

说明书
一种数据管理方法以及相关设备

本申请要求于 2024 年 01 月 11 日提交国家知识产权局的、中国申请号为 202410044120.5、发明名称为“一种数据管理方法、系统、设备和存储介质”的中国专利申请的优先权，以及要求于 2024 年 04 月 29 日提交国家知识产权局的、中国申请号为 202410532731.4、发明名称为“一种数据管理方法以及相关设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及数据处理技术领域，具体涉及一种数据管理方法以及相关设备。

背景技术

随着信息技术的不断发展，数据在各种业务领域有大量的应用需求，数据的购买也出现在各种应用场景中。例如，人工智能（Artificial Intelligence，AI）模型的性能提升充分依赖高质量的训练数据，目前，公开数据以及私有数据已难以满足 AI 模型的训练需要，因此，AI 开发者（也即是数据使用者）需要从第三方（也即是数据提供者）购买大量的高质量数据。

目前，数据提供者通常通过书面协议等形式来明确数据的使用权限，但数据使用者对数据的使用过程通常对数据提供者不够透明，导致数据提供者所提供的数据可能被滥用甚至泄密。

因此亟需一种能够对数据提供者提供的数据的使用过程进行管理的方法，以保证数据提供者所提供的数据被数据使用者安全合理地使用。

发明内容

本申请实施例提供了一种数据管理方法，可以对数据提供者提供的数据的使用过程进行管理，以保证数据提供者所提供的数据被数据使用者安全合理地使用。本申请还提供了相应的装置、设备、计算机可读存储介质和计算机程序产品等。

本申请第一方面提供一种数据管理方法，该方法应用于数据管理系统，数据管理系统设置于至少一个数据中心，数据管理系统位于隔离运行环境中，隔离运行环境通过应用程序编程接口（application programming interface，API）分别与数据提供者和数据使用者连接，数据管理系统中包括执行节点和存储节点，存储节点存储有数据提供者的第三方数据，该方法包括：执行节点运行业务程序，业务程序用于执行数据使用者的业务；执行节点向存储节点发送数据读取请求，数据读取请求指示运行的业务程序请求从存储节点读取第一数据，第一数据属于第三方数据；存储节点根据数据读取请求和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息。

在第一方面中，该数据管理系统位于隔离运行环境中，从而可以使得数据使用者难以将数据提供者提供的第三方数据直接传递至该隔离运行环境之外，使得第三方数据的使用被有效地限制在该隔离运行环境中，避免数据提供者提供的第三方数据被泄密。

并且，该第三方数据存储于数据管理系统的存储节点中，在数据管理系统的执行节点运行业务程序以执行数据使用者的业务的过程中，在需要使用第三方数据时，执行节点需要向存储节点发送数据读取请求，这样，存储节点可以基于数据读取策略控制业务程序对第三方数据的读取，使得数据使用者通过业务程序对第三方数据的使用受到存储节点的监控，从而通过存储节点对数据提供者提供的第三方数据的使用过程进行管理，以保证第三方数据被数据使用者安全合理地使用，减少数据泄露的风险。

在第一方面的一种可能的实现方式中，该方法还包括：监控数据使用者通过 API 对数据管理系统的访问行为，并监控数据管理系统通过 API 的数据输出行为。

该种可能的实现方式中，可以通过 API 网关或者其他形式的 API 管理工具对 API 进行管理，以通过 API 网关对 API 相关的数据交互进行监控。这样，通过 API 网关等，可以监控外部节点（数据使用者）通过 API 对内部节点（数据管理系统）的访问行为，并监控数据管理系统通过 API 的数据输出行为。

其中，可以通过监控访问频率、相应的响应大小、响应内容等中的一种或多种来监控访问行为，从

而可以识别异常访问行为。而监控数据输出行为可以是检测该数据输出行为所待输出的数据与数据提供者提供的第三方数据的相似性，从而避免第三方数据的泄露。

在第一方面的一种可能的实现方式中，数据读取策略由数据提供者确定，数据读取策略包括策略项和基于策略项确定的操作，策略项包括以下项中的一种或多种：数据策略项、用途策略项、数量策略项、时间策略项、主体策略项；其中，数据策略项用于描述第三方数据中能够被读取的数据范围；用途策略项用于描述业务程序中能够使用第三方数据的任务；数量策略项用于描述第三方数据的使用量阈值；时间策略项用于描述第三方数据的使用时间范围；主体策略项用于描述能够应用数据读取策略的主体类型。

该种可能的实现方式中，该数据读取策略可以通过一个或多个策略项从一个或多个方面进行灵活配置，配置方式清晰且易于实施，能够提高配置效率，便于数据使用者和数据提供者理解，从而便于数据使用者和数据提供者进行协商和统一。

在第一方面的一种可能的实现方式中，存储节点包括关于第三方数据的一个或多个数据读取策略；存储节点根据数据读取请求和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息，包括：存储节点将数据读取请求与一个或多个数据读取策略中的至少一个数据读取策略的策略项进行匹配；若存在与数据读取请求相匹配的数据读取策略，则存储节点根据与数据读取请求相匹配的数据读取策略中的操作，生成反馈信息并向执行节点发送反馈信息。

该种可能的实现方式中，待匹配的数据读取策略的数量可以有一个或多个。当待匹配的数据读取策略的数量有多个时，可以依次匹配该多个数据读取策略，具体地，可以根据数据读取策略中的策略项进行匹配。依次匹配该多个数据读取策略，直到存在相匹配的数据读取策略，并根据该数据读取策略中的操作，生成反馈信息，或者，直到确定所有的数据读取策略均不匹配，则可以根据默认操作生成反馈信息，或者可以生成指示拒绝数据读取的反馈信息。

在第一方面的一种可能的实现方式中，存储节点还包括数据权限策略，数据权限策略用于根据数据的来源，确定对数据的读写权限，其中，数据权限策略指示数据使用者对来自数据提供者的第三方数据具有读权限；在存储节点根据第一数据和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息之前，还包括：存储节点根据数据权限策略，基于第一数据属于第三方数据，确定业务程序对第一数据具有读权限。

该种可能的实现方式中，该数据权限策略可以为数据管理系统的系统级策略，从而可以通过该系统级策略，对存储节点中的数据的读写权限进行基础管理。该数据权限策略中可以包括指示数据使用者对来自数据提供者的第三方数据具有读权限的指示信息。这样，可以确定对第一数据具有读权限之后，再根据数据读取策略进行数据读取。

可见，该种可能的实现方式中，可以通过至少两级策略，更完善地对存储节点中的数据的读取操作进行管理。

在第一方面的一种可能的实现方式中，数据管理系统包括管理节点；该方法还包括：管理节点获取对业务程序的运行指令之后，创建对存储节点的访问凭证；管理节点指示执行节点运行业务程序，并向执行节点传递访问凭证的信息；执行节点向存储节点发送数据读取请求，包括：执行节点基于访问凭证的信息，向存储节点发送数据读取请求。

该种可能的实现方式中，可以通过管理节点管理执行节点对存储节点的访问凭证，也即是说，该管理节点可以管理执行节点与存储节点进行交互的权限。这样，执行节点无法自行获得对存储节点的访问权限，避免了执行节点在未被允许的情况下对存储节点的第三方数据进行数据读取和使用，从而避免了执行节点对第三方数据的泄露和不合理使用。

在第一方面的一种可能的实现方式中，该方法还包括：在执行节点停止运行业务程序之后，删除访问凭证。

该种可能的实现方式中，在执行节点停止运行业务程序之后，删除访问凭证可以保证访问凭证的有效性。

这样，执行节点在每次启动业务程序的运行时，均需要通过管理节点重新获得执行节点与存储节点进行交互的权限。这样，执行节点无法自行获得对存储节点的访问权限，也无法滥用已获得的访问权限，有效保证执行节点对存储节点的每次访问的可控性，避免了执行节点在未被允许的情况下对存储节点的

第三方数据进行数据读取和使用，从而避免了执行节点对第三方数据的泄露和不合理使用。

在第一方面的一种可能的实现方式中，在存储节点根据第一数据和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息之后，还包括：存储节点记录数据读取请求相关的数据访问行为。

5 该种可能的实现方式中，存储节点存储数据访问行为的信息之后，数据提供者可以方便灵活地查看其存储于存储节点的第三方数据的使用情况。

在第一方面的一种可能的实现方式中，该方法还包括：存储节点获取数据提供者输入的查询请求，查询请求用于请求查询第三方数据相关的访问记录；存储节点输出访问记录，访问记录包括数据读取请求相关的数据访问行为。

10 该种可能的实现方式中，数据提供者可以通过存储节点方便地查询第三方数据被数据使用者访问和使用的情况，使得第三方数据的使用透明化。此外，数据提供者还可以及时发现第三方数据的异常使用，并及时调整数据读取策略。数据提供者还可以通过数据管理系统，向数据使用者发起策略协商流程，以将数据管理系统作为中间平台，与数据使用者协商出新的数据读取策略，以满足数据提供者和数据使用者双方的要求。

15 在第一方面的一种可能的实现方式中，该方法还包括：在运行业务程序的过程中，执行节点向存储节点发送待存储数据；存储节点在接收到待存储数据之后，根据指定时间段内执行节点从存储节点读取的第三方数据，确定并存储待存储数据的衍生链路，指定时间段包括执行节点本次运行业务程序的时间段，衍生链路用于指示存储节点中，用于生成待存储数据的数据。

20 该种可能的实现方式中，执行节点在运行业务程序的过程中，在从存储节点读取到第三方数据之后，在对第三方数据进行应用的过程中可能产生衍生数据。为了保证数据的安全性，该种可能的实现方式中，可以通过衍生链路对该衍生数据进行跟踪和管理，避免衍生数据的泄露。

在第一方面的一种可能的实现方式中，该方法还包括：在执行节点停止运行业务程序之后，删除执行节点中业务程序相关的运行数据。

25 该种可能的实现方式中，可以有效保证运行业务程序的过程中使用第三方数据的使用痕迹被删除，避免执行节点在不合理地保留运行业务程序的过程中所使用的第三方数据，从而保证该第三方数据不被泄露和滥用。

