

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2025年1月9日(09.01.2025)



(10) 国際公開番号

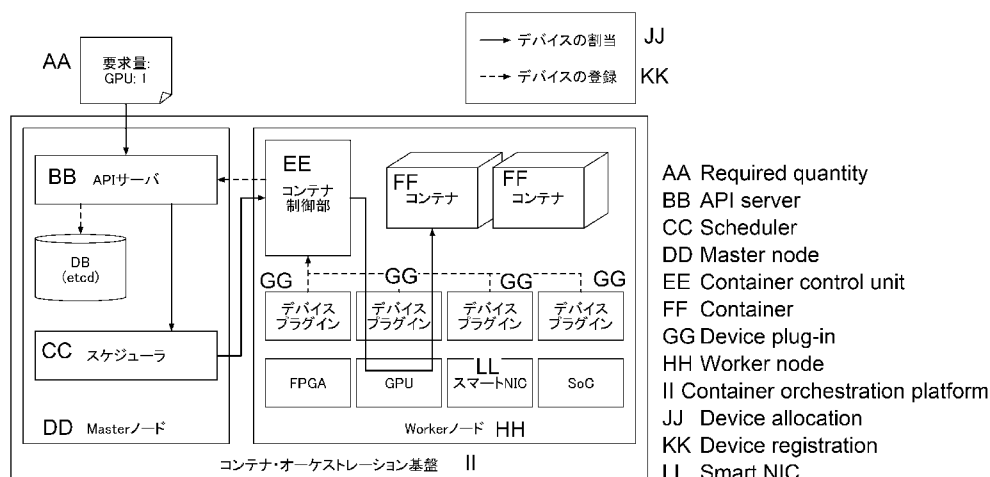
WO 2025/009152 A1

- (51) 国際特許分類:
G06F 9/50 (2006.01) *G06F 8/60* (2018.01) **Hiroki**; 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT知的財産センタ内 Tokyo (JP).
- (21) 国際出願番号: PCT/JP2023/025113 (74) 代理人: 伊東 忠重, 外 (**ITO**H, **Tadashige** et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 Tokyo (JP).
- (22) 国際出願日: 2023年7月6日(06.07.2023)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 日本電信電話株式会社 (**NIPPON TELEGRAPH AND TELEPHONE CORPORATION**) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 平井 志久 (**HIRAI**, **Shiku**); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT知的財産センタ内 Tokyo (JP). 馬場 宏基 (**BABA**,

(54) **Title:** DEVICE ALLOCATION CONTROL DEVICE, DEVICE ALLOCATION CONTROL METHOD, AND PROGRAM

(54) 発明の名称: デバイス割当制御装置、デバイス割当制御方法及びプログラム

[図1]



(57) **Abstract:** The present invention receives input information indicating the required performance of an application and whether the application is of a performance-guaranteed type. On the basis of information indicating device configuration and topology in a plurality of container orchestration platforms, information for estimating the performance of an application, and information indicating a device allocated to a container that has already launched an application and whether the application is of the performance-guaranteed type, and using, as constraint conditions, information that indicates the respective use states of devices and the feature in which a device to be allocated to a performance-guaranteed-type application container is not allocated to two or more containers, the present invention determines a device allocated to a container that launches an application, and determines an allocation quantity for the device, so that a cost calculated by

[続葉有]

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(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (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).

添付公開書類：

一 国際調査報告 (条約第21条(3))

an optimization algorithm from the allocation quantity for each device is minimized. The present invention instructs a container orchestration platform pertaining to a device allocated to the container to allocate the device.

(57) 要約：アプリケーションの要求性能と性能保証型であるか否かを示す入力情報を受け付け、複数のコンテナ・オーケストレーション基盤におけるデバイスの構成及びトポロジを示す情報、アプリケーションの性能を推定する情報、及び既にアプリケーションを起動しているコンテナに割り当てられているデバイスと当該アプリケーションが性能保証型であるか否かを示す情報に基づいて、デバイスそれぞれの使用状況を示す情報、及び性能保証型のアプリケーション・コンテナに割り当てるデバイスは2以上のコンテナに割り当てないということを制約条件として、デバイスごとの割当量から最適化アルゴリズムによって算出されるコストを最小化するように、アプリケーションを起動するコンテナに割り当てるデバイス及び割当量を決定し、当該コンテナに対して割り当てられるデバイスに係るコンテナ・オーケストレーション基盤に対し、当該デバイスの割り当てを指示する。

明 細 書

発明の名称：

デバイス割当制御装置、デバイス割当制御方法及びプログラム

技術分野

[0001] 本発明は、デバイス割当制御装置、デバイス割当制御方法及びプログラムに関する。

背景技術

[0002] K u b e r n e t e s（非特許文献1）をデファクト・スタンダードとするコンテナ・オーケストレーション基盤上では、コンテナ内のアプリケーションロジックの処理をハードウェア上にオフロードして高速化するため、G P U、F P G A、スマートN I C（例えば、R D M A対応N I C）、及びS o Cのようなアクセラレータのコンテナへの割り当てが可能になっている。

[0003] コンテナ・オーケストレーション基盤上のコンテナ制御部は、コンテナへの様々なデバイス（ハードウェアリソース）の割り当てを可能とするメタプラグイン機能（非特許文献2）を有する。デバイス毎の専用のデバイス・プラグインをこのメタプラグイン機能と連携させることで、デバイスの登録及び管理が可能となる。

[0004] コンテナにデバイスを割り当てる際には、図1に示されるように、ユーザからのアプリケーションのデプロイマニフェストをM a s t e rノード内のA P Iサーバを経由して当該M a s t e rノード内のスケジューラ（非特許文献3）が受信する。スケジューラは、デプロイマニフェスト中のデバイスの要求量に基づき、割り当て先のW o r k e rノード（物理サーバ又は仮想マシン）を決定する。当該W o r k e rノード内のコンテナ制御部は割り当て可能なデバイスを選択し、コンテナへのデバイスの割り当てが実行される。

[0005] なお、1つのノード（M a s t e rノード又はW o r k e rノード）は、1つの物理サーバ又は1つの論理サーバに相当する。

先行技術文献

非特許文献

[0006] 非特許文献1: "What is Kubernetes, Kubernetes", [online]、インターネット<URL: <https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/>>

非特許文献2: "Device Plugins, Kubernetes", [online]、インターネット<URL: <https://kubernetes.io/docs/concepts/extend-kubernetes/compute-storage-net/device-plugins/>>

非特許文献3: "Kubernetes Scheduler, Kubernetes", [online]、インターネット<URL: <https://kubernetes.io/docs/concepts/scheduling-eviction/kube-scheduler/>>

非特許文献4: CUDA、"Developing a Linux Kernel Module using GPUDirect RDMA", [online]、インターネット<URL: <https://docs.nvidia.com/cuda/gpudirect-rdma/>>

発明の概要

発明が解決しようとする課題

[0007] 現行のコンテナへのデバイスのスケジューリング方式では、ユーザからの単純なデバイスの要求量に基づき、割り当て可能なリソース量に余裕のある Worker ノード及びデバイス（ハードウェアリソース）が決定されている。

[0008] 今後は、5Gの発展などにより、ハードウェアリソースが潤沢なクラウド上のみならず、エッジ上でのアプリケーションの展開が想定される。したがって、コンテナに対して、アプリケーションの要求性能を満たす適切なデバイス（ハードウェアリソース）の割り当てが課題となる。

[0009] 本発明は、上記の点に鑑みてなされたものであって、アプリケーションの要求性能を満たすようにコンテナに対してデバイスを割り当てることを可能とすることを目的とする。

