



(12) 发明专利申请

(10) 申请公布号 CN 116303257 A

(43) 申请公布日 2023. 06. 23

(21) 申请号 202310193702.5

(22) 申请日 2023.02.23

(71) 申请人 度小满科技(北京)有限公司

地址 100085 北京市海淀区西北旺路10号
院西区4号楼6层606

(72) 发明人 李建鹏 潘丹 赵辉

(74) 专利代理机构 北京知帆远景知识产权代理
有限公司 11890

专利代理师 刘岩磊

(51) Int. Cl.

G06F 16/13 (2019.01)

G06F 16/182 (2019.01)

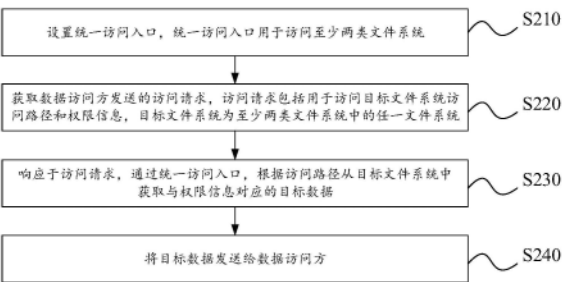
权利要求书2页 说明书11页 附图5页

(54) 发明名称

数据处理方法、装置、设备、介质及程序产品

(57) 摘要

本申请提供了一种数据处理方法、装置、设备、介质及程序产品,包括:设置统一访问入口,统一访问入口用于访问至少两类文件系统;获取数据访问方发送的访问请求,访问请求包括用于访问目标文件系统的访问路径和权限信息,目标文件系统为至少两类文件系统中的任意一文件系统;响应于访问请求,通过统一访问入口,根据访问路径从目标文件系统中获取与权限信息对应的目标数据;将目标数据发送给数据访问方。以提高数据访问效率,提高数据访问灵活性。



1. 一种数据处理方法,其特征在于,包括:

设置统一访问入口,所述统一访问入口用于访问至少两类文件系统;

获取数据访问方发送的访问请求,所述访问请求包括用于访问目标文件系统的访问路径和权限信息,所述目标文件系统为所述至少两类文件系统中的任意一个文件系统;

响应于所述访问请求,通过所述统一访问入口,根据所述访问路径从所述目标文件系统中获取与所述权限信息对应的目标数据;

将所述目标数据发送给所述数据访问方。

2. 根据权利要求1所述的方法,其特征在于,所述访问路径包括所述目标文件系统的集群名称,所述根据所述访问路径从所述目标文件系统中获取与所述权限信息对应的目标数据之前,还包括:

确定所述至少两类文件系统各自对应的集群名称;

根据所述至少两类文件系统各自对应的集群名称,设置所述至少两类文件系统各自对应的子访问接口;

所述根据所述访问路径从所述目标文件系统中获取与所述权限信息对应的目标数据,包括:

根据所述目标文件系统的集群名称,确定所述目标文件系统对应的子访问接口;

根据所述目标文件系统对应的子访问接口,从所述目标文件系统中获取与所述权限信息对应的目标数据。

3. 根据权利要求1或者2所述的方法,其特征在于,所述获取数据访问方发送的访问请求之前,还包括:

对所述数据发送方进行权限认证;

在对所述数据发送方进行权限认证通过时,确定所述数据发送方的权限信息,所述权限信息包括:对所述统一访问入口的使用权限、对所述至少两类文件系统中每个文件系统的访问范围;

将所述权限信息发送给所述数据发送方,以使所述数据发送方根据所述权限信息发送所述访问请求。

4. 根据权利要求3所述的方法,其特征在于,所述确定所述数据发送方的权限信息,包括:

确定所述数据发送方对应的目标组;

根据所述目标组确定所述目标组对应的权限信息;

将所述目标组对应的权限信息确定为所述数据发送方的权限信息。

5. 根据权利要求3所述的方法,其特征在于,所述对所述数据发送方进行权限认证,包括:

确定所述数据发送方的登录方式,所述登录方式包括以下任一项:基于命令行接口的登录方式、基于编程语言接口的登录方式;

根据所述登录方式确定所述数据发送方的登录信息;

根据所述登录信息对所述数据发送方进行权限认证。

6. 根据权利要求1或者2所述的方法,其特征在于,还包括:

根据所述至少两类文件系统各自的存储路径格式,确定通用访问路径格式;

根据所述通用访问路径格式确定所述至少两类文件系统各自对应的访问路径。

7. 一种数据处理装置, 其特征在于, 包括:

第一设置模块, 用于设置统一访问入口, 所述统一访问入口用于访问至少两类文件系统;

第一获取模块, 用于获取数据访问方发送的访问请求, 所述访问请求包括用于访问目标文件系统的访问路径和权限信息, 所述目标文件系统为所述至少两类文件系统中的任意一个文件系统;

第二获取模块, 用于响应于所述访问请求, 通过所述统一访问入口, 根据所述访问路径从所述目标文件系统中获取与所述权限信息对应的目标数据;

第一发送模块, 用于将所述目标数据发送给所述数据访问方。

8. 一种电子设备, 其特征在于, 包括:

处理器和存储器, 所述存储器用于存储计算机程序, 所述处理器用于调用并运行所述存储器中存储的计算机程序, 以执行权利要求1至6中任一项所述的方法。

9. 一种计算机可读存储介质, 其特征在于, 用于存储计算机程序, 所述计算机程序使得计算机执行权利要求1至6中任一项所述的方法。

10. 一种计算机程序产品, 其特征在于, 该计算机程序产品包括计算机指令, 使得电子设备执行权利要求1-6中任一项所述的方法。

数据处理方法、装置、设备、介质及程序产品

技术领域

[0001] 本申请实施例涉及数据处理技术领域,尤其涉及一种数据处理方法、装置、设备、介质及程序产品。

背景技术

[0002] 常用的存储方式包括:文件存储和对象存储。目前业界主流的文件存储为分布式文件系统(Hadoop Distributed File System,HDFS),业界主流的对象存储为百度智能云对象存储(Baidu Object Storage,BOS)、简单存储服务(Simple Storage Service,S3)。其中,由于HDFS为了提高数据吞吐量,而牺牲了获取数据的速度,因此HDFS不适用于低延迟访问数据的业务需求,且由于HDFS常将元数据存储在内存中,导致在需要存储许多小文件时,会占用较多的内存空间,闲置大量的硬盘空间,所以HDFS也不适用于大量的小文件存储。另外,对象存储不支持组权限管控。因此,一般会采用多种文件系统对不同存储需求的数据分别进行存储。