本申请第二方面提供一种数据管理系统，该系统具有实现上述第一方面或第一方面任意一种可能实现方式的方法的功能。该功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。该硬件或软件包括一个或多个与上述功能相对应的模块，例如管理节点、执行节点以及存储节点。

30 本申请第三方面提供一种计算设备集群，该计算设备集群包括至少一个计算设备，该至少一个计算设备包括处理器和存储器，至少一个计算设备的存储器中存储有可在处理器上运行的计算机执行指令，当计算机执行指令被处理器执行时，处理器执行如上述第一方面或第一方面任意一种可能的实现方式的方法。

35 本申请第四方面提供一种存储一个或多个计算机执行指令的计算机可读存储介质，当计算机执行指令被处理器执行时，处理器执行如上述第一方面或第一方面任意一种可能的实现方式的方法。

本申请第五方面提供一种存储一个或多个计算机执行指令的计算机程序产品，当计算机执行指令被处理器执行时，处理器执行如上述第一方面或第一方面任意一种可能的实现方式的方法。

40 本申请第六方面提供了一种芯片系统，该芯片系统包括处理器，用于支持处理器实现上述第一方面或第一方面任意一种可能的实现方式中所涉及的功能。在一种可能的设计中，芯片系统还可以包括存储器，存储器用于保存必要的程序指令和数据。该芯片系统，可以由芯片构成，也可以包含芯片和其他分立器件。

其中，第二方面至第六方面或者其中任一种可能实现方式所带来的技术效果可参见第一方面或第一方面的相关可能实现方式所带来的技术效果，此处不再赘述。

45 **附图说明**

图 1 是本申请实施例提供的一种示例性流程示意图；

- 图 2 是本申请实施例提供的 AI 平台的一种示例性系统框架示意图；
图 3a 是本申请实施例提供的逻辑多租场景中的开发空间的一种示例性示意图；
图 3b 是本申请实施例提供的物理多租场景中的开发空间的一种示例性示意图；
图 4 是本申请实施例提供的隔离运行环境的一示例性示意图；
5 图 5 是本申请实施例提供的在 AI 模型的训练场景中的管理节点、执行节点以及存储节点的一示例性示意图；
图 6 是本申请实施例提供的数据读取策略的策略项的一示例性示意图；
图 7 是本申请实施例提供的数据管理方法的一示例性示意图；
图 8 是本申请实施例提供的一种示例性流程示意图；
10 图 9 是本申请实施例提供的一种示例性流程示意图；
图 10 是本申请实施例提供的衍生链路的一示例性示意图；
图 11 是本申请实施例提供的衍生链路的一示例性示意图；
图 12 是本申请实施例提供的数据管理系统的一示例性示意图；
图 13 是本申请实施例提供的计算设备的一结构示意图；
15 图 14 是本申请实施例提供的计算设备集群的一结构示意图；
图 15 是本申请实施例提供的计算设备集群的一结构示意图。

具体实施方式

下面结合本申请实施例中的附图对本申请实施例进行描述。本申请的实施方式部分使用的术语仅用于对本申请的具体实施例进行解释，而非旨在限定本申请。

本领域普通技术人员可知，随着技术的发展和场景的出现，本申请实施例提供的技术方案对于类似的技术问题，同样适用。

本申请中，“至少一个”是指一个或者多个，“多个”是指两个或两个以上。“和/或”，描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 的情况，其中 A、B 可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项”或其类似表达，是指的这些项中的任意组合，包括单项或复数项的任意组合。本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换，这仅仅是描述本申请的实施例中对相同属性的对象在描述时所采用的区分方式。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，以便包含一系列单元的过程、方法、系统、产品或设备不必限于那些单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它单元。

在许多应用场景中，都涉及第三方数据的购买业务。

例如，在 AI 模型训练的场景中，AI 开发者（也即是数据使用者）需要从第三方（也即是数据提供者）购买大量的高质量数据。

目前，一种约束数据使用者对购买的数据的使用权限的方式为通过书面合约的方式来明确数据使用者对购买的数据的使用权限。然而，由于难以监管数据使用者对数据的使用过程，也即是说数据使用者对数据的使用过程通常对数据提供者不够透明，导致数据提供者所提供的数据（为了描述方便，后续称为第三方数据）可能被滥用甚至泄密。

目前，一种常用的数据使用控制方式为采用用户-角色-权限的授权模型。数据提供者给数据使用者分配最小权限，限制数据使用的权限。然而，这一控制方式较为简单，例如无法控制数据使用者对第三方数据的使用次数，也无法记录数据使用者对第三方数据的使用情况，使得数据提供者无法感知第三方数据的使用情况。此外，数据使用者依然可以在通过应用程序使用第三方数据的过程中对第三方数据及其衍生数据进行复制和输出，从而导致数据泄露。

而另一种数据使用控制方式为数据提供者通过权限管理服务，将对第三方数据的使用权限附加至第三方数据中，并加密。例如，对于文档类型的第三方数据，数据提供者可以对该第三方数据设置使用权

限并加密。然后，数据使用者可以获取该加密后的第三方数据并通过应用该文档的应用程序打开。该应用程序可以进行解密，获得解密后的文档，并根据文档中附加的使用权限控制数据使用者对该文档的使用。

然而，在该种数据使用控制方式中，需要对数据进行加解密，在数据较多时耗时较长，且数据使用者可以通过该应用程序获取到第三方数据的明文数据，因此导致该第三方数据容易被数据使用者通过该应用程序进行复制并转发，从而导致泄露，并且使得数据提供者难以对应用程序中的第三方数据以及衍生数据的使用和删除等进行管理。

可见，传统的数据使用控制方式均难以保证数据提供者所提供的第三方数据被数据使用者安全合理地使用。

基于此，本申请实施例提供了一种数据管理方法，可以对数据提供者提供的第三方数据的使用过程进行管理，以保证数据提供者所提供的第三方数据被数据使用者安全合理地使用。

如图1所示，本申请实施例的数据管理方法主要涉及以下方面中的一个或多个：

构建隔离运行环境、配置策略、使用策略进行数据读取、衍生数据管理、停止运行业务程序。

下面分别对各个方面进行示例性介绍。

一、构建隔离运行环境。

该数据管理方法可以应用于数据管理系统，该数据管理系统位于隔离运行环境中。

该隔离运行环境可以通过计算设备集群而实现，其中，该计算设备集群可以包括一个或多个计算设备。

其中，任一计算设备的类型在此不做限制。示例性地，任一计算设备可以为终端设备，也可以为服务器、服务器集群、容器（container）或者虚拟机（virtual machine）等。当计算设备集群包括多个计算设备时，不同的计算设备的类型可以相同，也可以存在差异。

该一个或多个计算设备可以包含于一个或多个数据中心中，而该数据管理系统可以为基于数据中心中的资源而构建的系统。

该数据管理系统所在的隔离运行环境的具体形式可以有多种。

例如，该隔离运行环境可以位于通过计算设备集群实现的云平台或者AI平台中，也可以部署于服务器并通过网络或应用程序编程接口（application programming interface, API）提供数据计算能力的云服务、应用程序，例如函数云服务、在线文档编辑器、网盘等。

下面以AI平台为例进行示例性介绍，需要说明的是，该隔离运行环境并不限于AI平台。

在AI模型的训练场景中，位于AI平台的数据管理系统可以执行的业务包括AI模型训练。

如图2所示，为AI平台的一种示例性系统框架示意图。

如图2所示的示例中，该AI平台可以对诸如数据使用者和数据提供者等用户提供访问接口。

例如，通过控制台（console）提供接口，用户可以通过浏览器等访问AI平台的console，从而访问AI平台；或者，用户的客户端设备中的应用可以通过API网关访问AI平台提供的API，从而访问AI平台。

该AI平台还可以提供数据管理服务、标注服务、训练服务、推理服务以及存储服务中的一个或多个，并可以对诸如数据使用者和数据提供者等用户提供访问接口。

标注服务、训练服务以及推理服务可以通过相应的应用程序，基于相应的算法等来提供各自的服务，并可以通过存储服务提供的存储软件开发工具包（software development kit, SDK）访问存储服务。

存储服务可以存储有数据，例如，可以存储有数据使用者的自有数据、数据提供者的第三方数据以及基于其他数据（例如自有数据和第三方数据等）而生成的衍生数据。

存储服务还可以存储有元数据，该元数据可以包括自有数据、第三方数据以及衍生数据的信息，还可以包括后续步骤中的相关策略，以用于对第三方数据等数据的使用进行控制。此外，存储服务还可以存储有对数据的访问记录等，以便于数据提供者等了解第三方数据的使用情况。

这样，数据使用者可以访问AI平台，以在该AI平台中，利用AI平台的资源进行AI模型训练。

并且，数据提供者可以通过AI平台存储并售卖该数据提供者的第三方数据。这样，数据使用者可以通过AI平台，购买数据提供者提供的第三方数据来进行AI模型的训练。

其中，本申请实施例中，该数据使用者可以理解为人、企业或者组织，也可以理解为客户端等设备、账号、网络地址或者通过其他形式的标识所表示的主体，该数据使用者的标识和具体形式可以有多种；类似地，该数据提供者可以理解为人、企业或者组织，也可以理解为客户端等设备、账号、网络地址或者通过其他形式的标识所表示的主体，该数据提供者的标识和具体形式可以有多种，申请实施例中对此不做限制。

在该示例中，可以在 AI 平台中构建隔离运行环境，并在该隔离运行环境中实现数据管理系统。

其中，如图 3a 所示，该 AI 平台向租户（可以为数据使用者）提供的开发空间是逻辑多租时，实现 AI 平台的计算设备集群可以处理多个租户的请求，也即是说，多个租户（例如图 3a 所示的租户 1 的开发空间 1 和租户 2 的开发空间 2）共享 AI 平台的资源，则该隔离运行环境可以为 AI 平台，其中，该隔离运行环境可以包括 AI 平台的网络资源、计算资源和存储资源。

