(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
2022 年 11 月 3 日 (03.11.2022)

WIPO | PCT

(10) 国际公布号
WO 2022/227719 A1

(51) 国际专利分类号:
G06F 11/14 (2006.01) *G06F 11/20* (2006.01)

(21) 国际申请号: PCT/CN2022/072427

(22) 国际申请日: 2022 年 1 月 17 日 (17.01.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110454477.7 2021年4月26日 (26.04.2021) CN
202110961109.1 2021年8月20日 (20.08.2021) CN

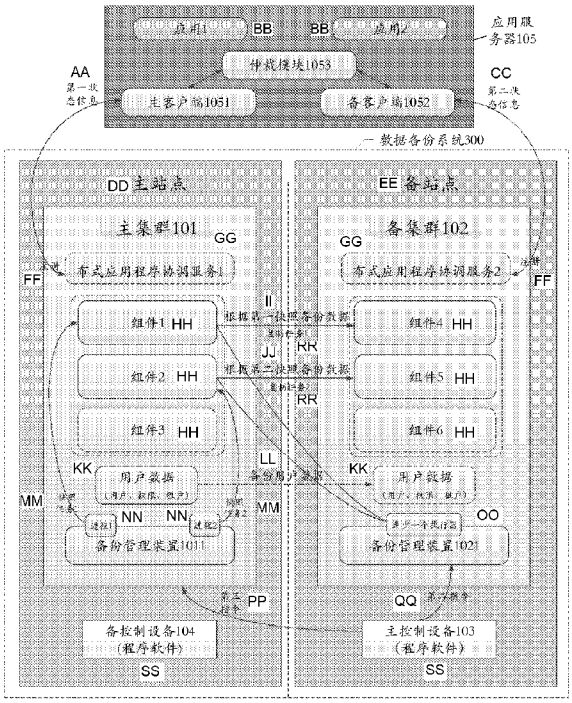
(71) 申请人: 华为云计算技术有限公司 (HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.) [CN/CN]; 中国贵州省贵安新区黔中大道交兴功路华为云数据中心, Guizhou 550025 (CN)。

(72) 发明人: 黄爽 (HUANG, Shuang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 巴图拉波罗门·雷迪 (BATTULA, Brahma Reddy); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 陈琪 (CHEN, Qi); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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, IT, JO, 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,

(54) **Title:** DATA BACKUP METHOD AND SYSTEM, AND RELATED DEVICE

(54) 发明名称: 数据备份方法、系统及相关设备



101 Master cluster
1011, 1021 Backup management apparatus
102 Standby cluster
103 Master control device
104 Standby control device
105 Application server
1051 Master client
1052 Standby client
1053 Arbitration module
300 Data backup system
AA First state information
BB Application
CC Second state information
DD Master site
EE Standby site
FF Register
GG Distributed application program coordination service
HH Assembly
II Back up data according to a first snapshot
JJ Back up data according to a second snapshot
KK User data (users, permissions and tenants)
LL Back up the user data
MM Snapshot task
NN Process
OO At least one actuator
PP Third instruction
QQ Second instruction
RR Replicate the task
SS Program software

图 5

(57) **Abstract:** Provided in the present application is a data backup method. The method is applied to a data backup system which comprises a master cluster, a standby cluster and a control device, wherein the control device controls, according to a first data backup policy, the master cluster or the standby cluster to back up, to the standby cluster, a plurality of data sets in the master cluster that are related to a first business at a first moment, with the first data backup policy comprising information of the plurality of data sets related to the first business, and the first moment. In this way, data backup can be realized between a master cluster and a standby

[见续页]

WO 2022/227719 A1

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, 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))。

cluster by taking a business as the granularity, and thus, a plurality of data sets, which are backed up to the standby cluster and are related to a first business, can be kept consistent in the time dimension, so that when the master cluster fails, the standby cluster can restore, on the basis of business data in the same time period, the running of a business, such that the quality of a data backup system providing a business service can be improved. In addition, further provided in the present application are a corresponding data backup system and a related device.

(57) 摘要：本申请提供了一种数据备份方法，该方法应用于包括主集群、备集群以及控制设备的数据备份系统，控制设备根据第一数据备份策略控制主集群或者备集群将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群，其中，第一数据备份策略包括与第一业务相关的多个数据集的信息和第一时刻。如此，可以实现在主集群以及备集群之间，以业务为粒度进行数据备份，从而备份至备集群上的与第一业务相关的多个数据集能够在时间维度上保持一致，以便备集群可以在主集群发生故障时能够基于同一时间段的业务数据恢复业务运行，进而可以提高数据备份系统提供业务服务的质量。此外，本申请还提供了对应的数据备份系统以及相关设备。

数据备份方法、系统及相关设备

技术领域

本申请涉及大数据技术领域，尤其涉及一种数据备份方法、系统及相关设备。

背景技术

随着大数据技术的发展，越来越多的用户（如企业等）逐步将业务数据迁移至大数据平台进行存储，相应的，用户对于大数据平台的容灾能力也越来越看重，即期望大数据平台在发生设备故障等灾难时能够保证用户存储于大数据平台的业务数据不会发生丢失。

目前，大数据平台通过数据备份系统实现对用户业务数据容灾，该数据备份系统包括主集群以及备集群。主集群可以利用组件处理用户的业务数据，如利用组件对该用户的业务数据进行封装和存储等，不同的组件可以处理同一用户的不同类型的业务数据。比如，主集群可以利用组件1存储用户的音视频、图片等业务数据，利用组件2存储该用户的表格形式的业务数据等。通常情况下，主集群可以周期性的将各个组件处理的业务数据备份至备站点，以便在主集群发生故障后，备集群能够基于备份的业务数据继续为用户提供业务服务。因此，在容灾切换后，如何尽可能避免备份至备集群的业务数据影响提供业务服务的质量，成为目前急需解决的问题。

发明内容

有鉴于此，本申请实施例提供了一种数据备份方法，该方法在数据备份时，以业务为粒度进行数据备份，以避免数据备份不一致导致的业务服务出错，保证了备份至备集群的业务数据不会影响提供业务服务的质量。本申请还提供了对应的数据备份系统、控制设备、计算设备、计算机可读存储介质以及计算机程序产品。

第一方面，本申请实施例提供一种数据备份方法，该方法应用于包括主集群、备集群以及控制设备的数据备份系统，具体实施时，控制设备根据第一数据备份策略控制主集群或者备集群将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群，其中，第一数据备份策略包括与第一业务相关的多个数据集的信息和第一时刻。

如此，可以实现在主集群以及备集群之间以业务为粒度进行数据备份，从而备份至备集群上的与第一业务相关的多个数据集能够在时间维度上保持一致。这样，当主集群发生故障时，备集群可以基于同一时间段的业务数据恢复该业务运行，从而可以避免数据备份系统因为备份至备集群上的业务数据在时间维度上不一致而发生提供业务服务出错的问题，进而可以提高数据备份系统为用户存储业务数据的可信度，提高业务服务质量。

在一种可能的实施方式中，控制设备根据第一数据备份策略控制主集群或者备集群将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群时，具体可以是控制设备向主集群发送第一指令，以指示主集群将与该第一业务相关的多个数据集在第一时刻的快照对应的数据发送至备集群。或者，控制设备向备集群发送第二指令，以指示备集群从主集群复制主集群中与第一业务相关的多个数据集在第一时刻的快照对应的数据。如此，控制设备可以通过向主集群或者向备集群发送指令的方式，控制主集群或者备集群根据快照实现数据备份过程。

在一种可能的实施方式中，控制设备向主集群发送第一指令或者向备集群发送第二指令之前，可以先向主集群发送第三指令，第三指令包括与第一业务相关的多个数据即的信息以

及第一时刻，用于指示主集群获取在第一时刻的与第一业务相关的多个数据集的快照。这样，后续主集群或者备集群可以根据该第一时刻对应的快照，实现将主集群中与第一业务相关的多个数据集备份至备集群，并且备份至备集群上的多个数据集均为第一时刻的主集群中与第一业务相关的多个数据集。通过采用对第一时刻的与第一业务相关的多个数据集进行快照的方式，可以更准确地获取并备份该第一时刻的数据集，避免了由于通信时延导致的数据备份时间不一致的问题。

在一种可能的实施方式中，控制设备还可以向主集群发送第四指令，该第四指令用于指示主集群将用户数据同步至备集群；或者，控制设备可以获取主集群以及备集群中存储的用户数据，并根据主集群中存储的用户数据对备集群中存储的用户数据进行调整，以使得主集群与备集群中存储的用户数据一致。这样，当备集群接管主集群上的业务时，可以根据备份至备集群的用户数据为用户提供相应的业务服务，从而无需由运维人员在备集群上进行人工配置用户数据。如此，不仅可以降低于运维人员的运维成本，而且，也能有效减小数据备份系统的恢复时间目标。

示例性地，上述用户数据例如可以是用户标识、用户权限、租户 标识中的至少一种，或者也可以是其它与用户相关的数据。

在一种可能的实施方式中，控制设备不仅可以为第一业务配置第一数据备份策略，还可以为第二业务配置第二数据备份策略，该第二数据备份策略包括与第二业务相关的多个数据集的信息和第二时刻；然后，控制设备根据第二数据备份策略控制主集群或备集群将第二时刻的主集群中与第二业务相关的多个数据集备份至所述备集群。其中，第二业务与第一业务属于不同业务，具体可以是属于同一用户的不同业务，或者可以是属于不同用户的不同业务等。如此，数据备份系统可以实现为多个不同的业务进行按照业务粒度的数据备份，以此实现支持多个业务的高质量服务。

在一种可能的实施方式中，与第一业务相关的多个数据集包括由主集群中的第一组件处理或存储的数据集和由主集群中的第二组件存储或者存储的数据集。示例性地，第一组件与第二组件可以用于对将封装成不同的格式，或者第一组件以及第二组件处理数据的性能存在差异等。如此，可以实现将第一时刻的主集群中由不同组件处理或者存储的属于同一业务的不同数据集备份至备集群。

另外，数据备份系统中备份至备集群的与第二业务相关的多个数据集，可以包括由主集群中的第一组件处理或存储的数据集、由主集群中的第二组件存储或者存储的数据集以及由主集群中的第三组件存储或处理的数据集等。处理或者存储不同业务的数据集的组件，可以存在差异。

在一种可能的实施方式中，控制设备包括主客户端以及备客户端，其中，主客户端用于检测主集群的第一状态信息，而备客户端用于检测备集群的第二状态信息，则控制设备还可以获取主客户端检测得到的第一状态信息以及备客户端检测得到的第二状态信息，并且，当第一状态信息指示主集群为备身份或者集群失效（如主集群因故障而失效等），且第二状态信息指示备集群为主身份时，控制设备确定备客户端为应用访问的客户端。如此，控制设备可以在主集群以及备集群的主备身份发生反转时，自动对访问集群的客户端进行切换，从而无需运维人员进行人工切换。

示例性地，在主集群故障之前，控制设备获取的第一状态信息可以指示主集群为主身份，而控制设备获取的第二状态信息可以指示备集群为备身份。

在一种可能的实施方式中，控制设备还可以向用户提示主集群故障的信息，以便用户基于该提示确定主集群发生故障，从而控制设备可以响应于用户针对该备集群的身份调整操作，将备集群的身份由备身份调整为主身份。如此，通过人工操作来调整主集群以及备集群进行身份反转，可以尽可能避免数据备份系统因为程序运行错误而出现主集群以及备集群的主备身份发生异常切换。

在一种可能的实施方式中，控制设备与主集群隔离部署，例如，控制设备可以与备集群共同部署于备站点，而主集群部署于主站点等。由于隔离部署了，当主集群故障时，控制设备不会发生故障，可以用于故障的切换。

在一种可能的实施方式中，控制设备、主集群以及备集群中设置有相同的时钟源，如此，控制设备控制主集群或备集群进行数据备份的时刻，与主集群或者备集群实际执行备份的时刻保持一致，以此避免因为时钟源不统一而导致数据备份出错的问题。提高了数据备份的时间一致性。

在一种可能的实施方式中，主集群和/或备集群包括基于hadoop架构构建的集群。

第二方面，本申请提供一种数据备份方法，该方法应用于数据备份系统，该数据备份系统包括主集群、备集群以及控制设备。具体实施时，主集群获取控制设备下发的指令，其中，该指令汇总包括与第一业务相关的多个数据集的信息和第一时刻，从而主集群根据该指令，将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群。如此，可以实现在主集群以及备集群之间，以业务为粒度进行数据备份，从而备份至备集群上的与第一业务相关的多个数据集能够在时间维度上保持一致。

在一种可能的实施方式中，主集群根据指令，将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群时，具体可以是根据与第一业务相关的多个数据集的信息和第一时刻，获取主集群中与第一业务相关的多个数据集在第一时刻的快照，从而主集群根据该快照，将该快照对应的数据（也即与第一业务相关的多个数据集）发送至备集群，以此实现将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群。

在一种可能的实施方式中，主集群还可以将用户数据备份至备集群。这样，当备集群接管主集群上的业务时，可以根据备份至备集群的用户数据为用户提供相应的业务服务，从而无需由运维人员在备集群上进行人工配置用户数据。如此，不仅可以降低于运维人员的运维成本，而且，也能有效减小数据备份系统的恢复时间目标。

示例性地，上述用户数据例如可以是用户标识、用户权限、租户 标识中的至少一种，或者也可以是其它与用户相关的数据。

在一种可能的实施方式中，主集群和/或备集群包括基于hadoop架构构建的集群。

第三方面，本申请提供了一种控制设备，该控制设备位于数据备份系统，数据备份系统还包括主集群以及备集群，控制设备包括：控制模块，用于根据第一数据备份策略控制主集群或备集群将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群，其中，第一数据备份策略包括与第一业务相关的多个数据集的信息和第一时刻。

在一种可能的实施方式中，控制模块，具体用于：向主集群发送第一指令，指示主集群将与第一业务相关的多个数据集在第一时刻的快照对应的数据发送至备集群，或者，向备集群发送第二指令，指示备集群从主集群复制主集群中与第一业务相关的多个数据集在第一时刻的快照对应的数据。

在一种可能的实施方式中，控制设备还包括：通信模块，用于向在控制设备向主集群发送第一指令，或者，控制设备向备集群发送第二指令之前，向主集群发送第三指令，第三指

令包括与第一业务相关的多个数据集的信息和第一时刻，第三指令用于指示主集群获取在第一时刻的与第一业务相关的多个数据集的快照。

在一种可能的实施方式中，控制设备还包括：通信模块，用于向主集群发送第四指令，第四指令指示主集群将用户数据同步至备集群；或者，控制模块，还用于获取主集群和备集群中存储的用户数据，并根据主集群中存储的用户数据对备集群中存储的用户数据进行调整。

在一种可能的实施方式中，控制设备还包括配置模块，用于根据用户输入的与第一业务相关的多个数据集的信息和第一时刻，为第一业务配置第一数据备份策略。

在一种可能的实施方式中，控制设备还包括配置模块，用于为第二业务配置第二数据备份策略，第二数据备份策略包括与第二业务相关的多个数据集的信息和第二时刻；控制模块，还用于根据第二数据备份策略控制主集群或备集群将第二时刻的主集群中与第二业务相关的多个数据集备份至备集群。

在一种可能的实施方式中，与第一业务相关的多个数据集包括由主集群中的第一组件处理或存储的数据集和由主集群中的第二组件处理或存储的数据集。

在一种可能的实施方式中，控制设备包括主客户端以及备客户端，主客户端用于检测主集群的第一状态信息，备客户端用于检测备集群的第二状态信息，控制设备还包括：通信模块，用于获取主客户端检测得到的第一状态信息以及备客户端检测得到的第二状态信息；确定模块，用于当第一状态信息指示主集群为备身份或集群失效，且第二状态信息指示备集群为主身份时，确定备客户端为应用访问的客户端。

在一种可能的实施方式中，控制设备还包括提示模块以及调整模块；提示模块，用于向用户提示主集群故障的信息；调整模块，用于响应于用户针对备集群的身份调整操作，将备集群的身份由备身份调整成主身份。

在一种可能的实施方式中，控制设备与主集群隔离部署。

在一种可能的实施方式中，控制设备、主集群和备集群中设置有相同的时钟源。

在一种可能的实施方式中，主集群和/或备集群包括基于hadoop架构构建的集群。

第四方面，本申请提供一种主集群，主集群位于数据备份系统，该数据备份系统还包括备集群以及控制设备，该主集群包括：通信模块，用于获取控制设备下发的指令，其中，指令中包括与第一业务相关的多个数据集的信息和第一时刻；备份模块，用于根据指令，将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群。

