

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7148033号  
(P7148033)

(45)発行日 令和4年10月5日(2022.10.5)

(24)登録日 令和4年9月27日(2022.9.27)

(51)国際特許分類	F I
H 0 4 L 67/1004(2022.01)	H 0 4 L 67/1004
H 0 4 L 67/1095(2022.01)	H 0 4 L 67/1095
H 0 4 L 67/1097(2022.01)	H 0 4 L 67/1097

請求項の数 24 (全32頁)

(21)出願番号	特願2020-567564(P2020-567564)	(73)特許権者	508140877 レベル スリー コミュニケーションズ, エルエルシー アメリカ合衆国,コロラド州 8 0 0 2 1,ブルームフィールド,エルドラド ブルバード 1 0 2 5
(86)(22)出願日	平成30年6月22日(2018.6.22)	(74)代理人	110000877弁理士法人R Y U K A国際 特許事務所
(65)公表番号	特表2021-526268(P2021-526268 A)	(72)発明者	ニュートン、クリストファー アメリカ合衆国,コロラド州 8 0 0 2 1,ブルームフィールド,エルドラド ブルバード 1 0 2 5 レベル スリー コミュニケーションズ,エルエルシー内
(43)公表日	令和3年9月30日(2021.9.30)	審査官	羽岡 さやか
(86)国際出願番号	PCT/US2018/039032		
(87)国際公開番号	WO2019/236113		
(87)国際公開日	令和1年12月12日(2019.12.12)		
審査請求日	令和3年1月21日(2021.1.21)		
(31)優先権主張番号	16/002,919		
(32)優先日	平成30年6月7日(2018.6.7)		
(33)優先権主張国・地域又は機関	米国(US)		

最終頁に続く

(54)【発明の名称】 スーパークラスタにわたる負荷分散

(57)【特許請求の範囲】

【請求項1】

第1の機械と、第2の機械と、第3の機械とを有する複数の機械と、  
前記複数の機械の各々がサブスクリブするように構成可能であるパブリッシュ - サブ  
スクリブ ( P u b - S u b ) チャンネルと

を備え、

前記複数の機械の各々は、

可用性ステータスおよびリソース割り当て情報を前記 P u b - S u b チャンネルにパブリ  
ッシュすること、

前記 P u b - S u b チャンネルを介して、前記複数の機械のうちの少なくとも別の1つの  
可用性ステータスおよびリソース割り当て情報を受信することであって、前記可用性ステ  
ータスは、コンテンツデータを格納するために前記複数の機械の各々が利用可能であるか  
どうかを示し、前記リソース割り当て情報は、前記複数の機械の各々により格納される前  
記コンテンツデータを特定する、受信すること、および、

前記 P u b - S u b チャンネル上の前記可用性ステータスおよび前記リソース割り当て情  
報に基づいて、クライアントから受信した、コンテンツデータの要求を前記複数の機械の  
うちの前記少なくとも別の1つへ転送することを行うように構成され、

前記第1の機械は、前記 P u b - S u b チャンネル上の前記可用性ステータスおよび前記  
リソース割り当て情報に基づいて、かつ、前記第1の機械が前記コンテンツデータを所有  
するのを前記第2の機械が許可したことに基づいて、前記第2の機械により所有されてい

10

20

る前記コンテンツデータを所有するように構成され、前記第1の機械は、前記コンテンツデータの中継された要求を受信したことに基づいて、前記コンテンツデータを前記第3の機械に提供するように構成される、

スーパークラスタ。

【請求項2】

前記Pub - Subチャンネルは、マルチキャストアドレスを有する、請求項1に記載のスーパークラスタ。

【請求項3】

前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいて、前記クライアントから受信した、前記コンテンツデータの前記要求を前記複数の機械のうちの前記少なくとも別の1つへ転送することは、

10

前記複数の機械のうちの前記第1の機械が前記コンテンツデータの前記要求を前記クライアントから受信することと、

前記第1の機械が、前記第1の機械は前記コンテンツデータを格納していない、と判断することと、

前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいて、前記コンテンツデータは前記複数の機械のうちの前記第2の機械に格納されている、と判断することと、

前記要求を前記第2の機械へ転送することと

を含む、

20

請求項1または2に記載のスーパークラスタ。

【請求項4】

前記第1の機械は、前記第2の機械により現在所有されている前記コンテンツデータを所有するために、

前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいて、前記第2の機械により所有されている前記コンテンツデータを特定し、

前記第2の機械により所有されている前記コンテンツデータを所有するための可用性を前記Pub - Subチャンネルにパブリッシュし、かつ、

前記第2の機械により所有されている前記コンテンツデータを所有する

ように構成される、

30

請求項1から3のいずれか一項に記載のスーパークラスタ。

【請求項5】

前記第1の機械は、前記第2の機械により現在所有されている前記コンテンツデータを所有するために、

前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいて、前記コンテンツデータを現在格納している第2の機械の数は閾値未満である、と判断し、

前記コンテンツデータを所有するための可用性を前記Pub - Subチャンネルにパブリッシュし、

前記Pub - Subチャンネルを介して、前記コンテンツデータを現在格納している前記第2の機械から許可を受信し、

40

前記コンテンツデータを所有し、

前記コンテンツデータの中継された要求を前記複数の機械のうちの前記第3の機械から受信し、かつ、

前記中継された要求に回答して、前記コンテンツデータを提供する

ように構成される、

請求項1から4のいずれか一項に記載のスーパークラスタ。

【請求項6】

前記複数の機械のうちの前記第1の機械は、別のスーパークラスタ内の第2の機械へのブリッジを確立するように構成され、

50

前記別のスーパークラスタは、前記スーパークラスタに近接し、

前記ブリッジは、前記別のスーパークラスタの少なくともいくつかの機械の各々の可用性ステータスおよびリソース割り当て情報を前記第1の機械へ中継するように構成される、請求項1から5のいずれか一項に記載のスーパークラスタ。

【請求項7】

前記第1の機械は、前記 Pub - Sub チャンネル上の前記複数の機械の前記可用性ステータスに基づいて、前記複数の機械のうちの利用不可能な機械の数は閾値を超えている、と判断するように構成され、

前記スーパークラスタは、前記利用不可能な機械の数は前記閾値を超えているという判断に回答して、前記ブリッジを確立するように構成される、

請求項6に記載のスーパークラスタ。

【請求項8】

前記ブリッジは、伝送制御プロトコル (TCP) 接続を含む、請求項6または7に記載のスーパークラスタ。

【請求項9】

前記第1の機械は、前記スーパークラスタに対する地理的な近接度に基づいて、前記別のスーパークラスタを特定するように構成される、請求項6から8のいずれか一項に記載のスーパークラスタ。

【請求項10】

前記第1の機械は、ドメインネームシステム (DNS) 要求を行うことにより、前記別のスーパークラスタが前記スーパークラスタに隣接していることを特定するように構成される、請求項6から9のいずれか一項に記載のスーパークラスタ。

【請求項11】

前記第1の機械は、前記 Pub - Sub チャンネル上の前記別のスーパークラスタの前記少なくともいくつかの機械の前記可用性ステータスおよび前記リソース割り当て情報をパブリッシュするように構成される、請求項6から10のいずれか一項に記載のスーパークラスタ。

【請求項12】

前記スーパークラスタの第3の機械は、

前記コンテンツデータの前記要求を前記クライアントから受信し、

前記 Pub - Sub チャンネル上の前記別のスーパークラスタの前記少なくともいくつかの機械の前記リソース割り当て情報に基づいて、前記コンテンツデータは前記別のスーパークラスタの第4の機械により格納されている、と判断し、かつ、

前記要求を前記第4の機械へ転送する

ように構成される、

請求項11に記載のスーパークラスタ。

【請求項13】

スーパークラスタの複数の機械にわたって負荷を分散するための方法であって、前記複数の機械は、第1の機械と、第2の機械と、第3の機械とを含み、前記方法は、

前記スーパークラスタの前記複数の機械の各々が、可用性ステータスおよびリソース割り当て情報をパブリッシュ - サブスクライブ (Pub - Sub) チャンネルにパブリッシュする段階であって、前記複数の機械の各々は、前記 Pub - Sub チャンネルをサブスクライブする、パブリッシュする段階と、

前記 Pub - Sub チャンネルを介して、前記複数の機械のうちの少なくとも別の1つの可用性ステータスおよびリソース割り当て情報を受信する段階であって、前記可用性ステータスは、コンテンツデータを格納するために前記複数の機械の各々が利用可能であるかどうかを示し、前記リソース割り当て情報は、前記複数の機械の各々により格納される前記コンテンツデータを特定する、段階と、

前記 Pub - Sub チャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を前記複数の機械の

10

20

30

40

50

うちの前記少なくとも別の1つへ転送する段階であって、前記第2の機械により所有されている前記コンテンツデータの前記第1の機械による所有は、前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいており、かつ、前記第1の機械が前記コンテンツデータを所有するのを前記第2の機械が許可したことに基づいており、前記第1の機械は、前記コンテンツデータの中継された要求を受信したことに基づいて、前記コンテンツデータを前記第3の機械に提供する、転送する段階と

を備える方法。

【請求項14】

実行された場合、スーパークラスタの複数の機械であって、第1の機械と、第2の機械と、第3の機械とを含む、複数の機械の各々のプロセッサに、

可用性ステータスおよびリソース割り当て情報をパブリッシュ - サブスクライブ (Pub - Sub) チャンネルにパブリッシュする手順であって、前記複数の機械の各々は、前記Pub - Subチャンネルをサブスクライブする、パブリッシュする手順と、

前記Pub - Subチャンネルを介して、前記複数の機械のうちの少なくとも別の1つの可用性ステータスおよびリソース割り当て情報を受信する手順であって、前記可用性ステータスは、コンテンツデータを格納するために前記複数の機械の各々が利用可能であるかどうかを示し、前記リソース割り当て情報は、前記複数の機械の各々により格納される前記コンテンツデータを特定する、手順と、

前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を前記複数の機械のうちの前記少なくとも別の1つへ転送する手順であって、前記第2の機械により所有されている前記コンテンツデータの前記第1の機械による所有は、前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいており、かつ、前記第1の機械が前記コンテンツデータを所有するのを前記第2の機械が許可したことに基づいており、前記第1の機械は、前記コンテンツデータの中継された要求を受信したことに基づいて、前記コンテンツデータを前記第3の機械に提供する、転送する手順と

を実行させるようなコンピュータ可読命令を備えるコンピュータプログラム。

【請求項15】

スーパークラスタにわたって負荷を分散するための方法であって、

第1のスーパークラスタの第1の機械が、前記第1のスーパークラスタ内の利用不可能な機械の数は閾値を超えている、と判断する段階と、

前記第1の機械が第2のスーパークラスタ内の第2の機械とのブリッジを確立する段階であって、前記第2のスーパークラスタは、前記第1のスーパークラスタに隣接する、確立する段階と、

前記第1の機械が、前記ブリッジを介して、前記第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報を受信する段階であって、前記可用性ステータスは、コンテンツデータを格納するために前記第2のスーパークラスタの前記複数の機械の各々が利用可能であるかどうかを示し、前記リソース割り当て情報は、前記第2のスーパークラスタの前記複数の機械の各々が担う前記コンテンツデータを特定する、段階と、

前記第1の機械が、前記第1のスーパークラスタのパブリッシュ - サブスクライブ (Pub - Sub) チャンネル上で、前記第2のスーパークラスタの前記複数の機械の各々の前記可用性ステータスおよび前記リソース割り当て情報をパブリッシュする段階であって、前記第2のスーパークラスタの前記第2の機械により所有されているコンテンツデータの前記第1の機械による所有は、前記Pub - Subチャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいており、かつ、前記第1の機械が前記コンテンツデータを所有するのを前記第2の機械が許可したことに基づいており、前記第1の機械は、前記コンテンツデータの中継された要求を受信したことに基づいて、前記コンテンツデータを前記第1のスーパークラスタの第3の機械に提供する、パブリッシュする段階と

を備える方法。

10

20

30

40

50

## 【請求項 16】

前記第1のスーパークラスタの前記 Pub - Sub チャンネル上の前記第1のスーパークラスタの複数の機械が前記第1のスーパークラスタの前記機械の各々の可用性ステータスおよびリソース割り当て情報をパブリッシュする段階をさらに備える、請求項 15 に記載の方法。

## 【請求項 17】

前記第1のスーパークラスタ内の前記利用不可能な機械の数は、前記第1のスーパークラスタの前記 Pub - Sub チャンネル上で提供される前記可用性ステータスに基づいて決定される、請求項 15 または 16 に記載の方法。

## 【請求項 18】

前記第1のスーパークラスタの前記 Pub - Sub チャンネルは、マルチキャストアドレスを含む、請求項 15 から 17 のいずれか一項に記載の方法。

## 【請求項 19】

前記第3の機械がコンテンツデータの要求をクライアントから受信する段階と、  
前記第3の機械が、前記 Pub - Sub チャンネル上の前記第2のスーパークラスタの前記複数の機械の各々の前記可用性ステータスおよび前記リソース割り当て情報に基づいて、前記コンテンツデータは前記第2のスーパークラスタの第4の機械により格納されている、と判断する段階と、

前記要求を前記第2のスーパークラスタ内の前記第4の機械へ転送する段階と  
をさらに備える、請求項 15 から 18 のいずれか一項に記載の方法。

## 【請求項 20】

前記ブリッジは、伝送制御プロトコル (TCP) 接続を含む、請求項 15 から 19 のいずれか一項に記載の方法。

## 【請求項 21】

前記第1の機械が、前記第1のスーパークラスタに対する地理的な近接度に基づいて、前記第2のスーパークラスタを特定する段階をさらに備える、請求項 15 から 20 のいずれか一項に記載の方法。

## 【請求項 22】

前記第1の機械は、ドメインネームシステム (DNS) 要求を行うことにより、前記第2のスーパークラスタが前記第1のスーパークラスタに隣接していることを特定する、請求項 15 から 21 のいずれか一項に記載の方法。

