

2015

大数据领域的创业思考

刘金玲

2015年9月12日

我们眼里的大数据



大数据最直接的意义就是让“随机性”的事务变得可提前预测；
以前做决定是“拍脑袋”决定，现在，做决定是根据数据结果；

目录

1 创业大环境

2 行业现状

3 风险预警

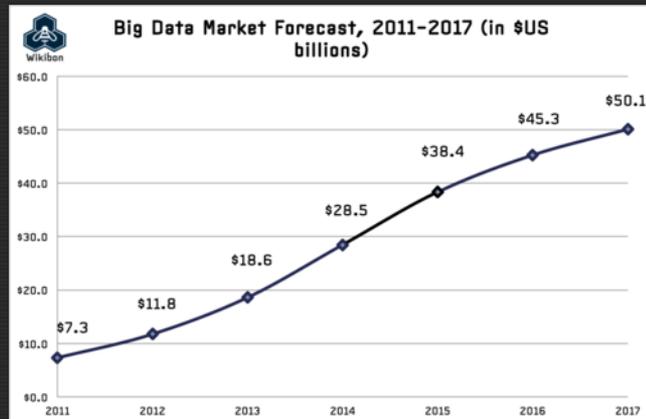
4 创业机会

PART ONE

2014年中国大数据市场规模达到767亿元，同比增长27.83%。
预计到2020年，中国大数据产业市场规模将达到8228.81亿元。

——贵阳大数据交易所《2015年中国大数据交易白皮书》

1 大数据市场规模巨大



根据Wikibon的报告，美国大数据产业的市场规模在2017年将达到**500亿美元**，这意味着大数据将为美国的很多行业带来千亿美元的价值。

根据易观智库7月30日发布的《中国行业大数据应用市场专题研究报告2015（简版）》报告来看，到2018年，中国大数据市场营销

规模将达到**258.6亿人民币**，环比增长率为37.2%。

规模达到

415亿美元

2 政策好

国务院总理李克强8月19日主持召开国务院常务会议，会议通过《关于促进大数据发展的行动纲要》，强调：

一要推动政府信息系统和公共数据互联互通，消除信息孤岛；

二要顺应潮流引导支持大数据产业发展，以企业为主体、以市场为导向，加大政策支持，着力营造宽松公平环境，建立市场化应用机制，深化大数据在各行业创新应用，催生新业态、新模式，形成与需求紧密结合的大数据产品体系，使开放的大数据成为促进创业创新的新动力。

三要强化信息安全保障，完善产业标准体系，依法依规打击数据滥用、侵犯隐私等行为。



表 4：各地方政府的大数据支持政策

地方	政策名称/措施
贵州	贵州省出台《关于加快大数据产业发展应用若干政策的意见》、《贵州省大数据产业发展应用规划纲要（2014-2020年）》。贵州省高校招生将优先考虑“大数据”专业 贵阳市发布《贵阳大数据产业行动计划》、《贵阳市关于加快推进大数据产业发展的若干意见》 贵安新区出台《贵安大数据产业基地发展规划》
杭州	发布《杭州市政务数据共享开放指导意见》
上海	发布《关于促进本市互联网金融产业健康发展的若干意见》，促进公共信用信息、金融信用信息、社会信用信息互动共用 经信委印发《2014年度上海市政府数据资源向社会开放工作计划》
武汉	《武汉市大数据产业发展行动计划（2014-2018）》 为鼓励企业使用大数据、从事相关开发研究，武汉出台“一揽子”扶持政策，每年投入2亿元，对重点项目进行补贴 打造云计算、大数据产业领域完整生态环境。在武汉未来科技城建设武汉大数据研究院、华中大数据交易市场
广东	2014年2月广东省政府印发的《广东省经济和信息化委员会主要职责内设机构和人员编制规定》，广东省大数据管理局为广东省经济和信息化委员会的21个内设机构之一 《广东省大数据发展规划（2015-2020年）》征求意见稿 广东省公布2014年首批推荐大数据应用示范项目名单
北京	打造京津冀大数据走廊
洛阳	出台《洛阳市电子信息产业行动计划》，建设国内一流大数据中心
湘潭	发布《关于建立数据资源管理制度的通知》，加强数据资源建设
湖北	发布《湖北省北斗卫星导航应用产业发展规划（2014-2020）》，推动大数据在交通领域应用发展

