

WebSAMの分析技術と応用例 ～インバリアント分析の特長と適用領域～

加藤 清志・矢吹 謙太郎

要 旨

統合運用管理ソフトウェアWebSAMでは、対象システムからさまざまな情報を収集して、俯瞰的にその状態を分析し、管理者通知や対処を実行します。このような分析の処理フローは、ビッグデータに求められる機能とも類似するものです。このWebSAMの性能分析製品 Invariant Analyzerでは、大量の性能データの相関関係を自動抽出し、隠れた障害の兆候を発見できます。本稿では、WebSAMの分析技術とビッグデータとの親和性について考察し、インバリアント分析技術の運用管理分野以外の領域への展開例を紹介します。

キーワード

●クラウド ●運用管理 ●WebSAM ●インバリアント分析技術 ●ビッグデータ ●機械学習

1. はじめに

近年、システムを安定稼働させるための運用管理システムの重要性が高まり、発生した障害を復旧するだけでなく、顕在化していない障害の予兆を検知し、サービスを止めずに回避する機能も求められています。WebSAMは、このような要求に対応し、大規模複雑化したシステムの運用をシンプル化する統合運用管理ソフトウェアです¹⁾。対象システムからさまざまな情報を収集し、俯瞰的にその状態を分析し、管理者通知や対処を実行します。

このような分析の処理フローは、ビッグデータに求められる機能とも類似しています。システム運用管理では、性能データやセンサデータなどの高頻度/大量の数値データをリアルタイムに分析することも必要ですが、WebSAMが採用するインバリアント分析技術は、このようなデータの相関関係を自動発見し、隠れた異常を見つけ出すことができます。

2. 分析基盤としてのWebSAM

2.1 運用管理分野の要件

システム運用管理では、基幹システムなどの重要なサーバ群を扱うため、管理者には大きな責任が伴います。もし対象システムが正常にサービス提供できなくなれば、サービスの事業者や利用者に大きな損害を与える可能性があります。このため、運用管理ソフトウェアにも可用性や信頼性などの高

いレベルの要件が課されますが、すべての要件をツールに委ねるわけではありません。現行のツールで解決できない事象については、人間側の管理体制で解決を図ることになります。

例えば、障害検知では、対象システムにさまざまな障害が発生しても、監視機能自体はダウンすることなく、常に確実に情報を収集することが求められます。しかし、未知障害のために高度な判断が必要になった場合などは、即座に管理者に正確な情報を提供し、既存の対処シナリオを適用するのか、人間側で対処するのかを判断してもらうことになります。

このような関係プレーを実現するには、システムと人間とのユーザーインターフェースも重要な要素です。誤認による二次障害を発生させないように、適切なタイミングで、適切な人に、ポイントを要約して伝えることも、ツールに求められる要件です。

2.2 WebSAMフレームワーク

システム運用管理では、さまざまな分野のシステムが対象になり、それぞれの分野でクリティカルになる要件が異なるため、監視項目や分析方法もそれに合わせてカスタマイズする必要があります。

WebSAMでは、前述した要件を満たしつつ、運用管理に必要な機能をプラグインとして組み替えられる共通フレームワークを提供しています(図1)。これにより、WebSAMでは、共通機能を含めた全体が監視、分析、制御の処理フローを実現する分析基盤として機能します。

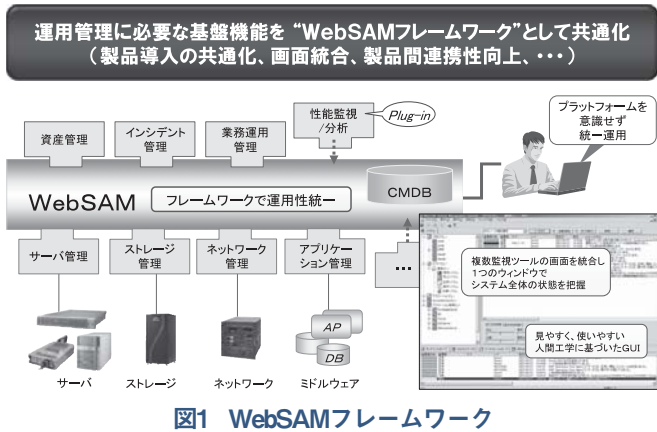


図1 WebSAMフレームワーク

1) 共通機能

- ・ 監視エージェントからマネージャーへの通信機能
- ・ メッセージ分類や通報などのメッセージ管理機能
- ・ ログ監視や性能閾値監視などの共通監視機能
- ・ 性能情報や構成情報などの共通データベース
- ・ 運用管理に共通な対話画面

