

DB4403

深 圳 市 地 方 标 准

DB4403/T XXX—XXXX

科研算力共享技术规范

Technical specifications for sharing scientific research computing power

(送审稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

 3.1 算力集群（Computing Power,CP） 1

 3.2 算力集群供给方（Computing Power Cluster Supplier,CPCS） 1

 3.3 安全外壳（Secure Shell,SSH） 2

 3.4 平台服务（Platform as a Service,PaaS） 2

 3.5 虚拟网络计算（Virtual Network Computing, VNC） 2

 3.6 A 类算力集群 2

 3.7 B 类算力集群 2

4 缩略语 2

5 科研算力共享规范技术架构 2

 5.1 概述 2

 5.2 技术架构图应用场景 3

 5.3 用户 4

 5.4 共享平台 4

 5.5 算力集群 5

 5.6 安全与可靠性要求 5

6 共享平台技术要求 5

 6.1 概述 5

 6.2 接口技术要求 6

 6.3 用户管理 6

 6.4 算力集群供给方管理 8

 6.5 算力集群管理 8

7 算力集群技术要求 10

 7.1 分类 10

 7.2 A 类算力集群技术要求 10

 7.3 B 类算力集群技术要求 11

参考文献 12

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。

本文件由深圳市工业和信息化局提出并归口。

本文件起草单位：深圳市高校教育信息化学会、深圳大学。

本文件主要起草人：张继群、谭帅帅、张胜利、钱恭斌、范士喜、姜来、杨海琨、赵永刚、江普、刘吉、陈兰、吴嘉亮、何海涛、赵晓栋、谢应双、阮华君、牟云强、林沐、陈园、陈树靖、邓敏学、李莎、朱博、韦植桐、刘航、陈锦繁、王又民、白熙、黄新宇、丘敏均。

科研算力共享技术规范

1 范围

本技术规范规定了深圳市科研算力共享体系的架构和技术要求，旨在明确科研算力资源的标准化管理和高效利用，确保科研活动的顺利进行和科研数据的准确处理。

本技术规范适用于深圳市行政区划范围内所有涉及科研算力共享的相关机构、单位和个人。

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 20984—2022 信息安全技术 信息安全风险评估方法
GB/T 22239—2019 信息安全技术 网络安全等级保护基本要求
GB/T 22240—2020 信息安全技术 网络安全等级保护定级指南
GB/T 25070—2019 信息安全技术 网络安全等级保护安全设计要求
GB/T 28448—2019 信息安全技术 网络安全等级保护测评要求
GB/T 35273—2020 信息安全技术 个人信息安全规范
GB/T 35293—2017 信息技术 云计算 虚拟机管理通用要求
GB/T 36326—2018 信息技术 云计算 云服务运营通用要求
GB/T 37988—2019 信息安全技术 数据安全能力成熟度模型
GB/T 39786—2021 信息安全技术 信息系统密码应用基本要求

ISO/IEC 27001:2022 信息安全、网络安全和隐私保护—信息安全管理—要求
(Information security, cybersecurity and privacy protection — Information security management systems — Requirements)

ISO/IEC 29100:2024 信息技术—安全技术—隐私框架 (Information technology — Security techniques — Privacy framework)

T/SZAS 86—2024 大湾区算力网络总体技术要求

3 术语和定义

下列术语和定义适用于本文件。

3.1 算力集群 (Computing Power, CP)

由多台高性能计算节点通过网络连接组成的大规模计算系统，配备高性能处理器、加速器、大容量内存及高速存储，以提供强大、灵活的计算能力。

3.2 算力集群供给方 (Computing Power Cluster Supplier, CPCS)

在算力共享服务体系中，拥有并提供算力资源的实体或组织，包括但不限于云服务厂商、电信运营商、超算与智算中心，以及拥有闲置算力资源的中小企业和个人。

3.3 安全外壳 (Secure Shell, SSH)

一种网络协议，用于在不安全的网络中为网络服务提供安全的传输层。SSH通过加密和认证机制，确保了数据传输的机密性和完整性，同时防止了中间人攻击和数据窃听。

[来源：GM/T 0129—2023，3.1]

