

“

# 营销 数据中台 白皮书

Market Guide for Marketing Data Platform

”



Version 20190218



Contact Us [Market@miaozhen.com](mailto:Market@miaozhen.com)



秒针营销科学院 MAMS , Miaozhen Systems

## 前言

“数据中台”是阿里提出，对标的国外“Data Lake”（数据湖）的概念，出现的背景是阿里的生态系中淘宝，天猫，蚂蚁金服，盒马鲜生等业务板块每天都产生着大量有价值的数据，要在不同业务群间做到数据的互联互通，对数据价值进行最大化挖掘，需要整合各业务群的数据建立集团层面的“数据中台”，统一管理和应用数据。

对于大部分广告主来说，“数据”还是一个相当陌生的词，虽然“数据驱动”代表了先进生产力，但缺少了数据，市场部也在正常运作，那花费大量成本搭建数据中台，对于广告主的价值是什么呢？



于勇毅  
秒针系统营销科学家



# 目录

一. 广告主对营销数据中台的期望是什么? .....	03
二. 和传统数据仓库对比, 数据中台有什么差别? .....	04
三. 什么是“知识图谱” .....	07
四. 数据中台的系统架构 .....	09
五. 数据中台的数据源 .....	11
六. 数据中台的三种形式 ( Data Lake, CDP, DMP ) .....	12
七. 构建数据中台的实现路径 .....	14
八. 数据中台实现的营销场景 .....	16
九. 数据中台实现的业务数字化转型场景 .....	20
十. 构建数据中台过程中需要知道的“坑” .....	22
十一. 营销数据中台的未来发展趋势 .....	26

## 广告主对

# 营销数据中台的期望是什么？

“数据中台”作为营销技术中最奢侈的投入，是只有大型广告主才需要的资源，价值在于：

### 1. 赋予广告主数字营销的精细化操作能力

当市场部承接的数字营销预算大到一定程度，就无法通过营销人员的个人经验对营销活动进行微观操作，在拥有数据中台后，可以依靠数据+技术，驱动整个营销体系的精细化操作

### 2. 提升营销执行的ROI

这是最常规的诉求，市场部的绝大部分预算是在营销执行层面，按照每年1亿的营销投入，如果通过数据能提升1%的精准度，就能节省100万的成本，这是企业能最直接看到的真金白银

### 3. 战略视角的营销策略

在打通生产，销售，电商，服务等数据后，市场部就能看到更加连贯的全局数据，可以站在更高维度，看待营销在公司战略布局中的定位和作用

### 4. 提升市场部内部运营的整合度

当市场部内部职能划分过细，需要通过数据来串接营销运营过程中的市场研究->市场策略->营销执行->效果考核，避免内部信息不对称，提升运营效率

### 5. 加强市场部和其他部门间协作

当企业内部组织架构达到一定复杂度，市场部需要通过数据对接其他部门的运作，在企业统一的考核体系下，在企业内部证明自身价值，争取更多资源

### 6. 支撑业务的数字化转型：

“数字营销”现在不只是营销词汇，数据中台所拥有的资源（数据/IT设施/考核规则/运营人员），除了支持营销场景，也可以构建各种数字化转型的业务场景，成为CMO和CEO/CGO/CDO对话的核心资本。很有意思今天讨论建立营销数据中台的，除了市场部和IT部门，有很多需求来自更高层的CEO，COO（首席运营官），CGO（首席增长官），这些高层的诉求是通过“数据中台”来解决业务问题（例如产能过剩，人员效能，获客），支持企业的创新业务（例如新零售，金融科技，数字化管理）

## 和传统数据仓库对比

# 数据中台有什么差别？

国外著名咨询公司 Gartner 把数据管理技术分为三种：

**Data Warehouse** – Supporting mostly known data (structured, transactional) and known questions (repeatable, for broad consumption) to deliver consensus for running the business.

**Data Lake** – Supporting unknown data (less organized, raw and/or exogenous) and unknown questions (discovery- and data science-oriented) to enable exploration and innovation.

**Data Hub** – Enabling manageable and governed sharing of data between producing and consuming systems and processes.

与存储“已知”结构化数据，解决“已知问题”的传统数据仓库（Data Warehouse）相比，数据中台存储了大量“未知”的原始数据，利用数据科学（Data Science）可以在应用层面进行更多探索，帮助企业解决更多“未知”的商业问题

数字技术的革命，使广告主可以收集的数据产生了爆发，因为数据的“量变”，催生了数据管理和应用的“质变”，这是“数据中台”出现的主要原因。如果说传统的“数据仓库”面对的是“小数据”，“数据中台”处理的则是真正的“大数据”。

回到5年前，广告主能收集到的营销数据大部分来自CRM，是基于消费者“人”的PII数据（Personal Identity information），传统营销数据是以下这样的：

消费者姓名	性别	年龄	历史购买金额	最近一次购买时间
张三	男	20-25	1000人民币	30天前

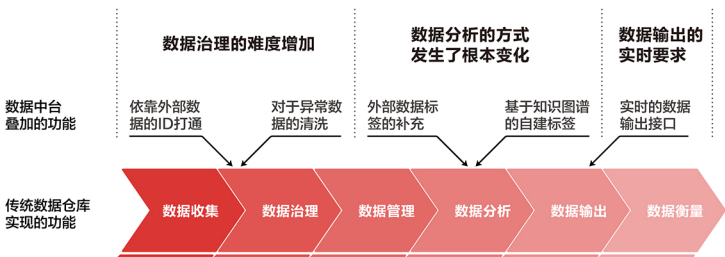
这些数据来自广告主的运营过程，数据量也相对较小，每年能收集的数据很难超过TB级别。数据的使用层面也相对简单，一个初级的数据分析师，可以依靠数据词典轻易读懂每条数据的含义，依靠传统统计学和算法工具就可以完成数据分析，支撑业务应用。

例如CMO想针对贡献了80%收入，但是过去2周没有任何采购的高消费用户群体做一次campaign，不到10行SQL语句就能抽取这些目标消费者数据。

今天广告主收集了大量描述消费者行为的“大”数据（在后文会讲数据中台的主要数据类型），这些数据是基于消费者“设备”的数字数据（Digital Data）：

设备ID	设备型号	浏览URL	访问时间	经度	纬度
A1B2C3	iPhone 8	www.sina.com	晚上11:30	1.222	2.111

消费者使用的数字设备（手机，电脑，Pad等），每天都产生百万级的行为数据，广告主很容易在几周内收集到TB级的数据，这些大数据的管理和应用对数据中台提出更高的要求，主要技术革新包括以下三点：



## ■ 数据中台的技术革新1：数据治理的难度增加

传统营销数据大部分是基于email地址，手机号和姓名，对消费者进行的识别，不同数据源的打通难度不大。但消费者的行为大数据是基于多种ID的（手机号，设备ID，Cookie ID，Mac等，具体在后文介绍），依靠广告主自有能力，很难实现ID的打通，打通的比率取决于广告主的数据量大小，在广告主的数据量没有达到足够海量前，需要依靠外部数据资源实现。

