Big Data Analysis Solutions for Driving Innovation in On-site Decision Making

● Tooru Shibata ● Yoichi Kurachi

Information systems have traditionally been constructed and used for increasing corporate value, developing competitive products, improving employee productivity, and accelerating the decision-making process. Today, however, dramatic improvements in hardware performance, advances in various types of sensors, and changes in the social environment such as consumer-driven dissemination of information via social networking services (SNSs) are generating a great need for "big data" analysis. In its early stages, the use of big data depended on information and communications technology (ICT) specialists and expert analysts located at support sites. Now, to speed up decision-making at sites directly involved in increasing sales and lowering costs, analysis based on big data is being performed at these frontline sites, and some marketing and planning departments have already begun to base their actions on big data. Moreover, high-function middleware that facilitates such big data usage has been appearing. However, issues arising in actual implementations of systems with such middleware have hindered advanced decision-making in the field. This paper introduces Fujitsu's approach to resolving these issues by providing integrated solutions that optimally combine big-data middleware and analysis scenarios and describes the effects of introducing these solutions.

1. Introduction

In the corporate world, information systems have traditionally been constructed and used for increasing corporate value, developing competitive products, improving employee productivity, and accelerating the decision-making process. Today, however, quantum leaps in hardware performance, the evolution of software technology for handling mass amounts of data, advances in various types of sensors, and changes in the social environment such as posting of consumer-generated information via social networking sites (SNSs) are increasing the need for "big data" analysis.

However, there are problems in implementing technologies for using big data: in addition to keeping up with advances in individual technologies, it is also necessary to integrate the benefits of multiple technologies to create a combined effect that includes ease of use for users in frontline business departments such as sales and marketing.

Fujitsu big data solutions put high value on ease of use through integration of analysis scenarios with middleware. The scenarios incorporate the data-related know-how that Fujitsu has obtained through its experiences in constructing actual information systems. These solutions can be quickly put to use by users.

This paper describes how Fujitsu's integrated solutions can solve the problems that arise in implementing a big-data system in a company and how these solutions can help grow a business.

2. Technology trends

The exponential growth in online sales means that many purchase histories are becoming available on e-commerce sites, and the growing use of social media means that a huge amount of information is being posted by consumers themselves. These trends, along with the growing use of smart devices, smart cards, and other types of devices, are enabling a massive amount of data to be collected. In short, the amount of information that can be used to extract knowledge useful for business is increasing at an explosive rate. The amount of available data is expected to reach 44 ZB

(1 zettabyte = 10^{21} bytes) by 2020 (**Figure 1**) .

Along with this dramatic increase in available data, the performance of CPUs, memory devices, and other types of hardware is likewise increasing at a rapid rate (**Figure 2**).

Another distinctive trend is the evolution of software technologies, including database infrastructure technologies, data processing technologies, and data representation technologies.

1) Database infrastructure technologies

• Column-oriented database

This type of database stores data in units of columns without duplication, thereby enabling efficient storing and retrieving of data for use in data warehouses.

• In-memory database

This type of database stores all data in main memory, thereby eliminating the need for disk access, resulting in high-speed data processing.

• NoSQL database

A NoSQL database stores only pairs of keys and values, for example. Some operations can thus be performed faster than with a relational database. Such databases are thus well suited for handling massive amounts of data, such as big data. NoSQL stands for "not only SQL." A NoSQL database can store data in various structures in addition to the SQL structure.

2) Data processing technologies

• Machine learning

Machine learning means the automatic uncovering of relationships or other latent features in the results generated from data (learning). These features are then used to forecast future events.

• Parallel distributed processing

Parallel processing on distributed platforms enables high-speed processing of large amounts of business data and computer log data.

• Complex event processing

Complex event processing (CEP) performs real-time analysis and evaluation of event information generated continuously from log data, sensor data, etc.

• Extreme transaction processing

Extreme transaction processing is high-speed, high-reliability processing of a massive (extreme) amount of transaction data, such as data from online trading and e-commerce.

