

基于数据资产全生命周期 估值与实践报告



序言

随着数字经济不断发展，数据作为一种“新型石油资源”价值逐渐凸显。2019年10月，党的十九届四中全会首次将数据确立为生产要素；2020年4月，《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》正式公布，明确完善要素市场化配置的具体举措，数据作为一种新型生产要素被列入其中，并指出“加快培育数据要素市场，建立健全数据产权交易和行业自律机制”。2022年12月，中共中央、国务院对外公开发布《数据二十条》，引起了社会各界的高度关注，着重强调了数据作为新型生产要素，具有无形性、非消耗性等特点，可以接近零成本无限复制，对传统产权、流通、分配、治理等制度提出新挑战，亟需构建与数字生产力发展相适应的生产关系，不断解放和发展数字生产力。2023年8月，财政部制定印发了《暂行规定》，明确数据资产明年起入表，是我国跨入数字化时代迈出实质性的一步。这些都展现了国家坚持促进数据合规高效流通使用、赋能实体经济这一主线，以充分实现数据资产价值的坚定决心。

在数据要素市场化持续探索的过程中，如何有效衡量数据价值，构建合理、健全、有效的数据资产评估方式，对数据交易流通具有重要意义，因此，为了促进数据市场交易、专家学者和业界纷纷都在探索数据资产评估模式，提出数据资产评估的方法。但是，由于数据资产与资本、土地、技术等其他传统生产要素相比，具有虚拟性、正外部性、规模报酬递增等独特性，使得数据价值评估成为难点，仍处于探索阶段。产业实践侧，互联网公司作为数字经济时代的代表性企业，其自身不仅拥有海量数据，又具备良好的数字化技术基础，是较早开展数据资产价值评估实践的行业之一。

本白皮书聚焦数据资产价值评估领域，从数据资产的概念和意义出发，分析了目前数据资产价值评估主流方法，以强化数据资产管理能力、推动数据资产广泛应用、加速数据资产高效流通为主要目标，提出了以成本法为主线、以数据管理要素和应用回报率作为调节因子的估值方法，并引入了包括血缘分析、层次分析法和超额收益法在内的方法，加强估值结果的准确度。此外，白皮书通过内部治理运营、外部数据流通两大类场景对估值方法进行阐述，使得估值方法更具有参考性和执行性。最后，白皮书对数据资产估值进行了总结和展望，希望以我们在数据资产价值评估领域研究工作的“一小步”推动我国数据要素市场发展的“一大步”。

目录

一、数据资产价值评估的意义和研究现状	4
1. 数据资产的概念和意义	4
2. 数据资产估值发展现状	4
二、数据资产价值评估主流方法	6
1. 数据资产价值评估的方法	6
2. 数据资产价值评估的核心因子说明	7
1) 成本计算因子	
2) 价值调节因子	
3. 数据资产价值评估的技术实现探索	10
1) 血缘传播用于数据成本评估	
2) 层次分析用于数据管理要素评估	
3) 超额收益法用于数据收益率评估	
三、数据资产价值评估应用场景与案例介绍	19
1. 数据资产价值评估应用场景	19
1) 内部治理运营	
2) 外部数据流通	
四、总结与趋势	26
1. 数据资产价值评估总结	26
2. 数据资产价值评估未来趋势	27

一、数据资产价值评估的意义和研究现状

1. 数据资产的概念和意义

数据要素市场化已成为建设数字中国不可或缺的一部分，而数据资产化是数据要素化的基础，在企业数字化转型和全面数据管理能力的基础上，进一步提升数据应用和数据流通，以最大化数据资产的价值。

数据资产是组织(政府机构、企事业单位等)与个人合法拥有或控制的数据，以电子或其他方式记录、存储，例如文本、图像、语音、视频、网页、数据库等结构化或非结构化数据。数据资产的构建和应用，对于企业发展，对于社会进步，都具有重大价值和意义¹。

数据资产估值在厘清数据资产权属、构建数据资产的定价和交易体系、促进数据要素市场发展具有至关重要的作用。**对企业自身发展而言**，数据资产估值结果将更加直观、丰富的展现数据给企业带来的业务和财务价值，进而推动企业构建数据文化，加大数据资产管理投入，调动各方参与数据资产管理工作的积极性，提高企业使用数据辅助决策的效率。**对社会经济发展而言**，数据资产的流通将极大的促进全社会的数据丰富度，创新更多数据应用场景，在此基础上， 社会各界可加快对数据资产价值的共识，加快形成不同数据资产的不同定价和多种交易模式，引导数据流通和交易健康有序发展。

2. 数据资产估值发展现状

国内各界自 2019 年起陆续关注数据资产价值评估领域，行业组织、科研院所和产业界相继发布了相关政策文件、研究成果或实践案例。中国资产评估协会于 2020 年 1 月发布了《资产评估专家指引第 9 号-数据资产评估》，在传统资产评估三种方法（成本法、收益法和市场法）的基础上进行了优化，提出了数据资产评估方法。德勤-阿里在 2019 年联合发布了《数据资产化之路-数据资产的估值与行业实践》，其基本思路与《资产评估专家指引第 9 号-数据资产评估》一致，同时深入分析了各个方法的优劣势。中国信息通信研究院与浦发银行在 2021 年联合发布了《商业银行数据资产管理体系建设

¹ 参考大数据技术标准推进委员会《数据资产管理实践白皮书（6.0 版）》

实践报告》，结合浦发银行数据资产管理实践，提出了内在价值、成本价值、业务价值、经济价值和市场价值的五大估值维度。光大银行于 2022 年发布的《商业银行数据资产会计核算研究报告》中从数据资产入表的视角出发提出了货币度量数据资产价值的方法。总结而言，国内在理论层面当前数据资产价值评估方法主要划分为成本法、收益法和市场法三种基本方法及其衍生方法。其中，**成本法**为数据资产的价值由生产该无形资产的必要劳动时间所决定，是从数据资产的重置角度考虑的一种估值方法，即新建该项数据资产所需花费的成本；**收益法**为数据资产的价值由其投入使用后的预期收益能力体现，是基于目标数据资产预期应用场景，对未来产生的经济收益进行求取现值的一种估值方法；**市场法**是基于相同或相似数据资产的市场可比交易案例的一种估值方法。

此外，数据资产估值也引起了国外学者的广泛关注，但是更多以数据资产市场化定价为出发点倒推数据资产估值思路的。例如，斯坦福大学的研究者提出了基于市场价值的数字估值方法，主要基于数据资产带来的消费者剩余价值，并研究生产者的市场表现等因素；英国的金融科技公司和数据服务提供商根据自身在数据交易市场中的实践，开发了基于数据流量、数据质量和数据影响力等指标的数据资产估值模型。

我们认为数据资产估值当前在理论研究层面已有初步成果，但仍然面临三大问题，一是数据资产估值目的、场景和相应方法尚未统一，二是估值过程主观性较大、缺少量化指标、技术应用程度低，使得数据资产估值结果不准确，三是缺少各行业数据资产估值实践案例，导致理论方法对于企业实际开展数据资产估值的指导性有限。

二、数据资产价值评估主流方法

1. 数据资产价值评估的方法

如前文所述，数据资产价值的评估方法主要有成本法、收益法和市场法三种基本方法及其衍生方法。阿里云及瓴羊在多年阿里内部实践及服务企业数字化转型的过程中沉淀了丰富的行业经验和数据技术，通过与中国信通院共同研究讨论，综合考虑主流的成本法，同时融合数据资产管理（包括数据质量和使用因素等）相关因素，提出了基于数据全生命周期的资产价值评估方法，从数据在不同阶段的影响因素出发，及完善数据资产价值评估框架。

