

永不停息的数据和永不停息的应用: DATAcoin by Streamr

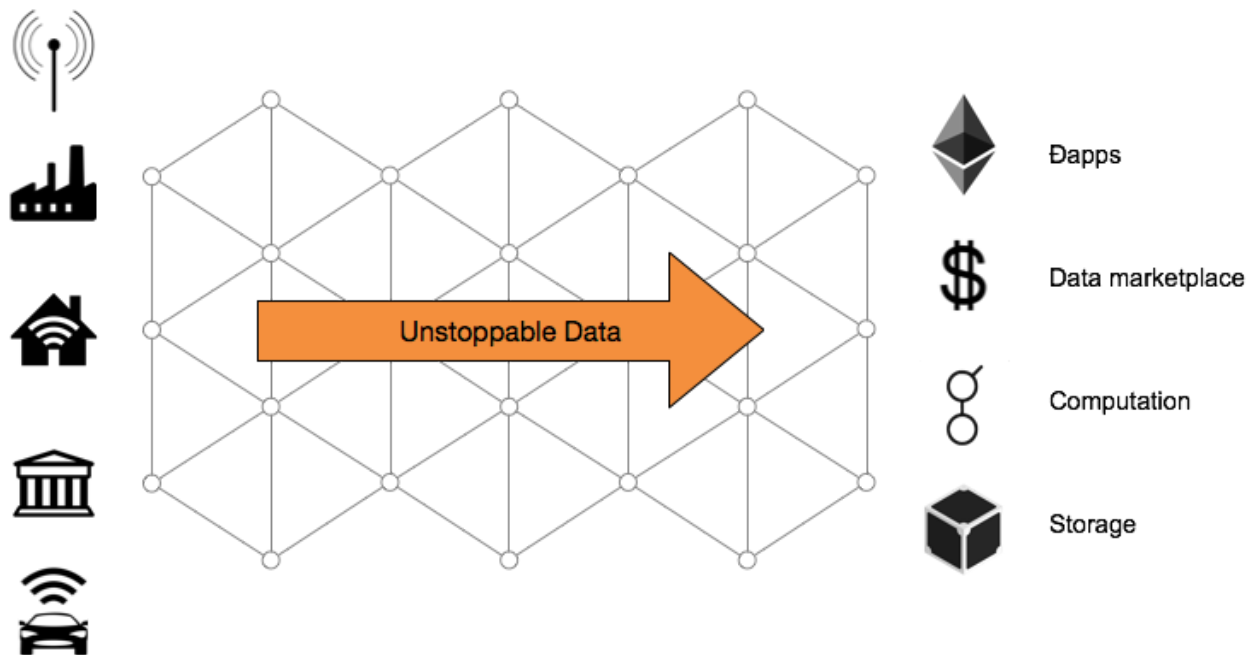
streamr

第一个公开初稿

2017年5月25日

版本 0.7

本白皮书仅供参考，不构成要约或任何形式的投资建议。
本白皮书的任何部份可能会随着项目的进一步发展而发生重大变化。



Streamr 愿景

Streamr将永不停息的数据提供给永不停息的应用程序。它是全球超级计算机的实时数据骨干。它是一个由DATAcoin代币操作的，去中心化，可扩展，低延迟，数据不可篡改并可持续的网络。任何人 - 或任何东西 - 可以将新数据发布到数据流，而其他人可以订阅这些数据流来为Dapps，智能合约，微服务和智能数据管道提供支持。

为了激励用户参与网络，有一个内置的数据获利机制。来自安全的交易所，连接的设备，物联网传感器和社交媒体的有价值数据可以提供给公司，开发者和公民。机器可以自主地出售他们的数据，获得报酬，并购买他们需要的数据。全球实时数据市场，以内置数据来源，加密和访问控制的形式出现。

除了去中心化的数据网络和市场，完整的Streamr系统还包括强大的分析引擎和用于快速开发实时数据Dapps的图形界面。数据流，智能合约和去中心化计算资源可以通过高级构建模块，使用少量代码的环境中进行互连。Streamr将是能最简单的创建实时的，数据驱动的和可信赖的块链应用程序的地方。

一场革命正在发生，中心化云服务逐渐被代币化的，去中心化的解决方案所取代。例如，Golem替代Azure Virtual Machine，而IPFS替代Azure Blob Storage。Streamr很自豪能加入这场革命。通过为消息传递和事件处理提供去中心化的解决方案，替代Azure EventHub和Azure Stream Analytics等平台。

1. 背景

实时数据将在未来几年逐渐变成商品。在制造业，服务业和整个供应链中，传感器和连接设备正在生成大量带时间戳的数据。这是现代经济的基础，其中大部分数据以流媒体方式产生。^{1,2}

随着物联网的增长和连接设备的普及，数据量呈指数级增长。在全球物联网市场，IHS Markit预测³ 安装设备的基地将从2015年的154亿台设备增加到2020年的307亿台，到2025年将达到754亿台。新生成的数据很多都是有价值的：它可以用来优化制造业流程，以更高的准确性跟踪资产，以高粒度对目标消费者服务，并创建全新的服务和商业模式。

同时，对下一代计算行业来说还有一个大趋势。在一个分布式的未来，去中心化应用程序—或Dapps⁴—的后台代码将在点对点网络中运行。以太坊本身是一个Dapp，Golem也是如此，还有更多Dapp的在开发中。

然而，Dapps不能孤立的运行：它们需要外部数据才能运行。就这样，现实世界数据的存储和分配仍然是中心化的，并且Dapps对所有已知的问题仍然需要负责：权力集中，缺乏稳健性，以及网络攻击时的脆弱性。

可以肯定的是，您可以将数据存储区块链中。还有分散的文件存储应用，如IPFS，Swarm和Storj。数据库如BigchainDB也开始出现。虽然这样的解决方案肯定是新的去中心化架构的一部分，但它们并不真正为任何需要大量实时数据的场景提供答案。该链并不是设计用于高吞吐量或低延迟，它不可扩展，并且存储昂贵。

需要的是一个本地的去中心化数据骨干，作为去中心化应用的补充。这个实时数据骨干将是缺少的链接，这恰是我们想要帮助提供的链接。我们创建的基础设施由一个技术栈组成，可帮助连接和激励全球点对点（P2P）网络中的计算机。这是一个提供低延迟，稳健和安全数据传输和持久性的网络，并且全部可扩展。Dapps未来将由数据推动，我们的使命是确保数据川流不息。

我们也创造了实时数据市场。在数据市场中，任何人都可以将事件发布到数据流，任何人都可以订阅数据流，并在去中心化应用中使用数据。许多数据是免费的，但如果不是这种情况，则使用条款将存储在以太坊智能合约中。我们需要一个数字代币 — DATAcoin来访问和操作数据市场，并补偿P2P网络中的节点。订阅用户使用代币支付数据，而数据生成器和网络参与者将自动和安全地得到补偿。

¹ Susan O'Brien: "5 Big Data Trends Shaping the Future of Data-Driven Businesses", Datameer, 11 May 2016 (<https://www.datameer.com/company/datameer-blog/5-big-data-trends-shaping-future-data-driven-businesses/>)

² Tony Baer: "2017 Trends to Watch: Big Data", Ovum, 21 November 2016 (https://ovum.informa.com/~media/Informa-Shop-Window/TMT/Files/Whitepapers/2017_Trends_to_Watch_Big_Data.pdf)

³ Sam Lucero: "IoT Platforms: enabling the Internet of Things", IHS Markit, March 2016 (<https://cdn.ihs.com/www/pdf/enabling-IOT.pdf>)

