

大数据行业发展和趋势

大数据应用实践和案例



大数据时代

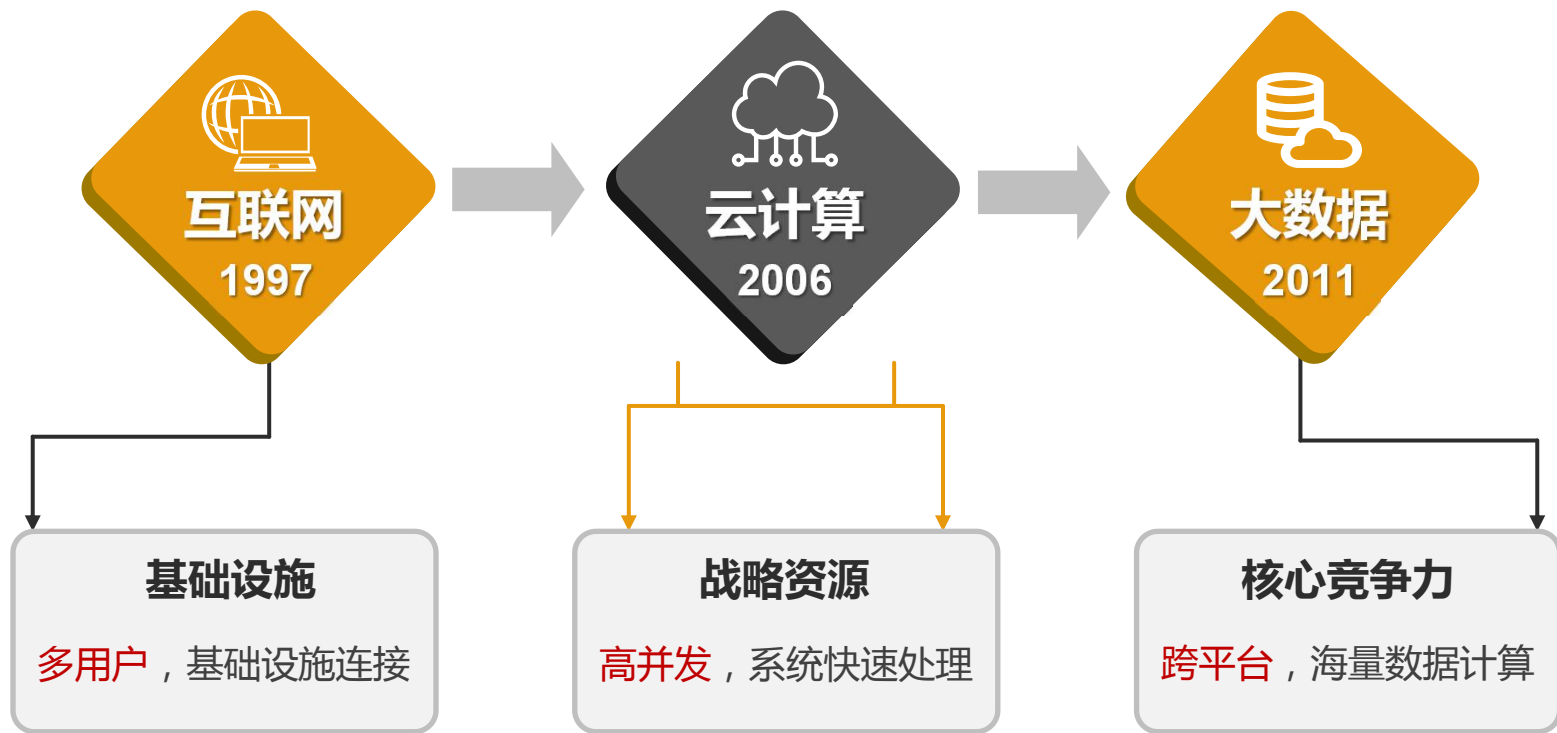
“

Part 1

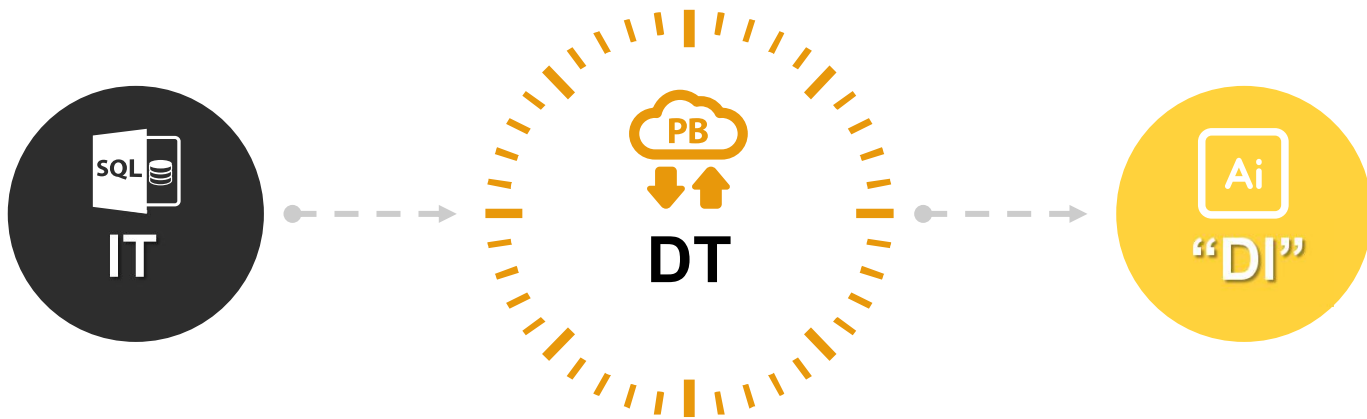
大数据行业发展和趋势

”

互联网、云计算、大数据发展路径



大数据产业发展的演变



信息驱动

- IT时代以自我控制、自我管理为主。
- 信息技术驱动生产，信息流为核心。
- 数据作为附加产物，碎片化严重。

数据驱动

- DT时代以服务、激发生产力为主。
- 数据驱动业务，以数据流为中心。
- 数据整合打通链路，加快产业升级。

智能驱动

- 智能技术应用产业、解放劳动力。
- 智能驱动决策，机器学习技术
- 赋能产业改变人与场景交互方式。

“大场景” 孵化 “大数据”



6年“数据处理能力”飙升

10万倍

2014年“双十一”：

总数据处理量：100PB

Oracle RAC
业务原始数据

GP
单BU数据

Hadoop
全集团数据

阿里云ODPS
生态圈数据

什么是大数据？



举个栗子

会员卡号是：342623
我想要一个海鲜披萨。

你怎么知道我会喜
欢吃这种的？

你们直接把比萨送我家吧，
你们多久会送到？

为什么？

那你们有什么推荐？

好。那要一个家庭特大号，多少钱？

**披萨店
为您服务。

海鲜披萨不
适合您。

您可以试试我们
的低脂健康比萨。

99元，它足够您一家六口吃。
但您母亲应该少吃。

您的会员
卡号是？

根据您的医疗记录，你的
血压和胆固醇都偏高。

您上星期一在国家图书馆
借了一本《低脂健康食谱》。

她上月刚做了心脏
手术，还在恢复期。

大约30分钟。如果您不想等，可以自己骑车来。根据我们CRM全球定位系统的车辆行驶自动跟踪系统记录，您登记有一辆车号为SV-568的摩托车，而您目前正在铁山路右侧骑着这辆摩托车.....



