

Banyan Network



Banyan Network

榕树网络

全球首个数据融合价值链网络

目录

1. 摘要.....	3
2. 缩略语	4
3. 项目背景	4
4. 业务架构	7
4.1 价值网络的构建环节	7
4.2 价值网络的支撑应用	10
5. 技术基础	11
6. 理论体系	12
6.1 数据价值三定律与其内含参数	12
6.2 数据质量参数	12
6.3 数据流动性参数	12
6.4 数据融合度参数	13
6.5 数据时效性参数	13
6.6 数据价值函数	13
6.7 相关参数的定义决策机制	13
7. 通证规则	14
7.1 通证介绍	14
7.2 通证场景	14
7.3 流通模型	14
7.4 持有与决策参与	16
8. 治理机制	16
8.1 底层运行	16
8.2 智能合约	16
8.3 共识机制	16
8.4 三方模式	17
9. 技术应用	17
9.1 区块链应用	17
9.2 人工智能应用	18
10. 感谢	18
参考文献	19

1. 摘要

当前，数据行业一方面困于安全管控日趋严格的政策态势，一方面限于信任体系缺失的竞合环境，数据孤岛变相加剧，数据价值难以释放^{[1][2]}。区块链技术的可信、可追溯、不可篡改和去中心化为重构数据价值链^{[3][4][5]}，创造数据领域各参与方诉求和利益的最大化提供了解决路径，并撬动万亿级大数据市场。

榕树网络（Banyan Network）是一个以区块链技术为特征的数据融合价值链网络（Data Fusion Value Chain Network，DVN），也是全球首个由数据领域参与者共筹共建的分布式数字经济生态体系。

DVN以“共生、共利、共治、共荣”为价值基准，专注区块链可信数据连接、第三方数据融合与治理、数据应用公开市场开发，通过建立标准、提供通道、确保安全，建立以价值为纽带的数据利益良性交互网络。DVN设有丰富使用场景的可延展、高成长通证体系^[6]，可在最广范围激发数据集聚，更高速度促进数据流动，从而提高可用性和融合价值，为数据提供者、加工者、使用者等所有广义贡献方提供价值倍增途径，创造无限价值分享空间。DVN采用先进的数据匹配、校正算法，精细的贡献、奖励机制，优先以超级数据源伙伴、大品牌客户等作为数据价值构建的联合体基础，适时为企业级数据源、个体数据源对等加入DVN生态提供便捷通道。

DVN涵盖从数据采集、清洗、融合、应用到服务的完整数据业务链，基于“数据价值三定律”理论和“三方模式”机制，并积极引入高性能区块链（High-performance Blockchain，HPB）、去中心化通

证交易撮合协议（LOOPRING）等最新技术方案，确保整个网络高效运行及逐步扩大。

2. 缩略语

BBN¹（Big Banyan Network）：大榕树网络

DVC（Data Fusion Value Chain）：数据融合价值链

DVN（Data Fusion Value Chain Network）：数据融合价值链网络

DPOS（Delegated Proof of Stake）：委任权益证明

IRC（Introducing Rewards Contract）：数据引入奖励合约

CRC（Cleaning Rewards Contract）：数据清理奖励合约

APC（Apply Pay Contract）：数据使用支付合约

TC（Transaction Contract）：撮合合约

API（Application Programming Interface）：应用程序编程接口

HPB（High-performance Blockchain）：高性能区块链

3. 项目背景

大数据已经深入，区块链正在到来。大数据和区块链共同的关键词是分布式，共同代表着未来由技术权威垄断到去中心化的本质转变。

数据产业是具有巨量市场潜力的超级产业集群。就规模看，数据基础服务、自动化交易等在百亿级；精准营销/泛营销、信息核

¹ BBN 为本项目通证名称

验、企业/政府/个人/市场咨询等在千亿级；而具体行业应用市场规模则达到万亿级。如数据应用典型产品之一“程序化营销”，2016年中国市场规模约115亿元，2017年达到250亿元，预计至2020年将保持20%年以上复合增长率；数据应用典型行业之一“数字营销”，2016年中国市场规模约2400亿元，2017年达到约2800亿元，预计至2020年将保持15%以上年复合增长率；行业总体方面，2016年中国大数据相关产品和服务业务收入约3400亿元，预计到2020年将突破1万亿元。

