

19) RÉPUBLIQUE FRANÇAISE
 INSTITUT NATIONAL
 DE LA PROPRIÉTÉ INDUSTRIELLE
 PARIS

11) N° de publication : **2 916 550**
 (à n'utiliser que pour les
 commandes de reproduction)

21) N° d'enregistrement national : **08 54903**

51) Int Cl⁸ : **G 06 F 12/08 (2006.01)**

12) **DEMANDE DE BREVET D'INVENTION** **A1**

22) Date de dépôt : 18.07.08.

30) Priorité :

43) Date de mise à la disposition du public de la demande : 28.11.08 Bulletin 08/48.

56) Liste des documents cités dans le rapport de recherche préliminaire : *Ce dernier n'a pas été établi à la date de publication de la demande.*

60) Références à d'autres documents nationaux apparentés : Division demandée le 18/07/08 bénéficiant de la date de dépôt du 20/07/04 de la demande initiale n° 04 08017.

71) Demandeur(s) : **HITACHI LTD — JP.**

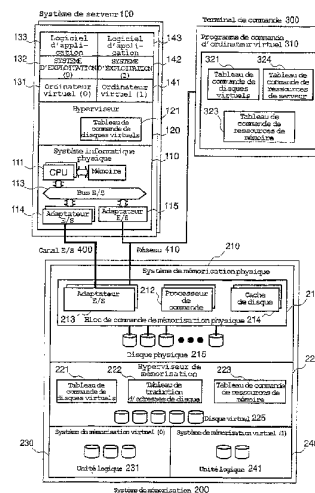
72) Inventeur(s) : **HASHIMOTO AKIYOSHI.**

73) Titulaire(s) :

74) Mandataire(s) : **NONY & ASSOCIES.**

54) **SYSTEME INFORMATIQUE, DISPOSITIF DE COMMANDE, SYSTEME DE MEMORISATION ET DISPOSITIF INFORMATIQUE.**

57) Système de mémorisation (200) comportant:
 une pluralité de ports,
 une pluralité de processeurs (212) qui commandent le transfert de données reçues par lesdits ports,
 une pluralité de ressources de mémoire (213, 214),
 une pluralité de groupes RAID (Réseau Redondant de Disques Indépendants), chacun d'eux étant associé à une pluralité d'unités de disque (215), et
 une pluralité de groupes de ressources (221, 222, 223), chacun desdits groupes de ressources étant divisé par au moins une partition logique, chaque groupe de ressources ayant une pluralité de types de ressources incluant au moins l'un desdits ports, au moins l'un desdits processeurs, au moins l'une desdites ressources de mémoire, et au moins l'un desdits groupes RAID,
 dans lequel une quantité desdites ressources dans chacun desdits groupes de ressources est attribuée sur la base d'une performance requise pour chacun desdits groupes de ressources.



FR 2 916 550 - A1



La présente invention concerne un système informatique et, plus particulièrement, une technologie de partitionnement logique qui implique des mémorisations du contenu de systèmes informatiques connectés à des systèmes de mémorisation.

Une approche pour améliorer les performances d'un système de traitement d'informations consiste à accroître le nombre d'ordinateurs contenus dans un système de traitement d'informations. Cependant, l'utilisation de nombreux ordinateurs dans un système pose le problème suivant : ceci nécessite une tâche difficile de commande des ordinateurs individuels, augmente la place occupée par les ordinateurs et consomme plus d'énergie électrique. En tant que solution à ce problème, une technologie qui partitionne de manière logique des ressources d'un ordinateur à l'aide d'une grande capacité de traitement (PARL : Partitionnement Logique) et rend possible d'utiliser les partitions logiques obtenues en résultat en tant d'ordinateurs virtuels indépendants a été proposée.

Cette technologie de partitionnement logique peut amener un ordinateur à ressembler à une pluralité d'ordinateurs virtuels. Lorsque l'allocation de ressources (processeur, mémoire, etc.) aux partitions est commandée, les performances de chaque ordinateur virtuel sont garanties. A l'aide de cette technologie, différents systèmes d'exploitation peuvent être librement installés dans des ordinateurs virtuels de sorte que chaque ordinateur virtuel peut être mis sous tension et hors tension ou dépanné indépendamment pour un fonctionnement flexible. De plus, l'utilisation d'un plus petit nombre de machines physiques offre des avantages en termes de commande de système, de place occupée et de consommation d'énergie. Ce type de technologie de partitionnement logique est décrite, par exemple, dans le document JP-A N° 157 177/2003.

Dans la technologie de partitionnement logique qui a été utilisée jusqu'ici pour les ordinateurs, les ressources d'ordinateurs telles que des processeurs et des mémoires sont partitionnées de manière logique et al-
5 louées à des ordinateurs virtuels.

Des systèmes de mémorisation qui sont utilisés avec des ordinateurs incluent non seulement un système de mémorisation directement connecté à un ordinateur hôte mais également un système de mémorisation partagé par une
10 pluralité d'ordinateurs via un réseau. La zone de mémoire d'un système de mémorisation connecté à un ordinateur est partitionnée et l'une des partitions obtenues en résultat est allouée à l'un des ordinateurs virtuels.

Lorsqu'un système de mémorisation a une fonc-
15 tion de système de fichiers, il est utilisé en tant que système de mémorisation qui permet un partage de fichiers entre différents serveurs, c'est-à-dire un système de Mé-
morisation en Réseau (NAS) en tant que système de mémori-
sation qui permet d'accéder aux fichiers à partir d'un
20 ordinateur. Une communication de données entre un système NAS et un ordinateur hôte a lieu fichier par fichier où chaque fichier doit avoir un nom et une structure pouvant être reconnu par le système d'exploitation s'exécutant
sur l'ordinateur hôte. Pour cette raison, en plus d'une
25 unité de disque qui mémorise des données et son contrô-
leur, le système NAS a un processeur et une mémoire pour le fonctionnement d'un système de fichiers qui convertit une entrée/sortie de fichier avec l'ordinateur hôte en une entrée/sortie de données avec l'unité de disque. Ce
30 type de système NAS ne prend pas en considération un par-
titionnement logique des ressources.

En outre, un système de réseau redondant de disques bon marché (RAID), qui est utilisé avec un grand système de mémorisation externe, ne présuppose pas un
35 partitionnement logique. Même lorsqu'un partitionnement

logique est autorisé dans ce type de système RAID, un système de serveur effectue simplement un partitionnement logique des ressources de mémoire pré-affectées et ne peut pas réaffecter les ressources du système de mémorisation et, par conséquent, l'allocation de ressources au système entier incluant le système de serveur et le système de mémorisation ne peut pas être optimisée.

Un but de la présente invention consiste à permettre une utilisation plus efficace d'un système de mémorisation partagé par une pluralité d'ordinateurs hôtes et à optimiser les performances du système entier y compris les ordinateurs hôtes et les dispositifs de mémorisation.

Conformément à un aspect de la présente invention, un système informatique comporte un dispositif informatique sur lequel un logiciel d'application s'exécute et un système de mémorisation qui mémorise des données requises pour le fonctionnement du dispositif informatique. Le dispositif informatique a un premier bloc de commande qui partitionne de manière logique des ressources informatiques du dispositif informatique et amène les partitions obtenues en résultat à fonctionner en tant qu'ordinateurs virtuels indépendants. Le système de mémorisation a un second bloc de commande qui partitionne de manière logique des ressources de mémoire du système de mémorisation et amène des partitions obtenues en résultat à fonctionner en tant que systèmes de mémorisation virtuels indépendants.

Le système comporte en outre une unité de gestion ayant : une première table de commande qui commande des ressources informatiques du dispositif informatique ; une deuxième table de commande qui commande des ressources de mémoire du système de mémorisation ; et une troisième table de commande qui commande les relations entre les ordinateurs virtuels et les systèmes de mémorisation

virtuels. Ici, le premier bloc de commande partitionne de manière logique les ressources informatiques conformément aux paramétrages de la première table de commande ; et le second bloc de commande partitionne de manière logique les ressources de mémoire conformément aux paramétrages de la deuxième table de commande.

Conformément à la présente invention, du fait que des ressources de mémoire peuvent être partitionnées de manière logique de manière à s'adapter au partitionnement logique de ressources de serveur, les ressources de système incluant des ressources de serveur et de mémoire peuvent être allouées de manière optimale.

Dans des systèmes habituels, l'état de ressources de mémoire autres que des disques (par exemple, des caches de disque) ne peut pas être contrôlé par le serveur. D'autre part, dans la présente invention, ces ressources, qui influencent considérablement les performances, peuvent également être allouées de sorte que l'allocation de ressources du système informatique est optimisée.

La présente invention va être plus particulièrement décrite en référence aux dessins annexés, sur lesquels :

- la figure 1 est un schéma fonctionnel représentant la configuration d'un système informatique conformément à un premier mode de réalisation de la présente invention,

- la figure 2 représente un tableau de commande de disques virtuels conformément à un mode de réalisation de la présente invention,

- la figure 3 représente un tableau de traduction d'adresses de disque conformément à un mode de réalisation de la présente invention,

- la figure 4 représente un tableau de commande de ressources de mémoire conformément à un mode de réalisation de la présente invention,

5 - la figure 5 représente un tableau de commande de ressources conformément à un mode de réalisation de la présente invention,

- la figure 6 est un ordinogramme représentant un processus d'allocation de ressources conformément à un mode de réalisation de la présente invention,

10 - la figure 7 est un ordinogramme représentant un processus d'entrée/sortie de données conformément à un mode de réalisation de la présente invention,

- la figure 8 représente la structure de couche d'un protocole de communication de canal d'entrée/sortie (E/S) conformément à un mode de réalisation de la présente invention,

15 - la figure 9 représente une communication de données entre un système de serveur et un système de mémorisation conformément à un mode de réalisation de la présente invention,

20 - la figure 10 représente un en-tête de communication d'hyperviseur conformément à un mode de réalisation de la présente invention,

- la figure 11 représente un écran de configuration de système informatique conformément à un mode de réalisation de la présente invention,

- la figure 12 représente un écran de configuration de système informatique conformément à un mode de réalisation de la présente invention,

30 - la figure 13 est un schéma fonctionnel représentant la configuration d'un système informatique conformément à un deuxième mode de réalisation de la présente invention,

- la figure 14 est un schéma fonctionnel représentant la configuration d'un système informatique

conformément à un troisième mode de réalisation de la présente invention,

- la figure 15 est un schéma fonctionnel représentant la configuration d'un système informatique
5 conformément à un quatrième mode de réalisation de la présente invention,

- la figure 16A représente les performances d'un disque virtuel constitué d'un disque physique et la figure 16B représente les performances d'un disque virtuel constitué de trois disques physiques conformément au
10 quatrième mode de réalisation de la présente invention,

- la figure 17 représente un tableau de commande de ressources de mémoire conformément au quatrième mode de réalisation, et

- la figure 18 représente un écran de configuration de système informatique conformément au quatrième mode de réalisation.

Des modes préférés de réalisation de la présente invention vont maintenant être décrits en se reportant aux dessins annexés.
20

En se reportant à la figure 1, le premier mode de réalisation de la présente invention est constitué d'un système de serveur 100 sur lequel un logiciel d'application s'exécute ; d'un système de mémorisation 200 qui mémorise des données requises pour le fonctionnement du système de serveur 100 ; et d'un terminal de commande 300 qui commande le fonctionnement de l'ensemble du système informatique.
25

Le système de serveur 100 a un système informatique physique 110 qui intègre des ressources telles qu'une unité centrale de traitement (CPU) 111, une mémoire 112, un bus E/S 113, et des adaptateurs E/S 114 et 115. La CPU 111 exécute des calculs pour le système d'exploitation (0) 132, le système d'exploitation (1) 142 et
30 les logiciels d'application 133 et 143 qui sont exécutés
35

dans le système de serveur 100. La mémoire 112 mémorise temporairement des programmes et des données requises pour le fonctionnement de la CPU 111. Le bus E/S 113 connecte la CPU 111 et les adaptateurs E/S 114 et 115 pour échanger des données. L'adaptateur E/S 114 est connecté au système de mémorisation 200 via un canal E/S (par exemple, Fibre Canal) 400 et transmet une demande pour une entrée/sortie de données au système de mémorisation 200 et reçoit des données mémorisées dans le système de mémorisation 200. L'adaptateur E/S 115 est connecté au terminal de commande 300 via un réseau 410 (par exemple, Ethernet (marque commerciale déposée)).

Dans le système de serveur 100, les systèmes d'exploitation 132 et 142 s'exécutent et les logiciels d'application 133 et 143 s'exécutent respectivement sur les systèmes d'exploitation (0) 132 et (1) 142. Les logiciels d'application 133 et 143 fournissent divers services tels qu'un service de base de données, un service de la toile mondiale aux terminaux clients (non-représentés) connectés au système de serveur 100.

Les ressources du système informatique physique 110 sont commandées par un hyperviseur 120. L'hyperviseur 120 est un logiciel de commande qui crée et commande des partitions logiques (c'est-à-dire des ordinateurs virtuels) dans le système de serveur 100. L'hyperviseur 120 s'exécute sur la CPU 111. L'hyperviseur 120 crée un ordinateur virtuel (0) 131 sur la base des ressources informatiques utilisées par le système d'exploitation (0) 132 et un ordinateur virtuel (1) 141 sur la base des ressources informatiques utilisées par le système d'exploitation (1) 142, dans le système informatique physique 110.

