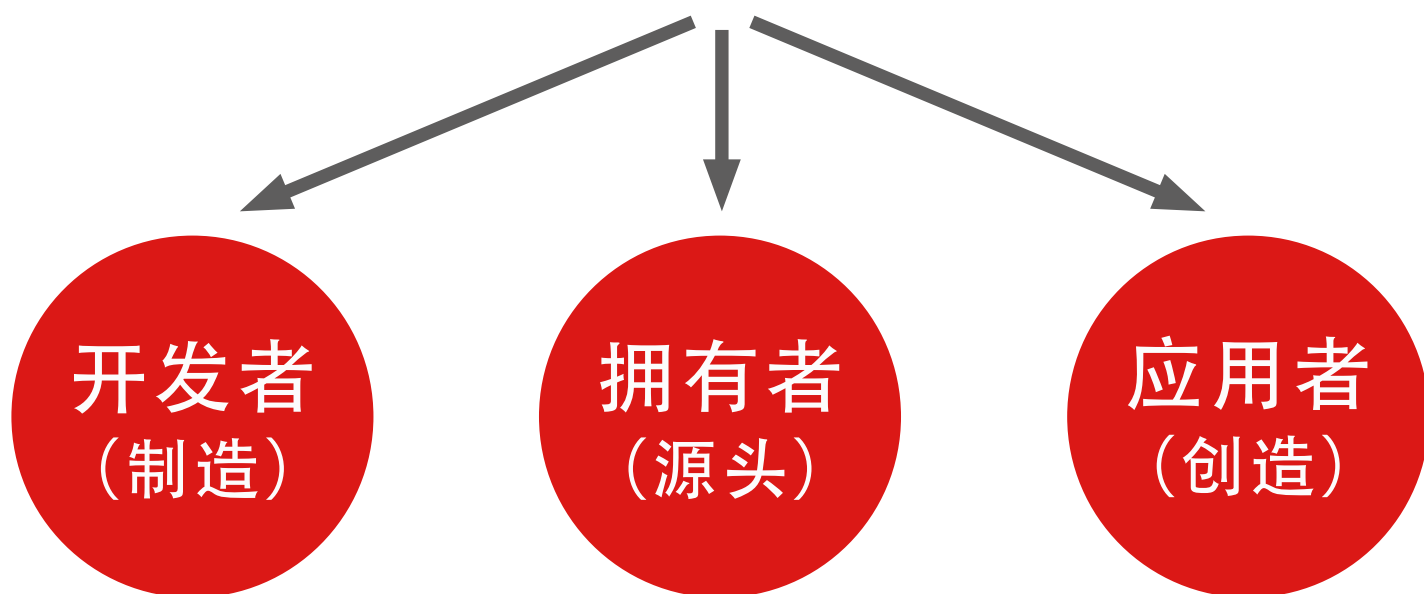


上医强生

大数据与电商整理

大数据中企业的角色说明

大数据角色



大数据应用的例子：

- 1, 佳能相机 RAW 照片格式，存储 1100 万光束，无需专业对焦知识，可拍摄后电脑上再调焦距。大数据改变了牌照模式。
- 2, 日本汽车防盗系统，根据大量车主的屁股与座椅的接触点，生成一套屁股接触点防盗系统，如果是小偷坐上去，车自动熄火。大数据改变了识别模式。
- 3, Google 输入法，通过网络上众多使用者输入错误后自己重新调整输入的行为，开发了自动提示和修正输入的功能，后来演变成智能输入提示预测功能。大数据改变了预测模式。
- 4, 美国个人消费信用评估系统 FICO，根据用户在线下和网络上的多种多样的信息，进行数据化处理后迭代分析，预测用户的信用评估。大数据改变了信用评估的方式。
- 5, Amazon 通过大数据预测用户可能感性兴趣并购买的图书和其他产品。
- 6, Netflix 通过大数据预测用户可能感兴趣的电影，代表产品《纸牌屋》。

开发者 (制造)

