

L-003

加速器科学仮想組織におけるグリッド環境自動監視・診断システム
 An automatic monitoring and diagnostic system of the grid environment
 in the Accelerator Science Virtual Organization

藤井峰夫† Mineo Fujii 長坂康史† Yasushi Nagasaka 渡瀬芳行‡ Yoshiyuki Watase 佐々木節‡ Takashi Sasaki 岩井剛‡ Go Iwai

1. はじめに

グリッド・ミドルウェアの一つである gLite を用いて、国内 6 研究機関で構成される加速器科学仮想組織が形成された。これにより各研究機関の計算資源を共有することが可能となった。しかし、これらの環境と機能を維持するためには、適切な管理および運用を行う必要がある。そのため、現状では手作業を伴う煩雑な作業を継続的に行っており、その維持は容易ではない。そこで本研究では、グリッド・システムの運用を支援するために自動監視・診断システムを開発し、仮想組織上にこれを展開した。このシステムを利用することで、管理者は管理している拠点の状態を迅速に把握でき、障害発生時に的確な処置を行えるなど、サービスの可用性を高めることが可能となる。

2. グリッド環境

欧州原子核研究機構 (CERN) で行われる高エネルギー加速器 (LHC: Large Hadron Collider) を利用した実験のデータを処理するためのグリッド環境を構築するプロジェクトを LCG (LHC Computing Grid Project) という。このプロジェクトで採用されているミドルウェアを用いて、国内でのグリッド環境を構築した。また、構築されたグリッド環境下で大量のデータを管理・処理することの出来る仮想組織 (VO: Virtual Organization) の基盤を確立した。

ここで定義する仮想組織とは、ネットワークで繋がれた資源 (CPU などのコンピュータ資源) を共有、もしくは、それを利用して共同作業をするグループのことである。仮想組織内にある各拠点では、すべての拠点にある計算資源を自由に使い解析を行うことが出来る。また、外からの攻撃に対しては共同で防衛線を張る必要がある。

3. グリッド・ミドルウェア

本仮想組織ではグリッド・コンピューティング環境を構築するためのミドルウェアとして EGEE (Enabling Grids for E-science) が開発した gLite[1] と呼ばれるグリッド・ミドルウェアを用いている。

3.1 gLite

gLite では複数の機能[2]が相互に連携することでグリッド環境を構築する。グリッド環境へのログイン・認証には電子証明書を利用する。また、gLite は仮想組織のユーザ管理機能や、ジョブのコントロール、アカウント管理、リソース情報の管理・共有などの機能を持つ。

3.2 グリッドの機能

gLite グリッド環境で構築されるそれぞれの機能の関係を図 1 に示す。仮想組織内にはいくつかの gLite で作られたグリッド環境が動作している。ユーザは UI (User Interface) からジョブの実行依頼などの操作を行う。UI は仮想組織内の RB (Resource Broker) にジョブを依頼す

る。RB では BDII (Barkley Distributed Information Index) からの情報を元にジョブを実行するのに適した CE (Computing Element) を選び、そこにジョブを渡す。BDII では、グリッド環境の情報を収集し、これを公開している。CE はいくつかの WN (Worker Node) (一般的にクラスター構成) を管理しており、RB から渡されたジョブを実行可能な WN に渡す。WN ではジョブが処理される。また、ストレージ資源を使う場合は SE (Storage Element) が使用される。LFC (LCG File Catalog) はジョブが使用するグリッド上のデータファイルの統合的カタログ管理を行う。これにより、同じデータを複製し分散配置することができ、各ユーザは、自分に近いデータを利用できる。

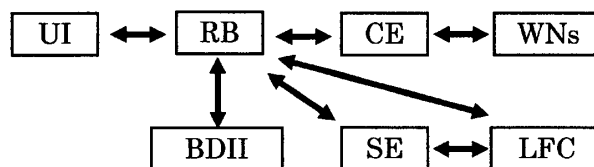


図 1. グリッドの機能

4. 自動監視・診断システム

前節で述べたグリッド環境・機能を維持するためには、ソフトウェアの更新を定期的に行い、適切な設定がなされた上で各拠点が正しくサービスを提供している必要がある。そこで、サービスの可用性を高めるため、グリッド・システムの運用を支援する自動監視・診断システムを開発し、加速器科学仮想組織上にこれを展開した。

