

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-198184
(P2010-198184A)

(43) 公開日 平成22年9月9日(2010.9.9)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 9/48 (2006.01)	G06F 9/46 452G	
	G06F 9/46 452B	
	G06F 9/46 452J	

審査請求 未請求 請求項の数 18 O L (全 12 頁)

(21) 出願番号	特願2009-40739 (P2009-40739)	(71) 出願人	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(22) 出願日	平成21年2月24日 (2009.2.24)	(74) 代理人	100065385 弁理士 山下 穰平
		(74) 代理人	100130029 弁理士 永井 道雄
		(72) 発明者	中川 和俊 東京都港区芝五丁目7番1号 日本電気株式会社社内

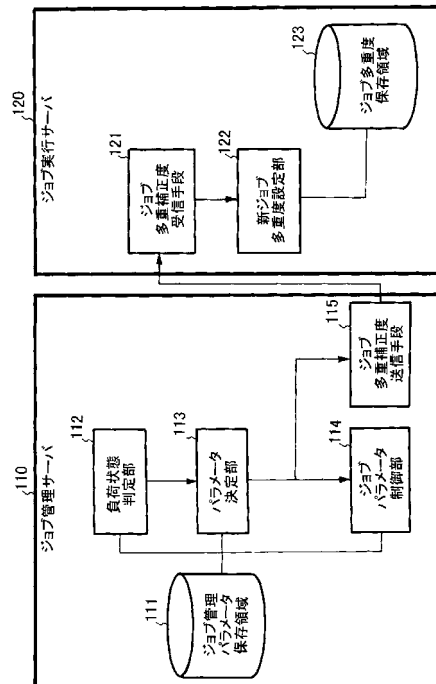
(54) 【発明の名称】 ジョブ管理システム、その方法及びそのプログラム

(57) 【要約】

【課題】ハードウェア資源の有効活用が可能なジョブ管理。

【解決手段】ジョブの実行時間の統計情報及び予想実行時間を保存し、前記ジョブ実行サーバの負荷の大きさを数値化した負荷度とジョブ多重補正度と実行時間補正度とを対応させた第1の表及び予想実行時間とジョブ優先度補正値とを対応させた第2の表を格納するジョブ管理パラメータ保存領域111と、実行時間の統計情報とジョブ実行サーバ120のCPU利用率とを基に負荷度を数値化し、この負荷度が予め決められた閾値を超えていた場合、ジョブ実行サーバ120が高負荷であると判断する負荷状態判定部112と、負荷度をキーとして、第1の表を参照し、ジョブ多重補正度及び実行時間補正度を決定するパラメータ決定部113と、実行時間補正度を基に変更した予想実行時間をキーとして、第2の表を参照し、ジョブ優先度を変更するジョブパラメータ制御部114とを備える。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

ジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラーの発生を防止するようにジョブ管理サーバが管理するジョブ管理システムにおいて、

前記ジョブ管理サーバは、

前記ジョブ実行サーバで実行されたジョブの実行時間の統計情報及び予想実行時間を保存し、前記ジョブ実行サーバの負荷の大きさを数値化した負荷度とジョブ多重度を補正するジョブ多重補正度と前記予想実行時間を補正する実行時間補正度とを対応させた第 1 の表及び前記予想実行時間と前記ジョブ実行サーバに投入されるジョブの優先度であるジョブ優先度を補正するジョブ優先度補正值とを対応させた第 2 の表を格納するジョブ管理パラメータ保存領域と、

10

前記実行時間の統計情報と前記ジョブ実行サーバの CPU 使用率とを基に前記負荷度を数値化し、該数値化した前記負荷度が予め決められた閾値を超えていた場合、前記ジョブ実行サーバが高負荷であると判断する負荷状態判定部と、

前記負荷度をキーとして、前記第 1 の表を参照し、前記ジョブ多重補正度及び前記実行時間補正度を決定するパラメータ決定部と、

高負荷であると判断された前記ジョブ実行サーバに投入されるジョブに対して、前記実行時間補正度を基に予想実行時間を変更し、かつ該予想実行時間をキーとして、前記第 2 の表を参照し、前記ジョブ優先度補正值を決定し、前記ジョブ優先度補正值に基づいて前記ジョブ優先度を変更するジョブパラメータ制御部と、

20

前記パラメータ決定部で決定された前記ジョブ多重補正度を高負荷と判断された前記ジョブ実行サーバに送信するジョブ多重補正度送信手段と、

を備え、

前記ジョブ実行サーバは、

前記ジョブ多重補正度を受信するジョブ多重補正度受信手段と、

該受信した前記ジョブ多重補正度から新しいジョブ多重度を計算する新ジョブ多重度設定部と、

該新しいジョブ多重度を保存するジョブ多重度保存領域と、

を備えることを特徴とするジョブ管理システム。

30

【請求項 2】

前記負荷状態判定部は、前記ジョブ実行サーバの CPU 使用率及び前記実行時間の統計情報と最新の実行時間との差分から前記負荷度を定期的に算出することを特徴とする請求項 1 に記載のジョブ管理システム。

