

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 6 月 6 日 (06.06.2024)



(10) 国际公布号
WO 2024/113705 A1

- (51) 国际专利分类号:
G06F 9/54 (2006.01)
- (21) 国际申请号: PCT/CN2023/095599
- (22) 国际申请日: 2023 年 5 月 22 日 (22.05.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202211518302.9 2022年11月30日 (30.11.2022) CN
- (71) 申请人: 苏州元脑智能科技有限公司 (SUZHOU METABRAIN INTELLIGENT TECHNOLOGY CO., LTD.) [CN/CN]; 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

- (72) 发明人: 郝振 (HAO, Zhen); 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。
- (74) 代理人: 北京康信知识产权代理有限公司 (KANGXIN PARTNERS, P.C.); 中国北京市海淀区知春路甲48号盈都大厦A座16层, Beijing 100098 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) Title: METHOD FOR ISSUING KAFKA MESSAGE DATA, AND RELATED APPARATUS

(54) 发明名称: 一种kafka消息数据的下发方法及相关装置

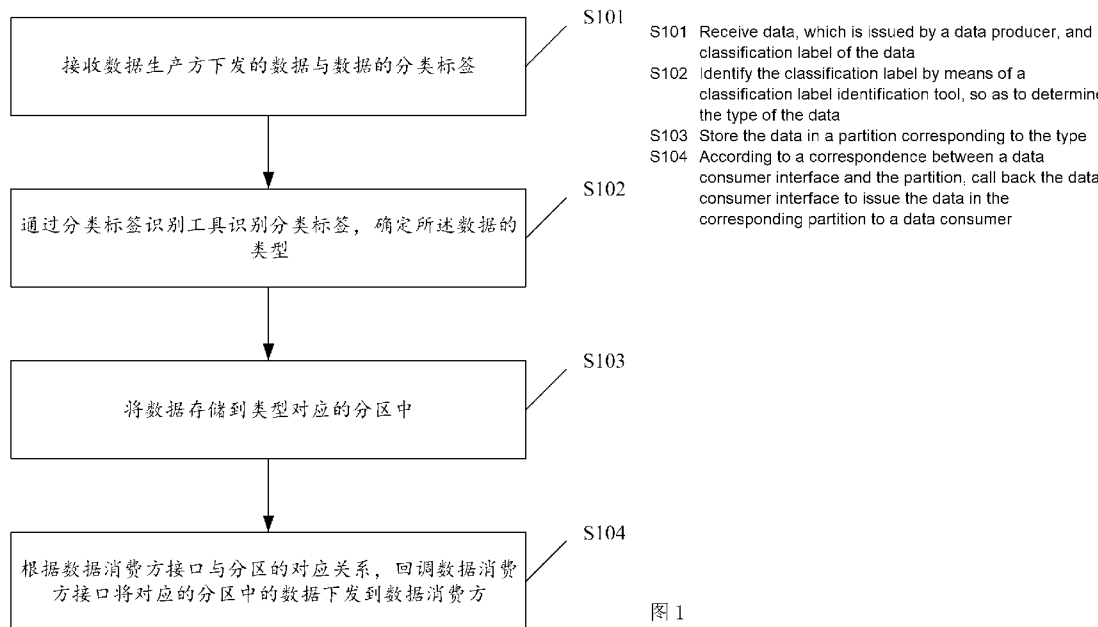


图 1

(57) Abstract: The present application relates to the technical field of computers. Disclosed is a method for issuing kafka message data. The method comprises: receiving data, which is issued by a data producer, and a classification label of the data; identifying the classification label by means of a classification label identification tool, so as to determine the type of the data; storing the data in a partition corresponding to the type; and according to a correspondence between a data consumer interface and the partition, calling back the data consumer interface to issue the data in the corresponding partition to a data consumer. The method can realize more detailed

SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚
(AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE,
BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,
HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO,
PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

本国际公布:
— 包括国际检索报告(条约第21条(3))。

classified storage; and a data type consumed by a data consumer can be actively controlled by a data producer, and when a new data consumer needs to consume a data type specified by the producer, the access of the consumer can be realized simply by configuring a correspondence, thereby making same efficient and convenient. Further disclosed in the present application are an apparatus and device for issuing kafka message data, a non-volatile computer-readable storage medium, and a kafka message integrated service system, which all have the technical effects above.

(57) 摘要: 本申请公开了一种kafka消息数据的下发方法, 涉及计算机技术领域, 包括: 接收数据生产方下发的数据与数据的分类标签; 通过分类标签识别工具识别分类标签, 确定数据的类型; 将数据存储到类型对应的分区中; 根据数据消费方接口与分区的对应关系, 回调数据消费方接口将对应的分区中的数据下发到数据消费方。该方法能够实现更加细致的分类存储, 能够由数据生产方主动控制数据消费方所消费的数据类型, 并且当有新的数据消费方需要消费生产方指定数据类型时, 只需配置对应关系即可实现消费方的接入, 高效便捷。本申请还公开了一种kafka消息数据的下发装置、设备、计算机非易失性可读存储介质以及kafka消息集成服务系统, 均具有上述技术效果。

一种 kafka 消息数据的下发方法及相关装置

相关申请的交叉引用

本申请要求于 2022 年 11 月 30 日提交中国专利局，申请号为 202211518302.9，申请名称为“一种 kafka 消息数据的下发方法及相关装置”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

本申请涉及计算机技术领域，特别涉及一种 kafka 消息数据的下发方法；还涉及一种 kafka 消息数据的下发装置、设备、计算机非易失性可读存储介质以及 kafka 消息集成服务系统。

背景技术

kafka 是一个分布式流处理平台，其主要有两大类别的应用：1、构造实时流数据管道，它可以在系统或应用之间可靠地获取数据（相当于消息队列）；2、构建实时流式应用程序，对流数据进行转换或者影响。针对第一种应用，在业务系统将 kafka 作为消息系统中间件进行使用时的场景下，传统技术方案通常根据 topic（主题）进行分类、存储和消费，完成数据下发，存在以下缺陷：

