

(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
2025 年 1 月 2 日 (02.01.2025)



(10) 国际公布号
WO 2025/001030 A1

(51) 国际专利分类号:
G06F 11/14 (2006.01)

(21) 国际申请号: PCT/CN2023/143183

(22) 国际申请日: 2023 年 12 月 29 日 (29.12.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202311701095.5 2023 年 12 月 12 日 (12.12.2023) CN

(71) 申请人: 天翼云科技有限公司 (CHINA TELECOM CLOUD TECHNOLOGY CO., LTD) [CN/CN]; 中国北京市东城区青龙胡同甲 1 号、3 号 2 幢 2 层 205-32 室, Beijing 100007 (CN)。

(72) 发明人: 雷光钰 (LEI, Guangyu); 中国北京市东城区青龙胡同甲 1 号、3 号 2 幢 2 层 205-32 室, Beijing 100007 (CN)。

(74) 代理人: 北京知汇林知识产权代理事务所 (普通合伙) (BEIJING ZHIHUILIN INTELLECTUAL PROPERTY LAW FIRM); 中国北京市海淀区海淀南路甲 21 号中关村知识产权大厦 A 座 7 层, Beijing 100080 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ,

(54) Title: HYBRID CLOUD DATA BACKUP MANAGEMENT METHOD BASED ON MULTI-OBJECTIVE OPTIMAL COPY MANAGEMENT POLICY

(54) 发明名称: 基于多目标最优副本管理策略的混合云数据备份管理方法

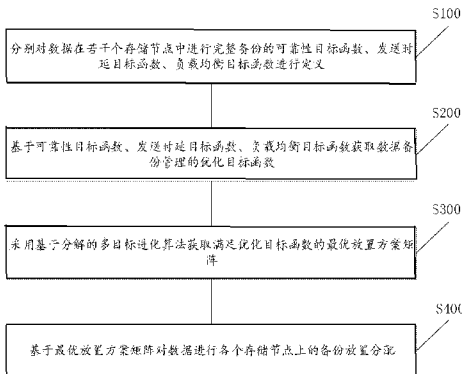


图 1

- S100 Respectively define a reliability objective function, a sending delay objective function and a load balancing objective function for complete backup of data in several storage nodes
- S200 Acquire an optimization objective function for data backup management on the basis of the reliability objective function, the sending delay objective function and the load balancing objective function
- S300 By using a decomposition-based multi-objective evolutionary algorithm, acquire an optimal placement scheme matrix that meets the optimization objective function
- S400 On the basis of the optimal placement scheme matrix, perform backup placement distribution of the data on each node

(57) Abstract: Provided in the present invention are a data backup management method and system based on a multi-objective optimal copy management policy. The method comprises: respectively defining a reliability objective function, a sending delay objective function and a load balancing objective function for complete backup of data; acquiring an optimization objective function for data backup management on the basis of the reliability objective function, the sending delay objective function and the load balancing objective function; by using a decomposition-based multi-objective evolutionary algorithm, acquiring an optimal placement scheme matrix that meets the optimization objective function; and on the basis of the optimal placement scheme matrix, performing backup placement distribution of the data on each node. In the present invention, by comprehensively considering the reliability of backup data, the delay of a backup system and the stability of the backup system, and by using a decomposition-based multi-objective optimization algorithm, a backup copy optimal placement scheme for a gold copy is obtained, thereby reducing the risk of data loss and damage.

(57) 摘要: 本发明提供了一种基于多目标最优副本管理策略的数据备份管理方法与系统, 包括: 分别对数据进行完整备份的可靠性目标函数、发送时延目标函数、负载均衡目标函数进行定义; 基于可靠性目标函数、发送时延目标函数、负载均衡目标函数获取数据备份管理的优化目标函数; 采用基于分解的多目标进化算法获取满足优化目标函数的最优放置方案矩阵, 基于最优放置方案矩阵对数据进行各个节点上的备份放置分配。本发明的综合考虑备份数据的可靠性、备份系统的时延和备份系统的稳定性, 通过基于分解的多目标优化算法, 得到黄金副本的备份副本最优放置方案, 降低数据丢失和损坏的风险。

LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN,
MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,
PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,
SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

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

本国际公布:

- 包括国际检索报告(条约第21条(3))。
- 在修改权利要求的期限届满之前进行, 在收到该
修改后将重新公布(细则48.2(h))。
- 根据申请人的请求, 在条约第21条(2)(a)所规定的
期限届满之前进行。

基于多目标最优副本管理策略的混合云数据备份管理方法

技术领域

本发明属于云计算灾备领域，特别是涉及一种基于多目标最优副本管理策略的混合云数据备份管理方法。

背景技术

随着大数据时代的发展，数据的重要性上升到关键资源的层次。云数据备份恢复在政府/企业上云过程中具有重要意义。它涉及将数据从本地机房备份到云端，或者将数据从云上备份到云端，以便在数据丢失、损坏或遭受攻击时进行恢复。对于确保数据安全、提高数据可访问性、降低成本、简化数据管理具有重要意义。但是海量数据的备份和恢复让传统的数据保护方法显得维护成本

高和恢复时间长，数据安全性降低。

数据副本管理策略（CDM）是云上一种高效的备份数据保护解决方案，是政府、企业和个人保护关键信息的重要组成部分。一个优秀的备份副本管理策略能够应对故障或自然灾害导致的数据丢失、保护业务连续性、应对网络安全威胁、提高数据恢复速度等。数据备份管理是确保数据安全和业务连续性的关键。通过实施有效的数据备份策略，可以降低数据丢失的风险，保护政府、企业和客户的利益，对于整个管理系统的数据稳定性、可靠性和相应速度等具有重大的影响。

在副本管理策略中，备份数据的可靠性是至关重要的。可靠性是指备份的数据应该是可靠的，能够在发生任何问题时都能被恢复。此外，在备份和恢复的过程中，还需要确保备份数据能够在出现错误时通过修复数据文件而得到恢复。

在副本管理策略中，备份系统的时延是其重要特性之一。云存储系统往往将文件以副本的形式存储多份并放在不同的节点上，以保证其存储信息的安全性和降低用户访问的延迟。在云存储系统中，当一个文件访问量较少，且该文件的副本数目过多时会增加系统不必要存储开销和存储成本；而当一个文件的