为什么会有大数据 - 大数据来源

机器产生的数据

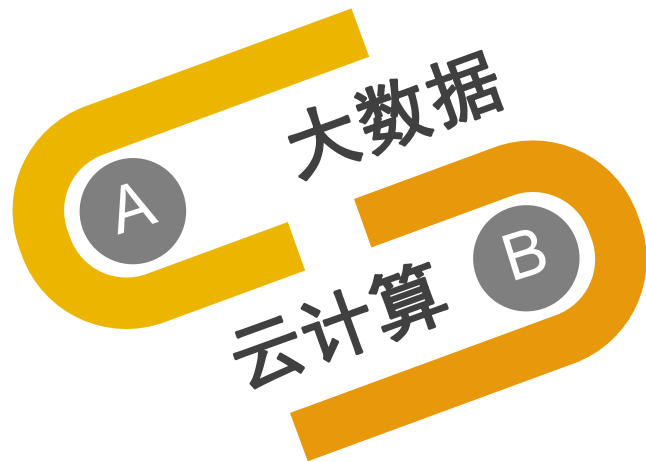
传感器感知和运输不断生产的数据

移动设备收集了大量的用户数据

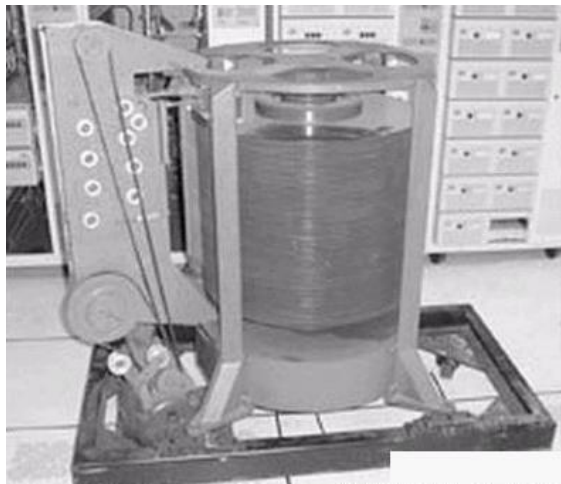


大数据的计算能力

- 云计算一般由**数量惊人的计算机群**构成，谷歌云计算拥有的服务器超过**100万台**；
- 云计算系统是一个极其庞大的资源池子，用户可以随时、随地、按需灵活地购买，就像购买煤气和自来水一样便利；
- 云计算让普通人体验每秒**10万亿次**的运算能力。而如此强大的计算能力，可以模拟核爆炸、预测气候变化和 market 发展趋势。



大数据的存储能力



1955年

世界上第一款硬盘——IBM商用硬盘存储器
一兆字节的存储量需要**6000多美元**



2010年

一兆字节的存储量
仅需要**0.005美分**



2020年

1T硬盘 = 一杯咖啡的钱
= 一个图书馆的全部信息

归类数据类型、有效分析组合

海量 Volume

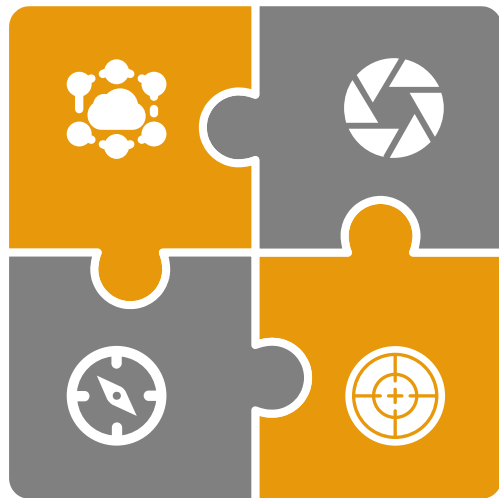
非结构化数据的超大规模和增长

- 占总数据量的80~90%
- 比结构化数据增长快10倍到50倍
- 是传统数据仓库的10倍到50倍

价值 Value

挖掘大量的**不相关信息**的价值

- 对未来趋势与模式的可预测分析
- 深度复杂分析（机器学习、人工智能、商务智能(咨询报告等)）



多样 Variety

大数据的**异构和多样性**

- 很多不同形式（文本、图像、视频、机器数据）
- 无模式或者模式不明显
- 不连贯的语法或句义

高速 Velocity

实时分析和离线分析

- 数据输入、处理与丢弃
- 互联网接入终端快速增长
- 快速计算、数据分析

Walmart
沃尔玛

1

商业

沃尔玛基于每月4500万网购数据，结合网上挖掘的对产品的大众评分，开发语义搜索引擎，方便浏览，在线购物者增加10—15%，**增加销售十多亿美元**。还通过对消费者购物行为分析，了解顾客购物习惯，优化商品陈列。

2

农业

硅谷Climate公司从美国政府获得30年的气候、60年的农作物收成、14TB的土壤数据，还收集250万个地点的气候数据，向农户提供**天气变化、作物、病虫害和灾害、肥料、收获、产量、市场价格等咨询和保险服务**，承诺每英亩的玉米利润增加100美元，如预测有误将及时赔付。

3

制造业

丰田利用数据分析在试制样车前**避免了80%的缺陷**；GE通过对2万台喷气引擎的数据分析，能够提前一个月预测其维护需求，预测准确率达到**70%**；百度将网民对汽车的各类搜索请求进行大数据挖掘，帮助一汽等车企深入了解消费者需求，设计新品及资源调配



