

“大数据”专题分析报告

专题报告

本报告的主要看点:

1. 揭示了大数据革命性的影响，阐明其对政府、互联网企业、传统企业、个人的价值。
2. 指出互联网企业独领风骚的本质，及未来的发展方向。
3. 指出传统企业面临的核心问题，并探讨突围之道。
4. 以大家耳熟能详的新浪微博为例，剖析大数据的特点。
5. 尽量避免使用缩写和术语，便于更多人士理解接受。
6. 受篇幅、时间所限，并未对各行业展开论述。将在后续报告中详细讨论。
7. 本报告来自分析师多年的行业经验以及与 Oracle、EMC、IBM、麦肯锡、神州数码以及诸多上市公司深度探讨所得，在此一并表示感谢。

大数据时代即将到来!

基本结论

- Big Data “大数据”是继云计算、物联网之后 IT 产业又一次颠覆性的技术变革。对国家治理模式、对企业的决策、组织和业务流程、对个人生活方式都将产生巨大的影响。
- 大数据时代网民和消费者的界限正在消弭；企业的疆界变得模糊；数据成为核心的资产，并将深刻影响企业的业务模式，甚至重构其文化和组织。大数据时代顺昌逆亡。如果不能利用大数据更加贴近消费者、深刻理解需求、高效分析信息并作出预判，所有传统的产品公司都只能沦为新型用户平台级公司的附庸，其衰落不是管理能扭转的。
- 云计算的核心是业务模式，本质是数据处理技术。数据是资产，云为数据资产提供了保管、访问的场所和渠道。如何盘活数据资产，使其为国家治理、企业决策乃至个人生活服务，是大数据的核心议题，也是云计算内在的灵魂和必然的升级方向。
- 企业内部的经营交易信息，物联网世界中商品、物流信息，互联网世界中人与人交互信息、位置信息等等是大数据的三个主要来源。其信息量远远超越了现有企业 IT 架构和基础设施的承载能力，其实时性要求则大大超越现有的计算能力。如果计划在大数据时代获益，必将引发新一轮的信息化投资和建设热潮。
- 企业在大数据时代，面临文化、战略、组织、流程、信息化、公共关系、人才培养方方面面的挑战，同时也迎来重大的转型机遇和飞跃契机。我们看好拥有大数据处理、挖掘技术的公司，看好拥有数据资产公司，看好拥有数据分析人才的公司。

投资标的

- 大数据技术提供商：超图软件、科大讯飞、拓尔思
- IT 系统咨询商：汉得信息、太极股份
- ERP/CRM/BI 改造商：用友软件、东方国信、久其软件
- 终端商向数据服务转型：广联达、大智慧、四维图新
- 新型展现方式：威创股份
- 信息安全：卫士通

重点推荐

超图软件、拓尔思、威创股份、广联达、用友软件、科大讯飞

赵国栋 联系人
(8639)10702427
zhaogd@gjzq.com.cn
易欢欢 分析师 SAC 执业编号：S1130511080002
(8621)61038267
yihh@gjzq.com.cn

图表1: 大数据概览



来源: 互联网 国金证券研究所

图表2: 投资标的一览表

代码	股票名称	11年EPS	12年EPS	13年EPS	11年P/E	12年P/E	11-12PEG	股价	11-12增长	12-13增长	年初至今相对涨跌幅%	一年相对涨跌幅%
002410.SZ	广联达	1.05	1.55	2.25	33.99	23.03	0.71	35.69	47.62%	45.00%	-27.83	-17.61
002405.SZ	四维图新	0.65	0.83	1.15	36.02	28.20	1.30	23.41	27.69%	38.00%	-53.67	-46.14
002368.SZ	太极股份	0.67	0.94	1.29	32.69	23.35	0.82	21.90	40.00%	38.00%	-21.14	-8.82
300036.SZ	超图软件	0.53	0.79	1.10	35.58	24.00	0.74	18.90	48.24%	40.00%	1.13	-5.09
002268.SZ	卫士通	0.55	0.78	1.10	45.85	31.91	1.05	25.04	43.66%	40.00%	-17.74	-17.18
600588.SH	用友软件	0.70	0.93	1.26	19.86	14.93	0.60	13.90	33.00%	35.00%	-15.59	-20.03
002308.SZ	威创股份	0.40	0.60	0.84	98.75	65.83	1.98	39.50	50.00%	40.00%	-18.50	-21.34
300229.SZ	拓尔思	0.67	0.88	1.17	44.31	33.57	1.38	29.69	32.00%	32.00%	79.17	79.17
300170.SZ	汉得信息	0.58	0.82	1.10	39.81	28.04	0.95	23.05	41.97%	34.18%	9.60	9.60
002279.SZ	久其软件	0.59	0.72	1.05	37.22	30.50	1.69	21.96	22.03%	45.83%	-24.89	-22.44
300166.SZ	东方国信	1.552	2.372	3.458	32.40	21.20	0.61	50.29	52.84%	45.78%	6.00	6.00
601519.SH	大智慧	1.464	1.919	-	9.39	7.16	0.30	13.74	31.08%	-	-45.39	-45.39
002230.SZ	科大讯飞	1.461	2.141	3.44	24.38	16.64	0.52	35.62	46.54%	60.67%	-32.18	-27.40

来源: 国金证券研究所

内容目录

为什么是新浪微博？	5
海量的数据	5
各种类型的数据	7
社会化关系是价值倍增器	7
实时的处理	8
稀疏但珍贵的价值	8
互联网企业演进逻辑和本质	8
互联网的本质是理解用户，并将向智能化演进	8
Google 精准满足用户需求	9
Facebook 创造需求——微力改变世界	10
传统产品制造企业之殇	12
“你的客户是谁？”——德鲁克	12
对消费品制造的影响	13
对软件开发模式的影响	13
对企业信息化的影响	13
当传统企业嫁接大数据	13
大数据引发的变革	14
大数据的内涵	14
大数据三个主要的数据来源	14
大数据赋予我们洞悉未来的能力	15
大数据四个主要特征	15
大数据技术发展的脉络和现状	16
大数据是迈向智慧世界的自然演进	16
IT 产业巨头的大数据技术及发展简介	18
大数据伴生的大机遇	21
产业规模	21
企业的投入	22
政府可能采取的行动	22
大数据的投资标的	23

图表目录

图表 1：微博推出后，新浪股价摆脱长期横盘的态势，一路暴涨	5
图表 2：新浪微博关注度 Top10 粉丝数	6
图表 3：Top10 传统媒体的发行量	6

图表 4: 生活在数据之中, 所有人都在制造和分享数据	6
图表 5: 关系本身就是价值。分析这个微博关系可以得出我将推荐哪支股票!	7
图表 6: 为了一点金子, 需要保存全部沙子。	8
图表 7 越来越真实的互联网, 正越来越智能.....	9
图表 8: Google 众多的产品线, 核心在于尽可能多掌握用户行为, 从而精确获 取用户需求	10
图表 9: Facebook 概况	11
图表 10: 奥巴马与马克·扎克伯格	12
图表 11: 社会认同效应.....	12
图表 12: 大数据的构成, 以数据融合、综合处理为方向	14
图表 13: 马云的判断来自于数据分析.....	15
图表 14: 阿里巴巴的规模保障了数据的代表性	15
图表 15: 大数据的四个主要特征	16
图表 16: “大数据”关键词搜索量处于激增阶段	16
图表 17: 相当于云计算 2008 年左右的水平	16
图表 18: Gartner 2011 年技术成熟度曲线, 大数据处于高速发展期.....	17
图表 19: 技术演进历史揭示未来是大数据驱动的智慧型经济模式.....	18
图表 20: 巨头在大数据时代的业务布局	19
图表 21: EMC 斩获了纽交所和 Nasdaq	19
图表 22: IBM 依然保持银行、电信行业的强势地位.....	20
图表 23: 微软在中小企业市场发力	20
图表 24: Oracle 扮演 IBM 狙击手的角色	21
图表 25: 大数据的市场空间及对社会的贡献.....	22
图表 26: 计算机和信息服务产业将在大数据浪潮中首先获益	22
图表 27: 投资标的一览表	23

为什么是新浪微博？

“从 Ipad 和微博中，我似乎看到了一点点未来……”

——默多克

- 微博最强大的力量来自于“小众的社会认同效应”。独立的网民因为某种兴趣、爱好、观点而集聚在一起，通过消息转发、评论的机制，广为扩散，其他网民迅速跟进，短时间内形成类似“飓风”的传播效应，从而形成广泛的社会影响力。
- 微博颠覆了传统的信息传播路径，使几千年来，单中心、单向的传播方式，向多中心、网状裂变传播方式转变。对企业而言，不管是主动拥抱这个变化，还是被动接受，都宣告了一个时代的大幕真正开启，消费者站上了舞台中央。
- 但此消费者是数字化的，他生活在网络中，他并不会直接告诉你他的需求，企业必须自己去收集、分析、跟踪、对比他在互联网上留下的各种“足迹”、评论、图片、视频等。谁首先在海量数据中沙里淘金，谁首先嫁接互联网基因，谁将在这轮技术变革中，脱颖而出。这就是大数据时代的核心命题！

图表3：微博推出后，新浪股价摆脱长期横盘的态势，一路暴涨



来源：雅虎财经 国金证券研究所

- 自 2 年前推出以来，新浪微博截止到 2011 年第三季度已经拥有 2.5 亿注册用户，新浪市值一度达到 94 亿美元。微博上最具人气的明星姚晨拥有 1400 多万粉丝。超过前五名报纸发行量的总和。
- 关于微博的价值，李开复博士可以写一本书。本报告主要关注大数据，因此，我们来分析，与新浪微博类似的网站，像 Facebook, Tweeter, 有哪些数据？这些数据有什么特点？通过分析这些数据我们可以获得什么？

