

团 体 标 准

T/AI 126.1—2023

数据湖 第 1 部分：磁光电混合媒体分布式存储系统

Data lake
Part 1: Distributed storage system with optical, magnetic and electrical hybrid media

2023 - 12 - 30 发布

2023 - 12 - 30 实施

中关村视听产业技术创新联盟 发布

T/AI 126.1-2023

目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 混合媒体存储分类	2
5.1 按照存储媒体分类	2
5.2 按照产品结构分类	3
6 技术架构	3
6.1 系统架构	3
7 功能要求	4
7.1 数据写入	4
7.2 数据读取	4
7.3 数据生存周期管理	4
7.4 数据分级存储	4
7.5 数据管理	4
7.6 存储服务和协议	5
7.7 存储扩展	5
8 性能要求	5
8.1 存储媒体长效性	5
8.2 存储容量	5
8.3 兼容性	5
8.4 I/O 响应时间	6
8.5 存取速度	6
9 信息安全要求	6
9.1 系统安全	6
9.2 数据安全	6
10 建设环境要求	6
10.1 电源适应性	6
10.2 系统物理安全	7
10.3 电磁兼容性	7
10.4 环境适应性	7
10.5 可靠性	8
10.6 能效要求	8
10.7 限用物质的限量	8

10.8 安装场地 8

T/AI 126.1-2023

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是T/AI 126《数据湖》的第1部分。T/AI 126拟发布了以下部分：

- 第1部分：磁光电混合媒体分布式存储系统；
- 第2部分：蓝光存储资源管理系统接口；
- 第3部分：数据资源管理系统；
- 第4部分：人工智能应用技术要求；
- 第5部分：城市治理水平评价模型；
- 第6部分：交通应急指挥与协调决策系统接口。

本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由新一代人工智能产业技术创新战略联盟AI标准工作组提出。

本文件由中关村视听产业技术创新联盟归口。

本文件起草单位：北京易华录信息技术股份有限公司、文安智能科技有限公司、博雅鸿图视频技术有限公司、华为技术有限公司、北京百度网讯科技有限公司、北京大学、中山大学。

本文件主要起草人：杨帅、魏占慧、王凌、赵阳、倪志云、李军浩、牛川、朱涛、陈颖、宋君、刘衡治、黄铁军、赵海英、崔晓冉、李艳梅、梁凡。

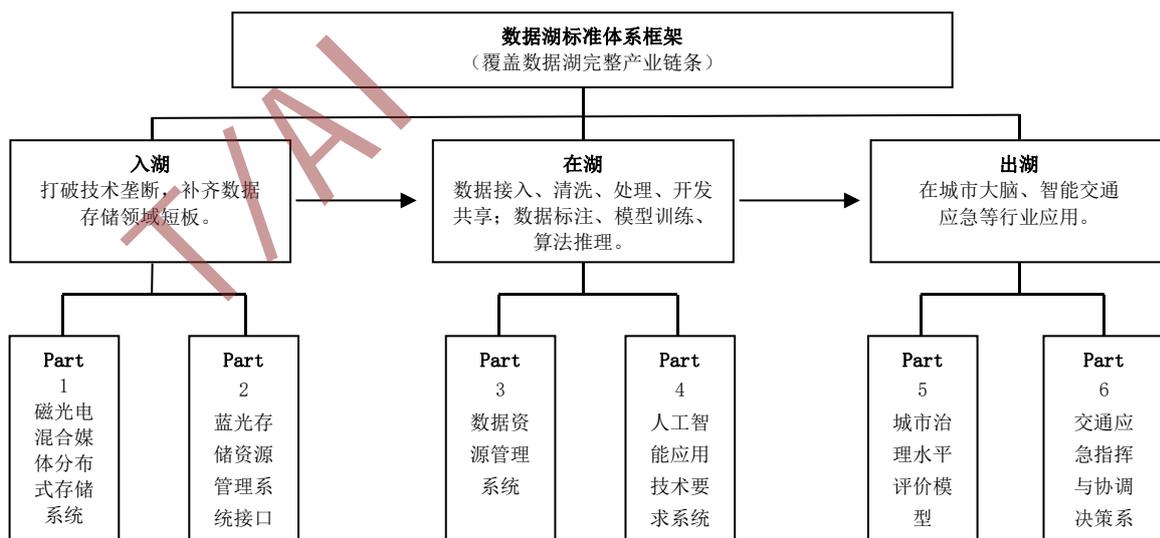
引 言

数据湖是用于存储、处理和分析，大量结构化、半结构化和非结构化数据的系统或存储库。城市数据湖是融合数据感知、存储、分析为一体，以光磁融合存储为依托，以人工智能为引擎、以区块链、云计算、大数据平台等技术为支撑，提供IDC、云计算、湖存储、数据增值、数据安全等运营服务的新一代数字经济基础设施。

