Cloud Fusion Concept

● Yoshitaka Sakashita ● Kuniharu Takayama ● Akihiko Matsuo
● Hidetoshi Kurihara

Cloud is expected to develop from a single-purpose cloud to a hybrid cloud that links clouds or existing systems, or to a fusion of two or more clouds in the future. Fujitsu Laboratories named this advanced form of coordination "Cloud Fusion" at the start of 2010. This paper explains the aim of this coordination and the direction in which research should head. It goes on to describe the relationship between Cloud Fusion and Fujitsu's and Fujitsu Laboratories' vision—enabling Human-Centric Intelligent Society. It describes the five pillars of research on Cloud Fusion and its outline. In particular, it introduces details about the development and execution environment that is one of the pillars.

1. Introduction

Fujitsu Laboratories started working on cloud computing (hereafter "cloud") together with Fujitsu in the spring of 2007 and started up the present Cloud Computing Research Center in April 2009. This organization is an aggregate of groups engaged in cloud-related research such as infrastructure control, system management and development and execution environment. It has been launched as a center with the nature of a cross-laboratory organization. Fujitsu and Fujitsu Laboratories had been researching themes such as grid computing and virtual system control since before the launch of the center. As an initial achievement, we developed a virtual system control engine, which provides the core of a cloud system.[1)] The engine was adopted as an in-house practice system of the On-demand Virtual System Service. It is capable of providing an infrastructure for the development of necessary systems such as user management and billing as well as operation and service design. Moreover, it contributed to the subsequent start of commercial service in October 2010.[note)] We have conducted research while closely working with the product development division and FUJITSU Family Association/LS Research Committee, a customers' research committee, from the beginning. In the process of studying what to realize with clouds and communicating with customers, we became aware that technology capable of linking existing systems and clouds was necessary. Hence, we started to work on achieving a hybrid cloud around the summer of 2009. With this idea still promoted as Fujitsu's cloud strategy, we have proposed the optimum system construction by linking public and private clouds and existing systems and offered products. Subsequently, we have proposed various types of clouds to give rise to healthcare, local government and other clouds. Under the circumstances, Fujitsu Laboratories felt it was necessary to develop a system linked with various clouds and existing systems, and named this form of coordination "Cloud Fusion"

note) In June 2011, the On-demand Virtual System Service was named as Fujitsu Global Cloud Platform "FGCP/S5" service in the Japanese market and Global Cloud Platform (GCP) service in overseas markets.

at the start of 2010.

This paper describes the aims and overall concept of this Cloud Fusion.

2. Aims of Cloud Fusion

In the cloud market, the public cloud business and software as a service (SaaS) business based on it were the mainstream in the initial phase, and we now expect customer-specific private cloud services will develop. IDC Japan released a forecast on April 7, 2011 saying that the scale of the software market for public clouds and that for private clouds will be almost equal at 145 billion yen in 2012, and the latter will overtake the former after that. Meanwhile, a form of sharing a cloud by an industry or community has appeared. At present, we are conducting research on media and engineering clouds as a service in addition to healthcare and local government clouds. In addition, we are currently linking clouds by using sensors and mobile phone handsets as input devices, which may be referred to as the Internet of Things.

Fujitsu and Fujitsu Laboratories set forth a vision of Human-Centric Intelligent Society in the spring of 2010 and have since been working for the construction of a cloud environment that supports this vision. Based on the trends described above, we estimated that a system for linking and mutually using private and public clouds would be important in terms of business as well. We aimed to link various clouds together by Cloud Fusion to link and share information, and thereby expand the fields of information and communications technology (ICT) application and create a new market. A similar concept is referred to as "inter-cloud" by the Smart Cloud Study Group of the Ministry of Internal Affairs and Communications (MIC).[2] While similar concepts have been recently announced by other companies as well, we believe that Fujitsu Laboratories has taken the lead in this research. Furthermore, Fujitsu Laboratories proposes a way to eliminate companies' anxiety of being locked in to a specific cloud by using Cloud Fusion to link different types of clouds. As technologies required for realizing these aims and ideas, we are intensively engaged in R&D under the following five themes.