海量的数据

- 每秒钟，人们发送 290 封电子邮件；每分钟人们在 youtube 上传 20 小时的视频；人们每月在总共在 facebook 上浏览 7000 亿分钟；移动互联网用户发送和上传的数据量达到 1.3exabytes，相当于 10 的 18 次方；每秒钟亚马逊处理 72.9 笔订单。

图表4: 新浪微博关注度 Top10 粉丝数

关注度TOP10	
1 姚晨	14,973,414
2 小S	14,080,436
3 谢娜	12,328,707
4 蔡康永	11,997,673
5 杨幂	11,920,236
6 赵薇	11,679,795
7 何炅	11,183,328
8 王力宏	10,743,473
9 李冰冰	9,744,137
10 李开复	9,213,678

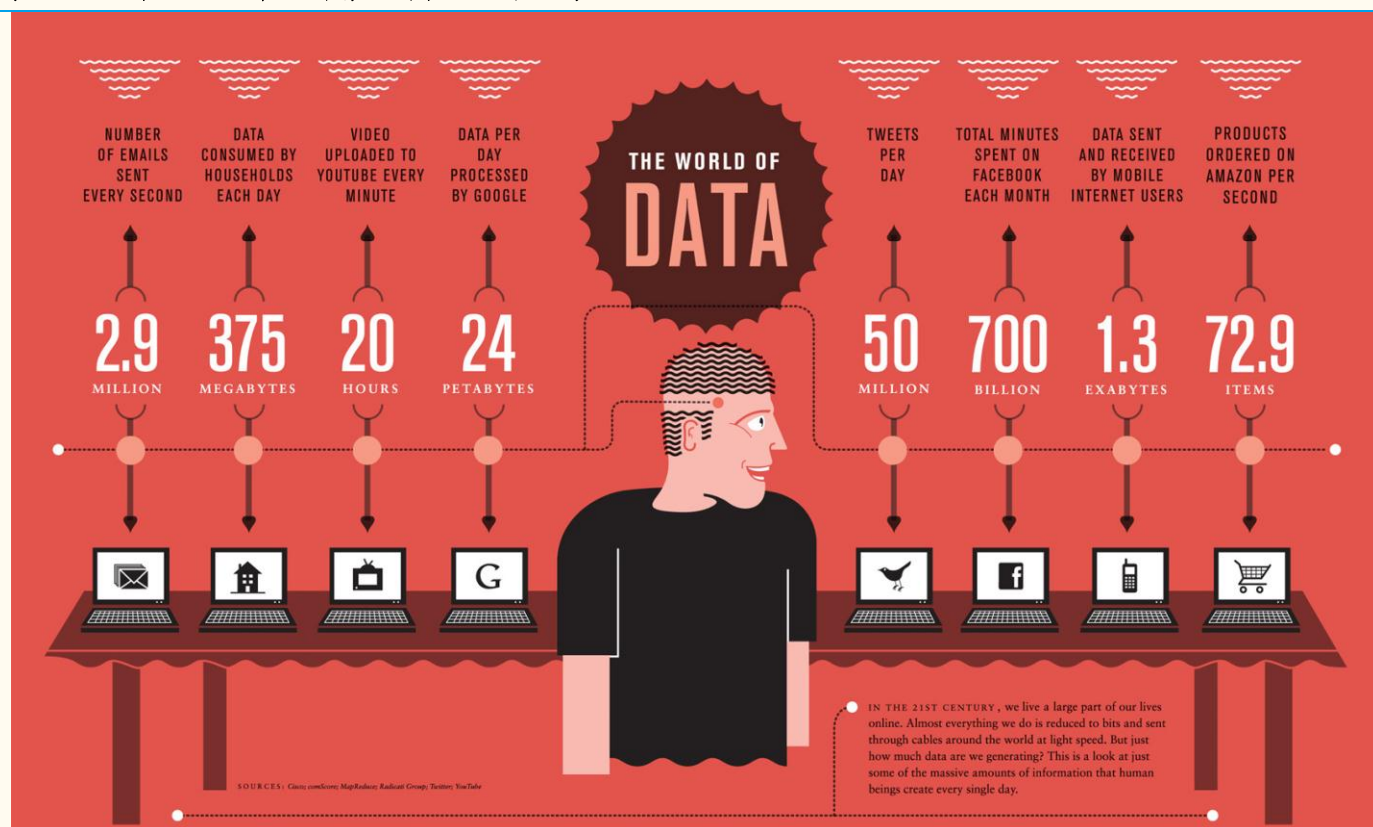
来源: 新浪微博 搜报网 国金证券研究所

图表5: Top10 传统媒体的发行量

序号	标题	发行量
1	参考消息	350万份
2	人民日报	280万份
3	环球时报	240万份
4	扬子晚报	220万份
5	中国电视报	200万份
6	广州日报	185万份
7	体坛周报	160万份
8	南方都市报	160万份
9	齐鲁晚报	150万份
10	信息时报	148万份

- 对这些海量数据的存储, 超过了任何一家传统企业的能力。但是对于互联网平台级的公司, 他们每时每刻在忙于把这些数据收集、整理、归类、保存。

图表6: 生活在数据之中, 所有人都在制造和分享数据



来源: IBM 国金证券研究所

- 麦肯锡几个研究团队曾研究过全世界产生、存储和消费的数据总量。尽管他们的估计数字以及由此产生的结果各不相同, 所有研究都表明, 未来数年会呈现指数增长。麦肯锡全球研究院 (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。1EB 数据相当于美国国会图书馆中存储的数据的 4000 多倍。事实上, 我们如今产生如此多的数据, 以至于根本不可能全部存储下来。例如, 医疗卫生提供商会处理掉

他们所产生的 90%的数据（比如手术过程中产生的几乎所有实时视频图像）。

各种类型的数据

- 传统的企业中，被有效的管理的数据都是以表格的形式保存在数据库中。所有的信息格式都一样。便于编程处理。但像新浪微博，海量数据中，包括各种各样的数据类型。最常见的普通文本、照片，视频等等，还有像位置信息、链接信息等 XML 类型的数据，每个信息都可能不同。还有“长微博”工具，可以把文字转换成图片，突破 140 字的限制，然后发布的微博中。往往这些图片形式存在的微博，包含大量的信息。
- 能够在不同的数据类型中，进行交叉分析的技术，是大数据的核心技术之一。语义分析技术、图文转换技术、模式识别技术、地理信息技术等等，都在大数据分析时获得应用。

图表7：关系本身就是价值。分析这个微博关系可以得出我将推荐哪支股票！



来源：新浪微博 国金证券研究所

社会化关系是价值倍增器

- 社会化的网络，为大数据提供了额外的价值维度。关系本身就是价值。一件商品，被笔者关注，和被姚晨关注，其效果完全不可同日而语。厂商可能会给姚晨付费，对于我最多只有一句“谢谢关注”而已。
- 酒类企业采用一种“盘中盘”的营销策略，就是主攻有社会地位人士，由他辐射其有“关系”的人，从而打开销路。这些人映射到微博空间，就是“意见领袖”、“草根大号”。

- 社会化关系之于数据的价值，是乘数的关系。数据价值 X 社会化关系=微博影响力。借助微博的关系属性，流言更是飞速传播，真相也会加速浮出水面。

实时的处理

- 一般而言，像数据仓库系统、BI 应用，对处理时间的要求并不高。因此这类应用往往运行 1、2 天获得结果依然可行的。但对于大数据应用而言，必须要在 1 秒钟内形成答案，否则这些结果可能就是过时的，无效的。
- 股价波动信息对金融从业人员感同身受，举另外一个常见的例子，导航应用已经飞入寻常百姓家，但是实时路况信息，还有待普及。路况的时效性非常强。路况信息约及时越有效。显而易见的是，如果信息服务商提供给你昨天的路况，估计大家除了不付钱之外，还会把他列入黑名单。实时的路况信息，对于改善物流企业的效率，降低运营成本显然有巨大的意义。
- 实时处理的要求，是区别大数据应用和传统数据仓库技术、BI 技术的关键差别之一。

稀疏但珍贵的价值

- 大数据无疑是有价值的。但是挖掘大数据的价值类似沙里淘金。譬如视频监控，每天产生 24 小时的视频数据，绝大部分都没有利用价值。可能是几秒镜头捕捉到某罪犯体貌特征，对公安部门而言就是弥足珍贵的。为了这几秒钟，必须保存全部的 24 小时。这也是大数据的一个典型特征。用术语来说，就是价值密度比较低。

