

浙江省疾病预防控制中心 大数据解决方案

十几分钟至十几秒钟¹，千万级数据量、多条件模糊查询质的巨大飞跃



“最终选择的方案经过我们谨慎的评估和测试。方案部署后，中心的系统性能和工作效率有显著提升。一方面，查询数据的速度较原有方案相比提升了几十倍，另一方面，能够将省市间跨平台数据实现很好的同步。新的方案对我们在疾病研究工作方面助力非常大，也极大地改善了中心工作人员的工作体验。”

浙江省疾病预防控制中心
大数据平台项目负责人
叶飞

浙江省疾病预防控制中心是直属于浙江省卫生厅的一类事业单位，是全省疾病预防控制中心的业务技术指导中心、检验检测中心、科研培训中心和信息管理中心，规格相当于副厅级。

中心共有在职在编人员 348 人，其中各类专业技术人员占 92.82%。中心的科研能力近年显著提高，已获得各类科技项目 439 项，发表各类学术论文 2072 篇；也进一步加强实验室能力和资质建设，现建有 2 套独立的 BSL-3 实验室，23 个 BSL-2 实验室和 6 套标准的 PCR 实验室。中心于 2002 年通过国家实验室认可，2004 年通过国家级计量认证，现有认可项目涉及 23 个领域 1045 项。

由于中心承担着全省疾病预防与控制、突发公共卫生事件应急处置、疫情与健康相关因素信息管理、健康危害因素监测与控制、健康教育与健康促进、实验室检测分析与评价、技术管理与应用研究等七大职能，原有的数据中心方案存在巨大挑战。

挑战

- **传统方案性能差，信息处理速度远远落后，传统硬盘有很大弱点。**
疾控中心数据库中存在医疗影像、电子病历等大量的非结构化数据或半结构化数据。传统方案使用机械硬盘，性能差，速度慢。千万级数据量、多条件模糊查询一次需十几分钟。出具一份分析报告，需要一周的时间。
- **传统方案无法承受大量访问和频繁检索，普通固态硬盘反应时间难保证。**
单个病人一次诊断即需要存储或者调阅数百张影像。在疾控中心研究过程中，大量的频繁检索和访问对数据存储设备的稳定性是巨大的挑战，普通固态硬盘难以满足需求。
- **无法实现多个省市平台跨数据库实时同步。**
单个大型医院中，存储数据量以每年几十个 TB 快速增长。而全省省市平台的海量数据的同步对接和跨数据库查询对读写速度性能要求非常高，必须采用新的数据库技术及存储方案。



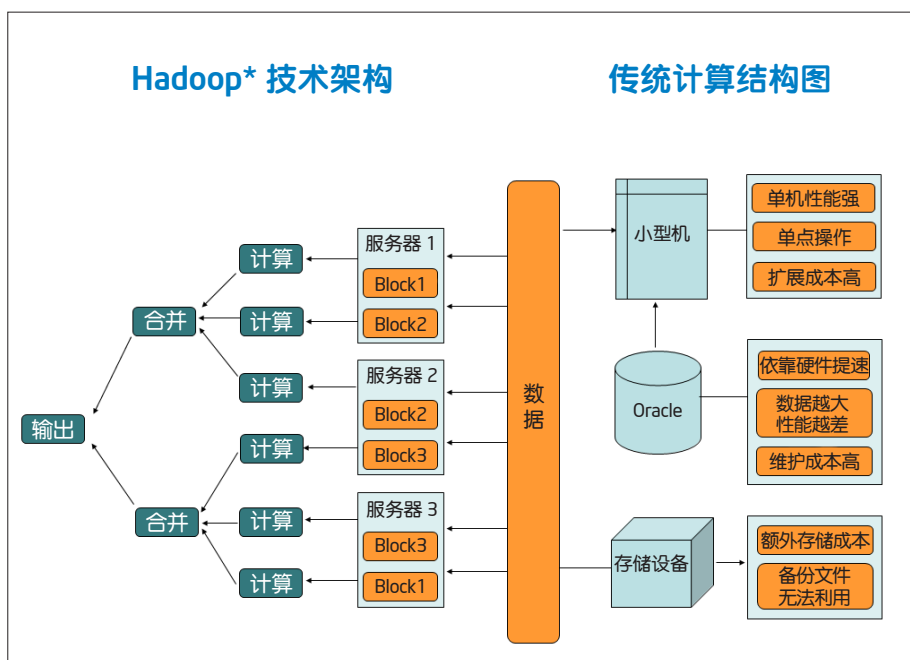
“应用英特尔® 固态硬盘 DC S3500 系列，网新易得基于 Hadoop 技术架构医疗大数据解决方案能够以高性能、高稳定性为客户带来更加出色和严谨的用户体验，同时也减少了方案后期维护成本，降低维护门槛。”