【請求項 3】

前記ジョブ管理サーバは、前記ジョブパラメータ制御部が決定した前記ジョブ優先度に基づいて、前記ジョブ実行サーバにジョブを投入することを特徴とする請求項 1 又は 2 に記載のジョブ管理システム。

【請求項 4】

前記ジョブ実行サーバは、前記ジョブ多重度保存領域に保存した前記新しいジョブ多重度に従ってジョブを処理することを特徴とする請求項 1 乃至 3 のいずれか 1 項に記載のジョブ管理システム。

40

【請求項 5】

前記パラメータ決定部は、高負荷状態と判断された前記ジョブ実行サーバの前記負荷度が前回計算した負荷度と比べて異なる場合、前記第 1 の表を参照して新たなジョブ多重補正度及び実行時間補正度を決定し、

前記ジョブパラメータ制御部は、該新しい実行時間補正度と、高負荷状態と判断される前にもともと設定されていたジョブの予想実行時間との積を新たな予想実行時間として設定し、前記第 2 の表を参照して新たなジョブ優先度補正值を決定し、該新たなジョブ優先度補正值と高負荷状態と判断される前にもともと設定されていたジョブの優先度との積を

50

新たなジョブの優先度として設定し、

前記ジョブ実行サーバは、前記ジョブ多重補正度送信手段及びジョブ多重補正度受信手段を通じて通知された該新しいジョブ多重補正度と、高負荷状態と判断される前にもともと設定されていたジョブ多重度との積を新たなジョブ多重度として前記ジョブ多重度保存領域に保存することを特徴とする請求項 1 乃至 4 のいずれか 1 項に記載のジョブ管理システム。

【請求項 6】

前記ジョブ管理サーバは、高負荷状態と判断された前記ジョブ実行サーバの前記負荷度が、平常状態と判断される閾値を下回っていた場合は、高負荷状態と判断されていた前記ジョブ実行サーバのジョブ多重度、該ジョブ実行サーバに投入されるジョブの予想実行時間及びジョブの優先度を高負荷状態と判断される前にもともと設定されていた値に変更し、平常状態と判断される閾値を高負荷状態と判断される閾値より低く設定することを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載のジョブ管理システム。

10

【請求項 7】

ジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラーの発生を防止するようにジョブ管理サーバが管理するジョブ管理の方法において、

前記ジョブ管理サーバは、

前記ジョブ実行サーバで実行されたジョブの実行時間の統計情報及び予想実行時間を保存し、前記ジョブ実行サーバの負荷の大きさを数値化した負荷度とジョブ多重度を補正するジョブ多重補正度と前記予想実行時間を補正する実行時間補正度とを対応させた第 1 の表及び前記予想実行時間と前記ジョブ実行サーバに投入されるジョブの優先度であるジョブ優先度を補正するジョブ優先度補正值とを対応させた第 2 の表を格納する手順と、

20

前記実行時間の統計情報と前記ジョブ実行サーバの CPU 使用率とを基に前記負荷度を数値化し、該数値化した前記負荷度が予め決められた閾値を超えていた場合、前記ジョブ実行サーバが高負荷であると判断する負荷状態判定手順と、

前記負荷度をキーとして、前記第 1 の表を参照し、前記ジョブ多重補正度及び前記実行時間補正度を決定するパラメータ決定手順と、

高負荷であると判断された前記ジョブ実行サーバに投入されるジョブに対して、前記実行時間補正度を基に予想実行時間を変更し、かつ該予想実行時間をキーとして、前記第 2 の表を参照し、前記ジョブ優先度補正值を決定し、前記ジョブ優先度補正值に基づいて前記ジョブ優先度を変更するジョブパラメータ制御手順と、

30

前記パラメータ決定手順で決定された前記ジョブ多重補正度を高負荷と判断された前記ジョブ実行サーバに送信するジョブ多重補正度送信手順と、

を備え、

前記ジョブ実行サーバは、

前記ジョブ多重補正度を受信するジョブ多重補正度受信手順と、

該受信した前記ジョブ多重補正度から新しいジョブ多重度を計算する新ジョブ多重度設定手順と、

該新しいジョブ多重度を保存するジョブ多重度保存手順と、

40

を備えることを特徴とするジョブ管理の方法。

【請求項 8】

前記負荷状態判定手順は、前記ジョブ実行サーバの CPU 使用率及び前記実行時間の統計情報と最新の実行時間との差分から前記負荷度を定期的に算出することを特徴とする請求項 7 に記載のジョブ管理の方法。

【請求項 9】

前記ジョブ管理サーバは、前記ジョブパラメータ制御手順が決定した前記ジョブ優先度に基づいて、前記ジョブ実行サーバにジョブを投入する手順を更に備えることを特徴とする請求項 7 又は 8 に記載のジョブ管理の方法。

