

Big Data Analytics in the Cloud - System Invariant Analysis Technology Pierces the Anomaly -

KATO Kiyoshi, MIYAZAKI Hiroyuki, SOMA Tomoya, TAKAGI Mayumi

Abstract

The system invariant analysis technology automatically creates behavior models from various systems by extracting the relationships which are invariant in normal system operation. It enables the early detection of a system's anomaly behavior. It helps to improve availability of information and communication technology (ICT) systems such as the cloud computing systems. Moreover, it also helps to optimize the maintenance cost by enabling failure prediction of electric power plants, structural health monitoring of bridges, etc. and improves safety and reliability of the systems performing as social infrastructures. This paper outlines the system as a big data analysis technology and considers analysis applications in the cloud.

Keywords



System invariant analysis technology, big data, cloud computing, failure prediction, machine learning

1. Introduction

Following the recent dissemination of innovative concepts such as cloud computing, ubiquitous computing and Internet of Things (IoT), an increasing number of ICT systems aggregates a large amount of data from various devices, stores and shares them in data warehouses (DWH) for later analysis. The large amount of detailed information of today's complex society is expected to be utilized widely by using the big data analysis technologies to reveal its characteristics with machine learning techniques and by the cloud computing environment providing an effective analysis platform. Pioneering applications such as the investment in artificial intelligence in the United States and the industry-government-university action plan called Industry 4.0 in Germany have provided useful references for Japan to advance its own utilization of information.

The system invariant analysis technology is one of the big data analysis technologies. It automatically learns the normal behaviors of various systems and monitors changes in behaviors in real time in order to improve the availability of social infrastructures and to reduce their operating costs. In this paper, we outline the impact of the architecture of this technology on the cloud environment and its analysis applications.

2. System Invariant Analysis Technology

2.1 Techniques and Feature

The system invariant analysis technology is a kind of machine learning technology that exhaustively extracts the relationships among the time series sequences of numeric values obtained as system performance data or plant sensor data¹⁾. The

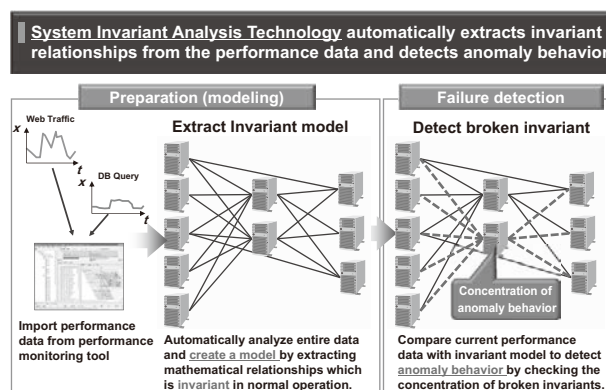


Fig. 1 System Invariant analysis technology.

relationships between sensor values are extracted as a normal behavior model of the target system. The early signs of failures, the time and place of the system's behavior change, can be detected by monitoring the broken relationships in the model (Fig. 1). In case of applying this technology to the system management process, it enables the detection of silent failures; for example, a kind of performance degradation unaccompanied by any error message is difficult to find by using conventional monitoring techniques such as threshold monitoring, baseline monitoring, etc. It enables an early workaround before the failure actualization and propagation, therefore, it reduces the business loss and the operational cost of service failures.

This technology can essentially use time series of numerical values as the target of analysis. Therefore, for example, it allows cross-domain analysis, analysis of all kind of performance data related to the service acquired from the servers, networks, databases and application software, which enables rapid triage of the failure according to the management organization of each of the components. Moreover, it will then be possible to apply the findings for visualizations or simulations over the entire range of the business objectives; it allows analysis of the facility status of the data center by using the sensor value of temperature, power consumption values, etc., and it also allows finding out relationships of the business indicators such as stock amounts and sales amounts, etc.

2.2 Target Systems of Analysis

(1) Analysis of systems in the cloud

The system performance analysis software for ICT systems has been released as the first application of this technology²⁾. Since ICT systems can be monitored and controlled by component parts of the existing integrated operation management software, linkages with their interfaces enable automation of processing operations; such as by performance data aggregation, model creation, failure detection and recovery (Fig. 2). Recently, the business

systems deployed in the cloud environment are increasing, as is their use in remote monitoring and distributed data center operations. Because of this, it has now become rare that delays in the monitoring control cause practical problems, except when extremely high speed is required in data processing.

For example, this technology allows service providers to visualize the health of the VMs (Virtual Machines) and of the applications installed on them, while also allowing the cloud infrastructure providers to visualize bottlenecks in the performances of platform components such as the servers and networks. It is also capable of comparing changes in behavior before and after maintenance work as well as anomaly detection during service operations. Therefore, it can provide supportive evidence indicating the system's normal behavior based on objective observational data, whereas normality judgments used to be dependent on the contents of operation manuals and the operator's expertise.