2) プラグイン機能

- ・ サーバ、アプリケーションなどの監視機能
- ・ インバリエント分析などの分析エンジン
- ・ ワークフローなどの制御機能
- ・ 各機能特有の対話画面
- ・ 他社ツールとの連携機能

例えば、インバリエント分析エンジンは、性能データの相関分析を行います。入力となる情報は各監視製品が収集し、共通機能を介してエンジンに受け渡されます。分析結果は共通画面に組み込まれて表示されるとともに、メッセージとして共通機能に受け渡され、他の異常メッセージと組み合わせた総合的な判断が行われます。また、共通機能を介して対象分野に応じた他社ツールとの連携も可能になります。

3. ビッグデータとWebSAM

3.1 ビッグデータの特徴

ビッグデータの技術的な定義は明確ではありませんが、一般に、以下の3要素のいずれかの特徴を持つものを指します。このような情報は、既存の計算機でそのまま扱うことが困難

なため、その収集、蓄積、分析、可視化のそれぞれの処理で特別な技術を必要とします。

1) 大量

データのサイズ自体が大きいこと。絶対的なサイズだけでなく、処理する計算機のメモリやI/O性能に対して相対的に大きいデータでは、分散処理などの特別な技術が必要とする。

2) 多種

データの構造が定まっていないこと。種類が多いことだけでなく、取り扱いに適した構造化がなされていない生データ（非構造文書やノイズを含む数値など）から必要な情報を抽出する困難さを含む。

3) 高速

データの発生頻度が高いこと。例えば、大量機器からのセンサ情報など、短い時間間隔で発生する時系列データから異常検知を行う場合、極めて短い時間間隔で多数のデータ間の相関分析を完了させることになる。

3.2 共通点：分析処理としての類似性

ビッグデータが注目される大きな理由は、これまで活用できていなかった情報を分析することで、何らかの新しい知見が得られる可能性があることです。例えば、構造化されていない大量の文書から何らかの傾向を見つけ出してビジネスの戦略に生かすことや、位置情報や電力使用量のような大量の時系列の数値データから共通点や特異点を見つけ出して危険を予測することなどです。そのためには、高速データベースや分散処理といった個別の技術だけでなく、収集、蓄積、分析の一連の処理を実現する必要があります。

前述したように、WebSAMは、このようなデータの一元管理、統合分析を指向した一連の分析処理を行う仕組みを提供します。現時点ではビッグデータに適用できる十分な性能が保証されているわけではありませんが、技術的な類似性は高いものになります。実際、大規模データセンターでは、大量のデータを扱う必要があり、スケーラビリティも重要な要件となっています。

また、ビッグデータの分析では、非構造データからの情報抽出が必要なため、対象分野の知識を持つドメインエキスパートや、分析手法の知識を持つ分析エキスパートが必要とされています。ツールによる自動化だけでなく、その目的達

成に人間との関係プレーが不可欠な点でも、運用管理ソフトウェアとの類似性は高いと言えます。

3.3 相違点：運用管理分野との要件の違い

ビッグデータの適用領域では、運用管理分野の要件に合わせて開発されたWebSAMとは、以下のような相違点が出てくる可能性があります。

(1) 扱うデータのサイズ

WebSAMでもデータの一元管理のためのスケーラビリティは確保していますが、広範な利用が想定されるビッグデータでは、オーダの違うほど巨大なデータの処理が求められる可能性があります。同じアーキテクチャのコンピュータを利用する限りは、処理できるデータサイズの限界も同等になりますが、ビッグデータ領域ではまったく異なるアーキテクチャとなる可能性もあります。

(2) リアルタイム性と網羅性

運用管理分野のメッセージ管理は、必要な情報を選別して分析する仕組みです。ビッグデータの領域であっても、リアルタイム性が要求される用途では、このような情報の絞り込みは不可避ですが、リアルタイム性よりも網羅性を優先する用途では、絞り込みを行わない分析手法が採用される可能性があります。WebSAMのインバリエント分析にもオフライン分析の機能はありますが、より大量の履歴データを網羅的に分析する手法などです。

