(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5675840号
(P5675840)

(45)発行日　平成27年2月25日(2015.2.25)　(24)登録日　平成27年1月9日(2015.1.9)

(51)Int.Cl.		F I		
G06F 17/30	(2006.01)	G06F	17/30	110B
G06F 12/00	(2006.01)	G06F	17/30	330B
		G06F	17/30	340Z
		G06F	12/00	513D

請求項の数 60　(全 23 頁)

(21)出願番号	特願2012-546227 (P2012-546227)
(86)(22)出願日	平成22年12月23日 (2010.12.23)
(65)公表番号	特表2013-516008 (P2013-516008A)
(43)公表日	平成25年5月9日 (2013.5.9)
(86)国際出願番号	PCT/US2010/061979
(87)国際公開番号	WO2011/079251
(87)国際公開日	平成23年6月30日 (2011.6.30)
審査請求日	平成25年12月20日 (2013.12.20)
(31)優先権主張番号	61/289,778
(32)優先日	平成21年12月23日 (2009.12.23)
(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　０２４２１　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　２０１
(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　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

最終頁に続く

(54)【発明の名称】クエリー管理

(57)【特許請求の範囲】

【請求項１】

１つ又は複数のデータ・ソースについて実行されるクエリーを管理するための方法であって、

少なくとも第１のクエリーを記憶媒体に記憶することと、

処理のために前記第１のクエリーを選択することと、

第１のクエリー間隔に対して前記１つ又は複数のデータ・ソースにおけるデータの第１の部分についての前記第１のクエリーを処理するようにクエリーエンジンに指示することと、

前記データの第１の部分についての前記第１のクエリーの処理に基づいて前記クエリーエンジンから第１の結果データを受信することと、

前記第１のクエリー間隔後に前記第１のクエリーの状態を前記記憶媒体に保存することと、

前記第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するように前記クエリーエンジンに指示することと、

前記第１のクエリーに関連付けられた優先順位を変更することと、

前記第２のクエリーの処理に基づいて前記クエリーエンジンから第２の結果データを受信することと、

前記第２のクエリーが中断され、前記クエリーエンジンによって処理されないように、前記第１のクエリーに関連付けられた前記優先順位に基づいて前記第２のクエリーを中断

することを決定することと、
前記第２のクエリー間隔後の第３のクエリー間隔の間に前記１つ又は複数のデータ・ソース内の第２のデータ部分についての前記第１のクエリーを処理するように前記クエリーエンジンに指示すること
を含む、方法。
【請求項２】
前記第１のクエリーに関係付けられた前記優先順位を前記記憶媒体に記憶することをさらに含み、
前記第１のクエリーに関係付けられた前記優先順位を変更することは、処理のために前記第１のクエリーを選択する前に行われ、
処理のために前記第１のクエリーを選択することが部分的に前記優先順位に基づいて前記クエリーを選択することを含む、請求項１記載の方法。
【請求項３】
前記第１のクエリー間隔が所定の時間量によって定義される、請求項１記載の方法。
【請求項４】
前記第１のクエリーに関連付けられた前記優先順位は、前記１つ又は複数のデータ・ソース内の前記データのうちのどのくらいの量が、前記第１のクエリー間隔に対して前記第１のクエリーが実行される前記データの第１の部分に含まれるか、に影響する、請求項３記載の方法。
【請求項５】
前記第１のクエリーを記憶することが、前記第１のクエリーを提供したリクエスタに通知される前に使用可能になるべき前記第１の結果データの量の通知しきい値を記憶することを含む、請求項１記載の方法。
【請求項６】
前記第１の結果データの前記量が前記通知しきい値を超えた時に前記リクエスタに通知することをさらに含み、前記第１のクエリーの前記状態を保存することが前記クエリーエンジンから受信した前記第１の結果データの前記量を記憶することを含む、請求項５記載の方法。
【請求項７】
前記リクエスタからの要求による前記第１の結果データを返すことと、前記リクエスタに返された前記第１の結果データの前記量を前記記憶媒体に記憶すること、をさらに含む、請求項６記載の方法。
【請求項８】
前記第１のクエリーを選択することが、前記クエリーエンジンから受信した前記第１の結果データの前記量と前記リクエスタに返された前記第１の結果データの前記量に基づくものである、請求項７記載の方法。
【請求項９】
前記第１のクエリーの前記状態を保存することが、
前記第１のクエリーを中断するよう前記クエリーエンジンに指示することと、
前記第１のクエリーが中断された後に前記第１のクエリーの状態を保存すること、
を含む、請求項１記載の方法。
【請求項１０】
前記第２のデータ部分に関する前記第１のクエリーを処理するように前記クエリーエンジンに指示することが、
前記第１のクエリーの前記保存状態をロードすることと、
前記第１のクエリーを再開するように前記クエリーエンジンに指示すること、
を含む、請求項９記載の方法。
【請求項１１】
前記第１のクエリーの前記状態を保存することが、二次索引へのオフセットを保存することを含む、請求項９記載の方法。

【請求項１２】
前記二次索引がブロック圧縮索引付きファイルの索引である、請求項１１記載の方法。
【請求項１３】
前記第１のクエリーを複数の副クエリーに分割することと、前記副クエリーのうちの少なくともいくつかを同時に処理するよう前記クエリーエンジンに指示すること、をさらに含む、請求項１記載の方法。
【請求項１４】
前記第１のクエリー間隔が始まった後で前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項１記載の方法。
【請求項１５】
前記第１のクエリー間隔が始まる前に前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項１記載の方法。
【請求項１６】
１つ又は複数のデータ・ソースについて実行されるクエリーを管理するためのコンピュータ・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・プログラムが、
少なくとも第１のクエリーを記憶媒体に記憶することと、
処理のために前記第１のクエリーを選択することと、
第１のクエリー間隔に対して前記１つ又は複数のデータ・ソース内のデータの第１の部分についての前記第１のクエリーを処理するようにクエリーエンジンに指示することと、
前記データの第１の部分についての前記第１のクエリーの処理に基づいて前記クエリーエンジンから第１の結果データを受信することと、
前記第１のクエリー間隔後に前記第１のクエリーの状態を前記記憶媒体に保存することと、
前記第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するように前記クエリーエンジンに指示することと、
前記第１のクエリーに関連付けられた優先順位を変更することと、
前記第２のクエリーの処理に基づいて前記クエリーエンジンから第２の結果データを受信することと、
前記第２のクエリーが中断され、前記クエリーエンジンによって処理されないように、前記第１のクエリーに関連付けられた前記優先順位に基づいて前記第２のクエリーを中断することを決定することと、
前記第２のクエリー間隔後の第３のクエリー間隔の間に前記１つ又は複数のデータ・ソース内の第２のデータ部分についての前記第１のクエリーを処理するように前記クエリーエンジンに指示することと、
をコンピュータに実行させるための命令を含む、コンピュータ可読記憶媒体。
【請求項１７】
１つ又は複数のデータ・ソースについて実行されるクエリーを管理するためのシステムであって、前記システムが、
少なくとも第１のクエリーを記憶する記憶媒体と、
前記１つ又は複数のデータ・ソース内のデータに関するクエリーを処理するように構成されたクエリーエンジンと、
サーバであって、
処理のために前記第１のクエリーを選択し、
第１のクエリー間隔に対して前記１つ又は複数のデータ・ソース内のデータの第１の部分についての前記第１のクエリーを処理するように前記クエリーエンジンに指示し、
前記データの第１の部分についての前記第１のクエリーの処理に基づいて前記クエリーエンジンから第１の結果データを受信し、
前記第１のクエリー間隔後に前記第１のクエリーの状態を前記記憶媒体に保存し、
前記第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するよ

