

大数据驱动的企业创新发展战略

张涵诚

中关村大数据交易产业联盟副秘书长
北京大学电子商务总裁班大数据课题组主任

培训议题

1. 大数据基础概论与大数据产业生态分析
2. 大数据在企业的应用与实践
3. 大数据产业促企业转型的作用与价值
4. 大数据政策分析和企业发展大数据的战略选择

传统行业

如何在被大数据和移动互联网颠覆之前

实现自我颠覆？

干掉

苹果 干掉 诺基亚

苹果的App Store

飞凡 **干掉** 商城信息解决方案商

免费的干掉收费的

AlphaGo干掉 柯洁

数据+超算+深度学习

京东 **干掉** 中关村电脑城

电商的效率价格

360 干掉 瑞星

免费商业的未来

阿里巴巴 干掉 批发商

马云：阿里巴巴是一家大数据公司

人工智能 **干掉** 一批工作

医生、律师、翻译、

西少爷 x 肉夹馍

黄太极 x 煎饼

懒人x家政

Uberx租车

.....

谁即将干掉我？

周鸿祎说：

有一个对手你永远无法打败它

趋势

用**未来**的自己打败**现在**的自己

百度干了广告的事！

淘宝干了超市的事！

阿里巴巴干了批发市场的事！

微博干了媒体的事！

微信干了通讯的事！

滴滴干了汽车的事情

不是外行干掉内行，是趋势干掉规模！先进的取代落后的！

在信息大爆炸时代你干了什么？

所以我们当时就选择了

大数据



大数据将帮助企业进行产业变革

云是妈，端是爹，养了个儿子是移动互联网。培养花了很大的功夫，移动互联网最终成为一个高富帅，最后看上一个白富美，叫**大数据**。最后会修成正果，结婚。

假如20年过去了，爹、妈、儿子发现，原来一切的一切最后都是围着这个儿媳妇（**数据**）转的。世界最后是属于这个叫大数据的儿媳妇的。他为大家服务，大家也都是为她服务的。

数据说话，数据决策，数据管理。

大数据可以实时获得竞争比业务的反应更快...

BI, 分析&经验决策

- 太多噪音数据
- 事后诸葛
- 依靠猜测

实时自动调节



SaaS部署实时更新升级

批准&部署

- 优先级问题
- 计划实施
- 与市场脱节

重-开发
重-发布
新-平台

每年
每季
每月



开发&QA

- 太多项目
- 矛盾的优先级
- 昂贵

少量IT参与



使用群体智慧

业务、产品&人员协作

- 周期长&高成本
- 专家偏见



大数据可以实时获得竞争比业务的反应更快...

批准&部署研究

- 优先级问题
- 错过截止日期
- 与市场脱节

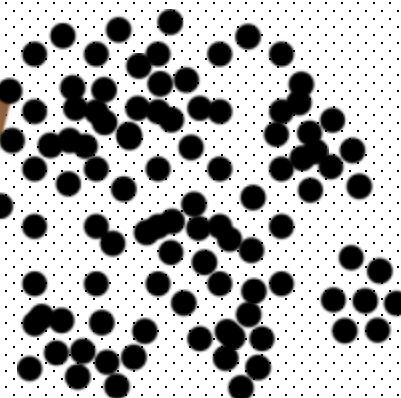
实时自动调节



BI, 分析&多变量测试

- 太多噪音数据
- 事后诸葛
- 依靠猜测

系统自动升级



重-设计
重-发布
新-平台

每年
每季
每月

开发&QA

- 太多项目
- 矛盾的优先级
- 昂贵

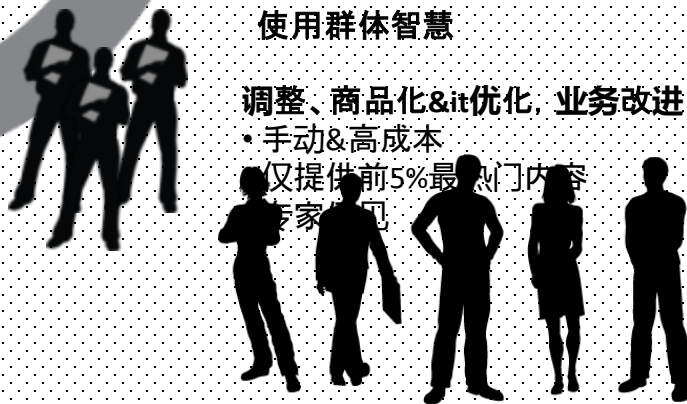
少量IT参与



使用群体智慧

调整、商品化&it优化, 业务改进

- 手动&高成本
- 仅提供前5%最热门内容
- 专家意见



我们的一言一行、一举一动都迅速的转变成大数据，然后转变成商家的洞察力和企业决策。大数据将做为仅次于能源的生产要素驱动所有产业进行革命，生产制造、品牌营销、商业模式。大数据新阶段，你准备好了吗？

一切都在数据化、这代表了新的生产力，生产关系.....

数据化、智能化转型是必然趋势！

从大数据案例说起

1. 杜蕾斯的雨夜之疯狂传播
2. 大数据进行抓小偷、抓鸡头
3. 大数据助力小米进行社会化营销
4. 用户评论数据助力海尔商城起航

目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

商业模式驱动



数据竞争的压力



能力向消费者转移



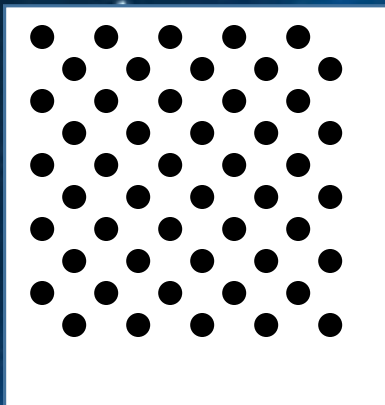
大数据的激增

“数据现在就像新型石油。数据就像原油一样，非常宝贵，但是如果未进行优化则毫无用处。” – 专家

“我们的经济是基于一种不仅可再生并且会自我生成的资源。用完不是问题，被淹没才是问题。” – John Naisbitt

全新的计算时代

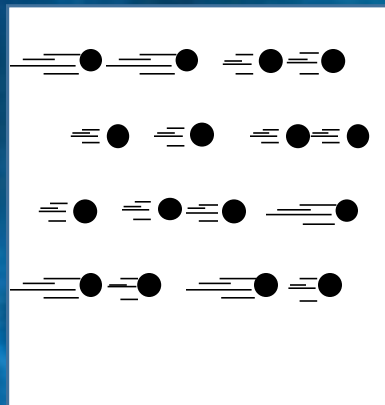
Volume多



Data at rest

Terabytes to
exabytes of
existing data to
process

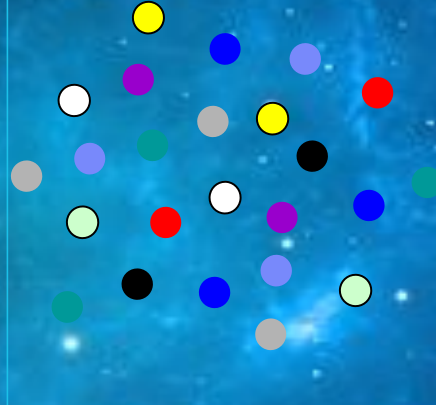
Velocity快



Data in motion

Streaming data,
milliseconds to
seconds to respond

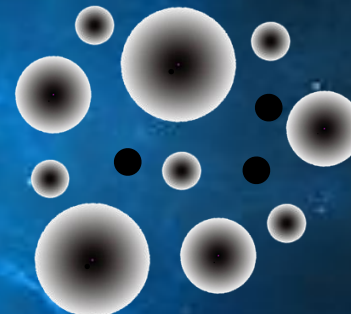
Variety杂



Data in many
forms

Structured,
unstructured, text,
multimedia

Veracity垃圾*



Data in doubt

Uncertainty due to
data inconsistency
& incompleteness,
ambiguities, latency,
deception, model
approximations

* Truthfulness, accuracy or precision, correctness

目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

何为大？—数据度量PB

- 1Byte = 8 Bit
- 1KB = 1,024 Bytes
- 1MB = 1,024 KB = 1,048,576 Bytes
- 1GB = 1,024 MB = 1,048,576 KB = 1,073,741,824 Bytes
- 1TB = 1,024 GB = 1,048,576 MB = 1,099,511,627,776 Bytes
- 1PB = 1,024 TB = 1,048,576 GB = 1,125,899,906,842,624 Bytes
- 1EB = 1,024 PB = 1,048,576 TB = 1,152,921,504,606,846,976 Bytes
- 1ZB = 1,024 EB = 1,180,591,620,717,411,303,424 Bytes
- 1YB = 1,024 ZB = 1,208,925,819,614,629,174,706,176 Bytes

大数据的“大”

- 《红楼梦》含标点87万字（不含标点853509字）
- 每个汉字占两个字节：1汉字=16bit = 2*8位=2bytes
- 1GB 约等于671部红楼梦
- 1TB 约等于631,903 部
- 1PB 约等于647,068,911部
- 美国国会图书馆藏书（151,785,778册）（2011年4月：收录数据235TB）
- 中国国家图书馆：2631万册
- 1EB = 4000倍美国国会图书馆存储的信息量
- 600美元的硬盘就可以存储全世界所有的歌曲
- MGI估计,全球企业2010 年在硬盘上存储了超过7EB(1EB 等于10 亿GB)
- 的新数据,同时,消费者在PC 和笔记本等设备上存储了超过6EB 新数据

阿里巴巴云梯—数据规模(2017)

- 集群容量
- —约20000台服务器
- —物理CPU ~60000核
- —内存 ~3000TB
- —磁盘 ~360000块
- —存储容量 ~2000PB



- 集群负载
- —每天Job数 150,000+道
- —每天hive query数 6,000+
- —每天扫描数据量 ~17.5PB
- —每天扫描文件数 ~4亿
- —存储利用率~80%
- —CPU利用率~65% 峰值80%

传统数据



全行业



全民



延迟性

01

传统数据分析平台建设周期长
传统数据分析以离线分析为主，分析周期长，无法实时响应市场、营销分析需求

有限洞察

02

传统数据分析所使用数据有限，以结构化数据为主，无法全面分析全量数据，对服务对象洞察能力有限

高门槛

03

传统数据分析使用专业平台和工具，分析平台使用技术门槛高，同时不易使用

高成本

04

传统数据分析平台多在大中型企业建设使用，采用“IOE”方式，建设周期长，投入大，ROI较低



大数据的优势

01



实时性高

- 大数据分析采用 Hadoop、Spark、流处理、内存、分布式等平台及技术，为实时数据分析提供了良好的技术支撑
- 大数据分析采用专业、可视化分析工具，实现“闪电式”分析

02



准确性强

- 大数据分析融合分析了企业内外部原本孤立的数据，分析数据覆盖更全面，维度更多样化
- 大数据分析采用R语言、Mhout、Mlib等专业数据分析工具、技术，分析成果更准确
- 大数据时代，分享成为一种常态，通过借鉴他人经验，如算法，可快速提升分析准确性

03



ROI高

- 大数据分析采用云计算、SaaS、大数据等相关技术，去“IOE”，降低建设成本和周期
- 分享精神，开源工具及技术使用，降低分析成本，提升数据分析ROI



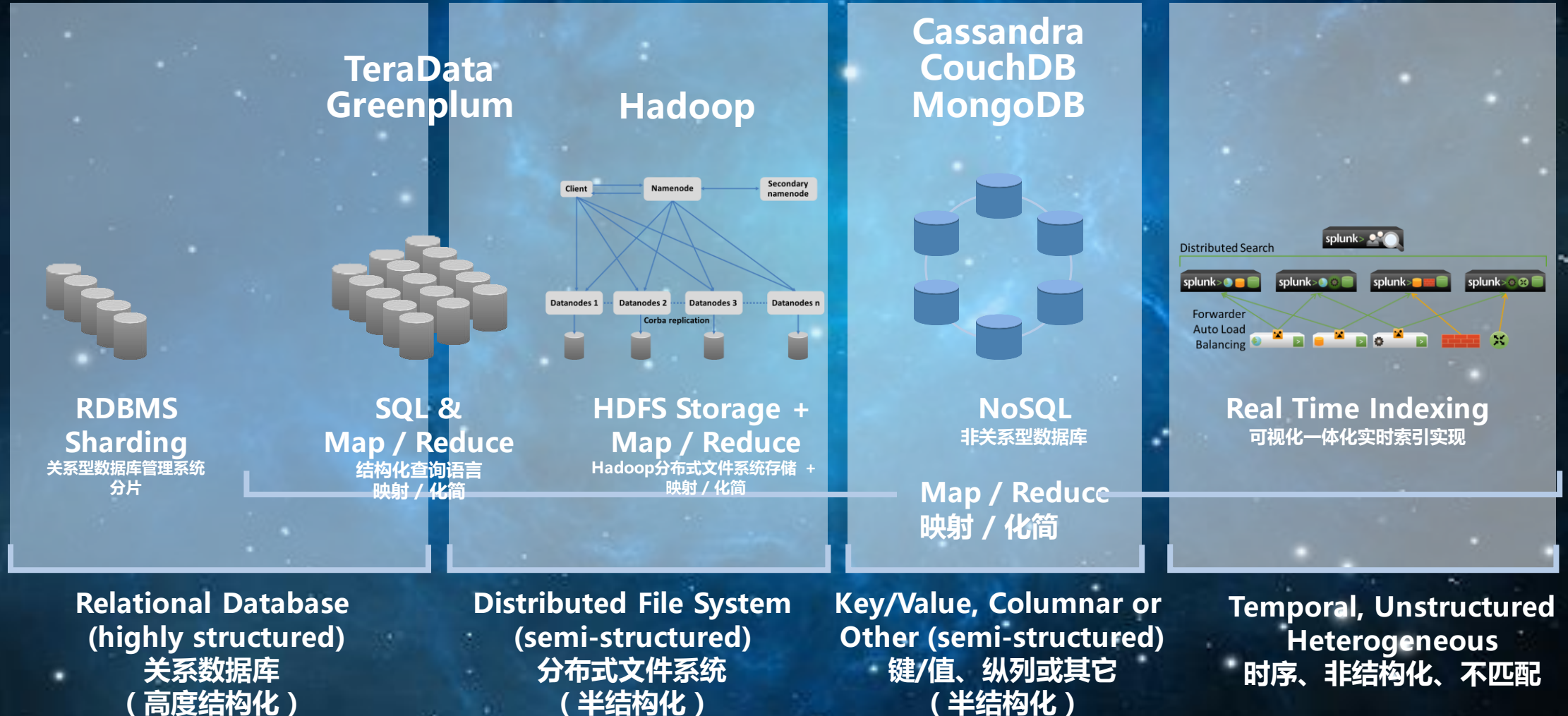
04



门槛低

- 可视化、移动化、拖拽式、专业化分析产品降低数据分析使用门槛
- 云分析平台、SaaS方式应用，降低数据分析建设门槛

Big Data Technologies 大数据技术



目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

大数据约90%是机器数据



商业应用程序数据2%

- 关系型数据、高度结构化、基于僵化模式
- 财务记录、多维数据、数据计算
- 月报，非实时事件



人为产生的数据8%

- 由人与人之间的互动而产生
- 包括电子邮件、即时通信、语音、视频和文本
- 储存在集中式公司服务器、文件共享和桌面中



机器产生的数据90%

- 时间序列非结构化数据，无预定义模式
- 由所有 IT 系统生成，大量不同类型的格式
- 巨量；快速导航和相关性最重要

互联网络



谷歌大数据中心：全球有8个（几十万台服务器）



全国几百万平方米几百个数据中心



目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

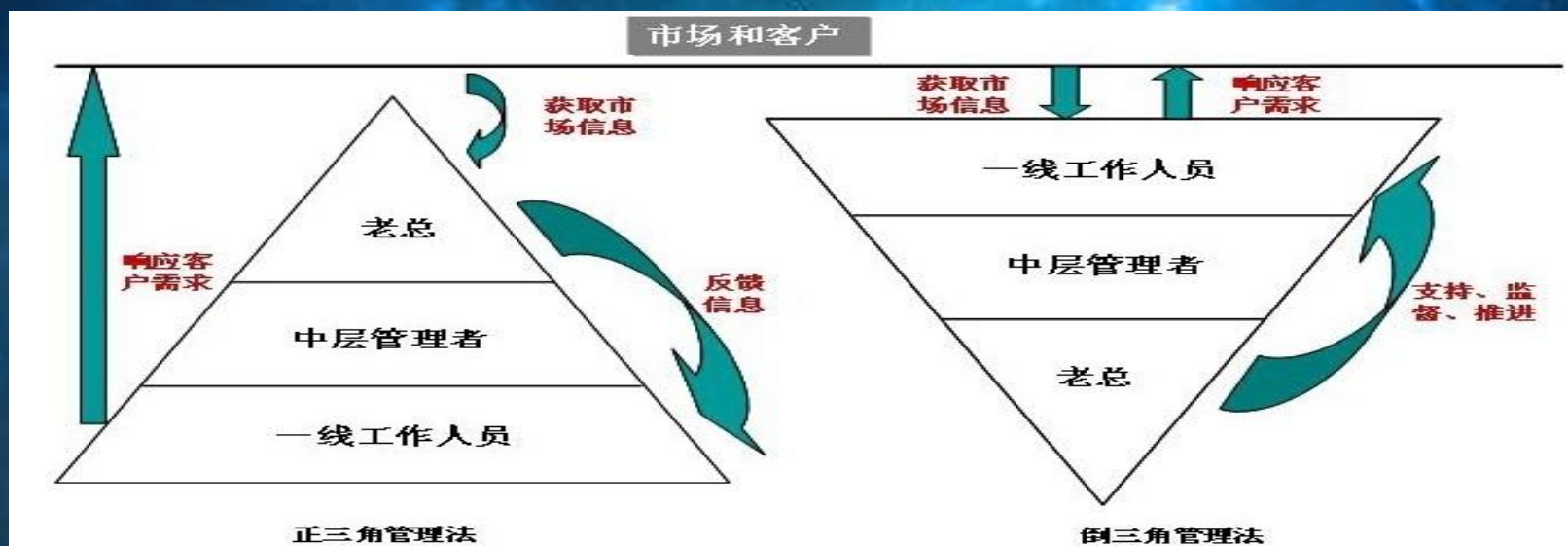
信息社会的发展&企业内部管理的变革

C端用户应用的网络发展：

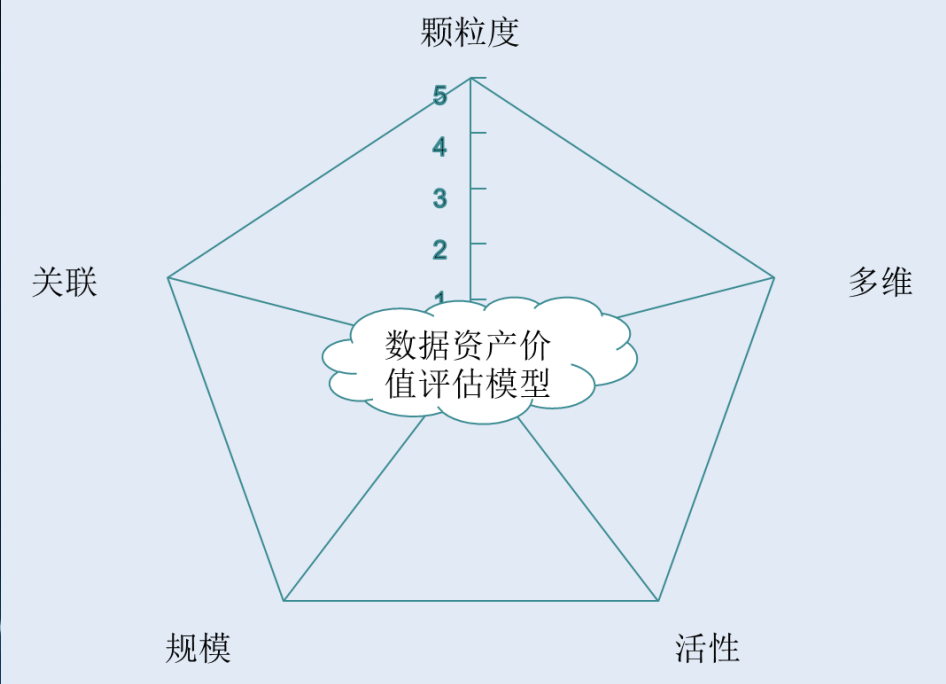
- 目录式（yahoo）-搜索式（google）-分享式（FaceBook）-推荐式（个性化信息服务平台）

B端企业内部管理变革

- 小农生产--小规模生产-大规模生产-大规模按需定制-大规模个性化生产-大规模协同生产
- 营销管理变革 大众-分众-微众-个性化 粗放-精细化



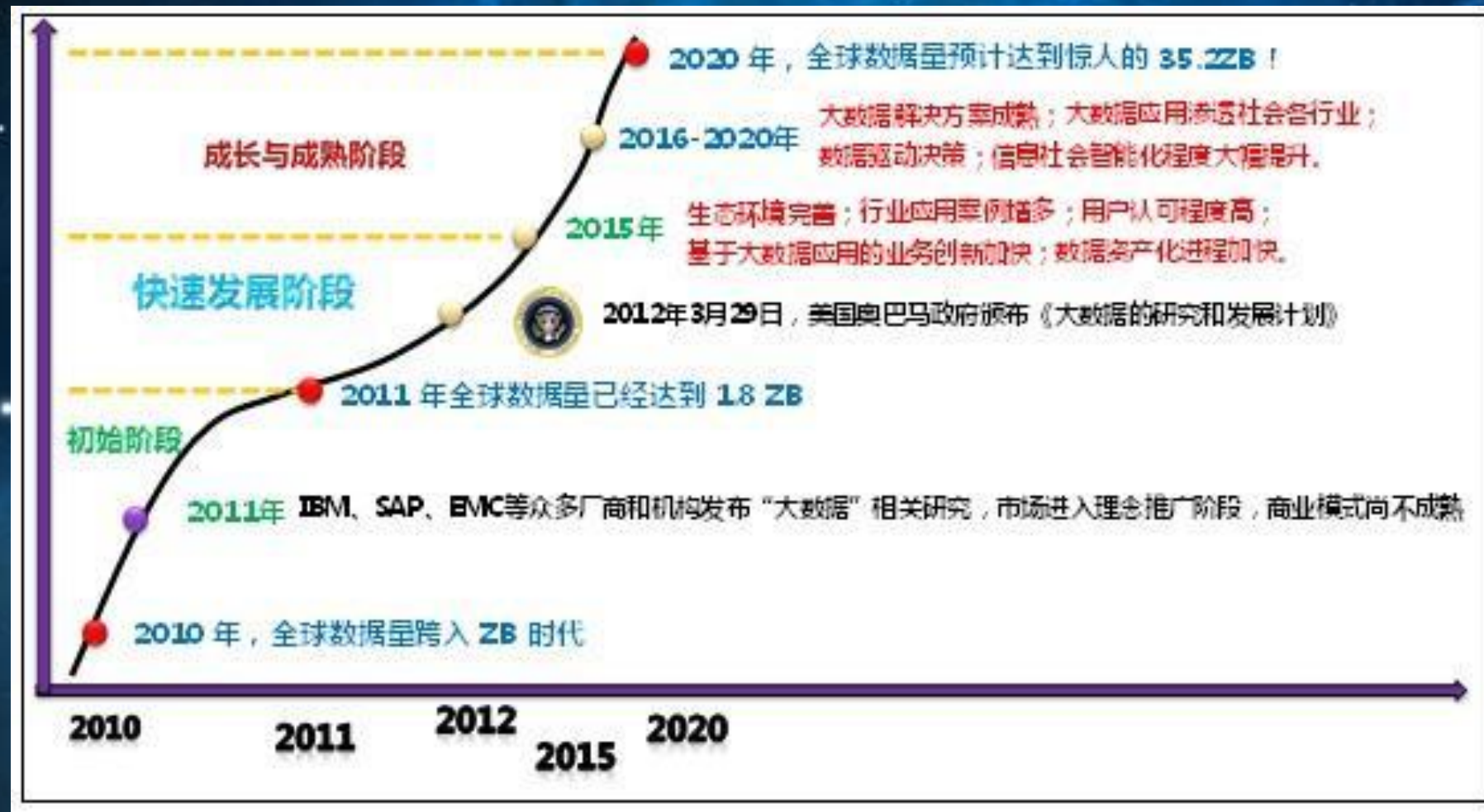
从采样到多维到全体样板



	银行	运营商	社交网络	电子商务	搜索引擎
数据标识	账号	手机号码	邮件账号	注册账号	无
身份真实性	●	●	◐	●	◐
规模	●	●	●	●	●
颗粒度	●	●	●	●	●
活性	◐	●	●	●	●
多维	◐	◐	●		●
关联性	◐	●		◐	◐

