# Social Innovation through Utilization of Big Data

Shuntaro Hitomi
Keiro Muro

*OVERVIEW: The analysis and utilization of large amounts of actual operational data collected from equipment and devices have made it practical to realize the possibilities of converging and advancing the control and information systems. An effective approach to analyze and utilize big data efficiently is to determine the events that must be discovered through data analysis to achieve business objectives, and then select the appropriate mathematical analysis algorithms based on available data, thereby building an analysis method. Hitachi has formed a team specializing in data analysis with extensive experience in product manufacturing, maintenance, and operations, while bringing together a tremendous amount of mathematical analysis technology research in order to promote the achievement of social innovation through the utilization of big data.*

## INTRODUCTION

ADVANCEMENTS in information technology (IT) have caused the amount of data we can access to increase dramatically. In other words, factors such as faster computational processing through IT, expanded network environments for data transmission including the Internet and wireless communications, and increased capacity in the storage devices used to store data for a certain period of time have made it a realistic prospect to handle larger amounts of actual data than previous systems could handle. As a result, major opportunities have opened up for collecting and analyzing these massive amounts of data for utilization in the improvement of business quality.

This article describes new services that are implemented in the public infrastructure sector through the collection and analysis of large amounts of data, as well as case studies.

## PUBLIC INFRASTRUCTURE AND BIG DATA

### Age of “Things” Big Data

#### Big data of “things,” “people,” and “concepts”

As the amounts of data that can be handled increase, the data types available are also expanding (see Fig. 1). Hitachi sees big data as being divided into three categories, “things,” “people,” and “concepts.” In addition to the analysis of production management, inventory management, and other types of back office

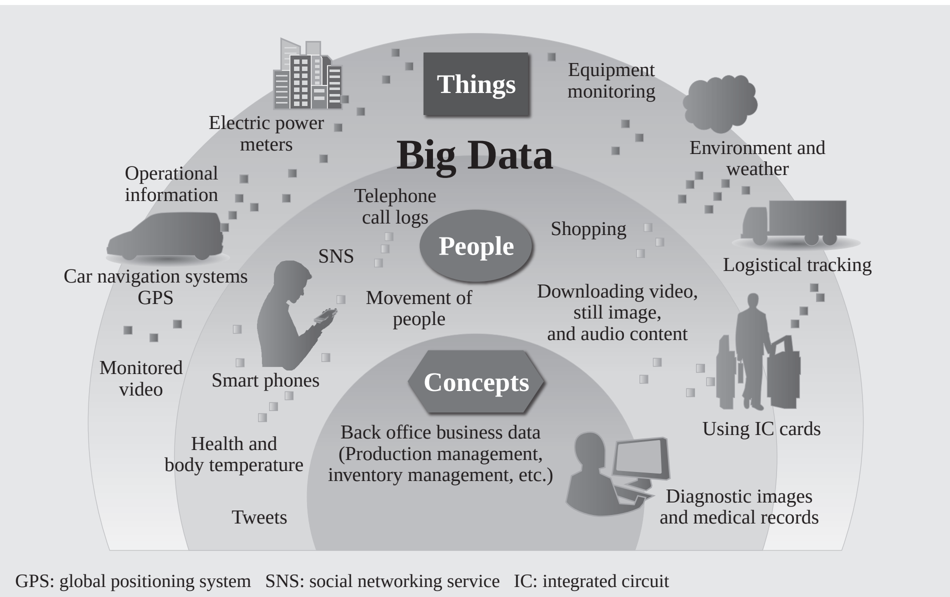


*Fig. 1—Big Data = Data of “Things, “People,” and “Concepts.”*
*The world of data that can be processed with information technology (IT) is growing day by day.*

business data ("concepts" big data, or abstract things), which IT has traditionally been strong in, the big data that is generated from the actions of individual people ("people" big data), such as text messages posted to social networking services (SNS) and other forms of social media as well as location, movements, and other information that can be acquired from smartphones and other devices, can now be utilized through the application of IT.

