ICS 03.080.99
CCS A 20

T/ACCEM

团　　体　　标　　准

T/ACCEM XXXX—XXXX

工程咨询 AI 大数据算力技术规范

Technical specification for AI big data computing power in engineering consulting

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中国商业企业管理协会　　发 布

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。

请注意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。

本文件由中德高路咨询(云南)股份有限公司提出。

本文件由中国商业企业管理协会归口。

本文件起草单位：中德高路咨询(云南)股份有限公司、XXX、XXX。

本文件主要起草人：XXX、XXX、XXX。

工程咨询AI大数据算力技术规范

1 范围

本文件规定了工程咨询AI大数据算力的术语和定义、系统技术要求、安全与隐私保护、测试与验证和实施与维护的内容。

本文件适用于程咨询AI大数据算力的设计、实施与验证。

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 32909 非结构化数据表示规范

GB/Z 40846-2021 工程咨询 基本术语

YD/T 3802-2020 电信网和互联网数据安全通用要求

YD/T 4255-2023 算力网络 总体技术要求

3 术语和定义

GB/Z 40846-2021界定的以及下列术语和定义适用于本文件。

3.1

算力 computing capability

网络中具有计算能力的节点通过对工程类数据的处理，实现特定咨询申请结果输出的能力，具体包括不限于计算、内存和存储能力，算力可以分布在网络边缘、云数据中心、连网终端、转发节点等各种形态的设备上。

[来源：YD/T 4255-2023，3.1，有修改]

3.2

算力资源 computing power resources

提供计算能力的软硬件基础设施的集合,其中硬件方面包括CPU、GPU集群等计算设备，以及分布式计算框架所依赖的硬件环境,软件方面涵盖了分布式计算框架以及云计算资源的管理和调度软件。算力资源。

4 系统技术要求

4.1 总体架构

系统应采用模块化分层设计，系统包含以下层级：

——数据层：应具备强大的数据接入和存储能力，支持多源异构数据的实时接入。通过高效的数据采集和传输技术，确保数据能够及时准确地存储到系统中。同时，采用合适的数据存储架构，满足不同类型数据的存储需求。

——算力层：应包括算力服务层、算力路由层、算网基础设施层和算网编排管理层，并提供弹性伸缩的算力资源池，能够根据业务需求动态调整算力资源的分配。支持CPU/GPU混合调度，充分发挥不同计算设备的优势，提高计算效率。通过智能的算力调度算法，实现算力资源的优化利用。

——算法层：应集成丰富的机器学习、深度学习模型库，涵盖各种常用的算法模型。支持模型的训练与推理过程，通过对大量数据的学习和分析，生成能够解决实际工程咨询问题的模型。同时，应提供模型管理和优化工具，方便对模型进行更新和改进。

——应用层：提供 API 接口及可视化交互界面，API 接口用于与其他系统进行数据交互和功能集成，实现系统的开放性和扩展性。可视化交互界面则应为用户提供直观、便捷的操作方式，覆盖工程咨询全业务流程，包括数据查看、分析结果展示、决策支持等功能。

4.2 数据管理要求

4.2.1 数据采集

4.2.1.1 结构化数据格式应符合 JSON 或 XML 规范，确保数据的规范性和一致性，便于数据的处理和传输。

4.2.1.2 非结构化数据表示应符合 GB/T 32909 的规定，并标注元数据（来源、时间、类型），通过标注元数据，能够更好地对非结构化数据进行管理和检索，提高数据的可用性。

