



(12) 发明专利申请

(10) 申请公布号 CN 115756549 A

(43) 申请公布日 2023. 03. 07

(21) 申请号 202211509504.7

(22) 申请日 2022.11.29

(71) 申请人 重庆长安汽车股份有限公司
地址 400023 重庆市江北区建新东路260号

(72) 发明人 魏鑫

(74) 专利代理机构 广州市华学知识产权代理有限公司 44245
专利代理师 赖晋儒

(51) Int. Cl.
G06F 8/65 (2018.01)
G06F 8/71 (2018.01)

权利要求书2页 说明书8页 附图3页

(54) 发明名称

一种大数据中台数据的下载方法、装置及存储介质

(57) 摘要

本发明属于大数据中台数据下载技术领域，具体涉及一种大数据中台数据的下载方法、装置及存储介质。该方法包括以下步骤：S1、数据源管理，通过数据源配置项，用于对数据源增删改查和数据源统一管理；S2、数据下载程序开发，基于大数据引擎开发数据下载程序；S3、数据下载任务提交，基于大数据引擎完成数据下载程序任务提交；S4、数据下载任务监控，基于大数据任务监控组件实现对数据下载程序任务的监控；S5、数据文件下载，在数据下载任务中，从数据文件临时存储系统目录下载结果数据到客户端。其目的是：能够实现数据源的同一管理，使海量数据的下载不受版本和数据源的影响，并提高数据下载效率。



1. 一种大数据中台数据的下载方法,其特征在于:包括以下步骤:

S1、数据源管理,通过数据源配置项,用于对数据源增删改查和数据源统一管理;

S2、数据下载程序开发,基于大数据引擎开发数据下载程序;

S3、数据下载任务提交,基于大数据引擎完成数据下载程序任务提交;

S4、数据下载任务监控,基于大数据任务监控组件实现对数据下载程序任务的监控;

S5、数据文件下载,在数据下载任务中,从数据文件临时存储系统目录下载结果数据到客户端。

2. 根据权利要求1所述的一种大数据中台数据的下载方法,其特征在于:所述S1步骤中,数据源包括但不限于数据库、数据文件、接口和消息总线,所述数据库包括但不限于mysql、oracle、hbase、hive和es,所述数据文件包括但不限于hdfs、ftp和s3,所述接口包括但不限于rest api、web sockets,所述消息总线包括但不限于kafka、rabbitmq、pulsar。

3. 根据权利要求2所述的一种大数据中台数据的下载方法,其特征在于:所述方法还包括,

在通过数据源的配置项时,通过配置连接地址、用户名、密码或认证密钥文件,将数据源统一管理,且能够用于对数据源增删改查。

4. 根据权利要求1所述的一种大数据中台数据的下载方法,其特征在于:所述方法还包括,

根据不同的设计引擎,对不同的数据源开发下载程序,下载程序根据任务执行规则,获取数据源连接必要参数、数据临时存储文件系统连接必要参数和数据下载需求信息,数据下载执行后,统一将下载信息写入临时数据文件系统,得到下载结果数据文件,并连接配置项。

5. 根据权利要求4所述的一种大数据中台数据的下载方法,其特征在于:所述方法还包括,

下载结果数据文件包括数据文件夹和json描述文件,所述的数据文件夹为一个或多个数据文件的压缩文件,所述json描述文件包括但不限于结构化数据字段名、数据类型、数据字段说明、数据任务id、数据文件压缩文件个数、数据任务流水号;

所述json描述文件还包括数据文件夹中各个数据文件压缩文件的数据文件后缀、数据文件压缩格式、数据文件名、数据文件大小、数据文件中数据集条数。

6. 根据权利要求4所述的一种大数据中台数据的下载方法,其特征在于:所述方法还包括,

在完成数据下载程序任务提交过程中,调用大数据引擎执行提交;或调用第三方服务执行提交;

根据数据下载程序中传入的数据源配置信息、数据下载临时存储路径、文件系统配置信息、数据下载规则调用大数据引擎或第三方服务。

7. 根据权利要求1所述的一种大数据中台数据的下载方法,其特征在于:所述方法还包括,

在数据下载任务监控过程中,使用大数据集群任务资源管理组件,并调用任务资源组件接口,监控数据下载任务的执行资源情况、执行状态、执行结果和执行日志。

8. 根据权利要求5所述的一种大数据中台数据的下载方法,其特征在于:所述方法还包

括，

当提交的下载任务已经执行过时，将下载结果数据文件中的json描述文件中任务相关信息修改后，与数据文件一起下载到客户端；

当数据下载任务未执行过，启动数据下载程序，监控任务执行完成后，从数据文件临时存储系统目录下下载数据到客户端。

9. 一种大数据中台数据的下载装置，其特征在于，所述装置包括：

数据源模块，用于通过数据源配置项或配置文件，实现增删改查功能和数据源统一管理；

执行模块，用于根据大数据引擎开发数据下载程序，并完成数据下载程序任务提交；

监控模块，用于根据大数据任务监控组件实现对数据下载程序任务的监控；

下载模块，用于将数据文件从数据文件临时存储系统目录下下载结果数据到客户端。

10. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质中存储有计算机程序，当所述计算机程序在计算机上运行时，使得所述计算机执行如权利要求1-7中任一项所述的方法。

