Next-generation IT Platforms Delivering New Value through Accumulation and Utilization of Big Data

—One Platform for All Data

Hiroaki Fujii

CHANGING IT TRENDS

THE spread of cloud computing has focused attention on what is being called "big data," the large quantities of data being accumulated on and circulated around the cloud. The conversion of this data (which was collected, stored, processed, and generated for a particular purpose) into knowledge that can be utilized in a range of different ways is the key to creating new value in areas such as customers' businesses or social infrastructure.

Hitachi is accelerating the actions it is taking to build IT (information technology) systems that support this creation of new value.

NEW TRENDS IN USE OF IT

A wide range of end users have familiarized themselves with IT since the rise of the Internet, and we are now living in a time when demand for this technology and its potential for disseminating information are shaping the nature of business and other parts of society like never before. For example, use of social networking services such as blogs or Twitter*1 continues to spread due to the arrival of smartphones and other mobile devices with advanced functions. Services like these now process daily volumes of data in the petabyte range. Many new business initiatives are emerging that involve making active use of the data that can be obtained through such services in marketing, particularly among companies engaged in consumer businesses.

It is also possible nowadays to collect large quantities of different types of digital data through the use of technologies such as RFID (radio-frequency identification) and sensors. Examples of the various ways these can be used to provide a visual user interface of the data on the real world include determining the status of goods during distribution; factory environmental conditions such as temperature and humidity; use of probe cars, ETCs (electronic toll collection systems), and other tools to determine traffic conditions; history data from smartcard ticketing; e-money purchasing records; electricity usage and other forms of automatic metering; and Earth observation data from satellites.

*1 Twitter is a trademark of Twitter, Inc. in the United States and other countries.

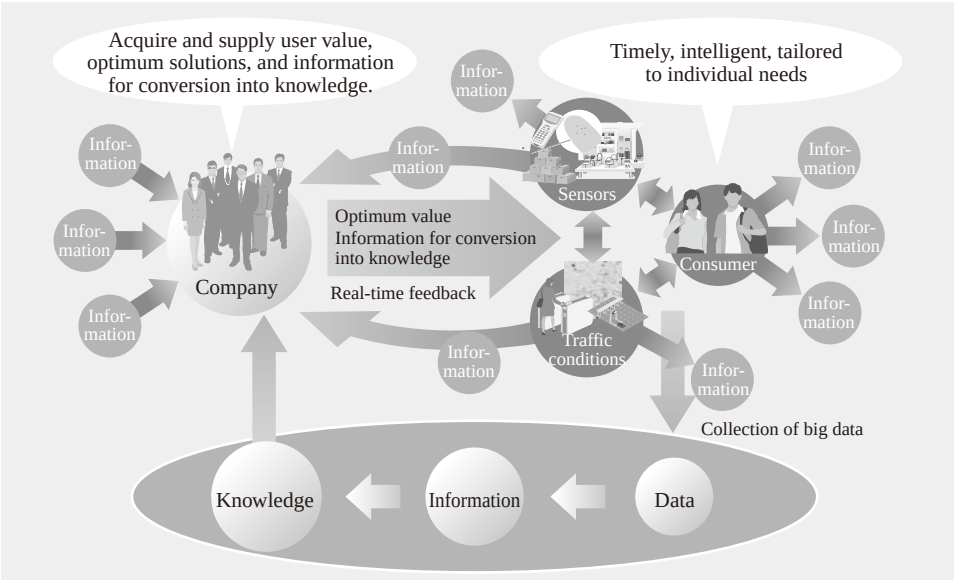
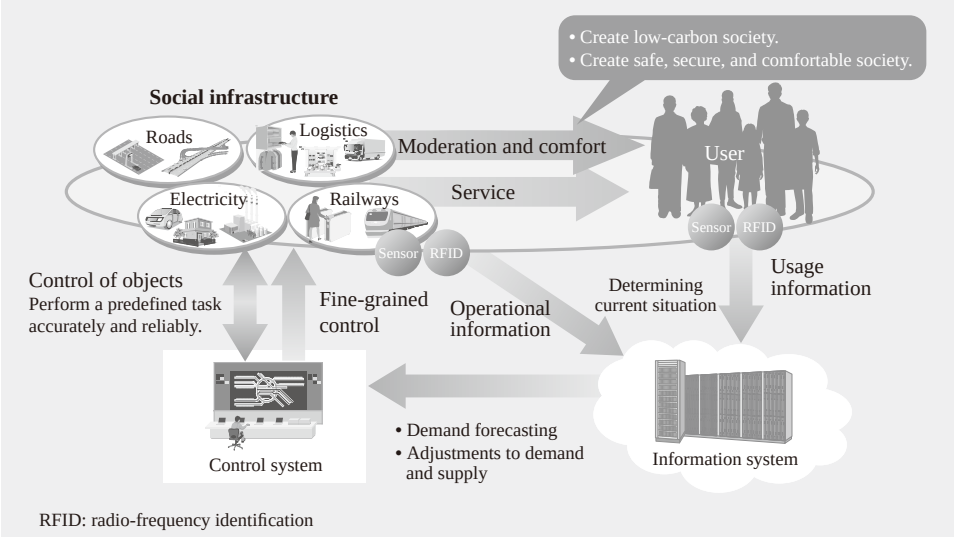