⁴ For a definition of Dapps, see Johnston et al.: The General Theory of Decentralized Applications (<https://github.com/DavidJohnstonCEO/DecentralizedApplications>)

我们的技术栈是建立在去中心化的传输层上的。除了更强大的稳健性，弹性和容错能力外，去中心化的功能有利于开放，透明和社区建设。数据权力不是Google，Amazon，Microsoft和IBM等大型企业所独有。该网络由多个数据生成器，数据消费者和其间的消息代理节点组成。您可以通过贡献数据并帮助运行网络给每个人带来好处从而获得名誉和回报。

我们认为，通过建立良好的可用性界面将促进块链社区的持续增长。我们需要工具，以便非专家的普通用户也可以创建安全的智能合约，并将这些合约和Dapps连接到可靠的数据源。我们将通过提供可视化编辑器，包装器和模板来构建所需的工具包。总之，我们希望成为一个能为任何人提供创造数据驱动的去中心化服务业务的地方。

在白皮书的剩余部分，我们将描述Streamr的技术栈，定义数字代币的角色，解释现状，展示研发路线图，并介绍团队。

2. Streamr 技术栈

去中心化的实时数据管道是构建在多层技术栈之上的：

- **Streamr 编辑器** 由可用界面和工具包构成，可以快速开发去中心化的，数据驱动的应用程序。
- **Streamr 引擎** 是一个高性能的事件处理和分析引擎，以去中心化的方式执行离链任务。它可以运行在一个去中心化的计算提供商，如Golem。
- **Streamr 数据市场** 是一个共享数据流平台，任何人都可以贡献和订阅数据流。
- **Streamr 网络** 是数据传输层，在去中心化的数据流管道中定义了用于消息传递的点对点激励网络。
- **Streamr 智能合约** 启用Streamr网络中的节点达成共识，保持流的元数据，处理权限和诚信检查，并促进代币的安全传输。

以下部分将以自上而下的方法详细介绍技术栈的每一层（见图1）。

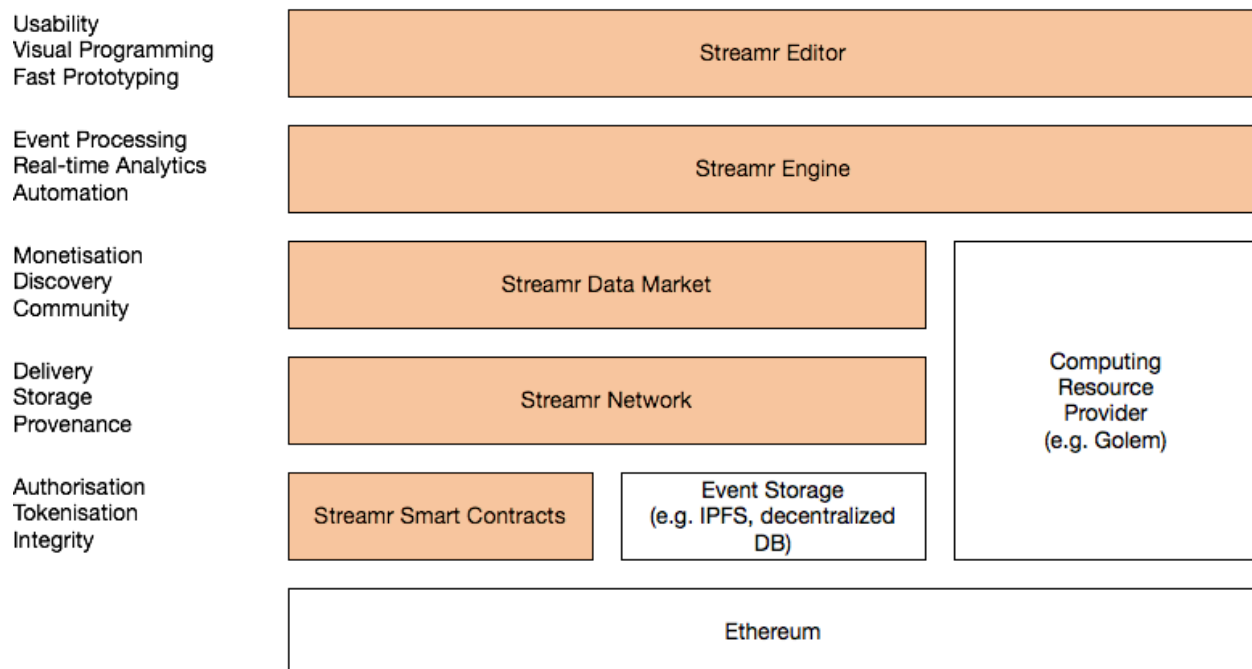


图 1. Streamr 技术栈。

2.1 Streamr 编辑器

Streamr 编辑器 可以快速开发数据驱动的智能合约，降低Dapp创建的门槛，并附带内置常见用例的现成模板。

商界对区块链和去中心化应用程序有相当大的兴趣，但落地应用仍然不多。这些都是早期阶段，可以设想许多想参与的人都不是以太坊，Solidity，加密，数据源和其他技术等细节方面的专家。

在我们看来，生态系统的商业发展需要一整套工具，让不是专家的人也能创建智能合约，连接到可靠的数据源，利用安全的离链模块进行数据过滤，聚合和细化，部署去中心化应用，跟踪智能合约的执行，并可视化输入数据和区块链事件的流程。

我们通过为领域专家和业务用户提供强大的工具（如方便使用的可视化编辑器），包装器和智能合约模板来满足使用层的需求。这些工具隐藏了高深的技术，处理数据集成和通信，并自动执行智能合约部署和监控中的常规步骤。

我们预见到含有一些使用平台和工具的生态系统。现有的Streamr平台已经实现了使用层的一些元素，在未来的几个月和更多的时间里将会增加了更多的功能。最终目标是到了让您可以在几分钟

内创建和部署有用和功能丰富的数据驱动的智能合约的那一刻。这不是幻想；我们在2017年2月在EDCON⁵ Paris 做的演示就是我们可以做得到的佐证（参见图2的例子）。