一种大数据中台数据的下载方法、装置及存储介质

技术领域

[0001] 本发明属于大数据中台数据下载技术领域,具体涉及一种大数据中台数据的下载方法、装置及存储介质。

背景技术

[0002] 随着数据量的快速增长,数据体量指数增长,面对海量数据必将出现数据难理解、难获取、难处理和难组织等四个难题。大数据处理技术其中涉及数据管理、数据治理、数据组织和数据服务等。数据下载服务是大数据中台数据管理与数据服务的重要部分功能,根据数据需求提供大数据中台数据下载服务,可以与其他系统和平台完成数据交换和数据下载。

[0003] 数据中台有多种数据源的海量数据的下载功能,面对海量数据,多任务并行下载时,普通下载方法不能满足该业务场景。并且通常数据下载功能与业务场景紧耦合,不同业务需求需要整合多种数据源的数据下载功能,数据下载功能存在重复开发,重复依赖等问题。

发明内容

[0004] 本发明的目的是:旨在提供一种大数据中台数据的下载方法、装置及存储介质,能够实现数据源的同一管理,使海量数据的下载不受版本和数据源的影响,并提高数据下载效率。

[0005] 为实现上述技术目的,本发明采用的技术方案如下:

[0006] 第一方面,本申请提供了一种大数据中台数据的下载方法,包括以下步骤:

[0007] S1、数据源管理,通过数据源配置项,用于对数据源增删改查和数据源统一管理;

[0008] S2、数据下载程序开发,基于大数据引擎开发数据下载程序;

[0009] S3、数据下载任务提交,基于大数据引擎完成数据下载程序任务提交;

[0010] S4、数据下载任务监控,基于大数据任务监控组件实现对数据下载程序任务的监控;

[0011] S5、数据文件下载,在数据下载任务中,从数据文件临时存储系统目录下载结果数据到客户端。

[0012] 结合第一方面,在一些可选的实施方式中,所述S1步骤中,数据源包括但不限于数据库、数据文件、接口和消息总线,所述数据库包括但不限于mysql、oracle、hbase、hive和es,所述数据文件包括但不限于hdfs、ftp和s3,所述接口包括但不限于rest api、websockets,所述消息总线包括但不限于kafka、rabbitmq、pulsar。

[0013] 结合第一方面,在一些可选的实施方式中,所述方法还包括,

[0014] 在通过数据源的配置项时,通过配置连接地址、用户名、密码或认证密钥文件,将数据源统一管理,且能够用于对数据源增删改查。

[0015] 结合第一方面,在一些可选的实施方式中,所述方法还包括,

[0016] 根据不同的设计引擎,对不同的数据源开发下载程序,下载程序根据任务执行规则,获取数据源连接必要参数、数据临时存储文件系统连接必要参数和数据下载需求信息,数据下载执行后,统一将下载信息写入临时数据文件系统,得到下载结果数据文件,并连接配置项。

[0017] 结合第一方面,在一些可选的实施方式中,所述方法还包括,

[0018] 下载结果数据文件包括数据文件夹和json描述文件,所述的数据文件夹为一个或多个数据文件的压缩文件,所述json描述文件包括但不限于结构化数据字段名、数据类型、数据字段说明、数据任务id、数据文件压缩文件个数、数据任务流水号;

[0019] 所述json描述文件还包括数据文件夹中各个数据文件压缩文件的数据文件后缀、数据文件压缩格式、数据文件名、数据文件大小、数据文件中数据集条数。

[0020] 结合第一方面,在一些可选的实施方式中,所述方法还包括,

[0021] 在完成数据下载程序任务提交过程中,调用大数据引擎执行提交;或调用第三方服务执行提交;

