



Lenovo 联想 + intel.

联想高性能计算和人工智能医疗行业 生命科学解决方案 白皮书



售前专线 400-819-6776
售后专线 400-106-8888

©2024 Lenovo. 保留所有权利。

供货情况：产品、价格、规格和供货情况可能发生变化，恕不另行通知。联想不对图片或排版错误承担责任。保修：如需获取适用保修的副本，请访问官方网站，对于第三方产品或服务，联想不作任何声明或担保。商标：Lenovo、Lenovo徽标、ThinkSystem是联想的商标或注册商标。英特尔、英特尔标识、至强和Xeon Inside是英特尔公司在美国和其他国家的商标。其他公司、产品和服务名称可能是其他公司的商标或服务标记。

目录

CONTENTS

01	联想高性能计算	01
02	联想生命科学行业解决方案	07
	联想生命科学 HPC 解决方案	07
	联想生命科学人工智能解决方案	42
03	联想高性能计算产品及特点介绍	45
	联想高性能计算硬件和环境配套	45
	联想高性能计算软件	72
04	联想高性能计算集群实施服务	83
05	联想生命科学客户案例	87
	某医院组学和深度影像学习平台	87
	北京大学高性能计算平台“未名生科一号”	93

“

近几年，为了推动生物医药产业的快速发展，加快构建生物产业链，国务院、国家发改委等部门相继发布了多项生物医药产业促进政策，支持生物技术药、化学药、中医药等细分领域的产品研发、成果转化，以及公共服务平台的建设等等。2022 年，工业和信息化部、国家发展改革委、科技部等九部门联合发布的《“十四五”医药工业发展规划》明确提出，国家将推动医药制造能力系统升级，以新一代信息技术赋能医药研发。

同时，随着第三代分子测序技术的发展，数据量快速增长，生物医学数据从 PB 组学时代迈入 EB 多维度大数据时代。从 CADD 到 AIDD，通过高性能计算资源进行虚拟药物筛选、分子对接、蛋白结构预测等工作，可有效缩短新药研发周期，降低科研成本。高性能计算正在影响着生命科学各研究领域的发展。

为了支撑科研人员的数据分析和计算需求，我国生命科学领域的高性能计算基础设施快速发展，且已初具规模，但生命科学领域企业应用高性能计算时依然存在算力、弹性、存储吞吐量不足，运维管理和成本大等挑战。如何降本增效，提高高性能计算的性价比，成为企业高度关注的问题。

联想集团作为全球高性能计算机（HPC）TOP500 和中国高性能计算机 TOP100 排行榜中领先的企业，不仅为生命科学各领域海量数据研究提供用于基因组学、蛋白质组学、药物研发等领域的超级计算机和高性能服务器以及大容量存储服务基础设施，也围绕高性能服务器基础技术打造了齐全的产品线和严格的质量控制体系，进而为生命科学领域提供了一套覆盖研发、数据存储、生产及认证、销售等环节的完整技术链。

当前，依托高性能计算平台，联想集团已经在全球为生命科学、基因测序等多个领域提供了成熟的高性能计算解决方案；联想集团交付的高性能计算集群，正在为研究机构、高等院校和企业的业务创新发挥着重要作用；针对医疗机构快速分析、共享海量元数据的需求，联想集团个性化构建了高性能的 IT 基础架构，提高医疗系统的性能、可靠性与灵活性，确保工作人员高效获取所需的科学计算资源，并轻松共享关键信息，从而节省了大量的时间和资源。

在以算力为基础数字经济时代，联想集团将基于自身传承了 30 余年的强大技术基因，持续升级算力基础架构，发展算力服务，继续致力于创新和深化在生命科学等领域的高性能计算解决方案，为生命科学行业打造“最强数字大脑”，推动行业快速发展。

联想集团政府纵队总经理 | 傅 强

联想中国基础设施业务群
高性能计算与人工智能产品营销主任技术顾问 | 郝常杰



联想高性能计算

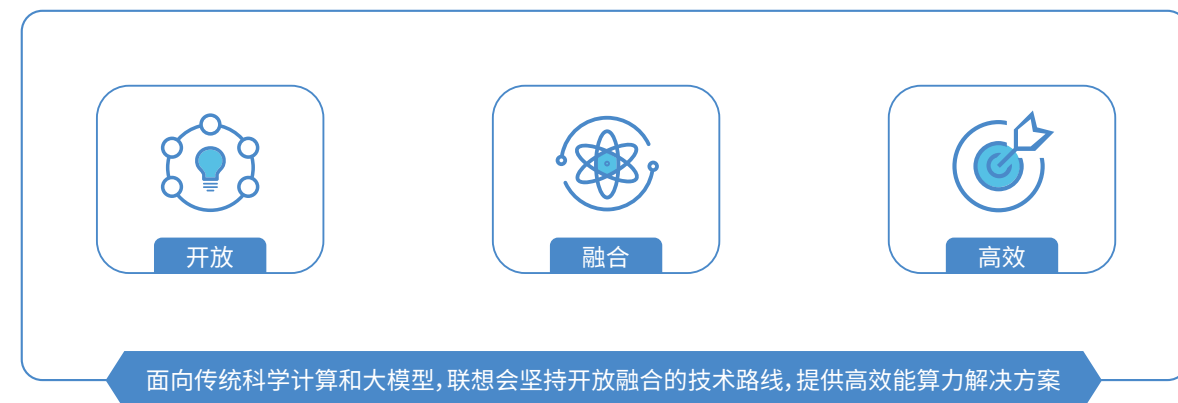
高性能计算多年来一直是科技综合实力竞争的至高点，也在一定程度上反映了各大服务器厂商系统研发方面的实力。作为行业的技术领先者，联想集团在这一领域积累了长达 20 多年的丰厚经验，并在关键技术领域不断创新，取得大量里程碑式的成果。

联想从 2001 年进军高性能服务器领域，是最早针对高等院校和科研院所进行产品开发与市场拓展的厂商，并在市场中一直处于技术领先地位。截止 2022 年，先后为数万个用户成功实施了高性能集群。曾经两次承担了中国科学院网络计算中心主节点的建设任务，并且成功地与威廉姆斯车队进行合作，成为国产品牌中最早将高性能业务拓展到海外的企业。

联想顺应国际主流技术发展趋势，以市场需求为驱动，吸收国内外最新技术成果，进行了大量创新性研发，突破包括系统设计与优化、系统基础架构、系统软件等在内的一大批高性能服务器的关键核心技术，开发出一系列可扩展、易管理、好使用、稳定可靠的高性能服务器产品，并配备可满足用户个性化需求的行业解决方案，提供从系统层到应用软件层的全面解决方案和技术服务。

联想高性能计算一贯秉承开放，融合和高效的理念，并以此来服务高性能计算各行各业，成就高性能计算客户。

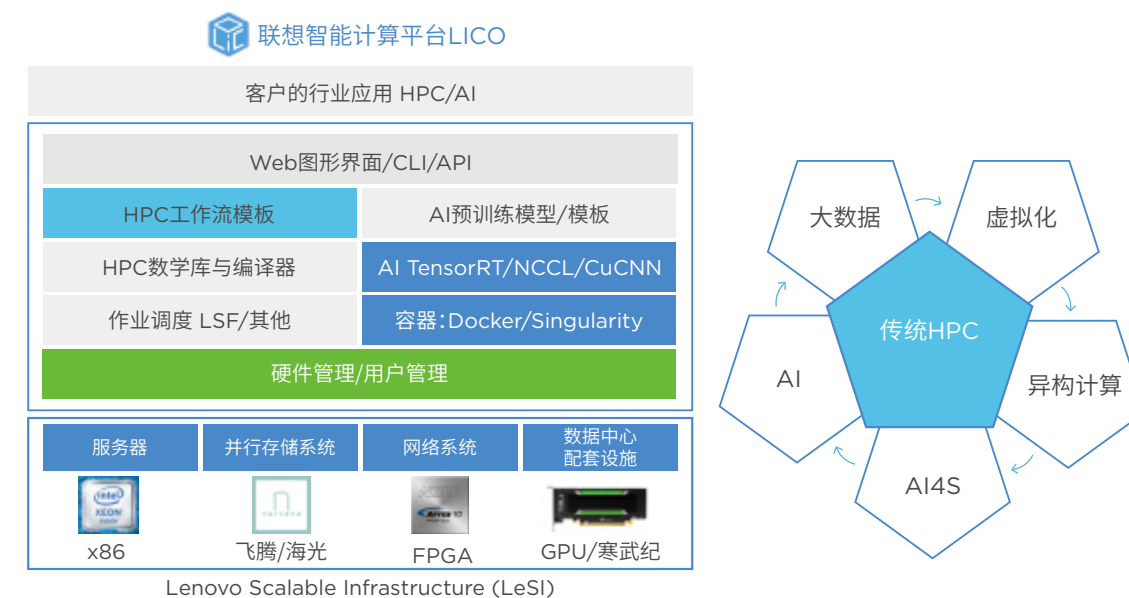
联想算力发展愿景



开放



融合



高效



2002 年 7 月，联想研制成功“深腾 1800”万亿次集群系统，安装在中科院数学与系统科学研究院。这是世界上第一个实际速度超过 1 万亿次的大规模集群系统。曾入选新华社 2002 年中国十大新闻及两院院士评选的 2002 年中国十大科技进展，并荣获 2004 年国家科技进步二等奖。2002 年末，另一套深腾 1800 大规模集群系统安装在中科院大气物理所国家重点实验室。2002 年 12 月 30 日，联想深腾 1800 中标大庆油田，使该油田第一次在国内实现三维叠前深度偏移地震资料处理。



2003 年，联想成功研制“国家网络主节点—联想深腾 6800 超级计算机”，安装在中科院计算机网络信息中心。这是当时世界上 Linpack 效率（78.5%）最高的高端通用计算机，其组合查询性能名列当时所有大型服务器的第四位，其典型应用 MM5 的测试结果在 2004 年 3 月列世界所有超级计算机的第一位。该机荣获 2005 年国家科学技术进步二等奖、2005 年国家重点新产品奖、2004 年信息产业重大技术发明奖。联想深腾 6800 自 2004 年初在网络中心对外服务以来，一直 7 × 24 小时稳定运行，在双星计划、气候模式计算、油藏模拟、材料科学计算、流体力学计算等领域取得了 150 多项重要计算成果。

联想深腾系列高性能计算机成为最早进入世界 TOP500 的一批国产计算机，分列当时世界 TOP500 的第 14、43、98 和 299 名。这是一个历史性的突破，联想深腾系列高性能计算机已成为国际知名国内主流的品牌。联想在推动高性能技术产业化方面取得了突破性进展，联想的高性能计算机广泛应用于许多关键领域，在国民经济和社会发展中发挥重要作用。

目前，集群已成为世界高性能计算机体系结构的主流，联想深腾 1800、深腾 6800 和深腾 7000 为这一趋势的形成做出了重要贡献。2002 年 8 月初，世界上主流并行编程环境 MPI-ch 的发明人、美国阿贡实验室 William Cropps 参观联想深腾 1800 后写道：“We see the future of clustering computing”。

联想在高性能服务器基础技术方面有着长期的积累，有齐全的产品线和严格的质量控制体系，为高性能计算机的研制和生产奠定了坚实的基础。

在产品设计上，联想坚持用户导向的原则，同时结合对新技术的深入理解和消化吸收，始终遵循模块化设计思想，在充分综合考虑各模块精密配合和整机系统合理整合的基础上，先设计出最佳性价比、最稳定的产品方案，然后对方案进行工程计算仿真，同时不断地结合验证性实验，最终才形成可行的开发方案，从而保证为用户在最短的时间里开发出最贴近的具有竞争力的产品。

在研究开发上，联想建立了与国际接轨的两级研发体系，即公司级研发平台和各事业部研发中心。公司级研发平台由联想研究院、软件中心、板卡中心和工业设计中心组成。事业部研发中心隶属于各事业部，直接承担具体的专项技术开发工作。联想在高性能服务器技术上已突破并拥有了自己的核心技术，拥有自主知识产权的系统设计与优化技术、系统监控技术、系统管理技术、高可用和负载均衡技术以及基础架构技术等关键技术，在高性能计算机系统技术方面已申请国家发明专利 85 项，其中，46 项已获授权。

在工程技术上，联想拥有针对服务器的部件及整机进行专业性测试的全套技术。部件测试包含外观、结构、功能、兼容性、可靠性、安全性、性能和环境 8 个方面的测试，以保证所有部件符合联想服务器技术特性和质量标准的要求，对服务器的核心部件如电源、内存，还建立了专业化的实验室，实现了部件的自动测试。如全球技术领先的自动电源测试实验室和自动内存测试实验室，国内功能最全面、技术最先进的系统测试实验室，以及高温实验室、电磁兼容检测实验室、噪音实验室、湿热实验室等等，所有产品需要在这些实验室中通过一系列的严格检测，只有通过了这一系列的严格检测的服务器产品，才可以顺利出厂，提供给客户。联想始终严格执行国际标准的质量控制体系，是国内最早通过 ISO9000 — 2000 版质量认证体系的服务器厂商。

在技术服务与方案上，联想服务器应用方案中心拥有雄厚的技术力量，在硬件平台、操作系统、数据库、软件、网络、存储、集群技术等方面有着多年的技术和经验积累，可以分别从不同的技术层面为用户提供有效的产品应用和方案支持服务。中心拥有先进的实验环境，包括方案集成实验室、性能评测实验室、数据中心、客户实验室四个部分，为用户提供方案开发、测试、方案移植、优化以及培训、咨询等服务，及时、快速、可靠地解决用户系统在使用过程中所遇到的技术问题，使客户的系统可以更加安全稳定地运行，以保障和促进客户业务的顺利开展并取得更大的成功。

2014 年 9 月 29 日，联想宣布完成对 IBM x86 业务的收购，从此，具有丰富的高性能计算方面经验的原 IBM x86 大批 HPC 专家加入了联想。

进入联想的原 IBM x86 部门熟悉应用的行业专家非常了解行业用户的需求，他们会针对行业的具体情况，与行业应用软件开发商密切配合，提供切实可行的解决方案，使得行业 HPC 用户的应用得以快速部署。

2016 年 7 月 1 日，从 ISC 2016 凯旋归来的联想集团再度吹响集结号，在北京隆重召开了以“开启 E 级计算新篇章”为主题的首届全球高性能计算峰会。本次大会联想正式发布了面向 E 级计算的高性能计算机系统深腾 x8800。

2018 年 6 月 30 日，联想正式升级高性能计算机系统为深腾 x9000，这是联想面向智能计算的统一平台，该平台涵盖传统高性能计算和人工智能技术。这是联想集团 3S 战略的具体方案落地，也是向人工智能方向迈出的一大步。

2021 年 12 月 14 日，上海交大校友、联想集团董事长兼 CEO 杨元庆捐建的高性能计算中心在上海交大李政道研究所揭牌启用。这套高性能计算设备被命名为“思源一号”，每秒运算可达 6 千万亿次，其算力为中国高校第 1，全球 HPC TOP500 第 132。

上海交大杨元庆科学计算中心坐落于张江科学城的李政道研究所实验楼内。李政道研究所实验楼于 2021 年 12 月 3 日正式启用，目前已聚集若干学术大师领衔的研究团队，吸引了一批科技创新活力的青年才俊，围绕粒子与核物理、天文与天体物理、量子基础科学等研究方向，开展了多个前沿领域原创性基础研究，实验平台建设正稳步推进。

这些领域的科学研究，恰恰最需要高性能计算支撑。作为全球第一的高性能计算解决方案提供商，截至 2022 年 11 月，联想连续十次蝉联全球高性能计算机 TOP500 榜单制造商份额榜首，九次登顶中国 HPC TOP100 年度数量总份额榜首。而联想的温水水冷技术，也成为业内绿色低碳的典范。

“思源一号”高性能计算机总计算力为 6 千万亿次 / 秒（双精度）。在今年 11 月发布的全球 TOP500 高性能计算机榜单上，“思源一号”的计算力排名第 206 位，算力超过哈佛、加拿大大学等国际名校。在中国高性能计算 TOP100 榜单上，它排名 15 位，中国高校高性能计算排名第一。此外，“思源一号”还具备超高计算密度与功耗密度、绿色节能、能源可循环利用三大技术特色。

联想 HPC 的发展历程



2023 年 11 月 14 日，TOP500 组织发布了最新的全球超级计算机 TOP500 榜单。榜单显示，在全球浮点运算性能最强的 500 台超级计算机中，联想制造的高性能计算入围 169 台的数量远超其他厂商，以 33.8% 的总体份额，再次名列全球高性能计算提供商份额第一名。

在注重高性能的同时，联想也非常重视绿色节能技术，联想以出众的节能技术打造了 GREEN500 榜单榜首的全球最节能高性能计算机。

2023 年 11 月的最新一届的全球最绿色的高性能计算机榜单，联想集团安装在美国纽约 Flatiron 研究所的 Henri 系统再次获得了 GREEN500 榜单头名，其能效达到了惊人的 65.40 Gflops/Watts，并三次蝉联榜首。此外，联想为韩国基础科学研究所打造的科学计算集群采用了联想 ThinkSystem SR675 V3 GPU 服务器，总核心数达到了 3936 个。该集群在最新 GREEN500 榜单中以 45.12 GFlops/Watts 的成绩一举跻身前十，并同样入围最新的 HPC TOP500 榜单。

目前，联想在全球部署的高性能计算机正在政府机构、商业公司、学术机构和医疗机构发挥力量。在科学计算、城市管理、人工智能开发等领域助力中国各行各业的研究，新 IT 赋能实体经济，助力构建新发展格局。

联想生命科学行业解决方案

联想生命科学 HPC 解决方案

生命科学作为 21 世纪最重要的科学分支之一，高性能计算在生命科学的研究和发展中起到非常重要的作用。从蛋白质结构中的应用辅助药物设计、疫苗研发，到生物信息学中的应用辅助基因数据处理与分析，再到提供精准医疗服务于肿瘤治疗、产前筛查等医疗技术，高性能计算 HPC 在生命科学研究中扮演着十分重要的角色。由于测序技术的飞速发展，人类发现的基因序列数目按照指数级增长，那么对于如此数量庞大的基因进行同源性搜寻、比对、分析、遗传发育分析等等，往往伴随着巨大的数据处理量和并行计算量。同时，由于生命科学的研究对象往往是蛋白质和 DNA 的大分子，对这些分子三维结构的预测，动力学特性、热力学特性、在生命过程中如何发生作用，这些科学问题也要借助于高性能计算机。所以高性能计算机在生命科学研究中，应用非常广泛，扮演着及其重要的角色。

同时，随着人工智能（Artificial Intelligence, AI）技术的兴起，高性能计算 HPC 中逐渐出现了 AI for Science，HPC+AI 极大地提升了生命科学领域的计算效率。随着生命科学行业的研究问题越来越复杂，加上 AI 的不断投入，构建 HPC+AI 平台成为满足生命科学行业日益暴涨的算力需求的基本基础架构。

生命科学行业涉及对微生物、动物和植物等所有生物进行研究的科学领域，同时包括生物伦理学等相关领域的考量。生命科学的研究对提高人类的生活品质有很大的助益。目前，生命科学已经在医疗、农业、保健、食品工业和制药等行业得到广泛应用，不断造福人类。虽然，当前生物学仍然是生物科学的中心，但随着分子生物学和生物技术的发展，生命科学已经成为一个专精化、多学科交叉的领域。

从全球范围内来看，21 世纪开始，全球生命科学领域的发展进入快车道，尤其是人类基因组计划的实施、干细胞研究的不断深入、克隆技术的不断发展等因素都将生命科学领域的发展推向了新的高度，与之相对应领域的研发投入也在不断增加。作为对科技信息技术有强依赖的典型代表，生命科学行业的药物研发和基因测序分析领域面临着计算资源紧缺、研发周期长等问题，为了解决这些问题，相关企业开始从传统 IT 转向上云，希望借助云计算平台的资源优势加速药物研发速度、提升测序数据分析效率，加快业务创新步伐，为用户创造更大价值。

新药的诞生通常需要经历药物发现、临床前研究、临床试验和审批等阶段后，最终才可以获批上市。在靶点发现、化合物合成等药物发现阶段以及化合物筛选等临床前研究阶段，往往需要借助高性能计算强大的计算能力才能加速研发过程辅助药物设计。

生命科学 HPC 应用分类

基因组学是指对生物体所有基因进行基因组作图、核苷酸序列比对分析、基因定位及基因功能分析的一门学科，而基因测序则是针对基因组学研究的检测分析技术。基因组学所对应的研究主要包括结构基因组学和功能基因组学两个方面。

结构基因组学代表基因组分析的早期阶段，一般以基因组测序为目的，得出生物样本的静态碱基序列，并构建生物体的基因图谱，一般基于高通量测序技术生成生物的碱基序列。

功能基因组学也称作后基因组学，基于结构基因组所提供的数据及信息，结合计算机科学，同时对多个基因片段进行系统性的分析研究，主要研究内容包括: 基因突变检测 (SNP, InDel)、基因表达分析及基因功能发现等。

基因测序的业务流程主要包括样本上机（测序仪）、测序文件生成、基因序列比对及结果分析（计算机），并将结果数据及报告交付至各科研医疗机构。其中，基因序列比对及分析环节极为耗时，涉及大量的生信领域专业软件，计算资源的算力性能及方案优化对生信研发效率起着至关重要的作用。



生命科学的主要软件

生物领域中的高性能计算最大的特点就是种类多。

- ① 生物分子的结构测定：Illumina 公司的 bcl2fastq，Genome Analys，HiSeq 2000，GAPipeline 等；ABI 公司的 Solid3，Solid4，Bioscope。电镜 EMAN，SPIDER；利用 Xray 方法测量用软件 CCP4、ARP/wARP，CNS 进行解析等；质谱仪 tandem(X!tandom) 等。
- ② 生物信息学类计算：最常用的有 BLAST，FASTA，HMMER，ClustalW，DNASTAR，PHYLIP，PAML，PAUP，T-Coffee，EMBOSS 等。
- ③ 分子动力学类计算：主要软件有 AMBER ，Charmm，NAMD，Gromacs 等，这类应用非常适合大规模并行。
- ④ 计算机辅助药物设计类计算：应用软件有 DOCK ，AutoDock，FlexX，Discovery Studio，ZDOCK，RDOCK，MORDOR 等。

主要生命科学学科的工作负载

学科	解决方案	数据/应用程序特点	主要应用程序
生物信息学-序列分析	搜索、对齐生物序列 (DNA和蛋白质) 并对其进行模式匹配	结构化数据整数占主导, 频率依赖性, 大量缓存和内存BW并非关键, 一些算法适合进行SIMD加速	NCBI BLAST, wuBLAST、ClustalW、HMMER、FASTA、Smith-Waterman
生物信息学-NGS	对齐并合并DNA短片段, 以重建原始序列	- 一些需要大量内存 - 许多为IO密集型任务 - 字符串分析和匹配算法	Phrap/phred、CAP3/PCAP、Velvet、ABYSS、SOAPdenovo、Newbler、MAQ、BOWTIE、BFAST、SOAP、BioScope、SAM工具、GATK、Bowtie、Bwa等二代测序的比对工具，Fastqc等数据预处理工具
生物化学-药物研发	筛选大型数据库, 查找具有所需生物活性的潜在药物	- 大多为浮点数据 - 计算密集型极高 - 高度并行任务	Dock、Autodock、GLIDE、FTDock、Ligandfit、Flexx
计算化学-分子模拟和量子力学	使用分子动力学和量子力学技术对生物分子进行建模	- 需要处理大量浮点数据 - 延迟至关重要 - 频率依赖性 - 可缩短到100秒以下	CHARMM/CHARMm、GROMACS、Desmond、AMBER、NAMD、Gaussian、GAMESS、Jaguar、NWCHEM
蛋白质组学	解读质谱分析数据, 将频谱与蛋白质数据库进行匹配	- 整数比较 - FFT频谱分析具有浮点数据 - 对通信要求不高	Mascot、Sequest、ProteinProspector、X!Tandem、OMSSA
结构生物学	用物理学方法, 配合生物化学和分子生物学方法研究生物大分子结构与功能的新学科	- 数据量大 - 计算密集型	Relio为典型代表(GPU优化)

生物信息学

近 20 年来，随着人类基因组计划的相继完成和各种模式生物的基因组计划的顺利实施，以及基因测序技术的不断进步，生命科学研究得到了巨大的发展。同时，海量的有关生物序列的数据不断涌现。这些数据具有丰富的内涵，其中蕴含着大量的人类尚且不知道的生物学奥秘。

生物信息学是研究生物信息的采集、处理、存储、传播、分析和解释等各方面的一门学科，与以观察和实验为主的传统生物学研究不同，它通过综合利用生物学、计算机科学和信息技术而揭示大量而复杂的生物数据所蕴含的生物学奥秘。生物信息学把 DNA 序列、蛋白质序列以及其它相关生物数据作为分析对象，力求揭示 DNA 编码区、蛋白质、RNA 基因以及其基因组中非编码序列的信息实质。

1. 测序技术介绍

DNA 测序技术是现代分子生物学研究中最常用的技术。自 1977 年第一代测序技术问世以来，经过三十多年的发展，DNA 测序技术取得重大进展，以高通量为特点的第二代测序技术逐渐走向市场，以单分子测序为特点第三代测序技术也已经出现，其分别在测序特点上占有不同的优势。测序技术的快速发展，使小型化 / 台式高通量测序仪成为现实，这意味着大规模基因组测序将不再是大型实验室或科研中心的专利，中小型实验室、公司、临床检验中心都将能够利用高通量测序技术快速高效的获取大量信息，进行科研或开发应用。

2. 第一代测序技术

20 世纪七十年代中期，Sanger 提出了经典的双脱氧核苷酸末端终止测序法，标志着第一代测序技术的形成。第一代测序技术完成人类基因组计划，花费了 30 亿美元巨资，耗时长达三年，测序成本高，测序速度比较慢。至今为止，应用最广泛的第一代 DNA 测序仪即是基于毛细管电泳和荧光标记技术的 3730 系列自动测序仪。

3. 第二代测序技术

随着人类基因组计划的完成，传统的测序方法已经不能满足深度测序和重复测序等大规模基因组测序的需求，这促使了以高通量为显著特征的第二代测序技术的诞生。第二代测序技术主要包括 454 公司的 GS FLX 测序平台、Illumina 公司的 Solexa Genome Analyzer 测序平台和 ABI 公司的 SOLiD 测序平台。

454 测序技术利用了焦磷酸测序原理。454 测序系统是第二代测序技术中第一个商业化运营的测序平台。其在 2005 年最早推出了第二代测序平台 Genome Sequence 20，完成支原体 Mycoplasm a genitalium 基因组测序。并在 2007 年推出性能更优的测序平台 GS FLX。2010 年秋，该公司自行研制的 GS Junior 测序仪上市，其在系统性能方面均得到提升。目前，GS Junior 测序仪平均读取长度长达 400bp，每次运行得到超过 35MB 高质量过滤后的数据，准确率达 99%，平均运行时间为 10 小时，更适合规模较小的实验室。与第二代测序平台相比，454 技术最大的优势在于较长的读取长度，使得后继的序列拼接工作更加高效、准确。但是，454 技术无法准确测量同聚物的长度，其技术的主要错误主要来源于核苷酸的插入或缺失。

Solexa 测序技术主要采用边合成边测序的方法 (SBS)。2009 年，Solex 推出了对读测序的方法，使得在技术层面上取得了进步。目前，新一代的 Illumina HiSeq 2000 测序仪的读取长度长达 2×100bp，每次运行能够得到大约 200Gbp 的数据，精确度达 99.5% 以上，使得在后续的序列拼接工作的计算量和难度上均有所增加。Solexa 技术在合成中每次只能添加一个 dNTP，很好的解决了同聚物长度的问题，其技术的主要错误主要来源是核苷酸的替换，其错误率大约在 1%-1.5% 之间。

SOLiD 技术利用了 DNA 连接酶测序的方法，通过连接反应进行测序。其基本原理是以四色荧光标记的寡核苷酸进行多次连接合成，取代传统的聚合酶连接反应。SOLiD 是 ABI 公司于 2007 年底推出的全新测序技术，目前已发展到 SOLiD 4 Plus，其读长达到 50bp，每次运行能够得到 80-100Gbp 的数据量。（sdarticle）最新研发出来的 5500xl solid 系统 (SOLiD4hp) 每次运行能够得到 240Gbp 数据量，其准确率达到 99.94%。然而，尽管新一代测序技术优势多，其局限性也不容忽视，测序速度提高了，但是测序产生的海量数据却为后续的分析与存储带来了巨大的挑战。