访问量很高，且该文件的副本较少时，则会增加用户访问该文件的访问时延，无法满足用户需求。

在副本管理策略中，备份系统的负载均衡也是需要考虑的重要因素。服务器的负载均衡率不仅影响了服务器的资源利用率，并对整个系统都有较大的影响。

然而，在实际的副本管理策略中，备份系统在三个方面的追求可能是潜在冲突的，无法同时达到最优。

发明内容

本发明的目的是提供一种基于多目标最优副本管理策略的混合云数据备份管理方法，用于弥补上述技术方案的不足。在备份恢复管理系统上，能够通过一个平台实现对不同类型数据的备份与恢复。在备份模块中，同时考虑备份数据的可靠性、系统整体时延和负载均衡问题，建立多目标优化模型，获取三种问题 Pareto 平衡的最优副本管理策略，基于最优策略将备份副本备份到系统底层存储空间中。实现了在简化用户操作提高业务效率的同时，提高备份数据的可靠性和备份系统的稳定性，降低备份系统时延，保证业务系统运行稳定。

一方面为实现上述目的，本发明提供了一种基于多目标最优副本管理策略的混合云数据备份管理方法，包括以下步骤：

分别对数据在若干个存储节点中进行完整备份的可靠性目标函数、发送时延目标函数、负载均衡目标函数进行定义；

基于所述可靠性目标函数、发送时延目标函数、负载均衡目标函数获取数据备份管理的优化目标函数；

采用基于分解的多目标进化算法获取满足所述优化目标函数的最优放置方案矩阵；

基于所述最优放置方案矩阵对数据进行各个存储节点上的备份放置分配。

可选地，所述可靠性目标函数如下所示：

$$s_1(TK) = \prod_{i=1} \left| 1 - \prod_{j=1} (tk_{ij} \times e^{-\lambda_j u}) \right|$$

其中, $s_1(k_i) = 1 - \prod (tk_{ij} \times e^{-\chi_j u})$, 表示第 i 个数据文件 k_i 的可靠性, χ_j 为存储集群 T 中第 j 个存储节点 t_j 的故障率, u 为存储节点的使用时间, $e^{-\chi_j u}$ 为存储节点 t_j 的失效率函数, K 表示数据文件, m 为存储节点数量, T 表示 m 个存储节点组成的集群, tk_{ij} 表示节点 t_j 上的副本。

5 可选地, 所述发送时延目标函数如下所示:

$$s_2(TK) = \frac{1}{\sum_{i=1}^n \sum_{j=1}^m A_{tol,i}^j} \times \sum_{i=1}^n \left\{ \sum_{j=1}^m \left[Tim\_del_j \times \sum_{a=1}^{A_{tol,i}^j} \left(\frac{a \times c_i}{v_j} \times tk_{ij} \right) \right] \right\}$$

其中, v_j 为备份节点 t_j 的数据发送速率; c_i 为数据文件 k_i 的大小, $A_{tol,i}^j$ 表示 k_i 在节点 t_j 上的总访问量; $s(k_i, a) = Tim\_del_j \times \frac{a \times c_i}{v_j} \times tk_{ij}$, 表示 k_i 的备份副本在存储节点 t_j 上的第 a 个访问的发送时延。

10 可选地, 所述负载均衡目标函数如下所示:

$$s_3(TK) = \sum_{j=1}^m \left\| \sigma_j \times \sum_{i=1}^n tk_{ij} - \sum_{i=1}^n tk_{ij} \right\|_2^2$$

其中, σ_j 为每个备份节点 t_j 的权重, 且 $\sum_{j=1}^m \sigma_j = 1$ 。

可选地, 所述优化目标函数如下所示:

$$\begin{aligned} \max \quad & S(X) = [f_1(X), f_2(X), f_3(X)]^T \\ \text{s.t.} \quad & \sum_{i=1}^n [\rho(k_i) \times tk_{ij}] \leq \Gamma(t_j) \\ & \sigma_j \times \sum_{i=1}^n tk_{ij} \geq \sum_{i=1}^n tk_{ij} \\ & \sum_{j=1}^m (tk_{ij}) = 1 \end{aligned}$$

15 其中, X 为决策变量, 此处为放置方案矩阵 TK , $f_1(X) = s_1(TK)$,

$f_2(X) = \frac{1}{s_2(TK)}$, $f_3(X) = \frac{1}{s_3(TK)}$, $\rho(k_i)$ 表示第 i 个数据文件 k_i 的大小, $\Gamma(t_j)$ 表

示第 j 个存储节点 t_j 的容量大小。

可选地, 采用基于分解的多目标进化算法获取满足所述优化目标函数的最优放置方案矩阵的过程包括:

5 初始化种群数量并随机生成与种群数量相同的权重向量;

基于Tchebycheff分解算法将所述优化目标函数进行分解为与种群数量相同的若干个子问题;

初始化进化次数, 随机生成种群数量相同的放置方案矩阵;

初始化前沿解集以及优化目标函数对应的参考点;

10 初始化邻域向量, 获取距离最后一个权重向量最近的若干个权重向量, 并基于若干个权重向量构建索引集;

设置最大迭代次数并进行迭代, 基于迭代次数对最后一个权重向量和对应的放置方案矩阵进行迭代更新, 当达到最大迭代次数时停止迭代, 获取所述最优放置方案矩阵。

15 可选地, 基于迭代次数对每个权重向量和对应的放置方案矩阵进行更新的过程包括:

基于所述索引集随机获取索引号, 获取所述索引号对应的放置方案矩阵;

基于对应的放置方案矩阵获取新放置方案矩阵, 基于所述新放置方案矩阵分别对参考点、种群以及非支配解集进行迭代更新;

20 当达到最大迭代次数时, 获取更新后的非支配解集并作为最优前沿解集, 在所述最优前沿解集中选择一个权衡解, 将所述权衡解作为最优放置方案矩阵。

可选地, 所述参考点的更新过程中, 若 $z_k^* < f_k(X_h^{p(new)})$, 则 $z_k^* = f_k(X_h^{p(new)})$, z_k^* 表示子问题 f_k 的参考点, $X_h^{p(new)}$ 表示新放置方案矩阵;

25 所述种群的更新过程中, 对于索引集中的每一个邻域个体 $X_h^{r_2}$, 若 $g^{te}(X_h^{p(new)} | \lambda, z^*) \leq g^{te}(X_h^b | \lambda, z^*)$, 则 $X_h^b = X_h^{p(new)}$, 其中 g^{te} 表示基于

Tchebycheff 分解算法得到的子问题, λ 表示邻域个体 X_h^b 的权重向量。

可选地, 所述非支配解集的更新过程中, 保留 $INL(h-1)$ 中支配 $X_h^{p(new)}$ 的解, 舍弃 $INL(h-1)$ 中所有被 $X_h^{p(new)}$ 支配的解, 若所有 $INL(h-1)$ 的解都不支配 $X_h^{p(new)}$, 将 $X_h^{p(new)}$ 加入 $INL(h)$, 其中, $INL(h-1)$ 表示上一次更新后的非支配解集, $X_h^{p(new)}$ 表示新放置方案矩阵, $INL(h)$ 表示当前的非支配解集。