数据全生命周期可以依据从生产到消费的链路分为“采”、“建”、“管”、“用”四个阶段，“采”代表汇聚异构数据的全过程，“建”代表数据治理的完备和连贯，“管”指全量数据资产的治理与管理，“用”则强调数据资产的应用与回报。基于前述的成本和使用维度，结合数据全生命周期的四阶段的相关因素，我们将单张表作为评估对象，评估模型基础公式(1)如下：

$$V_{t(k)} = \left(\sum_{n=1}^N D_{cost(n)} \right) * f\left(\sum_{j=1}^J \alpha_j M_j \right) * (1 + \bar{R}_K) \quad (1)$$

其中， $V_{t(k)}$ 代表在业务域 K 的单资产表的价值， $D_{cost(n)}$ 是通过血缘成本分摊算法得到的基础成本， α_j 是数据管理要素的系数， M_j 是数据管理要素的分数，R 是归属业务域 k 的市场投资回报率，和行业属性紧密相关。该公式来自于资产价值=成本*（1+收益率）因此对于所有表的数据管理要素的期望值 $E(f(\sum_{j=1}^J \alpha_j M_j))$ 应该为 1，即我们会采用对应的函数 $f(x)$ 变化该 $\sum_{j=1}^J \alpha_j M_j$ 项，使得整体数据管理要素满足对应分布的期望值为 1。该资产价值评估基础公式也展示出数据资产价值由数据成本计算因子和价值调节因子共同作用确定，其中数据成本会考虑为成本计算因子，数据管理要素和数据市场投资回报率为价值调节因子。

2. 数据资产价值评估的核心因子说明

本白皮书中定义的数据资产价值由数据成本计算因子和价值调节因子共同作用确定。其中，**成本计算因子**由本次数据资产评估的核心技术实现，即采用血缘分析方法完成对数据整体建设成本到单一数据表建设成本的分摊计算。**价值调节因子**是包含企业数据管理水平及市场交易等其他要素，以对成本计算结果进行优化调整。

1) 成本计算因子

a) 数据资产总投入成本

从数据资产全生命周期出发，数据资产总投入成本包括硬件建设成本、软件建设成本、运维成本、资源成本、数据采购成本、数据咨询成本、人力支出、场地费用等方面，覆盖了数据资产采集、购买、存储、计算、管理、应用等环节。各类成本具体成本子项和说明如表所示（表 1）。

表 1：数据资产总投入成本说明

成本分类	成本子项	说明
硬件建设成本	设备采购费用	数据采集，数据存储，数据计算相关设备（服务器，交换机，物理机）的采购费用
	设备折旧费用	数据采集，数据存储，数据计算相关设备（服务器，交换机，物理机）的折旧费用
	设备租赁费用	数据采集，数据存储，数据计算相关设备（服务器，交换机，物理机）的租赁费用
	云计算资源费用	数据采集，数据存储，数据计算相关的云计算资源费用
软件建设成本	软件许可费用	数据相关软件许可证的采购费，例如数据库套件，数据分析软件等
运维成本	设备维修费用	数据设备日常的维修和保养费用
资源成本	资源消耗费用	数据的影响到的水电的资源消耗
数据采购成本	外部数据采购费用	对非公司自有数据（政府数据，互联网数据）的单位期间采购费用

数据咨询成本	第三方外包咨询服务费用	数字化规划设计，给可行性研究报告，数据治理，数据应用等咨询类外包支出
人力支出	人员薪资和福利	数据采集，存储，开发管理人员的相关人力成本
场地费用	场地费用	数据资产设计场地租赁和购买费用

b) 基于血缘成本分摊成本

在实际数据链路中，从源头的数据采集或引入，到最终的数据消费使用，会经过一系列的数据处理与加工。因此，对于某个数据对象，其成本不能仅考虑当前对象生产或直接上游引用来源的成本，还应考虑层层上游传递的成本。因此，我们用基于数据仓库理论的血缘成本分摊成本，对应着公式中的 $D_{cost(n)}$ 。

在数据的“建”阶段，为了发挥数据价值而进行的数据处理和加工过程，可以基于数据血缘客观反映，呈现数据的来源以及数据随时间变化的位置。它有助于提供对分析管道的可视性，并将错误跟踪简化回源。血缘继承具有时效性和自更新特性。如果源端数据的加工逻辑和血缘关系发生变化，基于血缘继承的方法能及时自动更新价格试算，无需人为重新调优。

本白皮书采用血缘成本分摊和破圈算法来解决数据血缘对数据成本的影响问题。输入要素为上游血缘和下游血缘，在下文技术实现部分详细阐述。

2) 价值调节因子

数据资产最终价值除去已经投入产生的成本外，还与企业自身数据管理水平及市场交易现状相关。其中，企业数据管理水平主要影响数据资产价值的真实有效性，成熟的数据资产管理体系可以确保数据可靠、可信及可用，降低因质量、安全等问题带来的负面影响；而市场交易因素则反应了数据资产在市场中的稀缺水平，从外部供需交易视角将数据资产价值调整至合适范围。

a) 数据管理要素

在数据管理要素的选择和定义上, 本白皮书主要参考了《数据管理能力成熟度评估模型》(GB/T 36073-2018) 这一国家标准, 同时考虑指标获取难易程度, 确定了以下数据管理要素评价体系框架。

表 2: 数据管理要素评价体系

数据管理要素评价体系							
基本属性	质量属性					使用属性	
数据规模	准确性	一致性	完整性	规范性	时效性	可访问度	使用热度
数据量	内容准确率	元素赋值一致率	元素填充率	值域合规率	周期及时性	平台访问次数	输出次数
增长率	精度准确率	跨表取值一致率	记录填充率	格式合规率	实时及时性	平台访问用户数	输出系统数
...

企业在具体选择数据管理要素指标集合时, 主要受两大因素影响: 一方面应结合企业当前的数据基础设施的成熟水平, 确保所选取的要素指标均可被计算; 另一方面, 应结合企业当前的经营状况和行业属性, 选取与数据资产价值高度关联的评价指标。

本白皮书将通过层次分析法分别对每一项数据管理要素的权重和评分进行评价打分, 其中 α_j 代表每一类数据管理要素的权重, M_j 代表数据管理要素的分数。通过整合所有被选取的数据管理要素评价指标 (即 $\sum_{j=1}^J \alpha_j M_j$), 可最终获得企业整体的数据管理评价结果, 并参与数据资产价值评估的后续计算过程。

b) 数据资产应用回报率

数据资产的应用回报率 (即 \bar{R}_K) 与其自身的应用场景和价值高度相关, 即一项数据资产如果可以被广泛应用在多个不同的应用场景中, 则可以为企业带来更高的商业价值和回报水平。然而, 除自身建设水平外, 数据资产应用价值高度依赖于数据市场当前的供

需水平、历史交易等因素，但由于当前尚缺乏成熟的数据相关交易市场，导致缺乏足够数据以支撑企业获取数据回报率结果。

因此，为了尽可能准确地衡量数据在使用过程中产生的价值，在本白皮书充分借鉴金融学模型，并采取学界讨论较多的超额收益法来对整体交易方案进行补充，具体内容详见 2.3.3 章节。此外，考虑到数据应用回报率与数据交易市场、数据财会政策发展相关，本白皮书仅在此提供一些参考思路。