此外，消费者行为大数据中异常数据的比率远高于传统数据，例如广告主收集了1000万条浏览过自己主页的设备ID，这里面可能涉及到爬虫，无效流量，无效浏览等多种场景，真正有价值的消费者数据量甚至会少于异常数据，这时需要通过算法或者外部数据资源对这些无意义的异常数据进行清洗



## ■ 数据中台的技术革新2：数据分析的方式发生了根本变化

消费者行为大数据的解读没有以往这么“直接”，知道了消费者浏览的URL，知道了他们在每个页面的停留时间，知道了他们经常出现的经纬度，这些大数据如何和业务关联和使用呢？

如果把这些原始数据比喻成蔬菜，在端上饭桌实际应用前，需要经过一个“烹调”的过程，即把原始大数据简化成业务侧能读懂的标签，“烹调”的方式有2种：

### a. 基于广告主收集的ID，到外部直接采购现成标签：

例如广告主收集到浏览过自己官网的设备ID，想知道这些设备ID背后的消费者画像，可以对接外部数据源，对这些ID补充年龄，收入等标签，这个过程被称为Data enrichment

### b. 通过“知识图谱”进行数据结构后，建立自定义标签：

例如广告主收集了某消费者一天1000条位置数据，如果手上有全国所有小区经纬度，就能知道这个消费者晚上住在哪个小区。如果有每个小区房价，就能去猜测这个消费者的收入水平。如果有全国办公楼经纬度，就能知道这个消费者的大致工作。如果有全国高尔夫球场经纬度，就能知道这个消费者是否有打高尔夫的习惯…

以上这些对于原始数据结构化的“词典”，就被称为“知识图谱”（在后文中，单独有一章节解释），有意思的是，同样的行为数据，连接不同知识图谱后，可以获得不同的洞察结果和客户标签体系，知识图谱是广告主解读大数据，建立自己洞察体系的那把“钥匙”

## ■ 数据中台的技术革新3：数据输出的实时要求

传统从大型数据库中提起数据需要花费数分钟甚至数小时，而今天很多大数据的应用场景都是毫秒级别，例如某广告主想让不同消费者浏览自己主页的时候，看到不同的内容（千人千面），从技术上需要实现毫秒级别完成以下动作：

消费者ID识别->消费者画像提取->展示图片匹配->图片加载

当以上闭环无法在毫秒级完成，做不到实时输出，就会出现消费者几秒打不开企业官网，失去耐心而直接选择关闭的情况



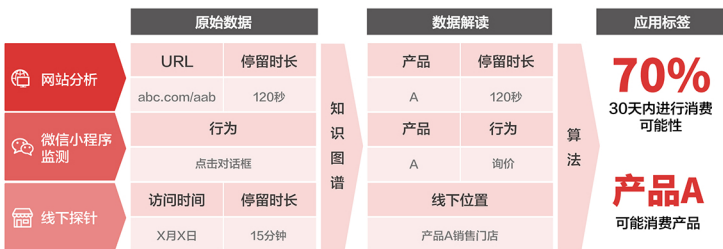
# 什么是“知识图谱”

在数据中台搭建过程中，最难的不是IT层面的数据管理，而是将海量大数据化繁为简，变成业务侧能看懂的标签的“分析”过程。

上文提及了分析的两种方式，现在绝大部分广告主大走的就是第一条路线：对于数据收集主要集中在消费者ID，再基于这些ID到外部匹配可用标签。

这种模式的好处是快速落地，缺点是外部标签成本高企，而且由于外部供应商缺少行业理解，标签缺乏精准度。从中长期来看，广告主在使用外部标签遇到瓶颈后，必定转向建立自身标签体系的第二条路线。

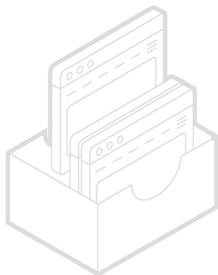
和传统基于统计学的分析不同，基于大数据的分析的第一步并不是“算法”，而是借助“知识图谱”对于底层数据的结构化，下面是一个例子：



## 某广告主收集到了消费者的三条行为数据：

通过“知识图谱”，发现：

- 访问了某URL，并停留了120秒（网站分析数据）  
第一条数据的URL代表的是产品A的介绍。
- 在微信某小程序中，点击了某个对话框（小程序监测）  
第二条数据，这个小程序的对话框是产品A的询价
- 出现在某线下销售门店中（智能探针数据）  
第三条数据，是专门销售产品A的线下门店



从定性角度已经可以初步判断这个消费者对于产品A的高度兴趣，但是有这样行为的消费者可能成千上万，在页面停留120秒，和在线下门店停留15分钟，这样的数据如何定量呢？

通过“算法”，可以发现有这样行为的消费者，在未来30天内进行购买A产品的可能性是70%，“A产品”+“70%”是业务侧最终能读懂的，并且横向比较应用的标签。

在今天的数字技术收集的大数据中，常规营销用的“知识图谱”包括：

- 网页URL的知识图谱
- APP行为的知识图谱
- 第三方平台（例如微信公众号）行为的知识图谱
- 地理位置的知识图谱
- 广告主自身产品标签化的知识图谱

不同颗粒度的“知识图谱”在解读同一条行为数据时候，得到的洞察深度也会不同，例如：

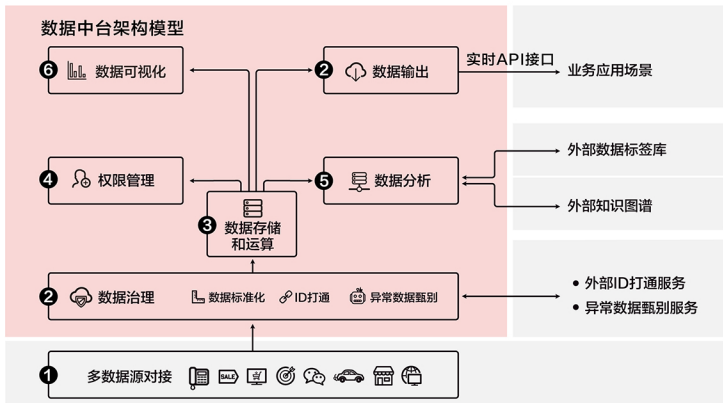
- 这个消费者正在看我的竞品
- 这个消费者正在看我的竞品的产品A
- 这个消费者正在看我的竞品的产品A，最新的促销
- 这个消费者正在看我的竞品的产品A，促销价格比我的对标产品低15%
- .....



