

(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
2024 年 3 月 7 日 (07.03.2024)



(10) 国际公布号
WO 2024/046188 A1

- (51) 国际专利分类号:
G06F 9/455 (2018.01) **G06F 9/50** (2006.01)
- (21) 国际申请号: PCT/CN2023/114511
- (22) 国际申请日: 2023 年 8 月 23 日 (23.08.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202211060455.3 2022年8月30日 (30.08.2022) CN
- (71) 申请人: 阿里巴巴(中国)有限公司(**ALIBABA (CHINA) CO., LTD.**) [CN/CN]; 中国浙江省杭州市余杭区五常街道文一西路 969 号 3 幢 5 层 554 室, Zhejiang 311121 (CN)。
- (72) 发明人: 巩小东 (**GONG, Xiaodong**); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号云谷园区, Zhejiang 310030 (CN)。
- (74) 代理人: 北京同钧律师事务所(**BEIJING TONGJUN LAW FIRM**); 中国北京市海淀区西直门北大街 32 号院 1 号楼 7 层 808 室, Beijing 100082 (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,

(54) **Title:** I/O UNLOADING METHOD AND SYSTEM IN CLOUD ENVIRONMENT, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种云环境下的 I/O 卸载方法、设备、系统及存储介质

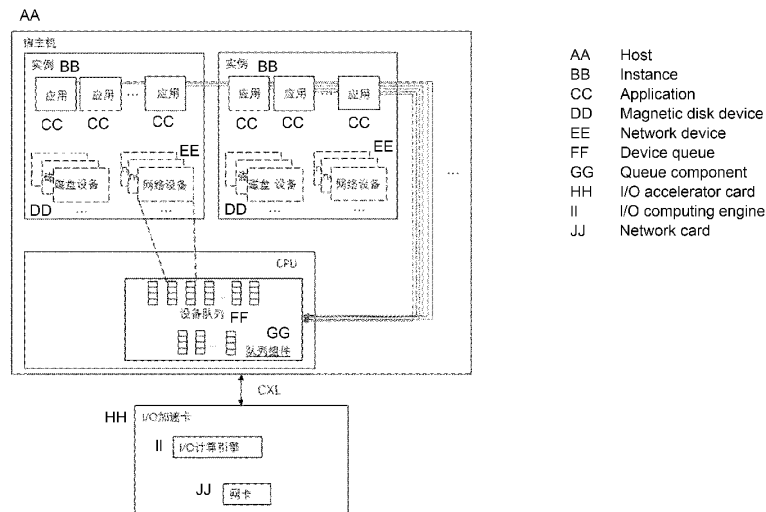


图 1

(57) **Abstract:** Embodiments of the present application provide an I/O unloading method and system in a cloud environment, and a device and a storage medium. In the embodiments of the present application, a queue component is additionally provided in a CPU of a host, and it is innovatively proposed that the queue component is used in an I/O processing process. On this basis, virtual I/O devices and device queues corresponding to the virtual I/O devices can be provided for instances on the host by means of the queue component; in the CPU, I/O requests generated between the instances and the virtual I/O devices are scheduled by means of the device queues;

[见续页]

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))。

and an original I/O unloading card only needs to be responsible for monitoring the I/O requests in the device queues and transmitting data related to the I/O requests. Therefore, the device queues in the I/O processing process can be moved up into the CPU, so that the device queues can be linked with each component in the CPU in an I/O scheduling process to obtain the real-time load of each core in the CPU, and I/O scheduling can be realized more reasonably without blind scheduling, thereby effectively improving the I/O processing efficiency.

(57) 摘要: 本申请实施例提供一种云环境下的I/O卸载方法、设备、系统及存储介质。在本申请实施例中, 宿主机的CPU中增配了队列组件, 且创新性地提出将队列组件用于I/O的处理过程中。基于此, 可通过队列组件为宿主主机上的实例提供虚拟I/O设备以及与虚拟I/O设备对应的设备队列; 在CPU内, 利用设备队列对实例与虚拟I/O设备之间发生的I/O请求进行调度; 而原本的I/O卸载卡则只需负责监听设备队列中的I/O请求并传递I/O请求所相关的数据即可。这样, 可将I/O处理过程中的设备队列上移到CPU内, 从而在I/O调度过程中可与CPU内各部件联动, 获取到CPU内各核心的实时负载, 从而可更加合理地实现I/O调度, 而不再盲目调度, 这可有效提高I/O的处理效率。