



团 体 标 准

T/BECC 002—2024

智算中心技术要求和评估方法

Technical requirements and evaluation method of artificial
intelligence computing center

2024-03-28 发布

2024-04-02 实施

北京电子商会 发布
中国标准出版社 出版

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 1

5 智算中心总体要求 2

 5.1 智算中心服务分类 2

 5.2 智算中心功能架构 2

 5.3 智算中心通用要求 3

6 智算基础设施 5

 6.1 IDC 机房 5

 6.2 网络交换机 9

 6.3 智算算力服务器 11

 6.4 通用计算服务器 12

 6.5 存储服务器 12

 6.6 安全服务器 13

7 智算基础平台 13

 7.1 算力服务 13

 7.2 容器服务 13

 7.3 容器镜像服务 13

 7.4 网络服务 14

 7.5 存储服务 14

 7.6 安全服务 14

8 机器学习平台 15

 8.1 总体要求 15

 8.2 数据预处理 15

 8.3 模型开发 16

 8.4 模型训练 16

 8.5 模型推理 16

 8.6 模型优化 17

 8.7 模型评估 17

 8.8 模型纳管 17

8.9 模型部署 18

9 大模型智算平台 18

9.1 总体要求 18

9.2 数据采集 19

9.3 大模型训练 19

9.4 大模型微调 19

9.5 大模型推理 20

9.6 大模型多模纳管 20

10 统一监控运营平台 20

10.1 平台业务流程 20

10.2 统一门户 21

10.3 用户控制台 21

10.4 资源管理 21

10.5 运营管理 22

10.6 运维管理 22

10.7 融合多租户管理 22

10.8 统一安全管理 23

10.9 自动化安装部署 23

10.10 通用基础平台 23

10.11 机器学习平台和大模型智算平台的运营 23

11 智算中心评估方法 23

11.1 评估原则 23

11.2 评估内容与指标 24

11.3 评估方法 24

11.4 评估结论与判定 24

附录 A（资料性） 智算中心典型场景案例 25

附录 B（资料性） 智算中心统一监控运营平台相关能力 32

参考文献 36

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。

本文件由北京电子商会提出并归口。

本文件起草单位：北京电子控股有限责任公司、北京电子数智科技有限责任公司、阿里云计算有限公司、新华三集团、首都信息发展股份有限公司、中国软件评测中心、北京电子城数智科技有限公司、中国电信股份有限公司北京分公司、中电云计算技术有限公司、浪潮（北京）电子信息产业有限公司。

本文件主要起草人：王琳、杨坤、徐文星、陆海翔、刘如冰、王磊、魏诗雅、郭磊、师宇、杨鑫、孙东明、李辉、于波、徐高杰、杨景、刘小健、刘俊伟、马辉、兰文丽、赵德智、郝杰。

技术审查专家：吴志刚、张群、张东、于铁强、蔡春久、胡良霖、曹建彤、朱大培、陈玉峰、张雅彬、马福忠。

引 言

智算中心是开展算力资源服务,支撑各领域开展数据运营的重要基础设施之一。目前,智算中心存在能力等级不清晰、度量指标不统一、服务商的服务能力不明确等问题。亟待解决的典型问题包括:

- a) 对于中等复杂模型算力,预训练、微调、检验、回归测试全流程的效率较低;
- b) 算力隔离问题,号称支持多个 CPU 隔离与实际效果不符,缺乏隔离评价机制;
- c) 如何保障存储、算力、负载全链路的高效可靠。

因此,急需制定本文件,从用户对算力服务的不同需求、数据应用服务的不同环节(采集、传输交互、加工处理、分析建模、应用服务等)出发,明确算力服务类别、功能性能要求和评估方法等,作为智算中心建设/评估参考依据:

- a) 便于用户依据应用场景、资金预算对算力服务的选型参考;
- b) 为算力服务供应链各方的产品研发和统一对接提供参考标准,实现高效、高质量、高性价比的算力服务;
- c) 为政府、企事业单位等各类算力用户/第三方评估机构开展算力服务评价提供参考依据。

智算中心技术要求和评估方法

1 范围

本文件规定了智算中心的服务分类、功能架构、通用要求、智算基础设施、智算基础平台、机器学习平台、大模型智算平台、统一监控运营平台的技术要求和评估方法。

本文件适用于指导智算中心的研发、设计、建设和运营。

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 50073—2013 洁净厂房设计规范

GB 50116—2013 火灾自动报警系统设计规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

智算中心 artificial intelligence computing center

具备智能计算、存储、高性能网络、容器、安全等基础设施和服务,通过机器学习平台和大模型智算平台为各类智算场景和应用提供人工智能算力、大模型开发训练和统一监控运营等服务的系统。

3.2

训练 training

利用数据,基于机器学习算法,建立或改进机器学习模型参数的过程。

[来源:ISO/IEC 22989:2022,3.3.15]

3.3

推理 inference

计算机根据已知信息进行分析、分类或诊断,做出假设,解决问题或者给出推断的过程。

[来源:GB/T 42018—2022,3.12]

4 缩略语

下列缩略语适用于本文件。

AIGC:生成式人工智能(Artificial Intelligence Generated Content)

API:应用程序编程接口(Application Programming Interface)

CPU:中央处理器(Central Processing Unit)

GPU:图形处理器(Graphics Processing Unit)