

IBM Big Data Platform



王云
IBM院士
中国研究院首席技术官

信息是新一轮机遇中的核心...

企业需要更深入的 远见卓识

未来的10
年里，数
据和内容
将增长

44倍

2020
35 zettabytes

2009
800,000 petabytes

Volume
Variety
Velocity

80%

的数据是非结构化的



1 in 3

商界领袖经常需要以不信任的
或不存在信息为基础进行决策

1 in 2

商界领袖无法获取做好本职
工作所需要的信息

83%

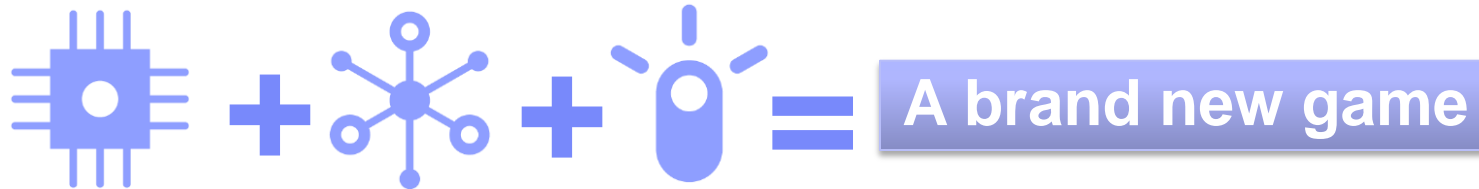
CIO们提到的“商业智能和分析”
作为有远见计划的一部分来提高
竞争力

60%

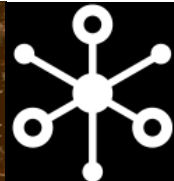
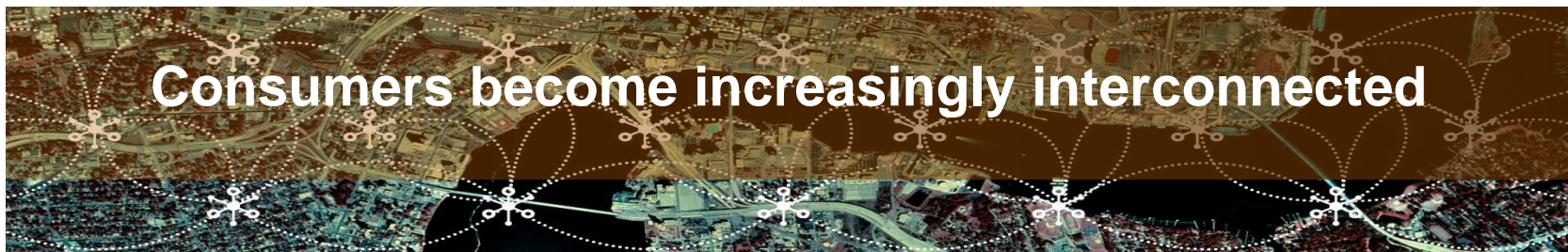
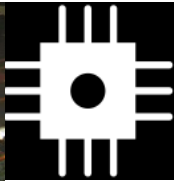
CEO们需要更迅速捕获和了解信息，
以便迅速作出业务决策来做取得更
好的工作成绩

