WeBank 微众银行 · 区块链

金融科技 微洞察

# 打造数据新基建 释放数据生产力

## —— 微众银行数据新基建白皮书

2020年12月

# 序言

当下，移动互联网、云计算、物联网、大规模存储、高性能计算和芯片等信息技术蓬勃发展，世界进入数据爆炸的“大数据时代”。与传统的资本、土地、劳动、技术等一样，数据已是重要生产要素之一，与算力、算法组合，作为一种新型社会生产力，在人们的生产生活中发挥显著作用。

但是，与历经千百年的传统生产要素不同，数据深入塑造人们生产生活的历史仅有二十余年。正如一个青春期的少年，它让我们“又爱又恨”，爱它方便和提升了我们的生活，恨它有时候因泄露、滥用、盗用等问题造成困扰。这些“成长的烦恼”阻碍数据发挥最大的价值，羁绊其成为可信赖、可持续的生产力。

生产力的发挥有赖于基础设施的进步，数据要素同样如此，需要合适的基础设施以克服这些问题，解放生产力。以人工智能、区块链、云计算、大数据等为代表的数字时代的最前沿科技，代表着先进生产力的方向，有望为数据要素构筑起新型的数字化基础设施。

作为国内首家互联网银行，我们在多年生产经营和服务客户的实践中深刻认识到，可靠的数据应用、规范的数据治理具有十分重要的意义。我们融合在金融科技领域积淀下来的能力和经验，为自己、为客户、也为社会提出一套完整有效的数据应用和治理解决方案，希望有助于全产业和社会深入认识和充分释放数据生产力。

本白皮书着眼于此，将展现我们的“数据新基建”构想、方案和实践。白皮书首先从生产要素释放生产力的基本条件出发，深入探讨束缚数据生产力的问题现象和成因。为了解决这些问题，白皮书将数据新基建的关键特性归纳为 3 个核心要求：安全存储、可信传输、协同生产。以此要求为指引，以数字科技为依托，白皮书随后介绍我们的“数据新基建”具体方案及应用案例。这些方案和应用来源于我们的多年实践，希望能为业界带来启发意义和引导价值。

我们期冀，透过这份“数据新基建”方案，激发高效稳定的数据生产力，成为引领经济发展的核心动能；搭建可靠安全的数据连接网，成为奠定生活美好的宝贵基石；体现与时俱进的数据治理观，成为推动社会前进的重要指南。我们谨以此白皮书为数据新基建的发展提供一定的思路和方法，望能携手各界伙伴，共同为我国数字经济的腾飞贡献力量。

# 目录

# 生产要素释放生产力的基本条件

历经农业经济、工业经济时代，人类社会而今迈入了数字经济时代。每一个大时代的跨越都以生产力的变革为标志，早期以农民自身体力、牲畜和简单工具为主，工业时代依靠石油电力和大机器来生产，数字经济时代则在计算机和互联网的推动下迅猛发展。生产力蕴含于生产要素（factors of production）之中。生产要素是指能够产出（output）产品和服务的经济资源投入（input）[1]，在人类历史的长河里，主要的生产要素同样经历了变革。

农业经济时代，主要的生产要素是土地和劳动。农民耕作在土地之上，土地的数量和质量很大程度上决定了农业产出的规模。除了农业，人们还利用各种各样金属工具来完成酿酒、制陶、水利、盖房等工作，酒品、器皿、工程、房屋等产出就凝结了人们的劳动。

工业革命之后，除了土地和劳动，资本和技术成为主要生产要素，在经济发展过程中起到非常重要的作用。蒸汽机、石油、电力、铁路、通讯、化工等新兴技术的出现，让人类社会生产力达到一个空前的高度。技术驱动下，生产从家庭小作坊走向工厂，需要规模化的投资，现代意义的工商企业开始发展，全国化、全球化的市场产生，不仅促成了大量设备和固定投资的形成，也带动了社会对金融资本的旺盛需求。以金融、设备、厂房等形式积淀的资本，成为推动经济发展的重

1　定义来自于《大英百科全书》

要动力。资本和技术对经济的拉动作用不断超过劳动。以 1760-1900 年工业时代的英国为例（图 1 所示），将单位劳动贡献的经济产出（Y/L，其中 L 代表劳动投入，Y 代表经济产出）拆解，代表技术贡献的全要素生产率（total factor productivity, TFP）对 Y/L 拉动巨大。此外，资本密度（单位劳动拥有资本，K/L）不断增加，对 Y/L 的拉动也大幅提高[1]。

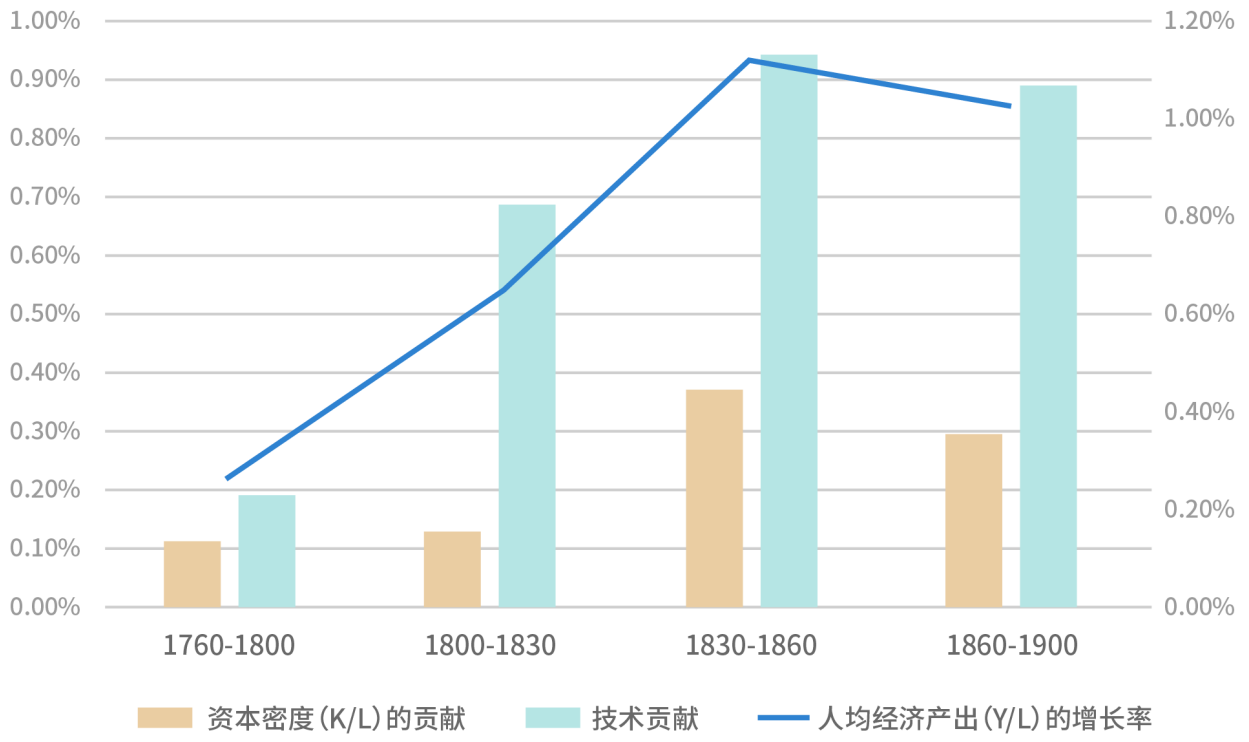



图 1：资本和技术对工业革命时代英国的经济增长拉动巨大

数据来源：Allen (2009)

进入信息经济时代，数据成为新的重要生产要素。互联网集中起分散在线下的各类数据信息，处理优化之后将信息呈现和传递给需求者，产生形形色色的应用。数据越丰富，从中挖掘和创造的价值就越大。例如，金融机构可以通过丰富的个人信息、购物信息、银行账户信息、社交信息等等更好地刻画一个人的信用肖像，增加风控和营销能力。据估计，目前每年产生的数据总量达 40ZB，若将这些数据储存在 DVD 里，将所有 DVD 叠加起来可绕地球 50 圈[2]，其中蕴含的宝藏价值不可想象。

1 Allen, Robert C. 2009. “Engels ’ pause: Technical Change, Capital Accumulation, and Inequality in the British Industrial Revolution.” Explorations in Economic History 46(4): 418–35.

2 IDC. Data Age 2025.

不论是土地、劳动、资本、技术还是最新的数据，生产要素要成为真正的生产力，在社会大生产范围内发挥价值，应当满足一定条件，否则只是局限在单一组织、单一个体内，不能在市场经济环境下产生应有的市场价值。这些条件可概括如下：

- **要素的产权可清晰界定**。新制度经济学揭示了，当产权清晰时，经济活动才是有效率的，对市场主体才能产生合适的激励[1]。这里的产权是“权利束”(a bundle of rights) 的广义概念，即产权实际上包含资产的所有权、使用权、收益权、抵押权、处置权等不同性质的一组权利，权利束的不同成分可能归属于不同的主体。例如，同一土地的使用权和所有权可以分离，是独立的权利。产权清晰是指要素的单个权利或整个权利束的拥有者清楚明晰，不存在归属模糊、多方同时拥有一个权利的情况。以技术要素为例，技术有可能通过申请专利予以保护，确定发明者对技术的产权，保护发明者的切身利益。

- **要素的价值可评估**。生产要素在多方之间交换流通，才能在更大范围内发挥自身价值。一个基本条件是要素价值应可靠、可信，参与各方有能力评估要素的价值，否则欺诈、逆向选择等现象会导致很高的交易成本。要素使用方也没办法评估投入产出比。传统生产要素通常有成熟的价值评估方法，例如股票的价值可以由财务报表反映，劳动力的价值可以由过往简历反映或可信第三方推荐，技术专利的价值可结合技术创新程度、专利保护期限、市场需求等因素来评估。

- **要素的价值可流通，具有一定通用性**。如果要素不可流通，只能局限在单一组织、单一场景和单一用途，在一方手中的要素很难被另一方所应用，就出现了一个个割裂的 “孤岛”，失去了要素的交换价值。当要素在一定范围内具有通用性，并且以合适的方式能够流通时，生产要素就可以在不同组织、场景、产业复用，使其生命周期内总价值（life time value）最大化。以技术要素为例，专利能够在不同企业间以转让或许可的形式流通，满足不同企业的需求。

- **要素的价值可存储，在一定时期内具有稳定性**。除了一类特殊的生产与消费同时发生的要素——电力，其他生产要素的生产和消费之间往往有一定时间间隔，如企业需要一段时间才能消耗所融到的资金，工业资本品的使用寿命很长。如果要素不能可靠地存储价值，不具有稳定性，那么它就很难流通交换以产生更多价值。

