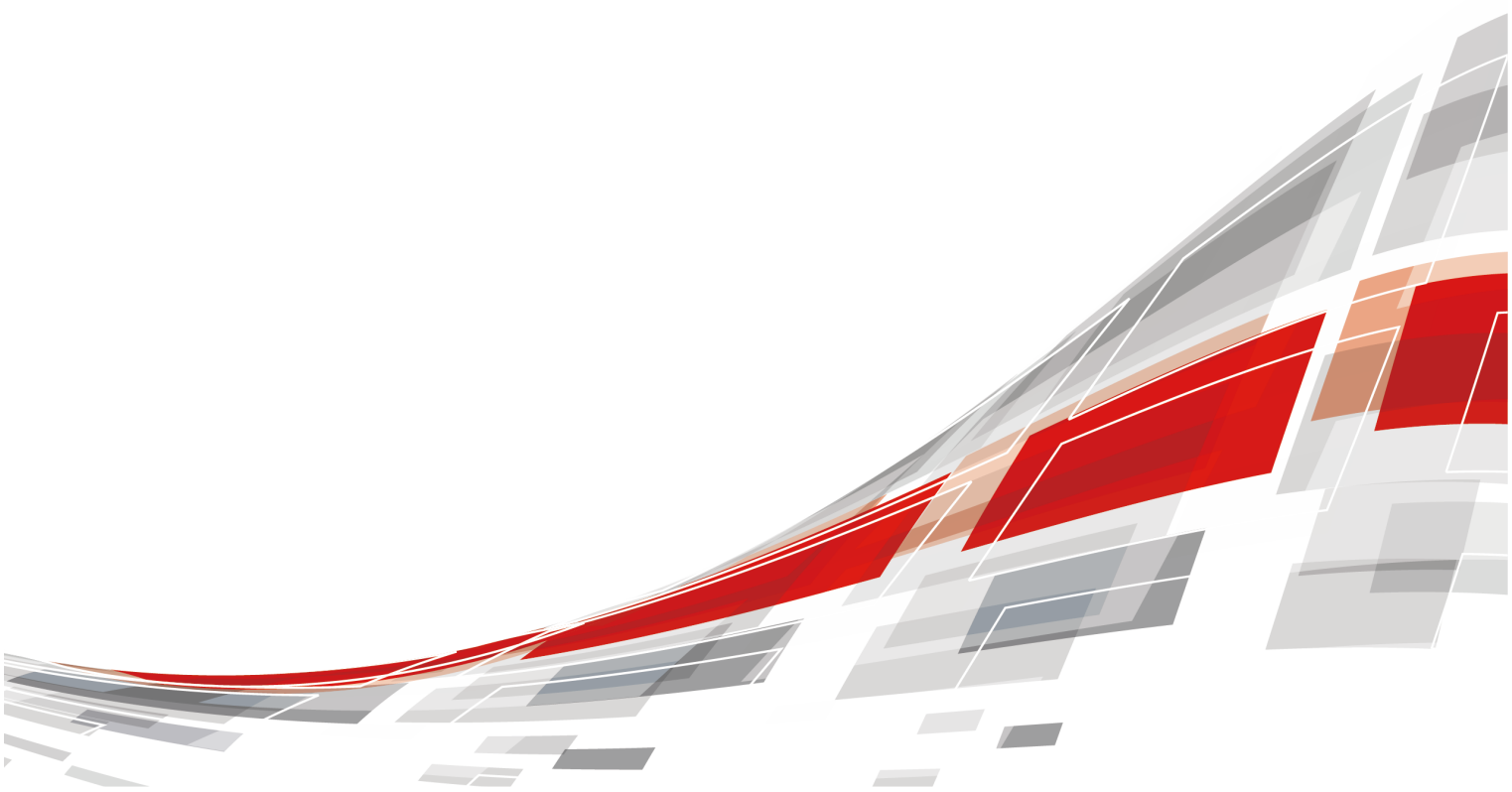
FusionOne HPC 23.0.3 技术白皮书

文档版本 01

发布日期 2024-01-30



超聚变数字技术有限公司

版权所有 © 超聚变数字技术有限公司 2024。 保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明

FUSION 和其他超聚变商标均为超聚变数字技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

本文中，只是为了描述的简洁和方便理解，用“xFusion”指代“xFusion Digital Technologies Co., Ltd.”，这并不代表“xFusion”还可以具备其它含义。基于本文中单独提及或描述的“xFusion”，不能用于“xFusion Digital Technologies Co., Ltd.”之外的理解或表达，超聚变数字技术有限公司也不承担因单独使用“xFusion”所带来的其它任何法律责任。

您购买的产品、服务或特性等应受超聚变数字技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，超聚变数字技术有限公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

超聚变数字技术有限公司

地址： 河南省郑州市郑东新区龙子湖智慧岛正商博雅广场1号楼9层 邮编：450046

网址： https://www.xfusion.com

前言

概述

本文档主要介绍服务器FusionOne HPC技术，读者可以通过本文档了解FusionOne HPC的技术方案。

读者对象

本文档供销售工程师、渠道商、高级业务经理以及希望使用FusionOne HPC解决方案和FusionServer服务器的客户查阅。

符号约定

在本文中可能出现下列标志，它们所代表的含义如下。

符号	说明
危险	表示如不避免则将会导致死亡或严重伤害的具有高等级风险的危害。
警告	表示如不避免则可能导致死亡或严重伤害的具有中等级风险的危害。
注意	表示如不避免则可能导致轻微或中度伤害的具有低等级风险的危害。
须知	用于传递设备或环境安全警示信息。如不避免则可能会导致设备损坏、数据丢失、设备性能降低或其它不可预知的结果。 “须知”不涉及人身伤害。
说明	对正文中重点信息的补充说明。 “说明”不是安全警示信息，不涉及人身、设备及环境伤害信息。

版权所有 © 超聚变数字技术有限公司

修改记录

文档版本	发布日期	修改说明
01	2024-01-30	第一次正式发布。

 版权所有 © 超聚变数字技术有限公司

目 录

前言 ii
1 产品概述 1
1.1 背景 1
1.2 应用场景 2
1.2.1 教育科研 2
1.2.2 制造行业 5
1.3 方案优势 6
2 技术方案 7
2.1 方案架构 7
2.2 硬件平台 7
2.2.1 计算节点 7
2.2.1.1 高密服务器 8
2.2.1.2 机架服务器 10
2.2.1.3 整机柜服务器 12
2.2.1.4 GPU 服务器 13
2.2.2 存储 13
2.2.2.1 FusionOne HPC NFS 13
2.2.2.2 FusionOne HPC DFS 14
2.2.2.2.1 功能架构 17
1. 对称式架构 17
2. 非对称式架构 17
3. 全闪架构 18
2.2.2.2.2 软件架构 19
2.2.2.2.3 组网规划 20
1. 对称式架构组网规划 21
2. 非对称式架构组网规划 22
2.2.2.2.4 关键特性 23
1. 数据冗余策略 23
2. 快速数据重构 25
3. 小文件聚合 26
4. 文件多协议互通 26
5. 快照和克隆 27

 版权所有 © 超聚变数字技术有限公司

6. 负载均衡.....28
7. 配额.....28
8. Cache 策略.....28
9. 文件流转跟踪.....29
10. 操作日志.....29
11. 访问权限控制.....29
12. 扩容.....30
13. 容量管理.....32
14. 数据迁移.....33
15. 集群监控.....33
16. 告警管理.....35
17. 集群双活.....35
18. 安全审计.....36
19. 升级维护.....36
20. 接口能力.....37
21. 数据扫描.....37
22. 数据隔离.....37
23. 传输协议安全.....37
24. QOS 管理.....37
25. 块设备管理.....37
26. 对象存储.....37
2.2.3 网络.....38
2.3 全栈软件.....39
2.3.1 FusionOne Center for HPC.....39
2.3.1.1 OneMind.....39
2.3.1.2 用户管理.....42
2.3.1.3 数据管理.....43
2.3.1.4 应用管理.....43
2.3.1.5 作业管理.....45
2.3.1.6 前后处理.....46
2.3.1.7 应用特征分析.....46
2.3.1.8 统计中心.....48
2.3.1.9 计费统计.....48
2.3.1.10 多集群.....49
2.3.1.11 弹性上云.....50
2.3.2 AI 平台.....51
2.3.2.1 数据管理.....51
2.3.2.2 算法开发.....52
2.3.2.3 训练管理.....53
2.3.2.4 模型管理.....54
2.3.2.5 推理服务.....55
2.3.2.5.1 在线服务.....55

 版权所有 © 超聚变数字技术有限公司

2.3.2.5.2 批量服务.....56
2.3.2.6 镜像管理.....56
2.3.2.7 租户与计费管理.....57
2.3.2.8 控制台.....57
2.3.2.9 统计报表.....58
2.3.3 FusionOne Center 平台.....58
2.3.3.1 统一监控.....58
2.3.3.2 统一告警.....59
2.3.3.3 统一日志.....59
2.3.3.4 智能故障管理.....59
2.3.3.5 服务器管理.....60
2.3.3.6 智能资产管理.....60
2.3.3.7 智能能效管理.....61
2.3.3.8 智能部署管理.....61
2.3.3.9 FusionOnline.....62
2.3.4 系统环境.....62
2.3.4.1 操作系统.....62
2.3.4.2 开发及并行运行环境.....62
3 安全可信.....63
3.1 安全.....63
3.1.1 行业特点与外部安全需求.....63
3.1.2 超聚变内部安全要求与产品安全能力.....63
3.1.3 FusionOne HPC 安全架构模型.....64
3.2 可靠.....67
4 服务与支持.....68
4.1 实施服务.....68
4.1.1 服务概要.....68
4.1.2 目标.....68
4.1.3 服务内容.....68
4.1.4 项目交付流程.....69
4.2 超凡服务.....69
4.2.1 服务概要.....69
4.2.2 服务内容.....70
4.2.2.1 保修技术支持服务.....70
4.2.2.2 介质保留服务.....70
4.2.3 超聚变服务理念.....70
4.2.4 技术支持与服务目标.....70
4.2.5 超聚变技术支持体系.....71
4.3 SNS 服务.....72
4.3.1 服务概要.....72
4.3.2 服务内容.....72

 版权所有 © 超聚变数字技术有限公司

1 产品概述

1.1 背景
1.2 应用场景
1.3 方案优势

1.1 背景

高性能计算HPC（High Performance Computing）即并行计算系统，能够执行一般电脑无法处理的海量数据与高速运算。HPC将大规模运算任务拆分并分发到各个服务器上进行并行运算，再将计算结果汇总得到最终结果，从而实现强大的计算功能。

HPC解决方案已经广泛应用于各行业，但仍然面临诸多挑战：

- 性能瓶颈

  随着高性能计算在各个行业领域的普及，越来越多的行业，越来越多的数据需要在HPC平台上进行计算和分析，对算力的需求越来越强烈。

  算力成本昂贵，如何最大限度提升算力资源特别是GPU资源的利用率。
- 能耗居高不下

  高性能计算发挥的效用越显著，随之带来的高能耗、高排放就越无法忽视。大多数传统风冷数据中心实际上处于PUE（数据中心总能耗/IT设备能耗）偏高的运行状态，数据中心低效的部署和运维造成了极大的能源损失。为此，围绕数据中心的“低碳”需求变得愈发迫切。
- 使用复杂
  - HPC集群用户多为非IT从业人员，不擅长使用命令行方式进行任务管理和数据管理操作。
  - 资源管理的维度不是基于客户视角设计，使用之前需要先学习概念。
  - 资源管理配置比较复杂，用户不能快速掌握，配置容易出错。
- 行业应用复杂

  HPC应用行业复杂，多种类型软件并存，比如碰撞模拟LS-Dyna、Simulia Abaqus explicit等软件需持续升级。

  HPC与AI融合趋势下，需要同时运行HPC作业和AI作业。
- 数据安全问题

  数据频繁传输，效率和安全性问题需要保障。

 版权所有 © 超聚变数字技术有限公司

- 异常监控

  需及时发现集群中的异常、实现多集群统一管理、用户异常行为监控，异常监控项多变。

为应对以上挑战，超聚变FusionOne HPC解决方案应运而生。HPC为客户提供一体化高性能计算平台，通过软硬件的深度融合和自研OneMind智慧引擎，为客户提供运行高效、安全可靠、运维简单和绿色节能的智慧HPC解决方案，满足教育科研、制造仿真等行业高性能计算业务场景。

超聚变FusionOne Center for AI一站式平台（下文简称AI平台）是超聚变面向AI模型生产的生命周期，推出的集数据管理(数据集管理、智能标注和数据增强)、模型开发、模型训练和模型管理等功能于一体的AI算法构建平台，旨在降低用户算法开发门槛，快速推出面向各行各业的智能算法。

1.2 应用场景

FusionOne HPC解决方案可用于教育科研、泛制造、生命科学、气象和海洋等多个行业，为科学突破、众多行业的产品开发等提供强大算力基础设施和算力服务。

1.2.1 教育科研

场景描述

高性能计算在科研领域有着广泛的应用，包括生物的基因工程、化学的分子研究、物理的天体物理、数学的大型计算任务等都会用到高性能计算，大量使用开源软件和免费软件，软件类别较多，对硬件的需求多样化。

如**表1-1**对部分软件特点进行简单介绍。

版权所有 © 超聚变数字技术有限公司

表 1-1 软件介绍

名称	概述	软件特点
VASP	VASP（Vienna Ab-inito Simulation Package）是维也纳大学Hafner小组开发的进行电子结构计算和量子力学-分子动力学模拟软件包。它是目前材料模拟和计算物质科学研究中十分流行的商用软件之一。VASP是基于赝势平面波基组的第一性原理密度泛函计算程序，VASP软件作为目前国内国际上权威的第一性原理计算软件，可以研究多种体系，包括金属及其氧化物、半导体、晶体、掺杂体系、纳米材料、分子、团簇、表面体系和界面体系等。	VASP软件通过近似求解Schrödinger方程得到体系的电子态和能量，既可以在密度泛函理论（DFT）框架内求解Kohn-Sham方程（已实现了混合泛函计算），也可以在Hartree-Fock（HF）的近似下求解Roothaan方程。此外，VASP也支持格林函数方法（GW准粒子近似，ACFDT-RPA）和微扰理论（二阶Møller-Plesset），支持屏蔽交换（Screened Exchange）。 VASP软件为基于第一性原理的量子力学和分子动力学（从头分子动力学，ab initio MD）计算软件，在VASP的分子动力学中，每一步都采用有效矩阵对角和Pulay方法混合求解瞬时电子基态，能够指定温度、压力、总时间、时间步长等参数，系综方面支持蒙特卡洛计算具代表性的正则系综（NVT）和分子动力学计算具代表性的微正则系综（NVE）。 VASP软件能够实现大规模的高效率并行计算，支持多核多节点并行计算，对核数和节点数均没有限制，支持单用户多用户同时使用。
GROMACS	GROMACS是用于研究生物分子体系的分子动力学程序包。它可以用分子动力学、随机动力学或者路径积分方法模拟溶液或晶体中的任意分子，进行分子能量的最小化，分析构象等。它的模拟程序包包含GROMACS力场（蛋白质、核苷酸、糖等），研究的范围可以包括玻璃和液晶、到聚合物、晶体和生物分子溶液。GROMACS是一个功能强大的分子动力学的模拟软件，其在模拟大量分子系统的牛顿运动方面具有极大的优势。	GROMACS进行了大量的算法的优化，使其计算功能更强大。例如：在计算矩阵的逆时，算法的内循环会根据自身系统的特点自动选择由C语言或Fortran来编译。GROMACS中对Altivec loops的计算，无论是在Linux还是MacOSX.系统上，它都要比其它软件快3-10倍，而且GROMACS提高计算速度的同时也保证了计算精度。GROMACS具有友好的用户界面，拓扑文件和参数文件都以文档的形式给出。在程序运行过程中，并不用输入脚本注释语言。所有GROMACS的操作都是通过简单的命令行操作进行的。而且运行的过程是分步的，随时可以检查模拟的正确性和可行性，可以减少时间上的浪费。GROMACS允许并行运算，使用标准的MPI通讯。

 版权所有 © 超聚变数字技术有限公司

名称	概述	软件特点
NAMD	NAMD（NAnoscale Molecular Dynamics）是用于在大规模并行计算机上快速模拟大分子体系的并行分子动力学代码。NAMD用经验力场，如Amber，CHARMM和Dreiding，通过数值求解运动方程计算原子轨迹。	与大多数同类软件不同，NAMD的并行模式建立在Charm++提供的并行框架之上。Charm++是一种面向对象、跨平台的C++库，为并行编程提供高层抽象。在Charm++的并行模型中，程序被划分成中等粒度的重复任务，由称为chare的对象执行，chare之间通过消息相互作用。Chare到物理处理单元的映射和其间消息的传递由Charm++运行系统负责完成。该模型通过消息驱动的异步并发模式实现通信与计算的重叠，另一方面通过对Chare的调度实现动态负载均衡。在此基础之上，NAMD采用了空间/力混合划分模式来划分并行任务，相对于其他主要采用空间划分实现并行的同类软件对计算的划分粒度更细。该模式在其后为Blue Matter和Desmond采用。在这种并行模式支持下的高可扩展性成为NAMD最重要的特色。
LAMMPS	LAMMPS（Large-scale Atomic/Molecular Massively Parallel Simulator）大规模原子分子并行模拟器，主要用于分子动力学相关的一些计算和模拟工作。	LAMMPS可以支持包括气态，液态或者固态相形态下、各种系统下、百万级的原子分子体系，并提供支持多种势函数，且LAMMPS有良好的并行扩展性，采用的是消息影响和模拟域的空间分解并行机制。
CP2K	CP2K是一个量子化学和固态物理软件包，可以对固态，液态，分子，周期性，材料，晶体和生物系统进行原子模拟。	CP2K为不同的建模方法提供了通用框架。支持的理论水平包括DFTB，LDA，GGA，MP2，RPA，半经验方法（AM1，PM3，PM6，RM1，MNDO等）和经典力场（AMBER，CHARMM等）。CP2K可以使用NEB或二聚体方法进行分子动力学、元动力学、蒙特卡洛、埃伦菲斯特动力学、振动分析、核心能谱、能量最小化和过渡态优化的模拟。

高校的HPC集群资源按规模递减有校级平台、院系级、教授自有等。校级平台是多学科、跨院系的，一般由信息中心的专业人员维护。院系级的业务领域较为单一，规模较小，一般由老师、学生自己维护。从全校的角度出发，各院系高性能计算资源建设成本很高，而计算资源和研究成果之间难以实现共享。建设一个高性能计算资源池，为各院系以及师生提供支持多种复杂高性能运算业务的公共计算平台，成为很多高校的诉求。

业务挑战

- 学科分散，各个学科的计算需求爆炸式增长。HPC与AI应用软件日新月异，计算工具种类繁多。

 版权所有 © 超聚变数字技术有限公司

- 运维复杂，运维人员需要看护应用、中间件、计算、存储、网络等多个部件，致使运维压力过大。
- 最终用户的IT能力弱，对简化集群管理和容器化有较强诉求。
- 院系级HPC资源往往无法有效统筹，不成规模，重复投资建设造成资源浪费。

1.2.2 制造行业

场景描述

随着数字化时代的到来，越来越多的制造类企业已经完成数字化转型，其中以工业仿真业务尤为突出。工业仿真一般分为前处理、求解和后处理三个阶段，对算力要求最高的是中间求解阶段，需要在高性能计算平台上进行计算，而前处理、后处理则需要进行CAE模型可视化，一般在高性能的图形工作站上完成。

以汽车研发为例，汽车研发流程中70%的时间是基于数字模型仿真。数字模型仿真能够缩短上市时间和提高产品质量，所以各车企在产品开发中都建立了相应的仿真流程。

借助计算机辅助工程CAE（Computer Aided Engineering），企业研发人员使用计算机辅助求解分析复杂工程和产品的力学性能，以及性能优化等，完成产品设计快速迭代，提升企业竞争力。

- 前处理

  对工程或产品进行建模，建立合理的有限元分析模型（采用各种CAD工具，建立几何模型，划分计算网格）。

- 求解

  指定荷载和边界条件，提交给HPC进行分析。按照计算每一时刻动力反应是否需要求解线性方程组，可将直接积分法分为隐式积分方法和显式积分方法两类。

  - 隐式积分法：根据当前时刻及前几时刻体系的动力反应值建立以下一时刻动力反应值为未知量的线性方程组，通过求解方程组确定下一时刻动力反应。隐式方法的研究和应用由来已久，常用的方法有线性加速度法、平均加速度法等。
  - 显式积分法：可由当前时刻及前几时刻的体系动力反应值直接外推下一时刻的动力反应值，不需要求解线性方程组，实现了时间离散的解耦。解方程组一般占整个有限元求解程序耗时的70%左右，因此，这一解耦技术对计算量的节省是很可观的。

- 后处理

  根据工程或产品模型与设计要求，对有限元分析结果进行用户所要求的加工、检查，并以图形方式提供给用户，辅助用户判定计算结果与设计方案的合理性（显示计算结果，评估产品性能）。

业务挑战

- 应用软件多样化对算力种类的挑战

  CAE领域应用软件众多，每款软件对算力形态的需求各不相同。一套功能完备、性能适配的高性能计算机集群系统建设，不仅需要与细分领域的仿真工程师充分沟通，了解工程师对算力的需求，还要结合细分领域软件对计算形态的需求。

- 前后处理的挑战

  CAE领域的建模、前后处理工具对图形工作站硬件要求较高（高主频CPU、大容量内存以及高端图形显示卡），这些硬件特性能够保障仿真工程师的图形操作体

 版权所有 © 超聚变数字技术有限公司

验，有效增强CAE仿真研发生产力。图形工作站一般放置于仿真工程师的办公桌面，每次CAE仿真求解时，需要将本地工作站上大量的建模数据通过园区网络上传到高性能计算集群系统，仿真求解完成后，将结果数据下载到本地工作站通过后处理工具进行查看或优化。往复的数据传输可能会导致仿真工程师无法及时提交计算任务，无法及时获取仿真结果，研发效率大打折扣。

- Portal远程可视化挑战

  CAE领域仿真过程中，仿真工程师习惯于将计算初始条件通过可视化的方式进行设置，同时在求解过程当中又需要密切关注求解过程产生的日志信息和收敛曲线变化情况。这些可视化的需求在一般的命令行作业调度方式中难以实现，导致仿真工程师在完成CAE仿真的工作之外，还需投入时间学习这些命令行作业调度工具的使用方法。

- 突发大量算力需求的挑战

  在CAE仿真集群建设前期进行大量的调研、统计和沟通，以满足企业日常仿真所需，但在企业的日常生产过程中总会出现某阶段仿真任务量非常集中，导致排队的仿真任务数量大量积压，仿真工程师无法及时获取仿真结果，影响研发效率。

1.3 方案优势

FusionOne HPC解决方案深入理解客户诉求，重点在下述方向提供更好能力，帮助客户应对自身HPC业务问题，产生更好的收益。

项目	描述
运行高效	● 提供业界最佳高密服务器、高性能基础硬件平台。 ● FusionOne HPC DFS高性能文件存储方案。 ● OneMind加速引擎，软硬协同创新，应用运行效率提升10%。
绿色节能	● OneMind应用感知智能降频、调度技术，整体节能10%。 ● FusionOne HPC全液冷方案，PUE低至1.1。
运维简单	● 提供FusionOne Center for HPC智能管理平台和FusionOne Center for AI深度学习平台。 ● 便捷的计算资源、应用资源管理调度，图形化与模板化作业和数据管理。 ● 全方位集群运维管理与运营分析，HPC及AI集群统一管理。 ● 易学习，快速上手。
安全可信	协议安全、数据安全多层级安全防护，异常入侵检测。
融合开放	开放兼容东西方产业生态，包括HPC相关芯片、关键硬件到OS、集群调度、行业应用软件全面开放兼容，积极参与行业和社区交流。

版权所有 © 超聚变数字技术有限公司

2 技术方案

2.1 方案架构
2.2 硬件平台
2.3 全栈软件

2.1 方案架构

HPC集群可以分为基础设施层、硬件平台层、全栈软件层、应用软件层。

- 基础设施层
  主要包含了机房基础设施，机房空调、动能设施、液冷设施等。
- 硬件平台层
  包含了计算、存储、网络子系统。
  - 计算子系统：指HPC集群中的管理节点、头节点、计算节点、胖节点和加速节点等各类服务器。
  - 存储子系统：包括IO服务器节点等HPC存储硬件。
  - 网络子系统：FusionOne HPC支持Infiniband网络，RoCE网络
- 全栈软件层
  系统环境：包含操作系统、文件系统和应用运行环境如编译器、数学库和MPI通信库等。
  管理调度软件：包含FOC统一平台、HPC集群管理、深度学习平台
- 应用软件层
  各行业的开源和商业版软件。这些软件通过管理调度软件进行管理和监控。

2.2 硬件平台

2.2.1 计算节点

FusionOne HPC FusionServer高性能计算服务器涵盖高密、机架、异构和整机柜服务器。

 版权所有 © 超聚变数字技术有限公司

2.2.1.1 高密服务器

FusionServer X6000 V6 服务器

图 2-1 X6000 V6（示例：24x2.5 英寸硬盘配置）



- 超高密度、节省空间

  相对于传统机架服务器的部署密度翻倍，节省机房空间。

  - 在相同机柜空间内，计算密度是传统1U机架服务器的2倍、传统2U机架服务器的4倍，大大提高机房空间利用率。
  - X6000 V6高度2U，最多支持4个服务器节点。节点支持顶配270W CPU，整框算力可达25TFlops。
- 汇聚管理、轻松维护

  采用全新高密度服务器系统架构，维护操作布局在冷风区，更具人性化。

  - 服务器节点管理网口具备整机汇聚管理功能，采用各服务器节点iBMC协同管理方式，协同管理系统风扇、电源和整机资产，融合了刀片服务器和机架服务器的优点。
  - 服务器节点后置、后出线，方便维护。
  - 全模块化设计，硬盘、风扇模块、服务器节点、电源模块、OCP3.0网卡支持热插拔，更换方便，大幅提升运维效率。共享架构、高效节能各服务器节点共享系统电源和散热，提高电源模块转化效率，降低系统散热能耗，最大化实现系统节能。
- 共享架构、高效节能