1、不对 topic 下的数据进行再次分类，对于只想要某个 topic 下的部分类型数据的消费方而言不友好。

2、很多 kafka 消息系统往往对 topic 有数量上的限制，仅用 topic 将数据进行分类，无疑是对 topic 资源的一种浪费。

3、数据生产方在不接入生产者客户端的情况下，无法主动更加细致地对数据进行分类存储和控制消费方所消费的数据类型。

因此，如何解决上述技术缺陷已成为本领域技术人员亟待解决的技术问题。

发明内容

本申请的目的是提供一种 kafka 消息数据的下发方法，能够实现更加细致的分类存储，并且能够主动控制数据消费方所消费的数据类型。本申请的另一个目的是提供一种 kafka 消息数据的下发装置、设备、计算机非易失性可读存储介质以及 kafka 消息集成服务系统，均具有上述技术效果。

为解决上述技术问题，本申请提供了一种 kafka 消息数据的下发方法，包括：

接收数据生产方下发的数据与数据的分类标签；

通过分类标签识别工具识别分类标签，确定数据的类型；

将数据存储到类型对应的分区中；

根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

可选的，通过分类标签识别工具识别分类标签包括：

通过正则表达式识别分类标签。

可选的，一种分类标签识别工具被设置为识别一类分类标签。

可选的，根据数据消费方接口与分区的对应关系，通过数据消费方接口将对应的分区中的数据下发到数据消费方前还包括：

确定数据消费方接口与分区的对应关系。

可选的，确定数据消费方接口与分区的对应关系包括：

根据分类标签识别工具与数据消费方接口的第一对应关系以及分类标签识别工具与分区的第二对应关系，确定数据消费方接口与分区的对应关系。

可选的，还包括：

绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

可选的，绑定分类标签识别工具与数据消费方接口，得到第一对应关系包括：

在数据库或配置文件中绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

可选的，还包括：

绑定分类标签识别工具与分区，得到第二对应关系。

可选的，绑定分类标签识别工具与分区，得到第二对应关系包括：

在数据库或配置文件中绑定分类标签识别工具与分区，得到第二对应关系。

可选的，绑定分类标签识别工具与分区，得到第二对应关系包括：

将数据存储到类型对应的分区后，绑定分类标签识别工具与分区，得到第二对应关系。

可选的，将数据存储到分区中包括：

将数据存储到类型对应的且为数据所属的 topic 对应的分区中。

可选的，还包括：

更新分类标签识别工具。

可选的，还包括：

记录分类识别工具识别分类标签成功与失败的次数。

可选的，还包括：

根据记录分类识别工具识别分类标签成功与失败的次数，计算分类识别工具的识别失败率。

可选的，还包括：

判断分类识别工具的识别失败率是否达到预设阈值；

若分类识别工具的识别失败率达到预设阈值，则提示修正分类识别工具。

可选的，还包括：

若分类识别工具的识别失败率达到预设阈值，则向数据生产方发送分类标签修正提示。

为解决上述技术问题，本申请还提供了一种 kafka 消息数据的下发装置，包括：

接收模块，被设置为接收数据生产方下发的数据与数据的分类标签；

识别模块，被设置为通过分类标签识别工具识别分类标签，确定数据的类型；

存储模块，被设置为将数据存储到类型对应的分区中；

下发模块，被设置为根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

为解决上述技术问题，本申请还提供了一种 kafka 消息数据的下发设备，包括：

存储器，被设置为存储计算机程序；

处理器，被设置为执行计算机程序时实现如上任一项的 kafka 消息数据的下发方法的步骤。

为解决上述技术问题，本申请还提供了一种计算机非易失性可读存储介质，计算机非易失性可读存储介质上存储有计算机程序，计算机程序被处理器执行时实现如上任一项的 kafka 消息数据的下发方法的步骤。

为解决上述技术问题，本申请还提供了一种 kafka 消息集成服务系统，包括：

数据生产方、服务方以及数据消费方；

服务方，被设置为接收数据生产方下发的数据与数据的分类标签；通过分类标签识别工具识别分类标签，确定数据的类型；将数据存储到类型对应的分区中；根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

本申请所提供的 kafka 消息数据的下发方法，包括：接收数据生产方下发的数据与数据的分类标签；通过分类标签识别工具识别分类标签，确定数据的类型；将数据存储到类型对应的分区中；根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

可见，本申请所提供的 kafka 消息数据的下发方法，数据生产方对 topic 下的数据进行

分类并为其设置分类标签，然后将数据以及对应的分类标签下发给服务方，服务方使用分类标签识别工具识别数据的分类标签，并将数据存储到相应的分区，最后服务方依据数据消费方接口与分区的对应关系进行数据下发，这样可以在通过 topic 分类的基础之上，实现更加细致分类存储，可以避免 kafka 消息集成服务系统仅通过 topic 对数据进行分类时对 topic 资源的浪费，数据生产方新增数据类型不必再次申请 topic，并且数据生产方能够主动控制数据消费方所消费的数据类型，另外当有新的数据消费方需要消费生产方指定数据类型时，只需配置对应关系即可实现消费方的接入，高效便捷。

本申请所提供的 kafka 消息数据的下发装置、设备、计算机非易失性可读存储介质以及 kafka 消息集成服务系统均具有上述技术效果。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对现有技术和实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请实施例所提供的一种 kafka 消息数据的下发方法的流程示意图；

图 2 为本申请实施例所提供的一种 kafka 消息集成服务系统的示意图；

图 3 为本申请实施例所提供的一种数据分类识别存储下发示意图；

图 4 为本申请实施例所提供的一种 kafka 消息数据的下发装置的示意图；

图 5 为本申请实施例所提供的一种 kafka 消息数据的下发设备的示意图。

具体实施方式

本申请的核心是提供一种 kafka 消息数据的下发方法，能够在通过 topic 分类的基础之上，实现更加细致的分类存储，并且能够主动控制数据消费方所消费的数据类型。本申请的另一个核心是提供一种 kafka 消息数据的下发装置、设备以及计算机非易失性可读存储介质，均具有上述技术效果。

为使本申请实施例的目的、技术方案和优点更加清楚，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

请参考图 1，图 1 为本申请实施例所提供的一种 kafka 消息数据的下发方法的流程示意

图，参考图1所示，该方法包括：

S101：接收数据生产方下发的数据与数据的分类标签；

本实施例所提供的kafka消息数据的下发方法的执行主体为kafka消息集成服务系统中的服务方。kafka消息集成服务系统可参考图2所示，包括数据生产方、服务方以及数据消费方。数据生产方对topic下的数据进行分类，生成分类标签。需要说明的是，区别于根据topic进行数据分类，本实施例中数据生产方进行数据分类不是根据topic进行数据分类，而是对topic下的数据根据业务类型、数据消费方类型等所进行的分类。

