MasterScope: Features and Experimental Applications of System Invariant Analysis Technology

KATO Kiyoshi, YABUKI Kentaro

Abstract

NEC's MasterScope middleware is an integrated operation management software suite. MasterScope collects operational and performance metrics from target IT systems and analyzes them comprehensively in order to detect and locate system failures. Detected events and failures will be notified to the operator and workarounds can be applied to recover from failures. There are similarities between such an analysis process for the operation management and that required to process big data. For Instance, the system performance analysis software "MasterScope Invariant Analyzer" automatically discovers important correlations from a large amount of performance data and proactively detects hidden performance anomalies, thereby avoiding serious system level damages. This paper describes the analysis technology of MasterScope which has similarity to the big data analysis technology, and then introduces experimental applications of the system invariant analysis technology in domains other than the operation management.

Keywords

cloud computing, operation management, MasterScope
system invariant analysis technology, big data, machine learning

1. Introduction

According to the increasing importance of maintaining the service level of today's IT systems, the operation management system is required to provide features not only to recover from system failures but also to detect hidden performance anomalies, so-called silent failures, so that the anomalies can be taken care of before they start causing serious damage to IT systems. MasterScope is an integrated operation management software suite that aims at simplifying the operation management of large-scale and complex IT systems [1]. It collects operational and performance metrics from target IT systems and analyzes them comprehensively in order to detect and locate system failures. Then, it notifies to the operator and applies workarounds to recover them.

There is a similarity between such an analysis process for operation management and that required to process big data. The operation management system also needs to analyze such large volume and frequently generated performance data as numerical values from servers, network devices, environment sensors, etc. Therefore, MasterScope employs the System Invariant Analysis Technology (SIAT) that automatically extracts important correlations from such large amounts of data and detects hidden system failures.

2. Analysis Platform of MasterScope

2.1 Requirements of Operation Management

Operators of the operation management system must assume a heavy responsibility because they operate important enterprise servers or network devices such as mission critical systems. Once the target system fails to provide its services appropriately, both service providers and users will suffer a heavy loss in their businesses. Therefore the operation management system must offer high availability and reliability. However, operators never entrust all requirements to the management products that are currently available. Operators have to solve incidents by themselves, if management products cannot deal with them.

For instance, in case of fault detection, the monitoring tool must collect effective information without interruption, even if various faults occur in the target system. Moreover, if a previously unrecognized fault occurs, the operation management system must provide accurate information to operators immediately. Then, operators take over the recovery process and make decisions whether applying existing workarounds or trying recovery operations manually.

The user interface (UI) is also important in achieving such cooperation between the system and the operator. It is required that the operation management system must report by properly summarizing information at the proper time to the proper person in order to prevent operator misunderstandings that can lead to secondary failure.

2.2 MasterScope Framework

The operation management system needs to be applicable in various kinds of business domains, and its monitoring and analyzing settings must be customizable because of the difference of requirements critical in each domain.

MasterScope provides the Unified Management Framework (UMF), a common software library enabling plug-in installation for operation management products (**Fig. 1**). Using this framework, the entire MasterScope system, each products and the UMF, will work as an analysis platform that forms an operation flow with monitoring, analyzing and control features.

(1) Common features

- Communication functions from monitoring agent modules to manager modules.
- Message management functions such as filtering, e-mail notification, etc.
- Common monitoring functions for log files, performance value with threshold, etc.
- Common management databases for performance value, configuration, etc.
- Graphical user interface (GUI) for common features.

(2) Plug-in features

- Monitoring products for servers, application, etc.
- Analysis products such as invariant analysis, etc.
- Optional features such as workflow, etc.
- Specific GUI for each product.
- Collaborative interface to other vendor's products.

