

IT Platforms for Utilization of Big Data

Yasutomo Yamamoto

OVERVIEW: The growing momentum behind the utilization of big data in social and corporate activity has created a demand for the IT platform that has secure and reliable collection and archiving of large amounts of different types of data, fast and simple means for its extraction and analysis, high levels of reliability and expandability, and advanced techniques for automating its management. Having set out its “One Platform for All Data” vision for its IT platform business, Hitachi is preemptively developing technologies that will be needed in the future, while also supplying products and services that support its customers’ business operations.

INTRODUCTION

THE era of big data has ushered in a major trend toward the utilization of the large amounts of data held in corporate systems or on the internet, and the diverse variety of information produced by real-world companies, people, and other elements of society.

Hitachi has set out its vision of “One Platform for All Data” as the basis for its information technology (IT) platform business that supports both social innovation and this utilization of big data (see Fig. 1).

This vision involves the collection and archiving of large amounts of different types of data on Hitachi’s

integrated platforms that utilize servers, storage, networks, and open middleware. Through search, statistics, and analysis, this information can be utilized in applications such as social infrastructure and corporate information systems that in turn generate new data. Hitachi envisages this data cycle on a single highly reliable platform that supports integrated management and incorporates advanced functions such as high-speed data processing and virtualization.

This article describes IT platform solutions that support social innovation and the utilization of big data.

MIDDLEWARE FOR UTILIZATION OF BIG DATA

Critical to the utilization of big data is the collection, archiving, and distribution of large amounts of real-world data, followed by its utilization in specific applications.

Here, “collection” means the gathering of large amounts of different types of data for realtime monitoring and immediate processing, “archiving” means the efficient storage of these large amounts of collected data, and “distribution” means providing fast and simple ways of extracting and forwarding required data. The “utilization” of this “distributed” data then requires quick and easy ways of applying it, including through statistical and other analyses. Meanwhile, new forms of innovation are made possible by providing realtime feedback of the knowledge or other findings extracted from the results of statistics and analysis to corporate or public applications.

Hitachi has developed a suite of open middleware products that implement technologies for overcoming the challenges presented by the utilization of field data (see Fig. 2).

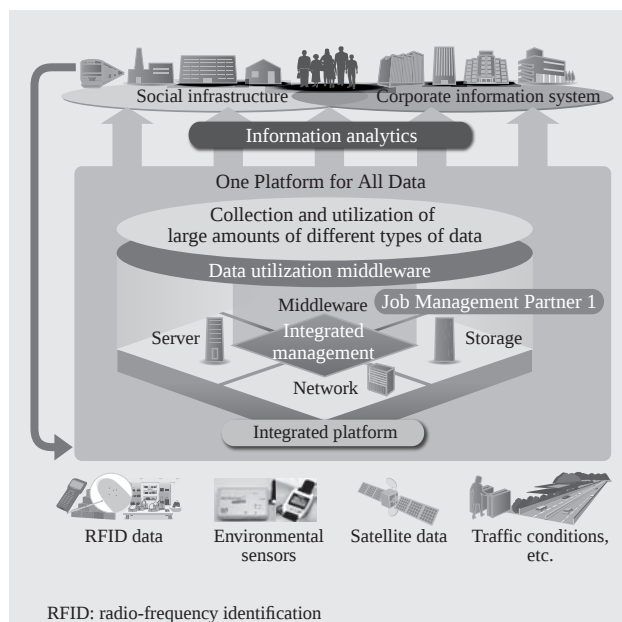


Fig. 1—“One Platform for All Data,” Hitachi’s IT Platform Business Vision.

A variety of data can be collected and archived on an integrated platform based on Hitachi hardware, and then the same platform can be used to retrieve, analyze, and utilize this data.