“知识图谱”的建立过程，往往是基于消费者全量数据，用穷举法去作分析得到的，考虑到数据的多样性，高精度的“知识图谱”必定是借助AI实现的，例如掌握了消费者访问的海量URL数据，需要拿爬虫工具去获取所有URL对应页面上的文字，通过“语义识别”技术，给每个URL贴上对应标签。

# 数据中台的系统架构

下图是数据中台的大致模型，从IT层面有以下几个模块：



## 1. 多数据源对接

从各种数据源提取数据，放入数据中台，数据类型包括：

第一方数据：	广告主自己系统上产生的数据，例如CRM，售后服务，会员系统等
第二方数据：	广告主在外部系统上产生的数据，由外部系统通过API提供，例如电商数据，广告监测数据，微信公众号数据等
第三方数据：	广告主直接获取的外部数据资源。需要强调的是，和业外人想象的不同，第三方数据交易并非一手交钱一手交数据，今天数据生态圈的法律合规要求，第三方数据交易不允许广告主直接采购消费者的ID，数据服务商智能基于广告主提供的消费者ID提供数据服务



## 2. 数据治理

在获取不同数据后，对数据的治理包括三个任务：

A. 数据标准化：	例如不同数据源对于消费者性别的描述有“男-女”，“先生-女士”等多种写法，在做数据整合的时候，需要统一不同数据源对于相同含义的描述值
B. ID打通：	上文也描述过，不同数据源对于消费者的识别是基于不同ID的，数据源的拼合需要ID打通。对于大部分广告主来说，无法拥有像BAT那样的数据量，BAT每天能收集十亿级别ID发生的千亿级别的行为数据，他们的确能做到依靠自己的数据就能打通不同数据源。大部分广告主在没有这么大量的数据和消费者活跃度的情况下，ID打通需要依赖外部数据源
C. 异常数据甄别	和ID打通一样，异常数据的甄别在广告主自身数据不够大的情况下，同样依赖外部数据。例如某个ID每天会点击2次某广告主的广告，这个行为相对正常。但是如果放到全行业，这个ID也许每天会点击1万次各个广告主的广告，这就很明显是流量作弊了

## 3. 数据存储和运算

在数据完成治理后，统一放入数据库进行管理和运算。由于数据量过大，在单一服务器上无法完成存储和运算，就涉及到大数据的分布式运算，云计算等复杂IT层面的管理。

按照数据存储和计算的地方，可以分为营销云（数据存在第三方的云平台上），自有云（广告主自己的IT环境中），混合模式（非敏感数据存放在第三方云，敏感数据在自有平台），以上不同的方式有着不同的成本，数据安全和合规要求等

## 4. 权限管理

数据中台的目标是支撑市场部甚至全公司不同业务场景，也意味着从公司高层到底层外包员工都需要从数据中台提取数据，为了防止数据泄露的问题，需要对不同用户，不同场景进行权限分级管理。例如负责接听电话的客服人员，在客服系统中就不可以看到消费者的全名和实际手机号。

## 5. 数据分析

上文也有描述，在使用“营销大数据”前，需要对数据通过分析生成业务侧用户能看得懂的标签，分析的过程包括非实时的传统数据挖掘，依靠AI人工智能的实时分析等。在分析过程中，广告主很难短期内积累自身的知识图谱和高质量的标签，需要依赖外部的数据能力。

## 6. 数据可视化








虽然数据中台是由技术背景的团队进行运维，但是实际使用的是对数据完全没有感觉的业务侧人员，对比成亿条原始数据，业务侧也许只需要一个饼图就能得到商业结论，因此可视化是真正让业务侧使用数据中台的基本工具。

## 7. 数据输出

数据中台的产出，除了数据可视化展现的洞察外，还会对接不同的业务系统，通过数据来驱动业务场景，例如程序化广告，新零售，动态定价等，很多业务场景需要毫秒级的查询和输出，也是IT层面需要解决的技术问题。

# 数据中台的数据源

今天广告主常规的数据源包括四大类：

	广告监测/网站分析	CRM	线下探针	平台数据
可识别ID	Device ID	手机号	MAC#	微信Open ID
	Cookie Id	—	人脸	其他平台ID
数据收集技术	 网站分析	 会员系统	 门店数据	 社交媒体数据
	 广告监测	 CRM数据	—	 第三方大数据

## 1. 基于设备ID，cookie的网站分析和广告监测数据：

描述的是消费者对于互联网广告，以及广告主官网，APP 等自由平台的浏览和点击行为

## 2. 基于手机号的PII数据：

包括会员系统，CRM 数据等，描述的是消费者的会员信息，历史采购记录等

## 3. 基于Mac#和人脸识别的线下数据：

通过智能探针技术，摄像头+人脸识别技术，收集到的消费者线下行为数据

## 4. 基于外部平台ID的平台数据：

包括微信的 Open ID，各个大数据方的自建 Super ID 等

这些数据的打通虽然有很多技术途径，例如在微信中建立SCRM会员体系，消费者在微信公众号中进行手机的实名认证，就能打通手机号和微信Open ID，例如一个消费者在手机和电脑上用同一个用户名登录了某个APP，就知道手机的设备ID和电脑浏览器的Cookie是同一个人等等。

考虑到一个消费者可能有多个终端，多个手机号，会经常换手机（设备ID变化），会借用别人终端登录APP，还有网吧共享电脑等复杂形式的存在，当广告主自身数据量不够大的时候，很难依靠以上这些技术手段达到很好的数据打通效果。



# 数据中台的三种形式

(Data Lake, CDP, DMP)

今天营销数据中台在技术上分为三种：

## Data Lake (数据湖) :

技术难度最重的一种，定位是企业业务层面的数据大集市，会整合全公司各种数据源，支撑的不只是营销场景，还包括企业个性化的业务场景，往往由企业的最高层直接领导，目标是帮助企业进行数字化转型。由于在数据对接和数据处理层面需要处理大量定制化数据源，因此构建过程往往以年为时间单位。

## CDP (Customer Data Platform) :

技术难度稍低的数据中台，定位是营销层面的数据大集市，目标是支撑各种利用广告主自有数据的营销场景。因为CDP通常只对接标准化数据源（例如两个广告主用的是同一款标准化CRM，他们的底层数据结构都是一样的），数据治理和数据管理相对容易，因此实施周期以月为单位。

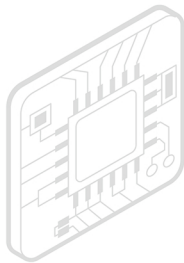
## DMP (Data Management Platform) :

定位是支撑以程序化广告为主的实时营销场景，和Data Lake, CDP的最大不同是毫秒级数据输出。因为DMP主要用到的是广告监测数据，网站分析数据和第三方大数据，数据格式相对固定，因此实施难度最低。

因为在程序化广告运营过程中，DMP的数据会暴露在公网上，被外部供应商和媒体调用，因此只能存放广告投放使用的匿名数据（ID和标签），不能存放其他敏感信息（姓名，手机号，地址，历史购买等），投放的标签也需要脱敏（例如某ID的标签是A，实际代表着过去3周没有购买的高消费客户，但这个定义只有广告主内部数据分析团队知道）。

在数据存储和运算中，DMP可以构建在广告主自己的IT环境里（称为第一方DMP），也可以放在第三方营销云上（称为SAAS DMP，或者第三方DMP）。因为第三方DMP会预先对接媒体，自带算法，标签和数据治理模块，能把DMP的实施时间缩小到几周，可以大大降低实施成本。

