



中国国际大数据大会
China International Big Data Summit

资料汇编

(仅供内部交流，请勿外传)

大数据分析驱动商业价值

朱辉

IBM中国研发中心
信息管理与大数据产品部 总经理



1 IBM 对于大数据与分析的观点

2 议程 大数据分析的5个高复制的使用场景

大数据来自哪里？

企业自身累积的数据



社交媒体产生的数据

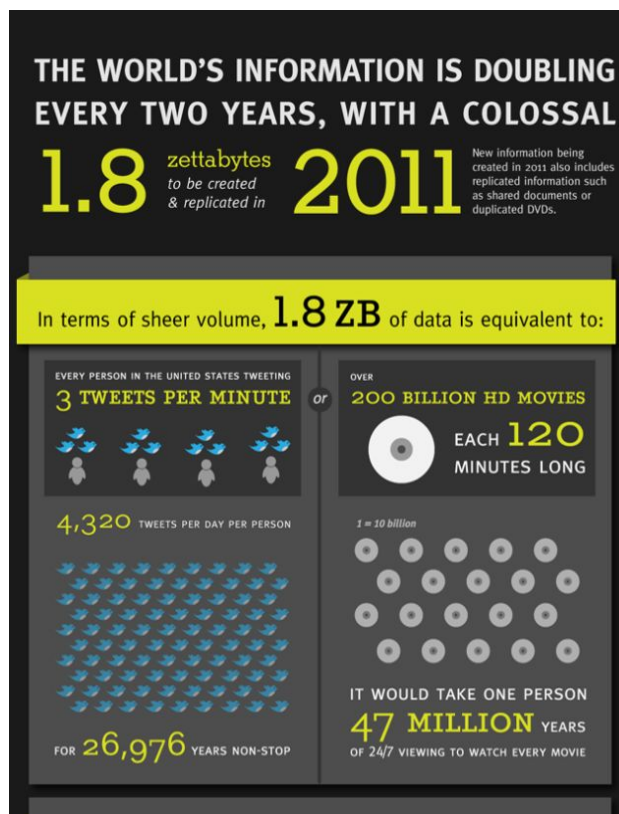
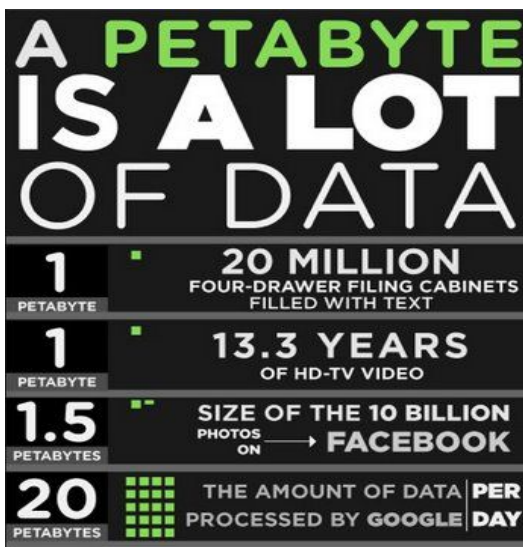


感知设备采集的数据



数据规模爆炸式地增长

人类产生的数据综合2020年将达到**40ZB**，2003年之前所有的数据，我们现在只要两天就可以产生，每天产生 5 Exabytes 数据



1 Gigabyte = 1,000,000,000
1 Terrabyte = 1,000,000,000,000
1 Petabyte = 1,000,000,000,000,000

1 Exabyte = 1,000,000,000,000,000,000
1 Zetabyte = 1,000,000,000,000,000,000,000
1 Yottabyte = 1,000,000,000,000,000,000,000,000