4

金融

华尔街的银行应用大数据分析风险时间从3月缩短到3小时，价格计算从48小时到20分钟，行为分析从72小时到20分钟；交通银行太平洋信用卡中心采用大数据信贷审批决策引擎，审批效率提高30%；阿里公司基于大数据建设投融资平台。



5

交通

美国UPS每天平均运输163万件包裹，用传感器跟踪46000辆车，2011年少跑了8500万英里路程，节约了840万加仑燃料；中远物流公司有100多个配送中心，3000个网点，装GPS的上万辆车每月产生2亿条信息，减排10%。



6

医疗

苹果公司总裁乔布斯花费几十万美元获得自身全部DNA，希望通过这种方式战胜癌症虽然愿望没有实现，但也将他的生命延长了几年；人体对药品代谢方式的差异取决于个体特定的基因、酶和蛋白质组合，因此基因信息对选择最优治疗非常关键。

大数据的行业应用 – 商业案例

DB
GEEK

IT大咖说
知识分享平台



农夫山泉每天新增的数据是100G，每月为3TB。



过去

采用传统数据仓库，再通过数据报表展现，这个过程长达**24小时**

24

现在

采用实时大数据计算，同等数据量的计算速度从过去的24小时缩短到了**0.67秒**



农夫山泉以**30%-40%**的年增长率，快速超越原先的三甲：娃哈哈、乐百氏和可口可乐。



北美最大的付费订阅视频网站——Netflix

2012年，Netflix准备推出自制剧。不过在决定拍什么、怎么拍上，Netflix推出了自己的秘密武器——大数据。

用大数据拍自制剧

01

收集

收集该网站上用户每天产生的行为，如收藏、推荐、回放、暂停等，还包括用户的搜索请求等。

02

预测

分析出凯文·史派西、大卫·芬奇和“BBC出品”这三种元素结合在一起的电视剧产品将会大火。

03

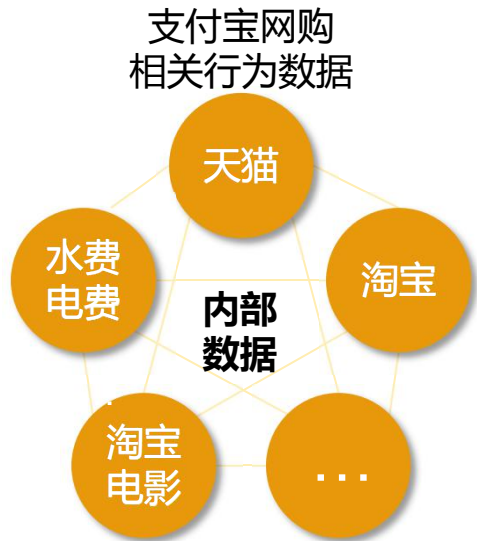
拍摄

融合三者拍摄了一部《纸牌屋》，结果大获成功，成为了2013年全球最火的美剧。

大数据行业应用 - 金融



支付宝“芝麻信用”——
“芝麻信用分”，授权开通后，每个支付宝用户都可以看到自己的芝麻信用分。分数越高代表信用程度越好，违约可能性越低。



芝麻信用分高

租车、住酒店无押金

签证无需财产证明

...

芝麻信用分低

找不到工作

租不到房子

贷不到款

...