数据产业也面临直观显性的发展瓶颈。数据源层面，数据获取门槛较高，数据不完整且割裂封闭，每个数据源只能提供部分可用信息；数据误差大，缺乏多重数据源校正，精确度难以保证。数据产品层面，产品化程度低，接口无标准且接入复杂；解决问题方式单一，效果无法衡量。数据安全层面，欠缺合规体系，数据来源难追溯，前置授权难获知。

在行业经营参与者视角，数据的体量、质量、活性是大数据价值的原生力，而数据的连接、聚合运用则是大数据价值的催化剂。当前，数据孤岛现实未解决，数据安全管控更严格，这种双重担忧反映了一个长期困扰行业的严肃话题——“数据信任体系”尚未且难于建立^[7]。区块链技术凭借可信、可追溯、不可篡改和价值共建正在各领域加速应用，构建基于区块链的数据融合价值网络将撬动并重构万亿级大数据市场。

从行业瓶颈解决者视角，数据的开放、连接和价值流通，不仅需要法律、制度、政策保障，更需要一个刚性解决方案，一个重要而可靠的载体。在日益强调透明性、安全性的大数据应用场景下，区块链技术通过多节点共同参与数据的计算和记录，并相互验证其有效性，从而让数据更值得信任、更有价值，其持续增长、按序整理成区块的链式数据结构将重塑数据行业生态，成为数字经济时代的未来基石。

对于数据融合价值的实现，区块链不仅可以帮助数据建立价值并传递、转移价值，也建立一种可信的数据连接基础，未来所有数据行业参与者都将被连接成网络，通过契约关系形成智能合约或智能资产，区块链成为数据价值链中的一个万能账本。在这一基础上，数据的确权、数据资源的资产化转变将依托这一支撑技术，正好契合了分享经济发展的一种模式。

基于区块链的数据融合价值链网络就是为了彻底解决上述问题。DVN 去中心化的核心优势之一是避免任何数据被非合规复制、截流、沉淀甚至修改，数据价值不存在被盗用弱化的可能，从而极大降低数据源对 DVN 的信任成本。另一个优势是突破地缘和时间限制，提高各参与方的透明性、延展性和效率，促使产生更大的数据流动性和更高的数据价值。

在此之上，以 DVN 数据为基础的人工智能产业将获得新生。DVN 将作为人工智能的数据供应基础设施，满足其应用开发的几乎所有数据源供给、数据服务采购以及高性能的分布式数据处理能力，

为人工智能的未来提供充足的燃料。

4. 业务架构

商用化的数据供给和数据需求有较大的差距，数据普遍缺乏流动和整合，甚至孤立、不流动，或者因未经清洗、加工而无法发挥价值。DVN 提供面向数据行业所有参与者的融合价值网络框架，提供贯通数据需求与数据供给的系统化构建方案。DVN 支持面向数据价值生命周期各环节的增值产品、服务开发，为各方更积极、主动参与 DVN 构建、优化提供各类基础或个性支持，以更高效促进价值增益。