我们正处于大数据的黎明期.....
.....新信息的不确定性和复杂性在不断增加

数量



大规模的数据

TB 级到 PB 级的数据

Volume

多样性



多种形式的

结构化、非结构化、
文本、多媒体

Velocity

速度



动态数据

对流数据的分析支持
在几分之一秒内制定
决策。

Variety

准确性



数据的不确定性

管理本质上不精确的
数据类型的可靠性和
可预测性。

Veracity

商业和社会价值来自对大数据的分析和应用

“数据是新的石油资源” --
石油还没有被加工处理之前，一文不值；一旦经过处理和提炼，它就产生巨大的价值

“Data is the new oil.” -- Clive Humby



大数据是下一个自然资源

“我们首次拥有一个基于关键资源（信息）的经济体制，不仅是可再生的而且是自发生成的。耗尽它没关系，但不要淹没在其中。”

— John Naisbitt

Forbes
.com

*“Big Data has arrived at Seton Health Care Family, fortunately **accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year...**”*

The New York Times

*“At the **World Economic Forum** last month in Davos, Switzerland, Big Data was a marquee topic. A report by the forum, “Big Data, Big Impact,” **declared data a new class of economic asset, like currency or gold.**”*

THE WALL STREET JOURNAL.

*“**Companies are being inundated with data** from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers.”*