数据资产评估模型

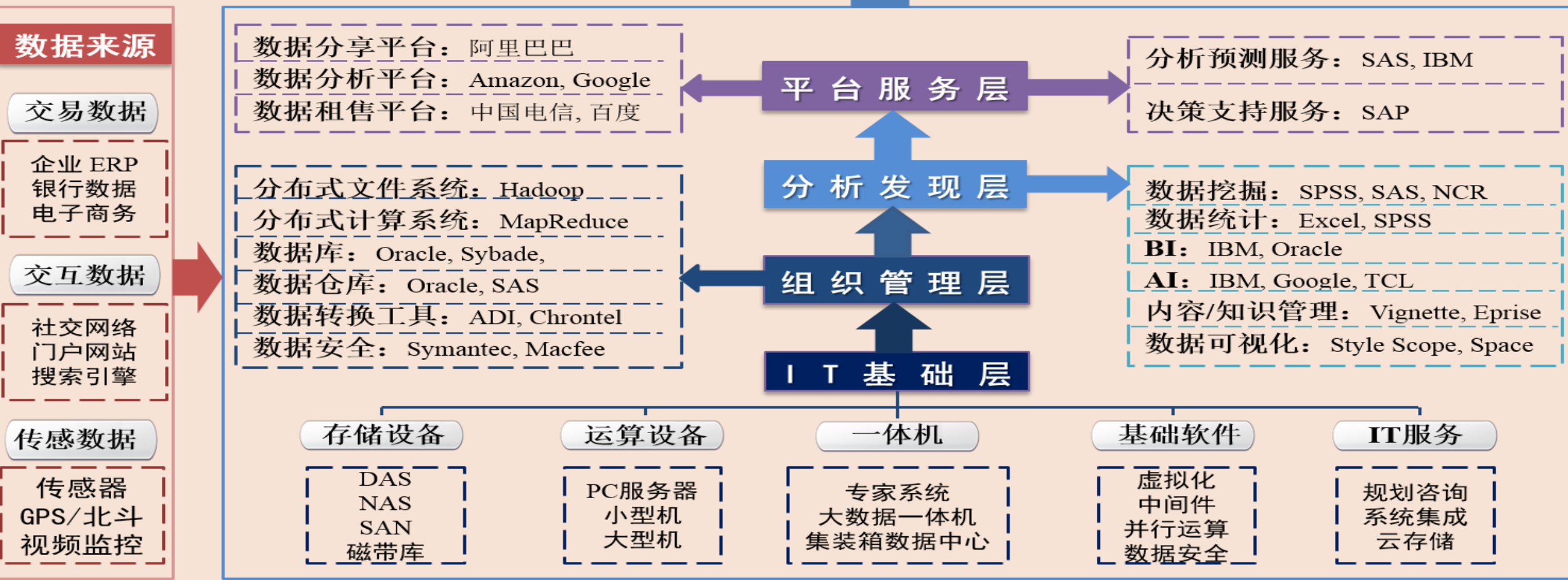
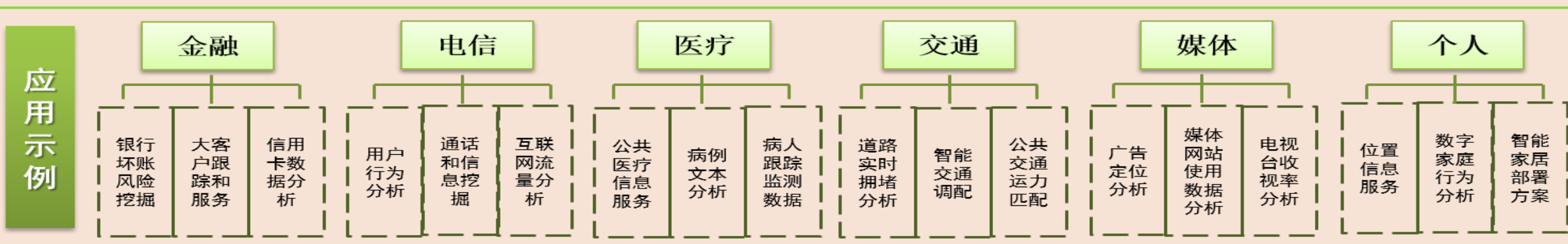
从初始期-发展期-成长成熟期



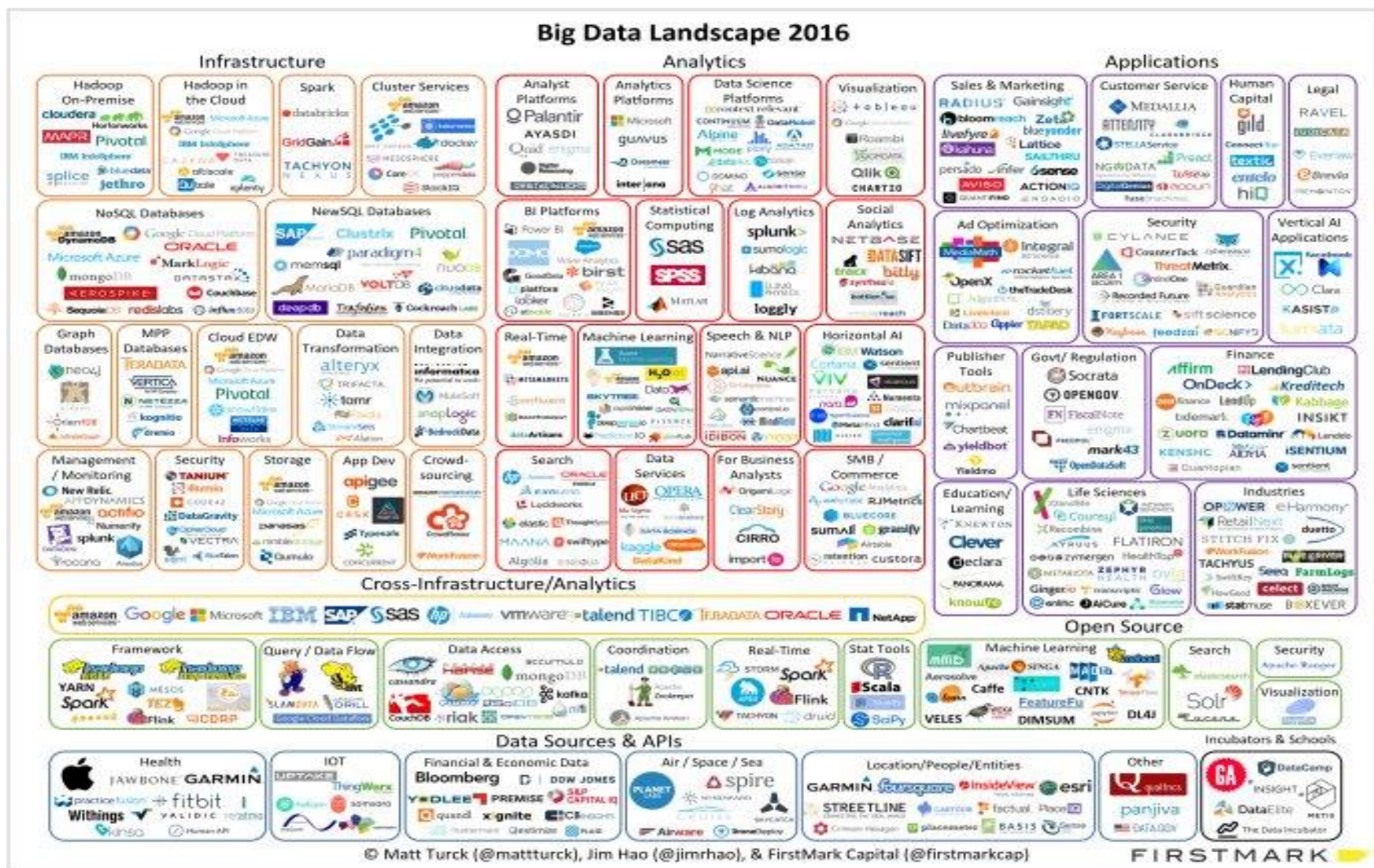
目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

大数据产业链全景图



大数据产品生态图谱



IT角度

大数据生态系统日臻完善

互联网大数据：互联网环境下蕴含丰富洞察、决策和优化能力的海量、高增长率和多样复杂化的信息资产

Analysys
易观智库

数据使用



数据管理



数据收集



垂直化应用

专注于实现某种垂直化功能的大数据应用产品服务提供者



行业化应用

专注于解决某些行业业务问题的大数据应用产品服务提供者



云存储平台

将互联网中大量不同类型的存储设备通过应用软件集合起来协同工作，共同对外提供数据存储和业务访问功能的平台



第三方数据平台

独立于数据源和数据使用者，提供专业化的、依托于网络的数据监测、数据交易、数据管理等服务的平台



技术服务

面向细分领域的专业化数据管理技术服务提供者



物联网

利用局部网络或互联网等通信技术把传感器、控制器、机器、人员和物等通过新的方式联系在一起，形成人与物、物与物相联，实现信息交换、业务管理控制和智能化的服务提供者



数据源

某种所需要数据的原始媒体服务提供者。数据源可以是任何数据类型，但必需可靠且具备更新能力



垂直化应用

垂直化应用

数据交易						
移动开发者服务						
商业WIFI						
语音智能识别						
互联网舆情						
WEB流量分析						
业务性能管理						
智慧城市						
地理位置数据应用						
精准营销						
信息安全						

行业化解决方案角度

行业化应用

电信行业



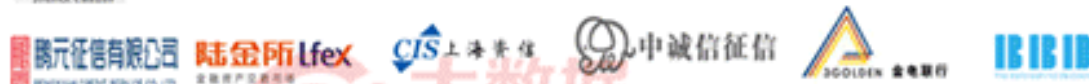
金融行业



个人征信行业



企业征信行业



医疗卫生行业



娱乐行业



农业行业



人才招聘行业



交通行业



安防行业



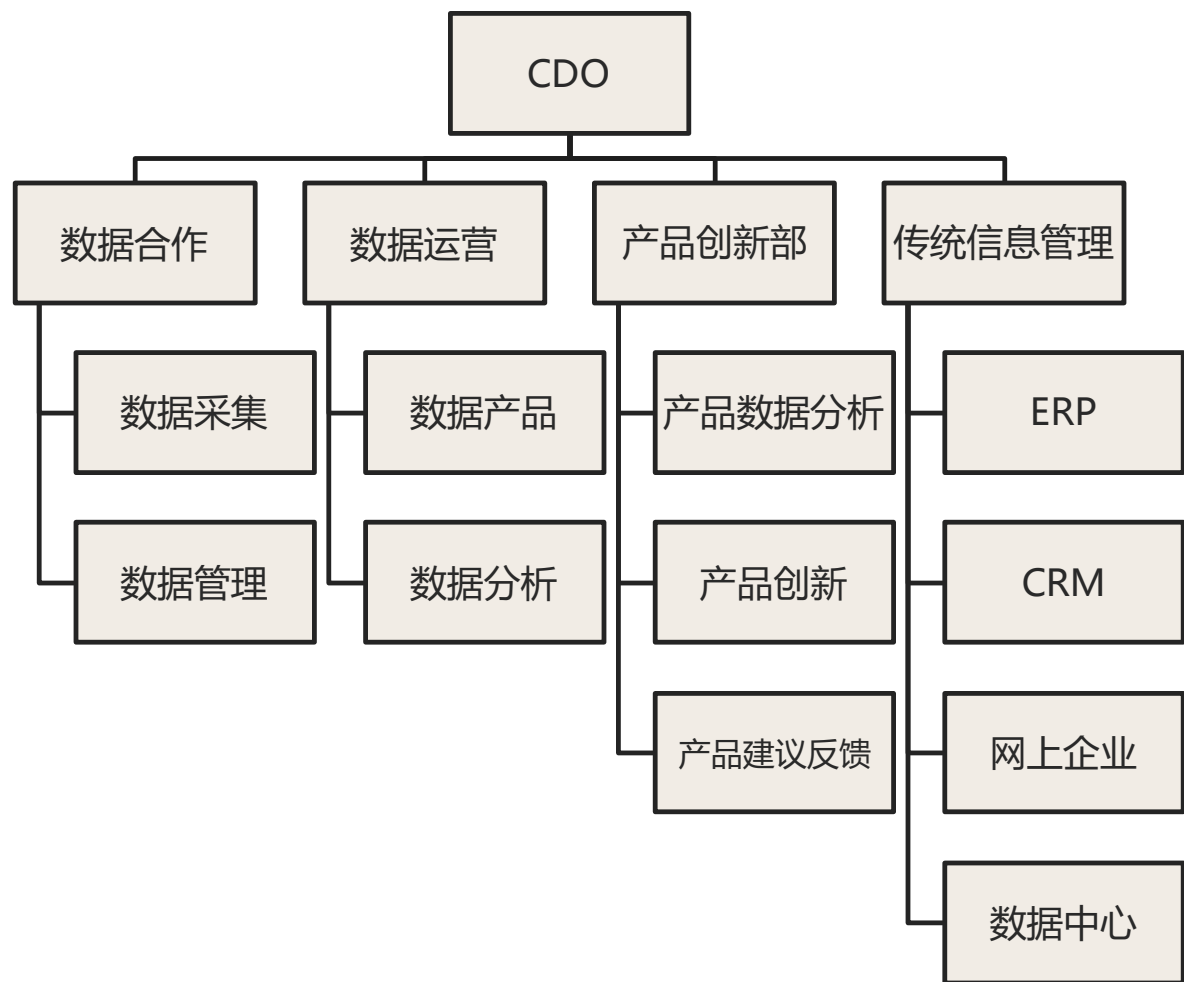
目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

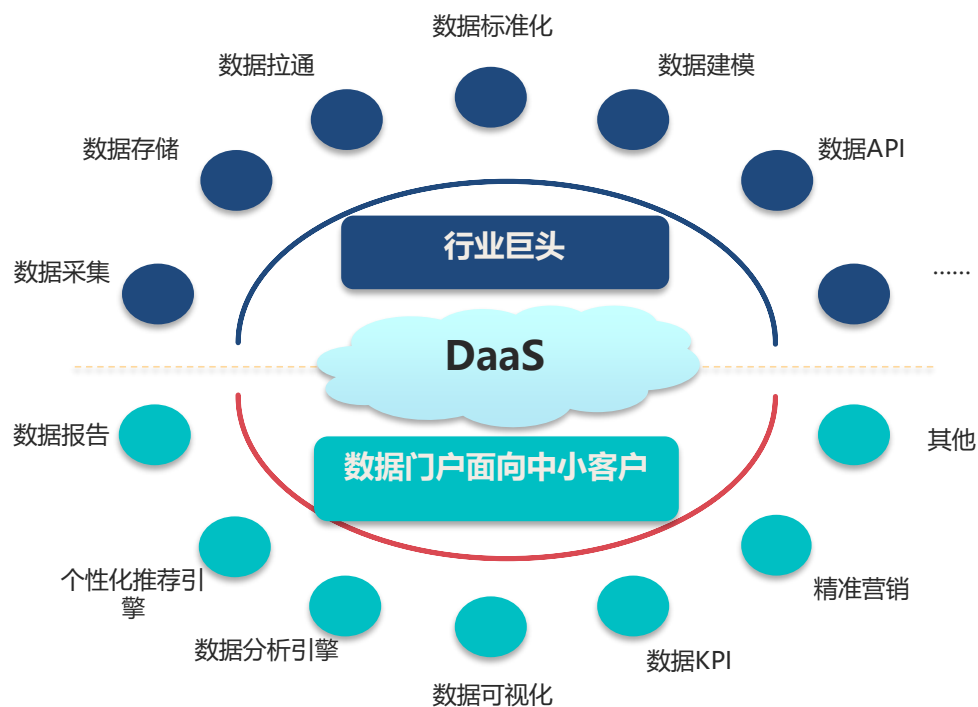
思维：大数据时代商业模式在发生变化

1. 卖产品产品转变为卖服务 企业自身的转型 粘性的增强
2. 服从管理转为创造客户价值 组织的再造 可持续
3. 利用数据成为平台型企业 数据连接世界 按客户需求聚散资源 企业无边界
4. 支持到微企业 自主经营体 管理无领导
5. 产品按需定制 持续创造客户需求 供应链无尺度
6. 数据驱动 资源共享 业务融合
7. 分享经济 所有的产品都可以共享 除了老婆孩子等亲人

人才：大数据驱动组织架构变化



企业需要数据门户+行业数据引擎



行业DaaS平台：服务行业巨头

需要释放数据价值获取更多数据

- 行业数据的提供
- 行业数据分析的结果
- 对数据进行分析
- 数据驱动产业变革

数据门户：服务小企业

需要DaaS新的云化有数据的环境

- 工具
- 数据

目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

大数据涉及的关键技术

	需求		关键技术	技术描述
大数据采集处理	海量数据分布式处理		Hadoop 生态系统	针对大量数据进行分布式处理的系统框架
	非结构化数据处理		文本处理技术；自然语言理解；多媒体处理技术...	文本内容分词与分析；图像、音视频分析
	实时数据处理		Streaming Data	流计算引擎
大数据分析	可视化交互界面		交互式可视化探索分析技术	通过交互式可视化界面辅助用户进行分析
	智能数据分析		大规模机器学习技术	计算机模拟人类学习行为，包括特征提取、图形生成等
存储、组织、管理	数据隐私保护		数据隐私防范保护措施与数据安全技术	保护隐私数据与信息个体的对应关系等安全技术
	高效存储和管理大规模数据		数据存储备份技术、数据放置和调度技术、数据溯源	存储、放置、调度大规模的数据

目录

1. 为什么要研究大数据 (WHY)
2. 什么是大数据 (WHAT)
3. 哪里有大数据 (WHERE)
4. 大数据现在什么阶段 (WHEN)
5. 谁在做大数据 (WHO)
6. 大数据也是一种思维 (HOW)
7. 大数据有哪些关键的技术 (Technology)
8. 大数据的应用场景 (Value)

大数据应用案例

Amazon个性化推荐系统

The screenshot displays the Amazon homepage with a focus on personalized recommendations. At the top, the Amazon logo and navigation links like 'Today's Deals', 'Gift Cards', and 'Help' are visible. A search bar and a 'Hello, Michelle' greeting are also present. The main content area features several promotional banners and product recommendations:

- Kindle:** A large banner for the Kindle e-reader, stating 'The Best-Selling e-Reader in the World' and 'From \$114'. It includes a 'Shop now' link.
- Used Textbooks:** A banner advertising 'Save up to 90% on Used Textbooks' with a 'Shop now' link. It shows a stack of books including 'PSYCHOLOGY', 'DISNEY WORLD', and 'FINANCIAL ACCOUNTING'.
- Top 100 Kindle Books:** A section titled 'Top 100 Kindle Books' with a 'Shop now' link.
- YOUR REWARDS ARE WAITING FOR YOU AT CHECKOUT:** A promotional box for Amazon Rewards, stating 'Redeem 100 points = \$1 towards your purchase' and 'Use any number of points - no minimums'. It includes an 'Amazon Gift Card' image and a 'Learn More' link.
- BEARS-WAR 3:** A promotional box for the game 'BEARS-WAR 3', stating 'Pre-order the standard edition and get a \$20 promotional credit'. It includes a 'Pre-order now' link.
- New For You:** A section titled 'New For You' showing a row of recommended products, including a gold-colored item, a black device, a book, a 'DRAKE HEADLINES' book, a smartphone, and a 'MAKING THINGS MOVE' book.
- Sell Back Your Copy:** A section titled 'Sell Back Your Copy' with a 'Reserve an Amazon.com Gift Card for 100 Items' and a 'Making Things Move' book.

个性化邮件推荐-亚马逊

在 Amazon.cn 购买 谭(Pang-Ning Tan) 的 "数据挖掘导论(英文版)", 可节省 22% [收件箱](#) | [X](#)

☆ Amazon.cn 发送至 我

[显示详细信息](#) 11月13日

[回复](#)



尊敬的卓越亚马逊客户,

作为曾经购买或评价过 费尔德曼(Feldman.R.) 的 [文本挖掘\(英文版\)](#) 的人, 您可能想知道现在可以订购 [数据挖掘导论\(英文版\)](#) 了。 点击下面的链接预订, 您只需支付 ¥ 46.00 (目录价的 ¥ 13.00)。



[数据挖掘导论\(英文版\)](#)
谭(Pang-Ning Tan)

市场价: ¥ 59.00

现价: **¥ 46.00**

为您节省: **¥ 13.00 (22%)**

由卓越亚马逊卖出并发货



添加到购物车

编辑推荐



《数据挖掘导论(英文版)》是经典原版书库。

案例：智能学习与实时更新



顾客在电子商务网站上所有行为都会被跟踪分析，实时反馈到推荐引擎中，实时改变给用户推荐的内容

商品列表

商品名称	属性	操作
 冬季女性内衣 ¥125.80元 删除 放入收藏夹		删除 放入收藏夹
 冬季女性内衣 ¥87.00元 删除 放入收藏夹		

购物车小计 ¥125.80元 比市场价 ¥147.00元 节省了 ¥21.20元 (14%)

当顾客再添加一件冬季的女性内衣时

推荐引擎会实时更新用户偏好，立刻推荐出更偏向于冬季的女性商品

基于您购物车中商品的推荐


¥255.000
MDX 冬季小衫 冬日暖阳心曲
商品属性: 颜色: 粉色 尺码: 均码
价格: 无货


¥250.000
L&L 冬季小衫 冬日暖阳心曲
商品属性: 颜色: 黑色 尺码: 均码
价格: 无货


¥1430.000
STOON 红色 手机
商品属性: 颜色: 红色 品牌: STOON
价格: 无货


¥155.000
MDX 冬季小衫 冬日暖阳心曲
商品属性: 颜色: 粉色 尺码: 均码
价格: 无货


¥105.000
【品牌】冬季小衫 冬日暖阳心曲
商品属性: 颜色: 黑色 尺码: 均码
价格: 无货


¥90.000
【品牌】冬季小衫 冬日暖阳心曲
商品属性: 颜色: 黑色 尺码: 均码
价格: 无货

powered by 百分点

女性相

大数据挖掘销售机会



outwomannana_宝贝凤凤宝贝我★: 我在乐峰网看百雀羚的东西, 哈哈, 好想买~



7月22日21:31 来自新浪微博

转发 | 收藏 | 评论(2)

消费客户



-舌甘-: 同一品牌的化妆品用了好几年, 想换换了, 你们都在用什么化妆品, 推荐一下嘛 @梧梧籽 @丫aaaa @躲在床下的壁虎 @LONGINES_83 @余晶洁 @幸福旋律1987

7月18日21:10 来自iPhone客户端

转发 | 收藏 | 评论(4)

意向客户



墨沫: #想换个化妆品大家给推荐个牌子? 24岁#

7月18日14:29 来自新浪微博

转发 | 收藏 | 评论

目标客户



陌轩紫岚: 聚美优品家东西涨价了? 为何同样的东西同样的分量, 价格却差了这么多, 只是因为新款么? 那请问新在哪里呢?