3) Data representation technologies

• Advanced data visualization (ADV)

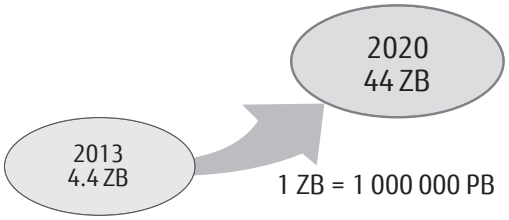
ADV presents data in a manner that is simpler and easier to understand than with conventional business intelligence tools. It uses advanced graphics and a wider variety of analysis axes and can perform high-speed, in-memory execution.

This evolution of software technologies in addition to the evolution of hardware means that data processing of a type that could not be done in the past—such as the direct handling of large quantities of itemized data—can now be done at a reasonable cost and in a practical period of time. The era of big data has truly arrived.

The optimal combination of these new software technologies is enabling big data to make real contributions to business growth.

3. User case

Company analysis of data accumulated either internally or externally has been constrained by the amount of data and the time needed for processing



Data source: IDC White Paper
"The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things" sponsored by EMC Corporation (April 2014)

Figure 1
Growth in available data.

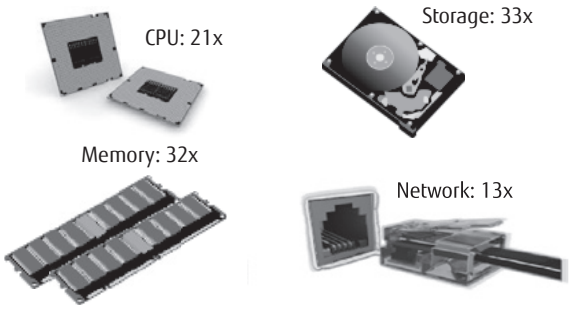


Figure 2
Increase in hardware performance of Fujitsu products from 2003 to 2013.

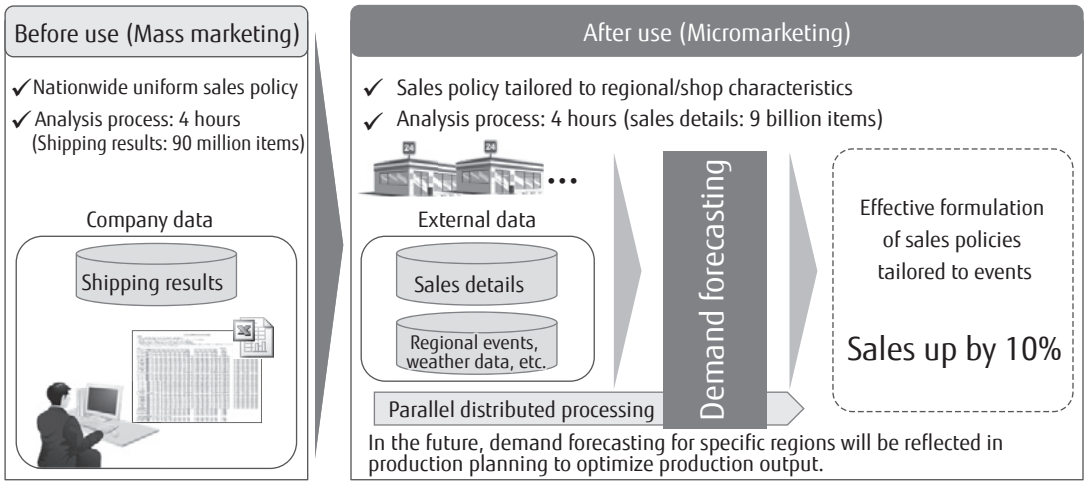


Figure 3
Case study of using big data.

that data, which is a significant problem when the volume of data spikes. In addition, complicated analysis methods have given companies no choice but to depend on the skills of specialists to execute analyses, and the results of analyses have had to be appropriately summarized to produce reports that can be easily understood. Furthermore, the analysis of large quantities of data can take up a considerable amount of time and can result in a great number of reports that would be difficult for anyone to review in total. All of these problems have made it difficult for frontline departments to make quick, on-site business decisions based on data analysis.

