**异构集群资源管理平台**

愿景和范围文档



项目名称： 异构集群资源管理平台

项目组别： G05组

小组成员： 钱佳苗 钱丁瑜 胡欣阳 黄子涵 李玉炜

指导教师： 杨枨老师

二〇二四年五月九日

|  |  |  |
| --- | --- | --- |
| 文件状态：  [√] 草稿  [ ] 正式发布  [ ] 正在修改 | 文件标识： | SRA2024-G05-愿景与范围文档 |
| 当前版本： | 0.2.0 |
| 作 者： | 钱佳苗 钱丁瑜 胡欣阳 黄子涵 李玉炜 |
| 完成日期： | 2024.05.09 |

**版本记录**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 修订日期 | 版本/状态 | 修订人 | 备注 |
| 01 | 2024.04.28 | 0.1.0 | 钱佳苗、胡欣阳 | 愿景与范围模板确定与初稿完成 |
| 02 | 2024.05.05 | 0.2.0 | 钱佳苗、胡欣阳 | 讨论完成初稿 |
| 03 | 2024.05.09 | 0.3.0 | 钱佳苗、胡欣阳 | 修改整体排版，完善特性树和部分细节内容。 |

目录

[1. 业务需求 0](#_Toc21854)

[1.1 背景 0](#_Toc6485)

[1.2 业务机遇 0](#_Toc11293)

[1.3 业务目标 0](#_Toc11372)

[1.4 成功指标 1](#_Toc17272)

[1.5 愿景陈述 1](#_Toc12443)

[1.6 业务风险 1](#_Toc9167)

[1.7业务假设与依赖 1](#_Toc26119)

[2. 范围与限制 2](#_Toc10455)

[2.1 主要特性 2](#_Toc27949)

[2.2 初始与后续发布的范围 2](#_Toc15946)

[2.3 限制与排除项 2](#_Toc25419)

[3. 业务上下文 3](#_Toc4908)

[3.1 干系人资料 3](#_Toc12043)

[3.2 项目优先级 3](#_Toc85)

[3.3 部署考虑 3](#_Toc26483)

# 业务需求

## 1.1 背景

我校已建成浙江省内公办高校首家校级超算中心，为我校师生提供支撑科学研究的信息化基础设施，打造支撑学校交叉创新研究和产业创新的基础平台，成为学校的金名片。

为更加高效地使用超算中心的资源，现阶段迫切地需要一个资源管理平台去提高集群管理员的管理效率，使算力使用者可以有更好地算力使用体验。

## 1.2 业务机遇

校级超算中心的建设为浙江省内公办高校带来了前所未有的信息化基础设施支持，成为学校交叉创新研究和产业创新的重要基础平台，彰显了学校在科技领域的领先地位。然而，随着超算中心资源的不断扩展和应用范围的拓展，高效管理和优化资源利用的需求日益凸显。

在当前阶段，面临的业务机遇是迫切需要开发一个异构集群资源管理平台，以提升集群管理员的管理效率，同时优化算力使用者的体验。这一资源管理平台将成为学校科研和教学工作的重要支撑，为用户提供高效、可靠的超算资源服务。

异构集群资源管理平台的功能包括对多个异构计算集群的资源管理、分配和状态监控。它将为管理员提供直观的管理界面，使其能够实时监测集群资源的使用情况、优化资源分配，从而更精确地满足不同用户的需求。对于算力使用者，平台将提供友好的用户界面和便捷的操作体验，让他们能够轻松地申请、使用和监控超算资源，有效提高科研和教学效率。

通过开发异构集群资源管理平台，学校将进一步提升超算中心的服务水平，推动科研和教学成果的转化和应用，为学校打造更具影响力和竞争力的金名片。

## 1.3 业务目标

BO-1：在初始发布之后的1个月内，使异构集群资源管理平台在浙大城市学院内推广开，具有一百名用户。

BO-2：在初始发布之后的3个月内，使异构集群资源管理平台的用户规模扩大至一千人。

BO-3：在初始发布之后的6个月内，使异构集群资源管理平台的用户规模达到五千人，涉及面包括各个学院。

## 1.4 成功指标

SM-1：异构集群资源管理平台成功投入校内超算管理。

SM-2：异构集群资源管理平台成为一大有力工具，成为其他高校乃至组织的优先考虑。

## 1.5 愿景陈述

我们的愿景是通过开发和实现异构集群资源管理平台，为浙大城市学院师生提供一流的计算资源管理和利用体验。我们致力于打造一个高效、智能的平台，让集群管理员能够轻松管理和优化计算资源，同时为算力使用者提供便捷、可靠的资源申请和使用方式。我们的目标是推动学校科研和教学的发展，促进学术交流和创新合作，成为学校信息化建设的重要支撑，助力学校在科技领域的领先地位。通过我们的努力，希望为学校师生创造更加开放、智能化的学习与研究环境，推动学校迈向信息化智能化发展的新高度。

