

# BIG DATA

## World Forum 2011

2011大数据世界论坛

The Business Value of BIG DATA

[BigDataWF.com](http://BigDataWF.com)

2011年7月7日

# 大数据遇到大智慧

## IBM 大数据平台概述

薛峰 Steven Sit  
-大数据计划主管



## 免责声明

© 版权所有 **IBM Corporation 2010**。保留所有权利。

美国政府用户享有的受限权利 - 使用、复制或披露受到与 **IBM Corp** 签订的 **GSA ADP** 时效合同的制约。

本演示文稿中所含的信息仅供参考使用。尽管我们已努力确保本演示文稿中所含的信息完整、准确，但这些信息“按原样”提供，不附带任何形式的明示或默示担保。此外，这些信息均基于 **IBM** 现行的产品计划和战略，**IBM** 可能会更改这些产品计划和战略，恕不另行通知。对于因使用本演示文稿或任何其他文档而造成的或以其他方式与这二者有关的任何损害，**IBM** 概不负责。本演示文稿中包含的任何内容均不能说明 **IBM**（或其供应商或许可商）做出了任何担保或陈述，不会修改约束 **IBM** 产品和/或软件使用行为的任何协议或许可协议的条款和条件，也不会起到做出这种说明或修改的作用。

**IBM**、**IBM** 徽标、**ibm.com**、**InfoSphere**、**Infosphere Streams** 和 **Infosphere BigInsights** 是 **International Business Machines Corporation** 在美国和/或其他国家/地区的商标或注册商标。在这些信息中，以上商标及其他带有 **IBM** 商标的词语第一次出现时会标有一个商标符号（® 或 ™），这两个符号表示在这些信息出版时它们所标记的词语是 **IBM** 所拥有的美国注册商标或普通法商标。此类商标可能也是在其他国家/地区的注册商标或普通法商标。**IBM** 网站上的“版权和商标信息”中提供了最新的 **IBM** 商标列表，网址为 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)



# 信息是新一轮机遇浪潮的核心...



**数量增长**  
每一天都会产生超过 15 PB 的新信息。数据量预计每 2 年就会翻一番。

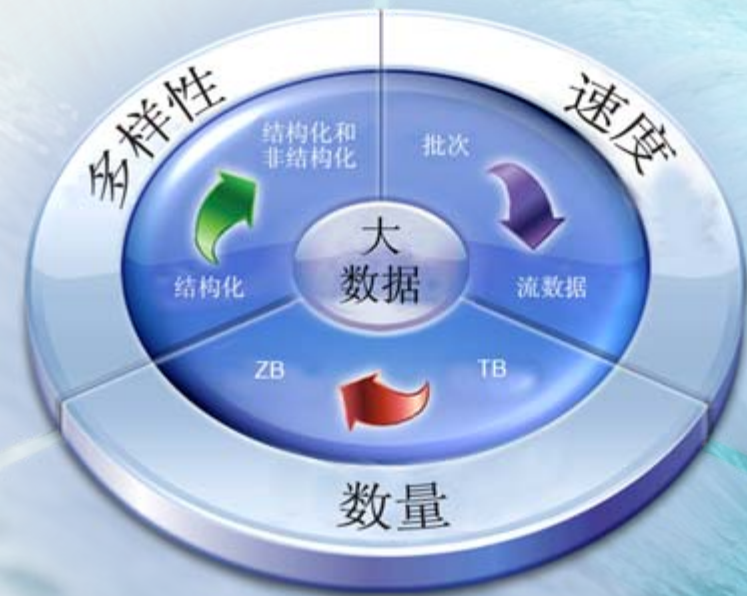
**多样性增长**  
80% 的新数据增长源自非关系数据类型和非传统数据类型，如电子邮件、文档、RFID 源、多媒体等