另一方面为实现上述目的, 本发明提供了一种基于多目标最优副本管理策略的混合云数据备份管理系统, 其特征在于, 所述系统包括存储器、处理器及存储在存储器上并能够在处理器上运行的计算机程序, 所述处理器执行所述计算机程序时实现上述基于多目标最优副本管理策略的混合云数据备份管理方法。

与现有技术相比, 本发明的有益效果在于, 本发明的基于多目标最优副本管理策略的混合云数据备份恢复管理方法, 综合考虑备份数据的可靠性、备份系统的时延和备份系统的稳定性, 通过基于分解的多目标优化算法, 得到黄金副本的备份副本最优放置方案, 降低数据丢失和损坏的风险。用户通过本发明可以对备份恢复策略进行设置, 能够灵活、快速的进行数据备份与恢复, 帮助管理员快速定位故障, 最大程度地提高系统的效率和性能。

附图说明

通过阅读下文优选实施方式的详细描述, 各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的, 而并不认为是对本发明的限制。而且在整个附图中, 用相同的参考符号表示相同的部件。在附图中:

图 1 为本发明实施例中的多目标最优副本管理方法流程图;

图 2 为本发明实施例中的基于多目标最优副本管理策略的混合云数据备份恢复管理系统设计原理示意图;

图 3 为本发明实施例中的基于多目标最优副本管理策略的混合云数据备份恢复管理系统架构图;

图 4 为本发明实施例中基于 CDM 黄金副本思想的多目标备份副本管理策略

设计示意图；

图 5 为本发明实施例中的多目标最优副本管理方法路线图。

具体实施方式

下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。需要说明的是，在不冲突的情况下，本发明中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本发明。

10 如图 1-5 所示，本实施例中提供一种基于多目标最优副本管理策略的混合云数据备份管理方法与系统，其中方法如图 1 所示，包括：

S100：分别对数据在若干个存储节点中进行完整备份的可靠性目标函数、发送时延目标函数、负载均衡目标函数进行定义；

15 S200：基于所述可靠性目标函数、所述发送时延目标函数、所述负载均衡目标函数获取数据备份管理的优化目标函数；

S300：采用基于分解的多目标进化算法获取满足所述优化目标函数的最优放置方案矩阵；

S400：基于所述最优放置方案矩阵对数据进行各个存储节点上的备份放置分配。

20 系统的工作原理如图 2 所示，包括：

步骤 1：用户和权限管理设置。实现用户认证和权限管理功能，并对备份恢复系统进行全生命周期管理。

步骤 1.1：用户认证功能实现。用户登陆时，需要将用户处的账户名密码，与系统中保存的账户名密码进行比较，若两者匹配，则表明用户通过认证，否则拒绝用户访问并要求重新进行认证。
25

步骤 1.2：权限管理功能实现。为不同的用户设置角色权限，如普通用户、

运维管理员等，对不同用户角色进行权限设置。

步骤 2：备份策略管理。对备份方式和备份策略进行设置，用于调度各种备份任务的运行。

步骤 2.1：备份方式设置。本发明支持实时备份和定时备份两种备份方式。
5 用户可以根据不同的需求，选择相应的备份方式。

步骤 2.2：备份策略设置。本发明支持完全备份、增量备份和差异备份三种备份策略。当用户第一次备份时，默认进行全量备份。后续备份操作，用户可根据需求进行备份策略选择。

步骤 3：数据备份。开始进行数据备份，从数据节点与备份节点建立加密
10 连接并传输数据，按照其申请的备份方式进行数据备份。如用户选择备份至第三方备份软件，则转至步骤 3.1；若用户选择备份到系统底层存储空间中，则转至步骤 3.2。

步骤 3.1：备份恢复管理系统对接第三方备份软件，根据用户设置的备份策略，直接备份至第三方备份软件存储空间。

15 步骤 3.2：系统底层存储空间采用多目标最优备份副本管理方案。基于 CDM 黄金副本思想，设计多目标备份副本管理策略的步骤及原理分别如图 3、图 4 所示：

对由 m 个异构且独立的存储节点组成的集群 T ， $T=(t_1, t_2, \dots, t_m)$ ，用户将其数据文件 K ， $K=(k_1, k_2, \dots, k_n)$ ，备份到存储系统中。

20 相同的副本不能放置在同一个存储节点上，数据副本在备份节点中的放置方案通过矩阵 TK 来表示：

$$TK = \begin{bmatrix} tk_{11} & tk_{12} & \dots & tk_{1m} \\ tk_{21} & tk_{22} & \dots & tk_{2m} \\ \vdots & \vdots & \dots & \vdots \\ tk_{n1} & tk_{n2} & \dots & tk_{nm} \end{bmatrix}_{n \times m}$$

其中， $tk_{ij} = \begin{cases} 1, & \text{表示节点 } t_j \text{ 上有数据块 } k_i \text{ 的副本} \\ 0, & \text{表示节点 } t_j \text{ 上没有数据块 } k_i \text{ 的副本} \end{cases}$ 。

步骤 3.2.1 如图 5 所示, 从备份系统可靠性、时延和负载均衡三方面, 设置相应的目标优化函数:

步骤 3.2.1.1 可靠性目标函数设计。记存储集群 T 中第 j 个存储节点 t_j 的故障率为 χ_j , 存储节点的失效率与其使用时间 u 有关 (u 的单位为小时), 则存储节点 t_j 的失效率函数为 $e^{-\chi_j u}$ 。数据文件 K , $K=(k_1, k_2, \dots, k_n)$, 对其每一个数据文件 k_i , 若在 m 个存储节点中, 至少拥有一个完整的备份副本, 则认为第 i 个数据文件 k_i 是可靠的。 k_i 的可靠性定义为:

$$s_1(k_i) = 1 - \prod_{j=1}^m (tk_{ij} \times e^{-\chi_j u})$$

整个系统, 对数据文件 K 的进行完整备份的可靠性目标函数定义为:

$$s_1(TK) = \prod_{i=1}^n \left| 1 - \prod_{j=1}^m (tk_{ij} \times e^{-\chi_j u}) \right|$$