[0022] 根据数据下载程序中传入的数据源配置信息、数据下载临时存储路径、文件系统配置信息、数据下载规则调用大数据引擎或第三方服务。

[0023] 结合第一方面,在一些可选的实施方式中,所述方法还包括,

[0024] 在数据下载任务监控过程中,使用大数据集群任务资源管理组件,并调用任务资源组件接口,监控数据下载任务的执行资源情况、执行状态、执行结果和执行日志。

[0025] 结合第一方面,在一些可选的实施方式中,所述方法还包括,

[0026] 当提交的下载任务已经执行过时,将下载结果数据文件中的json描述文件中任务相关信息修改后,与数据文件一起下载到客户端;

[0027] 当数据下载任务未执行过,启动数据下载程序,监控任务执行完成后,从数据文件临时存储系统目录下下载数据到客户端。

[0028] 第二方面,本申请还公开了一种大数据中台数据的下载装置,所述装置包括:

[0029] 数据源模块,用于通过数据源配置项,实现增删改查功能和数据源统一管理;

[0030] 执行模块,用于根据大数据引擎开发数据下载程序,并完成数据下载程序任务提交;

[0031] 监控模块,用于根据大数据任务监控组件实现对数据下载程序任务的监控;

[0032] 下载模块,用于将数据文件从数据文件临时存储系统目录下下载结果数据到客户端。

[0033] 第三方面,本申请还公开了一种计算机可读存储介质,所述计算机可读存储介质中存储有计算机程序,当所述计算机程序在计算机上运行时,使得所述计算机执行如上述所述的方法。

[0034] 采用上述技术方案的发明,具有如下优点:

[0035] 1.可以统一多数据源数据下载,大数据中台中统一的数据源管理,实现数据源管理增删改查,使数据源使用清楚明了,方便快捷;开发基于大数据引擎针对不同的数据源下载程序,由统一的任务调度服务执行,由统一的任务监控平台监控,实现统一管理,增加开发迭代效率;

[0036] 2.基于大数据引擎的数据下载程序,无下载数据量大小的瓶颈,能够下载处理海

量数据,并提高下载效率;

[0037] 3.数据下载程序任务执行后,下载的数据文件可写入多种数据文件系统,做临时存储,使相同数据下载任务不用重复执行数据下载程序,从数据文件系统中下载已执行过下载任务的数据即可,更加方便;

[0038] 4.统一的下载任务管理,基于大数据引擎的下载任务,可由大数据数据集群任务调度器实现统一任务管理,可对任务执行资源、执行进度、执行状态等统一监控,在数据下载服务中只需调用数据集群任务管理器接口,即可监控任务状态;

[0039] 5.基于大数据引擎的下载任务可灵活开发任务执行程序,灵活扩展,对不同数据源开发不同数据下载程序,对不同版本的数据源开发不同版本的数据下载程序,各种类型数据源,各版本数据源的数据下载程序间无关联,无相互影响。

附图说明

[0040] 本申请可以通过附图给出的非限定性实施例进一步说明。应当理解,以下附图仅示出了本申请的某些实施例,因此不应被看作是对范围的限定,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相关的附图;

[0041] 图1为本申请实施例提供的下载设计方法的流程示意图一;

[0042] 图2为本申请实施例提供的下载设计方法的流程示意图二;

[0043] 图3为本申请实施例提供的下载装置的框图;

[0044] 主要元件符号说明如下:

[0045] 下载装置200、数据源模块210、执行模块220、监控模块230、下载模块240。

具体实施方式

[0046] 以下将结合附图和具体实施例对本申请进行详细说明,需要说明的是,在附图或说明书描述中,相似或相同的部分都使用相同的图号,附图中未绘示或描述的实现方式,为所属技术领域中普通技术人员所知的形式。在本申请的描述中,术语“第一”、“第二”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0047] 请参考附图1所示,本申请实施例提供了一种大数据中台数据的下载方法,包括以下步骤:

[0048] 步骤110、数据源管理,通过数据源配置项,用于对数据源增删改查和数据源统一管理;

[0049] 步骤120、数据下载程序开发,基于大数据引擎开发数据下载程序;

[0050] 步骤130、数据下载任务提交,基于大数据引擎完成数据下载程序任务提交;

[0051] 步骤140、数据下载任务监控,基于大数据任务监控组件实现对数据下载程序任务的监控;

[0052] 步骤150、数据文件下载,在数据下载任务中,从数据文件临时存储系统目录下载结果数据到客户端。

[0053] 基于上述设计,统一多数据源数据下载,大数据中台中统一的数据源管理,实现数据源管理增删改查,使数据源使用清楚明了,方便快捷;且开发基于大数据引擎针对不同的数据源下载程序,由统一的任务调度服务执行,由统一的任务监控平台监控,能够无下载数

