



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)

大数据产业进展与发展策略思考

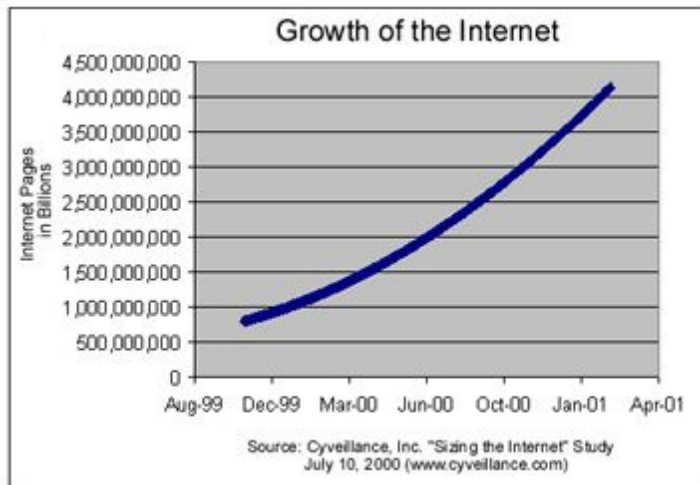
刘多

工业和信息化部电信研究院

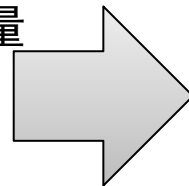
2014年5月



大数据的起源与断代



互联网内容的暴增是第一推动力量



大数据起源

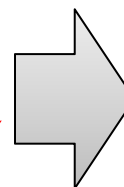
2000年前后，基于海量数据分析的搜索引擎逐步发展

大数据时代

2006后，电商、广告、SNS应用，2011大数据提出

1999-2000网页数量从5亿增长到40亿，每天新增700万！PB级非结构化数据

突变



分布式文件系统
分布式并行计算
分布式数据库

数据耦合

1950，计算机诞生，数据与应用捆绑，都存储在文件中

数据库

1960，数据与应用分离，事务处理数据库技术发展

数据分析

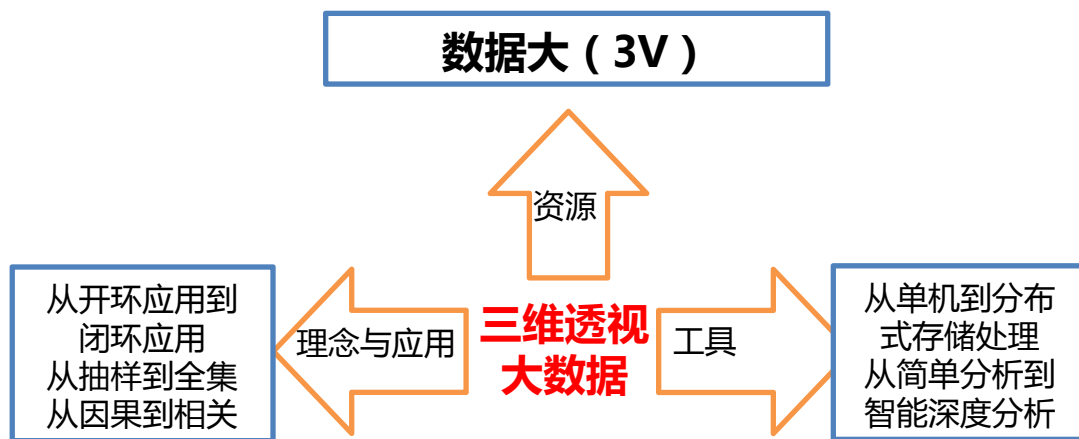
1990，提出数据挖掘和商业智能

今天：随着应用领域扩展，技术也在不断演进，为更广泛应用提供条件

大数据不仅是“大”，三位一体认识大数据

各方对大数据给出不同定义：维基百科、Gartner、IDC、NIST、IBM、麦肯锡...