7月23日16:11 来自新浪微博

转发 | 收藏 | 评论(3)

潜在客户

大数据分析在汽车行业的应用

User Segmentation

对用户标识的Tag进行聚类研究，
确定面向应用的用户分群



汽车

看车

买车

产品库
经销商

用车

用车服务
轮胎
油耗
保养
问答

汽车八卦图片



打酱油型



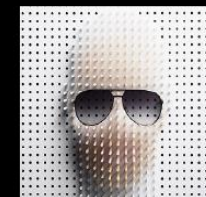
看热闹型



汽车兴趣型



买车念头型



买车计划型



快要出手型

计算机上虚拟育种：大数据+超级计算

Virtual Breeding: Big Data + Supercomputing



曙光星云**FLOP: 3P**

华大自主**FLOP: 200T**



天河一号**FLOP: 4.7P**



天河二号**FLOP: 110P**

5000株玉米, $R^2=0.78$, 需要**30T**的计算能力
5000株玉米, $R^2=0.9$, 需要**200T**计算能力
10000株玉米, $R^2=0.9$, 需要**5P**的计算能力

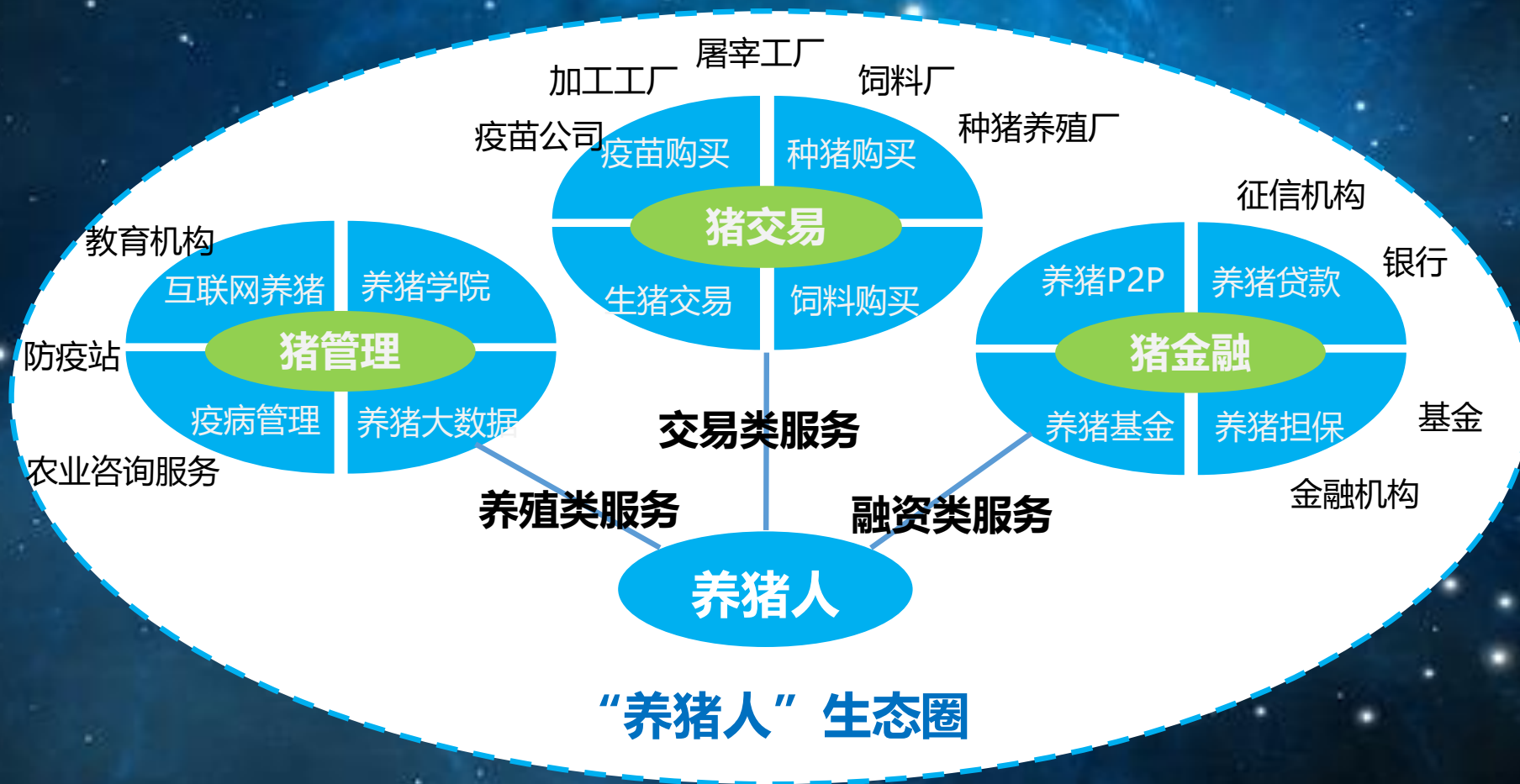


全生命周期、实时全景贯穿 Real Time , Whole picture of whole life
精准预测、该存就存、该防就防，该变就变，该补就补、该治就治、该死就死



农业大数据：养猪大数据

通过建立养猪人平台，聚合交易、养殖、融资、咨询、采购等一系列应用，为养猪人提供一站式服务，增强养殖户的粘性。



营销大数据：大数据让营销更为精准-有更强的可信度

- 1、微博客中汇聚着来自不同行业不同阶层的民众，通过微博客产生互动与联系，使社交关系网络更加透明，同时也真正让企业可以有机会零距离的与用户面对面的沟通。
- 2、基础信息、兴趣性息、情感信息，通过搜索和语义分析技术，让精准营销成为可能；



能源管理大数据

能耗有效管理始于精准、全面、标准化的能耗排放数据计量。

静态数据

- 基本配置信息：名称、编号、归属地市、位置信息（GPS经纬度）、面积、总载频数等；
- 设备信息：标识码、名称、厂家、设备类型、规格型号、用途、入网时间、额定功率、额定负荷 等；
- 节能措施：技术名称、实施时间、描述等；

动态数据

- 能耗数据：用能场所总耗电量、主设备耗电量、空调耗电量；

AI量：电流、电压、温度、湿度、压力、流量、振动；
DI量：告警信息；

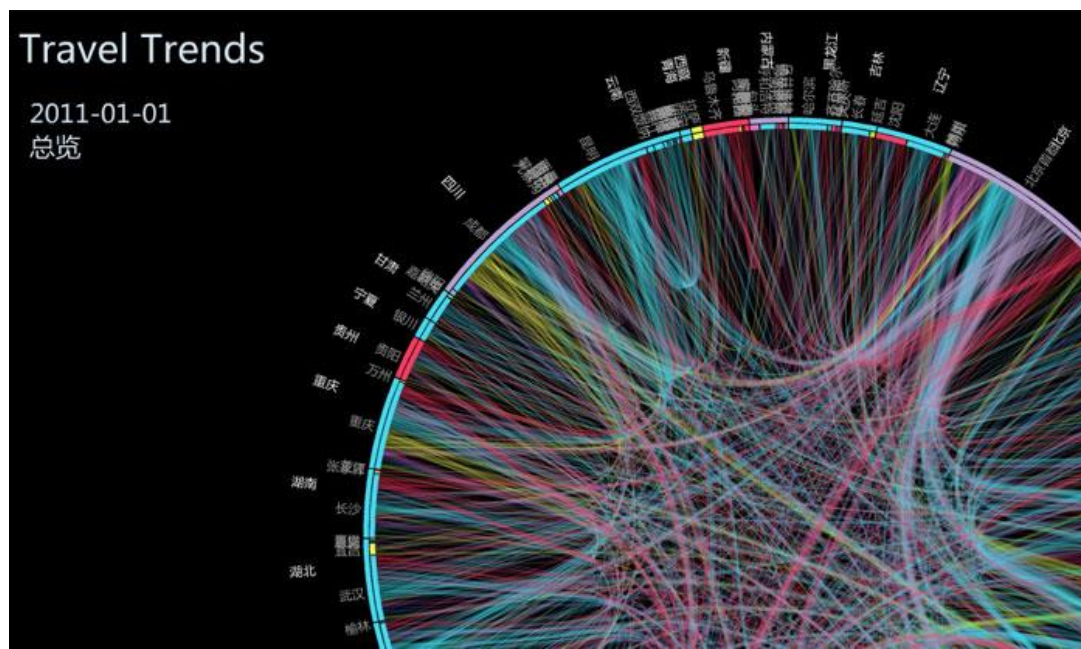
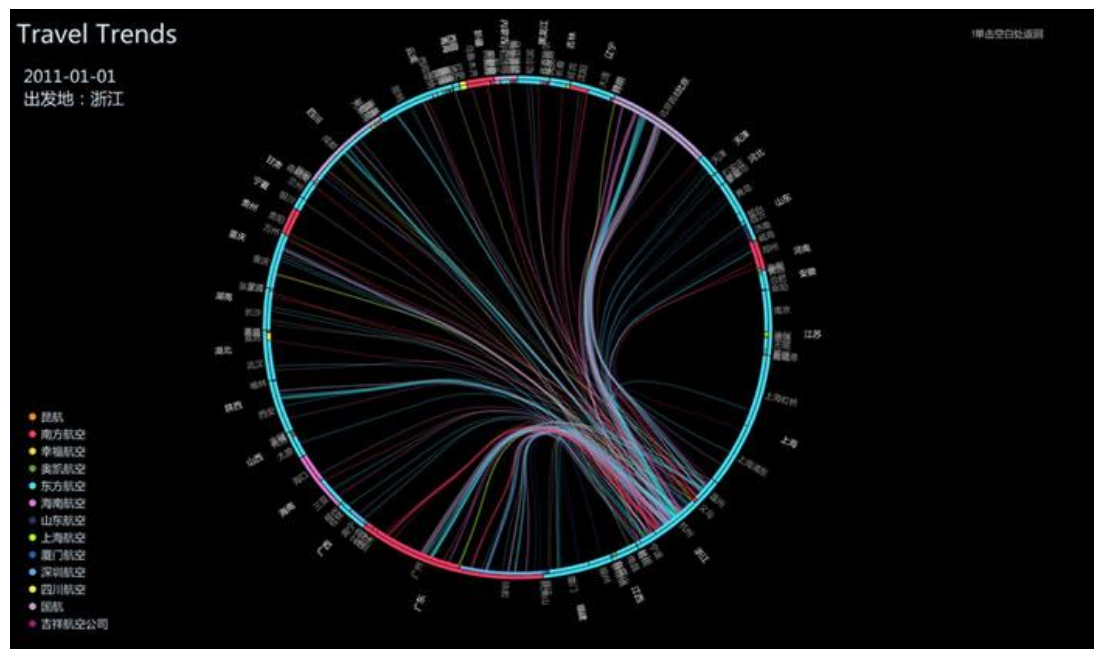
- 环境因素：室内外温湿度；
- 业务量：话务量、数据业务流量；

实现耗电量、PUE值、电费、告警的分时段、任意时段统计查询。

航空大数据可视化案例

航班走势

航班走势是全面展示某航空公司机票业务的新应用，通过该应用可以清晰的**看到指定时间点，人们乘坐飞机在哪些城市间往来**，在这些航线上**各航空公司客源的占比大小**。功能：通过应用右侧时间轴选定时间后应用会展示出这个时间内各个城市间机票业务（基于某航空公司机票数据），不同色彩的线代表不同航空公司，外围圆环所填充的色彩代表该地区占据主要客源的航空公司图例颜色。



固原网络舆情实时监测平台

2017-04-06 11:26:21 舆情监测点：364,154

固原舆情监测数据总量

342,070,458 条

今日新增舆情信息

635,998

今日舆情负面信息

55,998

固原舆情词云解析



固原舆情实时信息

- 1 套取资金发补贴一校长被查处 [新华网 04-27]
- 2 宁夏撤销所有政府还贷普通公路收 [新华网 04-27]
- 3 违规发放补贴!宁夏通报两起违反 [新浪网 04-27]
- 4 湖南永州市冷水滩区大力推进光伏 [资源网 04-27]
- 5 让宁夏航空运输更有“国际范儿” [新华网 04-27]
- 6 图片幸福航空走进银川开启 [新华网 04-27]
- 套取资金发补贴一校长被查处 [新华网 04-27]

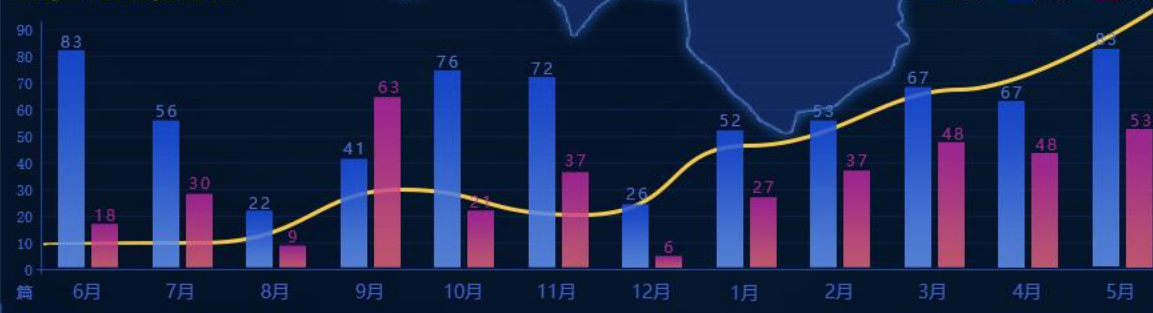
固原市舆情区域监测



固原舆情——情点监测



固原舆情数据视图



当日固原舆情预警



政府大数据：职住平衡分析

大型商务区居住地分布（金融街）

金融街上班人群中，以金融街为中心、半径十公里范围内的居住人群占69%，呈均衡的向心分布



大型居住区工作地分布（天通苑）

天通苑居住人群中，以天通苑为中心、半径十公里范围内的工作人群占39%，呈偏态失衡分布

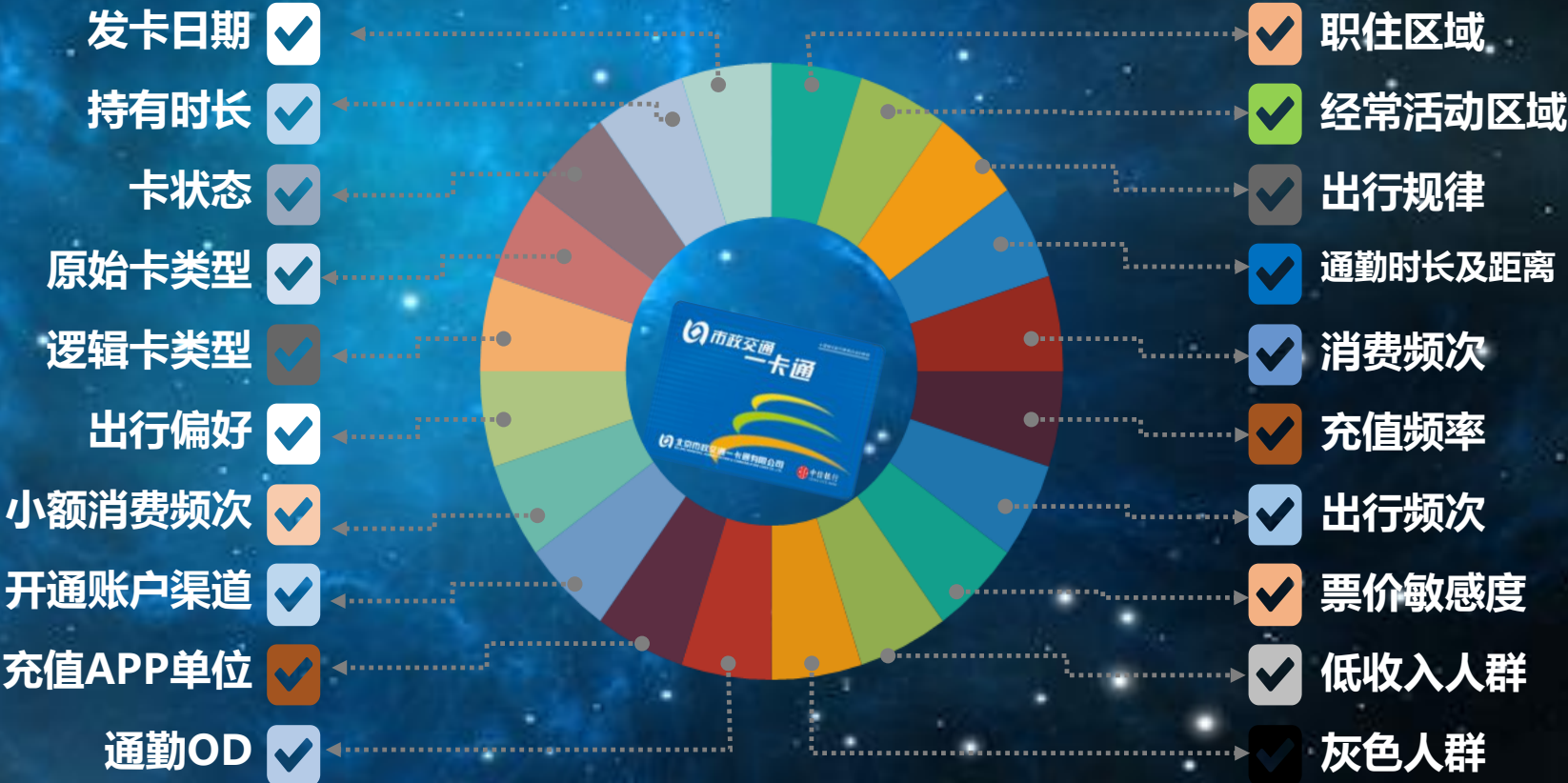


政府大数据：人群画像分析

数据治理体系---标签画像

标签应用

- 标签人群：3000万活跃用户
- 标签类别：公共交通
市政服务
外源融合
- 样本数据：70亿



政府大数据：突发事件处理



两抢一盗事件

- 对两抢一盗的重点人员进行实时监控, 防止该类人员再次犯案。结合已有的电动车防盗系统, 有效降低再案率。

聚众吸毒事件

- 对有吸毒史的人员进行实时监控, 一旦发现有吸毒人员异常聚集可主动告警, 预防案件发生

群体治安事件

- 对监控画面中人群异常聚集, 系统主动发出预警, 防止多人暴力案件的发生

固原网络舆情实时监测平台

2017-04-06 11:26:21 舆情监测点：364,154

固原舆情监测数据总量

342,070,458 条

今日新增舆情信息

635,998

今日舆情负面信息

55,998

固原舆情词云解析



固原舆情实时信息

- 1 套取资金发补贴一校长被查处 [新华网 04-27]
- 2 宁夏撤销所有政府还贷普通公路收 [新华网 04-27]
- 3 违规发放补贴!宁夏通报两起违反 [新浪网 04-27]
- 4 湖南永州市冷水滩区大力推进光伏 [资源网 04-27]
- 5 让宁夏航空运输更有“国际范儿” [新华网 04-27]
- 6 图片幸福航空走进银川开启 [新华网 04-27]
- 套取资金发补贴一校长被查处 [新华网 04-27]

固原市舆情区域监测



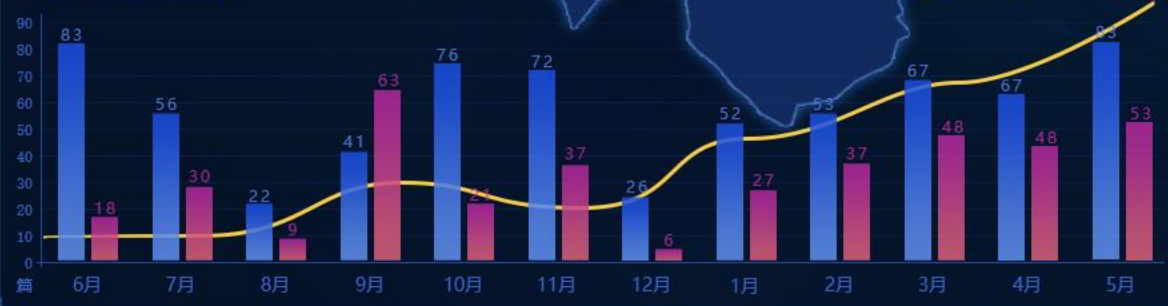
西吉县	活跃媒体
网媒舆情 (325)	人民网 (325)
论坛 (164)	铁血社区 (164)
博客 (35)	网易 (35)
微博 (24)	中国网 (24)
微信 (8)	固原市政府网 (8)
其他 (2)	新浪博客 (2)

固原舆情——情点监测

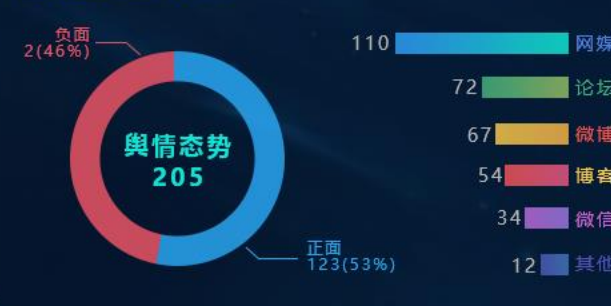


石嘴山市：3,544
银川市：13,648
吴忠市：4,925
中卫市：2,288
固原市：9,870

固原舆情数据视图



当日固原舆情预警



服务大数据：大数据与呼叫中心

每天5万通电话；5G/天大数据量的采集

语音

OA



Email

外部
舆情

语音数据到可分析文本的
解析转换(关键词85%准确率)



完成语义分析，挖掘
对客户、产品、销售的分析

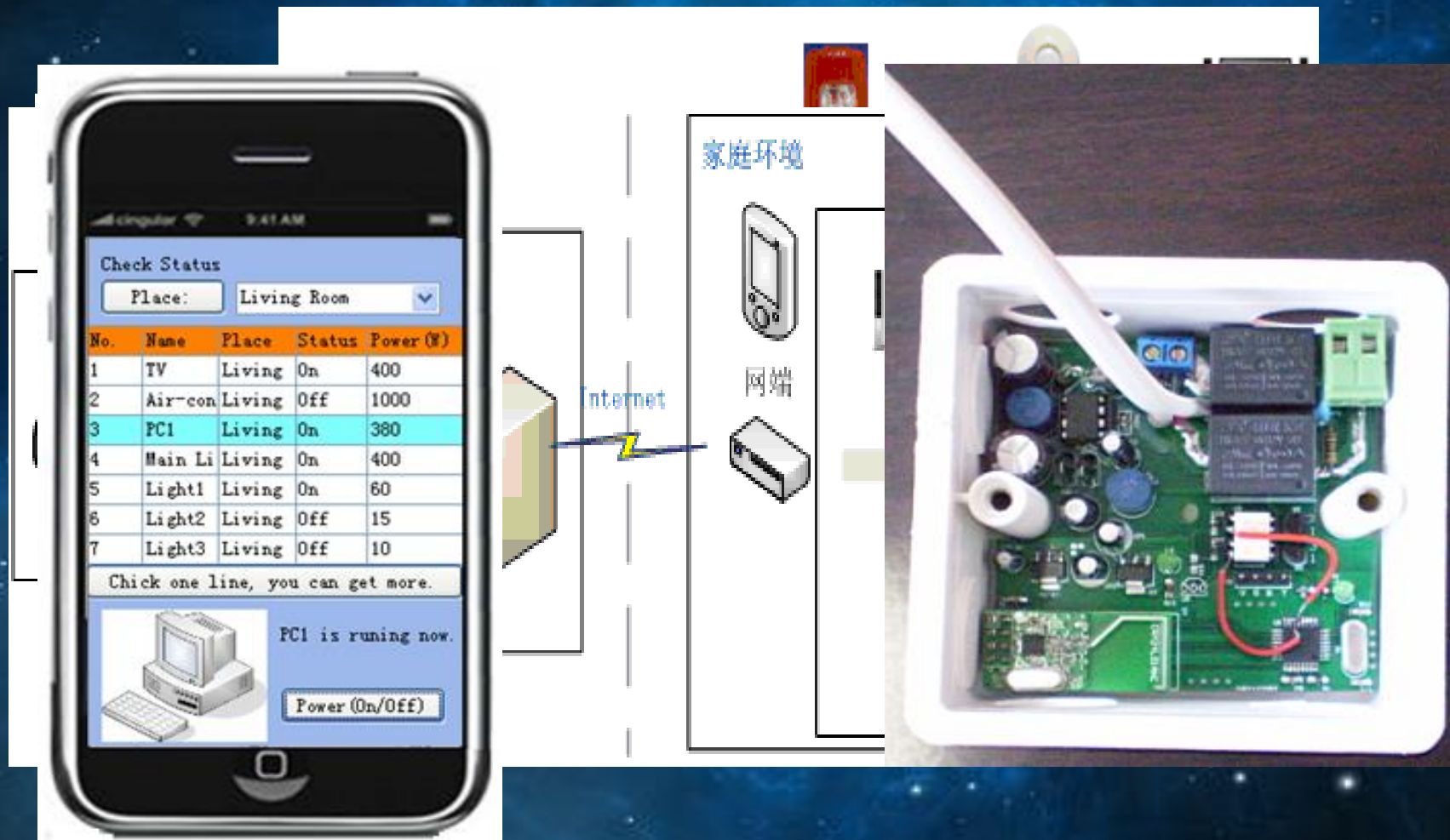
场景1：通过语义分析，提升坐席服务水平，降低作业成本，提高客户满意度

场景2：分析客户投诉情况，预测集中退保、预防群体性事件等

场景3：分析客户对销售和服务渠道喜好，挖掘针对性的销售机会和提供个性化服务

场景4：通过客户反馈，分析判断竞争对手产品、服务新动向

智能大数据：智能家居大数据管理系统



大数据广泛深度应用于日常

中央电视台对2016中国大数据技术大会进行了报道，展示了大数据技术在交通、医疗、金融等行业的深度应用。并详细介绍了应用大数据使令人头痛的套牌车无处遁形的应用案例。



广东广播电视台推出《新闻大数据》栏目

我国首个以大数据为核心的电视新闻栏目正式开播，该栏目全称《新闻大数据》，是由广东广播电视台率先推出的。据悉，该节目将不同领域的海量数据通过互联网以及专家的收集、分析处理之后，通过技术手段使其视听化和叙事化，进而创作成一种电视新闻报道的形式。一直以来，大数据给大多数“外行人”的印象都过于神秘且遥远，而该节目的播出，恰可以以一种生动、形象的方式让人们认识大数据，了解大数据，另一方面，也更加体现了大数据在当今社会的重要地位。



DaaS案例-农场主可以付费订阅这些分析情报。



John Deere直接卖数据

John Deere tractors are proprietary and the **data they collect sells to owners, operators, dealers and agricultural consultants.**

用于预测：Forecast demand correctly, crop yielding, determine land area and usage, etc.