职能与特性：

- 1, 存储：Cloud Server
- 2, 集群：Hadoop
- 3, 数据库：NoSQL
- 4, BI：商业智能分析
- 5, APP：采集
- 6, 智能设备：采集

代表：

- 1, Microsoft Azure：提供各种云服务和云存储，还有，MS Sql Server 2014 大数据存储数据库
- 2, Google BigQuery：针对PB级数据的极高效的处理引擎。
- 3, Intel Hadoop：Apache基金会的支持者，Hadoop商业化的推动者。
- 4, DataMeer：为Hadoop定制的大数据分析解决方案提供商，独有的可视化的视图系统，大大简化了商业分析和应用。

拥有者 (源头)

职能与特性：

- 1, 样本 = 全数据，不在基于样本抽样，而是全数据化存储。
- 2, 混杂数据，非结构化数据，非精确数据。
- 3, 相关性数据，非因果关系数据，非逻辑性数据。
- 4, 元数据，关于信息本身的数据。
- 5, 历史数据，长期保存，挖掘潜在价值。
- 6, 废弃数据，表面无用的数据，隐藏价值巨大。

代表：

- 1, WalMart：海量的产品与消费者购物数据等。
- 2, Google：海量搜索关键词数据、海量地图数据等。
- 3, 淘宝 /Amazon：海量商品，价格，评论，位置和用户数据等。
- 4, Jawbone：用于大量用户睡眠数据。
- 5, Nike：通过手机 Nike+Running App 采集了大量用户运动、心率、呼吸、气压、地理位置、天气状况等数据。

应用者 (创造)

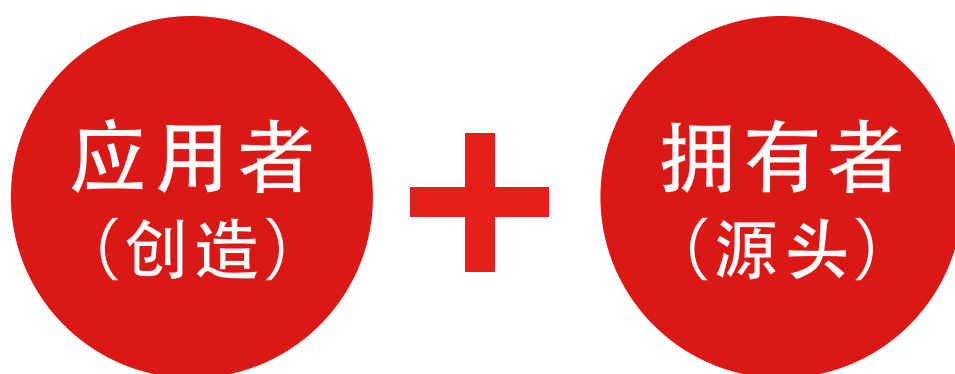
职能与特性：

- 1, 历史数据挖掘
- 2, 非结构化数据重组
- 3, 数据废弃再利用
- 4, 数据相关性发现
- 5, 个性化定制化数据推荐
- 6, 基于非精确大数据的精确预测
- 7, 开发数据授权应用
- 8, 数据价值评估

代表：

- 1, 淘宝指数：通过众多淘宝商品和用户数据，挖掘出特定地区的特殊群体对特殊产品的兴趣。处理速度非常快。
- 2, 淘宝数据魔方：通过淘宝海量数据，帮助卖家进行数据化运营，包括产品定位，行业定位，品牌定位，促销推荐等。
- 3, 百度指数：以百度海量网民行为数据为基础，分析市场动向，洞察网民兴趣点，需求点，协作舆情动向监控。
- 4, 惠惠购物助手：通过抽取和分析各大电商网站的产品、价格、评论等各种数据，进行比价和推荐的购物助手。
- 5, 腾讯路宝（刚上市）：通过车配件的物联网 + 互联网模式，根据路段、路况、车辆状况、驾驶员状况等众多信息，进行驾驶辅助。
- 6, 中国移动：通过全国大量手机信号反馈，微调基站信号性能。