3. 数据资产价值评估的技术实现探索

充分考虑数据资产在发挥价值的路径里面的所有对价值因素有意义的因素，对数据资产进行剖析，分类，以及针对不同类型的因素采用不同的技术化手段进行计算，得到了数据资产价值评估的完整且符合数据资产真实属性的评估体系，整体从数据成本识别和录入开始，以血缘图算法为主导的成本分配“计算器”为基础，数据资产要素评价体系和数据应用市场价值分析四大主要模块，对数据资产价值如何得出开展了突破性的落地研究（图 1），我们对该资产评估模型的算法进一步的阐述：

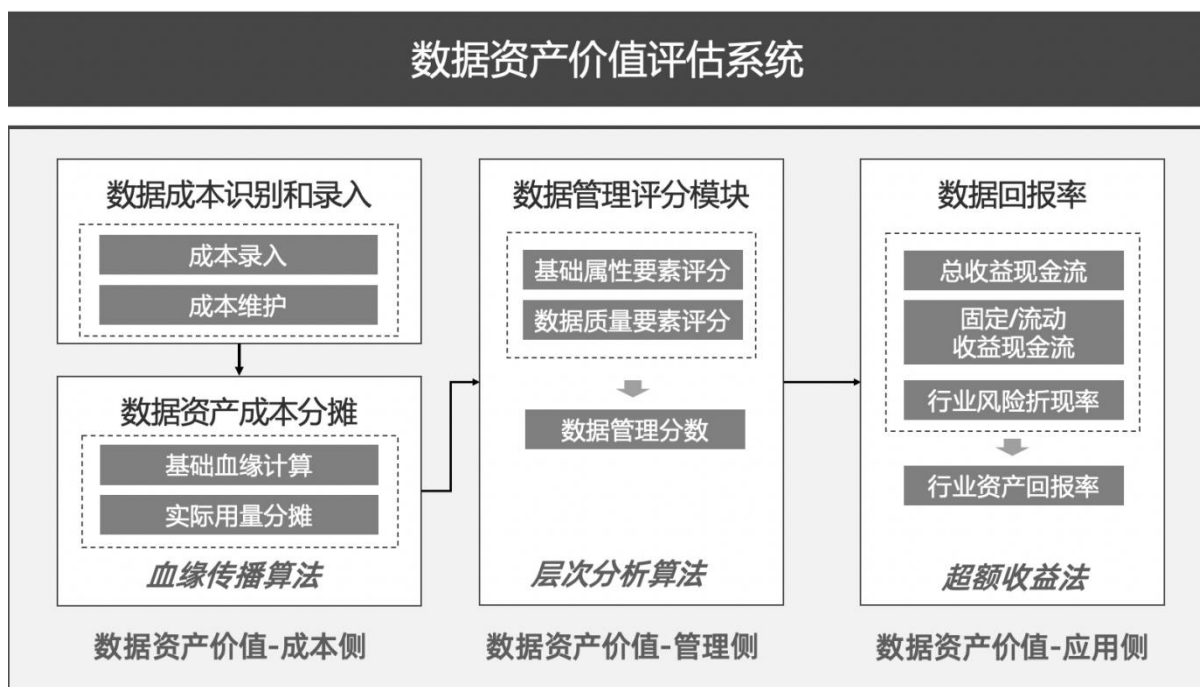


图 1：数据资产价值评估系统

1) 血缘传播用于数据成本评估：

血缘传播算法主要分为以下的两个模块进行考虑：

第一，原始成本核算&基于血缘依赖的成本传播在原始成本核算节点，需要计算所有数据生产过程中产生的建设成本，机器成本，软件成本，和知识发现成本和运维成本，为单个数据生产的全部成本进行核算。

对于数据 d 原始成本 $O_d = \sum_{i=0}^n c_i^d$ ，其中 c_i^d 为数据 d 生产过程中的每一项单项成本计算出每份数据自身的成本。

例如，对于数据 A，相关主要原始成本如下(表 3)：

表 3：数据 A 的相关原始成本

成本项	硬件建设成本	软件建设成本	资源成本	数据咨询成本	运维成本
数据 A	20 万	10 万	20 万	5 万	5 万

则数据自身成本 $O_A = \text{硬件建设成本} + \text{软件建设成本} + \text{资源成本} + \text{数据咨询成本} + \text{运维成本} = 60 \text{ 万}$

注：由于数据的各项成本是企业从总投资成本中识别、抽离出的数据建设成本，这一过程也是成本法数据资产评估中较为复杂的部分，这部分计算依赖于较为完整的数据平台能力建设。在实操中，可以通过详细统计数据的计算消耗、数据量大小来计算相关的资源、硬件和软件成本，均摊相关开发人员人力成本及上下游依赖的系统成本来计算数据的咨询和运维成本。

得到数据自身生产成本后，由于数据生产是一个有向无环图，数据生产是一个由上游数据逐步经过不同处理、合并等操作逐步产生下游数据的过程，最终被使用的表往往依赖了内部的加工节点被称作上游依赖节点，直接服务于业务的节点被称作叶子节点。成本分摊旨在将上游依赖节点的计算和存储费用分摊至叶子节点，从而将费用归属给各个产品。成本分摊的意义不仅仅在于它可以精确的计算业务的数据消耗，它还能反向的推动数据治理，数据优化，公共逻辑沉淀的进行。公式可以写作：对于数据 d 最终成本

$C_d = O_d + \sum_{i=0}^n (O_i^{d-1} * R_i^{d-1})$, 其中 O_i^{d-1} 为数据 d 上游的第 i 项数据, R_i^{d-1} 为 O_i^{d-1} 对数据 d 分摊的成本比例。

例如下图, 数据 A-H, 各自原始成本均为 60 (基于上述原始成本计算结果), 则数据总生产成本为 480, 同时数据 A、B、C 还支持了数据 D 和 G 的生产, 数据 D 支持了数据 E 和 F 的生产, 数据 E 和 F 支持了数据 H 的生产, 所有数据中只有数据 G 和 H 对外提供服务, 业务场景为 M 与 N。依赖图如下(图 2):