## 【請求項 23】

実行された場合、第1のスーパークラスタの第1の機械のプロセッサに、  
前記第1のスーパークラスタ内の利用不可能な機械の数は閾値を超えている、と判断する手順と、

第2のスーパークラスタ内の第2の機械とのブリッジを確立する段階であって、前記第2のスーパークラスタは、前記第1のスーパークラスタに隣接する、確立する手順と、

前記ブリッジを介して、前記第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報を受信する手順であって、前記可用性ステータスは、コンテンツデータを格納するために前記第2のスーパークラスタの前記複数の機械の各々が利用可能であるかどうかを示し、前記リソース割り当て情報は、前記第2のスーパークラスタの前記複数の機械の各々が担う前記コンテンツデータを特定する、手順と、

前記第1のスーパークラスタのパブリッシュ - サブスクライブ (Pub - Sub) チャンネル上で、前記第2のスーパークラスタの前記複数の機械の各々の前記可用性ステータスおよび前記リソース割り当て情報をパブリッシュする手順であって、前記第2の機械により所有されているコンテンツデータの前記第1の機械による所有は、前記 Pub - Sub チャンネル上の前記可用性ステータスおよび前記リソース割り当て情報に基づいており、かつ、前記第1の機械が前記コンテンツデータを所有するのを前記第2の機械が許可したことに基づいており、前記第1の機械は、前記コンテンツデータの中継された要求を受信したことに基づいて、前記コンテンツデータを第3の機械に提供する、パブリッシュする手

10

20

30

40

50

順と

を実行させるようなコンピュータ可読命令を備えるコンピュータプログラム。

【請求項 24】

複数の第 1 の機械と、

前記複数の第 1 の機械の各々がサブスクライブするように構成可能であるパブリッシュ  
- サブスクライブ ( P u b - S u b ) チャンネルと

を有する第 1 のスーパークラスタと、

複数の第 2 の機械

を有する第 2 のスーパークラスタと、

前記第 1 のスーパークラスタと前記第 2 のスーパークラスタとの間のブリッジと

を備え、

前記複数の第 1 の機械のうちの 1 つは、

前記ブリッジを介して、前記複数の第 2 の機械の各々の可用性ステータスおよびリソ  
ス割り当て情報を前記複数の第 2 の機械のうちの 1 つから受信することによって、前記可  
用性ステータスは、コンテンツデータを格納するために前記第 2 のスーパークラスタの前  
記複数の第 2 の機械の各々が利用可能であるかどうかを示し、前記リソース割り当て情報  
は、前記第 2 のスーパークラスタの前記複数の第 2 の機械の各々が担う前記コンテン  
ツデータを特定する、受信すること、および、

前記第 1 のスーパークラスタの前記 P u b - S u b チャンネル上で、前記複数の第 2 の機  
械の各々の前記可用性ステータスおよび前記リソース割り当て情報をパブリッシュするこ  
とを行うように構成され、

前記複数の第 1 の機械のうちの前記 1 つは、前記 P u b - S u b チャンネル上の前記可  
用性ステータスおよび前記リソース割り当て情報に基づいて、かつ、前記複数の第 1 の機  
械のうちの前記 1 つがコンテンツデータを所有するのを前記複数の第 2 の機械のうちの  
前記 1 つが許可したことに基づいて、前記複数の第 2 の機械のうちの前記 1 つにより  
所有されている前記コンテンツデータを所有するように構成され、前記複数の第 1 の機  
械のうちの前記 1 つは、前記コンテンツデータの中継された要求を受信したことに基  
づいて、前記コンテンツデータを前記第 1 のスーパークラスタの第 3 の機械に提供する、

システム。

【発明の詳細な説明】

【技術分野】

【0001】

[著作権表示]

本特許文献は、著作権保護の対象となる資料を含む。著作権者は、本特許文献または米  
国特許商標庁のファイルにおけるあらゆる関連資料の複製に対して異議を唱えるものでは  
ないが、そうでなければ、いかなる理由があるうとも全ての著作権を留保する。

[関連出願の相互参照]

【0002】

本特許協働条約 ( P C T ) 特許出願は、2018年6月7日出願された、「スーパーク  
ラスタにわたる負荷分散」と題する米国出願第 16 / 002,919 号の優先権を主張す  
る。当該米国出願は、参照により、全ての内容が本明細書に組み込まれる。

【0003】

本技術は概して、コンテンツ配信ネットワーク ( C D N ) などのネットワーク内のス  
ーパークラスタ内および当該スーパークラスタにわたるコンテンツの分散に関する。

【背景技術】

【0004】

データセンタは、データ格納およびデータ処理が可能な機械を各々が含み得る 1 または  
複数のスーパークラスタを含み得る。典型的には、各スーパークラスタは、データセン  
タ内の多数のラック ( または物理クラスタ ) に及ぶ。いくつかの実装において、1 または複  
数のスーパークラスタがコンテンツ配信ネットワーク ( C D N ) 内で用いられ得る。

10

20

30

40

50

## 【 0 0 0 5 】

C D Nは、サーバ（各々がC D Nノードと称される）の地理的に分散されたネットワークであり、オリジンサーバがオリジンサーバのコンテンツデータを、当該コンテンツデータを消費するクライアントへ配信するのを容易にするためのものである。オリジンサーバから遠く離れているクライアントへ配信するために、クライアントに地理的に近接するC D Nノードは、オリジンサーバに代わって、コンテンツデータをそれらのクライアントに提供できる。特に、C D Nノードは、オリジンサーバのコンテンツデータを複製およびキャッシュし、複製およびキャッシュしたコンテンツデータをクライアントに提供できる。いくつかのC D Nノードは、それらのC D Nノードがクライアントに近接していることから、「エッジノード」と称される。

10

## 【 0 0 0 6 】

スーパークラスタは、C D Nの1または複数のノードを実装するために用いられ得る。例えば、ノードは、スーパークラスタの機械により実装され得る。スーパークラスタは、C D Nのエッジノードを実装するために用いられ得る。それに関連して、スーパークラスタ内の各機械は、コンテンツデータの総ライブラリの少なくともいくつかの部分（例えば、少なくとも1つのスライス）を担うキャッシュである。総ライブラリは、1または複数のスーパークラスタに結び付けられるプロパティのセットにより定義される。そのため、総ライブラリを構成するスライスは、1または複数のスーパークラスタ内の機械に割り当てられる。

## 【 0 0 0 7 】

スーパークラスタ内の機械は、予め定められたノスケジューリングされた非可用性（例えば、メンテナンス等）および予期しないノ突然の非可用性（例えば、非応答性、処理遅延、電源遮断等）を含む様々な理由で利用不可能になり得る。所与のスライスを格納する一次機械が利用不可能である場合、当該スライスを格納する二次機械が、当該スライスをクライアントに提供するように選択され得る。所与のリソースノ機械が割り当てられるスライスを選択するために、一貫したハッシュアルゴリズムが用いられる。この一貫したハッシュアルゴリズムは、機械のリストにわたって反復する。リスト上での機械の順序は、ソースキー（C D N内の要求されるユニフォームリソースロケータ（U R L）の正規化バージョンなど）に基づいて決定され得る。本方式では、機械が利用不可能であると判断されたことに応答して、利用不可能な機械が担うコンテンツデータは、スーパークラスタ内の多数の（理想的には全ての）他の機械に拡散され得るので、ホットスポットが回避される。ホットスポットは、スーパークラスタの所与の機械により、当該機械がかなりの量のトラフィックおよび処理を負担するように格納および提供されている1または複数の一般的なスライスを指す。所与のスーパークラスタ内の十分な数の機械がオフラインになった場合、残りの機械は、リソースが利用不可能であることに起因して、追加の負荷に圧倒されてしまい得る。従来のデータセンタが機械ステータスと可用性情報との伝達に関連して提供する機械間または機械グループ間の通信は、不十分である。

20

## 【 発明の概要 】

## 【 0 0 0 8 】

複数の機械と、複数の機械の各々がサブスクライブするパブリッシュ - サブスクライブ（P u b - S u b）チャンネルとを含むスーパークラスタの実施形態が提供される。複数の機械の各々は、可用性ステータスおよびリソース割り当て情報をP u b - S u bチャンネルにパブリッシュし、P u b - S u bチャンネルを介して、複数の機械のうちの少なくとも別の1つの可用性ステータスおよびリソース割り当て情報を受信し、かつ、P u b - S u bチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を複数の機械のうちの別の1つへ転送するように構成される。

40

## 【 0 0 0 9 】

可用性ステータスは、コンテンツデータを格納するために複数の機械の各々が利用可能であるかどうかを示す。リソース割り当て情報は、複数の機械の各々により格納されてい

50

るコンテンツデータを特定する。Pub - Subチャンネルは、好ましくは、マルチキャストアドレスを含む。

【0010】

一態様によれば、Pub - Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を複数の機械のうちの別の1つへ転送することは、複数の機械のうちの第1の機械が、コンテンツデータの要求をクライアントから受信することと、第1の機械が、第1の機械はコンテンツデータを格納していない、と判断することと、Pub - Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、コンテンツデータは複数の機械のうちの第2の機械に格納されている、と判断することと、要求を第2の機械へ転送することを含む。

10

【0011】

複数の機械のうちの第1の機械は、Pub - Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、複数の機械のうちの第2の機械により所有されているコンテンツデータを所有し得る。

【0012】

第2の機械により現在所有されているコンテンツデータを所有するために、第1の機械は、Pub - Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、第2の機械により所有されているコンテンツデータを特定し、第2の機械により所有されているコンテンツデータを所有するための可用性をPub - Subチャンネルにパブリッシュし、Pub - Subチャンネルを介して、第2の機械からの許可を第2の機械から受信し、第2の機械により所有されているコンテンツデータを所有し、コンテンツデータの中継された要求を複数の機械のうちの第3の機械から受信し、かつ、中継された要求に回答して、コンテンツデータを提供し得る。

20

【0013】

第2の機械により現在所有されているコンテンツデータを所有するために、第1の機械は、Pub - Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、コンテンツデータを現在格納している第2の機械の数は閾値未満である、と判断し、コンテンツデータを所有するための可用性をPub - Subチャンネルにパブリッシュし、Pub - Subチャンネルを介して、コンテンツデータを現在格納している第2の機械から許可を受信し、コンテンツデータを所有し、コンテンツデータの中継された要求を複数の機械のうちの第3の機械から受信し、かつ、中継された要求に回答して、コンテンツデータを提供し得る。

30

【0014】

複数の機械のうちの第1の機械は、別のスーパークラスタ内の第2の機械へのブリッジを確立するように構成され得る。別のスーパークラスタは、好ましくは、スーパークラスタに近接しており、地理的に最も近いスーパークラスタであり得る。ブリッジは、別のスーパークラスタの少なくともいくつかの機械の各々の可用性ステータスおよびリソース割り当て情報を第1の機械へ中継するように構成される。

【0015】

第1の機械は、Pub - Subチャンネル上の複数の機械の可用性ステータスに基づいて、複数の機械のうちの利用不可能な機械の数は閾値を超えている、と判断し得る。ブリッジは、利用不可能な機械の数は閾値を超えているという判断に回答して確立され得る。ブリッジは、伝送制御プロトコル(TCP)接続を含み得る。

40

【0016】

第1の機械は、スーパークラスタに対する地理的な近接度に基づいて、他のスーパークラスタを特定し得る。第1の機械は、ドメインネームシステム(DNS)要求を行うことにより、他のスーパークラスタが当該スーパークラスタに隣接していることを特定し得る。第1の機械は、Pub - Subチャンネル上の他のスーパークラスタの少なくともいくつかの機械の可用性ステータスおよびリソース割り当て情報をパブリッシュし得る。

【0017】

50



スーパークラスタの第3の機械は、コンテンツデータの要求をクライアントから受信し、Pub-Subチャンネル上の別のスーパークラスタの少なくともいくつかの機械のリソース割り当て情報に基づいて、コンテンツデータは別のスーパークラスタの第4の機械により格納されている、と判断し、かつ、要求を第4の機械へ転送し得る。

**【0018】**

一態様によれば、スーパークラスタの複数の機械にわたって負荷を分散するための方法は、スーパークラスタの複数の機械の各々が可用性ステータスおよびリソース割り当て情報をPub-Subチャンネルにパブリッシュする段階であって、複数の機械の各々は、Pub-Subチャンネルをサブスクライブする、パブリッシュする段階と、Pub-Subチャンネルを介して、複数の機械のうちの少なくとも別の1つの可用性ステータスおよびリソース割り当て情報を受信する段階と、Pub-Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を複数の機械のうちの別の1つへ転送する段階とを備える。

10

**【0019】**

別の態様によれば、非一時的コンピュータ可読媒体は、実行された場合、スーパークラスタの複数の機械の各々のプロセッサに、可用性ステータスおよびリソース割り当て情報をPub-Subチャンネルにパブリッシュする手順であって、複数の機械の各々は、Pub-Subチャンネルをサブスクライブする、パブリッシュする手順と、Pub-Subチャンネルを介して、複数の機械のうちの少なくとも別の1つの可用性ステータスおよびリソース割り当て情報を受信する手順と、Pub-Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を複数の機械のうちの別の1つへ転送する手順とを実行させるようなコンピュータ可読命令を備える。

20

**【0020】**

別の態様によれば、スーパークラスタにわたって負荷を分散するための方法は、第1のスーパークラスタの第1の機械が、第1のスーパークラスタ内の利用不可能な機械の数は閾値を超えている、と判断する段階と、第1の機械が第2のスーパークラスタ内の第2の機械とのブリッジを確立する段階であって、第2のスーパークラスタは、第1のスーパークラスタに隣接する、確立する段階と、第1の機械が、ブリッジを介して、第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報を受信する段階と、第1の機械が、第1のスーパークラスタのPub-Subチャンネル上で、第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報をパブリッシュする段階とを備える。

30

**【0021】**

可用性ステータスは、コンテンツデータを格納するために第2のスーパークラスタの複数の機械の各々が利用可能であるかどうかを示し得る。リソース割り当て情報は、第2のスーパークラスタの複数の機械の各々が担うコンテンツデータを特定し得る。

**【0022】**

方法は、第1のスーパークラスタのPub-Subチャンネル上の第1のスーパークラスタの複数の機械が第1のスーパークラスタの機械の各々の可用性ステータスおよびリソース割り当て情報をパブリッシュする段階をさらに許可する。第1のスーパークラスタ内の利用不可能な機械の数は、第1のスーパークラスタのPub-Subチャンネル上で提供される可用性ステータスに基づいて決定され得る。第1のスーパークラスタのPub-Subチャンネルは、マルチキャストアドレスを含み得る。

