

团体标准

T/CESA XXXX—202X

计算中心 算力基础设施总体技术要求

Computing center- General technological requirements for computing power
infrastructure

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。
已授权的专利证明材料为专利证书复印件或扉页，已公开但尚未授权的专利申请证明材料为专利公开通知书复印件或扉页，未公开的专利申请的证明材料为专利申请号和申请日期。

202X-XX- XX 发布

202X-XX- XX 实施

中国电子工业标准化技术协会 发布



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以任何形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目 次

前 言..... III

1 范围..... 4

2 规范性引用文件..... 4

3 术语和定义..... 4

4 缩略语..... 5

5 总体框架..... 5

6 基础能力层..... 6

6.1 计算力..... 6

6.2 存储力..... 7

6.3 运载力..... 8

7 平台能力层..... 9

7.1 算力调度..... 9

7.2 平台软件..... 11

7.3 应用软件..... 11

8 配套能力层..... 11

8.1 绿色节能..... 11

8.2 风火水电基础设施..... 12

9 管理能力层..... 12

9.1 安全..... 12

9.2 运营..... 13

9.3 运维..... 13

附录 A（资料性）算力基础设施能力指标 14

参考文献..... 18

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。

本文件由曙光信息产业股份有限公司提出。

本文件由中国电子工业标准化技术协会高性能计算机标准工作委员会归口。

本文件起草单位：。

本文件主要起草人：。

计算中心 算力基础设施总体技术要求

1 范围

本文件规定了计算中心算力基础设施的基础能力、平台能力、配套能力、管理能力等方面的技术要求。

本文件适用于为计算中心算力基础设施的设计、建设和运营提供指标参考。

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 22239-2019 信息安全技术 网络安全等级保护基本要求