うに前記クエリーエンジンに指示し、
前記第１のクエリーに関連付けられた優先順位を変更し、
前記第２のクエリーの処理に基づいて前記クエリーエンジンから第２の結果データを受信し、
前記第２のクエリーが中断され、前記クエリーエンジンによって処理されないように、前記第１のクエリーに関連付けられた前記優先順位に基づいて前記第２のクエリーを中断することを決定し、
前記第２のクエリー間隔後の第３のクエリー間隔の間に前記１つ又は複数のデータ・ソース内の第２のデータ部分についての前記第１のクエリーを処理するように前記クエリーエンジンに指示する、
ように構成されたサーバと、
を含む、システム。
【請求項１８】
前記コンピュータ・プログラムが、
前記第１のクエリーに関係付けられた前記優先順位を前記記憶媒体に記憶すること、
をコンピュータに実行させる命令をさらに含み、
前記第１のクエリーに関係付けられた前記優先順位を変更することは、処理のために前記第１のクエリーを選択する前に行われ、
処理のために前記第１のクエリーを選択することが部分的に前記優先順位に基づいて前記クエリーを選択することを含む、請求項１６記載のコンピュータ可読記憶媒体。
【請求項１９】
前記第１のクエリー間隔が所定の時間量によって定義される、請求項１６記載のコンピュータ可読記憶媒体。
【請求項２０】
前記第１のクエリーに関連付けられた前記優先順位は、前記１つ又は複数のデータ・ソース内の前記データのうちのどのくらいの量が、前記第１のクエリー間隔に対して前記第１のクエリーが実行される前記データの第１の部分に含まれるか、に影響する、請求項１９記載のコンピュータ可読記憶媒体。
【請求項２１】
前記第１のクエリーを記憶することが、前記第１のクエリーを提供したリクエスタに通知される前に使用可能になるべき前記第１の結果データの量の通知しきい値を記憶することを含む、請求項１６記載のコンピュータ可読記憶媒体。
【請求項２２】
前記コンピュータ・プログラムが、前記第１の結果データの前記量が前記通知しきい値を超えた時に前記リクエスタに通知することをコンピュータに実行させる命令をさらに含み、前記第１のクエリーの前記状態を保存することが前記クエリーエンジンから受信した前記第１の結果データの前記量を記憶することを含む、請求項２１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・プログラムが、前記リクエスタからの要求に対して前記第１の結果データを返すことと、前記リクエスタに返された前記第１の結果データの前記量を前記記憶媒体に記憶することと、をコンピュータに実行させる命令を含む、請求項２２記載のコンピュータ可読記憶媒体。
【請求項２４】
前記第１のクエリーを選択することが、前記クエリーエンジンから受信した前記第１の結果データの前記量と前記リクエスタに返された前記第１の結果データの前記量とに基づく、請求項２３記載のコンピュータ可読記憶媒体。
【請求項２５】
前記第１のクエリーの前記状態を保存することが、
前記第１のクエリーを中断するよう前記クエリーエンジンに指示することと、

前記第１のクエリーが中断された後に前記第１のクエリーの状態を保存すること、
を含む、請求項１６記載のコンピュータ可読記憶媒体。
【請求項２６】
前記第２のデータ部分に関する前記第１のクエリーを処理するように前記クエリーエンジンに指示することが、
前記第１のクエリーの前記保存状態をロードすることと、
前記第１のクエリーを再開するように前記クエリーエンジンに指示することと、
を含む、請求項２５記載のコンピュータ可読記憶媒体。
【請求項２７】
前記第１のクエリーの前記状態を保存することが、二次索引へのオフセットを保存することを含む、請求項２５記載のコンピュータ可読記憶媒体。
【請求項２８】
前記二次索引がブロック圧縮索引付きファイルの索引である、請求項２７記載のコンピュータ可読記憶媒体。
【請求項２９】
前記コンピュータ・プログラムが、前記第１のクエリーを複数の副クエリーに分割することと、前記副クエリーのうちの少なくともいくつかを同時に処理するよう前記クエリーエンジンに指示すること、をコンピュータに実行させる命令を含む、請求項１６記載のコンピュータ可読記憶媒体。
【請求項３０】
前記第１のクエリー間隔が始まった後で前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項１６記載のコンピュータ可読記憶媒体。
【請求項３１】
前記第１のクエリー間隔が始まる前に前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項１６記載のコンピュータ可読記憶媒体。
【請求項３２】
前記サーバは、
前記第１のクエリーに関係付けられた前記優先順位を前記記憶媒体に記憶し、
処理のために前記第１のクエリーを選択する前に前記第１のクエリーに関係付けられた前記優先順位を変更し、
処理のために前記第１のクエリーを選択することが部分的に前記優先順位に基づいて前記クエリーを選択することを含む、
ようにさらに構成されている、請求項１７記載のシステム。
【請求項３３】
前記第１のクエリー間隔が所定の時間量によって定義される、請求項１７記載のシステム。
【請求項３４】
前記第１のクエリーに関連付けられた前記優先順位は、前記１つ又は複数のデータ・ソース内の前記データのうちのどのくらいの量が、前記第１のクエリー間隔に対して前記第１のクエリーが実行される前記データの第１の部分に含まれるか、に影響する、請求項３３記載のシステム。
【請求項３５】
前記第１のクエリーを記憶することが、前記第１のクエリーを提供したリクエスタに通知される前に使用可能になるべき前記第１の結果データの量の通知しきい値を記憶することを含む、請求項１７記載のシステム。
【請求項３６】
前記サーバは、前記第１の結果データの前記量が前記通知しきい値を超えた時に前記リクエスタに通知するようにさらに構成され、前記第１のクエリーの前記状態を保存することが前記クエリーエンジンから受信した前記第１の結果データの前記量を記憶することを含む、請求項３５記載のシステム。

【請求項３７】
前記サーバは、前記リクエスタからの要求に対して前記第１の結果データを返し、前記リクエスタに返された前記第１の結果データの前記量を前記記憶媒体に記憶するようにさらに構成されている、請求項３６記載のシステム。
【請求項３８】
前記第１のクエリーを選択することが、前記クエリーエンジンから受信した前記第１の結果データの前記量と前記リクエスタに返された前記第１の結果データの前記量に基づくものである、請求項３７記載のシステム。
【請求項３９】
前記第１のクエリーの前記状態を保存することが、
前記第１のクエリーを中断するよう前記クエリーエンジンに指示することと、
前記第１のクエリーが中断された後に前記第１のクエリーの状態を保存すること、
を含む、請求項１７記載のシステム。
【請求項４０】
前記第２のデータ部分に関する前記第１のクエリーを処理するように前記クエリーエンジンに指示することが、
前記第１のクエリーの前記保存状態をロードすることと、
前記第１のクエリーを再開するように前記クエリーエンジンに指示すること、
を含む、請求項３９記載のシステム。
【請求項４１】
前記第１のクエリーの前記状態を保存することが、二次索引へのオフセットを保存することを含む、請求項３９記載のシステム。
【請求項４２】
前記二次索引がブロック圧縮索引付きファイルの索引である、請求項４１記載のシステム。
【請求項４３】
前記サーバは、前記第１のクエリーを複数の副クエリーに分割し、前記副クエリーのうちの少なくともいくつかを同時に処理するよう前記クエリーエンジンに指示するようにさらに構成されている、請求項１７記載のシステム。
【請求項４４】
前記第１のクエリー間隔が始まった後で前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項１７記載のシステム。
【請求項４５】
前記第１のクエリー間隔が始まる前に前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項１７記載のシステム。
【請求項４６】
１つ又は複数のデータ・ソースについて実行されるクエリーを管理するためのシステムであって、
少なくとも第１のクエリーを記憶する手段と、
処理のために前記第１のクエリーを選択する手段と、
第１のクエリー間隔に対して前記１つ又は複数のデータ・ソース内のデータの第１の部分についての前記第１のクエリーを処理するようにクエリーエンジンに指示する手段と、
前記データの第１の部分についての前記第１のクエリーの処理に基づいて前記クエリーエンジンから第１の結果データを受信する手段と、
前記第１のクエリー間隔後に前記第１のクエリーの状態を前記記憶媒体に保存する手段と、
前記第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するように前記クエリーエンジンに指示する手段と、
前記第１のクエリーに関連付けられた優先順位を変更する手段と、
前記第２のクエリーの処理に基づいて前記クエリーエンジンから第２の結果データを受