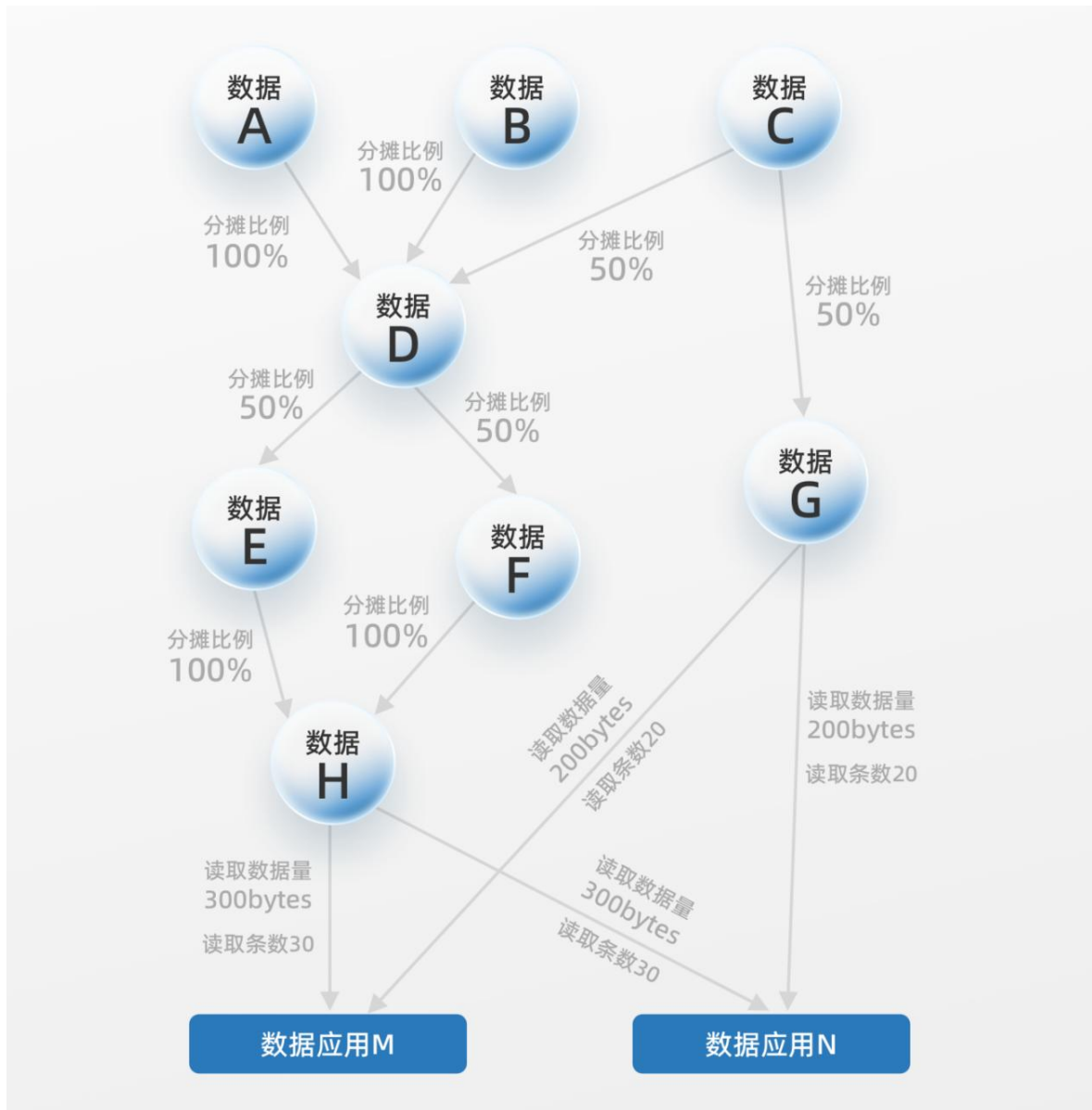


图 2: 数据依赖图

则数据成本间相互分摊的比例如下（表 4）：

表 4: 成本的分摊比例

分摊比例	数据 A	数据 B	数据 C	数据 D	数据 E	数据 F
数据 D	100%	100%	50%			
数据 E				50%		
数据 F				50%		
数据 G			50%			
数据 H					100%	100%

经过成本分摊，各个数据分摊后成本如下：

数据 A\B\C 由于没有上游分摊成本，则 $C_A = C_B = C_C = O_A = O_B = O_C = 60$

最终数据 D 传播后成本为 $C_D = O_A * 1.0 + C_B * 1.0 + C_C * 0.5 + O_D = 210$

最终数据 E 传播后成本为 $C_E = C_D * 0.5 + O_E = 165$

最终数据 F 传播后成本为 $C_F = C_D * 0.5 + O_F = 165$

最终数据 G 传播后成本为 $C_G = C_D * 0.5 + O_G = 90$

最终数据 H 传播后成本为 $C_H = C_E * 1.0 + C_F * 1.0 + O_H = 390$

通过以上分摊，则最终分摊后对外提供服务的表 G 和 H 的传播后成本 $C_G + C_H = \sum_{i=A}^H O_i = 480$ 可以将整体数据的生产成本完全分摊至对外使用的数据，从而保证数据成本价值估算与实际发生成本一致。

第二，基于实际用量的成本分摊：

在得到数据生产的基础成本之后，由于一份数据往往被多个下游使用，所以数据的生产成本需要被下游所有的使用方共同承担，而在实际使用场景中，不同业务对于数据的依赖程度往往是不同的。所以在进行成本分摊时不能采用简单的平均分摊，而需要结合下游使用实际用量为不同的业务方按照实际使用情况进行分摊金额的计算，在实际操作中，为了最大程度兼顾公平性。在实际操作时，可以基于实际业务场景基于下游对于数据使用总条数/使用总数据量/使用分区数等逻辑作为计算单元进行比例分摊，即使用数据量越大的业务方需要承担更多的数据生产成本，同时保障所有业务方共同承担数据全部的生产成本。

例如，对于数据 G 和 H 的使用场景 M 与 N，基于不同核算逻辑的分摊比例如下：

表 5：数据 M 和 N 的核算逻辑分摊

分摊比例	读取数据量	读取数据条数	访问量核算比例	读取条数核算比例
场景 M	100bytes	20	25%	40%
场景 N	300bytes	30	75%	60%

则对于场景 M，选择读取条数比例核算逻辑，则分摊成本为 $(C_H + C_H) * 40\% = 192$ 为场景 M 需要分摊的整体数据的成本，同样的，场景 N 的分摊成本为 $(C_H + C_H) * 60\% = 288$ 。通过以上方式，数据产出链路的总成本就可以以成本完全分摊的方式被实际业务承担。

基于该计算逻辑，可以得到保证在数据生产成本被全部分摊的前提下，兼顾下游使用方对于数据不同使用情况公平性的基础价值评估结果。在实际分摊计算中，应结合不同场景的不同特征，确定具体的分摊逻辑确定方法。例如数据记录中字段信息量较为一致，但每条数据长度差别较大的场景(如收货地址)，则更适合采用基于读取记录数作为分摊逻辑的方法；如果数据中字段信息较多，不同业务场景使用数据字段差别也较大，如包含很多字段的商家维度宽表，则更适合采用基于数据读取量计算分摊逻辑的方法。

2) 层次分析用于数据管理要素评估：

数据管理价值是基于成本价上其他一切非经济因素得到的数据价值评估，其主要的目标是为了评估数据管理活动对数据资产价值的主要影响。我们将详细阐述部分前文列举的

数据管理要素指标定义和 BI 口径(表 6), 具体的明细指标的定义还要基于相关的行业标准, 业务所使用的数据产品以及获取数据的难度程度决定, 大部分的数据都可以通过企业数仓的元数据和对应数据产品的使用底表获取。

表 6: 数据管理要素指标明细

	明细指标	明细指标定义
基础属性	数据量	数据集元素总数量, 指标口径为数据项数量*数据记录数量
	数据规模	数据集元素增加量与原数据集元素总数量之比, 指标口径为(本期数据集元素总数量-上期数据集元素总数量)/上期数据集元素总数量*100%

质量属性	数据质量规则触发率	取指定产品的数据, 近 30 天触发数据质量规则的比率, 指标口径为触发规则数量/总规则数量
	元素填充率	数据资产在基础属性、管理属性和业务属性上的信息完整率, 指标口径为非空项数量/表中指标数量

使用属性	平台访问次数	最近 30 天用户在全域资产平台上对此数据资产的访问、收藏、分享的次数之和
	平台访问用户数	最近 30 天在全域资产平台上对该数据资产有收藏、访问、分享的用户数去重之和

在评估时, 由于基于简单模型函数往往难以反应实际不同数据价值项在最终应用场景上产生的实际价值, 而大多数的非经济因素都是难以比较和量化的。

因此, 本白皮书采用层次分析法(Analytic Hierarchy Process, AHP) 来进行分析, 具体分为以下几个步骤:

