

# 金融行业数据库技术发展路径 探索与实践

中国工商银行金融科技研究院

2022年01月

# 版权声明

本文版权属于中国工商银行所有，并受法律保护。转载、摘编或利用其他方式使用本文观点或文字的，应注明来源。

违反上述声明者，将追究其相关法律责任。

# 编制说明

本报告由中国工商银行金融科技研究院编制。

# 前 言

“十四五”规划提出“加快数字化发展”的总体布局，数据库作为金融信息系统的关键基础设施，对于突破金融效率瓶颈、释放金融创新空间、提升金融服务水平、支撑金融行业数字化转型和高质量发展具有重要而深远的意义。

为进一步推动数据库技术在金融行业的应用与发展，中国工商银行金融科技研究院组织调研金融行业数据库技术应用现状、分析金融行业数据库架构转型中的重点和难点应用场景，编写了本报告，希望可以为金融同业及相关领域从业者提供参考。

本文主要聚焦于银行业为主的金融领域对关系型数据库的应用。第一章回顾金融行业发展历程中使用的数据库技术，并介绍当前数据库技术发展几种主流技术路线及在金融行业的应用。第二章着重分析当前金融行业数据库架构转型中的重点和难点应用场景，包括主机金融核心系统分布式迁移，以及开放平台传统应用迁移等场景。第三章展望数据库技术在金融行业的应用与推广。

# 目 录

一、 金融行业数据库技术的发展.....	1
(一) 数据库技术为金融行业发展提供技术支持.....	1
1、 金融行业数据库技术发展历程.....	1
2、 金融行业数据库技术应用现状.....	2
(二) 数据库技术路线及在金融行业的应用.....	3
1、 集中式数据库.....	3
2、 中间件架构数据库.....	5
3、 NewSQL 数据库.....	6
4、 云原生数据库.....	9
二、 金融行业数据库架构转型的探索.....	10
(一) 主机金融核心系统转型场景.....	11
1、 业务需求.....	11
2、 解决方案.....	12
3、 典型案例：工商银行核心应用分布式转型.....	14
(二) 开放平台传统应用转型场景.....	16
1、 业务需求.....	16
2、 解决方案.....	17
3、 典型案例：工商银行开放平台传统应用转型.....	18
三、 金融行业数据库应用未来展望.....	19
(一) 金融行业数据库技术应用的思考.....	19
1、 多种数据库产品并行发展.....	19
2、 各类技术路线相互融合.....	22
(二) 金融行业数据库技术应用的建议.....	23
1、 建设标准体系，引领金融行业数据库技术发展.....	23
2、 促进产用对接，推动产业生态发展成熟.....	24
3、 加强人才培养，激发金融行业创新活力.....	24
参考文献.....	26

# 一、金融行业数据库技术的发展

## （一）数据库技术为金融行业发展提供技术支撑

### 1、金融行业数据库技术发展历程

数据库作为金融信息系统的核心基础设施，历经数十年发展，为金融行业经营战略转型升级提供了有力的技术支撑。

改革开放后，我国金融行业面对信息技术革命滚滚大潮，积极学习和吸收世界金融行业技术革新成果，告别了手工记账时代，开启电子化、信息化建设进程，带来了金融行业工作方式和业务处理的巨大变革。随着金融体制改革不断深化，开始推动金融系统纵向统筹管控，金融行业走向数据大集中的发展道路。工行、农行、中行、建行等国有大行于世纪之交率先开启集中式金融信息系统技术体系建设，宣告中国大型金融机构步入集约化经营时代。1999年，工商银行启动“991”数据大集中工程，将全行30000多家下属机构经营数据全部集中到南北两个数据中心，实现数据共享大集中处理模式。这一时期，商用集中式数据库以其高效的数据存取效率、优异的系统稳定性，很好地契合了金融行业对数据存储管理的需求，为金融行业实现信息集中统计处理、财务集中改革、风险集中控制、业务集中管理等奠定了坚实的技术基础。

随着互联网技术兴起，金融行业出现电子银行、电子支付等新业态，金融行业开放平台技术体系涌现出多种不同的数据库产品，有效

实现金融服务的降本增效和创新发展。互联网金融的异军突起，促进了基于开源数据库的金融级解决方案发展与成熟，在互联网领域具备较多应用案例的开源数据库也逐步引入金融行业。金融行业吸收业界数据库技术应用的成功经验，并结合金融交易场景推陈出新，不断扩展现阶段平台数据库在金融行业的应用范围，有力支撑金融行业向信息化转型。

进入数字化时代，在有效处理超高峰值、海量数据的需求驱动下，分布式数据库蓬勃发展，涌现出众多与各类新的软硬件平台相适配的分布式数据库产品。目前，金融行业数字化转型不断深入，向数据库如何更好支撑金融业务创新提出了新的课题。各金融机构纷纷开展数据库架构转型，探索各种新型分布式数据库产品在金融应用场景试点和落地。

## **2、金融行业数据库技术应用现状**

商业集中式数据库以其较强的功能黏性、优秀的系统稳定性、良好的软硬适配能力，目前在金融行业的存量应用仍占据较大的份额。而 MySQL、PostgreSQL 等开源数据库近些年逐渐从金融外围系统向金融核心业务延伸。

为应对数字化转型带来的高并发、海量数据、超高峰值等挑战，近年来各金融机构纷纷开展分布式数据库的探索，多技术路线并行推进试点，分布式数据库在金融行业应用规模预计将有明显提升。