而Data Lake 和CDP上存储了广告主的敏感数据和商业机密（例如历史采购信息），因此只能构建在广告主自己的IT环境下，从技术角度更加复杂，成本也更高。



通过下表简单列举三种技术，四种形式的主要差别（分开第一方和第三方DMP）

	Data Lake	CDP	第一方DMP	第三方DMP/SAAS DMP
使用者	全公司	市场部	市场部	市场部
实施周期	年	月	月	周
目标业务场景	企业数字化转型	第一方数据支撑的营销场景	实时程序化广告	实时程序化广告
数据类型	企业内部各种数据	所有营销数据	匿名的广告投放数据	匿名的广告投放数据
数据存储环境	企业内部 IT 环境	企业内部 IT 环境	企业内部 IT 环境	营销云
数据输出	非实时	非实时	实时	实时
投资	千万级	百万级	百万级	十万级

通过上表可见，Data Lake和CDP相对复杂，但是在应用场景上有更多遐想空间，而DMP则就是为程序化广告而生，是最轻量的营销数据中台技术。三种形式无所谓谁优谁劣，但有着各自的适用对象：

### DMP:

适合在广告投放领域投入大量预算的广告主，例如快消，汽车，化妆品，娱乐等

### CDP

适合消费者复购比率高，自有体系可以收集大量消费者数据的广告主，例如奶粉，零售，运动用品，化妆品等

### Data Lake

适用于有大量数据驱动应用场景的广告主，例如快消，零售，汽车

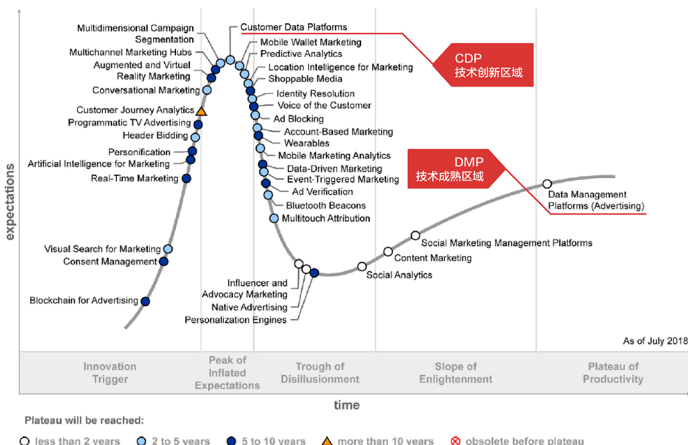
# 构建数据中台的实现路径

广告主对于数据中台的实际需求往往会模糊化三种技术的边界，笔者之前接到的需求，大部分是三者的平衡点。

例如某汽车广告主有大量预算在程序化广告，希望通过DMP来提升投放效果，同时又想把数据融入4S店的智慧门店体系，做到4S店的接待员在陌生消费者入店的那一刻，手持设备上就能显示这个消费者的客户画像，这个“消费者识别”场景需要接入店面智能探针数据，并且输出消费者标签数据到店员的手持设备，实际用到的是DMP+Data Lake的技术和理念。

这是Gartner 2018年公布的数字营销Hype Cycle图，描述了不同技术的成熟度，DMP已经被归类在技术成熟阶段（实施成功度高，成本低，有足够多的应用场景），而CDP仍在技术创新阶段（实施风险相对较高，成本较高，广告主有过高期望值），而Data Lake并非数字营销专有技术，因此并未被收录。

Figure 1. Hype Cycle for Digital Marketing and Advertising, 2018



Source: Gartner (July 2018)

在实施过程中，三种技术的难度和成本是 Data Lake>CDP>第一方DMP>第三方DMP，当广告主构建的数据中台需要融合以上三种技术时，技术路线有以下三种

## 1. 第三方DMP -> 第一方DMP-> CDP-> Data Lake

这是实现风险最小的安全路径，从技术和商业应用上都是由浅入深，整个过程的实施周期大约为2年，每3-6个月有清晰的里程碑，资源投入是一个从小到大的渐进过程，适合对于数据中台的态度是摸着石头过河的广告主

### 安全路径

周期大约为**2**年，每**3-6**个月有清晰的里程碑

- 通过几周时间搭建低成本，快速实施，高实施成功率的第三方DMP，看到数据中台对于广告投放效果的优化作用
- 经过3-6个月的验证，把第三方DMP迁移到广告主自有IT环境，形成第一方DMP
- 在3-6个月系统稳定后，追加对接更多营销数据源，寻找更多的营销应用场景，实现DMP + CDP的功能
- 再3-6个月的实施和探索，对接更多业务数据源，寻找更多的业务应用场景，实现DMP + Data Lake的功能

## 2. CDP -> DMP + Data Lake

这是实施周期和难度相对适中的路线，大约实施周期为1.5-2年，开始阶段内部营销数据的整合会有难度，但是控制在市场部内部，后期逐渐做大，适合数字营销主要基于自身数据展开的广告主（例如奶粉，化妆品等）：

### 平衡策略

周期大约为**1.5-2**年

- 使用6-9个月整合广告主内部数据形成CDP，主要营销方式通过电子邮件，短信，微信等进行
- 通过6-9个月时间，增加CDP的实时输出功能，以及寻找更多业务应用场景，对接更多业务侧数据源，使CDP进化为DMP+Data Lake功能

## 3. Data Lake + CDP + DMP

这是最复杂但也是最快实现的路线，大约1-1.5年，直接把三种技术涉及的各方面同时实施，对IT实施能力有很大的挑战，这样的项目往往会由企业的COO/CIO级别直接挂帅，需要短期内投入大量资源，只适合对数据中台有很高期望值，目标是短期内快速推进数字化转型的激进广告主。

### 快速实现

最复杂但最快路线，大约为**1-1.5**年

# 数据中台实现的营销场景

如果用一句话形容数据中台在营销的作用，就是“赋予了市场部在数字营销领域的微观操作能力”

今天广告主在数字营销的投入越来越大，但同时流量红利越来越稀薄，造成粗放式营销方式的效率越来越低，广告主在遇见数字营销的瓶颈后，需要利用数据驱动整个营销闭环，提升营销效率。具体落地的主要场景包括以下：

## 1. 程序化广告投放闭环

对于很多2C的大型广告主，有大量预算在互联网广告领域，传统的做法有两种：

- 在媒体的固定点位（广告位置）+CPT（包时间段）进行投放
- 广告主根据产品的客户画像，利用媒体/第三方数据源提供的数据标签进行精准投放。例如某母婴广告主的目标客户是居住在1-3线城市，25-35岁的女性，广告主会询问各个供应商谁有所在城市，年龄，和性别三个数据标签，然后对接媒体资源后进行投放

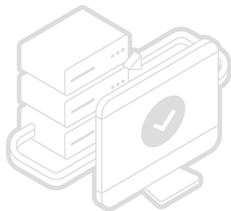
以上模式的主要弊端包括：

### 第三方数据标签质量无法控制：

外部数据源不会公开自己的算法，电商可能根据消费者经常购买女性产品，社交媒体可能根据消费者的注册性别为女，来判断这个消费者是女性，这种相对简单的逻辑和实际情况可能会有很大程度的偏差