However, as described above, the creation of environments that can realistically use big data is now making it possible for frontline departments to quickly perform effective data analyses for business purposes.

To give an example (**Figure 3**), sales staff at a beverage manufacturer can now perform detailed promotions at each sales outlet in their area through demand forecasting that takes into account not just detailed sales data from those retail stores but also external data such as regional weather and events. In the past, this beverage manufacturer performed nationwide mass marketing[note 1)] on the basis of in-house wholesale shipping data (90 million items). Now, in addition to detailed sales data (9 billion items) from those sales outlets targeted for wholesale shipments, this company has been performing demand forecasting that adds information on regional weather and events. As a result, it has been able to increase sales by 10% by proposing ways for sales outlets to arrange and display its products so as to attract customers attending nearby events. This is a strategy that could not previously be adopted with conventional mass-marketing techniques. To give a specific example, this company has been able to increase sales by switching product promotion to low-calorie drinks popular with women on the day of a concert that a large number of women are expected to attend. This is a case study of a frontline department that has achieved micromarketing[note 2)] and expanded business by introducing a big-data analysis environment.

By setting up such big-data usage environments in this way, Fujitsu has enabled users in frontline departments to use big data in a routine, matter-of-fact way as part of their daily work and to make quick decisions at the frontline of business. Fujitsu believes that this approach can revolutionize a company's decision-making process, improve its on-site capabilities, transform the workplace, and optimize business tasks.

note 1) A marketing method by which a single, mass promotion is directed at the general consumer. This method can reach many consumers using mass media such as television, radio, and newspapers, but it cannot address the particular needs of individual consumers.

note 2) A marketing method by which promotions are directed at small, specific groups of consumers, such as ones living in a specific region. Targeting consumers in this way can greatly increase sales.

4. Problems in implementing system using big data

Although there are great expectations for using big data in frontline business departments, there are several difficulties that arise in the actual implementation of a system using big data.

1) Difficulty of introducing new technologies

As described in Section 2, the evolution of various technologies including database technologies and data processing technologies has enabled the processing of big data. However, to understand and integrate these individual technologies and to use big data with sufficient ability, a company needs technically competent personnel for each of those technologies and the middleware on which they are implemented (Problem 1).

2) Difficulty of defining system requirements

Frontline business department users are not necessarily skillful in analysis techniques and have no other alternative but to refine the types and amounts of data processed and the analysis method used through a process of trial and error. Defining system requirements beforehand as is done in conventional business systems is consequently difficult (Problem 2-1). It is also necessary to see that both the data targeted for processing and the analysis method used evolve in step with changes in the business environment (Problem 2-2).

3) Difficulty of estimating effects of implementation

While it is possible to show the effects of implementing new technologies in a business system through work-efficiency gains and other parameters, it is often difficult to describe the return on investment of a new information system. This is why obtaining approval for investing in the creation of a big-data environment tends to be difficult (Problem 3).

5. Solutions for using big data

Fujitsu provides FUJITSU Business Application Operational Data Management & Analytics as an integrated solution for solving the problems described in the previous section and for enabling frontline business departments to make effective use of big data.[1)]

The Operational Data Management & Analytics product integrates analysis scenarios developed by Fujitsu in actual projects and the middleware needed to use and apply big data. This product obviates the need to obtain a deep understanding of individual technologies since it integrates, optimizes, and provides the technologies needed for using and applying big data in combination with those analysis scenarios. The user need only select an analysis scenario and define a data source and operation policy to begin using big data, which easily solves the problem of having to learn and use new technologies (Problem 1).

There are three main elements in the Operational Data Management & Analytics product (**Figure 4**).