图表8: 为了一点金子，需要保存全部沙子。



来源：国金证券研究所

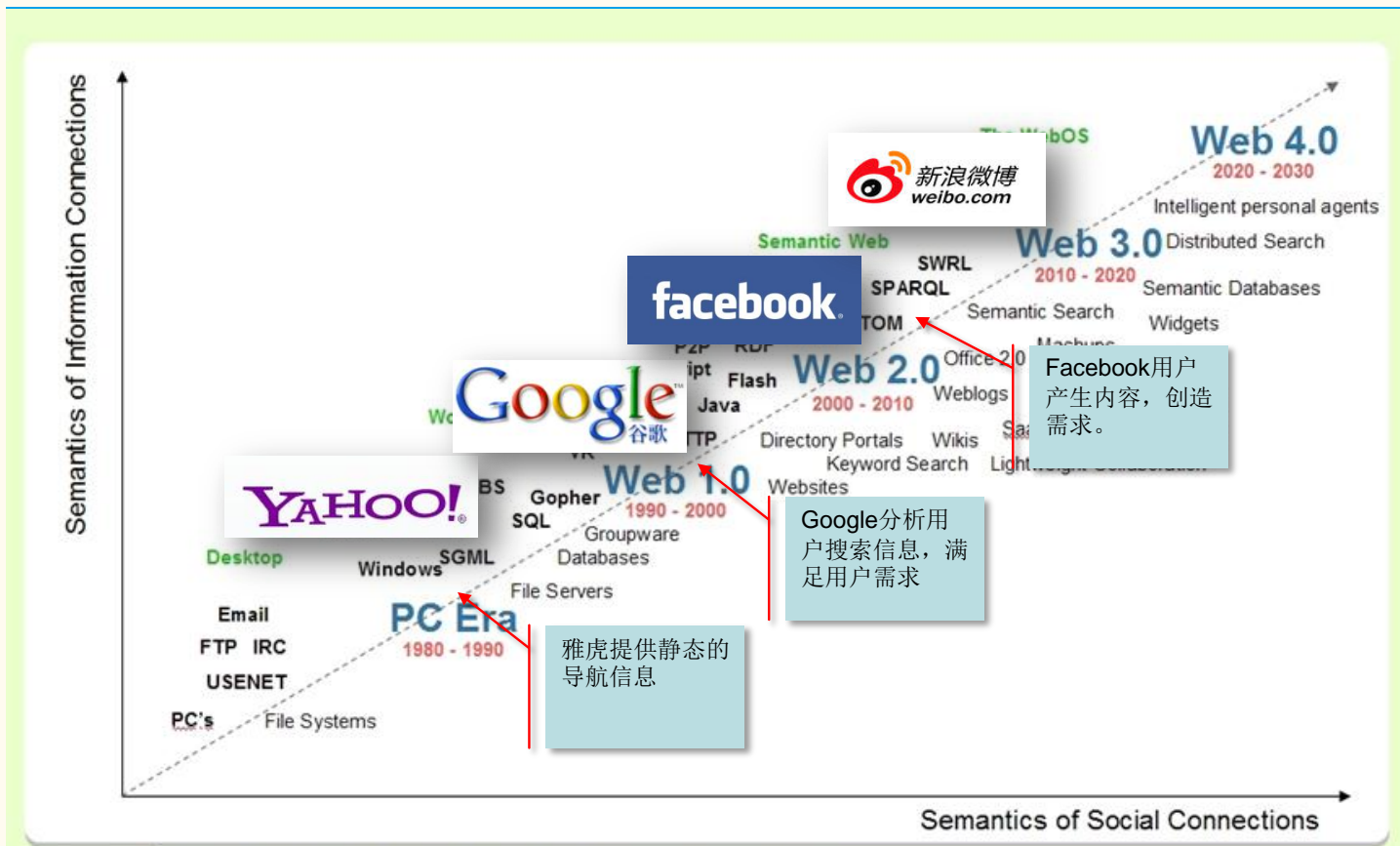
互联网企业演进逻辑和本质

- Google 是大数据技术的奠基人，但是他停留在满足用户需求这一步，我们难以将谷歌和传统企业联系起来，因此还称不上对传统企业颠覆性的影响。
- 发展到 Facebook 时代，我们发现网民和消费者的界限正在消弭，借助社会化传播的影响，消费者获得空前的地位和关注。研究、利用消费者在网上的行为，对于传统制造企业而言，无疑具备了预测未来的能力。大数据恰恰是为传统企业嫁接了互联网基因，为互联网企业深度介入供应链打开了方便之门。
- 本节将回溯互联网的发展历史，揭示互联网企业独领风骚的本质。

互联网的本质是理解用户，并将向智能化演进

- 互联网最为本质的一点，就是网民真实的展示了个人的想法，并能迅速为人所知。心理学家说人类的细微行为，直接暴露内心的想法。你可能对某商品横加指责，但反复端详了几分钟，那么你就可能是潜在的买家。所谓嫌货才是买货的人。网友在网络中的足迹、点击、浏览、留言直接反映他的性格，偏好，意愿。随时随地精准满足你圈定用户群的真实需求和潜在需求，就是互联网基因。至于快不快、创不创新是第二位的東西。

图表9 越来越真实的互联网，正越来越智能



来源：Radar Networks,国金证券研究所

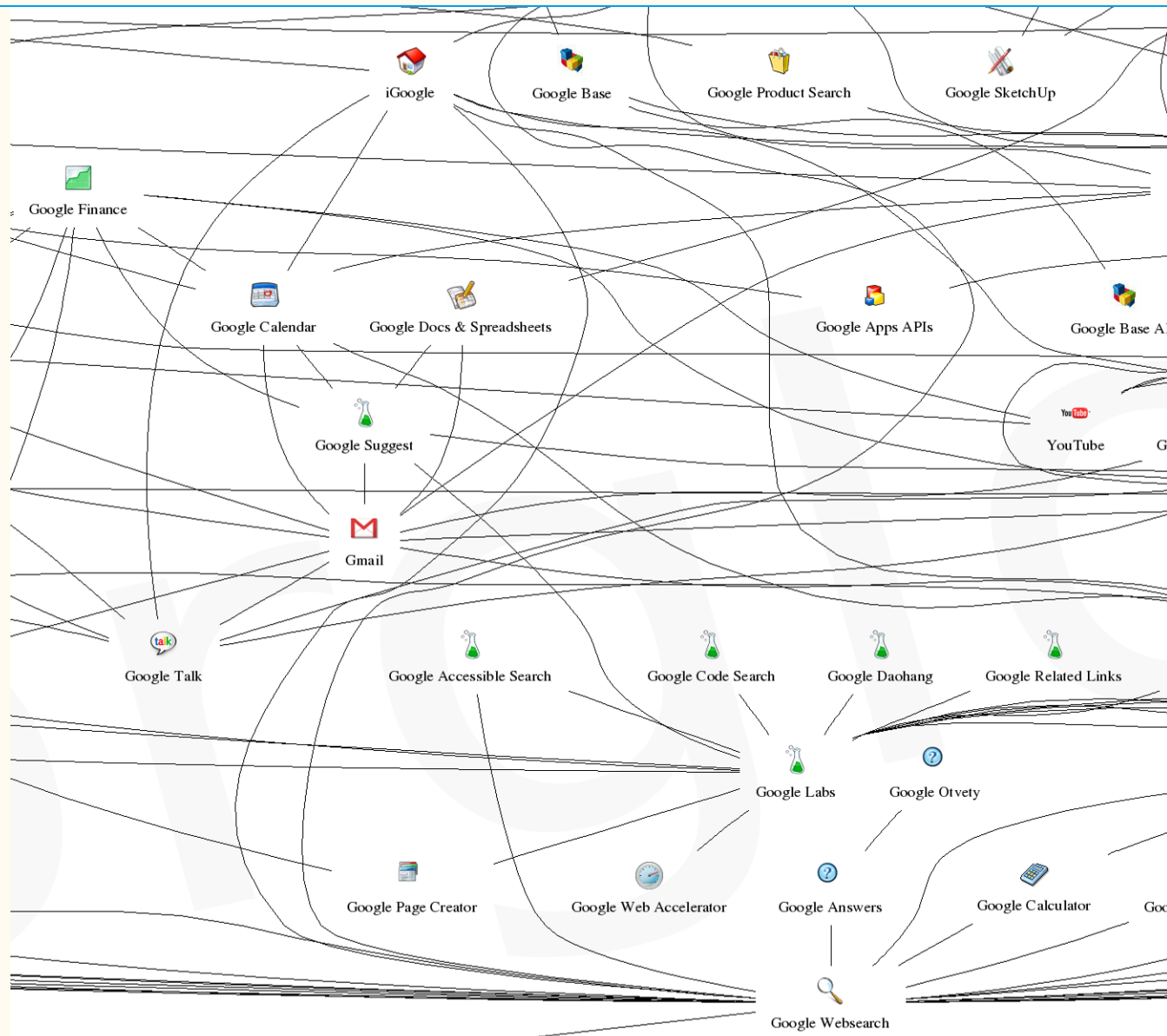
- 前瞻来看，随着互联网对网民的理解，网民对网络的反作用，互联网将变得越来越智能。它在满足你需求的同时，也在创造新的需求。前者的代表是 Google，后者的典型则是 Facebook。

Google 精准满足用户需求

- 谷歌是免费经济的最佳代表，对于个人用户而言，所有的应用都是免费的。Gmail、日历、照片、联系人、文档编辑等等，产品众多。有好事者整理了一副 Google 产品间的关系图。这幅图太大，我只是截取了部分贴在下面。
- 谷歌如何盈利呢？答案在于所有的软件应用都是在线的。用户在免费使用这些产品的同时，把个人的行为、喜好等信息免费的送给了 Google。因此 Google 的产品线越丰富，他对用户的理解就越深入，他的广告就越精准。广告的价值就越高。
- 这是正向的循环，谷歌好用的、免费得软件产品，换取对用户理解；通过精准的广告，找到生财之道。颠覆了微软卖软件拷贝赚钱的模式。成为互联网的巨擘。
- 尽管 Google 是互联网巨擘，也是“大数据”技术的奠基人，但是停留在满足用户需求这一步，我们难以将谷歌和传统企业联系起来，他们是广告

