

(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
2017 年 10 月 5 日 (05.10.2017)



(10) 国际公布号
WO 2017/167100 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2017/077722
- (22) 国际申请日: 2017 年 3 月 22 日 (22.03.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610202889.0 2016 年 3 月 31 日 (31.03.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 及
- (71) 申请人 (仅对美国): 陈逸喆 (CHEN, Yizhe) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (CN)。

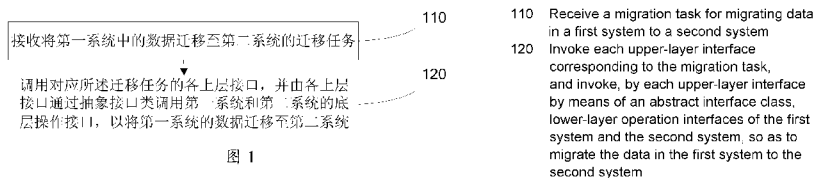
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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))。

(54) Title: DATA MIGRATION METHOD AND DEVICE

(54) 发明名称: 一种数据迁移方法和装置



(57) Abstract: Provided are a data migration method and device that relate to the field of computer technologies. The method comprises: receiving a migration task for migrating data in a first system to a second system (110, 210); and invoking each upper-layer interface corresponding to the migration task, and invoking, by each upper-layer interface by means of an abstract interface class, lower-layer operation interfaces of the first system and the second system, so as to migrate the data in the first system to the second system (120). By means of the invention, bidirectional data transmission and migration can be implemented between any two data terminals, and between any data terminal and another data terminal comprising cloud storage products that do not yet support data migration, such as Baidu object storage and Jinshan object storage.

(57) 摘要: 一种数据迁移方法和装置, 涉及计算机技术领域。所述方法包括: 接收将第一系统中的数据迁移至第二系统的迁移任务 (110, 210); 调用对应所述迁移任务的各上层接口, 并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口, 以将第一系统的数据迁移至第二系统 (120)。首先, 实现了任意两个数据端可以进行双向的数据传输和迁移。其次, 可以实现将任意的数据终端, 包括尚未支持数据迁移的云存储产品, 如百度对象存储、金山对象存储, 与其它任意一个数据端进行数据的双向传输迁移。



WO 2017/167100 A1

一种数据迁移方法和装置

技术领域

本申请涉及计算机技术领域，特别是涉及一种数据迁移方法和一种数据迁移装置。

5 背景技术

随着国内外云计算平台、市场的快速发展，云存储产品也越来越多。在云计算平台上，云服务器是最基础、最重要的产品，其次就是云存储。各大云服务商都有自己的云存储产品，如 AWS（Amazon Web Services，亚马逊云服务）的 S3、阿里云的 OSS（Object Storage Service，对象存储服务）、微软云计算平台 Azure 的存储数据服务（Blob）。另外，还有像七牛云存储这种专门提供云存储服务的服务商。

虽然云存储产品种类很多，它们在功能和使用方法上都有一些共通点，但具体到 API（Application Programming Interface，应用程序编程接口）调用、SDK（Software Development Kit，软件开发工具包）的编程上，又都有一些差异和各自的特点。在实际使用过程中，有时因为整体业务的搬迁，可能会从一个云服务商切换到另一个服务商，这时候就需要相应地把云存储的数据进行迁移同步；或者有时为了数据的安全性，不仅在同个服务商提供的多个云存储产品上同步备份数据，甚至会把数据备份到另一服务商提供的云存储产品，减少风险；又或者将客户端文件系统中的数据同步迁移到某一云服务商提供的云存储产品中。

但是，先有的数据迁移技术或工具，只能针对性地在某一服务商提供的云存储产品与用户之间进行数据传输，不支持其它云存储产品；或者是将其他存储系统中的数据迁移到某一服务商提供的云存储产品，不能进行任意两个存储系统之间的数据的双向迁移。

发明内容

鉴于上述问题，提出了本申请实施例以便提供一种克服上述问题或者至少部分地解决上述问题的一种数据迁移方法和相应的一种数据迁移装置。

5 本申请公开了一种数据迁移方法，其特征在于，包括：

接收将第一系统中的数据迁移至第二系统的迁移任务；

调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

10 优选地，所述调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤，包括：

确定属于获取数据的各上层接口，以及属于同步数据的各上层接口；

15 对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据；

对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统。

优选地，对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口的步骤，包括：

20 对于属于获取数据的各上层接口中的第一上层接口，以抽象接口类中与第一上层接口对应的第一抽象类接口作为底层操作接口以直接调用；所述第一上层接口为各系统通用的接口；

对于属于获取数据的各上层接口中的第二上层接口，通过抽象接口类生成针对第一系统的第一接口类，在所述第一接口类中调用与各第二上层接口对应的底层操作接口；所述第二上层接口不为各系统通用的接口。

25

优选地，对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口的步骤，包括：

对于属于同步数据的各上层接口中的第三上层接口，以抽象接口类中与第三上层接口对应的第二抽象类接口作为底层操作接口以直接调用；所述第三上层接口为各系统通用的接口；

对于属于同步数据的各上层接口中的第四上层接口，通过抽象接口类生成针对第二系统的第二接口类，在所述第二接口类中调用与各第四上层接口对应的底层操作接口；所述第四上层接口不为各系统通用的接口。

优选地，所述对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据的步骤，包括：

由管理组件的扫描器执行获取元数据的上层接口，所述获取元数据的上层接口通过抽象类接口调用第一系统的获取元数据的底层操作接口，从第一系统获取所述迁移任务对应的各元数据；

由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器；

由工作组件的接收器根据接收到同步任务，执行获取实际数据的上层接口，所述获取实际数据的上层接口通过抽象类接口调用第一系统的获取实际数据的底层获取接口，从第一系统获取所述迁移任务的实际数据。

优选地，所述对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统的步骤，包括：

由工作组件的同步器执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

优选地，所述管理组件在集群的一台服务器中，所述工作组件在集群

的各台服务器中。

优选地，所述由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器的步骤，包括：

提取元数据中记录的数据量大小；

- 5 如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

如果元数据记录的数据量不小于第一阈值，则将所述元数据作为同步任务分发给负载最小的服务器中的工作组件的接收器；

- 10 优选地，所述由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器的步骤，包括：

提取元数据中记录的数据量大小；

如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

- 15 如果元数据中记录的数据量不小于第一阈值但是小于第二阈值时，则将所述元数据生成同步任务分发给负载最小的服务器中的工作组件的接收器；

如果元数据中记录的数据量不小于第二阈值时，则调用对大文件切片的上层接口，所述对大文件切片的上层接口通过抽象接口类调用第二系统的对大文件切片的底层操作接口，将所述元数据对应的大文件切片；

- 20 将对所述元数据的不同切片的切片信息生成同步任务，并分发给负载低的多个服务器的工作组件的接收器。

优选地，所述集群的各台服务器中之间采用 JGroups 通信协议传输数据。

- 25 优选地，所述由工作组件的同步器执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将

所述实际数据同步至第二系统的步骤，包括：

由工作组件的同步器根据限速指令，调用与限速指令相应的线程执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

- 5 优选地，调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤之后，还包括：

判断将第一系统中的数据迁移至第二系统是否成功；

- 10 如果失败，则重复执行调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤，直至重试次数达到重试阈值或者在重试阈值内成功。

- 15 优选地，在调用对应所述迁移任务的各上层接口，通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的过程中，还包括：

判断是否接收到暂停指令；

如果接收到暂停指令，则暂停将第一系统的数据迁移至第二系统，并记录暂停时刻第一系统中已迁移的数据位置；

则进一步的，在暂停将第一系统的数据迁移至第二系统之后，还包括：

- 20 判断是否接收到重启指令；

如果接收到重启指令，则根据暂停时刻第一系统中已迁移的数据位置开始，重新调用对应所述迁移任务的各上层接口，通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤。

- 25 优选地，当第一系统为客户端文件系统时，在接收将第一系统中的数

据迁移至第二系统的迁移任务的步骤之前,还包括:

在客户端中安装应用程序,所述应用程序针对指定文件模拟各底层操作接口。

优选地,所述调用对应所述迁移任务的各上层接口,并由各上层接口
5 通过抽象接口类调用第一系统和第二系统的底层操作接口,以将第一系统的数据迁移至第二系统的步骤之后,还包括:

监控客户端文件系统是否与最近一次迁移到第二系统的数据出现变更;

如果出现变更,则调用对应所述迁移任务的各上层接口,并由各上层
10 接口通过抽象接口类调用第一系统和第二系统的底层操作接口,以将客户端文件系统的变更数据同步至第二系统。

优选地,所述对于属于获取数据的各上层接口,通过抽象接口类调用第一系统的底层操作接口,以从第一系统获取数据的步骤之后,还包括:

调用签名的上层接口,并由所述签名的上层接口调用第一系统的签名
15 的底层操作接口,对所述数据进行签名;

进一步的,对于属于同步数据的各上层接口,通过抽象接口类调用第二系统的底层操作接口,以将从第一系统获取的数据同步至第二系统的步骤之后,还包括:

调用签名认证的上层接口,并由所签名认证的上层接口调用第二系统
20 的签名认证的底层操作接口,对所述数据进行签名验证。

优选地,在调用对应所述迁移任务的各上层接口,并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口,以将第一系统的数据迁移至第二系统的步骤之后,还包括:

针对第一系统调用文件列表的上层接口,并由所述调用文件列表的上
25 层接口通过抽象接口类调用第一系统的底层操作接口,获取第一系统的文

件列表;

针对第二系统调用文件列表的上层接口,并由所述调用文件列表的上层接口通过抽象接口类调用第二系统的底层操作接口,获取第二系统的文件列表;

- 5 比较第一系统的文件列表和第二系统的文件列表,并将两者之间的差异提示给客户端。

优选地,所述迁移任务由客户端发送,所述迁移任务包括第一系统的标识信息、用户在第一系统的身份信息、待迁移数据在第一系统中的文件信息、第二系统的标识信息、用户在第二系统的身份信息。

- 10 优选地,在确定属于获取数据的各上层接口,以及属于同步数据的各上层接口的步骤之后,还包括:

针对第一系统调用身份验证的上层接口,并由所述身份验证的上层接口通过抽象接口类调用第一系统的底层操作接口,对用户在第一系统的身份信息进行验证;

- 15 针对第二系统调用身份验证的上层接口,并由所述身份验证的上层接口通过抽象接口类调用第二系统的底层操作接口,对用户在第二系统的身份信息进行验证。

本申请还公开了一种数据迁移装置,其特征在于,包括:

- 20 任务接收模块,适于接收将第一系统中的数据迁移至第二系统的迁移任务;

迁移模块,适于调用对应所述迁移任务的各上层接口,并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口,以将第一系统的数据迁移至第二系统。

优选地,所述迁移模块,包括:

- 25 上层接口确定子模块,适于确定属于获取数据的各上层接口,以及属

于同步数据的各上层接口；

数据获取子模块，适于对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据；

数据同步子模块，适于对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统。

优选地，所述数据获取子模块，包括：

第一底层操作接口确定子模块，适于对于属于获取数据的各上层接口中的第一上层接口，以抽象接口类中与第一上层接口对应的第一抽象类接口作为底层操作接口以直接调用；所述第一上层接口为各系统通用的接口；

第一底层操作接口调用子模块，适于对于属于获取数据的各上层接口中的第二上层接口，通过抽象接口类生成针对第一系统的第一接口类，在所述第一接口类中调用与各第二上层接口对应的底层操作接口；所述第二上层接口不为各系统通用的接口。

优选地，所述数据同步子模块，包括：

第二底层操作接口确定子模块，适于对于属于同步数据的各上层接口中的第三上层接口，以抽象接口类中与第三上层接口对应的第二抽象类接口作为底层操作接口以直接调用；所述第三上层接口为各系统通用的接口；

第二底层操作接口调用子模块，适于对于属于同步数据的各上层接口中的第四上层接口，通过抽象接口类生成针对第二系统的第二接口类，在所述第二接口类中调用与各第四上层接口对应的底层操作接口；所述第四上层接口不为各系统通用的接口。

优选地，所述数据获取子模块，包括：

元数据获取子模块，适于由管理组件的扫描器执行获取元数据的上层接口，所述获取元数据的上层接口通过抽象类接口调用第一系统的获取元数据的底层操作接口，从第一系统获取所述迁移任务对应的各元数据；

任务分发子模块，适于由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器；

实际数据获取子模块，适于由工作组件的接收器根据接收到同步任务，执行获取实际数据的上层接口，所述获取实际数据的上层接口通过抽象类接口调用第一系统的获取实际数据的底层获取接口，从第一系统获取所述迁移任务的实际数据。