L'hyperviseur 120 a un tableau de commande de disques virtuels 121 (figure 2). Le tableau de commande de disques virtuels 121 mémorise le même contenu que le tableau de commande de disques virtuels 221, c'est-à-dire

des données concernant la configuration de systèmes de mémorisation virtuels 230 et 240 du système de mémorisation 200.

Le système de mémorisation 200 a un système de
5 mémorisation physique 210 incluant ces ressources en tant
que bloc de commande de mémorisation physique 211 et dis-
ques physiques 215.

Le bloc de commande de mémorisation physique
211 intègre un processeur de commande (CPU) 212, un adap-
10 tateur E/S 213 et un cache de disque 214. Le processeur
de commande 212 commande l'entrée/sortie de données avec
les disques physiques 215 et également le fonctionnement
du système de mémorisation 200. Si le système de mémori-
sation 200 est un système NAS, le processeur de commande
15 212 gère un système de fichiers. L'adaptateur E/S 213 est
connecté au système de serveur 100 via le canal E/S 400.
Le cache de disque 214 mémorise temporairement des don-
nées lues à partir du disque physique 215 et des données
à écrire dans le disque physique 215 pour améliorer les
20 performances d'accès du système de mémorisation 200.

Le disque physique 215 est commandé par un hy-
perviseur de mémorisation 220. L'hyperviseur de mémorisa-
tion 220 est un logiciel de commande qui crée et commande
des partitions logiques dans le système de mémorisation
25 200. L'hyperviseur 220 s'exécute sur un processeur de
commande 212. L'hyperviseur de mémorisation 220 crée des
disques virtuels 225. Spécifiquement, l'hyperviseur de
mémorisation 220 partitionne le disque physique 215 en
une pluralité de disques virtuels 225 ou combine une plu-
30 ralité de disques physiques 215 en un disque virtuel 225
unique.

Le système de mémorisation 200 sélectionne un
ou plusieurs disques virtuels 225 et offre ceux-ci en
tant que zone de mémoire aux ordinateurs virtuels 131 et
35 141. Les disques virtuels ainsi sélectionnés sont appelés

unités logiques. Une unité logique désigne une unité qu'un système d'exploitation (OS) reconnaît en tant que disque.

L'unité logique intègre un Réseau Redondant de
5 Disques Bon Marché (RAID) pour rendre redondantes des données mémorisées. Par conséquent, même s'il existe un problème dans certains des disques physiques 215, des données mémorisées ne vont pas être perdues.

Les unités logiques en tant que disques vir-
10 tuels 225 sont divisées en un groupe d'unités logiques 231 pour le système de mémorisation virtuel (0) et en un groupe d'unités logiques 241 pour le système de mémorisation virtuel (1). Le système de mémorisation virtuel (0) fait l'objet d'un accès par l'ordinateur virtuel (0) 131
15 et le système de mémorisation virtuel (1) fait l'objet d'un accès par l'ordinateur virtuel (1) 141.

L'hyperviseur de mémorisation 220 a un tableau de commande de disques virtuels 221, un tableau de traduction d'adresses de disque 222 et un tableau de com-
20 mande de ressources de mémoire 223.

Le tableau de commande de disques virtuels 221 (figure 2) mémorise le même contenu qu'un tableau de commande de disques virtuels 321 intégré au terminal de commande 300.

25 Le tableau de traduction d'adresses de disque 222 (figure 3) définit les relations entre des disques virtuels et des disques physiques et également les relations entre des adresses de disque virtuel et des adresses de disque physique. Le tableau de traduction d'adresses de disque 222 convertit des adresses de disque vir-
30 tuel en adresses de disque physique et vice versa.

Le tableau de commande de ressources de mémoire 223 mémorise le même contenu que le tableau de commande de ressources de mémoire 323 intégré au terminal de com-
35 mande 300.

Le terminal de commande 300 est un dispositif informatique qui commande le système informatique de manière globale et exécute un programme de commande d'ordinateur virtuel 310. Le programme de commande d'ordinateur virtuel 310 a le tableau de commande de disques virtuels 321, le tableau de commande de ressources de mémoire 323 et le tableau de commande de ressources de serveur 324.

Le tableau de commande de disques virtuels 321 mémorise le même contenu que le tableau de commande de disques virtuels 221 intégré au système de mémorisation 200.

Le tableau de commande de ressources de mémoire 323 (figure 4) définit les relations entre les ressources du système de mémorisation 200 et les ordinateurs virtuels. Le tableau de commande de ressources de mémoire 223 commande l'allocation de ressources de mémoire.

Le tableau de commande de ressources de serveur 324 (figure 5) définit les relations entre les ressources du système de serveur 100 et les ordinateurs virtuels. Le tableau de commande de ressources de serveur 324 commande les ressources informatiques du système de serveur 100.

Le terminal de commande 300 est connecté au système de serveur 100 et au système de mémorisation 200 via un réseau 410. Le système de serveur 100, le système de mémorisation 200 et le terminal de commande 300 reçoivent ou envoient des informations de commande de système informatique (le contenu des tableaux de commande) via le réseau 410.

De manière concrète, le tableau de commande de disques virtuels 321 est créé par le programme de commande d'ordinateur virtuel 310 et transmis au système de mémorisation 200 pour devenir le tableau de commande de disques virtuels 221. Le tableau de commande de disques virtuels 321 définit la configuration de systèmes de mé-

morisation virtuels correspondant aux ordinateurs virtuels. Le tableau de commande de disques virtuels 321 détermine quel ordinateur virtuel peut accéder à quelle unité logique.

5 Le tableau de commande de ressources de mémoire 323 est également créé par le programme de commande d'ordinateur virtuel 310 et transmis au système de mémorisation 200 pour devenir le tableau de commande de ressources de mémoire 223. Les données mises à jour dans ces ta-
10 bleaux sont reçues ou envoyées via le réseau 410.

 Le canal E/S 400 est un support de transmission qui permet une communication conformément à un protocole adapté à une transmission de données, tel qu'un protocole Fibre Canal. Le système de serveur 100 et le système de
15 mémorisation 200 peuvent être connectés sur une base bi-univoque ou via un réseau (SAN).

 Le réseau 410 est conçu pour permettre une communication de données et d'informations de commande entre des ordinateurs, par exemple, conformément au protocole
20 TCP/IP. Par exemple, il utilise Ethernet.

 Dans le premier mode de réalisation décrit ci-dessus, on suppose qu'un système de serveur 100 est connecté à un système de mémorisation 200. Cependant, concernant l'un ou l'autre du système de serveur 100 et
25 du système de mémorisation 200 ou les deux, plus d'un système peut être utilisé.

 La description ci-dessus suppose qu'un ordinateur virtuel correspond à un système de mémorisation virtuel. Cependant, plus d'un ordinateur virtuel peut être
30 connecté à un système de mémorisation virtuel ou un ordinateur virtuel peut être connecté à plus d'un système de mémorisation virtuel.

 La figure 2 représente un tableau de commande de disques virtuels conformément à un mode de réalisation
35 de la présente invention.

Comme mentionné ci-dessus, le tableau de commande de disques virtuels 221 est créé dans le terminal de commande 300 par l'intermédiaire d'une opération de l'utilisateur du terminal de commande 300 et un tableau
5 ayant le même contenu est mémorisé en tant que tableau de commande de disques virtuels 121 dans le système de serveur 100 et en tant que tableau de commande de disques virtuels 221 dans le système de mémorisation 200.

Le tableau de commande de disques virtuels 221
10 contient des numéros d'ordinateur virtuel 401, des numéros d'unité logique 402 et des numéros de disque virtuel 403 de telle manière qu'ils correspondent entre eux. Un numéro d'ordinateur virtuel 401 correspond à un ordinateur virtuel dans le système de serveur 100. Un numéro
15 d'unité logique 402 est un numéro affecté à une unité logique en tant que disque virtuel 225 identifié par un numéro de disque virtuel 403.

Le tableau de commande de disques virtuels 221 détermine quel ordinateur virtuel peut accéder à quelle
20 unité logique (à savoir à quel disque virtuel).

La figure 3 représente un tableau de traduction d'adresses de disque conformément à un mode de réalisation de la présente invention. Le tableau de traduction d'adresses de disque 222 est créé dans le système de mémorisation 200 par l'hyperviseur de mémorisation 220 et mémorise les relations entre des disques virtuels et des disques physiques et des relations entre des adresses de
25 disque virtuel et des adresses de disque physique, comme défini ci-dessus.

Le tableau de traduction d'adresses de disque
30 222 contient des numéros de disque virtuel 501, des adresses de bloc virtuel 502, des numéros de disque physique 503 et des adresses de bloc physique 504 de telle manière qu'ils correspondent entre eux. Un numéro de dis-
35 que virtuel 501 est un numéro affecté à un disque virtuel

225 créé par l'hyperviseur de mémorisation 220 et correspond à un numéro de disque virtuel 403 mémorisé dans le tableau de commande de disques virtuels 221. Une adresse de bloc virtuel 502 est une adresse d'un disque virtuel
5 225. Une adresse de bloc virtuel 502 correspond à une adresse de bloc physique 504 d'un disque physique 215 identifié par un numéro de disque physique 503. Spécifiquement, une adresse de bloc virtuel 0x00000000 d'un numéro de disque virtuel 121 correspond à une adresse de
10 bloc physique 0x00000000 d'un numéro de disque physique 8. En outre, une adresse de bloc virtuel 0x80000000 d'un numéro de disque virtuel 121 correspond à une adresse de bloc physique 0x00000000 d'un numéro de disque physique 9. En d'autres termes, un disque virtuel 121 est consti-
15 tué de disques physiques 8 et 9. Le tableau de traduction d'adresses de disque 222 peut convertir des adresses de disque virtuel en adresses de disque physique et vice versa.

La figure 4 représente un tableau de commande
20 de ressources de mémoire conformément à un mode de réalisation de la présente invention.

Comme mentionné ci-dessus, le tableau de commande de ressources de mémoire 323 est créé dans le terminal de commande 300 par l'intermédiaire d'une opération
25 de l'utilisateur du terminal de commande 300 et un tableau ayant le même contenu est mémorisé en tant que tableau de commande de ressources de mémoire 223 dans le système de serveur 200.

Dans le deuxième mode de réalisation qui va
30 être décrit ultérieurement (figure 13), un tableau de commande de ressources de mémoire 223 est créé dans le système de mémorisation 200. Dans le troisième mode de réalisation qui va être décrit ultérieurement (figure 14), un tableau de commande de ressources de mémoire 223 est créé
35 dans le système de serveur 100.

Le tableau de commande de ressources de mémoire 323 contient des numéros d'ordinateur virtuel 601, des numéros de disque virtuel 602, des capacités de cache de disque 603, des numéros de processeur de commande 604 et des numéros d'adaptateur E/S 605 de telle manière qu'ils correspondent les uns aux autres. Le tableau de commande de ressources de mémoire 323 mémorise les relations entre les ressources du système de mémorisation 200 (disques virtuels 225, processeurs de commande 212, adaptateurs E/S 213 et caches de disque 214) et les ordinateurs virtuels.

Un numéro d'ordinateur virtuel 601 correspond à un ordinateur virtuel du système de serveur 100. Un numéro de disque virtuel 602 est un numéro affecté à un disque virtuel 225 créé par l'hyperviseur de mémorisation 220, qui désigne un disque virtuel alloué à un ordinateur virtuel identifié par un numéro d'ordinateur virtuel 601. Ce numéro de disque virtuel 602 correspond à un numéro de disque virtuel 403 mémorisé dans le tableau de commande de disques virtuels 221.

Une capacité de cache de disque 603 est la capacité d'un cache de disque 214 qui est allouée à un ordinateur virtuel identifié par un numéro d'ordinateur virtuel 601. Un numéro de processeur de commande 604 désigne un processeur de commande 212 qui commande l'accès à partir d'un ordinateur virtuel identifié par un numéro d'ordinateur virtuel 601 (à un disque virtuel identifié par un numéro de disque virtuel 602).

Un numéro d'adaptateur E/S 605 désigne un adaptateur E/S 213 qui est en charge de l'accès à partir d'un ordinateur virtuel identifié par un numéro d'ordinateur virtuel 601 (à un disque virtuel identifié par un numéro de disque virtuel 602).

Spécifiquement, trois disques virtuels 225 (numéros de disque 121 à 123) sont alloués à l'ordinateur

virtuel (0) 131. Pour avoir accès à ces disques virtuels 225 (numéros de disque 121 à 123), l'ordinateur virtuel (0) 131 peut utiliser 512 mégaoctets de cache de disque. Pour avoir accès à partir de l'ordinateur virtuel (0) 131
5 aux disques virtuels 225 (numéros de disque 121 à 123), trois adaptateurs E/S (numéros 0 à 2) sont utilisés. Trois processeurs de commande (CPU) (numéros 48 à 50) sont utilisés pour traiter l'accès à partir de l'ordinateur virtuel (0) 131 aux disques virtuels 225 (numéros
10 121 à 123).

La figure 5 représente un tableau de commande de ressources de serveur conformément à un mode de réalisation de la présente invention.

Comme mentionné ci-dessus, dans le premier mode
15 de réalisation, le tableau de commande de ressources de serveur 324 est créé dans le terminal de commande 300 par le programme de commande d'ordinateur virtuel 310.

