

OSS 性能・信頼性評価、障害解析ツール開発プロジェクト OSS ミドル評価手順書とLinux 障害解析ツールを開発

1. 背景

サーバ分野において、OSS を適用したシステムが普及・拡大している。数年前までは、Web サーバやメールサーバ、ネームサーバといったネットワークのフロントエンドサーバへの適用が中心であったが、最近では、アプリケーションサーバ、DB サーバから構成されるエンタープライズシステムへの適用ニーズも出てきている。また、大企業や社会インフラなどで、ミッションクリティカルシステムに OSS を適用する事例も出てきた。

元来、Linux は、Apache、Sendmail 等のネットワーク系ミドルウェアの前提 OS として発展してきたので、ネットワークのフロントエンドサーバとしての利用では、性能や信頼性に関する情報やシステム構築ノウハウは多数参照可能である。しかし、アプリケーションサーバや DB サーバといった分野は、OSS にとって比較的歴史の浅い新しい分野でもあるので、エンタープライズシステムへの適用について、以下のような問題がある。

- (1) システム設計・構築のための十分な情報が無く、各ベンダで個別に評価し、システム設計・構築を実施している。評価方法についても、各社バラバラで、ノウハウが分散しており OSS を適用した際に得られるコストメリットが十分に発揮されていない。アプリケーションサーバ、DB サーバといった分野は、発展途上でもあるため進化のスピードが早く、またバージョンアップも頻繁なため、せっかく評価したデータが、次のシステム構築では役に立たないといったことがよくあり、コストがかさむ。
- (2) エンタープライズシステムでは、迅速な障害対応が求められるが、Linux にはダンプやトレースといった障害解析のための標準的なツールが無く、障害発生時は、各社固有のノウハウで対応している。障害発生時に十分なデータが得られないため、解決に時間がかかったり、原因の特定に至らないケースも多々ある。

2. 目的

上記課題を受け、本プロジェクトでは、特に「サーバ Linux、OSS の更なる普及・拡大のためのベンダサイドの課題解決」を図るものとし、ベンダ内にある信頼性データ & OSS ノウハウを集約して公開することで、ユーザが安心して OSS を選択するための指標とすることを可能にすることを目的として実施した。

具体的には、以下の通りである。

- (1) ベンダ共同の OSS の性能・信頼性評価によるシステム設計・構築ノウハウの共有
結果だけでなく、手順やデータも共有し、共通化を図る
・日本 OSS 推進フォーラム等のサイトから公開することで、OSS の普及に貢献
・多様なノウハウを活用し、信頼性の高いシステムを構築
・ベンダにおける OSS 評価コストの低減（特に Linux カーネル 2.6、AP 層、DB 層などの新分野）
- (2) 障害情報解析ツールの開発（例えばダンプ解析、トレーサ等）とノウハウの共有
・ツールの利用ノウハウのベンダ間での共有とブラッシュアップ
・必要な機能は OSS 技術開発・評価コンソーシアムで開発し、日本 OSS 推進フォーラムの

- サイト等で公開することで OSS 適用システムの障害解析の容易化に貢献
 - 障害解析時間の短縮
 - ミッションクリティカルシステムへの適用ニーズに対応

3. 開発の内容

本プロジェクトにおいて、具体的に実施した研究・開発項目は以下のとおりである。

3.1 性能・信頼性評価

性能・信頼性評価では、Java アプリケーション層、DB 層、OS 層の評価を行い、結果の比較および性能限界の原因分析を行うことによって、各層毎に性能限界、解析手法を明らかにした。

3.1.1 Java アプリケーション層の評価

(1) Linux カーネル 2.4 と 2.6 の比較

Linux カーネル 2.4 および 2.6 それぞれの OS 上に、SPECjAppServer を用いた性能・信頼性測定環境を構築し、両者の性能・信頼性測定を行うことで、カーネルのバージョンの違いが、Java アプリケーション層のレベルにおいてはどのような形で現れてくるのかという点について検証した。