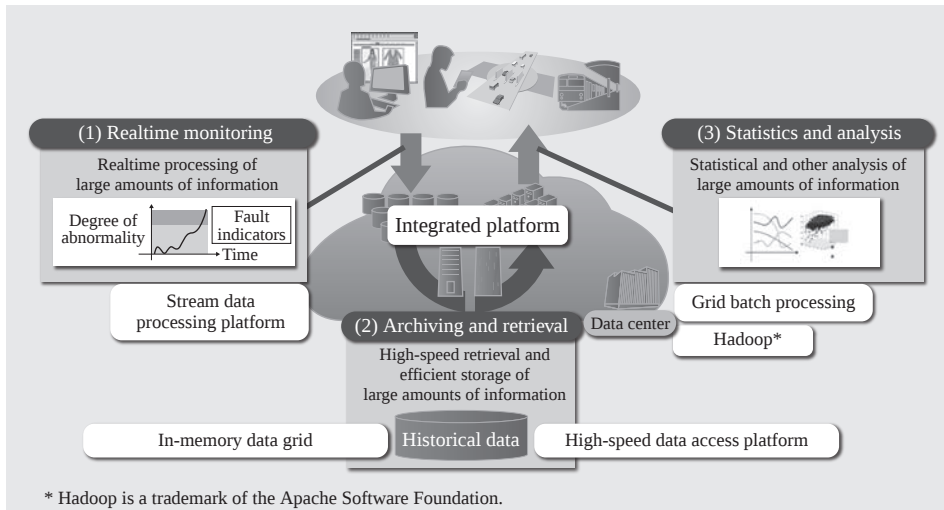


Fig. 2—Open Middleware for Utilization of Big Data. This suite of open middleware supports the requirements of big data utilization.

To meet the requirements for data collection, Hitachi supplies a platform for processing data streams and in-memory data grids that can process large amounts of data at high speed. Similarly, a high-speed data access platform^{*1} meets the needs of archiving and distribution, and Hitachi also supplies grid batch processing that accelerates the execution of existing batch programs to meet data utilization requirements.

High-speed Data Access Platforms

Hitachi is working with The University of Tokyo on the joint development of an ultrafast database engine that can maximize hardware performance. Technology produced by this work has already been incorporated into Hitachi products.

The ultrafast database engine uses an out-of-order execution principle^{*2} (the asynchronous processing of data without regard for the order in which input and output requests occur) that achieves highly parallelized execution by automatically splitting database retrieval

into blocks able to execute in parallel. This makes full use of the available server and storage capacity and delivers retrieval speeds that are 100 times faster than previous Hitachi systems.

This high-speed data retrieval performance allows applications such as the analysis of corporate data archived at a data warehouse to be performed ad hoc, without needing to set up a data mart, for example. This reduces the cost of batch processing and facilitates data utilization by providing the flexibility to analyze it in different ways (see Fig. 3).

^{*1} Utilizes the results of “Development of the Fastest Database Engine for the Era of Very Large Database and Experiment and Evaluation of Strategic Social Services Enabled by the Database Engine” (Principal Investigator: Prof. Masaru Kitsuregawa, The University of Tokyo/Director General, National Institute of Informatics), which was supported by the Japanese Cabinet Office’s FIRST Program (Funding Program for World-Leading Innovative R&D on Science and Technology).

^{*2} A principle devised by Masaru Kitsuregawa, Professor at The University of Tokyo and Director of the National Institute of Informatics, and Kazuo Goda, Project Associate Professor at The University of Tokyo.

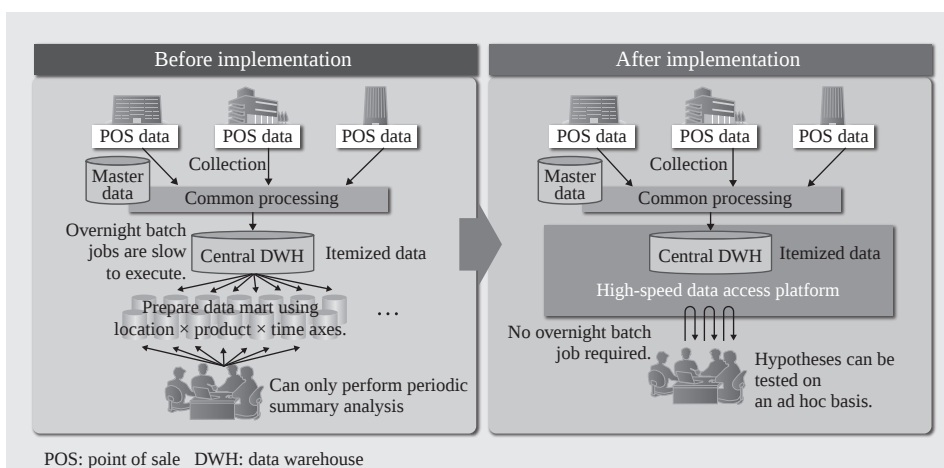


Fig. 3—Example Application of High-speed Data Access Platform.