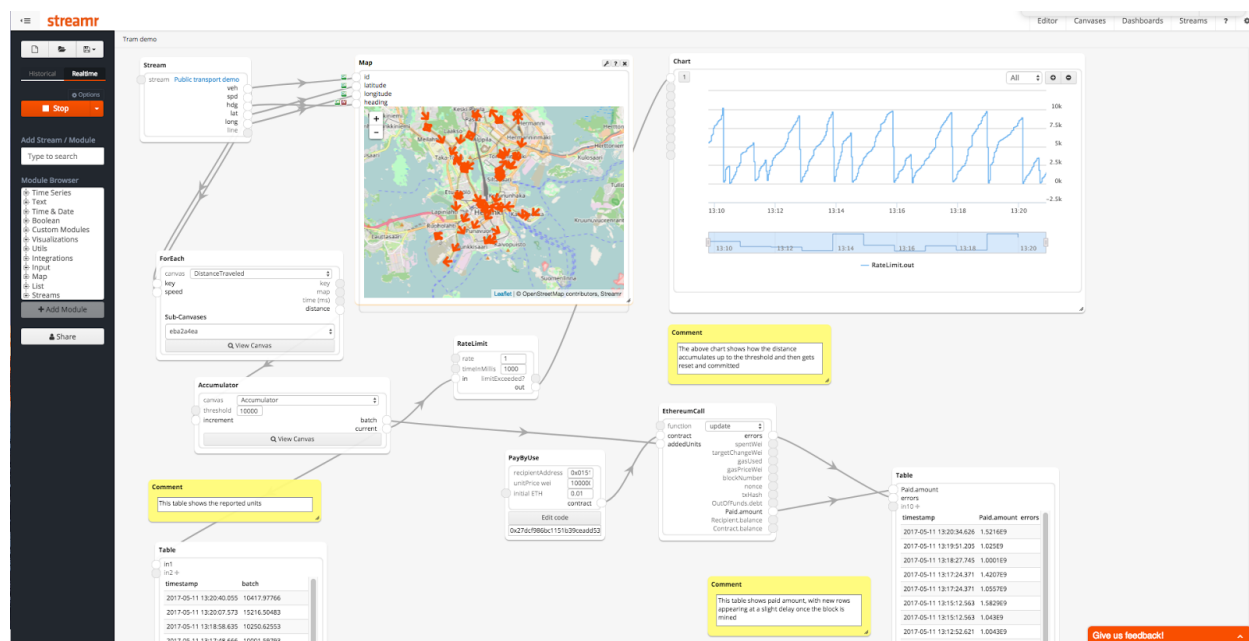


图 2. Streamr 编辑器的alpha版本.

这些是使用层的一些计划实现的功能：

- 一个用于创建智能合约，提供现实数据，并构建离链数据处理管道的可视化编辑器。
- 用于与智能合约进行通信并与区块链交互的模块。
- 用于离链处理的模块：数据过滤，细化和聚合，部署去中心化应用程序，跟踪智能合约的执行以及输入数据和区块链事件的可视化流程。
- Solidity编辑器，可以在上下文相关环境中编写和修改智能合约代码。
- 内置和已测试的开源Solidity模板，用于不同用途的以太坊智能合约。
- 回放功能，用于在部署之前模拟智能合约的功能，调试合约代码和测试功能。
- 一键部署，用于提交智能合约到测试网或主网。

2.2 Streamr 引擎

⁵ Henri Pihkala: "Connecting Ethereum with the real world: How to easily create data-driven smart contracts", European Ethereum Development Conference (EDCON), Paris, 17-18 February, 2017 (<https://www.youtube.com/watch?v=C1l0rcj-Fok>)

Streamr 引擎 是离链运行在去中心化计算提供商（例如，Golem上的Docker容器）上的高性能分析引擎。

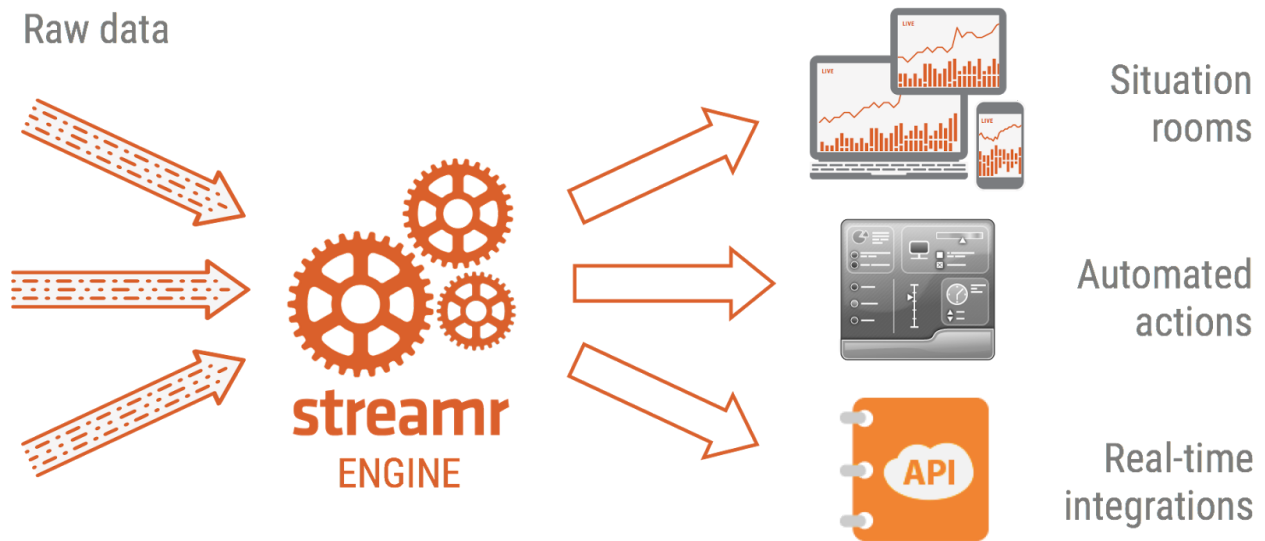


图 3. Streamr 分析引擎 的典型数据流模式和输出。

Dapps通常基于Web的前端界面和智能合约的后台，目前无法处理原始数据和转换其为资讯。一组物联网传感器或股票市场每秒可能产生数千甚至数百万个事件，数量之庞大和费用之贵，以致无法在任何区块链上进行计算。

需要一个流式分析层，将原始数据转换为精确的信息，并可以让Dapps和智能合约消费。原始数据可能需要进行过滤，下采样，聚合，与其他数据组合，运行异常检测算法，或由高级机器学习和模式识别模型处理。或者您可能想要做一些无法在智能合约上简单完成的工作，例如在处理流程上调用外部API。

Streamr 引擎 侦听Streamr网络上的事件，使用Streamr编辑器构建的模型会提炼传入数据并实时响应新事件。有很多方法可以做出响应，其中包括：

- 在Streamr网络中的另一个数据流中发布提炼的数据，也可能通过已连接到网络的Dapp UI进行实时显示。
- 与物联网设备进行交互，例如控制执行器，开锁，开灯或打电话。
- 通过电子邮件或推送发送警报。
- 在智能合约中调用功能函数。

使用Streamr网络作为在Dapps和引擎上的离链计算之间的消息粘合剂，产生了全新的去中心化应用程序类别：由非凡数据量驱动的应用程序。显然，产生的结果也可以被传统的中心化应用程序使用，同时仍可享受去中心化的消息传递和分析的好处。

2.3 Streamr 数据市场

Streamr数据市场是一个全球性的任何人都可以贡献和订阅的共享数据流市场。它是数据获利和机器对机器（M2M）数据交换的场所。数据市场支持匿名，但允许在需要时验证数字身份。

数据市场是数据生产商和数据消费者的汇聚场所。数据消费者在提供的数据中找到有价值数据，并希望访问它，以作为在Dapps，智能合约或传统应用程序中的数据输入。

数据在数据流中组织，数据市场的基本构建块和Streamr网络中的原始数据（参见下面的第2.4章）。数据流中的来自数据源的事件以规则或不规则的间隔持续发射新的数据点。以下是一些以数据流方式生成的实时数据的典型设置：

- 股票市场每当出现新的出价或报价，每次交易发生时都会产生新的事件。
- 公共交通工具每几秒广播它的身份，状态，速度，加速度，地理位置和去向。
- 当在其范围内检测到移动物体时，运动检测器发送信号。
- 在小工厂里，物联网传感器通过电力驱动测量驱动操作当中的温度，速读和震动。
- 空气质量传感器测量城市中的一氧化碳，二氧化硫，二氧化氮和臭氧水平。
- 地震仪测量火山活动区域的地面运动。
- 专业运动员佩戴的智能服装收集心跳，温度和加速度等生物特征数据。