  服务器节点共享系统电源和散热，提高电源模块转化效率，降低系统散热能耗，最大化实现系统节能。

  - 服务器节点共享4个电源模块，依据不同功耗配置实现电源模块1+1或2+2冗余，提高供电能力。
  - 服务器节点共享4个高风压对旋风扇模块，降低部署难度，提高风扇的利用效率，提升整机散热能力。
  - 采用动态能耗管理技术（DEMT），对能耗进行管理，降低系统能耗。
  - 无硬盘背板和系统背板，系统通风状况极致优化，支持最高规格处理器散热。
- 冗余设计、安全可靠

  高可靠的服务器系统架构，保障业务的长期稳定运行。

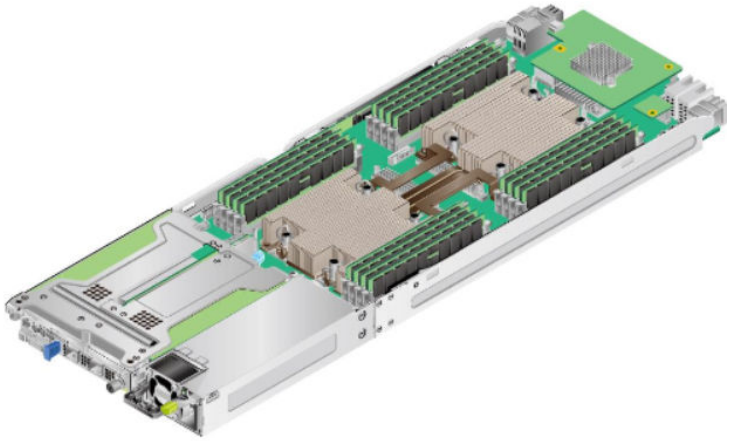
  - 支持4个电源模块，依据不同功耗配置实现电源模块1+1或2+2冗余。支持4个风扇模块，实现N+1冗余，提高系统整体可靠性。
  - 提供RAID控制扣卡和超级电容掉电数据保护，保障业务不中断、数据不丢失。

版权所有 © 超聚变数字技术有限公司

- 采用电信级别的器件和生产加工流程，产品更稳定、生命周期更长。

FusionServer XH321 V6 服务器节点

图 2-2 XH321 V6

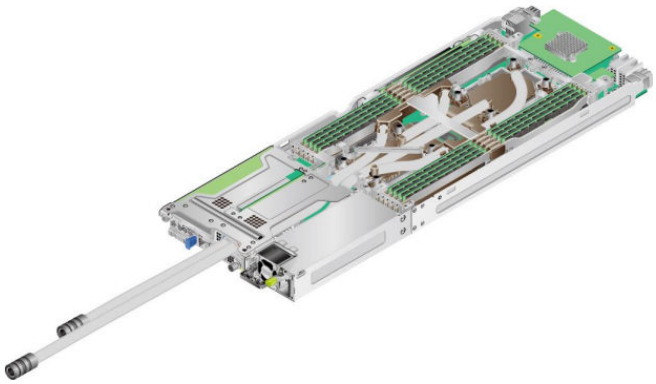


FusionServer XH321 V6服务器节点（以下简称XH321 V6）是X6000 V6服务器的节点，X6000 V6服务器机箱高度为2U，可以安装4个XH321 V6。XH321 V6为突破能源限制、提升系统存储密度而创新设计，具有更高的性能、超高的存储密度、易管理易维护等特点。

- XH321 V6支持2个M.2 SSD，最大支持16条DDR4内存，2张PCIe卡，1张OCP卡。
- XH321 V6具有强大的计算能力、大容量的本地存储能力和丰富的对外接口扩展能力。

FusionServer XH321C V6 液冷服务器节点

图 2-3 XH321C V6



FusionServer XH321C V6服务器节点（以下简称XH321C V6）是X6000 V6服务器的液冷节点，X6000 V6服务器机箱高度为2U，可以安装4个XH321C V6。

- XH321C V6为突破能源限制、提升系统存储密度而创新设计，具有更高的性能、超高的存储密度、易管理易维护等特点。

版权所有 © 超聚变数字技术有限公司

- XH321C V6支持2个SATA M.2 SSD，最大支持16条DDR4内存，1张PCIe卡，1张RAID控制扣卡，1张OCP卡。XH321C V6具有强大的计算能力、大容量的本地存储能力和丰富的对外接口扩展能力。
- XH321C V6为处理器、内存、VRD液冷配置节点。
  - 含处理器、内存、VRD液冷冷板。
  - 配套液冷机柜使用。
  - 配套液冷机柜使用时，PCIe卡、OCP卡及其他部件通过可选配的液冷门实现液冷。

2.2.1.2 机架服务器

- FusionServer 1288H V6

  一款1U2路机架服务器。1288H V6在1U空间里可配置2路英特尔®至强®可扩展处理器、32条DDR4内存及4*3.5"或10*2.5"的本地存储资源（可配置4/10个NVMe SSD硬盘）。

- FusionServer 2288H V6

  一款2U2路机架服务器。2288H V6可配置2路英特尔®至强®可扩展处理器、32条DDR4内存、14个PCIe扩展槽、支持大容量的本地存储资源。

- FusionServer 5288 V6

  一款4U2路机架服务器。在确保卓越计算性能的同时，提供灵活、超大容量的本地存储扩展能力，降低数据存储成本。5288 V6在4U空间里可配置2路英特尔®至强®可扩展处理器、32条DDR4内存及44*3.5"+4*NVMe SSD的本地存储资源。

- FusionServer 2488H V6

  一款2U4路机架服务器。相比于2台2U2路机架服务器，2488H V6可以带来更多的OPEX节省。2488H V6在2U空间内可配置4个第三代英特尔®至强®可扩展处理器，48条DDR4内存，11个PCIe扩展插槽以及最多25个2.5"的本地存储资源。

- FusionServer 1288H V7

  一款1U2路机架服务器，1/2个第四代英特尔®至强®可扩展处理器（Sapphire Rapids）、32DIMM DDR5内存、32*E1.S或20*E3.S SSD硬盘等多种本地大容量高性能存储配置，支持PCIE5.0、400G网卡、智能调频、整机相比业界最高节能8%。

- FusionServer 2288H V7

  一款2U2路机架服务器，1/2个第四代英特尔®至强®可扩展处理器（Sapphire Rapids），32DIMM DDR5内存，34*NVMe U.2 SSD硬盘等多种本地大容量高性能存储配置、PCIe5.0协议、2个OCP3.0网卡，支持4个双宽GPU加速卡或14个单宽GPU加速卡、智能调频、整机相比业界最高节能8%。

  在访存密集型场景中，亦可通过CXL扩展内存增加内存通道，提升内存带宽性能高达50%。

  以上服务器都集成DEMT智能功耗管理、FDM智能故障管理等专利技术，可选配FusionDirector全生命周期管理软件，能够有效降低运营成本、提升投资回报。

- FusionServer 2488H V7

  一款2U4路机架服务器。相比于2台2U2路机架服务器，2488H V7可以带来更多的OPEX节省。2488H V7在2U空间内可配置4个第三代英特尔®至强®可扩展处理器，64条DDR5内存，9个PCIe扩展插槽以及最多25个2.5"的本地存储资源。面向大容量内存需求场景，2488H V7最大可提供16TB内存容量，支持特殊应用的运算任务。

 版权所有 © 超聚变数字技术有限公司

- FusionServer 5288 V7

  一款4U2路机架服务器。在确保卓越计算性能的同时，提供灵活、超大容量的本地存储扩展能力，降低数据存储成本。5288 V7在4U空间里可配置2路英特尔®至强®可扩展处理器、32条DDR4内存及44*3.5"+4*NVMe SSD的本地存储资源。

- FusionServer 5885H V7

  一款4U4路机架服务器。在2U4路机架服务器的基础上，提供更大容量的本地存储扩展能力，降低数据存储成本。5885H V7在4U空间里可配置4路英特尔®至强®可扩展处理器、32条DDR4内存及44*3.5"+4*NVMe SSD的本地存储资源。

- FusionServer G8600 V7

  旗舰级GPU服务器，支持8GPU模组，具备12个PCIe标准卡和1个OCP卡的扩展能力，通过支持25块2.5寸硬盘和8个NVMe SSD盘获得超大容量或超高速存储。G8600 V7具有旗舰算力、极致能效、极致可靠、易于运维等优点，用于训练、科学计算、图片和数据分析等高性能场景。

- FusionServer G5500 V7

  一款4U2路GPU服务器，最多支持10个双宽GPU卡，具备4个PCIe标准卡和3个OCP卡的扩展能力，通过支持24个3.5寸硬盘或12个NVMe SSD盘获得超大容量或超高速存储。G5500 V7具有高性能、架构灵活、高可靠、易部署、易管理等优点，为AI训练、AI推理、HPC、图片视频分析和数据库等应用加速。

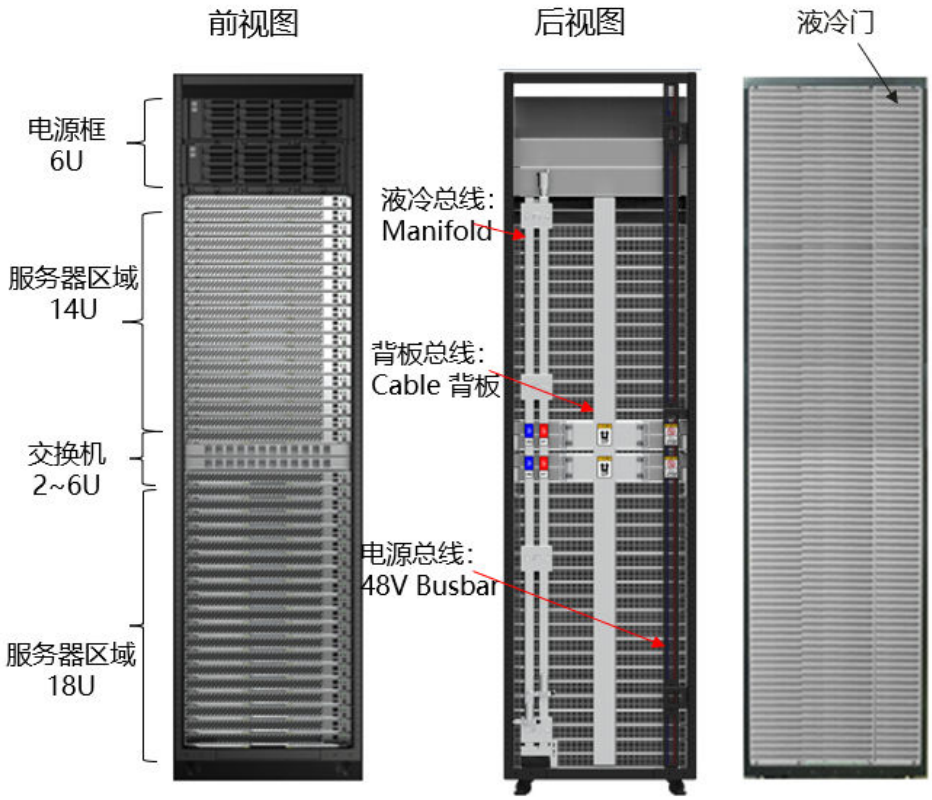
- FusionServer G5200 V7

  一款4U2路GPU服务器，可用于训练和推理工作负载，适用于语音、图像和视频分析等场景，具有强劲性能、可靠安全、高效节能、易管理、易部署等优点，面向大容量存储AI需求。

 版权所有 © 超聚变数字技术有限公司

2.2.1.3 整机柜服务器

图 2-4 FusionPod 液冷机柜



- **高密度**

  整机支持36个1U液冷服务器节点。

- **高性能**
  - 支持液冷环境，服务器节点可运行在更高频率，拥有超强计算能力。
  - 支持4台业务交换节点和1台带外管理交换节点配置，支持业务交换节点1+1备份。
- **高能效**
  - 支持市电直供。
  - 48V集中供电，最大97%电源转换效率。
  - 服务器节点支持液冷散热，最高支持45℃进水。
  - PUE：Power Usage Effectiveness低至1.2（不带液冷门）。
- **高可靠**
  - 支持2N配电。
  - 电源模块最大11+1配置，最大支持33kW供电。
  - 服务器节点风扇N+1配置，单风扇单转子失效时支持的最高工作温度下降5℃。
  - Cable背板无有源器件，失效率相较传统背板降低90%。

版权所有 © 超聚变数字技术有限公司

  - 服务器节点、机柜防漏液设计，支持柜级和节点级漏液监控，能够实现漏液自隔离。
  - 柜级部件满足至少10年的使用寿命。
- **一体化交付**

  支持除机柜门及其附件的整柜运输（含机柜、服务器节点、交换节点、管理模块、电源框、PDU、电源模块、传感器等）。
- **极简运维**
  - 服务器节点支持三总线盲插。
  - 服务器U位自动识别，极简运维。
  - 支持Redfish接口，第三方网管集成。
  - 支持FusionDirector等智能管理特性。

2.2.1.4 GPU 服务器

- FusionServer G8600 V7

新一代旗舰级GPU服务器，支持8GPU模组，具备12个PCIe标准卡和1个OCP卡的扩展能力，通过支持25块2.5寸硬盘和8个NVMe SSD盘获得超大容量或超高速存储。

- FusionServer G5500 V6

一款4U2路GPU服务器，支持多种规格GPU卡，具有计算性能卓越、配置灵活均衡、高效部署运维等特点，适合为AI推理、AI训练、HPC、视频分析和数据库等应用加速，针对深度学习的训练、推理和HPC等业务类型进行了优化，同时支持企业和公有云的部署。

2.2.2 存储

2.2.2.1 FusionOne HPC NFS

方案背景

在HPC场景，NFS典型用户为国内中小型长尾客户（如学校院级平台，小型车企、模具制造等），市场份额巨大。为应对市场需求，超聚变提供基于NVMe加速、NFS over RDMA协议技术性能优化的FusionOne HPC NFS存储方案，满足300TB及以下容量小型部署场景。

NFS 介绍

- NFS（Network File System），中文称为网络文件系统，NFS兼容主流Linux操作系统。它的主要功能是通过网络（一个局域网）让不同的主机系统之间可以共享文件或目录，可以通过挂载（mount）的方式将NFS服务器共享的数据目录挂载到NFS客户端本地，从客户端本地看，NFS服务器端共享目录就好像是客户端自己的磁盘分区或者目录一样，而实际上却是远端的NFS服务器的目录。
- NFS是第一个构建在IP协议之上的现代网络文件系统，在20世纪80年代，首先作为实验文件系统，由Sun Microsystems在内部开发完成，NFS协议归属与RFC标准，在随后演化为NFSv2，作为一个标准，由于NFS与其他客户端和服务器的互操作能力很好而发展迅速。
- 在CentOS 8中，支持NFSv3和NFSv4等版本，默认的NFS版本是4.2，支持访问控制列表（ACL），稀疏文件，空间保留，布局增强等能力。

 版权所有 © 超聚变数字技术有限公司

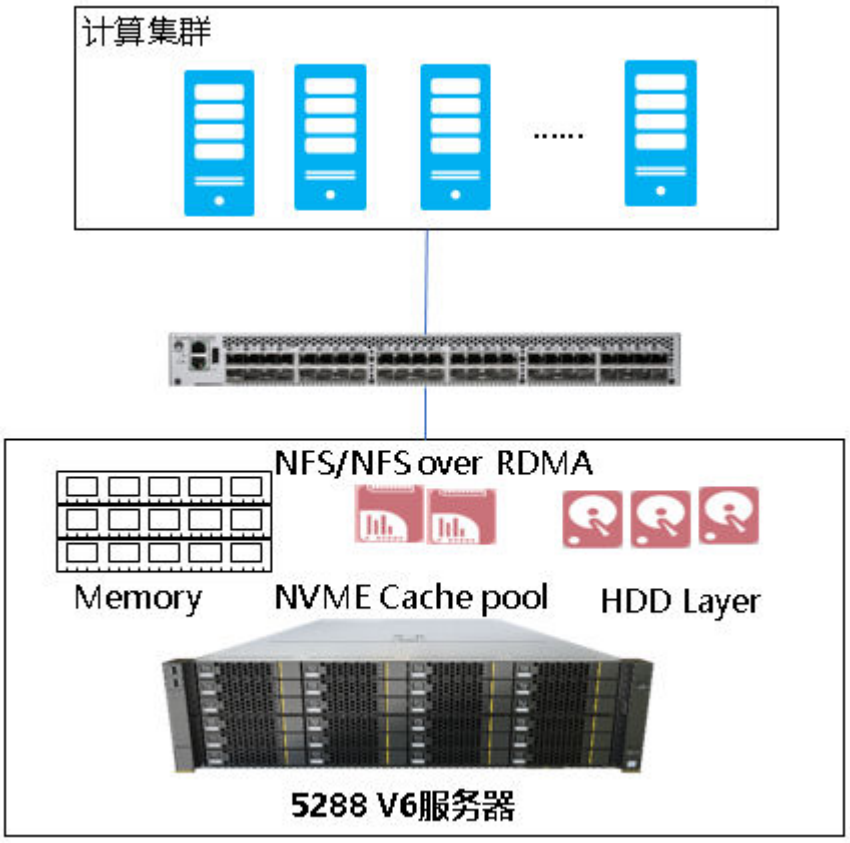
- NFSv4需要在IP网络上运行的传输控制协议（TCP）。NFSv3在早期CentOS版本中还可以采用用户数据报协议（UDP）。默认情况下，NFSv4在服务中禁用UDP。

NFS over RDMA（NFSoRDMA）介绍

NFS支持RDMA硬件实现NFS over RDMA，通过配置实现远程直接内存访问（RDMA）服务会在CentOS 8中运行。NFS over RDMA通过InfiniBand网络承载，实现内存到内存数据传输。RDMA可通过内存直接进行数据读写不受CPU干预，减轻了主机CPU的负担，减少了主机内存和I/O总线的争用。

功能架构

NFS over RDMA通过InfiniBand或RoCE网络承载，实现内存到内存数据传输。RDMA可通过内存直接进行数据读写不受CPU干预，减轻了主机CPU的负担，减少了主机内存和I/O总线的争用。



2.2.2.2 FusionOne HPC DFS

方案背景

FusionOne HPC DFS存储面向教育、制造及新型HPC场景（如AI，HPDA等），要求对称/非对称式架构，企业级特性和客户端兼容性场景。支持主流MPI-IO，兼容多类型客户端（Window，Linux和MacOS），提供快照、配额等存储特性。

版权所有 © 超聚变数字技术有限公司

FusionOne HPC DFS文件系统支持部署节点数量超过10000台，单一文件系统容量支持大于10PB，并行协议客户端数量支持大于100000台，单目录支持文件数量超过1亿，单套存储系统支持文件数量超过1000亿，容量支持EB级扩展。

- 针对海量数据存储应用而设计的大规模通用集群存储系统，采用通用硬件设备作为基本的构建单元，为应用提供全局统一的系统映像和完全POSIX兼容的API接口。
- 采用了自主研发的全系统规模数据高可用技术，彻底消除存储系统中的单点故障，结合特有的自动故障探测和快速故障恢复技术，确保用户的应用持续稳定地运行。
- 为用户提供PB级的存储容量，并根据用户应用发展的趋势，适时按需进行在线动态扩展。
- 世界领先的元数据服务器集群技术消除了现有存储系统中所存在的单目录下文件数量、小文件处理速度等限制，提供了近乎无限的文件存储数量和极高的文件检索速度，是业界唯一一款能够高效支持千万级大目录的存储系统（单目录下可轻松创建千万数量级的文件，并能对文件进行高速随机检索）。

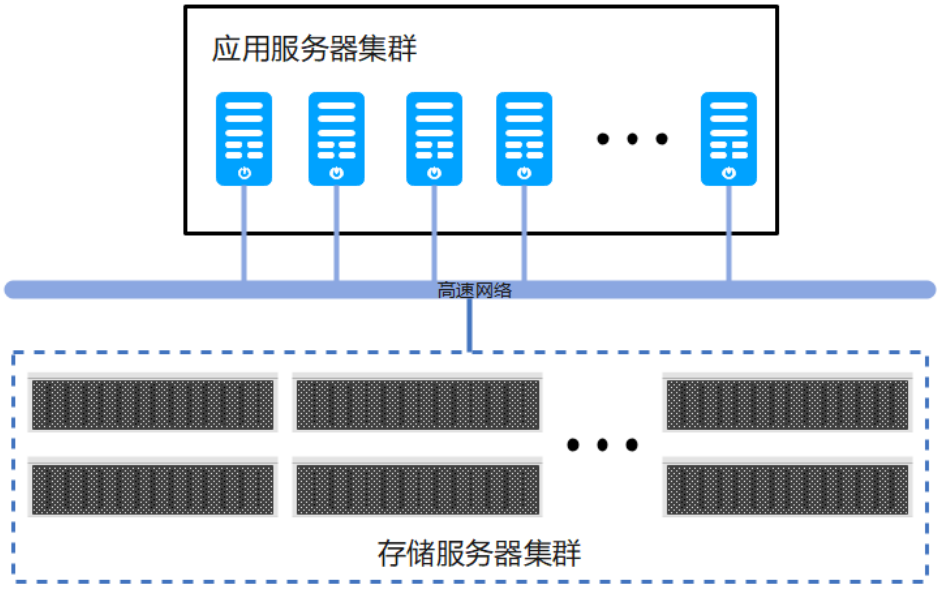
功能架构

FusionOne HPC DFS存储主要由应用服务器集群、元数据服务器集群和存储服务器集群组成，采用并行存储体系架构，元数据服务器与存储服务器采用本地磁盘作为数据存储构建统一分布式存储池，无需外挂磁盘阵列或JBOD盘框。

FusionOne HPC DFS存储架构灵活，支持对称架构和非对称架构部署。在对称式架构模式下，无单独元数据节点集群，元数据分布在各个存储节点集群上。在非对称架构模式下，元数据节点集群与存储节点集群物理分离，所有元数据节点和存储节点都同时在线提供服务，元数据集群在多活模式下工作。

- **对称式架构**

  FusionOne HPC DFS存储对称式架构，元数据和数据混合部署，无单独元数据节点（元数据分布在存储节点上），可提供轻量型和容量型2种方案。重点针对海量带宽场景，节省元数据服务器，降低方案成本。



  - 应用服务器集群

版权所有 © 超聚变数字技术有限公司

应用服务器集群（即FusionOne Center for HPC计算节点）上需要安装FusionOne HPC DFS存储客户端软件。运行客户应用程序的服务器集群，进行特定的计算、信息处理及业务服务等。

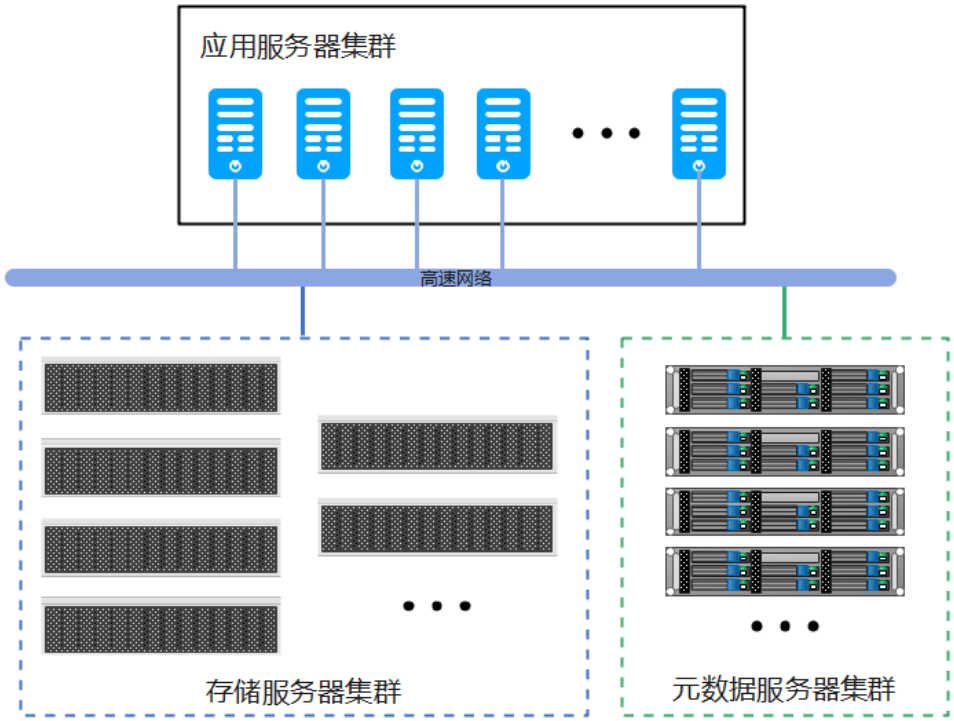
- 存储服务器集群

  存储用户的实际数据的服务器集群，是整个FusionOne HPC DFS存储系统的存储资源提供者。提供元数据和数据存储服务，管理整个存储系统的命名空间，对外提供单一的文件系统，其上安装集群存储系统管理工具，通过在Web页面上的简单操作即可达到对集群文件系统进行配置、管理以及监控的目的，以保证集群文件系统的正常运行。

- **非对称式架构**

  FusionOne HPC DFS非对称式架构，元数据服务器集群和存储服务器集群分离部署，重点针对海量小文件场景，元数据集中在单独元数据服务器节点，多活方式提供元数据服务，提升元数据访问效率。

  图 2-5 文件系统非对称架构



  - 应用服务器集群

    应用服务器集群（即FusionOne Center for HPC计算节点）上需要安装FusionOne HPC DFS存储客户端软件。运行客户应用程序的服务器集群，进行特定的计算、信息处理及业务服务等。

  - 元数据服务器集群

    管理文件系统的元数据的服务器集群（包括文件目录树组织、属性维护、文件操作日志记录、授权访问等），管理整个存储系统的命名空间，对外提供单一的文件系统。其上安装集群存储系统管理工具，通过在Web页面上的简单操作即可达到对集群文件系统进行配置、管理以及监控的目的，以保证集群文件系统的正常运行。