在一种可能的实施方式中，备份模块，具体用于根据与第一业务相关的多个数据集的信息和第一时刻，获取主集群中与第一业务相关的多个数据集在第一时刻的快照；根据快照，将快照对应的数据发送至备集群。

在一种可能的实施方式中，备份模块，还用于将用户数据同步至备集群。

在一种可能的实施方式中，主集群和/或备集群包括基于hadoop架构构建的集群。

第五方面，本申请提供一种数据备份系统，该数据备份系统包括控制设备、主集群以及备集群。其中，控制设备用于执行如第一方面或第一方面的任一种实现方式中的数据备份方法，主集群用于执行如第二方面或第二方面的任一种实现方式中的数据备份方法，备集群用于获取并存储从主集群备份的数据集。

第六方面，本申请提供一种控制设备，其中，控制设备包括处理器、存储器。所述处理器用于执行存储器中存储的指令，以使得所述控制设备执行如第一方面或第一方面的任一种实现方式中的数据备份方法。

第七方面，本申请提供一种主集群，其中，主集群包括至少一个处理器以及至少一个存

储器，所述至少一个处理器执行所述至少一个存储器中存储的指令，以使得所述主集群执行如第二方面或第二方面的任一种实现方式中的数据备份方法。

第八方面，本申请提供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指令，当其在计算设备上运行时，使得计算设备执行上述第一方面或第一方面的任一种实现方式所述的数据备份方法。

第九方面，本申请提供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指令，当其在至少一个计算设备上运行时，使得所述至少一个计算设备执行上述第二方面或第二方面的任一种实现方式所述的数据备份方法。

第十方面，本申请提供了一种包含指令的计算机程序产品，当其在计算设备上运行时，使得计算设备执行上述第一方面或第一方面的任一种实现方式所述的数据备份方法。

第十一方面，本申请提供了一种包含指令的计算机程序产品，当其在至少一个计算设备上运行时，使得所述至少一个计算设备执行上述第二方面或第二方面的任一种实现方式所述的数据备份方法。

本申请在上述各方面提供的实现方式的基础上，还可以进行进一步组合以提供更多实现方式。

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申请中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

图1为一种数据备份系统100的架构示意图；

图2为在数据备份系统100中备份业务数据的示意图；

图3为本申请实施例提供一种数据备份系统300的架构示意图；

图4为本申请实施例提供的在数据备份系统300中备份业务数据的示意图；

图5为本申请实施例提供的一种数据备份系统300的架构示意图；

图6为本申请实施例提供的一种集群配对界面的示意图；

图7为本申请实施例提供的一种策略配置界面的示意图；

图8为本申请实施例提供的在数据备份系统300中备份不同业务的数据集的示意图；

图9为本申请实施例提供的一种数据备份方法的流程示意图；

图10为本申请实施例提供的另一种数据备份方法的流程示意图；

图11为本申请实施例提供的一种控制设备的结构示意图；

图12为本申请实施例提供的一种主集群的结构示意图；

图13为本申请实施例提供的一种控制设备的硬件结构示意图；

图 14 为本申请实施例提供的一种主集群的硬件结构示意图。

具体实施方式

下面将结合本申请中的附图，对本申请提供的实施例中的方案进行描述。

本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换，这仅仅是描述本申请的实施例中对相同属性的对象在描述时所采用的区分方式。

参见图1，为一示例性的数据备份系统的架构示意图。如图1所示，数据备份系统100中包括主集群101以及备集群102。其中，主集群101以及备集群102可以通过至少一个设备（如服

务器、虚拟机、容器、存储设备等）实现。例如，主集群101以及备集群102均可以通过包括多个服务器的集群（如基于hadoop架构构建的集群等）实现。在部分场景中，主集群101以及备集群102也可以分别通过单个设备实现等。

实际应用场景中，用户可以通过应用服务器（具体可以是应用服务器中的客户端）访问主集群101，该访问包括对该用户的一个或者多个业务相关的数据进行读写等。为了描述简洁，以下将与一个业务相关的数据简称为业务数据。并且，主集群101可以周期性的将用户在该主集群101上保存的业务数据备份至备集群102。这样，当主集群101发生故障时，备集群102可以利用备份的业务数据继续为用户104提供相应的业务服务，以提高数据备份系统100为用户104存储业务数据的可靠性。

举例来说，假设主集群101以及备集群102基于hadoop架构进行构建，则主集群101可以包括hadoop分布式文件系统（hadoop distributed file system，HDFS）组件1011以及Hive组件1012，如图2所示。其中，HDFS组件用于对文件进行存储，将每个文件存储成一系列的数据块，可以提供高吞吐量的数据访问，适合应用于大规模数据集；Hive组件用于对数据进行提取、转化以及加载，以便在基于hadoop架构构建的集群中存储、查询数据。具体的，主集群101在存储多个用户的业务数据时，可以利用多个组件，将每个用户的业务数据按照不同的格式进行存储。其中，HDFS组件1011，用于将用户的音视频、图片等业务数据封装成文件格式并存储至主集群101，并且该文件格式的业务数据保存在相应的目录下，例如主集群101通过HDFS组件1011存储的属于用户106的业务数据位于目录3下（目录1以及目标2下的业务数据可以是其它用户的业务数据）。Hive组件1012，用于将用户的业务数据封装成结构化数据，并通过表格形式将其存储至主集群101，例如主集群101通过Hive组件1012存储的属于用户106业务数据为表格3（表格1以及表格2中的业务数据可以是其它用户的业务数据）。相应的，备集群102上可以包括HDFS组件1021、Hive组件1022，备集群102上的组件与主集群101上的组件的作用类似，在此不做赘述。当主集群101将其上的业务数据备份至备集群102时，相关技术中，在进行备份时，分别启动复制任务1以及复制任务2，并通过创建相应的进程执行这些复制任务，其中，复制任务1用于将与HDFS组件1011相关的数据（包括：目录1、目录2、目录3）备份至HDFS组件1021对应的存储区域中，复制任务2用于将与Hive组件1012相关的数据（包括：表格1、表格2、表格3）备份至Hive组件1022对应的存储区域中。

但是，上述以组件为粒度的数据备份，通常仅能使得主集群101与备集群102之间在单个组件上的业务数据保持一致，而备份至备集群102的多个组件中的数据在时间维度上可能并不一致。具体的，在备份业务数据的过程中，由于备份与HDFS组件1011相关的业务数据的过程以及备份与Hive组件1012相关的业务数据的过程相互独立，导致复制任务1以及复制任务2可能在不同时刻开始执行数据备份过程，这使得最终备份至HDFS组件1021对应的存储区域中的业务数据以及备份至Hive组件1022对应的存储区域中的业务数据，可能是主集群101存储的不同时间的用户104的业务数据。从而，当容灾切换后，备集群102基于不同时间的业务数据为用户提供业务服务，可能会导致备集群102提供的业务服务出错。本实施例中，与组件（如上述HDFS组件1011、Hive组件1022）相关的数据，例如可以是通过该组件进行封装等处理的数据，或者可以是通过该组件进行存储的数据等。

举例来说，当复制任务1在13:00:00时刻执行，而复制任务2在13:01:00时刻执行时，则最终备份至备集群102上HDFS组件1021对应的存储区域中的业务数据为多个用户（包括用户104）在13:00:00之前的业务数据，而备份至Hive组件1022对应的存储区域中的业务数据为多个用户（包括用户104）在13:01:00之前的业务数据。因此，如果在13:00:00~13:01:00期间，

主集群101通过HDFS组件1011和Hive组件1012保存了属于用户104的新的业务数据（如主集群101对已保存的部分业务数据进行了修改），则备份至HDFS组件1021中的业务数据缺失了13:00:00~13:01:00这一时间段内的新数据。这样，当主集群101发生故障时，用户104通过客户端103访问备集群102中备份的业务数据时，访问Hive组件1022可以获得最新的、正确的业务数据，而访问HDFS组件1021所获得的部分业务数据可能存在错误（为修改前的旧数据），从而可能导致数据备份系统100为用户104提供数据存储等业务服务时出错。比如，当数据备份系统100中存储有清理账单业务的数据时，用户104可能需要访问目录3下的账单明细数据以及表格3中的客户名单，并基于时间维度对目录3下的账单明细数据以及表格3中的客户名单进行比对校验，以便确定每个客户对应的账单。此时，若客户名单与账单明细数据属于不同时间段的数据，则可能存在部分客户已经结清的账单被认定为未结清，又或者账单明细数据中存在部分账单并不属于客户名单中的任意客户等错误，从而导致数据备份系统100存储的业务数据的可信度降低，也即降低了数据备份系统100提供数据存储服务的质量。

基于此，本申请实施例提供一种数据备份系统。参见图3，为本申请实施例提供的数据备份系统的架构示意图。如图3所示，在图1所示的数据备份系统100的基础上，图3所示的数据备份系统300中包括主集群101以及备集群102、主控制设备103、备控制设备104。并且，该数据备份系统300可以与应用服务器105连接。其中，主集群101以及备集群102可以通过至少一个设备（如服务器、虚拟机、容器、存储设备等）实现。例如，主集群101以及备集群102均可以通过包括多个服务器的集群（如基于hadoop架构的集群等）实现。在部分场景中，主集群101以及备集群102也可以分别通过单个设备实现等。主集群101用于为应用服务器105中部署的应用1以及应用2提供数据的存储和处理，并供应用服务器105进行数据的读写。图3中以应用服务器105包括两个应用进行示例，实际应用时，应用服务器105可以包括任意数量的应用。备集群102作为主集群101的备份集群，用于备份主集群101中的数据，在发生容灾切换后，由备集群102支持应用服务器105的数据读写。

图3示出的主控制设备103和备控制设备104用于控制主集群101和备集群102之间的数据备份。具体地，在主集群101发生故障之前，主控制设备103控制主集群101和备集群102之间进行数据备份；而在主集群101发生故障之后，备控制设备104控制备集群102与故障恢复后的主集群101（或者其它作为备集群102的冗余备份的集群）之间进行数据备份。示例性地，主控制设备103、备控制设备104具体可以是服务器、虚拟机或容器等。图3示出了主控制设备103与备集群102共同部署于备站点；备控制设备104与主集群101共同部署于主站点，其中，主站点和备站点可以是两个独立的区域（region）、机房或者是两个不同的局域网下的设备集群，通常主站点和备站点还具有独立的风、火、水、电控制系统。图3所示的主控制设备103和备控制设备104的部署方式仅是一种示例。另外，图3是以应用服务器105部署于数据备份系统300外部为例进行示例性说明，在其它可能的实现方式中，应用服务器105也可以部署于数据备份系统300中，此时，当主控制设备103、备控制设备104、以及应用服务器105为虚拟机或容器时，该主控制设备103以及备控制设备104还可以与应用服务器105部署于同一服务器，或者，主控制设备103可以与应用服务器105部署于同一服务器，或者，备控制设备104可以与应用服务器105部署于同一服务器。在部分实现场景中，数据备份系统300中还可以仅有一个控制设备，此时，该控制设备可以具有本申请实施例中主控制设备103以及备控制设备104的功能。应理解，图3所示的数据备份系统300仅作为一种示例性说明，本实施例对此并不进行限定。

实际部署时，数据备份系统300可以部署于云环境，例如数据备份系统300可以是基于云环境的多个区域构建等。或者，数据备份系统300也可以是部署于边缘环境，并通过边缘环境

的多个机房进行构建。应理解，本申请中的云环境表示云服务提供商设立的，用于为多个区域（region）（例如：华东区域、华北区域）的租户（上述用户106也即租户）提供服务（如数据存储服务等）的资源集合。云环境通常包括大量的资源，可以为各区域的租户提供基础资源服务和/或软件应用服务。本申请中的边缘环境表示针对特定地区的租户提供基础资源服务和/或软件应用服务的资源集合，边缘环境相比于云环境在物理位置上可以距租户更近，在提供一些业务时，可以更好地保证业务的低时延。

应理解，图3所示的数据备份系统300的架构仅作为一种示例性说明，实际应用时，数据备份系统300的架构也可以采用其它可能的实施方式。比如，数据备份系统300还可以包括其它设备，如包括用于对主集群101以及备集群102进行管理的设备等。又比如，数据备份系统300中的主集群101，除了可以包括图2所示的HDFS组件1011以及Hive组件1012之外，还可以包括更多数量的组件，如图2中主站点101还可以包括处理业务数据效率更高的SparkSQL组件，或者还可以包括轻量目录访问协议（Lightweight Directory Access Protocol，LDAP）组件或活动目录（active directory，AD）组件等。本实施例对于数据备份系统300的具体实现方式并不进行限定。

在图3所示的数据备份系统300中，主控制设备103在控制主集群101与备集群102之间备份业务数据时，具体是以业务为粒度进行数据的备份，以此实现备份至备集群102上的业务数据在时间维度保持一致。具体的，主控制设备103预先配置有针对业务的数据备份策略，该数据备份策略包括主集群101中存储的与该业务相关的多个数据集的信息以及第一时刻。示例性地，如图4所示，针对用户106的业务，与该业务相关的多个数据集中，部分数据集可以是图2中与HDFS组件1011相关的属于用户106的目录3下的业务数据的集合，而其余数据集可以是与Hive组件1012相关的属于用户106的表格3中的业务数据的集合。另外，数据备份策略中的第一时刻，用于指示备份至备集群102的业务数据为第一时刻的主集群101中已经存储的与该业务相关的数据。然后，当主集群101上该业务的数据需要被备份至备集群102时，主控制设备103可以根据为该业务配置的数据备份策略控制主集群101或备集群102，将第一时刻的主集群101中与该业务相关的多个数据集被备份至备集群102。例如，主控制设备103可以向主集群101发送指令，以指示主集群101或备集群102在第一时刻将与HDFS组件1011相关的目录3下的业务数据以及与Hive组件1012中相关的表格3中的业务数据备份至备集群102等。这样，以业务为粒度备份至备集群102上的数据，为在第一时刻主集群101中已经存储的与该业务相关的数据，从而备份的与该业务相关的数据能够在时间维度上保持一致。如此，当主集群101发生故障时，备集群102可以基于同一时间段的业务数据恢复该业务运行，从而可以尽可能避免数据备份系统300提供业务服务出错，进而可以提高数据备份系统300为用户存储业务数据的可信度，提高业务服务质量。

接下来，对数据备份系统的各种非限定性的具体实施方式进行详细描述。

在图3所示的数据备份系统300的基础上，参阅图5所示的数据备份系统300。主集群101可以包括一个或者多个用于处理或存储业务数据的组件，如图5所示，主集群101包括组件1、组件2以及组件3等，该组件1、组件2、组件3可以用于对一个或者多个用户的业务数据进行处理或存储。并且，实际应用时，任意一个或多个组件可以用于处理或存储同一个业务的不同数据，例如，组件1以及组件2可以将同一业务的不同数据封装成不同的数据格式并进行存储。当主集群101（以及备集群102）基于hadoop架构进行构建时，组件1例如可以是图2中的HDFS组件1011，用于将一个业务的部分业务数据封装成文件格式并进行存储；组件2例如可以是图

2中的Hive组件1012，用于将该业务的另一部分业务数据封装成结构化数据。或者，组件1以及组件2也可以是将同一业务的不同业务数据封装成相同格式进行存储。比如，组件1可以是图2中的Hive组件，而组件2可以是图2中的SparkSQL组件，并且，组件1以及组件2均可以将该业务的数据封装成结构化数据进行存储（其中，SparkSQL组件的数据读写性能通常高于Hive组件的数据读写性能）。本实施例中，组件1、组件2以及组件3可以是主集群101上用于封装并存储非流式数据的任意组件。并且，在处理或存储不同业务的数据时所使用的组件可以存在差异。比如，在处理或存储业务1的数据时，所使用的组件可以仅包括组件1；在处理或存储业务2的数据时，所使用的组件可以包括组件1以及组件2；在处理或存储业务3的数据时，所使用的组件可以包括组件1、组件2以及组件3。

备集群102中也可以包括一个或者多个组件，如图5所示，备集群102包括组件4、组件5以及组件6等。与主集群101类似。备集群102上的组件4、组件5以及组件6可以用于对一个或者多个用户的业务数据进行处理或存储，同一业务的不同数据可以被封装成相同或者不同的数据格式，并且，在处理或者存储不同业务的数据时所使用的组件可以存在差异等。其中，备集群102上的组件可以作为主集群101上的组件的备份。例如，备集群102上的组件4可以作为主集群101上的组件1的备份，组件5可以作为组件2的备份，组件6可以作为组件3的备份。相应的，备集群102在主集群101发生故障之前，周期性的对主集群101上的业务数据进行备份。在备份业务数据的过程中，主集群101上与各组件相关的业务数据可以被备份至备集群102上作为备份的组件中。比如，主集群101上与组件1相关的业务数据，可以被备份至集群102上的组件4对应的存储区域中；与组件2相关的业务数据，可以被备份至组件5对应的存储区域中；与组件3相关的业务数据，可以被备份至组件6对应的存储区域中。这样，当主集群101发生故障并由备集群102接管主集群101上的业务时，备集群102可以利用备份至组件4、组件5以及组件6分别对应的存储区域中的数据继续为用户提供读写业务数据的服务。