Furthermore, it has also become practical to gather, analyze, and utilize the data generated through the operation of "things," including the operational information of transportation systems and the global positioning system (GPS) information of car navigation systems, as well as actual operational data obtained from sensors installed with the purpose of monitoring and controlling a variety of different mechanical equipment ("things" big data), without the need to discard such information each time.

### Convergence of information and control as implemented through utilization of "things" big data

One of the major features of "things" big data is that it is a record of device operation derived from the actual use of devices in society, unlike the test result data that is from previously envisioned situations and a certain set length of time derived from sources such as equipment tests and trial operation. By appropriately analyzing this data, it is possible to accurately evaluate the current state of devices that are actually operating, and to analyze the effectiveness of a device maintenance business or the influence of operations.

Although in the past, it was difficult to obtain the actual detailed operational data of devices in certain fields, the individual know-how of those in charge of operations and maintenance businesses who directly handled the devices during actual operations has led through various trial-and-error processes and contrivances during device maintenance and operation to this know-how being both accumulated and passed down throughout organizations. The effects described below can be expected as the result of incorporating the analysis and utilization of this actual operational data, which is the "things" big data itself:

(1) The effects and influence of maintenance policies or operational changes can be evaluated and feedback can be created using actual operational data. This can be used to accelerate improvements in maintenance and operations.

(2) Appropriate maintenance and operation modification methods can be automatically predicted and determined based on an evaluation of the current state of devices, which is derived through data analysis. Also, maintenance and operational know-how that had to be passed down individually in the past can now be replaced with a method of analyzing the data of maintenance and operational know-how, which is an effective way of sharing and passing down know-how.

(3) The results of collecting and analyzing a wide range of device positional and operational information data, energy consumption data, and other data can be used to automate operations, balance supply and demand, support human activities, and otherwise converge and increase the sophistication of information and control systems.

## Hitachi's Efforts in Utilization of "Things" Big Data

### Challenges facing data analysis

A valid method for effectively implementing data analysis and utilization is to construct a data analysis procedure by following these steps: (1) determine the events that should be derived through data analysis in order to achieve business objectives, (2) obtain and select the data that can be used for this purpose, and (3) select and tune the necessary mathematical analysis algorithms.

While executing this series of steps, it is important to master the properties and usage methods of many mathematical analysis algorithms, to possess a wide range of knowledge, and to understand how usable data reflects device states and physical properties (information specific to the devices). In particular, evaluation and verification are indispensable parts of the algorithm selection and tuning process, and in some cases, a long period of trial and error is necessary. This knowledge and information is useful for correctly interpreting what physical or business significance the indices that act as criteria for discovering events have, and as a result, can be utilized for quickly discovering analytical methods commensurate with the objectives.

In the past, a system that was widely used included measuring data for mechanical devices under test or trial operation environments, securing device design quality, and applying manufacturing process control. Due to Hitachi's many years of experience in the research, development and manufacturing of countless mechanical devices, the group possesses a wide range of information regarding the physical properties of devices, and a wealth of knowledge about the mathematical analysis algorithms necessary for data analysis.

### Challenges specific to "things" big data

Two new challenges must be conquered in order to expand the scope beyond test result data obtained from equipment testing and trial operations, to the construction of data analysis methods for analyzing and utilizing actual data from operations.

(1) A wide range of external environmental factors affect data, and many different events related to actual operating devices can affect the actual operational data. This is why it is necessary to appropriately select which event's occurrence to focus on from among a large number of possible events, and to find out how that event is reflected in the data.

(2) In many cases, it is difficult to adjust sensors or add new ones in order to obtain data that is useful for the analysis, as devices are actually operating in different parts of society. Even under these conditions, it is still necessary to efficiently separate the occurrence of target events from the noise that also exists in the data, which is caused by other factors.