[0003] 然而,当需要获取多种文件系统下各自存储的数据时,服务器需要基于该多种文件系统各自对应的数据接口分别获取各自存储的数据,从而导致数据访问效率较低,数据访问灵活性较低。

发明内容

[0004] 本申请提供一种数据处理方法、装置、设备、介质及程序产品,以提高数据访问效率,提高数据访问灵活性。

[0005] 第一方面,提供一种数据处理方法,包括:设置统一访问入口,统一访问入口用于访问至少两类文件系统;获取数据访问方发送的访问请求,访问请求包括用于访问目标文件系统的访问路径和权限信息,目标文件系统为至少两类文件系统中的任意一个文件系统;响应于访问请求,通过统一访问入口,根据访问路径从目标文件系统中获取与权限信息对应的目标数据;将目标数据发送给数据访问方。

[0006] 第二方面,提供一种数据处理装置,包括:第一设置模块、第一获取模块、第二获取模块、第一发送模块,其中,第一设置模块用于设置统一访问入口,统一访问入口用于访问至少两类文件系统;第一获取模块用于获取数据访问方发送的访问请求,访问请求包括用于访问目标文件系统的访问路径和权限信息,目标文件系统为至少两类文件系统中的任意一个文件系统;第二获取模块用于响应于访问请求,通过统一访问入口,根据访问路径从目标文件系统中获取与权限信息对应的目标数据;第一发送模块用于将目标数据发送给数据访问方。

[0007] 第三方面,提供一种电子设备,包括:处理器和存储器,该存储器用于存储计算机程序,该处理器用于调用并运行该存储器中存储的计算机程序,执行如第一方面或其各实现方式中的方法。

[0008] 第四方面,提供一种计算机可读存储介质,用于存储计算机程序,计算机程序使得

计算机执行如第一方面或其各实现方式中的方法。

[0009] 第五方面,提供一种计算机程序产品,包括计算机程序指令,该计算机程序指令使得计算机执行如第一方面或其各实现方式中的方法。

[0010] 第六方面,提供一种计算机程序,计算机程序使得计算机执行如第一方面或其各实现方式中的方法。

[0011] 通过本申请技术方案,电子设备可以先设置统一访问入口,统一访问入口用于访问至少两类文件系统,然后,电子设备可以获取数据访问方发送的访问请求,访问请求包括用于访问目标文件系统的访问路径和权限信息,目标文件系统为至少两类文件系统中的任意一个文件系统,接着,电子设备可以响应于访问请求,通过统一访问入口,根据访问路径从目标文件系统中获取与权限信息对应的目标数据,最后,电子设备可以将目标数据发送给数据访问方。相比于现有技术中电子设备需要基于多种文件系统各自对应的数据接口分别获取各自存储数据的方式,本申请可以仅根据一个数据接口即上述统一访问入口完成对上述至少两类文件系统中任一文件系统的访问,从而可以提高数据访问效率,提高数据访问灵活性。

附图说明

[0012] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述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发明的一些实施例,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。

[0013] 图1为本申请实施例提供的一种应用场景图;

[0014] 图2为本申请实施例提供的一种数据处理方法的流程图;

[0015] 图3为本申请实施例提供的一种数据处理方法的示意图;

[0016] 图4为本申请实施例提供的另一种数据处理方法的示意图;

[0017] 图5为本申请实施例提供的再一种数据处理方法的示意图;

[0018] 图6为本申请实施例提供的又一种数据处理方法的示意图;

[0019] 图7为本申请实施例提供的又一种数据处理方法的示意图;

[0020] 图8为本申请实施例提供的一种数据处理装置800的示意图;

[0021] 图9是本申请实施例提供的一种电子设备900的示意性框图。

具体实施方式

[0022] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0023] 需要说明的是,本发明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数据在适当情况下可以互换,以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆

盖不排除他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或服务器不必限于清楚地列出的那些步骤或单元,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步骤或单元。

[0024] 如上所述,现有技术中当需要获取多种文件系统下各自存储的数据时,服务器需要基于该多种文件系统各自对应的数据接口分别获取各自存储的数据,从而导致数据访问效率较低,数据访问灵活性较低。

[0025] 为了解决上述技术问题,本申请可以设置统一访问入口,统一访问入口用于访问至少两类文件系统,使得电子设备可以仅根据一个数据接口即上述统一访问入口完成对上述至少两类文件系统中任一文件系统的访问,从而可以提高数据访问效率,提高数据访问灵活性。

[0026] 应理解的是,本申请技术方案可以应用于如下场景,但不限于:

[0027] 在一些可实现方式中,本申请技术方案可以应用于文件系统访问场景中,但不限于此。例如,在用户基于网络设备对文件系统中存储的金融数据进行查询时,电子设备可以获取网络设备发送的访问请求后,电子设备可以响应于该访问请求,从文件系统中获取金融数据,并将金融数据发送给网络设备,其中,文件系统可以是电子设备本地的文件系统,也可以是电子设备外部的文件系统。另外,该文件系统可以是一个文件系统,也可以是至少两类文件系统组成的一个文件系统。

[0028] 在一些可实现方式中,图1为本申请实施例提供的一种应用场景示意图,如图1所示,该应用场景中可以包括电子设备110和网络设备120。电子设备110可以通过有线网络或者无线网络与网络设备120建立连接。

[0029] 示例性的,电子设备110可以是服务器,例如,电子设备110可以是一台服务器,也可以是由多台服务器组成的服务器集群,还可以是云平台控制中心,本申请对此不做限定。网络设备120可以是手机、平板电脑、笔记本电脑、车载设备等终端设备或者服务器,但不限于此。在本申请实施例中,网络设备120可以向电子设备110发送访问请求,该访问请求包括用于访问目标文件系统的访问路径和权限信息,目标文件系统为至少两类文件系统中的任一文件系统,该至少两类文件系统可以是电子设备内部的文件系统,也可以是电子设备外部的文件系统,本申请对此不做限制。进一步地,电子设备110可以响应于上述访问请求,通过统一访问入口,根据访问路径从目标文件系统中获取与权限信息对应的目标数据,之后,电子设备110可以将目标数据发送给网络设备120。