1 Coase, Ronald H. 1988. The Firm, the Market, and the Law. ed. 陈郁 . Chicago, IL: University of Chicago Press.

产权可界定、价值可评估、价值可流通、价值可存储，是生产要素能在市场经济中流通、交换、应用，以释放生产力、最大化自身价值的 4 个基本条件。土地、劳动、资本、技术要素都在不同程度上满足这四个条件（表 1），因此成为了可靠的生产力。

表 1：传统生产要素释放生产力的 4 个基本条件

| | 产权可界定 | 价值可评估 | 价值可流通 | 价值可存储 |
|---|---|---|---|---|
| 土地 | 地契清晰载明土地产权 | 根据面积、区位、土地性质、土壤成分来确定价值 | 土地有招拍挂市场 | 土地的价值稳定性高 |
| 劳动 | 劳动力的产权属于劳动者本人 | 根据过往职业和教育履历可评估劳动力价值 | 人才可以在市场上自由流动 | 人才的技能在若干年内保持稳定 |
| 资本 | 证券的产权归属明晰 | 证券价值可通过公司业绩来评估 | 证券可流通 | 证券价值可存续，尽管有波动 |
| 技术 | 可申请注册专利清晰界定和保护产权 | 专利价值可根据创新程度、权利期限、市场需求等评估 | 专利能许可转让，满足不同企业的需求 | 专利作为无形资产，一定时期内其价值稳定 |

# 数据要素的特性与生产力释放难点

与传统生产要素相比，数据要素却在释放生产力的每个条件上都有一定困难，原因来自于它有着与传统生产要素差异很大的若干个特性。

- **数据具有易复制性（replicable）**，复制的边际成本极低。数据信息的生产和整理成本较高，通常需要做大量的搜集、清洗、分类、标注等工作，但一旦整理完毕，入库形成可用的数据集、标准的数据文件后，生产复制额外副本的成本可忽略不计。所以，数据具有高固定成本、低边际成本的特性。

- **数据具有非排他性（non-exclusive）和非竞争性（non-competitive）**，即一份数据可以同时供无数人使用，也不会因为用的人多而产生损耗。相反，有可能因为用户增多而让数据样本更加丰富、维度更多，从而提升数据的价值。这个特性让数据只要一公开，就成为“公共品”，人人可共享。

- **数据具有分散性（scattered）**，数据持续不断地从各个途径产生，来源非常分散。这个世界上每个人的衣食住行和工作都在产生数据，据 Intel CEO 估计，到 2020 年互联网用户个体每

天产生 1.5GB 的数据[1]。除了个人数据，每个企业和政府组织也在不断生产数据，如工厂生产状况、物流运输数据、经济统计数据等。即使是同一个属主的数据，一般也分布在不同渠道上。例如一个普通人的饮食外卖数据会在美团、饿了么上，打车出行数据会在滴滴上，工资数据会在银行上。

- **数据具有多样性（heterogeneous）**。数据种类杂、结构乱，既有个人数据，也有各种机器设备数据，既有结构化的表格数据，也有非结构化的图片、视频等。

- **数据具有价值聚合性（aggregated）**。单一少量数据只能反映一小部分样本的情况，统计意义不强，少数维度的数据往往反映了事物的一个侧面，这两种情况下，数据的应用价值受限。但当数据量和种类增加时，多维数据、海量数据的联合应用分析有助于揭示事物的完整特性，产生“1+1>2”的鲜明效果，显著提升应用价值。例如，银行可以聚合多方数据建立对小微企业的风险评估模型，从而授信小微企业。

- **数据具有价值认知多样性 (individually-respectable)**。同一类数据对于不同属主的价值可能存在巨大的差异，例如，敏感实体的财务信息其价值往往显著高于一般实体的同类数据，对于不同敏感等级的数据如果不进行分级处理，将难以尊重每一个属主的数据隐私诉求，势必会影响高价值属主参与数据协作的意愿。

数据的这些特性使得它产生如下问题，很难满足要素释放生产力的几个条件。

- **数据权属确认有难度，导致产权不够清晰**。相比于传统资产，数据要素的产权归属比较模糊，法律上饱受争议，缺乏明确结论[2]。其原因除了权利主体多元化、产权分割困难等法律性质外[3]，还有一个很重要的技术原因是非排他、非竞争性、易于复制。经典产权的基础来自资源稀缺性和排他性使用处置，但数据——尤其是互联网世界的数据——往往具有源源不断产生、共享开放访问等原生属性，且复制和存储成本几乎为零，很容易被复制和盗用，这就与产权的基础相矛盾。

1 https://www.twice.com/industry/intels-krzanich-has-seen-the-future

2 阿里研究院 . 2020. 数据生产力 .

3 闫立东 . 以“权利束”视角探究数据权利 . 东方法学 . 2019 年第 2 期 .

- **数据易泄露、易盗用、易滥用、难追踪，也导致产权保护差，数据属主和控制方的利益得不到保护。**数据从原始产生、记录、整理加工到使用，可能会有多个主体参与、经过多个环节和使用场景，又因为复制成本极低、非排他、非竞争性等原因，就很可能遭遇泄露、盗用等问题，流向难以追查。而且，目前数据被滥用、个人隐私被侵犯的事情时有发生。对于互联网机构和金融机构来说，客户丰富的数据信息通常具有机密性，一旦被泄露或被滥用，很可能给客户带来巨大损失和麻烦。目前，隐私信息泄露和非法利用的案件常常发生，基于隐私数据的商业探索仍常常触犯红线。

- **分散的数据之间尚不能做到互联互通，可信程度不够，导致流通性不足，价值评估较难。**数据要素的价值聚合性决定了集成多方数据资源会产生大的价值。然而，现实世界中数据高度分散，缺乏统一的授权、获取、存储、传输、验证及共享等交互标准，更重要的是各方不愿意、不能够共享数据，导致一个个“数据孤岛”（data silos）出现。即使数据能联通，可信程度也存有疑问。数据难流通、弱可信，还导致了各方之间不容易评估他方数据价值。

- **海量、异构、多维数据的存储和处理难度大。**在这个大数据时代，业务会面临着海量用户和交易的不定时冲击，带来海量数据的吞吐处理需求。系统通常在短时间内计算数据，以实现动态响应和策略应用；事务完成之后，必须按照合规要求安全可靠地长期存储数据。同时，多种多样的数据要求用不同类型的数据技术来存储和处理，比如传统的用户信息数据可以用经典的数据库，但社交网络关系数据用图数据库会更合适，增加了数据存储和处理的复杂性。

# 释放数据要素的生产力呼唤“数据新基建”

当生产要素不满足生产力释放条件时，合适的基础设施就有可能发挥重要作用，改变现实状况以满足其中一个或者几个条件。事实上，传统要素在历史变革中，也依赖于升级完善的基础设施以最大程度地满足条件、释放生产力。

在农业经济时代，土地和劳动要素相结合，产出粮食作物等农产品，不仅在本地消化，还需要运送到外地，让土地和劳动的价值最大化。但如果没有较好的道路基础设施，价值可流通的条件就会弱化。到了工业时代，随着铁路、港口等基础设施的扩展，凝结着劳动要素价值的产品更具有全球流通性；电报电话等基础设施同样为资本的全球流通创造了条件。在信息经济时代，通信和计算机基础设施为全球证券市场搭建了良好的资讯、估值、流通体系，极大发挥了资本的能量。

同样，数据要素需要合适的基础设施。而且与传统生产要素不同，由于数据的抽象、虚拟和数字特性，它的基础设施应该是一种基于数字技术的新型信息基础设施——本白皮书称为“数据新基建”。数据新基建依托人工智能（AI）、区块链（blockchain）、云计算（cloud computing）、数据科学（data science）等数字技术（可统称为“ABCD”），有助于克服数据要素的主要问题，满足如下几个特性：

- **有助于确认数据权属，追踪数据流动，从而清晰地界定和保护产权**。例如利用区块链技术防篡改、可追溯的特性，在重存证和溯源的业务场景中，新基建能够通过数据存证确立产权，并在后续数据流传过程中追踪数据的分享传播，更好地保护产权。

- **确保数据全生命周期过程中的合规，充分保护数据属主隐私**。结合人工智能、大数据等技术加强数据应用业务的监管能力，尤其在高度依赖用户数据的金融、互联网等行业，保障合规应用，不侵犯数据产权持有人应有的权益。隐私保护始终是数据应用的红线，而且许多业务场景中，敏感数据不可以流通交换，只能留在本地。为了解决隐私保护问题，以安全多方计算、联邦学习等为代表的一系列技术就有了用武之地，显著提升多方机构对敏感数据的合规应用和隐私保护能力。区块链和密码技术相结合，也能有效增进数据在传输过程中的安全隐私性。

- **有效实现碎片数据、孤岛数据之间的互联互通，并确保不同来源数据的可信可验**。区块链为连通碎片化、孤岛化的多方数据带来了可能性。各方节点将数据或其摘要上链存储，进一步结合密码算法在链上协同计算和完成交易。跨链技术还能将不同业务链上的数据相互连通，以便在更大范围内发挥数据的协同价值。联通之后，许多场合下数据属主或控制方不愿意披露具体内容，但又必须提供合适的手段证明自己数据的价值和可信度，满足对数据价值的高效安全评估。以零知识证明、同态加密、差分隐私等为代表的密码算法能让各方在数据机密不泄露的情况下计算和验证交易的正确性，实现了数据的可信检验。

- **保证数据的安全可靠存储和计算**。云计算和分布式架构等技术能重塑数据计算和存储的基础设施，以其广泛接入、资源共享、弹性伸缩、按需使用的特点为大数据提供了强大的计算和存储能力，实现了高弹性、高可用、低成本和低风险的服务。

以上 4 点理想特性实际上可归纳为三大核心要求：**安全存储（secure storage）**、**可信传输（trusted transfer）**、**协同生产（collaborative production）**。它们的内涵如下表所示。

表 2：数据新基建的三大核心要求内涵

| 内涵 | 安全存储 | 可信传输 | 协同生产 |
| --- | --- | --- | --- |
| 特性 | 可靠存储，安全计算 | 数据可信可验，可确权，可追踪 | 互联互通，隐私保护，合规应用 |
| 典型方法 | 外部托管、本地存储、可信执行环境（TEE） | 授权使用、加密传输、数据鉴证 | 联合建模、隐私计算、融合分析 |
| 技术要求 | 1. 用户可自主选择存储策略；<br>2. 数据加密、隔离；<br>3. 数据可恢复、可删除 | 1. 准确、高效、安全；<br>2. 用户授权、选择性披露；<br>3. 防篡改、可分布式验证、可追溯、可审计 | 1. 合法合规、隐私保护；<br>2. 根据场景选择方案；<br>3. 用户价值最大化、合理化 |

