

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6282728号
(P6282728)

(45) 発行日 平成30年2月21日 (2018. 2. 21)

(24) 登録日 平成30年2月2日 (2018. 2. 2)

(51) Int. Cl.	F I
G 0 6 F 12/00 (2006. 01)	G O 6 F 12/00 5 1 1 A
G 0 6 F 9/445 (2018. 01)	G O 6 F 9/06 6 5 O D
G 0 6 F 13/00 (2006. 01)	G O 6 F 13/00 5 4 O A

請求項の数 14 (全 32 頁)

(21) 出願番号	特願2016-515098 (P2016-515098)	(73) 特許権者	507303550
(86) (22) 出願日	平成26年5月22日 (2014. 5. 22)		アマゾン・テクノロジーズ・インコーポレ
(65) 公表番号	特表2016-520930 (P2016-520930A)		ーテッド
(43) 公表日	平成28年7月14日 (2016. 7. 14)		アメリカ合衆国・9 8 1 0 8 - 1 2 2 6 ・
(86) 国際出願番号	PCT/US2014/039209		ワシントン州・シアトル・ピーオーボック
(87) 国際公開番号	W02014/190190		ス・8 1 2 2 6
(87) 国際公開日	平成26年11月27日 (2014. 11. 27)	(74) 代理人	100098394
審査請求日	平成27年12月25日 (2015. 12. 25)		弁理士 山川 茂樹
(31) 優先権主張番号	13/900, 350	(74) 代理人	100153006
(32) 優先日	平成25年5月22日 (2013. 5. 22)		弁理士 小池 勇三
(33) 優先権主張国	米国 (US)	(74) 代理人	100064621
前置審査			弁理士 山川 政樹

最終頁に続く

(54) 【発明の名称】 サービスとしての効率的なデータ圧縮及び分析

(57) 【特許請求の範囲】

【請求項 1】

圧縮サービスを実装するよう構成される 1 つ以上のコンピューティングデバイスを備えるシステムであって、前記圧縮サービスが、

複数の圧縮エンジンであって、各々が、複数の圧縮技法の中から、少なくとも 1 つの圧縮技法を実行するように構成される、複数の圧縮エンジンと、

ルールベースの圧縮エンジンセレクトアであって、

クライアントから圧縮されるデータを受信し、

前記データの受信に応答して、

圧縮されるデータまたは前記データと関連付けられるメタデータにルールベースの分析を実行して、前記複数の圧縮技法の中から前記データに適用される 2 つ以上の圧縮技法を選択し、かつ

前記複数の圧縮エンジンのうちの 1 つ以上に、前記選択された 2 つ以上の圧縮技法のそれぞれに対して、1 つ以上のデータ圧縮候補を、前記選択された 2 つ以上の圧縮技法に従って、所与の圧縮サービス制限を超過するまで生成するよう指示する、ように構成される、圧縮エンジンセレクトアと、

応答生成モジュールであって、

圧縮選択基準に従って、生成された 2 つ以上の前記データ圧縮候補のうちの 1 つを選択して、得られた圧縮されたデータとして提供するように構成される、応答生成モジュールと、を備える、前記システム。

10

20

【請求項 2】

圧縮される前記データまたは前記データと関連付けられる前記メタデータにルールベースの分析を実行して、前記複数の圧縮技法の中から前記データに適用される前記2つ以上の圧縮技法を選択するために、前記ルールベースの圧縮エンジンセレクトが、最新の圧縮選択ルールのセットを、前記データの1つ以上のデータ特性に適用するように構成され、前記圧縮サービスが、

1つ以上の機械学習技法を過去の圧縮データに実行して、前記最新の圧縮選択ルールのセットを更新するように構成される、機械学習圧縮分析モジュールをさらに備える、請求項1に記載の前記システム。

【請求項 3】

前記応答生成モジュールが、

前記2つ以上の前記データ圧縮候補のうちの前記選択された1つのエントロピー測定値を決定し、かつ

前記2つ以上の前記データ圧縮候補のうちの前記選択された1つの前記エントロピー測定値がエントロピー閾値未満であると決定することに応答して、システム圧縮技法に従って、前記選択された1つを圧縮して、前記要求された圧縮されたデータとして送信される多段圧縮されたデータを生成するようにさらに構成される、請求項1に記載の前記システム。

【請求項 4】

1つ以上のコンピューティングデバイスによって、

クライアントから圧縮されるデータを示す圧縮要求を受信することと、

前記圧縮要求の受信に応答して、

圧縮される前記データまたは前記データと関連付けられるメタデータに分析を実行して、複数の圧縮技法の中から前記データに適用される2つ以上の圧縮技法を選択することと、

前記選択された2つ以上の圧縮技法のそれぞれに対して、及び、所与の圧縮サービス制限を超過するまで、前記選択された2つ以上の圧縮技法に従って、1つ以上のデータ圧縮候補を生成することと、

圧縮選択基準に従って、2つ以上の前記データ圧縮候補のうちの1つを選択して、要求された圧縮されたデータとして送信することと、

前記要求された圧縮されたデータを含む応答を送信することと、を実行することを含む、方法。

【請求項 5】

圧縮される前記データまたは前記データと関連付けられるメタデータの前記分析がルールベースの分析である、請求項4に記載の前記方法。

【請求項 6】

前記方法が、

1つ以上の機械学習技法を過去の圧縮データに実行して、前記2つ以上の圧縮技法を選択するために適用される最新の圧縮選択ルールのセットを更新することをさらに含み、

圧縮される前記データまたは前記データと関連付けられる前記メタデータに前記分析を実行して、前記複数の圧縮技法の中から前記データに適用される前記2つ以上の圧縮技法を選択することが、

前記更新された圧縮選択ルールのセットを、前記データの1つ以上のデータ特性に適用して、前記複数の圧縮技法の中から前記データに適用される前記2つ以上の圧縮技法を選択することを含む、請求項5に記載の前記方法。

【請求項 7】

圧縮される複数の他のデータを、複数のクライアントから受信することと、

前記複数の他のデータの各々について、

前記ルールベースの分析、前記生成、前記選択、及び前記送信を実行することと、

前記他のデータののための圧縮結果データ及びデータ特性を、前記過去の圧縮データの

10

20

30

40

50

部として記憶することと、をさらに含む、請求項 6 に記載の前記方法。

【請求項 8】

前記要求された圧縮されたデータを含む前記応答を送信する前に、システム圧縮技法に従って、前記選択されたデータ圧縮候補を圧縮して、前記要求された圧縮されたデータとして送信される多段圧縮されたデータを生成することをさらに含む、請求項 4 に記載の前記方法。

【請求項 9】

圧縮される前記データが、複数のデータチャンクを含むデータストリームであり、前記実行、前記生成、及び前記選択は、圧縮される前記複数のデータチャンクのうちの圧縮される第 1 のデータチャンクについて実行され、前記方法が、

前記複数のデータチャンクのうちの後続データチャンクの各々について、

前記第 1 のデータチャンクを圧縮するために適用される前記 1 つ以上のデータ圧縮技法に従って、圧縮されたデータチャンクを生成することと、

前記データチャンクを含む応答を送信することと、をさらに含む、請求項 4 に記載の前記方法。

【請求項 10】

前記要求された圧縮されたデータを含む応答を送信する前に、1 つ以上の暗号化技法に従って、前記選択されたデータ圧縮候補を暗号化することをさらに含む、請求項 4 に記載の前記方法。

【請求項 11】

ネットワークベースのサービスを実装するように構成される複数のコンピューティングデバイスを備えるシステムであって、前記ネットワークベースのサービスが、

2 つ以上の圧縮技法に従ってデータを圧縮するように構成されるデータ圧縮サービスモジュールと、

ネットワークベースのサービスインターフェースであって、

クライアントから圧縮されるデータを示す圧縮要求を受信し、

前記圧縮要求の受信に応答して、

前記圧縮要求と関連付けられる料金体系を決定し、

前記 2 つ以上の圧縮技法のそれぞれに対して、前記圧縮要求と関連付けられる前記料金体系を超過するまで、前記 2 つ以上の圧縮技法のそれぞれに従って、1 つ以上の圧縮データ候補を生成するよう指示し、

前記料金体系に従って、前記要求された圧縮されたデータの料金を発生させ、かつ

前記圧縮要求に従って、前記要求された圧縮されたデータを送信するように構成される、ネットワークベースのサービスインターフェースと、を備える、前記システム。

【請求項 12】

前記料金体系が前記圧縮要求のための前記圧縮サービス制限を示し、前記要求された圧縮されたデータを生成するために、前記データ圧縮サービスモジュールが、前記圧縮サービス制限内で前記要求された圧縮されたデータを生成するように構成される、請求項 11 に記載の前記システム。

【請求項 13】

前記圧縮要求が、前記示されたデータに適用される 1 つ以上のクライアント選択の圧縮技法をさらに示し、前記要求された圧縮されたデータを生成するために、前記データ圧縮サービスモジュールが、前記 1 つ以上のクライアント選択の圧縮技法を実行して、前記要求された圧縮されたデータを生成するように構成される、請求項 11 に記載の前記システム。

【請求項 14】

前記圧縮要求が、前記圧縮されるデータの圧縮分析をさらに要求し、前記 2 つ以上の圧縮技法のそれぞれに対して、前記 2 つ以上の圧縮技法のそれぞれに従って、1 つ以上の圧縮データ候補を生成するために、前記データ圧縮サービスモジュールが、

圧縮される前記データまたは前記データと関連付けられるメタデータに分析を実行して

10

20

30

40

50

、複数の圧縮技法の中から前記データに適用される2つ以上の圧縮技法を選択し、前記2つ以上の圧縮技法のそれぞれに対して、前記選択された2つ以上の圧縮技法に従って、1つ以上の圧縮データ候補を生成し、かつ

圧縮選択基準に従って、前記1つ以上の生成された圧縮データ候補のうちの1つを、前記要求された圧縮されたデータとして選択するように構成される、請求項11に記載の前記システム。

【発明の詳細な説明】

【背景技術】

【0001】

組織がデータを作成し、追跡し、保持する技術力が向上し続けるにつれ、増大し続ける情報を発信し、かつ記憶するための様々な異なる技術が開発されている。そのような技術の1つであるデータ圧縮によって、データを異なる様式で表すことにより、データサイズを縮小することができる。後に、データはさらなる処理のために復元することができる。多くの異なる圧縮技法に従って、多くの異なる型のデータを圧縮することができる。どの圧縮技法を適用するかを決定することが難しい場合が多い。いくつかの技法は、より効果的であり、いくつかのデータにデータサイズのさらなる縮小がもたらされ、一方、他のデータでは、異なる圧縮技法がより適する場合がある。データに適用する圧縮技法を選択する能力は、データの圧縮を所望するエンティティの制約内で、様々な分析をデータに実行するのに必要なリソースによって限定される場合が多い。同様に、異なる圧縮技法によって、圧縮技法を実行するために異なる負担が課される。したがって、データの圧縮を所望するエンティティは、時間、運用コスト、及び他の圧縮選択基準ソースの限定によって、データを効率的に圧縮することが限られてしまうことが多い。

【図面の簡単な説明】

【0002】

【図1】いくつかの実施形態による、サービスとしての効率的なデータ圧縮及び分析のデータフローブロック図を示す。

【図2】いくつかの実施形態による、データ圧縮サービスの例示的な動作環境を示すブロック図である。

【図3】いくつかの実施形態による、例示的なデータ圧縮サービスを示すブロック図である。

【図4】いくつかの実施形態による、効率的なデータ圧縮及び分析をサービスとして実行するための方法の高レベルフローチャートである。

【図5】いくつかの実施形態による、選択された一連の圧縮技法による、かつ所与の圧縮サービス制限内の、1つ以上のデータ圧縮候補を生成するための方法の高レベルフローチャートである。

【図6】いくつかの実施形態による、機械学習を実行して、圧縮されるデータのルールベースの分析を更新するための方法を示す、高レベルフローチャートである。

【図7】いくつかの実施形態による、多段圧縮を含むサービスとして、効率的なデータ圧縮及び分析を実行するための方法を示す、高レベルフローチャートである。

【図8】いくつかの実施形態による、データ圧縮をサービスとして実行するための方法を示す、高レベルフローチャートである。

【図9】いくつかの実施形態による、例示的なシステムを示す。

【0003】

実施形態は、いくつかの実施形態の実施例、及び例示的な図面によって本明細書中で説明されるが、実施形態がそれらの説明される実施形態及び図面に限定されるものではないことを、当業者は認識するであろう。図面及びその詳細な説明は、実施形態を開示される特定の形態に限定することを意図するものではないが、反対に、その意図は、添付の特許請求範囲によって定義される精神及び範囲内に含まれる全ての修正形態、同等物、及び代替物を網羅することである旨を理解されたい。本明細書中で使用される見出しは、単に組織的目的のものであり、説明の範囲または特許請求の範囲を限定するために使用されるこ

10

20

30

40

50

とを意味しない。本出願全体で使用されるとき、単語「することができる (may)」は、義務的な意味 (すなわち、～しなければならない、を意味する) ではなく、むしろ広範な意味 (すなわち、～する可能性を有する) で使用される。同様に、単語「含む (include)」、「含んでいる (including)」、及び「含む (includes)」は、含んでいることを意味するが、それらに限定されない。

【発明を実施するための形態】

【0004】

以下の詳細な説明では、特許請求された主題について完全な理解をもたらすために、特定の詳細が多数示されている。しかしながら、特許請求された主題は、それら特定の詳細なしでも実行可能であることが、当業者によって理解されよう。他の例では、特許請求された主題が曖昧にならないように、当業者によって知られている方法、装置、またはシステムは、詳細に説明されていない。

