Hitachi develops high performance data processing technology increasing data analytics speed by up to 100 times
A real-time analytics system prototype to realize self-service data analytics

Tokyo, August 3, 2016 – Hitachi, Ltd. (TSE: 6501) today announced the development of a database management system optimized for the high-speed embedded memory in the hardware (FPGA [*1]) and technology for high performance parallel data processing in FPGAs. Using these technologies, speed of data analytics was successfully increased by up to a maximum of 100 times compared with not using these technologies. Further, the two technologies developed were combined with "Pentaho Business Analytics," a business analytics software[*2] developed by Pentaho Corporation (a Hitachi Group company) to visualize business analytics results, and with flash storage for data storage, to create a prototype real-time data analytics system. The prototype will contribute to realizing self-service data analytics enabling employees in the field to easily and quickly execute data analytics on massive business data.

In recent years, self-service analytics that allow employees in the field to easily conduct big data analytics, usually conducted by experts such as data scientists, is gaining attention. One example might be that of a financial advisor, listening to a customer requirements, entering the information into the analytics system on the spot, and being able to suggest a financial product which matches the customer's needs. As can be imagined from this example, the data analytics system for self-service data analytics needs to produce results quickly, and thus must have high processing capabilities to execute data read and data analysis processes. By using flash storage[*3] instead of a hard disk drive to store data, the data read performance was increased by up to 10 to 100 times. Data analysis performance, however, has been unable to keep up with data read performance, thus creating a bottleneck in the analytics.

To overcome this issue, Hitachi developed a database management system optimized for the high-speed embedded memory in the hardware (FPGA) and technology to conduct high speed parallel data processing in the FPGAs, and succeeded in increasing data analytics speed by up to a maximum of 100 times. A real-time data analytics system prototype was then built by combining these two technologies with Pentaho Business Analytics for visualization of results, and flash storage for data store (Figure1).

The outline of the two technologies developed is as described below.

1. **Database management system optimized for high-speed memory in hardware (FPGA)**
   FPGA is equipped with small but high-speed internal memory (few MB), and connected to large but low-speed external memory (few GB). In the data format used in column-oriented or columnar databases,[*4] data management information which shows the location of data is larger than the internal memory and needed to be stored in the external memory. This management information, however, is required to determine the location of the data and frequently referred every time accessing the data. Thus, storing this information on large but low-speed external memory slows down the processing speed. In this research, a database management system was developed where the database was subdivided into multiple data segments so that the management information of each data segments could be handled by the FPGA internal memory, stored in the flash storage, and processed within the FPGA by each data segment. This database management system enables high-speed processing (Figure 2). (13 patent applications have been filed)

2. **Technology for high-speed parallel data processing by the hardware (FPGA)**
   Parallel data processing is widely adopted to conduct high-speed processing. In column-oriented or columnar database, however, this is difficult as the processing of one column must finish before the next column can be processed. To overcome this, a column processing method was developed to enable a set number of columns to be processed in turn. Parallel data processing was realized using this method together with a data filter circuit to select the data for analytics, and an aggregation circuit to group the data and calculate values such as total or average, to realize parallel data processing.