[0030] 此外,图1示例性地给出了一个电子设备和一个网络设备,实际上可以包括其他数量的电子设备和网络设备,本申请对此不作限制。

[0031] 在另一些可实现方式中,本申请技术方案也可以由上述电子设备110执行,或者,本申请技术方案还可以由上述网络设备120执行,本申请对此不做限制。

[0032] 应该理解,图1中的服务器和电子设备的数目仅仅是示意性的,实际上,根据实际情况需要可以设置任意数目的服务器和电子设备,本申请对此不做限制。

[0033] 在介绍了本申请实施例的应用场景之后,下面将对本申请技术方案进行详细阐述:

[0034] 图2为本申请实施例提供的一种数据处理方法的流程图,该方法可以由服务器执行,例如,该服务器可以是图1中的电子设备110,本申请对此不做限制,如图2所示,该方法

可以包括如下步骤:

[0035] S210:设置统一访问入口,统一访问入口用于访问至少两类文件系统;

[0036] S220:获取数据访问方发送的访问请求,访问请求包括用于访问目标文件系统的访问路径和权限信息,目标文件系统为至少两类文件系统中的任意一个文件系统;

[0037] S230:响应于访问请求,通过统一访问入口,根据访问路径从目标文件系统中获取与权限信息对应的目标数据;

[0038] S240:将目标数据发送给数据访问方。

[0039] 在一些可实现方式中,至少两类文件系统可以包括以下至少两项,但不限于此:BOS、HDFS、S3。

[0040] 在一些可实现方式中,电子设备可以确定针对至少两类文件系统的统一协议,在数据访问方的配置文件如core-site.xml中增加FsgBosFileSystem实现类配置,以提供访问至少两类文件系统如BOS和HDFS的相同入口即统一访问入口,其中,统一访问入口的名称为FsgBosFileSystem。另外,电子设备可以在电子设备的Hadoop-common包中Configuration构造方法增加DGS入口,其中,DGS入口可以理解为上述统一访问入口FsgBosFileSystem。

[0041] 需要说明的是,DGS是本申请提供的一种虚拟分布式文件系统,采用上述至少两类文件系统作为离线存储服务,例如,可以采用HDFS和BOS作为离线存储服务。DGS可以采用Etcd作为存储服务,以存储DGS协议即上述针对至少两类文件系统的统一协议、各个数据访问方或者数据访问方对应的组的访问权限等数据。数据访问方需要访问的至少两类文件系统可以是DGS中的文件系统,也就是说,DGS可以和电子设备进行通信连接,当数据访问方需要访问至少两类文件系统中的任一文件系统时,可以向电子设备发送访问请求,之后,电子设备可以获取该访问请求,并响应于访问请求,从DGS中的文件系统中获取访问请求对应的数据,并将获取的数据发送给数据访问方。本申请将在后续实施例中,对DGS的示意图等DGS的相关内容以及DGS协议做进一步介绍,本申请在此先不赘述。

[0042] 在一些可实现方式中,电子设备可以根据至少两类文件系统各自的存储路径格式,确定通用访问路径格式,然后,电子设备可以根据通用访问路径格式确定至少两类文件系统各自对应的访问路径。其中,通用访问路径即为DGS协议中的路径。

[0043] 示例性的,DGS服务端即本申请中的电子设备可以将Hive表分区路径及Hadoop路径进行整合,合并为DGS协议的路径,存储到DGS连接的Etcd中。其中,Hadoop是由Apache基金会开发的一个分布式系统基础架构,其框架的核心包括HDFS;Hive是一种基于Hadoop的数据仓库工具,用来进行数据提取、转化、加载,是一种可以存储、查询和分析存储在Hadoop中的大规模数据的机制。Hive表分区路径、Hadoop路径是两类文件系统各自的存储路径,可以反映出两类文件系统各自的存储单元、存储方式等的特征,电子设备可以根据该特征对上述两类文件系统各自的存储路径进行整合,得到通用访问路径,以便于电子设备可以基于统一访问入口,根据统一协议即DGS协议完成对两类文件系统的访问。

[0044] 在上述实施例中,电子设备可以通过设置针对多种文件系统的统一访问入口以及统一的协议即DGS协议,支持同时访问至少两类文件系统如BOS和HDFS存储服务,打通多种存储架构,可以屏蔽各个文件系统底层的存储引擎使用细节,对数据访问方如客户端实现无感知,使得数据访问方需要访问至少两类文件系统中任一文件系统时,电子设备可以仅

根据一个数据接口即上述统一访问入口完成对上述至少两类文件系统中任一文件系统的访问,从而可以提高数据访问效率,提高数据访问灵活性。

[0045] 在一些可实现方式中,电子设备在获取数据访问方发送的访问请求之前,可以对数据发送方进行权限认证,在对数据发送方进行权限认证通过时,电子设备可以确定数据发送方的权限信息,权限信息包括:对统一访问入口的使用权限、对至少两类文件系统中每个文件系统的访问范围,之后,电子设备可以将权限信息发送给数据发送方,以使数据发送方根据权限信息发送访问请求。其中,电子设备对数据访问方进行权限认证可以包括:电子设备可以先确定数据发送方的登录方式,登录方式可以包括以下任一项,但不限于此:基于命令行接口的登录方式、基于编程语言接口的登录方式,接着,电子设备可以根据登录方式确定数据发送方的登录信息,然后,电子设备可以根据登录信息对数据发送方进行权限认证。

[0046] 示例性的,电子设备可以基于keytab认证或者本地缓存认证对数据访问方进行权限认证。其中,本地缓存认证是基于kerberos认证的一种认证方式,kerberos认证又称门神系统,是业界标准的安全认证。DGS可以集成门神登陆系统即集成kerberos认证,当数据访问方需要从DGS中获取数据时,数据访问方可以先通过门神系统登录电子设备,电子设备可以根据数据访问方登录门神系统的登录方式获取到数据访问方的登录信息如principle信息,接着,电子设备可以从内核中取出kerberos凭据对数据访问方的登录信息进行认证,其中,内核中的kerberos凭据可以理解为电子设备本地缓存的一种登录认证凭证。如果该凭证存在于本地缓存中并且数据访问方的域名符合电子设备预先定义的域名,则电子设备可以确定该数据访问方通过权限认证,即电子设备可以完成登陆验证;否则,电子设备可以确定权限认证未通过,则不会对该数据访问方发送权限信息。