数据分析的演变历程

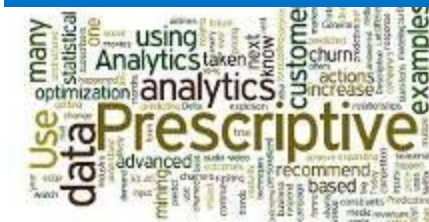
描述性分析



预测性分析



指令性分析



认知性分析



描述性

- 基于历史数据的“滞后的事实”分析
- 为企业或机构业务措施提供清晰的评估
- 适用与所有业务部门进行事实调查，及成功与失败的可视化

预测性

- 利用数据挖掘，统计和机器学习算法等来分析当前和历史数据来预测未来事件和业务成果。
- 从历史和交易数据发现规律，从而优化商业措施

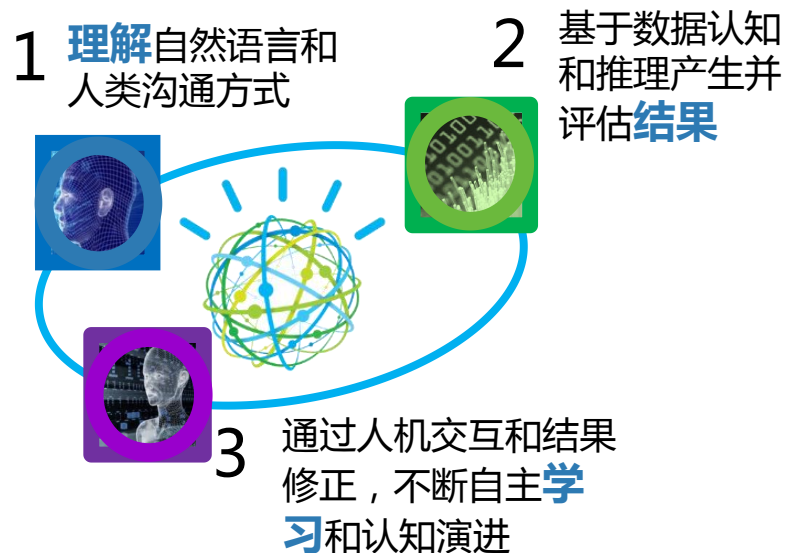
指令性

- 综合大数据，数学和计算科学，及业务规则，提出决策方案。
- 注意到未来的机遇或减轻未来风险，并显示每个决策选项的含义

认知性

- 关于感知的心理过程，记忆，判断，学习和推理。
- 用于了解某些类型的业务相关功能的不同分析策略范围
- 自然语言处理

认知计算为大数据分析提供新的能力



Watson Foundations

在数据分析领域的能力

帮助发现全新的洞察

- 预测和内容分析，揭示隐含尚未发现的规律
- 跨所有数据的交互式探索

高效及时的运作

- 实时分析流经一个组织的数据
- 比开源快数倍的企业级 Hadoop
- 与意念分析速度可比拟的内存计算能力

建立信任，制定有信心的决策

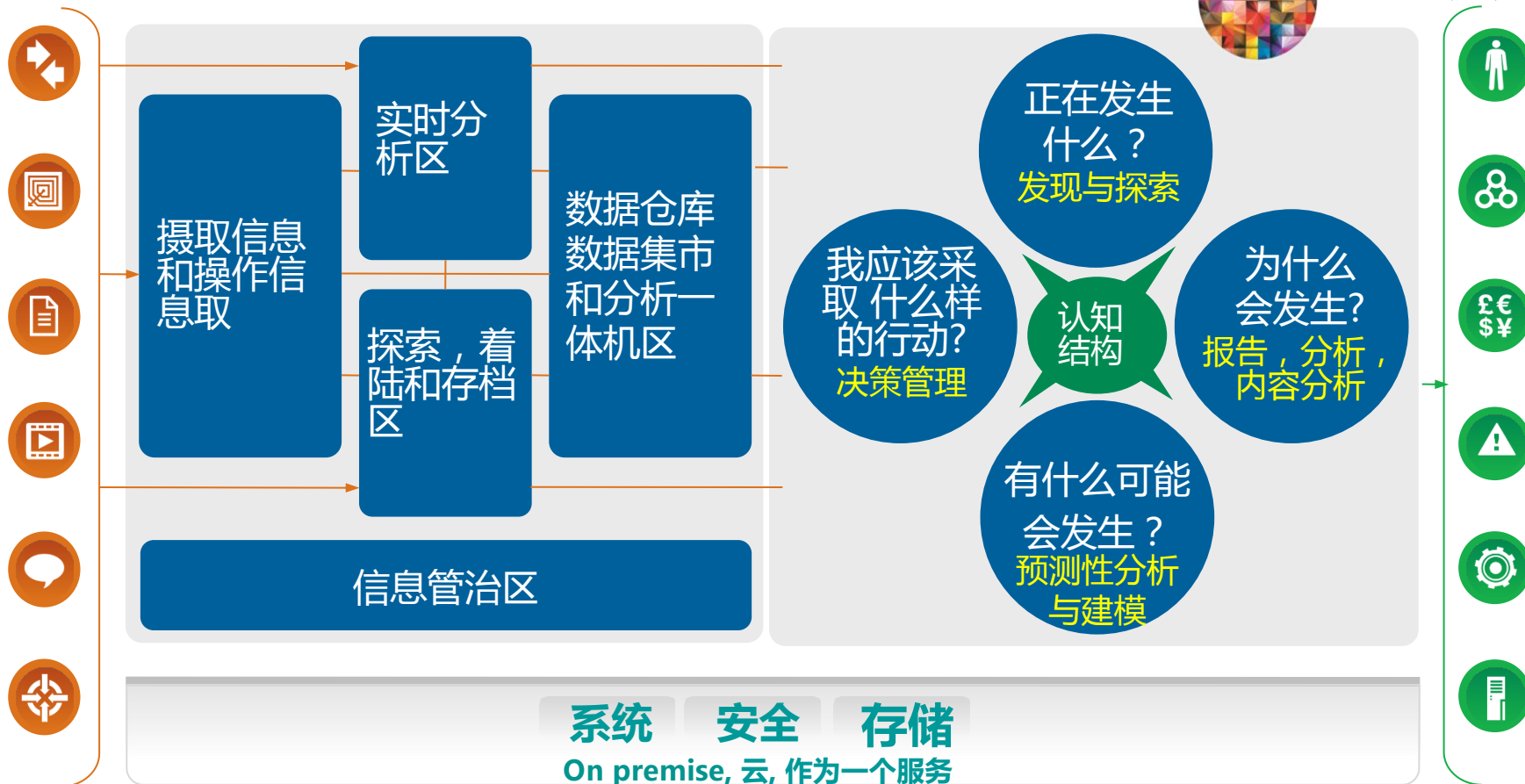
- 包括Hadoop在内的整个数据生命周期的管治
- 安全和隐私合规
- 透明度和上下文的决策过程