优选地，所述数据同步子模块，包括：

实际数据同步子模块，适于由工作组件的同步器执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

优选地，所述管理组件在集群的一台服务器中，所述工作组件在集群的各台服务器中。

优选地，所述任务分发子模块，包括：

数据量大小提取子模块，适于提取元数据中记录的数据量大小；

第一元数据分发子模块，适于如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

第二元数据分发子模块，适于如果元数据记录的数据量不小于第一阈值，则将所述元数据作为同步任务分发给负载最小的服务器中的工作组件的接收器。

优选地，所述任务分发子模块，包括：

数据量大小提取子模块，适于提取元数据中记录的数据量大小；

第一元数据分发子模块，适于如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

5 第三元数据分发子模块，适于如果元数据中记录的数据量不小于第一阈值但是小于第二阈值时，则将所述元数据生成同步任务分发给负载最小的服务器中的工作组件的接收器；

元数据切分子模块，适于如果元数据中记录的数据量不小于第二阈值时，则调用对大文件切片的上层接口，所述对大文件切片的上层接口通过抽象接口类调用第二系统的对大文件切片的底层操作接口，将所述元数据
10 对应的大文件切片；

将对所述元数据的不同切片的切片信息生成同步任务，并分发给负载低的多个服务器的工作组件的接收器。

优选地，所述集群的各台服务器中之间采用 JGroups 通信协议传输数据。

15 优选地，所述实际数据同步子模块，包括：

第一实际数据同步子模块，适于由工作组件的同步器根据限速指令，调用与限速指令相应的线程执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

20 优选地，在所述迁移模块之后，还包括：

迁移判断模块，适于判断将第一系统中的数据迁移至第二系统是否成功；如果失败，则重复执行迁移模块，直至重试次数达到重试阈值或者在重试阈值内成功。

优选地，在所述迁移模块之后，还包括：

25 暂停指令判断模块，适于判断是否接收到暂停指令；如果接收到暂停

指令，则进入暂停记录子模块；

暂停记录模块，适于暂停将第一系统的数据迁移至第二系统，并记录暂停时刻第一系统中已迁移的数据位置；

则进一步的，在所述暂停指令判断模块之后，还包括：

- 5 重启指令判断模块，适于判断是否接收到重启指令；如果接收到重启指令，则根据暂停时刻第一系统中已迁移的数据位置开始，重新进入迁移模块。

优选地，当第一系统为客户端文件系统时，在任务接收模块之前，还包括：

- 10 应用程序安装模块，适于在客户端中安装应用程序，所述应用程序针对指定文件模拟各底层操作接口。

优选地，在所述迁移模块之后，还包括：

数据监控模块，适于监控客户端文件系统是否与最近一次迁移到第二系统的数据出现变更；如果出现变更，则进入变更数据同步模块；

- 15 变更数据同步模块，适于调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将客户端文件系统的变更数据同步至第二系统。

优选地，在所述数据获取子模块之后，还包括：

- 20 签名子模块，适于调用签名的上层接口，并由所述签名的上层接口调用第一系统的签名的底层操作接口，对所述数据进行签名；

进一步的，在所述数据同步子模块之后，还包括：

签名验证子模块，适于调用签名认证的上层接口，并由所签名认证的上层接口调用第二系统的签名认证的底层操作接口，对所述数据进行签名验证。

- 25 优选地，在所述迁移模块之后，还包括：

第一文件列表获取模块，适于针对第一系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第一系统的底层操作接口，获取第一系统的文件列表；

第二文件列表获取模块，适于针对第二系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第二系统的底层操作接口，获取第二系统的文件列表；

差异提示模块，适于比较第一系统的文件列表和第二系统的文件列表，并将两者之间的差异提示给客户端。

优选地，所述迁移任务由客户端发送，所述迁移任务包括第一系统的标识信息、用户在第一系统的身份信息、待迁移数据在第一系统中的文件信息、第二系统的标识信息、用户在第二系统的身份信息。

优选地，在所述上层接口确定子模块之后，还包括：

第一身份验证模块，适于针对第一系统调用身份验证的上层接口，并由所述身份验证的上层接口通过抽象接口类调用第一系统的底层操作接口，对用户在第一系统的身份信息进行验证；

第二身份验证模块，适于针对第二系统调用身份验证的上层接口，并由所述身份验证的上层接口通过抽象接口类调用第二系统的底层操作接口，对用户第二系统的身份信息进行验证。

本申请实施例包括以下优点：

本申请实施例，接收将第一系统中的数据迁移至第二系统的迁移任务；然后调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

首先，实现了任意两个数据端可以进行双向的数据传输和迁移。

其次，本申请所述的数据迁移技术具有良好的扩展功能，可以实现将

任意的数据终端，包括尚未支持数据迁移的云存储产品，如百度对象存储、金山对象存储，与其它任意一个数据端进行数据的双向传输迁移。

附图说明

图 1 是本申请的一种数据迁移方法实施例的步骤流程图；

5 图 1A 是本申请的一种五个系统进行数据迁移的示意图；

图 1B 是本申请的一种五个系统的数据迁移过程接口调用示意图；

图 2 是本申请的一种数据迁移方法实施例的步骤流程图；

图 2C 是本申请的一种数据迁移结构示意图；

图 2D 是本申请的一种分布式模式数据迁移数据流示意图；

10 图 3 是本申请的一种数据迁移装置实施例的结构框图；

图 4 是本申请的一种数据迁移装置实施例的结构框图。

具体实施方式

为使本申请的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。

15 本申请实施例的核心构思之一在于，接收将第一系统中的数据迁移至第二系统的迁移任务；所述第一系统和第二系统为多个系统中的任意两个系统，然后调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

20 首先，实现了任意两个数据端可以进行双向的数据传输和迁移，包括数据端内部的 bucket 间迁移。

其次，本申请所述的数据迁移技术具有良好的扩展功能，可以实现将任意的数据终端，包括尚未支持数据迁移的云存储产品，如百度对象存储、金山对象存储，与其它任意一个数据端进行数据的双向传输迁移。

实施例一

参照图 1，示出了本申请的一种数据迁移方法实施例的步骤流程图，具体可以包括如下步骤：

步骤 110，接收将第一系统中的数据迁移至第二系统的迁移任务；所述
5 第一系统和第二系统为多个系统中的任意两个系统。

本申请的一种数据迁移方法是将一个系统中的数据迁移至另一个系统的方法，是针对多个系统中的任一两个系统。其中的系统可以是具有存储功能的服务器，或者存储空间、云存储产品等，例如客户端文件系统（Local）、AWS（Amazon Web Services，亚马逊云服务）的 S3、阿里云
10 的 OSS（Object Storage Service，对象存储服务）、微软云计算平台 Azure 的存储数据服务（Blob）、七牛云存储（QiNiu）等。如图 1A 为前述五个系统进行数据迁移的示意图。可以看出有，多个系统中的任意两个系统可以进行双向的数据传输和迁移，包括系统内部的不同存储区间（bucket）之间的数据迁移。其中，Bucket 是一个用户用来管理所存储对象的存储空间，类似于文件系统里面的根目录或者磁盘分区。而且本申请的数据迁移
15 方法具有良好的扩展功能，对于其它尚未加入图 1A 所示五个系统的云存储产品，如百度对象存储（Baidu Object Storage，BOS）、金山对象存储（Kingsoft Standard Storage Service，KS3）等，也可以很方便地加入，并且可以跟其它任意一个系统进行数据的双向传输与迁移。

20 为了区分两个系统，可以定义将第一系统中的数据迁移至第二系统。其中的第一系统和第二系统可以为两个相同类型的系统，例如都为阿里云的 OSS，那么此时是将一个 OSS 的数据迁移至另一个 OSS，比如将一个地区的 OSS 中的数据迁移至另一个地区的 OSS 中；当然，第一系统和第二系统也可以为两个不同类型的系统，例如分别为 OSS 和 Blob，那么此
25 时是将 OSS 的数据迁移至 Blob，或者也可以为同一系统中的两个不同的

存储区，那么此时是将同一系统中的一个存储区间的数据迁移至另一个存储区间。

对于具体的第一系统和第二系统，以及迁移数据，例如是将第一系统中的全部数据迁移至第二系统，还是只迁移一部分，可以在迁移任务中设定。接收数据迁移任务的服务器可以从迁移任务中获取上述信息。

在本申请的另一个优选地实施例中，所述迁移任务由客户端发送，所述迁移任务包括第一系统的标识信息、用户在第一系统的身份信息、待迁移数据在第一系统中的文件信息、第二系统的标识信息、用户在第二系统的身份信息。其中，第一系统的标识信息是指可以唯一标识第一系统的信息，例如第一系统的访问域名等；用户在第一系统的身份信息是指当前第一系统用户的身份信息，例如用户名等；第二系统的标识信息是指可以唯一标识第二系统的信息，例如第二系统的 IP 地址等；用户在第二系统的身份信息是指当前第二系统用户的身份信息，例如用户名等；待迁移数据在第一系统中的文件信息可以包括待迁移数据存储在第一系统中的文件的名称、存储位置、数据大小等信息。

在实际应用中，本申请实施例可以在客户端中提供一个迁移配置页面，用户可以在该迁移配置页面中配置相关信息，作为被迁移对象的第一系统的标识信息、用户在第一系统的身份信息、待迁移数据在第一系统中的文件信息、作为迁移目标对象的第二系统的标识信息、用户在第二系统的身份信息，当然还包括迁移对象标识。其中，如果用户需要迁移第一系统的全部数据，则可以配置迁移对象标识为 all，all 标识迁移第一系统的全部数据；如果用户需要迁移某个或者某些文件，则可以配置该文件或者该几个文件在第一存储系统中的存储路径作为迁移对象标识。当然，还可以配置其他需求的信息，比如有的存储系统是分区域，比如北京有一个存储中心，上海有一个存储中心，那么还可以配置该存储系统所在区域

Region。具体的配置项可以根据需求设置，本申请实施例不对其加以限制。

步骤 120，调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

- 5 在本申请实施例中，可以将多个服务器连接起来形成一个大的处理集群。整个集群有一台控制节点（Master）服务器和多台工作节点（Slave）服务器，Master 用以接收将第一系统中的数据迁移至第二系统的迁移任务，然后通过合理的分发策略，把接收到的迁移任务分发给 Slave 进行处理。Slave 处理完接收到的迁移任务后，可以把迁移结果反馈给 Master。
- 10 当然，本申请实施例也可以用单台服务器实现数据的迁移。优选的本申请数量采样前述集群的方式。

如图 1B 为对前述五个系统的数据迁移过程接口调用示意图。其中，上层（Peer）接口定义了每个具体的进行数据迁移的系统都必须具有的功能。例如，get()，put()，分别是指下载、上传数据，这是进行实际数据传输需要调用的接口；

15

mkRoot()，rmRoot()：建立、删除根目录。操作的对象就是系统中的 Bucket；Bucket 存储空间，是一个用户用来管理所存储 Object 的存储空间，类似于文件系统里面的根目录或者磁盘分区。

getObj()，getAllObjs()：获取单个元数据、获取全部元数据，主要就是利用这些接口，获取要传输的数据的元数据；元数据 Metadata 作为实际存储数据的描述型数据，只包含一些基础信息，如：标识数据唯一性的 Key、数据大小、时间戳等。

20

beginPartPut()，partPut()，endPartPut()：主要处理断点续传、大文件的切片传输功能；

25 rm()，rmAll()：删除单个数据，删除全部数据。

其中，对于某一些上层接口，如 `getAllObjs()`，`rmAll()`，它们的实现是通过调用其它底层操作接口如 `getObj()`，`rm()`来实现，这对于任何一个进行数据迁移的系统来说都是通用的，可以在图 2B 中所示的抽象接口类 (`AbstractPeer`) 层中实现。

5 所以在本申请实施例中，可以在 `AbstractPeer` 这一层实现调用底层操作接口，例如图 1B 中的 `Local`、`OSS`、`S3`、`QiNiu`、`Blob` 构成的底层的操作接口。所以，可以调用对应迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

10 例如，若某一迁移任务是要将 `Local` 中的数据迁移至 `OSS`，则可以调用对应该迁移任务的各上层接口，包括用以获取 `Local` 中数据的上层接口和用以将获取的数据同步到 `OSS` 中的上层接口，然后首先可以由获取 `Local` 中数据的上层接口通过与之对应的抽象接口类依次调用 `Local` 中的 `getObj()` 接口、`get()` 接口，以获取 `Local` 中的迁移数据，然后可以由用以将获取的数据同步到 `OSS` 中的上层接口通过与之对应的抽象接口类依次调用 `OSS` 中的 `put()` 接口，将从 `Local` 中获取的迁移数据传输至 `OSS`。

由于本申请设置一个统一的上层接口层 `peer`，在其中构建了各种操作接口。又构建了一个抽象接口类层 `AbstractPeer`，该抽象接口类层可以继承上层接口层。然后对于各个存储系统（如前述 `Local`、`OSS`、`S3`、`QiNiu`、
20 `Blob`）本身的底层操作接口，可以继承该 `AbstractPeer`，从而可以实现各个系统之间互相迁移数据的过程。同时，由于有抽象接口类层 `AbstractPeer`，对于新的存储系统，只需要将其底层的底层操作接口层继承抽象接口类层 `AbstractPeer`，即可实现该新的存储系统与之前的其他任意一个存储系统之间的数据迁移过程。