实际应用时，主集群101以及备集群102所在的云环境或者边缘环境包括的计算设备（如服务器等）较多，从而在云环境或者边缘环境可以构建多组主集群以及多组备集群。因此，可以预先在云环境或者边缘环境对主集群101以及备集群102进行配对，以构建图5所示的数据备份系统300。

作为一种实现示例，云环境或者边缘环境可以向管理员呈现如图6所示的集群配对界面，并且在该集群配对界面中不仅可以呈现有多组集群标识（如图6所示的集群1、集群2、集群3等），还可以呈现各集群的相关信息，如图6所示的集群的位置信息、资源规格信息等，从而管理员可以从集群配对界面呈现的多个集群中，选择相应的集群作为主集群101以及备集群102，以此构建数据备份系统300。实际应用时，云环境或者边缘环境也可以是根据用户针对主集群101以及备集群102的配对操作构建数据备份系统300等，本实施例对此并不进行限定。

在完成对于主集群101以及备集群102的配对后，主集群101以及备集群102之间可以进行数据通信，比如，备集群102可以周期性的将主集群101上的业务数据备份至备集群102。在进一步可能的实施方式中，主集群101以及备集群102之间在进行业务数据备份之前，可以预先完成通信认证过程，以使得主集群101与备集群102双方互信。示例性地，主集群101与备集群102之间可以基于传输控制协议/因特网互联协议（Transmission Control Protocol/Internet Protocol，TCP/IP）网络系统设计中可信的第三方认证协议进行通信认证，该第三方认证协议例如可以是Kerberos协议等。实际应用时，主集群101与备集群102之间可以自动执行通信认证过程，也可以在用户或者管理员的介入下完成双方通信认证等。如此，可以提高主集群101与备集群102之间进行数据通信的安全性以及可靠性。

在构建出数据备份系统300并且完成主集群101以及备集群102之间的通信认证后，主集群101以及备集群102之间可以以业务为粒度，将主集群101上的第一业务的数据备份至备集群102。下面以备份用户1的第一业务的数据为例进行示例性说明。

具体实现时，如图5所示，主控制设备103上可以运行用于配置数据备份策略和控制主备集群的程序软件，例如：该程序软件可以为用户1的第一业务配置第一数据备份策略，该第一数据备份策略中包括主集群101中与第一业务相关的多个数据集的信息以及第一时刻。其中，与第一业务相关的多个数据集，具体可以包括由组件1进行处理或存储的数据集以及由组件2处理或存储的数据集。本实施例中，与第一业务相关的多个数据集也可以称之为第一保护组。示例性地，第一保护组的信息具体可以是组件中的文件目录的标识，如图2中的目录3的名称等；或者可以是组件中的表格的标识，如图2中的表格3的名称等。第一时刻，用于指示备份至备集群102的业务数据为在该第一时刻主集群101中已经存储的第一业务相关的多个数据集。

在一种可能的实施方式中，主控制设备103可以基于用户1的操作实现为第一业务配置第一数据备份策略。例如，主控制设备103可以向用户1呈现如图7所示的策略配置界面，该策略配置界面中可以呈现有提示信息（如图7所示的“请指定备份的业务以及备份时刻”），以提示用户1输入所要配置的第一业务的标识（如第一业务的名称等）以及对该第一业务的数据进行备份的时刻。然后，主控制设备103可以根据用户1在控制策略配置界面上输入的第一业务的标识，提示是否为第一业务配置第一保护组。并且，在获知用户1确定为第一业务配置第一保护组后（如用户1在策略配置界面上点击“是”按钮），进一步呈现主集群101上存储的与第一业务相关的一个或者多个数据集的名称，例如图7所示的与组件1相关的数据集1以及与组件2相关的数据集2和数据集3的名称等，并提示用户1为第一业务配置第一保护组，也即提示用户1配置需要备份至备集群102上的与第一业务相关的一个或者多个数据集。这样，主控制设备103可以基于用户1对于数据集1至数据集3的选中操作或输入操作，确定第一业务的保护组包括数据集1、数据集2以及数据集3。同时，主控制设备103还可以获取用户1在策略配置界面中输入对第一业务的数据进行备份的第一时刻，从而主控制设备103基于该第一时刻以及上述第一保护组的信息生成第一业务对应的第一数据备份策略。实际应用时，备集群102可以周期性的针对主集群101上的第一业务的数据进行备份，此时，策略配置界面中可以同时呈现备份时刻的输入框以及备份周期的输入框，以便用户1在策略配置界面上针对备份时刻的输入框中输入第一时刻，也即输入该周期性备份的起始时刻。同时，用户1还可以在该策略配置界面上针对备份周期的输入框中输入备份周期，以便备集群102从第一时刻开始，按照用户1配置的备份周期，对主集群101上存储的第一业务的数据进行周期备份。相应的，在周期备份过程中的第二次备份（以及后续备份）时，主控制设备103可以根据第一时刻与备份周期，确定第二次备份的时刻，以便将在第二次备份的时刻的主集群中与第一业务相关的多个数据集备份至备集群102。

当然，上述示例仅作为示例性说明，并不用于限定主控制设备103配置第一数据备份策略的具体实现。比如，在其它示例中，用户1也可以是直接在策略配置界面中输入与第一业务相关的数据集的标识（如数据集的名称或者编号等），以指定第一业务的第一保护组；或者，主控制设备103也可以是自动将与第一业务相关的所有数据集确定为第一业务的第一保护组等。

在成功为第一业务配置第一数据备份策略后，主控制设备103可以根据该第一数据备份策略控制主集群101或备集群102，将第一时刻的主集群101中与第一业务相关的多个数据集备份

至备集群102。

作为一种实现示例，主控制设备103可以基于该数据备份策略生成第三指令，并将该第三指令发送给主集群101，以指示主集群101获取在第一时刻的与第一业务相关的多个数据集的快照。其中，第三指令中可以包括第一保护组的信息以及第一时刻，该第一时刻晚于主控制设备103向主集群101发送第三指令的时刻。相应的，主集群101在接收到该第三指令后，解析出第三指令中第一保护组的信息以及第一时刻，并根据该第一保护组的信息，确定主集群101中分别由组件1以及组件2处理或存储的与第一业务相关的多个数据集。然后，主集群101可以利用其上的备份管理装置1011创建进程1以及进程2，并启动包括第一时刻的快照任务1以及快照任务2。其中，进程1负责执行快照任务1，具体用于访问组件1，并在第一时刻对组件1处理或存储的与第一业务相关的数据集进行快照，得到第一快照；同时，进程2负责执行快照任务2，具体用于访问组件2，并在第一时刻对组件2处理或存储的与第一业务相关的数据集进行快照，得到第二快照。在其它实施例中，备份管理装置1011也可以利用主集群101上的至少一个执行器执行上述快照任务，其中，每个执行器例如可以是通过执行线程实现。

当完成业务数据的快照后，备份管理装置1011可以通知备集群102中的主控制设备103快照完成，从而主控制设备103可以向备集群102上的备份管理装置1021下发第二指令，以指示备集群102复制主集群101中与第一业务相关的多个数据集在第一时刻的快照所对应的数据。示例性地，该第二指令中可以包括上述第一快照以及第二快照的指示信息。备份管理装置1021可以根据该第二指令确定主集群101上的第一快照以及第二快照，并针对第一快照启动复制任务1以及针对第二快照启动复制任务2。在执行复制任务时，备集群102上可以包括至少一个执行器，该至少一个执行器可以执行复制任务1，具体可以是读取主集群101在第一时刻的第一快照，并根据该第一快照将由组件1处理或存储的与第一业务相关的数据集复制至备集群102，例如可以是将其写入备集群102上的组件4对应的存储区域中。并且，备集群102上的至少一个执行器可以执行复制任务2，具体可以是读取主集群101在第一时刻的第二快照，并根据该第二快照将由组件2处理或存储的与第一业务相关的数据集复制至备集群102，例如可以是将其写入备集群102上的组件5对应的存储区域中。如此，可以实现以业务为粒度将主集群101上属于第一业务的数据备份至备集群102。并且，备份至备集群102上的业务数据均为主集群101在第一时刻已经存储的第一业务的数据，从而对于同一业务的数据，备份后的业数据能够在时间维度上保持一致。

本实施例中，备集群102可以利用一个执行器，依次执行多个复制任务；或者，备集群102可以利用多个执行器，并行执行多个复制任务，提高数据备份效率。另外，主集群101上备份管理装置1011以及备集群102上的备份管理装置1021可以是运行在相应服务器上的应用程序。或者，备份管理装置1011以及备份管理装置1021可以是运行应用程序的硬件，例如：单独配置的处理器核、处理器以及服务器中的任意一种等，本实施例对于备份管理装置的具体实现方式并不进行限定。

上述实施方式中，是以备集群102主动备份业务数据为例进行示例性说明，实际应用时，也可以是由主集群101主动将与第一业务相关的数据集备份至备集群102。

作为一种实现示例，主控制设备103在确定主集群101完成对与组件1以及组件2相关的第一业务的多个数据集进行快照后，可以向主集群发送第一指令，以指示主集群101将与所述第一业务相关的所述多个数据集在第一时刻的快照所对应的数据发送至备集群102。示例性地，该第一指令中可以包括上述第一快照以及第二快照的指示信息。主集群101在接收到该第一指令后，可以根据该第一指令确定主集群101上的第一快照以及第二快照。然后，主集群101可

以通过一个或者多个执行器，根据第一快照，将由组件1处理或存储的与第一业务相关的数据集传输至备集群102；并通过该一个或者多个执行器，根据第二快照，将由组件2处理或存储的与第一业务相关的数据集传输至备集群102。其中，主集群101根据快照执行数据备份的具体实现过程，与前述备集群102根据快照执行数据备份过程的具体实现方式类似，可参照前述相关之处描述理解，在此不做赘述。

进一步的，在实现将与第一业务相关的多个数据集备份至备集群102后，备集群102还可以对备份的多个数据集进行快照，并对得到的快照进行存储。此时，备集群102针对备份的业务数据进行快照后所得到的快照，与主集群101在第一时刻对第一业务的数据进行快照所得到的快照通常一致。如此，备集群102可以在未来时刻利用该快照确定数据备份系统300在第一时刻所存储的第一业务的数据。

实际应用场景中，主集群101以及主控制设备103可能由不同的计算设备构建，从而主集群101与主控制设备103之间的时钟源可能并不统一。这样，主控制设备103在第三指令中所指示的第一时刻，与主集群101实际执行快照操作的时刻可能并不相同。比如，假设主控制设备103上的时钟比主集群101上的时钟快5秒，并且，主控制设备103在第三指令中所指示的第一时刻为13:00:00，则主集群101在基于该第一指令执行快照操作时，由于其时钟比主控制设备103慢5秒，这使得主集群101实际执行快照操作的时刻为主控制设备103的13:00:05，从而导致主控制设备103指示快照第一业务的数据的时刻与主集群101实际执行快照操作的时刻不一致。基于此，本实施例中主集群101以及主控制设备103之间可以设置相同的时钟源。例如，主控制设备103可以通过网络时间协议（Network Time Protocol，NTP）与主集群101进行时钟同步，以使得主控制设备103与主集群101之间具有相同的时钟源。当然，主集群101与主控制设备103也可以是通过其它方式实现时钟同步，本实施例对此并不进行限定。进一步地，主集群101与主控制设备103，还可以与备集群102进行时钟同步，以便实现在数据备份系统300中进行时钟统一。

实际应用场景中，主集群101不仅可以存储第一业务的数据，还可以存储其它业务的数据，并且不同业务的数据通常并不相同，不同的业务可以来自相同的用户或者不同的用户。以主集群101同时存储第一业务以及第二业务的数据为例，对于第一业务的数据，其可以是如图8所示的目录1下的业务数据以及表格1中的业务数据，而对于第二业务的数据，其可以是如图8所示的目录2下的业务数据、表格2中的业务数据以及表格3中的业务数据。因此，在以业务为粒度对业务数据进行快照和备份时，主控制设备103除了可以为第一业务配置第一数据备份策略之外，还可以为第二业务配置第二数据备份策略，该第二数据备份策略中包括主集群101中与第二业务相关的多个数据集的信息以及第二时刻。本实施例中，与第二业务相关的多个数据集也可以是称之为第二保护组，该第二保护组包括主集群101中由组件1处理或存储的与第二业务相关的数据集、由组件2处理或存储的与第二业务相关的数据集以及由组件3处理或存储的与第二业务相关的数据集（假设主集群101利用组件1、组件2以及组件3对第二业务的数据进行处理或存储）。这样，主控制设备103可以基于第二数据备份策略控制主集群101或备集群102将第二时刻的主集群中的第二保护组（也即与第二业务相关的多个数据集）备份至备集群102。其中，主控制设备103配置第二数据备份策略以及根据第二数据备份策略实现将主集群101上与第二业务相关的多个数据集备份至备集群102的具体实现过程，与上述主控制设备103配置第一数据备份策略以及根据第一数据备份策略实现将主集群101上与第一业务相关的多个数据集备份至备集群102的具体实现过程类似，可参照前述实施例的相关之处描述，在此不做赘述。

通常情况下，主集群101上的业务数据，可以被周期性的备份至备集群102。在一种实现方式中，每次备份至备集群102的业务数据，可以是主集群101上属于第一业务的所有业务数据。而在另一种实现方式中，主集群101以及备集群102之间进行第一次备份时，备份至备集群102上的业务数据可以是主集群101在第一时刻已经存储与第一业务相关的数据，而当主集群101以及备集群102之间进行第二次备份时，备份至备集群102上的业务数据可以是第一时刻至第三时刻期间，主集群101上的增量数据。其中，第三时刻为第二次对第一业务的数据进行备份的时刻。作为一种实现示例，主集群101在第一时刻完成对业务数据进行快照后，可以在第三时刻对由组件1存储或存储的与第一业务相关的数据集进行快照，得到第三快照，并同时在第三时刻对由组件2存储或存储的与第一业务相关的数据集进行快照，得到第四快照，该第三时刻晚于上述第一时刻。实际应用中，该第二时刻与第一时刻之间的间隔时长，例如可以是主集群101与备集群102之间的数据备份周期。或者，该第二时刻也可以是由主控制设备103进行指定等。主集群101在完成第二轮的快照操作后，可以通过备份管理装置1011将其通知给主控制设备103，并由主控制设备103指示备份管理装置1021执行第二轮的业务数据备份过程。与第一轮的业务数据备份过程类似，备份管理装置1021可以针对第三快照以及第四快照启动新的复制任务3以及复制任务4。然后，备集群102可以利用至少一个执行器执行复制任务3，具体可以是根据第一快照以及第三快照，确定主集群101上组件1在第一时刻至第三时刻这一时间段内处理或存储的与第一业务相关的增量数据（以下称之为第一增量数据），并将该第一增量数据备份至备集群102，例如可以是备份至组件4对应的存储区域中。并且，备集群102可以利用至少一个执行器执行复制任务4，具体可以是根据第二快照以及第四快照，确定组件2在第一时刻至第三时刻这一时间段内处理或存储的与第一业务相关的增量数据（以下称之为第二增量数据），并将该第二增量数据备份至备集群102，例如可以是备份至组件5对应的存储区域中。如此，后续每次备份业务数据时，主集群101以及备集群102之间可以仅传输增量数据，而可以无需将主集群101上的所有与第一业务相关业务数据均传输至备集群102，从而可以有效减少主集群101与备集群102之间的业务数据的传输量，这在提高备份效率的同时，也能减少备份业务数据所需的资源消耗。

上述是对主集群101与备集群102之间的数据备份过程进行介绍，下面介绍当主集群101发生故障时，数据备份系统300中的容灾切换过程。

在主集群101上的业务数据被备份至备集群102后，若主集群101发生故障，则主集群101可能难以继续为用户提供业务数据的读写服务，此时，备集群102可以对主集群101当前运行的业务进行接管，并利用预先备份的业务数据，继续为用户提供业务数据的读写服务，以此保证数据备份系统300对于存储用户的业务数据的可靠性。