【請求項 10】

50

前記ジョブ実行サーバは、前記ジョブ多重度保存手順で保存した前記新しいジョブ多重度に従ってジョブを処理する手順を更に備えることを特徴とする請求項7乃至9のいずれか1項に記載のジョブ管理の方法。

【請求項11】

前記パラメータ決定手順は、高負荷状態と判断された前記ジョブ実行サーバの前記負荷度が前回計算した負荷度と比べて異なる場合、前記第1の表を参照して新たなジョブ多重補正度及び実行時間補正度を決定し、

前記ジョブパラメータ制御手順は、該新しい実行時間補正度と、高負荷状態と判断される前にもともと設定されていたジョブの予想実行時間との積を新たな予想実行時間として設定し、前記第2の表を参照して新たなジョブ優先度補正値を決定し、該新たなジョブ優先度補正値と高負荷状態と判断される前にもともと設定されていたジョブの優先度との積を新たなジョブの優先度として設定し、

前記ジョブ実行サーバは、前記ジョブ多重補正度送信手順及びジョブ多重補正度受信手順を通じて通知された該新しいジョブ多重補正度と、高負荷状態と判断される前にもともと設定されていたジョブ多重度との積を新たなジョブ多重度として保存する手順を更に備えることを特徴とする請求項7乃至10のいずれか1項に記載のジョブ管理の方法。

【請求項12】

高負荷状態と判断された前記ジョブ実行サーバの前記負荷度が、平常状態と判断される閾値を下回っていた場合は、高負荷状態と判断されていた前記ジョブ実行サーバのジョブ多重度、該ジョブ実行サーバに投入されるジョブの予想実行時間及びジョブの優先度を高負荷状態と判断される前にもともと設定されていた値に変更し、平常状態と判断される閾値を高負荷状態と判断される閾値より低く設定することを特徴とする請求項7乃至11のいずれか1項に記載のジョブ管理の方法。

【請求項13】

ジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラーの発生を防止するようにジョブ管理サーバが管理するジョブ管理のプログラムにおいて、

前記ジョブ実行サーバで実行されたジョブの実行時間の統計情報及び予想実行時間を保存し、前記ジョブ実行サーバの負荷の大きさを数値化した負荷度とジョブ多重度を補正するジョブ多重補正度と前記予想実行時間を補正する実行時間補正度とを対応させた第1の表及び前記予想実行時間と前記ジョブ実行サーバに投入されるジョブの優先度であるジョブ優先度を補正するジョブ優先度補正値とを対応させた第2の表を格納する処理と、

前記実行時間の統計情報と前記ジョブ実行サーバのCPU使用率とを基に前記負荷度を数値化し、該数値化した前記負荷度が予め決められた閾値を超えていた場合、前記ジョブ実行サーバが高負荷であると判断する負荷状態判定処理と、

前記負荷度をキーとして、前記第1の表を参照し、前記ジョブ多重補正度及び前記実行時間補正度を決定するパラメータ決定処理と、

高負荷であると判断された前記ジョブ実行サーバに投入されるジョブに対して、前記実行時間補正度を基に予想実行時間を変更し、かつ該予想実行時間をキーとして、前記第2の表を参照し、前記ジョブ優先度補正値を決定し、前記ジョブ優先度補正値に基づいて前記ジョブ優先度を変更するジョブパラメータ制御処理と、

前記パラメータ決定処理で決定された前記ジョブ多重補正度を高負荷と判断された前記ジョブ実行サーバに送信するジョブ多重補正度送信処理と、

を前記ジョブ管理サーバに実行させ、

前記ジョブ多重補正度を受信するジョブ多重補正度受信処理と、

該受信した前記ジョブ多重補正度から新しいジョブ多重度を計算する新ジョブ多重度設定処理と、

該新しいジョブ多重度を保存するジョブ多重度保存処理と、

を前記ジョブ実行サーバに実行させることを特徴とするジョブ管理のプログラム。

【請求項14】

10

20

30

40

50

前記負荷状態判定処理は、前記ジョブ実行サーバのCPU使用率及び前記実行時間の統計情報と最新の実行時間との差分から前記負荷度を定期的に算出することを特徴とする請求項13に記載のジョブ管理のプログラム。

【請求項15】

前記ジョブパラメータ制御処理が決定した前記ジョブ優先度に基づいて、前記ジョブ実行サーバにジョブを投入する処理を、前記ジョブ管理サーバに実行させることを特徴とする請求項13又は14に記載のジョブ管理のプログラム。

【請求項16】

前記ジョブ多重度保存処理で保存した前記新しいジョブ多重度に従ってジョブを処理する処理を、前記ジョブ実行サーバに更に実行させることを特徴とする請求項13乃至15のいずれか1項に記載のジョブ管理のプログラム。