- ①通过大量的访谈定义影响数据资产价值的管理要素指标;
- ②通过专家打分计算所选的非经济因素的权重;
- ③通过权重计算价值分数。

在明细的两两打分方面，按照下表(表 7)进行开展：每次打分都会构成对比较矩阵的元素 a_{ij} ，表示的是第 i 个指标对于第 j 个指标的比较结果。两两比较的指标形成矩阵，进一步进行矩阵一致性检验。如果矩阵中的每个元素均大于 0 且为正互反矩阵，若正互反矩阵满足 $a_{ij} = 1/a_{ji}$ ，则为一致性矩阵。最后，基于上文算法整合所有信息得到非经济因素的重要性系数，可以得到单资产综合评价的分数。

表 7: AHP 评价标度

标度	含义
1	表示两个指标相比，具有同样重要性
3	表示两个指标相比，一个指标比另外一个指标稍微重要
5	表示两个指标相比，一个指标比另外一个指标明显重要
7	表示两个指标相比，一个指标比另外一个指标强烈重要
9	表示两个指标相比，一个指标比另外一个指标极端重要
2, 4, 6, 8	上述两相邻判断的中间值

为了将不同的数据管理要素评价指标进行融合，需要进行归一化；这样可以避免中长尾案例的分布过多而导致的缺乏区分度的问题，转化为 0-100 的分数。同时为了保障整体收益回报率和成本的关系依旧严格符合，即资产价值=成本*（1+收益率），我们将得到的价值分基于某种统计学的 $f(x)$ 分布（通常选用 Log Normal 分布），保证 $E(f(\sum_{j=1}^J \alpha_j M_j))$ 为 1，转化为了整体期望值为 1 的分数值用于体现数据管理的价值，以及公平的衔接下文的整体资产回报率（图 3）。



图 3: AHP 打分流程

3) 超额收益法用于数据收益率评估：

企业的整体资产回报率和所处行业以及该行业的数字化建设能力相关，同时数据交易的二级市场回报率也与数据交易所及市场建设相关。但由于目前尚缺乏有效的数据交易市场，在实操过程中尚无法通过实际交易的价格来计算数据资产的回报率和数据资产的收益。

因此, 关于数据资产回报率, 本白皮书主要借鉴金融学中的投资回报率模型进行计算(2)。其中, 总收益由数据宽表的总价值(DV)表示, 参考学界讨论较多的超额收益法计算获得; 总成本由从血缘成本分摊中得到的单表资产原始成本表示。

$$\overline{R_k} = \frac{\text{数据宽表总价值 (DV)} - \text{初始数据表成本}}{\text{初始数据表的成本}} \quad (2)$$

参考超额收益法计算逻辑，DV（数据宽表总价值）由该数据宽、表的净资产回报值、年度数、风险折现率、数据资产调整系数共同确定（如下图公式所示）。其中，该净资产回报值为总资产回报值（E1）与以下取值的差值：固定资产对总资产的贡献值（E2）、流动资产对总资产的贡献值（E2）、除待评估数据表之外的其它数据对总资产的贡献值

(E4)。风险折现率 (R_{mk}) 通过资本资产定价模型 (Capital Asset Pricing Model, 简称为 CAPM) 确定。K 为数据资产调整系数, 取值在 0~1 之间。

(3)

该回报率需要结合公司整体经营情况和内部 BI 体系对不同业务域的回报率进行分别计算, 并在未来依据数据交易所的成交价进行调整。

三、数据资产价值评估应用场景与案例介绍

1. 数据资产价值评估应用场景

数据资产价值评估目前在企业内部数据治理运营和外部数据资产市场流通领域都发挥着重要作用。依据我们过去的部分探索和实践，主要从以下的两个内容进行阐述。

- 在企业内部，通过数据资产估值，可以达到企业内部数据治理和运营的有效目的。在一些集团型统建数据体系中，对数据成本有准确、统一可接受的价值后，进行合理的成本分摊，可极大的节省企业的数据成本，达到成本集约化和构成合理化。
- 在外部数据资产市场流通上，通过数据资产的全生命周期估值方法，推动数据资产收益向数据价值和使用价值的创造者合理倾斜，确保在数据内容采集、加工、流通、应用等不同环节相关主体的投入有相应回报。

1) 内部治理运营

数据治理及建设方案评估

对特定公司内部由于组织架构庞大，业务线众多，产生了多样化的海量数据，通过对于数据价值的评估，可以很好地衡量数据建设产生的实际业务影响，可以很好评估已建设的数据是否得到较好应用，从而推动数据建设方式向更加合理与高效的方向进行改进，主要影响动作包括：

- 重点数据影响分析、指标监控及异常保障
- 数据架构及生产流程合并及优化
- 数据合规化方案选择及成本优化
- 无效或低效数据治理

具体案例

某集团公司，旗下拥有电商、地产、物业多个业态，通过数据采集及整合，构建了相对完善的数据体系。在完成数据分类分级、确保安全合规流通的前提下，为了促进数据流通，最大化发挥数据价值，该公司在内部建立了相关数据流通平台，致力于让不同的业务之间可以合规流通数据，其中的关键一步是设计数据流通过程中的价值评估方法，让数据生产和使用双方在数据使用过程中的成本和相关信息得到充分还原，并进一步指导数据治理工作。

如下图所示。首先，对公司内部生产并维护的所有数据，通过收集维护每张表产出过程中系统的存储、计算、运维等相关成本得到每张表的直接生产成本；然后，通过血缘分摊算法将所有成本分摊至对外使用数据方(图 4)。

具体地，如图左侧在整体的数据生产流程上，通过分摊算法最终数据 A 的成本为 90，数据 B 的成本为 60，数据 A 和数据 B 承载了整体数据生产流程的全部生产成本。该结果可以确保业务方对于数据 A 和数据 B 的使用可以直观反应实际数据消耗成本对于业务的流向情况。如图右侧所示，数据 A 和数据 B 带有数据分类分级标签，在保障合规流后，对具备资质的业务 BU1-BU4 开放数据使用，最终这 4 个业务单元均会承担数据生产成本，如图右下角所示。

基于这一方案，1. 可以实现对拥有复杂数据加工链路的公司进行精准且清晰的数据成本拆解；2. 基于血缘分摊算法的结果（即数据的成本）和治理 ROI 评估体系（不同业务不同场景 ROI 评估都不一样，比较简单的方法包含，成本/访问量，计算单次访问的成本，成本高于一个阈值可以认为是低价值），可以对低价值的数据或数据加工链路进行清理；3. 公司内部不同业务之间的数据交换，可基于分摊后的数据成本，实现较为公平的成本分摊。案对于数据生产成本进行有效分摊。