3.4 平台服务 (Platform as a Service, PaaS)

一种云计算服务模式，它提供了一个完整的开发、部署和运行应用程序的平台，保证应用程序的高可用性，而用户无需关心底层硬件和操作系统的复杂性。

[来源：GB/T 44067.2—2024，2.1]

3.5 虚拟网络计算 (Virtual Network Computing, VNC)

一种图形桌面共享系统，允许用户通过图形界面远程访问和控制另一台计算机。

3.6 A类算力集群

没有虚拟化层的物理服务器，用户独享整个服务器的硬件资源。

3.7 B类算力集群

在虚拟化环境中运行的服务器，多个用户可以共享同一物理服务器的资源。

4 缩略语

下列缩略语适用于本文件。

CPU：中央处理器 (Central Processing Unit)

GPU：图形处理器 (Graphics Processing Unit)

SSD：固态硬盘 (Solid State Disk)

HDD：机械硬盘 (Hard Disk Drive)

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

Ius：用户服务接口 (User Service Interface)

Irmids：资源监控数据服务接口 (Resource monitoring data service Interface)

Ivcpc：算力集群可视化接口 (Visualization interface for computing power cluster)

Ipcss：算力集群供给方服务接口 (Computing power cluster sliders service interface)

Itm：任务管理接口 (Task management interface)

CPcId：算力集群标识 (Computing power cluster Id)

CUDA：统一计算设备架构 (Compute Unified Device Architecture)

CUDNN：CUDA深度神经网络库 (CUDA Deep Neural Network library)

5 科研算力共享规范技术架构

5.1 概述

技术架构示意图详见图1，该架构由用户、共享平台、算力集群以及安全要求四大核心部分构成。用户层涵盖了个人用户、高等教育机构用户、科研机构用户及企业用户等多类型用户群体。共享平台作为整个技术架构的中枢环节，全面肩负起用户信息管理、算力集群供给方管理以及算力集群管理的核心职责。通过北向API，平台能够有效响应并满足用户需求；借助南向API，平台则实现了对算力资源的管理与任务监控。算力集群分为裸金属服务器（A类）与虚拟机服务器（B类）两大类。在此架构体系中，安全性被视为保障共享平台及算力集群持续稳定运行不可或缺的关键支撑要素。

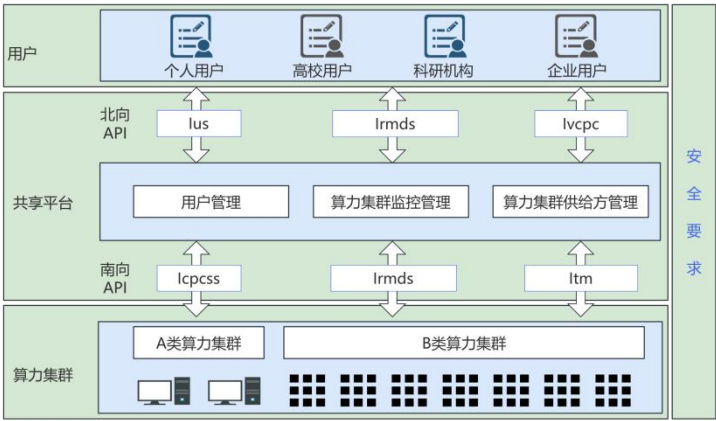


图 1 技术架构图

5.2 技术架构图应用场景

用户访问共享平台的应用场景如图2所示。

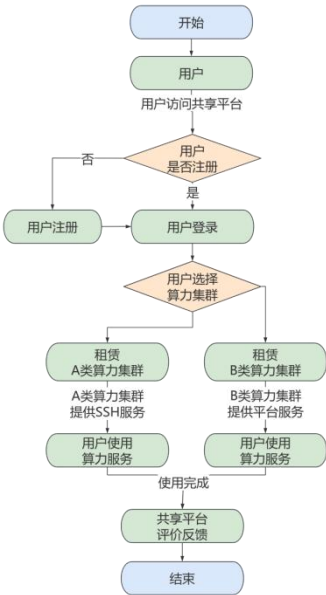


图 2 用户应用场景

用户首先检查自己的注册状态，若未注册则进行注册，随后登录平台查看可用的算力集群信息，并根据需求租赁相应的集群。对于A类算力集群，用户可通过SSH服务（或备选VNC服务）进行访问；而对于B类算力集群，则提供平台服务供用户使用。

