

先进存力中心研究报告

(2025 年)

中国信息通信研究院云计算与大数据研究所

2025年8月

版权声明

本报告版权属于中国信息通信研究院，并受法律保护。
转载、摘编或利用其它方式使用本报告文字或者观点的，应
注明“来源：中国信息通信研究院”。违反上述声明者，本院
将追究其相关法律责任。

前 言

在数字经济蓬勃发展的时代背景下，数据作为基础性战略资源和关键生产要素，数据要素价值释放成为推动经济社会高质量发展的核心动力。然而，一方面相较于逐年递增的数据产量，存力规模增速略显滞后，数据“应存未存”现象依然存在。另一方面，数据作为 AI 三大核心要素之一，其规模与质量直接影响 AI 技术应用与发展。在此背景下，先进存力中心应运而生，作为新型信息基础设施的重要组成部分，对于数据资源积累、数据资产转化、数据资本流通以及促进数字经济与实体经济深度融合具有深远意义。

先进存力中心作为一种新型信息基础设施，采用“1+4+N”架构体系，即 1 个支撑海量数据汇聚、存储、管理与流通的先进存力底座，提供数据托管、治理、开发加工与供给流通 4 大服务能力，赋能于 N 个应用场景，实现规模聚数、高效治数、安全供数、产业用数，打通行业数据孤岛、激发数据潜在价值、促进 AI 创新发展。先进存力中心不仅是推动我国 AI 发展的重要举措，也是贯彻落实国家数字经济发展战略的有效途径，有助于我国在全球数据竞争中占据有利地位。

本报告从信息基础设施建设发展的角度出发，探索先进数据存力在数据资源汇聚、数据资产治理以及流通应用等方面的建设模式，为业界提供参考。因时间和能力所限，报告内容若有所疏漏，实属难免，烦请各界不吝指正。如有意见或建议请联系 dceco@caict.ac.cn。

目 录

一、先进存力中心建设背景.....	1
（一）数据成为数字经济发展新阶段战略资源	1
（二）数据留存率与存力建设水平仍有待提升	2
二、先进存力中心定义与内涵.....	3
（一）先进存力中心定义内涵	3
（二）先进存力中心服务能力	5
（三）先进存力中心产业定位	7
（四）先进存力中心运营模式	7
三、先进存力中心建设意义.....	8
（一）多源多模态数据资源高效汇聚	8
（二）助力数据要素价值进一步释放	9
（三）为算力中心提供数据资源保障	10
四、先进存力中心关键技术.....	11
（一）全闪存储架构建设，突破数据应用效能瓶颈	11
（二）数据编织可视可管，打破分散数据孤岛效应	12
（三）虚拟资源智能调度，满足数据高效流通利用	13
（四）数据存储内生安全，保障数据资产高可用性	14
（五）AI 数据湖存储技术，实现多源数据统一归集	15
（六）全域数据分级归类，强化数据资源高效治理	16
（七）构建可信数据空间，促进数据资产供给流通	17
（八）数据挖掘与 AI 赋能，夯实行业数智转型基础	17
五、先进存力中心应用场景.....	18
（一）行业高质量数据集构建，构筑企业全域知识运营基座	19
（二）医疗影像共享空间建设，打造影像数据共享管理中枢	20
（三）智能网联汽车数据管理，促进新能源产业数字化转型	21
（四）公共数据授权运营开放，建设数据要素托管流通枢纽	22
（五）城市全域数据汇聚治理，形成民意感知研判响应闭环	24
六、发展建议.....	25

（一）加强国家顶层规划，统筹先进存力中心建设布局	25
（二）鼓励技术研发攻关，推进存力与算力的协同创新	25
（三）完善存力建设体系，保障基础设施可持续性发展	26
（四）筑牢人才培育根基，支撑数据存力产业生态建设	26

图 目 录

图 1 先进存力中心定义与特征.....	4
图 2 先进存力中心服务能力示意图.....	5
图 3 先进存力中心架构图.....	11

表 目 录

表 1 全国数据资源调查各维度对比.....	3
------------------------	---

一、先进存力中心建设背景

（一）数据成为数字经济发展新阶段战略资源

近年来，我国高度重视数字经济建设，出台一系列政策文件，为数字经济健康发展提供有力的政策支持与战略指引。2022 年 1 月，国务院印发《“十四五”数字经济发展规划》，明确到 2025 年，数字经济核心产业增加值占国内生产总值比重达到 10%，数据要素市场体系初步建立，产业数字化转型迈上新台阶，数字产业化水平显著提升。该规划从优化升级数字基础设施、充分发挥数据要素作用、大力推进产业数字化转型等八个方面对“十四五”时期我国数字经济发展做出总体部署。2023 年 2 月，中共中央、国务院印发《数字中国建设整体布局规划》，提出到 2025 年，数字中国建设取得重要进展，数字基础设施高效联通，数据资源规模和质量加快提升，数据要素价值有效释放，数字经济发展质量效益大幅增强，并强调优化数字化发展环境，建设公平规范的数字治理生态，构建开放共赢的数字领域国际合作格局。这些政策文件的出台，表明我国数字经济发展进入从基础设施建设到产业融合应用，从数据要素市场培育到数字化发展环境创新的全方位、多层次共同推进的全新阶段。

在数据资源价值挖掘与应用方面，国务院于 2022 年发布《关于构建数据基础制度更好发挥数据要素作用的意见（“数据二十条”）》指出数据作为新型生产要素，是数字化、网络化、智能化的基础，充分实现数据要素价值，为深化创新驱动、推动高质量发展、推进国家治理体系和治理能力现代化提供有力支撑。2024 年 11 月起，国家发