For instance, in case of correlation analysis of performance data by the invariant analysis engine, monitoring products acquire performance data from the target system as input, and the engine aggregates them through the UMF. Then, Analysis results of the engine are integrated into the common GUI window of the UMF. These results are also sent as information

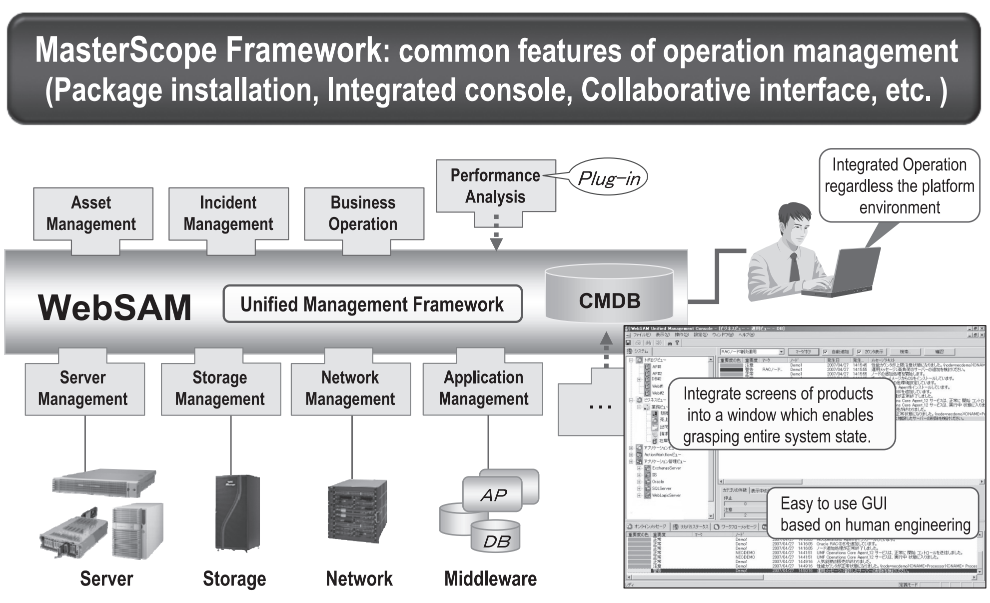

Fig. 1 MasterScope framework.

messages through the UMF, which will be judged comprehensively by the message correlation analysis feature with other error messages. And furthermore, the engine will be able to collaborate with the other vender's product through the UMF.

3. MasterScope and Big Data

3.1 Definition of Big Data

Though the technical definition of "big data" is something unclear, it may have at least one of the following three distinctive features [2]. Big data usually requires special technologies in order to capture, store, analyze and visualize it, because it includes data sets of a size beyond the processing ability of commonly-used computer systems.

1) **High-volume**
   Data size is very big. Special technologies, such as distributed processing, are required to process it from the aspect not only that the absolute data size is big but also that the relative data size is big with regard to the memory or I/O performance of the computer.
2) **High-variety**
   Data structure is not limited. It is difficult not only to recognize a wide variety of structures but also to extract effective data from unstructured raw data, such as unstructured documents or noisy numerical data.
3) **High-velocity**
   Data occurrence frequency is very high. For instance, in cases of detecting anomalies in time-series data with a short time interval, such as sensor data from a large amount of devices, correlation analysis of large amounts of data must be completed in very short intervals.

3.2 Similarity as Analysis Platform

Big data technology is remarkable for its possibility of discovering new kinds of knowledge from information that has not hitherto been utilized appropriately. For Instance, it may be useful for planning business strategies to discover trends in a large volume of unstructured documents. It may enable the hazard prediction to discover trends and anomalies in a large amount of numeric time-series data such as location information or electrical power consumption data. To achieve this, a total operation framework composed of aggregation, storage and analysis technologies is required as well as individual technologies such as a high-speed data base or distributed processing technologies, etc.

As described above, MasterScope provides an analysis framework that aims to unify data management and analysis. Though it does not yet ensure a sufficient performance to process big data, it uses a similar technology to that of big data. Actually, scalability is an important requirement of the operation management system because there is a large amount of monitoring data in a large-scale data center.

In addition, it is said that big data processing needs support from appropriate experts, the domain expert knows the target system and the analysis expert knows the analysis method, when extracting effective knowledge from unstructured information. Therefore, to succeed in analysis, collaborative work between operators and the system is indispensable as well as the automation features of the software tools. This point also indicates a similarity between the technology employed by the operation management system and that of big data.

3.3 Difference in Requirements of the Target Domain

Because MasterScope was developed to satisfy the requirements of the operation management domain, the following different design concept elements might be found in the big data application domain.