課題を解決するための手段

[0010] そこで上記課題を解決するため、デバイス割当制御装置は、アプリケーションの要求性能と性能保証型であるか否かとを示す入力情報を受け付けるように構成されている要求性能受付部と、複数のコンテナ・オーケストレーション基盤におけるデバイスの構成及びトポロジを示す情報、前記アプリケーションの性能を推定する情報、及び既にアプリケーションを起動しているコンテナに割り当てられている前記デバイスと当該アプリケーションが性能保証型であるか否かとを示す情報に基づいて、前記デバイスそれぞれの使用状況を示す情報、及び前記要求性能及び性能保証型のアプリケーション・コンテナに割り当てるデバイスは2以上のコンテナに割り当てないということを制約条件として、デバイスごとの割当量から最適化アルゴリズムによって算出されるコストを最小化するように、前記アプリケーションを起動するコンテナに割り当てる前記デバイス及び割当量を決定するように構成されているデバイス割当決定部と、前記アプリケーションを起動するコンテナに対して割り当てられる前記デバイスに係るコンテナ・オーケストレーション基盤に対し、当該デバイスの割り当てを指示するように構成されているデバイス割当指示部と、を有する。

発明の効果

[0011] アプリケーションの要求性能を満たすようにコンテナに対してデバイスを割り当てることを可能とすることができる。

図面の簡単な説明

[0012] [図1]コンテナへのデバイスの割り当てを説明するための図である。

[図2]第1の実施の形態におけるデバイス割当制御システムの機能構成例を示す図である。

[図3]第1の実施の形態におけるヘテロジニアス・コンピューティング設計制御部1として機能するコンピュータのハードウェア構成例を示す図である。

[図4]第1の実施の形態におけるデバイス割当制御システムにおいて実行される処理手順の一例を説明するための図である。

[図5]第2の実施の形態におけるデバイス割当制御システムの機能構成例を示す図である。

[図6]第2の実施の形態におけるデバイス割当制御システムにおいて実行される処理手順の一例を説明するための図である。

発明を実施するための形態

[0013] 本実施の形態では、アプリケーションごとの要求性能とデバイス構成及びトポロジ構成情報を組み合わせた最適なハードウェアリソース設計及び制御が行われる。また、トランスポート装置制御連携部17を構築することで、オフロードされたデータが複数デバイス間を跨ぎ、かつ、物理又は論理サーバ間を越える場合に、ルータ又はスイッチ等のトランスポート装置リソースプール4を管理するトランスポート制御装置管理部3との連携が行われる。

[0014] 以下、図面に基づいて本発明の実施の形態を説明する。図2は、第1の実施の形態におけるデバイス割当制御システムの機能構成例を示す図である。図2において、複数のコンテナ・オーケストレーション基盤2のそれぞれは、ヘテロジニアス・コンピューティング設計制御部1と通信可能である。ヘテロジニアス・コンピューティング設計制御部1は、トランスポート制御装置管理部3と通信可能である。トランスポート制御装置管理部3は、トランスポート装置リソースプール4と通信可能である。

[0015] なお、コンテナ・オーケストレーション基盤2、ヘテロジニアス・コンピューティング設計制御部1及びトランスポート制御装置管理部3は、同じコンピュータを用いて実現されてもよい。又は、これら3つのうちのいずれかが異なるコンピュータを用いて実現されてもよい。又は、これら3つのそれぞれが異なるコンピュータを用いて実現されてもよい。

[0016] コンテナ・オーケストレーション基盤2は、例えば、K u b e r n e t e sを例とするようなコンテナ・オーケストレーション基盤である。コンテナ・オーケストレーション基盤2は、図1に示したように、M a s t e rノード及びW o r k e rノードを含む。M a s t e rノードは、A P Iサーバ21を含む。W o r k e rノードは、1以上のコンテナを含む。各ノードのそれ

それは、異なる物理サーバ又は論理サーバ上に構成される。各ノードが論理サーバ上に構成される場合、各ノードは同じ物理サーバ上に構成されてもよい。

[0017] トランスポート装置リソースプール4は、コンテナの配備先の候補としての物理又は論理サーバ間を接続するネットワーク内のリソース（転送系のルータ又はスイッチ、伝送系の光伝送装置、及びそれら装置が管理する情報（VLAN IDや光波長等））で構成されるトランスポート装置及び情報群である。

[0018] トランスポート制御装置管理部3は、トランスポート装置リソースプール4内のリソース（ルータやスイッチ等）を管理し、物理又は論理サーバ間を跨いでデバイス間を接続する場合の、物理又は論理サーバ間の仮想専用ネットワークを提供するための制御を行う。

[0019] ヘテロジニアス・コンピューティング設計制御部1は、要求性能受付部11、デバイス情報取得部12、性能推定モデル取得部13、デバイス割当決定部14、デバイス情報監視部15、デバイス割当指示部16、トランスポート装置制御連携部17含む。これら各部は、同じコンピュータを用いて実現されてもよいし、一部又は全部が異なるコンピュータを用いて実現されてもよい。

[0020] 要求性能受付部11は、ユーザがコンテナ上で実行させたい1以上のアプリケーション（以下、「対象アプリケーション」という。）に関するアプリケーション要求性能プロファイルを受け付け、当該のアプリケーション要求性能プロファイルをデバイス割当決定部14に送信する。アプリケーション要求性能プロファイルとは、1以上のアプリケーションごとの要求性能情報が記載されたプロファイルをいう。要求性能情報は、例えば、内部処理遅延時間、スループット、及びその他アプリ固有の性能指標を含む。内部処理遅延時間は、平均遅延や許容遅延（Tail遅延、パーセンタイル遅延）等であり、いずれの指標であるかについてはアプリケーションごとに異なってもよい。

- [0021] デバイス情報取得部 1 2 は、デバイス割当決定部 1 4 からの要求に応じ、各コンテナ・オーケストレーション基盤 2 において配備及び管理されているデバイス（ハードウェアリソース）の構成やトポロジ（接続関係）を示す情報（以下、「デバイス情報」という。）をデバイス情報記憶部 1 2 1 から取得して、取得したデバイス情報をデバイス割当決定部 1 4 に送信する。デバイス情報記憶部 1 2 1 は、デバイス情報を記憶する。デバイス情報は、例えば、デバイス種別及び識別子、デバイスを配備した物理又は論理サーバの種別及び識別子、内部トポロジ構成、物理又は論理サーバの配置拠点、拠点内又は拠点間の物理又は論理ネットワーク接続情報を含む。
- [0022] 性能推定モデル取得部 1 3 は、デバイス割当決定部 1 4 からの要求に応じ、アプリケーションごとに作成された、アプリケーションの各性能指標を推定するモデルを示す情報（以下、「モデル情報」という。）を性能推定モデルリポジトリ 1 2 2 から取得して、取得したモデル情報をデバイス割当決定部 1 4 に送信する。性能推定モデルリポジトリ 1 2 2 は、アプリケーションごとに作成されたモデル情報を記憶する。
- [0023] デバイス情報監視部 1 5 は、デバイス割当決定部 1 4 からの要求に応じ、各コンテナ・オーケストレーション基盤 2 において配備及び管理されているデバイス（各 W o r k e r ノードにおいて配備及び管理されている、GPU、FPGA、スマートNIC（例えば、RDMA 対応 NIC）、及び SoC 等のハードウェアリソース）の使用状況に関する情報（以下、「デバイス使用状況情報」という。）を各コンテナ・オーケストレーション基盤 2 から取得し、取得したデバイス使用状況情報をデバイス割当決定部 1 4 に送信する。なお、デバイス使用状況情報とは、デバイスごとに、当該デバイスが特定のコンテナに既に割り当てられているか否かを示す情報や、当該デバイスの使用率等を含む情報である。
- [0024] デバイス割当決定部 1 4 は、要求性能受付部 1 1、デバイス情報取得部 1 2、性能推定モデル取得部 1 3 及びデバイス情報監視部 1 5 から取得した情報に基づいて、各対象アプリケーションを起動する各コンテナ（以下、「ア

アプリケーション・コンテナ」という。)の配備先のコンテナ・オーケストレーション基盤2と、アプリケーション・コンテナの配備先の物理又は論理サーバ(Workerノード)と、配備先の物理又は論理サーバごとの割当デバイス種別、割当デバイス識別子及び割当デバイス数量とを含む情報(以下、「デバイス割当情報」という。)を決定(計算)する。デバイス割当決定部14は、決定(計算)したデバイス割当情報をデバイス割当指示部16へ通知する。