T/AI 126-2023《数据湖》是指导城市数据湖建设和应用技术需求的基础性标准，数据湖应用技术涉及从数据分级存储、分析处理、AI需求定义、应用技术框架、行业应用业务落地等全方位覆盖产业链各环节。基于此，本标准从多维多源数据的入湖、在湖、出湖各场景应用技术需求给出标准系列，拟由六个部分构成。

- 第1部分：磁光电混合媒体分布式存储系统。数据湖中汇集了海量数据，磁光电混合媒体分布式存储系统作用于数据入湖阶段，目的在于指导数据湖中磁光电混合存储的应用系统设计、建设和应用，提供技术参考。
- 第2部分：蓝光存储资源管理系统接口。目的在于指导蓝光光盘存储的应用系统设计与开发，提供技术参考。
- 第3部分：数据资源管理系统。目的在于针对数据湖内数据的接入、存储、管理、共享交换等，提供技术参考。
- 第4部分：人工智能应用技术要求。目的在于指导数据湖中人工智能技术应用框架及人工智能技术应用API的设计与开发，提供技术参考。
- 第5部分：城市治理水平评价模型。目的在于针对数据湖城市大脑建设过程中对指标的统一规划、统一开发、统一运维等做一定程度的指导，提出技术参考建议。
- 第6部分：交通应急指挥与协调决策系统接口。目的在于明确定义了基于数据湖支撑的交通应急决策系统相关接口技术，为面向城市区域交通安全事件应急指挥与协调决策提供相关系统建设提供技术参考。

数据湖标准体系框架如下所示：



数据湖

第 1 部分：磁光电混合媒体分布式存储系统

1 范围

本文件作用于数据入湖阶段，为数据湖中作为存储底座的分布式存储系统提供标准规范。本文件给出了磁光电混合媒体分布式存储系统的技术架构，规定了功能要求、性能要求、信息安全要求和建设环境要求，为数据湖内数据的接入、存储等，提供技术参考。

本文件适用于磁光电混合媒体分布式存储的设计、制造和应用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB/T 2887-2011 计算机场地通用规范
- GB 4943.1-2022 音视频、信息技术和通信技术设备 第1部分：安全要求
- GB/T 9254.1-2021 信息技术设备、多媒体设备和接收机 电磁兼容 第1部分：发射要求
- GB/T 9254.2-2021 信息技术设备、多媒体设备和接收机 电磁兼容 第2部分：抗扰度要求
- GB/T 15934-2008 电器附件 电线组件和互连电线组件
- GB 17625.1-2022 电磁兼容 限值 第1部分：谐波电流发射限值（设备每相输入电流≤16A）
- GB/T 26572-2011 电子电气产品中限用物质的限量要求
- GB/T 41785-2022 磁光电混合存储系统通用规范
- GB/T 37973-2019 信息安全技术 大数据安全管理指南

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据湖 data lake

数据湖是用于存储、处理和分析，大量结构化、半结构化和非结构化数据的系统或存储库。

3.2

磁光电混合媒体分布式存储系统 distributed storage system for optical, magnetic and electrical hybrid media

带有光盘，同时包含磁盘、固态硬盘/卡中的一种及以上存储媒体，且通过软件管理对外提供统一存储空间的存储系统。

[来源：GB/T 41785-2022, 3.1]

3.3

I/O 响应时间 I/O response time

应用程序发出I/O请求，到接收到存储设备返回该请求数据的时间间隔。

3.4

存取吞吐率 access throughput rate

单位时间内存取的数据总量。

3.5

光盘库 optical disc library

一种带有光盘自动交换装置（机械手）的存储设备。

3.6

分布式存储 distributed storage

分布式存储是通过网络使用每台机器上的磁盘空间，并将这些分散的存储资源构成一个虚拟的存储设备。

3.7

块存储 block storage

块存储是一种裸设备，它是将存储设备以“块”的方式直接提供给客户。

3.8

文件存储 file storage

文件存储采用NFS或CIFS命令集访问数据，以文件为最小存储单元，通过TCP/IP实现网络化存储。

3.9

对象存储 object storage

对象存储是采用扁平数据组织形式并通过基于HTTP的接口访问的分布式存储系统。

4 缩略语

下列缩略语适合于本文件。

API: 应用编程接口 (Application Programming Interface)

CIFS: 网络文件共享系统 (Common Internet File System)

HTTP: 超文本传输协议 (Hyper Text Transfer Protocol)