相对于 Sanger 测序技术，第二代测序技术具有成本低、高度并行化和通量高速度快的优点。但第二代测序仪又有各自的特点：

454: 读长较长（可达 400bp），但通量较小，比较适合用来测量较长的 DNA 片段或全新测序 (de novo sequencing)，例如：可用它测序个人基因组——已经测序出 Watson 的基因组（7.4 的覆盖率）和 Neanderthal 的 DNA 序列。

Solexa: 读长较短（35-50 bp），但通量大，机器运行一次就可产生 1.5 Gb，用 pair-ended 库测序数据量可达 3Gb。Solexa 多数情况下用于测序小片段的核苷酸序列，如 ChIP-seq、mRNA-seq 等等；其深度测序也可用于人的基因组序列，但其覆盖率一般比较高（>30 倍）。

SOLiD: 性能参数与 Solexa 接近，但通量略高，机器运行一次可产生 3～6 Gb 的数据，但需花 6～10 天（而 Solexa 运行一次花 3 天）。

4. 第三代测序技术

近期出现的 Helicos 公司的 Heliscope 单分子测序仪、Pacific Biosciences 公司的 SMRT 技术和 Oxford Nanopore Technologies 公司正在研究的纳米孔单分子技术，被认为是第三代测序技术。与前两代技术相比，他们最大的特点是单分子测序。其中，Heliscope 技术和 SMRT 技术利用荧光信号进行测序，而纳米孔单分子测序技术利用不同碱基产生的电信号进行测序。

Helicos 公司的 Heliscope 单分子测序仪基于边合成边测序的思想，将待测序列随机打断成小片段并在 3' 末端加上 Poly(A)，用末端转移酶在接头末端加上 Cy3 荧光标记。用小片段与表面带有寡聚 Poly(T) 的平板杂交。然后，加入 DNA 聚合酶和 Cy5 荧光标记的 dNTP 进行 DNA 合成反应，每一轮反应加一种 dNTP。将未参与合成的 dNTP 和 DNA 聚合酶洗脱，检测上一步记录的杂交位置上是否有荧光信号，如果有则说明该位置上结合了所加入的这种 dNTP。用化学试剂去掉荧光标记，以便进行下一轮反应。经过不断地重复合成、洗脱、成像、淬灭过程完成测序。Heliscope 的读取长度约为 30-35 bp，每个循环的数据产出量为 21-28Gb。值得注意的是，

在测序完成前，各小片段的测序进度不同。另外，类似于 454 技术，Heliscope 在面对同聚物时也会遇到一些困难。但这个问题并不会十分严重，因为同聚物的合成会导致荧光信号的减弱，可以根据这一点来推测同聚物的长度。此外，可以通过二次测序来提高 Heliscope 的准确度，即在第一次测序完成后，通过变性和洗脱移除 3' 末端带有 Poly(A) 的模板链，而第一次合成的链由于 5' 末端上有固定在平板上的寡聚 Poly(T)，因而不会被洗脱掉。第二次测序以第一次合成的链为模板，对其反义链进行测序。对 Heliscope 来说，由于在合成中可能掺有未标记的碱基，因此其最主要的错误来源是缺失。一次测序的缺失错误率约为 2-7%，二次测序的缺失错误率约为 0.2-1%。相比之下替换错误率很低，一次测序的替换错误率仅为 0.01-1%。总体来说，采用二次测序方法，Heliscope 可以实现目前测序技术中最低的替换错误率，即 0.001%。

Pacific Biosciences 公司的 SMRT 技术基于边合成边测序的思想，以 SMRT 芯片为测序载体进行测序反应。SMRT 芯片是一种带有很多 ZMW(zero-mode waveguides) 孔的厚度为 100 nm 的金属片。将 DNA 聚合酶、待测序列和不同荧光标记的 dNTP 放入 ZMW 孔的底部，进行合成反应。与其他技术不同的是，荧光标记的位置是磷酸基团而不是碱基。当一个 dNTP 被添加到合成链上的同时，它会进入 ZMW 孔的荧光信号检测区并在激光束的激发下发出荧光，根据荧光的种类就可以判定 dNTP 的种类。此外由于 dNTP 在荧光信号检测区停留的时间（毫秒级）与它进入和离开的时间（微秒级）相比会很长，所以信号强度会很大。其它未参与合成的 dNTP 由于没进入荧光型号检测区而不会发出荧光。在下一个 dNTP 被添加到合成链之前，这个 dNTP 的磷酸基团会被氟聚合物（fluoropolymer）切割并释放，荧光分子离开荧光信号检测区。SMRT 技术的测序速度很快，利用这种技术测序速度可以达到每秒 10 个 dNTP。

Oxford Nanopore Technologies 公司正在研究的纳米孔单分子技术是一种基于电信号测序的技术。他们设计了一种以 α- 溶血素为材料制作的纳米孔，在孔内共价结合有分子接头环糊精。用核酸外切酶切割 ssDNA 时，被切下来的单个碱基会落入纳米孔，并和纳米孔内的环糊精相互作用，短暂地影响流过纳米孔的电流强度，这种电流强度的变化幅度就成为每种碱基的特征。碱基在纳米孔内的平均停留时间是毫秒级的，它的解离速率常数与电压有关，180 mV 的电压就能够保证在电信号记录后将碱基从纳米孔中清除。纳米孔单分子技术的另一大特点是能够直接读取甲基化的胞嘧啶，而不像传统方法那样必须要用重亚硫酸盐（bisulfite）处理，这对于在基因组水平研究表观遗传相关现象提供了巨大的帮助。纳米孔单分子技术的准确率能达到 99.8%，而且一旦发现替换错误也能较容易地更改，因为 4 种碱基中的 2 种与另外 2 种的电信号差异很明显，因此只需在与检测到的信号相符的 2 种碱基中做出判断，就可修正错误。另外由于每次只测定一个核苷酸，因此该方法可以很容易地解决同聚物长度的测量问题。该技术尚处于研发阶段，目前面临的两大问题是寻找合适的外切酶载体以及承载纳米孔平台的材料。

5. 测序技术应用

目前，第三代测序技术尚处于研发阶段，第二代测序技术已经应用于基因组学研究的各个方面，因此主要介绍第二代测序技术在基因组测序及转录测序等方面的应用。



6. DNA 测序

全基因组测序

新一代高通量测序技术的发展，使得物种全基因组测序变得速度快、效率和准确率更高，越来越多的物种基因组信息相继公布。全基因组测序主要应用与基因组序列未知的物种，DNA 片段测序后，用生物信息学软件对序列进行拼接、组装，从而获得该物种的基因组序列图谱。

基因组重测序

重测序是指该物种基因组序列已被测序，有参考基因组序列的测序工作。第 2 代高通量测序技术目前应用最为广泛的领域就是对已知基因组物种进行重新测序，基因组结构变异、单核苷酸多态性、群体多态性、突变热点等重要信息都是通过重测序研究发现的。

宏基因组研究

宏基因组学（Meta-Genomics）测序是近年来提出的一种新概念，即不再进行分离，而是从整体上研究整个微生物种群结构的特征，研究对象从单一基因组发展到基因组集合，摆脱了对于传统基因组研究的物种限制，开辟了微生物群体，特别是不可培养微生物菌群基因组学研究的新路径。该类研究的具体操作方法是在提取微生物种群的总 DNA 后，制备整个种群的 DNA 文库，然后进行高通量的测序，从而在整体上对样品群落进行分析。与传统的微生物研究相比，宏基因组研究跳出了实验室培养的局限，真实地描述了大自然生态群落的复杂性和多样性，对于人类更好地了解微生物群落有着重要的意义。目前，欧盟推出的人类肠道宏基因组计划（MetaHIT），就是通过研究人类肠道中所有的微生物种类，为明确肠道微生物与肥胖等人类疾病的关系提供重要的理论依据。

7. RNA 测序

转录组测序

生物体最主要的调控方式就是转录水平的调控，而作为目前从全基因组水平研究基因表达的主流方法，建立在高通量测序技术基础上的转录组研究已逐渐取代了基因芯片技术。该方法对大量样品同时进行测序，可获得样品之间的表达差异，而同一样品进行深度测序时可以捕获低表达的基因。此外，还可以获得转录本表达丰度、转录发生位点、转录本 SNP、可变剪切等重要信息，因此，转录组测序越来越多地用于各种生物的差异表达基因的筛选及可变剪切的鉴定等方面。

小分子 RNA 测序

小分子 RNA 近几年受到了科学界的广泛关注，它是一类长约 20 ～ 30 个核苷酸的非编码 RNA 分子，其介导的转录后基因调控是生物体的一种新型基因调控机制，它在生物体的生长发育和适应外界各种环境胁迫的过程中起着非常重要的作用。传统的对小分子 RNA 的研究方法主要包括克隆测序法、正向遗传学筛选、芯片技术等，而高通量测序给小分子 RNA 研究拓展了新的思路。高通量测序技术既能捕捉真实存在的小 RNA，同时也能对

没有注释的小片段 RNA 进行预测，更重要的是通量的增加甚至可以捕捉体内一些低表达的小 RNA，在小 RNA 检测方面确实具有极大的优势。

8. 表观遗传学

甲基化测序

DNA 甲基化是可改变染色质结构、DNA 稳定性及 DNA 与蛋白质相互作用的一种重要基因调控方式，以达到控制基因表达的目的。高通量测序技术在检测全基因组范围内的甲基化位点方面也提出了高效的解决方案。其原理为：由于在哺乳动物中甲基化一般发生在 CpG 的胞嘧啶 5 位碳原子上，所以可通过特异性结合甲基化 DNA 的蛋白 MBD2b 或 5' - 甲基胞嘧啶抗体富集高甲基化的 DNA 片段，用高通量测序对富集到的 DNA 片段进行测序，从而检测全基因组范围内的甲基化位点。

转录因子结合位点测序

转录因子是通过与 DNA 特定区域相结合，开启或关闭基因的表达以达到调控基因表达目的的一种生物体内常见的调控蛋白。染色质免疫共沉淀技术（ChIP）也称结合位点分析法，是研究体内蛋白质（转录因子）与 DNA 相互作用的有力工具，通常用于转录结合位点或组蛋白特异性修饰位点的研究。将 ChIP 与第 2 代高通量测序技术相结合的 ChIP-Seq 技术，能够高效地在全基因组范围内检测与组蛋白、转录因子等互作的 DNA 区段。ChIP-Seq 的原理是，首先通过染色质免疫共沉淀技术（ChIP）特异性地富集目的蛋白结合的 DNA 片段，并对其进行纯化与文库构建，然后对富集得到的 DNA 片段进行高通量测序。研究人员通过将获得的数百万条序列标签精确定位到基因组上，从而获得全基因组范围内与组蛋白、转录因子等互作的 DNA 区段信息。

9. 测序数据处理

基本的数据处理和分析涉及：

第一步，对测序获取的短序列进行比对拼接。如果是重测序，可以用 bowtie 进行参考基因组比对，即匹配测序短片段在参考基因组上的位置；如果是对新物种进行从头 (De Novo) 测序，用 velvet 进行拼接，即利用测序短片段重构基因组序列。

第二步，比对拼接后，进行全基因组基因注释。包括基因组组分分析，SNP 分析，编码基因预测，重复序列注释，Non-coding RNA 基因注释，Micro RNA 基因注释等。如 SNP 分析可以用 MAQ。

第三步，对预测的基因进行功能 (GeneOntology, Pathway 等) 注释。可以用 InterproScan，WEGO。

第四步，比较基因组和分子进化分析。如快速进化 (Rapid Evolution) 分析，共线性分析 (SyntenyBlock)，基因家族分析等。常用的进化树分析软件如 MEGA。

这个过程中，突出的问题有：

① 软件选择难。对应某一功能有上百种软件可选，随着仪器的更新换代，数据格式的变化，同一款软件的算法不断升级；

② 分析效率不高。多为领域专家依赖脚本语言和库写成的软件，未考虑与硬件资源使用的匹配。基本少有优化，并行化，串行或多线程软件居多；

③ 分析流程中多软件衔接难。多数的高通量测序数据分析需几个软件配合完成，各软件通过脚本和大数据的重复读写（数据格式也需匹配）来协调。例如，比对之后做 SNP 检测，那么比对结果将作为 SNP 分析的输入；

④ 各软件资源使用特征差异大。例如，拼接软件需要大量的内存消耗，比对则是典型的数据密集计算。

除了各分析算法上的不断优化，当前业界突出的两方面进展表现在 workflow 系统和云计算的应用。比如 UCSC 开发的针对第二代测序数据分析的应用系统 Galaxy，Notre Dame 大学仿 makefile 开发的用来在集群、云和网格中执行大而复杂任务的工作流引擎 Makeflow；计算大规模 RNA-seq 数据集基因差异表达的云计算工具 Myrna，基于序列片段数据进行 SNPcalling 的 MapReduce 软件 Crossbow。

10. de novo 序列拼接

新一代基因测序技术所产生的序列片段具有序列短、高覆盖率、额外的双端信息等特点。大量的 overlap（重叠区域）的产生，增加了组装算法的计算量，初始测序数据的庞大对算法内存的要求随之增加。这就使得序列组装软件不仅要处理重叠区域，片段间隔，还要处理短片段。而且短片段通常使得前两个问题更加严重。de novo 序列组装问题就是对测序产生的大量长度为 50bp 到 100bp 的重叠序列片段进行比对合并，以重构源序列。而基因组源序列的长度长达几个 Gbp。因此，如果没有高效的序列组装软件，对大基因组测序后产生的数以亿计重叠序列片段进行组装几乎是不可能的。

目前广泛使用的拼接策略大致分为三类：Greedyextension，Overlap-Layout-Consensus 和 de Bruijn graph。

基于 Greedy-extension 的拼接策略

Greedy-extension 拼接策略从一个 read 出发，将其作为种子序列，搜索所有其前缀与种子序列的后缀具有超过阈值重叠（overlap）的 read，或选择具有最大重叠长度的 read 进行延伸，或通过投票选择最大可能的碱基进行延伸，然后将延伸的结果作为种子序列，迭代延伸来完成拼接。Greedy-extension 拼接策略是一种贪心策略，可能出现错误延伸。

使用 Greedy-extension 拼接策略具有代表性的拼接工具包括 SHARCGS，SSAKE，VCAKE 和 QSRA。在 SSAKE 中，首先从 hash 表中选取数量最多的 read 作为种子，然后采用绝对贪婪的方法，检索出与种子匹配的 read，并按一定规则拼接，同时补充相应的反向互补序列，逐步形成双链 config（即根据 read 间的 overlap，由 read 拼接成的片段重叠群），从而实现序列的扩展。当源数据的覆盖度很高时，SSAKE 可以从头拼接一些简单病毒的 DNA，但是拼接长度较短，而且对于真实测序中产生的错误，SSAKE 无能为力。

VCAKE 是对 SSAKE 的修订，它将所有与种子有 overlap 的 read 考虑到，并且每次只扩展一个碱基。VCAKE

虽能利用高覆盖度来处理测序产生的错误，但其也只适合拼接病毒和细菌这些较短较简单的基因组。

SHARCGS 延续了 VCAKE 对拼接不确定的处理，并加入了对测序错误的修复，从而有效提高了拼接质量。与 SSAKE 和 VCAKE 相比，它在内存、速度、准确度上都有所改善。但这 3 种算法核心相似，所有它们的整体性能差别不大，都只能应用在简单病毒、细菌基因组的拼接上，且较为耗时。

基于 Overlap-Layout-Consensus 的拼接策略

Overlap-Layout-Consensus(OLC) 拼接策略一般分为以下 3 个步骤：

Overlap，对所有的 read 进行两两比对，以获得可能存在的重叠信息，并建立重叠图，通常将每个 read 看作图上的一个节点，当 read A 的后缀和 read B 的前缀具有超过阈值的重叠时，A 和 B 建立连接；

Layout，分析重叠图，尽可能寻找贯穿重叠图上所有节点的合适路径，但在实际执行中由于测序错误或重复序列等原因，寻找的结果是多条路径，即多个 DNA 序列片段；

根据多个序列片段比对得到最后的结果序列，称作“Consensus”序列。

OLC 策略理想的情况是在重叠途中找到一条路径能够遍历所有的节点，即寻找图的 Hamilton 路径，是一个 NP 问题，因此使用这种策略的拼接工具通常时间复杂度较高。

目前典型的拼接工具包括：CABOG，Edena，Newbler 和 Shorty。Newbler 只适用于 Roche /454 测序仪产生较长 read 数据 (250-400bp)，并设计特定的过程用于处理 454 测序仪在均聚物位置长度不确定问题；CABOG，Newbler 和 Shorty 都在拼接过程中对测序错误和 repeat 进行了处理。CABOG 也只能对 100bp 长度以上的 read 进行拼接方能实现较好的拼接性能，它使用了一种称为“rocks and stones”的技术实现对测序错误的纠正；Shorty 和 Edena 都适用于短 read 数据的拼接，Edena 建立 transitively reduced 重叠图降低图的复杂度；Shorty 能通过 300-500bp 长度的配对末端 read 数据来估计拼接得到的序列片段间的距离，目前它更多地针对于拼接 ABI 测序仪的 SOLiD 类型数据。

基于 de Bruijn graph 的拼接策略

现在使用最多的是 De Bruijn 图算法,其代表软件包括 Euler, allpaths, velvet,idba.soapdenovo 和 abyss 等。该算法思想是“反直觉”法：在将 reads 组装前，将其切分为长度为 k 的子串，每一个节点 N 代表了一系列长度为 k 的子串，称为 k-mers。相邻的 k-mers 之间存在 N-1 个重叠的碱基。基因序列排布信息由依次重叠的 k-mers 最后一个碱基的读取顺序决定。N 节点上的基因序列的内容表示为 s(N)。由于序列的测定的方向性并不明确，为了保证序列组装的正确性，对于每一个 N 节点，都有一个反向的 N' 节点而对应，N 与 N' 是关联在一起的，所有对于 N 的操作同时也对 N' 有效。De Bruijn 图算法的目的在于得到没有分支的最大路径，基于此最大路径，得到 contigs。De Bruijn 图算法的主要工作就是通过获得的原始数据构建一个有众多 N 节点的图，然后用“边”来连接这些节点，从而构建一个连续的序列信息。

在 Velvet 中，它先根据 k-mer 构建 de Bruijn 图，然后结合序列特征及图的结构特征，处理图中的错误 read，最后找到一条最优路径作为拼接结果。与之前的算法相比，在拼接速度上有所提高，但由于没有对数据进行压缩处理，所以占用内存较多。

需求特点

- ① allpaths，velvet 等软件均为单机多线程运行，对内存需求较大，对存储性能要求较高；
- ② abyss 可进行分布式运算，将数据分布式存储于每个节点上，采用 MPI 实现节点间通信。

11. 序列比对

生物序列比对的基本思想是基于分子生物学中的一条经验规则，即当两个分子享有相似的序列时，由于进化关系或者物理化学限制，它们将很有可能具有相似的结构和生物学功能。因此序列比对（Sequence Alignment）的基本问题是比较两个或两个以上符号序列的相似性或不相似性，尽可能确切的反映它们之间的相似性和不相似性，用于阐明序列之间的同源关系，通过序列比对，找出序列之间的相似性，发现与结构相联系的保守序列片段，以及检测新测定序列与数据库中已知结构和功能的序列之间的相似性关系，从而以足够的可信度确定新序列的结构和功能信息。

对于第二代测序平台数据的分析，最为重要的一步是将所产生读段通过序列比对形式完成序列相似性比较。由于第二代测序平台数据通量高，因此，第二代序列比对方法一般分为两步：首先对读段数据或者参考基因组进行归类整理等预处理；然后通过适当的算法，将短读段序列进行比对和定位。

12. 高通量数据预处理方法

对高通量的序列片段数据进行的预处理方法一般有两类，即基于哈希表的方法和基于后缀 trie 的 Burrows-Wheeler 转换思想。

基于哈希表的方法基本思想是对参考基因序列（也有少数工具是对读段序列）建立哈希表，使用定长的种子在参考基因序列中选取序列与查询序列比对。基于哈希表思想的方法的优点是具有较高的匹配敏感性和准确性，但也有如下不足：第一，占用的内存空间大；第二，产生的种子匹配过多。

为解决第一个问题，有学者提出了空位种子方法。该方法简单来说，就是在定长的种子中空出个别位置，延长种子长度。使用该方法的比对工具主要有：SOAP，MAQ，RMAP，Mosaik 等。

为解决第二个问题，有学者提出了后缀 Trie 方法。后缀 Trie 是一种 n 叉树，n 为字母表大小。每个节点表示从根节点到此节点所经过的所有字符组成的字符串，它的根结点不包含任何信息，是一种以牺牲存储空间来降低序列查询时间的字符串预处理方式。后缀 Trie 的主要特征是：对于任何叶子 i，从根节点到该叶子所经历的边的所有标识串联起来后恰好拼出 S 的从 i 位置开始的后缀，即 S[i，…，m]。树中节点的标识被定义为从根到该节点的所有边的标识的串联。后缀 trie 思想即通过 n 叉树的形式组织各序列，利用字符串集合中字符串的公共部分来降低时间开销以达到提高效率的目的，后缀 trie 就是包含了所给字段的所有后缀。

基于后缀 Trie 思想的 Burrows-Wheeler 转换可以用“循环、排序”四个字来概括。采用 Burrows-Wheeler 转换的代表是 Bowtie 和 SOAP2，BWA，BWASW。

13. 序列比对方法

目前已知的序列比对方法很多，依据不同的划分方式有不同的分类。根据同时进行比对序列的数目分为两两比对和多序列比对，从比对范围考虑分为全局比对和局部比对。

双序列比对

根据算法结构的不同，将双序列比对算法分为三类：动态规划的优化方法，启发式算法和大型数据库搜索设计的概率方法。

① 动态规划的优化算法

Needleman-Wunsch 算法是最早的序列比对算法，属于全局序列比对，在生物信息处理中应用广泛。Smith-Waterman 算法是一种局部相似性的动态规划算法，在识别局部相似性时具有很高的灵敏度，是双序列比对算法中最基本的算法。

② 启发式算法

1) FASTA 算法

FASTA 是双序列比对启发式算法，采用了改进的 wilbllr 和 Lipmall 算法以集中反映具有显著意义的比对结果。它的基本思想是：一个能揭示出真实序列关系的比对至少包含一个两条序列都拥有的片段，把查询序列中的所有片段编成 Hash 表，然后在数据库搜索时查询这个 Hash 表，以检索出可能的匹配，这样命中的片段就能很快地被鉴定出来。

2) BLAST 算法

BLAST 算法可以兼顾搜寻的速度以及搜寻结果的精确度，它比 FASTA 速度更快。它的基本思想是：产生比 FASTA 更少而更有意义的增强点，以提高整个算法的速度。BLAST 算法在不失敏感性的前提下大大提高了算法的效率。

3) BLAT 算法

BLAT 算法最初用于人类基因组拼接和注释过程中的大规模数据比对任务上。其速度快、共线性输出结果简单易读，存在的局限性是对于特殊的任务需要选择合适的软件，如：用于远亲缘物种间的核酸序列比对时，比对精度就不够高；在重复搜索短小匹配片段的同时，会产生过多的没有生物学意义的序列比对碎片。

③ 大型数据库搜索设计的概率方法为基础的算法

MUMmer 算法是一种基于后缀树数据结构的全基因组比对方法，利用后缀树的数据结构有效地将算法的时间和空间复杂度由 (N³) 降到了 (N)。与 BLAST 算法相比，其后缀树法在速度上快得多，且能处理大量的插入和删除片段，能识别重复片段和单核酸多态性等多种全基因组序列中的复杂片段。

多序列比对

多序列比对算法的基础是动态规划比对算法，但随着比对序列数目及长度的增加，问题的解空间也急剧增大。多序列比对的常用算法有渐进算法、隐马尔科夫模型、迭代比对法等。累进方法是最常用的启发式多序列比对算法。

① 渐进比对算法

渐进比对算法是最常用的启发式多序列比对算法。算法的基本假设是要比对的序列是同源的。算法的基本思想是由近至远将序列或子比对结果按双重比对算法逐步进行比对，重复这一过程直到所有序列都加入为止。这类算法的主要优点是：简单、快速，所占内存较少。缺点是在比对初期引进的空位插入错误无法在比对后期因加入其它序列而改正，易陷入局部最优解。

CLUSTALW 是一个使用最广的渐进比对程序，该算法主要由三个步骤组成：计算距离矩阵；构建指导树；依据指导树进行渐进比对。CLUSTALW 对于亲缘关系较近的序列比对效果较好，但是对于分歧较大的序列，比对的准确率明显降低。

T-COFFEE 是另一个有代表性的渐进比对算法，它的主要特点是将序列的两两局部及全局比对结果做成一个扩展比对信息库，再利用扩展比对信息库中的信息对序列进行比对，这样在每一步比对过程中都用到了所有序列之间的关系信息，在一定程度上提高了序列比对的准确率，尤其是对于存在大量空位插入的情况，效果更为明显。

② 迭代比对算法

迭代比对（Iterative alignment）算法是另一类有效的多序列比对算法，它基于一个能产生比对的算法，并通过迭代方式精细多序列比对，直到比对结果不再改进为止。这类算法不能提供获得优化比对结果的保证，但却具有鲁棒性和对比对序列个数不敏感等特性。

基于遗传算法的多序列比对 SAGA 算法是一种实用的迭代算法。该算法的思想是将序列集中不等长的序列以两端加空位方式补齐，构造初始群体中的个体；将初始群体中的个体按一定的概率进行遗传操作（复制、连锁互换、突变）产生新的个体构成新种群；对新种群的个体重复上述的遗传操作，直到满足终止条件。个体适应度函数用 WSP 度量。该算法的优点是：可以对任意多个序列同时比对，而不会受到限制。主要缺点是速度慢，易于陷入局域优化解。

③ 基于隐马尔可夫模型的序列比对

隐马尔科夫模型是目前较先进的多序列比对方法，跟常规的方法相比，它可以发现序列久远的同源性。

从原理上，可以将序列比对方法分为如下两类：

1) 空位种子片段索引思想：它将一个短读段分成了若干条长度相等的更短的片段——种子片段。如果整段短小片段可以与参考基因组序列完全配对，那么很显然所有的种子片段也理所应当与参考基因组序列完全配对。但如果其中有一处错配，例如 SNP，那么肯定有一条种子片段无法与参考基因组序列完全匹配。依次类推，如果出现了两处错配就会导致一条或两条种子片段无法与参考基因组序列完全匹配。因此，对所有种子片段两两

组合后的片段进行比对，就有可能找出该短小片段在基因组中最有可能的位点。采用空位种子片段索引法的代表是 MAQ，ELAND，SOAP2 等。

2) Smith-Waterman 思想：Smith-Waterman 思想可以概括为一个评价打分技术，它基于动态规划策略的局部序列比对技术，在一条搜索路径中分数可能增加、减少或者不变。通过相似性评估技术来对当前节点打分，相同节点则增分，不同节点则减分，并且必须有间隙（Gap）惩罚机制来处理片段空隙。比较有代表性的使用该策略的比对软件有：SHRIMP、BFAST、Mosaik 和 BWA-SW 等。

需求特点

① 计算主要为整数计算，基本无浮点计算。

② 计算的主要特点为频繁的 load 和 write（内存的读写），意味着需要较大的内存容量和内存带宽，需要直连架构 CPU。

③ 具有很好的扩展性，性能与参与比对的进程数几乎完全成线性关系，性能对网络的压力不大，建议使用具有最高性价比的千兆网，而不需求使用具有较高价格的 Infiniband 高速网。

14. 计算资源需求

内存

常用的序列拼接软件一般需要较大的内存，内存容量需 1TB 以上，一般配置内存槽位较多的四路或八路机架式服务器。

序列比对软件可手动调节问题规模，对内存容量的需求不是必要条件，但需要保证内存带宽，按照 CPU 的内存通道数量配置足够多内存条数。

存储

序列拼接软件在计算时会产生中间数据，存储性能的好坏对整个计算过程所需要的时间影响较大，对存储的性能要求较高，一般配置多块本地硬盘，计算过程中的临时数据通过本地硬盘读写，一方面保证性能，另一方面减少序列拼接程序运行时对共享存储的影响。也可以为序列拼接节点配置单独的磁盘阵列。

序列比对计算时间断性的读写硬盘，存储性能越高，计算速度越快，由于序列拼接可手动调节问题规模，一般配置性价比较高的两路服务器，所以节点数目较多，需配置并行文件系统。