The platform allows data to be analyzed ad hoc, without creating a data mart. This reduces the cost of overnight batch jobs and provides the flexibility to add different axes to the analysis.

The initial version of the high-speed data access platform was released in June 2012. This combined the ultrafast database engine with the Hitachi Compute Rack series and Hitachi storage systems that incorporate solid state drives (SSDs) and feature a high level of data access performance.

The second product, released in June 2013, was the DaTa SuperExpress^{*3}, a business intelligence (BI) appliance developed jointly with DTS Corporation. DaTa SuperExpress integrates DTS's BI NavigationStudio^{*4}, a BI tool featuring excellent ease-of-use, with the high-speed data access platform to create a data warehouse platform with ultra-high speed and large capacity, providing quick and simple support for the strategic analysis and utilization of big data.

In addition to working with customers and partners to work out best-practice models through measures such as conducting proof-of-concept (PoC) demonstrations of data utilization based on these high-speed data access platforms, Hitachi's future plans also include the supply of solutions and services targeted at overcoming specific business issues, such as a logistics analysis solution, for example.

New Integrated Data Platform

In the era of big data, the key to business development lies in the utilization of big data based on the collection, archiving, and distribution of large amounts of different types of data in ways that keep pace with changes in the business. Having up-to-date data is important for keeping up with what is currently happening in different corporate or public

^{*3} DaTa SuperExpress is a product name, trademark or registered trademark of DTS Corporation.

^{*4} BI NavigationStudio is a product name, trademark or registered trademark of DTS Corporation.

^{*5} TWX-21 is a trademark of Hitachi, Ltd.

situations, and for providing timely feedback to business operations. However, the productivity of activities like analysis and extraction that generate new value from big data is low, with time and effort being required for tasks such as re-modeling of the data being analyzed, redesign of data conversion, or revision of the data mart whenever the data needs to be analyzed from a different perspective. Similarly, collection of data from diverse sources inside and outside the company requires system reconfiguration work such as application software development or the installation of new hardware and software to handle specific data. As a result, making data more up-to-date is a problem.

Hitachi has provided a platform for the execution and management of collection, archiving, and distribution operations. Also planned for the future is the provision of products such as Global e-Service on TWX-21^{*5} in the form of a platform as a service (PaaS) that supports software as a service (SaaS).

This platform for the execution and management of collection, archiving, and distribution operations utilizes a high-speed data access platform for centralized management of data collected in a variety of formats, including structured, unstructured, semi-structured, non-realtime, and near-realtime. This provides fast and simple handling of sequentially generated data, together with planning that takes full account of what is happening in the field and the reliable transmission of these plans (see Fig. 4).

INTEGRATED PLATFORM FOR BIG DATA

Companies and other organizations hold data in diverse forms, such as databases, e-mail, images, and sensor data. The quantity of this data is growing explosively, currently in the petabyte (PB) and exabyte

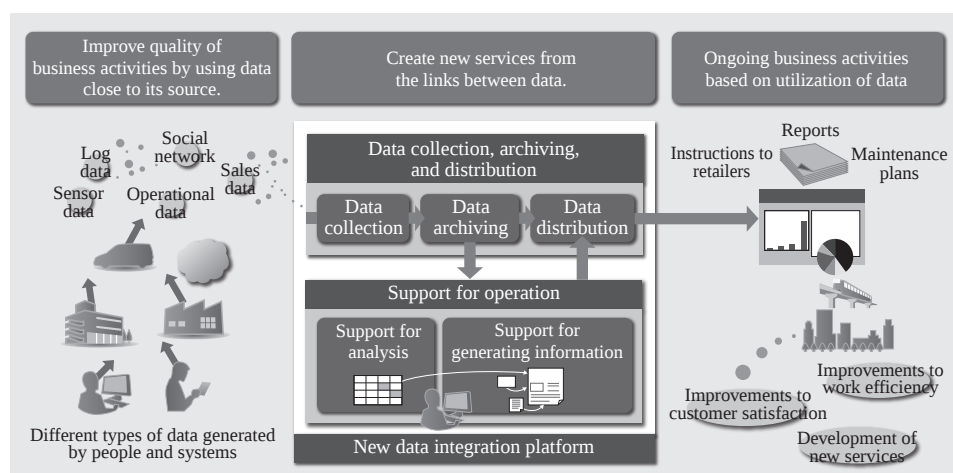


Fig. 4—New Data Integration Platform.