[0047] 示例性的,基于命令行接口的登录方式可以是基于统一的域名进行登录的方式、基于本地挂载服务的登录方式,基于编程语言接口的登录方式可以是基于JAVA语言的登录方式,如基于软件开发工具包(Software Development Kit,SDK)的登录方式。其中,基于本地挂载服务的登录方式可以根据Fuse实现,Fuse为开发者提供了两组接口:fuse_lowlevel_ops、fuse_operations。本申请提供一种基于本地挂载服务的登录方式,可以称为DGS-Fuse,DGS-Fuse可以基于DGS中的文件系统BOS实现,采用的协议为上述实施例中的DGS协议,可以采用Fuse中的fuse_operations接口,将DGS中的文件系统中的需要处理的读、写、创建、删除等函数以回调的方式注册到Fuse模块中。当数据访问方访问挂载目录时,DGS-Fuse的Fuse模块可以回调相应的注册接口,以完成数据访问方的登录。如图3所示,图3中的“Userapace”和“Kernel”分别表示数据访问方和电子设备,“Ls/tmp/test”和“./dgsfs-Ls/tmp/test”分别表示数据访问方通过DGS-Fuse的登录方式登录电子设备时的路径,“glibc”和“libfuse”分别表示数据访问方对应的接口,用于和电子设备进行通信,“VFS”和“FUSE”分别表示电子设备对应的接口,用于和数据访问方进行通信。

[0048] 在一些可实现方式中,上述电子设备确定数据发送方的权限信息可以包括:电子设备可以先确定数据发送方对应的目标组,根据目标组确定目标组对应的权限信息,然后,电子设备可以将目标组对应的权限信息确定为数据发送方的权限信息。其中,目标组也可以称为RCC组。权限信息可以表示数据访问方可以访问的文件系统有哪些、可以访问文件系统中的哪些内容等信息,本申请对此不做限制。

[0049] 示例性的,DGS服务端即电子设备可以先对Hive、Spark、Hadoop等服务均采用以组为单位授权:将Hive表分区路径及Hadoop路径进行整合,合并为DGS协议的路径,存储到DGS连接的Etcd中,并根据上述DGS协议的路径确定各组的权限信息如各组对应的存储空间(bucket)路径权限。接着,电子设备可以将执行用户即数据访问方替换为其所在的目标组即RCC组,并根据目标组从Etcd中获取到目标组对应的权限信息,将目标组对应的权限信息确定为数据发送方的权限信息。其中,Spark是一种专为大规模数据处理而设计的快速通用的计算引擎,是一种大数据服务组件,一般用于数据分析、数据建模等领域。

[0050] 在一些可实现方式中,如图4所示,电子设备对数据访问方进行认证时的各步骤具体如图4中的(b)所示,各步骤与箭头的对应关系如图4中的(a)所示。电子设备对数据访问方进行权限认证的步骤主要包括:密钥分配中心(Key-Distribution Center,KDC)认证、主节点(Gateway)认证与监控从节点(proxy)、RCC组校验、客户端(Client)与(从节点)Proxy交互,其中,KDC认证是门神系统的一部分,用于对数据访问方进行登录认证;Client是一种数据访问方;Gateway是电子设备基于Raft算法选取的主节点,用于对接KDC认证验证数据访问方身份,Raft算法是一种用于实现主从节点之间数据的强一致性的算法;Proxy是电子设备基于Raft算法选取的从节点,负责处理数据访问方的业务请求,如向数据访问方颁发STS,STS是电子设备针对访问BOS生成的一组具体特定操作权限、具有一定时效性的临时AK/SK,这组临时的AK/SK可以暴露给数据访问方直接使用。如图4中的(b)所示,电子设备在对Client进行权限认证之前,可以先通过keytab文件从KDC中获取票据(TGT),并将TGT发送给Gateway,已完成对Gateway的认证,确保后续可以根据Gateway对Client进行再次认证。接着,Client可以通过Kinit命令从KDC获取TGT或者通过GSS-API接口从KDC获取Gateway的密钥(Ticket),完成电子设备对Client的KDC认证。接着,Client可以根据user-group info即Client的RCC组、Ticket向电子设备的Gateway发起数据访问请求,接着,Gateway可以根据user-group info、Ticket对Client进行认证,Gateway可以获取Client对应的组信息即user-group info对Client的权限范围进行校验,在认证通过后,Gateway可以向Client返回(从节点信息)ProxyInfo、加密的(委派密钥)Delegation Token。接着,Client可以根据Delegation Token,通过FileSystem接口向ProxyInfo对应的Proxy发起数据访问请求,该数据访问请求包括user-group info和加密的Delegation Token,之后,Client可以从Proxy处获取STS。其中,Proxy需要向Gateway注册、汇报心跳,Gateway需要监控Proxy的状态等。

[0051] 在一些可实现方式中,在高并发场景下,为了保障数据访问的准确性,电子设备可以为每次请求提供单独的Token。

[0052] 需要说明的是,当数据访问方需要访问的数据来自DGS中的BOS时,数据访问方在完成KDC认证和Gateway认证后,还需要从Proxy处获取STS,从而才可以完成权限认证;当数据访问方需要访问的数据来自DGS中的HDFS时,数据访问方在完成KDC认证和Gateway认证后,即可完成权限认证。

[0053] 示例性的,如图5所示,假设数据访问方是客户端,客户端需要访问的目标数据是DGS中的BOS,那么在客户端通过上述实施例中的方法,通过FileSystem接口请求Gateway完成身份认证后,客户端可以获取到Gateway返回的Delegation Token信息,并根据该信息通过Thrift向电子设备申请访问目标数据,接着,电子设备的Proxy可以向DGS申请临时访问

凭证即STS,电子设备可以自定义需要获取的STS的使用权限和过期时间,获取到STS之后,可以将STS返回给客户端。接着,客户端可以根据STS构造API请求,以请求访问DGS,从DGS中的BOS获取目标数据。在高并发场景下,为了防止节点压力过大,数据流可以不经过DGS,客户端可以利用STS特性,在获取STS后可以直接访问BOS。其中,Thrift是一种接口描述语言和二进制通讯协议,被用来定义和创建跨语言的服务,本申请中其主要用于客户端模块和服务端模块即电子设备的信息传递。