## （二）数据库技术路线及在金融行业的应用

随着互联网经济的发展，金融业务线上化和服务场景生态化，金融应用系统面临处理海量数据、超高峰值等一系列挑战。为更好地支撑金融业务数字化转型，金融行业积极开展 IT 架构转型，数据库技术领域的最新发展为金融行业各类创新业务的快速增长提供了更好的技术支撑。

按照不同维度，数据库可以分为不同的类型。

按照技术架构维度，数据库可以划分为集中式数据库和分布式数据库，业界一般将分布式数据库进一步细分为基于中间件架构数据库、NewSQL 数据库、云原生数据库三条技术路线。

按照应用场景维度，数据库可以划分为 OLTP 数据库、OLAP 数据库、HTAP (Hybrid Transaction and Analytical Process, 混合事务和分析处理) 数据库等。在应用场景的需求驱动下，上述集中式数据库、中间件架构数据库、NewSQL 数据库和云原生数据库等 4 种技术路线都有产品发展出了一定程度的 HTAP 能力。

本文从技术架构维度对数据库技术发展路径及金融行业应用进行说明。

### 1、集中式数据库

#### （1）技术特征

集中式数据库典型技术特征：解析器、优化器、存储引擎等数据



库核心组件都集中部署在同一台服务器上。

目前集中式数据库技术发展主要有几个方向：

一是提升数据库系统处理能力。如 MySQL 8.0 宣称处理性能可达 MySQL 5.7 版本的两倍，但在企业实用环境中开展性能对比测试结果提升并不显著。实际上，金融行业数据库系统性能问题大多来源于存储或网络的资源瓶颈，正常情况下数据库内部处理性能的提升对于系统整体性能影响有限。

二是增强 SQL 查询功能，简化应用开发。如 DB2 11.5 引入机器学习优化器加快 SQL 查询速度；MySQL 8.0 引入更多窗口函数加强对统计分析类查询的支持。

三是优化系统管理，提高运维效率。如 Oracle 21c 提供更细粒度的多租户资源管理和更完善的数据库自治能力。

集中式数据库优化器、存储引擎等数据库核心组件的发展优化，对于分布式数据库中各个节点的处理能力也具有积极意义，有助于控制分布式数据库集群规模，降低数据库使用成本。

## **(2) 金融行业应用**

目前金融行业采用的数据库系统仍以集中式数据库为主。经过数十年的技术积累，集中式数据库的发展已经非常成熟，特别是金融行业广泛应用的商用集中式数据库，在产品易用性、稳定性、成熟度等方面都达到了较高水准，值得新型数据库产品对标学习。集中式数据库的主要不足在于系统扩展性方面受到限制，在硬件条件未发生根本

性突破的情况下已难以继续支撑金融创新业务的快速增长。

## 2、中间件架构数据库

### (1) 技术特征

为解决传统集中式数据库在扩展性上的不足，突破系统纵向扩展面临的硬件能力制约，数据库引入分库分表中间件实现数据容量的横向扩展，再经过分布式事务管理、高可用切换、自动化运维等企业级能力建设和产品化，形成了基于中间件架构的数据库产品。

中间件架构数据库的典型设计采用“传统数据库作为数据节点(DN, DataNode) + 分片路由中间件作为协调节点(CN, CoordinatorNode) + 全局事务管理器(GTM, GlobalTransactionManager)”架构，其技术特征是基于传统集中式数据库，通过中间件实现数据分片、分布式事务等分布式数据库能力，大多采用 MySQL/PostgreSQL 等开源数据库作为数据存储节点，继承了传统数据库的技术积累和成熟生态，数据库存储引擎内核稳定性相对较高，周边生态工具配套完善，可较好满足大型商业机构对系统稳定性的需求。

### (2) 金融行业应用

对比数据库领域各种前沿技术热点，中间件架构数据库较少从架构设计角度进行宣传。事实上，率先在金融应用场景落地的分布式数据库产品大部分属于这条技术路线。例如，某头部科技企业同时推出基于中间件架构数据库产品和云原生数据库产品，其金融核心系统最

初也是使用前者。基于中间件架构的数据库在金融行业的应用具备以下优势：

一是发展较早，在金融行业有一定技术积累。早期金融行业引入分布式数据库的选择较少，其它分布式数据库技术路线推出商业化产品的时间相对较晚。

二是采用经过广泛应用、长久检验的传统集中式数据库作为数据存储节点，数据库存储引擎内核稳定性相对较高。

三是技术栈迁移跨度较小，便于开发和运维人员理解掌握。SQL语法与传统数据库兼容性较好，数据备份恢复等运维管理工具可以基于传统数据库现有工具进行二次开发。

中间件架构数据库为金融信息系统实现性能容量横向扩展提供了一种成熟可落地的解决方案，在实际应用中仍存在诸多痛点，主要体现在于如何在分布式环境下整合数据库服务能力，比如银行批量作业的分布式执行、大规模分布式集群的集中统一管理。需对标金融行业核心系统广泛应用的商用集中式数据库服务水平，持续提升分布式环境下的企业级数据服务能力。

### **3、NewSQL 数据库**

#### **(1) 技术特征**

为应对管理超大规模 MySQL 集群带来的系统复杂性，Google 公司结合新闻广告业务场景，设计了牺牲 SQL、事务特性换取极致扩展性的 NoSQL 系统，并发表著名的 GFS、Map-Reduce、BigTable 三篇论