[0025] デバイス割当情報を決定(計算)するためのロジックについては、グリッドサーチ又は既存の最適化アルゴリズムを利用することが考えられる。一例として、デバイス割当決定部14は、デバイス情報及びモデル情報を用いて、デバイス(CPU、メモリ、その他デバイス等)ごとの割当量の大量の入力パターン(すなわち、配備先の物理又は論理サーバごとの割当デバイス種別、割当デバイス識別子及び割当デバイス数量の大量のパターン)について、配備対象のアプリケーション・コンテナに係るアプリケーションの各性能指標の推定値を導出する。デバイス割当決定部14は、導出した各性能推定値に対して、アプリケーション要求性能プロファイルやデバイス使用状況情報を制約条件として、デバイス(CPU、メモリ、その他デバイス等)ごとの割当量から最適化アルゴリズムによって算出されるコスト(設備コストや電力消費量)を目的関数として、当該制約条件を満たしつつ当該目的関数を最小化する入力パターンを探索する。最適化計算結果から、同一のアプリケーション・コンテナを複数拠点に分散配備する場合(コンテナのレプリカ数が複数の場合)や、単一アプリケーション・コンテナに複数のデバイスを割り当てる場合、複数の配備先(デバイス制御を指示する先)が決定される。

[0026] デバイス割当決定部14は、また、単一のアプリケーション・コンテナに複数のデバイスを割り当て、かつ、物理又は論理サーバ間を跨いで複数デバイス間を接続することを決定した場合(単一のコンテナに割り当てられるデバイスが複数の物理又は論理サーバに跨がる場合)に仮想専用ネットワークの提供をトランスポート装置制御連携部17に指示する。デバイス割当決定

部 1 4 は、トランスポート装置制御連携部 1 7 への指示に際し、接続元のデバイス及び接続先のデバイスのそれぞれのデバイス情報をトランスポート装置制御連携部 1 7 へ送信する。本実施の形態では、コンテナに割り当て可能なデバイスは、当該コンテナが起動した W o r k e r ノード上に配備されるデバイスに制限されず、他クラスタ及び他 W o r k e r ノード上のデバイスも割り当て可能である。なお、複数デバイス間の接続が物理又は論理サーバ間を跨がない場合、デバイス割当決定部 1 4 は、仮想専用ネットワークを提供しないことをトランスポート装置制御連携部 1 7 に指示する。

[0027] デバイス割当指示部 1 6 は、デバイス割当決定部 1 4 からの通知に基づき、各配備先のコンテナ・オーケストレーション基盤 2 の A P I サーバ 2 1 に対し、対象の各物理又は論理サーバ（配備先のコンテナ・オーケストレーション基盤 2 が管理する W o r k e r ノード）内のデバイス種別、デバイス識別子及びデバイス数量についての予約及び割り当てを指示する。同一アプリケーション・コンテナを複数拠点に分散配備する場合（コンテナのレプリカ数が複数の場合）や、単一アプリケーション・コンテナに複数のデバイスを割り当てる場合、デバイス割当指示部 1 6 は、各配備先のコンテナ・オーケストレーション基盤 2 の A P I サーバ 2 1 に対して指示を行う。

[0028] トランスポート装置制御連携部 1 7 は、デバイス割当決定部 1 4 からの指示（接続元のデバイス及び接続先のデバイスのそれぞれのデバイス情報）に基づき、物理又は論理サーバ間を跨いでデバイス間を接続する場合の仮想専用ネットワークの提供をトランスポート制御装置管理部 3 に指示する。この際、トランスポート装置制御連携部 1 7 は、トランスポート制御装置管理部 3 に対して、接続元のデバイス及び接続先のデバイスのそれぞれのデバイス情報を併せて送信する。トランスポート制御装置管理部 3 は、当該デバイス情報に基づき、デバイス間を接続する複数の仮想専用ネットワークの経路候補のうち、「遅延が最小になる経路」を探索し、該当する経路で仮想専用ネットワークを生成するよう、トランスポート装置リソースプール 4 内のリソースを制御する。

[0029] 図3は、第1の実施の形態におけるヘテロジニアス・コンピューティング設計制御部1として機能するコンピュータのハードウェア構成例を示す図である。図3のコンピュータ10は、それぞれバスBで相互に接続されているドライブ装置100、補助記憶装置102、メモリ装置103、プロセッサ104、及びインタフェース装置105等を有する。

[0030] コンピュータ10をヘテロジニアス・コンピューティング設計制御部1として機能させるプログラムは、CD-ROM等の記録媒体101によって提供される。プログラムを記憶した記録媒体101がドライブ装置100にセットされると、プログラムが記録媒体101からドライブ装置100を介して補助記憶装置102にインストールされる。但し、プログラムのインストールは必ずしも記録媒体101より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよい。補助記憶装置102は、インストールされたプログラムを格納すると共に、必要なファイルやデータ等を格納する。

[0031] メモリ装置103は、プログラムの起動指示があった場合に、補助記憶装置102からプログラムを読み出して格納する。プロセッサ104は、CPU若しくはGPU (Graphics Processing Unit)、又はCPU及びGPUであり、メモリ装置103に格納されたプログラムに従ってコンピュータ10に係る機能を実行する。インタフェース装置105は、ネットワークに接続するためのインタフェースとして用いられる。

[0032] なお、コンテナ・オーケストレーション基盤2やトランスポート制御装置管理部3も、図3に示されるコンピュータ10と同様のコンピュータにインストールされた1以上のプログラムが当該コンピュータに実行させる処理によって実現されてもよい。

[0033] 以下、デバイス割当制御システムにおいて実行される処理手順について説明する。図4は、第1の実施の形態におけるデバイス割当制御システムにおいて実行される処理手順の一例を説明するための図である。

[0034] ステップS110において、要求性能受付部11は、ユーザ（が利用する

端末) から送信されたアプリケーション要求性能プロファイルを受信する。

[0035] 続いて、要求性能受付部 11 は、当該アプリケーション要求性能プロファイルをデバイス割当決定部 14 に送信する (S 120)。デバイス割当決定部 14 は、当該アプリケーション要求性能プロファイルを受信すると、デバイス情報取得部 12、性能推定モデル取得部 13、デバイス情報監視部 15 のそれぞれに対し、デバイス情報、モデル情報、デバイス使用状況情報を要求する。

[0036] デバイス情報取得部 12 は、デバイス割当決定部 14 からの要求に応じ、デバイス情報をデバイス情報記憶部 121 から取得し (S 130)、当該デバイス情報をデバイス割当決定部 14 に送信する (S 140)。

[0037] 性能推定モデル取得部 13 は、デバイス割当決定部 14 からの要求に応じ、アプリケーションごとのモデル情報を性能推定モデルリポジトリ 122 から取得して (S 150)、当該モデル情報をデバイス割当決定部 14 に送信する (S 160)。

[0038] デバイス情報監視部 15 は、デバイス割当決定部 14 からの要求に応じ、デバイス使用状況情報を各コンテナ・オーケストレーション基盤 2 から取得し (S 170)、取得した各デバイス使用状況情報をデバイス割当決定部 14 に送信する (S 180)。なお、デバイス情報監視部 15 は、常時 (例えば、定期的に) デバイス使用状況情報を各コンテナ・オーケストレーション基盤 2 から取得していてもよい。この場合、ステップ S 180 において、デバイス情報監視部 15 は、各コンテナ・オーケストレーション基盤 2 から最後に取得したデバイス使用状況情報をデバイス割当決定部 14 に送信してもよい。

[0039] 続いて、デバイス割当決定部 14 は、アプリケーション要求性能プロファイル、デバイス情報、モデル情報及びデバイス情報に基づいて、デバイス割当情報を決定 (計算) する (S 190)。デバイス割当情報とは、上記したように、アプリケーション・コンテナの配備先のコンテナ・オーケストレーション基盤 2 と、アプリケーション・コンテナの配備先の物理又は論理サー