  - 存储服务器集群

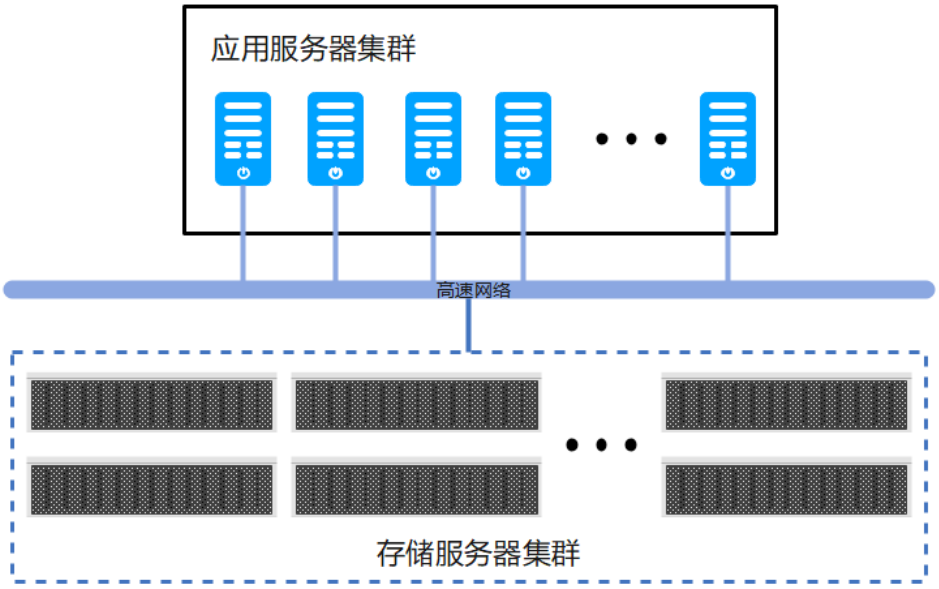
版权所有 © 超聚变数字技术有限公司

存储用户的实际数据的服务器集群，是整个FusionOne HPC DFS存储系统的存储资源提供者。

2.2.2.2.1 功能架构

?.1. 对称式架构

FusionOne HPC DFS存储对称式架构，元数据和数据混合部署，无单独元数据节点（元数据分布在存储节点上），可提供轻量型和容量型2种方案。重点针对海量带宽场景，节省元数据服务器，降低方案成本。



- 应用服务器集群

  应用服务器集群（即FusionOne Center for HPC计算节点）上需要安装FusionOne HPC DFS存储客户端软件。运行客户应用程序的服务器集群，进行特定的计算、信息处理及业务服务等。

- 存储服务器集群

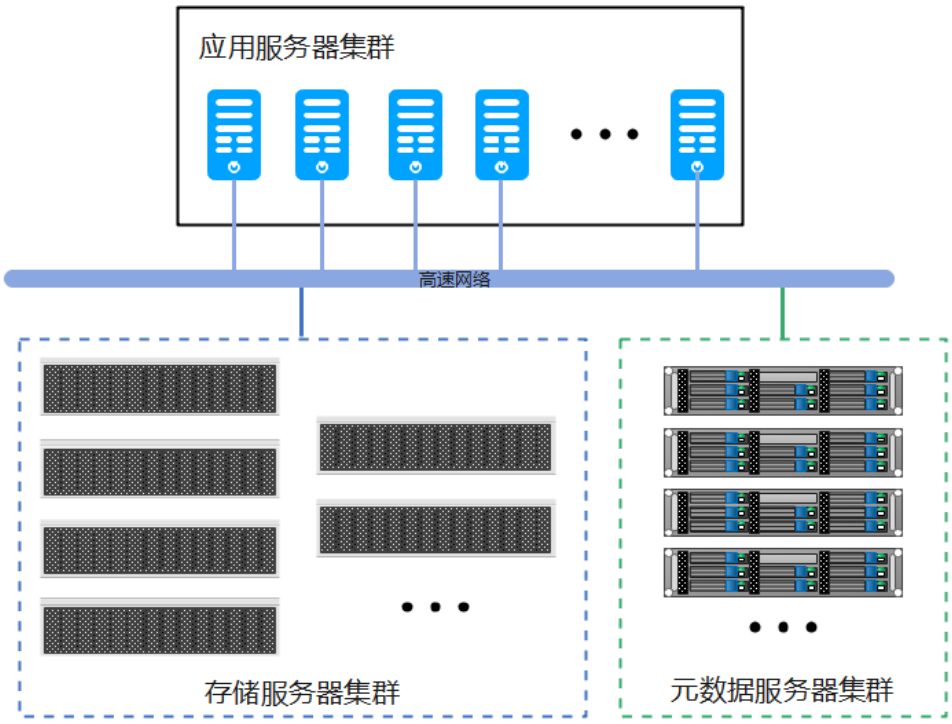
  存储用户的实际数据的服务器集群，是整个FusionOne HPC DFS存储系统的存储资源提供者。提供元数据和数据存储服务，管理整个存储系统的命名空间，对外提供单一的文件系统，其上安装集群存储系统管理工具，通过在Web页面上的简单操作即可达到对集群文件系统进行配置、管理以及监控的目的，以保证集群文件系统的正常运行。

?.2. 非对称式架构

FusionOne HPC DFS非对称式架构，元数据服务器集群和存储服务器集群分离部署，重点针对海量小文件场景，元数据集中在单独元数据服务器节点，多活方式提供元数据服务，提升元数据访问效率。

版权所有 © 超聚变数字技术有限公司

图 2-6 文件系统非对称架构



- 应用服务器集群

  应用服务器集群（即FusionOne Center for HPC计算节点）上需要安装FusionOne HPC DFS存储客户端软件。运行客户应用程序的服务器集群，进行特定的计算、信息处理及业务服务等。
- 元数据服务器集群

  管理文件系统的元数据的服务器集群（包括文件目录树组织、属性维护、文件操作日志记录、授权访问等），管理整个存储系统的命名空间，对外提供单一的文件系统。其上安装集群存储系统管理工具，通过在Web页面上的简单操作即可达到对集群文件系统进行配置、管理以及监控的目的，以保证集群文件系统的正常运行。
- 存储服务器集群

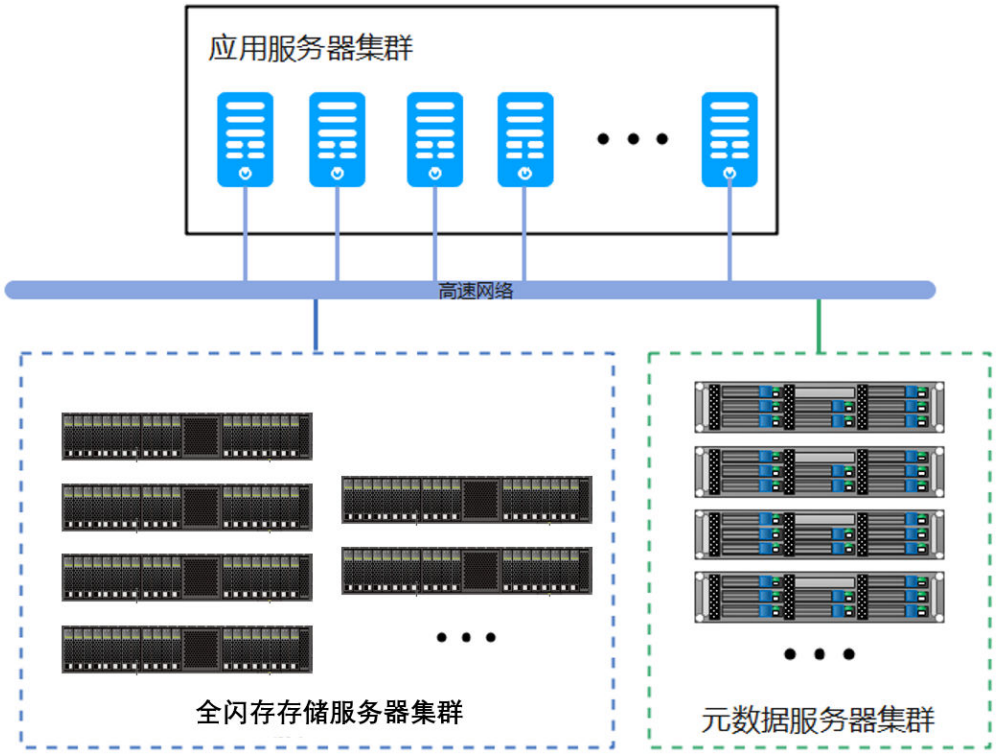
  存储用户的实际数据的服务器集群，是整个FusionOne HPC DFS存储系统的存储资源提供者。

?.3. 全闪架构

FusionOne HPC DFS提供全闪存架构方案，全闪存方案中节点配置NVME SSD盘，可提供较高的聚合带宽、IOPS以及较低的时延。

全闪存方案推荐采用非对称架构组网，提供单独的元数据服务器，其部署方案如下：

版权所有 © 超聚变数字技术有限公司



- 应用服务器集群

  应用服务器集群（即FusionOne Center for HPC计算节点）上需要安装FusionOne HPC DFS存储客户端软件。运行客户应用程序的服务器集群，进行特定的计算、信息处理及业务服务等。

- 元数据服务器集群

  管理文件系统的元数据的服务器集群（包括文件目录树组织、属性维护、文件操作日志记录、授权访问等），管理整个存储系统的命名空间，对外提供单一的文件系统。其上安装集群存储系统管理工具，通过在Web页面上的简单操作即可达到对集群文件系统进行配置、管理以及监控的目的，以保证集群文件系统的正常运行。

- 全闪存存储服务器集群

  存储用户的实际数据的服务器集群，是整个FusionOne HPC DFS存储系统的存储资源提供者，采用全闪存硬件类型以及配置NVME SSD盘。

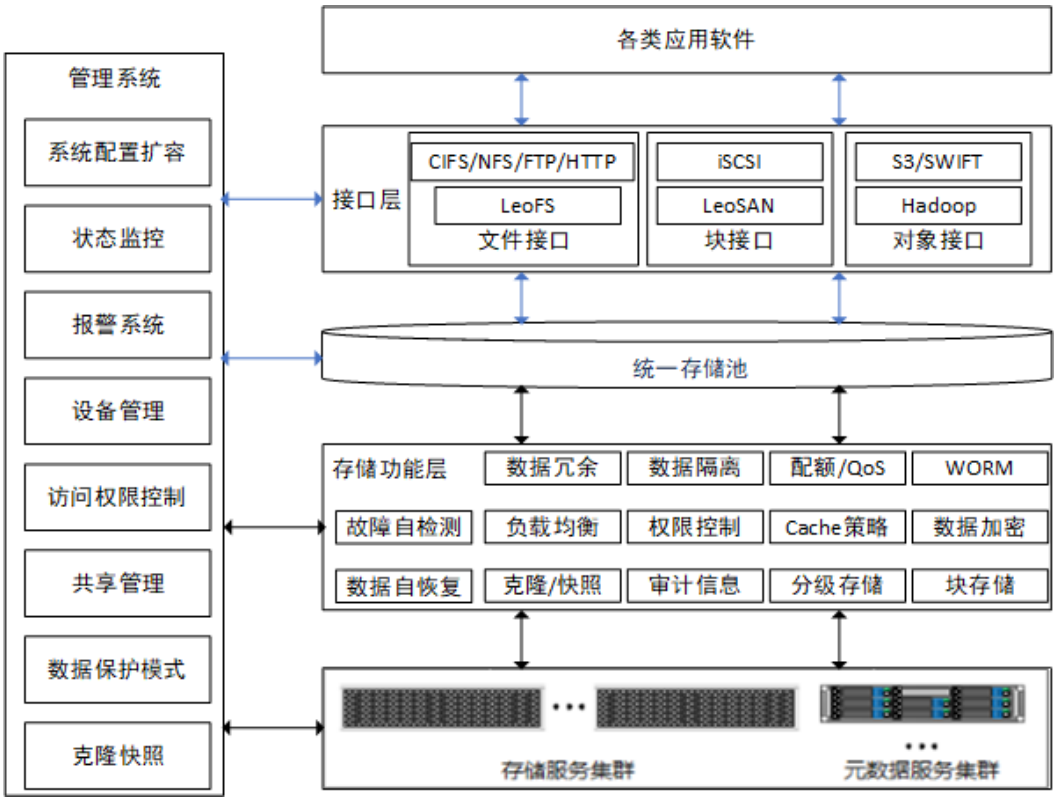
  说明

  全闪存架构是相对HDD硬盘而言，其本身也分对称架构和非对称架构，为了获得更高的性能，全闪存通常采用非对称架构。

2.2.2.2.2 软件架构

FusionOne HPC DFS存储软件可分为LeoFS协议层、接口层和管理系统。

版权所有 © 超聚变数字技术有限公司

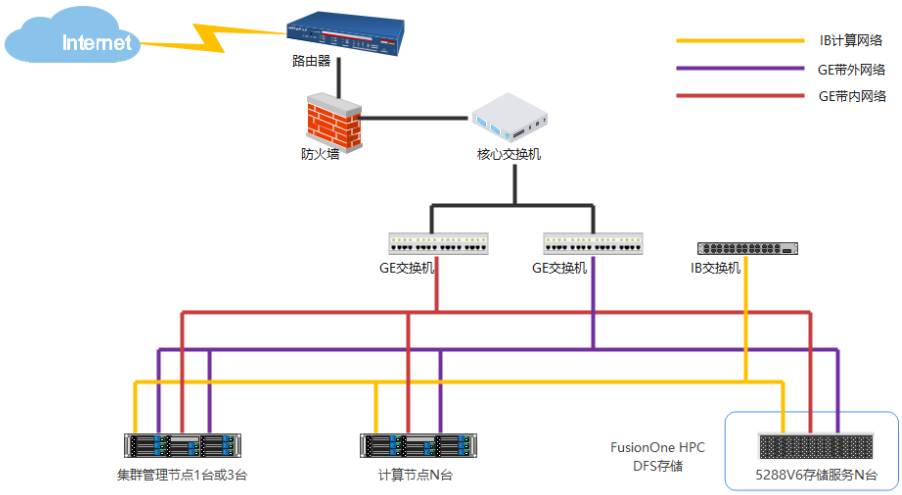


- LeoFS协议层是整个存储系统的核心，是保证数据安全、高性能以及弹性扩展的基础 。该层将底层的硬件系统资源进行整合，为上层应用提供了一个统一的共享存储池，并实现了纠删码和副本等数据冗余保护、负载均衡、克隆快照、块存储、对象存储等基本存储功能。
- 接口层实现了统一共享存储到各类应用的接口转换，以支持各种应用的数据访问。其中LeoFS和LeoSAN基于系统专有通信协议，具有出色的IO性能。对于公有访问协议用户，系统提供了相关协议的集群服务，较传统接口大大提升了可靠性和访问性能。
- 管理系统提供了存储系统、服务器及网络状态监控、各类报警信息提示，存储系统相关的所有管理、功能设置等操作均可通过管理界面完成。管理系统还是重要的审计渠道，管理系统界面中显示所有管理操作日志均记录，方便审计。管理系统支持三员模式，各角色权限互斥且各自独立，为管理安全提供了保障。

2.2.2.2.3 组网规划

版权所有 © 超聚变数字技术有限公司

?.1. 对称式架构组网规划



存储服务器、应用服务器都必须在同一个局域网内。避免将各服务器连接在不同的交换机上，存储网络支持100GE以太网络、IB网络；方案推荐采用EDR IB、HDR IB、NDR IB网络。方案组网分为计算网络、带外管理网络和业务管理网络。

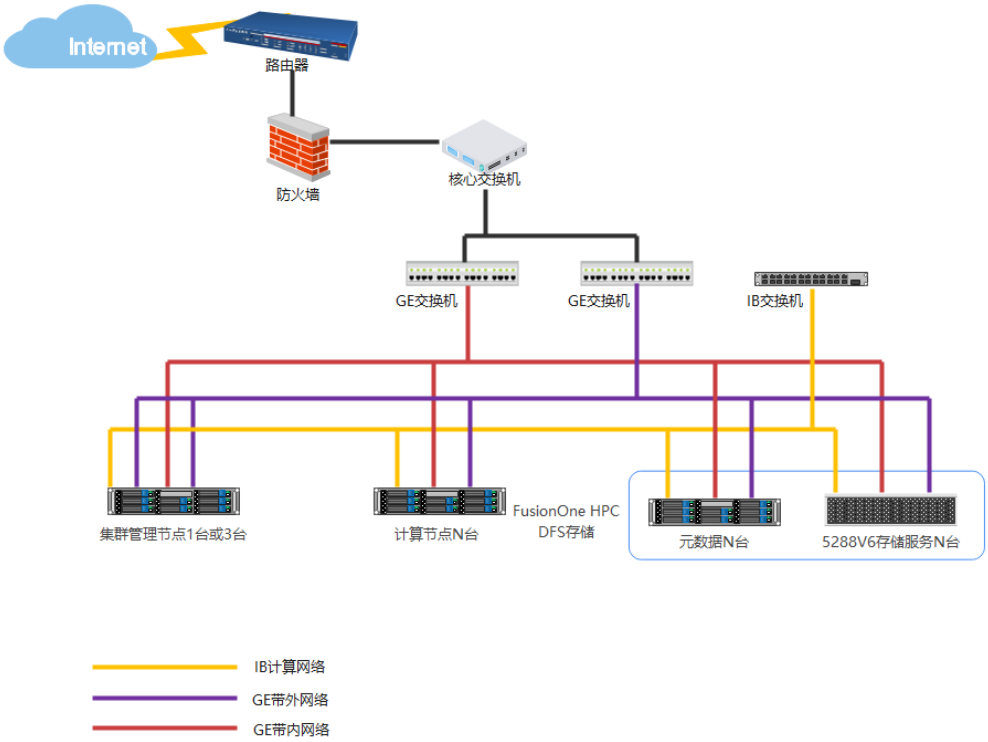
- 计算网络
  存储服务器请选择以下一种配置：
  - 配置1张或2张EDR IB网卡。
  - 配置1张或2张HDR IB网卡。
  - 配置1张NDR IB网卡。
  - 配置1张或2张100GE网卡。

    说明

    - HDR IB网卡支持100G或200G。
    - EDR IB网卡支持单口或双口。
    - 不支持同时配置2张双口EDR IB网卡。
- 带外管理网络
  存储服务器：1*IPMI口（远程上下电）。
- 业务管理网络
  存储服务器：配置2*1GE口（管理网口）。

版权所有 © 超聚变数字技术有限公司

?.2. 非对称式架构组网规划



元数据服务器、存储服务器、应用服务器都必须在同一个局域网内。避免将各服务器连接在不同的交换机上，存储网络支持以太网络、IB网络；方案推荐采用EDR IB、HDR IB、NDR IB网络。方案组网分为计算网络、带外管理网络和业务管理网络。

- 计算网络
  - 元数据服务器请选择以下一种配置：
    - 配置1张或2张EDR IB网卡。
    - 配置1张或2张HDR IB网卡。
    - 配置1张NDR IB网卡。
    - 配置1张或2张100GE网卡。

      说明

      - HDR IB网卡支持100G或200G。
      - EDR IB网卡支持单口或双口。
      - 不支持同时配置2张双口EDR IB网卡。
  - 存储服务器请选择以下一种配置：
    - 配置1张或2张EDR IB网卡。
    - 配置1张或2张HDR IB网卡。
    - 配置1张NDR IB网卡。
    - 配置1张或2张100GE网卡。

版权所有 © 超聚变数字技术有限公司

说明

- HDR IB网卡支持100G或200G。
- EDR IB网卡支持单口或双口。
- 不支持同时配置2张双口EDR IB网卡。

- 带外管理网络
  - 元数据服务器：1*IPMI口（远程上下电）。
  - 存储服务器：1*IPMI口（远程上下电）。
- 业务管理网络
  - 元数据服务器：配置2*1GE口（管理网口）。
  - 存储服务器：配置2*1GE口（管理网口）。

2.2.2.2.4 关键特性

?.1. 数据冗余策略

FusionOne HPC DFS存储系统提供了副本和纠删码（EC）数据保护策略，最高可允许4块磁盘或节点同时损坏而不丢失数据。在允许一块磁盘或一个节点损坏而不丢失数据的冗余配置下，磁盘使用率最高可达94%（N=16，M=1）。冗余模式的设置对象为目录，用户可根据数据的安全级别要求，对各级目录和子目录设置不同的冗余模式，目录深度不限。一个目录下的文件及子目录自动继承父目录的冗余模式，设置根目录的冗余模式即对全系统设置了冗余模式。更改父目录的冗余模式，其下子目录的冗余模式也随之更改。更改目录冗余模式只会影响新创建的文件，不会改变目录下已存在的文件的冗余模式，也不会影响设置过冗余模式的子目录。

副本模式

FusionOne HPC DFS存储系统提供了数据副本冗余模式。用户可根据数据的重要程度，以目录为单位设置N（N可为1、2、3、4）个副本的数据冗余保护模式，N副本模式下允许系统中同时有N-1个节点损坏而不丢失数据。

纠删码模式

纠删码级别有：N+1、N+2、N+3和N+4，N可设置为2、4、8、16，代表一个数据块将被分为N片，M代表通过对N个数据片进行计算，生成的校验片数量，N+M个数据分片将被分别存储在不同的节点或磁盘内，在任意M个节点或磁盘出现故障的情况下，仍然不丢失数据。

纠删码冗余配比如**表2-1**所示。

 版权所有 © 超聚变数字技术有限公司

表 2-1 纠删码冗余配比

冗余配比		空间利用率	最少存储节点数量	支持最大同时故障存储节点数	说明	配置命令示例
N+1	2+1	66%	3	1	2个数据片+1个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 2 1模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 2 1
	4+1	80%	5	1	4个数据片+1个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 4 1模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 4 1
	8+1	88%	9	1	8个数据片+1个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 8 1模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 8 1
	16+1	94%	17	1	16个数据片+1个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 16 1模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 16 1
N+2	2+2	50%	4	2	2个数据片+2个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 2 2模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 2 2
	4+2	66%	6	2	4个数据片+2个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 4 2模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 4 2
	8+2	80%	10	2	8个数据片+2个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 8 2模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 8 2

版权所有 © 超聚变数字技术有限公司

冗余配比		空间利用率	最少存储节点数量	支持最大同时故障存储节点数	说明	配置命令示例
	16+2	88%	18	2	16个数据片+2个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 16 2模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 16 2
N+3	4+3	57%	7	3	4个数据片+3个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 4 3模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 4 3
	8+3	72%	11	3	8个数据片+3个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 8 3模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 8 3
	16+3	84%	19	3	16个数据片+3个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 16 3模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 16 3
N+4	4+4	50%	8	4	4个数据片+4个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 4 4模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 4 4
	8+4	66%	12	4	8个数据片+4个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 8 4模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 8 4
	16+4	80%	20	4	16个数据片+4个校验片	1. 使用root用户登录应用节点。 2. 执行如下命令配置纠删码xd2 16 4模式。 /LeoCluster/bin/leofslayout5 -c *应用节点挂载路径* xd2 16 4

?.2. 快速数据重构

存储系统内置故障自动探测机制和数据自动恢复机制。当系统内任何磁盘、存储节点或网络发生故障，系统可以感知故障点，并自动完成隔离，不影响前端应用，同时在

版权所有 © 超聚变数字技术有限公司

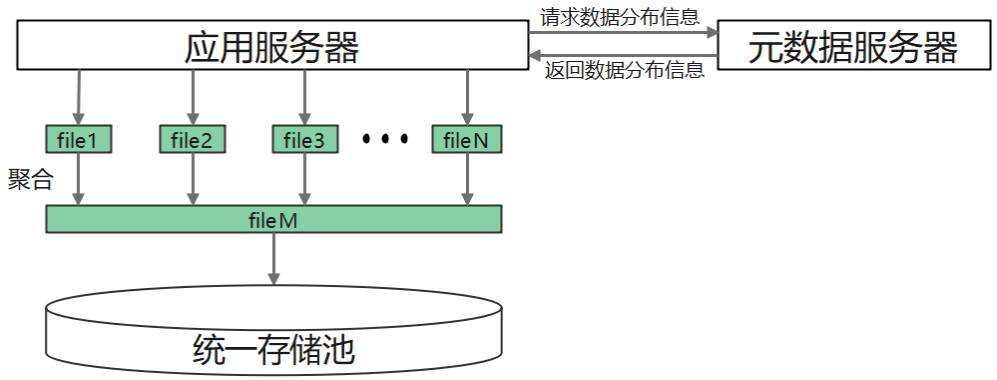
冗余策略允许的范围内自行启动数据恢复过程，把失效磁盘或存储节点上的数据恢复到系统内其他正常的设备上，以保障数据的冗余要求，整个节点宕机，业务不中断。

为了提高数据恢复的效率，存储系统仅针对有效数据进行恢复，而不像传统存储系统那样针对整个磁盘进行数据恢复，可有效减少恢复数据量。在数据存储过程中，不同的数据切片或副本在系统全局范围内均匀分布在每一个存储节点和每一块磁盘上，存储池所有节点参与数据重构流程，多块盘向多块盘重构，1T数据重建不超过30分钟。

?.3. 小文件聚合

在海量小文件场景，小文件的大小往往小于数据传输的最小单元，无论文件的有效数据量多大，都会以固定大小（数据传输的最小单元）传输，使得有效数据占比低，导致内存、带宽的无谓消耗。当用户读取大量小文件时，应用服务器频繁地与元数据服务器及存储服务器进行交互，导致数据访问产生较大延迟。