安全存储是指满足数据安全计算和可靠存储要求，这是数据要素释放生产力的基础。数据存储的通常方式包括本地存储、外部托管或第三方的可信执行环境（trusted execution environment, TEE），用户应能根据需要自主选择合适的存储策略。在存储时，数据应有相应的加密和隔离措施，可根据需要删除数据，或在误删之后有恢复补救机制。

可信传输是指数据在不同所有者和控制方传递过程中，能追踪数据全流程，保护好产权，并保证数据的可信任、可检验。这是数据要素释放生产力的必要步骤。在数据传输时，数据应只能流向获得合法授权的接收方，传输过程应有合适的加密方案和鉴证验真方案，确保所传递的数据不被篡改。因此，可信传输的基本技术要求既有准确、安全、高效，也包括用户授权、选择性披露，不能突破授权范围进行披露。传输过程应具有防篡改和多点验证的能力，全流程也应可审计和可追溯。

协同生产是指打通多方之间的可信数据，互联互通，让更广范围内的更多数据联合发挥更大价值，在此过程中同时注意隐私和合规。这是数据生产力释放的形态。常用的方法有数据融合分析、联合建模、隐私计算等。合法合规、隐私保护始终是数据要素投入生产时所遵守的基本前提，生产方应根据不同的使用场景选择合适的方案。协同生产还有一个重要问题，就是设计合理的流通激励机制，给予贡献者合理的回报，让用户的价值在协同过程中最大化、合理化，这样各方才有动力源源不断地分享贡献数据，促进生产。

基于 ABCD 等数字技术搭建的“数据新基建”将通过满足以上三大核心要求，使数据要素能克服固有缺陷，释放生产力。下文将详细阐释我们提出的解决方案。

# “数据新基建”解决方案

传统基础设施以钢铁、水泥、木材为材料，根据不同用途和场地情况，用浇筑、铸造、铆接等技术组合起各部件，满足人们生产、生活需要。类似地，“数据新基建”同样依托于 ABCD 等各项技术的融合，以这些技术为基础，根据现实场景的不同需求进行立体灵活的组合，解决数据要素生命周期中不同环节痛点。在此过程中，设计开发人员可采用开源开放的方式，持续提升这些基础技术，创新性地组建应用解决方案，以提升基础设施的性能、增加和优化功能，保证基础设施安全稳定，丰富技术和应用生态。

## 4.1 新基础设施的关键底层技术

### （一）区块链

从数据视角，区块链可定义为一种服务于多方的软件架构，作用在于促进跨组织边界的可信数据流通和传输[1]。区块链具有一系列鲜明的技术特点，使其能成为非常关键的数据基础设施底层技术。基于分布式系统技术，它可以用共同组网、共同维护数据的基本形式，构建多中心的协作模式；在协作过

1 Forrester. 2018. Emerging Technology Projection: The Total Economic Impact of IBM Blockchain.

程中，基于密码学技术，它能有效保护数据安全和隐私，保证数据不可篡改；基于智能合约技术，它能激励各方数据的交换流通，形成良好的系统治理，促进数据价值最大化。由此，区块链技术解决了分布式数据资产的存储和调用问题。

从现实商业应用看，在构建数据基础设施时，联盟链是主要技术路线。与比特币、以太坊等强调去中心化、去监管的公有链不同，联盟链更注重权限控制、监管治理、性能提升和安全保障，能满足现实商业场景的信任需求。目前国外有 Hyperledger Fabric 等联盟链底层框架，国内有 FISCO BCOS、AntChain 等框架。

## （二）隐私计算和人工智能

隐私计算是指面向隐私信息的采集、存储、处理、发布（含交换）、销毁等全生命周期过程的计算理论和技术，在保证数据提供方不泄露敏感数据的前提下，分析计算数据并能验证计算结果，安全地实现数据价值[1]。隐私计算并不指单一技术，而是包含了人工智能、密码学、数据科学等多学科的综合性技术体系。在目前实践中，隐私计算通常与人工智能的其他方法结合，应用于多方联合机器学习建模。

根据需求不同，隐私计算可用于保证数据输入或输出的信息保护。输入隐私保护（input privacy）是指计算方无法接触或分析出实际输入的真实数据，甚至中间计算结果，目标是防止数据泄露；输出隐私保护（output privacy）的目标是减少能从公开计算结果中还原出的真实输入，目标不在于防止数据泄露，而在于避免从公开结果中反推出原始数据[2]。前者又称为“数据计算过程保护”，后者又称为“数据计算结果保护”[3]。在隐私计算技术体系中，安全多方计算（secure multiparty computation）、联邦学习（federated learning）、同态加密（homomorphic encryption）、机密计算（confidential computation）等技术属于输入保护，差分隐私（differential privacy）属于输出保护。

国内已经有微众银行 FATE 和 WeDPR、腾讯 Angel PowerFL、蚂蚁集团 Morse、富数科技 Avatar、平安科技蜂巢等隐私计算框架，国外还有谷歌 Asylo、脸书 CrypTen 等。

1 李凤华 , 李晖 , 牛犇 , 陈金俊 . 2019. 隐私计算——概念、计算框架及其未来发展趋势 . 工程（英文版）. 第 5 卷，第 6 期 . 1179-1192.

2 UN handbook for privacy-reserving techniques.

3 中国信通院，阿里巴巴集团，数牍科技 . 2020. 隐私保护计算技术研究报告 .

## （三）大数据

在数据基础设施的关键技术中，适用于海量、多样化数据计算、存储、交换、分发的大数据底层技术框架和平台是重中之重。优质的大数据应用平台要具备可靠基础计算存储数据交换能力，具备支持机器学习的能力，具备高并发、高可用、多租户隔离和资源管控等执行与调度能力，具备让业务与数据快速实现互动、高效生产报告的能力，还要具备提供数据地图、数据脱敏工具、数据质量工具的能力。

为此，业内已经在开发各种各样优秀的大数据相关底层技术框架和软件应用。从开源的 Hadoop、Spark、Hbase、KubeFlow 等组件，到 TDSQL、Snowflake、TiDB、AWS Redshift 等知名数据库，以及 WeDataSphere 等数据管理平台套件。

## （四）云计算

在这个大数据和互联网时代，企业业务常常面临海量用户和海量交易的不定时冲击，带来海量数据的吞吐处理需求。系统通常在短时间内计算数据，以实现动态响应和策略应用；事务完成之后，必须按照合规要求安全可靠地长期存储数据。但是，传统的企业 IT 系统依赖本地化、集中式架构，扩展能力有限，建设和运维成本居高不下，难以满足高性能、高弹性、高可靠的计算和存储需求。

伴随着虚拟化、云平台、分布式资源管理、海量分布式存储、云安全等核心技术的发展，人们能快速、可靠、低成本、高效率地计算和存储数据，有助于构建数据要素市场的稳健基础设施。云计算以其广泛接入、资源共享、弹性伸缩、按需使用、容错恢复的特点为大数据提供了强大的计算和存储能力，实现了高弹性、高可用、低成本和低风险的服务，确保了业务交易和业务数据的稳定性与安全性。

## 4.2 面向三大核心要求的解决方案

### 4.2.1 安全存储

为了保障数据安全，首先需要为数据的所有者提供可靠的存储方案，保护数据在载体上的完整性、正确性，且可备份可恢复。其次，确保只有数据属主才能控制数据访问，数据不会泄露，不会被越权访问，不会被篡改，即做到“不丢、不错、不泄露、不篡改”。

根据位置和方式不同，数据存储所面临的挑战和需要使用的技术也有所不同。我们面向几个关键场景阐述安全存储的方案。

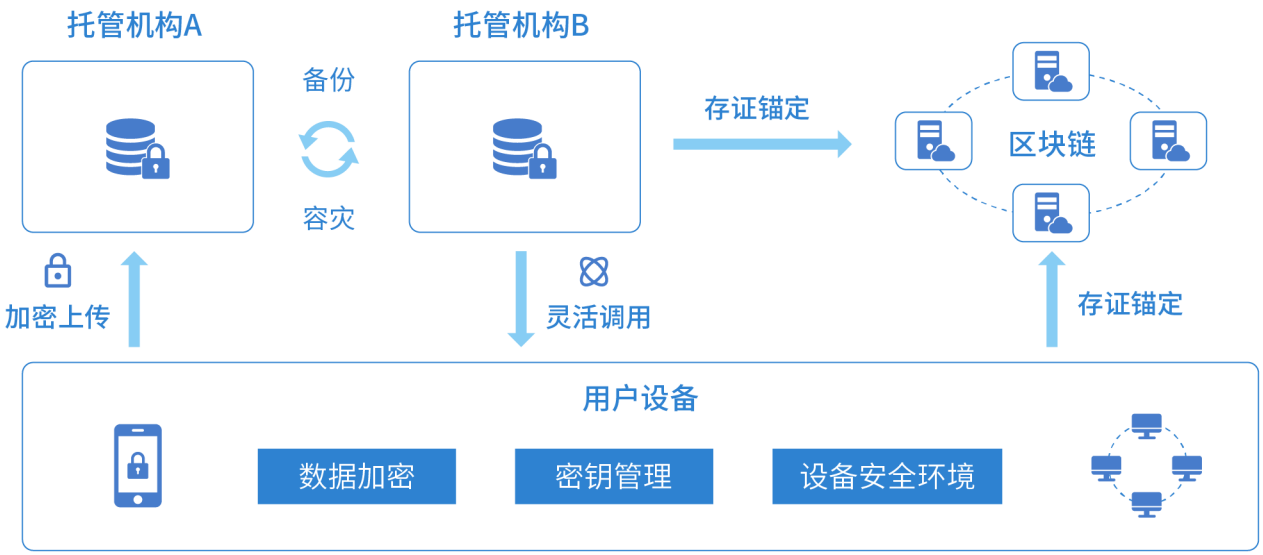



图 2 数据托管存储方案示意

#### （一）终端设备存储

个人数据往往存储在本地的设备上，如手机、个人电脑、物联网设备等，具体数据包括个人信息、与身份验证相关的证书和密钥、APP 产生的本地数据、照片视频等文档、由智能设备采集的信息如步数、环境信息等。

本地数据存储的安全挑战在于是否会被恶意程序访问、是否会出错和误删除。对于关键数据，可采用以下方案增强保护：

1) 加密存储：采用高强度的多重加密技术，将数据加密后存储，且在存储时加入数据水印、数据指纹等验证信息，避免数据出错和被篡改。相关密钥由用户掌握，存储于独立的空间。在用户使用密钥时，提供防窃取的输入方式，以及引入生物特征、手机短信等多因子验证的加强方案。另外，为用户提供密钥恢复或重置方式，避免密钥丢失或泄露。

