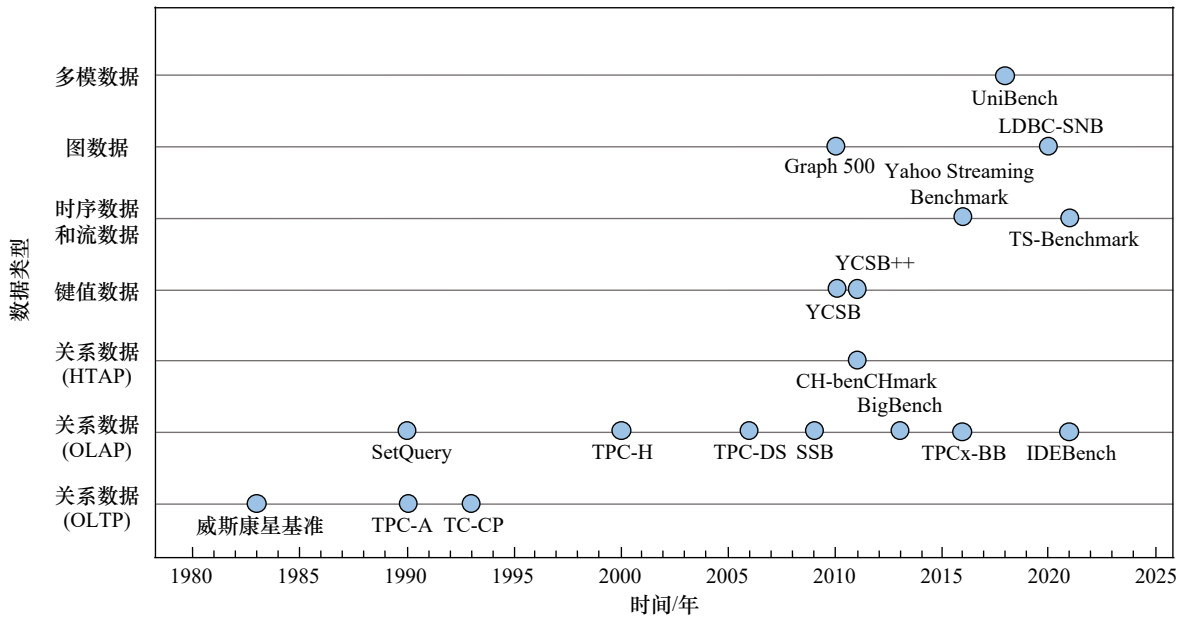


图2 数据库测试基准的发展历程

务处理系统的测试基准，其模拟了仓库订单管理场景 [15]。TPC-C 中包含 9 张数据表和 5 种交易事务，比较全面地涵盖了仓库订单管理业务；采用事务吞吐量作为度量指标来描述每分钟处理的事务数量，使用性价比来衡量系统的能耗。

2. 面向分析型数据库的测试基准

为提高分析效率和决策效果，一般通过建立数据仓库等方式来支持对大规模数据进行复杂的分析处理，因此 OLAP 数据库更关注复杂查询和聚集分析，侧重决策支持。

TPC-H: TPC-H 主要针对数据分析系统设计，模拟了一个贴近实际的商务采购应用场景 [6]。该测试基准基于第三范式 (3NF) 设计数据模式，包含 8 个数据表 (1 个事实表和 7 个维度表)，在每张数据表中的数据分布均为均匀分布。TPC-H 的工作负载中总共含有 22 个即席查询以及一些数据更新操作，这些查询操作从各方面模拟了商务采购应用中的典型业务逻辑。TPC-H 主要以每小时执行的查询数和性价比为度量指标对数据库系统进行评测。但是，由于 TPC-H 测试基准生成的数据为均匀分布，数据分布过于理想，研究表明 TPC-H 并不适合用于进行基数估计相关的评测 [26]；此外，随着数据和业务系统更加复杂，TPC-H 中的数据模式与工作负载也显得过于简单。

