(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7266354号
(P7266354)

(45)発行日　令和5年4月28日(2023.4.28)

(24)登録日　令和5年4月20日(2023.4.20)

(51)国際特許分類

Ｇ０６Ｆ　21/62　(2013.01)	Ｇ０６Ｆ	21/62	３５４
Ｇ０６Ｆ　16/24　(2019.01)	Ｇ０６Ｆ	16/24	
Ｇ０６Ｆ　16/28　(2019.01)	Ｇ０６Ｆ	16/28	

ＦＩ

請求項の数　9　（全22頁）

(21)出願番号　特願2020-545618(P2020-545618)
(86)(22)出願日　平成31年3月19日(2019.3.19)
(65)公表番号　特表2021-516811(P2021-516811A)
(43)公表日　令和3年7月8日(2021.7.8)
(86)国際出願番号　PCT/IB2019/052201
(87)国際公開番号　WO2019/180599
(87)国際公開日　令和1年9月26日(2019.9.26)
審査請求日　令和3年8月16日(2021.8.16)
(31)優先権主張番号　15/924,345
(32)優先日　平成30年3月19日(2018.3.19)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国１０５０４　ニューヨーク州　アーモンク　ニュー　オーチャード　ロード
Ｎｅｗ　Ｏｒｃｈａｒｄ　Ｒｏａｄ，　Ａｒｍｏｎｋ，　Ｎｅｗ　Ｙｏｒｋ　１０５０４，　Ｕｎｉｔｅｄ　Ｓｔａｔｅｓ　ｏｆ　Ａｍｅｒｉｃａ
(74)代理人　100112690
弁理士　太佐 種一

最終頁に続く

(54)【発明の名称】　データ匿名化

(57)【特許請求の範囲】
【請求項１】
コンピュータの情報処理により実行される、データベース・システムのデータ匿名化のための方法であって、
前記データベース・システムの第１のデータセットと第２のデータセットとが、前記第１のデータセット及び前記第２のデータセットの両方における値を有するエンティティを示す関係を有するかどうかを、ＢＩＮＤＥＲアルゴリズム又はＭｉｎｈａｓｈ技術を用いて判断することと、
前記第１のデータセット及び前記第２のデータセットの少なくとも一方に対する要求を、ユーザから受け取ることと、
前記第１のデータセットと前記第２のデータセットとが前記関係を有すると判断されたことに基づいて、前記要求から、匿名化アルゴリズムを用いて、修正された要求されたデータセットを生成することであって、前記修正された要求されたデータセットは、前記修正された要求されたデータセットにおいて統治ポリシーに違反するエンティティにアクセスできないように、前記第１のデータセット及び前記第２のデータセットの少なくとも一方を修正することによって生成され、前記統治ポリシーは、前記データベース・システムへの選択的アクセスを可能にする１つ又は複数の規則を含む、生成することと、
前記修正された要求されたデータセットを提供することと、
を含む、方法。
【請求項２】

前記第１のデータセット及び前記第２のデータセットはレコードを含み、前記レコードのそれぞれはそれぞれのエンティティの属性値の組み合わせであり、前記それぞれのエンティティは、前記第１のデータセット及び前記第２のデータセットの少なくとも一方のレコードの前記エンティティを示す前記関係を有する、請求項１に記載の方法。

【請求項３】

前記データベース・システムの全てのデータセットの間の関係を判断することと、判断した前記関係についての情報を含むメタデータ構造を提供することとをさらに含み、前記第１のデータセットと前記第２のデータセットとが前記関係を有するかどうかを判断することは、前記メタデータ構造を用いて行われる、請求項１又は請求項２に記載の方法。

【請求項４】

前記データベース・システムにおける変更に応答して、前記データベース・システムの前記第１のデータセットと前記第２のデータセットとの間の前記関係を再判断することと、それに応じて前記メタデータ構造を更新することとをさらに含む、請求項３に記載の方法。

【請求項５】

前記第１のデータセット及び前記第２のデータセットの少なくとも一方の修正を行うことは、ユーザによる前記エンティティへのアクセスが前記統治ポリシーに違反するとの判断に応答して行われる、請求項１から請求項４までのいずれかに記載の方法。

【請求項６】

前記要求を受け取ったこと、並びに前記第１のデータセット及び前記第２のデータセットがユーザによりアクセス可能であるとの判断に応答して、前記第１のデータセットと前記第２のデータセットとが前記関係を有するかどうかの判断を行うことをさらに含む、請求項１から請求項５までのいずれかに記載の方法。

【請求項７】

前記修正された要求されたデータセットを生成することは、前記第１のデータセット及び前記第２のデータセットの少なくとも一方の１以上のカラムをマスキングすることを含む、請求項１から請求項６までのいずれかに記載の方法。

【請求項８】

請求項１から請求項７までのいずれかに記載の方法をコンピュータに実行させるコンピュータ・プログラム。

【請求項９】

請求項８に記載のコンピュータ・プログラムを格納したコンピュータ可読ストレージ媒体。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、デジタル・コンピュータ・システムの分野に関し、より具体的には、データ匿名化（data anonymization）のための方法に関する。

【背景技術】

【０００２】

プライバシー規則は、特定の種類の分析のために個人のデータを使用することを、その個人たちがそうした使用への同意を明示的に宣言したのでない限り許可しない。一方、データが匿名化される限り、個人データを収集及び格納し、それを分析で使用することは容認し得る。例えば、コントローラにより処理されるデータが、コントローラに個人を特定させないものである場合には、データを分析することは可能である。しかしながら、これは、匿名化を正しくかつ十分に適用して、プライベート・データと関連付けられたエンティティの特定を防止することを必要とする。

【発明の概要】

【０００３】

種々の実施形態が、独立請求項の主題により説明されるデータ匿名化のための方法、コンピュータ・システム、及びコンピュータ・プログラム製品を提供する。従属請求項にお

いて、有利な実施形態が説明される。本発明の実施形態は、それらが相互排他的でない場合、互いに自由に組み合わせることができる。

【０００４】

１つの態様において、本発明は、データベース・システムのデータ匿名化のための方法に関する。この方法は、

（ａ）データベース・システムの第１のデータセットと第２のデータセットとが、２つのデータセットにおける値（例えば、属性値）を有するエンティティを示す関係を有するかどうかを判断することと、

（ｂ）第１のデータセット及び第２のデータセットの少なくとも一方に対する要求を、ユーザから受け取ることと、

（ｃ）第１のデータセットと第２のデータセットとが上記関係を有する場合、ユーザがエンティティを示すものにアクセスできないように、第１のデータセット及び第２のデータセットの少なくとも一方を修正することと、

（ｄ）要求されたデータセットを提供することと、

を含む。

【０００５】

別の態様において、本発明は、前述の実施形態による方法のステップの全てを実施するように構成されたコンピュータ可読プログラム・コードが具体化されたコンピュータ可読ストレージ媒体を含むコンピュータ・プログラム製品に関する。

【０００６】

別の態様において、本発明は、データベース・システムのデータ匿名化のためのコンピュータ・システムに関する。コンピュータ・システムは、

（ａ）データベース・システムの第１のデータセットと第２のデータセットとが、２つのデータセットにおける値を有するエンティティを示す関係を有するかどうかを判断することと、

（ｂ）第１のデータセット及び第２のデータセットの少なくとも一方に対する要求を、ユーザから受け取ることと、

（ｃ）第１のデータセットと第２のデータセットとが上記関係を有する場合、ユーザがエンティティを示すものにアクセスできないように、第１のデータセット及び第２のデータセットの少なくとも一方を修正することと、

（ｄ）要求されたデータセットを提供することと、

を行うように構成される。

【図面の簡単な説明】

【０００７】

以下に、本発明の実施形態が、図面を参照して、単なる例としてより詳細に説明される。

【０００８】

【図１】本開示による、ストレージ・システムのブロック図を示す。

【図２】データベース・システムのデータのデータ匿名化のための方法のフローチャートである。

【図３】本発明の実施形態による、クラウド・コンピューティング環境を示す。

【図４】本発明の実施形態による、抽象化モデル層を示す。

【発明を実施するための形態】

【０００９】

本発明の種々の実施形態の説明は、例証の目的のために提示されるが、これらは、網羅的であること、又は開示された実施形態に限定することを意図するものではない。当業者には、説明される実施形態の範囲及び趣旨から逸脱することなく、多くの修正及び変形が明らかであろう。本明細書で用いられる用語は、実施形態の原理、実際の適用、又は市場に見られる技術に優る技術的改善を最もよく説明するため、又は、当業者が、本明細書に開示される実施形態を理解するのを可能にするために選択された。

【００１０】

「データセット」又は「情報アセット（information asset）」という用語は、１以上のデータ要素の集合である。データ要素は、例えば、文書、データ値、又はデータ・レコードとすることができる。例えば、データセットは、ファイル内に含まれる関連するレコードの集合の形態で提供することができ、例えば、データセットは、クラス内の全ての学生のレコードを含むファイルとすることができる。レコードは、関連データ項目、例えば、学生のロール番号、生年月日、クラスの集合である。レコードはエンティティを表し、そこで、エンティティは、例えば１人の学生のような異なる別個の存在を有する。データセットは、例えば、データベースのテーブル又はＨａｄｏｏｐファイル・システムのファイル等とすることができる。別の例において、データセットは、ＨＴＭＬページ又は他の文書の種類などの文書を含むことができる。文書は、例えば、患者のデータを含むことができる。

