

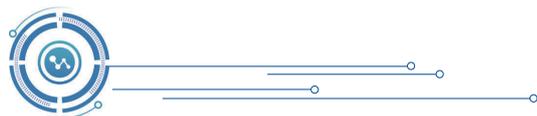
中国CDM技术 及市场应用 白皮书

目录

| | |
|----------------------------------|----|
| 1. 数字经济下的数据市场：数据价值与数据安全并重 | 2 |
| 1.1 数据成为数字经济发展的核心要素 | 2 |
| 1.2 数据安全法制化给数据管理提出新要求 | 3 |
| 1.3 信息技术应用创新对数据管理提出新要求 | 4 |
| 2. 当前数据管理面临的挑战 | 4 |
| 3. 副本数据管理（CDM）构建数据管理底座 | 8 |
| 3.1 CDM 的技术演进 | 8 |
| 3.2 CDM 典型应用场景 | 8 |
| 3.3 中国用户评估 CDM 的五大要素 | 10 |
| 4. 上讯全新的数据管理解决方案赋能数据降本增效 | 12 |
| 4.1 上讯 ADM 的四大数据管理解决方案 | 13 |
| 4.2 上讯敏捷数据治理平台（ADM）核心技术组成 | 14 |
| 5. 行业典型案例 | 19 |
| 总结 | 21 |



数字经济下的数据市场： 数据价值与数据安全并重

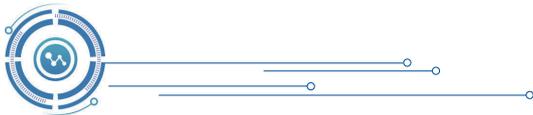


1.1 数据成为数字经济发展的核心要素

随着云计算、大数据、物联网、人工智能等数字技术的快速推进，在带来技术革新的同时，还带来指数级增长的海量数据。相关调研表明¹，2025年，新创建的数据总量将高达175.8ZB。数据在推动产业结构重构的同时，还带来业务模式革新、产品和服务的创新。数据成为驱动数字经济创新发展的核心要素。IDC指出：“在未来数字化基础设施中，企业将以数据驱动为特征，实现数据管理、数据即服务和数据分析。数据是数字经济的石油，是企业产出有效洞察的基础，可以实现流程和决策优化。”“十四五”规划纲要也强调要“激活数据要素潜能”，以推进数字经济新产业、新业态及新模式，实现各行业高质量的发展。

数据驱动企业快速发展。无论是应用的快速开发迭代来满足市场需求，还是高质量的应用快速交付以缩短产品面世周期，或通过更好的客户服务支持来提升客户体验和满意度，以及更快地通过海量数据实现睿智的商业洞察，都是通过数据驱动，来推动企业业务革新，提升企业的行业竞争力。

数据价值的获取，甚至决定了企业的生存与发展。然而，如果把企业所拥有的所有数据比喻为一座“冰山”，那么企业真正投入使用，实现价值挖掘的数据量只是冰山一角。分析公司 Forrester 表示²，企业有 73% 的数据从来没有得到任何利用过。这些数量庞大的“暗数据”³ 成为企业当前业务中最大的未开发的“金矿”。如何通过“暗数据”管理和运营，释放暗数据全部潜力，真正实现数据驱动发展，成为加速企业数字化转型步伐的关键。



1.2 数据安全法制化给数据管理提出新需求

数据在充分释放价值、促进经济发展的同时，也使得数据隐私和安全面临更大风险。每年都有很多企业因为数据泄露遭受严重的财务损失。如何在安全合规范围内充分发挥数据价值已经上升到国家高度。当前国际、国内纷纷颁布了数据隐私安全相关的法律法规，包括：

- ◎ 美国加利福尼亚州投票通过了《加州隐私权利法案》（“CPRA”），该法案在《加州消费者隐私法案》（“CCPA”）的基础上对个人隐私、数据安全、网络安全等提出了更为严格的要求。
- ◎ 欧盟委员会先后通过了《通用数据保护条例》（GDPR）、《欧盟数据战略》、《数据治理法案》，在建立完善的数据保护法律体系的同时，解决数据安全问题。
- ◎ 我国也先后发布了《国家安全法》、《网络安全法》、《数据安全法》、《中华人民共和国个人信息保护法》，分别从不同角度、不同程度对数据保护做了规定，为国内数据安全治理法制化和规范化提供可靠的政策保障。

十四五规划纲要也指出，要求统筹数据开发利用、隐私保护和公共安全，加快数据安全等基础制度和标准规范的建立。在国家宏观政策的引导和规范下，各行业也越加重视数据作为新型生产要素的重要性，数据的安全、合规使用和管理成为每一个企业的重中之重。为此，各行业也纷纷颁布了数据安全和数据隐私相关的法律条文，如：

金融服务行业

《银行业金融机构数据治理指引》，从数据管理、数据质量控制、数据价值实现等方面要求银行业金融机构提升数据治理能力，同时对数据合规提出了更高要求。

医疗服务行业

《国家医疗保障局关于加强网络安全和数据保护工作的指导意见》要求在数据全生命周期内实现有效的数据安全治理，确保医疗行业数据的安全和高效可用。

电信行业

2019年7月1日发布了《电信和互联网行业提升网络数据安全保护能力专项行动方案》，以强化网络数据安全全流程管理，及时整改消除重大数据泄露、滥用等安全隐患。

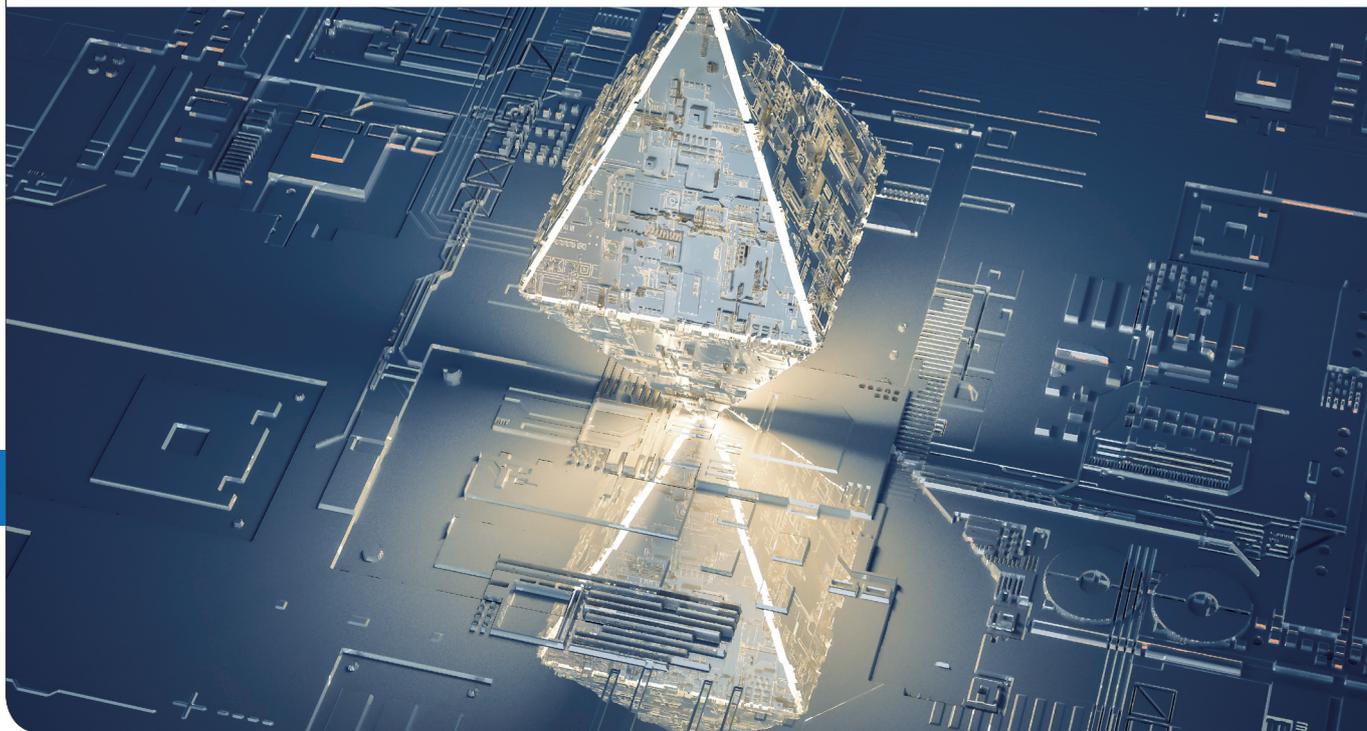


1.3 信息技术应用创新对数据管理提出新要求

信息安全部署已经成为国家重要战略，信息行业国产化及网络安全自主可控的需求更为凸显。自主可控是国产化信息技术网络安全的必要条件。如何依靠自身研发设计，全面掌握产品核心技术，构建可替代的产业链条，通过关键技术、零部件以及软件等的国产化，实现核心技术的自主可控，推动了我国对信息技术应用创新（以下简称“信创”）的迫切需求。