改委、国家数据局、中央网信办、工信部、公安部、国资委等连续发布《可信数据空间发展行动计划（2024—2028 年）》《关于促进企业数据资源开发利用的意见》《关于促进数据产业高质量发展的指导意见》，均强调数据对构建数据市场、推动数字经济发展的作用。2025 年 5 月 28 日国务院发布《政务数据共享条例》，要求政府部门建立健全政务数据全过程质量管理体系，推进政务数据共享标准化、规范化，激发数据要素资源价值潜力。此番举措既体现了我国数字经济发展全新阶段对数据要素的高度重视，也标志着数据作为基础性、战略性资源的地位不断加强。

（二）数据留存率与存力建设水平仍有待提升

国家多部委明确存力建设目标，存力建设进入由大体量迈向强韧性、绿色化、高安全的转型阶段。截至 2024 年底，我国存力规模达到 1580EB，先进存储容量占比超过 28%，我国存力建设呈现规模扩张与技术升级并进的良好态势。《算力基础设施高质量发展行动计划》将 2025 年目标锁定在存力规模达到 1800EB、先进存储占比 30%和重点行业灾备覆盖率 100%等方面，其本质上是在规模建设基础上进一步追求质量建设。当前我国存力建设正在加速迈向强韧性、绿色化、高安全的新阶段，后续重点将是在保障具备充足存力资源的前提下，进一步加强与计算、网络 and 应用的协同创新，推进算力基础设施高质量发展，充分发挥存力建设对数字经济的驱动作用。

我国数据生产量逐年增长，存力建设略显滞后，数据仍然存在“应

存未存”现象。据《全国数据资源调查报告》连续两年数据披露，在数据产量与存储容量的规模方面，2023 至 2024 年数据年产量由 32.85 ZB 增至 41.06 ZB，增速达到 25%，数据存储总量由 1.73 ZB 提升至 2.09 ZB，增速为 20.81%，数据存储增量落后于数据生产速度，意味着出现存力缺口。从数据留存与存储空间利用的角度来看，两年新增数据存储量虽从 0.95 ZB 提升至 1.15 ZB，但与数据产量对比分析，数据留存率由 2.89%下滑至 2.80%，同比下降 3.15%，数据“应存未存”现象依然存在。同时，存储空间利用率也从 2023 年的 59%升高至 61%，表明剩余可用存储空间被进一步压缩，详见表 1。综合来看，目前我国存力建设出现规模扩张滞后与有效存储空间受限相互叠加的局面，大量数据无法有效留存。因此，持续强化存力建设已成为避免数据价值大规模流失、保障数字经济稳健运行的关键举措。

表 1 全国数据资源调查各维度对比

	数据生产 总量	数据存储 总量	新增数据 存储量	数据存储 总空间	存储空间 利用率	数据留 存率
2023	32.85ZB	1.73ZB	0.95ZB	2.93ZB	59%	2.89%
2024	41.06ZB	2.09ZB	1.15ZB	3.43ZB	61%	2.80%
增速	24.99%	20.81%	21.05%	17.06%	3.39%	-3.15%

来源：《全国数据资源调查报告》

二、先进存力中心定义与内涵

（一）先进存力中心定义内涵

先进存力中心是集先进存储设备、数据资源管理和数据安全防护

等于一体，融合数据治理、开发流通等功能的新型基础设施，具备高可靠、高性能、高安全、绿色低碳等特征。可提供数据托管、数据治理、数据供给与流通、数据开发与加工等服务能力，实现规模聚数、高效治数、安全供数、产业用数，为大模型训练、智能体开发及数据交易提供高质量、高可信、高可用的数据供给，支撑数字经济创新。未来将与算力中心、网络传输中心协同发展，共同形成新型信息基础设施体系。

具体来讲，先进存力中心利用先进存储技术，实现海量数据高效汇聚，保障数据存储安全；运用数据治理手段，确保数据在存储、处理和应用过程中的一致性与完整性，并完成数据资产转化；通过联合数据供需方构建合规数据交易平台，支持数据在不同主体之间的安全、高效流通；结合数据开发与加工平台，对行业数据进行二次开发，赋能智能化应用。



图 1 先进存力中心定义与特征

（二）先进存力中心服务能力

先进存力中心以“采、存、治、管、用”全生命周期服务能力为核心，打造数据可信托管中心、数据治理中心、数据供给与流通中心、数据开发与加工中心，实现数据资源到数据资产的有效闭环。构建面向政务、行业、企业的数据要素供给枢纽，实现从原始数据汇聚到合规流通的高价值转化。

