

团 体 标 准

T/CQJR 032.1—2025

金融业算力应用平台的计算效用提升技术 要求及评估方法 第1部分：通用要求

Technical Requirements and Evaluation Methods for Computing Usage Efficiency
Improvement of Computing Power Application Platform in Finance - Part1:
General Requirements

2025 - 08 - 08 发布

2025 - 08 - 08 实施

目 次

前 言	2
引 言	3
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 金融业算力应用平台计算效用提升技术能力框架	5
6 金融业算力应用平台计算效用提升技术要求	6
6.1 计算机效用监控评估能力	6
6.2 算力资源全局调度优化能力	6
6.3 工作负载算力消耗优化能力	7
6.4 算力资源选择和迁移能力	7
7 金融业算力应用平台计算效用评价	7
7.1 概述	7
7.2 评价方法	8
附录 A （资料性附录） 金融业算力应用平台计算效用水平的评价	9
参 考 文 献	10

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/CQJR 032《金融业算力应用平台的计算效用提升技术要求及评估方法》的第1部分，T/CQJR 032已经发布了以下部分：

- 第1部分：通用要求；
- 第2部分：消费金融。

本文件由重庆蚂蚁消费金融有限公司、蚂蚁科技集团股份有限公司、重庆国家金融科技认证中心提出。

本文件由重庆市金融学会归口。

本文件起草单位：重庆蚂蚁消费金融有限公司、蚂蚁科技集团股份有限公司、重庆国家金融科技认证中心、浙江网商银行股份有限公司、重庆银行股份有限公司、重庆农村商业银行股份有限公司、重庆富民银行股份有限公司、重庆三峡银行股份有限公司、马上消费金融股份有限公司、重庆小米消费金融有限公司。

本文件主要起草人：刘义、彭晋、王义、孙曦、蒋俊杰、陈彬、秦逞、吴娟、李强、罗丹、吕科苇、胡元瑞、官仓璘、张成松、陈南。

引 言

在数字经济蓬勃发展的时代背景下，算力已成为支撑各行业创新发展的关键基础设施，金融领域尤为显著。随着大数据分析、人工智能、风险控制等技术的深度应用，金融行业算力需求呈爆发式增长，与之伴随的能耗总量亦急剧攀升。在全球积极推进“碳达峰、碳中和”目标的大环境下，实现算力的绿色低碳发展已成为金融行业可持续发展的必由之路。

然而，当前金融领域对算力绿色低碳的探索主要集中于数据中心基础设施层面，如可再生能源的应用和绿色数据中心建设，对业务应用环节中算力利用效率较低的问题重视不足。实际业务中，金融核心业务服务器利用率呈现明显的高峰低谷特征，导致大量算力资源在非高峰时段闲置，造成能源浪费。国际数据中心认证标准组织Uptime Institute研究显示，美国30%的服务器处于“休眠”状态，这一现象在重庆金融行业同样普遍存在，暴露出金融业在算力应用环节存在巨大的优化空间。

与此同时，金融行业尚未建立针对算力应用平台效率的统一标准，缺乏系统性提升计算效用的技术要求、评价指标及方法，严重制约了算力利用效率的进一步提升。为推动金融行业算力资源的高效利用，助力实现绿色低碳发展目标，亟需总结和提炼行业内算力应用平台计算效用提升经验，形成具有指导性和可操作性的技术标准，为金融机构优化资源配置、降低运营成本、提升服务质量提供有力支撑。

本文件拟由2个部分组成。

- 第1部分：通用要求，规定了用于金融业务的算力应用平台在提升计算效用过程中可参照执行的技术要求、衡量算力应用平台计算效用水平的评价指标和相应的评估方法。
- 第2部分：消费金融，规定了用于消费金融业务的算力应用平台在提升计算效用过程中可参照执行的技术要求、衡量算力应用平台计算效用水平的评价指标和相应的评估方法。

金融业算力应用平台的计算效用提升技术要求及评估方法

第1部分：通用要求

1 范围

本文件规定了用于金融业务的算力应用平台在提升计算效用的过程中可参照执行的技术要求以及衡量算力应用平台计算效用水平的评价指标和相应的评估方法。

本文件适用于金融业的算力应用平台的提供方和使用方参照使用进行相关软硬件的开发、运维、评估和持续优化,也适用于专业评测机构参考本文件定义的评价指标和评价方法对金融业算力应用平台的计算效用水平和提升程度开展评估测试。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅所注日期的版本适用于本文件。凡是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 34982—2017 云计算数据中心基本要求