例如，某个topic下的某类数据的分类标签为：“#Gender#Age#school#”。

数据生成方调用服务方接口，将数据与分类标签下发到服务方。

例如，将“#male#18#CentralSouthUniversity#”下发到服务方。

对于topic下的数据的分类方式可以进行差异性设置。例如，可以根据业务类型对topic下的数据进行分类，例如，根据性别、年龄、体重等进行分类。可以根据数据消费方类型对topic下的数据进行分类。

S102：通过分类标签识别工具识别分类标签，确定数据的类型；

分类标签识别工具与数据类型可一一对应，一个分类标签识别工具可用于识别一类分类标签。服务方设置多个分类标签识别工具，能够识别数据生产方下发的任意一种类型的分类标签。数据生产方每新增一种分类标签，服务方对应的增加能够识别此新增的分类标签的分类标签识别工具。

在一些实施例中，通过分类标签识别工具识别分类标签包括：

通过正则表达式识别分类标签。

本实施例中分类标签识别工具为正则表达式。

例如，正则表达式为“^#male#.*”，能够识别所有性别为“male”类型的数据。

正则表达式与数据类型一一对应，一个正则表达式对应一类数据，一个正则表达式用于识别一类分类标签。服务方设置多个正则表达式，能够识别数据生产方下发的任意一种类型的分类标签。

除了正则表达式外，分类标签识别工具还可以为其他形式的识别工具，能够确保准确识别出各种分类标签即可。

S103：将数据存储到类型对应的分区中；

本实施例中分区用于存储特定的某种类型的数据。服务方通过分类标签识别工具识别分类标签，确定数据的类型后，将数据存储到用于存储该种类型的数据的分区中。

例如，服务方通过正则表达式 “^#male#.*” 识别出性别为 “male” 类型的数据。唤起生产者客户端后，将性别为 “male” 类型的数据存储到对应的分区中。

其中，在一些实施例中，将数据存储到分区中包括：

将数据存储到类型对应的且为数据所属的topic对应的分区中。

本实施例中，分区用于存储特定的topic下的特定的某种类型的数据。例如，分区1与分区2均用于存储topic1下的数据，且分区1用于存储topic1下的业务消费方A的数据，分区2用于存储业务消费方B的数据。

S104：根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

每个数据消费方有自身的数据消费方接口。同一个数据消费方可以对应一个或多个数据消费方接口。同一个数据消费方接口可以对应一个或多个数据消费方。服务方向数据消费方下发数据时，唤起消费客户端，根据数据消费方接口，以及数据消费方接口与分区的对应关系，确定出要下发的数据所在的分区，然后从该分区中拉取数据，并回调数据消费方接口，将数据下发到数据消费方。

根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方的前提是存在数据消费方接口与分区的对应关系，因此，在一些实施例中，根据数据消费方接口与分区的对应关系，通过数据消费方接口将对应的分区中的数据下发到数据消费方前还包括：

确定数据消费方接口与分区的对应关系。

其中，确定数据消费方接口与分区的对应关系可以包括：

根据分类标签识别工具与数据消费方接口的第一对应关系以及分类标签识别工具与分区的第二对应关系，确定数据消费方接口与分区的对应关系。

本实施例根据分类标签识别工具与数据消费方接口的对应关系，分类标签识别工具与分区的对应关系，间接得到数据消费方接口与分区的对应关系。

在上述实施例的基础上，在一些实施例中，还包括：

绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

其中，绑定分类标签识别工具与数据消费方接口，得到第一对应关系包括：

在数据库或配置文件中绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

服务方将分类标签识别工具与数据消费方接口配置到服务方的数据库或配置文件中绑定，得到二者之间的第一对应关系。

同一个分类标签识别工具可以具有多个第一对应关系，即同一个分类标签识别工具可以分别与多个数据消费方接口绑定。每一个分类标签识别工具代表一类数据，每一个第一对应关系代表一种业务消费类型。

在上述实施例的基础上，在一些实施例中，还包括：

绑定分类标签识别工具与分区，得到第二对应关系。

其中，绑定分类标签识别工具与分区，得到第二对应关系包括：

在数据库或配置文件中绑定分类标签识别工具与分区，得到第二对应关系。

每个分类标签识别工具可以具有唯一的第二对应关系，即一个分类标签识别工具与一个分区绑定。每一个分类标签识别工具代表一类数据，每一个第二对应关系代表一种类型的数据存储在一个特定的分区中。

每一个分区中存储的数据的类型可以在数据存储到分区之前指定，也可以在某种类型的数据首次存储到分区之后指定。也就是说，可以在数据存储到分区之前，将用于识别某类数据的分类标签识别工具与用于存储该类数据的分区进行绑定。也可以在数据存储到分区之后，将用于识别此类数据的分类标签识别工具与存储了此类数据的分区进行绑定。

在一种可选的实施方式中，绑定分类标签识别工具与分区，得到第二对应关系包括：

将数据存储到类型对应的分区后，绑定分类标签识别工具与分区，得到第二对应关系。

在上述实施例的基础上，在一些实施例中，还包括：

更新分类标签识别工具。

当topic下的数据的分类方式改变，所生成的分类标签发生变化时，相应的进行分类标签识别工具的更新，确保服务方能够识别出数据生产方下发的所有的分类标签。数据生产方每新增一种分类标签，服务方对应的增加能够识别此新增的分类标签的正则表达式。

