3A-7

WAN接続クラスタ群をメモリ資源として利用するための メモリサーバ自動選定システム

鈴木 悠一郎† 緑川 博子† 市野 晴菜†

成蹊大学理工学部情報科学科†

1. はじめに

筆者らは，クラスタの遠隔ノードのメモリを利用し，仮想的に大容量のメモリ空間を逐次プログラムに提供するシステム，DLM(Distributed Large Memory)を構築，評価してきた[1].

DLMにはメモリサーバを常駐型プロセスにした，マルチクライアント向けDLM-Mシステム[2]があり，クラスタ内の複数の常駐メモリサーバから自動で利用メモリサーバを選択し，負荷分散を行う管理プロセスをシステムに導入している[3][図1]．このシステムをさらに拡張し，WANで接続したクラスタ群の中から，メモリ資源として適切なクラスタ，メモリサーバノード，計算ノードを自動選定して実行するクラスタ自動選定システムを提案している[4].

今回，クラスタ自動選定システムをIntriggerシステム[5]上に実際に構築し，直接的なクラスタ接続環境がないユーザのパソコンからもDLMを使用したプログラムの実行ができるような，ポータルサイトアクセスを可能にするWebインターフェースも構築した.

2. 稼働環境

本システムは，以下の環境を持つ WAN で接続されたクラスタ群で利用できる.

a. クラスタ間でのユーザアカウントは同一.
b. 同一クラスタ内のノードはホームディレクトリが共通.
c. クラスタ内のノードはグローバル IP を持ち，WAN での遠隔アクセスが可能.

本報告では，多くの大学のクラスタを結合した分散コンピューティングシステムであるInTrigger[5]を用いた.

3. クラスタ自動選定システム(DLM-WAN-ADMIN)

WAN 接続クラスタ群で，大容量のメモリ空間を必要とする逐次プログラムを実行するにあたって，メモリ資源として条件の良いクラスタを自動選定する，クラスタ自動選定システム(DLM-ADMIN) を構築した.

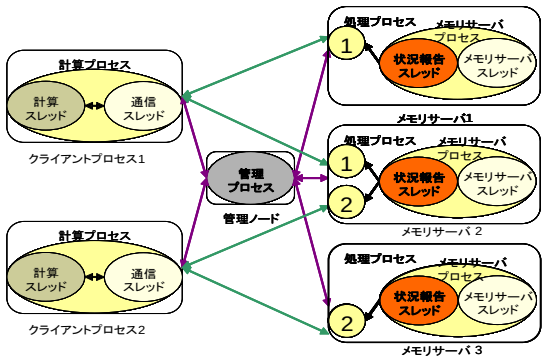

図1. DLM-M System

クラスタ内の情報は本システム用に拡張したDLM-M システムの管理プロセス(LAN-ADMIN)が管理する．WAN-ADMIN はクラスタごとの LAN-ADMIN と通信をすることにより，クラスタ内の情報取得および実行を行う.

クラスタの選択ポリシーは以下の通り.

a. 計算ノードのローカルメモリ量のみで実行できるクラスタを優先的に選択する．ユーザの要求メモリ量より，利用されていないメモリ量(MemFree)が多く，CPU の load が低いクラスタを選択する．MemFree のみでは不足する場合は，解放される予定のメモリ量(Inactive)との合算値で選択する.
b. 計算ノードのローカルメモリ量で要求メモリ量を満たせない場合は，遠隔ノードを利用する DLM-M システムでプログラムの実行を行う．１クラスタ内の全メモリサーバの MemFree と Inactive の合算値が，ユーザのメモリ要求量を満たすクラスタのうち，値が低いクラスタから選択する.

運用状況に合わせて，以下の Large と Fast の２つのモードがクラスタごとに設定できる．また，組み合わせた使用もできる.

a. Large モード
ユーザが要求しているメモリ量を持つクラスタの中で，最小のメモリ量を持つクラスタを選択するモード．できるだけ空き容量

Automatic memory server allocation for sequential large data processing using remote memory on WAN-connected clusters

†Yuichiro Suzuki, Hiroko Midorikawa, Haruna Ichino
†Department of Computer and Information Science, Seikei University

が少ないクラスタから使用していき，空メモリ量が多いクラスタを残す．後続の容量の大きいメモリを使用するプログラムの実行ができるようにしている．

b. Fast モード
クラスタ内での DLM の利用ユーザ数をクラスタ間でできるだけ均等に割り振るモード．DLM を使用するプログラムが 1 クラスタに集中すると，一般的にノード間通信に負荷がかかるため動作が遅くなる．このモードではできるだけ通信に負荷がかからないようにし，各ユーザへのレスポンスを早くするようにしている．

4. WAN での全体構成

WAN 全体の構成は，図 2 の通りである．ブラウザのみで DLM システムを使用することが可能となっている．クラスタの一つは WAN-ADMIN が動作するノード（WAN-ADMIN ノード）を持つ．

