



智慧医疗

终端与大数据

黄庆春

英特尔数字医疗 中国区经理

Qingchun.huang@intel.com

无所不在的大数据时代



Netflix公司成立于1997年，目前在美国拥有2700万用户，在全球拥有3300万用户

Netflix分析了3000万次用户体验，包括观众何时暂停、后退和快进，分析了400万条用户评价，300万次用户搜索操作，以及用户观看视频的时间和使用终端等。

发现鬼才导演大卫·芬奇和演技派演员凯文史派西在用户中受关注程度很高，而且英国的一部政治剧《纸牌屋》很受欢迎。于是，便有了现在这部根据观众口味制作的美版《纸牌屋》，创作者按照美国的政治生态对剧情进行了全新创作。

从啤酒/尿布到谷歌流感趋势

在识别流感疫情时，
谷歌比疾病控制和
预防中心更有效



大数据 ， 四大区别于传统数据的特征



大数据技术将被设计用于在成本可承受（economically）的条件下，通过非常快速（velocity）的采集、发现和分析，
从大数据量（volumes）、多类别（variety）的数据中提取价值（value），但是传统关系型数据库存在性能、存储、成本、IO瓶颈、等等的原因，无法支撑这4个V+1个E的要求，解决方法就是分布式技术、廉价的X86平台、本地存储，点出分布式技术是大数据处理的核心

大数据，无法在传统架构上发挥.....

