# NEC IT Infrastructure Transforms Big Data into New Value

In step with rapid spread and adoption of the Internet and ICT in recent years, our world is amassing an ever-increasing quantity of data, and the vast majority lies unused in storage with its data unexploited. This article will examine the new challenges posed by Big Data, explain how NEC responds by enabling added value businesses, and give the reader an overview of the products and technologies behind NEC's solutions.

General Manager 3rd IT Software Division ITOU Akinori

#### **1** Preface

Accompanying the spread of the Internet and ICT, the volume of data accumulated by society is increasing at a dramatic rate. By 2020, the global data load is forecast to reach about 200 times the level in 2006. It is also interesting to note that according to a survey by the Ministry of Public Management, Home Affairs, Posts and Telecommunications, over 90% of the corporate-stored data are left unused and wasted like "buried treasure."

It has been previously pointed out that new business opportunities can be identified by mining the mountains of stored corporate data and using the insights to formulate strategies. In the future, it will be necessary to respond to increasing speed, volume and diversity of data. Putting the huge stores of unused data to work in a way that will reveal new knowledge and added value, which in turn leads to new business, will be one of the biggest challenges in the future of corporate strategy.

Also due widespread adoption of mobile terminals such as smart phones and mobile phones and increasing use of ubiquitous sensors, the drastic rise in data traffic has become a serious issue. Results of a survey conducted by the Ministry of Public Management, Home Affairs, Posts and Telecommunications reported that the volume of mobile data traffic in Japan approximately doubled in just one year from 2010 to 2011. Responding to this rapid rise in traffic is already an issue for not only the telecommunications carriers. This is also a challenge facing all information and service providers that use mobile terminals and sensors.

This explosion of data is called Big Data and is the focus of increasing attention.

This article will explain some of the new issues presented by Big Data as well as the business opportunities arising from them and provide the reader with an overview of how NEC is approaching the development of solutions.

## 2 The Advent of the Age of Big Data and the Issues Facing Corporations

Big Data is already present in every imaginable domain of business. In a study conducted by McKinsey & Company, the three sectors that handle and store the largest amount of data are Discrete Manufacturing, Government and Communications & Media as of 2009<sup>1)</sup>. Even in the case of the Construction sector (the sector with the smallest amount of aggregate data in the study), the volume was estimated to be as much as 51 petabytes (PB). In every sector and field of business, Big Data and how each enterprise or organization responds to its challenge are keys to differentiation from the competition and success.

Unlike conventional Business Intelligence (BI) in which the data are analyzed by a certain expert from a certain industry



Fig. 1 Example of a high-speed financial information service.



Fig. 2 Example of a sensor data-based traffic information service.

perspective, Big Data analytics requires analysis from multiple and diverse perspectives - from the viewpoints of users from various sectors and occupations. Anticipating the arrival of the Age of Big Data, NEC has developed a variety of platform technologies and products that facilitate the analysis of diverse types of data. The next chapter in this article will introduce the shape of new businesses created through the exploitation of Big Data.

# **3** Business Opportunities Created by Big Data Exploitation

The conventional utilization of data requires the design of a system based on a variety of predetermined conditions such as the purpose of the data usage, collection methodology, format, processing flow and so on. However, in the case of exploiting Big Data, it is necessary to mine data for knowledge and added value - a purpose for which the data were not originally

envisioned to be used, and to set up a flexible processing flow that can handle unstructured and irregular data and perform real-time or batch processing according to the form of utilization. Therefore, compared with conventional data usage, the degree of a Big Data system environment's freedom to handle data must be overwhelmingly expanded.

Through the exploitation of Big Data by such a flexible system environment, the following three categories of added value can be gained.

#### (1) Accelerated provision of information

Businesses that can process tremendous quantities of data in real time become possible. For example, it is feasible to develop a business that collects the vast quantities of constantly fluctuating information from trading systems such as stock and forex exchanges, processes them at high-speed in real time, and distributes the information as a service to private individual and institutional investors in milliseconds (**Fig. 1**).

## (2) Identification of new correlationships

Businesses that can identify new correlationships and rules from the unexploited cumulative data become possible. For example, through the collection and analysis of various information such as sensor data from vehicle terminals, road traffic congestion data and road construction/accident information, it is conceivable to develop a service that provides highly detailed traffic information for general drivers, dealers and manufacturers capable of providing diverse value from route optimization to part replacement notification and even accident diagnostics (**Fig. 2**).

#### (3) Forecasting

Businesses that can forecast human and event behavior and proactively take appropriate preemptive action become possible. For example, a service can be developed to collect and analyze massive amounts of data (minute vibrations, temperature, etc.) from various types of sensors installed in a factory and automatically build a "factory behavior model" for normal operating conditions. Then by comparing this model with actual conditions in real time, the service can monitor, detect deviations from the norm, and predict possible factory abnormalities (**Fig. 3**).