GB/T 39204-2022 信息安全技术 关键信息基础设施安全保护要求

GB 40879-2021 数据中心能效限定值及能效等级

GB 50174-2017 数据中心设计规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

计算中心 **computing center**

以风火水电等基础设施和IT软硬件设备为主要构成，用于处理和管理大规模计算任务的集中化组织。

注：通用数据中心、智能计算中心、超算中心等通常属于该组织的一种。

3.2

算力基础设施 **computing power infrastructure**

用于实现数据信息的集中计算、存储、传输与应用的集计算力、运载力、存储力于一体的新型基础设施。

3.3

计算力 **computing power**

计算中心计算设备对数据处理并实现结果输出的综合能力。

3.4

存储力 **storage power**

计算中心存储设备在数据存储容量、性能表现、安全可靠和绿色低碳四方面的综合能力。

3.5

运载力 network power

计算中心网络设备在网络架构、网络带宽、传输时延、智能化管理与调度等方面的综合能力。

3.6

先进存储 advanced storage

应用全闪存阵列、SSD 等先进存储部件，采用存算分离、高密等先进技术，单位容量数据操作能力达到万IOPS（每秒读写次数）以上的存储模块。

3.7

算力调度 computing scheduling

以计算、存储、网络高度协同为主，包括算力接入、算力交易、算力调度等在内的综合能力。

4 缩略语

下列缩略语适用于本文件。

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

ASIC：专用集成电路（Application Specific Integrated Circuit）

CE：算效（Computational Efficiency）

CPU：中央处理器（Central Processing Unit）

CUE：碳利用效率（Carbon Usage Effectiveness）

ERE：能源再利用效率（Energy Reuse Effectiveness）

FPGA：现场可编程逻辑门阵列（Field Programmable Gate Array）

GPU：图形处理单元（Graphics Processing Unit）

HPL：测试高性能计算集群系统浮点性能的基准程序（High-Performance Linpack）

HPL-MxP：测试HPC和AI的融合性能的基准程序（High-Performance Linpack Mixed-Precision）

IOPS：每秒进行读写操作的次数（Input/Output Operations Per Second）

MLPerf™：对硬件、软件和服务的训练和推理性能进行评估的基准程序（Machine Learning Performance）

NAS：网络附属存储（Network Attached Storage）

NPU：神经网络处理器（Neural Network Processing Unit）

POSIX：可移植操作系统接口（Portable Operating System Interface）

PUE：能效（Power Usage Effectiveness）

RER：可再生能源利用率（Renewable Energy Ratio）

WUE：水利用效率（Water Usage Effectiveness）

5 总体框架

算力基础设施总体框架见图1。

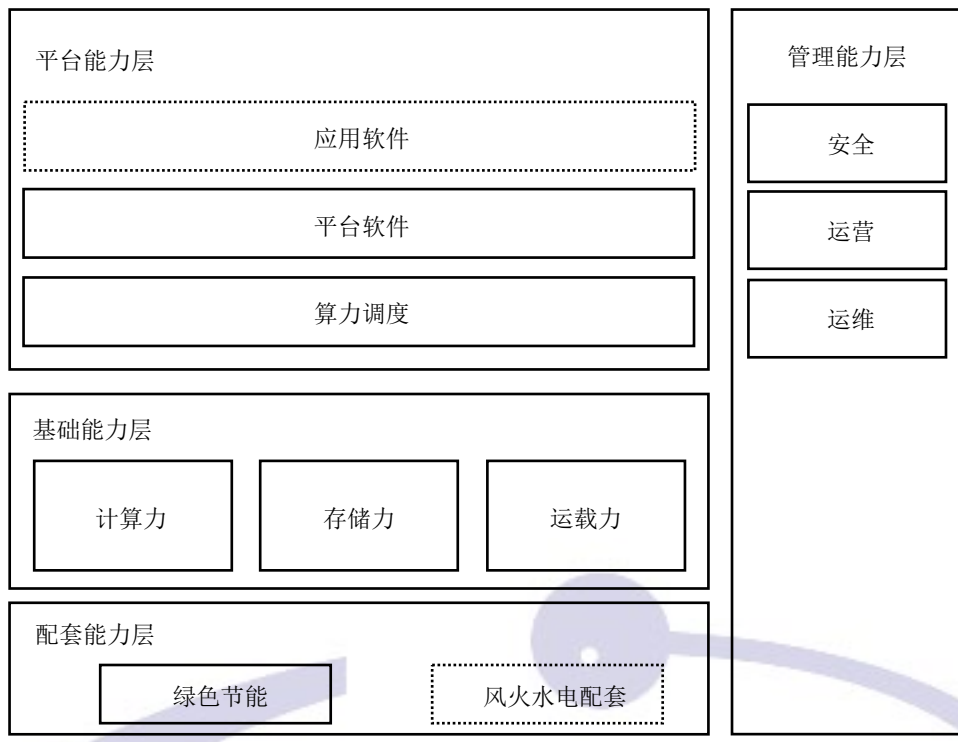


图1 算力基础设施总体框架

注：其中风火水电配套、应用软件不是本文件重点展开的内容，但作为算力基础设施的重要组成部分，以虚框标识。

图1中算力基础设施总体框架主要包括配套能力层、基础能力层、平台能力层、管理能力层：

- a) 配套能力层主要包含绿色节能能力和风火水电配套基础设施能力。绿色节能能力保障算力基础设施满足绿色低碳发展。风火水电配套为算力基础设施提供基本的配套运行环境；
- b) 基础能力层是算力基础设施的核心生产力，主要包括计算力、存储力、运载力，基础能力层的目标是为上层应用提供强大的数据计算能力、高效的数据处理能力和高速的数据传输能力；
- c) 平台能力层主要包含算力基础设施计算调度能力、平台软件能力和应用支撑能力，主要负责管理和调度基础能力层资源，以支持各种应用和服务的需求，构建一体化算力服务体系；
- d) 管理能力层主要包含算力基础设施安全能力、运营能力、运维能力。安全能力是一个全方位的体系，实现数据高效流转和数据安全防护。运营能力指计算中心的运营服务、计量计费能力。运维能力指对算力基础设施的日常管理和维护，以确保其正常运行。