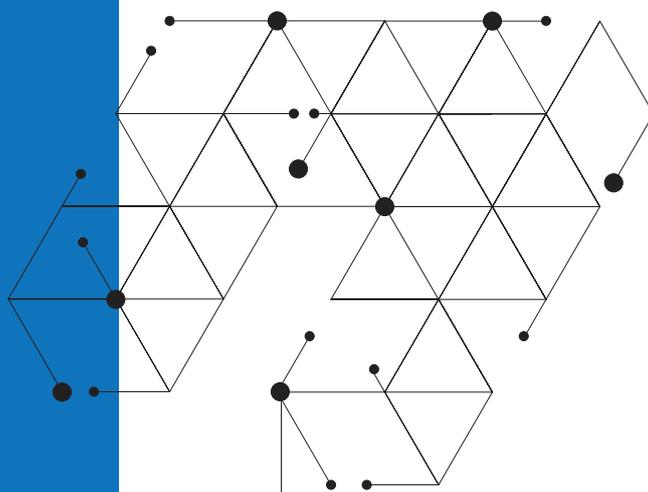
信创不仅是我国当前的重要发展战略，也是我国 IT 产业发展升级的长久之计。在 2020 年，科技部、工信部等国家机构与地方政府已推出多项政策助力信创产业落地。如何通过构建自己的 IT 产业标准和生态，确保 IT 产品和技术安全可控，从而赋能各行业用户实现数字化转型就绪，成为当前信创产业的发展重点。各行各业将信创作为未来 IT 和企业战略布局谋篇重要的考虑因素。

据安全和隐私已经成为国家层、行业层以及产业层共同关注的主题，这也推动了数据管理从传统的针对生产、测试、商业智能分析、信息化分离数据管理，向企业全流程、系统化数据管理平台化的演进，加速了跨企业甚至于跨产业上下游，在数据管理全流程中嵌入数据安全的需求。



2

数据管理面临的挑战

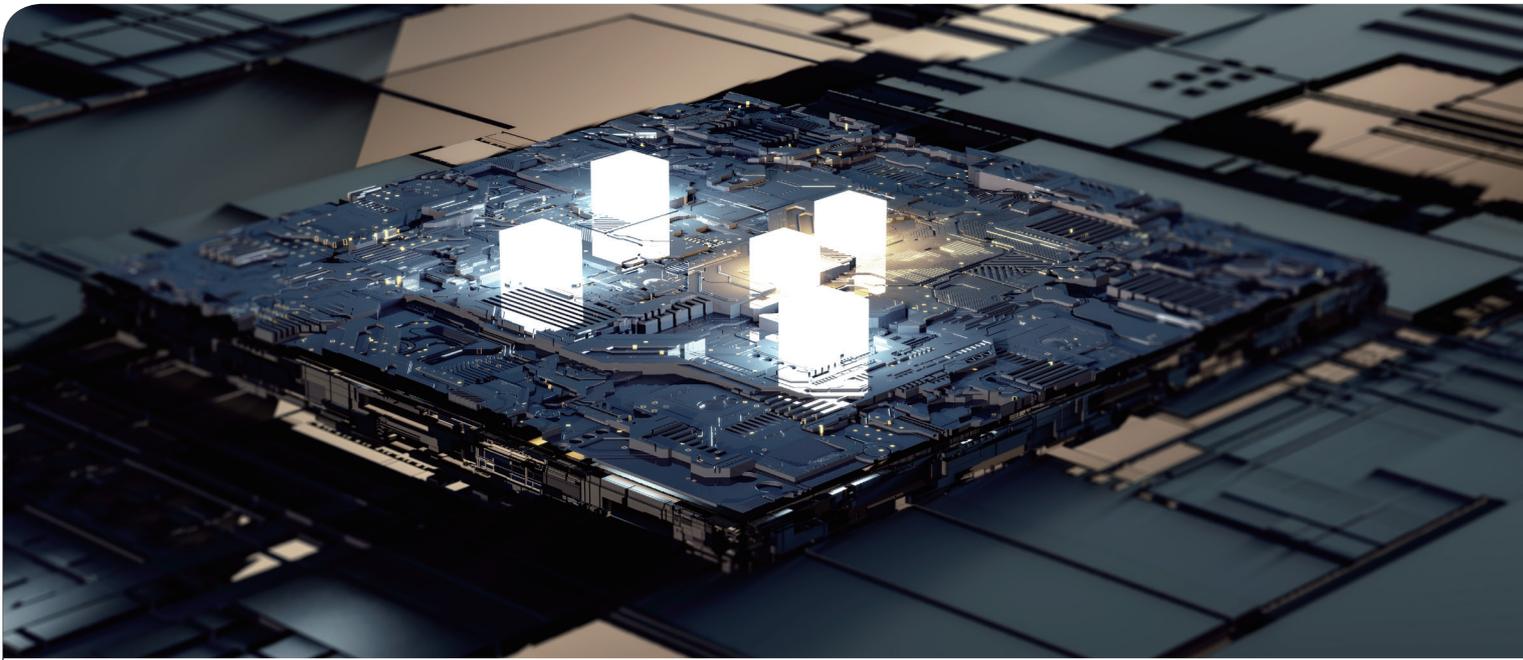


数字经济下“数据为王”，构建数据驱动型企业已经成为大趋势。企业对数据的要求不仅仅是支持核心业务系统，来保证生产系统的稳定和连续，更需要满足敏捷测试开发、审计、品控、事件分析等需求。而上述各种使用场景，都意味着一份生产数据的复制，从而带来数据量的快速增长。对于数据前沿行业，如银行、电信、政府和大型企业，在受益于数据实现业务赋能的同时，如何有效管理这些复制数据，同时确保所有数据的安全性、隐私性和可用性，成为企业面临的挑战。

生产数据可靠、快速恢复的挑战

企业业务需要保持全天候运营，对核心业务系统进行数据备份成为必需。然而传统的备份机制，备份数据格式与原格式不同，备份数据需要通过恢复才能直接使用。不仅备份数据是否可恢复存在疑问，且难以在指定的时间内完成大量的数据恢复，生产数据的连续可用难以保证。再者，随着数据量的快速增长，无法在指定备份窗口完成数据备份，还增加了生产数据的丢失风险。

企业需要具有可恢复效验的数据管理机制，在业务系统发生故障时，能够分钟级完成生产数据恢复，保障业务系统的连续性。



数据准备周期长、交付效率低 通过 DevOps 提升软件交付效率，快速实现产品和服务迭代，是企业构建核心竞争能力的关键。在这一过程中，生产数据需要频繁被调用，以满足软件快速开发迭代的需求。如在金融的“信创改造”背景下，金融测试部门在国产化架构和系统上，进行新的信息系统的研发测试，这就带来从功能、性能到兼容性等一系列数据测试需求。然而，传统数据准备流程下，从数据审批到获取，手续复杂，时间往往按天或月为单位。并且很多开发测试的工作环境还需要不断更新和重新生成数据，这都导致数据准备时间长。再者，传统方式下，数据的导入和导出通常是手工操作，大大增加数据管理员的工作强度。此外，这种手动数据从生产环境到测试环境的迁移，增加数据风险，限制了根据市场和业务需求快速支撑项目的执行能力。

企业需要简化数据准备流程，并通过自动化的数据导出和导入，实现分钟级数据快速交付，以更快响应市场需求，为信创提供自动化测试平台支撑。

副本数据泛滥导致高 TCO 数据的成本管理中，存储成本往往占比最高。传统的数据管理方式下，每一份副本数据的创建，都需要消耗存储资源。以金融服务行业为例，一个 5TB 的数据库，搭建 8-10 套测试环境，传统物理复制需要 40TB-50TB 的存储容量。IDC 调研显示⁴，77% 的企业在当前的生产环境中拥有 200 个数据库实例，包括 Oracle、SQL Server 等。而 82% 的企业每个数据库有 10 个以上的副本。这就意味着企业需要管理的数据库副本数量超过 2000。这些数量庞大的副本数据，意味着高昂的存储资源和维护开支。根据 IDC 的调查⁵，企业的副本数据消耗了 45% 到 60% 的存储空间，并且带来额外的存储软件及产品的采购开支。

企业需要以单个副本数据，即可满足开发、测试、备份、恢复、数据脱敏、统计分析等多种工作负载需求，同时还需要高效的数据压缩技术，来进一步节省存储空间，减少企业不必要的存储开支。

敏感数据的泄露风险 不论是企业的开发测试、财务报表、绩效考核，还是经营管理、审计等，往往都是直接使用的生产或类生产数据。传统的数据管理方案，由于各类数据使用部门的独立性，使得这些来自于业务系统的生产数据缺乏统一、标准的数据安全管控管理系统，难以根据各种数据安全合规政策快速实现企业级数据管理的系统性调整，保证数据合规需求，增加了数据泄露风险和业务违规风险。很多代码效验或测试开发、敏感数据生命周期管理的细粒度和权限管理的缺失，增加了业务合规风险。由于测试环节涉及人员冗多、过程较为复杂，生产数据在导出测试环境的过程不可控，测试库安全防范手段单一，加大了敏感数据的泄露风险。