Hitachi supplies platforms that handle different types of data and support everything from collection to distribution.

(EB) range, and expected to reach the zettabyte (ZB) range in the future. Meanwhile, with corporate IT budgets remaining tight, there is a need to maximize customers' return on their IT investment by boosting their business operations through the use of high-performance IT platforms with advanced functions and low total cost of ownership (TCO) that can efficiently process these large amounts of different types of data.

Hardware Solutions for High-speed Processing

This growth in the quantity of data is driving demand for product enhancements, including improvements in the processing performance of server and storage products, larger memory sizes, and better energy efficiency.

In the case of servers, Hitachi is quick to release new products that take advantage of the latest processors, operating systems, and other technologies to support the high-speed processing of large amounts of data and the use of server virtualization for the large-scale consolidation of applications. This has included the development of a new Hitachi compute blade series (released in November 2012) that supports the latest Windows Server^{*6} 2012 operating system (OS) and features higher processing performance and memory capacity, and Hitachi Compute Rack series of personal computer (PC) servers (released in July 2013) that incorporate the latest Intel^{*7} Xeon^{*7} processor and deliver approximately 1.1 times the performance of previous models.

In the case of storage, Hitachi has expanded its range of products that use flash memory media, a form of storage that is recognized for its potential in the high-speed processing of large amounts of data.

In August 2012, Hitachi added a flash acceleration function to its Hitachi Virtual Storage Platform (high-end storage). By upgrading the operation of the storage controller to suit data processing on flash memory, the function achieves data reading rates of more than one million input/output operations per second (IOPS).

Meanwhile, Hitachi Accelerated Flash (HAF), a flash module developed using proprietary Hitachi technology that combines low cost and high performance, was released worldwide in November 2012. Thanks to a flash memory controller developed by Hitachi, HAF achieves approximately twice the data processing speed of previous models with an installation cost that is roughly 30% lower (see Fig. 5). With the release in July 2013 of a version of HAF intended for use with Hitachi Unified Storage VM, Hitachi has also started supplying new models that support an all-flash configuration.

Content Cloud Solutions

The data used within companies exists in a variety of forms, including structured data such as databases and unstructured data such as e-mail and documents. This data is spread across different sites and departments, and on systems such as staff PCs and mobile devices, and is continually increasing in volume. Hitachi supplies content cloud solutions that provide centralized management and easy access to content and other files in various different formats.

Hitachi Data Ingestor installed at an operational site or department acts as an on-ramp to the cloud, allowing for automatic consolidation and centralized management of site data on a Hitachi Content Platform storage system located at a data center. Users can access data without concern for whether it is held at their own site or at a remote data center. This reduces TCO and improves return on investment (ROI) by overcoming the problem of needing to administer

^{*6} Windows Server is a registered trademark of Microsoft Corporation in the United States and other countries.

^{*7} Intel and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

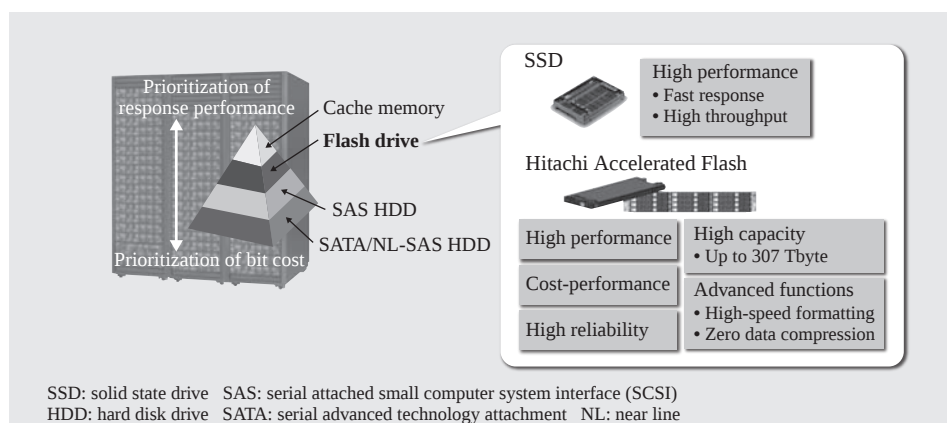


Fig. 5—High-speed Solution Using Flash Memory. This platform achieves high-speed data processing through use of Hitachi Accelerated Flash, a flash module based on proprietary Hitachi technology.