大数据行业应用 – 医疗案例



医疗机构以利用Watson规范医疗费用

在医疗、医药行业可以帮助做复杂疾病的诊断和数据分析



利用认知计算技术缩短治疗方案制定周期

案例：Watson通过比对**2000万份**癌症研究论文，在**10分钟**得出了诊断结果：患者得了一种罕见白血病。

帮助分析疾病风险，保证理赔过程的合规性、合理性，防止滥用和欺诈



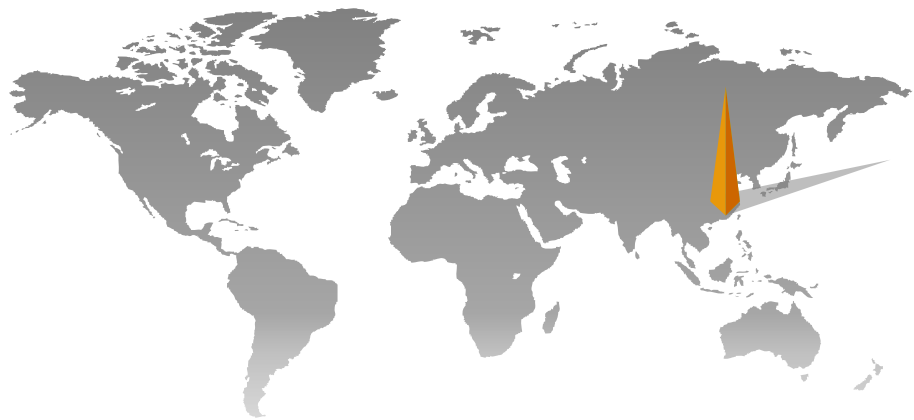
大数据时代

“

Part 2

大数据应用实践和案例

”



01

企业大数据应用

02

企业大数据实践

03

典型案例分享

大部分企业已意识到数据分析对企业发展的重要性

59.2%的受访企业已经成立了数据分析相关部门；同时，27.3%的企业正在计划成立数据部门。

01



02

大数据应用得到进一步普及

35.1%的受访企业已经应用了大数据；同时，22.9%的企业计划将在未来1年内应用大数据，发展前景较为乐观。

企业应用大数据实现智能决策、提升运营效率和改善风险管理

其中应用大数据实现智能决策的企业占比最高(55.8%)，其次为提升了运营效率(48.2%)以及更好的管理风险(25.7%)。

03



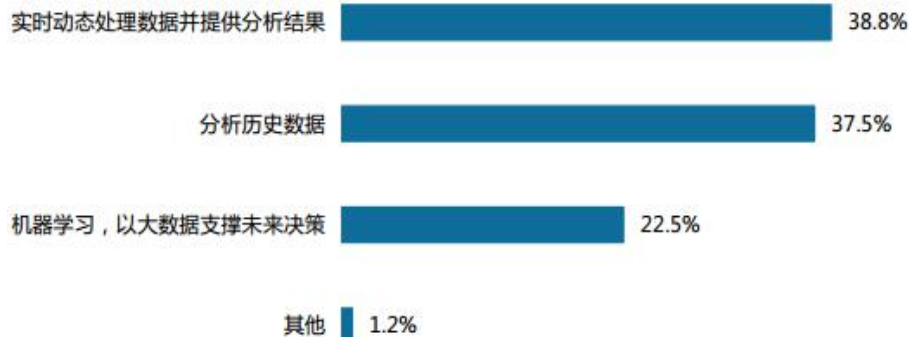
04

企业将进一步加大在大数据领域的投入

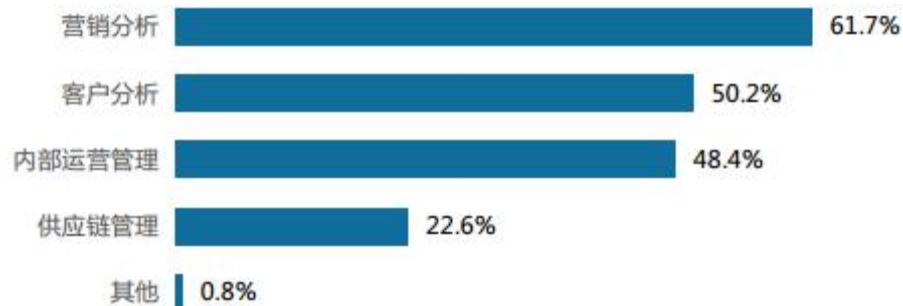
一半以上的受访企业未来计划加大对大数据的投入，其中近20%的企业预计投入增长在50%以上。