Fig. 1—Framework for Use of Big Data in Consumer Businesses.
More appropriate products and services can be provided by utilizing data generated by consumer purchasing activity.



RFID: radio-frequency identification

Fig. 2—Framework for Use of Big Data in Social Infrastructure.
The use of various real-world data associated with social infrastructure can lead to the creation of a society that provides greater safety, peace of mind, and comfort.

Effective utilization of this rapidly growing data is seen as a core requirement of IT in the future.

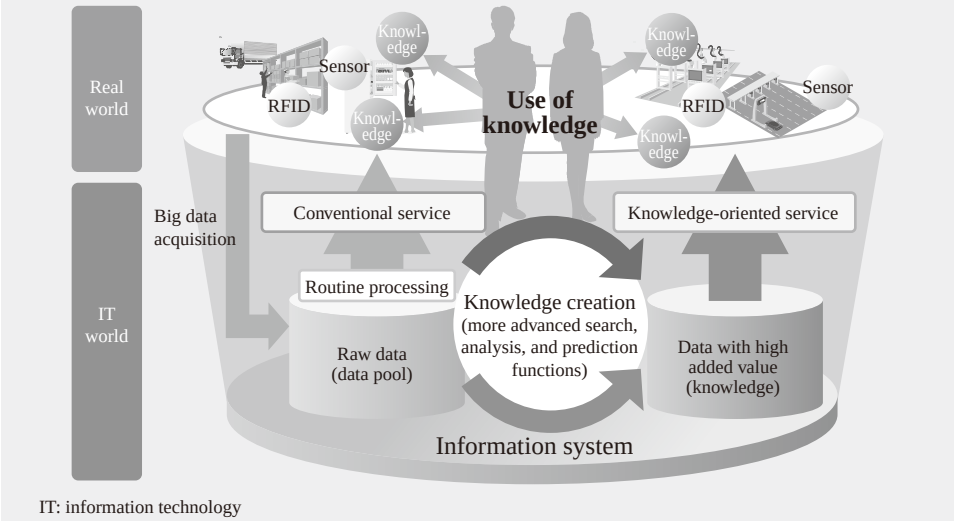
Fig. 1 outlines a framework for use of data in business.

For example, purchasing activity generates a variety of data at consumer businesses, which can then be made available. Turning this data into information, and then into knowledge, facilitates the identification of user needs at companies that supply products and services, and allows them to deliver value in a way that best matches these needs.

Similarly, Fig. 2 outlines a framework for use of data in social infrastructure.