- Secure data connection
- Big data parallel distributed processing
- Cloud development and execution environment
- Cloud mangement
- Managed network

Figure 1 shows the background, concepts

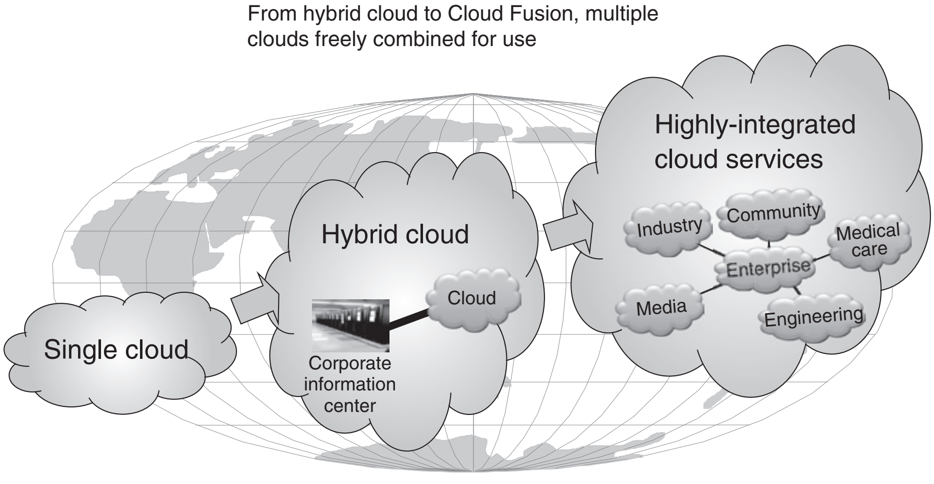

Figure 1
Fujitsu Laboratories' approach to cloud.

and major research themes. The following sections describe the R&D for realizing Cloud Fusion with the focus on the development and execution environment. For other themes, see the respective papers contained in this special issue.

3. Technologies for realizing Cloud Fusion

An overview of the technologies for realizing Cloud Fusion is shown in **Figure 2**.

Cloud is a synthesis of technologies widely ranging from infrastructures such as hardware and networks to technologies that use them to make contributions as business and social infrastructure. Accordingly, Fujitsu Laboratories does not only conduct world-class R&D in the respective fields of these technologies but also links other companies' technologies and open source materials. It does so by vertical integration in an attempt to sophisticate the entire system. For that purpose, aside from the research group for Cloud Fusion, we have other groups such as one engaged in research on Green Internet Data Center (IDC), for example, which puts together servers, storage, networks and facilities. Above Cloud Fusion, there is also a group working on the technologies for enabling Human-Centric Intelligent Society, which is Fujitsu's and Fujitsu Laboratories' vision. In this field of research, Cloud Fusion takes charge of the development and execution environment for high-speed accumulation and analysis of various pieces of event data and other big data. Human-Centric Intelligent Society supports data collection in the actual field and analysis, prediction and proposal for use in the field.

The Cloud Fusion group has defined five pillars of R&D as mentioned in the previous section. The following outlines the respective pillars of research.