【0005】

また、用語、第1の、及び第2の等を本明細書中で使用して、様々な要素について説明することができるが、これらの要素は、それらの用語によって限定されるべきではないということも理解されたい。それらの用語は、1つの要素を別のものと区別するためにのみ使用される。例えば、本発明の範囲から逸脱することがなく、第1の接点を第2の接点と称することができ、同様に、第2の接点を第1の接点と称することもできる。第1の接点及び第2の接点はいずれも接点であるが、それらは同じ接点ではない。

【0006】

本発明の明細書中の説明で使用される専門用語は、特定の実施形態について説明することのみが目的であり、本発明を限定することを意図するものではない。本発明及び添付の特許請求の範囲の説明で使用されるとき、単数形「1つの (a)」、「1つの (an)」、及び「その (the)」は、文脈上で別途明確に示されない限り、複数形も含むことを意図する。また、用語「及び/または」は、本明細書中で使用されるとき、列挙される関連物品の1つ以上の、任意のまたは全ての可能な組み合わせを指し、かつ包含することとも理解されたい。さらに、用語「含む (includes)」、「含んでいる (including)」、「備える (comprises)」、及び/または「備えている (comprising)」は、本明細書中で使用されるとき、記載される特性、整数、ステップ、動作、要素、及び/または構成要素の存在を特定するが、1つ以上の他の特性、整数、ステップ、動作、要素、構成要素、及び/またはそれらの群の存在または追加を除外するものではない、ということを理解されたい。

【0007】

本明細書中で使用されるとき、用語「もし (if)」は、文脈に応じて、「とき (when)」、または「後 (upon)」、もしくは「決定にตอบสนองして (in response to determining)」、あるいは「検出にตอบสนองして (in response to detecting)」を意味すると解釈することができる。同様に、慣用句「であることが決定される場合 (if it is determined)」、または「(記載される条件または事象) が検出される場合 (if [a stated condition or event] is detected)」は、文脈に応じて、「決定した後に (upon determining)」、「決定にตอบสนองして (in response to determining)」、または「(記載された条件または事象) を検出した後に (upon response to detecting [the stated condition or event])」、または「(記載された条件または事象) の検出にตอบสนองして (in response to detecting [the stated condition or event])」を意味すると解釈することができる。

【0008】

サービスとしての効率的なデータ圧縮及び分析の様々な実施形態が本明細書中で説明される。圧縮要求は、圧縮されるデータを含むクライアント (例えば、クライアントシステ

10

20

30

40

50

ム、サービス、デバイス、及びユーザー等)から受信することができる。圧縮されるデータ、またはそのデータと関連付けられるメタデータにルールベースの分析を実行して、複数の圧縮技法の中から、そのデータに適用される1つ以上の圧縮技法を選択することができる。選択される圧縮技法に従って、データ圧縮候補を生成することができる。圧縮データオブジェクトの生成を抑制する圧縮サービス制限を施行することができる。次いで、圧縮選択基準に従って、データ圧縮候補のうちの1つを選択して、要求された圧縮されたデータとして送信することができる。

【0009】

本明細書はまず、効率的なデータ圧縮及び分析の実施例について説明する。様々なウェブサービスの一部として実装される圧縮サービス等の圧縮サービスは、効率的なデータ圧縮及び分析を実装するように構成することができる。例示的な圧縮サービスの説明には、圧縮サービス及び様々な他のサービスの様々な態様が含まれており、それらは、データベースサービス等の圧縮サービスと通信することができる。次いで、本明細書は、サービスとしての効率的なデータ圧縮及び分析の方法の、様々な実施形態のフローチャートについて説明する。次いで、本明細書は、開示された技法を実装することができる例示的なシステムについて説明する。本明細書全体で、様々な異なる実施例を提供することができる。

【0010】

多くの異なる理由から、データを作成、生成、発信、管理、修正、記憶、または、さもないければ、操作することができる。それらのエンティティ(例えば、消費者、組織、ユーザー、クライアント、及びシステム等)がそのデータを圧縮して、そのデータをより効率的に記憶、伝送、または、さもないければ、管理することは珍しいことではない。報道機関は、例えば、オーディオまたはビジュアルファイルを作成して、消費者に分配することができる。それらのオーディオまたはビジュアルファイルの圧縮版を消費者に送信することができるが、それらは、より小さくなるほど、より容易にかつ迅速に伝送することができる(例えば、回線容量をほとんど消費しない)。データ記憶機構は、多くの異なる記憶クライアントのための大量のデータを保持することができる。記憶されたデータの安全性及び信頼性を向上させるために、データの複数の複製を保持することができる。それらのデータの複数の複製を圧縮形式で記憶することによって記憶領域はほとんど消費されず、それによって、より信頼性の高い記憶を提供するためのコストを低減させることが可能になる。好ましいデータ圧縮を明示する多数の他の実施例についても考察することができるので、上述の実施例は限定することを意図するものではない。

【0011】

しかしながら、データの量が増大するにつれ、及び/またはデータを圧縮するための様々な技法が向上するにつれ、最も効果的な圧縮の形式をデータに実行することが次第に困難になり得る。例えば、上述の報道機関について考察する。データを受信かつ消費することができる様々な技術が拡大するにつれ、データを記憶するために使用されるデータフォーマット(例えば、mp3、jpeg、及びmpeg等のファイルフォーマット)の数が増加し続けている。適切な圧縮技法を選択してデータに適用する負担はより複雑なものになり、より多くのリソースが必要とされる。同様に、上述の例示的なデータ記憶機構は、様々な異なる種類の圧縮の影響を受けやすい、多くの異なる型のデータを受信することができる。新規のまたは珍しい型の圧縮されるデータを受信する場合等では、最も効率的な圧縮技法を選択して適用することが困難であると証明され得る。それらのシナリオ及び他のものにおいては、圧縮されたデータを実現するための代替の方法及び技法が有利である。

【0012】

図1は、いくつかの実施形態による、効率的なデータ圧縮及び分析のデータフローブロック図を示す。データ圧縮サービス100は、図1中のコンピュータシステム1000に関して以下に記載される、分散システム、または任意の他の型のコンピューティングシステムもしくはデバイスの1つ以上のノード等の、1つ以上のコンピューティングシステムまたはデバイスによって実装可能である。データ圧縮サービスは、様々な異なるシステム

、デバイス、またはクライアントから、データ102を含む圧縮要求104を受信するように構成され得る。それらのクライアントは、データ圧縮サービス100を制御する同じエンティティによって制御される他のサービスまたはシステムであってよい、またはいくつかのサードパーティーによって制御されてもよい。

【0013】

データ102は、データ圧縮サービス100で受信可能な、任意の型のデータであってよい。例えば、データ圧縮サービス100がネットワーク接続全体で圧縮クライアントと通信するように構成される場合、データ圧縮サービス100は、そのようなネットワーク接続全体に発信され得る、任意の型のデータを圧縮することができる。データを1つ以上の個別のチャンク、パケット、または他の部分に分割して、それらを、全体として、または個別の部分内でのいずれかで処理することができる。いくつかの実施形態では、データ102は、1つ以上の圧縮技法に従って既に圧縮されているデータであってよい。

10

【0014】

データ102に適用される複数の異なる圧縮技法の中からいくつかの圧縮技法を選択するために、圧縮分析110が存在し得る。データ102を分析することによって、圧縮分析を実行することができる。分析は、またはデータの可能値の特定のサイズドメインまたは通常の反復値について調査することによってデータ型、フォーマット、及びサイズ等のデータの様々な特性を決定することを含み得る。多くの異なる型のデータ分析が実行可能であるため、以前の実施例は限定することを意図するものではない。

【0015】

20

また、圧縮されるデータについて記載するメタデータを分析することによって、圧縮分析110を適用することもできる。また、いくつかの実施形態に従って、圧縮されるデータと関連付けられるメタデータを得ることもできる。例えば、メタデータは、データの発信元または宛先（例えば、クライアント）、及びデータが受信された時間（例えば、タイムスタンプ）について記載し、またはより一般的には、データの型または特性を示す。このメタデータは、圧縮サービス100で受信されるときに圧縮要求またはデータとともに含まれる圧縮要求または他の情報フォーマット内に、データヘッダとして含まれ得る。また、識別されたクライアントから受信したデータと関連付けられるメタデータを記憶する特定のクライアント識別子等の、特定のデータ型またはクライアントと関連付けられる圧縮サービスにおいて、メタデータを記憶することもできる。メタデータは、クライアントによって生成される、またはクライアントから受信する記述的情報であってよい。例えば、メタデータは、データがウェブサイトの1日あたりのトラフィック値であり得ると認識することができる。

30

【0016】

メタデータ及び/またはデータのルールベースの分析、または他のいくつかの動的分析技法は、様々な実施形態で実行可能である。ルールベースの分析は、圧縮サービスのための知識ベースを表すルールセットを含むことができる。それらのルールは、様々な型のデータのエン트로ピー測定値、及び記録されたエン트로ピー測定値を実現するためのデータに適用される1つ以上の圧縮技法等の、以前に圧縮されたデータと関連付けられる過去のデータに基づいて決定され得る。ルールベースの分析は、データ特性を決定することを含み、そのデータ特性を用いて、圧縮されるデータを評価し、次いで、データ及びメタデータの以前のデータ分析によってより早い段階で得た情報等の、データに関する既知の情報に適用される圧縮選択ルールのセットを識別して、データに適用される1つ以上の圧縮技法を選択することができる。少なくともいくつかの実施形態では、それらの圧縮技法を特定の順序で順序付けすることができる。いくつかの実施形態では、各々が異なる圧縮技法を適用してデータ圧縮候補を生成するように機能される複数のノードまたはコンピューティングデバイス等によって、圧縮技法を並行して適用することができる。

40

【0017】

少なくともいくつかの実施形態では、機械学習を圧縮分析110の一部として適用させて、データ圧縮サービス100の知識ベースを更新することができる。例えば、管理学習

50

技法は、圧縮されるデータのデータ特性に類似する１つ以上のデータ特性を有する過去の圧縮されたデータに、管理学習技法を適用することができる。データのルールベースの分析において圧縮されるデータに適用される最新の圧縮選択ルールのセットを、この機械学習技法の結果として、更新、修正、または変更することができる。例えば、特定の圧縮技法には、以前に適用された未修正のルールのセット下でもたらされたデータの優先度よりも高いデータの優先度がもたらされ得る。

【 0 0 1 8 】

データ圧縮候補 1 2 0 を生成するための１つ以上の圧縮技法を実行するために、複数の異なる圧縮エンジン 1 3 0 を構成することができる。それらの圧縮エンジンは、選択された１つ以上の圧縮技法に従って、圧縮されるデータを圧縮することができる。少なくともいくつかの実施形態では、所与の圧縮サービス制限による圧縮技法の特定の順番に従ってデータを圧縮するように圧縮エンジン 1 3 0 を構成することができる。圧縮サービス制限によって、時間及びコストが限定され得る、または、候補圧縮データオブジェクトの生成に対して、いくつかの他の抑制がもたらされ得る。例えば、選択される１つ以上の圧縮技法には優先順位がもたらされ得て、最も効率的に機能する可能性が高いそれらの圧縮技法が、圧縮分析 1 1 0 によって決定された、効率性が少ない圧縮技法の前に順序付けされる。したがって、所与の圧縮サービス制限が時間制限を適用する場合、最も効率的な圧縮技法に従って、時間制限内にデータ圧縮候補を生成することができる一方で、効率的に圧縮されたデータを生成する可能性が低い圧縮技法を所与の時間内で実行してもよい、または実行しなくてもよい。別の実施例では、サービス料金または他のいくつかのコストを、データオブジェクトを圧縮する圧縮サービスによって実行されたワークに割り当てることができる。クライアントは、サービス料金の上限をデータの圧縮に適用することのみを要求することができる。それら２つの実施例は、施行可能な多くの異なる型のサービス制限のうちの単にいくつかにすぎないため、以前の実施例は限定することを意図するものではない。

【 0 0 1 9 】

様々な実施形態では、圧縮選択基準に従って、データ圧縮候補 1 2 0 のうちの１つを選択することができる。例えば、いくつかの実施形態では、圧縮選択基準は、データ圧縮候補のサイズ等の単一測定に基づいてもよい。図 1 に示されるように、データ 1 0 2 のデータ圧縮候補 1 2 0 は、候補 1 2 2、1 2 4、及び 1 2 6 を含む。圧縮選択基準がサイズである場合、図 1 に示されるように、データ圧縮候補 1 2 4 を送信される要求された圧縮されたデータとして選択することができる。データ圧縮候補を生成するために必要な時間、またはデータ圧縮候補を生成するためのリソースの数等の他の測定値は、選択に基づき得る圧縮選択基準の構成要素として示すことができる。選択されたデータ圧縮候補を、要求された圧縮されたデータとして、圧縮要求を提出した元のクライアント、または別のリモートシステムしくはエンティティ等の様々な異なる位置に送信することができる。いくつかの実施形態では、応答 1 4 2 によって、例えば、選択された圧縮されたデータオブジェクト 1 2 4（または、圧縮されたデータオブジェクト 1 2 4 の位置）が示され得る。

【 0 0 2 0 】

効率的なデータ圧縮及び分析の実施形態は、様々な異なるサービスまたはシステムの一部として実装可能である。また、データベースサービス等のデータ管理サービスは、クライアントに提供されて、データ管理サービス内に記憶されるクライアントデータのためのデータ圧縮サービスを実装することもできる。データ圧縮サービスは、より広範なウェブまたはネットワークベースのサービスのセットの一部であってよく、内部ならびに外部クライアント及び／または消費者に提供され、もしくは、代替として、スタンドアロンサービスとして提供される。いくつかの実施形態では、ネットワークベースのサービスは、１つ以上の料金体系、サービスプラン、または性能オプションを、データ圧縮サービスを利用するクライアントに提供することができる。それらの料金体系、サービスプラン、または性能オプションは、クライアントに提供される、より大きなネットワークベースのサービスのセットのためのもの、またはその一部であってもよい。より一般的には、データを