In the past, social infrastructure (such as electricity; rail, road, and other transport infrastructure; and logistics systems) provided services to users with support from control systems, the primary role of which was to perform certain predefined tasks accurately and reliably. However, future social infrastructure will be able to provide new value through the incorporation of information systems. Specifically, information relating to things such as the operating status of the social infrastructure itself, or use of the social infrastructure services by users, will be acquired using sensors, RFID, and other new device technologies in order to provide an understanding of what is happening in the field. More fine-grained control of social infrastructure than was possible in the past will be enabled by converting this data into information in the form of demand forecasts, supply and demand adjustments, or other forms that can be used by control systems.

This will inject added value into the social infrastructure services supplied to users in the form of things like moderation or comfort. That is, it will become possible to supply feedback on users' views to the social infrastructure services supplied to those



IT: information technology

Fig. 3—How Big Data will be Used in the Future.
Future use of big data will involve collecting raw data from the real world and then analyzing and using it to create knowledge. This will lead to the supply of new services, which will use this data with high added value that has been turned into knowledge.

users. The result will be to help achieve objectives such as a low-carbon society or a society that provides safety, peace of mind, and comfort. The rebuilding of social infrastructure in Japan, as part of the recovery from the Great East Japan Earthquake, is taking a variety of forms. The use of data to add value is seen as an essential requirement in this process. Hitachi believes it is contributing to the recovery by supplying such solutions.

Supporting these processes for converting data into information and knowledge will be a new role for IT in the future. The traditional role of IT has been to input data from a particular scope and process it in a predefined way and this was intended to improve business efficiency, especially for routine tasks. As shown in Fig. 3, the future role of IT will involve using large quantities of real-world data in digital formats, not only to improve business efficiency, but also to produce knowledge that leads to business innovation. The huge quantities of raw data collected will be used to produce knowledge in the form of data with high added value by using advanced IT functions such as search, analysis, and prediction, and this new knowledge will then be provided as feedback to different real-world situations where it will be used to produce services or new products with even higher quality. In broad terms, this is how IT that supports the utilization of data will work.

ADVANCES IN IT PLATFORMS THAT UNLOCK NEW VALUE

Form of Infrastructure for Processing of Big Data

Hitachi uses the term “One Platform for All Data” to define the future form to be taken by IT platforms in their new role of utilizing big data to improve knowledge productivity (see Fig. 4).

Functions for the unified management and processing of data as well as flexible and scalable systems are important for the accumulation of diverse types of big data so that it can be processed in a variety of ways for use. To satisfy these requirements, Hitachi is enhancing and equipping its platform products based on the following:

(1) IT infrastructure layer

Configuration of systems consisting of a number of servers, storage units, and other devices, and the ability to operate these systems without having to take account of constraints, such as processing capacity and data storage capacity, in order to implement virtualization and automation.

(2) Content management layer

Allow unified and intelligent management and processing above the IT infrastructure layer without having to take into account the type or other attributes of the data or content.

One Platform for All Data seamlessly integrates the IT infrastructure and content management layers to create new value through the use of big data.

IT Infrastructure Layer

As mentioned above, the underlying technologies that support the IT infrastructure layer are virtualization and automation.

Hitachi storage businesses have been working consistently to develop advanced virtualization technologies. Hitachi Universal Volume Manager virtualization technology for multiple heterogeneous storage devices provides unified storage management