与广告主的关系。还称不上对传统企业颠覆性的影响。Google 在互联网圈子里面称雄。海尔们在传统市场里称霸。井水不犯河水。

图表10: Google 众多的产品线, 核心在于尽可能多掌握用户行为, 从而精确获取用户需求



来源: 国金证券研究所

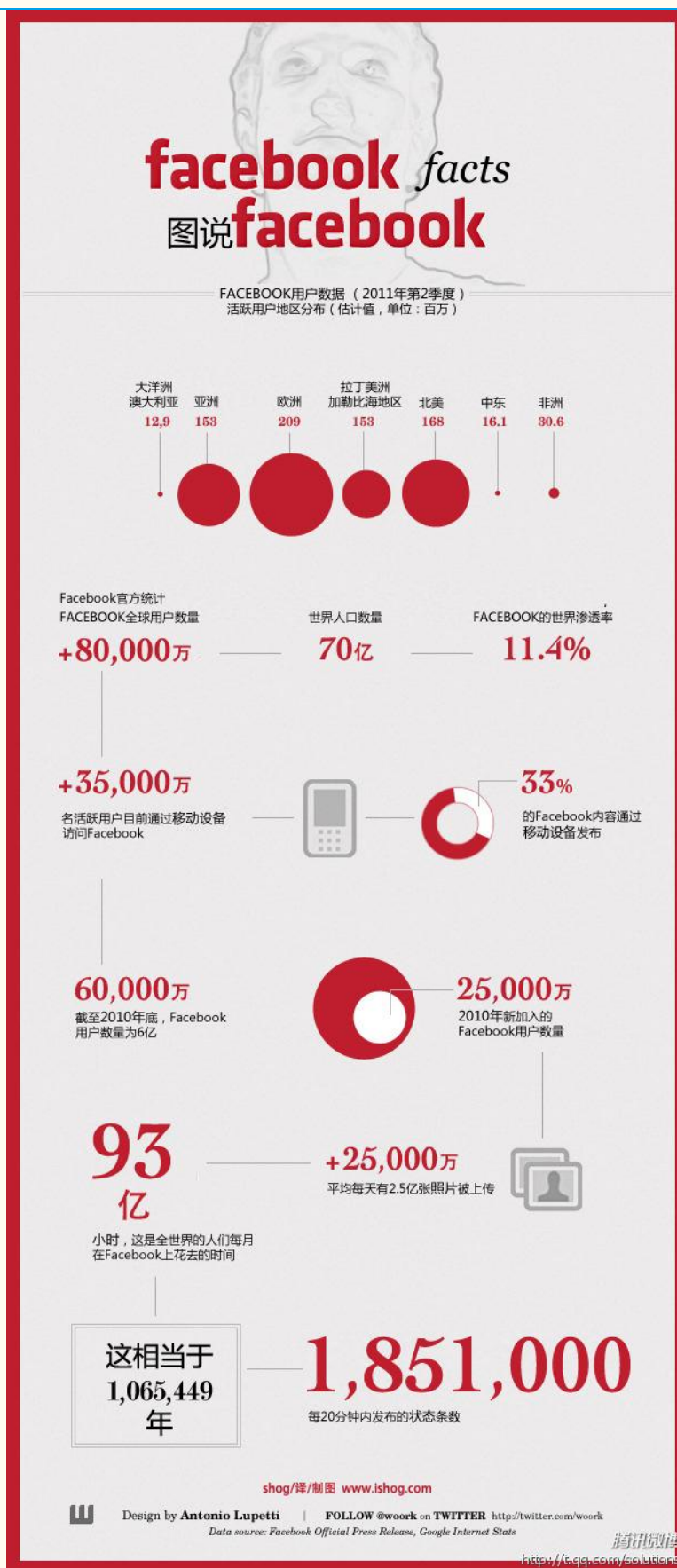
Facebook 创造需求——微力改变世界

“让这个世界上的人们自己组织起来, 才是最重要的事情”

——马克·扎克伯格

- 智者的思想总是能穿越时空! 我惊诧于托夫勒在《第三次浪潮》中的预言。它正在 Facebook 中成为现实。第三次浪潮的特征是小众、体验化, 而非第二次浪潮(工业革命)以来的大众化、标准化。Facebook 恰恰把“小众”们组织起来, 借助“社会认同”效应, 迅速放大他们的影响。甚至成为影响大选结果、影响组织决策的有力工具。具备了颠覆性的力量。
- 为了对 Facebook 有直观的印象, 我还是用一些数字描述这个帝国, 名副其实的帝国。截止到 2011 年 2 月份, 它拥有 8 亿用户, 分布在除南极洲外的所有大陆, 如果它是一个国家, 仅次于中国、印度而位列第三大人口大国; 3.5 亿用户通过移动设备访问 Facebook; 平均每天上传 2.5 亿张照片。

图表11: Facebook 概况



来源: 国金证券研究所

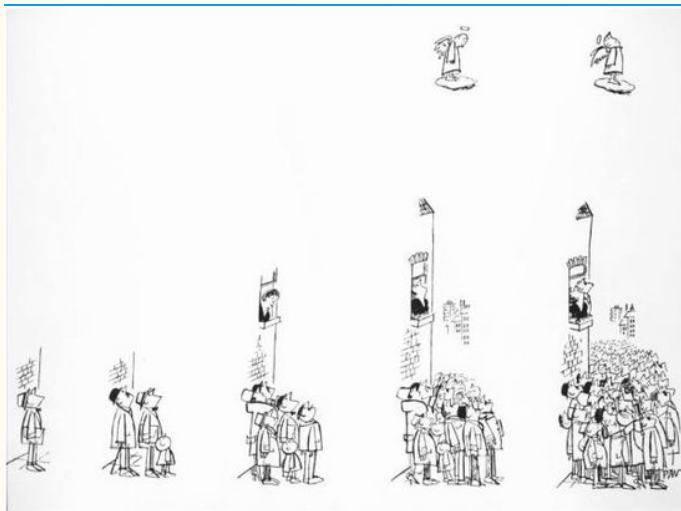
- Facebook 通过精心的控制与良好的用户体验，打造了一个实名帝国。用户在上面“生活”，产生大量的“内容”。因此 Facebook 不必像 Google 那样开发各种各样、丰富多彩的应用程序，而是吸引用户、第三方公司自主创造。
- 行文至此，开始区分两个概念：网民和消费者。网民泛指在网络上聊天、游戏、购物等等具有上网行为的自然人。消费者是相对于企业，尤其是制造企业的概念。

图表12：奥巴马与马克·扎克伯格



来源：国金证券研究所

图表13：社会认同效应



- Facebook 最强大的力量来自于“小众的社会认同效应”。独立的网民因为某种兴趣、爱好、观点而集聚在一起，通过消息沟通的机制，迅速传播，其他网民迅速跟进，短时间内形成类似“飓风”的传播效应，从而形成广泛的社会影响力。
- Facebook 颠覆了传统的信息传播路径，使几千年来，单中心、单向的传播方式，向多中心、网状裂变传播方式转变。对企业而言，不管是主动拥抱这个变化，还是被动接受，都宣告了一个时代的大幕真正开启，消费者站上了舞台中央。
- 但此消费者是数字化的，他生活在网络中，他并不会直接告诉你他的需求，企业必须自己去收集、分析、跟踪、对比他在互联网上留下的各种“足迹”、评论、图片、视频等。谁首先在海量数据中沙里淘金，谁首先嫁接互联网基因，谁将在这轮技术变革中，脱颖而出。这就是大数据时代的核心命题！

传统产品制造企业之殇

“你的客户是谁？”——德鲁克

“所有传统的产品公司都只能沦为这种新型用户平台级公司的附庸，其衰落不是管理能扭转的”——索尼前总裁出井伸之

- 传统企业很难获得他们用户的需求，更谈不上真实。曾经有服装企业想调查他们顾客的购买意愿，看那件衣服顾客拿起来了，哪件试穿过。又要安装摄像头，又要选择样本，没有几亿下不来，要想省钱的话，结果就会因为样本量太小而失去参考意义。但是在网上，顾客在网页停留时间，点击衣服款式，收藏的品类，最好哪些放到“购物车”中，都翔实的记录在网站中。可以使用所有样本，全面的分析用户需求，而且成本近乎于“零”。

- 出井伸之先生概括一针见血指出“新一代基于互联网 DNA 企业的核心能力在于利用新模式和新技术更加贴近消费者、深刻理解需求、高效分析信息并作出预判，所有传统的产品公司都只能沦为这种新型用户平台级公司的附庸，其衰落不是管理能扭转的。互联网的魅力就是“the power of low end””

对消费品制造的影响

- 任正非《让听得见炮声的人来决策》一文，其核心思想围绕一线人员重组了业务流程和组织架构。因为一线人员最了解客户需求。无独有偶，张瑞敏也提出在互联网时代，必须从大规模制造过渡到大规模定制。过去一个产品型号可以生产几万件到几百万件，而现在可能需要几十万个型号。现在许多工业装备已经具备了这样的能力，对企业的影响主要是在业务流程的再造和组织结构调整方面的挑战。
- 原来企业的结构是自定而下总经理、管理人员、员工的金字塔结构。在大数据时代必须是倒置的。最顶层是消费者，其次是员工、管理人员、总经理。

对软件开发模式的影响

- 传统的大型软件应用往往采取瀑布模型，这是软件工程的基本方法。首先要进行需求调研，接下来一系列的步骤包括设计、开发、测试、交付上线等，上一步没有完成，原则上不能进行下一步。严格采取瀑布模型周期会比较长。用户除了在需求阶段有发言权外，基本被排除在整个流程之外。而互联网时代，产品的推出往往以天计算，几天就有升级，吸收大量用户的反馈，迅速的完善产品。同样地需要重新思考软件开发的指导思想，这无疑也会引起业务流程和组织结构的变化。
- 互联网天然的面向海量用户，平台战略往往成了大型互联网企业的最终选择。一旦成为平台，在吸引海量用户的同时，汇聚大批的第三方独立的软件开发商、极客，从而围绕这个平台形成“生态系统”。就像房地产行业的万达。建一个城市广场，吸引众多业态，不通业态之间相互带来客流。不同的是，互联网的平台厂商不用像万达到处买地盖楼，只要添置几台服务器，就能无限的扩张疆域。

对企业信息化的影响

- 互联网大幅提升了企业接收订单的效率，对一般意义上的网站而言，就是大量的点击请求。但如果内部的处理流程依然依赖手工，无疑是一场灾难。十年前.com 泡沫，这也是原因之一。在用户端，看起来下单的效率提高了，但是网站后台处理依然靠人工干预、审核，最终的结果还是效率低下。现在我们看到一些购物网站推出的大型促销，譬如“秒杀”，对后台信息系统的处理能力是极大的考验。
- 再者，前文提过“获取用户的真实需求”是互联网的基因。用户不是通过语言告诉你需求，而是在一系列的行为中，不经意的透露了需求。这一系列的行为必须能够相关参照、关联，才能得出答案。解决这个问题的技术就是“大数据”。现在有些网站已经有数据分析师岗位。
- 还有，电商不是孤立存在的。商品由数量众多的各类供应商分别提供。用户下订单后，如果缺货怎么快速的通知供应商呢？要采购多少？什么时候运到？等等这些问题都是供应链系统需要处理的。客观上渠道商都会要求和供应商联网。这就促进了供应商的信息化。实时监控货物状态的要求，同样促进物流行业的信息化，催生了智能立体仓库，虚拟仓库的概念。