我们的认识：大数据是具有体量大、结构多样、时效强等特征的数据。处理大数据需采用新型计算架构和智能算法等新技术。大数据的应用强调以新的理念应用于辅助决策、发现新的知识，更强调在线业务流程优化。



大数据不能只强调“大”，用从资源、技术、应用和理念多个角度全面认识。

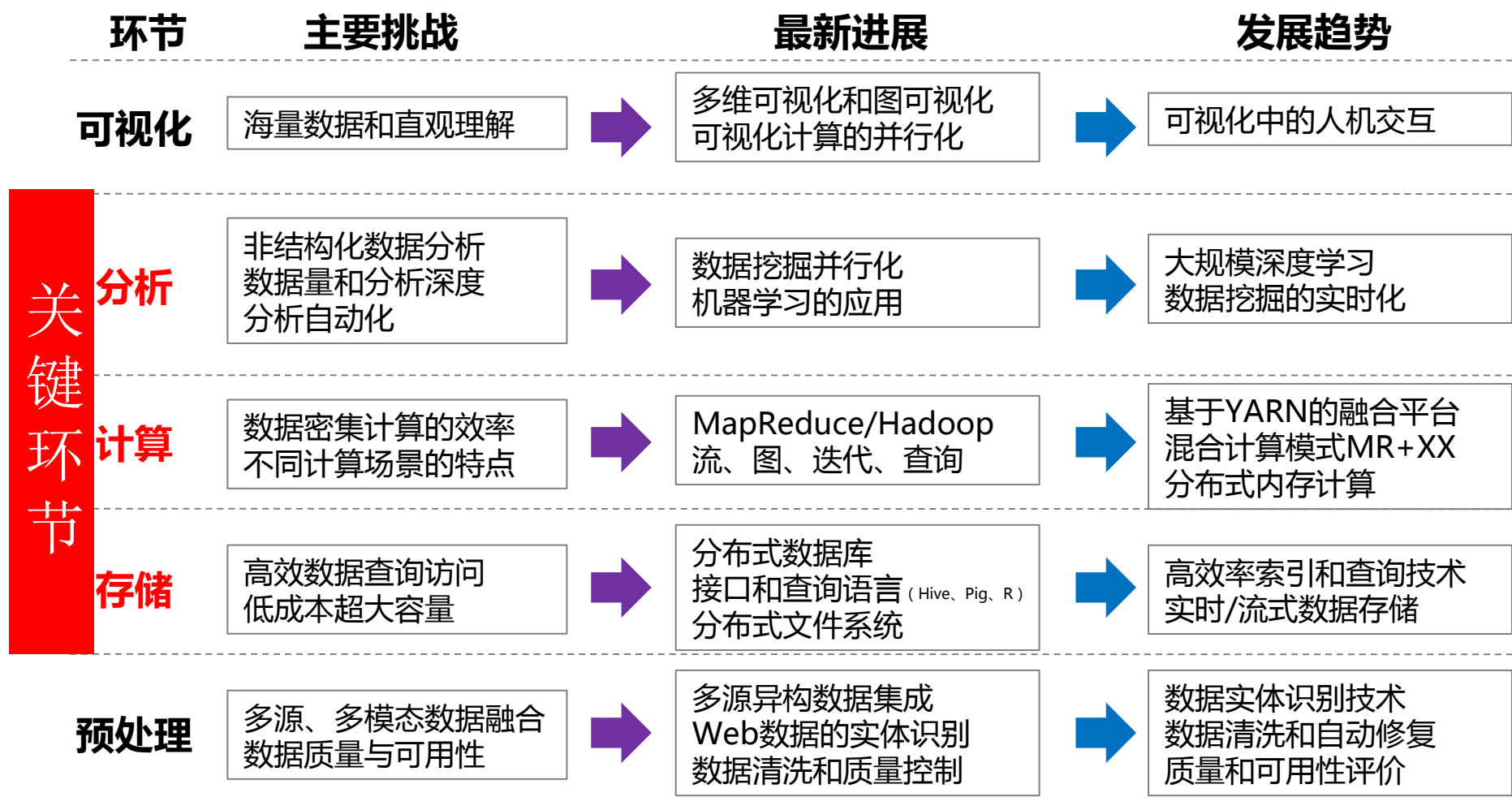
大数据更重要的意义是启发全社会重新审视**数据的价值**。

提纲



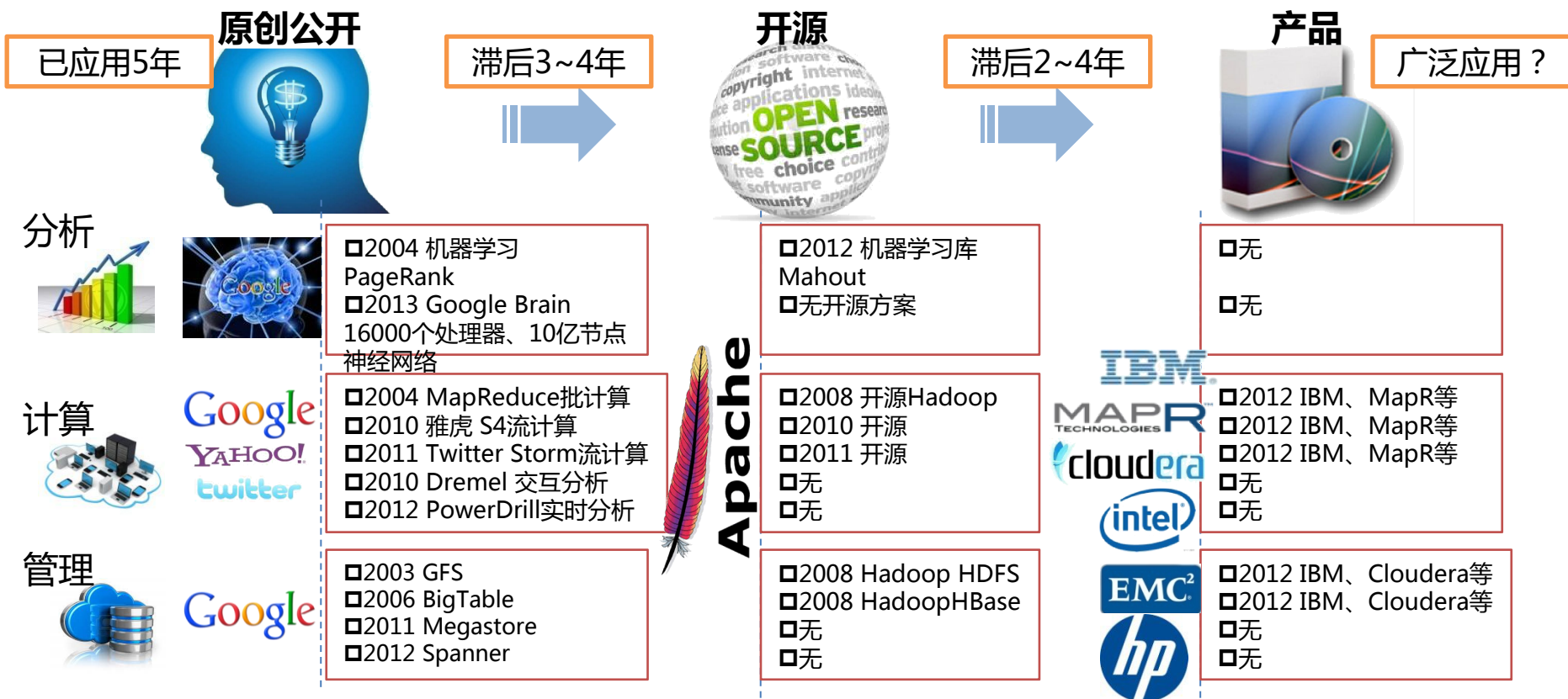
1. 大数据发展脉络与概念探讨
2. 大数据技术、产业和应用进展
3. 大数据发展政策及几点思考