关系数据库 vs. Hadoop/Hive

Use Case: Intel Silicon Design Environment Usage Analysis

Data: 40M records/day



医疗行业数据同样面临这些大数据的挑战

数据量

检验结果，费用数据，影像，设备产生的感应数据，基因数据等

类型

- 结构化数据，遵循标准的数据标准(如, HL7)
- 非结构化数据，如口述、手写、照片、影像等

价值

基于现有数据库中的数据进行分析，来支持不同类型的业务：如费用及报销、患者病史、归档影像分析、实时临床决策支持（数据分析）

速度

- 实时数据分析，而非传统的批量处理分析
- 数据以流的方式进入系统，进行抽取和分析
- 对于实时运行中的每个时间节点产生影响，而不是事后处理

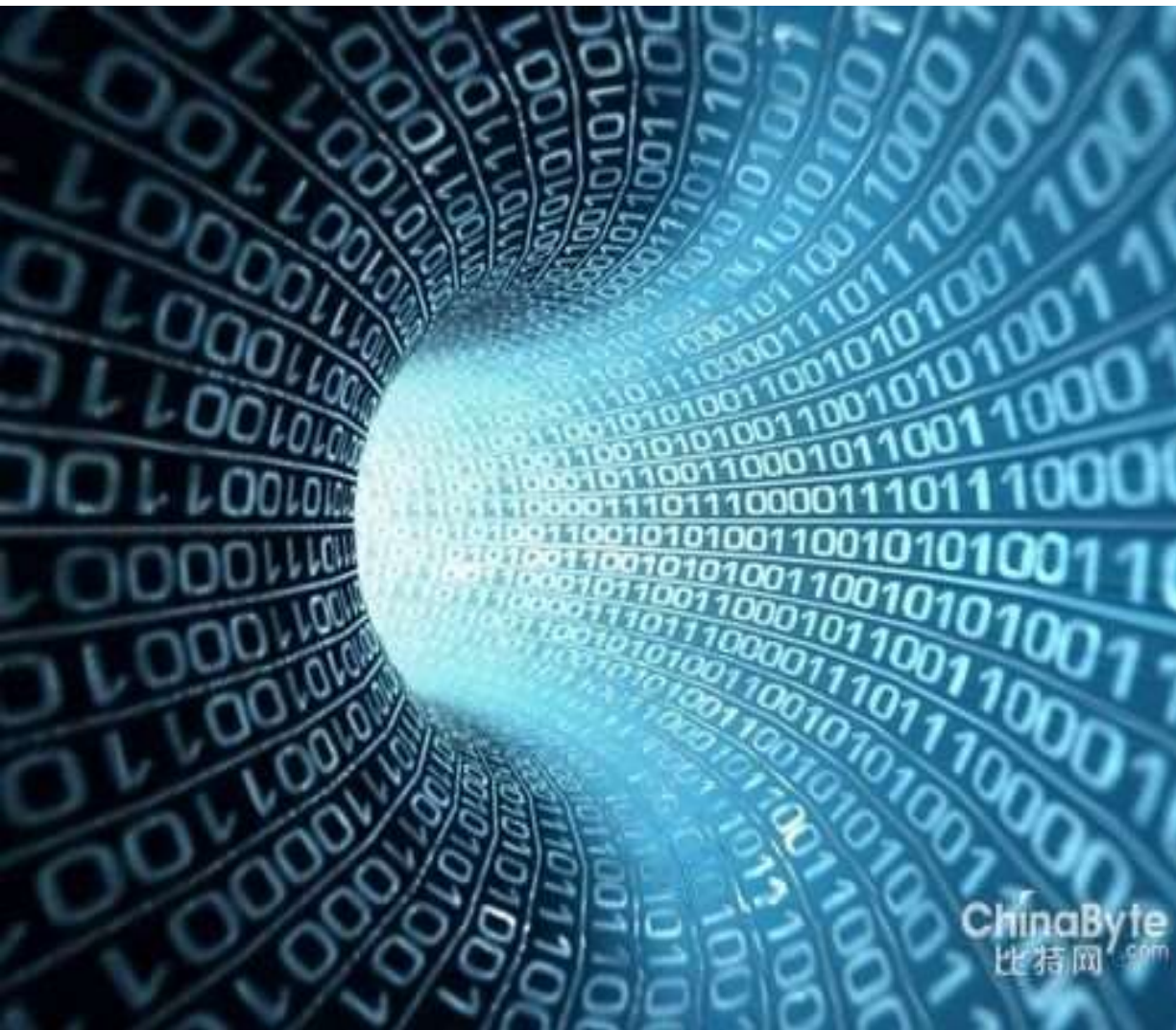
在传统的解决方案之上，引入新的数据及分析模型和技术，
实时有效的商业价值

ONC (Office of the National Coordinator) 谈到医疗信息技术时认为 ...

- “大数据将改变医疗卫生服务模式”, ONC在谈到医疗信息技术未来五年战略规划时这样说
- 通过一个知识型的医疗系统, 正确的信息将用于支持诊疗的决策. 它可以是有关提高单个病人治疗, 给药的后续效果, 也可以是预测全国性的流行疾病发展趋势, 或者是用于决定是否开展研究和开发新的潜在的治疗方案