据量大小的瓶颈,能够下载处理海量数据,并提高下载效率,从而实现统一管理,增加开发迭代效率。

[0054] 作为一种可选的实施方式,在步骤110中,数据源包括但不限于数据库、数据文件、接口和消息总线,所述的数据库包括但不限于mysql、oracle、hive、hbase和es;所述的数据文件包括但不限于hdfs、ftp和s3;所述的接口包括但不限于rest api、web sockets;所述的消息总线包括但不限于kafka、rabbitmq、pulsar。

[0055] 可以理解的,将多种不同类型的数据源整合统一,有利于对不同的下载进行服务,不必再根据不同的数据源开发不同的下载程序,实现了资源的统一管理。

[0056] 作为一种可选的实施方式,方法还可以包括,

[0057] 在通过数据源的配置项时,通过配置连接地址、用户名、密码或认证密钥文件,将数据源统一管理,且能够用于对数据源增删改查。

[0058] 可以理解的,当数据源为hive时,配置url连接地址、用户名、密码或者hive kerberos认证密钥文件;当hive数据源服务中没有驱动时,配置hive jdbc驱动,当hive数据源服务中有驱动时,则不配置驱动。

[0059] 当数据源为Mysql时,配置url连接地址、用户名、密码,当Mysql数据源服务中没有驱动时,配置mysql jdbc驱动;当Mysql数据源服务中有驱动时,则不配置驱动。

[0060] 当数据源为Hbase时,配置zookeeper连接地址、用户名、密码或者hbase kerberos认证密钥文件。

[0061] 可以理解的,上述的数据源配置仅为一个示范性实施方式,根据不同的数据源,通过配置不同的连接地址、用户名、密码或认证密钥文件,从而对数据源进行增删改查,进而能够在下载服务时,登录数据源进行不同数据的下载。

[0062] 可以理解的,数据源还可以通过配置文件将数据源统一,配置文件可以是认证密钥文件、认证的票据文件,也可以是本领域常用的其他认证文件。

[0063] 作为一种可选的实施方式,方法还可以包括,

[0064] 在步骤120中,根据不同的设计引擎,对不同的数据源开发下载程序,下载程序根据任务执行规则,获取数据源连接必要参数、数据临时存储文件系统连接必要参数和数据下载需求信息,数据下载执行后,统一将下载信息写入临时数据文件系统,得到下载结果数据文件,并连接配置项或配置文件。

[0065] 在本实施例中,设计引擎包括但不限于spark、flink、mr,可以根据不同的设计引擎,相应的开发数据下载程序,使开发的数据下载程序,能够实现对数据库、数据文件、接口和消息总线等数据源中的数据进行下载。

[0066] 在本实施例中,开发的数据下载程序在根据任务执行规则,获取必要参数时,采用如下规则,

[0067] 根据相应的数据源,获取数据源的配置项,从而从在数据源中获取数据。以下通过具体的数据源进行说明,当数据源为source时,获取source数据源的连接地址、用户名、密码或者认证密钥文件,在source数据源中获取数据。当数据源为sink数据存储系统时,获取sink数据存储系统的连接地址、用户名、密码或者认证密钥文件,将要下载的数据写入数据存储系统,从而获取想要的数。

[0068] 可以理解的,在不同的数据源中,根据开发的数据下载程序,控制要下载的数据内

容和控制数据描述的基本信息,开发的数据下载程序也可以进行离线下载。

[0069] 作为一种可选的实施方式,方法还可以包括,

[0070] 在步骤120中,下载结果数据文件包括数据文件夹和json描述文件,所述的数据文件夹为一个或多个数据文件的压缩文件,所述json描述文件包括但不限于结构化数据字段名、数据类型、数据字段说明、数据任务id、数据文件压缩文件个数、数据任务流水号;

[0071] 所述json描述文件还包括数据文件夹中各个数据文件压缩文件的数据文件后缀、数据文件压缩格式、数据文件名、数据文件大小、数据文件中数据集条数。

[0072] 可以理解的,在开发下载程序任务执行规则时,控制下载数据字段、下载数据压缩方式、和每个数据文件的大小等,且控制json描述文件中描述的基本信息,从数据下载规则中获取,数据字段名、数据类型、数据字段说明和任务id等,从而能够准确的执行下载任务。

[0073] 作为一种可选的实施方式,方法还可以包括,

[0074] 在步骤130中,在完成数据下载程序任务提交过程中,调用大数据引擎执行提交;或调用第三方服务执行提交;

[0075] 根据数据下载程序中传入的数据源配置信息、数据下载临时存储路径、文件系统配置信息、数据下载规则调用大数据引擎或第三方服务。

[0076] 在本实施例中,当采用flink引擎时,数据下载程序可基于客户端提交方式,当采用spark引擎时,数据下载程序可基于第三方服务livy提交等。

[0077] 可以理解的,数据源配置信息和文件系统配置信息为数据库、数据文件、接口和消息总线等数据库的连接地址、用户名、密码或者认证密钥文件等。数据下载规则为下载程序任务执行规则。