通常情况下，用户可以通过应用服务器105上的客户端对主集群101或备集群102上存储的业务数据进行访问。示例性地，如图5所示，应用服务器105中包括主客户端1051以及备客户端1052。其中，在主集群101故障之前，用户可以通过应用服务器105上的主客户端1051访问主集群101，而在主集群101故障之后，用户可以通过应用服务器105上的备客户端1052访问备集群102。

实际应用时，应用服务器105可以在主集群101发生故障后，自动将访问集群的客户端由主客户端1051调整为备客户端1052。

首先，应用服务器105在运行时，可以通过主客户端1051以及备客户端1052感知发生故障的集群。作为一种示例，主客户端1051可以向主集群101上的分布式应用程序协调服务1（如zookeeper服务等）注册检测事件，并接收该分布式应用程序协调服务1反馈的检测结果，该检

测结果可以指示主集群101是否发生故障（如失效等）。另外，备客户端1052也可以向备集群102上的分布式应用程序协调服务2（如zookeeper服务等）注册检测事件，并接收该分布式应用程序协调服务2反馈的检测结果，该检测结果可以指示备集群102是否发生故障（如失效等）。比如，在备集群102接管主集群101故障之前所运行业务的过程中，应用服务器105可以通过备客户端1052确定备集群102是否发生故障，并在确定备集群102未发生故障后，指示备集群102接管主集群101在故障之前运行的业务。

然后，应用服务器105可以在感知到主集群101发生故障后，依据集群的状态信息切换访问集群的客户端。具体的，应用服务器105还可以包括仲裁模块1053，并可以从主客户端1051以及备客户端1052中分别获取指示主集群101以及备集群102身份的信息。其中，主客户端1051在向分布式应用程序协调服务1注册检测事件后，分布式应用程序协调服务1可以向主客户端1051反馈第一状态信息，该第一状态信息用于指示本集群为主身份还是备身份（也即指示本集群当前为主集群还是备集群），或者指示本集群是否发生失效（如因集群故障而失效等）。同时，分布式应用程序协调服务2可以向备客户端1052反馈第二状态信息，该第二状态信息用于指示为主身份还是备身份，或者指示本集群是否发生失效。然后，仲裁模块1053可以分别从主客户端1051中获取第一状态信息，从备客户端1052中获取第二状态信息。当第一状态信息用于指示主集群101为备身份或者集群失效，而第二状态信息用于指示备集群102为主身份时，仲裁模块1053可以确定应用服务器105访问集群的客户端切换为备客户端1052。值得注意的是，在主集群101故障之前，第一状态信息用于指示主集群101为主身份，第二状态信息用于指示备集群102为备身份。而在主集群101发生故障后，主控制设备103可以向用户或管理员提示主集群故障的信息，以便由用户或者管理员通过主控制设备103或者其它设备针对备集群的身份调整操作，将备集群102的身份由备身份调整为主身份。如此，可以尽可能避免数据备份系统300因为程序运行错误而出现主集群101以及备集群102的主备身份发生异常切换。进一步的，用户或者管理员还可以对主集群进行身份调整，具体可以是将主集群101的身份由主身份调整为备身份（或者集群失效），以此实现主集群101与备集群102之间的身份反转。此时，仲裁模块1053从主客户端1051中获取的第一状态信息用于指示主集群101为备身份或者集群失效，从备客户端1052中获取的第二状态信息用于指示备集群102为主身份。

如此，在主集群101发生故障后，应用服务器105能够将访问数据的集群由主集群101自动切换至备集群102，无需人工介入对应用服务器105所访问的集群进行切换，从而可以提高数据备份系统300的灵活性，降低人工运维成本。

值得注意的是，本实施例上述描述中，是以应用服务器105与主控制设备103独立部署，且主客户端1051和备客户端1052以及仲裁模块1053部署在应用服务器105为例进行示例性说明。在其它可能的数据备份系统300中，应用服务器105可以与主控制设备103集成部署，也即：由一个设备实现应用服务器105和主控制设备103的功能，该设备可以称为控制设备或称为应用服务器，即该控制设备或者应用服务器中可以集成有如图4所示的主客户端1051、备客户端1052以及仲裁模块1053，并由该控制设备或者应用服务器执行上述应用服务器105所执行的访问集群以及感知集群故障等操作。在另一些实施例中，当应用服务器105和主控制设备103独立部署时，上述应用服务器105中的主客户端1051、备客户端1052以及仲裁模块1053的功能也可以由控制设备103实现，也即控制设备103执行上述主备客户端自动切换的操作。由于主备客户端在控制设备103中，应用服务器105产生的数据的读写请求会发送至控制设备103中，由控制设备103中的当前为主身份的主客户端（或备客户端）执行对主集群（或备集群）中的数据的读写。

进一步的，当第一状态信息以及第二状态信息均指示各自集群为主集群时（如故障恢复后的主集群101在被主控制设备103管理之前，主集群101上的分布式应用程序协调服务1向主客户端1051指示本集群为主集群），应用服务器105仍然采用当前的访问策略对集群进行访问，即可以不对当前所访问的集群进行切换。

本实施例中，备集群102在接管主集群101故障之前所运行业务的过程中，不仅需要与这些业务相关的数据，还需要这些所属用户的数据，如用户名称、用户权限等。

示例性地，在应用服务器105将其访问的集群由主集群101切换至备集群102后，可以由运维人员在备集群102上配置相应的用户数据，以便备集群102基于该用户数据提供相应的数据读写服务。其中，所配置的用户数据例如可以包括用户标识、用户权限、用户所属租户的标识等数据中的至少一种。通常情况下，运维人员的人工配置过程，如在备集群102上执行创建用户、租户、申请权限的操作等，会增大数据备份系统300的恢复时间目标（recovery time object，RTO），也即增大了灾难（主集群101故障）发生后，数据备份系统300暂停业务至恢复业务之间的时长。

基于此，在另一种实现方式中，在主集群101故障之前，备集群102不仅可以将主集群101上的业务数据备份至备集群102，还可以将主集群101上的用户数据也备份至备集群102。具体实现时，主控制设备103可以向主集群101发送第四指令，以指示主集群101根据该第四指令将用户数据同步至备集群102。或者，主控制设备103可以获取主集群101以及备集群102中存储的用户数据，并根据主集群101中存储的用户数据对备集群102中存储的用户数据进行调整，以使得主集群101与备集群102中存储的用户数据一致，如根据主集群101中存储的用户数据在备集群102中添加未存储的用户数据，或者对备集群102中存储的用户数据进行修改等。这样，当备集群102接管主集群101上的业务时，可以根据备份至备集群102上的用户数据为用户提供相应的业务服务，从而无需由运维人员在备集群102上进行人工配置。如此，不仅可以降低于运维人员的运维成本，而且，也能有效减小数据备份系统300的恢复时间目标。

值得注意的是，上述实施方式中，是以主集群101发生故障为例进行示例性说明，实际应用场景中，当主集群101未发生故障时，若接收到用户或者管理员所指示的集群切换命令，则应用服务器105可以将访问的集群由主集群101自动切换至备集群102,，其实现集群切换的具体实现过程，与上述实现过程类似，可参照前述相关之处描述进行理解，在此不做赘述。

参阅图9，为本申请实施例中一种数据备份方法的流程示意图。该方法可以应用于图5所示的数据备份系统300。该方法以主集群和备集群为基于Hadoop架构构建的集群为例进行介绍，本实施例中，图5所示的数据备份系统300中的组件1以及组件4具体是HDFS组件，组件2以及组件5具体是Hive组件，而组件3以及组件6具体是SparkSQL组件。其中，备控制设备104可以在应用服务器105所访问的集群切换至备集群102时，将新存储至备集群102上的业务数据备份至主集群101（该主集群101例如可以是在故障后完成恢复等）。实际应用时，备控制设备104的功能与主控制设备103的功能类似，具体可以参考下述关于主控制设备103的功能描述。下面以备份T_0时刻的业务数据为例进行说明，图9所示的数据备份方法具体可以包括：

S901：主控制设备103将配置信息同步至备控制设备104。

本实施例中，在主集群101正常运行的过程中，主控制设备103可以控制实现将主集群101上的业务数据备份至备集群102。相应的，当主集群101故障后，备集群102可以接管主集群101上的业务，在此过程中，如果主集群101故障恢复，则备控制设备104可以控制实现将备集群102上的业务数据备份至（故障后恢复的）主集群101。为此，主控制设备103可以预先将自身

相关的配置信息同步至备控制设备104，以便于当主集群101发生故障后，备控制设备104能够控制实现相应的业务数据备份过程，从而无需由运维人员重复进行人工配置。其中，主控制设备103中的配置信息，可以在设备部署过程中由管理员配置在该主控制设备103中，以便主控制设备103可以根据配置的信息控制主集群101与备集群102之间进行数据备份。

示例性地，主控制设备103同步的配置信息，可以包括数据备份系统300的相关信息，如主集群101与备集群102之间的配对关系、数据备份系统300所包括的资源、数据备份系统300当前所备份的业务数据的时间点等信息。

应理解，上述S801是可选的，在一些实施例中，备控制设备104也可以主动获取上述配置信息，或许由管理员对备控制设备104进行配置。

S902：主控制设备103根据用户106的设置操作，为第一业务配置数据备份策略，该数据备份策略包括保护组的信息以及备份第一业务的数据的T_0时刻，其中，保护组的信息用于指示由HDFS组件、Hive组件以及SparkSQL组件存储的与第一业务相关的数据集。

实际应用时，用户106可以根据第一业务的数据在主集群101上存储时所生成的数据集，设置第一业务对应的保护组，该保护组的信息指示了主集群101中与该第一业务相关的多个数据集，以便于后续对第一业务的数据进行备份时，将保护组所指示的多个数据集备份至备集群102上。其中，用户106为第一业务创建保护组的具体实现过程，可参见前述实施例的相关之处描述，在此不做重述。

本实施例中，以保护组包括由HDFS组件、Hive组件以及SparkSQL组件存储的与第一业务相关的多个数据集为例，该保护组的信息例如可以是该多个数据集的标识。在其它可能的实施例中，用户106所设置的保护组也可以是仅包括由其中任意2个组件所存储的与第一业务相关的多个数据集，如用户106设置的保护组也可以是仅包括由HDFS组件以及Hive组件存储的与第一业务相关的多个数据集等。

另外，用户106还可以指示数据备份系统300对第一业务的数据进行备份的时刻为T_0时刻，从而后续数据备份系统300可以在该T_0时刻对第一业务的数据进行快照和备份。其中，保护组的信息与T_0时刻可以构成主控制设备103为第一业务配置的数据备份策略，其具体实现过程可参见前述实施例的相关之处描述。

S903：主控制设备103根据配置的数据备份策略，在T_0时刻之前向主集群101上备份管理装置1011发送第三指令，该第三指令中包括T_0时刻以及保护组的信息。

应理解，主控制设备103在T_0时刻之前，生成第三指令并将其发送给主集群101，以便主集群101能够在即将到来的T_0时刻对业务数据进行快照处理。

S904：备份管理装置1011在接收到第三指令后，创建多个进程，并利用该多个进程分别访问保护组的信息指示与第一业务相关的多个数据集所对应的HDFS组件、Hive组件以及SparkSQL组件。

S905：备份管理装置1011利用多个进程在T_0时刻对这些组件存储的与第一业务相关的多个数据集进行快照。

作为一种实现示例，备份管理装置1011可以创建进程1、进程2以及进程3。其中，进程1可以负责访问HDFS组件，并在T_0时刻对由HDFS组件存储的与第一业务相关的HDFS目录（也即与第一业务相关的数据集）进行快照，并且，HDFS组件存储的第一业务的数据通过文件形式保存在该HDFS目录中。进程2可以负责访问Hive组件，并可以在T_0时刻通过数据抽取命令从数据库（data base，DB）中获取由该Hive组件存储的业务数据的元数据，从而根据该元数据对指示第一业务的数据实际存储位置的HDFS目录进行快照。通常情况下，由Hive组件存储

的结构化数据，是通过文件格式保存在相应的HDFS目录中。进程3可以负责访问SparkSQL组件，并可以在T_0时刻通过数据抽取命令从数据库中获取由该SparkSQL组件存储的业务数据的元数据，从而根据该元数据对指示第一业务的数据实际存储位置的HDFS目录进行快照。并且，由SparkSQL组件存储的结构化数据，也是通过文件格式保存在相应的HDFS目录中。

在其它实施例中，备份管理装置1011也可以是利用至少一个执行器对各组件存储的业务数据进行快照等。其中，本实施例中所描述执行器的实现方式与前述实施例中执行器的实现方式类似。

S906：在完成数据快照后，主控制设备103向备集群102上的备份管理装置1021下发第二指令，以指示备集群102复制主集群101中的与第一业务相关的多个数据集在第一时刻的快照对应的数据。

实际应用时，主集群101上备份管理装置1011在利用多个进程完成对于通过HDFS组件、Hive组件以及SparkSQL组件存储的业务数据的快照后，可以向主控制设备103返回快照成功的通知，从而主控制设备103在确定快照结束后，通过下发第二指令指示备集群102将主集群101上的业务数据备份至备集群102。

示例性地，第二指令中可以包括主集群101中对由HDFS组件、Hive组件以及SparkSQL组件存储的与第一业务相关的数据集对应的快照的指示信息，以便于备份管理装置1021确定对哪些数据集进行备份。

S907：备份管理装置1021根据第一业务的保护组的信息启动多个复制任务，每个复制任务用于实现对一个组件存储的与第一业务相关的数据集进行备份。

S908：备份管理装置1021通过至少一个执行器执行该多个复制任务，根据主集群101上T_0时刻的快照，将由各个组件存储的与第一业务相关的数据集备份至备集群102。

作为一种示例，假设备集群102上运行有3个执行器，分别为执行器1、执行器2以及执行器3。其中，执行器1用于执行复制任务1，通过访问主集群101获得在T_0时刻针对HDFS组件存储的与第一业务相关的数据集（如HDFS目录）的快照，从而可以根据该快照将HDFS组件存储的与第一业务相关的数据集备份至备集群102上的HDFS组件对应的存储区域中。类似的，执行器2用于执行复制任务2，根据在T_0时刻针对Hive组件存储的与第一业务相关的数据集的快照，实现将通过Hive组件存储的与第一业务相关的数据集至备集群102上的Hive组件对应的存储区域中；执行器3用于执行复制任务3，根据在T_0时刻针对SparkSQL组件中与第一业务相关的数据集的快照，实现将通过SparkSQL组件存储的与第一业务相关的数据集备份至备集群102上的SparkSQL组件对应的存储区域中。

其中，在将Hive组件以及SparkSQL组件对应的第一业务的数据集备份至备集群102时，可以将该业务数据在备集群102上的元数据保存至备集群102上的数据库中，以便于后续在备集群102上根据数据库中的元数据查询该第一业务的数据。

此时，备份至备集群102上的业务数据，均为主集群101在T_0时刻所存储的第一业务相关的多个数据集，从而实现备集群102上的第一业务的数据在时间维度保持一致。

S909：备份管理装置1021通过至少一个执行器对备份至备集群102上的第一业务的数据进行快照。

在将业务数据由主集群101备份至备集群102后，备份管理装置1021还可以利用该执行器对备份的第一业务的数据进行快照。此时，在备集群102上的快照，与主集群101在T_0时刻的业务数据一致。

值得注意的是，本实施例中，是以备集群102主动从主集群101上备份业务数据为例进行

示例性说明，实际应用时，也可以是由主集群101主动将业务数据备份至备集群102上，例如，主控制设备103可以向主集群101发送第一指令，以指示主集群101将由HDFS组件、Hive组件以及SparkSQL组件存储的与第一业务相关的数据集备份至备集群102。此时，主集群101可以在完成快照后，可以通过相应的执行器，根据T_0时刻的快照将第一业务的数据备份至备集群102。

进一步的，本实施例中，不仅主集群101上第一业务的数据可以被备份至备集群102，主集群101上的用户数据也可以被备份至备集群102。例如，主控制设备103可以向所述主集群101发送第四指令，以便基于该第四指令指示主集群101将用户数据同步至备集群102。或者，主控制设备103可以获取主集群101和备集群102中存储的用户数据，并根据主集群101中存储的用户数据对备集群102中存储的用户数据进行调整，以使得主集群101与备集群102中存储的用户数据一致。这样，当备集群102接管主集群101上的业务时，可以根据备份至备集群102上的用户数据为用户提供相应的业务服务，从而无需由运维人员在备集群102上进行人工配置。如此，不仅可以降低于运维人员的运维成本，而且，也能有效减小数据备份系统300的恢复时间目标。为此，本实施例还可以包括下述步骤S910:

