

ifenxi



# 小步快跑， 数据库持续敏捷迭代

2024爱分析·数据库应用实践报告

04. 2024

## 报告编委

### 报告指导人

张扬 爱分析 联合创始人&首席分析师

### 报告执笔人

孟晨静 爱分析 分析师

### 外部专家（按姓氏拼音排序）

张贤孝 中通快运 信息中心副总监

数据研发部 中通快运 信息中心

## 特别鸣谢（按拼音排序）



# 目录

1. 报告综述	6
2. 分析型数据库	9
3. 云原生数据库	17
4. 展望	20
关于爱分析	21
研究咨询服务	22
法律声明	23

CHAPTER

# 01

## 报告综述

## 1. 报告综述

在信创政策、技术创新和业务需求等因素共同驱动下，我国数据库市场保持快速发展，据大数据技术标准推进委员会测算，2022 年中国数据库市场规模为 403.6 亿元，预计 2027 年将达到 1286.8 亿元，年复合增长率为 26.1%。

图 1：2022-2027 年中国数据库市场规模



### 政策端：信创自主可控推动数据库国产化加速

信创已经成为国家战略之一，通过发展信创产业实现信息技术领域的自主可控，保障国家信息安全是国家经济稳定发展的前提。随着大量中央及地方信创相关政策的出台，信创在“2+8+N”行业加速落地。其中数据库作为承载企业数据存储和管理的基础设施，是信创基础软件的替换重点，国产数据库市场迎来加速发展的黄金时期。

在信创政策推动下，以金融、电信、政府、制造、交通为代表的行业对国产数据库需求旺盛，未来，随着信创在更多行业的落地，国产数据库厂商将迎来巨大的市场空间，这也为国产数据库突破核心技术、储备技术人才、丰富产品形态提供发展契机。

## **技术端：AI、云计算与数据库融合加深，大数据、IoT 技术催生数据库品类日益丰富**

随着大模型落地加快，大模型与数据库的融合场景愈加成熟，如将生成式 AI 技术与数据库结构设计、架构设计、数据分析挖掘等场景结合，能有效提升数据库开发、运维和分析效率，正成为数据库厂商智能工具开发新方向。

上云是企业数字化转型的重要战略。为适应云应用的研发需求，数据库厂商正联合云厂商推出云数据库产品，为用户提供高效、便捷的数据库服务，如基于云计算的数据库即服务（DBaaS），支持用户在云端访问和使用数据库系统，可提供灵活的数据库管理解决方案。进一步，云计算中的无服务架构（Serverless）技术和服务模式逐渐成熟，其极致的可扩展性和资源细粒度自动控制的特点，使 Serverless 成为云厂商的重点布局。

此外，大数据、IoT 技术的快速发展，使企业数据体量呈爆炸式增长，同时带来异常丰富的数据类型，如时序、GIS、图像、视频、文本等数据类型日益丰富，推动图数据库、时序数据库、时空数据库、文档数据库等各种专用数据库涌现并快速发展。

## **需求端：企业分析需求多元化促进分析型数据库技术更新迭代**

随着数据资产的积累，企业正将数据分析广泛的应用到各个业务中，对分析型数据库的需求也从结构化数据、T+1 周期分析，变成海量、多源异构、高并发、实时等复合需求，推动数据库从性能、可扩展性、架构等多方面持续迭代。

在以上背景下，本报告选择分析型数据库和云原生数据库市场作为重点研究对象，围绕两个市场的典型应用实践展开研究。

CHAPTER

02

# 分析型数据库

## 2. 分析型数据库

随着市场的快速变化和数据资产的快速增长，企业对数据分析的应用愈加广泛，对数据分析的性能要求也越来越高。在场景方面，除固定报表外，自助式分析、即席分析、复杂分析、预测分类场景也日益普遍，正成为企业日常经营的必要手段。在性能方面，以精准营销、风险控制、经营统计为代表的企业实时和准实时业务决策场景在快速增加。而基于 Oracle、MySQL 搭建的传统查询分析引擎在以上场景中表现严重不足，如在对海量数据高并发、实时查询场景响应缓慢，难以适应跨系统的多源异构数据的联邦查询，难以支持复杂分析的准实时场景等。