【００１１】

例えば、２つのデータセットにおける値(values)を有するエンティティは、例えば、第１のデータセット及び第２のデータセットの少なくとも一方における少なくとも１つのレコードを有するエンティティであり得る。例えば、関係は、第１のデータセット内のレコードを有する患者Ｘが、第２のデータセット内の別の関連するレコードも有する（例えば、患者Ｘの２つのレコードは、そのアドレスによりリンクされることができ、そこで、第１のレコードはフルネームを有さないが、第２のレコードは患者Ｘのフルネームを含む）ことを示し得る。従って、患者Ｘは、第１のデータセットと第２のデータセットとの間の上記関係により指し示されるエンティティである。

【００１２】

２つのデータセット間の上記関係は、データセットの１以上のカラム（column）／属性と他のデータセットの他の１以上のカラムとの間のリンクとすることができる。例えば、上記関係は、主キー・外部キー（ＰＫ－ＦＫ）関係とすることができる。別の例において、上記関係は、同じエンティティに関するより多い情報を含む別の文書を指し示すＸＭＬ文書からのリンク、又は関連情報を含む非構造化文書からのリンクなどの、１つの文書から別の文書へのリンクを含むことができる。データがトリプルストア内のトリプレット（例えば、ＲＤＦデータ）として格納されるか、又はデータがグラフＤＢ内に格納される場合、上記関係は、１つのエンティティと別のエンティティとの間のリンクとすることができる。上記関係は、例えば、第１のデータセットと第２のデータセットを結合するのを可能にできるので、同じエンティティを表す第１のデータセットのレコードと第２のデータセットのレコードは、同じエンティティについての２つのデータセット内に含まれる組み合わせられた情報を表す結合されたデータセットの新しい単一のレコード内に併合される。２つのデータセットはそれぞれ匿名化されているが、組み合わせられた情報は、そのエンティティの秘密（confidential）のデータを明らかにし得る。

【００１３】

「ユーザ」という用語は、エンティティ、例えば、個人、コンピュータ、又はコンピュータ、コンテナ、ファイル・システム、ディレクトリ上で実行されているアプリケーションなどを指す。ユーザは、例えば、ユーザのグループを表すことができる。データベース・システムは、データセットを格納するための１以上のストレージを含むことができる。データベース・システムは、例えば、文書ストア、トリプルストア、グラフＤＢ、及びリレーショナル・データベースの少なくとも１つを含むことができる。第１のデータセット及び第２のデータセットは、データベース・システムの同じ又は異なるストレージ上に格納され得る。

【００１４】

処理のために情報アセットにアクセスするとき、例えば、チェックを行って、アセットが機密（sensitive）情報を含み、匿名化を必要とするかどうかを確かめることがある。必要である場合、所定のデータ・マスキング技術を用いる適切な匿名化方策が適用される。組み合わせられた使用のために２以上の情報アセットがマーク付けされる場合、それが匿名化の潜在的な違反をもたらすかどうかのチェックが行われる。違反をもたらす場合、所

定のマスキング技術を用いる適切な匿名化方策が適用される。チェックは、情報アセットのために確立され、情報ガバナンスカタログ（information governance catalog）内に登録された技術メタデータに対して実施され得る。

【００１５】

本方法は、それぞれ個別に十分に匿名化された２以上の情報アセットが集められるが、一緒に用いられる情報アセットによりデータ匿名化が破られた場合に、データ匿名化の喪失を防止することができる。本方法は、データレイク内の大規模Ｈａｄｏｏｐクラスタのような単一のシステムが、データ・サイエンティストが使用可能な共に分析されるべき、および事前に作成することができない多くの組み合わせを有する何万もの情報アセットを格納する場合に、特に、ビッグデータ及び大規模データレイク・アーキテクチャの場合において、特に有利であり得る。こうしたシステムのために、本方法は、１度のデータ分析において、情報アセットを一緒に用い得るかを予想する手法を提供することができる。例えば、本本法は、匿名化の違反を回避することができ、そこで、匿名化の違反は、以下の特性を有し得る。２以上の情報アセットにわたり、１以上の属性を結合することができる。２以上の情報アセットにわたり、１つの情報アセットにおいて、情報ガバナンス・ポリシーに従って、特定のエンティティ又は属性グループ又は属性をマスキングする必要があった。また、２以上の情報アセットにわたり、少なくとも１つの情報アセットにおいて、他のアセットの１つにおいて保護されるドメインがマスキングされず、ひとたび結合されると匿名化に違反することが可能になる。

【００１６】

別の利点は、本発明が、必要なところのデータ匿名化を保証し、事前対応の（pro-active）自動化されたデータ保護方策を可能にでき、データ保護とデータ有用性との間の最適なバランスを提供することであり得る。例えば、２つのデータセットは、それぞれ匿名化することができ、各々が互いに独立してアクセスされた場合にはいずれの機密情報のソースともすることができない。しかしながら、ユーザは、第１のデータセット及び第２のデータセットを組み合わせてエンティティにアクセスできるので、本方法は、２つのデータセットの１つの要求の受け取り時に既に修正を行うことによって、事後に（a posteriori）動作し、それにより、ユーザが後の段階において第２のデータセットを別個に要求できる場合に備える。言い換えれば、これは、データ・サイエンティストが匿名化されたデータセットを取得し、各データセットがガバナンス・ポリシーに適合するが、複数のポリシー適合データセットを一緒に結合し、もはやポリシーに適合しない結果とし得られるデータセットを取得することにより、匿名化に対処できてしまう場合を防止することができる。

【００１７】

本方法はさらに、信用の喪失もしくは罰金又はその両方をもたらす恐れがある意図的でないデータ漏洩を防止することができる。

【００１８】

別の利点は、本方法は、プロセスを、完全なデータ系列及び他の監査証跡（audit trail）を含む情報ガバナンスカタログに基づくメタデータ駆動とすることができるので、監査が容易な改善された規則適合性を提供する。１つの実施形態によると、第１のデータセット及び第２のデータセットは、各々がそれぞれのエンティティの属性値の組み合わせであるレコードを含み、関係により指し示されるエンティティは、第１のデータセットもしくは第２のデータセット又はその両方の少なくとも１つのレコードのエンティティである。例えば、データベース・システムのデータ匿名化のための例示的方法を提供することができる。例示的方法は、データベース・システムの第１のデータセットと第２のデータセットとが、２つのデータセットの少なくとも一方のレコードのエンティティを示す関係を有するかどうかを判断することと、第１のデータセット及び第２のデータセットの少なくとも一方に対する要求を、ユーザから受け取ることと、第１のデータセットと第２のデータセットとが上記関係を有する場合、ユーザがエンティティを示すものにアクセスできないように、第１のデータセット及び第２のデータセットの少なくとも一方を修正することと、要求されたデータセットを提供することとを含む。

【００１９】
１つの実施形態によると、方法は、データベース・システムの全てのデータセットの間の関係を判断することと、判断した上記関係に関する情報を含むメタデータ構造を提供することとをさらに含み、第１のデータセットと第２のデータセットが上記関係を有するかどうかを判断することは、メタデータ構造を用いて実行される。このように、方法は、上記関係の判断に対する明白なオンデマンドの必要性なしに、自動的に実行することができる。このことは、単位時間毎に多くのデータ要求を受け取る大規模なシステムにおいて特に有利であり得る。メタデータ構造が今回に限り作成され、各々の受け取った要求について再処理されないので、これは処理時間の節約になり得る。
【００２０】
１つの実施形態によると、方法は、データベース・システムにおける変更に応答して、データベース・システムのデータセット間の上記関係を再判断することと、これに応じてメタデータ構造を更新することとをさらに含む。このことは、最新の情報ソースを提供することができ、従って、正確な匿名化を実行することにより、データへのアクセスをさらにセキュア保護することができる。
【００２１】
例えば、所定の関数（例えば、ＰＫ－ＦＫ関係発見アルゴリズム）を用いて、データベース・システム内の全てのデータセット間の全ての可能な上記関係（例えば、ＰＫ－ＦＫ関係）を判断し、全てのこれらの判断した上記関係をメタデータ構造に格納することができる。新しいデータセットが付加されるときには、上記関係の少なくとも一方の側が新しいデータセット内にある場合の上記関係の識別に焦点を合わせて、同じ関数を再実行することができる。それに応じて、メタデータ構造を更新することができる。データセットが除去されるときには、除去されたデータセットに関わる可能な上記関係又は全ての上記関係をリストから除去することができる。この実施形態は、バックグラウンドにおいて連続的に実行して、データベース・システムにおいて利用可能なデータベースのリストにおける変更を検出することができ、ＰＫ－ＦＫ関係発見アルゴリズムをトリガして、変更が検出されるや否や、上記関係のリストを更新することができる。
【００２２】
１つの実施形態によると、修正は、ユーザによるエンティティへのアクセスが所定の統治（governing）ポリシー（又は規則）に違反するとの判断に応答して行われる。これは、データへの選択的アクセスを可能にし、従って、データ・アクセスの最適な制御を可能にする。
【００２３】
一例として、統治ポリシーは、「ユーザがロールＡを持ち、データセットが、機密の個人を特定できる情報（personally identifiable information）であるとしてカタログ内にフラグが立てられたカラムを含み、データセットは、識別子又は準識別子としてフラグが立てられたカラムも含む場合、データセットを匿名化する必要がある」と指定することができる。例えば、所与の（特権をもつ）ユーザについては、修正が行われなくてもよく、修正なしにデータを提供することができる。しかしながら、他の（信頼できない）ユーザについては、修正が行われる。この場合、統治ポリシーは、エンティティ（例えば、個人のフルネーム）が、他のタイプのユーザによってではなく、所与のタイプのユーザによってアクセスされることを求める。
【００２４】
１つの実施形態によると、上記関係がエンティティを示すかどうかを判断することは、第２のデータセットのそれぞれのターゲット・カラムを参照する第１のデータセットの１以上のソース・カラムを識別することと、ソース・カラム及びターゲット・カラムを組み合わせることと、組み合わせ結果に基づいて、上記関係がエンティティを示すか又は示さないかを判断することとを含む。１つの実施形態によると、組み合わせは、１以上のＳＱＬ結合（join）演算を用いて行われる。これは、これらの実施形態と、こうしたシステム内のデータへのアクセスをセキュア保護するための既存のデータベース・システムとのシ

