

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5902185号
(P5902185)

(45) 発行日 平成28年4月13日 (2016. 4. 13)

(24) 登録日 平成28年3月18日 (2016. 3. 18)

(51) Int. Cl.

F I

G 0 6 F 9 / 4 4 (2006. 01)

G 0 6 F 9 / 0 6 6 2 0 E

請求項の数 44 (全 22 頁)

(21) 出願番号 特願2013-535149 (P2013-535149)
 (86) (22) 出願日 平成23年10月25日 (2011. 10. 25)
 (65) 公表番号 特表2013-541784 (P2013-541784A)
 (43) 公表日 平成25年11月14日 (2013. 11. 14)
 (86) 国際出願番号 PCT/US2011/057623
 (87) 国際公開番号 W02012/061109
 (87) 国際公開日 平成24年5月10日 (2012. 5. 10)
 審査請求日 平成26年10月23日 (2014. 10. 23)
 (31) 優先権主張番号 61/406, 438
 (32) 優先日 平成22年10月25日 (2010. 10. 25)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 509123208
 アビニシオ テクノロジー エルエルシー
 アメリカ合衆国 O 2 4 2 1 マサチュー
 セッツ州 レキシントン スプリング ス
 トリート 2 0 1
 (74) 代理人 100079108
 弁理士 稲葉 良幸
 (74) 代理人 100109346
 弁理士 大貫 敏史
 (72) 発明者 ラーソン, ブロンド
 アメリカ合衆国, マサチューセッツ州 O
 2 0 6 7, シャロン, ランターン レーン
 4 9

最終頁に続く

(54) 【発明の名称】 コンピュータプログラムを表すデータフローグラフ内のデータセットオブジェクトの管理

(57) 【特許請求の範囲】

【請求項 1】

グラフに基づくデータ処理のためにデータセットオブジェクトを管理する方法であって、

1 つ又は複数のデータセットオブジェクトのグループをデータ記憶システムに記憶することであって、前記データセットオブジェクトのそれぞれは各データセットを表し、前記グループは第 1 のデータセットオブジェクトを含み、前記第 1 のデータセットオブジェクトは、データセットからデータレコードを読み取るように前記第 1 のデータセットオブジェクトを構成するデータソースモードと、前記データセットにデータレコードを書き込むように前記第 1 のデータセットオブジェクトを構成するデータシンクモードとを少なくとも含む複数のモードを含むこと、

前記第 1 のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第 1 のノードとの間の関連付けを生成することであって、前記第 1 のノードは、前記データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、前記第 1 のデータセットオブジェクトは、異なる変換ロジックが前記第 1 のノードで処理されるデータに適用される複数のモードを含み、前記第 1 のデータセットオブジェクトは、前記複数のモードのうちのあるモードにしたがって前記データフローグラフの前記第 1 のノードと対話することを含む方法。

【請求項 2】

10

20

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記データフローグラフに対応するデータ構造にアクセスすることであって、前記データフローグラフは、前記第 1 のノードを含む複数のノード及び前記ノードを接続するリンクを含み、前記リンクは、前記ノード間のデータフローを表すこと、

ユーザ入力に基づいて前記グループから選択された前記第 1 のデータセットオブジェクトを受信することであって、前記第 1 のデータセットオブジェクトは前記複数のモードを含み、各モードは、変換ロジック及び入力ポート又は出力ポートのうちの少なくとも 1 つを含む、受信すること、

ユーザ入力に基づいて前記第 1 のデータセットオブジェクトに対して選択されたモードを受信すること、

前記選択されたモードを受信した後、前記データフローグラフに前記第 1 のノードを生成することであって、前記第 1 のノードは、前記選択されたモードの変換ロジック、任意の入力ポート、及び任意の出力ポートを含むこと、

リンクを使用して、前記第 1 のノードを前記データフローグラフ内の少なくとも 1 つのノードに接続すること、
を含む、請求項 1 に記載の方法。

【請求項 3】

前記第 1 のデータセットオブジェクトは少なくとも 1 つのパラメータをさらに含み、前記第 1 のノードは前記パラメータの値をさらに含む、請求項 2 に記載の方法。

【請求項 4】

前記複数のモードのうちの第 1 のモードの前記変換ロジックは、データストアからデータレコードを読み取る変換ロジックを含む、請求項 2 に記載の方法。

【請求項 5】

前記複数のモードのうちの第 2 のモードの前記変換ロジックは、データレコードを前記データストアに書き込む変換ロジックを含む、請求項 4 に記載の方法。

【請求項 6】

前記第 1 のノードは、前記第 1 のデータセットオブジェクトを識別する情報を含む、請求項 2 に記載の方法。

【請求項 7】

ユーザ入力に基づいて前記グループから選択された前記第 1 のデータセットオブジェクトを受信すること、

複数のデータフローグラフを識別することであって、前記データフローグラフのそれぞれは、前記第 1 のデータセットオブジェクトを識別するノードを含むこと、
をさらに含む、請求項 1 に記載の方法。

【請求項 8】

前記データセットオブジェクトの各モードは、ノード及び前記ノードを接続するリンクを含む別個のデータフローグラフで表され、前記リンクは前記ノード間のデータフローを表す、請求項 1 に記載の方法。

【請求項 9】

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記第 1 のノードを含む前記データフローグラフを含む複数のデータフローグラフに対応するデータ構造にアクセスすることであって、各データフローグラフは、複数のノード及び前記ノードを接続するリンクを含み、前記ノードは変換ロジックを含み、前記リンクは前記ノード間のデータフローを表すこと、

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別することであって、少なくとも 1 つの識別されたノードは 1 つ又は複数のパラメータを含むこと、

データセットオブジェクトを生成することであって、

10

20

30

40

50

前記識別されたノードの前記変換ロジックに基づいて、前記生成されたデータセットオブジェクトの複数のモードを識別すること、

前記識別されたノードに関連付けられたパラメータに基づいて、前記生成されたデータオブジェクトのパラメータを識別すること

を含むこと、

前記生成されたデータセットオブジェクトを前記グループに記憶すること、
を含む、請求項 1 に記載の方法。

【請求項 10】

前記生成されたデータセットオブジェクトを識別する情報に基づいて、前記複数のノードのそれぞれを更新して、前記データストアにアクセスすることをさらに含む、請求項 9 に記載の方法。

10

【請求項 11】

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別することは、

前記複数のデータフローグラフのそれぞれを実行すること、

前記データフローグラフの実行中、前記複数のデータフローグラフの前記ノードのパラメータに割り当てられたパラメータ値を記録すること、

前記記録されたパラメータ値に基づいて前記複数のノードを識別すること、
を含む、請求項 9 に記載の方法。

【請求項 12】

20

グラフに基づくデータ処理のためにデータセットオブジェクトを管理するコンピュータプログラムを記憶するコンピュータ可読記憶媒体であって、前記コンピュータプログラムは、計算システムに、

1 つ又は複数のデータセットオブジェクトのグループをデータ記憶システムに記憶することであって、前記データセットオブジェクトのそれぞれは各データセットを表し、前記グループは第 1 のデータセットオブジェクトを含み、前記第 1 のデータセットオブジェクトは、データセットからデータレコードを読み取るように前記第 1 のデータセットオブジェクトを構成するデータソースモードと、前記データセットにデータレコードを書き込むように前記第 1 のデータセットオブジェクトを構成するデータシンクモードとを少なくとも含む複数のモードを含むこと、

30

前記グループ内の前記第 1 のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第 1 のノードとの間の関連付けを生成することであって、前記第 1 のノードは、前記データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、前記第 1 のデータセットオブジェクトは、異なる変換ロジックが前記第 1 のノードで処理されるデータに適用される複数のモードを含み、前記第 1 のデータセットオブジェクトは、前記複数のモードのうちのあるモードにしたがって前記データフローグラフの前記第 1 のノードと対話すること
を行わせる命令を含む、コンピュータ可読記憶媒体。

【請求項 13】

グラフに基づくデータ処理のためにデータセットオブジェクトを管理する計算システムであって、

40

1 つ又は複数のデータセットオブジェクトのグループを記憶するデータ記憶システムであって、前記データセットオブジェクトのそれぞれは各データセットを表し、前記グループは第 1 のデータセットオブジェクトを含み、前記第 1 のデータセットオブジェクトは、データセットからデータレコードを読み取るように前記第 1 のデータセットオブジェクトを構成するデータソースモードと、前記データセットにデータレコードを書き込むように前記第 1 のデータセットオブジェクトを構成するデータシンクモードとを少なくとも含む複数のモードを含む、データ記憶システムと、