网络

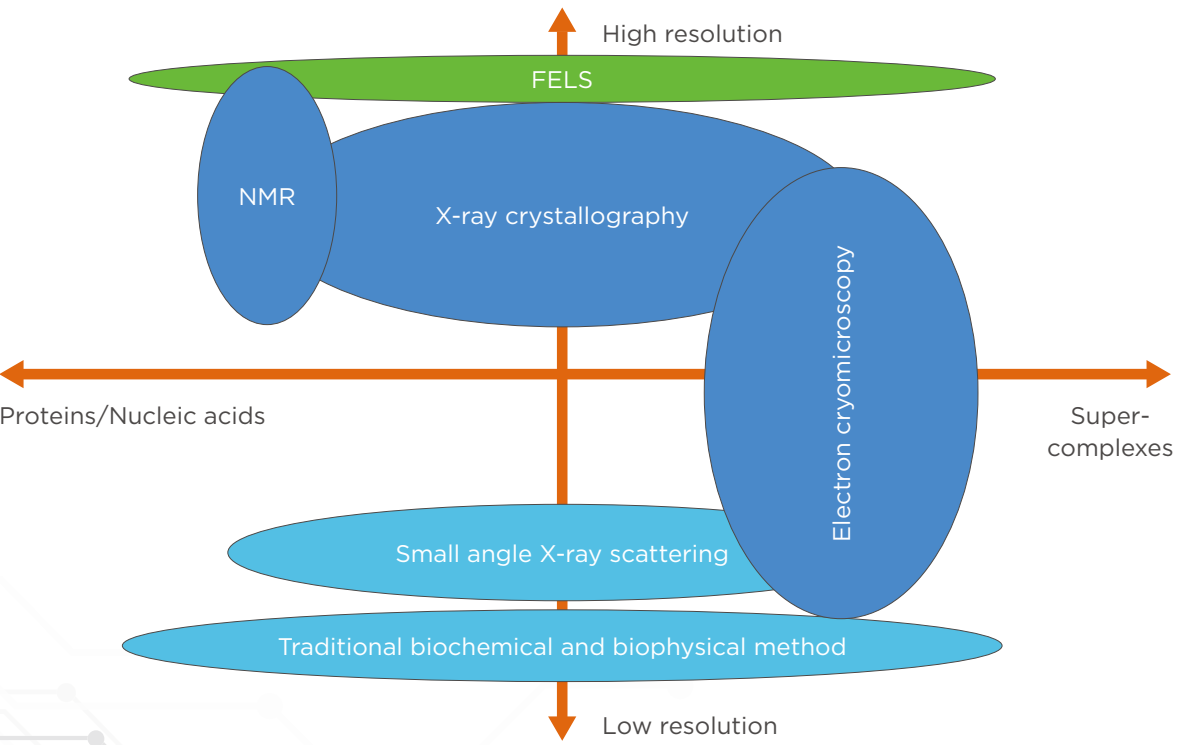
序列拼接、比对等软件大部分为单机运行，对网络的需求主要来自于访问存储的要求。一般配置千兆以太网即可，如果配置高速 Infiniband 网络，访问存储的性能会提高，减少数据读写的时间，从而提高计算效率。

GPGPU

目前已有部分软件实现了 GPGPU 加速，比如 GPU ClustalW，MUMmerGPU，SWAMP，CUDASW++，MSA-CUDA，GPU-BLAST，SOAP3，G-Aligne，BarraCUDA 等，可根据实际运行软件情况选配 GPGPU 卡。

蛋白质结构研究

结构生物学是通过研究生物大分子的结构与运动来阐明生命现象的科学，药物设计、疫苗开发和蛋白质分子性能改造等应用领域都以结构生物学的研究成果为基础。X 射线晶体学、核磁共振波谱学、电子显微三维重构（亦称电镜三维重构）是结构生物学的三大研究手段，具有不同的优势。核磁共振波谱学可以获得蛋白质在溶液中的三维结构，能够分析蛋白质的动态变化，但是研究对象的分子量通常难以超过 20kD。X 射线晶体学通常可以获得生物大分子原子分辨率的三维结构，但对于分子量较大、较复杂的生物复合分子体系，其晶体难以获得，结构解析难度也较大。当研究对象的尺度越来越大的时候，电子显微三维重构技术，特别是低温电子显微三维重构技术，就显示出其重要的作用，利用该技术我们可以获得分子量巨大（200kD 以上）的超分子复合体系的纳米分辨率三维结构，弥补晶体学和核磁共振分析技术的不足，架起从蛋白质、蛋白质复合体、超分子复合体系到亚细胞系统的三维结构研究的桥梁。



1. 电子显微三维重构

基本原理

电子显微三维重构技术的基本原理基于中央截面定理：三维物体沿电子束方向投影的傅立叶变换是该物体所对应的傅立叶空间中通过中心且垂直于投影方向的一个截面。那么一个物体的完备投影的二维傅立叶变换一定能够完全填充该物体的三维傅立叶空间，因此物体的完备投影与物体的三维结构是等价的。实验中，我们收集到大量全同颗粒在不同方向的投影（单颗粒），或同一样品在不同角度的投影（断层成像），首先通过等价线（common line）、投影匹配等方法确定每张投影图像的取向和中心（断层数据角度已知，只需精修即可），然后对每张投影图进行傅立叶变换，按照投影方向填充到三维傅立叶空间对应的切面，并进行差值计算得到倒空间的网格点数值，最后再进行反傅立叶变换，就可得到实空间的三维结构。三维重构理论的建立使得利用电子显微镜来解析蛋白质等生物大分子的三维结构成为可能，但是需要解决的问题是电子束打到生物样品上时会带来比较严重的辐射损伤，从而丢失了许多结构信息，在低温电镜（低温电镜的目的是提高生物样品对电子辐射的耐受力，同时减少图像热噪声）和低剂量辐照成像技术出现以前，人们只能采用重金属负染的方法来获得生物大分子的嵌影结构（不是真正的分子结构）。到了 80 年代，快速冷冻含水生物样品的制备技术和低温电子显微镜技术的快速发展并实用化，为利用电子显微技术研究蛋白质三维结构奠定了基础。特别是 1984 年，Dubochet 等发表了第一张病毒的低温电子显微镜照片，开创了低温电镜的研究时代。自此，在低温电子显微术、低剂量辐照成像术和电子显微三维重构理论的基础上，一门新的学科领域形成了，即低温电镜结构生物学，这一领域无论是在方法学上还是在蛋白质结构研究上，在近二十多年间都获得了快速发展，取得许多重要成果。利用低温电子显微术和三维重构技术解析蛋白质超分子复合物的结构迄今已分化发展为三种具有不同特点和适用范围的方法：电子晶体学（Electron Crystallography），主要处理二维晶体样品；单颗粒三维重构技术（Single Particle Analysis），主要解析具有全同性的蛋白质分子的三维结构；电子断层三维重构技术（Electron Tomography），主要针对不具有全同性的超分子体系和亚细胞体系。

第一种：电子晶体学

电子晶体学技术利用电子显微镜的成像和电子衍射的功能，从生物大分子的二维晶体获取结构信息，解析其三维结构。生物大分子在空间中有序排列，可以形成三维晶体，也可以形成二维晶体（特别是膜蛋白分子）。对于二维晶体来说，其只在 X-Y 平面内具有平移对称性，电子波照射到二维晶体上时能够发生衍射，根据电子显微镜记录的二维图像来确定相位，利用二维晶体的衍射图谱来确定振幅，从而通过反傅立叶变换计算出大分子的密度投影，之后再利用三维重构技术获得大分子的三维结构图，从而解析出生物大分子的三维结构。该方法的特点是解析分辨率较高，目前可达到近原子分辨率。相对来说膜蛋白更易形成二维晶体，因此电子晶体学在膜蛋白研究中更具有潜力。此外，该技术也可用于研究可溶蛋白质的三维结构。除了二维晶体外，电子晶体学技术还可以处理一类特殊的晶体结构，即具有螺旋对称性的样品结构。在某些条件下，蛋白质等生物大分子能够形成纤维状、具有螺旋对称性的微观结构，这种结构在电子显微镜下是可以观察到的，而且电子束经过这种螺旋结构可以发生衍射（helical diffraction），利用这种衍射现象并结合电镜的成像数据，我们就可以解析出这些大分子的三维结构。

第二种：单颗粒三维重构

该技术也叫做单颗粒分析，主要适用于结构具有全同性的生物大分子的结构解析，这些生物大分子可以是具有高度对称性的病毒颗粒（具有二十面体对称性），也可以是不具有对称性的蛋白质颗粒。蛋白质的分子量通常要求在 100kD 以上，在颗粒数目足够多的情况下，理论上其分辨率可以达到原子水平。蛋白质等生物大分子的溶液样品被加载到含有微孔的碳膜上，通过快速冷冻的方法，这些大分子颗粒被包埋在玻璃态的冰层当中，它们具有全同的三维结构，只是方位取向不同，其空间关系可以用一系列的刚体运动来描述。利用透射电子显微镜大量采集这些大分子的投影图像，通过图像处理和三维重构计算就可以获得该生物大分子的三维结构。通常的计算过程如下：从原始的电镜照片中将颗粒图像挑选出来，对其进行二维图像对中、分类和平均，然后通过计算等价线的方法推算各分类图的取向，利用傅立叶重构法建立初始三维结构模型，通过对原始图片或分类平均图与结构模型投影的匹配，优化取向参数，进而得到更准确的三维结构模型，如此反复对初始结构模型进行修正，直到收敛获得最终的结果。单颗粒三维重构在二维图像分类分析、取向参数求解和三维重构计算等方面根据研究对象特点的不同有不同的算法策略，因而演化出多种不同的单颗粒分析软件，常用的有 EMAN，SPIDER，IMIRS，FREALIGN，IMAGIC，AUTO3DEM，XMIPP 等。

第三种：电子断层三维重构

电子断层三维重构技术就是对样品做微观 CT。具体步骤为：将制备好的样品放在透射电子显微镜中，以一定角度间隔对样品进行旋转，并用 CCD 或胶片记录样品在不同角度下的二维投影图像——样品的倾转范围通常为－70°～＋70°，间隔角为 1°或 2°；然后对收集的这些原始数据进行匹配，利用加权背投影、代数迭代（ART）或同步迭代（SIRT）等方法进行三维重构获得样品的三维图像。由于电子断层三维重构的研究尺度可以很大——可以对细胞水平上的超微结构进行重构、观察和分析，因此我们也称该技术为电子断层三维成像技术。该技术适宜对细胞器、亚细胞组装体甚至整个细胞的三维结构进行研究，分辨率为 5～20 nm。低温快速冷冻技术和自动化数据收集技术的发展是电子断层三维重构技术发展过程中的两个关键技术，尤其是低温快速冷冻技术避开了树脂包埋和负染技术中的假象，能够保持样品接近天然状态的结构。截至目前，利用电子断层三维重构技术，已经有大量文献报导了有关整个细胞、细胞骨架、病毒、细胞器等超微结构。如 Baumeister 等利用该技术获得了真核细胞 dictyostelium 的三维结构，其分辨率达到了 5～6 nm，并在原位对核糖体和 26S 蛋白酶体等大分子复合物进行了定位观察。

在上面所述的多种电子显微三维重构方法中，发展最快应用最广的是单颗粒三维重构技术，近年来不断有文章报道利用此技术所获得的大分子复合物的三维结构，其中许多具有高对称性的病毒颗粒结构已经达到近原子分辨率。



数据收集与处理

① 单颗粒分析原始数据的自动化收集

蛋白质等生物大分子样品固定在玻璃态冰层中，对电子束辐射非常敏感，其原始图像数据的收集需要在低剂量（low dose）模式下进行，常用的电子剂量为 10～30 e/Å²，在该模式下的图像采集分为三步：首先在低倍（search 模式，辐照剂量非常低）下寻找冰层厚度合适的区域；然后进行第二步操作——聚焦（focus 模式），将放大倍数设置为曝光时所需倍数或者更大，在选定曝光区域的 1～2μm 附近的碳膜上确定正焦点；然后设定合适的欠焦值（－1～－3μm）和曝光时间；最后完成第三步操作——曝光（exposure 模式），在选定区域进行曝光记录图像。由于整个过程中目标区域只经过一次电子束低剂量照射，只有在成像后才能获知图像的质量以及所包含颗粒的数目和分布情况，因此要完成单颗粒分析，必须收集大量的原始图像。要收集如此大数目的颗粒，手动数据收集方式将成为瓶颈，必须寻找有效的自动化收集方法。近五年来，单颗粒原始数据的自动化收集技术得到很快发展，一方面适合于自动化数据收集的含有有序排列微筛的碳支持膜被开发出来并实现了商业化，如 QuantifoilTM（<http://www.quantifoil.com>）和 C-flatTM（<http://www.emsdiasum.com/microscopy/products/grids/cflat.aspx>）；另一方面若干单颗粒自动化数据收集软件被开发出来并逐渐成熟，主要有 Leginon，AutoEM 和 JADAS 等软件，这些自动化软件能够自动识别成像区域，自动完成样品对中和自动对焦，经过一定的参数设置后可以实现大量数据的自动化收集，一个典型的例子是 Scripps 研究所的 Carragher 研究组利用 Leginon 系统在 25 小时内自动完成了 284742 个分子伴侣 GroEL 的低温单颗粒数据收集工作，并且非常轻松地获得了好于 8 Å 分辨率的三维重构结果。

② 单颗粒三维重构的自动化颗粒挑选和图像处理流水线

单颗粒三维重构技术需要从原始图片中将蛋白质等生物大分子颗粒选取出来，手动挑选数十万的颗粒是一项非常耗时费力的工作，因此自动化的颗粒挑选显得十分重要。到目前为止，已经有很多颗粒自动挑选的软件，主要算法有基于模板的方法（template-based methods）、基于边缘检测的方法（edge detection-based）、灰度值比较的方法（intensity comparison）、基于纹理的方法（texture-based）和神经网络方法（neural network）等，识别率能够达到 80%～90% 左右。

软件名称	算法及特点
TYSON	局部平均，模板匹配，局部方差等三种
SLEUTH	模式识别
FindEM	基于模板匹配
SwarmPS	相关和边缘检测
DoG Picker	高斯差分图像变换（Difference of Gaussians）
Cyclops	自动排除碳膜区域，基于旋转空间采样
SIGNATURE	基于模板，分等级筛选（hierarchical screening）
ETHAN	灰度值强度比较，适用于球形颗粒
Xmipp	机器学习技术
SPIDER	基于模板匹配，局部标准化相关系数算法

③ 电子断层三维重构的自动化数据收集

电子断层成像时要在样品的特定部位连续收集上百张不同角度的照片，这就要求数据收集必须在极低剂量下进行，否则样品将被电子束损伤。一般而言在进行数据收集时，先在低倍下寻找样品，然后在高倍下样品周围的区域聚焦，最后在准备收集数据的区域曝光，这样可以有效地减少辐射损失。电子断层数据收集存在许多实际问题：1) 样品由一个角度转向另一个角度时，由于机械原因，往往会有一定的偏移，在数据收集过程中要进行匹配对中（alignment），所以实际在收集过程中必须额外拍摄一些用于匹配的图像，这势必造成额外的剂量增加；2) 由于样品和仪器本身的特点，电子断层只能收集到 ±70 度左右的数据，在数据收集时应采用合理的策略以使所得的数据包括尽量多的信息，基于这样的考虑产生了等角度、等斜率以及 saxton 三种常用的收集策略；3) 由于冷冻样品的成像衬度较低，从高角度开始收集数据有可能无法找到样品，所以一般采用从 0 度开始逐渐向高角度收集数据，然后再退回 0 度向反方向收集数据；4) 在数据收集过程中，样品不同部位欠焦量并不一致，甚至可能某些部位处于过焦而另一些区域处于正焦或欠焦，并且欠焦量也可能随着角度的变化而不同；5) 样品厚度也是影响电子断层数据收集的一个重要因素，厚样品能够包含更丰富的结构信息，尤其是对一些较大的细胞器，但厚样品衬度不如薄样品好，而且随着角度增大，衬度会越来越差，这为数据收集带来困难，重构分辨率也比薄样品低，所以样品厚度要在两者之间权衡。针对这些实际问题，开发自动化的电子断层数据收集软件就非常有必要，目前广泛使用的软件有 UCSF tomography, TOM software toolbox, SerialEM, Leginon 以及 FEI 公司的 Explor3D 等，这些自动化数据收集程序（如 Xplore3D）能够利用最初几张图像的偏移算出整个数据收集过程中图像的偏移轨迹，从而大大降低了额外增加的剂量对样品的损伤，并且能够比较精确地控制电镜和 CCD 相机，完成自动聚焦功能，能够实现多种收集策略和模式。

计算资源需求

许多电镜三维重构程序目前是实现的任务级别的并行，如 BSOFT、FREALIGN、IMOD 和 PRIISM/IVE 等。而 AUTO3DEM、IMAGIC、UCSF TOMOgraphy 等则完全采用了 MPI 并行方式。著名的单颗粒三维重构软件 EMAN 目前除了 GPU 外，其它几种并行手段都已实现，SPIDER 和 IMIRS 程序则采用了 OpenMP 和 MPI 并行方式，但 SPIDER 保留了任务级的并行方式，而 IMIRS 程序没有。XMIPP 采用了 MPI 和 pthreads 方式。

程序包	方法	并行任务	并行方式
AUTO3DEM	单颗粒	定取向；重构	MPI
BSOFT	单颗粒	重构	任务级
	电子断层	重构；滤波；分辨率估计	任务级
EMAN	单颗粒	分类；定取向；重构；结构识别	任务级；MPI；OpenMPI；多线程
FREEALIGN	单颗粒	定取向	任务级
IMAGIC	单颗粒	定取向；重构	MPI
IMIRS	单颗粒	定取向；重构	MPI；OpenMP
IMOD	电子断层	CTF 矫正；重构；滤波；双轴数据合并	任务级
PRIISM/IVE	电子断层	重构；双倾数据匹配	任务级；GPU _s
SPIDER	单颗粒	重构；定取向；模板匹配	OpenMP；MPI；任务级
	电子断层	重构；模板匹配	OpenMP；任务级
UCSF TOMOGRAPHY	电子断层	重构	MPI
XMIPP	单颗粒	分类和匹配；定取向；重构	MPI；多线程

① 需求总结

- 1) 主要为浮点计算，其中大量的单精度 FFT 计算。
- 2) 软件加速比接近线性，对网络依赖低。
- 3) 软件内存需求和 IO 需求较大，一般需要配置并行文件系统。

2. 质谱仪原始资料处理

发展现状

质谱方法（Mass Spectroscope, MS）是通过正确测定蛋白质分子的质量而进行蛋白质分子鉴定、蛋白质分子的修饰和蛋白质分子相互作用的研究，基本原理是带电粒子在磁场或电场中运动的轨迹和速度依粒子的质量与携带电荷比（质荷比，M/Z）的不同而变换，从而可以据此判断粒子的质量及特性。但长期以来，质谱方法仅限于小分子和中等分子的研究，因为要将质谱应用于生物大分子需要将之制备成气相带电分子，然后在真空中物理分解成离子。但如何使蛋白分子经受住离子化过程转成气相带电的离子而又不丧失其结构形状是个难题。20 世纪 70 年代，解吸技术的出现成功地将蛋白分子转化成气相离子。尔后快原子轰击与其紧密相关的溶液基质二次离子质谱法使得具有极性的、热不稳定的蛋白分子可经受住电离过程。但这些方法仅限于 10kD 以下蛋白分子的研究。80 年代电喷雾电离 (ESI) 和软激光解吸 (SLD) 电离技术的发展则使得质谱方法应用于高分子量蛋白分子的研究。

电喷雾电离 (ESI) 原理可按电荷残留模型予以描述，带电液滴蒸发，液滴变小，液滴表面相斥的静电荷密度增大。当液滴蒸发到某一程度，液滴表面的库仑斥力使液滴爆炸。产生的小带电液滴继续此过程。随着液滴的水分子逐渐蒸发，就可获得自由徘徊的质子化和去质子化的蛋白分子。针对电喷雾电离所产生的多电荷状态，Fenn 将多电荷状态理解为对分子质量进行多次独立的测量，并基于联立方程解的平均方法，获得对分子质量的正确估量，解决了多电荷离子信息的问题，使蛋白分子质量测量精度获得极大的提高，并于 1988 年首次成功地测量了分子量为 40 kD 的蛋白质分子，精确度达到 99.99%。

软激光解吸 (SLD) 是指从激光脉冲中获得能量后，样品分子以完整的低电荷分子离子释放，然后由电场加速。运用激光解吸电离蛋白分子时，激光的能量和波长、化学 / 物理基质的吸收和热传递特性，与基质中分析物的分子结构之间需要作合理的选择调配。Tanaka 选用了低能量氮激光和含有胶状颗粒的甘油作基质，成功地测定了高分子量的糜蛋白酶原、梭肤酶 -A 以及细胞色素。由于 Tanaka 成功的开创性工作，SLD 技术迅速发展。目前占主导的方法是基质辅助激光解吸电离 (MALDI)。这一方法是将样品掺入一种低分子量的结晶基质，基质的最大吸收与激光脉冲波长匹配。由于 MALDI 产生的是低电荷的完整气相大分子，可用于检测纯度不高的生物分子。

生物质谱技术由于其高灵敏度、高通量和高精度等特点，目前是大规模蛋白质鉴定的主流技术。在典型的自底向上蛋白质组学研究策略中，蛋白质样品被酶解成肽段混合物，后者通过色谱 - 质谱联用生成串联质谱。从串联质谱重构出肽段序列，是蛋白质鉴定的核心计算问题。目前，最成功和最常用的解决方法是用串联质谱搜索蛋白质序列数据库，将数据库中的序列做理论酶切和理论碎裂，然后将预测的谱图跟实验谱图匹配，从而鉴定肽序列，进而鉴定整个蛋白。基于蛋白质序列库搜索的蛋白质鉴定，实际上是一个检索系统，其核心计算问题是谱图匹配的肽打分算法。同时，为了得到正确的鉴定结果，蛋白质鉴定系统还必须对检索结果的可靠性进行统计评估。蛋白质的修饰给蛋白质鉴定检索系统的速度和精度都带来了更大挑战。

计算资源需求

① 常用软件

在用于蛋白质分析的质谱技术中，蛋白质样品首先被选定的蛋白酶水解，形成多肽。不同质量电荷比的多肽离子被质谱仪分离、检测出来，得到一级质谱。这些肽离子可以进一步被打碎，形成碎片离子。碎片离子被分离和检测便得到串联质谱。用质谱鉴定蛋白质的方法因此分为两大类：

第一类是基于一级质谱的，称作肽质量指纹作图。这类方法搜索已知蛋白质数据库，用指定的酶对蛋白质进行模拟水解，得到理论一级质谱。理论一级质谱与实验一级质谱进行比较，结果按照匹配的程度排序。这样的系统有 MOWSE，Mascot，ProFound，PeptIdent，MS-Fit 等等。肽质量指纹作图适用于蛋白质样品包含一种蛋白质或简单的混合物的情况。其缺点是由于蛋白质混合物和污染物、部分酶解、残基修饰（所谓氨基酸残基是指去掉一个水分子的氨基酸）、质量精度等因素的影响，往往误差较大，导致搜索结果错误。

第二类是基于串联质谱的。这类方法首先利用串联质谱技术 (Tandem Mass Spectrometry，MS/MS) 准确测定肽的氨基酸序列，再通过肽序列鉴定蛋白质的序列。所以，这种方法可以用于鉴定复杂的蛋白质混合物或者验证肽质量指纹搜索的结果，是目前最常用最有效的主流方法。

为了鉴定蛋白质，从串联质谱鉴定肽的氨基酸序列是中心问题。从串联质谱鉴定肽序列的计算方法有三种。数据库搜索方法、从头测序方法、序列标签查询方法。其中最常用的是数据库搜索方法。在这种方法中，数据库中的蛋白质序列被理论水解和碎裂，生成理论串联质谱。把理论质谱与实验质谱相比较，从而找到生成实验质谱的肽序列。

在利用串联质谱鉴定蛋白质的方法中，蛋白质鉴定问题归约为更基本的肽鉴定问题。而数据库搜索方法是目前普遍采用的肽鉴定方法。给定实验串联质谱，对数据库中的候选肽进行匹配打分是肽鉴定算法的核心。对肽打分鉴定结果的评价，即识别出正确鉴定的肽序列，也是必不可少的一步。

所谓“肽打分”是指： 给定实验串联质谱，对候选肽产生该质谱的可能程度做出评分，从而对所有候选肽进行排序。用信息检索的语言描述，这里的串联质谱即是输入的查询，候选肽即是数据库中保存的对象，而肽打分函数实际上就是检索函数或称排位函数。肽打分函数的功能就是对候选肽进行排位，把最可能产生实验质谱的肽序列排在首位。可以把肽打分函数按构造方式分为三类。

第一类基于谱向量点积；



第二类基于概率；

如 Mascot，SCOPE，ProbID 和 PepSearch 等等。

Mascot 是除了 SEQUEST 之外，另一个广泛采用的商业蛋白质鉴定软件。Mascot 试图计算实验串联质谱由候选肽随机生成的概率 p ，而候选肽分数为 $-\log(p)$ 。Mascot 的概率打分算法综合考虑了肽长度的分布、酶切位点遗漏概率、质量误差分布以及离子强度等因素。

SCOPE 是 Celera 公司设计的打分算法。它利用贝叶斯模型进行打分，对于给定的质谱求每个序列的后验概率。SCOPE 通过用两步随机过程模拟串联质谱生成的过程：1) 根据概率分布生成肽的碎片；2) 根据仪器测量误差，从碎片生成质谱。

ProbID 试图计算实验串联质谱由候选肽随机生成的贝叶斯后验概率。但是 ProbID 计算的概率不能算作真正意义上的概率，而只不过是若干因素的简单乘积。其中包括亚胺离子的出现情况、肽序列酶切点是否满足酶的特异性、匹配和不匹配的谱峰以及连续和互补离子的匹配情况等等。

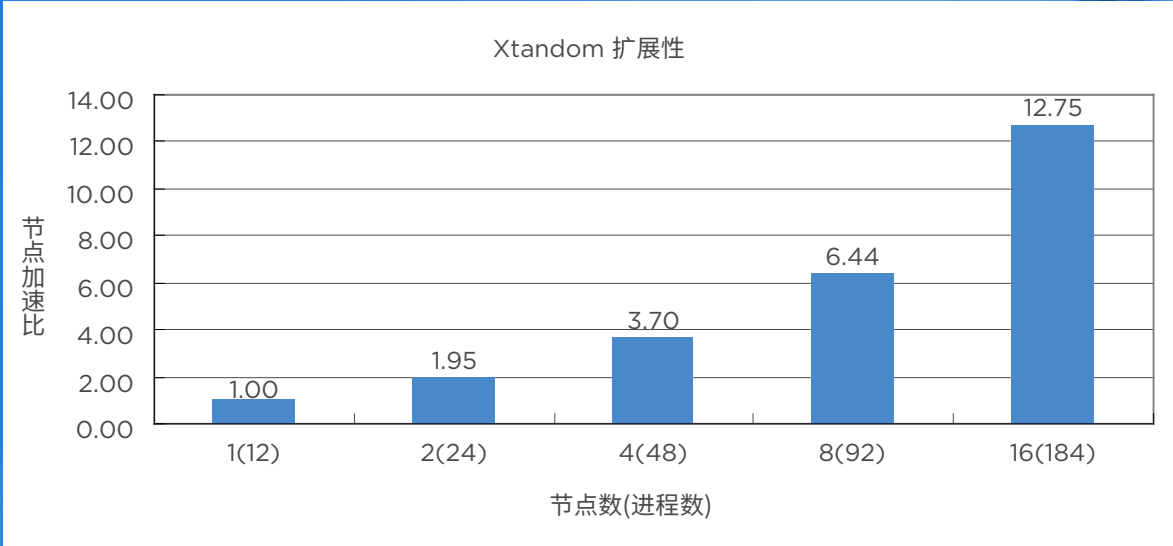
第三类基于机器学习，或者说基于模式分类。

肽鉴定本质上可以看作将候选肽分为“正确”和“不正确”的两类分类问题。在基于机器学习的肽打分函数中，候选肽与实验质谱间的多种匹配信息被表示成特征向量的形式，然后利用机器学习方法从序列已知的质谱训练数据中学习出一个打分函数。

对于每个质谱，在数据库搜索之后，尽管总会有一个得分最高的候选肽，但是这个候选肽不一定是正确的。对肽鉴定结果可靠性的评估早期使用的是经验阈值法，如 SEQUEST。经验阈值法的好处是简单直观，但缺点也很明显，那就是阈值的指定只凭经验，缺少理论上的依据。当数据库规模增大时，错误候选肽的最高分也会水涨船高。并且，根据阈值过滤出的结果，其可靠性没有定量的估计。使用经验阈值是一种武断的做法，实际上，无论肽鉴定结果的得分有多高，都带有或多或少的不确定性。为了有效估计鉴定结果的可靠性，必须利用统计手段。目前，使用最多的鉴定结果可靠性统计度量指标是针对单谱鉴定的期望值和针对多谱鉴定的假发现率。