算力集群供给方访问共享平台的应用场景如图3所示。

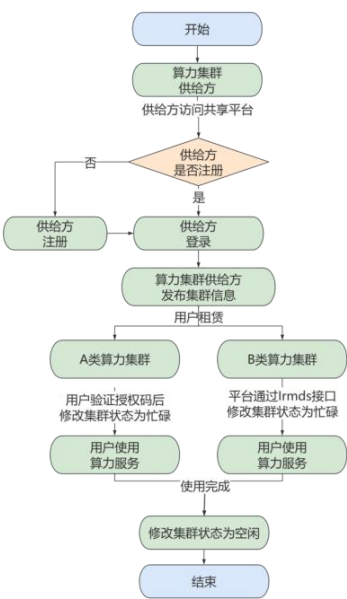


图 3 供给方应用场景

算力集群供给方也需先检查注册状态并完成注册，登录后发布其可用的集群信息。当用户选择租赁某集群时，A类供给方在用户验证授权码后修改集群状态为忙碌，而B类供给方则通过Irmuds接口自动完成此操作。待用户使用完算力服务后，供给方需再次修改集群状态，以反映其当前可用性。

5.3 用户

用户：架构的顶层是用户层，它包括了多类用户群组，如“个人用户”“高校用户”“科研机构”和“企业用户”。

5.4 共享平台

5.4.1 概述

共享平台作为技术架构的核心组成部分，全面且深入地覆盖了用户管理、算力集群供给方管理以及算力集群管理的各项关键功能。通过北向 API 的集成，平台为用户提供了包括用户服务（Ius）、资源监控（Irmuds）以及算力集群可视化（Ivcpc）在内的多元化接口，旨在全方位满足用户的多样化需求。同时，借助南向 API 的支撑，平台为算力集群供给方构建了包括供给方服务（Icpcss）和任务管理（Itm）在内的精细化接口，以实现对算力资源的优化管理和任务调度的智能化监控。

5.4.2 共享平台接口

5.4.2.1 北向 API

Ius：用户服务接口；Irmuds：资源监控数据服务接口；Ivcpc：算力集群可视化接口。

5.4.2.2 南向 API

Icpcss：算力集群供给方服务接口；Itm：任务管理接口。

5.4.3 共享平台功能

用户管理：用户注册，登录，信息修改以及注销功能。

算力集群供给方管理：算力集群供给方注册，登录，信息修改以及注销功能。

算力集群管理：算力集群发布，管理，修改，下架，筛选，任务管理以及可视化功能。

5.5 算力集群

算力集群主要分为A类算力集群：裸金属服务器；B类算力集群：虚拟机服务器。

建议A类算力集群为用户提供SSH远程登录服务或VNC远程控制工具软件，实现算力集群共享。

建议B类算力集群为用户提供平台服务，实现算力集群共享。

5.6 安全与可靠性要求

5.6.1 总则

安全与可靠性要求是确保共享平台及算力集群稳定运行的关键。

5.6.2 共享平台安全要求

共享平台安全要求应包括以下内容：

- a) 应遵守《中华人民共和国网络安全法》《中华人民共和国数据安全法》等法律法规；
- b) 应满足网络安全等级保护制度的要求，并符合 GB/T 22239 的规定；
- c) 应建立严格的访问控制机制，实施用户身份认证与权限审核流程，确保用户只能访问其被授权的资源和服务；
- d) 应定期进行安全漏洞扫描与修复；
- e) 应确保用户数据和算力集群数据在收集、存储、处理、传输等全生命周期内的安全与保密；
- f) 宜引入多因素认证（MFA）机制，提高账户安全性；
- g) 宜部署防火墙、入侵检测系统（IDS）或入侵防御系统（IPS）等网络安全设备。