[0054] 在上述实施例中,电子设备可以在获取数据访问方发送的访问请求之前,先对数据访问方进行权限认证,可以保证数据访问的安全性,而且,电子设备可以以组为单位对数据访问方进行存储空间的访问限制和鉴权,对不同组实现不同的访问路径权限,从而可以保证数据的隔离性、权限的正确性和安全性,在存储的数据为金融数据时,是符合金融安全规范和金融法规标准如金融4A规范的,保证数据访问的安全。

[0055] 另外,电子设备可以支持多种登录方式如基于命令行接口的登录方式、基于编程语言接口的登录方式等,以提高数据访问方访问数据时登录电子设备的多样性。

[0056] 在一些可实现方式中,访问路径可以包括目标文件系统的集群名称,电子设备在根据访问路径从目标文件系统中获取与权限信息对应的目标数据之前,还可以先确定至少两类文件系统各自对应的集群名称,然后,根据至少两类文件系统各自对应的集群名称,设置至少两类文件系统各自对应的子访问接口,接着,电子设备可以根据访问路径中的目标文件系统的集群名称,确定目标文件系统对应的子访问接口,根据目标文件系统对应的子访问接口,从目标文件系统中获取与权限信息对应的目标数据。从而确保可以较准确地从两类文件系统中确定出目标文件系统,减少故障频率。

[0057] 示例性的,当电子设备设置统一访问入口之后,可以通过集群名称区分至少两类文件系统系统中的每个文件系统。例如,假设至少两类文件系统分别是HDFS和BOS,电子设备可以针对HDFS和BOS分别设置其集群名称为baihua、baize,针对HDFS和BOS分别设置其访问路径为dgs://baihua/user/、dgs://baize/user/等,其中,“dgs”表示访问数据时使用的协议为DGS协议。另外,电子设备还可以根据至少两类文件系统各自的集群名称等信息确定DGS协议中的文件系统协议,以便于对至少两类文件系统进行统一的操作。

[0058] 示例性的,针对BOS,电子设备基于访问路径获取集群名称的方式可以是:通过统一访问接口,使用请求语法的方式如get BucketName获取集群名称,并构建(存储空间名称)bucketname,以便于后续可以从BOS中获取与权限信息对应的目标数据。针对HDFS,电子设备基于访问路径获取集群名称的方式可以是:先在配置文件如hdfs-site.xml中配置集群名称,然后通过命令如dfs.nameservices获取集群名称,并发起HDFS IPC操作,以便于后续可以从HDFS中获取与权限信息对应的目标数据,HDFS IPC操作是HDFS中的一种通信机制。

[0059] 示例性的,假设至少两类文件系统分别是HDFS和BOS,电子设备可以确定其各自对应的子访问接口可以分别是HDFS client接口和BOS接口。同时,统一访问入口对应的类可以同时支持HDFS client接口操作和BOS接口操作,以使得电子设备可以通过统一访问入口,基于子访问接口确定目标数据。

[0060] 在一些可实现方式中,如图6所示,图6中的“Guest”可以表示数据访问方,“FsgBosFileSystem”表示统一访问入口,DFSCClient是一种使用HDFS各项功能的起点的类,

DFSCClient可以连接到HDFS,对外提供管理文件、管理目录、读写文件以及管理与配置HDFS系统等功能,电子设备可以将HDFS的子访问接口设置在DFSCClient中,“NativeFileSystemStore”是一种使用BOS各项功能的起点的类,NativeFileSystemStore可以连接到BOS,对外提供管理文件、管理目录、读写文件以及管理与配置BOS系统等功能,电子设备可以将BOS的子访问接口设置在NativeFileSystemStore中。数据访问方可以向电子设备发送包括权限信息和访问路径的访问请求,电子设备在获取到访问请求后,可以通过统一访问入口,根据访问路径确定集群名称,并根据集群名称确定目标文件系统是HDFS还是BOS,进而确定子访问接口,从而通过子访问接口,从目标文件系统中获取与权限信息对应的目标数据。

[0061] 在一些可实现方式中,如图7所示,电子设备可以包括认证层、接入层、管理层和存储层,其中,存储层可以是DGS,DGS可以包括多类文件系统、节点组群和元数据存储仓库,多类文件系统可以包括BOS和HDFS,节点组群在图7中使用“DGS Raft Group”表示,节点组群可以包括主节点(Gateway)和(从节点)Proxy,元数据存储仓库在图7中使用“DGS metaStore”表示,元数据存储仓库可以包括在图7中使用“Etcd Node”表示的多个用于存储元数据的节点,与DGS相关的内容可以参考上述实施例,本申请在此不做赘述。认证层可以对数据访问方进行初始认证如上述权限认证,初始认证时可以根据接入层获取的数据访问方的登录方式,确定登录信息,根据登录信息完成初始认证,与初始认证相关的内容可以参考上述实施例,本申请在此不做赘述。在初始认证结束之后,电子设备可以基于管理层对数据访问方进行二次认证如组级别的认证,以确定数据访问方的权限信息,其中,组级别的认证可以理解为根据数据访问方对应的目标组确定其权限信息,其具体过程可以参考上述实施例,本申请在此不做赘述。在完成对数据访问方的二次认证之后,电子设备可以根据数据访问方的权限信息访问DGS中的文件系统。从而,本申请实施例提供了一套完整的认证及授权机制,保证了数据的隔离性和安全性。

[0062] 图8为本申请实施例提供的一种数据处理装置800的示意图,如图8所示,该装置800包括:

[0063] 第一设置模块801,用于设置统一访问入口,统一访问入口用于访问至少两类文件系统;

[0064] 第一获取模块802,用于获取数据访问方发送的访问请求,访问请求包括用于访问目标文件系统的访问路径和权限信息,目标文件系统为至少两类文件系统任一文件系统;

[0065] 第二获取模块803,用于响应于访问请求,通过统一访问入口,根据访问路径从目标文件系统中获取与权限信息对应的目标数据;

