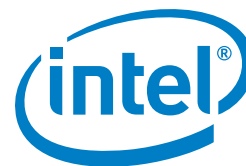


成功案例
英特尔® 至强™ 处理器 E5-2600 系列
区域医疗
大数据 / 云服务



区域医疗大数据应用案例

锦州, 户籍人口规模约 310 万, 辽宁西部城市群区域性中心城市和重要港口城市。2012 年 3 月, 锦州市政府正式启动“健康城市”战略。建立以城市为单位建设的区域卫生数据中心, 以居民健康卡为纽带, 覆盖卫生信息化的主要业务领域, 存储的数据将包括居民健康档案、电子病历、公共卫生及综合管理等相关数据, 20 年左右的数据量预计将达到 PB 级。随着“健康城市”理念的提出, 个人健康数据将伴随着居民对于个人健康的关注, 以及健康管理相关服务的推广而大幅增加。

卫生部于 2010 年编制了全国卫生信息化“十二五规划”, 简称“3521-2”工程。具体来说, “3”代表国家、省、市三级卫生信息平台; “5”代表医疗服务、公共卫生、医疗保障、药品保障和综合管理五大卫生信息化业务领域; “2”代表健康档案和电子病历两大基础资源库; “1”代表卫生信息专网; 后一个“2”代表信息安全体系和标准规范体系。其中, 市级卫生信息平台, 即区域卫生信息化, 是卫生信息化的核心内容之一。

用友
yonyou



我们在锦州市区域医疗项目上基于健康卡和健康档案的“一卡一档”的基础性建设中, 感受到大数据实现了过去不敢期望的性价比及实时响应性要求。由于健康档案属于基础性, 长期存续建设项目, 要管理从市民的出生到死亡的全过程。因此前期规划非常重要。英特尔公司在大数据解决方案方面起步比较早, 在这个项目上英特尔给了我们很大的帮助。

——徐春华

大数据在区域医疗卫生的应用

目前，大数据技术最具实际意义的应用在于健康档案数据的管理和服务。“健康档案是个人全生命周期的医疗 / 健康数据的管理”。

- 从医生及卫生行政管理人员的角度来看，全生命周期的健康档案调阅有着现实意义。例如，对于慢性病患者，以往病程的变化，治疗的过程都对医生诊断和处置有着重要的辅助作用。过敏史，不良反应这些数据对避免出现医疗差错和事故也有着积极的作用。
- 对于海量的医疗及健康数据进行统计和分析，为管理决策、监管实施，提供了更为科学的依据。
- 传统的临床科研往往基于抽样调查进行，而随着健康档案数据的丰富可以大幅减轻工作量，同时提高科研数据的质量和数量以及数据处理的效率。

未来的大数据的利用前景是十分广阔的，不仅用于临床诊断，临床科研，而且为政府公共卫生决策及个人管理健康都会发挥积极的作用。

挑战

用友医疗卫生信息系统有限公司（简称：用友医疗），作为锦州“健康城市”战略区域卫生数据中心的主要建设方，拥有大量省市级以及区县智慧健康城市区域卫生解决方案项目建设成功经验。用友医疗在采用传统关系型数据库在锦州的区域卫生数据中心建设中，逐步发现一些潜在的建设问题：

数据量庞大

- 区域卫生数据中心以城市为单位，将存储健康档案、电子病历、卫生管理三大数据库，锦州作为 300 万人口的中等规模城市，其卫生数据中心的规模在 20 年后预计会达到 PB 级：
- 传统关系型数据在大数据存储实现中存在局限性，在一张表中存储 500GB 之上的数据存在性能方面的问题。

数据类型复杂多变

锦州的区域卫生数据中心，将会存储大量的非结构化数据和半结构化数据。如果采用传统关系型数据库，则面临众多不易解决的问题：

- PACS 影像，B 超、病理分析等业务产生的非结构化数

据大小不一，从数百 KB 到数百 MB 都有，病人的一次诊断活动中需要存储、调阅数百张影像数据；

- 临床电子病历数据一般采用符合 HL7 CDA 标准的 XML 文件格式，文件格式随着时间变化，会不断的演变；
- 由于医疗卫生行业的业务复杂性，很难制定统一的数据标准，这就给数据访问和交换带来新的挑战。