信する手段と、
前記第２のクエリーが中断され、前記クエリーエンジンによって処理されないように、前記第１のクエリーに関連付けられた前記優先順位に基づいて前記第２のクエリーを中断することを決定する手段と、
前記第２のクエリー間隔後の第３のクエリー間隔の間に前記１つ又は複数のデータ・ソース内の第２のデータ部分についての前記第１のクエリーを処理するように前記クエリーエンジンに指示する手段と
を含む、システム。
【請求項４７】
前記第１のクエリーに関係付けられた前記優先順位を前記記憶媒体に記憶する手段をさらに含み、
前記第１のクエリーに関係付けられた前記優先順位の前記変更は、処理のために前記第１のクエリーを選択する前に行われ、
処理のために前記第１のクエリーを選択することが部分的に前記優先順位に基づいて前記クエリーを選択することを含む、請求項４６記載のシステム。
【請求項４８】
前記第１のクエリー間隔が所定の時間量によって定義される、請求項４６記載のシステム。
【請求項４９】
前記第１のクエリーに関連付けられた前記優先順位は、前記１つ又は複数のデータ・ソース内の前記データのうちのどのくらいの量が、前記第１のクエリー間隔に対して前記第１のクエリーが実行される前記データの第１の部分に含まれるか、に影響する、請求項４８記載のシステム。
【請求項５０】
前記第１のクエリーを記憶することが、前記第１のクエリーを提供したリクエスタに通知される前に使用可能になるべき前記第１の結果データの量の通知しきい値を記憶することを含む、請求項４６記載のシステム。
【請求項５１】
前記第１の結果データの前記量が前記通知しきい値を超えた時に前記リクエスタに通知する手段をさらに含み、
前記第１のクエリーの前記状態を保存することが前記クエリーエンジンから受信した前記第１の結果データの前記量を記憶することを含む、請求項５０記載のシステム。
【請求項５２】
前記リクエスタからの要求に対して前記第１の結果データを返し、前記リクエスタに返された前記第１の結果データの前記量を前記記憶媒体に記憶する手段をさらに含む、請求項５１記載のシステム。
【請求項５３】
前記第１のクエリーを選択することが、前記クエリーエンジンから受信した前記第１の結果データの前記量と前記リクエスタに返された前記第１の結果データの前記量に基づくものである、請求項５２記載のシステム。
【請求項５４】
前記第１のクエリーの前記状態を保存することが、
前記第１のクエリーを中断するよう前記クエリーエンジンに指示することと、
前記第１のクエリーが中断された後に前記第１のクエリーの状態を保存すること、
を含む、請求項４６記載のシステム。
【請求項５５】
前記第２のデータ部分に関する前記第１のクエリーを処理するように前記クエリーエンジンに指示することが、
前記第１のクエリーの前記保存状態をロードすることと、
前記第１のクエリーを再開するように前記クエリーエンジンに指示すること、

を含む、請求項５４記載のシステム。
【請求項５６】
前記第１のクエリーの前記状態を保存することが、二次索引へのオフセットを保存することを含む、請求項５４記載のシステム。
【請求項５７】
前記二次索引がブロック圧縮索引付きファイルの索引である、請求項５６記載のシステム。
【請求項５８】
前記第１のクエリーを複数の副クエリーに分割し、前記副クエリーのうちの少なくともいくつかを同時に処理するよう前記クエリーエンジンに指示する手段をさらに含む、請求項４６記載のシステム。
【請求項５９】
前記第１のクエリー間隔が始まった後で前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項４６記載のシステム。
【請求項６０】
前記第１のクエリー間隔が始まる前に前記第２のクエリーが受信され、前記記憶媒体に記憶される、請求項４６記載のシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願の相互参照
本出願は、参照により本明細書に組み込まれる２００９年１２月２３日出願の米国特許出願第６１／２８９７７８号に対する優先権を主張するものである。
【０００２】
本明細書はクエリー（queries）の管理に関する。
【背景技術】
【０００３】
いくつかのデータ記憶システム（例えば、データベース）は、多数のクエリーの処理をサポートするような何らかの方法で、記憶された大量のデータを記憶する。例えば、いくつかのシステムは、並列記憶装置、並列クエリー処理エンジン、又はその両方の使用による並列処理能力を備える。
【発明の概要】
【課題を解決するための手段】
【０００４】
ある一つの態様では、一般に、１つ又は複数のデータ・ソースについて実行されるクエリーを管理するための方法は、少なくとも第１のクエリーを記憶媒体に記憶することと、処理のために第１のクエリーを選択することと、第１のクエリー間隔に対して１つ又は複数のデータ・ソース内のデータの第１の部分についての第１のクエリーを処理するようにクエリーエンジンに指示することと、データの第１の部分についての第１のクエリーの処理に基づいてクエリーエンジンから結果データを受信することと、第１のクエリー間隔後に第１のクエリーの状態を記憶媒体に保存することと、第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するようにクエリーエンジンに指示することと、第２のクエリー間隔後の第３のクエリー間隔の間に１つ又は複数のデータ・ソース内のデータの第２の部分についての第１のクエリーを処理するようにクエリーエンジンに指示することを含む。
【０００５】
諸態様は以下の特徴のうちの１つ又は複数を含むことができる。
【０００６】
この方法は、第１のクエリーに関連する優先順位を記憶媒体に記憶することと、処理のために第１のクエリーを選択する前に第１のクエリーに関連する優先順位を変更すること

をさらに含み、処理のために第１のクエリーを選択することが部分的に優先順位に基づいてクエリーを選択することを含む。

【０００７】

第１のクエリー間隔は所定の時間量によって定義される。

【０００８】

第１のクエリーの優先順位は、１つ又は複数のデータ・ソース内のデータのうちのどのくらいの量が、第１のクエリー間隔に対して第１のクエリーが実行されるデータの第１の部分に含まれるか、に影響する。

【０００９】

第１のクエリーを記憶することは、第１のクエリーを提供したリクエスタに通知される前に使用可能になるべき結果データの数量の通知しきい値を記憶することを含む。

【００１０】

この方法は、結果データの数量が通知しきい値を超えた時にリクエスタに通知することをさらに含み、第１のクエリーの状態を保存することがクエリーエンジンから受信した結果データの数量を記憶することを含む。

【００１１】

この方法は、リクエスタからの要求次第で結果データを返すことと、リクエスタに返された結果データの数量を記憶媒体に記憶することをさらに含む。

【００１２】

クエリーを選択することは、クエリーエンジンから受信した結果データの数量とリクエスタに返された結果データの数量に基づくものである。

【００１３】

第１のクエリーの状態を保存することは、第１のクエリーを中断するようクエリーエンジンに指示することと、第１のクエリーが中断された後に第１のクエリーの状態を保存することを含む。

【００１４】

第２のデータ部分に関する第１のクエリーを処理するようクエリーエンジンに指示することは、第１のクエリーの保存状態をロードすることと、第１のクエリーを再開するようクエリーエンジンに指示することを含む。

【００１５】

第１のクエリーの状態を保存することは、二次索引（secondary index）へのオフセットを保存することを含む。

【００１６】

二次索引はブロック圧縮索引付きファイル（block compressed indexed file）である。

【００１７】

この方法は、第１のクエリーを複数の副クエリーに分割することと、その副クエリーのうちの少なくともいくつかを同時に処理するようクエリーエンジンに指示することをさらに含む。

【００１８】

第１のクエリー間隔が始まった後で第２のクエリーが受信され、記憶媒体に記憶される。

【００１９】

第１のクエリー間隔が始まる前に第２のクエリーが受信され、記憶媒体に記憶される。

【００２０】

他の態様では、一般に、コンピュータ可読媒体は、１つ又は複数のデータ・ソースについて実行されるクエリーを管理するためのコンピュータ・プログラムを記憶する。このコンピュータ・プログラムは、少なくとも第１のクエリーを記憶媒体に記憶することと、処理のために第１のクエリーを選択することと、第１のクエリー間隔に対して１つ又は複数のデータ・ソース内のデータの第１の部分についての第１のクエリーを処理するようにク

エリーエンジンに指示することと、データの第１の部分についての第１のクエリーの処理に基づいてクエリーエンジンから結果データを受信することと、第１のクエリー間隔後に第１のクエリーの状態を記憶媒体に保存することと、第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するようにクエリーエンジンに指示することと、第２のクエリー間隔後の第３のクエリー間隔の間に１つ又は複数のデータ・ソース内のデータの第２の部分についての第１のクエリーを処理するようにクエリーエンジンに指示することをコンピュータに実行させるための命令を含む。

【００２１】

他の態様では、一般に、１つ又は複数のデータ・ソースについて実行されるクエリーを管理するためのシステムが提供される。このシステムは、少なくとも第１のクエリーを記憶する記憶媒体を含む。このシステムは、１つ又は複数のデータ・ソース内のデータに関するクエリーを処理するように構成されたクエリーエンジンを含む。また、このシステムは、処理のために第１のクエリーを選択し、第１のクエリー間隔に対して１つ又は複数のデータ・ソース内のデータの第１の部分についての第１のクエリーを処理するようにクエリーエンジンに指示し、第１のデータ部分についての第１のクエリーの処理に基づいてクエリーエンジンから結果データを受信し、第１のクエリー間隔後に第１のクエリーの状態を記憶媒体に保存し、第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するようにクエリーエンジンに指示し、第２のクエリー間隔後の第３のクエリー間隔の間に１つ又は複数のデータ・ソース内のデータの第２の部分についての第１のクエリーを処理するようクエリーエンジンに指示するように構成されたサーバも含む。