S910: 备集群102将主集群101上的用户数据备份至备集群102；或者，主集群101主动将用户数据备份至备集群102；或者，主控制设备103根据主集群101中存储的用户数据对备集群102中存储的用户数据进行调整。

示例性地，主集群101上的用户数据，例如可以包括在主集群101上创建的用户（包括用户106）的标识、租户的标识以及为用户申请的权限中的至少一种。

在主集群101上的业务数据被备份至备集群102后，若主集群101发生故障，则备集群102可以对主集群101当前运行的业务进行接管，并利用预先备份的业务数据，继续为用户提供业务数据的读写服务，以此保证数据备份系统300对于存储用户的业务数据的可靠性。

值得注意的是，为了描述简洁，本实施例中重点是对步骤S902至步骤S910的执行操作进行介绍，针对各个步骤的具体实现方式，具体可以参见前述实施例中的相关之处描述，本实施例中对此不在进行赘述。并且，本实施例中，用户106可以通过应用服务器105，实现对于主集群101或者备集群102的访问，具体可以是在主集群101故障之前，用户通过应用服务器105上的主客户端1051访问主集群101，而在主集群101故障之后，应用服务器105可以自动切换访问集群的客户端，从而用户通过应用服务器105上的备客户端1052访问备集群102。其具体实现过程，可以参见前述实施例的相关之处描述，在此不再重复赘述。

上述实施例中，是以在T_0时刻对主集群101上的业务数据进行备份为例进行示例性说明，实际应用时，主集群101与备集群102之间可以周期性的进行数据备份。比如，用户106在策略配置界面上配置起始备份的时刻为T_0时刻的同时，还配置了主集群101与备集群102之间的备份周期，从而在执行第一次数据备份后，经过备份周期的时长时执行第二次数据备份过程。因此，上述实施例中的T_0时刻也即周期备份的起始时刻。此时，备集群102每次可以按照上述实施例所描述的类似过程将主集群101上的所有业务数据均备份至备集群102。而在其它实施例中，备集群102在第二次以及后续的备份过程中，可以仅将主集群101上的增量数据备份至备集群102。下面，以主集群101与备集群102之间进行第二轮数据备份为例进行说明，其中，第二轮备份的业务数据为主集群101在T_0~T_1时间段内新增至主集群101上的业务数据（以下称之为增量数据）。参见图10，示出了本申请实施例中又一种数据备份方法的流程示意图，该方法具体可以包括:

S1001：主控制设备103在T_1时刻之前，向主集群101上备份管理装置1011发送第五指令，该第五指令中包括T_1时刻以及第一业务的保护组信息。

S1002：备份管理装置1011在接收到第五指令后，创建多个进程（或者沿用第一轮数据备份时已创建的多个进程），并利用该多个进程分别访问保护组的信息指示的数据集所对应的HDFS组件、Hive组件以及SparkSQL组件。

S1003：备份管理装置1011利用多个进程在T_1时刻对这些组件存储的与第一业务相关的数据集进行快照。

作为一种实现示例，备份管理装置1011所创建的进程1，可以负责访问HDFS组件，并在T_1时刻对由HDFS组件存储的与第一业务相关的HDFS目录下的数据集进行快照。其中，HDFS组件存储的第一业务的数据通过文件格式保存在该HDFS目录中。进程2可以负责访问Hive组件，并可以在T_1时刻通过数据抽取命令从数据库中获取该Hive组件存储的与第一业务相关数据集的元数据，从而根据该元数据对指示该数据集实际存储位置的HDFS目录进行快照。其中，Hive组件存储的结构化数据，是通过文件格式保存在相应的HDFS目录中。进程3可以负责访问SparkSQL组件，并可以在T_1时刻通过数据抽取命令从数据库中获取通过该SparkSQL组件存储的与第一业务相关数据集的元数据，从而根据该元数据对指示该数据集的实际存储位置的HDFS目录进行快照。其中，SparkSQL组件存储的结构化数据，也是通过文件格式保存在相应的HDFS目录中。

S1004：在完成数据快照后，主控制设备103向备集群102上的备份管理装置1021下发第六指令，以指示备集群102将T_1时刻的主集群101中的与第一业务的数据集备份至备集群102。

S1005：备份管理装置1021根据第一业务的保护组的信息启动多个复制任务，每个复制任务用于实现对一个组件存储的与第一业务相关的数据集进行备份。

S1006：备份管理装置1021通过至少一个执行器执行该多个复制任务，根据主集群101上T_0时刻的快照以及T_1时刻的快照，将各个组件存储的第一业务的增量数据备份至备集群102。

作为一种示例，假设备集群102上运行有3个执行器，分别为执行器1、执行器2以及执行器3。其中，执行器1用于执行复制任务1，通过访问主集群101获得在T_0时刻针对HDFS组件的第一业务对应的快照以及在T_1时刻针对HDFS组件的第一业务对应的快照，从而可以根据T_0时刻以及T_1时刻的快照，确定HDFS组件在T_0~T_1时间段内存储的第一业务的增量数据，并将该增量数据备份至备集群102上的HDFS组件对应的存储区域中。类似的，执行器2用于执行复制任务2，根据在T_0时刻以及T_1时刻针对Hive组件的第一业务对应的快照，确定Hive组件在T_0~T_1时间段内存储的第一业务的增量数据，并将该增量数据备份至备集群102上的Hive组件对应的存储区域中；执行器3用于执行复制任务3，根据在T_0时刻以及T_1时刻针对SparkSQL组件的第一业务对应的快照，确定SparkSQL组件在T_0~T_1时间段内存储的第一业务的增量数据，并将该增量数据备份至备集群102上的SparkSQL组件对应的存储区域中。

其中，在将Hive组件以及SparkSQL组件对应的增量数据备份至备集群102时，可以将该增量数据在备集群102上的元数据保存至备集群102上的数据库中，以便于后续在备集群102上根据数据库中的元数据查询相应的第一业务的数据。

此时，备份至备集群102上的业务数据，为主集群101在T_0时刻的第一业务的数据以及在T_0~T_1时间段内该第一业务新增的业务数据，这也就是主集群101上在T_1时刻的业务数据。

S1007：备份管理装置1021通过至少一个执行器对备份至备集群102上的第一业务的增量数据进行快照。

如此，后续每次备份业务数据时，主集群101以及备集群102之间可以仅传输第一业务的

增量数据，而可以无需将主集群101上的所有业务数据均传输至备集群102，从而可以有效减少主集群101与备集群102之间的业务数据的传输量，这在提高备份效率的同时，也能减少备份业务数据所需的资源消耗。

值得注意的是，为了描述简洁，本实施例中重点是对步骤S1001至步骤S1007的执行操作进行介绍，针对各个步骤的具体实现方式，具体可以参见前述实施例中的相关之处描述，本实施例中对此不在进行赘述。

以上结合图1至图10对本申请实施例提供的数据备份系统以及数据备份方法进行介绍，接下来结合附图对本申请实施例提供的用于执行上述数据备份方法的设备进行介绍。

参见图11，示出了本申请实施例提供的一种控制设备的结构示意图。其中，图11所示的控制设备1100可以用于实现上述各实施例中主控制设备103所执行的数据备份方法，图11所述的控制设备1100位于数据备份系统，如上述图5所示的数据备份系统300等，该数据备份系统还包括主集群以及备集群，该控制设备1100包括:

控制模块1101，用于根据第一数据备份策略控制所述主集群或所述备集群将第一时刻的所述主集群中与第一业务相关的多个数据集备份至所述备集群，其中，所述第一数据备份策略包括与所述第一业务相关的所述多个数据集的信息和所述第一时刻。

在一种可能的实施方式中，所述控制模块1101，具体用于:

向所述主集群发送第一指令，指示所述主集群将与所述第一业务相关的所述多个数据集在所述第一时刻的快照对应的数据发送至所述备集群，或者，向所述备集群发送第二指令，指示所述备集群从所述主集群复制所述主集群中与所述第一业务相关的所述多个数据集在所述第一时刻的快照对应的数据。

在一种可能的实施方式中，所述控制设备1100还包括:

通信模块1102，用于向在所述控制设备向所述主集群发送第一指令，或者，所述控制设备向所述备集群发送所述第二指令之前，向所述主集群发送第三指令，所述第三指令包括与所述第一业务相关的所述多个数据集的信息和所述第一时刻，所述第三指令用于指示所述主集群获取在所述第一时刻的与所述第一业务相关的所述多个数据集的快照。

在一种可能的实施方式中，所述控制设备1100还包括:

通信模块1102，用于向所述主集群发送第四指令，所述第四指令指示所述主集群将用户数据同步至所述备集群;

或者，所述控制模块1101，还用于获取所述主集群和所述备集群中存储的用户数据，并根据所述主集群中存储的用户数据对所述备集群中存储的用户数据进行调整。

在一种可能的实施方式中，所述控制设备1100还包括配置模块1103，用于根据用户输入的与所述第一业务相关的多个数据集的信息和所述第一时刻，为所述第一业务配置所述第一数据备份策略。

在一种可能的实施方式中，所述控制设备1100还包括配置模块1103，用于为第二业务配置第二数据备份策略，所述第二数据备份策略包括与所述第二业务相关的多个数据集的信息和第二时刻;

所述控制模块1101，还用于根据所述第二数据备份策略控制所述主集群或所述备集群将所述第二时刻的所述主集群中与所述第二业务相关的所述多个数据集备份至所述备集群。

在一种可能的实施方式中，与所述第一业务相关的所述多个数据集包括由所述主集群中的第一组件处理或存储的数据集和由所述主集群中的第二组件处理或存储的数据集。

在一种可能的实施方式中，所述控制设备1100包括主客户端以及备客户端，所述主客户端用于检测所述主集群的第一状态信息，所述备客户端用于检测所述备集群的第二状态信息，所述控制设备1100还包括：

通信模块1102，用于获取所述主客户端检测得到的第一状态信息以及所述备客户端检测得到的第二状态信息；

确定模块1104，用于当所述第一状态信息指示所述主集群为备身份或集群失效，且所述第二状态信息指示所述备集群为主身份时，确定所述备客户端为应用访问的客户端。

在一种可能的实施方式中，所述控制设备1100还包括提示模块1105以及调整模块1106；

所述提示模块1105，用于向用户提示所述主集群故障的信息；

所述调整模块1106，用于响应于所述用户针对所述备集群的身份调整操作，将所述备集群的身份由备身份调整成主身份。

在一种可能的实施方式中，所述控制设备1100与所述主集群隔离部署。

在一种可能的实施方式中，所述控制设备1100、所述主集群和所述备集群中设置有相同的时钟源。

在一种可能的实施方式中，所述主集群和/或所述备集群包括基于hadoop架构构建的集群。

根据本申请实施例的控制设备1100可对应于执行本申请实施例中描述的方法，并且控制设备1100的各个模块的上述和其它操作和/或功能分别为了实现上述实施例中主控制设备103所执行的相应流程，为了简洁，在此不再赘述。

参见图12，示出了本申请实施例提供的一种主集群的结构示意图。其中，图12所示的主集群1200可以用于实现上述各实施例中主集群101所执行的数据备份方法，图12所述的主集群1200位于数据备份系统，如上述图5所示的数据备份系统300等，该数据备份系统还包括备集群以及控制设备，该主集群1200包括：

通信模块1201，用于获取控制设备下发的指令，其中，指令中包括与第一业务相关的多个数据集的信息和第一时刻；

备份模块1202，用于根据指令，将第一时刻的主集群中与第一业务相关的多个数据集备份至备集群。

在一种可能的实施方式中，备份模块1202，具体用于：

根据与第一业务相关的多个数据集的信息和第一时刻，获取主集群中与第一业务相关的多个数据集在第一时刻的快照；

根据快照，将快照对应的数据发送至备集群。

在一种可能的实施方式中，备份模块1202，还用于将用户数据同步至备集群。

在一种可能的实施方式中，主集群和/或备集群包括基于hadoop架构构建的集群。

根据本申请实施例的主集群1200可对应于执行本申请实施例中描述的方法，并且主集群1200的各个模块的上述和其它操作和/或功能分别为了实现上述实施例中主集群101所执行的相应流程，为了简洁，在此不再赘述。

图13提供了一种控制设备。如图13所示，控制设备1300具体可以用于实现上述图11所示的控制设备1100的功能。

控制设备1300包括总线1301、处理器1302和存储器1303。处理器1302、存储器1303之间

通过总线1301通信。

总线1301可以是外设部件互连标准（peripheral component interconnect，PCI）总线或扩展工业标准结构（extended industry standard architecture，EISA）总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图13中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。

处理器1302可以为中央处理器(central processing unit，CPU)、图形处理器（graphics processing unit，GPU）、微处理器（micro processor，MP）或者数字信号处理器（digital signal processor，DSP）、神经网络处理器（neural network processing unit, NPU）等处理器中的任意一种或多种。

存储器1303可以包括易失性存储器(volatile memory)，例如随机存取存储器(random access memory，RAM)。存储器1303还可以包括非易失性存储器(non-volatile memory)，例如只读存储器(read-only memory，ROM)，快闪存储器，机械硬盘（hard drive drive，HDD）或固态硬盘（solid state drive，SSD）。

存储器1303中存储有可执行的程序代码，处理器1302执行该可执行的程序代码以执行前述实施例中主控制设备103所执行的数据备份方法。

图14提供了一种主集群。如图14所示，主集群1400具体可以用于实现上述图12所示的主集群1200的功能。

主集群1400包括至少一个处理器以及至少一个存储器，其中，该至少一个处理器与至少一个存储器可以位于一个或者多个计算设备中。示例性地，本实施例中以至少一个存储器以及至少一个存储器位于多个计算设备为例进行说明。其中，每个计算设备可以包括总线1401、处理器1402和存储器1403。处理器1402、存储器1403之间通过总线1401通信。

总线1401可以是PCI总线或EISA总线等。总线可以分为地址总线、数据总线、控制总线等。为便于表示，图14中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。

处理器1402可以为CPU、GPU、MP或者DSP、NPU等处理器中的任意一种或多种。

存储器1403可以包括易失性存储器(volatile memory)，例如RAM。存储器1403还可以包括非易失性存储器(non-volatile memory)，例如ROM，快闪存储器，HDD或SSD。

每个计算设备中的存储器1403可以存储有可执行的程序代码，并且，各个计算设备中的处理器1402执行该可执行的程序代码后，使得主集群1400执行前述实施例中主集群101所执行的数据备份方法。

本申请实施例还提供了一种计算机可读存储介质。所述计算机可读存储介质可以是计算设备能够存储的任何可用介质或者是包含一个或多个可用介质的数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘）等。该计算机可读存储介质包括指令，所述指令指示计算设备执行上述主控制设备103或主集群101所执行的数据备份方法。

本申请实施例还提供了一种计算机程序产品。所述计算机程序产品包括一个或多个计算机指令。在计算设备上加载和执行所述计算机指令时，全部或部分地产生按照本申请实施例所述的流程或功能。

所述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计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机或数据中心通过有线（例如同轴电缆、光纤、数字用户线（DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机或数据中心进行传输。

所述计算机程序产品可以为一个软件安装包，在需要使用前述对象识别方法的任一方法的情况下，可以下载该计算机程序产品并在计算设备上执行该计算机程序产品。

上述各个附图对应的流程或结构的描述各有侧重，某个流程或结构中没有详述的部分，可以参见其他流程或结构的相关描述。

权 利 要 求 书

1.一种数据备份方法，其特征在于，所述方法应用于数据备份系统，所述数据备份系统包括主集群、备集群以及控制设备，所述方法包括：

所述控制设备根据第一数据备份策略控制所述主集群或所述备集群将第一时刻的所述主集群中与第一业务相关的多个数据集备份至所述备集群，其中，所述第一数据备份策略包括与所述第一业务相关的所述多个数据集的信息和所述第一时刻。

2.根据权利要求 1 所述的方法，其特征在于，所述控制设备根据第一数据备份策略控制所述主集群或所述备集群将第一时刻的所述主集群中与第一业务相关的多个数据集备份至所述备集群，包括：

所述控制设备向所述主集群发送第一指令，指示所述主集群将与所述第一业务相关的所述多个数据集在所述第一时刻的快照对应的数据发送至所述备集群；或者，

所述控制设备向所述备集群发送第二指令，指示所述备集群从所述主集群复制所述主集群中与所述第一业务相关的所述多个数据集在所述第一时刻的快照对应的数据。

3.根据权利要求 2 所述的方法，其特征在于，在所述控制设备向所述主集群发送第一指令，或者，所述控制设备向所述备集群发送所述第二指令之前，所述方法还包括：

所述控制设备向所述主集群发送第三指令，所述第三指令包括与所述第一业务相关的所述多个数据集的信息和所述第一时刻，所述第三指令用于指示所述主集群获取在所述第一时刻的与所述第一业务相关的所述多个数据集的快照。