ームレスな統合を可能にすることができる。
【００２５】
１つの実施形態によると、判断される上記関係は、主キー・外部キー関係である。ＰＫ－ＦＫ関係は、カラムの対又はカラムのグループの対で構成され、それらの間に包含従属性が存在する。付加的に、主キーを形成するカラム及びカラムのグループは一意であり得る。データセットのグループにおける包含従属性の検索は、例えば、それらをＭｉｎＨａｓｈ又はドメイン署名技術と組み合わせ、カラムの濃度を用いて、上記関係の一方の側が一意又はほぼ一意である場合の組み合わせに検索を制限することにより実行することができる。これは、完全に自動的な方法で、妥当な期間に、データセットのグループの全ての可能なＰＫ－ＦＫ関係、従って、データセットのグループ内のデータを結合する全ての可能な方法を判断するのを可能にする機構を可能にし得る。
【００２６】
１つの実施形態によると、方法は、要求を受信することと、第１のデータセット及び第２のデータセットがユーザによりアクセス可能であると判断することとに応答して、第１のデータセットと第２のデータセットとが上記関係を有するかどうかの判断を行うことをさらに含む。これは、特定の条件下でのみ、データセットの判断が行われる要求ごとの手法を可能にし得る。これは、データの一部のみが使用される場合、全てのデータについての上記関係を自動的に判断する必要がないので、データへのアクセス頻度が低いシステムの場合、特に有利であり得る。これは、処理リソースを節約することができる。
【００２７】
１つの実施形態によると、第１のデータセットもしくは第２のデータセット又はその両方は、要求されたデータセットの１以上のカラムをマスキングすることを含む。第１のデータセットもしくは第２のデータセット又はその両方の修正は、匿名化アルゴリズムを用いて行われる。匿名化アルゴリズムは、以下の、一般化（generalization）、黒塗り（redaction）、抑制、サンプリング、ランダム化、データ・スワッピング、マスキング、列挙（enumeration）のうちの少なくとも１つである。
【００２８】
この実施形態は、関心あるデータセット内の個人を特定できる情報を、例えば単一の人を特定できる情報を省略し、同時に、分析に有用な情報を保持できるような方法で修正することができるという利点を有することができる。
【００２９】
１つの実施形態によると、上記関係の判断は、自動的に行われる。例えば、上記関係の判断は、ある期間ごとに行うことができる。
【００３０】
１つの実施形態によると、上記関係の判断は、データベース・システムにおける変更の検出に応答して自動的に行われる。例えば、データベース・システムにおける変更は、データベース・システムへの第１のデータセットもしくは第２のデータセット又はその両方の少なくとも１つの付加、又は第１のデータセットもしくは第２のデータセット又はその両方における変更を含むことができる。
【００３１】
図１は、本開示に含まれる方法ステップを実施するのに適した一般的なコンピュータ化されたシステム１００を表す。
【００３２】
本明細書で説明される方法は、少なくとも部分的に非対話型であり、サーバ又は組み込みシステムなどのコンピュータ化されたシステムによって自動化されることが理解されるであろう。しかしながら、例示的実施形態においては、本明細書で説明される方法は、（部分的に）対話型システムで実施することができる。これらの方法はさらに、ソフトウェア１１２、１２２（ファームウェア１２２を含む）、ハードウェア（プロセッサ）１０５、又はその組み合わせで実施してもよい。例示的実施形態において、本明細書で説明される方法は、実行可能プログラムとしてソフトウェアで実施され、パーソナル・コンピュー

タ、ワークステーション、ミニコンピュータ、又はメインフレーム・コンピュータなどの専用又は汎用デジタル・コンピュータにより実行される。従って、最も一般的なシステム１００は、汎用コンピュータ１０１を含む。

【００３３】

例示的実施形態において、ハードウェア・アーキテクチャの点で、図１に示されるように、コンピュータ１０１は、プロセッサ１０５、メモリ・コントローラ１１５に結合されたメモリ（主メモリ）１１０、及びローカル入力／出力コントローラ１３５を介して通信可能に結合された１以上の入力もしくは出力又はその両方の（Ｉ／Ｏ）デバイス（又は機器）２０、１４５を含む。入力／出力コントローラ１３５は、これらに限定されるものではないが、当技術分野で知られるような、１以上のバス、又は他の有線もしくは無線接続とすることができる。入力／出力コントローラ１３５は、通信を可能にするための、コントローラ、バッファ（キャッシュ）、ドライバ、中継器、及び受信機などの付加的な要素を有し得るが、それらは簡単にするために省略される。さらに、ローカル・インターフェースは、上述のコンポーネント間での適切な通信を可能にするために、アドレス、制御もしくはデータ接続又はそれらの組み合わせを含むことができる。本明細書で説明されるように、Ｉ／Ｏデバイス２０、１４５は、一般に、当技術分野で知られている任意の一般化された暗号カード又はスマートカードを含むことができる。

【００３４】

プロセッサ１０５は、特にメモリ１１０内に格納されるソフトウェアを実行するためのハードウェア・デバイスである。プロセッサ１０５は、任意の特注又は市販のプロセッサ、中央処理ユニット（ＣＰＵ）、コンピュータ１０１と関連付けられた幾つかのプロセッサの中の補助プロセッサ、半導体ベースのマイクロプロセッサ（マイクロチップ又はチップセットの形態の）、マクロプロセッサ、又は一般的にソフトウェア命令を実行するための任意のデバイスとすることができる。

【００３５】

メモリ１１０は、揮発性メモリ素子（例えば、ランダム・アクセス・メモリ（ＤＲＡＭ、ＳＲＡＭ、ＳＤＲＡＭ等のようなＲＡＭ））及び不揮発性メモリ素子（例えば、ＲＯＭ、消去可能プログラム可能読み出し専用メモリ（ＥＰＲＯＭ）、電子的消去可能プログラム可能読み取り専用メモリ（ＥＥＰＲＯＭ）、プログラム可能読み取り専用メモリ（ＰＲＯＭ））のいずれか１つ又はそれらの組み合わせを含むことができる。メモリ１１０は、分散アーキテクチャを有することができ、種々のコンポーネントが、互いから遠隔に位置するが、プロセッサ１０５によりアクセスできることに留意されたい。

【００３６】

メモリ１１０内のソフトウェアは、１以上の別個のプログラムを含むことができ、その各々は、論理関数、とりわけ本発明の実施形態に含まれる関数を実施するための実行可能命令の順序付きリストを含む。図１の例において、メモリ１１０内のソフトウェアは、例えば、データベース管理システムなどのデータベースを管理するための命令などの命令１１２を含む。

【００３７】

メモリ１１０内のソフトウェアは、典型的には、適切なオペレーティング・システム（ＯＳ）１１１も含むことになる。ＯＳ１１１は、本明細書で説明される方法を実施するための潜在的なソフトウェア１１２など、他のコンピュータ・プログラムの実行を本質的に制御する。

【００３８】

本明細書で説明される方法は、ソースプログラム１１２、実行可能プログラム１１２（オブジェクト・コード）、スクリプト、又は実行される命令１１２のセットを含む任意の他のエンティティの形とすることができる。ソースプログラムの場合、プログラムは、ＯＳ１１１と関連して適切に動作するように、メモリ１１０中に含まれていても又は含まれていなくてもよいコンパイラ、アセンブラ、インタープリタ等を介して変換する必要がある。さらに、方法は、データ及び方法のクラスを有するオブジェクト指向プログラミング