而如图 3b 所示，该 AI 平台向租户（可以为数据使用者）提供的开发空间是物理多租时，租户从该 AI 平台租用的开发空间的资源为该租户独享，AI 平台中的各个租户间的资源充分隔离。也即是说，如图 3b 所示的示例中，租户 1 的开发空间 1 中的网络资源、计算资源和存储资源为隔离运行环境 1，而租户 2 的开发空间 2 中的网络资源、计算资源和存储资源为隔离运行环境 2，隔离运行环境 1 与隔离运行环境 2 之间相互隔离。

可见，本申请实施例中，该隔离运行环境可以从网络、计算以及存储中的一个或多个方面来实现。下面分别进行介绍。

1、网络隔离

如图 4 所示的示例中，该隔离运行环境通过 API 与该隔离运行环境之外的其他节点连接，例如，该隔离运行环境通过该 API 分别与数据提供者和数据使用者连接。

这样，该隔离运行环境中的内部节点（例如数据管理系统中的执行节点和存储节点）与隔离运行环境外的外部节点（例如数据使用者的客户端设备）不能直接相通。

这样，可以通过监控 API，监控内部节点与外部节点之间的信息传输。

具体地，在一些实施例中，可以监控数据使用者通过 API 对数据管理系统的访问行为，并监控数据管理系统通过 API 的数据输出行为。

示例性地，可以通过 API 网关或者其他形式的 API 管理工具对 API 进行管理。数据管理系统可以将 API 注册到 API 网关，以通过 API 网关对 API 相关的数据交互进行监控。

API 网关对内可以访问隔离运行环境中的内部节点。该 API 网关可以部署于该隔离运行环境中，例如，该 API 网关可以为云平台中的弹性负载均衡（elastic load balance, ELB）服务，或者，该 API 网关也可以位于该隔离运行环境之外，通过虚拟私有云（virtual private cloud, VPC）对等连接、源地址转换（source network address translation, SNAT）、弹性 IP（elastic Internet Protocol, EIP）等技术连通该隔离运行环境。

此外，API 网关对外可以被外部节点访问，例如，可以通过云平台中的 ELB 服务连通外部节点所在的外部网络。

并且，还可以通过 API 网关等控制隔离运行环境的网络流。例如，该网络流可以是单向的，可以是外部节点到 API 网关，再由 API 网关到内部节点。也即是说，仅可以由外部节点通过 API 向内部节点发出请求，再由内部节点处理该请求后，通过 API 向外部节点发送响应，而不能由内部节点在未接收到来自外部节点的请求的情况下，主动通过 API 向外部节点发送数据。

这样，通过 API 网关等，可以监控外部节点（数据使用者）通过 API 对内部节点（数据管理系统）的访问行为，并监控数据管理系统通过 API 的数据输出行为。

其中，可以通过监控访问频率、相应的响应大小、响应内容等中的一种或多种来监控访问行为，从而可以识别异常访问行为。而监控数据输出行为可以是检测该数据输出行为所待输出的数据与数据提供者提供的第三方数据的相似性，从而避免第三方数据的泄露。

该隔离运行环境的具体形式在此不做限制。

示例性地，在云平台中，该隔离运行环境可以为使用虚拟私有云（virtual private cloud, VPC）创建独立的网络环境。在非云平台中，该隔离运行环境可以为小型网络环境，也可以为使用防火墙隔离

出的独立的网络区域。

2、计算隔离

在一些示例中，隔离运行环境中的数据管理系统包括管理节点、执行节点和存储节点，该管理节点可以用于执行数据管理系统的系统操作，该执行节点用于运行业务程序，该业务程序用于执行数据使用者的业务，该存储节点可以用于提供存储服务，例如，存储数据提供者提供的第三方数据。

其中，管理节点、执行节点以及存储节点等节点的具体形态，上述各个节点的形态可以相同，也可以不同，在此不做限制。

示例性地，任一节点可以为数据管理系统中的软件模块或者插件；或者，可以为虚拟机或者容器；或者，也可以为实现数据管理系统的计算设备集群中的硬件，例如，执行节点为计算设备集群中的服务器或者服务器集群，存储节点为计算设备集群中的存储设备。

本申请实施例中，执行数据使用者的业务（例如进行 AI 模型训练）的业务程序由执行节点执行，以便于数据管理系统对业务程序的运行和数据交互进行隔离管理，避免业务程序在执行过程中泄露第三方数据以及相关的衍生数据等。

并且，管理节点还可以对执行节点所运行的业务程序进行监控，例如，可以监控使用频繁的业务程序，以及及时发现业务程序的异常。

此外，数据管理系统还可以限制用户对于业务程序的执行结果、执行日志等执行数据的查看权限，例如，可以限制用户对执行数据的查看规模和查看频率等，举例来说，用户在指定时间段内最多可以查看 1000 条执行结果，或者，最多可以查看 10MB 的执行日志等。

3、存储隔离

该数据管理系统中的数据管理系统的存储节点可以提供存储服务。

其中，数据使用者可以向存储节点传输、下载、删除数据使用者的自有数据。

数据提供者可以向存储节点传输数据提供者提供的第三方数据，并通过存储节点存储该第三方数据。

这样，数据使用者可以向存储节点申请对第三方数据的使用权限，例如，可以通过购买的方式，获得对第三方数据的使用权限。这样，可以通过存储节点实现对数据提供者的第三方数据的使用管理。

在实际场景中，存储节点中存储的第三方数据以及相关衍生数据等不能被直接下载，若数据使用者或者其他用户需要从存储节点下载第三方数据，可以通过数据管理系统，向数据提供者发起对第三方数据的下载请求，当该第三方数据的数据提供者通过数据管理系统审批通过该下载请求之后，才允许数据使用者或者其他用户从存储节点下载相应的第三方数据。

如图 5 所示，以该数据管理系统为 AI 平台为例，对管理节点、执行节点以及存储节点在 AI 模型的训练场景中的相关功能进行示例性介绍。

在图 5 所示的示例中，AI 平台的管理节点可以提供平台级的 AI 服务。

AI 服务可以基于数据使用者的指令，触发执行节点执行业务服务。

具体地，AI 服务在执行业务程序时，先初始化执行节点，然后将业务程序分配到执行节点上执行。业务程序可以使用 AI 平台的基础设施（例如图形处理器（graphics processing unit， GPU）集群等）进行 AI 模型的训练和推理。

在执行业务程序的过程中，执行节点可以将执行过程中产生的临时数据写入到本地存储中，在业务程序执行完毕后，本地存储被清空释放。执行节点可以将执行过程中产生的日志写入到日志服务中，或者写入到本地存储由日志代理采集到日志服务。执行节点还可以将 AI 模型等持久化数据写入到存储服务中。用户需要通过 AI 服务查看日志、持久化数据，在 AI 服务中配置策略，以进行数据的使用控制等。

二、配置策略。

本申请实施例中，数据管理系统中可以包括以下策略中的一个或多个：

数据权限策略、数据读取策略。

通过上述策略，数据管理系统可以对存储节点中的数据的读取操作进行管理。

下面分别对上述策略进行示例性介绍。

1、数据权限策略

在该示例中，数据权限策略用于根据数据的来源，确定对数据的读写权限。

在一些示例中，该数据权限策略可以为系统级的策略，也即是说，该数据权限策略可以应用于数据管理系统，并由数据管理系统的开发者预先进行配置。

其中，数据的来源可以包括以下中的一种或多种：自有数据、第三方数据、衍生数据。其中，自有数据为数据使用者提供的数据，第三方数据为数据提供者提供的数据，衍生数据为基于第三方数据等数据而获得的数据。在5 一些示例中，存储节点可以将执行节点执行业务程序时要待存储至存储节点的数据（也称为待存储数据）认为是衍生数据。

在一些示例中，数据权限策略还可以用于根据数据的内容，确定对数据的读写权限。

其中，数据的内容可以有多种分类方式，举例来说，数据的内容可以包括普通数据和模型数据，除10 模型数据之外的数据均可以认为是普通数据。

此外，数据权限策略还可以用于根据数据的用途，确定对数据的读写权限。

下面通过一个具体示例，说明该数据权限策略的一种示例性内容。

表 1：数据权限策略

用途	读/写	数据类型	行为	备注
推理用途	读	模型数据	允许	
	读	普通数据	拒绝	
	写	模型数据和模型数据	拒绝	推理服务不应当写入数据。
其他用途	读	自有数据	允许	
	写	自有数据	拒绝	自有数据应通过数据管理系统的管理节点投递，而不是在通过执行节点的业务程序写入存储节点。
	读、写	衍生数据	允许	
	读	第三方数据	具有权限，行为基于数据读取策略确定	
	写	第三方数据	拒绝	第三方数据应通过数据管理系统的管理节点投递，而不是在通过执行节点的业务程序写入存储节点。

15 可见，在表 1 所示的示例中，数据权限策略可以指示对来自数据提供者的第三方数据具有读权限，并从而可以在后续根据数据读取策略，确定业务程序能够从存储节点读取到的第三方数据。

2、数据读取策略

本申请实施例中，该数据读取策略可以为对相应的第三方数据的读取策略。该第三方数据的具体类型在此不做限制。例如，该第三方数据可以为一个文件，也可以为一组文件（例如目录、压缩数据包等）。

20 该第三方数据可以包括类型、名称、路径、行数、存储大小等属性。

本示例中，某一第三方数据对应的数据读取策略可以由该第三方数据的数据提供者确定。具体地，可以由数据管理系统提供配置界面，并由第三方数据的数据提供者通过该配置界面输入配置信息，以配置该数据读取策略；或者，也可以是数据管理系统向该第三方数据的数据提供者提供候选数据读取策略（可以由历史数据读取策略或者默认数据读取策略而得到），并由该数据提供者将该候选数据读取策

略确定为对该第三方数据的数据读取策略。

该数据读取策略包括策略项和基于策略项确定的操作。该策略项的数量和内容可以有多种情况。其中，任一策略项的对象可以为数据本身，也可以是数据中的文件，每个策略项的对象的具体类型可以预先配置，例如，通过 PolicyType 来定义，具体地，PolicyType 为 DATA 指示策略项的对象是数据本身，PolicyType 为 FILE 指示策略项的对象是文件。基于策略项确定的操作为策略项匹配后所指示的动作，例如，可以包括允许读取、拒绝读取等。