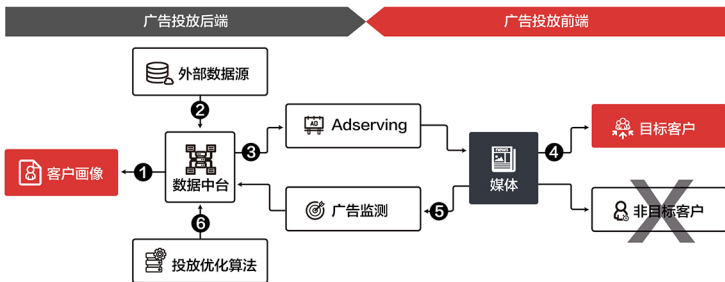
### 广告主无法输入个性化的需求：

例如某母婴广告主希望接触正在购买一段奶粉的客户，这时除了广告主自己CRM的客户，外部数据源是没有能力提供这个字段



当广告主拥有数据中台后，可以形成不断优化的程序化广告投放闭环：

- 1 广告主通过调研，历史经验等方式假设目标客户画像
- 2 在数据中台提取广告投放人群包，在建立初期数据不完备的情况下，需要部分依赖外部数据源
- 3 数据中台连接Ad-serving（广告服务器）后，设置投放策略（广告投放的物料，频次控制，匹配媒体等）
- 4 Ad-serving和媒体做对接，只有广告主指定的目标消费者出现在媒体的时候，才进行广告投放
- 5 通过广告监测，回收广告投放中产生的数据回到数据中台
- 6 通过投放优化算法，在数据中台中优化客户画像，再进入下一次运转



以上是数据中台的DMP模式最常规实现的场景，具体应用以汽车行业为例：某广告主给目标客户群进行大范围的广告推送后，某消费者有过一次深度广告浏览（例如点击了车型询价，并且页面停留时间超过1分钟），当消费者的ID和行为被捕捉后，广告主可以在数据中台中标示这个ID作为重要目标，当这个ID出现在任何一个媒体时，都投入资源让这个消费者只看到自己的广告。而从消费者视角角度，就是一次深度浏览某车企的广告之后，看到所有的广告都可能是这个车企的，直到完成销售。

## 2. 贯穿业务全链的客户洞察体系

在企业业务链条上有三个节点关心客户洞察：

- 1 产品研究：需要清晰对产品进行定位，指出产品的目标客户
- 2 营销人员：需要为每次营销选择目标客户
- 3 销售人员：需要了解已经进入购买环节的客户画像

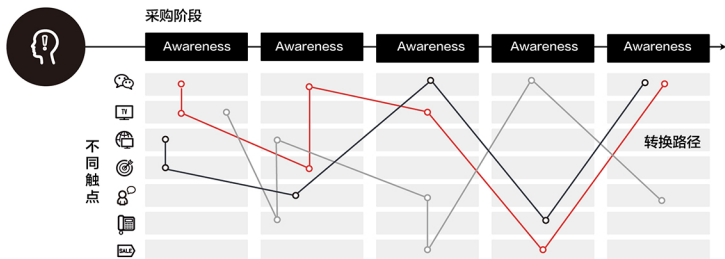
在实际运营过程中，经常会发现大家讨论的客户洞察并不在同一个口径中，例如某企业定位了一款针对“运动人群”的产品，对于产品研究人员来说，“运动人群”意味着每天跑步，打球的人，而对于营销人员来说，广告投放的“运动人群”可能是喝着可乐躺在沙发上，在手机上看足球比赛的胖子，而对销售人员来说，“运动人群”就是购买这款产品的客户。

由于口径的不一致，使得业务链条产生了前后端的信息不对称，降低了运营效率。而数据中台的建立，可以以统一数据为基础，结合调研，内部经营分析等方法，确保各部门统一口径的客户洞察体系。



### 3. 客户体验和消费者转换路径的最优化

在消费者的采购决策链上，可以从很多触点获取信息，在数据中台上整合多种数据后，广告主可以基于消费者的采购阶段，在最适合的触点上，以最适合的传播频率，传递针对性的内容，达到更好的“客户体验”（Customer Experience），实现消费者转换路径的最优化。



客户体验的实现包括了以下几个层面：



**客户画像和传播内容的匹配：**在今天信息过剩时代，要找到正确的内容来打动消费者，需要通过算法连接客户画像，和海量广告主能掌握的内容素材资源



**触点的最优化：**不同触点有着各自的覆盖面和成本，广告主可以根据消费者的重要性，选择触点（例如电话外呼单个成本在20元，短信0.1元，互联网广告小于0.01元），频次（单日接触客户次数的上限），竞价策略（很多媒体的程序化广告需要通过多广告主竞价），触点的最优化是营销成本精细化操作的重要课题



**接触频度控制：**数据中台实现的另外一个场景，对同一个消费者的过度投资，不光造成成本浪费（例如一天让同一个消费者看100次同样的广告），还会引起消费者的反感。过往的研究，发现同一个消费者每天在不同触点接收到3-7次同一品牌的广告有最佳效果



**失联消费者的召回：**当消费者在某个触点改变了联系方式（例如更换手机号，微信取消关注等），广告主仍能保有和这个消费者在其他触点形式的接触，对比发掘新消费者，失联消费者的召回在成本上更低

## 4. 和业务对接的营销考核体系

对于企业来讲，收入是唯一有真正价值的考核标准，但是在大部分行业中，营销对于收入的直接拉升作用无法定量证明，因此传统的营销考核体系，都是建立在和销售数字有一定距离的中间变量上。在拥有数据中台后，广告主打通了前端营销和后端销售数据，可以基于销售的数据验证营销的效果。



**自有销售平台的验证：**在打通营销ID（设备号，微信Open ID等）和销售ID（手机号）后，可以直接验证接受过营销的消费者带来的最终销售



**第三方电商的验证：**京东和阿里都提供了Ad to sales的销售验证体系，广告主可以从数据中台提取接受过营销的消费者ID，倒给电商平台后，由电商出具这些ID在看到广告后的一段时间内，在电商平台上购买此商品的数量和总金额（但是无法出具清单级具体购买的消费者ID）



**线下到店的验证：**和第三方电商的Ad to sales模式类似，掌握了消费者线下定位数据的供应商（例如在线地图），可以验证接受过营销的消费者ID，在固定时间内出现在线下店铺（通过地址围栏技术圈定范围）的消费者数量（但是无法验证这些消费者出现在店铺后，是否有购买，也无法提供消费者的ID清单）

## 5. 行业个性化营销场景

在不同行业的数字营销中，除了“精准营销”外有很多场景可以通过数据中台来做优化。例如在汽车，教育，B2B等行业，“商机”是营销的最终产出物，在消费者递交需求（商机）后，广告主的外呼团队会跟进核实，最终转交给销售团队跟进。

广告主在这个场景下遇到的问题，是低质量的商机往往会浪费大量呼叫中心成本（平均每个电话将近20元），如果能通过前期数据的筛选，选择高质量的商机跟进，是广告主提升ROI的主要诉求。

通过数据中台上的网站分析数据，能知道这个消费者递交商机时的操作行为，例如如果消费者从打开页面到递交商机只有0.1秒，或者同一个设备号，关联了多个手机号进行商机递交，很明显是虚假商机，根本不需要跟进。