クライアントから受信する、ならびにクライアントに発信する任意のシステム、及び/または他のサービス、システム、もしくは消費者は、効率的なデータ圧縮及び分析の様々な実施形態をサービスとして実装することができ、したがって、以前の実施例は、想定される様々な他のシステムを限定するものである必要がない。

【0021】

データ圧縮サービスのネットワークベースのサービスとしての実装

以上に考察されるように、様々なクライアント（または消費者、組織、エンティティ、もしくはユーザー）は、データ圧縮サービスを用いたデータの圧縮を所望することができる。図2は、いくつかの実施形態によるデータ圧縮サービスをクライアントに提供することができる、例示的な動作環境を示す。クライアントは、ネットワークベースのサービスプラットフォームの一部として提供されるデータ圧縮サービスと、ネットワーク上で通信することができる。また、ネットワークベースのサービスプラットフォームの一部として実装される他のサービスは、データ圧縮サービスからのサービスと通信することもできる、及び/またはそれを得ることもできる。

【0022】

複数のユーザーまたはクライアントは、圧縮されるデータを、データ圧縮サービスに送信することができる。いくつかの実施形態では、クライアント250a~250nは、様々なクライアントシステム、ユーザー、クライアントアプリケーション、及び/またはデータネットワークベースのサービス加入者を含むことができる。例えば、クライアントシステムは、コンテンツプロバイダー、またはデータ管理、もしくは記憶サービスを含んでもよい。このサービスは、コンテンツをクライアントまたは他のサービスもしくはプロセスに提供する前に、提供されたコンテンツまたは記憶されたデータを、データ管理サービスを介して転送するように構成されるシステムもしくは構成要素を含んでもよい。

【0023】

クライアント250a~250n等のクライアントは、デスクトップコンピュータ、ラップトップコンピュータ、タブレットコンピュータ、パーソナルデジタルアシスタント、モバイルデバイス、サーバー、または任意の他のコンピューティングシステム、もしくは図9に関して以下に記載される圧縮されるデータとともに、要求をデータ圧縮サービス220に送信する、及び/またはデータ圧縮サービス220からの圧縮されたデータ等の応答を受信するように構成されるコンピュータシステム1000等の他のデバイスを介して、データ圧縮サービス220と通信することができる。例えば、要求を、パラメーター、及び/またはデータ圧縮サービス220によって圧縮されるデータと関連付けられるメタデータを含むメッセージとしてフォーマットすることができる。Extensible Markup Language (XML) 等の特定のマークアップ言語に従って、そのようなメッセージをフォーマットすることができる、及び/またはSimple Object Access Protocol (SOAP) 等のプロトコルを使用してカプセル化することができる。アプリケーションプログラマーインターフェース (API) を実装して、標準化されたメッセージフォーマットを、クライアントがデータ圧縮サービス220と通信しているときなどに、クライアントに提供することが可能である。

【0024】

少なくともいくつかの実施形態では、クライアント250a~250nは、ネットワークベースのサービスプラットフォーム200との通信の一部として、データ圧縮サービス220と通信することができる。ネットワークベースのサービスプラットフォーム200は、1つ以上のサービスをクライアント250a~250nに提供することができ、そのサービスには、データ圧縮サービス220、及び様々なホスティング、記憶、計算、ならびに他のサービスを含み得る他のサービス230が含まれる。少なくともいくつかの実施形態では、ネットワークベースのサービスは、クラウドベースのサービスを含むことができる。ネットワークベースのサービスプラットフォーム200は、様々なクライアントアカウントを追跡する、または記憶することができ、それによって、クライアントの様々な料金、料金体系、使用記録、及びクライアントがデータ圧縮サービス220等のネットワ

10

20

30

40

50

ークベースのサービスと交信することに関する他の情報、ならびに他のサービス230を保持することができる。少なくともいくつかの実施形態では、ネットワークベースのサービスプラットフォーム200は、クライアント250a~250n間のインターフェースとして動作することができ、一方、他の実施形態では、クライアント250a~250nは、各々のサービスと直接通信することができる。

【0025】

クライアント250a~250nは、ネットワーク260上の方法等の様々な異なる通信方法を用いて、データ圧縮サービス220、他のサービス230、またはネットワークベースのサービスプラットフォーム200と通信することができる。ネットワーク260は、広域ネットワーク(Wide Area Network)(WAN)(例えば、インターネット)であってよい。しかしながら、プライベートネットワーク、イントラネット、及び通信技術の他の形式もまた、クライアントとデータ圧縮サービス220との間の通信を容易にすることができる。例えば、他のサービス230は、データ圧縮サービス220を利用して、その独自のサービスに関するデータを圧縮、分析、または復元することができ、かつ、プライベートネットワークまたは通信チャネルを利用することができる。いくつかの実施形態では、クライアントは、圧縮要求を含むメッセージをアセンブルして、そのメッセージを、ネットワークエンドポイント(例えば、データ圧縮サービス230に対応するユニフォームリソースロケータ(Uniform Resource Locator)(URL))に伝達することができる。例えば、クライアント250aは、ウェブクライアント等のローカルソフトウェアアプリケーションを起動させるデスクトップコンピュータを介して通信することができる、つまり、ネットワーク260を介して、ハイパーテキストトランスファープロトコル(HTTP)要求をデータ圧縮サービス230に送信するように構成される。同様の方式で、クライアントに送信される応答または他のデータをフォーマットすることができる。

【0026】

上述のように、少なくともいくつかの実施形態では、ネットワークベースのサービスプラットフォーム200は、データ圧縮サービス220及び/または他のサービス230のインターフェースとして動作することができる。以下でさらに詳細に考察される図8は、いくつかの実施形態によると、データ圧縮サービスをネットワークベースのサービスとして提供するための様々な方法及び技法について説明する。いくつかの実施形態では、クライアントから圧縮されるデータを示す圧縮要求を、ネットワークベースのサービスプラットフォーム200で受信することができる。ネットワークベースのサービスプラットフォーム200は、要求と関連付けられる料金体系を決定することができる。ネットワークベースのサービスプラットフォーム200は、決定された料金体系に従って、データ圧縮サービス220等のデータ圧縮サービス/モジュールに、圧縮されたデータを生成するように指示することができる。また、料金体系に従って、要求された圧縮されたデータに対して、料金を発生させることもできる。次いで、圧縮要求に従って、要求された圧縮されたデータを送信することができる。

【0027】

図3を参照すると、データ圧縮サービス220を、複数の異なるネットワークベースのサービス(例えば、ウェブサービス)、またはスタンドアロンサービスの一部として実装することができる。1つ以上のサーバー、もしくはコンピューティングシステム1000に関して以下の図9中に記載される任意の他のデバイス等の、1つ以上のコンピューティングシステムまたはデバイスは、データ圧縮サービス220を実装するように構成され得る。データ圧縮サービスは分散様式で実装可能であり、その様式では、分散システムの複数のノードは、データ圧縮サービス220の1つ以上の異なる構成要素を実装することができる。例えば、1つ以上のノードは、データ圧縮エンジン330を実装するように構成され得る。様々なコンピューティングシステムまたはデバイスに加えて、様々な異なる他のハードウェア、ソフトウェア、またはハードウェア構成要素とソフトウェア構成要素との組み合わせを用いて、図3に示される様々な構成要素を実装することができ、したがっ

て、以下の説明は限定することを意図するものではない。

【0028】

様々な異なるクライアントから圧縮されるデータを示す圧縮要求302は、データ圧縮サービス220で受信され得る。いくつかの実施形態では、データは、要求と連動して、要求とともに含まれてもよい、または要求によって示されてもよい。例えば、要求は、別のシステム、サービス、または圧縮されるデータを取得し得る記憶位置を示すことができる。次いで、そのデータ圧縮サービスは、示された圧縮位置からデータを得ることができる。上述のように、様々な異なる方式で、かつ多くの異なるプロトコルに従って、データ圧縮要求をフォーマットすることができる。例えば、APIを使用して、圧縮要求、圧縮されるデータ、データと関連付けられる様々なメタデータ、または圧縮要求に対する圧縮サービス制限をフォーマットすることができる。圧縮されるデータは、多くの異なる型のものであってよく、多くの異なるユーザー、サービス、またはクライアントにとって有用な、またはそれらによって有用となる様々なテキストデータもしくはメディアデータを含むがそれらに限定されず、それらのユーザー、サービス、またはクライアントには、記憶サービス、コンテンツプロバイダー、及び通信サービス（例えば、メッセージまたは通知サービス）等が含まれるがそれらに限定されない。圧縮されるデータ302は、大きなデータオブジェクト、または異なるデータチャンクに分割されるデータであってよい。データ302は、データストリーム、またはいくつかの他のグルーピング、もしくは（例えば、圧縮されるデータを構成する1つ以上のデータパケットで）関連付けられるデータの論理配列として見なすことができる。

10

20

【0029】

いくつかの実施形態では、データ圧縮サービス220は、データ、及び/または圧縮されるデータを圧縮せよとの要求を受信する圧縮サービスインターフェース310を実装することができる。いくつかの実施形態では、圧縮サービスインターフェースは、圧縮要求を分割するように、及び/またはさらなる処理のために、データ圧縮サービス220の1つ以上の他の構成要素に、圧縮要求から得た情報を提供するように構成され得る。例えば、圧縮サービスインターフェースは、圧縮されるデータの型または特性を示す圧縮要求から抽出されたメタデータを圧縮エンジンセクタ320に提供して、ルールベースのまたはいくつかの他の形式の分析を実行することができる。例えば、特定のクライアントアカウントにリンクされるクライアント識別子は、圧縮サービスインターフェース310によって識別可能である。圧縮サービスインターフェース310もまた、特定のクライアントによるサービスの使用を追跡すること、様々な請求もしくは他の料金、コスト、または評価技法を実行すること、及び図2中のネットワークベースのサービスプラットフォーム200の一部として実装される、請求またはアカウント管理サービス等のデータ圧縮サービス220によって利用可能な他のサービスと交信すること等の、様々な異なる他のタスクを実行して、データ圧縮サービス220を実装するように構成され得る。

30

【0030】

様々な実施形態では、データ圧縮サービス220の一部として、圧縮エンジンセクタ320を実装することができる。圧縮エンジンセクタ320は、圧縮されるデータ、またはそのデータと関連付けられるメタデータに1つ以上の分析技法を実行するように構成され得る。例えば、いくつかの実施形態では、圧縮エンジンセクタ320は、データのサブセットまたは部分をサンプリングする、スキャンする、またはレビューして、そのデータの1つ以上のデータ特性を決定することができる。それらのデータの特性は、データ型、フォーマット、サイズ、または可能値もしくは通常の反復値のセットの特定のサイズドメインを含むことができるが、それらに限定されない。データ中のデータ値の変化を示す、エントロピー測定またはいくつかの他の測定を行うことができる。このデータ分析に基づいて、圧縮されるデータの1つ以上のデータ特性を決定することができる。

40

【0031】

また、同様に、圧縮エンジンセクタ320によって様々な技法を利用して、圧縮されるデータと関連付けられるメタデータを得ることができる。例えば、クライアント（例え

50

ば、小売業者、データ管理プロバイダー、及び安全性または暗号化サービス等）について記載するメタデータにリンクされ得るクライアント識別子または他のソース識別子を調査すること等によって、圧縮されるデータの発信元を判別することができる。データが送信または受信される時間のタイムスタンプ等の時間情報、データ値のドメイン等の記述的データ（例えば、男性／女性、50の州、ジップコード、及び年齢幅等）、またはデータもしくはクライアントあるいは圧縮されたデータの受信者（圧縮クライアントであってもよい、または圧縮クライアントではなくてもよい）について記述している任意の他の型のデータ。データそれ自体と同様に、データのデータ特性は、圧縮エンジンセクタ320によって識別可能であり、圧縮エンジン選択及び／または分析の一部として利用可能である。

10

【0032】

圧縮エンジンセクタ320は、複数の圧縮技法の中から、データに適用される1つ以上の圧縮技法を選択するように構成され得る。1つ以上の分析技法を実行することによって、圧縮技法の選択を実装することができる。例えば、この選択を行うために、いくつかの実施形態では、データ、またはそのデータと関連付けられるメタデータ（例えば、データ特性）に、ルールベースの圧縮分析を実行することができる。いくつかの実施形態では、データに対するデータ特性を、メタデータ及び／またはデータそれ自体の一部として識別することができ、それらのデータ特性に圧縮選択ルールのセットを適用して、どの圧縮技法を選択するかを決定することができる。例えば、圧縮選択ルールのセットによって、データ値のドメインは2に限定されることを示すデータ特性のルールを決定することができ、選択する効率的なデータ圧縮技法は、ランレングス圧縮技法であってもよい。他のルールによって、データの異なるデータ特性に基づいて、異なる圧縮技法が選択される。少なくともいくつかの実施形態では、圧縮選択ルールによって、圧縮技法の順序付けが特定の順番で決定され得る。この順番は、ルールベースの分析に基づいて、圧縮効率の可能性によって順序付けすることができる。また、圧縮エンジンセクタ320は、ランダムに選択されたまたは決定された圧縮技法を、選択された圧縮技法のうちの1つとして含み、圧縮されたデータ候補を生成することもできる。ランダムに選択された圧縮技法によって、より優れた圧縮をもたらす可能性がある、または可能性がない異なる技法を試すことなく、ルールベースの分析等の分析が同様の圧縮技法の結果によって過度に影響されるのを防ぐことができる。例えば、少なくともいくつかの実施形態では、データ圧縮候補を過去のデータの一部分として使用して、機械学習を実行することができる。ランダムに選択された圧縮技法を用いて生成される圧縮候補によって、極小値、または、過去のデータから効果的な機械学習を行うことを限定し得る、他の型の局在化因子を抑制することができる。