数据市场提供了大量可选可供订阅的可信赖的时间戳数据。一些数据源自成熟专业的数据供应商和再分配者，一些来自公共，开放数据源。重要的是，该平台允许任何人贡献数据和从中获利。虽然公司从传感器和设备中传输有价值的数据，但公民也同时在生产有价值的信息。

例如，穿戴智能手表的人可能会将心率数据放到数据市场上。数据可以匿名提供，因此不会侵犯隐私权。谁会对这些数据感兴趣？那么一家制药公司可能会购买它来进行研究，或者一家公共医疗机构可能会使用它来了解人们进行运动的频率，或者是公众的压力水平。智能手表制造商可能会购买它来获取他们的心率传感器的功能的诊断。数据生产者只是通过提供数据来就可以赚取日常收入。

数据市场中的订阅不应由软件开发人员，数据工程师或数据科学家发起。事实上，分散市场可能最终以机器对机器交易为主。自动机器，机器人，智能家电都在操作中需要数据，并对生态系统其他参与者生产有价值的的数据。

自动增值优化模式将会出现。AI订阅股票市场原始数据，应用专有模式识别来生成交易信号，并在同一数据市场上出售这些信号。

虽然数据市场中的大部分内容将免费提供给所有用户，但仍有的数据需要支付，有的数据需要终端用户提交许可证。在这种情况下，需要一种订阅许可证。许可证给予一种权限在特定时间段内

，在特定条件下访问数据，并且是收费的。这与流媒体音乐有着非常的相似之处：您不需要拥有订阅的数据，您只要在Spotify上听或从iTunes下载歌曲即可拥有歌曲的权限。

数据许可证在智能合约上实现（见第2.5.4节）。区块链的巨大好处是它提供了一种无需信任和去中心化的方式来存储使用条款和访问权限，并确保按照约定进行支付。

在更广阔的背景下，在市场上可能有强大的网络效应。内容越多，数据贡献者和数据消费者对某主题数据的吸引力越大。在Streamr数据市场中，门户网站（实现为Dapp）有助于发现存在的数
据，为数据流的创建和管理提供了一个全面的工具包，并且可以轻松订阅所选择的数据流。

2.4 Streamr 网络

Streamr Network是技术栈中的数据传输层。它由建立点对点网络的Streamr 代理节点组成。此网络托管一套发布/订阅机制，并支持事件的去中心化存储。网络吞吐量随参与节点数量呈线性增长，每秒可处理数百万个事件。

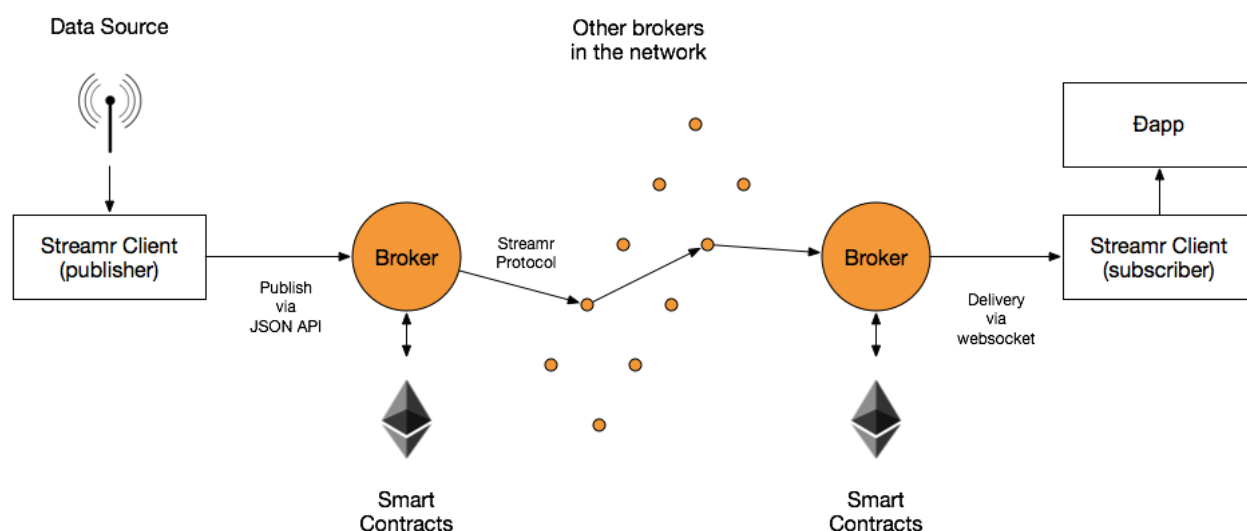


图 4. 通过代理网络从数据源到订阅的Dapp的一个事件流程例子

Streamr网络（图4）是Streamr技术栈中的传输层。网络处理去中心化数据管道中的所有消息。该层由元数据（事件和流）和代理节点组成。代理节点对元数据进行操作，代理节点的集合构成处理去中心化存储和去中心化消息传递的点对点网络。

基础设施层使用Ethereum的底层堆栈进行操作。节点协调需要强有力的共识，这是通过智能合约来实现的。原始事件数据本身往往不会进入区块链，这样隔离能使Streamr网络扩展到每秒数百万个事件的处理量。

Streamr网络将可扩展的基于云计算的实时数据传输（例如Kafka, ZeroMQ, ActiveMQ）的最佳实践与去中心化的点对点/加密货币（Whisper, Bitmessage）中的最佳实践相结合。基于云的框架使用高效的分片和持久性方案来达到高吞吐量，但只能在可信的本地网络环境中使用。点对点协议展示了路由，节点发现，NAT遍历，位置混淆等的有效策略，但无法提供数据密集型实际应用所需的吞吐量。

2.4.1 事件

事件⁶ 是时间戳的信息。每个事件都包含标题和内容。标题指定事件的元数据，例如其时间戳，来源和内容类型。事件协议支持内容有效载荷的任意类型和格式，例如JSON消息或二进制映像。内容类型表明内容的格式。事件标题和内容以二进制格式进行编码以进行传输。

Streamr网络上的所有事件都经过加密签名。所有的事件都有来源，例如一个以太坊地址。签名从私钥和消息的其余部分计算出来。签名用于证明消息的来源以及消息的可靠信。由于事件格式允许任何种类的来源和签名，系统是面向未来的。

一下表格列出了一个事件中包含的信息。

字段	描述
version	事件协议的版本号
stream	Stream id (stream 智能合约的以太坊地址)
partition	数据流分区 (请查看分区章节)
timestamp	事件时间戳 (ISO 8601)
contentType	解析内容的指示 (e.g. JSON)
encryptionType	用于加密内容的加密算法
content	数据有效载荷
originType	指代如何翻译原文
origin	数据发起者
signatureType	指代如何翻译签名
signature	加密签名，以证明消息的来源和完整性

2.4.2 流

⁶ Streamr events should not be confused with events in Ethereum smart contracts.