10

【請求項17】

前記パラメータ決定処理は、高負荷状態と判断された前記ジョブ実行サーバの前記負荷度が前回計算した負荷度と比べて異なる場合、前記第1の表を参照して新たなジョブ多重補正度及び実行時間補正度を決定し、

前記ジョブパラメータ制御処理は、該新しい実行時間補正度と、高負荷状態と判断される前にもともと設定されていたジョブの予想実行時間との積を新たな予想実行時間として設定し、前記第2の表を参照して新たなジョブ優先度補正値を決定し、該新たなジョブ優先度補正値と高負荷状態と判断される前にもともと設定されていたジョブの優先度との積を新たなジョブの優先度として設定し、

20

前記ジョブ多重補正度送信処理及びジョブ多重補正度受信処理を通じて通知された該新しいジョブ多重補正度と、高負荷状態と判断される前にもともと設定されていたジョブ多重度との積を新たなジョブ多重度として保存する処理を、前記ジョブ実行サーバに更に実行させることを特徴とする請求項13乃至16のいずれか1項に記載のジョブ管理のプログラム。

【請求項18】

高負荷状態と判断された前記ジョブ実行サーバの前記負荷度が、平常状態と判断される閾値を下回っていた場合は、高負荷状態と判断されていた前記ジョブ実行サーバのジョブ多重度、該ジョブ実行サーバに投入されるジョブの予想実行時間及びジョブの優先度を高負荷状態と判断される前にもともと設定されていた値に変更し、平常状態と判断される閾値を高負荷状態と判断される閾値より低く設定することを特徴とする請求項13乃至17のいずれか1項に記載のジョブ管理のプログラム。

30

【発明の詳細な説明】

【技術分野】

【0001】

ジョブ管理サーバが管理しているジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラー等が発生するのを防止するジョブ管理システム、その方法及びそのプログラムに関する。

【背景技術】

【0002】

一般的なジョブ管理システムでは、サーバの負荷が一時的に増大するなどの現象が発生していてもジョブの実行に影響がないことが求められている。この技術的要件を満たすためには、サーバが高負荷状態になった場合に、高負荷状態をどのようにして緩和するか、高負荷状態によるジョブ実行遅延の影響をどのようにして抑止するかといった課題が存在する。

40

【0003】

そこで、特許文献1には、CPU負荷量毎にジョブ実行を動的に変化させるジョブスケジューリング方式が開示されている。

【0004】

特許文献2には、業務内の各ジョブに、統計情報のCPU使用時間に応じて割り当てる

50

CPU使用率を自動で算出するCPU使用率最適化方式が開示されている。

【0005】

特許文献3には、サーバのCPU使用率が閾値下限を越えている場合には、サーバが過負荷状態と判断するCPU余裕管理とトランザクション優先度による負荷分散方法が開示されている。

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開平05-173807号公報

【特許文献2】特開平09-305417号公報

【特許文献3】特開2006-113827号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、特許文献1乃至3に開示されている発明は、単にCPU使用率が高いだけでは、ハードウェア資源を有効に活用しているからCPU使用率が高いのか、過剰にジョブが投入された結果、CPU使用率が高いのかが判断できないという問題点がある。

【0008】

又、CPU使用率が低くても、ハードディスクへの読み込み又は書き込みが頻繁に発生する等の他の要素でジョブが遅延する可能性もあり、そのような場合には、CPU使用率だけではハードウェア資源の負荷の度合いが正確に把握できないという問題点もある。

【0009】

本発明は上記に鑑みてなされたもので、ハードウェア資源の有効活用が可能なジョブ管理システム、その方法及びそのプログラムを提供することを目的とする。

【課題を解決するための手段】

【0010】

上述の問題を解決するため、本発明に係るジョブ管理システムは、ジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラーの発生を防止するようにジョブ管理サーバが管理するジョブ管理システムにおいて、前記ジョブ管理サーバは、前記ジョブ実行サーバで実行されたジョブの実行時間の統計情報及び予想実行時間を保存し、前記ジョブ実行サーバの負荷の大きさを数値化した負荷度とジョブ多重度を補正するジョブ多重補正度と前記予想実行時間を補正する実行時間補正度とを対応させた第1の表及び前記予想実行時間と前記ジョブ実行サーバに投入されるジョブの優先度であるジョブ優先度を補正するジョブ優先度補正值とを対応させた第2の表を格納するジョブ管理パラメータ保存領域と、前記実行時間の統計情報と前記ジョブ実行サーバのCPU使用率とを基に前記負荷度を数値化し、該数値化した前記負荷度が予め決められた閾値を超えていた場合、前記ジョブ実行サーバが高負荷であると判断する負荷状態判定部と、前記負荷度をキーとして、前記第1の表を参照し、前記ジョブ多重補正度及び前記実行時間補正度を決定するパラメータ決定部と、高負荷であると判断された前記ジョブ実行サーバに投入されるジョブに対して、前記実行時間補正度を基に予想実行時間を変更し、かつ該予想実行時間をキーとして、前記第2の表を参照し、前記ジョブ優先度補正值を決定し、前記ジョブ優先度補正值に基づいて前記ジョブ優先度を変更するジョブパラメータ制御部と、前記パラメータ決定部で決定された前記ジョブ多重補正度を高負荷と判断された前記ジョブ実行サーバに送信するジョブ多重補正度送信手段と、を備え、前記ジョブ実行サーバは、前記ジョブ多重補正度を受信するジョブ多重補正度受信手段と、該受信した前記ジョブ多重補正度から新しいジョブ多重度を計算する新ジョブ多重度設定部と、該新しいジョブ多重度を保存するジョブ多重度保存領域と、を備えることを特徴とする。