評価の結果、低負荷状態では性能差が見られなかったが、高負荷状態ではカーネル 2.6 の性能が優れることがわかった。

(2) 分散処理性能の評価

本評価では、Java アプリケーション層をクラスタ化し、クラスタ台数を増加させた場合の性能・信頼性測定等を行った。これにより、OSS を採用したシステムにおける、クラスタ構成のような大規模システムの構築の是非を検討した。

1台から4台の構成で測定した結果、クラスタ台数を増やすことで、より高負荷に耐えられるシステムを構築できることがわかった。

(3) トランザクション特性の解析

SPECjAppServer は、典型的なエンタープライズ Java アプリケーションを総合的に模した仕様となっているため、逆にミクロな視点である Java オブジェクトレベルでの処理性能解析やトランザクションレベルでの性能解析において弱い部分がある。

本評価では、SPECjAppServer をベースに、メソッドレベルの実行回数、時間を測定できる新たな測定ツール (EJB profiler) を開発することにより、ボトルネックの解析を試み、ツールの有効性を確認できた。

(4) 商用ソフトウェアとの比較

アプリケーションサーバの商用ソフトウェアとして WebLogic を選択し、これと JBoss との比較評価を行った。JBoss に比べて、WebLogic は高負荷時でも安定して動作することがわかった。

3.1.2 DB 層の評価

OSDL で公開されている DB 性能評価ツール DBT-1 と DBT-3 をベースとした共通性能評価手順を作成した。

(1) Web システムのためのデータベース性能評価の共通手順作成および評価

DBT-1 を利用した共通手順を作成し、DBT-1 + PostgreSQL、MaxDB での性能評価を実施した。またディストリビューションを変えて、Linux カーネルの違いに関する評価も実施した。

評価の結果、Java アプリケーション層評価と同様に、低負荷状態では性能差が見られなかったが、高負荷状態ではカーネル 2.6 の性能が優れることがわかった。

- (2) DSS(意思決定支援システム)のためのデータベース性能評価の共通手順作成および評価
DBT-3 を利用した共通手順の作成、DBT-3 + PostgreSQL での性能評価を実施した。
評価の結果、PostgreSQL 8.0 は 7.4 に比べ、10%ほど性能が向上していることがわかった。

(3) 大規模データベース環境における性能検証

バックアップとリカバリ 大容量データのロード、インデックスの再構築に要する時間の性能評価、および PostgreSQL 8.0 の Point In Time Recovery(PITR)を使用し、クラッシュリカバリに要する時間の性能評価も測定した。

評価の結果、PITR を使用したリカバリは、バックアップからのリカバリに比べ 1/3 程度の時間でリカバリできることが分かった。

(4) DBT-1 のスコープ拡大のための Oracle 向け改変作業及び MySQL 向け改変調査作業

現状は、SAP-DB 及び PostgreSQL のストアプロシジャで実装されている DBT-1 の Oracle9i 向け移植を行い、より広範囲の評価モデルとして利用できるようにした。

3.1.3 OS 層の評価

OS 層の評価では、サイジングに役立つツールの整備とツール利用方法の確立を行った。

(1) DBMS を模したマイクロ性能評価ツールを用いた評価

Oprofile 等によるカーネル内部プロファイリングを実施した。また、DBMS のファイル I/O およびメモリコピーに近似するマイクロ性能評価ツール(diskio2.0)を開発し、ファイル読み書き、ユーザ空間 カーネル空間のメモリコピーにおける Linux カーネルの特性を測定した。

また、DB 層の DBT-3 評価との比較によって相関関係を確認し、diskio の評価結果の実システムにおけるサイジングへの利用可能性を考察した。

(2) OS の限界値付近での動作評価

大規模メモリ搭載(主メモリ4GB以上)の動作評価を実施した。D など高負荷状態にして OS 限界値付近での動作評価をおこなった。カーネル内部プロファイリングの実施により、性能限界を決めるカーネル内部コンポーネントの特定手法を確立し、CPU ネット、I/O ネット、ロックネットの場合について、それぞれ実際の解析を試みた。