**需要作出“更明智的”决策**  
70% 的高管认为，未及时作出决策以及所作出的决策欠佳对其公司的业绩产生了不利影响。



## 大数据机遇

从数量庞大、多样化的高速数据中联系上下文提取真知灼见，从而做到以前所不可能做到的事情。



**多样性:** 管理多种关系和非关系数据类型及架构的复杂性

**速度:** 流数据和大批量数据移动

**数量:** 从 TB 扩展到 ZB

## IBM 大数据平台的定义...

### 1. 针对 V<sup>3</sup> 的平台 – 多样性、速度、数量

- 多样性 - “按原样”管理数据和内容
- 可处理任何速度 - 低延迟流和大批次
- 数量 – 大量静态数据和流数据

### 2. 针对 V<sup>3</sup> 的分析功能

- 以数据来源原本的格式分析它们 - 文本、视频、音频
- 分析所有数据 - 而不是其中的一部分
- 动态分析 – 自动调整和操作

### 3. 开发人员和用户可轻松使用

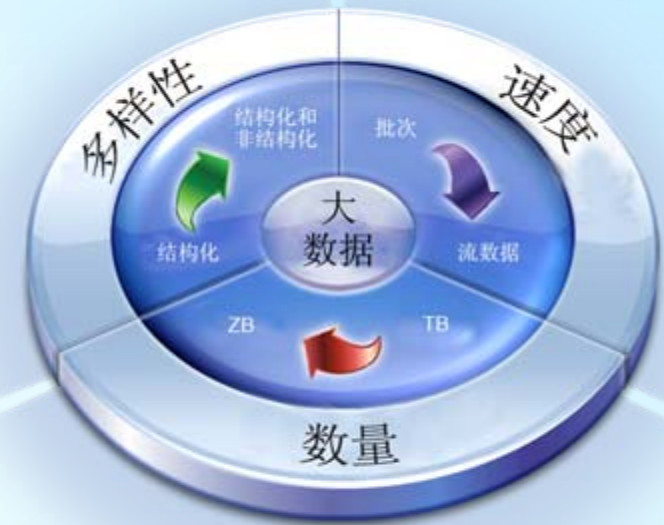
- 开发人员用户界面、通用语言和自动优化
- 最终用户用户界面和直观显示

### 4. 企业级

- 容错能力、安全性和隐私性
- 扩展成本经济合算

### 5. 丰富的集成功能

- 可集成众多种来源
- 可利用企业集成技术

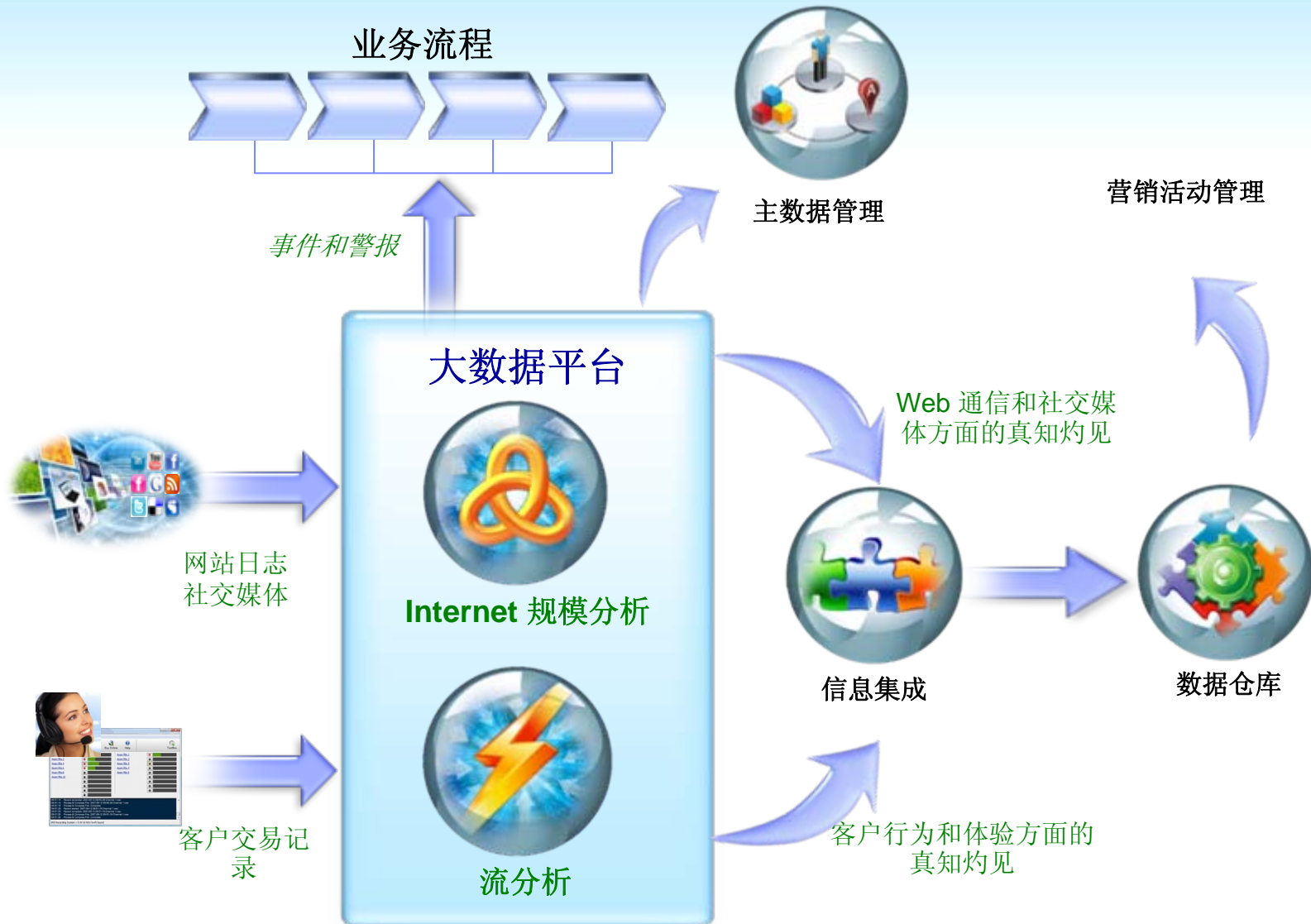


# IBM 的大数据平台愿景

将大数据引入企业

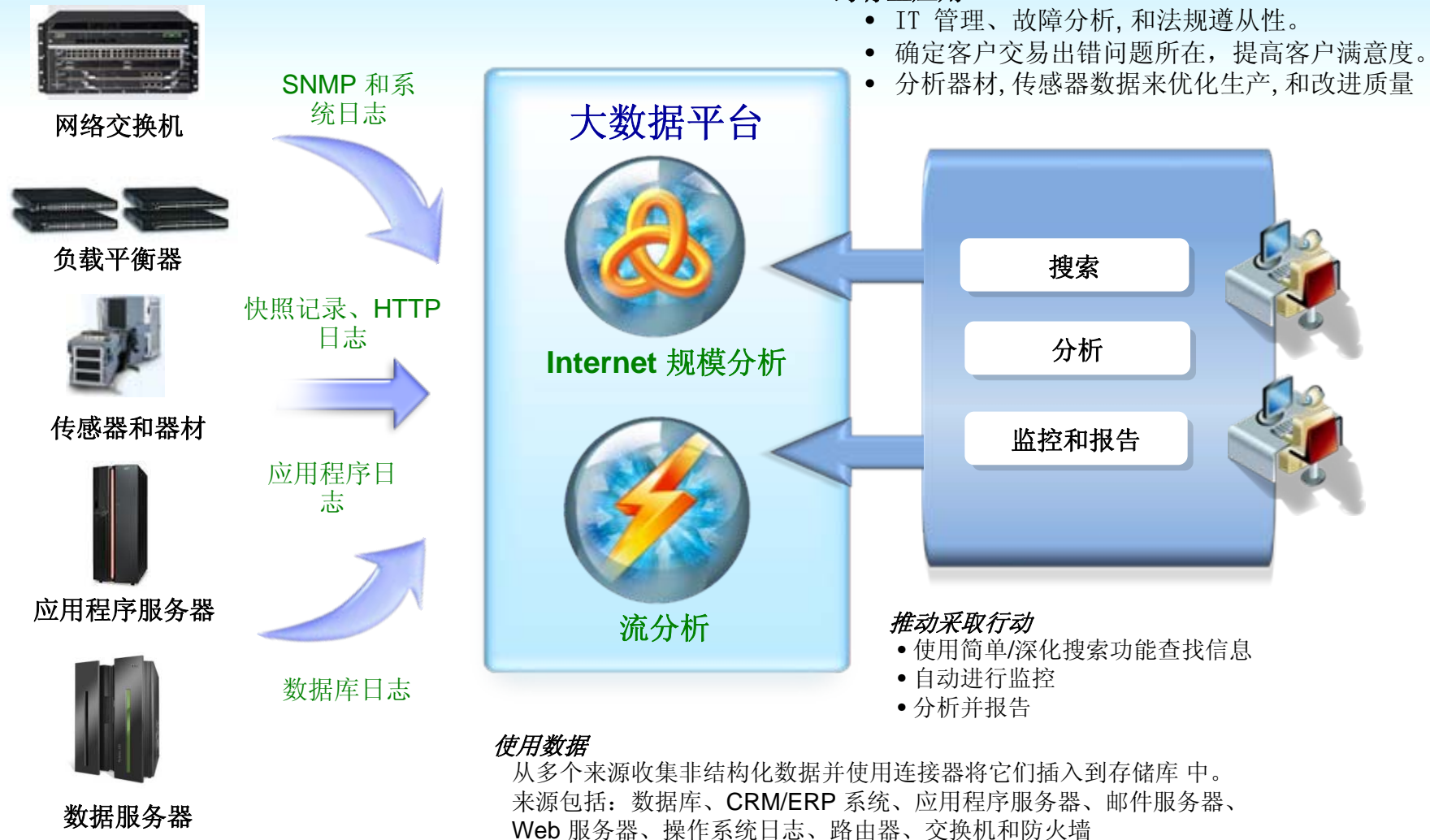


# 示例 - 客户行为和产品看法分析





## 示例 - 运营效率,客户满意度,和 生产质量



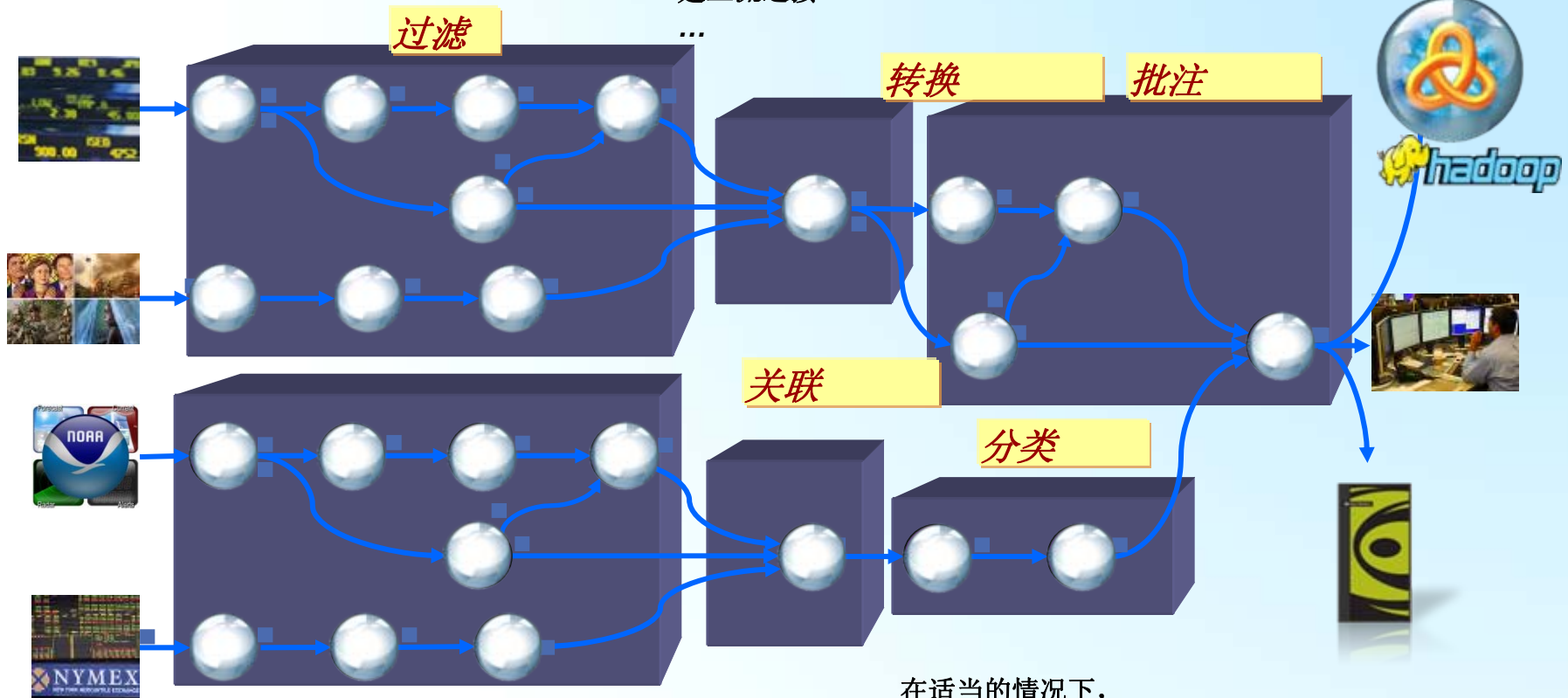
▪ **跨行业应用:**

- IT 管理、故障分析, 和法规遵从性。
- 确定客户交易出错问题所在, 提高客户满意度。
- 分析器材, 传感器数据来优化生产, 和改进质量

# 流分析 (InfoSphere Streams)

- 持续吸纳
- 持续分析

基础架构提供相应服务，用于  
跨硬件节点调度分析功能  
建立流连接



实现扩展，方法是  
将应用程序划分成组件  
跨通过流连接的硬件节点进行分配

在适当的情况下，  
可以将各元素“熔合”在一起  
以便降低通信延迟

# Internet 规模分析 (InfoSphere BigInsights)

## 支持开源体系

Hadoop、MapReduce 等

## 性能和安全性

用作企业级文件系统的 GPFS-SNC

## 编程接口

Java、Hive、PIG 和 JAQL

用于分析 UDF 的 JAQL 运行库

## workflow 协调和优先级排定

完全可恢复的工作流管理

基于策略的可配置调度程序

- 群集内分析

分布式文本和数据挖掘库 (systemT、systemML)