ISO/IEC 23544:2021 信息技术 数据中心 应用平台能源效率(APEE) (Information Technology - Data centres - Application Platform Energy Effectiveness (APEE))

3 术语和定义

下列术语和定义适用于本文件。

3.1

算力应用平台 `computing power application platform`

一种从算力应用视角定义的平台系统,通过为业务应用提供管理和使用算力资源的机制,以支持各类业务应用工作负载的执行。

3.2

计算效用 `computing usage efficiency`

算力应用平台上,硬件算力资源被实际业务应用有效利用的程度。

3.3

持续性能分析 `continuous profiling`

一种在生产环境中持续收集和分析应用程序性能数据的技术,通过低开销的采样方法,定期收集和分析系统资源的使用情况,以期识别性能瓶颈和优化点。

3.4

工作负载 `workload`

一系列资源和代码，面向应用场景完成承载商业价值的业务计算任务，如面向客户的应用程序等。

3.5

碳排放量 carbon emissions

在特定时空边界内，由人类活动直接或间接产生的温室气体总量，一般以二氧化碳当量（CO₂e）为计量单位。

4 缩略语

下列缩略语适用于本文件。

SCI：软件碳强度（Software Carbon Intensity）

CPU：中央处理器（Central Processing Unit）

GPU：图形处理器（Graphic Processing Unit）

PUE：电能利用效率（Power Usage Effectiveness）

FLOPS：每秒浮点运算次数（Floating-point Operations Per Second）

TOPS：每秒万亿次运算（Tera Operations Per Second）

SLA：服务水平协议（Service Level Agreement）

IOPS：每秒的读写次数（Input/Output Operations Per Second）

5 金融业算力应用平台计算效用提升技术能力框架

金融业算力应用平台计算效用提升技术能力框架见图1。

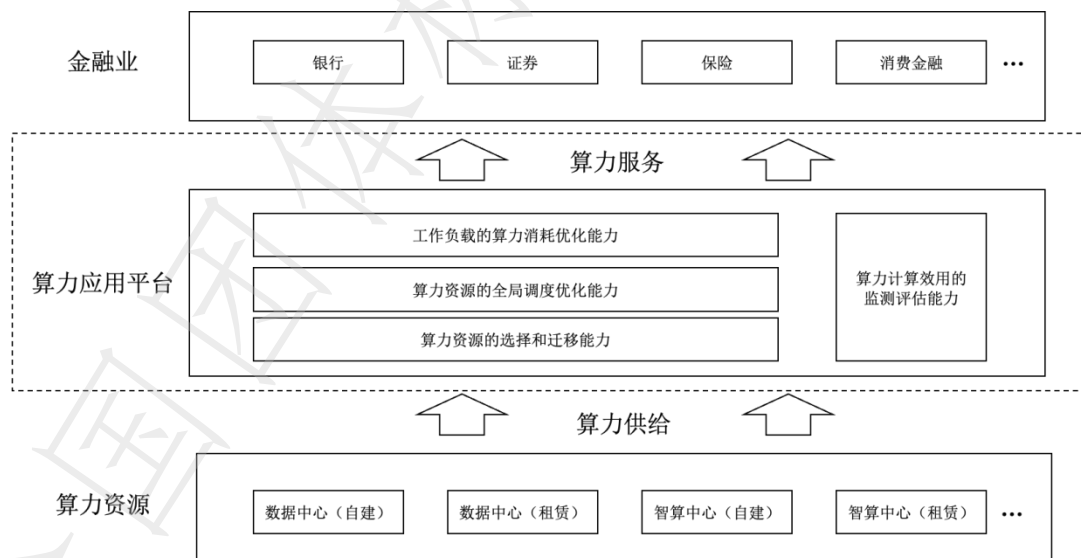


图1 金融业算力应用平台计算效用提升技术能力框架

如图1所示，在将算力资源应用到金融业时，算力应用平台一方面可对接和整合各类算力资源的供给，包括但不限于数据中心（可自建或租用）、智算中心（可自建或租赁）等；另一方面，面向金融业各子业态（如银行、证券、保险、消费金融等）的典型业务场景，算力应用平台应屏蔽底层算力资源的类型和来源差异，并提供统一的算力服务来支撑具体业务场景的实现。