步骤 3.2.1.2 时延目标函数设计。记存储集群 T 中第 j 个存储节点 t_j 的网络访问总时延为 Tim_del_j (包括处理时延、排队时延、发送时延、传播时延)。考虑数据文件响应时间和传输时间, k_i 的备份副本在存储节点 t_j 上的第 a 个访问的发送时延:

$$s(k_i, a) = Tim\_del_j \times \frac{a \times c_i}{v_j} \times tk_{ij}$$

其中, v_j 为备份节点 t_j 的数据发送速率; c_i 为数据文件 k_i 的大小。 k_i 在不同备份节点上的副本的访问是相互独立的, 则 k_i 的总访问时延函数定义为:

$$s_2(k_i) = \sum_{j=1}^m \left[Tim\_del_j \times \sum_{a=1}^{A_{tol,i}^j} \left(\frac{a \times c_i}{v_j} \times tk_{ij} \right) \right]$$

其中, $A_{tol,i}^j$ 表示 k_i 在节点 t_j 上的总访问量。则整个系统, 对数据文件 K 的平均时延目标函数定义为:

$$s_2(TK) = \frac{1}{\sum_{i=1}^n \sum_{j=1}^m A_{tol,i}^j} \times \sum_{i=1}^n \left\{ \sum_{j=1}^m \left[Tim\_del_j \times \sum_{a=1}^{A_{tol,i}^j} \left(\frac{a \times c_i}{v_j} \times tk_{ij} \right) \right] \right\}$$

步骤 3.2.1.3 负载均衡目标函数设计。采用加权轮询法, 每个备份存储节

点 t_j ，根据其硬件承载能力，为其附上相应的权重 σ_j ， σ_j 满足 $\sum_{j=1}^m \sigma_j = 1$ 。本发明将所有存储节点 t_j 的实际负载与基于硬件承载能力计算出来的理论负载差的绝对值作为衡量系统负载均衡的标准。标准差越小，负载越均衡。定义负载均衡目标函数：

$$s_3(TK) = \sum_{j=1}^m \left\| \sigma_j \times \sum_{i=1}^n tk_{ij} - \sum_{i=1}^n tk_{ij} \right\|_2^2$$

步骤 3.2.2 定义多目标副本管理策略优化目标函数。在进行备份时，既要提供可靠的云备份服务、保障时延，还要避免某个存储节点的负载过高而出现故障。基于此，定义优化目标函数：

$$s.t. \sigma_j \times \sum_{i=1}^n tk_{ij} \geq \sum_{i=1}^n tk_{ij}$$

其中， X 被称为决策变量，此处为放置方案矩阵 TK 。经过多目标优化可搜寻到最优放置方案矩阵。 $f_1(X) = s_1(TK)$ ， $f_2(X) = \frac{1}{s_2(TK)}$ ， $f_3(X) = \frac{1}{s_3(TK)}$ ， $\rho(k_i)$ 表示第 i 个数据文件 k_i 的大小， $\Gamma(t_j)$ 表示第 j 个存储节点 t_j 的容量大小。 $\rho(k_i)$ 和 $\Gamma(t_j)$ 的计算单位要求保持一致。 σ_j 表示根据存储节点硬件承载能力为其分配的权重。

步骤 3.2.3 利用基于分解的多目标进化算法解决上述多目标问题。具体步骤为：

步骤 3.2.3.1 初始化种群大小为 P ，随机生成 Tchebycheff 分解算法均匀分布的 P 个权重向量 $\lambda^1, \dots, \lambda^P$ ， $\lambda^p = (\lambda_1^p, \lambda_2^p, \dots, \lambda_l^p)$ ， l 是目标函数个数。

步骤 3.2.3.2 利用 Tchebycheff 分解算法将 l 个目标构成的多目标问题分解为 P 个子问题，每个子问题为：

$$g^{te}(X|\lambda, z^*) = \max_{1 \leq k \leq l} \left\{ \lambda_k^p |f_k(X) - z_k^*| \right\}, p = 1, 2, \dots, P$$

其中， $z^* = (z_1^*, z_2^*, \dots, z_l^*)$ ， z^* 为多目标函数对应的参考点。 z_k^* 是使得子问

题 $f_k(X)$ 成立的最小值，表示子问题 $f_k(X)$ 的参考点。

步骤 3.2.3.3 初始化进化次数 $h=0$ ，从可行域中随机生成 P 个放置方案矩阵 $\{X^1(0), \dots, X^p(0), \dots, X^P(0)\}$ 。其中， $X^p(0)$ 表示初始化时第 p 个放置方案矩阵，可行域为满足约束条件的放置方案矩阵中各数值取值范围。

5 步骤 3.2.3.4 初始化 Pareto 前沿解集 $INL(0)=\Phi$ ，初始化参考点。

步骤 3.2.3.5 初始化邻域向量，找出距离第 p 个权重向量 λ^p 最近的 Neb 个权重向量，并将与 λ^p 最近的 Neb 个权重向量的索引集记为 $B(p)=\{p_1, p_2, \dots, p_{Neb}\}$ ， $p_{Neb}=1, 2, \dots, P$ 。

10 步骤 3.2.3.6 设置最大迭代次数 H ，从 $h=1$ 开始迭代，设置 $h=H$ 时迭代终止，对每一个权重向量 λ^p 和其对应的个体 X_h^p ， $p=1, 2, \dots, P$ 根据迭代次数 h 作更新操作：

步骤 3.2.3.6.1 进化算法更新个体：

从 $B(p)$ 中随机选择出索引号 c 和 d ，找到其对应的个体 $X_h^{p_c}$ 和 $X_h^{p_d}$ ，由 $X_h^{p_c}$ 和 $X_h^{p_d}$ 产生新解 $X_h^{p(new)}$ ：

$$15 \quad X_h^{p(new)} = \xi \cdot X_h^{p_c} + (1 - \xi) \cdot X_h^{p_d}$$

其中， ξ 为进化参数。

步骤 3.2.3.6.2 更新参考点 z^* ：对每一个分量 $k=1, 2, \dots, l$ ，若 $z_k^* < f_k(X_h^{p(new)})$ ，则 $z_k^* = f_k(X_h^{p(new)})$ ， z_k^* 表示子问题 f_k 的参考点。

20 步骤 3.2.3.6.3 更新种群：对于 $B(p)$ 中的每一个邻域个体 $X_h^b, b \in B(p)$ ，若 $g^{te}(X_h^{p(new)} | \lambda, z^*) \leq g^{te}(X_h^b | \lambda, z^*)$ ，则 $X_h^b = X_h^{p(new)}$ ；其中 g^{te} 表示基于 Tchebycheff 分解算法得到的子问题， λ 表示邻域个体 X_h^b 的权重向量。