随机分值的概率分布决定了期望值的计算方法。依据概率分布来源的不同，期望值的计算方法可分为三类，即经验拟合法、理论建模法和穷举计算法。经验拟合法是通过拟合实际的分值分布数据估计概率分布，理论建模法是通过假定某种随机概率模型从理论上推导分值的随机分布，穷举计算法是通过穷举所有可能候选肽得到分值的真实分布。通过经验拟合法计算期望值的搜索引擎有 Sonar，X!Tandem，pFind，及 RAld_DbS 等，通过理论建模法计算期望值的搜索引擎有 Mascot 和 OMSSA，穷举计算法则是最近由金（Kim）等人提出的。三种方法各有优缺点，经验拟合法适用于任意类型的打分函数，但要求必须有足够的候选肽规模以拟合分值分布以及恰当的分布形式假设；理论建模法对于任意给定的一条肽序列都可以计算其期望值，但是只适用于基于概率的打分函数，准确性取决于概率模型的准确性；穷举计算法能够直接计算出分值的真实分布，但是只适用于加和性的打分函数，并且计算复杂度较高。

② 应用测试



图：Xtandom 扩展性测试

③ 需求特点

- 1) 软件加速比接近线性，对网络依赖低。
- 2) 系统计算量大，计算时间长，系统稳定性，可靠性要求高。

计算资源需求

① 常用软件

1) NAMD

NAMD 是一个用于生物大分子大规模分子动力学的并行软件，支持 Charmm、Gromacs 和 Amber 等多种力场，由美国 Illinois 大学生物物理系和计算机系联合开发，旨在开发出高效的分子动力学并程序，可支持 MPI 和 Charm++ 并行。NAMD 具有非常强的大规模并行计算能力，已经实现了在具有上千个处理器上的并行计算机上，对包含超过三十万个原子的的大分子系统进行模拟。NAMD 注册后可免费下载使用。

官方网站：<http://www.ks.uiuc.edu/Research/namd/>

2) AMBER

Amber (Assisted Model Building with Energy Refinement) 是一个常用的分子动力学模拟套件，特别适用于生物大分子。同时，Amber 力场也是常用的分子力场。

官方网站：<http://ambermd.org/>

自从版本 10 开始，Amber 被拆分成了两部分。AmberTools：摘除了核心程序（如 sander）的工具集，以 GPL 开源发布；Amber：Amber 核心程序，需要付费购买。

3) GROMACS

GROMACS 是一个开源的、高效的、基于标准 MPI 并行环境设计的大规模分子动力学并程序，基于 GROMACS 分子力场，由荷兰的 Groningen 大学 Department of Biophysical Chemistry 开发。GROMACS 主要执行生物大分子如蛋白质、核酸、磷脂等物质的计算，也可研究非生物的有机大分子系统，如高分子多聚体等。可以免费下载使用。

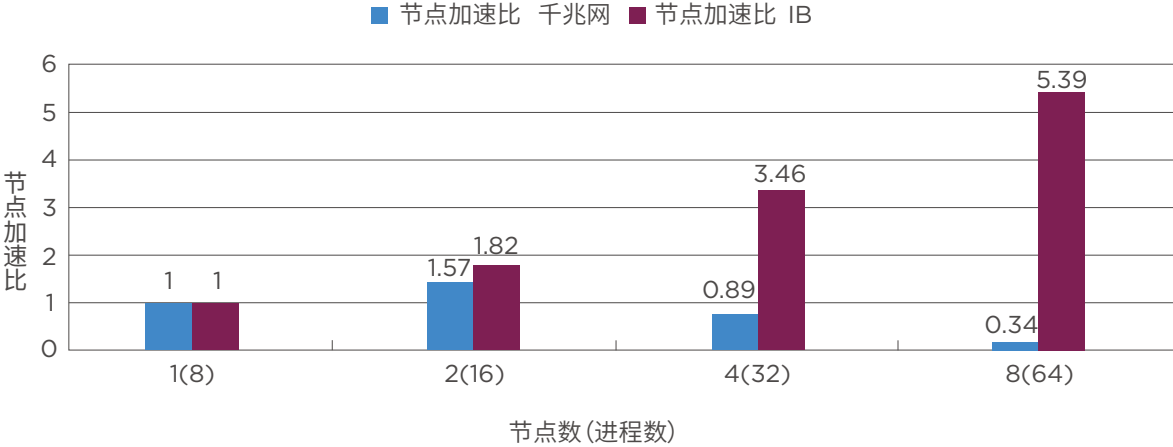
官方网站：<http://www.gromacs.org/>

4) CHARMM

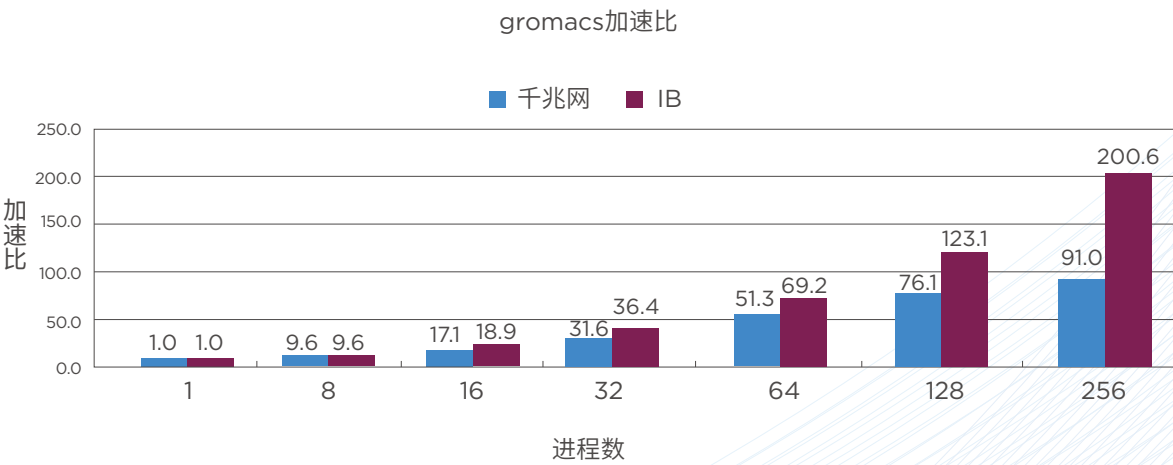
CHARMM (Chemistry at HARvard Macromolecular Mechanics) 是一种常用的分子立场，CHARMm 是相应的分子动力学软件。CHARMm 常用于生物大分子的模拟，包括能量最小化，分子动力学和蒙特卡罗模拟等。

官方主页：<http://www.charmm.org/>

② 应用测试



图：NAMD 扩展性及网络对比测试



图：Gromacs 扩展性及网络对比测试

③ 需求特点

- 1) 主要是浮点计算为主。
- 2) 非常适合大规模并行，但是对网络的需求比较大，节点间需要使用低延时，高带宽的 Infiniband 高速网。

计算机辅助药物设计

1. 发展现状

计算机辅助药物设计根据受体的结构是否已知，分为直接药物设计和间接药物设计。

计算机辅助药物设计的方法始于 1980 年代早期。当今，随着人类基因组计划的完成、蛋白组学的迅猛发展，以及大量与人类疾病相关基因的发现，药物作用的靶标分子急剧增加；同时，在计算机技术推动下，计算机药物辅助设计在近几年取得了巨大的进展。

计算机辅助药物设计的一般原理是，首先通过 X 一单晶衍射技等技术获得受体大分子结合部位的结构，并且采用分子模拟软件分析结合部位的结构性质，如静电场、疏水场、氢键作用位点分布等信息。然后再运用数据库搜寻或者全新药物分子设计技术，识别得到分子形状和理化性质与受体作用位点相匹配的分子，合成并测试这些分子的生物活性，经过几轮循环，即可以发现新的先导化合物。因此，计算机辅助药物设计大致包括活性位点分析法、数据库搜寻、全新药物设计。

（一）活性位点分析法

该方法可以用来探测与生物大分子的活性位点较好地相互作用的原子或者基团。用于分析的探针可以是一些简单的分子或者碎片，例如水或者苯环，通过分析探针与活性位点的相互作用情况，最终可以找到这些分子或碎片在活性部位中的可能结合位置。由活性位点分析得到的有关受体结合的信息对于全新药物的设计具有指导性。目前，活性位点分析软件有 DRID、GREEN、HSITE 等。另外还有一些基于蒙特卡罗、模拟退火技术的软件如 MCSS、HINT、BUCKETS 等。

其中，GRID 由 Goodford 研究小组开发，其基本原理是将受体蛋白的活性部位划分为有规则的网格点，将探针分子（水分子或甲基等）放置在这些网格点上，采用分子力场方法计算探针分子与受体活性部位各原子的相互作用能，这样便获得探针分子与受体活性部位相互作用的分布情况，从中可发现最佳作用位点。GRID 最初运算的例子是用水分子作为探针分子，搜寻到了二氢叶酸还原酶（DHFR）活性部位中水的结合位点以及抑制剂的氢键作用位点。由此软件成功设计的药物有抗 A 型感冒病毒药物 4 一胍基 Neu5Ac2en（GG167，Relenza™）。该化合物有很强的抗感冒病毒能力，克服了以往抗感冒病毒药物的耐药性缺陷，具有很好的市场前景。

MCSS 是 Miranker 和 Karplus 在 CHARMM 力场基础上发展而来，它的基本要点是在运用 CHARMM 力场进行分子动力学模拟时，取消溶剂分子间的非键相互作用。这样，在分子动力学模拟时，溶剂在能量合适的区域叠合在一起，从而提高了搜寻溶剂分子与受体分子结合区域的效率。小分子碎片（如水和苯分子）可当作溶剂分子，运用上述动力学方法搜寻出分子碎片与受体的结合区域，然后对每个碎片选择 100 — 1000 个拷贝，在低能碎片结合域进行能量优化。在最后的能量搜寻过程中，可以用随机取样或网格点的方法来实施。搜寻时每个碎片的各个拷贝可以作刚性转动，最后直接比较每个碎片各个拷贝与受体的结合能，以此选择碎片的最佳作用位点。2001 年 Adlington 等利用 MCSS 对前列腺特异性免疫抗原（PSA）的活性位点进行了详细分析，以此对已有的 PSA 抑制剂进行结构优化，从而得到了迄今为止活性最高的 PSA 抑制剂。

2. 数据库搜寻

目前数据库搜寻方法分为两类。一类是基于配体的，即根据药效基团模型进行三维结构数据库搜寻。该类方法一般需先建立一系列活性分子的药物构象，抽提出共有的药效基团，进而在现有的数据库中寻找符合药效基团模型的化合物。该类方法中比较著名的软件有 Catalyst 和 Unity，而以前者应用更普遍。

另一类方法是基于受体的，也称为分子对接法，即将小分子配体对接到受体的活性位点，并搜寻其合理的取向和构象，使得配体与受体的形状和相互作用的匹配最佳。在药物设计中，分子对接方法主要用来从化合物数据库中搜寻与受体生物大分子有较好亲和力的小分子，从而发现全新的先导化合物。分子对接由于从整体上考虑配体与受体的结合效果，所以能较好地避免其他方法中容易出现局部作用较好，整体结合欠佳的情况。

刚性对接：刚性对接方法在计算过程中，参与对接的分子构象不发生变化，仅改变分子的空间位置与姿态，刚性对接方法的简化程度最高，计算量相对较小，适合于处理大分子之间的对接。

半柔性对接：半柔性对接方法允许对接过程中小分子构象发生一定程度的变化，但通常会固定大分子的构象，另外小分子构象的调整也可能受到一定程度的限制，如固定某些非关键部位的键长、键角等，半柔性对接方法兼顾计算量与模型的预测能力，是应用比较广泛的对接方法之一。

柔性对接：柔性对接方法在对接过程中允许研究体系的构象发生自由变化，由于变量随着体系的原子数呈几何级数增长，因此柔性对接方法的计算量非常大，消耗计算机时很多，适合精确考察分子间识别情况。

目前具代表性的分子对接软件主要有 DOCK、FlexX 和 GOLD。

DOCK 由 Kuntz 小组于 1982 年开发，最新版本为 DOCK 6.11。DOCK 的开发经历了一个由简单到复杂的过程：DOCK 1.0 考虑的是配体与受体间的刚性形状对接；DOCK 2.0 引入了“分而治之”算法，提高了计算速度；DOCK 3.0 采用分子力场势能函数作为评价函数；DOCK 3.5 引入了打分函数优化以及化学性质匹配等；DOCK 4.0 开始考虑配体的柔性；DOCK 5.0 在前面版本基础上，采用 C++ 语言重新编程实现，并进一步引入 GB/SA 打分。DOCK 程序现已成功地应用于药物分子设计领域。Kuntz 等利用 DOCK 程序研究 HIV — 1 蛋白酶，根据分子相似性对剑桥晶体数据库进行搜寻，得到化合物 haloperidol，通过测试，其对 HIV — 1 蛋白酶的 Ki 值为 100μmol/L；进一步的结构改造得到化合物 thiolethal，其 IC50 高达 15μmol/L。DesJarlais 利用 DOCK 程序的一个改进版 target-DOCK 搜寻 HIV — 1 蛋白酶抑制剂，得到一系列 HIV — 1 蛋白酶抑制剂，其中活性最高的化合物其 Ki 值为 7μmol/L。

AUTODOCK 是 Scripps 的 Olson 科研小组开发的分子对接软件包，AUTODOCK 采用模拟退火和遗传算法来寻找受体和配体最佳的结合位置，用半经验的自由能计算方法来评价受体和配体之间的匹配情况。软件免费，2007 年发布了 Version 4。在 AUTODOCK 中，配体和受体之间结合能力采用能量匹配来评价。在 1.0 和 2.0 版本的软件中，能量匹配得分采用简单的基于 AMBER 力场的非键相互作用能。非键相互作用来自于三部分的贡献：范德华相互作用，氢键相互作用，以及静电相互作用。在 3.0 版中，AUTODOCK 提供了半经验的自由能计算方法来评价配体和受体之间的能量匹配。

FlexX 是一种快速、精确的柔性对接算法，在对接时考虑了配体分子的许多构象。 FlexX 首先在配体分子中选择一个核心部分，并将其对接到受体的活性部位，然后再通过树搜寻方法连接其余片断。FlexX 的评价函数采用改进的结合自由能函数。FlexX 的对接算法建立在逐步构造策略的基础之上，分以下三步：第一步是选择配体的一个连接基团，称为核心基团；第二步将核心基团放置于活性部位，此时不考虑配体的其他部分；最后一步称为构造，通过在已放置好的核心基团上逐步增加其他基团，构造出完整的配体分子。FlexX 对接一个典型的药物分子大约需要 3 分钟，表明它可用于中等规模的三维数据库搜寻；此外，由于其采用了经验结合自由能函数进行评价，结果可能要优于以相互作用能为评价函数的分子对接方法。因此，FlexX 是一个非常有前途的药物设计方法，近年来发展迅速。

3. 全新药物设计

数据库搜寻技术在药物设计中广为应用，该方法发现的化合物大多可以直接购买得到，即使部分化合物不能直接购买得到，其合成路线也较为成熟，可以从专利或文献中查得，这都大大加快了先导化合物的发现速度。但是，数据库搜寻得到的化合物通常都是已知化合物，而非新颖结构。近年来，全新药物设计越来越受到人们的重视，它根据受体活性部位的形状和性质要求，让计算机自动构建出形状、性质互补的新分子，该新分子能与受体活性部位很好地契合，从而有望成为新的先导化合物；它通常能提出一些新的思想和结构类型，但对所设计的化合物需要进行合成，有时甚至是全合成。全新药物设计方法出现的时间虽然不长，但发展极为迅速，现已开发出一批实用性较强的软件，其主要软件有 LUDI、Leapfrog、GROW、SPROU 等，其中 LUDI 最为常用。

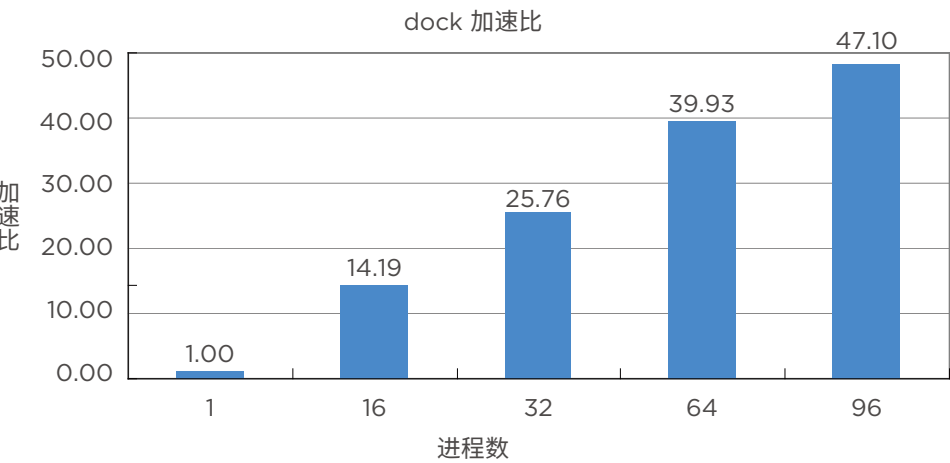
LUDI 是由 Bhöm 开发的进行全新药物设计的有力工具，已广泛地被制药公司和科研机构使用，其特点是以蛋白质三维结构为基础，通过化合物片段自动生长的方法产生候选的药物先导化合物。它可根据用户确定好的蛋白质受体结合部位的几何形状和物理化学特征（氢键形成能力、疏水作用位点），通过对已有数据库中化合物的筛选并在此基础上自动生长或连接其他化合物的形式，产生大量候选先导化合物并按评估的分值大小排列，供下一步筛选；可以对已知的药物分子进行修改，如添加 / 去除基团、官能团之间的连接等。在受体蛋白质结构未知的情况下，此模块也可以根据多个已知的同系化合物结构的叠合确定功能团，再根据功能团的空间排列和理化性质推测可能的蛋白质受体结合部位特征，根据此特征进行新型药物设计。目前研究人员利用 LUDI 设计出数十个针对不同疾病的活性化合物。

4. 计算资源需求

常用软件

- ① DOCK 免费，半柔性对接，支持 MPI 并行，应用广泛
- ② AutoDock 免费软件，半柔性对接（不支持并行）
- ③ FlexX 商业软件
- ④ Discovery Stuio 商业包含多种尺度的分子对接方法（创腾公司代理）
- ⑤ ZDOCK，RDOCK（刚性对接，蛋白质对接）学术版免费
- ⑥ MORDOR，（柔性对接）

应用测试



图：DOCK 扩展性测试

需求特点

Dock 扩展性良好，dock 并行模式为将各个小分子平均分配到每个进程中，所以对网络的性能要求不高，扩展性主要看看配体小分子的数量和它们计算规模的均匀性。

生命科学应用特点及解决方案

仪器后处理类的应用

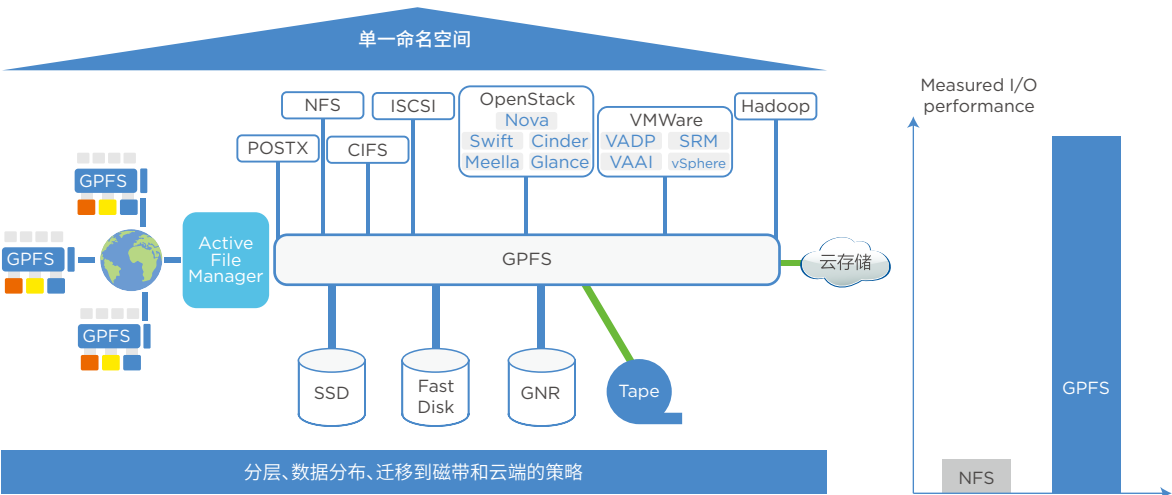
- ① 计算量和数据量巨大，可靠性要求高
- ② 大容量存储系统和并行文件系统

基因序列比对，组装；分子对接，药物筛选

- ① 计算任务相对独立，进程间通信量很小，计算网线速千兆或非线速万兆即可
- ② 较大内存要求，较高的数据 I/O 吞吐量
- ③ 集群系统：普通两路服务器节点 SRXX 系列服务器，高密度刀片服务器 SD530 V3/SD550 V3，或 SMP 胖节点 SR8XX/SR9X，或者水冷服务器 SD650 V3
- ④ 大容量存储 + 并行文件系统

联想并行存储解决方案

使用 DSS-G 的联想并行存储解决方案，可用于在站点间快速访问、分层和迁移生命科学数据



- 测量数据表明，可显著提高性能，进而缩短生命科学分析时间
- 可以在联想合作伙伴的帮助下，通过联想 HPC 团队实施基于 IBM GPFS 的归档解决方案

联想并行存储解决方案

DSS高性能并行存储联想解决方案	特点
 • 冗余双路2U机架式服务器 • 84插槽JBOD存储设备 • 首选高速IB/OPA网络技术IBM GPFS 具有管理GUI并支持CIFS/SAMBA • 支持NVMe • 支持SSD和HDD混合 • 超过800GB/s的持续读写带宽 • 优异无与伦比的单流性能	• 并行访问 • 冗余组件 • 可扩展性能和容量(PB) • 存储分层 • 进行加密以提高安全性(仅限GPFS) • 开发人员API(GPFS) • 元数据优化 • 分钟级数据重构技术

- 以智能群集的形式交付预先集成的 GPFS 存储服务器
- 大容量和接近线性的扩展能力，面向大型环境
- 分摊 RAID 重建时间，降低网络受到的影响

分子动力学

- ① 集群系统：高密度刀片服务器集群，4GB/Core，两个本地硬盘
- ② 线速无阻塞 InfiniBand 高速网络

生命科学 HPC 参考架构

联想生命科学解决方案



联想加速生命科学工具 GOAST

联想加速生命科学工具——GOAST 基因组优化和可扩展工具

Genomics

Optimization

And

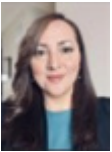
Scalability

Tool

双管齐下的解决方案:

1、GOAST架构

2、GOAST可扩展工具



Mileidy Giraldo, Ph.D.
Global Lead, Genomics Research and Development
Lenovo HPC & AI

Previous affiliations: Bioinformatics Scientist (15+ years) NCBI, NIH. National Cancer Institute (NCI), NIH National Institute of Allergy and infectious Diseases (NIAID),NIH

Expertise:

- Large-Scale Sequence Analysis
- Next-generation sequencing (NGS)
- High Performance Computing for Life Sciences
- Large-scale Biological Databases
- Biomarker discovery
- Cancer Genomics
- Infectious Diseases

Current Focus:
Dr.Giraldo leads Lenovo's Genomics R&D Group focusing on evaluating the latest HPC technologies to bring innovation and reduce time to scientific insight. With a vertical team of researchers, engineers, and expanding partnerships, Dr. Giraldo's group has built robust capabilities around Population-level Next Generation Sequencing and Precision Medicine.

GOAST 架构在 GATK4 上的性能表现

序列读取

基因组学分析

Variants



	WGS	WES
Typical genomics user	150hr.	4hr.
Intel's Original Architecture (NVMe + FPGA)	10.8hr.	25min.
Lenovo's GOAST Architecture (Open source software, No NVMe, No FPGA, no specialty hardware)	5.5hr.	4min.
GOAST Speed up	27X	40X

First solution provider to obtain



基因组数据中心可以处理

4.5全基因组 或者 343全外显子

4.5 Whole genomes or 343 Exomes

在GOAST架构上一个节点一天

GOAST 可扩展工具

我们如何为生命科学确定 HPC 的规模，并针对不同的容量、工作负载组合、预算和需求扩展 GOAST 体系结构。

联想生命科学人工智能解决方案

人工智能在生命科学领域

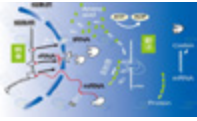
AI 技术在基因数据分析和医疗影像识别中的应用



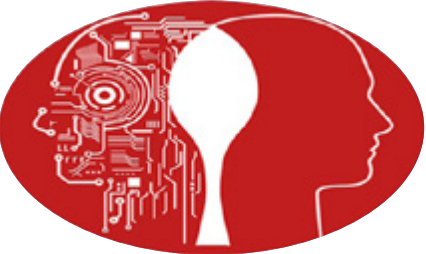
医疗影像识别



基因突变和疾病诊断

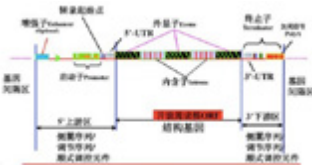


基因表达分析

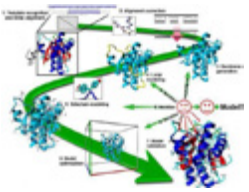


面面向临床

面面向科研



鉴别基因中的组分



蛋白同源性分析和结构预测

算法和框架

- 合理的算法和模型设计十分关键
- 可以把算法设计/优化问题转化为深度学习模型设计/优化问题，从而借鉴深度学习其它领域的成果
- 可以利用 TensorFlow，Caffe 等框架来简化模型/算法开发时间

数据

- DNA 测序价格的持续走低为我们提供了海量的数据
- 传统的统计分析方法跟不上数据爆炸式增长的节奏
- AI 为我们提供了挖掘数据价值的更佳方式

计算能力

- V100 等 GPU 加速芯片可以为模型训练提供强力支撑
- T4 等 GPU 加速芯片和设备可以在推理时加速数据处理，应对海量数据挑战

AI 开发平台面临的问题与挑战

医疗行业 AI 业务在扩大，投入在持续增加



市场发展变化快
业务更新频繁



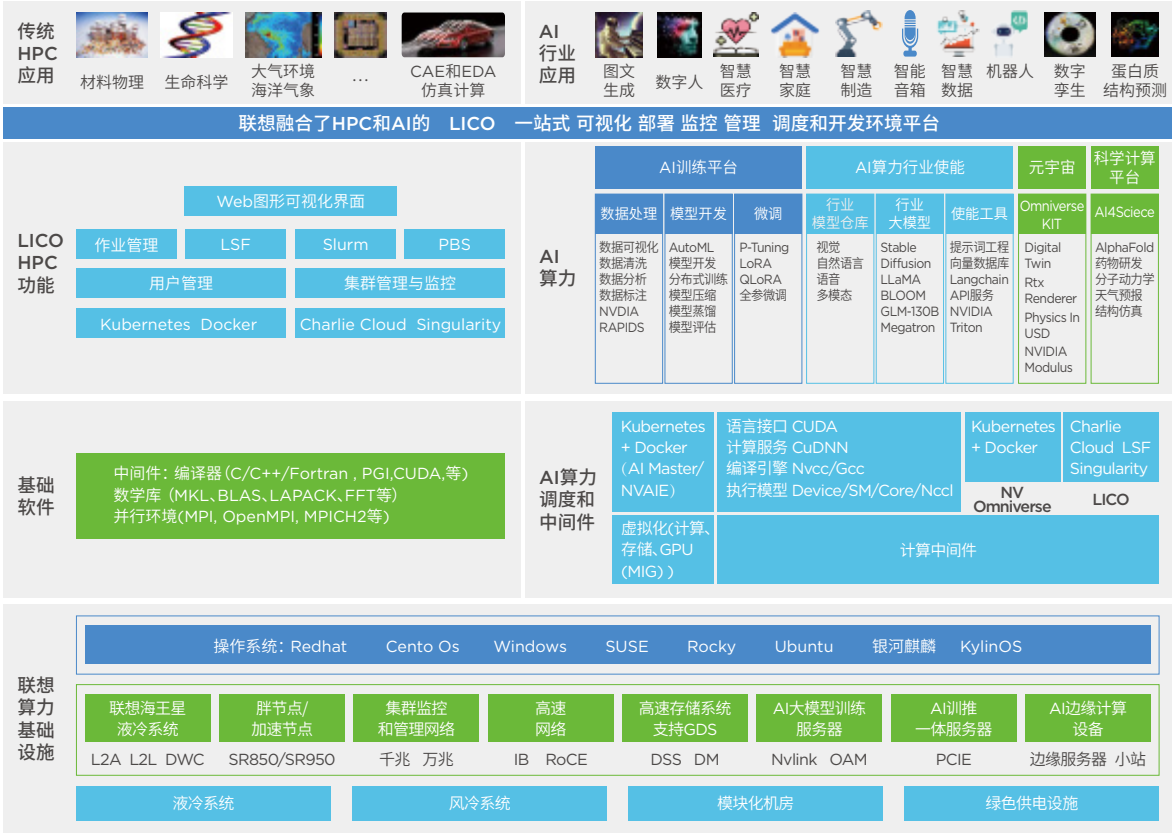
数据种类多，数据量大
潜在价值高



业务复杂，计算资源多
应用种类多

联想人工智能全栈产品

联想算力平台全栈全域框架



联想 AI 算力全栈全域框架



联想高性能计算产品 及特点介绍

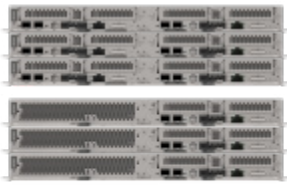
基于 Intel 新一代处理器 ThinkSystem V3 服务器



- ThinkSystem ST650 V3



- ThinkSystem SR630 V3
- ThinkSystem SR650 V3
- ThinkSystem SR850 V3
- ThinkSystem SR860 V3
- ThinkSystem SR950 V3