当传统企业嫁接大数据

- 传统企业难以了解消费者，而以新浪微博、Facebook 为代表的互联网公司天然的了解网民。当网民和消费者的交集爆发式扩张时，任何一家传统企业都难以忽略来自网民的影响；

- 当传统企业利用大数据，嫁接互联网基因，那么他将获得新的发展动力。互联网公司也通过与传统企业的合作，进一步增强其平台的粘性。获得源源不断的发展动力。
- 企业信息化部门将从一个辅助角色，一跃而为决策的力量，一切都将以精确的大数据分析为导向。
- 我们再下一章节，将详细说明大数据的内涵，和引发的变革。

大数据引发的变革

在开始本章之前，简要回顾上述三章的主要思路。在大数据概念提出之前，互联网企业沿着固有的脉络一路繁荣，网民主要是沟通、娱乐、消费。传统企业一直忙于供、研、产、销。基本上平行的发展，鲜有交集。而大数据是两者发展的交集，为互联网企业嫁接了“供应链”，为传统企业嫁接“互联网基因”，引发消费模式、制造模式、管理模式的巨大变革。对政府、对企业、对个人影响深远！

大数据的内涵

大数据不是云计算，是云计算的灵魂和升级方向。

- 云计算的核心是业务模式，本质是数据处理技术。如果失去对大数据理解和运用，云计算只是房地产的代名词。数据是资产，云为数据资产提供了保管、访问的场所和渠道。如何盘活数据资产，使其为国家治理、企业决策乃至个人生活服务，是大数据核心议题，也是云计算的灵魂和必然的升级方向。

图表14：大数据的构成，以数据融合、综合处理为方向



来源：国金证券研究所

大数据三个主要的数据来源

- 企业内部的经营交易信息，物联网世界中商品、物流信息，互联网世界中人与人交互信息、位置信息等等是大数据的三个主要来源。其信息量远远

超越了现有企业 IT 架构和基础设施的承载能力，其实时性要求则大大超越现有的计算能力。

- 企业内部的信息主要包括联机交易数据和联机分析数据。就数据本身的格式来讲，是结构化的，通过关系型数据进行管理和访问。这些数据价值秘密高，但都是历史的、静态的数据。通过对这些数据的分析，我们只能过去发生了什么，很难说未来将发生什么。
- 来自于社交网站，新浪微博、Facebook 等的的数据，是大量的、鲜活的，代表了一个个具体网民的想法，反映了他们想做的事情。这些数据价值密度低，但事关未来。

大数据赋予我们洞悉未来的能力

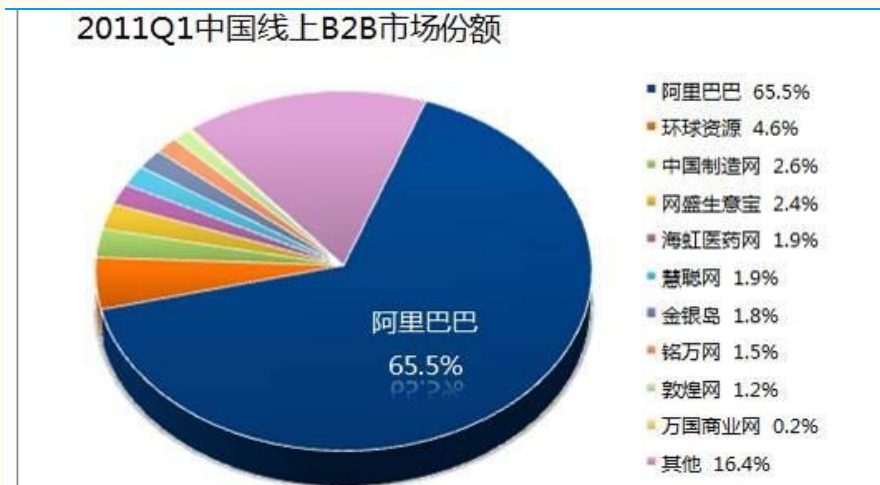
- 这两类数据的有效融合将是大势所趋。因为它让我们获得洞悉未来的能力。
- 话说马云成功预测 2008 年经济危机，并帮助他成千上万的中小制造商准备过冬的粮食。让马云赢得了崇高的声誉。
 - “2008 年初,阿里巴巴平台上整个买家询盘数急剧下滑，欧美对中国采购在下滑。海关是卖了货,出去以后再获得数据;而我们提前半年时间从询盘上推断出世界贸易发生变化了。”

图表15: 马云的判断来自于数据分析



来源：易观国际 国金证券研究所

图表16: 阿里巴巴的规模保障了数据的代表性

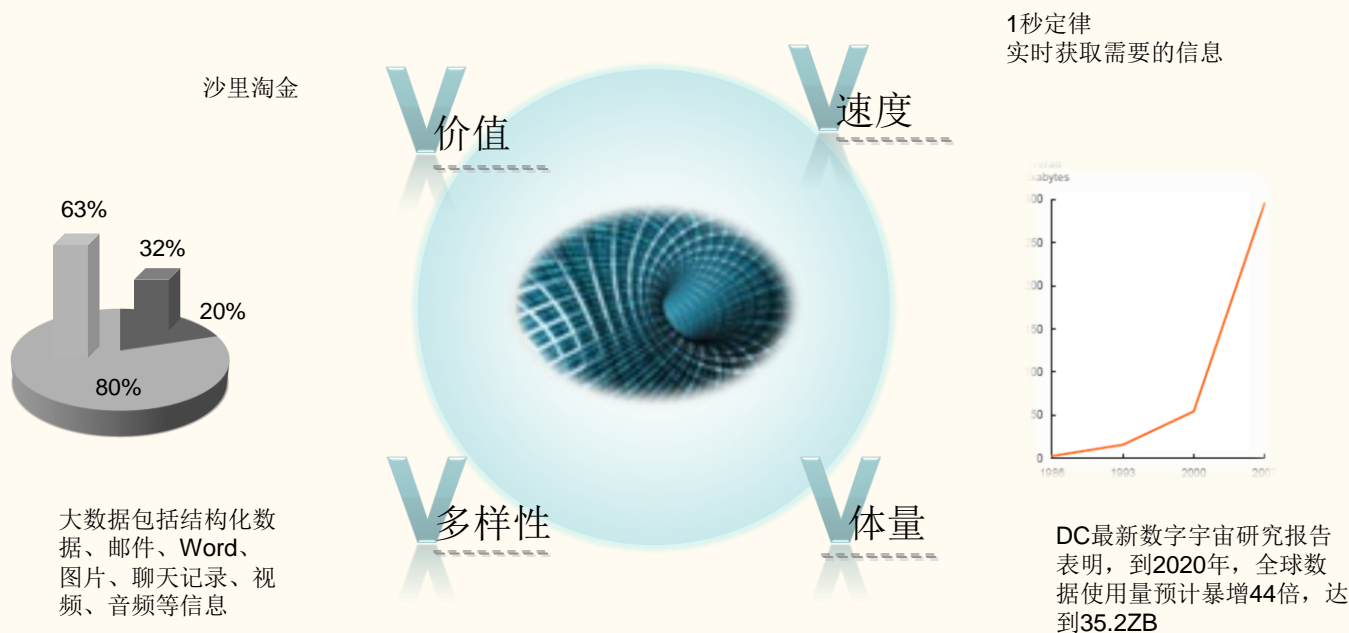


- 马云对未来的预测，是建立在对用户行文分析的基础上。通常而言，买家在采购商品前，会比较多家供应商的产品，反映到阿里巴巴网站统计数据中，就是查询点击的数量和购买点击的数量会保持一个相对的数值，综合各个维度的数据可建立用户行为模型。因为数据样本巨大，保证用户行为模型的准确性。因此在这个案例中，询盘数据的下降，自然导致买盘的下降

大数据四个主要特征

- 大数据有 4 个典型的特征。所谓的 4 个“V”，多样性(variety)、体量(volume)、速度(velocity)和价值(Value)。多样性是指，数据应包含结构化的和非结构化的数据。体量是指聚合在一起供分析的数据量必须是非常庞大的。而速度则是指数据处理的速度必须很快。

图表17: 大数据的四个主要特征



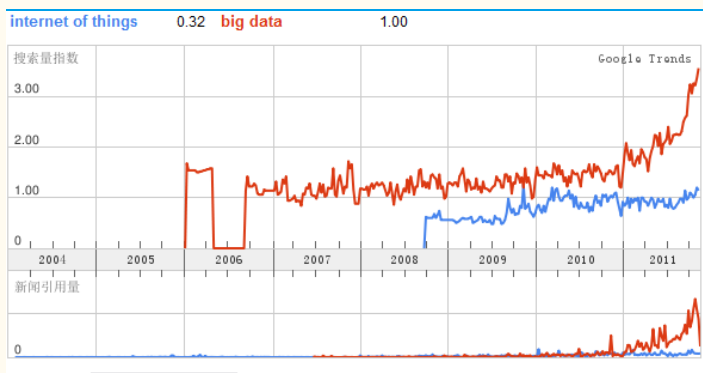
来源: 国金证券研究所

大数据技术发展的脉络和现状

大数据是迈向智慧世界的自然演进

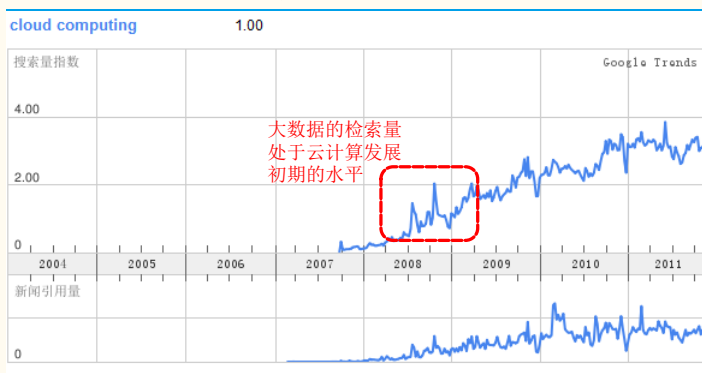
- 我们沿用托夫勒《第三次浪潮》一书中的提法。如果说 IBM 的主机拉开了信息化革命的大幕,那么“大数据”才是第三次浪潮的华彩乐章。
- 依据大数据的预测逻辑,我们利用 Google 趋势工具,发现在全球范围内,“Big data”搜索指数呈现陡峭上升的趋势,其热度超过物联网,处于云计算 2008 年前后的关注水平。传统的 IT 巨头和新兴的行业巨擘,在大数据的应用方面。