Dans le deuxième mode de réalisation qui va être décrit ultérieurement (figure 13), un tableau de
20 commande de ressources de serveur 224 est créé dans le système de mémorisation 200. Dans le troisième mode de réalisation qui va être décrit ultérieurement (figure 14), un tableau de commande de ressources de serveur 124 est créé dans le système de serveur 100.

25 Le tableau de commande de ressources de serveur contient des numéros d'ordinateur virtuel 701, une allocation de CPU (pourcentage) 702, des capacités de mémoire 703 et des numéros d'adaptateur E/S 704 de telle manière qu'ils correspondent entre eux. Le tableau de commande de
30 ressources de serveur 324 mémorise les relations entre les ressources du système de serveur 100 (CPU 111, mémoire 112 et adaptateur E/S 114).

Un numéro d'ordinateur virtuel 701 correspond à un ordinateur virtuel du système de serveur 100. L'allocation de CPU 702 est la proportion de la CPU du système
35

de serveur 100 qui est allouée à cet ordinateur virtuel. Une capacité de mémoire 703 est la capacité de la mémoire 112 qui est allouée à cet ordinateur virtuel. Un numéro d'adaptateur E/S 704 désigne un adaptateur E/S 213 qui
5 est en charge de l'accès à partir de l'ordinateur virtuel au système de mémorisation 200.

La figure 6 représente un processus d'allocation de ressources conformément à un mode de réalisation de la présente invention.

10 Tout d'abord, l'utilisateur actionne le terminal 300 pour allouer les ressources informatiques du système de serveur 100 (CPU 111, mémoire 112, adaptateur E/S 114, etc) et les ressources du système de mémorisation 200 (CPU 212, adaptateur E/S 213, cache de disque 214, et
15 disque virtuel 225) aux ordinateurs virtuels individuels afin de mettre à jour le tableau de commande de ressources de serveur 324 (étape S101). Le terminal de commande 300 transmet les données d'allocation de ressources au système de serveur 100 (étape S102).

20 Lorsque le système de serveur 100 reçoit les données d'allocation de ressources en provenance du terminal de commande 300, il alloue les ressources informatiques du système de serveur 100 pour créer des ordinateurs individuels (étape S103). Après la création des ordinateurs virtuels, il informe le terminal de commande
25 300 de la création des ordinateurs virtuels (étape S104).

Lorsque le terminal de commande 300 reçoit la notification de création des ordinateurs virtuels en provenance du système de serveur 100, il transmet des données d'allocation de ressources (données destinées à mettre à jour le tableau de commande de ressources de mémoire) au système de mémorisation 200 (étape S105).

Lorsque le système de mémorisation 200 reçoit les données d'allocation de ressources en provenance du terminal de commande 300, il met à jour le tableau de
35

commande de ressources de mémoire 223 et le tableau de commande de disques virtuels 221 conformément aux données d'allocation pour allouer les ressources du système de mémorisation 200 (étape S106). Lorsque nécessaire, le tableau de commande de disques virtuels 221 et le tableau de traduction d'adresses de disque 222 sont mis à jour pour créer ou mettre à jour des systèmes de mémorisation virtuels (étape S106). Après la création des systèmes de mémorisation virtuels, le système de mémorisation 200 informe le terminal de commande 300 de la création de systèmes de mémorisation virtuels (étape S107).

La figure 7 représente le processus d'entrée/sortie de données avec le système de mémorisation 200.

Le système de mémorisation 200 reçoit une instruction d'entrée/sortie en provenance du système de serveur 100 (étape S111). Cette instruction d'entrée/sortie est transmise à l'hyperviseur de mémorisation 220. L'hyperviseur de mémorisation 220 lit un numéro d'ordinateur virtuel source 1302 et un numéro d'ordinateur virtuel de destination 1303 qui sont inclus dans l'instruction d'entrée/sortie (en-tête de communication de l'hyperviseur 1203 (voir figures 9 et 10) (étape S112)). L'hyperviseur de mémorisation 220 transmet une charge utile de communication de l'hyperviseur 1204 à un système de mémorisation virtuel correspondant au numéro d'ordinateur virtuel de destination 1303 (étape S113). Dans ce mode de réalisation, la charge utile de communication de l'hyperviseur 1204 inclut une instruction E/S de disque que le système de mémorisation virtuel exécute.

Le système de mémorisation virtuel acquiert le numéro du disque virtuel devant faire l'objet d'un accès et identifie le disque virtuel approprié 225 et accède à celui-ci (étape S114).

L'accès au disque virtuel 225 est accepté par l'hyperviseur de mémorisation 220. L'hyperviseur de mémorisation 220 utilise le tableau de traduction d'adresses de disque 222 pour identifier l'adresse de bloc physique du disque physique correspondant à l'adresse de bloc virtuel du disque virtuel devant faire l'objet d'un accès et traduit l'accès au disque virtuel 225 en un accès au disque physique 215. Ensuite, l'hyperviseur de mémorisation 220 accède au disque physique 215 et lit ou écrit des données (étape S115).

A la fin de l'entrée/sortie de données avec le disque physique 215, l'hyperviseur de mémorisation 220 informe le système de mémorisation virtuel du résultat de l'entrée/sortie de données (étape S116). Lorsque le système de mémorisation virtuel reçoit le résultat de l'entrée/sortie de données en provenance de l'hyperviseur de mémorisation 220, il informe l'ordinateur virtuel du résultat de l'entrée/sortie de données via l'hyperviseur de mémorisation 220 et l'hyperviseur 110 (étapes S117, S118, S119).

Ensuite, on va décrire la manière avec laquelle le système de serveur 100 et le système de mémorisation 200 traitent une instruction d'entrée/sortie. Une communication entre le système de serveur 100 et le système de mémorisation 200 est établie via le canal E/S 400. Une communication via le canal E/S 400 est effectuée par un protocole ayant une structure de couche telle qu'un protocole Fibre Canal ou Ethernet par exemple.

La figure 8 représente la structure de couche d'un protocole de communication pour le canal E/S 400.

Lorsque le système d'exploitation (0) 132 sur l'ordinateur virtuel (0) 131 accède à une unité logique du système de mémorisation 200, une entrée/sortie a lieu conformément à un protocole E/S de disque (par exemple, SCSI). Dans ce mode de réalisation, une couche de proto-

cole E/S de disque est appelée une "couche E/S de disque" 1100, 1106. Une instruction E/S de disque émise par le système d'exploitation (0) 132 est reçue par l'hyperviseur 120 et une couche de protocole de communication existe entre l'hyperviseur 120 et l'hyperviseur de mémoire 220. Ceci est appelé une "couche de communication d'hyperviseur" 1101, 1105. En outre, dans ce mode de réalisation, une couche pour une communication générale via le canal E/S 400 est appelée une "couche de protocole de canal E/S" 1102, 1104. Une couche matérielle telle qu'un support physique est appelée une "couche physique" 1103. Grâce à cette structure de couche, les couches E/S de disque 1100 et 1106 et les couches de communication d'hyperviseur 1101 et 1105 ne sont pas affectées par un changement dans le support physique du canal E/S 400.

Une instruction E/S de disque émise par le système d'exploitation (0) 132 est transmise à l'ordinateur virtuel (0) 131. L'ordinateur virtuel (0) 131 envoie l'instruction E/S au système de mémorisation virtuel (0). En réalité, l'hyperviseur 120 reçoit l'instruction E/S. L'hyperviseur 120 ajoute des informations à l'instruction E/S de disque (voir figure 9) et transmet celle-ci à l'hyperviseur de mémorisation 220. L'hyperviseur de mémorisation 220 reçoit cette instruction, extrait l'instruction E/S de disque de celle-ci et transmet l'instruction au système de mémorisation virtuel (0) 230. Lorsque la structure de couche est utilisée pour établir une communication de cette manière, le système d'exploitation (0) 132 identifie celle-ci comme s'il s'agissait d'une communication directement avec le système de mémorisation virtuel (0) 230.

La figure 9 représente une communication de données entre le système de serveur 100 et le système de mémorisation 200.

Dans ce mode de réalisation, une communication via le canal E/S 400 est effectuée trame par trame 1200 comme via le protocole Fibre Canal ou Ethernet. Une trame 1200 est constituée d'un en-tête de protocole de canal E/S 1201 et d'une charge utile de protocole de canal E/S 1202. L'en-tête de protocole de canal E/S 1201 contient des informations de commande requises pour une communication via les couches de protocole de canal E/S 1102 et 1104. Bien que ceci ne soit pas représenté, les informations de commande peuvent être un identifiant de source ou un identifiant de destination. La charge utile de protocole de canal E/S 1202 est constituée de données qui sont transmises via les couches de protocole de canal E/S 1102 et 1104. Les couches de protocole de canal E/S 1102 et 1104 ne sont pas concernées par les données.

La charge utile de protocole de canal E/S 1202 est constituée d'un en-tête de communication d'hyperviseur 1203 et d'une charge utile de communication d'hyperviseur 1204. L'en-tête de communication d'hyperviseur 1203 contient des informations de commande requises pour établir une communication via les couches de communication d'hyperviseur 1101 et 1105 (décrites ultérieurement). La charge utile de communication d'hyperviseur 1204 est constituée de données qui sont transmises via les couches de communication d'hyperviseur 1101 et 1105. Les couches de communication d'hyperviseur 1101 et 1105 ne sont pas concernées par les données.

Dans ce mode de réalisation, la charge utile de communication d'hyperviseur 1204 est constituée d'informations nécessaires pour établir une communication entre les couches d'E/S de disque 1100 et 1106. Spécifiquement, les informations incluent des instructions E/S de disque ou des données à transmettre. Dans ce mode de réalisation, la charge utile de communication d'hyperviseur 1204 inclut des informations concernant les couches E/S de

disque 1100 et 1106 du fait que les couches E/S de disque sont situées au-dessus des couches de communication d'hyperviseur 1101 et 1105. Cependant, si une communication est établie entre l'hyperviseur et l'hyperviseur de mémoire, des informations autres que des informations de couche d'E/S de disque sont incluses.

La figure 10 représente le contenu de l'en-tête de communication d'hyperviseur 1203.

L'en-tête de communication d'hyperviseur 1203 est unique aux modes de réalisation de la présente invention. Il est constitué d'un numéro d'hyperviseur source 1300, d'un numéro d'hyperviseur de destination 1301, d'un numéro d'ordinateur virtuel source 1302 et d'un numéro d'ordinateur virtuel de destination 1303. Dans ce mode de réalisation, des identifiants uniques sont attribués à l'hyperviseur et à l'hyperviseur de mémorisation pour faire face à un système informatique qui a une pluralité de systèmes de serveur 100 et de systèmes de mémorisation 200.

Le numéro d'hyperviseur source 1300 est un identifiant d'un hyperviseur ou d'un hyperviseur de mémorisation qui envoie la trame.

Le numéro d'hyperviseur de destination 1301 est un identifiant d'un hyperviseur ou d'un hyperviseur de mémorisation qui reçoit la trame.

Le numéro d'ordinateur virtuel source 1302 est un identifiant d'un ordinateur virtuel ou d'un système de mémorisation virtuel qui envoie la trame.

Le numéro d'ordinateur virtuel de destination 1303 est un identifiant d'un ordinateur virtuel ou d'un système de mémorisation virtuel qui reçoit la trame.

Les figures 11 et 12 représentent des écrans de configuration du système conformément à un mode de réalisation de la présente invention.

Dans la partie supérieure de l'écran, des pages spécifiant des ressources allouées à chaque ordinateur virtuel sont affichées. Dans la partie inférieure de l'écran, une fenêtre "ressources" représentant toutes les
5 ressources du système de serveur 100 et du système de mémorisation 200 est affichée. En plus de toutes ces ressources, la fenêtre peut afficher des ressources qui ne sont pas utilisées (ou déjà en utilisation).

L'administrateur peut spécifier des ressources
10 pour chaque ordinateur virtuel en écrivant des ressources du système de serveur ou du système de mémorisation dans chaque page de la partie supérieure de l'écran ou en déplaçant des ressources depuis la fenêtre "ressources" de la partie inférieure de l'écran.

15 En outre, l'administrateur peut spécifier des performances requises pour un ordinateur virtuel (et un système de mémorisation virtuel) sans avoir besoin d'exécuter la tâche d'allocation de ressources à chaque ordinateur virtuel et à chaque système de mémorisation vir-
20 tuel de sorte que les ressources requises pour les performances sont calculées et établies pour l'ordinateur virtuel et le système de mémorisation virtuel.

Par exemple, pour un ordinateur virtuel qui met l'accent sur les performances de lecture de données, une
25 valeur supérieure doit être établie pour la capacité du cache de disque 214 qui est allouée à un système de mémorisation virtuel correspondant. Si toutes les ressources du cache de disque 214 sont petites en termes de quantité et que la capacité du cache de disque 214 allouée au sys-
30 tème de mémorisation virtuel est petite, une zone de mémoire plus grande doit être allouée à l'ordinateur virtuel. D'autre part, si toutes les ressources du cache de disque 214 sont grandes en termes de quantité et que la capacité du cache de disque 214 allouée au système de mé-

morisation virtuel est petite, une zone de mémoire plus petite est allouée à l'ordinateur virtuel.