[0078] 作为一种可选的实施方式,方法还可以包括,

[0079] 在步骤140中,在数据下载任务监控过程中,使用大数据集群任务资源管理组件,并调用任务资源组件接口,监控数据下载任务的执行资源情况、执行状态、执行结果和执行日志。

[0080] 在本实施中,大数据集群任务资源管理组件包括但不限于Yanr,以Yanr为例进行说明,原生Yanr有提供查询RESTful API接口,在下载服务中调用接口,传入任务id通过http定时查询,获取任务进度。具体的,每秒带着任务id去调接口,第一次查询返回10%,第二次20%,直到完成数据下载任务。

[0081] 作为一种可选的实施方式,方法还可以包括,

[0082] 在步骤150中,当提交的下载任务已经执行过时,将下载结果数据文件中的json描述文件中任务相关信息修改后,与数据文件一起下载到客户端;

[0083] 当数据下载任务未执行过,启动数据下载程序,监控任务执行完成后,从数据文件临时存储系统目录下下载数据到客户端。

[0084] 在本实施例中,数据已经下载过一次时,两次或者多次下载数据文件是相同的时,json描述文件中的描述信息不一样,更新任务id、流水账号、下载用户的用户名、下载任务时间等信息。系统自动根据任务规则跟新,Json描述文件重新跟新,生成数据任务不用重复执行。

[0085] 可以理解的,当数据下载后,再需要下载数据时,不用再从数据源去重新下载数据,只需要在数据文件临时存储系统目录下查询,直接从数据文件临时存储系统中下载即

可。

[0086] 请参考附图2,下面对大数据中台数据的下载方法进行具体阐述,如下:

[0087] S1、数据源管理,通过设置数据库、数据文件、接口和消息总线等数据源的连接地址、用户名、密码或认证密钥文件等配置项,将不同的数据源统一管理,且通过配置项,能够用于对数据源中内容的增删改查;

[0088] S2、数据下载程序开发,基于spark、flink、mr等任一个引擎,开发基于大数据引擎的数据下载程序,对不同数据源开发下载程序,用于实现将不同数据源中的数据下载;在开发数据下载程序时,开发数据下载程序的任务执行规则,即根据开发的数据下载程序,控制要下载数据内容和控制数据描述的基本信息;数据下载程序从任务执行规则中获取数据源连接必要参数、数据临时存储文件系统连接必要参数和数据下载需求信息,数据下载程序执行后,统一将文件写入临时数据文件系统,得到下载结果数据文件,并连接配置项,其中,必要参数包括但不限于连接地址、用户名、密码或者认证密钥文件;其中,下载结果数据文件包括数据文件夹和json描述文件,所述的数据文件夹为一个或多个数据文件的压缩文件,所述的json描述文件包括但不限于结构化数据字段名、数据类型、数据字段说明、数据任务id、数据文件压缩文件个数、数据任务流水号;所述的json描述文件还包括数据文件夹中各个数据文件压缩文件的数据文件后缀、数据文件压缩格式、数据文件名、数据文件大小、数据文件中数据集条数;

[0089] S3、数据下载任务提交,根据数据下载程序中传入的数据源配置信息、数据下载临时存储路径、文件系统配置信息、数据下载规则调用大数据引擎或第三方服务执行;

[0090] S4、数据下载任务监控,使用大数据集群任务资源管理组件,对大数据集群中提交的数据下载任务监控,在数据下载服务中,调用任务资源组件接口,监控数据下载任务的执行资源情况,执行状态,执行结果,执行日志等监控项;

[0091] S5、数据下载服务,在数据下载服务中从数据文件临时存储系统目录下载结果数据到客户端。如提交的下载任务已经执行过,将数据结果文件夹中json描述文件中任务相关信息修改后,与数据文件一起下载到客户端;如数据下载任务未执行过,启动数据下载程序,监控任务执行完成后,从数据文件临时存储系统目录下下载数据到客户端。

[0092] 在上述实施方式中,通过配置配置项将不同数据源的统一管理,再通过不同的设计引擎开发数据源的数据下载程序,使数据下载程序能够下载不同的数据源里面的数据,再将数据下载程序导入的数据源配置信息、数据下载临时存储路径、文件系统配置信息、数据下载规则,使大数据引擎或第三方服务执行下载程序,从而将数据下载到客户端,从而实现海量数据的下载不受版本和数据源的影响,并提高数据下载效率。