DTCC2012

群体和个体之间交互沟通的方式发生了
翻天覆地的变换由此带来的信息量的巨增

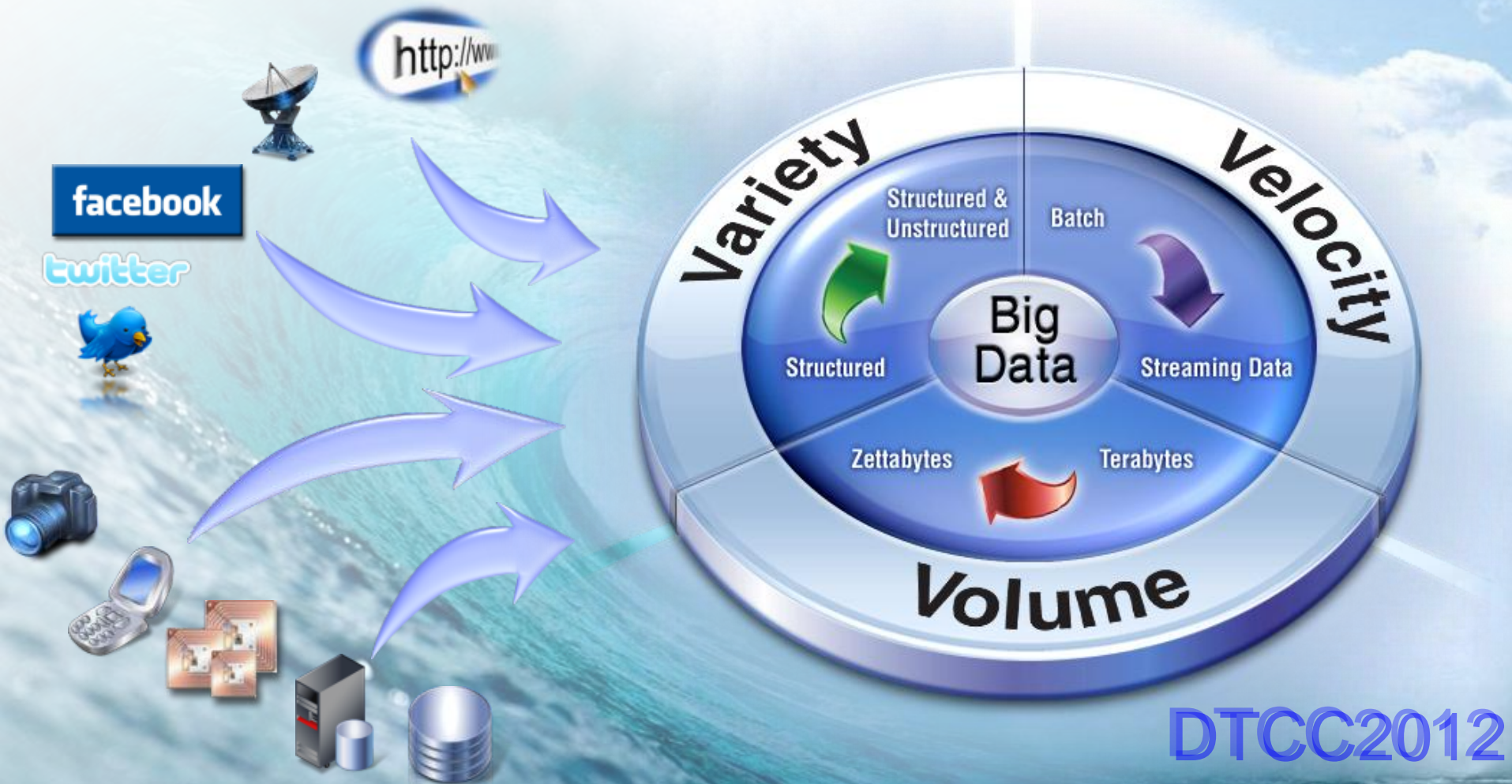


new
rules:
new
game



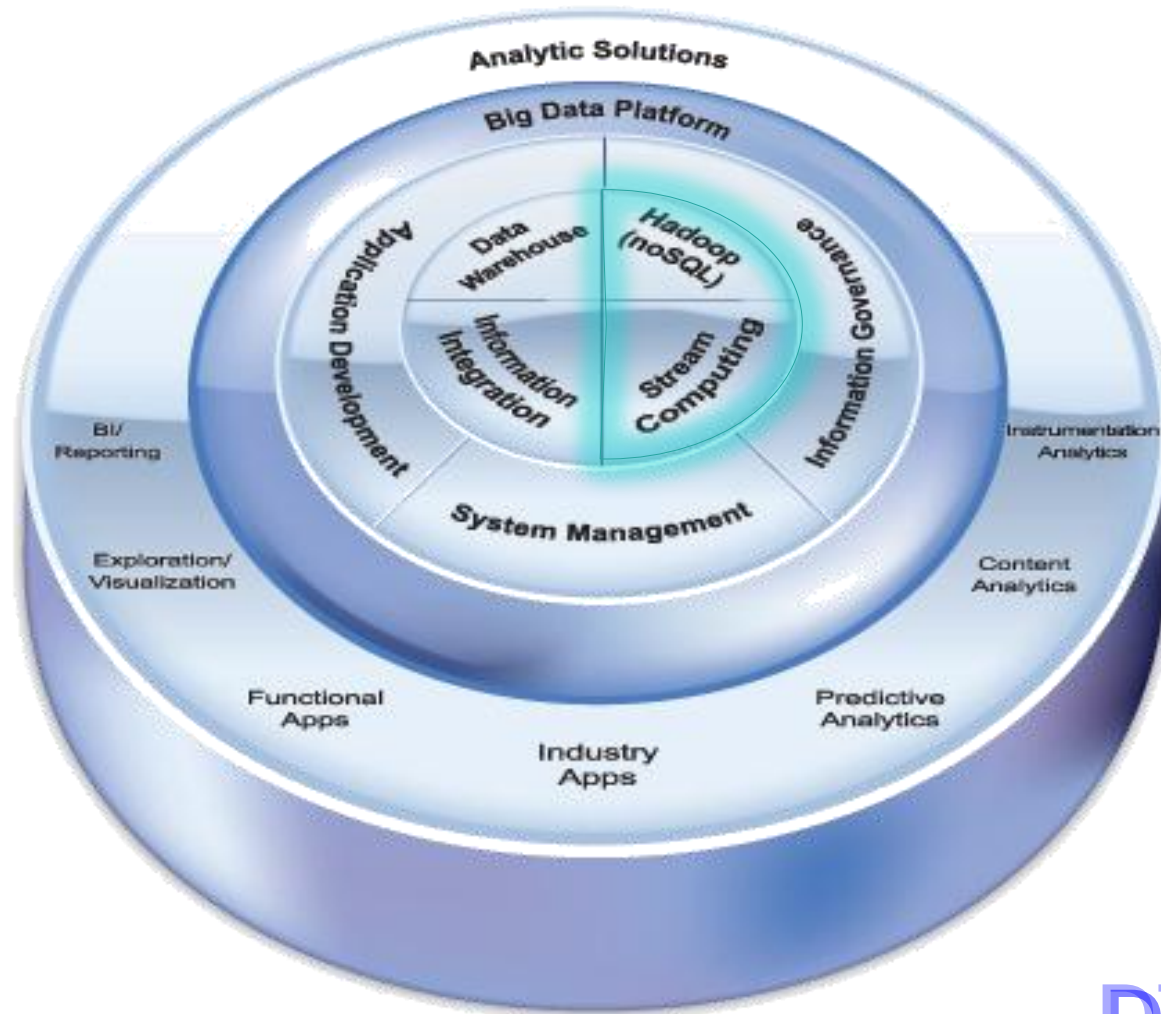
Big Data 的机遇

从巨大、以无与伦比的速度增长和多样化的数据中
提取远见卓识，而这些是以前无法做到的



DTCC2012

The IBM Big Data Platform