バ（Workerノード）と、配備先の物理又は論理サーバごとの割当デバイス種別、割当デバイス識別子及び割当デバイス数量とを含む情報である。

[0040] 続いて、デバイス割当決定部14は、アプリケーション・コンテナの配備先のコンテナ・オーケストレーション基盤2と、アプリケーション・コンテナの配備先の物理又は論理サーバ（Workerノード）と、配備先の物理又は論理サーバごとの割当デバイス種別、割当デバイス識別子及び割当デバイス数量とをデバイス割当指示部16へ通知する（S200）。デバイス割当決定部14は、また、物理又は論理サーバ間を跨いで複数デバイス間を接続する場合（コンテナに割り当てられるデバイスが複数の物理又は論理サーバに跨がる場合）に、仮想専用ネットワークの提供をトランスポート装置制御連携部17に指示する（S210）。

[0041] デバイス割当指示部16は、デバイス割当決定部14からの通知に基づき、各配備先のコンテナ・オーケストレーション基盤2のAPIサーバ21に、対象の各物理又は論理サーバ（配備先のコンテナ・オーケストレーション基盤2が管理するWorkerノード）内のデバイス種別、デバイス識別子及びデバイス数量についての予約及び割り当てを指示する（S220）。各配備先のコンテナ・オーケストレーション基盤2は、指示通りに配備先のコンテナに対してデバイスの割り当てを実行する。

[0042] トランスポート装置制御連携部17は、デバイス割当決定部14からの指示（接続元のデバイス及び接続先のデバイスのそれぞれのデバイス情報）に基づき、物理又は論理サーバ間を跨いでデバイス間を接続する場合の仮想専用ネットワークの提供をトランスポート制御装置管理部3に指示すると共に、接続元のデバイス及び接続先のデバイスのそれぞれのデバイス情報をトランスポート制御装置管理部3へ送信する。（S230）。

[0043] 続いて、トランスポート制御装置管理部3は、トランスポート装置制御連携部17からの情報に基づき、デバイス間を接続する複数の仮想専用ネットワークの経路候補のうち、「遅延が最小になる経路」を探索し、該当する経路で仮想専用ネットワークを生成するよう、トランスポート装置リソースプ

ール4内のリソースを制御する（S240）。その結果、当該仮想専用ネットワークが提供される。

[0044] 上述したように、第1の実施の形態によれば、デバイス構成・トポロジ構成情報等を活用することで、アプリケーションの要求性能End-to-Endの性能要件）を満たすようにコンテナに対してデバイスを割り当てることを可能とすることができる。

[0045] なお、ユースケースやアプリケーションによっては、単一のデバイスのみならず、複数のデバイスに跨ってデータ処理のオフロードが発生する場合があるため、デバイス間を跨ぐ際の単一の物理又は論理サーバ内のデバイス構成及びトポロジ構成や、複数の物理又は論理サーバ間の仮想ネットワーク接続構成が性能に大きな影響を与えるようになる。

[0046] このような問題に対し、本実施の形態によれば、オンデマンドなトランスポート連携により、リモート・デバイスが割り当て可能となることで、特定ドメイン（クラスタ又はサーバ）内のローカルデバイスに制限されることのない、柔軟なアプリケーションの展開を可能にすることができる。これにより、エッジ環境のようなリソースの乏しい環境に迅速にアプリケーションを展開することで、設備コストを削減することができる。

[0047] 次に、第2の実施の形態について説明する。第2の実施の形態では第1の実施の形態と異なる点について説明する。第2の実施の形態において特に言及されない点については、第1の実施の形態と同様でもよい。

[0048] 第1の実施の形態において、アプリケーション・コンテナが遠隔のコンテナ・オーケストレーション基盤2上で登録・管理されているデバイスを利用する場合（例えば、遠隔のNVMe over RDMA社GPUに対してGPUDirect RDMA（非特許文献4）を使用してデータを直接送信し、遠隔GPU上で処理を実行する場合等）には、当該遠隔のコンテナ・オーケストレーション基盤2上でその利用の実態を把握することができない。具体的には、当該コンテナ・オーケストレーション基盤2上から、デバイス情報監視部15に対して送信される情報には、単純なデバイス使用に関するメトリクス（GPU使用率等）

のみであり、各メトリクスの割当先（つまり、どのアプリケーション・コンテナによりそのデバイスがどの程度使用されているか）を把握できない。したがって。例えば、性能保証型で物理リソースの専有化を求めるアプリケーションに係るアプリケーション・コンテナに対して既に或るデバイスが割り当てられている場合、当該デバイスのリソースに余裕があったとしても、当該アプリケーションの性能を保証するために、他のアプリケーションに対して当該デバイスを割り当てないようにする、といった制御を行うのは困難である。

[0049] 第2の実施の形態では、このような課題を解決した例について説明する。

[0050] 図5は、第2の実施の形態におけるデバイス割当制御システムの機能構成例を示す図である。図5中、図2と同一部分には同一符号を付し、その説明は省略する。

[0051] 図5において、ヘテロジニアス・コンピューティング設計制御部1は、更に、デバイス利用情報管理部18を有する。

[0052] デバイス利用情報管理部18は、デバイス割当決定部14が決定（計算）したデバイス割当情報（アプリケーション・コンテナの配備先のコンテナ・オーケストレーション基盤2と、アプリケーション・コンテナの配備先の物理又は論理サーバ（Workerノード）と、配備先の物理又は論理サーバごとの割当デバイス種別、割当デバイス識別子及び割当デバイス数量を含む情報）と、当該デバイス割当情報に基づくデバイスの割り当ての実行によって変化したデバイス使用状況情報と、当該デバイス割当情報の決定（計算）に利用されたデバイス情報とが関連付けられた情報（以下、「デバイス利用情報」という。）をデバイス利用情報記憶部123に記録する。但し、デバイス情報は、デバイス利用情報に含められなくてもよい。デバイス利用情報のうちのデバイス割当情報及びデバイス使用状況情報は、新たなアプリケーション要求性能プロファイルが入力された際に、デバイス割当決定部14がデバイス割当情報を決定（計算）するために利用される。

[0053] 図6は、第2の実施の形態におけるデバイス割当制御システムにおいて実

行される処理手順の一例を説明するための図である。図6中、図4と同ーステップには同ーステップ番号を付し、その説明は適宜省略する。

[0054] まず、デバイス割当制御システムの運用開始後の最初のアプリケーション要求性能プロファイルの入力時には、第1の実施の形態において説明したステップS110～S240が実行される。但し、第2の実施の形態において、アプリケーション要求性能プロファイルがアプリケーションごとに含む要求性能情報（入力情報）は、更に、当該アプリケーションが性能保証型であるか性能非保証型（ベストエフォート型）を示す情報（以下、「性能保証情報」という。）を含む。

[0055] また、ステップS240の実行により仮想専用ネットワークが提供された後でステップS250が実行される。ステップS250において、デバイス割当決定部14は、ステップS200においてデバイス割当指示部16へ通知したデバイス割当情報と同じデバイス割当情報（アプリケーション・コンテナの配備先のコンテナ・オーケストレーション基盤2と、アプリケーション・コンテナの配備先の物理又は論理サーバ（Workerノード）と、配備先の物理又は論理サーバごとの割当デバイス種別、割当デバイス識別子及び割当デバイス数量とを含む情報）と、ステップS140において取得したデバイス情報とをデバイス利用情報管理部18へ通知する。デバイス割当情報は、更に、デバイスが割り当てられたアプリケーション・コンテナに係るアプリケーションの性能保証情報をも含む。当該性能保証情報は、アプリケーション要求性能プロファイルに基づいてデバイス割当決定部14が特定可能である。

[0056] ステップS250に応じ、デバイス利用情報管理部18は、デバイス情報監視部15に対してデバイス使用状況情報（各デバイス使用に関するメトリクス）の取得を要求する。デバイス情報監視部15は、デバイス利用情報管理部18からの要求に応じ、デバイス使用状況情報を各コンテナ・オーケストレーション基盤2から取得し（S251）、取得した各デバイス使用状況情報をデバイス利用情報管理部18に送信する（S252）。第1の実施の

形態のステップS 1 7 0 及びS 1 8 0 において説明したように、デバイス情報監視部 1 5 は、常時（例えば、定期的に）デバイス使用状況情報を各コンテナ・オーケストレーション基盤 2 から取得していてもよい。この場合、デバイス情報監視部 1 5 は、各コンテナ・オーケストレーション基盤 2 から最後に取得したデバイス使用状況情報をステップS 2 5 2 においてデバイス利用情報管理部 1 8 に送信してもよい。