【００２２】

他の態様では、一般に、１つ又は複数のデータ・ソースについて実行されるクエリーを管理するためのシステムが提供される。このシステムは、少なくとも第１のクエリーを記憶する記憶媒体を含む。このシステムは、１つ又は複数のデータ・ソース内のデータについてのクエリーを処理するように構成されたクエリーエンジンを含む。このシステムは記憶媒体内のクエリーを管理するための手段を含み、その管理は、第１のクエリー間隔に対して１つ又は複数のデータ・ソース内のデータの第１の部分についての第１のクエリーを処理するようにクエリーエンジンに指示することと、データの第１の部分についての第１のクエリーの処理に基づいてクエリーエンジンから結果データを受信することと、第１のクエリー間隔後に第１のクエリーを記憶媒体に保存することと、第１のクエリー間隔後の第２のクエリー間隔の間に第２のクエリーを処理するようにクエリーエンジンに指示することと、第２のクエリー間隔後の第３のクエリー間隔の間に１つ又は複数のデータ・ソース内のデータの第２の部分についての第１のクエリーを処理するようクエリーエンジンに指示することを含む。

【発明の効果】

【００２３】

諸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。

【００２４】

部分的にクエリーに関連する優先順位に基づいてクエリーを選択すると、並列クエリー処理システムにおいて効率的な処理が可能になる。クエリーの各部分を部分的に処理し、次に中断することができる複数の間隔に時間をスライスすると、いくつかのクエリーをより速やかに処理することができ、特に優先順位の高いクエリーの場合、システム内の潜在的なバックログが低減される。

【００２５】

本発明のその他の特徴及び利点は、以下の説明並びに特許請求の範囲から明らかになるであろう。

【図面の簡単な説明】

【００２６】

【図１】クエリー処理を描写する概略図である。

【図２】クエリー処理を描写する概略図である。

【図３】データ記憶システムのブロック図である。
【図４】索引付き圧縮データ記憶装置の概略図である。
【図５Ａ】クエリーの処理に関連する時間間隔を示す図である。
【図５Ｂ】クエリーの処理に関連する時間間隔を示す図である。
【図６Ａ】クエリーの処理に関連する時間間隔を示す図である。
【図６Ｂ】クエリーの処理に関連する時間間隔を示す図である。
【図７Ａ】クエリーの処理に関連する時間間隔を示す図である。
【図７Ｂ】クエリーの処理に関連する時間間隔を示す図である。
【図７Ｃ】クエリーの処理に関連する時間間隔を示す図である。
【図７Ｄ】クエリーの処理に関連する時間間隔を示す図である。
【図８】スライスされたクエリー処理の概略図である。
【図９】索引付き圧縮データ記憶装置のクエリー処理を示す概略図である。
【図１０】クエリーを管理するためのプロセスのフローチャートである。
【図１１】クエリーを管理するためのプロセスのフローチャートである。
【発明を実施するための形態】
【００２７】
１　概要
図１を参照すると、いくつかの問題が分散クエリー管理において発生し得る。例えば、先入れ先出し法でデータ記憶システムのクエリーエンジンにクエリーが引き渡されると、システムはバックログになる可能性がある。いくつかのケースでは、引き渡されたクエリーは、あまりリソースを必要とせずに迅速に実行される短いクエリー１０２、１０４、１０８、１１２、１１８と、実行するのにより長い時間を必要とし、大量のシステム・リソースを使用する長いクエリー１１０、１１４、１１６と、短いクエリーと長いクエリーの間のどこかに入るクエリーとを含む可能性がある。特定のクエリーが実行される前にそのクエリーが要求するシステム・リソースの量をあらかじめ決めることは実用的ではない可能性がある。図１は、複数のクエリーエンジンを使用してクエリーを処理するためのシステムの一例を示している。クエリーは、非同期的に受信されて、待ち行列１０１に記憶され、データ記憶システムのクエリーサーバ１００上で実行されるクエリーエンジンによって処理される機会を待つ。この例では、初めに、長いクエリー１１６が処理のために第１のクエリーエンジン１２０に割り当てられ、短いクエリー１１８が処理のために第２のクエリーエンジン１２２に割り当てられる。図２を参照すると、短い時間の後に短いクエリー１１８は完了した可能性があり、次に並んでいるクエリーである長いクエリー１１４が空いているクエリーエンジン１２２に割り当てられる。この時点で残りのクエリー１０２、１０４、１０８、１１０、１１２は、長いクエリー１１６、１１４のうちの一方が処理を完了し、クエリーエンジン内の処理リソースを解放するまで待つ。この現象は短いクエリーの待ち時間を増やすものであり、迅速応答が期待されているクエリーにおいて受け入れがたい遅延を引き起こす可能性がある。
【００２８】
図３を参照すると、データ記憶システム３００は、クエリーを実行するための要求を受信するために、フロントエンド・サービス３０２、例えば、ウェブ・サービスを提供するように構成される。仲介サーバ３０４は複数のクエリーエンジン３１２によるクエリー実行をスケジュールする。各クエリーは割り当てられた期間に対して実行が許可され、その期間は、例えば、時間（例えば、ＣＰＵクロックによって測定されたもの）、持続時間、処理された行数、又は検索された行数によって測定され得る。クエリーエンジン３１２は、１つ又は複数のデータ・ソース３１０Ａ、３１０Ｂ、３１０Ｃからのデータにアクセスし、処理セット３１４を生成するためにクエリーを処理する。データ・ソースを提供する記憶装置は、例えば、仲介サーバ３０４を実現するコンピュータに接続された記憶媒体上に記憶されていて、システム３００にとってローカルなもの（例えば、ハードディスク・ドライブ）である場合もあれば、リモート接続により通信接続しているリモート・システム上でホストとして処理され、仲介サーバ３０４にとってリモートなもの（例えば、メイ

ンフレーム）である場合もある。

【００２９】

仲介サーバ３０４は結果セット３１４を管理する。仲介サーバ３０４は、クエリーに関する追加情報、例えば、クエリーの優先順位、要求された行数、クエリーから返された行数、リクエスタに返された行数、クエリーがどのように使用されるかを示す表示要素（indication）、一度に必要な行数、クエリーがクエリーエンジンによって最後に実行された時間、及び状況標識を記憶することができる。状況標識は、クエリーが待っていること、実行していること、中断されていること、割り込まれていること、又は完了したことを示すことができる。いくつかの編成（arrangements）では、クエリー状態は、クエリー処理中に発生したエラー条件の存在も示すことができる。待ち状態にあるクエリーは、現在実行されていないが、実行する資格のあるものである。実行状態にあるクエリーは、クエリーエンジンによって現在処理されている。中断状態にあるクエリーは、クライアントによって現在要求されている行数を仲介サーバがすでに返したので、実行する資格のないものである。割り込み状態にあるクエリーは、優先順位がより高いクエリーによって先取りされたので、実行する資格のないものである。完了状態にあるクエリーは実行を完了したものである。いくつかの編成では、追加の状況がサポートされる。

【００３０】

仲介サーバ３０４は、クエリー結果が処理の準備ができていることをいつ、どのようにリクエスタに通知しなければならないかを同定する情報も記憶し得る。いくつかの編成では、複数のクエリーエンジンが単一クエリーの種々の部分について独立して動作し得る。各クエリーエンジンは仲介データベースを独立に更新する。トリガ・イベントが発生した時、例えば、結合されたクエリーエンジンが要求された行数を返した時に、通知イベントがトリガされる。

【００３１】

いくつかの実装例では、仲介サーバは仲介論理モジュール３０６と仲介データベース３０８とを含む。いくつかの実装例では、仲介論理モジュール３０６は、個々のクエリーエンジン３１２、フロントエンド・サービス３０２に組み込まれる場合もあれば、複数のサービスに分割される場合もある。クエリーエンジン３１２は、結果セット３１４で使用可能な行数について仲介データベース３０８を更新し得る。

【００３２】

いくつかの実装例では、図４を参照すると、データ・ソース３１０は、索引付き圧縮データ記憶装置４０２を含む。索引付き圧縮データ記憶装置は、例えば、ファイル内に記憶された複数の圧縮データ・ブロック４０４を含む。それぞれの圧縮データ・ブロックは、その圧縮データ・ブロック内のデータの位置特定を可能にする、少なくとも１つの索引４０６に関連付けられている。いくつかの実装例では、第１のキー（例えば、一次キー）に基づいてサーチ可能な一次索引が提供され、その他のキー（例えば、外部キー）に基づいてサーチ可能な１つ又は複数の二次索引が提供される。索引のうちのいくつかは、それぞれのキー値が固有のものであるサロゲート・キーで構成することができ、その他の索引は、キーの値がデータ・セット内で固有のものではない可能性がある自然キーに基づくものである。いくつかの実装例では、自然索引を結合して、単一連結索引を作成することができる。索引付き圧縮データ記憶技法及びシステムについては、参照により本明細書に組み込まれる米国特許出願公報第２００８／０１０４１４９Ａ１号により詳細に記載されている。