文。开源社区随后推出 HBase、MongoDB 等 NoSQL 数据库，将 NoSQL 的应用推向热潮。经过一段时间的探索，业界重新认识到基于坚实数学基础的关系模型所具备的价值，SQL 和事务在大量应用场景都不可或缺。2012 年 Google 发表 Spanner 论文，在 NoSQL 基础上引入关系模型和线性一致性的分布式事务，发展成为 NewSQL 数据库（也被称为“原生分布式数据库”，强调其分布式处理能力优势）。Google 官方已确认将 Spanner 作为 BigTable 的下一代数据存储。开源社区也推出了 TiDB、CockroachDB 等 NewSQL 架构分布式数据库产品。

NewSQL 数据库的典型架构设计基于 KV 存储引擎，引入 SQL 关系模型和分布式事务 ACID 特性，主要技术特征包括：

一是基于 Paxos、Raft 等共识算法实现分布式架构下数据一致性达到线性强一致。

二是追求极致扩展性。具备媲美 NoSQL 平台的海量数据存储容量 (100PB 级别)，可实现对应用几乎完全透明的弹性扩缩容。

三是实现全局统一的开发视图。提供与分片键无关的全局唯一索引，可以将 NewSQL 数据库看成一个逻辑上的超大单体数据库来使用，简化应用研发模型。

在交易性能方面，NewSQL 数据库早期设计性能偏低，Google Spanner 基于新闻广告业务场景，设计上侧重于读和顺序写操作的优化，性能上侧重于吞吐量 (TPS 指标) 而非交易响应时间 (Latency 指标)。随着 NewSQL 数据库产品面向通用场景和交易性能持续优化提升，目前已在事务处理性能委员会 (TPC, Transaction Processing Perfor

mance Council)认证的性能测评上取得佳绩。

## (2) 金融行业应用

近年来 NewSQL 数据库在金融行业已有应用案例。因 NewSQL 数据库在海量数据存储、统计分析类查询能力等方面具备一定优势，部分金融机构采用 NewSQL 数据库作为金融核心批量场景的补充解决方案。

根据实践经验，金融行业使用 NewSQL 数据库需关注如下问题：

一是采用全新技术栈实现带来的数据库功能差异。例如在数据库事务隔离级别方面，隔离级别影响并发事务的数据可见性，与应用并发处理逻辑紧密相关。SQL 标准通过禁止出现的现象来定义隔离级别，给出的是一个范围定义而非精确定义。不同数据库的隔离级别实现并不完全一致，比如 Oracle 与 MySQL 的 RC 隔离级别就存在细微差异，在数据库迁移场景下可能导致应用处理逻辑出现预料之外的结果，数据库迁移必须通过完备的测试，保障应用的逻辑正确性。NewSQL 数据库采用全新技术栈实现，与传统数据库的事务模型存在，需更加重视应用逻辑的充分测试验证。

二是高性能场景下慎用对应用透明的数据分布。NewSQL 数据库有能力实现数据分布和系统扩缩容对应用完全透明，从应用视角可视作为一个巨大的单库，从而简化应用程序设计。但实现这种扩展性需在性能方面做出牺牲。金融应用实践中发现，使用分布式数据库需要通过高内聚、低耦合的数据规划尽可能减少分布式事务，降低跨节点交

互对交易性能的影响，在要求高性能的场景下，应用不能将数据分布完全托管给数据库。对此，NewSQL 数据库引入复制表、表分组(table group)等特性提升本地化数据访问占比，从而提升交易性能。

## 4、云原生数据库

### (1) 技术特征

随着云虚拟化技术的蓬勃发展，数据库也迁移到云上，提供关系型数据库服务(RDS, Relational Database Service)。依托云底座，可以实现数据库环境快速供给和多租户等云化服务能力。早期的数据库云化服务并没有改变数据库的架构设计，数据库高可用、扩展性等能力没有本质提升，从全技术栈角度，存在数据库层和云存储层功能设计冗余、资源利用率低等不足。

在 SIGMOD 2017 会议上发表的 Aurora 论文正式提出“Cloud-Native (云原生)”概念。“原生”与“迁移”相对，表示与之前的数据库迁移上云不同。云原生是一种架构设计思想，而不是数据库部署方式，与数据库云服务、云数据库是不同概念。

云原生数据库指基于云与数据库引擎的深度融合，实现数据库计算和存储资源的池化管理，从而提升数据库高可用、弹性扩展等能力。典型技术特征有：

一是数据库层和云存储层的一体化设计，采用 Shared Storage 的计算存储分离架构；

二是数据库层为无状态（或少状态）节点，基于“日志即数据库”

理念，向云存储只刷日志不刷数据，降低存储网络负载，提升性能；

三是云存储层不仅提供存储 IO 能力，更是与数据库存储引擎深度融合，通过数据分片的细粒度组织实现数据库的高可用能力提升和弹性扩展能力。

云原生数据库借鉴了包括 NewSQL 在内的各种数据库设计思想，通过数据库与云的深度融合，对比传统集中式数据库，能够提供更好的性能、更高的可用性和更具弹性的扩展能力。目前是数据库行业重点投入方向，数据库各技术发展路线都有与云融合的趋势。