企业大数据应用场景和效果

企业数据分析方式



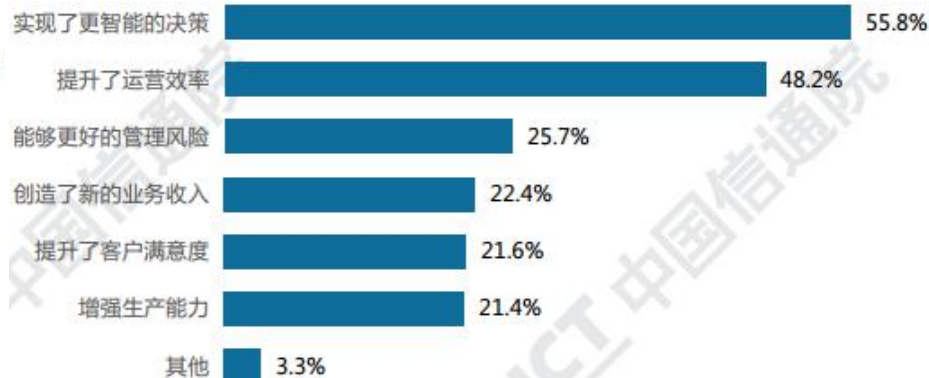
企业大数据应用场景



企业数据分析结果展现方式



企业应用大数据带来的效果



数据驱动技术，技术促进商业



数据质量和管

高质量的数据处理和管
理是输出高质量结果
的前提

挖掘分析建模

深入数据内部，用云
计算去做人工做不到
的数据价值发现

人工智能

从非结构化数据中提
取信息，引入语义分
析、图像识别等技术

图形可视化

图形化展示数据分
析结果，为使用者
提供决策依据

预测分析能力

根据可视化和数据挖
掘的结果做出预测，
风险预估和市场判断



01

/ 企业大数据应用

02

/ **企业大数据实践**

03

/ 典型案例分享



阿里大数据业务支撑

- ✓ 几千名 大数据工程师
- ✓ 几万台 服务器
- ✓ 数百PB 的海量数据
- ✓ 几十个 业务板块



新零售大数据来源

DB
GEEK

IT大咖说
知识分享平台



搜索引擎

搜索、提问数据



互联网

用户点击行为数据



社交网络

社交行为数据



电子商务

资金流、信息流数据



电子地图

地理位置、流数据



影音资料

模拟形式数据

通过“**研、定、投、优**”的闭环运营，实现对客户的**可识别、可洞察、可触达、可服务**

优：可服务

- 客户服务提升满意度
- 运营效果评估优化经营策略
- 问题溯源解决实际问题



研：可识别

- 采集集关于主体的多源数据
- 整合数据，实现统一存储和分析
- 研究分析了解主体属性与关系

投：可触达

- 实现数据驱动运营
- 新客挖掘精准营销
- 线下门店经营指导

定：可洞察

- 构建主体标签体系
- 构建企业全域数据逻辑模型
- 全景洞察企业数据资产

从需求到线索

研

数据采集与整合

内部

- 顾客数据
- 门店数据
- 经销商数据
- 商品数据

...

外部

- 楼盘数据
- 区域商圈数据
- 消费者特征数据

...

从线索到订单

定

构建企业全域数据

资产视图

标签体系

- 客户标签
- 资产标签
- 门店标签

...

全景洞察

- 商圈画像
- 消费者画像
- 门店选址

...

从订单到交付

投

精准营销

- 精准铺货
- 潜客发掘
- 营销活动
- 智能推荐
- 经营指导

交付能力提升

- 服务网点选址
- 服务能力预估
- 销售预测
- 库存优化

从问题到解决

优

客户服务提升

- 门店评估
- 服务能力评估
- 客户维系
- 经营决策

研

定

投

优

数据源系统

财务 DRP

客服 CRM

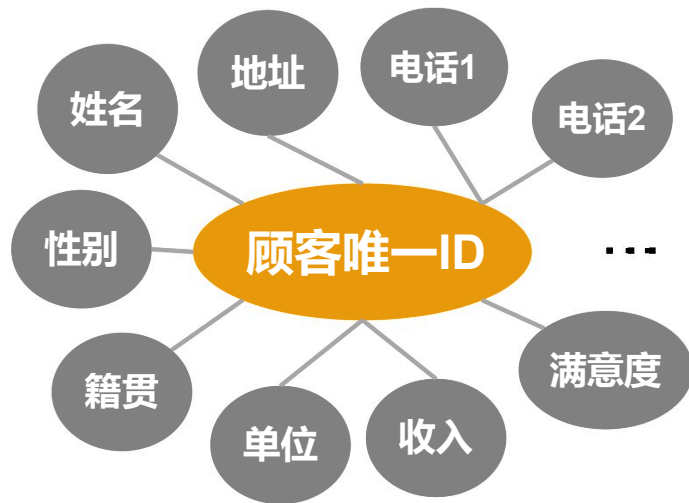
客户体验 CIF

品牌 CMS

O2O会员

...