在上述实施例的基础上，在一些实施例中，还包括：

记录分类识别工具识别分类标签成功与失败的次数。

根据记录分类识别工具识别分类标签成功与失败的次数，计算分类识别工具的识别失败率。

判断分类识别工具的识别失败率是否达到预设阈值；

若分类识别工具的识别失败率达到预设阈值，则提示修正分类识别工具。

若分类识别工具的识别失败率达到预设阈值，则向数据生产方发送分类标签修正提示。

分类识别工具的识别失败率达到预设阈值的原因可能是分类识别工具自身，因此，当分类识别工具的识别失败率达到预设阈值时，可以提示修正分类识别工具，以得到高识别成功

率的分类识别工具，确保能够准确识别分类标签。另外，分类识别工具的识别失败率达到预设阈值的原因可能是分类标签存在错误，因此，当分类识别工具的识别失败率达到预设阈值时，可以向数据生产方发送分类标签修正提示，以使数据生产方进行分类标签的修正，得到准确的可以被分类识别工具识别出的分类标签。

参考图3所示，以下通过一个示例阐述一种可选的分类、识别、存储以及下发流程：

数据生产方对topic下的数据进行分类，得到A、B、C、D四类，数据a为A类数据，为数据a添加上A类标签，数据b为B类数据，为数据c添加上C类标签，数据d为D类数据，为数据d添加上D类标签。

服务方接收到数据生产方下发的数据d与D类标签后，通过正则表达式regular4进行识别，得出数据d为D类数据，将数据d存储到分区partition4。

数据消费方接口URL（Uniform Resource Locator，统一资源定位器）4与分区partition4相对应，从分区partition4拉取数据，回调数据消费方接口URL4，将所拉取的数据下发给对应的数据消费方。

综上所述，本申请所提供的kafka消息数据的下发方法，数据生产方对topic下的数据进行分类并为其设置分类标签，然后将数据以及对应的分类标签下发给服务方，服务方使用分类标签识别工具识别数据的分类标签，并将数据存储到相应的分区，最后服务方依据数据消费方接口与分区的对应关系进行数据下发，这样可以在通过topic分类的基础之上，实现更加细致分类存储，可以避免kafka消息集成服务系统仅通过topic对数据进行分类时对topic资源的浪费，数据生产方新增数据类型不必再次申请topic，并且数据生产方能够主动控制数据消费方所消费的数据类型，另外当有新的数据消费方需要消费生产方指定数据类型时，只需配置对应关系即可实现消费方的接入，高效便捷。

本申请还提供了一种kafka消息数据的下发装置，下文描述的该装置可以与上文描述的方法相互对应参照。请参考图4，图4为本申请实施例所提供的一种kafka消息数据的下发装置的示意图，结合图4所示，该装置包括：

接收模块10，被设置为接收数据生产方下发的数据与数据的分类标签；

识别模块20，被设置为通过分类标签识别工具识别分类标签，确定数据的类型；

存储模块30，被设置为将数据存储到类型对应的分区中；

下发模块40，被设置为根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

在上述实施例的基础上，作为一种可选的实施方式，识别模块20被设置为：

通过正则表达式识别分类标签。

在上述实施例的基础上，作为一种可选的实施方式，一种分类标签识别工具用于识别一类分类标签。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

对应关系确定模块，被设置为确定数据消费方接口与分区的对应关系。

在上述实施例的基础上，作为一种可选的实施方式，对应关系确定模块被设置为：

根据分类标签识别工具与数据消费方接口的第一对应关系以及分类标签识别工具与分区的第二对应关系，确定数据消费方接口与分区的对应关系。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

第一绑定模块，被设置为绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

在上述实施例的基础上，作为一种可选的实施方式，第一绑定模块被设置为：

在数据库或配置文件中绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

第二绑定模块，被设置为绑定分类标签识别工具与分区，得到第二对应关系。

在上述实施例的基础上，作为一种可选的实施方式，第二绑定模块被设置为：

在数据库或配置文件中绑定分类标签识别工具与分区，得到第二对应关系。

在上述实施例的基础上，作为一种可选的实施方式，第二绑定模块被设置为：

将数据存储到类型对应的分区后，绑定分类标签识别工具与分区，得到第二对应关系。

在上述实施例的基础上，作为一种可选的实施方式，存储模块3被设置为：

将数据存储到类型对应的且为数据所属的topic对应的分区中。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

更新模块，被设置为更新分类标签识别工具。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

记录模块，被设置为记录分类识别工具识别分类标签成功与失败的次数。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

计算模块，被设置为根据记录分类识别工具识别分类标签成功与失败的次数，计算分类识别工具的识别失败率。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

判断模块，被设置为判断分类识别工具的识别失败率是否达到预设阈值；

提示模块，被设置为若分类识别工具的识别失败率达到预设阈值，则提示修正分类识别工具。

在上述实施例的基础上，作为一种可选的实施方式，还包括：

发送模块，被设置为若分类识别工具的识别失败率达到预设阈值，则向数据生产方发送分类标签修正提示。

本申请所提供的kafka消息数据的下发装置，数据生产方对topic下的数据进行分类并为其设置分类标签，然后将数据以及对应的分类标签下发给服务方，服务方使用分类标签识别工具识别数据的分类标签，并将数据存储到相应的分区，最后服务方依据数据消费方接口与分区的对应关系进行数据下发，这样可以在通过topic分类的基础之上，实现更加细致分类存储，可以避免kafka消息集成服务系统仅通过topic对数据进行分类时对topic资源的浪费，数据生产方新增数据类型不必再次申请topic，并且数据生产方能够主动控制数据消费方所消费的数据类型，另外当有新的数据消费方需要消费生产方指定数据类型时，只需配置对应关系即可实现消费方的接入，高效便捷。

本申请还提供了一种kafka消息数据的下发设备，参考图5所示，该设备包括存储器1和处理器2。

存储器1，被设置为存储计算机程序；

处理器2，被设置为执行计算机程序实现如下的步骤：

接收数据生产方下发的数据与数据的分类标签；通过分类标签识别工具识别分类标签，确定数据的类型；将数据存储到类型对应的分区中；根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

对于本申请所提供的设备的介绍请参照上述方法实施例，本申请在此不做赘述。

本申请所提供的kafka消息数据的下发设备，数据生产方对topic下的数据进行分类并为其设置分类标签，然后将数据以及对应的分类标签下发给服务方，服务方使用分类标签识别工具识别数据的分类标签，并将数据存储到相应的分区，最后服务方依据数据消费方接口与分区的对应关系进行数据下发，这样可以在通过topic分类的基础之上，实现更加细致分类存储，可以避免kafka消息集成服务系统仅通过topic对数据进行分类时对topic资源的浪费，数据生产方新增数据类型不必再次申请topic，并且数据生产方能够主动控制数据消费方所消费的数据类型，另外当有新的数据消费方需要消费生产方指定数据类型时，只需配置对应关系即可实现消费方的接入，高效便捷。