## 1.6 业务风险

RI-1: 异构集群包含不同类型的计算资源和硬件设备，系统需要处理多样化的技术要求和兼容性问题，增加了开发和集成的复杂性。

RI-2：管理敏感数据和用户信息可能面临安全威胁和隐私泄露风险，如数据泄露、身份认证问题等。

RI-3：用户对新平台的接受度和使用意愿不确定，可能面临推广和普及的挑战。

RI-4：异构集群资源管理平台需要处理大量的计算任务和数据请求，可能面临性能瓶颈和系统稳定性问题。

## 1.7业务假设与依赖

AS-1：假设学校管理层和相关部门支持和认可异构集群资源管理平台的开发和推广，提供必要的资源和支持。

AS-2：假设对集群管理员和算力使用者的用户需求调研和分析准确无误，平台功能设计能够有效满足用户期望和需求。

DE-1：依赖学校系统和异构计算集群的数据接入和集成，确保平台能够获取和处理所需的计算资源信息和用户数据。

# 范围与限制

## 2.1 主要特性

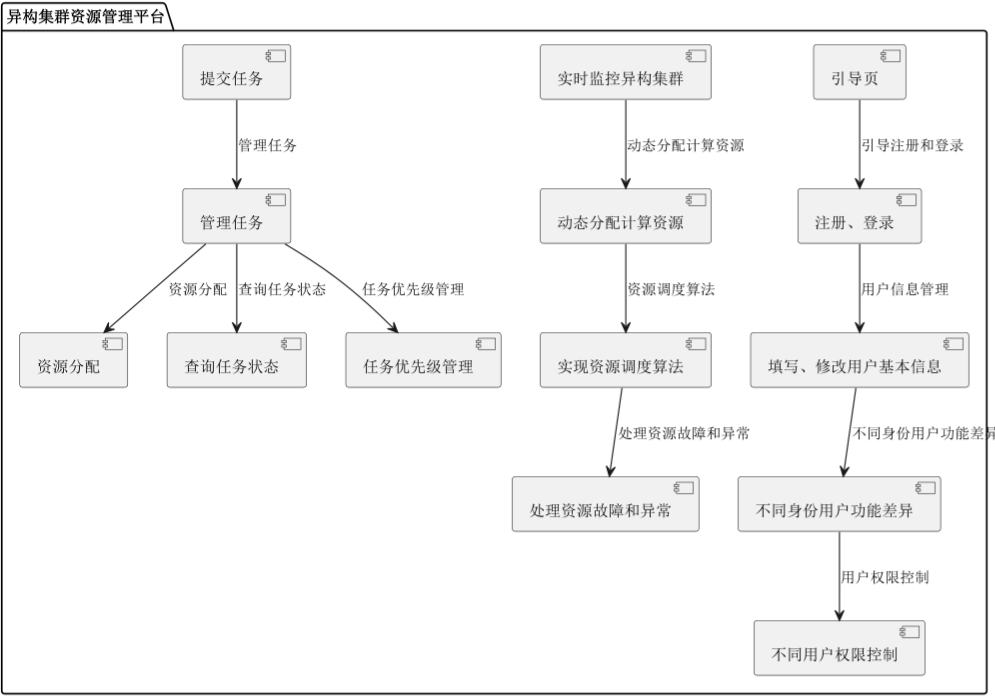
FE-1：实现对异构计算集群资源的管理和调度，包括CPU、GPU、存储等资源的分配和优化，确保资源的高效利用和平衡。