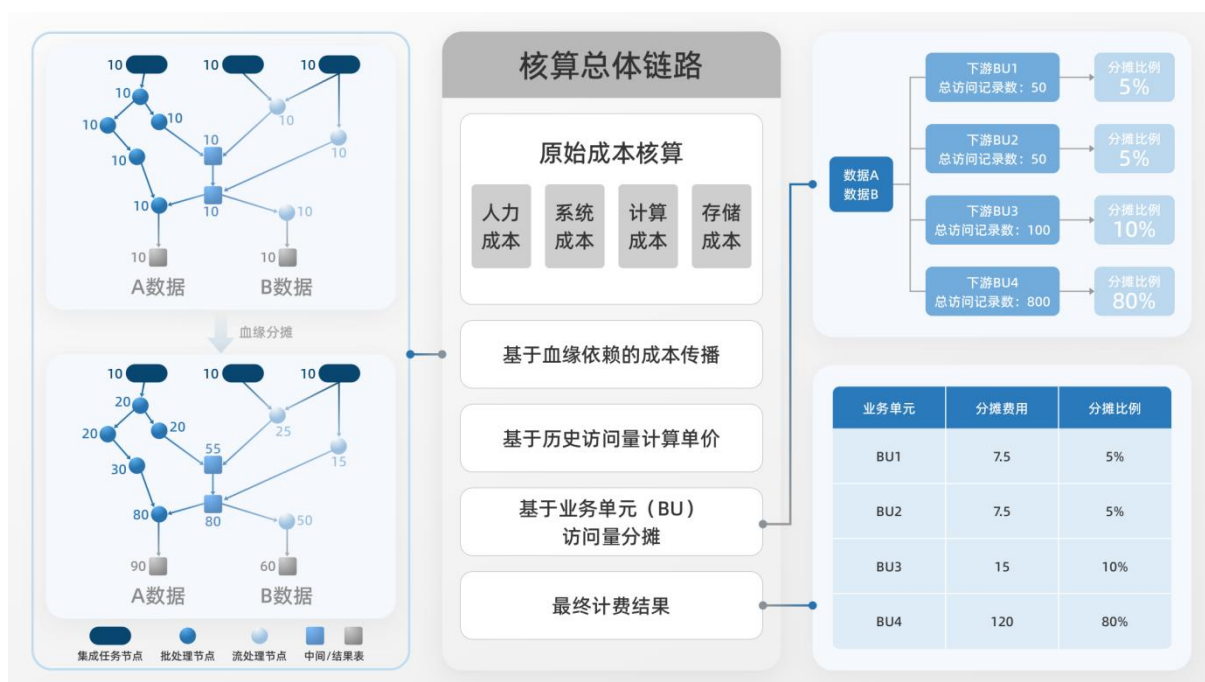


图 4：数据生产成本承担

数据运营及业务策略制定

通过对于数据产生价值所影响的业务线、业务场景进行关联分析，可以对于公司当前运营状况及数据化程度进行衡量，从而影响企业内部业务决策及战略治理，主要影响动作包括：

- 评估重点业务场景、数据产品及其商业价值，为企业战略决策提供依据
- 基于数据价值流通链路调整组织结构关系，保持企业各部门工作顺畅进行
- 确认数据产品价值模块，优化数据产品商业模式
- 了解内部各个部门数据化建设状态，拉齐数据化建设水位，推动内部数据价值流动
- 对于数据相关部门工作目标制定提供方向及考核标准
- 业务价值分配及绩效衡量指标依据

具体案例：

在获取完整的数据流向后，企业可以进一步基于数据血缘分摊相关信息，得到数据对于企业内部各个数据产品、业务场景、相关团队的支持情况。并进一步得到每份数据的实际业务价值贡献，从而得到数据的价值分层。同时，也可以对于高价值数据流向的业务场景、数据产品及团队进行追踪，从而为企业相关业务运营提供指导和战略决策依据（图 5）。

- 业务可以通过分摊后的血缘数据，感知生产链路的变化。如果业务使用表的数量没有大的变化而分摊成本发生剧烈变化，那么从血缘分摊的明细数据可以得知，业务所使用表的成本组成（包含自身生产+存储费用和上游链路分摊下来的费用），继而定位生产链路的变化：包含上游加工链路增删改的变化导致的费用变化，上游任务成本变化导致的费用变化等。这可以一定程度上反应数据链路的稳定性。
- 基于不同评估体系（图 6），筛选出高价值的数据，进行数据监控。通常一个公司的监控资源是有限的，因此，将监控资源放在重点业务和重点数据上是必要的。通过分摊后的数据成本以及数据各维度的打分，计算出数据表的 ROI，继而选出高价值数据进行保障。
- 基于数据表下游服务的业务场景，数量和稳定性，业务可以选出“可信赖的优质表”。通常在一个庞大的数仓中，存在许多相似的表，或者相同的表。“使用哪一张表比较可靠？”这个问题对业务或者数据加工者均是一个难题。因此，拥有丰富的描述信息（包含不限于下游服务的业务场景，数量和稳定性）对业务和数据加工者的帮助是巨大的，这些信息可以帮助他们筛选出质量过硬，价值更高，链路稳定性更强的数据。