针对以上挑战，分析型数据库厂商采用多种手段满足企业复杂场景下的性能需求，如优化存储和计算架构，存储上采用存算分离架构，计算上采用分布式计算、MPP 或全内存等架构；如通过外表联邦查询、湖仓一体满足对跨系统多元异构数据的融合分析；以及通过预算算、索引优化、物化视图、向量化执行引擎等多种方式提升固定报表、复杂查询的查询速度等。

### 案例：中通快运重构数据中心，满足业务多种复杂分析需求

中通快运成立于 2016 年，是中通品牌旗下快运企业，聚焦数智物流新趋势，提供面向企业及个人客户的全链路一站式物流服务。目前中通快运全国揽派件网点有 21000 余家，分拨中心超 80 个，自有运输车辆 9600 余台，区县覆盖率达 99%。

#### 传统架构难以适应业务快速发展

中通快运率先在业内提出“数字快运”概念，在产品可追溯、物流智能管理、智慧供应链、辅助决策系统等方面持续投入。中通快运在数字化之初，搭建了基于 Oracle 的数据分析环境。近年来，中通快运业务规模保持快速增长，日快递单量能达到 50 万票，考虑到下单、揽件、运输、分拣、中转、签收等各个节点的操作，单表的最大写入量能达到 20 亿条，业务对海量数据、高并发、实时分析的复杂需求日益高涨，这为中通快运的 IT 系统带来严峻挑战，具体如下：

## 1. 传统架构设计混乱，横向扩容受阻，系统性能面临瓶颈

中通快运传统架构在建立之初，为解决运营操作问题，软件开发中的数据分析和业务操作都在大运主库 Oracle 中实现，分析数据和业务操作数据杂糅，业务逻辑和数据逻辑均通过 Java 应用程序实现，所有表之间存在复杂的级联关系。并且，IT 部门普遍以大运主库 OGG 同步的方式开发数据服务，使得底层数据架构的复杂程度进一步升级，成为主库横向扩容的阻碍。虽然中通快运通过对主库进行了纵向升级，但 OGG 同步路线的快速增长已经逐步消耗掉主库纵向升级带来的性能提升。此外，中通快运传统架构中以宽表的方式进行数据建模，一旦业务运营模式发生变化，需要对宽表进行调整或重构形成复杂的数据模型，如多个表之间存在多对多、嵌套关系等，导致数据同步困难，难以满足业务对数据分析的时效要求。

## 2. 难以满足业务日益广泛的复杂分析需求

为保证业务正常运行，主库 Oracle 中只能存储 3-6 个月的数据，更早的数据则以文件格式进行备份。然而业务部门对数据分析的应用场景愈发普遍，并涌现出更多复杂分析场景，如车辆调度、路线规划、仓储管理、用户需求预测等越来越多的场景需要基于大跨度的历史数据进行预测分析，但备份中的数据难以调用。

## 3. 开发和部署方式呈现烟囱式，无法快速响应业务变化

中通快运的系统之间相互独立，多年来对业务需求的支持以烟囱式开发为主，缺乏统一的数据规范，使得数据开发过程需花费大量时间处理数据格式、数据一致性等质量问题，造成数据开发效率缓慢，同时也带来重复开发、资源浪费的问题。

在对业务和数据反复思考过程中，中通快运大数据团队意识到，只有推动系统架构设计和业务运作流程同时演进，才能实现数据赋能。这意味着大数据团队不能仅仅作为一个支撑部门，更要具有主动赋能的意识。

为推动业务流程变革，中通快运大数据团队开始推动集团高层重新定位部门价值，将数据部门确定为业务赋能核心，纳入业务战略规划中。同时，数据团队与业务部门紧密合作，理解业务需求，提供相应的数据服务和解决方案，实现数据驱动的业务决策和优化。

在技术支持上，IT 部门对数据中心的系统架构进行重构，并着重解决大数据量、高并发、实时场景下的 OLAP 分析问题。在陆续使用过 HBase、Clickhouse 等数据库后，中通快运将目光转向开源数据库，并有两点核心诉求：1) 功能上，数据库需具备强大的联表查询能力，且支持大数据量、高并发、实时分析等场景；2) 运维支持上，中通快运 IT 资源有限，并且需要将更多资源投入业务支撑中，因此要求数据库具有简易运维的特点，且有商业厂商做支撑。综合考虑数据库功能和中投入产出比，中通快运最终选择与基于 StarRocks 的商业化公司镜舟科技一起解决核心 OLAP 分析问题。