[0057] 続いて、デバイス利用情報管理部 1 8 は、ステップS 2 5 0 において通知されたデバイス割当情報及びデバイス情報と、ステップS 2 5 2 において受信したデバイス使用状況情報とを関連付けて含むデバイス利用情報をデバイス利用情報記憶部 1 2 3 に保存する（S 2 5 3）。当該デバイス使用状況情報は、当該デバイス割当情報に基づくデバイス割当（S 2 2 0）が反映された後のデバイス使用状況情報である。当該デバイス情報は、当該デバイス割当情報の計算に利用されたデバイス情報である。

[0058] 次に、アプリケーション要求性能プロファイルの 2 回目以降の入力時について図 6 を参照しつつ説明する。この場合、ステップS 1 1 0 及びS 1 2 0 は、図 4 と同様に実行される。なお、アプリケーション要求性能プロファイルが性能保証情報を含むのは 2 回目以降においても同様である。

[0059] デバイス割当決定部 1 4 は、当該アプリケーション要求性能プロファイルを受信すると、デバイス情報取得部 1 2、性能推定モデル取得部 1 3、デバイス利用情報管理部 1 8 のそれぞれに対し、デバイス情報、モデル情報、デバイス利用情報を要求する。すなわち、第 2 の実施の形態では、デバイス情報監視部 1 5 に対してデバイス使用状況情報が要求される代わりに、デバイス利用情報管理部 1 8 に対してデバイス利用情報が要求される。デバイス利用情報は、デバイス使用状況情報をも含むため、デバイス情報監視部 1 5 からの情報の取得は不要だからである。

[0060] デバイス情報取得部 1 2 は、デバイス割当決定部 1 4 からの要求に応じ、デバイス情報をデバイス情報記憶部 1 2 1 から取得し（S 1 3 0）、当該デバイス情報をデバイス割当決定部 1 4 に送信する（S 1 4 0）。

- [0061] 性能推定モデル取得部 13 は、デバイス割当決定部 14 からの要求に応じ、第 1 の実施の形態と同様に、アプリケーションごとのモデル情報をデバイス割当決定部 14 に送信する（S 150、S 160）。
- [0062] デバイス利用情報管理部 18 は、デバイス割当決定部 14 からの要求に応じ、これまでにデバイス利用情報記憶部 123 に保存（記憶）した全てのデバイス利用情報のデバイス割当情報と、最後に保存したデバイス利用情報のデバイス使用状況情報とをデバイス利用情報記憶部 123 から取得し（S 181）、取得した情報をデバイス割当決定部 14 に送信する（S 182）。
- [0063] 続いて、デバイス割当決定部 14 は、アプリケーション要求性能プロファイル、ステップ S 182 において取得したデバイス割当情報及びデバイス使用状況情報、並びにモデル情報に基づいて、デバイス割当情報を決定（計算）する（S 190）。第 1 の実施の形態と異なる点は、デバイス利用情報が含む、コンテナ・オーケストレーション基盤 2 へ反映済みのデバイス割当情報がデバイス割当決定部 14 に対する入力情報として追加される点である。
- [0064] デバイス割当決定部 14 が実行するロジックは、基本的には第 1 の実施の形態と同様であるが、デバイス利用情報が含むデバイス割当情報が考慮される点が第 1 の実施の形態と異なる。デバイス割当情報は、各アプリケーション・コンテナがどのコンテナ・オーケストレーション基盤 2 上のどのデバイスをどの程度利用しているかを示すとともに、各アプリケーション・コンテナに係るアプリケーションの性能保証情報（性能保証型又は性能非保証型（ベストエフォート型））を含む情報である。したがって、当該デバイス割当情報によれば、既にデバイスが割り当てられた既存の割当先がアプリケーション・コンテナにおいて起動されているアプリケーションが、性能保証型及び性能非保証型のいずれのアプリケーションであるかが分かる。そこで、デバイス割当決定部 14 は、性能保証型のアプリケーションに係るアプリケーション・コンテナにはデバイスを専有化させ、性能非保証型のアプリケーション・コンテナ間では同一のデバイスを時分割でシェアすることを許容して、デバイス割当情報を決定（計算）する。

[0065] 例えば、デバイス割当決定部14は、デバイス情報及びモデル情報を用いて、デバイス（CPU、メモリ、その他デバイス等）ごとの割当量の大量の入力パターン（すなわち、配備先の物理又は論理サーバごとの割当デバイス種別、割当デバイス識別子及び割当デバイス数量の大量のパターン）について、配備対象のアプリケーション・コンテナに係るアプリケーションの各性能指標の推定値を導出する。デバイス割当決定部14は、導出した各性能推定値に対して、アプリケーション要求性能プロファイル及びデバイス使用状況情報に加えて、性能保証型のアプリケーション・コンテナに割り当てるデバイスは（リソースに余剰があったとしても）2以上のコンテナに割り当てないということを制約条件として、デバイス（CPU、メモリ、その他デバイス等）ごとの割当量から最適化アルゴリズムによって算出されるコスト（設備コストや電力消費量）を目的関数として、当該制約条件を満たしつつ当該目的関数を最小化する入力パターンを探索することでデバイス割当情報を決定（計算）する。

[0066] 続いて、第1の実施の形態と同様にステップS200～S240が実行される。この際、上記したようにステップS250も実行されて、ステップS250に応じてステップS251～S253が実行される。その結果、デバイス利用情報記憶部123が記憶するデバイス利用情報は、今回のデバイス割当情報が反映された内容となる。

[0067] 上述したように、第2の実施の形態によれば、新たなアプリケーション・コンテナに対して割り当てるデバイスを決定する際に、既にデバイスが割り当てられている既存のアプリケーション・コンテナのデバイス利用状況を考慮することができる。その結果、新たなアプリケーションの要求性能をより高い精度で満たすようにコンテナに対してデバイスを割り当てることを可能とすることができる。

[0068] なお、ヘテロジニアス・コンピューティング設計制御部1として機能するコンピュータ10は、デバイス割当制御装置の一例である。

[0069] 以上、本発明の実施の形態について詳述したが、本発明は斯かる特定の実

施形態に限定されるものではなく、請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

符号の説明

[0070]	1	ヘテロジニアス・コンピューティング設計制御部
	2	コンテナ・オーケストレーション基盤
	3	トランスポート制御装置管理部
	4	トランスポート装置リソースプール
	1 1	要求性能受付部
	1 2	デバイス情報取得部
	1 3	性能推定モデル取得部
	1 4	デバイス割当決定部
	1 5	デバイス情報監視部
	1 6	デバイス割当指示部
	1 7	トランスポート装置制御連携部
	1 8	デバイス利用情報管理部
	1 0 0	ドライブ装置
	1 0 1	記録媒体
	1 0 2	補助記憶装置
	1 0 3	メモリ装置
	1 0 4	プロセッサ
	1 0 5	インタフェース装置
	1 2 1	デバイス情報記憶部
	1 2 2	性能推定モデルリポジトリ
	1 2 3	デバイス利用情報記憶部
	B	バス

請求の範囲

【請求項1】 アプリケーションの要求性能と性能保証型であるか否かとを示す入力情報を受け付けるように構成されている要求性能受付部と、

 複数のコンテナ・オーケストレーション基盤におけるデバイスの構成及びトポロジを示す情報、前記アプリケーションの性能を推定する情報、及び既にアプリケーションを起動しているコンテナに割り当てられている前記デバイスと当該アプリケーションが性能保証型であるか否かとを示す情報に基づいて、前記デバイスそれぞれの使用状況を示す情報、及び前記要求性能及び性能保証型のアプリケーション・コンテナに割り当てるデバイスは2以上のコンテナに割り当てないということを制約条件として、デバイスごとの割当量から最適化アルゴリズムによって算出されるコストを最小化するように、前記アプリケーションを起動するコンテナに割り当てる前記デバイス及び割当量を決定するように構成されているデバイス割当決定部と、

 前記アプリケーションを起動するコンテナに対して割り当てられる前記デバイスに係るコンテナ・オーケストレーション基盤に対し、当該デバイスの割り当てを指示するように構成されているデバイス割当指示部と、