言語、又はルーチン、サブルーチン、もしくは関数又はそれらの組み合わせを有する手続き型プログラミング言語として記述することができる。

【００３９】

例示的実施形態において、従来型のキーボード１５０及びマウス１５５を入力／出力コントローラ１３５に結合することができる。Ｉ／Ｏデバイス１４５など、他の出力デバイスは、例えば、これらに限定されるものではないが、プリンタ、スキャナ、マイクロホン等などの入力デバイスを含むことができる。最後に、Ｉ／Ｏデバイス２０、１４５は、入力及び出力の両方を通信するデバイス、例えば、これらに限定されるものではないが、（他のファイル、デバイス、システム、又はネットワークにアクセスするための）ネットワーク・インターフェース・カード（ＮＩＣ）又は変調器／復調器、無線周波数（ＲＦ）、又は他の送受信機、電話インターフェース、ブリッジ、ルータ等をさらに含むことができる。Ｉ／Ｏデバイス２０、１４５は、当技術分野で知られている任意の一般的な暗号カード又はスマートカードとすることができる。システム１００は、ディスプレイ１３０に結合されたディスプレイ・コントローラ１２５をさらに含むことができる。例示的実施形態において、システム１００は、ネットワーク１６５に結合するためのネットワーク・インターフェースをさらに含むことができる。ネットワーク１６５は、コンピュータ１０１と任意の外部サーバ、クライアント等との間の広帯域接続を介した通信のためのＩＰベースのネットワークとすることができる。ネットワーク１６５は、コンピュータ１０１と外部システム３０との間でデータを送信及び受信し、これら外部システムは、本明細書で説明される方法のステップの一部又は全てを実行することに関与することが可能である。例示的実施形態において、ネットワーク１６５は、サービス・プロバイダによって管理される管理（managed）ＩＰネットワークとすることができる。ネットワーク１６５は、例えば、ＷｉＦｉ、ＷｉＭａｘなどの無線プロトコル及び技術を用いて、無線方式で実施することが可能である。また、ネットワーク１６５は、ローカル・エリア・ネットワーク、広域ネットワーク、メトロポリタン・エリア・ネットワーク、インターネット・ネットワーク、又は他の類似のタイプのネットワーク環境など、パケット交換網ネットワークとすることもできる。ネットワーク１６５は、固定無線ネットワーク、無線ローカル・エリア・ネットワーク（ＬＡＮ）、無線広域ネットワーク（ＷＡＮ）、パーソナル・エリア・ネットワーク（ＰＡＮ）、仮想私設ネットワーク（ＶＰＮ）、イントラネット、又は他の好適なネットワーク・システムとすることができ、信号を受信及び送信するための機器を含む。

【００４０】

コンピュータ１０１がＰＣ、ワークステーション、インテリジェント・デバイス等である場合には、メモリ１１０内のソフトウェアは、基本入力出力システム（ＢＩＯＳ）１２２をさらに含むことができる。ＢＩＯＳは、立ち上げ時にハードウェアを初期化及びテストし、ＯＳ１１１を開始し、ハードウェア・デバイスの中にあるデータの転送をサポートする基本的なソフトウェア・ルーチンのセットである。ＢＩＯＳは、コンピュータ１０１が起動されたときにＢＩＯＳが実行できるように、ＲＯＭの中に格納される。

【００４１】

コンピュータ１０１が動作しているとき、プロセッサ１０５は、メモリ１１０内に格納されたソフトウェア１１２を実行し、メモリ１１０との間でデータを通信し、ソフトウェアに従ってコンピュータ１０１の動作を全般的に制御するように構成される。本明細書で説明される方法及びＯＳ１１１は、全体的に又は部分的にだが一般的には後者で、プロセッサ１０５によって読み取られ、恐らくはプロセッサ１０５内にバッファされ、次いで実行される。

【００４２】

図１に示されるように、本明細書で説明されるシステム及び方法がソフトウェア１１２に実装される場合、これら方法は、何らかのコンピュータ関連システム又は方法によって、又はこれと関連させて使用するために、ストレージ１２０など、任意のコンピュータ可読媒体上に格納することができる。ストレージ１２０は、ＨＤＤストレージなどディスク・ストレージを含むことができる。

【００４３】
システム１００は、データベース・システム１５０をさらに含む。データベース・システム１５０は、１５１．１～１５１．Ｎを含む。メタデータ記述又はデータセット１５１．１～Ｎを示すものは、カタログ１５３に格納することができる。カタログ１５３は、例えば、データセット１５１．１～Ｎのデータ・プロファイルを含むことができる。データ・プロファイルは、どの意味領域内に、特定の属性又は属性のグループが属するかを指し示すことができる。カタログ１５３は、データセット１５１．１～Ｎに関する分類情報をさらに含むことができる。例えば、所定のデータ分類分析関数は、データセット１５１．１～Ｎの各カラムをカテゴリに割り当てることができ、例えば、各カテゴリを分類識別子により識別することができる。カタログ１５３は、例えば、各々の分類識別子及び関連したカラムを格納することができる。カタログ１５３は、ガバナンス・ポリシーをさらに含むことができる。ガバナンス・ポリシーは、例えば、どの属性が匿名化による保護を必要とするか（例えば、クレジットカード番号、個人の名前及びアドレス等）、並びにどの匿名化アルゴリズムを使用するかを指し示すことができる。カタログ１５３は、データモデルをさらに含むことができ、データモデルは、データがどのように構造化され、マッピングされ、リンクされるかについての詳細を提供する。データベース・システム１５０は、単に例示のために単一のコンポーネントとして示される。しかしながら、データベース・システムの他の例を用いることもできる。例えば、データベース・システム１５０は、複数のストレージを含むことができる。複数のストレージは、互いに接続されていても又は接続されていなくてもよい。

【００４４】
図２は、データベース・システム１５０のデータのデータ匿名化のための方法のフローチャートである。

【００４５】
ステップ２０１において、データベース・システム１５０の、例えば１５１．２などの第１のデータセットと、例えば１５１．４などの少なくとも１つの第２のデータセットとが、２つのデータセットの少なくとも１つのレコードのエンティティを示す（又は、２つのデータセットの少なくとも１つにおける値を有するエンティティを示す）関係を有するかどうかを判断することができる。エンティティを示すもの（例えば、個人のフルネームなど）は、ガバナンス・ポリシーを満たさず、従って、匿名化を必要とし得る。データベース・システムが複数のストレージを含む場合には、第１のデータセット及び第２のデータセットを同じ又は異なるストレージ上に格納することができる。上記関係は、例えば、第１のデータセット及び第２のデータセットにおける属性値を結合することを可能にし、結合結果が同じエンティティを表すようにできる（例えば、以下に説明される通話詳細レコードの例を参照されたい）。結合した属性値は、両方のデータセット内の示されるエンティティに属すること又は２つのデータセットの一方における示されるエンティティに属することができ、他のデータセットは、例えば示されるエンティティに関連する別のエンティティの属性値を含むことができる。例えば、第１のデータセットは、所与のアドレスを有する患者Ｘの属性値を含む患者のカルテであり、第２のデータセット（例えば、ソーシャルメディア・プロファイル）は、ファーストネーム及び同じ所与のアドレスと関連付けられた患者Ｘの親類の属性値を含むことがある。２つのデータセットの組み合わせは、患者Ｘのファーストネームを明らかにすることがある。この例において、示されるエンティティは患者Ｘであり、他のエンティティは親類である。

【００４６】
例えば、ステップ２０１の判断は、例えばステップ２０３の要求の受信時にオンデマンドで、２つのデータセット１５１．２及び１５２．４に対して行うことができる。別の例において、ステップ２０１の判断は、データベース・システム１５０の全てのデータセット１５１．１～Ｎの上記関係の全体の判断の一部として行うことができる。これは、データベース・システム１５０の２つのデータセット１５１．２及び１５２．４、並びに他のデータセットに対して自動的に行うことができる。例えば、ステップ２０１の自動実行は

、例えば毎日など時間で、又は例えばデータセットが変更され、新しいデータセットが付加されるなど、データベース・システム１５０における変更の検出時に行うことができる。

【００４７】

例えば、ステップ２０１は、最初に、第１のデータセット１５１．２及び第２のデータセット１５１．４の少なくとも一方だけを匿名化方式で使用できるかどうかをチェックすることにより、行うことができる。このチェックは、所定の統治ポリシーもしくはカタログ１５３のデータモデル又はその両方に対して行うことができる。第１のデータセット１５１．２及び第２のデータセット１５１．４のいずれも匿名化方式で使用可能でない場合には、方法は停止する。

【００４８】

１つの例において、ステップ２０１の判断は、匿名化されていない第１のデータセット１５１．２又は第２のデータセット１５１．４の少なくとも一方の属性が、意味的に同じ企業体（business entity）又は属性グループ又は属性を表すかどうかをチェックすることにより、行うことができ、それを用いて、２つのデータセット１５１．２及び１５１．４にわたって個々のレコードを結合することができる。これは、例えばデータセットにわたるカタログ１５３のデータ用語分類情報を用いて、同じ用語分類器がデータセットにわたって使用されるかどうかを判断すること、もしくはデータセット１５１．２及び１５１．４にわたるデータ・プロファイリングの結果を用いて、データセットの間にＰＫ／ＦＫ制約（例えば、包含従属性）が見られるかどうかを判断すること、又はその両方によって、行うことができる。これは、特定の属性についての情報アセット間の結合動作を実行するために特定の属性を用いることができることを示唆し得る。