另外，未来的数据处理也将面临巨大挑战，譬如未来海量数据的存储，备份，扩容等，快速的检索特定数据，以及高效的数据交换等等，这也是锦州区域卫生数据中心未来必会产生的新的需求。



基于 Hadoop 实现海量数据的处理和分析

为了解决上述诸多问题，英特尔公司协助用友医疗进行了合理的架构分析和指导，对于基于大数据分析的解决方案进行了深入的探索和研究。并且制定了基于英特尔大数据解决方案的区域卫生数据中心建设目标：

- 在海量数据情况下，数据中心必须具有文档快速检索的性能；
- 随着未来医疗系统的升级，医疗标准的版本升级，数据中心必须能够解决存储模式如何满足数据模式的更新的问题；
- 数据中心必须具备水平扩展能力和对应用程序透明的能力，要求能做到底层扩展对上层业务的隔离，通过更多的服务器成比例的透明化扩展容量和性能。

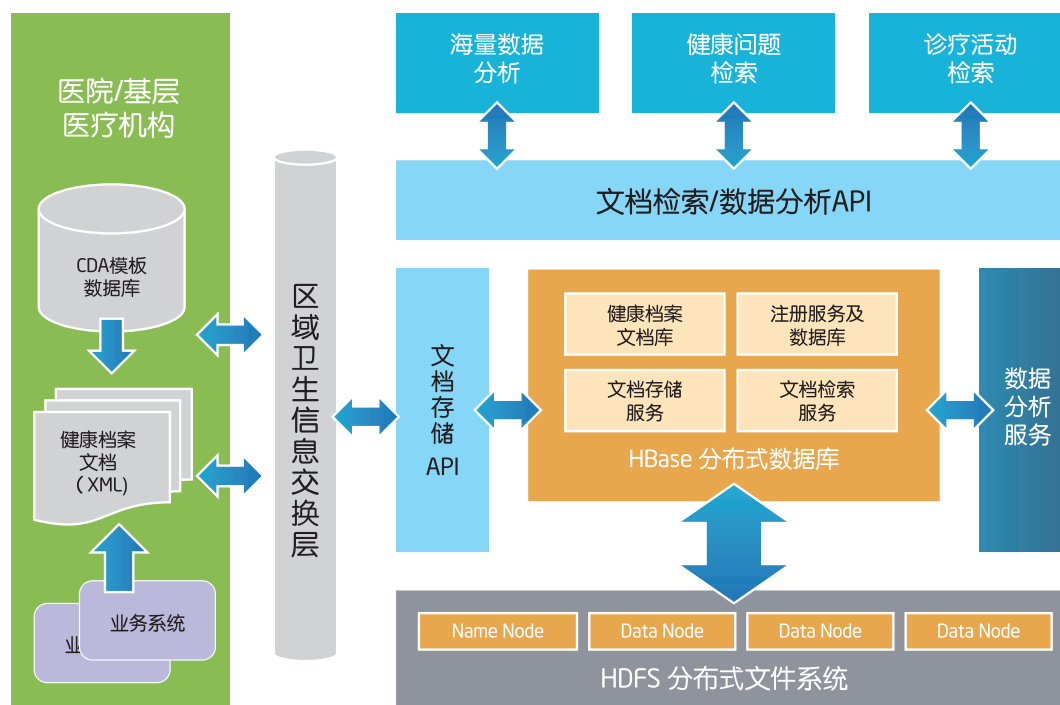


图 3 基于 Hadoop 的区域卫生信息平台数据处理流程

个人基本健康信息登记：1 千万条

门诊摘要：3 千万条

处方记录：1 亿条

糖尿病专档：2 百万条

糖尿病随访：2 千 4 百万

区域内的各医疗卫生服务机构，调用区域卫生信息平台提供的服务，将遵从 HL7 CDA 模板设计的健康档案文档（XML 格式），上传至区域卫生信息平台。采集的 CDA 文档类型包括个人基本健康记录登记、门诊摘要、处方记录、糖尿病专档和糖尿病随访等不同类型的文档（文档）。文档上传后，通过 XML 解析，提取关键数据元素（META-DATA），调用文档存储服务将关键数据元素和 XML 文档存储在 hbase 数据库中，形成健康档案文档库。