我们希望解决的问题是



从大海捞针到寻找同质的模式，在现有的分散的数据中（包括非结构化数据）进行模式匹配，行为预测，发现非正常事物，推送数据服务

大数据价值的产生需要一系列的过程

Assess the present, Model the future

大数据基础设施

Hadoop
Framework

有效
可信
负载能力
管理
工具
网络
计算
存储

数据采集和存储

数据处理

Data
Science

机器学习
算法
分析
整合
查询效率
转换
入库

利用算法处理数据

数据使用

Domain
Expertise

业务策略
KPIs
业务发展
报表
可视化
结构

提出合适问题

= Value



发挥大数据的价值依赖于数据处理的质量，精确度和效率

大数据在医疗行业中主要发展方向

•药品研发

对药品实际作用进行分析；实施药品市场预测

•基因测序

•分布式计算加快基因测序计算效率

1. 制药企业/生命科学

2. 支撑临床诊疗服务 & 其他临床应用（包括诊断相关的影像信息）

•临床数据比较

比较同类型/异类型的病人的诊疗方案及用药

•临床质量监控

利用规则和数据实时分析给出智能提示

•开展典型疾病的研究

•公共卫生实时统计分析

发现公共卫生疫情及公民健康状况

•新农合基金数据分析

及时了解基金状况，预测风险辅助制定农合基金的起付线，赔付病种等

•基本药物临床应用分析

分析基本药物在处方中的比例

3. 费用报销，利用率和 欺诈监管

4. 患者行为/社交网络

•创新健康管理

采集并分析病人随身携带仪器数据，给出智能建议

•人口统计学分析

对不同群体人群的就医，健康数据实施人口统计分析

•了解病人就诊行为

发现病人的特定就诊行为，分配医疗资源

移动终端+可穿戴设备=多样化非结构数据



更好的点对点的
服务和管理



与外围设备的
联接



工作的实时性



支持多类型
工作



随时随地的
接入查看



移动医疗智能终端带来大数据采集和使用

数据使用consume

- 查看检查结果
- 查看EMR/HER
- 查看病人管理系统
- 获得影像数据
- 医疗参考和计算
- 药品数据库



数据创建create

- 获得记录
- 护士和医生的备注
- 图表
- 自动生成文档
- 药品管理

团队协作collaborate

- 护理团队协作
- 提出咨询或会诊
- 视频会议或会诊
- 护理计划
- 临床路径

数据流通circulate

- 电子处方
- 医嘱录入
- 执行或跟踪转移
- 床边远程报警
- 接收检验结果

移动医疗智能终端带来大数据采集和使用（续）



	智能手机	平板电脑	笔记本/超级本	台式机
移动性	●	●	●	◐
生命体征, 输入输出	◐	●	◐	◐
药品管理	◐	◐	◐	◐
固定格式数据输入	◐	●	●	●
自由格式数据输入	◐	◐	●	●
大影像数据	◐	◐	◐	●
数据查看	●	●	●	●
已管理性	◐	●	●	●



Ultrabook is a trademark of Intel Corporation in the U.S. and/or other countries.

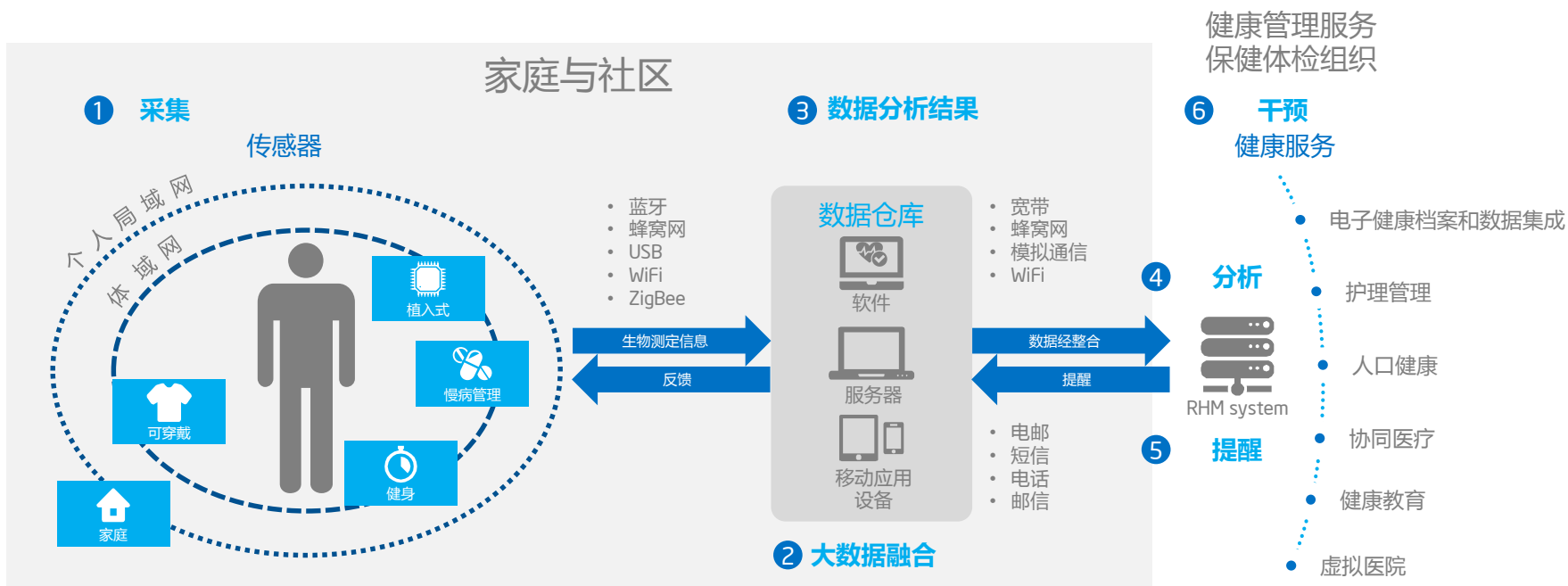
Ultrabook™ products are offered in multiple models. Some models may not be available in your market. Consult your Ultrabook™ manufacturer. For more information and details, visit intel.com/ultrabook.