2) 访问控制：将关键数据隔离存储于诸如 TEE 等安全区域，精细化的控制应用层对数据的访问权限，拒绝非授权的读写和未经允许的网络传输，并对数据的访问和传输留痕审计，对可疑操作提出警示。

3) 存证追溯：基于区块链等技术对数据进行存证，在可信网络上留存数据指纹和归属信息，进一步还可对数据的访问和操作记录存证，在本地数据出错时可以用数据指纹校验；对数据的所有权、使用情况有争议时，可以得到区块链网络的可信背书。

### （二）托管存储

托管存储意味着用户的数据离开了本地设备，通过特定的程序存到了其他存储空间。这个空间可以是其他设备，也可以是云端服务或者分布式文件系统。托管方案须满足以下特性：

1) 便利性：数据可以跨设备、跨网络使用，可生成多个备份，且能拓展用户的存储容量。为此，应关注数据导出、传输、恢复操作的体验和效率，降低相关成本开销，为用户提供友好便捷的体验。

2) 安全性：数据导出必须经过用户明确授权，传输过程应采用 SSL 等机制加密，在其他设备或云端服务存储时也进行高强度加密，托管方并不掌握数据的密钥，仅负责数据的存储而无法查看数据明文。有严密机制防止越权访问，仅有用户或已授权的第三方可访问数据。保留数据操作和数据访问的日志，结合区块链存证等机制支持追溯。

3) 可靠性：生成多个数据副本分片冗余存储，借助纠删码等技术，在优化存储效率的前提下保证数据可检测、可恢复，确保存储的可靠性。在支持多方共同托管的网络里，结合区块链和分布式存储方案，由多方共同管理数据，避免单点失效。

4) 可控性: 即使数据被托管，数据的控制权依旧是在用户手里，属主通过密钥、私钥、身份验证等机制，确保自己可以访问数据。用户可以灵活选择不同的托管方案、不同的托管服务者，自由地将数据从一个托管服务迁移到另一个托管服务，可以彻底删除在某一个托管服务上的数据，托管方无法禁止用户访问、迁移、删除数据。

#### （三）机构端存储

机构数据远比个人数据更加海量、复杂度更高，可能包含诸多用户数据。机构的数据存储必须符合行业规范满足安全性、可靠性、合规性等要求，建立完善的管理机制，把控相关角色、流程、系统，避免操作风险和道德风险，抵抗内外部攻击，此外应能支持第三方审计。

对大型和专业的机构来说，基础的数据存储已经相对成熟。在数字化时代，机构更需要关注的是对个人数据的采集和存储过程能否确保用户隐私保护。例如，在群体数据采集时，经过用户授权后，机构可采用差分隐私算法采集和处理用户数据，得到去除了个体特征的数据。另外，机构对个人数据的存储需要保证“可遗忘权”，在用户明确指示或数据生命期结束后，需要清理相关数据。

### 4.2.2 可信传输

在金融、政务、工业、个人应用等场景，业务通常由多个参与者共同完成，在业务流程中会互相交换数据，如发送信用证、合同、票据、扫描件等，其类型包括凭据或明文形式等。数据在个人和个人之间、个人和机构之间、机构和机构之间传输，面临参与者身份、数据合规性和完整性、网络通信等方面的风险和挑战。在传输过程中的主要风险及解决思路如下：

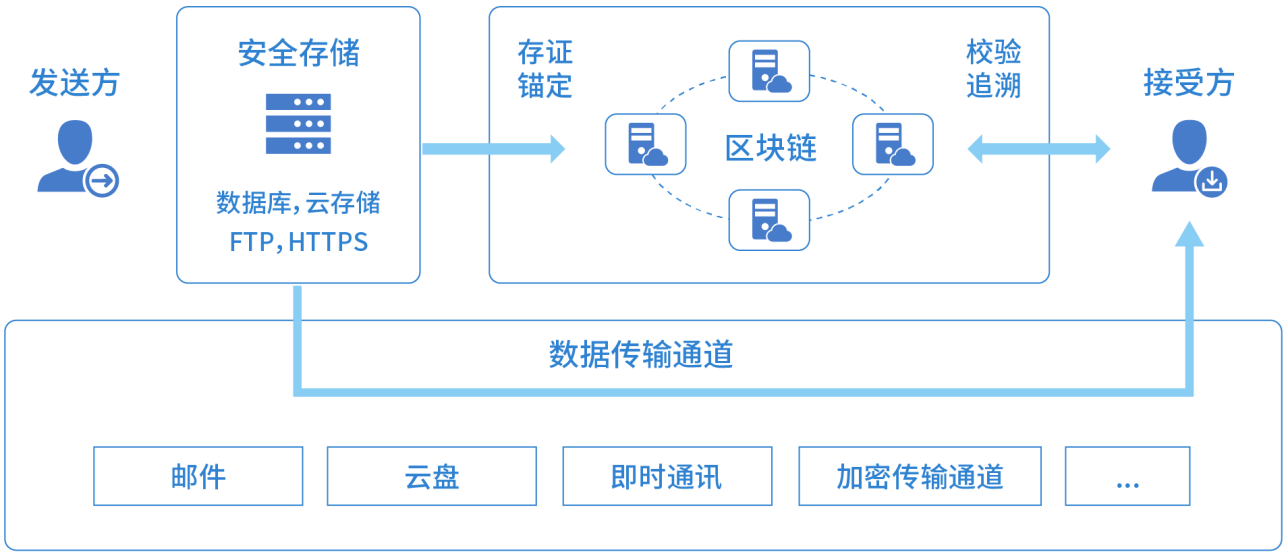



图 3 数据可信传输方案示意

### （一）参与者的身份是否可信？

首先，参与者包括数据拥有方、接收方以及参与到传输和验证的各环节参与者。可信传输方案要求参与者身份可知可验。微众银行推出了分布式数字身份解决方案 WeIdentity，结合区块链、公钥基础设施（PKI）和去中心化公钥基础设施（DPKI）体系，首先对参与者进行 KYC（Know Your Client）验证，然后在网络上为各方分配唯一的身份标识，该标识可以在链上进行分布式验证，在保护隐私的前提下，身份可控匿名，可信可审计。

### （二）发送者和接受者是否有权操作和访问数据？

对于个人数据，用户使用自己的证书或私钥进行数字签名，即宣称了自身对数据的所有权。对诸如身份证明、作品版权、医疗检查结果等需要权威机构背书的数据，则由权威机构和用户一起共同确权，用户对数据拥有合法的所有权和操作权，并将其生成凭据在区块链上进行存证。在传输时，发送方向接收方给出明确授权，通过数字签名向接收方表明身份，接收方才可访问数据，且通过数字签名验证数据的权属，相关授权记录可在区块链上存证和审计。微众银行的 WeIdentity 方案亦为凭据的存证和访问提供了可行可靠的解决方法。

### （三）是否可验证数据的正确性、完整性？

在传输的过程中如果出现干扰、通信错误或人为干预，会导致数据出错、丢失甚至被篡改。可信数据传输方案采用区块链构建可信网络，使得数据的发送方、接收方、各方确权授权信息以及数据本身的指纹，都在区块链上进行存证。基于 Hash 等数字指纹算法的单向性、可校验性和数字签名的不可篡改性，接收方在链上验证传输过程的各维度信息，确保其正确性、完整性，确信在传输过程未被篡改。

### （四）传输模式是否会导致数据泄露？

不恰当的传输操作会导致数据泄露。常见的通信模式包括点对点传播、接力式传播、广播式传播等，直接的点对点的传播是比较可控的；接力式传播代表在发送方和接收方之间增加了第三方环节；广播式传播代表更多第三方不必要地收到了数据。后两种方式中，即使是加密过的数据，依旧面临被暴力破解的风险。所以，将数据直接发往区块链上，用智能合约承载、借助交易广播、区块同步机制传递数据，并不是最推荐的方式。在可信传输方案中，数据的明文并不会在区块链网络上广播，仅将数据的数字指纹锚定在区块链上，如涉及明文，则采用链下方式点对点传输。即使是需要代理传播以应对复杂的网络环境，也需要审慎的选择路由、构建安全的通道，且和通道上的参与者签订相应的协议，对数据的转发和留存动作可记录可审计，如有问题可追责。

### （五）传输的实现方式是否安全？

数据可以由发送方通过灵活的方式传输到接收方，包括点对点网络、邮件、社交工具等，也可以托管到文件服务器、云盘等位置，授权接收方去获取。无论什么方式，均须采用高效的网络传输协议保证时效性，并做到全链路的物理隔离和加密保护，防止数据被窃取。发送方可采用“一次一密”的方式加密数据本身，并通过密码信封将授权信息和解密密钥交给接收方，接收方在限定时间内用指定的方式和发送方建立点对点的传输通道接受数据，或到托管服务器上拉取数据，托管服务器和读写方之间均支持加密通信，可防窃听、防攻击。

### （六）数据信任是否可以跨边界？

随着应用场景的丰富、参与者的增加，协作关系会变得更加庞杂，出现多层级、多应用、跨地域的生态圈，如国家级网络和省级网络分层分组，或者多个联盟链应用进行互联，信任的传递均突破了原有

边界，技术和治理模式迥异。针对这种情况，无论原来是否已经使用区块链技术的系统，都可以基于区块链的可信机制，在链上锚定数据和资产内容、追溯数据的授权和使用记录。然后引入跨链互联方案，将多个区块链网络连接在一起，运用简易支付验证（SPV）、零知识证明、哈希时间锁定、分布式事务控制等技术，实现数据可信证明、以及安全稳妥地完成事务。对于未接入区块链的独立领域，如提供汇率、天气等信息的数据源，可以借助“预言机机制”，打通链上链下通道，使链下经过确权、筛选的信息可以锚定到区块链上，作为可信数据使用。结合跨链、预言机等机制，可以使数据可信性得以跨边界传递。

### 4.2.3 协同生产

现实生产生活场景中，不同的参与主体可能存在复杂的竞争与合作关系。协同生产的目标在于使不同参与主体通过自组织的运行方式，在满足安全存储和可信传输的必要前提下，协调一致地发掘和转化数据要素价值，以实现数据价值回路闭环和正反馈效应。