通过数据中台上的客户画像，就能判断消费者的购买能力和未来可能购买的时间点，帮助广告主排定商机跟进的优先级。

## 6. 和大型互联网企业的营销资源互换

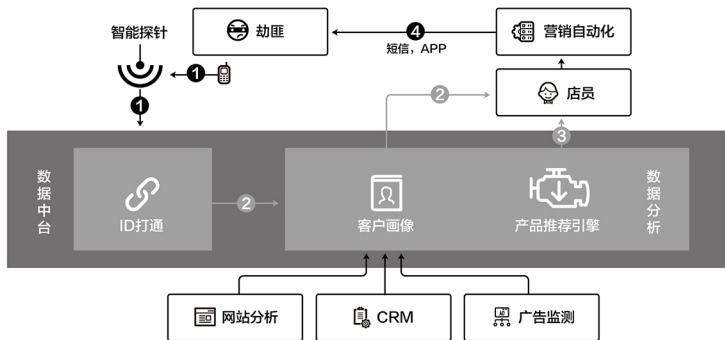
在大型广告主和掌握了数据，媒体，流量等核心资源的大型互联网企业（例如BAT）的合作中，双方可能不再是简单的甲方-乙方，而是复杂的共生-共荣-博弈的关系。

今天互联网企业都有着深入不同行业的诉求，而对于行业的理解，体现在对数据的收集，分析，运用的过程中，获取这些行业大数据，是互联网企业触摸各个行业实质的最快途径。当大型广告主建立数据中台后，数据中台上的数据和运营体系，可以成为进行个性化营销资源互换的核心筹码，例如在车联网，新零售等场景。

## 数据中台实现的业务

## 数字化转型场景

前文也描述过，为了营销目的创建的数据中台，同样可以驱动数字化转型，以下是智慧门店的场景，来自于全球领先的营销技术公司Adobe：



### 1. 劫匪进入银行抢劫

当他们进入银行的那一刻，随身携带手机和银行智慧门店体系中的智能探针进行了交互，完成了客户身份的识别

**技术路径解读：**智能探针技术可以收集进入店面，并且开着热点搜索的手机Mac#（基于网卡的唯一识别号），作为消费者众多ID之一，要读取其他数据源的消费者行为数据，需要打通Mac#和消费者其他ID，因为在后面场景中，店员叫出了劫匪的名字，而实名数据对于银行的合规收集路径，只能是CRM数据，所以至少需要匹配上劫匪的手机号和银行账号。

## 2. 在识别了客户后

银行柜员叫出了劫匪的名字，并且知道劫匪甲的诉求是房贷，劫匪乙的诉求是学费贷款

**技术路径解读：**要实现这个场景，数据中台至少对接了

- CRM：劫匪的实名数据
- 网站分析：劫匪甲在银行官网上的浏览行为，主要集中在房贷信息
- 广告监测：劫匪乙的电子邮件浏览和点击行为，主要集中在学费贷款信息

要识别劫匪的需求，首先要打通门店收集的Mac#要和网站分析，广告监测的Cookie/device ID。其次要甄别劫匪海量行为数据中的真实诉求，需要依赖知识图谱+算法：

- 劫匪浏览的URL和银行产品的关联关系，通过“知识图谱”，去解读劫匪每个行为对标银行的产品
- 在海量的行为数据中去除噪声数据，例如劫匪停留时长很短的网页不代表真实需求。在众多行为中，甄别出被关注最多的产品，虽然原理很简单，实际对于海量的数据需要复杂的算法支持

## 3. 数据中台的算法模块在不需要柜员思考的情况下，实时推荐了最适合劫匪，在公开渠道看不到的利率方案

**技术路径解读：**可以想象银行有着庞大而复杂的产品体系，一个前台柜员不可能了解所有产品，要实时匹配消费者的需求，需要银行有着强大的AI算法体系支撑。

这个场景最妙的地方，是柜员给出的利率是在公开渠道看不到的定制化产品，在识别劫匪是高度价格敏感型的情况下（否则也不会来打劫），能实时给出点对点的定价，保证了这个价格的市场不透明性。

说到“精准营销”，营销人普通想到的都是“千人千面”，给不同类型的消费者推送不同的营销内容，其实“精准”还有另一个极致“个性化定价”，利用数据完成产品定价的信息不对称，对于广告主来说，是真实能看到的真金白银。

## 4. 银行柜员直接通过短信把利率给到劫匪，并且在APP中锁定这个利率让劫匪可以办理

**技术路径解读：**为了确保营销质量，营销触点的执行往往集中在中央的市场部，但在这个场景下，通过营销自动化工具，银行将营销触点下沉到一线销售人员。

营销自动化工具连接了消费者数据，AI算法驱动的产品推荐，以及触点工具（短信和APP），一线销售人员在被框定的范围内仍保有一定的灵活性，可以根据对于消费者的主观判断，实时选择发送的内容，趁着消费者的冲动期，推进销售的迅速落地。

这个场景已经很难定义是数字营销还是业务场景，数据中台起的作用，是打通了多种数据，驱动了智慧门店+营销自动化两个应用场景的实时发生，实际效果也有目共睹。

## 构建数据中台过程中

# 需要知道的“坑”

至此，笔者介绍的都是数据中台对于营销和数字化转型的正面帮助部分，那么在构建数据中台的过程中，是否有潜在的“坑”，最终造成项目的失败呢？以下是笔者所见，或者被最常被问及的几个问题：

### 1. 数据中台驱动的精准确营销，造成营销ROI的降低

看到这个标题您可能会一头雾水，之前描述的都是数据中台的作用，“精准营销”代表的是先进生产力，目标不就是提升ROI吗？

举个例子，当广告主按照常规的CPM方式（千次展示计费）进行广告投放的时候，按照CPC（每次点击费用）计算效率，如果粗放式经营CPM媒体计价20元，每个CPM带来10次点击，那效果CPC=2元。

当建设完数据中台后，CPM仍然按照20元计算，投放效果提升3倍，每个CPM带来30次点击，那CPC是不是等于0.67元，同样提升3倍呢？

我们再考虑两个额外成本：

#### 数据成本：

每条精准数据标签按照0.2元计算，复用5次，那就是说每个CPM叠加了40元成本（ $0.2 \times 1000 / 5$ ）

#### 系统和人员成本：

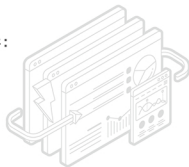
包括数据中台搭建，和维护人员成本，按照一年200万计算。再考虑这个广告主每年投放10万CPM（200万的广告投放经费），每个CPM叠加了20元成本（ $200万 / 10万$ ）

当叠加了数据，系统，人的成本后，精准投放的总CPM=80元（20元媒体+40元数据+20元系统和人），按照30次点击，实际CPC=2.67元，高于原来粗放式经营的2元，而且这还不考虑数据中台系统实施失败的风险。此外，当广告主提出要基于自己数据做广告投放的时候，媒体也知道这都是优质流量，如果优质流量都卖完了，长尾流量都卖给谁呢？因此媒体对于精准营销往往会提出涨价的要求。