ISCSI: 互联网小型计算机系统接口 (Internet Small Computer System Interface)

NFS: 网络文件系统 (Network File System)

NVMe: 非易失性存储器 (Non-Volatile Memory express)

OSD: 对象存储设备 (Object Storage Device)

PMR HDD: 垂直磁记录硬盘驱动器 (Perpendicular Magnetic Recording Hard Disk Drive)

QoS: 服务质量 (Quality of Service)

S3: 简单存储服务 (Simple Storage Service)

SMR HDD: 叠瓦磁记录硬盘驱动器 (Shingled Magnetizing Recording Hard Disk Drive)

SSD: 固态硬盘驱动器 (Solid State Drives)

TCP/IP: 传输控制协议/互联网协议 (Transmission Control Protocol/Internet Protocol)

WORM: 一次写入多次读取 (Write Once Read many)

5 混合媒体存储分类

5.1 按照存储媒体分类

5.1.1 光盘+固态硬盘

包含光盘和固态硬盘二种存储媒体，可通过纵向或横向扩展方式实现存储容量与性能扩展的存储系统。

5.1.2 光盘+机械式磁盘

包括光盘和机械式磁盘二种存储媒体，可通过纵向或横向扩展方式实现存储容量与性能扩展的存储系统。

5.1.3 光盘+磁带

包括光盘和磁带二种存储媒体，可通过纵向或横向扩展方式实现存储容量与性能扩展的存储系统。

5.1.4 光盘+机械式磁盘+固态硬盘

包含光盘、机械式磁盘和固态硬盘三种存储媒体，可通过纵向或横向扩展方式实现存储容量与性能扩展的存储系统。

5.1.5 光盘+磁带+固态硬盘

包含光盘、磁带和固态硬盘三种存储媒体，可通过纵向或横向扩展方式实现存储容量与性能扩展的存储系统。

5.1.6 光盘+磁带+机械式磁盘

包含光盘、磁带和机械式磁盘三种存储媒体，可通过纵向或横向扩展方式实现存储容量与性能扩展的存储系统。

5.1.7 光盘+磁带+机械式磁盘+固态硬盘

包含固态硬盘、机械式磁盘、磁带和光盘四种存储媒体，可通过纵向或横向扩展方式实现存储容量与性能扩展的存储系统。

5.2 按照产品结构分类

5.2.1 一体机

集成光盘、磁带、机械式磁盘和固态硬盘中的一种或几种存储媒体的单机柜产品形态。

5.2.2 集群

集成光盘、磁带、机械式磁盘和固态硬盘中的一种或几种存储媒体的多机柜产品形态。

6 技术架构

6.1 系统架构

磁光电混合媒体分布式存储系统的系统架构见图1。



图1 系统架构

磁光电混合媒体分布式存储系统由软件层和硬件层组成，软件层是系统的主要部分，硬件层为软件层提供物理介质支撑。通过块存储、文件存储和对象存储等标准存储协议中的一种或几种与外部系统进行数据读写。

- a) 软件层应包括以下三个功能区：
 - 1) 接口功能区：应提供通用的软件接口，提供块存储（iSCSI 协议）、文件存储（NFS 和 CIFS 协议）或对象存储（标准 S3 协议）中的一种或几种；
 - 2) 系统控制功能区：应管控存储区并保障其正常工作，提供数据写入、数据读取、存储扩展、数据存储存储和数据生存周期管理等功能的统一调度；
 - 3) 存储管理功能区：应提供数据管理功能，管理存储模块中的存储盘或存储节点应对外提供统一存储空间，应实现不同存储媒体间的数据迁移。
- b) 硬件层应包括以下二级存储区：
 - 1) 第一级存储区：为系统控制功能模块提供电/磁媒体物理存储空间，硬件设备为服务器，可纵向或横向扩展；
 - 2) 第二级存储区：为系统控制功能模块提供光/磁带媒体物理存储空间，硬件设备为磁带库或蓝光库，可纵向或横向扩展。

7 功能要求

7.1 数据写入

磁光电混合媒体分布式存储系统应通过接口功能区接收到写入数据请求，由系统控制功能区执行写入命令，把数据写入存储区相应的存储媒体，由存储管理功能区记录数据的存储位置。数据写入功能应符合下列要求：

- a) 提供块、文件和对象等标准存储协议中一种或几种写入数据；
- b) 将数据直接写入第一级存储；
- c) 将数据直接写入第二级存储；
- d) 通过第一级存储将数据写入第二级存储。