(3) 推論機能と確実性

運用管理分野では、管理者をミスリードする情報は、致命的な二次障害を引き起こす可能性があるため敬遠されます。インバリエント分析でも、事実としての相関関係のディスカバリとエビデンスの提示は行いますが、確実性が低下する推論は行いません。推論に基づいて対処を行った後に、実は正解は違うものだったという状況になることを避けるためです。しかし、ビッグデータの領域では、確実性よりも知的な分析結果が優先され、仮説に基づく推論の機能が採用される可能性があります。

4. インバリエント分析技術

4.1 WebSAM Invariant Analyzer

WebSAM Invariant Analyzerは、システムの性能モデルを自

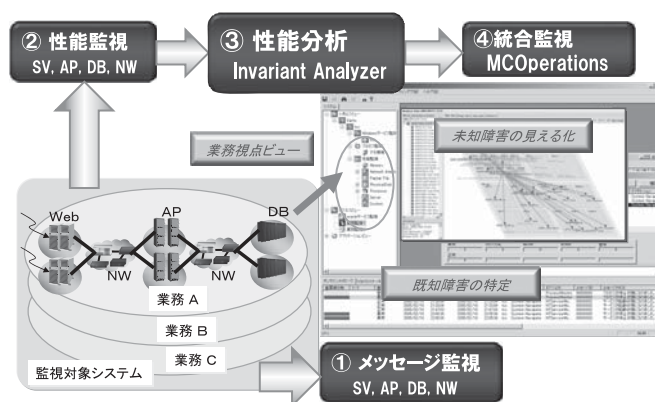


図2 システム性能分析の概要

動抽出し、従来の分析方式では見つけづかったサイレント障害を発見する性能分析製品です（図2）。時系列の数値データを分析対象とし、正常な期間の数値データから性能モデルを学習し、リアルタイムに得られる数値データから異常を発見する、機械学習の手法を採用しています。

インバリエント分析の手法を採用することで、性能分析における監視設定の手間を大幅に軽減できます。異常の発見は、性能モデルとして抽出した相関関係が維持されているかどうかで判断するため、性能データごとの特性に応じて閾値を設定する必要がありません。また、性能データ間の値が相関関係に従って変化していれば、その値の大小によらず正常と判定するので、平常時と繁忙日のような負荷の違いがあっても、同じ設定で利用できます。

4.2 機械学習手法の特徴

数値データから特徴を抽出する機械学習手法をおおまかに分類すると、数学モデル型（ダイナミックスレッシュホールド、システム同定など）と汎用知識型（ルールリファイン、事例ベース推論など）があります。

数学モデル型では、与えられたデータの特徴を数学的な式として学習します。この式に新たなデータを代入することで異常判定を行います。このため、数学関数で表現できないものは学習できませんが、異常判定は算術演算だけなので非常に高速です。

一方、汎用知識型では、さまざまな特徴をif-thenルールのような知識表現で学習します。得られたデータがルールに合致

しているかどうかで異常判定します。このため、さまざまな知識を学習できますが、異常判定では比較的重いマッチング処理をルールの数だけ繰り返すことになります。

インバリエント分析は、数学モデル型のシステム同定手法になります。異常判定が高速なため、リアルタイム分析に適しています。また、ダイナミックスレッシュOLDのように周期性などの稼働パターンを学習するものではないため、対象領域やデータ発生元の機器によらず、さまざまな数値データ間の相関を分析することができるという特徴があります。

5. インバリエント分析技術の他領域展開

前述したように、インバリエント分析技術は、数値データであれば、運用管理分野以外の分析にも適用できる技術です。ここでは、現在進めている他領域への展開例を紹介します。

5.1 システム運用管理との融合領域

(1) 電力マネジメント

2011年の震災以降、ピークカット対策やHEMS、BEMSなどが注目されています。外気冷却を採用したデータセンターでは、温度と電力のセンサデータから空調を制御する用途があります。インバリエント分析でセンサデータ間の相関抽出やいつもと違う利用を検知できます。

(2) ソフトウェア開発

運用と開発の一体化（DevOps）という考え方から、運用時に得られた情報を設計にフィードバックするものです。扱うデータはシステム運用と同様ですが、詳細度と多様性が飛躍的に大きくなります。インバリエント分析は、性能データの単位で異常部位を特定できます。