移动医疗智能终端带来大数据采集和使用（续2）



移动医疗智能终端带来大数据采集和使用

----- 社区与居民



新一代可穿戴生物技术



牙嵌入式传感器使牙医了解对象饮食习惯

保健数据流



可视化运动装备实时反应佩戴者的活动水平

运动反应



脉冲响应手链发送热电加热或冷却，调节体温

主动生理调节



Kimono* 和服传感器监测婴儿的温度、呼吸和与父母沟通。

睡眠监控



智能手机的可摄入密码丸设备

人体身份验证



可穿戴相机自主捕获社交拍照

云分享

大数据主要技术平台



Greenplum



Hadoop 大数据平台

发行版组件

稳定的企业级hadoop发行版

利用硬件新技术进行优化

HBase改进和创新，为Hadoop提供实时数据处理能力

针对行业的功能增强，应对不同行业的大数据挑战

Intel Hadoop Manager 2.2

安装、部署、配置、监控、告警和访问控制

Sqoop

1.4.1

关系型数据库
同步工具

Mahout 0.6

机器学习

R - statistics

数据统计

Hive 0.9.0

交互式数据仓库

Pig 0.9.2

数据流处理语言

Map/Reduce 1.0.3

稳定高效的分布式计算框架

分布式、高维数据库Hbase

HBase 0.94的改进和创新，提供即时数据处理

HDFS 1.0.3

可靠的分布式文件系统

Zoo

keeper

3.4.5

分布式
协作服务

Flume

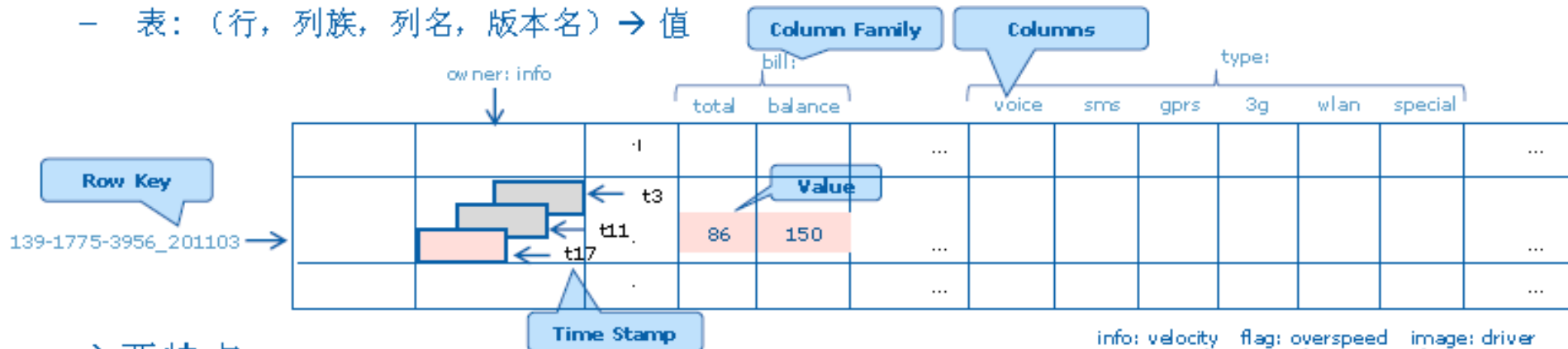
1.1.0

日志收集工具

HBase 分布式数据库

- HBase是一个分布式的、按列存储的、多维表结构的实时数据库，为高速在线数据服务而设计

- 表：（行，列族，列名，版本名）→ 值



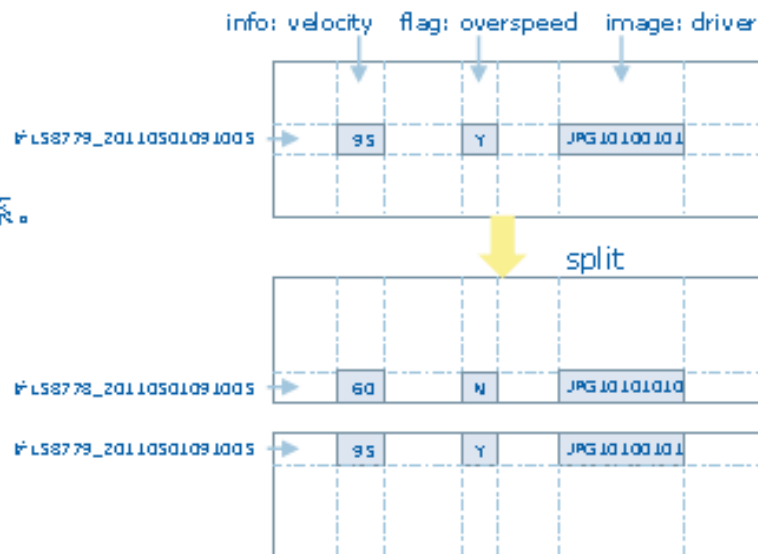
- 主要特点

- NoSQL

- 面向列、可压缩，有效降低磁盘I/O，提高利用率。
- 多维表，四个维度，其中三个维度可变，适合描述复杂嵌套关系。
- 灵活的表结构，可动态改变和增加（包括行、列和时间戳）。
- 支持单行的ACID事务处理

- 分布式系统

- 高性能，支持高速并发写入和高并发查询；
- 可扩展，数据自动切分和分布，可动态扩容，无需停机；
- 高可用性，建立在HDFS分布式文件系统之上。



医疗大数据案例一

Humedica 数据分析平台

拥有30多个国家近25万患者数据

包括药物，实验室，生命体征，患者基础数据，住院，门诊，医生笔记
数据来自IDN，学术机构，医院及医院集团

帮助纽约长老会医院利用数据分析降低了近25%的致命的血凝事件

帮助赛盾家庭医院关注充血性心力衰竭（congestive heart failure）
预测模型，使得以上提前感知CHF病人的风险。

┌ NewYork-Presbyterian
└ Lower Manhattan Hospital



医疗大数据案例二（公共医疗）



1. **Survey of Patient's Hospital Experience:** Percent of respondents who indicated top box response (e.g., "always;" overall rating of 9-10; Yes, Definitely recommend.) across seven customer experience questions and two patient loyalty questions.