上医强生大数据具体说明



角色定位：大数据拥有者 + 大数据应用者，而不是大数据开发者，因为随着技术的进步，大数据开发技术成本会越来越低，外包的可能性增加。



应用者 (创造)

作为大数据应用者：提高深度、精确度的同时，也要增加广度。采集的数据越精确预测结果越精确，采集的数据越广阔，非结构化数据越多样，预测结果也会越精确。（相关性而不是因果性）

- 1，历史数据：用户以前的就诊信息，就诊周期，诊后行为等对比分析，可以医院展开合作。
- 2，数据重组：天气，性别，年龄，地理位置，饮食状况，病史病例等于脉诊数据等进行重组，分析相关性。
- 3，数据废气：收集用户脉诊后的行为数据，是否遵循推荐的食材，推荐的养生方式，是否又住院，是否改变饮食习惯，是否坚持等等，而不是舍弃这些数据。
- 4，数据预测：配合大数据的相关性结果，针对用户进行个性化推荐，并收集用户随后行为。
- 5，数据授权：与中医药、研究院、政府部门等展开合作，共同挖掘数据价值，改善国民健康状况。



拥有者 (源头)

作为大数据拥有者：增加数据采集的多样性。

云数据采集：

- 1, App：高互动性，采集用户行为。
- 2, 智能设备：脉诊仪，生命环等设备智能采集。
- 3, GPS 定位：采集地理位置，周边信息，交通信息，工作位置等。
- 4, 填表：用户针对问题填表。
- 5, 多方合作：与中医院，养生馆，餐馆餐厅等合作，共同采集数据。
- 6, 数据抓取：网络信息、社交信息、健康问题数据收集。
- 7, 元数据：采集表面上没有明显因果关系的相关性数据。

合作需要的大数据技术：

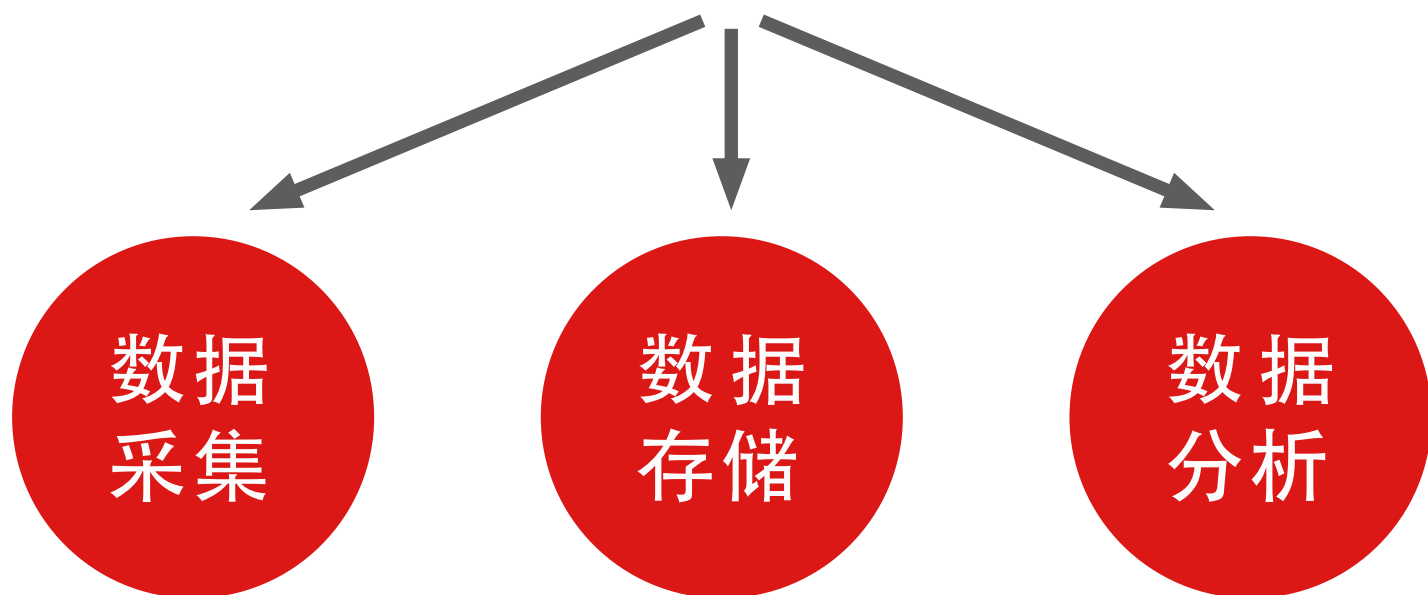
Hadoop, Azure, NoSQL, Redis, TTServer, Linux, PHP, Java, SLB, OSS



