

联智科技

高性能计算管理平台

CHESS

教育版技术白皮书

版本 V7.0

2023-4-4

尊敬的用户：

版权©联智科技（北京）有限公司所有

在未经联智科技(北京)有限公司事先书面许可的情况下，严禁以任何形式复制，传递，分发和存储本文档中的任何内容。

联智科技（北京）有限公司遵循可持续发展的原则，因此，联智科技（北京）有限公司保留在未预先通知的情况下，对本文档中描述的任何内容进行修改和改进的权利。

技术服务电话：+86 010 82608832

网 址：<http://www.lianzhitech.cn>

地 址：北京市海淀区知春路 1 号
学院国际大厦 1801 室

邮 编：100191

前言

本文档主要介绍了 CHESS 的设计理念、架构和功能特性。

读者对象

本文档主要适用于以下读者对象：

- 技术支持工程师
- 售前/售后工程师

术语和缩略语

术语	说明性定义
CHESS	联智科技高性能计算管理平台
SLURM	Simple Linux Utility for Resource Management

目录

前言	2
读者对象	2
术语和缩略语	2
第 1 章 产品简介	5
1.1. 产品概述	5
1.2. 关键技术特性	5
第 2 章 设计理念	7
第 3 章 系统架构	8
第 4 章 适配硬件	11
第 5 章 功能特性	13
5.1. 集群部署模块	13
5.2. web portal	13
5.3. 安全控制模块	14
5.3.1. 三员管理	14
5.3.2. 项目管理	16
5.4. 用户管理模块	16
5.5. 集群调度模块	18
5.6. 应用中心	20
5.7. 远程 3D 可视化模块	21
5.8. 数据管理模块	22
5.9. 集群资源管理模块	22
5.8.1. 节点列表	22
5.8.2. GPU 列表	23

5.8.3. 资源限制	23
5.8.4. 队列管理	24
5.8.5. 配额管理	25
5.10. 人工智能模块	25
5.11. 监控运维模块	26
5.12. 计费模块	28
5.13. 集群报表模块	30
5.14. 系统设置模块	32
第 6 章 应用场景	35
6.1. CAE 仿真	35
6.2. 人工智能	36
6.3. 基础学科	36
6.4. 生命科学	37
6.5. 石油勘探	38
6.6. 气象行业	40

第1章 产品简介

1.1. 产品概述

CHESS (ClusterTech HPC Environment Software Stack) 是联智科技自主研发的高性能集群管理软件, 采用 B/S 架构, 通过浏览器进行操作实现集群资源统一部署、管理、调度、监控、报警、报表和计费, 可将松散堆叠的服务器变成一整套 HPC 集群系统, 大幅提高集群效率, 简化集群管理。

在操作界面上采用了桌面式效果, 把所有的管理操作均使用桌面式效果展示, 并根据多年积累的经验和收到的用户反馈, 重新进行架构设计, 在易用性、稳定性、可靠性、可扩展性、可维护性、安全性等各方面也都提出了更高的要求, 提升了用户体验, 满足了各行业用户的需求。

功能模块主要包括集群部署、安全控制、数据管理、用户管理、资源管理、集群监控、作业调度管理、作业提交、计费系统、许可证管理、集群报表、知识库、多集群, 系统管理员可根据用户使用情况为每个用户分配功能模块, 灵活使用, 严格控制用户权限。

此外, 联智科技在高性能计算集群系统软件领域有十余年的经验, 可为用户完善其并行环境, 帮助用户安装函数库, 编译器, 消息传递接口以及用户的应用软件, 完成实施后还将为用户进行基准测试, 评估用户集群效率, 进行系统调优, 为用户搭建一套完善、高效的高性能计算集群系统。

1.2. 关键技术特性

CHESS 的关键技术特性有:

✧ 私有集群管理: 快速部署, 集中管理, 统一调度

- ✧ 动态调度、按需申请、自动释放，提高资源利用率，降低计算成本
- ✧ 公私深度融合：与公有云、私有云深度融合，轻松构建公私及混合云集群
- ✧ 独立的模块化设计，可根据用户需求自由组合
- ✧ 支持 HA，避免单点故障带来的时间和经济上不可逆转的损失
- ✧ GPU 监控与调度：支持 GPU 单卡&多卡的共享、独占
- ✧ 一站式管理服务：为用户优化、定制应用软件和集群管理软件
- ✧ 最高支持上万节点的规模，最大可承载 100 万个作业的管理和调度
- ✧ 支持多种架构：支持 X86、ARM 和 AIX 架构服务器
- ✧ 支持多种 AI 开源框架：支持 Tensorflow、Caffe 等框架
- ✧ 提供独立的计费系统、强大的集群监控和完善的报表系统
- ✧ 先进性：采用了先进的 B/S 结构，运用了数据传输技术，数据存储、处理、分析技术，界面设计本着用户的工作特点、行为心理学的原理、人性化的理念
- ✧ 灵活性：平台拓扑结构层次灵活，视用户的需要和管理方便，可采用二层或多层结构，随时进行管理和设置
- ✧ 可视化、智能化、远程化：通过技术手段把远程的应用分析计算的过程以 3D 可视化的方式智能的呈现出来，便于用户理解和监控。

总结起来，CHESS 具有如下特点：

- 功能全
- 易使用
- 高可靠
- 可扩展

第2章 设计理念

在近代科学研究中，单靠理论和实验解决问题的难度逐渐增大，数值运算的方法被用来模拟物理世界，以求解复杂的问题，计算科学成为自然科学研究的必备工具。随着求解问题规模的越来越大，对计算能力的需求成为驱动高性能计算发展最直接的动力。

早期的高性能计算机主要应用于解决军事领域的计算问题，如美国在 1960 年代使用 CDC 超级计算机进行弹道计算、火箭设计等工作。20 世纪 90 年代中期以后，随着机群技术构建的高性能计算机的普及，高性能计算的成本和编程的复杂度大幅度下降，为高性能计算的广泛使用创造了条件。现在高性能计算已经渗透到各个学科领域，不仅在气候模拟、石油勘探、天体物理这些传统应用领域保持强劲的生命力，在生命科学、人工智能、大数据处理这些新兴领域也有广泛的应用。

高性能计算系统利用大量处理单元的聚合计算能力来满足应用巨大的计算需求，其关键问题是实现众多计算节点的大规模集成和高效协同计算，核心技术涉及高性能计算机和大规模并行应用程序。随着高性能计算系统的不断升级和硬件需求的不断扩充，衍生了高性能计算集群。

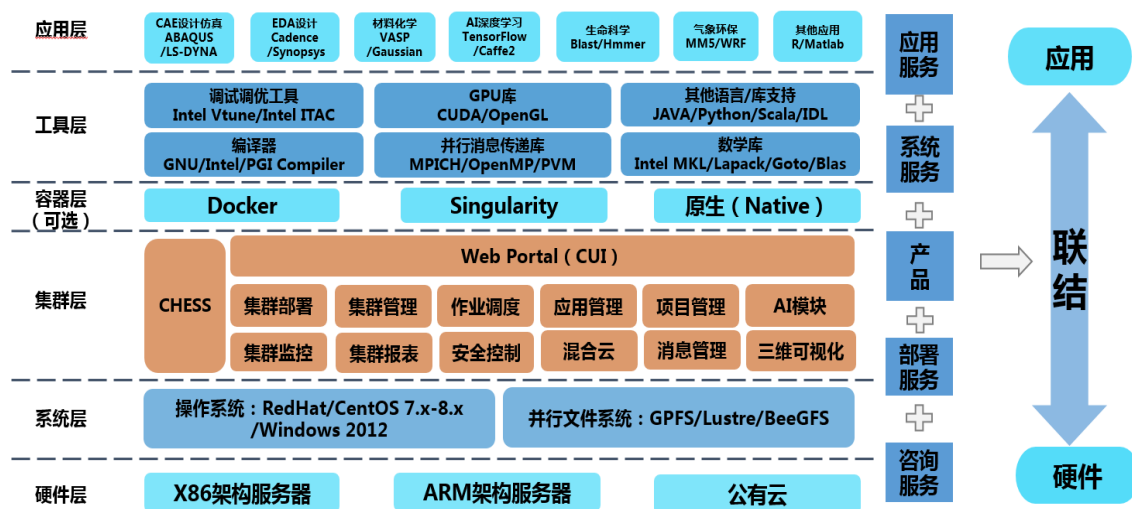
CHESS 高性能计算管理平台的设计围绕着高性能计算的特点、以人为核心，可持续发展为基本出发点，用主动、积极的方式，一种系统的、整体的观念，把需求、人机工程学、美学及环境因素等有机的融合，进行综合分析，以此确立产品设计的理念和目标。

CHESS 产品的设计理念和目标概述如下：

- ✧ 产品架构：采用 B/S 架构，通过浏览器实现了集群资源的统一部署、管理、调度等功能。
- ✧ 数据存取：针对数据安全存取及高可用行进行设计和实施。
- ✧ 安全配置：支持匹配等保规范等安全审计及三员设计，支持用户及密码强度匹配相关的策略设计。
- ✧ 计算调度：支持并行、异构、多核及串行计算任务，及 AI 实训的开发代码调试与机器学习任务。
- ✧ 用户管理：界面化的用户增删改管理设计，和灵活的权限分配。
- ✧ 计时计费：基于用户、部门或任务等多维度的统计和计费。

第3章 系统架构

CHESS 的系统架构如图所示：



各层介绍如下：

硬件层：

- ◆ CHESS 支持 X86 架构服务器、ARM 架构服务器，混合云可与公有云对接，够成 HPC 底层计算资源。
- ◆ 联智科技的咨询服务可以根据用户的基本需求，所运行的应用软件名称、版本、数据规模 and 用户数等情况提供 HPC 集群的软、硬件设计方案。