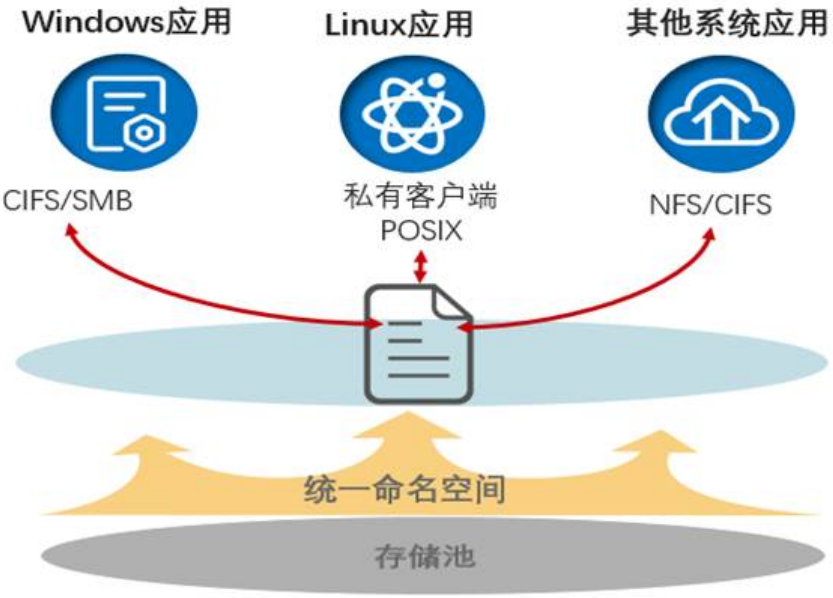
针对小文件存取的效率问题，FusionOne HPC DFS存储系统提供小文件聚合传输和存储机制。将若干个小文件合并成一个数据块进行传输，减轻读写数据压力，避免带宽消耗，有效降低磁盘操作次数及数据访问延迟。在顺序读取大量小文件时，内存命中率高达96%以上，有效缩短了IO路径。在修改和删除操作时，仍以文件为操作单元，仅更新修改的部分，删除某个文件后即释放其存储空间。



?.4. 文件多协议互通

存储系统共享统一存储池，通过单一命名空间对外全局提供文件存储服务。存储系统支持HDFS、NFS、CIFS/SMB标准文件存储协议和POSIX/专有客户端协议，支持CSI容器存储，各种文件协议支持融合互通访问。一份文件数据多协议访问，数据零迁移拷贝，避免多份数据重复存储，提高访问效率。

版权所有 © 超聚变数字技术有限公司

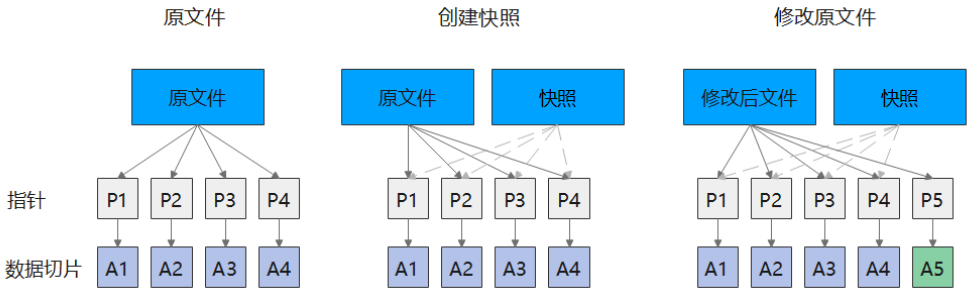


?.5. 快照和克隆

快照

存储系统提供针对目录、文件的快照功能，支持对文件系统内的任意目录、文件创建快照（文件系统根目录除外），可在秒级完成快照，方便用户对重要数据进行备份且不影响业务。当存储设备发生故障或者文件损坏时，可以快速将数据恢复至某个可用时间点的状态。

存储系统通过ROW（Redirect On Write）写时重定向技术实现快照。快照指针基于文件切片创建，在创建快照时不进行数据拷贝，仅仅创建一组指向原文件各数据片的指针，空间占用小。当原文件内容修改时才将相关的数据片进行复制修改，之后存于新的位置，并创建相应指针。单系统及单目录均支持无限数量的快照存储。



克隆

克隆即内部拷贝，系统的内部拷贝基于ROW方式，实现副本索引的快速创建，可以秒级完成TB级别的数据拷贝。拷贝后，如果原数据和拷贝的数据均无修改，则拷贝的数据不会占用任何存储空间，仅增加一组索引信息。当原数据或拷贝数据有修改时，会进行相关数据切片的复制操作，节约了大量存储空间。

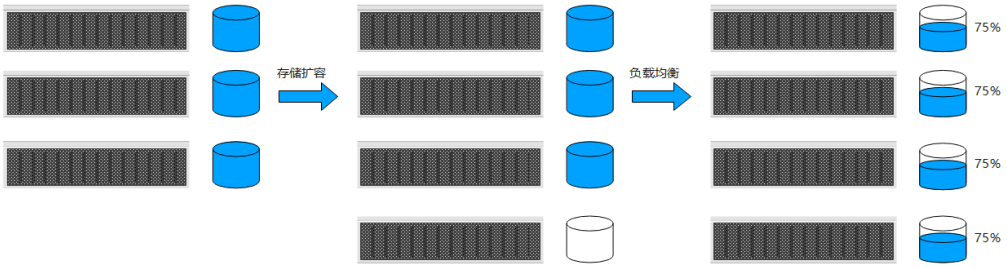
版权所有 © 超聚变数字技术有限公司

?.6. 负载均衡

负载均衡是保证系统性能和空间使用率达到理论值的重要前提，当系统无负载均衡功能时，会导致部分存储设备尚有较大可用空间，部分存储设备已满，数据切片无法写入的情况，同时系统写带宽也会随着可提供服务的存储设备数量的减少而下降。

本存储系统元数据服务接收到应用服务器的数据写入请求后，在全局范围内选择数据量最小的存储设备分配空间，确保在全局范围内数据负载均衡，即每个存储设备上的数据量均衡。

在某些特殊情况下会有负载失衡的情形，例如系统新扩容存储服务器或存储设备后，仅依赖前端数据写入时的元数据服务器进行存储空间选择，很难使系统在短时间内回归全局负载均衡。因此，需要额外的负载均衡机制，负载均衡的原理是元数据服务根据系统中所有磁盘上的数据分布情况统计出数据量最大的存储设备，在满足数据的冗余要求的前提下将其上数据搬移至数据量最小的存储设备上，并更新相关文件的位置信息，最终使得全系统存储设备上数据均衡存储节点间数据量差别小于3%。负载均衡任务会占用一定的元数据服务资源，建议该任务在系统相对空闲时执行。



如上图所示，当系统扩容一台存储服务器后，虽然系统总容量增加，但系统的写性能并没有因为扩容而线形提升；在某些需要同时向多个节点写入数据的冗余模式下，实际可用容量仍然为零，扩容失去意义。因此需要系统管理员手动启动负载均衡操作，使数据均匀分布于当前系统中的所有存储设备上。

?.7. 配额

配额是一种限制或者统计所使用的存储空间和文件数量的技术。为了避免个别用户对存储资源的过度使用，管理员通过配置合适的配额策略，控制目录、用户、用户组设置存储空间或文件数量使用上限，并在资源使用量达到配置的阈值时发送告警通知。

配额支持嵌套设置，默认子目录继承父目录的配额限制设置。配额对所有访问接口有效，并可在线设置、在线更改和取消，且即时生效。

用户和用户组可使用存储系统内部的用户模式，也可导入当前场景中已有的AD域用户和LDAP域用户进行配额设置和管理。所有配额相关设置、信息查询均可在基于Web的管理界面上进行。

目录配额的设置对象为目录，目录的配额设置记录在该目录的元数据信息中。当元数据服务器收到数据写入请求时，元数据服务器依次查询该目录的各种配额设置（目录、用户、用户组配额），分析出本次写操作受影响的配额配置项，并逐项进行配额门限值比较，根据门限值允许或禁止此次文件操作。如果受影响的所有配额控制项允许，写请求才能继续进行，同时更新配额的统计值，一旦写操作尝试超过配额限制，则阻止本次写入。

?.8. Cache 策略

DFS支持SSD Cache技术，通过SSD Cache功能可以提升存储部分场景下的性能，达到性能加速的目的。DFS每个节点支持两块NVME SSD硬盘，通过软RAID构建

版权所有 © 超聚变数字技术有限公司

RAID10后将逻辑盘提供给DFS作为高速缓存盘。SSD Cache方案可以将热点数据驻留在SSD盘中，提升热点数据的性能。

SSD Cache有三种设备：

- BCache设备：驱动虚拟出来的一个设备节点，应用程序可直接读写访问BCache设备。
- Cache设备：存储热数据的高速设备。
- Backing设备：后端设备，传统机械硬盘或低速设备。

SSD Cache通过软件驱动能够动态的调整，将热数据存储在Cache设备上，温/冷数据存储在Backing设备上，通过BCache设备的接口访问时，可以判断哪些是热数据，哪些是温/冷数据。如果热数据已经存储在Cache设备里，从Cache设备读取数据，将直接返回给应用程序，这样应用程序看到的IO性能就是SSD的IO性能，不是传统机械硬盘的IO性能。因此，当应用的请求多数为热数据时，用户使用的就是低延迟和高吞吐的高速设备，提升了整体IO性能。

目前SSD Cache支持透写模式：

- Writethrough：写通模式，在写入时，既写入到Cache设备里，又写入到Backing设备里，因此当读取时，可直接从Cache设备中读取。该模式的优点是在设备掉电时，能尽可能降低丢失Cache设备里的脏数据的概率。
- None：与Cache无关，直接将后端设备当成传统的磁盘或慢速SSD使用。

?.9. 文件流转跟踪

存储系统支持准确定位某个文件的数据写入客户端ID、数据删除客户端ID、删除时间等详细记录。传统文件下系统只能记录存储系统上的操作，而DFS存储系统可以记录从客户端到存储系统端到端的高危操作，协助系统管理员定位定则。

?.10. 操作日志

存储系统提供完备的日志功能，日志记录了系统管理员的管理操作以及用户对文件的操作，日志信息包含文件名及所执行的具体操作。

管理操作日志是指管理员登录管理系统后对存储系统进行的管理操作，该日志由管理系统采集并记录在数据库中，内容包括管理用户名、操作时间、操作详情、操作结果。管理操作日志只有审计用户可以查看并进行删除操作，其他用户无法访问。

用户对文件及目录的操作日志由元数据服务进行记录并保存在元数据服务日志中。默认情况下该操作日志功能关闭，用户可根据需要，通过修改配置文件打开创建、删除、修改、重命名文件及目录中的一个或多个操作的日志开关。

?.11. 访问权限控制

存储系统在原有基于操作系统用户的访问控制之外，提供基于系统内部客户端（应用服务器）和用户的精细化访问控制机制，且操作系统用户、管理用户以及其它特权用户均无法逾越。

管理员可配置用户和客户端对目录的读（R）、写（W）、删除（D）、重命名（Ren）、创建硬链接（Ln）、显示目录内容列表（Lst）以及追加写权限（ApdW），以此满足用户的精细化权限访问控制的需求，实现数据隔离，保证数据的访问安全，防止系统入侵或误操作导致的数据损失。

 版权所有 © 超聚变数字技术有限公司

说明

精细化权限访问控制功能默认关闭，所有客户端及用户均可访问存储系统的任何目录。

当访问控制开关打开时，所有用户及客户端按照配置好的访问权限对存储系统目录进行访问，没有配置访问权限的客户端和用户则不能对存储系统进行访问。因此在打开权限开关前，须先配置客户端和用户对相应目录的访问权限，否则会影响正在运行的数据服务。

表 2-2 存储系统客户端和用户访问权限对应值

权限值（十进制）	64	32	16	8	4	2	1
权限值（二进制）	2^6	2^5	2^4	2^3	2^2	2^1	2^0
访问权限	ApdW	Lst	Ln	Ren	D	W	R

如表2-2所示，目录的访问权限值用字节，即8位2进制数表示，从低位到高位分别代表读、写、删除、重命名、创建硬链接、显示目录内容列表以及追加写权限，共7位，有权限用1表示，没有权限用0表示，总127种权限组合。将此二进制权限值转化为十进制数，以及相应的客户端或用户信息保存于该目录的元数据中。当某个客户端去访问文件时，元数据会将该文件所有的权限信息包括操作系统权限和存储系统私有访问控制权限返回给客户端。客户端获取权限信息后首先校验操作系统权限，确认操作系统权限允许后再进行存储系统访问权限校验，若不允许则拒绝该访问请求，若允许则会提供后续的访问服务。

?.12. 扩容

DFS存储对称式和非对称式两个架构都支持节点在线扩容和离线扩容，根据存储系统实际情况和企业运行业务的运行需求选择相应的扩容硬件。DFS支持对元数据节点、数据节点分别扩容或者同时扩容，具体支持能力如下表：

存储架构	**扩容方式**	**扩容的原则**	**备注**
非对称式架构	元数据节点扩容 存储节点扩容	• 元数据节点采用x86/arm架构服务器扩展，节点配置（扩展元数据盘，系统盘） • 存储节点采用x86/arm架构服务器扩展，节点满配（36块HDD），磁盘容量、个数、类型与现网配置保持一致 • 扩容硬件不能低于现网硬件配置（CPU，内存，系统盘/元数据盘，RAID卡等）	支持基于节点满配扩容，可扩展元数据（成对扩展，最低扩展1对节点）和存储节点（最低扩展1节点），非对称式架构只能扩展为非对称式架构。

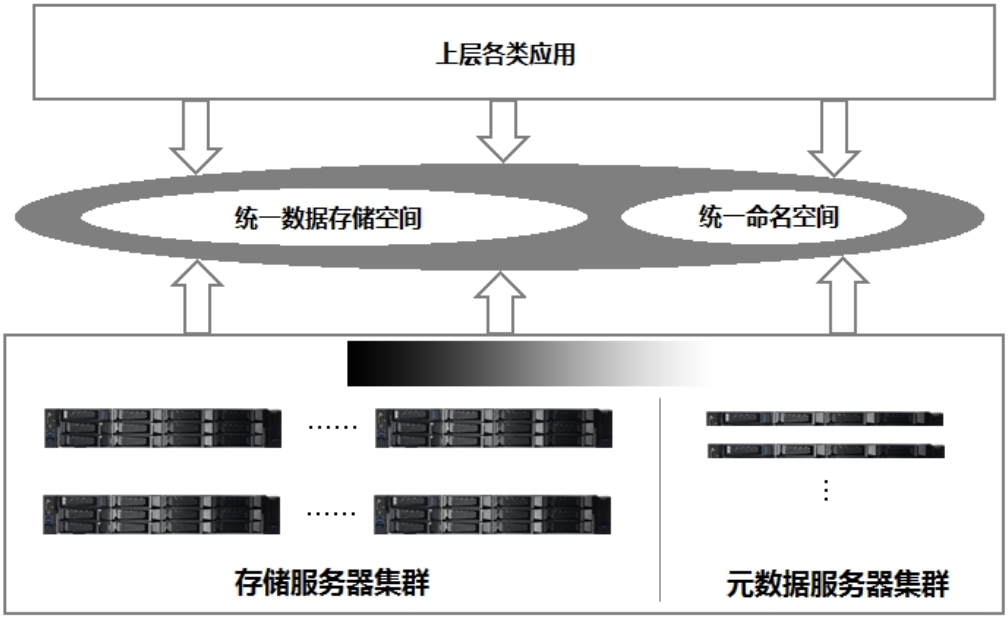
版权所有 © 超聚变数字技术有限公司

存储架构	扩容方式	扩容的原则	备注
	元数据节点和存储节点同时扩容	• 网络组网/网卡与现网保持一致 • 存储版本信息/DFS版本与现网保持一致 • 新扩容部分容量的SNS应与现网剩余服务时间保持一致 • 支持同一个存储池扩容（EC不变），暂不支持不同存储池扩容 • 新扩容硬件维保年限及级别建议与现网硬件剩余服务时间保持一致	
对称式架构	混合节点扩容 存储节点扩容 混合节点和存储节点同时扩容	• 存储节点采用容量型/轻量型扩展存储，节点满配（36/12块HDD），磁盘容量、个数、类型与现网配置保持一致； • 混合节点采用容量型/2288H V6轻量型扩展存储、元数据，节点满配（36/12块HDD），磁盘容量、个数与现网配置保持一致，元数据盘； • 启用现网原容量型/轻量型存储节点的元数据盘（成对）启用扩展元数据； • 扩容硬件不能低于现网硬件配置（CPU，内存，系统盘/元数据盘，RAID卡等） • 网络组网/网卡与现网保持一致 • 存储版本信息/DFS版本与现网保持一致 • 新扩容部分容量的SNS应与现网剩余服务时间保持一致 • 支持同一个存储池扩容（EC不变），暂不支持不同存储池扩容 • 新扩容硬件维保年限及级别建议与现网硬件剩余服务时间保持一致	DFS对称式架构针对元数据和存储节点融合（部署元数据盘的节点称为混合节点，没有部署元数据盘的节点称为存储节点），分为容量型（采用5288 V6）和轻量型（采用2288H V6）两种类型。支持基于节点扩容，可扩展元数据（成对扩展，最低扩展1对节点）和存储节点（最低扩展1节点），对称式架构只能扩展为对称式架构。

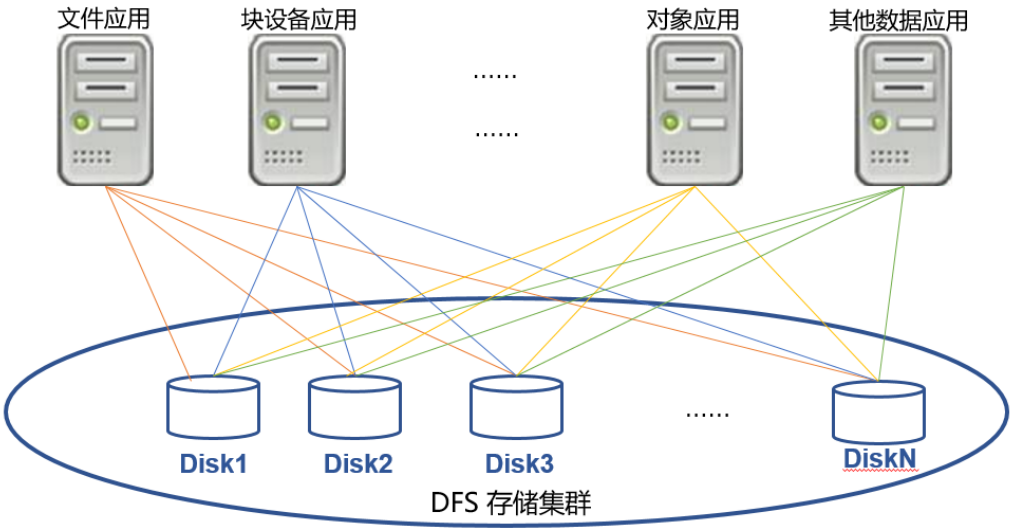
 版权所有 © 超聚变数字技术有限公司

?.13. 容量管理

DFS存储系统通过LeoFS协议层对底层硬件资源的整合，为上层应用提供一个统一的数据存储空间和命名空间。所有应用端均可在线访问全空间实时数据。元数据和数据分开存储，各成集群，可通过单独增加存储服务器或元数据服务器来扩容数据存储空间或命名空间，以及提升相应的数据或元数据访问性能，可同时满足元数据密集型存储需求和数据密集型存储需求，避免了空间浪费。单系统（卷）支持EB级数据及不限数量的文件存储，实际上线系统单卷支持最大文件数超过5000亿。



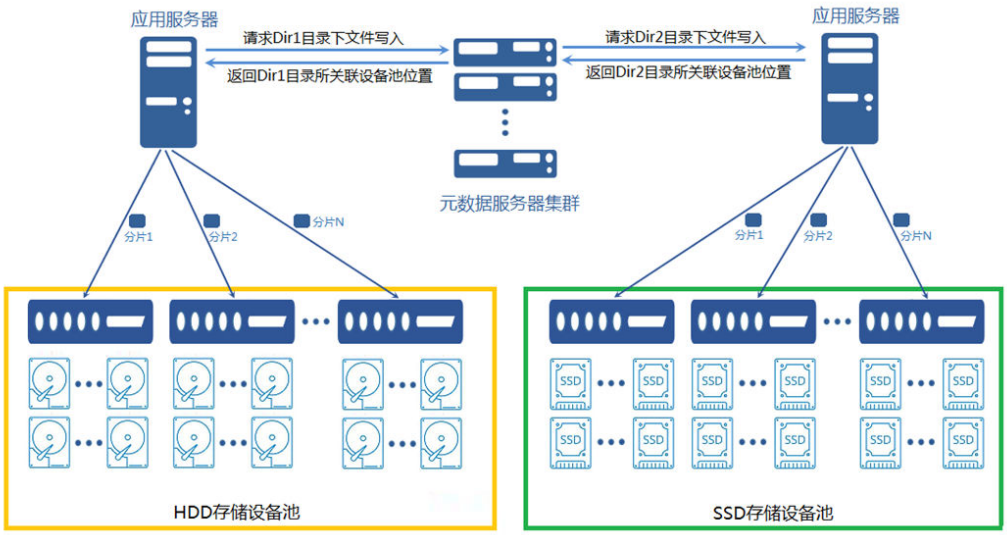
DFS存储单卷可同时支持文件、块以及对象类应用，各类应用能够同时共享使用集群中的所有硬盘，系统支持块设备在线动态扩展，块设备的容量大小不做限制；数据库支持 Oracle、SQL Sever 等数据库以 Oracle RAC的模式进行挂载使用，同时支持国产数据库。



版权所有 © 超聚变数字技术有限公司

?.14. 数据迁移

默认情况下，DFS存储系统只有一个存储池，包含系统中所有存储设备，此默认模式适合全系统使用单一类型的存储设备的场景，例如系统全部使用HDD构建。DFS系统还支持使用不同种类的存储设备（HDD、SSD、SAS）混建，且可根据介质、性能及容量等指标，将不同种类的存储设备进行分组，构成多个数据存储池，通过目录与相应的存储池进行关联，达到指定数据写入相应存储设备组的目的，在全局统一命名空间的条件下，同时满足高安全级别、高性能以及大容量的数据存储需求，且可有效控制硬件成本。



如上图所示，目录Dir1关联HDD存储设备池，目录Dir2关联SSD存储设备池，目录和存储设备池的关联信息，以及每个存储设备的存储池所属信息均保存在元数据服务器中。在应用服务器向元数据服务器发送该目录的写入请求时，元数据服务器会根据该目录与存储设备池的关联信息在相应的存储设备池中选择负载最小的存储设备分配空间，并将存储设备的位置信息返回给应用服务器，应用服务器随即与相应的存储服务器建立通信，并把数据存放在指定位置，由此完成指定设备类型的数据写入。因此，数据在存储池范围内实现了负载均衡。（在没有设备分组的情况下，数据在全局范围内实现负载均衡。）同理，当发生设备故障时，数据恢复也是在存储池范围内进行，故障设备上的数据会被自动恢复到同一存储池的其它设备上；在系统扩容时仍然以存储池为单位，通过增加相应类型的存储设备完成存储池的扩容，更加灵活和节约成本。

数据可在不同存储池间进行数据迁移，保持数据在目录结构中的路径不变，只改变其物理存储位置，且对前端应用透明。通常，高性能存储池用于存放访问频率高或对性能要求高的数据，低性能存储设备池用于存放访问频率低或性能要求低的数据。当数据的访问频度或性能要求发生变化时，可以根据需求将原先存放于高性能存储池的数据迁移至低性能存储池以释放高成本存储空间，反之亦可将低性能存储池的数据迁移至高性能存储池以提高数据服务性能。

?.15. 集群监控

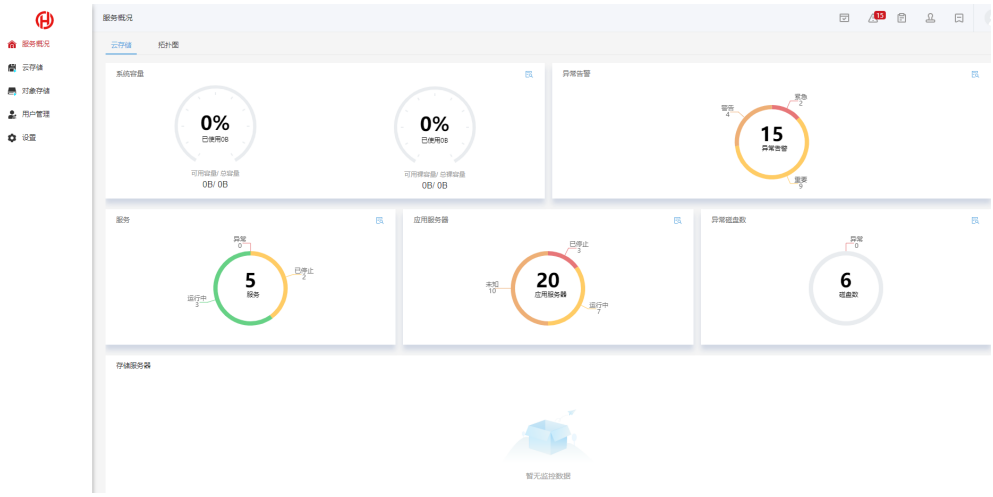
DFS存储系统提供了基于Web的图形化管理界面，可在该界面上进行存储系统的状态监控、系统配置和运维管理的所有的操作。