步骤 3.2.3.6.4 更新非支配解集 $INL(h)$ ：在 $INL(h)$ 中，保留 $INL(h-1)$ 中支配 $X_h^{p(new)}$ 的解，舍弃中 $INL(h-1)$ 所有被 $X_h^{p(new)}$ 支配的解，若所有 $INL(h-1)$ 的解都不支配 $X_h^{p(new)}$ ，将 $X_h^{p(new)}$ 加入 $INL(h)$ 。

25 步骤 3.2.3.6.5 如果 $h=H$ ，得到的非支配解集 $INL(H)$ 即为放置方案矩阵 TK 的 Pareto 最优前沿解集，迭代终止，转至步骤 3.2.3.6；否则 $h=h+1$ ，转

至步骤 3.2.3.1

步骤 3.2.3.6.6 根据决策者的决策偏好或者随机从 Pareto 最优解集选择一个权衡解 $TK^{optimal}$ ，该 $TK^{optimal}$ 即为满足可靠性、延时和负载均衡三种目标 Pareto 最优的放置方案矩阵。根据 $TK^{optimal}$ 完成备份副本在各个节点上的放置分配。

步骤 3.2.4：备份完成后，备份系统通知管理节点数据备份已完成，管理节点通过备份信息生成备份日志并进行存档，并通知用户备份完成。

步骤 4：数据恢复。备份系统收到数据恢复请求，按照请求中的恢复时间点，通过备份选择算法计算出合适的备份数据，进行数据恢复。具体步骤为：

10 步骤 4.1：备份节点将数据向恢复节点进行传输。在发送端和接收端建立连接后，发送端先将数据进行加密。接收端接收到数据后进行解密操作，并对解密后的数据进行校验值检测。如果通过了校验检测，证明数据传输正确无误，可以将数据写入到磁盘当中。若传输过程中出现错误数据，接收端可以要求发送端对错误数据进行重新发送。

15 步骤 4.2：数据恢复完成后，恢复节点通知管理节点恢复已完成，管理节点通过数据恢复信息生成恢复日志并进行存档，并通知用户数据恢复完成。

步骤 5：日志管理。对每一个备份任务流程和恢复任务流程分别建立一个日志文件，跟踪任务流程的信息：记录周期备份和恢复的时间、任务错误信息等，便于查找任务记录和错误定位。

20 步骤 6：监控管理。备份恢复管理系统提供统一的监控大屏，对整个备份恢复流程进行监控，用户可以实时查看数据备份恢复情况，对系统底层存储空间可用容量、各存储节点状态进行监控。

本发明基于多目标优化的思想，将备份系统的可靠性、时延问题和负载均衡问题抽象成多目标优化问题。以向量的形式同时优化系统 3 个方面的性能。
25 在保持负载均衡的前提下，提升系统的可靠性并使系统时延最低。获取的 Pareto 最优解，使系统能够达到在多个目标下的性能最优，从而提升系统在多个方面的性能表现。

本发明相对现有技术而言，所具有的优点和效果：

5 (1) 本专利发明设计了一种混合云数据备份恢复管理系统，把不同类型的数据和备份空间在一个系统平台上进行对接整合。用户可以在平台上通过管理界面进行操作，根据其需求将本地/云端上不同的类型数据备份到第三方存储软件或者系统底层存储空间，简化用户操作，提高业务效率。

(2) 本专利发明的基于多目标最优副本管理策略的混合云数据备份恢复管理系统，综合考虑备份数据的可靠性、备份系统的时延和备份系统的稳定性，通过基于分解的多目标优化算法，得到黄金副本的备份副本最优放置方案，降低数据丢失和损坏的风险。

10 (3) 本专利发明的基于多目标最优副本管理策略的混合云数据备份恢复管理系统，以图形化的界面进行管理，实时记录日志信息，并且提供统一的监控平台，用户可以对备份恢复策略进行设置，能够灵活、快速的进行数据备份与恢复，帮助管理员快速定位故障，最大程度地提高系统的效率和性能。

15 最后应当说明的是：以上实施例仅用以说明本发明的技术方案而非对其限制，尽管参照上述实施例对本发明进行了详细的说明，所属领域的普通技术人员应当理解：依然可以对本发明的具体实施方式进行修改或者等同替换，而未脱离本发明精神和范围的任何修改或者等同替换，其均应涵盖在本发明的权利要求保护范围之内。

权 利 要 求 书

1. 一种基于多目标最优副本管理策略的数据备份管理方法，其特征在于，包括以下步骤：

5 分别对数据在若干个存储节点中进行完整备份的可靠性目标函数、发送时延目标函数、负载均衡目标函数进行定义；

基于所述可靠性目标函数、所述发送时延目标函数、所述负载均衡目标函数获取数据备份管理的优化目标函数；

采用基于分解的多目标进化算法获取满足所述优化目标函数的最优放置方案矩阵；

10 基于所述最优放置方案矩阵对数据进行各个存储节点上的备份放置分配。

2. 根据权利要求1所述的基于多目标最优副本管理策略的数据备份管理方法，其特征在于，所述可靠性目标函数如下所示：

$$s_1(TK) = \prod_{i=1}^m \left| 1 - \prod_{j=1}^m (tk_{ij} \times e^{-\chi_j u}) \right|$$

15

其中， $s_1(k_i) = 1 - \prod_{j=1}^m (tk_{ij} \times e^{-\chi_j u})$ ，表示第*i*个数据文件 k_i 的可靠性， χ_j 为存储集群*T*中第*j*个存储节点 t_j 的故障率， u 为存储节点的使用时间， $e^{-\chi_j u}$ 为存储节点 t_j 的失效率函数， K 表示数据文件， m 为存储节点数量， T 表示*m*个存储节点组成的集群， tk_v 表示节点 t_j 上的副本。

20 3. 根据权利要求1所述的基于多目标最优副本管理策略的数据备份管理方法，其特征在于，所述发送时延目标函数如下所示：

$$s_2(TK) = \frac{1}{\sum_{i=1}^n \sum_{j=1}^m A_{tol,i}^j} \times \sum_{i=1}^n \left\{ \sum_{j=1}^m \left[Tim\_del_j \times \sum_{a=1}^{A_{tol,i}^j} \left(\frac{a \times c_i}{v_j} \times tk_{ij} \right) \right] \right\}$$

其中, v_j 为备份节点 t_j 的数据发送速率; c_i 为数据文件 k_i 的大小, $A_{tol,i}^j$ 表示 k_i 在节点 t_j 上的总访问量; $s(k_i, a) = Tim\_del_j \times \frac{a \times c_i}{\dots} \times tk_{ij}$, 表示 k_i 的备份副本在存储节点 t_j 上的第 a 个访问的发送时延。