40

**【0023】**

方法は、第1のスーパークラスタの第3の機械がコンテンツデータの要求をクライアントから受信する段階と、第3の機械が、Pub-Subチャンネル上の第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報に基づいて、コンテンツデータは第2のスーパークラスタの第4の機械により格納されている、と判断する段階と、要求を第2のスーパークラスタ内の第4の機械へ転送する段階とをさらに備え

50

得る。このブリッジは、TCP接続を含み得る。方法は、第1の機械が、第1のスーパークラスタに対する地理的な近接度に基づいて、第2のスーパークラスタを特定する段階を備えることをさらに許可する。第1の機械は、DNS要求を行うことにより、第2のスーパークラスタが第1のスーパークラスタに隣接していることを特定し得る。

【0024】

別の態様によれば、非一時的コンピュータ可読媒体は、実行された場合、第1のスーパークラスタの第1の機械のプロセッサに、第1のスーパークラスタ内の利用不可能な機械の数は閾値を超えている、と判断する手順と、第2のスーパークラスタ内の第2の機械とのブリッジを確立する段階であって、第2のスーパークラスタは、第1のスーパークラスタに隣接する、確立する手順と、ブリッジを介して、第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報を受信する手順と、第1のスーパークラスタのPub-Subチャンネル上で、第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報をパブリッシュする手順とを実行させるようなコンピュータ可読命令を備える。

10

【0025】

別の態様によれば、システムは、複数の第1の機械を有する第1のスーパークラスタと、複数の第1の機械の各々がサブスクリブするPub-Subチャンネルと、複数の第2の機械を有する第2のスーパークラスタと、第1のスーパークラスタと第2のスーパークラスタとの間のブリッジとを備える。複数の第1の機械のうちの1つは、ブリッジを介して、複数の第2の機械の各々の可用性ステータスおよびリソース割り当て情報を複数の第2の機械のうちの1つから受信し、かつ、第1のスーパークラスタのPub-Subチャンネル上で、複数の第2の機械の各々の可用性ステータスおよびリソース割り当て情報をパブリッシュするように構成される。

20

【0026】

これらの特徴および他の特徴ならびにそれらのオペレーションの編成および方式は、添付図面と共に解釈された場合、以下の詳細な説明から明らかになるであろう。

【図面の簡単な説明】

【0027】

【図1】本開示による、スーパークラスタ内および複数のスーパークラスタ間での通信を提供するためのシステムの例を示す図である。

30

【0028】

【図2】スーパークラスタの機械の例を示すブロック図である。

【0029】

【図3】スーパークラスタ内での通信を提供するための例示的な方法を示すフロー図である。

【0030】

【図4】スーパークラスタ内での通信を提供するための例示的な方法を示す別のフロー図である。

【0031】

【図5】スーパークラスタ内での通信を提供するための例示的な方法を示すさらに別のフロー図である。

40

【0032】

【図6】複数のスーパークラスタ間での通信を提供するための例示的な方法を示すフロー図である。

【発明を実施するための形態】

【0033】

本明細書において説明する構成は、所与のスーパークラスタの機械にわたる、または2つまたはそれよりも多くのスーパークラスタの機械にわたるコンテンツ（例えば、コンテンツデータの総ライブラリ）の共有またはそうでなければ分散に関連してハイブウェアネス（hive-awareness）を提供するためのシステム、装置、方法および非

50

一時的コンピュータ可読媒体に関する。各機械は、コンテンツの一部分（例えば、総ライブラリのスライスまたはスロット）を格納し得る。各機械は、同じスーパークラスタ内の他の機械および/または別のスーパークラスタ内の他の機械と通信して、機械により格納された一部分を特定し得る。

#### 【0034】

データセンタは、2つまたはそれよりも多くのスーパークラスタを含み得る。本明細書において説明する構成により、スーパークラスタ内の機械は、リソースの割り当ておよび提供に関し、説明する方式で互いに通信できるようになる。さらなる構成により、2つのスーパークラスタの機械は、リソースの割り当ておよび提供に関し、互いに通信できるようになる。そのため、所与のスーパークラスタ内のかなりの数の機械がオフラインになったことに応答して1つから別のものへロールオーバーする当該スーパークラスタの機械ではなく、代わりに、当該スーパークラスタの機械により処理されるいくつかのスライスが、少なくとも1つの他のスーパークラスタ内の機械により提供されるサービスを受け得る。この少なくとも1つの他のスーパークラスタは、隣接するスーパークラスタであり得る。いくつかの例において、所与のデータセンタが、単一のスーパークラスタを含み得る。その場合、最も近いスーパークラスタが、別のデータセンタにおいてホスティングされ得る。

10

#### 【0035】

したがって、本明細書において説明する構成は、（例えば、CDNの文脈において）コンテンツをクライアントに提供し（第1のスーパークラスタ内の機械（ノード）間でリソース割り当てデータを伝達することを含む）、1または複数の適切な近接するスーパークラスタ（本明細書において、第2のスーパークラスタと称される）を決定し、かつ、第1のスーパークラスタ内の失敗または劣化したノードの数は閾値を超えているという判断に応答して第1のスーパークラスタおよび第2のスーパークラスタを統合するデータセンタに関連する様々な技術的問題に対処する。

20

#### 【0036】

従来、総ライブラリの各スライスにDNS名が提供されている。各機械は、DNS名に対応するスライスの格納を担う。DNS解決は、要求されたスライスを格納する特定の機械へとクライアントを誘導できる。多数のスライスを有し得る大きい総ライブラリの場合、そのような従来の実装が管理不能になってしまう。本明細書において説明する構成は、総ライブラリ内の各スロットに割り当てられるDNS名を必要としない。代わりに、スーパークラスタ内の機械は、可用性ステータスおよびリソース割り当てに関して互いに通信して、要求されたスライスにどの機械がサービスを提供できるかを判断し得る。同じスーパークラスタ内の機械が、要求されたスライスを提供できない場合、当該要求は、スーパークラスタ間で共有される可用性ステータスまたはリソース割り当て情報に基づいて、別のスーパークラスタ内の機械へ転送され得る。

30

#### 【0037】

図1は、本開示のいくつかの構成によるスーパークラスタ内および複数のスーパークラスタ間での通信を提供するための例示的なシステム100の図である。図1を参照すると、CDNが、システム100を用いて実装され得る。CDNは、（各々がCDNノードである）サーバまたは機械の地理的に分散されたネットワークであり、オリジンサーバ（不図示）がオリジンサーバのコンテンツデータを、当該コンテンツデータを消費するクライアント（例えば、クライアント105）へ配信するのを容易にするためのものである。オリジンサーバから遠く離れているクライアントへオリジンサーバからコンテンツデータを配信するために、クライアントに地理的に近接するCDNノードが、オリジンサーバに代わって、コンテンツデータをクライアントに提供できる。特に、当該CDNノードは、オリジンサーバのコンテンツデータ（またはその少なくとも一部分）を複製およびキャッシュし、複製およびキャッシュしたコンテンツデータをクライアントへ提供できる。CDNの顧客は、オリジンサーバに関連し得る。

40

#### 【0038】

50

システム 100 は少なくとも、第 1 のスーパークラスタ 110 および第 2 のスーパークラスタ 140 を含む。各スーパークラスタは、複数の機械を含み得る。例えば、第 1 のスーパークラスタ 110 は、機械 114 a - 114 e と機械 116 - 126 とを含む。第 2 のスーパークラスタ 140 は、機械 142 - 150 を含む。同じスーパークラスタの機械は、同じ地理的位置にあり、および/または同じ物理位置に配置される。例えば、第 1 のスーパークラスタ 110 の機械 114 a - 114 e と機械 116 - 126 とは、第 1 の位置にあるが、第 2 のスーパークラスタ 140 の機械 142 - 150 は、第 1 の位置とは異なる第 2 の位置にある。他の例において、第 1 の位置および第 2 の位置は、同じ位置であり得る（例えば、スーパークラスタ 110 および 140 は、ない交ぜになったラックがあるかまたはない同じデータセンタに配置される）。第 1 のスーパークラスタ 110 および第 2 のスーパークラスタ 140 は、CDN の文脈において、互いに近接または隣接する。したがって、第 1 のスーパークラスタ 110 の機械（例えば、機械 120）と第 2 のスーパークラスタ 140 の機械（例えば、機械 142）との間の通信は、同じ CDN の第 1 のスーパークラスタ 110 の機械と別のスーパークラスタの機械との間の通信よりも速い。第 1 のスーパークラスタ 110 および第 2 のスーパークラスタ 140 は、同じデータセンタに属する。代替的に、第 1 のスーパークラスタ 110 および第 2 のスーパークラスタ 140 は、異なるデータセンタに属する。

#### 【0039】

さらに、各スーパークラスタ内の機械は、機械のラックであり得る。例えば、第 1 のスーパークラスタ 110 は、機械 114 a - 114 e のラック 112 を含む。機械 114 a - 114 e は、バーチャル IP アドレスを介して共に境界が定められる。機械 114 a - 114 e は、同じコンテンツデータを格納し得る。代替的に、機械 114 a - 114 e のうちの少なくとも 1 つは、機械 114 a - 114 e のうちの他のものにより格納されるコンテンツデータとは異なるコンテンツデータを格納する。ラック 112 は、機械 114 a - 114 e のスタック内の負荷均衡機能を提供し得る。例えば、クライアントの要求にตอบสนองして、ラック 112 内のスイッチが、機械 114 a - 114 e のうちの 1 つを選択することで、クライアントの要求に対してサービスを提供できる。図 1 における第 1 のスーパークラスタ 110 について 1 つのラック 112 のみが示されているが、第 1 のスーパークラスタ 110 および第 2 のスーパークラスタ 140 は、任意の数の機械ラックを有し得る。第 1 のスーパークラスタ 110 および/または第 2 のスーパークラスタ 140 内のラック構成の各々は、ラック 112 と同様の方式で構造化され得る。機械 116 - 126 と機械 142 - 150 との各々は、好ましくは、物理 IP アドレスと個別にアドレス指定される。

#### 【0040】

各スーパークラスタは、それらと関連する機械の結び付けにより定義される。一例において、ホスト名と機械との間のマッピングが、スーパークラスタの定義またはそうでなければ形成のために用いられ得る。CDN の文脈において、ホスト名および機械は、CDN の少なくとも 1 つの顧客の同じセットにサービスを提供するように構成される。各顧客は、オリジンサーバに関連し得る。機械 114 a - 114 e と機械 116 - 126 とは、（ホスト名と機械 114 a - 114 e および機械 116 - 126 との間の適切なマッピングを通じて）共に結び付けられ、第 1 のスーパークラスタ 110 を形成する。同様に、機械 142 - 150 は、（ホスト名と機械 142 - 150 との間の適切なマッピングを通じて）共に結び付けられ、第 2 のスーパークラスタ 140 を形成する。加えて、スーパークラスタが異なる位置に配置され得ると仮定すると、スーパークラスタは、CDN が必要とする任意の地理的制限（例えば、特定のコンテンツが特定の地理的位置においてサービスを受けることができるかまたはできない）および/または地理ベースセキュリティプロトコルを実装するために用いられ得る。

#### 【0041】

クライアント（例えば、クライアント 105）は、CDN のノードをサポートする機械のうちの 1 つに接続できるエンドユーザである。CDN の各ノードは、機械 114 a - 1

10

20

30

40

50

14eと、機械116-126と、機械142-150とのうちの1つに対応する。CDNは、不図示の追加のスーパークラスタおよび機械を用いて実装され得る。第1のスーパークラスタ110の機械114a-114eと機械116-126とは、クライアント105を含む近くのクライアントにサービスを提供する、CDNのエッジノードである。同様に、機械142-150も、第2のスーパークラスタ140に地理的に近接するクライアントにサービスを提供する、CDNのエッジノードであり得る。クライアント105は、限定されるものではないがDNSルックアップまたはドメイン名解決などのDNSオペレーションを用いてラック112または機械116-126のうちの1つに接続するよう誘導され得る。示されるように、クライアント105による要求に基づくDNSルックアップの結果により、クライアント105は、要求されたコンテンツデータを機械122から受信するよう誘導される。クライアント105がCDNの別のスーパークラスタの別の位置よりも第1のスーパークラスタ110の位置により近いと仮定すると、クライアント105による要求に基づくDNSルックアップは、CDNの第2のスーパークラスタ140または別のスーパークラスタ内の機械へ解決しない。当該要求は、少なくともURLを含む。

10

#### 【0042】

スーパークラスタ内の各機械は、関連するCDNの少なくとも1つの顧客のセット用のコンテンツの総ライブラリの一部分を格納する。例えば、第1のスーパークラスタ110内の機械114a-114eと機械116-126との各々は、総ライブラリの少なくとも1つのスライスを担うキャッシュである。第2のスーパークラスタ140内の機械142-150の各々は、総ライブラリの少なくとも1つのスライスを担うキャッシュである。総ライブラリは、1または複数のスーパークラスタに結び付けられるプロパティのセットにより定義される。ライブラリは概して、1または複数のスーパークラスタにより格納される、ライブラリのワーキングセットを指す。そのため、スーパークラスタ110または140の各々は、総ライブラリを格納し、総ライブラリは、(例えば、部分で、またはスライスで)スーパークラスタ110または140の各々における機械に割り当てられる。言い換えると、総ライブラリは、スーパークラスタ110および140の両方により格納される。総ライブラリの同じスライスは、冗長性のために、同じスーパークラスタの複数の機械により格納され得る。他の構成では、スーパークラスタの1または複数のノードまたは機械は、格納機能を有さないことがあるが、依然としてスーパークラスタの一部である。