[0066] 第一发送模块804,用于将目标数据发送给数据访问方。

[0067] 在一些可实现方式中,访问路径包括目标文件系统的集群名称,装置800还包括:第一确定模块805、第二设置模块806;第一确定模块805用于确定至少两类文件系统各自对应的集群名称;第二设置模块806用于根据至少两类文件系统各自对应的集群名称,设置至少两类文件系统各自对应的子访问接口;第二获取模块803具体用于:根据目标文件系统的集群名称,确定目标文件系统对应的子访问接口;根据目标文件系统对应的子访问接口,从目标文件系统中获取与权限信息对应的目标数据。

[0068] 在一些可实现方式中,装置800还包括:认证模块807、第二确定模块808、第二发送模块809;认证模块807用于对数据发送方进行权限认证;第二确定模块808用于在对数据发送方进行权限认证通过时,确定数据发送方的权限信息,权限信息包括:对统一访问入口的使用权限、对至少两类文件系统中每个文件系统的访问范围;第二发送模块809用于将权限信息发送给数据发送方,以使数据发送方根据权限信息发送访问请求。

[0069] 在一些可实现方式中,第二确定模块808具体用于:确定数据发送方对应的目标组;根据目标组确定目标组对应的权限信息;将目标组对应的权限信息确定为数据发送方的权限信息。

[0070] 在一些可实现方式中,认证模块807具体用于:确定数据发送方的登录方式,登录方式包括以下任一项:基于命令行接口的登录方式、基于编程语言接口的登录方式;根据登录方式确定数据发送方的登录信息;根据登录信息对数据发送方进行权限认证。

[0071] 在一些可实现方式中,装置800还包括:第三确定模块810、第四确定模块811;第三确定模块810用于根据至少两类文件系统各自的存储路径格式,确定通用访问路径格式;第四确定模块811用于根据通用访问路径格式确定至少两类文件系统各自对应的访问路径。

[0072] 应理解的是,装置实施例与方法实施例可以相互对应,类似的描述可以参照方法实施例。为避免重复,此处不再赘述。具体地,图8所示的装置800可以执行上述方法实施例,并且装置800中的各个模块的前述和其它操作和/或功能分别为了实现上述各个方法中的相应流程,为了简洁,在此不再赘述。

[0073] 上文中结合附图从功能模块的角度描述了本申请实施例的装置800。应理解,该功能模块可以通过硬件形式实现,也可以通过软件形式的指令实现,还可以通过硬件和软件模块组合实现。具体地,本申请实施例中的方法实施例的各步骤可以通过处理器中的硬件的集成逻辑电路和/或软件形式的指令完成,结合本申请实施例公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。可选地,软件模块可以位于随机存储器,闪存、只读存储器、可编程只读存储器、电可擦写可编程存储器、寄存器等本领域的成熟的存储介质中。该存储介质位于存储器,处理器读取存储器中的信息,结合其硬件完成上述方法实施例中的步骤。

[0074] 图9是本申请实施例提供的电子设备900的示意性框图。

[0075] 如图9所示,该电子设备900可包括:

[0076] 存储器910和处理器920,该存储器910用于存储计算机程序,并将该程序代码传输给该处理器920。换言之,该处理器920可以从存储器910中调用并运行计算机程序,以实现本申请实施例中的方法。

[0077] 例如,该处理器920可用于根据该计算机程序中的指令执行上述方法实施例。

[0078] 在本申请的一些实施例中,该处理器920可以包括但不限于:

[0079] 通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等等。

[0080] 在本申请的一些实施例中,该存储器910包括但不限于:

[0081] 易失性存储器和/或非易失性存储器。其中,非易失性存储器可以是只读存储器

(Read-Only Memory,ROM)、可编程只读存储器(Programmable ROM,PROM)、可擦除可编程只读存储器(Erasable PROM,EPROM)、电可擦除可编程只读存储器(Electrically EPROM,EEPROM)或闪存。易失性存储器可以是随机存取存储器(Random Access Memory,RAM),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的RAM可用,例如静态随机存取存储器(Static RAM,SRAM)、动态随机存取存储器(Dynamic RAM,DRAM)、同步动态随机存取存储器(Synchronous DRAM,SDRAM)、双倍数据速率同步动态随机存取存储器(Double Data Rate SDRAM,DDR SDRAM)、增强型同步动态随机存取存储器(Enhanced SDRAM,ESDRAM)、同步连接动态随机存取存储器(synch link DRAM,SLDRAM)和直接内存总线随机存取存储器(Direct Rambus RAM,DR RAM)。

[0082] 在本申请的一些实施例中,该计算机程序可以被分割成一个或多个模块,该一个或者多个模块被存储在该存储器910中,并由该处理器920执行,以完成本申请提供的方法。该一个或多个模块可以是能够完成特定功能的一系列计算机程序指令段,该指令段用于描述该计算机程序在该电子设备中的执行过程。

[0083] 如图9所示,该电子设备还可包括:

[0084] 收发器930,该收发器930可连接至该处理器920或存储器910。

[0085] 其中,处理器920可以控制该收发器930与其他设备进行通信,具体地,可以向其他设备发送信息或数据,或接收其他设备发送的信息或数据。收发器930可以包括发射机和接收机。收发器930还可以进一步包括天线,天线的数量可以为一个或多个。

[0086] 应当理解,该电子设备中的各个组件通过总线系统相连,其中,总线系统除包括数据总线之外,还包括电源总线、控制总线和状态信号总线。

[0087] 本申请还提供了一种计算机存储介质,其上存储有计算机程序,该计算机程序被计算机执行时使得该计算机能够执行上述方法实施例的方法。或者说,本申请实施例还提供一种包含指令的计算机程序产品,该指令被计算机执行时使得计算机执行上述方法实施例的方法。

[0088] 当使用软件实现时,可以全部或部分地以计算机程序产品的形式实现。该计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行该计算机程序指令时,全部或部分地产生按照本申请实施例该的流程或功能。该计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。该计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输,例如,该计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(digital subscriber line,DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。该计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。该可用介质可以是磁性介质(例如,软盘、硬盘、磁带)、光介质(例如数字视频光盘(digital video disc,DVD))、或者半导体介质(例如固态硬盘(solid state disk,SSD))等。

[0089] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的模块及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。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。

[0090] 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,该模块的划分,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,例如多个模块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,装置或模块的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0091] 作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理模块,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模块来实现本实施例方案的目的。例如,在本申请各个实施例中的各功能模块可以集成在一个处理模块中,也可以是各个模块单独物理存在,也可以两个或两个以上模块集成在一个模块中。

