(12) International Application Status Report

Received at International Bureau: 08 February 2018 (08.02.2018)
Information valid as of: 13 July 2018 (13.07.2018)
Report generated on: 22 September 2019 (22.09.2019)

(10) Publication number:
WO2018/144292

(43) Publication date:
09 August 2018 (09.08.2018)

(26) Publication language:
English (EN)

(21) Application Number:
PCT/US2018/015119

(22) Filing Date:
25 January 2018 (25.01.2018)

(25) Filing language:
English (EN)

(31) Priority number(s):
15/423,276 (US)

(31) Priority date(s):
02 February 2017 (02.02.2017)

(31) Priority status:
Priority document received (in compliance with PCT Rule 17.1)

(51) International Patent Classification:
G06F 9/455 (2018.01); ***G06F 9/50*** (2006.01)

(71) Applicant(s):
MICROSOFT TECHNOLOGY LICENSING, LLC [US/US]; One Microsoft Way Redmond, Washington 98052-6399 (US) *(for all designated states)*

(72) Inventor(s):
HOPPERT, Hadden Mark; Microsoft Technology Licensing, LLC One Microsoft Way Redmond, Washington 98052-6399 (US)
HUYBREGTS, Christopher L.; Microsoft Technology Licensing, LLC One Microsoft Way Redmond, Washington 98052-6399 (US)
OSHINS, Jacob Kappeler; Microsoft Technology Licensing, LLC One Microsoft Way Redmond, Washington 98052-6399 (US)

(74) Agent(s):
MINHAS, Sandip, S.; Microsoft Technology Licensing, LLC One Microsoft Way Redmond, Washington 98052-6399 (US)

(54) Title (EN): GRAPHICS PROCESSING UNIT PARTITIONING FOR VIRTUALIZATION

(54) Title (FR): PARTITIONNEMENT D'UNITÉ DE TRAITEMENT GRAPHIQUE POUR VIRTUALISATION

(57) Abstract:
(EN): Techniques for graphics processing unit (GPU) partitioning for virtualization are described herein. In one or more implementations, a GPU partitioning manager of a host device obtains a request for a virtual machine having GPU functionality. In particular, the request specifies the GPU functionality in terms of different GPU capabilities. These different capabilities correspond to segments of a GPU model that represents GPU functionality and is used to govern interactions between virtual machines and GPUs. The GPU partitioning manager determines whether GPUs of the host device are available to satisfy the request based on the specified capabilities. If so, the GPU partitioning manager allocates a portion of the determined available GPUs to the virtual machine to configure the virtual machine with a GPU partition having the functionality. The virtual machine configured with the GPU partition can then be exposed to provide GPU-processed data to a GPU partition requestor.

(FR): L'invention concerne des techniques de partitionnement d'unité de traitement graphique (GPU) pour virtualisation. Dans un ou plusieurs modes de réalisation, un gestionnaire de partitionnement de GPU d'un dispositif hôte obtient une demande pour une machine virtuelle ayant une fonctionnalité de GPU. En particulier, la demande spécifie la fonctionnalité de GPU en termes de capacités de GPU différentes. Ces différentes capacités correspondent à des segments d'un modèle de GPU qui représente une fonctionnalité de GPU et est utilisé pour gouverner des interactions entre des machines virtuelles et des GPU. Le gestionnaire de partitionnement de GPU détermine si des GPU du dispositif hôte sont disponibles pour satisfaire la requête sur la base des capacités spécifiées. Si tel est le cas, le gestionnaire de partitionnement de GPU attribue une partie du GPUs disponible déterminé à la machine virtuelle pour configurer la machine virtuelle avec une partition GPU ayant la fonctionnalité. La machine virtuelle configurée avec la partition GPU peut ensuite être exposée pour fournir des données traitées par GPU à un demandeur de partition GPU.

International search report:

Received at International Bureau: 23 April 2018 (23.04.2018) [EP]

International Report on Patentability (IPRP) Chapter II of the PCT:
Not available

(81) Designated States:
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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW
European Patent Office (EPO) : 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
African Intellectual Property Organization (OAPI) : BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG
African Regional Intellectual Property Organization (ARIPO) : BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW
Eurasian Patent Organization (EAPO) : AM, AZ, BY, KG, KZ, RU, TJ, TM

Declarations:
Declaration made as applicant's entitlement, as at the international filing date, to apply for and be granted a patent (Rules 4.17(ii) and 51bis.1(a)(ii)), in a case where the declaration under Rule 4.17(iv) is not appropriate
Declaration made as applicant's entitlement, as at the international filing date, to claim the priority of the earlier application, where the applicant is not the applicant who filed the earlier application or where the applicant's name has changed since the filing of the earlier application (Rules 4.17(iii) and 51bis.1(a)(iii))