- 监控系统运行

  用户可在管理界面上获取存储系统运行的详细信息，包括存储系统整体空间使用情况、系统服务带宽、服务器运行状况、服务器的网络、磁盘、操作系统、内核

版权所有 © 超聚变数字技术有限公司

等硬件信息，以及当服务器、存储设备或网络出现异常时的报警信息。用户可根据报警提示采取相应的措施。

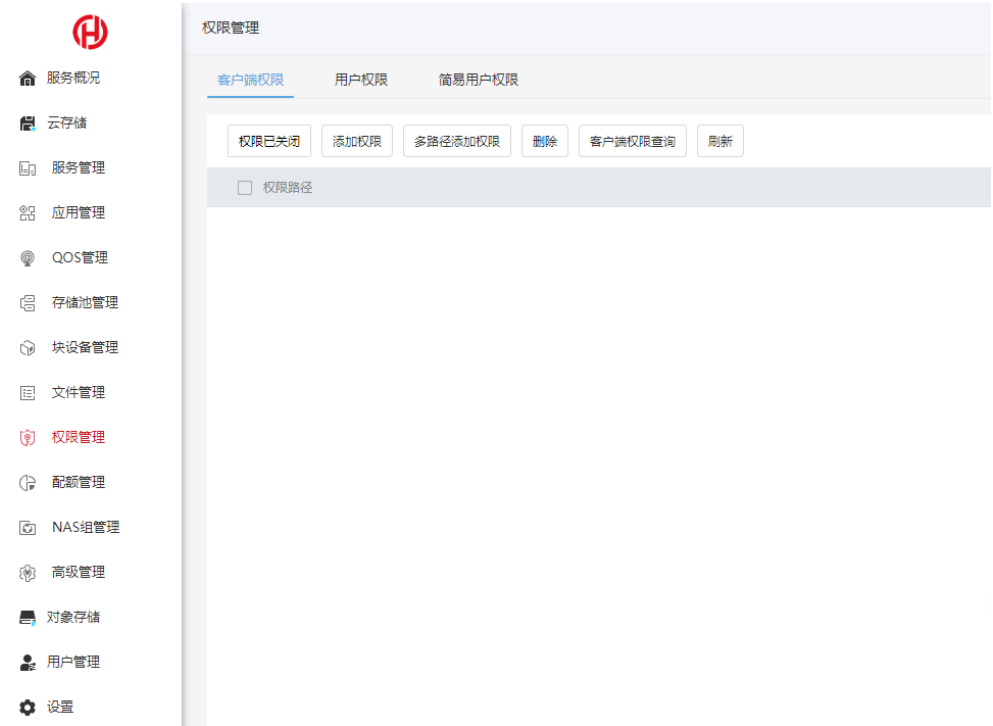


- 配置存储系统

  用户能够通过图形化管理界面完成从配置集群存储系统开始到设置完备给最终用户使用的全部操作。包括添加存储服务器及元数据服务器并启动服务、添加应用服务器、配置数据隔离及权限控制策略、配置分级存储，关联相关目录、设置目录冗余模式、设置用户权限以及多种协议文件共享等。

- 运维管理

  DFS存储系统内置管理向导功能，将所有常规运维操作以向导的方式进行引导，降低了运维难度。可实现单点管理多处数据中心。在管理界面上可进行存储系统提供的所有功能的配置和管理，用户可根据具体的应用场景合理规划文件存储空间，必要时对数据进行逻辑隔离或物理隔离，设置客户端、内部用户及域用户的权限等访问控制机制。



版权所有 © 超聚变数字技术有限公司

?.16. 告警管理

DFS存储系统告警平台用于监测存储系统服务器宕机、网卡掉线、网卡降速，操作系统盘只读，存储系统坏盘、目录空间达到上限、超过阈值、部分元数据服务未启动等异常情况，并在异常情况发生时产生报警信息显示在管理系统图形化界面上，以便相关人员及时获知系统状况以便采取相应措施。用户可对存储系统的各种异常自行设置告警级别（紧急、重要、警告）和报警方式（普通报警信息或弹窗报警）。所有报警信息均会纪录在历史告警板块中，可随时查看相应时段的告警信息内容。

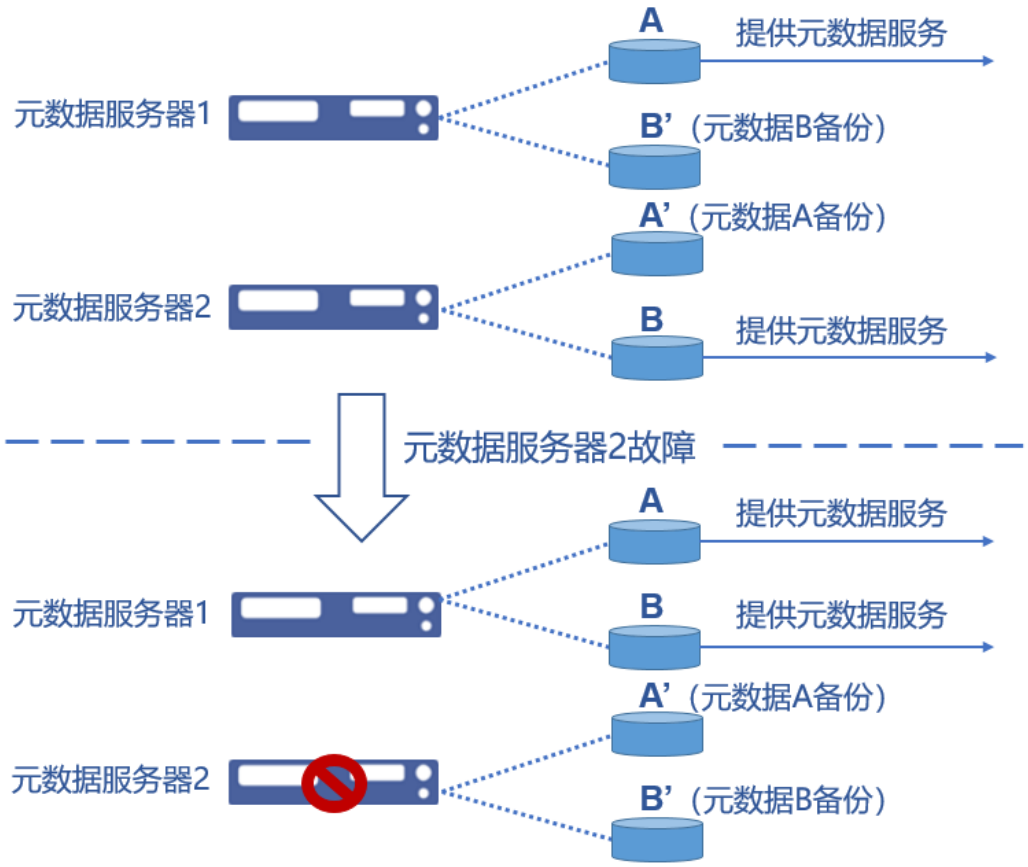


?.17. 集群双活

DFS存储系统采用双集群（存储服务器集群和元数据服务器集群）架构来保证系统的高可靠。元数据服务器与存储服务器分离各成集群的架构有助于系统容量和性能的有效扩展，可独立扩充数据的命名空间、存储空间以及服务性能，有力应对元数据操作密集型应用和数据操作密集型应用的增长需求，且避免了空间浪费。系统性能随容量扩充呈近线形增长。

存储服务器集群架构配合副本和纠删码数据冗余模式可应对系统中单点乃至多点故障而保证数据不丢失，可容忍多达4台存储服务器同时发生故障。DFS集群内嵌故障自动检测以及数据自动恢复机制，在服务器、磁盘或网络发生故障时，系统可以自动感知故障并启动数据恢复进程，无需人工干预，极大提高了系统可靠性，同时降低了运维人员的工作强度。

版权所有 © 超聚变数字技术有限公司



元数据服务器集群采用偶数互备双活元数据集群技术，即每两台元数据服务器配对工作同时对外提供服务，每台元数据服务器提供本组内50%的元数据访问服务，同时采用组内同步机制备份另50%的元数据。如上图所示，当有一台元数据服务器故障时，正常的元数据服务器将接管其元数据服务，持续对外提供服务且对上层应用透明。当故障元数据修复后，可人工或自动切换回正常元数据负载均衡工作模式。一个最小规模的元数据服务集群至少要包含2台元数据服务器，后期可根据存储系统的实际需求动态在线扩容。元数据服务器集群采用偶数扩容机制，每次必须添加偶数台元数据服务器，即配即用，无须停止元数据服务，系统元数据处理性能与元数据服务器集群规模呈近线形关系。

?.18. 安全审计

DFS存储系统提供了数据访问日志功能，可在元数据服务日志中记录用户对文件/目录的创建、删除、重命名、修改操作的详细信息，方便审计。该功能默认关闭，可通过修改配置文件的方式开启该功能。

?.19. 升级维护

DFS存储系统支持在管理界面中存储升级，该功能用于在监控端对底层存储软件包进行升级，监控端升级方式属于同步升级，在升级前会进行配置文件和软件包的自动备份工作，并对升级的软件包进行校验检查，目前没有回滚设置，因为同步升级只有成功或不成功，不成功会自动回滚老版本，因此没有回滚设置；如果是滚动升级那就需要运维人员在后端手动进行逐台机器的软件包更换停止和启动服务的操作来进行滚动升级操作。

版权所有 © 超聚变数字技术有限公司

?.20. 接口能力

DFS存储系统支持对接SNMP Trap服务器，指定的SNMP Trap服务器可以接收到系统服务器状态、系统报警信息等监控信息。

?.21. 数据扫描

数据扫描为了保护冷数据在长期不使用的情况下损坏恢复。可以开启/停止数据扫描，并展示扫描进度、扫描状态以及扫描日志信息。可以根据需求，设置扫描周期，进行自动扫描。

?.22. 数据隔离

DFS存储系统在默认情况下，所有数据资源被所有应用服务器全局共享。为了隔离不同应用之间的数据，避免其访问不相关的目录而带来非法访问以及误操作等风险， 存储系统提供了导出子目录功能，即通过建立存储系统的某个子目录和相关应用服务器的关联，使该应用服务器挂载后只能看到该子目录中的内容，有效限制了其对其它目录的访问，从而避免了非法访问及误操作。

?.23. 传输协议安全

DFS存储系统自带数据加密与数据块序列组合策略。所有待存储的文件都将在客户端进行自动切片封装加密，之后通过网络发送到各个存储服务器，在服务器端实现密文保存，文件的组合以及明文的获取只能通过存储系统客户端程序获取，其它应用程序不管是通过网络非法截取或者非法获取硬件设备，都无法对数据进行解密和数据块的组合。并提供专用、高速、安全的私有客户端、 POSIX标准分布式文件访问协议进行数据的安全访问。

?.24. QOS 管理

存储系统提供用户的服务质量（QoS）设置功能。可设置客户端访问存储目录的IOPS、最大写、最大读的服务上限值，使存储系统资源得到合理的分配，保证业务核心业务不受其它业务影响，从而为用户提供更好的服务。

?.25. 块设备管理

存储系统提供基于文件的分布式块存储服务，使用LeoSAN块设备的应用服务器在挂载了集群存储系统后可看到映射到本机的块设备，并可以像使用本地块设备一样使用LeoSAN块设备。

存储系统的块设备支持Oracle RAC、MongoDB、MyAQL等数据库应用，支持asmlib方式

存储系统支持创建无限个块设备，并支持动态扩展，块设备的容量大小没有限制，最大可扩至文件系统的总容量。

?.26. 对象存储

存储系统提供的分布式对象存储服务，用户可通过标准AWS S3或Swift接口将对象数据存放在存储系统中，并与文件访问协议实现数据互通。

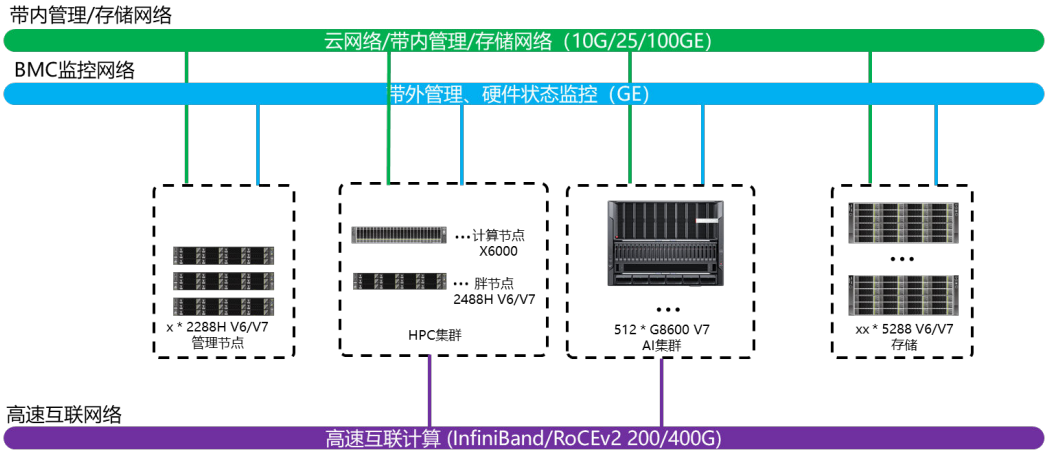
DFS对象存储支持多租户应用，可为不同的租户分配不同冗余策略的存储池。支持存储桶策略，并可与用户权限配合使用，以控制用户对桶及其中对象的操作权限。支持租户及存储桶配额设置、防盗链设置以及CORS跨域资源共享，支持对象多版本控制，支持定时删除对象或者转换存储类别等生命周期规则设置功能。

 版权所有 © 超聚变数字技术有限公司

存储系统支持大文件分片传输及直通存取，有效地降低了大文件的网络超时及重传概率，提高了数据的传输和存储效率。为用户提供了安全、稳定、高效、易用的图片处理服务，用户可以通过API接口在任何时间、任何地点对对象存储中的图片进行瘦身、剪切、缩放、水印、格式转换等处理，并且可以快速获取到处理后的图片。

2.2.3 网络

图 2-7 FusionOne HPC 组网图



- 计算网络
  承载HPC服务器协同计算的通信流量以及访问高速并行文件存储的流量，FusionOne HPC支持HDR/NDR IB组网以及100G RoCE V2组网。
- 业务管理网络
  用于集群系统带内管理或集群登录，承载所有服务器和网络的管理流量。
- 带外管理网络
  用于带外管理监控，承载所有服务器的BMC管理流量。

版权所有 © 超聚变数字技术有限公司

2.3 全栈软件

图 2-8 FusionOne HPC 软件栈



2.3.1 FusionOne Center for HPC

FusionOne Center for HPC面向于高性能计算及人工智能计算环境，集中供给计算资源（CPU/GPU）、应用资源（高性能应用程序、深度学习框架），提供WEB桌面化集群操作界面，实现应用程序的便捷化服务化管理；同时满足全面的运维管理与运营分析需求，涵盖用户管理、资源管理、应用管理、作业管理等功能，提高集群运维效率，使用效率，为决策层的投资、采购决策提供详实的数据支撑。

提供HPC与AI训练等作业处理能力，完整的为业务用户提供所需的计算资源调度能力、数据管理文件传输能力和性能分析能力，让业务用户完全告别繁杂的命令行，真正提升工作效率。

2.3.1.1 OneMind

OneMind智慧引擎是在传统HPC/AI计算集群系统中额外增加的一个增强软件组件，通过联动硬件BIOS，BMC，操作系统，调度器实现应用和业务负载感知的性能增强，节能优化，可靠性增强，以及安全性增强，并可以辅助调度器实现易用性的提升。

性能加速

通常集群在初期部署阶段会设定一套固定的硬件参数，如BIOS、OS内核配置等。而实际运行中，每个应用对硬件平台的需求都有差异化的要求。超聚变在过往的项目中

版权所有 © 超聚变数字技术有限公司

积累了大量的应用调优经验，将经验固化总结成一套体系方法，并将方法论和经验工具化，OneMind应运而生。OneMind感知应用特性，与HPC作业调度器联动，根据应用调配参数模板对硬件和系统参数调优，提升应用性能，大大提升HPC集群资源利用率。当用户在提交作业时，OneMind通过调度器插件感知到作业调度到相应的节点，在该节点生效相关模板参数，提升应用性能；作业运行完成后，恢复原有的参数。通过自适应动态调节系统运行参数的方式，OneMind将用户的平台环境进行调整，发挥出硬件的潜力，使系统资源利用率最大化。

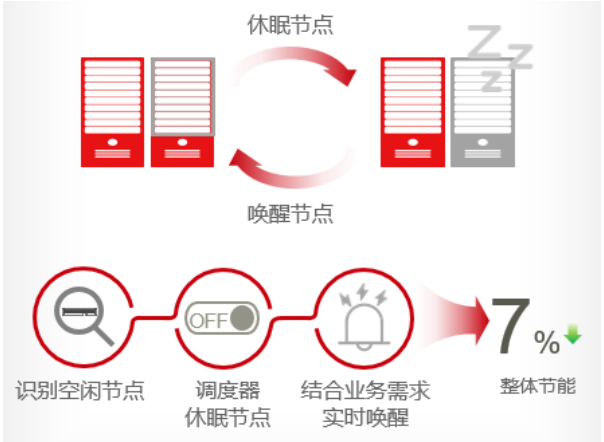
OneMind加速引擎通过分析系统和应用运行状况，可以对BIOS、OS、编译器、数学库、运行时和IO等超过200+关键参数进行调优，通过超聚变实验室数据，在Gromacs和NAMD两个主流分子力学应用软件中，性能提升均超10%。OneMind的工作原理是在应用启动运行时，将应用所在服务器的硬件，操作系统，运行时以及IO相关的参数动态调整实现每个应用在运行时都可以处于一个最佳的环境，从而实现性能提升。对于不在OneMind应用列表中的应用，也可以通过在现场POC测试将性能优化的参数预存在OneMind中，后续提交计算作业时就可以自动进行性能调优。



智能节能

OneMind提供了多项智能节能技术：

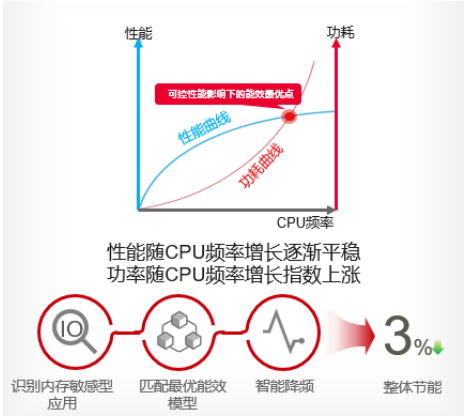
- **自动开关机**：对于集群负载不是非常高的场景，当集群中节点空闲后，OneMind可以联动调度器和服务器的BMC，自动对空闲的服务器关机。当有计算业务需要更多服务器时，OneMind自动启动服务器，无需管理员和使用者干预。在80%的集群平均负载下可以获得7%的整体节能效果。



- **智能降频**：对于访存密集型的应用，可以在不影响应用性能的前提下，通过降低CPU频率的方式获得节能效果，可以得到至少3%的节能。已知在典型的流体仿真应用OpenFOAM中可以在3%性能损失范围内获得20%的节能收益。对于GPU应用，可以在下发作业的时候指定功耗或者频率运行应用，以达成预期的功耗或者性能。

版权所有 © 超聚变数字技术有限公司

- OneMind感知作业下发，指定CPU的功耗/频率百分比，作业完成后恢复CPU功耗。
- OneMind感知作业下发，指定GPU的功耗/频率百分比，作业完成后恢复初始状态。



- **功率封顶**：集群系统配置的功率上限，最初，该功率平均分布在所有计算节点上。然后，调度器根据监控的计算节点实际功耗并酌情重新分配功率，即使用低于其上限的节点降低功率上限，并在其他节点之间重新分配该功率。提高或降低节点功率上限的阈值是可配置的。此外，在节点上启动作业会立即触发将节点的功率上限重置为更高级别。

  在基础功率封顶的基础上，也可以结合高峰时段、平时段、低谷时段电价设置不同的功率封顶策略。再将设置好的策略应用到不同的时段当中，从而达到节约电费的目的。在应用功率封顶策略的同时，也可以在设置界面中同时开启动态节能功能，根据每个节点的作业特征动态调整降频策略，以达到最小性能损失甚至性能无损下的节能目标。

队列资源自动选择

- 用户不指定队列资源提交作业，OneMind依据作业特征自动匹配到最佳队列中运行，提升用户体验和集群资源利用率。
- OneMind定期轮巡作业队列，针对队列资源不足或者等待时间过久的作业，重新调度匹配最佳队列，保证作业及时执行。

异常检测

OneMind实现了非法作业的智能监测功能，通过联动监测硬件功耗、服务器负载、调度状态，可以秒级监测到非法的计算任务，并将检测到的告警实时上报给FusionOne Center，及时呈现给管理员处理。

弹性调度

Onemind支持在Slurm和k8s之间设置共享资源池进行资源弹性调度

支持自动调度，可配置多种自动调度策略：均衡调度（默认）、slurm优先、k8s优先

支持手工调度，该特性失效自动调度，可手工批量/单个指定共享资源池的节点，向slurm或者k8s进行调度；

参数配置

OneMind提供前端界面参数配置，支持与HPC&AI调度器联动，自动感知业务应用，自动匹配业务特征，通过智能引擎调整与之匹配的系统参数，极大提升应用运行效率；

版权所有 © 超聚变数字技术有限公司

通过感知主频不敏感的应用，在不影响应用性能的情况下降低CPU频率进行节能；对空闲节点进行下电处理，节省空闲功耗，在业务启动时唤醒节点，流畅运行。

2.3.1.2 用户管理

- 用户在登录WebUI时，通过输入用户名和密码进行登录。
- 系统会根据配置的认证方式对用户进行认证。
  - 认证成功后将使用用户相关的信息创建会话、登录FusionOne Center for HPC。
  - 认证失败的用户将无法登录FusionOne Center for HPC，根据配置，用户在一段时间内连续登录失败次数达到系统安全策略中设置的次数，系统将自动锁定。锁定时间达到安全策略设置的时长后，系统将自动解锁。

    说明

    - 系统默认连续登录失败次数：3次
    - 系统默认锁定时长：5min
- 系统为用户提供双因素认证和分权分域功能。
  - 通过双因素认证，实现对用户的增强认证。
  - 通过分权分域功能，实现对权限的细粒度控制。

双因素认证

启用双因素认证功能后，用户在登录WebUI时，需要输入对应的密码，还需要输入验证码，增强了系统的安全性。验证码由系统随机生成，发送到指定的邮箱中，配置邮箱之前请确保已经开启SMTP功能。

用户权限管理

系统为用户提供了多种角色权限，每种角色对应权限如下。

表 2-3 表 1 用户角色及权限列表

角色	对应操作权限说明
超级管理员	具备所有权限。
业务操作员	具备HPC集群上作业下发及管理权限
只读	具备所有资源的只读权限。
设备管理员	具备所有资源类型的读权限，以及设备管理所需的相关权限（包括设备导入/删除、设备升级、设备相关配置的创建/修改/删除、以及设备相关告警信息的处理）。
固件管理员	具备所有资源类型的读权限，以及所有设备的固件升级相关权限。
作用域管理员	具备所有资源类型的读权限，以及管理作用域的相关权限（包括作用域的创建/修改/删除、添加/删除作用域中的资源）。
告警上报操作员	用于接收服务器iBMC上报的告警事件，不在WebUI中显示。
文件传输操作员	用于系统上的文件传输，不在WebUI中显示。

 版权所有 © 超聚变数字技术有限公司

用户作用域管理

系统支持用户根据自己的实际需要将系统资源进行分块，定义为不同的作用域。在创建用户时，可以为其指定适配的作用域。

DFS用户管理增强

支持用户从FusionOne Center for HPC登录到DFS管理界面。

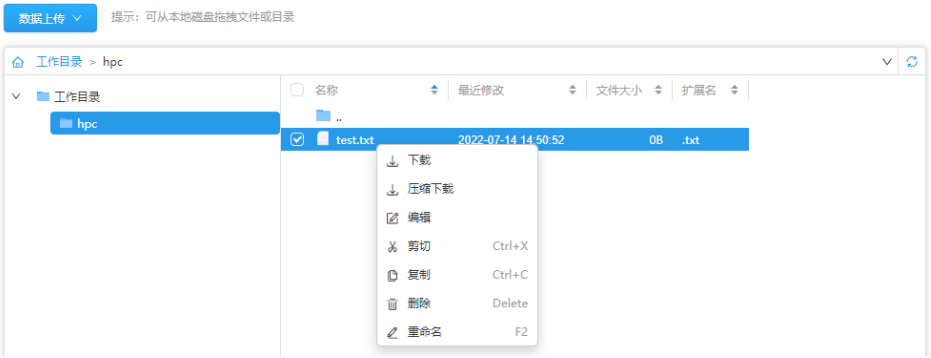
2.3.1.3 数据管理

平台支持数据管理，在平台中，每个用户都有自己的数据目录。

平台支持通过Web页面上传下载数据和管理文件。同时对于开放了SSH端口的集群，用户可以通过SSH账户来传输数据。

功能

- 文件上传下载
  - 快传：在同一界面拖拽本地文件上传到集群目录或从集群目录下载文件到本地。文件传输提供端到端数据完整性校验，确保数据正确落盘，并支持断点续传。同时可以查看传输文件的传输进度、速度、状态等信息。



  - 第三方传输工具：支持通过常见的FileZilla和WinScp等文件传输工具进行传输，延续用户以往的使用习惯。
- 作业数据同步：提供共享存储的算例文件主动传输到云端或作业结果文件回传至共享存储。此过程完全自动，无需人工干预。
- 文件操作：支持拖拽式文件管理操作。支持通过界面对个人文件夹数据文件进行复制、移动、删除、重命名、压缩、解压缩等操作，支持在线查看、编辑文本文件。

2.3.1.4 应用管理

平台支持根据客户需求预集成常用应用，商业应用须客户提供License和安装包。平台对用户提供各种应用程序及算法模块的管理。

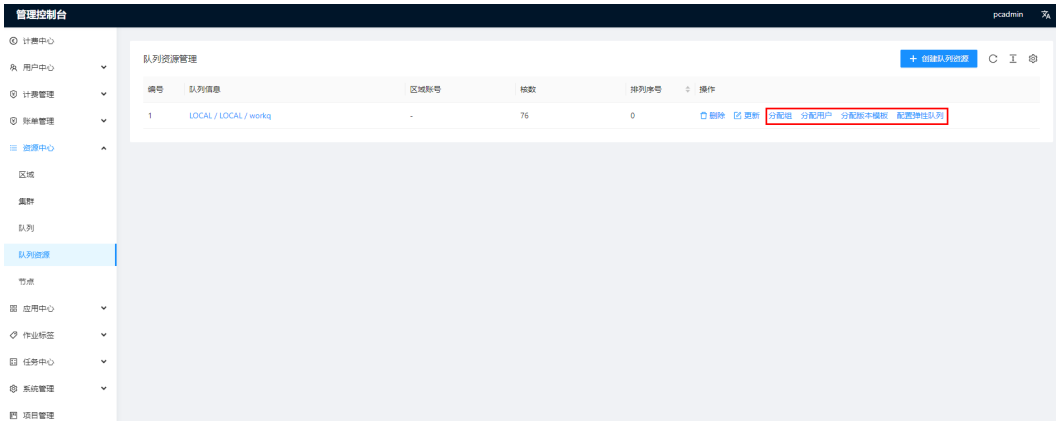
功能

- 应用集成：FusionOne Center for HPC支持调度软件如OpenPBS、Slurm等。FusionOne Center for HPC通过在后台提供“模版技术”使得传统的HPC应用可以便捷的集成到平台上，用户在前台仅需要通过图形界面设置必要的业务参数即可轻松地使用相关应用，而集群内部的复杂命令行、调度系统等概念被FusionOne Center for HPC隐藏起来。FusionOne Center for HPC可以根据用户