TPC-DS: TPC-DS 是 TPC 组织推出的用于替代

TPC-H 的决策支持系统测试基准 [7]。TPC-DS 采用雪花模型设计，包含 7 张事实表和 17 张维度表，工作负载中包含 99 个查询语句和 12 个数据维护操作。与 TPC-H 相比，TPC-DS 选用了更多的数据表，其中的数据分布也是模拟真实数据生成的，具有一定的偏斜性，与真实数据更为贴近；工作负载也更为复杂，其中包括即席查询、报表查询、迭代查询、数据挖掘等多种查询类型，从多个角度对真实的业务逻辑进行模拟，大部分的测试案例都有很高的输入/输出 (I/O) 负载和中央处理器 (CPU) 计算需求，能够较为全面地对数据库系统进行评测。由于 TPC-DS 适用范围广，更加符合现代商务应用场景，能够比较客观地反映系统的真实状况，因此是目前认可度较高的测试基准。

Star Schema Benchmark (SSB): SSB 是由麻州大学波士顿分校的研究人员提出的，主要用于评测决策支持系统的性能 [27]。SSB 将 TPC-H 中的 3NF 模型转化为星型模型，包含 1 张事实表和 4 张维度表；SSB 的工作负载中包含 13 条查询测试语句，其将 TPC-H 的工作负载中的一部分复杂查询用例改为结构更加固定的 OLAP 查询，并根据查询类型以及查询选择率将其分为 4 类查询模式，以提供更适用的功能覆盖和选择率覆盖测试。

TPCx-BB: TPCx-BB 的前身是 BigBench [28]，后于 2016 年被 TPC 接受并命名为 TPCx-BB [23]。

TPCx-BB 是一个端到端的大数据测试基准, 提供了 30 个模拟大数据处理、存储、分析场景的测试用例, 用以衡量大数据系统的软硬件性能。TPCx-BB 弥补了 TPC-DS 的不足, 它不仅包含了 TPC-DS 中结构化的零售商数据, 还添加了半结构化的网络日志数据和非结构化的商品评论数据, 更加贴近真实的应用场景。此外, TPCx-BB 从 1 个商业维度和 3 个技术维度对工作负载进行设计, 不仅能够度量大数据引擎的性能, 还符合大数据领域中的商务应用特征。

IDEBench: 随着数据量的快速增长, 在大型数据集集中进行交互式数据探索已经成为一个典型的分析应用和研究热点。研究工作者们提出了许多相关的查询处理技术以提高查询处理效率, 如基于采样原理的近似查询处理技术。然而, 当前流行的测试基准(如 TPC-DS、SSB)的工作负载和度量指标都是针对传统商业应用场景进行设计的, 并没有捕捉到交互式数据探索中的独特性质, 因此, 针对交互式数据探索的测试基准 IDEBench 被提出来 [29]。IDEBench 基于真实数据集来生成测试数据, 可将用户自定义的数据集缩放到任意大小并保持其原始特性; IDEBench 将交互式数据探索中的用户查询分为多种类型, 通过定义查询之间的链接关系来模拟用户探索行为之间的依赖关系, 生成的工作负载是有时间顺序的聚合查询序列。

3. 面向混合负载数据库的测试基准

HTAP 数据库是能同时提供 OLTP 和 OLAP 的混合关系型数据库 [16]。其中, 一部分 HTAP 数据库通过使用行列混存或者行列转化技术来支持 OLTP 和 OLAP 查询。

CH-benCHmark: CH-benCHmark 融合了 TPC-C 和 TPC-H 两种测试基准 [30]。CH-benCHmark 选用了 TPC-H 中的 3 个数据表对 TPC-C 中的数据模式进行扩充, 共包含 12 个数据表。此外, 针对改动后的数据模式, CH-benCHmark 以 TPC-C 的工作负载为基础, 结合重写后的 TPC-H 的工作负载, 构建了一个同时包含 OLAP、OLTP 查询的混合作业负载。同时, 为了评测数据库产品在混合作业负载下的性能, CH-benCHmark 使用事务吞吐量和每小时执行的查询数之间的比率作为度量指标。

(三) 面向非关系型数据库 (NoSQL) 的测试基准

除关系型数据库外, 随着近年来 NoSQL 数据

库的大量涌现, 新的数据库测试基准也在不断发展。本文选取了 4 类较为典型的 NoSQL 数据库测试基准进行分析介绍。

1. 面向键值数据与列簇式存储的测试基准

键值数据库与列簇式数据库是 NoSQL 数据库中较为常见的两类数据库, 主要用于加速系统中数据查找效率。具体而言, 键值数据库是对键值集合进行存储、检索和管理的一类 NoSQL 数据库, 其中键值对集合类似于映射或字典, 每个键在集合中具有唯一性, 键对应的值可以包含各种字段或数据; 列簇式数据库则是一类使用表、行和列进行数据存储的 NoSQL 数据库, 其针对快速检索数据列进行了优化, 通常用于联机分析处理。