前記データ記憶システムに結合され、前記グループ内の前記第 1 のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第 1

50

のノードとの間の関連付けを生成するように構成される少なくとも1つのプロセッサであって、前記第1のノードは、前記データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、前記第1のデータセットオブジェクトは、異なる変換ロジックが前記第1のノードで処理されるデータに適用される複数のモードを含み、前記第1のデータセットオブジェクトは、前記複数のモードのうちのあるモードにしたがって前記データフローグラフの前記第1のノードと対話する、少なくとも1つのプロセッサと、
を含む、計算システム。

【請求項14】

グラフに基づくデータ処理のためにデータセットオブジェクトを管理する計算システムであって、

10

1つ又は複数のデータセットオブジェクトのグループを記憶する手段であって、前記データセットオブジェクトのそれぞれは各データセットを表し、前記グループは第1のデータセットオブジェクトを含み、前記第1のデータセットオブジェクトは、データセットからデータレコードを読み取るように前記第1のデータセットオブジェクトを構成するデータソースモードと、前記データセットにデータレコードを書き込むように前記第1のデータセットオブジェクトを構成するデータシンクモードとを少なくとも含む複数のモードを含む、手段と、

前記グループ内の前記第1のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第1のノードとの間の関連付けを生成する手段であって、前記第1のノードは、前記データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、前記第1のデータセットオブジェクトは、異なる変換ロジックが前記第1のノードで処理されるデータに適用される複数のモードを含み、前記第1のデータセットオブジェクトは、前記複数のモードのうちのあるモードにしたがって前記データフローグラフの前記第1のノードと対話する、手段と、
を含む、計算システム。

20

【請求項15】

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記データフローグラフに対応するデータ構造にアクセスすることであって、前記データフローグラフは、前記第1のノードを含む複数のノード及び前記ノードを接続するリンクを含み、前記リンクは、前記ノード間のデータフローを表すこと、

30

ユーザ入力に基づいて前記グループから選択された前記第1のデータセットオブジェクトを受信することであって、前記第1のデータセットオブジェクトは前記複数のモードを含み、各モードは、変換ロジック及び入力ポート又は出力ポートのうちの少なくとも1つを含む、受信すること、

ユーザ入力に基づいて前記第1のデータセットオブジェクトに対して選択されたモードを受信すること、

前記選択されたモードを受信した後、前記データフローグラフに前記第1のノードを生成することであって、前記第1のノードは、前記選択されたモードの変換ロジック、任意の入力ポート、及び任意の出力ポートを含むこと、

40

リンクを使用して、前記第1のノードを前記データフローグラフ内の少なくとも1つのノードに接続すること、

を含む、請求項12に記載のコンピュータ可読記憶媒体。

【請求項16】

前記第1のデータセットオブジェクトは少なくとも1つのパラメータをさらに含み、前記第1のノードは前記パラメータの値をさらに含み、請求項15に記載のコンピュータ可読記憶媒体。

【請求項17】

前記複数のモードのうち第1のモードの前記変換ロジックは、データストアからデー

50

タレコードを読み取る変換ロジックを含む、請求項 15 に記載のコンピュータ可読記憶媒体。

【請求項 18】

前記複数のモードのうちの第2のモードの前記変換ロジックは、データレコードを前記データストアに書き込む変換ロジックを含む、請求項 17 に記載のコンピュータ可読記憶媒体。

【請求項 19】

前記第1のノードは、前記第1のデータセットオブジェクトを識別する情報を含む、請求項 15 に記載のコンピュータ可読記憶媒体。

【請求項 20】

前記コンピュータプログラムは、前記計算システムに、
ユーザ入力に基づいて前記グループから選択された前記第1のデータセットオブジェクトを受信すること、

複数のデータフローグラフを識別することであって、前記データフローグラフのそれぞれは、前記第1のデータセットオブジェクトを識別するノードを含むこと、
を行わせる命令をさらに含む、請求項 12 に記載のコンピュータ可読記憶媒体。

【請求項 21】

前記データセットオブジェクトの各モードは、ノード及び前記ノードを接続するリンクを含む別個のデータフローグラフで表され、前記リンクは前記ノード間のデータフローを表す、請求項 12 に記載のコンピュータ可読記憶媒体。

【請求項 22】

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記第1のノードを含む前記データフローグラフを含む複数のデータフローグラフに対応するデータ構造にアクセスすることであって、各データフローグラフは、複数のノード及び前記ノードを接続するリンクを含み、前記ノードは変換ロジックを含み、前記リンクは前記ノード間のデータフローを表すこと、

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別することであって、少なくとも1つの識別されたノードは1つ又は複数のパラメータを含むこと、

データセットオブジェクトを生成することであって、

前記識別されたノードの前記変換ロジックに基づいて、前記生成されたデータセットオブジェクトの複数のモードを識別すること、

前記識別されたノードに関連付けられたパラメータに基づいて、前記生成されたデータオブジェクトのパラメータを識別すること

を含むこと、

前記生成されたデータセットオブジェクトを前記グループに記憶すること、
を含む、請求項 12 に記載のコンピュータ可読記憶媒体。

【請求項 23】

前記コンピュータプログラムは、前記計算システムに、

前記生成されたデータセットオブジェクトを識別する情報に基づいて、前記複数のノードのそれぞれを更新して、前記データストアにアクセスすること
を行わせる命令をさらに含む、請求項 22 に記載のコンピュータ可読記憶媒体。

【請求項 24】

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別することは、

前記複数のデータフローグラフのそれぞれを実行すること、

前記データフローグラフの実行中、前記複数のデータフローグラフの前記ノードのパラメータに割り当てられたパラメータ値を記録すること、

前記記録されたパラメータ値に基づいて前記複数のノードを識別すること、

を含む、請求項 2 2 に記載のコンピュータ可読記憶媒体。

【請求項 2 5】

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記データフローグラフに対応するデータ構造にアクセスすることであって、前記データフローグラフは、前記第 1 のノードを含む複数のノード及び前記ノードを接続するリンクを含み、前記リンクは、前記ノード間のデータフローを表すこと、

ユーザ入力に基づいて前記グループから選択された前記第 1 のデータセットオブジェクトを受信することであって、前記第 1 のデータセットオブジェクトは前記複数のモードを含み、各モードは、変換ロジック及び入力ポート又は出力ポートのうちの少なくとも 1 つを含む、受信すること、

ユーザ入力に基づいて前記第 1 のデータセットオブジェクトに対して選択されたモードを受信すること、

前記選択されたモードを受信した後、前記データフローグラフに前記第 1 のノードを生成することであって、前記第 1 のノードは、前記選択されたモードの変換ロジック、任意の入力ポート、及び任意の出力ポートを含むこと、

リンクを使用して、前記第 1 のノードを前記データフローグラフ内の少なくとも 1 つのノードに接続すること、

を含む、請求項 1 3 に記載の計算システム。

【請求項 2 6】

前記第 1 のデータセットオブジェクトは少なくとも 1 つのパラメータをさらに含み、前記第 1 のノードは前記パラメータの値をさらに含む、請求項 2 5 に記載の計算システム。

【請求項 2 7】

前記複数のモードのうちの第 1 のモードの前記変換ロジックは、データストアからデータレコードを読み取る変換ロジックを含む、請求項 2 5 に記載の計算システム。

【請求項 2 8】

前記複数のモードのうちの第 2 のモードの前記変換ロジックは、データレコードを前記データストアに書き込む変換ロジックを含む、請求項 2 7 に記載の計算システム。

【請求項 2 9】

前記第 1 のノードは、前記第 1 のデータセットオブジェクトを識別する情報を含む、請求項 2 5 に記載の計算システム。

【請求項 3 0】

前記少なくとも 1 つのプロセッサは、

ユーザ入力に基づいて前記グループから選択された前記第 1 のデータセットオブジェクトを受信すること、

複数のデータフローグラフを識別することであって、前記データフローグラフのそれぞれは、前記第 1 のデータセットオブジェクトを識別するノードを含むこと、

を行うようにさらに構成される、請求項 1 3 に記載の計算システム。

【請求項 3 1】

前記データセットオブジェクトの各モードは、ノード及び前記ノードを接続するリンクを含む別個のデータフローグラフで表され、前記リンクは前記ノード間のデータフローを表す、請求項 1 3 に記載の計算システム。