 を有することを特徴とするデバイス割当制御装置。

【請求項2】 前記デバイス割当決定部が決定した前記アプリケーションを起動するコンテナに割り当てる前記デバイス及び割当量と当該アプリケーションが性能保証型であるか否かを示す情報を記憶部に記憶するように構成されているデバイス利用情報管理部を有し、

 前記デバイス割当決定部は、前記要求性能受付部が新たなアプリケーションについて前記入力情報を受け付けた場合に、前記記憶部が記憶する情報を前記既にアプリケーションを起動しているコンテナに関する情報として利用して、前記新たなアプリケーションを起動するコンテナに割り当てる前記デバイス及び割当量を決定するように構成さ

れている、

ことを特徴とする請求項1記載のデバイス割当制御装置。

[請求項3]

アプリケーションの要求性能と性能保証型であるか否かを示す入力情報を受け付ける要求性能受付手順と、

複数のコンテナ・オーケストレーション基盤におけるデバイスの構成及びトポロジを示す情報、前記アプリケーションの性能を推定する情報、及び既にアプリケーションを起動しているコンテナに割り当てられている前記デバイスと当該アプリケーションが性能保証型であるか否かを示す情報に基づいて、前記デバイスそれぞれの使用状況を示す情報、及び前記要求性能及び性能保証型のアプリケーション・コンテナに割り当てるデバイスは2以上のコンテナに割り当てないということを制約条件として、デバイスごとの割当量から最適化アルゴリズムによって算出されるコストを最小化するように、前記アプリケーションを起動するコンテナに割り当てる前記デバイス及び割当量を決定するデバイス割当決定手順と、

前記アプリケーションを起動するコンテナに対して割り当てられる前記デバイスに係るコンテナ・オーケストレーション基盤に対し、当該デバイスの割り当てを指示するデバイス割当指示手順と、

をコンピュータが実行することを特徴とするデバイス割当制御方法。

[請求項4]

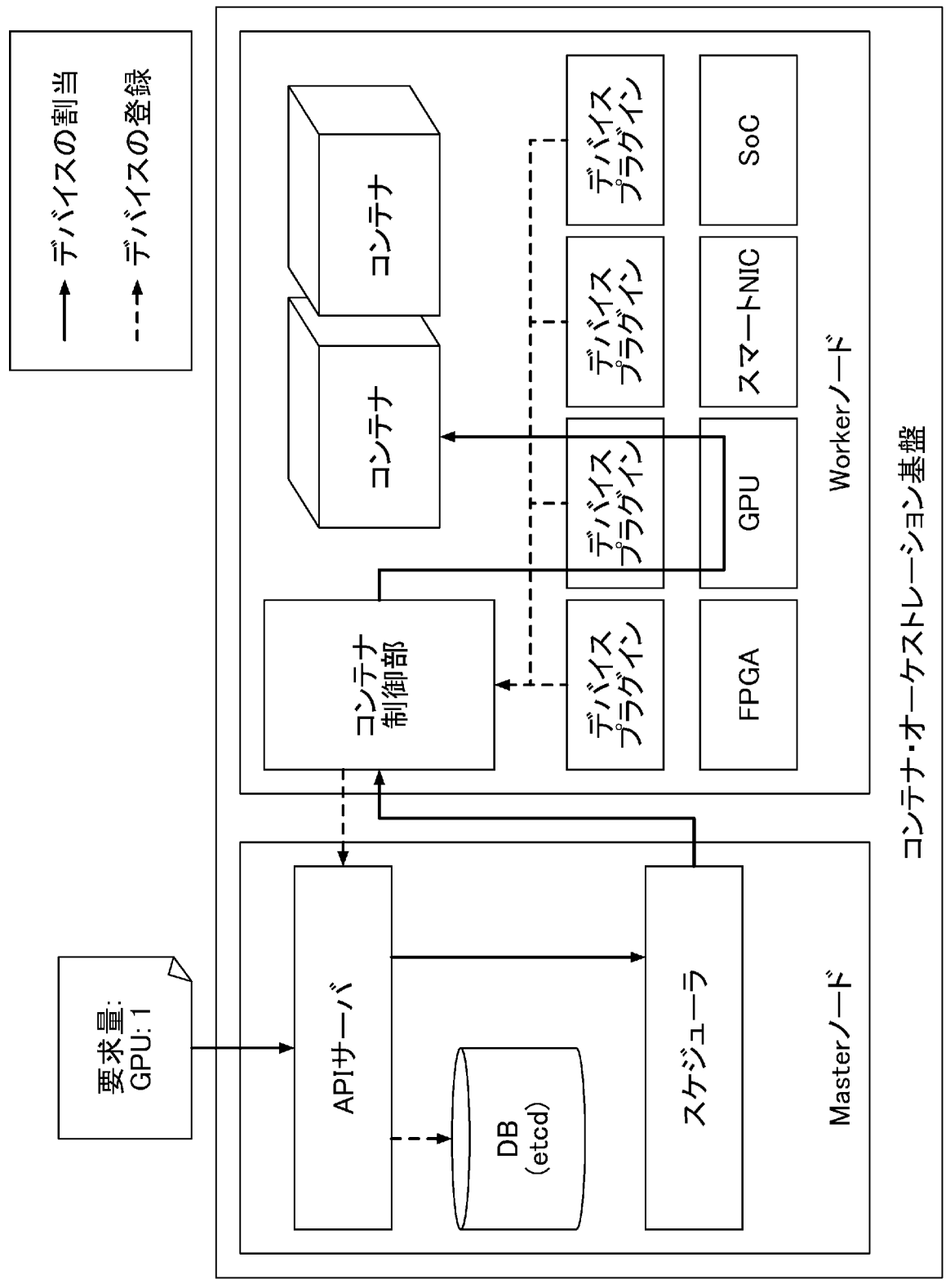
アプリケーションの要求性能と性能保証型であるか否かを示す入力情報を受け付ける要求性能受付手順と、

複数のコンテナ・オーケストレーション基盤におけるデバイスの構成及びトポロジを示す情報、前記アプリケーションの性能を推定する情報、及び既にアプリケーションを起動しているコンテナに割り当てられている前記デバイスと当該アプリケーションが性能保証型であるか否かを示す情報に基づいて、前記デバイスそれぞれの使用状況を示す情報、及び前記要求性能及び性能保証型のアプリケーション・コンテナに割り当てるデバイスは2以上のコンテナに割り当てないとい