6 基础能力层

6.1 计算力

6.1.1 算力类型

算力类型符合下列要求：

- a) 算力类型按照应用场景划分，应提供包括不限于通用算力、智能算力、超算算力等计算能力；
- b) 算力宜采用融合架构，即超算算力（双精度）和智算算力（半精度）同时存在，双精度与半精度峰值综合算力之比为1:100。

6.1.2 算力架构

算力资源按照架构可划分为通用算力架构和异构算力架构，算力架构符合下列要求：

- a) 通用算力架构，应提供以通用处理器（如CPU）承载应用及程序的能力；
- b) 异构算力架构，应提供以异构处理器（如GPU、FPGA、ASIC、NPU等）承载应用及程序的能力。

6.1.3 峰值综合算力

算力应具备执行多种数据运算的能力，峰值综合算力包括但不限于：双精度浮点FP64峰值算力、单精度浮点FP32峰值算力、半精度浮点FP16峰值算力和整型INT8/INT4峰值算力。

峰值综合算力计算方式参见附录A.1。

6.1.4 有效综合算力

应能通过有效综合算力工具（如HPL、HPL MxP、MLPerf等）评价实际用于计算业务的计算能力。

6.1.5 算力安全可信能力

算力安全可信能力符合下列要求：

- a) 宜采用安全可信的芯片及服务器提供计算能力，安全可信算力占比宜不低于总算力的60%；
- b) 处理器宜支持国密算法；
- c) 处理器应具可信计算功能；
- d) 处理器应具备隐私计算功能；
- e) 宜采用安全可信部件。

6.1.6 算力支撑环境

算力支撑环境应支持并行计算，软件栈应提供成熟的并行计算库和工具。

6.2 存储力

6.2.1 存储介质

算力基础设施存储介质符合下列要求：

- a) 存储系统应包括但不限于SSD和HDD存储介质；
- b) 若存在长期离线归档的业务需求，存储系统除了采用HDD大容量存储介质以外，宜采用蓝光存储、磁带等介质。

6.2.2 存储容量

存储容量应符合下列要求：

- a) 存储裸容量满足各类数据的有效存储需求；
- b) 存储可用容量满足各类数据的实际存储需求。

6.2.3 先进存储

根据应用场景，存储介质宜采用部分先进存储，先进存储容量占比宜不低于总存储裸容量的30%。

6.2.4 存储扩展能力

存储系统应支持在线容量扩展，满足存储集群系统单一命名空间。

6.2.5 存储带宽和存储 IOPS

存储系统应满足业务的性能要求，如存储读写带宽、存储IOPS等。

6.2.6 存储访问协议

存储系统应根据不同业务提供存储访问协议，包括但不限于文件接口（NAS、POSIX）、块接口、对象接口、大数据接口等。

6.2.7 存储高可靠性

存储系统符合下列要求：

- a) 应具备数据冗余的设计，支持数据副本或纠删码等数据冗余策略；
- b) 应具备物理硬件冗余的设计，支持关键部件（包括但不限于电源、控制器、风扇、系统盘等）冗余；
- c) 应具备数据修复与磁盘巡检的能力；
- d) 应采用冗余路径访问等容错技术，提高访问可用性；
- e) 宜具备数据加密及认证的能力。

6.2.8 存储容灾性

存储容灾性符合下列要求：

- a) 应具备数据快照或克隆的能力；
- b) 宜具备数据备份恢复等技术的逻辑故障容灾能力；
- c) 宜具备异步远程复制、同步远程复制、双活等存储技术实现数据的物理故障容灾能力。