- ThinkSystem SD650 V3
- ThinkSystem SD650-I V3

全新管理系统	增强安全性	领先的性能	专为 AI 设计	场景化算力覆盖
更快的算力	更可靠的算力	行业领先的性能	排名第一 高性能计算机提供商	V3 全新产品系列
• 财务分析快 70% • AI 模型开发速度快 2 倍	• 连续 8 年在所有 x86 服务器中正常运行时间排名第 1	• 从 2001 年开始累计创造了 800 多个性能世界纪录，一直位居世界纪录榜首	• 用超级计算机帮助客户解决人类面临的最大挑战	• 涵盖从 SMB 到超大规模数据中心的各种工作负载

一个久经验证的产品组合，在您需要的时候提供性能！

联想高性能计算硬件和环境配套

高性能计算服务器产品

Lenovo ThinkSystem

Lenovo® ThinkSystem™ 是对 Lenovo ThinkServer® 系列的一个有益补充，它极具经济效益，以创新的方式实现最大可用密度。经过优化的 ThinkSystem 可处理大量需要灵活性的工作负载，通过快速采购、部署和灵活的选件加快业务速度。这款简单但功能强大的系统可处理各种应用程序，从技术计算、GRID 部署、分析工作负载到大型云和虚拟化基础架构，几乎无所不能。



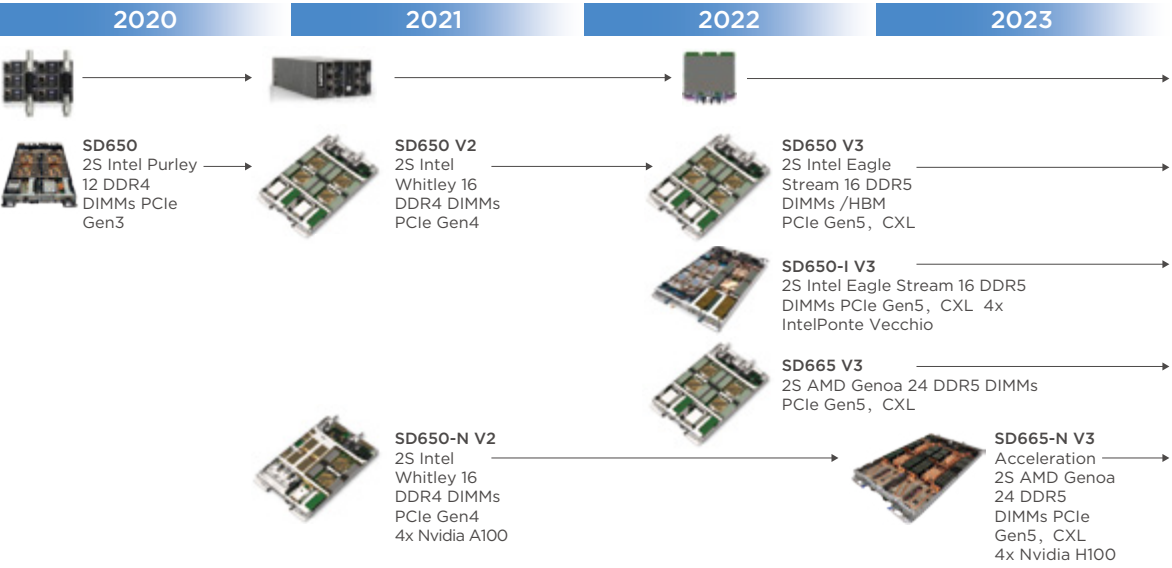
HPC&AI 2023/24 Server Portfolio

ThinkSystem	SR635 V3/ SR655 V3	SR650 V3/ SR665 V3	SR630 V3/ SR645 V3	SD665 V3	SD665-N V3	SD650 V3	SD650-I V3	SR675 V3
								
Form Factor	1U/2U	2U	1U	6U12N	6U6N	6U12N	6U6N	3U
CPU Sockets	1S AMD 320W	2S Intel 350W/ 2S AMD 320W	2S Intel 350W/ 2S AMD 320W	2S AMD 400W	2S AMD 400W	2S Intel 350W	2S Intel 350W	2S AMD 320W
DIMM Slots				24	24	16	16	24
2.5" Drive Slots	up to 10/32	up to 40	up to 12	up to 4 (NVMe)	up to 2 (NVMe)	up to 4 (NVMe)	up to 2 (NVMe)	up to 8 (NVMe)
Acceleration	3x SW 150W 3x DW 300W or 6x SW 75W	3x DW 300W or 6x SW 150W	2x SW 150W or 1x FHHL 150W	n/a	4x 600W SXM (Hopper)	n/a	4x 600W OAM (Ponte Vecchio)	4x DW 300W or 8x SW 150W 8x DW 300W 4x 400W SXM4
Cooling	Air (A4;A2 full perf.)	Air (A4;A2 full perf.)	Air (A4;A2 full perf.)	Water (A4;W5 full perf.)	Water (A4;W5 full perf.)	Water (A4;W5 full perf.)	Water (A4;W5 full perf.)	Air (A4;A2 full perf.)
Cabling	Rear	Rear	Rear	Front	Front	Front	Front	Front/Rear
Typical HPC Workload	HPDA AI Inference Technical Computing	HPDA AI Inference Technical Computing	Technical Computing	Supercomputing	Exascale Supercomputing AI Training	Supercomputing	Exascale Supercomputing AI Training	AI Training AI Inf. M&E Technical Computing

水冷服务器

液冷刀片：联想直接水冷刀片系统是业界最新的高密度计算解决方案。以其灵活、开发和简化数据中心的特点为高性能计算、网格计算、仿真和分析以及大规模云计算等应用提供了优秀的基础架构。

Lenovo HPC&AWATERCOOLEDRoadmap



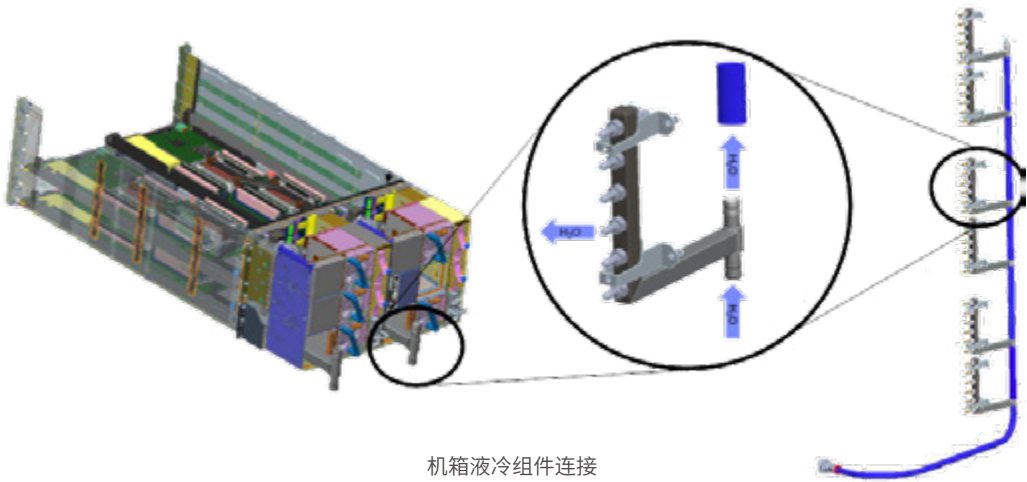
- 第五代联想 Neptune 温水水冷技术，全并联水路连接，节点的 98% 热量通过水冷散热
- 单个 42U 企业级机柜支持 72 个节点，并且支持 Intel 、NVIDIA 和 AMD 的多种技术
- 利用业界领先且经验证的基础设施设计和标准化交付（LESI）实现液体冷却，全球部署超过 60000 个节点

直接水冷刀片系统面对目前数据中心的局限性进行密度和性能优化而设计，可以在一个传统 42U 机柜内轻松部署 72 个节点，以及相应的网络设备。

直接水冷刀片系统设计运行于“温水”环境，节点冷却进水温度在 18-50 摄氏度。这意味着用户可以大幅降低 IT 系统的制冷能源消耗，最大化提高能效，降低 TCO 总拥有成本。

液冷型机箱：由于采用液冷设计，使得刀片机箱可以抛弃传统刀片机箱中转为刀片散热而提供的大量散热风扇，这一方面降低了风扇消耗的能耗，另一方面大幅降低了系统运行的噪音，测试表明在 0.5 米距离上，刀片机箱的运行噪音仅为 56 分贝。

机箱后部已取消散热风扇，只有六个电源模块，管理模块和供水、回水接口。

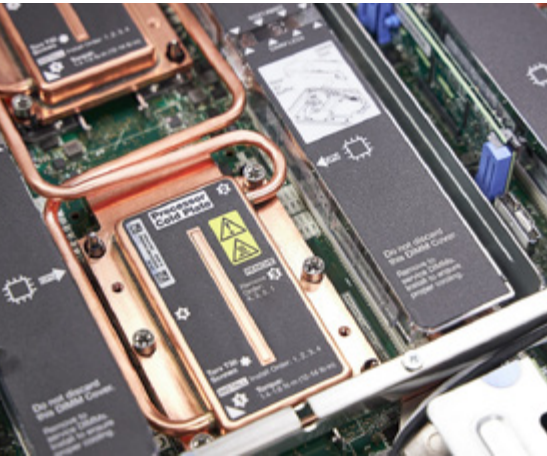


机箱液冷组件连接

节点插入机箱后，直接通过快速接头与液冷组件连接。

一个机柜中六个机箱的液冷组件直接连接，构成整个机柜的供水与回水通道，两个管路再分别连接到 CDU（冷水分配单元）的供水与回水管路，构成完整的二次侧水循环。

液冷节点：



液冷节点内部视图

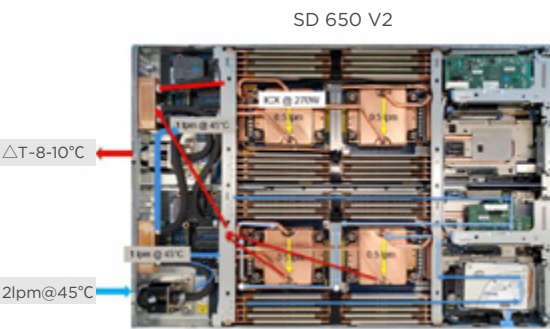


机柜视图

1 台液冷机箱内部可以安装 12 个液冷节点，每两个节点为一组。

液冷是“冷板式”设计，即发热部件不直接接触液体，而是先把热量传输给装有液体的铜制冷板，再通过液体循环带出设备。

节点的主要发热部件：CPU、内存和 I/O 板卡，本地硬盘，机箱电源都有导热件通过，节点中大部分的热量都通过冷却液体带走。



液冷节点后视图

每对内存中间插入一个金属导热件，以高效吸收热量，并传递到水冷管件。每对内存条上有一个塑料盖，用以确保没有空气流入来吸收热量。

液冷节点总体效果：通过在机箱和节点上的多重液冷设计手段，ThinkSystem 液冷系统可以将高达 98% 的发热量通过液冷系统带走，而剩余的 2% 左右的热量只需要机房环境空调即可轻松解决。

整机柜交付：联想 ThinkSystem 采用整机柜交付方式，一个机柜提供 6 个水冷机箱，可安装 72 个水冷节点，管理网和监控网的千兆连接在机柜内实现，整个机柜对外连接最多可达 72 个 25Gb 端口和 72 个 200Gb IB /400Gb IB 高速数据连接。



整机柜交付水冷单柜

联想问天服务器

2023 年 2 月 8 日，联想在京举办以“智算无限 思天逐梦”为主题的服务器品牌战略暨新品发布会，正式推出以联想问天+联想 ThinkSystem 双品牌布局的服务器战略及系列新品，通过品牌双翼强化新 IT 全栈全周期服务，更好地为金融、能源、教育等各行各业提供智能化转型的数字底座，进而为经济高质量发展贡献力量。

联想问天+联想ThinkSystem重新定义新算力

联想问天 + 联想 ThinkSystem



联想 ThinkSystem



联想问天



高性能

联想V3服务器性能提升**53%**,
IO带宽**翻倍**,传输延时**减半**



高扩展

最高扩展 **8** CPU, **128** DIMM
18 PCIe **48** HDD **32** NVMe



高可靠

双Flash冗余设计, **11**项 **NCTC**
国家级认证, 连续8年 **ITIC No.1**



低能耗

海神技术使PUE低于**1.06**, 总能耗
降低**40%**, TCO降低**47%**

联想服务器新品发布

Lenovo
ThinkSystem

AI 服务器
SR 系列
WA 系列



ThinkSystem SR675 V3
8 卡 PCIe

标准机架服务器
SR 系列
WR 系列



ThinkSystem SR850/860 V3
ThinkSystem SR950 V3



ThinkSystem SR665/655 V3



ThinkSystem SR630 V3
ThinkSystem SR645/635 V3

高密度服务器
SD 系列



ThinkSystem SD650/650-I V3
ThinkSystem SD665/665-N V3

塔式服务器
ST 系列



ThinkSystem ST650 V3

联想问天



联想问天 WA7880 G3
8 卡 NVLink



联想问天 WA5480 G3
8 卡 PCIe



联想问天 WR5220 G3



联想问天 WS5420 G3

高性能计算 GPU 产品

协处理器是提升应用性能的重要手段，联想高性能计算系统的 GPU 节点通常选用 NVIDIA GPU。协处理器的性能提升效果是严重依赖于应用本身代码的改造和数学模型 / 算法的特点的，因此不能保证协处理器会在任意应用和算法上都带来与标称性能相当的应用性能提升幅度。

高性能计算存储产品

联想高性能存储系统

联想HPC存储解决方案概览

DM	DSS-D	DSS-G
 高性能NAS 适用于 EDA/CAE/VDI	 高性能并行存储 适用于 学院/研究所/企业	 大规模分布式存储 适用于 中大型高性能计算中心
联想拥有全面领先的高性能存储方案，适用于不同的HPC应用场景		

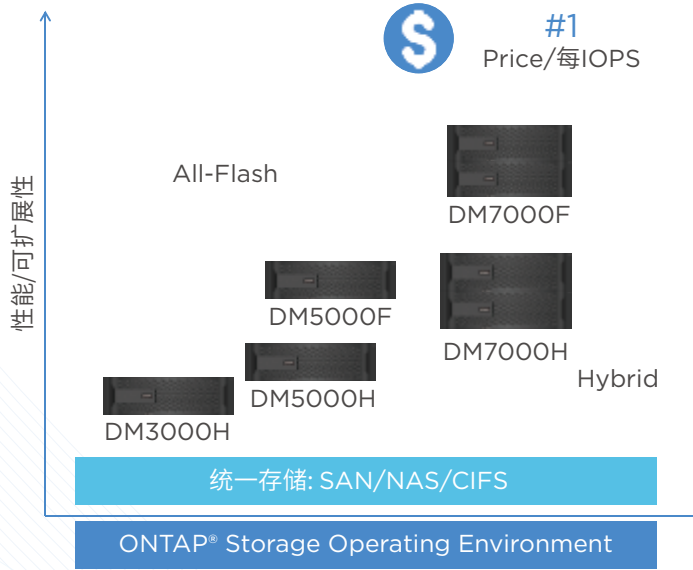
联想DM系列统一存储(iSCSI/NAS/SAN)

极致性能

- ONTAP数据管理软件
- 延迟敏感型应用
- 云分层/关键数据远程容灾
- 支持集群, 12对HA pair
- 650k随机I/O(8k), 单套最大4.8M

最佳用途

- 人工智能
- 大数据分析
- EDA工程设计



Lenovo (DSS-D) 存储系统

Lenovo DSS-D 整合了联想凌拓的高性能存储和业界领先的 GPFS 文件系统，以及灵活扩展的存储系统，可以为客户带来高性能、高带宽和高容量的存储系统解决方案。DSS 是基于联想 DE 系列存储和 GPFS 技术的软硬件一体化解决方案。

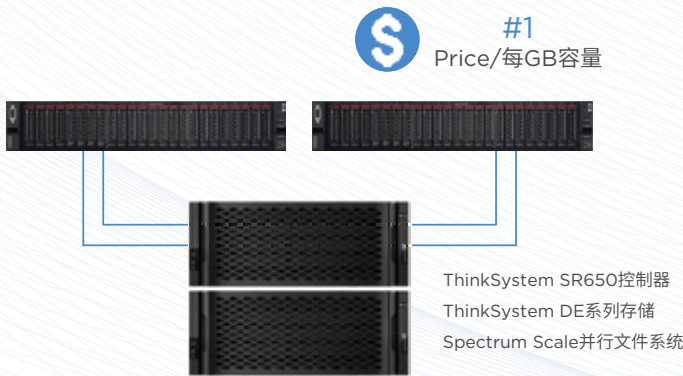
联想 DSS-D 系列高性能存储

高性价比

- 高存储密度, 4U/60硬盘
- 卓越的性能
- 单个构建模块6.72PB容量
- 单个构建模块>20GB/s
- 灵活的构建模块, 支撑横向扩展

最佳用途

- 高性能计算
- 数据密集型场景
- 高性能NAS
- 适用于中小型HPC/AI场景



Lenovo DSS-G

联想的 DSS (Distributed Storage Solution)。DSS 基于内置先进 RAID 技术的 GPFS 软件，和以可靠性和性能著称的联想 ThinkSystem 服务器，不需要专用磁盘阵列控制器，简化了存储系统结构，提高了可靠性。同时 DSS 不像普通的分布式存储系统需要使用多份数据拷贝方式来保证数据安全性，保证了磁盘利用率与基于磁盘阵列的解决方案相同。

联想 DSS-G 分布式存储

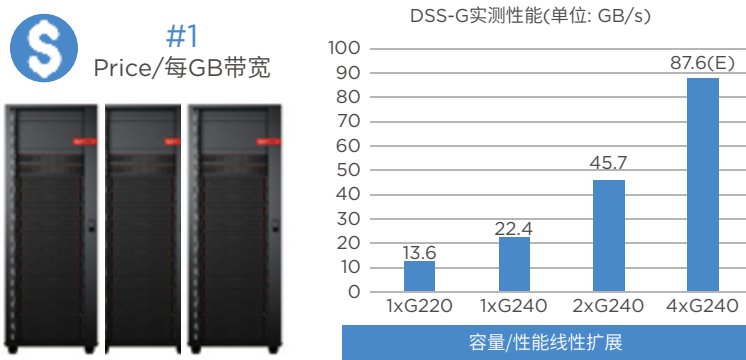
DSS-G是适用于HPC/AI/大数据/云的可扩展存储解决方案

极致性能

- 卓越的性能
- 快速重建, 重建时不影响应用性能
- 实测性能超过 2500GB/s
- 灵活的构建模块, 支撑横向扩展
- 线性容量/性能扩展
- 多种数据访问方式: 客户端/块/文件/对象

最佳用途

- 大容量分布式存储
- 极致存储性能需求
- 适用于中大型HPC/AI场景



DSS-G 的另一个优势是秒级的 rebuild 时间，这使得磁盘故障和修复过程导致的系统性能下降，预报作业时间延长等现象不再存在。

借助于 GPFS 方便灵活的分级存储功能，选择不同配置的 DSS 组合在一起，就可以获得一个具有优秀分级存储功能的存储集群。

磁盘盘组中的磁盘和 2 个存储节点上的多个 SAS 接口构成多个环路，磁盘故障不会中断 SAS 环路，不会导致其余磁盘不可访问。磁盘盘组本身无控制器电路，只提供磁盘安装和环路连接，以及冗余的磁盘供电和风扇，也不构成单点故障点。

由于我们提供了热备空间和分布 RAID，我们的 rebuild 方式更加先进，从被动变成主动。每台 DSS 的一半盘中同时损坏三块磁盘后开始进行 critical rebuild，现有配置下需时 20-50 秒，通常为 30 秒。此过程不须更换磁盘，整个存储系统基本不会因为磁盘故障有性能损失，整个过程全自动化。

内置世界领先的 Decluster 技术，硬盘故障时，所有磁盘参与重建，将重构过程中的读写操作分散到各磁盘中，实现并行 I/O，从而提高重建速度及减小对存储性能的影响。块数据支持 8+2P，8+3P 的 Reed-Solomon 算法，元数据支持 3 副本，4 副本复制。

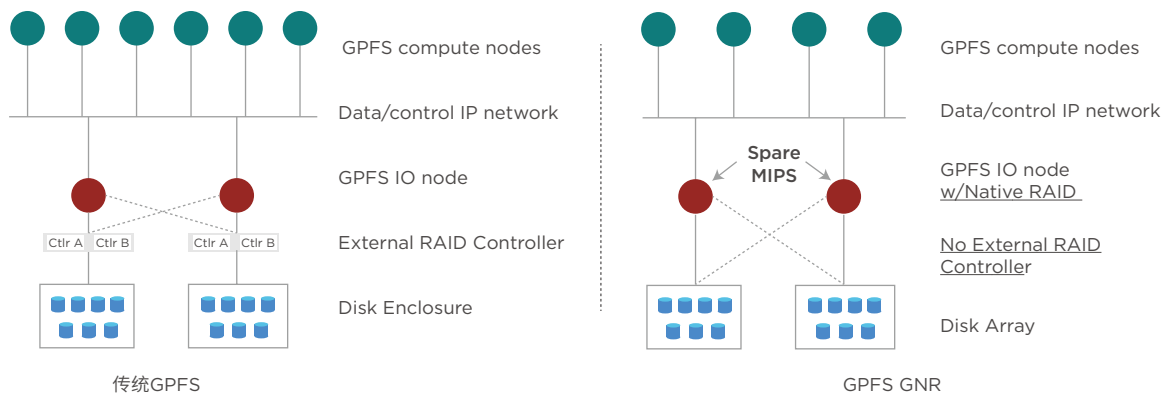
在数据安全性方面：

软件 RAID 模式支持同时坏掉 3 块硬盘而不中断业务，允许一个扩展柜损坏而数据不丢失、应用不中断；

自动检测硬盘响应时间，当出现 timeout，I/O error 等情况能自动判断是硬盘本身还是路径原因导致，与此同时，系统重建由于错误而丢失的数据信息；

自动检测和维修硬盘的扇区错误，端对端的检错功能，完全避免磁盘“无声数据失效”。

基于 GPFS GNR (General Native RAID) 技术的存储系统：



GPFS GNR 较传统的 GPFS 而言，省去了磁盘阵列控制器。而是采用 JBOD 的方式直接连接磁盘。这种连接方式不仅节约了成本，还在 I/O 性能和可靠性方面有更好的表现。尤其是在磁盘出故障需要重构时，GPFS GNR 的重构时间较传统磁盘阵列有明显优势。

较传统 RAID 而言，GPFS 采取 d-cluster 方式组织磁盘。这种方式能大大缩短磁盘损坏后系统重构的时间。上图分表描述了传统 RAID1 和 GPFS GNR RAID1 配置。黄色磁盘为预留的 spare 盘。在传统 RAID1 系统重构过程中，所有数据将从同一磁盘拷贝到 spare 磁盘。所需时间较长。而 GPFS GNR 中，每一对数据拷贝都分布在不同磁盘对中。当某个磁盘失效时，可以从各磁盘中读取相应的数据块，复制到预留的数据块中即可完成重构。其优势在于能将重构过程中的读写操作分散到各磁盘中，实现并行 I/O，从而缩短重构时间。

GPFS GNR 运行在 x86 平台下，采取的是 8+3p 模式。即 8 个数据盘加上 3 块校验盘。每份数据保存有 4 份拷贝，采用 d-cluster 的方式分布在 47 块磁盘上。当一组磁盘失效时，GPFS GNR 能很快重构，几乎不影响用户数据访问。即使有 3 组磁盘同时失效（非常少见），GPFS GNR 也能很快将其恢复在 2 组磁盘失效的模式，然后进行重构。GPFS GNR 在 x86 平台下，较传统的 RAID 而言在系统重构方面有很大优势。

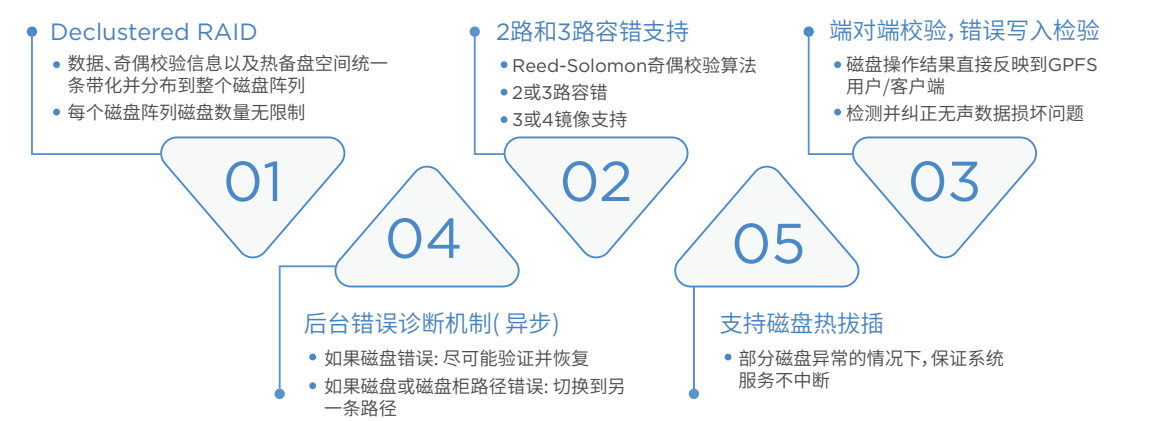


研究发现，现行存储系统由于数据传输路径较长。从主机 CPU，到主机内存，光纤卡，存储控制器，最终到磁盘。在数据的传输和存储过程中，有一些细微的数据错误（如比特位翻转）很难通过常用的工具检测到。GPFS GNR 能提供企业级的点到点数据校验检测。存放在磁盘上的所有数据块都有相应的校验位。GPFS GNR 能对数据在传输和读写过程进行全面的检测，保证数据有效性和可靠性。

GPFS GNR 提供统一管理接口。磁盘重构，负载均衡，点对点数据校验检测等都有后台自动进行。很少需要用户手动干预。GPFS GNR 还能动态配置系统重构对整个磁盘带宽占用的比例，能保证关键应用的磁盘带宽。

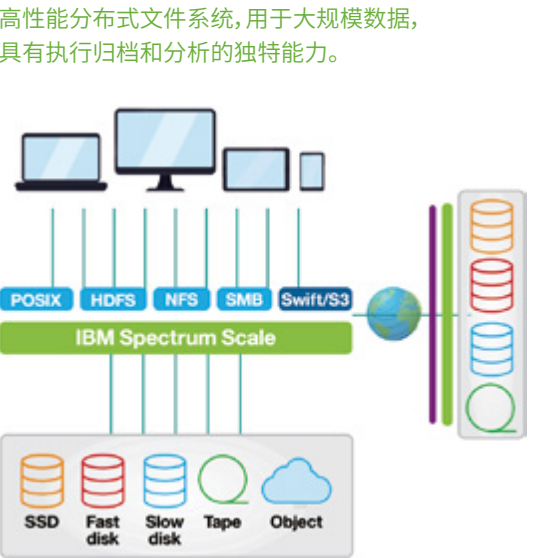
GPFS GNR 提供全面的磁盘管理功能。能检测并管理磁盘连接，介质错误，磁盘失效等问题。针对介质错误或者点对点校验失效等错误，GPFS GNR 能在后台自动修复。根据磁盘工作状态，记录其“健康”状况，能在磁盘失效前，进行预警。