[0092] 以上仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以该权利要求的保护范围为准。

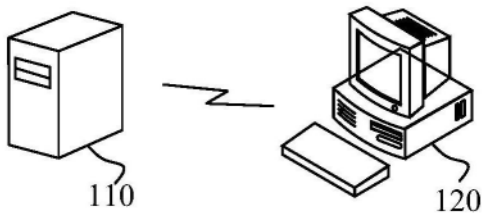


图1

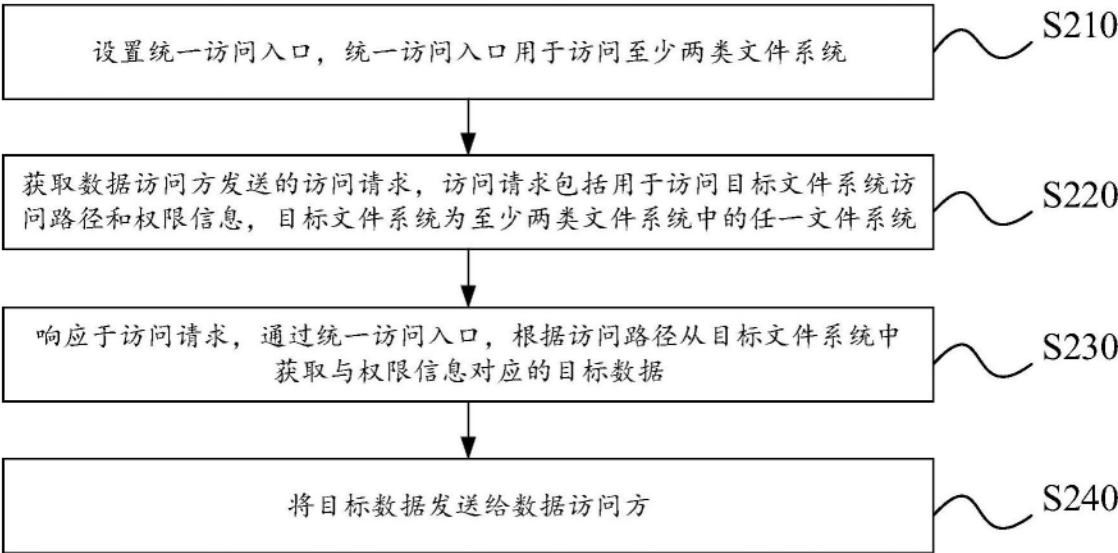


图2

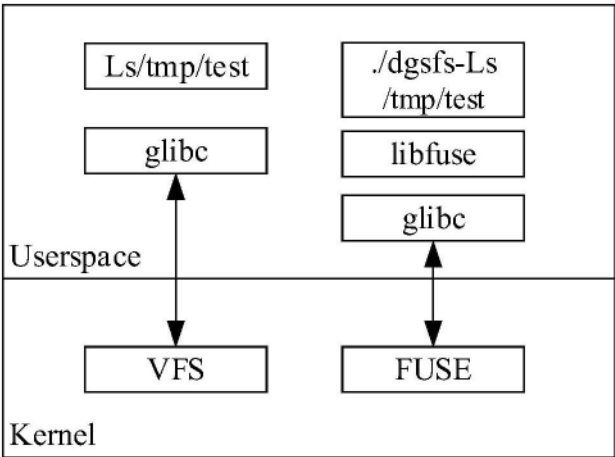


图3