5.6.3 算力集群可靠性要求

算力集群可靠性要求应包括以下内容：

- a) 宜定期对集群数据、业务数据和关键配置进行备份，并制定有效的数据恢复策略；
- b) 宜部署实时安全监控与预警系统，监测集群运行状态与安全状况，及时预警潜在威胁；
- c) 应加强物理环境的安全防护措施，包括机房门禁、视频监控、防火、防雷击、防水和防潮等；
- d) 宜取得 ISO/IEC 27001 等信息安全国际标准认证。

6 共享平台技术要求

6.1 概述

共享平台的核心功能如图 4 所示，该平台紧密围绕三大核心模块构建：用户管理、算力集群供给方管理以及算力集群管理。同时，它还集成了北向 API 与南向 API，以实现对外部服务与内部资源的高效对接与整合。

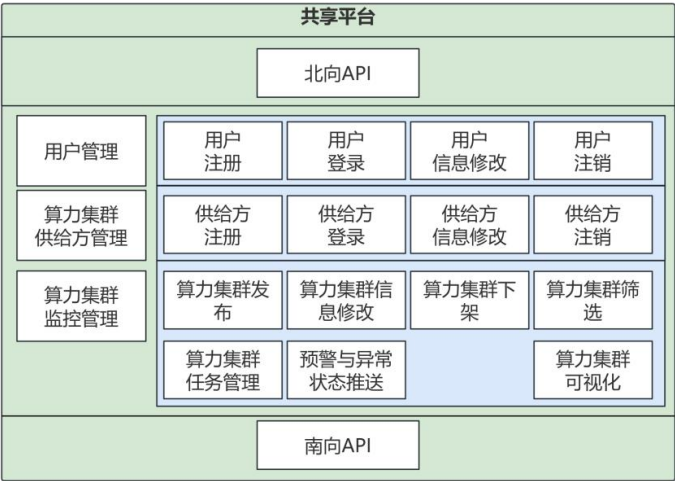


图 4 共享平台核心功能

6.2 接口技术要求

6.2.1 北向 API

Ius：用户服务接口。Ius实现用户注册、用户登录、用户信息修改以及用户注销等功能。此外，Ius还预留算力集群地域信息字段，使得用户和算力集群之间能够按照区域进行匹配。

Ivcpc：算力集群可视化接口。在算力集群发布后，Ivcpc负责将资源的详细信息，包括资源类型、型号、容量、内存、虚拟核心数、单价、计费模式、供应商、位置信息及本地存储配置等，回传给用户层进行展示。

Irmids：资源监控数据服务接口。Irmids负责实时采集正在运行的算力集群的各项指标，如CPU使用率、内存占用、GPU性能等。为后期B类算力集群的服务评价奠定数据基础。

6.2.2 南向 API

Icpcss：算力集群供给方服务接口。Icpcss实现供给方注册、供给方登录、供给方信息修改以及供给方注销等功能。

Itm：任务管理接口。Itm允许系统高效地创建并记录各种任务数据，任务数据包括但不限于算力集群信息、用户信息、计费信息等。

Irmids：资源监控数据服务接口。

6.2.3 API 实现要求

数据标准化：确保所有接口返回的数据都遵循统一的格式和单位标准，便于前端展示和跨平台使用。

文档与测试要求：编写详细的接口文档，包括接口描述、请求参数、响应格式、错误码等信息，并进行充分的测试，确保接口的正确性和可靠性，符合中国信通院发布的《2023年API安全发展白皮书》和《安全防范人脸识别应用程序接口规范》。

6.3 用户管理

6.3.1 概述

用户管理涵盖了用户注册、登录、信息修改及注销等操作。在此，我们将重点阐述用户注册与登录流程中涉及的用户属性信息及其实施细节。

6.3.2 用户注册

6.3.2.1 用户注册信息

用户注册信息包括但不限于：用户名（或邮箱/手机号）、密码、确认密码、姓名、身份证号、验证码等（用户身份证号等敏感信息采用哈希加密算法存储，遵循相关数据保护法规，保障用户信息安全），用户类型：个人用户，高校用户，科研机构，企业用户。