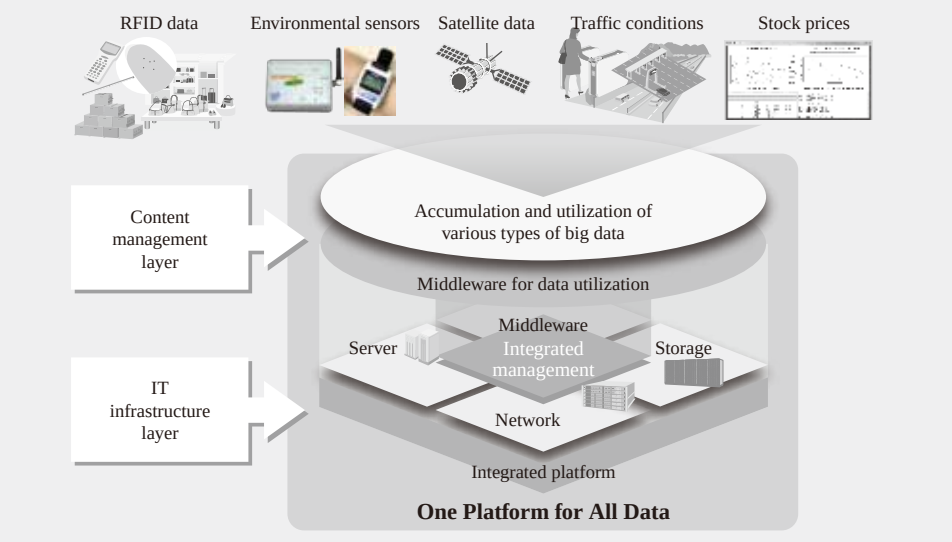


Fig. 4—One Platform for All Data.
This is the new shape of the IT platforms that Hitachi aims to provide for the unified and efficient accumulation and utilization of various types of big data.

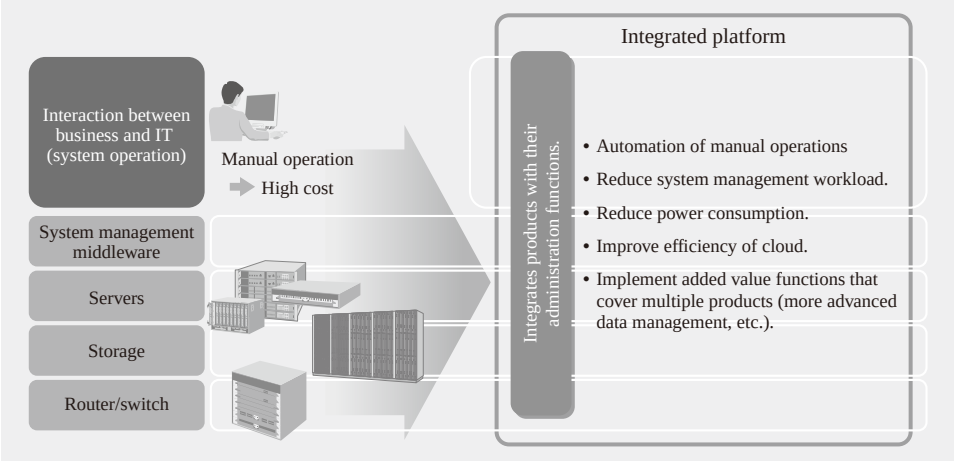


Fig. 5—Platform with High Added Value through Integration, Virtualization, and Automation.
Customer TCO (total cost of ownership) is reduced due to the added value produced by a shift from providing individual products to providing vertically integrated platforms.

and makes operation easier. Meanwhile, the Hitachi Dynamic Provisioning volume capacity virtualization function simplifies complex volume capacity design and administration. To satisfy various cost-performance criteria in an optimum way, Hitachi Virtual Storage Platform automates system management using the Hitachi Dynamic Tiering storage hierarchy virtualization function, which supports a range of different device types including HDDs (hard disk drives) and SSDs (solid state drives).

Using its proprietary logical partitioning for Hitachi compute products, which incorporates virtualization technology built up through experience with mainframes, Hitachi has transformed Hitachi Compute Blade into a highly reliable platform suitable for cloud computing. The aims for the future include virtualization of I/O (input/output), the provision of server resource pools, and more flexible platforms.