1) Operational analytics

Operational analytics (**Figure 5**) provides a solution for accelerating the use and application of big data in a frontline business department. It constitutes a self-service type of data usage environment in which frontline users can make use of analysis results on their own. This is made possible by automating the various data analysis procedures developed by Fujitsu in actual projects and providing them to users in the form of analysis scenarios. As a result, the burden of analysis is greatly reduced. Providing the functions needed for data analysis in an all-in-one configuration significantly shortens the time needed for system implementation. In short, even a business department without specialized knowledge in analytics can make good use of analysis results in a relatively simple manner.

Providing analysis scenarios that can be directly applied enables users to define system requirements through "fit/gap analysis," thereby solving Problem 2-1.

2) Operational data management

Operational data management (**Figure 6**) provides a big-data analysis platform that enables real-time use of information in a frontline business department. This includes mechanisms for integrating and managing various types of data including structured data, raw on-site data, and external data and for analyzing that data in an all-in-one configuration. The platform incorporates database-infrastructure and data-processing technologies such as column-oriented data storage, parallel distributed processing, and CEP, which can be used as needed depending on the type of data processing desired. In addition, Fujitsu's proprietary data-confidentiality protection technology, which can be applied in units of data items on the basis of user access rights, makes for high-speed and safe information sharing.

The above features enable all sorts of data to be managed and processed at high speed, which makes it

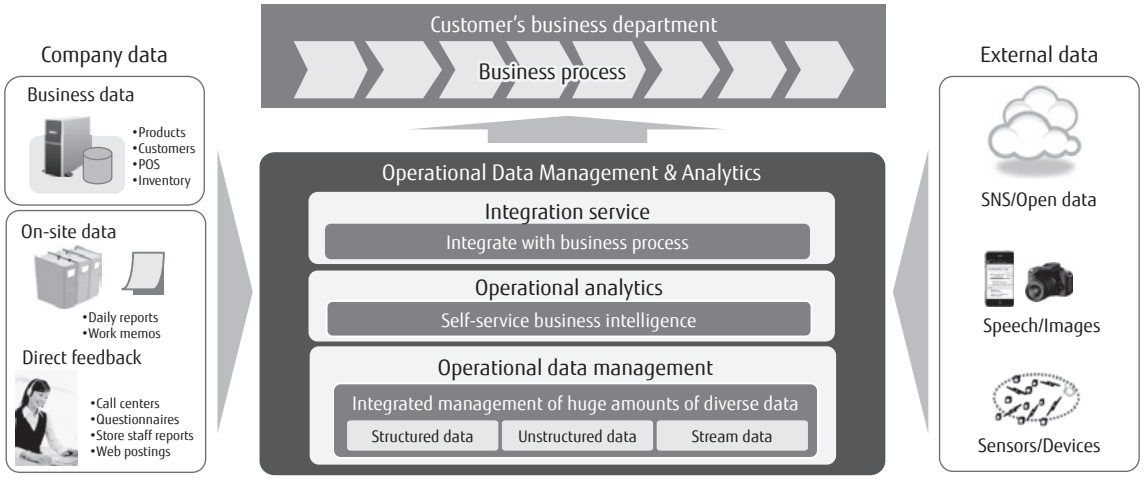


Figure 4
Three main elements of Operational Data Management & Analytics.