根据参与主体类型不同，协同生产可能会涉及两大类数据价值回路：一是以个人数据为主的个人数据应用回路，二是以机构数据为主的机构间数据协同回路。

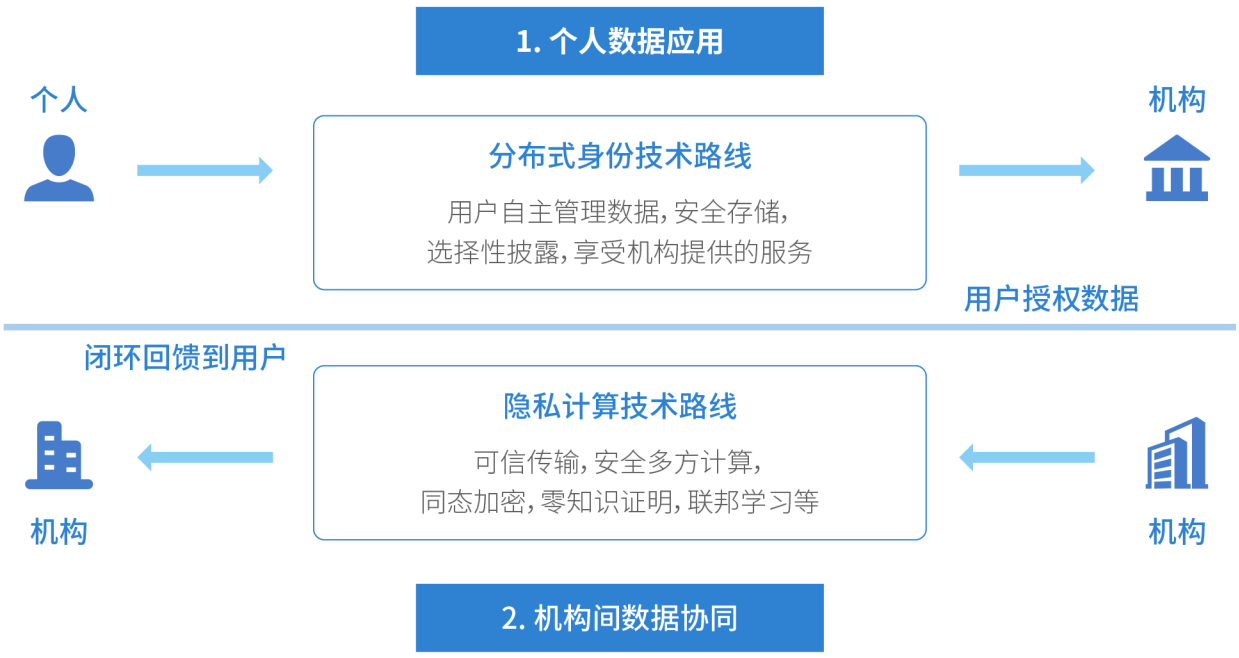



图 4 数据协同生产的两类价值回路

## （一）个人数据应用回路

个人数据应用回路的主要参与主体为个人和机构。其中个人一般作为数据生产主体，将产生的个人数据以可信的方式传递给机构；而机构作为数据服务主体，根据接收到的数据，为个人提供对应的服务。这个过程中，激励个人参与协同生产的重要条件是：能否有效保障个人对数据用途的掌控，控制数据使用所伴生的隐私风险？为此，派生出两类主要的协同模式：

A）托管模式：对于高度可信且不存在竞争关系的数据服务机构，个人将数据明文全权委托给数据服务机构，数据服务机构通过合同或信誉等担保，不将数据用于约定之外的用途；

B）自管模式：由于存在竞争关系或数据服务机构信誉不高，个人只对必要的数据进行选择性披露，并且通过数据密文化等手段，限定数据的用途，个人自主保存数据相关密钥，数据服务机构不具有对数据的完全掌控权。

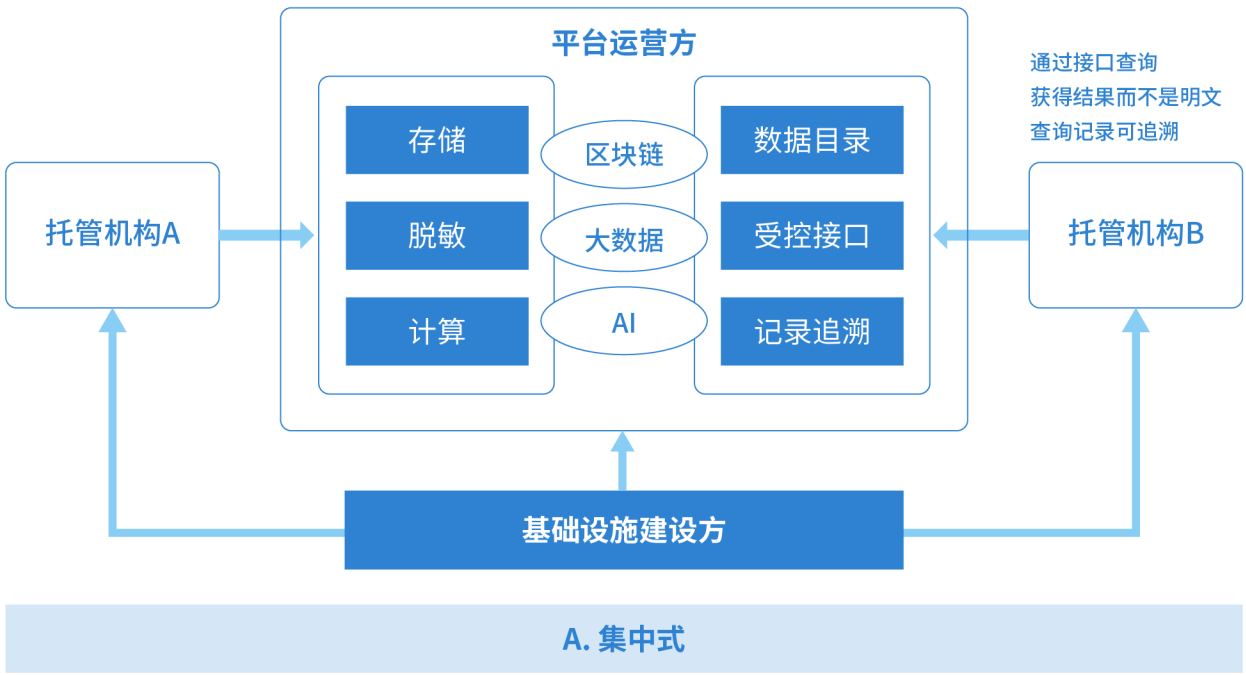

图 5 个人数据应用的两种协同模式

## （二）机构间数据协同回路

机构间数据协同回路的主要参与主体为机构和机构。机构既可以作为数据生产主体，也可以作为数据消费主体。不同于个人数据应用回路处理个体的数据，机构间数据协同回路所处理的数据对象通常是包含大量个体的批量数据，因此激励机构参与协同生产的重要条件是：能否有效保障机构对数据交换之后的合理价值分配，避免数据交换的合规风险？为此，派生出两类主要的协同模式：

A）集中式：存在被多数机构认可资质和实力的单一平台机构，其他机构将自身数据汇总到平台机构，并由平台机构统一进行数据融合并提供数据使用服务；

B）分布式：所有参与机构都能自主选择自身角色，参与到数据价值平台网络的组建中。网络中的任意机构可以和另一在网络中的机构，进行点对点隐私数据交易，或者通过安全多方计算、联邦学习等技术手段提供用途受控的数据 SaaS 服务，在保护自身商业秘密的前提下完成协作目标。该网络采用联盟委员会机制审查机构准入，按需组成不同联盟，以满足合规要求。



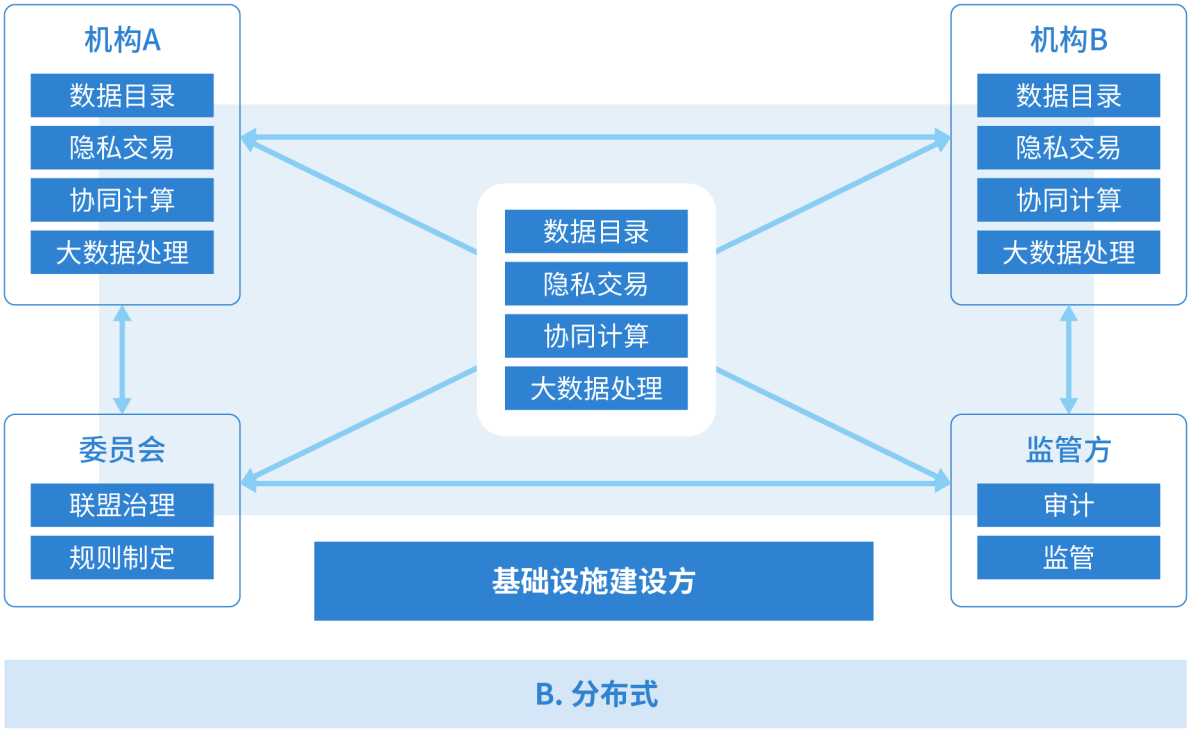



图 6 机构间数据协同的两种模式

值得注意的是，在未来 5G、物联网技术驱动的万物互联时代，良好的数据新基建势必要具备低延时、高吞吐的海量高维隐私数据处理能力。而目前隐私性强的通用计算技术，无论在理论上还是工程上，都存在显著的性能瓶颈，尚不足以支撑数据新基建的建设。因此，需要引入场景化的技术解决方案，对核心共性场景进行特制优化。

