案例研究

高性能计算
高性能存储
转化医学
英特尔® 傲腾™ 持久内存 + DAOS

打破存储瓶颈，加速生信数据挖掘
瑞金医院借傲腾™ 持久内存推进转化医学实践

案例简介

基础研究与临床实践分离的状况正制约着现代医学的发展——这种分离导致越来越多的基础研究逐渐偏离“解决实际需求”这一出发点，使得研究成果难以在临床实践中得到充分利用，因此无法更好地作用于医学的发展和进步。

转化医学正是解决这一问题的良方，它的出发点就是要将基础医学研究与临床实践更为紧密地结合起来。其基本特征是通过多学科交叉合作来实现“从实验室到临床、再从临床到实验室”这种双向且高效的转化。具体地说，就是要把临床提出的问题快速转化为基础研究项目，而后再将研究项目的成果同样高效地转化为针对临床患者疾病的精准预防、诊断、治疗及预后评估等一系列方案。

转化医学涉及众多学科，不仅包含了各个医学专科和药物开发领域，还融合了基因组学、IT 和以分子检测为重心的生物信息学。因此，IT 平台正是转化医学连接、打通和融合多学科以及开展以患者为中心的研究的基石。其在数据，尤其是海量生物信息数据（以下或简称生信数据）处理分析上的性能表现，是决定转化医学双向转化效率的一大关键要素。

作为中国首个，也是截至目前国内唯一建成的国家级综合性转化医学中心，上海交通大学医学院附属瑞金医院转化医学中心（下文或简称为“瑞金医院转化医学中心”）在进行转化医学实践与探索过程中深刻认识到了 IT 平台性能的重要性。他们携手英特尔，导入其至强® 可扩展平台，利用该平台独有的英特尔® 傲腾™ 持久内存，以及可充分激发这一创新存储硬件产品优势的开源分布式异步对象存储（Distributed Asynchronous Object Storage，以下简称 DAOS）并行文件系统，经过深度定制和优化，打造了可打破存储瓶颈、更适用于海量生信数据挖掘的 ASTRA 高性能计算平台，大大提升了瑞金医院转化医学中心生信数据访问和处理的整体效率，实实在在地推进了中国转化医学的实践和发展。

“生物标记物和新药研发平台是转化医学国家重大科技基础设施技术支撑系统的重要组成之一，而 ASTRA 高性能计算平台则是为生物标记物和新药研发提供“生信”及“计算”的基础设施。无论是面向患者真正个性化、精准化诊疗方案的定制，或是新型靶向药物的快速开发以及国家重大公共卫生安全事件的敏捷应对，都深度依赖生信分析和生信计算平台提供的数据和算力。我们很高兴 IT 行业的领先产品、技术和方案能够应用于建设海量生物信息数据的分析挖掘能力，为转化医学的加速落地提供全周期、全流程的支持。相信这种跨界的协作和创新，还将在未来的医疗行业中碰撞和激发出更多创新的火花，为全行业的持续变革和演进提供源源不断的能量和助力。”

陈赛娟
中国工程院院士
上海交通大学医学院附属瑞金医院
国家转化医学研究中心（上海）主任

“转化医学中心的建立，对瑞金医院乃至全国范围的医学发展而言都有着举足轻重的影响。转化医学涉及海量生物信息数据分析，不仅在计算上存在复杂多样的需求，在数据存储方面，更是需要满足对海量数据进行高时效访问的要求。为此，我们导入了英特尔® 至强® 可扩展平台，利用其独有的英特尔® 傲腾™ 持久内存的低时延、高带宽、大容量、长寿命、非易失等特性，再搭配专为高性能存储硬件设计的 DAOS 开源并行文件系统，成功构建了适用于转化医学应用场景的高性能临床医学深度挖掘系统。该系统正用于支撑我们的一体化 ASTRA 高性能计算平台，在实现更优存储和计算能效的同时，推进我们在转化医学领域的探索和实践。”