In order to meet the challenges described above, it is important to have extensive knowledge and information regarding the site of actual operations or the maintenance business where the "things" or devices are being used. It is also important to have a comprehensive data utilization procedure in order to effectively use this knowledge and information.

Hitachi has extensive experience in maintenance and after-sales support businesses for its own products, and also has a line of products and services where it provides operational monitoring support that allows it to remotely gather and accumulate operational data. It is through its experience analyzing information from the sites of maintenance and operations as well as accumulated operational data that Hitachi has built the powerful data utilization techniques (mathematical analysis methods, IT processing technologies, and data analysis and utilization method construction processes) necessary for utilizing "things" big data.

## BIG DATA UTILIZATION SOLUTION CASE STUDIES

Hitachi's utilization of "things" big data is described below, along with case studies.

### Remote Monitoring of Gas Turbines

In order to provide equipment operation support and quick troubleshooting via long-term service agreements (LTSA), Hitachi has been offering remote monitoring services since 2003 with a focus on gas turbine plants (see Fig. 2).

With a remote monitoring system, analog sensor data from the control panels of power generation equipment delivered to the customer as well as digital data from equipment control signals and other sources is acquired in one-second intervals and forwarded to the monitoring center for storage once per day. A total of between approximately 300 and 2,000 points of analog and digital data is generated by each plant in a single day, amounting to from several hundred megabytes to several gigabytes of time-series data. This time-series data is utilized in creating monthly reports, responding to maintenance inquiries, and assisting design. As the data is retained for a long period of time (15 years or more), its accumulated amount can be in the order of the terabytes (thousands of gigabytes).

It will be necessary to efficiently utilize these huge amounts of data in the future in order to increase

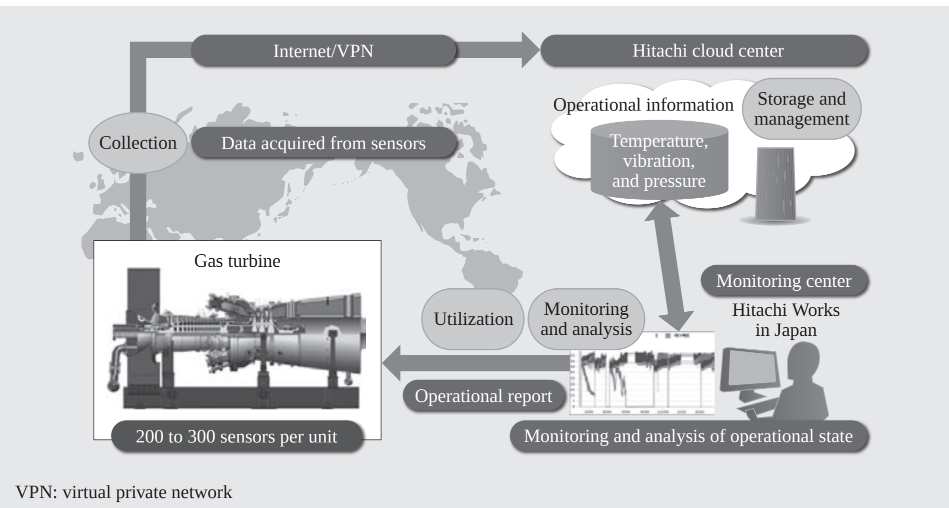


*Fig. 2—Gas Turbine Remote Monitoring and Maintenance Business.*
*Trial operation monitoring and preventive maintenance based on data monitoring and analysis contribute to improvements in rates of operation.*

the sophistication of monitoring and diagnosis so that customer service be improved. For this reason, Hitachi is developing technology that can store this time-series data in a highly compressed form while at the same time enabling fast search, and is introducing this technology at monitoring centers for demonstration. Using this technology makes it possible to provide services with a high level of added value, and both lower in cost and faster than before to launch. Hitachi is also researching and developing technology that can provide early anomaly detection, in order to further improve the reliability of plant operations.