如图 6 所示，在一些示例中，该策略项包括以下项中的一种或多种：
数据策略项、用途策略项、数量策略项、时间策略项、主体策略项。
下面，分别对多种示例性的策略项进行示例性介绍。

1) 数据策略项 (Data Policy Item)

数据策略项用于描述第三方数据中能够被读取的数据范围。该数据策略项可以对数据进行切分，并切分后的数据中确定出能够被读取的数据。

示例性地，该数据策略项可以包括路径过滤(Path Filter)、选择列(Columns)和行过滤(Row Filter)。

路径过滤指从第三方数据包含的多个文件中筛选出部分文件，默认设置为选择全部文件。具体地，可以使用文件扩展名、通配符、正则表达式等方式进行过滤。其中，通配符可以使用 Ant 风格等较为简单易用的路径匹配表达式。

选择列指选定 1 个或多个列，默认设置为选择所有列。对逗号分隔值 (comma-separated values, CSV)、JS 对象简谱 (JavaScript object notation, JSON)、Excel 等数据类型的第三方数据，这些第三方数据在传输至数据管理系统时可以定义列，若某一第三方数据未定义有列，则不适用选择列这一内容。例如图片、视频等二进制文件类型的第三方数据不适用选择列这一内容。

行过滤指对相应的第三方数据按行进行过滤，只能读取满足条件的行，如 “name in (‘beijing’, ‘nanjing’) AND age > 30”，其中 name 和 age 是数据的列。对 CSV、JSON、Excel 等数据类型的第三方数据，这些第三方数据在传输至数据管理系统时可以定义列，若某一第三方数据未定义有行，则不适用行过滤这一内容。对图片、视频等二进制文件类型的第三方数据，可以对相应的元数据进行过滤，如 “imageSize < 10MB AND imageWidth < 1024”，其中 imageSize 描述图片大小，imageWidth 描述图片宽度。

示例性地，可以通过以下代码，对数据策略项 (Data Policy Item) 进行配置：

```
struct DataPolicyItem {  
    DataPathFilter pathFilter;  
    DataSetColumns columns;  
    DataSetRowFilter rowFilter;  
}
```

2) 用途策略项 (UsagePolicy Item)

用途策略项用于描述业务程序中能够使用第三方数据的任务，也即是用于描述第三方数据能够被哪些功能使用。

例如，在 AI 模型的应用场景中，所涉及的功能可以包括以下中的一种或多种：

- a) Labeling: 标注;
- b) Feature: 特征工程;
- c) Train: 训练;
- d) Reasoning: 推理。

示例性地，可以通过以下代码，对用途策略项 (UsagePolicy Item) 进行配置：

```
struct UsagePolicyItem {  
    String[] usageList;  
}
```

3) 数量策略项 (Count Policy Item)

数量策略项用于描述第三方数据的使用量阈值，也即是说，数量策略项用于描述第三方数据中能够

被使用的量。该使用量可以有多种情况，该数据策略项的对象为数据或者文件时，该使用量的具体内容可以不同。

在一种示例中，PolicyType 为 DATA，指示策略项的对象是数据本身。此时，数据策略项的使用量可以为使用次数，具体地，该数据策略项用于描述第三方数据的使用次数阈值（maxUseTimes）。

5 该使用次数阈值可以是该第三方数据最多能够被使用的次数。其中，该使用次数的计算规则可以由用户或者数据管理平台定义，该计算规则包括但不限于：

a) 仅算一次：算法程序在一次运行期间，读取数据一次或以上时，useTimes 仅累加一次。多次运行会累计计算。

10 b) 一段时间内多次使用仅算一次：算法程序在一段时间读取数据一次或以上时，useTimes 仅累加一次。

c) 一次使用算作一次：算法程序每使用一次数据，useTimes 就累加一次。

而在另一种示例中，PolicyType 为 FILE，指示策略项的对象是文件。此时，使用量阈值可以包括以下中的一种或多种：

15 a) 最大行数（maxRowCount）指单个文件最多可使用的行数。而对于图片、视频等非结构化的第三方数据，不适用该使用量阈值。

b) 最大存储大小（maxStorageSize）指单个文件最多可使用的数据量。

示例性地，可以通过以下代码，配置数量策略项（Count Policy Item）：

```
20 struct CountPolicyItem {  
    int maxRowCount;  
    int maxStorageSize;  
    int maxUseTimes;  
}
```

4) 时间策略项（Time Policy Item）

时间策略项用于描述第三方数据的使用时间范围。

25 例如，该时间策略项可以指定具体的截止时间，例如可以采用 RFC 3339 日期格式，并且可以包含时区，例如该介质时间可以设置为 2023-12-01T12:21:32Z。

或者，该时间策略项可以指定使用时长，例如，可以从第三方数据传递至数据管理系统时的时间为起始时间，该使用时长可以有多种表示方式，例如使用 RFC 3339 中的 Duration 来表示，举例来说，该使用时长表示为 P2DT3H4M，指示使用时长为“2 天，3 小时和 4 分钟”。

30 示例性地，可以通过以下代码，配置时间策略项（Time Policy Item）：

```
struct TimePolicyItem {  
    ZonedDateTime deadline;  
    Duration duration;  
}
```

35 5) 主体策略项（Subject Policy Item）

主体策略项用于描述能够应用数据读取策略的主体类型，从而可以明确相应的数据读取策略用于的主体对象的类型。

考虑到数据管理系统可能被多个用户使用，通过主体策略项可以为不同的用户指定不同的数据读取策略。

40 该主体类型具体可以有多种情况，可以根据具体的应用场景而确定。

示例性地，该主体类型可以包括以下中的一种或多种：

a) 用户 User：数据管理系统所涉及的账号，例如可以为数据管理系统所在的云管理平台或者 AI 平台中的用户，具有独立的身份标识。

45 b) 用户组 Group：用户组是用户的集合，可以通过主体策略项确定对适用于用户组的数据读取策略，从而简化相关读取权限的授权操作。

c) 角色 Role：角色是一组权限的集合，根据用户的工作职能进行定义。

d) 组织单元 `OrganizationalUnit`: 组织单元可以映射为企业或者组织的部门、子公司、项目族等。用户可以归属到一个组织单元中。

示例性地, 可以通过以下代码, 对主体策略项 (`Subject Policy Item`) 进行配置:

```
struct SubjectPolicyItem {  
    UserId[] users;  
    GroupId[] groups;  
  
    RoleId[] roles;  
    OrganizationalUnit[] orgUnits;  
}
```

基于上述任一示例, 可以对第三方数据配置一个或多个数据读取策略。

其中, 该数据读取策略的表述方式可以有多种。

例如, 面向数据管理系统以及其中的业务程序等应用时, 数据读取策略可以通过开放数字版权语言 (`open digital rights language, ODRL`)、可扩展标记语言 (`extensible markup language, XML`)、`JSON`、`YAML` 等方式进行表述。

而面向用户 (例如数据提供者) 时, 可以使用文本进行表述, 以便于用户理解。

例如, 通过文本表述数据读取策略时, 可以使用 “【允许/拒绝】【主体】在【时间】使用【数量】的数据用于【用途】” 的格式, 如 “允许 `XXX` 在 3 月内使用不超过 1GB 的头像图片数据用于训练”。

可见, 该数据读取策略可以通过一个或多个策略项从一个或多个方面进行灵活配置, 配置方式清晰且易于实施, 能够提高配置效率, 便于数据使用者和数据提供者理解, 从而便于数据使用者和数据提供者进行协商和统一。

基于上述策略, 存储节点中可以记录有每个数据对应的元数据信息。

示例性地, 每个数据的元数据信息的具体内容可以如表 2 所示。

表 2: 存储节点中存储的数据的元数据信息

元数据名称	说明	备注
来源类型	自有数据、第三方数据、衍生数据	
内容类型	普通数据、模型数据	
结构定义	定义结构化信息, 例如列的名称、类型、位置等	可选。未进行结构定义时, 表示相应数据为二进制文件。
策略	策略的面向应用的表述	<code>ODRL</code> 、 <code>XML</code> 、 <code>JSON</code> 、 <code>YAML</code> 等格式
衍生自	数据 ID	仅来源类型为 “衍生数据” 时存在。
创建时间		创建时间为使用时间范围的初始时间。 第三方数据的创建时间是存储至存储节点的时间。 衍生数据的创建时间与其来源的创建时间一致, 或者为其来源的创建时间中的最近时间

这样, 可以通过存储节点中的数据的元数据信息, 对数据的使用进行管理。具体的管理方式可以参考使用策略进行数据读取的处理阶段等相关实施例, 在此不再赘述。

三、使用策略进行数据读取。

在本申请实施例中, 在实际应用场景中, 可以根据数据读取策略和/或数据权限策略, 对数据使用者对存储节点中的第三方数据的读取操作进行管理。

具体地，如图 7 所示，在一些实施例中，该方法包括步骤 701-703。

步骤 701，执行节点运行业务程序。

业务程序用于执行数据使用者的业务。

本申请实施例中，执行节点用于运行业务程序，以执行数据使用者的业务。

5 该业务程序的具体功能在此不做限制。

例如，在 AI 模型训练的场景中，该业务程序用于进行 AI 模型的训练，例如，可以调用 AI 平台的 GPU 集群等资源，以及存储节点中存储的训练数据进行 AI 模型训练。

在一些示例中，该执行节点可以根据数据管理系统中的管理节点的指令而创建，并开始执行业务程序。

10 例如，具体地，管理节点可以接收来自数据使用者的指令，以指示执行指定的业务。

管理节点接收到该指令之后，可以初始化执行节点，并指示执行节点执行业务程序，以执行数据使用者的业务。

在执行节点完成该业务程序的本次执行之后，可以删除执行节点中存储的关于本次执行的相关运行数据，例如临时数据等，但该执行节点可以保留，以执行该数据使用者的其他业务，或者在后续再次执行该业务程序；或者，在执行节点完成该业务程序的本次执行之后，可以销毁该执行节点。