北京镜舟科技有限公司致力于帮助中国企业建立卓越的数据分析系统，形成自己的“数据护城河”，旗下产品镜舟分析型数据库和镜舟湖仓分析引擎是基于 StarRocks 发展起来的企业级商用数据库，支持从各类实时和离线的数据源高效导入数据，也支持直接分析数据湖上各种格式的数据，同时具备水平扩展、高可用、高可靠、易运维等特性。镜舟科技服务 70 亿元以上客户数量超过 350 家，包括不限于腾讯、携程、平安银行、中原银行、中信建投、招商证券、众安保险、大润发等企业。

## 重构数据中心架构，建设实时、准实时和大跨度大数据量三层数据服务能力

中通快运数据中心系统架构设计上，主要实现了两个关键点：1. 对操作和分析进行解耦，将数据查询任务与 DML 任务隔离。2. 对数据服务能力进行分层，按照实时、准实时和大跨度大数据量三层的服务能力。具体来看：

图 2：中通快运数据中心建设方案



## 1. 对操作和分析解耦

针对操作和分析杂糅的情况，中通快运对操作和分析进行解耦。

- 1) 针对 DML 操作，由 TiDB 做业务实时数据的落地库，提供实时插入、更新、删除等数据操作服务，TiDB 实时宽表可实现 10 秒及以内的实时响应。
- 2) 针对数据分析，通过镜舟分析型数据库、湖仓分析引擎和 Paimon 数据湖搭建批流一体架构，由镜舟的分析型数据库和湖仓分析引擎承担核心 OLAP 能力。

镜舟分析型数据库能支持大规模的多表关联查询，其提供的 CBO（基于代价的优化器）、向量化引擎、前缀索引等技术，能实现高效的多维度聚合操作与明细数据的拉取，此外，镜舟分析型数据库提供的物化视图可以在明细查询的基础上做不同维度的融合操作，最终实现应用层上的多维灵活分析。

批流一体架构提供准实时和大跨度大数据量的数据分析服务：

- 面向（一年期内）大数据、高并发、实时要求高的分析场景，镜舟湖仓分析引擎从 Paimon 中读取并存储一年期数据，提供准实时响应，响应速度在 10s 左右。
- 面向跨年的长期历史数据的大规模复杂查询，由镜舟分析型数据库直接对 Paimon 历史数据进行分析，整体从业务端数据传输到服务端的响应时间控制在 2-3 分钟内。

## 2. 建立数仓分层架构，实现数据管理

镜舟湖仓分析引擎中的物化视图能简化湖仓分层建模，如中通快运通过物化视图可将 Paimon 数据湖中数据作为 ODS 层，进一步加工处理形成 DWD、DWS 层。湖上数据可以由镜舟湖仓分析引擎直接进行分析，也能通过物化视图将湖上数据写入到镜舟湖仓分析引擎中加速查询。数仓分层架构下，不同层级的数据能为不同的应用程序提供查询服务。

## 3. 使系统具备弹性扩展能力

通过镜舟分析型数据库的 MPP 执行框架以及存算分离的特点，中通快运数据中心具备了弹性扩展能力，计算、存储能分别独立扩展。

## 4. 简化数据开发操作，加速数据开发效率

借助镜舟科技的数据库产品，中通快运大幅提升数据开发效率，如在数据集成上，镜舟分析型数据库提供多种数据导入方式，包括实时数据导入、批量数据导入、流式数据导入等，以满足不同业务场景下的数据导入需求。在数据建模方式上，镜舟分析型数据库提供了星型、雪花、大宽表、预聚合等方多种数据建模方式。此外，在升级后的批流一体架构中，镜舟湖仓分析引擎支持通过 Paimon Catalog 组件直接查询 Paimon 数据，以及结合 insertinto 操作实现数据的转换和导入，能极大简化数据集成和处理流程。