(2) Analysis of external systems

For the application of this technology to targets other than ICT systems, we provide a plant failure sign detection solution (Fig. 3). Plants handling electric power, chemicals and steel are monitored by plant management systems, so their sensor data can be collected through the interfaces of these systems. Since the operation policy and domain knowledge are different in each application domain, we provide generic and customizable software components; a component can adapt its analysis logic and display screen to each application domain and enable provision of a solution in each domain.

For example, it can provide displays focused on prediction alerts for plant operators (Fig. 4) and displays enabling detailed parameter configurations for the expert operations and analyses in order to create models and to analyze with past data (Fig. 5). When this solution is combined with an existing operational system, behaviors

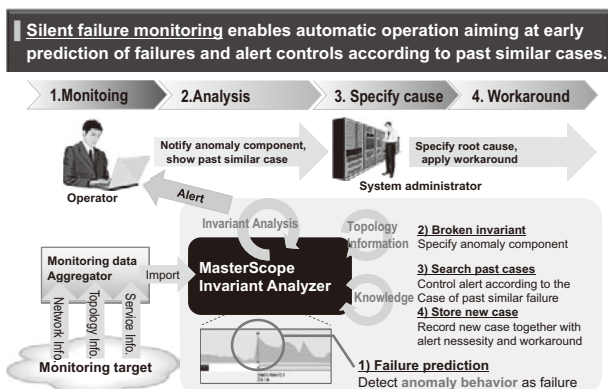


Fig. 2 Silent failure monitoring.

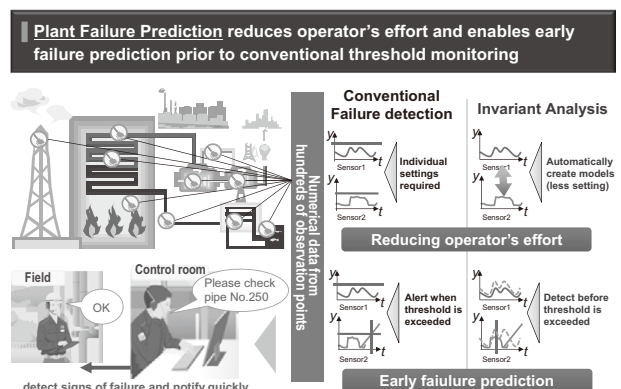


Fig. 3 Plant failure sign monitoring.

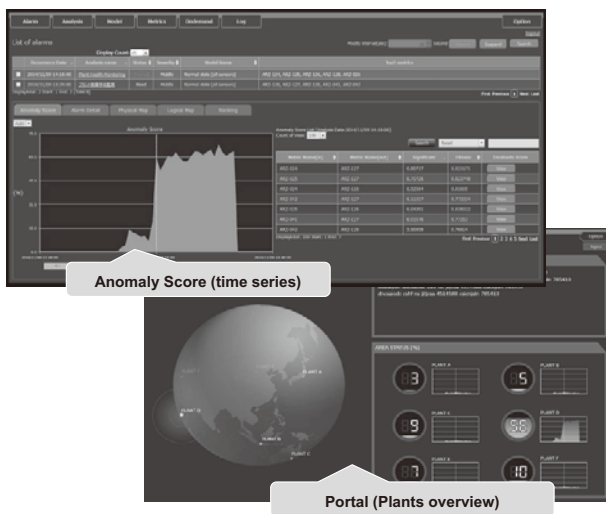


Fig. 4 Examples of failure alert displays.

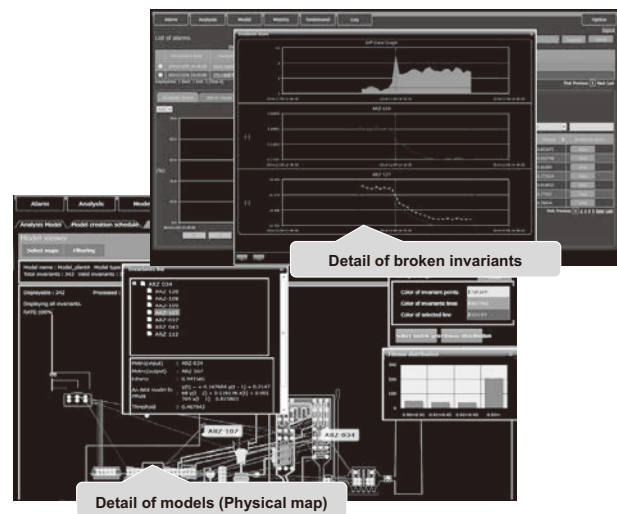


Fig. 5 Examples of model detail displays.

preceding failures that have previously been hard to find can be detected early on. So the maintenance cost can be optimized and losses caused by the interruption of the service can be reduced.

