Technologies for Real-time Use of Information through Integration of Diverse Forms of Internal and External Data

● Kazushige Nakamura ● Shinya Echigo ● Takehiko Aoyagi

As markets continue to diversify, the amount of internal and external data and the amount of processing required are growing explosively, and the quality of the target data is changing as well. This situation is creating a demand for systems that can quickly and easily analyze and use massive amounts of data so that market needs can be accurately assessed and sound business decisions can be made. FUJITSU Integrated System Analytics Ready is a hardware appliance product integrating a series of functions that can extract the data deemed necessary from a huge amount of stored data and analyze and use that information. These functions include "easy-to-implement system," "high-speed database as a platform for analyzing and using data," "operational data storage for storing huge amounts of data by using simple settings," "data linking for collecting data from diverse data sources," "data processing by using an intuitive graphical user interface (GUI)," and "safe sharing of consolidated data." The last one is the most needed function because it protects stored data from a variety of threats without a loss in customer convenience. This paper introduces technologies for high-speed and safe use of data in response to market needs.

1. Introduction

Today, the business environment is becoming increasingly diversified, and the volume of information that is being generated is of massive proportions. Under these conditions, business managers need a means of analyzing and using this huge amount of information to quickly uncover buried knowledge that could be used to grow their businesses. To meet this need for analyzing and using a massive amount of diverse types of information in business, Fujitsu released a hardware appliance product called FUJITSU Integrated System Analytics Ready.[1)] It is a platform that is easy to deploy and operate and that enables information to be analyzed at high speed and put to use quickly.

In this paper, we describe the configuration of Analytics Ready and introduce the platform technologies that simplify its deployment and operation and that enable high-speed analysis and use of information.

2. Information-usage needs

Today, a company seeking to grow its business must be able to use not only internal data stored within the company but also various forms of external data including "open data." Furthermore, to make appropriate use of such diverse forms of external data, there is a need for a data storage function[2)] that can link that data with internal data and store it "as-is."

The need has also arisen for collecting and combining internal and external data to extract knowledge and to accelerate management decision-making by executing that cycle repeatedly. This can be accomplished through a function that can process data in an intuitive manner and a function that can tabulate results at high speed.

3. Configuration of Analytics Ready

Analytics Ready provides a series of functions, from data collection by data linking to data usage, that are needed for using information as described in the previous section. Analytics Ready consists of the following components (**Figure 1**).

1) Data linking

The function of this component is to collect data by linking with a variety of data sources and to store

that data. This series of data linking and collecting processes can be defined using Microsoft Excel, a program familiar to most users. Such a definition enables data stored in databases and files on different platforms to be collected at optimal times.

2) Data storage

An operational data store (**Figure 2**) is used to store and manage the diverse forms of data collected by data linking. Using FUJITSU Software Interstage Shunsaku Data Manager[3)] technology cultivated by Fujitsu over many years, this operational data store serves as a data depository that can directly store comma-separated-value (CSV) files without having to design complicated schemas or create indexes and that can perform high-speed data searching.

It can also store and search CSV files containing a different number of data items and different data formats without having to create special applications as long as headers are attached to those files. Additionally, CSV files with a data format that has been changed can be accommodated by simply modifying their headers

Furthermore, Fujitsu's proprietary fjcx high-compression technology can be used to compress data to 1/10–1/25 its original size (compared to 1/2–1/5 using the general gzip format), which makes for more effective use of disk resources.

3) Data processing

This function provides key processes, namely extracting data from the operational data store, processing the various forms of data stored therein, and storing the results of that processing in a data warehouse (Figure 2) in the form of components that can be interconnected and combined by drag-and-drop operations on a graphical user interface (GUI). This makes

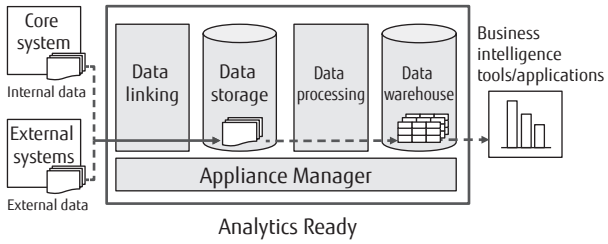

Figure 1
Configuration of Analytics Ready.