企业不仅需要高效、迅捷地为不同环境提供数据交付使用，通过集中管控防范数据泄露风险，同时还需要对生产数据进行脱敏规则设定，保证脱敏数据支撑开发、测试、审计、品控等不同业务场景。

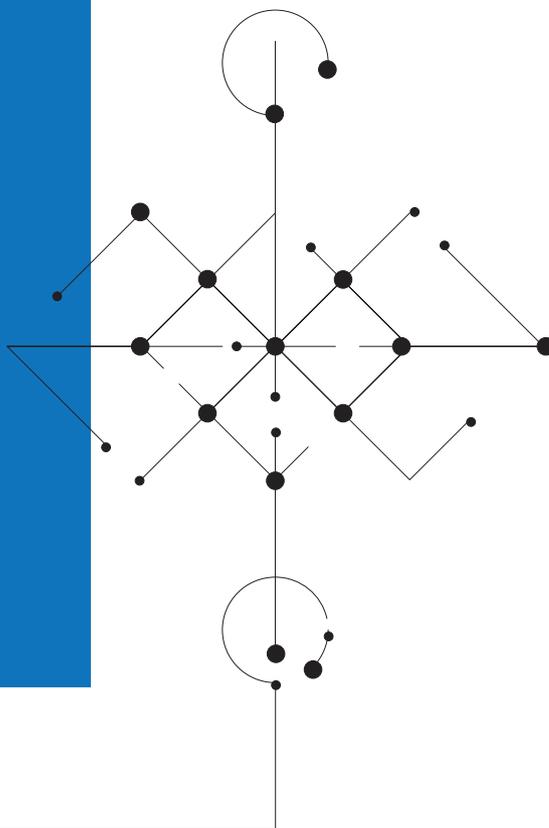
测试数据的“真实性”不足 基于场景和用户需求，测试环境下充足的测试数据量、测试数据的快速更新，以及数据的真实性，决定着是否能通过产品和服务的迭代推动业务创新，拓宽客户入口和提升用户体验。随着微服务化、容器化、DevOps 和 CI/CD 的快速普及，应用研发、测试等环节对测试数据量提出了多样化需求。同时，信创对于系统兼容性的要求，将进一步加大企业对测试的需求。然而，在测试环境下，大规模生产数据实时导入导出来进行测试，鉴于对生产环境稳定性和安全性的影响，几乎是不可能做到的。而采用仅几 GB 的子集数据进行测试，又损失了“类生产”数据的真实大小和保真度，并且这些副本数据与生产系统之间缺乏数据更新机制，往往难以真实模拟或反映出生产系统的特性，导致偏差，从而导致测试结果不真实，甚至测试结果没有参考价值。

数字经济的发展以及我国市场的信创需求，都要求通过仿真类生产数据生成，帮助测试开发环境尽可能贴近真实运行环境，提高开发测试质量，这成为企业的业务重点。

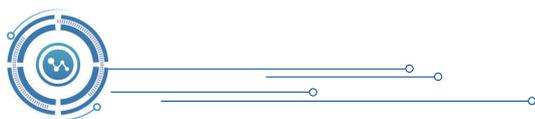
数字经济需要新一代功能多元化、部署智能化，以及自动化、智能、全流程数据安全的数据管理平台。这一平台通过有效的数据管理流程，降低数据环境搭建对人员、存储、带宽资源的消耗；精细化的数据管理，保证生产数据备份的一致性，满足敏捷测试、开发、审计和品控，以及事件分析的需求；自动化和系统化的测试开发、审计、质量流程，降低数据环境对生产资源的资源争用可能带来的业务不稳定性，同时降低数据管理开支，保证各种业务场景下数据的安全合规。



副本数据管理 (CDM) 构建数据管理底座



早期副本数据管理主要基于手动设置管理，由于应用数量较少，数据量不大，不论是时间、人力还是存储成本，企业还能承受。然而随着业务的不断发展，应用系统的数量快速增长，数据量也呈指数级的增长，庞大的副本数据数量成了企业发展的重负。手动副本管理方式，不仅带来副本泛滥、DBA 工作强度加大、存储开支激增，同时增加了数据安全和业务合规风险。数字经济的发展推动了副本数据管理 (CDM) 技术升级，且成为敏捷数据管理平台最重要的核心技术。



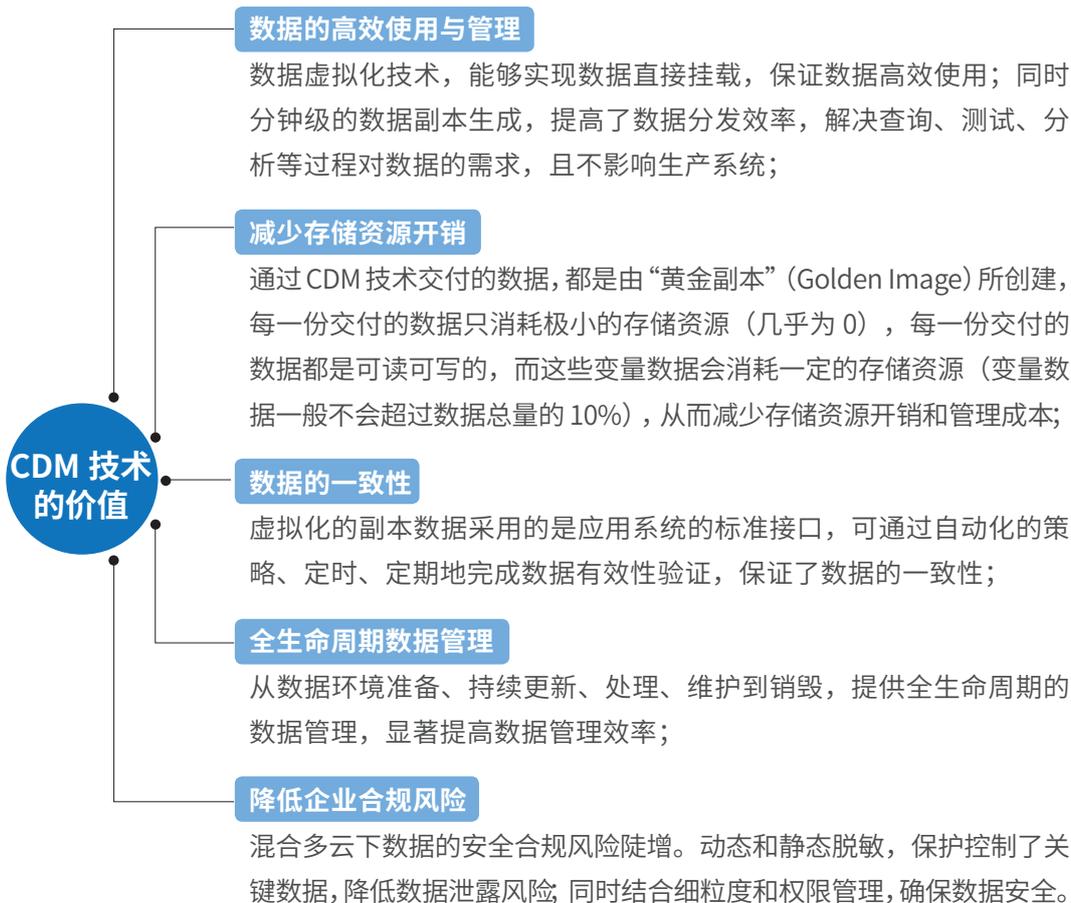
3.1 什么是 CDM

Gartner 定义 CDM (Copy data management, 副本数据管理)：CDM 从生产环境通过快照技术获取有应用一致性保证的数据，在非生产存储上生成“黄金副本” (Golden Image)。这个“黄金副本”数据格式是原始的磁盘格式，原格式副本可直接打开使用，也可再虚拟化成多个副本直接挂载给服务器，分别用于备份恢复、容灾或开发测试。

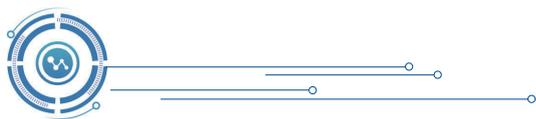
目前在中国市场的以下两类产品——以存储为核心的 CDM 和备份软件提供的 CDM 功能，按照以上 Gartner 的定义，都属于“类” CDM。

- ❌ 以存储为核心提供的 CDM，虽然能够提供快照和克隆功能，但很难实现跨异构存储构建企业级规则驱动的副本数据平台，并缺乏丰富的数据服务能力，无法提供自动化、自服务等功能。
- ❌ 备份软件提供的 CDM 功能，利用备份技术所创建的副本数据，是一种磁盘快照技术，无法保证数据的一致性，并且同一时间的磁盘快照只能挂载一份，难以满足多应用场景的数据需求。
- ✅ 真正的 CDM：一种端到端的创新技术，把单纯面向恢复的应用场景变成了面向数据使用的应用场景，打破了数据备份和数据使用之间的壁垒。通过副本数据在企业各个业务环节的即时可用，为更多业务场景提供数据支撑，在盘活暗数据的同时，实现业务赋能。