4.2.2 数据存储

4.2.2.1 热数据存储响应时间≤1 s，以满足实时数据处理和分析的需求。对于频繁访问的数据，宜采用高速存储设备和优化的存储架构，确保数据能够快速读取和写入。

4.2.2.2 冷数据归档周期≤24 h，对于不经常访问的数据，及时进行归档处理，节省存储资源。

4.2.2.3 备份的数据应确保数据的安全性和可靠性。

4.3 算力资源配置

4.3.1 硬件要求：

4.3.1.1 单节点算力应确保每个计算节点具备足够的计算能力，以支持复杂的 AI 算法和大数据分析任务。

4.3.1.2 网络带宽应大于 10Gbps，保证数据在节点间的快速传输，减少数据传输延迟。

4.3.1.3 节点间延迟应不大于 2 ms，确保分布式计算环境下节点之间的高效通信。

4.3.2 软件要求：

4.3.2.1 算力网络架构及接口功能要求应符合 YD/T 4255-2023 第 6 章规定。

4.3.2.2 应支持主流分布式框架，能够满足不同类型的大数据处理和 AI 模型训练需求。

4.3.2.3 应通过容器化技术和 Kubernetes 进行部署和管理，提高系统的可移植性、可扩展性和可维护性。

5 安全与隐私保护

5.1 数据安全

5.1.1 数据传输

数据传输安全应符合YD/T 3802-2020第9章的规定，应禁用弱密码套件，防止因使用不安全的密码套件而导致数据泄露的风险。

5.1.2 数据存储

5.1.2.1 数据存储应符合 YD/T 3802-2020 第 10 章以及 GB/T 37973-2019 第 5 章的规定。

5.1.2.2 敏感数据安全保护应符合 YD/T 3865-2021 第 8 章的规定。

5.1.2.3 个人信息的存储应符合 GB/T 35273-2020 第 6 章的规定。

5.2 系统安全

5.2.1 访问控制

5.2.1.1 数据的使用应符合 YD/T 3802-2020 第 11 章的规定。

5.2.1.2 数据开放共享应符合 YD/T 3802-2020 第 12 章的规定。

5.2.1.3 个人信息的使用应符合 GB/T 35273-2020 第 7 章的规定。

5.2.1.4 基于 RBAC 模型，角色权限细分至数据字段级别，通过对用户角色的定义和权限分配，实现对系统资源的细粒度访问控制，确保只有授权用户能够访问和操作相应的数据和功能。

5.2.2 审计日志

对系统中的所有操作应进行详细记录，以便进行审计和追溯。日志应保留6个月以上，满足安全审计和合规性要求。

6 测试与验证

6.1 性能测试

6.1.1 算力评估

算力评估指标及测试方法应符合表1的规定。

表 1 算力评估指标

项目	类型	单位	测试方法
整数计算速率	Int	KB/seconds of cpu time	平台算法测试
半精度浮点计算速率	Float	GFLOPS、TFLOPS	业界常用方法
单精度浮点计算速率	Float	GFLOPS、TFLOPS	业界常用方法
双精度浮点计算速率	Float	GFLOPS、TFLOPS	业界常用方法
哈希计算速率	Float	Hash/s	业界常用方法

6.1.2 通信能力

通信能力评估指标及测试方法应符合表2的规定。

表 2 通信能力评估指标及测试方法

项目	类型	单位	测试方法
网络带宽	Int	Mbps	厂商获取
DPDK L3转发能力	Float	Mbps or Gbps	厂商获取
FIB能力	Float	Mbps or Gbps	厂商获取
IPSec能力	Float	Mbps or Gbps	厂商获取
虚拟网络能力	Float	Mbps or Gbps	厂商获取
防火墙损耗	Float	Mbps or Gbps	厂商获取

6.1.3 内存能力

内存能力评估指标及测试方法应符合表3的规定。

表 3 内存能力评估指标及测试方法

项目	类型	单位	测试方法
内存容量	Int	MB	厂商获取
内存带宽	Int	MB/s	STREAM测试程序
内存访问延时	Float	ns	Lmbench测试程序

6.1.4 存储能力

存储能力评估指标及测试方法应符合表4的规定。

表 4 存储能力评估指标及测试方法

项目	类型	单位	测试方法
存储容量	Int	GB或TB	厂商获取
存储带宽	Float	MB/s	存储带宽=读写字节数/存取周期
每秒进行读写操作的次数	Int	Ops/s	FIO工具

6.2 模型验证

6.2.1 交叉验证

采用5折交叉验证，将数据集划分为5份，每次使用4份作为训练集，1份作为测试集，重复5次，以评估模型的泛化能力。数据集划分比例 7:3，即训练集占 70%，测试集占 30%，确保训练集和测试集具有代表性。

6.2.2 生产环境 A/B 测试

新模型上线应通过显著性检验，通过 A/B 测试比较新模型和旧模型在生产环境中的性能表现，只有当新模型的性能具有显著优势时，才允许新模型上线，确保模型的改进能够真正提高系统的性能和服务质量。

7 实施与维护

7.1 版本管理

主版本号按“年.月”格式表示，便于清晰地标识系统的版本更新时间和顺序。紧急修复版本追加后缀（如2024.01.1），当系统出现紧急问题需要修复时，通过追加后缀的方式标识修复版本，方便用户了解系统的更新情况。

7.2 文档要求

7.2.1 应提供《系统部署手册》，详细描述系统的部署环境、安装步骤、配置方法等信息，方便系统的部署和安装。

7.2.2 应提供《API 接口文档》，对系统提供的 API 接口进行详细说明，包括接口的功能、参数、返回值等信息，便于开发人员进行系统集成和二次开发。

7.2.3 应提供运维标准操作程序，包含系统的日常运维操作、故障排除流程、性能优化方法等内容，指导运维人员进行系统的维护和管理。

7.3 培训要求

每年至少开展2次技术培训，培训内容包括系统的使用方法、技术原理、维护技巧等，确保系统的使用人员和维护人员都能够接受培训，提高他们对系统的掌握程度和使用效率。