6.2.9 存储智能管理

存储智能管理符合下列要求：

- a) 应具备自动精简配置、智能QoS、分级存储、SSD缓存加速等功能；
- b) 应具备可视化的存储管理方式，支持状态监控、资源监控、运维管理等功能；
- c) 宜具备IO模型分析、智能运维、存储容量及性能预测等功能；
- d) 宜具备存储性能持续监控、监控数据长期保存、回顾等功能。

6.2.10 存储绿色节能

存储设备绿色节能的技术包括节能设计、降低散热、高效存储、数据压缩等方面，以提高设备的可靠性和稳定性，降低能源消耗和对环境的影响。符合下列要求：

- a) 宜支持采用低功耗的硬件部件（如SSD）或高效存储技术（如数据压缩、重复数据删除等）；
- b) 宜具备智能电源管理功能；
- c) 宜支持多级工作能耗模式设定，应用可根据业务需求选择适当的能耗模式，达到节能目的；
- d) 宜支持液冷散热技术。

6.3 运载力

6.3.1 网络可靠性

网络可靠性符合下列要求：

- a) 计算中心内部计算网络宜采用无损 (Lossless) 网络, 保证业务数据无丢包;
- b) 网络应支持流量控制与拥塞管理技术, 保证上层业务数据完整性。

6.3.2 网络协议支持

计算网络协议宜支持通用协议, 如: 以太网、InfiniBand、RoCE等。

6.3.3 网络传输速率

计算网络应根据业务需要, 提供高带宽的计算网络, 网络传输速率包括但不限于10Gbps、25Gbps、100Gbps、200Gbps、400Gbps、800Gbps等。

6.3.4 网络传输时延

网络传输时延符合以下要求:

- a) 城市级算力基础设施间时延宜不高于1ms;
- b) 区域级算力基础设施间时延宜不高于5ms;
- c) 国家枢纽节点内算力基础设施间时延宜不高于5ms, 国家枢纽节点间算力基础设施间时延宜不高于20ms。

6.3.5 网络可扩展性

计算网络应能够快速适应不同的应用场景和业务需求, 并能够支持未来的扩展和升级。

6.3.6 网络运维

计算网络宜支持多种运维手段, 包括但不限于管理智能化、体验可视化、实时快速定位网络问题等方式。

6.3.7 网络组网

网络组网符合下列要求:

- a) 计算中心网络应包括但不限于计算网络、存储网络、管理网络;
- b) 宜支持采用虚拟化技术实现逻辑网络业务隔离;
- c) 计算中心网络与外部网络互联方式应支持专线、VPN等方式满足数据传输需求。

7 平台能力层

7.1 算力调度

7.1.1 统一管理

算力调度系统应能够统一管理底层资源 (通用算力资源和异构算力资源), 为不同应用提供灵活的资源调度服务, 满足用户多样化业务需求。

7.1.2 算力接入

算力接入符合下列要求:

- a) 应具备第三方资源接入能力;
- b) 宜采用API、消息传递等技术, 满足计算中心与其它算力基础设施的互相接入能力;

- c) 传输方式应具备数据安全加密要求，非明文跨中心传输数据；
- d) 接入应具备用户身份认证功能；
- e) 接入应具备获取接入资源的调度管理功能，算力平台能够将任务传递给接入资源，并在其正常运行。

7.1.3 算力协同调度

算力协同调度应符合下列要求：

- a) 支持算力资源跨中心调度和管理；
- b) 支持应用工作流跨中心协同调度；
- c) 支持跨中心作业调度时，数据可共享及可控管理；
- d) 支持跨中心服务统一编排和管理；
- e) 采用与硬件资源解耦技术，不受硬件资源限制。

7.1.4 存力协同调度

存力协同调度应符合下列要求：

- a) 支持数据资源跨中心调度、管理和存储；
- b) 支持存力协同调度能力，即数据从一个计算中心存储系统流向另一个计算中心存储系统的能力；
- c) 能够调度不同的存储资源(NAS\块\并行文件\对象等其中的2种)，实现算力、存力的统一身份认证。

