

【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 06 F 17/30 (2006.01)

【F I】

G 06 F 17/30 1 1 0 C

G 06 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】0031

【補正方法】変更

【補正の内容】

【0031】

図3は、設定ファイルDB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HDFS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】0035

【補正方法】変更

【補正の内容】

【0035】

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、ソートキーには「4BYTE~8BYTE」が設定されており、これは、入力ファイル01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】0056

【補正方法】変更

【補正の内容】

【0056】

Shuffle処理部25a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25aは、マスタ計算機10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25aは、受信したReduceタスク情報のMap結果情報をしたがって、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DB22aに格納する。一例を挙げると、Shuffle処理部25aは、MapタスクID「task_m_2」のMap処理の結果をスレーブ計算機30から取得し、MapタスクID「task_m_1」のMap処理の結果を自装置内の中間ファイルDB22aから取得する。なお、割当される手法は、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】0059

【補正方法】変更

【補正の内容】

【0059】

例えば、Reduce初期化部25b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25bは、「ENVNAME.01=TRAN」と指定して、入力ファイル01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】0065

【補正方法】変更

【補正の内容】

【0065】

図7に示すように、マスタ計算機10の設定読込部14は、設定ファイルDB12aから設定情報を読み込む(S101)。続いて、Mapタスク管理部15は、入力DBサーバ3から入力ファイルが予め読み込まれたHDFSなどの共有ファイルシステムから入力ファイルを1つ読み込み(S102)、データを所定サイズごとに分割し(S103)、タスクリストに登録する(S104)。例えば、Mapタスク管理部15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】0091

【補正方法】変更

【補正の内容】

【0091】

その後、Reduce初期化部25b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(S407)。例えば、Reduce初期化部25b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】0097

【補正方法】変更

【補正の内容】

【0097】

その後、Reduce初期化部25b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(S417)。例えば、Reduce初期化部25b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