技术：Use the open source programming [language R](#)

农业机械生产商John Deere，他为所有拖拉机配备了传感器，这些传感器会收集机械、土壤以及庄稼状况的信息，并将这些流数据传到MyJohnDeere.com与Farmsight服务上；从何时该订购备件，到哪里种植庄稼一应俱全，**农场主可以付费订阅这些分析情报。**

制造大数据：汽车制造

当问起汽车的制造过程，大多数人脑子里随即浮现的是各种生产装配流水线和制造机器。然而在福特，在产品的研发设计阶段，大数据就已经对汽车的部件和功能产生了重要影响。

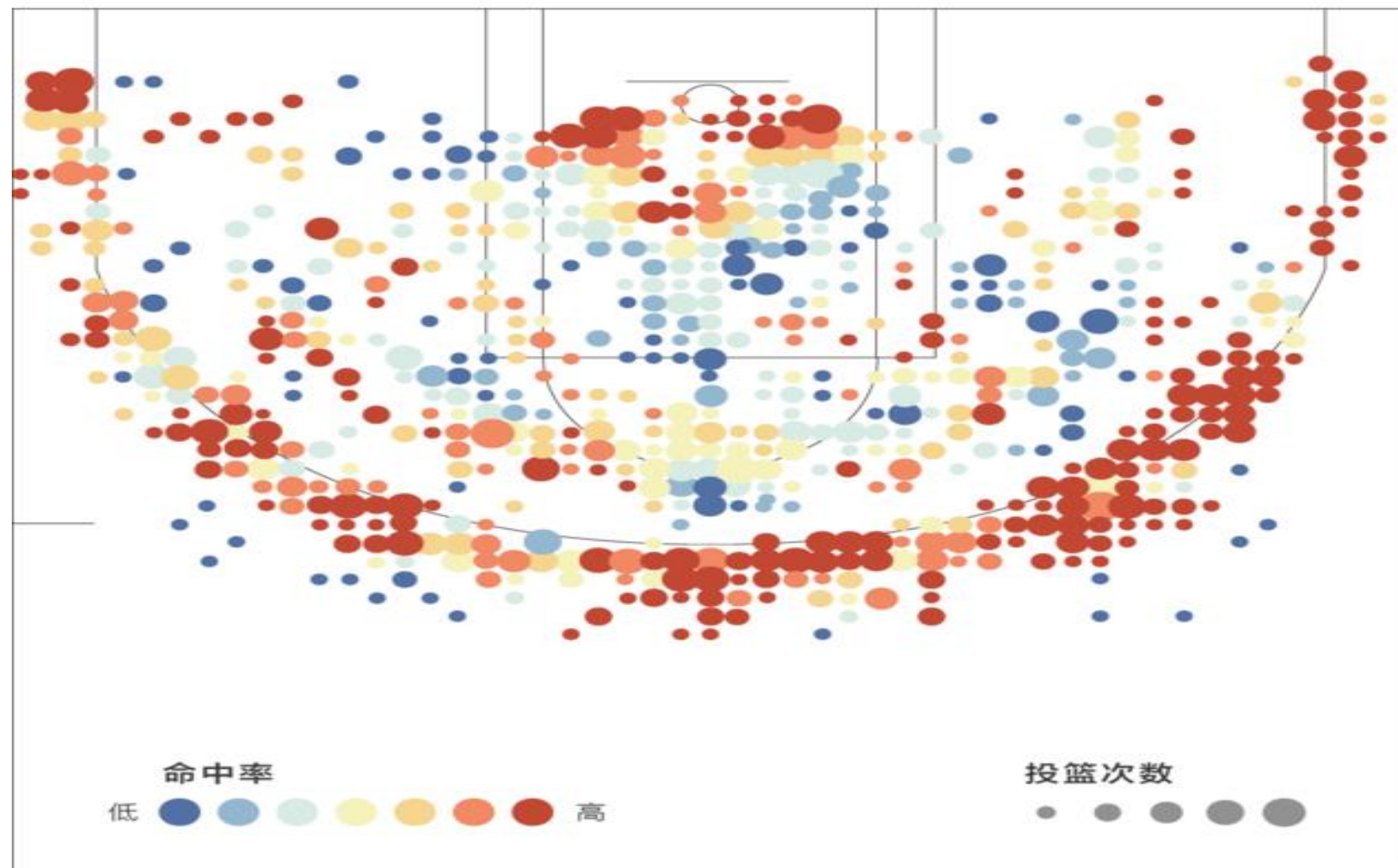
比如，福特产品开发团队曾经对SUV是否应该采取掀背式(即手动打开车后行李箱车门)或电动式进行分析。如果选择后者，门会自动打开、便捷智能，但这种方式会影响到车门开启有限的困扰。此前采用定期调查的方式并没有发现问题，但后来根据对社交媒体的关注和分析，发现很多人都在谈论这些问题。



善于三分球的球队更有可能夺得冠军。这就是大数据的魅力

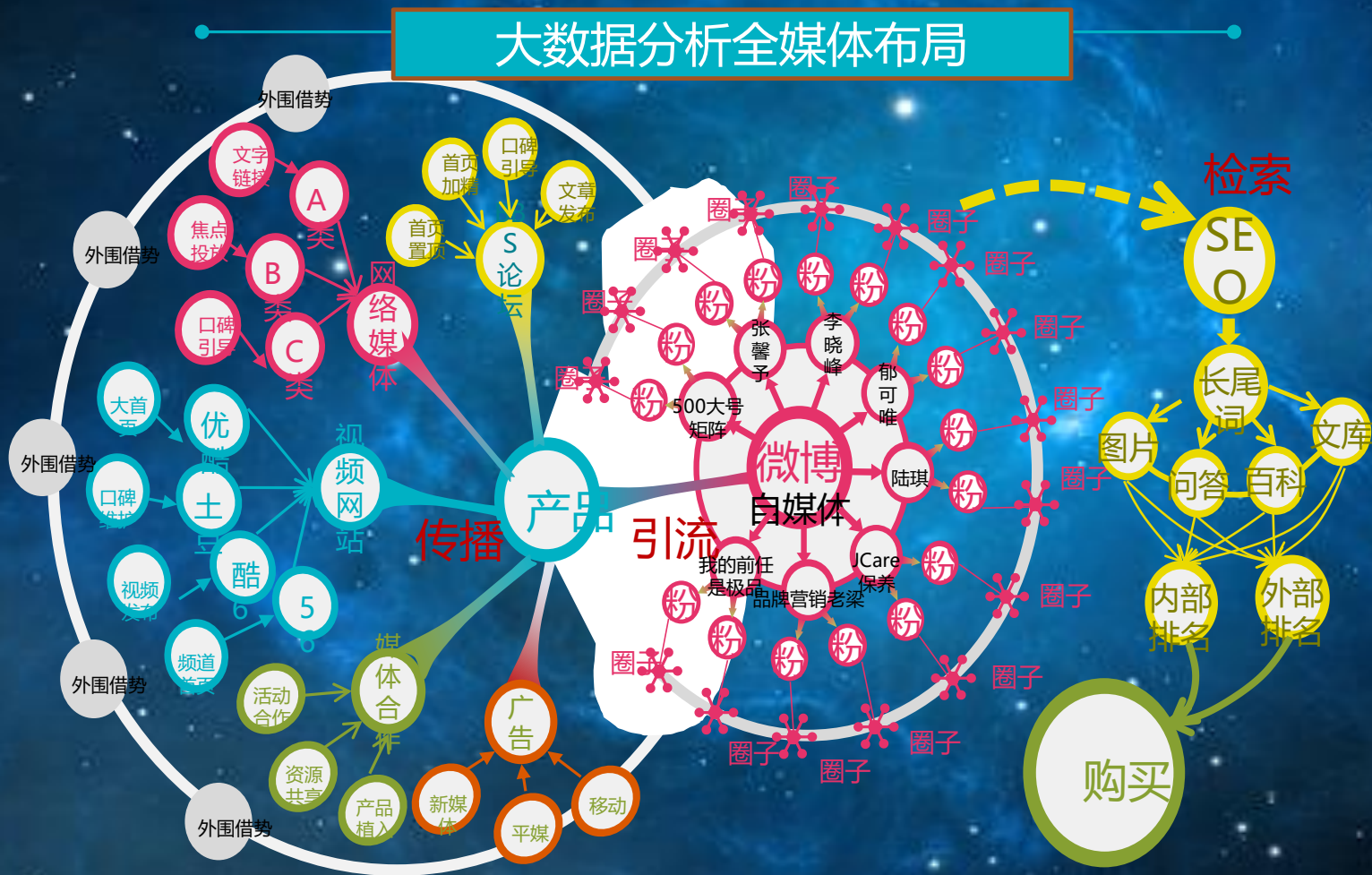
NBD 图数馆

库里13-14赛季 投篮热力图



<http://www.nbd.com.cn/articles/2016-06-20/1014344.html>

融媒大数据：大数据分析全媒体布局



交通大数据：共享汽车

目前，滴滴平台上每天产生超过50TB数据（相当于5万部电影），超过90亿次路径规划次数。在过去一年中，滴滴出行平台完成14.3亿订单，这相当于在中国平均每个人都使用滴滴打过一次车；累计行驶里程达128亿公里，相当于环绕中国行驶29万圈，累计行驶时间达4.9亿小时，相当于昼夜不歇的行驶56000年。所以滴滴“数据大脑”对弈的是现实出行的海量数据，通过对每天24小时不间断产生的新数据，以及检测这些数据本身产生的二度数据，包括ETA、路径规划、实际路线、匹配时间等等，进行研究、学习，最终实现订单匹配效率的提升，司机取得更多收入，乘客更加快捷出行。



金融大数据：Kabbage 用大数据开辟新路径

Kabbage 是一家为网店店主提供营运资金贷款服务的创业公司，总部位于美国亚特兰大，截至目前已经成功融资六千多万美元。Kabbage 的主要目标客户是 eBay、亚马逊、雅虎、Etsy、Shopify、Magento、PayPal 上的美国网商。

Kabbage 与“阿里小贷”的经营模式类似，通过查看网店店主的销售和信用记录、顾客流量、评论以及商品价格和存货等信息，来最终确定是否为他们提供贷款以及贷多少金额，贷款金额上限为4 万美元。店主可以主动在自己的 Kabbage 账户中添加新的信息，以增加获得贷款的概率。Kabbage 通过支付工具 PayPal 的支付 API 来为网店店主提供资金贷款，这种贷款资金到账的速度相当快，最快十分钟就可以搞定。

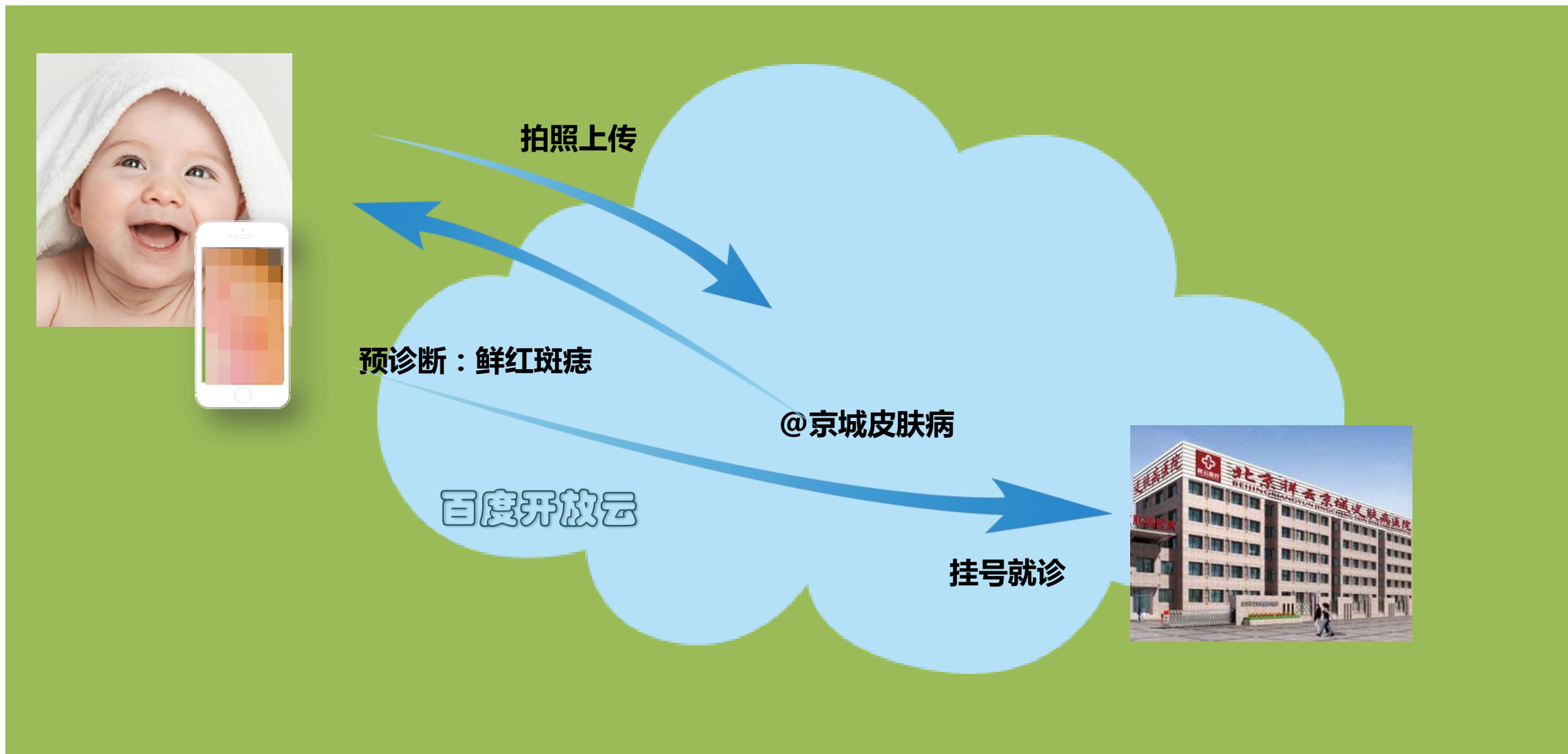
Kabbage 用于贷款判断的支撑数据的来源除了网上搜索和查看外，还来自于网上商家的自主提供，且提供的数据多少直接影响着最终的贷款情况。同时，Kabbage 也通过与物流公司 UPS、财务管理软件公司 Intuit 合作，扩充数据来源渠道。

目前，使用 Kabbage 贷款服务的网店店主已达近万家，Kabbage 的服务范围目前仅限于美国境内，不过公司打算利用这轮融资将服务拓展至其他国家。



案例解析：基于大数据的商业模式创新过程有两个核心环节：一是数据获取；二是数据的分析利用。在本案例中，Kabbage 与阿里金融的区别在于数据获取方面，前者是从多元化的渠道收集数据，后者则是借助旗下平台的数据积累，其中网上商家可自主提供数据且其数据的多少直接决定着最终的贷款额度与成本，这充分体现出大数据的资产价值，就如同传统的抵押物一样可以换取资金。

智能医疗



丰富的场景模型，自动感知和智能服务推荐

智能感知办事服务需求

市民实名登录，根据用户身份模型和业务模型特征，系统自动感知当前用户所需待办事项与服务，进行主动推荐。

孕妇族

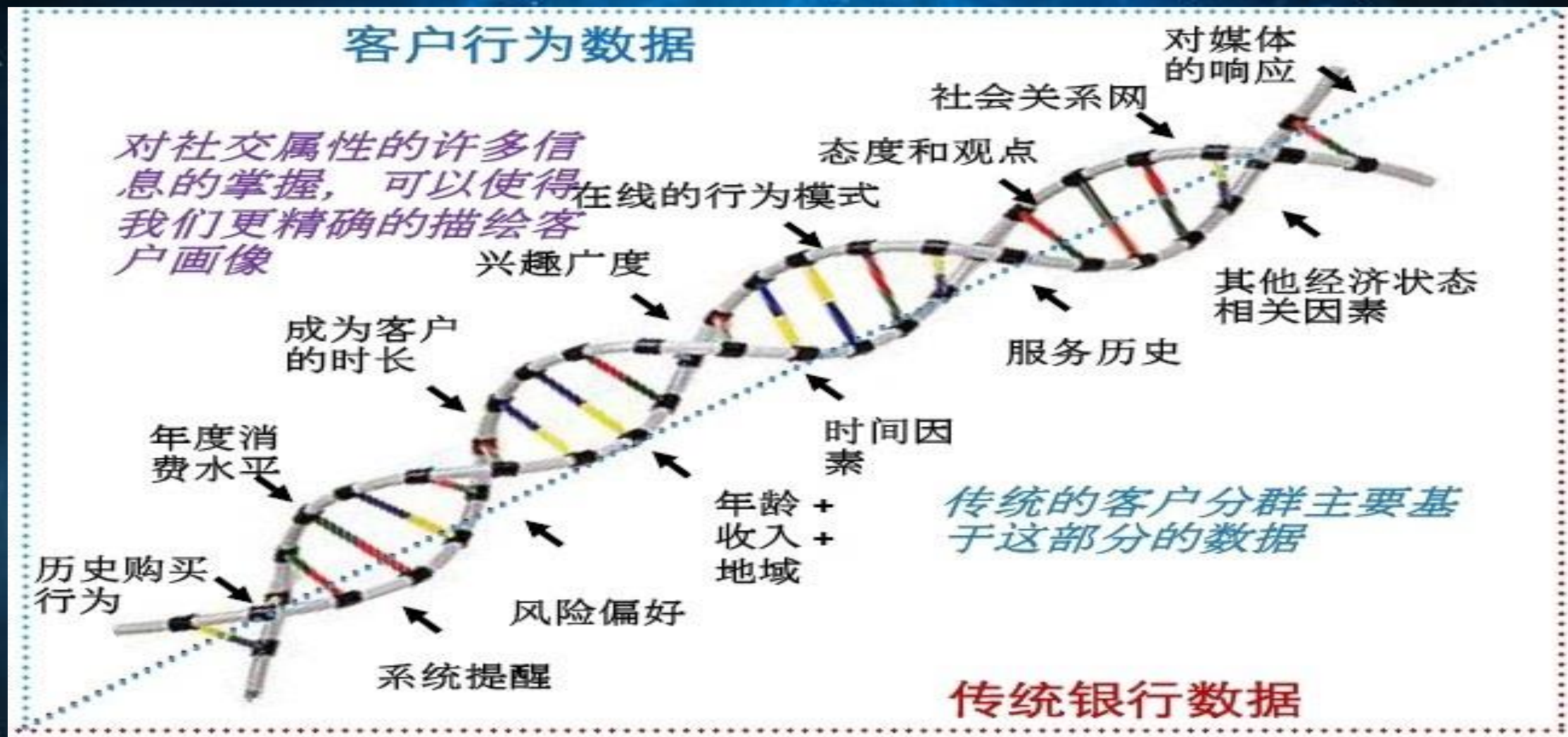
二孩准生证办事指南
产检计划管理
统一预约挂号

妈妈族

婴幼儿疫苗接种
学前教育
民生食品价格查询



金融大数据：构建客户全景视图



教育大数据：利用预测分析大数据帮助学生提高成绩

“希维塔斯学习”是一家专门聚焦于运用预测性分析、机器学习从而提高学生成绩的年轻公司。CivitasLearning提供了一套应用程序，学生和老师可以在其中规划自己的课程和安排。“希维塔斯学习”各种基于云的智能手机第三方应用程序(APP)都是用户友好型的，能够根据高校的需要个性化。这意味着高校能聚焦于各自不同的对象，相互不同地用这家公司的分析工具开展大数据工作。

该公司在高等教育领域建立起最大的跨校学习数据库。通过这些海量数据，能够看到学生的分数、出勤率、辍学率和保留率的主要趋势。通过使用100多万名学生的相关记录和700万个课程记录，这家公司的软件能够让用户探测性地知道导致辍学和学习成绩表现不良的警告性信号。此外，还允许用户发现那些导致无谓消耗的特定课程，并且看出哪些资源和干预是最成功的。



教育大数据：美国利用大数据检测学生作业完成情况

纽约州波基普西市玛丽斯特学院(MaristCollege)与商业数据分析公司Pentaho合作发起开源学术分析计划(TheOpenAcademicAnalyticsInitiative)，旨在一门新课程开始的两周内预测哪些学生可能会无法顺利完成课程。

该计划基于Pentaho的开源商业分析平台(BusinessAnalyticsPlatform)开发了一个分析模型，通过收集分析学生的学习习惯——例如点击线上阅读材料、是否在网上论坛中发言、完成作业的时长——来预测学生的学业情况、及时干预帮助问题学生，从而提升毕业率。



教育大数据：利用LinkedIn数据选择心中的大学

以美国的著名高校卡内基梅隆大学和普渡大学为例：对这两所高校，领英都收集了60000多名毕业生的职业生涯数据。数据量之庞大，足以在其中看出清晰的规律。输入

“MIT”，你很快就会看到这所高校的毕业生一般会在谷歌、IBM和甲骨文公司找到工作。输入“普渡”，你会发现礼莱、康明斯和波音是毕业生的首选。

这类信息对于中学的高年级生和低年级学生都是一座金矿，运用领英的这个工具，对太阳能、编剧、或者医疗器械感兴趣的学生，就可以挑选那些毕业生最容易进入相关领域的大学报考了。当年我们要是这款工具，可能很多人就不会错过自己喜欢的专业或学校了。



教育大数据：利用大数据来择优录取学生

美国伊萨卡学院自2007年开始收集学生的社交网络数据。该学院为申请者设立了一个类似Facebook的网站ICPEERS，让申请者得以通过网站联系学院教师和彼此。

伊萨卡使用IBM统计分析系统来收集ICPEERS上产生的数据，研究拥有怎样的网络行为的学生更有可能选择就读伊萨卡。收集的数据包括申请者上传了多少张账户照片、拥有多少名ICPEERS好友。研究人员认为，这能反映出申请者对这所学院有多感兴趣，对哪些专业感兴趣。



教育大数据：南京理工大学根据学生饭卡消费额帮学生充饭卡

每个月在食堂吃饭超过60顿、一个月总消费不足420元的，被列为受资助对象。据报道，南京理工大学教育基金会通过数据分析，每个月在食堂吃饭超过60顿、一个月总消费不足420元的，被列为受资助对象。南京理工大学还采取直接将补贴款打入学生饭卡的方式，学生无需填表申请，不用审核。



培训议题

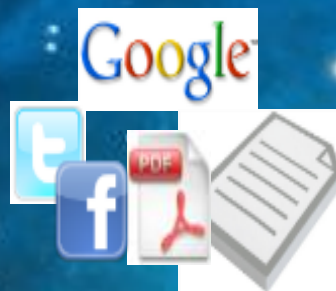
1. 大数据基础概论与大数据产业生态分析
2. 大数据在企业的应用与实践
3. 大数据产业促企业转型的作用与价值
4. 大数据政策分析和企业发展大数据的战略选择

思考：具体哪些数据可以用来赚钱，如何赚钱？



商业应用程序数据2%

- 关系型数据、高度结构化、基于僵化模式
- 财务记录、多维数据、数据计算
- 月报，非实时事件



人为产生的数据8%

- 由人与人之间的互动而产生
- 包括电子邮件、即时通信、语音、视频和文本
- 储存在集中式公司服务器、文件共享和桌面中
- 访问网页、



机器产生的数据90%

- 时间序列非结构化数据，无预定义模式
- 由所有 IT 系统生成，大量不同类型的格式
- 巨量；快速导航和相关性最重要
- 底层数据：流数据、wifi等

企业5大数据价值分析

1. 用户评论数据
2. Cookie数据
3. 微博微信的数据价值分析
4. Wifi数据价值分析
5. 第一方数据价值分析

数据获取

数据分析

数据应用

拉新客户、老客户服务等

建立完整的某种数据的价值生态

个性化推荐引擎

系统的智能大脑

什么是个性化推荐

个性化推荐就是

在**最合适**的时间、

以**最恰当**的方式、

向用户推荐他**最需要**的资讯、产品或服务

You Don't Know Jack



推荐什么?
目标意图和背景!