数字经济下，CDM 技术所实现的价值主要体现在：



相对于早期数据复制方案，新一代 CDM 技术从“被动响应”到“主动布局”，通过分钟级生产数据恢复，为生产数据提供应急保障，保证全天候运营业务的连续稳定；通过智能自动规则驱动，支撑从测试、开发、审计、分析到报表等全流程数据服务，提高数据交付效率的同时显著降低 TCO，且通过规则驱动自定义脱敏和权限管理，确保数据安全。



3.2 CDM 典型应用场景

A 备份数据的恢复验证

各金融机构（银行、证券、保险、基金等）需要定期投入大量的人力和资源，对备份数据进行手动恢复测试和有效性验证，以保证备份数据的可恢复性。传统的基于手动的恢复验证，不仅需要投入大量人力，并且所需恢复的数据量越大，数据恢复的时间就越长，可谓是耗时耗力，又效率低下；再者，没能在指定备份窗口完成备份，意味着生成数据还存在丢失的风险，故障时间点与备份时间点跨度越长，丢失数据越多；此外，即使是在完整备份数据和定期校验有效恢复条件下，备份数据必须经过恢复方能使用，整个恢复过程缓慢且操作繁琐，这意味着一旦业务数据发生故障，企业由于缺乏应急接管能力，直接面临“业务中断”的风险。

CDM 通过规则驱动自动化备份数据的恢复验证，简化了整个备份数据的验证流程，保证了数据高可恢复性，轻松应对监管需求。同时自动化的恢复演练能够将恢复演练周期缩短至数天，实现恢复演练的简单化。

B 测试数据的管理

随着企业应用数量的快速增加和数据量的持续增长，传统测试数据管理流程成为业务创新的瓶颈。应对测试环境下多个应用、多个功能性测试同步进行，缺少系统化和流线化的测试数据管理平台，导致来自不同应用、多版本、多时间点、数量各异的副本数据，不仅消耗大量的存储资源，还易形成数据孤岛，难以实现集中管控；由于测试资源和管理资源的限制，还使得测试数据难以快速交付和流转，不利于版本的敏捷迭代，以及保证测试数据与生产数据的一致性。

CDM 自动化的流程管理，能够简化测试数据管理复杂度，同时对测试数据的集中管控，提高了数据的管理效率。此外，对不同版本数据的切换管理，满足了日常测试工作的需要，提高了测试数据版本管理的效率。

CDM
典型应用场景

C 脱敏数据的安全可用

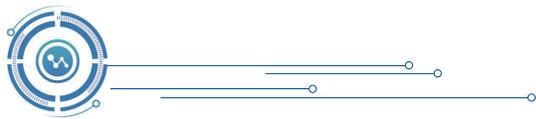
基于生产数据创建的副本数据应用于开发、测试、培训等环境时，由于企业不同机构、部门数据库运维、管理、测试、开发人员的复杂性（内部和外包）、流动性，使得数据面临泄露的风险。例如金融行业，相关监管机构直接提出，用于测试的生产数据要进行脱敏处理，严格防止敏感信息泄露。然而传统的脱敏流程，需要管理员手工 / 脚本修改副本数据，审核后开放给开发、测试环境，不仅数据准备时间长，且数据分发效率低，对于系统运维，传统“一刀切”的脱敏已经无法满足需求，而且进一步增加了副本数据的安全风险。

CDM通过脱敏策略配置实现脱敏源数据的自动化准备，显著减少数据准备时间，基于属性的敏感数据访问控制，通过细粒度的管理，确保数据安全、可控的同时，保证了运维、服务的效率。

D 研发、测试环境数据的构建

测试过程往往需要准备大量的测试数据。测试数据有时可以直接使用历史数据，然而由于基于历史生产数据建立的测试数据可能出现内容缺失、数据量级不够、数据涉密无法导出、数据已加密无法参与计算等情况，这时就需要根据用户的业务需求、数据设置条件、数据间关联等条件，来“造”数据。此外，新系统的研发测试过程，也需要“造”数据。以信创为例，大量国产化系统的研发测试没有历史数据可用，更需要“造”数据来满足需求。

CDM能够根据测试人员的需求“造”出测试使用的数据，实现测试、研发环境下数据的快速构建，提高测试效率，降低测试数据开支。



3.3 中国用户评估 CDM 的五大要素

数字经济和数据隐私及安全的法制化，加速了中国企业级用户对 CDM 的需求。应对数据使用管理的挑战，中国企业级用户评估 CDM 的五大要素为：

① 自主可控

相对于 IT 算力和网络资源，数据治理企业对技术自主可控的要求更高。金融服务、政府、医疗等越来越多的行业在评估数据管理技术时，首选国产操作系统。以对信息安全要求极高的金融行业为例，信创需求更为突出，以安全、可靠、自主可控的国产系统构建金融信创应用体系，成为金融信创应用的重点。

② 安全性

随着应用数量的快速增长，应对不同业务场景对数据服务性能要求的不同，通过 SLA 规则驱动敏捷脱敏和细粒度的权限管理，企业快速根据新的法律法规做安全性调整，保证业务持续合规和安全。

1

2

③ 开放灵活

CDM 技术的部署灵活性，对各种云计算的支撑，以及与各种业务系统、产品的兼容，决定着用户是否需要“推到重来”。以金融行业为例，由于金融信创涉及国内众多厂商，CDM 能否兼容各中国软硬件厂商的产品，助力业务服务创新，直接影响到金融信创体系的顺利构建。

3

④ 自动化

规则驱动实现副本数据多功能、自动化管理、数据获取、备份数据恢复验证、数据交付、智能敏感数据扫描、数据脱敏等操作的流程化，降低敏感数据泄露风险，逐步实现企业级、自动化的统一数据管理。

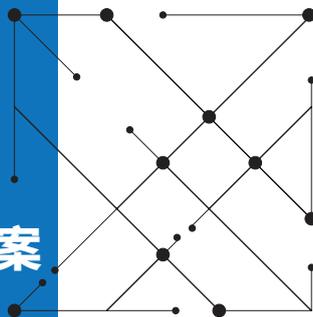
4

5

⑤ 平台化

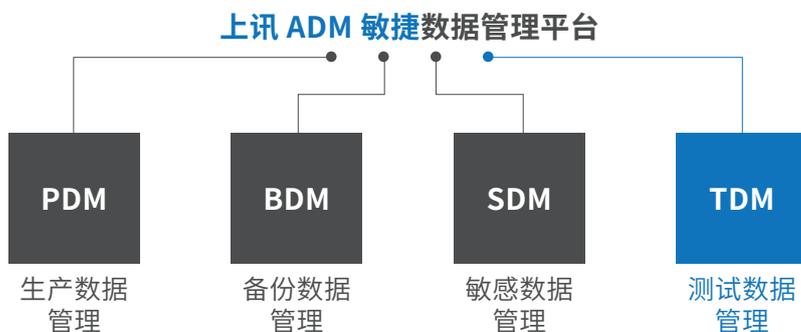
数字经济驱动企业从早期功能分离的数据管理（生产、开发测试、信息管理相对独立）向平台化发展。企业级用户数据管理平台，从实现分钟级恢复，到现代化数据服务（原始数据格式分钟级数据服务），以及满足研发、测试、培训、审计、数据分析等多功能、全流程数据管理需求。

上讯全新的数据管理解决方案 赋能数据降本增效

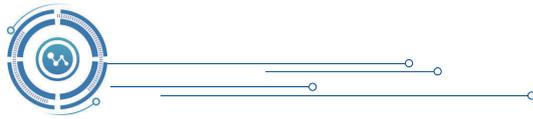


应对企业从以功能为导向到以平台为导向的数据管理需求转变，上海上讯信息技术股份有限公司（以下简称上讯），自主开发推出全新的、整体的数据管理解决方案——敏捷数据管理平台（ADM）。这一综合、统一、自动化的数据管理平台，实现从数据获取、数据存储、数据构建、数据使用，到数据归档、数据销毁的数据闭环式全生命周期管理。统一的副本数据管理，不仅让企业数据管理成本下降 80%-90%，且分钟级的数据交付，支撑多场景应用需求，加快服务和产品推出；同时数据的全生命周期管理，充分释放以前未被利用的数据所蕴藏的价值，推动企业业务的长足发展。

上讯整体数据管理解决方案——ADM 敏捷数据管理平台涵盖以下数据管理方案，即 PDM 生产数据管理解决方案、BDM 备份数据管理解决方案、SDM+DDM 敏感数据管理解决方案、TDM 测试数据管理解决方案。



- * 四大方案都能够以独立功能模块的形式，嵌入到企业的业务系统，企业只需要通过 API 接口连接，即可直接调用各个功能模块，满足备份、恢复、开发测试等各业务场景需求。
- * 企业也可以直接选用上讯完整的 ADM 自动化、闭环式敏捷数据管理平台，以满足各种数据管理需求。



4.1 上讯 ADM 的四大数据管理解决方案

银行业正在加速进行数字化转型以及信创国产化平台替代,实现核心业务的降本增效,同时有效利用数据分析和挖掘,支持业务创新成了当务之急。因此,建设一个高效、高可用性、安全的智能化数据管理平台必不可少。

随着医疗数字化的不断深入发展,HIS、PACS 等关键业务系统已成为医院的“中枢”。一旦 HIS、PACS 等系统宕机,医院的门诊、挂号、收费、影像等业务都将面临中断风险,给医院的正常运营带来巨大损失。如何保障 HIS、PACS 等核心业务应用的高效稳定运行与业务连续性,保障数据安全以及备份数据可恢复性,成为医疗行业面临的首要问题。

当前,汽车制造行业正蓬勃发展,开发测试环节中数据的统计分析,要使用脱敏的生产数据,而传统的脱敏方式由于采用人工脚本,不仅效率低,且数据交付时间长。那么有没有一套完整的解决方案,既能解决生产数据的备份问题,又能安全、自动化地利用脱敏数据的解决方案呢?