20

30

【0033】

少なくともいくつかの実施形態では、機械学習圧縮分析モジュール360は、データ圧縮サービス220の一部として実装され得る。機械学習圧縮分析モジュール360は、1つ以上の機械学習技法を過去の圧縮データ350に実行することができる。少なくともいくつかの実施形態では、過去の圧縮されたデータは、以前に受信された圧縮要求のための圧縮技法の結果、及び圧縮されたデータのデータ特性を含むことができる。例えば、同じクライアントもしくはデータの型、及び圧縮されていないもしくは初期データの同じサイズ等の、同様のデータ特性を有する以前の要求が受信される場合、機械学習技法を使用して、同様のデータ特性を有する最新の圧縮要求にとってより良好であった圧縮技法を識別することができる。管理学習、非管理学習、及び／または半管理学習は、過去の圧縮データ350に適用可能な、様々な機械学習技法のいくつかの実施例である。それらの技法のうちの1つ以上を用いて、複数の他のデータ圧縮要求の以前の圧縮と関連付けられる様々な異なるデータ特性または他のデータポイントを分析して、圧縮エンジンセクタ320によって適用される圧縮選択ルールのセットを更新、修正、または変更することができる。例えば、1つ以上の機械学習技法による、クライアントAから受信したデータの型は、クライアントBからのデータの型に類似しており、特定の圧縮技法はクライアントBのデータに非常に効果的であったという観測に基づいて、圧縮技法が適用可能な優先度または

40

50

順序を変更することができる。

【0034】

圧縮エンジンセクタ320は、1つ以上のデータ圧縮エンジン330に、選択された1つ以上の圧縮技法を実行してデータ圧縮候補を生成するよう指示することができる。各々のデータ圧縮エンジンは、バイトディクショナリ、テキスト255、テキスト32k、デルタ、ランレングス、mostly n、ランレングス圧縮、Lempel-Ziv、Lempel-Ziv-Oberhumer、gzip、または、より一般的には、ロスレスもしくはロッシーデータ圧縮技法の任意の他の形式等であるが、それらに限定されない1つ以上のデータ圧縮技法を適用するように構成され得る。それらの圧縮技法のうちの異なる1つに従って、異なる圧縮エンジン330は、データを圧縮することができる、またはデータを圧縮するように構成され得る。いくつかの実施形態では、複数の圧縮技法を適用して、データ圧縮候補を生成することができる。例えば、識別される「最も優れた」圧縮技法をデータに適用し、次いで、二次共通またはシステムレベルの技法を圧縮されたデータに続けて適用して、多段圧縮されたデータを生成することができる。多くの異なる圧縮技法は当業者に周知であり、したがって、以前の実施例は限定することを意図するものではない。データ圧縮エンジンを分散様式で実装することができ、それにより、選択された圧縮技法の各圧縮技法を並行してまたはほぼ並行して適用することができる。いくつかの実施形態では、圧縮技法の特定の順序付けに従って、異なる圧縮技法もまた、連続的に適用され得る、または並行性能のために選択され得る。

10

【0035】

少なくともいくつかの実施形態では、圧縮エンジンセクタ320は、1つ以上のデータ圧縮エンジン330が、特定の順序または順番に従って、1つ以上のデータ圧縮技法を適用するように指示することができる。例えば、圧縮エンジンセクタ320は、圧縮されるデータがある順序でデータ圧縮エンジン330に送信することができ、それによって、特定の順序または順番に従って、圧縮されるデータを生成することができる。代替として、別のロードバランサ、キューマネージャー、ノード、またはいくつかの他の構成要素は、圧縮されるデータを、異なるデータ圧縮エンジンで圧縮するために、異なるキュー内に配置ことができ、それによって、データ圧縮が特定の順番で生じ得る。

20

【0036】

少なくともいくつかの実施形態では、圧縮エンジンセクタ構成要素320、または圧縮サービスインターフェース310もしくは応答生成モジュール330等のいくつかの他の構成要素は、圧縮サービス制限を施行するように構成され得る。圧縮サービス制限は、データ圧縮候補の生成、または所与のデータのために実行される全てのサービス等の、データ圧縮サービス220内の別の機能もしくは構成要素を限定し得るルール、上限、リソース限定、または境界の任意の形式であってよい。例えば、いくつかの実施形態では、圧縮サービス制限は、時間制限または限界であってよい。圧縮エンジンセクタ320は、時間制限内に完了可能なデータ圧縮候補の生成を指示するよう構成可能である。同様に、応答生成構成要素340等のいくつかの他の構成要素は、特定のサービス上限料金内で生成可能なデータ圧縮候補等の、時間制限または他のデータ圧縮制限を満たす、1つ以上のデータ圧縮候補を単に選択するように構成され得る。また、圧縮サービス制限は、大きいデータの圧縮要求または複数の消費者からの複数の要求にわたるその独自のリソースを保存する、またはそのリソースのバランスを保つためにデータ圧縮サービス220によって施行される、1つ以上のルールであってもよい。例えば、受信された圧縮されるデータが特定のサイズを超える場合、特定のリソース限定（例えば、データ圧縮候補を生成するよう指示され得るノードまたはデータ圧縮エンジンの数等）を課すことができる。多くの他の異なる型の制限が想定されるので、以前の実施例は限定することを意図するものではない。

30

40

【0037】

データヘッダ等のインジケータをデータ圧縮候補に追加して、データ圧縮候補を生成するために適用される1つ以上の圧縮技法を識別することができる。例えば、データに適用

50

される圧縮技法がディクショナリ圧縮である場合、ディクショナリ圧縮技法の指標となるべく使用されるデータ値は、データヘッダ内に記憶されて容易に復元され得る。

【 0 0 3 8 】

少なくともいくつかの実施形態では、データ圧縮サービス 220 もまた、応答生成モジュール 340 を実装することができる。応答生成モジュール 340 は、圧縮選択基準に従って、データ圧縮候補のうちの 1 つを選択し、それに応じて、要求された圧縮されたデータとして送信するよう構成可能である。そのデータ圧縮候補は、応答生成モジュールでの入力として受信され得る。データ圧縮候補を生成するために使用される、1 つ以上の圧縮技法の性能を示すデータ圧縮候補の性能特性を決定することができる。例えば、性能特性は、データ圧縮候補を生成するのに 2 . 7 秒要したことを示すことが可能である。他の性能特性は、データ圧縮候補のサイズ、データ圧縮候補を生成するために使用されるリソースの数（例えば、ノードの数）、コスト、またはデータ圧縮候補に割り当てることができるいくつかの他のサービス料金、もしくは復元時間 / コスト（オブジェクトを復元するのに要した時間またはコストの量（例えば、計算コスト））を含むことができるが、それらに限定されない。データ圧縮候補のうちの 1 つを選択するために使用される圧縮選択基準は、その性能特性が圧縮選択基準を最もよく満たすデータ圧縮候補のうちの 1 つであってよい。したがって、性能特性によって示される最小のサイズを有するデータ圧縮候補は、送信される要求されたデータ圧縮として選択され得る。いくつかの実施形態では、複数の性能特性を使用して、どのデータ圧縮候補が圧縮選択基準を最もよく満たすかを判定することができる。例えば、圧縮選択基準は最もコスト効率の良いデータ圧縮であってよい。そのような基準を適用して、コストによって分割されるデータ圧縮候補のサイズを決定し、発生する時間または使用されるリソースの数等の、データ圧縮候補を生成することができる。したがって、本実施例では、小さいが最小ではない圧縮サイズを有していた可能性があるが、より速い圧縮時間を有していた可能性があるデータ圧縮候補は、圧縮されたデータオブジェクトとして選択され得る。同様に、わずかに効果的ではない圧縮技法は、さらにより短い復元時間を有することができ、したがって、送信される圧縮されたデータとして選択可能である場合、データ圧縮候補のサイズもまた、復元時間と併せて分析され得る。

【 0 0 3 9 】

少なくともいくつかの実施形態では、応答生成モジュール 340 は、選択されたデータ圧縮候補を生成して、またはフォーマットして、圧縮されたデータの受信者に伝送することができる。例えば、送信前に、1 つ以上の暗号化技法を、選択されたデータ圧縮候補に適用することができる。以下に記載されるように、システム圧縮技法等の他の圧縮技法は、選択されたデータ候補にも適用することができる。代替として、少なくともいくつかの実施形態では、別の構成要素（図示せず）は、データ暗号化モジュールまたはサービス（例えば、図 2 中の他のサービス 230）を実装することができ、それらに対して、圧縮サービスインターフェース 310 を介して送信する前に、圧縮されたデータが送信されて暗号化される。例えば、圧縮サービスインターフェース 310 は、フォーマットされ、圧縮されたデータを受信して、圧縮されたデータ 304 を受信者に送信することができる。いくつかの実施形態では、圧縮されたデータ 304 の受信者は、圧縮されたデータを要求したクライアントであってよい。代替として、受信者は、圧縮クライアントではなく、別のリモートシステムまたはサービスであってよい。例えば、記憶クライアントはまず、データ圧縮サービス 220 にデータを送信してそのデータを対象の受信体を用いて圧縮し、その対象の受信体は、圧縮されたデータを記憶するためのデータ記憶サービスとなる。

【 0 0 4 0 】

いくつかの実施形態では、応答生成モジュール 340 は、エントロピー測定値、または選択されたデータ圧縮候補のデータ値の変化を示す、いくつかの他の測定を決定することができる。このエントロピー測定値に少なくとも一部基づいて、応答生成モジュール 340 は、システム圧縮技法に従って、選択されたデータ圧縮候補を再度圧縮して、データを

10

20

30

40

50

さらに圧縮することができる。そのようなシステム圧縮技法は、上述の1つ以上の圧縮技法のうちのいずれかであってよい。例えば、エントロピー閾値を実装することができ、それによって、エントロピー閾値よりも低いエントロピー測定値を有する選択されたデータ圧縮候補では、システム圧縮技法を適用して、多段圧縮されたデータを生成することができる。代替として、少なくともいくつかの実施形態では、システム全体の圧縮技法を、選択されたデータ圧縮候補（またはデータ圧縮候補のセット）に適用して、多段圧縮されたデータを生成することができる。

【0041】

いくつかの実施形態では、過去の圧縮データ350を、データ蓄積、または他の形式の記憶デバイスもしくは記憶サービスとして実装して、データの以前の圧縮と関連付けられる情報を追跡する、記録する、または保持することができる。この情報は、データ特性、性能特性、または受信者に送信された圧縮されたデータの圧縮と関連付けられる任意の他のデータを含み得る。いくつかの実施形態では、複数のデータ圧縮候補を生成することができ、得たデータ、及びデータ圧縮候補の生成と関連付けられる他のデータ特性もしくは他の性能特性は、要求された圧縮されたデータとして送信される、選択されたデータ圧縮候補と関連付けられるデータ特性、及び/または性能特性に加えて、過去のデータ圧縮350内に記憶することができる。それらのデータ特性は、データ圧縮サービス220の様々な他の構成要素、例えば、圧縮エンジンセクタ320、データ圧縮エンジン330、応答生成モジュール340、またはデータの圧縮と関連するデータを得る任意の他の構成要素等から得ることができる。例えば、クライアントから受信された以前のデータの特性、適用された圧縮技法、及びデータが送信された受信者を示す過去の圧縮データ内に、クライアント識別子を記憶することができる。したがって、同じクライアントから他のデータが受信されるとき、圧縮エンジンセクタ320は、過去の圧縮データ350内に示されるように、適用する最も優れた技法は、以前に適用された圧縮技法であると判定することができる。

【0042】

少なくともいくつかの実施形態では、復元されるデータを示す復元要求306は、データ圧縮サービスで受信可能である。いくつかの実施形態では、復元されるデータは、要求とともに要求に含まれてもよい、または、復元されるデータを得るための位置、システム、またはサービスを示すこと等によって、要求によって示されてもよい。この復元要求によって、圧縮されたデータを生成するために適用される1つ以上の圧縮技法を識別することができる。圧縮サービスインターフェース310、または他の構成要素のいくつかは、1つ以上の圧縮技法に従ってデータを復元するように構成される1つ以上のデータ復元エンジン370に、圧縮されたデータ306を復元するように指示することができる。次いで、応答生成モジュール340は、復元されたデータ308を受信者に送信することができる。上述のように、受信者は、復元を要求したクライアントと同じリモートシステムまたは位置のものであってもよい、または同じでなくてもよい。

【0043】

ネットワークベースのサービスのセットの一部として提供されるデータ圧縮サービスの文脈において、図2及び3について説明されて示されているが、図2及び3中で示されて説明される様々な構成要素は、データ圧縮をサービスとして提供することを所望し得る他のシステムに容易に適用することができる。さらに、示される構成要素、システム、またはデバイスの構成は、他の可能な構成について限定しない。例えば、図3は、データ圧縮サービスを、1つ以上の機能的構成要素またはモジュールとして示すが、いくつかの実施形態では、それらの構成要素またはモジュールは、様々な方式で、異なるコンピューティングデバイスまたはノード全体に分散され得る。したがって、図2及び3は、データ圧縮サービスの実施形態に関して、限定することを意図するものではない。

【0044】

サービスとしての効率的なデータ圧縮及び分析のワークフロー

以上で考察されるように、データ圧縮サービスは、より一般的には、任意の型のクライ

10

20

30

40

50

アント管理、記憶、またはデータ伝送に効率利点をもたらすことができる。図4は、いくつかの実施形態による、効率的なデータ圧縮及び分析をサービスとして実行するための方法の高レベルフローチャートである。様々な異なるシステム及びデバイスは、以下で説明する様々な方法及び技法を実装することができる。上述の図3中に記載されるデータ圧縮サービス220等のデータ圧縮サービスは、様々な方法及び技法を実装することができる。しかしながら、示される方法を実行するものとして参照される、上述の実施例、及びまたは任意の他のシステムもしくはデバイスは、他の異なる個体、またはシステムならびにデバイスの構成に関して、限定するものではない。