25 在本申请的另一个优选的实施例中，步骤 120 包括：

子步骤 121, 确定属于获取数据的各上层接口, 以及属于同步数据的各上层接口。

在本申请实施例中, 定义用以从第一系统中获取数据的上层接口为属于获取数据的上层接口, 用以将获取的数据同步至第二系统的上层接口为属于同步数据的上层接口。在实际应用中, 根据具体需求的不同, 属于获取数据的上层接口和属于同步数据的上层接口都可能为一个或多个上层接口, 对此本申请实施例不加以限制。

在实际应用中, 接收到一个迁移任务后, 首先要确定去某个存储系统获取数据要调用的各上层接口, 以及将数据存储至另一系统要调用的上传接口。

比如, 对于迁移任务中包括了要从中迁移数据的第一系统的标识信息, 以及待迁移的一个文件的文件信息, 如果不需要断点续传, 则需要调用图 1B 的 peer 中的上层接口 `getObj()` 获取元数据, 还要调用图 1B 的 peer 中上层接口 `get()` 获取实际数据, 那么 `getObj()` 和 `get()` 属于获取数据的上层接口。当然实际应用中, 还可以包括对第一系统的用户身份信息, 如用户在第一系统中的账户名和密码, 进行身份验证的上层接口。

而对于第二系统, 由于该文件是从第一系统迁移过来, 通常情况下第二系统是没有相应文件夹的, 那么需要调用图 1B 的 peer 中的 `mkRoot()` 建立目录, 以及图 1B 的 peer 中的 `put()` 上传数据。那么 `mkRoot()` 和 `put()` 属于同步数据的上层接口。当然实际应用中, 还可以包括对第二系统的用户身份信息, 如用户在第二系统中的账户名和密码, 进行身份验证的上层接口。

子步骤 122, 对于属于获取数据的各上层接口, 通过抽象接口类调用第一系统的底层操作接口, 以从第一系统获取数据。

如前述, 在实际应用中, 上层接口是通过抽象接口类调用第一系统的底层操作接口, 以将第一系统的数据迁移至第二系统。在将第一系统的数

据迁移至第二系统的过程中包括从第一系统中获取数据。此时，可以利用属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据。

在实际应用中，本申请实施例的服务器可以将第一系统的标识信息作为参数之一传给获取数据的上层接口，那么上层接口再调用抽象接口类，抽象接口类则可以根据该标识信息去调用第一系统的相应的底层操作接口。

在本申请的另一个优选的实施例中，子步骤 122 包括：

子步骤 1221，对于属于获取数据的各上层接口中的第一上层接口，以抽象接口类中与第一上层接口对应的第一抽象类接口作为底层操作接口以直接调用；所述第一上层接口为各系统通用的接口。

在实际应用中，对于属于获取数据的各上层接口中包括的各系统通用的接口，例如，`getAllObjs()`，`rmAll()`，可以称为第一上层接口。它们的实现是通过调用其它底层操作接口来实现的，因此对于属于获取数据的各上层接口中的第一上层接口，可以以抽象接口类中与第一上层接口对应的第一抽象类接口作为底层操作接口以直接调用。

例如，对于第一上层接口 `getAllObjs()`，在抽象接口类中第一抽象类接口 `getAllObjs()` 是通过调用底层操作接口 `getObj()` 实现的，所以，此时可以直接将第一抽象类接口 `getAllObjs()` 作为底层操作接口调用，相当于调用了底层操作接口 `getObj()`。

本申请实施例的服务器可以将第一系统的标识信息作为参数之一传给获取数据的上层接口，那么上层接口再调用抽象接口类，如果抽象接口类中有该上层接口对应的第一抽象类接口，则该第一抽象类接口可以根据该标识信息等参数执行。

子步骤 1222，对于属于获取数据的各上层接口中的第二上层接

口，通过抽象接口类生成针对第一系统的第一接口类，在所述第一接口类中调用与各第二上层接口对应的底层操作接口；所述第二上层接口不为各系统通用的接口。

而对于为各系统通用的各上层接口，可以称为第二上层接口，对于此类上层接口，则需要通过抽象接口类生成针对第一系统的第一接口类，然后在第一接口类中调用与各第二上层接口对应的底层操作接口。

本申请实施例的服务器可以将第一系统的标识信息作为参数之一传给获取数据的上层接口，那么上层接口再调用抽象接口类，如果抽象接口类中没有该上层接口对应的第一抽象类接口，则由抽象接口类根据该标识信息生成对应第一系统的第一接口类，然后在该第一接口类中调用上层接口对应的底层接口。

子步骤 123，对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统。

如前述，在实际应用中，上层接口是通过抽象接口类调用第一系统的底层操作接口，以将第一系统的数据迁移至第二系统。在将第一系统的数据迁移至第二系统的过程中包括将从第一系统获取的数据同步至第二系统。此时，可以利用属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统。

在实际应用中，本申请实施例的服务器可以将第二系统的标识信息作为参数之一传给获取数据的上层接口，那么上层接口再调用抽象接口类，如果抽象接口类中没有对应该上层接口的第一抽象类接口，抽象接口类则可以根据该标识信息去调用第二系统的相应的底层操作接口。

在本申请实施例中，对于步骤 122，在第一系统的获取数据的底层操作接口获取到需要迁移到第二系统的实质数据后，本申请实施例可以直接

将该实质数据传给第二系统的同步数据的存储接口，从而可以实现两个存储系统内部的 bucket 间的迁移。比如，在调用第一系统的 get () 获取到数据后，直接将该数据传给第二系统的 put ()，其处理过程均在内存中进行，那么该数据则直接迁移到第二系统，不会存储到本地磁盘中。在实际应用中，本申请实施例可以设定一个第一函数监控 get () 是否获取到数据，如果获取到数据，则将该数据传给 put ()。在本申请实施例中，获取到数据的底层接口传输给同步数据的底层操作接口后，在获取到数据的底层接口和同步数据的底层操作之间架构了一个传输管道，用于传输两类接口之间的数据，其每份数据传输后，不会在本地存储，也不会累积到内存中。

在本申请的另一个优选的实施例中，子步骤 123 包括：

子步骤 1231，对于属于同步数据的各上层接口中的第三上层接口，以抽象接口类中与第三上层接口对应的第二抽象类接口作为底层操作接口以直接调用；所述第三上层接口为各系统通用的接口。

此时，将属于同步数据的各上层接口中各系统通用的上层接口称为第三上层接口，以抽象接口类中与第三上层接口对应的第二抽象类接口作为底层操作接口以直接调用。本申请实施例的服务器可以将第二系统的标识信息作为参数之一传给获取数据的上层接口，那么上层接口再调用抽象接口类，如果抽象接口类中有该上层接口对应的第二抽象类接口，则该第二抽象类接口可以根据该标识信息等参数执行。该步骤与前述的步骤 1221 类似，在此不再赘述。

子步骤 1232，对于属于同步数据的各上层接口中的第四上层接口，通过抽象接口类生成针对第二系统的第二接口类，在所述第二接口类中调用与各第四上层接口对应的底层操作接口；所述第四上层接口不为各系统通用的接口。

此时，将属于同步数据的各上层接口中不为各系统通用的上层接口称为第四上层接口，通过抽象接口类生成针对第二系统的第二接口类，在所述第二接口类中调用与各第四上层接口对应的底层操作接口。在实际应用中，本申请实施例的服务器可以将第二系统的标识信息作为参数之一传给
5 获取数据的上层接口，如果抽象接口类中没有对应该上层接口的第二抽象类接口，那么上层接口再调用抽象接口类，抽象接口类则可以根据该标识信息去调用第二系统的相应的底层操作接口。该步骤与前述的步骤 1222 类似，在此不再赘述。

在本申请的另一个优选的实施例中，在步骤 120 之后，还包括：

10 步骤 130，判断将第一系统中的数据迁移至第二系统是否成功；如果失败，则重复执行步骤 120，直至重试次数达到重试阈值或者在重试阈值内成功。

在实际应用中，在数据迁移的过程中可能会出现干扰或错误等，导致数据迁移不成功，所以在本申请实施例中，会判断将第一系统中的数据
15 迁移至第二系统是否成功，所谓迁移成功，是指将需要迁移指令中需要进行迁移的数据全部迁移至第二系统中。具体地，可以通过判断已经迁移的数据大小是否等于迁移指令中设定的数据大小，或者是其他现有的可用技术比较已经迁移的数据和迁移指令中设定的数据是否一致，对此本申请实施例不加以限定。

20 当然，判断将第一系统中的数据迁移至第二系统是否成功可以采用多种方式，比如：可以通过监控第一系统下载成功与否的通知，以及监控第二系统上层成功与否的通知，如果收到任意一个不成功的通知，则确认迁移失败。当然，还有其他方式，本申请实施例不对其加以限制。

而且，在本申请实施例中，可以设置一个重试阈值，如果重试的次数
25 超过重试阈值仍未成功迁移，则放弃执行当前迁移任务，进而可以检查迁

移不成功的原因。重试阈值可以根据需求设定，对此本申请实施例不加以限定。

在本申请的另一个优选的实施例中，在步骤 120 之后，还包括：

步骤 140，判断是否接收到暂停指令；如果接收到暂停指令，则进入

5 步骤 150。

在数据迁移过程中，若因为某些原因需要暂停数据迁移，则此时可以发送暂停指令至执行迁移指令的服务器。所以，可以判断是否接收到暂停指令。

步骤 150，暂停将第一系统的数据迁移至第二系统，并记录暂停时刻
10 第一系统中已迁移的数据位置。

如果接收到暂停指令，则会暂停继续将第一系统的数据迁移至第二系统，同时，为了避免在再次开始将第一系统的数据迁移至第二系统时，又从初始情况开始，将第一系统中已经迁移至第二系统的数据再次重复迁移至第二系统，从而浪费时间，可以在暂停将第一系统的数据迁移至第二系
15 统的同时，记录暂停时刻第一系统中已迁移的数据位置，以便从已迁移的数据位置开始，继续进行数据迁移。

在实际应用中，本申请实施例中，也可以调用暂停的上层接口，然后通过抽象接口类分别调用第一系统的暂停的底层操作接口和第二系统的暂停的底层操作接口，实现迁移暂停过程。

20 步骤 160，判断是否接收到重启指令；如果接收到重启指令，则根据暂停时刻第一系统中已迁移的数据位置开始，重新进入步骤 120。

在接收到暂停指令之后，可以进一步判断判断是否接收到重启指令，如果接收到重启指令，则根据暂停时刻第一系统中已迁移的数据位置开始，重新执行调用对应所述迁移任务的各上层接口，并由各上层接口通过
25 抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数

据迁移至第二系统。此时第一系统的数据是指第一系统中的尚未被迁移至第二系统中的数据。

在实际应用中，本申请实施例中，也可以调用重启的上层接口，然后通过抽象接口类分别调用第一系统的重启的底层操作接口和第二系统的重启的底层操作接口，实现迁移暂停。

在本申请的另一个优选的实施例中，当第一系统为客户端文件系统时，在步骤 110 之前，还包括：

步骤 100，在客户端中安装应用程序，所述应用程序针对指定文件模拟各底层操作接口。

若第一系统为客户端文件系统，那么此时将客户端文件系统中的数据迁移至第二系统，可以理解为将客户端文件系统中的数据备份至第二系统。对于客户端而言，本申请为了实现客户端与各个系统之间是互通的，为了适应图 1B 的逻辑架构，本申请可以通过在客户端中安装应用程序，为客户提供服务。所以，在本申请实施例中，若第一系统为客户端文件系统，则首先在客户端中安装应用程序，所述应用程序针对指定文件模拟各底层操作接口，如图 1B 中的 local 下的各个底层操作接口。进而，客户端用户可以通过访问应用程序，调用可编辑该应用程序对应的指定文件的各底层操作接口。

在本申请的另一个优选的实施例中，在步骤 120 之后，还包括：

步骤 170，监控客户端文件系统是否与最近一次迁移到第二系统的数据出现变更；如果出现变更，则进入步骤 180。

在实际应用中，客户端的应用程序可能会被多个用户多次登录，而且在登录过程中，可以对客户端文件系统中的数据进行修改，如若客户端文件系统中的数据被变更，则相应地需要将变更部分更新至第二系统中。所以，针对客户端文件系统，可以监控客户端文件系统是否与最近一次迁移

到第二系统的数据出现变更，如果出现变更，则需要将变更更新至第二系统中，具体地，可以调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将客户端文件系统的变更数据同步至第二系统。

- 5 步骤 180，调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将客户端文件系统的变更数据同步至第二系统。

在本申请的另一个优选的实施例中，在子步骤 122 之后，还包括：

- 子步骤 124，调用签名的上层接口，并由所述签名的上层接口调用
10 第一系统的签名的底层操作接口，对所述数据进行签名。

在本申请实施例中，为了保证数据迁移前后的一致性，也即保证在执行完一次数据迁移任务，是将第一系统中的迁移数据完整的迁移至第二系统中，可以首先对第一系统中的需要进行迁移的数据进行签名，以标识需要迁移至第二系统的数据。具体地，可以调用签名的上层接口，并由签名的上层接口调用第一系统的签名的底层操作接口，对所述数据进行签名。
15 可以在第一系统中对数据进行签名，也可以在将数据移出第一系统，且尚未同步至第二系统时，对数据进行签名，对此本申请实施例不加以限定。