【0011】

上述の問題を解決するため、本発明に係るジョブ管理の方法は、ジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラーの発生を防

10

20

30

40

50

止するようにジョブ管理サーバが管理するジョブ管理の方法において、前記ジョブ管理サーバは、前記ジョブ実行サーバで実行されたジョブの実行時間の統計情報及び予想実行時間を保存し、前記ジョブ実行サーバの負荷の大きさを数値化した負荷度とジョブ多重度を補正するジョブ多重補正度と前記予想実行時間を補正する実行時間補正度とを対応させた第1の表及び前記予想実行時間と前記ジョブ実行サーバに投入されるジョブの優先度であるジョブ優先度を補正するジョブ優先度補正值とを対応させた第2の表を格納する手順と、前記実行時間の統計情報と前記ジョブ実行サーバのCPU使用率とを基に前記負荷度を数値化し、該数値化した前記負荷度が予め決められた閾値を超えていた場合、前記ジョブ実行サーバが高負荷であると判断する負荷状態判定手順と、前記負荷度をキーとして、前記第1の表を参照し、前記ジョブ多重補正度及び前記実行時間補正度を決定するパラメータ決定手順と、高負荷であると判断された前記ジョブ実行サーバに投入されるジョブに対して、前記実行時間補正度を基に予想実行時間を変更し、かつ該予想実行時間をキーとして、前記第2の表を参照し、前記ジョブ優先度補正值を決定し、前記ジョブ優先度補正值に基づいて前記ジョブ優先度を変更するジョブパラメータ制御手順と、前記パラメータ決定手順で決定された前記ジョブ多重補正度を高負荷と判断された前記ジョブ実行サーバに送信するジョブ多重補正度送信手順と、を備え、前記ジョブ実行サーバは、前記ジョブ多重補正度を受信するジョブ多重補正度受信手順と、該受信した前記ジョブ多重補正度から新しいジョブ多重度を計算する新ジョブ多重度設定手順と、該新しいジョブ多重度を保存するジョブ多重度保存手順と、を備えることを特徴とする。

10

【0012】

20

上述の問題を解決するため、本発明に係るジョブ管理のプログラムは、ジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラーの発生を防止するようにジョブ管理サーバが管理するジョブ管理のプログラムにおいて、前記ジョブ実行サーバで実行されたジョブの実行時間の統計情報及び予想実行時間を保存し、前記ジョブ実行サーバの負荷の大きさを数値化した負荷度とジョブ多重度を補正するジョブ多重補正度と前記予想実行時間を補正する実行時間補正度とを対応させた第1の表及び前記予想実行時間と前記ジョブ実行サーバに投入されるジョブの優先度であるジョブ優先度を補正するジョブ優先度補正值とを対応させた第2の表を格納する処理と、前記実行時間の統計情報と前記ジョブ実行サーバのCPU使用率とを基に前記負荷度を数値化し、該数値化した前記負荷度が予め決められた閾値を超えていた場合、前記ジョブ実行サーバが高負荷であると判断する負荷状態判定処理と、前記負荷度をキーとして、前記第1の表を参照し、前記ジョブ多重補正度及び前記実行時間補正度を決定するパラメータ決定処理と、高負荷であると判断された前記ジョブ実行サーバに投入されるジョブに対して、前記実行時間補正度を基に予想実行時間を変更し、かつ該予想実行時間をキーとして、前記第2の表を参照し、前記ジョブ優先度補正值を決定し、前記ジョブ優先度補正值に基づいて前記ジョブ優先度を変更するジョブパラメータ制御処理と、前記パラメータ決定処理で決定された前記ジョブ多重補正度を高負荷と判断された前記ジョブ実行サーバに送信するジョブ多重補正度送信処理と、を前記ジョブ管理サーバに実行させ、前記ジョブ多重補正度を受信するジョブ多重補正度受信処理と、該受信した前記ジョブ多重補正度から新しいジョブ多重度を計算する新ジョブ多重度設定処理と、該新しいジョブ多重度を保存するジョブ多重度保存処理と、を前記ジョブ実行サーバに実行させることを特徴とする。