系统层：

- ◆ CHESS 支持 RedHat 7. x 和 CentOS 7. x 版本，并行文件系统支持 GPFS、BeeGFS 和 Lustre。
- ◆ 联智科技的部署服务可以将 HPC 集群中的管理节点、登录节点、IO 节点和计算节点等的操作系统进行快速部署，并支持系统的备份和恢复；根据用户需求配置集群中的并行文件系统。
- ◆ 联智科技提供的四项基本测试服务可以在初次安装或者年度巡检时，

诊断 CPU、内存、网络 and IO 性能是否正常，HPC 整机的性能是否达到最优。

集群层：

- ◆ 通过 CHESS 提供集群部署、集群管理、集群监控、作业调度、应用管理、用户管理、集群报表、消息管理、项目管理、计费、混合云等功能，同时支持 HPC 和 AI 的融合应用。

容器层（可选）：

- ◆ 有些应用程序是以容器的形式发布的，CHESS 支持 Docker 和 Singularity 容器。
- ◆ 联智科技的系统服务可将用户的容器做成应用模板，将用户最常用的参数设置为交互参数，方便用户提交容器作业。

工具层：

- ◆ CHESS 的安装过程可选的 HPC 工具集包含并行命令、调试调优工具、消息传递库、数学库和编译器，具体内容如下表所示：

功能	工具
并行命令	dvt, dsh, pcp, distcc
调试、调优工具	Intel Vtune , Intel Trace Collector/Analyser, Total View
消息传递库	MPICH, MPICH2, MVAPICH, OpenMPI 等
数学库	Intel MKL, Lapack, Scalapack, Goto, Blas...
编译器	GNU 编译器, Intel 编译器

注：Intel 编译器、Intel MKL、Intel Vtune、Intel Trace Collector/Analyser 和 Total View 为商业软件。

- ◆ 联智科技的系统服务可以安装和配置好所有工具。

应用层：

- ◆ CHESS 支持 CAE 仿真设计、EDA 仿真设计、气象海洋、石油勘探、生命科学、计算化学、人工智能等行业的各种应用。
- ◆ 联智科技的应用服务包括应用的移植、调优、并行化，将应用软件在推荐的硬件平台上性能达到最优。

第4章 适配硬件

CHESS 平台部署在合适的硬件环境中，可大大提高应用的并行计算能力。目前 CHESS 能够匹配市场上主流的通用硬件平台，包括 x86 架构的服务器平台及国产信创计算平台产品，均能稳定高效的运行，CHESS 平台为独立软件平台产品，不刻意绑定硬件平台，拥有最广泛的、最专业的跨品牌、跨平台硬件支持能力，能够最大限度保护用户投资，实现多品牌、跨架构设备共存共管。

高性能计算集群基础硬件架构主要由服务器设备、带内管理网络、带外管理网络、高速计算/存储网络、存储平台等几部分构成，以下做依次说明：

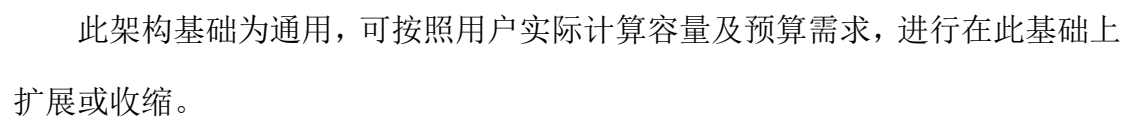
服务器：主流 Intel、AMD 的 x86 架构服务器品牌，包括且不仅限于 DELL、联想、浪潮、曙光等国内外知名品牌或其他服务器品牌产品，同时支持基于华为鲲鹏及国内的信创产业平台下的飞腾、龙芯 CPU 平台，支持基于 Power9 平台的 AIX 操作系统；

分布式存储：CHESS 能够提供基于 NFS、Lustre、GPFS、BeeGFS 的高性能计算集群分布式文件系统的架构设计、部署、调优、维护与支持升级服务支持工作，提供一篮子集群文件系统解决方案，保障用户数据安全，提升性能；

高速互联网络：支持基于 1Gb / 10Gb / 25Gb 以太网及 RoCE 计算存储及管理网络，支持基于 Mellanox 的 56Gb / 100Gb / 200Gb 高速 RDMA 计算互联网络及 Intel 主导的 100Gb OPA 高速网络；

带外/带内管理网络：支持基于 1Gb / 10Gb / 25Gb 的 TCP 协议下的以太网络的计算存储及管理网络；支持基于 Mellanox 的 56Gb / 100Gb / 200Gb 高速 RDMA 计算互联网络及 Intel 主导的 100Gb OPA 高速 TCP 协议网络；

典型的通用高性能计算硬件架构如下：

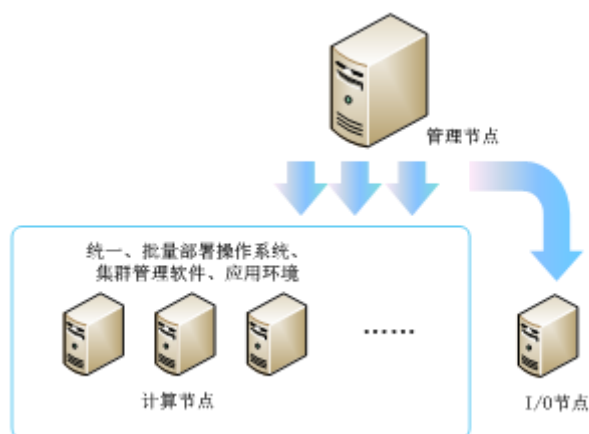


第5章 功能特性

5.1. 集群部署模块

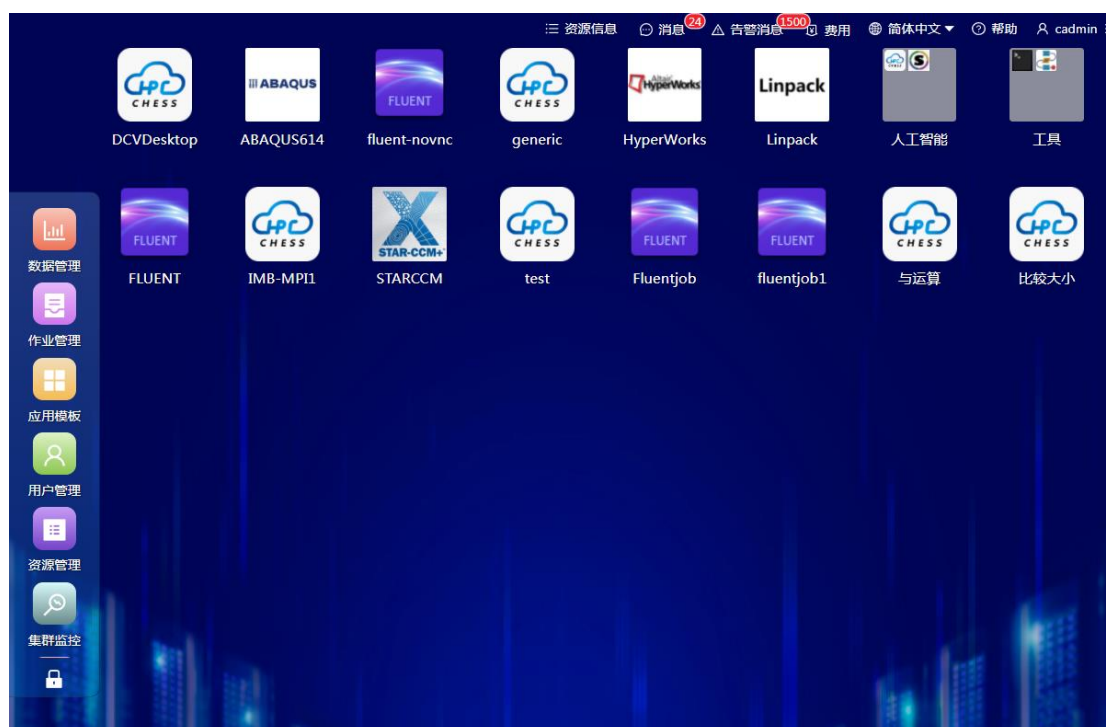
CHESS 集群部署模块帮助系统管理员快速部署数百台节点的集群系统，完成集群节点的操作系统和软件的部署。集群部署模块的优势有：

- 批量安装、快速部署
- 节点弹性扩展、动态伸缩
- 批量系统的开、关机管理功能
- 针对不同节点分发相应系统镜像、定制软件包
- 操作系统、管理软件、应用环境统一部署



5.2. web portal

CHESS web portal 采用桌面式方式实现，每个应用以图标形式展现，逻辑简单，操作方便。用户可通过系统默认或自定义应用模板，定制每一个应用的交互参数，满足不同应用或不同场景的各种使用习惯。



用户登录后，在首页左侧展示导航菜单栏，中间展示发布的应用，右上角的菜单栏包含了资源信息、消息、告警消息、用户费用、语言切换等功能按钮。

5.3. 安全控制模块

CHESS 的安全控制模块由三员管理、项目管理和密级管理组成。通过权限、项目和秘密级别三种层面进行安全把控。

5.3.1. 三员管理

三员管理实现了三种管理员权限分离，“三员”为内置的角色，其权限也是内置的，提供三种管理员不同的页面视图，提供用户安全审核功能，能够批量进行操作，包括冻结、解冻等操作，还提供用户和管理员操作的全面日志记录和审计。

