

【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】平成26年6月19日(2014.6.19)

【公開番号】特開2014-78085(P2014-78085A)

【公開日】平成26年5月1日(2014.5.1)

【年通号数】公開・登録公報2014-022

【出願番号】特願2012-224604(P2012-224604)

【国際特許分類】

G 0 6 F 17/30 (2006.01)

【F I】

G 0 6 F 17/30 1 1 0 C

G 0 6 F 17/30 4 1 5

【手続補正書】

【提出日】平成26年4月24日(2014.4.24)

【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 1

【補正方法】変更

【補正の内容】

【0 0 3 1】

図 3 は、設定ファイル D B に記憶される情報の例を示す図である。なお、図 3 では、入力ファイルが 2 つであり、出力ファイルが 2 つである場合を例示している。入力ファイルとは、突合プログラムが処理対象とする種類の異なる複数のファイルであり、出力ファイルとは、各スレーブ計算機が外部プログラムを実行した結果を格納する H D F S 上の共有ディレクトリ上の共有ファイルである。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 5

【補正方法】変更

【補正の内容】

【0 0 3 5】

さらに、設定ファイルには、Reduceアプリケーションやソートキーが設定される。例えば、Reduceアプリケーションには「a.out %in01 %in02 %out01 %out02」が設定されており、これは、外部プログラムに入力ファイルを渡す際に、1 番目の入力ファイルは第 1 引数で渡し、2 番目の入力ファイルは第 2 引数として渡すことを示す。また、ソートキー 0 1 には「4BYTE~8BYTE」が設定されており、これは、入力ファイル 0 1 の「4BYTE~8BYTE」をソートキーとして使用することを示す。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0 0 5 6

【補正方法】変更

【補正の内容】

【0 0 5 6】

Shuffle処理部 2 5 a は、Hadoopのreduce side joinなどで実行されるシャッフルソート処理を実行する処理部である。例えば、Shuffle処理部 2 5 a は、マスタ計算機 1 0 に対してReduceタスク要求を送信して、Reduceタスク情報を受信する。そして、Shuffle処理部 2 5 a は、受信したReduceタスク情報のMap結果情報にしたがって、自装置がシャッフルソート処理の対象とするMap処理結果を各スレーブ計算機から収集して中間ファイ

ル D B 2 2 a に格納する。一例を挙げると、Shuffle処理部 2 5 a は、Mapタスク I D 「task\_\_m\_\_2」のMap処理の結果をスレーブ計算機 3 0 から取得し、Mapタスク I D 「task\_\_m\_\_1」のMap処理の結果を自装置内の中間ファイル D B 2 2 a から取得する。なお、割当てられる手法は、Hadoopの分散手法を用いる。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0 0 5 9

【補正方法】変更

【補正の内容】

【0 0 5 9】

例えば、Reduce初期化部 2 5 b は、「a.out %in01 %in02 %out01 %out02」と指定して、1 番目の入力ファイルを外部プログラムの第 1 引数に渡す。あるいは、Reduce初期化部 2 5 b は、「ENVNAME.01=TRAN」と指定して、入力ファイル 0 1 のファイル名をTRAN環境変数に設定して、外部プログラムを呼び出すと設定する。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0 0 6 5

【補正方法】変更

【補正の内容】

【0 0 6 5】

図 7 に示すように、マスタ計算機 1 0 の設定読込部 1 4 は、設定ファイル D B 1 2 a から設定情報を読込む ( S 1 0 1 )。続いて、Mapタスク管理部 1 5 は、入力 D B サーバ 3 から入力ファイルが予め読み込まれた H D F S などの共有ファイルシステムから入力ファイルを 1 つ読込み ( S 1 0 2 )、データを所定サイズごとに分割し ( S 1 0 3 )、タスクリストに登録する ( S 1 0 4 )。例えば、Mapタスク管理部 1 5 は、入力ファイルの番号で、フォーマット処理およびMap処理のクラス名を切替えて登録する。

【手続補正 6】

【補正対象書類名】明細書

【補正対象項目名】0 0 9 1

【補正方法】変更

【補正の内容】

【0 0 9 1】

その後、Reduce初期化部 2 5 b は、ジョブ設定で指定されたReduceアプリケーションの引数の「%inNN」を入力用名前付きパイプ NN のファイル名で置き換える ( S 4 0 7 )。例えば、Reduce初期化部 2 5 b は、Reduceアプリケーション「a.out %in01 %in02 %out01 %out02」を「a.out ./tmp/in01.txt ./tmp/in02.txt %out01 %out02」に書き換える。

【手続補正 7】

【補正対象書類名】明細書

【補正対象項目名】0 0 9 7

【補正方法】変更

【補正の内容】

【0 0 9 7】

その後、Reduce初期化部 2 5 b は、ジョブ設定で指定されたReduceアプリケーションの引数の「%outNN」を出力用名前付きパイプ NN のファイル名で置き換える ( S 4 1 7 )。例えば、Reduce初期化部 2 5 b は、Reduceアプリケーション「a.out ./tmp/in01.txt ./tmp/in02.txt %out01 %out02」を「a.out ./tmp/in01.txt ./tmp/in02.txt ./tmp/out01.txt ./tmp/out02.txt」に書き換える。

【手続補正 8】

【補正対象書類名】図面

【補正対象項目名】図 3

【補正方法】変更

【補正の内容】

【図 3】

設定ファイルDBに記憶される情報の例を示す図

ジョブ設定	
入力データファイル名01=	/usr/transaction.txt
入力データファイル名02=	/usr/master.txt
入力フォーマット処理クラス名01=	固定長フォーマット
入力フォーマット処理オプション01=	レコード長:80BYTE
入力フォーマット処理クラス名02=	改行ありフォーマット
入力フォーマット処理オプション02=	改行コード:¥r¥n
Reduceアプリケーション=	a.out %in01 %in02 %out01 %out02
入力ファイル環境変数名01=	TRAN
入力ファイル環境変数名02=	MASTAR
出力データディレクトリ名=	/output/
出力ファイル名01=	processed.txt
出力ファイル名02=	error.txt
出力ファイル環境変数名01=	PROCOUT
出力ファイル環境変数名02=	ERROUT
ソートキー01=	4BYTE~8BYTE
ソートキー02=	12BYTE~16BYTE