图表18: “大数据”关键词搜索量处于激增阶段



来源: Google 国金证券研究所

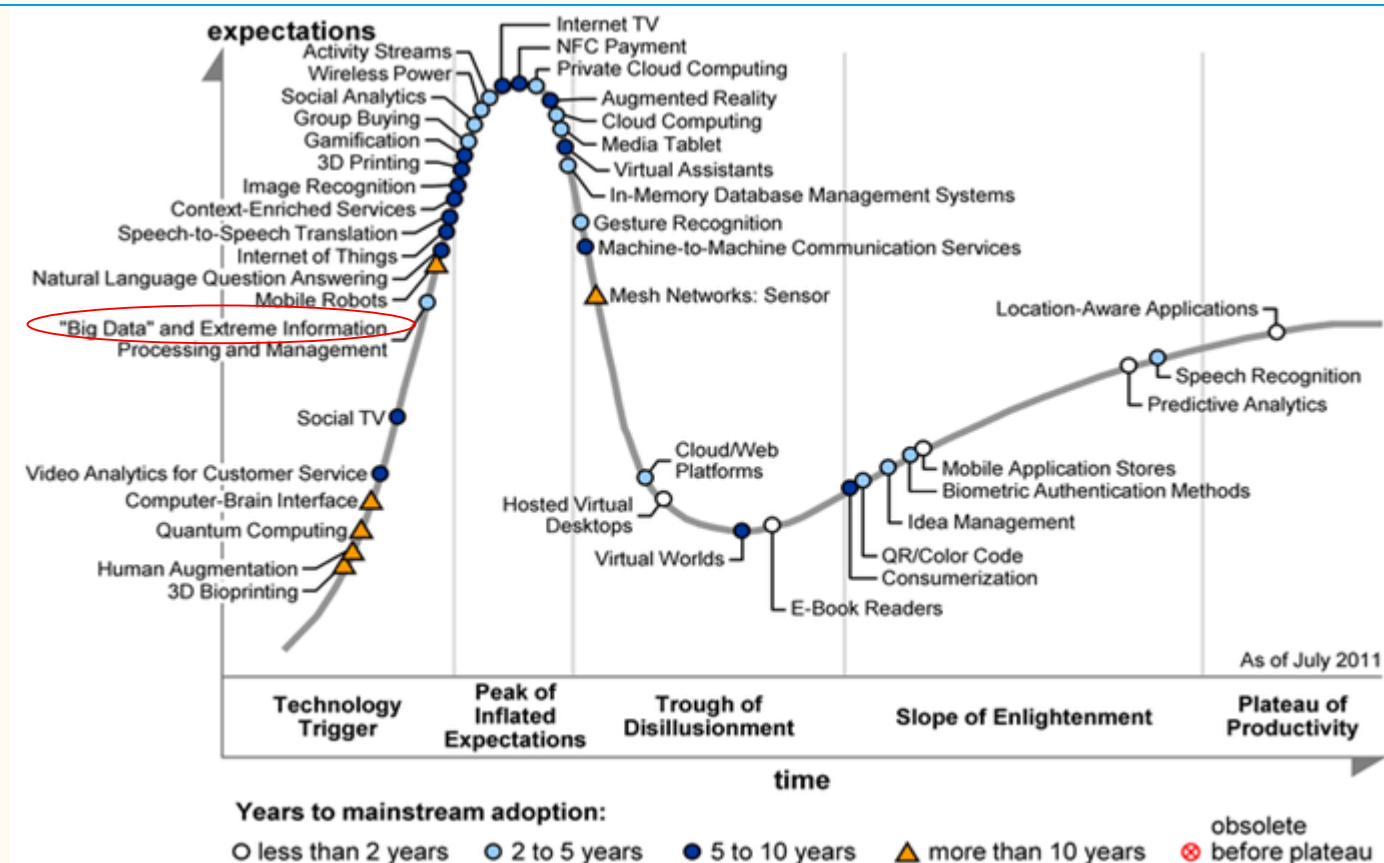
图表19: 相当于云计算 2008 年左右的水平



- 根据 Gartner 的预测,2011 年,大数据技术处于高速发展期。不断取得技术上得突破,产品密集发布或其他能产生重大产业利益的项目快速大量出现。

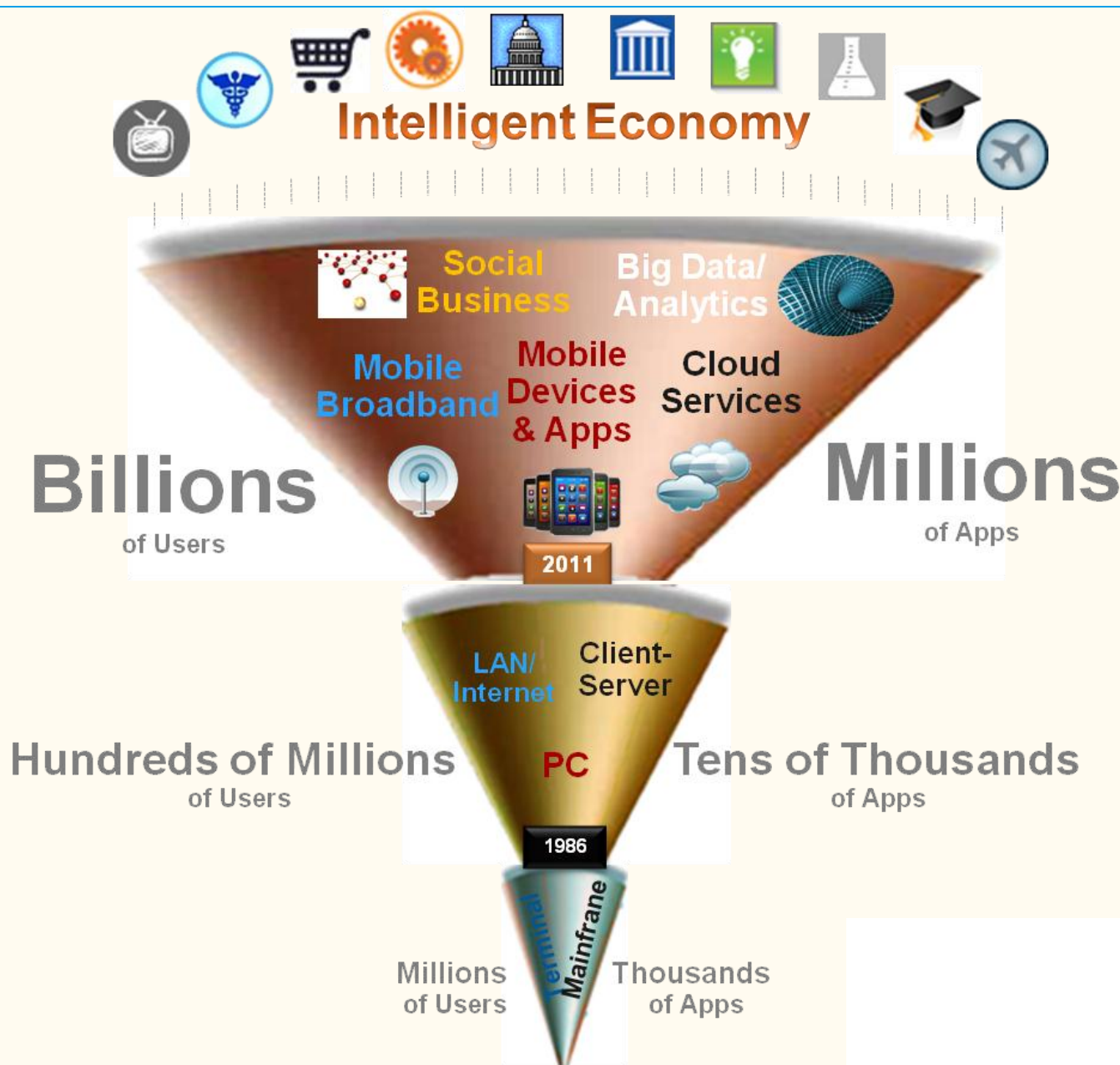
- 今年 11 月 30 号，业界召开了大数据研讨会新浪、百度、阿里巴巴、腾讯等公司与会，其中像 Hadoop、NoSQL、数据分析与挖掘、数据仓库、商业智能、开源云计算架构等最受关注的技术热点。

图表20: Gartner 2011 年技术成熟度曲线，大数据处于高速发展期



来源：Gartner 国金证券研究所

图表21：技术演进历史揭示未来是大数据驱动的智慧型经济模式



来源：国金证券研究所

IT 产业巨头的大数据技术及发展简介

- 我们选择 EMC、IBM、Oracle、微软来观察大数据技术的发展，暨在各行各业的应用情况。总体而言。我国大数据的利用情况，落后北美 5 年。因此我国产业具备后发优势，投资空间巨大。
- 这些产业巨头无一例外开展了并购行动。在大数据时代谋篇布局。