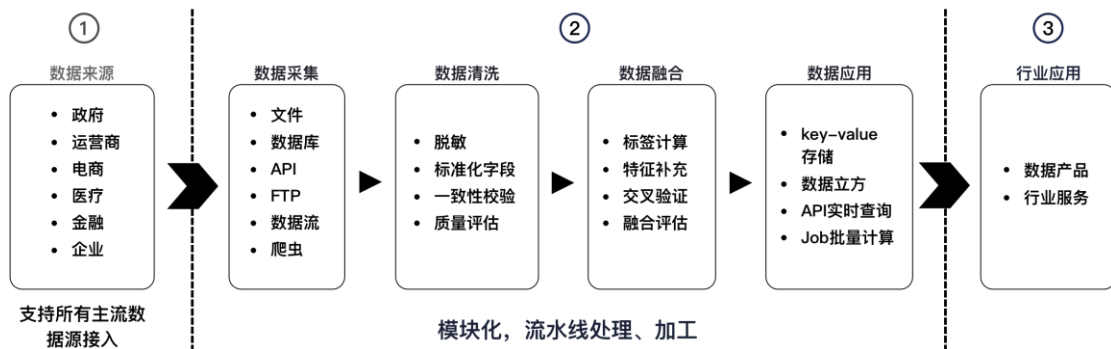
双螺旋上升引擎：DVN 数据价值链路具有“数据流”、“价值流”双螺旋结构。“数据流”促进数据汇聚、连接、打通、耦合，“价值流”促进价值挖掘、评估、记录、分配，“横截面”形成连接数据需求与数据供给的闭环结构，“纵切面”形成促进价值催生与价值加速的指数上升膨胀结构，最终形成了一个由 DVN 所有参与者的数据贡献、价值回馈构成的双螺旋上升的数据价值大爆炸驱动引擎。

4.1 价值网络的构建环节

面向所有数据行业参与者需求，构建数据融合价值链网络 DVN。通过业务流上的价值共建冲破传统“数据来源”到“行业应用”的距离，将传统的数据采集、清洗、融合、应用环节进行分布式的节点化连接、融合（见下图），形成一条条数据融合价值链（DVC，

Data Fusion Value Chain)，多个 DVC 交互交叠共同构成线条密集但无边界的数据融合价值链网络（DVN）。DVN 中，各方纳入到统一的数据价值生态中，依托区块链技术及其配套机制保证各参与方利益。DVN 中，各方归并到统一的数据安全体系内，通过安全多方计算框架（SMC，Secure Multi-Party Computation）的协议安全、零知识证明等确保数据、记录、流程的安全合规。

数据融合价值链



4.1.1 多元异构采集 DVN-Collection

支持多元数据快速便捷接入，解决不同数据源间数据存储方式、格式差别造成的异构数据接入难题。

支持的数据源：政府、金融、运营商、电商、互联网、垂直行业、企业、个体等。

支持的数据采集方式：文件、数据库（关系型数据库，NoSQL 数据库，分布式数据库等）、API、流数据、FTP、爬虫等。

支持的数据解析格式：文本、JSON、XML、非结构化数据、自定义解析器等。

4.1.2 多级串联清洗 DVN-Cleaning

对数据错误值、缺失值、异常值、可疑数据等进行多级“清理”，通过字段串联形成格式化数据，解决数据预处理和质量提升难题。

除狭义数据清洗外，还包括数据脱敏、数据一致性检查、数据质量评估等；同时设置虚拟数据治理岗位，吸收组织或个人参与到数据质量提升过程中。

4.1.3 多源升维融合 DVN-Fusion

对多个数据源进行融合打通，实现跨数据源标签计算和升维，解决数据多源整合、融合应用难题。

数据融合过程中，用一个数据源的特征补充完善另一个数据源，并进行多源数据的交叉验证，更全面地刻画数据的真实属性；可评估数据源的数据质量，对于数据融合效果进行评估。将融合后的数据升维成多维度的数据立方体（Data Cube）以备使用。