[0093] 请参考附图3,本申请实施例还提供了一种大数据中台数据的下载装置,下载装置200包括至少一个可以软件或固件(Firmware)的形式存储于存储模块中或固化在操作系统(Operating System,OS)中的软件功能模块。例如下载装置200所包括的软件功能模块及计算机程序等。

[0094] 下载装置200可以包括数据源模块210、执行模块220、监控模块230和下载模块240,各单元具有的功能可以如下;

[0095] 数据源模块210,用于通过数据源配置项,实现增删改查功能和数据源统一管理;

[0096] 执行模块220,用于根据大数据引擎开发数据下载程序,并完成数据下载程序任务

提交；

[0097] 监控模块230,用于根据大数据任务监控组件实现对数据下载程序任务的监控；

[0098] 下载模块240,用于将数据文件从数据文件临时存储系统目录下载结果数据到客户端。

[0099] 通过数据源模块210,将通过设置配置项,不同的数据源统一管理,再通过执行模块220,用于存储开发了的数据下载程序,并能够将数据下载任务提交到数据源模块210,根据数据下载任务,能够在数据源中找到对应数据的位置和描述,再通过下载模块240,将数据下载到客户端,完成数据的快速下载,在数据下载的过程中,通过监控模块监控数据下载的进度以及完成情况等,有利于实时显示下载进度。

[0100] 可选的,数据源模块210,还可以用于,

[0101] 根据数据库、数据文件、接口和消息总线等数据源配置包括但不限于的连接地址、用户名、密码或认证密钥文件的等配置项,还可以添加配置文件,实现不同数据源的统一管理,且能够对不同数据源中的数据内容进行增删改查。

[0102] 可选的,执行模块220,还可以用于,

[0103] 开发基于不同大数据引擎的数据下载程序,数据下载程序的任务执行规则,根据任务执行规则获取数据源连接必要参数、数据临时存储文件系统连接必要参数和数据下载需求信息,数据下载执行后,统一将下载信息写入临时数据文件系统,得到下载结果数据文件,并连接配置项。

[0104] 可选的,执行模块220,还可以用于,

[0105] 包含数据文件夹和json描述文件,数据文件夹为一个或多个数据文件的压缩文件,json描述文件包括但不限于结构化数据字段名、数据类型、数据字段说明、数据任务id、数据文件压缩文件个数、数据任务流水号,json描述文件还包括数据文件夹中各个数据文件压缩文件的数据文件后缀、数据文件压缩格式、数据文件名、数据文件大小、数据文件中数据集条数。

[0106] 可选的,下载模块240,还可以用于,

[0107] 当提交的下载任务已经执行过时,将下载结果数据文件中的json描述文件中任务相关信息修改后,与数据文件一起下载到客户端；

[0108] 当数据下载任务未执行过,启动数据下载程序,监控任务执行完成后,从数据文件临时存储系统目录下下载数据到客户端。

[0109] 在本实施例中,存储模块可以是,但不限于,随机存取存储器,只读存储器,可编程只读存储器,可擦除可编程只读存储器,电可擦除可编程只读存储器等。在本实施例中,存储模块可以用于存储数据源模块210中的数据内容、执行模块220、监控模块230和下载模块240的工作状态等。当然,存储模块还可以用于存储程序,处理模块在接收到执行指令后,执行该程序。

[0110] 本申请实施例还提供一种计算机可读存储介质。计算机可读存储介质中存储有计算机程序,当计算机程序在计算机上运行时,使得计算机执行如上述实施例中所述的下载设计方法。

[0111] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到本申请可以通过硬件实现,也可以借助软件加必要的通用硬件平台的方式来实现,基于这样的理解,本申

请的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中,包括若干指令用以使得一台计算机设备(可以是个人计算机,制动设备,或者网络设备等)执行本申请各个实施场景所述的方法。

[0112] 综上所述,本申请实施例提供一种大数据中台数据的下载方法、装置及存储介质。在本方案中,通过配置项统一多数据源数据下载,大数据中台中统一的数据源管理,实现数据源管理增删改查,使数据源使用清楚明了,方便快捷;且开发基于大数据引擎针对不同的数据源下载程序,由统一的任务调度服务执行,由统一的任务监控平台监控,能够无下载数据量大小的瓶颈,能够下载处理海量数据,并提高下载效率,从而实现统一管理,增加开发迭代效率。

[0113] 在本申请所提供的实施例中,应该理解到,所揭露的装置和方法,也可以通过其它的方式实现。以上所描述的装置和方法实施例仅仅是示意性的,例如,附图中的流程图和框图显示了根据本申请的多个实施例的方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。另外,在本申请各个实施例中的各功能模块可以集成在一起形成一个独立的部分,也可以是各个模块单独存在,也可以两个或两个以上模块集成形成一个独立的部分。