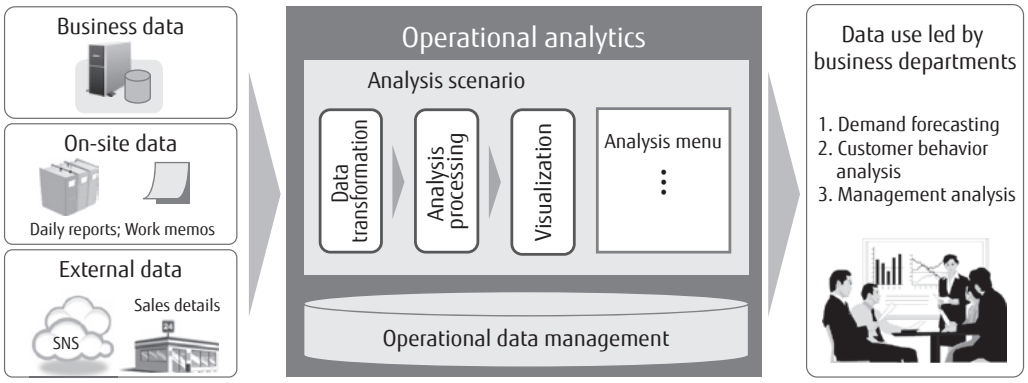
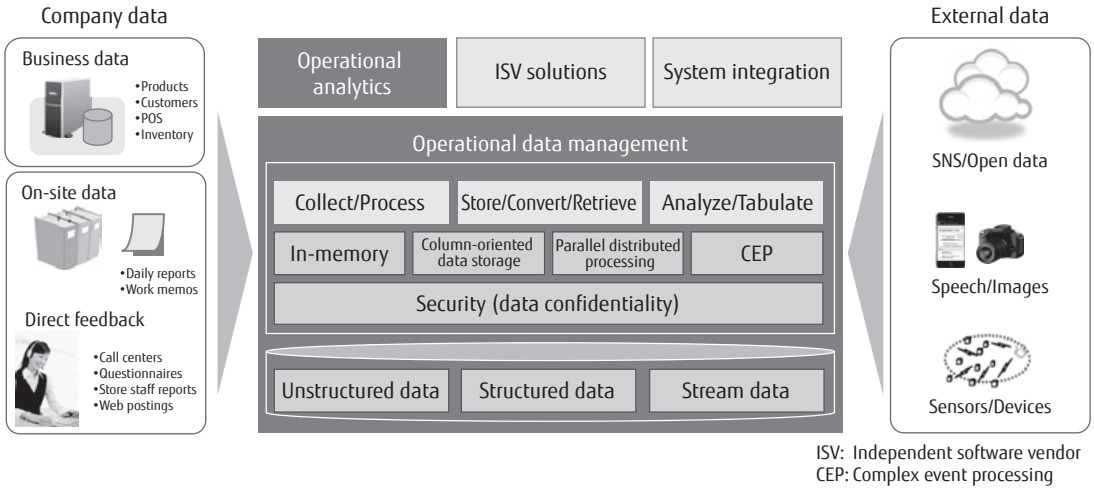


Figure 5
Operational analytics.



ISV: Independent software vendor
CEP: Complex event processing

Figure 6
Operational data management.

possible to adopt an approach of step-by-step system extension after initial deployment, thereby solving Problem 2-2.

3) Integration service

The Operational Data Management & Analytics product includes an integration service that provides a full range of support from consulting to operation for business departments that wish to use and apply data in new ways. This service proposes optimal analysis systems to fit the customer's new information-usage needs and provides ongoing support that matches the evolution in the way that the customer uses the information.

Specifically, in addition to design, start-up, and operation/management support, this integration service can add analysis scenarios and new types of data as the customer's analysis requirements evolve. It also incorporates a service that supports the customization of big-data usage and a service that leverages Fujitsu's experience in system construction to quickly achieve linking with external systems.

Operational Data Management & Analytics is provided as a product that fully integrates the above three elements for on-premise use. However, it can also be provided via software as a service (SaaS) to enable the customer to minimize system startup costs.

Applying this service via SaaS enables actual operation to commence once those effects have been tested, thereby solving Problem 3. In this way, the customer can deploy a system that provides true value to its business.

As described above in the section on operational analytics, the Operational Data Management & Analytics product integrates the analysis know-how that Fujitsu has accumulated through actual projects into a solution that a frontline business department can use immediately. The product is initially targeting three types of business models, as described below.

1) Demand forecast model

Targeting the sales departments of consumer goods manufacturers, this solution supports advanced demand forecasting using external data such as sales details from retail stores as well as regional event information and weather data.