【0045】

410に示されるように、圧縮されるデータを示す圧縮要求は、クライアントから受信されてもよい。このデータは、図1及び3に関する上述のように、多くの異なるフォーマット、型、及びサイズのものであってよい。いくつかの実施形態では、データは、要求と連動して、要求とともに含まれてもよい、または要求によって示されてもよい。例えば、要求は、別のシステム、サービス、または圧縮されるデータを取得し得る記憶位置を示すことができる。次いで、そのデータ圧縮サービスは、示された圧縮位置からデータを得ることができる。メタデータは、一般的に、受信または取得可能な圧縮されるデータの型または特性を示すことができる。図1及び3に関して以上で考察される実施例で説明されるように、型/フォーマット、クライアントの身元、クライアントの素性、及びデータが送信ならびに受信された時間に関する時間情報は全て、データと関連付けられるメタデータの実施例のものであってよい。例えば、メタデータを得ることには、圧縮要求を分析または解析することが含まれてもよく、その圧縮要求には、圧縮されるデータを含むヘッダまたは他のフォーマットもしくはメッセージ内の、メタデータが含まれてもよい。いくつかの実施形態では、特にデータ圧縮サービスによるAPIの一部として確立されている記述子または他のインジケータを使用するデータのために、メタデータを生成することができる。例えば、クライアントがデータのドメイン内の固有の値の数を識別するのを可能にする、値ドメインフラグを設定することができる。

【0046】

420に示されるように、データまたはそのデータと関連付けられるメタデータに分析を実行して、複数の圧縮技法の中から、データに適用される1つ以上の圧縮技法を選択することができる。データそれ自体を、全体で、またはいくつかのサブセットで、あるいはデータの部分で、分析、サンプリング、スキャン、またはレビューして、そのデータの1つ以上のデータ特性を決定することができる。それらのデータの特性は、データ型、フォーマット、サイズ、または可能値もしくは通常の反復値のセットの特定のサイズドメインを含むことができるが、それらに限定されない。エントロピー測定またはいくつかの他の測定を行うことができ、それらは、データ内のデータ値の変化も判定できることを示す。データ分析に基づいて、圧縮されるデータの1つ以上のデータ特性を決定することができる。また同様に、上述のように、そのデータと関連付けられるメタデータを使用して、そのデータのデータ特性を決定することができる。

【0047】

圧縮エンジンセクタ320に関して以上で考察されるように、ルールベースの分析またはいくつかの他の型の分析は、圧縮選択ルールのセット等の知識ベースを動的に使用することができる、または圧縮されるデータのデータ特性に適用することができる。圧縮ルールのセットを、圧縮されるデータのデータ特性に適用して、適用される1つ以上の圧縮技法を選択することができる。例えば、選択ルールは、整数等のデータ型から構成されるデータを示すことができ、1つ以上の効率的な圧縮技法を有することができる、したがって、1つ以上の効率的な圧縮技法が選択されて適用されるであろう。圧縮選択ルールを適用するとき、複数のデータ特性を使用することができる。例えば、データ値の型をフィルターとして使用して、可能な圧縮技法を、可能な圧縮技法の限定されたセットに狭めることができ、次いで、データ内のデータの分布が表示されることを利用して、適用される圧縮技法のより小さなセットを決定することができる。例えば、データ値の型は整数であって

よく、圧縮を圧縮技法のサブセットに限定し、次いで、値の小範囲を示す分布が1つ以上のディクショナリベースの圧縮技法の使用を示し得る場合等では、データ内の整数値の分布が表示されることによって、圧縮技法のサブセットが、圧縮技法の特定のセットにさらに狭められ得る。以下で考察される図6は、長い時間をかけてルールベースの分析を更新するまたは適合させるために使用可能である、様々な機械学習技法のさらなる考察をもたらす。ルールベースの分析の様々な実装は当業者に周知であるため、上述の実施例は限定を意図するものではない。

【0048】

いくつかの実施形態では、430に示されるように、1つ以上の選択された圧縮技法に従って、1つ以上のデータ圧縮候補を生成することができる。それらの圧縮技法は、様々な周知の、またはあまり知られていない、もしくはカスタマイズされた圧縮技法を含んでよく、その圧縮技法には、バイトディクショナリ、テキスト255、テキスト32k、デルタ、ランレングス、mostly n、ランレングス圧縮、Lempel-Ziv、Lempel-Ziv-Oberhumer、gzip、または、より一般的には、任意の他の形式のロスレスもしくはロシーデータ圧縮技法が含まれるが、それらに限定されない。いくつかの実施形態では、複数の圧縮技法を適用して、データ圧縮候補を生成することができる。多くの異なる圧縮技法は当業者に周知であり、したがって、以前の実施例は限定することを意図するものではない。いくつかの実施形態では、図3に関して以上で考察されるもの、及び図5に関する以下のもの等の圧縮サービス制限を施行することができる。

【0049】

440に示されるように、圧縮選択基準に従って、1つ以上のデータ圧縮候補のうちの1つを選択することができる。いくつかの実施形態では、圧縮選択基準は、サイズに従って最小のデータ圧縮候補を選択するための基準であってよい、または、時間によって分割されるデータ圧縮候補のサイズ、またはデータ圧縮候補を生成するために使用されるリソース等の、データ圧縮の性能特性のいくつかの組み合わせであってもよい。したがって、例えば、より小さな圧縮サイズのデータ圧縮候補よりも著しく時間を要さない場合は、わずかにより大きなサイズのデータ圧縮候補を選択することができる。

【0050】

次いで、それに応じて、450に示されるように、選択されたデータ圧縮候補を、要求された圧縮されたデータとして送信することができる。圧縮されたデータの受信者は、クライアントと同じリモートシステムまたは位置のものであってもよい、またはクライアントとは異なる位置またはリモートシステムのものであってもよい。また、要求された圧縮されたデータを送信する前に、選択されたデータ圧縮候補に対して、他の修正、変更、またはさらなる操作を実行することができる。例えば、いくつかの実施形態では、1つ以上の暗号化技法を選択されたデータ圧縮候補に適用することができる。同様に、システム圧縮技法等の別の圧縮技法を適用して多段圧縮されたデータを生成し、それらのデータを要求された圧縮されたデータとして送信することができる。

【0051】

上述のように、複数の異なるクライアントからの異なるデータのために、要素410～450を繰り返し実行することができる。少なくともいくつかの実施形態では、受信されるデータは、同じクライアントからのもの、及び同じ型ならびにフォーマットのもの、またはデータの他の同様の特性のものであってもよい。以前に、ルールベースの分析、1つ以上のデータ圧縮候補の生成、または送信するデータ圧縮候補のうちの1つの選択を実行することなく適用された技法に従って、そのようなデータを圧縮することができる。例えば、いくつかの実施形態では、受信されるデータは、複数のデータチャンクを含むデータストリームであってよい。第1のデータチャンクについて、要素410～450を実行して、データチャンクを圧縮することができる。しかしながら、データストリームの後続のデータチャンクでは、第1のデータチャンクに適用される圧縮技法に従って、各データチャンクを圧縮することができる。したがって、追加の分析を実行することなく、それらのデ

10

20

30

40

50

ータチャンクを圧縮して送信することができる。

【 0 0 5 2 】

図5を参照すると、いくつかの実施形態では、圧縮サービス制限を施行することができる。このサービス制限は、特定の順番の1つ以上の圧縮技法の選択と組み合わせることができる。これによって、最も効率的なデータ圧縮候補を生成して、圧縮制限を満たすことができる。図5は、いくつかの実施形態による、一連の選択された圧縮技法に従って、かつ所与の圧縮サービス制限内で、1つ以上のデータ圧縮候補を生成するための方法の、高レベルフローチャートである。

【 0 0 5 3 】

510に示されるように、実行されるルールベースの分析データ、または圧縮されるデータと関連付けられるメタデータを実行して、複数の圧縮技法の中から、データに適用される1つ以上の圧縮技法の順序を選択することができる。いくつかの実施形態では、この順番は、優先順位を表すことができ、また試行すべき最も効率的な圧縮技法の指標を他にいくつか表すことができる。例えば、520に示されるように、データに適用される圧縮技法の順序付けに従って、試行されていない圧縮技法を選択することが可能である。したがって、まず、最も効率的な圧縮を実行する可能性が最も高いデータ圧縮技法を選択することができる。いくつかの実施形態では、最も効率的な圧縮によって、最小のサイズの圧縮されたデータが圧縮技法により生成可能であることが示され得る。次いで、530に示されるように、圧縮技法のうちで選択された試行されていない技法に従って、データ圧縮候補を生成することができる。

【 0 0 5 4 】

図5に示されるように、圧縮サービス制限を超過するまで、540に示されるように、要素520及び530を繰り返すことが可能である。上述のように、圧縮サービス制限は、データ圧縮候補の生成を限定し得る任意の形式のルール、上限、リソース限定、もしくは境界、または所与のデータのために実行される全てのサービス等の、データ圧縮サービス220内の別の機能もしくは構成要素であってよい。例えば、いくつかの実施形態では、圧縮サービス制限は、サービス料金上限であってよく、その上限によって、圧縮データ候補の生成が、特定のサービス料金内で生成可能なものに限定される。図5等々に示されるように、圧縮技法の順序付けは連続的に適用可能であるが、データ圧縮候補が並行して生成される場合、圧縮技法の順序付けはシステム内で適用可能である。特定の数の圧縮エンジン（例えば、ノード、システム、及びバーチャルコンピュータインスタンス（*virtual compute instances*）等）を実装して、データ圧縮候補を（例えば、データ圧縮制限に従って）生成することができる、例示的なシステムについて考慮する。データ圧縮候補を生成するときにデータ圧縮エンジンによって適用される圧縮技法は、圧縮技法の順序付けによって決定することができる。現在使用される圧縮エンジンの1つ以上が圧縮データ候補の生成を完了させている場合は、追加の圧縮技法を試行することができ、それ故に、圧縮エンジンは、圧縮技法の順序付けに従って、動作と並行して、データ圧縮候補の生成も行うことができる。540の否定方向に示されるように、圧縮サービス制限を超過するとき、新たなデータ圧縮候補の生成はもはや継続されない可能性がある。以上の440で説明される選択等の選択が生じる可能性があり、550に記載されるように、選択された圧縮されたデータを送信することができる。

【 0 0 5 5 】

示される圧縮サービス制限及び圧縮技法の順番を適用するための技法はともに示されるが、それらは、個別に、または図4に関して以上で考察されるもの等の他の方法もしくは技法と組み合わせて実行可能であることを、理解されたい。

【 0 0 5 6 】

図6を参照すると、いくつかの実施形態では、データのルールベースの分析及び/または圧縮されるデータと関連付けられるメタデータは、1つ以上の機械学習技法に基づいて、修正、更新、または変更可能である。図6は、いくつかの実施形態による、機械学習を実行して圧縮されるデータのルールベースの分析を更新するための方法を示す、高レベル

10

20

30

40

50

フローチャートである。610に示されるように、圧縮データサービスにおいて以前に圧縮されたデータのための過去の圧縮データを得ることができる。図3中の過去の圧縮データ350に関して以上で考察されるように、この過去の圧縮データは、様々な異なるデータ特性、性能特性、及び/またはデータの圧縮の以前の性能と関連付けられる任意の他のデータを含むことができる。例えば、個人のクライアントのための圧縮されたデータの履歴を保持することができる。

【0057】

620に示されるように、過去の圧縮データに1つ以上の機械学習技法を実行して、圧縮されるデータのデータ特性に適用される圧縮選択ルールの最新のセットを更新することができる。例えば、圧縮されるデータは、データ型、クライアントの種類、及びサイズを含むデータ特性のセットを有し得る。同じまたは類似する特性を有するデータの、以前に圧縮されたデータのための過去の圧縮データを得ることができる。管理、非管理、または半管理学習等の1つ以上の機械学習技法を過去の圧縮データに適用することができる。機械学習技法によって識別される学習に基づいてデータに適用される、選択ルール。例えば、異なる圧縮技法を用いて既知のデータ型をよりよく圧縮可能であることが一般的に知られているとしても、この種類のクライアント、及びデータのサイズでは、特定の圧縮技法によって、データがより効率的に圧縮されると考えられると判定され得る。機械学習からのこの観測に基づいて圧縮選択ルールを選択することにおいて、特定の圧縮技法の優先度は上昇し得る。機械学習は当業者に周知であるため、以前の実施例は限定することを意図するものではない。次いで、630に示されるように、更新された圧縮選択ルールのセットを、圧縮されるデータのデータ特性に適用して、データに適用される1つ以上の圧縮技法を選択することができる。