吕纲
ASTRA 高性能计算平台共同负责人
上海交通大学医学院附属瑞金医院
国家转化医学研究中心（上海）

转化医学国家重大科技基础设施（上海）简介

近年来，转化医学不仅在欧美等发达国家得到了大力倡导和发展，在国内也获得了政府和相关行业的高度重视及投入[1]。上海交通大学和瑞金医院共同承建的转化医学国家重大科技基础设施（上海）是中国“十二五”重点规划的十六项重大科技基础设施之一，其重点研究方向聚焦肿瘤（包括白血病）、代谢性疾病和心脑血管疾病等领域，采用患者招募模式，提供约 300 张病床。所有床位都可通过智能设备及信息化技术组成多功能智慧临床研究集群，每个床单元都具备自动感知以及临床研究数据自主集采功能。所获数据通过 ASTRA 高性能计算平台进行分析，分析结果供相关团队制定针对性治疗方案，由此推动了我国医疗行业基础研究和临床实践的良性互动和同步提升，造福更多国民[2]。

海量数据：转化医学的双刃剑

如何理解转化医学，特别是它与传统医学的不同之处？

其实用一个来自实践的例子就可以说明：与完成病理分析及少量分子检测后即可实施手术或化疗的传统癌症治疗方案不同，转化医学会对患者的整个基因组进行分析，以指导药企及临床医生进行更有针对性的药物及治疗方案研究，从而为患者提供更为精准、有效的治疗。

由此可见，生物信息学研究是转化医学研究的基石，而对它更为全面和充分的利用，也是转化医学与传统医学的主要差别之一。这种研究会涉及对蛋白质、DNA 和 RNA 等生物分子的研究，也包含了生物信息的获取、加工、储存、分配、分析、解释等方方面面[3]。这就使得瑞金医院在开展转化医学研究的进程中，必须面对以下三个环环相扣的挑战：

一、庞大、复杂且持续增长的数据：由于转化医学中心在实际工作中需要围绕基因组测序、转录组测序、蛋白质组学、代谢组学、药物筛选等各种先进组学检测技术与平台，紧密结合各种患者生理生化指标、组织病理检查、器官影像检查、家族遗传背景、疾病历史诊疗结果等信息形成多维度的数据流，然后再通过各种生物信息学手段进行原始数据的综合分析和挖掘，提供与疾病诊断和治疗相关的遗传和临床信息，待形成特征数据后再通过机器学习、深度学习和人工智能等方法整合特征数据与临床诊疗方案，成就真正意义上的精准分析、辅助诊断和个性化医疗，因此，它在此过程中就必须面对**体量庞大、复杂且持续增长的数据**。例如，仅单个人类全基因组测序分析涉及的数据就高达 870 GB[4]。

二、数据处理速度至关重要：如上述转化医学工作流程所述，在收集海量数据的基础上，瑞金医院转化医学中心还需要全力提升数据的处理速度。这是因为中心招募的患者通常病情都比较危急，需要以尽可能快的速度基于患者生信数据分析结果给出针对性的创新疗法。毕竟，在针对患者各类数据进行分析和探索新疗法或尝试新药物的过程中，如果前期分析研究阶段花费太多时间，就会相应地缩短后期临床实践的过程。这不仅会延长患者治疗周期、影响治疗效果，还可能会导致错失最佳治疗时机。因此，数据处理和分析的速度，或者说效率，就成了能否成功挽救患者生命、同时探索出有效疗法的关键。

三、高并行访问需求不容忽视：在加速数据处理的同时，中心还需要满足不同转化医学研究团队高效并行访问海量生信数据的需求。

为应对上述三大业务层面的挑战，瑞金医院转化医学中心着手搭建了一个集存、传、算、用为一体的定制化超算平台。在初期实践中，中心的 IT 技术团队发现：该超算平台强劲的 CPU 算力和高效的算法固然可以保障平台的计算性能，但其存储系统却难以满足转化医学实时、高频和高效的数据访问和处理需求。换言之，该超算平台的存储系统不仅要具备存储海量数据的能力，还必须要具备更出色的 I/O 和吞吐能力，才能成为提升整个超算平台数据处理能力的助力，而非瓶颈。

以白血病为例，更好地认识转化医学

瑞金医院在白血病研究与诊治领域成绩斐然。例如总治愈率能达 90% 以上、被媒体誉为“上海方案”的著名急性早幼粒细胞白血病(APL)治疗方案就是该医院研究团队的重要成果[5]，也是从基础研究成功走到临床转化应用的转化医学实践范例。