为什么您不需要了解Jack



基于适当的背景和意图, 我们都是先知。

**人类是背景的努力！
这是本能！**



更多可读文章: *Stanford Prison Study, On Intelligence, Outliers*

智能学习与实时更新



顾客在电子商务网站上所有行为都会被跟踪分析，实时反馈到推荐引擎中，实时改变给用户推荐的内容

商品列表

商品名称	属性	操作
 夏季新款红色连衣裙 韩版修身 显瘦大方		¥75.80元 删除 加入收藏
 夏季新款白色文胸 韩版修身 显瘦大方		¥57.00元 删除 加入收藏

购物金额小计: ¥123.80元, 比市场价 ¥147.00元 节省了 ¥23.20元 (16%)

当顾客再添加一件冬季的女性内衣时

推荐引擎会实时更新用户偏好，立刻推荐出更偏向于冬季的女性商品

基于您浏览过的商品的推荐


¥239.000
KDM 韩版小童 冬季加厚外套 韩版修身 显瘦大方 价格: 无货


¥239.000
KDM 韩版小童 冬季加厚外套 韩版修身 显瘦大方 价格: 无货


¥1430.000
STOOX 红色 黑色 KDM 韩版小童 冬季加厚外套 韩版修身 显瘦大方 价格: 无货


¥199.000
KDM 韩版小童 冬季加厚外套 韩版修身 显瘦大方 价格: 无货


¥108.000
【韩版修身显瘦】KDM 韩版小童 冬季加厚外套 韩版修身 显瘦大方 价格: 无货


¥90.000
KDM 韩版小童 冬季加厚外套 韩版修身 显瘦大方 价格: 无货

女性相

powered by 百合点

个性化是互联网发展的大势所趋

- Facebook COO 雪莉·桑德伯格断言：
未来3-5年内，所有媒体平台商城都将实现个性化！
 - 如果一个网站无法针对个别用户的兴趣进行调整，那么将不符合时代发展的潮流。
 - 人们不想看到针对全世界发布的内容，他们希望看到和知道的是能够反映自己兴趣的内容。



Facebook首席运营官雪莉·桑德伯格(腾讯科技配图)

您的网站不能输在起跑线上！

个性化推荐引擎——网站系统必备引擎

- 权威调研机构ChoiceStream 调查显示：
- 45%的用户希望访问有个性化内容推荐的网站
- 56%的用户愿意回访有个性化内容推荐的网站。
- 71%的用户喜欢个性化推荐的内容！
- 亚马逊表示：其35%的销售额来自个性化推荐引擎！

个性化推荐引擎让您赢在起跑线！

推荐引擎帮助商城-提高转化率、客单价、复购率



商城个性化商品推荐部署机制

购物流程	部署页	推荐方法	解决的重要问题	线下导购话术
注意	首页	猜你喜欢	首页跳出率比较高	欢迎光临、主动沟通
		购买历史、浏览历史	节省购买时间	您做这里、主动沟通
兴趣	导航页	个性化热销榜	快速了解用户	能帮您介绍下吗？
		个性化热栏榜	产生购买冲动	最近我们有一些新品
		浏览历史	直接购买	您以前购买这个
搜索	单品页	组合购买	减少搜索时间	产品在这里
欲望		浏览过本商品的顾客还看过商品	减少搜索时间\发掘潜在需求	您是不是也需要这个
		浏览了本商品的顾客还买过	进行交叉销售	其他客户购买了这个
		买过还买过	发掘需求、提高客单价	买过的用户还买过
		基于浏览历史的推荐	客单价提高	这样购买会更便宜
挑选	购物车	购买历史	增加购买商品的种类数和客单价	您还可以购买这个
		浏览历史	节省时间	您浏览过这个商品
		买过还买过	交叉销售	看看这个是否适合
		猜你喜欢（购买+浏览）	发掘需求、提高客单价	喜欢就多买点吧
付款	订单	基于购买历史的推荐	提醒购买、可高客单价	您忘记购买这个了
	会员中心	基于浏览历史的推荐	节省时间直接购买	您可以直接购买
		相似用户购买	发掘客户需求、增强购买信心	预计这个是更好的选择
售后服务	EDM	个性化EDM	在售后的每一个阶段	期待客户满意并再次购买

推荐算法简介

- ◆ 基于内容的推荐

- ◆ 内容的“基因”和用户的偏好相匹配，需要分析商品的各种内容信息（如文字、图片、音频、视频等）
- ◆ 和其他用户无关 举例：Pandora音乐推荐

- ◆ 基于用户行为的推荐

- ◆ 群体智慧，综合利用群体用户的行为偏好。如协同过滤、兴趣扩散等算法
- ◆ 用户之间会相互影响，更加符合现实世界中的用户行为 举例：Amazon商品推荐

- ◆ 基于SNS的推荐

- ◆ 利用人和人的直接关联进行推荐用户A在红孩子买过X类型的商品，在后玛特网上超市买过Y类型的商品，在一号店买过Z类型的商品，但是X、Y、Z是类型完全不同。用户B第一次上一号店，一号店从自身数据无法判断A的偏好。但推荐引擎知道B在红孩子买过X，在后玛特网上超市买过Y，从而认为B与A类似。由于A在一号店上买过类型Z的商品，因此也会将类型Z的商品推荐给B

例：Facebook，Twitter，Linkedin

- ◆ 会综合利用以上所有的推荐方法

从阅读客户思想中获益：

意识科技基础

算法引擎基于内容的推荐算法——文本挖掘简介

- 分析用户历史行为，为每个用户建立兴趣档案：
 - 每个用户有上百个兴趣类别，每个类别有其权重
 - 每个兴趣类别由数十个兴趣标签组成，每个标签有其权重。

UID: 内部用户编号

UID: 10002625

C1: 该用户的第一个兴趣类别及其权重

篮球: 品类及其权重

Adidas: 品牌及其权重

C1:0.369698 姚明:0.044142 火箭:0.020832 球员:0.016843 赛季:0.014822 得分:0.014604 篮球:0.012036 湖人:0.011830 7:0.011566 Adidas:0.010967
C2:0.179218 曹操:0.066031249 演义:0.015009 诸葛:0.013467 周瑜:0.011830654 董卓:0.008789
C3:0.100697 能力:0.024372 方面:0.019771 发展:0.014979 水平:0.011126 问题:0.010699 重要:0.010010 优势:0.009816 竞争:0.009755 相同:0.009429 相对:0.007367
C4:0.074551 比赛:0.043478 分钟:0.029947 球员:0.019427 进攻:0.018428 防守:0.017881 后卫:0.014467 禁区:0.013883 裁判:0.012925 前锋:0.010618 队员:0.010512
C5:0.034295 篮球:0.106530 队员:0.049118 比赛:0.048301 球场:0.040743 打球:0.029569 足球:0.017044 高尔夫:0.015253 流川:0.013408 场地:0.012288 训练:0.011663
C6:0.023639 减肥:0.148964 体重:0.040139 运动:0.024173 脂肪:0.021469 效果:0.019686 饮食:0.016506 方法:0.016500 坚持:0.015308 公斤:0.015199 健康:0.014670
C7:0.017490 经典:0.235002 一句:0.060571 语录:0.036155 鼓掌:0.024037 牙膏:0.020529 大力:0.017509 台词:0.016446 喝彩:0.014213 名言:0.008025 对白:0.007678
C8:0.014392 高手:0.128542 各位:0.113176 指点:0.077414 请教:0.072011 帮忙:0.045795 多谢:0.042004 小弟:0.033171 求助:0.029809 感激:0.029339 指教:0.022325
C9:0.013863 德国:0.120609 世界:0.061885 巴西:0.047551 阿根廷:0.040824 意大利:0.038858 荷兰:0.032905 法国:0.031137 英格兰:0.024837 西班牙:0.024512 欧洲:0.019861
C10:0.013847 驴友:0.089014 青海:0.040733 自驾游:0.023754 防滑轮胎:0.021876 切诺基:0.020793 体育:0.016367 金牌:0.016236 参加:0.014796 成绩:0.014106 训练:0.013513

麦包包数字解读

- 平均客单价的增长：体现个性化推荐引擎对收入的带动
- 平均购买商品种类数的增长：体现了个性化推荐引擎对销售多样性的带动。更深一步，体现了我们越来越能够更好地满足客户的多样化需求，或者帮助客户发掘自己隐藏的多样化需求。该数字越高，说明电商网站越人性，用户体验越好。这是互联网长尾的本质映射到的电子商务这个特定领域的体现（实体店在这方面很难做好）

麦包包数字解读

在2月17日这一周，相对于非推荐产生的客单价，为何推荐产生的客单价如此之高？

麦包包在2015-02-17这一周开始推出高端包



[JAMIE MOORE]爵士名伶系列复古牛皮两用包
棕色

麦包价: **980.00**元 编 号: 1110011701

市场价: 2850.00元

品 牌: [JAMIE MOORE]

包体大小: 40x30x18 重量: 1.57kg

颜 色:



我要买: 件 热销中, 请尽快抢购!



立即购买

加入收藏

出售: 74 件 收藏: 358 次

麦包包数字解读

女士包包

男士包包

旅行运动

功能包

价格选包

按材质分类

买过本商品的麦友还买过



¥ 398.00

百分点推荐引擎

看过本商品的麦友还看过



¥ 298.00



¥ 298.00



¥ 228.00



¥ 980.00



麦包价: **298.00**元

编 号: 1109029501

市场价: 799.00元

品 牌: [DUDU(头层牛皮)]

包体大小: 35×27×18

重量: 1.03kg

颜 色:







我要买: 件 热销中, 请尽快抢购!

立即购买

加入收藏

出售: 1310 件 收藏: 623 次

浏览: 73859 次

评分: ★★★★★ (已有39人评论)

出于对个性化推荐的信任, 浏览中低端包的顾客更倾向于从个性化推荐栏中购买高端包

查看全部大图

分享到: 

商品展示

商品参数

售前咨询

评价详情

销售记录

如何购买

售后服务

请注意: 包的颜色请选择确认后下单, 系统按订单颜色发货, 不接受留言选色!

麦包包数字解读

- 女士包包
- 男士包包
- 旅行运动
- 功能包
- 价格迭包
- 按材质分类

买过本商品的麦友还买过



¥ 190.00



¥ 980.00



¥ 168.00



¥ 980.00



¥ 220.00

百分点推荐引擎



麦包价: 138.00元 编号: 1108010324

市场价：378.00元

品 牌：[浪美]

包体大小: 38x25x14 重量: 0.79kg

颜色:



我要买: 件 **热销中, 请尽快抢购!**

 [立即购买](#)

[加入收藏](#)

出售: 203 件 收藏: 174 次

浏览: 27212 次

评分: ★★★★★ (已有2人评论)

出于对个性化推荐的信任，购买中低端包顾客更倾向于从个性化推荐栏中购买高端包

 [查看全部大图](#)

分享到:

商品展示

商品参数

售前咨询

评价详情

销售记录

如何购买

售后服务

麦包包数字解读

- 麦包包以前的商品平均价格在150元左右，2月份运营团队决定小规模尝试销售价格接近1000元的高端包。如何才能更有效、更自然地让用户接受高端包？
- 解读：
- 使用推荐栏的用户中购买高端包的比例相对全站用户中购买高端包的比例更大
- 浏览或者购买低端包的用户，更加倾向于从推荐栏中购买高端，说明推荐栏更加让用户信赖
- 猜测：倾向于使用推荐栏的用户，更加容易成为高端客户？如果这个结论正确，那么这个结论可以指导我们日后对麦包包客户进行差异化营销

个性化EDM的种类---事件触发邮件

- 弃单找回和休眠订单激活
- 订单确认和发货确认邮件
- 关联商品推荐（聚焦式推荐）
- 交叉商品推荐（发散式推荐）
- 促销活动通知邮件
- 可能感兴趣的商品促销
- 关联商品，交叉商品促销
- 帐户改动确认邮件（如密码重置等）
- 根据购买历史，可能感兴趣的交叉商品推荐

个性化EDM的种类顾客生命周期触发邮件——顾客激活

- 潜在顾客（已注册但未进行任何购买）
- 如无任何浏览历史，发送全站促销优惠券信息以及热销商品
- 如有浏览，搜索，收藏，加入购物车等行为，发送：
- 收藏夹和购物车内商品的降价，打折，优惠券通知
- 收藏夹和购物车内商品的相近商品推荐和交叉商品推荐
- 浏览和搜索的相近商品推荐

个性化EDM的种类顾客生命周期触发邮件——提升忠诚度

- 新顾客（历史较短，已有购买）
- 发送可能感兴趣门类的促销和优惠券信息
- 购买商品的关联推荐，交叉推荐
- 收藏夹和购物车内商品的降价，打折，优惠券通知
- 收藏夹和购物车内商品的相近商品推荐和交叉商品推荐
- 可能感兴趣的新品推荐

顾客生命周期触发邮件—— 定期发送

- 稳定顾客（历史较长，活跃，多次购买）
- 发送可能感兴趣门类的促销和优惠券信息
- 购买商品的关联推荐，交叉推荐，周期推荐，序列推荐
- 收藏夹和购物车内商品的降价，打折，优惠券通知
- 收藏夹和购物车内商品的相近商品推荐和交叉商品推荐
- 可能感兴趣的新品推荐
- -----流失预警
- 可能流失的顾客（历史较长，活跃度显著下降）
- 发送专门针对的代金券、优惠券
- 收藏夹和购物车内商品的降价，打折，优惠券通知
- 购买商品的关联推荐，交叉推荐，周期推荐，序列推荐

个性化EDM的种类产品周期触发邮件

- 新品
- 发送给可能感兴趣的，特别是追新的顾客（注意节奏）
- 热销商品
- 发送给可能感兴趣的，特别是追热门顾客
- 促销商品
- 发送给可能感兴趣的，价格敏感度较高的顾客
- 库存商品
- 针对有清仓需求的商品，找到可能对其感兴趣的顾客

用户评论的大数据营销



评价维、集、分一体化整合营销：

- 》通过各种Web Apps, 整合网站、电子邮件、网购渠道、微博、搜索引擎、移动客户端
- 》精准与规模结合、即时性与转化力结合、个性化与精准结合、社交与销售结合
- 》低成本高收益，而且收益持久，最重要的，被消费者需要和喜欢的感觉真好
- 》数据监测、分析、优化

建立云口碑中心



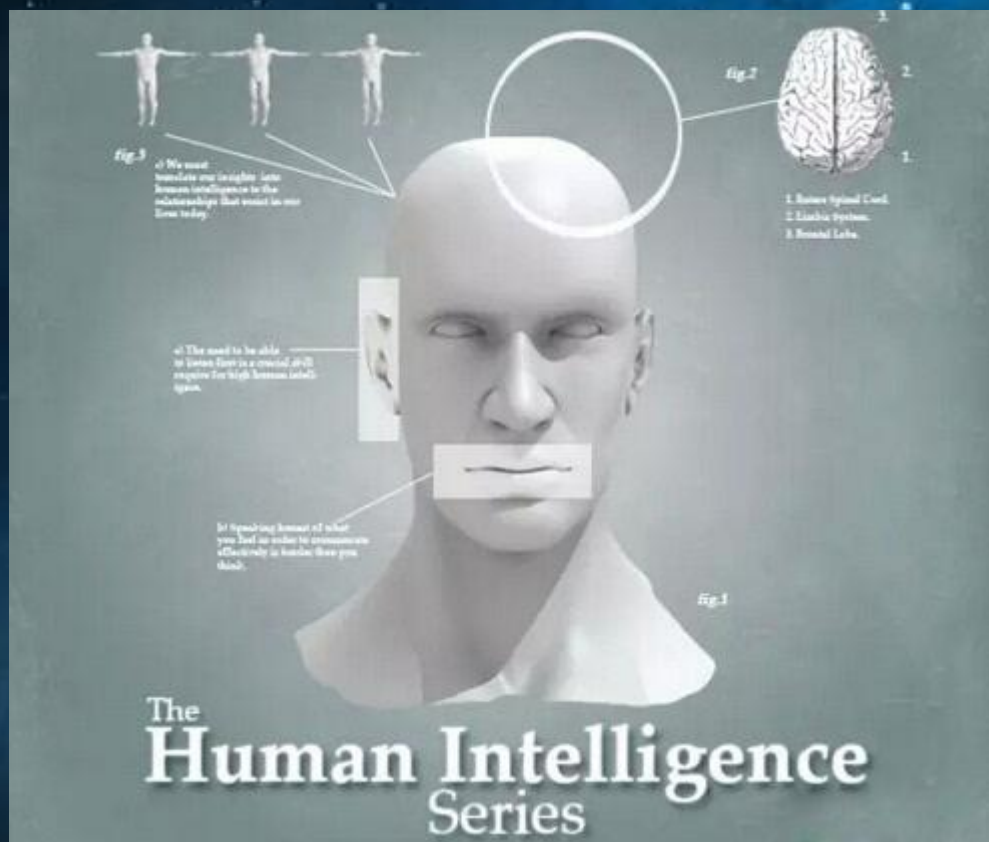
[illegible]

1	人工智能	6	专业的数据加工者咨询研究报告
2	传统互联网公司	7	专业的数据数据营销者
3	数据征信评价机构	8	大数据工具的生产者
4	基因大数据指导生命科学	9	数据开放平台API经济
5	专业领域的数据共享者	10	数据交易平台

数据变现独角兽可能的10种商业模式

一、人工智能，数据+物体=智能（1/2）

人工智能是数据变现的最好方式，但是目前是2B的智能买单意愿更强 **GDP*20%**



- 1) 数据是为人服务的，人接触最多的是物体；
- 2) 数据的智慧将延伸人的五官，拓展人的四肢，这些都依赖硬件；
- 3) 数据作为一种软资源，必须借助物体才能更好的发挥它的价值；
- 4) 物体是数据的最佳载体；

从国内外的互联巨头的投资动向不难看出，传统的盈利的大数据公司开始涉足硬件市场，利用其固有的软件技术整合硬件厂商快速的占据市场的有利位置。硬件是连接线上与线下的重要组成手段。所以笔者以为智能硬件这才是大数据正在的用武之地，才是大数据最终的价值所在！

数据变现独角兽可能的10种商业模式

一、人工智能，数据+物体=智能（2/2）

人工智能是数据变现的最好方式，但是目前是2B的智能买单意愿更强 **GDP*20%**

所以随着大数据所涉及的数据采集、数据管理、数据分析等技术的发展：

- 1) 未来，所有物体都会拥有智能。
- 2) 未来，所有的物体都会成为类人脑；
- 3) 未来，所有的物体都会联网；
- 4) 未来，所有的物体会相互制约发展，不是以原始生态制约，而是以商业制约；
- 5) 未来，物体和人的对话将无处不在；



数据变现独角兽可能的10种商业模式

二、互联网巨头（BATFGW）

搜索、信息展示、社交-广告 **万亿左右市值**



用大数据与大故事创造共鸣，那么，如何才能创造有共鸣的解决方案？品牌必须选择与消费者站在一起，融入后者的生活方式。从这层意义上讲，在大数据的技术路径之上所进行的品牌推广已不再是“广告”，而是生活“内容”。广告人不是在创造广告，而是在创造生活本身，如果不能从这一点上去建立自己的认知，这个行业最终会失去自身存在的价值。

移动社交的基础是共同的话题与兴趣。“共鸣”成为了社交媒体广告沟通的入口。传统意义上仅通过细分市场来建立品牌区隔的方法已很难奏效。在大数据的技术路径之上所进行的品牌推广已不再是“广告”，而是生活“内容”。要做到这一点，大数据与大故事(Big Story)皆不可或缺。

BATFGW市值（截止2016.9.7）



602.85亿美元



2382.92亿美元



220390.44亿港币



3600.56亿美元



3570.23亿美元

2000年是数字广告发展的一个分水岭。

那一年，Google推出了名为AdWords的搜索广告产品，这项业务连同2004年推出的Adsense一起，终结了以Yahoo等门户网站为代表的第一代数字广告模式。

数据变现独角兽可能的10种商业模式

三、数据征信评价机构（1/2）

通过数据加快贷款、通过数据降低风险 万亿左右市值

BAT巨头纷纷进入大数据征信市场，也正是看中了这一千亿级的蓝海市场。据平安证券估计：中国征信行业未来市场规模将达千亿元，其中企业征信市场规模有百亿元，个人征信市场规模有千亿元。有着国企背景的中诚征信则更加progressive，给出了未来市场过万亿的预期。

美国征信市场由传统征信机构、商业信息服务机构、创新型的金融科技企业三种力量组成。

- 传统征信机构以全球最大的个人征信机构Experian、全球第二大征信机构Equifax、征信数据挖掘公司FICO为代表，基于掌握的消费者和支付数据提供征信服务。
- 商业信息服务机构Dun & Bradstreet以庞大的全球商业数据库-全世界最大的企业信用数据库知名，基于其全球化的发展战略，主推风险管理服务（贡献营收62.7%）和销售及市场拓展（37.3%），利用征信业务的规模经济获取高毛利率。
- 创新金融企业Zestfinance则以技术输出为主要手段，利用传统的信贷记录等数据、大量交易信息、法律记录、租赁信息、网购信息等数据（第三方、网络、调研），使用机器学习的大数据分析模型进行信用评估，取得不错的实效，将信贷的成本降低了25%。

数据变现独角兽可能的10种商业模式

三、数据征信评价机构（2/2）

通过数据加快贷款、通过数据降低风险 **万亿左右市值**

个人征信和企业征信的模式区别

个人征信和企业征信因其目标客户、数据来源、数据产出形式的差异，而具有不同的商业模式：

个人征信：服务面向个人借款需求的互联网P2P平台为主，通过基本的征信报告服务和基于互联网技术的个人信用综合解决方案,如贷款决策系统外包、信用卡实时信用分析数据传送等盈利;