- 负责注册和创建用户账号、注册角色名称、维护组织结构的系统管理员；
- 负责为用户和角色赋权并使账号生效，审计用户和其他管理员的行为的安全

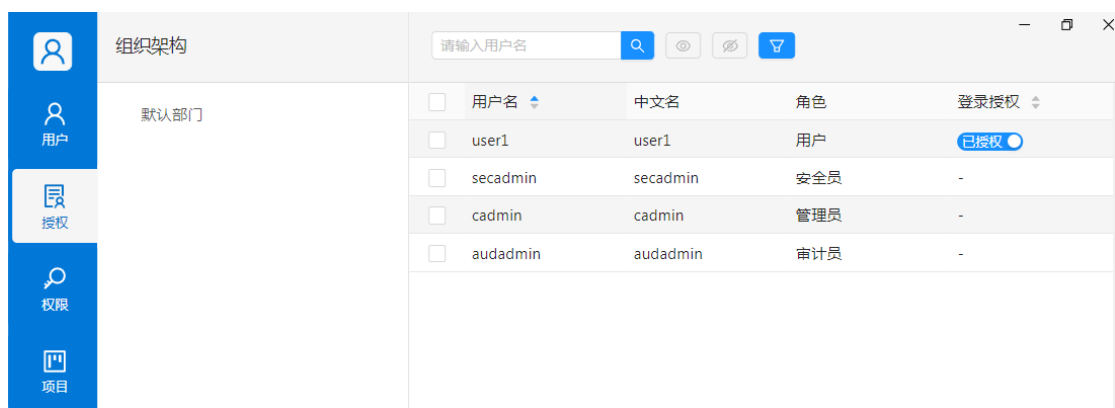
管理员；

- 负责审计系统管理员及安全管理员的行为的审计管理员。如下图所示：



用户名	中文名	角色	登录授权
secadmin	secadmin	安全员	● 已授权
cadmin	cadmin	管理员	● 已授权
audadmin	audadmin	审计员	● 已授权

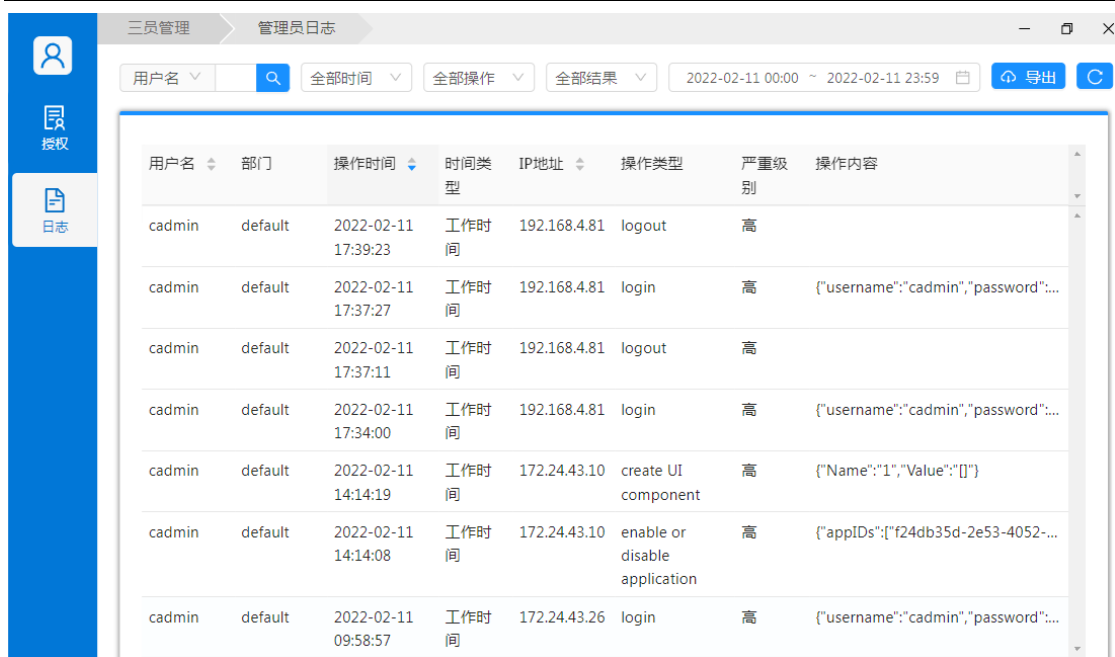
三员管理中的用户是操作系统或 OPENLDAP 或 AD 域中创建的用户。通过自动同步和手动同步两种方式同步到集群中，再对不同角色的用户进行登录授权和功能权限设定，只有授权了的用户才能登录平台，如图所示：



☐	用户名	中文名	角色	登录授权
☐	user1	user1	用户	已授权
☐	secadmin	secadmin	安全员	-
☐	cadmin	cadmin	管理员	-
☐	audadmin	audadmin	审计员	-

授权后，也只有赋予了权限的模块才能访问和操作，确保访问的安全性。

在三员管理中，日志审计也是必不可少的操作。审计日志内容包含操作者、操作日期、操作地点（IP）、操作类型、操作内容（诸如修改的用户名称，作业信息和文件信息等），为安全提供有据可循。如图所示：



管理员日志

用户名: 部门: 操作时间: 全部操作: 全部结果: 2022-02-11 00:00 ~ 2022-02-11 23:59 导出

用户名	部门	操作时间	时间类型	IP地址	操作类型	严重级别	操作内容
cadmin	default	2022-02-11 17:39:23	工作时间	192.168.4.81	logout	高	
cadmin	default	2022-02-11 17:37:27	工作时间	192.168.4.81	login	高	{"username":"cadmin","password":..."}
cadmin	default	2022-02-11 17:37:11	工作时间	192.168.4.81	logout	高	
cadmin	default	2022-02-11 17:34:00	工作时间	192.168.4.81	login	高	{"username":"cadmin","password":..."}
cadmin	default	2022-02-11 14:14:19	工作时间	172.24.43.10	create UI component	高	{"Name":"1","Value":[]}"}
cadmin	default	2022-02-11 14:14:08	工作时间	172.24.43.10	enable or disable application	高	{"appIDs":["f24db35d-2e53-4052-..."}
cadmin	default	2022-02-11 09:58:57	工作时间	172.24.43.26	login	高	{"username":"cadmin","password":..."}

5.3.2. 项目管理

项目管理模块主要用于添加、修改项目和删除项目，给项目分配涉及到的用户，修改项目结束时间，查看项目已有用户等操作。如图所示：



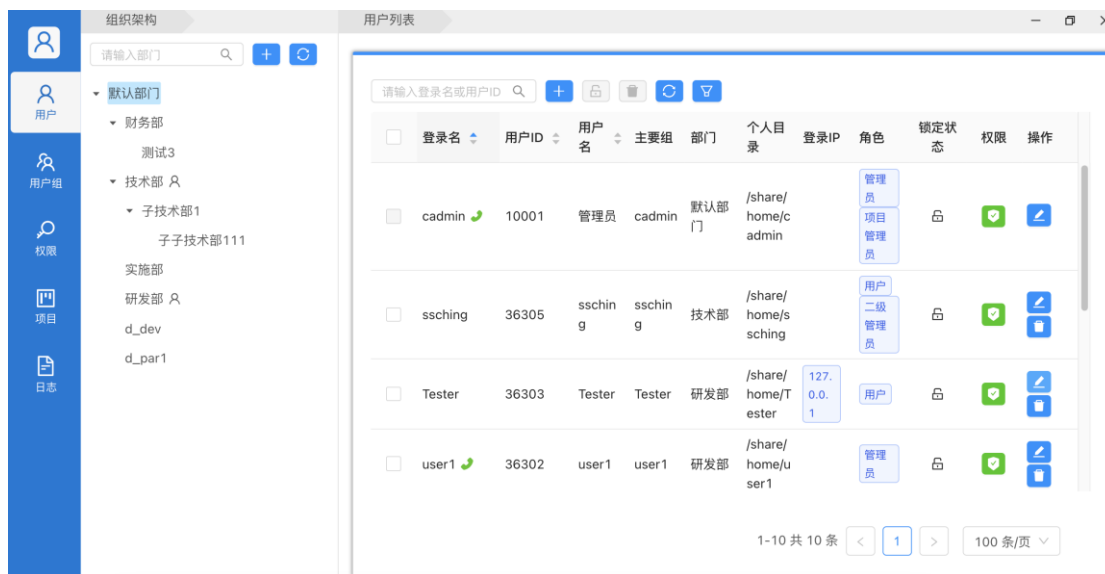
请输入英文名称

英文名称	中文名称	开始时间	结束时间	项目管理员	项目状态	操作
project1	项目1	2022-02-11 17:40:52	2023-03-31 17:40:52	-	正常	用户 设置
default	默认项目	2021-11-12 16:21:06	2038-01-01 00:00:00	-	正常	用户 设置