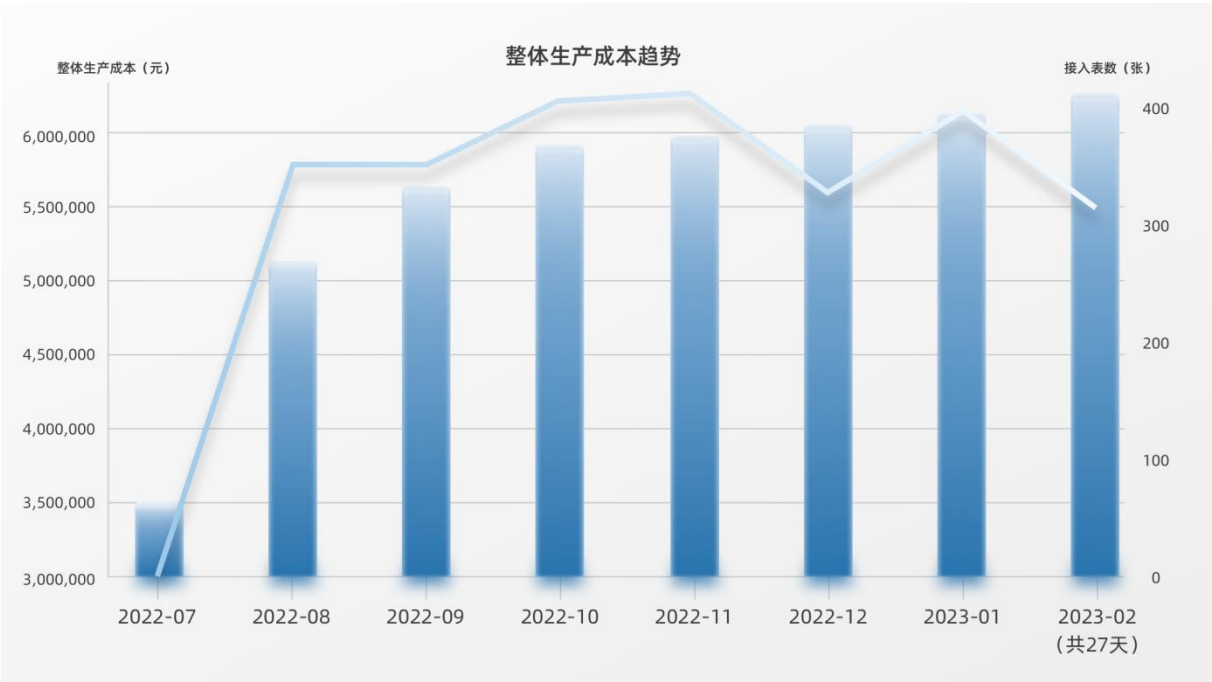


图 5：数据成本生产趋势

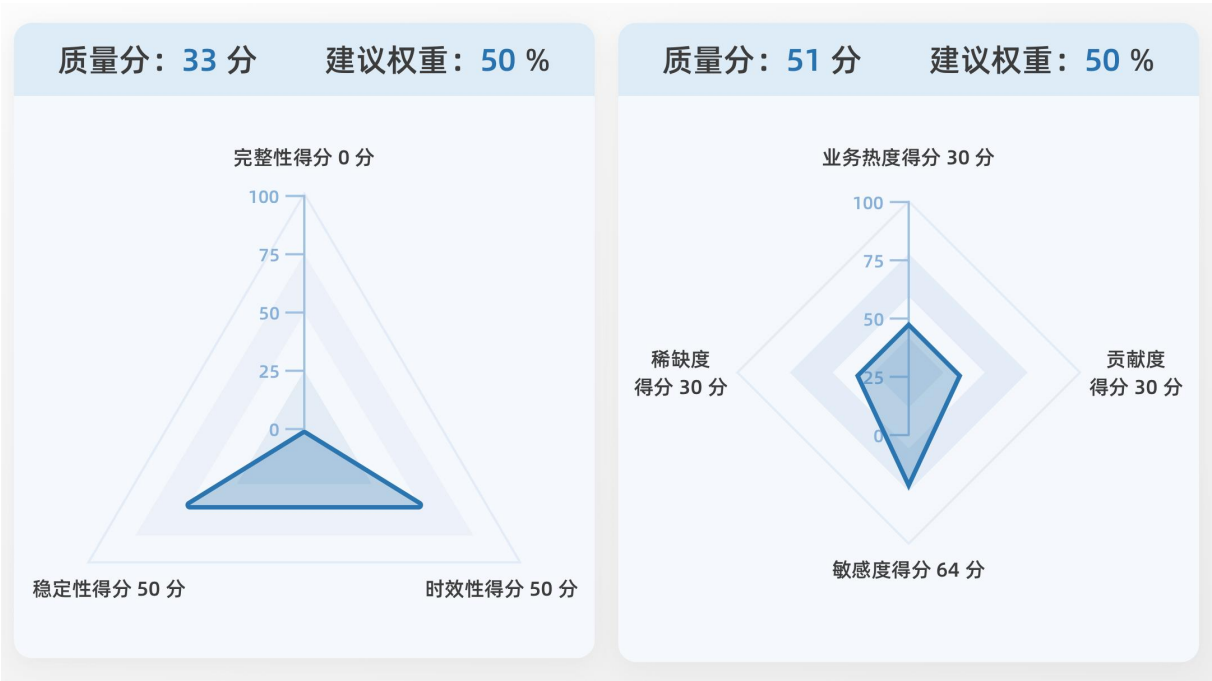


图 6：数据的评估体系

2) 外部数据流通

近年来，各地掀起了新一波数据交易所建设热潮，北京、上海、深圳等地成立了新一批数据交易所，在数据登记、合规审查、估值定价、交易流通等领域进行了深入的探索。本方法论旨在为估值定价的应用进行方法论上的阐述。

前文已经介绍了如何通过“采建管用”在企业内部实现数据全生命周期的价值评估，面对后期全社会多方协同的数据流通体系下，有必要提倡在企业内部治理及外部流通中建立一套统一的价值评估体系，作为数据产品上架交易所时的价值评估，以及数据衍生产品的二次价值评估，具体流程如下(图 7)：

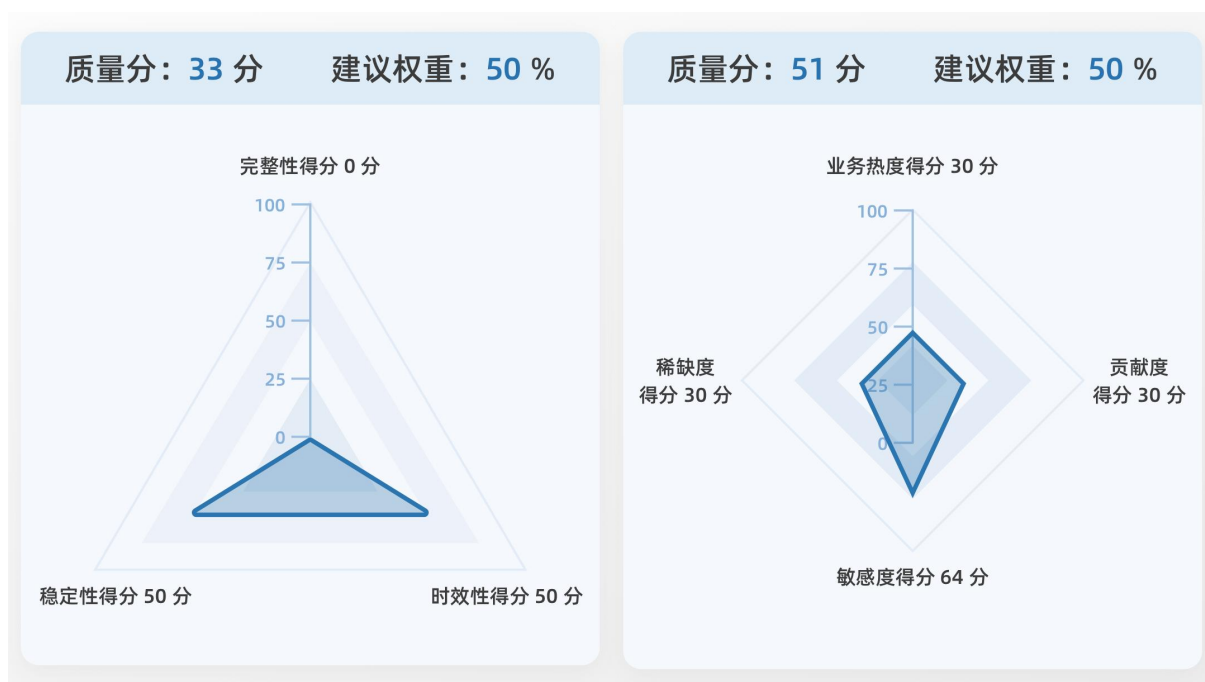


图 7：数据交易资产价值评估体系

- 首先数据持有方（如企业与政府）在内部进行数据产品加工时，依赖价值评估公式，计算基础成本、管理系数、管理要素及回报率，作为数据产品的基本属性在上架到交易所时进行登记，数据产品中每张表的价值之和构成了数据产品整体价值，数据持有方通过该流程可以做到企业内部的数据账本分明。
- 其次在数据交易场所，基于已上架的数据产品进行二次开发时，也同样依赖价值评估公式计算在可信数据空间内成本及回报率。通过交易场所不断产生新的数据产品，

让更多的真实数据参与计算市场收益率，增加价值评估的准确性。所有价值评估的结果作为数据产品定价的参考，让流通交易规则更加量化。

数据可以实现对内成本和权属可知，对外权属和价值释放的过程，为后续在社会上流通提供有依据的衡量标准。

以上场景均可通过如智能数据建设与治理 Dataphin 等第三方产品的资产建设、治理、运营和隐私计算等能力来支持。



图 8: Dataphin 资产评价模块示意

四、总结与趋势

1. 数据资产价值评估总结

数据资产市场目前正处于高速发展的阶段，对于数据资产价值评估领域的关注也在不断加强。政府部门和监管机构开始关注并积极推动相关研究和规范制定，同时越来越多的企业也意识到自身拥有大量有价值的数据资产。然而，尽管已有大量学术研究、行业报告等理论沉淀，但均未突破理论与实践的鸿沟，导致数据资产价值评估体系尚未在企业内部得到有效的落地与应用。