企业征信：定位于消费型企业及政府公共管理市场，服务于从事融资业务的金融机构，通过向机构输出报告、提供信用担保等服务盈利。

国内征信机构已形成重资产模式、轻资产模式、数据整合与评分模型模式，以及互联网模式等几种商业模式。各机构的模式分析如下：

芝麻信用 【一分多用】	<ul style="list-style-type: none">模式：互联网模式公测阶段：和支付宝钱包/获得用户授权的租车、租房、酒店等多个领域商户网页和 APP 合作，提供体验性服务对用户暂时不收费，未来的信用评级验证后才会考虑收费方案可能的收费方案：查询收费-与查询量和数据质量挂钩（商户使用越频繁，数据回流越多，对于“芝麻分”和其他信用产品的有效性质数据就越大，收费就会更便宜）未来：由商业场景拓展金融服务，和外部数据双向流动，形成信用商业生态，如花呗、借呗、微贷；信用住计划，近 5500 家酒店加入
腾讯征信 【腾讯信用 （一分多用）】	<ul style="list-style-type: none">模式：互联网模式2B：机构接入腾讯征信，把社交数据放入信用评估模型；账户级及账户产品：面向银行、证券、保险、消费金融、小贷、p2p 等商业机构，已经开始接入合作机构2C：通过公众号查询信用报告记录和评分；推出实名认证、信用产品等技术创新：人脸识别判定真实身份对个人免费，对机构收费未来：金融生态-虚拟信用卡、小额贷款，如京东白条
深圳前海征信中心 【好信通 （一分多用）】	<ul style="list-style-type: none">全牌照金融企业，依托平安旗下保险、信托等业务链数据用户：前期为金融机构和部分个人，主要切入借贷场景目前：并不将征信公司定位为牟利手段，而作为综合金融的组成部分，是综合金融的基础建设未来：商业模式尚不明确
鹏元征信 【鹏元 800（一分多用）、深圳市个人信用征信系统】	<ul style="list-style-type: none">中国最早成立的商业征信机构之一，深圳市政府背书模式：重资产模式代表2B 主要：政府、银行、小额贷款公司、公用事业单位、电商平台等机构2B 其他：政府资质评定、人事落户、银行信用卡发放、贷款审批、电话开户、求职招聘、投资担保、典当融资
中诚信证 【2C 万象分 （一分多用）】	<ul style="list-style-type: none">国企背景模式：重资产模式代表前期：暂时只提供查询个人“万象分”服务；业务重点是信贷机构；联合 P2P、小贷公司、电商以及银行成立互联网征信联盟，共享信息未来：征信报告、风险信息查询、互联网身份验证服务收费：不同产品或服务组合阶段定价；联盟会员部分免费服务，超出条件或免费范畴的，给予会员优惠折扣（阶段定价机制）；定制化服务盈利情况：征信业务从 2007 年以来就持续增长和盈利，每年 5000 多万元收入，1000 多万元净利润，其中约分之一来自个人征信，其余来自企业征信
中智诚征信	<ul style="list-style-type: none">民营第三方征信公司主做民放征信（一分多用）数据整合与评分模型模式的代表创始人及团队投身中国征信行业十年之久
拉卡拉信用管理 【考拉信用分 考拉商户分 （一分多用）】	<ul style="list-style-type: none">联想系，第三方支付公司2B：信用分接入拉卡拉平台，面向小微商户申请贷款（POS 贷）2C：信用分接入拉卡拉平台未来：与政府、金融机构、电商企业等展开合作
北京华信征信 【2C：随随分 （零分专用）】	<ul style="list-style-type: none">前期：检验租房者信用情况未来：租房市场

数据变现独角兽可能的10种商业模式

四、基因大数据指导生命科学

通过数据加快贷款、通过数据降低风险 万亿左右市值

目前华大基因净利润在1亿元左右，不过深圳不少基金经理认为，作为基因测序的龙头，华大基因上市估值可能一步到位，其市值或直接到1000亿左右。华大基因的招股说明书显示，2015年上半年归属于母公司的净利润为7565万元，2014年度，2013年度的净利润分别为2456万元、13588万元。

生命经济的发展才是未来：面向人类最根本需求的经济形态和创新会是最大趋势。实现从后工业时代到生命经济时代的转变，需要大众转变观点、政策扶持以及科研机构的多方推动。未来，以国家基因库作为支点，围绕生命科学发展的产业，会走入从科学研究到产业化的发展之路，最终实现为人类服务的目标。

在未来社会发展上，影响人类社会经济和生命质量上有三个重大的问题。一是出生缺陷，二是代谢性疾病和心脑血管，三是肿瘤。这三个疾病导致人类医疗费用的支出70%到80%，而这三个疾病的防控唯一的办法就是用现代科技和大数据的支撑才能够解决这样的问题。

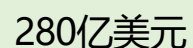
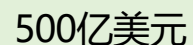
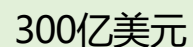
我们依靠基因科学技术，产生的大数据来引领着未来的大发展，来支撑着小康社会建设，以一个前所未有的高科技来作为支撑和引领我们一定能在某些领域走在世界前沿。

五、专业领域的数据共享者（1/2）

2000亿的市值

快递业、家政服务业、教育行业、培训业、个人服务行业、新闻业、
租赁业、广告&创业业、医疗业、色情业

AUD估值 (截止2016.9.7)



数据变现独角兽可能的10种商业模式

六、专业领域的数据共享者（2/2）

共享经济企业：企业共享、房屋共享、技能共享、办公空间共享

2000亿的市值

在过去几年间，对全球范围近几百家共享经济企业进行深入研究，发现成功的共享经济企业的商业模式运作在以下六个方面有明显优势：

- 挖掘充裕而稀缺的资源
- 激发网络效应的平台
- 突破引爆点的用户
- 建设共情的社群
- 维护基于信任的秩序
- 满足供需高效的匹配

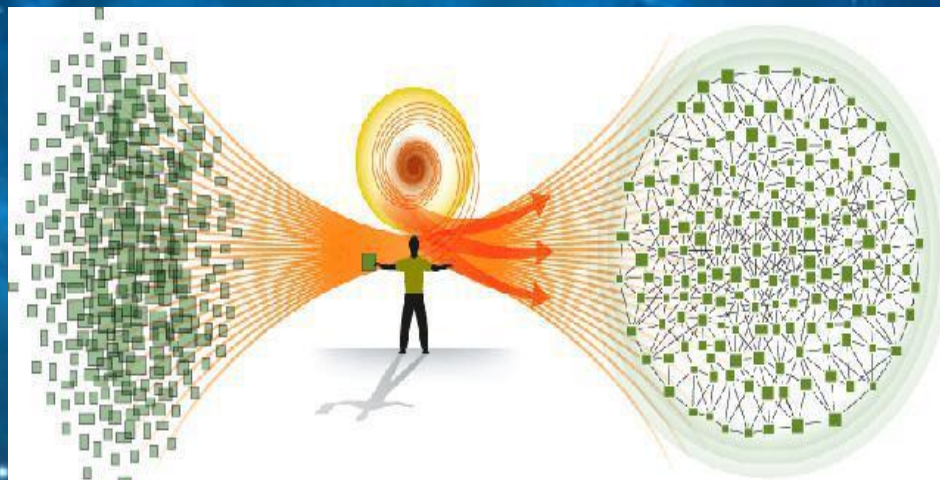


数据变现独角兽可能的10种商业模式

六、专业的数据加工者咨询研究报告

汤森路透、万德、尼尔森、艾瑞、易观、中经社 1000亿的市值

汤森路透得总裁吉姆·史密斯说：大数据对汤森路透非常重要。从某些方面来说，我们已经长期在管理大型复杂的数据了。我们面临的挑战与其他大型科技公司不一样，过去近25年里，我们一直在管理和整合我们所服务的不同行业领域的各类数据。我们投入了大量的资金来整合众多的数据，集成数据库，让客户可以简单地掌握和搜索所需要的数据资料，而不必再花时间去了解来源或复杂性。



市值（截止2016.9.7）



7.45亿美元



600亿人民币



192.52亿美元



暂无



4.5亿人民币

万德数据服务（Datafeed）这样描述自己：中国市场的精准金融数据服务供应商，为量化投资与各类金融业务系统提供准确、及时、完整的落地数据，内容涵盖：股票、债券、基金、衍生品、指数、宏观行业等各类金融市场数据，助您运筹帷幄，决胜千里

为客户提供标准的结构化数据，支持模块化订阅，同时满足客户个性定制需求，实现合作伙伴式的落地数据服务。

数据变现独角兽可能的10种商业模式

七、专业的数据数据营销者（1/2）

精准营销DSP+短信、email、私信（暴力广告，获得线索，客单价较高的产品，如地产） **500亿的市值**

DSP产业链

- DSP行业产业链上的角色包括广告主、广告代理商、DSP、广告交易平台、DMP、SSP、广告网络、广告联盟、媒体以及受众。
- 广告主或代理商通过DSP进行投放，DSP帮助广告主或代理商通过搜索引擎、广告网络以及广告联盟进行投放，同时DSP可以接入多个广告交易平台或可以接入多个SSP来获取媒体受众资源，而广告主则通过DSP对广告交易平台中的流量进行基于受众的购买。

广告主



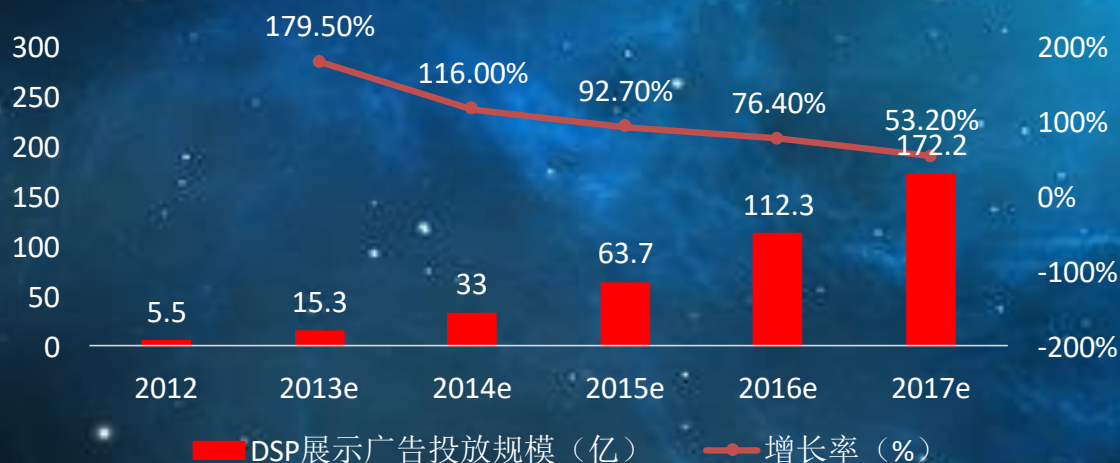
受众

数据变现独角兽可能的10种商业模式

七、专业的数据数据营销者（2/2）

- 2012年是中国DSP发展的元年，经过过去3年多的酝酿，去年出现了大量的DSP服务商和技术提供商。并且在一些巨头的广告交易平台的推出影响下，DSP所能够投放的广告的量迅速增长。2013年更多的广告平台出现、更多的媒体接入这些平台，同时提升了广告供给量、刺激了广告主的兴趣，市场获得非常高速的增长。在市场上RTB的购买方式是主流。另外，移动端的DSP初露端倪，未来极具成长空间。
- 能够为广告主、代理公司提供全面服务的服务商，有艾维邑动、爱点击、璧合网络、传漾、好耶、互动通、晶赞科技、聚胜万合、派择、派瑞威行、品友互动、随视传媒、泰一指尚、新数网络、亿玛、亿赞普、易传媒、悠易互通等。

2012-2017年中国DSP展示类网络广告
投放市场规模



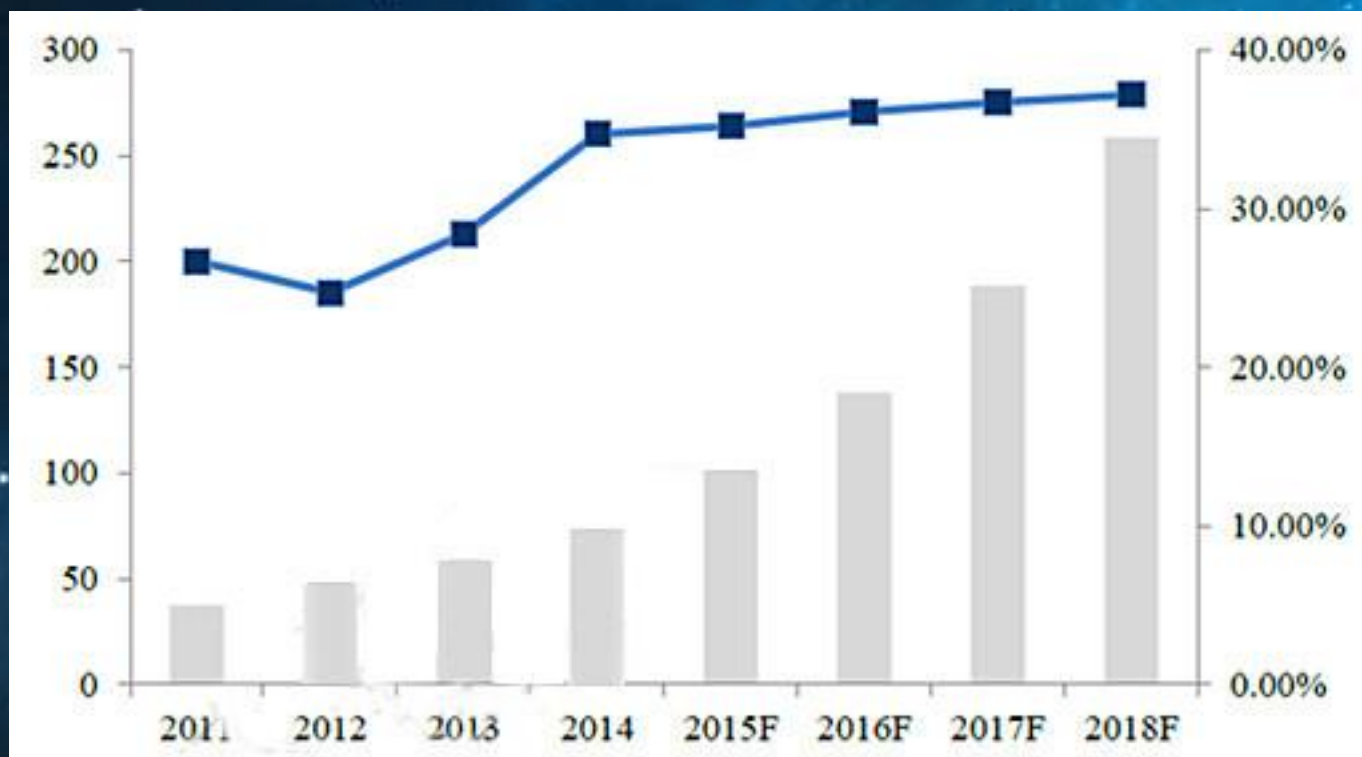
2012-2017年中国DSP展示类网络广告
投放市场规模



数据变现独角兽可能的10种商业模式

八、大数据工具制造商Taradata\SAS\Hadoop\Hana\Wason

第一方数据的加工利用绝对是最好的数据变现间接方式 **千亿的市值**



大数据在全球范围内的市场规模同样巨大，根据IDC发布最新研究结果，预测到2018年全球大数据技术和服务市场的2018年的复合年增长率将达到26.4%，规模达到415亿美元，是整个IT市场增幅的6倍。从行业结构来看，大数据应用主要集中在金融、通信、销售和政府领域，在医疗和旅游行业也有应用，但占比相对较低。

数据变现独角兽可能的10种商业模式

九、数据开放平台，API经济模式

如Salesforce，AWS，百度数据开放平台、腾讯数据开放平台等 **百亿左右市值**

开放类型	开放APP	开放login	开放数据源
说明	引入第三方应用在自己的平台上，赚的更多的用户和流量，还可以和第三方分成	降低用户进入第三方的注册门槛，同时也能圈住自己的核心用户群	引入第三方优质数据源填充自己的搜索结果，会做一定的展现方式优化，第三方要流量，平台要数据，
代表型企业			

BAT开放平台的特点

- 一、腾讯的开放是产品层面的开放，核心资源不可能开放
- 二、百度的开放是技术层面的开放，过度开放，对百度来说是风险
- 三、阿里的开放是产业链的开放，但生态的封闭

数据变现独角兽可能的10种商业模式

十、数据交易市场-中国的虚假繁荣（1/3）

B2B大数据交易所 几十亿

1. 个截止目前，中国总共有三个大数据交易平台，三个大数据交易所，十个大数据交易中心：

中关村数海大数据交易平台，数据堂，聚合数据、贵阳大数据交易所，武汉长江大数据交易所，湖北华中大数据交易所

贵阳现代农业大数据交易中心，武汉东湖大数据交易中心，华东江苏大数据交易中心，杭州钱塘大数据交易中心，浙江大数据交易中心，河北数据交易中心，上海数据交易中心，哈尔滨数据交易中心，淮南数据交易中心，广州大数据交易中心

2. 这些交易所未来会成为数据交易的主要场所，前景很光明，但是路也很遥远，这依赖技术和法律的成熟。

数据变现独角兽可能的10种商业模式

十、数据交易市场 – 问题（2/3）



数据变现独角兽可能的10种商业模式

十、数据交易市场 – 行业数据运营商（3/3）



- 数据源：专注在IT能力比较弱，但是数据价值较高的行业客户上；
- 数据云化：强调数据管理的云化；
- 数据产品SaaS化：为可以接受SaaS服务的客户提供数据产品

数据变现从数据资产登记开始

数据资产应该以金融的手段开始，不打破原来数据的流通价值，不参与企业原有IT系统的竞争，用登记，买卖的方式促进数据的合法合规流通。



培训议题

1. 大数据基础概论与大数据产业生态分析
2. 大数据在企业的应用与实践
3. 大数据产业促企业转型的作用与价值
4. 大数据政策分析和企业发展大数据的战略选择

趋势：大数据基因植入传统企业

A green rectangular road sign with rounded corners and a white border of reflective dots. The word "Change" is written in large, white, sans-serif capital letters. The sign is mounted on two wooden posts. The background is a bright blue sky with scattered white clouds.

Change

大数据冲击传统市场的表现

- 1、卖产品产品转变为卖服务 企业自身的转型 粘性的增强
- 2、服从管理转为创造客户价值 组织的再造 可持续
- 3、互联网思维、数据思维、消费者驱动 商业模式的核心思维

大数据最大的机会在于具有互联网基因的传统企业

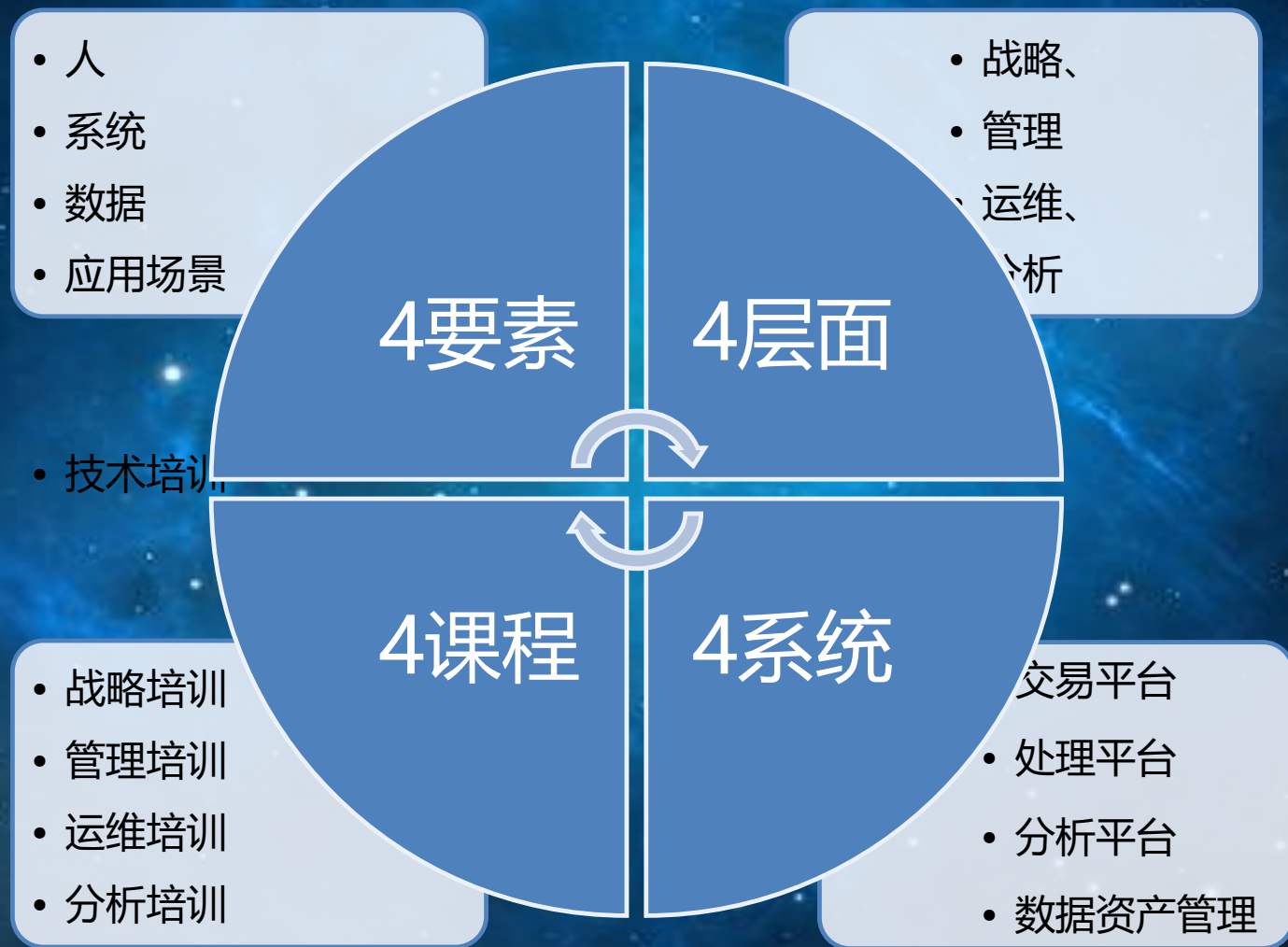
大数据基因植入更多企业

- 1、平台型企业 按客户需求聚散资源 企业无边界
- 2、微企业 自主经营体 管理无领导
- 3、产品按需定制 持续创造客户需求 供应链无尺度
- 4、数据驱动 资源共享 业务融合
- 5、大数据驱动分享经济 所有的产品都可以共享 除了老婆孩子等亲人

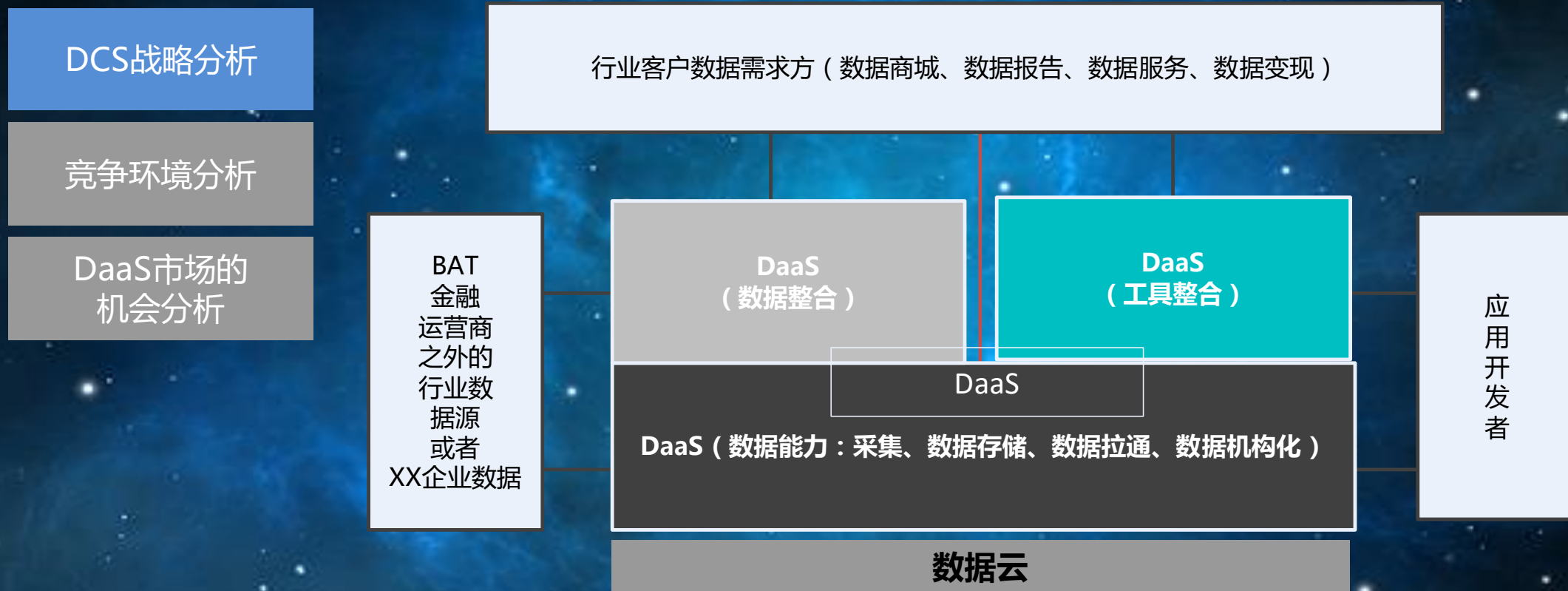
培训议题

1. 大数据基础概论与大数据产业生态分析
2. 大数据在企业的应用与实践
3. 大数据产业促企业转型的作用与价值
4. 大数据政策分析和企业发展大数据的战略选择

需要建立4V一体的咨询体系



DCS战略：数据源优先战略、数据源云化战略和数据服务SaaS化战略



- 数据源：专注在IT能力比较弱，但是数据价值较高的行业客户上；
- 数据云化：强调数据管理的云化；
- 数据产品SaaS化：为可以接受SaaS服务的客户提供数据产品