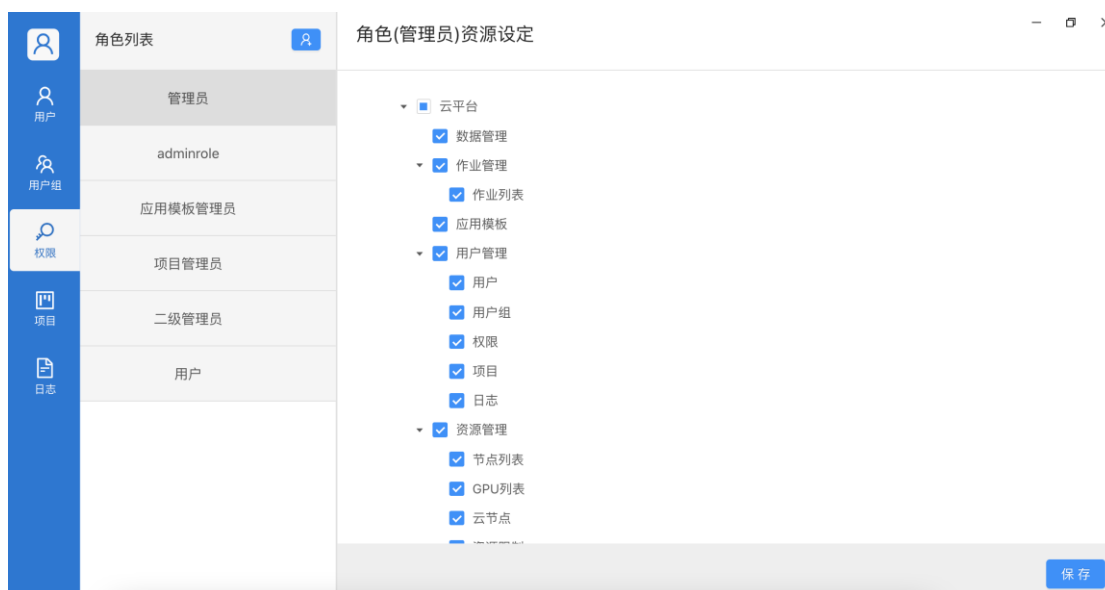
5.4. 用户管理模块

用户管理模块支持用户和用户组的增、删、改、查操作，角色和权限设置，操作日志查看和导出。用户列表中，包含登录名、用户 ID、用户名、主要组、

部门、个人目录、登录 IP、角色、锁定状态、权限和操作按钮，如图所示：



在权限中，内置了管理员、二级管理员、项目管理员和用户角色的默认权限。不同的角色默认拥有的权限不同。管理员也可以根据需要，增减相应模块的权限，如图所示：



日志页记录了用户在平台的所有操作日志，包含登录、注销、修改、计费等操作日志，可以根据筛选条件过滤日志，导出日志。

5.5. 集群调度模块

应用调度模块实现了应用调度和应用管理功能。

应用调度模块支持四种调度器：CHESS CAS 、LSF、SLURM、PBS Professional 和 SGE。

CHESS CAS 是联智科技自主研发的企业级应用程序工作负载调度器，支持各种 HPC 和分析应用类程序，可将多个作业同时调度到集群中运算，提高计算效率，缩短运算时长。

IBM Platform LSF 是一款性能强大的工作负载管理平台，适用于要求苛刻的、分布式的，以及任务关键型高性能计算环境。它提供一整套全面的智能化，策略驱动的计划功能，可以跨异构分布式计算环境对工作负载进行管理，并可以加快工作负载处理速度。可以帮助您利用所有计算基础架构资源，并确保应用程序发挥最佳性能。借助高度可伸缩且可用的体系结构，您可以调度复杂的工作负载，管理高达千万亿次规模的资源。

SLURM （Simple Linux Utility for Resource Management）是一种可用于大型计算节点集群的高度可伸缩和容错的集群管理器和作业调度系统。

PBS Professional 是一个用于复杂和高性能计算环境（比如网格）的工作负载管理器和调度器。

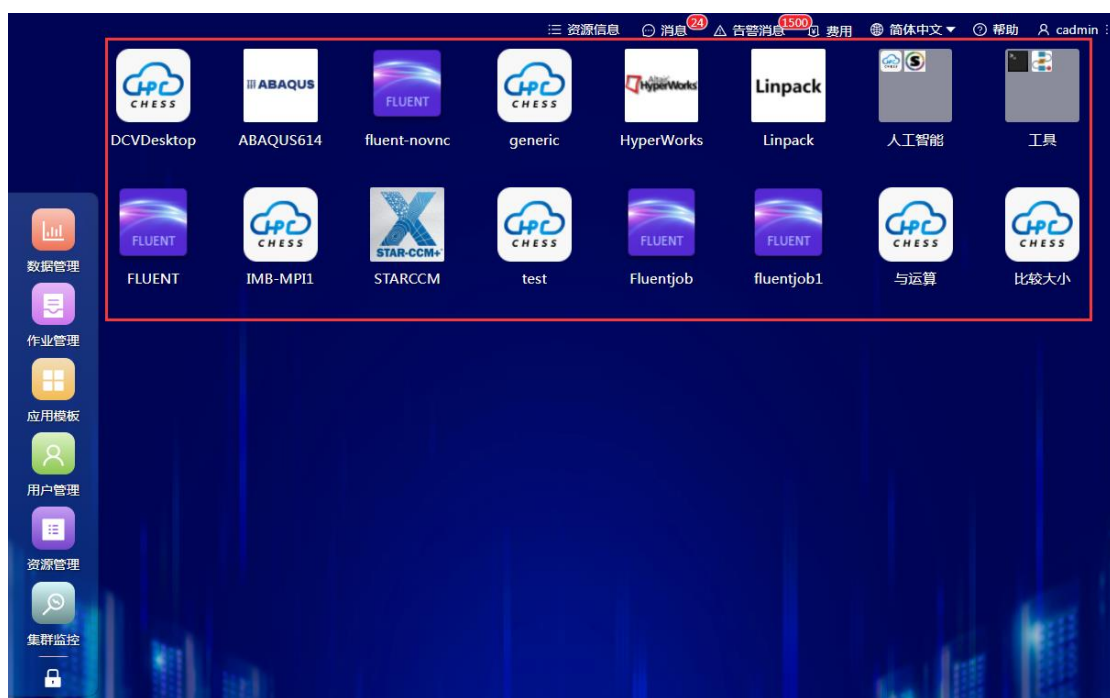
Sun 网格引擎（Sun Grid ，SGE）是一种来自于 SUN Microsystem 的分布式资源管理和调度系统，它用来在基于 UNIX 的计算环境中优化软件和硬件资源的使用。SGE 能用于查找资源池内的闲置资源并利用这些资源；它同样用于通常的一些事务中，例如管理和调度作业到可用资源中。

应用调度模块支持以下功能：

- ✧ 管理自定义的静态和动态资源，可以用于调度系统的资源匹配，如操作系统类型、应用软件许可证等；

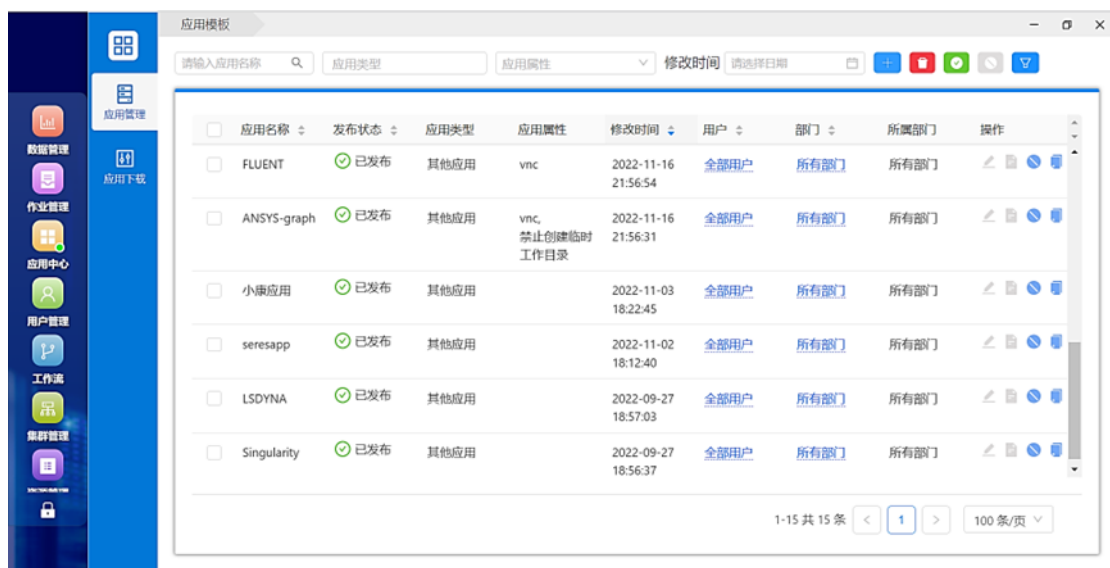
- ✧ 支持多种调度算法，包括先来先服务、Fairshare 公平调度、CPU/内存/应用软件许可证等多种资源的高级预约功能、抢占、独占等；
- ✧ 提供节点和队列的管理功能，提供开启和关闭功能，提供作业的控制、杀死、停止、继续、改变作业排队顺序等；
- ✧ 支持进行进程的跟踪、资源的收集并从而控制 CPU、内存的使用限制；
- ✧ 支持排队作业的余额检测，当余额少于所设限制时自动终止作业运行；
- ✧ 支持 GPU 调度，支持单节点单任务，单节点多任务，多节点多任务等调度方式；支持 MIG 调度。
- ✧ 支持作业组任务的提交和展示。
- ✧ 支持定时任务的设置和自动提交；
- ✧ 支持作业记录的查看作业详情、重新提交等操作。
- ✧ 支持应用后处理计算结果的图形展示。
- ✧ 支持 VNC 图形界面查看。

应用调度模块还集成了多种应用模板，大大提升了应用交付的效率，避免了系统管理员为不同的用户对同一种应用进行反复配置的简单重复性劳动。用户可根据需要打开应用模板，提交作业。集成的应用模板列表如图所示：



5.6. 应用中心

应用调度模块的应用管理提供了应用的自定义配置、删除、发布、禁用、编辑用户&部门、搜索及编辑等应用模板管理操作，还提供了支持用户根据需要自行下载或移除应用的应用下载功能。如图所示：