15 步骤 702，执行节点向存储节点发送数据读取请求。

数据读取请求指示运行的业务程序请求从存储节点读取第一数据，第一数据属于第三方数据。

本申请实施例中，执行节点在执行业务程序的过程中，需要使用属于第三方数据的第一数据时，可以生成并向存储节点发送数据读取请求。

20 可见，执行节点对第三方数据的使用受到存储节点的控制，也即是说，执行节点无法在未被监控的情况下，获取到第三方数据，相比于通过使用第三方数据的业务程序来控制对第三方数据的使用的传统数据控制方法，本申请实施例的方案可以更好地对使用第三方数据的业务程序进行监控，从而更好地保证第三方数据不被滥用和泄密。

步骤 703，存储节点根据数据读取请求和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息。

25 该存储节点中，可以存储有一个或多个数据读取策略，此外，在一些示例中，还可以存储有数据权限策略。

本申请实施例中，可以根据数据读取请求，匹配关于第三方数据的数据读取策略，还可以匹配关于第三方数据的数据权限策略。根据匹配结果的不同情况，该反馈信息的具体内容可能不同。

30 例如，在根据匹配结果确定允许数据读取的情况下，该反馈信息可以包括允许读取的数据；而在根据匹配结果确定不允许数据读取的情况下，该反馈信息可以指示拒绝数据读取请求的数据读取操作。

下面数据权限策略和/或数据读取策略相关的匹配过程进行示例性介绍。

1、进行数据权限策略匹配。

具体地，在一些实施例中，存储节点还包括数据权限策略，数据权限策略用于根据数据的来源，确定对数据的读写权限，其中，数据权限策略指示数据使用者对来自数据提供者的第三方数据具有读权限；在存储节点根据第一数据和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息之前，还包括：

存储节点根据数据权限策略，基于第一数据属于第三方数据，确定业务程序对第一数据具有读权限。

40 本申请实施例中，该数据权限策略的相关内容可以参考上述配置策略部分的相关内容，在此不再赘述。

该数据权限策略可以为数据管理系统的系统级策略，从而可以通过该系统级策略，对存储节点中的数据的读写权限进行基础管理。

在本申请实施例中，该数据权限策略中可以包括指示数据使用者对来自数据提供者的第三方数据具有读权限的指示信息。这样，可以确定对第一数据具有读权限之后，再根据数据读取策略进行数据读取。

45 可见，本申请实施例中，可以通过至少两级策略，更完善地对存储节点中的数据的读取操作进行管理。

2、进行数据读取策略匹配。

具体地，在一些实施例中，存储节点包括关于第三方数据的一个或多个数据读取策略；

存储节点根据数据读取请求和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息，包括：

- 5 存储节点将数据读取请求与一个或多个数据读取策略中的至少一个数据读取策略的策略项进行匹配；

若存在与数据读取请求相匹配的数据读取策略，则根据与数据读取请求相匹配的数据读取策略中的操作，生成反馈信息并向执行节点发送反馈信息。

本申请实施例中，数据读取策略包括策略项和基于策略项确定的操作。

- 10 待匹配的数据读取策略的数量可以有一个或多个。当待匹配的数据读取策略的数量有多个时，可以依次匹配该多个数据读取策略，具体地，可以根据数据读取策略中的策略项进行匹配。依次匹配该多个数据读取策略，直到存在相匹配的数据读取策略，并根据该数据读取策略中的操作，生成反馈信息，或者，直到确定所有的数据读取策略均不匹配，则可以根据默认操作生成反馈信息，或者可以生成指示拒绝数据读取的反馈信息。

- 15 此外，在一些示例中，为了更好地对执行节点的数据读取操作进行管理，可以使得执行节点本身无法自行获得与存储节点进行交互的权限，而是由数据管理系统管理执行节点与存储节点进行交互的权限，当数据管理系统为执行节点授权，以使得执行节点获得与存储节点进行交互的权限之后，执行节点才能向存储节点发送数据读取请求。

具体地，在一些实施例中，数据管理系统包括管理节点；

- 20 管理节点获取对业务程序的运行指令之后，创建对存储节点的访问凭证；

管理节点指示执行节点运行业务程序，并向执行节点传递访问凭证的信息；

执行节点向存储节点发送数据读取请求，包括：

执行节点基于访问凭证的信息，向存储节点发送数据读取请求。

- 25 管理节点可以认为是数据管理系统的系统级节点，该管理节点可以管理执行节点与存储节点进行交互的权限。数据管理系统通过管理节点，对存储节点的访问凭证进行系统级管理，而不是由执行节点自行申请存储节点的访问凭证，避免数据使用者通过执行节点不正当地访问第三方数据。

可见，本申请实施例中，可以通过管理节点管理执行节点对存储节点的访问凭证，也即是说，该管理节点可以管理执行节点与存储节点进行交互的权限。这样，执行节点无法自行获得对存储节点的访问权限，避免了执行节点在未被允许的情况下对存储节点的第三方数据进行数据读取和使用，从而避免了执行节点对第三方数据的泄露和不合理使用。

- 30 其中，该访问凭证的具体形式可以有多种情况，在此不做限制。

示例性地，该访问凭证可以为存储会话，具体地，该访问凭证可以通过存储会话的标识来唯一地表示。或者该访问凭证也可以为指定密钥或者其他形式的凭证。

如图8所示，为管理节点、执行节点以及存储节点之间进行信息交互的一种示例性流程示意图。

- 35 在图8的示例中，可以包括以下步骤：

1) 数据使用者向管理节点发送启动任务的启动指令，以指示通过业务程序执行指定的业务。

2) 管理节点在接收到启动指令之后，可以向存储节点发送指示存储节点创建存储会话的指示信息。

该指示信息可以包括管理节点所在的隔离运行环境的ID（例如所在的开发空间的ID）、存储会话对应的主体和存储会话对应的用途，以请求该主体可以通过该存储会话，获得从存储节点读取用于该用途的数据的权限。

- 40 3) 存储节点在接收到指示信息之后，可以向管理节点返回存储会话的ID，这样，相应主体类型的主体可以通过该存储会话的ID，获得从存储节点读取用于该用途的数据的权限。

4) 管理节点在接收到存储会话的ID之后，指示执行节点运行业务程序，并向业务程序传递存储会话的ID，例如，可以通过参数、环境变量等方式传递。

- 45 5) 执行节点在运行业务程序时，根据存储会话的ID，向存储节点请求列举存储节点的数据包括的文件。

6) 存储节点在不校验策略的情况下, 返回文件列表。

7) 执行节点在运行业务程序时, 根据存储会话的 ID, 向存储节点发送数据读取请求, 数据读取请求指示运行的业务程序请求从存储节点读取第一数据, 第一数据属于第三方数据。

8) 存储节点根据数据读取请求, 读取第一数据的元数据。

5 9) 存储节点根据第一数据的元数据、存储会话对应的主体、存储会话对应的用途等, 校验策略, 例如, 校验数据权限策略, 再校验数据读取策略。

10) 根据匹配结果, 返回反馈信息。

其中, 匹配结果可能有以下情况:

A、若某一数据读取策略不匹配, 判断下一个数据读取策略;

10 B、若某一数据读取策略匹配, 则:

(a) 若数据读取策略中的操作为允许 (ALLOW), 则确定该数据读取策略为匹配的策略, 并根据该数据读取策略反馈允许读取的文件内容, 并停止执行后续的策略匹配操作。

(b) 若数据读取策略中的操作为拒绝 (DENY), 则拒绝访问第一数据, 停止执行后续的策略匹配操作。

C、若全部不匹配, 则存储节点可以采用默认操作, 该默认操作可以由数据管理系统指定。

15 具体地, 判断某一数据读取策略是否匹配, 可以从数据读取策略的策略项来实现。

例如, 某一数据读取策略包括数据策略项、用途策略项、数量策略项、时间策略项、主体策略项。其中, 每个策略项的示例性匹配方式如下:

A. 数据策略项:

(a) 路径过滤: 数据读取请求所指定的文件路径需要匹配指定条件;

20 (b) 选择列: 不参与匹配, 而是在读取内容时应用;

(c) 行过滤: 不参与匹配, 而是在读取内容时应用。

B. 用途策略项:

数据读取请求所指定的需在用途策略项定义的使用寿命列表中。

C. 数量策略项:

25 数据读取策略的控制对象可以是数据自身 (DATA), 也可以是数据中的文件 (FILE), 通过 PolicyType 定义。第一数据的元数据中的 PolicyType 可以为 DATA 或 FILE。

(a) 元数据中的 PolicyType 为 DATA 时:

使用次数: 按数量策略项中用户指定的计算规则计算第一数据的使用次数, 判断是否超出范围。其中, 一个存储会话内对某一数据读取多次, 可以仅算一次使用。

30 (b) 元数据中的 PolicyType 为 FILE 时:

① 行数: 不参与匹配, 而是在读取内容时应用;

② 存储大小: 不参与匹配, 而是在读取内容时应用。

D. 时间策略项:

当前时间需要在使用时间范围内。

35 E. 主体策略项:

数据读取请求对应的主体能够匹配主体策略项的定义, 如主体的 userId 在主体策略项定义的 userId 中。

若存储节点根据匹配结果确定允许数据读取, 则返回允许读取的数据。

其中, 存储节点返回允许读取的数据时, 需要基于最终匹配的数据读取策略对读取的内容进行控制, 40 主要有以下控制方式中的一种或多种:

选择列: 剔除未选择的列;

行过滤: 对内容进行过滤;

行数: 限制最多可读取的行数;

存储大小: 限制最多可读取的大小。

45 这样, 可以完成存储节点对数据读取请求的反馈。

在一些实施例中, 为了保证访问凭证的有效性, 在执行节点停止运行业务程序之后, 删除访问凭证。

这样，执行节点在每次启动业务程序的运行时，均需要通过管理节点重新获得执行节点与存储节点进行交互的权限。这样，执行节点无法自行获得对存储节点的访问权限，也无法滥用已获得的访问权限，有效保证执行节点对存储节点的每次访问的可控性，避免了执行节点在未被允许的情况下对存储节点的第三方数据进行数据读取和使用，从而避免了执行节点对第三方数据的泄露和不合理使用。