30

40

【発明の効果】

【0013】

以上説明したように、本発明によれば、ジョブ実行サーバの負荷の大きさを数値化した負荷度に基づいて制御することにより、ハードウェア資源の有効活用が可能なジョブ管理システム、その方法及びそのプログラムを提供することができる。

【図面の簡単な説明】

【0014】

【図1】本発明の第1の実施の形態に係るジョブ管理システムの構成図である。

【図2】本図2は、本発明の第1の実施の形態に係るジョブ管理システムの負荷度の算出

50

からジョブ実行サーバのジョブ多重度を設定する処理の流れを示すブロック図である。

【図3】本発明の第1の実施の形態に係るジョブ管理システムの高負荷状態と判断されたジョブ実行サーバへのジョブ投入を行う際のパラメータの設定処理の流れを示すブロック図である。

【発明を実施するための形態】

【0015】

以下、図面を参照しながら、本発明の実施の形態について詳細に説明する。

【0016】

[第1の実施の形態]

図1は、本発明の第1の実施の形態に係るジョブ管理システムの構成図である。この図1を参照すると、本発明の第1の実施の形態に係るジョブ管理システムは、ジョブ管理サーバ110と、ジョブ実行サーバ120とから構成されている。

10

【0017】

ジョブ管理サーバ110は、各ジョブの実行時間の統計情報及び予想実行時間を保存し、ジョブ実行サーバ120の負荷の大きさを数値化した負荷度とジョブ多重度を補正するジョブ多重補正度と予想実行時間を補正する実行時間補正度とを対応させた第1の表及び予想実行時間とジョブ実行サーバ120に投入されるジョブの優先度であるジョブ優先度を補正するジョブ優先度補正值とを対応させた第2の表を格納する領域であるジョブ管理パラメータ保存領域111と、ジョブ管理サーバ110が管理している各ジョブ実行サーバ120がどれくらい高負荷であるかをジョブ管理パラメータ保存領域111に保存されている実行時間の統計情報と、各ジョブ実行サーバ120のCPU使用率とを基に投入されたジョブがどの程度遅延しているかを示す負荷度として数値化する負荷状態判定部112と、負荷状態判定部112で決定される負荷度をキーとして、ジョブ管理パラメータ保存領域111に保存されている第1の表を参照し、ジョブ多重度及び予想実行時間をそれぞれの程度補正するかのパラメータ(ジョブ多重補正度、実行時間補正度)を決定するパラメータ決定部113と、縮退運用中に高負荷と判断されたジョブ実行サーバ120に投入されるジョブに対して、実行時間補正度を基に予想実行時間を変更し、この予想実行時間をキーとして第2の表を参照して予想実行時間に応じてどの程度優先度を変更するかのパラメータ(ジョブ優先度補正值)を決定し、このジョブ優先度補正值に基づいてジョブ優先度を変更するジョブパラメータ制御部114と、パラメータ決定部113で決定されたジョブ多重補正度を高負荷と判断されたサーバに送信するジョブ多重補正度送信手段115と、を備える。

20

30

【0018】

ジョブ実行サーバ120は、ジョブ多重補正度送信手段115から送信されるジョブ多重補正度を受信するジョブ多重補正度受信手段121と、ジョブ多重補正度受信手段121で受信したジョブ多重補正度から新しいジョブ多重度を計算し、ジョブ多重度保存領域123に保存する新ジョブ多重度設定部122と、ジョブ実行サーバ120のジョブ多重度が保存されているジョブ多重度保存領域123で構成されている。

【0019】

[第1の実施の形態の動作]

本発明の第1の実施の形態に係るジョブ管理システムの動作の流れを図面を用いて詳細に説明する。

40

【0020】

最初に、図2を用いて負荷度の算出からジョブ実行サーバのジョブ多重度を設定する処理の説明を行う。図2は、本発明の第1の実施の形態に係るジョブ管理システムの負荷度の算出からジョブ実行サーバのジョブ多重度を設定する処理の流れを示すブロック図である。

【0021】

負荷状態判定部112が、ジョブ管理サーバ110が管理している各ジョブ実行サーバ120のCPU使用率と、各ジョブ実行サーバで実行されたジョブの実行時間の統計と最

50

新の実行時間の差分から負荷情報を分析し、各ジョブ実行サーバの負荷度というパラメータを定期的に算出する処理を行う。ここで、負荷状態判定部 112 は、各ジョブ実行サーバの負荷度が予め決められた閾値を超えていた場合、当該サーバが高負荷状態であると判断する。

【0022】

なお、各ジョブ実行サーバ 120 の CPU 使用率は、各ジョブ実行サーバ 120 にインストールされている OS (Operating System) が測定したものをジョブ管理サーバ 110 がネットワークを通じて取得するものとする。