7.1.5 算力输出

算力输出应符合下列要求：

- a) 支持开放必要的网络端口，允许运行代理或开放API与算力平台对接；
- b) 支持向算力平台提供本计算中心可输出的算力资源；
- c) 支持算力平台的用户在权限范围内向本计算中心上传、下载数据；
- d) 支持通过算力平台提交的计算任务在本计算中心运行，提供管理计算任务的接口；
- e) 支持隔离本计算中心内不同用户的数据，保障数据安全。

7.1.6 算力交易

算力交易应符合下列要求：

- a) 支持接入平台的算力资源进行量化和定价；
- b) 支持居间方撮合算力供给方与需求方进行交易；
- c) 支持为平台方分别与算力供给方及需求方进行交易明细对账；
- d) 支持对算力资源使用情况进行计量计费，计费粒度可精确到每个计算任务；
- e) 支持生成和导出费用账单；
- f) 支持记录来自算力平台的计算任务使用资源明细，以便与算力平台对账。

7.1.7 统一认证

统一认证功能应符合下列要求：

- a) 支持业界通用的统一认证方式，包括但不限于OAUTH2.0、CAS协议、SAML2.0、RESTFUL API等；
- b) 支持用户登录后可在登录有效期内使用多个计算中心的资源，无需单独登录每个计算中心。

7.1.8 应用商店

应用商店应符合下列要求：

- a) 支持应用软件、数据、服务发布为标准商品用于交易；
- b) 支持用户可以订阅/购买商品，能在有效期内、有效范围内使用，并获得相应的售后服务。

7.1.9 服务保障

服务保障应符合下列要求：

- a) 支持用户的服务级别管理；
- b) 提供在线帮助手册、技术文档；
- c) 提供算力资源的开发指南、调优手册；
- d) 提供售后服务支持能力。

7.2 平台软件

7.2.1 基础软件环境

根据不同算力应用服务，提供相应的基础软件环境。基础软件环境符合下列要求：

- a) 科学计算基础环境，应包括操作系统、编译器、数学库、并行编程MPI等；
- b) 工程计算基础环境，除科学计算基础环境外，还应具备相关远程图形交互软件，如：三维云桌面、VNC或DCV等；
- c) 云计算基础环境，应包括操作系统、数据库、中间件、虚拟化技术、分布式存储等技术；
- d) 智能算力基础环境，应包括操作系统、容器化技术、数据集管理、模型管理等服务，应支持人工智能框架，如TensorFlow、PyTorch、Keras、MXNet、PaddlePaddle等框架之一。

7.2.2 平台软件环境

平台软件环境符合下列要求：

- a) 算力基础设施平台软件宜包含科学计算平台、工程计算平台、云计算平台、人工智能平台等；
- b) 科学计算平台应支持计算作业管理能力，实现了算力资源、存力资源、数据资源的共享；
- c) 工程计算平台应支持远程图形交互界面、仿真工程软件、作业管理等功能；
- d) 人工智能平台应支持人工智能计算环境快速部署，提供训练服务、数据服务、算法服务、资产服务、容器镜像管理、权限管理等功能；
- e) 云计算平台应支持提供虚拟化、存储、网络、数据库、应用中间件、大数据、容器及安全等通用云计算服务统一管理、资源调度、监控日志及运营运维等一整套云资源管理能力。支持租户管理能力完成对资源管控隔离。