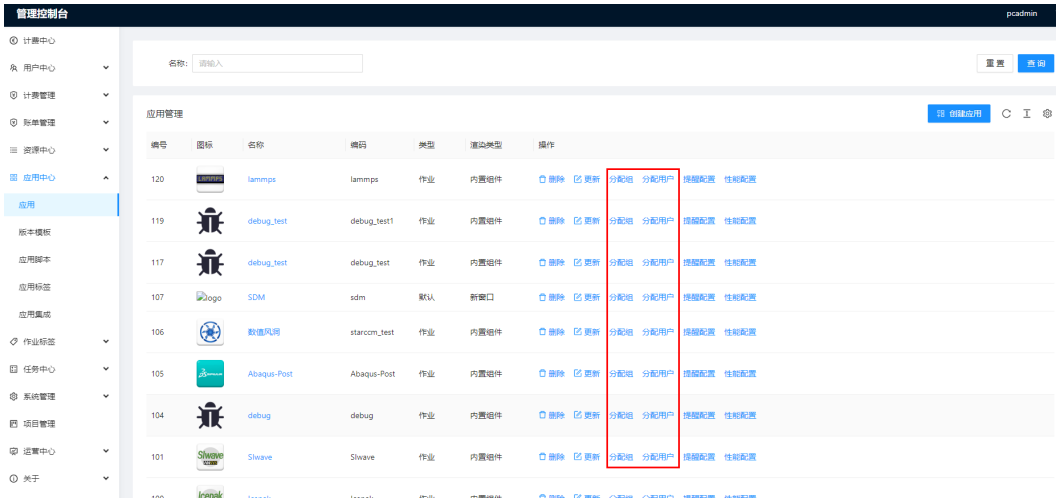
版权所有 © 超聚变数字技术有限公司

需求集成多个应用软件，默认内置了多种应用软件的模板。如果没有匹配应用软件的内置模板，还可以自定义开发。

- 收藏常见应用：用户可以从应用中心选择应用添加到登录后的桌面。以便后续快速访问，避免应用过多时搜索特定应用的繁琐。
- 应用分组：特定的应用软件或版本只能给指定的用户和用户组使用。这样可以对购买License的软件使用做好管理，仅提供给有权限的用户使用，避免License滥用。

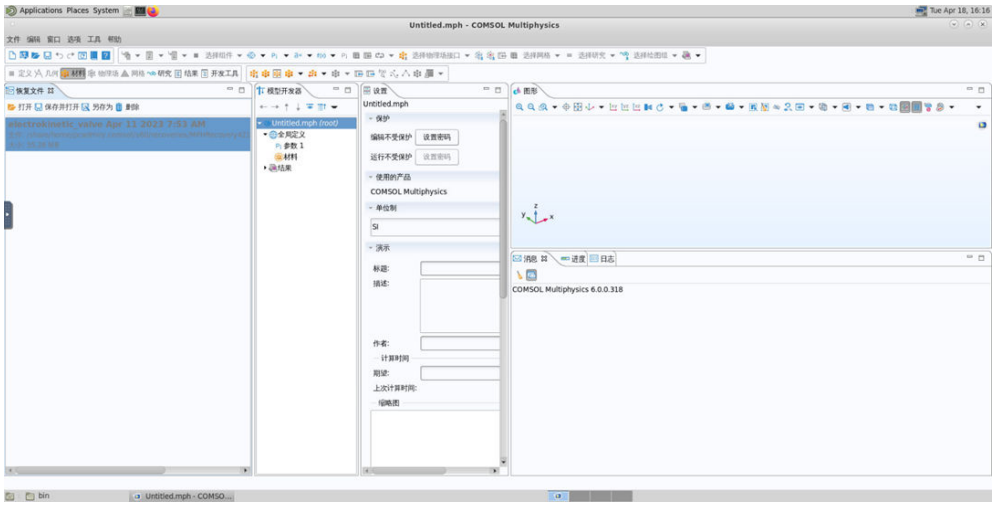


- 资源分配：集群管理员在管理界面管理队列资源的分配。合理化分配资源，有利于资源的最大化利用。



- 容器化部署虚拟应用：应用通过容器镜像方式封装，一个应用一个镜像，通过singularity容器把镜像拉起为linux虚拟机环境实现快速部署。

版权所有 © 超聚变数字技术有限公司

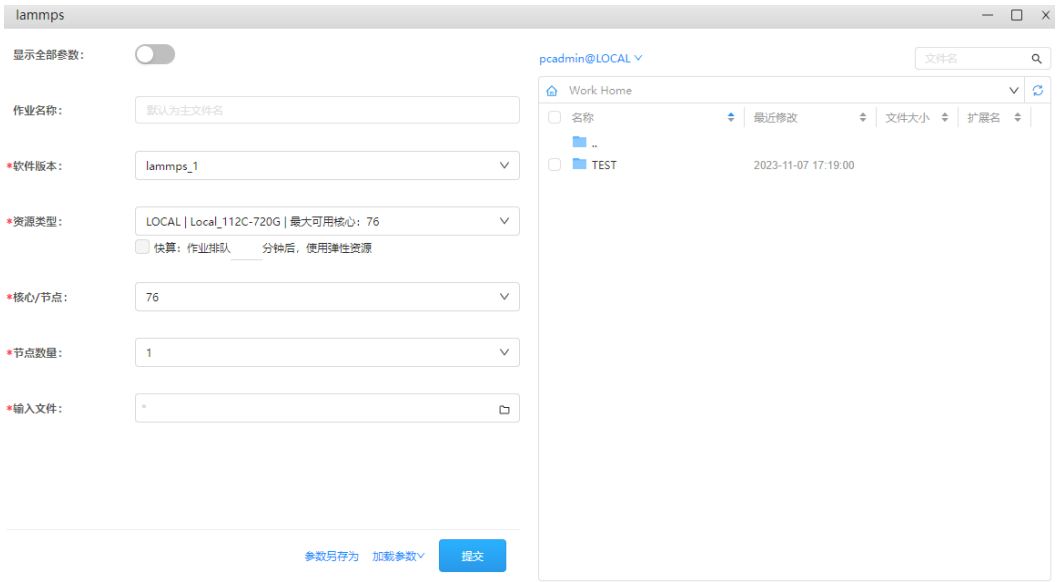


2.3.1.5 作业管理

作业管理是对作业生命周期进行管理、维护、分类统计等。

功能

- 提交作业
  - 命令行提交：支持通过webshell或者putty连接到集群，进行作业命令行提交。
  - 集成界面提交：FusionOne Center for HPC集成了各种应用软件，单击应用软件图标即可打开作业提交页面。用户可以选择软件的版本、资源类型、需要的节点数量、以及算例文件。如果需要删除选项，可以在应用集成发布界面进行修改。



  - 图形交互式作业：对于某些需要看到应用软件页面的应用软件任务，FusionOne Center for HPC提供了图形交互式提交功能。
- 作业控制管理：用户提交作业后，可以在任务中心查看当前运行作业或历史作业，对作业进行各种操作，比如：取消作业、删除作业、重新提交，如果作业在同步过程还可以暂停同步、恢复同步、重新同步。

 版权所有 © 超聚变数字技术有限公司

- 查看作业信息：可以查看作业名称、作业状态、作业运行时间、进程数量、回传大小等信息。
- 查看结果文件：可以查看到作业的结果文件、以及文件的同步情况，了解作业结果是否从HPC集群回传至本地存储。
- 调度策略：调度是线程或进程访问系统资源（比如处理器时间）的方法，通常用于有效地使系统达到负载平衡。调度是工作负载管理软件的核心，它需要一次执行多个进程。FusionOne Center for HPC支持以下调度算法：
  - 基于队列优先权的先进先出（FIFO）：目标是将CPU利用率最大化。在这个策略中，只有当一个即将运行的线程具有较高优先权时，它才会优先运行。
  - 作业和队列循环：类似于FIFO，但是基于可配置的时间总量，在一个循环时间片内，实现一个基于优先权的优先策略。
  - 公平共享：基于使用和共享值来调度作业。
  - 负载平衡：在时间共享节点和循环通过的节点之间平衡负载。
  - 专用的时间/节点：在特定的时间将作业调度到特定的节点。
  - 此外还支持资源预约、作业回填、作业抢占、作业基于优先级调度等。
- 资源控制：具备对个别用户可使用资源数量进行限制的能力。提供ACL式的安全管理控机制，管理者可允许或拒绝使用者对资源队列权限，能限制用户提交的作业数量或CPU总量，以及单个作业占用CPU核数等。
- 资源预留：支持为用户预留所需计算资源，资源包括：CPU时间、内存使用量、CPU核心、作业时长、节点数量、计算节点的类型等。

2.3.1.6 前后处理

前后处理区别于求解计算，它要求的不是多CPU核心的并行计算能力，而是GPU的显示能力，满足用户在2D/3D可视化下划分网格或是查看结果文件的使用场景。

功能

- 设计仿真一体化：FusionOne Center for HPC利用求解计算和前后处理整合，形成了一站式的前后处理与计算求解一体化体验。用户在进行网格划分、参数配置等前处理时将文件存储在FusionOne Center for HPC管辖的共享存储上，然后在FusionOne Center for HPC的应用软件界面输入算例文件进行求解，再通过后处理应用对求解结果进行渲染、观察、调试，这些步骤在FusionOne Center for HPC中一气呵成。用户不需要对文件进行任何搬运即完成了此次前后处理和计算求解过程，大幅降低了用户为了一次计算在各个不同计算环境、软件环境中搬运文件、调试环境的成本。
- 图形会话管理：用户可以查看每个打开的会话，包含会话数量、会话状态（活动、已断开连接等）、会话开启时长、运行在那台服务器上、已经打开的那个应用软件等。用户开启的会话如果长时间不在线可以自动断开。还可以对话会进行控制，比如：注销、端口、连接等。

2.3.1.7 应用特征分析

FusionOne Center for HPC应用特征分析功能好比HPC系统的CT机，将“黑匣子”的运行数据通过可视化实时呈现，并提供数据采集分析、管理控制、性能优化等功能。其价值主要体现在以下方面：

 版权所有 © 超聚变数字技术有限公司

- **日常运维：**图形可视化节点管理，提供各种日常管理工具，比如：并行Top命令，一键异常显示等功能，即使没有经验的管理员也能够快速上手。
- **监控报警：**每秒刷新集群状态，节点、存储、网络等集群部件的状态，还可以监控到CPU微架构等信息，异常情况通过不同颜色报警，让管理员第一时间知晓并处理。
- **性能优化：**自动学习历史作业特征，当某个作业性能异常时可以快速识别，同时给出性能瓶颈所在，减少计算资源浪费，由被动运维变成主动运维。
- **应用特征：**所有作业的特征均自动保存，分析历史作业寻找规律，积累应用运行特征库，深入理解HPC系统，还能指导硬件选型。

功能

- 集群状态实时监控：能够可视化监控管理节点、网络、存储、作业、进程的不同状态，可从宏观角度了解集群的整理使用情况，也可从微观视觉了解每个CPU核心的运行情况。
- 系统级性能数据：可以实时采集和显示包括CPU整体利用率、CPU系统利用率、内存利用率、SWAP利用率、磁盘读速率、磁盘写速率、网络接收速率和网络发送速、协处理器利用率，万兆位以太网和InfiniBand网络高速互联的网络设备、网络文件系统吞吐速率，服务器功耗等信息。应用运行特征收集器软件以“秒级”刷新采集这些数据，达到实时准确的效果。
- 协处理器性能数据：可以实时采集和显示每块GPU卡利用率、显存使用率、温度等指标，提供了快捷一键显示GPU的配置模式。
- 微架构级性能数据：采集显示的微架构数据包括浮点计算性能、实际整型计算性能、实时内存带宽、代码执行效率、向量化比例、最后一级缓存失效（LLCM）等。
- 应用级性能数据：采集显示了每个CPU核上所运行的进程名称和所对应的用户名称。
- 函数级性能数据：实时采集和显示函数级性能数据，给出函数热点分布。通过此功能开发者可以直接看到，当前程序运行时函数占用比例，定位函数运行时的系统开销。
- 快速显示异常节点：具有一键显示异常节点功能，帮助管理员快速发现集群中的异常问题，快速分析诊断，提升集群管理效率。可捕捉异常包括坏节点、宕机节点、磁盘满节点、低效节点等，包括CPU缺失、内存缺失、网络降速、节点不可用、磁盘利用率过高、磁盘等待时间过长等异常状态。
- 进程管理：提供对大规模集群的并行进程管理功能，前台实时显示大规模集群的进程状态，并支持按条件快速搜索查询，针对出现异常的进程提供便捷的管理功能。能快速识别残留作业（作业运行结束进程未正常退出占用CPU资源）或非法作业（比如：没有通过作业调度系统运行作业）等。能够查看集群进程，快速识别僵尸进程（CPU100%，systemtime100%；作业退出，进程残留），一键杀除节点中的异常进程。
- 智能识别作业异常性能：可以监控集群内运行的作业。可以查看作业号，作业名称，用户，使用的节点数量，进程名字等信息。可以自动学习历史作业运行情况，设置作业性能差值范围，当前运行中的作业超出性能异常范围时会通过红色报警提醒。
- 作业运行特征数据保存：能够自动收集每个作业的应用特征数据，以秒为单位收集作业生命周期运行特征：包括硬件配置、应用名称、作业信息、节点性能信息（CPU利用率、Gflop值等）。可对收集的数据进行离线保存，作为故障定位及系统优化等离线分析的依据。
- 作业运行特征分析及优化：对采集的所有作业的特征数据进行并列历史线条显示，通过多节点并列的对比分析以及资源的依赖程度分析，可迅速发现系统故障

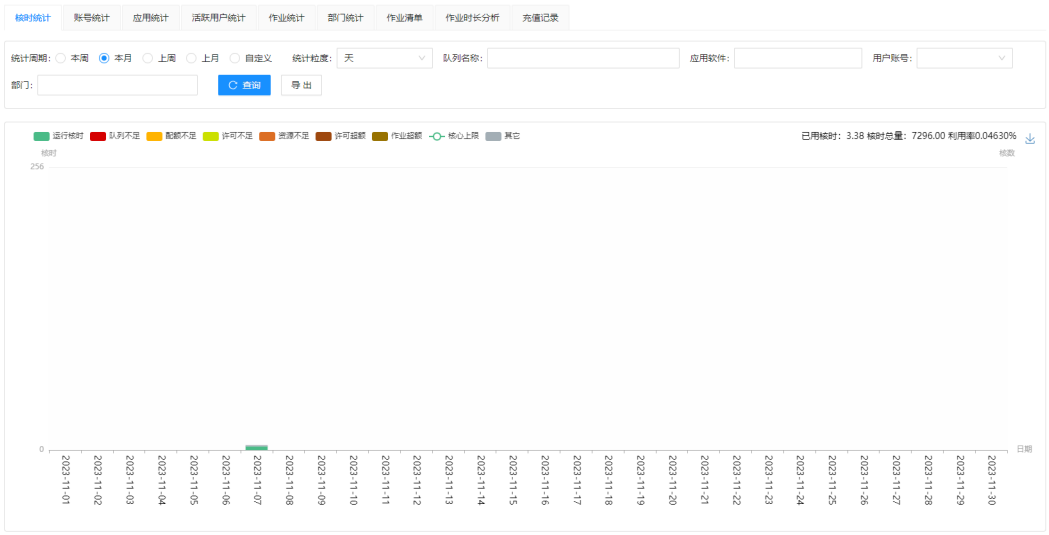
 版权所有 © 超聚变数字技术有限公司

或瓶颈所在，分析出不同应用软件内存用量、内存带宽、以及不同进程规模下的性能差异，对于应用优化具有精确的指导意义。

- 支持应用的硬件配置选型：通过对采集的所有作业的特征数据进行对比分析和资源的依赖程度分析，为不同的应用软件选择合适的硬件配置，实现应用使用效率最大化。

2.3.1.8 统计中心

FusionOne Center for HPC使用大数据流式处理技术，根据业务的需要将实时监控指标聚合、关联成系统所需的分析指标，为系统使用者提供各种指标维度、时间维度、业务维度的监控与统计分析数据，并以最合理的方式图形化展现给用户。



功能

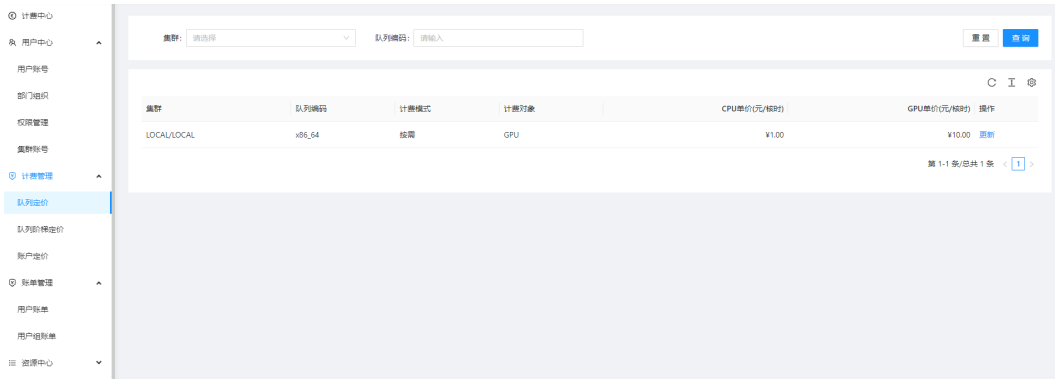
- 核时统计：可以按月/天/小时/查看历史上每个时刻的运行/排队作业数量、运行/排队核时数量，以及不同排队原因的统计，为运营决策提供数据支持。
- 账号统计：可以按账号统计每个用户和用户组的作业核时、作业数量等。
- 应用统计：按应用统计每个应用的作业核时和作业数量。
- 活跃用户统计：可以按月/天/小时/自定义时间查询，展示不同节点活跃用户数量信息。
- 作业统计：可以按月/天/小时/自定义时间查询，展示不同节点作业核时和作业数量信息。
- 部门统计：可以按月/天/小时/自定义时间/队列类型/部门查询，展示不同部门作业核时和作业数量信息。
- 作业清单：可以按作业号、用户账号、对列、应用软件名称查询搜索作业。
- 作业时长分析：展示不同时长作业所占数量统计。
- 充值记录：展示充值记录统计信息。

2.3.1.9 计费统计

面对超算场景的运营需求以及提高计算资源的利用率，计费系统是集群管理平台必不可少的功能模块。FusionOne Center for HPC为集群管理员提供了平台计费规则查看和管理平台计费策略相关的配置功能。平台针对计算作业计费，主要以队列维度为设

版权所有 © 超聚变数字技术有限公司

置不同的计费规则及计费单价；也可以按照作业运行占用 CPU 资源的规模进行阶梯定价。针对已有的队列的定价，根据用户使用的规模定义不同规模时的折扣。同时，计费系统可以查看计费账户的信息、绑定的队列固定定价策略、队列弹性定价策略以及费用历史记录信息。



功能

- 定价策略：支持队列定价，计时/独占模式灵活切换；支持资源规模定价，阶梯定价策略适应多样化场景。
- 账户计费：支持账户类型分组，批量管理，可将不同用户（组）关联至队列策略，按需动态调整；支持用户单点 / 批量操作，充值、补偿、奖励等多种计费方案。
- 账单管理：支持账单管理、查询功能，根据查询结果进行账单信息生成导出；支持费用历史信息统计。

2.3.1.10 多集群

为了支持用户建设分布式多集群的HPC混合云架构，FusionOne Center for HPC支持多集群的管理。根据不同集群认证系统、调度系统的集成，FusionOne Center for HPC可以在企业内形成服务中心与资源集群的概念，用户只需在平台上登录，即可根据各自的权限同时访问多个集群以及该集群上的应用，使得资源管理的集中化、智能化成为可能。FusionOne Center for HPC在权限上对各个集群资源的访问有着非常精细的控制，以便管理员更好的通过自动化手段管理多个集群。

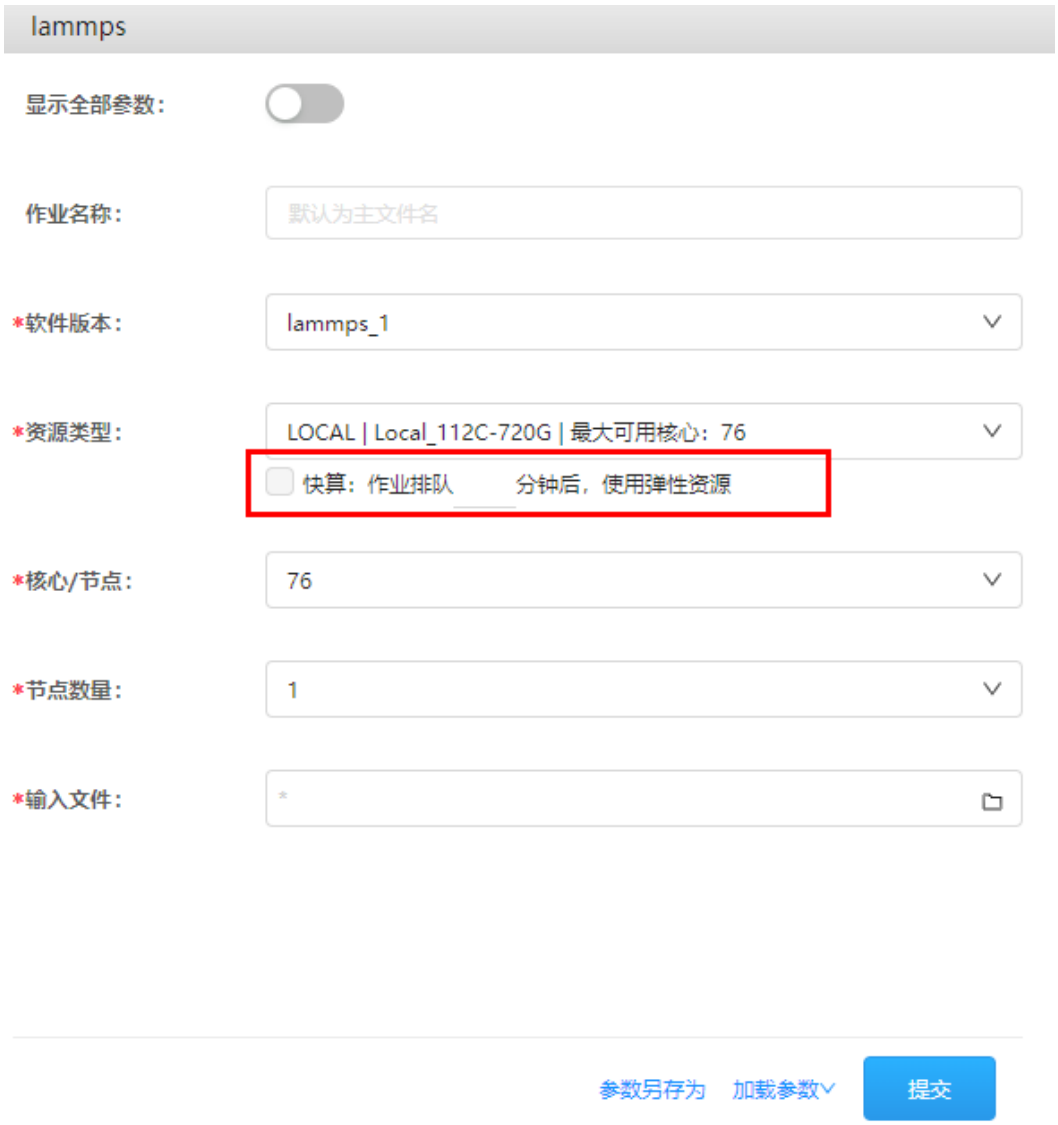
 版权所有 © 超聚变数字技术有限公司



2.3.1.11 弹性上云

弹性计算作为FusionOne Center for HPC实现混合云平台的重要功能，它结合公有云的弹性特性，同时保证数据安全和用户习惯。用户在提交作业时勾选“快算”，系统会优先使用固定队列的资源。如果固定队列资源不足，出现排队且时间超过设定的阈值，系统就会将作业发送到弹性资源的池子。作业获得资源后马上开始计算，作业完成后结果文件回传至本地存储。

版权所有 © 超聚变数字技术有限公司



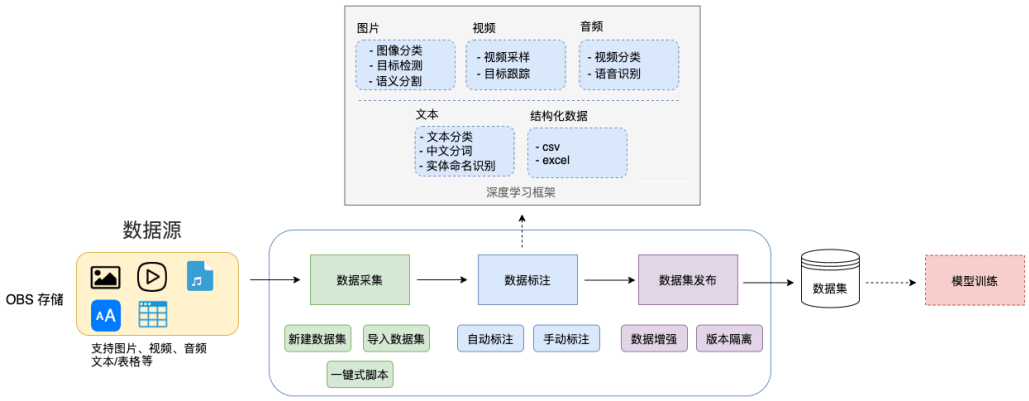
2.3.2 AI 平台

2.3.2.1 数据管理

AI开发过程中经常需要处理海量数据，数据准备与标注通常占据整体开发一半以上时间。数据管理功能提供了一套高效便捷的管理和标注数据框架。支持图片、文本、语音、视频等多种数据类型，涵盖图像分类、目标检测、音频分割、文本分类等多个标注场景，可适用于各种AI场景，如计算机视觉、自然语言处理、音视频分析等；同时提供数据筛选、数据分析、数据处理、智能标注以及版本管理等功能，AI开发者可基于该框架实现数据标注全流程处理。