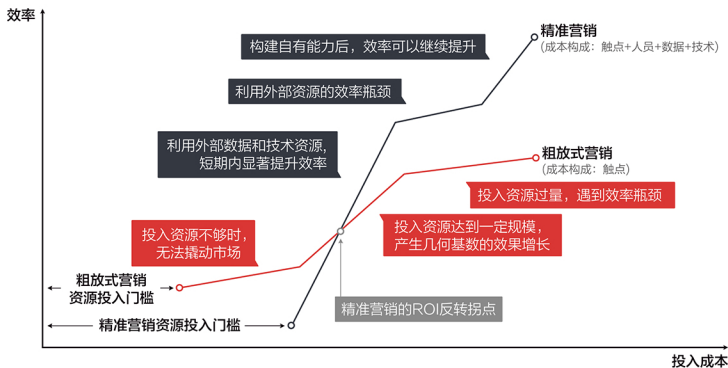
综上，当广告主希望依靠数据中台提升营销ROI的时候，需要达成以下条件：

- 价格足够低廉的精准数据标签
- 足够的数据复用度，摊薄数据成本
- 足够的营销次数，摊薄系统和人员成本



我们把符合以上三个条件，精准营销的效果真正超越粗放式营销的资源投入节点称为“精准营销的ROI反转拐点”，这个节点对不同广告主是不一样的，因此在构建数据中台时，如果主要以营销ROI提升为目标，笔者建议是第七章中提及的数据中台建设“安全路径”：先建设小成本的第三方DMP，当摸索到这个定量拐点后，再考虑是否追加投入建设更复杂的CDP和Data Lake。

以下是精准营销和传统粗放式营销的成本/效率对比图：



- 精准营销的成本构成比粗放式营销更复杂，因此资源投入门槛更高
- 在广告主的资源投入在“精准营销的ROI反转拐点”之下，粗放式营销的效率高于精准营销
- “精准营销的ROI反转拐点”的资源投入定量，需要根据具体广告主进行摸索
- 粗放式营销是反Z字结构，当投入不够，无法撬动市场的时候效率很差；当资源达到一定体量，就发生效率的飞升；当资源投入过量的时候，遇到瓶颈
- 精准营销是闪电型结构，初期投入利用外部资源（数据，人员，系统）可以获取效率的快速提升；但是外部供应商无法真正了解广告主的行业特性，很快会碰到效率瓶颈；当广告主构建自身的能力后，效率出现第二次飞升，而且有很大的遐想空间

## 2. 数据中台里的数据孤岛

在上文中介绍了广告主能收集到的清单级数据，按照消费者ID可以分为四大类，在建立数据中台的过程中，打通不同数据源看上去是一个天经地义的事情，其实却是最大的技术难点。在国外有专业的ID打通公司（Data Onboarding），虽然成本高昂，但是多数据源的高打通率仍有可能，而国内由于数据合规等原因，并没有供应商提供单一的ID打通服务。

建立5W-1H的客户画像是数据中台最基本的应用，但是现实当中，低打通率造成的数据无法有效整合，得到的客户画像仍然是割裂的。数据中台起到的作用只是“把很多数据放在一个数据库中”，数据和数据之间，仍然处于无法打通的孤岛状态。

理想情况下的客户画像



- 网站分析
- 会员系统
- 广告监测
- CRM数据
- 门店数据
- 社交媒体
- 第三方大数据

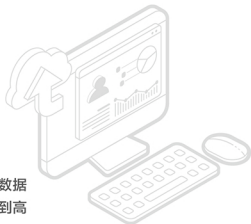
现实中低打通率的客户画像



广告主依靠自己的能力，常规的数据打通方式包括：

- 微信SCRM中的消费者手机认证，可以打通手机号和微信open ID
- 会员APP监测，可以打通Mac#和设备ID
- 广告监测数据，基于IP分析，可以打通不同设备ID
- 各种活动页面上留下注册手机号

以上办法，不同数据源的打通比率，高度依赖广告主收集到消费者的数据量，由于几乎没有广告主的数据量能达到BAT级别，无法在自有体系内达到高打通率，因此有效的数据打通，需要依赖外部供应商。



### 3. 数据源过多，造成数据治理陷入泥潭

数据中台对于多数据源的整合，并非合并原有数据库，而是从原有数据库中抽取需要的数据，通过ETL工具进行治理后，放在新的数据库中。当数据中台对接的上游数据源过多的时候，就出现了以下问题：

- 不同数据源的格式统一，例如有的数据通过“男，女”来标识消费者性别，而有的数据源通过“先生，女士”，当对接新的数据源后，会面临需要在原有定义中插入新值的过程
- 不同数据源的优先逻辑，例如数据中台对接了10个数据源，8个描述某消费者为“男性”，2个描述消费者为“女性”，那在数据中台中最终的描述应该是什么？这背后需要建立不同数据源优先级关系的复杂逻辑
- 当某个上游数据源发生了变化（例如版本迭代，造成数据格式的变化），数据中台的数据治理模块就需要重新配置，如果数据中台对接的数据源过多，这样的重新配置就会永无止境



因此，对接数据源过多，是数据中台项目时间远超预期的最大技术风险。

### 4. 缺乏足够应用场景

在数据中台规划阶段，广告主往往会寄予厚望，遐想足够多的应用场景。当数据中台上线后，广告主会发现原来设想的场景效果不明显，运营成本过高，数据安全，内部利益冲突等多方面问题，最终留存落地的场景远少于设想。

为了避免缺乏应用场景的现象发生，在数据中台规划阶段，广告主需要拿出最精力的精力，去研究同行业的成功案例，对于场景实现的成本和收益有足够精算，特别是重量级的Data Lake项目，笔者建议在规划阶段，广告主需要拿出固定成本做好数据中台的前期咨询项目，还有最重要一点：得到内部的认同感，所以往往行业里的重型Data Lake项目往往是企业一把手来主持。



# 营销数据中台的未来发展趋势

如果数十年后，史学家编写中国数字营销的发展史，2018年可以作为数字营销“上半场”和“下半场”的分水岭。

在过去十年的“上半场”，国内的数字营销生态圈经历了每年30%的市场增速，至今已达近4000亿的规模。由于数字营销是一个庞大的生态体系，一个有经验的操盘手，一些局部领域的差异化资源，一些信息不对称的打法，再加上一些运气就可以帮助广告主实现以小（资源）博大（成绩）的效果，如果用一词形容，就是“红利”。

而在2018年，整个生态圈的一些变化从量变到了质变规模：

## 1. 头部媒体的进化

作为给消费者传递信息，数字营销最后一公里的媒体，现在优质流量都掌握在BAT，头条+抖音，门户网站，垂直媒体（例如汽车之家）等头部媒体手中，在过去1-2年中，这些头部媒体逐渐完成自身能力的进化，除了扮演好“触点”的角色，还延伸到内容，数据，技术等领域，构建了自己的数字营销能力闭环，期望直接和广告主交易，提供一站式服务。媒体的进化打破了生态圈平衡，挤压生态圈的中间角色（广告公司，媒体采购公司等），实现了“去中心化”。最直观的结果是头部媒体的溢价能力越来越强，广告越来越贵。