## 高效分析支撑业务场景，为业务发展赋能

### 1. 高效满足业务多种分析需求

1) 针对实时分析场景，实时数据导入、更新和计算，可以广泛应用于物流订单状态更新、TP 数据库同步、多流 join 写入宽表等场景。

2) 支持复杂分析的准实时场景，以往中通快运使用 Oracle 存储数据时，业务流程经常崩溃，数据服务难以维系。数据中心重构后，中通快运利用镜舟分析型数据库强大的多表关联查询能力，能在量本利、运营统计等各种复杂分析场景下对一年期内数据实现秒查，有效支撑业务需求。以中通快运对在线分拨中心的货物滞留计算为例，传统 Oracle 存储下，对 60 天的滞留数据计算需每两小时计算一次，每次计算需要将数据拆分为 60 次进行循环计算，经常由于资源不足或性能限制导致计算任务无法完成。在使用镜舟分析型数据库后，中通快运对 60 天滞留数据一次性查询，计算时间平均为 8 分钟。

3) 支持跨年长期历史数据的大规模复杂分析。如进行回溯对账分析，或是基于强化学习对车辆调度、路线规划、仓储管理等场景进行预测分析，提升经营效率。

未来，中通快运计划利用镜舟分析型数据库的异步物化视图能力进一步优化加速，减少外部调度任务，进一步降低运维成本，实现业务响应和成本控制的双重提升。

## 2. 数据架构优化，系统实现弹性扩展

数据操作和数据分析的解耦，使中通快运能优化业务流程、集中管理数据以及灵活进行资源分配，减少系统故障和停机时间，提高业务的连续性和可靠性。

## 3. 有效提高数据开发效率

镜舟分析型数据库中数据集成的便捷性、数据建模的灵活性，多表关联对数据处理流程的简化以及物化视图对数据 ETL 的简化等，都极大提升了中通快运的数据开发效率，使数据应用能快速响应业务需求。

## 服务治理、业务流程再造是实现数据赋能的关键路径

面对业务的快速迭代，中通快运大数据团队积极应对业务变化，保持敏捷性，对基础数据架构和业务流程进行思考和迭代。在这个过程中，中通快运总结了自身数字化建设经验供企业参考。

中通快运认为，在组织数据驱动经营决策的目标下，IT 部门要以业务应用价值为核心，通过数据赋能业务，满足业务战略发展需求。其中要数据赋能的路径包括以下两点：

第一，重视服务治理。将数据作为业务架构设计的核心，建立数据之间的关系和流程，且要具备一定的前瞻性，尽早将操作和分析两种业务解耦。

第二，业务流程再造。业务操作流程要和技术手段同步演进，只有业务流程发生变革，才能实现数据赋能。

在这个过程中，IT 部门要改变以往支撑部门的角色定义，而是作为赋能部门主动推动业务流程改造，才能实现新一代信息技术驱动下的业务、管理和商业模式的深度变革和重构。

CHAPTER  
**03**

# 云原生数据库

### 3. 云原生数据库

数字化转型深入为传统企业带来艰巨的数据挑战，具体体现在：

- 1) 传统企业硬件资源无法灵活扩展难以适应业务快速变化，也无法满足业务流量波动需求，而靠服务器堆积的方式又会带来巨大的资源浪费。
- 2) 传统数据库单点故障或是对容灾备份的配置疏漏难以保证数据的可用性和可靠性。
- 3) 单一数据库难以满足企业业务需求，企业普遍部署事务型数据库、分析型数据库、数据仓库、数据湖等多种基础设施，使数据处理流程愈发复杂的同时也带来了高昂的运维成本。

受限于 IT 资源，传统企业希望能在有限的资源和人力投入下解决上述挑战。

云原生数据库弹性扩展、高性能、高可用、易于维护等特点能灵活满足企业高时效性、高流量的业务需求，保障数据安全可用，并能帮企业节约大量运维开发成本，使企业专注于业务逻辑和创新。在此基础上，云原生厂商也在融合 Serverless、实时灾备、日志即数据、智能弹性、云原生 HTAP 等技术对云数据库的弹性伸缩、资源粒度、计算性能和可用性等方面持续优化。