【００４９】

別の例において、ステップ２０１の判断は、データセット１５１．２及び１５１．４にわたって、それらのいずれかが、匿名化方式のみでの使用が許可される１以上の領域を有する少なくとも１つの情報アセットと同じビジネス上の意味を有するデータを含むかどうかをチェックすることにより、行うことができる。これは、カタログ１５３の分類情報を用いて、行うことができる。

【００５０】

ステップ２０１の判断は、例えば、ＢＩＮＤＥＲアルゴリズム又はＭｉｎｈａｓｈ技術を用いて、行うことができる。

【００５１】

ステップ２０３において、第１のデータセット及び第２のデータセットの少なくとも一方に対する要求を、ユーザから受け取ることができる。要求は、ステップ２０１の前に又は後に受け取ることができる。

【００５２】

ステップ２０５において、ユーザがエンティティを示すものにアクセスできないように、第１のデータセットもしくは第２のデータセット又はその両方の少なくとも一部を修正することができる。例えば、２つのデータセットの一方がデータ・マスキング要件を有し、他方のデータセットがデータ・マスキング要件を有さない場合には、他方のデータセットを、それが要求されるものではなかったとしても、修正することができる。

【００５３】

ステップ２０７において、要求されたデータセットをユーザに提供することができる。要求されたデータセットは、ステップ２０５の修正されたものであることも又は修正されたものでないこともある。

【００５４】

ＰＫ－ＦＫ関係の場合、例示的方法を次のように行うことができる。データ・サイエンティストが１つのデータセット又はデータセットのグループと連携したいと望むとき、本方法は、これらのデータセットと、同じデータ・サイエンティストに利用可能な他のデータセットとの間の全ての可能なＰＫ－ＦＫ関係の事前計算されたリストをチェックすることができる。前のステップで取得された全ての可能な上記関係のリストは、要求されたデ

ータセットとデータ・サイエンティストに利用可能な他のデータセットとの間にこれらの上記関係を有するように構築することができる全ての可能な結合により取得することができる、全ての結果セットのメタデータを得るために用いられる。例えば、結果セットは、同じ要求されたデータセットから構築することができる最大結合組み合わせをシミュレートすることにより、最悪の場合のシナリオに従って生成され得る。カタログのガバナンス・ポリシーとのこれらの可能な結合により取得することができる可能な結果セットの適合性のチェックが行われる。前のチェックの結果に基づいて、データ・サイエンティストは、それらがガバナンス・ポリシーに適合するように（例えば、最悪の場合のシナリオ結合結果でさえ、ガバナンス・ポリシーに適合するように）匿名化された要求されたデータセットを受け取ることができるか、又はデータ・サイエンティストは、新しい要求されたデータセットをより低レベルの匿名化でロードできる前に、同じく利用可能な何らかのデータセットが除去されるべきであるとの示唆を受け取ることができる。

【００５５】

以下は、本方法の利点を示す例である。

【００５６】

例えば、ガバナンス・ポリシーは、通話詳細レコード（ＣＤＲ：Call Detail Record）の匿名化された詳細のみを格納するよう要求する。例えば、第１のデータセットは、例えば、属性顧客名、顧客アドレス及び顧客電話の値のマスキングなど、修正により匿名化される以下のＣＤＲを含む。

（ａ）顧客ＩＤ：１１２２３３４４５５
（ｂ）顧客名：ａｂｃ７８７８ｄｆ３４３
（ｃ）顧客アドレス：ｆｇｈ７８７８ｅｒ９０
（ｄ）顧客電話：ｉｏｐ７８７８ｔｚ１１
（ｅ）デバイス：ＩＤｘｙｚを有するＡｐｐｌｅ　ｉＰｈｏｎｅ６
（ｆ）通話開始：２０１６年１０月２５日午後２時４０分
（ｇ）通話終了：２０１６年１０月２５日午後２時５０分
（ｈ）持続時間：１０分間
（ｉ）通話した電話：００４９－（０）７０３１－８８８－９９１１
（ｊ）通話中に使用したアンテナ：５２．５２００°Ｎ、１３．４０５０°Ｅ

【００５７】

また、第２のデータセットは、以下の属性値を有するプラットフォーム（例えば、Ｔｗｉｔｔｅｒ、Ｆａｃｅｂｏｏｋ、ＬｉｎｋｅｄＩｎ）上に投稿されるソーシャルメディア投稿を含む。：

（ａ）投稿時間：２０１６年１０月２５日午後２時３９分
（ｂ）位置：５２．５２００°Ｎ、１３．４０５０°Ｅ
（ｃ）デバイス：ＩＤｘｙｚを有するＡｐｐｌｅ　ｉＰｈｏｎｅ６
（ｄ）投稿のＩＤ：Ｊｏｈｎ　Ｓｍｉｔｈ
（ｅ）コンテンツ：「～についてのこの格好いいものをチェックしてみて」

【００５８】

単に通話の位置（アンテナ位置）及びタイミングと組み合わせられる、投稿タイミング、位置等を有するソーシャルメディア投稿におけるメタデータのＣＤＲの匿名化バージョンとの重ね合わせにより、多くの場合、ＣＤＲの８０％又はそれより多くの匿名化解除（de-anonymize）が可能になり、匿名化ＣＤＲを、識別を可能にするソーシャルメディア・プロファイルに結合することにより、ＣＤＲの背後の個人を知ることができるようになることが、研究により示されている。

【００５９】

従って、各々がそれ自体では損害を与えない２つのデータセットが与えられるものの、それらを一緒にすることで、２つのデータセットの一方の匿名化の努力が破られる。本方法により、第１のデータセットが要求される場合に、単独で得られるその第１のデータセットは発呼側の識別を含むことがなくても、デバイスや通話の時間を識別するカラムをマ

スキング又は一般化して、ユーザの特定を可能にする付加的な情報を与えることになってしまう第１のデータセットと第２のデータセットの結合のためにもはや使用できないようにすることができる。

【００６０】

一例において、別の例示的方法が提供される。方法は、データセットのグループの１つのデータセットに対する要求を、ユーザから受け取ることと、どの更なるデータセットがユーザに利用可能かを判断することと、（ｉ）データセット又はデータセットのグループと（ｉｉ）使用するユーザに利用可能な更なるデータセットの組み合わせについての全ての可能な主キー／外部キー関係を判断することと、全ての可能な主キー／外部キー関係についての結合の可能な結果セットのメタデータを判断することと、可能な結果セットの、ガバナンス・ポリシーとの適合性をチェックすることと、適切な場合は、要求されたデータセット又はデータセットのグループへのアクセスを提供する前に、ガバナンス・ポリシーと適合するように、要求されたデータの一部を匿名化することとを含む。全ての考えられる主キー／外部キー関係の判断は、例えば、以下のように実行することができる。すなわちデータセットのグループにおける全ての可能な単一カラム及びマルチカラムの包含従属性を判断すること（例えば、ＢＩＮＤＥＲアルゴリズムを用いて）、可能な対を決定してカラム値（the column values）に基づいて計算された特性（シグネチャー）を用いてキー関係を構築すること、カラムの濃度に基づいて、全ての可能な主キー／外部キー関係をもたらす、１つのカラムは一意であるか又はほぼ一意である（つまり、全てのその値が異なる）カラムの対を識別すること、を実行することができる。

【００６１】

種々の実施形態が、以下の番号付き箇条において特定される。

【００６２】

１．データベース・システムのデータ匿名化のための方法であって、

データベース・システムの第１のデータセットと第２のデータセットとが、２つのデータセットにおける値（values）を有するエンティティを示す関係を有するかどうかを判断することと、

第１のデータセット及び第２のデータセットの少なくとも一方に対する要求を、ユーザから受け取ることと、

第１のデータセットと第２のデータセットが上記関係を有する場合、ユーザがエンティティを示すものにアクセスできないように、第１のデータセット及び第２のデータセットの少なくとも一方を修正することと、

要求されたデータセットを提供することと、

を含む、方法。

【００６３】

２．第１のデータセット及び第２のデータセットはレコードを含み、各レコードはそれぞれのエンティティの属性値の組み合わせであり、上記関係により示されるエンティティは、第１のデータセットもしくは第２のデータセット又はその両方の少なくとも１つのレコードのエンティティである、箇条１に記載の方法。

【００６４】

３．データベース・システムの全てのデータセットの間の上記関係を判断することと、判断した上記関係についての情報を含むメタデータ構造を提供することとをさらに含み、第１のデータセットと第２のデータセットとが上記関係を有するかどうかを判断することは、メタデータ構造を用いて行われる、上述の箇条のいずれかに記載の方法。

【００６５】

４．データベース・システムにおける変更に応答して、データベース・システムのデータセット間の上記関係を再判断することと、それに応じてメタデータ構造を更新することとをさらに含む、箇条３に記載の方法。