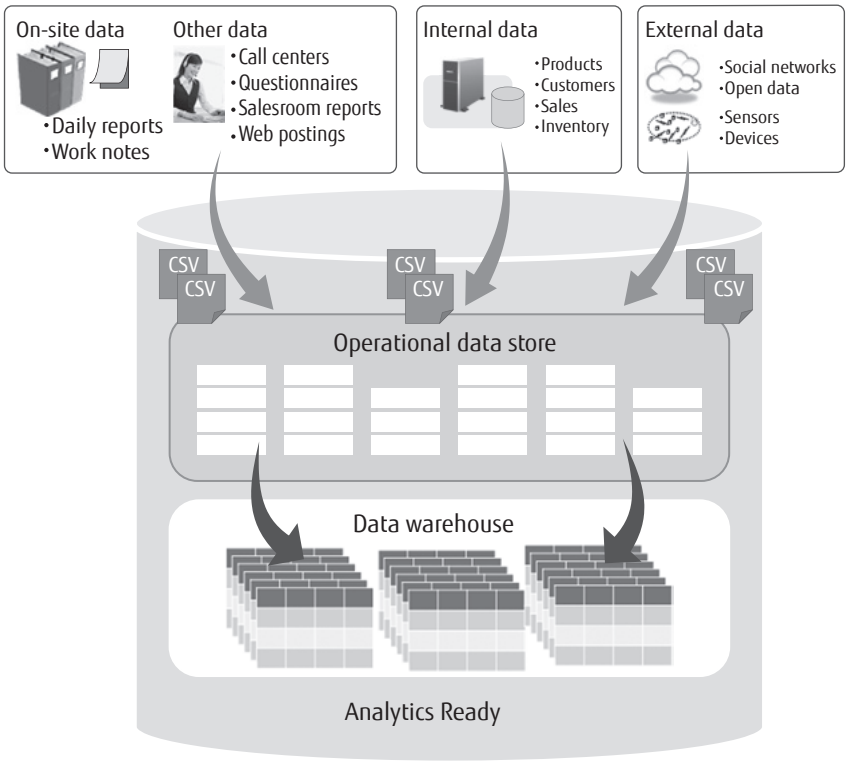

Figure 2
Roles of operational data store and data warehouse.

it easy for the user to define a series of processes as a data processing flow (**Figure 3**). Also provided are data-processing templates that define frequently used processing flows. These templates can be customized as the need arises, thereby making it even easier for the user to define a desired processing flow.

Processing flows defined in the above way can be saved and reused in Analytics Ready, which enables frontline staff who analyze and use data in the field to share their know-how. In contrast to the conventional approach of relying on a company's information systems department to analyze and use data, this data processing function in Analytics Ready enables frontline staff themselves to process the data they need and to store that data in a data warehouse. In short, this function enables data to be quickly analyzed and put to use at the front lines of business.

4) Data warehouse

The Analytics Ready data warehouse incorporates a high-performance query engine that can handle a columnar table (in which columns are managed separately and independently for storing data). This enables data to be accessed in the column direction, which is optimal for tabulation processing.

Storing data in a columnar table means that only those columns of data that are currently required for tabulation purposes need to be accessed. This can reduce I/O processing by 90% on average (based on actual measurements by Fujitsu). Compared to ordinary relational databases, this can improve processing performance by up to 500 times (based on actual measurements using a Fujitsu model) so that tabulations of huge amounts of data that would normally take several hours to process can be completed within several seconds (**Figure 4**).

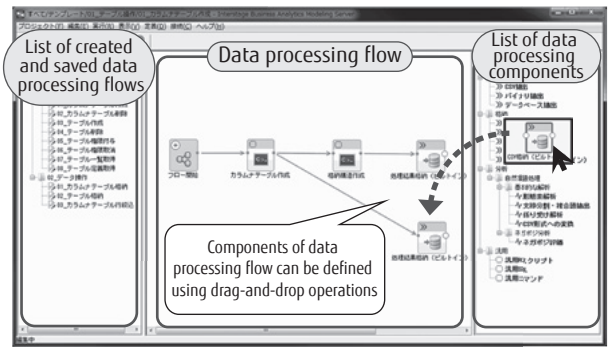

Figure 3
Data processing function.

This component of Analytics Ready also prepares indexes applicable to different types of applications such as extraction, joining, tabulating, and grouping. It can define multiple indexes for a single column so that appropriate indexes can be used depending on the type of query and high-speed processing can be performed.

These features enable data to be analyzed and used safely at high speeds by business intelligence (BI) tools and various types of applications.