大数据技术前沿不断推进，创新非常活跃



参考资料：CCF大数据白皮书，2013.11

大数据技术创新呈现“原创-开源-产品化”的阶梯格局



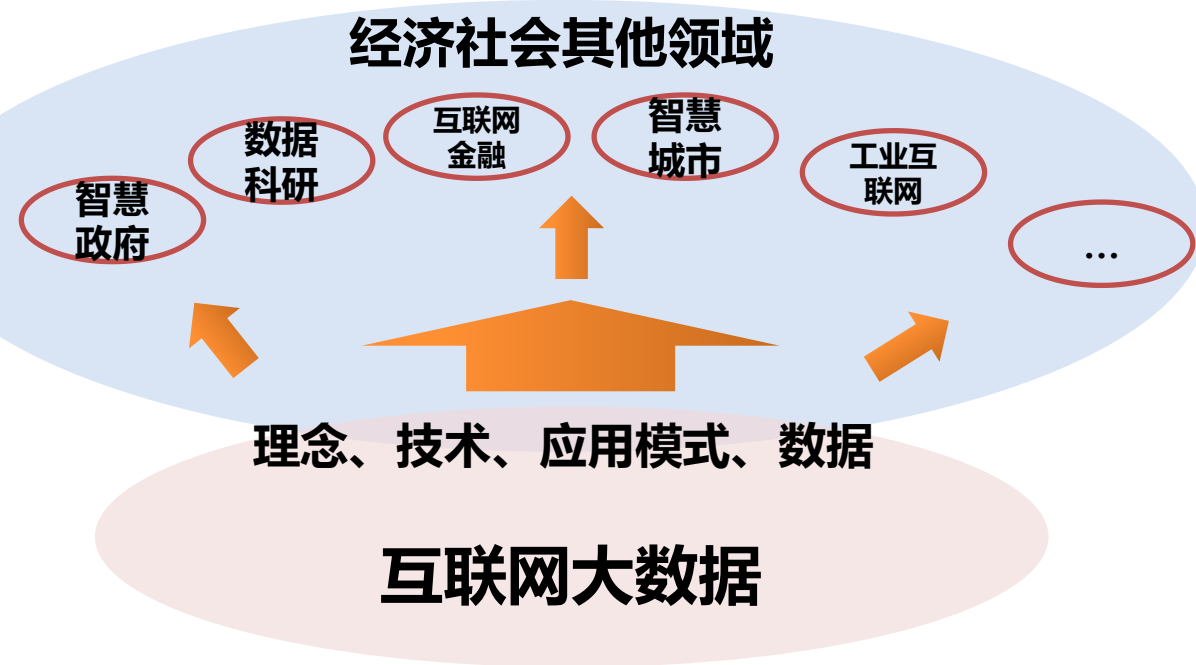
- 1、丰富的数据和强大的平台是大数据创新的基础条件；
- 2、大数据的应用效益不是飞跃突进的，必须依靠长期的不断累积；
- 3、累积效益的获取，主要靠持续不断的技术迭代；
- 4、技术和应用一体化组织，是快速迭代的保障；
- 5、开源模式加速研发和技术扩散。