2. **General Hospital Information:** Describes the hospital type and the owner.
3. **Outcome Measures:** Includes three mortality rates and three readmission rates for: heart attack, heart failure, and pneumonia
4. **Process of Care Measures:** 12 measures related to surgical care improvement
5. **Hospital Acquired Condition (HAC) Measures:** Percent of patients who acquire HAC.
6. **Medicare Spend per Patient:** This measure shows whether Medicare spends more, less, or about the same per Medicare patient treated in a specific hospital, compared to how much Medicare spends per patient nationally.

Map of US Hospitals and their Patient Experience Ratings



The Obama administration recently announced their "Big Data Research and Development Initiative," in which they are committing more than \$200 million in new commitments to Big Data projects.

医疗大数据案例三

研究者在洛杉矶儿童医院利用Hadoop采集和分析医疗传感数据

问题在于:

需要从传感器或医疗仪器中获取大量的（几十亿条）生命体征数据，并且和病人关系起来。护士需要定期的捕捉这些信息并存入EMR中（大约一天四次）。然而研究者需要获取整个历史数据。

解决方案：

持续的将传感数据存储
在HDFS中，而且利用
分布式存储大大降低了
成本。系统可以通过实
时分析，监控患儿发生
的任何情况，并且提前
预判和报警。



医疗大数据案例四：药品不良反映上的应用（ADE）

Beth Israel Deaconess Medical Center

拥有病人病历庞大的数据库连接起来的搜索引擎，能够让医院员工测试到底是什么原因导致一种疾病的发生，例如测试药物、饮食或生活习惯这些可能导致疾病发展的变量。

该库包含了220万名患者的200万个数据点，包括药物服用、诊断和实验室值。

在临床查询的帮助下，临床医生或研究人员可以搜索记录，以找出多少名乳腺癌患者也服用ACE抑制剂，这是一类用于治疗高血压的药物。如果药物和恶性肿瘤之间显示出很强的相关性，医院可以做更深入的分析，并建立一个正式的研究项目来做调查。