 版权所有 © 超聚变数字技术有限公司

图 2-9 数据管理业务架构



数据管理模块集成了数据导入、数据筛选、数据标注、数据增强、版本管理等一站式数据服务。提供自动标注、数据增强等一系列数据加工方案，拥有高质量的数据标注处理算法，输出高品质的数据，支持下游AI数据训练获得更优的训练效果。

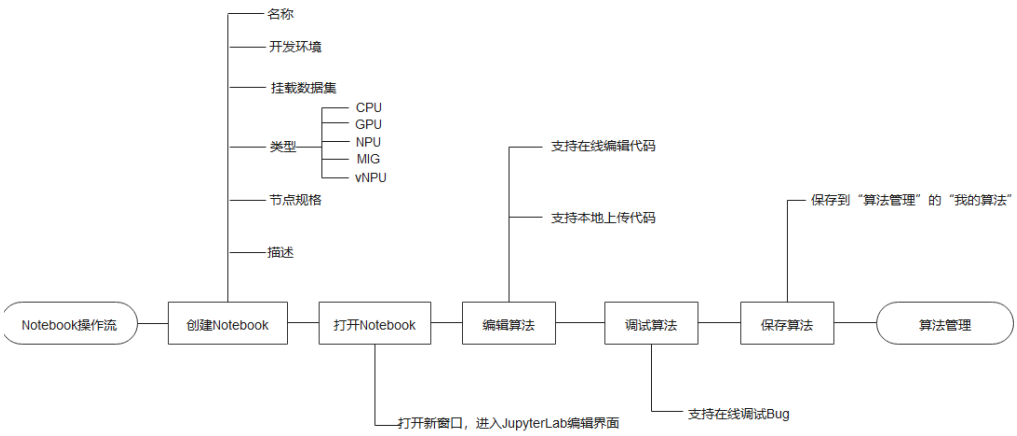
数据管理功能目前支持图片、视频、文本、表格、音频、自定义等数据类型，支持图像分类、目标检测、语义分割、目标跟踪、文本分类、中文分词、命名实体识别等数据标注功能。内置数据增强功能（针对图片类型），可以针对已有数据集进行快速扩充，获得更好的数据多样性。

2.3.2.2 算法开发

算法开发模块是AI平台提供给用户开发算法的模块，包含Notebook和算法管理两部分：

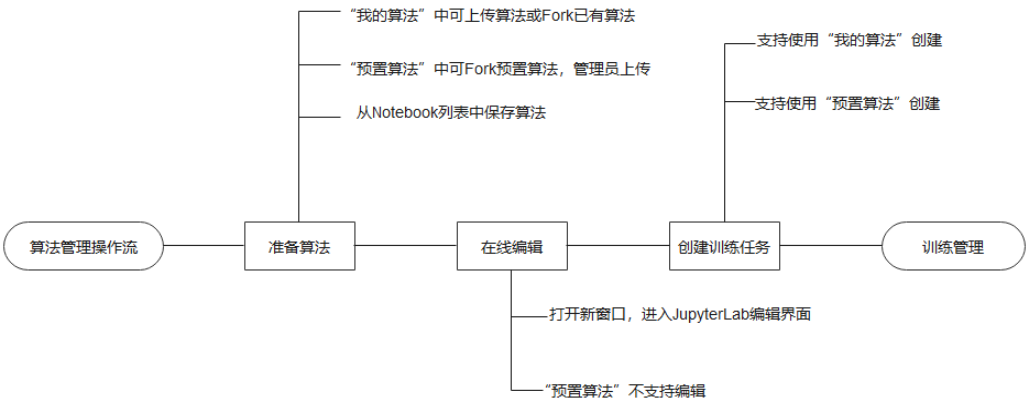
- Notebook是一种交互式编程环境，AI开发者可以在云端进行算法的开发。该模块集成了开源的JupyterLab，可支持开发者在线编辑、调试、运行代码，同时预置了PyTorch，TensorFlow等多种深度学习框架，用户可在多种框架之间自由切换。每个Notebook都是一个独立的编程环境，用户可以对Notebook实现创建、打开、停止、启动、删除等操作。算法开发完成之后，还可以将其保存到“算法管理”的“我的算法”中，进行后续的训练工作。
- 算法管理用于保存一些平台预置的算法以及用户自己开发的算法。对于已经保存的算法，用户可以进行在线编辑、创建训练任务、下载、Fork、删除等操作。

图 2-10 Notebook 操作流程



 版权所有 © 超聚变数字技术有限公司

图 2-11 算法管理操作流程



2.3.2.3 训练管理

AI平台为每一个人工智能的训练任务都单独分配了一个虚拟容器进行训练，各个训练任务之间相互隔离，互不干扰，提高了训练任务的可靠性。训练任务支持TensorFlow、PyTorch等多种深度学习框架，支持使用预置算法或用户自定义算法进行云端训练。模型管理具有多版本控制功能，用户可基于现有版本进行修改，动态调整算法超参数，从而得到一个满意的模型。

支持对训练任务进行创建、停止、删除操作，支持查看任务的ID、名称、任务类型、版本数、训练时长、状态、创建时间等信息。

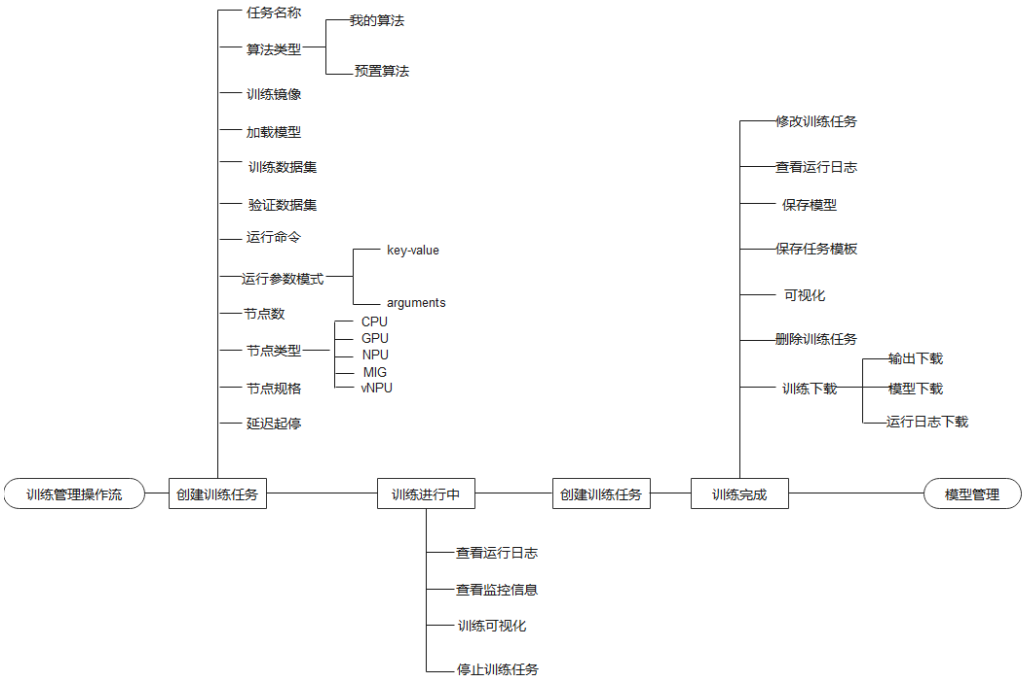
支持任务模板管理，支持创建、编辑和删除任务模板，并通过模板创建训练任务。

支持训练过程TensorBoard可视化，提供一系列直观的图表和可视化界面，帮助用户监控训练过程中的指标、可视化模型结构、嵌入数据等，为深度学习模型训练及调优提供直观的参考。

版权所有 © 超聚变数字技术有限公司

训练管理操作流程

图 2-12 训练管理流程图



2.3.2.4 模型管理

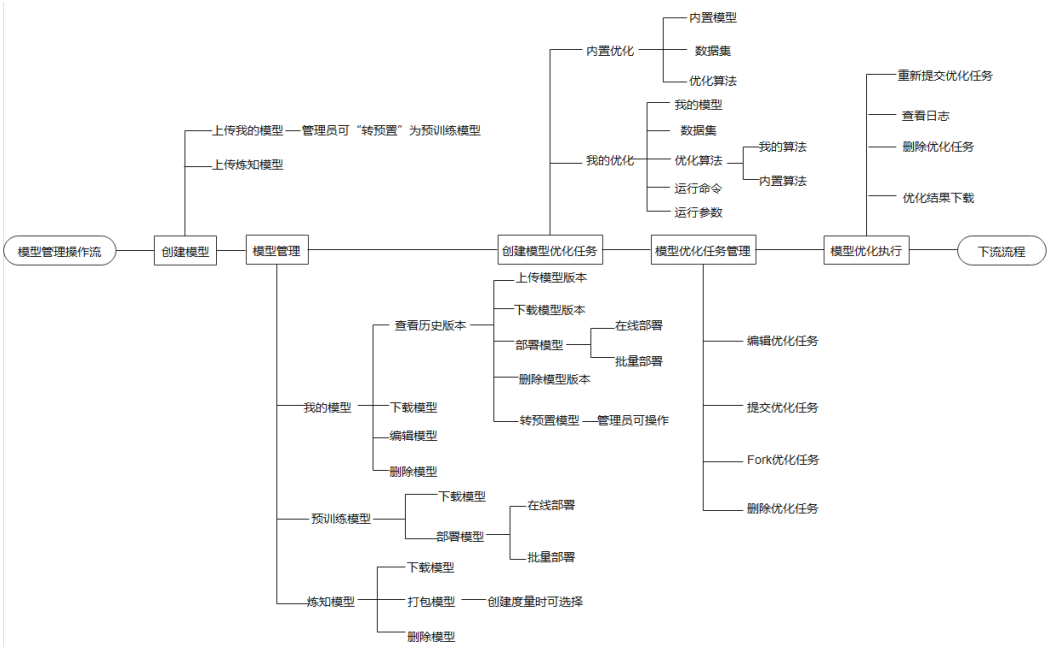
模型管理可以导入训练生成的模型，并对模型的版本迭代进行统一管理。用户可以从本地上传模型文件，也可以从“训练管理”界面中导入。用户可在此进行模型的管理，支持的操作包括创建模型、查询模型、历史版本、下载、编辑、删除等。

模型管理具有多版本控制功能，用户可基于现有版本进行修改，动态调整算法超参数，从而得到一个满意的模型。

版权所有 © 超聚变数字技术有限公司

模型管理操作流程

图 2-13 模型管理流程图



2.3.2.5 推理服务

完成模型训练及模型格式转化后，模型管理功能会保存指定格式的模型，用户可部署推理服务运行这些模型。

推理服务模块特点如下：

- 支持对TensorFlow，PyTorch两种深度学习框架训练的模型部署。
- 支持HTTP、GRPC两种通信方式。
- 支持多节点部署。
- 支持灰度发布。
- 支持基本的图片预处理。
- 支持自定义推理脚本。
- 支持使用自制镜像。

支持在线服务和批量服务这两种方式进行模型部署。

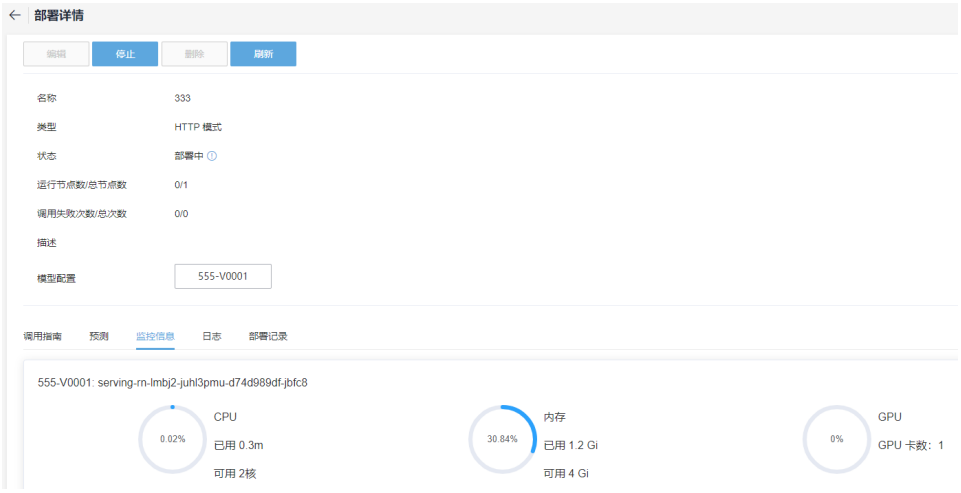
- 在线服务：将AI应用部署为一个Web Service，并且提供在线的测试UI与监控功能。
- 批量服务：批量服务可对批量数据进行推理，完成数据处理后自动停止。

2.3.2.5.1 在线服务

在服务列表可以查看到服务名称、服务描述、状态、运行节点数/总节点数、调用失败次数/总次数、服务类型、创建时间、操作(编辑、启动、停止、删除、预测、回滚)等信息。

在部署详情中可以查询在线服务信息和调用指南、预测、监控信息、日志、部署记录板块。

版权所有 © 超聚变数字技术有限公司



支持灰度发布分流，支持自定义推理脚本。

2.3.2.5.2 批量服务

可以查看服务 ID、服务名称、服务描述、状态、任务开始时间、任务结束时间等信息。

支持创建、启停、编辑、Fork、删除批量服务。

2.3.2.6 镜像管理

镜像是算法的载体。平台支持镜像的新建、上传、制作、分享、下载等操作。在界面中可以查看到镜像名称、镜像版本号、镜像用途、镜像描述、上传时间等信息。支持镜像分类如Notebook镜像、训练镜像、推理镜像、数据标注镜像等。提供Pytorch、TensorFlow、MindSpore、PaddlePaddlle、Keras、Caffe等基础镜像，有效提升用户的镜像使用效率。

版权所有 © 超聚变数字技术有限公司



2.3.2.7 租户与计费管理

为了方便不同的用户使用，AI平台支持多租户管理、配额管理、计算管理。当机构内有组织、多用户时，可显著提升计算资源的使用效率，方便系统管理员对资源进行管理。

- 支持多租户管理，每个租户有各自独立的用户空间，避免因多租户导致系统的安全风险。
- 支持针对不同的租户设置不同的用户权限，提升计算资源的管理水平和安全水平。
- 支持组织管理和部门管理，支持部门成员管理，有效的提升系统管理效率。
- 支持不同的租户设置配额和计费策略，提升计算资源管理效率。
- 支持用租户配额查询，租户账单导出，方便管理员的日常管理。

2.3.2.8 控制台

为了方便系统的管理，AI平台提供了统一的控制台，为管理员提供用户管理、菜单管理、字典管理，集群状态、资源规格管理、回收站等功能。

管理员可以通过用户管理为普通用户修改配额，包括CPU、内存、GPU、MIG、NPU等资源，以及Notebook自动关闭时间。

版权所有 © 超聚变数字技术有限公司

支持调度器配置，支持volcano和kube-scheduler队列管理。

通过菜单管理和字典管理，可以修改目录和页面及等标题名称，调整布局方式。

资源规格管理是为notebook、训练任务、在线服务等提供预定资源的，在这里创建好资源规格后，创建对应的任务或服务可以直接选择相应的资源规格进行使用。

集群状态可以查看集群的状态、CPU使用情况、内存使用情况、GPU使用情况等信息。展开详情可以查看具体的每个pod的状态和CPU、内存、GPU占用情况。支持配置给指定用户独占节点资源。



回收站是对用户删除文件的暂存，数据默认是在执行操作时间进行清理，用户可以在这个时间之前进行恢复或者自行彻底删除。

2.3.2.9 统计报表

系统提供了丰富的报表查询功能，用户可以快捷的查询硬件资源的使用情况，作业完成情况，有效的提升了系统的易用性。

- 平台支持按照集群或者硬件节点维度，生成硬件资源使用率和使用明细统计报表。
- 支持查询AI作业的状态、运行时长、硬件资源使用等信息，并导出相关的报表。
- 支持查询已经完成作业统计和明细，方便用户查询历史作业情况。

2.3.3 FusionOne Center 平台

2.3.3.1 统一监控

实时监测HPC集群的服务器信息、资源使用及作业运行，通过不同的视图方式进行组织，让用户更快更准确的找到所关注内容。时刻了解集群的运行情况。

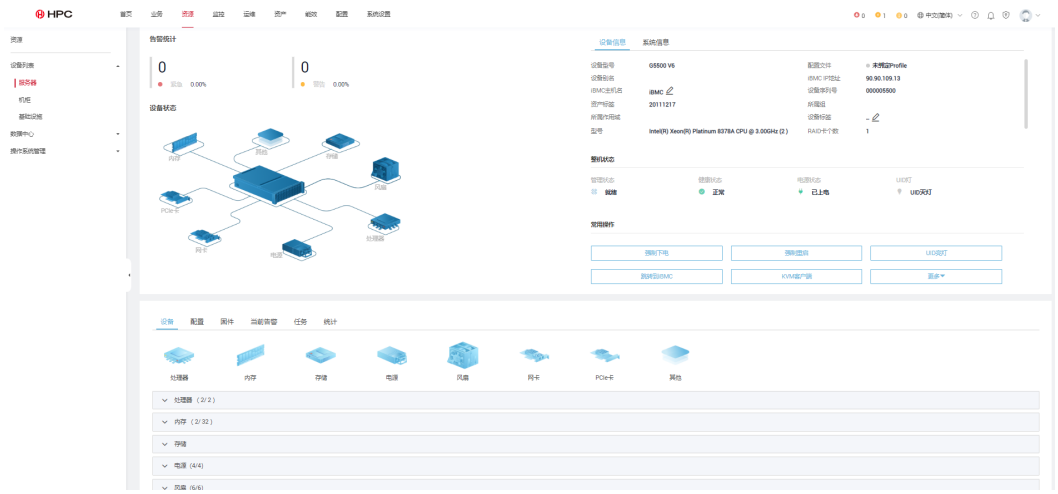
概况

可以看到集群的整体情况包括集群和设备数量、资源占用率统计（CPU、GPU、内存、存储）、节点的开关机个数统计、作业的运行数和排队数。服务器的使用率TOP5、功耗TOP5、服务器温度TOP5及最新的告警信息。

服务器信息

支持对已接入的服务器设备部件信息进行360度呈现和管理，包括电源信息、风扇信息、处理器信息、内存信息、存储信息、网卡信息、端口信息、整机信息等。

 版权所有 © 超聚变数字技术有限公司



大屏展示

- 调度器大屏：
  - 集群信息：包括集群名称、调度器类型、调度器版本
  - 作业状态监控：包括作业的实时数据和历史数据
  - 节点状态监控：包括节点的实时数据和历史数据
  - 队列状态监控：包括队列的实时数据和历史数据
  - 资源占用率监控：包括CPU核数使用率、GPU卡数使用率、内存使用率、共享存储使用率
- GPU大屏：GPU分配个数、显存分配、GPU温度、功率等。

2.3.3.2 统一告警

通过告警功能实现提前发现预防问题的效果，告警系统支持自定义告警策略，可以根据CPU、GPU、内存、文件系统等使用情况（包括CPU、GPU温度）进行设置，可以设置告警的等级等更细粒度的设置并且为了防止报警风暴可以设置告警抑制时间间隔。

支持通过Email、短信等将告警信息按照客户指定的规则通知到运维人员。

2.3.3.3 统一日志

日志是定位问题的有效手段，支持在Web界面上的统一日志管理，包括节点日志和管理平台及调度器日志。

日志搜索过滤：支持按时间对日志进行搜索，支持自定义字段进行日志过滤。

日志查看：在Web界面查看各类日志文件，快速定位问题。

2.3.3.4 智能故障管理

服务器及关键部件生命周期上限是5年，但有些部件生命周期是比较短的，部件使用时间到达上限时，故障率比较高。为了保障系统的可用性和可靠性，核心在于如何有效的降低服务器故障的影响。通过故障预测功能，可以帮助用户快速发现问题并修复故障。

通过软件内置的硬盘失效预测模型，实现硬盘“亚健康”状态在线推理分析。支持提前20天准确预测即将失效的硬盘，从而实现主动式故障预测。

 版权所有 © 超聚变数字技术有限公司

支持预测功能开启和关闭，风险信息总览，硬盘健康趋势统计，硬盘风险统计分析，帮助用户全面掌握硬盘运行状况。

支持显示已纳入系统管理的内存健康状况和预告警数量统计信息，并提供内存列表和服务器列表帮助您更好地了解内存的详细信息。

2.3.3.5 服务器管理

常用操作

对已纳管的服务器，您可以通过服务器页面了解和修改每个服务器的基本信息，并对其进行上下电、强制重启、UID点灯、跳转到iBMC、跳转到KVM客户端、删除、刷新、修改纳管帐号密码、iBMC IP配置、下线、导出日志和格式化磁盘等操作。此外，还提供服务器分组接口，便于服务器批量维护。

服务器监控

支持查看服务器的告警统计、设备状态、设备信息、系统信息、整机状态等。

- 设备：支持查看服务器主要部件（处理器、内存、存储、电源、风扇、网卡、PCIe部件等）的健康状态、基本技术参数、部件编码等信息，帮助您更快了解服务器当前状态。
- 配置：支持查看服务器iBMC、BIOS、RAID以及网卡的当前配置情况。
- 固件：支持查看服务器当前存在的组件类型、组件名称和当前固件版本。
- 当前告警：支持查看服务器的当前告警列表，并提供刷新功能。
- 任务：支持查看服务器接入系统后执行过以及正在执行的任务列表。

2.3.3.6 智能资产管理

能够实时管理机柜设备的物理位置和型号等信息，实现自动U位识别、部件级精细化资产管理、秒级自动盘点资产，盘点准确率100%，提升资源利用率。

资产概览

资产概览展示已纳入系统管理的所有数据中心的基本资产统计信息，包括概览、设备统计、部件统计、资产汇总情况、资产变化情况、资产变更事件、资产变更趋势以及机柜空间利用率等信息显示。

资产明细

- 设备明细：展示全部或指定数据中心的设备统计情况、各类设备的具体型号和数量，并且支持查看指定设备信息。
- 部件明细：展示全部或指定数据中心的部件统计情况、各类部件的具体型号和数量，并且支持查看指定部件信息。
- 支持导出资产明细信息。

资产变更

展示全部数据中心近一年的设备和部件的变更统计情况。

在变更记录表中，展示变更的设备或部件类型、变更时间、变更事件类型、序列号、型号、资产位置、所在数据中心名称、设备标签以及变更事件的描述信息。

电子保单

 版权所有 © 超聚变数字技术有限公司

- 支持查看、导入、导出纳管设备的维保信息。
- 支持配置服务年限告警。开启服务年限告警功能后，系统会在每个整点定时检测维保状态，并上报服务年限相关告警信息。

2.3.3.7 智能能效管理

系统提供功耗管理、能耗统计和能耗优化功能，支持显示所管理的智能机柜的功耗封顶情况，并提供功耗封顶相关配置功能；支持对已管理的数据中心进行能耗统计，持续记录数据中心的功耗数据，从配电、温度、低载服务器以及空间优化等角度进行能耗优化分析，根据统计信息以及参数列表，对各项参数针对性优化，以达到最优配置，实现能耗优化管理。

功耗管理

系统支持展示所管理的智能机柜的功耗封顶情况，并提供设置功耗封顶参数、开启功耗封顶、关闭功耗封顶、配置削峰填谷的功能，方便用户自行设置每台智能机柜的封顶功率及封顶策略，防止机柜功耗过大。

能耗统计

系统支持对已管理的所有数据中心进行能耗统计。用户可以根据需要选择查看全部数据中心、单个数据中心、数据中心中的单个机柜或单台服务器的功耗数据。

能耗优化

系统可对纳入管理的设备进行能耗优化分析，包括配电、温度、低载服务器以及空间优化等角度进行分析，帮您更好地了解设备能耗、优化能耗管理方案。

在各个功能页签中，您可以查看系统中最新的分析结果，可以根据统计信息以及参数列表，对各项参数针对性优化，以达到最优配置，实现能耗优化管理。

- 配电风险分析：支持查看系统中机柜供电的数量统计信息以及机柜列表，您可以根据机柜情况设置阈值，针对分析及时调整机柜状态，保证配电最优，节约能源。
- 温度风险分析：支持查看系统中服务器高低温数量以及服务器列表，您可以自行设置温度风险阈值，并根据分析对服务器进行上下电操作。
- 低载服务器分析：支持查看系统中低负载服务器的数量统计信息以及服务器列表，您可以自行设置低载服务器的统计阈值，并根据分析对服务器进行上下电操作。
- 空间优化分析：支持查看系统中机柜空间利用率统计信息以及机柜列表，您可以根据机柜情况设置阈值，根据分析及时调整，以保证空间利用最优化。

2.3.3.8 智能部署管理

支持将一台设备的相关配置形成一个Profile配置管理文件，支持批量配置，同时配置文件可导入、导出，可以快速复制到其它设备，实现硬件配置的灵活变更，提高故障设备更换、服务器扩容的效率。

配置管理

- 服务器模型：主要是配置设备的型号，高度以及设备中各项资源参数，提前规划待管理设备需要使用的模型，在配置服务器配置文件和执行上线规划时，直接选用已配置好的设备模型，有利于提高操作效率。
- 服务器配置模型：查看已有服务器模型，可通过修改设备模型详细信息，使其与待规划的管理设备匹配。提高规划和管理效率。