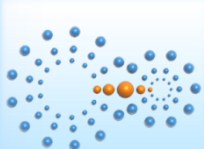
DTCC2012

Big Data 平台能够做什么事情？



分析多样化的信息

基于大范围混合信息的上下文分析，而这种分析此前根本不能执行



分析流动中的信息

流数据分析
针对数据爆炸和ad-hoc模式的分析



分析极端大量的数据

成本最优化地处理和分析PB量级的信息
管理、分析大容量的结构化和关系型数据



发掘和探索

Ad-hoc模式的分析、数据发现和验证



计划和管控

加强数据结构的完整性和控制，以确保可重复查询的一致性

DTCC2012

Big Data 丰富了整个信息管理生态系统



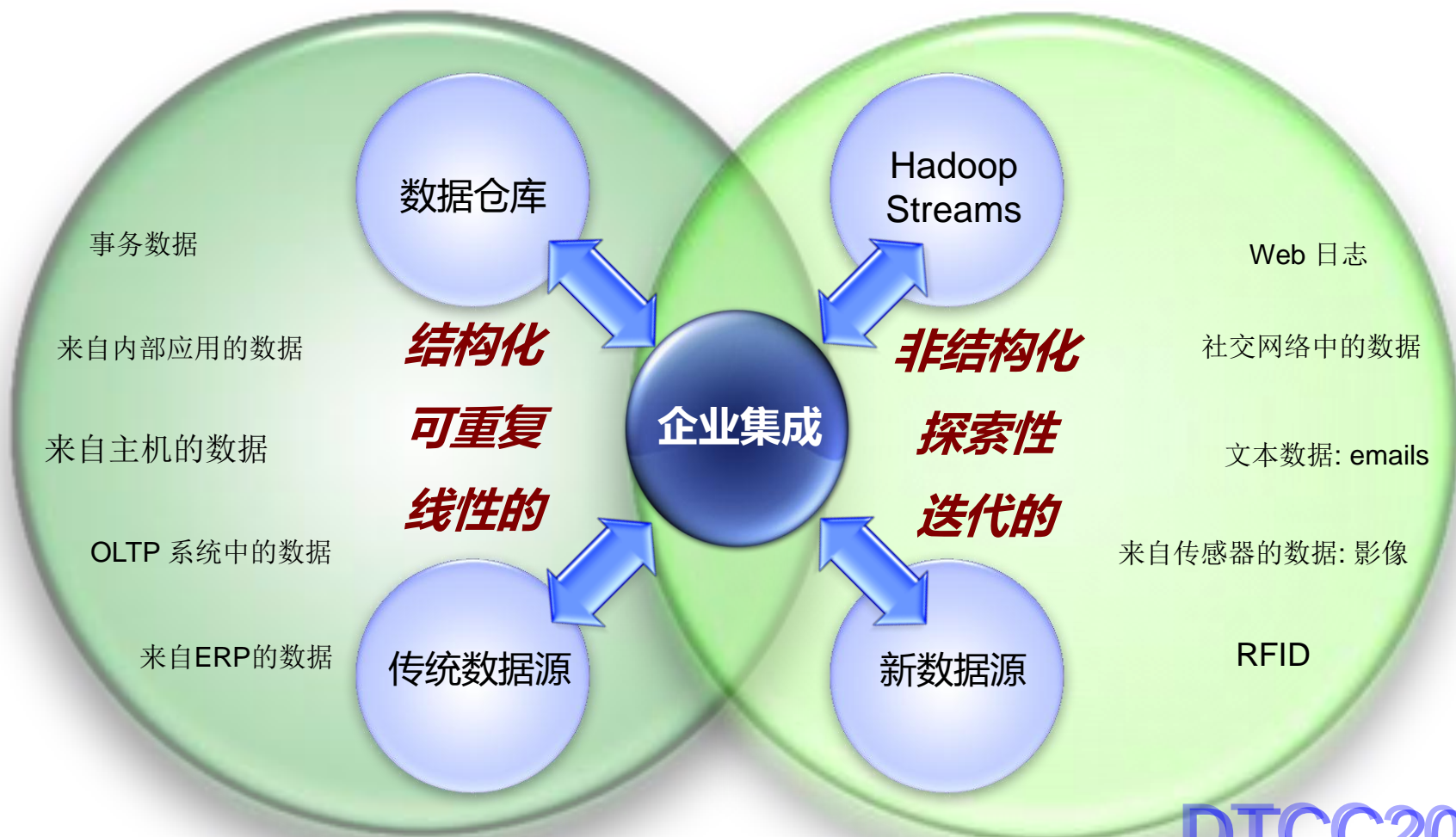
互为补充的分析能力

传统的做法

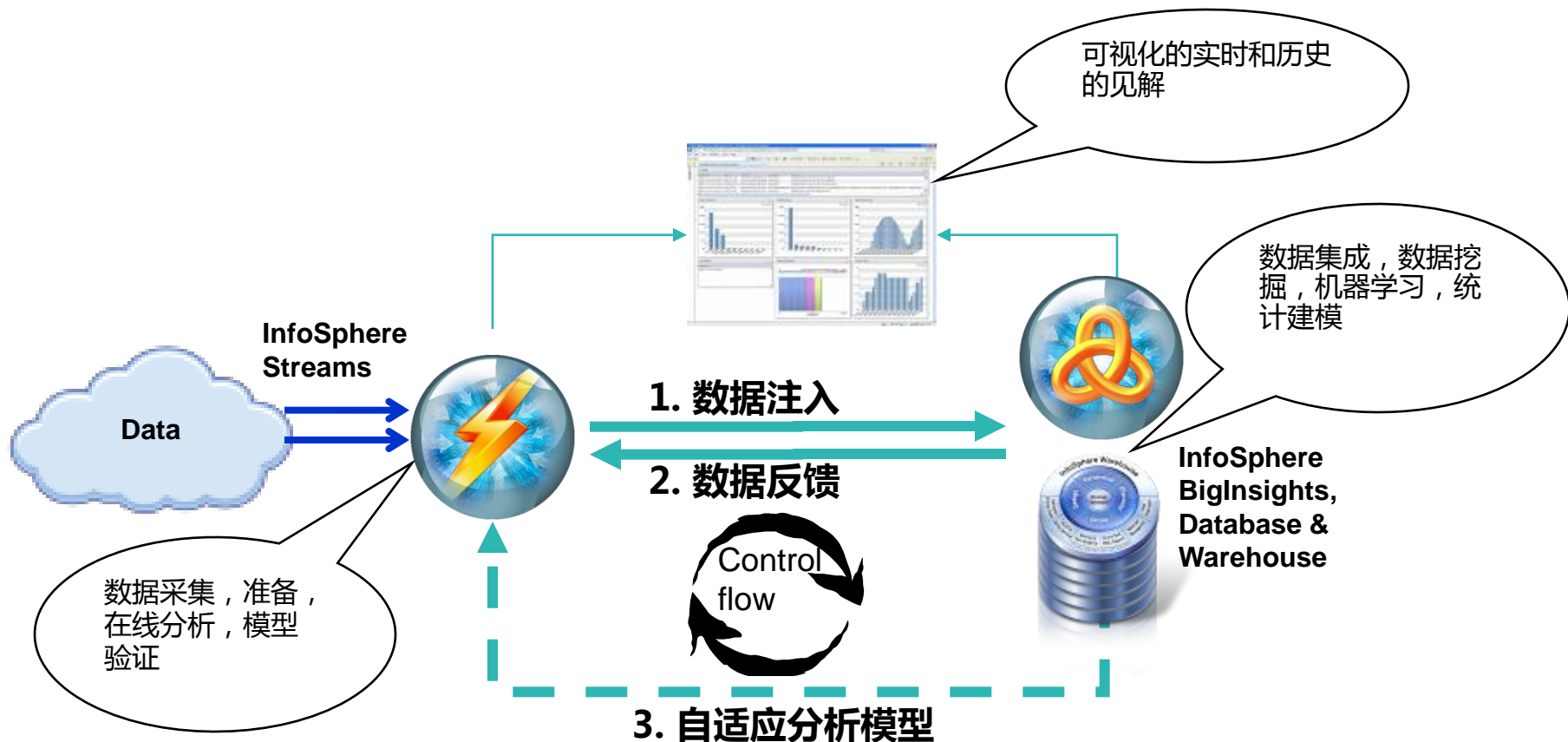
结构化, 可分析, 逻辑性

新的路径

创造性, 整体思维, 知觉



Streams 和 BigInsights – 沉淀的数据和流动数据的无缝集成