医疗大数据案例五：区域医疗平台（RHIN）

案例分享——上海健康信息网云计算应用示范工程



- 覆盖全市所有的医疗卫生机构
- 为市民开展自我健康管理，享有优质医疗卫生服务提供信息支撑。
- 市平台首期投资近8000万，各区及机构平均投资3000~4000万
- 健康档案调阅由10台刀片虚拟为40台服务器提供服务
 - VMWare+EMC
- 部署Intel Hadoop
- 10亿级别的数据量
- 普通查询1秒内响应
- 数据挖掘30秒内完成



医疗大数据案例六：医院临床数据仓库（CDR）

- 医院信息化现状

- 借助业务系统简单查询统计功能；
- 基于住院病案首页数据统计分析；
- 信息中心临时提供各种数据提取；
- 借助BI工具进行数据分析与展现；



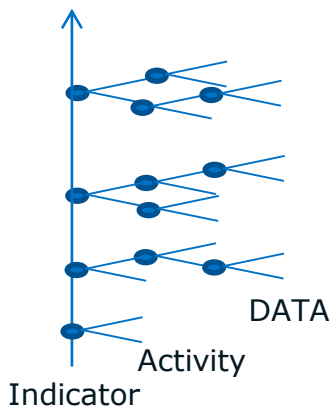
- 业务要求

- 1989年11月29日
 - 医院分级管理与评审工作正式启动
 - 医院服务，患者安全，医疗质量安全管理与持续改进，护理管理与质量持续改进，医院管理，日常统计学评价
- 2011年12月24日
 - 颁布《三级综合医院评审标准实施细则(2011年版)》
 - 7章73节378条标准与监测指标
- 2012年5月11日
 - 颁布《二级综合医院评审标准实施细则(2012年版)》



医疗大数据案例六：医院临床数据仓库（CDR）续

解决方案



- 集中式CDR系统的部署
- 利用Hadoop的存储结构和分布式计算能力
- 整合数据管理和数据分析功能
- 良好的性能价格比以及扩展性



hadoop



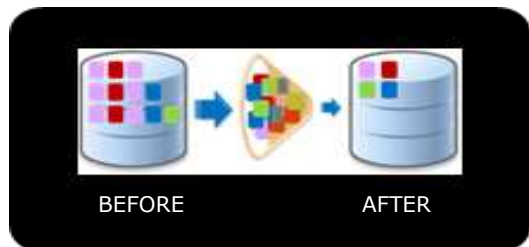
2010年，BCG(波士顿咨询集团)相关调研报告显示，“嘉和美康”与其他3家电子病历厂商一起占有了国内电子病历产品60%以上的市场份额。---波士顿

公司拥有众多的遍布全国各省市的分销商网络。产品销售覆盖全国，已形成很好的销售网络。公司产品销售额成逐年上升趋势，公司在市场上份额也是逐年上升，目前公司占据市场份额的50%以上。--嘉和美康网站

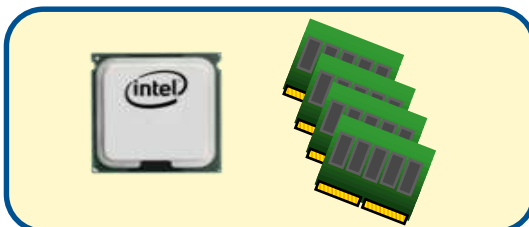
其他大数据案例.....

业务应用		
临床数据比对	匹配同类型的病人，用药	
临床决策支持	给出智能提示	
实时统计分析	展示公共卫生统计数据	
远程病人数据分析	监控，分析临床监护数据	
人口统计学分析	对不同体质人群分类	
就诊行为分析	跟踪健康卡数据，分析病人就诊行为	
基本药物临床应用分析	分析基本药物在处方中的比例	
药品研发	基于大数据的药品市场预测	
新农合基金数据分析	辅助制定农合基金的起付线，赔付病种等	
新的服务模式	根据临床数据分析，提供个性化服务及自助服务等新模式	
。 。 。		