4.1.4 多态并发应用 DVN-DataApp

识别多种数据结果形式，自动封装成不同形态的数据应用，同步并发提供服务，解决数据最终使用难题。

计算后的标签画像数据以 Key-Value 形式保存，系统将其导入到分布式数据库中，对外提供 API 实时查询。通过数据立方体对外提供 Job 批量计算服务。

4.2 价值网络的支撑应用

4.2.1 榕树评分 **Banyan Score**

DVN 基础评分系统，具有标准、指标体系、算法、发布机制等，针对数据源、数据产品和服务提供公开评分服务。

4.2.2 榕树虫洞 **Banyan Wormhole**

DVN 统一清洗系统，具有数据分类分拣、缺陷检测、优化算法、补全规则等，快速、高效进行数据标准化封装、可用性提升服务。

4.2.3 榕树标签库 **Banyan TagLib**

DVN 标准标签体系，具有数据标签打通、匹配、关联机制等，提供基于融合数据的跨源、多维组合、联合调用标签使用服务。

4.2.4 榕树市场 **BanyanMaket**

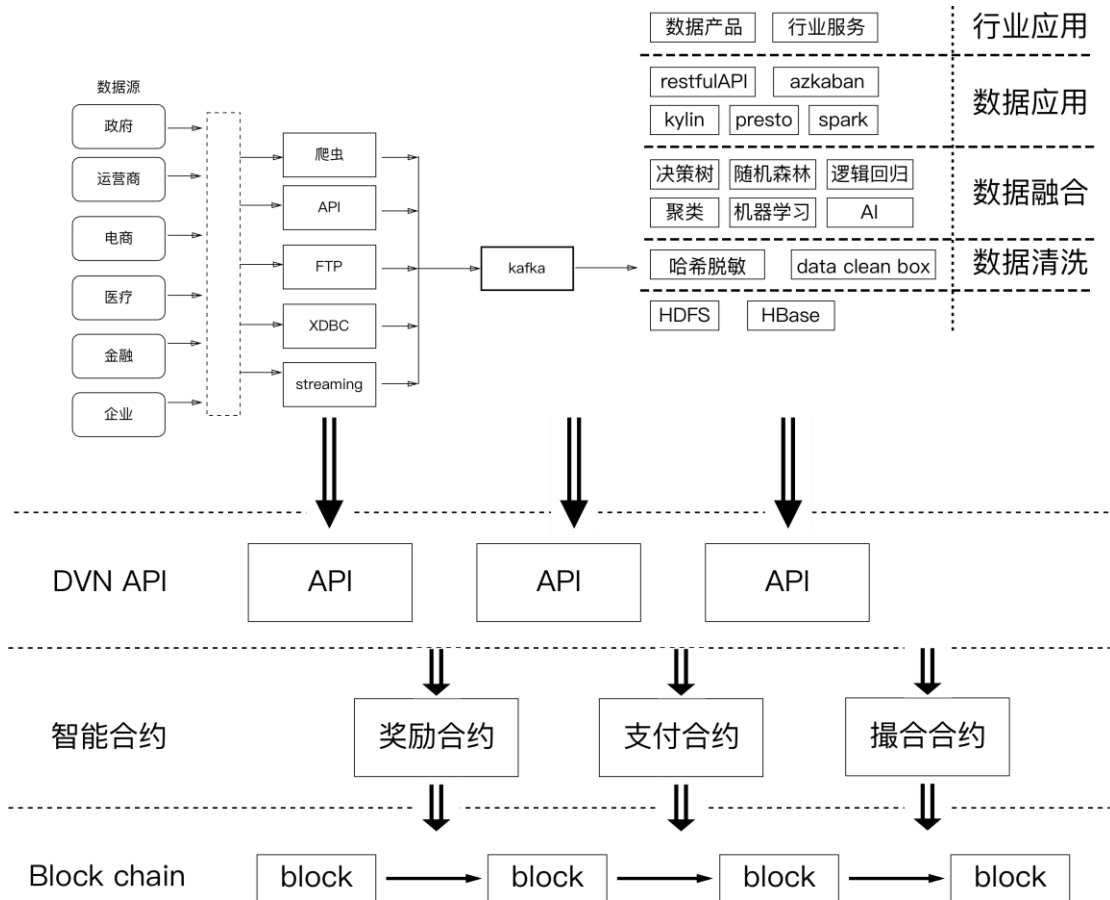
DVN 开放应用市场，具有数据供给推介、需求发布、价值撮合、中间存管等，提供基于融合数据的开放型应用市场。