InfoSphere BigInsights

Platform for **volume, variety, velocity** -- V^3

- 增强的Hadoop方案

Analytics for V^3

- 文本分析的工具箱
- 机器学习与预测分析

Usability

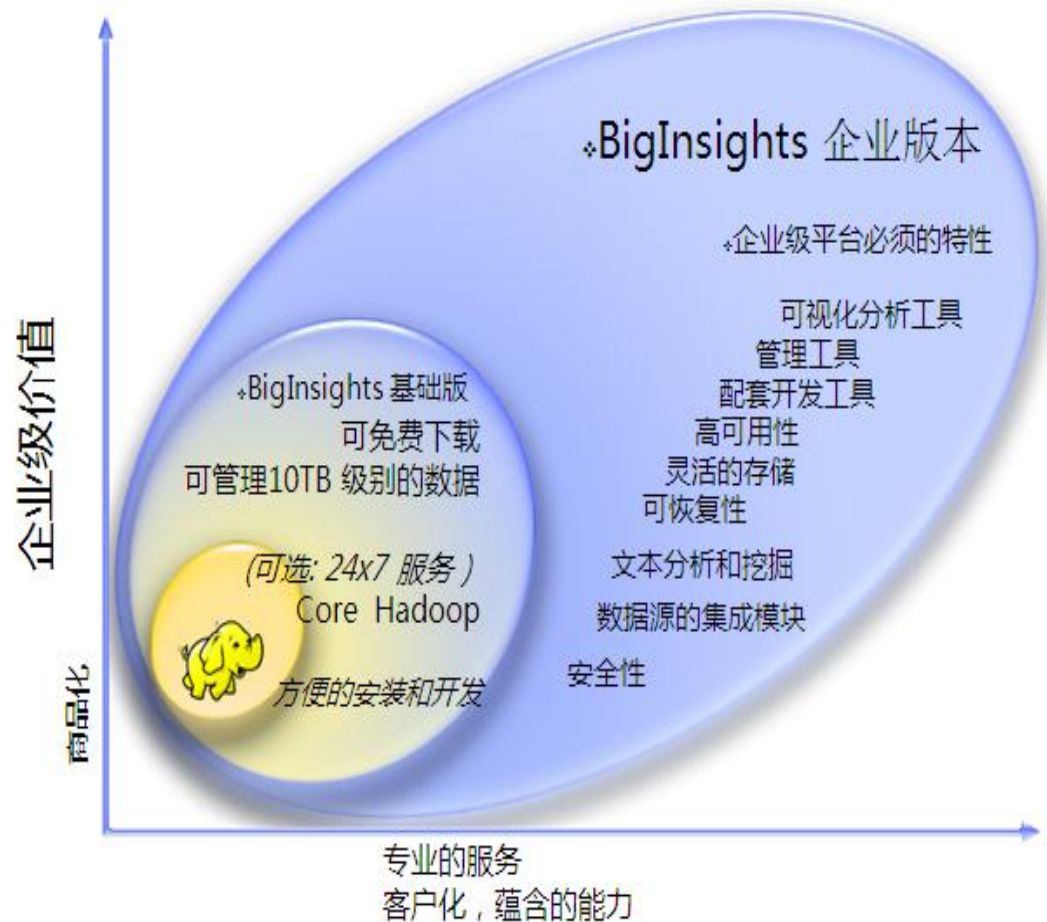
- Web console
- 集成安装
- 可视化工具
- 大数据分析的apps, 如apps Store