4.根据权利要求 1-3 任一项所述的方法，其特征在于，所述方法还包括：

所述控制设备向所述主集群发送第四指令，所述第四指令指示所述主集群将用户数据同步至所述备集群；或者，

所述控制设备获取所述主集群和所述备集群中存储的用户数据，所述控制设备根据所述主集群中存储的用户数据对所述备集群中存储的用户数据进行调整。

5.根据权利要求 1-4 任一项所述的方法，其特征在于，所述方法还包括：

所述控制设备根据用户输入的与所述第一业务相关的多个数据集的信息和所述第一时刻，为所述第一业务配置所述第一数据备份策略。

6.根据权利要求 1-5 任一项所述的方法，其特征在于，所述方法还包括：

所述控制设备为第二业务配置第二数据备份策略，所述第二数据备份策略包括与所述第二业务相关的多个数据集的信息和第二时刻；

所述控制设备根据所述第二数据备份策略控制所述主集群或所述备集群将所述第二时刻的所述主集群中与所述第二业务相关的所述多个数据集备份至所述备集群。

7.根据权利要求 1-6 任一项所述的方法，其特征在于，与所述第一业务相关的所述多个数据集包括由所述主集群中的第一组件处理或存储的数据集和由所述主集群中的第二组件处理或存储的数据集。

8.根据权利要求 1-7 任一项所述的方法，其特征在于，所述控制设备包括主客户端以及备客户端，所述主客户端用于检测所述主集群的第一状态信息，所述备客户端用于检测所述备集群的第二状态信息，所述方法还包括：

所述控制设备获取所述主客户端检测得到的第一状态信息以及所述备客户端检测得到的第二状态信息；

当所述第一状态信息指示所述主集群为备身份或集群失效，且所述第二状态信息指

示所述备集群为主身份时，所述控制设备确定所述备客户端为应用访问的客户端。

9.根据权利要求 8 所述的方法，其特征在于，所述方法还包括：

所述控制设备向用户提示所述主集群故障的信息；

所述控制设备响应于所述用户针对所述备集群的身份调整操作，将所述备集群的身份由备身份调整成主身份。

10.根据权利要求 1-9 任一项所述的方法，其特征在于，所述控制设备与所述主集群隔离部署。

11.根据权利要求 1-10 任一项所述的方法，其特征在于，所述控制设备、所述主集群和所述备集群中设置有相同的时钟源。

12.根据权利要求 1-11 任一项所述的方法，其特征在于，所述主集群和/或所述备集群包括基于 hadoop 架构构建的集群。

13.一种数据备份方法，其特征在于，所述方法应用于数据备份系统，所述数据备份系统包括主集群、备集群和控制设备，所述方法包括：

所述主集群获取所述控制设备下发的指令，其中，所述指令中包括与第一业务相关的多个数据集的信息和第一时刻；

所述主集群根据所述指令，将所述第一时刻的所述主集群中与所述第一业务相关的所述多个数据集备份至所述备集群。

14.根据权利要求 13 所述的方法，其特征在于，所述主集群根据所述指令，将所述第一时刻的所述主集群中与所述第一业务相关的所述多个数据集备份至所述备集群，具体包括：

所述主集群根据与所述第一业务相关的所述多个数据集的信息和所述第一时刻，获取所述主集群中与所述第一业务相关的所述多个数据集在所述第一时刻的快照；

所述主集群根据所述快照，将所述快照对应的数据发送至所述备集群。

15.根据权利要求 13 或 14 所述的方法，其特征在于， 所述方法还包括：

所述主集群将用户数据同步至所述备集群。

16.根据权利要求 13-15 任一项所述的方法，其特征在于，所述主集群和/或所述备集群包括基于 hadoop 架构构建的集群。

17.一种控制设备，其特征在于，所述控制设备位于数据备份系统，所述数据备份系统还包括主集群以及备集群，所述控制设备包括：

控制模块，用于根据第一数据备份策略控制所述主集群或所述备集群将第一时刻的所述主集群中与第一业务相关的多个数据集备份至所述备集群，其中，所述第一数据备份策略包括与所述第一业务相关的所述多个数据集的信息和所述第一时刻。

18.根据权利要求 17 所述的控制设备，其特征在于，所述控制模块，具体用于：

向所述主集群发送第一指令，指示所述主集群将与所述第一业务相关的所述多个数据集在所述第一时刻的快照对应的数据发送至所述备集群；或者，

向所述备集群发送第二指令，指示所述备集群从所述主集群复制所述主集群中与所述第一业务相关的所述多个数据集在所述第一时刻的快照对应的数据。

19.根据权利要求 18 所述的控制设备，其特征在于，所述控制设备还包括：

通信模块，用于在向在所述控制设备向所述主集群发送第一指令或所述控制设备向所述备集群发送所述第二指令之前，向所述主集群发送第三指令，所述第三指令包括与所述第一业务相关的所述多个数据集的信息和所述第一时刻，所述第三指令用于指示所

述主集群获取在所述第一时刻的与所述第一业务相关的所述多个数据集的快照。

20.根据权利要求 17-19 任一项所述的控制设备，其特征在于，所述控制设备还包括：

通信模块，用于向所述主集群发送第四指令，所述第四指令指示所述主集群将用户数据同步至所述备集群；或者，

所述控制模块，还用于获取所述主集群和所述备集群中存储的用户数据，并根据所述主集群中存储的用户数据对所述备集群中存储的用户数据进行调整。

21.根据权利要求 17-20 所述的控制设备，其特征在于，所述控制设备还包括配置模块，用于根据用户输入的与所述第一业务相关的多个数据集的信息和所述第一时刻，为所述第一业务配置所述第一数据备份策略。

22.根据权利要求 17-21 任一项所述的控制设备，其特征在于，所述控制设备还包括配置模块，用于为第二业务配置第二数据备份策略，所述第二数据备份策略包括与所述第二业务相关的多个数据集的信息和第二时刻；

所述控制模块，还用于根据所述第二数据备份策略控制所述主集群或所述备集群将所述第二时刻的所述主集群中与所述第二业务相关的所述多个数据集备份至所述备集群。

23.根据权利要求 17-22 所述的控制设备，其特征在于，与所述第一业务相关的所述多个数据集包括由所述主集群中的第一组件处理或存储的数据集和由所述主集群中的第二组件处理或存储的数据集。

24.根据权利要求 17-23 任一项所述的控制设备，其特征在于，所述控制设备包括主客户端以及备客户端，所述主客户端用于检测所述主集群的第一状态信息，所述备客户端用于检测所述备集群的第二状态信息，所述控制设备还包括：

通信模块，用于获取所述主客户端检测得到的第一状态信息以及所述备客户端检测得到的第二状态信息；

确定模块，用于当所述第一状态信息指示所述主集群为备身份或集群失效，且所述第二状态信息指示所述备集群为主身份时，确定所述备客户端为应用访问的客户端。

25.根据权利要求 24 任一项所述的控制设备，其特征在于，所述控制设备还包括提示模块以及调整模块；

所述提示模块，用于向用户提示所述主集群故障的信息；

所述调整模块，用于响应于所述用户针对所述备集群的身份调整操作，将所述备集群的身份由备身份调整成主身份。

26.根据权利要求 17-25 任一项所述的控制设备，其特征在于，所述控制设备与所述主集群隔离部署。

27.根据权利要求 17-26 任一项所述的控制设备，其特征在于，所述控制设备、所述主集群和所述备集群中设置有相同的时钟源。

28.根据权利要求 17-27 任一项所述的控制设备，其特征在于，所述主集群和/或所述备集群包括基于 hadoop 架构构建的集群。

29.一种数据备份系统，其特征在于，所述数据备份系统包括控制设备、主集群和备集群；

所述控制设备用于执行前述方法权利要求 1-12 任一项所述的方法；

所述主集群用于执行前述方法权利要求 13-16 任一项所述的方法；

所述备集群用于获取并存储从所述主集群备份的数据集。

30.一种控制设备，其特征在于，所述计算设备包括处理器、存储器；

所述处理器用于执行所述存储器中存储的指令，以使所述计算设备执行如权利要求1至12任一项所述的方法。

31.一种计算机可读存储介质，其特征在于，包括指令，当所述指令在计算设备运行时，使得所述计算设备执行如权利要求1至12中任一项所述的方法。

32.一种包含指令的计算机程序产品，当其在计算设备上运行时，使得所述计算设备执行如权利要求1至12中任一项所述的方法。

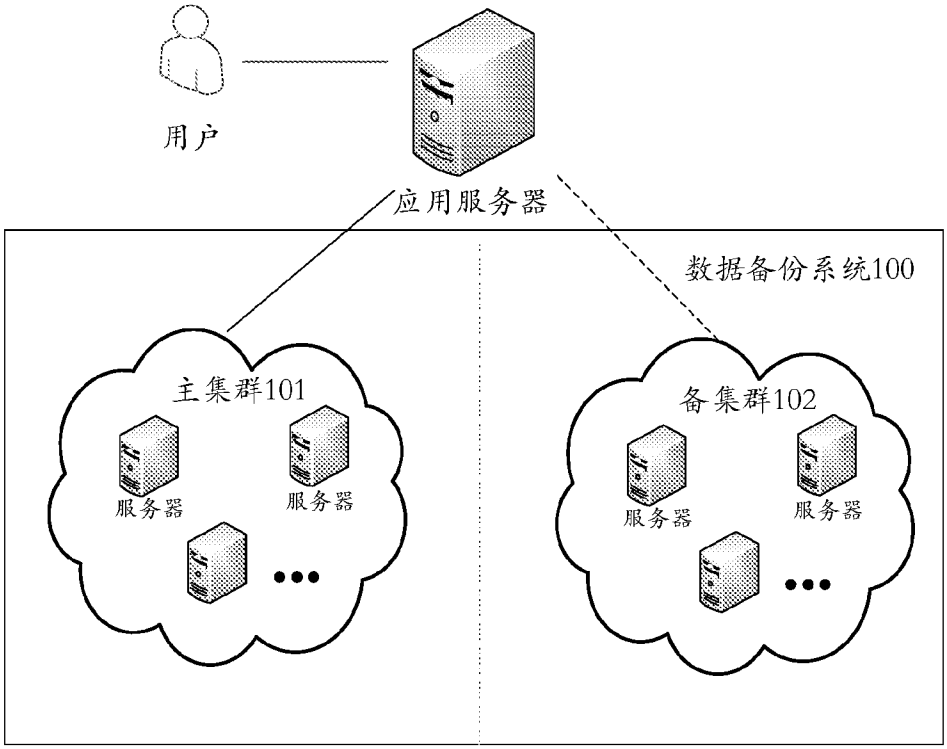

用户
应用服务器
数据备份系统100
主集群101
服务器
服务器
备集群102
服务器
服务器

图 1

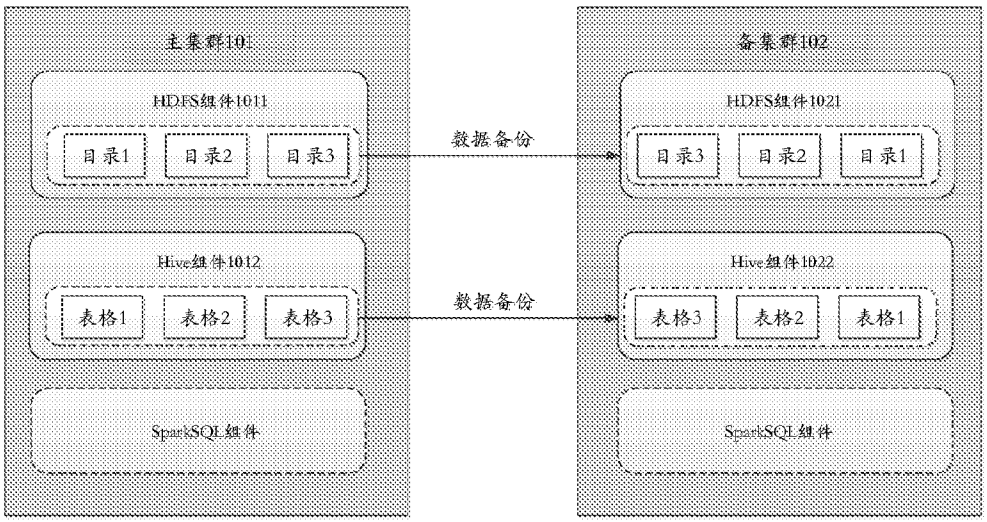

主集群101
HDFS组件1011
目录1
目录2
目录3
Hive组件1012
表格1
表格2
表格3
SparkSQL组件
数据备份
数据备份
备集群102
HDFS组件1021
目录3
目录2
目录1
Hive组件1022
表格3
表格2
表格1
SparkSQL组件

图 2

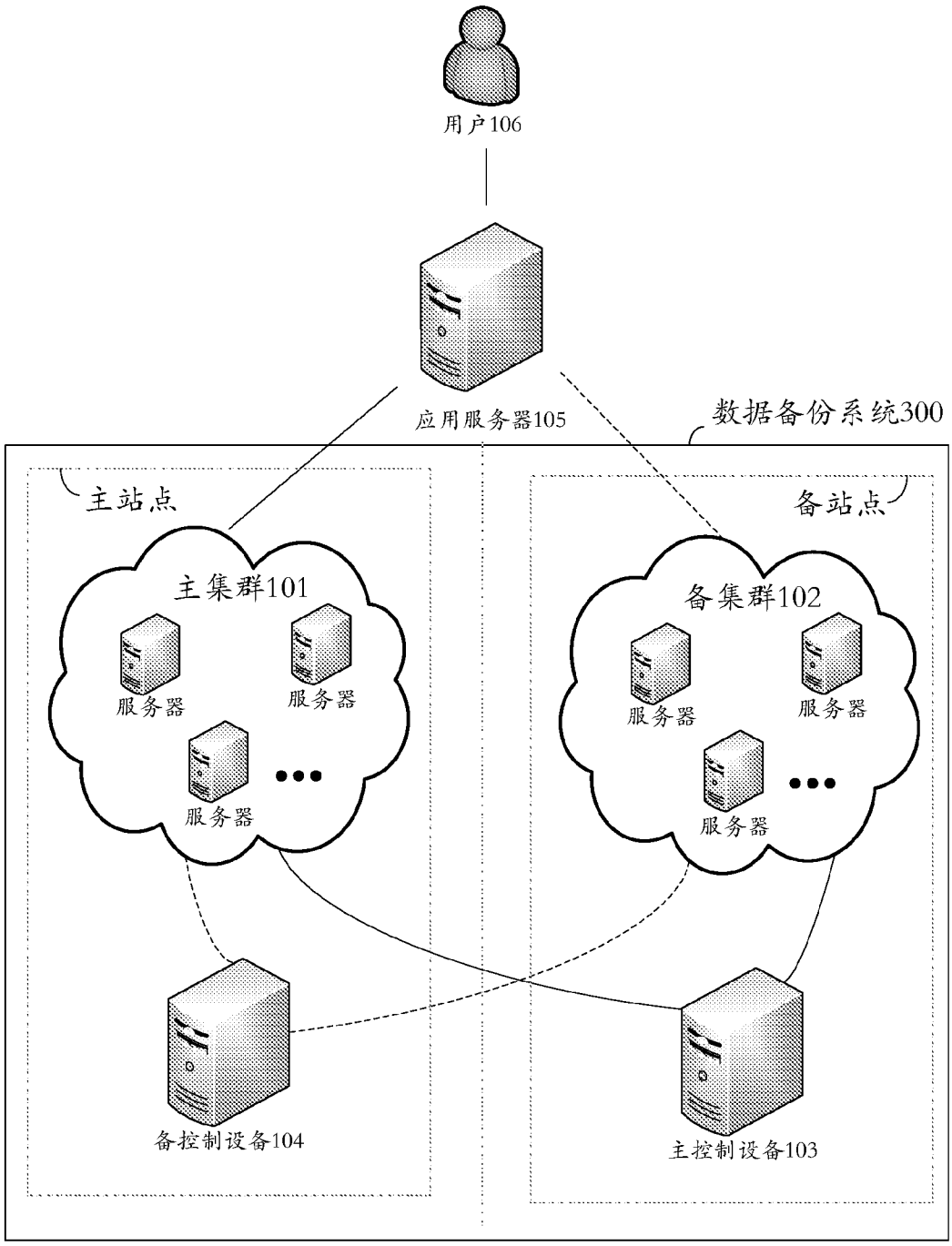

用户106
应用服务器105
数据备份系统300
主站点
备站点
主集群101
服务器
服务器
服务器
备集群102
服务器
服务器
服务器
备控制设备104
主控制设备103