【請求項 3 2】

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記第 1 のノードを含む前記データフローグラフを含む複数のデータフローグラフに対応するデータ構造にアクセスすることであって、各データフローグラフは、複数のノード及び前記ノードを接続するリンクを含み、前記ノードは変換ロジックを含み、前記リンクは前記ノード間のデータフローを表すこと、

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別

10

20

30

40

50

することであって、少なくとも1つの識別されたノードは1つ又は複数のパラメータを含むこと、

データセットオブジェクトを生成することであって、

前記識別されたノードの前記変換ロジックに基づいて、前記生成されたデータセットオブジェクトの複数のモードを識別すること、

前記識別されたノードに関連付けられたパラメータに基づいて、前記生成されたデータオブジェクトのパラメータを識別すること

を含むこと、

前記生成されたデータセットオブジェクトを前記グループに記憶すること、
を含む、請求項13に記載の計算システム。

10

【請求項33】

前記少なくとも1つのプロセッサは、

前記生成されたデータセットオブジェクトを識別する情報に基づいて、前記複数のノードのそれぞれを更新して、前記データストアにアクセスすること
を行うようにさらに構成される、請求項32に記載の計算システム。

【請求項34】

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別することは、

前記複数のデータフローグラフのそれぞれを実行すること、

前記データフローグラフの実行中、前記複数のデータフローグラフの前記ノードのパラメータに割り当てられたパラメータ値を記録すること、

前記記録されたパラメータ値に基づいて前記複数のノードを識別すること、
を含む、請求項32に記載の計算システム。

20

【請求項35】

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記データフローグラフに対応するデータ構造にアクセスすることであって、前記データフローグラフは、前記第1のノードを含む複数のノード及び前記ノードを接続するリンクを含み、前記リンクは、前記ノード間のデータフローを表すこと、

ユーザ入力に基づいて前記グループから選択された前記第1のデータセットオブジェクトを受信することであって、前記第1のデータセットオブジェクトは前記複数のモードを含み、各モードは、変換ロジック及び入力ポート又は出力ポートのうちの少なくとも1つを含む、受信すること、

30

ユーザ入力に基づいて前記第1のデータセットオブジェクトに対して選択されたモードを受信すること、

前記選択されたモードを受信した後、前記データフローグラフに前記第1のノードを生成することであって、前記第1のノードは、前記選択されたモードの変換ロジック、任意の入力ポート、及び任意の出力ポートを含むこと、

リンクを使用して、前記第1のノードを前記データフローグラフ内の少なくとも1つのノードに接続すること、

40

を含む、請求項14に記載の計算システム。

【請求項36】

前記第1のデータセットオブジェクトは少なくとも1つのパラメータをさらに含み、前記第1のノードは前記パラメータの値をさらに含む、請求項35に記載の計算システム。

【請求項37】

前記複数のモードのうちの第1のモードの前記変換ロジックは、データストアからデータレコードを読み取る変換ロジックを含む、請求項35に記載の計算システム。

【請求項38】

前記複数のモードのうちの第2のモードの前記変換ロジックは、データレコードを前記データストアに書き込む変換ロジックを含む、請求項37に記載の計算システム。

50

【請求項 39】

前記第1のノードは、前記第1のデータセットオブジェクトを識別する情報を含む、請求項35に記載の計算システム。

【請求項 40】

ユーザ入力に基づいて前記グループから選択された前記第1のデータセットオブジェクトを受信する手段と、

複数のデータフローグラフを識別する手段であって、前記データフローグラフのそれぞれは、前記第1のデータセットオブジェクトを識別するノードを含む、手段と、
をさらに含む、請求項14に記載の計算システム。

【請求項 41】

前記データセットオブジェクトの各モードは、ノード及び前記ノードを接続するリンクを含む別個のデータフローグラフで表され、前記リンクは前記ノード間のデータフローを表す、請求項14に記載の計算システム。

【請求項 42】

前記関連付けられたデータセットオブジェクトと前記関連付けられたノードとの間の関連付けを生成することは、

前記第1のノードを含む前記データフローグラフを含む複数のデータフローグラフに対応するデータ構造にアクセスすることであって、各データフローグラフは、複数のノード及び前記ノードを接続するリンクを含み、前記ノードは変換ロジックを含み、前記リンクは前記ノード間のデータフローを表すこと、

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別することであって、少なくとも1つの識別されたノードは1つ又は複数のパラメータを含むこと、

データセットオブジェクトを生成することであって、

前記識別されたノードの前記変換ロジックに基づいて、前記生成されたデータセットオブジェクトの複数のモードを識別すること、

前記識別されたノードに関連付けられたパラメータに基づいて、前記生成されたデータオブジェクトのパラメータを識別すること

を含むこと、

前記生成されたデータセットオブジェクトを前記グループに記憶すること、
を含む、請求項14に記載の計算システム。

【請求項 43】

前記生成されたデータセットオブジェクトを識別する情報に基づいて、前記複数のノードのそれぞれを更新して、前記データストアにアクセスする手段をさらに含む、請求項42に記載の計算システム。

【請求項 44】

同じデータストアにアクセスする前記複数のデータフローグラフの複数のノードを識別することは、

前記複数のデータフローグラフのそれぞれを実行すること、

前記データフローグラフの実行中、前記複数のデータフローグラフの前記ノードのパラメータに割り当てられたパラメータ値を記録すること、

前記記録されたパラメータ値に基づいて前記複数のノードを識別すること、
を含む、請求項42に記載の計算システム。

【発明の詳細な説明】**【技術分野】****【0001】**

関連出願の相互参照

本願は、2010年10月25日に出願された米国仮特許出願第61/406,438号の優先権を主張するものであり、この特許出願を参照により本明細書に援用する。

【0002】

背景

本記載はデータセットオブジェクトの管理に関する。

【背景技術】

【0003】

データフローグラフはデータを処理する。データフローグラフは、データソース、データシンク、及びコンポーネントを含む。データソース、データシンク、及びコンポーネントは、データのあるコンポーネントから別のコンポーネントに、又はデータソースからコンポーネントに、又はコンポーネントからデータシンクに流れるようにする接続により接続される。データフローグラフは、接続を表すリンク（弧とも呼ばれる）により接続されるデータソース、データシンク、及びコンポーネントを表すノード（頂点とも呼ばれる）を有する有向グラフで視覚的に表すことができる。

10

【発明の概要】

【課題を解決するための手段】

【0004】

概要

一態様では、一般には、グラフに基づくデータ処理のためにデータセットオブジェクトを管理する方法は、1つ又は複数のデータセットオブジェクトのグループをデータ記憶システムに記憶することであって、データセットオブジェクトのそれぞれは各データセットを表す、記憶すること、及びグループ内の少なくとも第1のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第1のノードとの間の関連付けを生成することを含み、第1のノードは、データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、第1のデータセットオブジェクトは、異なる変換ロジックが第1のノードで処理されるデータに適用される複数のモードを含む。

20

【0005】

態様は以下の特徴のうちの1つ又は複数を含み得る。

【0006】

関連付けられたデータセットオブジェクトと関連付けられたノードとの関連付けを生成することは、データフローグラフに対応するデータ構造にアクセスすることであって、データフローグラフは、第1のノードを含む複数のノード及びノードを接続するリンクを含み、リンクは、ノード間のデータフローを表す、アクセスすること、ユーザ入力に基づいてグループから選択された第1のデータセットオブジェクトを受信することであって、第1のデータセットオブジェクトは複数のモードを含み、各モードは、変換ロジック及び入力ポート又は出力ポートのうちの少なくとも1つを含む、受信すること、ユーザ入力に基づいて第1のデータセットオブジェクトに選択されたモードを受信すること、選択されたモードを受信した後、データフローグラフに第1のノードを生成することであって、第1のノードは、選択されたモードの変換ロジック、任意の入力ポート、及び任意の出力ポートを含む、生成すること、並びにリンクを使用して、第1のノードをデータフローグラフ内の少なくとも1つのノードに接続すること、を含む。