## **(2) 金融行业应用**

云原生数据库在金融行业落地的主要障碍是与云基础设施高度耦合。金融行业正在经历专用技术栈向开放技术栈的转型阵痛，对于数据库与单一厂商云基础设施强绑定持谨慎态度。要解决产业界技术发展与金融行业技术选型的这一矛盾，需进一步提升云基础设施和云原生数据库的标准化水平，实现不同云原生数据库产品与云平台之间兼容。

为适应金融行业数据库应用需求，部分云原生数据库发展出不依赖云的纯软产品形态。但云原生数据库脱离云的支撑，难以完美发挥架构设计优势，需要补齐周边配套支撑能力。

## **二、金融行业数据库架构转型的探索**

结合金融科技发展趋势与金融业务创新需求，各金融机构积极加

大数据数据库转型试点力度，以实现灵活弹性的系统扩展能力，支撑金融业务规模持续增长。目前金融行业数据库转型正处于多技术路线并行试点阶段，典型场景包括基于主机的金融核心系统从传统集中式架构向分布式架构转型、与特定数据库专有特性高度耦合的开放平台传统应用转型，这也是金融行业数据库架构转型的重点和难点。针对以上典型场景，金融机构与科技企业开展联合创新，共同探索解决方案，树立典型示范案例，为分布式数据库在金融行业的全面推广提供参考。

## （一）主机金融核心系统转型场景

### 1、业务需求

金融核心系统承载金融行业最核心的业务，如商业银行管理客户信息、处理客户账户及核心总账、提供基础存款/贷款/支付服务的系统。随着金融行业数字化转型不断深化，传统金融业务逐渐走向线上化，金融服务场景生态化，金融创新业务快速扩张，传统基于主机的金融核心系统面临如下痛点问题：

一是性能容量压力。随着数字化时代的来临，金融服务更加广泛和便利，移动终端用户和交易频次大幅增加，越来越多第三方系统接入，金融核心系统需要提供更高的并发处理能力和海量数据存储能力。传统集中式架构面临摩尔定律失效的硬件限制，不具备横向扩展能力，难以支撑金融业务规模持续增长。

二是业务连续性要求。随着移动互联网的发展，新生代客户金融



习惯发生改变，金融服务呈现多渠道、全天候等特点，对金融核心系统的业务连续性提出更高要求。

三是快速交付能力。为应对激烈的市场竞争，金融行业加快响应客户需求，提升金融服务水平和产品竞争力。传统开发模式下，应用内部模块、应用与应用之间的耦合度较高，导致软件开发和交付的时间周期较长。

四是成本控制。为践行“服务实体、让利实体”的金融使命，金融行业大力推进企业经营降本增效，在支撑金融创新业务爆发式增长的同时，系统运营成本必须得到有效控制。在金融产品秒杀抢购等业务场景，交易峰值出现跨数量级的突增，且持续时间较短。传统集中式架构需要大量资源超配以满足业务高峰请求，成本很难得到有效控制。

在这样的背景下，金融行业 IT 架构转型的整体诉求是通过优化应用架构、数据架构、技术架构，建立起灵活开放、高效协同、安全稳定的 IT 架构体系，强化对业务快速创新发展的技术支撑。金融行业数据库架构转型需要与上述业务需求和转型战略相适配。

## **2、解决方案**

为支撑金融业务规模持续增长，主机金融核心系统数据库架构转型的整体策略方向是集中式架构向分布式架构转型。

对于高性能、大数据量的金融核心系统，通过采用分布式数据库技术，可以引入一种系统级的分布式解决方案，避免应用层大规模分

布式改造的成本和风险，实现系统性能容量的弹性扩展。通过金融级高可用、数据分片路由、数据库治理、云化服务管理等企业级数据服务能力建设，实现更好的业务连续性、更快的交付能力、更优的成本控制，有效满足海量数据、超高峰值等业务场景需求。

金融核心系统分布式迁移还需考虑数据库对批量业务处理的支持。银行核心应用不仅提供联机交易服务，同时也需支撑日终对账、报表统计等批量服务。批量服务的特点是数据统计量大、逻辑相对简单、日间联机小批量时效性要求高，同时需尽可能避免大型批量作业影响联机实时交易，这是典型的 HTAP 应用场景。金融核心系统迁移的批量业务处理可选择如下几种解决方案：