1) Secure data connection

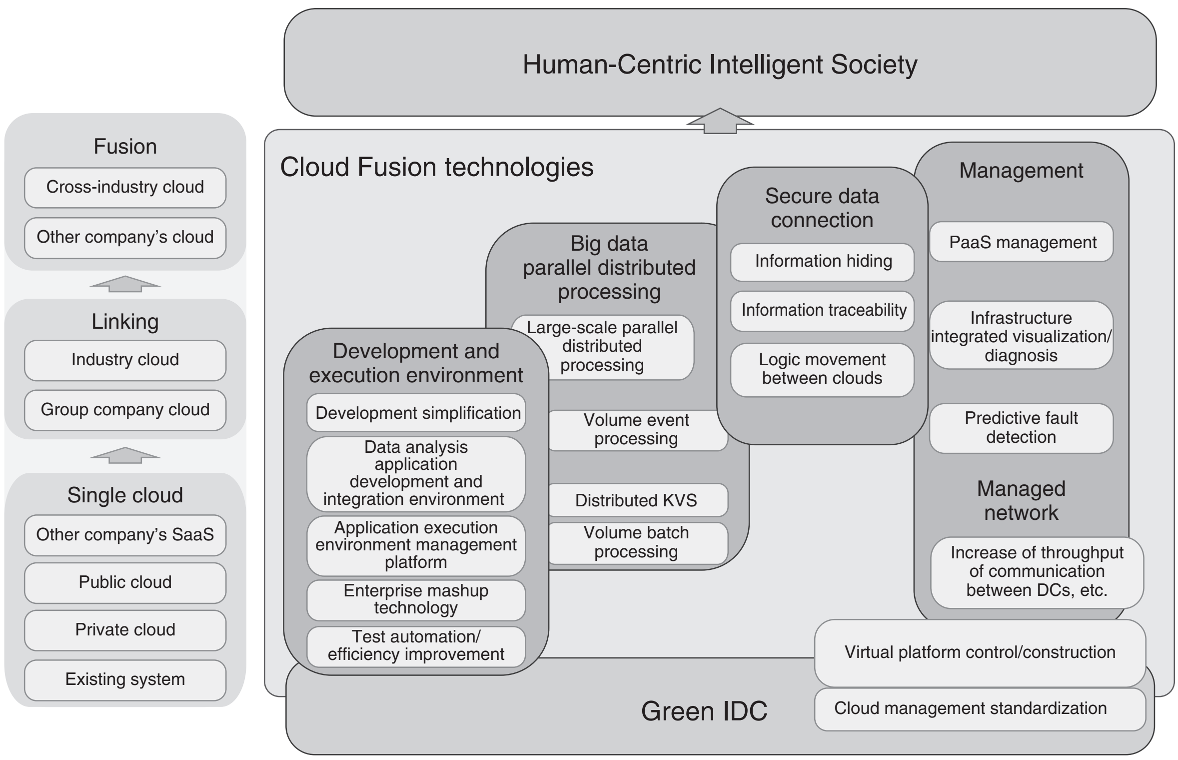

Figure 2
Fujitsu Laboratories' overall technology for Cloud Fusion.

Security, which is said to be a major hindering factor to the application of cloud, may have a significant impact if information is leaked and a system to allow use with more security than before is required. We want customers to entrust Fujitsu's cloud with their information without anxiety. Hence, we aim to be able to hide information between the existing systems and clouds and to identify how information has been used (this function will be commercialized in the middle of FY 2012). In addition, we need to be able to process data without handing them to a cloud. Therefore, we have researched and developed a technology to have SaaS applications downloaded to the own center and impose a restriction to prevent the applications from sending information outside.

2) Big data parallel distributed processing

The essential aim of Cloud Fusion is to link and share information to expand the fields of ICT application and create a new market. To do this, big data processing is a key technology. We are basing our work on technologies such as the key value store (KVS) technology, suitable for storing big data, and open source Hadoop, capable of quickly processing big data. And we are investigating even more advanced parallel distributed processing for close linkage with Human-Centric Intelligent Society.

3) Cloud development and execution environment

This involves research on technology to allow customers to easily connect clouds with existing systems and big data analysis platforms. The following section gives the details.

4) Cloud management

Up to now, we have conducted research on the efficient control and management of virtualization systems. We have also proposed the management interface for On-demand Virtual System Service. It became commercially available in October 2010, and we proposed it to the Distributed Management Task Force (DMTF), an industry group promoting the standardization of IT management, aiming for its standardization. In addition, we have commercialized part of the technology to comprehensively visualize and diagnose the performance of network and server infrastructure at the end of 2010. Since the beginning of 2011, we have been aiming to apply predictive fault detection technology to the real environment. To this end, we have been moving ahead with new research on PaaS management in consideration of application life cycles.

5) Managed network

In this age of cloud, we believe networks will experience a dramatic increase in data volume, and this has given rise to demand for research to address various challenges. For example, research on the performance and management technology of a network between clouds and between clouds and terminals is necessary and a new perspective is required.

4. Overview of development and execution environment

The essence of cloud is aggregation and distribution. That is, cloud is intended to aggregate on a large scale customers' systems and data to realize economies of scale and their distributed processing by commodity infrastructure (such as servers, storage and networks) to realize cost efficiency.

In terms of execution environment, aggregating systems on a cloud provides cost and quality benefits. For example, it eliminates the need for infrastructure on the customer's site, makes systems available for use only when and as much as required, and acts as measures for disaster recovery by using a backup cloud. Meanwhile, aggregating data on a cloud also provides cost, quality and efficiency benefits in terms of development environment. For instance, it centralizes the development environment and makes it easy to ensure governance on the development processes and results. With this situation in the background, software development and execution environment on a cloud, or the so-called platform as a service (PaaS), has recently been evolving and becoming widespread.