Enterprise Class

- 企业存储, 安全, 集群管理

Integration

- 与DB2, Netezza, JDBC数据的互联互通



流数据分析

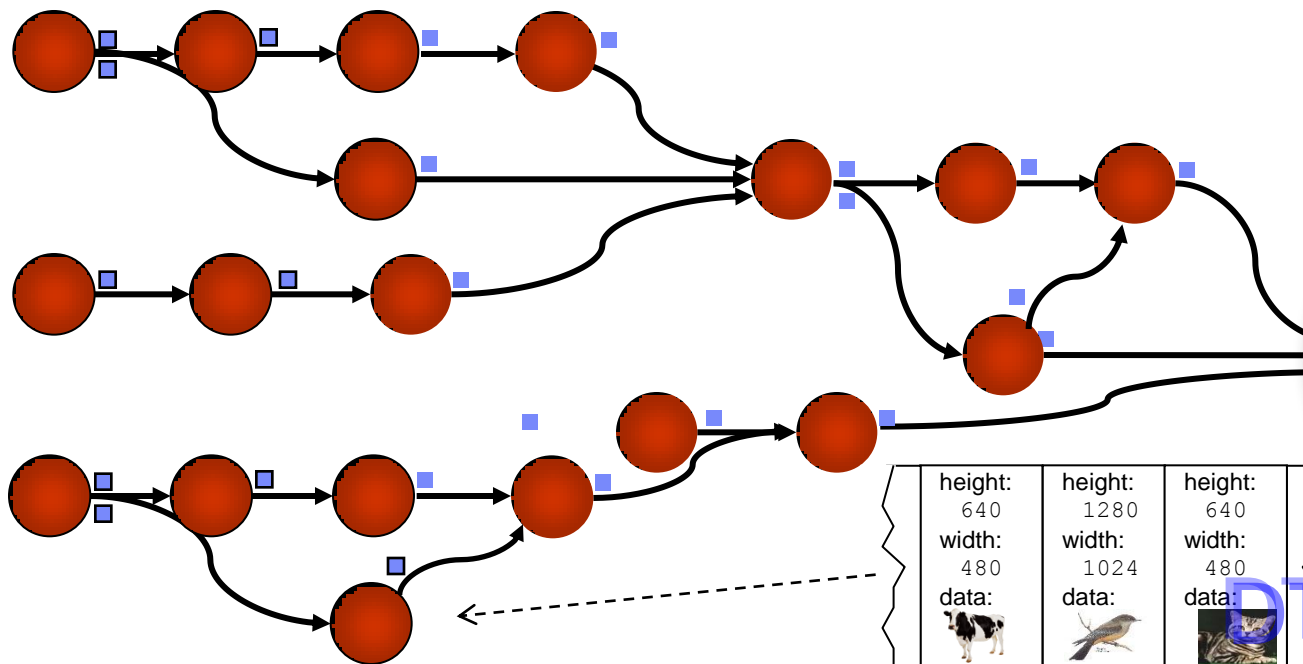
Algorithmic
Trading *Forum*

分析和关联 **500万** 市场消息/秒，
以 **30微秒** 的平均延迟执行算法期权
交易。

ipdr

以 **500K/秒** 的速度，每天分析
6B+ IPDRs，每年分析 4 PBs.
sustaining 1GBps.