YCSB: YCSB 的全称是 Yahoo! Cloud Serving Benchmark, 是一个用于比较 NoSQL 数据库性能的开源测试基准, 目标是促进新一代云数据服务系统的性能比较 [31]。目前, YCSB 已为广泛使用的 4 个系统 (Cassandra、HBase、PNUTS 和 MySQL) 定义了一组核心基准并报告了测试结果。此外, YCSB 是可扩展的, 支持定义新的工作负载。

YCSB++: 研究人员对 YCSB 进行扩展后提出了 YCSB++ [32], 其包括并行性测试、弱一致性测试、块上传测试等方面。YCSB++ 为多个测试客户机之间提供分布式同步, 并能够测量最终的一致性、批量加载以及批量写入的优化效果。此外, YCSB++ 还可以测量数据库中额外功能(如访问控制)的性能开销, 并收集每个集群节点上的资源监控信息。

2. 面向时间序列数据 (时序数据) 与流数据的测试基准

TS-Benchmark: 时序数据在物联网、数字金融、天气预报等领域都有广泛应用。随着 InfluxDB、IoTDB 等时序数据库的出现, 面向时序数据库的测试基准 TS-Benchmark [24] 应运而生。基于真实的物联网应用场景, TS-Benchmark 通过模拟风力涡轮机设备中不同传感器之间的物理模型来设计时序数据之间的相关性, 能够在数据加载、数据注入和数据获取这 3 类工作负载模式下系统测试时序数据库的性能。

Yahoo Streaming Benchmark: 随着大数据和云计算的不断发展, 基于流数据的应用产品数量激增, 这同时也产生了一大批新的流计算引擎(如 Storm、Flink、Spark Streaming)。为了便于用户对

不同的流计算引擎进行对比评测、选择合适的产品，研究人员提出了 Yahoo Streaming Benchmark 这一流计算引擎测试基准并得到业界广泛认可 [33]。Yahoo Streaming Benchmark 构建了一个完整的测试流程，使用 Kafka、Redis 进行数据获取和存储，模拟真实的广告分析流程并设计实现了一个贴近实际生产环境的流处理工作负载。

3. 面向图数据的测试基准

Graph 500: Graph 500 主要针对高性能计算中图算法部分进行评测，用以度量超级计算机的计算性能 [25]。Graph 500 侧重并发搜索、单源最短路径和最大独立集 3 个问题，涉及网络安全、医疗信息学、社交媒体等与图相关的业务领域，可充分评测超级计算机的数据处理能力。

LDBC-SNB: LDBC-SNB 是一个针对社交网络场景进行模拟的测试基准 [34]，根据社交网络中的特征设计生成数据模式并通过两种典型的工作负载对图数据库进行评测，能够较好地覆盖当前社交网中的一般业务，具备一定的普适性和代表性。其中，该测试基准有两种典型的工作负载，一种是交互式工作负载，包括给定节点下的邻居查询、数据插入等任务；另一种是商业智能工作负载，包括聚合函数与多图连接等复杂查询，相关查询会涉及到图中的大部分节点。

4. 面向多模数据库的测试基准

UniBench: 不同于围绕单一数据模型构建的数据库，多模数据库能够同时支持多种不同数据模型。为全面评估多模数据库产品，多模数据库测试基准 UniBench 应运而生 [35]。UniBench 是一个旨在针对多模数据库进行评测的测试基准，由一组模拟社交商务应用程序的混合数据模型组成，其中的数据生成器提供包括可扩展标记语言 (XML)、键值、文本、图等在内的多种数据格式；工作负载由一组多模型只读查询和读写事务组成，这些事务至少涉及两个数据模型，涵盖多模型数据管理的基本方面。