NEC has introduced Big Data-related products and technologies one after another, and draws on extensive experience and a record of achievements in diverse fields from large-volume event processing and image data recognition to sensor data acquisition (**Fig. 4**).

# 4 NEC Big Data-related Products and Technologies

Sensor data and all other kinds of information are being con-

verted into electronic data; the volume of data continues to grow exponentially; and the necessity to collect, process and analyze diverse types of data is growing dramatically. The technological issues facing the usage of Big Data can be broadly divided into the following three categories:

- (1) Acquisition of the large volume of data generated by sensors and devices,
- (2) Data management and processing that can respond to the rapidly increasing volume of data, and
- (3) Data analysis of the broad diversity of data types including image and voice data.

Concretely responding to these issues, the platform architecture that will support Big Data business will be structured in 3 layers: (1) Data Collection Platform, (2) Data Management & Processing Platform, and (3) Data Analysis Platform (**Fig. 5**). Already very conscious of the approaching advent of the Age of







Fig. 3 Example of a service that monitors factory/building status and predicts potential abnormalities/failures.



Fig. 5 NEC's Big Data product and technology map.

Big Data, NEC had begun tackling the development of related technologies one after another since 2006, and is already systemizing various Big Data technologies by these platform layers. Flexibility combining these product groups to optimally respond to the unique needs of each customer, NEC can provide Big Data solutions. This chapter will introduce the reader to the NEC product lineup for each platform layer.

#### (1) Data collection platform

The Data Collection Platform layer facilitates the collection of data by offering multiple receiving channels for input from various types of sensors and devices, and standardizing the sensor data access method.

NEC products that support this platform layer include CONNEXIVE (M2M platform that provides standardized handling of sensors with different access methods), and agricultural sensors, radiation sensors, ultrasensitive vibration sensors, "smart" electric power distribution boards and other sensors that provide realtime acquisition of "real world" information.

Enabling connectivity with all types of sensors and devices, CONNEXIVE is a data collection and storage platform for various types of information. By providing an allpurpose application interface that can assimilate the differences in the interfaces of diverse connected devices, CONNEXIVE enables the building of applications without being conscious of inter-device differences.

Also among the various types in the sensor lineup, the ultra-sensitive vibration sensor which adopts an originally developed oscillation amplification mechanism is especially remarkable. It can measure a broader frequency range of vibration with high sensitivity that is 20 times that of conventional vibration sensors, and it achieves this performance at 1/10th the cost.

### (2) Data management & processing platform

Depending on the intended purpose of the processing, NEC's Data Management & Processing Platform layer provides two different high-speed, large-quantity data processing methods: Real-time Processing of Large-scale Streamed Data and High-speed Batch Processing of Stored Data. Through this approach, it is possible to swiftly respond to real world fluctuations with real-time processing and uncover patterns in stored data with distributed processing. Also for the Data Management & Processing Platform layer, NEC can provide a server that can efficiently process large quantities of data by scaling out, storage and flow control as hardware for large-volume processing.

Products that make possible real-time processing of largescale data streams include InfoFrame Table Access Method (TAM) and Blockmon (high-speed processing platform for data streams).

TAM can perform continuous processing of large-volume data using a memory table access method, achieve about 10 times the processing performance of conventional mission-critical systems (thousands of events per sec. during a peak period) and is adopted in systems that real-time processing of large quantities of transactions. Blockmon, a software platform that is under research and development by NEC Laboratories Europe, adopts a highly flexible modular architecture. This flexibility and performance of this platform makes it ideal for various stream data analytics including a wide range of traffic monitoring, use by telecom carriers to detect unauthorized users, and corporate intranet traffic statistical analysis.

Products for high-speed batch processing of stored data Include InfoFrame Relational Store (IRS), InfoFrame DataBooster, and InfoFrame DWH Appliance. Combining the merits of conventional relational database (RDB) technology and the "scaling out" advantage of key value store (KVS) technology, IRS is the new database software for the Age of Big Data and has the following 3 features:

- "Think Big, Start Small and Scale Fast"-ready architecture (abbreviated "Small Start" in Japan) results in a 50% reduction in initial deployment costs compared with conventional systems;
- Compatibility with SQL (the international standard language for relational database development) enables effective utilization of existing legacy of database design; and
- 3) Adoption of originally developed "MicroShard" technol-

ogy delivers the high reliability demanded by missioncritical tasks.

Also InfoFrame DataBooster significantly shrinks batch processing time by high-speed in-memory processing of stored data, enabling a reduction of execution cycles from months to days.

