

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
【部門区分】第 6 部門第 3 区分
【発行日】令和 3 年 7 月 26 日 (2021.7.26)

【公表番号】特表 2021-508389 (P2021-508389A)
【公表日】令和 3 年 3 月 4 日 (2021.3.4)
【年通号数】公開・登録公報 2021-012
【出願番号】特願 2020-529139 (P2020-529139)
【国際特許分類】

G 0 6 N 20/00 (2019.01)

【F I】

G 0 6 N 20/00

【手続補正書】

【提出日】令和 3 年 5 月 18 日 (2021.5.18)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

データ処理プラットフォーム上でジョブを処理するための方法であって、
処理のために第一ジョブを受け取るステップであって、前記第一ジョブは、前記データ処理プラットフォームの処理パイプラインによる前記第一ジョブの実行のためのパラメータ構成のセットを有する、ステップと、
前記処理パイプラインの一つ以上の処理モジュールを介して、前記第一ジョブの少なくとも一部分を実行するステップと、
を含み、
前記実行するステップは、第二ジョブの一部分の実行中に生成された少なくとも一つのデータ・シャードを使用するステップと、前記一つ以上の処理モジュールのうちの少なくとも一つの処理モジュールの出力を使用するステップであって、前記出力は、前記第二ジョブの一部分の実行中に生成される、ステップとを含む、
方法。

【請求項 2】

前記第二ジョブの一つ以上の部分は、前記第一ジョブのいずれの部分を実行するよりも前に処理される、請求項 1 に記載の方法。

【請求項 3】

前記処理モジュールは、一つ以上のアプリケーション・プログラミング・インタフェース (API) を含む、請求項 1 に記載の方法。

【請求項 4】

前記第一ジョブおよび前記第二ジョブのパラメータ構成は、前記第一ジョブおよび前記第二ジョブに関連するそれぞれのパラメータ・ファイルにおいて定義される、請求項 1 に記載の方法。

【請求項 5】

前記第二ジョブの一部分の実行中に生成された前記少なくとも一つのデータ・シャードを含むデータ・シャードは、中央データ・サービスによって管理される、請求項 1 に記載の方法。

【請求項 6】

前記中央データ・サービスは仮想化されている、請求項 5 に記載の方法。

【請求項 7】

前記中央データ・サービスは、中央データ・リポジトリおよび中央メタデータ・レジストリを含む、請求項 5 または 6 に記載の方法。

【請求項 8】

前記中央メタデータ・レジストリは、前記中央データ・リポジトリに記憶されたデータ・シャードについてのメタデータ・ファイルのセットを含み、前記処理パイプラインの処理モジュールの出力についての追加のメタデータ・ファイルのセットをさらに含む、請求項 7 に記載の方法。

【請求項 9】

前記第二ジョブの前記一部分の実行中に生成された前記少なくとも一つのデータ・シャードを、前記データ・シャードについてのメタデータ・ファイルに示される位置からプルするステップと、

既知のデータ・シャードによって定義されない任意のデータを、生データ・ソースからプルするステップと

をさらに含む、請求項 1 に記載の方法。

【請求項 10】

前記少なくとも一つのデータ・シャードをプルするステップ、および既知のデータ・シャードによって定義されない前記データをプルするステップは、グラフィカル・ユーザ・インタフェース (GUI) を介したユーザの選択に基づいてトリガされる、請求項 9 に記載の方法。

【請求項 11】

データ処理プラットフォーム上でジョブを処理するためのコンピュータ・プログラムであって、請求項 1 から 10 のいずれか一項に記載の方法における各ステップをプロセッサに実行させる、コンピュータ・プログラム。

【請求項 12】

請求項 11 に記載のコンピュータ・プログラムを記録した、コンピュータ可読ストレージ媒体。

【請求項 13】

データ処理プラットフォーム上でジョブを処理するためのシステムであって、プロセッサと、

コンピュータ・プログラムを保存したメモリと、
を備え、

前記コンピュータ・プログラムは、請求項 1 から請求項 10 のいずれか一項に記載の方法における各ステップを前記プロセッサに実行させる、システム。

【請求項 14】

データ処理プラットフォーム上でジョブを処理するための方法であって、

処理のために第一ジョブを受け取るステップであって、前記第一ジョブは、前記データ処理プラットフォームの処理パイプラインによる前記第一ジョブの実行のためのパラメータ構成のセットを有する、ステップと、

前記処理パイプラインの一つ以上の処理モジュールを介して、前記第一ジョブの少なくとも一部分を実行するステップであって、前記処理モジュールは、一つ以上のアプリケーション・プログラミング・インタフェース (API) を含む、ステップと、

を含み、

前記実行するステップは、第二ジョブの一部分の実行中に生成された少なくとも一つのデータ・シャードを使用するステップを含み、前記一つ以上の処理モジュールのうちの少なくとも一つの処理モジュールの出力を使用するステップであって、前記出力は、前記第二ジョブの一部分の実行中に生成される、ステップをさらに含み、

前記第二ジョブの一つ以上の部分は、前記第一ジョブのいずれの部分を実行するよりも前に処理され、

前記第二ジョブの一部分の実行中に生成された前記少なくとも一つのデータ・シャード

を含むデータ・シャードは、中央データ・リポジトリと中央メタデータ・レジストリとを含む中央データ・サービスによって管理され、

前記第一ジョブの前記少なくとも一部分を実行するステップは、

第二ジョブの一部分の実行中に生成された少なくとも一つのデータ・シャードを、データ・シャードについてのメタデータ・ファイルに示される位置からプルするステップと、

既知のデータ・シャードによって定義されない任意のデータを、生データ・ソースからプルするステップと

をさらに含む、

方法。