



## 建设数据中台，赋能改革创新

成本效应成为了企业及政府机构数据化的绊脚石，管理层必须理解，从信息化到数据化再到应用，必须有强大的技术支持、灵活的政策保障以及开放的生态支撑，方可马到功成

文 / 车品觉

**信**息及数据的收集及处理自古有之，几千年来变化不大，直到个人电脑的普及才开始出现了极其微妙的改变。首先是带动着数据化的领先企业包括 Google、Amazon、Facebook 及中国的 BAT，然后是移动应用和人工智能所引起的智能热潮。再加上物联网的加入，让我们在不知不觉之间每天都在生产着数据，甚至消费着数据。可以说，几乎全民都参与在大数据的生命周期之中。大数据的组成和几年前很不一样，因为需求的细分及终端的发达变得越来越零散。同时数据安全的规管也改变了大家分享数据的形式及难度，汇聚数据的成本正在不断提高。但是我们都清楚大数据的核心价值来自分享，经过数据的分享增加了预判及还原现实的可能性。所

以成本效应成为了企业及政府机构数据化的绊脚石。

如今企业都意识到数据是重要的资产，也意识到有效数据治理是数据资产化的前提，近年本人在国内做市政府和一些互联网企业的咨询专家及顾问，都涉及到如何建立横向的数据治理系统。这个经验最早来自阿里巴巴，当时在企业飞速的发展过程中，我们发现数据使用面临着如下挑战：

——各部门低水平重复开发数据集，浪费大量的存储和计算资源；

——数据资源缺乏沉淀机制，导致计算能力的提升和进化非常低效；

——数据割据，算法分离，带来混乱和质量的不确定性；

——业务变更时，数据及数据产品反应不及时；

——组织架构制约了数据的共建和共享，缺乏标

准及激励机制。

经过内部总结发现，数据的“汇管用”过程中伴随着三个现象特点：数字业务变化速度非常快、数据处理技术及方法都很类似、数据及算法中间层能产生巨大效能。因此，做好数据治理工作成为当时笔者在阿里巴巴的主要任务，也随之诞生了阿里数据中台。

无独有偶，企业内的一些数据治理问题，在各地的政府机构内部也在重复发生。在缺乏顶层设计之下，数字化步伐都在追随各个职能部门的发展，数据体系也是基于业务单元垂直积累，从而形成了烟囱式体系。垂直式数据体系的优点是紧贴场景反应敏捷，缺点是数据分散、不标准，难以共用关联成为合力，大数据价值优势被削弱。此外，烟囱式数据体系还会造成混乱的数据调用和拷贝，以及系统功能建设和维护带来的重复投资，不仅造成人力、财力、资源的浪费，更重要的是时间浪费以及数据质量的参差不齐！在目前高速发展的互联网市场大环境下，商机是稍纵即逝的。

在认同大数据是未来创新核心的前提下，需要把数据战略的先进性、前瞻性放到优先考虑位置。否则大数据的能力会随着粗放式运营而变得停滞不前，沦为有名无实；数据愈乱，建立大数据的能力门槛愈高，从信息化到数据化的时间节点都会影响治理难度。管理层必须理解，从信息化到数据化再到应用，必须有强大的技术支持、灵活的政策保障以及开放的生态支撑，方可马到功成。

## 数据共享是数据生态的核心

为了促进企业内部对于数据挖掘、更新、使用的效率，本人在阿里内部尝试建立了数据公共层，首先是对于线上及离线的数据按交叉使用量、紧缺风险等进行盘点，基于现状及未来需要作中长期战略预估。这里汇聚了阿里内部共用得最频



繁或者最关键的数据，这些资源有如生产中所需要的必备部件，可以加快生产速度及降低重复性。公共层作为数据中台的核心部分，积累了最关键的数据资源，同时也是最具备品质保障的主数据。