GNR 功能概览：



Spectrum Scale - 高性能, 高度可扩展的分布式文件系统
DSS-D和DSS-G集成了最新的Spectrum Scale分布式文件系统

- 消除性能瓶颈
- 确保数据可用性、完整性和安全性
- 支持QoS(Quality of Service)
- 支持本地写缓存(HAWC)/读缓存(LROC)
- 策略驱动在线压缩
- 分布式的元数据, 支持闪存加速
- 支持最新的低延迟、高带宽NVMe
- 多协议支持
- POSIX,GPFS,NFS v4.0,SMB v3.0
 - Big data and analytics: Hadoop MapReduce
 - Cloud: OpenStack Cinder (block), OpenStack Swift (object), S3 (object)
- 支持云和本地混合
- Amazon S3,OpenStack Swift 和 与Amazon S3兼容的云存储



1. Lenovo Intelligent Cluster 的介绍

要点

- ① 复杂解决方案变得简易
- ② 行业最佳的技术，优化的解决方案设计
- ③ 享受 Lenovo 端到端支持的端到端解决方案

部署适用于技术计算、高性能计算（HPC）、存储和云环境的解决方案可能会给 IT 带来极其沉重的负担。Intelligent Cluster 利用 Lenovo 数十年的丰富经验，凭借预先集成、交付、得到全面支持的解决方案来降低部署复杂性，这些解决方案将行业最佳的组件与优化的解决方案设计紧密结合。借助 Lenovo Intelligent Cluster，客户可以集中精力最大限度提高业务价值，而不是耗费宝贵的资源来设计、优化、安装和支持满足业务要求所需的基础架构。

行业最佳的技术，优化的解决方案设计

Intelligent Cluster 解决方案采用行业领先的 Lenovo® System x® 服务器、存储设备、软件和第三方组件，可以在集成交付的解决方案中支持广泛的技术选项。Lenovo 针对可靠性、互操作性以及最高性能彻底测试和优化了每个解决方案，因此客户可以迅速部署系统并使投入工作，从而实现他们的业务目标。

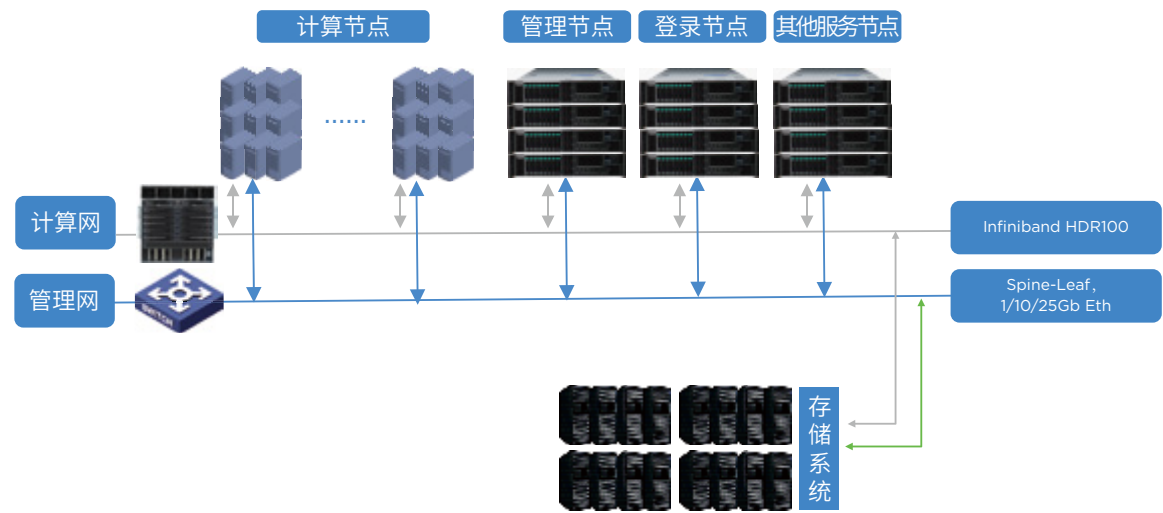
享受 Lenovo 端到端支持的端到端解决方案

Intelligent Cluster 解决方案由 Lenovo 构建、测试、交付和安装，并作为单一解决方案进行支持，而不是作为数以百计单独组件进行处理。Lenovo 提供了包括 Lenovo 和第三方组件二者、单一联系点、解决方案层面的支持，可以在系统的整个生命周期内实现最高的系统可用性，因此客户可以花较少的时间维持系统，花更多的时间实现成果。

高性能计算网络产品

高性能计算网络部分包含多个不同协议产品，他们分别是：（1）高速网络 100Gb/200Gb InfiniBand 网络：包括 36 端口、216 端口和 648 端口交换机系列；也包括 40 端口 HDR200 和 800 端口 HDR200 交换机以及 NDR400 128 端口和 2048 端口交换机系列；（2）高速网络 Intel 100Gb OPA：包括 48 端口和 768 端口的交换机；（3）16/32Gb FC 网络；（4）12Gb SAS 网络；（5）万兆网络；（6）千兆网络。

HPC 网络概况



高性能计算机房环境配套产品 - 配套的水冷和模块化数据中心

联想模块化数据中心

1. 产品概述

目前数据中心基础设施形态千差万别，但一体化、标准化及模块化已经成为数据中心建设的主流趋势。在此背景下，联想推出了模块化数据中心（MDC），它将数据中心所必需的电气、制冷、机柜、监控、消防、布线、IT 设备及操作平台等软硬件集成在一个封闭的模块化空间内，在高度集成了计算能力的同时，还大大降低了对空间和能耗的需求，在具备高可靠性的同时提供极其灵活的可扩展能力。

联想针对不同客户类型，提供了完整的模块化数据中心解决方案，包括了模块化数据中心 Lenovo Smart Aisle、小型数据中心 Lenovo Smart Row，以及微型数据中心 Lenovo Smart Cabinet 解决方案等。

智慧数据中心

Lenovo Smart Cabinet



机柜内整合的智能型基础架构，适用于小型网络间，远程站点。
容量:3.5KW, 单机柜

Lenovo Smart Row



整排机柜整合的智能型基础架构，适用于小型数据中心，灾后恢复。
容量:12.5KW, 1-6个机柜

Lenovo Smart Aisle



针对中大型数据中心基础设施的标准化、模块化解决方案，可快速部署、在线扩容、在线改造。
容量:25-200KW, 7-35个机柜

2. Lenovo Smart Cabinet 介绍

Lenovo Smart Cabinet 是联想数据中心解决方案产品，应用于微小型数据中心、办公区域等室内环境。为客户提供完整的 UPS 不间断电源、机架配电、热量管理、机柜支撑和监控管理；600mm 宽 1200mm 深 2000mm 高，42U 高机柜（IT 可用空间 29U），重量约 300Kg，满足 3.5kW 设备散热及供电需求；采用机架式 UPS，满足后备 30 分钟；全封闭机柜，带来机房的进一步节能。

具有如下特点：

- 防尘降噪，高效节能
- 智能监控
- 高度集成、节省空间
- 良好的人机界面
- 快速交付，品牌无忧



3. Lenovo Smart Row 介绍

Lenovo Smart Row 系列解决方案，是在已获得巨大市场成功的 Lenovo Smart Row 系列的基础上进行重大升级的全新解决方案，其应用仍然定位于 10-100 平米小型数据中心。符合用户对移动管理的需求、对机房环境的依赖度更低、适应更宽广的 IT 热密度范围。

为客户提供完整的 UPS 不间断电源、机架配电、热量管理、机柜支撑和监控管理，满足客户机房的基础架构需求。采用变频 12.5kW 的制冷空调，机架式 UPS，满足后备 30 分钟；采用全封闭机柜，带来机房的进一步节能；整体化方案具有高可靠、高智能和快速部署等优势；可为用户提供多种搭配方案。

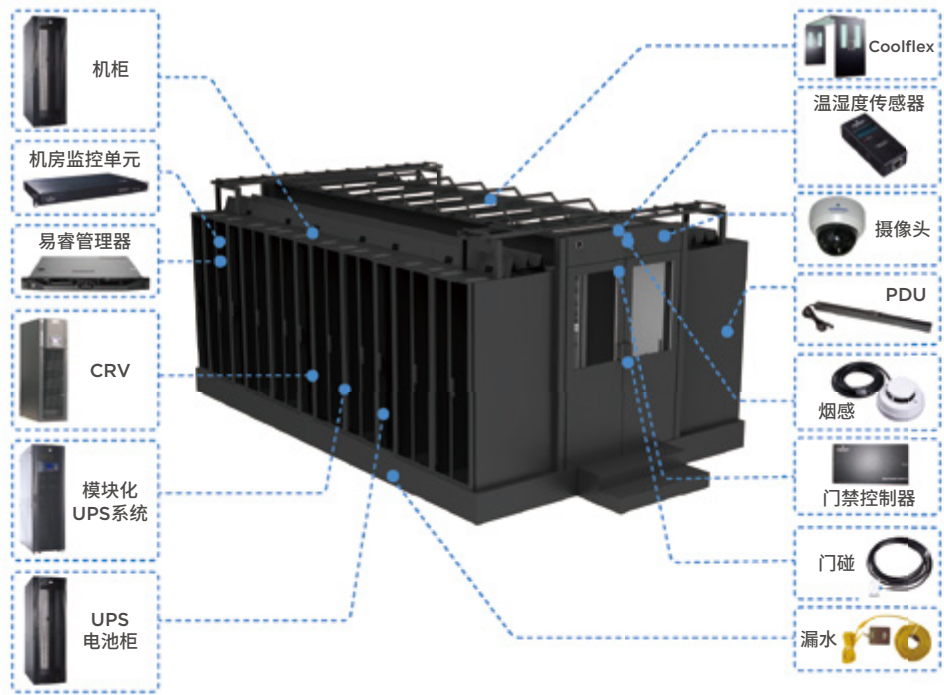
具有如下特点：

- 灵活设计与扩容
- 低建设成本
- 低运营成本
- 交付周期短
- 良好的人机界面
- 绿色节能



4. Lenovo Smart Aisle 介绍

Lenovo Smart Aisle 按照不同机柜数量、机柜密度，集成行间制冷，冷通道封闭，监控，机柜及配电并保证消防的一体化设计；供电上可灵活的采用集中 UPS 供电、分散模块化 UPS 供电等多种方式，实现 UPS 单路、双总线供电模式，并根据后期业务需要可逐渐扩容。



Lenovo Smart Aisle 模块化整体方案设计包括四大系统：供配电系统、空气调节系统、服务器机柜及附件系统和动力环境监控系统。具有如下特点：

- ① 降低系统复杂度，提高系统可靠性 — 与传统数据中心相比，模块化设计可预先验证，可复制性强，可靠性不随系统数量增加而降低；
- ② 部署速度快 — 模块化数据中心可以非常迅速地部署，通常只需要几周时间；相比之下，建造传统的数据中心通常得需要 1 到 2 年，甚至几年时间；
- ③ 降低运维成本 （OPEX） — 相同的模块可以采用相同维护规程，降低维护人员要求，共享备品备件；采用行间空调及封闭冷通道技术，可以大大降低 PUE 值，带来数据中心整体运营成本降低。
- ④ 易于部署 — 模块化数据中心可采用集装箱或预制单位构建方式部署在指定的任何位置，甚至可以移动；易于扩展 — 如果需要，可以添加更多的模块。

5. 模块化数据中心整体方案特点

- ① 高可用性：可用性超过 99.999%。通过工程产品化，实现整体设计及交付，避免各系统之间耦合问题，大大提高数据中心整体可用性。
- ② 经济性：降低机房 TCO。采用分期建设的思路，减少初期建设成本 CAPEX 占用；可提升单机柜容量，带来占地 40%，方案美观紧凑；运营成本 OPEX 降低 30%，更高效节能。
- ③ 协同性：高度一体化。各系统风格、协议统一，一站式服务，实现快速部署和在线扩容，保障客户关键业务永远在线。
- ④ 智能性：管理简单方便。通过统一监控平台实现机房智能监控、智能管理、智能供电和智能制冷，保障了对 IT 环境的智能控制。
- ⑤ 售后服务：联想提供整体机柜三年售后服务，且全国均有售后网点。

直接水冷解决方案

1. 设计规范、标准或依据

《Think M5 WCT solution implementation general guide V3》
《Thermal Guidelines for Liquid Cooled Data Processing Environments》

2. 主要设计参数和设计工况

设计指标及工况如下表：

NO.	项目	参数	备注
1	单机柜额定热负荷	33.5KW	节点按照最高功耗 CPU 算（165W） 规定其中 85% 的热负荷需要由水冷设备承担制冷
2	机柜数量	2	
3	单机柜最小热负荷	10KW	以 2*150W CPU、8*16GB MEM、1*1B、2*2TB 节点的 25% 负载估算
4	机柜进水温度	45℃	
5	单机柜额定水流量	≥36lpm	36x compute trays (dual node), 6x switches: Minimum flow rate = 36 liters per minute
6	单机柜水阻力	2.4psi	rack manifold and 6x DWC chassis
7	机柜工作水压范围	≤50psi	Pressure should be below 50psi
8	室外温度（额定点）	40℃	
9	电源形式	380V 50Hz 3N~	最大为 4 路输入
10	室外温度范围	-14℃~ 40℃	在该范围内满足全负荷制冷要求，冬季防冻

NO.	项目	参数	备注
11	冷水机组配置	不标配	实现完全自然冷
12	水质要求	参看文件	
13	机柜进出水接口形式	参看文件	
14	水过滤器要求	50μm	300 目

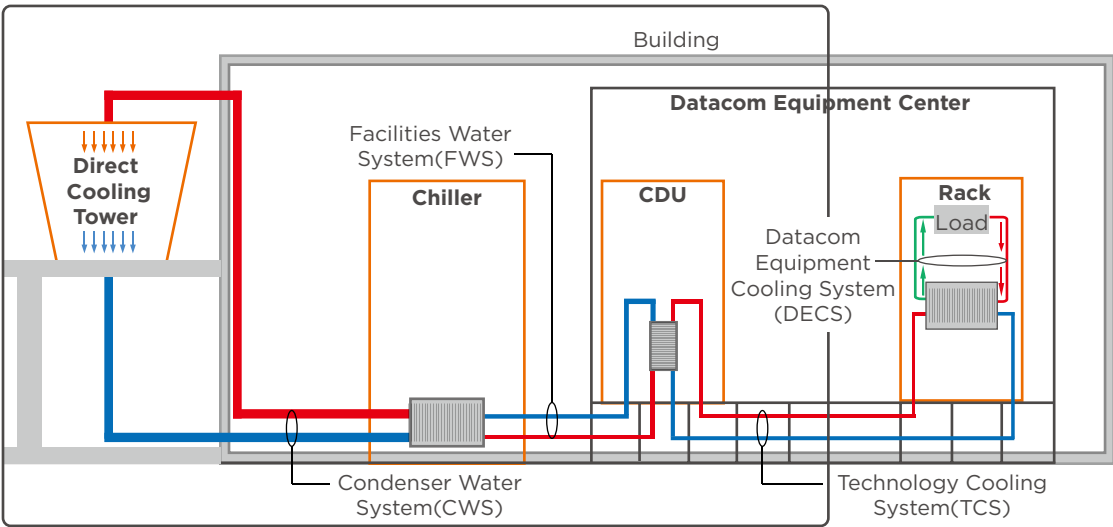
表 设计参数

3. 通用设计要求

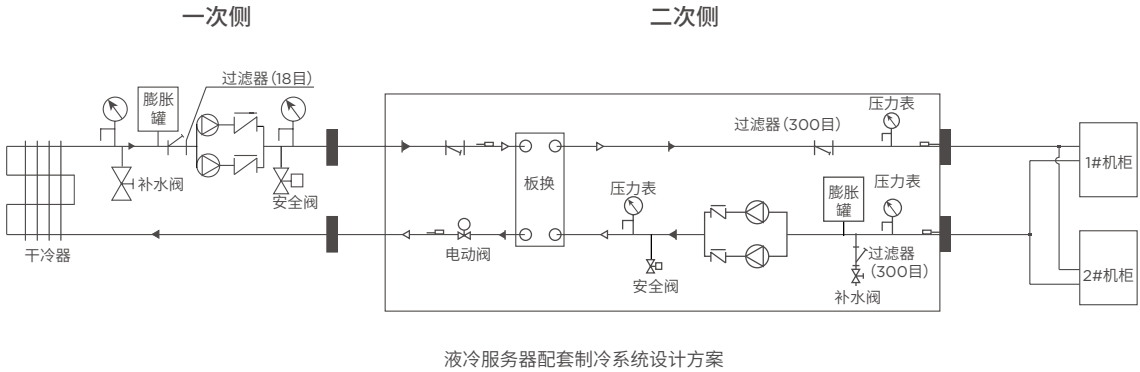
- 制冷系统具备高可靠性，满足液冷服务器长时间不间断运行。
- 制冷系统能够有效预防水温过低，防止在液冷服务器内出现凝露。
- 制冷系统具有必要的智能监控功能，能够对系统出现的异常情况作出告警和处理。

4. 制冷系统设计方案

液冷服务器配套制冷系统如下图所示，本文的设计范围包括左边框体内的所有设备或功能，具体即中间换热单元、室外干冷器等设备或功能。



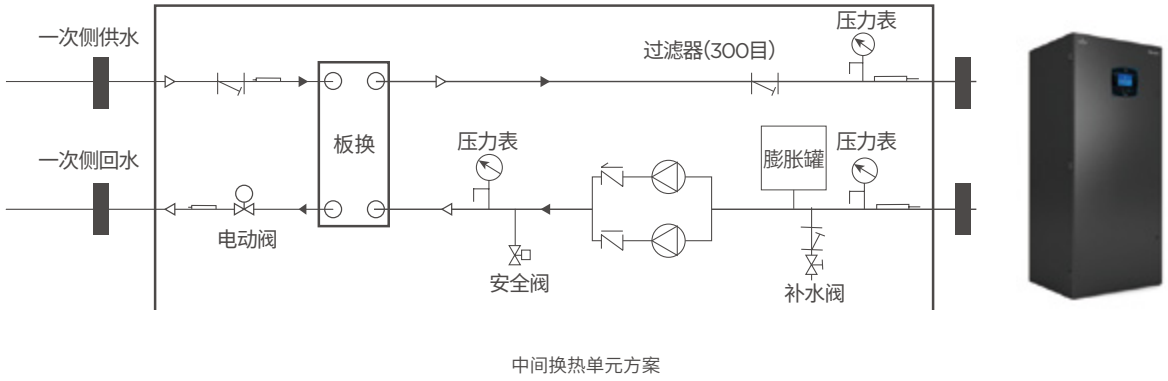
液冷服务器配套制冷系统



图所示为本建议书的设计方案，中间换热单元提供一定流量、一定温度的冷却水进入液冷服务器的换热芯体，带走 CPU 发出的热量，被加热的冷却水流回中间换热单元的板式换热器，将热量释放到室外侧循环水，该部分热量再通过室外干冷器或者冷水机组释放到室外环境中，完成对液冷服务器的热量管理。

5. 中间换热单元方案

中间换热单元在二次侧通过调节送入液冷服务器换热芯体的水温 and 水流量从而向末端机柜提供冷量，起到冷量分配的作用，此外为了便于实施最佳的室外侧方案及保持液冷服务器换热芯体内的水质，中间换热单元也起到系统隔离的作用。



6. 中间换热单元系统设计

中间换热单元主要包括二次侧板式换热器、循环水泵、过滤器、膨胀罐、阀件、管路及电气和控制系统，如所示。

7. 二次侧板式换热器

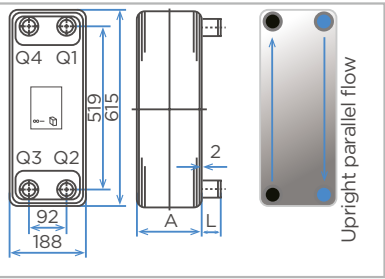
板式换热器选用钎焊式，采用国际知名品牌，不低于 SWEP、Danfoss。

为保证纯水水质，板式换热器采用 316L 不锈钢材质、换热器与外管路采用螺纹连接。

板式换热器表面敷有保温棉，消除壁板结露现象。

为了最大限度的使用自然冷源，中间换热单元按照 45℃出水、55℃进水进行选型计算，2 台机柜的散热量为 57kW（假设单机柜 33.5kW 的发热量 85% 全部被冷却水带走）。

Item	二次侧	一次侧
进水温度 °C	55	43
出水温度 °C	45	48
水流量 L/min	83	175
总压降 kPa	2.5	10.7
换热量 kW	68	



中间换热单元板换选型（一次侧工质为 30% 乙二醇水溶液）

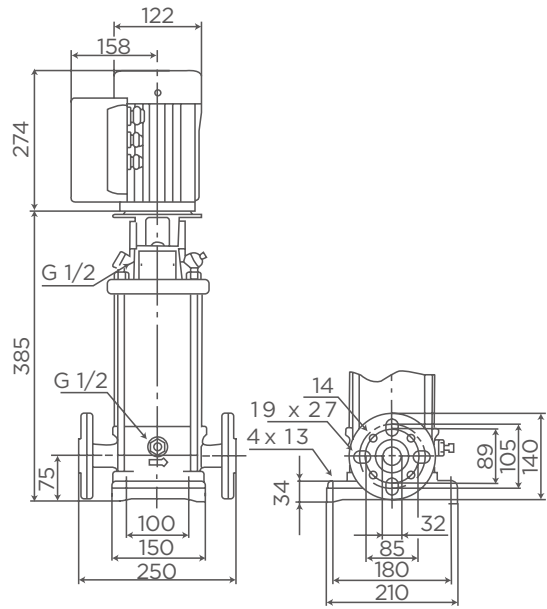
8. 二次侧循环水泵

采用一主一备双泵设计。

水泵采用高效、节能、紧凑的离心式清水泵。

水泵选用国际知名品牌，水泵电机可调速。

设计流量为 5.1m3/h，克服软管（假设 30m）、过滤器、机柜等器件的阻力，设计扬程为 33m。



二次侧循环水泵

9. 电动调节水阀

电动调节阀采用国际知名品牌：Siemens、Belimo、Johnson Controls 等品牌。

电动调节阀采用等百分比球阀形式。

执行机构应采用 24V AC/DC 供电，0.5 ~ 10VDC 或 4 ~ 20mA 信号控制，具有位置信号 (0 ~ 10VDC 或 4 ~ 20mA) 反馈。调节精度高，稳定可靠。

10. 二次侧膨胀罐

膨胀罐选用隔膜式膨胀罐，作为定压膨胀装置，提供系统定压、缓冲水泵启停产生的水锤、容纳水温变化导致的管路内容积变化。

膨胀罐选用国际知名品牌，不低于 Aquasystem 和 Varem。

11. 过滤器

过滤器选用 Y 型过滤器，不低于埃美珂、OR、天津格莱式等品牌。

12. 橡胶软管

橡胶软管材质应为三元乙丙橡胶（EPDM）。

橡胶软管满足系统可能出现的最高压力。

橡胶软管采用国际知名品牌：Paker、Hansa、Manuli 等。

13. 中间换热单元控制系统

中间换热单元采用艾默生机房空调专用控制器。

中间换热单元控制系统具有信号采集、逻辑计算、自检告警等功能，以保证系统的正常运行。

中间换热单元具有人机操作界面、数据通讯传输等交互途径，提供快捷的控制方式。

14. 信号采集

- 一次侧进出水温度采集
 - 二次侧进出水温度采集
 - 二次侧进出水压力采集
- 漏水检测采集
 - 远程温湿度采集
 - 电动调节水阀开度采集

15. 二次侧出水温度控制

通过对一次侧电动调节水阀的控制，调节一次侧的水流量，使二次侧的出水温度达到设定的要求。

16. 二次侧水流量控制

通过二次侧供回水压差、温差及进水温度对二次侧循环水泵的转速进行控制，向末端服务器提供充足的水流量。

17. 防凝露控制

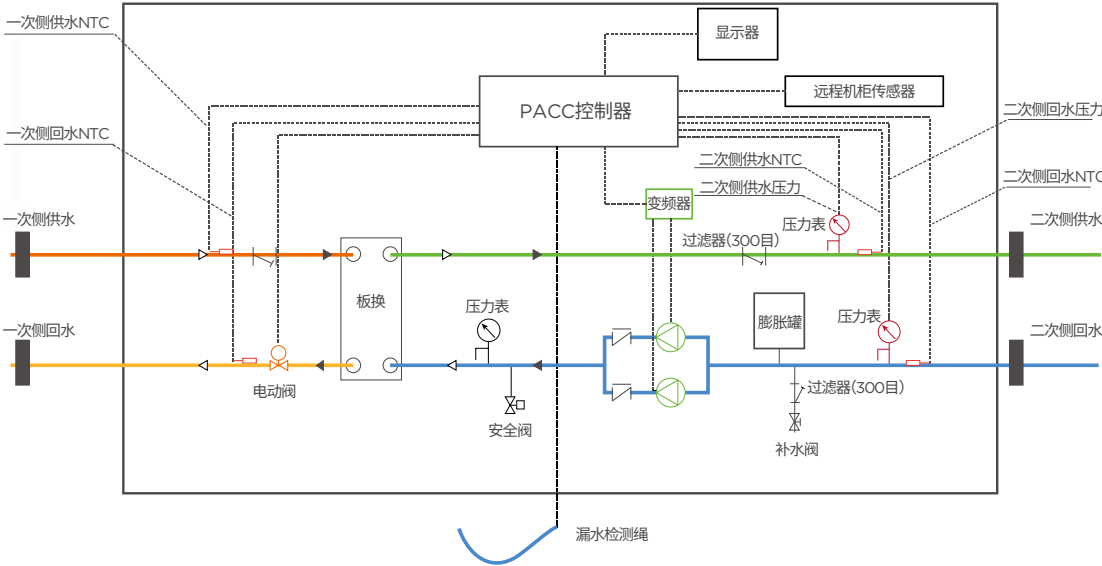
通过采集服务器周围环境的温湿度，实时监测出水温度使其高于露点温度。

18. 告警

可设置高低水温告警、漏水告警、泵故障告警等多种属性告警。

19. 水泵主备切换

当检测到运行水泵故障告警时，自动切换至备用水泵运行。



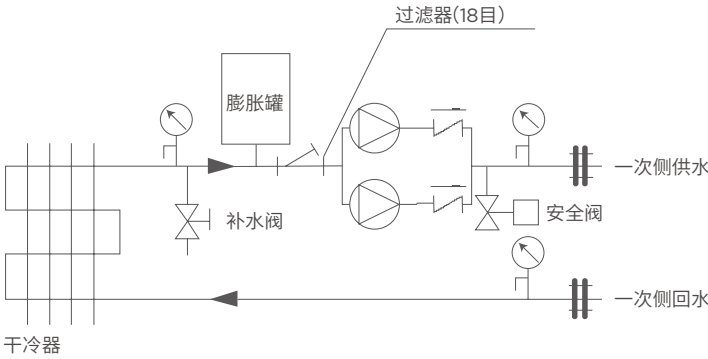
中间换热单元控制系统

20. 室外散热方案

室外散热系统将中间换热单元传递出来的热量释放到室外环境中。

21. 室外散热系统设计

室外散热系统主要包括干冷器、一次侧膨胀罐、水泵、过滤器、管路、电动调节水阀及电气和控制系统。如图所示。



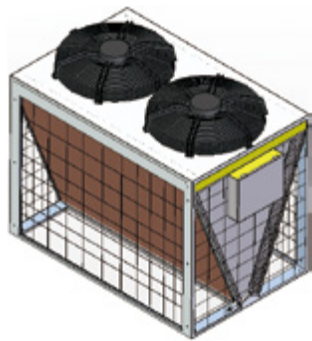
室外散热系统

22. 干冷器设计

按照室外温度 35℃条件下能够提供满负荷散热量进行设计计算。

Item	空气侧	水侧
进水（风）温度 °C	40	48
出水（风）温度 °C	43.6	43
流量	54,000m³/h	203L/min
换热量 kW	65.3	