5) Appliance Manager

Fujitsu ships the Analytics Ready product with hardware, the OS, and software already initialized, thereby greatly reducing the work required for design and implementation. In addition, the Appliance Manager, which integrates a full range of functions from implementation to operation, can also reduce the work associated with implementation and operation to a great extent. This enables the user to begin business operations with the Analytics Ready product about one hour after installation through a simple setup procedure in wizard format, including manager-ID and network settings. The Appliance Manager also enables tasks like hardware and software monitoring, setting of automatic backups, and batch execution of OS and software patches to be easily performed from an integrated GUI, thereby reducing the workload of the system manager.

Furthermore, while multiple data warehouses may exist to accommodate different departments, lines of business, regions, etc., Appliance Manager can reduce the total cost of operating those warehouses through integrated operation management while maintaining prior processing performance through the use of a high-speed query engine.

In conventional systems, operation and management work is performed separately for each data warehouse, so total operation cost is proportional to the number of warehouses. Analytics Ready, in contrast, can create up to four database instances for a single data warehouse. This means that an environment of multiple data warehouses can be constructed with only one Analytics Ready implementation by migrating existing data warehouses to those database instances. This mechanism reduces the total cost of operating data warehouses.

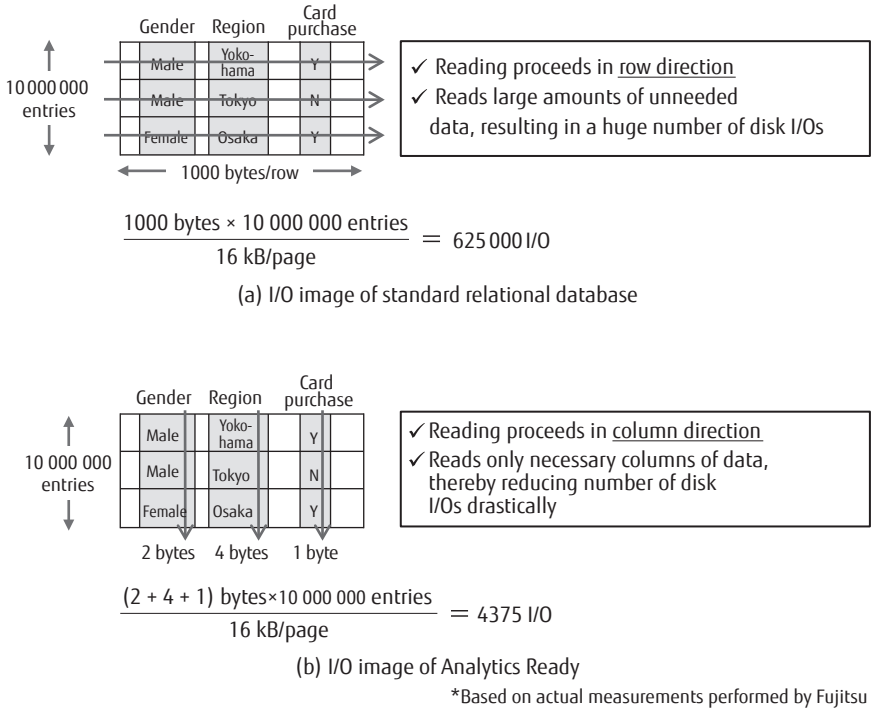

Figure 4
Conventional database table versus columnar table.

4. Changes in information usage

Converting the data stored within a company into reports to assist management in making sound decisions is one form of information usage that is already being applied. This process uses BI tools such as FUJITSU Software Interstage Navigator Server.[4] In conventional systems, a very basic access control function that allows data access to only those users involved in creating reports and that denies data access to unauthorized users is sufficient.

Recent years, however, have seen a growing need for speeding up the decision-making process at frontline business departments. Such departments are therefore receiving internal company data and being encouraged to analyze and use those data on their own. This transition to a greater number of data users has generated an even greater need for "safe sharing of consolidated data" by using technology that can finely control the scope of released data and minimize security risks such as information leaks.

5. Safe data sharing

The most important requirement among the various processes involved in using information is data protection. Analytics Ready addresses this need for safe data sharing.

The data warehouse consolidates various types of corporate data, and the referencing of much of this data is restricted in accordance with the affiliation and attributes of each user desiring access.

In the case of an ordinary relational database, access rights can be set for each table to restrict data referencing. However, if only some columns in a table need to be restricted, a new table has to be created that excludes those columns. This has the unfortunate effect of duplicating data since data in the unrestricted columns will appear in both the original table and the new table.

Analytics Ready solves this problem by using a data concealment function that enables the data returned by queries to be controlled by simply setting a data concealment policy on a table-by-table basis. Such a policy, as described below, covers target, type, condition, and format as concealment parameters (**Figure 5**).