(1) **Maximum data size**
   MasterScope achieves scalability for unifying data management in the large-scale data center. However, there are some cases where an extraordinary degree of processing ability is needed in the big data application domain. Of course, the limitation of the data size to be processed will be the same as long as the same computing architecture is used. However, a completely different architecture may be employed in the big data domain.

(2) **Real-time processing vs. completeness**
   Message management features in the operation management domain adopt the filtering approach to analyze data. It is essential in real-time processing to narrow down information. However, a technique without filtering information may be employed in a case where completeness of analysis is preferable rather than the real-time performance. For instance, big data applications will require an analysis technique that allows a much larger volume of historical data, while the MasterScope's invariant analysis engine also provides offline analysis features.

(3) **Inference vs. accuracy**
   Notifications that mislead the system operator must be

avoided in the operation management domain, because it may cause fatal secondary faults. Though the invariant analysis engine discovers correlations and shows them as a fact and provides the evidence, it does not employ the inference technology, which decreases the accuracy of the result. Because we must avoid the situation in which operators find that the result of inference is wrong, after the workaround according to it has been done. However, big data applications that prefer intelligent analysis rather than an accurate one may employ the inference technology based on the hypothesizing and evaluating approach.

4. System Invariant Analysis Technology

4.1 MasterScope Invariant Analyzer

MasterScope Invariant Analyzer [3] is a system performance analysis product that automatically extracts performance models of the system and detects "silent failure" that used to be difficult to find with a conventional analysis technique (**Fig. 2**). It employs a machine learning technology that analyzes time-series numerical data, learns normal system behavior as a performance model from a period of performance data in which the system works normally, and detects anomalies from numerical data collected in real time.

The system invariant analysis technology can reduce the operator's burdensome efforts for performance monitoring setting. Operators have no need to set different thresholds depending on the properties of each performance data, because anomalies are automatically detected by analyzing whether the correlations in the performance model are valid or not. Moreover, the same model can be used even if the system workload is different, such as between a busy day and an ordinary day. This is because the analysis engine regards the system as normal behavior as long as the correlations in the model are valid, regardless of the value of performance data.

4.2 Types of Machine Learning Technology

Machine learning technologies, which can be applied to numerical data can be roughly classified into types. The first is the mathematical model type involving dynamic threshold and system identification etc. And the second is the general knowledge type involving production rule refining and case-based reasoning, etc.

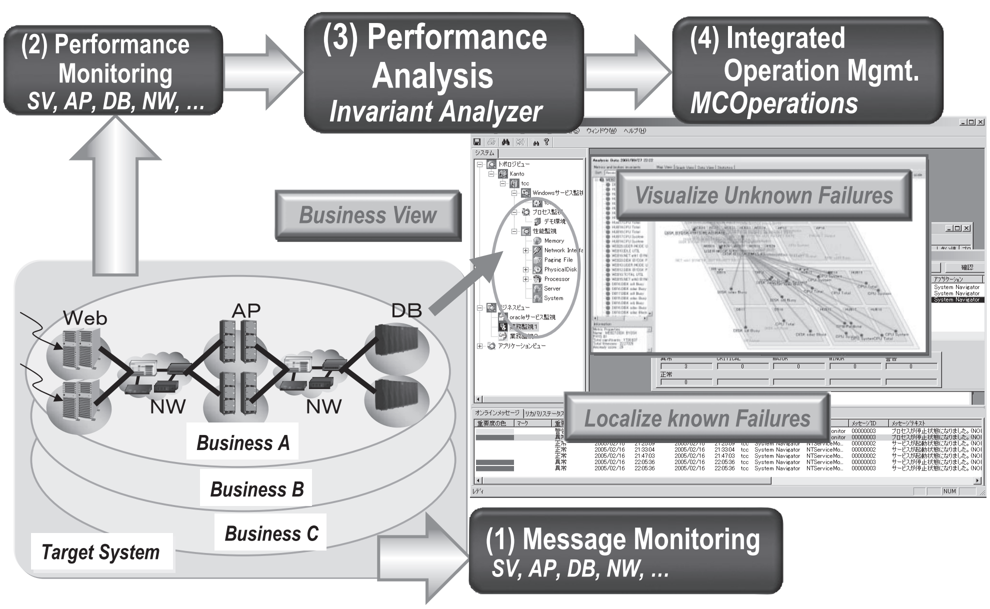

Fig. 2 Outline of system performance analysis.