4.1 システムの機能

グリッド環境下で自動監視・診断システムに求められる機能は以下の 5 つにまとめることが出来る。

4.1.1 監視

監視には、各マシンが正常動作しているかどうかの監視とグリッド・システムとして正しく機能しているかの監視と二つに分類することができる。

前者は、ネットワークを介して接続されている複数のマシンを監視し、各マシンで稼動している基本的なサービスが正常か否かを検知する機能である。また後者は、gLite で構成されたグリッド・システム環境そのものを監視するものである。そのためには、そのグリッド環境に合わせた機能 (コマンド) を利用する必要がある。そこで、このような理由から、本システムでは、監視機能をグリッド環境の操作をすることが出来る UI 上に展開した。

4.1.2 診断

監視しているマシンに異常を発見した場合、異常がどのようなものか診断し、データベースからその異常に対する診断情報を取得し、管理者に提供するとよいと考えられる。診断機能ではそのような機能を提供する。また、データベースに含まれない未知の異常の場合、対処した管理者がその対策をデータベースに逐次蓄えることが出来るのが望ましい。

† 広島工業大学大学院工学研究科

‡ 高エネルギー加速器研究機構計算科学センター

4.1.3 通知

通知機能は、監視しているマシンの異常をメールなどで管理者に伝える機能である。複数の管理者が登録され、異常の種類やレベルにより、情報通知の相手を自動的に選択し通知できることが望ましい。

4.1.4 記録 (アカウンティング)

障害に対しては、同じ障害を起こさないためにも、その障害が起こった過程を分析し、原因の追及、そして、対応策を練る。そのためには、さまざまな情報を記録し、分析・原因究明などに役立てていく。本機能はその役割を担う。

4.1.5 表示

システムの監視は通常、ネットワークに接続されたコンピュータで行う。また、多くのコンピュータで同時に閲覧する必要がある場合もある。そのため、ネットワーク経由で web ブラウザなどを使い表示出来ると良い。また、ユーザが複雑な操作を必要とせず、簡潔に情報を得られることが望ましい。

4.2 開発システム

本研究では上記の要件を満たすシステムの開発基盤として nagios (Nagios Ain't Gonna Insist On Sainthood Nagios) [3]を使用した。nagiosとは、システムとネットワークの総合監視オープンソースソフトウェアである。

Nagiosには監視のためのプラグインが豊富にあるため、基本的な機能の開発コストを削減できる。また、自分で機能追加を行う場合、使える言語が多彩という利点もある。

このプラグインは、nagiosが監視しているマシンに対して行う動作を決定する。また、nagiosではマシンやサービスに障害が起こった場合などに通知する機能を持つ。さらに、web インタフェースを持っており、容易に web ブラウザを使ってマシンの状態を監視することが出来る。

本システムでは、シェルスクリプトを用いて gLite の機能を監視するプラグインを作成し、グリッド・システムそのものに対する監視機能を充実させた。表示や通知は nagios の機能を用いた。また、各プラグインの動作記録のログを取り、記録機能とした。なお、動作記録をリアルタイムで使用してモニタリングも行える。

4.2.1 監視機能

本研究で開発した機能追加で自動監視することのできる項目を以下に示す。これらの追加機能は、プラグインとして認識され、実際にサービスを動かすときに nagios が読み込んで使用する。

(1) 応答監視

各マシンに ping や ssh を利用して応答を確認する。また、各マシンに必要なポートの開閉が正しく行われているかどうかもチェックする。

(2) グリッド機能監視

グリッドの環境をチェックする特有のプラグインとして、CE で gLite のジョブを正しく実行できるかをチェックする。SE ではディスク容量や、SE の情報が正しく取れているか、グリッド FTP は正常かどうかをチェックする。また、LFC ではファイルカタログの操作が正しく行えるかをチェックする。