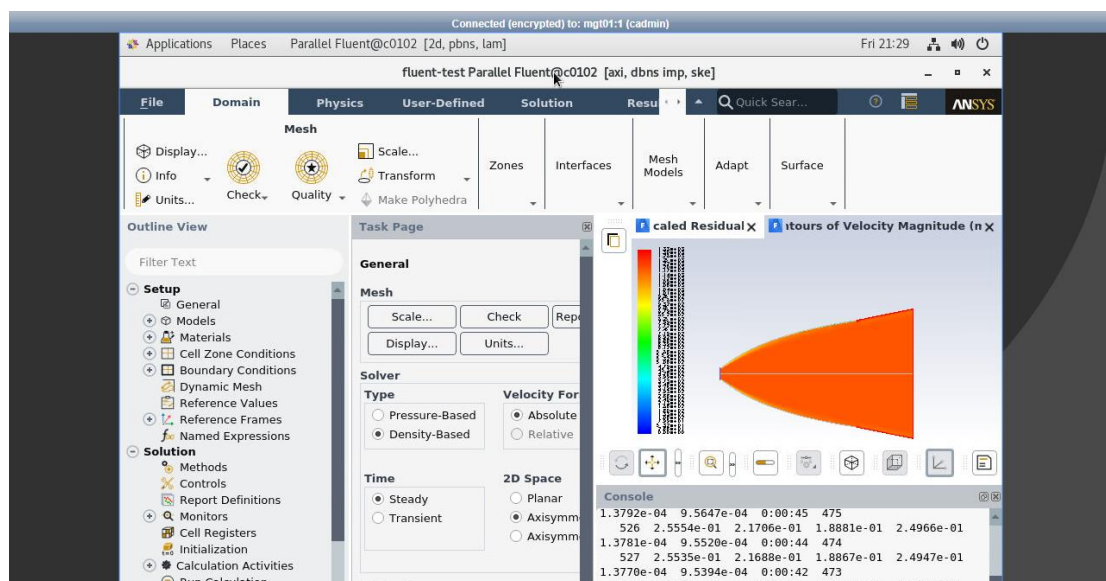
通过拖拽移动组件即可完成模板组件选择，自定义应用模板便于用户快捷的创建个性化应用模板。自定义应用模板创建完成后，发布即可使用该模板提交作业。发布操作如图所示：



5.7. 远程 3D 可视化模块

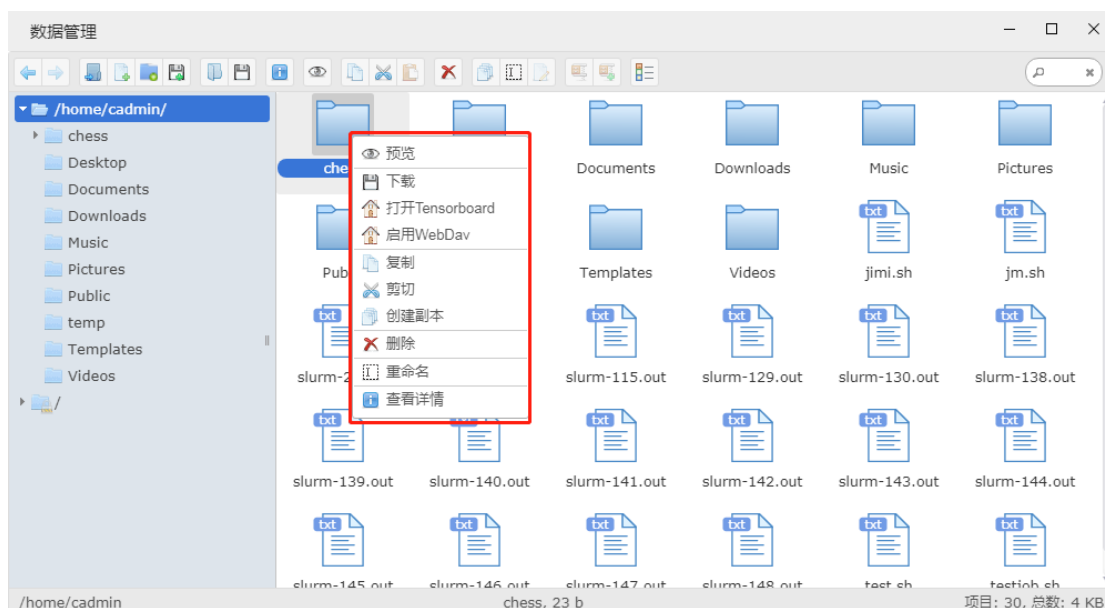
CHESS 采用 VNC 或 DCV 技术实现应用的远程 3D 可视化，将复杂抽象的应用计算过程和数据信息，以合适的视觉元素及视角去呈现，方便大家理解、监控和记忆，且操作简单、易理解。

用 VNC 或 DCV 实现的可视化操作是预先安装上 VNC 或 DCV，然后打开一个应用模板，输入相应的运算参数后，提交即可在互动窗口看到该应用的计算过程，如图所示：



5.8. 数据管理模块

数据管理模块提供用户数据文件的管理功能。用户可通过左侧的文件目录树找到路径下的所有文件和目录，根据需要打开或者浏览目标文件。也可使用文件管理功能栏对文件或文件夹进行各种操作，如“预览”、“下载”、“复制”、“剪切”、“创建副本”、“删除”和“查看详情”等。如图所示：



为了用户便捷的下载常用的软件，平台还支持管理员定制常用软件离线下载页，把软件安装包上传到服务器指定的目录中，修改指定的脚本内容，即可完成常用软件离线下载页的定制。用户在下载页点【下载】，选择本地路径，就能把相应软件包下载到指定文件夹中。

5.9. 集群资源管理模块

集群资源管理模块包含了云节点、CPU 和 GPU 节点列表、资源限制、队列配置、配额管理的功能。

5.8.1. 节点列表

在节点列表中展示了节点状态、名称、服务状态、资源使用情况和作业分布，

并且可以在页面上对节点进行关机、重启、VNC、硬关机、硬开机、控制台的操作，还提供了允许提交作业和拒绝提交作业的设置操作。如图所示：

节点状态	系统类型	主机名	服务状态	CPU核数	CPU使用率	空闲内存	作业分布	负载诊断
☐	linux	mgmt.chess.com	在线	已用:0 总数:8	0.2%	10%	暂无数据	较高
☐	linux	node1	在线	已用:0 总数:12	0%	60%	暂无数据	正常
☐	linux	node2	在线	已用:0 总数:4	22.5%	10%	暂无数据	正常
☐	linux	node3.chess.com	在线	已用:0 总数:4	1.8%	10%	暂无数据	正常

5.8.2. GPU 列表

GPU 列表展示了每个节点上所有的 GPU 卡信息，包括 GPU 所在主机名、GPU 名称、使用率、温度、已使用显存、显存频率、处理器频率及 PCIe 读写带宽等信息，如图所示：

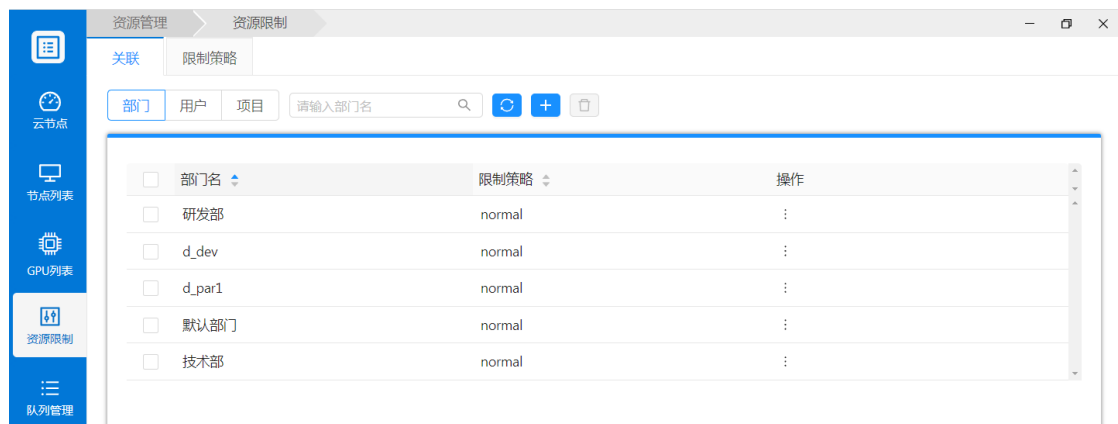
主机名	GPU名称	GPU使用率	GPU温度	已使用显存	显存频率	处理器频率	PCIe读写带宽
node1	GeForce GTX 1070	0%	48°C	0%	405MHz	139MHz	0MB/s 0MB/s
node2	Quadro M2000	0%	43°C	0%	405MHz	135MHz	0MB/s 0MB/s