某保险公司作为全国财产险排名前列的保险公司，积极打造高可靠、高安全、高扩展、及时响应业务需求的基础设施，推进数字化、智能化转型。

在保险改革的大背景下，如何压缩成本、更精准的风险识别和定价、更精准地获客和留住客人，成为了财险公司在车险市场激烈竞争中拓展生存空间的必要能力。为寻找新的业务增长点，该保险公司计划建设基础数据平台。

传统基础数据平台的建设交付一次性投入高、试错成本大，且会造成资源浪费。一方面 IT 资源的规划需要考虑系统峰值，但大量 IT 资源在系统建设初期无法得到充分利用；另一方面，IT 资源无法弹性扩展，难以适应系统快速发展或是紧急情况，极易导致服务性能下降或业务中断。

同时，该保险公司的封闭商业数据库的运维成本高昂，数据同步操作频繁且只能手工操作，工作量大且繁琐，投入成本高。

考虑到传统平台的局限性以及业务对数据库易用和高性能的需求，该保险公司希望通过建设云上金融平台，强化数据赋能。经多方评估后，该保险公司最终与某云厂商达成合作。

云厂商为该保险公司建设金融云平台，提供平滑无感的上云方案帮助该保险公司完成数据迁移，并在云平台上部署云数据库满足该保险公司对高性能数据库的需求。

通过与云厂商合作，该保险企业完成主流商业数据库切换，实现数据库查询性能 7 倍提升，海量访问无压力；确保系统在高访问场景下数据不丢失，同时降低了运维压力，实现 RPO=0、RTO 秒级，整体资源投入节省了 25%。

CHAPTER

# 04

展望

## 4. 展望

随着大模型的广泛落地，企业对数据库 AI 能力的关注将逐渐提升，未来数据库智能化、自动化能力将成为企业选型的重要指标。企业对数据库 AI 能力的需求体现将在三个方面：1) 数据库支持对关系型、文档型、图形型、时序型等多元异构数据的存储和查询，提高数据处理的灵活性和效率，以满足愈发广泛的 AI 应用场景；2) 库内 AI 引擎支持端到端数据清洗、特征工程、模型构建和模型训练，使机器学习训练更高效；3) 数据库与 AI 协作使数据库管理更加智能高效。

## 关于爱分析

爱分析致力于成为中国领先的数字化市场专业服务平台，为企业用户提供数字化规划与落地全流程服务。

爱分析以咨询为牵引，依托实践经验、专家建议和方法论工具，为企业提供最适配的解决方案和落地服务，帮助企业提高项目成功率、优化投资收益，以及控制风险。

爱分析深耕数字化市场 10 余年，在央国企、金融、消费品、制造等行业积累诸多标杆案例和丰富实践经验。

## 研究咨询服务

技术研究	客户洞察	行业研究
新兴技术研究，厂商能力调研，助力数字化最优决策	企业用户需求及实践调研，辅助制定业务与市场策略	行业数字化趋势与实践研判，辅助业务与战略决策
商业研究	品牌&营销	投资研究
基于研究、数据和案例调研积累，辅助业务可靠落地	权威背书，树立行业地位；教育市场，精准触达客户	成熟方法论，一手数据，助力研判机会、稳健投资

## 法律声明

此报告为爱分析制作，报告中文字、图片、表格著作权为爱分析所有，部分文字、图片、表格采集于公开信息，著作权为原著者所有。未经爱分析事先书面明文批准，任何组织和个人不得更改或以任何方式传送、复印或派发此报告的材料、内容及其复印本予任何其它人。

此报告所载资料的来源及观点的出处皆被爱分析认为可靠，但爱分析不能担保其准确性或完整性，报告中的信息或所表达观点不构成投资建议，报告内容仅供参考。爱分析不对因使用此报告的材料而引致的损失而负上任何责任，除非法律法规有明确规定。客户并不能仅依靠此报告而取代行使独立判断。

北京爱分析科技有限公司 2024 版权所有。保留一切权利。



咨询/合作

微 信: ifenxi888

网 址: www.ifenxi.com

北京市朝阳区东四环中路56号 远洋国际A座15层-1510B室



如欲了解更多爱分析精彩洞见  
请关注我们的微信公众号

©北京爱分析科技有限公司2024版权所有



ifenxi

数字化市场专业服务平台