互联网行业是大数据应用的领跑者



此外还有公益性大数据应用（如流感趋势）和研发类大数据应用（如A/B测试）

大数据应用加速向传统领域拓展，还处于初期



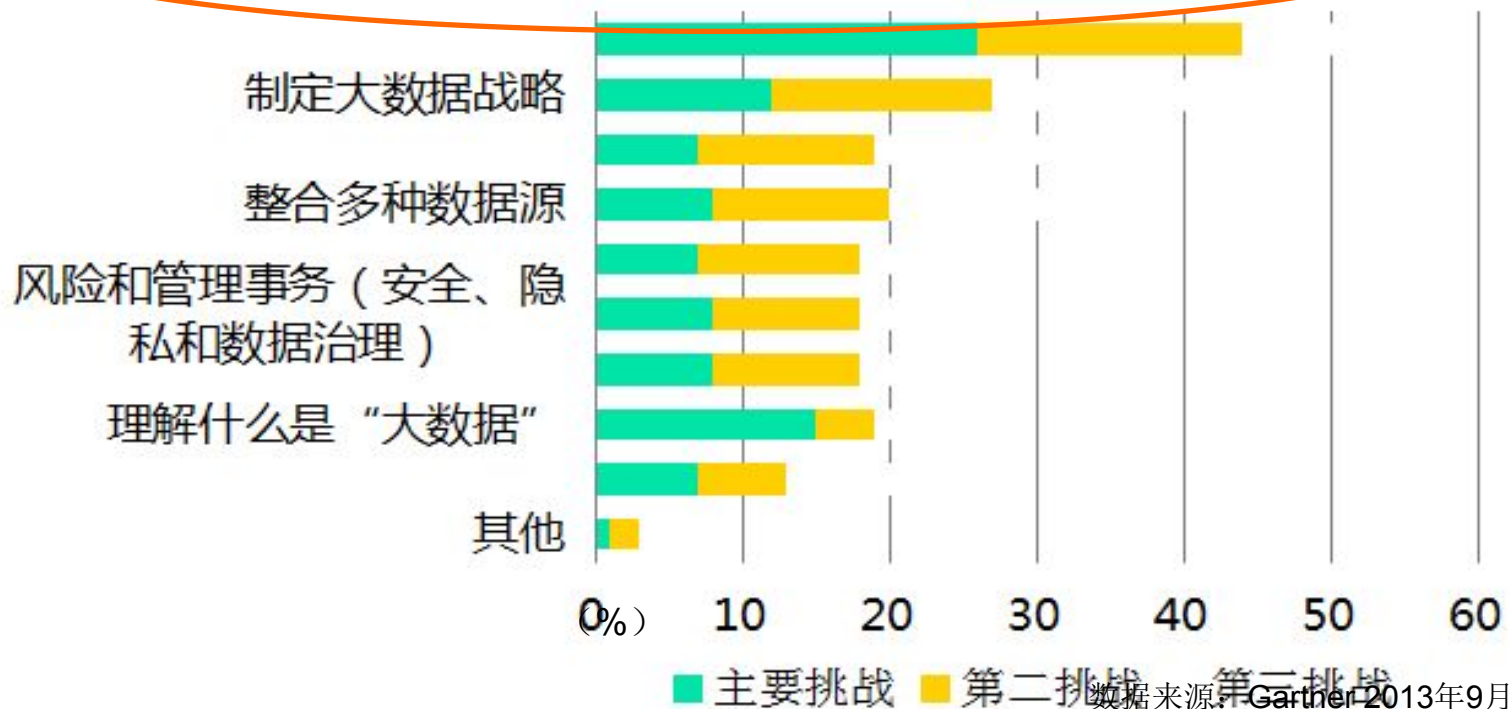
整体：初级阶段

谈概念的多，真正用的少
案例散发，未成燎原之势

- 大数据理念、技术、应用模式、数据正在向更广阔的领域扩散
- 大数据技术与模式的改造和适配还要经历长期过程

大数据火爆，但传统企业仍不知如何应用

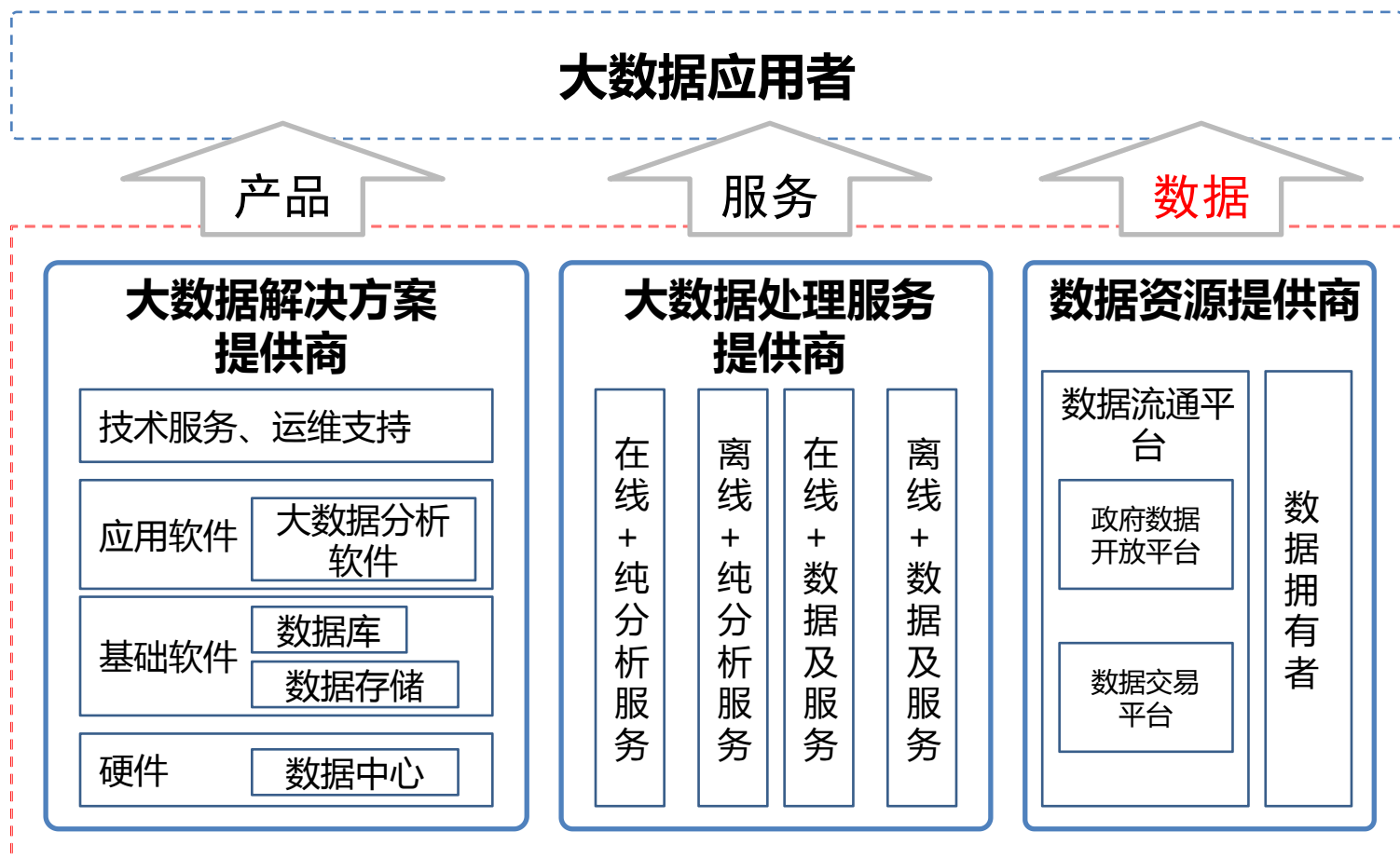
传统企业应用大数据的主要障碍Gartner2013



不知道如何能从大数据中获取价值，成为阻碍大数据应用在多数传统企业和机构普及的最主要障碍

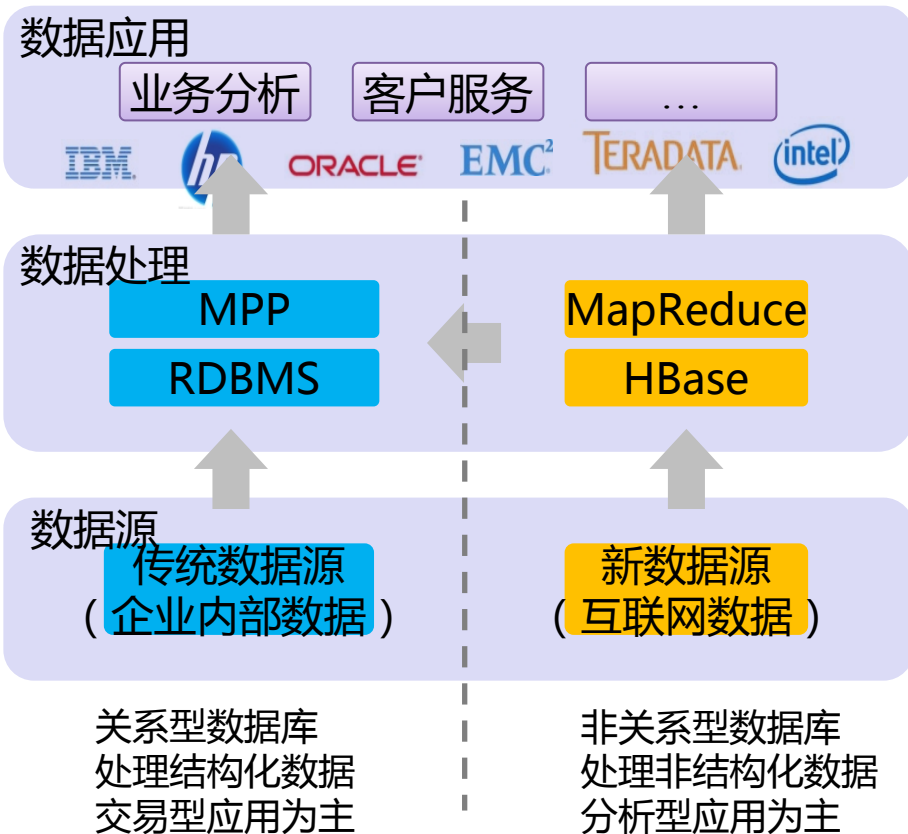
大数据产业活跃，正围绕数据形成新的产业生态

2012年全球大数据市场总体规模为114亿美元，2013年达到181亿美元，2012-2017年的年度复合增长率为31%。（数据来源：Wikibon 2013）