从转化医学的角度出发，白血病的治疗包括以下几个步骤：

1) 细胞形态学研究：在显微镜下看细胞形态，确定细胞属于癌细胞还是正常细胞；
2) 细胞遗传学研究：看染色体是否正常；
3) 免疫表型（流式）研究：研究患者表现异常的细胞群（如 T 细胞/B 细胞）；
4) 分子检测：对患者的整个基因组进行测序，确定从出生到现在患者的基因变化（是否存在融合基因、染色体变异、基因突变等），从而基于这些研究进行更有针对性的新药物开发或治疗方案制定；
5) 免疫表型 HLA：这关系到后续如果需要移植是否有合适配型，是否有合适的药物；
6) 基于以上结果对白血病患者进行分类（如：淋系/髓系），并匹配合适的临床试验；
7) 结合研究结果及临床实践开展治疗，为患者提供更精准的诊治。

其中第四步的分子检测既是转化医学的支点，也是转化医学与传统医学的区别所在。

加速存储！导入傲腾™ 持久内存 + DAOS

明确存储是 ASTRA 高性能计算平台需要攻克的瓶颈后，瑞金医院在规划该平台二期部署时，就开始与 IT 业界领先的产品与技术提供商们开展了有针对性的攻关。在与英特尔工程师进行了深度探讨和研究后，其 IT 技术团队最终决定采用“英特尔® 傲腾™ 持久内存 + DAOS”的方案，来破解这一瓶颈。

英特尔® 傲腾™ 持久内存是英特尔® 至强® 可扩展平台中面向存储应用的核心组件，也是颠覆传统内存-存储架构的明星产品。它与 DRAM 相近的性能、远超 DRAM 的容量、相比 DRAM 更优的成本和对 DRAM 所不具备的数据持久性的支持，使得其使用者可以在更靠近 CPU 的地方持久存储和高效访问更大体量的数据，从而为 ASTRA 高性能计算平台带来更加出色的整体性能、敏捷性、可用性和经济性。

与英特尔® 傲腾™ 持久内存搭档的 DAOS 分布式异步对象存储系统，是英特尔针对非易失性存储器（NVM）技术全面优化的、面向百亿亿次级（Exascale）超算存储堆栈的基础。它可为高性能计算应用提供具备更高带宽、更低时延和更高 IOPS 的存储容器。DAOS 在随机小 I/O（<=16 KB）、元数据方面拥有出色的性能表现，可支持结合了仿真、数据分析及人工智能的新一代以数据为中心的工作流程，在医疗系统动辄处理大量的半结构化及非结构化数据的应用场景中，也是大有用武之地。