1) Concealment target
   Specifies those columns to be concealed.
2) Concealment type

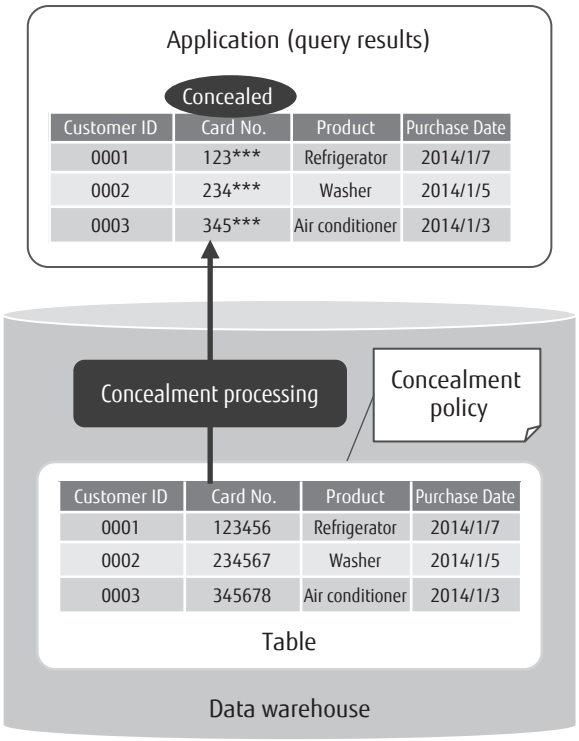

Figure 5
Data concealment function.

Specifies the range of data targeted for concealment with options being the entire column, part of a column, or a regular expression. For example, in the case of a character string like an e-mail address with a length that may vary, a regular expression could be used to specify that the local part of the address preceding the @ sign be replaced by asterisks (*).

3) Concealment condition

Specifies the condition for concealment using an expression that returns a Boolean result. For example, if data are to be concealed for only the user designated as "user1," the expression "current_user = user1" can be used to specify the concealment condition.

4) Concealment format

Specifies the concealment method and characters to be displayed when a concealment condition is true.

Analytics Ready executes data concealment in accordance with the settings described above, so there is no need to modify any applications. In addition, the data targeted for concealment are not altered in any way.

This data concealment function makes it possible to broaden the range of data analysis and use from diverse perspectives without having to prepare multiple tables corresponding to each range of data to be referenced by users. It also makes data sharing safe when users download data since the original data cannot be recreated from the information subjected to concealment processing.

6. Conclusion

This paper introduced the FUJITSU Integrated System Analytics Ready hardware appliance product, which is an extremely effective platform for analyzing and using massive amounts of various types of information in business applications.

Today, the move toward using big data in business is accelerating, and whether a company can promptly obtain useful knowledge from that data can affect its business success. The need for effective database technologies in the field of information systems is consequently greater than ever. In addition, the need is growing for information and communications technology (ICT) systems that can repeat plan-do-check-act (PDCA)-based decision-making in even shorter cycles by analyzing and using business information on-site in contrast to conventional information systems controlled by information systems departments.

Within this trend, Fujitsu considers safe data sharing to be a vital factor in expanding business, and, with this in mind, it intends to continue developing and enhancing Analytics Ready technology and enhancing the security of its information systems. In this way, Fujitsu aims to support a flowering of innovation in its customers' business enterprises.

References

1) Fujitsu: FUJITSU Integrated System Analytics Ready—Vertically Integrated Data Warehouse Infrastructure (in Japanese).
http://software.fujitsu.com/jp/symfoware/products/analyticsready/
2) H. Nagakura: AsIs Data Searching Using Database Engine "Shunsaku Data Manager." *FUJITSU*, Vol. 57, No. 2, pp.134–139 (2006) (in Japanese).
3) Fujitsu: FUJITSU Software Interstage Shunsaku Data Manager, an XML-based Database Engine (in Japanese).
http://interstage.fujitsu.com/jp/shunsaku/
4) Fujitsu: FUJITSU Software Interstage Navigator Server—Business Intelligence Server (in Japanese).
http://interstage.fujitsu.com/jp/navigator/

Kazushige Nakamura
Fujitsu Ltd.
Mr. Nakamura is engaged in the sales promotion of business products.

Takehiko Aoyagi
Fujitsu Ltd.
Mr. Aoyagi is engaged in the development of information-application products.

Shinya Echigo
Fujitsu Ltd.
Mr. Echigo is engaged in the development of information-application products.