5

此外，在一些实施例中，存储节点可以记录对第三方数据的数据访问行为，以便于数据提供者查看。

具体地，在一些实施例中，在上述步骤 703 之后，该方法还包括：

存储节点记录数据读取请求相关的数据访问行为。

数据访问行为包括成功实现数据读取的访问行为，也可以包括数据读取失败的访问行为。

10 本申请实施例中，数据访问行为可以包括数据读取请求、数据读取请求相关的用户以及数据读取请求对应的响应（例如反馈信息等）中的一种或多种信息。

示例性地，数据访问行为可以包括以下信息中的一种或多种：

访问时间、用户（包括数据使用者和/或数据提供者）、用途、操作（例如读操作或写操作）、数据、文件、行数、存储大小。

15 例如，访问记录可以如表 3 所示。

表 3：访问记录

字段名称	数据类型	备注
开发空间的 ID	String	
存储会话 ID	String	
时间	DateTime	
读写	String	读（R）/写（W）
数据 ID	String	

存储节点存储上述数据访问行为的信息之后，数据提供者可以方便灵活地查看其存储于存储节点的第三方数据的使用情况。

20 具体地，在一些实施例中，该方法还包括：

存储节点获取数据提供者输入的查询请求，查询请求用于请求查询第三方数据相关的访问记录；

存储节点输出访问记录，访问记录包括数据读取请求相关的数据访问行为。

本申请实施例中，数据提供者可以通过客户端等，基于 API 向数据管理系统输入查询请求，以请求查询数据提供者提供的第三方数据相关的访问记录。存储节点在接收到该查询请求之后，可以根据该数

25 据提供者的标识和/或第三方数据的标识，从存储的数据访问行为信息中，查询关于该第三方数据相关的数据访问行为信息，以获得并向数据提供者和/或其他设备输出对第三方数据的访问记录，此时，该访问记录可以包括数据读取请求相关的数据访问行为。

这样，数据提供者可以通过存储节点方便地查询第三方数据被数据使用者访问和使用的情况，使得第三方数据的使用透明化。

30 此外，数据提供者还可以及时发现第三方数据的异常使用，并及时调整数据读取策略。数据提供者还可以通过数据管理系统，向数据使用者发起策略协商流程，以将数据管理系统作为中间平台，与数据使用者协商出新的数据读取策略，以满足数据提供者和数据使用者双方的要求。

四、衍生数据管理。

35 本申请实施例中，执行节点在运行业务程序的过程中，在从存储节点读取到第三方数据之后，在对第三方数据进行应用的过程中可能产生衍生数据。

为了保证数据的安全性，本申请实施例中，可以对该衍生数据进行跟踪和管理，避免衍生数据的泄露。

具体地，如图 7 所示，在一些实施例中，该方法还包括：

40 步骤 704，在运行业务程序的过程中，执行节点向存储节点发送待存储数据。

步骤 705，存储节点在接收到待存储数据之后，根据指定时间段内执行节点从存储节点读取的第三方数据，确定并存储待存储数据的衍生链路。

指定时间段包括执行节点本次运行业务程序的时间段，衍生链路用于指示存储节点中，用于生成待存储数据的数据。

5 本申请实施例中，待存储数据可以有多种情况。

例如，在一种示例中，执行节点在运行业务程序的过程中可以检测到根据从存储节点读取到的第三方数据而生成的衍生数据，并将该衍生数据作为待存储数据。

在另一种示例中，执行节点在执行业务程序的过程中，可能会生成临时数据以及需要持久化存储的持久化数据。其中，临时数据可以在执行节点中进行存储，例如，可以通过执行节点的缓存、内存等进行存储。而持久化数据需要通过存储节点进行存储。因此，执行节点可以将该持久化数据作为待存储数据并向待存储节点发送。在本示例中，并不实际地检测该待存储数据是否真实地根据第三方数据而生成。

10 存储节点在接收到待存储数据之后，可以认为该待存储数据很可能是基于指定时间段内执行节点从存储节点读取的第三方数据而生成的，因此，根据指定时间段内执行节点从存储节点读取的第三方数据，确定并存储待存储数据的衍生链路。

15 其中，该指定时间段可以包括执行节点本次开始运行业务程序的时间点到接收到存储节点接收到该待存储数据的时间点之间的时间段。

例如，如图 9 所示的流程示意图中，管理节点指示执行节点运行业务程序，并向执行节点传递存储会话的 ID 以作为访问凭证之后，可以包括以下步骤：

20 执行节点从存储节点读取数据 A，数据 A 为第三方数据；
存储节点记录数据 A 相关的访问行为，例如，记录存储会话的 ID、数据 A 的信息、时间等；
执行节点从存储节点读取数据 B，数据 B 为第三方数据；
存储节点记录数据 B 相关的访问行为，例如，记录存储会话的 ID、数据 B 的信息、时间等；
执行节点向存储节点写入数据 C，该数据 C 可以认为是待存储数据；
存储节点可以根据记录的访问行为，获取创建本次的存储会话的 ID 之后到本次写入操作之前的指
25 定时间段内，执行节点从存储节点读取的第三方数据；

根据该第三方数据的信息，更新存储节点中的数据 C 的元数据中，“衍生自”字段的信息，以指示该数据 C 衍生至数据 A 和数据 B；

30 存储节点确定并存储数据 C 的衍生链路；
存储节点写入数据 C 的内容，并向执行节点返回信息，以指示已写入数据 C；
执行节点从存储节点读取数据 D，数据 D 为第三方数据，相关读取操作不再赘述；
执行节点向存储节点写入数据 E，相关读取操作不再赘述。

在执行上述步骤之后，存储节点中可以生成并存储如图 10 所示的衍生链路，其中，数据 C 衍生自数据 A 以及数据 B，数据 E 衍生自数据 A、数据 B、数据 C 以及数据 D。

35 示例性地，衍生链路可以通过表 4 所示的字段进行记录。

表 4：衍生链路的字段

字段名称	数据类型	备注
空间 ID	String	
源数据 ID	String	
目标数据 ID	String	

其中，数据 C 的衍生链路中，数据 C 可以认为是目标数据，数据 A 以及数据 B 则为数据 C 的源数据。

此外，一个开发空间内，执行节点中的业务程序多次执行后，相关衍生链路还可以进行自动合并更新，如图 11 所示，可以合并有存储会话 1 相关的衍生链路以及存储会话 2 相关的衍生链路。

40 当数据提供者或者数据使用者查看数据 A 的衍生链路时，根据衍生链路的记录，可以找到数据 A 之后的衍生链路以及相关衍生数据。

此外，数据提供者还可以请求删除数据提供者提供的第三方数据，还可以请求删除该第三方数据的

衍生数据。

具体地，数据提供者可以向存储节点发起删除请求。同时，存储节点可以给予一定时间作为缓冲。例如，数据使用者可以收到删除请求通知，然后可以根据查看待删除的第三方数据对应的所有衍生链路，以及及时调整业务程序。缓冲时间到期后，数据管理系统可以根据上述衍生链路实施第三方数据及相关衍生数据的删除操作，并通知数据使用者。若数据使用者在缓冲时间内无法完成，可以要求延期，数据提供者同意后以新时间为准。

五、停止运行业务程序。

具体地，在一些实施例中，该方法还包括：

在执行节点停止运行业务程序之后，删除执行节点中业务程序相关的运行数据。

在执行节点完成该业务程序的本次执行之后，可以删除执行节点中存储的关于本次执行的相关运行数据，例如临时数据等，但该执行节点可以保留，以执行该数据使用者的其他业务；或者在后续再次执行该业务程序；或者，在执行节点完成该业务程序的本次执行之后，可以销毁该执行节点。

这样，可以有效保证运行业务程序的过程中使用第三方数据的使用痕迹被删除，避免执行节点在不合理地保留运行业务程序的过程中所使用的第三方数据，从而保证该第三方数据不被泄露和滥用。

以上，从多个方面介绍了本申请实施例提供的的数据管理方法，下面结合附图，介绍本申请实施例提供的的数据管理系统。

如图 12 所示，本申请实施例提供一种数据管理系统 120，数据管理系统 120 设置于至少一个数据中心，数据管理系统 120 位于隔离运行环境中，隔离运行环境通过应用程序编程接口 API 分别与数据提供者和数据使用者连接，数据管理系统 120 中包括执行节点 1201 和存储节点 1202，存储节点 1202 存储有数据提供者的第三方数据。

执行节点 1201 用于：

运行业务程序，业务程序用于执行数据使用者的业务；

向存储节点发送数据读取请求，数据读取请求指示运行的业务程序请求从存储节点读取第一数据，第一数据属于第三方数据；

存储节点 1202 用于：根据数据读取请求和关于第三方数据的数据读取策略，生成反馈信息并向执行节点发送反馈信息。

可选地，系统 120 还包括管理节点 1203；

管理节点 1203 用于监控数据使用者通过 API 对数据管理系统的访问行为，并监控数据管理系统通过 API 的数据输出行为。

可选地，数据读取策略由数据提供者确定，数据读取策略包括策略项和基于策略项确定的操作，策略项包括以下项中的一种或多种：

数据策略项、用途策略项、数量策略项、时间策略项、主体策略项；

其中，数据策略项用于描述第三方数据中能够被读取的数据范围；

用途策略项用于描述业务程序中能够使用第三方数据的任务；

数量策略项用于描述第三方数据的使用量阈值；

时间策略项用于描述第三方数据的使用时间范围；

主体策略项用于描述能够应用数据读取策略的主体类型。

可选地，存储节点 1202 包括关于第三方数据的一个或多个数据读取策略；

存储节点 1202 用于：

将数据读取请求与一个或多个数据读取策略中的至少一个数据读取策略的策略项进行匹配；

若存在与数据读取请求相匹配的数据读取策略，则根据与数据读取请求相匹配的数据读取策略中的操作，生成反馈信息并向执行节点发送反馈信息。

可选地，存储节点 1202 还包括数据权限策略，数据权限策略用于根据数据的来源，确定对数据的读写权限，其中，数据权限策略指示数据使用者对来自数据提供者的第三方数据具有读权限；