资料来源：贵阳大数据交易所，申万宏源研究

3

资本关注热

时间	公司名称	公司类型	融资阶段	融资金额	领头方
2015/1/13	永洪科技	商业智能分析	A轮	\$ 5000万	经纬创投
2015/1/17	九次方大数据	金融大数据分析	未知	¥ 2亿	博信资本领投
2015/3/2	麦客CRM	信息、数据收集	A轮	\$ 400万	红杉资本
2015/3/3	闪银奇异	金融大数据、信用评估	B轮	\$ 2000万	海纳亚洲创投基金
2015/3/25	艾曼科技	娱乐影视大数据分析	A轮	¥ 6000万	荣联创富基金和创势资本
2015/3/30	DataEye	大数据服务商	A轮	¥ 5000万	亚商资本，深创投跟投
2015/4/9	UCloud	数据存储、云计算	C轮	\$ 1亿	君联资本
2015/5/22	同盾科技	大数据反欺诈	B轮	\$ 3000万	启明资本
2015/7/28	华云数据	大数据存储	C轮	\$ 1亿	海通投资
2015/8/28	数联铭品	金融大数据分析	B轮	¥ 1亿	鼎晖投资
2015/8/31	时趣Social Touch	大数据营销	D轮	近亿美元	联想控股
2015/9/8	百分点	大数据技术与应用服务	D轮	¥ 4亿	光大证券



PART TWO

“Big Data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...” – Dan Ariely

1 市场尚未饱和 竞争不激烈

尚未出现垄断性行业巨头

除了数据堂在新三板上市之外，新创大数据企业，没有一家在美股、港股或深交所上市。

现有大数据企业扎堆北京

上海、广东和浙江以及二线城市潜力巨大。

大数据在生活/商业渗透性较弱

目前的大数据企业多为纯技术和数据资源平台，细分领域应用层面大有可为。

没有直接的变现模式

与网络游戏/电子商务不同，大数据没有最直接的变现模式，值得探索。



2

人才紧缺



中国

据中国商业联合会数据分析专业委员会统计，未来中国基础性数据分析人才缺口将达到**1400万**，而在BAT企业招聘的职位里，60%以上都在招大数据人才。2015年-2016年是大数据人才最为匮乏的两年。

Gartner预测，到2016年，25%的全球大型企业将部署大数据分析系统；到2015年，全球大数据人才需求将达到440万人；调查结果表明，全球64%的企业已经开始向大数据项目注资，或者打算在2015年6月之前将计划付诸实践。

全球

3 数据量增长超快

1

49亿

咨询公司 Gartner 预测今明两年互联物品的涨幅将达30%。分析师在指出，截止到2015年，全球物联网设备数量将从38亿飙升至49亿

2

250亿

分析师预测，传感器的普及将大大加速智能设备的开发、生产进度。到2020年，将会有大约250亿部智能装置出现在全球市场。

3

84%

在 2014 年年底，国内网络上集中存储的数据已经达到 1ZB，2到 2020 年时，当年的新增数据量将会达到 15.45ZB，整体的网络上数据存储量将会达到 39ZB，未来 6 年的年复合增长率达到了 84%。