3. Invariant Analysis in the Cloud

3.1 Comparison between On-premise and Cloud Analysis

This technology is capable of automatic and fast analyses of the relationships between numerical data. When an analysis is performed in the location of the target system (on-premise analysis), detailed data can be acquired and analyzed so that the action according to the analysis results can be executed immediately. On the other hand, when the data aggregated from a remote data center is analyzed (cloud analysis), various kinds of data can be analyzed although there are delays in data monitoring and in the alert notification.

In the case of a plant failure prediction it is essential to collect detailed data at high frequencies in order to detect any slight sign of failure such as a minor change of relationship between sensor values from the large amount of plant facilities. In order to adopt proper workarounds prior to any increasing damage due to a failure, anomaly alerts should be notified promptly to the field operators in the plants. For this purpose, on-premise analysis is suitable because of its capability of tight linkages with the plant control systems monitoring sensor values and with the incident management system notifying anomaly alerts.

When performing the cloud analysis, it is necessary to carry out data communications between the remote cloud environment and the analysis target system. Such a condition may drop the performance less than for the on-premise analyses

in case of the real time analysis of a large amount of data. Instead, it enables deeper analyses of various kinds of data for an experienced analysis expert. For example, aggregating data from several plants enables a detailed comparison of the operational status of each plant. Analyzing the relationships including environmental sensor data and/or business data from outside of the plant helps to consider the mid/long-term improvement plans. Even if analysis experts cannot be assigned for each plant, analysis experts in the cloud side may provide the monitoring setting of each plant and, if necessary, perform remote monitoring in place of field operators.

3.2 Features in each Application Domain

Both on-premise and cloud analysis have their respective advantages and disadvantages depending on the purpose of the analysis and the assignment of the analysis experts as described above. It is therefore required to select whether utilizing either or both of them accords to the requirements and achievement level of each application domain. Following is the explanation of some examples of the domains currently undergoing experiments.

(1) Failure prediction for electric power plants

Currently, the major part of requirements in this domain is concerned with detecting slight signs of failures from a relatively stable plant. Therefore, an on-premise analysis is suitable because of its potential for high speed and frequency monitoring of sensor data from the plant control system. However, in the future, a cloud environment, of at least enterprise-wide size, will be required to enable the integrated analysis of information aggregated from each plant in order to optimize and reduce the total costs of power generation and for the maintenance of related

plants such as LNG and coal, etc. Particularly for the power plants in overseas countries where the generation and transmission of power are separated commercially, there are great variances between the operations of different plants, so the provision of objective analysis as an external service is keenly awaited.

(2) Failure predictions for manufacturing plants

For the process plants of the manufacturing industries, petrochemical, steel, etc., the main requirements are the on-premise analyses as for the electric power plants. In this domain, monitoring settings are often modified according to changes of products or other environmental factors. However, it is difficult to assign an analysis expert in each plant because of the added demand of reducing manufacturing cost. One of the strategies adopted to deal with this situation is to combine the cloud analysis in supporting the fine tuning of several plants. However, this may require enhancement of the linkages between the on-premise and cloud analysis, for example, by appropriately allocating the tasks to the field operators and to remote analysis experts.

(3) Failure analysis of large-scale facilities

With large-scale complex facilities such as aircraft, advanced considerations of the analysis environment designed for built-in facilities is required in case of the on-premise analysis. Therefore, the analysis is often started on the cloud environment aiming at improved efficiencies in fuel consumption and parts replacement. Visualization of facility behaviors and prediction of unknown failures are expected rather than detection of known failures because systems in this domain often adopt multiplex safety design and monitoring of features with many sensors.

In future, on-premise analysis will be available sequentially starting from the features that have been proven effective in the cloud environment by the verification of the design and analysis experts. In the case of verification in the cloud environment, it will also be necessary to achieve features not only of provisioning the traditional ICT system but also of simulating the actual behavior of facilities.

(4) Structural health monitoring

Aiming at reducing diagnostic cost and improving the efficiency of repairs, new degradation diagnosis techniques are emerging that use environment sensors for engineered structures. While the on-premise analysis environment is capable in cases of the diagnosis of buildings, it is actually hard to install analysis servers permanently on bridges or other civil engineering structures. Therefore, the diagnosis of engineered structures is essentially made in the cloud environment with remote monitoring. Even when on-premise analysis is applied, patrolling vehicles

or other moving equipment may be considered instead of the diagnostic equipment installed permanently on the structures.