(四) 现有测试基准在金融场景中面临的新问题

数据库产品在金融业务场景中应用广泛，在银行及其他金融机构中的对公、对私业务线中广泛应用，是关系国计民生的重要领域，更是各大数据数据库厂商竞争的热点。目前，已有的数据库产品种类繁多、解决方案层出不穷，所提供的业务覆盖范围与服务性能都各有不同，致使银行用户难以做出最适合的选择。虽然日趋丰富的数据库测试基准已经显著推动了数据库技术的进步和研发（如 TPC-C 和 TPC-H），但是针对金融数据库产品及数据服务解决方案的评测却仍有较大空白。进一步，随着金融业数据应用服务不断发展，所涉及的数据类型、业务场景更加多样化，现有测试基准在金融数据服务中面临诸多问题，面向金融场景的数据库测试基准亟待发展和完善，以推动金融数据库技术的蓬勃发展。

1. 业务场景适配问题

从业务场景方面来看，由于金融应用场景中的业务逻辑（工作负载）与现有测试基准所模拟的商务应用场景有所不同，因此直接使用现有的测试基准在金融场景下对数据库产品及数据服务解决方案进行评测存在着许多的不足之处。例如，通过总结转账、存款、取款、账户查询、代发工资和资产盘点等 6 种金融场景中常见的业务逻辑 [36]，研究者们发现在这些业务中需要对数据进行多次访问操作以及安全验证，而现有的测试基准如 TPC-H、TPCxBB 等都与之相差甚远，不能有效地评价出数据库产品及数据服务解决方案在金融场景下的可用性。

2. 数据模式多样化问题

从测试基准所支持的数据模式来讲，为了满足互联网金融业务的需求，数据服务平台需要对多种不同模式的数据同时进行处理（如关系型数据、图数据、流数据等），而现有的测试基准通常仅针对单一数据模式，是一种单模的、孤立的、封闭的测试基准，不能很好地反映出金融业务的真实应用场景。如图 3 所示，针对贷款风险计算场景，银行需要查询贷款发放以后 2 个月期间，是否存在客户 N 人 ($N > 3$) 将资金转至同一账号（交易最大层数 3 层）。此类多层复杂路径查询问题使用图数据库进行处理将比使用关系型数据库更为高效。随着非关系型数据及其查询分析在金融业务中的需求不断上升，金融场景下的数据库产品及数据服务解决方案需要覆盖的数据模式也应当更加全面，而现有的测试基准在这一方面存在一定的不足。

3. 测试数据生成问题

现有的测试基准一般通过模拟真实数据中的数据结构、数据规模、数据分布、数据相关性等因素来生成测试数据，为数据库产品的性能测试提供数

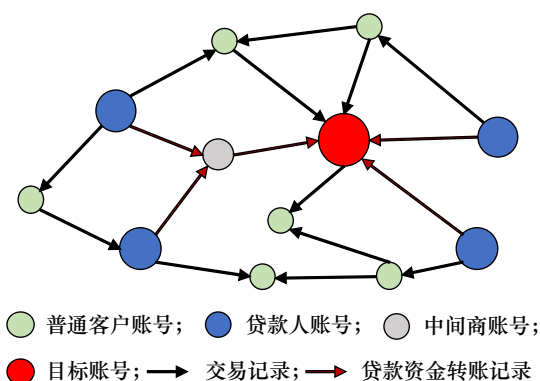


图3 贷款风险计算示意图

据支撑。然而，这种基于规则的数据生成方法难以发现真实数据中潜在的特征信息，不能生成与真实数据高度相似的测试数据；金融行业高度重视数据安全，生成的测试数据在模拟真实数据相关特征的同时，需要对一些涉及国家安全与用户隐私的数据特征进行遮掩，而通过人为地去总结数据特征来设计数据生成规则需要耗费大量的人力物力，并且不可避免的会出现错漏，难以在保证金融数据安全的同时生成符合金融数据特征的测试数据。

4. 兼容性和可移植性的测试问题

现有的测试基准主要是对各类数据库产品在一个给定应用场景下的性能指标进行测试，不能准确地对两个数据库产品之间的兼容性和可移植性进行度量，但这正是金融业目前迫切需要一个度量指标。随着金融业内数据库国产化替代的不断推进，金融从业者亟需一个度量指标来对数据库产品之间的兼容性和可移植性进行客观准确的评测，以帮助其从种类繁多的数据库产品及数据服务解决方案中选择出能够在保证业务连续性、数据完整性的同时具备高性价比的产品，最小化数据库替换对自身业务所带来的不利影响。