渗透在大数据的整个生命周期咨询方案



整个企业业务生命周期的大数据咨询

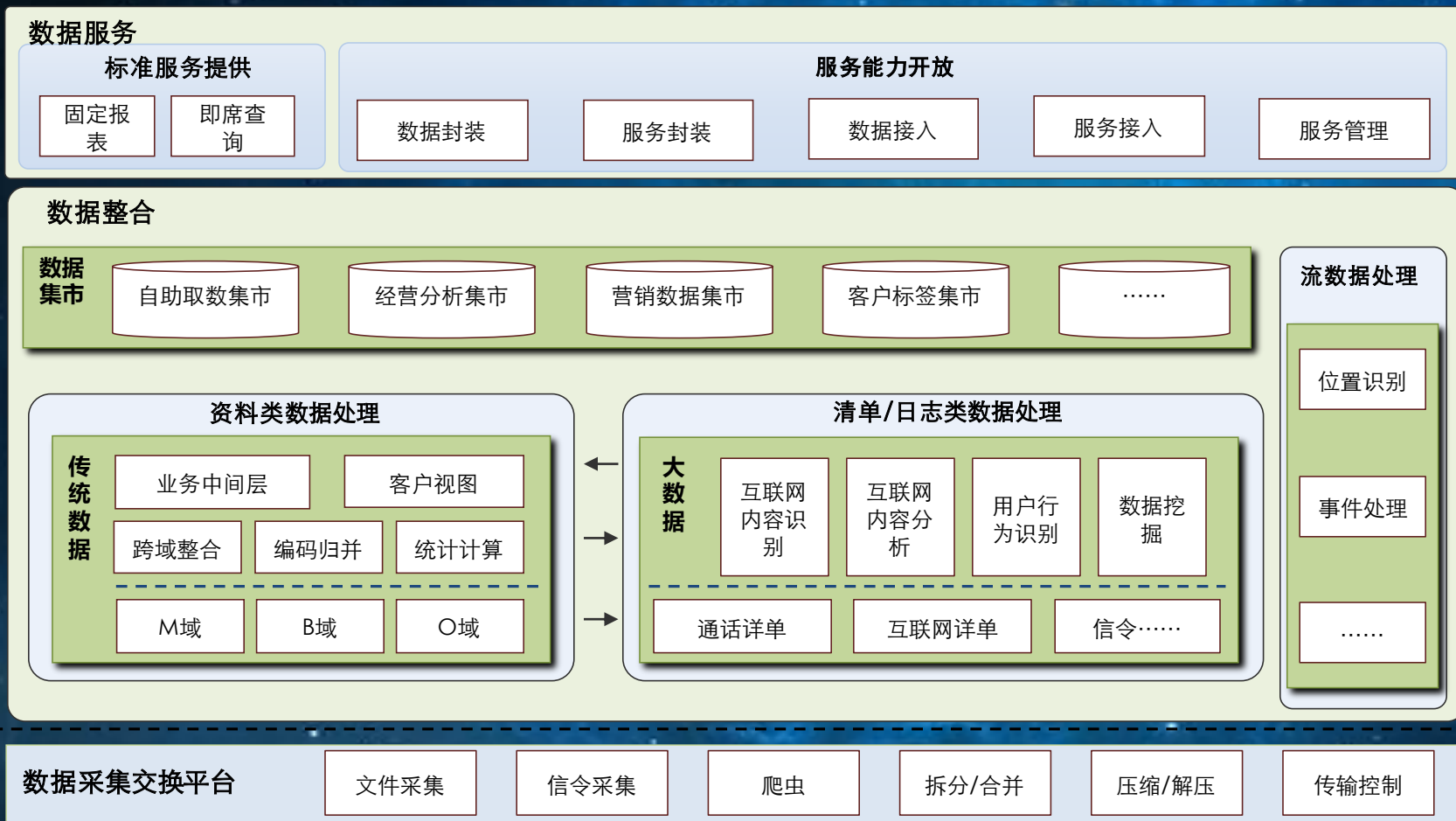


The background of the image is a deep space scene. It features a large, vibrant blue nebula with wispy, filamentary structures that glow with varying intensities of blue. Scattered throughout the dark, black void are numerous white stars of different sizes and brightnesses, some appearing as sharp points of light while others are slightly blurred.

系统

大数据服务体系架构

数据整合服务平台



□ 大数据服务体系由数据采集交换平台和数据整合服务平台组成。

□ 数据采集交换平台：负责对源数据进行清洗转换、对互联网日志进行预处理，以及提供各系统之间的数据交换功能。

□ 数据整合服务平台：源数据按照数据模型进行加工，生成应用层数据，供应用平台使用，并对外部系统提供多种数据服务方式，实现数据的共享。

数据源

业务支撑系统
BSS/MSS/OSS

业务平台

通信网络
信令/事件/日志...

互联网信息
内容/用户行为...

外部合作伙伴
数据/知识...



场景

创收：业务类大数据应用

垂直化应用

数据交易						
移动开发者服务						
商业WIFI						
语音智能识别						
互联网舆情						
WEB流量分析						
业务性能管理						
智慧城市						
地理位置数据应用						
精准营销						
信息安全						

节省成本：分析决策类应用

系统巡
检

智能
运维

安全合
规

舆情

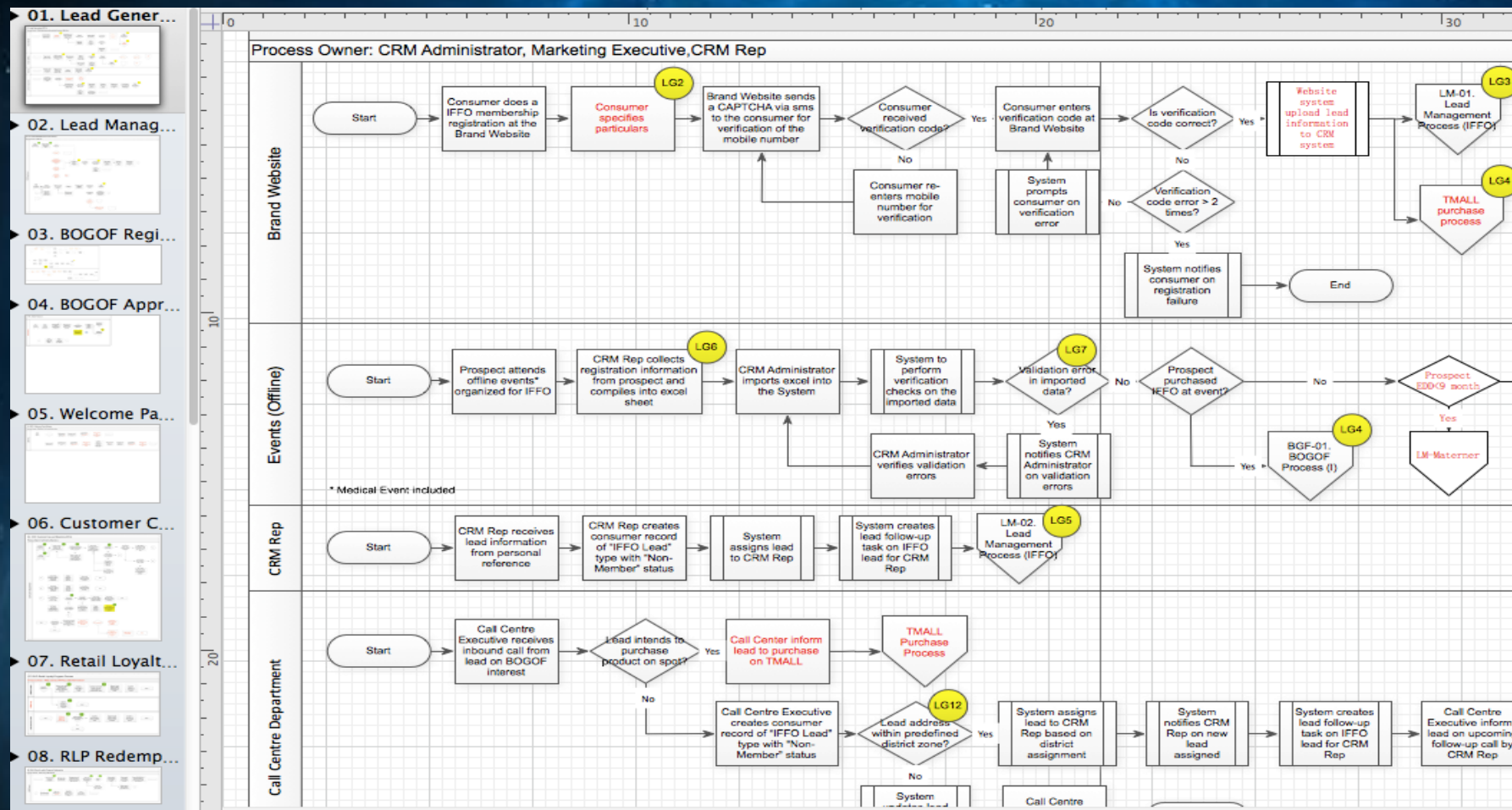
业务决
策分析

Web 流
量
分析

数据对企业KPI的的价值

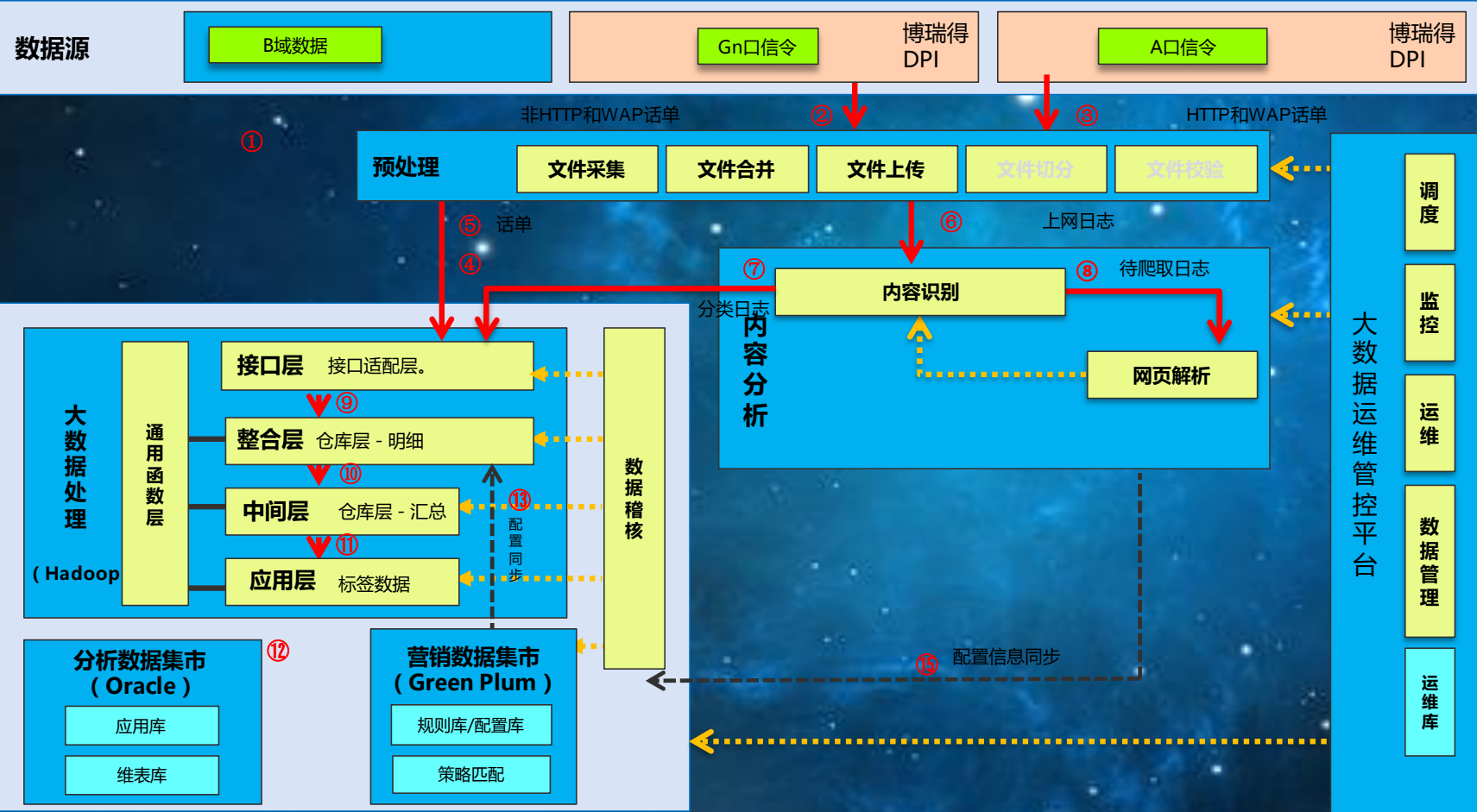


对内我们和客户的接触是如何管理的？



数据

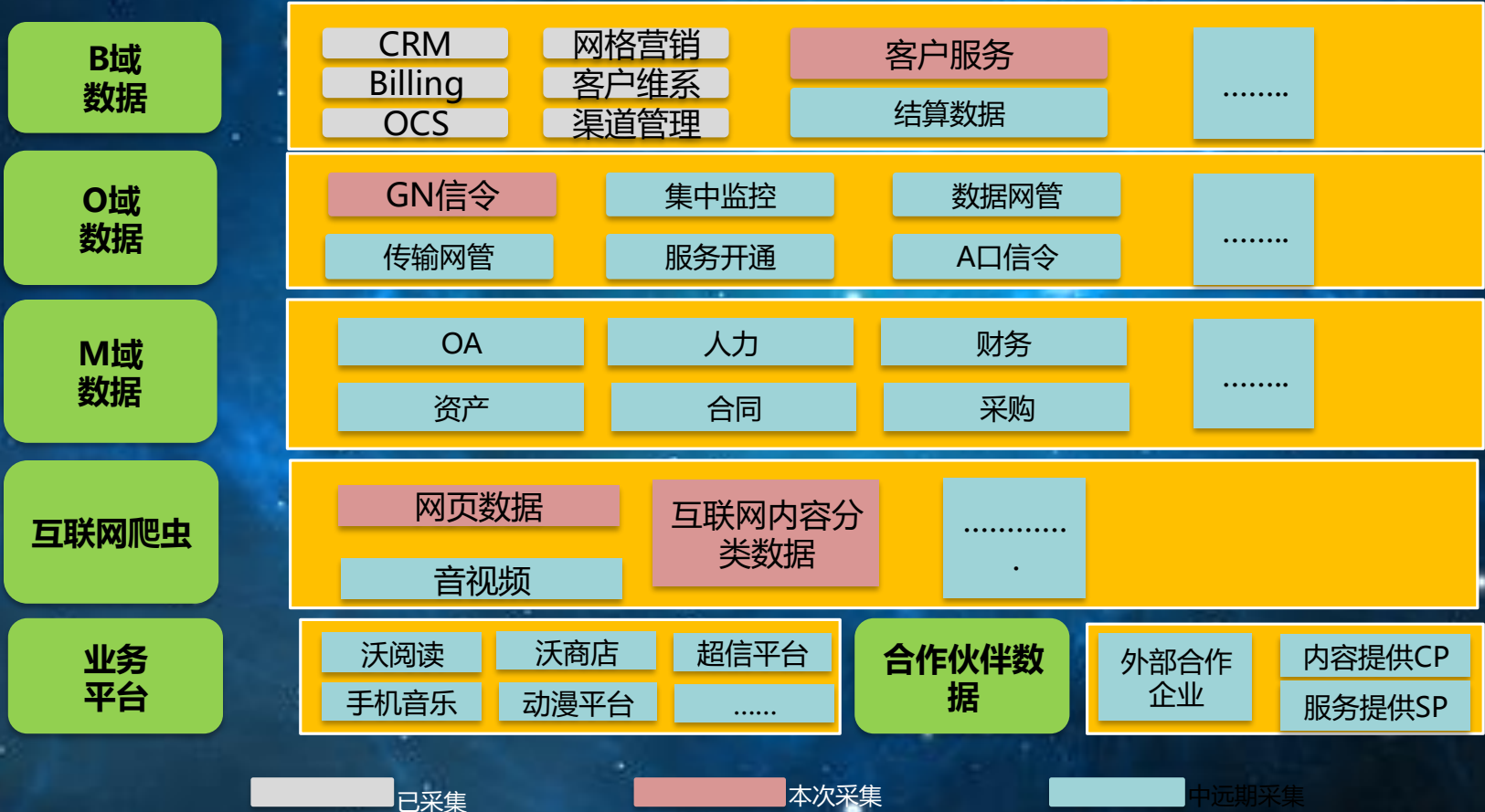
大数据中心内部数据流向



- 1、B域数据
- 2、O域Gn口信令数据
- 3、O域A口信令数据
- 4、用户资料、账单
- 5、话单
- 6、上网日志
- 7、分类日志
- 8、待爬取日志
- 9、明细数据
- 10、轻度汇总数据
- 11、基础中层数据
- 12、应用报表层数据
- 13、配置同步数据
- 14、用户特征标签
- 15、配置信息同步

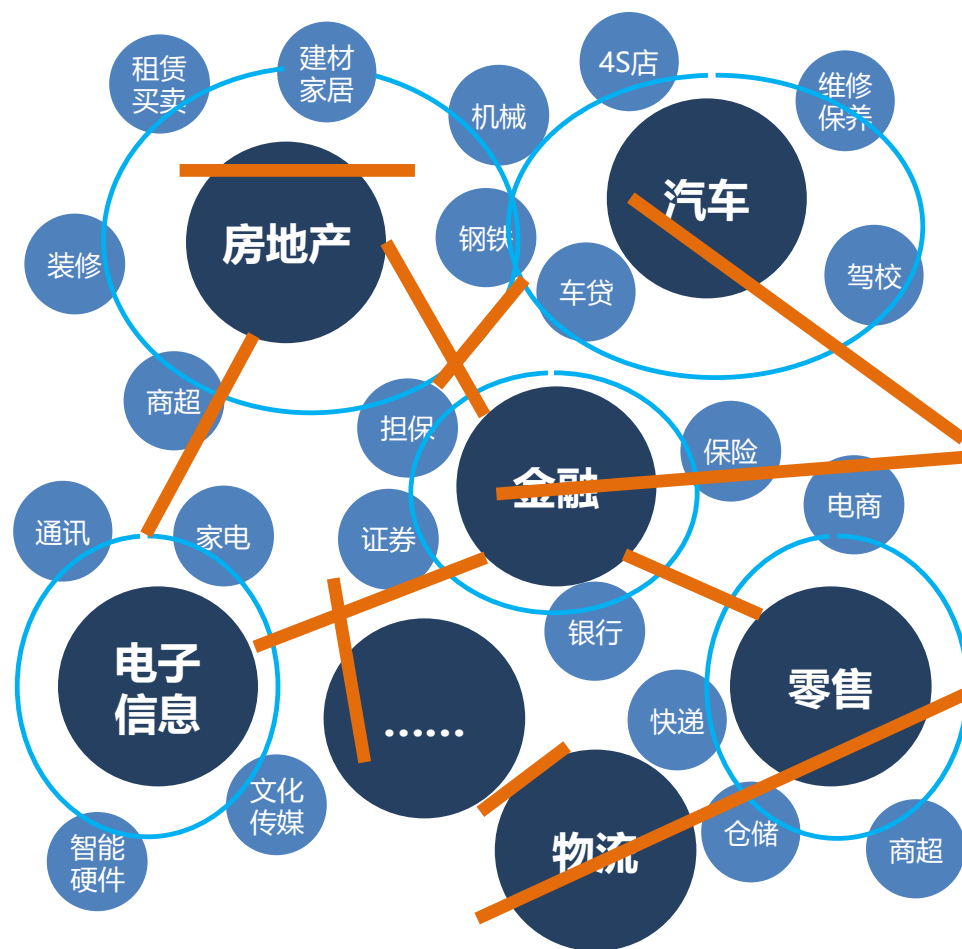
数据采集交换平台——外部数据引入

- 数据范围扩展的原则：按照应用/需求驱动为原则、高价值数据优先原则、用户相关数据优先原则
- 数据范围扩展的目标：近期以形成用户全视图、企业运营全视图为目标；中远期逐步涵盖企业全部的数据源。

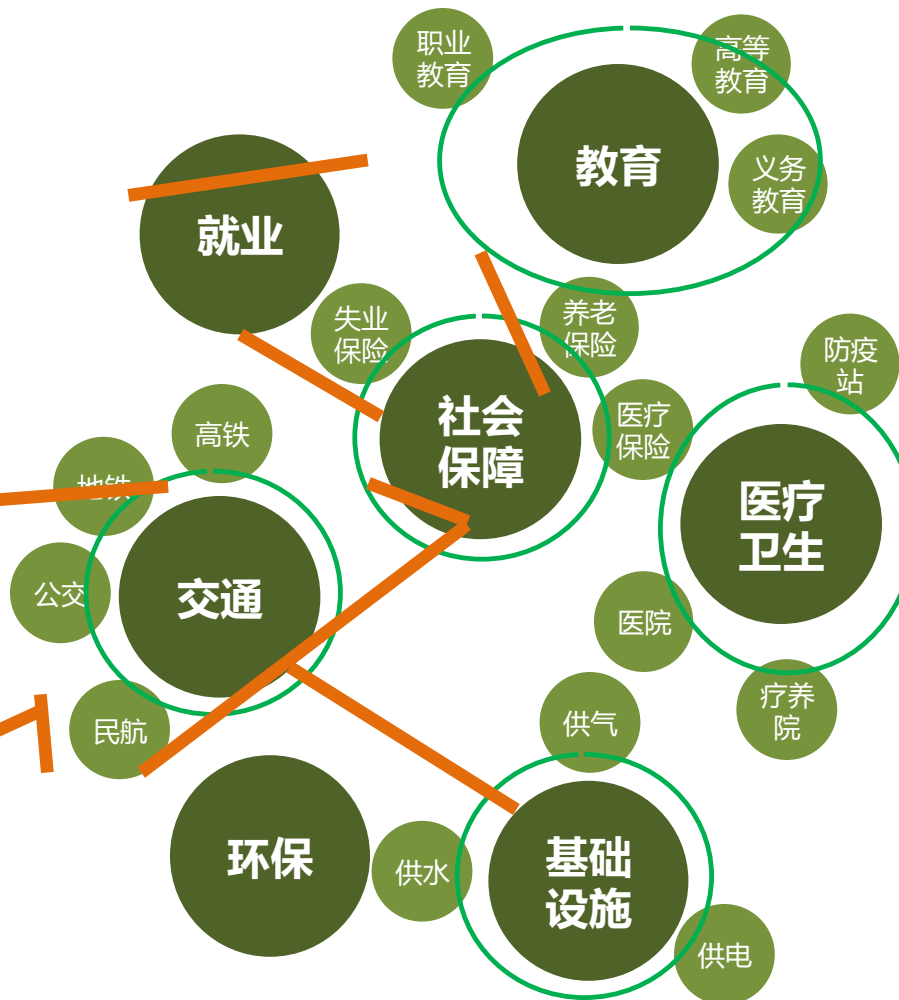


数据经过跨界融合，将会释放更大的价值

促进“数据化+”战略落地



引导“智能化”建设



在数据交易市场进行数据交易



如何做一个行业的数据产品？



一级分类

承载方式	科技领域	科技主题
• 论文	• 工业	• 十三五规划
• 期刊	• 信息行业	• 科技发展规划
• ...	• ...	• ...

二级分类

时间	链接	标签	国别
机构	人物	关联	知识产权
成果转化	科技投资	政策法规	...

搜索流程

```

graph TD
    A((高级搜索)) --> C{是否命中结果}
    B((开始搜索)) --> C
    C -- 否 --> D[条件筛选]
    D --> C
    C -- 是 --> E[结果展示]
    E --> F[相关文章]
    E --> G[相关主题]
    E --> H[结果详情]
    E --> I[...]
  