上讯 ADM 敏捷数据管理平台可以充分满足各个行业生产、测试等不同应用环境,对数据备份和保护、处理、分析、管理等的全面需求。

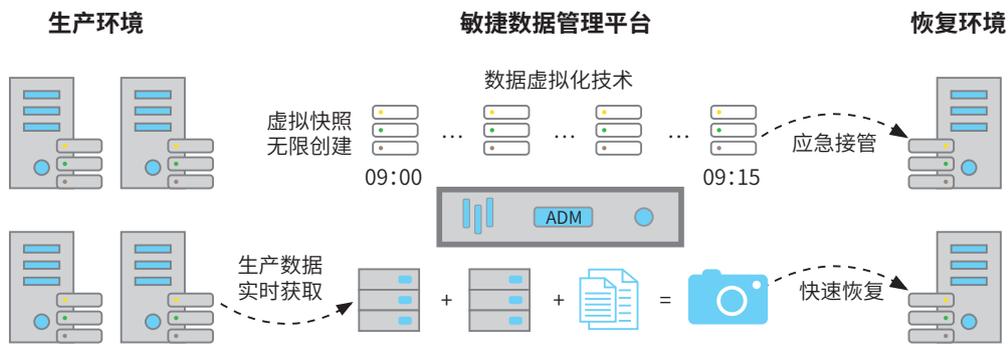


PDM

生产数据管理解决方案 (PDM)

上讯生产数据管理解决方案 (PDM) 解决了生产数据恢复时间长而导致的生产数据不可用, 无法在备份窗口完成备份而导致数据丢失, 以及备份数据一致性带来的挑战。上讯 PDM 基于原始格式的数据复制技术, 能够通过自动化策略设定, 将通过数据一致性检测的副本数据, 安全存储到 ADM 的存储池中, 确保数据的一致性; 全量备份复制, 实现生产数据的实时获取, 简化备份流程, 降低传统备份可能导致的数据丢失风险。一旦业务数据发生故障, PDM 通过生产数据副本能够实现数据的任意时间点恢复, 为生产数据提供分钟级恢复的“应急”保障。

拓扑示意图



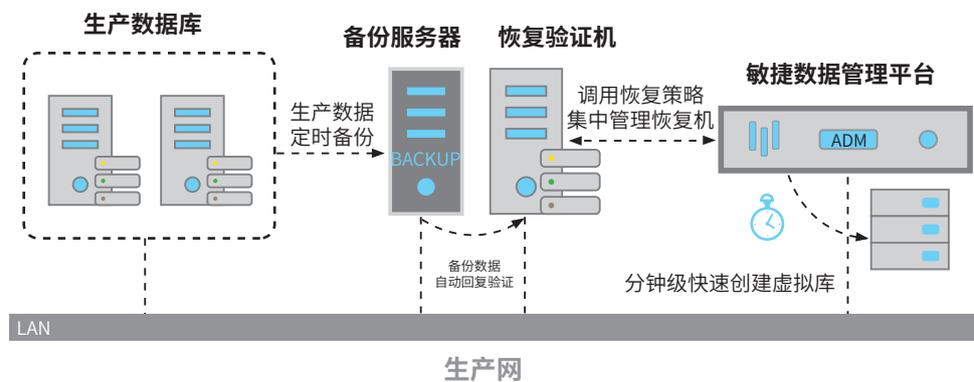


BDM

高效备份数据管理解决方案 (BDM)

上讯 BDM 解决了传统备份流程复杂、缺少校验导致备份数据不可恢复的挑战，提高了备份数据安全管理的能力和水平。BDM 备份数据管理的集中统一化，贯穿数据获取、存储和使用的各个环节，实现闭环式自动化管理，不仅简化了备份管理流程，实现了策略驱动的恢复集中管理，确保了企业级数据的恢复性，而且实现了备份数据有效性验证的自动化，提高用户现有备份系统或新上线备份系统恢复有效性验证的效率，规避人为原因导致的数据备份泄漏风险。

拓扑示意图

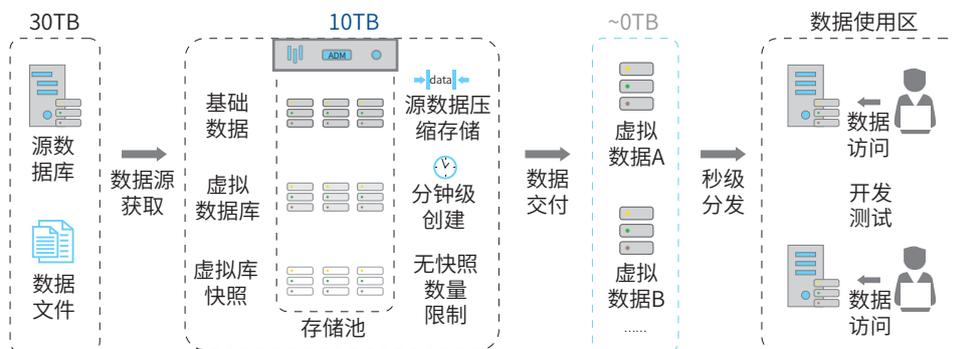


TDM

测试数据管理解决方案 (TDM)

上讯 TDM 测试数据管理解决方案，提供流程化、自动化、集中统一的测试数据管控，在显著减少存储资源占用的同时，实现测试数据的快速交付和流传。TDM 基于虚拟数据库技术，一份数据副本即可在分钟级拉起多份虚拟数据库，不仅降低了物理数据副本导致的副本泛滥，同时显著缩短数据交付周期；数据副本快照的即时挂载、分钟级完成仿真生产环境的部署，加快了应用开发；为了满足版本管理需求，TDM 提供不限量快照版本保留，让用户经济高效地实现测试开发版本的管理。

拓扑示意图

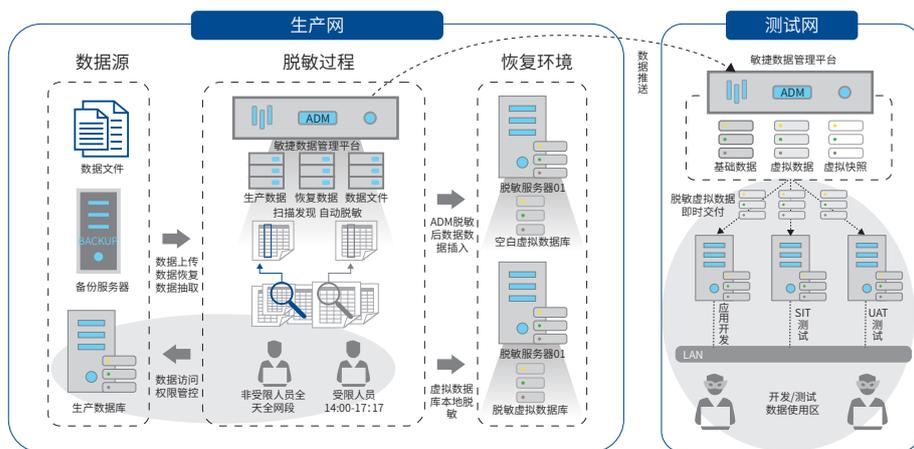


SDM+DDM

敏感数据管理解决方案 (SDM+DDM)

上讯 SDM 敏感数据管理解决方案，结合业内领先的动态脱敏和静态脱敏功能，覆盖敏感数据使用和管理的全部场景。SDM 支持多种敏感数据源，自动扫描发现敏感数据并进行脱敏，确保敏感信息脱敏无遗漏；基于动态稳定因子的随机数算法，保证数据脱敏后原有的数据特征、业务规则和数据关联性，确保脱敏数据的一致性；结合虚拟数据库，脱敏后的数据能够实现秒级分发，从而确保脱敏数据的快速交付。此外，DDM 通过虚拟账户，结合细粒度化访问权限管理，严格控制访问行为，在消除访问隐患的同时，提高数据访问安全的管理水平，降低数据泄露风险。