网新易得产品经理
董锴

解决方案：基于英特尔® 固态硬盘 DC S3500 系列的 Hadoop* 架构

一款合适的固态硬盘，能有效的解决高效数据读取和性能稳定的问题。英特尔® 固态硬盘 DC S3500 系列 160G 在高稳定性和高性能的卓越表现，能够非常好地

解决这个问题。配合网新易得采用的 Hadoop* 分布式数据库，海量医疗影像、电子病历等的跨数据库实时同步、快速检索得到实现。



影响

▪ 配备英特尔® 固态硬盘 DC S3500 系列，方案性能极大提高，免去漫长焦躁等待。

基于 Hadoop 技术架构,部署英特尔固态硬盘 DC S3500 系列之后，方案性能可以支撑上万 IOPS 的要求,将千万级数据库多条件模糊检索时间，从十几分钟缩短到十几秒¹。出具分析报告的时间，由原来的一周缩短到仅仅 1 小时¹。免去长时间焦躁等待，极大地提高用户体验。

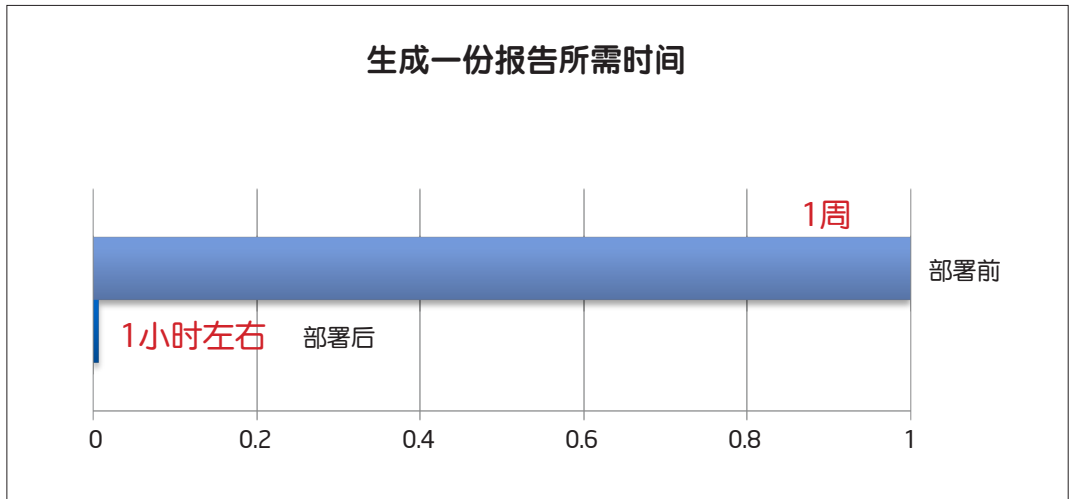
▪ 大量访问和频繁检索，性能稳定如初。

英特尔固态硬盘 DC S3500 系列使得 Hadoop 数据库服务更加稳定，其小数据块的高稳定的读写性能保证了在大量频繁访问和检索中的持续高性能。

▪ 多个跨数据中心实时同步，真正进入大数据时代。

面对高速增长的数据存量，复杂的数据结构以及大规模的数据运算，基于英特尔固态硬盘 DC S3500 系列的 Hadoop 数据库服务，使各级省市平台实现无缝对接，实时同步¹。这不仅极大程度满足救治病人和疾病研究的需求，也减小了数据库的维护成本和极大降低了维护门槛。





数据来自网新易得*

解决方案介绍

上海浙大网新易得科技发展有限公司（简称“网新易得*”）是由国内著名 IT 上市公司浙江浙大网新科技股份有限公司（沪市代码：600797）旗下的核心成员。作为一家常年专注于研发、生产、销售基于英特尔® 架构的品牌服务器的厂商，网新易得致力于为客户提供领先的应用解决方案及满意的售后服务。为了满足医疗行业数据中心的系统需求，网新易得采用了先进的英特尔 Hadoop 技术架构，与原有的传统技术架构相比，解决了医疗行业传统方案难以做到的跨数据中心读写、高速稳定的文件存储、大数据计算要求和数据安全保障等问题。

网新易得* 采用配备英特尔® 固态硬盘 DC S3500 系列的 Hadoop* 分布式数据库，帮助浙江疾控中心实现对系统的各项要求，疾控中心综合实力大大增强。

	传统数据库	配备英特尔® 固态硬盘 DC S3500 系列的 Hadoop* 分布式数据库
多数据库读取	<ul style="list-style-type: none"> 无法解决跨数据库读写 数据量越大性能越差，难以突破数据上限 合并数据时，无法解决数据冲突问题，数据标识不可重复 	<ul style="list-style-type: none"> 基于 hBase 分布式数据库，轻松实现大数据读写 数据规模可平滑扩展，突破数据极限 hBase 采用键值方式存储数据，同一标识支持多个不同版本
文件存储	<ul style="list-style-type: none"> 单文件读写，性能极差 备份需要额外昂贵的设备和软件支持 备份数据并不被使用，造成数据浪费 	<ul style="list-style-type: none"> 基于 HDFS 分布式文件系统，自动备份文件 无需额外设备和软件开发 通过分布式读写，数据被日常使用，极大提高读写速度
数据处理	<ul style="list-style-type: none"> 难以应对大数据的计算需求 单服务器结构成为性能瓶颈 	<ul style="list-style-type: none"> 基于大规模分布式计算框架，提供对大数据的提取和计算 数据热点功能，自动均衡系统内数据分布，避免单节点瓶颈
数据安全	<ul style="list-style-type: none"> 需独立昂贵的存储设备支持和软件支持 数据恢复过程，需停机维护，造成业务中断 	<ul style="list-style-type: none"> 基于 HDFS 的自动备份机制，无需额外操作 智能算法自动修复数据损坏，保证数据均衡与安全