所有事件属于流。可以有任何数量的流，每组组合在一起的事件都是逻辑相关并且按时间升序的顺序存储。流元数据存储以太坊智能合约中。每个流由以太坊的智能合约地址确定。对于可扩展性，事件（即实际数据点）不存储在智能合约或区块链中。

数据流持有一组权限。权限控制谁可以从流中读取事件（订阅），以及谁可以将新事件写入流（发布）。流拥有者控制这些权限，但她也可以在必要时向第三方授予或委派权限控制。

一下表格列出了流的元数据。

字段	描述
id	流 id (以太坊地址)
name	流名字
description	流描述
owner	流拥有者
permissions	从以太坊地址到权限级别的映射

2.4.3 发布/订阅

网络中的数据传送遵循发布/订阅范例。发布到流的事件将及时发送到流的所有已授权和连接的订阅者。订阅流可以仅限于某些用户，也可以向公众免费。类似的，将内容发布到流的权限可以由一个，多个或所有人持有。

发布/订阅范例可以实现在实际应用中使用的许多信息拓扑：

- 1对多 (例如， 新闻渠道或股票行情)
- 多对多 例如， 群聊或多人游戏)
- 1对1 (例如， 私聊或分析管道)
- 多对1 (例如， 投票系统)

请注意，发布事件并不意味着将事件传递给任何客户端：因为有可能是没有订阅者的情况。尽管如此，该事件仍然存在并传递给多个代理节点进行多重备份。

技术上讲，有两种类型的订阅用户。大多数数据流通过直接连接到代理节点以连接网络的用户（见下文第2.4.4节）。例如，它们可以是Dapps的Web前端，在Streamr引擎上运行的事件处理链，或者来自网络的数据控制的物联网设备。

智能合约是Streamr网络支持的特殊类型的订阅用户。网络中的代理节点受到激励以向订阅的智能合约传递事件。在这种情况下，当然可以应用区块链扩展性限制。该机制允许网络作为一个预

言机，意味着数据可以推送到智能合约而没有第三方预言机的帮助。由于网络中的所有数据都在源头中签名，因此可以随时验证和信任。

2.4.4 代理节点

Streamr 代理节点是网络中的核心软件组件。代理节点处理诸如发布事件，订阅流，处理存储以及通过JSON 远程调用与以太坊节点进行通信等任务。代理节点通过API向连接的应用程序开放其功能。

代理API可以通过任何编程语言的标准HTTP和Websocket库的应用程序中使用。为了易于使用，我们将提供多种语言的实现参考。主要客户端库平台将使用JavaScript编写。它可以用于将数据传递到在浏览器中运行的基于Web的Dapps以及运行node.js的后台应用程序。Websocket API处理从数据源到网络以及从网络到客户端Dapps的事件传递。对于流管理，使用JSON API。

Websocket 流 API负责以下任务：

- 认证会话
- 发布事件
- 订阅流中的事件
- 传递事件到订阅用户
- 查询流中的历史事件

JSON API开放了以下功能：

- 创建一个流
- 设置流
- 删除流
- 得到流的资料
- 搜索流
- 发布事件 (替代 Websocket API)
- 查询流中历史事件 (替代 Websocket API)

代理之间的大部分流量由事件消息组成，但也有与路由和节点发现相关的流量。代理之间的重要协调任务是任务的分配，这必须达成可靠的共识。这一机制将借助以太网络的力量通过智能合约的形式实现（见下文第2.4.5节）。

2.4.5 分区 (分片)

整个网络中的事件流量划分为各独立部份，称为分区。每个代理节点处理一组分区的流量。这是如何实现可扩展性：并非所有节点都处理所有流量。这类似于分区发现方案，例如 Apache Kafka。

事件的分区通过计算流ID的哈希值来得到。这是一个快速的操作，并在本地完成。使用流ID作为分区的键意味着特定流中的所有事件总是转到同一个分区。这让网络在同一个流中维护事件的顺序，并且有效地存储它们。

流可能会发生单个代理无法处理完的消息量。在这种情况下，另一轮分区被用于流，并且流内的流量被分割成独立的部分。在这种情况下，我们对（流ID，流分区）元组进行哈希以分配网络分区，并且发布者提供将事件分配给流内分区的分区密钥。流分区密钥的事件顺序将被保留。

网络中的分区数量保持不变，直到随时间推移自动增加。如下一节所述，有一个协调者智能合约来控制网络分区。分区数量与参与网络的代理节点数成正比。

2.4.6 节点协调

在分布式数据系统（如Apache Kafka和Apache Cassandra）中，通常通过使用Apache Zookeeper等组件来实现节点协调。像领导节点选举的过程中，有一个中心化的过程来建立共识。或者，一些系统在网络中具有特殊权限使用手动的方式分配协调器节点。

在一个去中心化网络中，这种中心化或特权组件将不能存在。相反，Streamr网络使用以太坊底层网络为点对点网络中的节点协调建立共识。

关键协调任务是将网络分区分配给网络中的代理节点，并在节点出现或消失时维护此信息的更新。不像Zookeeper这样的中心化组件，这个任务是由智能合约实现的：网络协调者。网络协调者合约部署在以太坊区块链上。代理节点通过观察和查询智能合约来查看当前的网络状态。只需切换到新的网络协调者合约即可实现网络升级。

重新平衡分区分配是网络协调者合约的任务之一。只有有必要才会更改，如果没必要，将什么都不做。当网络不平衡时，调用功能将DATAcoin授予呼叫者。这种激励措施能确保在需要进行网络重新平衡。

分配给分区的节点接收该分区的所有数据。部分或全部计算数据上的滚动校验和，并以一定的间隔将校验和报告给网络协调员智能合约。在一个大的公共网络中，每个分区都有足够的节点使得它们难以串通。网络协调者智能合约的分区分配也很难影响。

2.4.7 激励

订阅用户是网络中的数据消费者。DATAcoin，网络中的使用代币，让用户订阅流。其他方通过向网络做出贡献来获取DATAcoin：代理节点（该网络的“矿工”）和数据发布者。

代理节点被鼓励做两件事情：将分配的分区分区校验报告给网络协调者智能合约（见上文第2.4.6节），并向任何智能合约订阅户提供数据（参见第2.4.3节）。这两个操作都会消耗一些以太坊gas并由代理节点支付。这成本由代理节点收到的用于进行网络功能的DATAcoin所涵盖。

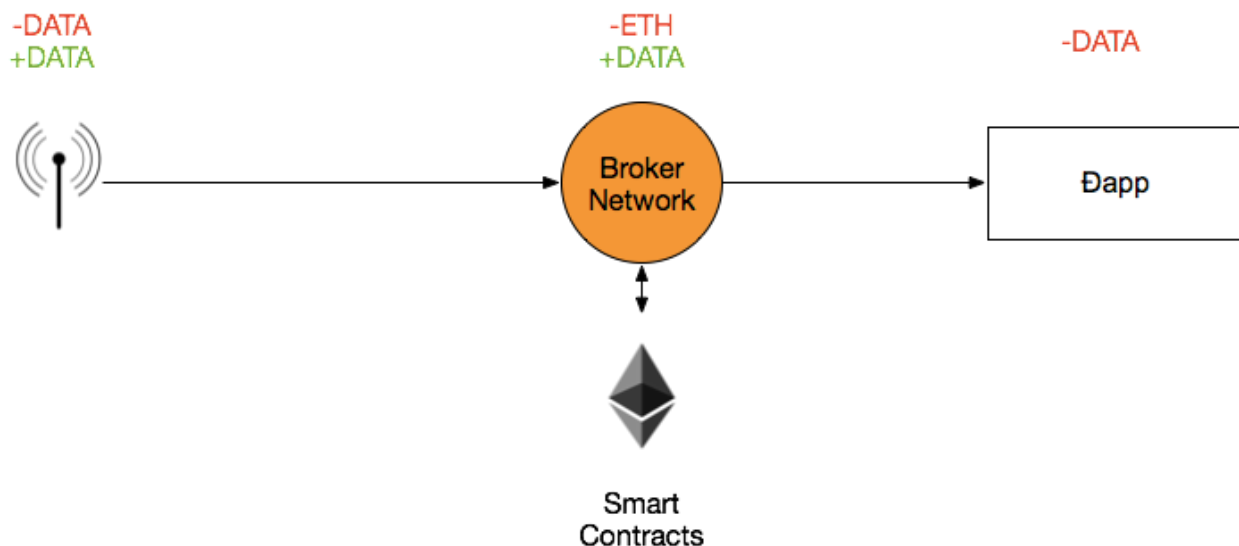


图 5. Streamr 网络 激励结构示意图.