拓扑示意图



上讯基于数据库虚拟化的综合的 ADM 敏捷数据管理解决方案，让用户可以灵活选择功能模块，快速消除副本数据泛滥，降低存储开支，提高测试、开发、验收等的数据交付管理效率，提升敏感数据防泄露能力。同时，上讯 ADM 提供生产数据实时备份 \ 数据恢复有效性效验和分钟级恢复，保证了生产数据的可恢复性。



4.2 上讯敏捷数据治理平台 (ADM) 核心技术组成



虚拟数据库技术 数据库虚拟化是上讯的核心专利技术之一，可以实现迅捷、经济、高效且多场景的数据服务交付。上讯虚拟数据库通过首次全量复制、周期性增量快照，快速消除副本数据泛滥；同时，可以生成多个具有读写功能的副本，用户可以同步实现多种数据服务交付，例如生产数据恢复效验、多种测试数据交付、实时数据备份等。通过读写分离和动态加速，保证多场景同步数据服务的交付性能要求。

实时增量快照可读写技术 上讯增量快照可读写技术，提供可读写的虚拟数据库副本，让用户只需要一份基础数据源，即可在分钟级内快速交付多份 TB 量级虚拟数据库，显著缩短数据准备时间，满足开发测试、查询分析、合规审计、应急容灾等场景对数据交付效率的要求；此外，实时快照读写技术的无限量快照保留，提高了测试版本的管理效率。



自动化、流程化数据管理平台 ADM 自动化、流程化数据管理平台为用户提供集中管控平台和自助式数据管理，降低 DBA 强度，从而释放资源，推动业务迭代升级；集中的数据使用流程，能够对数据流进行跟踪，实现数据版本的精细化管理；即使在测试环境崩溃的情况下，运用上讯的数据库虚拟化，也能够分钟级实现数据恢复，重新进行测试，从而显著缩短测试的数据准备时间。



跨多数据库敏捷脱敏 ADM 敏感数据管理具有敏感数据自动识别（扫描）功能，防止脱敏不完整，同时结合即时和定时脱敏，静态脱敏和动态脱敏提升测试和生产系统运维对脱敏数据交付质量和管理的效率。同时，仿真脱敏与数据交付为一体，覆盖敏感数据使用和管理的全部场景，通过自定义策略的脱敏，保证脱敏后数据的完整性、仿真性和一致性，满足开发、测试、统计分析等场景对数据脱敏处理的全部需求。

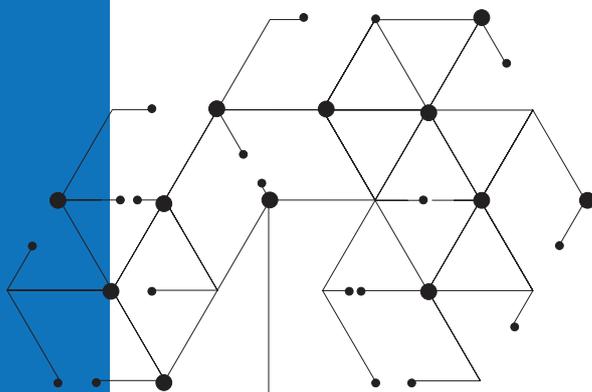
细粒度的权限管理 能够根据用户角色、访问时间、终端 IP 以及数据库访问工具制定个性化数据访问规则，全面控制数据访问操作行为，提高数据的安全管控。



在国家和地方政策的推动下，信创产业得到快速发展，特别是在金融及运营商等信息安全要求极为苛刻的行业，对信创所涉及的基础架构、应用架构进行改造已经成为迫切需求。上讯作为坚持自主研发、产品安全可控的国内企业，持续在应用技术创新、核心技术创新上进行投入，当前在用户环境中实际兼容适配的国产数据库包括巨杉、星环、GoldenDB、OceanBase、浪潮，其他国产数据库及操作系统也在持续兼容中。

5

行业典型案例



金融行业

金融机构尤其是银行对数据的安全合规性要求极高，一方面是监管指引高标准要求，且有严格的日常督查管理制度，另一方面也是由其业务属性决定的。另外，银行在信创方面也走在行业的前端，实现国产化替代的需求十分迫切。随着银行线下业务线上化，从测试管理不当、生产环境暴露，到线上系统延迟、宕机等，都对银行数据的敏捷使用和安全管控提出了严苛要求。

用户：某银行

挑战：

数字经济的发展加大了银行测试开发压力，银行广泛采用外包资源加入测试开发。测试开发团队动则数千人。某银行软件开发中心承担着该银行应用软件的开发、测试和运维等工作。在软件测试开发过程中，测试数据准备是第一步，然而当前该机构面临以下挑战。

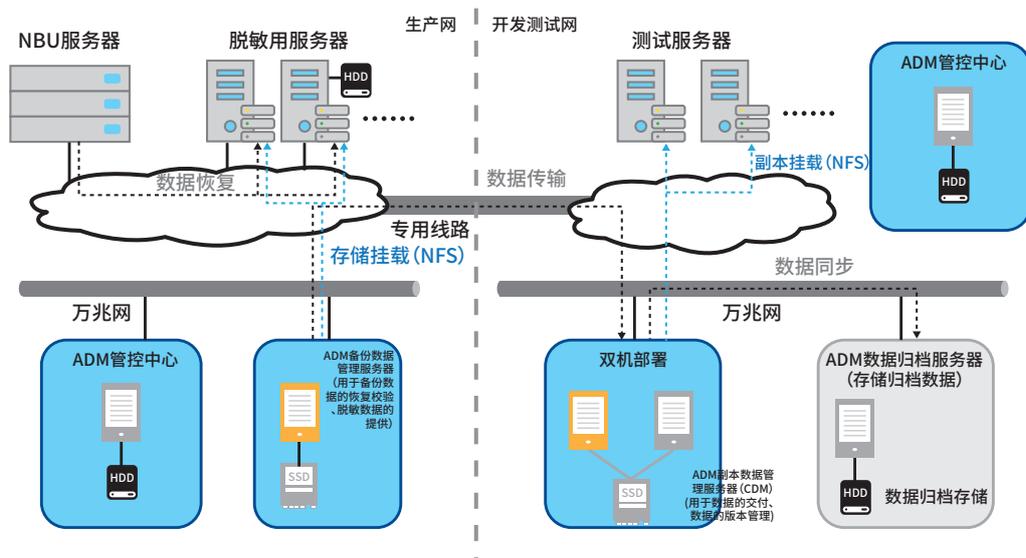
- ◎ **开发测试数据管理、使用效率低：**该银行的测试开发数据，一般从生产环境直接导出，数据准备一般需要4个小时左右，且极易影响生产数据库的正常运行；为满足多应用、多功能的测试需求，测试部门需要频繁进行数据交付，然而测试数据管理和应用的长周期、高开支，成为银行业务创新和迭代优化用户体验的瓶颈。

- ◎ **副本数据泛滥，导致存储空间耗费：**面对测试环境下多场景同步测试的需求，该银行的测试数据环境往往需要创建多个副本数据，这不仅带来大量新增存储开支，还易形成数据孤岛，亟需有效治理来防止副本泛滥。
- ◎ **测试版本管理无法满足业务需求：**早期基于物理数据副本建立测试环境，为了降低已有测试环境开支，无法实现无限副本保留，也大幅度增加了测试版本的管理难度。
- ◎ **测试数据安全性难以保证：**由于测试环境下管理人员、测试需求等不同，存在多份测试数据。然而当前的数据安全，缺乏规范的管理，导致测试数据安全性得不到保障。
- ◎ 在信创的大潮下，该银行希望以国产化的数据管理系统实现安全、可控、高效的数据管理。



方案：

在不改变该银行现有主机和网络架构前提下，某银行在生产和测试网中各部署一套 ADM（管理节点和控制节点分开部署），实现了从生成到测试的自动化，最大限度降低测试数据准备对生产环境的性能压力和安全风险。ADM 集中管控中心（如拓扑图所示）实现任务参数配置、下发任务以及对任务的可视化管理，测试开发人员实现自服务数据服务，降低 DBA 工作强度，保证测试数据管理效率。





价值：

- ◎ **测试数据交付时间减少到 3 分钟，数据交付效率提升近百倍：**ADM 敏捷数据管理平台依托数据库虚拟化技术，能够在 3 分钟内创建完成虚拟数据库，近百倍的数据交付效率提升，让测试部门使用最新的数据进行数据准备成为日常，有效地保证了系统测试的效果。
- ◎ **节省存储空间，降低成本：**ADM 内置的高效压缩存储池，存储比可达 4:1，显著降低数据存储成本；ADM 的虚拟数据库技术，大幅度降低多场景测试下的存储容量需求；此外，一次脱敏、多次分发技术，进一步降低同步测试的数据准备开支。
- ◎ **动态复制提高版本管理效率：**ADM 基于 Golden Image 虚拟数据库，提供不限量快照服务，让用户通过动态快照，提高不同版本的数据管理效率。
- ◎ **测试数据闭环管理，确保数据脱敏和安全使用：**ADM 从数据源头进行管控，规范整个数据管理流程，并结合 ADM 的可视化管理，让管理员轻松掌控所有测试数据状态；可灵活自定义敏感数据类型，并且具备丰富的脱敏算法，可结合具体需求实现静态脱敏或动态脱敏，满足业务使用需求。
- ◎ ADM 兼容国内主流的操作系统和数据库软件，可以满足银行对国产化替代的要求。