20

30

#### 【0043】

ライブラリの「ワーキングセット」は、アクティブに用いられているのでキャッシュに充填済みであろうサブスライバの総ライブラリの一部分を指す。第1のスーパークラスタ110の機械114a-114eと機械116-126との各々は、2つの格納部を含み得る。第1の格納部は、総ライブラリの割り当てられた部分またはスライス(ワーキングセットのスライスと称される)を格納する。第2の格納部は、総ライブラリの一般的なコンテンツデータを格納する。第1のスーパークラスタ110に接続するクライアントにより総ライブラリの所与のスライスが頻繁に要求された場合、頻繁に要求されたスライスは、一般的なコンテンツデータになり、第1のスーパークラスタ110内の他の機械の第2の格納部に格納される。第1のスーパークラスタ110の機械114a-114eと機械116-126との各々は、要求を受信した機械が要求されたスライスを格納しているかどうかにかかわらず、総ライブラリの任意のスライスの要求(例えば、クライアント105からの要求)を受信できる。

40

#### 【0044】

一例において、第1のスーパークラスタ110内の機械122がオブジェクトの要求をクライアント105から受信したことに応答して、機械122は、当該スライスを格納している場合、オブジェクトをクライアントに提供する。本明細書において称される場合、「要求されたスライス」は、クライアント105により要求された、オブジェクトに関連するスライスを指す。要求されたスライスは、第1の格納部(ワーキングセット)または

50

第2の格納部（一般的なコンテンツ）のいずれかに格納され得る。一方、要求されたオブジェクトに関連する要求されたスライスが機械122が格納していない場合、機械122は、要求されたスライスを格納している、第1のスーパークラスタ110内の別の1つの機械から要求されたスライスを要求する。言い換えると、機械122は、要求を第1のスーパークラスタ110内の別の機械へ転送する。

【0045】

要求を第1のスーパークラスタ110内の別の機械へ転送すべく、機械114a-114eと機械116-126とが可用性ステータスおよびリソース割り当て情報を互いに通信することを可能にするために、第1のスーパークラスタ110の機械114a-114eと機械116-126とのための通信チャンネルが提供され得る。可用性ステータスは、コンテンツデータ（例えば、少なくとも1つのスライス）を格納および/または提供するために関連する機械が利用可能であるかどうかを示す。リソース割り当て情報により、関連する機械により格納されているコンテンツデータ（例えば、少なくとも1つのスライス）が特定される。

10

【0046】

説明したように、クライアント（例えば、クライアント105）からの要求は、DNSオペレーションに基づいて、第1のスーパークラスタ110内の機械114a-114eと機械116-126とのいずれかに解決され得る。第1のスーパークラスタ110が多数の機械を有する場合、要求されたスライスが、要求をクライアントから最初に受信した機械（例えば、機械122）により格納されていない可能性が高い。

20

【0047】

従来のスーパークラスタに伴う問題は、スーパークラスタの機械が様々な理由でオンラインになってはオフラインになることに関連している。例えば、スーパークラスタ内の機械は、限定されるものではないがメンテナンスなどの予め定められた/スケジュールリングされた非可用性が原因でオフラインになり得る。スーパークラスタ内の機械は、限定されるものではないが非応答性、処理遅延および電源遮断などの予期しない/突然の非可用性が原因でオフラインになり得る。

【0048】

要求されたスライスを格納していることが認識されている機械がオフラインである場合、要求を最初に受信する機械122は、第1のスーパークラスタ110内でチェックし、要求されたスライスを第1のスーパークラスタ110内の別の機械が提供できるかどうかを確かめる。例えば、機械122が、要求されたスライスを格納していることが認識されている機械116はオフラインである、と判断した場合、機械122は、第1のスーパークラスタ110内でチェックして、機械118は要求されたスライスも格納している、と判断する。

30

【0049】

それに関連して、機械114a-114eと機械116-126とは、Pub-Subチャンネル130をサブスクライブする。一例において、Pub-Subチャンネル130は、機械114a-114eと機械116-126とがサブスクライブできるデータストリームである。機械114a-114eと機械116-126との各々は、その独自の可用性ステータスおよびリソース割り当て情報をPub-Subチャンネル130上でパブリッシュできる。また、機械114a-114eと機械116-126との各々は、第1のスーパークラスタ110の他の機械の可用性ステータスおよびリソース割り当て情報をPub-Subチャンネル130から取得できる。そのため、機械114a-114eと機械116-126とのいずれも、要求されたスライスを提供するために利用可能である、第1のスーパークラスタ110内の機械をいつでも特定できる。Pub-Subチャンネル130の例がマルチキャストアドレスである。第2のスーパークラスタ140に関連するPub-Subチャンネル160は、Pub-Subチャンネル130と同じ方式で構成され得るデータストリームである。例えば、機械142-150の各々は、その独自の可用性ステータスおよびリソース割り当て情報をPub-Subチャンネル160上でパブリッシュで

40

50

きる。また、機械 1 4 2 - 1 5 0 の各々は、第 2 のスーパークラスタ 1 4 0 の他の機械の可用性ステータスおよびリソース割り当て情報を Pub - Sub チャンネル 1 6 0 から取得できる。そのため、Pub - Sub チャンネル 1 3 0 および Pub - Sub チャンネル 1 6 0 はそれぞれ、第 1 のスーパークラスタ 1 1 0 内および第 2 のスーパークラスタ 1 4 0 内のハイブアウェアネスを提供する。

#### 【 0 0 5 0 】

いくつかの例において、スーパークラスタの単一の機械が、所与のスライスを担当するキャッシュである。当該機械がオフラインである場合、スーパークラスタの第 2 の機械は、構成情報に基づいて決定される。同じ第 2 の機械が、当該スライスのスーパークラスタの全てのメンバにより選択される。いくつかの例において、スーパークラスタが、例えば、機械 1 1 6 がオフラインであることを検出したとき、第 1 のスーパークラスタ 1 1 0 の別の機械（例えば、機械 1 1 8）は、当該スliceを担当する（例えば、当該スライスの所有を引き継ぐ）キャッシュになる。機械 1 1 8 は、機械 1 1 8 が当該スライスを担当するキャッシュであることを Pub - Sub チャンネル 1 3 0 上でアナウンスし得る。その時まで、いくつかの例において、第 1 のスーパークラスタ 1 1 0 の各機械は、第 1 のスーパークラスタ 1 1 0 内でのスライスの再割り当てが完了するまで、または、要求されたオブジェクトのスライスを特定するために用いられるハッシュアルゴリズムが第 1 のスーパークラスタ 1 1 0 内の一連の機械を戻すまで、CDN 階層を上げて CDN の次のティアへ移行することで、当該スライスを取得し得る。例えば、要求されたオブジェクトに対応するスライスを特定する際、ハッシュアルゴリズムは、スライス番号（1、2、3...x）のリストを戻すことができる。リストの最初の要素（例えば、1）は、要求されたオブジェクトの好ましいスライスである。すなわち、スライス 1 は、機械 1 1 6 により現在「所有されて」いる。機械 1 1 6 はオフラインであるという判断に回答して、要求は、スライス 2 を「所有して」いる機械 1 1 8 へ送信される。

#### 【 0 0 5 1 】

第 1 のスーパークラスタ 1 1 0 内でオフラインの機械の数が閾値を超えている場合、第 1 のスーパークラスタ 1 1 0 内の機械の残りには、過負荷がかかり得る。したがって、第 1 のスーパークラスタ 1 1 0 内の機械は、近くのまたは近接するスーパークラスタ（例えば、第 2 のスーパークラスタ 1 4 0）内の機械と通信でき、その近くのまたは近接するスーパークラスタ内の機械へ要求を転送できる。スーパークラスタ 1 1 0 とスーパークラスタ 1 4 0 との間の通信は、ブリッジ 1 7 0 により実現され得る。ブリッジ 1 7 0 の例が TCP 接続である。

#### 【 0 0 5 2 】

ブリッジ 1 7 0 は、第 1 のスーパークラスタ 1 1 0 の機械（例えば、機械 1 2 0）と第 2 のスーパークラスタ 1 4 0 の機械（例えば、機械 1 4 2）と間に確立され得る。機械 1 2 0 は、地理的な近接度に基づいて、第 2 のスーパークラスタ 1 4 0 または機械 1 4 2 を特定できる。例えば、機械 1 2 0 は、近接度要求（例えば、DNS 要求）を行うことにより、第 2 のスーパークラスタ 1 4 0 および / または機械 1 4 2 が第 1 のスーパークラスタ 1 1 0 に隣接していることを特定できる。いくつかの構成において、CDN のクライアントがランデブーシステムを用いて、第 2 のスーパークラスタ 1 4 0 および / または機械 1 4 2 は第 1 のスーパークラスタ 1 1 0 に隣接している、と判断できるように、近接要求は、DNS 要求のランデブーシステムと同じまたは同様のランデブーシステムを用い得る。第 1 のスーパークラスタ 1 1 0 および第 2 のスーパークラスタ 1 4 0 の両方の機械の可用性ステータスおよびリソース割り当て情報は、ブリッジ 1 7 0 を介して共有され得る。例えば、機械 1 2 0 は、第 1 のスーパークラスタ 1 1 0 の機械 1 1 4 a - 1 1 4 e と機械 1 1 6 - 1 2 6 との可用性ステータスおよびリソース割り当て情報を Pub - Sub チャンネル 1 3 0 から抽出し、ブリッジ 1 7 0 を介して、可用性ステータスおよびリソース割り当て情報を機械 1 4 2 へ送信できる。機械 1 4 2 は、機械 1 1 4 a - 1 1 4 e と機械 1 1 6 - 1 2 6 との可用性ステータスおよびリソース割り当て情報を Pub - Sub チャンネル 1 6 0 にパブリッシュすることにより、第 2 のスーパークラスタ 1 4 0 の機械 1 4 4 - 1 5

10

20

30

40

50

0が当該データを利用可能になるようにできる。同様に、機械142は、第2のスーパークラスタ140の機械142-150の可用性ステータスおよびリソース割り当て情報を抽出し、ブリッジ170を介して、可用性ステータスおよびリソース割り当て情報を機械120へ送信できる。機械120は、機械142-150の可用性ステータスおよびリソース割り当て情報をPub-Subチャンネル130にパブリッシュすることにより、第1のスーパークラスタ110の機械114a-114eと機械116-126とが当該データを利用可能になるようにできる。いくつかの例において、本明細書において説明する近接度の選択は、最も近いスーパークラスタ（例えば、第2のスーパークラスタ140）を判断する際に、第1のスーパークラスタ110は損なわれ、第2のスーパークラスタ140は損なわれない状況を考慮に入れている。その意味では、ブリッジ170は、負荷およびスライス割り当て情報を1つの方向に（例えば、機械142から機械120へ）伝達するだけでよいことがある。そのため、第1のスーパークラスタ110内の機械114a-114eと機械116-126とは、コンテンツを機械114a-114eと機械116-126とに提供するために第2のスーパークラスタ140のどの機械が利用可能であるかを知る必要があるが、第2のスーパークラスタ140内の機械142-150は、第2のスーパークラスタ140内の機械142-150からスライスを取得し続ける。

10

#### 【0053】

図2は、いくつかの構成による機械200を示すブロック図である。図1および図2を参照すると、機械200は、機械114a-114eと、機械116-126と、機械142-150とのいずれかの例である。示されるように、機械200は、処理回路210、Pub-Sub回路220、ブリッジ回路230、ステータスデータベース240およびコンテンツデータベース250のうちの1つまたは複数を含む。

20

#### 【0054】

処理回路210は、本明細書において説明する様々な機能を機械200に対して実行するように構成される。処理回路210は、プロセッサ212およびメモリ214を含む。プロセッサ212は、汎用プロセッサ、特定用途向け集積回路（ASIC）、1または複数のフィールドプログラマブルゲートアレイ（FPGA）、デジタル信号プロセッサ（DSP）、処理コンポーネントのグループ、または他の適切な電子処理コンポーネントを用いて実装され得る。メモリ214は、ランダムアクセスメモリ（RAM）、リードオンリメモリ（ROM）、不揮発性RAM（NVRAM）、フラッシュメモリ、ハードディスクストレージまたは別の適切なデータストレージユニットを用いて実装され得る。メモリ214は、プロセッサ212により実行される様々な処理を容易にするためのデータおよび/またはコンピュータコードを格納する。さらに、メモリ214は、有形の非一時的揮発性メモリまたは不揮発性メモリであるか、または有形の非一時的揮発性メモリまたは不揮発性メモリを含む。したがって、メモリ214は、データベースコンポーネント、オブジェクトコードコンポーネント、スクリプトコンポーネント、または本明細書において説明する様々な機能をサポートするための任意の他の種類の情報構造を含む。

30

#### 【0055】

Pub-Sub回路220は、Pub-Subチャンネル（例えば、Pub-Subチャンネル130または160）を介して通信し、当該チャンネルを確立し、および/または当該チャンネルに接続するように構成された適切なハードウェアおよびソフトウェアを含む。例えば、Pub-Sub回路220は、（セルラー規格用に構成された）セルラートランシーバ、（802.11X、ZigBee（登録商標）、Bluetooth（登録商標）またはWi-Fi（登録商標）等用の）ローカル無線ネットワークトランシーバ、有線ネットワーク（例えば、Ethernet（登録商標）、SONET、DWDM、OTNなどの光ネットワーク）インタフェース、それらの組み合わせ（例えば、セルラートランシーバおよびBluetoothトランシーバの両方）および/またはこれらと同様のものを含む適切なネットワークインタフェースであり得る。Pub-Sub回路220は、無線ネットワークマルチキャスト、データリンクレイヤマルチキャスト（例えば、Ethernetマルチキャスト）、インターネットレイヤマルチキャスト（例えば、インターネ

40

50



ットプロトコル（IP）マルチキャスト、アプリケーションレイヤマルチキャストおよび別の適切な種類のマルチキャストのうちの1つまたは複数を有効にし得る。それに関連して、Pub-Subチャンネル130または160は、スーパークラスタの機械が接続する1または複数の適切なネットワーク（例えば、セルラーネットワーク、ローカル無線ネットワーク、有線ネットワーク、それらの組み合わせおよび/またはこれらと同様のもの）をさらに含み得る。Pub-Sub回路220は、処理回路210および/または別の適切な処理回路により提供される処理機能を含み得る。