一是可采用具备较强 HTAP 能力的分布式数据库产品进行承接，或通过数据复制引入具备 HTAP 能力的辅库进行批量处理：

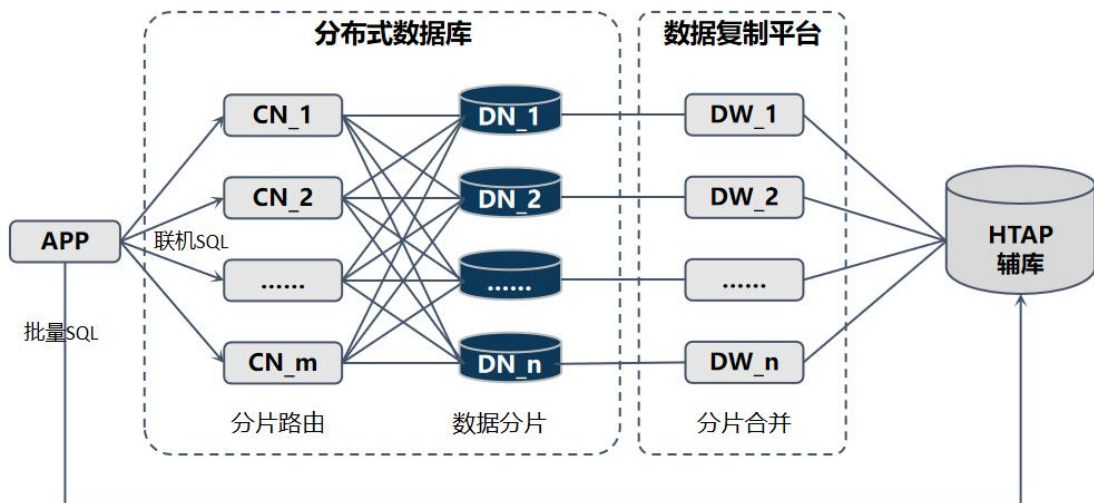


图 1 批量业务处理解决方案

这种解决方案采用统一的 SQL 接口和统一的数据主备复制协议，无需数据 ETL (Extract-Transform-Load) 过程，降低应用研发技术栈

的复杂度。

二是可通过大数据平台实现对数据的批量处理。

### 3、典型案例：工商银行核心应用分布式迁移

工商银行于 2015 年开始推动 IT 架构转型，考虑到当时数据库产品商业配套尚不完善，而 MySQL 数据库作为流行度第一的开源数据库，在内核稳定性、产品易用性、生态工具完善度等方面具备显著优势，在互联网高并发可扩展的分布式场景下已有较多成功案例，因此在 2016 年开放平台 OLTP 数据库转型规划中，工商银行明确“基于开源 MySQL 成熟生态建设分布式数据库解决方案，解决好横向扩展和便捷运维等问题，有效降低应用转型成本。”

工商银行基于开源 MySQL 数据库开展自主研发，针对大规模分布式数据库集群带来的管理挑战，持续建设金融级高可用、数据路由、云化服务管理、数据库治理等企业级数据服务能力，对标金融核心系统数据服务水平，实现分布式数据库集群的集中统一自动化管理，打造了与“分布式 + 云计算”体系相适配的金融级分布式数据库解决方案。经过 4 年的发展，于 2020 年作为工商银行分布式数据库产品正式发布。该产品架构设计如下图所示：

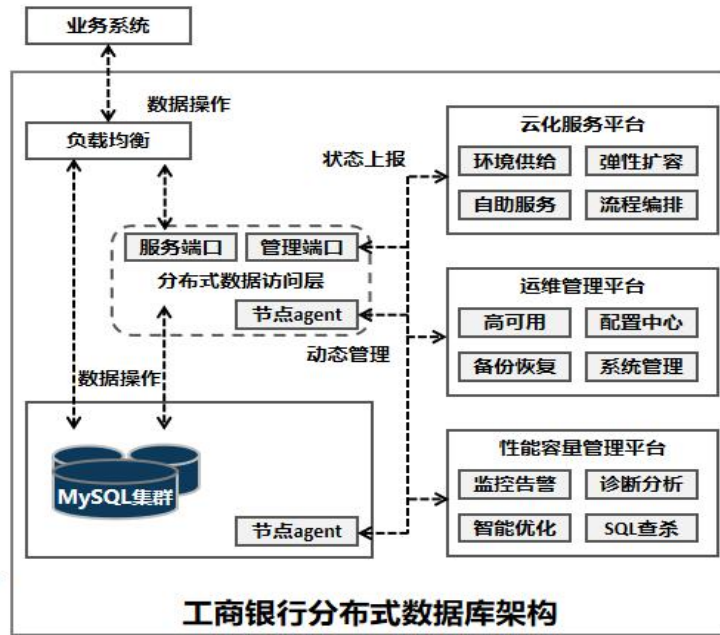


图 2 工商银行分布式数据库架构图

工商银行将自研分布式数据库创新应用于借记卡系统、客户信息系统等核心银行系统，构建了支撑开放平台核心银行系统的数据存储服务体系，实现大型银行 IT 架构转型关键性突破。其中：

- 借记卡系统管理 10 亿个人结算账户及介质，日平均提供 2.5 亿笔结算服务、1 亿笔查询管理类服务，TPS 峰值达 2.5 万笔以上，平均响应时间小于 50 毫秒，全链路交易耗时小于 250 毫秒；
- 客户信息系统为全行业务系统提供客户信息维护与查询服务，管理 6 亿个人客户和 1000 多万对公客户，数据量超过 160 亿条记录，日平均提供 2 亿笔客户信息维护与查询，最高并发 7500TPS，平均交易耗时小于 30 毫秒。

## （二）开放平台传统应用转型场景

### 1、业务需求

目前，金融行业开放平台也存在部分大型传统集中式应用面临扩展性不足的问题，因与数据库特性高度耦合，成为数据库架构转型中的难点。开放平台传统应用转型存在如下几点需求：

一是需满足业务系统性能要求。开放平台传统应用与数据库高度耦合的优势就在于减少了系统组件之间的交互开销，此类应用的业务场景往往对系统性能有较高要求，需在满足业务性能指标要求的前提下开展数据库架构转型。

二是需保障迁移过程中的系统稳定性。在金融业务创新、应用逻辑重构的过程中，也在不断实现底层系统架构的迁移和迭代，随着金融行业数字化转型不断深入，存量的开放平台传统应用往往具有历史比较久远、业务长期稳定、关联应用较多等特点，必须有效控制迁移风险，保障迁移过程中应用服务平稳运行。