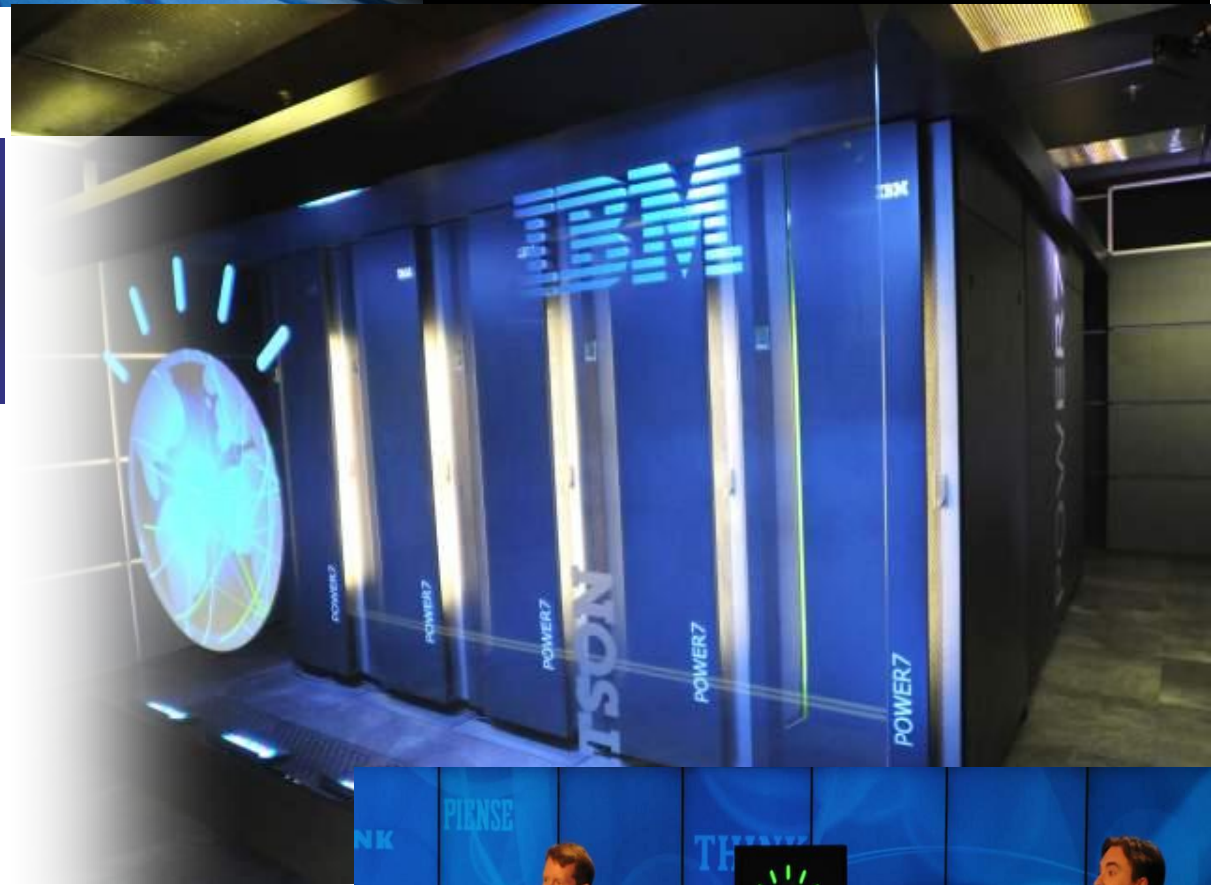
Consider: 数据不用存储, 不用遵循信息生命周期管理, 极大量的节省存储等资源消耗



