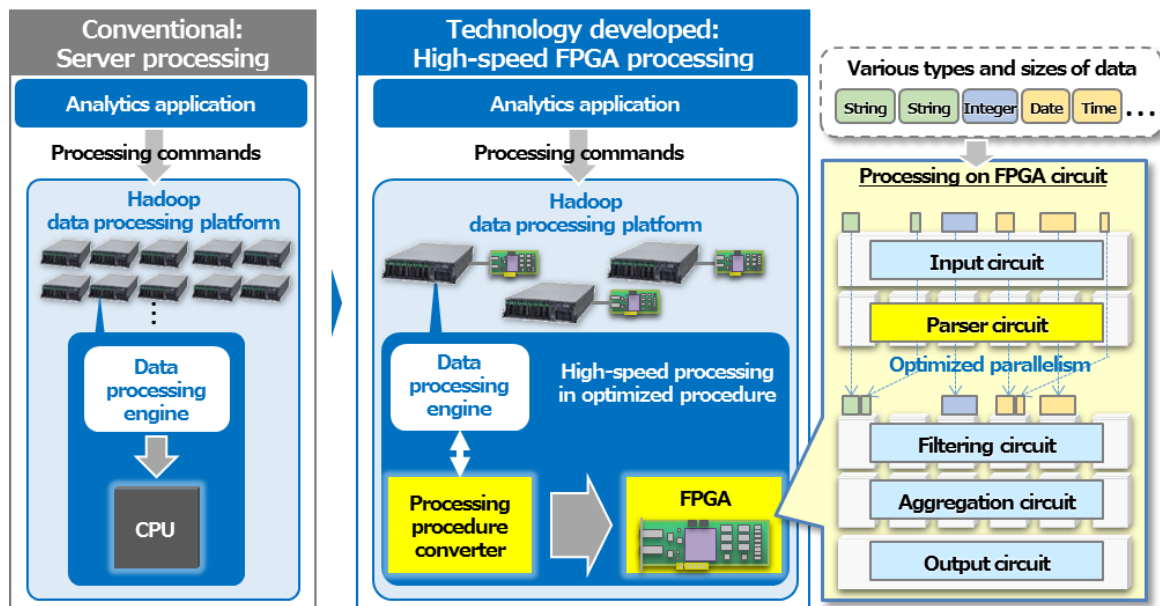


# News Release

FOR IMMEDIATE RELEASE

## Hitachi develops open source software based big data analytics technology to increase speed by up to 100 times

*For a high-speed analytics system with lower IT investment*



Overview of the technology developed

**Tokyo, November 14, 2017** --- Hitachi, Ltd. (TSE: 6501, Hitachi) today announced the development of the technology increasing the speed of big data analytics on an open source software Hadoop-based distributed data processing platform<sup>(1)</sup> (“Hadoop platform”) by a maximum of 100 times that of a conventional system. This technology converts data processing procedure generated for software processing in conventional Hadoop data processing, to that optimized for parallel processing on hardware, to enable high-speed processing of various types of data in FPGA<sup>(2)</sup>. As a result, less number of servers will be needed when conducting high-speed big data analytics, thus minimizing IT investment while enabling interactive analytics by data scientists, quick on-site business decision making, and other timely information services. This technology will be applied to areas such as finance and communication, and through verification tests, will be used to support a platform for data analytics service.

In recent years, big data analytics for interactively analyzing large amounts of various types of data from sources such as sensor information in IoT, financial account transaction records and social media, under various conditions and from various perspectives for business and services, is becoming increasingly important. The open source Hadoop platform is widely used for such analytics, however as many servers

are required to raise processing speed, issues existed in terms of equipment and management costs.

In 2016, Hitachi developed high performance data processing technology using FPGA<sup>(3)</sup>. As this technology however was developed for Hitachi's proprietary database, it could not easily be applied to the Hadoop platform as it employed a different data management method and used customized database management software.

To address this issue, Hitachi developed technology to realize high-speed data processing on the Hadoop platform utilizing FPGA<sup>(4)</sup>. Features of the technology developed are outlined below.

**(1) Data processing procedure conversion technology to optimize FPGA processing efficiency**

The Hadoop platform data processing engine optimizes data processing using the CPU to serially execute software to retrieve, filter and compute. Simply executing this procedure however does not fully exploit the potential of the hardware to achieve high-speed processing through parallel processing. To overcome this, the Hadoop processing procedures were analyzed, and taking into consideration distributed processing efficiency, technology was developed to convert the order of the processing commands to that optimized for parallel processing on FPGA. This will enable the FPGA circuit to be efficiently used without loss.

**(2) Logic circuit design to analyze various data formats and enable high-speed processing in FPGA**

Conventionally in FPGA processing, to facilitate processing on the hardware, the formats of different types of data, such as date, numerical value and character string, was restricted, and dedicated processing circuits were required for each type of data. The Hadoop platform however needs to deal with multiple data formats even for the same item, for example, even with dates there is the UNIX epoch day expression as well as the Julian day expression among others. Thus, as many dedicated processing circuits would be needed, the limited FPGA circuitry could not be effectively used with conventional FPGA processing. To resolve this issue, a logic circuit was designed to optimize parallel processing in FPGA, using parser circuits that clarify various data types and sizes<sup>(5)</sup> and depending on the data type and size, packs multiple data to be processed in one of the circuits. As a result, it is possible to not only handle various data formats but also realize parallel processing fully utilizing filtering and aggregation circuits for efficient high-speed data processing.

The technology developed was applied to the Hadoop platform. When analytics was performed on sample data, it was found that data processing performance improved by up to 100 times. The results suggest it will be possible to reduce the cost of Hadoop-based big data analytics as the number of servers required for high-speed processing can be significantly reduced. Hitachi will now conduct verification tests together with customers as it works towards the commercialization of this technology.

The technology developed will be on exhibit at SC17 - The International Conference for High Performance Computing, Networking, Storage and Analysis, to be held from 13th to 16th November 2017 in Denver, Colorado, USA.

- (1) Hadoop-based distributed data processing platform: A computation platform for storing and analyzing large amount of data on distributed servers using open source software, "Hadoop"
- (2) 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.
- (3) 3<sup>rd</sup> August 2016 News Release: "Hitachi develops high performance data processing technology increasing data analytics speed by up to 100 times"  
<http://www.hitachi.com/New/cnews/month/2016/08/160803a.html>
- (4) 10 related international patents pending
- (5) Supports the standard format "Parquet," generally used in open source data processing platforms such as Hadoop

### **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 2016 (ended March 31, 2017) totaled 9,162.2 billion yen (\$81.8 billion). The Hitachi Group is a global leader in the Social Innovation Business, and it has approximately 304,000 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>.

###

---

Information contained in this news release is current as of the date of the press announcement, but may be subject to change without prior notice.

---