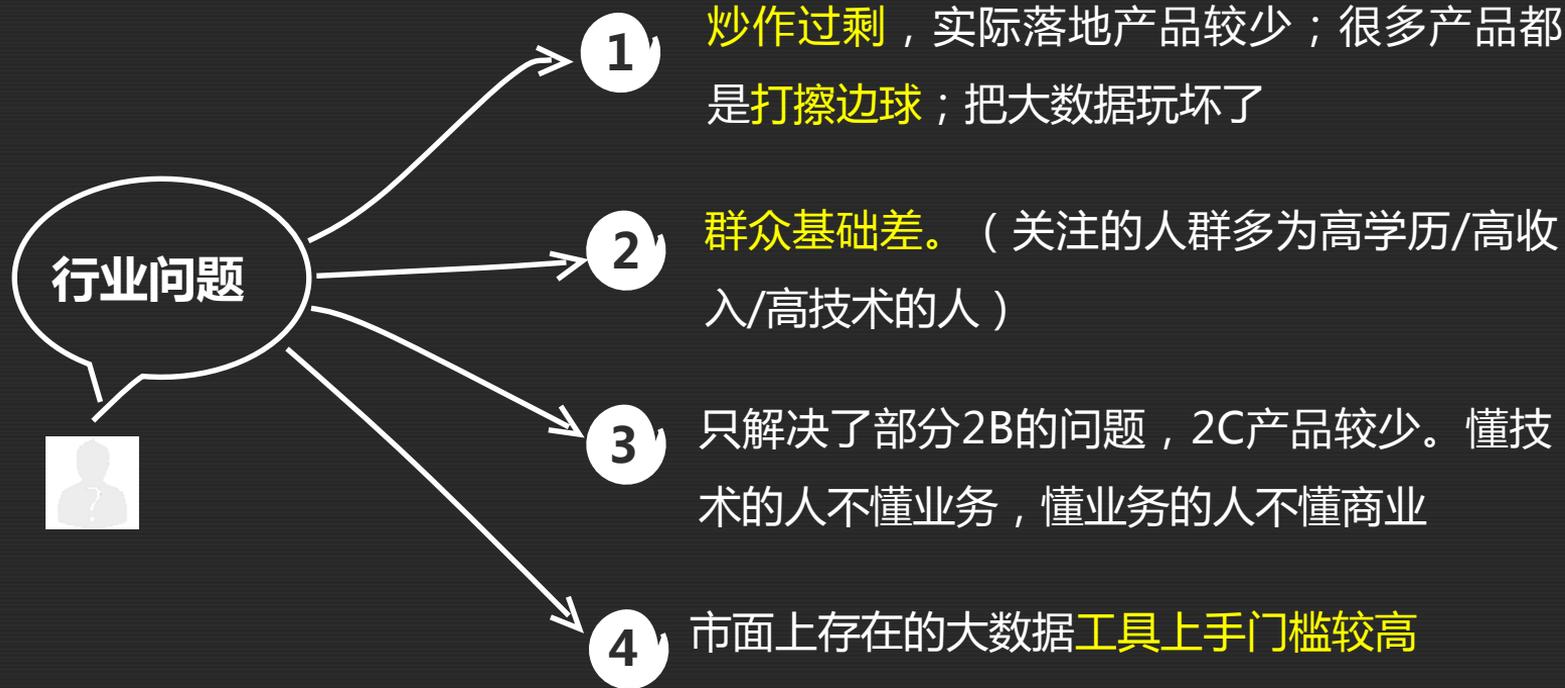
预测到 2020 年，平均每个中国人每年产生的数据约为 4.1GB。

图 3：中国产生的数字内容的数据存储累计量预测（ZB）



资料来源：IBM 商业研究院，中信证券研究部预测

5 行业目前存在问题



3

PART THREE

1

成本较高

01.人才成本较高

在美国，在R、NoSQL和MapReduce方面需求的专业人才薪水达到了每年约11万5千美元，在中国，大数据人才一将难求，创业公司不容易招大数据技术人才，即使招到，人才方面支出也较高。包括高薪、期权和股票等等。

03.项目启动资金高

不是30-50十万就可以玩起来的，比移动互联网APP创业项目启动资金要求高。

大数据创业门槛

02.存储硬件成本高

自建机房比云存储成本高很多倍，但是考虑到数据归属和安全性，大数据公司一般不会数据存在云上。

04.用户少/获取成本高

大数据目前的产品多为2B，但很多传统企业不懂大数据

没有明确的商业变现模式，也是目前大数据创业的最大门槛。
拼数据，你拼不过阿里百度腾讯，拼钱，这还是别拼了.....

