

国产数据库先行者ArgoDB——以自主创新技术，加速全行业国产化替代

作者：杜思雨 编辑：杜思雨 2022-12-22 11:46 [IT168网站 原创](#) [分享](#)

2022年的党政信创试点即将进入尾声。党政信创国产化试点的成功加速了其他行业国产替代的脚步。预计从2023年起，数据库国产化替代工作将在各行业的信息化、数字化建设中遍地开花。随着信息化建设的推进，业务数据不断增长，业务处理愈发复杂，传统的单机或者集中式数据库逐渐无法满足对原有业务和新兴业务的处理需求。由于传统数据库在功能、架构等的局限性因素，难以支撑更高并发、更低延时的业务诉求，同样也难以满足企业多模融合、湖仓一体、隐私计算等新兴业务需求。

在国际上，我们不难发现，在2022年的俄乌冲突中，冲突双方通过限制底层通用技术的使用，可以干预并控制民生、政府、金融等行业的正常运行。在Oracle宣布暂停俄罗斯所有业务后，大量企业的ERP、财务系统等一系列基于Oracle数据库的业务系统在很长一段时间内均处于瘫痪状态，相关业务无法正常开展。

同样在国内，从美国近年来多次发动的经济制裁与技术封锁行动中，我们也深刻认识到科技战必将成为未来IT领域的新常态。作为信息化支撑的基础软件，数据库的国产化替代已迫在眉睫。但是数据库国产化替代往往涉及到IT底层架构、业务应用等各个方面，并不是一蹴而就的事情。

星环科技坚持核心技术自主研发与创新，打造先进的分布式分析型数据库ArgoDB，深耕国内市场，推动各行业国产化替代落地，助力企业加速数字化转型。

坚持自主研发，力求产品安全可控

在国产化替代的大背景下，受利好政策和用户需求的催化，近年来“国产数据库”如雨后春笋般涌现。但是剥开“国产数据库”华丽的外壳，究其内里，我们不禁要问，“国产数据库”真的能够实现数据库技术的“完全自主可控”吗？其实，我们不难发现，大多数所谓的“国产数据库”中，绝大多数是基于开源MySQL或PostgreSQL数据库进行加工改造，“基于开源”并非原创，“抄作业”始终难以内化为自己的东西。缺少对数据库核心技术的掌控，又如何真正实现“完全自主可控”呢？

ArgoDB自诞生以来，星环科技始终坚持数据库技术的自主研发路线，一步一个脚印，历时长达9年时间，逐步积累了一系列数据库的核心技术，如分布式技术、实时数据处理技术、多模型数据的统一处理技术、湖仓一体化、隐私计算技术等。根据工信部电子第五研究所的代码扫描结果显示——ArgoDB代码自主率（行数）超过90%，完全满足了各行业对国产数据库“自主可控”的诉求。此外，ArgoDB还通过了信通院分布式数据库产品能力基础能力和性能测试2项测评认证，进一步验证了使用ArgoDB数据库建设自主可控的分析业务数字化底座的可行性。

得益于自主创新技术，2019年8月，经过国际事务处理性能委员会（简称TPC）官方审计，ArgoDB成为全球第四个通过TPC-DS基准测试的数据库产品。

当前，ArgoDB已与国内主流软硬件信创厂商完成了兼容适配互认证，并且深度适配飞腾、鲲鹏等国产CPU及麒麟、UOS等国产操作系统，全面融入国产化生态。

基于技术创新，为数字化转型铺路

ArgoDB坚持核心技术全面自主创新，并积累了一系列核心技术。不止能够帮助企业实现国产化替代，而且加速企业数字化转型，满足未来对于更多业务场景的需求。

1 高性能、高扩展、高可用的分布式计算框架

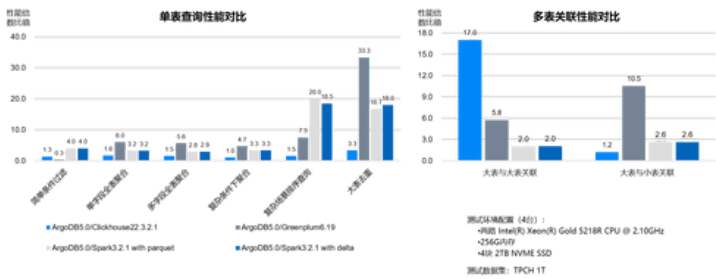
1 高吞吐、低时延的实时分析能力

1 支持10种数据模型，轻松实现多模型数据融合分析

1 支持多方言SQL兼容，协助业务极简改造

高性能、高扩展、高可用的分布式计算框架

基于分布式计算框架，ArgoDB能完美支持交互式查询、基于索引方式的点查等，协助用户构建大型的、低延迟的数据分析应用程序。ArgoDB凭借Native、向量化计算引擎和多模存储引擎的双重利器，性能大幅度提升。典型应用场景已全面优于Clickhouse、Greenplum、Spark等常见OLAP引擎。在相同的测试环境配置下，ArgoDB的OLAP场景性能是Clickhouse、Greenplum、Spark的2~10倍。