图表22：巨头在大数据时代的业务布局

公司	时间	收购
EMC	2010年7月	收购数据库软件供应商 Greenplum，花费 3 亿美元
EMC	2009年7月	收购数据复制解决方案提供商 Data Domain，花费 24 亿美元
EMC	2009年起	陆续收购 Archer Technologies, SourceLabs, FastScale Technology, Configuresoft, and Varonis Systems
IBM	2010年9月	收购数据库分析供应商 Netezza 公司，花费 17 亿美元
IBM	2010年10月	收购网络分析软件供应商 Coremetrics
IBM	2009年10月	收购数据分析和统计软件提供商 SPSS，花费 12 亿美元
IBM	2009年1月	收购业务规则管理软件供应商 ILOG，花费 3 亿 4 千万美元
IBM	2007年	花费 20 亿美元收购商务智能软件供应商 Cognos
SAP	2007年	68 亿美元收购了全球商业智能软件老大 Business Object
Oracle	2008年11月	33 亿美元收购商业智能解决方案提供商海波龙 (Hyperion) 公司
Oracle	2009年7月	收购专注于数据复制和实时数据集成解决方案的私人企业 GoldenGate Software
HP	2011年	花费 100 亿美元收购英国软件公司 Autonomy
微软	2006年	收购商业智能软件公司 ProClarity
微软	2008年	收购数据仓库产品厂商 DATAllegro

来源：国金证券研究所

■ EMC 斩获了纽交所和 Nasdaq。

图表23：EMC 斩获了纽交所和 Nasdaq

背景	收购
<p>EMC的大数据解决方案专注于使组织更有效地使用他们从不同来源产生的数据，包括网络上，网页上，消费者，监控系统 and 传感器。</p> <p>EMC的数据计算产品事业部正在开发分析工具以解决大数据现象。</p> <p>EMC的大数据解决方案包括40多个产品。</p>	<ul style="list-style-type: none"> 2010年7月收购数据库软件供应商Greenplum，花费3亿美元 2009年七月收购数据复制解决方案提供商Data Domain，花费24亿美元 不一定和大数据完全相关，EMC从2009年起收购了Archer Technologies, SourceLabs, FastScale Technology, Configuresoft, and Varonis Systems。
提供的产品及服务	合作伙伴
<ul style="list-style-type: none"> EMC Greenplum的数据计算设备，是使用Greenplum的大规模并行处理（MPP）架构的一个集成的数据仓库系统 最近公布了 EMC Greenplum的社区免费版，同时还公布了免费的分析算法和数据挖掘工具 EMC Greenplum的社区免费版包括 Greenplum Database CE, an MPP database product <ul style="list-style-type: none"> MADlib, 开放源码的分析算法库 Alpine Miner, 提供快速建模分析能力的第三方分析工具 	<ul style="list-style-type: none"> 2010年九月与Cloudera结盟，Cloudera是提供基于Hadoop数据管理软件的领先企业 2010年12月与Havas Digital合作，来增强Artemis Analytics 实验室的研究能力，用于研发处理大规模大数据的私有云技术。
	客户
	<ul style="list-style-type: none"> EMC/ Greenplum在全球拥有超过200家客户，包括NASDAQ, NYSE, hitwise, O'Reilly, and Sun 美国亚利桑那州立大学地球和空间探索学院（SESE），使用EMC的Isilon向外扩展NAS，以确定理想的着陆点，还有月球表面的永久阴影和照明领域。

来源：MGI 国金证券研究所

■ IBM 依然保持银行、电信行业的强势地位

图表24: IBM 依然保持银行、电信行业的强势地位

背景	收购
<p>IBM的策略是提供一个全面的方法来解决前所未有的信息爆炸提出的挑战,因为信息量无论在流量、种类、速度还是活力上都是爆炸式增长</p> <p>IBM一直致力于扩大对包括数据仓库中的大数据、信息流和结构化数据的分析</p>	<p>在过去四年中, IBM已经投入超过120亿美元进行了23项相关并购, 其中包括:</p> <ul style="list-style-type: none"> 2010年9月收购数据库分析供应商Netezza公司, 花费17亿美元 2010年10月收购网络分析软件供应商Coremetrics 2009年10月收购数据分析和统计软件提供商SPSS, 花费12亿美元 2009年1月收购业务规则管理软件供应商ILOG, 花费3亿4千万美元 2007年花费20亿美元收购商务智能软件供应商Cognos
提供的产品及服务	合作伙伴
<ul style="list-style-type: none"> IBM大数据提供的服务包括数据分析, 文本分析, 蓝色云杉(混搭供电合作的网络平台);业务事件处理; IBM Mashup Center的计量, 监测, 和商业化服务(MMMS) IBM的大数据产品组合中的最新系列产品的InfoSphere bigInsights, 基于Apache Hadoop。 该产品组合包括 打包的Apache Hadoop的软件和服务,代号是bigInsights核心,用于开始大数据分析 <ul style="list-style-type: none"> 软件被称为bigsheet, 软件目的是帮助从大量数据中轻松、简单、直观的提取、批注相关信息 为金融, 风险管理, 媒体和娱乐等行业量身定做的行业解决方案 	<ul style="list-style-type: none"> 2010年5月, 选择与使用开源技术设计海量数据分析软件的Apache Hadoop合作
	客户
	<ul style="list-style-type: none"> 全球25000多分析客户, 其中包括22个全球前24大商业银行, 18个前22电信运营商和11个美国排名前12位的专业零售商 赢得新客户, 包括Avis, FUJIFILM Imaging Colorants,西班牙的社会服务机构, and Hildebrand to extract insights from data 2010年8月, 北卡罗来纳州使用IBM大数据技术, 以简化搜索和匹配潜在的大学研究项目的投资和合作机会的过程 其他客户: Hertz, 大英图书馆

来源: MGI 国金证券研究所

■ 微软在中小企业市场发力

图表25: 微软在中小企业市场发力

背景	收购
<p>微软提供了高性能计算能力, 并在2005年靠Windows Compute Cluster服务器进入相关市场</p> <p>最近, 微软的HPC部门开发了该公司的Dryad 并行处理技术社区技术预览(CTP), 第一步是向Windows HPC Server的用户提供处理大数据工具</p>	<ul style="list-style-type: none"> 无
提供的产品及服务	合作伙伴
<ul style="list-style-type: none"> 微软提供的Windows HPC Server2008是一种基于Windows Server技术的高性能计算解决方案, 这个方案可以在HPC环境内部署、管理、调度监控软件工具 微软开发的Dryad通用的分布式计算引擎和抽象化DryadLINQ简化了基于Dryad的应用程序进程 微软还开发了基本局部比对搜索工具(BLAST)用来帮助进行生命科学的研究, 是在Azure上运行搜索引擎来扫描化合物的大型数据库以寻求匹配数据 微软宣布他在日本的超级计算机Tsubame 2打破了界限, 能够同时运行Windows HPC Server2008 R2和Linux 	<ul style="list-style-type: none"> 2011年1月与惠普(具体而言是HP数据库综合应用部门)合作目标是开发了一系列能够提升生产力和提高决策速度的设备。
	客户
	<ul style="list-style-type: none"> HPC Server客户包括Lloyds TSB支持风险管理计算 联邦机构

来源: MGI 国金证券研究所

■ Oracle——ExaData 凶猛

图表26: Oracle 扮演 IBM 狙击手的角色

背景	收购		
甲骨文大数据提供的数据库和数据库软件主要用于配合Sun的硬件，特别是它的最高端服务	<ul style="list-style-type: none"> 2009年7月收购专注于数据复制和实时数据集成解决方案的私人企业GoldenGate Software 		
提供的产品及服务	合作伙伴		
<ul style="list-style-type: none"> 甲骨文在大数据市场有三个核心产品 Exadata: 软硬件结合提供的解决方案专注于大型数据处理，主要适用于石油和天然气勘探领域(附注，年运营收入15亿美元) Exalogic:通过云计算补充Exadata，一个Exalogic可以每秒处理一百万个HTTP请求，两个就可以处理全球的Facebook。 Sun SPARC Superclusters: 在RAC环境运行甲骨文数据库的完整基础设施解决方案 最近，甲骨文为它的Exadata X2-8增加了5TB的固态硬盘。 Recently, Oracle added 5TB of solid state disk to its Exadata X2-8 offering (固态硬盘是一种快速存储设备，常用做大型数据的缓存) 	<div>客户</div> <ul style="list-style-type: none"> 美洲银行正在测试两个Exadata系统智能商务部分中人力资源和培训的功能。 <p>Exadata 的客户包括</p> <table border="0"> <tr> <td> <ul style="list-style-type: none"> Allegro Group Banca Transilvania BNP Paribas Garanti Bank Hong Kong housing authority Commonwealth Bank Dialog Semiconductor </td> <td> <ul style="list-style-type: none"> Fast Retailing Hotwire, Inc. LinkShare, Inc. SK Telecom Softbank Targetbase Turkcell </td> </tr> </table>	<ul style="list-style-type: none"> Allegro Group Banca Transilvania BNP Paribas Garanti Bank Hong Kong housing authority Commonwealth Bank Dialog Semiconductor 	<ul style="list-style-type: none"> Fast Retailing Hotwire, Inc. LinkShare, Inc. SK Telecom Softbank Targetbase Turkcell
<ul style="list-style-type: none"> Allegro Group Banca Transilvania BNP Paribas Garanti Bank Hong Kong housing authority Commonwealth Bank Dialog Semiconductor 	<ul style="list-style-type: none"> Fast Retailing Hotwire, Inc. LinkShare, Inc. SK Telecom Softbank Targetbase Turkcell 		

来源: MGI 国金证券研究所

大数据伴生的大机遇

“这是一个伟大的机会，我明白该把自己的钱投向哪里”

——前网景公司创始人马克·安德森

产业规模

- 大数据将在政府公共服务、医疗服务、零售业、制造业、以及涉及个人位置服务等领域得到广泛应用，并产生巨大的社会价值和产业空间。以麦肯锡评估西方产业数据为例，美国医疗服务业价值约 3000 亿美元，欧洲公共部门管理支出约为 3500 亿欧元，GIS 服务商收入将在 1000 亿没有左右，而位置服务中个人最终的价值将达到 7000 亿美元。大数据将使得美国零售业净利润增长水平可达到 60%。降低制造业产品开发、组装成本达 50%。详见下图。