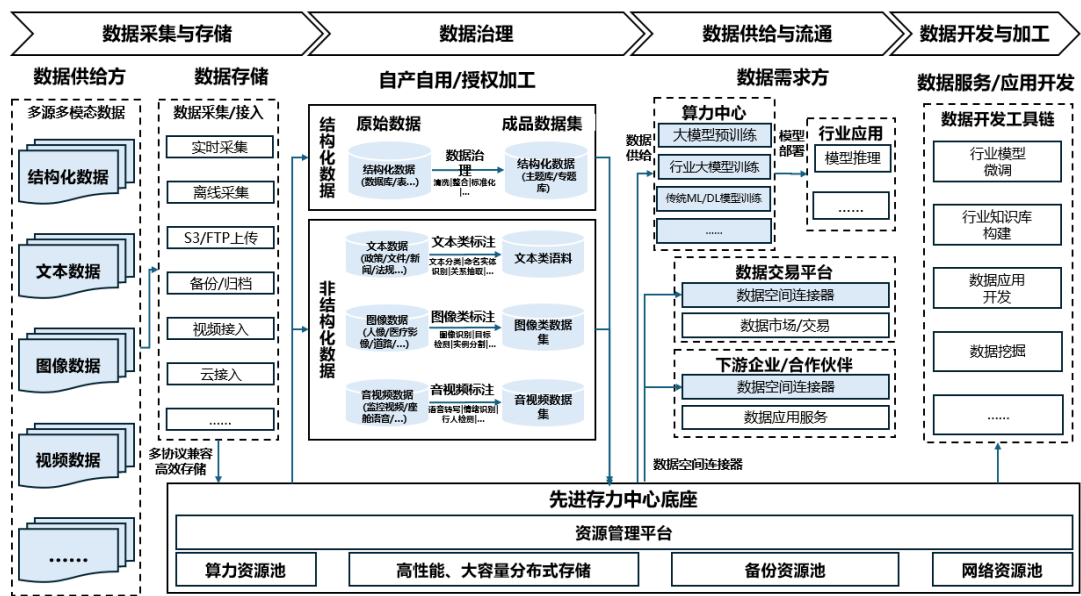


图 2 先进存力中心服务能力示意图

先进存力中心作为数据托管中心，实现跨域数据高效汇聚，保障全局数据可视、可管、可用，完成数据资源的有效积累。在数字经济时代，数据的安全存储和可靠管理是基本要求。先进存力中心通过部署多源数据采集接口，兼容多模态数据传输协议，并提供先进数据存储与管理技术，确保数据在物理和逻辑层面的可靠性，同时结合数据加密、防勒索、防病毒等手段，保障数据的安全性与可用性。不仅可降低数据存储与维护难度，也提升数据管理的效率和质量。

先进存力中心作为数据治理中心，通过统一管理与合规开放，将分散数据转化为可信、可用的公共资产，加速数据资源到数据资产的高效转化。数据治理的关键是将庞杂数据变为可信资产，将静态数据变为活跃服务。通过制定统一规则和自动化流程，对全域数据进行盘点、分级、清洗与编目，形成可检索、可授权、可计量的高价值数据资产。结合标准化接口和可视化工具，将数据封装成可调用的服务应用。同时引入多方安全机制，实现数据使用时，使用方无法看到原始明文内容，且每一次使用都能被完整记录和追溯的安全防护体系。

先进存力中心作为数据供给与流通中心，通过构建软硬协同的可信数据空间，保证数据资产的安全流通，为数据价值释放提供关键保障。数据可信流通是深化数据要素市场化改革的关键环节，先进存力中心通过联合业界多方主体共同建立安全可信的数据交换平台和针对行业应用特点的数据安全保护机制，确保数据在供需方之间高效、安全流通。通过部署加密存储、访问控制、权限分级等存储内生安全技术，有效降低数据在流通过程中的泄露和篡改风险。同时，通过建立数据交易和共享机制，为数据需求方提供高效、稳定、安全的高质量数据供给，促进数据要素市场繁荣发展。

先进存力中心作为数据开发与加工中心，实现将原始数据在本地安全环境内加工为可复用的数据服务。先进存力中心的数据开发体系以“数据留在本地、能力向外辐射”为原则，将汇聚的原始数据转化为可直接支撑决策和服务的智能成果。通过自动化、流水线式的手段对多源数据进行归集与提纯，可用于 AI 模型的行业微调与场景应用。

整个体系可实现数据不出域即可完成训练、测试、发布和运营环节，既消除敏感信息外泄风险，也为数据需求方提供适用于应用场景的数据服务，形成数据价值循环。

（三）先进存力中心产业定位

先进存力中心是新型信息基础设施体系发展的重要方向。目前算力中心作为新型信息基础设施的重要组成部分，主要提供计算资源，而先进存力中心则专注于数据的汇聚、治理、开发加工与流通供给。推进先进存力中心建设不仅有助于充分提升数据存储和管理能力，也有助于实现数据资源与算力资源的协同发展，两者相辅相成，可提升信息基础设施体系的综合效能，共同支撑数字经济的高质量发展。

先进存力中心将与算力中心协同发展，通过数据汇聚、治理、供给流通、开发加工，实现以存汇数，以数促算。先进存力中心通过与通算中心、超算中心、智算中心等现有信息基础设施深度协同，实现数据存储与计算资源的优化配置。先进存力中心不仅是建设汇聚数据要素的载体，也是突破人工智能、数据科学等领域发展瓶颈的有效方式，对带动传统行业数字化、智能化转型有积极影响。

（四）先进存力中心运营模式

在建设实施层面，以地方政府与企业市场化合作共建模式夯实区域数字经济底座。以产业投资控股、城市投资、各地方数据企业出资建设为区域数据产业发展提供坚实的硬件基础，也为数据要素汇聚、存储、流通和应用创造良好条件，可进一步推动地方数字经济蓬勃发

展，助力地方经济转型升级。通过成立合资公司方式联合开展的合作模式，既可充分发挥多方资源优势，也可借助运营公司的专业能力，实现优势互补、互利共赢，推动先进存力中心可持续发展。

在数据资源供需匹配方面，先进存力中心可提供定制化数据供给，同时深度挖掘数据价值，精准满足各行业的数据需求。通过与各类数据要素相关企业联合，精准把握市场对数据存储、处理、分析等方面需求，从而提供有针对性的定制化服务。在供给侧，先进存力中心汇聚行业海量数据资源，经过数据标准化、清洗、安全处理以及数据标注等数据生命周期的治理工作，为各行业提供多样化、高质量的数据资源，满足不同场景下的数据需求，推动数据要素的市场化配置和高效利用。

三、先进存力中心建设意义

（一）多源多模态数据资源高效汇聚

充分积累数据资源，突破数据管理困境，加速推进我国从数字大国迈向数字强国。数字强国建设不仅需要海量数据，更需要高质量的数据管理和应用能力。当前，我国在数据存储和管理方面仍面临诸如多源数据高效采集、海量多模态数据治理、数据安全存储与可信流通等方面挑战。先进存力中心通过联合业界多方主体，深度结合行业数据特点，以自身先进存储技术、数据加密算法和多样化治理手段为基础，实现数据资源汇聚、数据资产转化。