5.8.3. 资源限制

资源限制中支持配置集群的最大运行作业数、最大提交作业数、CPU/内存的总资源限制等指标，如下图所示：



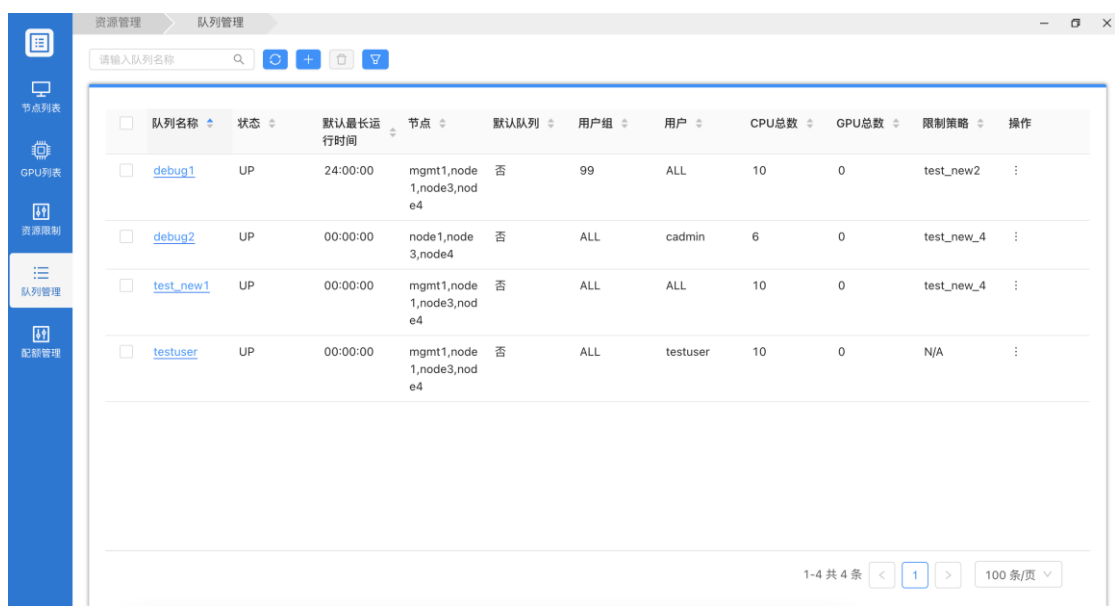
把设置的资源限制与用户或部门或队列进行关联，从而实现针对整个集群优化资源使用的目的。如图所示：



资源限制设置完成后，用户在首页的“资源信息”中可以查看到队列和存储的限制的资源信息。

5.8.4. 队列管理

队列管理支持集群队列的增、删、改、查，设置队列的使用节点、用户组或用户、限制策略等，如下图所示：



队列名称	状态	默认最长运行时间	节点	默认队列	用户组	用户	CPU总数	GPU总数	限制策略	操作
debug1	UP	24:00:00	mgmt1,node1,node3,node4	否	99	ALL	10	0	test_new2	:
debug2	UP	00:00:00	node1,node3,node4	否	ALL	admin	6	0	test_new_4	:
test_new1	UP	00:00:00	mgmt1,node1,node3,node4	否	ALL	ALL	10	0	test_new_4	:
testuser	UP	00:00:00	mgmt1,node1,node3,node4	否	ALL	testuser	10	0	N/A	:

5.8.5. 配额管理

配额管理则可对 Lustre、GPFS、NFS 这些文件系统设置用户的配额。先在页面上创建文件系统的关联信息，再在配额限制中设置用户的配额。如图所示：

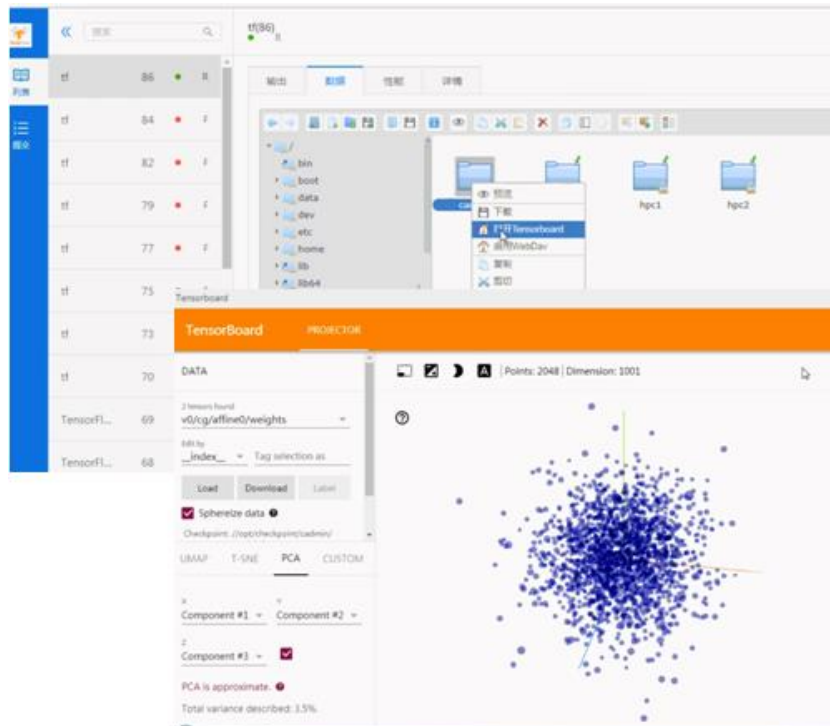


名称	挂载点	类型	已使用空间	总空间	设备名	挂载选项	启用
150home	/home	xfs	26.89%	900G	192.168.4.150:/home	rw,relatime,ver...	启用
150share	/share	xfs	21.74%	1.9T	/dev/sda2	rw,relatime,att...	启用
ansys	/ansys_inc	xfs	-	-	-	-	启用
gpfsdata1	/gpfsdata1	gpfs	0.63%	932G	gpfsdata1	rw,relatime ...	启用
gpfsdata2	/gpfsdata2	gpfs	0.46%	932G	gpfsdata2	rw,relatime ...	启用

5.10. 人工智能模块

CHESS 平台支持 AI 相关应用的集成，支持基于传统模式运行 AI 应用，也支持基于容器模式的 AI 应用调度集成。支持多种 AI 框架，例如 Tensorflow、PyTorch、

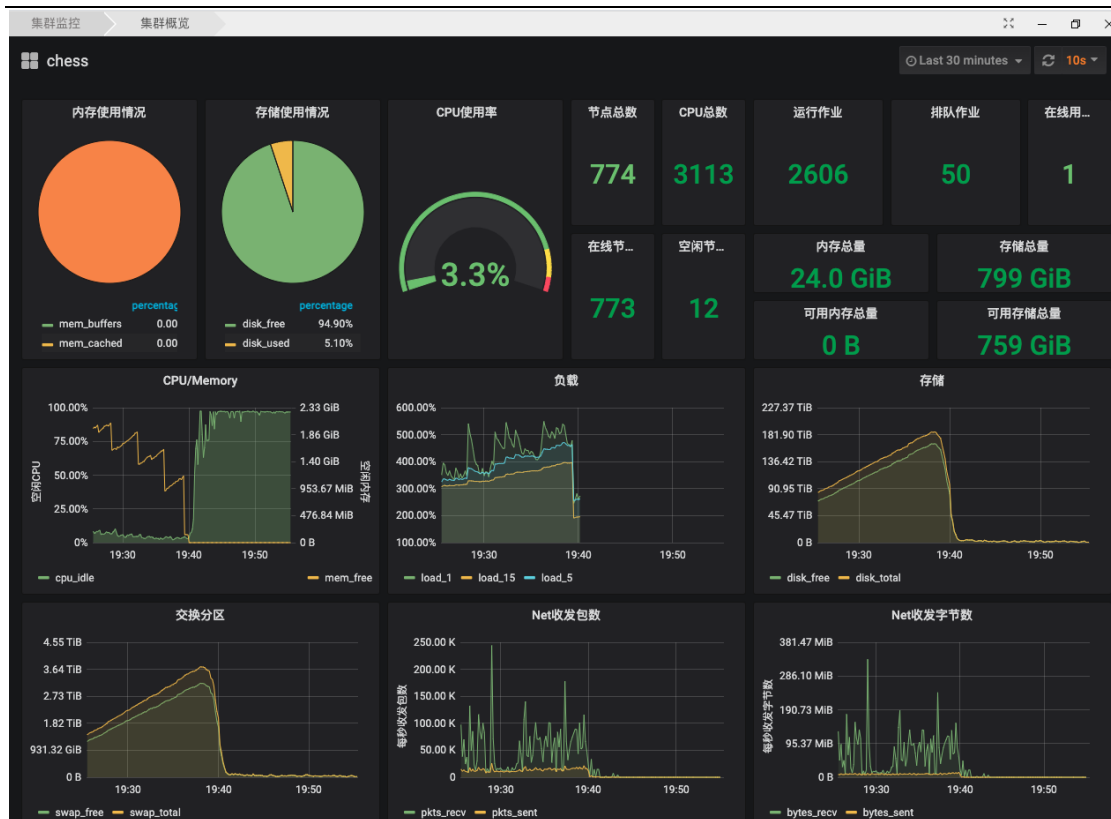
Caffe、MXNet 等，可与 Tensorboard 集成实现人工智能应用的集成调度与管理，支持容器模式及跨节点的推理操作，包括基于多机多卡、多机单卡或纯 CPU 多机多核模式的调度和计算：



同时，通过定制化的提交脚本和应用程序模版，可定制前后处理工作，实现容器服务的即起即用，不用不起的功能，减少资源占用。实现容器服务的自动起停。

5.11. 监控运维模块

监控运维模块是对目标监控对象实时数据的监控展示，主要为监控节点状态、限制节点资源、管理节点的常规开关机操作以及机柜图的相关信息的设置功能。从机柜图、节点列表、GPU 列表这些维度监控集群服务器的运行情况。



支持 AVX 监控的展示。

支持各 GPU 卡的监控，支持 IB 监控。

支持按照机房的服务器上架位置设定机柜图。在机柜图中实现了集群中各个节点的内存利用率、CPU 利用率、网络收发速率、swap 等信息进行监控。

当鼠标右键点击运行中状态的节点主机时，还提供关机、重启、VNC、硬关机、硬开机、控制台和控制台（新窗口）共 7 个功能选项，来为用户提供操作节点的便捷途径。

监控运维模块也支持监测平台的服务状态，当服务状态不正常时，可重启服务。

告警列表则可以设置资源监控的告警阈值，一旦资源占用达到阈值范围时，则产生告警信息，告警信息展示在告警列表中，在首页的右上角也有告警信息提示，若配置了邮件服务器和收件箱，也能自动发送告警邮件。告警记录如下图所示：