【００３３】

２　クエリースライシング

図５Ａを参照すると、一連のクエリーＡ５０２、Ｂ５０４、Ｃ５０６、及びＤ５０８が、異なるクエリーに関係付けられた時間間隔を示す図に示されている。クエリーが引き渡された順序で実行される場合、クエリーＡは間隔５０２の間に実行されて完了し、次にクエリーＢは間隔５０４の間に実行されて完了し、次に間隔５０６の間のクエリーＣと間隔５０８の間のクエリーＤが続く。これらの条件下では、クエリーＡは時間５１０で完了す

るまで結果を返さず、クエリーＢは時間５１２で完了するまで結果を返さず、クエリーＣは時間５１４で完了するまで結果を返さず、クエリーＤは時間５１６で完了するまで結果を返さないであろう。クエリーＤは短いクエリーであるが、たまたま他の長いクエリーの後ろに位置していたので結果を返すのに長い時間を要する。

【００３４】

仲介サーバ３０４のいくつかの実装例では、クエリーを完了まで必然的に順次実行する代わりに、仲介サーバは１つのクエリーを複数の異なる小さい部分に分割する。クエリーエンジン３０４は、特定の間隔の間にクエリーを実行するよう指示される。この間隔は、期間、返すべき行数、処理された行数、又はその他の何らかの基準に基づいて定義することができる。この手法を使用して、図５Ｂを参照すると、クエリーＡは間隔５２８の間に実行され、クエリーＢは間隔５３０の間に実行され、クエリーＣは間隔５３２の間に実行され、クエリーＤは間隔５３４の間に実行され（完了する）、次にクエリーＡは第２の間隔の間にもう一度実行される。いくつかのケースでは、１つのクエリーが処理されるそれぞれの間隔後にそのクエリーをサブミットしたプロセスに対し、そのクエリーによるいくつかの結果が返される可能性がある。例えば、クエリーＡによるいくつかの結果は時間５２０後に返され、クエリーＢ、Ｃ、及びＤによるいくつかの結果は、それぞれ時間５２２、５２４、５２６後に返される可能性がある。これらのクエリーを小さい実行間隔に分割することにより、システム３００は、他のクエリーが実行される前にクエリーが完了するのを待たなければならない場合より速やかにより多くのクエリーに関するいくつかの結果を生成することができる。さらに、他のクエリーを遅延させるというトレードオフにより、いくつかのクエリーは、本来完了したと思われる時期より速やかに完了することができる。この例では、クエリーＤは時間５２６で完了し、クエリーＣは時間５４０で完了し、クエリーＡは時間５４２で完了し、クエリーＢは時間５４４で完了する。従って、この例では、長いクエリーＡ及びＢを遅延させるという犠牲を払って、短いクエリーＣ及びＤの方がより速やかに完了する。

【００３５】

クエリーの分割方法の決定は、システムに望ましい動作特性に依存しうる。例えば、時間に基づいてクエリーを分割することは、それぞれのクエリーが特定の量の作業を実行できることを保証する可能性があるが、その作業がどのくらいの長さの暦時間を費やせるかという保証はなく、１つの実行間隔でどのくらいの行数が返されるかについても保証はない。対照的に、いくつかの行が返されるまでクエリーを実行できるようにすることにより、いくつかの結果を生成するためにどのくらいの数の実行間隔が必要になるかが決定されるが、１つの間隔がどのくらい長く持続するかについての保証はない。いくつかの行が処理されるまでクエリーを実行できるようにすることは、システムに、クエリーを完了するのにどのくらいの数の実行間隔が必要になるかを同定することを可能にするが、特定の数の行を返すのにどのくらいのサイクル数が必要であるか又は具体的に特定の実行サイクルがどのくらいの長さの時間を要するかを知らせることはない。

【００３６】

クエリーを処理するための時間は、単一クエリーのみが処理されている場合でも複数の実行間隔（又は「クエリー間隔」）に分割することができる。あるクエリー間隔の終わりに、新しいクエリーが到着している場合、処理されているクエリーは中断され、次のクエリー間隔を使用して新しいクエリーを処理する。代わって、そのクエリー間隔の終わりに新しいクエリーが到着していない場合、処理されているクエリーは追加のクエリー時間の間に処理を続行することができる。例えば、図６Ａの例では、クエリーＢはクエリーＡの処理中の時間６１０に到着し、図６Ｂの例では、クエリーＡ又はクエリーＢのいずれかの処理が始まる前に両方のクエリーＡ及びＢが到着する。

【００３７】

図６Ａの例では、クエリーＡは間隔６０２の間に実行され、クエリーＡが間隔６０２の終わりに完了していない場合、システムは、追加のクエリー間隔の間にクエリーＡを処理しなければならないかどうか又は他のクエリーが処理を待っているかどうかを判断するた

めにチェックする。クエリーＢは間隔６０２の終わりにまだ到着していないので、クエリーＡはクエリー間隔６０４の間に処理される。同様に、クエリーＡは次のクエリー間隔６０６の間にも処理される。しかし、クエリー間隔６０６の終わりに、システムは、時間６１０に到着したクエリーＢを間隔６０８の間に処理しなければならないと判断する。次に、それぞれが完了するまで（この例では、クエリーＡは時間６１２に完了し、クエリーＢは時間６１４に完了する）交互の間隔においてクエリーＡ及びＢが処理される。図６Ｂの例では、クエリーＡは間隔６２０の間に実行され、クエリーＡが間隔６２０の終わりに完了していない場合、システムは、追加のクエリー間隔の間にクエリーＡを処理しなければならないかどうか又は他のクエリーが処理を待っているかどうかを判断するためにチェックする。クエリーＢは間隔６２０の終わり以前にすでに到着しているので、クエリーＢはクエリー間隔６２２の間に処理される。次に、それぞれが完了するまで交互の間隔においてクエリーＡ及びＢが処理される。

【００３８】

クエリー間隔の終わりにクエリーを中断することは、クエリーの状態を仲介データベースに保存することを含む。ある編成では、１つの間隔後に仲介データベース内でクエリー状態を「中断」又は他の状態に更新し、そのクエリーが実行する資格のないものであることを示すことができる。所定の間隔後に、クエリーの状況を「待ち」に更新し、そのクエリーをもう一度実行できるようにすることができる。他の編成では、仲介サーバは自動的に所定の間隔後に直ちにクエリーをスケジュールする。

【００３９】

３　クエリーの優先順位付け及び再優先順位付け

仲介データベースは個々のクエリーに関係付けられた優先順位を記憶し得る。この優先順位は、クエリーが実行される頻度及び方法に影響し得る。図７Ａを参照すると、優先順位の高いクエリーＡには、クエリーＢ（間隔７０４の間に処理される）又は優先順位が低いクエリーＣ（間隔７０６の間に処理される）より大きい実行間隔７０２が提供され得る。この例では、優先順位の高いクエリーＡにはクエリーＢに提供される実行間隔７０４より大きい実行間隔７０２が提供され、クエリーＢには優先順位の低いクエリーＣに提供される実行間隔７０６より大きい実行間隔７０４が提供される。或いは、図７Ｂを参照すると、優先順位の高いクエリーＡには標準的な優先順位のクエリーＢ（間隔７１０の間に処理される）より高い頻度の実行間隔７０８が提供され、標準的な優先順位のクエリーＢには優先順位の低いクエリーＣ（間隔７１２の間に処理される）より高い頻度の実行間隔が提供され得る。図７Ｃを参照すると、ある状況では、クエリーＡには、クエリーＡが（間隔７１４後に）実行を完了するまで他のクエリーＢ及びＣの処理が中断されるように十分高い優先順位が提供され、実行を完了した時点で、それぞれ間隔７１６と７１８の間で交互に、中断されたクエリーＢ及びＣの実行が再開される。

【００４０】

また、仲介データベースは、クエリーが実行している間、クエリーに再優先順位付けを可能にする。例えば、図７Ｄを参照すると、優先順位の高いクエリーＡは（間隔７２０の間）、通常の優先順位のクエリーＢ（間隔７２２の間）及び優先順位の低いクエリーＣ（間隔７２４の間）とともに、仲介データベースによってスケジュールされる。時間７２６で、優先順位の高いクエリーＡは通常の優先順位レベルに再優先順位付けされる。その時点で、仲介データベースは、新しい優先順位付けに基づいてクエリーのスケジューリングを調整する。再優先順位付け後に前進すると、次に、通常の優先順位のクエリーＡには通常の優先順位のクエリーＢに提供される間隔７２２と同様のサイズの実行間隔７２８が提供される。

【００４１】

再優先順位付けは、要求しているプロセスによってなされた判断により生じる場合もあれば、それ自体の基準に基づいて仲介サーバ内で生じる場合もある。例えば、仲介サーバにはクエリーを完了するための期限が設けられる可能性があり、期限が近づくにつれて、サーバは適時完了を保証するためにクエリーの優先順位を高め得る。いくつかのケースで