IDMapping 技术实现数据整合



信息转换

全景画像

人群透视

人群放大

投 - 新零售精准营销

DB GEEK

IT大咖说
知识分享平台

研

定

投

优

1

数据整合

WIFI热点

- 店外客流
- 店内客流
- Mac地址

线下环境

- 交通
- 天气
- 市政规划
- 基础设施

客群特征

- 性别、年龄
- 行为偏好
- 消费偏好

ERP\CRM

- 会员信息
- 订单和收银
- 门店渠道
- 供应链

2

门店区域消费者洞察



3

营销规则制定

● 人群筛选规则

- 年龄：36~46
- 客源地：大陆地区
- 消费等级：高
- 信用等级：优

2

算法驱动精准营销



5

线下交易提升，数据回流沉淀



4

多渠道客户触达引流

线上广告投放

- 新闻门户
- 移动APP
- 视频贴片
- Banner
- 全屏广告
- 开屏广告
- 插屏广告
- 应用墙广告



线下广告投放

- 楼宇LCD视频资源
- 公交候车亭资源
- 报刊亭资源
- 户外立柱大牌资源
- LED大屏资源
- 移动媒体车体资源
- 机场、火车站、停车场灯箱.....



目录 Contents

DB
GEEK

DBSTACK.COM
IT大咖说
知识分享平台



01

企业大数据应用

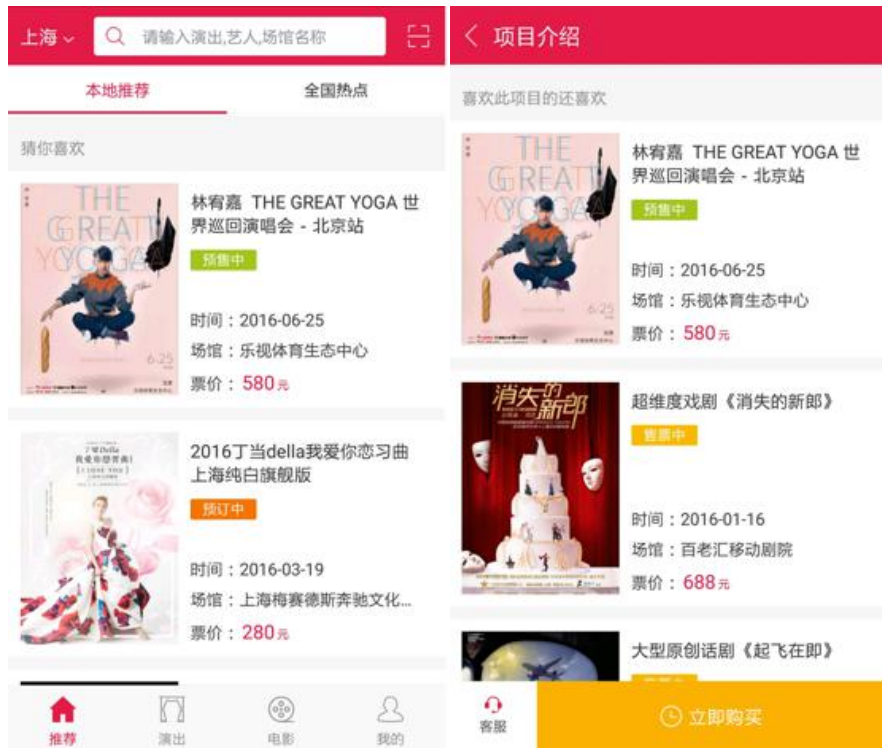
02

企业大数据实践

03

典型案例分享

案例分享 - 电商个性化推荐



大麦网个性化推荐的成果



占大麦网
整体销售额



案例分享 - 茅台云商

DB
GEEK

IT大咖说
知识分享平台

多角色

消费者 | 经销商 | 网点 | 社会化营销员 | 营销员 | 客服 | 运营 | 领导

多终端

APP | WEB | 微信 | 钉钉 | 400 | TV | 门店 (体验店)