基于场景化这一核心理念，微众银行针对以上两大数据价值回路、四大主要协同模式，提供了丰富的场景式数据新基建解决方案。其中，两大核心技术能力源自自研的 FATE 和 WeDPR 方案。FATE 作为全球首个工业级联邦学习框架，支持广义线性模型、神经网络、决策树等齐全的机器学习建模算法，并实现了隐私数据模型训练和推演的集群部署；WeDPR 作为场景式数据隐私保护解决方案的先驱，对金融、政务、医疗、商业等数字化经济核心领域中的常用隐私数据协作流程提供了完备的即时可用高效解决方案模板，支持任意多数量参与主体之间的对等数据协作和零信任的安全机构计算节点部署，并对海量数据处理能力做了进一步升级。


图 7 微众银行数据协同生产的全栈技术架构

融合一系列前沿技术能力，根据数据从入到出的视角，微众银行为协同生产提供了自下而上、六个层次的全栈技术架构。

1) 第 1 层为**数据接入层**，具体包括对账户数据、行为明细、事件通知、统计数据、大数据集、文件等一系列源数据的高性能访问和安全存储；

2) 第 2 层为**数据预处理层**，具体包括抽取、脱敏、标准化、结构化、分类、聚类等高频数据预处理模块；

3) 第 3 层为**数据计算层**，具体包括隐私计算密码库、联邦学习、安全多方计算等多方数据协同计算方案；

4) 第 4 层为**数据管理层**，具体包括分布式身份、数据目录、接入控制、授权管理、可信传输管理、安全态势感知等关键数据安全管理平台；

5) 第 5 层为**数据调用层**，具体包括区块链应用服务、协同训练服务、统一数据接口调用服务、分布式存储服务等数据调用入口；

6) 第 6 层为**数据应用层**，具体包括营销定价、信用评级、排名定序、投票决策、存证溯源等多方数据融合应用。

从最底层的数据接入到最顶层的数据调用和实际应用，该解决方案全方位整合了计算资源和数据资源，为数据价值的转换和发掘提供了全生命周期的安全隐私保障，为全面解放数据生产力提供了强有力的技术基础设施。

## 4.3 现有方案和实践案例

### 4.3.1 安全存储：开源的一站式金融级数据管理平台

当金融与互联网相结合之后，银行 IT 架构要承受智能金融时代的数据洪峰，数据的安全存储和高效计算非常关键。可靠的数据存储和管理方案须满足多个需求，比如应能兼容不同来源和结构的数据，能审核、管理异质来源的数据质量，能支持便捷分析数据甚至支持机器学习，能做好数据生命周期管理和资源成本管控，能支持数据隐私安全保护等。

为满足这些需求，微众银行开发了名为 WeDataSphere 的开源一站式金融级数据管理平台套件，涵盖了数据管理开发全流程。该平台套件的基础功能层基于 Hadoop、Spark、Hbase 等各种开源组件，构建可靠的基础计算存储数据交换能力及强大的机器学习能力。在基础功能层之上，套件包含平台工具、数据工具、应用工具三大层次。其中，平台工具包含平台门户、数据计算中间件和运营管理系统，数据工具包含数据地图、数据脱敏工具、数据质量工具和跨 Hadoop 集群的数据传输工具，应用工具则包含开发探索工具、图形化工作流调度系统、数据展现 BI 工具和机器学习支持系统。这三大层次工具关注用户各类功能需求的工具实现，形成了完整的大数据平台技术体系，有力地支持了数据安全存储和计算。


图 8 WeDataSphere 框架图

### 4.3.2 可信传输：跨政务机构个人数据可信流通

数据难流通、安全难保障是当前智慧政务信息建设过程中面临的主要问题。首先，政务数据信息化程度不一，互联互通程度较低，“信息孤岛”现象普遍存在，使得不少政务服务仍需依赖人工操作，大大影响了政务协同效率，难以真正实现“最多跑一次”。再者，政务数据对数据安全和隐私保护具有很高要求，可信授权机制的缺乏、数据传输过程中的安全问题、数据协作的不可追溯，都会提高多部门数据融合的风险。尤其是跨机构的个人数据流通仍然存在不少痛点，例如，公民办理各项事务时经常需要出示自己的纸质证书或证明。但接收机构很难验证这些个人材料的真伪，通过人工或第三方验证的方式往往耗时长、效率低，还可能因道德风险或操作风险导致用户隐私信息泄露。因此，不难发现，机构间信息壁垒、数据真实性存疑、用户隐私保护难等问题，广泛存在于跨政务机构个人数据流通的各类场景中。

针对上述难题，微众银行推出面向智慧政务的一系列数据要素解决方案，助力政务数据可信流通。通过分布式数字身份方案 WeIdentity 为用户构建统一的链上数字身份（WeID），每个参与方都可以在不依赖中心机构或者第三方的情况下进行 WeID 的身份验证，从而为多部门间的数据流通提供了更安全的分布式身份认证管理体系。在数据传输方面，采用高效的网络传输协议保证传

输时效，并加强全链路的物理隔离和加密保护，以确保传输通道安全，同时，利用哈希等数字指纹算法的单向性、可校验性和数字签名的不可篡改性，保障政务数据在可信流通过程中的完整性和安全性。此外，用户可以通过细粒度授权机制实现信息的选择性披露或最小化披露，在海量数据场景下，还可结合联邦学习和安全多方计算，实现“数据可用不可见”，更好地保护用户隐私。在实践中，区块链电子证照平台可实现跨部门、跨区域、跨层级的电子证照可信流通互认，方便用户在保护隐私的前提下一键授权证照信息，真正实现“指尖办事”。


图 9 智慧政务数据协同示意图

2019 年，澳门政府与微众银行签订合作协议，以区块链等创新技术作为主要支点，推进澳门特区的智慧城市建设。澳门政府基于微众银行自主研发的 WeIdentity 开源方案推出了证书电子化项目，实现安全高效的跨机构身份标识和数据流通：首先，通过分布式多中心的身份可信协议，为每个用户生成唯一的 WeID 身份标识；第二，结合可验证数字凭证技术和加密技术，把线下的证明类文件转换成真实可验证的电子化数据凭证 credential，证书发行方对 credential 的信息摘要进行加密、签名后上链存储，证书接收方在链上验证数据的真实性、完整性，确信在传输过程未被篡改；第三，让用户参与到数据交换的过程中，由用户定向授权并自主发起数据传输，发送方、接收方都需要在表明身份并得到明确授权的前提下才能进行相关操作，授权及操作记录亦可在区块链上存证。

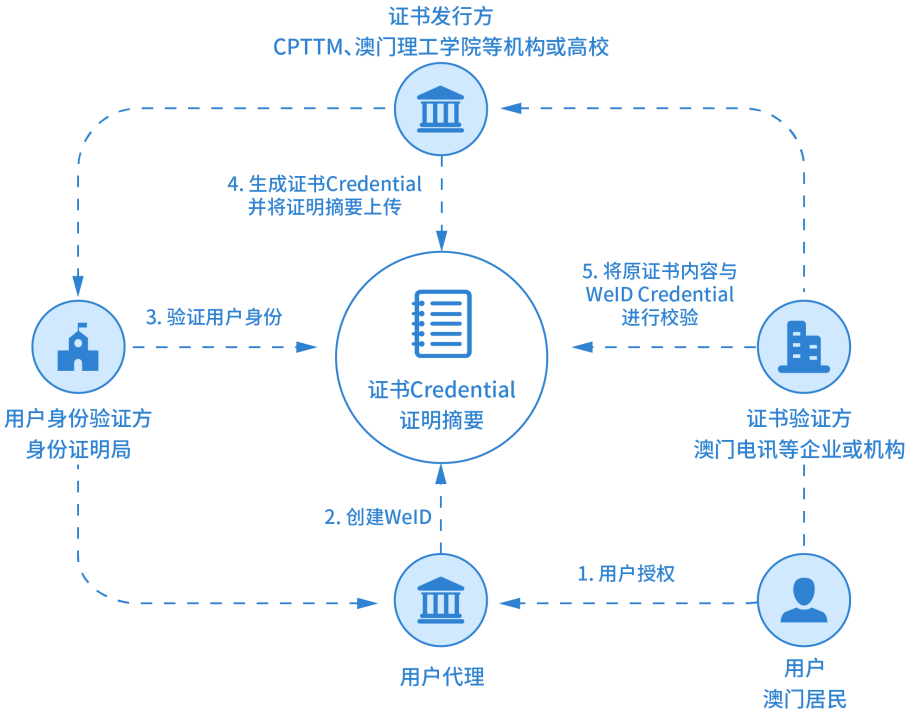



图 10 电子证书发行与验证流程示意图

### 4.3.3 可信传输：粤澳两地健康码跨境互认

2020 年 5 月，澳门特区和广东省启动了 " 粤康码 " 与 " 澳门健康码 " 互认系统。为加快恢复内地与澳门人员正常往来工作，7 月 15 日起，两地居民持粤康码通关凭证以及有效核酸检测阴性结果即可正常通关，无需隔离 14 天。截至 2020 年 11 月，持粤康码通关凭证通关累计超 2800 万人次。

粤澳两地健康码互认，涉及两大难题。首先，是健康码生成、使用过程中的用户信息安全和隐私保护问题，其安全隐私标准应符合两地各自用户隐私保护的相关法规要求；其次，由于居民的个人信息及核酸检测信息等只有本地权威机构有能力验证，而澳门和内地相关机构需在用户数据不直接传输和交换的前提下，验证用户提交信息的真实有效性，搭建跨地区的数据真实性核验通道。

粤澳两地利用区块链技术优势解决了这个难题。系统采用国产开源的底层框架 FISCO BCOS，并使用 WeIdentity 将健康码相关信息转化为加密的可验证数字凭证，两地机构在后台不互联的情况下依然可以验证信息的真实有效性。当用户需要跨境验证健康码时，不需要在多个平台重复填写信息，系统在获得授权后将自动为用户转码。

粤澳健康码互转互认及粤康码“通关凭证”的应用，首次实现了粤澳防疫数据互信互认，这种由用户自行驱动的数据提交和核验机制，有效促进数据要素在用户知情、授权和主导下的跨地区流通。而且，因技术具备可扩展性强的特点，除了粤澳两地之外，其他国家及地区也可以非常容易并快速地加入其中。