进一步地，在子步骤 123 之后，还包括：

- 子步骤 125，调用签名认证的上层接口，并由所签名认证的上层接口
20 调用第二系统的签名认证的底层操作接口，对所述数据进行签名验证。

理论上，若将需要迁移至第二系统中数据进行签名，则在数据迁移至第二系统后，可以对已迁移的数据进行签名验证。若签名验证成功，则表明已迁移至第二系统中的数据与第一系统中需要迁移至第二系统中的数据一致，而若签名验证不成功，例如已迁移至第二系统中的部分数据与其
25 在第一系统中的签名不对应等情况，则表明此时迁移失败，可以重复执行

步骤 120, 直至重试次数达到重试阈值或者在重试阈值内成功。

具体地, 可以调用签名认证的上层接口, 并由所签名认证的上层接口调用第二系统的签名认证的底层操作接口, 对数据进行签名验证。

在本申请的另一个优选的实施例中, 在步骤 120 之后, 还包括:

- 5 步骤 190, 针对第一系统调用文件列表的上层接口, 并由所述调用文件列表的上层接口通过抽象接口类调用第一系统的底层操作接口, 获取第一系统的文件列表。

- 10 步骤 1110, 针对第二系统调用文件列表的上层接口, 并由所述调用文件列表的上层接口通过抽象接口类调用第二系统的底层操作接口, 获取第二系统的文件列表。

步骤 1120, 比较第一系统的文件列表和第二系统的文件列表, 并将两者之间的差异提示给客户端。

在本申请的另一个优选的实施例中, 在子步骤 121 之后, 还包括:

- 15 子步骤 126, 针对第一系统调用身份验证的上层接口, 并由所述身份验证的上层接口通过抽象接口类调用第一系统的底层操作接口, 对用户在第一系统的身份信息进行验证。

子步骤 127, 针对第二系统调用身份验证的上层接口, 并由所述身份验证的上层接口通过抽象接口类调用第二系统的底层操作接口, 对用户

20 在第二系统的身份信息进行验证。

- 20 在本申请实施例中, 可以接收将第一系统中的数据迁移至第二系统的迁移任务, 所述第一系统和第二系统为多个系统中的任意两个系统, 然后调用对应所述迁移任务的各上层接口, 并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口, 以将第一系统的数据迁移至第二系统。

- 25 从而实现了任意两个数据端可以进行双向的数据传输和迁移; 其次,

本申请所述的数据迁移技术具有良好的扩展功能，可以实现将任意的数据终端，包括尚未支持数据迁移的云存储产品，如百度对象存储、金山对象存储，与其它任意一个数据端进行数据的双向传输迁移。

实施例二

5 参照图 2，示出了本申请的一种数据迁移方法实施例的步骤流程图，具体可以包括如下步骤：

步骤 210，接收将第一系统中的数据迁移至第二系统的迁移任务；所述第一系统和第二系统为多个系统中的任意两个系统。

10 如图 1A 中的示例，用户可以选择其中的任意两个系统，将其中一个系统作为第一系统，另一个作为第二系统。

步骤 220，确定属于获取数据的各上层接口，以及属于同步数据的各上层接口。

15 步骤 230，由管理组件的扫描器执行获取元数据的上层接口，所述获取元数据的上层接口通过抽象类接口调用第一系统的获取元数据的底层操作接口，从第一系统获取所述迁移任务对应的各元数据。

20 如图 2C 为本申请的一个数据迁移结构示意图。其由两大独立的组件构成：Master 和 Slave。Master 主要负责扫描数据源和任务分发，Slave 负责接收任务，并执行真正的数据迁移工作。Master 包括扫描器（Scanner）、分发器（Dispatcher）；Slave 包括接收器（Receiver）、同步器（Syncer）。
25 在本申请实施例中，可以称由扫描器、分发器构成的 Master 为管理组件，由接收器和同步器构成的 Slave 为工作组件。该结构既可以在单机环境运行，也支持分布式数据传输。在单机模式下，Master 和 Slave 作为一个进程内的两个组件同时提供服务，而在分布式模式下，整个集群里面有一台服务器中同时存在 Master 和 Slave，其余的服务器中只存在 Slave；或者是集群里面有一台服务器中只存在 Master，其余的服务器中只存在 Slave。

因为 Master 本身并不占用太多内存空间，所以 Master 和 Slave 可以共存于一个服务器中，以节省集群中的服务器成本。本申请实施例优选的采样集群的方式，即一个带 Master+Slave 的服务器和各个带 Slave 的服务器。

其中，扫描器负责遍历获取数据源的元数据（metadata），并存放到
5 队列“Metadata Queue（元数据队列）”里面。Metadata 作为实际存储数据的描述型数据，只包含一些基础信息，如：标识数据唯一性的 Key、数据量大小、时间戳等。另外，在本申请另一个优选地实施例中，Metadata Queue 可以为阻塞队列，阻塞队列与普通队列的区别在于，当队列是空的时候，从队列中获取数据的操作将会被阻塞，或者当队列是满时，往队列里
10 添加数据的操作会被阻塞。同样，试图往已满的阻塞队列中添加新数据的线程同样也会被阻塞，直到其他的线程使队列重新变得空闲起来，如从队列中移除一个或者多个数据，或者完全清空队列。则在数据迁移的过程中，首先将元数据按序存放至阻塞队列，直至阻塞队列已满，然后依次将阻塞队列中的元数据移出并同步至第二系统，每移出一个元数据则相应地可以
15 从第一系统中提取一个元数据存入阻塞队列，从而可以有效的均衡数据迁移的速度。

在本申请实施例中，管理组件的扫描器可以执行获取元数据的上层接口，进而获取元数据的上层接口可以通过抽象类接口调用第一系统的获取元数据的底层操作接口，从第一系统获取当前迁移任务对应的各元数据。

20 分发器负责从 Scanner 的 Metadata Queue 里面取出 metadata，根据分发算法，把 metadata 作为任务派送给 Slave 组件的接收器（Receiver）。每分发一个任务都会进行记录，Slave 完成一个任务后都会汇报给 Master 进行汇总，Master 可以实时知道当前整个集群里面所有 Slave 节点的任务数和服务器负载情况。

25 其中，在单机模式下可以直接把任务分发给本机的 Slave；而在分布式

模式下 Dispatcher 的分发算法，可以如下：

如果数据小于 1MB，就直接分发给本服务器的 Slave。因为对于小数据而言，通过网络分发到其它服务器执行，所耗费的总时间还不如直接在本服务器执行。

- 5 服务器接收器负责接收来自 Dispatcher 分发的任务，存放在自己的 Metadata Queue 里面，等待处理。

同步器真正执行数据传输的组件。它不断地从 Receiver 的 Metadata Queue 里面获取要传输数据的 metadata，放到不同的工作线程池里面执行传输任务。

- 10 其中，工作池线程是一种多线程处理形式，处理过程中将任务添加到队列，然后在创建线程后自动启动这些任务。线程池线程都是后台线程。使用工作池线程的好处包括：减少在创建和销毁线程上所花的时间以及系统资源的消耗；如不使用线程池，有可能造成系统创建大量线程而导致消耗过多系统内存以及需要在不同线程之间“过度切换”。

- 15 在本申请实施例中，工作线程池（Work Thread Pool）可以有两个，一个负责处理比较小的数据，一次性可以完成整个元数据对应的文件的传输。另一个则负责处理比较大的数据，例如接收比较大的元数据对应的大文件的切片数据，对于大文件而言，需要把所有切片传输完成后，才能在第二系统中完成整个大文件的拼接还原，其中，可以预置一个阈值，若元
- 20 数据小于该阈值，则由负责处理比较小的数据的工作线程池处理该元数据，若元数据不小于该阈值，则由负责处理比较大的数据的工作线程池处理该元数据。当然，工作线程池可以只有一个，对全部数据统一处理，例如若不对大文件进行切片处理，对此本申请实施例不加以限定。

- 25 如图 2D 为分布式模式数据迁移数据流示意图。其中，将云存储 A 中的数据迁移至云存储 B，主要可以分为以下四步：

(1) Scanner 扫描数据源“云存储 A”，获取 metadata。

(2) Dispatcher 从 Scanner 获取 metadata，根据分发算法派送给 Slave 的 Receiver。

(3) Syncer 从 Receiver 获取 metadata，开始真正的数据传输。从数据源“云存储 A”下载获取实际数据，不经过任何中间文件保存，直接上传到目的数据端“云存储 B”，实现端到端的直接传输。

(4) Syncer 把每个传输结果回传给 Receiver，Receiver 再汇报给 Master 的 Dispatcher 进行汇总处理。

在本申请另一个优选的实施例中，所述管理组件在集群的一台服务器中，所述工作组件在集群的各台服务器中。

在本申请另一个优选的实施例中，所述集群的各台服务器中之间采用 JGroups 通信协议传输数据。

JGroups 由纯 Java 编写实现，是一个可靠的消息传播工具包。利用它可以简单地构建起一个多节点的通信集群，各个节点间可以互相通信，进行协作。利用 JGroups，用户可以自己搭建一个对等节点网络，或者是 Master-Slave 的集群模式。

本申请的分布式集群，采用的是 Master-Slave 模式。利用 JGroups 的发现协议，只要预先在每个节点配置好 Master 节点的 IP+端口地址，然后约定好一个统一的集群名称。先启动 Master 节点后，再依次启动各个 Slave 节点，这样每个 Slave 节点就都能发现找到 Master 节点并自动加入到这个集群里面。如果某个 Slave 节点退出，利用 JGroups 的 FD(Failure detection) 协议能够自动探测到该退出的 Slave 节点，并把该节点的信息从整个集群中移除。

在实际应用中，JGroups 提供两种消息传播机制：

(1) 字节流的消息直传。一个节点可以往另外一个节点发送字节

(byte)流数据。这种模式一般需要发送端先把要传输的数据序列化成 byte 数组，然后在接收端再反序列化回数据。

(2) 类方法的直接 RPC (Remote Procedure Call Protocol, 远程过程调用协议) 调用。JGroups 支持在一个节点直接远程调用另外一个节点里
5 面某个类的方法，就跟普通的方法调用一样。这种方法可以避免对于消息的序列化处理，更加方便。

因此，本申请的 Dispatcher 和 Receiver 之间的数据传输就是通过 JGroups 的 RPC 功能来实现的。

步骤 240，由管理组件的分发器，基于各元数据生成同步任务并分发
10 给各工作组件的接收器。

同步任务是数据迁移任务的一种，同步任务中同样可以包括第一系统的标识信息、用户在第一系统的身份信息、待迁移数据在第一系统中的文件信息、第二系统的标识信息、用户在第二系统的身份信息。其中的待迁移数据是指个元数据。

15 在本申请另一个优选的实施例中，步骤 240，包括：

子步骤 241，提取元数据中记录的数据量大小。

如前述，在各元数据中可以包含数据量大小的参数，因此，可以直接从元数据中提取其中记录的数据量大小。

子步骤 242，如果元数据中记录的数据量小于第一阈值，直接将所
20 述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器。

如果元数据中记录的数据量小于第一阈值，则说明该元数据不会给服务器增加过多的负载量，所以可以直接将该元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器。其中，第一阈值可以根据需求设定，对此本申请实施例不加以限定。
25

比如，如果数据小于 1MB，Dispatcher 就直接分发给本服务器的 Slave 的 Receiver。因为对于小数据而言，通过网络分发到其它服务器执行，所耗费的总时间还不如直接在本服务器执行。

子步骤 243，如果元数据记录的数据量不小于第一阈值，则将所述元数据作为同步任务分发给负载最小的服务器中的工作组件的接收器。

对于分布式模式下的数据迁移，如果元数据记录的数据量不小于第一阈值，此时为了均衡集群中各服务器的负载量，即让各服务器的负载量不会相差太大，可以将该元数据作为同步任务分发给负载最小的服务器中的工作组件的接收器。

比如，如果数据大于等于 1MB，Dispatcher 就将该数据分发给其他服务器的 slave 的 Receiver。

在本申请另一个优选的实施例中，步骤 240，包括：

子步骤 241，提取元数据中记录的数据量大小。

子步骤 242，如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器。

子步骤 243，如果元数据中记录的数据量不小于第一阈值但是小于第二阈值时，则将所述元数据生成同步任务分发给负载最小的服务器中的工作组件的接收器。

在本申请实施例中，为了避免过大的元数据在传输的过程中效率太低，所以可以另外设置一个第二阈值，同样的，第二阈值可以根据需求设定，但是第二阈值必须大于第一阈值，对此本申请实施例不加以限定。