集群监控 告警列表

选择状态

状态	时间	主机名	警报类型
紧急告警	2021-04-01 05:05:00	node3	占用磁盘
紧急告警	2021-04-01 05:05:00	mgmt2	占用磁盘
紧急告警	2021-04-01 05:00:00	node4	占用磁盘
普通告警	2021-04-01 05:00:00	mgmt2	CPU利用率
紧急告警	2021-04-01 05:00:00	mgmt2	占用磁盘
紧急告警	2021-04-01 05:00:00	mgmt1.chess	占用磁盘
紧急告警	2021-04-01 05:00:00	node3	占用磁盘
紧急告警	2021-04-01 04:55:00	mgmt1.chess	占用磁盘
紧急告警	2021-04-01 04:55:00	node3	占用磁盘
普通告警	2021-04-01 04:55:00	mgmt2	CPU利用率
紧急告警	2021-04-01 04:55:00	node4	占用磁盘

1-100 共 18399 条

1 2 3 4 5 ... 184

100 条/页

设置页提供用户根据服务器物理位置自定义设置机柜图的展示和远程 IPMI 开关机的配置。

5.12. 计费模块

计费模块提供了详单、报表、余额和定价的功能。详单中展示了用户作业产生的实时费用信息。如果开启了用户扣费，并给用户充值了，还可看到用户的当前余额。如下图所示：

作业ID	开始时间	结束时间	CPU核数	CPU定价	内存	内存定价	GPU卡数	扣费单位	运行时间	收费(元)
270915	2021-06-03 10:07:35	2021-06-03 10:09:16	1	1	126MB	0	0	cadmin	00:01:41	0.03
270914	2021-06-03 10:03:52	2021-06-03 10:05:33	1	1	126MB	0	0	cadmin	00:01:41	0.03

余额: 999.94元 1-2 共 2 条 100 条/页 总费用: 0.06元

余额中提供了预充值功能，开启预付后，可分别给不同的部门或用户开启扣费，并充值。若余额开启预付后，部门或用户又不充值，则不能提交作业。支持部门或用户余额的增、减，和余额变更查看，支持短信和邮件的余额提醒设置。

定价则从用户、部门或队列的维度，分别设置不同的计算资源的计费单价，也支持从文件系统、用户或部门的维度，结合不同的阶梯系数，设置存储的计费单价。如下图所示：

队列	CPU 定价(元/(核·小时))	内存 定价(元/(GB·小时))	GPU 定价(元/(卡·小时))	操作
*	1	0	1	-

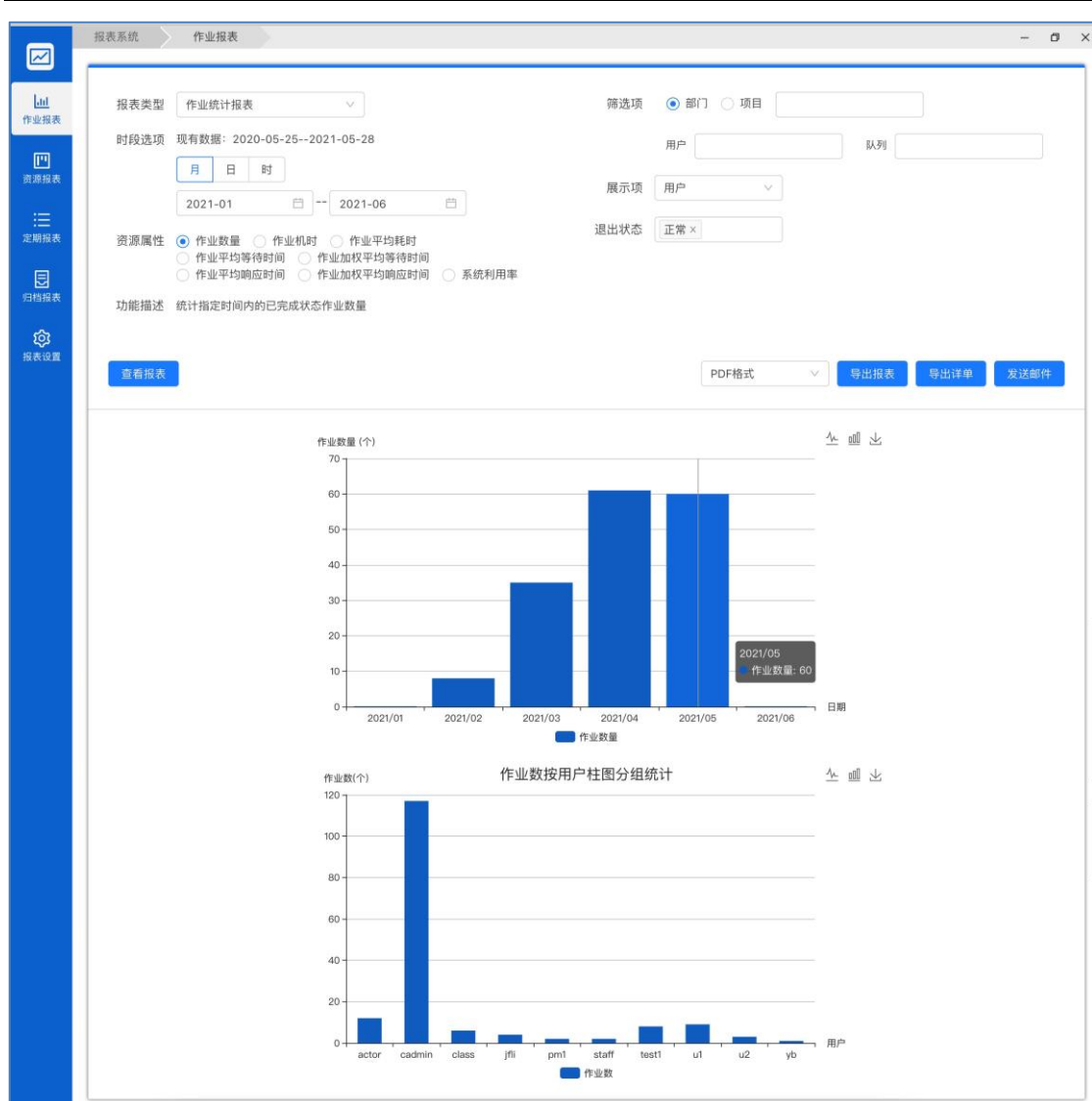
计费报表则能按时段统计作业收费报表和存储收费报表。提供小时报、日报、

月报和年报表，报表导出支持：HTML、PDF 和 excel 详单三种格式。报表支持组织架构或扣费单位两种维度展示。组织架构展示是指按用户管理的部门层级结构中，把产生了费用的部门全部显示出来；而扣费单位则是指仅仅把在余额中开启了扣费的部门或用户，同时产生了费用的部门或用户显示出来。如下图所示：



5.13. 集群报表模块

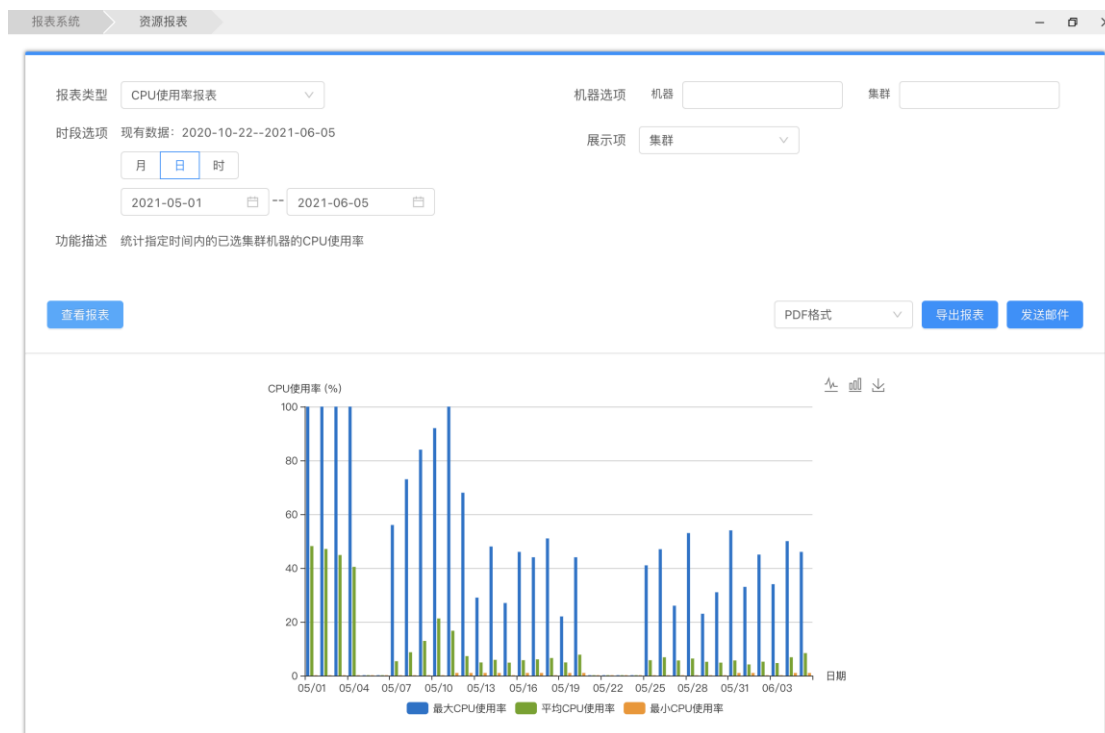
集群报表中记录了集群中所有用户的作业数据，可根据用户、部门、队列和项目这些搜索条件，从小时、天、月的维度查询相应的作业数据展示。如图所示：