30

【0007】

第1のデータセットオブジェクトは少なくとも1つのパラメータをさらに含み、第1のノードはパラメータの値をさらに含む。

40

【0008】

複数のモードのうちの第1のモードの変換ロジックは、データストアからデータレコードを読み取る変換ロジックを含む。

【0009】

複数のモードのうちの第2のモードの変換ロジックは、データレコードをデータストアに書き込む変換ロジックを含む。

【0010】

第1のノードは、第1のデータセットオブジェクトを識別する情報を含む。

50

【 0 0 1 1 】

この方法は、ユーザ入力に基づいてグループから選択された第1のデータセットオブジェクトを受信すること、及び複数のデータフローグラフを識別することであって、データフローグラフのそれぞれは、第1のデータセットオブジェクトを識別するノードを含む、識別することをさらに含み得る。

【 0 0 1 2 】

データセットオブジェクトの各モードは、ノード及びノードを接続するリンクを含む別個のデータフローグラフで表され、リンクはノード間のデータフローを表す。

【 0 0 1 3 】

関連付けられたデータセットオブジェクトと関連付けられたノードとの関連付けを生成することは、第1のノードを含むデータフローグラフを含む複数のデータフローグラフに対応するデータ構造にアクセスすることであって、各データフローグラフは、複数のノード及びノードを接続するリンクを含み、ノードは変換ロジックを含み、リンクはノード間のデータフローを表す、アクセスすること、同じデータストアにアクセスする複数のデータフローグラフの複数のノードを識別することであって、少なくとも1つの識別されたノードは1つ又は複数のパラメータを含む、識別すること、データセットオブジェクトを生成することであって、識別されたノードの変換ロジックに基づいて、生成されたデータセットオブジェクトの複数のモードを識別すること、及び識別されたノードに関連付けられたパラメータに基づいて、生成されたデータオブジェクトのパラメータを識別することを含む、生成すること、生成されたデータセットオブジェクトをグループに記憶することを含む。

10

20

【 0 0 1 4 】

この方法は、生成されたデータセットオブジェクトを識別する情報に基づいて、複数のノードのそれぞれを更新して、データストアにアクセスすることをさらに含み得る。

【 0 0 1 5 】

同じデータストアにアクセスする複数のデータフローグラフの複数のノードを識別することは、複数のデータフローグラフのそれぞれを実行すること、データフローグラフの実行中、複数のデータフローグラフのノードのパラメータに割り当てられたパラメータ値を記録すること、及び記録されたパラメータ値に基づいて複数のノードを識別することを含む。

30

【 0 0 1 6 】

別の態様では、一般に、コンピュータ可読記憶媒体は、グラフに基づくデータ処理のためにデータセットオブジェクトを管理するコンピュータプログラムを記憶する。コンピュータプログラムは、計算装置に、1つ又は複数のデータセットオブジェクトのグループをデータ記憶システムに記憶することであって、データセットオブジェクトのそれぞれは各データセットを表す、記憶すること、及びグループ内の少なくとも第1のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第1のノードとの関連付けを生成することであって、第1のノードは、データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、第1のデータセットオブジェクトは、異なる変換ロジックが第1のノードで処理されるデータに適用される複数のモードを含む、生成することを行わせる命令を含む。

40

【 0 0 1 7 】

別の態様では、一般に、グラフに基づくデータ処理のためにデータセットオブジェクトを管理する計算システム。計算システムは、1つ又は複数のデータセットオブジェクトのグループを記憶するデータ記憶システムであって、データセットオブジェクトのそれぞれは各データセットを表す、データ記憶システムと、データ記憶システムに結合され、グループ内の少なくとも第1のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第1のノードとの関連付けを生成するように構成される少なくとも1つのプロセッサであって、第1のノードは、データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、第1のデー

50

タセットオブジェクトは、異なる変換ロジックが第1のノードで処理されるデータに適用される複数のモードを含む、少なくとも1つのプロセッサとを含む

【0018】

別の態様では、一般に、グラフに基づくデータ処理のためにデータセットオブジェクトを管理する計算システム。計算システムは、1つ又は複数のデータセットオブジェクトのグループを記憶する手段であって、データセットオブジェクトのそれぞれは各データセットを表す、手段と、グループ内の少なくとも第1のデータセットオブジェクトと、データ処理システムでデータを処理するデータフローグラフの少なくとも第1のノードとの関連付けを生成する手段であって、第1のノードは、データフローグラフ内のリンクにより表されるデータフロー内のデータのソース又はシンクを表し、第1のデータセットオブジェクトは、異なる変換ロジックが第1のノードで処理されるデータに適用される複数のモードを含む、手段とを含む。

10

【0019】

態様は、以下の利点のうちの1つ又は複数を含むことができる。データセットオブジェクトを、複数のデータフローグラフで再使用し得る。データレコードの更新を簡易化し得る。データの従属性を特定し得る。アプリケーション開発を簡易化し得る。

【0020】

本発明の他の特徴及び利点が以下の説明及び特許請求の範囲から明らかになるう。

【図面の簡単な説明】

【0021】

20

【図1】データセットオブジェクト技法を使用することができる例示的なデータ処理システム100を示す。

【図2A】データセットオブジェクトの管理に使用することができる例示的な手順200のフローチャートを示す。

【図2B】データセットオブジェクトの管理に使用することができる例示的な手順200のフローチャートを示す。

【図3】開発環境のユーザインタフェース例を示す。

【図4】データセットオブジェクトに関連付けられた例示的なパーソナリティを示す。

【図5】カタログ内のデータフローグラフとデータセットオブジェクトとの関係を示す。

【図6】カタログ内のデータセットオブジェクトへの既存のノードの関連付けを示す。

30

【図7】パラメータ値に基づいてデータセットオブジェクトへのノードの関連付けを示す。

【発明を実施するための形態】

【0022】

説明

図1は、データセットオブジェクト技法を使用することができる例示的なデータ処理システム100を示す。システム100はデータソース102を含み、データソース102は、記憶装置又はオンラインデータストリームへの接続等の1つ又は複数のデータソースを含み得、各データソースは、任意の様々な記憶フォーマット（例えば、データベーステーブル、スプレッドシートファイル、平文ファイル、又はメインフレームで使用されるネイティブフォーマット）でデータを記憶し得る。実行環境104は、事前実行モジュール106及び実行モジュール112を含む。実行環境104は、UNIX（登録商標）オペレーティングシステム等の適したオペレーティングシステムの制御下で1つ又は複数の汎用コンピュータでホストし得る。例えば、実行環境104は、ローカルである（例えば、SMPコンピュータ等のマルチプロセッサシステム）か、又はローカルに分散する（例えば、クラスタとして結合された複数のプロセッサ又はMPP）か、又はリモートであるか、又はリモートに分散する（例えば、ローカルエリアネットワーク（LAN）及び/又は広域ネットワーク（WAN）を介して結合された複数のプロセッサ）か、又はこれらの任意の組み合わせである複数の中央演算処理装置（CPU）を使用するコンピュータシステムの構成を含む複数ノード並列計算環境を含むことができる。

40

50

【 0 0 2 3 】

事前実行モジュール 1 0 6 は、データフローグラフ及びデータフローグラフのコンパイル、実行環境 1 0 4 がアクセス可能なデータ記憶システム 1 1 6 への / からのコンパイルされたデータフローグラフの記憶 / ロード等の他の実行可能プログラム、又はデータフローグラフの管理に関連付けられた他のタスクを実行する準備として、様々なタスクを実行するように構成される。実行モジュール 1 1 2 は、データをデータソース 1 0 2 から読み取り、データフローグラフの形態で表現されるプログラムを含む実行可能プログラムを使用してデータを処理する。

【 0 0 2 4 】

データソース 1 0 2 を提供する記憶装置は、実行環境 1 0 4 にローカルであってもよく、例えば、実行環境 1 0 4 を実行するコンピュータに接続された記憶媒体（例えば、ハードドライブ 1 0 8 ）に記憶されてもよく、又は実行環境 1 0 4 からリモートであってもよく、例えば、リモート接続を介して実行環境 1 0 4 を実行するコンピュータと通信するリモートシステム（例えば、メインフレーム 1 1 0 ）でホストしてもよい。