三是需平衡知识产权风险和迁移成本。部分商业数据库产品的高级特性和软件包存在知识产权壁垒，国际化金融机构在数据库迁移过程中需更加重视目标数据库产品的知识产权风险，保障技术供应链的安全合规和稳定可靠。完全兼容商业数据库、无需应用层改造的迁移方案是否适合金融行业大规模推广使用，目前业界尚无定论。

## 2、解决方案

金融行业与数据库技术栈高度耦合的开放平台传统应用转型，可借助异构数据库迁移辅助工具自动化进行语法分析和转换，快速低成本地实现从商业专用平台向开放创新平台迁移；后续根据业务发展需要，随业务创新逐步重构相关应用模块，最终实现应用业务处理逻辑与数据库技术栈的解耦。

在具体实施上，需从开发成本控制、服务连续性保障、测试效率提升、数据复制优化、灵活部署等方面开展技术攻关，确保开放平台大型传统应用转型的便捷性和稳定性。为提升数据库版本升级和故障期间的服务连续性，可采用同城双集群解耦部署方案，通过磁盘级增量日志强同步，实现集群间切换  $RPO = 0$ ， $RTO < 60$  秒，如下图所示：

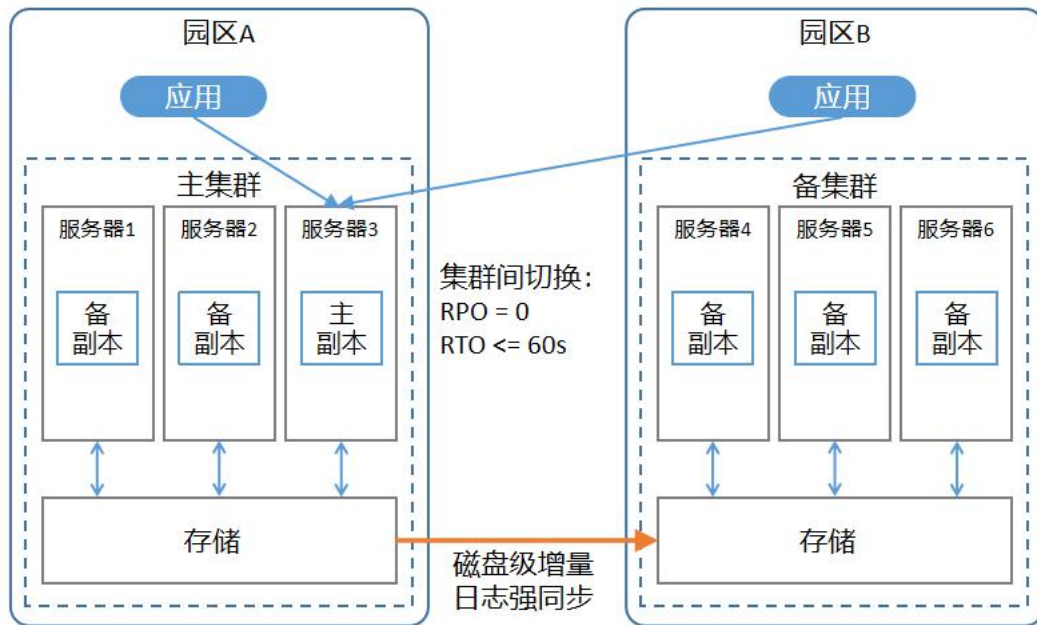


图 3 同城双集群解耦部署方案

通过建设配套的自动化数据库迁移工具，对应用 SQL 代码进行语法分析，提前评估和规划迁移工作进程，识别迁移风险，再通过工具

自动化进行语法转换和逻辑校验，降低迁移成本、控制迁移风险。



图 4 数据库迁移工具

### 3、典型案例：工商银行开放平台传统应用转型试点

针对金融行业传统集中式数据库与应用耦合度高的转型难点，工商银行积极探索金融与科技融合新模式，以联合创新课题研究等方式加强与头部企业合作，开展技术攻关，加快分布式数据库产品的试点应用，推动数据库产业生态发展。

一是结合金融应用场景提升数据库产品成熟度。针对银行大型业务系统数据库转型过程中出现的新场景新问题，对标高等级金融应用性能容量和容灾要求，提出并被采纳 400 余项有效优化需求，促进分布式数据库产品能力提升，更好地适配金融行业应用场景。

二是推动攻关大型集中式业务系统跨异构数据库的低成本迁移解决方案。针对部分传统集中式数据库特有的非标准 SQL 语法，提供迁移解决方案及配套迁移工具，显著降低金融行业大规模存量传统应用的数据库转型成本。

三是共同建设数据库技术金融应用的实施指引。结合工行实施的典型案例，开展“分布式数据库在金融级复杂核心账务系统场景实施指引”等课题研究，探索分布式数据库产品在金融核心应用场景的最



佳实践。

工商银行自 2019 年引入商用分布式数据库以来，已在包括对公理财、实物贵金属等 20 多个业务系统试点上线，覆盖办公系统、一般业务系统和关键业务系统各类典型业务场景。通过在信贷业务和平台核心系统等大型业务系统开展数据库转型技术攻关，聚焦传统数据库与应用耦合度高的难点进行突破，形成了一套高效可控低成本的数据数据库转型技术方案及配套工具，确保系统交易的连续性、数据迁移的完整性、数据库切换的安全性，为金融行业数据库转型提供参考案例。