基于Watson Foundations的认知性分析

所有数据

IBM Watson Foundations

新型/增强应用

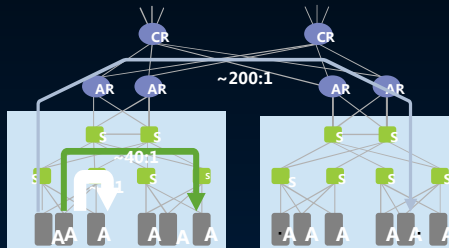


IBM 大数据 & 分析 基础设施

“云计算”是大数据分析的基石



支撑对海量数据的分析洞察，对存储设备的容量和吞吐量、计算能力等提出更高要求。随着网络带宽的快速提升，通过软件定义网络，将跨地域的数据中心虚拟为扁平的大“云”，为大数据分析提供强有力的基础设施保障。



360度全方位的安全管控和合规保证，是在云端进行大数据分析的重要前提和必须面对的挑战

挑战一	挑战二	挑战三
<p>繁杂的、离散的安全控制点</p>	<p>不断演进的黑客技术和病毒威胁</p>	<p>海量的连接设备、产生更多的攻击入口</p>

基于大数据和分析产生新的业务价值和行业解决方案



IBM 提供了全面、整合化的大数据和分析方法

IBM 协助企业实现:

- 收集和组合各种相关信息
- 借助智能可视化平台来发现和探索信息
- 通过分析、预测和自动化获取更准确的答案
- 采取措施并自动化流程
- 优化分析性能和 IT 成本
- 降低基础架构复杂性和成本
- 管理、治理和保护信息



议程

1

IBM 对于大数据与分析的观点

2

大数据分析的5个高复制的使用场景

5个高复制的使用场景为各行各业带来新的商业价值





探索和挖掘大数据，以寻找业务关注的相关内容，从而实现更好的决策制定

需求

- 对新数据源进行探索，以便在无假设的前提下发现现有企业数据和内容以外的潜在价值
- 提供假设，挖掘与企业的某项当务之急相关的内容
- 在上下文中关联并不相似的信息源，并且评估非结构化内容的业务价值
- 利用可视化、算法和处理，发现关注的模式
- 在探索时避免泄漏敏感信息

1. 大数据探索

- 客户服务知识门户
- 保险灾难建模
- 汽车功能与定价优化
- 化学品和石油基于条件的维护
- 疾病诊断
- 生命科学药品有效性
- 赛事预测
- ...

案例：大数据和分析在体育



IBM 大数据分析 网球应用中的精华

- IBM SlamTracker技术

通过IBM SPSS预测分析技术提供的预测分析，使得球员、球迷、教练可以更加透彻地理解比赛

案例：大数据和分析在医疗业



“最快可提早 24 小时帮助检测出对生命有威胁的情况”

Ontario Institute of Technology 大学 (UOIT) 更早检测出新生儿患者的症状

利用的功能：流计算

- 使用新生婴儿的生理数据执行实时分析
- 持续关联来自医疗监视器的数据，以检测微小的变化，并更早提醒医院员工
- 早期的警告使护理人员能够主动处理并发症

成果

- 最快可提早 24 小时帮助检测出对生命有威胁的情况
- 降低发病率并改善患者护理水平





全面了解客户，从而优化每一次客户交互

需求

- 尝试整合内部信息和外部信息，从而建立互联互通的客户视图
- 关联和挖掘企业数据、主数据、呼叫记录、企业内容和新信息源，从而获得可采取行动的洞察
- 分析社交媒体和外部数据源，以发现客户对于您的产品和企业的观点
- 通过优化每一次客户交互来增添价值

2. 增强的 360 度全方位客户视图

行业示例

- 智能仪表分析
- 电信运营商客户关系管理
- 零售营销优化
- 旅行和运输客户分析和忠诚度营销
- 金融服务的后续最佳举措和客户保留
- 汽车质保索赔
- ...