如果元数据中记录的数据量不小于第一阈值但是小于第二阈值时，则将元数据生成同步任务分发给负载最小的服务器中的工作组件的接收器。

子步骤 244，如果元数据中记录的数据量不小于第二阈值时，则调

用对大文件切片的上层接口，所述对大文件切片的上层接口通过抽象接口类调用第二系统的对大文件切片的底层操作接口，将所述元数据对应的大文件切片。

而若元数据中记录的数据量不小于第二阈值，则说明此时该元数据的数据量过大，为了保证数据迁移的速度，可以将该元数据对应的大文件进行切片处理，将大文件分为若干个较小的切片，相当于将元数据分为相应个数的数据切片。具体地，可以通过调用对大文件切片的上层接口，对大文件切片的上层接口可以通过抽象接口类调用第二系统的对大文件切片的底层操作接口，将所述元数据对应的大文件切片。

其中，可以将大文件分成若干个数据量大小相同的切片，也可以随机切成若干个数据量大小不完全相同的切片，但是需要保证各切片的数据量小于第二阈值，若其中存在数据量不小于第二阈值的切片，则需要对该切片再次进行分切。而且，每次切片的数量可以根据需求设定，对此本申请实施例不加以限定。

子步骤 245，将对所述元数据的不同切片的切片信息生成同步任务，并分发给负载低的多个服务器的工作组件的接收器。

在将元数据对应的大文件切片后，则可以对该元数据的不同切片的切片信息，也即各切片中包含的数据，生成不同的同步任务，然后可以按照当前集群中各服务器的负载从低到高的顺序，分别将各同步任务分发给负载从低到高，对应同步任务个数的服务器的工作组件的接收器。例如，若生成 5 个同步任务，则按照当前集群中各服务器的负载从低到高的顺序，选出负载最低的 5 个服务器，然后将 5 个同步任务分别分发到 5 个服务器的工作组件的接收器，使每个服务器的工作组件的接收器都接收到一个同步任务，且任意两个服务器的工作组件的接收器接收到的同步任务不为同一个同步任务。

其中，可以随机将 5 个同步任务分发到 5 个服务器的工作组件的接收器，也可以按照一定顺序，例如按照同步任务的数据量从大到小，服务器的负载从低到高的对应关系，将 5 个同步任务分发到 5 个服务器的工作组件的接收器，对此本申请实施例不加以限定。

5 步骤 250，由工作组件的接收器根据接收到同步任务，执行获取实际数据的上层接口，所述获取实际数据的上层接口通过抽象类接口调用第一系统的获取实际数据的底层获取接口，从第一系统获取所述迁移任务的实际数据。

10 在本申请实施例中，同步任务中包含了待迁移的元数据第一系统中的文件信息，但是同步任务中并不包含待迁移的元数据，也即当前同步任务所需要迁移的实际数据。同步任务所需要迁移的实际数据仍存储于第一系统中。所以，需要从第一系统中获取同步任务所需要迁移的实际数据，具体地，可以由负责执行该同步任务的服务器的工作组件的接收器在收到同步任务后，执行获取实际数据的上层接口，获取实际数据的上层接口即可以
15 通过抽象类接口调用第一系统的获取实际数据的底层获取接口，从第一系统获取迁移任务的实际数据，也即同步任务所需要迁移的实际数据。

步骤 260，由工作组件的同步器执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

20 在获取了同步任务对用需要迁移的实际数据之后，即将实际数据同步至第二系统，具体地，可以由工作组件的同步器执行同步数据的各上层接口，同步数据的各上层接口即可以通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

在本申请另一个优选的实施例中，步骤 260，包括：

25 子步骤 261，由工作组件的同步器根据限速指令，调用与限速指令

相应的线程执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

5 为了调节数据迁移过程中的数据传输速度，在本申请实施例中，可以输入限速指令，其中包含需要可以接受的数据迁移过程中的数据传输速度，其可以为具体的限速数值，也可以为一个速度范围，进而可以由工作组件的同步器根据限速指令，调用与限速指令相应的线程执行同步数据的各上层接口，同步数据的各上层接口可以通过抽象接口类调用第二系统的底层操作接口，将所述实际数据按照限速数值的速度或者是以速度范围内的数据传输速速将实际数据同步至第二系统。其中，与限速指令相应的线程可以存在于前述的工作线程池中，也可以为另外创建的一个线程，对此本申请实施例不加以限定。

15 在本申请实施例中，可以接收将第一系统中的数据迁移至第二系统的迁移任务，所述第一系统和第二系统为多个系统中的任意两个系统，然后调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

20 从而实现了任意两个数据端可以进行双向的数据传输和迁移，包括数据端内部的 bucket 间迁移。其次，本申请所述的数据迁移技术具有良好的扩展功能，可以实现将任意的数据终端，包括尚未支持数据迁移的云存储产品，如百度对象存储、金山对象存储，与其它任意一个数据端进行数据的双向传输迁移。

25 另外，本申请可以支持分布式数据传输，可以极大提升数据迁移的效率。比以第一系统和第二系统为两个地区的 OSS 为例，从一个地区 OSS 的一个 bucket 的数据传输到另一个地区的 bucket 中，测试方法采用普通网

络路径，先从一个 bucket 下载出来，通过内网络链路，再上传存储到另外一个 bucket。其传输采用分布式集群环境：阿里云杭州 ECS(Elastic Compute Service,云服务器)，单台服务器配置为 4 核 CPU、4G 内存、500Mbps 内网带宽，总共 5 个节点。其测试示例图下表（一）：

数据大小	单机		分布式（5 节点）	
	时间	效率	时间	效率
10.5G	00:03:20	53MB/s	00:00:53	202MB/s
172G	01:01:46	47MB/s	00:12:37	232MB/s
995G	no test	no test	01:06:56	257MB/s

表（一）

其中，效率一列单位是 MB/s（兆比特每秒）。

可以看到，随着数据规模的增加，本申请的分布式传输的效率优势更加明显，跟集群的节点规模成正比例关系。

需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本申请实施例所必须的。

实施例三

参照图 3，示出了本申请的一种数据迁移装置实施例的结构框图，具

体可以包括如下模块:

任务接收模块 310, 适于接收将第一系统中的数据迁移至第二系统的迁移任务; 所述第一系统和第二系统为多个系统中的任意两个系统。

5 迁移模块 320, 适于调用对应所述迁移任务的各上层接口, 并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口, 以将第一系统的数据迁移至第二系统。

在本申请另一个优选地实施例中, 所述迁移模块 320, 包括:

上层接口确定子模块, 适于确定属于获取数据的各上层接口, 以及属于同步数据的各上层接口。

10 数据获取子模块, 适于对于属于获取数据的各上层接口, 通过抽象接口类调用第一系统的底层操作接口, 以从第一系统获取数据。

数据同步子模块, 适于对于属于同步数据的各上层接口, 通过抽象接口类调用第二系统的底层操作接口, 以将从第一系统获取的数据同步至第二系统。

15 在本申请另一个优选地实施例中, 所述数据获取子模块, 包括:

第一底层操作接口确定子模块, 适于对于属于获取数据的各上层接口中的第一上层接口, 以抽象接口类中与第一上层接口对应的第一抽象类接口作为底层操作接口以直接调用; 所述第一上层接口为各系统通用的接口。

20 第一底层操作接口调用子模块, 适于对于属于获取数据的各上层接口中的第二上层接口, 通过抽象接口类生成针对第一系统的第一接口类, 在所述第一接口类中调用与各第二上层接口对应的底层操作接口; 所述第二上层接口不为各系统通用的接口。

在本申请另一个优选地实施例中, 所述数据同步子模块, 包括:

25 第二底层操作接口确定子模块, 适于对于属于同步数据的各上

层接口中的第三上层接口，以抽象接口类中与第三上层接口对应的第二抽象类接口作为底层操作接口以直接调用；所述第三上层接口为各系统通用的接口。

第二底层操作接口调用子模块，适于对于属于同步数据的各上层接口中的第四上层接口，通过抽象接口类生成针对第二系统的第二接口类，在所述第二接口类中调用与各第四上层接口对应的底层操作接口；所述第四上层接口不为各系统通用的接口。

在本申请另一个优选地实施例中，在所述迁移模块 320 之后，还包括：迁移判断模块，适于判断将第一系统中的数据迁移至第二系统是否成功；如果失败，则重复执行迁移模块，直至重试次数达到重试阈值或者在重试阈值内成功。

在本申请另一个优选地实施例中，在所述迁移模块 320 之后，还包括：暂停指令判断模块，适于判断是否接收到暂停指令；如果接收到暂停指令，则进入暂停记录子模块。

暂停记录模块，适于暂停将第一系统的数据迁移至第二系统，并记录暂停时刻第一系统中已迁移的数据位置。

则进一步的，在所述暂停指令判断模块之后，还包括：

重启指令判断模块，适于判断是否接收到重启指令；如果接收到重启指令，则根据暂停时刻第一系统中已迁移的数据位置开始，重新进入迁移模块。

在本申请另一个优选地实施例中，当第一系统为客户端文件系统时，在任务接收模块 310 之前，还包括：

应用程序安装模块 300，适于在客户端中安装应用程序，所述应用程序针对指定文件模拟各底层操作接口。

在本申请另一个优选地实施例中，在所述迁移模块 320 之后，还包括：

数据监控模块，适于监控客户端文件系统是否与最近一次迁移到第二系统的数据出现变更；如果出现变更，则进入变更数据同步模块。

变更数据同步模块，适于调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以

5 将客户端文件系统的变更数据同步至第二系统。

在本申请另一个优选地实施例中，在所述数据获取子模块之后，还包括：

签名子模块，适于调用签名的上层接口，并由所述签名的上层接口调用第一系统的签名的底层操作接口，对所述数据进行签名。

10 进一步的，在所述数据同步子模块之后，还包括：

签名验证子模块，适于调用签名认证的上层接口，并由所签名认证的上层接口调用第二系统的签名认证的底层操作接口，对所述数据进行签名验证。

在本申请另一个优选地实施例中，在所述迁移模块 320 之后，还包括：

15 第一文件列表获取模块，适于针对第一系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第一系统的底层操作接口，获取第一系统的文件列表。

第二文件列表获取模块，适于针对第二系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第二系统的底

20 层操作接口，获取第二系统的文件列表。

差异提示模块，适于比较第一系统的文件列表和第二系统的文件列表，并将两者之间的差异提示给客户端。

在本申请另一个优选地实施例中，在所述上层接口确定子模块之后，还包括：

25 第一身份验证模块，适于针对第一系统调用身份验证的上层接口，

并由所述身份验证的上层接口通过抽象接口类调用第一系统的底层操作接口，对用户在第一系统的身份信息验证。

第二身份验证模块，适于针对第二系统调用身份验证的上层接口，并由所述身份验证的上层接口通过抽象接口类调用第二系统的底层操作接口，对用户在第二系统的身份信息验证。

本申请实施例可以接收将第一系统中的数据迁移至第二系统的迁移任务；所述第一系统和第二系统为多个系统中的任意两个系统，然后调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

首先，实现了任意两个数据端可以进行双向的数据传输和迁移，包括数据端内部的 bucket 间迁移。

其次，本申请所述的数据迁移技术具有良好的扩展功能，可以实现将任意的数据终端，包括尚未支持数据迁移的云存储产品，如百度对象存储、金山对象存储，与其它任意一个数据端进行数据的双向传输迁移。

实施例四

参照图 4，示出了本申请的一种数据迁移装置实施例的结构框图，具体可以包括如下模块：

任务接收模块 410，适于接收将第一系统中的数据迁移至第二系统的迁移任务；所述第一系统和第二系统为多个系统中的任意两个系统。

迁移模块 420，适于调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。具体包括：

上层接口确定子模块 421，适于确定属于获取数据的各上层接口，以及属于同步数据的各上层接口。

数据获取子模块 422，适于对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据。具体包括：

元数据获取子模块 4221，适于由管理组件的扫描器执行获取元数据的上层接口，所述获取元数据的上层接口通过抽象类接口调用第一系统的获取元数据的底层操作接口，从第一系统获取所述迁移任务对应的各元数据。

任务分发子模块 4222，适于由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器。

10 在本申请另一个优选地实施例中，所述任务分发子模块 4222，包括：

数据量大小提取子模块，适于提取元数据中记录的数据量大小。

第一元数据分发子模块，适于如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器。

第二元数据分发子模块，适于如果元数据记录的数据量不小于第一阈值，则将所述元数据作为同步任务分发给负载最小的服务器中的工作组件的接收器。

在本申请另一个优选地实施例中，所述任务分发子模块 4222，包括：

20 数据量大小提取子模块，适于提取元数据中记录的数据量大小。

第一元数据分发子模块，适于如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器。

25 第三元数据分发子模块，适于如果元数据中记录的数据量不

小于第一阈值但是小于第二阈值时，则将所述元数据生成同步任务分发给负载最小的服务器中的工作组件的接收器。

元数据切分子模块，适于如果元数据中记录的数据量不小于第二阈值时，则调用对大文件切片的上层接口，所述对大文件切片的上层接口通过抽象接口类调用第二系统的对大文件切片的底层操作接口，将所述元数据对应的大文件切片。