InfoFrame DWH Appliance is a data warehouse (DWH)/ appliance product that optimally combines NEC's highperformance and high-reliability server and storage systems with the advanced Netezza data retrieval/analysis software, and delivers about 10 to 100 times the performance of conventional large-scale DWH systems.

Express 5800 Scalable HA Server supports a large-scale memory to realize high-speed in-memory processing. SIGMABLADE facilitates the addition of nodes by adoption of a "scaling out" system. And the Storage M Series of SAN storage arrays optimize arrangement of autonomous data according to access frequency to accelerate data processing. In addition, the UNIVERGE PF Series of network products adopts next-generation OpenFlow network technology for flow control with optimized routing depending on data content.

# (3) Data analysis platform

The Data Analysis Platform layer provides highly sophisticated analytics technology and know-how for unstructured data such as image and voice data. Examples of analytics technology are biometrics that use fingerprint or face matching, and fault detection that uses machine learning.

NEC's fingerprint identification engine boasts a FAR (false acceptance rate) of less than 0.00001% - the top performing system in the world, and has been adopted for HANIS, the Republic of South Africa's citizen ID system - one of largest scale ID databases in the world with 45 million people.

Featuring Generalized Learning Vector Quantization using face confidence maps based on the generation of multiresolution images and Non-frontal Recognition technology, NEC's NeoFace face detection/recognition software achieves 10 times the accuracy of competitive systems and leads the world in performance.

In addition, NEC's fault detection technology that uses machine learning automatically generates an invariance model of the total system based on vast quantities of monitored data, and then by performing a comparison of real-time monitored data against this model, the technology can detect system performance anomalies. This technology is incorporated in MasterScope Invariant Analyzer - a key operations management tool.

### Conclusion

This completes our explanation of the issues and business opportunities presented by the advent of the Age of Big Data, and an overview of how NEC has been responding with the development of a total architecture consisting of the Data Collection Platform, Data Management/Processing Platform, and Data Analysis Platform.

This special issue of the NEC Technical Journal will give the reader a more detailed explanation of the various products that comprise NEC's Big Data solutions and the cutting-edge technologies from our laboratories that support them. While exploiting our established technological edge in processing large volume and unstructured data and providing our customers with cloud services and state-of-the-art ICT products, NEC shall continue to endeavor to strengthen competitive advantage of our customers with solutions and services that anticipate and exceed their needs and expectations.

\*Hadoop is a registered trademark or trademark of The Apache Software Foundation. \*Netezza is a registered trademark or trademark of International Business Machines Corporation.

#### Reference

 McKinsey&Company:"Big data: The next frontier for innovation, competition, and productivity," 2011.5 http://www.mckinsey.com/insights/mgi/research/technology\_and\_innovation/big\_data\_the\_next\_frontier\_for\_innovation

# Information about the NEC Technical Journal

Thank you for reading the paper.

If you are interested in the NEC Technical Journal, you can also read other papers on our website.

# Link to NEC Technical Journal website



# Vol.7 No.2 Big Data

Remarks for Special Issue on Big Data NEC IT Infrastructure Transforms Big Data into New Value

# $\Diamond$ Papers for Special Issue

# Big data processing platforms

Ultrahigh-Speed Data Analysis Platform "InfoFrame DWH Appliance" UNIVERGE PF Series: Controlling Communication Flow with SDN Technology InfoFrame Table Access Method for Real-Time Processing of Big Data InfoFrame DataBooster for High-speed Processing of Big Data "InfoFrame Relational Store," a New Scale-Out Database for Big Data Express5800/Scalable HA Server Achieving High Reliability and Scalability OSS Hadoop Use in Big Data Processing

#### Big data processing infrastructure

Large-Capacity, High-Reliability Grid Storage: iStorage HS Series (HYDRAstor)

#### Data analysis platforms

"Information Assessment System" Supporting the Organization and Utilization of Data Stored on File Servers Extremely-Large-Scale Biometric Authentication System - Its Practical Implementation MasterScope: Features and Experimental Applications of System Invariant Analysis Technology

#### Information collection platforms

M2M and Big Data to Realize the Smart City Development of Ultrahigh-Sensitivity Vibration Sensor Technology for Minute Vibration Detection, Its Applications

#### Advanced technologies to support big data processing

Key-Value Store "MD-HBase" Enables Multi-Dimensional Range Queries Example-based Super Resolution to Achieve Fine Magnification of Low-Resolution Images Text Analysis Technology for Big Data Utilization The Most Advanced Data Mining of the Big Data Era Scalable Processing of Geo-tagged Data in the Cloud Blockmon: Flexible and High-Performance Big Data Stream Analytics Platform and its Use Cases

# $\diamondsuit$ General Papers

"A Community Development Support System" Using Digital Terrestrial TV



Vol.7 No.2 September, 2012