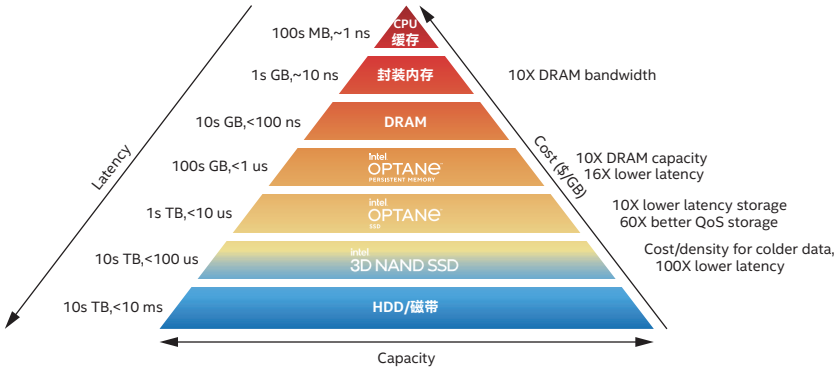

图 1. 从性能和容量两个维度上看，英特尔® 傲腾™ 持久内存的出现都是对传统内存-存储架构“缺口”的补足

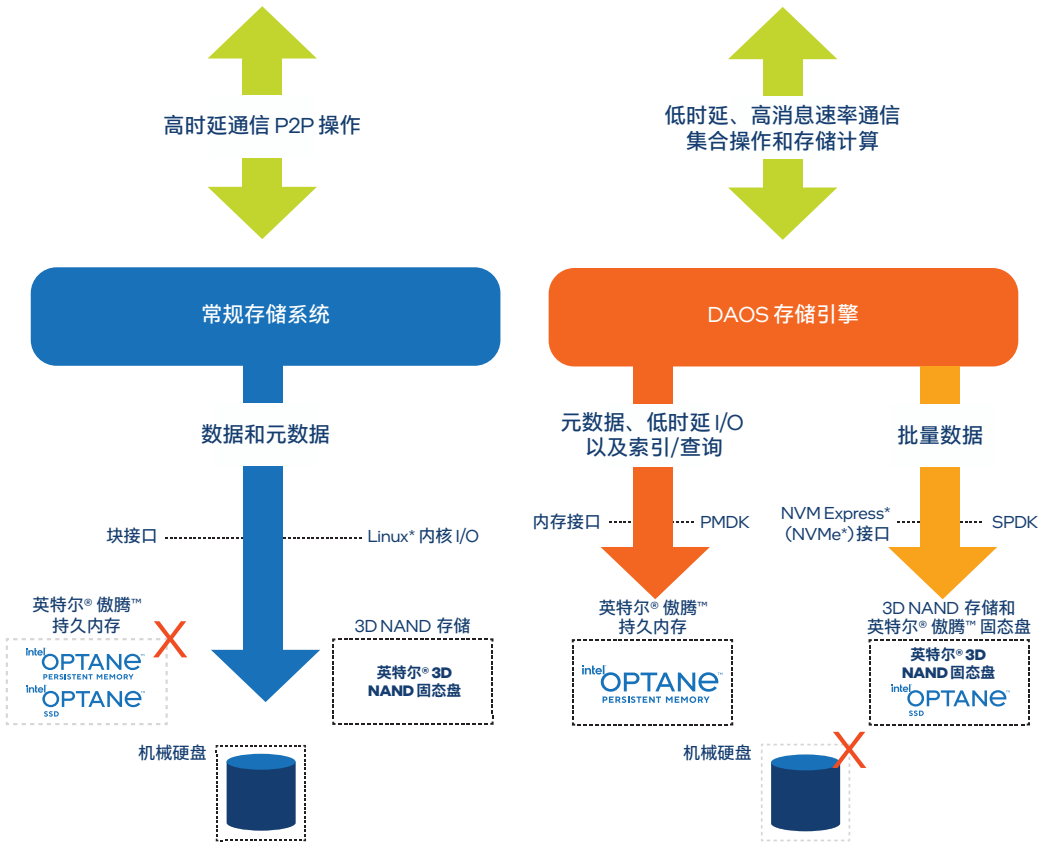

图 2. 从架构层面了解传统存储系统（左）与搭载英特尔® 傲腾™ 持久内存的 DAOS 系统（右）的差异[6]

作为存储加速硬件底座的英特尔® 傲腾™ 持久内存，在与 DAOS 搭配使用后能够全面融合两者的优势，输出优于其他存储方案的性能表现。早在 2019 年全球 IO-500 榜单中已有采用“英特尔® 傲腾™ 持久内存 + DAOS”方案的系统上榜[7]，2021 年最新排名前十的系统中更是有四套采用了 DAOS 方案，这充分验证了该方案在存储加速及文件系统性能提升方面的领先地位[8]。

“英特尔® 傲腾™ 持久内存 + DAOS”方案的这种领先性，正是它赢得瑞金医院转化医学中心青睐的核心原因。该中心基于这一方案以下四大特点，有效地应对了前文所述的业务挑战，具体包括：

一、让海量生信数据的存取兼顾容量和速度优势。由于英特尔® 傲腾™ 持久内存能提供比 DRAM 更大的容量规格，并兼具接近 DRAM 内存的性能，因此可让瑞金医院转化医学中心将更多生信数据存储到更靠近算力的地方，在提升存储容量的同时不牺牲性能。