#### 【0056】

Pub-Sub回路220は、機械200の可用性ステータスおよびリソース割り当て情報をPub-Subチャンネル130または160にパブリッシュできる。機械200がオンラインになり、オフラインになりそうであることに応答して、Pub-Sub回路220は、オンライン通知およびオフライン通知のうちのそれぞれをPub-Subチャンネル130または160にパブリッシュする。例えば、機械200が良好にかつ順序正しくシャットダウンされている場合、機械200は、オフライン通知を送信でき、機械200が現在担っているスライスの順序正しい遷移が同じスーパークラスタの他の機械に対して行われ得るように「排出」状態に移る。しかしながら、機械200は、良好にかつ順序正しくシャットダウンされるのではなく急に終了されることが頻繁にある。機械200が急に終了される状況において、機械200からのメッセージが、予め定められた期間（例えば、10秒、30秒、1分、5分等）にわたってないと、同じスーパークラスタの別の機械、または管理デバイスは、スーパークラスタの他の機械が通知されるようにスーパークラスタのPub-Subチャンネルを介してオフライン通知を生成および送信するようトリガされ得る。さらに、Pub-Sub回路220は、機械200のうちの1つまたは複数がオンラインになり、少なくとも1つのスライスが機械200に割り当てられたことに応答して、リソース割り当て情報、割り当てられたスライスに対するあらゆる修正（例えば、スライスの除去または変更）等をパブリッシュする。送信されたリソース割り当て情報は、機械200によりコンテンツデータベース250に格納されている現在のスライスに対応する。Pub-Sub回路220は、オンライン通知、オフライン通知およびリソース割り当て情報を、Pub-Subチャンネル130または160に対応するマルチキャストアドレスへ送信し得る。次に、送信された情報のコピーが、限定されるものではないが、例えばルータ、スイッチ、セルラーネットワーク基地局等の関連するネットワーク要素により自動的に、生成され、マルチキャストアドレスによりカバーされる受信者（例えば、同じスーパークラスタ内の機械）へ送信され得る。それに関連して、Pub-Subチャンネル130または160は、関連するネットワーク要素をさらに含む。

#### 【0057】

加えて、Pub-Sub回路220は、機械200と同じスーパークラスタ内の他の機械のオンライン通知、オフライン通知およびリソース割り当て情報をPub-Subチャンネル130または160から受信できる。Pub-Subチャンネル130または160に関連するマルチキャストアドレスが機械200の適切なアドレスも特定することにより、Pub-Sub回路220は、同じスーパークラスタ内の他の機械のオンライン通知、オフライン通知およびリソース割り当て情報を受信できる。

#### 【0058】

ブリッジ回路230は、ブリッジ（例えば、ブリッジ170）を介して通信し、当該ブリッジを確立し、および/または当該ブリッジに接続するように構成された適切なハードウェアおよびソフトウェアを含む。例えば、ブリッジ回路230は、（セルラー規格用に構成された）セルラートランシーバ、（802.11X、ZigBee（登録商標）、BluetoothまたはWi-Fi（登録商標）等用の）ローカル無線ネットワークランシーバ、有線ネットワーク（例えば、Ethernet（登録商標）、SONET、DWDM、OTNなどの光ネットワーク）インタフェース、それらの組み合わせ（例えば、セルラートランシーバおよびBluetoothランシーバの両方）および/またはこれらと同様のものを含む適切なネットワークインタフェースであり得る。ブリッジ回路2

10

20

30

40

50

30 および Pub - Sub 回路 220 は、セルラートランシーバ、ローカル無線ネットワークトランシーバおよび/または有線ネットワークインタフェースのうちの1つまたは複数共有し得る。ブリッジ回路 230 は、例えば、別のスーパークラスタの別の機械(限定されるものではないが機械 200 など)との TCP 接続を可能にし得る。それに関連して、ブリッジ回路 230 は、IP 規格をサポートする。ブリッジ回路 230 は、処理回路 210 および/または別の適切な処理回路により提供される処理機能を含み得る。

【0059】

ブリッジ回路 230 は、機械 200 と同じスーパークラスタ内の機械の可用性ステータスおよびリソース割り当て情報をステータスデータベース 240 から受信し、ブリッジ 170 を介して(例えば、TCP 接続を介して)、そのような可用性ステータスおよびリソース割り当て情報を別のスーパークラスタの機械へ送信できる。

10

【0060】

ステータスデータベース 240 は、(スーパークラスタの Pub - Sub チャンネルを介して受信される)機械 200 と同じスーパークラスタ内の各機械の可用性ステータスおよびリソース割り当て情報ならびに(ブリッジを介して受信される)機械 200 のスーパークラスタとは異なる少なくとも別のスーパークラスタの1または複数の機械の可用性ステータスおよびリソース割り当て情報を編成および格納するために実装されたデータ構造を有するメモリデバイスである。Pub - Sub 回路 220 およびブリッジ回路 230 が可用性ステータスおよびリソース割り当て情報を受信したことに応答して、可用性ステータスおよびリソース割り当て情報は、ステータスデータベース 240 に格納される。そのため、可用性ステータスおよびリソース割り当て情報のリアルタイムでの更新がステータスデータベース 240 に格納され得ることにより、ハイブアウェアネスが可能になる。

20

【0061】

コンテンツデータベース 250 は、機械 200 に割り当てられるコンテンツデータの部分またはスライスを編成および格納するために実装されたデータ構造を有するメモリデバイスである。いくつかの例において、コンテンツデータベース 250 は、2つの別個の格納部を含む。第1の格納部は、総ライブラリの割り当てられた部分またはスライス(「ワーキングセット」と称される)を格納する。第2の格納部は、総ライブラリの一般的なコンテンツデータを格納する。総ライブラリの所与のスライスがクライアントにより頻繁に要求された場合、頻繁に要求されたスライスは、一般的なコンテンツデータになり、コンテンツデータベース 250 の第2の格納部に格納される。

30

【0062】

図3は、様々な構成による、スーパークラスタ 110 内での通信を提供するための方法 300 の例を示すフロー図である。図1から図3を参照すると、方法 300 は、Pub - Sub チャンネル 130 を用いた、スーパークラスタ 110 内でのハイブアウェアネスの提供に関する。さらに、Pub - Sub チャンネル 130 およびブリッジ 170 により、2つの近接するスーパークラスタ 110 および 140 にわたるハイブアウェアネスが可能になる。当業者であれば、第1のスーパークラスタ 110 および第2のスーパークラスタ 140 の各機械が、方法 300 に関連して特定の機械について説明した機能を実行できることを理解し得る。

40

【0063】

310 において、機械(例えば、機械 200 に関して説明した方式で実装され得る機械 122)が、Pub - Sub チャンネル 130 をサブスクライブする。いくつかの例において、機械 122 の Pub - Sub 回路 220 は、Pub - Sub チャンネル 130 に対応するマルチキャストアドレスを介してデータの送信および受信を構成することにより、Pub - Sub チャンネル 130 をサブスクライブする。

【0064】

320 において、機械 122 が、機械 122 の可用性ステータスおよびリソース割り当て情報を Pub - Sub チャンネル 130 にパブリッシュする。機械 122 の Pub - Sub 回路 220 は、機械 122 の可用性ステータスおよびリソース割り当て情報をマルチキ

50

キャストアドレスへ送信する。そのような情報は、複製され、第1のスーパークラスタ110内の他の機械に対応するアドレスへ配信され得る。第1のスーパークラスタ110内の他の機械に対応するアドレスは、マルチキャストアドレスに対応する。同様に、機械122は、Pub-Subチャンネル130を介して、第1のスーパークラスタ110内の他の機械の可用性ステータスおよびリソース割り当て情報を受信できる。

#### 【0065】

330において、機械122が、Pub-Subチャンネル130上の可用性ステータスおよびリソース割り当て情報に基づいて、クライアント105から受信した、コンテンツデータの要求を第1のスーパークラスタ110の別の機械へ転送する。例えば、機械122の処理回路210は、クライアント105により要求されたスライスがコンテンツデータベース250に格納されているかどうかをチェックする。当該スライスは機械122のコンテンツデータベース250に格納されていないという判断に回答して、機械122は、Pub-Subチャンネル130を介して受信した第1のスーパークラスタ110内の他の機械の可用性ステータスおよびリソース割り当て情報に基づいて、要求されたスライスは第1のスーパークラスタ110の別の機械（例えば、機械126）に格納されている、と判断する。例えば、処理回路210は、要求されたスライスを基準として用いてステータスデータベース240を検索することで、要求されたスライスを格納している機械126を特定する。機械122の処理回路210は、機械126を特定すると、当該要求を機械126へ転送する。機械126は、任意の適切な方式で、要求されたスライス（例えば、当該スライスに関連するオブジェクト）をクライアント105に提供する。いくつかの例において、当該要求を転送する段階は、機械122が機械126のプロキシとしてサービスを提供しているように、機械122が当該要求を機械126へプロキシする段階を含む（例えば、当該要求は機械126へ転送され、応答が機械122を通る）。いくつかの例において、当該要求を転送する段階は、接続が機械126へ移行されることにより、機械126が応答（例えば、オブジェクト）をクライアント105に直接提供できる段階を含む。いくつかの例において、当該要求を転送する段階は、機械126がDirect Server Returnを用いて応答（例えば、オブジェクト）をクライアント105へ配信する段階を含む。いくつかの例において、当該要求を転送する段階は、機械122がリダイレクトメッセージをクライアント105へ戻すことにより、クライアント105が機械126にリダイレクトされ得る段階を含む。

#### 【0066】

図4は、いくつかの構成による、第1のスーパークラスタ110内での通信を提供するための方法400の例を示すフロー図である。図1から図4を参照すると、スーパークラスタの機械は、当該スーパークラスタのPub-Subチャンネル上の可用性ステータスおよびリソース割り当て情報に基づいて、スーパークラスタの別の機械により格納されているコンテンツデータ（例えば、1または複数のスライス）を格納できる。例えば、第1の機械（例えば、機械116）は、方法400を用いて、同じスーパークラスタ110内の少なくとも第2の機械（例えば、機械118）をオフロードし得る。

#### 【0067】

405において、機械116がオンラインになる。例えば、機械116は、スケジューリングされた、または予期しないシャットダウンの後にオンラインになり得る。この時点で、機械116のコンテンツデータベース250は、いかなるコンテンツデータも格納していない。機械116のPub-Sub回路220は、説明した方式で、Pub-Subチャンネル130をセットアップする。機械116は、第1のスーパークラスタ110内の他の機械の可用性ステータスおよびリソース割り当て情報のPub-Subチャンネル130からの受信を開始する。機械116は、第1のスーパークラスタ110内の他の機械の可用性ステータスおよびリソース割り当て情報の機械116のステータスデータベース240への格納を開始できる。

#### 【0068】

410において、機械116の処理回路210が、Pub-Subチャンネル130上の

10

20

30

40

50

可用性ステータスおよびリソース割り当て情報に基づいて、第1のスーパークラスタ110の第2の機械(例えば、機械118)により格納されているコンテンツデータを特定する。特定されたコンテンツデータは、機械116へオフロードされ得る。一例において、機械116の処理回路210が、機械118は第1のスーパークラスタ110の機械の間で最も多くの数のスライスまたは一般的なスライスを現在格納していると判断したこと、または、機械118は予め定められた閾値を超える数のスライスを格納していると判断したことに応答して、機械118のコンテンツデータは特定される。さらに別の例において、機械116の処理回路210が、機械118は第1のスーパークラスタ110の機械の間で最も長くオンラインになっていると判断したこと、または、機械118は予め定められた閾値を超える期間にわたってオンラインになっていると判断したことに応答して、機械118のコンテンツデータは特定される。さらに別の例において、機械116の処理回路210が、機械118はホットスポットであると判断したことに応答して、機械118のコンテンツデータは特定される。それに関連して、機械118は、クライアントからの要求で機械118が過負荷になっていることをPub-Subチャンネル130上でパブリッシュし得る。さらに別の例において、機械116の処理回路210が、機械118はオフラインになりそうであると判断したことに応答して、機械118のコンテンツデータは特定される。それに関連して、機械118は、機械118がオフラインにされているか、または速やかにオフラインになることをPub-Subチャンネル130上でパブリッシュし得る。さらに別の例において、機械116の処理回路210は、第1のスーパークラスタ110の利用可能な機械のうちの1つをランダムまたは擬似ランダムで選択する。いくつかの場合において、オフロードされるコンテンツデータ(例えば、スライス)は、ランダムまたは擬似ランダムで選択され得る。他の場合において、オフロードされるコンテンツデータは、コンテンツデータベース250の第2の格納部に格納されている一般的なコンテンツデータであってもよく、第2の格納部に格納されていないが機械118にホットスポットを生じさせているコンテンツデータであってもよい。

#### 【0069】

415において、機械116のPub-Sub回路220が、機械118により格納されているコンテンツデータを格納またはオフロードするために機械116は利用可能であることをPub-Subチャンネル130上でパブリッシュする。例えば、機械116のPub-Sub回路220は、Pub-Subチャンネル130に関連するマルチキャストアドレスへ可用性通知を送信し得る。当該通知は、機械118により格納されているコンテンツデータを格納するために機械116が利用可能であることを示す。

#### 【0070】

代替的に、ブロック410が省略されてよく、機械116のPub-Sub回路220は、第1のスーパークラスタ110のあらゆる利用可能な機械のあらゆるコンテンツデータを格納するために機械116が利用可能であることをPub-Subチャンネル130上でパブリッシュする。減速問題またはホットスポットを経験している機械は、許可を付与し得る。

#### 【0071】

420において、機械116の処理回路210が、機械116のPub-Sub回路220は機械118から許可を受信したかどうかを判断する。許可は、Pub-Subチャンネル130上で受信される任意の適切なメッセージであり得る。言い換えると、機械118は、Pub-Subチャンネル130上で許可をパブリッシュすることにより応答する。許可は受信されていない(420:いいえ)という判断に応答して、機械116の処理回路210は、425において、第1のスーパークラスタ110の別の機械をオフロードに設定し、410において、別の機械が特定される。一方、許可は受信されている(420:はい)という判断に応答して、機械116の処理回路210は、430において、機械118のコンテンツデータを格納する。機械118により格納されているコンテンツデータの全てのうちのいくつかは、機械116のコンテンツデータベース250に格納され得る。