助力构建全国一体化数据市场，实现数据资源优化配置与高效利

用。先进存力中心围绕国家算力枢纽节点、数字经济重点区域建设存力中心，优先建立区域数据体系，促进当地数字经济市场发展，未来形成覆盖全国的中心节点，实现数据跨域流通、数据资源配置优化。因此，建设先进存力中心不仅是数字经济发展过程中的自然演化方向，也是充分发挥我国数据资源优势，推进全国一体化数据市场建设，助力我国在全球数据竞争中占据有利地位，提升我国数据资源利用效率的关键举措。

（二）助力数据要素价值进一步释放

数据资源高效治理，为行业智能化转型提供关键力量。先进存力中心将汇聚的行业数据进行标准化处理和质量把控，确保数据准确性、一致性和完整性，为通用人工智能发展以及行业应用提供丰富、多样的训练语料，推动 AI 技术在各行业的深度应用和创新发展，加速我国人工智能产业赋能。同时可利用数据治理与分析工具，为政府决策、企业运营和社会治理提供数据依据。

AI 数据集构建，为打造人工智能强国夯实数据基础。数据标注是监督学习的重要环节，带有行业应用特点的高质量训练数据集更是推广 AI 技术产业化的重要资源，不仅可提升人工智能模型训练效果，也为模型优化和迭代提供支撑。先进存力中心通过部署数据清洗、标注、特征融合等数据科学工具，可将内部存储的大量行业数据资源转化为用于 AI 训练的高质量数据集，为人工智能模型训练提供精准、丰富的“养料”，提升模型性能和应用效果。同时，可对标注后的训

练数据进行管理和维护，确保数据质量的持续优化和更新。

（三）为算力中心提供数据资源保障

利用高效数据归集和存储能力，为算力资源高效利用提供数据基础。当前，大量的行业数据因未能及时发现其潜在价值而不被重视，也因一些中小企业存储能力不足而未能有效保存，大量有价值的数据未能被妥善保存和利用，逐渐变成“数据废料”。先进存力中心不仅可为产业界的海量数据提供高可靠、高性能、高安全、低成本的数据归集和存储空间，也可利用数据管理手段，确保数据的完整性和可用性，将海量数据的存储优势对外辐射，推进数据资源行业应用与流通，为算力资源高效利用提供数据保障。

先进存力中心以存算资源协同调度为算力中心提供低延迟、高吞吐的数据流，提升资源利用率。在涉及数据流任务启动前，先进存力中心根据任务特征提前完成数据格式转换与副本优化，并通过高速网络将数据处理为可支持挂载的远程存储卷或可下载的对象存储文件，为算力中心提供标准化数据服务。在任务运行阶段，存力中心通过预置副本策略降低跨域访问延迟，所有数据交互均基于开放协议与标准 API 实现，从而为算力资源提供高可靠、可扩展、低延时的数据支持。由此将先进存力中心从“先存后算”的数据资源仓库升级为与算力中心任务耦合的外部存储与数据资源调度平台，共同构筑高效、弹性、绿色的新一代信息基础设施。

四、先进存力中心关键技术

先进存力中心技术架构分为物理层与应用层，物理层以先进介质、存算分离、数据编织构建弹性 AI 数据湖存储池，向下屏蔽多源异构差异，向上提供可按需伸缩的存力资源，夯实先进存力底座。功能层以内生安全、全域治理、可信流通、开发加工四大能力域为核心，打通数据从汇聚到价值释放的完整链路。

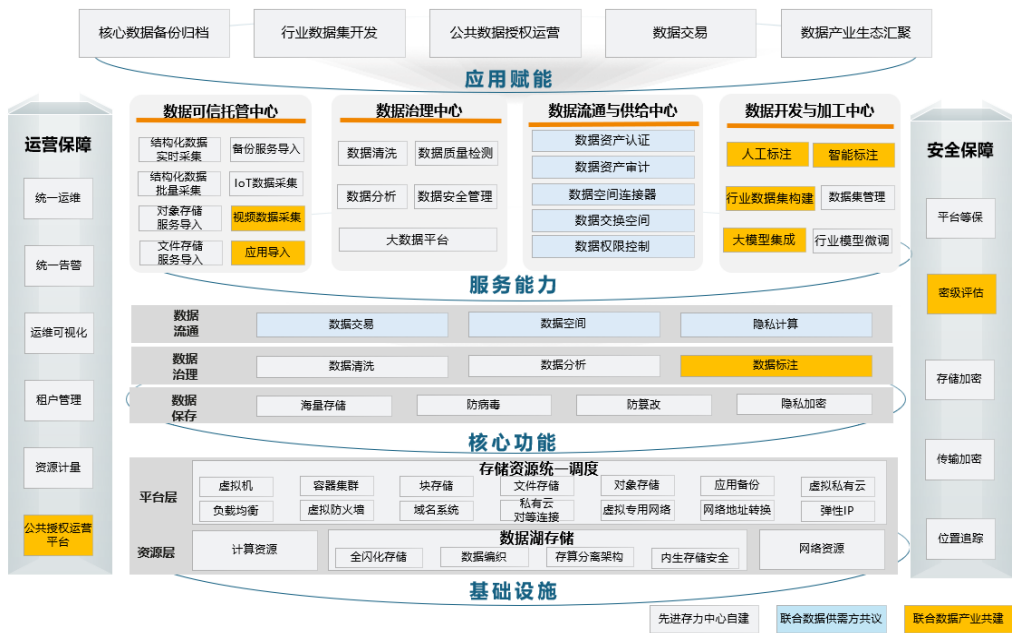


图 3 先进存力中心架构图

（一）全闪存储架构建设，突破数据应用效能瓶颈

闪存介质结合先进算法，提供高可靠、高性能、高效率的存储方案。全闪化建设具有显著的高性能、低延迟和高可靠性等特点，适用于高并发、大规模数据处理等要求严格的应用场景。同时，闪存的低延迟特性可显著加速数据检索和处理过程，为实时数据处理提供有力支持。在耐用性和稳定性方面，结合数据压缩与去重技术的应用，可