Si un logiciel d'application s'exécutant sur un ordinateur virtuel accède de manière aléatoire à une large zone sur le disque, le cache est moins efficace et ainsi l'allocation de la capacité du cache de disque doit être petite. Pour un logiciel d'application qui offre la fonction de lecture au fil de l'eau d'images mobiles ou d'autres fonctions multimédias, la capacité du cache de disque allouée au système de mémorisation virtuel doit être grande et la capacité de la mémoire allouée à l'ordinateur virtuelle doit être également grande.

Lorsque le nombre de systèmes de serveur ou de systèmes de mémorisation augmente ou diminue, des ordinateurs virtuels et des systèmes de mémorisation virtuels peuvent être configurés sur cet écran.

Ainsi, le premier mode de réalisation de la présente invention est résumé de la manière suivante. Il a un tableau de commande de ressources de serveur, un tableau de commande de ressources de mémoire et un tableau de commande de disques virtuels. L'hyperviseur partitionne de manière logique des ressources informatiques conformément aux paramètres du tableau de commande de ressources de serveur et amène les partitions obtenues en résultat à fonctionner indépendamment en tant qu'ordinateurs virtuels. L'hyperviseur de mémorisation partitionne de manière logique les ressources de mémoire conformément aux paramètres du tableau de commande de ressources de mémoire et amène les partitions obtenues en résultat à fonctionner indépendamment en tant que systèmes de mémorisation virtuels. Par conséquent, les ressources du système informatique incluant le système de serveur et le système de mémorisation peuvent

être entièrement commandées et allouées de manière optimale.

Lors de la reconfiguration d'un ordinateur virtuel, un système de mémorisation virtuel correspondant
5 peut être reconfiguré. Cela signifie que l'ordinateur virtuel et le système de mémorisation virtuel n'ont pas besoin d'être configurés séparément et que les ressources de l'ordinateur virtuel et du système de mémorisation virtuel peuvent être paramétrées, en prenant en considé-
10 ration les performances globales du système informatique. Des ressources telles que le cache de disque 214 qui ne peuvent pas être commandées par le terminal de commande 300 dans la technique habituelle peuvent être définies au même moment en tant que ressources informatiques virtuel-
15 les.

Dans ce mode de réalisation, l'utilisateur peut effectuer un paramétrage détaillé pour un "disque" sur l'écran de configuration représenté sur la figure 11 en appelant une fenêtre de paramétrage détaillé. Il va sans
20 dire que la présente invention ne repose pas sur le procédé d'affichage de l'écran.

La figure 12 représente une fenêtre de paramétrage détaillé.

La fenêtre de paramétrage détaillé (figure 12)
25 peut être appelée pour chaque ordinateur virtuel en cliquant sur le bouton "détail" représenté sur la figure 11. Dans ce mode de réalisation, une unité logique 0 est constituée de deux disques physiques (disques physiques 8 et 9). L'information "10 000 tpm" qui est indiquée à côté
30 de chaque numéro de disque physique indique que les disques physiques 8 et 9 sont des supports d'enregistrement magnétique tels que des disques magnétiques qui effectuent 10 000 tours par minute. Le nombre de tours par minute du disque magnétique est un facteur important qui
35 définit les performances en tant que disque physique.

Pour une application qui nécessite des performances élevées, l'utilisateur peut sélectionner un disque physique à hautes performances dans cette fenêtre pour constituer une unité logique. L'utilisateur peut également sélectionner plus de disques physiques pour accroître les performances de l'unité logique.

Comme décrit ci-dessus, conformément à la présente invention, des ressources de mémoire peuvent être allouées en association avec des ordinateurs virtuels et il est possible d'allouer des ressources d'un système informatique entier incluant un système de serveur et un système de mémorisation de manière optimale.

La figure 13 représente la configuration d'un système informatique conformément à un deuxième mode de réalisation de la présente invention.

Contrairement au premier mode de réalisation (figure 1), le deuxième mode de réalisation n'utilise pas de terminal de commande 300 et dispose à la place de la même fonction que celle du terminal de commande 300 du premier mode de réalisation, dans le système de mémorisation 200. Les mêmes éléments que ceux du premier mode de réalisation sont désignés par les mêmes références numériques et leur description détaillée est omise.

Conformément au deuxième mode de réalisation, un système informatique est constitué d'un système de serveur 100 sur lequel un logiciel d'application s'exécute, d'un système de mémorisation 200 qui commande le système informatique entier et mémorise des données requises pour le fonctionnement du système de serveur 100 et d'un terminal de commande 350 qui envoie des instructions au système de mémorisation 200 pour le fonctionnement du système informatique entier.

Le système de serveur 100 a un système informatique physique 110 qui intègre des ressources telles qu'une CPU 111, une mémoire 112, un bus E/S 113 et des

adaptateurs E/S 114 et 115. La configuration et le fonctionnement du système de serveur 100 sont identiques à ceux du premier mode de réalisation.

Le système de mémorisation 200 a un système de
5 mémorisation physique 210 incluant des ressources telles qu'un bloc de commande de mémorisation physique 211 et des disques physiques 215.

L'hyperviseur de mémorisation 220 a un tableau
10 de commande de disques virtuels 221, un tableau de traduction d'adresses de disque 222, un tableau de commande de ressources de mémoire 223 et un tableau de commande de ressources de serveur 224.

Le tableau de commande de disques virtuels 221
15 (figure 2), le tableau de traduction d'adresses de disque 222 (figure 3) et le tableau de commande de ressources de mémoire 223 (figure 4) sont identiques à ceux du premier mode de réalisation. Le tableau de commande de ressources de serveur 224 (figure 5) définit les relations entre les ressources du système de serveur 100 et les ordinateurs
20 virtuels. Le tableau de commande de ressources de serveur 224 est utilisé pour commander les ressources informatiques du système de serveur 100.

L'hyperviseur de mémorisation 220 commande de
25 manière globale le système informatique en utilisant les tableaux de commande 221, 223 et 224.

Un programme de commande d'ordinateur virtuel
qui commande entièrement le système informatique en utilisant les tableaux de commande 221, 223 et 224 s'exécute dans l'hyperviseur de mémorisation 220.

30 Le terminal de commande 350 est un dispositif informatique qui est utilisé pour établir des informations de commande pour le système informatique. Il est connecté au système de mémorisation 200. Par conséquent, l'administrateur peut mettre à jour le tableau de commande
35 de ressources de mémoire 223 et le tableau de com-

mande de ressources de serveur 224 en actionnant le terminal de commande 350.

Ainsi, en plus des effets mentionnés ci-dessus du premier mode de réalisation, le deuxième mode de réalisation a pour effet que les systèmes de mémorisation virtuels peuvent être commandés d'une manière à s'adapter aux ordinateurs virtuels, sans terminal de commande séparé, du fait que la même fonction que celle du terminal de commande 300 est fournie dans le système de mémorisation 200.

La figure 14 représente la configuration d'un système informatique conformément à un troisième mode de réalisation de la présente invention.

Contrairement au premier mode de réalisation (figure 1) ou au deuxième mode de réalisation (figure 13), le troisième mode de réalisation n'utilise pas le terminal de commande 300 et dispose à la place de la même fonction que celle du terminal de commande 300 du premier mode de réalisation, dans le système de serveur 100. Les mêmes éléments que ceux du premier mode de réalisation sont désignés par les mêmes références numériques et leur description détaillée est omise.

Conformément au troisième mode de réalisation, un système informatique est constitué d'un système de serveur 100 qui a un logiciel d'application s'exécutant sur celui-ci et commande le système informatique entier et d'un système de mémorisation 200 qui mémorise des données requises pour le fonctionnement du système de serveur 100.

Le système de serveur 100 a un système informatique physique 110 qui intègre des ressources telles qu'une CPU 111, une mémoire 112, un bus E/S 113 et des adaptateurs E/S 114 et 115. La configuration du système informatique physique 110 est identique à celle du premier mode de réalisation.

Les ressources du système informatique physique 110 sont commandées par un hyperviseur 120. L'hyperviseur 120 crée un ordinateur virtuel (0) 131 sur la base des ressources informatiques utilisées par le système d'exploitation (0) 132 et un ordinateur virtuel (1) 141 sur la base des ressources informatiques utilisées par le système d'exploitation (1) 142, dans le système informatique physique 110. L'hyperviseur 120 a un tableau de commande de disques virtuels 121, un tableau de commande de ressources de mémoire 123 et un tableau de commande de ressources de serveur 124.

Le tableau de commande de disques virtuels 121 mémorise le même contenu qu'un tableau de commande de disques virtuels 221 dans le système de mémorisation 200.

Le tableau de commande de ressources de mémoire 123 (figure 4) définit les relations entre les ressources du système de mémorisation 200 et les ordinateurs virtuels. Le tableau de commande de ressources de mémoire 223 commande l'allocation de ressources de mémoire.

Le tableau de commande de ressources de serveur 124 (figure 5) définit les relations entre les ressources du système de serveur 100 et les ordinateurs virtuels. Le tableau de commande de ressources de serveur 224 est utilisé pour commander les ressources informatiques du système de serveur 100.

Un programme de commande d'ordinateurs virtuels qui commande entièrement le système informatique en utilisant les tableaux de commande 121, 123 et 124 s'exécute dans l'hyperviseur 120. Par conséquent, l'administrateur peut mettre à jour les paramètres du tableau de commande de ressources de mémoire 123 et du tableau de commande de ressources de serveur 124 en utilisant le système de serveur 100.

Le système de mémorisation 200 inclut un système de mémorisation physique 210 ayant des ressources

telles qu'un bloc de commande de mémorisation physique 211 et des disques physiques 215. La configuration du système de mémorisation 200 est identique à celle du premier mode de réalisation. Le tableau de commande de res-
5 sources de mémoire 223 mémorise le même contenu que le tableau de commande de ressources de mémoire 123 dans le système de serveur 100.

Ainsi, en plus des effets mentionnés ci-dessus du premier mode de réalisation, le troisième mode de ré-
10 alisation a pour effet que les systèmes de mémorisation virtuels peuvent être commandés de manière à s'adapter aux ordinateurs virtuels, sans terminal de commande séparé du système de serveur 110, du fait que la même fonction que celle du terminal de commande 300 est fournie
15 dans le système de serveur 100.

La figure 15 représente la configuration d'un système informatique conformément à un quatrième mode de réalisation de la présente invention.

Le quatrième mode de réalisation est différent
20 des modes de réalisation ci-dessus en ce qui concerne la structure du bloc de commande de mémorisation physique 1100. Dans le bloc de commande de mémorisation physique 1100, un ou plusieurs adaptateurs de canal 1101, un ou plusieurs adaptateurs de disque 1102, un ou plusieurs caches de disque 1103 et un ou plusieurs processeurs de
25 commande 212 sont connectés via un réseau interne 1104. Les adaptateurs de canal commandent la communication avec le système de serveur 100 et les adaptateurs de disque 1102 commandent les disques physiques.

30 Dans le bloc de commande de mémorisation physique 1100 ayant le réseau interne 1104, la largeur de bande du réseau 1104 est un facteur important qui influence les performances du système de mémorisation 200. Pour cette raison, dans ce mode de réalisation, l'hyperviseur de mémorisation 220 effectue l'allocation de la
35

largeur de bande du réseau interne 1104 entre l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141 et le processeur de commande 212 traite l'entrée et la sortie conformément à l'allocation. Divers procédés de commande de largeur de bande sont disponibles mais la présente invention ne repose pas sur le procédé de commande de largeur de bande.

La structure des disques virtuels 225 influence également les performances. Comme mentionné précédemment, les disques virtuels 225 sont des zones de mémorisation des disques physiques 215 que l'hyperviseur de mémorisation 220 fait reconnaître en tant que disques physiques à l'ordinateur virtuel (0) 131 et à l'ordinateur virtuel (1) 141. Un procédé de création d'un disque virtuel 225 ayant des performances d'entrée/sortie améliorées consiste à extraire des parties de zones de mémoire d'une pluralité de disques physiques 215 et de combiner celles-ci dans un disque virtuel 225. Ceci est dû au fait que les demandes d'entrée/sortie de l'ordinateur virtuel (0) 131 et de l'ordinateur virtuel (1) 141 sont traitées par l'intermédiaire d'une opération parallèle simultanée de plusieurs disques physiques 215.

Cette approche est décrite ci-dessous en se reportant aux figures 16(a) et 16(b).

Comme représenté sur la figure 16(a), un disque virtuel 1200 est constitué d'un disque physique 1201. D'autre part, comme représenté sur la figure 16(b), un disque virtuel 1202 est constitué de parties de zones de mémorisation de trois disques physiques 1203, 1204 et 1205. Les performances du disque physique 1201 peuvent être exprimées par le nombre de processus d'entrée/sortie exécutés dans une unité de temps. Lorsque x représente ce nombre, les performances d'entrée/sortie du disque virtuel 1200 sont exprimées sous la forme x . Au contraire, en supposant que l'ordinateur virtuel (0) 131 et l'ordi-

nateur virtuel (1) 141 accèdent à toutes les zones de mémorisation du disque virtuel 1202 de manière égale, les performances du disque virtuel 1202 sont exprimées sous la forme 3x du fait que les disques physiques 1203, 1204
5 et 1205 fonctionnent simultanément en parallèle. Ainsi, les performances du disque virtuel 1202 dépendent en grande partie du nombre de disques physiques 215 qui constituent celui-ci.