对于金融业算力应用平台提供方来说，为能够建立起计算效用提升的持续改进流程，应着重建设下述核心能力（如图1中的虚线框所示），包括但不限于：

- a) 计算效用的监控评估能力：具备对平台的计算效用水平进行观测和度量的能力，并能够转化成相应的碳排放量；
- b) 算力资源的全局调度优化能力：通过算力应用平台上的各类操作系统、中间件等高质量软件，对算力资源的全局分配与调度进行优化的能力；
- c) 工作负载的算力消耗优化能力：结合金融典型业务场景，通过软硬件协同的设计思路，使得完成相同规模业务量所消耗算力更低的能力；
- d) 算力资源的选择和迁移能力：具备在不同算力资源供给间的灵活选择及迁移能力。

6 金融业算力应用平台计算效用提升技术要求

6.1 计算机效用监控评估能力

金融业算力应用平台的计算效用监控评估能力要求包括但不限于：

- a) 应设定计算效用的优化目标，并使用定量的评估指标来识别优化目标的实现情况；
 - 1) 宜使用章节 7.1 所定义的评估指标进一步衡量算力应用平台的计算效用水平；
 - 2) 可统计算力应用平台在运营过程中所产生的碳排放量。
- b) 确定优化目标和评估指标后，应相应地构建计算效用的监测与统计机制，对算力应用平台的计算效用进行有效监测；
- c) 算力应用平台在统计运营过程中所产生的碳排放量时；
 - 1) 可按照范围 1、范围 2 或范围 3 分别计算碳排放量；
 - 2) 宜按照一定周期定期生成碳排放报告，如按照月度、季度或者年度；
 - 3) 宜使用可视化方式呈现碳排放数据情况，帮助理解和优化碳排放。
- d) 宜计算出在计算效用提升情况下等效减少的碳排放量。

6.2 算力资源全局调度优化能力

金融业算力应用平台的算力资源全局调度优化能力要求包括但不限于：

- a) 硬件虚拟化；
 - 1) 对于不同品牌、不同性能的底层算力硬件，应使用虚拟化等技术实现对不同算力硬件的封装；
 - 2) 应向上层提供统一的算力硬件访问接口，屏蔽底层硬件差异；
 - 3) 应支持对虚拟化资源的生命周期管理，包括创建、销毁、迁移等；
 - 4) 应实现对 CPU、GPU 等硬件资源的虚拟化隔离，确保安全性和性能隔离。
 - 5) 应支持硬件资源热插拔，实现资源动态调度。

示例：以 CPU 为例，在对底层 CPU 硬件芯片差异进行封装后，会按照 CPU 核数作为统一的算力单位对上层金融业务提供服务。基于业务性能需求，为业务分配相应的 CPU 核数（算力单位），实际上该业务可能运行在不同 CPU 上。

- b) 算力抽象；
 - 1) 应定义明确的算力单位，如 FLOPS、TOPS 等；
 - 2) 应按照所支持的算力单位对上层应用提供算力调用服务；
 - 3) 宜对算力硬件的算力归一化程度进行评估，并能够通过优化措施减少算力硬件归一化后的算力评估差异。

示例：可通过相关的性能测试工具对归一化之后的 CPU 硬件算力进行评估，评估值与实际值差异越小，表示归一化

程度越好。

- c) 算力调度；
 - 1) 应构建统一的算力调度框架，实现跨异构硬件的算力编排；
 - 2) 应构建算力按需分配能力，根据不同的业务需求匹配提供不同的算力类型；
 - 3) 应具备在集群内灵活调度不同工作负载的能力，并对调度时间进行评估。调度的时间越短，则表示能力越强；
 - 4) 应实现业务应用软件与算力资源的解耦，即业务应用软件的运行不依赖于特定来源的算力；
 - 5) 宜建立算力需求的预测能力，并对预测效果进行评估，例如预测出来的算力需求量和实际的业务需求量相比差值越小，预测能力越佳。
- d) 宜采用云原生技术，实现业务与算力资源的灵活匹配与调度，并按需分配计算、存储、网络等算力资源，具体可使用的云原生关键技术包括但不限于：
 - 1) 通过容器化技术实现被配置和调度的业务应用对象的标准化管理；
 - 2) 通过容器编排技术实现容器调度能力和算力资源自动弹性伸缩的能力；
 - 3) 通过服务网格技术实现快速的应用流量调度和微服务弹性管理。
- e) 宜支持离线混合部署技术，将在线应用和离线任务编排部署在一个节点，通过安全容器隔离、资源共享、资源抢占等技术保障在线应用运行时不受离线任务干扰，并利用离线任务充分使用在线应用的空闲算力资源。