4.2.5 榕树安全套件 **Banyan Security Suite**

DVN 分布式安全套件，具有数据治理、流程管理、风险评估、认证标示等，提供面向所有参与方的数据安全管控服务。

5. 技术基础

DVN 将大数据技术和区块链技术进行深度融合，支持数据应用生命周期管理。整个数据生命周期中的所有数据行为都通过 DVN API 与 DVN 链路进行交互，DVN API 接收到数据请求后，通过智能合约将奖励、支付等写入区块链中记录保存。参考技术架构如下：



参考技术架构中涉及的相关技术如下：

爬虫、API、FTP、XDBC、Streaming 等多数据交互协议；

HDFS 等分布式文件系统、HBase 等分布式数据库；

哈希算法、data clean box 等工具集；

决策树、随机森林、逻辑回归、聚类等数据挖掘算法；

apache kylin、presto、Impala、sparkSQL 等大数据平台工具。

6. 理论体系

6.1 数据价值三定律与其内含参数

数据的价值体现在应用，只有在流通与使用中的数据其价值才得以体现。通过成熟的商业模式验证，总结形成“数据价值三定律”：

定律一，沉默的数据加速贬值；

定律二，流通的数据产生价值；

定律三，融合的数据促进增值。

以上定律中，数据价值用 V 表示，与数据年龄相关的时效性用 A 表示，数据流动性用 D_y 表示，数据的融合性用 F 表示。

6.2 数据质量参数

数据质量 Q 乃参考了现行国际标准（如 ISO8000）以及国家及行业标准规范所制定的相关参数，反映了数据的有序性、一致性、精准性以及完整性。

6.3 数据流动性参数

数据的流动性 D_y 与数据被加工、交易、流通、应用的频次相关。其中，加工、交易与流通都可使数据的价值（ V ）与质量（ Q ）发生改变；而应用的频次 F_q 乃依据行业经验制定的数据使用次数等级划分，可表示为以下函数：

$$D_y = f_1(\Delta V, \Delta Q, F_q)$$

6.4 数据融合度参数

数据的融合性 F 与不同来源数据相互打通、关联的程度有关， F 的参数大小等于数据链中实际产生关联的维度。

6.5 数据时效性参数

数据的年龄 A 反映了数据的时效性，与数据生成的时间 T ，数据被清洗、整理导致数据质量发生改变的时间 $T_{\Delta Q}$ ，数据由于被加工、交易、流通造成数据流动性发生改变的时间 $T_{\Delta Dy}$ ，以及数据融合发生的时间 $T_{\Delta F}$ 有关，具体的函数如下：

$$A=f_2(T, T_{\Delta Q}, T_{\Delta Dy}, T_{\Delta F})$$

其中，数据生成的时间 T 越长，数据的时效性越差，数据年龄 A 越大；数据质量、流动性以及融合发生改变的时间越长，数据的时效性越差， A 越大。

6.6 数据价值函数

根据以上定义，可将数据价值 V 表达为以下函数：

$$V=f_3(Q, Dy, F, A)$$

其中， V 与 Q 、 Dy 、 F 的大小成正比，与 A 的大小成反比。

6.7 相关参数的定义决策机制

数据质量 Q 、 f_1 、 f_2 、 f_3 、 f_4 （参见 7.6）、 f_5 （参见 7.6）的定义与参数大小，依通证持有人取得共识（参见 7.7）。

7. 通证规则

7.1 通证介绍

DVN 通证是一种为 DVN 提供支持的通证，代号 BBN (Big Banyan Network)，其使用场景包括数据接入奖励、数据清洗奖励和数据使用支付等。

7.2 通证场景

BBN 的作用是对生态系统中开发和使用者进行激励，以提升整个网络中数据的质量、融合度和可用性，从根本上促进数据行业的规范、阳光化发展，为最终的数据应用环节提供坚实的支撑和保障。

在 DVN 中，通过区块链的特性解决了诸多实际问题，例如数据清洗、融合、产品交付中的可追溯和防篡改，也就是说存在大量的硬性场景需要消耗 BBN。例如，参与者在数据的治理、产品化各环节中全部使用 BBN 参与注册、充值，完成数据治理或产品加工后获得奖励同样通过 BBN 来支付。因此，数据行业的繁荣，就是 BBN 价值的支撑。