案例：大数据和分析在零售业

Automercados Plaza 的数据分析提供客户行为洞察，收入增加 30%

需求

- 轻松访问每个店铺的运营数据
- 及时、全面的运营视图，以改进业务决策

获益

- 年度收入增加了 30%，年度利润增加了 7 百万美元
- 编制销售税数据所需的时间减少了 98%
- 防止易腐货物的损失，这在所有产品中约占 35%



案例：大数据和分析在电信运营商

Ufone 降低用户流失率并保持让用户满意，帮助确保营销活动高度有效和及时

需求

- 确保其营销活动在客户退出网络之前瞄准正确的客户
- 通过营销活动提供适合客户的服务和计划，保持其更高使用率的客户的满意度

获益

- 预计预测分析将营销活动的响应率从大约 25% 提高到至少 50%
- CDR 的分析可以在30 秒内完成, 而不再需要至少一天的时间
- 预计流失率减少大约 15-20%





对机器数据应用分析，从而提高运营效率

需求

- 在保证亚秒级延迟的前提下分析海量机器数据，以便在发生关注的事件时及时识别
- 应用预测模型和规则，在发生潜在的反常现象或出现机遇时及时识别
- 整合运营和企业数据，从而实时了解服务水平
- 监控系统，以主动提高运营效率，避免服务降级或中断

3. 行业示例及运营分析

- 智慧交通监控
- 化学品和石油基于条件的维护
- 能源和公用事业基于条件的维护
- 电信市场活动管理
- 旅行和运输实时预测维护
- ...

案例：大数据和分析在电力行业

太平洋西北地区智能电网示范项目

利用的功能

流计算

数据仓库设备

成果

- 证明从 100 到 50 万个家庭的可扩展性，同时保留 10 年的历史数据
- 在 5 个州有 6 万户电表客户
- 可容纳价格波动、能耗档案、风险、欺诈检测、电网状况等专用分析

案例：大数据和分析为智慧城市

都柏林城市中心；稳健和高
效的全市交通信息系统，增
强对事故的快速行动

需求

- 符合预算效益的解决方案，改进交通信息系统
- 带来事件检测、推断交通条件（道路速度）和公交车到站预测的准确性。
- 挑战是要正确地分析 GPS 数据，通常是高数据吞吐量，并且难以捕获

获益

- 每天监视 150 条路线上的 600 辆公交车
- 使用 InfoSphere Streams 分析每秒 50 个公交车位置更新
- 收集、处理和可视化所有公共交通工具的位置数据





利用技术进步从现有的数据仓库投资获得更多的价值，同时降低成本

需求

- 将流式传输和非结构化的数据源添加到现有的数据仓库投资
- 优化数据仓库存储，并提供可查询的归档
- 合理化数据仓库，实现更好的简单性及更低的成本
- 提供更好的查询性能，支持复杂的分析应用程序
- 为业务经营提供洞察，实现实时决策
- 利用预测分析和商业智能实现性能和扩展

4. 数据仓库能力扩展

- 预处理中心
- 可查询的归档
- 探索性分析
- 运营报告
- 实时评分
- 细分和建模

案例：大数据和分析在金融业



Fiserv 在五年内节省 8 百万美元，并帮助银行提高业务成果

利用的功能：*数据仓库*

需求

- 中小型银行和信用社寻求吸引、留住和发展可盈利的客户关系，同时利用分析功能与新的大型银行竞争

获益

- 将数十亿美元的交易转化为可操作的洞察，帮助这些银行提供目标性更强的产品，并使其营销资金发挥最大作用
- 预期有针对性的市场营销活动的响应率增加 100% 或更高
- 预期在五年内 IT 节省 8 百万美元
- 提供快速提供客户关键高质量可靠的分析洞察
- 减少客户流失及推广新产品使用

fiserv.

案例：大数据和分析在汽车制造业

某汽车制造商 搭建全球化数据仓库

需求

- 在全球整合现有 DW 项目
- 提供实时的运营报告
- 跨所有数据源获得新洞察

获益

- 单一基础架构，以整合结构化、半结构化和非结构化数据
- 经过证明的企业级功能，可以实现快速部署并且更易于管理

谢谢！