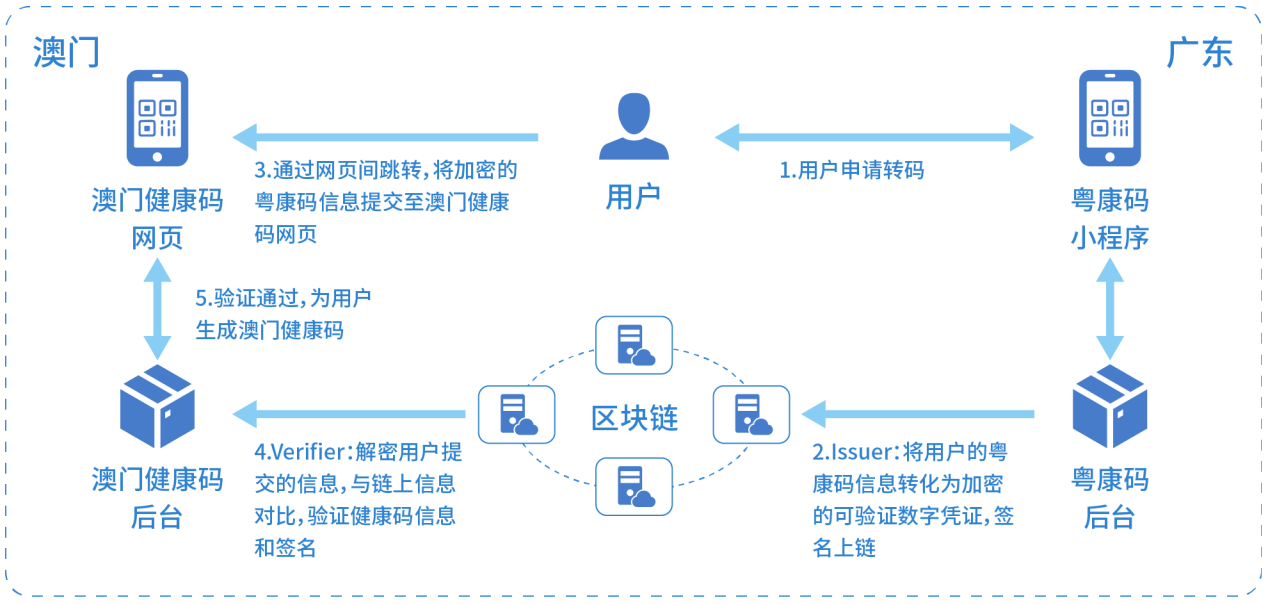



图 11 转码过关业务流程图（以粤康码转换为澳门健康码为例）

### 4.3.4 可信传输：医疗处方线上流转

处方流转是医药分开综合改革中的一项重要内容，政府希望通过医保对接、处方信息互联互通等方式，形成医生诊断病情开具处方、患者凭处方到药店购药的模式，对医院药房形成实质上的“分流”。处方流转本质是用户授权和主导下医疗机构与药店间的数据流通与协同。电子处方应该具有可消耗、可追溯等特性，以提高药品流通效率，并避免处方被重复使用可能带来的法律风险。同时，电子处方流转对数据授权和可信传输提出了更高的要求，医疗机构和药店既要确保处方真实不被篡改，也需要在发生纠纷时，自证处方流转是在用户授权之下合法合规进行的。

微众银行为此提供了底层区块链开源技术，可进一步开发电子处方流转方案：第一步，结合 WeIdentity 分布式数字身份解决方案和 KYC 等身份验证手段，做到医疗机构、医生、药房、用户等参与方身份的可信、可控；第二步，处方生成和传输阶段，由医生对电子处方进行私钥签名，生成处方凭证，由用户发起处方流转，生成授权凭证，相关凭证均上链存储，需要说明的是，患者的处方数据仍存储在医疗机构的数据库中，电子处方的明文并不会在区块链网络上广播；第三步，处方使用阶段，药店在收到处方后，需要验证 WeID 和凭证信息，验证通过后为用户提供药品，并记录电子处方的使用情况。不难发现，方案很好地解决了医疗机构、药店各自独立的业务状态带来的信息不互通问题，保障了电子处方的真实可验证，同时，相关操作都将存储在区块链上，全过程可追溯，利于监督审计。

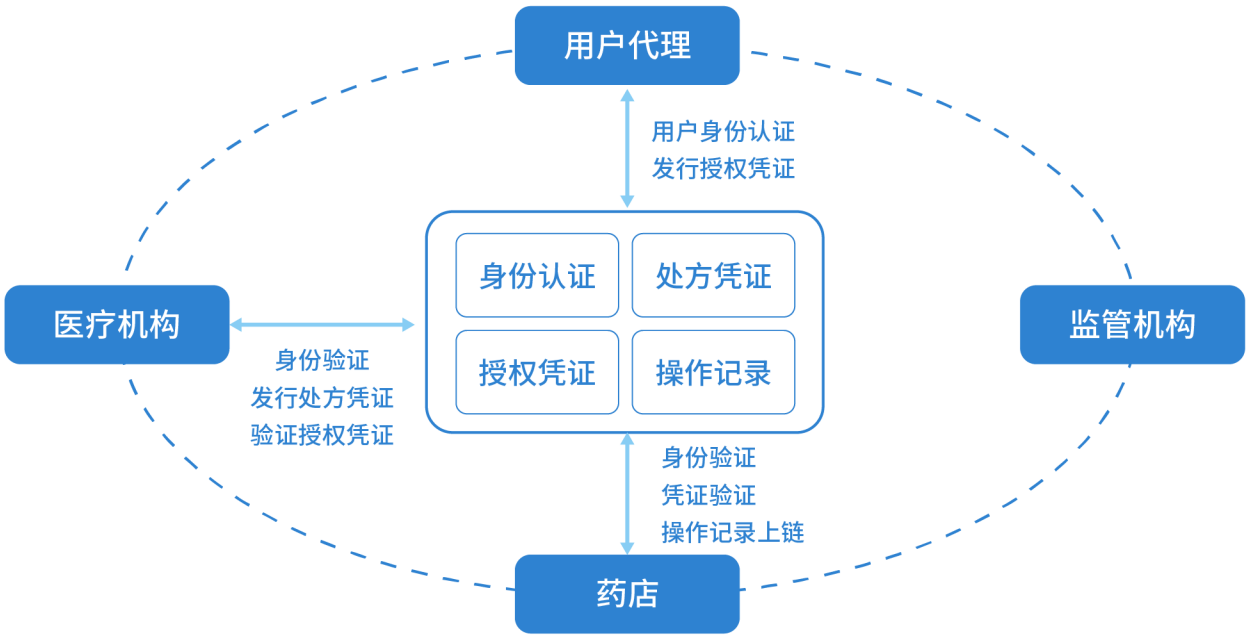



图 12 电子处方流转示意图

### 4.3.5 协同生产：绿色出行普惠平台激励碳减排

“推动绿色低碳发展、解决污染防治”是新时代国家攻坚战和重要发展战略之一。政府鼓励低碳出行，交通运输部、生态环境部等部委分别提出绿色出行相关建议。推动绿色低碳发展，培养公民绿色生活方式成为社会各界共同努力的方向。然而，如何连接、协同和激励参与各方，如何通过市场化激励手段鼓励大众绿色出行，仍是社会文明治理中需要关注的痛点。

在此情况下，微众银行推出基于区块链技术的社会治理框架“善度”，将参与方分为发行者、分发者、赞助者、监管者和终端用户等共七大善度角色，角色间有效分工，高效协作，借助 5G、物联网、区块链等技术对善行进行全面的度量、激励、跟踪和监督，在兼顾绿色创新和风险防控的前提下通过开放合作，实现绿色生态闭环。绿色出行普惠平台便是在这样的机制下应运而生，采用 WeIdentity 分布式数字身份解决方案管理注册用户的链上身份，利用车联网设备对用户车辆停驶减排行为进行精准度量，将碳减排量转化为碳积分后，一并分发至用户的链上账户中，可供用户兑换商品、服务和公益活动，实现正向反馈，价值共赢，且全流程公开透明，相关记录可随时追溯查证。

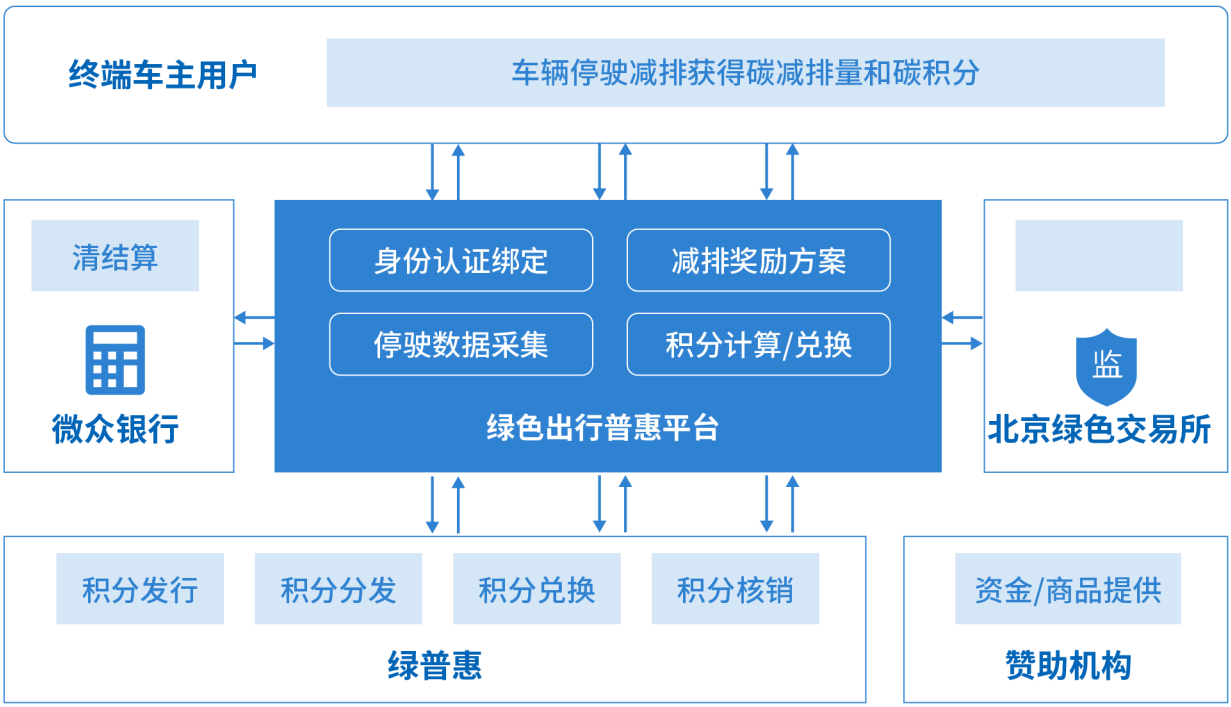



图 13 绿色出行普惠平台示意图

### 4.3.6 协同生产：联合营销中的隐私保护

广告主（如微众银行的某个信贷产品）在营销过程中，常常需要在媒体广告平台（如腾讯）投放产品广告，这时往往面临着隐私保护的难题。首先，传统广告技术只能根据展现、点击、到达等前端环节数据优化，但没有后端数据。为了优化广告投放的效果，广告主需要向广告平台回传尽可能多的后端用户转化数据，以优化广告策略。但是，由于用户隐私保护问题，广告主不能直接将用户数据分享给广告平台，而且数据回传过程也存在着泄露可能性。其次，广告主本身只有用户的主营业务单一数据，缺乏丰富多样的其他数据以优化用户画像，精准投放产品，但由于隐私法规限制和商业考虑，很多第三方机构也不愿意将自有数据与广告主共享。