### Maintenance, Operation, and Management of Chillers

Hitachi already provides a service to acquire and remotely monitor the operational data of industrial chillers, and is proceeding with efforts to analyze this operational data further for use in improving its equipment operation and maintenance business (see Fig. 3). Sensor data acquired from inside the chiller for the purpose of controlling the equipment is used as actual operational data, and is collected and accumulated for a long period of time. Concurrently, by-the-minute records are being collected regarding equipment adjustment and maintenance work, and this makes it possible to know the long-term history of which chillers were maintained and adjusted when and how.

Actual operational data for chillers has traditionally been the target of automatic monitoring. An alarm would be triggered at a remote monitoring center whenever a threshold value was exceeded, and this was useful for equipment maintenance and inspection businesses. With simple threshold monitoring, however, it is only possible to begin maintenance after the anomaly has occurred, such as abnormal equipment stoppages or obvious performance degradations.

This is why Hitachi has been analyzing previous chiller alarm and maintenance history logs as a test to determine the causal relationships before and after problems occur in actual operational data, on an experimental basis. As a result of this analysis, it was shown that the actual operational data of the equipment before the alarm occurred would clearly approach the alarm condition, and Hitachi worked to create indices to clarify this type of trend. By monitoring these indices, it is possible to evaluate equipment degradation and assist with timely maintenance before an anomaly occurs.

These indices are still under development, and the challenge that remains before they can be put to practical use is to ensure that as long as the equipment model for which they were designed is being used, that they can diagnose both accurately and with a high level of reliability the degree of anomalous behavior and the amount of time left before an anomaly occurs, regardless of the installation or usage conditions. Efforts to establish these indices and put them to practical use are continuing.

## HITACHI'S BIG DATA ANALYSIS AND UTILIZATION SERVICES

In data analysis efforts, it is important to clarify the goal of determining what information must be extracted from the data to realize concrete benefits for actual business. To this end, it is also important to receive the support of a team of data analysis specialists who are proficient in a variety of different

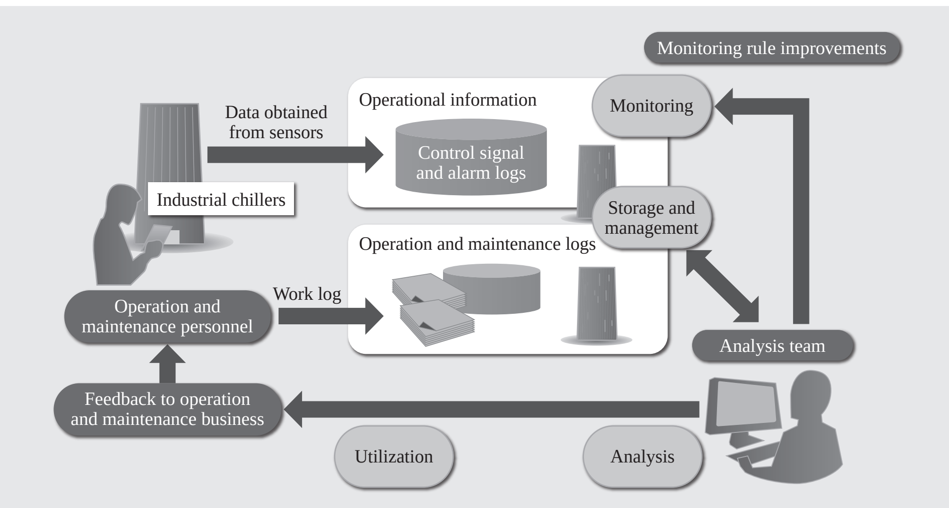


*Fig. 3—Maintenance, Operation, and Management of Chillers.*
*Equipment monitoring, operation, and maintenance businesses are being improved by analyzing actual operational data and work log data.*