【００６６】

５．修正を行うことは、ユーザによるエンティティへのアクセスが所定の統治ポリシーに

違反するとの判断に応答して行われる、上述の箇条のいずれかに記載の方法。
【００６７】
６．上記関係がエンティティを示すかどうかを判断することは、第２のデータセットのそれぞれのターゲット・カラムを参照する第１のデータセットの１以上のソース・カラムを識別することと、ソース・カラムとターゲット・カラムを組み合わせることと、組み合わせ結果に基づいて、上記関係がエンティティを示すか又は示さないかを判断することとを含む、上述の箇条のいずれかに記載の方法。
【００６８】
７．組み合わせは、ＳＱＬ結合演算を用いて行われる、箇条６に記載の方法。
【００６９】
８．判断した上記関係は、主キー・外部キー関係である、上述の箇条のいずれかに記載の方法。
【００７０】
９．要求を受け取ったこと、並びに第１のデータセット及び第２のデータセットがユーザによりアクセス可能であるとの判断に応答して、第１のデータセットと第２のデータセットとが上記関係を有するかどうかの判断を行うことをさらに含む、上述の箇条のいずれかに記載の方法。
【００７１】
１０．要求されたデータセットの修正は、要求されたデータセットの１以上のカラムをマスキングすることを含む、上述の箇条のいずれかに記載の方法。
【００７２】
１１．上記関係の判断は、自動的に行われる、上述の箇条のいずれかに記載の方法。
【００７３】
１２．上記関係の判断は、データベース・システムにおける変更の検出に応答して自動的に行われる、箇条１１に記載の方法。
【００７４】
本発明の態様は、本発明の実施形態による方法、装置（システム）及びコンピュータ・プログラム製品のフローチャート図もしくはブロック図又はその両方を参照して説明される。フローチャート図もしくはブロック図又はその両方の各ブロック、並びにフローチャート図もしくはブロック図又はその両方におけるブロックの組み合わせは、コンピュータ可読プログラム命令によって実装できることが理解されるであろう。
【００７５】
本発明は、システム、方法、もしくはコンピュータ・プログラム製品又はその組み合わせとすることができる。コンピュータ・プログラム製品は、プロセッサに本発明の態様を実行させるためのコンピュータ可読プログラム命令をその上に有するコンピュータ可読ストレージ媒体（単数又は複数）を含むことができる。
【００７６】
コンピュータ可読ストレージ媒体は、命令実行デバイスにより使用される命令を保持及び格納できる有形デバイスとすることができる。コンピュータ可読ストレージ媒体は、例えば、これらに限定されるものではないが、電子記憶装置、磁気記憶装置、光学記憶装置、電磁気記憶装置、半導体記憶装置、又は上記のいずれかの適切な組み合わせとすることができる。コンピュータ可読ストレージ媒体のより具体的な例の非網羅的なリストとして、以下のもの：すなわち、ポータブル・コンピュータ・ディスケット、ハードディスク、ランダム・アクセス・メモリ（ＲＡＭ）、読み出し専用メモリ（ＲＯＭ）、消去可能プログラム可能読み出し専用メモリ（ＥＰＲＯＭ又はフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（ＳＲＡＭ）、ポータブル・コンパクト・ディスク読み出し専用メモリ（ＣＤ－ＲＯＭ）、デジタル多用途ディスク（ＤＶＤ）、メモリ・スティック、フロッピー・ディスク、命令がそこに記録された機械的にエンコードされたデバイス、及び上記のいずれかの適切な組み合わせが挙げられる。本明細書で使用される場合、コンピュータ可読ストレージ媒体は、電波、又は他の自由に伝搬する電磁波、導波管若しくは

他の伝送媒体を通じて伝搬する電磁波（例えば、光ファイバ・ケーブルを通る光パルス）、又はワイヤを通って送られる電気信号などの、一時的信号自体として解釈されない。

【００７７】

本明細書で説明されるコンピュータ可読プログラム命令は、コンピュータ可読ストレージ媒体からそれぞれのコンピューティング／処理デバイスに、又は、例えばインターネット、ローカル・エリア・ネットワーク、広域ネットワーク、もしくは無線ネットワーク又はその組み合わせなどのネットワークを介して外部コンピュータ又は外部ストレージ・デバイスにダウンロードすることができる。ネットワークは、銅伝送ケーブル、光伝送ファイバ、無線伝送、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ、もしくはエッジ・サーバ又はその組み合わせを含むことができる。各コンピューティング／処理デバイスにおけるネットワーク・アダプタ・カード又はネットワーク・インターフェースは、ネットワークからコンピュータ可読プログラム命令を受け取り、コンピュータ可読プログラム命令を転送して、それぞれのコンピューティング／処理デバイス内のコンピュータ可読ストレージ媒体に格納する。

【００７８】

本発明の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セットアーキテクチャ（ＩＳＡ）命令、機械命令、機械依存命令、マイクロコード、ファームウェア命令、状態設定データ、又は、Ｓｍａｌｌｔａｌｋ、Ｃ＋＋などのオブジェクト指向プログラミング言語、及び、「Ｃ」プログラミング言語若しくは類似のプログラミング言語などの従来の手続き型プログラミング言語を含む１以上のプログラミング言語の任意の組み合わせで記述されるソース・コード又はオブジェクト・コードとすることができる。コンピュータ可読プログラム命令は、完全にユーザのコンピュータ上で実行される場合もあり、一部がユーザのコンピュータ上で、独立型ソフトウェア・パッケージとして実行される場合もあり、一部がユーザのコンピュータ上で実行され、一部が遠隔コンピュータ上で実行される場合もあり、又は完全に遠隔コンピュータ若しくはサーバ上で実行される場合もある。最後のシナリオにおいて、遠隔コンピュータは、ローカル・エリア・ネットワーク（ＬＡＮ）若しくは広域ネットワーク（ＷＡＮ）を含むいずれかのタイプのネットワークを通じてユーザのコンピュータに接続される場合もあり、又は外部コンピュータへの接続がなされる場合もある（例えば、インターネットサービスプロバイダを用いたインターネットを通じて）。幾つかの実施形態において、例えば、プログラム可能論理回路、フィールド・プログラマブル・ゲート・アレイ（ＦＰＧＡ）、又はプログラム可能論理アレイ（ＰＬＡ）を含む電子回路は、本発明の態様を実施するために、コンピュータ可読プログラム命令の状態情報を利用することによって、コンピュータ可読プログラム命令を実行して、電子回路を個別化することができる。

【００７９】

本発明の態様は、本発明の実施形態による方法、装置（システム）及びコンピュータ・プログラム製品のフローチャート図もしくはブロック図又はその両方を参照して説明される。フローチャート図もしくはブロック図又はその両方の各ブロック、並びにフローチャート図もしくはブロック図又はその両方におけるブロックの組み合わせは、コンピュータ可読プログラム命令によって実装できることが理解されるであろう。

【００８０】

これらのコンピュータ可読プログラム命令を、汎用コンピュータ、専用コンピュータ、又は他のプログラム可能データ処理装置のプロセッサに与えて機械を製造し、それにより、コンピュータ又は他のプログラム可能データ処理装置のプロセッサによって実行される命令が、フローチャートもしくはブロック図又はその両方の１以上のブロック内で指定された機能／動作を実施するための手段を作り出すようにすることができる。これらのコンピュータ・プログラム命令を、コンピュータ、他のプログラム可能データ処理装置、もしくは他のデバイス又はその組み合わせを特定の方式で機能させるように指示することができるコンピュータ可読媒体内に格納し、それにより、そのコンピュータ可読媒体内に格納された命令が、フローチャートもしくはブロック図又はその両方の１以上のブロックにお

いて指定された機能／動作の態様を実施する命令を含む製品を含むようにすることもできる。

【００８１】

コンピュータ・プログラム命令を、コンピュータ、他のプログラム可能データ処理装置、又は他のデバイス上にロードして、一連の動作ステップをコンピュータ、他のプログラム可能データ処理装置、又は他のデバイス上で行わせてコンピュータ実施のプロセスを生産し、それにより、コンピュータ又は他のプログラム可能装置上で実行される命令が、フローチャートもしくはブロック図又はその両方の１以上のブロックにおいて指定された機能／動作を実行するためのプロセスを提供するようにすることもできる。

【００８２】

図面内のフローチャート及びブロック図は、本発明の様々な実施形態による、システム、方法、及びコンピュータ・プログラム製品の可能な実装の、アーキテクチャ、機能及び動作を示す。この点に関して、フローチャート内の各ブロックは、指定された論理機能を実装するための１以上の実行可能命令を含む、モジュール、セグメント、又はコードの一部を表すことができる。幾つかの代替的な実装において、ブロック内に示される機能は、図に示される順序とは異なる順序で生じることがある。例えば、連続して示される２つのブロックは、関与する機能に応じて、実際には実質的に同時に実行されることもあり、又はこれらのブロックはときとして逆順で実行されることもある。ブロック図もしくはフローチャート図又はその両方の各ブロック、及びブロック図もしくはフローチャート図又はその両方におけるブロックの組み合わせは、指定された機能又は動作を実行する、又は専用のハードウェアとコンピュータ命令との組み合わせを実行する、専用ハードウェア・ベースのシステムによって実装できることにも留意されたい。

【００８３】

本開示はクラウド・コンピューティングについての詳細な説明を含むが、本明細書に記載される教示の実装は、クラウド・コンピューティング環境に限定されないことを理解されたい。むしろ、本発明の実施形態は、現在既知の又は後で開発される他のいずれかのタイプのコンピューティング環境と共に実施することができる。