元数据说明：全方位的混杂性数据，非结构化数据，非精确数据

- 1, 与脉诊相关：寸关尺，浮中沉，温度，湿度等。
- 2, 与用户相关：性别，年龄，职位，病史，家庭，性格，收入，饮食习惯，习气（抽烟饮酒）等。
- 3, 与环境相关：天气，气候，气压，地点，省份，人口，交通，平均收入等。
- 4, 与行为相关：脉诊频度，周期，间隔，反馈，脉诊后行为，推荐后行为等。
- 5, 与医疗相关：看病周期，中医西医，就诊环境，病发周期，症状，遗传。

上医强生大数据执行 参考要点



数据采集

要点

- 1, 采集大量混杂性, 非结构性数据。
- 2, 增加元数据的采集。
- 3, 不能只收集样本, 而要全息数据, 尽可能多的采集。
- 4, 更新频繁, 吞吐率高, 必须保证高性能。
- 5, 混杂数据多, 数据格式多样, 变更快, 必须保证灵活适应。

数据存储

要点

- 1, 云端存储, 私有云用于内部集群, 公有云用于外部集群, 也可以开放接口。
- 2, 数据更新频繁, 要求分布式集群系统, 保证并发, 高效, 可靠。(Hadoop, HDFS, HBase)
- 3, 混杂数据多, 非机构化数据多, 采用 NoSQL。(MongoDB, Redis, TTServer)
- 4, DataCenter 保证周期性备份, 数据不丢失。

数据分析

要点

- 1, 表格化, 原始数据变成人可理解和查看的表格数据。
- 2, 视图化, 多样丰富地显示数据的相关性。
- 3, 关系化, 相关性数据分析与对比, 生成变量函数。
- 4, 自助化, BI 应用人员可以根据自定义查询条件查看分析视图。