10

20

30

40

50

## 【 0 0 7 2 】

機械 1 1 8 は、オフロードされたコンテンツデータの格納を続け得る。コンテンツデータまたはスライスに関連する URL はここで、機械 1 1 6 および 1 1 8 の両方に関連している。代替的に、機械 1 1 8 は、オフロードされたコンテンツデータをもはや格納しなくてもよく、当該コンテンツデータが機械 1 1 6 へオフロードされた後はオフラインであってもよい。そのような場合、当該コンテンツデータは、機械 1 1 6 へ移行されたものとみなされる。当該コンテンツデータまたはスライスに関連する URL は、機械 1 1 6 へ迂回される。

## 【 0 0 7 3 】

4 3 5 において、機械 1 1 6 が、機械 1 1 6 により格納されている、中継または転送されたコンテンツデータ要求を第 3 の機械（例えば、機械 1 2 2）から受信し得る。要求されたコンテンツデータは、機械 1 1 6 が機械 1 1 8 からオフロードしたコンテンツデータである。機械 1 1 6 が機械 1 1 8 をオフロードしていると仮定すると、機械 1 2 2 は、当該要求をクライアント 1 0 5 から受信し、当該要求を機械 1 1 6 へ中継する。機械 1 2 2 は、機械 1 2 2 が Pub - Sub チャンネル 1 3 0 上のデータストリームのモニタリングおよび格納もしているため機械 1 1 6 が機械 1 1 8 をオフロードしているという認識を有する。Pub - Sub チャンネル 1 3 0 をサブスクライブする第 1 のスーパークラスタ 1 1 0 内の全ての機械が、Pub - Sub チャンネル 1 3 0 上のデータストリームを利用可能である。

10

## 【 0 0 7 4 】

4 4 0 において、機械 1 1 6 が、中継または転送された要求毎にコンテンツデータを提供し得る。4 4 5 において、機械 1 1 6 の処理回路 2 1 0 が、配信閾値は満たされているかどうかを判断する。配信閾値は、機械 1 1 6 が格納および配信できるコンテンツデータの限度を指す。配信閾値は、予め定められてもよく、動的に設定されてもよい。配信閾値は、コンテンツデータベース 2 5 0 のサイズ、処理回路 2 1 0 の処理機能、ネットワークレイテンシ、スーパークラスタ内の利用可能な機械の数等に基づいて、決定または設定され得る。

20

## 【 0 0 7 5 】

機械 1 1 6 の処理回路 2 1 0 が、配信閾値は満たされている（4 4 5 : はい）と判断したことに応答して、機械 1 1 6 は、4 3 5 において、コンテンツデータベース 2 5 0 に格納されているコンテンツデータの追加の中継された要求を第 1 のスーパークラスタ 1 1 0 の機械から受信し得る（または要求をクライアントから直接受信し得る）。機械 1 1 6 は、第 1 のスーパークラスタ 1 1 0 の他の機械からの追加のコンテンツデータをもはやオフロードしなくてもよく、総ライブラリの追加のコンテンツデータを格納してもよい。

30

## 【 0 0 7 6 】

一方、機械 1 1 6 の処理回路 2 1 0 が、配信閾値は満たされていない（4 4 5 : いいえ）と判断したことに応答して、機械 1 1 6 は、4 1 5 において、機械 1 1 8（または第 1 のスーパークラスタ 1 1 0 の別の機械）により格納されている追加のコンテンツデータを格納するための可用性を Pub - Sub チャンネル 1 3 0 にパブリッシュし得る。

## 【 0 0 7 7 】

さらに、スーパークラスタ内の利用可能な機械は、スーパークラスタ内の利用可能な過負荷になっている機械をオフロードできる。図 5 は、いくつかの構成による、第 1 のスーパークラスタ 1 1 0 内での通信を提供するための方法 5 0 0 の例を示すフロー図である。図 1 から図 5 を参照すると、5 0 5 において、第 1 の機械（例えば、機械 1 1 6）の処理回路 2 1 0 は、Pub - Sub チャンネル 1 3 0 上の可用性ステータスおよびリソース割り当て情報に基づいて、コンテンツデータ（例えば、1 または複数のスライス）を現在格納している第 1 のスーパークラスタ 1 1 0 内の機械の数は閾値未満である、と判断する。第 1 のスーパークラスタ 1 1 0 は、冗長性のために、同じスライスを格納する 2 つまたはそれよりも多くの機械を有し得る。あるスライスを格納する機械が無いと、当該スライスを格納する機械にホットスポットが生じ得る。これに応答して、機械 1 1 6 は、方法 5 0 0

40

50

の残りをを用いて、当該スライスを格納している残りの機械をオフロードする。ブロック 5 1 0 - 5 4 5 の各々は、ブロック 4 1 0 - 4 4 5 のうちの 1 つに対応する。いくつかの構成において、スライスは典型的には、単一の機械によりキャッシュされる。大量のピアトラフィックに起因してある機械がホットスポットになった場合、当該機械は、同じスーパークラスタ上のいくつかの他の機械上へ特定のスライスを放出し得る。その意味では、わずかなトラフィックを受信しているか、またはトラフィックを受信していないスライスを移動させても、状況が改善されないことがあり、最も頻繁に要求されたスライスを移動させると、ホットスポットを同じスーパークラスタ内の別の機械へ移動させてしまう虞があり得る。したがって、いくつかの例において、最も頻繁に要求されたスライス、または要求されることが最も少なかったスライスのいずれでもないスライスが、本明細書において

10

#### 【 0 0 7 8 】

図 6 は、本開示のいくつかの構成による、複数のスーパークラスタ（例えば、図 1 のスーパークラスタ 1 1 0 および 1 4 0）間での通信を提供するための方法 6 0 0 の例を示すフロー図である。図 1 から図 6 を参照すると、Pub - Sub チャンネル 1 3 0 および Pub - Sub チャンネル 1 6 0 により保持される可用性ステータスおよびリソース割り当て情報は、ブリッジ 1 7 0 を介して共有され得る。機械 1 2 0 および 1 4 2 はブリッジ 1 7 0 をサポートする機械であるが、当業者であれば、ブリッジ（限定されるものではないがブリッジ 1 7 0 など）が第 1 のスーパークラスタ 1 1 0 の任意の機械と第 2 のスーパークラスタ 1 4 0 の任意の機械との間で確立され得ることを理解できる。

20

#### 【 0 0 7 9 】

6 1 0 において、第 1 の機械（例えば、第 1 のスーパークラスタ 1 2 0 の機械 1 2 0）が、第 1 のスーパークラスタ 1 1 0 内の利用不可能な機械の数は閾値を超えている、と判断する。第 1 のスーパークラスタ 1 1 0 内の利用不可能な機械の数は、第 1 のスーパークラスタ 1 1 0 の Pub - Sub チャンネル 1 3 0 上で提供される可用性ステータスに基づいて決定される。

#### 【 0 0 8 0 】

6 2 0 において、機械 1 2 0 が、第 1 のスーパークラスタ 1 1 0 内の利用不可能な機械の数は閾値を超えているという判断に回答して、第 2 のスーパークラスタ 1 4 0 内の第 2 の機械（例えば、機械 1 4 2）とのブリッジ 1 7 0 を確立する。ブリッジ 1 7 0 の例が TCP 接続である。機械 1 2 0 は、第 1 のスーパークラスタ 1 1 0 に対するネットワーク近接度に基づいて、第 2 のスーパークラスタ 1 4 0 および/または機械 1 4 2 を特定する。ネットワーク近接度は、パケットが機械 1 2 0（または第 1 のスーパークラスタ 1 1 0）と機械 1 4 2（または第 2 のスーパークラスタ 1 4 0）との間を移動する距離に対応する。第 2 のスーパークラスタ 1 4 0 は、第 1 のスーパークラスタ 1 1 0 に隣接する。例えば、機械 1 2 0 は、近接度要求（例えば、DNS 要求、または DNS 要求のランデブーシステムと同じまたは同様のランデブーシステムを用いる要求）を行うことにより、第 2 のスーパークラスタ 1 4 0 および/または機械 1 4 2 が第 1 のスーパークラスタ 1 1 0 に隣接していることを特定する。機械 1 2 0 は、DNS 要求への（例えば、IP アドレスを含む）応答を適切な DNS サーバから受信し、当該応答に関連する位置情報（例えば、IP アドレス）を分析して、第 2 のスーパークラスタ 1 4 0 を最も近い近接物と特定できる。いくつかの構成において、機械 1 2 0 は、最も近い近接するスーパークラスタの数（例えば、2、3、5 または 1 0）を特定すると共に、最も少ないホットスポット、最も高い処理機能、最も多い利用可能な機械、最も少ないレイテンシ等を有する 1 つを選択する。いくつかの例において、第 2 のスーパークラスタ 1 4 0（例えば、機械 1 4 2 または別の適切な機械）は、第 2 のスーパークラスタ 1 4 0 がブリッジ 1 7 0 を確立する前に第 1 のスーパークラスタ 1 1 0 をオフロードすることが可能である、と判断する。例えば、第 2 のスーパークラスタ 1 4 0（例えば、機械 1 4 2 または別の適切な機械）は、Pub - Sub チャンネル 1 6 0 上で伝達される情報を用いた、第 2 のスーパークラスタ 1 4 0 内の利用不可能な機械の数は予め定められた閾値（例えば、3 0 %、5 0 %、6 0 % 等）を超えてい

30

40

50

るという判断に応答してブリッジ 170 が確立されることを可能にする。別の例において、第 2 のスーパークラスタ 140 (例えば、機械 142 または別の適切な機械) は、第 2 のスーパークラスタ 140 の追加のブリッジの数は予め定められた閾値 (例えば、0、1、2 等) を超えていないという判断に応答してブリッジ 170 が確立されることを可能にする。

【0081】

630 において、機械 120 が、ブリッジ 170 を介して、第 2 のスーパークラスタ 140 の機械 142 - 150 の各々の可用性ステータスおよびリソース割り当て情報を受信する。

【0082】

640 において、機械 120 が、第 2 のスーパークラスタ 140 の機械 142 - 150 の各々の可用性ステータスおよびリソース割り当て情報を第 1 のスーパークラスタ 110 の Pub - Sub チャンネル 130 上でパブリッシュする。例えば、機械 120 は、第 2 のスーパークラスタ 140 の機械 142 - 150 の各々の可用性ステータスおよびリソース割り当て情報を、Pub - Sub チャンネル 130 に関連するマルチキャストアドレスへ送信する。そのため、機械 114a - 114e と機械 116 - 126 とは、マルチキャストアドレスに基づく第 2 のスーパークラスタ 140 の機械 142 - 150 の各々の可用性ステータスおよびリソース割り当て情報を受信および格納できる。

【0083】

いくつかの構成において、第 1 のスーパークラスタ 110 内のブリッジ 170 の端部を形成する機械 120 は、第 1 のスーパークラスタ 110 内の利用可能な機械がサポートできないそれらのスライスのみが Pub - Sub チャンネル 130 に対して利用可能になるように、第 2 のスーパークラスタ 140 から受信するデータをフィルタリングできる。そのような処理は移り変わり得る。なぜなら、スライスを第 2 のスーパークラスタ 140 へとオフロードするために必要なことが原因で第 1 のスーパークラスタ 110 がまず閾値に近づくが、全てのスライスが依然として第 1 のスーパークラスタ 110 内の機械からサービスを受けていると思われるからである。ブリッジ 170 が確立されているので、第 1 のスーパークラスタ 110 の機械がさらに劣化しても、第 1 のスーパークラスタ 110 の残りの機械の間で追加のスライスが割り当てられることは一切ない。代わりに、第 2 のスーパークラスタ 140 の適切な機械から追加のスライスが充填される。ブリッジ 170 が確立される時点で、または当該時点よりも前に、最近再割り当てされたスライスのうちのいくつかをロールバックすることにより、したがって、それらのスライスが第 2 のスーパークラスタ 140 から充填されるようにすることにより、第 1 のスーパークラスタ 110 の残りの機械のある程度の軽減を (第 1 のスーパークラスタ 110 内の追加の機械の喪失なしでも) 実現させることが可能である。例を用いて示すと、機械 1 がスライス 1、11、21、31 等を有し、機械 2 がスライス 2、12、22 等を有するように、100 個のスライスおよび 10 個の機械がスーパークラスタ内に存在する。機械 1 がオフラインになり、次に、スライス 1 が機械 2 に再割り当てされ得る場合、例えば、スライス 11 は、機械 2 に再割り当てされてよく、スライス 21 は、機械 3 に再割り当てされてよい。オフラインになっている機械 2、3、4 および 5 に対してそのような処理が繰り返されている場合、残りの機械の各々は、通常の 2 倍の数のスライスにサービスを提供している。ブリッジに接続するための閾値に到達した場合、これは、機械 4 および 5 からのスライス、例えば、スライス 4、14、24 および 34 ならびにスライス 5、15、25 および 35 がスーパークラスタにおいて「利用不可能」とみなされ得るということであり得る。代わりに、それらのスライスは、別のスーパークラスタから処理される。すなわち、本明細書において説明したシステムは、5 個の機械がスーパークラスタ内でオフラインになるまでブリッジ 170 が実際に生成されなくても、ブリッジを生成するための閾値がスーパークラスタ外の 3 個の機械であるかのように挙動し得る。

【0084】

第 1 のスーパークラスタ 110 の機械 114a - 114e と機械 116 - 126 とは、

10

20

30

40

50

ブリッジ170を介して受信した第2のスーパークラスタ140の機械142-150の各々の可用性ステータスおよびリソース割り当て情報に基づいて、クライアントから受信した要求を第2のスーパークラスタ140の機械へ転送できる。例えば、機械122は、コンテンツデータ(例えば、スライス)の要求をクライアント105から受信する。機械122は、機械122のコンテンツデータベース250内のPub-Subチャンネル130を介して受信した第2のスーパークラスタ140の機械142-150の各々の可用性ステータスおよびリソース割り当て情報を保存する。

【0085】

機械122は、Pub-Subチャンネル130上の機械142-150の各々の可用性ステータスおよびリソース割り当て情報に基づいて、コンテンツデータ(例えば、要求されたスライス)は第2のスーパークラスタ140の機械(例えば、機械146)により格納されている、と判断する。例えば、機械122の処理回路210は、要求されたスライスを基準として用いて機械122のステータスデータベース240を検索することで、要求されたスライスを格納している機械146を特定する。これに応答して、機械122は、要求を機械146へ転送する。