分区的校验由多个代理节点计算和报告，代理节点只有在协调者智能合约中设定的一致性阈值上同意校验才能得到奖励（例如，90%的指定代理节点需要报告特定的校验来证明有效性）。如果一个节点报告了不正确的校验，没校验，或不一致的校验，则不会获得奖励，并且此违规节点未来不太可能被分配给分区。

如所讨论的，智能合约可以是流的订阅用户。订阅者在DATAcoin中设置一个奖金用于将事件交付给智能合约。该奖金将由最先交付数据的一方获得。通常代理节点直接连接到发布者，这是因为该代理节点处于进行交付的先行者位置。其他节点或外部用户可以观看此过程，并确定传递数据的机会，否则由通常的嫌疑节点传送。

还需要一种机制来防止网络中的洪泛攻击。最低成本必须与所有发布操作和订阅者交付相关联。网络可以将成本汇总并提交给底层的区块链进行扩展，类似于在某些区块链网络中运行的状态通道或微支付通道。

2.4.8 事件持久化

数据流中的事件被持久化在点对点网络中。这有效地将网络转变为去中心化的时间序列数据库。去中心化带来了许多优点，包括更强的鲁棒性，容错性，更低的成本和匿名的潜力。

使用流作为事件序列，最简单的存储形式是事件日志。事件日志可以存储在任何区块存储器中，例如节点本身上的文件系统或去中心化的对象存储器，例如 Swarm⁷，IPFS⁸ 或 Storj⁹。

⁷ Viktor Trón et al: "Introduction to Swarm" (<http://swarm-guide.readthedocs.io/en/latest/introduction.html>)

⁸ See "IPFS - The Permanent Web" (<https://github.com/ipfs/ipfs>)

⁹ See "Storj: A Peer-to-peer Cloud Storage Network" (<https://storj.io/storj.pdf>)

对于具有更多粒度和查询功能的存储介质，去中心化数据库如 BigchainDB¹⁰ 正在出现。这样的解决方案是 Streamr 网络中事件存储的候选者。然而，行业正在迅速变化，我们目前不会承诺使用特定的存储解决方案。

2.4.9 数据源

只要外部数据用作 Dapps 和智能合约的输入，安全和数据源就是关键性问题。由于区块链交易是不可撤销的，因此有明确的激励给诚实者以确保输入是值得信赖的。还会有一些动机不诚实的人以及不道德的黑客来操纵数据以获得利益。

在 Streamr 网络中，每个数据点由数据源拥有者的私钥进行加密签名。该领域正在快速发展，有许多方法是可实现的。事件可以用如以太坊私钥，或由物联网传感器的 X.509 证书签名，可信硬件使用英特尔的 SGX 芯片和 Town Crier¹¹ 的继电器，或者用 TLSNotary¹² 服务来建立 web API 的数据的桥梁。

根据设计，考虑 Streamr 网络用于证明数据源的方法是无意义的，它确实可以支持现在或将来可用的任何方法。网络上的事件总是携带签名本身以及用于验证签名的方法的信息。发布和订阅事件的客户端库可以逐步增加对不同方法的支持，从开发人员的角度抽象出每种方法的内部工作，并使签名验证变得容易。

最初支持的签名算法与以太坊使用的 secp256k1 ECDSA 相同。这样便于网络将任何已发布的数据可靠地映射到一个以太坊地址。

2.4.10 数据保密

鉴于任何人都可以通过运行代理节点参与 Streamr 网络，Streamr 网络中的非公开流的事件有效载荷将使用非对称密钥进行强加密。只有持有授权私钥的一方才能读取数据。流智能合约持有允许访问该流的公钥。在发布时，授权人的公钥用于加密数据，这样只有授权人可以访问数据。

组播加密^{13,14} 技术可用于平衡消息大小。内置的加密支持还可以实现直接的数据获利，因为可以创建诸如 Streamr 数据市场的服务来销售或租用对流内容的访问权。发布者可以重新输入数据流，以选择性地拒绝访问。例如防止他们抓住用户在网络外转售他们的数据。

¹⁰ See “BigchainDB: A Scalable Blockchain Database”
(<https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf>)

¹¹ Fan Zhang et al.: “Town Crier: An Authenticated Data Feed for Smart Contracts”, Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS), Vienna, Austria, October 24-28, 2016, pp. 270-282 (<https://eprint.iacr.org/2016/168.pdf>)

¹² “TLSNotary - a mechanism for independently audited https sessions”, whitepaper, September 10, 2014
(<https://tlsnotary.org/TLSNotary.pdf>)

¹³ Micciancio, Daniele and Saurabh Panjwani. “Multicast Encryption: How to maintain secrecy in large, dynamic groups?” (<http://cseweb.ucsd.edu/~spanjwan/multicast.html>)

¹⁴ Duan, Yitao and John Canny. Computer Science Division, UC Berkeley. “How to Construct Multicast Cryptosystems Provably Secure Against Adaptive Chosen Ciphertext Attack”
(<http://www.cs.berkeley.edu/~jfc/papers/06/CT-RSA06.pdf>)

去中心化方案与加密相结合带来了安全性：数据被分散到多个未知的物理位置，并在传输和存储过程中受到强加密的保护。该设计解决了公司和组织对通过物理访问数据中心和存储设施的数据风险的潜在担忧。

2.5 Streamr 智能合约

许多以太坊智能合约支持Streamr网络和数据市场的运作。Streamr网络使用智能合约进行激励，协调，权限和完整性检查。数据市场筑建在网络提供的数据许可和获利的功能之上。DATACoin，一个ERC20代币，被用于激励，声誉象征，以及付款方式。

2.5.1 流

流智能合约保存有关流的信息（参见第2.4.2节）。除了持有静态信息，它还持有流的权限信息。特别地，它持有加密流的授权人的公钥，这可能与数据许可证相关（见下文）。

2.5.2 流注册表

流注册表合约保存网络中已知流的信息。流可以添加到注册表以供查找。流注册表还可以在ENS（以太坊名称服务）上注册流。

2.5.3 网络协调者

网络协调者合约将分区分配给代理节点（参见第2.4.6节）。代理节点向协调者注册，并通过观看智能合约接收网络状态的更新。

2.5.4 数据许可证

数据许可合约代表了Streamr数据市场中列出的产品。以DATACoin作为回报，合约通过在流中注册购买者的公钥来授予对相关流的访问权限。数据许可证可以在一段时间内生效。许可证过期后，购买者将无法再访问流上发布的新数据。

许可合约的存在价值是提供接收方有权使用特定和不变的使用条款访问数据流的证据，同时保证数据提供者在发布时候收到约定的实时数据资金。使用条款可以直接存储在数据许可合约中（并根据需要进行哈希），或作为链接储存在内容寻址存储系统中，如IPFS。合约还可包含关于履行法律要求的证明，例如“认识你的客户”（KYC）流程的结果。

代理节点报告事件计数和滚动哈希到流智能合约，从而可以将其报告给相关的许可证合约。许可证智能合约几乎可以实现任意的保护措施，以防止发布者欺诈。例如，他们可以锁定付款直到在流上发布一定数量的事件。款项也可以随着时间的推移获得或以逐个事件的发布获得。还有一种机制让订阅用户可以标记坏数据，从而对发布者的声誉产生不利影响（参见第3节的DATAcoin和其机制）。这些安全属性可确保发布者若无法按照承诺提供优质数据则无法获得报酬。