6.3.2.2 用户注册流程

用户发起注册请求：首先用户提交必要的注册信息字段。同时，提供注册协议和隐私政策链接，确保用户了解注册过程中的条款和条件。注册协议和隐私政策的制定应符合国家关于个人信息保护和网络安全的相关要求，可见《中华人民共和国个人信息保护法》《中华人民共和国网络安全法》等相关法律规定。

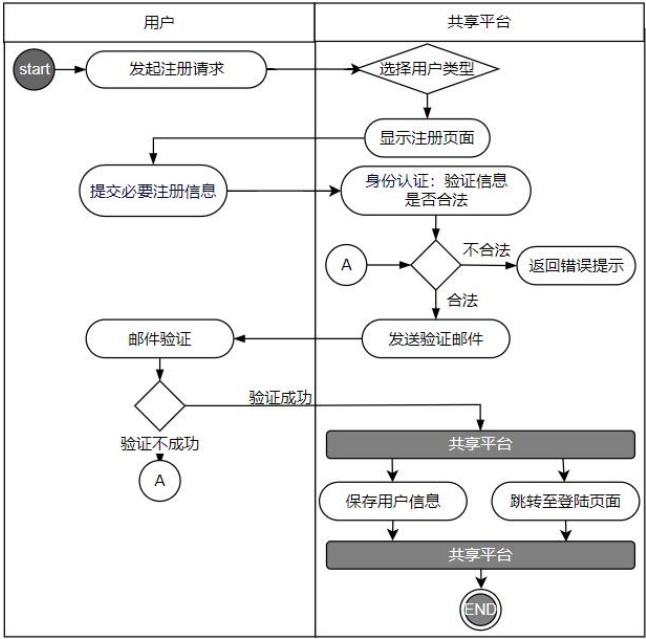


图 5 用户注册流程图

共享平台进行身份认证：在前端进行初步的数据验证，如检查输入是否为空、密码强度是否达标、两次输入的密码是否一致、身份信息是否正确等，对高校用户、科研机构和企业用户实施更严格的身份审核流程（提交单位营业执照（法人签印））。提交表单后，在后端进行更严格的数据验证，包括检查用户名（或邮箱/手机号）是否已存在、验证码是否正确、营业执照合法。

账户激活：共享平台发送注册成功通知给用户，包含账户激活链接。未经验证的账户无法登录或访问某些功能。

6.3.3 用户登录

登录验证：用户提交登录信息后，平台在后端进行验证。首先检查用户名（或邮箱/手机号）是否存在，若密码一致，则登录成功。

会话管理：登录成功后，平台生成一个会话令牌，并将其返回给用户。用户需要在后续的请求中携带此令牌以进行身份验证。平台应支持会话的创建、存储、验证和销毁，确保会话的安全性。

多因素认证（建议要求）：为了提高安全性，可以引入多因素认证机制，如手机验证码、指纹识别等。

6.4 算力集群供给方管理

6.4.1 概述

算力集群供给方管理涵盖了供给方注册、登录、信息更新及注销等操作。在此，我们将重点阐述供给方注册流程中涉及的供给方属性信息及其实施细节。

6.4.2 算力集群供给方注册

6.4.2.1 算力集群供给方注册信息

算力集群的供给方可被划分为两类：企业供给方与个人供给方。对于企业供给方所需提交的包括但不限于：企业名称、法定代表人的身份信息、注册资本的具体数额、统一社会信用代码这一官方识别码、有效的联系方式、营业执照副本的提交以及用于对公付费的对公账户（或个人实名认证账户，视具体情况而定）等核心要素。