Hitachi plans to exhibit these technologies at the Flash Memory Summit 2016, to be held from 9-11 August 2016 in Santa Clara, California, USA.
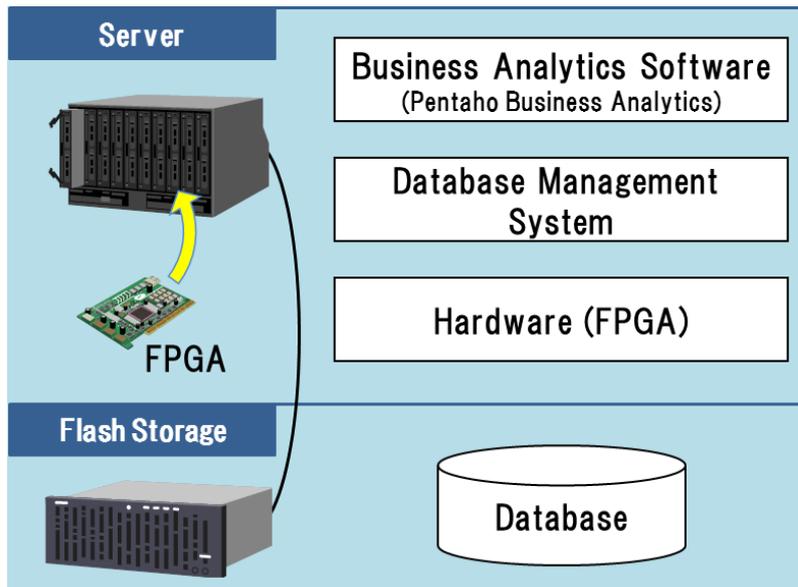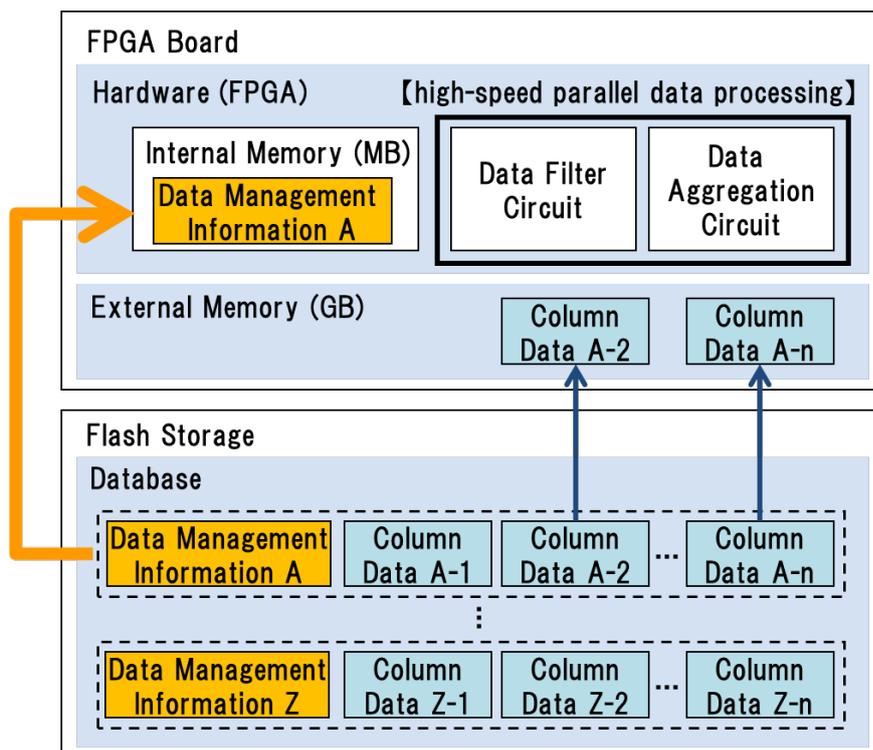
Fig. 1 Configuration of the data analytics system

Fig.2 Details of the technologies developed

Notes:

*1 FPGA (Field Programmable Gate Array): An integrated circuit manufactured to be programmable by the purchaser. In general, FPGA is inexpensive compared to application specific circuits.

*2 Business analytics software: Software to specify how to analyze the data and view the results.

*3 Flash storage: Storage equipment employing flash memory as the storage device.

*4 Column-oriented or columnar database: A database designed for efficient processing of data related to items, e.g. transaction data for a given item.

About Hitachi, Ltd.

Hitachi, Ltd. (TSE: 6501), headquartered in Tokyo, Japan, delivers innovations that answer society's challenges. The company's consolidated revenues for fiscal 2015 (ended March 31, 2016) totaled 10,034.3 billion yen ($88.8 billion). The Hitachi Group is a global leader in the Social Innovation Business, and it has approximately 335,000 employees worldwide. Through collaborative creation, Hitachi is providing solutions to customers in a broad range of sectors, including Power / Energy, Industry / Distribution / Water, Urban Development, and Finance / Government & Public / Healthcare. For more information on Hitachi, please visit the company's website at http://www.hitachi.com.

# # #