(3) ビジネスインパクト解析

経営指標や金融取引などの情報から特異点を抽出し、リスク要因の特定や不正取引の発見を行います。例えば、情報システムの投資対効果を見る場合、金額の情報だけでなく、稼働状況などの運用データとも突き合わせる必要があり、さまざまな種類のデータを合わせたクロスドメイン分析ができるインバリエント分析が有効です。

5.2 物理システムの領域

(1) 構造物劣化診断

設計データに基づくシミュレーションを行う診断手法は以前からありますが、建造物の個体差が大きい場合は、性能モデルを自動抽出できるインバリエント分析が有効です。例えば、橋梁診断では、振動センサの検出値の変化から劣化を発見し、必要性に応じた効率的なメンテナンスを行うことができます²⁾。

(2) 製造プラントでの異常検知

メンテナンスの効率化のため、センサの情報から製造物または製造装置の故障検知を行います。複雑なプラントでは、依存関係がすべて把握できていない場合もあり、専門家向けに相関関係のディスカバリも求められています。

(3) テレマティクス

自動車は製造工程を含め、最も先進の技術が集約されている領域です。ナビゲーションシステム以外にも多数のマイコンが設置されており、センサ情報の相関分析による故障の予測など、それらの情報を有効活用する方法が求められています。

6. おわりに

以上、説明したように、WebSAMの分析機能はビッグデータと親和性があります。現状のWebSAMはシステム運用の効率化を実現するものですが、今後、インバリエント分析技術の領域展開などの取り組みを進め、ビッグデータを含むさまざまな分野のシステムの効率運用に貢献していきます。

参考文献

- 1) 加藤清志ほか：“WebSAM Ver.8が実現するクラウド時代のデータセンター運用,” NEC技報, Vol.63, No.2, pp.75-78, 2010.4
- 2) 金哲佑ほか：“模型桁車両走行実験における異常診断の可能性の検討,” 応用力学論文集, 土木学会, Vol.14, pp.1.833-1.842, 2011.8

執筆者プロフィール

加藤 清志
ITソフトウェア事業本部
第2ITソフトウェア事業部
マネージャー

矢吹 謙太郎
ITソフトウェア事業本部
第2ITソフトウェア事業部
マネージャー

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報(日本語)

NEC Technical Journal(英語)

Vol.65 No.2 ビッグデータ活用を支える 基盤技術・ソリューション特集

ビッグデータ活用を支える基盤技術・ソリューション特集よせて
ビッグデータを価値に変えるNECのITインフラ

◇ 特集論文

データ管理/処理基盤

超高速データ分析プラットフォーム [InfoFrame DWH Appliance]
SDN 技術で通信フローを制御する [UNIVERGE PF シリーズ]
大量データをリアルタイムに処理する [InfoFrame Table Access Method]
大量データを高速に処理する [InfoFrame DataBooster]
ビッグデータの活用最適なスケールアウト型新データベース [InfoFrame Relational Store]
高い信頼性と拡張性を実現した Express5800/ スケーラブル HA サーバ
大規模データ処理に対する OSS Hadoop の活用
大容量・高信頼グリッドストレージ iStorage HS シリーズ (HYDRAStor)

データ分析基盤

ファイルサーバのデータ整理・活用を支援する [Information Assessment System]
超大規模バイオメトリック認証システムとその実現
WebSAMの分析技術と応用例～インバリエント分析の特長と適用領域～

データ収集基盤

スマートな社会を実現する M2M とビッグデータ
微小な振動を検知する超高感度振動センサ技術開発とその応用

ビッグデータ処理を支える先進技術

多次元範囲検索を可能とするキーバリューストア [MD-HBase]
高倍率・高精細を実現する事例ベースの学習型超解像方式
ビッグデータ活用のためのテキスト分析技術
ビッグデータ時代の最先端データマイニング
ジオタグ付きデータをクラウドでスケラブルに処理するジオフェンシングシステム
柔軟性と高性能を備えたビッグデータ・ストリーム分析プラットフォーム [Blockmon] とその使用事例

◇ 普通論文

地デジ TV を活用した「まちづくりコミュニティ形成支援システム」

◇ NEC Information

NEWS

スケールアウト型新データベース [InfoFrame Relational Store] が 2 つの賞を受賞



Vol.65 No.2
(2012年9月)

特集TOP