图 3

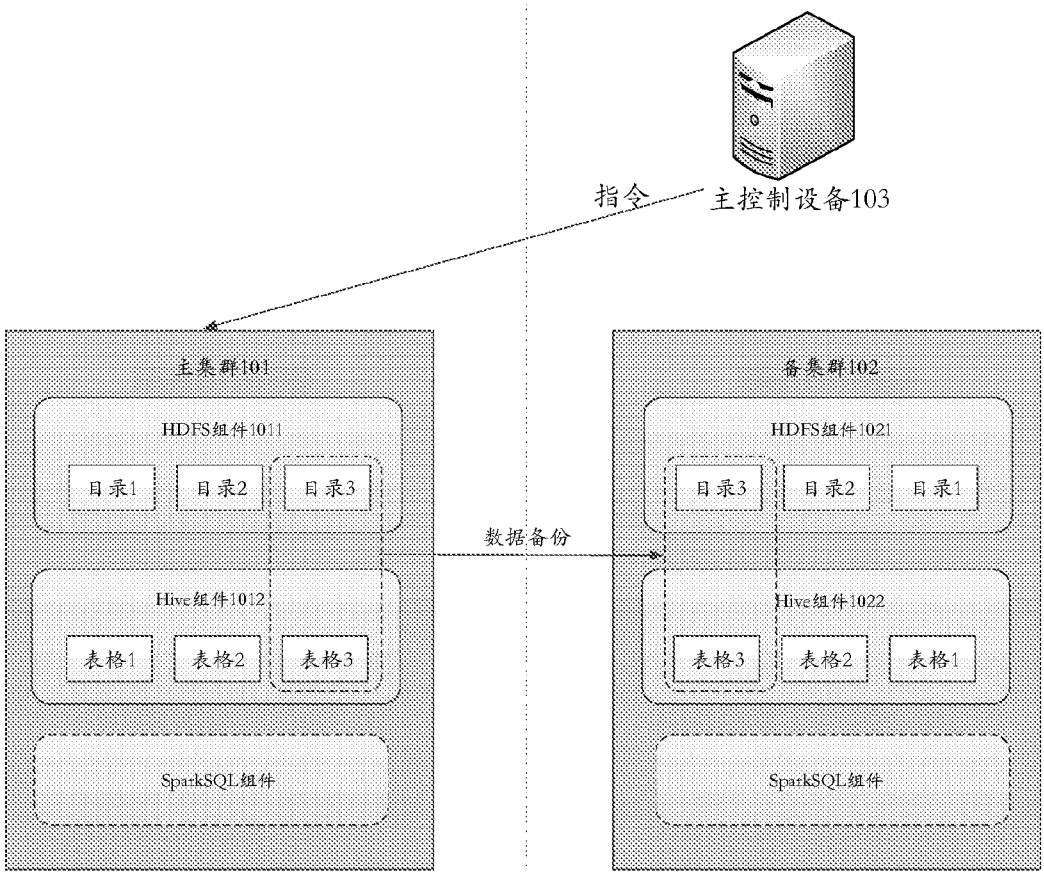

主控制设备103
指令
主集群101
HDFS组件1011
目录1
目录2
目录3
Hive组件1012
表格1
表格2
表格3
SparkSQL组件
数据备份
备集群102
HDFS组件1021
目录3
目录2
目录1
Hive组件1022
表格3
表格2
表格1
SparkSQL组件

图 4

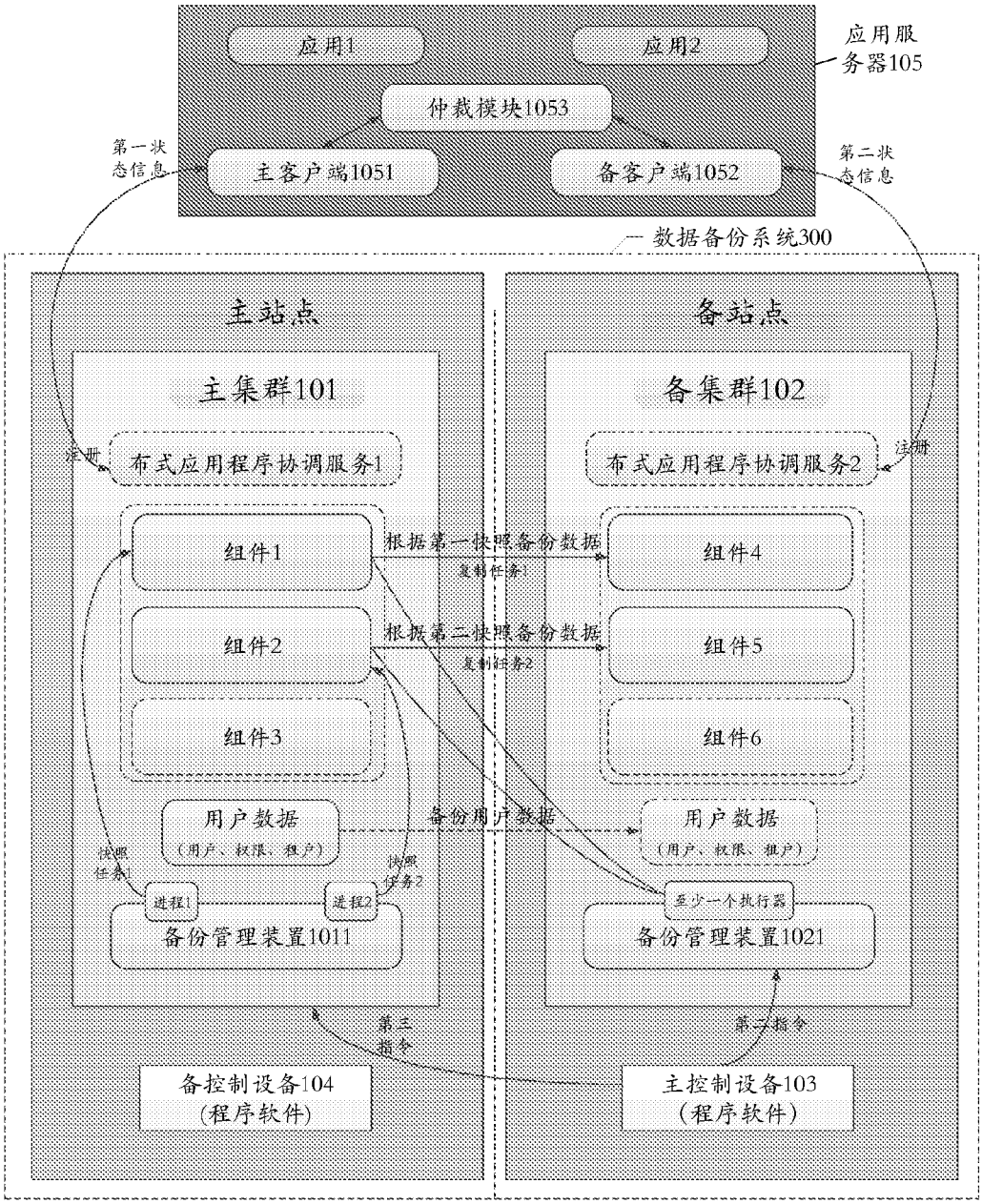

应用1
应用2
应用服务器105
仲裁模块1053
第一状态信息
主客户端1051
备客户端1052
第二状态信息
数据备份系统300
主站点
备站点
主集群101
备集群102
注册
布式应用程序协调服务1
布式应用程序协调服务2
注册
组件1
组件2
组件3
根据第一快照备份数据
复制任务1
根据第二快照备份数据
复制任务2
组件4
组件5
组件6
用户数据
（用户、权限、租户）
备份用户数据
用户数据
（用户、权限、租户）
快照任务1
快照任务2
进程1
进程2
至少一个执行器
备份管理装置1011
备份管理装置1021
第三指令
第二指令
备控制设备104
（程序软件）
主控制设备103
（程序软件）

图 5

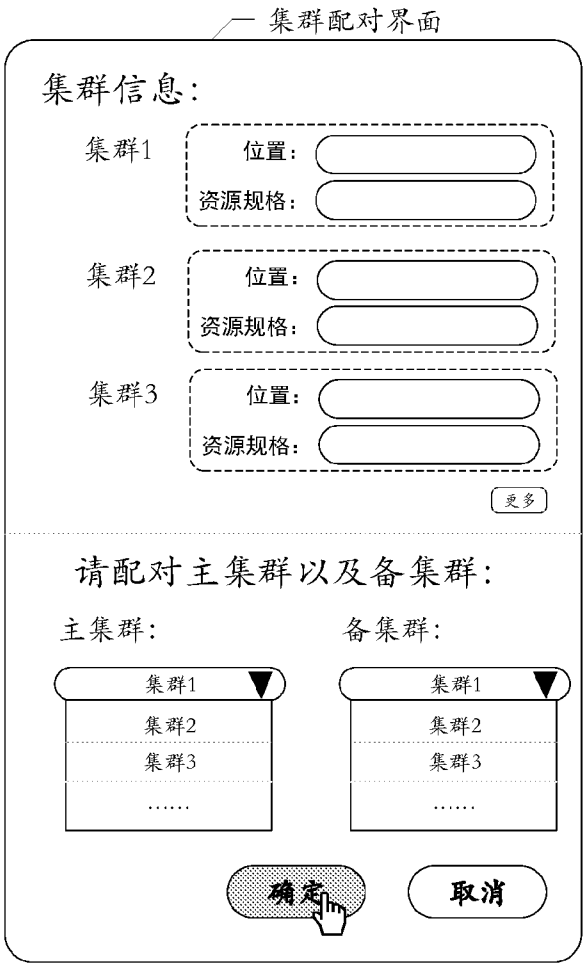

集群配对界面
集群信息:
集群1
位置:
资源规格:
集群2
位置:
资源规格:
集群3
位置:
资源规格:
更多
请配对主集群以及备集群:
主集群:
备集群:
集群1
集群2
集群3
……
集群1
集群2
集群3
……
确定
取消

图 6

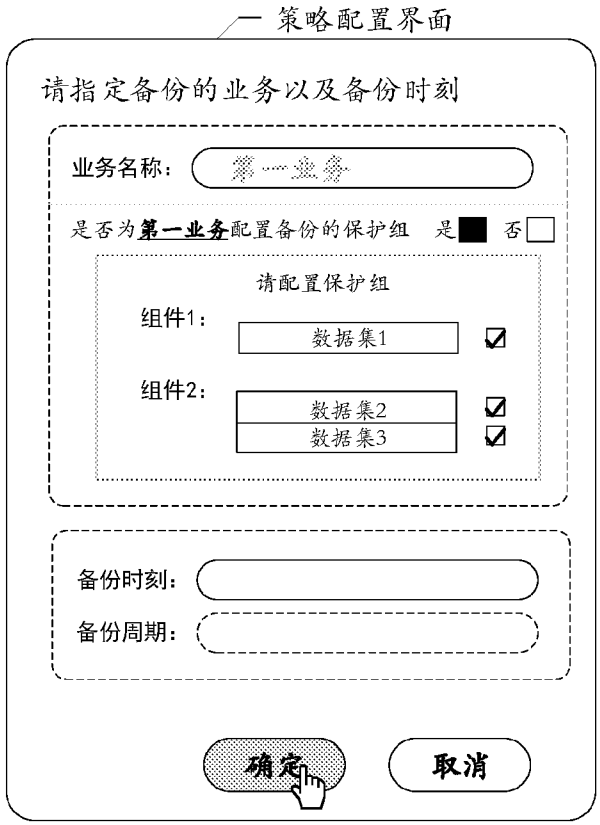

策略配置界面
请指定备份的业务以及备份时刻
业务名称：
第一业务
是否为第一业务配置备份的保护组 是 否
请配置保护组
组件1：
数据集1
组件2：
数据集2
数据集3
备份时刻：
备份周期：
确定
取消

图 7

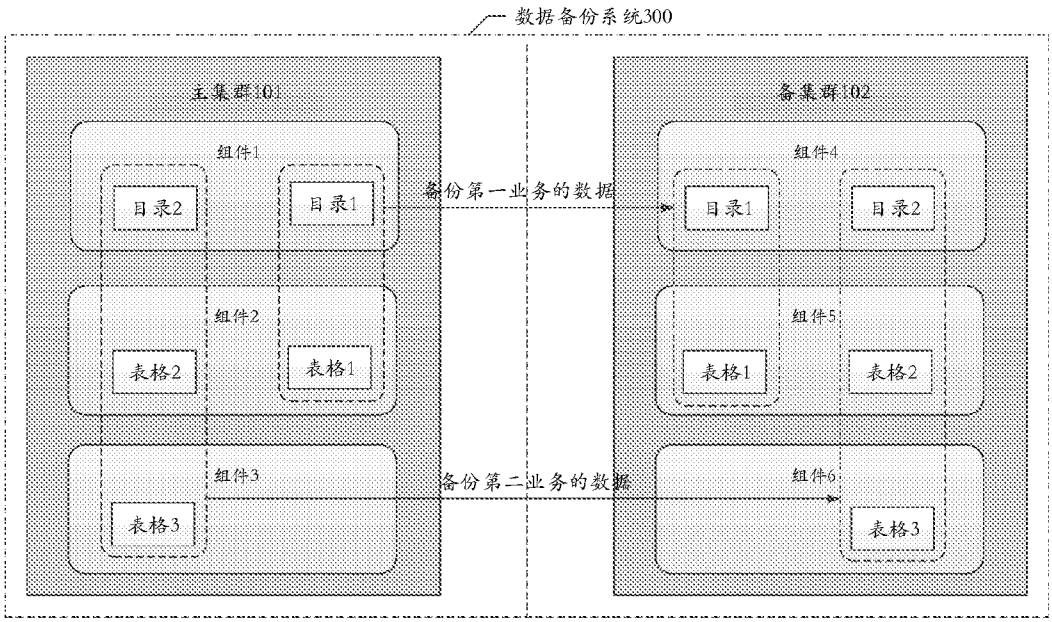

数据备份系统300
主集群101
备集群102
组件1
目录2
目录1
组件2
表格2
表格1
组件3
表格3
备份第一业务的数据
备份第二业务的数据
组件4
目录1
目录2
组件5
表格1
表格2
组件6
表格3