While conventional mass marketing is based on wholesale shipping data, this solution supports a micromarketing approach in which detailed sales policies are formulated in accordance with demand forecasting on a store-by-store basis. Furthermore, in addition to providing analysis scenarios that the sales department of a consumer goods manufacturer can easily put to use, it comes equipped with a parallel distributed processing mechanism[note 3)] that supports highly accurate forecasts at high speed. It also supports more than 200 types of sales data, enabling quick on-site implementation[note 4)].

2) Customer behavior analysis model

Targeting the marketing departments of retail stores, this solution supports detailed purchasing behavior analysis using conventional purchasing data on customers, products, etc. as well as external data such as information from sales personnel on their interaction with customers, information from SNSs, etc.

This solution provides 32 types of purchasing-behavior analysis scenarios that are of most interest to the marketing departments of retail enterprises. It enables the use of data that include natural language, such as in customer-servicing reports from the sales floor and other types of external data by applying the Japanese-language analysis technology of FUJITSU Software ATLAS, which Fujitsu has been enhancing for many years. In contrast to uniform customer servicing, which has been the norm up to now, this solution enables detailed customer interaction tailored to the interests and preferences of individual customers.

---

note 3) In addition to providing parallel execution for multiple forecast algorithms, this mechanism can automatically decide which forecast algorithm is optimal for each type of product, thereby enabling high-accuracy demand forecasting to be performed in a short period of time. For example, to perform demand forecasting in units of stores or products for 9 billion items of detailed sales data from 20 000 stores, a conventional system would require 400 hours while this mechanism would need only 4 hours to complete the task.

note 4) Supports as standard more than 200 types of detailed sales data in different formats such as those used by chain stores, thereby negating the need for application development or master-data design, which has traditionally been extremely time consuming. Supports weather data and event data as well, and can greatly reduce system implementation time from the usual 3 to 6 months to about one month.

3) Management analysis model

This solution targets the sales departments of globally expanding manufacturers and supports the sales activities of overseas bases by assessing the business conditions at those bases in real time. It provides 24 types of globally uniform analysis scenarios such as profit-and-loss analysis for regional, base, and customer segments. By converting data on sales and on sales-related activities from each base into an international standard code, this solution enables globally uniform analysis and helps to fortify sales by linking bases with sales policies.

This model has been applied in Fujitsu's global sales department and has contributed to expanding sales in global markets by sharing information about information on sales-related activities with overseas bases on a daily basis and supporting sales activities at each base.

Fujitsu is presently rolling out these three business models for the marketing and corporate-planning fields for which big-data usage is attracting the most attention. Going forward, it will continue to release more models for many types of businesses and business tasks.

6. Effects of introducing solutions

Having frontline business departments use Operational Data Management & Analytics on their own has enabled them to quickly plan and execute a variety of actions based on analysis results that are appropriate to frontline operations. There are several examples of how the use of this product has led to a growth in business.

1) Marketing/Planning: Increased sales due to formulating sales policies and optimizing promotions for each region and store.
2) Store management: Increased sales due to optimizing the floor layout and enhancing customer interaction through customer behavior analysis.
3) Sales activities: Increased sales due to providing individual customers with optimal recommendations.
4) Production management: Optimized production due to forecasting demand (reduce lost opportunities in accepting orders and decrease excessive inventory).
5) Facility maintenance: Optimized maintenance costs due to reducing maintenance frequency.
6) Inventory management: Optimized inventory due to increasing the accuracy of demand forecasting (reduce out-of-stock losses and decrease waste).

Example 6) relates to the actual use of big data by Fujitsu itself. In this case, enhancing the accuracy of demand forecasting through big data analysis resulted in a reduction in parts inventory by more than 15%.