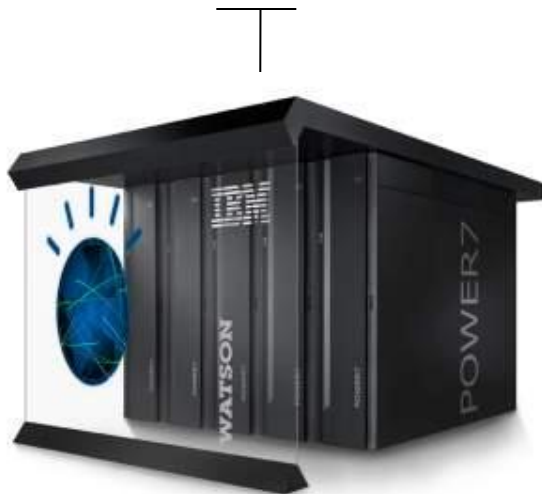
DTCC2012

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，完全得益于它处理的信息的速度和质量。



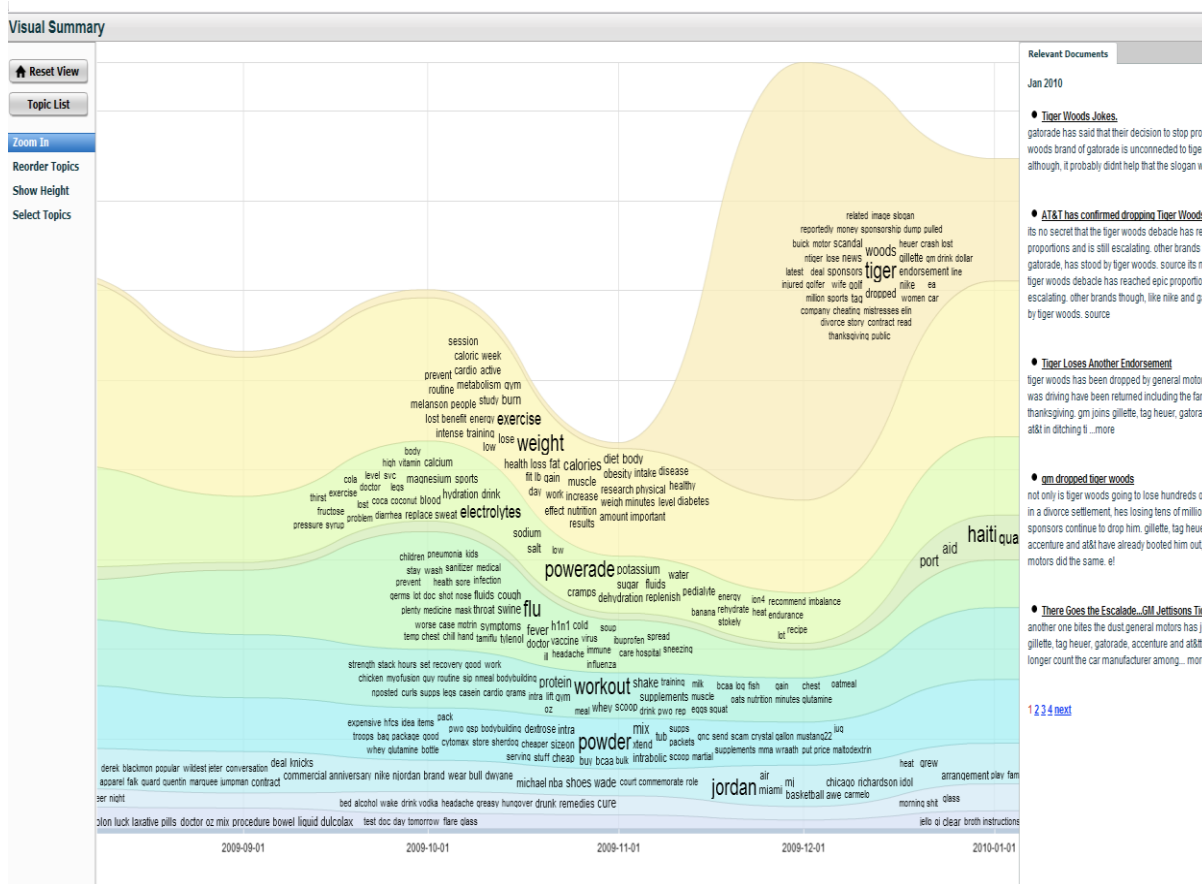
InfoSphere Streams
Massively parallel analysis



InfoSphere BigInsights
“Big Data” analysis (Hadoop)



基于BigData 平台的预测分析



tokens

documents

0	3	0	0	0	0	1	0	0	0
1	0	0	0	0	7	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	2	0	0	0	0	0	0	7
0	0	0	5	0	0	0	8	0	0
2	0	0	0	0	0	6	0	0	0
0	0	0	0	9	0	0	0	4	0
0	1	0	7	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	4
0	0	7	0	0	0	0	5	0	0



topics

words

documents

V

≈

W

x

topics

H

Learning the parts of objects by non-negative matrix factorization

Daniel D. Lee* & H. Sebastian Seung*†

* Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

† Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

```

while (i < max iteration) {
  H = H * (WT V / WT W H);
  W = W * (V HT / W H HT);
  i = i + 1;
}

```

DTCC2012

各个行业BigData 应用的适用范围和场景

Smarter Healthcare



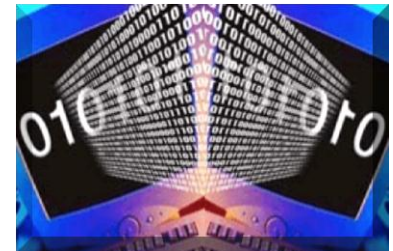
Multi-channel



Finance



Log Analysis



Homeland Security



Traffic Control



Telecom



Search Quality



Manufacturing



Trading Analytics



Fraud and Risk



Retail: Churn, NBO



DTC2012

IBM Big Data Solutions

客户和合作伙伴方案



Big Data 加速器

文本 统计数据 财经 地理信息 音频信息

影像/视频 挖掘 时间序列 数学信息

连接器

应用

蓝图

Big Data 企业引擎

InfoSphere Streams

InfoSphere BigInsights

提升和优化生产效率

工作负载管理和优化

配置

工作流

工作时间表

作业跟踪

数据摄入

管理

管理工具

配置管理器

事件监控

身份和访问管理

数据保护

规则 / 业务流程管理

iLog & Lombardi

数据仓库

InfoSphere Warehouse

数据仓库一体机

IBM & non-IBM

主数据管理

InfoSphere MDM

数据库

DB2 & non-IBM

内容分析

ECM

业务分析

Cognos & SPSS

营销

Unica

数据增长管理

InfoSphere Optim

INTEGRATION

Information Server

DTCC 2012

IBM big data • IBM big data • IBM big data

THINK

BIG

BIG

IBM big data • IBM big data • IBM big data

DTCC2012

IBM big data • IBM big data

IBM big data • IBM big data