[0114] 以上所述仅为本申请的实施例而已,并不用于限制本申请的保护范围,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

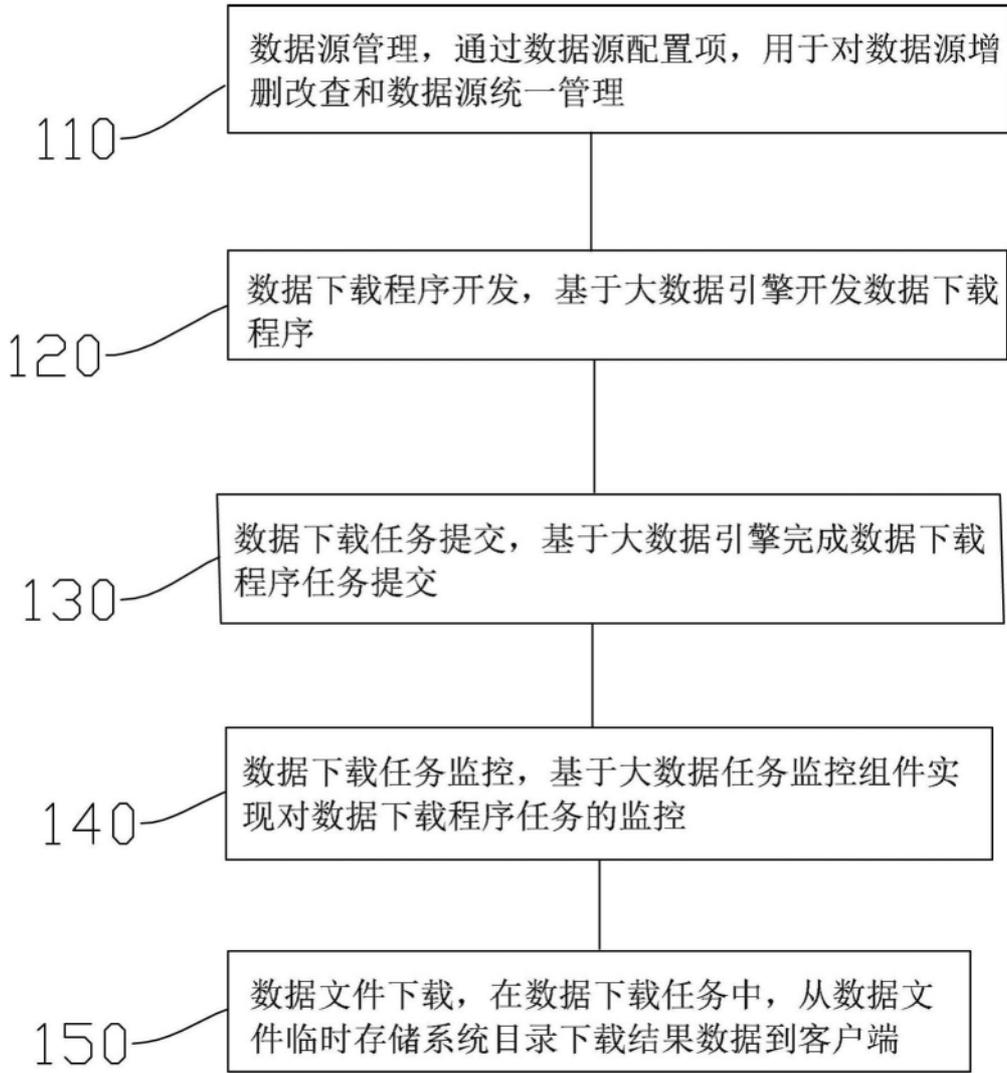