图 8

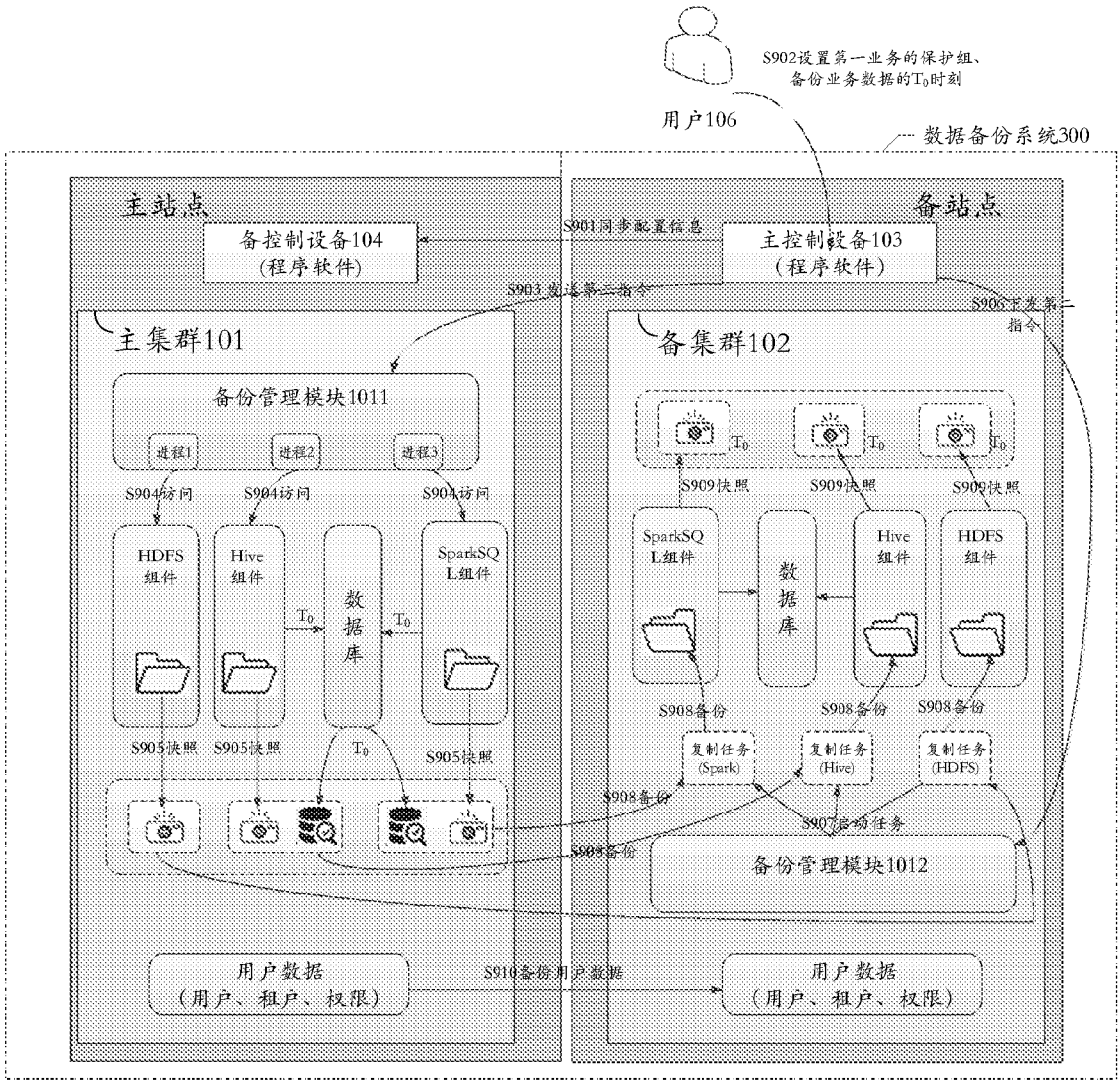

S902设置第一业务的保护组、
备份业务数据的T0时刻
用户106
数据备份系统300
主站点
备站点
备控制设备104
(程序软件)
S901同步配置信息
主控制设备103
(程序软件)
S903发送第二指令
S906下发第二指令
主集群101
备集群102
备份管理模块1011
进程1
进程2
进程3
S904访问
S904访问
S904访问
HDFS组件
Hive组件
数据库
SparkSQL组件
T0
T0
S905快照
S905快照
T0
S905快照
T0
T0
T0
S909快照
S909快照
S909快照
SparkSQL组件
数据库
Hive组件
HDFS组件
S908备份
S908备份
S908备份
复制任务(Spark)
复制任务(Hive)
复制任务(HDFS)
S908备份
S907启动任务
S908备份
备份管理模块1012
用户数据
(用户、租户、权限)
S910备份用户数据
用户数据
(用户、租户、权限)

图 9

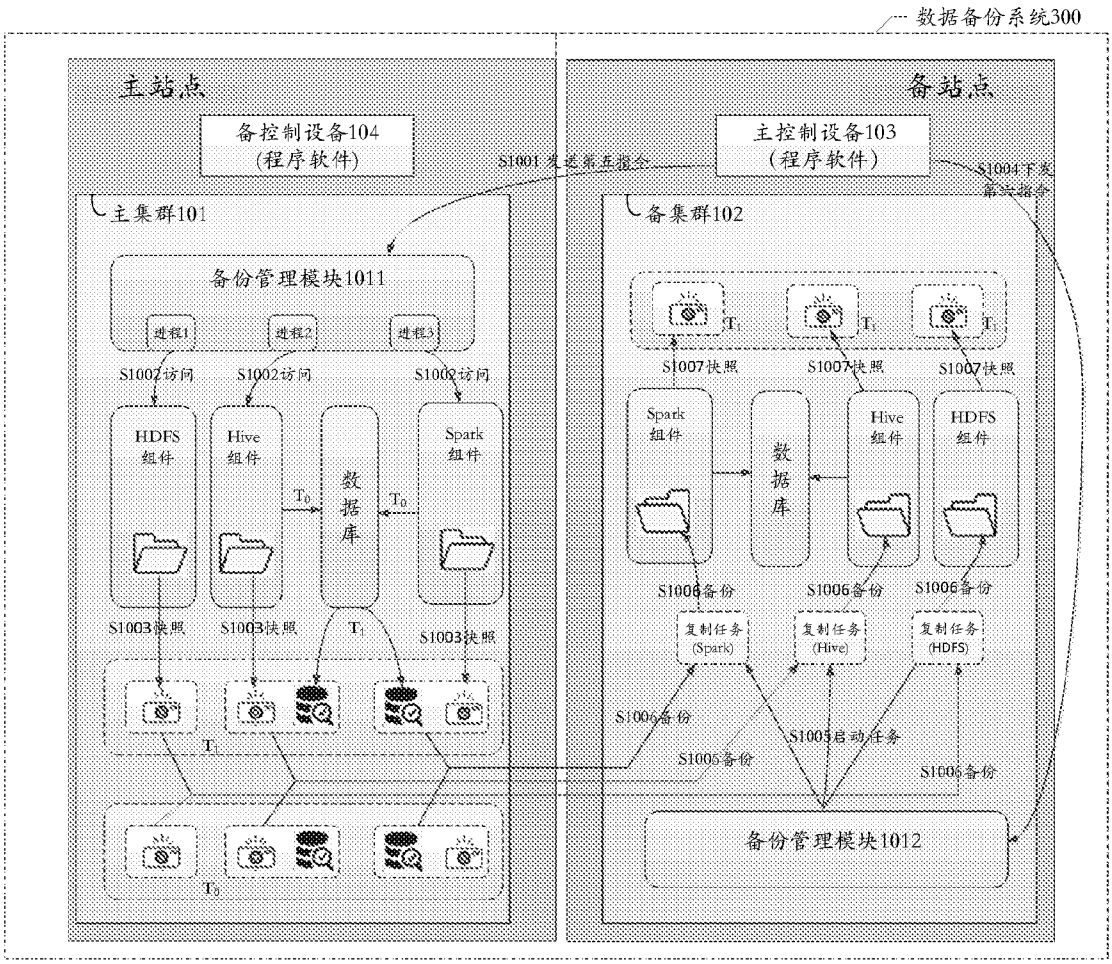

数据备份系统300
主站点
备站点
备控制设备104
(程序软件)
主控制设备103
(程序软件)
S1001发送第一指令
S1004下发第二指令
主集群101
备集群102
备份管理模块1011
进程1
进程2
进程3
S1002访问
S1002访问
S1002访问
HDFS组件
Hive组件
数据库
Spark组件
T0
T0
S1003快照
S1003快照
T1
S1003快照
T1
T0
T1
T1
T1
S1007快照
S1007快照
S1007快照
Spark组件
数据库
Hive组件
HDFS组件
S1006备份
S1006备份
S1006备份
复制任务(Spark)
复制任务(Hive)
复制任务(HDFS)
S1006备份
S1005启动任务
S1005备份
S1006备份
备份管理模块1012

图 10

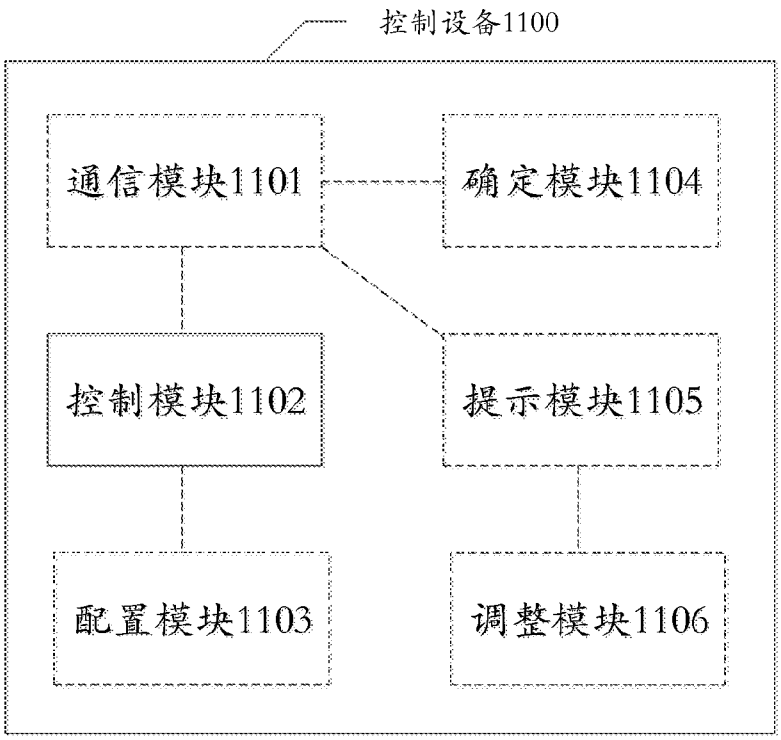

控制设备1100
通信模块1101
确定模块1104
控制模块1102
提示模块1105
配置模块1103
调整模块1106

图 11

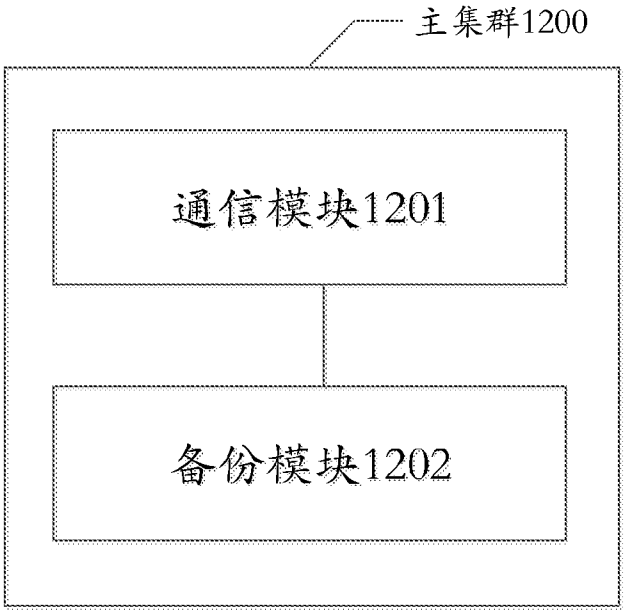

主集群1200
通信模块1201
备份模块1202

图 12

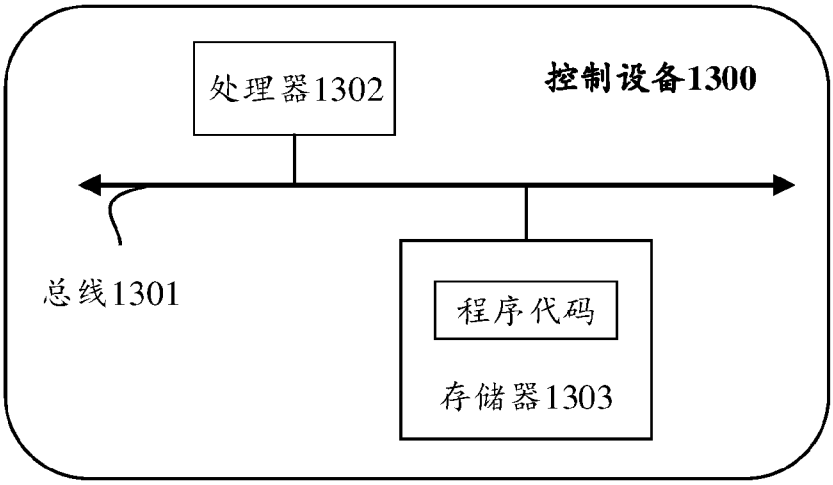

处理器1302
控制设备1300
总线1301
程序代码
存储器1303

图 13

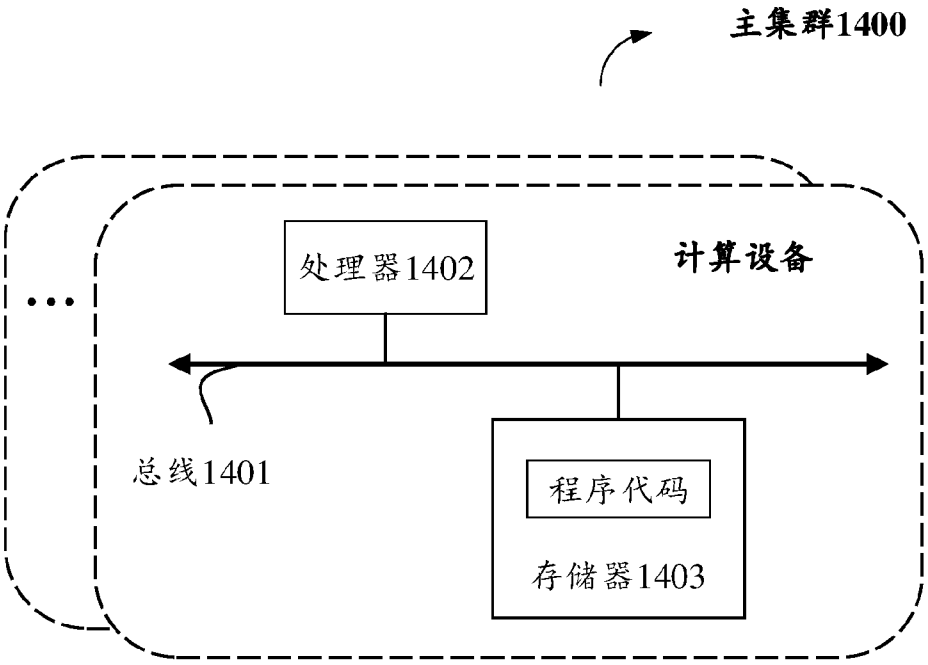

主集群1400
处理器1402
计算设备
总线1401
程序代码
存储器1403

图 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/072427

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/14(2006.01)i; G06F 11/20(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; H04L

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; WPI; EPODOC; CNKI; IEEE: 控制, 主集群, 集群, 业务, 相关, 数据, 备份, 同步, control, main, primary, master, cluster, operation, data, backup, synchro+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 112527567 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 19 March 2021 (2021-03-19) description, paragraphs [0020]-[0072]	1-32
Y	CN 102982085 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.) 20 March 2013 (2013-03-20) description, paragraphs [0022]-[0050]	1-32
Y	CN 102917072 A (BEIJING QIHOO TECHNOLOGY CO., LTD. et al.) 06 February 2013 (2013-02-06) description, paragraphs [0023]-[0051]	1-32
A	CN 112380067 A (WEST CHINA HOSPITAL, SICHUAN UNIVERSITY) 19 February 2021 (2021-02-19) entire document	1-32
A	CN 109739690 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 10 May 2019 (2019-05-10) entire document	1-32
A	US 2019050306 A1 (T-MOBILE USA, INC.) 14 February 2019 (2019-02-14) entire document	1-32

☐ 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
"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	Date of mailing of the international search report
02 April 2022	**13 April 2022**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/072427

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 112527567 A	19 March 2021	None	
CN 102982085 A	20 March 2013	None	
CN 102917072 A	06 February 2013	None	
CN 112380067 A	19 February 2021	None	
CN 109739690 A	10 May 2019	None	
US 2019050306 A1	14 February 2019	None	

国际检索报告

国际申请号
PCT/CN2022/072427

A. 主题的分类

G06F 11/14(2006.01)i; G06F 11/20(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; H04L

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;WPI;EPODOC;CNKI;IEEE: 控制，主集群，集群，业务，相关，数据，备份，同步，control，main，primary，master，cluster，operation，data，backup，synchro+

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 112527567 A (北京百度网讯科技有限公司) 2021年3月19日 (2021-03-19) 说明书第[0020]-[0072]段	1-32
Y	CN 102982085 A (北京奇虎科技有限公司 等) 2013年3月20日 (2013-03-20) 说明书第[0022]-[0050]段	1-32
Y	CN 102917072 A (北京奇虎科技有限公司 等) 2013年2月6日 (2013-02-06) 说明书第[0023]-[0051]段	1-32
A	CN 112380067 A (四川大学华西医院) 2021年2月19日 (2021-02-19) 全文	1-32
A	CN 109739690 A (平安科技深圳有限公司) 2019年5月10日 (2019-05-10) 全文	1-32
A	US 2019050306 A1 (T-MOBILE USA, INC.) 2019年2月14日 (2019-02-14) 全文	1-32

☐ 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	国际检索报告邮寄日期
2022年4月2日	2022年4月13日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 金霞 电话号码 86-(10)-53961524

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2022/072427

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 112527567 A	2021年3月19日	无	
CN 102982085 A	2013年3月20日	无	
CN 102917072 A	2013年2月6日	无	
CN 112380067 A	2021年2月19日	无	
CN 109739690 A	2019年5月10日	无	
US 2019050306 A1	2019年2月14日	无	