```

The diagram illustrates the architecture of the Science and Technology Big Data Platform. It is organized into four main vertical sections from left to right:

- Data Sources (Data Input):** A vertical stack of boxes representing various data inputs: 自有数据 (Self-owned data), 互联网数据 (Internet data), 科技文献 (Scientific literature), 科技论文 (Scientific papers), 其他地区数据 (Data from other regions), 第三方数据 (Third-party data), and ... (More data).
- 科技大数据平台 (Science and Technology Big Data Platform):** A central vertical bar representing the core platform.
- 科技进步指数分类 (Science and Technology Progress Index Classification):** A vertical bar representing the classification of the index.
- 自定义科技指数 (Custom Science and Technology Index):** A vertical bar representing the custom index.

On the right side, there are five horizontal boxes representing the index's functions and applications, connected to the central platform and classification bar:

- 分析研判 (Analysis and Judgment)
- 决策辅助 (Decision Support)
- 指数可视化 (Index Visualization)
- 指数监测 (Index Monitoring)
- 指数上报 (Index Reporting)

Below these functions, there are five horizontal boxes representing the index's applications, connected to the central platform and classification bar:

- 建立统一的指数标准及统计口径 (Establish unified index standards and statistical口径)
- 建立具有地区特色的科技指数体系 (Establish a science and technology index system with regional characteristics)
- 统一归口管理指数数据 (Unified management of index data)
- 通过指数辅助领导决策 (Use index to assist leadership decision-making)
- 辅助制定科技规划及人才政策 (Assist in formulating science and technology planning and talent policies)

The diagram illustrates the architecture of the Technology Big Data Platform. On the left, a vertical stack of data sources includes: 自有数据 (Self-owned data), 互联网数据 (Internet data), 科技文献 (Scientific literature), 科技论文 (Scientific papers), 科技地图数据 (Science map data), 第三方数据 (Third-party data), and ... (More data). These sources feed into a central vertical bar labeled 科技大数据平台 (Technology Big Data Platform). From this platform, data is processed through three main algorithmic paths: NLP遗忘算法 (NLP forgetting algorithm), 全文索引比对算法 (Full-text indexing and comparison algorithm), 相似度检测算法 (Similarity detection algorithm), 多语言检测算法 (Multi-language detection algorithm), and 文章片段检测算法 (Article fragment detection algorithm). The results of these processes are consolidated into a vertical bar labeled 查重结果 (Duplication results). This bar then feeds into a final box labeled 科技诚信 (Technology integrity), which is part of a larger system labeled 科技诚信档案库 (Technology integrity archive library). To the right of the 查重结果 bar, a list of five key features or outcomes is provided: 1. 利用多源数据进行关键因子查重 (Utilizing multi-source data for key factor duplication), 2. 利用大数据技术提高查重效率 (Utilizing big data technology to improve duplication efficiency), 3. 具备多维度重复率控制指标调节能力 (Having the ability to adjust control indicators for repetition rate from multiple dimensions), 4. 具备查重作业调度机制和过程可视化机制 (Having a scheduling mechanism for duplication work and a process visualization mechanism), and 5. 查重结果与科技诚信档案库互通，查重不合格信息记入科技诚信档案 (Duplication results are interoperable with the technology integrity archive library, and information that fails the duplication check is recorded in the technology integrity archive).

```

graph LR
    subgraph DataSources [ ]
        direction TB
        A[自有数据]
        B[互联网数据]
        C[科技文献]
        D[科技论文]
        E[科技地图数据]
        F[第三方数据]
        G[...]
    end
    DataSources --> H[科技大数据平台]
    H --> I[NLP遗忘算法]
    H --> J[全文索引比对算法]
    H --> K[相似度检测算法]
    H --> L[多语言检测算法]
    H --> M[文章片段检测算法]
    I --> N[查重结果]
    J --> N
    K --> N
    L --> N
    M --> N
    N --> O[科技诚信]
    O --> P[科技诚信档案库]
    
```

- 利用多源数据进行关键因子查重
- 利用大数据技术提高查重效率
- 具备多维度重复率控制指标调节能力
- 具备查重作业调度机制和过程可视化机制
- 查重结果与科技诚信档案库互通，查重不合格信息记入科技诚信档案

科技大数据平台

- 历史信用记录
- 申报单位
- 项目成员
- 评审专家
- 学术腐败
- 学术早就
- 异常结项
- 科技泄密
- 学术不端
- ...
- 互联网征信数据
- ...

- 一企一档
- 一人一档

- 建立一人一档、一企一档的科技诚信档案
- 通过大数据技术对科研单位、科技人员、评审专家等进行多维度诚信记录，建立科技诚信档案
- 依托科技诚信档案数据限制失信人参与科技项目申报，获取科研经费支撑，限制进入重点实验室、限制科技设备使用等控制措施

Diagram illustrating the data flow and analysis process for high-tech enterprises:

- Data Sources (Left):** 高科技企业 (High-tech Enterprise), 科技孵化器 (Science and Technology Incubator), 科研基地平台 (Science Research Base Platform), 实验室 (Laboratory), 工程研究中心 (Engineering Research Center),
- Intermediate Data (Middle):** 科技成果 (Science and Technology Achievement), 科技实力 (Science and Technology Power), 位置分布 (Location Distribution), 机构详情 (Institution Details), 机构专家 (Institution Experts),
- Final Data (Right):** 高科技企业 (High-tech Enterprise), 高科技企业 (High-tech Enterprise), 高科技企业 (High-tech Enterprise)

Supporting Functions (Right):

- 统一归口地区科技资源数据 (Unified regional science and technology resource data)
- 对机构的科技成果、科技实力等信息进行数据关联 (Data linkage for science and technology achievements, science and technology power, etc. of institutions)
- 对科技资源进行分类检索 (Classification search for science and technology resources)
- 对科技资源进行可视化展现 (Visualization of science and technology resources)
- 对科技资源的变化情况进行可视化跟踪 (Visualization tracking of changes in science and technology resources)

4

30天调研，7天交流，2天策划会，助力大数据项目应用落地



数据产品倒逼产业链健康发展



四大行业数据产品客户需求调研

行业	应用场景		应用部门	带来价值
地产	建房决策	1. 楼盘选址。获取拟建楼盘周边的人群生活轨迹、生活习惯和消费能力，结合当地政策、经济、环境等因素，进行土地估值、置业吸引力分析，辅助楼盘选址，确定项目定位； 2. 建筑设计。通过消费解码、品牌识人、媒体兴趣、上网习惯，洞察客群的消费观和价值观，还原客群的生活方式，指导产品的设计和创新。	决策部门/产品设计部门	楼盘选址和设计更加符合当地购房需求
	楼盘营销	1. 人群定向。通过用户的行为轨迹、活动热力图进行地域定向，通过置业情况、家庭结构、购买能力、关注楼盘等进行人群定向，通过地产垂直媒体浏览行为进行媒体定向。 2. 广告投放。通过电话、短信、媒体、户外广告等通道触达用户，进行精准营销。	市场部	精准获客，促进销售
	售后服务	1. 掌握业主全面的兴趣偏好，如喜欢吃什么、喜欢买什么产品、经常买什么药品、家里几口人、是否有家政服务、房屋装修、教育培训需求等； 2. 对这些业主提供个性化的社区服务，如房屋装修、家政、医疗、餐饮、房屋租赁、二手房交易、社区金融、社区商铺、生鲜配送等，提升品牌忠诚度，扩大盈利规模。	售后	提升品牌忠诚度，扩大盈利规模
媒体	素材采集	1. 通过外部新闻数据爬取，建立素材库，帮助编辑进行原创、伪原创和专题制作； 2. 获取专业词库进行文本分析，帮助编辑提炼文章摘要、发现事件热点。	采编部	扩充自身内容量，提高采编效率
	广告主招募	1. 分析媒体上的高价值用户，产出分析报告，包括用户的兴趣偏好、消费能力； 2. 将报告呈现给品牌广告主，用高价值用户吸引广告主进行广告投放。	广告业务部	促进广告位销售，提高广告主的满意度
文娱	投资决策	1. 市场大环境分析，拍摄前对市场环境、竞争环境、票房及收视率进行分析； 2. 同类题材估值预测，对同类题材的受众画像、关注点、口碑声誉、对明星的关注度进行全面分析，辅助投资决策。	决策部门	辅助投资拍摄前决策
	影视制作	1. 前期通过用户对明星的印象为明星打标签，再根据影视剧要求进行大数据选角； 2. 中期通过观众对剧情的喜好与反馈，调整下一步的剧情走向； 3. 后期根据用户对角色、剧情的偏好，设计电影衍生品。	制片部	满足受众需求，提高收视率
	舆情监测	1. 监控节目开播前、中、后受众的舆论走向、话题焦点，以及评价正负面； 2. 根据监测情况进行话题焦点跟踪，问题改进。	市场部	影视剧口碑监测
医疗	精准营销	1. 全面获取健康领域生态圈的数据，如母婴数据、体检数据、医院数据； 2. 基于这些数据针对性举办营销活动、健康讲座，并对用户进行跟踪管理。	市场部	营销获客
	新药研发	抓取论文数据，分析最新的技术成果，使用新技术研发新产品。	研发部	技术改进

四大行业媒体数据产品外部需求调研

行业	解决问题		应用部门	带来价值
媒体	媒体资源汇聚平台	媒体资源汇聚平台采用分布式技术，以任务调度、采集集群和解析集群相结合的方式，实现对境内外网站、微博以及移动新闻客户端的实时采集。平台在既能实现本地化采集的同时也能够提供基于云端数据服务，依托于拓尔思建设的数据中心，可以面向媒体提供所需的全网数据推送和新闻热点分析服务，为报社的选题策划和编前会提供量化数据支撑。	编辑部	
	全媒体资源管理平台	全媒体资源管理平台实现了对各种渠道不同来源的稿件和信息资源进行统一汇总管理，将多来源的本地资源、互联网资源和新媒体资源进行汇总集中使用，采用大数据管理系统将数据以多维分类的方式提供资源使用服务。通过领域分类、多维聚类、智能关联以及自动专题汇聚等多种智能化方式将资源信息抽取为相互关联的决策知识。使用碎片化和标签化的智能技术，将所有类型的稿件知识抽取，形成媒体资源中心潜在的知识关联网络。融合管理使用全媒资源内容。	广告部、编辑部、媒介合作	
	传媒运营业务支撑平台	媒体智能分析决策平台采用大数据的智能分析决策平台支撑整体媒体应用解决方案的所有信息挖掘分析服务，所有基于语义面向内容的智能化和面向行为分析的智能化技术都融合在媒体智能分析决策平台。本平台还从媒体业务应用的多个维度入手，将智能服务通过组件化封装，提供给媒体业务平台的产前、产中和产后等各环节实现新媒体转型流程的全方位的智能支撑服务体系服务。	广告	
	媒体传播效果分析平台	1.剧本内容语义分析：剧本名称、剧本题材、故事梗概、故事结构、情节逻辑、出入人物、人物关系等 一体化采编平台是面向媒体转型的未来的媒体内容创作平台，面向多个渠道的定制编辑发布，采用工作流引擎实现不同媒体的自定义流程，同时兼顾现有的媒体传统采编流程。平台无需定制开发，即可实现兼容未来可能产生的各种新型媒体的编辑、发布渠道和方式。一体化采编发实现了生产制作一体化、编辑加工一体化、数据共享一体化、渠道监控一体化、渠道发布一体化。	广告部、编辑部	
	用户资产分析平台	用户资产分析平台通过对各渠道用户信息的采集，存储，进行统一管理，建设统一的用户管理平台。平台采用行为分析工具及挖掘模型完成数据的统计分析，实现信息的个性化精准传播，为服务对象不同的需求提供有针对性的生产特色信息产品。	广告部、编辑部	

四大行业地产数据产品外部需求调研

行业	解决问题		应用部门	带来价值
媒体	建房决策	1. 楼盘选址。获取拟建楼盘周边的人群生活轨迹、生活习惯和消费能力，结合当地政策、经济、环境等因素，进行土地估值、置业吸引力分析，辅助楼盘选址，确定项目定位； 2. 建筑设计。通过消费解码、品牌识人、媒体兴趣、上网习惯，洞察客群的消费观和价值观，还原客群的生活方式，指导产品的设计和创新。	决策部门/ 产品设计 部门	楼盘选址和设计更加符合当地购房需求
	楼盘营销	1. 人群定向。通过用户的行为轨迹、活动热力图进行地域定向，通过置业情况、家庭结构、购买能力、关注楼盘等进行人群定向，通过地产垂直媒体浏览行为进行媒体定向。 2. 广告投放。通过电话、短信、媒体、户外广告等渠道触达用户，进行精准营销。	市场部	精准获客，促进销售
	售后服务	1. 掌握业主全面的兴趣偏好，如喜欢吃什么、喜欢买什么产品、经常买什么药品、家里几口人、是否有家政服务、房屋装修、教育培训需求等； 2. 对这些业主提供个性化的社区服务，如房屋装修、家政、医疗、餐饮、房屋租赁、二手房交易、社区金融、社区商铺、生鲜配送等，提升品牌忠诚度，扩大盈利规模。	售后	提升品牌忠诚度 扩大盈利规模

XX企业做SaaS数据产品的几种战略选择

媒体：选题宝

1. 价值：媒体选题监测系统
2. 功能：分析媒体所有的内容
3. 客户对象：主编、采编、记者、运营总监。
4. 运营模式：线上运营，费用低廉，自动升级完善
5. 前景分析：主编和采编工作量巨大、机器能帮到制作内容很重要

医疗：医查查

1. 价值：为药厂、保险公司分析医院和具体病人的数据查询系统
2. 功能：聚合医疗数据，分析药厂对于数据的需求，分析保险公司对于一个人的病情了解。
3. 客户对象：药厂的市场部门，保险公司理赔人员
4. 运营模式：线上运营，费用低廉，自动升级完善
5. 前景分析：全国保险公司理赔线下成本太高，意义重大

地产：业主通

1. 价值：为地产企业提供他们业主和商户经验分析
2. 功能：业主分析，周边商业分析、业主舆情分析
3. 线上运营，费用低廉，自动升级完善
4. 前景分析：地产企业众多，物业公司众多，前景很大

文娱：影视查

1. 价值：预测电视电影的市场前景。
2. 功能：查询新电影可能的影响力，目标客户，区域，观众喜欢指数等：
3. 客户对象：导演，制片人
4. 运营模式：线上运营，按照数据量收费，自动升级完善
5. 前景分析：电影市场前景广阔

四大行业数据产品评估

XX企业数据产品评估					
项目	媒体数据产品（选题宝）	医疗数据产品（医查查）	地产数据产品（业主通）	文娱数据产品（影视查）	权重
数据源	9	6	6	6	30%
需求迫切度	5	9	7	8	20%
客户价值	4	9	9	8	20%
数据加工能力	8	5	8	8	5%
数据产品价格	5	8	6	6	5%
开发周期	8	5	5	6	5%
行业经验	6	5	6	5	5%
竞争环境	5	8	9	6	5%
客户关系	6	5	6	5	5%
总分	6.4	7.0	6.9	6.8	100%

风险分析 - 人

人才架构

技术架构



以行业大数据项目需求养产品，以爆品数据产品打行业数据市场



从行业巨头来说

需要释放数据价值获取更多数据

- 行业数据的提供
- 行业数据分析的结果
- 对数据进行分析
- 数据驱动产业变革

从中小企业来说

需要DaaS新的云化有数据的环境

- 数据产品

当前大数据的发展趋势

- 数据行业化深度应用
- 数据产品SAAS化并进行数据交换
- 有独立数据源的公司成为数据变现核心企业
- 数据处理成本在贬低
- 数据驱动业务、数据改变流程、数据辅助管理正在形成
- 数据资产贷款、数据股权融资、数据金融体系正在建立

大数据的独角兽到底在哪里？业务在哪里数据价值就在哪里



国家机密

Plantir分析了全球最有价值的FBI、联邦政府反恐、金融客户如高盛、中情局等核心的数据，估值超过200亿美金



BAT + Google + Facebook + AWS

以量取胜涉及民生，解决社会问题，引领科技发展的平台

资源独断强于数据垄断以业务驱动数据，以数据驱动业务，行业巨头的数据在业务中价值极大的发挥出来



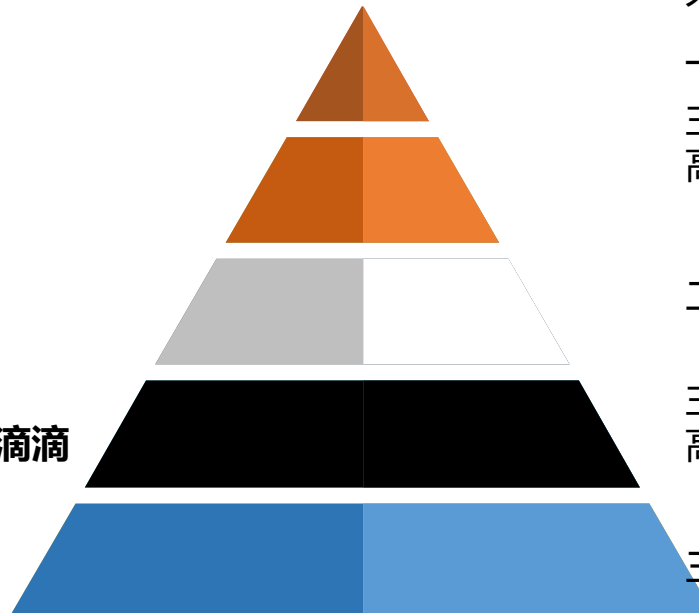
新世界：分享经济龙头企业Airbnb+Uber+滴滴

数据价值延伸到物品价值，成为新的商业模式，打破传统的商业模式



垂直的云平台

企业的核心业务数据从内部到外部，随着法律法规健全，数据交换意识提升，云平台会代表企业和其他云进行数据交换，达到双赢，行业共赢的局面



大数据价值规律：

一是以数据驱动的**决策**：
主要通过提高预测概率，来提高决策成功率；

二是以数据驱动的**流程**
(data-driven processes)，
主要是形成营销闭环战略，提高销售漏斗的转化率；

三是以数据驱动的**产品**
(data-driven products)，
在产品阶段，强调个性化；在产品运营阶段，则强调迭代式创新。

2017年大数据发展10大预测

1. 大数据在政府应用将增多，项目增长迅猛，政府从电子政务到智慧政府需要的是数据思维，技术和数据源的补充。
2. 大数据在人工智能的应用将爆发，特别是智能APP
3. 大数据驱动产业转型升级方案将成为热点，大数据+产业+金融概念盛行；
4. 早期的大数据企业讲概念的、融不到资本的将死亡，有客户但不能交付的也将死亡。
5. 大数据在运营商、金融、政府、互联网、医疗、交通、媒体广告、之外的行业开始产生巨大影响；
6. 大数据培训和咨询持续火爆，不在新鲜，而是成为一种培训机构的一般性课程，常态化。百度问咖、在行、厅客、分答等技能分享平台颠覆咨询产业；
7. 大数据人才稀缺严重、随着竞争加剧，很多企业想要进入大数据产业但是缺少好的数据架构师，分析师将缺少竞争力；
8. 大数据驱动智能自助服务快速发展；
9. 迫于难交易，交易难的问题，大数据交易的立法和技术解决手段将得到发展；
10. 基于物联网传感器的数据采集应用将增加成为新的投资热点；

当前大数据的十个观点

1. 数据价值依赖用户数
2. 数据价值释放必须依赖产品，且和业务紧密向管理，数据产品的加工需要学习汤森路透
3. 大数据+超算+深度的学习会革新所有产业
4. 大数据是量子力学，但解决问题还是需要牛顿力学。
5. 心中有数为智慧，物中有数为智能，人工智能是数据变现的绝佳方式数据
6. 资产贷款、数据股权融资、数据金融体系正在建立
7. 大数据的最佳商业模式是共享经济，数据被共享利用的次数越多，数据价值越大。
8. 数据的思维，技术和数据的流通共享将是政府工作转型升级的抓手，是官员自我能力提升必要路径，不懂产业、不懂数据如何驱动政府管理变革，不是合格政府领导。
9. 数据目前对于很多企业来说是负债不是资产，1个T的数据每年的计算、存储、宽带运营成本是2000美金，数据利用收益低于运营成本，数据就是负债，目前谈数据是资产还是比较早，大数据处于非常初期的阶段。
10. 未来每个企业都会产生一个数据合作部门



三大体系

- 法律法规体系
- 政策支持体系
- 数据干部体系

5大工程

- 大数据基础平台建设
- 数据应用展示中心建设
- 数据共享平台建设
- 数据可视化工程建设 数据惠民工程

先行先试

- 数据投行 数据展示中心 数据惠民中心

产品说话

- 企查查、人查查、医查查、交通查、政务查

1

顶设：发展大数据的顶层设计体系

2

培训：建立高效的指导思想体系

3

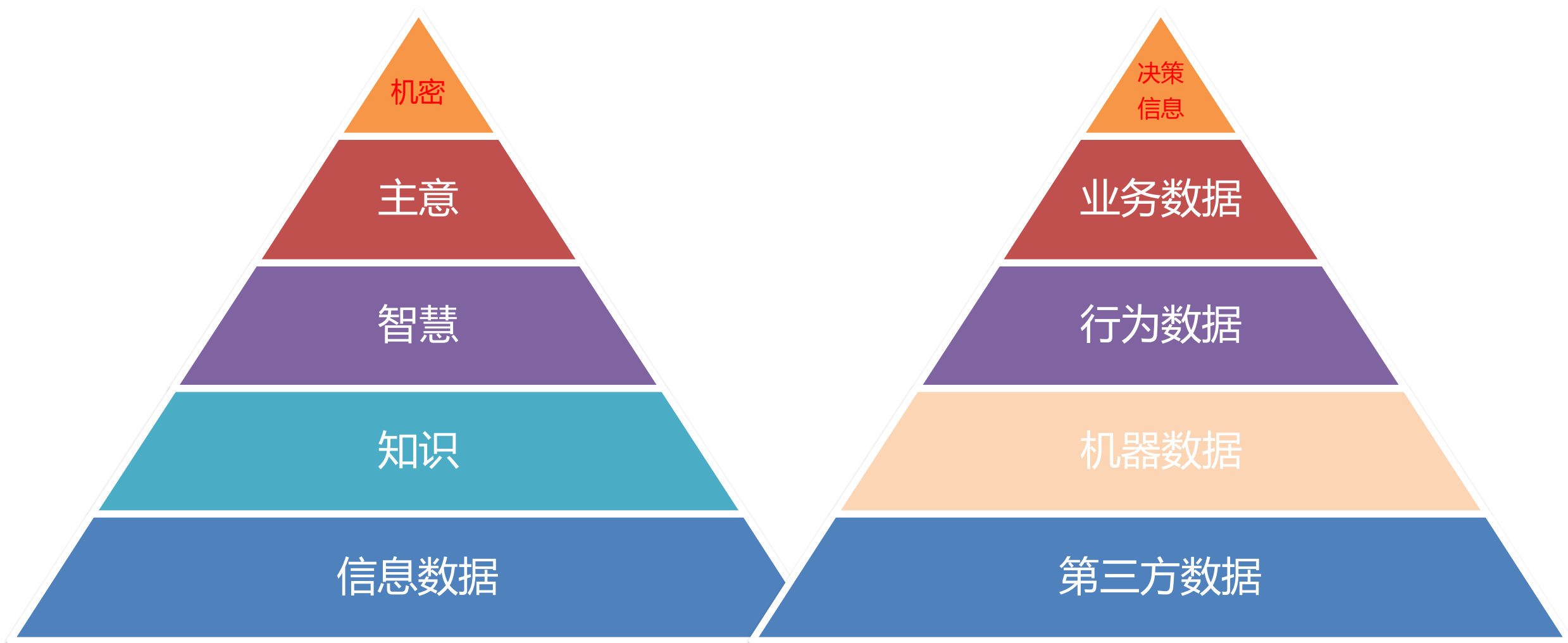
项目：基础支撑、应用带动、产业突破

4

亮点：先行先试，创新思路、野蛮发展

5

能力：人数据系统场景缺一不可



城市数据产品生产者

城市数据产品生产者

思考：广州有哪些数据资源？产业发展布局是什么？可以做哪些数据产品？

电力
数据
产品

水利
数据
产品

交通
数据
产品

招商
数据
产品

公共
服务
数据
产品

能力
开放
平台



驾驭大数据掌控未来

“集中群体智慧，激活人员动力”——构建数据推动的商业体系