图表27：大数据的市场空间及对社会的贡献

海量数据可以在各个部门创造重大财物价值

美国医疗服务业 <ul style="list-style-type: none"> • 每年价值3000亿美元 • 大约0.7%的年生产率增长 	欧洲公共部门管理 <ul style="list-style-type: none"> • 每年价值2500亿欧元（约3500亿美元） • 大约0.5%的年生产率增长 	全球个人位置数据 <ul style="list-style-type: none"> • 服务提供商收入1000亿美元或以上 • 最终用户价值达7000亿美元
美国零售业 <ul style="list-style-type: none"> • 可能的净利润增长水平为60%或以上 • 0.5~1.0%的年生产率增长 	制造业 <ul style="list-style-type: none"> • 产品开发、组装成本降低达50% • 运营资本降低达7% 	

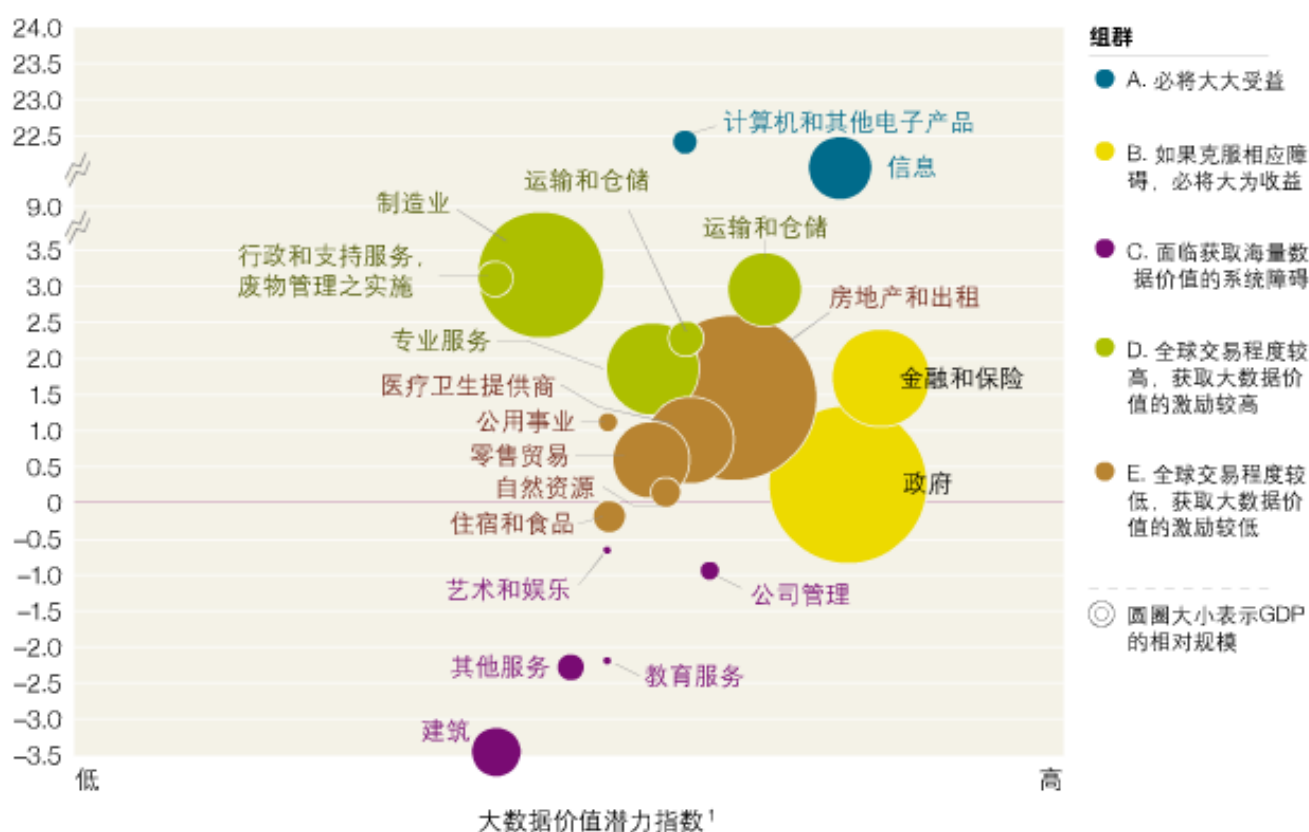
来源：麦肯锡全球研究院 国金证券研究所

企业的投入

- 传统企业如果有效利用大数据，需要重构其 IT 架构，投资包括租用足够空间、购买分析软件，并聘用合适的人才。第一波受益者将是具备提供大数据服务能力信息服务企业。

图表28：计算机和信息服务产业将在大数据浪潮中首先获益

美国2000—2008年历史生产率增长（百分比）



来源：麦肯锡全球研究院 国金证券研究所

政府可能采取的行动

- 重视应用大数据技术，盘活各地云计算中心资产。把原来大规模投资产业园，物联网产业园从政绩工程，改造成智慧工程。
- 在安防领域，应用大数据技术，提高应急处置能力和安全防范能力。
- 在民生领域，应用大数据技术，提升服务能力和运作效率，以及个性化的服务。如医疗、卫生、教育等部门。
- 在金融、电信领域，数据分析一直得到极大的重视。但囿于存储能力和计算能力的限制，局限在交易型数据的统计分析。据我们了解，部分银行已经开始用户行为分析。无疑，这两个领域将是大数据施展拳脚的舞台。
- 政府投入将形成示范效应，大大推动大数据的发展。

大数据的投资标的

图表29：投资标的一览表

代码	股票名称	11年EPS	12年EPS	13年EPS	11年P/E	12年P/E	11-12PEG	股价	11-12增长	12-13增长	年初至今相对涨跌幅%	一年相对涨跌幅%
002410.SZ	广联达	1.05	1.55	2.25	33.99	23.03	0.71	35.69	47.62%	45.00%	-27.83	-17.61
002405.SZ	四维图新	0.65	0.83	1.15	36.02	28.20	1.30	23.41	27.69%	38.00%	-53.67	-46.14
002368.SZ	太极股份	0.67	0.94	1.29	32.69	23.35	0.82	21.90	40.00%	38.00%	-21.14	-8.82
300036.SZ	超图软件	0.53	0.79	1.10	35.58	24.00	0.74	18.90	48.24%	40.00%	1.13	-5.09
002268.SZ	卫士通	0.55	0.78	1.10	45.85	31.91	1.05	25.04	43.66%	40.00%	-17.74	-17.18
600588.SH	用友软件	0.70	0.93	1.26	19.86	14.93	0.60	13.90	33.00%	35.00%	-15.59	-20.03
002308.SZ	威创股份	0.40	0.60	0.84	98.75	65.83	1.98	39.50	50.00%	40.00%	-18.50	-21.34
300229.SZ	拓尔思	0.67	0.88	1.17	44.31	33.57	1.38	29.69	32.00%	32.00%	79.17	79.17
300170.SZ	汉得信息	0.58	0.82	1.10	39.81	28.04	0.95	23.05	41.97%	34.18%	9.60	9.60
002279.SZ	久其软件	0.59	0.72	1.05	37.22	30.50	1.69	21.96	22.03%	45.83%	-24.89	-22.44
300166.SZ	东方国信	1.552	2.372	3.458	32.40	21.20	0.61	50.29	52.84%	45.78%	6.00	6.00
601519.SH	大智慧	1.464	1.919	-	9.39	7.16	0.30	13.74	31.08%	-	-45.39	-45.39
002230.SZ	科大讯飞	1.461	2.141	3.44	24.38	16.64	0.52	35.62	46.54%	60.67%	-32.18	-27.40

来源：国金证券研究所

预告：

在下一篇报告中，我们将讨论大数据产业链分工情况，探讨大数据技术提供商、系统集成商、咨询公司、软件公司在大数据时代的发展之道，揭示不同的行业可能孕育的投资机会。尤其值得关注的一个现象是终端提供商向数据运营商的演进路径。在大数据时代，提供终端展示产品、集中控制产品的公司亦有大机遇。还将紧密关注政府、行业动态，研判发展趋势。并解答路演过程中，投资者关注的问题。

参考文献：

- 【1】麦肯锡《Big data: The next frontier for innovation, competition, and productivity》2011年5月 James Manyika、Michael Chui 等
- 【2】麦肯锡《大数据展望》2011年3月 Nidhi Aggarwal、William Forrest 等
- 【3】托夫勒《第三次浪潮》
- 【4】大卫·柯克帕特里克《Facebook 效应》
- 【5】克里斯·安德森《长尾理论》
- 【6】克里斯·安德森《免费：商业的未来》
- 【7】亚当·斯密《国富论》
- 【8】吴军《浪潮之巅》

特别声明:

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证，对由于该等问题产生的一切责任，国金证券不作出任何担保。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。本报告亦非作为或被视为出售或购买证券或其他投资标的邀请。

证券研究报告是用于服务机构投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告反映编写分析员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，且收件人亦不会因为收到本报告而成为国金证券的客户。

本报告仅供国金证券股份有限公司的机构客户使用；非国金证券客户擅自使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

上海

电话: (8621)-61356534

传真: (8621)-61038200

邮箱: researchsh@gjzq.com.cn

邮编: 201204

地址: 上海浦东新区芳甸路 1088 号紫竹国际大厦 7 楼

北京

电话: 010-6621 6979

传真: 010-6621 5599-8803

邮箱: researchbj@gjzq.com.cn

邮编: 100032

地址: 中国北京西城区金融街 27 号投资广场 B 座 4 层

深圳

电话: 0755-33516015

传真: 0755-33516020

邮箱: researchsz@gjzq.com.cn

邮编: 518000

地址: 深圳市福田区福中三路 1006 号诺德金融中心 34B