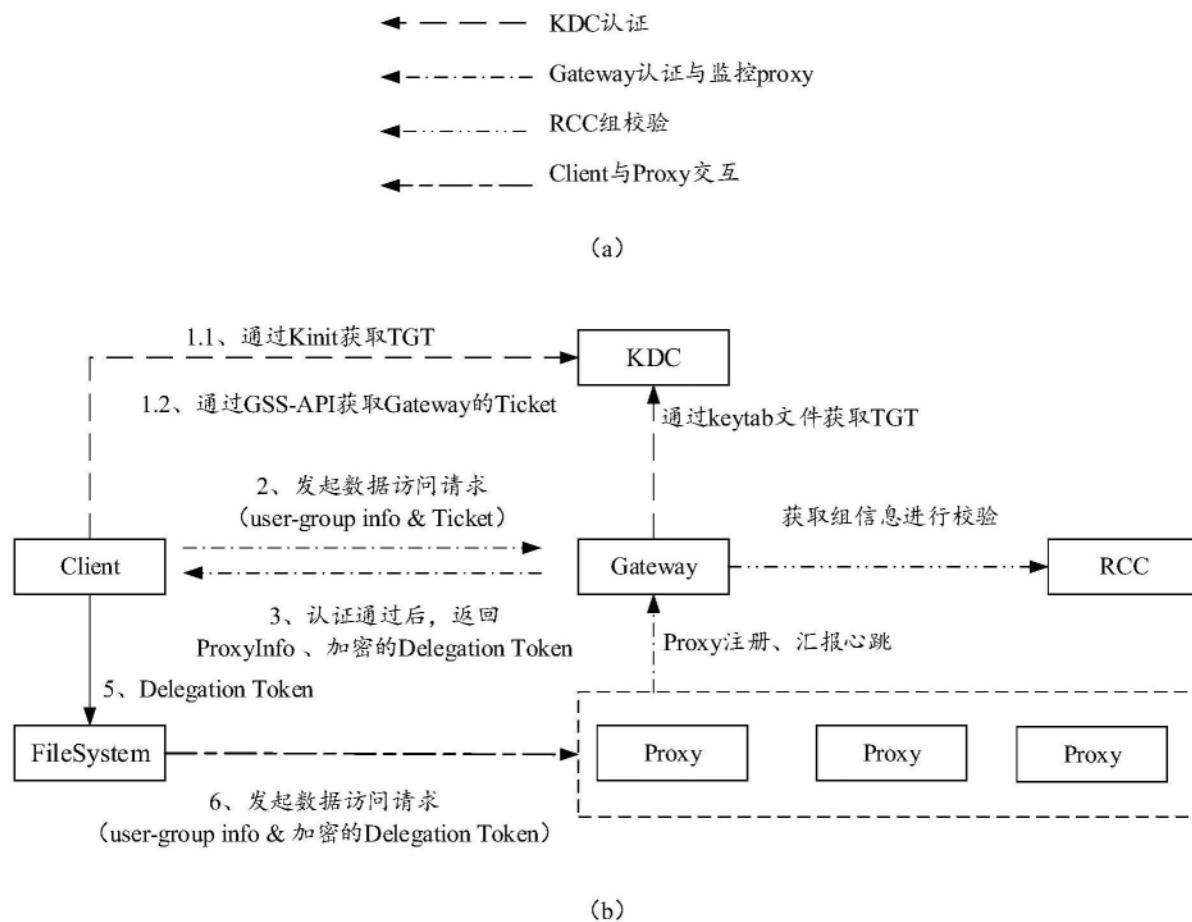


图4

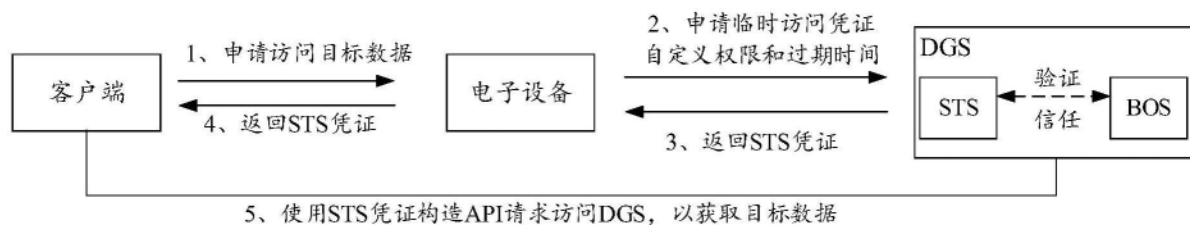


图5

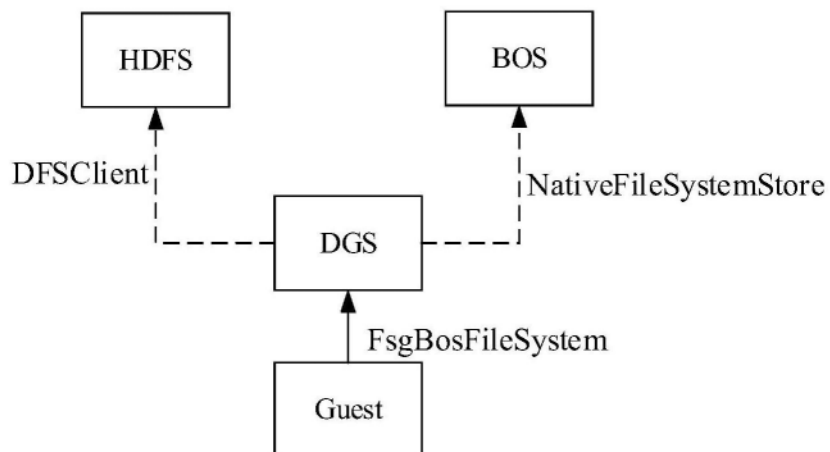


图6

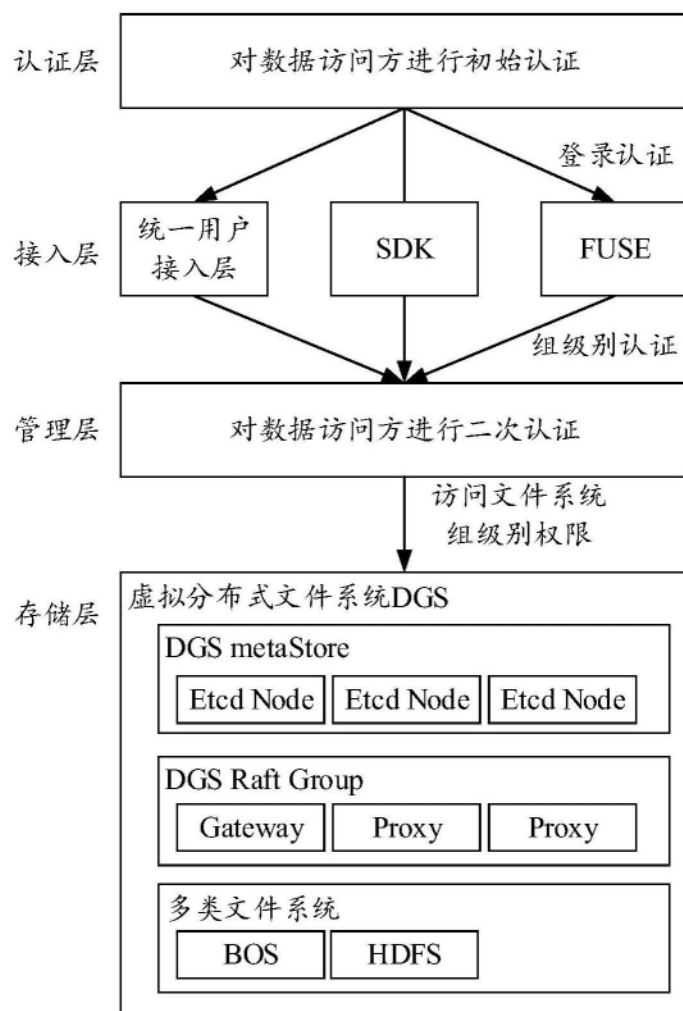


图7



图8

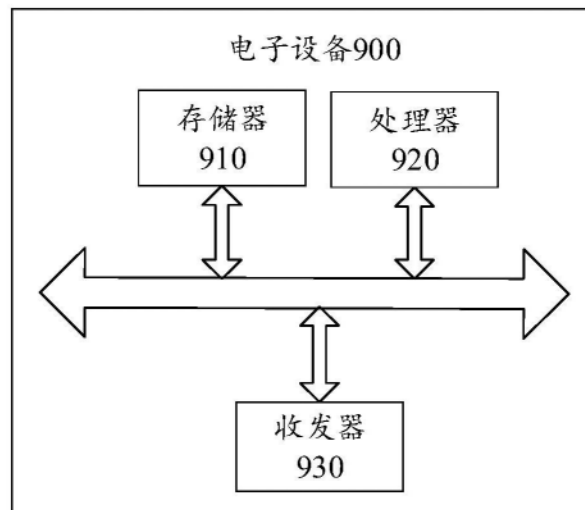


图9