【0058】

図7を参照すると、いくつかの実施形態では、多段圧縮は、圧縮のために受信されたデータに適用することできる。図7は、いくつかの実施形態による、多段圧縮を含むサービスとして、効率的なデータ圧縮及び分析を実行するための方法を示す、高レベルフローチャートである。710に示されるように、1つ以上のデータ圧縮技法に従って、データ圧縮候補を生成可能である。いくつかの実施形態では、このデータ圧縮候補は選択されたデータ圧縮候補であり、要求された圧縮されたデータとして送信される。次いで、720に示されるように、データ圧縮候補のために、エントロピー測定値を決定することができる。データ圧縮候補内に記憶されたデータ値の変化を示す、エントロピー測定値を決定することができる。次いで、730に示されるように、データ圧縮候補のエントロピー測定値がエントロピー閾値未満であるか否かが判定され得る。閾値未満である場合は、740に示されるように、システム圧縮技法に従って、データ圧縮候補を再度圧縮して、圧縮されたデータとして送信される多段圧縮されたデータを生成することできる。システム圧縮技法は、図3及び4に関する上述のもの等の、任意の圧縮技法であってよい。次いで、750に示されるように、この多段圧縮されたデータを送信することできる。図示されていないが、この多段圧縮されたデータを調査して、多段圧縮されたサイズが、以前に圧縮されたサイズよりも大きいかが否かを判定することでき、大きい場合は、多段圧縮を解除して、760に示されるようにデータを送信することできる。代替として、エントロピー測定値が閾値以上である場合は、760等に示されるように、データ圧縮候補を送信することできる。しかしながら、図7は、エントロピー閾値を適用可能であることを示すが、いくつかの実施形態では、送信されるデータとして選択された全てのデータ圧縮候補は、システム全体の圧縮技法を用いて圧縮可能である。

【0059】

図2に関して以上で考察されるように、いくつかの実施形態では、データ圧縮を、ネットワークベースのサービスとして、クライアント250a~250n等の様々な異なるクライアントに提供することできる。図8は、いくつかの実施形態による、データ圧縮をサービスとして実行するための方法を示す高レベルフローチャートである。様々な異なるシステム、構成要素、及びアーキテクチャを利用して、以下に記載される技法を実装する

10

20

30

40

50

ことができる。少なくともいくつかの実施形態では、図 2 中のネットワークベースのサービスプラットフォーム 200 等の、ネットワークベースのサービスプラットフォームは、記載される要素をいくつかまたは全て実装することができる。しかしながら、以下に示される実施例は、ネットワーク系データ圧縮サービスを実装する任意の特定のアーキテクチャ、構成、または構成要素を限定することを意図するものではない旨を理解されたい。また、要素の特定の順序付けまたは配列も、要素の配列の任意の他の順序付け、または追加の要素の実行、もしくは図 8 に示す要素の除去を限定するものではない。

【0060】

810 に示されるように、いくつかの実施形態では、クライアントから圧縮されるデータを示す圧縮要求を受信することができる。図 3 及び 7 に関する上述の様々な圧縮要求として、圧縮されるデータを、圧縮要求とともに、示す（例えば、データの位置）、含む、または送信することができる。圧縮要求は、圧縮要求と関連付けられる 1 つ以上の識別子を含むことができ、その識別子によって、クライアント、またはその要求が送信されるクライアントアカウントが識別される。例えば、クライアントアカウント番号の要求を含む、いくつかのフィールドを定義する API、または他の型のプロトコル、もしくはインターフェースに従って、識別子をフォーマットすることができる。同様に、データ圧縮サービスから提供される、様々な型の圧縮サービスまたはオプションの要求等の圧縮要求において、他のクライアント情報を示すことができる。例えば、圧縮要求は、圧縮要求に対する特定の料金体系を識別することができ、次いで、その料金体系によって、いかにして圧縮されたデータを生成することができるか、及び要求と関連付けられる料金はいくら発生し得るかを、判定することができる。同様に、特定の型の圧縮技法、分析、クライアント履歴、または任意の他の種類の情報を、圧縮要求の一部として示すことができる。例えば、クライアントは、単に同じクライアントからの以前のクライアント圧縮要求に依存して、示されるデータを分析かつ圧縮するというオプションとともに、圧縮要求を提出することができる。

【0061】

様々な実施形態では、ネットワーク系データ圧縮サービスは、様々な異なる料金体系を、圧縮クライアントに提供することができる。それらの料金体系には、1 つ以上の圧縮サービス因子及び/または制限を決定することが含まれ得る。例えば、いくつかの実施形態では、リソース制約、時間制約、及びコスト制約は、料金体系に含まれる 1 つ以上の圧縮サービス因子であってよい。例えば、様々な時間制約が料金体系に含まれ得、その料金体系によって、特定のサービスレベルの合意に適合するもの等の、合意された完了時間（例えば、要求につき 2 分）が、圧縮要求に提供される。料金上限等のコスト制約、または図 7 に関して以上で考察されるもの等の他の圧縮サービス制限を、料金体系の一部として含むことができる。また、同様に、リソース制約も含むことができる。例えば、（例えば、利用可能であるとき、要求に応じて、予約されたとき、または特定の値段で）圧縮要求を実行するまたは処理するための、特定の数のリソース（例えば、ノード）。

【0062】

次いで、様々な実施形態では、820 に示されるように、圧縮要求と関連付けられる料金体系を決定することができる。料金体系の決定は、例えば、圧縮要求に含まれる 1 つ以上の識別子（例えば、アカウント番号）を介して、クライアントを識別することによって実行され得る。また、データの型、データが送信された時間、及び図 4 に関する上述のメタデータ等の、特定のデータのために含まれるメタデータ等の他の因子を含んで、データ圧縮サービスを決定する際に使用することができる。いくつかの実施形態では、データ圧縮サービスによって、単一の料金体系をクライアントに提供することができる。しかしながら、少なくともいくつかの実施形態では、同じクライアントまたは異なるクライアントからの異なる要求に対して、複数の料金体系を決定する場合がある。

【0063】

830 に示すように、決定された料金体系に従って、圧縮されたデータを生成することができる。図 3 及び 4 に関して以上で考察されるように、様々な異なる圧縮技法を、圧縮

10

20

30

40

50

要求のデータに適用することができる。少なくともいくつかの実施形態では、圧縮要求は、示されたデータに適用するための、選択された1つ以上の圧縮技法を示すことができる。代替として、いくつかの実施形態では、圧縮要求は、例えば、ルールベースの分析、または機械学習によって通知されるもしくは機械学習なしで通知される他の形式の動的分析を適用すること等の、図3及び4に関する上述の様々な分析技法のうちの1つ以上を要求することができる。図5に関する以上の考察と同様に、決定された料金体系に従って、要求された圧縮されたデータを生成することができる。例えば、いくつかの実施形態では、決定された料金体系によって、データ圧縮候補を生成するための時間制限、またはリソース、もしくは料金上限が適用され得る。同様に、料金体系によって、異なる圧縮技法の値段を異なる値段で設定することができ、ユーザーが選択した圧縮技法が、要求に対する料金体系によって決定された上限または限定を超過する場合は、代わりに類似する圧縮技法を圧縮要求に対して実行することができる（または、選択された圧縮技法の使用に対してより高い料金を発生させることができる）。より一般的には、いくつかの実施形態では、圧縮されたデータを生成するとき、決定された料金体系によって、1つ以上の圧縮サービス制限を圧縮要求に課すことができる。例えば、圧縮サービス制限がクライアントに対して単に圧縮要求を実行するものである場合は、データ圧縮のコストが（例えば、要求を実行するためのノード等のリソースの利用可能性に基づいて）特定の値段であるとき、圧縮サービス制限を満たすことができるまで、要求がキューされ得る。他の多くの異なる圧縮サービス制限及び料金体系の種類を想定することができるため、以前の実施例は限定することを意図するものではない。

【0064】

少なくともいくつかの実施形態では、840に示されるように、料金体系に従って、圧縮されたデータに対して料金を発生させることができる。例えば、料金体系によって、圧縮する時間、圧縮するデータの量、圧縮するデータの型、データを圧縮するためのリソースの数（例えば、ノードの数）、及び圧縮技法の型等の様々な因子に対する特定の価格またはコストを類型化することができる。様々な料金体系によって、特定の数の圧縮要求を、所与の期間内（例えば、10日）で提供することができる、または特定の量の圧縮されるデータを、特定の期間内（例えば、1ヶ月あたり10ギガバイト）で提供することができる。データの動的またはルールベースの分析等の、上述の様々なオプションまたは圧縮サービスに対して、追加の請求、料金、または値段を加えることができる。少なくともいくつかの実施形態では、図2中のネットワークベースのサービスインターフェース200に関する上述のもの等の、クライアントの会計及び請求情報を保持するデータ蓄積内の料金を記録すること等によって、発生する料金をクライアントと関連付けることができる。

【0065】

次いで、850に示されるように、圧縮要求に従って、要求された圧縮されたデータを送信することができる。図3及び4に関して以上で考察されるように、圧縮されたデータを、別のクライアント、システム、またはサービス等の、データの発信元とは異なるリモート位置に送信することができる。例えば、圧縮されたデータを、保管または他の永続データ蓄積サービスに送信することができる。代替として、圧縮されたデータを、要求するクライアントに返送することができる。

【0066】

いくつかの実施形態では、復元要求は、上述の図8中のものに類似する様式の、クライアントのためのプロセスであってよい。圧縮要求に対する料金体系を決定することができ、その料金体系に従ってデータを復元することができ、復元の料金を発生させることができ、かつ圧縮要求に従って、復元されたデータを送信することができる。

【0067】

例示的なシステム

本明細書中に記載の効率的なデータ圧縮及び分析の実施形態は、様々な他のデバイスと交信可能な1つ以上のコンピュータシステムに実行することができる。1つのそのようなコンピュータシステムは、図9によって示される。様々な実施形態では、コンピュータシ

10

20

30

40

50

システム１０００は、パーソナルコンピュータシステム、デスクトップコンピュータ、ラップトップ、ノートパソコンもしくはネットブックコンピュータ、メインフレームコンピュータシステム、ハンドヘルドコンピュータ、ワークステーション、ネットワークコンピュータ、カメラ、セットトップボックス、モバイルデバイス、消費者デバイス、ビデオゲーム機器、ハンドヘルドビデオゲームデバイス、アプリケーションサーバー、記憶デバイス、スイッチ、モデム、及びルーター等の周辺機器、または、一般に、任意の型のコンピュータティングもしくは電子デバイスを含むが、それらに限定されない様々な型のデバイスのうちのいずれかであってよい。

【００６８】

例示された実施形態では、コンピュータシステム１０００は、入力／出力（Ｉ／Ｏ）インターフェース１０３０を介してシステムメモリ１０２０に接続される、１つ以上のプロセッサ１０１０を含む。コンピュータシステム１０００は、Ｉ／Ｏインターフェース１０３０に接続されるネットワークインターフェース１０４０、及びカーソル制御デバイス１０６０、キーボード１０７０、ならびにディスプレイ（複数可）１０８０等の１つ以上の入力／出力デバイス１０５０をさらに含む。ディスプレイ（複数可）１０８０は、標準コンピュータモニター（複数可）、及び／または他のディスプレイシステム、技術、もしくはデバイスを含むことができる。少なくともいくつかの実装では、入力／出力デバイス１０５０もまた、パッドもしくはタブレット等のタッチもしくはマルチタッチ可能デバイスを含むことができ、それらを介して、ユーザーは、スタイラス型デバイス及び／または１つ以上のデジットを用いて入力を行う。いくつかの実施形態では、コンピュータシステム１０００の単一のインスタンスを用いて、実施形態を実装することができる旨が企図されるが、他の実施形態では、コンピュータシステム１０００を構成する複数のそのようなシステムまたは複数のノードを構成して、実施形態の異なる部分またはインスタンスをホストすることができる。例えば、一実施形態では、他の構成要素を実装するノードとは異なる、コンピュータシステム１０００の１つ以上のノードを介して、いくつかの要素を実装することができる。

【００６９】

様々な実施形態では、コンピュータシステム１０００は、１つのプロセッサ１０１０、またはプロセッサ１０１０をいくつか（例えば、２つ、４つ、８つまたは別の好適な数）含むマルチプロセッサシステムを含む、ユニプロセッサシステムであってよい。プロセッサ１０１０は、インストラクションを実行することができる、任意の好適なプロセッサであってよい。例えば、様々な実施形態では、プロセッサ１０１０は、x 8 6、Power PC、SPARC、またはMIPS ISA、もしくは任意の他の好適なISA等の、様々なインストラクションセットアーキテクチャ（ISA）のいずれかを実装する、汎用または組み込みプロセッサであってよい。マルチプロセッサシステムでは、各々のプロセッサ１０１０は、通常、同じISAを実装することができるが、必ずしも同じISAを実装しなくてもよい。

【００７０】

いくつかの実施形態では、少なくとも１つのプロセッサ１０１０は、グラフィック処理ユニットであってよい。グラフィック処理ユニット、すなわちGPUは、パーソナルコンピュータ、ワークステーション、ゲーム機器、または他のコンピュータティングもしくは電子デバイス専用のグラフィックレンダリングデバイスであってよいとされる。最新のGPUは、非常に効率的にコンピュータグラフィックを操作かつ表示することができ、その高並列構造によって、グラフィカルアルゴリズムの範囲が複雑である典型的なCPUよりも、効果的なものとなり得る。例えば、グラフィックプロセッサは、複数のグラフィックプリミティブ動作を実装することができるが、その実装の実行は、ホスト中央処理ユニット（CPU）を用いてスクリーンに直接描画するよりもさらに速い方式で行われる。様々な実施形態では、グラフィックレンダリングを、そのようなGPUのうちの１つにおいて実行されるように、または２つ以上において並行実行されるように構成されるプログラムインストラクションによって、少なくとも一部実装することができる。GPU（複数可）は

、プログラマーがGPU（複数可）の機能性を起動するのを可能にする、1つ以上のアプリケーションプログラマーインターフェース（API）を実装することができる。好適なGPUは、NVIDIA Corporation、及びATI Technologies（AMD）等のベンダーから、商業的に入手可能であり得る。