四、面向金融场景的下一代数据库测试基准研究初探

金融业对测试基准提出了新的需求。面向金融数据服务领域的特征，金融从业者需要一套能够全面覆盖多样化金融业务场景的测试基准。通过使用这一测试基准对各家厂商所提供的数据库产品及数据服务解决方案进行统一、标准的评测，来帮助金融从业者选择出最适合自己实际业务需求的数据库

解决方案。作为第一批对金融场景下数据库性能进行评测的测试基准，中国信息通信研究院发布了面向 OLTP 数据库的性能测试工具 DataBench-T [37]、面向 OLAP 数据库的性能测试工具 DataBench-A，根据金融行业内的特殊要求，制定并设计了一系列符合金融业务需求的数据模式、工作负载与度量指标。但这些测试基准仍然存在着一一定的问题，如 DataBench-T 和 DataBench-A 缺乏对混合负载的评测能力，支持的数据模式较为单一，目前仅支持关系型数据库测试。因此，新一代金融数据库测试基准需要从数据模式、工作负载、度量指标以及技术架构等方面，全方位适配金融业务场景，全面评估各类数据库产品及数据服务解决方案在金融场景中的性能指标。具体而言，未来需要在以下 7 个方面进一步对金融场景下的测试基准进行研究。

（一）支持混合工作负载的测试能力

金融业务领域混杂着大量的 OLTP 工作负载和 OLAP 工作负载，一方面是由于金融业务的特殊性，需要考虑使用各种资源隔离手段来隔离不同业务负载；另一方面，又要考虑同时支持这两种业务负载所带来的数据新鲜度等指标。因此，在设计新一代金融数据库测试基准中的工作负载时，需要同时考虑这两种负载的特征并选取适当的度量指标，以在模拟真实业务场景的同时对不同类型的工作负载进行统一评测。

（二）支持多模数据库的测试能力

随着金融业务的不断创新，除关系型数据以外，还存在着大量的非关系型数据。因此，在设计新一代金融数据库测试基准中的数据模式时，需要同时考虑关系型数据与非关系型数据，即需要对多模数据库具备一定的测试能力，并且考虑到金融业务未来的发展情况，在测试基准中留下充足的扩充空间，以便根据需要进行扩展。

（三）支持分布式架构适配性测试的能力

金融行业特别是银行业的数据库分布式改造正在如火如荼地进行。分布式系统区别于一般的单机系统和集中式系统，系统构造更复杂，影响因素更多。所以，在设计新一代金融数据库测试基准时，需要重点考虑如何评测数据库产品以及数据服务解

决方案对分布式架构的适配能力。

（四）支持金融仿真测试数据生成的能力

高保真的测试数据是模拟真实应用场景、实现更准确测试的保证。但由于合规约束严、隐私保护难、泄漏风险高等因素，致使利用真实业务数据进行测试是不可行的。目前在测试中大多采用模拟生成的测试数据进行仿真测试。传统的测试数据生成方法基本采用基于规则的方式（如 TPC-H），缺点在于规则需要人工定义、复杂规则定义困难、生成的数据容易丢失关键特征、数据仿真度不高，从而导致测试结果不够准确、无法真正模拟真实业务场景等。而现有研究表明，通过使用深度学习方法从真实数据中自动学习数据特征来生成仿真数据，能够较好地模拟真实数据特征，同时可以有选择性地在一些数据特征中添加噪声信息以保证最终生成的数据不会造成隐私泄露 [38]。因此，新一代金融数据库测试基准需要大力发展基于人工智能的金融仿真测试数据生成的能力，以生成具有安全性保障的高仿真金融数据。

（五）提供更全面的可靠性和安全性测试能力

金融业对数据库产品有更高的可靠性和安全性要求。金融数据需要提供严格的安全保障；数据存储系统应当具备高可靠性、提供备份和恢复的能力；基于业务特点，金融数据库需要具备更强的容灾能力。然而，现有的测试基准在这方面的测试能力有所欠缺，难以评价一个数据库产品是否符合金融业务中的可靠性与安全性要求。因此，新一代金融数据库测试基准需要有针对性地提高对数据库可靠性和安全性进行测试的能力。

（六）加强对数据库读写性能的测试能力

金融业对数据库产品有更高的性能需求，互联网金融数据服务平台需要具备更强的读写能力以应对高并发业务的压力。这些压力中的一部分来自于庞大的用户量，与电商类应用类似；另外一部分是因为金融数据服务需要在一次业务逻辑中进行多次数据一致性验证，以保障人民群众的财产安全，这与现有测试基准所模拟的应用场景有很大的不同。所以，新一代金融数据库测试基准需要有针对性地加强对读写能力方面的测试。