传统厂商加快向大数据解决方案提供商转型

模式1：传统IT厂商提供“双栈”方案



模式2：新兴厂商主攻Hadoop优化



云服务商成为大数据处理服务提供商的主体

在线的大数据处理云服务将成为主流 云服务提供商拥有数据和平台，将成为服务主体

模式1：纯分析服务 + 在线

主要以SaaS形式提供给用户，服务商有互联网企业、大数据分析软件商和新创企业等。



模式2：数据及服务 + 在线

主要以SaaS形式提供给用户，服务商以拥有海量用户数据的大型互联网企业为主。



模式3：纯分析服务 + 离线

为企业提供专业、定制化的大数据咨询服务和技术支持，服务商有大数据咨询公司、软件商等。



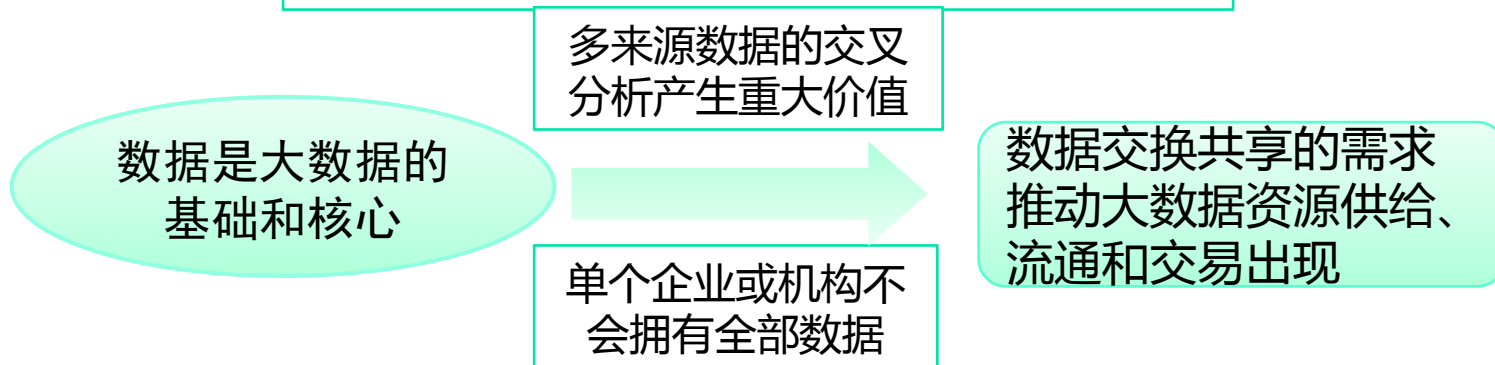
模式4：数据及服务 + 离线

服务商以拥有大量用户数据的行业领先企业为主，目前主要集中在信息化水平较高的行业。



大数据资源提供商应运而生，交易市场萌芽

大数据流通是发展的必然产物



大数据流通平台的模式

政府数据开放平台



数据交易市场



大数据交易规则有待完善

产权界定
交易规则
定价机制
转售控制
隐私保护

提纲

1. 大数据发展脉络与概念探讨
2. 大数据技术、产业和应用进展
3. 大数据发展政策及几点思考



各国积极推动政府和公共数据开放

美国领衔开放数据

事件：2009年1月，美国总统奥巴马签署了《开放透明政府备忘录》，要求建立更加开放透明、参与、合作的政府。**同年，数据门户网站Data.gov上线。**

八国开放数据声明

事件：2011年9月20日，巴西、印度尼西亚、墨西哥、挪威、菲律宾、南非、英国、美国等八个国家联合签署《开放数据声明》，成立开放政府合作伙伴。

63个国家/地区开放数据

事件：2013年6月，八国集团首脑签署《开放数据宪章》，法国、美国、英国、德国、日本、意大利、加拿大和俄罗斯承诺在2013年年底前，制定开放数据行动方案。**目前已有63个国家（地区）加入了开放数据运动。**