【0071】

プロセッサ1010によってアクセス可能なプログラムインストラクション及び/またはデータを記憶するために、システムメモリ1020を構成することができる。様々な実施形態では、静的ランダムアクセスメモリ（SRAM）、同期式動的RAM（SDRAM）、不揮発性/フラッシュ型メモリ、または任意の他の型のメモリ等の、任意の好適なメモリ技術を使用して、システムメモリ1020を実装することができる。例示された実施形態では、上述のような望ましい機能を実装するプログラムインストラクション及びデータは、システムメモリ1020内に、プログラムインストラクション1025及びデータ記憶1035としてそれぞれ記憶されることが示されている。他の実施形態では、システムメモリ1020、またはコンピュータシステム1000から分離している異なる型のコンピュータアクセス可能なメディアもしくは同様のメディアにおいて、プログラムインストラクション及び/またはデータを受信、送信、または記憶することができる。一般的に言えば、コンピュータアクセス可能な媒体は、磁気もしくは光学メディア、例えば、I/Oインターフェース1030を介してコンピュータシステム1000に接続される、ディスクもしくはCD/DVD-ROM等の、記憶メディアまたはメモリメディアを含むことができる。コンピュータアクセス可能な媒体を介して蓄積されるプログラムインストラクション及びデータは、伝送メディア、または電気、電磁、もしくはデジタル信号等の信号によって、発信することができ、それらの伝送メディアまたは信号は、ネットワークインターフェース1040等を介して実装可能なネットワーク及び/または無線リンク等の、通信媒体を介して伝達することができる。

【0072】

一実施形態では、I/Oインターフェース1030を構成して、プロセッサ1010と、システムメモリ1020と、入力/出力デバイス1050等のネットワークインターフェース1040あるいは他の周辺機器インターフェースを含むデバイス内の任意の周辺機器デバイスと間の、I/Oトラフィックを調整することができる。いくつかの実施形態では、I/Oインターフェース1030は、任意の必須のプロトコル、時間、または他のデータ変換を実行して、データ信号を、1つの構成要素（例えば、システムメモリ1020）から、別の構成要素（例えば、プロセッサ1010）によって使用されるのに好適であるフォーマットに変換することができる。いくつかの実施形態では、I/Oインターフェース1030は、例えば、周辺機器構成要素相互接続（PCI）バス標準またはユニバーサルシリアルバス（USB）標準の変形等の、様々な型の周辺機器バスを介して取り付けられるデバイスのための支援を含むことができる。いくつかの実施形態では、I/Oインターフェース1030の機能を、例えば、ノースブリッジ及びサウスブリッジ等の、2つ以上の個別の構成要素に分割することができる。さらに、いくつかの実施形態では、システムメモリ1020へのインターフェース等の、I/Oインターフェース1030の機能のいくつかまたは全てを、プロセッサ1010内に直接組み込むことができる。

【0073】

ネットワークインターフェース1040を構成することによって、データを、コンピュータシステム1000と、他のコンピュータシステム等のネットワークに接続される他のデバイスとの間で、またはコンピュータシステム1000のノード間で交換することが可能となる。様々な実施形態では、ネットワークインターフェース1040は、任意の好適な型のイーサネット（登録商標）ネットワーク等の有線または無線一般データネットワークを介して、例えば、アナログ音声ネットワークもしくはデジタルファイバー通信ネットワーク等の電気通信/電話ネットワークを介して、ファイバーチャネルSAN等の記憶領域ネットワークを介して、または任意の他の好適な型のネットワーク及び/もしくはプロトコルを介して、通信を支援することができる。

【 0 0 7 4 】

いくつかの実施形態では、入力／出力デバイス 1 0 5 0 は、1 つ以上のディスプレイ端末、キーボード、キーパッド、タッチパッド、スキャニングデバイス、音声もしくは光学認識デバイス、または、1 つ以上のコンピュータシステム 1 0 0 0 によってデータを入力するもしくは取り出すのに好適である、任意の他のデバイスを含むことができる。複数の入力／出力デバイス 1 0 5 0 は、コンピュータシステム 1 0 0 0 内に存在してもよい、またはコンピュータシステム 1 0 0 0 の様々なノードにおいて分配されてもよい。いくつかの実施形態では、同様の入力／出力デバイスは、コンピュータシステム 1 0 0 0 から分離されてもよく、有線もしくは無線接続を介して、例えば、ネットワークインターフェース 1 0 4 0 を介して、コンピュータシステム 1 0 0 0 の 1 つ以上のノードと通信することができる。

10

【 0 0 7 5 】

図 9 に示されるように、メモリ 1 0 2 0 は、効率的なデータ圧縮及び分析の実施形態を本明細書中に記載のサービスとして実装するように、かつプログラムインストラクション 1 0 2 5 によってアクセス可能な様々なデータを含むデータ記憶 1 0 3 5 を実装するように構成される、プログラムインストラクション 1 0 2 5 を含むことができる。一実施形態では、プログラムインストラクション 1 0 2 5 は、本明細書中に記載される、かつ図中に示される実施形態のソフトウェア要素を含むことができる。データ記憶 1 0 3 5 は、実施形態内で使用可能なデータを含むことができる。他の実施形態では、他のまたは複数のソフトウェア要素及びデータを含むことができる。

20

【 0 0 7 6 】

コンピュータシステム 1 0 0 0 は単に例示的なものであり、本明細中に記載の実施形態の範囲を限定することを意図するものではない旨を、当業者は理解するであろう。特に、コンピュータシステム及びデバイスは、示される機能を実行することができるハードウェアまたはソフトウェアの任意の組み合わせを含むことができ、それらのハードウェアまたはソフトウェアには、コンピュータ、パーソナルコンピュータシステム、デスクトップコンピュータ、ラップトップ、ノートパソコンもしくはネットブックコンピュータ、メインフレームコンピュータシステム、ハンドヘルドコンピュータ、ワークステーション、ネットワークコンピュータ、カメラ、セットトップボックス、モバイルデバイス、ネットワークデバイス、インターネット器具、PDA、無線電話、ページャー、消費者デバイス、ビデオゲーム機器、ハンドヘルドビデオゲームデバイス、アプリケーションサーバー、記憶デバイス、スイッチ、モデム、ならびにルーター等の周辺機器デバイス、または、一般に、任意の型のコンピューティングもしくは電子デバイスが含まれる。また、コンピュータシステム 1 0 0 0 は、示されない他のデバイスに接続可能である、または、代わりに、スタンドアロンシステムとして動作可能である。さらに、いくつかの実施形態では、示される構成要素によってもたらされる機能性は、より少ない構成要素と組み合わせることができる、または追加の構成要素内に分配させることができる。同様に、いくつかの実施形態では、いくつかの示される構成要素の機能性を提供しなくてもよい、及び／または、他の追加の機能性を利用してもよい。

30

【 0 0 7 7 】

また、様々な物品は、使用される間、メモリまたは記憶内に記憶されるものとして示されるが、メモリ管理及びデータの完全性を目的として、それら物品またはそれら物品の部分を、メモリと他の記憶デバイスとの間で移動させることができる旨も、当業者は理解するであろう。代替として、他の実施形態では、ソフトウェア構成要素のいくつかもしくは全ては、別のデバイスのメモリ内で実行可能であり、かつ、コンピュータ間通信によって、示されるコンピュータシステムと通信可能である。また、システム構成要素またはデータ構造のうちのいくつかまたは全てを、適切なドライブによって読み込まれるコンピュータアクセス可能な媒体またはポータブル物品上で、（例えば、インストラクションまたは構造化データとして）記憶することもできるが、そのドライブの様々な実施例は、以上に記載されるものである。いくつかの実施形態では、伝送メディア、またはネットワーク及

40

50

び／もしくは無線リンク等の通信媒体を介して伝達される電気、電磁、もしくはデジタル信号等の信号を介して、コンピュータシステム１０００から分離するコンピュータ可読媒体上に記憶されるインストラクションを、コンピュータシステム１０００に発信することができる。このコンピュータ可読記憶媒体は、持続性であってよい。様々な実施形態は、上記説明に従ってコンピュータアクセス可能な媒体に実装されるインストラクション及び／またはデータを、受信すること、送信すること、または記憶することをさらに含んでもよい。したがって、本発明は、他のコンピュータシステム構成を用いて実施することができる。

【００７８】

また、以下の付記の観点からも、上述についてよりよく理解することができる。

10

１．圧縮サービスを実装するよう構成される１つ以上のコンピューティングデバイスを備えるシステムであって、前記圧縮サービスが、

複数の圧縮エンジンであって、各々が、複数の圧縮技法の中から、少なくとも１つの圧縮技法を実行するように構成される、複数の圧縮エンジンと、

ルールベースの圧縮エンジンセレクトアであって、

クライアントから圧縮されるデータを受信し、

前記データの受信に応答して、

圧縮される前記データまたは前記データと関連付けられるメタデータにルールベースの分析を実行して、前記複数の圧縮技法の中から前記データに適用される１つ以上の圧縮技法を選択し、かつ

20

前記選択された１つ以上の圧縮技法に従って、かつ所与の圧縮サービス制限に準じて、前記複数の圧縮エンジンのうちの１つ以上に、１つ以上の圧縮データ候補を生成するように指示するように構成される、圧縮エンジンセレクトアと、

応答生成モジュールであって、

圧縮選択基準に従って、前記１つ以上のデータ圧縮候補のうちの１つを選択して、要求された圧縮されたデータとして送信するように構成される、応答生成モジュールと、を備える、前記システム。

２．圧縮される前記データまたは前記データと関連付けられる前記メタデータにルールベースの分析を実行して、前記複数の圧縮技法の中から前記データに適用される前記１つ以上の圧縮技法を選択するために、前記ルールベースの圧縮エンジンセレクトアが、最新の圧縮選択ルールのセットを、前記データのための１つ以上のデータ特性に適用するように構成され、

30

前記圧縮サービスが、

１つ以上の機械学習技法を過去の圧縮データに実行して、前記最新の圧縮選択ルールのセットを更新するように構成される、機械学習圧縮分析モジュールをさらに備える、付記１に記載の前記システム。

３．前記応答生成モジュールが、

前記１つ以上のデータ圧縮候補のうちの前記選択された１つに対してエントロピー測定値を決定し、かつ

前記１つ以上のデータ圧縮候補のうちの前記選択された１つに対して前記エントロピー測定値がエントロピー閾値未満であると決定することに応答して、システム圧縮技法に従って、前記選択された１つを圧縮して、前記要求された圧縮されたデータとして送信される多段圧縮されたデータを生成するようにさらに構成される、付記１に記載の前記システム。

40

４．１つ以上のコンピューティングデバイスによって、

クライアントから圧縮されるデータを示す圧縮要求の受信を実行することと、

前記圧縮要求の受信に応答して、

圧縮される前記データまたは前記データと関連付けられるメタデータに分析を実行して、複数の圧縮技法の中から、前記データに適用される１つ以上の圧縮技法を選択することと、

50

前記 1 つ以上の圧縮技法に従って、1 つ以上のデータ圧縮候補を生成することと、
圧縮選択基準に従って、前記 1 つ以上のデータ圧縮候補のうちの 1 つを選択して、要求された圧縮されたデータとして送信すること、かつ

前記要求された圧縮されたデータを含む応答を送信することと、を実行することを含む、方法。

5 . 前記 1 つ以上の圧縮技法に従う前記 1 つ以上のデータ圧縮候補の該生成が、
所与の圧縮サービス制限を超過するまで、前記 1 つ以上の圧縮技法の各々についてデータ圧縮候補を生成することを含む、付記 4 に記載の前記方法。

6 . 前記所与の圧縮サービス制限が時間制限である、付記 5 に記載の前記方法。

7 . 圧縮される前記データまたは前記データと関連付けられるメタデータの分析がルールベースの分析である、付記 4 に記載の前記方法。

8 . 前記方法が、

1 つ以上の機械学習技法を過去の圧縮データに実行して、前記 1 つ以上の圧縮技法を選択するために適用される最新の圧縮選択ルールのセットを更新することをさらに含み、

圧縮される前記データまたは前記データと関連付けられる前記メタデータの前記分析を実行して、前記複数の圧縮技法の中から、前記データに適用される前記 1 つ以上の圧縮技法を選択することが、

前記更新された圧縮選択ルールのセットを、各データの 1 つ以上のデータ特性に適用して、前記データに適用される前記複数の圧縮技法の中から前記 1 つ以上の圧縮技法を選択することを含む、付記 7 に記載の前記方法。

9 . 前記メタデータが、圧縮されるデータの型または他の 1 つ以上の特性を示し、前記 1 つ以上の機械学習技法が、前記圧縮されるデータの型、または前記他の 1 つ以上の特性に少なくとも一部基づいて実行される、付記 8 に記載の前記方法。

10 . 圧縮される前記データまたは前記データと関連付けられる前記メタデータの分析を実行して、前記複数の圧縮技法の中から、前記データに適用される前記 1 つ以上の圧縮技法を選択することが、前記選択された 1 つ以上の圧縮技法に、ランダムに選択された圧縮技法を含むことをさらに含む、付記 8 に記載の前記方法。

11 . 圧縮される複数の他のデータを、複数のクライアントから受信することと、

前記複数の他のデータの各々について、

該ルールベースの分析、該生成、該選択、及び該送信を実行することと、

前記他のデータのための圧縮結果データ及びデータ特性を、前記過去の圧縮データの一部として記憶することと、をさらに含む、付記 8 に記載の前記方法。

12 . 前記要求された圧縮されたデータを含む前記応答を送信する前に、前記システム圧縮技法に従って、前記選択されたデータ圧縮候補を圧縮して、前記要求された圧縮されたデータとして送信される多段圧縮されたデータを生成することをさらに含む、付記 4 に記載の前記方法。

13 . 前記圧縮されるデータが、複数のデータチャンクを含むデータストリームであり、該実行すること、該生成すること、及び該選択することが、前記複数の圧縮されるデータチャンクのうちの前記第 1 の圧縮されるデータチャンクに対して実行され、前記方法が、