Verizon发布的《2015年数据泄露调查报告》显示，79790个安全事件中已有2122个确认的数据泄露。值得关注的是在2015年的报告中新增了一个统计模型，用以帮助企业评估到底每笔数据泄露，要损失多少钱。如果泄露1000条记录时，有95%的可能会损失5.2万-8.7万。泄露1千万数据记录的花费介于210万到520万之间，但最多可能到7390万。

- ✓ 2014年5月，800万小米用户数据或被泄露
- ✓ 2014年12月，12306大量用户信息遭泄露
- ✓ 2015年4月，超30省市曝管理漏洞数千万社保用户信息或泄露
- ✓ 2015年5月，携程网全面瘫痪疑似数据库物理删除
- ✓ 2015年6月，美国人事管理局（OPM）被指出大量工作人员信息泄露
- ✓ 2015年8月，婚外情网站Ashley Madison数据遭泄露



大数据的安全性，是部署大数据架构和大数据创业最大的挑战之一！

3、大数据隐私

美国

美国的《隐私法案》
美国与欧盟之间的安全港
隐私权的社会保护-隐私声明

中国

规定“公民的个人数据不得非法搜集、传输、处理和利用
《民法通则》并未将隐私权作为一项独立的人格权加以保护
立法模糊

未来

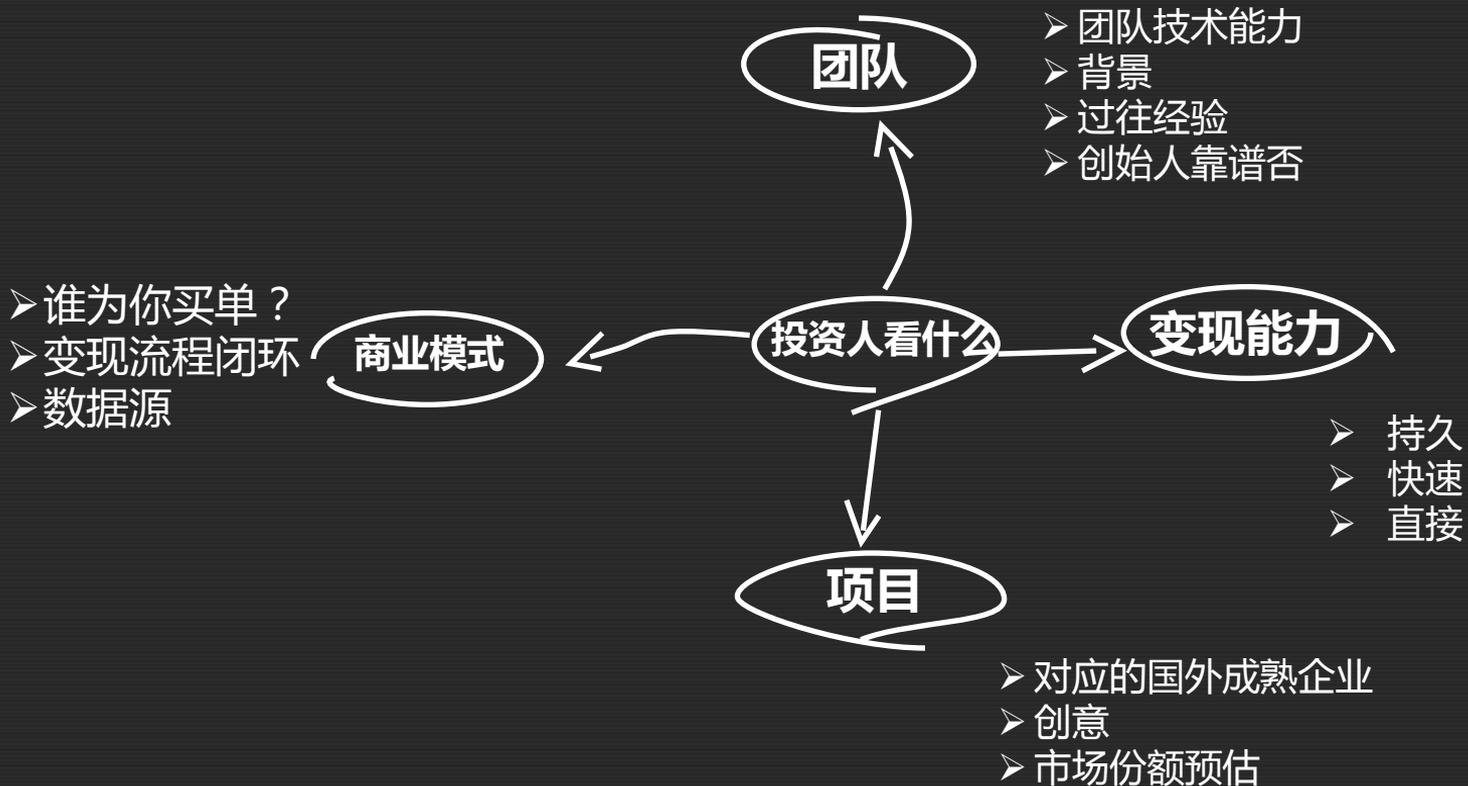
国家或将出台新的关于个人数据隐私保护的的政策法规
具有不确定因素。〈参考新的广告法〉