二、针对快速数据处理，满足硬件层面上低时延的需求。由于 DAOS 依靠 OFI (OpenFabric Interface) 绕过操作系统将 DAOS 操作交付给 DAOS 存储服务器，可充分利用架构中的任何远程直接内存访问 (Remote Direct Memory Access, RDMA) 功能进行低时延、高消息速率的用户空间通信，并将数据存储在英特尔® 傲腾™ 持久内存和 NVMe 固态盘中，因此新方案可以大大提升数据存储、获取、处理和分析的速度。此外，不同于传统的缓冲区，DAOS 可提供独立的高性能容错存储层，不依赖其他层来管理元数据并能提供数据恢复能力，因此新方案也有助于更好地保护数据和加快数据恢复速度[9]。

三、满足各科研部门对数据的高并行访问需求。在 DAOS I/O 引擎中，存储可静态地跨越多个 Target 分区。为避免竞争，每个 Target 都有各自的存储、服务线程池以及网络上下文，这些上下文可以直接通过结构寻址，而不依赖于托管在同一存储节点上的其他 Target。这一特性可大大增强系统的并发能力，有利于更好地应对高并行访问需求。

此外，在多个分布式事务出现冲突访问的情况下，与 POSIX 采用悲观锁方式确保数据一致性不同，DAOS 采用的是类似于数据库的较乐观 (Optimistic) 的机制：DAOS 的 I/O 操作有各自独立的时间戳 (Epoch)，应用编程接口 (API) 在处理有冲突的 I/O 操作时会按照时间戳顺序执行，并且要求冲突事务进行重试以避免冲突。这种机制既可以避免使用分布式锁带来的额外开销，也可以避免分布式锁和随之而来的页对齐锁边界所带来的伪冲突，因而能达到提高应用并发度和提升常规 I/O 性能的目的。

四、满足医院对科研数据的可靠性及管理性需求。最新发布的 DAOS 新版本中增加了数据保护策略。传统数据副本的保护策略虽然数据冗余度高，但写带宽和存储空间的额外消耗很大，对于高速存储来说性价比较低。DAOS 的纠错码 (Erasure Code) 则是类似于 RAID 的基于软件层的数据保护策略，其存储空间和写带宽的额外消耗远低于多副本。当用户使用纠错码时，应用的数据会被分片存储到多个数据引擎并分布到不同的节点上，对应的纠错码也会存储到另外的节点中。当系统出现故障，造成存储引擎数据丢失，或者出现存储设备损坏的情况时，DAOS 会根据纠错码来修复丢失的数据，从而实现系统的高可用性。

强大生态系统为英特尔® 傲腾™ 持久内存应用保驾护航

自 2019 年发布英特尔® 傲腾™ 持久内存以来，英特尔就积极与 OEM、ISV 和 ODM 等合作伙伴携手，针对该产品打造各类优化方案。而今这种合作已经促成了强大且成熟的生态系统。该系统中的合作伙伴可为用户使用傲腾™ 持久内存提供有力的支持。例如，作为英特尔 OEM 合作伙伴的宁畅就在本次瑞金医院转化医学中心的项目中发挥了重要作用，有效降低了瑞金医院部署新方案所面临的复杂性。

实战性能测试与验证

前文提及的 IO-500 是全球高性能计算领域针对存储性能最权威的排行榜之一。自 2017 年 11 月开始，该榜单每年都会在美国 Supercomputing Conference (SC) 和德国 International Supercomputing Conference (ISC) 上发布和更新。IO-500 榜单包含总榜单及 10 节点榜单两大类别，其中 10 节点榜单因更接近实际并行程序可能达到的规模，能更准确地反映出存储系统可为实际程序提供的 I/O 性能，所以具备更高的参考价值。

瑞金医院携手英特尔及其合作伙伴宁畅一同在 ASTRA 高性能计算平台中部署和导入了“英特尔® 傲腾™ 持久内存 + DAOS”优化方案，通过冲刺 IO-500 榜单的方式验证了自身的性能优势——这个将高性能计算和存储技术与医学研究相融合的高性能生信大数据计算平台，在国际超级计算大会 SC21 公布的 IO-500 榜单上，以高达 87.50 GiB/s 和 2984.61 kIOP/s 的带宽和吞吐性能拿下了 10 节点榜单第八名的排位[10]，并且成为中国及全球生信领域惟一一个打进 10 节点榜单前十名的系统，其在 IO-500 总榜单上的排名也高达第 14 名[11]。