- 5 4. 根据权利要求1所述的基于多目标最优副本管理策略的数据备份管理方法, 其特征在于, 所述负载均衡目标函数如下所示:

$$s_3(TK) = \sum_{j=1}^m \left\| \sigma_j \times \sum_{i=1}^n tk_{ij} - \sum_{i=1}^n tk_{ij} \right\|_2^2$$

其中, σ_j 为每个备份节点 t_j 的权重, 且 $\sum_{j=1}^m \sigma_j = 1$ 。

- 10 5. 根据权利要求1所述的基于多目标最优副本管理策略的数据备份管理方法, 其特征在于, 所述优化目标函数如下所示:

$$\begin{aligned} \max \quad & S(X) = [f_1(X), f_2(X), f_3(X)]^T \\ \text{s.t.} \quad & \sum_{i=1}^n [\rho(k_i) \times tk_{ij}] \leq \Gamma(t_j) \\ \text{s.t.} \quad & \sigma_j \times \sum_{i=1}^n tk_{ij} \geq \sum_{i=1}^n tk_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^m (tk_{ij}) = 1 \end{aligned}$$

其中, X 为决策变量, 此处为放置方案矩阵 TK , $f_1(X) = s_1(TK)$, $f_2(X) = \frac{1}{s_2(TK)}$, $f_3(X) = \frac{1}{s_3(TK)}$, $\rho(k_i)$ 表示第 i 个数据文件 k_i 的大小, $\Gamma(t_j)$ 表示第 j 个存储节点 t_j 的容量大小。

- 15 6. 根据权利要求1所述的基于多目标最优副本管理策略的数据备份管理方法, 其特征在于, 采用基于分解的多目标进化算法获取满足所述优化目标函数

的最优放置方案矩阵的过程包括：

初始化种群数量并随机生成与种群数量相同的权重向量；

基于Tchebycheff分解算法将所述优化目标函数进行分解为与种群数量相同的若干个子问题；

5 初始化进化次数，随机生成种群数量相同的放置方案矩阵；

初始化前沿解集以及优化目标函数对应的参考点；

初始化邻域向量，获取距离最后一个权重向量最近的若干个权重向量，并基于若干个权重向量构建索引集；

10 设置最大迭代次数并进行迭代，基于迭代次数对最后一个权重向量和对应的放置方案矩阵进行迭代更新，当达到最大迭代次数时停止迭代，获取所述最优放置方案矩阵。

7. 根据权利要求6所述的基于多目标最优副本管理策略的数据备份管理方法，其特征在于，基于迭代次数对每个权重向量和对应的放置方案矩阵进行更新的过程包括：

15 基于所述索引集随机获取索引号，获取所述索引号对应的放置方案矩阵；

基于对应的放置方案矩阵获取新放置方案矩阵，基于所述新放置方案矩阵分别对参考点、种群以及非支配解集进行迭代更新；

当达到最大迭代次数时，获取更新后的非支配解集并作为最优前沿解集，在所述最优前沿解集中选择一个权衡解，将所述权衡解作为最优放置方案矩阵。

20 8. 根据权利要求7所述的基于多目标最优副本管理策略的混合云数据备份管理方法，其特征在于，所述参考点的更新过程中，若 $z_k^* < f_k(X_h^{p(new)})$ ，则 $z_k^* = f_k(X_h^{p(new)})$ ， z_k^* 表示子问题 f_k 的参考点， $X_h^{p(new)}$ 表示新放置方案矩阵；

所述种群的更新过程中，对于索引集中的每一个邻域个体 X_h^{r*} ，若 $g^{te}(X_h^{p(new)}|\lambda, z^*) \leq g^{te}(X_h^b|\lambda, z^*)$ ，则 $X_h^b = X_h^{p(new)}$ ，其中 g^{te} 表示基于
25 Tchebycheff分解算法得到的子问题， λ 表示邻域个体 X_h^b 的权重向量。

9. 根据权利要求7所述的基于多目标最优副本管理策略的混合云数据备份

管理方法，其特征在于，所述非支配解集的更新过程中，保留 $INL(h-1)$ 中支配 $X_h^{p(new)}$ 的解，舍弃 $INL(h-1)$ 中所有被 $X_h^{p(new)}$ 支配的解，若所有 $INL(h-1)$ 的解都不支配 $X_h^{p(new)}$ ，将 $X_h^{p(new)}$ 加入 $INL(h)$ ，其中， $INL(h-1)$ 表示上一次更新后的非支配解集， $X_h^{p(new)}$ 表示新放置方案矩阵， $INL(h)$ 表示当前的非支配解集。

- 5 10. 一种基于多目标最优副本管理策略的混合云数据备份管理系统，其特征在于，所述系统包括存储器、处理器及存储在存储器上并能够在处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述权利要求 1 至 9 任一项所述的方法。