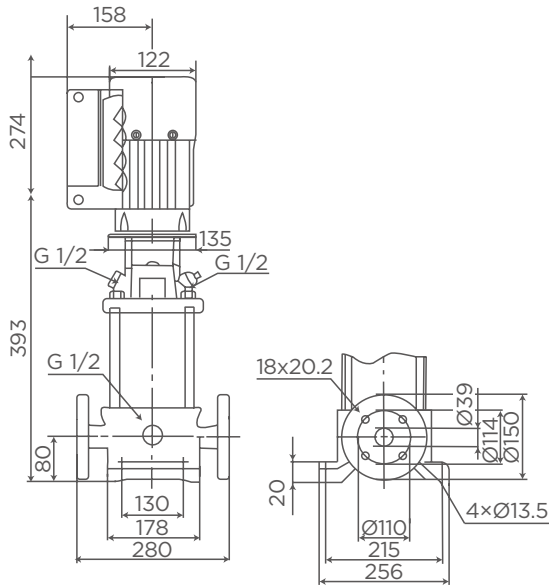
40℃室外温度干冷器设计（30% 乙二醇水溶液）



干冷器

23. 一次侧循环水泵

采用一主一备双泵设计。设计流量为 12.2m³/h，克服管路、过滤器、干冷器等器件的阻力，设计扬程为 22m。其他要求同二次侧水泵。



一次侧循环水泵

24. 室外散热控制系统

室外散热系统的运行逻辑要尽可能的使用自然冷，以二次侧的出水温度为控制目标，一次侧循环水泵、干冷器风机、水阀按照由高到低的优先级进行调节，尽可能的降低能耗。

25. 水质处理

二次侧循环水需要线监控水质、PH、硬度、腐蚀率等参数，一旦超出水质要求，就定期更换或者采用一部分水旁滤并进行补水的方法使总体水质满足要求。对二次侧系统进行补水时，先经过一定的水处理，经过软化、去离子、反渗透（RO）等技术处理，由专业的水处理设备实现。

26. 干冷器和闭式冷却塔的对比

干冷器和闭式冷却塔都是典型的室外散热设备，最本质的不同在于干冷器直接与室外空气进行热交换，而闭式冷却塔是在干冷器的基础上增加了一套喷淋设备，利用喷淋水的蒸发起到增强换热的作用。鉴于该特点：

- ① 闭式冷却塔需要专用的喷淋水供水系统，包括水泵、管路等，以使其能利用源源不断的喷淋水来制冷，同时该系统还需要水处理设备，以保证喷淋水不会在换热器上产生结垢、腐蚀等问题。因此闭式冷却塔的优点在于强化了换热，在室外温度较高时也能带来相当的制冷量，尤其是在炎热干燥的地区效果更明显，但相应的也带来了明显的缺点，冷却塔需要持续供应喷淋水，需要有水源，并且也产生了用水的费用，如果水质达不到要求，则需要定期的除垢换水，日常的维护工作较多，控制较复杂，另外在冬季低温时，还要将喷淋系统放水防冻。
- ② 干冷器没有喷淋系统，因此不存在由此带来的工程施工、额外设备投入等问题，不需要提供水源以及不会产生持续的用水费用，干冷器日常的维护也较简单。但由于干冷器没有喷淋水强化换热，冷源仅来自于与空气的强制对流换热，在室外温度较高时，需要更大的换热器才能达到更大的制冷量。

对于液冷服务器的散热需求，从应用上来看，当机柜数量较少即散热需求较小时，适合使用干冷器，投入小，施工维护简单；当机柜数量较多即总体散热需求较大时，需根据应用场地的环境温度、水源有无、占地面积等综合考虑选择冷却塔或是干冷器。

模块	设备名称	参数
中间换热单元 系统部分	板式换热器	68kW（一次侧 55/45℃，二次侧 43/48℃）
	变频水泵	5m ³ /h, 330kPa 2-9m ³ /h 调节
	膨胀罐及安全阀	2L, 泄压阀
	补水阀	1 1/2
	止回阀	DN40
	水过滤器	300 目（50μm），DN40
	水过滤器	18 目, DN65
	球阀及执行器	DN50 螺纹, Kvs=40m ³ /h
	橡胶软管	EPDM, 1", 15m
中间换热单元 控制部分	快速接头	1"
	控制器 & 显示器	PACC 控制器
	NTC	一次侧供回水，二次侧供回水
	远程温度传感器	1# 机柜, 2# 机柜
	压力传感器	二次侧供回水
	漏水检测绳	
室外干冷器	配电器件	主空开、变频泵空开 / 接触器、汇流排、成套线缆
	风机	27000m ³ /h, 50Pa
室外水力模块	翅片管换热器	
	水泵	12.2m ³ /h, 220kPa
	单向阀	DN65
	补水阀	1 1/2
	膨胀罐及安全阀	8L
	水过滤器	18 目, DN65
室外控制系统	压力表	
	控制板	PACC 控制器
	温度传感器	室外温度
	配电器件	主空开、泵空开 / 接触器、风机空开 / 接触器、冷水机组空开 / 接触器、汇流排、成套线缆

表 关键设备清单

联想高性能计算软件

联想智能计算平台 LICO

联想 LICO（Lenovo Intelligent Computing Orchestration）智能计算平台是联想基础设施方案业务集团（ISG）基于用户实际需求和联想大量行业应用经验所开发的一套高性能计算中心管理软件，是符合国家战略和拥有自主知识产权的商业软件，同时 LICO 是联想高性能计算产品 x9000 的核心。LICO 是联想 HPC 的一站式解决方案，适用于各种规模的高性能集群，使用 LICO 可以快速安装部署好一个 HPC 集群，同时 LICO 针对管理员和普通用户提供易用的管理平台。

软件平台由节点操作系统、集群软件系统、编程环境、应用软件、存储系统等 6 个子系统组成。



联想智能计算平台 LICO-HPC 包含但不限于以下功能：

部署功能

集群管理软件平台系统应具有批量部署操作系统的功能，以方便同时部署多个服务器的操作系统。

集群管理软件平台系统应具有批量更改系统配置的功能，比如根据规则设置主机名、IP 地址等。

集群管理软件平台系统应具有备份操作系统或者操作系统配置的功能，以便需要重新部署环境时，能够迅速恢复原有环境。

集群管理软件平台应为图形化的使用方式，以方便系统管理员使用。

监控功能

集群管理软件平台应具有对硬件系统的监控功能。

对于服务器，应监控到以下指标：

- 系统整体运行状态
- CPU 数量、内存数量、网络状态、硬盘状态
- 系统运行温度，包含但不限于 CPU、内存、硬盘
- 系统状态异常情况

控制功能

集群管理软件平台应具有远程控制功能，控制功能主要包含：

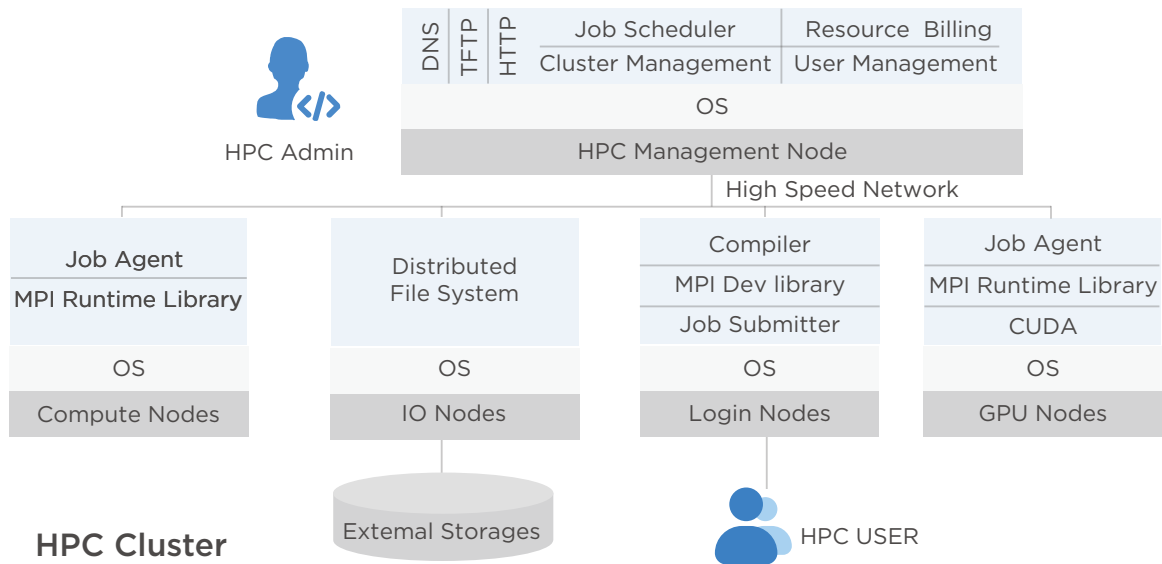
- 远程开关机（上电、下电）操作
- 动态调整 CPU 的运行频率
- 动态调整风扇的运行速度
- 动态调整能耗，以使系统运行在兼顾性能和节能的状态下

分布式文件系统管理功能

分布式文件系统的创建，磁盘空间管理，存储节点 HA 配置，文件的操作。

二次开发能力

所有 HPC 软件接口直接向用户开放，用户可以非常方便的要求定制或进行二次开发。LICO 解决方案架构示意如下：



LICO 解决方案示意图

针对智能计算平台的联想 LICO-AI

随着互联网的普及和 IT 业的高速发展，高性能计算（HPC）已经不再是少数大公司或大型科研机构的专属要求，而是被越来越多的包括政府、教育科研、石油石化、制造、军工和生命科学类的客户所需要和接受，针对 HPC 方案的复杂性，用户需要一个易用的 HPC 集群管理和作业管理平台。

同时，最近几年人工智能（AI）快速发展，大多数人工智能模型的训练需要使用 GPU，而如 NVIDIA Tesla A100 等 GPU 通常比较昂贵，不同的部门和人员有不同的模型训练需求，为每个部门和人员购买独占的 GPU 是对资源的浪费，所以一个集中共享式的 AI 模型训练平台也成为了越来越多客户的需求。

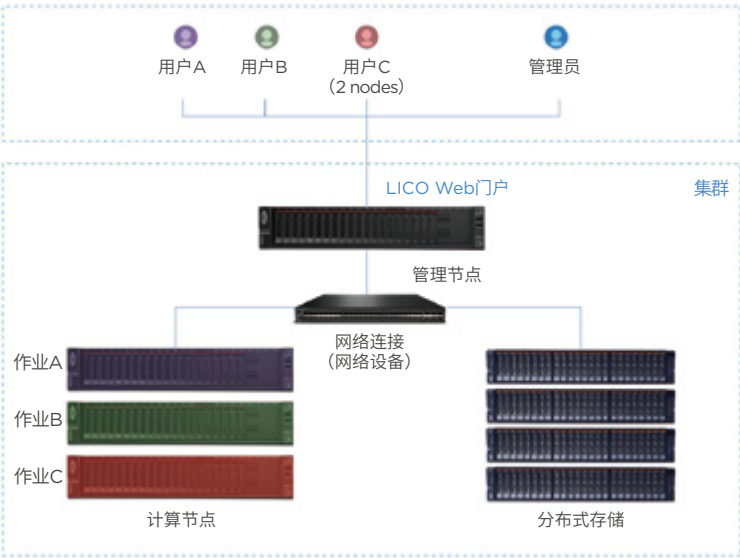
联想智能计算平台 AI 增强版本（Lenovo intelligent Computing Orchestration 以下简称 LICO）是基础设施方案业务集团（ISG）开发的，针对高性能计算（HPC）和人工智能（AI）的一站式解决方案，在一套集群中通过统一的资源调度，可以同时支持 HPC 作业和 AI 作业的运行。

LICO-AI 集成了集群需要的集群调度软件、监控软件、计算库、分布式文系统等，使用 LICO-AI 可以快速的部署好一个 HPC 和 AI 集群。

LICO-AI 提供了统一的 web 访问接口，集群管理员可以使用 LICO-AI 方便的管理集群，HPC 用户可以使用 LICO 方便的提交和管理 HPC 作业，AI 用户可以使用 LICO-AI 进行 AI 模型的训练。（以下介绍 LICO-AI 简称 LICO）

在如下的一个物理集群上：

- LICO 可以作为一个 HPC 的平台
- LICO 可以作为一个 AI 模型训练的平台
- LICO 可以作为一个 HPC+AI 模型训练的平台



LICO 平台提供了统一的 Web 门户：

- 集群管理员可以使用 LICO 管理集群
- AI 用户可以使用 LICO 进行 AI 模型的训练
- HPC 用户可以使用 LICO 提交和管理 HPC 作业
- LICO 平台提供了开放 API

节点操作系统

联想高性能集群系统中的服务器节点支持的操作系统有：Red Hat、Enterprise、Linux、CentOS、SUSE 、Windows Server。

集群软件系统

整个集群的部署、监控、管理和计算资源的分配和计算任务的调度。集群软件系统需要包括集群部署、集群监控、集群管理、作业调度等几个主要的功能模块。这些功能组件合力对集群系统的软硬件资源整合管理，使得用户无需关心某一资源单元，而只要集中于生产应用即可。

IBM Platform 解决方案

原 Platform 公司拥有 60% 世界 500 强的用户，用户行业遍布航空航天、电子、制造、教育科研、生物、海洋、能源等。

IBM（国际商业机器）公司于 2011 年 10 月 11 日宣布收购 Platform 公司。位于加拿大多伦多、中国北京和西安的研发团队，也全部并入 IBM 国际商业机器（中国）投资有限公司。于 2012 年 7 月 1 日，IBM 完成了与 Platform 公司的整合，完成了 IBM Platform 这个产品品牌的简历，这次强强联合，使 IBM 拥有了在高性能计算领域从硬件到软件的整体解决方案。2014 年 10 月，联想完成了对 IBM System X 的收购，从此联想转售 Platform 软件产品。

IBM Platform 产品系列包括多款适用于 HPC、大数据、云计算等各大领域软件产品系列，如 Platform Cluster Manager，用于部署、监控和管理 HPC 集群；Platform LSF，强大的负载管理平台等。

其中 Platform LSF 是 HPC 领域非常权威和应用广泛的一款产品。Platform LSF 产品家族实现了软硬件资源共享调度，将所有软硬件资源有机地组合在一起，根据事先定义的调度策略，统一管理，结合 Platform LSF 产品家族成员，建立先进的高性能计算平台，提高软硬件资源的利用率。

什么是 IBM Spectrum Computing LSF Suite？

IBM Spectrum LSF 套件提供了高度可扩展、高可用的体系结构，支持传统的高性能计算和高吞吐量工作负载，以及大数据、感知、GPU 机器学习和容器化工作负载。

Platform LSF

Platform LSF 是一个管理和加速高性能计算的系统软件、尤其对用户的关键任务和作业提供运行保障。借助 Platform LSF，你将能够智能地调度并行和串行作业，轻松解决大型、复杂的计算问题，同时最大限度地利用可用的计算资源。Platform LSF 全面地利用了超级计算机的高性能网络互连资源，让用户得到利益的最大化。Platform LSF 是一个经过实践证明的可靠的解决方案，在 TOP500 高性能计算项目中，目前，在排名前 10 的超级计算机都安装和运行了 Platform LSF。

1. Platform LSF 主要特点如下：

基于拓扑结构的调度算法 - 具有一致性的、最优的性能

Platform LSF 与系统厂商和第三方具有高性能互连结构的应用进行了高度集成，以期达到最佳性能的负载调度。基于拓扑的调度方法可根据每个处理器间的高性能互连的特性来确定哪些处理器的连接距离最短。这保证了每次运行一个应用时都能确保最佳的运行性能和一致性。

IT 资源虚拟化 - 自我管理和可靠的服务

用 Platform LSF，用户可共享虚拟的 IT 资源池，无需为每个批量作业指定特定的 IT 资源。Platform LSF 的自适应功能能够动态地分配资源，以适应不断变化的 IT 基础设施，减少了管理成本和相关成本。

采用强大、可靠、基于标准的高级技术

Platform LSF 是创建在经生产证明的、开放的、具有网格能力的虚拟执行机 (VEM) 结构上，拥有 11 年的分布计算经验，可提供强大、可靠的负载管理解决方案。Platform LSF 最大限度地利用了全球分布的异构计算资源。Platform LSF 的高级技术提供了无可比拟的可扩展性和性能，为企业组织提供了他们所需的速度和灵活性，使他们能够满足预期的功能和项目要求。

智能的、由策略驱动的 HPC 负载处理

通过 Platform LSF 管理和加速高性能计算 (HPC) 中重要的、计算密集和数据密集的负载。以智能方式对并行和串行负载进行排程，以便克服主要困难，并最大限度地利用计算资源。Platform LSF 还提供了丰富的用户工具，如 PAC，WEB Portal，集成应用后可以直接从网页提交作业；Reports，提供报表功能，供用户以报表形式查看资源使用情况。

GPFS 并行文件系统介绍

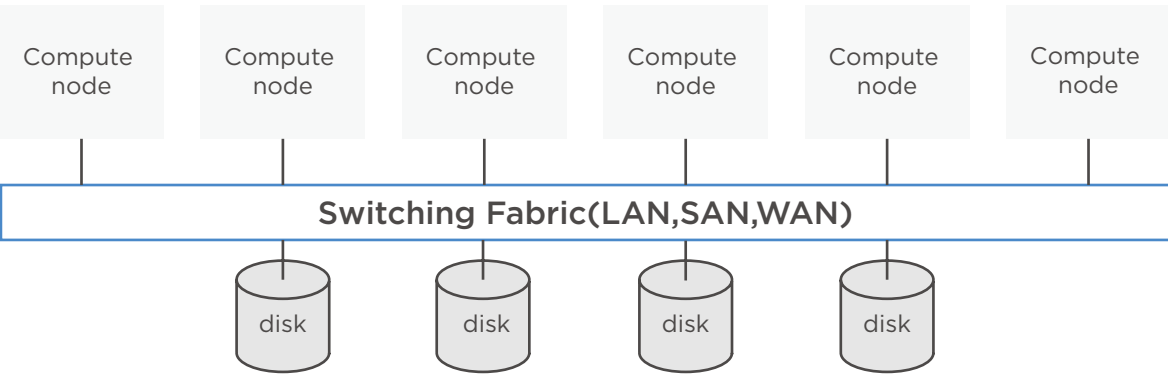
GPFS（General Parallel File System）是行业领先的并行分布式通用并行集群文件系统，GPFS 从 1993 年开始研发，1995 年投入商用（最早投入使用在多媒体处理行业，所以我们现在看到 GPFS 的很多目录和命令为什么都有 MM[multimedia] 两个字母，就是根据这个背景延续下来的），其中经过了多次版本的改进和修订，到 2013 年，GPFS 的版本为 3.5。到 2015 年，GPFS 已经到 4.2 版本。目前 GPFS 的版本已经全面升级到了 V5.x。

GPFS 是一个共享的磁盘文件系统，集群内的所有节点 (NODE) 可以并行的访问所有共享磁盘，并通过分布式的令牌管理机制和条带化技术来管理和优化各个节点的访问。

GPFS 的应用范围非常广泛，从多节点的文件共享服务到大型的高性能计算机群，都可以看到 GPFS 的优秀表现。GPFS 也应用在众多的企业应用系统中，比如高可用的解决方案：HACMP，ORACLE RAC，SAP 的共享文件系统等，GPFS 在这些应用里面都表现出了出色的性能和高可用性。

1. GPFS 是可扩展的、并行化的、高可用的、高性能的文件系统

- 群集：数千个节点、快速可靠的通信、通用管理域
- 共享磁盘：用户可在分布式锁定服务的协调之下从任何阶段接入磁盘上面所有的数据和元数据
- 并行：数据和元数据在所有节点及所有磁盘之间并行流动；文件分布在所有磁盘上



2. GPFS 是高可扩展的：（实际超过个 20000 个节点）

- 对称的、可扩展的软件架构
- 分布式元数据管理
- 允许轻松地逐渐扩展系统（节点和磁盘空间）

3. GPFS 是高性能的文件系统

- 大规模数据块（可调）支持与（跨越所有节点和磁盘）的广泛分布。对于高带宽环境来说，GPFS 可通过单次操作读/写大块数据，从而最大限度地降低 I/O 运营开销
- 从多个节点并行接入文件
- 全面的令牌增强（扇区范围）和令牌管理
- 高效的深度预提：预读和迟写。GPFS 能够从缓冲池中智能地预提数据，从而并行向足够多的磁盘发送 I/O 请求，以便充分利用基本的存储硬件基础架构的带宽

- 识别接入模式（自适应机制）。GPFS 可识别多个 I/O 模式，包括顺序、逆序及其他各种形式的分布式接入模式
- 高效的多线程后台程序
- 面向 MPI-IO 及其他应用的数据迁移模式

4. GPFS 是高可用的和容灾的

- 支持多个数据保护机制，包括日志记录、复制、镜像和影子等（都是标准的文件系统技术）
- GPFS 可配置用于消除单点故障。您可将这个文件系统配置成在磁盘或服务器发生故障时自动保持可用性
- 心跳机制可在发生多个磁盘、节点和连接故障时恢复正常运行。各层均已实施了恢复软件机制

5. GPFS 提供明显的性能和扩展优势

- 具有公认可扩展性的单一文件系统
- 专门开发用于高性能超级计算群集应用
- 非常成熟和稳定 – 最基本的基础架构能够支持包含 20000 多个节点的群集，提供卓越的性能及稳定性
- 当与 SSD 一起使用时，可提供巨大的性能优势
- 当主数据丢失时可继续存活，自动使用复制好的数据
- 群集文件系统，群集可在节点丢失时继续存活
- 支持高可用性解决方案

Lustre 并行文件系统介绍

Lustre 由三个部分组成：客户端（Client）、对象存储服务器（Object Storage Target，OST）和元数据服务器（MetaData Server，MDS）。

客户端通过标准的 POSIX 接口向用户提供对文件系统的访问。对于客户端而言，Lustre 是一个透明的文件系统。它无需知道具体数据所在的位置，可以透明地访问整个文件系统中的数据。

OST 负责实际数据的存储，处理所有客户端和物理存储之间的交互。这种存储是基于对象（Object -based）的，OST 将所有的对象数据放到物理存储设备上，并完成对每个对象的管理。OST 和实际的物理存储设备之间通过设备驱动程序来实现交互。通过驱动程序的作用，Lustre 可以继承新的物理存储技术以及文件系统，实现对物理存储设备的扩展。

MDS 负责元数据的管理。MDS 构造、管理文件分布的视图，允许客户端访问对象。通过 MDS 的文件和目录访问管理，Lustre 可以控制客户端对文件系统中文件的创建、删除、修改以及对目录的创建、删除、修改等访问控制。通过 MDS，客户端得到数据所在的 OST，并与其建立连接，此后的读写操作就在客户端同 OST 之间

进行，除非有对命名空间的修改，将不再同 MDS 有关系，这样就降低了 MDS 的负载。在多个客户端的情况下，由于有多个 OST 存在，上述的工作模式就把对文件系统的访问转换为并行操作，从而可以较好地提高性能。

对象存储结构的优点：

性能：对象存储体系结构为计算节点提供可以直接、并行访问存储设备的方法。通过把文件数据 stripe 到多个 OST 上实现并行访问。

传统存储结构的元数据服务器有两个主要功能。（1）负责为计算节点提供存放数据的逻辑视图（vfs 层），提供文件列表以及目录的组织结构。（2）组织数据在物理存储媒介上的存放方式（inode 层）。

对象存储结构把存放数据的逻辑视图和物理视图分开，把 inode 层的工作分给具有处理能力的智能 OST 存储设备。实现了元数据的分布存放，避免了传统元数据服务器的性能瓶颈问题。

可扩展性：将负载分布到多个智能的 OST，并用网络 and 软件将它们有机结合起来，消除了可扩展问题。一个对象存储系统有内存、处理器、磁盘系统等，允许它们增加其存储处理能力而与系统其它部分无关。如果对象存储系统没有足够的存储处理能力，可以增加 OST，确保线性增加性能。

易管理性：智能化的分布对象存储结构可以简化存储管理任务，可以简化数据优化分布的任务。例如，新增存储容量可以自动合并到存储系统中，因为 OST 可以接受来自计算结点发出的对象请求。系统管理员不需要创建 LUN，不需要重新调整分区，不需要更新文件服务器等。RAID 块可自动扩展到新的对象，充分利用新增的 OST。

安全性：对象存储结构在每个级别都提供安全功能，主要包括存储设备的身份认证，计算结点的身份认证，计算结点命令的身份认证，所有命令的完整性检查，基于 IPSec 的私有数据和命令等。

高性能文件系统结构：文件系统由三部分构成，包括客户端 Client、元数据服务器 MDS 和存储服务器 OSS。元数据节点和 OSS 节点接入 InfiniBand 网络。

基于 **Lustre** 构建文件系统，对特定应用采用相应的策略优化，并侧重于易管理性和可恢复性。完成的高性能文件系统具有如下主要特色：

- 基于对象存储，使存储更具智能化
- 对单个目录下的元数据操作进行优化
- 极佳的可扩展性。基于对象存储的体系结构使系统可以支持几万个客户节点，存储容量支持 Petabytes 级别
- 可靠性。由于 Lustre 已经部署到多个不同规模大小的实际机群系统中，可满足商业和安全应用对正常运行时间的要求
- 开源及开放的标准，Lustre 作为一个开源软件开发和维护，满足开放的网络协议和 Posix 文件系统语义，对工业标准平台和异构网络环境进行广泛支持
- 支持多种主流的网络连接，包括 TCP/IP，Myrinet，InfiniBand，Grandrics 等

应用软件

联想高性能计算系统可以集成和支持多个领域的多种商业和开源的科学计算应用软件，包括：商业应用软件如 MOLPRO、GAUSSIAN、Q-CHEM、MATLAB、VASP、TURBOMOLE、WIEN2K、SPARTAN、Materials Studio、Linda、LS-DYNA、FLUENT、CRYSTAL、ADF 等；

开源应用软件如 GROMACS、GrADS、ABINIT、GMT、CASTEP、CPMD、DL_POLY、DOCK、GRAPES、MM5、NAMD、AutoDock、VMD、TINKER、OpenMX、VENUS、WRF 等。详细介绍请参考 3.1.5 节。

编程环境

为用户提供应用程序运行环境、用户编程环境以及集群系统软件。

集群编程环境主要是为用户提供软件开发环境，同时为集群软件系统提供运行环境。除了有很开源的集群编程环境项目以外，各硬件厂商会根据自己的硬件架构设计提供整套优化的编程开发工具。主流集群编程环境如下所示：

- Intel OneAPI
 - NVIDIA CUDA Toolkit
 - PGI Compilers & Tools
 - GNU
- MPICH
 - OpenMPI
 - OpenMP