PART FOUR

我觉得真的是不缺钱，想法也满天都是。中国缺的是有一个想法，并且能够持之以恒把这个想法不断坚持做下去的人。
—— 马云

1

资本层面



2

现有商业模式

01

广告/营销

03

做工具

05

跨界和融合

直接卖数据

卖报告和解决方案

02

04

3

垂直热门投资领域

Cloudera、
Hortonworks和MapR
星环科技

Hadoop 商业化

Datastax
MongoDB

NoSQL 数据库

DataSift、RelateIQ、
RocketFuel

行业大数据应用



基于应用场景下的数据框架

SQL on Hadoop

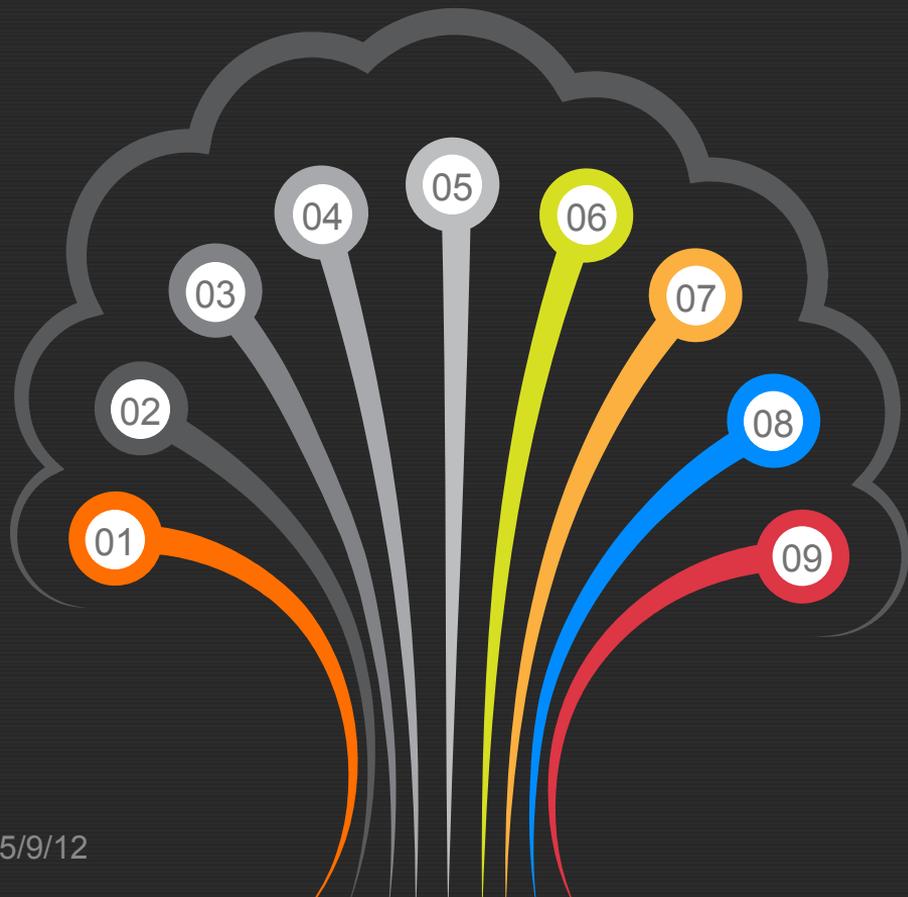
查询引擎、存储引擎、计算模型

分析和可视化

Tableau、Datameer

行业	应用价值	当前应用状况	未来应用场景
医疗	各地医院统一协调	药品研发	智能医疗平台化
	保健咨询服务系统	疾病治疗	电子健康服务生态系统
	新药研发	公共卫生管理	个人健康管理及预警
	便于医患沟通，提高诊疗效率	居民健康管理	个性化医疗保险
	疾病基因分析	健康危险因素分析	
金融	客户行为分析	银行大数据	金融产品垂直搜索和销售平台
	金融风险分析	保险大数据	全民征信
	运营分析	证券大数据	个人理财及投资电子化平台
电子商务	市场营销	广告追踪和优化	数据化运营
	个性化导购	产品分析	行业应用垂直整合
		用户行为分析	数据资产化
		推荐系统	商品来源可追溯及智能比价