【0023】

各ジョブ実行サーバ 120 は、この CPU 使用率をジョブ管理サーバ 110 に送信する専用の手段を別途備え、又ジョブ管理サーバ 110 は、CPU 使用率を受信する専用の手段を別途備えてもよいが、各ジョブ実行サーバ 120 とジョブ管理サーバ 110 とがサーバとして備えている既存の通信手段を援用してもよい。

10

【0024】

次に、パラメータ決定部 113 は、高負荷と判断されたジョブ実行サーバ 120 に関してジョブ管理パラメータ保存領域 111 に保存された負荷度とジョブ多重補正度と実行時間補正度との第 1 の表を参照して、高負荷なジョブ実行サーバのジョブ多重補正度、実行時間補正度のパラメータを決定する。

【0025】

負荷度とジョブ多重補正度と実行時間補正度とを対応させた第 1 の表はサイジング等によって予め決定し、ジョブ管理パラメータ保存領域 111 に登録する。その例を表 1 に示す。

20

【0026】

【表 1】

負荷度	ジョブ多重補正度	実行時間補正度
4	0.7 度	1.3 倍
5	0.6 度	1.4 倍
6	0.5 度	1.5 倍
...

30

次に、ジョブ多重補正度送信手段 115 は、高負荷状態と判断されたジョブ実行サーバに、ジョブ多重補正度のパラメータを送信する。

40

【0027】

ジョブ多重補正度受信手段 121 は、ジョブ多重補正度のパラメータを受信する。

【0028】

新ジョブ多重度設定部 122 は、ジョブ実行サーバの現在のジョブ多重度とジョブ多重補正度の積を新しいジョブ実行サーバのジョブ多重度としてジョブ多重度保存領域 123 に保存する。

【0029】

第二に、図 3 を用いて高負荷状態と判断されたジョブ実行サーバへのジョブ投入を行う際のパラメータの設定処理の説明を行う。図 3 は、本発明の第 1 の実施の形態に係るジョブ管理システムの高負荷状態と判断されたジョブ実行サーバ 120 へのジョブ投入を行う

50

際のパラメータの設定処理の流れを示すブロック図である。

【 0 0 3 0 】

ジョブパラメータ制御部 1 1 4 は、高負荷状態と判断されたジョブ実行サーバ 1 2 0 に投入されるジョブに関して、現在のジョブの予想実行時間と実行時間補正度の積を新しいジョブの予想実行時間として設定する。

【 0 0 3 1 】

次に、予想実行時間とジョブの優先度をどの程度補正するかのパラメータ（ジョブ優先度補正值）との第 2 の表を参照し、新たに設定した予想実行時間に対応するジョブ優先度補正值を取得し、得られたジョブ優先度補正值と投入されるジョブの優先度の積を新しいジョブの優先度として設定する。この負荷度を基にした予想実行時間とジョブ優先度補正值とを対応させた第 2 の表は予め設定しておき、ジョブ管理パラメータ保存領域 1 1 1 に登録する。その例を表 2 に示す。

10

【 0 0 3 2 】

【表 2】

予想実行時間	ジョブ優先度補正值
20秒未満	1.5倍
20秒以上 30秒未満	1.3倍
30秒以上 40秒未満	1倍
...	...

20

30

ジョブ管理サーバ 1 1 0 は、ジョブパラメータ制御部 1 1 4 が決定したジョブ優先度に基づいて、ジョブ実行サーバ 1 2 0 にジョブを投入する。又、ジョブ実行サーバ 1 2 0 は、ジョブ多重度保存領域 1 2 3 に保存した新しいジョブ多重度に従ってジョブを処理する。これにより、ジョブ実行サーバの高負荷状態の緩和、高負荷状態が原因の予想実行時間超過によるジョブのエラー等が発生するのを防止するジョブ管理システムを実現できる。

【 0 0 3 3 】

[第 2 の実施の形態]

本発明の第 2 の実施の形態として、図 2 において、負荷状態判定部 1 1 2 で負荷度を算出した際、高負荷状態と判断されたジョブ実行サーバ 1 2 0 の負荷度が前回計算した負荷度と比べて異なる場合と、負荷度が平常状態と判断される負荷度の閾値を下回っていた場合とで、それぞれどのような動作になるかを示す。

40

【 0 0 3 4 】

高負荷状態と判断されたジョブ実行サーバの負荷度が前回計算した負荷度と異なる場合は、パラメータ決定部 1 1 3 は、再びジョブ管理パラメータ保存領域 1 1 1 に保存されている負荷度とジョブ多重補正度と実行時間補正度との第 1 の表を参照し、新たなジョブ多重補正度及び実行時間補正度を決定する。

【 0 0 3 5 】

その後、ジョブ多重補正度送信手段 1 1 5、ジョブ多重補正度受信手段 1 2 1 を通じてジョブ実行サーバに通知された新しいジョブ多重補正度と、高負荷状態と判断される前に