【 0 0 2 5 】

実行モジュール 1 1 2 は、実行環境 1 0 4 がアクセス可能なデータ記憶システム 1 1 6 に記憶されたデータセットオブジェクトのカatalog 1 1 4 を使用して、データフローグラフを実行する。Catalog 1 1 4 は、事前処理モジュール 1 0 6 により管理されて、より詳細に後述するように、異なるモードに構成することができるデータセットオブジェクトを提供する。Catalog は、データセットオブジェクトのグループをデータ構造又は特定のロケーション（例えば、ファイルシステムディレクトリ）に記憶し得、又は様々なロケーションに記憶されたデータオブジェクトのグループを識別する情報を含み得る。いくつかの実施態様では、Catalog は、検索又は他の動作を可能にするインデックス又はメタデータ等の情報を含む。開発者 1 2 0 がデータフローグラフを開発し、Catalog 1 1 4 と対話することが可能な開発環境 1 1 8 も、データ記憶システム 1 1 6 にアクセス可能である。

【 0 0 2 6 】

開発環境 1 1 8 は、いくつかの実施態様では、有向リンク（作業要素の流れを表す）で接続される頂点（コンポーネント又はデータセットを表す）を含むデータフローグラフとしてアプリケーションを開発するシステムである。例えば、そのような環境については、「Managing Parameters for Graph-Based Applications」という名称の米国特許出願公開第 2 0 0 7 / 0 0 1 1 6 6 8 号により詳細に説明されており、この公開出願を参照により本明細書に援用する。そのようなグラフに基づく計算を実行するシステムは、米国特許第 5 , 5 6 6 , 0 7 2 号、EXECUTING COMPUTATIONS EXPRESSED AS GRAPHSに記載されており、この特許を参照により本明細書に援用する。このシステムに従って作成されたデータフローグラフは、グラフコンポーネントにより表される個々のプロセス内外への情報を取得する方法、例えば、プロセス間で情報を移動させ、プロセスの実行順序を定義する方法を提供する。このシステムは、プロセス間通信方法を選ぶアルゴリズムを含む（例えば、グラフのリンクに従う通信路は、TCP / IP 又は UNIX（登録商標）ドメインソケットを使用することができ、又は共有メモリを使用して、プロセス間でデータを渡すことができる）。

【 0 0 2 7 】

実行モジュール 1 1 2 は、異なる形態のデータベースシステムを含む様々なタイプのシステムからデータを受信することができる。データは、ヌル値を含むことが可能な、各フィールド（「属性」又は「列」とも呼ばれる）に値を有するレコードとして編成し得る。データソースから最初にデータを読み取る際、実行モジュール 1 1 2 は通常、そのデータソース内のレコードについてのいくつかの初期フォーマット情報で開始する。いくつかの状況では、データソースのレコード構造は、最初に分からず、代わりに、データソースの分析後に特定し得る。レコードについての初期情報は、別個の値を表すビット数、レコード内のフィールドの順序、及びビットで表される値のタイプ（例えば、文字列、署名付き / 署名なし整数）を含むことができる。

10

20

30

40

50

【 0 0 2 8 】

開発環境 1 1 8 は、実行環境 1 0 4 で実行すべきデータフローグラフの構築及び / 又は編集を行うために開発者 1 2 0 が使用するインタフェースを提供する。データフローグラフには、データ記憶システム 1 1 6 に記憶される実行可能データ構造が関連付けられ、データフローグラフは、開発環境 1 1 8 において、ノード及びノードを接続するリンクを含む有向グラフとして表現される。開発環境は、開発中のデータフローグラフに対応するデータ構造にアクセスすることが可能である。データフローグラフのノードは、データソース、データシンク、及びデータ処理コンポーネントを表す。リンクは、ノード間のデータフローを表す。開発環境 1 1 8 は、データセットオブジェクトのカatalog 1 1 4 へのアクセスを提供することも可能である。Catalog 1 1 4 は、データ記憶システム 1 1 6 内のデータストア又は別のデータストア内に配置し得る。

10

【 0 0 2 9 】

Catalog 1 1 4 は、ユーザ入力に基づいて選択することができる 1 つ又は複数のデータセットオブジェクトを含む。データセットオブジェクトは、データセットオブジェクトが提供する機能を決める複数のモード（「パーソナリティ」と呼ばれる）を含むデータフローグラフノードである。例えば、データセットオブジェクトは、データソースパーソナリティ及びデータシンクパーソナリティを含み得る。データセットオブジェクトが、データフローグラフ内のノードとして使用されるように選択される場合、データセットオブジェクトは、例えば、ユーザ入力に基づいてパーソナリティの 1 つを採用する。例えば、開発環境 1 1 8 は、データフローグラフ（例えば、リンクで接続されたノード）の表現を含む Catalog 及びキャンバスの表現を含むグラフィカルユーザインタフェース（GUI）を含み得る。ユーザ（例えば、開発者 1 2 0）は、データセットオブジェクトを Catalog からキャンバスにドラッグアンドドロップすることができる。データセットオブジェクトの表現がキャンバスに配置されると、ユーザはデータセットオブジェクトのパーソナリティを選択することができる。

20

【 0 0 3 0 】

選択されたパーソナリティを使用して、データフローグラフに新しいノードを生成する。選択されたパーソナリティには、何らかの方法でデータを変更する変換ロジックが関連付けられる。例えば、データソースパーソナリティは、データレコードをデータストア（例えば、データベーステーブル、ファイル、又は他のデータストア）から読み取り、何らかの変換を実行する変換ロジックを含むことができる。データシンクパーソナリティは、何らかの変換を実行した後、データレコードをデータストアに書き込む変換ロジックを含むことができる。いくつかの実施態様では、データセットオブジェクトのパーソナリティの変換ロジックは、データフローグラフを使用して実施することができる。いくつかの実施態様では、データセットオブジェクトの各パーソナリティには、ノード及びノードを接続するリンクを含む異なる各データフローグラフが関連付けられる。

30

【 0 0 3 1 】

各パーソナリティは、入力ポート（例えば、データシンクパーソナリティ）を有してもよく、出力ポート（例えば、データソースパーソナリティ）を有してもよく、入力ポート及び出力ポートの両方（例えば、データ変換パーソナリティ）を有してもよく、又はポートを全く有さなくてもよい（例えば、ルックアップテーブルパーソナリティ）。データセットオブジェクトに選択されたパーソナリティを受信することにより生成される新しいノードは、選択されたパーソナリティの変換ロジック及びポートを反映する。いくつかの実施態様では、データセットオブジェクトは、データソースのロケーションを識別する。データセットオブジェクトは、データソースのロケーションを識別するパラメータを含み得る。いくつかの実施態様では、パラメータは、データフローグラフの実行中に決定される値を含む動的パラメータを使用して、データソースのロケーションを識別する。データソースにアクセスするデータセットオブジェクトは、データのフォーマットを識別するとともに、データが圧縮又は暗号化されているか否かを識別することができる。

40

【 0 0 3 2 】

50

生成されたノードは、リンクを使用して、生成されたノードをデータフローグラフ内の少なくとも1つのノードに接続することにより、データフローグラフに組み込むことができる。いくつかの実施態様では、生成されたノードは、データソースからのデータをデータフローグラフに提供するか、又はデータフローグラフからのデータをデータストアに記憶する。

【0033】

図2Aは、データセットオブジェクトを管理するシステムにより実施することができる例示的なプロセス200のフローチャートを示す。便宜上、プロセスについてプロセスを実行するシステムに関連して説明する。

【0034】

プロセス200は、データフローグラフに対応するデータ構造にアクセスする202。データ構造は、ノード及びノードを接続するリンクを含み、リンクはノード間のデータフローを表す。

【0035】

プロセス200は、データセットオブジェクトのカタログを提供する204。上述したように、カタログは1つ又は複数のデータセットオブジェクトを含むことができる。

【0036】

プロセス200は、カタログから選択されたデータセットオブジェクトを受信する206。例えば、ユーザは、データセットオブジェクトをカタログからキャンバスにドラッグすることにより、データカタログからデータセットオブジェクトを選択することができる。上述したように、データセットオブジェクトは複数のパーソナリティを含み、各パーソナリティは、変換ロジックと、少なくとも1つの入力ポート及び/又は少なくとも1つの出力ポートを含む。