零售	市场营销	数据营销	消费者全过程数据描述
	商品管理	精准管理	产业链营销重构
	提升运营效率	建立新市场增长点	着眼于线下数据的采集
	供应链环节		线上线下数据融合和系统对接
	新商业模式		基于消费者喜好定制化产品
通信	提升网络服务质量/提升书卷气	支持营销运营管理应用	社交网络分析
	精准洞察客户需求	客户体验管理	客户体验分析
	升级商业信息化解决方案	促进智能管道运营应用	客户价值分析
政府公共服务	加强统筹规划	智慧交通	政府的智能化
	提高智慧城市感知水平	智慧安防	数据驱动政府公共服务
	提高经济社会智慧化水平		促进政府的数据治理（反恐/环保/防灾等）

大数据应用的现状和机会



- 01 想清楚谁为你买单（找用户）
- 02 痛点是什么？（找需求）
- 03 稳定/独特的数据源（找数据）
- 04 靠谱的人做靠谱的事（找人才）
- 05 考虑2C的产品方向
- 06 忘记科技行业过往经验
- 07 将大数据产品化（小而美）
- 08 试错和迭代
- 09 深耕一个领域

6

谁为你买单

B2C+B2B 企业用户/个人用户

阿里巴巴-百度-腾讯
政府公开数据

或给政府用 (如: Palantir)

B

B2C, 个人用户能够用到的大数据产品较少

如: 大姨妈、百度指数

A

B2B, 现有的大数据公司/产品用户多为2B

如: 星图数据, Hortonworks, Cloudera, 星环科技、Talkingdata

C

更需要傻瓜式，如同Excel那样的简单工具.....

苹果没有教大家用iPhone.....