此外，随着新兴硬件的不断发展，为了充分发掘新兴硬件的更多潜能，ArgoDB结合新硬件技术，满足更复杂的业务场景需求。如基于闪存存储硬件进行深度优化，相比机械

硬盘分析性能有显著提升，让用户真正感受到新兴硬件带来的巨大性能优势。

基于分布式计算框架，ArgoDB还能够提供相对于传统并行计算架构更好的扩展能力。大多数基于并行计算的数据库（如Teradata、Greenplum等MPP架构数据库）在实现集群扩展时，对IO资源消耗较大，直接影响现有业务的运行。部分基于并行计算架构的数据库在调整数据服务节点过程中，会出现元信息读取的不可用，导致必须停服而中断业务。而对于基于分布式计算框架的ArgoDB数据库而言，由于通过TDDMS（星环分布式数据管理系统）进行了元信息的统一管理，集群规模调整过程中，不会因为元信息同步导致对数据表无法读写，在集群规模变更时，仅需将小部分数据移动到新节点即可。

高吞吐、低时延的实时分析能力

对于上游业务数据库（如MySQL、Oracle等）以日志形式实时传递数据的场景，传统数据库一般通过对日志解析，再以SQL重放的形式实现数据加载，在小数据量场景下可以实现近实时的数据同步和数据分析。但随着上游业务源的增多，同步数据对象变多，需同步的数据量随之增大，如果仍采用SQL重放的方式处理数据的实时同步，在分析库端的SQL提交、SQL解析等过程中会出现严重的性能瓶颈，随之就影响了分析库中实时分析业务的性能。同样的实时写入分析场景，如果采用流式写入的方式，在应对大量表数据需要同步场景时，对于流式写入中Kafka Topic的管理也会成为处理瓶颈，并且对上千任务的调度和维护成本也急剧增长。高并发、高吞吐的实时数据写入分析型数据库，并同时对外提供极致高效的查询分析服务是对分析型数据库能力的一个重大挑战。

基于上述实时写入和实时分析的业务场景，ArgoDB 以新的数据写入接口方式提供了新的实时数据接入方案，该接口具备对数据的Insert、Update、Delete和Upsert操作能力。对于大规模的数据实时同步场景，客户端可通过对接口的调用，直接将数据写入ArgoDB的存储引擎中，在实时更新的场景下进行高效的数据分析。

支持10种数据模型，轻松实现多模型数据融合分析

为解决多种数据库或平台混合使用的混合架构中的数据孤岛问题，ArgoDB提供了多模型融合分析处理能力。在编译器层、计算引擎层、数据存储管理层、资源管理层实现了统一重构，提供了统一的 SQL 引擎、统一的计算引擎、异构的数据存储模型、统一的分布式数据管理系统和统一的资源管理系统的技术架构，能够支持关系型、搜索、文本、对象、图等10种数据模型。

ArgoDB 通过统一的元数据管理、统一的事务管理和完备的多模优化器支撑了对不同模式数据的统一读取调度，支持跨模型联合分析，满足更多“一库多用”场景。例如传统反欺诈解决方案由于不同数据模型分散存储在不同的数据库（例如关系型数据库和图数据库）等原因，在实际业务中需要大量数据转换操作，应用实施成本高，实时性有待提高。ArgoDB为解决此类业务中的数据关联难问题，将多模型数据存储、数据融合处理能力进行迭代加强，用户可将关系型数据和图数据库进行统一存储，用户只需通过一个 SQL 即可关联查询分析关系型数据和图数据，在数据免搬迁、减少人工操作的同时，提升业务效率。



支持多方言SQL兼容，协助业务极简改造

Oracle、IBM DB2、Teradata等国外数据库往往已经在企业中使用多年，国产化替代往往涉及到IT底层架构、业务应用等各个方面，因此迁移过程比较复杂。大多数国产数据库

存在SQL 语法无统一规范、对 ANSI SQL 标准和传统关系型数据库方言支持度较低、兼容的存储过程语法有限等问题。这些问题导致国产化替代落地时，开发人员需要耗费大量人力进行业务改造，数据库使用人员需要重新学习产品用法。

为了使得各行业国产化替代时，业务平滑迁移，降低迁移成本。星环科技自研编译技术，实现了用统一的 SQL 语言操作不同模型（如表、文档、图等）的数据。与其他国产数据库相比，SQL 编译技术达到业界先进水平，支持绝大部分 ANSI92、99、2003 SQL 标准，兼容传统关系型数据库方言，如 Oracle、IBM DB2、Teradata等，支持存储过程，从而大幅降低了大数据应用开发的难度和成本。

从Oracle、IMB DB2、Teradata等数据库迁移至ArgoDB时，开发人员无需对原有业务进行过多的改造，使用人员无需重新学习产品用法，大幅度降低了迁移至ArgoDB的技术与人力成本，快速实现业务的平滑迁移。