图 5 工商银行典型试点场景

## 三、金融行业数据库应用未来展望

### (一) 金融行业数据库技术应用的思考

#### 1、多种数据库产品并行发展

与金融行业使用的商业集中式数据库相比，分布式数据库产品发展时间普遍较短，技术成熟度和产品稳定性有待金融行业生产环境长时间平稳运行的检验，目前还没有绝对优势产品出现，金融应用场景面临多种数据库产品选择的局面还将持续。



考察金融行业数据库技术应用发展历程，从商用数据库 DB2、Oracle、SQL Server 到开源数据库 MySQL、PostgreSQL，以及文档数据库、大数据平台的引入，再到近些年繁荣发展的各种分布式数据库产品，金融行业使用的数据库产品和技术日趋多样化，这是金融业务不断创新发展在技术层面的体现。随着金融行业数字化转型逐步深入开展，金融应用持续创新，金融服务不断优化，金融业务场景的广度和深度都大幅扩展，对金融数据的使用方式日益多样化，单一数据库产品已难以支撑金融行业所有应用场景。多种数据库产品并行发展的现状将在金融行业持续存在，需针对具体应用场景对数据库能力的需求和侧重，选择合适的数据库产品。

金融行业数据库架构转型的重点难点场景，使得分布式数据库成为业界关注的焦点，但集中式数据库的应用场景并未消失，集中式与分布式数据库各有其适用的应用场景。分布式数据库解决了集中式数据库性能容量扩展能力不足的问题，相应地也在系统层和应用层付出了多方面的成本。

在系统层面：从专用大型机服务器迁移到通用服务器，单体设备可靠性降低，采用分布式架构可以实现更高的可用性和扩展性，同时也带来冗余备份、网络交互等方面的开销，硬件节点使用规模快速扩张为大型数据中心的节能减排和机房规划带来较大压力。大规模分布式集群的系统复杂度呈指数级上升，必须具备与之适配的运维管理能力作为支撑，需要在运维管理配套的系统能力建设和人才储备方面加大投入。

在应用层面：根据业界实践，将分布式系统完全封装成一个逻辑单库的解决方案虽然能够简化应用开发模型，但系统性能开销太大，因此适用场景有限。为充分发挥分布式架构优势，应用层也需要投入更多的研发设计成本，一是需要进行合理的数据分片设计，通过高内聚低耦合的数据规划，尽可能减少跨节点访问；二是需要在系统架构设计中充分考虑节点故障的容错及柔性事务的处理。

综合来看，集中式与分布式数据库的使用成本可表示如下图：

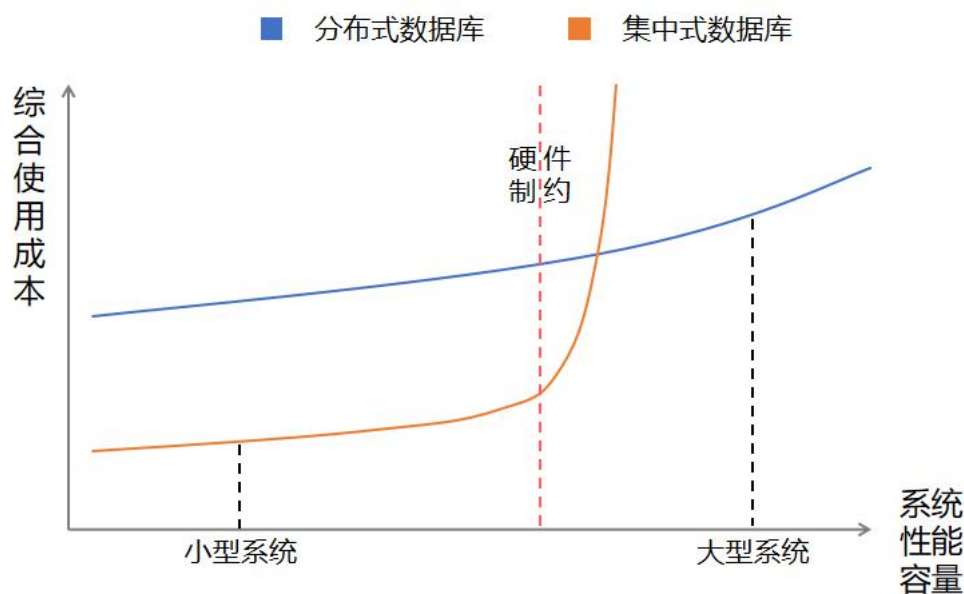


图 6 数据库综合使用成本示意图

因此，对于小型系统和业务规模稳定的应用，可优先考虑集中式架构能否满足要求。基于同样原因，大多数分布式数据库产品都支持集中式单体部署模式，在业务规模较小时规避分布式架构的成本开销，同时保留了随业务规模增长而横向扩展的灵活性。

## 2、各类技术路线相互融合

作为金融信息系统关键核心基础设施，数据库一直是学术界和产业界持续创新的重点领域，技术热点频现，产品形态众多，各数据库技术路线高速演进，呈现出相互融合发展的趋势。

一是市场竞争推动技术力量整合。部分研发实力较强的数据库厂商，前期并行投入多条技术路线。随着数据库市场竞争白热化，不约而同开始推动数据库品牌统一和产品线融合。

二是单一产品跨多条技术路线发展。从数据库产品的层面，各种产品互相借鉴引入其它技术路线数据库的能力，通过大量工程优化弥补架构设计上的不足，在适用场景上有很程度的重叠，已经出现了一款数据库产品同时跨多条技术路线的情况，比如部分 NewSQL 数据库往云原生方向发展、部分中间件架构数据引入原生分布式一致性协议。