（七）增加对兼容性和可移植性的测试能力

在数据库国产化替代过程中，为了实现业务的平稳过渡，必须要考虑到新数据库对既有数据库的兼容性和可移植性，以保证数据库替代的过程中业务不间断及数据库替代后系统内数据的完整性，而这正是现有的测试基准所无法做到的。因此，兼容性和可移植性测试也是新一代金融数据库测试基准必须要考虑的方面。

五、结语

随着金融业务的快速扩展和我国对数据安全的更加重视，以银行为代表的金融业逐渐开始进行数据库国产化替代以应对更加复杂的业务需求，保证业务系统自主可控。然而，面对种类繁多的数据库产品，现有的数据库测试基准无法满足金融业对数据库产品及数据服务解决方案的测试需求，因此亟需一个能够适应金融业务场景的测试基准来助力金融从业者做出更准确的选择，同时引导数据服务厂商的健康发展。基于此，本文经过系统调研总结了银行业数据库国产化替代过程中所面临的三大挑战，分析了国内外主要的数据库测试基准在金融场景下所面临的四大问题，论述了构建面向金融场景的下一代数据库测试基准的意义与价值，并展望了该领域未来的研究方向。

在未来的研究工作中，需要进一步深入分析金融场景下数据库产品及数据服务解决方案的测试需求，从工作负载、数据模型定义、度量指标、仿真数据生成、测试平台架构等方面研究具体的实现方案，构建能满足金融场景测试需求的下一代数据库测试基准，以更全面、更准确地评测金融数据库产品及数据服务解决方案，为国家相关管理部门、金融从业者数据库选择提供参考，助力我国金融科技水平提升。

利益冲突声明

本文作者在此声明彼此之间不存在任何利益冲突或财务冲突。

Received date: June 19, 2022; **Revised date:** July 14, 2022

Corresponding author: Jing Yinan is an associate professor from the School of Computer Science, Fudan University. His major research field is database. Email: jingyn@fudan.edu.cn

Funding project: Chinese Academy of Engineering project “Strategic Research on Financial Data Security Governance Intelligentization” (2022-XY-12); National Natural Science Foundation of China project

(92046024, 92146002)