The results in each of the above examples reflect the introduction of Operational Data Management & Analytics, which integrates the scenarios and middleware needed for data analysis. The all-in-one configuration of this product means that it can be easily implemented without having to depend on expert engineers. In addition, analysis-scenario and analysis-menu operations enable the user to understand what can be done, which means that the definition of system requirements can be made less troublesome by using fit/gap analysis. Furthermore, using Operational Data Management & Analytics in a SaaS environment enables system startup costs to be minimized through a "small start" and enable the use of Operational Data Management & Analytics while testing effects.

7. Middleware supporting use of big data

Fujitsu has been providing middleware products for using and applying big data for some time. These products incorporate the practical knowledge and technologies that have come to be used for operating cloud services related to data usage and provide a high level of connectibility and reliability for use in a corporate system. They add Fujitsu's high-reliability and high-performance technologies that have been nurtured in customers' business systems to new open software technologies.[2)]

The following introduces the features of the Fujitsu middleware products that were integrated into the Operational Data Management & Analytics product to improve its speed of operation and ease of use.

1) FUJITSU Software Interstage Big Data Parallel Processing Server[3)]

Based on Apache Hadoop, this parallel distributed processing middleware (**Figure 7**) greatly improves system reliability and performance. It is capable of processing huge volumes of data through a scale-out configuration using general-purpose servers,

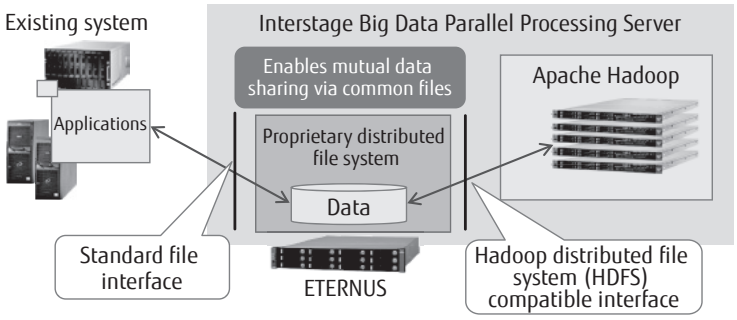


Figure 7
Features of Big Data Parallel Processing Server.

and, in conjunction with the cluster technology refined in FUJITSU Software PRIMECLUSTER, it implements a proprietary distributed file system that improves system reliability even further. This distributed file system uses the high-reliability shared disks of FUJITSU Storage ETERNUS instead of server-mounted disks. It features redundancy at the hardware level and high-speed switchover in the event of a server failure.

The most outstanding feature of this distributed file system is its high processing performance when linked with a business system. There are situations in which bottlenecks can form in data transfers between a business system and a Hadoop system in large-volume data processing. With this product, however, files created in a business system can be used as-is by Hadoop, thereby eliminating such bottlenecks. Testing using a test model showed that batch processing time, including the time for data linking, can be reduced by up to 80% in some cases.

By applying this product, Operational Data Management & Analytics achieves a high-reliability and high-performance analysis environment for big data while improving linkage with business systems.

2) FUJITSU Software Interstage Big Data Complex Event Processing Server[4)]

This is a CEP middleware product (**Figure 8**) with high performance that is easy to use. It achieves high processing performance (about four times conventional values) by combining a high-speed filter that applies Fujitsu's proprietary high-speed data matching technology with an open-standards CEP engine. It also enables easy linking with master data used in business.

In general, an event passed on from a device incorporates a code value such as a product code as

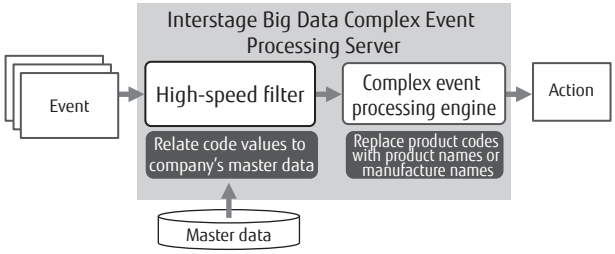


Figure 8
Configuration of Big Data Complex Event Processing Server.

opposed to a product name. Code values, however, can make it difficult to formulate event-processing rules that are easy to understand. It is possible, though, to relate such code values to the company's master data used for business so that product codes can be replaced with product names or manufacturer names, enabling clear and understandable rules to be formulated. Moreover, as the high-speed filter is also used in this integration process, master-data linking, which could easily degrade performance, can be achieved in a straightforward and high-speed manner.