は、仲介サーバが優先順位のより高いトラフィックをチェックできるようにするために、単一のより大きい実行間隔の代わりに、複数のより小さい実行間隔がクエリーに提供され得る。その他のケースでは、仲介サーバは、優先順位のより高いクエリーが実行できるようにするために、実行中のクエリーの実行間隔に割り込むことができる。

【００４２】

いくつかのケースでは、クエリーは、前のクエリーの実行前又は実行中のいずれかの実行に先立って次の間隔に入る次のクエリーとともに、スケジューリングされ得る。いくつかのケースでは、実行のためにスケジュールされるべき次のクエリーは、選択基準に基づいて実行直前に選択され得る。

【００４３】

４　並列クエリー処理

多くのシステムにとって、複数のクエリーを一度に実行することが有利であり得る。例えば、単一システム上で実行されている２つのクエリーは、１つのシステム上で実行される単一クエリーよりも改善されたパフォーマンスを実現（experience）しうる。これは、例えば、第２のクエリーが異なるリソースを使用している間に一方のクエリーが１つのコンピューティング・リソースを使用できるために生じ得る。一度に両方のクエリーを実行することによって、スループットは改善される。いくつかの実装例では、図８を参照すると、優先順位の高いクエリー８０２は、複数のクエリースライス８０４、８０６、８０８、８１０、８１２に分割される。各スライスは、個別のクエリーエンジン８１４、８１６、８１８、８２０、８２２によって処理され得る。

【００４４】

優先順位の高いクエリー８０２は、上記のように処理すべき行数に基づいてスライスされ得る。クエリーを完了するためにどのくらいの数の実行間隔が必要になるかを判断するために、パーティション化情報は、クエリーのターゲットである索引付き圧縮データ記憶装置の二次索引と比較され得る。これは、それぞれのクエリースライスによって索引付き圧縮データ記憶装置のどの部分が処理されるかを同定することにもなる。例えば、図９を参照すると、索引付き圧縮ファイル９０２は複数のデータ・ブロック９０４、９０６、９０８、９１０を含み、各データ・ブロックは複数のデータ・レコードを含む。索引付き圧縮ファイル９０２は、データ・ブロックを参照する索引９１２に関係付けられる。いくつかの編成では、この索引は、それぞれのデータ・ブロックに関する１つの索引レコード９２２を含む可能性があり、他の編成では、索引９１２はデータ・ブロックより少ない索引レコード９２２を含み得る。いくつかの編成では、各索引レコード９２２はデータ・ブロック９０４、９０６、９０８、９１０を参照し、他の編成では、各索引レコード９２２はデータ・ブロックの第１のデータ・レコードを参照する。仲介サーバは、索引９１２を検討し、索引レコードに基づいてクエリー実行間隔（又は「クエリースライス」）を決定する。この例では、クエリーエンジンは、索引９１２に基づいて４つのクエリースライス９１４、９１６、９１８、９２０を作成することを選択する。１つのクエリースライス９１４は、ブロック１：９０４から始まるデータ・レコードを処理し、ブロック１０（図示せず）の終わりで終了し、クエリースライス９１６は、ブロック１１：９０６から始まるデータを処理し、ブロック２０（図示せず）の終わりで終了し、クエリースライス９１８は、ブロック２１：９０８から処理を開始し、ブロック３０（図示せず）の終わりで終了し、最後にクエリースライス９２０は、ブロック３１：９１０から処理を開始し、索引付き圧縮ファイル９０２の終わりで処理を終了する。この例では、仲介サーバは、索引９１２内の索引レコード９２２の数によってのみ制限される、任意の数のクエリースライスを作成することを選ぶことができる。

【００４５】

図８を参照すると、クエリーの各スライスは、クエリーエンジン８１４、８１６、８１８、８２０、８２２のそれぞれ異なる１つによって同時に処理され得る。例えば、クエリースライス８０４はクエリーエンジン８１４によって処理され、クエリースライス８０６は実質的に同時にクエリーエンジン８１６によって処理される。同時に、クエリースライ

ス８０８はクエリーエンジン８１８によって処理され、クエリースライス８１０はクエリーエンジン８２０によって処理され、クエリー８１２はクエリーエンジン８２２によって処理される。各クエリーエンジンは、そのクエリーパーティションに関する結果セットを生成する。すべての結果セットが生成されると、結果セットは、クエリー全体に対する完全な結果セットを形成するように、結合されうる。この方法を使用すると、優先順位の高いクエリーは、通常、その動作を完了するのに要する時間の一部で完了することができる。

【００４６】

５　コールバック

あらかじめ指定された基準によって定義されたトリガが満たされると、システムは通知を行う。図３を参照すると、新しいクエリーがフロントエンド・サービス３０２にサブミットされる場合、このサブミットは、条件が満たされた時に（フロントエンド・サービス３０２を介して）リクエスタに通知するよう仲介サーバ３０４に要求する情報を含み得る。ある編成では、この条件は、特定の数の結果データ要素がリクエスタによってアクセスできる状態になった時の通知であり得る。例えば、リクエスタは、１００個の結果レコードの準備ができた時に通知され得る。いくつかケースでは、リクエスタは、通知の前に準備ができなければならない結果データ要素の数を指定し得る。他のケースでは、リクエスタは、リクエスタが通知を受ける前に満たさなければならない他の基準を提供し得る。例えば、リクエスタは、クエリーが中断された時又はすべての処理が完了した時に通知を受けたいと希望する可能性がある。いくつかのケースでは、トリガ基準は、仲介データベース３０８内で追跡された状態情報に制限されるかも知れない。他のケースでは、トリガ基準は制限がないかも知れない。トリガは、いくつかの異なる方法で仲介サーバ３０４に提供され得る。例えば、トリガは、それぞれのクエリー間隔後に仲介サーバ３０４が実行するスクリプトとして、又は所定のアプリケーション・プログラミング・インターフェース（ＡＰＩ）に適合するコンパイル済みクラスとして、提供され得る。いくつかのケースでは、例えば、１００個の結果レコードが発見されたという条件など、その条件が一度しか発生しない可能性がある。他の編成では、例えば、１００個の追加の結果レコードが発見されるごとに通知を求める要求など、その条件が再発する可能性もある。

【００４７】

いくつかのケースでは、トリガ条件のサブミットはアクション定義も含み得る。このアクションは、トリガとともに仲介データベース３０８に記憶され得る。アクションは、条件が満たされた時に仲介サーバ３０４がどのように応答するかを定義するものである。アクションは、例えば、通知、要約など、所定の１組の可能なアクションのうちの１つにすることができる。アクションは、仲介サーバ３０４上で実行されるスクリプトにすることができる。例えば、１つのアクションは、返された結果をクエリーパラメータとして使用して、追加のクエリーをシステムにサブミットすることができる。また、アクションは、事前確立されたＡＰＩに適合するコンパイル済みクラスとして提供することができる。

【００４８】

６　クエリーの中断

いくつかの実装例では、仲介サーバ３０４はクエリーの処理を中断することができる。仲介サーバは、そのクエリーに中断というマークを付けることができ、そのクエリーが再開されるまでいかなる処理も行われない。クエリーの中断とともに、クエリーサーバはクエリーの状態を保存し得る。この状態はクエリーのプロセスの表示要素である。例えば、この状態は索引付き圧縮データ・ストアへのオフセットになる場合もあれば、Ｂ木内で最後に処理されたノードを含む場合もある。

【００４９】

いくつかのケースでは、仲介サーバはそれ自体のイニシアチブについてのクエリーを中断することを選択しうる。これは、例えば、１つのクエリーがいくつかのレコードを結果セット内に生成し、リクエスタに引き渡されるのを待っている結果セット内の行数がしきい値を超えた時に生じ得る。

【００５０】
例えば、クエリーをサブミットするリクエスタは、固定数の行の引き渡しを後で要求し得る（例えば、ユーザ・インターフェースが画面を埋め尽くす（populate）ために２５行分のデータの「ページ」を要求するならば、システムはクエリーから２５行分のデータを要求する。）。その後、ユーザがより多くのクエリー結果を見ることを希望していることを示した場合、システムはクエリー結果の次の「ページ」又は２６～５０の結果を要求することができる。仲介データベースは、クエリーから返された結果の数と、ユーザに返された結果の数を追跡する。例えば、クエリーは３００行を返した可能性があるが、２５行がリクエスタに送信された可能性がある。クエリーから返された行数がマージン（例えば、２５、５０、７５、又は１００行）によってリクエスタに送信された行数を超える場合、仲介データベースはそのクエリーの処理を中断し得る。これは、仲介データベース内でそのクエリーに中断というマークを付けることによるか、又はそのクエリーの次の実行をスケジュールする前のチェックを介して、成し遂げられる。

