

工商银行大数据技术及应用

白皮书

中国工商银行软件开发中心

2024 年 10 月



版权说明

本白皮书版权属于中国工商银行软件开发中心，并受法律保护。转载、摘编或利用其它方式使用本白皮书文字或者观点的，应注明“来源:中国工商银行软件开发中心”。违反上述声明者，我们将追究其相关法律责任。

目 录

第一章 概述	2
(一) 大数据技术发展历程	2
(二) 大数据产业发展现状	3
(三) 大数据产业演进趋势	9
第二章 工商银行大数据发展现状	13
(一) 大数据平台产品视图	13
(二) 大数据平台总体架构	14
(三) 基于大数据平台打造企业级数据中台	25
第三章 工商银行大数据应用案例	27
(一) 基于流批一体的个人储蓄存款准实时计算方案	27
(二) 基于实时数仓的报送解决方案	28
(三) 内外部数据融合的风控业务模式	29
(四) 数据驱动的码上赢新型营销模式	29
(五) 面向营销经理/运营经理的地理空间信息助手	30
第四章 总结与展望	32

第一章 概述

(一) 大数据技术发展历程

随着大数据技术的不断迭代优化，数据处理速度、存储容量及分析精度均实现了显著提升，这些进展极大地提升了计算能力，引发了对算力需求的显著增长。同时，大数据与人工智能、机器学习等前沿技术的深度融合，正逐步解锁数据的深层价值，为各行各业提供了前所未有的洞察力和决策支持。在这一背景下，金融行业作为数据密集型行业，其大数据业务迎来了前所未有的创新和发展机遇，不仅促进了风险管理、客户洞察、产品创新等方面的精准化与智能化，还加速了金融业态与服务模式的深刻变革，预示着大数据技术在金融领域的未来趋势将更加注重实时性、智能化与个性化服务的提升。

1、萌芽期：传统数仓（2002-2012）

企业需要对数据进行全局性经营分析，由此诞生了一体机架构下的传统数仓。

2、发展期：离线大数据架构（2012-2016）

随着互联网的急速发展，数据规模爆发式增长，一体机架构无力以可接受的成本提供算力，大数据进入开源时代。

3、爆发期：Lambda/Kappa 架构（2016-2021）

大数据逐渐成为业务的重要支撑，业务逐渐产生高时效的数据感知、分析决策、行动和反馈的需求，大数据向实时化演进。

4、转型期（2021-2025）：云原生实时架构

业务复杂度进一步增加，数据源越发多样化，更方便的用数成为业务的主要诉求，大数据开始拥抱云原生、AI 等其他领域。

5、变现期（2025+）：用数智能

数据作为核心生产要素，与企业竞争力直接挂钩，如何将数据价值快速变现成为关注重点。

(二) 大数据产业发展现状

自进入“十四五”时期以来，大数据产业已迈入一个新的发展阶段，产业内部逐渐细化，形成**数据存储与计算、数据管理、数据应用、数据安全**四大核心领域。数据源通过数据存储与计算实现压缩存储和初步加工，通过数据管理提升质量，通过数据应用直接释放价值，并由数据安全技术进行全过程的安全保障。四大核心领域当前呈现不同发展现状，在数据要素大战略的新形势下，发展方向均进一步明确。

大数据产业四大核心领域		
核心领域	发展现状	发展方向
数据存储与计算	实现了海量数据的高效存储、计算	降低运维成本，提升处理效率
数据管理	头部行业实现关键数据的管理	各行业均实现全域数据管理
数据应用	支撑核心业务分析和顶层决策	与全域业务深度融合
数据安全	推进外规内化与风险治理	将安全治理能力嵌入业务

大数据技术在数据存储与计算、数据管理、数据应用、数据安全四大产业领域有不同的发展，**存算分离架构、融合一体衍进、智能增强分析、数据研发运营一体、内生安全技术** 5 项成为关注和发展焦点，国内外市场均涌现不少相关产品和方案。

1. 数据存储与计算领域

【存算分离架构】存算一体架构下，大数据数据处理产品的存储节点、计算节点功能耦合或者部署耦合，存在部署规模上限低、计算弹性扩缩困难、存算故障隔离弱等问题。存算分离的弹性计算、高规模上限、存算故障隔离等架构优势，能够提供更好的算力保障和运维灵活性，有助于应对大规模数据分析的挑战，从而提供更好的数据驱动决策的支持。

近年来，业界存算分离方面的主流产品和样板技术也日益涌现。主流产品方面有 Apache Doris、阿里巴巴 Hologres、腾讯 TCHouse、偶数科技 HashData