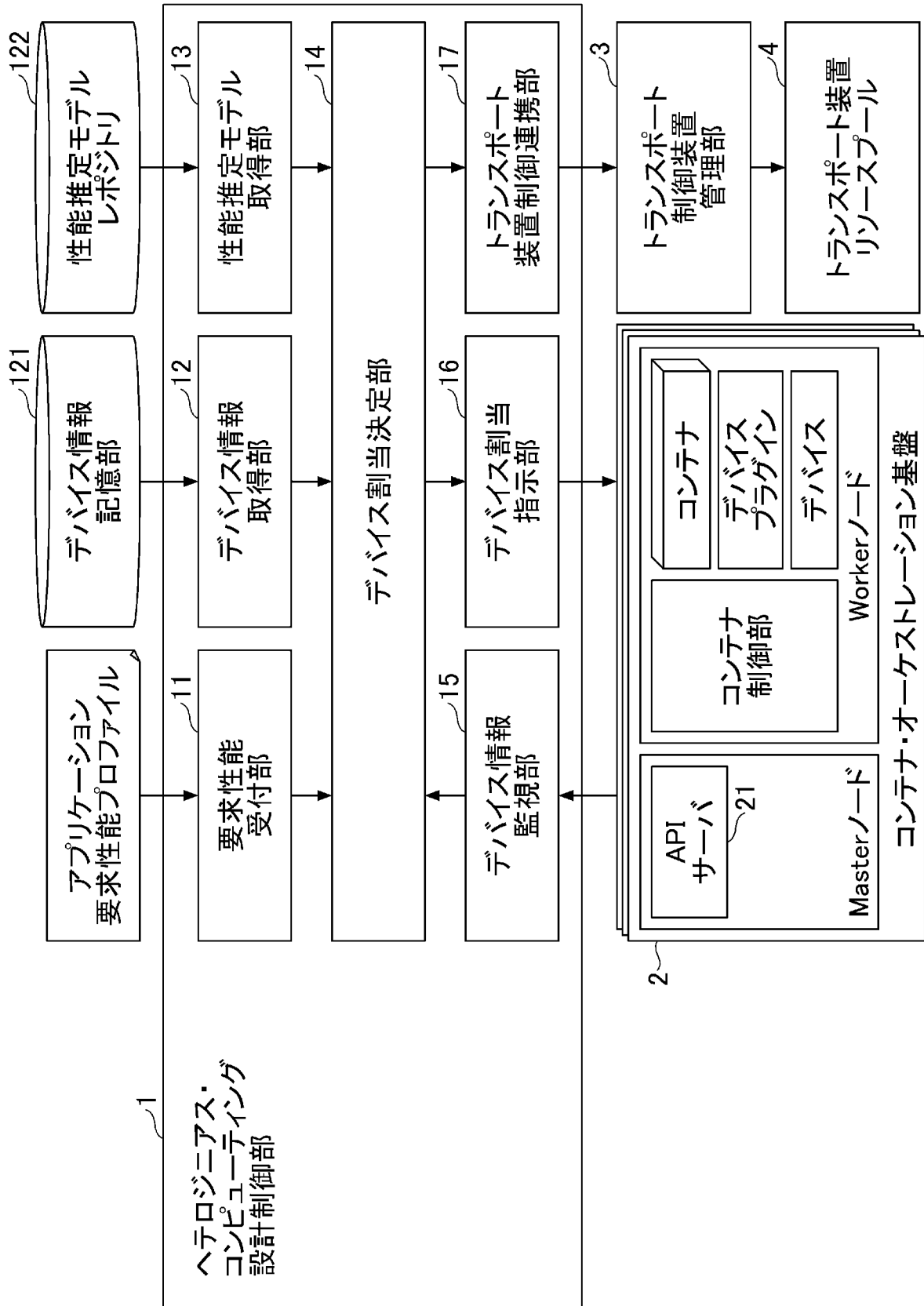
うことを制約条件として、デバイスごとの割当量から最適化アルゴリズムによって算出されるコストを最小化するように、前記アプリケーションを起動するコンテナに割り当てる前記デバイス及び割当量を決定するデバイス割当決定手順と、

前記アプリケーションを起動するコンテナに対して割り当てられる前記デバイスに係るコンテナ・オーケストレーション基盤に対し、当該デバイスの割り当てを指示するデバイス割当指示手順と、
をコンピュータに実行させることを特徴とするプログラム。

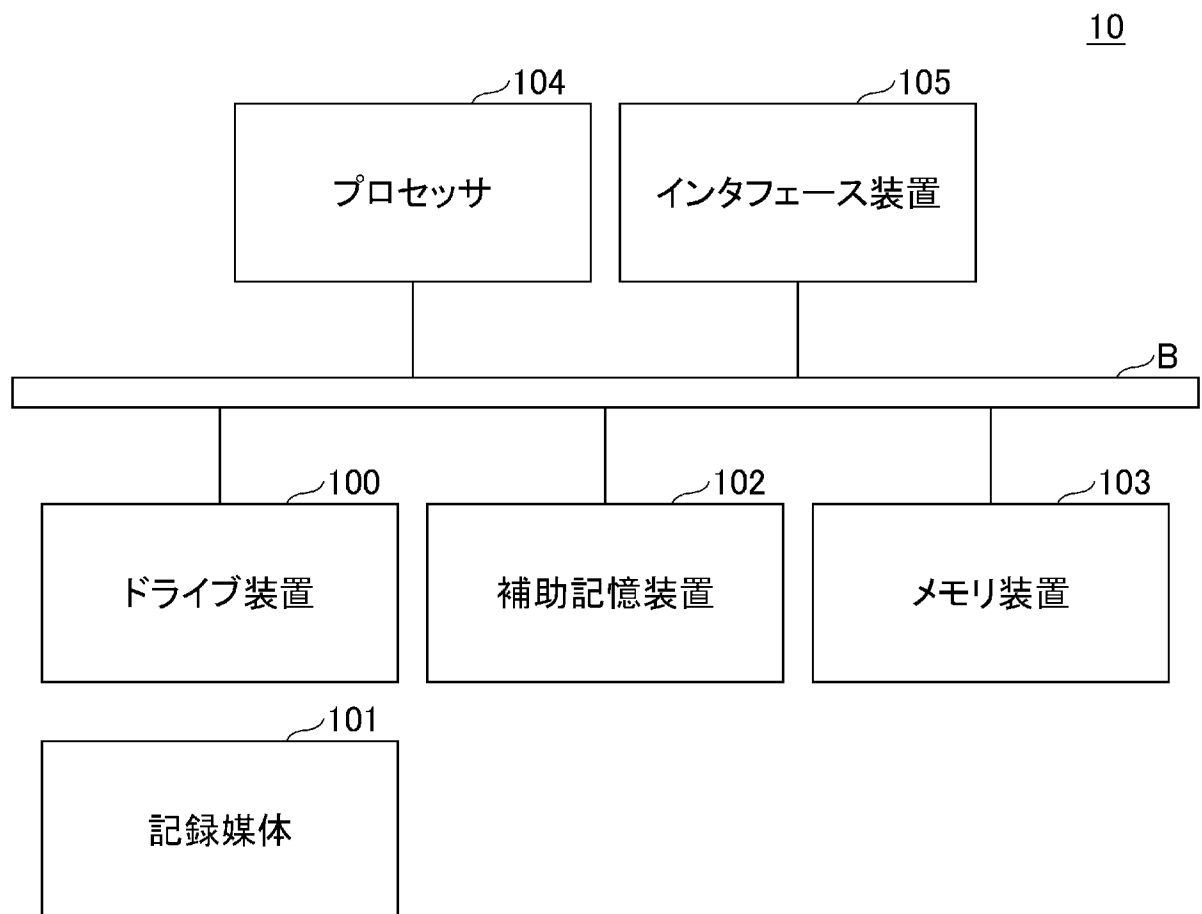
[図1]



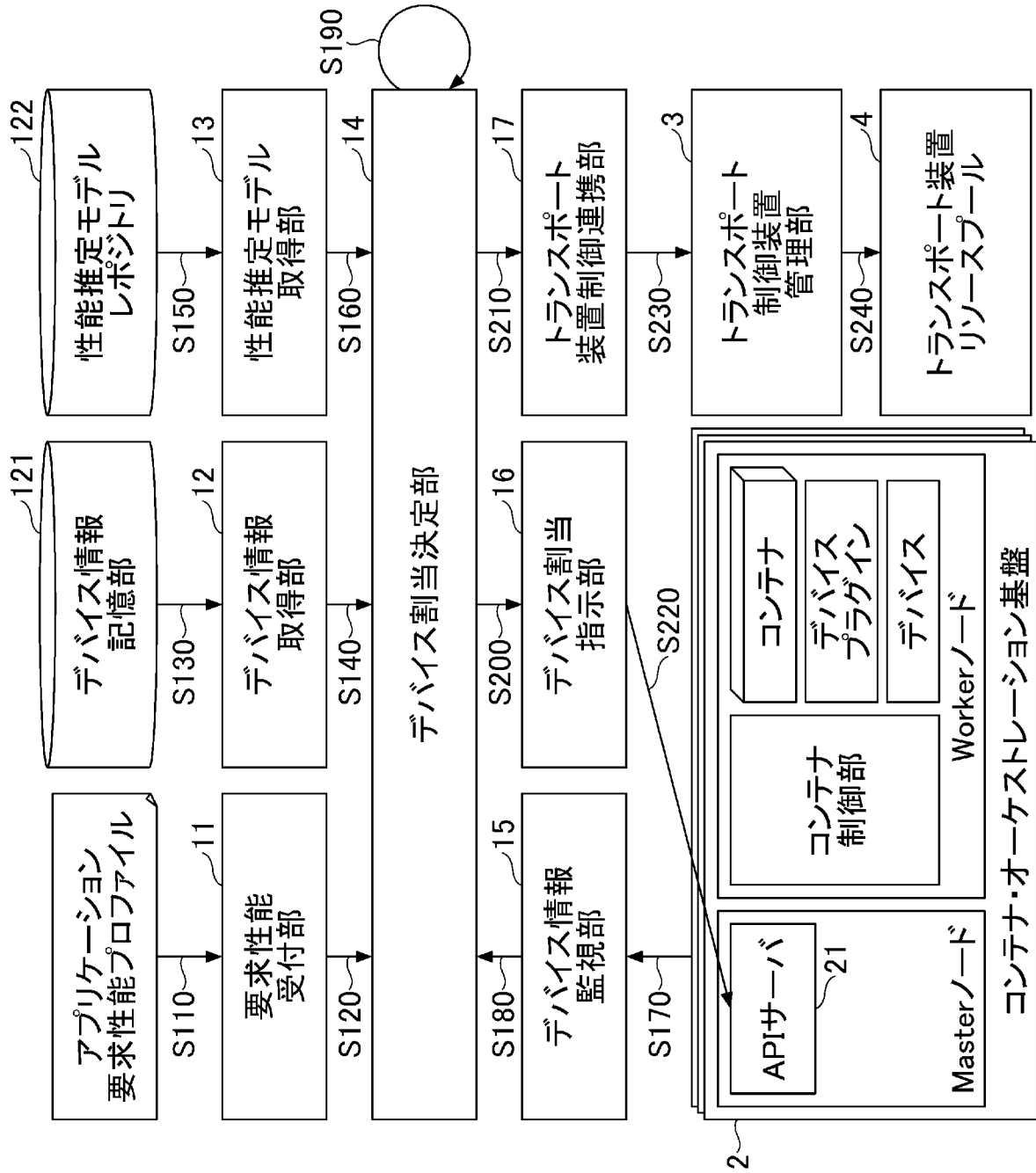
[図2]