本申请还提供了一种计算机非易失性可读存储介质，该计算机非易失性可读存储介质上存储有计算机程序，计算机程序被处理器执行时可实现如下的步骤：

接收数据生产方下发的数据与数据的分类标签；通过分类标签识别工具识别分类标签，确定数据的类型；将数据存储到类型对应的分区中；根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

该计算机非易失性可读存储介质可以包括：U盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

对于本申请所提供的计算机非易失性可读存储介质的介绍请参照上述方法实施例，本申请在此不做赘述。

本申请所提供的计算机非易失性可读存储介质，数据生产方对topic下的数据进行分类并为其设置分类标签，然后将数据以及对应的分类标签下发给服务方，服务方使用分类标签识别工具识别数据的分类标签，并将数据存储到相应的分区，最后服务方依据数据消费方接口与分区的对应关系进行数据下发，这样可以在通过topic分类的基础之上，实现更加细致分类存储，可以避免kafka消息集成服务系统仅通过topic对数据进行分类时对topic资源的浪费，数据生产方新增数据类型不必再次申请topic，并且数据生产方能够主动控制数据消费方所消费的数据类型，另外当有新的数据消费方需要消费生产方指定数据类型时，只需配置对应关系即可实现消费方的接入，高效便捷。

本申请还提供了一种kafka消息集成服务系统，包括：

数据生产方、服务方以及数据消费方；

服务方，被设置为接收数据生产方下发的数据与数据的分类标签；通过分类标签识别工具识别分类标签，确定数据的类型；将数据存储到类型对应的分区中；根据数据消费方接口与分区的对应关系，回调数据消费方接口将对应的分区中的数据下发到数据消费方。

在上述实施例的基础上，作为一种可选的实施方式，服务方被设置为：

通过正则表达式识别分类标签。

在上述实施例的基础上，作为一种可选的实施方式，一种分类标签识别工具被设置为识别一类分类标签。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

确定数据消费方接口与分区的对应关系。

在上述实施例的基础上，作为一种可选的实施方式，服务方被设置为：

根据分类标签识别工具与数据消费方接口的第一对应关系以及分类标签识别工具与分区的第二对应关系，确定数据消费方接口与分区的对应关系。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

在上述实施例的基础上，作为一种可选的实施方式，服务方被设置为：

在数据库或配置文件中绑定分类标签识别工具与数据消费方接口，得到第一对应关系。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

绑定分类标签识别工具与分区，得到第二对应关系。

在上述实施例的基础上，作为一种可选的实施方式，服务方被设置为：

在数据库或配置文件中绑定分类标签识别工具与分区，得到第二对应关系。

在上述实施例的基础上，作为一种可选的实施方式，服务方被设置为：

将数据存储到类型对应的分区后，绑定分类标签识别工具与分区，得到第二对应关系。

在上述实施例的基础上，作为一种可选的实施方式，服务方被设置为：

将数据存储到类型对应的且为数据所属的topic对应的分区中。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

更新分类标签识别工具。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

记录分类识别工具识别分类标签成功与失败的次数。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

根据记录分类识别工具识别分类标签成功与失败的次数，计算分类识别工具的识别失败率。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

判断分类识别工具的识别失败率是否达到预设阈值；

若分类识别工具的识别失败率达到预设阈值，则提示修正分类识别工具。

在上述实施例的基础上，作为一种可选的实施方式，服务方还被设置为：

若分类识别工具的识别失败率达到预设阈值，则向数据生产方发送分类标签修正提示。

对于本申请所提供的kafka消息集成服务系统的介绍请参照上述方法实施例，本申请在此不做赘述。

本申请所提供的kafka消息集成服务系统，数据生产方对topic下的数据进行分类并为其

设置分类标签，然后将数据以及对应的分类标签下发给服务方，服务方使用分类标签识别工具识别数据的分类标签，并将数据存储到相应的分区，最后服务方依据数据消费方接口与分区的对应关系进行数据下发，这样可以在通过topic分类的基础之上，实现更加细致分类存储，可以避免kafka消息集成服务系统仅通过topic对数据进行分类时对topic资源的浪费，数据生产方新增数据类型不必再次申请topic，并且数据生产方能够主动控制数据消费方所消费的数据类型，另外当有新的数据消费方需要消费生产方指定数据类型时，只需配置对应关系即可实现消费方的接入，高效便捷。

说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置、设备以及计算机非易失性可读存储介质而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

专业人员还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、只读存储器（ROM）、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

以上对本申请所提供的kafka消息数据的下发方法、装置、设备、计算机非易失性可读存储介质以及kafka消息集成服务系统进行了详细介绍。本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想。应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请原理的前提下，还可以对本申请进行若干改进和修饰，这些改进和修饰也落入本申请权利要求的保护范围。

权 利 要 求 书

1、一种kafka消息数据的下发方法，其特征在于，包括：

接收数据生产方下发的数据与所述数据的分类标签；

通过分类标签识别工具识别所述分类标签，确定所述数据的类型；

将所述数据存储到所述类型对应的分区中；

根据数据消费方接口与分区的对应关系，回调所述数据消费方接口将对应的所述分区中的数据下发到数据消费方。

2、根据权利要求1所述的kafka消息数据的下发方法，其特征在于，所述通过分类标签识别工具识别所述分类标签包括：

通过正则表达式识别所述分类标签。

3、根据权利要求1所述的kafka消息数据的下发方法，其特征在于，一种所述分类标签识别工具用于识别一类所述分类标签。

4、根据权利要求1所述的kafka消息数据的下发方法，其特征在于，所述根据数据消费方接口与分区的对应关系，通过所述数据消费方接口将对应的所述分区中的数据下发到数据消费方前还包括：