By applying this CEP middleware, Operational Data Management & Analytics achieves real-time analysis of large volumes of data received from sensors and devices.

3) QlikView[5)]

This ADV product features intuitive operations targeting general users in frontline business departments that usually have no information and communications technology (ICT) specialists. It also features in-memory data management, thereby eliminating the need for constructing data marts in data storage. The QlikView product is therefore easy to implement, enabling

frontline users themselves to use and apply big data as needed.

Operational Data Management & Analytics provides templates that enable users to customize this intuitive QlikView interface as they see fit.

8. Future developments

Looking to the future, Fujitsu plans to meet diverse needs by providing a full lineup of solutions for using big data. These solutions will incorporate various types of know-how that Fujitsu has so far accumulated as data analysis scenarios and will combine them in optimal ways for different types of businesses and business tasks. The plan is to support the use of big data in new fields such as the Internet of Things (IoT) and digital marketing for such purposes as detecting/predicting changes in equipment and facilities that use information from sensor devices, promoting innovation in manufacturing, and analyzing customer experiences using image processing technology.

Fujitsu also aims to achieve an advanced information integration environment that goes beyond information systems to include even business systems. In this environment, the focus will be on the use of information from the viewpoint of all corporate activities and the linking of intra-company and inter-company processes through data. Although intra-company business processes have become highly efficient through ongoing efforts to raise efficiency, they are also vertical in nature. Fujitsu believes that the information-usage solutions presented here can closely link business processes through information and thereby achieve an environment in which frontline business departments themselves can use visualized data to enhance corporate business processes. By putting this belief into practice, Fujitsu aims to continuously apply the value in big data and in information in general to formulating a company's business strategies as well as to transforming frontline business processes. The overall goal is to contribute to greater speeds and efficiencies in corporate activities.

9. Conclusion

Fujitsu is focused on "data value," which originates in the actions of people and the movement of things in the real world. In the use of big data, it places prime importance on the derivation of new value from huge amounts of data that could not be collected or fully used in the past.

Going forward, Fujitsu plans to develop and provide information-usage solutions through co-creation and joint implementations with customers. These solutions will provide value to frontline business processes in the form of forecasts and knowledge obtained from the analysis of big data. In this way, Fujitsu is committed to supporting frontline business departments in making real-time decisions based on big data.

References

1) Fujitsu: Fujitsu Provides Big-Data Solutions to Transform the Front Lines of Business.
http://www.fujitsu.com/global/about/resources/news/press-releases/2014/0513-01.html
2) Fujitsu: FUJITSU Software Big Data.
http://www.fujitsu.com/global/products/software/middleware/application-infrastructure/interstage/solutions/big-data/
3) Fujitsu: FUJITSU Software Interstage Big Data Parallel Processing Server.
http://www.fujitsu.com/global/products/software/middleware/application-infrastructure/interstage/solutions/big-data/bdpp/
4) Fujitsu: FUJITSU Software Interstage Big Data Complex Event Processing Server (in Japanese).
http://interstage.fujitsu.com/jp/bigdataceps/
5) Fujitsu: Fujitsu Enhances Big Data Analytics Solution to Facilitate Decision-Making on the Front Lines of Business.
http://www.fujitsu.com/global/about/resources/news/press-releases/2014/0814-01.html

Tooru Shibata
Fujitsu Ltd.
Mr. Shibata oversees the solutions business in the field of business innovation.

Yoichi Kurachi
Fujitsu Ltd.
Mr. Kurachi is engaged in strategy formulation and planning in the field of business innovation centered about the use of big data.