【0037】

プロセス200は、データセットオブジェクトに選択されたパーソナリティを受信する208。上述したように、ユーザは、例えば、ラジオボタン又はドロップダウンメニュー上の値を選択することにより、選択されたデータセットオブジェクトのパーソナリティを選択することができる。

【0038】

プロセス200は、データフローグラフ内にノードを生成する210。生成されたノードは、選択されたパーソナリティの変換ロジック及び任意の入力及び出力ポートを含む。

【0039】

プロセス200は、生成されたノードをデータフロー内の少なくとも1つのノードに接続する212。

【0040】

図2Bは、データセットオブジェクトを管理するシステムにより実施することができる例示的なプロセス230のフローチャートを示す。便宜上、プロセスについて、プロセスを実行するシステムに関連して説明する。

【0041】

プロセス230は、複数のデータフローグラフにアクセスする(232)。上述したように、各データフローグラフは、ノード及びノードを接続するリンクを含む。ノードは変換ロジックを含み、リンクはノード間のデータフローを表す。

【0042】

プロセス230は、同じデータストアにアクセスする複数のノードを識別する234。各ノードはパラメータ及び変換ロジックを含む。

【0043】

プロセス230は、データセットオブジェクトを生成する236。データセットオブジェクトを生成するために、プロセスは、複数のノードの変換ロジックに基づいて複数のパーソナリティを識別する238。プロセス230は、ノードに関連付けられたパラメータに基づいてパラメータも識別する240。プロセス230は、生成されたデータセットオ

10

20

30

40

50

プロジェクトをデータカタログに記憶する 2 4 2。

【 0 0 4 4 】

図 3 は、開発環境、例えば、図 1 の開発環境 1 1 8 のユーザインタフェース例 3 0 0 を示す。ユーザインタフェース 3 0 0 は、複数のデータセットオブジェクト 3 0 4 a ~ e を含むデータセットオブジェクトのカタログ 3 0 2 を含む。ユーザインタフェース 3 0 0 はキャンバス 3 0 6 も含み、キャンバス 3 0 6 において、ユーザはデータフローを作成又は変更することができる。キャンバスは、データフローグラフに含まれるノードの表現を表示する。この例では、データフローグラフは、第 1 のデータソース 3 1 0、第 2 のデータソース 3 1 2、第 1 のコンポーネントノード 3 1 4、第 2 のコンポーネントノード 3 1 6、データシンク 3 2 0、及びルックアップテーブル 3 0 8 を含む。データフローグラフを通してのデータの流れは、リンクにより定義される。例えば、第 1 のデータソース 3 1 0 は出力ポート 3 2 2 を含み、出力ポート 3 2 2 は、リンク 3 2 6 によりコンポーネントノード 3 1 4 の入力ポート 3 2 4 に接続される。

10

【 0 0 4 5 】

ユーザ（図示せず）は、データセットオブジェクト（例えば、データセットオブジェクト 3 0 4 b）をカタログ 3 0 2 からキャンバス 3 0 6 にドラッグして、新しいノード 3 1 8 を生成する。ユーザは、例えば、属性インタフェース要素 3 3 2 上のラジオボタンを使用してパーソナリティを選択することにより、新しいノード 3 1 8 のパーソナリティを選択する。パーソナリティが選択されると、新しいノード 3 1 8 は、パーソナリティに関連付けられた入力ポート、出力ポート、及び / 又は変換ロジックを提示する。後述するように、新しいオブジェクトは、ユーザが設定し得るパラメータを含むこともできる。

20

【 0 0 4 6 】

図 4 は、データセットオブジェクトに関連付けられた例示的なパーソナリティを示す。データセットオブジェクト 4 0 0 には、1 つ又は複数のパラメータ 4 1 4 を関連付けることができる。例えば、データセットオブジェクト 4 0 0 には、ラベル、ファイル名、レコードフォーマット、データベーステーブル名、又は他の任意の情報を定義する値を有するパラメータを関連付けることができる。いくつかの実施態様では、パラメータ値を実行時に動的に決定することができる。

【 0 0 4 7 】

データセットオブジェクト 4 0 0 は複数のパーソナリティを含むことができ、各自の特徴を有する各パーソナリティは、各自の入力ポート、出力ポート、及び / 又は変換ロジックを含む。例えば、データセットオブジェクトはデータソースパーソナリティ 4 0 4 を含むことができ、データソースパーソナリティ 4 0 4 は、ファイル、データベーステーブル、又は他のデータ構造からデータを読み取り、データフローグラフに提供する。データセットオブジェクト 4 0 0 はデータシンクパーソナリティ 4 1 0 を含むこともでき、データシンクパーソナリティ 4 1 0 は、入力ポートでデータレコードフローを受け入れ、データストア、例えば、データベーステーブル、ファイル、又は他の何らかのデータストアに記憶する。データセットオブジェクト 4 0 0 はパラレルデータソースパーソナリティ 4 0 2 を含むことができ、パラレルデータソースパーソナリティ 4 0 2 は、データを複数のパラレルデータ構造から読み取り、データフローグラフに提供する。同様に、データセットオブジェクト 4 0 0 は、データレコードストリームを複数のデータストアに提供するパラレルデータシンクパーソナリティ 4 1 2 を含むことができる。

30

40

【 0 0 4 8 】

データセットオブジェクト 4 0 0 は、ルックアップテーブルパーソナリティ 4 1 8 を含むことができる。一般に、ルックアップテーブルは入力ポート又は出力ポートを含まない。ルックアップテーブルパーソナリティは、他のノードでの高速データ参照のためにデータソースをメモリに読み出す。

【 0 0 4 9 】

データセットオブジェクト 4 0 0 は、1 つ又は複数のコンポーネントパーソナリティ、例えば、パーソナリティ 4 0 6 を含むことができる。一般に、コンポーネントパーソナリ

50

ティは、少なくとも1つの入力ポート及び少なくとも1つの出力ポートを含み得、データフローに対して演算を実行する。例えば、コンポーネントは、データフローをフィルタリングして、特定の値を有するデータレコードを識別し得る。

【0050】

いくつかの実施態様では、パラメータにデータセットオブジェクトパーソナリティを関連付けることができる。例えば、パラレルデータシンクパーソナリティ412に、データフローを分割すべきファイルの名前を定義するパラメータ416が関連付けられる。

【0051】

いくつかの状況では、データセットオブジェクトは複数のデータシンク及びデータソースを含むことができる。例えば、データセットオブジェクトは、データをデータベーステーブルに書き込む第1のデータシンクパーソナリティ、データをフラットファイルに書き込む第2のデータシンクパーソナリティ、及びファイル転送プロトコル(FTP)を使用してデータをリモートサーバに書き込む第3のデータシンクパーソナリティを含むことができる。同様に、データセットオブジェクトは、データベーステーブルからデータを読み取る第1のデータソースパーソナリティ、データをフラットファイルから読み取る第2のデータソースパーソナリティ、及びFTPを使用してリモートサーバ上のデータにアクセスする第3のデータソースパーソナリティを含むことができる。いくつかの実施態様では、他のプロトコル及びデータアクセス方法を使用することができる(例えば、ハイパーテキスト転送プロトコル(HTTP)及び拡張可能マークアップ言語(XML)ドキュメント)。いくつかの実施態様では、データセットオブジェクトのすべてのデータソースパーソナリティは、同一の出力を生成し、データセットオブジェクトのすべてのデータシンクパーソナリティは、同一の入力を受け入れる。したがって、どのデータソースパーソナリティ又はデータシンクパーソナリティがデータフローグラフに含まれるかは、データフローグラフ内のその他のノードにトランスペアレントであり得る。

【0052】

上述したように、いくつかの実施態様では、データセットパーソナリティには、データに対して変換を実行するデータフローグラフとして実施される変換ロジックを関連付けることができる。例えば、データセットオブジェクトのデータソースパーソナリティは、XMLファイルを読み取り、処理のために、階層XMLデータを一連のシーケンシャルレコードに変換し得る。同様に、データセットオブジェクトのデータシンクパーソナリティは、一連のシーケンシャルレコードを受け入れ、レコードを階層XMLデータに変換して、XMLファイルに記憶し得る。

【0053】

データセットオブジェクトに関連付け得る他のパーソナリティは、データソースからのデータをリフレッシュするパーソナリティ及びデータソースにおいてレコードを更新するパーソナリティを含む。