The mathematical model type handles knowledge which can be expressed as an algebraic equations. Anomalies can be detected by assigning actual values to the equations to see the difference between the model's expected value and the actual data obtained. The kind of knowledge which this type can handle will be limited to class of knowledge which can be expressed in mathematical equation, but computational cost of detecting anomaly is very low, since it only requires arithmetic calculations.

The general knowledge type acquires various kinds of knowledge as if-then type production rules. Anomaly detection is made by evaluating the rule condition as true or not with the acquired data. Though it is applicable to various kinds of knowledge, the processing speed of anomaly detection is rather slow because the matching procedure has to be repeated for each rule.

The system invariant analysis technology comes into the category of system identification in the mathematical model type. It is effective in real-time analysis because its anomaly detection is very fast. Moreover, it is different from other techniques such as the dynamic threshold that learns the periodical properties of peculiar systems. Therefore, it can be widely applicable to analyze correlations of various kinds of numeric data regardless of target domains or devices generating data.

5. Invariant Analysis for Other Domains

As described in the previous section, the system invariant analysis technology is not bound only to system operation management domain, but it can also be applied to other domains which analyzes numerical data. The following experimental applications for other domains are now in progress.

5.1 Domains Related to System Operation Management

(1) Electric power management
Since the Tohoku earthquake in 2011, HEMS, BEMS and the peak shift control program are receiving attention. In a data center with an outdoor air cooling system, it is required to control air conditioners according to values acquired from temperature and electric power sensors. It can be expected that the invariant analysis extracts correlations from sensor data and discovers "unusual" usage patterns.

(2) Software development
According to "DevOps," which is a response to the interdependence of software development and IT operations, information acquired by the operation management system will be fed back to the design process. Though the kind of data is the same as that for the system operation, it is required to analyze widely diverse and detailed information. The invariant analysis allows identifying malfunctions of the granularity of performance data.

(3) Business impact analysis
It is required to identify risk factors or detect acts of fraud by extracting singular points from such information as business indexes or financial transactions. For instance, regarding the ROI (return on investment) for IT systems, it is necessary to compare cost information with performance data in operations. Invariant analysis is effective because it allows cross-domain analysis, an integrated analysis using various kinds of numeric data.

5.2 Physical System Domain

(1) Structural health monitoring
Though there are some simulation techniques for structural diagnostics using design data from before, the invariant analysis that can automatically extract a performance model is more suitable in a case when there is a wide difference in the design of each structure. For instance, in the case of condition monitoring of bridges, it enables timely maintenance as required by detecting anomaly values from vibration sensors and localizing the degradation of a part of the bridge.

(2) Failure detection at manufacturing plants
Providing effective maintenance, it is required to detect failures of both products and production lines by analyzing information from built-in sensors. Regarding complex plants, it is also required to point out discovered correlations to the expert of such plants because they can hardly be expected to understand all behavior related to such large numbers of components.

(3) Telematics
The application domain related to automobiles and their factories uses most integrated and advanced technologies. A large number of microcomputers and built-in sensors are embedded in devices such as Sat Nav systems. Therefore, technologies utilizing information from them are required, such as the correlation analysis technology

for sensor data to enable the failure detection of automobiles.

6. Conclusion

As described in this paper, there are similarities between the analysis features of MasterScope and those of big data. Currently, MasterScope's primary application domain is targeted for effective system operation management of IT systems. We intend to expand application of MasterScope's system invariant analysis technology, which will contribute to efficient analysis of big data sets used in various domains.

References

1) "MasterScope"
http://www.nec.com/en/global/prod/masterscope/
2) "Big data"
http://en.wikipedia.org/wiki/Big_data
3) "Invariant Analyzer"
http://www.nec.com/en/global/prod/masterscope/invariantanalyzer/

Authors' Profiles

KATO Kiyoshi
Manager
2nd IT Software Division
IT Software Operations Unit

YABUKI Kentaro
Manager
2nd IT Software Division
IT Software Operations Unit

Information about the NEC Technical Journal

Thank you for reading the paper.
If you are interested in the NEC Technical Journal, you can also read other papers on our website.

Link to NEC Technical Journal website

Japanese

English

Vol.7 No.2 Big Data

Vol.7 No.2

September, 2012

Special Issue TOP