进一步提高存储容量的利用率，降低存储成本。

加速数据分析与 AI 训练，满足全业务高性能存储需求。在当前业务支撑方面，全闪存建设展现出多方面显著优势，目前已经被金融、运营商、制造、医疗等行业广泛采纳。对于金融、电信等行业的核心业务，全闪存架构能够满足行业对存储读写性能、稳定性、安全性和业务连续性的高要求，确保关键业务的高效运行。此外，全闪存系统能够大幅提升数据分析任务的执行速度，为企业的数据驱动决策提供更快速的支持。面向人工智能大模型训练和推理场景，全闪存系统能够满足其对海量数据高效读写的要求，确保数据资源的高效供给。

优化资源利用，助力先进存力中心绿色节能与可持续发展。全闪存系统能够高效支持多样化的数据处理需求，加速混合存储负载的处理，提升先进存力中心的综合性能。通过数据压缩和去重技术，全闪存系统优化了存储资源利用率，节省存储空间并降低成本。在能耗方面，全闪介质具备更高的能效比，可帮助企业减少存储系统节点数量，降低存储服务器部署和运行成本，减少空间占用并节约能耗，符合先进存力中心绿色节能和可持续发展的建设目标。

（二）数据编织可视可管，打破分散数据孤岛效应

数据编织与数据湖技术结合，共同构建数据管理和治理生态系统。数据编织是将数据通过集中化数据访问，实现多源数据的治理、审计、追踪、安全和监控的统一管理，支持多云、混合云和边缘部署，为数据提供统一的业务语义和安全性保障。数据湖提供大规模、低成本、

灵活的数据存储和处理能力，支持多种数据类型（结构化、半结构化、非结构化）存储，保留数据原生格式，提升数据应用的灵活性和敏捷性。这种结合不仅可优化数据存储和处理，也提升了数据的可理解性和可用性，确保数据的质量和一致性。

湖仓一体架构与统一接口治理，提高数据利用效率。基于数据湖为先进存力中心提供的底层数据存储和处理能力，结合数据仓库形成湖仓一体架构，提升数据调用能力与业务效率。额外利用数据编织实现统一的数据访问接口和治理功能，可优化数据管理流程，提供灵活的数据访问和管理能力，确保数据的质量和一致性。

（三）虚拟资源智能调度，满足数据高效流通利用

硬件资源虚拟化管理，实现先进存力中心资源高效整合与灵活调度。在先进存力中心的构建中，硬件资源虚拟化管理是实现算力资源高效利用的关键技术。通过资源虚拟化，将先进存力中心计算资源、存储资源和网络资源进行抽象化处理，形成统一资源池，实现资源的动态分配和灵活调度，可根据业务需求快速调整资源分配，确保资源高效利用。同时，对硬件资源进行全生命周期管理，从资源的接入、分配、监控到回收，实现智能化的运维管理，确保硬件资源的高效稳定运行。此外，虚拟化资源具备良好的扩展性，支持从小规模存储节点到大规模集群的平滑扩展，满足业务的持续增长需求，为先进存力中心的可持续发展提供基础。

数据资源智能化运维，为保障数据高效供给提供核心支撑。数据

资源运维是先进存力中心发挥效能的另一关键环节。通过数据融合技术实现多种协议共享，将块存储和文件存储统一管理，使得数据无需搬迁即可高效访问，提升数据调用效率。在数据治理方面覆盖数据清洗、标注、整合等全流程，将数据转化为高价值数据资产，满足算力资源对数据的高质量需求。运维管理方面搭载智能运维系统，可基于风险控制策略进行统一管控，快速定位并解决问题，确保数据资源的稳定供应。

（四）数据存储内生安全，保障数据资产高可用性

抵抗数据存储风险，确保企业核心数据资产安全。先进存力中心通过全链路加密与密钥闭环管理实现数据存储安全可审计、不可逆、零泄露。在保障数据存储安全维度，先进存力中心以数据写入加密、存储落盘加密、密钥分管、即时销毁的安全机制，将每一次数据的增删改查过程都纳入可审计闭环，通过软件层与硬件层的双重密文转换不仅割裂了明文与存储介质的直接联系，也使得密钥的每一次更新同步触发旧密文的不可恢复性，从而在无需业务系统感知的前提下，实现安全策略的自动化、不可逆和全域可追溯，构建覆盖数据全生命周期的端到端防护能力，筑牢数据资产安全防线。

关键数据秒级恢复，保障核心业务连续性。先进存力中心采用跨地域加密副本与秒级密钥同步确保灾备恢复时业务连续性与数据完整性。在关键数据灾备恢复维度，中心通过跨地域副本与实时日志协同，将加密后的数据连同密钥状态一并同步到远端隔离区，实现任何

主中心级别故障都可秒级切换到副本节点，密钥完整性与一致性由专用通道在灾备切换过程中同步完成校验，确保恢复过程既满足合规保密要求，又保持时间窗口最小化的业务连续，使数据资产在灾备场景下依然保持同等级别的可信水准。