前記複数のデータチャンクのうちの後続データチャンクの各々について、

前記第 1 のデータチャンクを圧縮するために適用される前記 1 つ以上のデータ圧縮技法に従って、圧縮されたデータチャンクを生成することと、

前記データチャンクを含む応答を送信することと、をさらに含む、付記 4 に記載の前記方法。

14 . 圧縮されたデータを示す別のクライアントからの復元要求であって、該復元要求が、前記圧縮されたデータを生成するために適用される 1 つ以上の圧縮技法を示す、前記復元要求を受信することと、

復元されたデータを生成するために適用される、示される 1 つ以上の圧縮技法に従って、前記圧縮されたデータを復元して、復元されたデータオブジェクトを生成することと、

10

20

30

40

50

前記復元されたデータを含む前記他のクライアントに、応答を送信することと、をさらに含む、付記 4 に記載の前記方法。

15．要求された圧縮されたデータを含む応答を送信する前に、1 つ以上の圧縮技法に従って、前記選択された圧縮されたデータ候補を暗号化して送信することをさらに含む、付記 4 に記載の前記方法。

16．前記 1 つ以上のコンピューティングデバイスがともに機能して、ネットワーク系データ圧縮サービスを実装し、前記前記圧縮要求が、ネットワーク系圧縮サービスのためのインターフェースに従って受信され、

前記方法が、

前記圧縮要求の受信に応答して、

前記圧縮要求に少なくとも一部基づいて、前記圧縮要求に対する料金体系を決定することと、

前記決定された料金体系に基づいて、前記要求された圧縮されたデータと関連付けられる料金を発生させることと、をさらに含む、

前記データ、または前記圧縮されるデータと関連付けられるメタデータでの前記分析の該実行、前記 1 つ以上の圧縮技法に従う前記 1 つ以上のデータ圧縮候補の該生成、及び前記 1 つ以上のデータ圧縮候補のうちから 1 つの該選択が、前記決定された料金体系に従って実行される、付記 4 に記載の前記方法。

17．ネットワークベースのサービスを実装するように構成される複数のコンピューティングデバイスを備えるシステムであって、前記ネットワークベースのサービスが、

1 つ以上の圧縮技法に従ってデータを圧縮するように構成されるデータ圧縮サービスモジュールと、

ネットワークベースのサービスインターフェースであって、

クライアントから圧縮されるデータを示す圧縮要求を受信し、

前記圧縮要求の受信に応答して、

前記圧縮要求と関連付けられる料金体系を決定し、

前記要求と関連付けられる前記料金体系に従って、前記データ圧縮サービスモジュールに要求された圧縮されたデータを生成するよう指示し、

前記料金体系に従って、前記要求された圧縮されたデータに対する料金を発生させ、かつ

前記圧縮要求に従って、前記要求された圧縮されたデータを送信するように構成される、前記ネットワークベースのサービスインターフェースと、を備える、前記システム。

18．前記料金体系が、圧縮要求に対する圧縮サービス制限を示し、前記要求された圧縮されたデータを生成するために、前記データサービス圧縮モジュールが、前記要求された圧縮されたデータを前記圧縮サービス制限内で生成するよう構成される、付記 17 に記載の前記システム。

19．前記圧縮要求が、前記示されたデータに適用される 1 つ以上のクライアント選択の圧縮技法をさらに示し、前記要求された圧縮されたデータを生成するために、前記データサービス圧縮モジュールが、前記要求された圧縮されたデータを生成するために、1 つ以上のクライアント選択の圧縮技法を実行するように構成される、付記 17 に記載の前記システム。

20．前記圧縮要求が、前記圧縮されるデータの圧縮分析をさらに要求し、1 つ以上の圧縮技法に従って前記データを圧縮するために、前記データ圧縮サービスモジュールが、

圧縮される前記データまたは前記データと関連付けられるメタデータに分析を実行して、複数の圧縮技法の中から、前記データに適用される 1 つ以上の圧縮技法を選択し、

前記 1 つ以上の圧縮技法に従って、1 つ以上のデータ圧縮候補を生成し、かつ

圧縮選択基準に従って、前記 1 つ以上のデータ圧縮候補のうちの 1 つを、前記要求された圧縮されたデータとして選択するように構成される、付記 17 に記載の前記システム。

21．プログラムインストラクションを記憶する持続性コンピュータ可読記憶媒体であって、1 つ以上のコンピューティングデバイスによって実行されるとき、

10

20

30

40

50

クライアントから圧縮されるデータを示す圧縮要求の受信を実行すること、
前記圧縮要求の受信に応答して、

前記データ、または前記圧縮されるデータと関連付けられるメタデータに実行して、前記データに適用される複数の圧縮技法の中から、1つ以上の圧縮技法を選択すること、

前記1つ以上の圧縮技法に従って、1つ以上のデータ圧縮候補を生成すること、

圧縮選択基準に従って、前記1つ以上のデータ圧縮候補のうちの1つを選択して、要求された圧縮されたデータとして送信すること、かつ

前記要求された圧縮されたデータを含む応答を送信すること、を実装する、前記持続性コンピュータ可読記憶媒体。

22．前記1つ以上の圧縮技法に従う、前記1つ以上のデータ圧縮候補の該生成において、前記プログラムインストラクションが、前記1つ以上のコンピューティングデバイスによって実行されるとき、

10

所与の圧縮サービス制限を超過するまで、前記1つ以上の圧縮技法のそれぞれのデータ圧縮候補を生成することを実装し、該所与の圧縮サービス制限が圧縮サービス料金上限である、付記14に記載の前記持続性コンピュータ可読記憶媒体。

23．前記データに適用される前記選択された1つ以上の圧縮技法が特定の順番で順序付けされ、前記1つ以上の圧縮技法に従う、前記1つ以上のデータ圧縮候補の該生成において、前記プログラムインストラクションが、前記1つ以上のコンピューティングデバイスによって実行されるとき、前記1つ以上の圧縮技法の前記特定の順番に従って、前記1つ以上のデータ圧縮候補を生成することを実装する、付記14に記載の前記持続性コンピュータ可読記憶媒体。

20

24．前記プログラムインストラクションが、前記1つ以上のコンピューティングデバイスによって実行されるとき、

1つ以上の機械学習技法を過去の圧縮データに実行して、前記1つ以上の圧縮技法を選択するために適用される最新の圧縮選択ルールセットを更新することをさらに実装し、

前記ルールベースの分析を、前記データ、または前記圧縮されるデータと関連付けられる前記メタデータに実行して、前記データに適用される前記複数の圧縮技法の中から前記1つ以上の圧縮技法の該選択を行うことにおいて、前記プログラムインストラクションが、前記1つ以上のコンピューティングデバイスによって実行されるとき、

前記更新された圧縮選択ルールセットを、各データの1つ以上のデータ特性に適用して、前記データに適用される前記複数の圧縮技法の中から前記1つ以上の圧縮技法を選択することを実装する、付記14に記載の前記持続性コンピュータ可読記憶媒体。

30

25．前記要求された圧縮されたデータを含む前記応答が、前記クライアントとは異なるリモートシステムに送信される、付記14に記載の前記持続性コンピュータ可読記憶媒体。

【0079】

結論

様々な実施形態は、上記説明に従ってコンピュータアクセス可能な媒体に実装されるインストラクション及び/またはデータを、受信すること、送信すること、または記憶することをさらに含んでもよい。一般的に言えば、コンピュータアクセス可能な媒体は、磁気もしくは光学メディア、例えば、ディスクもしくはDVD/CD-ROM、RAM（例えば、SDRAM、DDR、DRAM、及びSRAM等）等の不揮発性メディア、ならびにROM等の記憶メディアもしくはメモリメディア、及びネットワークならびに/または無線リンク等の通信媒体を介して伝達される、電気、電磁、もしくはデジタル信号等の伝送メディアもしくは信号を含んでもよい。

40

【0080】

図に示され、かつ明細書中に記載される様々な方法は、方法の例示的な実施形態を表す。本方法は、ソフトウェア、ハードウェア、またはそれらの組み合わせ内に実装してもよい。方法の順序を変更することができ、かつ様々な要素を追加して、再度順序付けする、組み合わせる、除外する、及び修正すること等を行うことができる。

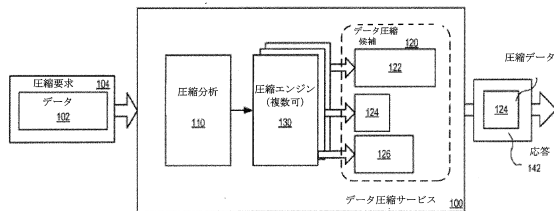
50

【 0 0 8 1 】

本開示の利益を享受する当業者にとって明白である、様々な修正及び変更を加えることができる。本発明は、そのような全ての修正及び変更を包含し、したがって、以上の記載は、限定的な意味ではなく、例示的な意味で考慮されるべきものであることを意図する。

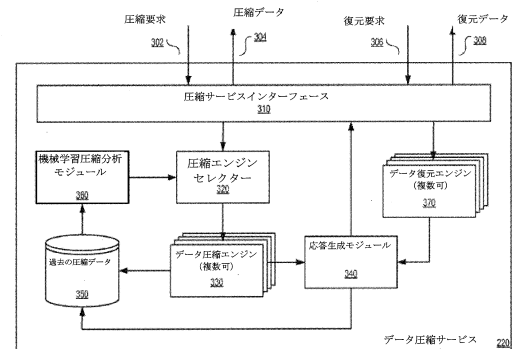
【 図 1 】

図 1



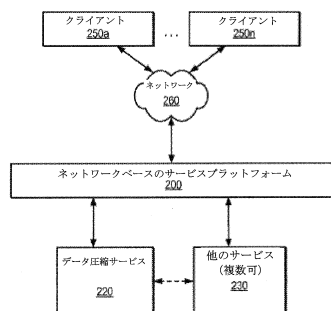
【 図 3 】

図 3



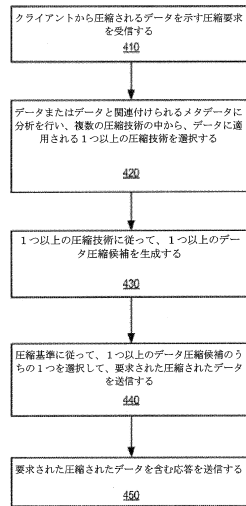
【 図 2 】

図 2



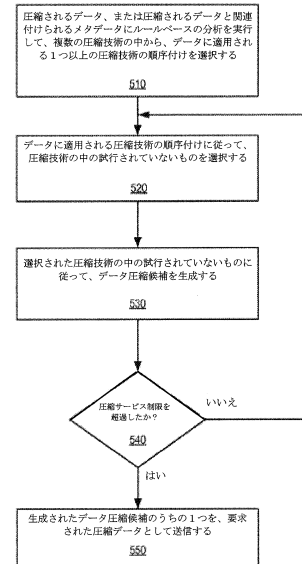
【図 4】

図 4



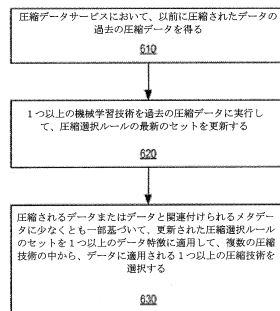
【図 5】

図 5



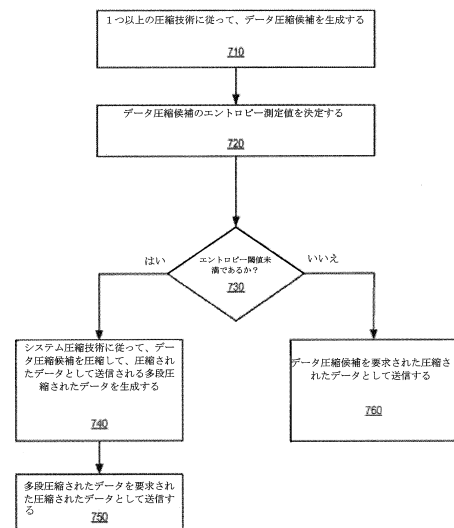
【図 6】

図 6



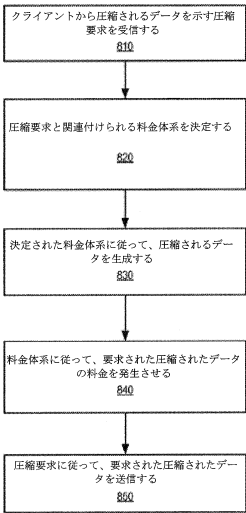
【図 7】

図 7



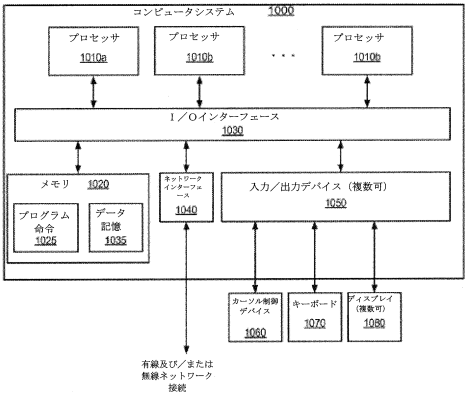
【図 8】

図8



【図 9】

図9



フロントページの続き

(72)発明者 グブタ, アヌラグ・ウィンドラス

アメリカ合衆国・98109-5210・ワシントン州・シアトル・テリー アヴェニュー ノース
・410

審査官 月野 洋一郎

(56)参考文献 特開2012-039198(JP, A)

米国特許出願公開第2012/0284239(US, A1)

特開2007-043673(JP, A)

特開2005-288157(JP, A)

特開2004-112497(JP, A)

特開2002-223220(JP, A)

特開2005-128969(JP, A)

国際公開第2011/129818(WO, A1)

米国特許出願公開第2010/0281079(US, A1)

(58)調査した分野(Int.Cl., DB名)

G06F 12/00

G06F 9/445

G06F 13/00