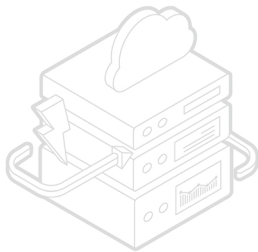
## 2. 线下流量的重新崛起

随着线上流量的越来越贵，过去10年通行的线上流量+电商的模式逐渐走入困局，甚至在很多行业，流量成本已经占了产品售价的15-20%。重新发掘线下流量价值，和现有线下销售渠道的整合，成为传统大型广告主在2018年数字化转型的大课题，最具代表性的是“新零售”概念的火热。

## 3. 数据合规法令的冲击

数据在数字营销中的核心作用无需再过多详述，过去对于数据的收集，使用还处于“法无禁止即可行”的粗放阶段，“数据合规”和“个人隐私”并没有受到足够的重视。在2018年，GDPR，网安法，GB/T 35273等一系列法律法规的公布和严苛执行，使得广告主收集和使用数据的方式发生根本变化，外部数据供应商为了应对风险，需要升级自身的IT系统，甚至停止一些数据服务形式，最直接的结果是广告主使用外部数据的条件越来越苛刻，成本越来越高。

要适应以上变化，大型广告主到了时间点去构建自身的数字营销运营闭环，来更好地管理资源，建设区别于竞争对手的差异化竞争力。



如果把数字营销“上半场”比作不停寻找“红利”，以小博大的“特种兵作战”，那“下半场”就是比拼资源体量和整合度的“集团军作战”，一城一池的得失不再影响大局，整体作战的效率和精细化程度，是长期胜负的关键因素。用一个词来形容数字营销的“下半场”，就是“运营”。

作为广告主在数字营销能力构建中投入最大，也是最复杂的数据中台，可以被称为广告主数字营销能力的“镇国利器”，决定了广告主的整体运营效率，在未来3年，我们可以看到数据中台的几个发展趋势：

## 1. 第一方数据运营闭环的构建：

数据合规的制约因素，造成了外部数据使用成本的不断增加，广告主需要构建自有的第一方数据运营闭环，在中长期降低数据使用成本，并且通过自有数据资源，获得和掌握了触点和内容资源的媒体对话时更大的话语权。

## 2. 数据驱动的新O2O模式：

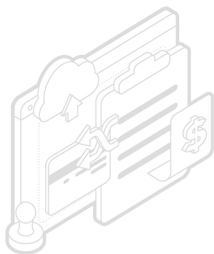
今天营销数据中台中收集和使用的更多是线上数据（广告监测，网站分析，微信SCRM等），在5G，线下探针，人脸识别，声纹识别等线下数据收集技术不断成熟，使用成本不断降低的情况下，广告主需要探索数据中台线上和线下数据的融合，打通线上线下的营销和销售渠道，尝试创建新O2O模式，这也是零售，汽车，服务业等拥有大量线下渠道的传统企业数字化转型的重要部分。

## 3. 第一方数据分析能力的构建，和对数据价值的深度挖掘：

对比数据运营能力的成熟，广告主对于数据价值的挖掘能力远远不够，今天大部分对于数据分析的方法论和实践还是处于传统的CRM分析+广告投放分析。在数据中台拥有更多数据，大量有行业经验的数据分析师成熟后，广告主可以构建自身的第一方数据分析团队，更好地挖掘数据价值，产生更多的方法论和实践。

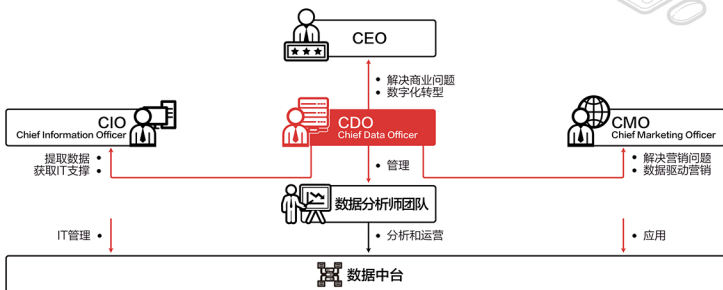
## 4. 营销技术的真正兴起

营销技术（Martech）这个词出现在国内上已经超过3年时间，虽然一直保持很高的热度，但是对比国外，国内的广告主并没有在营销技术领域投入核心资源，原因是现在大部分广告主并没有数据基础，来支撑各种营销技术的高阶应用。在广告主拥有数据中台后，广告投放管理系统（Ad Serving），营销自动化工具（MAT），A/B测试（A/B testing），动态主页（Dynamic Website）等国外已经成熟的技术才能真正在国内落地和兴起，帮助广告主实现实时的“千人千面”甚至“一人千面”的个性化营销场景。



## 5. CDO的出现和崛起

国内数据中台项目的主导权，在不同广告主会归属IT部门，营销部门，企业的管理层（CEO）等不同部门，并没有明确的归属。在这些角色中，IT部门精通技术却不在业务一线，营销部门懂得需求却对于技术一知半解，管理层虽然有强大资源调动能力却无法把握细节。国外也遇到同样问题，最终的结果是广告主内部出现了一个新角色：CDO（首席数据官），定位是比IT部门更懂业务，比营销部门更懂技术，掌握数据中台的分析和运营资源来应对企业管理者的诉求。CDO这个专业角色的出现和崛起，将配合CIO和CMO，把数据中台的价值最大化。



## 6. 更加严苛的IT数据合规和安全要求

随着数字技术的发展，利用数据和技术“作恶”的门槛越来越低，数据泄露造成各种民事和刑事案件越来越多，这是国内外不断更新数据合规立法，对于数据拥有方越来越严苛的背景。在享受数据中台带来的优势的同时，在IT层面遵守数据合规以及数据安全是大型广告主的一道紧箍咒，按照GDPR的要求，严重的合规问题最高处罚2000万欧元，或者前一财年全球收入的4%。在数据中台建设的时候，甚至需要考虑预留数据合规以及数据安全的成本，特别是可能涉及到海外数据，正在出海的国内大型广告主。

## MIAOZHEN SYSTEMS

### 关于秒针营销科学院 (MAMS)

秒针营销科学院 (MAMS) 是秒针系统集结行业资源设立的研究型组织, 着眼携手众合作伙伴的数据技术和智慧, 推动数字营销更科学, 更有序发展。

以下成员直接参与了这版“中国营销数据中台的解读”的制作:



赵洁

秒针系统 总裁



于勇毅

秒针系统 营销科学家

秒针系统 UED: 贾宇, 梁颖

秒针系统市场部: 王龙本, 项久大

如果您有任何问题, 希望和笔者沟通, 请发送邮件到

[yuyongyi@miaozhen.com](mailto:yuyongyi@miaozhen.com)

如果您希望了解秒针的数据中台产品和案例, 请发送邮件到

[market@miaozhen.com](mailto:market@miaozhen.com)

Miaozhen®  
Systems



MIAOZHEN.COM

■ 数据创造信任