3. DATAcoin

DATAcoin是数据生产者和消费者之间的补偿手段。这也是在点对点网络中运行代理节点的激励代币。DATAcoin是运转的基础，是社区的声誉度量。在更大的图景中，这是一种获取数据作为有价值商品的方式。

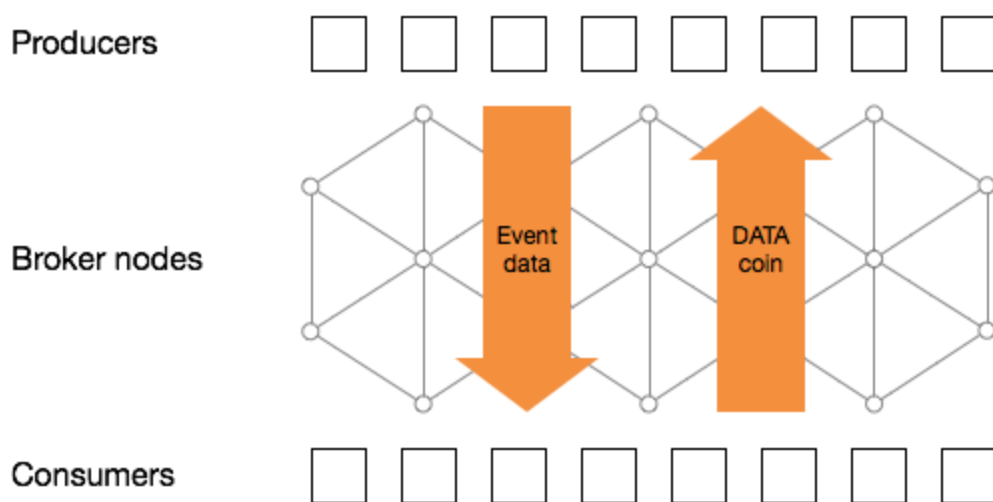


图 6. DATAcoin 往数据的相反方向流动.

在去中心化的数据管道中，数字代币是不可或缺的。DATAcoin是Streamr网络的使用代币。DATA是代币的符号。

- 维护和运作点对点网络会消耗资源：时间，电力，计算能力和通信带宽。第2.4.7节描述了参与代理节点的激励。
- DATAcoin是生产者和消费者之间的补偿手段。换句话说，它实现了数据生产者的货币化机制。这激励数据供应商加入和帮助社区成长为每个人的利益。
- DATAcoin是业务的基础，是数据生产者，数据消费者和消息代理在社区中的声誉的度量。各方从DATAcoin交易中获得回报：发布数据，使用数据以及运行网络的代理节点。数据生产者在将其发布的事件发送给用户时获得代币回报。而订阅者通过接收事件来获得回报。代理节点通过帮助数据传递和持续性而获得回报。这记账很容易：回报等于交换的DATAcoin的数量。差异在于订阅者回报最终会过期，而代币结余则不会。

DATACoin是在以太坊上实现的ERC20代币。代币智能合约维护DATACoin结余，并确保以可靠和安全的方式处理付款。遵循ERC20标准确保与钱包和其他代币的互通性。

DATACoin将在当前计划于2017第3季的代币生成活动（TGE）中创建。其细节，条款和详细的时间表将在稍后公布。

4. 现况

已经有一个很先进的平台用于创建数据管道，可视化和离链计算微服务。该平台提供了一个功能起点，但要实现完全的去中心，则必须重新安装在去中心化容器中运行，并使用新的Streamr网络层进行消息传输。

我们不是从头开始。已经有一个功能齐全和先进的平台用于创建数据管道，可视化，离链处理和以太坊智能合约。该软件是具有可扩展性，集成度和容错能力的云环境构建的。诸如Kafka和Cassandra这样的大数据框架用于数据处理和消息传递。Streamr平台于2017年2月在EDCON展出。

再追溯身世，我们在5年前就创建了第一个版本的软件供我们自己在高频交易算法中使用。高管都来自金融背景，量化交易，交易系统开发者，统计数据交易者，以及以上提到的都有相关背景。量化金融是一个自动处理大量实时数据，在过去10-15年或以上已经成为现实的领域。只有过去几年，同样的操作方式和同样的工具和平台正在进入到物联网，IoT和现在进入区块链的世界。

当前的平台是功能丰富的，可扩展的并正在被企业客户使用。然而，大多数组件不直接转化到新的世界。存储需要去中心化，消息，发布/订阅功能，和传输层内置的数据获利和加密，以及与节点协调和激励一起建立的点对点网络。下一节将介绍如何做这些事情的路线图。

5. 路线图

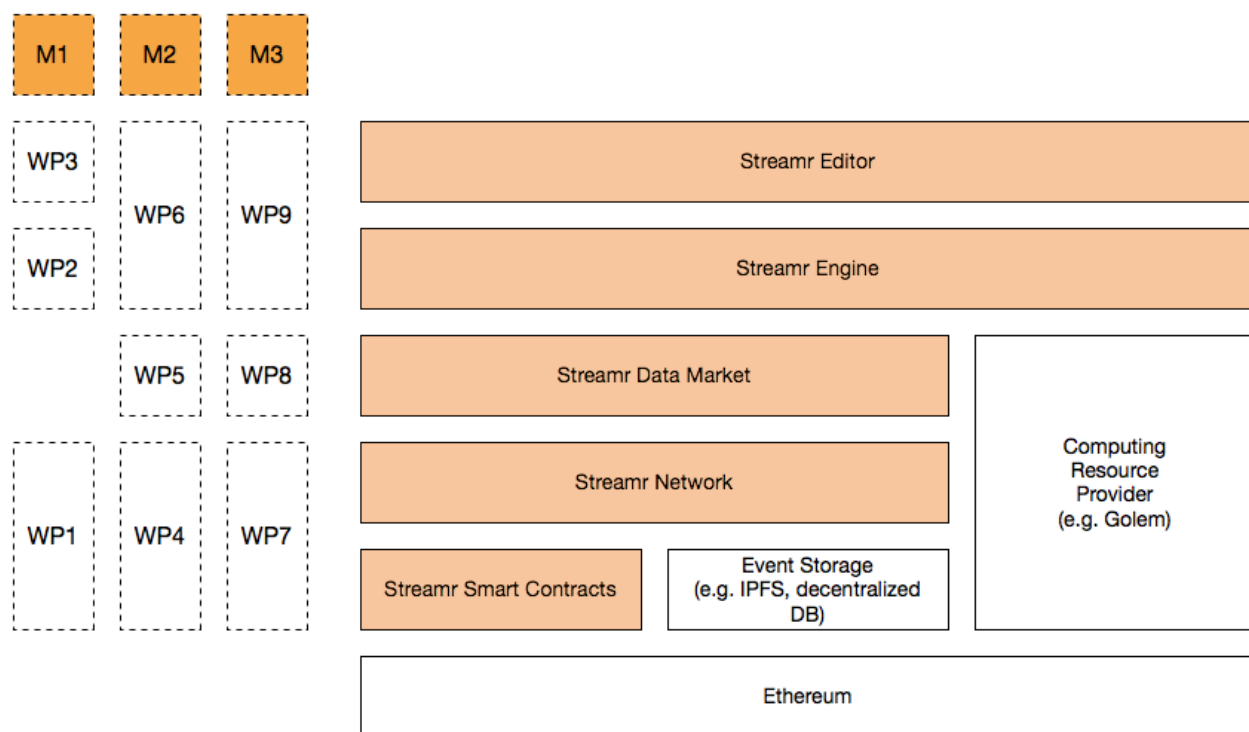


图7. Streamr 项目研发路线图.