7.3 流通模型

DVN 设计了一整套完整的 BBN 流通机制，丰富的 BBN 使用场景，通过一系列的智能合约保障 BBN 的流通性。

数据引入奖励合约 (Introducing Rewards Contract, IRC)：负责

计算和确认数据源数据引入的奖励（Reward）。数据在 DVN 上形成数据产品和服务，在使用后，按照治理委员会制定的比例分配给数据源一定额度的奖励。

数据清理奖励合约（Cleaning Rewards Contract, CRC）：负责计算和确认数据治理参与方获得的奖励（Reward）。申请数据清理需要付出一定的押金（Deposit）才能取得一定数量的数据加工优化处理权限。其处理工具由 DVN 提供，节点只提供运算资源。在数据被使用后，按照治理委员会制定的比例分配给数据清理方一定额度的奖励。具体函数如下：

$$\text{Deposit} = f_4(V_0)$$

其中 V_0 为出押金时的数据价值 V

$$\text{Reward} = f_5(\Delta V, \text{time})$$

其中 ΔV 为加工后数据价值的提升，若在承诺的时间（time）内无法完成数据清理，则没收押金（防止数据源被长时间占用得不到治理）。

数据使用支付合约（Apply Pay Contract, APC）：负责计算和确认数据使用方使用数据服务后的支付。具体支付额度按照治理委员会制定的标准执行。

撮合合约（Transaction Contract, TC）：负责计算和确认数据治理和产品化使用的手续费用。

此外，随着数据治理岗位和数据应用场景等越来越丰富，不排除在系统运行过程中增加其他合约的可能性。

7.4 持有与决策参与

任何持有 BBN 者都可以将通证抵押获得治理委员会的投票权。投票权益与抵押数量和抵押时间均正相关。

治理委员会的治理内容包括但不限于数据源引入终止、数据价值评价、数据治理奖励方案、数据使用支付方案等。这些调整可以通过治理系统中共同投票参与决定。

8. 治理机制

8.1 底层运行

DVN 底层区块链基于以太坊^{[8] [9]}并公网运行，以太坊作为目前最成熟的支持智能合约的平台，社区活跃且基金会运转良好。未来不排除采用 AntShares、HPB 等其他或自主独立开发技术。

8.2 智能合约

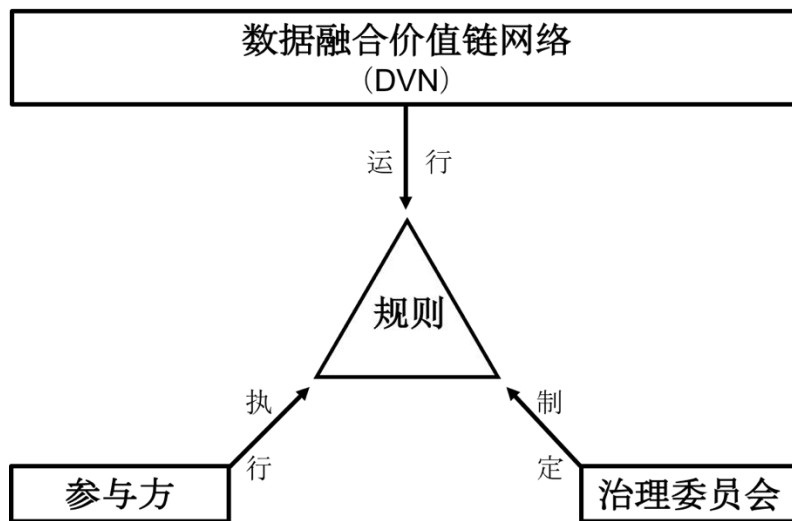
对网络内通用性强的业务模型与流程，制定通用智能合约模版，并在 DVN 中绝大多数业务场景统一使用。针对特定类别或后续新增模式，开发专类的智能合约储备库并公示，供灵活配置高效运行。

8.3 共识机制

DVN 里任何数据的引入、治理、加工、使用以及与之相关产生的奖励、支付等通证的流通，都通过 DPOS（Delegated Proof of Stake）共识机制^{[10] [11]}记录和写入。