7.3 应用软件

7.3.1 算力应用支撑

针对不同的应用场景，算力基础设施应提供不同的算力应用支持，包括科学计算服务、工程计算服务、智能计算服务、云计算服务、存储服务至少3种以上。

8 配套能力层

8.1 绿色节能

8.1.1 一般要求

满足绿色节能及可持续发展要求，各项指标应按照国家现有相关规定执行。

8.1.2 绿色节能指标

绿色节能指标符合下列要求：

- a) 宜监控算力基础设施能源的使用情况，包括但不限于算效（CE）、存效（SE）、能效（PUE）、水资源利用效率（WUE）、碳利用效率（CUE）、可再生能源利用率（RER）、服务器设备能效（SPUE）等。指标计算方法参见A.2绿色节能指标；
- b) 新建能效应符合GB 40879-2021中第4.2节规定的数据中心能效限定值要求，即能效限定值为能效等级2级、能效应不大于1.30；同时还应符合计算中心国家政策要求和所在地的地方标准规定；
- c) 改扩建计算中心能效应符合GB 40879-2021中第4.2节规定的数据中心能效限定值要求，即能效应不大于1.40；同时还应符合计算中心国家政策要求和所在地的地方标准规定；水资源利用效率（WUE）应符合计算中心所在地的地方标准规定，没有限值规定时水资源利用效率值（WUE）宜不大于1.6L/kWh；
- d) 宜充分利用地域、环境优势优先采用可再生能源，提升可再生能源利用率（RER）水平；
- e) 算效（CE）宜考虑细分特定精度下的技术指标，供不同的应用场景参考；
- f) 存效（SE）应考虑结合单位功率所对应的存储性能（如IOPS）。

8.1.3 节能技术

节能技术符合下列要求：

- a) 采用空调制冷时，宜采用如列间空调、背板空调等近端制冷方式提升制冷系统能效，宜采用间接蒸发冷却、高温冷冻水、氟泵空调、磁悬浮相变空调及冷却水直接供冷等技术充分利用自然冷源；
- b) 对于新建或改扩建的计算中心，宜采用液冷技术方案（如冷板式液冷技术、浸没式液冷技术等）；新建计算中心液冷技术方案宜满足全年全地域自然冷却条件；
- c) 宜采用余热回收技术提高能源利用率，宜采用节水及水处理等技术节约水资源；
- d) 制冷系统宜采用人工智能监控调优等技术根据IT负载、室外温湿度等条件调节运行参数，提升制冷系统整体能效；
- e) 供电系统宜采用高压直流供电技术降低供电系统能量损耗；
- f) 末端柜机、板载电源宜提高DC-DC供电的电压等级来满足容量需求，并减少中间降压环节，避免损耗；
- g) IT系统宜采用高能效设备，并通过高效调度运维等方式提高计算中心算效、存效水平。

8.2 风火水电基础设施

可靠性分级、环境要求、人流物流配套、照明消防配套等满足GB 50174-2017中的要求。

9 管理能力层

9.1 安全

9.1.1 信息安全

信息安全方面应符合GB/T 39204-2022中第7章规定的安全防护要求。

9.1.2 网络安全

网络安全方面应符合GB/T 22239-2019中第8章规定的第三级安全要求。

9.2 运营

9.2.1 运营服务

运营服务应符合下列要求：

- a) 应制定统一的服务机制；
- b) 具备统一的运营服务流程；
- c) 具备统一的服务级别；
- d) 具备统一的资源接入及分发管理标准；
- e) 具备统一运营服务评价流程；
- f) 具备统一的计费管理。

9.2.2 租户管理

租户管理应具备权限管理、算力隔离、数据隔离、合规与审计功能。

9.3 运维

9.3.1 运维服务

运维服务应建立适合的服务管理流程、服务活动指导文件或实施规则，以保障运维服务过程的执行规范化。

9.3.2 运维服务对象

运维服务对象应包含但不限于机房基础设施、物理资源、虚拟资源、平台资源、应用资源和数据资源。

9.3.3 运维服务连续性

运维服务连续性应符合下列要求：

- a) 保障业务连续性；
- b) 制定有效的运行维护策略来保证服务交付的质量，并确保按SLA服务协议提供长期、满足需求的服务。

附录 A

(资料性)

算力基础设施能力指标

A.1 峰值综合算力

浮点运算指令具有不同的精度格式，浮点数通常由符号位、指数位、尾数位组成，指数位决定了浮点数的表达范围，尾数位决定了数的精度。算力基础设施常用的浮点精度如表A.1所示：