表格来自网新易得*

而在 Hadoop 分布式数据库中，为了实现大数据读写、提供对大数据的计算，必须要求硬盘很高的读写能力和稳定性。传统的企业级机械硬盘，无论是采用 SATA 接口还是 SAS 接口，在 I/O 能力上严重不足。而一般的固态硬盘，则存在稳定性的瓶颈，难以在大量频繁操作的情况下仍旧保持高性能。因此，使用新的存储技术和模式势在必行。

网新易得产品经理董锴说道：“原有的传统技术架构采用 SATA/SAS 机械硬盘作为存储介质，虽然可以满足容量的需求，但是故障率和 IO 能力都会成为制约整体解决方案的瓶颈。而普通的固态硬盘，性能极限虽有大幅提高，但高性能却无法稳定地持续。我们必须寻求新的产品，来满足 IO 方面和稳定性的需求。”借助英特尔® 固态硬盘 DC S3500 系列的产品和服务，网新易得* 有效地提高了 Hadoop* 数据库服务的性能，稳定性也获得了显著的增强。

在比较市场上各品牌型号的固态硬盘的信息之后，网新易得最终选择了具备高稳定性技术的英特尔固态硬盘 DC S3500 系列

来构建新的解决方案。英特尔固态硬盘 DC S3500 系列产品稳定的高性能，使网新易得的新解决方案得到了更有效的保障。而英特尔® 固态硬盘技术人员专业而严谨的前期产品测试、积极响应的售后服务方案和强大的渠道供货能力，更是提供了强有力的后盾。

“在我们实际测试中，基于英特尔® 固态硬盘 DC S3500 系列的方案可以提供高达百万级别的 IOPS，尤其是在数据读取性能方面，较机械硬盘提升了 40%¹。这给我们的系统性能提供了很大的支持。而英特尔技术人员提供的更严谨和全面的技术测试，更是让我们相信英特尔® 固态硬盘团队在技术以及服务方面的专业性。”网新易得产品经理董锴说道。

对于固态硬盘而言，影响读写能力的最大因素在于固态硬盘的控制芯片和固件算法。英特尔在芯片技术和固件算法上的优势让英特尔固态硬盘 DC S3500 系列在处理读写任务时拥有了得天独厚的优势。英特尔固态硬盘 DC S3500 系列具有 50 微秒的一半顺序读取延迟，在 99.9% 的时间内 500 微

秒的最大读取延迟³，以及 4KB 随机读取性能高达 75,000 的每秒输入输出操作次数 (IOPS)²。同时具备较低的有功读取功率 (小于 1.3 瓦)²。这些特点保证了基于英特尔固态硬盘的医疗行业解决方案数据读取的出色能力以及在频繁操作情况下的性能稳定性。

作为一款优秀的企业级固态硬盘，除了高性能优势之外，英特尔固态硬盘 DC S3500 系列还以全面的特性防止数据丢失和损坏。英特尔固态硬盘 DC S3500 系列提供端到端安全性保护，从进入硬盘到离开硬盘，数据时刻出于保护之中。此外，英特尔固态硬盘将 NAND、SRAM 和 DRAM 内存的高级错误纠正方案结合起来，还通过奇偶校验、循环冗余检查 (CRC) 和逻辑块地址 (LBA) 标记验证来保护传输中的数据。它还使用高级加密标准 (AES) 和增强的电源损耗保护进一步增强数据的安全性。

未来，网新易得还将继续与英特尔紧密合作，凭借性能出色的企业级固态硬盘，有效增强解决方案能力，为用户提供更优质、更可靠的服务。

英特尔和英特尔标识是英特尔公司在美国和/或其他国家或地区的商标。

*文中涉及的其它名称及商标属于各自所有者资产。

性能测试和等级评定均使用特定的计算机和/或组件进行测量。这些测试反映了英特尔产品的大致性能。系统硬件、软件设计或配置的任何不同都可能影响实际性能。购买者应进行多方咨询，以评估他们考虑购买的系统或组件的性能。如欲了解有关性能测试的英特尔产品性能的更多信息。请访问：http://www.intel.com/performance/resources/benchmark_limitations.htm

¹ 数据来自网新易得*

² 基于英特尔® 固态硬盘 DC S3500 系列产品规格

³ 设备测量 使用 Iometer (4K 随机写入且队列深度为 32) 在整个固态硬盘范围上进行。延迟使用传输大小为 4KB (4,096 字节) 且队列深度为 1 的设置测量

版权所有 © 英特尔公司。所有权利受到保护。