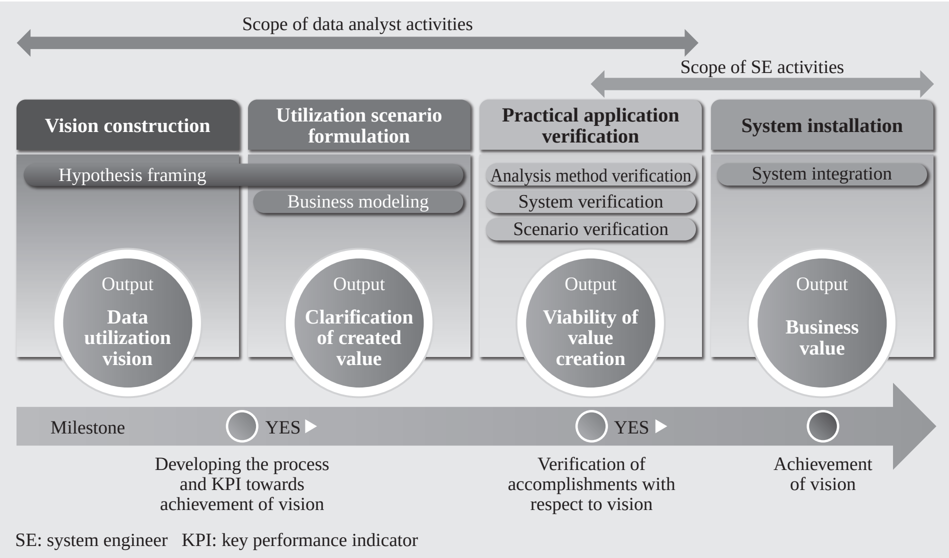


*Fig. 4—Details of Hitachi's Data Analytics Service Activities.*
*The data analysis method is constructed and verified in order to understand the customer's trade and resolve customer's business challenges (achieve the customer's vision).*

data analysis methods in addition to knowledge and understanding of related to business.

Hitachi announced a "data analysis service" in June 2012 in order to support the creation of new business value from big data. This service constructs and provides data analysis methods that match business details and actual data that can be used. Its objective is supporting the resolution of the customer's business challenges by teams of data analysis specialists who bring together Hitachi's extensive store of mathematical analysis technology, data analysis experience, product manufacturing experience, and knowledge derived from constructing systems for a variety of different customers (see Fig. 4).

This data analysis service is aimed at a wide range of fields including finance, the public sector, communications, and logistics, and works to achieve a convergence of information and control through the aggressive utilization of "things" data in both industry and control. Hitachi brings together and applies its many years of accumulated knowledge in the manufacturing, operation, and maintenance of "things" in order to provide this service, which achieves the utilization of "things" big data.

## CONCLUSIONS

This article discussed a new service that is implemented through the collection and analysis of large amounts of data in the public infrastructure field, and described case studies.

The utilization of "things" data is a key technology that will open up possibilities in the improvement and optimization of maintenance and operation businesses based on actual data, and in the application of diagnostic results to control in real time, which will enable sophisticated operational automation.

Hitachi will continue striving to promote the utilization of big data by applying its extensive knowledge regarding the manufacturing industry with teams of specialists in mathematical analysis technology, as part of its overall efforts to achieve social innovation through the analysis of data in the public infrastructure field, based on the utilization of "things" data.

## ABOUT THE AUTHORS

Shuntaro Hitomi

*Joined Hitachi, Ltd. in 1999, and now works at the Emerging Business Laboratory, Business Innovation, Smart Information Systems Division, Information & Telecommunication Systems Company. He is currently engaged in the development of big data analytics services.*

Keiro Muro

*Joined Hitachi, Ltd. in 1993, and now works at the Social Information Systems Research Department, Central Research Laboratory. He is currently engaged in research into time-series data management systems and time-series data analysis systems for machine maintenance and energy management systems. Mr. Muro is a member of the Information Processing Society of Japan (IPSJ).*