 版权所有 © 超聚变数字技术有限公司

- 服务器配置文件：主要为设备配置BIOS、RAID、iBMC等参数。当配置文件应用于相应设备后，设备会按照配置参数进行配置。

OS 部署管理

OS镜像

- 支持查看FusionDirector中已存在的镜像总数、存储空间以及使用率等情况。
- 支持手动导入OS镜像和自定义文件。

OS部署

支持使用iBMC带外网络，通过Smart Provisioning软件挂载镜像，提供批量基础安装OS功能。OS部署前，需先配置好RAID组，并且设置启动盘。

2.3.3.9 FusionOnline

FusionOnline为超聚变公司的Call Home产品，在取得客户授权后，通过采集客户设备及业务的信息（含设备列表、告警、日志等），并报送到超聚变服务支持中心，服务支持中心能够在第一时间监测到设备信息，并协调进行处理，做到主动、及时的处理问题。

详情请参见**《FusionOnline 用户指南》**。

2.3.4 系统环境

2.3.4.1 操作系统

FusionOne HPC集群系统中服务器节点的操作系统支持CentOS、RockyLinux、FusionOS。

2.3.4.2 开发及并行运行环境

为用户提供应用程序运行环境、用户编程环境，MPI库等。除了开源的集群编程环境项目外，各芯片厂商会根据自己的硬件架构提供整套优化的编程工具，包括GNU编译器，Intel OneAPI，NVIDA CUDA-X，OpenMPI等。

版权所有 © 超聚变数字技术有限公司

3 安全可信

3.1 安全
3.2 可靠

3.1 安全

3.1.1 行业特点与外部安全需求

HPC业内对集群系统以及中间件软件的安全要求较低，这主要与集群系统所处的网络环境相关。

HPC集群部署在企业或数据中心的专用机房，处于数据中心的内部网络中。面向客户主要有以下特点：

- 工作环境封闭

  专用机房，封闭环境，有物理防护（防火墙），有安全管理措施（认证鉴权等），近端接触有难度。
- 网络安全隔离

  与外网隔离，外部访问、跨数据中心连接需通过专用VPN与防火墙，保证网络传输安全。
- 不涉及隐私数据

  HPC处理各行业特有的科学计算数据，业务处理过程中不涉及隐私数据。
- 安全需求不强

  因HPC的运行环境与网络连接的特殊性，业内暂无特殊的安全需求。

3.1.2 超聚变内部安全要求与产品安全能力

虽然HPC行业较为特殊，业内无较多安全诉求，但是HPC解决方案面向全球市场各个行业销售。作为重要的计算系统，可能承担着国家级的机密计算任务，如果HPC集群受到攻击导致业务运转受损甚至宕机，带来的影响将是巨大的。另一方面，FusionOne HPC也考虑到外部环境日益复杂、网络安全问题日益突出，需要防止误操作或恶意攻击带来的安全问题，所以HPC解决方案特别是软件中间件（调度器、集群管理、文件系统等），满足超聚变公司的网络安全要求，在不影响功能使用及用户体验的前提下，提供安全可靠的软硬件服务质量显得尤为重要。

 版权所有 © 超聚变数字技术有限公司

- 软件安全

  HPC软件栈包括集群管理软件、调度器、文件系统等，基于内外部可能的安全攻击和危险，全方位进行安全保护。主要的安全能力有用户认证与鉴权、关键数据传输加密、软件完整性保护、支持日志审计、密钥管理、证书管理、安全隔离与权限最小化及Web安全等。

- 硬件安全

  包括计算节点、存储节点、网络设备，提供从电路安全设计、硬件接口安全设计到设备物理防护、安全启动与可信度量等一系列安全能力。

- 安全配置

  HPC会向用户提供安全加固建议与指导，特别是HPC中非超聚变发布及部署的组件，如操作系统、数据库等，这些组件的安全虽然不由超聚变负责，但其属于HPC的一部分，出现安全问题或受到攻击，一样可能影响到整个系统。

3.1.3 FusionOne HPC 安全架构模型

根据X.805模型的分层思想，安全技术可以分为三个主要的层次：应用层安全、系统层安全、网络层安全，同时加上贯穿所有层次的重要安全因素——安全管理（包括技术和非技术的方法），组成了分层的安全架构模型。

应用层

在应用层，采用的安全技术包括账号和口令安全、认证安全、访问控制与授权、会话管理、密钥管理、证书管理、输入校验、Web页面输出编码、接口协议安全、日志保护等。

- 账号和口令安全
  - 对接第三方账号管理系统时，本身不维护账号信息，故只涉及OS账号和第三方和开源软件的默认账号管理。通过提供的《帐户一览表》和OS加固资料确保所有的账号都可被系统管理（用户可见、口令可配置等），并对不同类型的账号提出不同的使用要求。
  - 对系统可以管理的账号口令，提供口令复杂度检测功能，对第三方和开源软件的默认账号口令提供首次登录后修改的机制，并提示使用缺省口令的风险。
- 认证安全
  - 认证接口提供防暴力破解机制，当重复输入口令次数超过阈值时锁定登录的账号，同时为了避免基于账号的锁定方式会被利用来进行DoS攻击，采用了流控机制来限制认证请求数。
- 证书管理
  - 使用OpenSSL进行统一的证书签发、吊销、维护，确保全网身份信息的一致性和有效性。
  - 使用X.509 v3格式的证书，证书私钥文件加密存储，并提供证书到期检查和更新机制。
- 秘钥管理

  密钥的安全管理对于整个系统的安全性至关重要，密钥管理的安全性直接影响受密钥保护的产品资产安全性。如果使用不恰当的密钥管理方式，强密码算法也无法保证系统的安全。密钥管理的核心诉求是保证密钥的机密性和完整性，同时降低密钥泄露引起的安全风险，中间件采用业界公认的组件化的密钥管理方案KMC，对密钥执行分层以及全生命周期管理，提供安全防护能力。

 版权所有 © 超聚变数字技术有限公司

  - 工作密钥：工作密钥对本地保存的敏感数据和需要在不安全信道上传输的数据提供机密性、完整性保护，还可提供认证和签名等密码学服务。工作密钥直接被上层应用程序所使用，包括存储加密密钥、会话密钥、MAC密钥、签名密钥等。
  - 密钥加密密钥（KEK）：每个KEK都各自对上层工作密钥提供机密性保护，其自身受到根密钥的保护。对于较为简单、安全等级要求不高的密码应用系统，密钥加密密钥的职能可以直接由根密钥兼任。
  - 根密钥：位于密钥管理分层结构的最底端，用于对上层所有密钥（如多个密钥加密密钥）的机密性进行保护。
- 会话管理
  - 对用户正在访问的页面，服务器需要判断会话是否已认证，并且用户是否有权限访问。
  - 通过服务器端认证后，服务器端使用安全随机数算法生成维持当前连接的会话标识。
  - 用户通过认证后，更换会话标识，以防止会话固定攻击。并且会话标识要安全传输（HTTPS、SSL或IPSec）以防止会话劫持情况的发生。
  - 提供会话超时机制，避免攻击者通过猜测或暴力破解的方式获取有效的会话标识来模拟用户和服务器进行通信。
  - 对B/S应用，用于维持会话的Cookie设置正确的属性，避免可能造成信息泄露的风险。
- 访问控制与授权
  - 对高低风险模块在安全设计上进行隔离，通过不同的进程实现。
  - 为了确保不同进程之间权限隔离，对外提供不同网络服务的进程使用不同的UID。
  - 应用权限最小化，确保运行软件程序的账号要尽可能是OS低权限账号，使用业务功能所必需的最小权限。
- 堆栈保护和防漏洞利用

  为了防止缓冲区溢出、运行库注入和逆向分析等漏洞利用技术，需要针对编译器和操作系统提供适当的安全编译选项来加固软件的安全性，满足公司的安全编译选项要求。

系统层

系统层安全主要是指为业务系统提供运行支持的操作系统、数据库系统及LDAP、Web等应用服务器的安全保护。

系统层安全的保护包括最小化安装、最小服务限制、帐户管理、口令策略、安全审计、认证与授权保护、内核参数安全调整、配置安全工具（如SSH、lsof等）、补丁管理、病毒保护、备份恢复等多方面的内容。

操作系统安全：

- HPC产品不交付OS，用户可以选用安全稳定的CentOS版本（兼容OS版本见版本配套表），通过使用可靠稳定的OS，可以确保应用程序运行在安全的环境中。
- 操作系统供应商在官网上会定期测试、发布操作系统补丁，客户应该及时升级补丁以防安全漏洞被攻击者利用。
- 操作系统在安装部署前需经业界主流漏洞扫描软件扫描，确保不存在高风险级别的漏洞。

 版权所有 © 超聚变数字技术有限公司

- 操作系统安装部署后应该进行安全加固，关闭不安全的协议和非业务必须的端口、开启系统漏洞防御机制等。

数据库安全：

- 数据库使用最新版本，确保没有高级别的安全漏洞，以解决基本的安全问题。
- 数据库系统补丁需要定期进行更新，以修补已知漏洞。
- 对数据库系统进行安全配置，包括：目录与文件权限、参数设置、用户Profile配置、用户权限配置、数据库网络安全等。
- 数据库系统经业界主流漏洞扫描软件扫描后，高风险级别的漏洞在出厂前处理。
- 对数据库进行安全加固。

应用服务安全：

对要集成使用的客户LDAP系统提供加固指导。

网络层

网络层安全就是通过采用一系列安全措施，使得网络系统得到应有的安全保护，为在该网络平台上运行的业务系统提供应用的支持，包括一切访问网络资源或使用网络服务相关的安全保护。HPC系统通常部署在私有局域网络，在标准组网环境中，会部署防火墙来防御外部攻击。为了提升整个服务的高可用性，通常会对关键性服务采用双机主备机制。

网络层安全技术包括网络拓扑安全设计、网络设备保护、网络隔离、网络边界保护（如防火墙）、网络安全检测（如IDS）、网络数据加密（如VPN）、网络安全扫描和网络安全管理等多个方面。

- 在业务网络中引入防火墙，通过安全域划分，合理的ACL策略配置进行访问控制。
- 设备的管理平面和近端维护终端、网管维护终端间传输的管理流量采用加密传输，使用合适的安全传输协议（如HTTPS等）。
- 使用SSHv2、TLS1.2等安全协议，对系统进行访问和维护，同时支持关闭不安全协议（FTP、Telnet等）。
- 提供完整的端口通信矩阵，明确标识哪些端口系统需要使用，便于用户更好地管理网络。
- 对非系统运行和维护所必需开启的端口，必须默认关闭，将攻击面降低到最小。
- 跨越信任网络的敏感数据加密传输，防止传输过程中泄露。
- 跨越信任网络传输数据时，实现敏感数据（如口令、密钥等）的完整性校验，防止敏感数据传输过程中被篡改。
- 对于系统所使用的中间件，避免使用默认的端口，并且对外提供服务的都配置了合适的防火墙策略。

管理层

安全管理是指对所有计算机网络应用体系中各个方面的安全技术和产品进行统一的管理和协调，进而从整体上提高计算机系统的安全保护能力。

安全管理包括两个方面的内容，一是从技术上建立高效的管理平台，协调各个安全技术和产品的统一管理，为实现安全策略和度量安全效果提供便利；另一方面是非技术方面，如人员管理、制度管理和安全策略规范等方面的内容，以这些措施来提高产品的安全能力。

 版权所有 © 超聚变数字技术有限公司

- 账号与权限管理
  - 每人每账号实现。系统账号按角色分权设置，实现账号分级管理，每人每账号授权，消除共享账号。
  - 清除多余账号与组。清除操作系统与数据库多余的账号与组（含默认）。
  - 本系统不提供用户密码、口令等管理能力。OS系统帐户密码策略受限于OS系统，第三方帐户密码策略如口令长度、口令复杂性，受限于第三方系统。
- 日志审计
  - 支持对系统有影响的操作记录日志，包括用户相关的操作（登录、注销、锁定、解锁、账号管理等）和用户登录后所进行的非查询类操作。
- 软件完整性保护
  - 对外发布的软件包采用大包签名进行完整性保护。
  - 在产品版本软件构建阶段进行CMS内层签名，并在升级阶段进行完整性校验。
- 提供完整的安全资料，便于客户运维人员进行安全相关运维操作。
- 通过合适的安全策略、标准、过程和补丁管理等来规避系统攻击带来的风险。
- 对系统使用者进行安全意识和赋能培训，提升团队整体安全素质和能力。

3.2 可靠

- 分布式设计

  HPC采用分布式管理结构，计算节点由管理节点进行管理，单个计算节点故障不会影响整个系统的正常运行。
- 主备冗余设计
  - HPC解决方案的管理节点硬件采用主备用、冗余配置等可靠性设计方法，确保了硬件系统的可靠性。
  - 集群管理软件采用主备模式冗余技术，在正常情况下，主用节点处理业务请求，当单个节点出现故障时，另一节点还可继续提供服务，业务不会中断。
- 网络路径全冗余

  HPC解决方案通过在网络各层采用不同的冗余技术来保证全网连接的可靠性。
  - 在服务器上，通过采用多网卡绑定，避免单个网卡故障引发的业务中断。
  - 在网络接入层、汇聚层、核心层，通过采用交换机堆叠或集群技术，来保证网络路径的冗余。

 版权所有 © 超聚变数字技术有限公司

4 服务与支持

4.1 实施服务
4.2 超凡服务
4.3 SNS服务

4.1 实施服务

4.1.1 服务概要

HPC（高性能计算）专业服务交付涉及到和L1基础设施的配合，L2的硬件安装及连接，软件的安装配置等，需要专业的HPC实施服务来完成以上工作。超聚变将提供以下服务内容：

- 客户环境调研
- HPC方案实施方案设计
- 项目交付
- 高性能计算调测&性能调优
- 项目验收&知识移交

4.1.2 目标

随着HPC中标项目日益增多，项目交付的工作量和压力也越来越大。为了保障项目的高效率、高质量交付。在此梳理了整个HPC项目交付的流程、关键节点以及详细的操作步骤。

4.1.3 服务内容

HPC解决方案实施服务属于超聚变专业服务内容，超聚变提供的实施内容包含以下几种：

- 硬件安装
- 设备上电
- 组网连线

 版权所有 © 超聚变数字技术有限公司

- 硬件配置
- 集群管理软件安装
- IB网络配置
- 中间件安装
- 集群环境一致性检查
- 集群性能验收测试

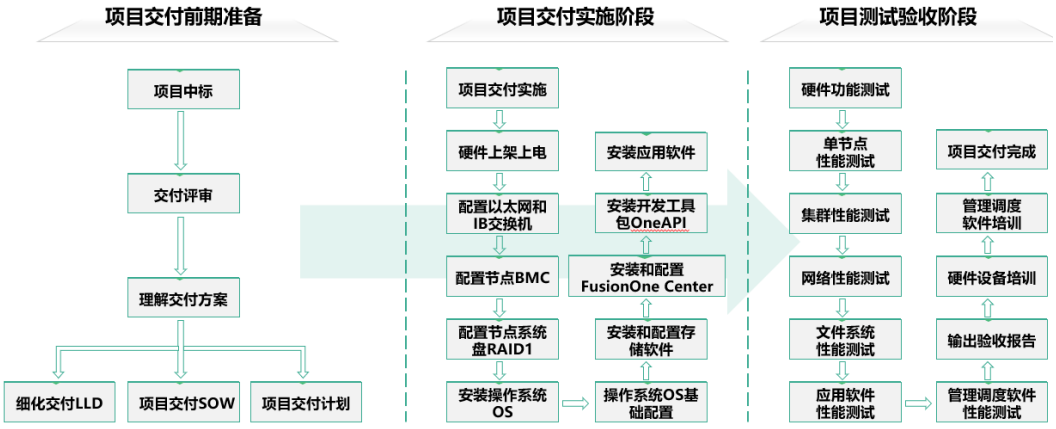
4.1.4 项目交付流程

整个HPC项目交付分为是三个阶段：

第一阶段：交付前期准备，交付工程师从项目中标后开始介入，参与项目方案技术评审，理解学习项目技术方案、了解项目的交付中可能的风险以及关键技术点。完全理解项目方案后输出项目的LLD、交付SOW以及初步的项目交付计划，并联系相关工程师进行评审。

第二阶段：项目交付阶段，设备到达客户现场后，指导本地服务工程师进行设备的拆装、上架，按照LLD进行组网连线、打标签，完成硬件的配置和信息收集，配合集群管理软件厂商部署集群调度软件，配置MPI环境、安装驱动等。软件部署完成后，检查集群配置信息是否一致。配合软件厂商完成应用软件的部署以及应用集成到集群管理软件中。

第三阶段：项目测试验收阶段，整个集群安装调试完成后，与客户一同对集群进行验收测试，包括基础的硬件功能测试、单节点性能测试、集群性能测试、网络性能测试、文件系统性能测试、应用软件性能测试以及集群管理软件功能测试，验收测试通过后。准备对客户进行培训，包括硬件设备的维护培训、集群管理调度软件培训、应用集成培训，实际项目中根据客户的需求进行灵活增减。



4.2 超凡服务

4.2.1 服务概要

Fusion-Care超凡维保服务是超聚变为购买其产品的客户专门打造并提供的高级维护保障服务解决方案，助客户维护更加高效、稳定的网络环境，提高网络生产力。

版权所有 © 超聚变数字技术有限公司

4.2.2 服务内容

4.2.2.1 保修技术支持服务

设备保修服务是超聚变为了满足客户对网络安全服务需求而提供的，旨在保障客户网上设备的稳定、高效运行，提高客户维护人员技术水平的服务方案。

保修服务包括：咨询服务、电话支持服务、远程技术支持服务、现场技术支持服务、紧急故障排除服务、软件补丁服务、故障件维修服务、紧急备件支持服务。

保修期内，超聚变保修技术服务的服务内容和界面是统一的，保修服务方案、内容、界面相同。

4.2.2.2 介质保留服务

超聚变介质保留服务是为购买超聚变服务器产品的客户专门打造并提供的专项服务解决方案；通过该服务，客户在获取更换的硬盘和（或）内存时可选择保留故障硬盘和（或）内存，以进一步满足客户提升信息安全等级、保护机密、私有信息安全的需求。

- 硬盘介质保留服务：硬盘介质保留服务是指硬盘发生故障后客户可以保留该介质，并根据自身的数据安全策略进行处理。
- 内存介质保留服务：内存介质保留服务是指内存发生故障后客户可以保留该介质，并根据自身的数据安全策略进行处理。
- 整机部件介质保留服务：整机部件介质保留服务是指服务器存储、内存、主板等所有部件发生故障后客户可以保留该部件，并根据自身的数据安全策略进行处理。

4.2.3 超聚变服务理念

- 实现客户满意

树立以客户为中心的工作作风，强化服务意识和服务技能，以优质服务切实保障产品运行质量，赢得客户满意。

- 追求服务领先

不断完善服务内容，追求服务的专业化、标准化和多元化。注重主动服务和个性化服务，塑造优质服务品牌，实现业界领先。

- 促进持久双赢

关注客户系统的整体效能，抢先一步发现客户价值提升点，不断提升对系统的完整支撑和持久保障，借助多方位的维护合作，促进与客户的双赢发展。

4.2.4 技术支持与服务目标

超聚变公司秉承“为用户服务是超聚变存在的唯一理由，用户需求是超聚变发展的原动力”的经营理念，奉行用户至上和诚信的企业文化，贴近用户，持续追求用户满意度的最大化，实现与用户的长期共赢。

- 为用户提供全方位的快捷服务与支持。
- 保障用户网络的运行质量，提高用户网络的整体效能。
- 帮助用户迅速解决网络问题。

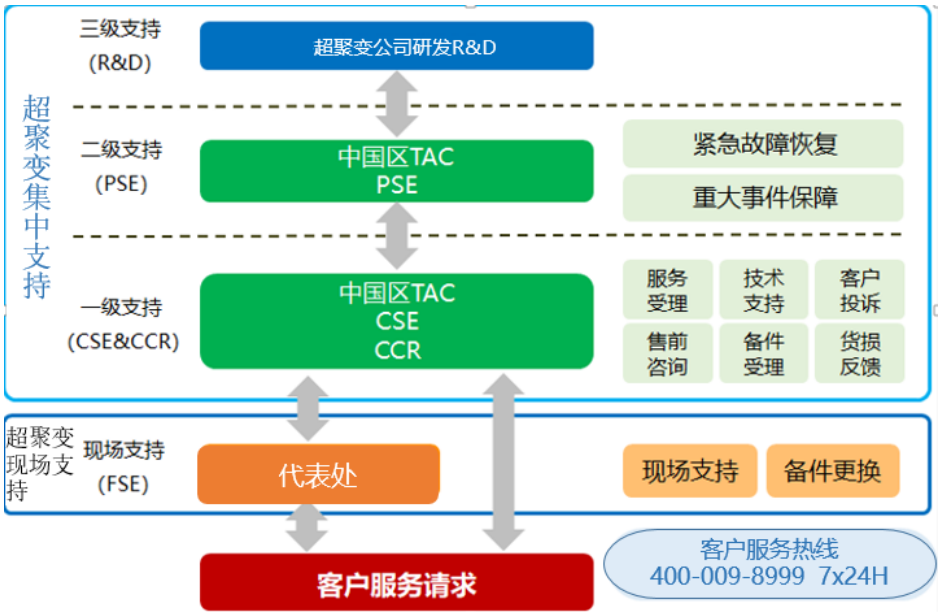
版权所有 © 超聚变数字技术有限公司

- 帮助用户树立网络竞争优势,优化网络性能，增加用户业务收入。
- 协助用户培养优秀维护人员。

“技术精湛、工程优质、响应及时、服务周到” ，通过不懈努力，追求卓越服务，协助我们的用户不断提高网络运行质量，降低运营维护成本，赢得最终用户的满意。

4.2.5 超聚变技术支持体系

超聚变具备完善的三级技术支持服务响应体系，设有具备7*24小时支持能力全国技术支持中心。超聚变企业客户服务中心电话为400-009-8999，是售前咨询、技术支持、备件受理、服务回访和客户投诉等业务的集中处理平台，实现了客户联系界面的统一、服务过程与业务数据的集成化，可为客户提供优质高效的一站式服务支持。



超聚变三级技术支持体系

超聚变的技术支持体系分为代表处/业务部、公司总部技术支持中心和研发R&D三级。代表处作为前端平台均配备维修车辆及先进的测试仪器，建有备品备件库，可以快速响应客户的服务请求，以专业化的装备随时为客户提供技术支援。公司总部和研发则作为后援平台为代表处提供更高一级技术支持与服务。各级技术支援部门统一管理，利用先进的网络技术达到信息和资源的共享，为客户提供主动服务和被动服务相结合的各项服务。

- 一级支持

  代表处/业务部为服务体系的前端平台，作为现场维护工程师的派出机构（一线），负责第一时间向客户提供服务，负责日常技术支持、技术交流、服务交付、问题处理、备件更换。37个（国内大陆）技术支援中心是服务体系的基本平台，每个技术支援中心均配备维修车辆及先进的测试仪器，建有备品备件库，通过INTERNET、PSTN互连，可以快速响应客户需求。代表处/业务部配置有服务经理和强大的售前/售后工程师技术团队，服务经理直接贴近客户，是最终客户满意度的责任人和客户需求的代言人，负责从客户角度审视超聚变服务的问题，组织专家分析网络问题并实施网络优化，有权调动公司一切资源满足客户需求。
- 二级支持

  超聚变总部全国技术支持中心，位于四川成都，是服务体系的管理平台（二线），面向政企客户提供7x24x365的不间断的技术支持服务。配置资深产品及解

 版权所有 © 超聚变数字技术有限公司

决方案专家团队和TAC实验室，负责疑难问题分析、定位和处理，为一线提供技术支撑。提供网络高级技术支持与服务、紧急故障处理、重大事件保障。通过问题管理系统、培训认证系统、备件管理系统等技术支持管理，快速响应客户的服务需求。

- 三级支持

  研发R&D是服务体系的后援平台，是产品的研发中心，可提供代码级的问题分析及定位。超聚变在国内深圳、北京、上海、南京、武汉、成都、西安、杭州、苏州等地设置有研发机构和研发实验室，是本服务体系的后援平台（三线），提供疑难问题的分析、处理和解决方案，并支撑二线、一线进行问题修复。设立维护优化部直接面向客户需求，负责远程故障解决及全网优化工作。超聚变建立技术支援与研发R&D的双向考核机制，确保在产品可服务性、可操作性的实现，确保客户问题的及时、有效处理。

对于通过电话支持服务不能解决的问题，超聚变在征得客户同意后，在协议约定的服务等级规定的响应时间内，通过远程终端登陆到客户申告问题的系统中进行调查和收集数据，然后进行故障诊断，查找故障出现的原因，进行相应的处理，指导现场维护人员排除设备故障。

4.3 SNS 服务

4.3.1 服务概要

超聚变SNS（软件订阅与保障基础）服务是超聚变为购买其产品的客户提供的软件基础维护保障服务解决方案。该服务方案主要包括“软件支持”和“软件订阅”两部分内容，提供软件产品的远程问题处理和软件更新/升级许可等服务，以帮助您实现软件在生命周期内的长期价值。适用于超聚变HPC解决方案中转售的相关软件。

4.3.2 服务内容

远程问题处理

超聚变工程师在接到软件故障申报后，将首先进行远程故障分析与处理，及时排除故障。远程问题处理包括电话支持和远程接入。

- 电话支持

  超聚变接到服务请求后，将在服务等级规定的响应时间内通过电话支持进行响应，帮助您对问题进行分析、诊断以及定位，提供解决问题的方案，并引导您实施。

- 远程接入

  对于通过电话支持服务不能解决的故障或问题，超聚变在征得您的同意后，通过远程终端登录到故障产品中调查和收集数据，分析故障原因，提出解决方案，指导您实施，必要时可以提供远程操作。

版本更新和升级

超聚变向您提供软件版本更新和升级下载许可，主要包含软件补丁、软件更新、软件维护版本、软件新特性和软件升级版本，以确保您所购买的软件始终处于最优性能，满足更多应用场景和更高业务需求，持续为您带来收益。

 版权所有 © 超聚变数字技术有限公司

在线自助服务

超聚变公司网站提供了大量的产品和技术资料，如产品手册、配置指南、组网案例、维护经验汇总等，通过为您开通网站相应权限，使您可以访问公司网站并下载相关资料，及时掌握最新的维护经验和技巧、获得最新的产品知识；有权登录网站自助服务工具，便捷快速的获取维护相关服务支持。

超聚变公司技术支持网站：**https://support.xfusion.com/**

版权所有 © 超聚变数字技术有限公司