Par conséquent, il est souhaitable que le nombre de disques physiques 215 qui constituent un disque virtuel 225 puisse être spécifié dans le terminal de commande 300 conformément au logiciel d'application etc. que l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141 exécutent. Par exemple, si l'ordinateur virtuel (0)
15 131 et l'ordinateur virtuel (1) 141 exécutent un logiciel d'application qui permet un accès aléatoire à une grande zone du disque, le cache de disque 214 est moins efficace que celui décrit précédemment. Dans ce cas, les performances d'accès du disque physique 215 sont un facteur dominant qui détermine les performances du disque virtuel
20 225. Pour cette raison, le nombre de disques physiques 215 qui constituent un disque virtuel 225 est accru afin d'améliorer les performances du disque virtuel 215.

Le processeur de commande 212 est également
25 l'un des facteurs qui détermine les performances d'entrée/sortie du système de mémorisation 200. Il est également souhaitable que l'utilisateur puisse spécifier le taux d'allocation du processeur de commande 212 dans le terminal de commande 300 conformément aux performances
30 d'entrée/sortie requises pour l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141 et le logiciel d'application. En fonction de la manière selon laquelle le système de mémorisation 200 est constitué, il est également possible que l'adaptateur de canal 1101 et l'adaptateur
35 de disque 1102 intègrent tous les deux un processeur

de commande 212. Si ceci est le cas, l'adaptateur de canal 1101 et l'adaptateur de disque 1102 qui sont en charge de l'entrée/sortie de données avec l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141 sont spécifiés dans le terminal de commande.

Le tableau de commande de ressources de mémoire 223 doit être modifié de sorte que les ressources (réseau interne 1104, disques physiques 215, processeurs de commande 212, etc.) du système de mémorisation 200 peuvent être spécifiées dans le terminal de commande 300 comme ci-dessus.

La figure 17 représente un tableau de commande de ressources de mémoire 223 conformément au quatrième mode de réalisation de la présente invention.

Le tableau représenté sur la figure 17 contient une colonne "largeur de bande de réseau interne" 1300 en tant que colonne supplémentaire. Cette colonne est utilisée pour spécifier le taux d'allocation de la largeur de bande du réseau interne 1104 pour l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141. Dans ce mode de réalisation, le taux d'allocation est exprimé sous forme de pourcentage par rapport à la largeur de bande totale. Le processeur de commande 212 surveille la largeur de bande du réseau interne utilisée par l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141, et retarde des processus d'entrée/sortie lorsque nécessaire pour empêcher la largeur de bande du réseau interne de dépasser un niveau pré-établi.

Des processeurs de commande sont alloués via l'utilisation de la colonne "processeur de commande" 604 du tableau de commande de ressources de mémoire 223. Cette colonne permet de spécifier les processeurs de commande 212 qui sont en charge de l'entrée/sortie avec l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141. On s'attend à ce que plus le nombre de processeurs

de commande alloués à un ordinateur virtuel est important, plus les performances d'entrée/sortie qu'il délivre sont élevées. Il est également possible qu'un processeur de commande 212 soit en charge de l'entrée/sortie avec à la fois l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141. Si ceci est le cas, le processeur de commande 212 surveille le temps de CPU que chaque ordinateur virtuel utilise et commande ainsi l'allocation de temps de CPU entre l'ordinateur virtuel (0) 131 et l'ordinateur virtuel (1) 141.

L'allocation des disques physiques est commandée par le tableau de commande de disques virtuels 221 de la même manière que dans les modes de réalisation ci-dessus.

La figure 18 représente un écran de configuration de système informatique conformément au quatrième mode de réalisation de la présente invention.

Un champ destiné à l'entrée du nombre de processeurs de commande 212 pour l'ordinateur virtuel (0) 131 est situé à droite du mot "CPU" dans la fenêtre supérieure. Un champ destiné à l'entrée de la capacité du cache de disque qui est allouée à l'ordinateur virtuel (0) 131 est situé à droite des mots "cache de disque". Un champ destiné à l'entrée de la largeur de bande (taux d'allocation) du réseau interne 1104 du système de mémoire 200 qui est alloué à l'ordinateur virtuel (0) 131 est situé à droite des mots "largeur de bande du réseau interne". Un champ destiné à l'entrée du nombre d'unités logiques 231 qui sont allouées à l'ordinateur virtuel (0) 131 est situé à droite du mot "disque". Une fenêtre de paramétrage détaillé (figure 12) qui représente des disques physiques en tant que constituants de chaque unité logique et permet un paramétrage détaillé est appelée en cliquant sur le bouton "détail" situé sur la ligne "disque".

REVENDICATIONS

1. Système de mémorisation (200) comportant :
- une pluralité de ports conçus pour recevoir des données en provenance d'au moins un dispositif de traitement d'informations,
 - une pluralité de processeurs (212) qui commandent le transfert de données reçues par lesdits ports,
 - une pluralité de ressources de mémoire (213, 214) qui mémorisent des données sur la base de la commande par lesdits processeurs,
 - une pluralité de groupes RAID (Réseau Redondant de Disques Indépendants), chacun d'eux étant associé à une pluralité d'unités de disque (215), lesdites unités de disque mémorisant une pluralité de données et des données de parité associées aux données mémorisées dans lesdites unités de disque de l'un desdits groupes RAID, et
 - une pluralité de groupes de ressources (221, 222, 223), chacun desdits groupes de ressources étant divisé par au moins une partition logique, chaque groupe de ressources ayant une pluralité de types de ressources incluant au moins l'un desdits ports, au moins l'un desdits processeurs, au moins l'une desdites ressources de mémoire, et au moins l'un desdits groupes RAID,
- dans lequel une quantité desdites ressources dans chacun desdits groupes de ressources est attribuée sur la base d'une performance requise pour chacun desdits groupes de ressources.
2. Système de mémorisation selon la revendication 1, dans lequel :
- ladite quantité inclut le nombre desdits processeurs dans un premier groupe de ressources desdits groupes de ressources, attribué sur la base de ladite performance requise.

3. Système de mémorisation selon la revendication 1, dans lequel :

ladite quantité inclut le nombre desdites ressources de mémoire dans un premier groupe de ressources desdits groupes de ressources, attribué sur la base de ladite performance requise.

4. Système de mémorisation selon la revendication 1, dans lequel :

ladite quantité inclut le nombre desdits ports dans un premier groupe de ressources desdites groupes de ressources, attribué sur la base de ladite performance requise.

5. Système de mémorisation selon la revendication 1, dans lequel :

au moins un niveau RAID dans un premier groupe de ressources desdits groupes de ressources est attribué sur la base de ladite performance requise.

6. Système de mémorisation selon la revendication 1, dans lequel :

lesdits ports dans un premier groupe de ressources desdits groupes de ressources sont couplés à au moins un premier dispositif de traitement d'informations et sont attribués sur la base de ladite performance requise, et

ladite performance requise est déterminée sur la base d'une performance requise pour le traitement de données envoyées par ledit premier dispositif de traitement d'informations.

7. Système de mémorisation selon la revendication 1, dans lequel :

lesdits ports dans un premier groupe de ressources desdits groupes de ressources sont couplés à au moins un premier dispositif de traitement d'informations et sont attribués sur la base de ladite performance requise, et

ladite performance requise est déterminée sur la base d'une performance requise pour mémoriser des données envoyées par ledit premier dispositif de traitement d'informations à au moins un premier groupe RAID parmi lesdits groupes RAID dudit premier groupe de ressources.

8. Système de mémorisation selon la revendication 1, dans lequel :

lesdits ports dans un premier groupe de ressources desdits groupes de ressources sont couplés à au moins un premier serveur virtuel, lequel est créé en divisant ledit dispositif de traitement d'informations par au moins une partition logique et dispose d'une pluralité de types de ressources dans ledit dispositif de traitement d'informations, et sont attribués sur la base de ladite performance requise, et

ladite performance requise est déterminée sur la base d'une performance requise pour traiter des données envoyées par ledit premier serveur virtuel.

9. Système de mémorisation selon la revendication 1, dans lequel :

lesdits ports dans un premier groupe de ressources desdits groupes de ressources sont couplés à au moins un premier serveur virtuel, lequel est créé en divisant ledit dispositif de traitement d'informations par au moins une partition logique et dispose d'une pluralité de types de ressources dans ledit dispositif de traitement d'informations, et sont attribués sur la base de ladite performance requise, et

ladite performance requise est déterminée sur la base d'une performance requise pour mémoriser des données envoyées par ledit premier serveur virtuel à au moins un premier groupe RAID parmi lesdits groupes RAID dudit premier groupe de ressources.

10. Système de mémorisation selon la revendication 1 comportant en outre :

un dispositif de gestion qui affiche des informations desdits ports, desdits processeurs, desdites ressources de mémoire et desdits groupes RAID, et qui est conçu pour recevoir au moins une demande pour créer les-

5 dits groupes de ressources,

 dans lequel ladite au moins une demande dispose d'informations de ladite performance requise.

 11. Système de mémorisation selon la revendication 1 comportant en outre :

10 un dispositif de gestion qui affiche des informations desdites ressources dans un premier groupe de ressources parmi lesdits groupes de ressources.

 12. Système de mémorisation selon la revendication 1, dans lequel :

15 un premier groupe de ressources parmi lesdits groupes de ressources est créé conformément à une demande de création dudit premier groupe de ressources.

 13. Système de mémorisation (200), comportant :
20 au moins un port conçu pour recevoir des données en provenance d'au moins un dispositif de traitement d'informations,

 au moins un processeur (212) qui commande le transfert de données reçues par ledit au moins un port,

 au moins une ressource de mémoire (213, 214)
25 qui mémorise des données sur la base de la commande par ledit au moins un processeur,

 au moins un groupe RAID (Réseau Redondant de Disques Indépendants), qui est associé à une pluralité d'unités de disque (215), lesdites unités de disque mémorisant une pluralité de données et des données de parité associées aux données mémorisées dans lesdites unités de
30 disque de l'un dudit au moins un groupe RAID, et

 une pluralité de groupes de ressources (221, 222, 223), chaque groupe parmi ladite pluralité de groupes de ressources étant divisé par au moins une partition
35

logique, chaque groupe de ressources ayant au moins l'un dudit au moins un port, au moins l'un dudit au moins un processeur, au moins l'une de ladite au moins une ressource de mémoire, et au moins l'un dudit au moins un

5 groupe RAID,

dans lequel une quantité de ressources dans chacun desdits groupes de ressources est attribuée sur la base d'une performance requise pour chacun desdits groupes de ressources.

10 14. Système de mémorisation (200), comportant :
une pluralité de ports conçus pour recevoir des données en provenance d'au moins un dispositif de traitement d'informations,

15 une pluralité d'unités logiques (231, 241) établies pour accéder à des données par l'intermédiaire dudit dispositif de traitement d'informations et associées à une pluralité de régions de mémorisation,

20 une pluralité de processeurs (212) qui commandent le transfert de données reçues par lesdits ports vers lesdites régions de mémorisation,

une pluralité de ressources de mémoire (213, 214) qui mémorisent des données sur la base de la commande par lesdits processeurs,

25 une pluralité d'unités de disque (215) ayant lesdites régions de mémorisation et mémorisant des données mémorisées dans lesdites ressources de mémoire, et

30 une pluralité de groupes de ressources (221, 222, 223), chacun ayant au moins l'un desdits ports, au moins l'une desdites unités logiques, au moins l'un desdits processeurs, et au moins l'une desdites ressources de mémoire, et chacun desdits groupes de ressources étant partitionné par au moins une partition logique,

dans lequel un rapport de partitionnement de chaque ressource dans chacun desdits groupes de ressource-

ces partitionnés est déterminé sur la base d'une performance requise pour chacun desdits groupes de ressources.

15. Système de mémorisation selon la revendication 14, dans lequel

5 le nombre desdites unités logiques dans un premier groupe de ressources desdits groupes de ressources est attribué sur la base de ladite performance requise.

16. Système de mémorisation (200) comportant :

10 au moins un port conçu pour recevoir des données en provenance d'au moins un dispositif de traitement d'informations,

15 au moins une unité logique (231, 241) établie pour accéder à des données par l'intermédiaire dudit au moins un dispositif de traitement d'informations et associée à au moins une région de mémorisation,

au moins un processeur qui commande le transfert de données reçues par ledit au moins un port vers ladite au moins une région de mémorisation,

20 au moins une ressource de mémoire (213, 214) qui mémorise des données sur la base de la commande par ledit au moins un processeur,

25 une pluralité d'unités de disque (215) ayant ladite au moins une région de mémorisation et mémorisant des données mémorisées dans ladite au moins une ressource de mémoire, et

une pluralité de groupes de ressources (221, 222, 223) chacun ayant une pluralité de types de ressources parmi ledit au moins un port, ladite au moins une unité logique, ledit au moins un processeur, et ladite au moins une ressource de mémoire, et chaque groupe de ressources étant partitionné par au moins une partition logique,

30 dans lequel un rapport de partitionnement de chaque ressource dans chacun desdits groupes de ressour-

ces partitionnés est déterminé sur la base d'une performance requise pour chacun desdits groupes de ressources.