- 存储节点 1202 还用于：
根据数据权限策略，基于第一数据属于第三方数据，确定业务程序对第一数据具有读权限。
可选地，管理节点 1203 用于：
获取对业务程序的运行指令之后，创建对存储节点的访问凭证；
5 指示执行节点运行业务程序，并向执行节点传递访问凭证的信息；
执行节点 1201 用于：
基于访问凭证的信息，向存储节点发送数据读取请求。
可选地，执行节点 1201 还用于：在执行节点停止运行业务程序之后，删除访问凭证。
可选地，存储节点 1202 还用于：记录数据读取请求相关的数据访问行为。
10 可选地，存储节点 1202 还用于：
获取数据提供者输入的查询请求，查询请求用于请求查询第三方数据相关的访问记录；
输出访问记录，访问记录包括数据读取请求相关的数据访问行为。
可选地，执行节点 1201 还用于：在运行业务程序的过程中，向存储节点发送待存储数据；
存储节点 1202 还用于：在接收到待存储数据之后，根据指定时间段内执行节点从存储节点读取的
15 第三方数据，确定并存储待存储数据的衍生链路，指定时间段包括执行节点本次运行业务程序的时间段，
衍生链路用于指示存储节点中，用于生成待存储数据的数据。
可选地，执行节点 1201 还用于：在执行节点停止运行业务程序之后，删除执行节点中业务程序相关的运行数据。
本申请实施例中，执行节点、存储节点和管理节点均可以通过软件实现，或者可以通过硬件实现。
20 示例性的，接下来以执行节点为例，介绍执行节点的实现方式。类似的，存储节点和管理节点的实现方式可以参考执行节点的实现方式。
模块作为软件功能单元的一种举例，执行节点可以包括运行在计算实例上的代码。其中，计算实例可以包括物理主机（计算设备）、虚拟机、容器中的至少一种。进一步地，上述计算实例可以是一台或者多台。例如，执行节点可以包括运行在多个主机/虚拟机/容器上的代码。需要说明的是，用于运行该
25 代码的多个主机/虚拟机/容器可以分布在相同的区域（region）中，也可以分布在不同的 region 中。
进一步地，用于运行该代码的多个主机/虚拟机/容器可以分布在相同的可用区（availability zone，AZ）中，也可以分布在不同的 AZ 中，每个 AZ 包括一个数据中心或多个地理位置相近的数据中心。其中，通常一个 region 可以包括多个 AZ。
同样，用于运行该代码的多个主机/虚拟机/容器可以分布在同一个虚拟私有云（virtual private
30 cloud，VPC）中，也可以分布在多个 VPC 中。其中，通常一个 VPC 设置在一个 region 内，同一 region 内两个 VPC 之间，以及不同 region 的 VPC 之间跨区通信需在每个 VPC 内设置通信网关，经通信网关实现 VPC 之间的互连。
模块作为硬件功能单元的一种举例，执行节点可以包括至少一个计算设备，如服务器等。或者，执行节点也可以是利用中央处理单元（central processing unit，CPU）实现，也可以利用专用集成电路
35 （application-specific integrated circuit，ASIC）实现、或可编程逻辑器件（programmable logic device，PLD）实现的设备等。其中，上述 PLD 可以是复杂程序逻辑器件（complex programmable logical device，CPLD）、现场可编程门阵列（field-programmable gate array，FPGA）、通用阵列逻辑（generic array logic，GAL）、数据处理单元（data processing unit，DPU）、神经网络处理单元（Neural Network Processing Unit，NPU）、片上系统（system on chip，SoC）、卸载卡、加速卡或其任意组合实现。
40 执行节点包括的多个计算设备可以分布在相同的 region 中，也可以分布在不同的 region 中。执行节点包括的多个计算设备可以分布在相同的 AZ 中，也可以分布在不同的 AZ 中。同样，执行节点包括的多个计算设备可以分布在同一个 VPC 中，也可以分布在多个 VPC 中。其中，多个计算设备可以是服务器、ASIC、PLD、CPLD、FPGA、GAL、DPU、NPU、SoC、卸载卡、加速卡等计算设备的任意组合。
需要说明的是，在其他实施例中，执行节点可以用于执行数据管理方法中的任意步骤，存储节点可以
45 用于执行数据管理方法中的任意步骤，管理节点可以用于执行数据管理方法中的任意步骤，执行节点、存储节点、以及管理节点负责实现的步骤可根据需要指定，通过执行节点、存储节点、以及管理节点

点分别实现数据管理方法中不同的步骤来实现数据管理系统的全部功能。

本申请实施例还提供一种计算设备 130。如图 13 所示，计算设备 130 包括：总线 132、处理器 134、存储器 136 和通信接口 138。处理器 134、存储器 136 和通信接口 138 之间通过总线 132 通信。计算设备 130 可以是服务器或终端设备。应理解，本申请不限定计算设备 130 中的处理器、存储器的个数。

5 总线 132 可以是外设部件互连标准（peripheral component interconnect, PCI）总线或扩展工业标准结构（extended industry standard architecture, EISA）总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图 13 中仅用一条线表示，但并不表示仅有一根总线或一种类型的总线。总线 134 可包括在计算设备 130 各个部件（例如，存储器 136、处理器 134、通信接口 138）之间传送信息的通路。

10 处理器 134 可以包括中央处理器（central processing unit, CPU）、图形处理器（graphics processing unit, GPU）、微处理器（micro processor, MP）或者数字信号处理器（digital signal processor, DSP）等处理器中的任意一种或多种。

存储器 136 可以包括易失性存储器（volatile memory），例如随机存取存储器（random access memory, RAM）。存储器 136 还可以包括非易失性存储器（non-volatile memory），例如只读存储器（read-only memory, ROM），快闪存储器，机械硬盘（hard disk drive, HDD）或固态硬盘（solid state drive, SSD）。

存储器 136 中存储有可执行的程序代码，处理器 134 执行该可执行的程序代码以分别实现前述执行节点和存储节点的功能，从而实现上述实施例中应用于计算设备集群的数据管理方法，也即，存储器 136 上存有用于执行上述实施例中应用于计算设备集群的数据管理方法的指令。

20 通信接口 138 使用例如但不限于网络接口卡、收发器一类的收发模块，来实现计算设备 130 与其他设备或通信网络之间的通信。

本申请实施例还提供了一种计算设备集群。该计算设备集群包括至少一台计算设备。该计算设备可以是服务器，例如中心服务器、边缘服务器，或者是本地数据中心中的本地服务器。在一些实施例中，计算设备也可以是台式机、笔记本电脑或者智能手机等终端设备。

25 如图 14 所示，计算设备集群包括至少一个计算设备 130。计算设备集群中的一个或多个计算设备 130 中的存储器 136 中可以存有相同的用于执行数据管理方法的指令。

在一些可能的实现方式中，该计算设备集群中的一个或多个计算设备 130 的存储器 136 中也可以分别存有助于执行数据管理方法的部分指令。换言之，一个或多个计算设备 130 的组合可以共同执行用于执行数据管理方法的指令。

30 需要说明的是，计算设备集群中的不同的计算设备 130 中的存储器 136 可以存储不同的指令，分别用于执行数据管理方法的部分功能。也即，不同的计算设备 130 中的存储器 136 存储的指令可以实现执行节点以及存储节点中的一个或多个模块的功能。

在一些可能的实现方式中，计算设备集群中的一个或多个计算设备可以通过网络连接。其中，网络可以是广域网或局域网等等。图 15 示出了一种可能的实现方式。如图 15 所示，两个计算设备 130A 和 130B 之间通过网络进行连接。具体地，通过各个计算设备中的通信接口与网络进行连接。在这一类可能的实现方式中，计算设备 130A 中的存储器 136 中可以存有执行执行节点的功能的指令。同时，计算设备 130B 中的存储器 136 中可以存有执行存储节点的功能的指令。或者，计算设备 130A 中的存储器 136 中可以存有执行存储节点的部分功能的指令。同时，计算设备 130B 中的存储器 136 中可以存有执行存储节点的另一部分功能的指令。

40 应理解，图 15 中示出的计算设备 130A 的功能也可以由多个计算设备 130 完成。同样，计算设备 130B 的功能也可以由多个计算设备 130 完成。

本申请实施例还提供了另一种计算设备集群。该计算设备集群中各计算设备之间的连接关系可以类似的参考图 14 和图 15 计算设备集群的连接方式。不同的是，该计算设备集群中的一个或多个计算设备 130 中的存储器 136 中可以存有相同的用于执行数据管理方法的指令。

45 在一些可能的实现方式中，该计算设备集群中的一个或多个计算设备 130 的存储器 136 中也可以分别存有助于执行数据管理方法的部分指令。换言之，一个或多个计算设备 130 的组合可以共同执行用于

执行数据管理方法的指令。

需要说明的是，计算设备集群中的不同的计算设备 130 中的存储器 136 可以存储不同的指令，用于执行数据管理方法的部分功能。也即，不同的计算设备 130 中的存储器 136 存储的指令可以实现执行节点以及存储节点中的一个或多个模块的功能。

5 本申请实施例还提供了一种包含指令的计算机程序产品。计算机程序产品可以是包含指令的，能够运行在计算设备上或被储存在任何可用介质中的软件或程序产品。当计算机程序产品在至少一个计算设备上运行时，使得至少一个计算设备执行数据管理方法。

10 本申请实施例还提供了一种计算机可读存储介质。计算机可读存储介质可以是计算设备能够存储的任何可用介质或者是包含一个或多个可用介质的数据中心等数据存储设备。可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘）等。该计算机可读存储介质包括指令，指令指示计算设备执行数据管理方法。

 本申请实施例还提供了一种芯片系统，该芯片系统包括处理器，该处理器用于实现上述计算设备集群所执行的步骤。在一种可能的设计中，芯片系统还可以包括存储器，存储器，用于保存必要的程序指令和数据。该芯片系统，可以由芯片构成，也可以包含芯片和其他分立器件。

15 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统，装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