表A.1 浮点精度表

浮点格式	符号位bit数	指数位bit数	尾数位bit数	总bit数
IEEE-754 FP64 64位双精度浮点	1	11	52	64
IEEE-754 FP32 32位单精度浮点	1	8	23	32
IEEE-754 FP16 16位半精度浮点	1	5	10	16
BFloat16 (BF16) 64位半精度浮点	1	8	7	16

不同的计算场景会根据具体算法和精度要求，使用不同的浮点指令格式。科学和工程计算通常使用FP64双精度浮点和FP32单精度浮点，人工智能场景通常使用FP32单精度浮点、FP16或BF16半精度浮点，人工智能驱动的科学计算（AI for Science）等场景需要混合使用多种浮点指令。另外，不同指令在计算芯片上的实现代价是不同的，通常和位宽的平方成比例。

为了综合性评价算力基础设施的算力能力，定义如下“峰值综合算力”指标，单位仍然为FLOPS，该指标对不同精度的算力性能根据指令位宽进行了修正。公式如下：

$$CR_{peak} = \frac{64^2}{16^2} \times R_{fp64} + \frac{32^2}{16^2} \times R_{fp32} + R_{fp16} \dots\dots (1)$$

式中：

CR_{peak} ——峰值综合算力，单位为TFLOPS；

R_{fp64} ——峰值双精度浮点算力，单位为TFLOPS；

R_{fp32} ——峰值单精度浮点算力，单位为TFLOPS；

R_{fp16} ——峰值半精度浮点算力，单位为TFLOPS。

A.2 绿色节能指标

A.2.1 算效（CE）

算效指算力基础设施总算力与IT设备总功率的比值，即“算力基础设施每瓦功率所产生的算力”，是同时考虑算力基础设施计算性能与能效的一种效率。数值越大，代表单位功率得到的算力越强，效能越高。提升算效，可在相同算力水平下降低算力基础设施整体供电负担，降低整体的能耗。公式如下：

$$CE = \frac{CR_{\max}}{\Sigma PIT} \times 100\% \quad \dots\dots (2)$$

式中：
 CE——算效，单位为TFLOPS/kW；
 CR_{max}——有效综合算力，单位为TFLOPS；
 ΣPIT——算力基础设施在实际运算速度达到有效综合算力水平时，所有IT设备的功率，单位为kW。

A. 2. 2 存效（SE）

存效是衡量存储电能使用效率的指标，是存储有效容量（TB）除以存储设备总功率。公式如下：

$$SE = \frac{\text{Usable Capacity}}{CUN} \quad \dots\dots (3)$$

式中：
 SE——存效，单位为TB/KW；
 Usable Capacity——存储可用容量，单位为TB；
 CUN——算力基础设施在实际运算速度达到有效综合算力水平时，所有存储设备的功率，单位为千瓦（kW）。

A. 2. 3 能效（PUE）

能效是衡量算力基础设施的电能使用效率的指标，是整个算力基础设施的用电量与IT设备用电量的比值，其值越低，说明除IT设备外其他用电设备的用电量越少，算力基础设施越节能。由于不同季节的制冷设备的用电量有比较大的差别，因此需要基于全年的用电情况测算。公式如下：

$$PUE = \frac{E_{\text{total}}}{E_{IT}} \quad \dots\dots (4)$$

式中：
 PUE——算力基础设施电能利用效率；
 E_{total}——计算中心消耗总电能，单位为千瓦时（kWh）；
 E_{IT}——IT设备消耗电能，单位为千瓦时（kWh）。

如算力基础设施产生的热量进行余热回收，测算在同一个电量统计年度内余热回收的热能，并换算成电能E_{recovery}，单位为千瓦时（kW·h），此时能源再利用效率(Energy Reuse Effectiveness, ERE)的公式如下：

$$ERE = \frac{E_{\text{total}} - E_{\text{recovery}}}{E_{IT}} \quad \dots\dots (5)$$