(3) 通信監視

プラグインから得られるパフォーマンスデータを利用して、RTT の監視を行う。RTT の推移から、混雑している時間帯を避けてジョブを投入することが出来る。

(4) 内部監視

監視対象に外部から操作できないマシンの内部情報を返す機能を導入し、プラグインからマシンの CPU やメモリ、ディスク容量の使用率などを監視する。

4.2.2 診断機能

障害が発生した場合、プラグインからの出力にその障害に対する対処法 (診断情報)などを記述した wiki (診断用データベース)へ誘導するリンクを付加している。これにより既存の障害ならば迅速に対処できる。また、wiki に書かれていない障害の場合、対処を行った管理者が wiki に記述することも出来る。

5. システムの運用

Scientific Linux 3.0.8 上で nagios 3.0a4 を運用した。監視状態の表示画面の例を図 2 に示す。図 2 では、各プラグインからの監視結果や、各マシンやプラグインの設定情報を nagios がまとめて一覧表示している。

Service Name	Status	Last Update	Plugin Name
CE-01	OK	2007-03-14 14:29:13	check_ping
CE-02	OK	2007-03-14 14:29:13	check_ping
CE-03	OK	2007-03-14 14:29:13	check_ping
CE-04	OK	2007-03-14 14:29:13	check_ping
CE-05	OK	2007-03-14 14:29:13	check_ping
CE-06	OK	2007-03-14 14:29:13	check_ping
CE-07	OK	2007-03-14 14:29:13	check_ping
CE-08	OK	2007-03-14 14:29:13	check_ping
CE-09	OK	2007-03-14 14:29:13	check_ping
CE-10	OK	2007-03-14 14:29:13	check_ping
CE-11	OK	2007-03-14 14:29:13	check_ping
CE-12	OK	2007-03-14 14:29:13	check_ping
CE-13	OK	2007-03-14 14:29:13	check_ping
CE-14	OK	2007-03-14 14:29:13	check_ping
CE-15	OK	2007-03-14 14:29:13	check_ping
CE-16	OK	2007-03-14 14:29:13	check_ping
CE-17	OK	2007-03-14 14:29:13	check_ping
CE-18	OK	2007-03-14 14:29:13	check_ping
CE-19	OK	2007-03-14 14:29:13	check_ping
CE-20	OK	2007-03-14 14:29:13	check_ping

図 2. Nagios Service 表示例

本システムでは、全ての監視項目の監視を自動的に行い、異常があれば通知する。すなわち、手作業の部分を大幅にカットでき、システムの運用に割り当てる人的リソースを削減できることになる。各プラグインは一定間隔で指定されたマシンやサービスを監視し、それが正常に動作しているか否かを示す。管理者はその情報を元にマシンを管理する。システムに障害が発生した場合は、診断情報から障害を迅速に認識し的確な処置を行うことが出来る。

これらのことから、少ない人的リソースで可用性の高いシステムの運用が可能となる。

6. まとめ

本研究ではグリッド環境を自動監視し、診断を行うシステムを開発した。本システムを用いることで、障害が起こった場合にどのような障害かを診断し、必要な情報を管理者に通知、更に、wiki ページなどで障害に関する情報を提供することで、迅速な対応が行えるようになった。

今後の課題として、監視を行うことによるシステム全体に対する負荷の増大があげられる。これは、監視は任意の間隔で gLite の機能を使用して行うからである。今後は、負荷の少ない動作が行えるようプラグインを改善していくとともに、機能をさらに充実させていく。

参考文献

- [1] gLite : <http://glite.web.cern.ch/glite/>
- [2] 中田秀基, 佐藤仁, 佐賀一繁, 畑中正行, 佐伯裕治, 松岡聡 : “NAREGI ミドルウェア β -gLite 間における相互ジョブ起動実験”, 情報処理学会研究報告 : Vol.2007, No.17 pp.269-274
- [3] nagios : <http://www.nagios.org/>