10

【0086】

本明細書において説明した構成は、図面を参照して説明されている。図面は、本明細書において説明したシステム、方法およびプログラムを実装する特定の構成の特定の詳細を示す。しかしながら、図面を用いた構成の説明は、図面に存在し得るいかなる限定も本開示に課すものと解釈されるべきではない。

20

【0087】

本明細書における特許請求の範囲のいかなる要素も、当該要素が「ための手段」という文言を用いて明示的に記載されていない限り、米国特許法第212(f)条の規定に基づいて解釈されるべきではないことを理解されたい。

【0088】

本明細書において用いられる場合、「回路」という用語は、本明細書において説明した機能を実行するように構造化されたハードウェアを含み得る。いくつかの構成において、対応する「回路」の各々は、本明細書において説明した機能を実行するようにハードウェアを構成するための機械可読媒体を含み得る。回路は、処理回路、ネットワークインタフェース、周辺デバイス、入力デバイス、出力デバイス、センサ等を含むがこれらに限定されない1または複数の回路コンポーネントとして具現化され得る。いくつかの構成において、回路は、1または複数のアナログ回路、電子回路(例えば、集積回路(IC)、ディスクリット回路、システムオンチップ(SOC)回路等)、電気通信回路、ハイブリッド回路および任意の他の種類の「回路」の形態を取り得る。これに関連して、「回路」は、本明細書において説明したオペレーションの実現を達成するか、または容易にするための任意の種類のコンポーネントを含み得る。例えば、本明細書において説明した回路は、1または複数のトランジスタ、論理ゲート(例えば、NAND、AND、NOR、OR、XOR、NOT、XNOR等)、抵抗器、マルチプレクサ、レジスタ、コンデンサ、インダクタ、ダイオード、ワイヤ等を含み得る。

30

【0089】

「回路」は、1または複数のメモリまたはメモリデバイスに通信可能に結合された1または複数のプロセッサも含み得る。これに関連して、1または複数のプロセッサは、メモリに格納された命令を実行し得るか、または、本来であれば1または複数のプロセッサにアクセス可能な命令を実行し得る。いくつかの構成において、1または複数のプロセッサは、様々な態様で具現化され得る。1または複数のプロセッサは、少なくとも本明細書において説明したオペレーションを実行するのに十分であるように解釈され得る。いくつかの構成において、1または複数のプロセッサは、複数の回路により共有され得る(例えば、回路Aおよび回路Bが、いくつかの例示的な構成においてメモリの異なるエリアを介して格納されるか、またはそうでなければアクセスされる命令を実行し得る同じプロセッサを備え得るか、またはそうでなければ共有し得る)。代替的に、または追加的に、1また

40

50



は複数のプロセッサは、1または複数のコプロセッサとは無関係に特定のオペレーションを行うように、またはそうでなければ実行するように構造化され得る。他の例示的な構成において、2つまたはそれよりも多くのプロセッサがバスを介して結合されることで、独立した、並行の、パイプライン化またはマルチスレッド化された命令の実行が可能になり得る。各プロセッサは、1または複数の汎用プロセッサ、ASIC、FPGA、DSP、またはメモリにより提供されされる命令を実行するように構造化された他の適切な電子データ処理コンポーネントとして実装され得る。1または複数のプロセッサは、シングルコアプロセッサ、マルチコアプロセッサ（例えば、デュアルコアプロセッサ、トリプルコアプロセッサ、クアドコアプロセッサ等）、マイクロプロセッサ等の形態を取り得る。いくつかの構成において、1または複数のプロセッサは、装置の外部にあってよく、例えば、1または複数のプロセッサは、リモートプロセッサ（例えば、クラウドベースプロセッサ）であってよい。代替的に、または追加的に、1または複数のプロセッサは、装置の内部および/またはローカルであってよい。これに関連して、所与の回路またはそのコンポーネントが、ローカルに（例えば、ローカルサーバ、ローカルコンピューティングシステム等の一部として）、またはリモートで（例えば、クラウドベースサーバなどのリモートサーバの一部として）配置され得る。そのために、本明細書において説明した「回路」は、1または複数の位置にわたって分散されたコンポーネントを含み得る。

10

#### 【0090】

当該構成のシステム全体または部分を実装するための例示的なシステムは、処理ユニットと、システムメモリと、システムメモリを含む様々なシステムコンポーネントを処理ユニットに結合させるシステムバスとを含むコンピュータの形態の汎用コンピューティングコンピュータを含み得る。各メモリデバイスは、非一時的揮発性記憶媒体、不揮発性記憶媒体、非一時的記憶媒体（例えば、1または複数の揮発性メモリおよび/または不揮発性メモリ）等を含み得る。いくつかの構成において、不揮発性媒体は、ROM、フラッシュメモリ（例えば、NAND、3D NAND、NOR、3D NOR等のフラッシュメモリ）、EEPROM、MRAM、磁気ストレージ、ハードディスク、光ディスク等の形態を取り得る。他の構成において、揮発性記憶媒体は、RAM、TRAM、ZRAM等の形態を取り得る。上述のものの組み合わせも、機械可読媒体の範囲内に含まれる。これに関連して、機械実行可能命令は、例えば、汎用コンピュータ、専用コンピュータまたは専用処理機械に特定の機能または機能のグループを実行させる命令およびデータを含む。対応するメモリデバイスの各々は、プロセッサ命令および関連データを含む、1または複数の関連する回路により実行されるオペレーションに関する情報（例えば、データベースコンポーネント、オブジェクトコードコンポーネント、スクリプトコンポーネント等）を本明細書において説明した例示的な構成に従って維持するように、またはそうでなければ格納するように動作可能であってよい。

20

30

#### 【0091】

本明細書において説明した「入力デバイス」という用語は、キーボード、キーパッド、マウス、ジョイスティックまたは同様の機能を実行する他の入力デバイスを含むがこれらに限定されない任意の種類の入力デバイスを含み得ることにも留意されたい。比較的、本明細書において説明した「出力デバイス」という用語は、コンピュータモニタ、プリンタ、ファクシミリ機または同様の機能を実行する他の出力デバイスを含むがこれらに限定されない任意の種類出力デバイスを含み得る。

40

#### 【0092】

本明細書における図は方法の段階の特定の順序および構成を示し得るが、これらの段階の順序は示されているものとは異なり得ると理解されることに留意されたい。例えば、2つまたはそれよりも多くの段階が、同時に、または部分的に同時に実行され得る。また、別個の段階として実行されるいくつかの方法の段階が組み合わてよく、組み合わせられた段階として実行されている段階が別個の段階に分離されてよく、一連の特定の処理が逆にされるか、またはそうでなければ変更されてよく、別個の処理の性質または数は変更されたり変化させられたりしてよい。任意の要素または装置の順序または順番は、代替的な構成

50

に従って変更されたり置き換えられたりしてよい。したがって、全てのそのような修正は、添付の特許請求の範囲において定義される本開示の範囲内に含まれるよう意図されている。そのような変更は、選択される機械可読媒体およびハードウェアシステムと、設計者の選択とに依存する。全てのそのような変更は本開示の範囲内であることが理解される。同様に、本開示のソフトウェアおよびウェブの実装は、ルールベースロジックと様々なデータベース検索段階、関連付け段階、比較段階および決定段階を実現する他のロジックとを伴う標準的なプログラミング技術を用いて実現され得る。

【 0 0 9 3 】

構成についての前述の説明を例示および説明の目的で提示した。包括的なものとするとしても、本開示を開示された正確な形式に限定することも意図されておらず、修正および変更が、上述の教示に照らして可能であるか、または本開示から取得され得る。当該構成は、本開示の原理およびその実際の用途を説明することで、様々な構成を当業者が利用することを可能にすべく、かつ、様々な修正が、意図された特定の用途に適合するよう、選択および説明されている。添付の特許請求の範囲において表される本開示の範囲から逸脱することなく、他の置き換え、修正、変更および省略が、当該構成の設計、動作条件および配置で行われ得る。

[ 項目 1 ]

複数の機械と、

上記複数の機械の各々がサブスクライブするパブリッシュ - サブスクライブ ( P u b - S u b ) チャンネルと

を備え、

上記複数の機械の各々は、

可用性ステータスおよびリソース割り当て情報を上記 P u b - S u b チャンネルにパブリッシュし、

上記 P u b - S u b チャンネルを介して、上記複数の機械のうちの少なくとも別の 1 つの可用性ステータスおよびリソース割り当て情報を受信し、

上記 P u b - S u b チャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を上記複数の機械のうちの上記別の 1 つへ転送する

ように構成される、

スーパークラスタ。

[ 項目 2 ]

上記可用性ステータスは、コンテンツデータを格納するために上記複数の機械の各々が利用可能であるかどうかを示し、

上記リソース割り当て情報は、上記複数の機械の各々により格納される上記コンテンツデータを特定する、

項目 1 に記載のスーパークラスタ。

[ 項目 3 ]

上記 P u b - S u b チャンネルは、マルチキャストアドレスを有する、項目 1 に記載のスーパークラスタ。

[ 項目 4 ]

上記 P u b - S u b チャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、上記クライアントから受信した、上記コンテンツデータの上記要求を上記複数の機械のうちの上記別の 1 つへ転送することは、

上記複数の機械のうちの第 1 の機械が上記コンテンツデータの上記要求を上記クライアントから受信することと、

上記第 1 の機械が、上記第 1 の機械は上記コンテンツデータを格納していない、と判断することと、

上記 P u b - S u b チャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、上記コンテンツデータは上記複数の機械のうちの第 2 の機械に格納されて

10

20

30

40

50

いる、と判断することと、

上記要求を上記第2の機械へ転送することと

を含む、

項目1に記載のスーパークラスタ。

[項目5]

上記複数の機械のうちの第1の機械は、上記Pub - Subチャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、上記複数の機械のうちの第2の機械により所有されているコンテンツデータを所有する、項目1に記載のスーパークラスタ。

[項目6]

上記第1の機械は、上記第2の機械により現在所有されている上記コンテンツデータを所有するために、

上記Pub - Subチャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、上記第2の機械により所有されている上記コンテンツデータを特定し、

上記第2の機械により所有されている上記コンテンツデータを所有するための可用性を上記Pub - Subチャンネルにパブリッシュし、

上記Pub - Subチャンネルを介して、上記第2の機械からの許可を上記第2の機械から受信し、

上記第2の機械により所有されている上記コンテンツデータを所有し、

上記コンテンツデータの中継された要求を上記複数の機械のうちの第3の機械から受信し、かつ、

上記中継された要求に回答して、上記コンテンツデータを提供する、

項目5に記載のスーパークラスタ。

[項目7]

上記第1の機械は、上記第2の機械により現在所有されている上記コンテンツデータを所有するために、

上記Pub - Subチャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、上記コンテンツデータを現在格納している第2の機械の数は閾値未満である、と判断し、

上記コンテンツデータを所有するための可用性を上記Pub - Subチャンネルにパブリッシュし、

上記Pub - Subチャンネルを介して、上記コンテンツデータを現在格納している上記第2の機械から許可を受信し、

上記コンテンツデータを所有し、

上記コンテンツデータの中継された要求を上記複数の機械のうちの第3の機械から受信し、かつ、

上記中継された要求に回答して、上記コンテンツデータを提供する、

項目5に記載のスーパークラスタ。

[項目8]

上記複数の機械のうちの第1の機械は、別のスーパークラスタ内の第2の機械へのブリッジを確立するように構成され、

上記別のスーパークラスタは、上記スーパークラスタに近接し、

上記ブリッジは、上記別のスーパークラスタの少なくともいくつかの機械の各々の可用性ステータスおよびリソース割り当て情報を上記第1の機械へ中継するように構成される、

項目1に記載のスーパークラスタ。

[項目9]

上記第1の機械は、上記Pub - Subチャンネル上の上記複数の機械の上記可用性ステータスに基づいて、上記複数の機械のうちの利用不可能な機械の数は閾値を超えている、と判断するように構成され、

上記利用不可能な機械の数は上記閾値を超えているという判断に回答して、上記ブリッジは、確立される、

10

20

30

40

50

項目 8 に記載のスーパークラスタ。

[ 項目 1 0 ]

上記ブリッジは、伝送制御プロトコル ( T C P ) 接続を含む、項目 8 に記載のスーパークラスタ。

[ 項目 1 1 ]

上記第 1 の機械は、上記スーパークラスタに対する地理的な近接度に基づいて、上記別のスーパークラスタを特定するように構成される、項目 8 に記載のスーパークラスタ。

[ 項目 1 2 ]

上記第 1 の機械は、ドメインネームシステム ( D N S ) 要求を行うことにより、上記別のスーパークラスタが上記スーパークラスタに隣接していることを特定する、項目 1 1 に記載のスーパークラスタ。

10

[ 項目 1 3 ]

上記第 1 の機械は、上記 P u b - S u b チャンネル上の上記別のスーパークラスタの上記少なくともいくつかの機械の上記可用性ステータスおよび上記リソース割り当て情報をパブリッシュするように構成される、項目 8 に記載のスーパークラスタ。

[ 項目 1 4 ]

上記スーパークラスタの第 3 の機械は、  
上記コンテンツデータの上記要求を上記クライアントから受信し、  
上記 P u b - S u b チャンネル上の上記別のスーパークラスタの上記少なくともいくつかの機械の上記リソース割り当て情報に基づいて、上記コンテンツデータは上記別のスーパー  
クラスタの第 4 の機械により格納されている、と判断し、かつ、  
上記要求を上記第 4 の機械へ転送する  
ように構成される、  
項目 1 3 に記載のスーパークラスタ。

20

[ 項目 1 5 ]

スーパークラスタの複数の機械にわたって負荷を分散するための方法であって、  
上記スーパークラスタの上記複数の機械の各々が、可用性ステータスおよびリソース割り当て情報をパブリッシュ - サブスクライブ ( P u b - S u b ) チャンネルにパブリッシュする段階であって、上記複数の機械の各々は、上記 P u b - S u b チャンネルをサブスクライブする、パブリッシュする段階と、  
上記 P u b - S u b チャンネルを介して、上記複数の機械のうちの少なくとも別の 1 つの可用性ステータスおよびリソース割り当て情報を受信する段階と、  
上記 P u b - S u b チャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を上記複数の機械のうちの上記別の 1 つへ転送する段階と  
を備える方法。