17. Système de mémorisation comportant :

5 une pluralité de ports conçus pour recevoir des données en provenance d'au moins un dispositif de traitement d'informations,

10 une pluralité d'unités logiques (231, 241) établies pour accéder à des données par l'intermédiaire du dit dispositif de traitement d'informations et associées à une pluralité de régions de mémorisation,

une pluralité de processeurs qui commandent un transfert de données reçues par lesdits ports vers lesdites régions de mémorisation,

15 une pluralité de ressources de mémoire (213, 214) qui mémorisent des données sur la base de la commande par lesdits processeurs,

une pluralité d'unités de disque (215) ayant lesdites régions de mémorisation et mémorisant des données mémorisées dans lesdites ressources de mémoire, et

20 une pluralité de groupes de ressources (221, 222, 223) chacun ayant une pluralité de types de ressources parmi au moins l'un desdits ports, au moins l'une desdites unités logiques, au moins l'un desdits processeurs, et au moins l'une desdites ressources de mémoire, et chacun desdits groupes de ressources étant partitionné

25 par au moins une partition logique,

dans lequel une quantité desdites ressources dans chacun desdits groupes de ressources est attribuée sur la base d'une performance requise pour chacun desdits

30 groupes de ressources.

18. Système de mémorisation (200) comportant :

au moins un port conçu pour recevoir des données en provenance d'au moins un dispositif de traitement d'informations,

au moins une unité logique (231, 241) établie pour accéder à des données par l'intermédiaire dudit au moins un dispositif de traitement d'informations et associée à au moins une région de mémorisation,

5 au moins un processeur qui commande le transfert de données reçues par ledit au moins un port vers ladite au moins une région de de mémorisation,

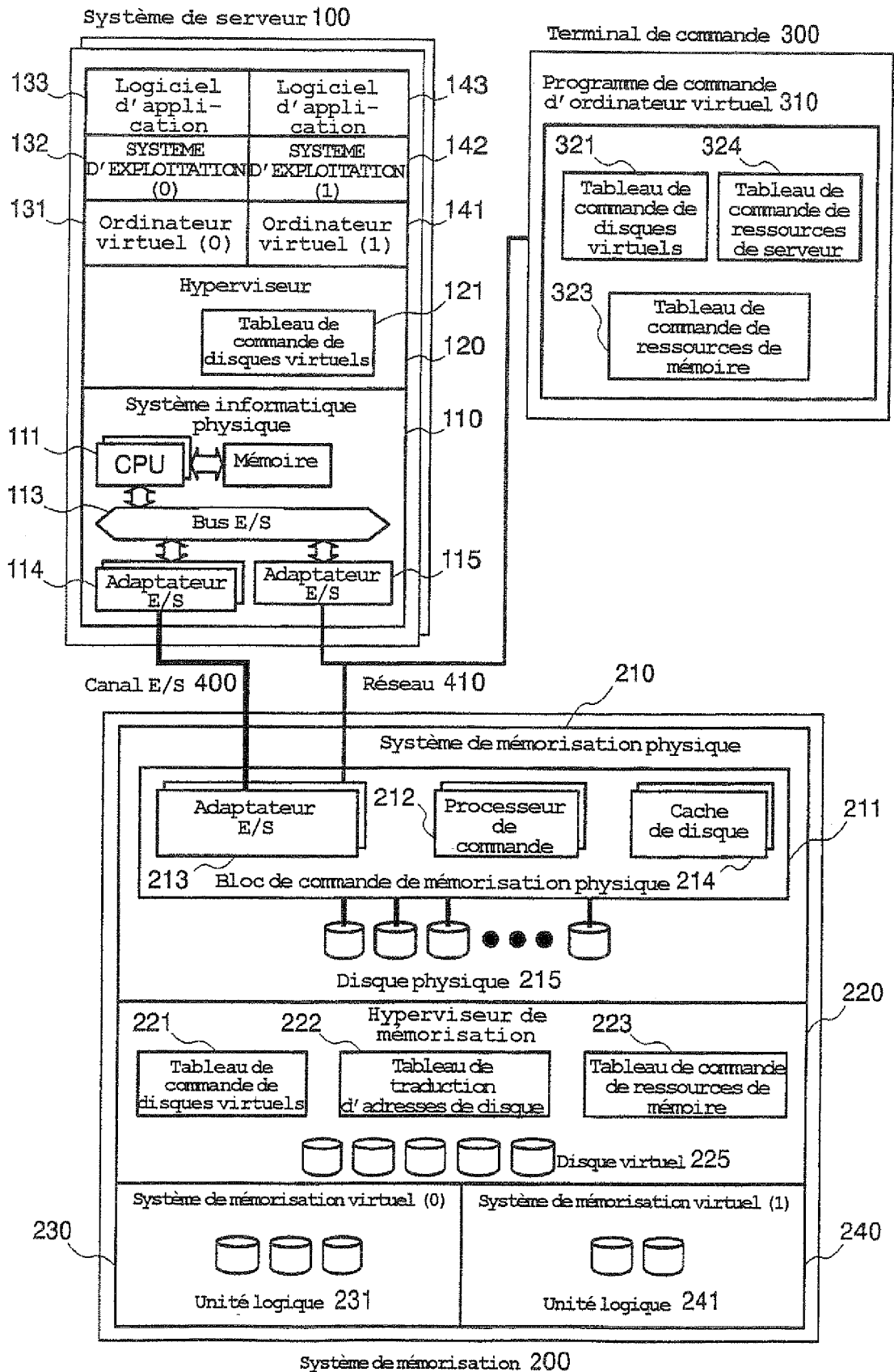
 au moins une ressource de mémoire (213, 214) qui mémorise des données sur la base de la commande par
10 ledit au moins un processeur,

 une pluralité d'unités de disque (215) ayant ladite au moins une région de mémorisation et mémorisant des données mémorisées dans ladite au moins une ressource de mémoire, et

15 une pluralité de groupes de ressources (221, 222, 223) chacun ayant au moins l'un dudit au moins un port, au moins l'une de ladite au moins une unité logique, au moins l'un dudit au moins un processeur, et au moins l'une de ladite au moins une ressource de mémoire,
20 et chacun desdits groupes de ressources étant partitionné par au moins une partition logique,

 dans lequel une quantité de ressources dans chacun desdits groupes de ressources est attribuée sur la base d'une performance requise pour chacun desdits groupes de ressources.
25

FIGURE 1



2/18

FIGURE 2

Numéro d'ordinateur virtuel	Numéro d'unité logique	Numéro de disque virtuel	Numéro de système de mémorisation virtuel
0	0	121	0
	1	122	
	2	123	
1	3	16	1
	4	17	
	5	18	
• • •	• • •	• • •	

Tableau de commande de disques virtuels

221, 121, 321

3/18

FIGURE 3

501 Numéro de disque virtuel	502 Adresse de bloc logique	503 Numéro de disque physique	504 Adresse de bloc physique
121	0x00000000	8	0x00000000
	0x80000000	9	0x00000000
122	0x00000000	10	0x10000000
• • •	• • •	• • •	• • •

Tableau de traduction d' adresses de disque

222

4/18

FIGURE 4

Numéro d'ordinateur virtuel	Numéro de disque virtuel	Capacité de cache de disque	Processeur de commande	Adaptateur E/S
0	121	512 MO	48	0
	122		49	1
	123		50	2
1	16	256 MO	112	3
	17		113	4
	18			
• • •	• • •	• • •	• • •	• • •

Tableau de commande de ressources de mémoire

323, 223, 123

5/18

FIGURE 5

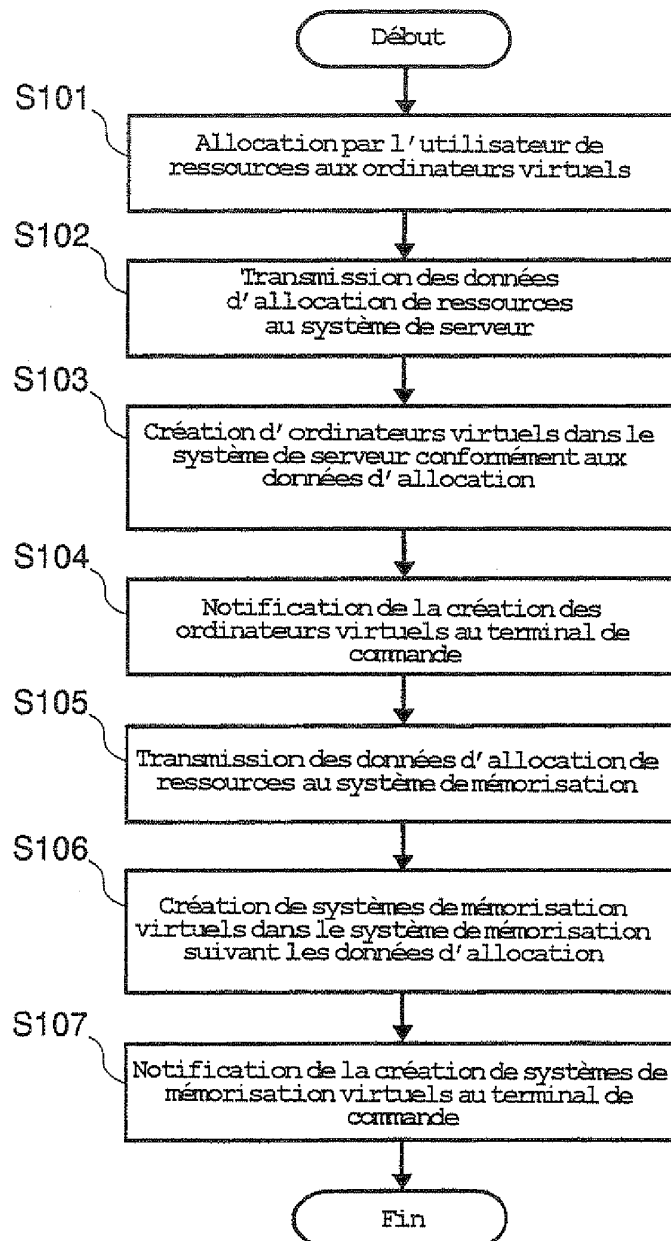
701 Numéro d'ordinateur virtuel	702 Allocation de CPU	703 Capacité de mémoire	704 Adaptateur E/S
0	20%	512MO	0 1 2
1	30%	128 MO	3 4
• • •	• • •	• • •	• • •