确定数据消费方接口与分区的对应关系。

5、根据权利要求4所述的kafka消息数据的下发方法，其特征在于，所述确定数据消费方接口与分区的对应关系包括：

根据所述分类标签识别工具与所述数据消费方接口的第一对应关系以及所述分类标签识别工具与所述分区的第二对应关系，确定所述数据消费方接口与分区的对应关系。

6、根据权利要求5所述的kafka消息数据的下发方法，其特征在于，还包括：

绑定所述分类标签识别工具与所述数据消费方接口，得到所述第一对应关系。

7、根据权利要求6所述的kafka消息数据的下发方法，其特征在于，所述绑定所述分类标签识别工具与所述数据消费方接口，得到所述第一对应关系包括：

在数据库或配置文件中绑定所述分类标签识别工具与所述数据消费方接口，得到所述第一对应关系。

8、根据权利要求5所述的kafka消息数据的下发方法，其特征在于，还包括：

绑定所述分类标签识别工具与所述分区，得到所述第二对应关系。

9、根据权利要求8所述的kafka消息数据的下发方法，其特征在于，所述绑定所述分类标签识别工具与所述分区，得到所述第二对应关系包括：

在数据库或配置文件中绑定所述分类标签识别工具与所述分区，得到所述第二对应

关系。

10、根据权利要求8所述的kafka消息数据的下发方法，其特征在于，所述绑定所述分类标签识别工具与所述分区，得到所述第二对应关系包括：

将所述数据存储到所述类型对应的分区后，绑定所述分类标签识别工具与所述分区，得到所述第二对应关系。

11、根据权利要求1所述的kafka消息数据的下发方法，其特征在于，所述将所述数据存储到分区中包括：

将所述数据存储到所述类型对应的且为所述数据所属的topic对应的分区中。

12、根据权利要求1所述的kafka消息数据的下发方法，其特征在于，还包括：更新所述分类标签识别工具。

13、根据权利要求1所述的kafka消息数据的下发方法，其特征在于，还包括：记录所述分类识别工具识别所述分类标签成功与失败的次数。

14、根据权利要求13所述的kafka消息数据的下发方法，其特征在于，还包括：根据记录所述分类识别工具识别所述分类标签成功与失败的次数，计算所述分类识别工具的识别失败率。

15、根据权利要求14所述的kafka消息数据的下发方法，其特征在于，还包括：判断所述分类识别工具的识别失败率是否达到预设阈值；若所述分类识别工具的识别失败率达到所述预设阈值，则提示修正所述分类识别工具。

16、根据权利要求15所述的kafka消息数据的下发方法，其特征在于，还包括：若所述分类识别工具的识别失败率达到所述预设阈值，则向所述数据生产方发送分类标签修正提示。

17、一种kafka消息数据的下发装置，其特征在于，包括：接收模块，被设置为接收数据生产方下发的数据与所述数据的分类标签；识别模块，被设置为通过分类标签识别工具识别所述分类标签，确定所述数据的类型；

存储模块，被设置为将所述数据存储到所述类型对应的分区中；

下发模块，被设置为根据数据消费方接口与分区的对应关系，回调所述数据消费方接口将对应的所述分区中的数据下发到数据消费方。

18、一种kafka消息数据的下发设备，其特征在于，包括：存储器，被设置为存储计算机程序；

处理器，被设置为执行所述计算机程序时实现如权利要求1至16任一项所述的kafka消息数据的下发方法的步骤。

19、一种计算机非易失性可读存储介质，其特征在于，所述计算机非易失性可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求1至16任一项所述的kafka消息数据的下发方法的步骤。

20、一种kafka消息集成服务系统，其特征在于，包括：

数据生产方、服务方以及数据消费方；

所述服务方，被设置为接收所述数据生产方下发的数据与所述数据的分类标签；通过分类标签识别工具识别所述分类标签，确定所述数据的类型；将所述数据存储到所述类型对应的分区中；根据数据消费方接口与分区的对应关系，回调所述数据消费方接口将对应的所述分区中的数据下发到所述数据消费方。