30

[ 項目 1 6 ]

実行された場合、スーパークラスタの複数の機械の各々のプロセッサに、  
可用性ステータスおよびリソース割り当て情報をパブリッシュ - サブスクライブ ( P u b - S u b ) チャンネルにパブリッシュする手順であって、上記複数の機械の各々は、上記  
P u b - S u b チャンネルをサブスクライブする、パブリッシュする手順と、  
上記 P u b - S u b チャンネルを介して、上記複数の機械のうちの少なくとも別の 1 つの可用性ステータスおよびリソース割り当て情報を受信する手順と、  
上記 P u b - S u b チャンネル上の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、クライアントから受信した、コンテンツデータの要求を上記複数の機械のうちの上記別の 1 つへ転送する手順と  
を実行させるようなコンピュータ可読命令を備える非一時的コンピュータ可読媒体。

40

[ 項目 1 7 ]

スーパークラスタにわたって負荷を分散するための方法であって、  
第 1 のスーパークラスタの第 1 の機械が、上記第 1 のスーパークラスタ内の利用不可能

50

な機械の数は閾値を超えている、と判断する段階と、

上記第1の機械が第2のスーパークラスタ内の第2の機械とのブリッジを確立する段階であって、上記第2のスーパークラスタは、上記第1のスーパークラスタに隣接する、確立する段階と、

上記第1の機械が、上記ブリッジを介して、上記第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報を受信する段階と、

上記第1の機械が、上記第1のスーパークラスタのパブリッシュ - サブスクリプション ( P u b - S u b ) チャンネル上で、上記第2のスーパークラスタの上記複数の機械の各々の上記可用性ステータスおよび上記リソース割り当て情報をパブリッシュする段階と

を備える方法。

[ 項目 1 8 ]

上記可用性ステータスは、コンテンツデータを格納するために上記第2のスーパークラスタの上記複数の機械の各々が利用可能であるかどうかを示し、

上記リソース割り当て情報は、上記第2のスーパークラスタの上記複数の機械の各々が担う上記コンテンツデータを特定する、

項目 1 7 に記載の方法。

[ 項目 1 9 ]

上記第1のスーパークラスタの上記 P u b - S u b チャンネルの上記第1のスーパークラスタの複数の機械が上記第1のスーパークラスタの上記機械の各々の可用性ステータスおよびリソース割り当て情報をパブリッシュする段階をさらに備える、項目 1 7 に記載の方法。

[ 項目 2 0 ]

上記第1のスーパークラスタ内の上記利用不可能な機械の数は、上記第1のスーパークラスタの上記 P u b - S u b チャンネル上で提供される上記可用性ステータスに基づいて決定される、項目 1 9 に記載の方法。

[ 項目 2 1 ]

上記第1のスーパークラスタの上記 P u b - S u b チャンネルは、マルチキャストアドレスを含む、項目 1 7 に記載の方法。

[ 項目 2 2 ]

上記第1のスーパークラスタの第3の機械がコンテンツデータの要求をクライアントから受信する段階と、

上記第3の機械が、上記 P u b - S u b チャンネルの上記第2のスーパークラスタの上記複数の機械の各々の上記可用性ステータスおよび上記リソース割り当て情報に基づいて、上記コンテンツデータは上記第2のスーパークラスタの第4の機械により格納されている、と判断する段階と、

上記要求を上記第2のスーパークラスタ内の上記第4の機械へ転送する段階と

をさらに備える、項目 1 7 に記載の方法。

[ 項目 2 3 ]

上記ブリッジは、伝送制御プロトコル ( T C P ) 接続を含む、項目 1 7 に記載の方法。

[ 項目 2 4 ]

上記第1の機械が、上記第1のスーパークラスタに対する地理的な近接度に基づいて、上記第2のスーパークラスタを特定する段階をさらに備える、項目 1 7 に記載の方法。

[ 項目 2 5 ]

上記第1の機械は、ドメインネームシステム ( D N S ) 要求を行うことにより、上記第2のスーパークラスタが上記第1のスーパークラスタに隣接していることを特定する、項目 2 4 に記載の方法。

[ 項目 2 6 ]

実行された場合、第1のスーパークラスタの第1の機械のプロセッサに、

上記第1のスーパークラスタ内の利用不可能な機械の数は閾値を超えている、と判断する手順と、

10

20

30

40

50

第2のスーパークラスタ内の第2の機械とのブリッジを確立する段階であって、上記第2のスーパークラスタは、上記第1のスーパークラスタに隣接する、確立する手順と、  
 上記ブリッジを介して、上記第2のスーパークラスタの複数の機械の各々の可用性ステータスおよびリソース割り当て情報を受信する手順と、

上記第1のスーパークラスタのパブリッシュ-サブスクライブ(Pub-Sub)チャンネル上で、上記第2のスーパークラスタの上記複数の機械の各々の上記可用性ステータスおよび上記リソース割り当て情報をパブリッシュする手順と

を実行させるようなコンピュータ可読命令を備える非一時的コンピュータ可読媒体。

[項目27]

複数の第1の機械と、

上記複数の第1の機械の各々がサブスクライブするパブリッシュ-サブスクライブ(Pub-Sub)チャンネルと

を有する第1のスーパークラスタと、

複数の第2の機械

を有する第2のスーパークラスタと、

上記第1のスーパークラスタと上記第2のスーパークラスタとの間のブリッジと

を備え、

上記複数の第1の機械のうちの1つは、

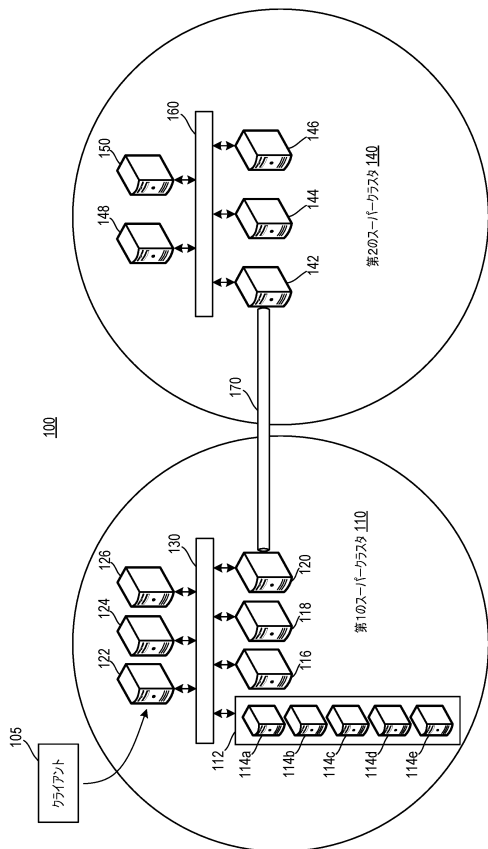
上記ブリッジを介して、上記複数の第2の機械の各々の可用性ステータスおよびリソース割り当て情報を上記複数の第2の機械のうちの1つから受信し、かつ、

上記第1のスーパークラスタの上記Pub-Subチャンネル上で、上記複数の第2の機械の各々の上記可用性ステータスおよび上記リソース割り当て情報をパブリッシュするように構成される、

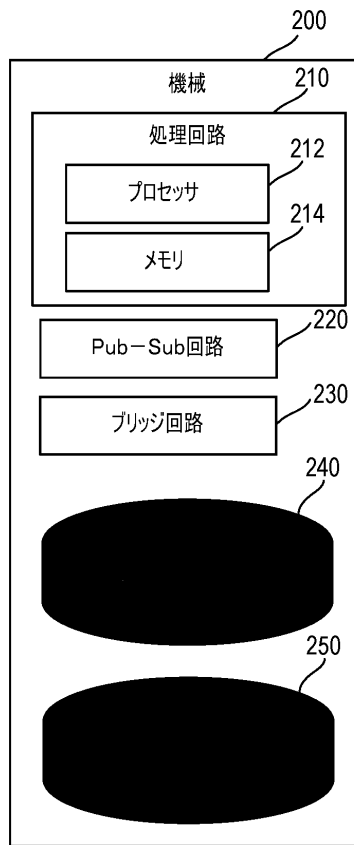
システム。

【図面】

【図1】



【図2】



10

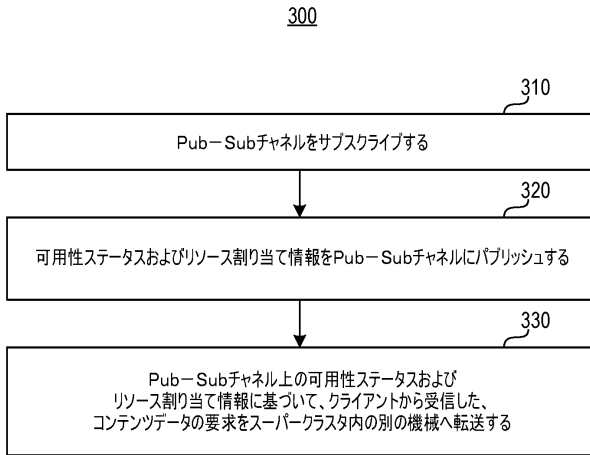
20

30

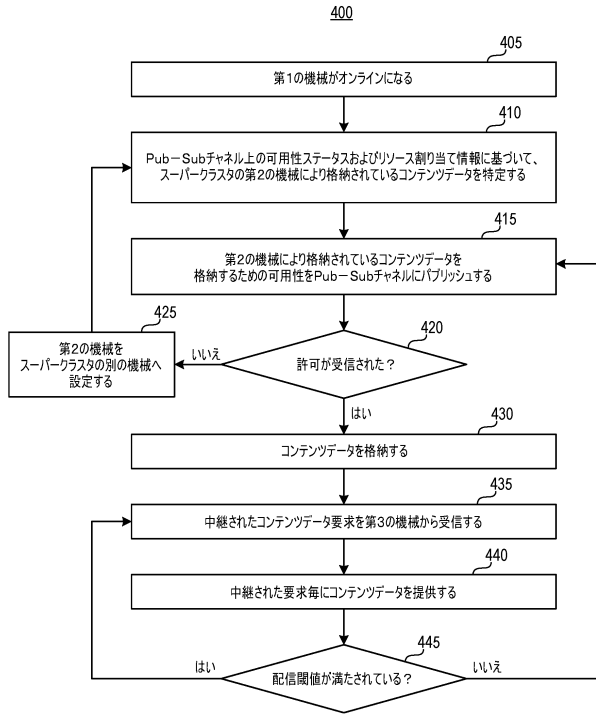
40

50

【 図 3 】



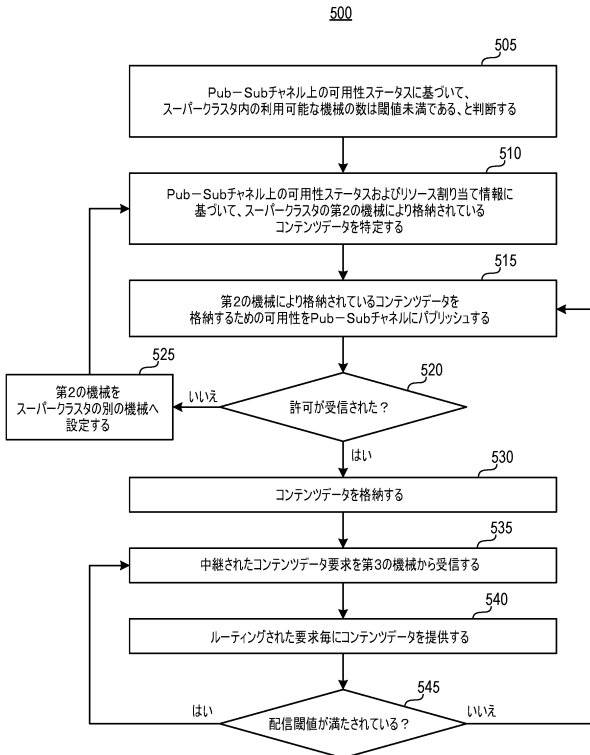
【 図 4 】



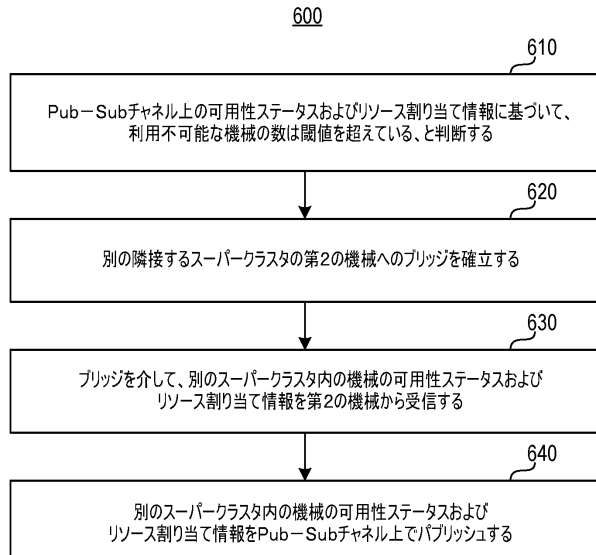
10

20

【 図 5 】



【 図 6 】



30

40

50

---

フロントページの続き

- (56)参考文献 特開 2 0 0 6 - 1 7 1 8 2 2 ( J P , A )  
特開 2 0 1 1 - 1 1 8 5 9 3 ( J P , A )  
特開 2 0 0 6 - 3 0 1 7 6 9 ( J P , A )  
米国特許出願公開第 2 0 1 3 / 0 1 7 3 8 0 6 ( U S , A 1 )  
特開 2 0 1 2 - 1 6 9 7 8 9 ( J P , A )  
特開 2 0 1 3 - 0 5 8 0 5 6 ( J P , A )  
特表 2 0 0 5 - 5 3 7 6 8 7 ( J P , A )  
韓国登録特許第 1 7 7 3 7 1 6 ( K R , B 1 )  
米国特許出願公開第 2 0 1 2 / 0 3 1 7 3 0 7 ( U S , A 1 )
- (58)調査した分野 (Int.Cl. , D B 名)  
H 0 4 L 6 7 / 0 0 - 6 7 / 7 5