It must be noted that implementing existing business applications on a cloud has a scale-out problem. With the conventional infrastructure, business applications with large amounts of accesses and processing are often built by using high-performance servers and databases, which is called scale-up. On the other hand, clouds are composed of large numbers of similar commodity servers and databases and cannot be scaled up. Accordingly, scale-out is necessary, in which processing is distributed among many servers and databases. Relational databases (RDBs), which are typical conventional databases, are good at complicated data retrieval and aggregation. However, they take their processing steps one at a time so as to maintain consistency, which makes them unsuitable for distributing multiple accesses for parallel processing. In clouds, on the other hand, new types of data storage are used such as KVS, which is capable of high-speed parallel distributed processing of a simple data structure with sets of keys and values. For this reason, we have developed optimum deployment technology for data storage. It automatically selects between multiple types of data storage such as RDB and distributed KVS in a cloud system according to the characteristics of the data or accesses handled. In this way, it appropriately switches data storage accesses from Web applications.[3),4)]

This paper has so far covered the existing areas of business. New business areas that take advantage of big data such as sensor data and business logs are being launched. To analyze large amounts of data and develop event-driven services, we need a system capable of easily and comprehensively analyzing and developing batch analysis processing of big data and real-time processing of event data. We have developed a big data analysis PaaS for that purpose.

In the future, we expect that linking and fusion of various application systems will be important including the existing and new areas. With general Web applications, linking is roughly classified into linking in the front end in charge of interaction with the user and that in the back end, which takes charge of processing and interchange with databases according to the access from the front end. One well-known technology of the former is mashup. For the latter, service-oriented architecture (SOA) is a popular technology. One important challenge in such linking is how to effectively use the existing systems. Accordingly, we have developed WebAPI creation assistance technology, which semi-automatically extracts application programming interfaces (APIs) from the existing systems to allow them to be mashed up with other applications.

Of these, big data analysis PaaS and WebAPI creation assistance technology will be explained in more detail in the following sections.

5. Big data analysis PaaS

From now on, it will be important to make use of data gathered in large volumes from sensors and terminals to propose new services and solutions that could not be readily realized with the conventional ICT. In the analysis and development environment based on Cloud Fusion, we will need to create new services together with customers by combining data across the boundaries of business categories and industries starting from the big data owned by customers.

However, analyzing big data and developing event-driven services requires skills to widely master and implement various processing technologies including data parallel distributed processing (such as Hadoop), event processing (such as Esper) and analysis processing (such as Mahout). An analysis and development environment that comprehensively supports them has not been realized.

To address this need, we have test-built the environment shown in **Figure 3** that supports the cycle of service planning by data analysis [2) in Figure 3], simplified development of service applications [3) in Figure 3] and service evaluation [4) in Figure 3]. In this environment, batch analysis processing of big data and real-time processing of event data can be handled in an integrated manner. As an interface, a data flow diagram (DFD) editor environment (HTML5) is provided [1) in Figure 3]. The architecture allows the defined DFD to be configured as a program according to Service Component Architecture (SCA), a standard component specification, for deployment and execution.

The data analysis personnel can easily execute analysis and event processing without needing knowledge about the individual implementation technologies, simply by deploying input/output (I/O) data and processing components as a DFD and providing execution parameters. To realize this system, we have developed as a differentiating technology a new function to convert components into the optimum implementation component according to the types of processing and data.

We have taken a real case of business as a reference to develop cluster analysis processing and coupon issuing application and evaluated this environment. We have confirmed that with this environment it is possible to reduce the period of time taken to provide the service, which was conventionally two months, to two weeks.

6. WebAPI and mashup technology

More than one business system is operated in an enterprise and a user often needs to use multiple systems to perform a series of operations. Along with the expansion of cloud, companies now need to link with applications provided by external clouds in addition to those systems. For this reason, we have developed WebAPI creation assistance technology to call existing systems as Web services without making changes to allow linking by mashup.