外部
第三方
平台

酒一茶一食品一大健康

业务
& 大数据
应用



能力
开放

云商
大中台

业务中台

Aliware阿里中间件

数据中台

数加阿里大数据平台

资源平台

云计算基础资源服务

安全
管控
体系



订单数

18,772 单

茅台云商经营概况

自提时效



客单价 PRICE 单位: 元

交易额 806,417,099 元

服务网点销售排行榜

排名	交易额	网点名称
01	5.79 million	深圳市仁美贸易有限公司深圳市专卖店
02	3.34 million	汇川区兴美洋酒类经营部的经销商
03	2.77 million	集团电商公司贵阳店
04	2.28 million	哈尔滨金宇酒业有限公司哈尔滨市专卖店
05	2.19 million	吉林省金利源商贸有限公司长春市专卖店
06	1.82 million	吉林省四方商贸有限公司长春市专卖店
07	1.50 million	福建省长汀县长泰商贸有限公司龙岩市专卖店



用户画像 USERS



热销宝贝排行

宝贝名称	交易额
01 茅台酒新飞天53度 500ml	586 million
02 茅台酒30年 500ml	54 million
03 茅台酒16年 500ml	48 million

宝贝名称	订单数
01 茅台酒新飞天53度 500ml	12046
02 茅台酒53度 500ml贵州多特酒 (订酒两年)	2769
03 茅台酒16年 500ml	949

今日指数

10:00:00 13:50:00

昨日指数

17.26.44

贵州茅台酒 (新飞天) 53%vol 500ml
 用户ID: 156****2090
 发货时间: 23/03/2017 0:03:50 AM
 送达时间: 21/03/2017 4:27:59 PM
 退货耗时: 36分钟

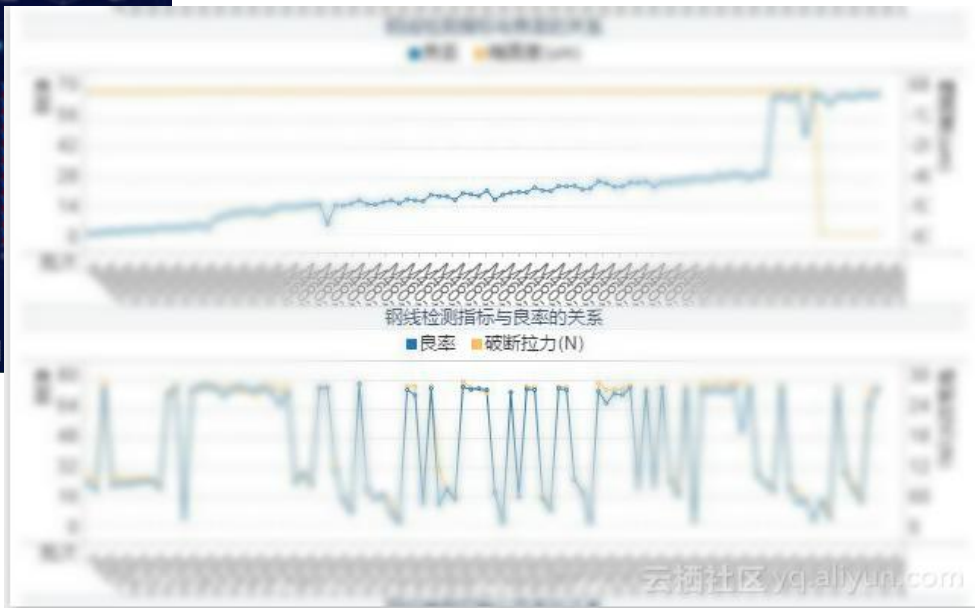
沈阳市 交易额 332,000 元

演示数据, 仅供参考

案例分享 - 协鑫光伏



良品率提升**1个百分点**
节省生产成本**上亿元**



- ✓ 良品率预测模型
- ✓ 生产监控大屏
- ✓ 关键参数监控模型
- ✓ 多维统计分析

大数据统计，分析

成交分析

客单价，成交金额，
退款金额，支付失败
订单，支付时长，支
付转化率，复购率

分享

分享人数，分享金额

流量运营

客户端/页面的pv,uv,新增用户,注册用
户,留存率,页面漏斗，转化率等

产品体验

投诉率，买家评价，
消息到达

项目运营

新增项目，在筹项目，成功
项目，失败项目等对应的数
量和金额等

数据仓库

制定数据采集策略，整合多源海量数据初步实现数据集中，依据业务主题设计并构建企业数据仓库，搭建企业大数据分析底盘，提升大数据统计分析效率，**为其其他大数据应用提供支撑。**

更**快**的时间、更**高**的效率、更**低**的成本



数据智能，让未来变成现在