面对这样的难题，隐私计算就能发挥作用，促进各方之间的协同生产。在业内，微众银行是“联邦学习”（federated learning）技术的重要领军者，将该技术引入智能营销，形成“联邦广告”和“联邦推荐”方案，在保障隐私安全的同时，大幅提升在线广告拉新和存量客户促活的效率。基于联邦广告方案，广告主可以将广告点击和转化数据利用“差分隐私”（differential privacy）进行混合加噪，然后再做数据加密、与广告平台进行加密条件下的联合建模。在这个过程中，广告主的转化数据、广告平台的用户标签都留在原地，而广告模型的效果得到优化，从而提升广告投放的效率。基于联邦推荐方案，广告主与第三方数据源可以在各自数据不出本地的条件下一起协同构建推荐系统，通过参与方之间交换加密参数的方式避免原始训练数据泄露和传输。在保证隐私的前提下，联邦广告和联邦推荐实现了高效贷款广告投放、理财产品推荐，广告效率提高 20%，理财产品推荐效率提升 30%，有效地实现了用户拉新促活的目标。

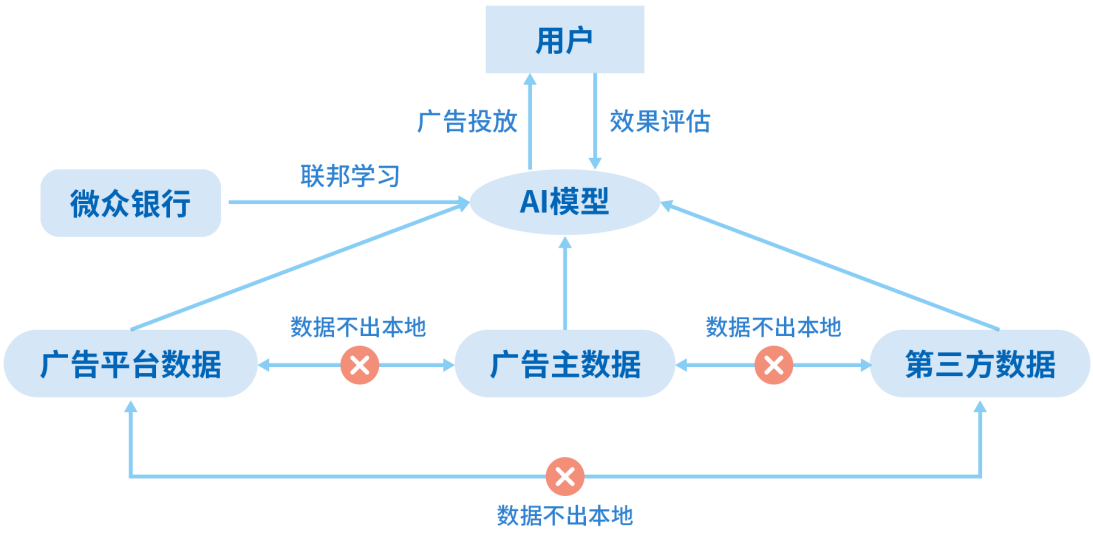



图 14 联邦广告与联邦推荐的应用流程示意

# 结语和展望

生产要素释放价值需要有可界定的产权，需要满足价值可存储、可评估和可流通的特性。由于数据要素具有易复制性、非竞争和非排他性、分散性、价值聚合性和价值认知多样性等特点，难以满足生产要素价值释放的条件。为了最大化发挥数据要素的价值，我们需要基于人工智能、区块链、云计算、大数据等数字技术的“数据新基建”解决方案。“新基建”的核心功能在于实现数据的安全存储、可信传输、协同生产，从而解放数据生产力。


图 15 数据要素新基建

本白皮书的“数据新基建”解决方案致力于解决数据在实际生产中可能面临的几个经典矛盾——如数据所有者与控制者、使用者之间收益分配的权利矛盾，满足个人隐私需求与因公共利益进行中心化监管之间的合规矛盾，单方保护数据与多方需要数据之间的透明度矛盾等。**这些矛盾归根究底是关于数据的“信任”问题**，即如何让数据的全生命周期透明可信，如何相信应有的权益能得到落实保护，如何设计激励相容的机制让各方彼此信任地分享使用数据。本白皮书解决方案较好地解决了这些信任问题。

**第一，方案满足了多主体共建信任的需求，实现了必要的透明可信**。这种透明可信来自于对技术的信任和对关系的信任。通过底层框架和应用软件的全面开源，解决方案打消了藏在黑匣里的隐患，让问题更容易暴露和修复，做到了技术全面可信。通过基于区块链技术的多个解决方案，数据不再被中心化地掌控，而是分布在不同主体手上，同样能安全高效平等地满足多方沟通协作的需求，做到了分布式信任。

**第二，方案满足了多层次表达信任的需求，实现了必要的权益保护**。在实际生产过程中，由于主体不同、关系不同、场景不同、数据不同，各方对于隐私保护、安全防护等问题的要求也不一样，人们对信任

的需求表达就会呈现多个层次，有丰富细致差别。为此，本白皮书的方案提出了基于场景式的隐私保护解决方案，根据场景需求的差异性来极速适配相应方案。同时，基于方案公开可验证的特性，用户可以根据需求自主验证方案效果，进行相关风险审计，确定自己是否真正控制和保护了应有的权益。

**第三，方案满足了多环节传递信任的需求，实现了必要的激励相容**。数据的使用是一个社会大生产过程，需经过多个环节，但多环节之间的信任并不能凭空产生，只能在不同人和组织之间传递，逐渐形成。为此，要设计合适的方法来保障各方行为与其利益一致，信任能够延续，不会单方面破坏传递的链条。在传递过程中，身份是最基础的数据，是其他数据权利的依托。本白皮书的方案通过打造分布式数字身份基础设施，实现了对身份数据的标识、证明、授权和验证等全流程管控，让身份及其之上的凭证等其他数据都能被对方信任。基于此，解决方案可度量、记录、查证和共享生产环节的数据，每一个数据、每一个参与者都被真实记录，有助于进一步设计和搭建多方信任的激励机制，更好地保证信任的传递与产生。

如今的互联网世界已经迈入了数据爆炸与隐私泛滥的时代，迈入了竞争白热与开放协同的时代。一方面越来越多的主体掌握了海量数据，试图独占更多数据，试图从数据富矿中挖掘利用价值；另一方面它们面临着向伙伴开放分享数据以发挥协同价值的机会，又面临着数据安全和隐私保护愈加健全的约束。因此，解决这些矛盾的数据要素基础设施变得极为重要，将是数据发挥社会生产力的基石。

我们相信，未来的数据要素基础设施一定要满足四个条件：**具备可管控的底层技术**，使基础设施从源头上被信任；**具备可验证的安全性和稳定性**，使基础设施真正可靠；**具备高效率的计算能力**，使基础设施能有效支撑数据发挥作用；**具备健全完备的数据流通商业模型和激励机制**，使基础设施可持续运转。本白皮书所提出的解决方案是朝这些方向迈出的第一步，前方的道路虽远而光明，一个数据和谐涌动的美好数字社会正在招手。

## 微众银行区块链

微众银行高度重视新兴技术的研究和探索，自 2015 年开展联盟链领域技术研究和应用实践以来，已研发一整套含括底层技术、中间件、分布式数字身份、数据隐私保护、跨链、消息协作、数据治理等在内的技术方案支撑产业应用，并持续为数字经济时代提供数据的安全存储、可信传输、协同生产等一系列数字化基础设施，释放数据价值，解放数据生产力。

在实现区块链关键核心技术自主研发的同时，微众银行自 2017 年起主动面向全球开源，至今已开源七大技术方案，牵头多方共建起最大最活跃的国产开源联盟链生态圈。生态圈内汇集 4 万余名社区用户、2000 多家企业及机构共建区块链产业生态，数百应用项目基于 FISCO BCOS 研发，其中超 120 个应用已在生产环境中稳定运行。

## 金融科技·微洞察

“金融科技·微洞察”是微众银行运营的金融科技研究品牌，聚焦国内外金融科技领域的技术发展、标准制定及产业应用，把握当下金融科技热点话题与政策动向，洞察未来领先的金融形态和商业模式。

微众银行作为国内首家互联网银行，自 2014 年成立之初即将“科技、普惠、连接”作为银行的三大发展愿景，将积极运用科技创新探索普惠金融新模式、新业态作为银行重要的发展方向，致力于为普罗大众、微小企业提供差异化、有特色、优质便捷的金融服务。自立行至今，微众银行在金融科技“ABCD”（人工智能、区块链、云计算、大数据）等四大领域积极探索，2017 年即已成为国内首家获评“国家级高新技术企业”的商业银行，截至 2019 年末共申请国家及国际专利数超过 1000 余件，拥有自身所有重要业务和技术系统的知识产权，有效实现了银行业信息化安全可控的战略目标。

# 免责声明

在任何情况下，本白皮书中的信息或所表述的意见并不构成对任何人的投资建议，本白皮书所载的资料、工具、意见及推测只作参考之用，并非作为或被视为出售或购买证券或其他投资标的邀请或向人作出邀请。在任何情况下，白皮书的编著机构不对任何人因使用本白皮书中的任何内容所引致的任何损失负任何责任。



所载资料仅供一般参考用，并非针对任何个人或团体的个别情况而提供。虽然我们已致力提供准确和及时的资料，但我们不能保证这些资料在阁下收取时或日后仍然准确。任何人士不应在没有详细考虑相关的情况及获取适当的专业意见下依据所载资料行事。

## 联合出品



## 报告出品人

马智涛

## 报告统筹

范瑞彬　姚辉亚

## 报告作者

张开翔　徐 磊　李辉忠　严 强　韩丹　陈明艳

## 美术编辑

黄秋云

微众银行区块链

金融科技·微洞察