具体来说，IO-500 通常会进行两组测试。第一组是理想状况下对 IO Easy: write/read 和 MDTest Easy: write/stat/delete 进行测试，ASTRA 的成绩见表 1。

理想状况下，存储系统的最优性能（例如大文件读写）		
IO	EASY WRITE	97.88 GiB/s
	EASY READ	102.35 GiB/s
MDTest	EASY WRITE	4896.24 kIOP/s
	EASY STAT	5041.12 kIOP/s
	EASY DELETE	2179.37 kIOP/s

表 1. 理想状况下 ASTRA 的 IO Easy 和 MDTest Easy 测试结果[12]

第二组是评估存储系统在极端场景下的性能底线，对 IO Hard: write/read 和 MDTest Hard: write/stat/read/delete、FIND 索引等指标进行测试。表 2 所示为 ASTRA 的测试结果：

设置一组苛刻的测试流程（如随机读写 3901Byte 数据），以及海量小文件的读写		
IO	HARD WRITE	71.71 GiB/s
	HARD READ	81.57 GiB/s
MDTest	HARD WRITE	1754.72 GiB/s
	HARD STAT	4392.24 kIOP/s
	HARD READ	3456.88 kIOP/s
	HARD DELETE	1561.46 kIOP/s
FIND	HARD	2813.66 kIOP/s

表 2. 极端情况下，ASTRA 的 IO Hard 和 MDTest Hard 测试结果[12]

展望未来：进一步优化性能并分享创新成果

在 IO-500 榜单上名列前茅，对瑞金医院转化医学中心 ASTRA 高性能计算平台而言只是“啼声初试”。接下来该中心还将继续深化与英特尔和宁畅的合作，针对更多实际应用场景来优化和验证该平台的性能表现。同时，作为国内首个建成并率先应用“英特尔® 傲腾™ 持久内存 + DAOS”方案的转化医学中心，瑞金医院转化医学中心也乐于与业界同仁分享相关经验，进而帮助更多医院和医疗研究机构高效挖掘生信大数据中蕴藏的可观价值，推进基础研究和临床实践的对接和融合，加速更多转化医学中心或项目的落地。

intel®

[1] 数据引自：https://www.fx361.com/page/2016/1222/419749.shtml
[2] 数据引自：https://baijiahao.baidu.com/s?id=1686482679324347723&wfr=spider&for=pc
[3] 数据引自：https://blog.csdn.net/qq_40459859/article/details/106521974
[4] 数据由瑞金医院提供，更多详情请咨询瑞金医院。
[5] 数据引自：http://www.360doc.com/content/18/0826/10/42202575_781290249.shtml
[6] 数据引自：https://www.intel.cn/content/www/cn/zh/high-performance-computing/daos-high-performance-storage-brief.html
[7] 数据引自：https://io500.org/list/sc21/historical
[8,11] 数据引自：https://io500.org/
[9] 数据引自：https://blog.csdn.net/nidongla/article/details/115352797
[10] 数据引自：https://io500.org/list/sc21/ten
[12] 数据引自：https://io500.org/submissions/view/584

英特尔并不控制或审计第三方数据。请您审查该内容，咨询其他来源，并确认提及数据是否准确。

英特尔技术特性和优势取决于系统配置，并可能需要支持的硬件、软件或服务得以激活。产品性能会基于系统配置有所变化。没有任何产品或组件是绝对安全的。更多信息请从原始设备制造商或零售商处获得，或请见 intel.cn。
性能测试中使用的软件和工作负荷可能仅在英特尔微处理器上进行了性能优化。诸如 SYSmark 和 MobileMark 等测试均系基于特定计算机系统、硬件、软件、操作系统及功能。上述任何要素的变动都有可能导致测试结果的变化。请参考其他信息及性能测试（包括结合其他产品使用时的运行性能）以对目标产品进行全面评估。更多信息，详见 www.intel.cn/benchmarks。
描述的成本降低情景均旨在特定情况和配置中举例说明特定英特尔产品如何影响未来成本并提供成本节约。情况均不同。英特尔不保证任何成本或成本降低。
英特尔、英特尔标识以及其他英特尔商标是英特尔公司或其子公司在美国和/或其他国家的商标。