FE-2：提供完善的用户权限管理和身份认证机制，确保只有授权用户能够访问和使用平台的功能和资源。

FE-3：设计直观、用户友好的用户界面，包括管理员后台界面和算力使用者前端界面，简化操作流程，提升用户体验。

FE-4：允许算力使用者提交计算任务，并提供任务监控和状态查询功能，实时反馈任务执行情况和结果。

FE-5：引入智能的资源调度算法，根据不同任务的需求和优先级，动态调整资源分配，优化资源利用效率。



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **类别** | **模块名称** | **主要功能** | **功能描述** | **备注** |
| 平台 | 用户管理模块 | 用户注册和登录 | 引导页 | 小程序加载，界面同时显示欢迎界面。 |
| 注册、登录 | 注册页面、登录页面、忘记密码页面。 |
| 用户信息管理 | 填写、修改用户基本个人信息 | 昵称、姓名、性别、手机号码、专业、年级、身份（学生、教师、管理员） |
| 身份验证和授权 | 不同身份用户功能不同 | 根据用户角色和权限管理策略，授权用户访问不同的系统功能和资源。 |
| 用户角色和权限管理 | 不同用户权限不同 | 管理用户的权限，限制不同角色用户能够执行的操作和访问的资源。 |
| 资源管理模块 | 资源监控和统计 | 实时监控异构集群 | 实时监控异构集群中各种资源的使用情况和性能指标，包括CPU利用率、内存占用、存储容量等。  统计资源的历史使用情况，生成报表和图表，帮助管理员了解资源的利用率和趋势。 |
| 资源分配和调度 | 动态分配计算资源 | 根据系统负载和用户需求，动态分配计算资源，确保资源的合理分配和高效利用。 |
| 实现资源调度算法 | 实现资源调度算法，根据任务优先级和资源需求，优化资源的分配策略，提高系统的性能和效率。 |
| 容错和负载均衡 | 处理资源故障和异常情况 | 实施容错机制，处理资源故障和异常情况，保障系统的稳定性和可靠性。  实现负载均衡策略，分配任务和资源，避免单点故障和资源瓶颈。 |
| 任务调度模块 | 任务提交和管理 | 提交 | 允许用户提交各种类型的计算任务，包括数据分析、科学计算、机器学习等任务。  支持任务的参数设置和描述，包括任务类型、资源需求、优先级等。 |
| 任务队列管理 | 管理任务 | 管理任务队列，按照优先级和调度策略对任务进行排队和调度。  支持任务的暂停、恢复、取消等操作，确保任务执行的顺序和有效性 |
| 资源调度和分配 | 分配 | 根据任务的资源需求和优先级，动态调度计算资源，确保任务能够及时得到执行。  实现资源分配算法，优化资源的利用效率，提高任务执行的性能和响应速度。 |
| 任务监控和状态查询 | 查询 | 提供任务执行状态的实时监控和查询功能，让用户了解任务的执行进度和结果。  显示任务的状态信息，包括排队中、运行中、完成等状态，以及任务的执行日志和报告。 |
| 任务优先级和调度策略 | 优先级 | 支持任务优先级的设置，根据任务的重要性和紧急程度进行调度。  实现灵活的调度策略，根据系统负载和资源可用性调整任务的执行顺序和分配策略。 |

## 2.2 初始与后续发布的范围

|  |  |  |  |
| --- | --- | --- | --- |
| 特性 | 发布1 | 发布2 | 发布3 |
| FE-1,基本用户功能 | 允许用户创建账户并登录系统，提供用户个人资料的基本管理功能。 | 基本的身份验证和权限管理，确保只有授权用户可以访问系统功能 | 引入更多个性化设置和权限控制，支持不同用户角色的功能定制 |
| FE-2,资源管理基础功能 | 实时监控异构集群：展示集群中关键资源的使用情况和状态。 | 动态分配计算资源：基本的资源分配功能，确保资源的合理利用和分配。 | 优化资源分配算法，提升资源利用效率和系统性能。 |
| FE-3,任务调度基础功能 | 允许用户提交计算任务到系统中。 | 管理任务：基本的任务管理功能，包括任务列表和状态查询。 | 增加更多的任务管理功能，如任务优先级设置和任务暂停恢复等。 |
| FE-4,系统稳定性和可靠性 | 基本的错误处理和容错机制，确保系统在面对异常情况时能够保持稳定运行 | 基本的系统性能优化，确保系统能够满足基本的运行需求。 | 加强系统的安全性，包括数据加密、访问控制等措施，保护用户数据安全。 |