上医强生大数据 案例模拟



脉诊仪
数据采集



脉诊仪

性别

血压

脉搏

体温

地点

时间

饮食

气压

心情

年龄

特性：精度低，数据混杂，非结构化

举例：少年，中年，老年，广州，浙江，尺脉高，尺脉低，吃素，吃肉，饮酒少量，大量饮酒，心情愉快，悲伤，愤怒



数据化



非结构化数据
的元数据
提取

元数据因素
量迭代

表格化

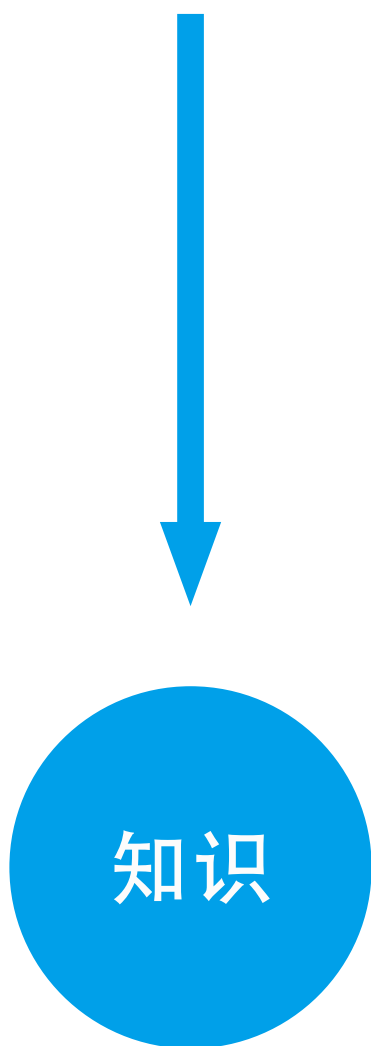
特性：提取并迭代

举例：年龄 (40-50)，尺脉数据 (5.5)，饮酒频度 (2 两 / 天)，心情等级 (2) 地点：山东省淄博市



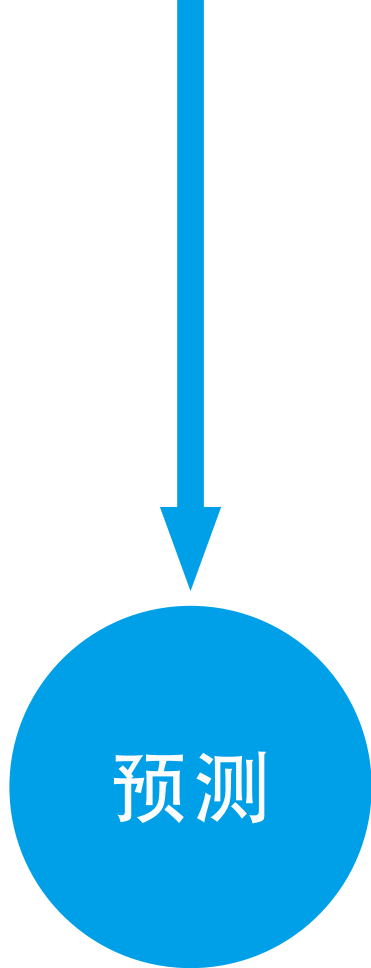
特性：不查找因果，查找数据相关性。

举例：40-50 岁男性，常饮酒，心情悲伤的人，常出现尺脉低沉无力。



特性：根据相关性形成特殊知识。

举例：40-50 岁男性，常饮酒，心情悲伤的人，常出现尺脉低沉无力，推荐吃薏仁，可让尺脉增强。



特性：大数据的目的

举例：40-50 岁男性，常饮酒，心情悲伤的人，常出现尺脉低沉无力，推荐吃薏仁，可让尺脉增强，增强概率高达 98%，方案采用率高达 99%。

大数据的成败 究竟何在？



资金？ 技术？ 数据？ 应用？ 可视化？
设备？ 存储？ 私有？ 决策？ 智能化？



对数据（信息）本质的理解
理解的越深走得越远

Google

VS

Microsoft

Google 的 拼 写 纠 正：

根据人们网络输入关键词的自我修正和对搜索结果的点击来动态调整方案。成本低，升级易，维护简单，与时俱进。

Microsoft 的 拼 写 纠正：

根据大英词典来修正用户的输入。成本高，版税高，升级难，维护复杂，耗时耗力。

上医强生大数据 合作方案

微软：Microsoft Azure Hadoop + MS SQL Server 2014，适宜商用，价格高，有售后，运行稳定，客服响应慢。

英特尔：Apache Hadoop。今年会推出中国市场。

用友：UAP + BI。没有成型的商用系统，工程师团队可定制开发，费用很高。

谷歌 Big Query：30 天免费试用，速度快，使用简单，对中国大陆不友好，网络传输不便。

星环科技：Intel Hadoop 负责人离职后创建的公司，技术能力强，定制开发能力一般，需要我方人员配合完成工作。

汉得科技：擅长 ERP,CRM 系统，提供大型系统定制服务，费用高。

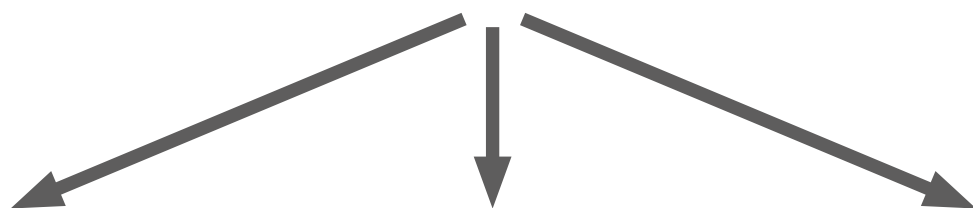
随身科技：汉得子公司，擅长移动数据化采集，需要进一步沟通。

QlikView：可视化 BI 分析产品，老外的产品，大陆服务差。

DataMeer：硅谷的可视化 BI 分析产品，功能强大，老外的产品，大陆没有服务。

清华大学研究生院：张硕，主攻大数据方向，可以考虑合作开发。

电商运营平台 分析



独立电
商平台
运营

第三方
渠道运

线下渠
道运营



独立电商平台运营

仓储：

货物统一仓储，统一出库，直销模式，可以自建厂或者托管，样式灵活，易于统一管理和监控。

订单：

可统一管理，第三方订单（淘宝，天猫等）可统一汇总进来，可实现大区订单跟踪与分析。