经过9年的自主研发，ArgoDB已全面超越了国外数据库产品，可以为用户提供更优的解决方案。

对比项	<u>ClickHouse</u>	Impala	<u>Greenplum</u>	ArgoDB
易用性	不支持事务 支持 SQL，但是不完全支持数据库方言	不支持事务 使用类似于 <u>HiveQL</u> 的 SQL 查询语言	完整的 ACID 事务能力支持 支持 SQL 2003	完整的 ACID 事务能力支持 完整的 SQL 支持，兼容多种数据库方言和存储过程
高可用	SQL 级别重试	SQL 级别重试	SQL 级别重试	Task 级别重试
高扩展	支持数百节点部署	支持数百节点部署	支持数百节点部署	支持数千节点部署
融合分析	支持列式数据存储	支持半结构化 JSON、JSONB、XML 数据等	支持半结构化 JSON、JSONB、XML 数据等	支持关系型、搜索、文本、对象、图等 10 种数据模型
实时入库吞吐量	数十兆/秒	支持离线分析	数十兆/秒	数百兆/秒
实时分析性能	秒级~分钟级		秒级~分钟级	秒级

深耕业务模型，助力客户架构升级

经过多年发展，ArgoDB 已经在金融、政府等领域有数十个落地案例，其中包括诸多成功替换 Oracle、Teradata、IBM DB2 等传统数据库的案例。

某银行替换Oracle数据库，构建新一代实时数据仓库

某行是当地规模最大的商业性地方金融机构，行内原基于 Oracle 数据库建设业务系统。随着行内在移动互联渠道提供的客户服务逐渐丰富，客户点击流量增长快速，客户的体验要求也在提高，对数据库的查询和分析能力也提出了较高要求。原 Oracle数据库在业务支撑能力方面有以下问题：

- a) 随历史数据不断增长，现存统一收单系统的交易明细查询结果输出缓慢，无法快速精确查询到相关交易；
- b) 业务需要实时统计分析用户的相关信息，改进业务质量和产品的服务水平。基于 Oracle 的复杂分析能力的欠缺在业务中的掣肘愈发明显，且经常出现数据加工延迟的情况；
- c) Oracle 以交易型业务为主，分析型业务挤占交易业务的资源，给日常业务的稳定性带来隐患；
- d) Oracle支撑场景单一，无法满足诸如基于实时数据处理等的创新型业务。

基于星环 ArgoDB 数据库平台，支持满足仓集一体、实时数仓等建设需求，同时具备完整的 SQL 及 PL/SQL 支持。通过支持 Oracle 方言，极大降低了 Oracle 数据库业务数据和现有分析型业务的迁移成本。在分析型业务方面以更低成本、更高性能完整替代了传统 Oracle 数据仓库，确保分析型业务与交易型业务的隔离。

新平台满足了行内包括历史明细数据查询、交易流水查询、实时交易大屏、大额交易提醒等十多个关键查询业务场景需求。平台针对各类分析型业务的自动性能优化，保障了

多用户高并发场景下的性能要求。平台将源数据库 Oracle 的增量数据以秒级延时快速同步到 ArgoDB 实时数仓，尤其确保了对源系统数据有删改的经常性调账退款业务数据能即时反映在分析系统中。基于实时落库的业务数据，实现多流水表多维度数据整合的交互式复杂分析，将原基于 Oracle 的离线级分析能力提升到秒级的准实时级交互式分析能力，为行内未来多种复杂的分析型业务应用的拓展与更高的实时性要求打下坚实的技术基础。

某银行替换Teradata数据库，构建湖仓一体化架构平台

某银行自 2012 年起建设了基于 Teradata 的数据仓库。随着银行业务的演进，业务数据规模不断扩大，随之待处理数据激增，原Teradata 数据仓库因功能单一，且数据处理效率低下，难以与更多数据源及上下游系统对接，导致出现了新业务上线困难等问题。

在该银行数据仓库解决方案中，针对客户现状进行数据/体系架构设计，在原有系统基础上提出了优化方案。在新数据仓库系统中，基于分布式分析型数据库 ArgoDB构建，不仅支持 SQL 99/2003 标准语法，并且完整兼容原 Teradata SQL 语法，方便开发人员熟悉使用。在原有基于 Teradata 的数据仓库整体迁移到星环科技大数据基础平台过程中，完成了临时数据层迁移、操作镜像层迁移、临时历史层迁移、基础模型层迁移、汇总层迁移、应用集市层/视图迁移等迁移工作，制定并落实数据标准管理办法及流程、数据质量管理办法及流程、元数据管理办法及流程、配套管控系统功能模块设计与实现，完善整个数据管控与治理体系，完成了数千张数据表、近万个脚本、近千个数据服务接口的完整迁移，推动了该银行的大数据战略实施。通过ArgoDB替换Teradata，从根源上解决了该行在数据仓库领域存在的问题，提升银行对业务发展的数据支撑能力。

 0

相关文章