【００５１】
いくつかのケースでは、しきい値はシステム３００によって定義され、他のケースでは、そのクエリーがどのように使用されるかに依存して各クエリーについて個別にしきい値が定義され得る。例えば、固定数の項目を有するＷｅｂページ上にデータ・リストを表示するためにその結果が使用されるクエリーは、４ページ分のデータが待っている時に、中断し得る。対照的に、そのクエリーによって返されるすべてのデータの要約レポート、例えば、月末レポートを作成するためにその結果が使用されるクエリーは、けっして中断し得ない。いくつかのケースでは、しきい値は、リクエスタに通知する前に収集するための行数から推測され得る。

【００５２】
いくつかのケースでは、クエリーは、仲介データベース内のそのクエリーの状態情報を更新することによって、明示的に中断され得る。例えば、優先順位のより高いクエリーを実行できるようにするために、クエリーに中断というマークが付され得る。他のケースでは、仲介サーバのスケジューリング・アルゴリズムは、中断されたクエリーの状態を有するクエリーがスケジュールされないようになっているので、クエリーは、暗黙的に中断され得る。クエリーの中断は、クエリーがサブミットされ、その後、そのクエリーが完了する前に呼び出しプログラムが終了する時にリソースの浪費を最小限にするという追加の利点を有する。仲介サーバは、リクエスタが定義済み期間の間に結果にアクセスしなかった場合にクエリー及び結果セットを削除することを選ぶことができる。

【００５３】
７　仲介サーバの処理
図１０を参照すると、フローチャート１０００は、外部要求なしにクエリーの処理を中断すべきかどうかに関する判断を含む、仲介サーバ３０４の動作の模範的な編成を表している。

【００５４】
動作は、実行のためにクエリーを選択すること１００２を含む。一例では、クエリーは、仲介サーバによって確立された所定のスケジュールの一部として選択され得る。他の例では、クエリーは、クエリーの優先順位及びクエリーが最後に実行された時間を含み得る何らかの基準に基づいて選択され得る。ある編成では、仲介サーバは実行を待っている（例えば、待ち状態にある）クエリーについて繰り返す。クエリーのそれぞれはクエリーエンジン上で実行するようにスケジュールされる。待っているすべてのクエリーが実行されると、仲介サーバは、依然として実行を待っているクエリーについてプロセスを繰り返す。他の編成では、仲介サーバは、実行のために最も長い間隔の間、待っていたクエリーを選択する。他の編成では、仲介サーバは、実行のために優先順位が最も高いクエリーを選択する。

【００５５】
動作はまた、クエリーエンジン上でクエリーを実行すること１００４を含む。一例では

、選択されたクエリーはクエリーエンジンに割り当てることができ、そのクエリーエンジンはデータ・レコードに対してクエリーを実行し、結果セットを更新し、返された行数を仲介サーバに通知する。

【００５６】

動作はまた、引き渡されるのを待っている行数のチェック１００６を含む。引き渡されるのを待っている行数が通知しきい値を超える場合、仲介サーバはリクエスタへのコールバック１００８を実行する。

【００５７】

動作はまた、リクエスタがアクセスするのを待っている行数が中断しきい値を超えるかどうかのチェック１０１０を含み、その場合、１０１２でクエリーが中断される。クエリーが中断されるかどうかにかかわらず、仲介サーバは次に処理するクエリーの選択に移行する。

【００５８】

図１１を参照すると、フローチャート１１００は、例えば、結果がアクセスできる状態になっていることをコールバックがリクエスタに通知した後に、クエリーによって返された結果セットの一部にリクエスタがアクセスしたことに応じて、仲介サーバ３０４の動作の模範的な編成を表している。

【００５９】

動作は、リクエスタがクエリーからの結果を要求すること１１０２を含む。いくつかの編成では、リクエスタは、返すべき行数の指示、例えば、２５行を返すよう求める要求を送信し得る。他の編成では、リクエスタは特定の範囲の結果を返すよう要求し得る。例えば、リクエスタは、５０～１２６の結果を返すよう要求し得る。さらに他の編成では、リクエスタは収集したすべての結果を返すことを要求し得る。

【００６０】

動作はまた、結果を返し、レコードを更新すること１１０４を含む。要求に応答して、仲介サーバは要求された行へのアクセスを提供し得る。いくつかの編成では、仲介サーバはまた、そのクエリーが依然として追加の結果を処理しているという表示要素をリクエスタに送信し得る。他の編成では、仲介サーバは、追加の結果が即時引き渡しに使用可能であるという表示要素も提供し得る。

【００６１】

動作はまた、そのクエリーが現在中断されているかどうかを判断するためのチェック１１０６を含む。クエリーが中断されている場合、制御は次の動作１１０８に移行する。そうではない場合、プロセスは完了する。

【００６２】

動作はまた、引き渡しを待っている行数が中断しきい値未満であるかどうかを判断するためのチェック１１０８を含む。そうである場合、クエリーは１１１０で再開され、仲介サーバによる処理のためにスケジュールされ得る。

【００６３】

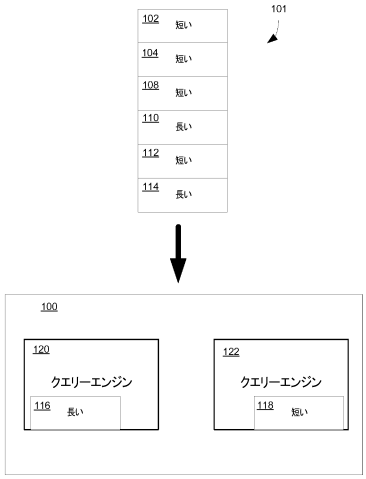
上記のクエリー管理手法は、コンピュータ上で実行するためにソフトウェアを使用して実現することができる。例えば、このソフトウェアは、１つ又は複数のプログラム式又はプログラム可能コンピュータ・システム（分散、クライアント／サーバ、又はグリッドなどの様々なアーキテクチャのものにすることができる）上で実行される１つ又は複数のコンピュータ・プログラム内の手順を形成し、それぞれのコンピュータ・システムは少なくとも１つのプロセッサと、少なくとも１つのデータ記憶システム（揮発性及び不揮発性メモリ及び／又は記憶素子を含む）、少なくとも１つの入力装置又はポート、並びに少なくとも１つの出力装置又はポートを含む。このソフトウェアは、例えば、計算グラフの設計及び構成に関連するその他のサービスを提供する、より大きいプログラムの１つ又は複数のモジュールを形成することができる。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・モデルに適合するその他の組織化されたデータとして実現することができる。

【００６４】
このソフトウェアは、汎用又は特殊目的プログラム可能コンピュータによって読み取り可能なＣＤ－ＲＯＭなどの記憶媒体上で提供するか、或いはそれが実行されるコンピュータへのネットワークの通信媒体により配布する（伝搬信号にコード化する）ことができる。すべての機能は、特殊目的コンピュータ上で又はコプロセッサなどの特殊目的ハードウェアを使用して実行することができる。このソフトウェアは、ソフトウェアによって指定された計算の異なる部分が異なるコンピュータによって実行されるという分散方法で実現することができる。それぞれのこのようなコンピュータ・プログラムは、好ましくは、汎用又は特殊目的プログラム可能コンピュータによって読み取り可能なストレージ・メディア又はデバイス（例えば、ソリッドステート・メモリ又はメディア或いは磁気又は光メディア）上に記憶されるか又はそれにダウンロードされ、本明細書に記載された手順を実行するためにそのストレージ・メディア又はデバイスがコンピュータ・システムによって読み取られた時にそのコンピュータを構成し操作する。また、本発明のシステムは、コンピュータ・プログラムとともに構成されたコンピュータ可読記憶媒体として実現されるものと見なすことができ、このように構成された記憶媒体は本明細書に記載された機能を実行するためにコンピュータ・システムを具体的かつ定義済みの方法で動作させる。

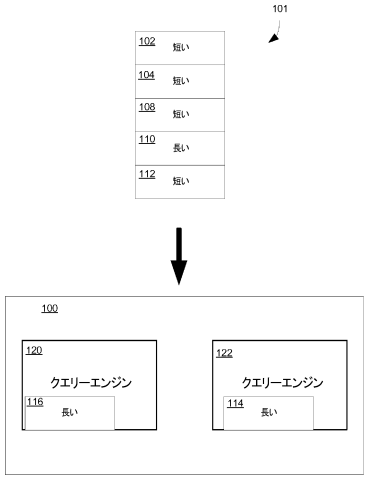
【００６５】
本発明のいくつかの実施形態について説明してきた。それにもかかわらず、本発明の精神及び範囲を逸脱せずに様々な変更が可能であることが理解されるであろう。例えば、上記の諸ステップのうちのいくつかは順序とは無関係なものにすることができ、従って、上記のものとは異なる順序で実行することができる。