对于个人供给方所需提交的包括但不限于：身份证号码、真实姓名、居住地址、有效的联系方式以及所属单位（如有）等在内的个人信息。

6.4.2.2 算力集群供给方注册流程

算力集群供给方的注册流程如图6所示，其详细步骤与6.3.2.2部分所述相似，在此处不再进行重复阐述。

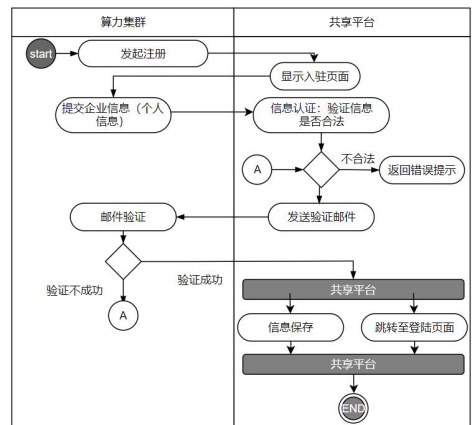


图 6 算力集群注册流程图

6.5 算力集群管理

6.5.1 概述

围绕供给方（即算力集群的供给方）与用户（即算力集群的需求方）的需求而展开的。供给方通过共享平台实现算力集群的发布、信息修改及下架；用户则通过共享平台筛选功能，筛选满足需求的集群。

同时，共享平台实现任务管理，跟踪并记录忙碌状态下算力集群的使用情况。

6.5.2 算力集群发布

算力集群供给方应将其算力集群信息上传至共享平台进行发布。算力集群发布要求包括以下内容：

- a) 算力集群信息应包括但不限于以下内容：
 - 1) 算力集群类型（如A类、B类）；
 - 2) 资源类型（如GPU）；
 - 3) 资源型号（如8*A100 80GB）；
 - 4) 容量（如80GB）；
 - 5) 内存容量（如1TB）；
 - 6) 虚拟核心数（如128核）；
 - 7) 单价（如42,500.00元/月）；
 - 8) 计费模式（如包年包月）；
 - 9) 供应商（如华为云）；
 - 10) 位置信息（如深圳市）；
 - 11) 本地存储配置（包括SSD类型，如3.84TB*2数据盘与480GB系统盘）。
- b) 算力集群中的每张显卡应视为一个独立的算力单位，并可单独设置其状态（如“空闲”与“忙碌”）；
- c) 算力集群供给方应综合多种因素合理制定并提供单价信息，并及时在平台更新；
- d) 共享平台应允许用户自主选择算力服务计费模式；
- e) A类算力集群的供给方还应提供算力集群的具体使用方式；
- f) 当供需双方成功达成使用协议后，用户应依据供给方提供的授权码，在共享平台上查看调用方式（如SSH远程调用的专用账户及密码）；
- g) 用户成功获取调用方式后，此算力集群的状态应即时更新为“忙碌”，标志着该集群资源已处于被占用状态；
- h) 算力集群供给方发布算力集群信息后，平台应为该集群生成唯一标识CPcId。

6.5.3 算力集群信息修改

随着业务需求或集群状态（“空闲”与“忙碌”）的变化，供给方需能够灵活调整集群信息，如更新集群配置、描述或性能参数等，以确保信息的准确性和时效性（A类算力集群供给方与用户达成租赁协议后，用户在平台校验授权码后，平台更新当前集群状态为忙碌。B类算力集群供给方通过Irmids接口自动更新集群状态）。