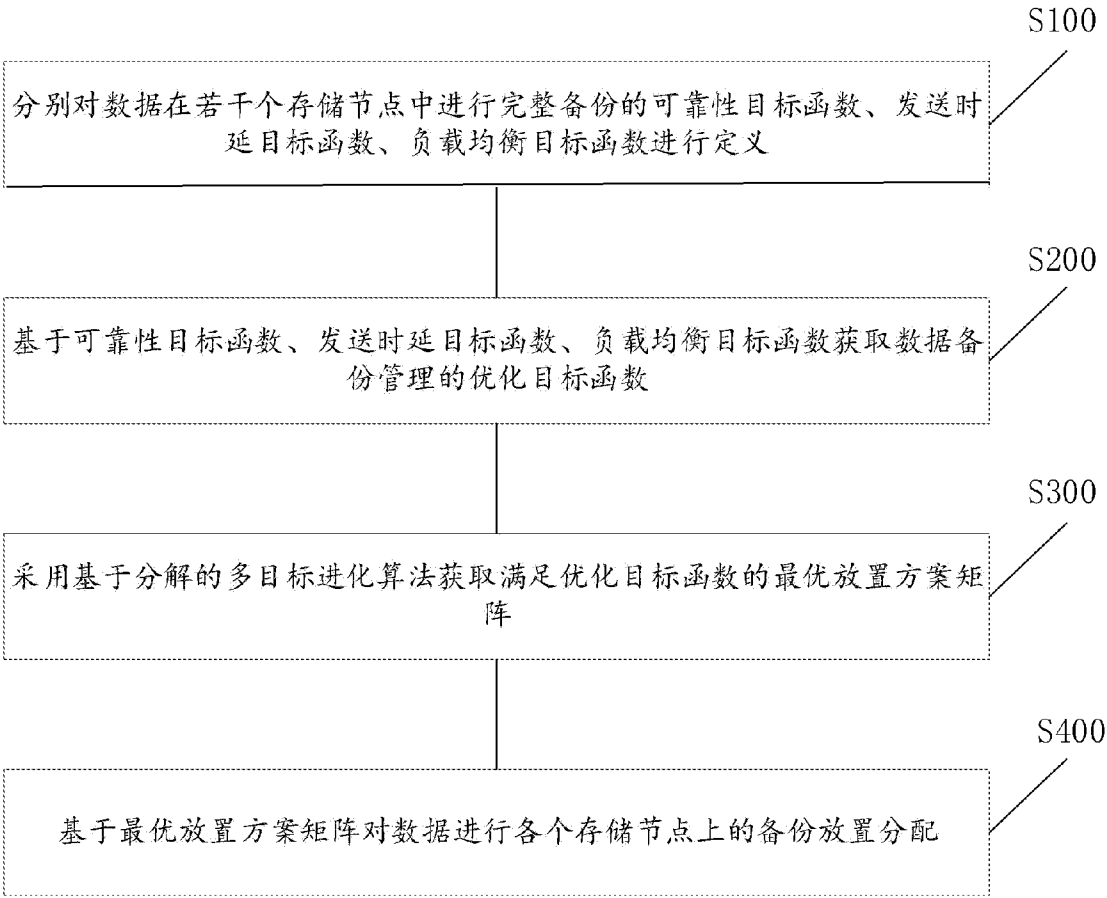


图 1

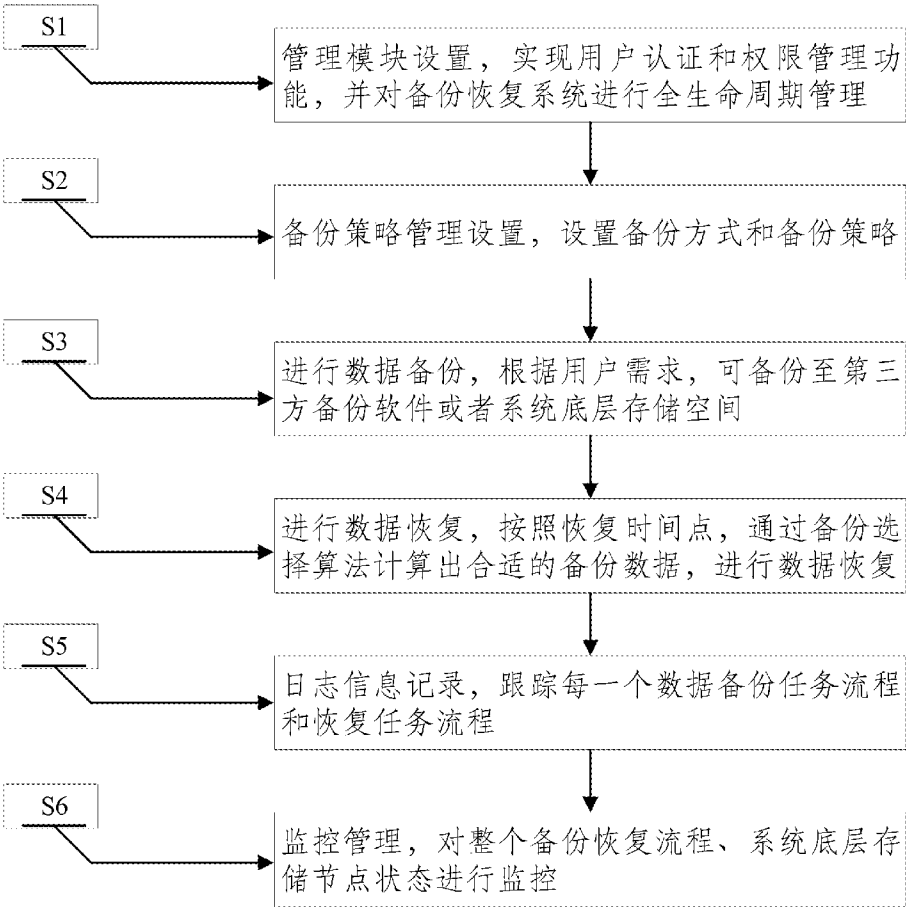


图 2

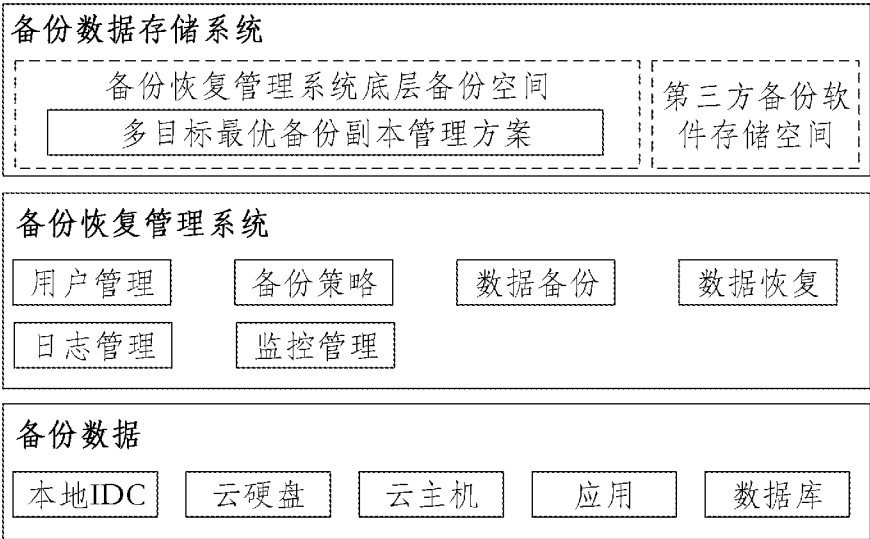


图 3

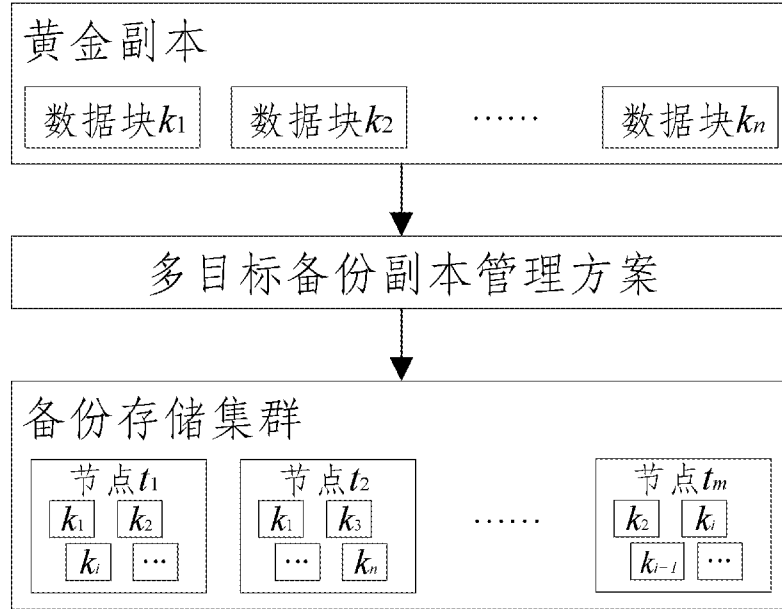


图 4

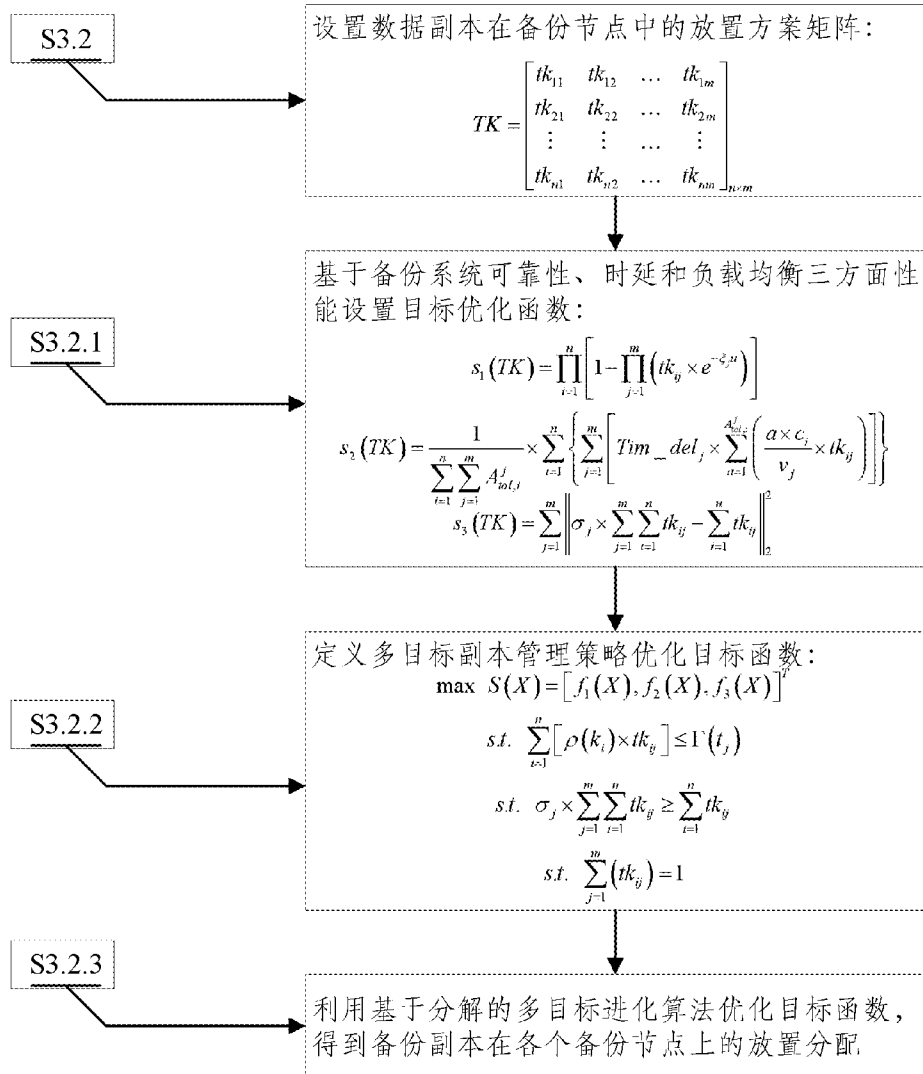


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/143183

A. CLASSIFICATION OF SUBJECT MATTER

G06F11/14(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 混合云, 数据, 备份, 最优, 副本, 策略, 函数, 时延, 负载均衡, 矩阵, 放置, hybrid, cloud, data, backup, optimal, copy, strategy, function, time, delay, load, balancing, matrix, placement

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 109889573 A (WUHAN UNIVERSITY OF TECHNOLOGY) 14 June 2019 (2019-06-14) claims 1-10, and description, paragraphs [0047]-[0110]	1, 10
A	CN 111124762 A (AEROSPACE KEGONG NETWORK INFORMATION DEVELOPMENT CO., LTD.) 08 May 2020 (2020-05-08) entire document	1-10
A	CN 111859703 A (JINAN UNIVERSITY) 30 October 2020 (2020-10-30) entire document	1-10
A	US 2015347430 A1 (INTERNATIONAL BUSINESS MACHINES CORP.) 03 December 2015 (2015-12-03) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

31 July 2024

Date of mailing of the international search report

13 August 2024

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/143183

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109889573	A	14 June 2019	None			
CN	111124762	A	08 May 2020	None			
CN	111859703	A	30 October 2020	None			
US	2015347430	A1	03 December 2015	US	2015363413	A1	17 December 2015