On the other hand, system management middleware has the core role in automating the operation of the IT infrastructure layer. IT resource management function in Job Management Partner 1 handles the various stages of the operational lifecycle (design, implementation, and capacity monitoring of IT resource allocation) to optimize use of virtualized and shared IT resources across business systems. IT resource management also requires the consistent management of mixed virtualization architectures. When IT resources are managed across different business systems, not only system management of the various platform devices, but also integrated system management that incorporates the service level management provided by IT service level management function in Job Management Partner 1 become increasingly important as foundations for supporting virtualized platforms.

The IT infrastructure layer of One Platform for All Data integrates the individual platform products such as servers and storage with functions for their administration (see Fig. 5).

Demand for vertically integrated platforms is growing even in the European and North American markets, where the standard practice in the past was to purchase the best product for each type of component and then combine them into a system. Reducing the cost of managing IT systems (which has been increasing year by year) has become an important issue, and is one of the drivers behind this trend. As progress is made on the consolidation, integration, and virtualization of IT systems, consolidation and integration reduce the amount of system management work. Greater use of virtualization, on the other hand, makes the environment more complex and results in higher costs for system management that involves the use of administration tools for specific components (servers, storage, and other devices).

In response, the integrated platform that Hitachi aims to provide will reduce the TCO (total cost of ownership) by automating system management functions such as ensuring service levels or configuring the system environment based on the use of virtualization. For example, the value delivered by the integrated platform will include operating with low power consumption and more efficient cloud operation.

Hitachi Unified Compute Platform is a product that realizes this integrated platform and it is already being deployed in pilot applications in North America. This product can also be positioned as a basic component in a private cloud[(a)] run from a corporate data center. Details can be found in the “Integrated IT Platform

(a) Private cloud
A cloud computing system implemented internally by a company or other organization. In contrast, standard cloud systems that are provided over the Internet are often called “public clouds.”

for Highly Efficient Operation in Virtual Server Environment" article in this edition.

Content Management Layer

Various cloud services are currently being operated via the Internet and the quantity of content is growing rapidly. Here, "content" refers to a range of different unstructured data(b) including files, e-mail, images, and video. Because they are handled by separate applications, complex processing is currently required for interoperation between these different types of content. The primary role of the content management layer is to allow the unified management and utilization of this diverse content.

To implement this unified management of data efficiently, it is necessary to integrate and consolidate file servers, NAS (network attached storage), and other conventional devices as well as organize the data hierarchically as in lifecycle management and automate its administration. A new solution called virtual file platform (VFP), which incorporates these functions was released on the Japanese market in October 2010. VFP provides the high level of scalability and access performance needed to consolidate large quantities of content data. It also supports data access functions for a number of different protocols to allow access from a wide range of applications. To handle large quantities of data efficiently, it also provides various data system management functions including data management functions based on policy control (see Fig. 6).

Fig. 7 shows an example cloud on-ramp solution for data management on the cloud that combines VFPs with Hitachi Content Platform, which provides large-capacity file storage. The VFPs installed at individual sites or departments interoperate with Hitachi Content Platform at the data center to consolidate data automatically in a way that is transparent to the user. The solution also supports unified data management at the data center including backups, archiving, compression, and elimination of duplication, thus reducing the cost of storage capacity monitoring and administration at each site or department.

To facilitate future content management layer implementation, Hitachi is providing a comprehensive range of functions for utilizing data in more advanced ways, which include the software functions provided by VFP. The content management layer uses these functions to provide content virtualization, whereby the consolidation and management of data is independent of its physical location, and to allow data to be used in ways that are independent of applications. The content management layer also includes middleware for data retrieval, search, analysis, and utilization (repurposing) and can interoperate with applications (see Fig. 8).

(b) Unstructured data
Types of data such as text, e-mail, images, audio, and video unsuitable for storage in databases. In contrast, information such as accounting or sales data that can be stored and administered in a general-purpose database is called "structured data."