ユーザは，ブラウザを通して Web サーバにアクセスし，実行プログラムやソースファイル，入力ファイルをアップロードする．Web サーバは，さらに WAN-ADMIN ノードへファイルをアップロードし，WAN-ADMIN とやりとりを行う．WAN-ADMIN は 3. で述べたように，ユーザプログラムに条件の良いクラスタと計算ノードを選定し，選定クラスタへ実行するファイルを転送する．転送先の計算ノードでは，DLM-M システムを使用したプログラムの実行が行われ，実行が終了すると出力ファイルと実行結果ファイルが，WAN-ADMIN ノードへ送られる．ユーザは，ブラウザを通して実行結果ファイルと出力ファイル等ユーザのディレクトリに入っているファイルをダウンロードすることができる．

図 3 に，Web インターフェースの例を挙げる．ユーザは，ファイルのアップロードとコマンド指定等で，容易に利用が可能となっている．

5. おわりに

本システムによって，ユーザは大容量のメモリ空間を必要とする逐次プログラムを，複数のクラスタやその中のノードを意識せずに，遠隔メモリを用いて実行することが可能となった．

また，Web インターフェースを使うことにより，ユーザはクラスタへの接続環境なしに Web ブラウザのみで利用することも可能となり，ユーザへの利便性の向上ができた．

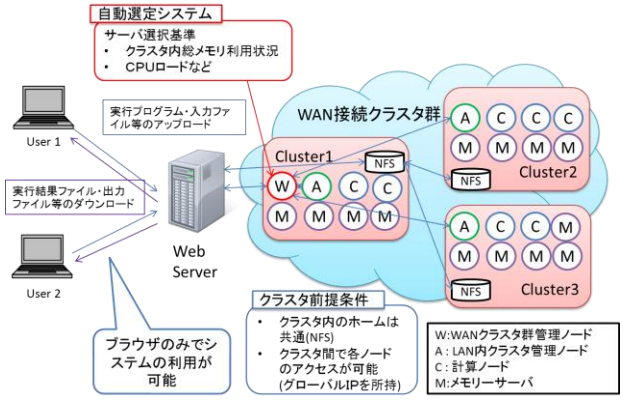

図 2. WAN での DLM システム全体図

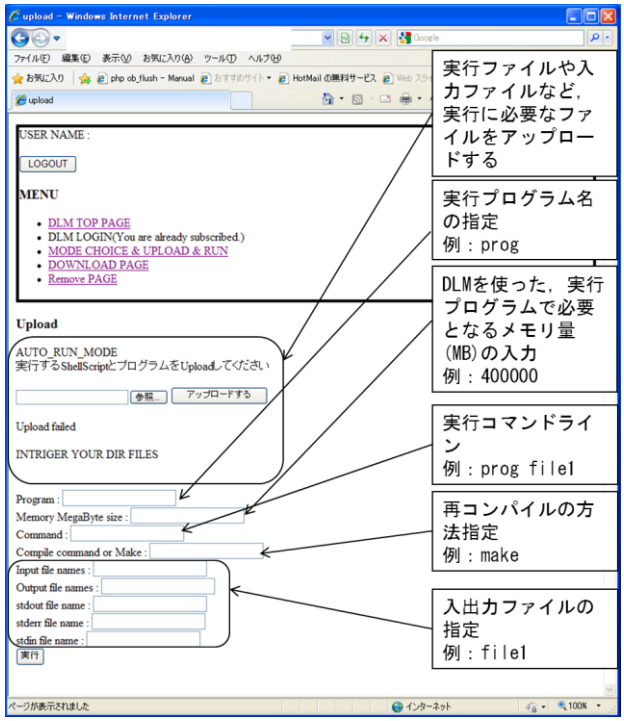

図 3. Web インターフェース例

参考文献

[1] 緑川,黒川,姫野: "遠隔メモリを利用する分散大容量メモリシステム DLM の設計と 10GbEthernet における初期性能評価", 情報処理学会論文誌 ACS, Vol.1,No 3,pp 136-157（2008,12）

[2] 齋藤, 緑川, 甲斐:"マルチクライアント向け分散型大容量メモリシステム DLM-M の設計と実装",FIT2008 論文集,C－003,pp.199-200,（2008,9）

[3] 三浦,緑川,甲斐:" クラスタをメモリ資源として利用するための動的メモリ提供システムの提案",FIT2009 論文集,B-029, pp.421-422,（2009,9）

[4] 鈴木，緑川："分散大容量メモリ DLM の WAN 接続クラスタ群への適用－クラスタ・サーバ自動選定システムの提案－", SACSIS2010 論文集，pp101-102,(2010,5)

[5] 田浦: "InTrigger：オープンな情報処理・システム研究プラットフォーム",情報処理 Vol.49 No.8,pp939-944,(2009.8)