Mashup is a technique of combining multiple Web services published on the Internet to build Web applications with new value, as exemplified by Google Maps. Application developers can easily use functions such as data

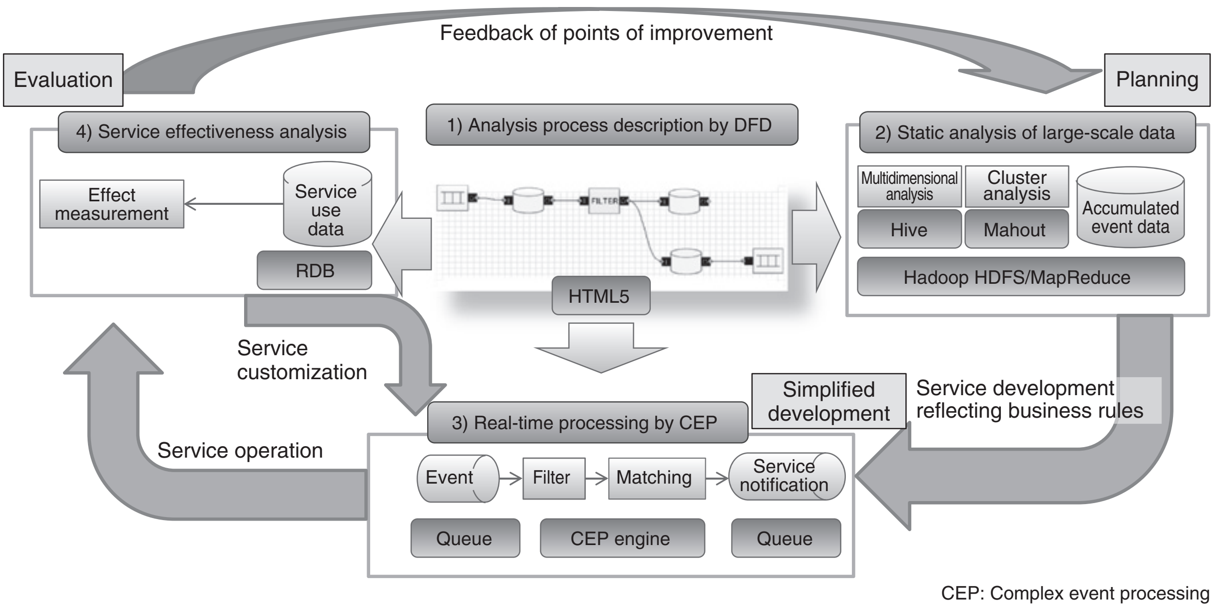

Figure 3
Outline of PaaS for large-scale data analysis.

acquisition and screen display/operation simply by executing calls according to the APIs offered by Web services. It may be possible to apply this mashup technique to enterprise systems to use the functions of existing systems and applications that meet the needs of user departments. If such an environment can be built easily then higher operations efficiency can be expected.[5)]

To use system functions by mashup, however, the functions must be provided as APIs and callable from a different system. Existing systems in enterprises are often not provided with such APIs and are unavailable from outside as components. Making changes to systems to add APIs or data-linking function requires many person-hours for development and testing. This has made it difficult up to now to use the functions of existing systems by mashup.

To deal with this challenge, we have developed WebAPI creation assistance technology. It adds APIs without making changes to existing Web applications to allow them to be used from mashup applications. With this technology, the user's screen operations on the system are reproduced to obtain the results of processing from the screen, thereby adding WebAPIs without making changes to existing systems. **Figure 4** shows an overall view of this technique.

To create an HTTP request queue as a reproduction of screen operations, this technology has the user execute the operations of the applications concerned and records data sent and received between the browser and Web applications. This eliminates the need to develop a program for generating a request queue and operations that need many screens before producing results can be easily turned into WebAPIs. In addition, it has a function to edit the recorded request queue for extraction of parameters to be provided from outside such as the user ID, password and search key. Hence, authentication information essential to business systems can be provided by mashup applications.