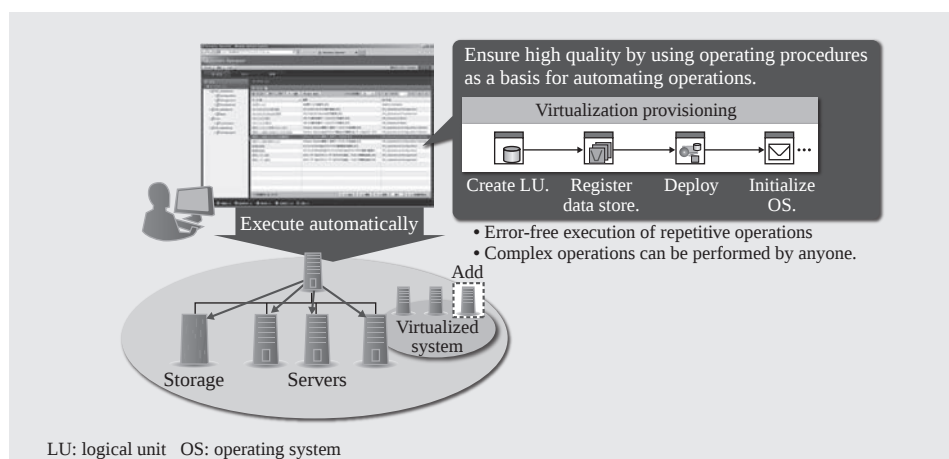


Fig. 6—Job Management Partner 1/Automatic Operation Product for Automating IT Operations. Job Management Partner 1/Automatic Operation reduces administrator workloads by executing IT operations automatically in accordance with operating procedures.

data separately at each site, and also the challenges of data management and capacity enlargement for the complex storage hardware required to cope with larger data volumes.

Outside Japan, Hitachi Data Systems Corporation released Hitachi Content Platform Anywhere in May 2013. Aimed at corporate users, this secure file synchronization and sharing solution for mobile devices makes it easy to share files between registered devices, enhances device security and reliability, and helps provide a better IT experience for staff.

Integrated System Management for More Efficient Administration

Urgent challenges faced by corporate IT departments include the growth in the size of information systems, their increasingly complex administration, expanding operational workloads, and tighter IT budgets.

Hitachi supplies a range of solutions for reducing the cost of IT system administration through its Job Management Partner 1 Version 10 integrated system management products. These include Job Management Partner 1/Automatic Operation for automating IT operation, the Job Management Partner 1 system monitoring service that incorporates system management technology and know-how developed through Job Management Partner 1, and the Job Management Partner 1 smart device management service that provides a SaaS application for handling security and asset management of a company's smart devices (see Fig. 6).

To reduce installation costs for IT systems, Hitachi supplies the Hitachi Unified Compute Platform, an all-in-one package that combines administration middleware with proven server, storage, and network hardware. The proven performance of these products shortens the time taken from installation until users

are able to start using them (the “service-in time”). The management software bundled with the platform also reduces operating costs by helping simplify and automate the operation of virtualization.

Furthermore, to provide for the quick and easy configuration of server virtualization on small and medium-sized systems, Hitachi released a simple virtualization model for the Hitachi Unified Compute Platform in July 2013 that incorporates server virtualization software.

In the future, Hitachi intends to help reduce the cost of IT systems for a wide range of customers by providing a greater diversity of configurations that combine software and hardware to suit specific end uses.

CONCLUSIONS

This article has described IT platform solutions that support social innovation and the utilization of big data.

The aims of these solutions are to use IT to process real-world data in order to determine the current and past situation on the ground, to enhance people's creativity, and to provide a view of the future. Through the accurate processing of data collected from the field, and by using it to provide feedback to the processes of public and corporate activity, it will be possible to create new forms of value.

ABOUT THE AUTHOR



Yasutomo Yamamoto
Platform Strategy, Strategy Planning & Development
Office, Information & Telecommunication Systems
Company, Hitachi, Ltd. He is currently engaged in
strategy planning and the development of IT platforms.