本白皮书打破了数据资产估值理论研究和企业系统落地实践的壁垒难题，通过对数据资产的深度剖析，以及对不同技术手段的探究，明确了对企业内部中所处“通关用”阶段数据表采用“以血缘分摊为核心、管理及交易要素为调节”的核心思路。

数据是通过企业内部采集、存储、加工、应用等一步步加工链路完成价值的逐步积累与释放，而血缘关系则准确地记录了数据在企业内部的价值传递链路。因此，本白皮书通过血缘分析视角，完整、清晰的为企业揭示每一项数据从源端直至应用过程中成本价值积累至释放的详细过程。进而帮助企业有针对性地识别、分析高价值、高成本流程或节点，进一步优化企业内部的数据资源配置。

同时，利用血缘分摊的思路，也可以实现以场景价值为输入的数据资产评估场景。通过记录、分析参与到每一类具体应用场景中的数据血缘关系，可以将应用场景中的总价值逐层、逐步的分摊至各项数据节点，进而为企业提供了以应用收益为主视角的数据资产价值。通过进一步结合成本分摊结果，可最终帮助企业评估每一项数据资产的净收益情况，开展更多维度的分析与评价。

最后，本白皮书结合血缘成本分摊的特点，从企业内外视角，分析了估值系统的具体应用场景和实践案例，旨在帮助企业更高效、顺利地在内部完成数据资产估值系统的落地实践，帮助企业更好地感知、理解数据资产估值对数据管理工作的价值提升作用。同时，确保数据资产估值这一创新体系成果可以更好地融入并协同其他数据管理工作模块，实现为数据为企业发展赋能的价值。

2. 数据资产价值评估未来趋势

随着数据资产估值系统在企业内部得到落地与应用，数据资产管理工作将实现由“人员主观判断”向“量化客观评价”的质变转型。数据资产价值评估将帮助企业科学、高效、精准地区分出内部的“低投入、高产出的优势资产”和“高投入、低产出的劣势资产”。这将使企业能够进一步丰富、拓展自身的运营管理手段，持续深化优势数据资产的成熟水平，并在巩固优势地位的同时，有效迁移成熟经验，赋能优化其余数据资产。同时，企业也可以及时采取止损手段，停止对劣势数据资产的持续投入，优化内部资源配置。通过逐步打造企业自身的核心数据资产竞争力，真正实现数据资产对企业战略发展的赋能与提升。

另外，数据资产估值体系的实践经验将加速数据要素市场整体发展的速度，为未来数据资产核算入表、市场交易定价等数据流通场景提供丰富而关键的理论基础和实践参考，真正实现数据资产的变现与流通。例如，在财政部发布的《企业数据资源相关会计处理暂行规定（征求意见稿）》中则明确说明成本核算是数据资产登记入表的基础与前提。同时，未来的数据资产估值体系也将超越企业的视角，立足于社会、国家乃至国际的视野，通过评估数据资产要素在社会生产经营活动中的具体价值，确保国家和社会可以更好地把握数据资产的经济规律，为人类社会的发展提供更加高效且长效的动力支持。

同时，数据资产估值技术并不是一项独立的技术领域。面向大模型，数据资产在其发展中具有关键性的作用，数据决定了模型的训练质量、性能表现和应用领域的广度与深度，高价值数据资产的有效供给是保障大模型技术的关键。未来高价值的数据一定是和人工智能和专业领域行业知识结合，充分促进企业的数字化发展，即数据会有更高阶的杠杆价值。在面向未来的数据资产价值评估的框架中，需要充分考虑到面向应用落地的智能技术的因素影响。

另一方面，大模型也可以助力于智能的资产价值评估体系的构建。例如，在未来大模型可以通过搜集并分析海量的市场交易数据，精准揭示数据市场供需关系，帮助决策者制定合理的定价策略；大模型也可以基于企业历史数据资产价值数据，建立预测模型有效预测数据资产未来价值变化趋势，为投资者、管理层等提供决策参考、数据资产配置及优化等策略内容，进一步丰富、释放数据资产估值领域的技术应用场景与价值。

总体而言，数据资产价值评估市场有着巨大的潜力和发展空间。随着技术的进步和监管环境的改善，这个市场将继续壮大并提供更多创新的解决方案。未来，将会有更多企业和个人积极地参与到这个市场中，以实现数据资产最大化利用的价值。



钉钉扫码加入「数智俱乐部」
与 5000+数据人交流互动

编者团队

本次报告由瓴羊、阿里云和中国信息通信研究院等机构合力完成，在此向撰写本报告及对本报告提出研究指导的团队
成员致以感谢。

董芳英 瓴羊数据技术产品总监

黄欢欢 阿里云技术架构总监

叶笔长 瓴羊数据产品经理

李冬青 阿里云技术架构师

刘吟啸 淘天集团算法工程师

余亿 阿里云业务架构师

李林洋 阿里云技术架构师

游杰 爱橙科技算法工程师

庄隆生 阿里云业务架构师

潘霏晨 阿里云解决方案架构师

邓镭 阿里云技术架构师

李铭洋 阿里云技术架构师

马永刚 瓴羊技术工程师

宗文健 爱橙科技工程师

姜春宇 中国信通院云计算与大数据研究所主任

王妙琼 中国信通院云计算与大数据研究所副主任

李雨霏 中国信通院云计算与大数据研究所高级业务主管

艾博焕 中国信通院云计算与大数据研究所工程师

马闻达 中国信通院云计算与大数据研究所工程师

邱梦媛 中国信通院云计算与大数据研究所工程师

About

瓴羊

瓴羊是阿里巴巴全资子公司，也是阿里云智能集团的重要业务，致力于将阿里巴巴沉淀十余年的数字化服务经验，系统化、产品化地全面对外输出给千行百业。按照企业的不同行业和具体情况，瓴羊提供一系列数字化产品、解决方案和服务，充分激发数据要素作用，帮助企业完成有效的、彻底的、全流程的数字化建设，实现数实融合。瓴羊是阿里巴巴通过数字技术促进实体经济和数字经济融合的核心成果，也是阿里巴巴携手多方，推进数字中国建设中的重要力量。瓴羊已服务了上百家知名企业和众多中小企业的数字化建设，其中包括一汽红旗、现代斗山、极氪等制造业，宝洁、星巴克、麦当劳等跨国企业，伽蓝、三只松鼠等新国货消费品牌。

About

阿里云

阿里云是一家云计算及人工智能科技公司，坚持让计算成为公共服务，通过提供安全、可靠的计算和数据处理能力，让云计算和人工智能成为普惠科技。目前，阿里云是亚太第一、全球前三的云服务商。从全面上云，到云上创新，阿里云彻底改变了云的使用方式和云上应用开发方式，让不懂得代码的人也能使用云的能力，大大降低了创新的门槛。我们和全球数字创新者一起，以科技应对挑战，为美好社会建设带来积极改变。

About

中国通信标准化协会大数据技术标准推进委员会

中国通信标准化协会大数据技术标准推进委员会（CCSA TC601）自2018年成立以来，紧贴国家数据要素重点政策导向，围绕数据要素产业发展关键问题，开展大数据技术产品、数据资产管理与流通、大数据行业应用方面的标准研制、专题研讨、行业调研。TC601目前下设17个工作组，成员单位数量已超过200余家，服务政府、通信、金融、能源、制造等多个行业。TC601将密切关注成员单位技术创新和生态培育需要，通过搭建行业交流平台，持续推动数字经济与实体经济深度融合。