参考文献

- [1] 林毅夫, 付才辉, 任晓猛. 金融创新如何推动高质量发展: 新结构经济学的视角 [J]. 金融论坛, 2019, 24(11): 3-13.
Lin Y F, Fu C H, Ren X M. How financial innovation promotes high quality development: A perspective of new structural economics [J]. Finance Forum, 2019, 24(11): 3-13.
- [2] 中国人民银行. 2021 年末我国金融业机构总资产 381.95 万亿元 [EB/OL]. (2022-03-15)[2022-06-22]. <http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/4507972/index.html>.
The People's Bank of China. Total assets of financial institutions reached 381.95 trillion yuan at end-2021 [EB/OL]. (2022-03-15)[2022-06-22]. <http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/4507972/index.html>.
- [3] 甲子光年智库. 中国金融科技系列报告 [R/OL]. (2020-08-11)[2022-06-06]. https://www.jazzyyear.com/study_list.html?classifyName2=金融科技&classifyName3=全部&classifyName4=全部.
Jazzyyear. China fintechreport series [R/OL]. (2020-08-11)[2022-06-06]. https://www.jazzyyear.com/study_list.html?classifyName2=金融科技&classifyName3=全部&classifyName4=全部.
- [4] 中国人民银行. 中国人民银行印发《金融科技发展规划(2022—2025 年)》[EB/OL]. (2022-01-04)[2022-06-06]. <http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/4438627/index.html>.
The People's Bank of China. The People's Bank of China issued the *Fintech development plan for 2022—2025* [EB/OL]. (2022-01-04)[2022-06-06]. <http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/4438627/index.html>.
- [5] 胡利明. 分布式数据库在金融行业的应用和展望 [J]. 金融科技时代, 2020 (5): 25-33.
Hu L M. Application and prospect of distributed database in financial industry [J]. FinTech Time, 2020 (5): 25-33.
- [6] Poess M, Floyd C. New TPC benchmarks for decision support and web commerce [J]. ACM Special Interest Group on Management of Data Record, 2000, 29(4): 64-71.
- [7] Nambiar R O, Poess M. The making of TPC-DS [C]. Seoul: Proceedings of the 32nd International Conference on Very Large Data Bases, 2006.
- [8] 中国信息通信研究院. 数据库发展研究报告(2021 年) [R]. 北京: 中国信息通信研究院, 2021.
China Academic of Information and Communications Technology. Database development research report (2021) [R]. Beijing: China Academic of Information and Communications Technology, 2021.
- [9] ITpub 技术栈. 激荡三十年: 银行数据库的发展与变迁 [EB/OL]. (2021-04-02)[2022-06-06]. <https://z.itpub.net/article/detail/CE307F44933F633B8EB297FE3CF7379E>.
ITpub Technology Stack. Thirty years of turbulence: The development and changes of bank databases [EB/OL]. (2021-04-02)[2022-06-06]. <https://z.itpub.net/article/detail/CE307F44933F633B8EB297FE3CF7379E>.
- [10] 中国人民银行. 中国人民银行印发《金融科技(FinTech)发展规划(2019—2021 年)》[EB/OL]. (2019-08-22)[2022-06-06]. <http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/3878634/index.html>.
The People's Bank of China. The People's Bank of China issued *The fintech development plan for 2019—2022* [EB/OL]. (2019-08-22)[2022-06-06]. <http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/3878634/index.html>.
- [11] 全国金融标准化技术委员会. 《分布式数据库技术金融应用规范 技术架构》等 3 项金融行业标准正式发布 [EB/OL]. (2020-12-25)[2022-06-06]. <https://www.cfsc.org/jinbiao/2929436/2978097/index.html>.
China Financial Standardization Technical Committee. Three financial industry standards including *Financial application specification of distributed database technology technical architecture* were officially released [EB/OL]. (2020-12-25)[2022-06-06]. <https://www.cfsc.org/jinbiao/2929436/2978097/index.html>.
- [12] 王飞鹏. 追求卓越 砥砺前行——中信银行 GoldenDB 分布式数据库转型实践 [J]. 金融电子化, 2020 (2): 76-78.
Wang F P. Pursue excellence and forge ahead—China CITIC Bank GoldenDB distributed database transform-ation practice [J]. Financial Computerizing, 2020 (2): 76-78.
- [13] 李肇宁. 分布式数据库金融应用稳步有序推进 [J]. 金融电子化, 2020 (12): 34-35.
Li Z N. Distributed database financial applications advance steadily and orderly [J]. Financial Computerizing, 2020 (12): 34-35.
- [14] 戴功旺. 构建“新生态”, 探索金融行业分布式数据库发展之路 [J]. 中国金融电脑, 2021 (7): 85-86.
Dai G W. Build A “New ecology” and explore the development of distributed databases in the financial industry [J]. Financial Computer of China, 2021 (7): 85-86.
- [15] Leutenegger S T, Dias D M. A modeling study of the TPC-C benchmark [C]. Washington DC: Proceedings of the 1993 ACM International Conference on Management of Data, 1993.
- [16] 计算机学会数据库专业委会, 清华大学, 墨天轮社区. 数据库系统的分类和测评研究 [EB/OL]. (2021-12-22)[2022-06-06]. <https://www.modb.pro/doc/52857>.
CCF Technical Committee on Database, Tsinghua University, Modb.pro. Research on classification and evaluation of database system [EB/OL]. (2021-12-22)[2022-06-06]. <https://www.modb.pro/doc/52857>.
- [17] 金澈清, 钱卫宁, 周敏奇, 等. 数据管理系统评测基准: 从传统数据库到新兴大数据 [J]. 计算机学报, 2015, 38(1): 18-34.
Jin C Q, Qian W N, Zhou M Q, et al. Benchmarking data management systems: From traditional database to emergent big data [J]. Chinese Journal of Computers, 2015, 38(1): 18-34.
- [18] 闫义博, 朱文强, 杨全, 等. 大数据系统 Benchmark 测试综述 [J]. 网络新媒体技术, 2018, 7(3): 6-13.
Yan Y B, Zhu W Q, Yang T, et al. Overview on benchmark test of big data system [J]. Network New Media Technology, 2018, 7(3): 6-13.
- [19] Bitton D, DeWitt D J, Turbyfill C. Benchmarking database systems—A systematic approach [R]. Madison: University of Wisconsin-Madison, 1983.
- [20] Xin R, Mokhtar M. Databricks sets official data warehousing performance record [EB/OL]. (2021-11-02)[2022-06-06]. <https://databricks.com/blog/2021/11/02/databricks-sets-official-data-warehousing-performance-record.html>.
- [21] Dageville B, Cruanes T. Industry benchmarks and competing with integrity [EB/OL]. (2021-11-12)[2022-06-06]. <https://www.snowflake.com/blog/industry-bench-marks-and-competing-with-integrity/>.