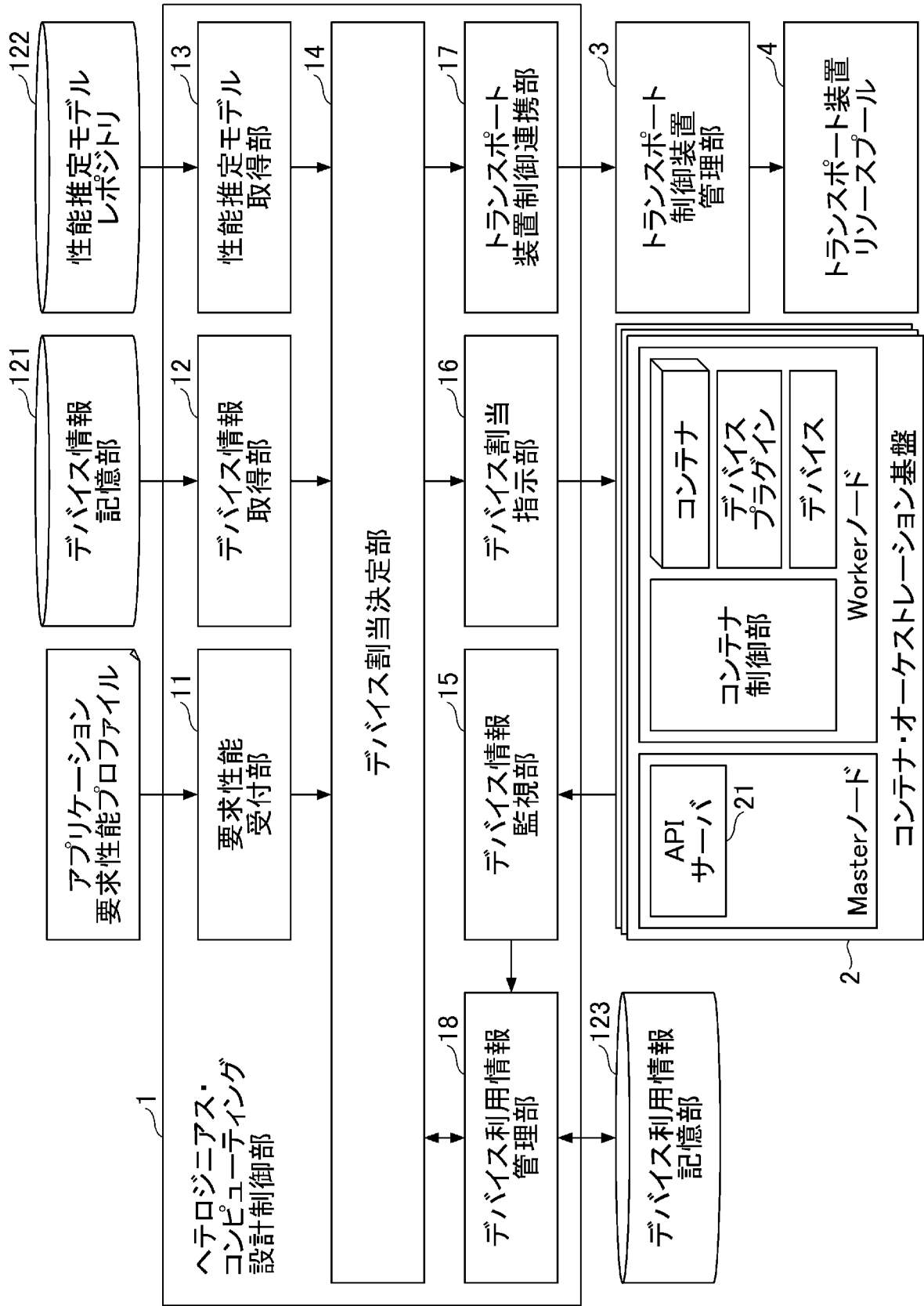
[図3]



[図4]



[図5]



[illegible]

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/025113

A. CLASSIFICATION OF SUBJECT MATTER**G06F 9/50**(2006.01)i; **G06F 8/60**(2018.01)i

FI: G06F9/50 120A; G06F8/60

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)

G06F9/50; G06F8/60

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

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2023

Registered utility model specifications of Japan 1996-2023

Published registered utility model applications of Japan 1994-2023

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2017-37403 A (HITACHI, LTD.) 16 February 2017 (2017-02-16) claims	1-4
A	CN 115169810 A (NARI TECHNOLOGY DEV CO., LTD.) 11 October 2022 (2022-10-11) paragraph [0051]	1-4
A	JP 2022-511177 A (INTEL CORPORATION) 31 January 2022 (2022-01-31) paragraphs [0833]-[0864]	1-4



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

21 August 2023

Date of mailing of the international search report

12 September 2023

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)

3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915

Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2023/025113

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2017-37403	A	16 February 2017	(Family: none)	
CN	115169810	A	11 October 2022	(Family: none)	
JP	2022-511177	A	31 January 2022	US 2021/0263779 A1	
				paragraphs [0849]-[0879]	
				CN 112955869 A	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 9/50(2006.01)i; G06F 8/60(2018.01)i FI: G06F9/50 120A; G06F8/60		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） G06F9/50; G06F8/60 最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2023年 日本国実用新案登録公報 1996-2023年 日本国登録実用新案公報 1994-2023年 国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2017-37403 A（株式会社日立製作所） 16.02.2017（2017-02-16） 【特許請求の範囲】	1-4
A	CN 115169810 A（国電南瑞科技股フン有限公司） 11.10.2022（2022-10-11） [0051]	1-4
A	JP 2022-511177 A（インテル・コーポレーション） 31.01.2022（2022-01-31） 【0833】～【0864】	1-4
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技术水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献 “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献		
国際調査を完了した日 21.08.2023	国際調査報告の発送日 12.09.2023	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 田中 幸雄 5B 9191 電話番号 03-3581-1101 内線 3545	

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2023/025113

引用文献			公表日	パテントファミリー文献			公表日
JP	2017-37403	A	16.02.2017	(ファミリーなし)			
CN	115169810	A	11.10.2022	(ファミリーなし)			
JP	2022-511177	A	31.01.2022	US	2021/0263779	A1	
				[0849]-[0879]			
				CN	112955869	A	