2009

2011

2013

时间

各国数据开放的特点

1.统一门户

建设统一的政府开放数据门户是各国普遍做法

2.机器可读

重点开放可机读的**数据集、应用程序（APPS）**等资源

3.广泛参与

开放数据在各国都获得**各个部门的广泛支持**，美国175个部门

技术和应用政策侧重基础技术和公共领域

以美国为例，2012年2亿美元大数据战略，主要投向基础研究和公共领域应用

技术研究：开源基础框架

应用：侧重公共领域

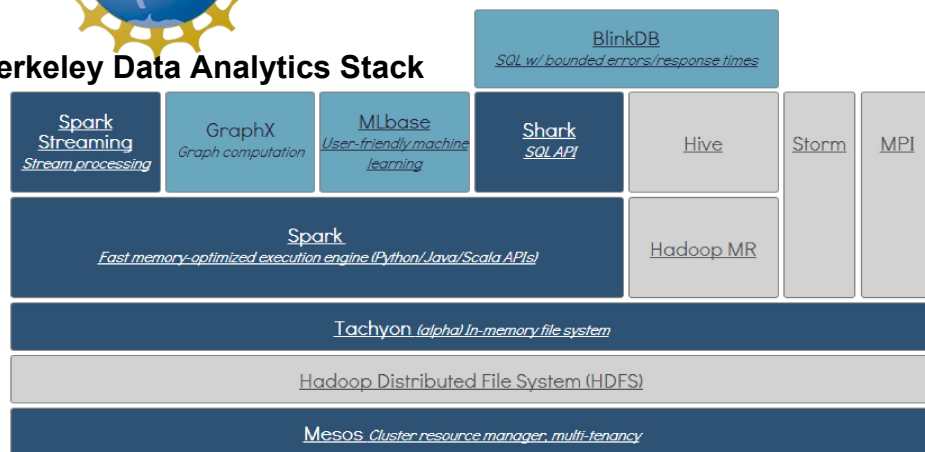
支持基础技术创新，通过开源加强技术扩散

大力资助在国土安全、医疗、能源、环保等公共领域应用



DoD、NIH、DoE、USGS

Berkeley Data Analytics Stack



■ Supported Release ■ In Development ■ Related External Project



例：NSF资助下，加州大学开发出完整的大数据分析完整软件栈，且全部开源，有望替代Hadoop。华为等30多家企业赞助

隐私保护：传统制度面临较大挑战

传统模式

数据在用户知情下主动收集

个人数据的定义是明确的

数据的收集有明确的用途

用户是数据的主体

个人在法律上同意，但不参与

政策框架注重降低个人风险

转变

大数据模式

数据在用户不知情时被动自动收集

个人数据的定义随着场景和习惯变化

数据的用途在数据融合和后续应用才体现

用户可能是主体，也可能是控制或处理者

个人参与数据的使用和价值的产生

政策框架需要平衡保护和创新，以促进发展

传统的知情、同意隐私保护原则将制约大数据价值发挥，放松或加强管制面临两难选择。

世界经济论坛

2013年1月

Unlocking the Value
of Personal Data:
From Collection to Usage

大数据写入政府工作报告



设立新兴产业创业创新平台，在新一代移动通信、集成电路、**大数据**、先进制造、新能源、新材料等方面赶超先进，引领未来产业发展。

——2014年《政府工作报告》

政府在大数据发展中的三点定位

从数据（数据开放、数据资产管理）、技术、人才等方面为国家构筑大数据时代的能力基础



解决大数据应用发展过程中面临的个人隐私保护、数据安全和数据产权制度设计等法律制度环境问题

在公共管理和服务领域积极应用大数据，改善效率、节约成本，提高政府管理和服务能力

对我国大数据发展策略的几点思考

我国发展大数据有很好的基础，互联网企业应用推陈出新，技术局部有所突破，经济转型升级对大数据技术有强劲需求。

谢谢！