【００８４】

クラウド・コンピューティングは、最小限の管理労力又はサービス・プロバイダとの対話で迅速にプロビジョニング及び解放することができる構成可能なコンピューティング・リソース（例えば、ネットワーク、ネットワーク帯域幅、サーバ、処理、メモリ、ストレージ、アプリケーション、仮想マシン、及びサービス）の共有プールへの、便利なオンデマンドのネットワーク・アクセスを可能にするためのサービス配信のモデルである。このクラウド・モデルは、少なくとも５つの特徴、少なくとも３つのサービス・モデル、及び少なくとも４つのデプロイメント・モデルを含むことができる。

【００８５】

特徴は、以下の通りである。

【００８６】

オンデマンド・セルフサービス：クラウド・コンシューマは、必要に応じて、サーバ時間及びネットワーク・ストレージ等のコンピューティング機能を、人間がサービスのプロバイダと対話する必要なく自動的に、一方的にプロビジョニングすることができる。

【００８７】

広範なネットワーク・アクセス：機能は、ネットワーク上で利用可能であり、異種のシン又はシック・クライアント・プラットフォーム（例えば、携帯電話、ラップトップ、及びＰＤＡ）による使用を促進する標準的な機構を通じてアクセスされる。

【００８８】

リソース・プール化：プロバイダのコンピューティング・リソースは、マルチ・テナント・モデルを用いて、異なる物理及び仮想リソースを要求に応じて動的に割り当て及び再割り当てすることにより、複数のコンシューマにサービスを提供するためにプールされる。コンシューマは、一般に、提供されるリソースの正確な位置についての制御又は知識を

持たないという点で、位置とは独立しているといえるが、より抽象化レベルの高い位置（例えば、国、州、又はデータセンタ）を特定できる場合がある。

【００８９】

迅速な弾力性：機能は、迅速かつ弾力的に、場合によっては自動的に、プロビジョニングして素早くスケール・アウトし、迅速にリリースして素早くスケール・インさせることができる。コンシューマにとって、プロビジョニングに利用可能なこれらの機能は、多くの場合、無制限であり、いつでもどんな量でも購入できるように見える。

【００９０】

計測されるサービス：クラウド・システムは、サービスのタイプ（例えば、ストレージ、処理、帯域幅、及びアクティブなユーザ・アカウント）に適した何らかの抽象化レベルでの計量機能を用いることによって、リソースの使用を自動的に制御及び最適化する。リソース使用を監視し、制御し、報告し、利用されるサービスのプロバイダとコンシューマの両方に対して透明性をもたらすことができる。

【００９１】

サービス・モデルは以下の通りである。

【００９２】

Ｓｏｆｔｗａｒｅ ａｓ ａ Ｓｅｒｖｉｃｅ（ＳａａＳ）：クラウド・インフラストラクチャ上で動作しているプロバイダのアプリケーションを使用するために、コンシューマに提供される機能である。これらのアプリケーションは、ウェブ・ブラウザ（例えば、ウェブ・ベースの電子メール）などのシン・クライアント・インターフェースを通じて、種々のクライアント・デバイスからアクセス可能である。コンシューマは、限定されたユーザ固有のアプリケーション構成設定の考え得る例外として、ネットワーク、サーバ、オペレーティング・システム、ストレージ、又は個々のアプリケーション機能をも含めて、基礎をなすクラウド・インフラストラクチャを管理又は制御しない。

【００９３】

Ｐｌａｔｆｏｒｍ ａｓ ａ Ｓｅｒｖｉｃｅ（ＰａａＳ）：プロバイダによってサポートされるプログラミング言語及びツールを用いて生成された、コンシューマが生成した又は取得したアプリケーションを、クラウド・インフラストラクチャ上にデプロイするために、コンシューマに提供される機能である。コンシューマは、ネットワーク、サーバ、オペレーティング・システム、又はストレージなどの基礎をなすクラウド・インフラストラクチャを管理又は制御しないが、配備されたアプリケーション、及び場合によってはアプリケーション・ホスティング環境構成に対して制御を有する。

【００９４】

Ｉｎｆｒａｓｔｒｕｃｔｕｒｅ ａｓ ａ Ｓｅｒｖｉｃｅ（ＩａａＳ）：コンシューマが、オペレーティング・システム及びアプリケーションを含み得る任意のソフトウェアを配備及び動作させることができる、処理、ストレージ、ネットワーク、及び他の基本的なコンピューティング・リソースをプロビジョニングするために、コンシューマに提供される機能である。コンシューマは、基礎をなすクラウド・インフラストラクチャを管理又は制御しないが、オペレーティング・システム、ストレージ、配備されたアプリケーションに対する制御、及び場合によってはネットワーク・コンポーネント（例えば、ホストのファイアウォール）選択の限定された制御を有する。

【００９５】

デプロイメント・モデルは以下の通りである。

【００９６】

プライベート・クラウド：クラウド・インフラストラクチャは、ある組織のためだけに運営される。このクラウド・インフラストラクチャは、その組織又は第三者によって管理することができ、オンプレミス又はオフプレミスに存在することができる。

【００９７】

コミュニティ・クラウド：クラウド・インフラストラクチャは、幾つかの組織によって共有され、共通の関心事項（例えば、任務、セキュリティ要件、ポリシー、及びコンプラ

イアンス上の考慮事項）を有する特定のコミュニティをサポートする。クラウド・インフラストラクチャは、その組織又は第三者によって管理することができ、オンプレミス又はオフプレミスに存在することができる。

【００９８】

パブリック・クラウド：クラウド・インフラストラクチャは、一般公衆又は大規模な業界グループに利用可能であり、クラウド・サービスを販売する組織によって所有される。

ハイブリッド・クラウド：クラウド・インフラストラクチャは、固有のエンティティのままであるが、データ及びアプリケーションの移行性を可能にする標準化された又は専用の技術（例えば、クラウド間の負荷分散のためのクラウド・バースティング）によって結び付けられる２つ以上のクラウド（プライベート、コミュニティ、又はパブリック）の混成物である。

クラウド・コンピューティング環境は、ステートレス性、低結合性、モジュール性、及びセマンティック相互運用性に焦点を置くことを指向するサービスである。クラウド・コンピューティングの中心は、相互接続されたノードのネットワークを含むインフラストラクチャである。

【００９９】

ここで図３を参照すると、例証的クラウド・コンピューティング環境５０が示される。示されるように、クラウド・コンピューティング環境５０は、例えば、携帯情報端末（ＰＤＡ）又は携帯電話５４Ａ、デスクトップ・コンピュータ５４Ｂ、ラップトップ・コンピュータ５４Ｃ、もしくは自動車コンピュータ・システム５４Ｎ又はその組み合わせ等といった、クラウド・コンシューマによって用いられるローカル・コンピューティング・デバイスと通信できる１以上のクラウド・コンピューティング・ノード１００を含む。ノード１００は、互いに通信することができる。これらのノードは、プライベート・クラウド、コミュニティ・クラウド、パブリック・クラウド、若しくはハイブリッド・クラウド、又はこれらの組み合わせなど、１以上のネットワークにおいて物理的又は仮想的にグループ化することができる（図示せず）。これにより、クラウド・コンピューティング環境５０が、クラウド・コンシューマがローカル・コンピューティング・デバイス上にリソースを保持する必要のないサービスとして、インフラストラクチャ、プラットフォーム、及び／又はソフトウェアを提供することが可能になる。図３に示されるコンピューティング・デバイス５４Ａ～Ｎのタイプは単に例示であることを意図し、コンピューティング・ノード１００及びクラウド・コンピューティング環境５０は、任意のタイプのネットワーク及び／又はネットワーク・アドレス指定可能な接続上で（例えば、ウェブ・ブラウザを用いて）、任意のタイプのコンピュータ化されたデバイスと通信できることを理解されたい。

【０１００】

ここで図４を参照すると、クラウド・コンピューティング環境５０によって提供される機能抽象化層４００のセットが示される。図４に示されるコンポーネント、層、及び機能は単に例示であることを意図し、本発明の実施形態はそれらに限定されないことを予め理解されたい。示されるように、以下の層及び対応する機能が提供される。

【０１０１】

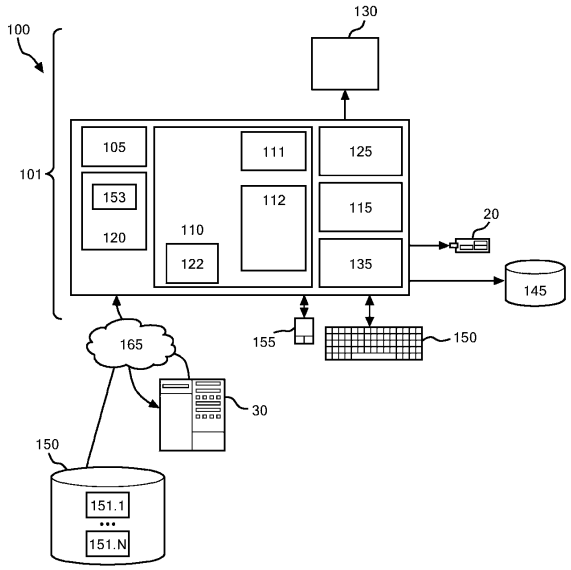
ハードウェア及びソフトウェア層６０は、ハードウェア及びソフトウェア・コンポーネントを含む。ハードウェア・コンポーネントの例として、メインフレーム６１と、ＲＩＳＣ（Reduced Instruction Set Computer、縮小命令セット・コンピュータ）アーキテクチャ・ベースのサーバ６２と、サーバ６３と、ブレードサーバ６４と、ストレージ・デバイス６５と、ネットワーク及びネットワーク・コンポーネント６６と、が含まれる。幾つかの実施形態において、ソフトウェア・コンポーネントは、ネットワーク・アプリケーション・サーバ・ソフトウェア６７と、データベース・ソフトウェア６８とが含まれる。