社区：

自主维护，可以针对用户提供定制化、个性化服务，可以针对用户的问题和商品进行互动，可统一管理粉丝群体。

控价：

可自主把控，易于统一价格，不易乱价，也可灵活进行促销和限时等各类活动。

优势：

模式灵活，互动性强，统一管理，粉丝粘性强，营销活动灵活，易于推广品牌。

劣势：

前期开发投入高，流量小（相比于天猫、京东等第三方电商平台），评论少，不容易形成商品信任，维护成本高，设备成本高（需要自建云服务器），技术门槛高，人员技能要求高。

举例：

小米：独立电商 + 社区 + 第三方电商平台直营模式

小米有大量的粉丝通过小米社区服务；小米商城官网独立运营经常举办各类促销活动和限购限时等预约活动；小米商品都是统一仓储，统一出货，官网订单和第三方电商平台产生的订单都统一汇总，统一管理，价格严格把控；京东、当当、亚马逊等非直营大电商平台，小米普遍抬高价格，只有自家最低价，但是自家商品基本都需要预约，饥饿营销的方式很好地实现了粉丝从大电商平台到小米官网的导流。小米即很好地利用了独立电商的灵活优势和粉丝粘性，又通过社区服务了大量粉丝形成了品牌效应和口碑传播，还能利用第三方电商平台的大量流量进行官网导流，大量用户评论加强了产品信任度，此方式是目前国内互联网电商的最佳方式，但是投入成本不低。

第三方
电商平台运营

直营

仓储：

货物统一仓储，统一出库，与独立电商相同，易于统一管理和监控。

订单：

依赖于第三方电商平台的订单管理机制，可打通独立电商平台的订单管理后台，实现第三方订单（淘宝，天猫等）的统一管理、统一汇总，可实现大区订单跟踪与分析。

社区：

依赖于第三方电商平台的活动规则来做活动，不足够灵活，粉丝无法聚拢，用户粘性低，客户服务需要依赖第三方工具，效果差。

控价：

可自主把控，易于统一价格，不易乱价，也可进行促销和限时等各类活动。

优势：

第三方电商平台流量大，短期增长明显，技术门槛低，设备成本低，易于进入，维护成本低，人员技能水平要求低，大量用户评论易于产生商品信任，增强品牌。

劣势：

模式不灵活，第三方电商平台产品数量众多，营销成本非常高，中后期营销需要购买大量的商家服务，第三方电商平台的手续费、支付平台费和各种杂费，销量越多，费用越高。

举例：

苹果公司 AppleStore：独立电商 + 社区 + 第三方电商平台直营模式 + 线下直营 + 线下授权店

AppleStore 官网与第三方直营能够做到价格完全统一，不轻易降价，活动也做到完全统一，出货、订单、人员培训、售后、维修、服务等都是公司总部统一管理和执行，在流程和授权监管上非常到位，授权商胆敢乱价马上终身取消代理资格，AppleStore 针对电商管理负责等因素，完全采用直营模式，天猫店基本上与苹果官网如出一辙，运营和服务人员也是苹果公司员工，只是支付渠道不同，流量不同，用户群体不同。针对渠道众多，用户难以跟踪等问题，苹果通过独有的 iTunes 系列软件把用户统一到一起，无论从哪里购买，苹果都能完全掌握用户信息和行为。线下苹果官方直营店更是集购买，服务，培训，售后，活动等众多功能于一身，打造成苹果品牌体验一体化。此模式需要大量的监管，执行必须非常到位，不适合早期采用。（很多国内大型公司也无法做到苹果这样，比如海尔，美的）