6.5.4 算力集群下架

当集群不再满足业务需求或需进行维护时，供给方需将其从共享平台下架，以防止用户选择不可用的资源。

6.5.5 算力集群筛选

用户可根据自身需求，如计算能力、存储容量、价格等，在共享平台上对算力集群进行筛选。

6.5.6 算力集群任务管理

算力集群任务管理流程如图7所示。

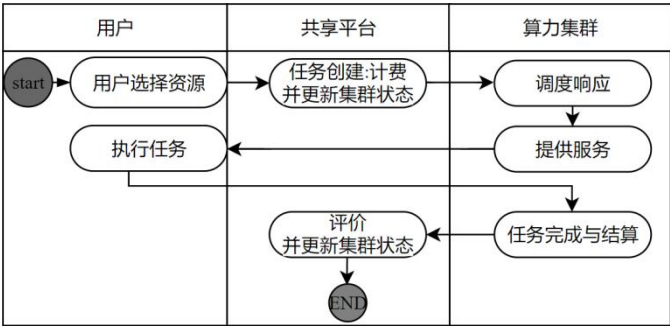


图 7 任务管理流程

用户选择资源：用户根据资源信息：包括算力集群类型（A，B）、资源类型、资源型号、内存容量、虚拟核心数、单价、位置信息等，选择所需资源。

任务创建：平台接收用户资源请求，并创建任务实例并计费（共享平台展示的算力集群均为空闲状态。A类算力集群供给方与用户达成租赁协议后，用户在平台校验授权码后，平台更新当前集群状态为忙碌。B类算力集群供给方通过Irmads接口自动更新集群状态为忙碌）。

算力集群的响应与服务：算力集群迅速响应并提供算力服务。

任务执行：建议B类算力集群提供平台服务，用户在虚拟化平台能执行任务（建议A类算力集群提供SSH账号服务，用户通过分配的SSH账号执行任务）。

评估并更新集群状态：共享平台依托Irmads接口，实时获取并分析算力集群的监控数据日志（使用率，掉卡等关键指标）。平台根据监控数据日志分析结果，向用户和算力集群提供评价与反馈（对于A类算力集群共享平台不进行评价），同时更新集群状态为空闲。

6.5.7 算力集群可视化

可视化要求如下：显示算力集群信息，包括但不限于：算力集群类型（A，B）、算力集群状态（空闲，忙碌）、资源类型、资源型号、内存容量、虚拟核心数、单价、位置信息等。

针对B类算力集群的实时可视化要求如下（A类算力集群不进行监控）：

应支持算力集群使用情况可视化，包括但不限于 CPU 利用率、GPU 利用率、运行温度、电压等；应支持网络状态监控可视化，包括但不限于流量、时延、丢包率等。

6.5.8 算力集群预警与状态推送功能

预警与状态推送功能要求如下：共享平台通过邮件、短信等渠道，实时向用户和算力供给者推送算力集群的异常预警信息及状态，保障双方能即时掌握集群最新动态。

7 算力集群技术要求

7.1 分类

算力集群分为两类：A类算力集群和B类算力集群。

7.2 A类算力集群技术要求

应提供用户使用算力集群服务，建议A类算力集群提供SSH服务或VNC服务。

宜采用最新的服务器硬件配置，如Intel最新一代全线服务器CPU产品、DDR4内存、NVMe SSD硬盘等。

宜安装以下软件工具：配置Python 3.x环境及CUDA、CUDNN加速库（如果配备NVIDIA GPU）。安装模型训练框架，如TensorFlow、PyTorch或MXNet。

7.3 B类算力集群技术要求

应支持对算力集群（如CPU、GPU、内存、存储等）的实时监控，包括但不限于：资源使用率、负载情况、健康状态等，并提供数据传输接口。

应提供用户使用算力集群服务，建议B类算力集群提供平台服务。用户注册登录虚拟化平台后，可选预配置开发环境开发应用，并一键部署至自动管理资源的虚拟环境。

应支持定期备份算力集群数据，算力集群能够灵活地根据其自身的资源状况来优化配置存储系统的副本数量，宜采用云存储或网络附加存储方式，同时，副本数至少为2份。

参 考 文 献

- [1] GB/T 0129—2023 SSH 密码协议规范
- [2] GB/T 9361—2011 计算机场地安全要求
- [3] GB/T 29265.206—2017 信息技术 信息设备资源共享协同服务 第206部分：远程访问服务平
- 台
- [4] GB/T 36630.2—2018 信息安全技术信息技术产品安全可控评价指标 第2部分：中央处理器
- [5] GB/T 44067.2—2024 工业互联网平台技术要求及测试方法 第2部分：工业PaaS平台
- [6] GA/T 1326—2017 安全防范 人脸识别应用 程序接口规范