路线图（图7）分为三个里程碑（M1-M3）。每个里程碑迭代地将新功能带入技术栈中。三个里程碑中的每一个都包含工作包（WP1-WP9）。在一个里程碑中完成所有的WP则完成了该里程碑。每个WP都着重关注技术栈中的某些层。一个里程碑中的所有WP都将大致同时进行，但下一个里程碑中的WP将不会在当前里程碑完成之前开始。

我们选择了迭代全栈的方法，具体如下：

- 我们可以从第一天开始向社区提供可以使用的东西。
- 我们从现有的技术栈开始。技术栈是模块化的，所以在路线图上任何一点都可升级任何一层。
- 无论我们达到哪个里程碑，都将实现明确和有用的可交付成果。社区不会因为中途解决方案而落不了地；将会有功能性技术来满足许多商业用例。

5.1 里程碑 1

里程碑1发布激励数据传输网络的第一版和底层智能合约。开发与以太坊生态系统完全整合的基础之上的其他层。

WP1 (网络, 智能合约): 去中心化代理网络原型

- 创建去中心化代理节点原型的第一个版本

- 整合libp2p
- secp256k1 ECDSA 签名和验证
- 事件协议
- JSON 和Websocket APIs
- 流智能合约
- 流注册表智能合约
- 网络协调者智能合约
- 分区分配
- 校验报告
- 通过发布／订阅费用防止攻击
- DATAcoin 基本激励计划

旧的“云”消息层和新的去中心化网络将并存一段时间，直到去中心化的代理节点达到规模和稳定。现有的生产引擎和编辑器将用旧代理运行，直到网络层升级。

WP2 (引擎): 引擎到以太坊

- 支持以太坊和Streamr互动的Streamr-Web3 桥
- 智能合约开发
- 本地和交易功能调用
- 事件观察
- ABIs
- 密钥和账号管理
- 支持测试网和主网
- 签名和签名验证
- Docker引擎给Golem 或其他去中心化计算提供商

WP3 (编辑器): 无缝的链上和离链计算

- 改进可视化编辑器以完全支持在引擎上所有以太坊相关的特性
- 集成Solidity 编辑器用来编写自定义智能合约
- 用于最常见的用例（付款，投注，SLA监控，预测等）的内置精选的智能合约模板，并能够轻松地将真实世界的数据流应用于这些模板。

5.2 里程碑 2

在这个里程碑中，我们推出了第一版数据市场，以及底层网络层所需的功能。

WP4 (网络, 智能合约): 支持数据激励和加密

- 第一个稳定版本
- 支持几本的加密算法

- 数据许可证智能合约
- 作为一个订阅目标支持智能合约
- 支持进一步的数字签名方法, 例如 基于 X.509, SGX
- 在去中心化区块储存或者去中心化数据库上的基础存储
- 实现业务逻辑
- 压力测试, 优化可扩展性

WP5 (数据市场): 数据市场发布

- 创建一个数据市场Dapp, 让用户能够发现数据流, 订阅流, 并以免费或DATAcoin付费的方式提供新的数据流。
- 使用来自不同垂直的实时数据流来占据市场, 如下:
 - 金融市场数据: 股票价格, 期权价格, 商品价格, 加密货币, 法定汇率, 交易量和公司行为。
 - 社交媒体数据: Twitter, Instagram, Facebook, reddit, flickr等。
 - 运输数据: 出发, 到达, 地理位置, 速度, 飞机, 船舶, 火车和当地运输。
 - 天气资料: 气温, 降水, 湿度, 云层, 风速, 目前的和预测的。

WP6 (引擎, 编辑器): 实现去中心化

- 迁移新的网络层
- 部署到 Golem 或其他容器提供者
- 在去中心化环境中实现容错和故障恢复

5.3 里程碑 3

里程碑3的目标是完成Streamr的所有愿景。里程碑 3 中的研发一路上一定会改变, 因为之前的里程碑中会出现新的想法和社区会请求新的功能。里程碑3还将包含大量的营销来推动技术栈的应用。

WP7 (网络): 高级路由, 位置混淆

- 位置混淆
- 组播加密
- 压力测试, 优化可扩展性
- 解决大型部署中遇到的问题
- 研发社区请求的功能
- 与最新的平台集成

WP8 (数据市场): 社区建设

- 在数据市场中增加更多的数据流

- 流愿望清单, 为加速实现的赏金计划
- 社区建设

WP9 (引擎, 编辑器): 产品化

- 改进工具的 UI/UX
- 新手上路, 教学视频, 帮助资料
- 集成到相关的区块链平台, 物联网, AI, 或其他地方

6. 项目管理团队



Henri Pihkala, 理学硕士

Henri是一位软件工程师，连续的创业者，也是一个前算法程序交易者。他带领开发了两个高频算法交易平台，并设计和管理创建分布式Streamr云分析平台。Henri热衷于复杂的架构，可扩展性，可用性和区块链。



Risto Karjalainen, 博士

Risto是一位取得沃顿商学院博士学位的数据科学家和金融专业人士。他是自动化，系统化交易和机构资产管理方面的国际人才。Risto的兴趣包括实时计算，机器学习，进化算法，行为金融，区块链和金融科技。



Nikke Nylund, 理学学士

Nikke是一个前低延时算法交易的策略师。作为信息通信和科技公司的创始人和/或连续投资者，他具有20年的管理和创业经验，以及多项成功的战绩。Nikke拥有赫尔辛基经济学院的财务和创业学士学位。



Michael Malka, 理学硕士

Michael是一名企业家和技术爱好者，具有20年的从软件开发人员到首席执行官的不同的经验。他从事过初创企业，银行和电信的不同领域的软件项目。Michael在开始他的第一家软件公司之前，曾在赫尔辛基大学学习计算机科学。

顾问

待定

7. 结论

本白皮书概述了我们提供给去中心化应用程序的强大的本地去中心化实时数据骨干网的愿景。我们认为，实时数据市场与数据管道的结合将为以太坊智能合约开发者和整个Dapp生态系统带来巨大变革。我们的目标是建立一个深思熟虑和专业实现的技术栈，满足用户未来的需求，为永不停息的应用程序提供永不停息的数据。

我们的技术栈是分层和模块化的，它建立在去中心化的传输层上。这里存在由激励代理节点组成的点对点网络。网络托管发布/订阅机制，并支持加密事件的去中心化存储。吞吐量与参与节点的数量呈线性关系，网络可以每秒处理数百万个事件。

数据骨干是M2M经济的理想推动者，这里有自主机器，机器人和机器人购买或出售微量数据。这种想法已经被提出，机器将会交换诸如存储，处理能力，通信带宽等资源。我们相信使用DATAcoin会导致交易成本低于物物交换。

Streamr是计算革命的一部分，其中单片解决方案正在被去中心化计算所取代。在分布式计算中，Golem替代了Azure虚拟机。在块存储中，IPFS，Storj等替代Azure Blob存储。在数据管道和消息传递中，Streamr取代了中心化的消息传递和事件处理平台，如Azure EventHub和Azure Stream Analytics。公司和企业权力正在转移给独立公民，自主代理，应用程序和机器。从而提高隐私权，效率，弹性和容错能力，最终为互联社会的良好公民提供更高的福利。