式中：
 ERE——能源再利用效率；
 E_{total}——算力基础设施消耗总电能，单位为千瓦时（kWh）；
 E_{recovery}——全年余热回收热量的电能当量值，单位为千瓦（kW）；
 E_{IT}——IT设备消耗电能，单位为千瓦时（kWh）。

A. 2. 4 水利用效率（WUE）

水利用效率（Water Usage Effectiveness, 简称WUE）为算力基础设施用水的全年消耗量与算力基础设施IT设备全年耗电量的比值，单位为L/kW·h，用来评价算力基础设施用水状况的指标，公式如下：

$$WUE = \frac{W_{total}}{E_{IT}} \quad \dots\dots (6)$$

式中：

WUE——水利用效率，单位为升/千瓦时（L/kWh）；

W_{total} ——全年用水量，单位为升（L）；

E_{IT} ——IT设备消耗电能，单位为千瓦时（kWh）；

WUE计算包括两种概念，分别为源头WUE和站点WUE。通常使用站点WUE作为测量评价指标。

A. 2. 5 碳利用效率（CUE）

碳利用效率（Carbon Usage Effectiveness, 简称CUE）是可量化的算力基础设施碳排放指标，用于衡量算力基础设施的碳排放量，是一个重要的可持续性指标，直接与能源生产来源联系在一起。CUE为算力基础设施CO₂总排放量与IT负载耗电量的比值，CUE 单位是每千瓦时（kWh）的二氧化碳千克数（kgCO₂eq），公式如下：

$$CUE = \frac{E_{CO_2}}{E_{IT}} = \frac{CEF \times E_{total}}{E_{IT}} = CEF \times PUE \quad \dots\dots (7)$$

式中：

CUE——碳利用效率；

E_{CO_2} ——算力基础设施CO₂全年排放量，指核算各个源头的能源(如电、天然气、柴油等)经过CO₂当量转换的排放量，单位为kgCO₂eq；

E_{IT} ——IT设备消耗电能，单位为千瓦时（kWh）；

CEF——基于政府发布的当年运营区域算力基础设施的碳排放因子，单位为kgCO₂eq/kWh；

A. 2. 6 可再生能源利用率（RER）

RER(Renewable Energy Ratio, 可再生能源利用率)用于衡量算力基础设施利用可再生能源的情况，以促进可再生、无碳排放或极少碳排放的能源利用。RER的计算公式如下：

$$RER = \frac{E_{renew}}{E_{total}} \quad \dots\dots (8)$$

式中：

RER——可再生能源利用率

E_{renew} ——可再生能源全年供电量，单位为千瓦时（kWh）；

E_{total} ——算力基础设施消耗总电能，单位为千瓦时（kWh）。

A. 2. 7 服务器设备能效（SPUE）

服务器设备能效用于衡量服务器设备运算能耗的能力。

$$SPUE = \frac{E_{IT}}{E_{operation}} \quad \dots\dots (9)$$

式中：

SPUE——服务器设备能效；

E_{IT} ——IT设备消耗电能，单位为千瓦时（kWh）；

$E_{operation}$ ——运算相关能耗总电能，包括CPU、服务器主板、内存条、磁盘、IO等功能组件的能源消耗，单位为千瓦时（kWh）。



参考文献

- [1] DB11/T 2052—2022 绿色数据中心评价指标与方法
- [2] T/CCUA 016—2021 超级计算数据中心设计要求
- [3] 工信部联通信〔2023〕180号 算力基础设施高质量发展行动计划
- [4] 发改数据〔2023〕1779号 关于深入实施“东数西算”工程加快构建全国一体化算力网的实施意见
- [5] 绿色网格TGG, 2010:6-8 ERE：衡量数据中心能源再利用效益的指标（第29号白皮书）