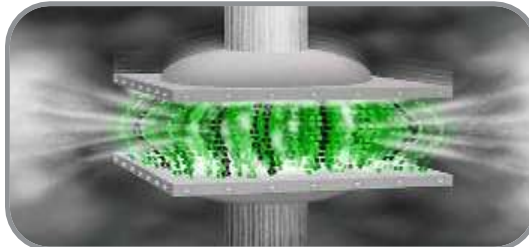
大数据对信息化基础设施带来的变化



存储：从传统的集中式存储转向分布式存储



计算力：依赖集群的计算能力而不是单机的能力



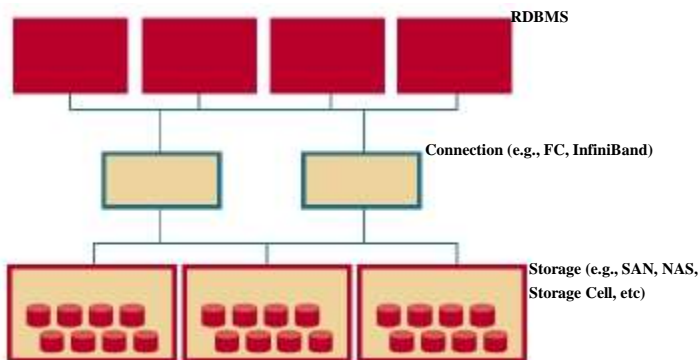
网路：万兆网络促进了集群节点间的通讯



扩展性和容错：更强调扩展性同时兼顾容错

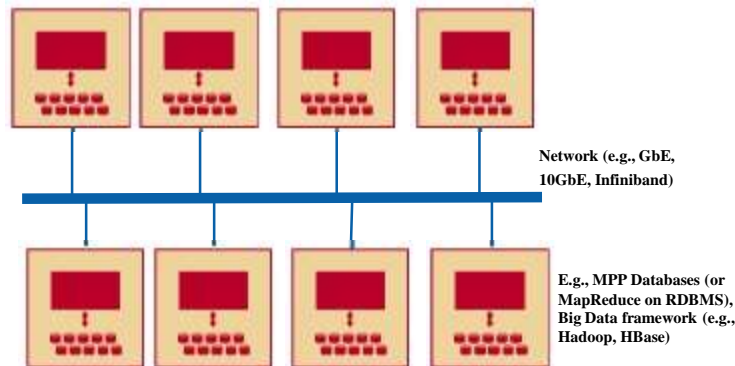
传统数据处理与Hadoop大数据架构对比

RDBMS



- 多个独立的关系数据库服务器，访问共享的存储资源池
- 优势：
 - 采用多个关系数据库服务器，多个存储，与原有的架构相比，扩展了存储容量和计算能力；
- 劣势：
 - 计算与存储分离，数据访问存在竞争和带宽瓶颈；
 - 支持的关系数据库服务器数量有限；
 - 只能向上扩展(scale-up), 不能横向扩展(not scale out)
- 适合复杂的需要事务处理的应用

NewSQL 与 NoSQL



- 由大量独立的服务器通过网络互连形成集群，每个服务器带存储。
- 优势：计算与存储融合，支持横向扩展，更好的扩展性
- 劣势：解决数据冲突时需要节点间协作
- 适用范围：
 - 数据仓库和离线数据分析 (MPP, Hadoop/HBase)
 - 大规模在线实时应用 (单行事务处理能满足的场景) (HBase)

医疗大数据相关解决方案总结

健康信息服务

基础医疗服务

个人健康管理

老龄社会

新兴的医疗服务
应用

临床决策支持

个体化医疗

肿瘤基因组学

数据分析及
视觉化处理

类SQL的检索

机器学习

医疗影像分析

数据处理/
管理

医疗记录

基因数据

医疗影像

分布式平台

存储优化

安全和隐私

影像数据处理加速

The background of the slide features a photograph of medical professionals in a clinical setting. In the foreground, a young woman with dark hair, wearing a white lab coat and a stethoscope, smiles warmly at the camera. In the background, two other individuals, a man and a woman, are blurred, appearing to be engaged in a task, possibly a medical procedure or a discussion, in a brightly lit room.

黄庆春
Qingchun.huang@intel.com
13600464413