## 2.3 限制与排除项

限制项

1资源限制：

确定可用资源的限制，包括人力、财力、设备和技术等方面的资源限制。

2技术限制：

确定可用的技术和工具的限制，例如特定的编程语言、开发框架或硬件设备。

3范围限制：

确定项目的功能和范围限制，明确哪些功能不在当前项目范围内。

**排除项**

1功能排除：

明确列出不在项目范围内的功能和特性，避免项目范围膨胀和需求不清晰。

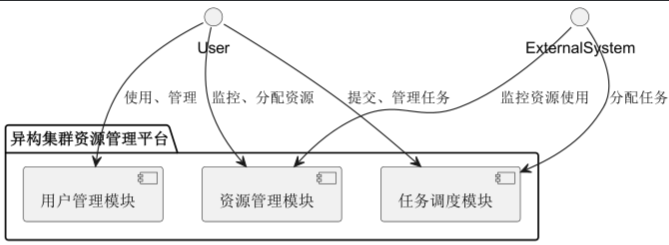
2技术排除：

排除不适合或不可用的技术和工具，避免在项目中使用未经验证或不熟悉的技术。

3资源排除：

排除不可用或不可获得的资源，避免因资源不足而影响项目进展和质量。

# 业务上下文



## 3.1 干系人资料

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 干系人 | 主要价值 | 态度 | 主要兴趣 | 约束 |
| 项目发起人 | 通过项目的经营活动，获得实际利益。 | 已确认项目 | 按照项目发起时规定的要求执行项目。 | 按照课程里程碑完成各个阶段的事务与内容。 |
| 项目经理 | 对项目全权负责，负责项目按照预期执行。 | 对项目预期乐观，准备后续执行。 | 全权负责项目的整体推进与各阶段内容的审核。 | 依照评审条目对各阶段成果进行核验。 |
| 需求记录者 | 每次需求获取的记录者。 | 对项目预期乐观，准备后续执行。 | 记录每次需求沟通时目标用户的需求与建议，形成文档供小组成员沟通。 | 依照评审条目对各阶段成果进行核验。 |
| 界面原型设计者 | 有一定美术美工能力的界面原型制作者。 | 对项目预期乐观，准备后续执行。 | 不断完善项目原型，使之能够被项目发起人和项目经理满意。 | 根依照评审条目对各阶段成果进行核验。 |

## 3.2 项目优先级

由于项目的课程要求，荐课小程序首要的任务是完成课程推荐的主要功能，在此之上建立用户体系，并根据用户体系实现社区功能。

|  |  |  |  |
| --- | --- | --- | --- |
| 维度 | 约束 | 驱动 | 自由度 |
| 特性 | 所有排入1.0的特性必须完全可操作 |  |  |
| 质量 | 用户验收测试通过率必须超过95%；安全测试必须全部通过 |  |  |
| 排期 |  |  | TBD |
| 成本 |  |  | 在无赞助方评审的情况下，可以接受不超过15%的预算超支。 |
| 人员 |  | 团队包括一名项目经理和四名项目开发人员 |  |

## 3.3 部署考虑

1硬件需求：

确定部署所需的服务器和计算资源，包括处理器、内存、存储等。考虑到异构集群的特性，可能涉及到不同类型的处理器和加速器（如CPU、GPU、FPGA等），因此需要根据系统的需求来配置硬件资源。

2软件环境：

确保部署环境具备所需的软件依赖和运行环境。这可能涉及操作系统的选择（如Linux、Windows等）、数据库管理系统、应用服务器和其他运行时环境。

3网络架构：

设计和配置网络架构，确保系统各个组件之间的通信畅通。考虑到异构集群可能涉及到分布式计算和跨网络资源调度，网络带宽和延迟也是需要考虑的因素。

4安全性：

实施安全措施，保护系统免受网络攻击和数据泄露。包括访问控制、身份验证、数据加密、防火墙和安全审计等措施，确保系统数据和用户信息的安全性。

5高可用性和容错性：

部署高可用性和容错性架构，确保系统在面对硬件故障或网络中断时能够保持稳定运行。使用负载均衡和故障转移技术，实现系统的高可用性和可靠性。

6监控和管理：

部署监控和管理工具，实时监测系统的运行状态和资源利用情况。采用日志记录和报警机制，及时发现和处理系统问题。