Actually, the location where sensors should be arranged and the way that sensor data should be aggregated remain as challenges to be solved in the future. Especially so in cases of degradation diagnosis using accelerometers or other vibration sensors, as a large amount of high frequency data is generated by sensors continuously and the patrolling vehicles or cloud servers will be required to aggregate the data without losses or delays. Some technologies concerned with the IoT may offer a capable solution and linkage with such a data aggregation platform is also required for the cloud analysis.

(5) Financial and business analysis

This is an application with which various kinds of data is aggregated into the DWH and utilized for the business situation analysis and fraud detection etc. As various kinds of analysis software tools, business intelligence (BI) tools, have been provided for analysis experts and cloud analysis is suitable in this domain because of the ease of linkage with such tools. When an analysis expert uses this technology to extract new relationship from data, it is essential to enable a shared use of input data and analysis results with the existing BI analysis environment. If visualization of the business situation is required, it is also necessary to provide data linkages with the existing enterprise business systems and new portal displays enabling a real time overview of the analysis results.

4. Conclusion

As described above, the system invariant analysis technology detects anomaly behavior of various systems and supports stable system operations. From now on, we will advance the development of solutions to resolve issues such as the linkage between on-premise and cloud analyses and linkage with the data collection platform, etc. in order to contribute to the achievement of the safety and security of the social infrastructures.

Reference

- 1) "Complex System Analytics", <http://www.nec.com/en/global/rd/research/cl/siat.html>
- 2) "Invariant Analyzer", <http://www.nec.com/en/global/prod/mastertscope/invariantanalyzer/>

Authors' Profiles

KATO Kiyoshi

Manager
Big Data Strategy Division

MIYAZAKI Hiroyuki

Assistant Manager
Big Data Strategy Division

SOMA Tomoya

Manager
Big Data Strategy Division

TAKAGI Mayumi

Big Data Strategy Division

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.9 No.2 Special Issue on Future Cloud Platforms for ICT Systems

Remarks for Special Issue on Future Cloud Platforms for ICT Systems

NEC's Approach to Orchestrating the Cloud Platform

NEC C&C cloud platforms ? NEC Cloud IaaS Services

Portal Services Integrate Multi-Cloud Environments

A Hybrid Server Hosting Which Have Broader Range of Applications

Network Service That Offers a Versatile Network Environment

Dependable Security Service That Takes Advantage of Internal Control Methodology

Data Center Service That Supports Cloud Infrastructure

Products and latest technologies supporting NEC C&C cloud platforms

MasterScope Virtual DataCenter Automation - Entire IT System Cost Optimization by Automating the System Administration

Integrated Operation and Management Platform for Efficient Administration by Automating Operations

Micro-modular Server and Phase Change Cooling Mechanism Contributing to Data Center TCO Reduction

iStorage M5000 Providing a High-Reliability Platform for the Cloud Environment

The iStorage HS Series Features the Superior Data Compression and High-Speed Transmission Capabilities that are Essential Functions of Big Data Storage

SDN Compatible UNIVERGE PF Series Supports Large-Scale Data Centers by Automating IT System Management

Phase Change Cooling and Heat Transport Technologies Contribute to Power Saving

Future technology for NEC's C&C cloud platforms

Accelerator Utilization Technology That Cuts Costs, Reduces Power Consumption, and Shrinks Hardware Footprint

Scalable Resource Disaggregated Platform That Achieves Diverse and Various Computing Services

Support Technology for Model-Based Design Targeted at a Cloud Environment

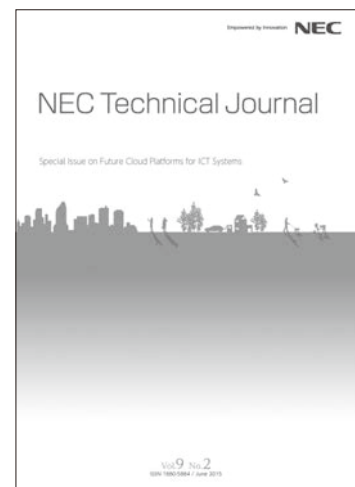
Cloud-based SI for Improving the Efficiency of SI in the Cloud Computing by Means of Model- Based Sizing and Configuration Management

Big Data Analytics in the Cloud - System Invariant Analysis Technology Pierces the Anomaly -

Case Studies

Using Cloud Computing to Achieve Stable Operation of a Remote Surveillance/Maintenance System Supporting More Than 1,100 Automated Vertical Parking Lots throughout Japan
Meiji Fresh Network's Core Business Systems are Transitioned to NEC Cloud IaaS NEC's Total Support Capability is Highly Evaluated.

Sumitomo Life Insurance Uses NEC's Cloud Infrastructure Service to Standardize IT Environments across the Entire Group and Strengthen IT Governance



Vol.9 No.2

June, 2015

Special Issue TOP

NEC Information

NEWS

2014 C&C Prize Ceremony