- [22] Mokhtar M, Tavakoli-Shiraji A, Xin R, et al. Snowflake claims similar price/performance to data-bricks, but not so fast! [EB/OL]. (2021-11-15)[2022-06-06]. <https://databricks.com/blog/2021/11/15/snowflake-claims-similar-price-performance-to-databricks-but-not-so-fast.html>.
- [23] Cao P, Gowda B, Lakshmi S, et al. From BigBench to TPCx-BB: Standardization of a big data benchmark [C]. New Delhi: 8th TPC Technology Conference, 2016: 24–44.
- [24] Hao Y, Qin X, Chen Y, et al. TS-Benchmark: A benchmark for time series databases [C]. Chania: 37th IEEE International Conference on Data Engineering, 2021.
- [25] Murphy R C, Wheeler K B, Barrett B W, et al. Introducing the graph 500 [J]. Cray Users Group, 2010, 19: 45–74.
- [26] Dresler M, Boissier M, Rabl T, et al. Quantifying TPC-H choke points and their optimizations [J]. Proceedings of the VLDB Endowment, 2020, 13(8): 1206–1220.
- [27] O’Neil P E, O’Neil E J, Chen X, et al. The star schema benchmark and augmented fact table indexing [C]. Lyon: First TPC Technology Conference, 2009.
- [28] Ghazal A, Rabl T, Hu M, et al. Bigbench: Towards an industry standard benchmark for big data analytics [C]. New York: The 2013 ACM International Conference on Management of Data, 2013.
- [29] Eichmann P, Zraggen E, Binnig C, et al. IDEBench: A benchmark for interactive data exploration [C]. Portland: The 2020 ACM International Conference on Management of Data, 2020.
- [30] Funke F, Kemper A, Krompass S, et al. Metrics for measuring the performance of the mixed workload CH-benCHmark [C]. Seattle: Third TPC Technology Conference, 2011.
- [31] Cooper B F, Silberstein A, Tam E, et al. Benchmarking cloud serving systems with YCSB [C]. Indianapolis: The 1st ACM Symposium on Cloud Computing, 2010.
- [32] Patil S, Polte M, Ren K, et al. YCSB++: Benchmarking and performance debugging advanced features in scalable table stores [C]. Cascais: ACM Symposium on Cloud Computing in conjunction with SOSp 2011, 2011.
- [33] Chintapalli S, Dagit D, Evans B, et al. Benchmarking streaming computation engines: Storm, flink and spark streaming [C]. Chicago: 2016 IEEE International Parallel and Distributed Processing Symposium Workshops, 2016.
- [34] Angles R, Antal J B, Averbuch A, et al. The LDBC social network benchmark [EB/OL]. (2022-06-06)[2022-06-16]. <http://arxiv.org/abs/2001.02299>.
- [35] Zhang C, Lu J H, Xu P F, et al. UniBench: A benchmark for multi-model database management systems [C]. Riode Janeiro: 10th TPC Technology Conference, 2018.
- [36] 田稼丰, 姜春宇. 基于金融场景的数据库性能评估工具 [J]. 信息技术与政策, 2020, 46(4): 85–90.
Tian J F, Jiang C Y. Database performance evaluation tool under financial scenario [J]. Information and Communications Technology and Policy, 2020, 46(4): 85–90.
- [37] Jiang C, Tian J, Ma P. Databench-T: A transactional database benchmark for financial scenarios [C]. Shenyang: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications, 2021.
- [38] Liew S P, Takahashi T, Ueno M. PEARL: Data synthesis via private embeddings and adversarial reconstruction learning [EB/OL]. (2022-03-08)[2022-06-16]. <https://openreview.net/pdf?id=M6M8BEmd6dq>.