Tableau de commande de ressources de serveur

324, 124, 224

6/18

FIGURE 6



7/18

FIGURE 7

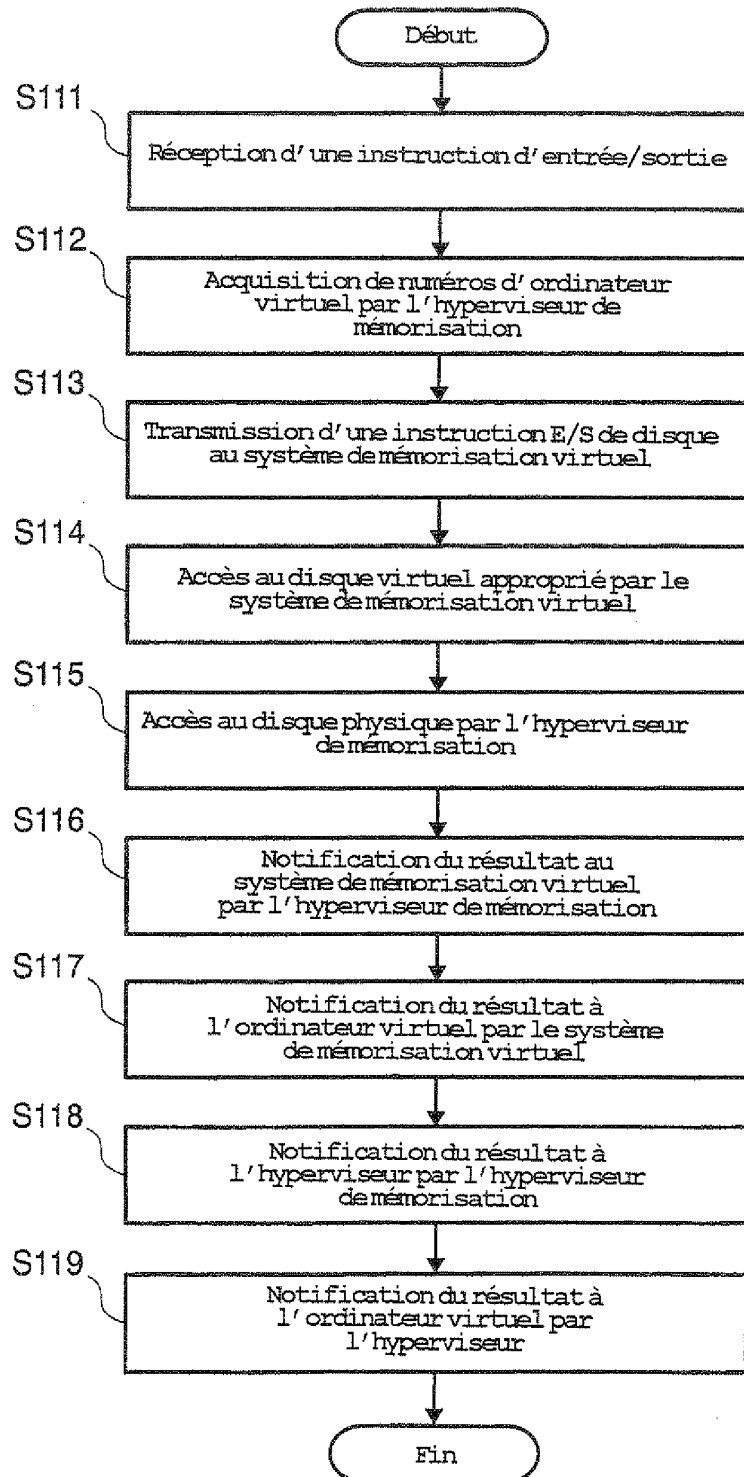
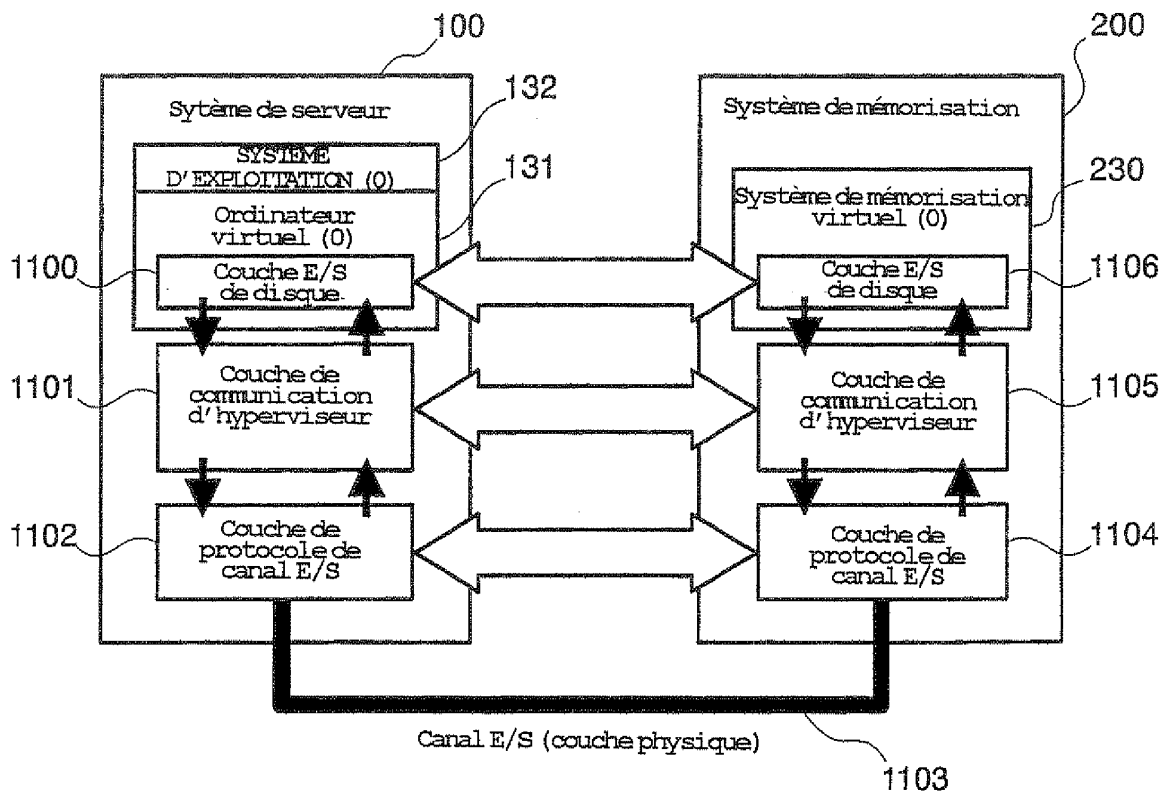
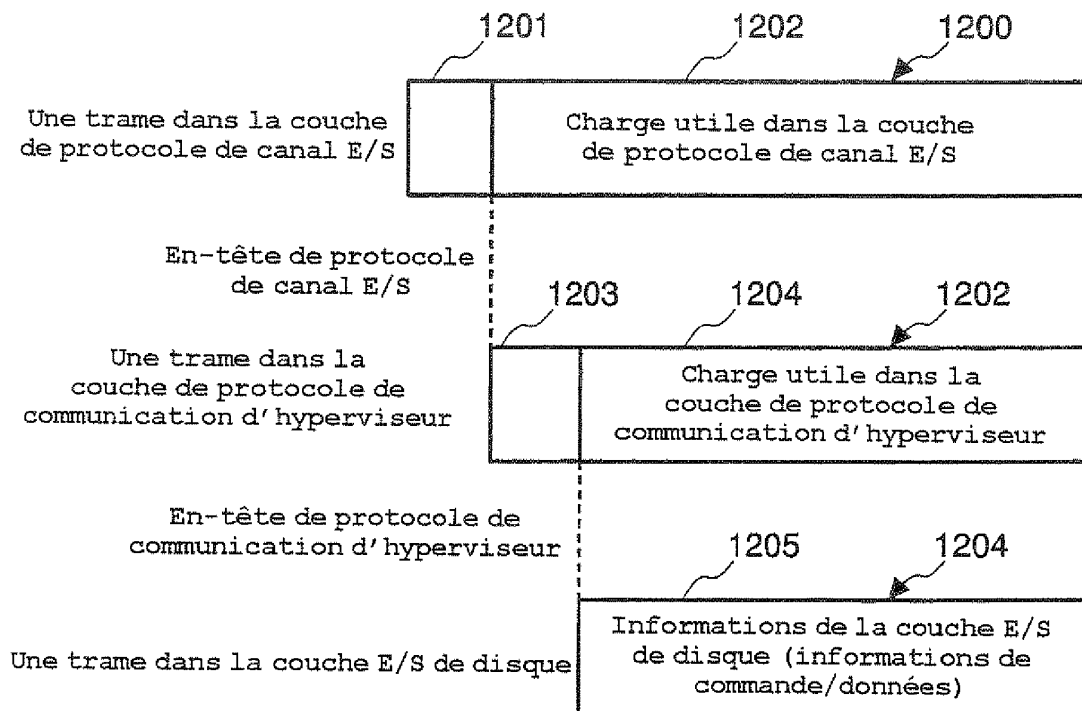


FIGURE 8



9/18

FIGURE 9



10/18

FIGURE 10

Numéro d'hyperviseur source	Numéro d'hyperviseur de destination
Numéro d'ordinateur virtuel source	Numéro d'ordinateur virtuel de destination

1300

1203

1301

1302

1303

The diagram shows a 2x2 table with the following content:

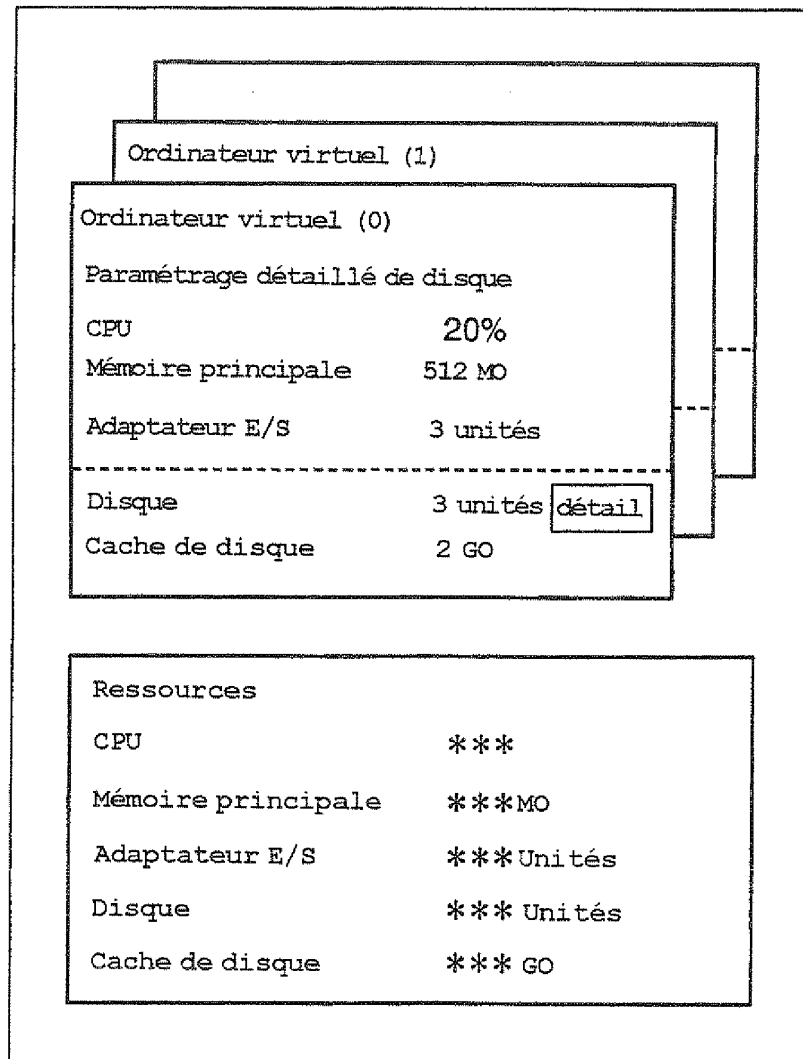
- Top-left cell: Numéro d'hyperviseur source
- Top-right cell: Numéro d'hyperviseur de destination
- Bottom-left cell: Numéro d'ordinateur virtuel source
- Bottom-right cell: Numéro d'ordinateur virtuel de destination

Labels and their positions:

- 1300: Points to the top-left corner of the table.
- 1203: Points to the top-right corner of the table.
- 1301: Points to the top-right corner of the table.
- 1302: Points to the bottom-left corner of the table.
- 1303: Points to the bottom-right corner of the table.