【0054】

図5は、カタログ内のデータフローグラフ間及びデータセットオブジェクト間の関係を示す。カタログからのデータセットオブジェクトが、データフローグラフ内のノードの生成に使用される場合、コンポーネントとデータセットオブジェクトとの関係が維持される。この例では、データセットオブジェクト504を使用して、データフローグラフ506a内にノード508aを生成した。同様に、同じデータセットオブジェクト504を使用して、データフローグラフ505b~e内でノード508b~eを生成した。

【0055】

線510aで表されているように、カタログ502はデータセットオブジェクト504とノード508aとの関係を維持する。線510b~eで表されるように、データセットオブジェクト504とノード508b~eとの関係も、線510b~eで表されるように、同様に維持される。いくつかの実施態様では、ノード508a~eは、ノードの生成に使用されたデータセットオブジェクト504の識別情報を維持する。

【0056】

10

20

30

40

50

いくつかの実施態様では、データフローグラフ506a~eは、データセットオブジェクト504を参照することによりノード508a~eを実施する。この実施態様では、データセットオブジェクト504のパーソナリティに行われた変更は、データフローグラフ506a~eの実行に自動的に反映される。

【0057】

開発環境、例えば、図1の開発環境118のユーザは、維持された関係を使用して、データフローグラフ間の従属性を分析することができる。例えば、ユーザは、データセットオブジェクト504及びデータセットオブジェクトにより識別されるデータソースへの変更により影響を受けるデータフローグラフ506a~eのそれぞれを識別することができる。識別されたデータフローグラフは、変更を達成するために、必要に応じて変更することが

10

【0058】

図6は、カタログ内のデータセットオブジェクトへの既存のノードの関連付けを示す。いくつかの状況では、ユーザは、カタログからのデータセットオブジェクトを使用せずにデータストアにアクセスする、生成されたノード（例えば、データソース、データシンク、又はルックアップテーブル）を有し得る。これは、データフローグラフの作成時に適切なデータセットオブジェクトが存在しなかったか、又は同様に、開発者が、提供されたデータフローグラフの使用を怠ったために発生し得る。図1を参照すると、データ記憶システム116及び潜在的に実行環境104と併せて開発環境118は、これらのノードを識別し、後述するように、適切なデータセットオブジェクトに関連付けることができる。

20

【0059】

図6を再び参照すると、データシンク612は、データストア618のシンクであり、データストア618にアクセスする変換ロジックを含む。開発環境は、カタログ内のデータセットオブジェクト620が、データストア618のシンクでもあるデータシンクパーソナリティ（図示せず）を含むことを特定する。矢印624で表されるように、データシンク612にはデータセットオブジェクト620が関連付けられる。いくつかの実施態様では、データシンク612内の変換ロジックは、データセットオブジェクト620のデータシンクパーソナリティ内の変換ロジックで置換される。他の実施態様では、変換ロジックは置換されない。いくつかの状況では、データシンク612は、別個のパーソナリティとしてデータセットオブジェクト620に追加し得る（例えば、データセットオブジェクト620がデータシンクパーソナリティを含まない場合）。

30

【0060】

別の状況では、データシンク614はデータストア626のシンクであり、データストア626にアクセスする変換ロジックを含む。開発環境は、データシンク614に対応するデータセットオブジェクトがカタログ622に存在しないと特定する。その結果、開発環境は、データシンク614に基づいてデータシンクパーソナリティ（図示せず）を有する新しいデータセットオブジェクト630を作成する。新しいデータセットオブジェクト630は、カタログ622に追加される。新しいデータセットオブジェクト630にはデータシンク614が関連付けられる。

【0061】

40

いくつかの実施態様では、開発環境は、ノードのパラメータの値に基づいてノード（データソース、データシンク、ルックアップテーブル、又は他のコンポーネント）を関連付けることができる。例えば、データセットオブジェクトのパラメータ及びパラメータ値が、ノードのパラメータ及びパラメータ値に一致する場合、ノードにデータセットオブジェクトを関連付けることができる。いくつかの実施態様では、パラメータ及びパラメータ値が厳密に一致しない場合であっても、ノード及びデータセットオブジェクトを関連付け得る。例えば、いくつかのパラメータは、必要なパラメータとして見なされ得る（例えば、ノード及びデータセットオブジェクトを関連付けるには、パラメータ及びパラメータ値の両方が一致しなければならない）。他のパラメータは、存在する必要がないか、又はデータセットオブジェクトに関連付けるべきノードのデータセットオブジェクトのパラメータ

50

値に一致する必要がない任意選択的なパラメータと見なされ得る。

【 0 0 6 2 】

上述したように、いくつかのパラメータは、グラフが実行されるまで決定されない値を有することができる。この状況では、システムは、実行中、データフロウグラフのノードのパラメータの値を追跡することができる。実行中に決定されるパラメータ及びパラメータ値は、上述したように、データセットオブジェクトのパラメータ及びパラメータ値と比較することができる。

【 0 0 6 3 】

図 7 を参照すると、いくつかの実施態様では、実行中に決定されるパラメータ及びパラメータ値は、実行中、データセットオブジェクトに関連付けられた他のノードのパラメータ及びパラメータ値と比較される。例えば、データカタログ 7 0 2 は、値「\$inputfile」を有するパラメータ 7 0 6 「FileName」を有するデータセットオブジェクト 7 0 4 を含み、「\$inputfile」は、パラメータ 7 0 6 が、値が実行中に決定される動的パラメータであることを示す。接続線 7 0 8 により表されるように、データセットオブジェクト 7 0 4 には、データセットオブジェクト 7 0 4 のデータソースパーソナリティとして以前に生成されたデータソース 7 1 0 が関連付けられる。実行中、パラメータ 7 1 2 「FileName」の値は、「X:\Input_20101001.dat」であると決定される。データセットオブジェクト 7 0 4 に関連付けられないデータシンク 7 1 4 は、値「X:\Input_20101001.dat」を有するパラメータ 7 1 6 「FileName」を含む。一致するパラメータ 7 0 6 及び 7 1 6 並びにデータセットオブジェクト 7 0 4 にすでに関連付けられた別のパラメータ 7 1 2 の値に一致するパラメータ 7 1 6 の値に基づいて、データシンク 7 1 4 は、データセットオブジェクト 7 0 4 のデータシンクパーソナリティとして、データセットオブジェクト 7 0 4 に関連付けられると決定される。一般に、データセットオブジェクトへのノード（データソース、データシンク、又はルックアップテーブル）の一致では、複数のパラメータ及びパラメータ値が一致する必要があり得る。例えば、データセットオブジェクトへのノードの関連付けでは、データストアのロケーション及びレコードフォーマットを識別するパラメータが一致する必要があり得る。

【 0 0 6 4 】

上述したデータセットオブジェクト手法は、コンピュータで実行されるソフトウェアを使用して実施することができる。例えば、ソフトウェアは、それぞれが少なくとも 1 つのプロセッサ、少なくとも 1 つのデータ記憶システム（揮発性及び不揮発性メモリ及び / 又は記憶要素を含む）、少なくとも 1 つの入力装置又はポート、及び少なくとも 1 つの出力装置又はポートを含むプログラムされるか、又はプログラム可能な 1 つ又は複数のコンピュータシステム（分散、クライアント / サーバ、又はグリッド等の様々なアーキテクチャであり得る）で実行される 1 つ又は複数のコンピュータプログラム内のプロシージャを形成する。ソフトウェアは、例えば、データフロウグラフの設計及び構成に関連する他のサービスを提供するより大きなプログラムの 1 つ又は複数のモジュールを形成し得る。グラフのノード及び要素は、コンピュータ可読媒体に記憶されたデータ構造又はデータリポジトリに記憶されたデータモデルに準拠する他の編成データとして実施することができる。