【００６６】
上記の説明は本発明を例示するためのものであって、本発明の範囲を限定するためのものではなく、本発明は特許請求の範囲の範囲によって定義されることを理解されたい。例えば、上記の機能ステップのうちのいくつかは、処理全体に実質的に影響せずに異なる順序で実行することができる。その他の諸実施形態は特許請求の範囲の範囲内である。

【図１】



【図２】



【図３】

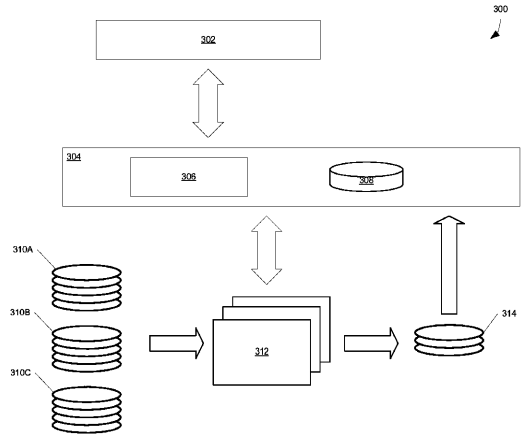


FIG. 3

【図４】

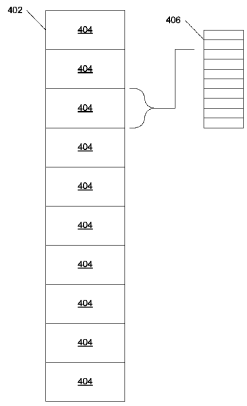


FIG. 4

【図５Ａ】

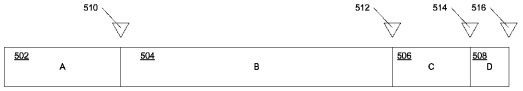


FIG. 5A

【図５Ｂ】

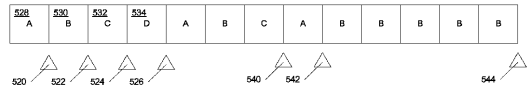


FIG. 5B

【図６Ａ】

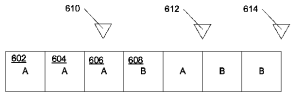


FIG. 6A

【図６Ｂ】

FIG. 6B

【図７Ａ】

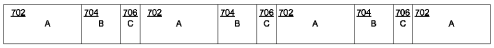


FIG. 7A

【図７Ｂ】

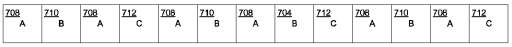


FIG. 7B

【図７Ｃ】

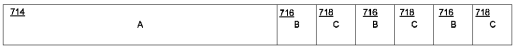


FIG. 7C

【図７Ｄ】

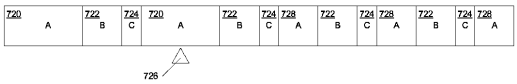


FIG. 7D

【図８】

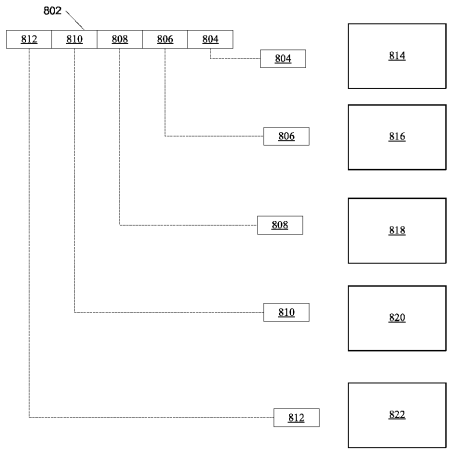
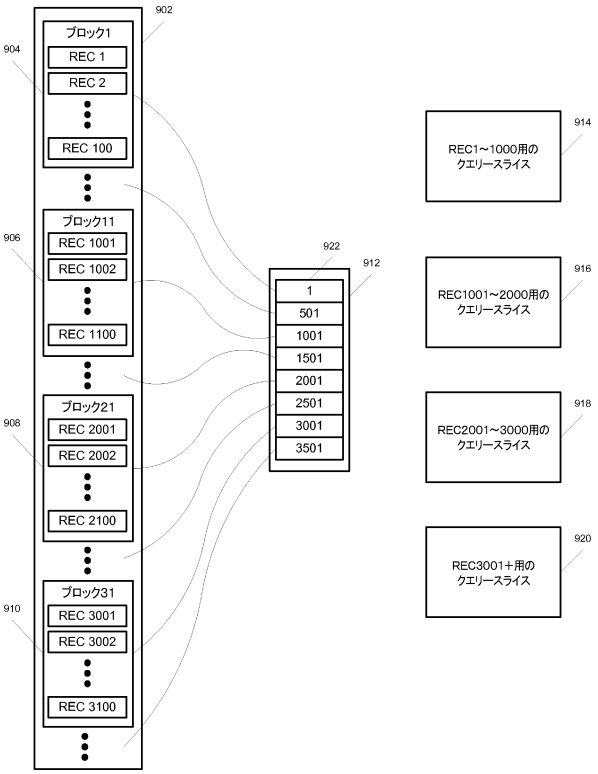
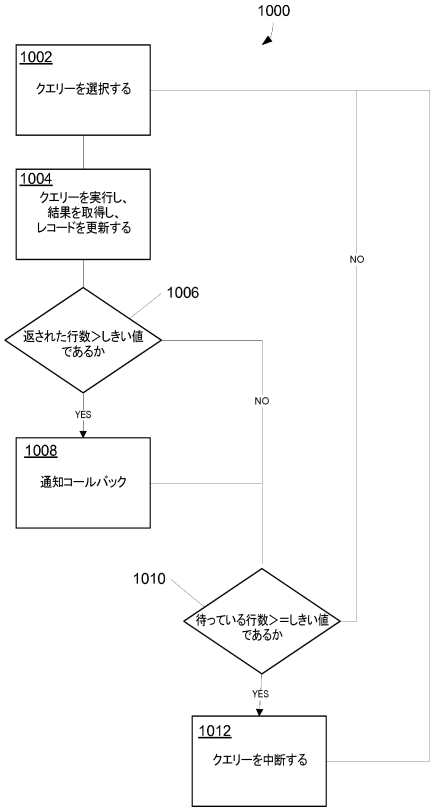


FIG. 8

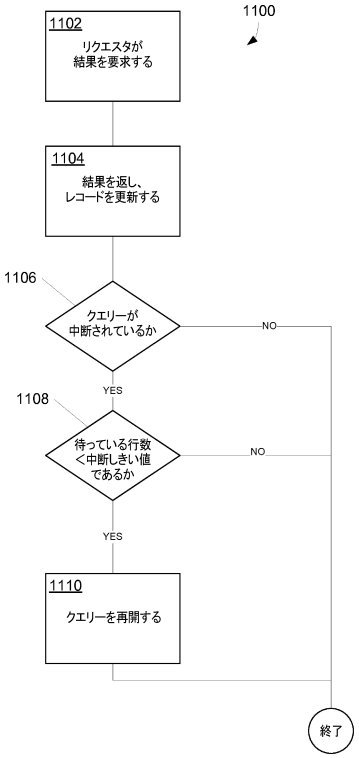
【図９】



【図１０】


1000
1002
クエリーを選択する
1004
クエリーを実行し、
結果を取得し、
レコードを更新する
1006
返された行数>しきい値
であるか
YES
NO
1008
通知コールバック
1010
待っている行数>=しきい値
であるか
YES
NO
1012
クエリーを中断する

【図１１】


1100
1102
リクエスタが
結果を要求する
1104
結果を返し、
レコードを更新する
1106
クエリーが
中断されているか
NO
YES
1108
待っている行数
<中断しきい値
であるか
NO
YES
1110
クエリーを再開する
終了

フロントページの続き

(72)発明者　マクリーン，ジョン
イギリス国，スコットランド　ピーエー５　９ビージェイ，エルダースライ，エルダー　グローブ，ニューランドクレイグス　アベニュー　３４

審査官　早川　学

(56)参考文献　米国特許出願公開第２００８／００３３９２０（ＵＳ，Ａ１）
特開平０８－３１４９６５（ＪＰ，Ａ）
特開２００６－２６０５１１（ＪＰ，Ａ）
米国特許出願公開第２００８／０１０４１４９（ＵＳ，Ａ１）
特開平８－２９２９５６（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１７／３０
Ｇ０６Ｆ　１２／００