将对所述元数据的不同切片的切片信息生成同步任务，并分发给负载低的多个服务器的工作组件的接收器。

实际数据获取子模块 4223，适于由工作组件的接收器根据接收到同步任务，执行获取实际数据的上层接口，所述获取实际数据的上层接口通过抽象类接口调用第一系统的获取实际数据的底层获取接口，从第一系统获取所述迁移任务的实际数据。

数据同步子模块 423，适于对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统。具体包括：

实际数据同步子模块 4231，适于由工作组件的同步器执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

在本申请另一个优选地实施例中，所述实际数据同步子模块 4231，包括：

第一实际数据同步子模块，适于由工作组件的同步器根据限速指令，调用与限速指令相应的线程执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

在本申请实施例中，可以接收将第一系统中的数据迁移至第二系统的

迁移任务，所述第一系统和第二系统为多个系统中的任意两个系统，然后调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

5 从而实现了任意两个数据端可以进行双向的数据传输和迁移，包括数据端内部的 bucket 间迁移。其次，本申请所述的数据迁移技术具有良好的扩展功能，可以实现将任意的数据终端，包括尚未支持数据迁移的云存储产品，如百度对象存储、金山对象存储，与其它任意一个数据端进行数据的双向传输迁移。

10 另外，本申请同时支持单机数据传输以及分布式数据传输，可以极大提升数据迁移的效率。

对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说
15 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的部分互相参见即可。

本领域内的技术人员应明白，本申请实施例的实施例可提供为方法、装置、或计算机程序产品。因此，本申请实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请
20 实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用同步介质(包括但不限于磁盘同步器、CD-ROM、光学同步器等)上实施的计算机程序产品的形式。

在一个典型的配置中，所述计算机设备包括一个或多个处理器 (CPU)、输入/输出接口、网络接口和内存。内存可能包括计算机可读介质中的非永久性同步器，随机存取同步器(RAM) 和/或非易失性内存等形式，如只读同步器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息同步。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的同步介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取同步器 (SRAM)、动态随机存取同步器 (DRAM)、其他类型的随机存取同步器 (RAM)、只读同步器 (ROM)、电可擦除可编程只读同步器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读同步器(CD-ROM)、数字多功能光盘(DVD)或其他光学同步、磁盒式磁带，磁带磁磁盘同步或其他磁性同步设备或任何其他非传输介质，可用于同步可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非持续性的电脑可读媒体 (transitory media)，如调制的数据信号和载波。

本申请实施例是参照根据本申请实施例的方法、终端设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/

或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个服务器，使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或

5 方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可同步在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读同步器中，使得同步在该计算机可读同步器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

10 这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上，使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

15 尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。

最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术

语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些
5 要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

以上对本申请所提供的一种数据迁移方法和一种数据迁移装置，进行
10 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；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

权 利 要 求 书

1、一种数据迁移方法，其特征在于，包括：

接收将第一系统中的数据迁移至第二系统的迁移任务；

调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口

5 类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

2、根据权利要求 1 所述的方法，其特征在于，所述调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤，包
10 括：

确定属于获取数据的各上层接口，以及属于同步数据的各上层接口；

对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据；

对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统。
15

3、根据权利要求 2 所述的方法，其特征在于，对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口的步骤，包括：

对于属于获取数据的各上层接口中的第一上层接口，以抽象接口类中与第一上层接口对应的第一抽象类接口作为底层操作接口以直接调用；所述
20 述第一上层接口为各系统通用的接口；

对于属于获取数据的各上层接口中的第二上层接口，通过抽象接口类生成针对第一系统的第一接口类，在所述第一接口类中调用与各第二上层接口对应的底层操作接口；所述第二上层接口不为各系统通用的接口。

4、根据权利要求 2 或 3 所述的方法，其特征在于，对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口的步骤，
25

包括:

对于属于同步数据的各上层接口中的第三上层接口,以抽象接口类中与第三上层接口对应的第二抽象类接口作为底层操作接口以直接调用;所述第三上层接口为各系统通用的接口;

- 5 对于属于同步数据的各上层接口中的第四上层接口,通过抽象接口类生成针对第二系统的第二接口类,在所述第二接口类中调用与各第四上层接口对应的底层操作接口;所述第四上层接口不为各系统通用的接口。

- 5、根据权利要求 2 所述的方法,其特征在于,所述对于属于获取数据的各上层接口,通过抽象接口类调用第一系统的底层操作接口,以从第一系统获取数据的步骤,包括:

由管理组件的扫描器执行获取元数据的上层接口,所述获取元数据的上层接口通过抽象类接口调用第一系统的获取元数据的底层操作接口,从第一系统获取所述迁移任务对应的各元数据;

- 15 由管理组件的分发器,基于各元数据生成同步任务并分发给各工作组件的接收器;

由工作组件的接收器根据接收到同步任务,执行获取实际数据的上层接口,所述获取实际数据的上层接口通过抽象类接口调用第一系统的获取实际数据的底层获取接口,从第一系统获取所述迁移任务的实际数据。

- 20 6、根据权利要求 5 所述的方法,其特征在于,所述对于属于同步数据的各上层接口,通过抽象接口类调用第二系统的底层操作接口,以将从第一系统获取的数据同步至第二系统的步骤,包括:

由工作组件的同步器执行同步数据的各上层接口,所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口,将所述实际数据同步至第二系统。

- 25 7、根据权利要求 5 或 6 所述的方法,其特征在于,所述管理组件在

集群的一台服务器中，所述工作组件在集群的各台服务器中。

8、根据权利要求 7 所述的方法，其特征在于，所述由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器的步骤，包括：

5 提取元数据中记录的数据量大小；

如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

如果元数据记录的数据量不小于第一阈值，则将所述元数据作为同步任务分发给负载最小的服务器中的工作组件的接收器；

10 9、根据权利要求 7 所述的方法，其特征在于，所述由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器的步骤，包括：

提取元数据中记录的数据量大小；

15 如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

如果元数据中记录的数据量不小于第一阈值但是小于第二阈值时，则将所述元数据生成同步任务分发给负载最小的服务器中的工作组件的接收器；

20 如果元数据中记录的数据量不小于第二阈值时，则调用对大文件切片的上层接口，所述对大文件切片的上层接口通过抽象接口类调用第二系统的对大文件切片的底层操作接口，将所述元数据对应的大文件切片；

将对所述元数据的不同切片的切片信息生成同步任务，并分发给负载低的多个服务器的工作组件的接收器。

25 10、根据权利要求 8 所述的方法，其特征在于，所述集群的各台服务器中之间采用 JGroups 通信协议传输数据。

11、根据权利要求 6 所述的方法，其特征在于，所述由工作组件的同步器执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统的步骤，包括：

- 5 由工作组件的同步器根据限速指令，调用与限速指令相应的线程执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

12、根据权利要求 1 或 2 所述的方法，其特征在于，调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤之后，还包括：

判断将第一系统中的数据迁移至第二系统是否成功；

15 如果失败，则重复执行调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤，直至重试次数达到重试阈值或者在重试阈值内成功。

13、根据权利要求 1 或 2 所述的方法，其特征在于，在调用对应所述迁移任务的各上层接口，通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的过程中，还包括：

20 判断是否接收到暂停指令；

如果接收到暂停指令，则暂停将第一系统的数据迁移至第二系统，并记录暂停时刻第一系统中已迁移的数据位置；

则进一步的，在暂停将第一系统的数据迁移至第二系统之后，还包括：

判断是否接收到重启指令；

25 如果接收到重启指令，则根据暂停时刻第一系统中已迁移的数据位置

开始，重新调用对应所述迁移任务的各上层接口，通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤。

14、根据权利要求 1 或 2 所述的方法，其特征在于，当第一系统为客户端文件系统时，在接收将第一系统中的数据迁移至第二系统的迁移任务的步骤之前，还包括：

在客户端中安装应用程序，所述应用程序针对指定文件模拟各底层操作接口。

15、根据权利要求 14 所述的方法，其特征在于，所述调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤之后，还包括：

监控客户端文件系统是否与最近一次迁移到第二系统的数据出现变更；

15 如果出现变更，则调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将客户端文件系统的变更数据同步至第二系统。

16、根据权利要求 2 所述的方法，其特征在于，所述对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据的步骤之后，还包括：

调用签名的上层接口，并由所述签名的上层接口调用第一系统的签名的底层操作接口，对所述数据进行签名；

进一步的，对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统的步骤之后，还包括：

调用签名认证的上层接口，并由所签名认证的上层接口调用第二系统的签名认证的底层操作接口，对所述数据进行签名验证。

17、根据权利要求 1 所述的方法，其特征在于，在调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统的步骤之后，还包括：

针对第一系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第一系统的底层操作接口，获取第一系统的文件列表；

10 针对第二系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第二系统的底层操作接口，获取第二系统的文件列表；

比较第一系统的文件列表和第二系统的文件列表，并将两者之间的差异提示给客户端。

15 18、根据权利要求 2 所述的方法，其特征在于，所述迁移任务由客户端发送，所述迁移任务包括第一系统的标识信息、用户在第一系统的身份信息、待迁移数据在第一系统中的文件信息、第二系统的标识信息、用户在第二系统的身份信息。

20 19、根据权利要求 18 所述的方法，其特征在于，在确定属于获取数据的各上层接口，以及属于同步数据的各上层接口的步骤之后，还包括：

针对第一系统调用身份验证的上层接口，并由所述身份验证的上层接口通过抽象接口类调用第一系统的底层操作接口，对用户在第一系统的身份信息进行验证；

25 针对第二系统调用身份验证的上层接口，并由所述身份验证的上层接口通过抽象接口类调用第二系统的底层操作接口，对用户第二系统的身份

份信息进行验证。

20、一种数据迁移装置，其特征在于，包括：

任务接收模块，适于接收将第一系统中的数据迁移至第二系统的迁移任务；

5 迁移模块，适于调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将第一系统的数据迁移至第二系统。

21、根据权利要求 20 所述的装置，其特征在于，所述迁移模块，包括：

10 上层接口确定子模块，适于确定属于获取数据的各上层接口，以及属于同步数据的各上层接口；

数据获取子模块，适于对于属于获取数据的各上层接口，通过抽象接口类调用第一系统的底层操作接口，以从第一系统获取数据；

15 数据同步子模块，适于对于属于同步数据的各上层接口，通过抽象接口类调用第二系统的底层操作接口，以将从第一系统获取的数据同步至第二系统。

22、根据权利要求 21 所述的装置，其特征在于，所述数据获取子模块，包括：

20 第一底层操作接口确定子模块，适于对于属于获取数据的各上层接口中的第一上层接口，以抽象接口类中与第一上层接口对应的第一抽象类接口作为底层操作接口以直接调用；所述第一上层接口为各系统通用的接口；

25 第一底层操作接口调用子模块，适于对于属于获取数据的各上层接口中的第二上层接口，通过抽象接口类生成针对第一系统的第一接口类，在所述第一接口类中调用与各第二上层接口对应的底层操作接口；所述第二

上层接口不为各系统通用的接口。

23、根据权利要求 21 或 22 所述的装置，其特征在于，所述数据同步子模块，包括：

第二底层操作接口确定子模块，适于对于属于同步数据的各上层接口
5 中的第三上层接口，以抽象接口类中与第三上层接口对应的第二抽象类接口作为底层操作接口以直接调用；所述第三上层接口为各系统通用的接口；

第二底层操作接口调用子模块，适于对于属于同步数据的各上层接口
10 中的第四上层接口，通过抽象接口类生成针对第二系统的第二接口类，在所述第二接口类中调用与各第四上层接口对应的底层操作接口；所述第四上层接口不为各系统通用的接口。

24、根据权利要求 21 所述的装置，其特征在于，所述数据获取子模块，包括：

元数据获取子模块，适于由管理组件的扫描器执行获取元数据的上层
15 接口，所述获取元数据的上层接口通过抽象类接口调用第一系统的获取元数据的底层操作接口，从第一系统获取所述迁移任务对应的各元数据；

任务分发子模块，适于由管理组件的分发器，基于各元数据生成同步任务并分发给各工作组件的接收器；

实际数据获取子模块，适于由工作组件的接收器根据接收到同步任
20 务，执行获取实际数据的上层接口，所述获取实际数据的上层接口通过抽象类接口调用第一系统的获取实际数据的底层获取接口，从第一系统获取所述迁移任务的实际数据。

25、根据权利要求 24 所述的装置，其特征在于，所述数据同步子模块，包括：

25 实际数据同步子模块，适于由工作组件的同步器执行同步数据的各上

层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

26、根据权利要求 24 或 25 所述的装置，其特征在于，所述管理组件在集群的一台服务器中，所述工作组件在集群的各台服务器中。

5 27、根据权利要求 26 所述的装置，其特征在于，所述任务分发子模块，包括：

数据量大小提取子模块，适于提取元数据中记录的数据量大小；

第一元数据分发子模块，适于如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

第二元数据分发子模块，适于如果元数据记录的数据量不小于第一阈值，则将所述元数据作为同步任务分发给负载最小的服务器中的工作组件的接收器。

15 28、根据权利要求 26 所述的装置，其特征在于，所述任务分发子模块，包括：

数据量大小提取子模块，适于提取元数据中记录的数据量大小；

第一元数据分发子模块，适于如果元数据中记录的数据量小于第一阈值，直接将所述元数据作为同步任务分发给管理组件所在的服务器中的工作组件的接收器；

20 第三元数据分发子模块，适于如果元数据中记录的数据量不小于第一阈值但是小于第二阈值时，则将所述元数据生成同步任务分发给负载最小的服务器中的工作组件的接收器；

元数据切分子模块，适于如果元数据中记录的数据量不小于第二阈值时，则调用对大文件切片的上层接口，所述对大文件切片的上层接口通过抽象接口类调用第二系统的对大文件切片的底层操作接口，将所述元数据