（五）AI 数据湖存储技术，实现多源数据统一归集

将多源异构数据统一采集、汇聚与预处理，形成可治理、可授权、可流通的 AI 数据湖。AI 数据湖存储平台通过批量同步与实时接入双通道，将结构化、半结构化及非结构化数据由多源节点引入，实现异构数据源的无缝对接与增量更新。在数据入湖前，系统依据统一的数据标准对表、字段、格式进行标准化映射，并同步完成敏感字段识别与脱敏标记，从源头确保数据质量与合规性。归集后的原始数据以分区、分表形式落地于文件存储系统与数据库构成的湖仓一体架构，既可满足高吞吐的历史数据存储，又支持低延迟的实时查询需求，为后续数据治理与授权加工奠定高可靠、高可用的资源基础。

元数据自动捕获、注册与血缘构建，以统一数据目录实现“数据—元数据—目录”一体化管理。在数据归集过程中，中心将同步完成元数据的自动捕获、注册与血缘构建，实现“数据—元数据—目录”三者的一体化管理。元数据管理模块通过正向建模与接口注入两种方式，将数据的逻辑模型、物理模型、业务标签、生命周期策略等信息集中注册至统一目录，形成面向数据需求方的可视化数据视图。数据目录作为归集成果的统一出口，可按照主题、业务域、安全分级等多

维方式呈现，使数据需求方能够快速定位检索。通过上述数据归集与目录化机制，先进存力中心不仅实现海量数据的高效汇聚，更为后续授权加工、服务封装及安全流通提供可追踪、可审计的完整链路。

（六）全域数据分级归类，强化数据资源高效治理

先进存力中心以“规范—加工—管控”为主线，将原始数据转化为可授权、可流通的高质量数据集。先进存力中心的数据治理体系通过一站式治理平台将原始数据转化为符合法律法规、可授权流通的高质量数据集。依托数据规范模块构建统一的数据标准、数据模型与数据分层分类体系，以关系建模和维度建模驱动，实现数据从归集层到标准层的逐级提炼，随后通过数据批处理、流处理及策略编排的自动化加工，确保数据在抽取、转换、加载全过程中的准确性、一致性与可溯源性。治理过程同步生成的元数据与数据资产信息被实时注册至统一目录，形成面向业务、面向监管的完整数据链路可视化，为后续授权、开放、共享提供可信依据。

通过细粒度权限控制、动态脱敏与分级分类管理，在保障数据安全与合规的前提下，实现数据价值深度释放。经过治理的数据资源将支持表级、字段级的访问授权，可对不同敏感等级的数据配置差异化脱敏策略与加密算法，并结合如水印、审计日志、防盗链、多因子认证等多重安全机制，确保数据在存储、传输、使用全生命周期内的可追溯、可审计、可防御。在此基础上，数据资产目录会呈现业务口径、字段含义及样例数据，使数据需求方能够“可见即可理解、可申即可

使用”，从而支撑普惠金融、信易贷等场景的快速创新与规模化落地。

（七）构建可信数据空间，促进数据资产供给流通

先进存力中心通过搭建数据交易平台，与多方主体联合制定数据跨域流通的安全范式。面向数据流通场景，先进存力中心搭载控制中台、数据连接器、审计清算平台、认证平台、空间管理平台等工具，使数据在供给方与需求方之间形成端到端的受控通道。借助数据权限组件完成身份、属性、策略的统一编排，兼容企业既有账号体系并支持邮箱、手机等终端轻量级接入，既保证高安全等级的身份核验，又兼顾数据服务快速上线的业务需求，从而让数据供给方、数据授权方、数据需求方在同一套技术体系下被清晰定义、实时校验与持续生效。

保障数据可用且全程自证可信，为先进存力中心成为数据要素供给枢纽奠定可行性基础。在数据流通追溯层面，先进存力中心将数据约束细化为可解析的策略指令，并叠加数字探针与区块链日志实现“事前-事中-事后”全链路可证。将时间、次数、地域、资质等多维条件抽象为左操作数、右操作数的量化表达式，再映射到允许、禁止、义务三类动作，形成覆盖使用、交换、订阅、资质四大策略域的动态策略树。将数字探针嵌入数据流，实时捕获元数据、内容指纹与行为日志，并同步上链固化，既支撑运营方对数据调用的清算，也满足需求方的自证合规。

（八）数据挖掘与 AI 赋能，夯实行业数智转型基础

先进存力中心以数据采集、开发、应用流程打通行业知识零门槛、

私有化、可持续交付的完整闭环。先进存力中心将数据开发与加工能力沉淀为一条端到端的本地化数据价值链，多模态原始数据可在数据应用平台中完成格式归集、语义清洗、版本管理与质量评估，形成高质量语料，这些语料随即流入数据需求方提供的模型中，在基础模型训练与分布式推理框架下完成行业大模型面向特定应用场景的参数微调，并通过模型仓库与自动评测实现知识资产化。最终，利用行业知识 RAG 编排与私域安全部署功能将模型能力封装为本地化行业应用，开发者可在数据本地完成应用创建、调测、发布与监控流程。

先进存力中心以资源一体化与安全治理一体化双轮驱动，实现高并发、多场景、合规可控的数据加工与模型孵化。先进存力中心通过统一资源池与弹性调度架构，将算力、存储与网络资源抽象为可编排的服务单元，使数据加工、模型微调及应用推理共享同一套资源底座，既避免重复建设，又实现性能与成本的最优平衡。同时，平台内置的安全沙箱与访问控制机制确保数据、模型及接口在任何环节均处于可监管、可隔离、可撤销的受控状态，实现多租户并行开发与迭代而不产生交叉泄露风险。

五、先进存力中心应用场景

先进存力中心在应用过程中形成以数据汇聚、存储、管理、治理为核心能力，以数据流通供给、开发加工为辅助功能的协同应用范式，利用先进存储技术实时汇聚多源异构数据，经内生加密完成高效存储。并联合多方数据产业主体进行按需标注、治理、开发等，形成高可信