运行环境

应用软件：各领域用户生产软件

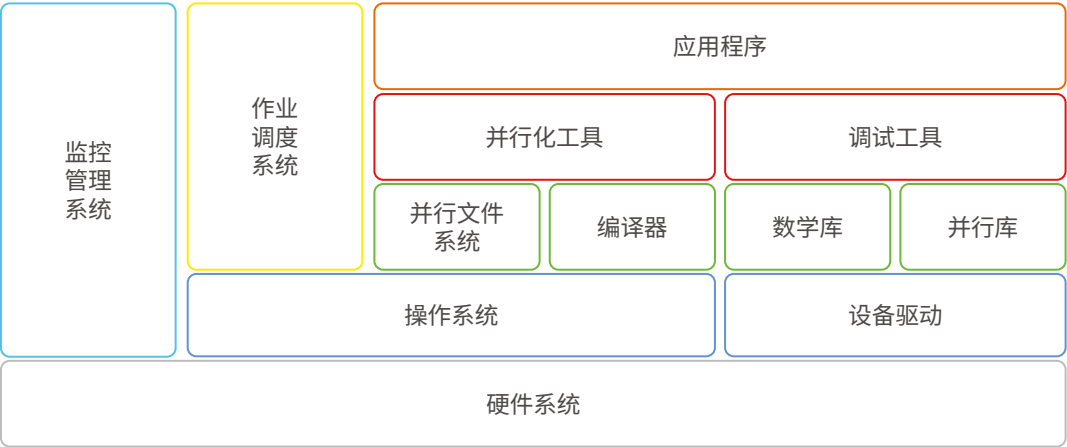
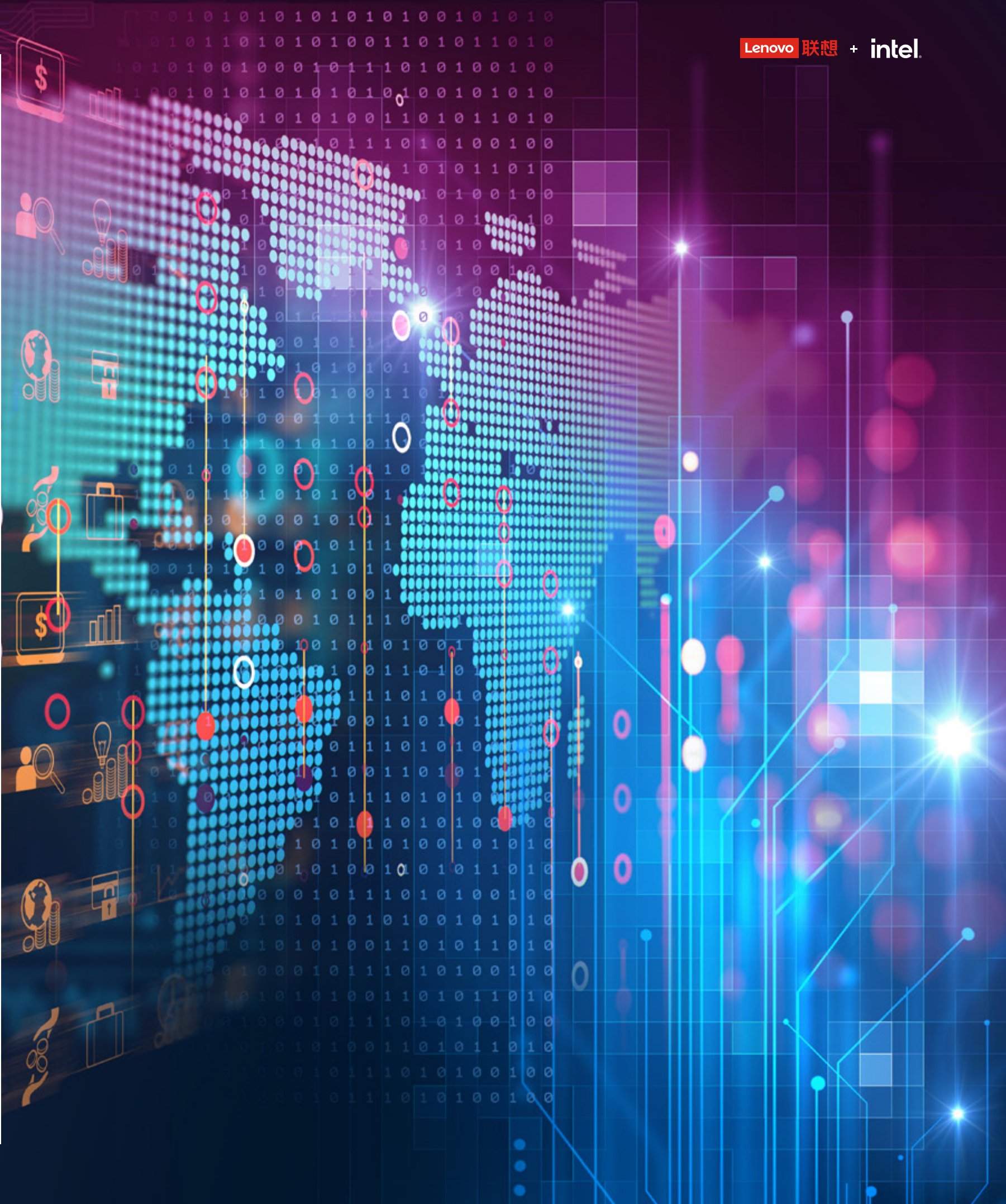


图 软件系统层次图



联想高性能计算集群实施服务

此高性能计算集群实施和培训服务会将已经完成上架的各种硬件设备进行系统软件的安装和配置，从而真正成为一套可用于实际工作的集群系统。联想企业级业务集团专业服务团队的高性能计算实施工程师通过快速配置集群以满足客户的需求，并且针对集群日常管理所需的技能对客户进行培训。集群可由各种联想自有产品或联想 OEM 的产品组成，包括 ThinkSystem 机架式服务器、高密度刀片式服务器、GPU 服务器以及存储产品和网络产品等。联想企业级业务集团专业服务团队拥有专业的高性能计算领域的知识和技能，快速帮助客户将各种硬件组合成集群系统。

在项目的前期阶段就可以引入专业服务团队的 HPC 实施工程师，和客户进行充分的沟通和交流，了解客户的需求和想法，同时可以基于我们的实践经验，为客户提供最佳实践的意见和建议。而后针对客户的需求和想法，我们会撰写工作说明书，以明确实施服务的主要工作内容和范围。在和客户确认工作说明书后，我们就可以评估整个项目实施服务所需的工作量，从而提供报价给到联想的销售人员。

在高性能计算集群的硬件和服务下单后，我们会得到通知。实施工程师即可安排和客户沟通，制定具体实施计划以确保实施过程顺利。在实施阶段，根据集群规模大小，一个或者多个实施工程师会到达客户数据中心，完成所需的实施服务。在项目尾声，实施工程师会将正常工作的集群系统交付给客户，并且提供相应的安装实施文档以及交付培训。

实施速度——我们的实施工程师会凭借他们丰富的经验以最快的速度完成集群的实施工作。他们会使用经过多次验证的最佳实施经验以及软件组件版本，避免反复的试验和错误，快速让集群系统可以进入工作状态。

实施质量——我们的实施工程师不断的积累最佳实践方法，改进安装配置技术。他们还与联想的研发团队以及各地客户紧密配合，以帮助完善产品和实施服务的质量。

减少风险——我们的实施工程师会帮助您避免典型错误导致的性能问题、兼容性问题或保修问题。

选择联想的专业实施服务将帮助客户充分利用新系统的能力，而不会忽略一些重要的特性。客户不需要花费大量的时间去学习集群的知识，而是快速的从实施工程师的丰富经验中得到所需的知识和信息。

高性能计算集群实施服务通常包含以下内容：

每一个高性能计算集群实施服务都有特殊的地方，但是多数通常都包含以下内容：

- 一个准备和计划会议
- 管理节点的安装和配置。包含 RAID 配置、操作系统安装、BIOS/UEFI 设置、微码更新、群集管理软件安装（LICO/Confluent）
- 以太网配置和验证
- 节点（计算节点、存储节点、登录节点、其它用途节点等）RAID 配置
- 向节点（计算节点、存储节点、登录节点、其它用途节点等）分发操作系统，并配置 BIOS/UEFI、微码更新、驱动和软件包更新

根据硬件配置不同的可选服务包含：

- 高速网络（InfiniBand 或高速以太网）的配置和验证，以及相匹配的软件包安装，例如 OFED
- 存储系统配置和验证
 - 在管理节点上安装存储管理软件
 - 根据存储需求配置多路径软件
 - 存储控制器或 SAN 交换机 Zone 配置
 - 创建 RAID、LUNs 和文件系统
- 特殊用途节点配置，如 GPU 节点

根据软件配置不同的可选服务包含：

- 资源管理 / 调度软件的实施，如 Torque/Maui, Torque/Moab, PBS-Pro, LSF, Platform HPC, SLURM 等
- 集群监控或报警软件实施，如 Ganglia, Nagios
- GPFS 并行文件系统实施
- 特殊用途工具、软件、库的实施，如 Intel OneAPI, Intel MKL, 特殊的 MPI bindings.

额外可选的附加服务包含：

- 协助解决方案的架构设计，审阅建议的解决方案
- 硬件测试：加电开机测试，包括 STREAM, PALLAS, IOzone, HPL 在内的基准测试
- 协助性能测试和验收测试
- 对于集群中各种组件的深入技术培训
- 协助客户进行应用软件和许可证管理器的安装
- 协助项目管理：实施进度安排、硬件和软件的订购和交付、预算管理

要订购此服务，请联系我们的商机管理员根据客户的实施需求撰写工作说明书（SOW，Statement of Work）。在中国地区，此服务可以通过以下编号进行订购。具体订购数量将由商机管理员根据工作说明书来确定。

了解更多联想企业级业务集团专业服务，请访问以下页面：

<http://shop.lenovo.com/us/en/systems/services/>

联想生命科学客户案例

某医院组学和深度影像学习平台

总医院是集医疗、保健、教学、科研于一体的大型现代化综合性医院。医院的信息化建设一直走在全国前列，取得了对行业发展有重要影响的系列成果，拥有一支素质高、能力强的技术人才队伍，积累了丰富的数据资源。

为了充分挖掘医院的数据资源潜力，支撑医学科技创新发展，医院于 2016 年在行业中率先成立了医疗大数据中心，系统化开展数据资源开发利用工作。2017 年，医院更获批为国家发展改革委“医疗大数据应用技术国家工程实验室”建设单位，面向医疗行业开展大数据应用创新。

医疗大数据中心的主要发展方向是：围绕医疗大数据整合、数据处理与分析、智能化辅助诊疗、医院管理决策、安全隐私保护等开展技术创新、标准创新、产品创新和应用创新。为医疗大数据技术的研发应用提供一流的支撑保障。

大数据中心已经拥有大量的组学、蛋白组、影像等医学数据，客观上存在深入研究的数据基础，但是在实际研究中发现基因组学研究需要大量的计算能力、存储能力，医疗影像分析需要支撑机器学习的计算能力、图形处理能力以及相应的算法基础平台。这些需求促使我中心进行高性能集群建设。

某总医院

组学计算平台
- 使用科室：大数据中心自用 - 研究范围：基因、蛋白质、生物序列、分子 - 应用：clc genomics workbench - 数据源：使用自有组学数据和第三方机构数据 - 目的：辅助临床

深度影像学习平台
- 使用科室：大数据中心自用 - 研究范围：肝、肺影像资源，肿瘤 AI 标的 - 基于 tensorflow pytorch 自研系统 - 数据源：使用本院以及医院体系院的影响数据 - 目的：辅助临床

学科平台需要支撑运行的软件

学科	解决方案	数据 / 应用程序特点	主要应用程序
生物信息学 - 序列分析	搜索、对齐生物序列（DNA 和蛋白质）并对其进行模式区配	结构化数据整数占主导，频率依赖性，大量缓存和内存 BW 并非关键，一些算法适合进行 SIMD 加速	NCBI BLAST、wuBLAST、ClustalW、HMMER、FASTA、Smith-Waterman
生物信息学 - NGS	对齐并合并 DNA 短片段，以重建原始序列	- 一些需要大量内存 - 许多为 IO 密集型任务 - 字符串分析和匹配算法	Phrap/phred、CAP3/PCAP、Velvet、ABYSS、SOAPdenovo、Newbler、MAQ、BOWTIE、BFAST、SOAP、BioScope、SAM 工具、GATK
计算化学 - 分子模拟和量子力学	使用分子动力学和量子力学技术对生物分子进行建模	- 需要处理大量浮点数据 - 延迟至关重要 - 频率依赖性 - 可缩短到 100 秒以下	CHARMM/CHARMm、GROMACS、Desmond、AMBER、NAMD、Gaussian、GAMESS、Jaguar、NWChem
蛋白质组学	解读质谱分析数据，将频谱与蛋白质数据库进行匹配	- 整数比较 - FFT 频谱分析具有浮点数据 - 对通信要求不高	Mascot、Sequest、ProteinProspector、X!Tandem OMSSA
结构生物学	用物理学方法，配合生物化学和分子生物学方法研究生物大分子结构与功能	- 数据量大 - 计算密集型	Relio 为典型代表（GPU 优化）

根据中心研究课题安排，建设两个集群：基因组学集群，深度影像学习平台。

基因组学集群：

- 用途：基因、蛋白质、生物序列，分子研究
- 数据源：自有数据及第三方公开数据集
- 主要应用：clc genomics workbench
- 用途：临床辅助

深度影像学习集群：

- 用途：肝肺等、肿瘤 AI 研究
- 数据源：自有数据及各医学研究中心相关影像数据
- 主要应用：基于 Tensorflow，Pytorch 自研应用
- 用途：临床辅助



高性能计算集群建设要求

架构要求

- 通用性：有大型案例验证的通用性架构，能够使用业内常用的框架
- 可扩展性：未来会根据研究成果和科研需求对集群进行扩建，所用架构必须具有技术持续性，未来扩展在原集群扩建即可

网络要求

- 高带宽：特别是组学分析，对带宽要求高
- 低时延：多任务并发，时延问题将会放大
- 无阻塞：计算存储要求网络一致，做到无阻塞连接，防止形成集群短板

计算能力要求

- 经调研，组学集群需 50TFlops 以上浮点运算能力
- 经调研，深度影像集群需 160TFlops 以上浮点运算能力，选用训练能力和推理能力均衡的 NVIDIA Tesla V100 GPU。

存储要求

- 容量要求：组学分析需不少于 200TB 容量，深度影像学习平台需不少于 100TB 容量
- 带宽要求：组学分析对带宽要求比较高，要求多流读写达到 7GB/s 以上，单流读写 3GB/s 以上
- 文件系统要求：非开源的商业版文件系统，有成熟应用案例，技术支持有保障
- 可扩展要求：随着未来集群的成熟使用，数据一定会有海量增加，存储架构要可扩展到海量容量；同时由于现阶段，机房空间和供电有限，需要考虑整体功耗，未来扩展给机房带来的空间和功耗压力

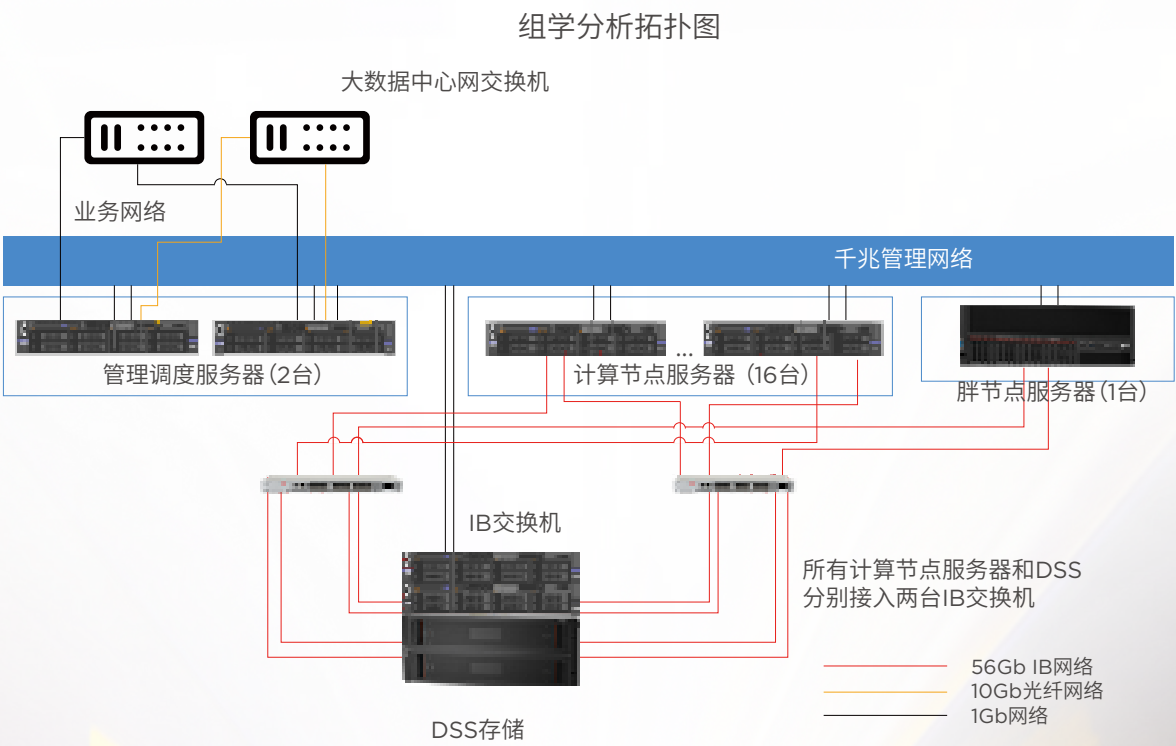
集群管理平台要求

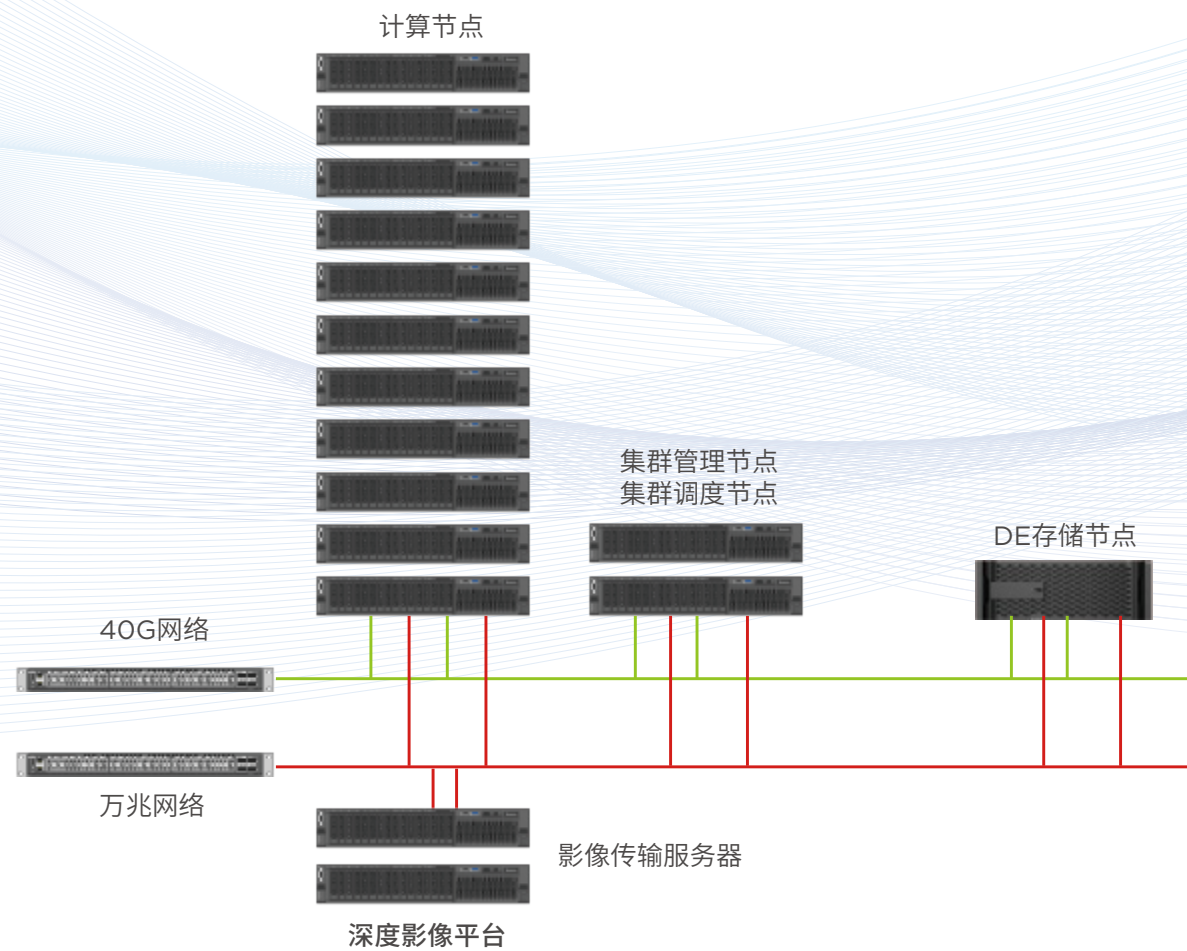
- 具有管理、调度、监控功能，兼容常用的调度系统，具有良好的交互能力
- 最好两个平台使用的集群管理平台为同一个（兼顾高性能与 AI 框架），减少学习成本

其他要求

供货时间、实施时间、调优及培训

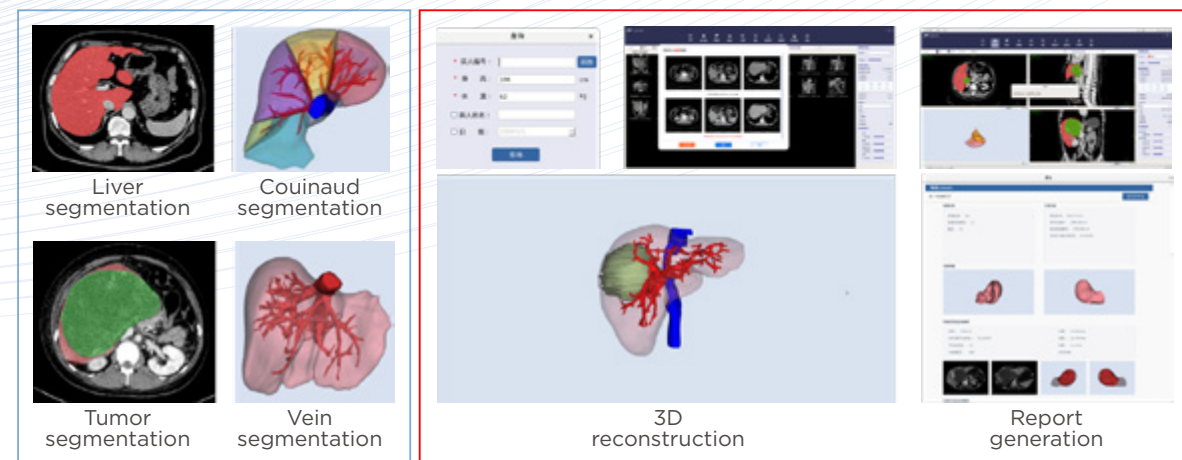
系统架构图：





合作利用 AI 技术协助肝癌的诊断

- 我国新发肝癌病例 46.6 万例，占世界肝癌病例的 55%。我国肝癌死亡病例 42.2 万例，占世界肝癌死亡病例的 45%-50%
- 中国有 1.2 亿乙肝病毒携带者
- 我国肝癌住院治疗费用 81.4 亿元



□ Main techs □ Prototype deployed in top hepatobiliary department in China



北京大学高性能计算平台“未名生科一号”

经过紧张的安装调试和测试运行，北京大学高性能计算平台“未名生科一号”高性能计算系统于 2018 年 10 月 26 日正式上线运行。至此，北京大学用不到一年的时间，完成了“未名一号”“未名教学一号”“未名生科一号”**等三套联想高性能计算系统的建设**，通过机制体制创新，建成了一个融教学科研计算、海量数据处理、信息管理服务于一体的高性能计算校级公共服务平台，能够较好地满足学校各学科领域对于大规模数据处理和大规模科学计算的普遍需求、专门学科领域对于高性能计算的特殊需求、各院系日常教学对于高性能计算的基本需求，通过组建一支高素质的专家管理团队和专业技术队伍，为学校重点学科建设和高素质人才培养提供了有力支撑，取得了丰硕的科研成果，实现了学校建设高性能计算校级公共服务平台的初衷和目标。

“未名生科一号”是学校根据生命科学及其相关前沿交叉学科建设的需要，在北大高性能计算校级公共平台上建设的第三套高性能计算系统，主要用于冷冻电镜平台的数据处理和存储。“未名生科一号”将优先服务于冷冻电镜平台的科研工作，同时也为其他学科领域提供不低于 30% 的公共机时。“未名生科一号”的建成，将有效缓解高性能计算平台首套系统“未名一号”的排队情况，为广大师生营造更加优质的高性能计算环境。



“未名生科一号” 整装待发

“未名生科一号”采用联想深腾 X8810 超级计算机系统建设，由 150 个节点组成，共计 4688 个 CPU 核心，28 块 GPU 卡，总体计算能力理论峰值为 587.8TFLOPS，其中 CPU 节点计算能力为 343.4TFLOPS，GPU 等其他节点总计算能力为 244.4TFLOPS，存储容量为 10.2PB。计算节点 CPU 支持 AVX-512 指令集，大幅度提升了处理器的浮点运算能力，GPU 节点采用最新的 NVIDIA Tesla V100 型号 GPU，并采用 NVLink 方式连接，单卡 GPU 双精度浮点运算速度理论值可达 7.8TFLOPS。计算网络采用 omni-path 架构，安装有 intel、gcc 等编译器，作业管理采用 slurm 调度系统，集群管理采用计算中心自主开发的集群监控管理套件。



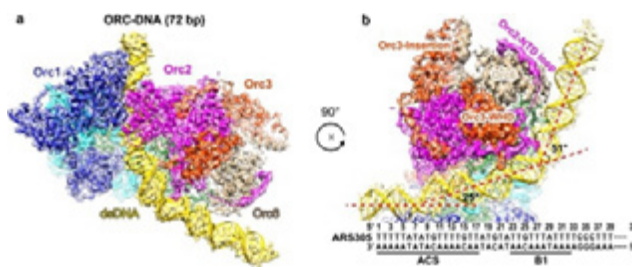
技术人员正在紧张调试

“未名生科一号”于 2018 年 6 月 21 日投入试运行，累计为 12 个项目 35 位科研人员提供了计算支持，包括生命科学学院高宁和李宁宁团队，物理学院欧阳颀、毛有东和李新征团队，分子医学研究所陈雷团队，北京国际数学中心鄂维南团队。试运行期间，系统的平均使用率为 94.7%。

北京大学高性能计算校级公共平台体制先进、技术领先、运行平稳、作业饱满、效率卓越、节能环保、成果丰硕，取得了良好的工作成效。自 2018 年 1 月 3 日正式投入运行以来，累计为学校 26 个院系的 500 余名师生提供了计算服务，支撑科研项目 166 个，支持发表高水平科研论文 57 篇，其中绝大多数为 SCI 一区或学科顶级会议论文，包括 3 篇 Nature、1 篇 PNAS、1 篇 JACS、一篇 PRL 和 3 篇 Nature 子刊，并获得两项专利。成果分布于数学、物理、化学、生物、信科、工学、医学、药学、地学、大气、材料等 10 余个学科领域。



王恩哥团队在 Nature 发文揭示水合离子的微观结构和幻数效应



高宁团队在 Nature 发文揭示 DNA 复制起点识别复合物 3-Å 分辨率的结构

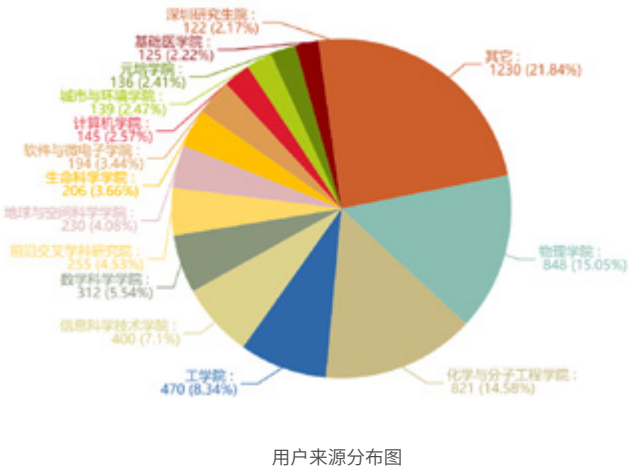
平台概况——用户覆盖

5633 人
师生用户

99 个
院系单位

20 位
两院院士

6 位
高校校长



用户来源分布图

高性能计算平台自第一套高性能计算“未名一号”建成以来，因其卓越的技术优势、良好的推广宣传和周到的技术服务，迅速在业界和校内外取得了巨大反响，技术团队多次受邀在重要学术会议做大会交流，很多兄弟院校前来参观学习，校内科研团队上机积极踊跃，平台迅速成为北大科研工作的重要支撑。与此同时，很多在平台获益的老师希望在教学中也使用系统，为课程实习和科研训练提供上机环境。为了不影响“未名一号”上的科研任务，计算中心克服困难，通过综合资源调配，快速搭建了一个专门服务于日常教学的小型高性能计算系统“未名教学一号”，为全校师生提供免费上机服务。“未名教学一号”于2018年5月4日北大120周年校庆期间正式上线运行，不仅解除了老师们搭建系统的后顾之忧，更是体现了平台服务学校核心需求的大局意识，有效支持了北大人才培养工作。目前已经有204位用户在“未名教学一号”上机实习，欢迎有需要的师生前来使用。



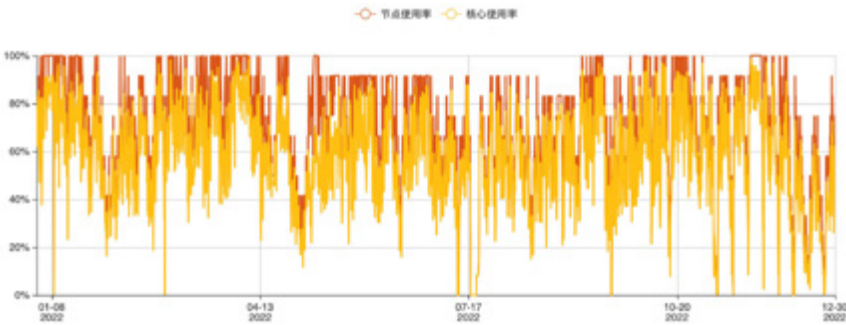
兄弟院校前来参观交流



材料计算科学与工程任课老师与选课同学在机房合影

平台资源使用状况

未名一号
年均使用率
87.58%



未名生科一号
年均使用率
81.30%