图1

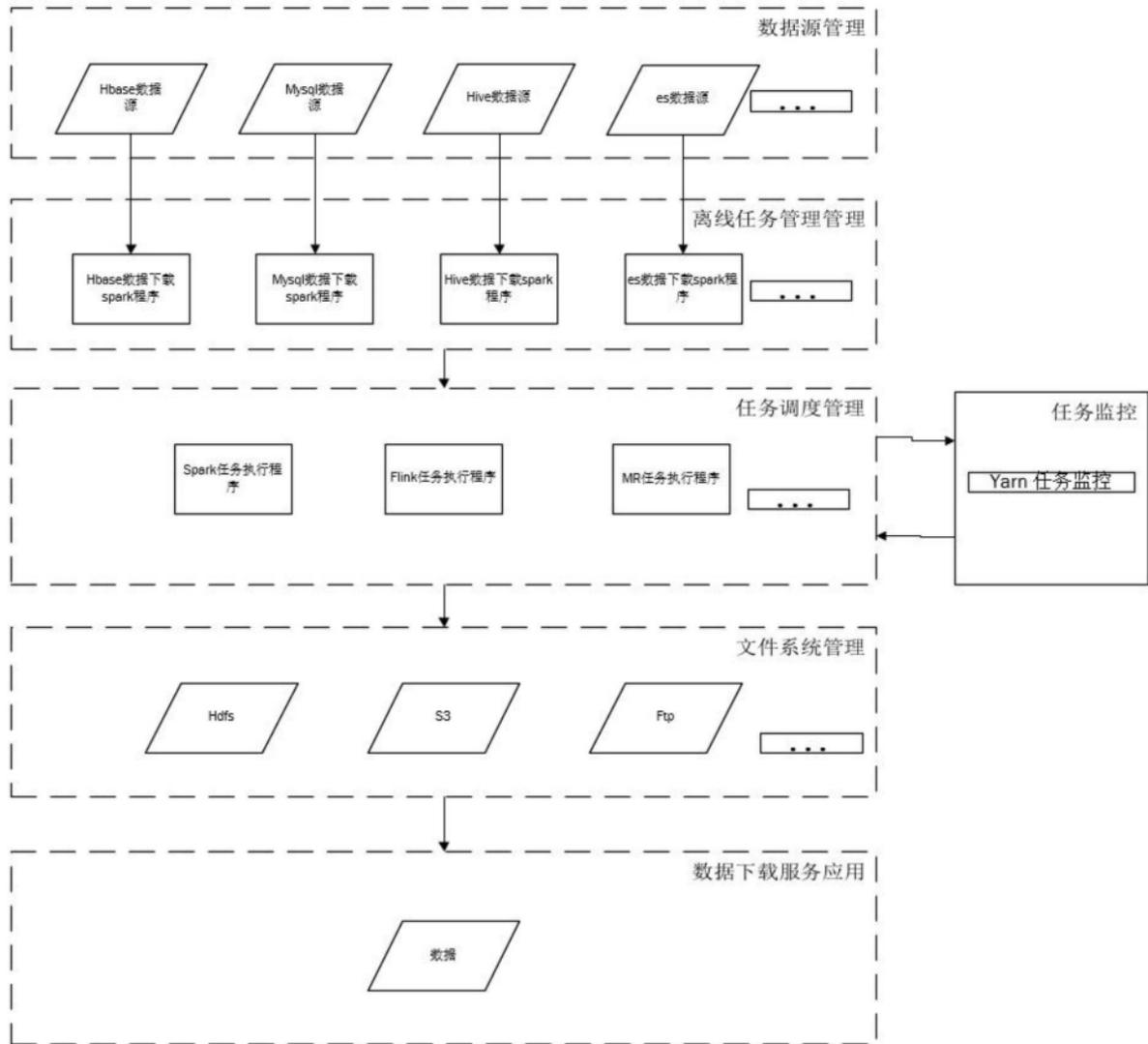


图2

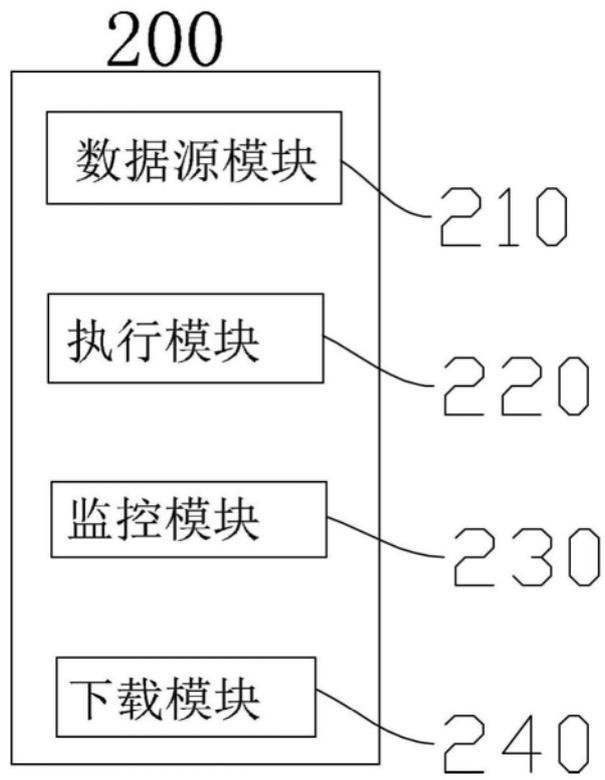


图3