8.4 三方模式

采用网络共治三方模式，包括数据融合价值链网络（DVN）、治理委员会、参与方，通过合理、适当且有限度的权利分配、运行机制确保整个价值体系运行。其中，治理委员会负责“规则制定”，参与方负责“规则执行”，DVN 负责“规则运行”（如下图）。



9. 技术应用

9.1 区块链应用

现行主流的各种数字通证，包括比特币和以太坊，均以区块链作为其核心技术，以确保货币交易纪录之验证确认、不可篡改和去中心化区块验证储存。

DVN 通证（BBN）同样以区块链作为其生成流通的技术支撑，除了确保通证在流通交易时的安全、可信与方便，还能依此客观条件让 BBN 使用者建立市场需求信心。特别地，BBN 有 DVN 生态体系内丰富、充分的应用场景作为使用支撑，使之有机会能成为如比

代币和以太币一样的，大量流通且保值增值的主流数字通证。

9.2 人工智能应用

以深度学习方法为主流的人工智能是大数据高附加价值应用的热门领域，也是大数据采集、清洗、融合等前端加工过程的利器^[12]。

DVN 一方面将利用成熟数据作为“训练”人工智能的素材基础，发展各种场景的人工智能应用作为市场产品。一方面利用人工智能对数据工程师们实施“人工”数据采集、清洗、融合过程中的人脑判断流程进行模拟，并开发实施上述业务的自动化工具或程序包。以此提升流程效率，并显著降低数据在加工过程的人为影响所产生的泄密安全风险和其他人为疏忽。

10. 感谢

DVN 的设计开发除了核心团队自身对数据行业的深耕理解和深刻洞察，更得到了业内标杆客户、合作伙伴、行业组织的支持和信任，以及来自区块链、人工智能领域的工程技术专家和市场资深人士的贡献与付出。

在此，对过往参与 DVN 建设的技术、业务、管理等专业贡献者，当下参与项目的模式、平台、资金等资源贡献者致以诚挚感谢，期待与有志于此的所有参与者一起，共同构筑数据融合无限未来。

榕树网络团队

官方网站: <http://www.banyanbbt.org/> 电报交流群: <http://t.me/bbnfans>

参考文献

- [1] *M Serrato , J Ramirez*The Strategic Business Value of Big Data. Springer International Publishing , 2017
- [2] *L Furtado , M Dutra , D Macedo*. Value Creation in Big Data Scenarios: A Literature Survey. Journal of Industrial Integration & Management , 2017 :1750002
- [3] *Satoshi Nakamoto*. Bitcoin: A Peer-to-Peer Electronic Cash System. White Paper, October 31, 2008
- [4] *Antonopoulos, Andreas*. "Bitcoin security model: trust by computation". Radar. O'Reilly. Retrieved 19 November 2016.
- [5] *Lakhani, Karim R.*. "The Truth About Blockchain". Harvard Business Review. Harvard University. Retrieved 2017-01-17.
- [6] *Thompson, Jeffrey*. "The Rise of Bitcoins, Altcoins—Future of Digital Currency". The Epoch Times. Retrieved 2013-12-29.
- [7] *J Sanger, C Richthammer, S Hassan, G Pernul*. Trust and Big Data: A Roadmap for Research. Workshop on Security in Highly Connected It Systems. , 2014 :278-282
- [8] *Popper, Nathaniel*. "Business Giants to Announce Creation of a Computing System Based on Ethereum" – via NYTimes.com.
- [9] Ethereum. URL <https://en.m.wikipedia.org/wiki/Ethereum>
- [10] Proof-of-stake. URL <https://en.m.wikipedia.org/wiki/Proof-of-stake>
- [11] Delegated Proof of Stake (DPOS) . URL <https://bitshares.org/technology/delegated-proof-of-stake-consensus/>
- [12] DE O'Leary. Artificial Intelligence and Big Data. IEEE Intelligent System. 2013: 28 (2)