25

对应的大文件切片；

将对所述元数据的不同切片的切片信息生成同步任务，并分发给负载低的多个服务器的工作组件的接收器。

29、根据权利要求 27 所述的装置，其特征在于，所述集群的各台服务器中之间采用 JGroups 通信协议传输数据。

30、根据权利要求 25 所述的装置，其特征在于，所述实际数据同步子模块，包括：

第一实际数据同步子模块，适于由工作组件的同步器根据限速指令，调用与限速指令相应的线程执行同步数据的各上层接口，所述同步数据的各上层接口通过抽象接口类调用第二系统的底层操作接口，将所述实际数据同步至第二系统。

31、根据权利要求 20 或 21 所述的装置，其特征在于，在所述迁移模块之后，还包括：

迁移判断模块，适于判断将第一系统中的数据迁移至第二系统是否成功；如果失败，则重复执行迁移模块，直至重试次数达到重试阈值或者在重试阈值内成功。

32、根据权利要求 20 或 21 所述的装置，其特征在于，在所述迁移模块之后，还包括：

暂停指令判断模块，适于判断是否接收到暂停指令；如果接收到暂停指令，则进入暂停记录子模块；

暂停记录模块，适于暂停将第一系统的数据迁移至第二系统，并记录暂停时刻第一系统中已迁移的数据位置；

则进一步的，在所述暂停指令判断模块之后，还包括：

重启指令判断模块，适于判断是否接收到重启指令；如果接收到重启指令，则根据暂停时刻第一系统中已迁移的数据位置开始，重新进入迁移

模块。

33、根据权利要求 20 或 21 所述的装置，其特征在于，当第一系统为客户端文件系统时，在任务接收模块之前，还包括：

5 应用程序安装模块，适于在客户端中安装应用程序，所述应用程序针对指定文件模拟各底层操作接口。

34、根据权利要求 33 所述的装置，其特征在于，在所述迁移模块之后，还包括：

数据监控模块，适于监控客户端文件系统是否与最近一次迁移到第二系统的数据出现变更；如果出现变更，则进入变更数据同步模块；

10 变更数据同步模块，适于调用对应所述迁移任务的各上层接口，并由各上层接口通过抽象接口类调用第一系统和第二系统的底层操作接口，以将客户端文件系统的变更数据同步至第二系统。

35、根据权利要求 21 所述的装置，其特征在于，在所述数据获取子模块之后，还包括：

15 签名子模块，适于调用签名的上层接口，并由所述签名的上层接口调用第一系统的签名的底层操作接口，对所述数据进行签名；

进一步的，在所述数据同步子模块之后，还包括：

20 签名验证子模块，适于调用签名认证的上层接口，并由所签名认证的上层接口调用第二系统的签名认证的底层操作接口，对所述数据进行签名验证。

36、根据权利要求 20 所述的装置，其特征在于，在所述迁移模块之后，还包括：

25 第一文件列表获取模块，适于针对第一系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第一系统的底层操作接口，获取第一系统的文件列表；

第二文件列表获取模块，适于针对第二系统调用文件列表的上层接口，并由所述调用文件列表的上层接口通过抽象接口类调用第二系统的底层操作接口，获取第二系统的文件列表；

5 差异提示模块，适于比较第一系统的文件列表和第二系统的文件列表，并将两者之间的差异提示给客户端。

37、根据权利要求 21 所述的装置，其特征在于，所述迁移任务由客户端发送，所述迁移任务包括第一系统的标识信息、用户在第一系统的身份信息、待迁移数据在第一系统中的文件信息、第二系统的标识信息、用户在第二系统的身份信息。

10 38、根据权利要求 37 所述的装置，其特征在于，在所述上层接口确定子模块之后，还包括：

第一身份验证模块，适于针对第一系统调用身份验证的上层接口，并由所述身份验证的上层接口通过抽象接口类调用第一系统的底层操作接口，对用户在第一系统的身份信息进行验证；

15 第二身份验证模块，适于针对第二系统调用身份验证的上层接口，并由所述身份验证的上层接口通过抽象接口类调用第二系统的底层操作接口，对用户第二系统的身份信息进行验证。