3.2 OSS 適用システムの障害解析ツール開発

OSS 適用システムの障害解析ツールとして、ダンプ解析ツール(ALICIA)、LKST によるカーネル性能評価ツール、ディスク割り当て評価ツール(DAV)を開発した。

3.2.1 ダンプデータ解析ツールの開発

ダンプ解析の高速化・省力化を実現する手段としてダンプ解析手順をスクリプト化できる機能を開発した。本機能は、既存解析ツールに対する統合インタフェース(今回は crash インタフェースのみ)として位置付けられる機能であり、ダンプ解析手順を簡単にスクリプト化でき、ダンプ編集の高速化、解析手順の蓄積・再利用を可能にする機能である。また、ダンプ解析スクリプトの作成を支援するためのサブルーチン群も含まれる。本機能を利用してダンプ解析スクリプトを開発・蓄積することにより、ダンプ解析ノウハウの共有化機能としても利用可能とした。

3.2.2 LKST(Linux Kernel State Tracer)によるカーネル性能評価ツールの開発

アプリケーション特有の処理や過負荷状態で発生する性能低下を分析する仕組みを開発

した。また、アプリケーション実行時に生じている性能ボトルネックの解析に必要なカーネル内の処理 (システムコール、スケジューラ、メモリ管理システム、ソフトウェア割り込み、ファイルシステム処理、ロック処理) に関して性能測定を可能にした。

3.2.3 ディスク割り当て評価ツールの開発

ファイルシステム全体、および特定ファイル毎にディスク上におけるブロック割り当ての状況を表示するツールを開発し、ファイルアクセスの性能低下が検知された際の原因分析手法を確立した。

4. 従来の技術 (または機能) との相違

性能・信頼性評価については、OSS には、これまで明確な評価指標が存在しなかった。また、障害解析ツールも今回開発した機能 (ダンプにおけるスクリプト機能、トレーサにおけるきめ細かなカーネルトレース情報の取得、ディスク割当評価ツールにおけるフラグメンテーション表示機能、等) は存在していなかった。

5. 期待される効果

OSS は進化していくものである。今回作成した評価手順に従って評価をしていくことで、今後の進化が定量的に評価できる。また、障害解析ツールにより、障害解析が容易化できるとともに、Linux の高信頼化にも貢献できる。

6. 普及 (または活用) の見通し

開発成果は、OSS 推進フォーラムの Web ページから公開した。北東アジア OSS 推進フォーラムの日中韓合同技術開発評価 WG などを通じて北東アジア地域、あるいは OSDL を通じて世界のコミュニティとブラッシュアップしていく予定である。

7. 開発者名 (所属) (各社代表のみ記述)

(1) 幹事会社

- ・ 鈴木友峰 (株) 日立製作所 ソフトウェア事業部 大型プロジェクト推進室 担当部長)

(2) 参加企業

- ・ 石井達夫 (株) SRA 開発サービスカンパニー コンサルティング部 OSS ビジネスプロジェクト長)
- ・ 三浦広志 (株) NTT データ オープンソース開発センター シニアエキスパート)
- ・ 澤藤宗彦 (新日鉄ソリューションズ (株) IT インフラソリューション事業部 マーケティング部 グループリーダー)
- ・ 菅原孝二 (住商情報システム (株) 生産技術センター 副センター長)
- ・ 西片公一 (株) 野村総合研究所 情報技術本部 システム技術一部 オープンソースソリューションセンター グループマネージャ)
- ・ 吉岡弘隆 (ミラクル・リナックス (株) 取締役)
- ・ 高橋秀樹 (ユニアデックス (株) プロダクト事業グループ ソフトウェアプロダクト統括部 第一グループマネージャ)

(参考) 成果物 URL

<http://www.ipa.go.jp/software/open/forum/DevInfraWG.html>