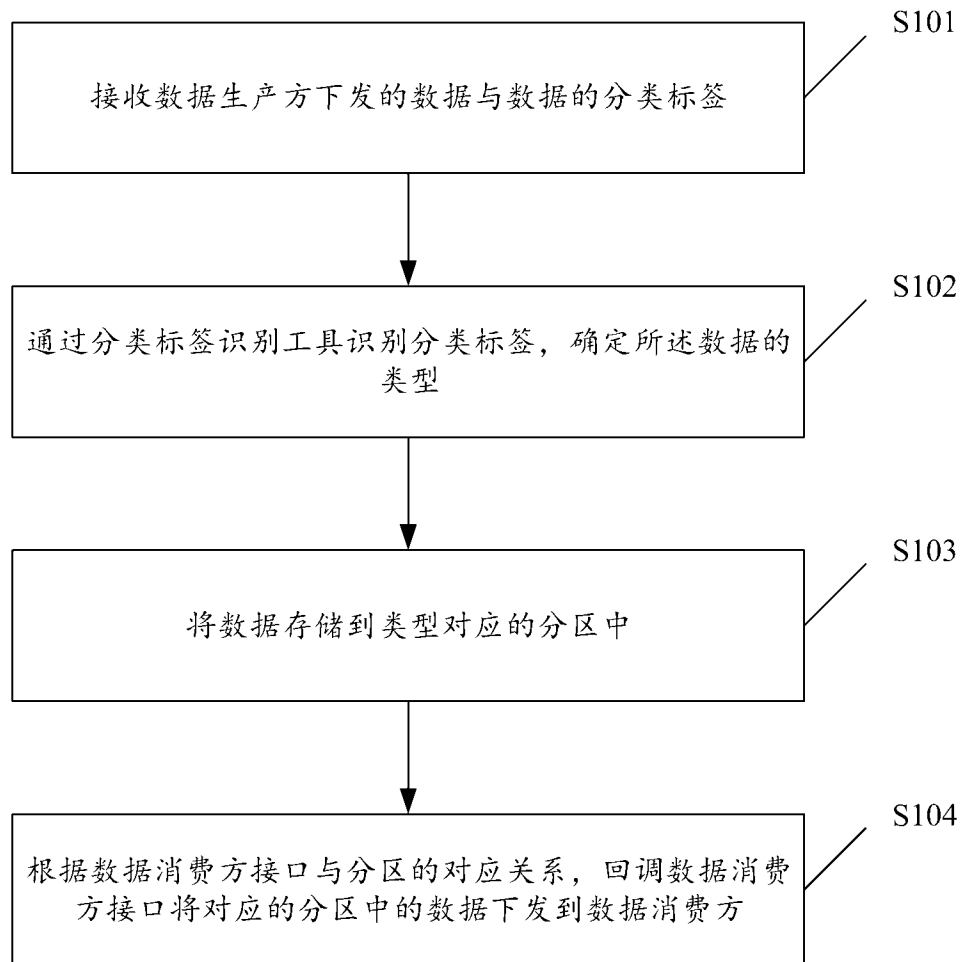


图 1

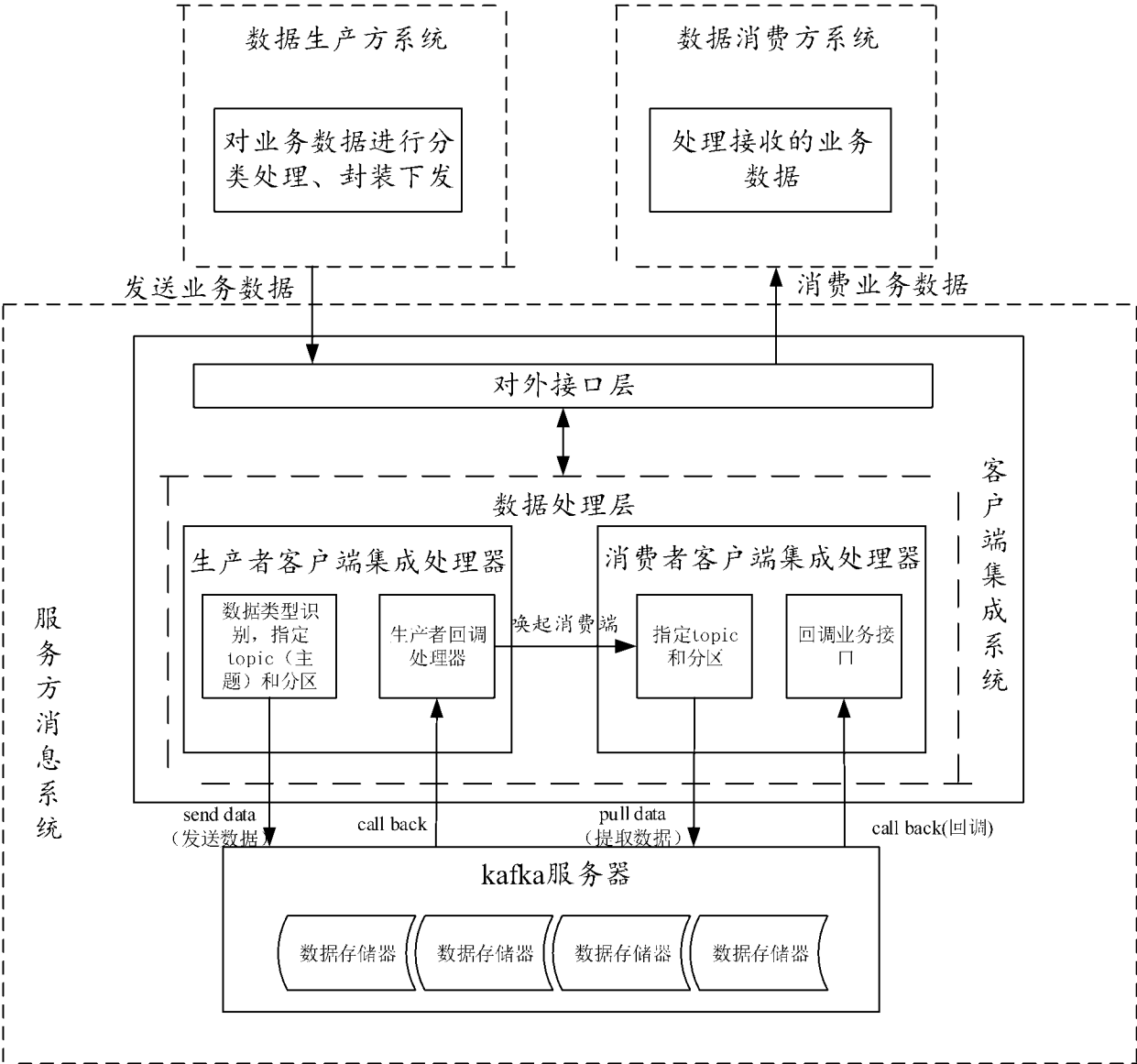


图 2

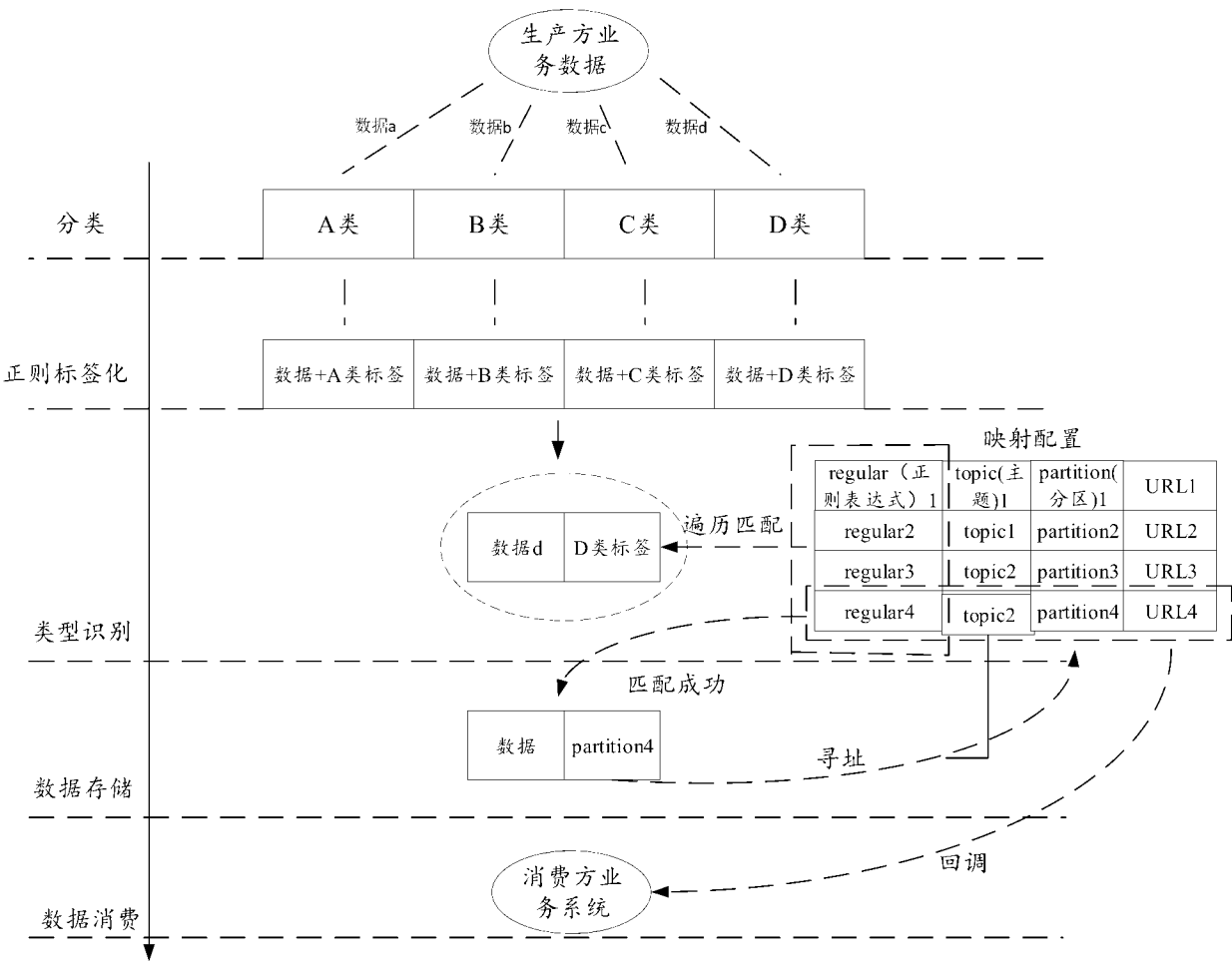


图 3

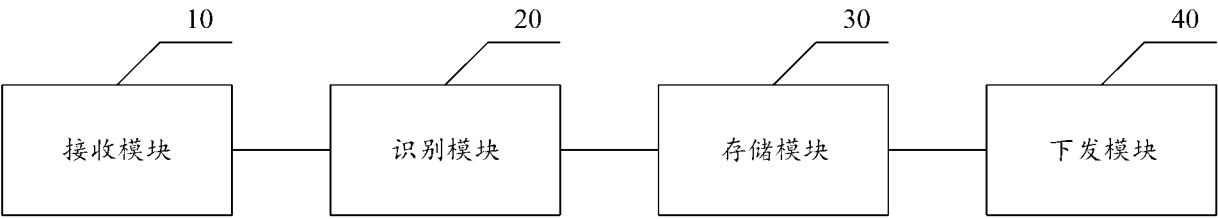


图 4

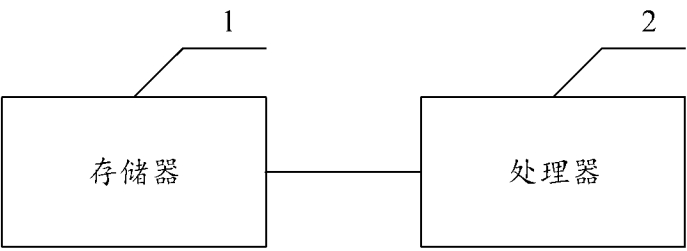


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/095599

A. CLASSIFICATION OF SUBJECT MATTER G06F 9/54(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC: G06F Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNTXT, ENTXT, ENTXTC, DWPI, CNKI: 标签, 分类, 消息, message, class+, label, tag		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 115543662 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 30 December 2022 (2022-12-30) claims 1-20	1-20
X	CN 115328678 A (JINAN INSPUR DATA TECHNOLOGY CO., LTD.) 11 November 2022 (2022-11-11) description, paragraphs [0038]-[0068], and figures 1-2	1-20
A	CN 112380040 A (BEIJING MEISHE NETWORK TECHNOLOGY CO., LTD.) 19 February 2021 (2021-02-19) entire document	1-20
A	US 2011099234 A1 (JHA RUCHIR P et al.) 28 April 2011 (2011-04-28) entire document	1-20
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: “A” document defining the general state of the art which is not considered to be of particular relevance “D” document cited by the applicant in the international application “E” earlier application or patent but published on or after the international filing date “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) “O” document referring to an oral disclosure, use, exhibition or other means “P” document published prior to the international filing date but later than the priority date claimed “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art “&” document member of the same patent family		