- 集成

双向 DB2/RDBMS 和 ETL 集成

- 安装和管理

集成式 Web 安装和管理工具



## IBM Netezza 分析：大数据遇到大智慧

- 目标定制的分析引擎
- 集成的数据库、服务器和存储



- 标准接口
- 总体拥有成本低



- 快速：比传统系统快 10-100 倍
- 简单性：管理和优化操作最少

- 可扩展性：PB 级用户数据容量
- 智能：高性能高级分析

无约束分析



## IBM Watson



**IBM Watson** 是在分析创新方面的一项重大突破，但它之所以取得成功，完全得益于它所处理的信息的质量。

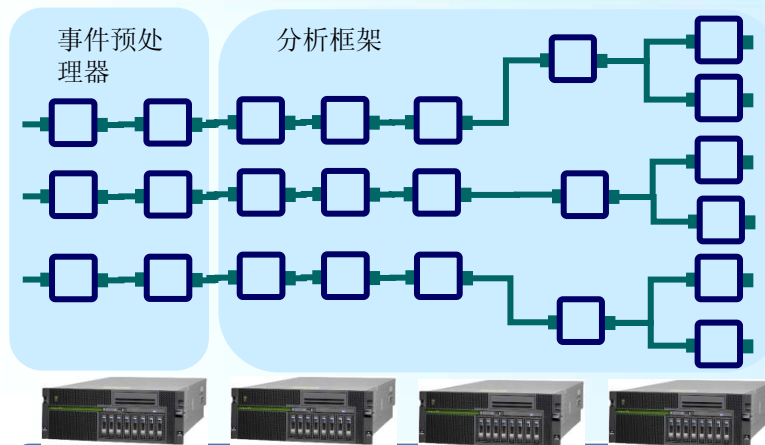


# SickKids

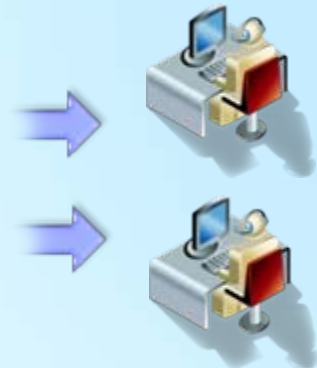
- 使用案例
  - 新生儿监护
  - 提前 24 小时预测重症监护室中的感染
- 解决方案
  - 监护的 120 名婴儿：每秒 12 万条消息，每天十亿条消息
  - 试验范围扩大，纳入了美国和中国的医院



传感器网络



基于流的分布式可互操作  
医护基础架构



解决方案  
(应用程序)

# 大数据分析正在发挥作用

## 金融服务

- 改善了风险决策
- 客户看法分析
- AML

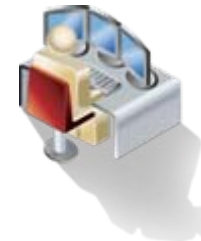


## 公用事业

- 天气对发电的影响分析
- 智能电表数据分析

## 交通运输

- 天气和交通状况对物流和燃油消耗量的影响



## IT

- 针对多个交易系统进行的交易日志分析

## 呼叫中心

- 进行语音至文本挖掘以便了解客户行为



## 电子商务

- 分析在 Internet 上的行为和购买特征
- 数字资产隐私保护

## 电信

- 根据设备、传感器和 GPS 输入进行的运营和故障分析



## 多渠道集成

- 集成式客户行为建模



धन्यवाद

印地语

多謝

繁体中文

ขอบคุณ

泰语

Спасибо

俄语

Gracias

西班牙语

Thank

英语

شكراً

阿拉伯语

You

Obrigado

巴西葡萄牙语

감사합니다

韩语

Danke

德语

Grazie

意大利语

Merci

法语

நன்றி

泰米尔语

ありがとうございました

日语

多谢

简体中文





# BIG DATA

## World Forum 2011

2011大数据世界论坛

The Business Value of BIG DATA

[BigDataWF.com](http://BigDataWF.com)