7.2 数据读取

磁光电混合媒体分布式存储系统应通过接口模块接收到读取数据命令，由系统控制模块执行读取命令，由存储管理模块查找到请求数据的存储位置，并读出数据发送给接口模块。数据读取功能应符合下列要求：

- a) 提供块存储、文件存储和对象存储等存储协议中的一种或几种读取数据；
- b) 直接从第一级存储读取数据；
- c) 直接从第二级存储读取数据；
- d) 通过第一级存储读取第二级存储中的数据。

7.3 数据生存周期管理

磁光电混合媒体分布式存储系统应提供生存周期管理，数据按照生存周期进行过期删除或者定期转储。

7.4 数据分级存储

磁光电混合媒体分布式存储系统应提供数据分级存储：可根据不同的业务类型，按照数据的访问频率的高低，将数据分级适配存储在不同媒体上：将访问频率高的数据，存储在读写性能高的存储媒体上；将访问频率低的数据，存储在成本低、安全性高、稳定性强、寿命长的存储媒体上。数据访问频率的高低根据业务场景划分，常规意义上可将半年内访问过的数据划分为访问频率高的数据，半年内没有访问过的数据划分为访问频率低的数据。存储媒体可包括：NVMe、SSD、PMR HDD、SMR HDD、磁带和蓝光光盘。

7.5 数据管理

磁光电混合媒体分布式存储系统数据管理功能应包含以下功能：

- a) 提供系统数据管理，采用系统统一命名空间设计，通过统一的目录视图找到所保存的文件数据；
- b) 提供批量数据自动写入、迁移或恢复；

- c) 提供增量数据自动备份；
- d) 提供对存储集群的监控、告警和日志审计，提供数据统计和分析，应对数据存储、读写、检测、备份和迁移情况分时段统计；
- e) 提供多版本：同一数据按存储时间存储多份；
- f) 对数据压缩处理；
- g) 对数据加密处理；
- h) 提供缓存加速：SSD 缓存加速；
- i) 提供小文件聚合：小文件合并后写入存储媒体，提高文件的写入效率；
- j) 提供 WORM：一次写入，多次读取模式，保护数据安全；
- k) 提供负载均衡，保证数据高可用。

7.6 存储服务 and 协议

应提供多种存储方式和相关协议：块存储（iSCSI协议）、文件存储（NFS协议和CIFS协议）、对象存储（标准S3协议）。

7.7 存储扩展

7.7.1 纵向扩展

纵向扩展技术要求如下：

- a) 通过专用接口实现单机的纵向扩展；
- b) 提供系统纵向扩展前后的各组成部分状态展示。

7.7.2 横向扩展

横向扩展技术要求如下：

- a) 提供横向扩展模块实现集群的横向扩展；
- b) 实现集中管理、配置和控制集群；
- c) 提供系统横向扩展前后的各组成部分状态展示。

8 性能要求

8.1 存储媒体长效性

磁光电混合媒体分布式存储系统中第二级存储媒体贮存寿命一般应不低于10个自然年。

8.2 存储容量

磁光电混合存储应标注第一级存储、第二级存储和系统的标称容量。
存储集群的最大可用容量应不低于其标称容量的90%。

8.3 兼容性

8.3.1 数据

数据兼容性要求如下：

- a) 提供存储数据迁移环境、工具和接口；
- b) 通过第三方设备或系统正确读取单张光盘数据；
- c) 提供第二级存储媒体在第三方系统批量读取。

8.3.2 硬件

硬件兼容性要求如下：

- a) 第一级存储和第二级存储可向前兼容；
- b) 第一级存储和第二级存储可兼容至少二种品牌、容量和型号的相应存储媒体。

8.3.3 软件

软件兼容性要求如下：

- a) 存储管理功能区可兼容至少二种不同品牌的操作系统和数据库；
- b) 系统控制功能区核心部件可兼容至少二种不同品牌和型号的产品。

8.4 I/O 响应时间

系统响应时间与相同容量热磁存储系统，在相同使用和环境条件下，平均延迟应不超过50%。

8.5 存取速度

在相同使用和环境条件下，对于日常小规模数据，系统存取速度应不低于同等容量热磁存储系统存取速度的80%；对于后台存取大规模数据，系统存取速度应不低于同等容量热磁存储系统存取速度的50%。