各技术路线在融合与交织发展中，共同向着解决用户不断发展的业务需求和使用痛点这一目标动态收敛。分布式数据库产品在金融行业的应用与推广，无论选择哪条技术路线，都需更加重视产品化能力提升。满足金融行业在高可用容灾、数据一致性、业务连续性和系统可扩展等方面的更高要求，提升分布式环境下对应用研发和系统运维的支持能力，是金融级数据库最核心的竞争力。

例如，分布式数据库产品不仅需要提供金融级高可用能力，在节点级/园区级异常故障场景下保证数据服务可用性，还需充分考虑新旧数据库系统迁移期间、数据库版本升级期间、云底座或网络等基础

设施升级变更期间、应用版本数据库对象投产期间、大批量作业执行期间等各类实际落地的应用场景，提供完整的业务连续性解决方案。

技术驱动型产业需依次经历技术、产品、生态三个发展阶段，分布式数据库目前已转入产品竞争阶段。分布式数据库在技术上屡有突破，在多数单项技术指标上都能实现对传统数据库的超越。而在产品化方面，分布式数据库因为发展时间普遍较短，分布式环境下又引入了新的挑战，产品化程度对比商业集中式数据库还存在一定差距，需要通过实际的应用场景持续打磨，提升产品成熟度。

## **（二）金融行业数据库技术应用的建议**

### **1、建设标准体系，引领金融行业数据库技术发展**

标准化在推动金融科技发展和金融治理能力现代化中发挥着基础性、引领性作用。建议金融行业在人民银行、金标委等相关单位指导下，加快推动数据库金融应用系统标准体系建设。工商银行作为北京金融科技产业联盟分布式数据库专业委员会主任单位，一直以来坚持“金融科技应用先进可控”的工作目标，积极统筹规划，牵头组织研制分布式数据库金融应用标准规范，从技术架构、安全防护、灾难恢复等方面确定金融行业标准要求，联合金融同业和科技产业共同推动数据库产业生态发展。

面对数据库领域的高速发展，金融行业数据库应用未来主流技术路线存在不确定性，金融行业应加强数据库应用开发规范指引的完善与落实，严控新增应用模块使用特定数据库产品专有特性，引导应用

随业务重构向着标准化方向演进，逐渐实现金融应用系统与特定数据库技术栈解耦，为未来 IT 架构发展保留灵活性，避免重复迁移改造。

## **2、促进产用对接，推动产业生态发展成熟**

依托北京金融科技产业联盟，10 余家人行相关机构、80 余家金融机构、140 余家金融科技公司及相关服务厂商、10 余家科研院所联合组建分布式数据库专业委员会，共同促进分布式数据库产业发展成熟，加快推动分布式数据库在金融领域的技术攻关和应用落地，结合金融业务转型需求，为数据库新产品新技术提供金融级实战场景和实施案例，加快实现关键核心技术重大突破。

近年来，多款商业分布式数据库产品宣布开源，引发广泛关注。开源作为软件构建和生态培育的一种重要方式，有助于用户广泛参与到产品反馈和创新，促进产品的成熟和生态的发展，在推动金融机构数据库转型方面也起到了积极作用。

## **3、加强人才培养，激发金融行业创新活力**

人才是创新发展的第一动力。金融行业需加强数据库专业人才的培养，通过创新人才激励机制，深化与国家战略科技力量的产学研用联合创新，全方位建设金融行业数据库技术人才资源池，为数字化转型健康发展提供保障和动能。

当前宏观经济环境发生深刻变化，数据库等基础信息平台面临跨

越式发展的历史机遇，金融行业也正处于数字化转型关键时期。“道阻且长，行则将至。”工商银行愿与金融同业及科技产业携手前行，共建标准体系，加强应用创新与技术创新的循环驱动，促进数据库关键技术成熟完善，赋能金融行业数字化转型。

## 参考文献

- [1] 中国人民银行办公厅, 中央网络安全与信息化委员会办公室秘书局, 工业和信息化部办公厅, 中国银行保险监督管理委员会办公厅, 中国证券监督管理委员会办公厅. 关于规范金融业开源技术应用与发展的意见. 2021
- [2] 王珊, 萨师瑄. 数据库系统概论(第5版). 2014
- [3] 中国信通院. 数据库发展研究报告. 2021
- [4] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The Google File System.
- [5] DEAN, J., AND GHEMAWAT, S. MapReduce: Simplified data processing on large clusters. In Proc. of the 6th OSDI (Dec. 2004), pp. 137 - 150.
- [6] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Michael Burrows, Tushar Chandra, Andrew Fikes, Robert Gruber. Bigtable: A Distributed Storage System for Structured Data. OSDI 2006:205-218.
- [7] J. C. Corbett, J. Dean, et al. Spanner: Google's globally-distributed database. OSDI 2012.
- [8] Alexandre Verbitski, Anurag Gupta, Debanjan Saha, Murali Brahmadesam, Kamal Gupta, Raman Mittal, Sailesh Krishnamurthy, Sandor Maurice, Tengiz Kharatishvili, Xiaofeng Bao. A

mazon Aurora: Design Considerations for High Throughput Cloud-Native Relational Databases. SIGMOD 2017.