Date of the actual completion of the international search 16 August 2023		Date of mailing of the international search report 28 August 2023
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/095599

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	115543662	A	30 December 2022	CN	115543662	B	21 March 2023
CN	115328678	A	11 November 2022	None			
CN	112380040	A	19 February 2021	CN	112380040	B	07 May 2021
US	2011099234	A1	28 April 2011	US	8082313	B2	20 December 2011

A. 主题的分类

G06F 9/54 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

IPC: G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNXTX、ENTXT、ENTXTC、DWPI、CNKI:标签, 分类, 消息, message, class+, label, tag

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 115543662 A (苏州浪潮智能科技有限公司) 2022年12月30日 (2022 - 12 - 30) 权利要求1-20	1-20
X	CN 115328678 A (济南浪潮数据技术有限公司) 2022年11月11日 (2022 - 11 - 11) 说明书[0038]-[0068]段、图1-2	1-20
A	CN 112380040 A (北京美摄网络科技有限公司) 2021年2月19日 (2021 - 02 - 19) 全文	1-20
A	US 2011099234 A1 (JHA RUCHIR P et al.) 2011年4月28日 (2011 - 04 - 28) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年8月16日

国际检索报告邮寄日期

2023年8月28日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

冯慧萍

电话号码 (+86) 62411838

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/095599

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	115543662	A	2022年12月30日	CN 115543662 B	2023年3月21日
CN	115328678	A	2022年11月11日	无	
CN	112380040	A	2021年2月19日	CN 112380040 B	2021年5月7日
US	2011099234	A1	2011年4月28日	US 8082313 B2	2011年12月20日