6.3 工作负载算力消耗优化能力

金融业算力应用平台的工作负载算力消耗优化能力要求包括但不限于：

- a) 应对算力应用平台上的主要工作负载类型进行识别和划分，并统计和分析当前金融业务场景的主要工作负载情况；
- b) 应针对主要的工作负载类型的特点，采取相应的技术手段优化算力消耗。

6.4 算力资源选择和迁移能力

金融业算力应用平台应具备算力资源的选择和迁移能力，技术要求主要包括但不限于：

- a) 应能识别出不同算力资源的碳强度，识别方式包括但不限于：
 - 1) 基于算力资源供给方提供的查询接口获取；
 - 2) 根据算力资源所使用的上游可再生绿色能源占比、算力产生过程中所使用的不同硬件基础功耗水平等进行折算。
- b) 宜能够根据识别出的不同算力资源的碳强度情况，支持对工作负载的灵活调度或迁移，优先使用具有更低碳强度的算力资源来支撑业务；
- c) 在不同算力资源上进行迁移时应保证业务的 SLA 不受影响，即在更换算力资源时，所支撑的金融业务和数据在逻辑上应不受到影响，并能够符合相应的数据安全、备份与容灾要求、业务连续性要求等。

7 金融业算力应用平台计算效用评价

7.1 概述

可以从不同角度对金融业算力应用平台的计算效用水平进行观测，包括但不限于：

- a) 基于某一类型的算力硬件利用效率进行观测，如 CPU 利用效率、GPU 利用效率、存储 IOPS 等；

- b) 基于多种类型的算力硬件利用效率进行观测，并加权计算出总体的利用效率；
- c) 在某类金融子业态（如银行、证券、保险、消费金融等）下，基于某个金融典型业务场景进行观测，即完成该类金融子业态的某个业务场景的单位业务所需要消耗算力水平；
- d) 在某类金融子业态（如银行、证券、保险、消费金融等）下，综合该金融子业态的所有典型业务场景进行观测，即结合业务实际运营过程中各典型业务场景的占比情况，通过加权计算出单位业务所需消耗的算力水平。

7.2 评价方法

对适用于不同金融子业态（如银行、证券、保险、消费金融等）的计算效用评价指标的具体定义，可见本系列标准的后续部分。

可使用附录A中的评估方法，对金融业算力应用平台计算效用水平进行评价。

附 录 A
(资料性附录)
金融业算力应用平台计算效用水平的评价

对金融业算力应用平台计算效用水平的评价，可参考下表定义的计算方法得到相应的得分。

表 A.1 金融业算力应用平台计算效用水平的评价打分规则

分类	指标名称	分数
C1: 计算效用监控评估能力	C1.1 优化目标制定	20分
	C1.2 计算效用监测与统计落实	
	C1.3 平台碳排放量统计	
	C1.4 平台碳减排量统计	
	C1.5 工作负载算力消耗监测与统计	
	C1.6 细颗粒度消耗监测与统计	
	C1.7 算力智能趋势预测	
C2: 算力资源全局调度优化能力	C2.1 硬件虚拟化	20分
	C2.2 算力抽象	
	C2.3 算力调度	
	C2.4 云原生水平	
	C2.5 在离线混合部署	
	C2.6 算力容灾	
C3: 工作负载算力消耗优化能力	C3.1 工作负载类型识别与划分	10分
	C3.2 依工作负载特点优化算力消耗	
C4: 算力资源选择和迁移能力	C4.1 不同算力来源的碳强度识别	10分
	C4.2 工作负载层面的算力灵活迁移	
	C4.3 总体层面的算力平滑迁移	
	C4.1 不同算力来源的碳强度识别	
M1: 计算效用评价指标	M1.1 采样范围	40分
	M1.2 采样周期	
	M1.3 评价指标绝对值水平	
	M1.4 评价指标改进相对值水平	

参 考 文 献

- [1] GB/T 34982-2017 云计算数据中心基本要求
- [2] ISO/IEC 23544:2021 Information Technology — Data centres — Application Platform Energy Effectiveness (APEE)
-

全国团体标准信息平台