第三方
电商平台运营

授权

仓储：

多方仓储货物，非统一出库，物流和配送都不统一，很难统一管理和监控。

订单：

依赖于第三方电商平台的订单管理机制，各家自我管理订单，很难统一管理和汇总，很难实现订单跟踪与分析。

社区：

依赖于第三方电商平台的活动规则来做活动，不足够灵活，多家授权商各自招揽自己的用户，粉丝无法聚拢，用户粘性非常低，客户服务效果完全依赖于授权商本身，无法统一管理品牌。

控价：

难以把控，难以统一价格，乱价现象严重，各家独自进行促销和限时等各类活动，各家直接互相竞争。各家授权商也经常以低于官网的价格销售商品。

优势：

外包授权商运营成本低，容易大面积铺开。

劣势：

乱价非常严重，严重影响品牌，仓储和订单管理比较混乱。此模式属于传统商家进入电商领域的过度模式，完全沿用线下的模式，只是进行互联网运营而已，线下的所有劣势在互联网上都会被放大。

举例：

海尔、美的、TCL：独立电商 + 社区 + 第三方电商平台授权模式 + 线下渠道运营

海尔、美的、TCL 的最严重问题就是乱价极为严重，代理商、分销商、经销商、直营店等等给予的折扣政策都不同，造成官网的价格为最高，用户很少去官网买，需要大量的搜索和调研才敢购买，用户分散，电器类没有软件可以统一用户，很难聚集粉丝。此模式对品牌的负面影响比较严重，而且需要极大力度的监管和执行，成本很高。

线下渠道运营

仓储：

分销商，经销商，代理商，直营店，精品店等模式不同仓储形式也可能不同，难以统一管理。

订单：

订单种类多样，难以管理和跟踪。

社区：

周边环境依赖性强，促销活动非常不统一，各自为营，会员卡机制也难奏效。

控价：

定价不统一，各大经销商、代理商等等折扣不统一，政策多样，极易乱价。

优势：

地域优势，品牌店门店优势，房地产增值等。

劣势：

引流艰难，拓展难度大，地价高昂，人员雇佣成本高。

举例：

中关村电脑城：大部分门店都已经关闭，从中关村走出来的刘强东，通过电商平台京东已经走向了纳斯达克。

上医强生电商 合作方

选择依据：

- 1, 开源 or 商用
- 2, 授权 or 版权
- 3, 有无二次开发服务
- 4, 纯电商 or 团购 or 博客 + 电商 or 论坛 + 电商
- 5, 托管服务 or 自建私有云
- 6, 价格

可选择合作方：

1, ECShop：国内开源电商平台，使用需授权，可提供二次开发服务，

2, osCommerce：国际开源电商平台，需要自己二次开发，无需授权

3, WordPress + osCommercePlugin：博客 + 电商平台，开源，免费，可扩展性强，需要自己二次开发，无需授权

4, Shopex：国内商用电商平台，提供定制二次开发，授权使用。

5, zhuyun：国内商业电商平台，提供定制二次开发，授权使用。

6, HiShop：国内商业电商平台，提供定制二次开发，授权使用。

7, V5Shop：国内商业电商平台，提供定制二次开发，授权使用。

8, ShopNC：国内商业电商平台，提供定制二次开发，授权使用。

9, xpshop：国内商业电商平台，提供定制二次开发，授权使用。

10, Discuz：论坛 + 电商平台，开源，免费，需要自己二次开发，无需授权

11, PhpWind：论坛 + 电商平台，开源，免费，需要自己二次开发，无需授权

注：

其中 ECShop, Shopex, zhuyun, HiShop, V5Shop, ShopNC, xpshop 都属于国内电商平台，类型比较相同，不同点在于价格, 服务范围, 技术优劣等, 需要再详细调查。