简单地看，数据中台有点像一条生产流水线，从原始数据收集，到提炼成稳定的生产流程。在这个制作过程中，需要有一套生产管理流程体系，用以保证数据品质、时效性、一致性等关键点。但与生产流水线的差别在于，数据中台不仅需要关注数据生产过程中的效率问题，中台实际上还具备以下能力：1. 如何收集数据被消费之后的反馈闭环；2. 解决多源异构的数据组合的效率；3. 具备业务发生变化时的快速自适应力；4. 保障数据服务的稳定性。

数据中台围绕数据生命周期的各个阶段（产生、存储、增强、使用、传输、共用共创、更新、销毁等）而建立，服务的对象可以是IT研发者、数据科学专家、产品经理、分析师、决策管理者等。使用者会因需要而加工数据，情况有点像石油提炼一样。而数据生产过程中还有一种极其重要的数据，被称为元数据，又叫数据中的数据。对元数据管理得当，就可以让数据在生产过程变得更精准、稳定及可被追溯。元数据管理须记录生产过程中各项数据因素，包括生命周期、调度情况、品质保障、安全监控、数据字典、数据血缘关系等。元数据是数据中台的精髓，有利于数据在生命周期中的监督、成本管理或分摊、追踪数据价值。因此，一般数据中台的价值体现可以根据数据开发能否化繁为简作为考核，进一步理解就是开发成本的节省（第五种能力）。

## 数据服务赋能快速创新

一切数据都是因业务目标驱动而形成，产生于业务且又服务于业务。通过松耦合的数据服务带来业务



的复用，例如淘宝和天猫有着各自的买家评价服务，但在防止刷屏的时候会使用相同的数据模型鉴别虚假评价。所以即便业务场景不一样，但很多的基础数据模型及算法可以被重复使用服务。

经过清晰的沉淀，算法可以通过重新编排、组合，成为服务接口响应业务的基本需求。由于具备快速编排、组合数据服务的能力，企业可以以较小的成本投入来构建出一个创新的前端业务。这是传统模式构建的系统中前所未有的，容许快速试错，更适合今天具有互联网精神的轻公司模式。

### 需要相应的组织架构与激励机制


任何完善的体系建设依靠的不仅是技术工具，缺乏完善的组织结构及激励机制便不可能令中台顺畅运行。多年来的经验证明了技术架构和治理组织的建立同样重要，而近年很多企业及政府也纷纷设立独立数据治理委员会。前面提到数据中台核心理念是“以通促用，以用带存”，这里的“通”不仅是数据的联通，

也关乎人为组织结构的联通，而且是横（功能部门之间）、纵（数据生命周期）都要通。

与此同时，数据中台管理需要制定并形成有效的规范，由治理小组从实例中由下而上地提炼出大纲，并由固定团队负责推进、制定工作机制（互惠互利及激励方式）、优先资源配置等。

### 构建符合互联网大数据时代的大数据中台

数据技术“中台”之所以成为攻坚大数据能力的重要途径，一来因为数据中台确实解决了数据竖井（之前各自为政）问题，其次是更有利于推动轻盈的前台业务创新，同时能把应用中的数据回流，形成更丰富的中台资源。数据中台作为推动数据化营运的利器，同时也能成为营运数据的中心（两者结合为闭环）。多年的经验笔者可以大胆地说，数据中台的建立刻不容缓，因为在大数据时代，业务与数据之间是强联系，但数据的内容及结构更新速度非常快；数据算法上的技术很类同，但各师各法；数据质量人人都说重要，但应该由谁负责任？

数据及其服务能力的汇聚与集中管控，很大程度会促进企业一体化运维的能力，归纳起来讲，互联网大数据时代的中台特点是：一方面具备海量多源异构数据的整合能力；另一方面促进创新且变化多端的业务前端服务能力。大数据中台的建设及行业普及到如今还是摸着石头过河的状态，任何中台都是在不断互动回馈的过程中成长出来的，而非统一搭建而成功。但是可以肯定，这是个“一把手”工程，必须秉持打破传统管理的决心，做好长期斗争的准备。

（作者为香港特别行政区创新及再工业化委员会委员、红杉中国专家合伙人、原阿里巴巴集团副总裁及数据委员会会长）

