複合システムにおけるジョブモニタの実装

4H－8

岩崎　元一　（株）東芝　青梅工場

1　はじめに

いくつかの計算機を高速の通信機構を介して結合し、これを単一のシステムとして運用する複合システム構成（以下、単に複合システムと呼ぶ）において、システム下で実行するジョブ群を一括して監視し、実時間で実行状況を表示し、また実行情報を蓄積し、これを編集して、実行履歴をレポートする機能（以下、ジョブモニタと呼ぶ）を開発したので報告する。

今回開発したジョブモニタが監視対象とする複合システムは、

（1）主メモリは各ノードに固有（共有しない）。

（2）各ノードから斉一なインタフェースで共有ファイルをアクセス可能。

（3）各ノードは、それぞれのＯＳによって独立に動作。

（4）各ノードのプロセス同士は特別な通信プリミティブによって容易に通信可能。

という特徴を持っており、本ジョブモニタは、この条件の下で以下に示す要求仕様を満たすように設計されている。

2　ジョブモニタの要求仕様

今回開発したジョブモニタに対する要求仕様は、次のようなものである。

（1）複合システムで実行されるジョブを一括して監視することができる。

（2）実行履歴情報は、複合システム全体で過去８１９２ジョブ分を保持する。

（3）各計算機（以下、ノードと呼ぶ）の主メモリの使用量は１ＭＢ程度以下とする。

（4）ジョブモニタ情報採取は、ジョブの起動終了時間にほとんど影響を与えない。

（5）どのノードからでもジョブモニタ機能を利用できる。

（6）実行状況の表示、実行履歴のレポートにおいて、条件による選択（ある特定のシステムコードで異常終了したジョブの検索など）を行うことができ、かつこれは、通常のシステム負荷の範囲内であれば実用に耐える程度の時間で終了する。

（7）複合システムを構成する計算機の一つがダウンしたとき、他の計算機でのジョブモニタ機能の情報提供サービス、及び実行中のジョブの実行および情報の収集は継続する。さらに、上記計算機がダウンから復帰したとき、ダウン前、ダウン中の情報も含めてジョブモニタ機能を利用することができる。

Job Monitor on Loosely Coupled Computers

Motokazu Iwasaki

Ome Works、TOSHIBA Corporation

3　情報の管理形態

本ジョブモニタにおいては、要求仕様と、対象とする複合システムの特徴とを考慮して、次のような情報管理形態を採用した。

（1）実行履歴情報の量（８１９２ジョブ分）と、使用メモリの制限（１ＭＢ）とから、情報の実体のすべてを各ノードの主メモリに置くことはできないので、実行履歴情報の実体は共有のランダムアクセスファイル（以下、実行履歴情報ファイルと呼ぶ）上に置く。ただし、実行状況の表示、実行履歴のレポートにおける“選択”の高速性のため、全システムの実行履歴情報をマージした情報に関する実行履歴情報管理構造を、マスタノードの主メモリ上に置く。実行履歴情報管理構造は、選択用のキー情報と、実行履歴情報ファイル上の情報の実体へのポインタとを持つ。

（2）各ノードのジョブモニタシステムはほとんど独立に動作し、実行履歴情報ファイルのレコードの更新は、各ジョブごとに行う。あるノードに投入されたジョブは、自身のノードのジョブモニタシステムに実行履歴情報管理構造のエントリ及び実行履歴情報ファイルのレコードの獲得を要求する。要求を受けたジョブモニタシステムは、マスタノードのジョブモニタシステム（以下、マスタジョブモニタシステムと呼ぶ）に依頼し、当該ジョブが使うべきレコードへのポインタを獲得して、これをジョブに返却する。以降、所定の事象（プロセスの生成、消滅など）が発生するたびに、ジョブは自身のエントリ及びレコードに情報を書き込む（この処理においては、レコードの更新の排他制御は不要なので、各ジョブは並列して情報を更新することができる）。また、ジョブモニタシステムは、この情報更新に実行履歴情報管理構造の変更が伴うとき、これをマスタジョブモニタシステムに通知する。通知を受けたマスタジョブモニタシステムは、主メモリ上の実行履歴情報管理構造を更新する。

（3）実行状況の表示、実行履歴のレポートを要求されたとき、ジョブモニタシステムは、必要なら“選択”用のキー情報を付けてマスタジョブモニタシステムに依頼し、実行履歴情報ファイル上の読むべきレコードへのポインタのリストを獲得する。この後、ジョブモニタシステムは、このレコードへのポインタのリストを使って、ファイルから必要な情報の実体のみを読み込んで、利用者に情報を返却する。

（4）複合システムを構成するノードの一つがダウンしたとき、これがマスタジョブモニタシステムのノードではないときには、単にそのノードからの情報更新が行われなくなるだけである。マスタジョブモニタシステムのノードがダウンしたときには、複合システム内に新にマスタジョブモニタシステムが選出され、新マスタジョブモニタシステムは、必要な管理情報の再構成を行った後、実行履歴情報ファイルのレコードの割り当てと、実行履歴情報管理構造の更新処理を再開する。計算機がダウンから復帰するときは、実行履歴情報ファイルから主メモリ上の実行履歴情報管理構造を再構成する。こののち、ジョブモニタ機能から、ダウン前、ダウン中の情報も含めて、実行状況の表示、実行履歴のレポートを行うことができる。

4　まとめ

主メモリ非共有、ファイル共有の複合システム上で、各ノードのジョブを一括して監視できるようなジョブモニタシステムを開発した。本方式は、ノード数が少数のネットワークにおいても適用することができる。