9 信息安全要求

9.1 系统安全

系统安全要求如下：

- a) 应监控异常访问等情况；
- b) 应设置异常或故障告警；
- c) 提供权限管理和日志输出功能。

9.2 数据安全

数据安全要求如下：

- a) 保密性要求
 - 1) 应保障数据传输的保密性，使用不同的安全协议保障数据采集、分发等操作中的传输保密；
 - 2) 应保障数据存储的保密性，使用访问控制、加密机制等；
 - 3) 应加密数据的运算，例如使用同态加密等算法；
 - 4) 应保护数据汇聚时的敏感性，提供支持对关键或敏感数据的加密写入和解密读出；
 - 5) 应通过数据匿名化使得个人信息主体无法被识别，提供支持用户身份鉴别，防止假冒人员登录。
 - 6) 所采用的密码设备和算法应符合国家有关规定，系统密码功能应通过密码测评机构相关检测。
- b) 完整性要求
 - 1) 应验证数据来源，确保数据来自于已认证的数据源；
 - 2) 应保障数据传输的完整性，确保数据传输安全；
 - 3) 应保障数据存储的完整性，提供数据冗余技术：支持多副本和纠删码技术；
 - 4) 密钥生存周期应与和数据的存储生存周期保持一致。
- c) 可靠性要求
 - 1) 提供拓扑规划和故障域隔离：支持自主规划集群物理设备拓扑，允许多级故障隔离；
 - 2) 提供不中断维护：硬件维护、软件升级和容量扩充时，保证业务不中断；
 - 3) 提供磁盘漫游：更换节点中硬盘的位置，OSD 可以正常启动并提供服务；
 - 4) 提供磁盘重建：原有的数据盘损坏重新更换时，可以使用新硬盘重建 OSD，并将数据重新恢复至新硬盘上；
 - 5) 提供数据恢复 QoS：支持不同级别的数据恢复控制。
- d) 供数据安全增强服务时符合 GB/T 37973 相关要求如下：
密钥生存周期应与和数据的存储生存周期保持一致。

10 建设环境要求

10.1 电源适应性

电源适应性应符合 GB/T 41785 相关要求：

- a) 对于交流供电的产品，应能在 220V±22V、50Hz±0.5Hz 条件下正常工作。
- b) 对于直流供电的产品，应能在直流电压标称值的±5%条件下正常工作。标称值应在产品铭牌或随机说明书中规定。对于电源有特殊要求的产品应在随机说明书中加以说明。
- c) 电线组件应符合 GB 15934 的规定。

10.2 系统物理安全

系统的物理安全要求应符合 GB 4943.1 的规定。

10.3 电磁兼容性

10.3.1 电磁兼容发射

产品的发射应符合 GB/T 9254.1 中对 A 级设备的限值要求。

10.3.2 谐波电流

谐波电流应符合 GB 17625.1 中对 A 类的限值要求。

10.3.3 抗扰度

产品的抗扰度应符合 GB/T 9254.2 的规定。

10.4 环境适应性

10.4.1 气候环境适应性

系统气候环境适应性应符合表 1 的规定。

表 1 气候环境适应性

气候条件		参数
温度	工作 °C	0~40
	贮存温度 °C	-20~60
相对湿度	工作	20%~80%
	贮存运输	20%~93% (40°C)
大气压 kPa		86~106

[来源：GB/T 41785-2022, 6.12.1]

10.4.2 机械环境适应性

系统的振动适应性、碰撞适应性和运输包装件跌落适应性应分别符合表 2、表 3 和表 4，满足应符合 GB 4943.1 的规定。

表 2 振动适应性

试验项目	试验内容	具体要求
初始和最后振动响应检查	频率范围 Hz	5~35
初始和最后振动响应检查	扫频速度 oct/min	≤1
	驱动振幅 mm	0.15
定频耐久试验	驱动振幅 mm	0.15
	持续时间 min	10

表 2 振动适应性 (续)

试验项目	试验内容	具体要求
扫频耐久试验	频率范围 Hz	5~35~5
	驱动振幅 mm	0.15
	扫频速度 oct/min	≤1
	循环次数	2
注：表中驱动振幅为峰值。		

表 3 碰撞适应性

峰值加速度 m/s ²	脉冲持续时间 ms	碰撞次数	碰撞波形
100	16	1000	半正弦波

表 4 运输包装件跌落适应性

包装件质量 kg	跌落高度 mm
≤15	1000
15~30	800
30~40	600
40~45	500
45~50	400
>50	300

[来源：GB/T 41785-2022, 6.12.2]

10.4.3 其他环境适应性

特殊环境条件应在产品标准中规定。

10.5 可靠性

产品的平均失效间隔时间的不可接受值应不小于9000h。

10.6 能效要求

系统消耗功率应不超过相同容量热磁存储系统在相同使用和环境条件下的20%。同时，系统应根据运行负载和环境的变化，自动调整能耗功率，达到节能的效果。

10.7 限用物质的限量

限用物质的限量应符合GB/T 26572的要求。

10.8 安装场地

大型系统设备按GB/T 2887要求。