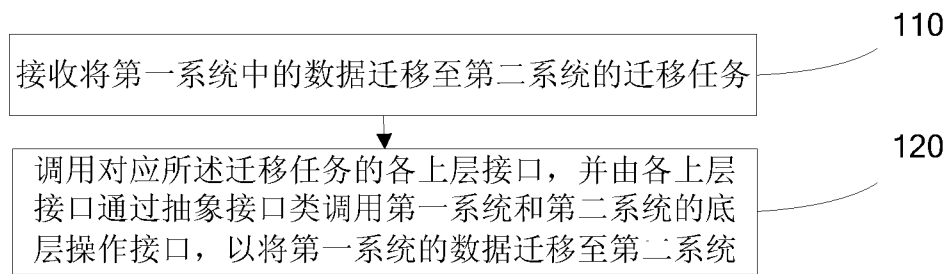


图 1

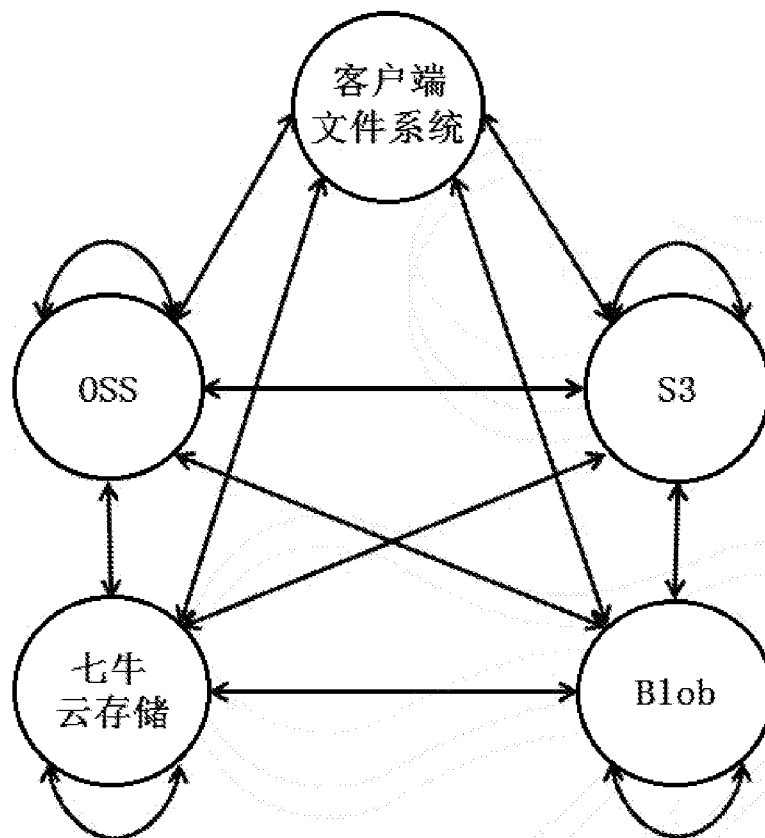


图 1A

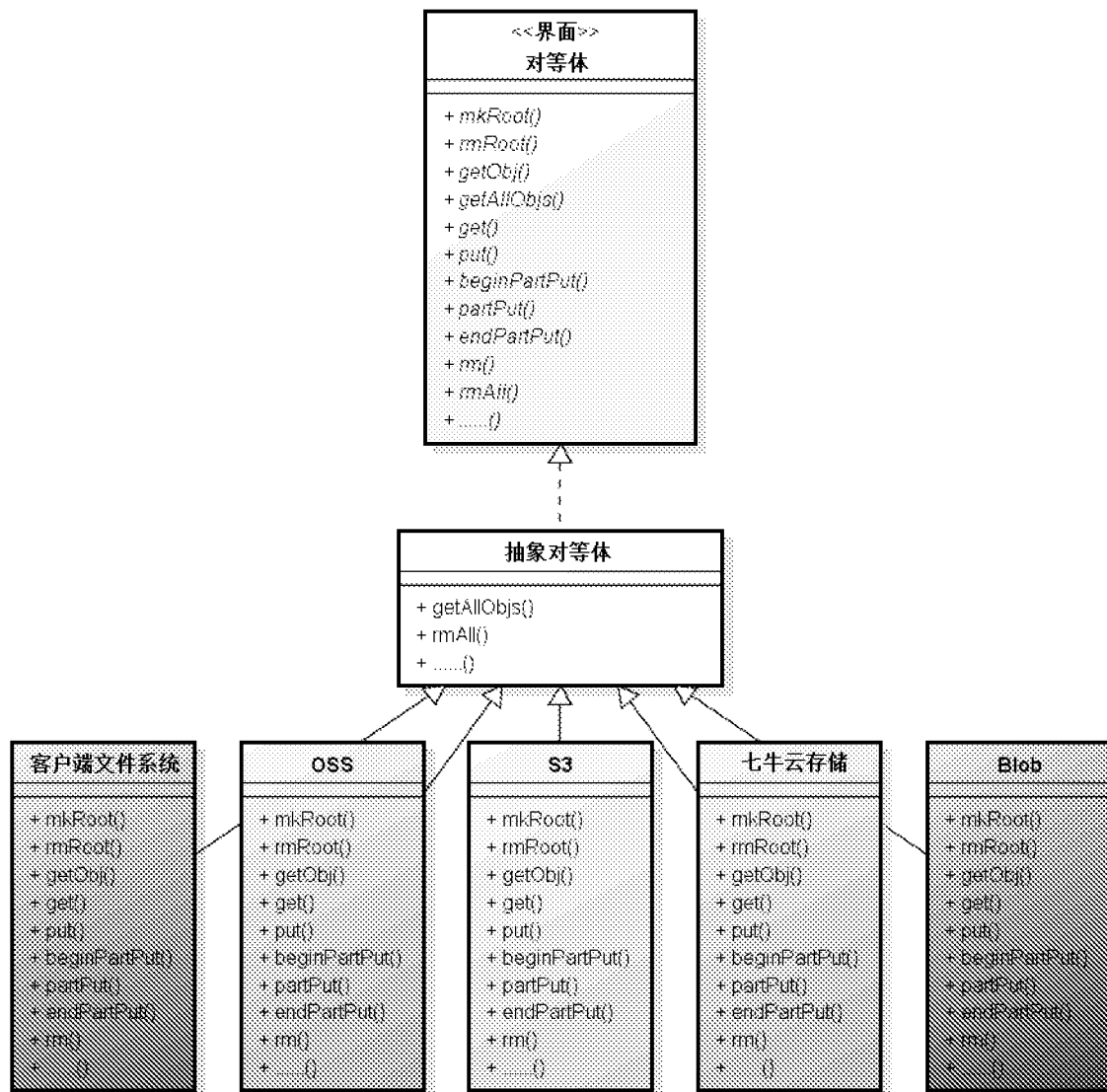


图 1B

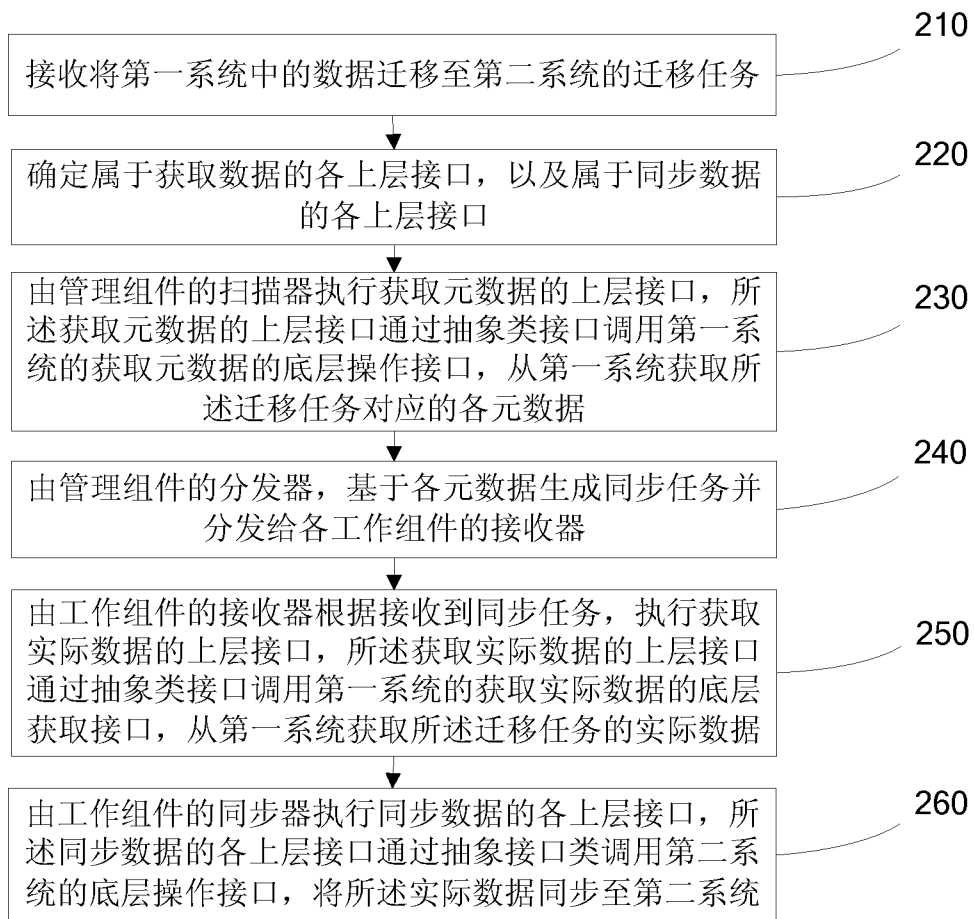


图 2

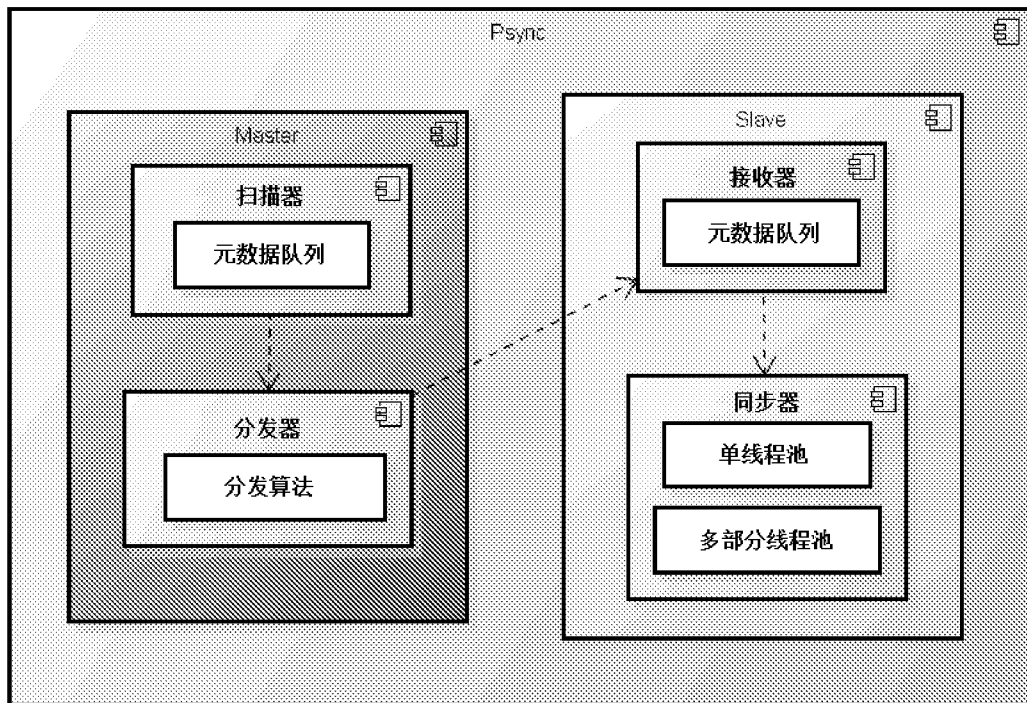


图 2C

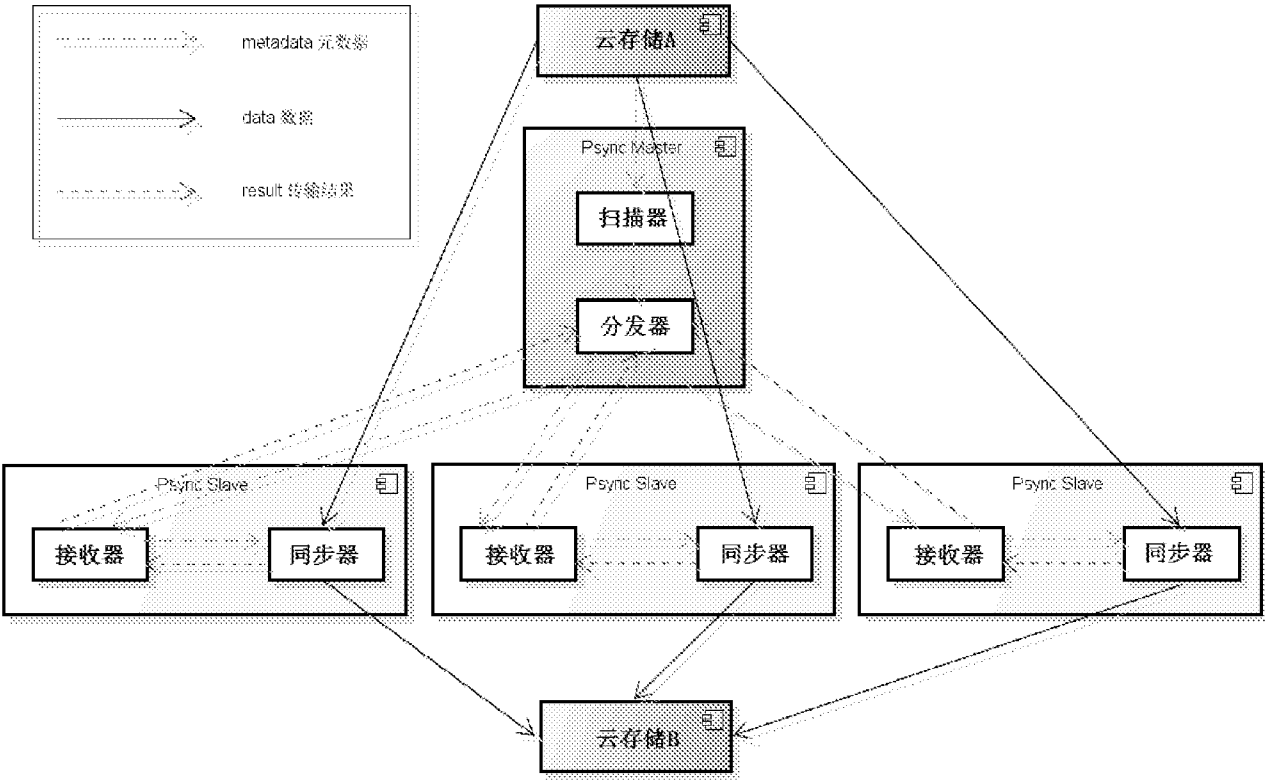


图 2D

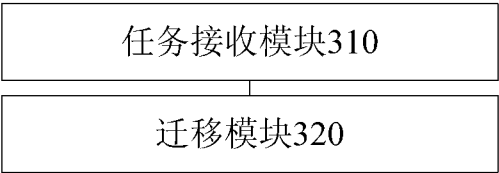


图 3

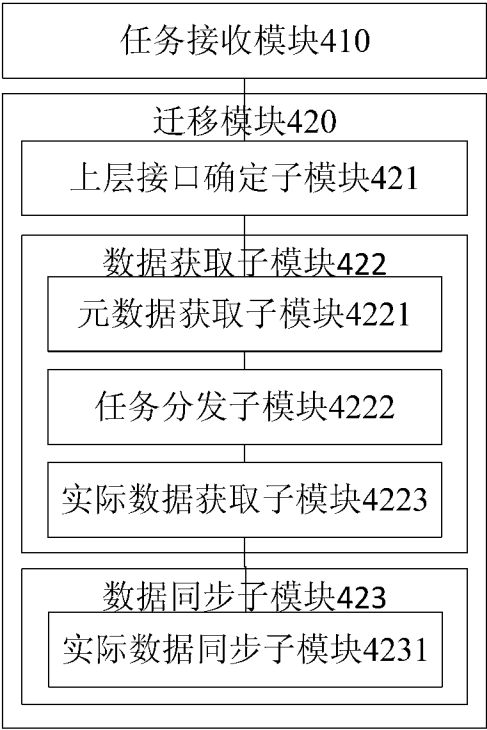


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/077722

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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)

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)

CNPAT, EPODOC, WPI, GOOGLE, CNKI: data, migration, function, dispatch, synchronization, metadata, interface

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101079902 A (TSINGHUA UNIVERSITY) 28 November 2007 (28.11.2007) claim 1	1, 2, 5-8, 10-21, 24-27, 29-38
Y	CN 103605728 A (INSPUR (BEIJING) ELECTRONIC INFORMATION INDUSTRY CO., LTD.) 26 February 2014 (26.02.2014) description, paragraphs [0071]-[0075]	1, 2, 5-8, 10-21, 24-27, 29-38
Y	CN 104424283 A (ALIBABA GROUP HOLDING LTD.) 18 March 2015 (18.03.2015) description, paragraphs [0086] and [0103]-[0107]	13, 15, 17, 32, 34, 36
A	CN 103856724 A (CHENGDU SOBEY DIGITAL TECHNOLOGY CO., LTD.) 11 June 2014 (11.06.2014) the whole document	1-38
A	US 2013080488 A1 (ALIBABA GROUP HOLDING LIMITED) 28 March 2013 (28.03.2013) the whole document	1-38

☐ 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	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“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
“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	“&” document member of the same patent family

Date of the actual completion of the international search 07 June 2017	Date of mailing of the international search report 22 June 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer TANG, Na Telephone No. (86-10) 62413997

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/077722

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101079902 A	28 November 2007	CN 100571281 C	16 December 2009
CN 103605728 A	26 February 2014	CN 103605728 B	12 October 2016
CN 104424283 A	18 March 2015	None	
CN 103856724 A	11 June 2014	CN 103856724 B	29 March 2017
US 2013080488 A1	28 March 2013	EP 2758880 A1	30 July 2014
		TW 201314470 A	01 April 2013
		CN 103019614 B	25 November 2015
		TW I521359 B	11 February 2016
		CN 103019614 A	03 April 2013
		WO 2013043514 A1	28 March 2013
		EP 2758880 B1	11 January 2017
		US 9053021 B2	09 June 2015
		JP 2014531668 A	27 November 2014
		HK 1179366 A1	19 August 2016
		HK 1179366 A0	27 September 2013

国际检索报告

国际申请号

PCT/CN2017/077722

A. 主题的分类

G06F 17/30(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)

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))

CNPAT, EPDOC, WPI, GOOGLE, CNKI: 数据, 迁移, 函数, 调度, 同步, 元数据, 接口, data, migration, function, dispatch, synchronization, metadata, interface

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
Y	CN 101079902 A (清华大学) 2007年 11月 28日 (2007 - 11 - 28) 权利要求1	1, 2, 5-8, 10-21, 24-27, 29-38
Y	CN 103605728 A (浪潮北京电子信息产业有限公司) 2014年 2月 26日 (2014 - 02 - 26) 说明书第[0071]段至第[0075]段	1, 2, 5-8, 10-21, 24-27, 29-38
Y	CN 104424283 A (阿里巴巴集团控股有限公司) 2015年 3月 18日 (2015 - 03 - 18) 说明书第[0086]段、第[0103]段至第[0107]段	13, 15, 17, 32, 34, 36
A	CN 103856724 A (成都索贝数码科技股份有限公司) 2014年 6月 11日 (2014 - 06 - 11) 全文	1-38
A	US 2013080488 A1 (ALIBABA GROUP HOLDING LIMITED) 2013年 3月 28日 (2013 - 03 - 28) 全文	1-38

☐ 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

“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

2017年 6月 7日

国际检索报告邮寄日期

2017年 6月 22日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

唐娜

传真号 (86-10)62019451

电话号码 (86-10)62413997

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/077722

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	101079902	A	2007年 11月 28日	CN	100571281	C	2009年 12月 16日
CN	103605728	A	2014年 2月 26日	CN	103605728	B	2016年 10月 12日
CN	104424283	A	2015年 3月 18日	无			
CN	103856724	A	2014年 6月 11日	CN	103856724	B	2017年 3月 29日
US	2013080488	A1	2013年 3月 28日	EP	2758880	A1	2014年 7月 30日
				TW	201314470	A	2013年 4月 1日
				CN	103019614	B	2015年 11月 25日
				TW	I521359	B	2016年 2月 11日
				CN	103019614	A	2013年 4月 3日
				WO	2013043514	A1	2013年 3月 28日
				EP	2758880	B1	2017年 1月 11日
				US	9053021	B2	2015年 6月 9日
				JP	2014531668	A	2014年 11月 27日
				HK	1179366	A1	2016年 8月 19日
				HK	1179366	A0	2013年 9月 27日

表 PCT/ISA/210 (同族专利附件) (2009年7月)