数据资产。通过与数据供需方共议，搭建数据安全可信的访问渠道，实现数据跨主体安全共享，并利用容器化开发工具与可计量接口，支撑数据二次加工与合规流通。目前先进存力中心已在制造业转型、个人医疗影像空间构建、网联汽车数据管理、公共数据授权运营、城市数据汇聚治理等场景应用落地。

（一）行业高质量数据集构建，构筑企业全域知识运营基座

先进存力中心在企业智改数转进程中主要应用体现在全域非结构化数据的统一运营。以制造业为例，通过持续汇聚研发设计、工艺文件、质量记录、设备日志、供应链影像、售后资料等多源异构内容，形成企业级知识库。先进存力中心以全闪化存储为核心，实现微秒级延迟、百万级 IOPS，同时以冷热分级、加密多副本机制保障海量文件长期可靠留存并兼顾高并发实时调阅。依托内置的内容管理框架，对文件进行元数据自动抽取、语义标签生成、版本与权限模型绑定，实现从原始资料到可检索、可复用、可审计的企业“知识资产”的分钟级转化。针对企业“不敢转、不会转”的痛点，先进存力中心提供零代码插件式能力，业务系统通过标准化接口即可调用存取、转换、预览、标注、审批等服务，无需重复建设存储与安全模块。同时将文档管理、知识图谱、安全外发、电子档案、质量体系等业务组件封装为可插拔微服务，支持按产线、按车间、按项目快速部署，满足企业多租户隔离及分区域容灾需求。内容安全体系贯穿采集、存储、流转、外发全链路，动态水印、透明加密、敏感词扫描、智能定密、离线外

发管控等手段确保核心工艺与商业机密在内外部协同过程中始终处于受控状态。此外，可通过用户行为与文件语义的双维度分析，实时生成知识热点图、协作网络图、风险预警图，为企业管理层提供研发效率、质量缺陷、知识沉淀、合规风险的量化视图，进而驱动业务流程优化与商业决策升级。

先进存力中心将企业中原本散落在个人终端、业务系统、纸质档案中的非结构化数据转化为可共享、可增值、可持续迭代的高质量数据集，为企业从自动化到智能化、从单点应用到生态协同的跃迁提供坚实的数据底座与持续的创新动能。

（二）医疗影像共享空间建设，打造影像数据共享管理中枢

先进存力中心在医疗影像领域的核心功能聚焦于海量医学影像数据的高效存储与管理，同时在远程医疗、人工智能辅助诊断等多方面发挥作用，贯穿影像数据从采集、存储、共享到利用的全流程。在海量影像数据存储方面，通过对接区域内各级医疗机构的影像采集系统，将分散在医院内部、医学研究工作站及第三方影像中心的原始影像数据以标准化格式持续汇入，形成医疗影像资源池，结合内置的冷热分级机制与加密存储策略，确保数据在长期保存过程中既保持高性能在线访问能力，也可满足低成本归档与合规留痕要求，为医学影像数据安全、高效存储提供坚实基础。在远程医疗协同方面，中心在取得数据提供方授权后，跨机构医生可在先进存力中心存储的医疗影像

数据副本中进行并发诊断，无需重复上传数据即可发起云端会诊，为远程医疗高效开展提供数据支撑。在医学教学与培训场景中，中心将经过脱敏与结构化处理的历史影像转化为可检索的教学素材，支持院校与住培基地在线开展虚拟阅片、病例考核与能力评估，形成持续更新的影像教学资源库。在科研场景中，平台以专病维度对影像数据进行标签化组织，向医院、高校、药械企业及人工智能研发机构提供研究级数据集，并配套数据清洗、匿名化、格式转换工具，显著缩短科研前期数据准备周期。

通过以上功能的系统化集成，先进存力中心将原本分散、封闭的医疗影像数据转化为安全、合规、可持续运营的区域级数字资产，为远程医疗、患者服务、医学教育、科研创新等应用场景提供统一、可信、高效的医疗影像数据底座。

（三）智能网联汽车数据管理，促进新能源产业数字化转型

先进存力中心在智能网联汽车领域应用聚焦车辆从出厂到退役全周期所产生的高频运行、电池、充电、用户行为等多源异构数据的汇聚、治理与价值释放。平台存储池承载持续涌入的车端报文、日志、视频、图片等非结构化数据，通过冷热分级、加密切片、多副本及国密算法，在保障秒级回传实时写入的同时实现低成本长期归档与合规留痕。数据落盘即触发完整性、真实性、及时性校验，异常样本自动隔离并生成质量报告，确保后续分析与监管可信。监管侧依托“数字

孪生车辆”视图，监管方可按区域、品牌、车型、电池厂商等多维视角实时监测车辆运行态势、故障及安全预警，预警模型基于历史缺陷库与实时特征流持续迭代，风险信号一键下发至车企、充电运营商与应急响应单位，实现安全隐患早识别、早处置。企业侧通过数据服务管理平台将脱敏车况、驾驶行为、电池健康指数封装为标准商品，支持主机厂、保险公司、金融机构在线订阅、调用与再加工，助力精准保险定价、残值评估及金融产品创新。账户、白名单、交易对账、支付结算均以微服务开放，保证运营安全、合规。面向车主，提供个人驾驶画像与车辆健康报告，异常耗电、潜在故障、保养建议通过用户终端推送，授权共享数据可无缝衔接换电、维修、二手车交易等服务。在科研与孵化环节，中心以容器化组件库开放实时聚合分析库、特征加工工具、模型镜像及 API 网关，高校与算法公司可快速构建电池健康预测、充电负荷调度、长尾场景挖掘等应用，实现数据、算力、算法即取即用。