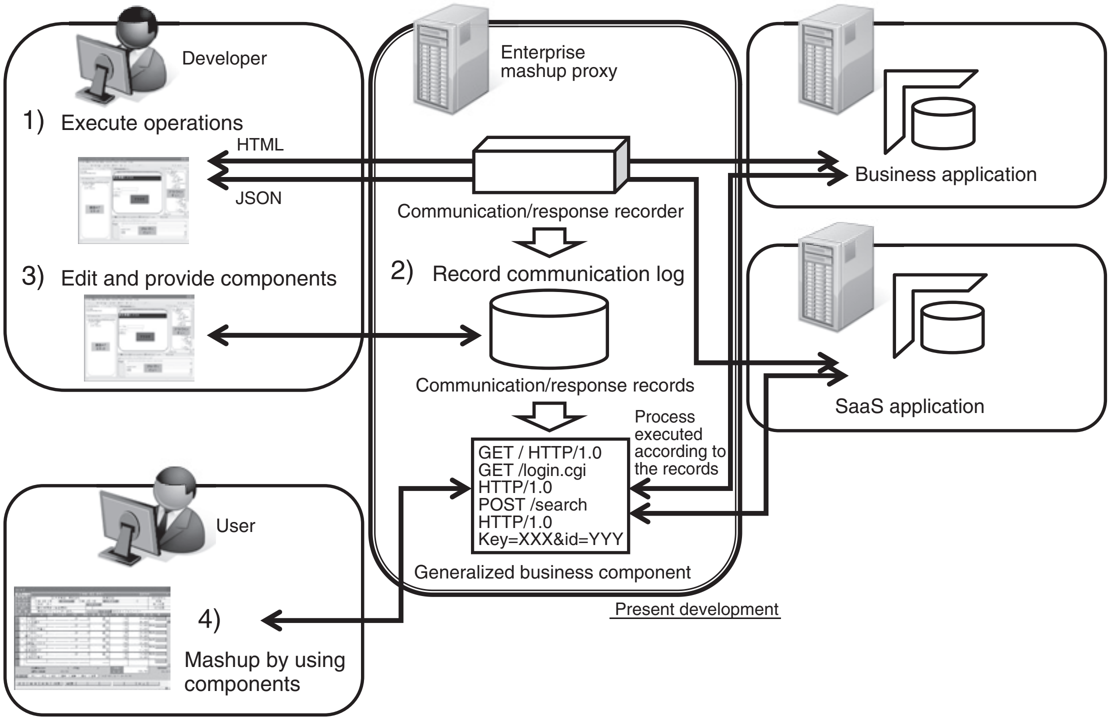

Figure 4
Overall support technology for making WebAPI.

In a trial using this technology, we have confirmed that WebAPI can be created with one-tenth the person-hours as compared with adding WebAPI to an existing system by manually creating a request queue.

7. Conclusion

Fujitsu Laboratories' approach to the area of cloud presented here is based on the vision called Cloud Fusion. At the beginning of the development of On-demand Virtual System Service in October 2010, it was referred to as the Service-Oriented Platform. We thought of a "site" where information is gathered to conduct new communication and business as a "platform." We have conducted research based on this idea since before the name "cloud" became popular and we think that the essence of Cloud Fusion is the same. In the future, cloud is expected to change the business, social infrastructure and lifestyles of people in the world more than we can imagine. Fujitsu Laboratories intends to continue to lead the development of cloud technology so as to help create and pioneer new business.

References

1) M. Kishimoto: Fujitsu's Trusted-Service Platform. (in Japanese), Cloud Computing Technology to Grasp the Other Side of the World of Clouds. ASCII, 2009, pp. 58–65.
2) MIC: Smart Cloud Study Group Report—Smart Cloud Strategy—(May 2010). *http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/councilreport/pdf/100517_1.pdf*
3) M. Adachi et al.: Fujitsu's Approach to the Age of Cloud—Development of Construction and Operation Technology for a Cloud System. (in Japanese), ASCII technologies, June Issue, pp. 62–69 (2010).
4) Y. Mizobuchi et al.: Automatic Optimized Use of Multiple Datastores in Cloud Computing Era. (in Japanese), *IPSJ Special Interest Group on Software Engineering Report*, Vol. 2010-SE-170, No. 14, pp. 1–8 (2010).
5) A. Matsuo et al.: LivePoplet: Technology That Enables Mashup of Existing Applications. *Fujitsu Sci. Tech. J.*, Vol. 45, No. 3, pp. 304–312 (2009).

Yoshitaka Sakashita
Fujitsu Laboratories Ltd.

Mr. Sakashita is currently engaged in research strategy and promotion for entire cloud computing.

Akihiko Matsuo
Fujitsu Laboratories Ltd.

Mr. Matsuo is currently engaged in research and development of inter-cloud linking technology and technology to improve efficiency of software maintenance.

Kuniharu Takayama
Fujitsu Laboratories Ltd.

Mr. Takayama is currently engaged in research and development of application development and execution envi-ronment in cloud.

Hidetoshi Kurihara
Fujitsu Laboratories Ltd.

Mr. Kurihara is currently engaged in research and development of software development environment and up-stream process technology.