【 0 0 6 5 】

ソフトウェアは、汎用又は専用プログラマブルコンピュータにより可読の CD - ROM 等の記憶媒体に提供してもよく、又はコンピュータにネットワークの通信媒体を介して送られ（伝播信号に符号化される）、コンピュータで実行されてもよい。すべての機能は、汎用コンピュータで実行されてもよく、又はコプロセッサ等の専用ハードウェアを使用して実行されてもよい。ソフトウェアは、ソフトウェアにより指定される計算の異なる部分が異なるコンピュータにより実行される分散様式で実施してもよい。そのような各コンピュータプログラムは、好ましくは、汎用又は汎用プログラマブルコンピュータにより可読の記憶媒体又は装置（例えば、固体状態メモリ、固体状態媒体、磁気媒体、又は光学媒体）に記憶又はダウンロードされて、記憶媒体又は装置がコンピュータシステムにより読み取られ、本明細書に記載のプロシージャを実行する場合、コンピュータを構成し動作させ

10

20

30

40

50

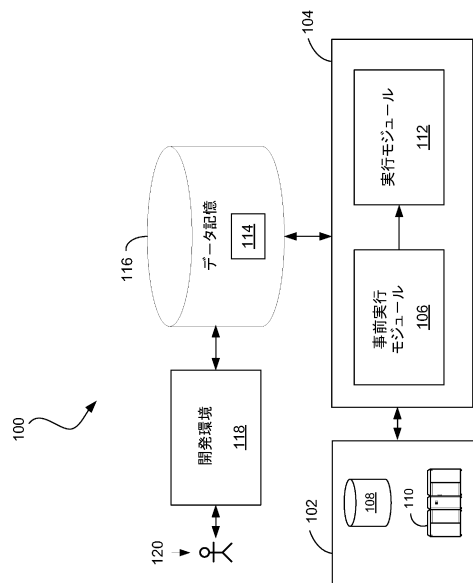
る。本発明によるシステムは、コンピュータ可読記憶媒として実施され、コンピュータプログラムが構成されるものと見なすこともでき、そうして構成された記憶媒体は、コンピュータシステムを特定の事前定義された様式で動作させて、本明細書に記載の機能を実行させる。

【 0 0 6 6 】

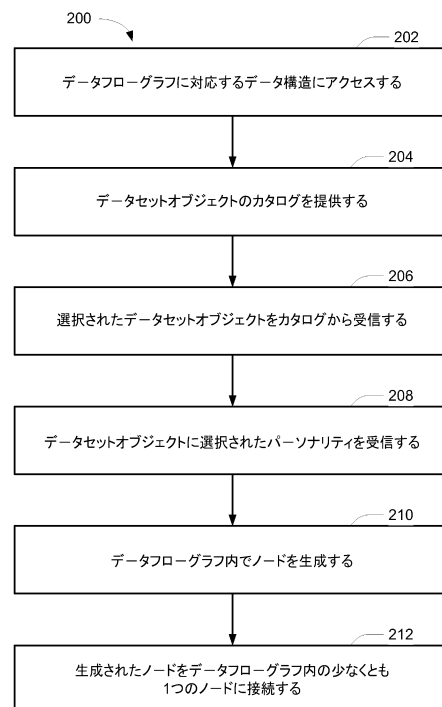
本発明のいくつかの実施形態について説明した。それに関わらず、本発明の趣旨及び範囲から逸脱せずに様々な変更を行い得ることが理解されよう。例えば、上述したステップのうちのいくつかは順序から独立し得、したがって、説明された順序と異なる順序で実行することもできる。上記説明が例示を目的とし、本発明の範囲を限定せず、本発明の範囲が、添付の特許請求の範囲により定義されることを理解されたい。例えば、上述したいくつかの機能ステップは、全体処理に実質的に影響せずに異なる順序で実行し得る。他の実施形態も以下の特許請求の範囲内にある。

10

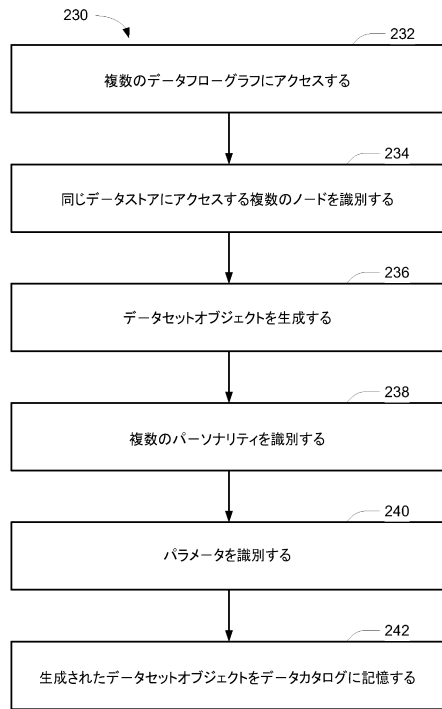
【圖 1】



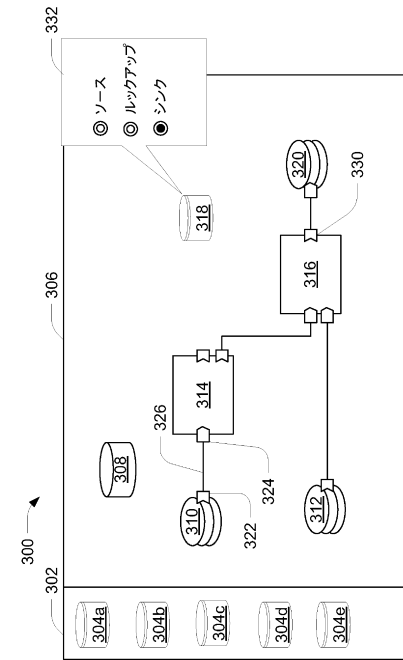
【 図 2 A 】



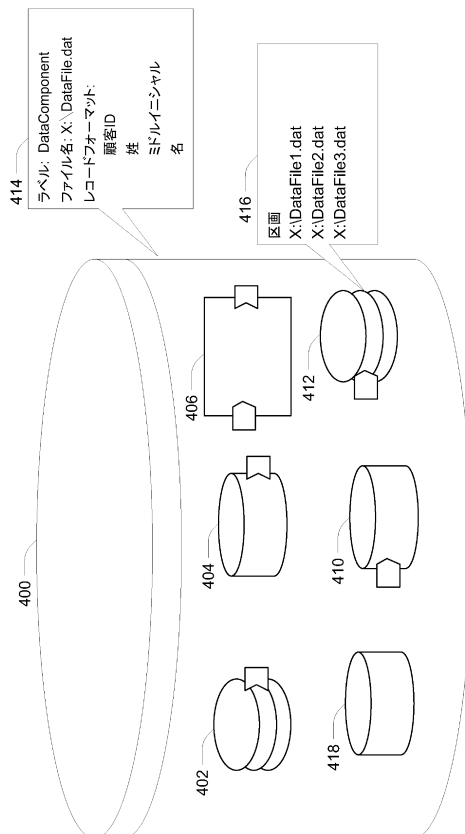
【図 2 B】



【図 3】



【図 4】



【図 5】

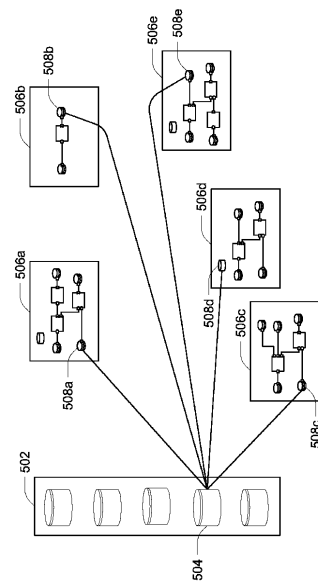


FIG. 5

フロントページの続き

- (72)発明者 シャピロ, リチャード エー .
アメリカ合衆国, マサチューセッツ州 02474, アーリントン, ホッジ ロード 39
- (72)発明者 スタンフィル, クレイグ ダブリュー .
アメリカ合衆国, マサチューセッツ州 01773, リンカーン, ハックルベリー ヒル ロード
43
- (72)発明者 ウェイス, アダム
アメリカ合衆国, マサチューセッツ州 02420, レキシントン, ローソン アヴェニュー 1
5

審査官 坂庭 剛史

- (56)参考文献 特開平07-044368(JP, A)
特表平05-507376(JP, A)
特表2008-524671(JP, A)
米国特許出願公開第2003/0016246(US, A1)
特表2008-547134(JP, A)
国際公開第2010/056867(WO, A1)
国際公開第2010/065511(WO, A1)
米国特許第07765529(US, B1)

- (58)調査した分野(Int.Cl., DB名)
G06F 9/44