 在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

20 作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

25 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

 集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM，read-only memory）、随机存取存储器（RAM，random access memory）、磁碟或者光盘等各种可以存储程序代码的介质。

35

权 利 要 求 书

1. 一种数据管理方法，其特征在于，应用于数据管理系统，所述数据管理系统设置于至少一个数据中心，所述数据管理系统位于隔离运行环境中，所述隔离运行环境通过应用程序编程接口 API 分别与数据提供者和数据使用者连接，所述数据管理系统中包括执行节点和存储节点，所述存储节点存储有所述数据提供者的第三方数据，所述方法包括：

所述执行节点运行业务程序，所述业务程序用于执行所述数据使用者的业务；
所述执行节点向所述存储节点发送数据读取请求，所述数据读取请求指示运行的所述业务程序请求从所述存储节点读取第一数据，所述第一数据属于所述第三方数据；
所述存储节点根据所述数据读取请求和关于所述第三方数据的数据读取策略，生成反馈信息并向所述执行节点发送所述反馈信息。

2. 根据权利要求 1 所述的方法，其特征在于，所述方法还包括：
监控所述数据使用者通过所述 API 对所述数据管理系统的访问行为，并监控所述数据管理系统通过所述 API 的数据输出行为。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，所述数据读取策略由所述数据提供者确定，所述数据读取策略包括策略项和基于所述策略项确定的操作，所述策略项包括以下项中的一种或多种：
数据策略项、用途策略项、数量策略项、时间策略项、主体策略项；
其中，所述数据策略项用于描述所述第三方数据中能够被读取的数据范围；
所述用途策略项用于描述所述业务程序中能够使用所述第三方数据的任务；
所述数量策略项用于描述所述第三方数据的使用量阈值；
所述时间策略项用于描述所述第三方数据的使用时间范围；
所述主体策略项用于描述能够应用所述数据读取策略的主体类型。

4. 根据权利要求 3 所述的方法，其特征在于，所述存储节点包括关于所述第三方数据的一个或多个数据读取策略；
所述存储节点根据所述数据读取请求和关于所述第三方数据的数据读取策略，生成反馈信息并向所述执行节点发送所述反馈信息，包括：
所述存储节点将所述数据读取请求与所述一个或多个数据读取策略中的至少一个数据读取策略的策略项进行匹配；
若存在与所述数据读取请求相匹配的数据读取策略，则所述存储节点根据与所述数据读取请求相匹配的数据读取策略中的操作，生成所述反馈信息并向所述执行节点发送所述反馈信息。

5. 根据权利要求 1-4 任一项所述的方法，其特征在于，所述存储节点还包括数据权限策略，数据权限策略用于根据数据的来源，确定对所述数据的读写权限，其中，所述数据权限策略指示所述数据使用者对来自所述数据提供者的所述第三方数据具有读权限；
在所述存储节点根据所述第一数据和关于所述第三方数据的数据读取策略，生成反馈信息并向所述执行节点发送所述反馈信息之前，还包括：
所述存储节点根据所述数据权限策略，基于所述第一数据属于所述第三方数据，确定所述业务程序对所述第一数据具有读权限。

6. 根据权利要求 1-5 任一项所述的方法，其特征在于，所述数据管理系统包括管理节点，所述方法还包括：
所述管理节点获取对所述业务程序的运行指令之后，创建对所述存储节点的访问凭证；
所述管理节点指示所述执行节点运行所述业务程序，并向所述执行节点传递所述访问凭证的信息；

所述执行节点向所述存储节点发送数据读取请求，包括：
所述执行节点基于所述访问凭证的信息，向所述存储节点发送所述数据读取请求。

7. 根据权利要求 6 所述的方法，其特征在于，所述方法还包括：
5 在所述执行节点停止运行所述业务程序之后，删除所述访问凭证。

8. 根据权利要求 1-7 任一项所述的方法，其特征在于，在所述存储节点根据所述第一数据和关于所述
10 所述第三方数据的数据读取策略，生成反馈信息并向所述执行节点发送所述反馈信息之后，还包括：
所述存储节点记录所述数据读取请求相关的数据访问行为。

9. 根据权利要求 8 所述的方法，其特征在于，所述方法还包括：
所述存储节点获取所述数据提供者输入的查询请求，所述查询请求用于请求查询所述第三方数据相
15 关的访问记录；
所述存储节点输出所述访问记录，所述访问记录包括所述数据读取请求相关的数据访问行为。

10. 根据权利要求 1-9 任一项所述的方法，其特征在于，所述方法还包括：
在运行所述业务程序的过程中，所述执行节点向所述存储节点发送待存储数据；
所述存储节点在接收到所述待存储数据之后，根据指定时间段内所述执行节点从所述存储节点读取
20 的第三方数据，确定并存储所述待存储数据的衍生链路，所述指定时间段包括所述执行节点本次运行所
述业务程序的时间段，所述衍生链路用于指示所述存储节点中，用于生成所述待存储数据的数据。

11. 根据权利要求 1-10 任一项所述的方法，其特征在于，所述方法还包括：
在所述执行节点停止运行所述业务程序之后，删除所述执行节点中所述业务程序相关的运行数据。

12. 一种数据管理系统，其特征在于，所述数据管理系统设置于至少一个数据中心，所述数据管理
25 系统位于隔离运行环境中，所述隔离运行环境通过应用程序编程接口 API 分别与数据提供者和数据使用
者连接，所述数据管理系统中包括执行节点和存储节点，所述存储节点存储有所述数据提供者的第三方
数据；

所述执行节点用于：
30 运行业务程序，所述业务程序用于执行所述数据使用者的业务；
向所述存储节点发送数据读取请求，所述数据读取请求指示运行的所述业务程序请求从所述存储节
点读取第一数据，所述第一数据属于所述第三方数据；
所述存储节点用于：根据所述数据读取请求和关于所述第三方数据的数据读取策略，生成反馈信息
并向所述执行节点发送所述反馈信息。

13. 根据权利要求 12 所述的系统，其特征在于，所述系统还包括管理节点；
所述管理节点用于监控所述数据使用者通过所述 API 对所述数据管理系统的访问行为，并监控所述
35 数据管理系统通过所述 API 的数据输出行为。

14. 根据权利要求 12 或 13 所述的系统，其特征在于，所述数据读取策略由所述数据提供者确定，
40 所述数据读取策略包括策略项和基于所述策略项确定的操作，所述策略项包括以下项中的一种或多种：
数据策略项、用途策略项、数量策略项、时间策略项、主体策略项；
其中，所述数据策略项用于描述所述第三方数据中能够被读取的数据范围；
所述用途策略项用于描述所述业务程序中能够使用所述第三方数据的任务；
45 所述数量策略项用于描述所述第三方数据的使用量阈值；
所述时间策略项用于描述所述第三方数据的使用时间范围；

所述主体策略项用于描述能够应用所述数据读取策略的主体类型。

15. 根据权利要求 14 所述的系统，其特征在于，所述存储节点包括关于所述第三方数据的一个或多个数据读取策略；

5 所述存储节点用于：
 将所述数据读取请求与所述一个或多个数据读取策略中的至少一个数据读取策略的策略项进行匹配；

 若存在与所述数据读取请求相匹配的数据读取策略，则根据与所述数据读取请求相匹配的数据读取策略中的操作，生成所述反馈信息并向所述执行节点发送所述反馈信息。

10 16. 根据权利要求 12-15 任一项所述的系统，其特征在于，所述存储节点还包括数据权限策略，数据权限策略用于根据数据的来源，确定对所述数据的读写权限，其中，所述数据权限策略指示所述数据使用者对来自所述数据提供者的所述第三方数据具有读权限；

 所述存储节点还用于：

15 根据所述数据权限策略，基于所述第一数据属于所述第三方数据，确定所述业务程序对所述第一数据具有读权限。

 17. 根据权利要求 12-16 任一项所述的系统，其特征在于，所述系统还包括管理节点；
 所述管理节点用于：

20 获取对所述业务程序的运行指令之后，创建对所述存储节点的访问凭证；
 指示所述执行节点运行所述业务程序，并向所述执行节点传递所述访问凭证的信息；
 所述执行节点用于：
 基于所述访问凭证的信息，向所述存储节点发送所述数据读取请求。

25 18. 根据权利要求 17 所述的系统，其特征在于，
 所述执行节点还用于：在所述执行节点停止运行所述业务程序之后，删除所述访问凭证。

 19. 根据权利要求 12-18 任一项所述的系统，其特征在于，
 所述存储节点还用于：记录所述数据读取请求相关的数据访问行为。

30 20. 根据权利要求 19 所述的系统，其特征在于，
 所述存储节点还用于：
 获取所述数据提供者输入的查询请求，所述查询请求用于请求查询所述第三方数据相关的访问记录；
 输出所述访问记录，所述访问记录包括所述数据读取请求相关的数据访问行为。

35 21. 根据权利要求 12-20 任一项所述的系统，其特征在于，
 所述执行节点还用于：在运行所述业务程序的过程中，向所述存储节点发送待存储数据；
 所述存储节点还用于：在接收到所述待存储数据之后，根据指定时间段内所述执行节点从所述存储节点读取的第三方数据，确定并存储所述待存储数据的衍生链路，所述指定时间段包括所述执行节点本次运行所述业务程序的时间段，所述衍生链路用于指示所述存储节点中，用于生成所述待存储数据的数据。

40 22. 根据权利要求 12-21 任一项所述的系统，其特征在于，
 所述执行节点还用于：在所述执行节点停止运行所述业务程序之后，删除所述执行节点中所述业务程序相关的运行数据。

45

23. 一种计算设备集群，其特征在于，包括至少一个计算设备，所述至少一个计算设备包括处理器和存储器；

所述处理器用于执行所述存储器中存储的指令，以使得所述计算设备集群执行如权利要求 1-11 任一所述的方法。

5

24. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机程序，当所述计算机程序在处理器上运行时，使得所述处理器执行如权利要求 1-11 任一所述的方法。

25. 一种包含指令的计算机程序产品，其特征在于，当所述指令被处理器执行时实现权利要求 1-11 任一项所述的方法。

10