11/18

FIGURE 11



12/18

FIGURE 12

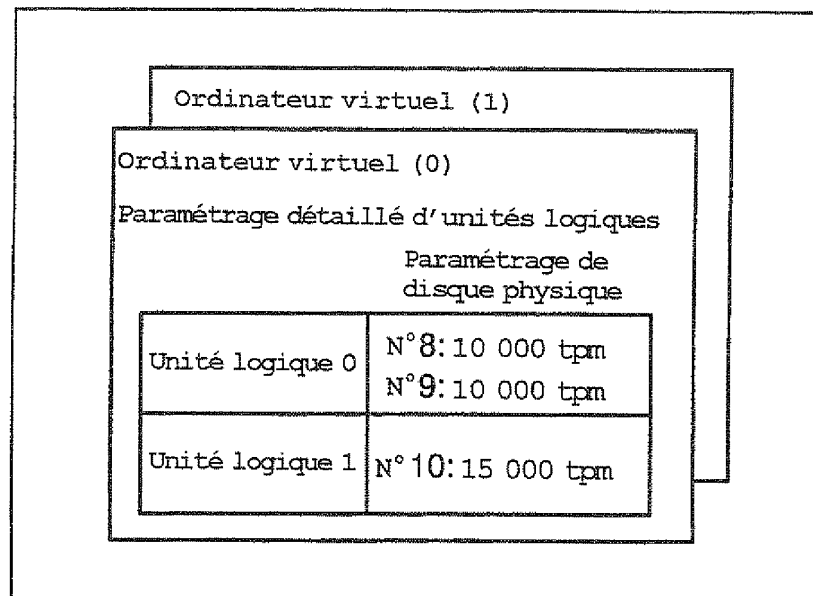


FIGURE 13

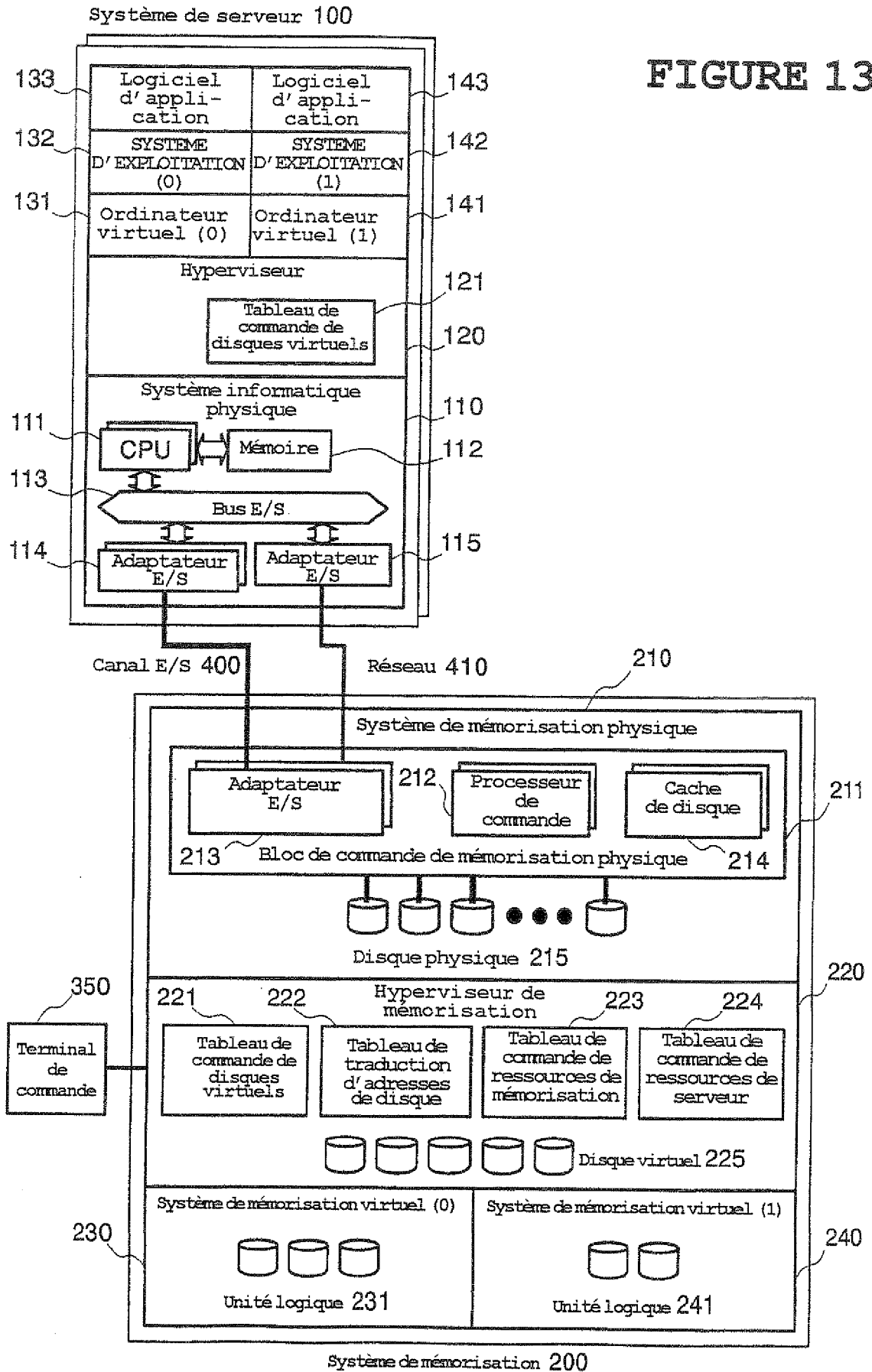


FIGURE 14

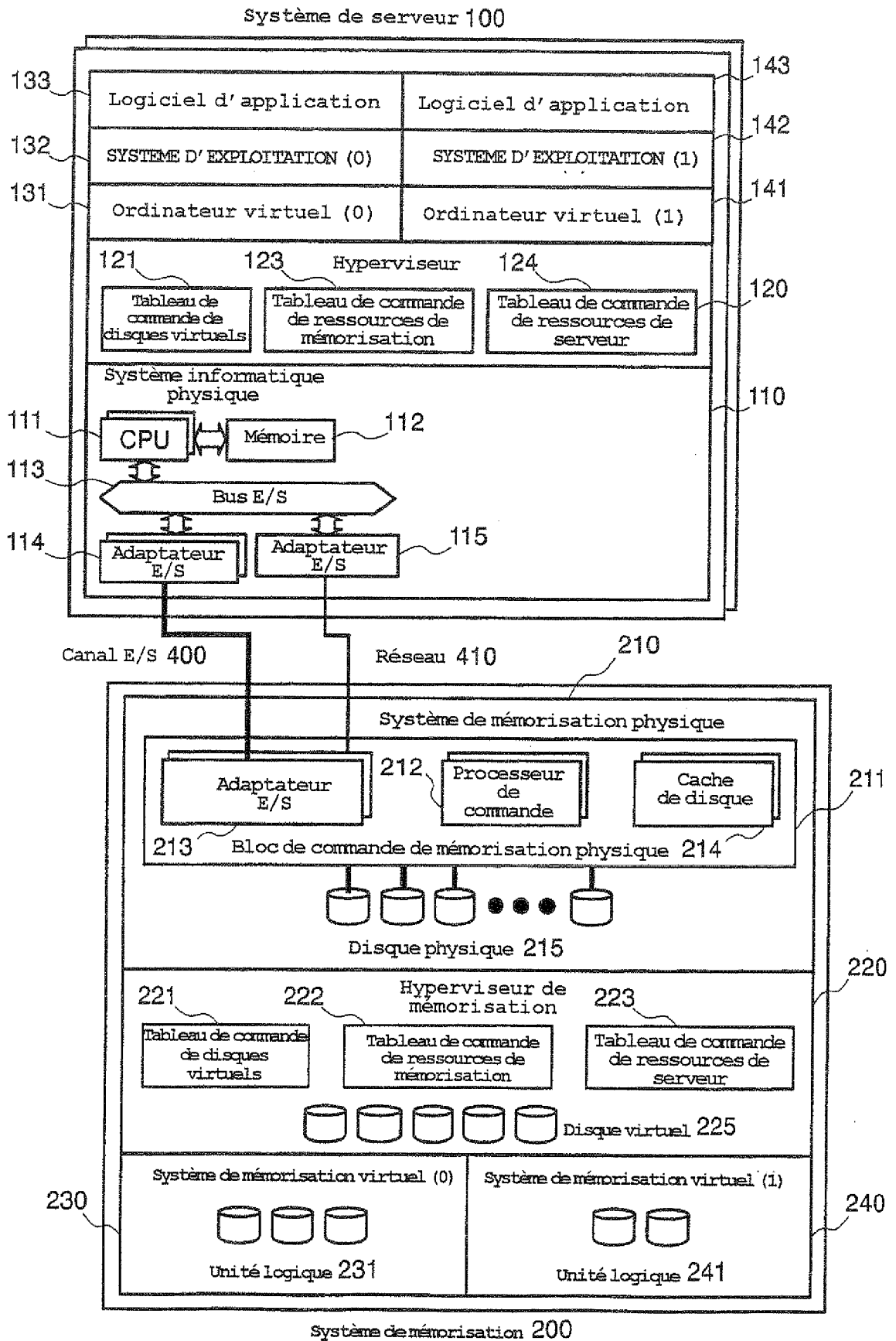
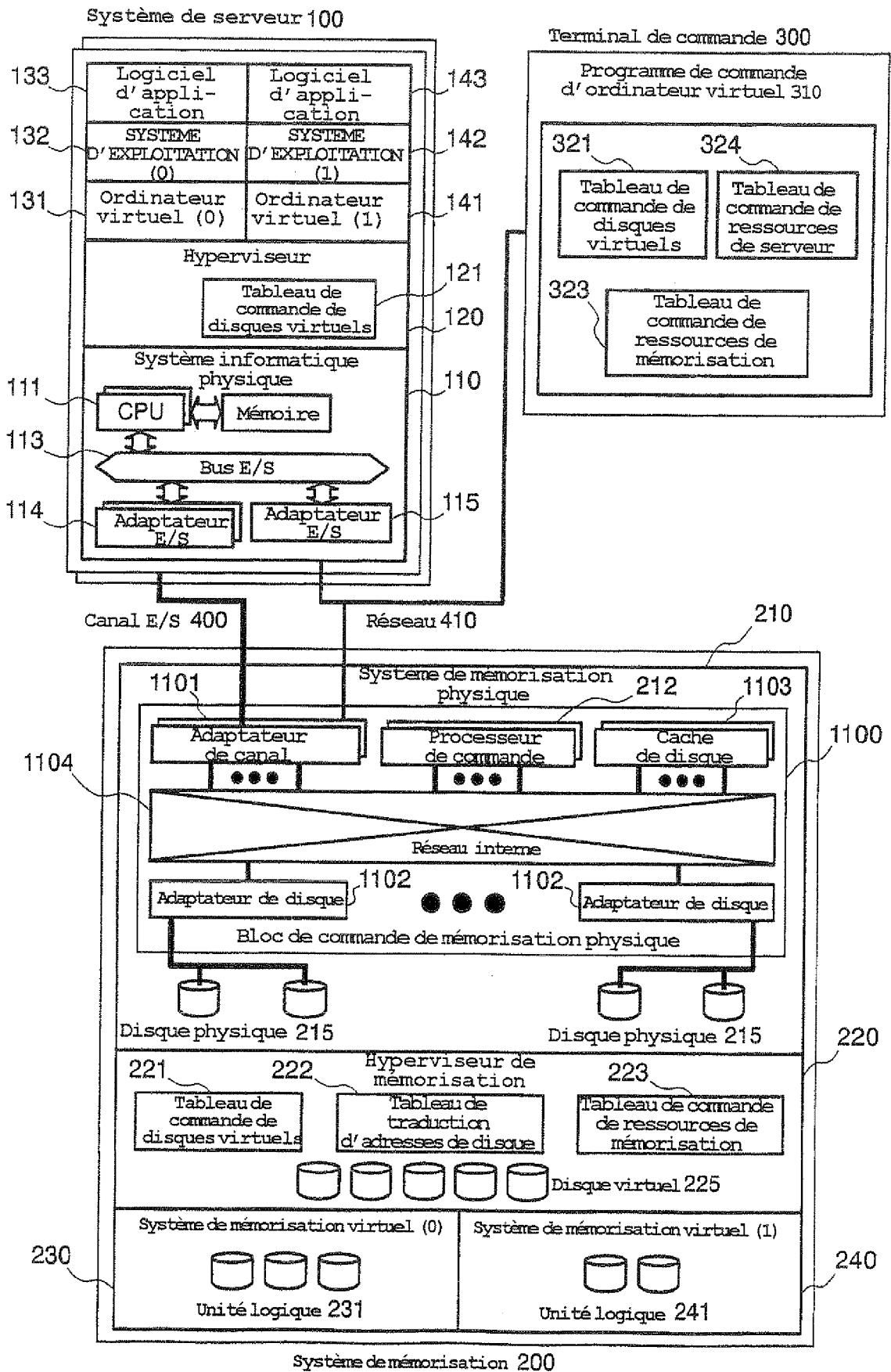


FIGURE 15



16/18

FIGURE 16

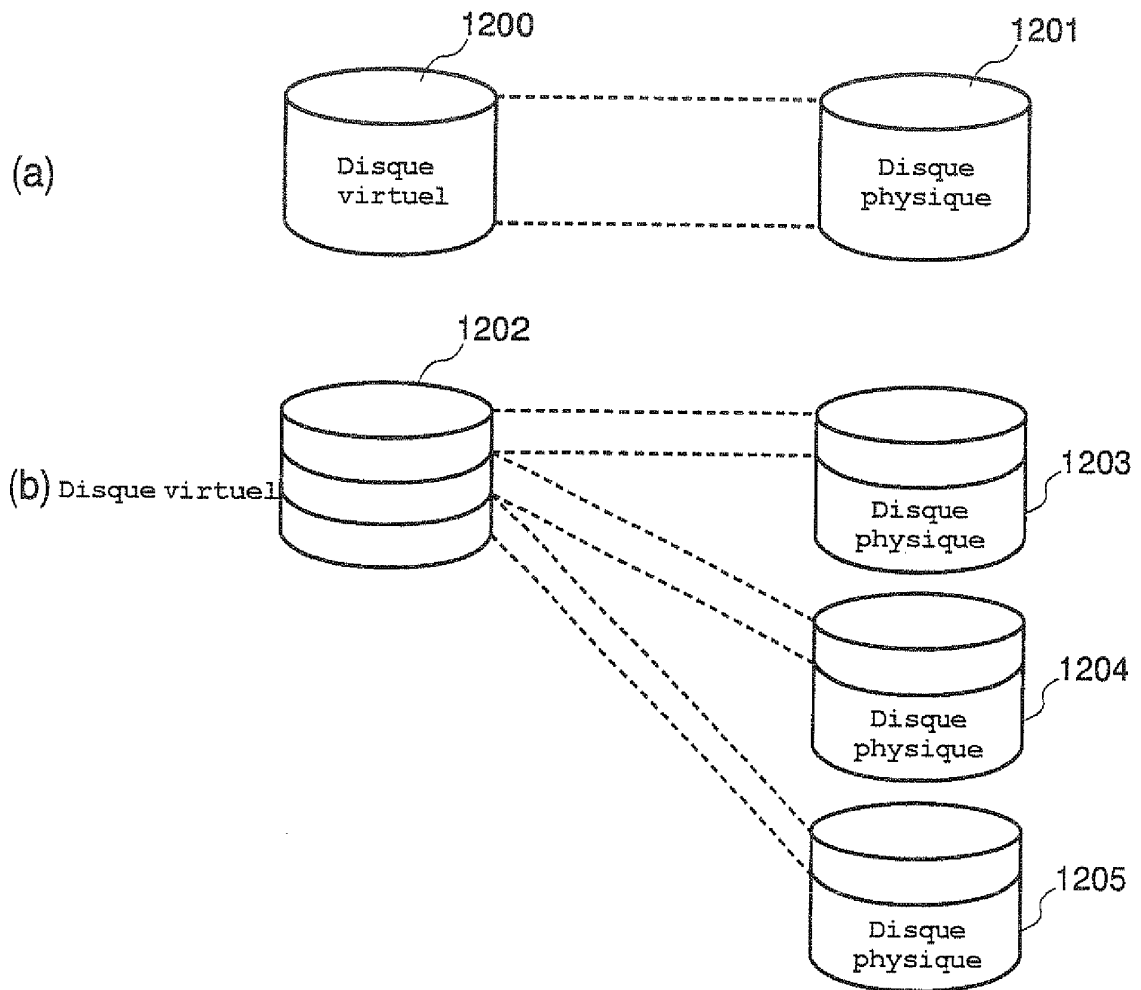


FIGURE 17

601 Numéro d'ordinateur virtuel	602 Numéro de disque virtuel	603 Capacité de cache de disque	604 Processeur de commande	605 Adaptateur de canal	1300 Largeur de bande de réseau interne
0	121	512 MO	48	0	20%
	122		49	1	
	123		50	2	
1	16	256 MO	112	3	10%
	17		113	4	
	18				
•	•	•	•	•	
•	•	•	•	•	
•	•	•	•	•	

323, 223
Tableau de commande de ressources de rétro-allocation

18/18

FIGURE 18