上讯 ADM 解决了该银行目前测试开发环境的数据管理问题，降低数据管理开支，提高数据管理自动化水平，保证敏感数据安全合规。Golden Image 虚拟数据库技术保证了生产数据在永远不离开生产系统的前提下，能够安全合规释放生产数据的创新价值。利用 ADM 平台，该银行可以构建从生产、测试开发到质量保证、大数据分析的全自动化数据管理，这种高效、轻松的数据使用和管理，在为用户提供极佳的使用体验的同时，能够让其将更多精力投入到业务创新。

通信行业

随着移动互联网、5G 等的快速部署和应用，如电信运营商等通信行业企业不仅要为庞大数量的终端提供综合应用服务，同时还要搭建应用平台，为行业用户提供服务。此外，还要不断进行新业务的开发，以拓宽业务范围。随着数据的快速增长，内部系统快速开发交付的要求以及外部病毒、黑客、外部设备等众多数据安全挑战，促使通信企业需要从数据管理、数据访问控制、数据脱敏等维度，构建一整套的数据管理流程，保证数据的可用、可控和可信。

用户：某通信公司

在全面上云的战略部署下，某通信公司基于华为云、天翼云和联通云搭建混合多云业务平台，以支撑其生产环境、开发环境和验收环境，满足其各类业务应用系统的部署和运营的同时，还能更好地支持新应用、新业务的敏捷开发和测试需求。目前管理的系统有大约 130 多套，其中包括 Oracle 60 多套、MySQL 40 多套，虚拟机 5000 台左右，总存储量 7PB。在当前的数据管理方面，该用户遇到以下挑战。



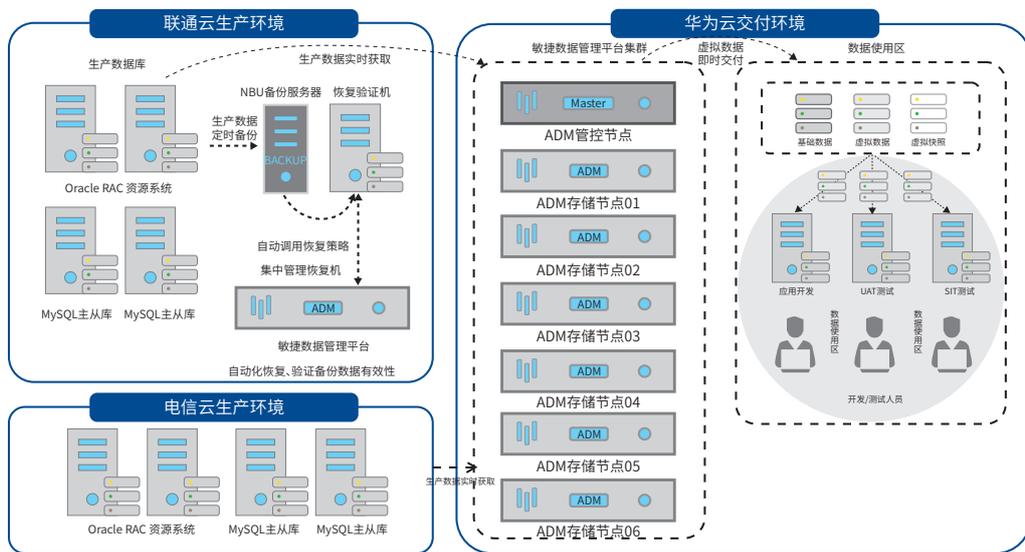
挑战：

- ◎ **测试数据需要高频交付：**在当前 6 人配置下，测试数据准备需要 1 个月的时间。随着开发测试、质量保证等数据需求的不断增加，测试数据准备长周期快速成为业务发展瓶颈；
- ◎ **副本数据分散导致高昂数据开支：**基于多场景数据需求，一套生产系统大约拥有 4 套副本数据，且副本数据分散在本地和云端，亟需构建统一的副本数据管理平台，在为本地和云端数据提供安全防护的同时，降低 IT 成本。
- ◎ **数据可恢复难以确保：**一旦核心系统数据库需要进行数据恢复，传统的基于手动的数据恢复机制，不仅耗人耗时，且可恢复性难以得到保证；



方案：

某通信公司引入上讯的 ADM 敏捷数据管理方案，构建了副本数据集中交付管理平台。借助智能化的副本数据管理和备份数据恢复校验，不仅解决了该公司测试环境下测试数据极速交付的问题，提高了该公司的业务开发和测试效率，同时提高了备份数据的可用性，降低了存储资源开支。



价值：

- ◎ **测试数据交付效率提高 15 倍：**ADM 敏捷数据管理平台依托数据库虚拟化技术，结合“全量 + 增量 + 归档日志”的数据备份策略，能够从云环境实时获取数据，并在几分钟内快速创建虚拟库交付给准生产环境。这种自动化、智能化的数据交付，将原来 6 个人、1 个月的数据准备时间缩减到只需要 2 天时间。
- ◎ **数据存储空间节省一半：**ADM 内置的高效压缩引擎，能够降低约一半的存储空间，并且用于准生产环境的虚拟库，几乎不占用任何存储空间，进一步节省存储开支。
- ◎ **自动化恢复校验确保备份数据高可用：**部署在联通云环境下的 ADM，对接 NetBackup 备份系统，自动化、智能化地完成备份数据的有效性验证，能够将核心数据库的恢复时间从以前的小时级缩短到分钟级，显著提升恢复效率和可靠性。

上讯 ADM 可从容应对 PB 级数据容量，借助数据库虚拟化技术，分钟级创建可追溯到任意时间点的副本数据，显著提升了数据交付效率；同时一整套的数据管理流程，实现了数据从获取、存储到使用的闭环式自动化管理，在降低 DBA 管理强度和复杂性的同时，为 DBA 提供了良好的使用体验。面对 CDM 平台的诸多优势，以及不断增长的数据量，该公司表示，随着后期业务规模的扩大，CDM 平台的扩容复购需求旺盛。

医疗行业

随着智慧医疗产业服务的快速发展，各种数字化系统建设和运营，如电子病历、临床诊疗管理信息系统、居民健康档案等，带来数据类型和数据量的快速增长。与此同时，等保 2.0 提高了医院诊疗全流程数据管理和业务合规的需求。这就对当前医疗行业的业务连续性，以及智能的数据管理提出了新挑战。

用户：某三甲医院

作为综合性三甲医院，某医院承担着辖区以及周边县市 170 多万人的医疗、预防、保健、教学和康复等任务。随着医疗改革的不断深入，以及该医院业务规模的不断拓展，如何管理多个关键性业务系统，包括 HIS、PACS、RIS、LIS、EMR 等的数据安全和高可用，成为该医院的第一要务。



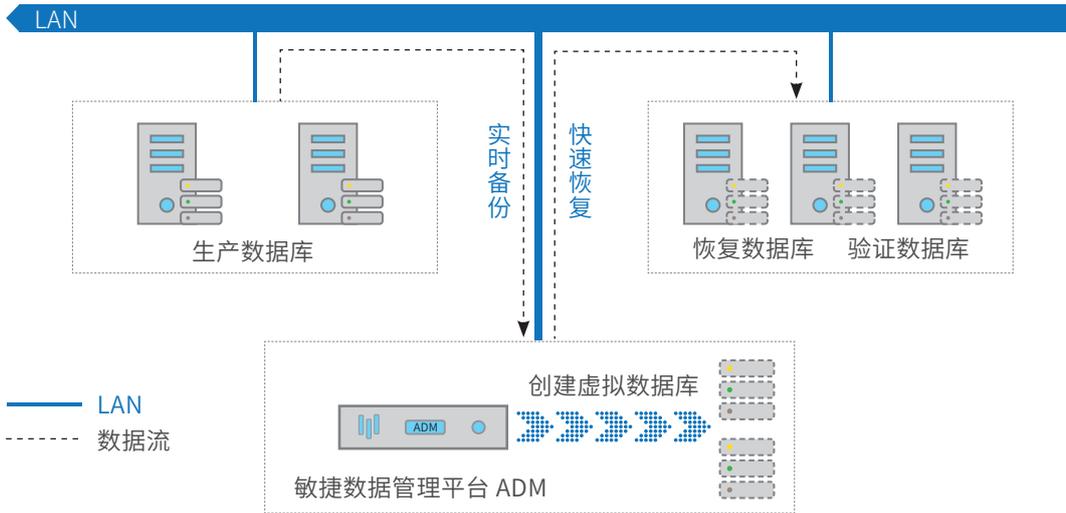
挑战：

- ◎ 关键业务系统数据的持续数据保护，以保证业务连续性，7x24 小时不间断运营。
- ◎ 随着关键业务数据的高速增长，尽量减少额外的存储资源消耗，简化管理。
- ◎ 数据恢复效率不高，集中管控能力弱，难以满足法律法规对于备份数据恢复的有效性验证的要求。
- ◎ 医疗数据十分敏感，且隐私性极强，脱敏处理不可或缺。