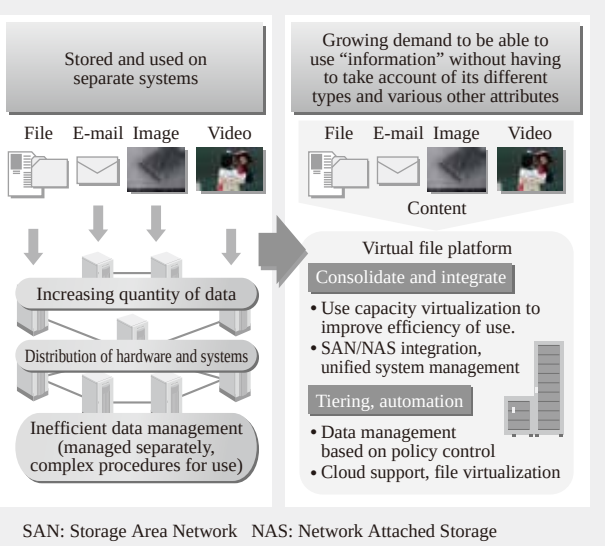


Fig. 6—Solution for Accumulation of Big Data.
Virtual file platform consolidates and integrates large quantities of content data to allow unified management and utilization of the data.

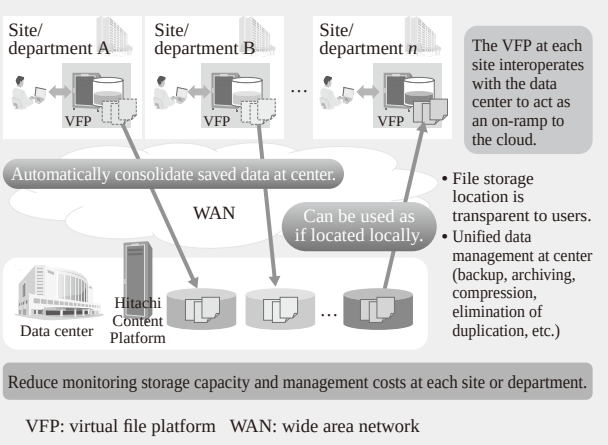


Fig. 7—Example Solution Using Virtual File Platform.
The file virtualization functions provide unified management of data on the cloud.

PUTTING BIG DATA TO USE

For companies and other parts of society to put big data to use and gain value from it, the data must be processed efficiently so that a result can be produced in a realistic length of time. In addition to the IT infrastructure layer of the platform described

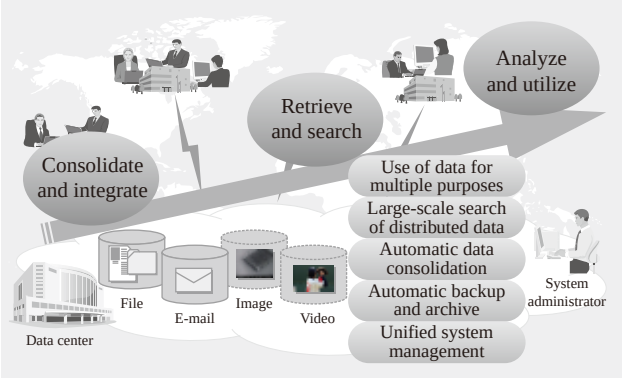


Fig. 8—Stages in Evolution of Functions Provided by Content Management Layer.
Hitachi intends to progressively add more functions to the content management layer so that data can be used in more advanced ways.

above being able to deliver performance capacity in a scalable way, this also requires software technologies that can utilize this performance efficiently.

One such technology is parallel and distributed computing. The open source Hadoop(c)*2 software has attracted attention in recent years as a platform for the parallel and distributed processing of big data. Examples in which Hadoop is applied to big data already exist, such as generating purchase recommendations from access logs at consumer web sites. Hitachi provides a support service for Hadoop in the Japanese market. Meanwhile, for applications to which Hadoop is not suited, Hitachi also supplies its own parallel and distributed computing platform.

