

(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



(10) 国际公布号
WO 2025/140269 A1

(43) 国际公布日
2025 年 7 月 3 日 (03.07.2025)

(51) 国际专利分类号:
G06F 9/455 (2018.01)

哲(GONG, Zhe); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。

(21) 国际申请号: PCT/CN2024/142182

(74) 代理人: 北京亿腾知识产权代理事务所(普通合伙)(E-TONE INTELLECTUAL PROPERTY FIRM (GENERAL PARTNERSHIP)); 中国北京市海淀区中关村紫金数码园3号楼707 100190 (CN)。

(22) 国际申请日: 2024 年 12 月 25 日 (25.12.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202311797508.4 2023年12月25日 (25.12.2023) CN
202410256871.3 2024年3月6日 (06.03.2024) CN

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

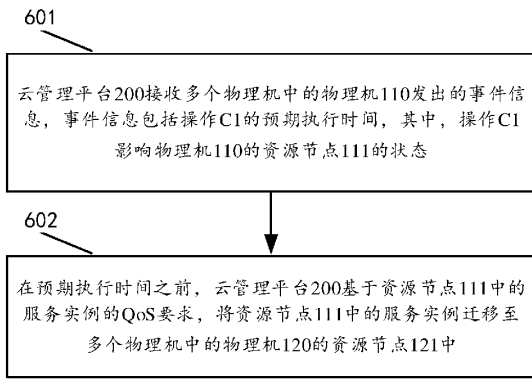
(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR,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

(72) 发明人: 叶会(YE, Hui); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。
曹昱(CAO, Yu); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。 龚

(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,

(54) Title: SERVICE INSTANCE MANAGEMENT METHOD BASED ON CLOUD TECHNOLOGY, CLOUD MANAGEMENT PLATFORM, AND CLUSTER

(54) 发明名称: 基于云技术的服务实例管理方法、云管理平台及集群



601 A cloud management platform 200 receives event information sent by a physical machine 110 among a plurality of physical machines, wherein the event information comprises an expected execution time of an operation C1, and the operation C1 affects the state of a resource node 111 of the physical machine 110
602 Before the expected execution time, the cloud management platform 200 migrates a service instance in the resource node 111 to a resource node 121 of a physical machine 120 among the plurality of physical machines on the basis of a QoS requirement of the service instance in the resource node 111

图 6

(57) Abstract: Provided in the present application are a service instance management method based on cloud technology, a cloud management platform, and a cluster. The method comprises: a cloud management platform receiving event information sent by a first physical machine among a plurality of physical machines, wherein the event information comprises an expected execution time of an operation, and the operation affects the state of a first resource node of the first physical machine; and before the expected execution time, the cloud management platform migrating a service instance in the first resource node to a second resource node of a second physical machine among the plurality of physical machines on the basis of a QoS requirement of the service instance in the first resource node. By means of the method, the QoS of a service instance can be prevented from being affected by an operation for a resource node where the service instance is located.

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))。

(57) 摘要: 本申请提供了一种基于云技术的服务实例管理方法、云管理平台及集群。该方法包括: 云管理平台接收多个物理机中的第一物理机发出的事件信息, 事件信息包括操作的预期执行时间, 其中, 操作影响第一物理机的第一资源节点的状态; 在预期执行时间之前, 云管理平台基于第一资源节点中的服务实例的QoS要求, 将第一资源节点中的服务实例迁移至多个物理机中的第二物理机的第二资源节点中。该方法可避免服务实例的QoS受该服务实例所在资源节点的操作的影响。