【０１０２】

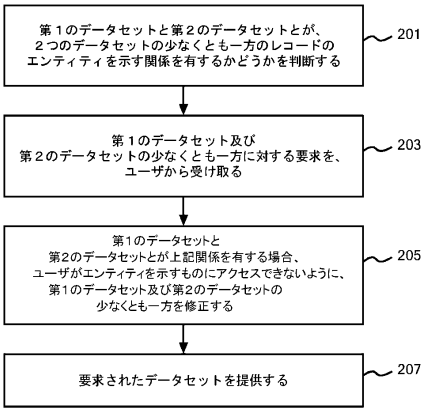
仮想化層７０は、抽象化層を提供し、この層により、仮想エンティティの以下の例、すなわち、仮想サーバ７１、仮想ストレージ７２、仮想プライベート・ネットワークを含む仮想ネットワーク７３、仮想アプリケーション及びオペレーティング・システム７４、並

びに仮想クライアント７５を提供することができる。
【０１０３】
一例において、管理層８０は、以下で説明される機能を提供することができる。リソース・プロビジョニング８１は、クラウド・コンピューティング環境内でタスクを実行するために利用されるコンピューティング・リソース及び他のリソースの動的な調達を提供する。計量及び価格決定８２は、クラウド・コンピューティング環境内でリソースが利用される際のコスト追跡と、これらのリソースの消費に対する課金又は請求とを提供する。一例においては、これらのリソースは、アプリケーション・ソフトウェア・ライセンスを含むことができる。セキュリティは、クラウド・コンシューマ及びタスクに対する識別情報の検証と、データ及び他のリソースに対する保護とを提供する。ユーザ・ポータル８３は、コンシューマ及びシステム管理者のために、クラウド・コンピューティング環境へのアクセスを提供する。サービス・レベル管理８４は、要求されるサービス・レベルが満たされるように、クラウド・コンピューティング・リソースの割り当て及び管理を提供する。サービス・レベル・アグリーメント（Service Level Agreement、ＳＬＡ）の計画及び履行８５は、ＳＬＡに従って将来の要件が予測されるクラウド・コンピューティング・リソースの事前配置及び調達を提供する。
【０１０４】
ワークロード層９０は、クラウド・コンピューティング環境を利用することができる機能の例を提供する。この層から提供することができるワークロード及び機能の例として、マッピング及びナビゲーション９１、ソフトウェア開発及びライフサイクル管理９２、仮想教室教育配信９３、データ分析処理９４、トランザクション処理９５、及びデータ匿名化処理９６が挙げられる。データ匿名化処理９６は、データベース・システムの全てのデータセットの間の上記関係を判断し、判断した上記関係についての情報を含むメタデータ構造を提供することによる、データベース・システムのデータ匿名化に関することができ、判断は、メタデータ構造を用いて行われる。データベース・システムにおける変更に応答して、データ匿名化処理９６は、データベース・システムのデータセット間の上記関係を再判断し、それに応じてメタデータ構造を更新する。
【０１０５】
本発明の種々の実施形態の説明は、例証の目的のために提示されたが、これらは、網羅的であること、又は本発明を開示した実施形態に限定することを意図するものではない。当業者には、説明される実施形態の範囲及び趣旨から逸脱することなく、多くの修正及び変形が明らかであろう。本明細書で用いられる用語は、実施形態の原理、実際の適用、又は市場に見られる技術に優る技術的改善を最もよく説明するため、又は、当業者が、本明細書に開示される実施形態を理解するのを可能にするために選択された。
【符号の説明】
【０１０６】
２０、１４５：入力／出力（Ｉ／Ｏ）デバイス
３０：外部システム
１００：システム
１０１：コンピュータ
１０５：プロセッサ
１１０：メモリ
１１１：ＯＳ
１１２：ソフトウェア
１１５：メモリ・コントローラ
１２０：ストレージ
１２２：基本入力出力システム（ＢＩＯＳ）
１２５：ディスプレイ・コントローラ
１３０：ディスプレイ
１３５：入力／出力コントローラ

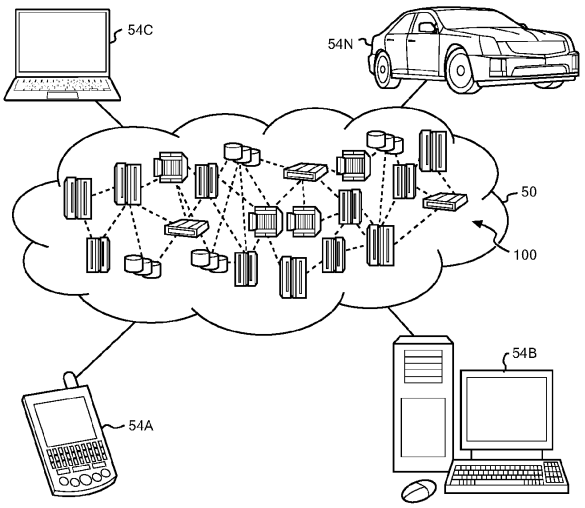
１５０：キーボード
１５０：データベース・システム
１５１．１～１５１．Ｎ：データセット
１５３：カタログ
１５５：マウス
１６５：ネットワーク
【図面】
【図１】



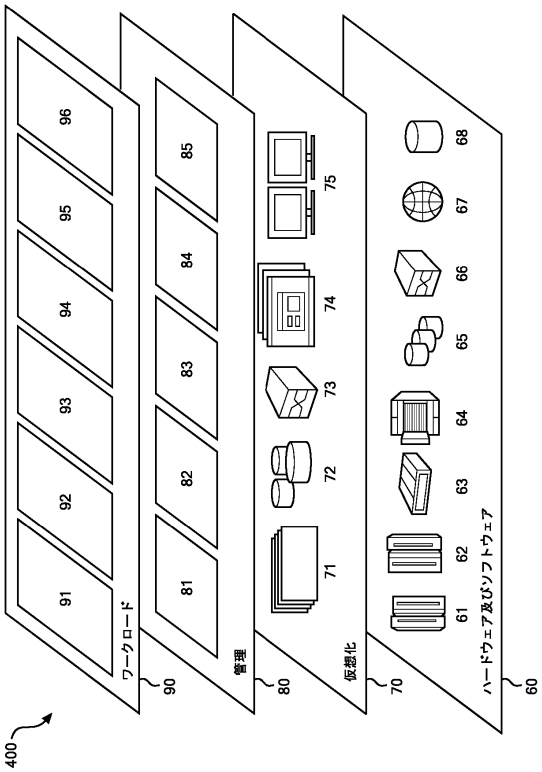
【図２】



【図３】


54C
54N
50
100
54B
54A

【図４】


96
95
94
93
92
91
ワークロード
90
85
84
83
82
81
管理
80
75
74
73
72
71
仮想化
70
68
67
66
65
64
63
62
61
ハードウェア及びソフトウェア
60
400

フロントページの続き

(72)発明者　オベルホファー、マルティン
ドイツ連邦共和国　７１０３２　ベーブリンゲン　シェーナイヒャー・シュトラーセ　２２０
(72)発明者　マイアー、アルベルト
ドイツ連邦共和国　７１０３２　ベーブリンゲン　シェーナイヒャー・シュトラーセ　２２０
(72)発明者　サイエ、ヤニック
ドイツ連邦共和国　７１０３２　ベーブリンゲン　シェーナイヒャー・シュトラーセ　２２０
審査官　岸野　徹
(56)参考文献　特開２００４－３１８３９１（ＪＰ，Ａ）
特開２０１４－０１６６７５（ＪＰ，Ａ）
特開２０１３－２４６５４７（ＪＰ，Ａ）
特表２０１４－５００５４４（ＪＰ，Ａ）
特開２００８－１４０２０２（ＪＰ，Ａ）
国際公開第２０１３／１２１７３８（WO，Ａ１）
国際公開第２０１２／１６５５１８（WO，Ａ１）
特開２００６－１８９９２６（ＪＰ，Ａ）
特開２０１３－１４３１１４（ＪＰ，Ａ）
米国特許出願公開第２００９／０１００５２７（ＵＳ，Ａ１）
米国特許出願公開第２０１４／００１３０６５（ＵＳ，Ａ１）
伊奈 優樹，同一個人データの存在する水平分割データベースのダミー値追加による匿名化，マルチメディア，分散，協調とモバイル（ＤＩＣＯＭＯ２０１７）シンポジウム論文集 情報処理学会シンポジウムシリーズ Ｖｏｌ．２０１７ Ｎｏ．１ ［ＣＤ－ＲＯＭ］ IPSJ Symposium Series，日本，一般社団法人情報処理学会，2017年06月21日，第2017巻，pp.661-668
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　２１／６２
Ｇ０６Ｆ　１６／２４
Ｇ０６Ｆ　１６／２８