Real-time performance is also required for purposes such as detecting abnormalities in plant and equipment using the never-ending flow of large quantities of sensor data, for example. The stream data processing function is designed for this purpose and can perform real-time in-memory processing of these large quantities of data as they are produced.

Hitachi is also undertaking research and development and other work aimed at overcoming the performance limits of commercial databases with respect to big data.

For the content management layer, Hitachi is also working on incorporating added-value middleware for data retrieval, search, analysis, and utilization, with the functional-layer software platform described here acting as a foundation.

(c) Hadoop
A software framework for the efficient distributed processing of big data developed and published by the Apache Software Foundation open source software community.
*2 Hadoop is a trademark of the Apache Software Foundation.

Meanwhile, when considering how best to deliver these solutions for using big data, it is anticipated that cloud services will be used in a variety of forms, both as a place to store this big data and as a way of providing the computing power required for big data processing. Interest in using the cloud has also been intensified by the recent Great East Japan Earthquake. For this reason also, Hitachi intends to direct its efforts toward providing cloud services aimed at the utilization of big data.

DATA DRIVES OUR WORLD

This article has described the form of the next-generation One Platform for All Data that Hitachi seeks to create for the utilization of big data, which will be at the core of future corporate and social innovation. Fig. 9 shows the future direction of Hitachi's platform solutions, which use One Platform for All Data as a base. As indicated by the slogan "data drives our world," we are moving toward a knowledge-based society in which data has the power to change the world. Using data in this way will achieve innovation and create new value in business and other parts of society.

By supplying platform solutions that support this sort of value creation, Hitachi aims to continue being the best solution partner for its customers. To

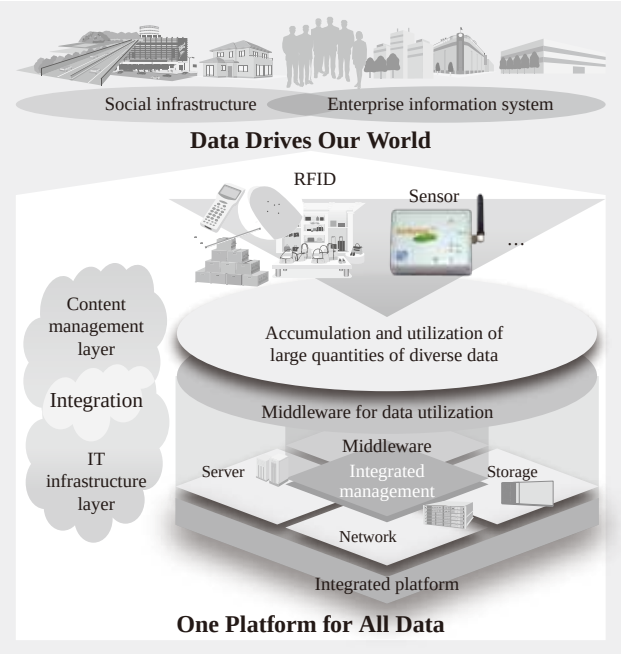


Fig. 9—Future Direction for Hitachi's Platform Solutions.
Hitachi aims to achieve innovation and create new value in business and other parts of society based on the next-generation One Platform for All Data designed for utilizing big data.

achieve this, Hitachi intends to continue preparing and developing the IT platforms that will turn One Platform for All Data into a reality through further advances made by bringing together the know-how and technologies acquired from experience with the supply of platform solutions based around servers and storage.

ABOUT THE AUTHORS

Hiroaki Fujii

Joined Hitachi, Ltd. in 1991, and now works at the Platform Business Strategy Planning Department, Strategic Business Development Division, Strategy Planning & Development Office, Information & Telecommunication Systems Company. He is currently engaged in the planning of platform business strategy. Mr. Fujii is a member of the Association for Computing Machinery (ACM), the IEEE, and the Information Processing Society of Japan (IPSJ).