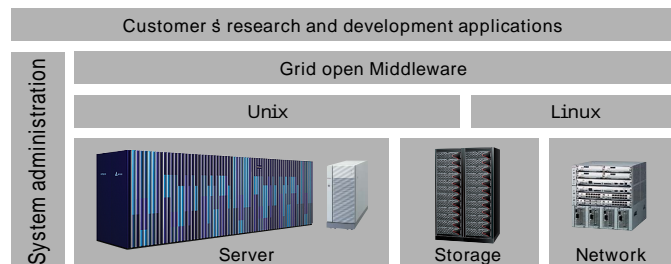


## Science Grid Computing

Grid computing is a technology which virtualizes the resources of multiple computer systems connected to a network. Grid computing is beginning to be developed and introduced in universities and various research laboratories for the purpose of sharing computer system resources. Hitachi is providing a science grid with solutions by concentrating on hardware technologies, software technologies and system operation technologies that have been accumulated in scientific computation fields. A hierarchical figure of a scientific grid is shown below. Grid open middleware virtualizes and integrates multiple servers in which the SR11000 model J1 plays a major role.



## Storage Equipment

The SR11000 model J1 supports large-capacity, high-performance, high-reliability storage equipment. The Hitachi Thunder 9500™ V Series is a disk array-based storage system which can play a major role in a file system. It uses 2Gbps Fibre Channel for both host interfaces and disk interfaces, and thus has a high throughput of up to 7.4 GB/sec. In addition, it has a large maximum capacity of 60 TB\* (449 disks, RAID-5 configuration) and up to 8GB of cache memory. Redundant RAID configuration (RAID-5), dual controllers, and hot-swappable components result in high reliability. Thunder 9500™ V Series systems are available in storage rackmounts consisting of controllers and a single rack of 14 disks. They can be upgraded with additional racks containing maximum 15 disks each. Users can choose the appropriate configuration to meet application requirements.

\*Thunder 9585V case



## SR11000 model J1 Hardware Specifications

	Number of nodes*	4	8	16	32	64	128	256	512	
System	Theoretical peak performance	486.4GFLOPS	972.8GFLOPS	1,945.6GFLOPS	3,891.2GFLOPS	7,782.4GFLOPS	15,564.8GFLOPS	31,129.6GFLOPS	62,259.2GFLOPS	
	Inter-node transfer speed	4 / 8 / 12GB / s (in each direction) × 2								
	Maximum total memory capacity	512GB	1,024GB	2,048GB	4,096GB	8,192GB	16,384GB	32,768GB	65,536GB	
	External interface	Ultra320 SCSI, Ethernet / Fast Ethernet, Gigabit Ethernet, Fibre Channel								
Node	Theoretical peak performance	121.6GFLOPS								
	Memory capacity	32GB / 64GB / 128GB								
	Maximum I / O performance	8GB / s								

\*HITACHI SR11000 model J1 can be configured with an arbitrary number of nodes between 4 and 512. A system with over 128 nodes will be treated individually.

- AIX, AIX 5L, IBM, LoadLeveler and POWER5 are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.
- Ethernet is a registered trademark of Xerox Corp.
- Linux is a trademark or registered trademark of Linus Torvalds in the United States and other countries.
- MPI is an interface convention set by MPI Forum in the U.S.
- UNIX is a registered trademark of The Open Group in the United States, other countries or both.
- Thunder 9500 is a trademark of Hitachi Data Systems Corporation.
- Other product and company names mentioned in this document may be the trademarks of their respective owners.

# HITACHI

For more information on the SR11000, please visit Hitachi's Web site at [http://www.hitachi.co.jp/Prod/comp/hpc/SR\\_e/11ktop\\_e.html](http://www.hitachi.co.jp/Prod/comp/hpc/SR_e/11ktop_e.html) or please contact the nearest Hitachi office.

### Hitachi Europe Ltd.

High Performance Computer Group  
Hitachi Europe GmbH  
Dornacher Strasse 3,  
85622 Feldkirchen  
bei Muenchen,  
Germany  
Tel : +49 (0)89 99 180-113  
Fax : +49 (0)89 99 180-356

### Hitachi America, Ltd.

50 Prospect Avenue  
Tarrytown, NY 10591  
U.S.A.  
Tel : +1 (914) 333-2991  
Fax : +1 (914) 333-2786

### Hitachi, Ltd.

Solution Business (IXL)  
Global Business Planning & Operations Division  
Information & Telecommunication Systems  
Hitachi Omori 2nd Bldg, 27-18,  
Minami Oi 6-chome, Shinagawa-ku Tokyo,  
140-8572 Japan  
Tel : +81 3 3763 2411  
Fax : +81 3 5471 2571

Design & specifications are subject to change without notice.  
To ensure safety and normal operation, be sure to read all instruction manuals before using the equipment.