基于大数据的区域卫生信息平台提供文档检索服务和数据分析服务，通过开放的文档检索 / 数据分析接口，针对个人（居民、患者）、医生和卫生管理人员提供不同类型的

数据和业务服务，满足不同用户类型对于海量数据访问、处理和分析的需求。

结论

经过多次反复单业务负载测试、大数量测试和调优等多种技术相结合的手段，英特尔公司和用友医疗在锦州区域卫生数据中心，利用基于英特尔® 至强™ E5 服务器平台，以及英特尔® Hadoop 发行版（Intel Distribution），成功的搭建了完整的区域医疗大数据计算架构。该架构可以满足海量数据（一亿以上记录数）的高并发检索和实时数据分析的性能要求。证明了基于英特尔® 至强™ E5 服务器平台以及英特尔® Hadoop 发行版的大数据整体解决方案，可以作为基于大数据分析的智慧健康云服务平台提供数据处理、检索和分析等服务的基础架构。

价值

业务价值

英特尔与用友医疗合作的基于大数据量优化的整体解决方案：

- 使得医生可以快速检索就诊患者的个人基本信息、既往病史、就诊及处方等重要信息。通过辅助信息，帮助

医生做出更准确的诊断，并且有效避免重复用药、药物不良反应等；

- 支持高并发的大数据检索，使得个人（居民、患者）通过便捷快速的方式访问自己的健康档案成为可能。与居民健康教育相结合，可以帮助居民建立起个人健康管理理念，从治疗转向预防；
- 支持灵活的数据分析建模和数据分析挖掘，为区域卫生管理者提供了接近实时的数据分析结果，可以广泛应用于疾病控制、医疗行为监管、医疗质量管理等领域。

IT 价值

英特尔与用友医疗合作的基于大数据量优化的软硬件解决方案，实现了：

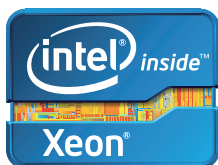
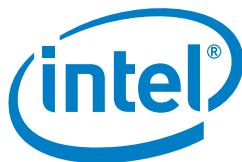
- 海量数据存储：应对超过一亿条记录（文档）的存储需求；
- 数据格式动态扩展：Hbase 提供的列式存储方式，可以

轻松应对数据格式的灵活调整；

- 海量数据快速检索：高并发的并行海量数据检索，满足居民和医生实时检索健康档案的需求；
- 统计分析：开放的统计分析框架，接近实时的统计分析能力；
- 平滑扩容：横向扩展性能卓越，应对未来业务和数据量的不断增长。

经济价值

- 降低主机存储建设成本：与传统数据库和小型机的方案相比，英特尔大数据解决方案具有更好的开放性和更好的经济性；
- 高可扩展性，集成的图形化界面提供了便捷的节点管理和横向扩展功能，大幅降低了管理成本；
- 充分利用现有硬件资源，降低升级成本。



法律声明：

本文档提供有关英特尔® 产品的信息。本文档并未以明示或暗示，或以禁止反言或其他方式授予任何知识产权许可。除英特尔在其产品的销售条款和条件中声明的责任之外，英特尔概不承担任何其他责任。并且，英特尔对英特尔® 产品的销售和 / 或使用不作任何明示或暗示的担保，包括对产品的特定用户适用性、适销性或对其任何专利权、版权或其他知识产权的侵权责任等，均不作担保。除非已获得英特尔的书面认可，否则英特尔® 产品并非设计为或用于任何英特尔® 产品故障可能导致人身伤害或生命危险的应用。

版权所有 © 2013 英特尔公司。保留所有权利。英特尔、Intel 标识、Xeon、至强、Xeon inside 是英特尔公司在美国和其他国家（地区）的商标。

* 文中涉及的其他名称及商标属于各自所有者资产。