方案：

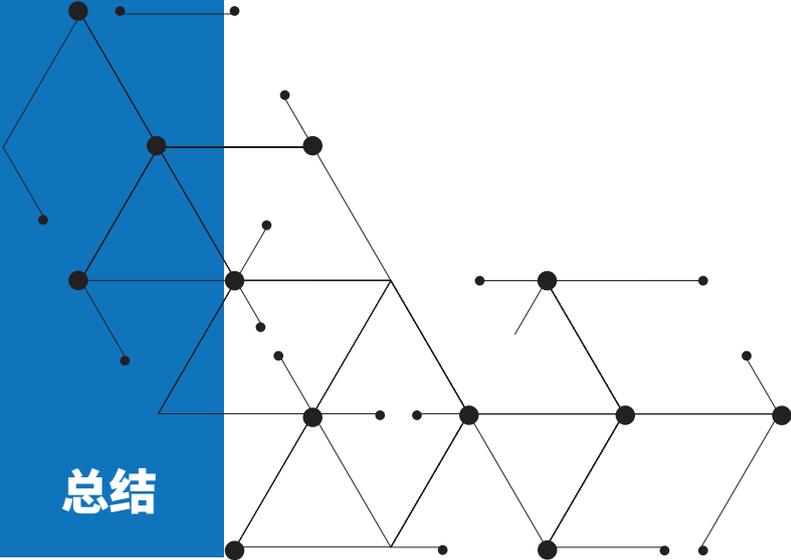
某三甲医院引入了上讯的 ADM 敏捷数据管理方案为业务提供持续性保障。借助数据库虚拟化技术，结合全量 + 增量备份，某三甲医院保证了备份数据的完整性和一致性；同时基于分布式集群架构的管理平台，能够灵活扩展，满足关键业务系统不断增长的存储容量需求。



价值：

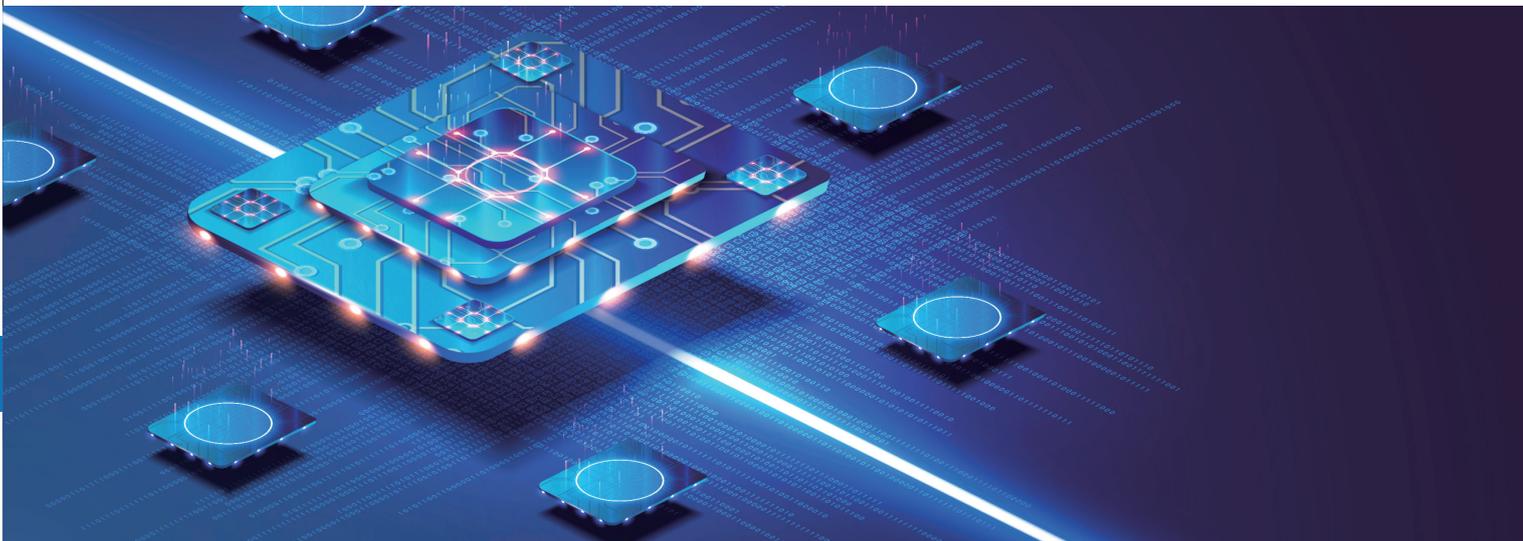
- ◎ **对生产数据实现实时保护，确保业务连续性：** 上讯 ADM 的数据库虚拟化技术，结合源端数据的首次全量 + 周期增量备份，以及实施日志抓取功能，确保了对生产数据的实时保护，进一步保证了数据的完整性和一致性；在业务系统发生故障时，可在分钟级创建虚拟数据库，快速恢复生产业务到最近的时间点，实现应用的快速接管，确保业务连续性。
- ◎ **数据存储空间节省 2/3：** ADM 数据库虚拟化技术，能够有效压缩备份数据，压缩比高达 1:3，从而减少存储空间需求，降低存储成本的同时，简化操作和维护。
- ◎ **数据备份和恢复的集中管控：** 数据恢复验证的自动化，提高备份数据恢复的频率和有效性。
- ◎ **数据脱敏处理，有效保护隐私：** ADM 的拓扑结构较为灵活，可以支持横向扩展，实现敏感信息的静态脱敏，完美实现 HIS 等关键业务系统的数据全生命周期管理。

上讯 ADM 敏捷数据管理平台为该医院带来了关键业务系统的全面安全防护，同时提高了数据恢复的效率，并且显著减少了数据备份恢复过程的存储成本。



总结

以数据为驱动的全新数字化服务模式正快速推动企业的数字化转型，以及推动新产业、新业态、新模式不断涌现。数据成了企业的重要资产和持续创新的驱动力。数据对市场的响应速度和创新的支撑，以及保证企业全流程数据的安全合规，成为评估数据价值的重要指标。随着应用数量快速增加和数据量的持续增长，企业应在以核心生产数据确保企业正常运营的同时，满足应用开发、质量控制、审计、大数据分析、验收、备份等IT场景的需求，让数据创造更多价值，推动数据管理技术和平台的演进升级。然而传统的副本数据管理，不论是数据的可恢复性，还是数据准备、使用和管理，都难以满足企业业务的创新需求。在这种背景下，企业数据管理从以功能为导向转向以平台为导向成了必然。





上讯综合数据管理解决方案ADM,通过功能多元化、部署智能化、流程自动化、管理智能化,让企业快速实现已有数据管理方案升级,逐步过渡到构建敏捷的全流程、综合式管理平台。多功能统一平台可以满足敏捷测试、开发、审计和品控,以及培训、事件分析的需求,加速企业构建系统化和流程化的管理平台,实现数据流线化高效管理。通过虚拟数据库消除副本泛滥;通过规则驱动,实现副本数据的精细化服务管理,以及数据全生命周期的安全合规;分钟级可恢复,更好地实现生产数据的“应急”机制,保证生产数据的高可用。此外,面对信创带来的机遇,上讯将积极践行创新驱动,通过应用技术创新,推动国家信息安全的科技发展。

数据引用:

1. 全球数据圈, IDC
2. <https://www.forbestechcouncil.com/2020/09/25/illuminating-dark-data-in-enterprises/?sh=78d95693c36a>
3. 暗数据: Gartner对于暗数据的定义指在日常商业活动中收集、处理和存储信息资产,但通常不会用于其他目的。
4. Copy Data Management, IDC InfoBrief
5. Copy Data Management, IDC InfoBrief



北京信诺中桥咨询有限公司（Sino-Bridges），成立于 2006 年，专注于 IT 基础设施、IT 架构、新技术等相关领域的调研、咨询和内容创作等服务。致力于为 IT 厂商和 IT 专业人士提供前瞻性、可信赖的市场和技术趋势参考、战略和策略咨询。

所有商标和公司名称是其各自公司的财产。本出版物中包含的信息是由中桥调研咨询认为可靠的来源提供的，但不保证其可靠性。本出版物包含中桥调研咨询的观点，这些观点随时间可能会有所改变。本出版物的版权归中桥调研咨询所有，未经中桥调研咨询的明确许可，不得复制或转载。有问题请联系 contact@sino-bridges.com。网站：www.sino-bridges.com。