50

もともと設定されていたジョブ多重度との積を新たなジョブ多重度としてジョブ多重度保存領域 1 2 3 に保存する。

【 0 0 3 6 】

同様に、ジョブパラメータ制御部 1 1 4 でも、新しい実行時間補正度と、高負荷状態と判断される前にもともと設定されていたジョブの予想実行時間との積を新たな予想実行時間として設定する。又、ジョブの優先度についても、新たな負荷度ごとの予想実行時間とジョブ優先度補正值との第 2 の表を参照し、新たなジョブ優先度補正值を決定する。その後、新たなジョブ優先度補正值と高負荷状態と判断される前にもともと設定されていたジョブの優先度との積を新たなジョブの優先度として設定する。

【 0 0 3 7 】

高負荷状態と判断されたジョブ実行サーバ 1 2 0 の負荷度が、平常状態と判断される閾値を下回っていた場合は、高負荷状態と判断されていたジョブ実行サーバ 1 2 0 のジョブ多重度、当該サーバに投入されるジョブの予想実行時間及びジョブの優先度を高負荷状態と判断される前にもともと設定されていた値に変更する。平常状態と判断される閾値と、高負荷状態と判断される閾値は異なり、平常状態と判断される閾値の方を高負荷状態と判断される閾値より低く設定する。

【 0 0 3 8 】

なお、本発明は、装置のみならず方法及びプログラムの発明としても実施できる。

【 産業上の利用可能性 】

【 0 0 3 9 】

本発明は、縮退運用方式でのジョブ管理に適しており、かかる方式でのハードウェア資源の有効活用が可能なジョブ管理システムに利用することができる。

【 符号の説明 】

【 0 0 4 0 】

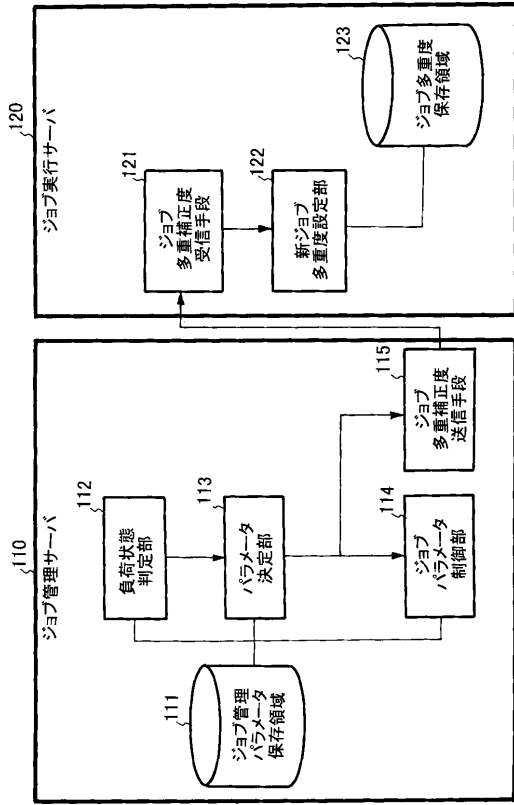
- 1 1 0 ジョブ管理サーバ
- 1 1 1 ジョブ管理パラメータ保存領域
- 1 1 2 負荷状態判定部
- 1 1 3 パラメータ判定部
- 1 1 4 ジョブパラメータ制御部
- 1 1 5 ジョブ多重補正度送信手段
- 1 2 0 ジョブ実行サーバ
- 1 2 1 ジョブ多重補正度受信手段
- 1 2 2 新ジョブ多重度設定部
- 1 2 3 ジョブ多重度保存領域

10

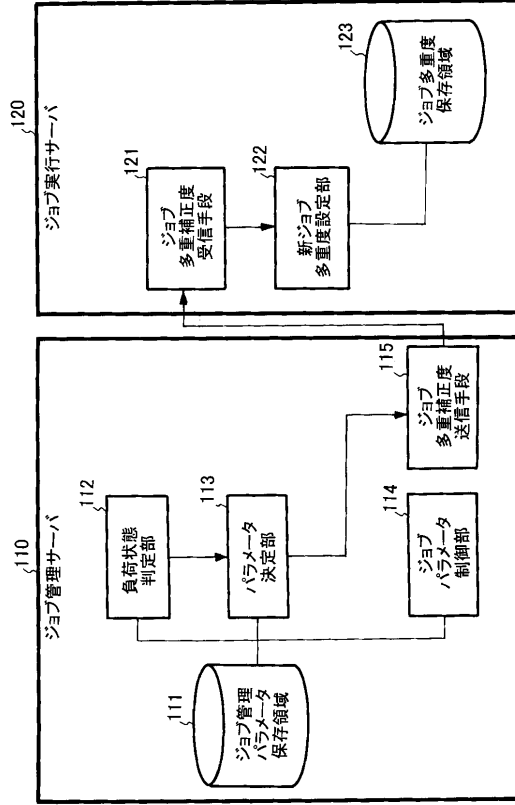
20

30

【 図 1 】



【 図 2 】



【 図 3 】