支持导出 PDF 或 HTML 格式的报表。

还可以导出报表的详细数据，保存为 excel 文档。

资源报表包含了集群的 CPU、GPU、内存、存储使用率报表和机器可用性报表。分别从不同的维度对集群资源的使用情况进行了统计。如下图所示，CPU 使用率报表展示了一段时间范围内，集群的 CPU 的使用情况：

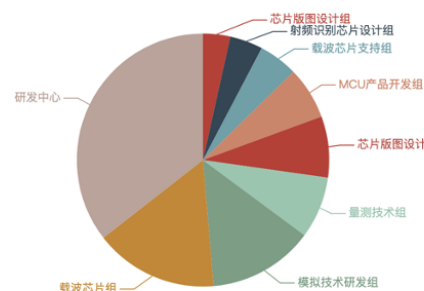
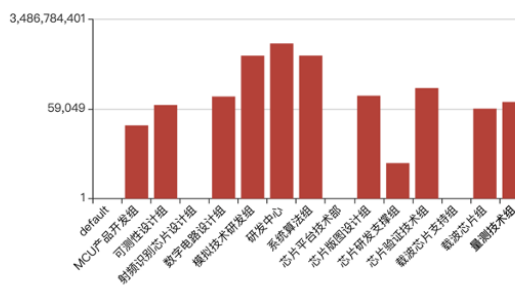


内存使用率报表则展示了一段时间范围内，集群的总内存使用情况。

还可以根据用户的需求编写 SQL 查询语句，定制报表，如下图所示，展示了定制的各部门的 CPU 和资源成本统计报表：

各部门 CPU 成本对比图

各部门资源成本占比统计图



5.14. 系统设置模块

系统设置模块提供了日志级别的设置、许可授权的更新、系统消息管理和在

线升级的功能。日志设置中可以设置输出路径、日志级别和日志的展示格式，如图所示：

系统设置 > 日志

日志设置

日志等级： 错误 警告 消息 **调试** 跟踪

输出路径： ☒ 启用

Syslog: UDP IP ☐ 禁用

Format: **text** json

许可授权中提供更新平台 License 的功能。

消息管理提供了管理员通知和系统公告的增删改查的功能，如图所示：

标题	类型	接收用户	接收部门	内容	发送时间	已读人数	操作
数据自动清理通告	管理员通知	user1	所有部门	您90天前的历史数据将在7天后的202...	2021-12-20 23:02:25	0/1	删除
notice	首页通知	所有用户	所有部门	System upgrade, Please	2021-11-08 18:22:06	所有用户	删除
notify	首页通知	所有用户	所有部门	Notify Test	2021-11-08 18:20:46	所有用户	删除

在线升级提供了平台的不断业务的情况下，对平台版本的升级功能，如图所示：





日志



许可授权



消息管理



在线升级

系统设置

在线升级

版本

6.6.0

更新安装包

[点击这里可上传新版本的安装包](#)

第6章 应用场景

6.1. CAE 仿真

CAE 主要是指在计算机上进行的基于现代计算力学理论的数值仿真技术，广泛应用于航空航天、汽车、船舶、机械、建筑、电子等行业领域，CAE 领域软件大部分采用 MPI 并行编程环境进行编译，采用并行计算的速度较原来有较大的提升。

应用：采用 CHESS 高性能计算管理平台可为 CAE 提供一套完整的高性能计算环境软件包，将 CAE 应用软件与调度系统相结合，提供应用模板，方便作业提交和查看结果。

常用软件：

HFSS	高频结构仿真
COMSOL	物理场建模与仿真
Feko* (EMSS)	三维全波电磁仿真软件
ABAQUS	工程模拟有限元软件
ANSYS	大型通用有限元分析软件
LS-DYNA	瞬态响应动力学分析软件
FLUENT	通用计算流体力学分析软件
Maxwell	有限元分析软件。
Hspice	集成电路性能分析模拟程序

6.2. 人工智能

人工智能 (Artificial Intelligence)，英文缩写为 AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。它是计算机科学的一个分支，目的是了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来，理论和技术日益成熟，应用领域也不断扩大，可以设想，未来人工智能带来的科技产品，将会是人类智慧的“容器”。人工智能可以对人的意识、思维的信息过程的模拟。

应用：CHESS 在支持原有 HPC 应用基础上，扩充支持人工智能框架。可支持基于 Docker 容器的人工智能框架的调度，也可支持基于 Singularity 容器的分布式人工智能框架的调度。可以在 TensorFlow 运行时使用集成 TensorBoard 工具随时查看训练过程。

常用软件：

TensorFlow	深度学习框架
Caffe	深度学习框架
MXNet	深度学习框架

6.3. 基础学科

基础科学研究是指认识自然现象、揭示自然规律，获取新知识、新原理、新方法的研究活动，主要应用领域包括高能物理、计算化学、纳米材料等。理论、实验、计算是当今科研创新的三大重要手段，高性能计算主要通过模拟仿

真的技术把目前通过实验无法实现的想法得以实现，在降低研发成本的基础上加快人们对自然世界的认识。例如，使用并行程序进行密度泛函理论（DFT）计算已经成为材料科学、固体物理、计算化学等领域内必不可少的研究手段之一。并行计算已经广泛应用于基础学科的研究，用户主要为高校以及科研院所。

应用：CHESS 高性能计算管理平台针对高校以及科研院所的使用特点提供一套完整的软件解决方案，支持多种复杂并行环境与应用程序，多种作业调度策略，为用户提供软件支撑平台。高校与科研院所有用户多、管理复杂、数据统计、以及用户水平参差不齐等问题，CHESS 完善的集群管理、丰富的数据报表统计，以及 B/S 架构采用 Web Portal 的使用方式可帮助用户解决此类问题，让用户更多的精力专注于自己的科研成果，帮助用户搭建一套高效、稳定的集群系统。

常用软件：

VASP	原子尺度材料模拟的计算机程序包
QCHEM	量子化学计算软件
NWCHEM	量子化学计算软件
GAUSSIAN	量子化学计算软件
MOLPRO	量子化学计算软件
Materials Studio	分子模拟软件
GAMESS	量子化学计算软件

6.4. 生命科学

生命科学是研究生命现象、生命活动的本质、特征和发生、发展规律，以及各种生物之间和生物与环境之间相互关系的科学。运用大规模高效的理论模

型和数值计算来识别基因组序列中代表蛋白质的编码区，破译隐藏在核酸序列中的遗传语言规律。用于有效地控制生命活动，能动地改造生物界，造福人类生命科学，与人类生存、人民健康、经济建设和社会发展有着密切关系，是当今在全球范围内最受关注的基础自然科学。

应用：CHESS 高性能计算管理平台可为生命科学高性能计算用户配备完备的并行软件开发和运行环境，配置 CHESS Monitor 和 CHESS Schedule 模块进行集群监控以及作业调度，包含针对不同应用软件的模板设置，提高用户的使用效率、降低高性能计算机的使用门槛。

常用软件：

BLAST	序列相似性检索程序
FASTA	序列搜寻
ClustalW	对序列对比
Hmmer	搜索序列蛋白质序列与序列对比
DOCK	分子半柔性对接
ZDOCK	刚性对接
RDOCK	蛋白质对接
MORDOR	柔性对接

6.5. 石油勘探

高性能计算集群广泛适用于石油勘探行业，主要应用于油藏模拟分析、地震油层分析、油层数据可视化方面。其中地震油层分析是利用地震波法用炸药在地面激起人工地震波，这种地震波可传入地下深处，并在碰到不同形态的地质构成时形成不同的反射波，这些反射波经地面的检波器收集、转变成电子信

号后可存储为数据，而后利用高性能服务器进行地震油层分析任务，利用加工处理野外观测所得地震波数据，将地震数据变成地质语言——地震剖面图或构造图，处理后的地震数据可根据地下岩层的产状和构造关系，找出有利的含油气地区。

应用：CHESS 高性能计算管理平台广泛应用于石油行业，可监控所有计算资源，实现集群远程管理，同时可提供 CGG 和 Omega 等软件的运行情况和使用报表。

常用软件：

ProMax	地震资料处理
SeisSpace	
Geodepth	
Omega	
Geocluster	
Nexus	油藏模拟
Eclipse	
RMS	
Geoprobe	油层数据可视化
Petrel	
VoxelGeo	

6.6. 气象行业

气象行业对高性能计算需求十分强烈，其需求主体是数值天气预报模式的科学研究和业务运行。天气预报的准确率、时效性、精细化程度和服务水平提高都与高性能计算集群密不可分。数值天气预报涉及的应用主要包括气象模式、海洋模式、气候模式、环境模式及卫星、遥感资料处理程序等。各种模式在流程上有一定的相似性，都包括数据前处理，主模式计算和结果后处理三部分。数据前处理包括地形和地表资料、气象预报背景场资料、各种气象观测资料(观测站、雷达、卫星等)的处理和同化，为模式计算提供边界场和初值，主模式计算是主要计算量所在，通过并行计算进行数值方程的迭代求解。

应用：气象、海洋模式对计算以及 I/O 要求较高，联智科技在气象行业有丰富的行业经验，CHESS 高性能计算管理平台解决用户常规管理、集群监控，为用户配备完备的并行计算开发和运行环境，针对不同模式可采用不同的应用模板进行提交作业；针对气象过程模拟涉及海量小文件的处理，为气象行业提供内存文件系统高带宽、高 IOPS 的解决方案，为用户解决 I/O 瓶颈；提供常用数值模式的安装、调试和培训服务，并配合用户完成数值预报业务系统的安装调试。

常用软件

MM5	中尺度模式
GRAPES	气象预报模式系统
CESM	社区地球系统模型
APRS	高级区域预报系统
WRF	中尺度预报模式
CAM	大气环流模式

Regcm3	区域气候模式
CCM3	大气环流模式
IFS	集成预报系统模式
CGCM	全球海气耦合模式