先进存力中心可将车端分散数据转化为统一、可持续运营的数字资产，为政府监管、企业研发、金融创新提供可信底座，并为车主创造安全、便捷、经济的出行体验，形成新能源汽车产业数字化转型的闭环生态。

（四）公共数据授权运营开放，建设数据要素托管流通枢纽

公共数据授权运营场景以数据要素全生命周期价值释放为主线，

贯通公共数据从采集、存储、治理、开发、定价、交易、应用全过程，先进存力中心在这一链条中被定位为“端到端”数据托管流通枢纽，凭借高性能存储底座、数据国密加密，实现公共数据长期、可靠、绿色保存，并通过可信数据空间与隐私计算确保跨主体流通。开发统一门户入口，面向数据需求方提供数据集、数据产品、开发工具、算力资源的分类检索和在线订阅，需求缺口径授权后直接在中心数据湖仓、标注平台、数据加工，再上架完成闭环。运营管理后台则依托微服务中台，将目录管理、资产确权、合规审核、订单管理、计量计费、风险监控、监管报送等功能封装成可插拔的服务模块，分别向平台运营方、场景运营方、数据提供方、数据服务方和数据监管方开放，支持按角色视角的实时运营可视和全链路追溯。运营方通过大屏实时掌握硬件资源、数据湖仓容量、工具使用、商品交易情况。场景运营方在开发中心完成数据抽取、清洗、建模和封装后，由审核中心完成从开发环境到生产环境的注册、上线审批。数据提供方在供应中心完成目录挂接和资产卡片维护，确保权属清晰、属性完整。数据服务方调用资产估值、质量评估等能力，支撑数据资产入表、质押融资等增值服务。监管方则通过监管视图实时查看授权、订阅、日志、链上存证，实现公共数据授权运营的在线化、透明化和可审计。

整个过程既保障公共数据长期、可靠、低成本保存，又可通过可信数据空间和隐私计算技术确保数据跨主体流通过程中安全可控，形成公共数据要素市场化配置的可持续生态。

（五）城市全域数据汇聚治理，形成民意感知研判响应闭环

先进存力中心在城市数据治理场景中，以全域、多模态、实时化的数据获取与治理体系为支撑，依托高性能计算与智能分析，实现对城市运行态势的精准感知、深度分析与闭环处置。系统通过 API 接口、OCR 识别、音视频解析等技术手段，对社交媒体公开评论、政府服务热线、线上投诉平台及线下服务窗口等多源异构数据进行全天候采集。在数据处理环节，采用标准化清洗、冗余过滤、关键字段标注及知识图谱关联等方法，将文本、语音、视频等多模态信息统一转化为结构化数据，并与政策法规库进行语义映射，显著提升数据治理效率与准确性。在数据分析层面，基于大模型的提示工程与监督微调机制，对市民反馈进行事件类型识别、地理位置定位、情绪倾向判断及合法性校验，结合机器学习模型实现民生需求趋势预测与实时舆情预警，最终结果通过可视化数据大屏动态呈现，为城市管理者提供科学决策依据，形成“问题发现、诊断、解决、反馈”的完整闭环，切实提升公共服务质量与城市治理现代化水平。

通过这一闭环，先进存力中心将城市零散的民意信息转化为可执行、可复盘、可持续改进的治理指令，确保民意感知无盲区，保障政策落地有依据，成为城市治理现代化进程中的重要数字底座。

六、发展建议

（一）加强国家顶层规划，统筹先进存力中心建设布局

在国家层面制定统一的先进存力建设规划，面向未来对存力规模、先进存储占比、核心数据灾备能力等提出中长期目标，为各地区制定本地存力产业规划提供参考框架。结合国家算力基础设施总体布局，统筹考虑东西部资源禀赋差异，支持东部地区通过技术升级、绿色改造提升存力资源利用效率，引导西部地区利用能源、土地和气候优势参与数据存储资源优化配置，形成优势互补、协同发展的全国一体化建设格局。同时，建议依托现有算力基础设施标准体系，适时启动先进存力中心建设规范研究，推动产业有序、规范发展。

（二）鼓励技术研发攻关，推进存力与算力的协同创新

聚焦闪存技术、架构创新、AI 技术融合、存储内生安全等关键技术，加大研发投入，突破技术瓶颈，提升先进存力中心的性能与效率。鼓励企业与科研机构合作，加速技术成果转化，加强存力与算力的协同发展，探索如量子存储、DNA 存储等先进存储技术应用，打造存算一体化示范项目，提升数据全生命周期管理效率，推动存储技术的持续创新与应用。

（三）完善存力建设体系，保障基础设施可持续性发展

建议围绕先进存力中心的建设、运维、应用全生命周期制定设施级管理体系。鼓励从规划、建设到运营阶段逐步引入先进存力应用、数据质量、资源可用性、灾备覆盖率以及存力中心能耗、碳排等关键指标，并探索建立第三方评估与持续监测机制。同时支持行业开展存储介质绿色回收、数据清除、灾备演练等共性需求开展技术研究，降低环境与数据泄露风险，为先进存力中心长期稳定运行和绿色低碳发展提供制度保障，为后续数据治理与流通提供可靠的基础设施支撑。

（四）筑牢人才培养根基，支撑数据存力产业生态建设

围绕先进存力中心技术和产业链需求，完善数字人才评价与培养体系，形成需求牵引、多元供给、有序流动的培育环境。鼓励各地在重大项目、示范工程中引导高校、职业院校、行业组织与企业共同培养适应数据存储产业发展需求的技术骨干与复合型人才。同时，借助职业资格评价、技能竞赛、国际交流等方式，逐步完善优秀队伍挖掘与激励机制，为数字经济高质量发展提供坚实的人力支撑。

中国信息通信研究院 云计算与大数据研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62300095

传真：010-62300095