交付的工具应该是有漂亮UI，功能按键等完整产品，而不是一堆代码



B2B更多的是做工具和服务（数据可视化/商务智能/CRM等）

我的钱花哪去了，哪些可以省下来

个人理财

该不该跳槽？现在薪水到底合适不合适？

职业发展/自我量化

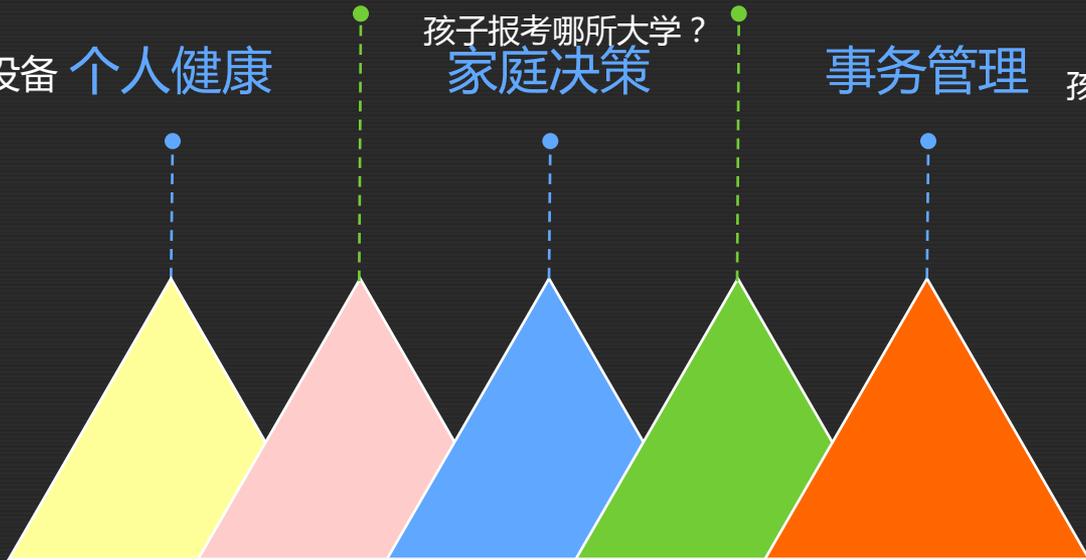
连接智能/可穿戴设备 个人健康

孩子报考哪所大学？

家庭决策

事务管理

孩子家庭作业、考试分数

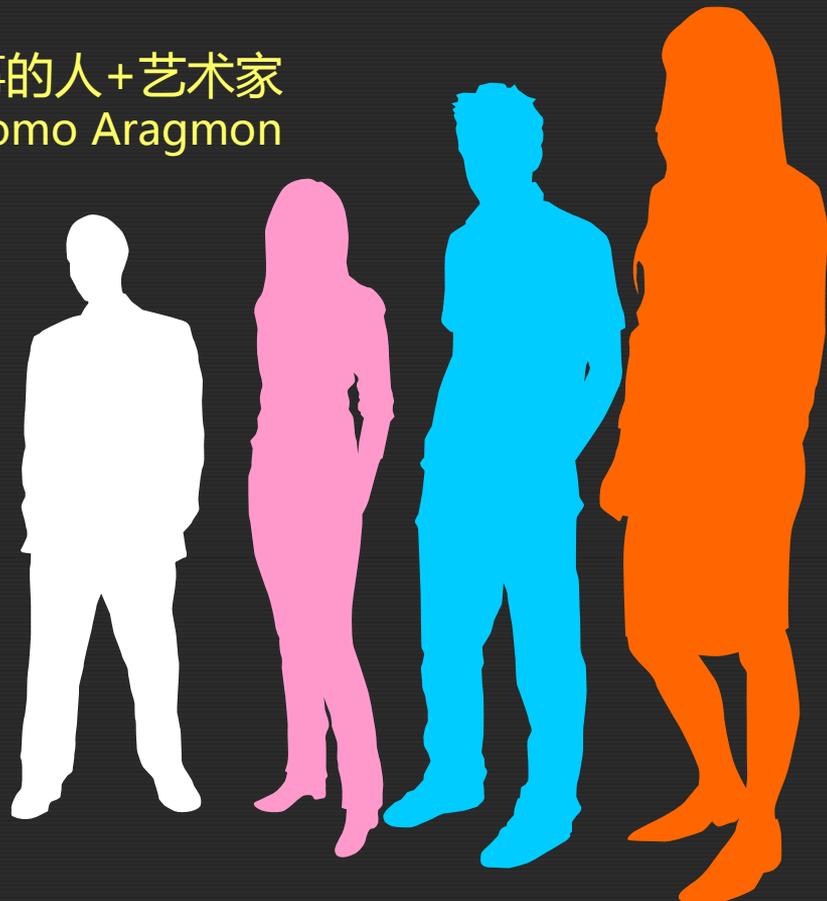


B2C更多做得是产品应用

数据科学家是一个团队，不是一个人。

数据科学家 = 统计学家 + 程序员 + 讲故事的人 + 艺术家
– Shlomo Aragon

Spark 数据建模
python 机器学习
统计学
概率学 算法 D3 神经网络
Java C+ NoSQL
Hadoop R SAS
MapReduce



大数据与小数据

01

大数据的结果是小数据

02

客户/用户不会在乎数据大小，只在乎有没有用

03

大数据的落地是小数据

04

数据本无大小，但运用数据的立场却分大小

当你在寻思如何抉择大数据，还是小数据时，先搁置这事儿。思量一下，你的公司是否擅长利用数据创造价值，如果你的公司还没有达到这个境界，那先把这事解决了再说。

谢谢

End.