A. 主题的分类

G06F11/14(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 混合云, 数据, 备份, 最优, 副本, 策略, 函数, 时延, 负载均衡, 矩阵, 放置, hybrid, cloud, data, backup, optimal, copy, strategy, function, time, delay, load, balancing, matrix, placement

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
X	CN 109889573 A (武汉大学) 2019年6月14日 (2019 - 06 - 14) 权利要求1-10, 说明书第[0047]-[0110]段	1、10
A	CN 111124762 A (航天科工网络信息发展有限公司) 2020年5月8日 (2020 - 05 - 08) 全文	1-10
A	CN 111859703 A (暨南大学) 2020年10月30日 (2020 - 10 - 30) 全文	1-10
A	US 2015347430 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2015年12月3日 (2015 - 12 - 03) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

★ 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

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

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“p” 公布日先于国际申请日但迟于所要求的优先权日的文件

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

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

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

“&” 同族专利的文件

国际检索实际完成的日期

2024年7月31日

国际检索报告邮寄日期

2024年8月13日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

张涛

电话号码 (+86) 010-53961356

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/143183

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	109889573	A	2019年6月14日	无			
CN	111124762	A	2020年5月8日	无			
CN	111859703	A	2020年10月30日	无			
US	2015347430	A1	2015年12月3日	US	2015363413	A1	2015年12月17日