CC-E317P 1004  
Printed in Japan(H)

## SUPER TECHNICAL SERVER HITACHI SR11000 Model J1

**HITACHI**  
Inspire the Next

SUPER TECHNICAL SERVER  
**SR11000**

This catalog has been created using Acrobat PDF format from Adobe Systems Incorporated.  
All Rights Reserved. Copyright © 2003, Hitachi, Ltd.

The quest for scientific knowledge whether illuminating the depths of outer space or mapping the genomes of life requires enhanced capabilities. The engines of remarkable breakthroughs in many fields, scientific and technical computers are now being transformed by exponential increases in performance. This progress has opened up new fields of application generating even more demand. The critical need is high-performance computers offering higher processing speeds, greater scalability and operational availability.

Hitachi's focus is unwavering. A relentless strategy of raising performance to meet this demand. Our ultra-powerful parallel computer, Super Technical Server SR11000 offers massive amounts of computational power and outstanding system balance by combining TFLOPS-class\*1 processing with a high-performance inter-node network.

The Hitachi SR11000. The machine designed to empower man's insatiable desire for scientific enlightenment. \*1: 1TFLOPS is the ability to execute one trillion floating-point arithmetic operations per second.

# Empowering the Pursuit of Scientific Knowledge

The SR11000 model J1 is a parallel computer in which each node\*2 is an SMP,\*3 and excels especially in the following points.

\*2 Nodes are independent processing units that are combined to form a parallel computer. \*3 SMP-Symmetric Multi Processor

#### Node architecture which has high performance and high scalability

Automatic parallelizing compiler obtains maximum performance.

#### High scalability architecture and excellent system balance

Inter-node network which has high performance and high scalability obtains optimal system performance.

#### Compact packaging

World class theoretical peak performance per unit floor area. A large high-speed computer with TFLOPS-level processing capability can be installed in a small area.

#### Wide variety of user applications

It can be applied to scientific and engineering applications that use various programming models. Physical random number generators which have world class generating speed can be used.

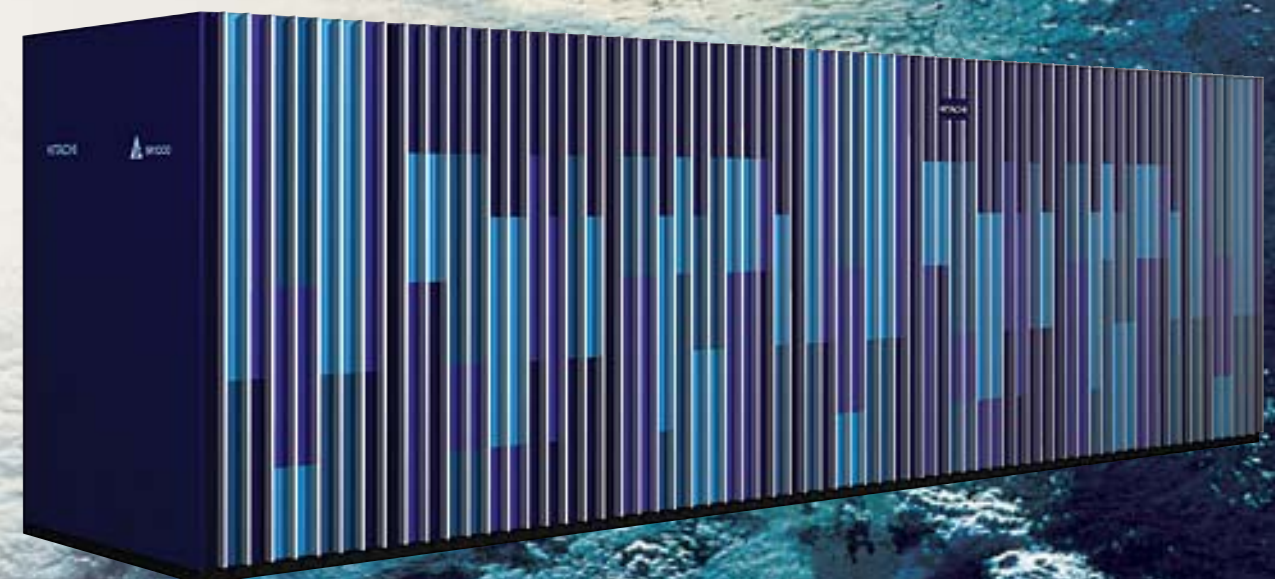
#### Advanced technologies for centralized administration

Budget and result management, automatic system operation and node maintenance to keep the system operational.

#### SR11000 Applications

Fluid analysis   Crash analysis   Weather forecasting   Atomic energy   Petroleum exploration  
New materials development   Aeronautics and Astronautics   Financial Engineering   Molecular design  
Life sciences   Structural analysis   And many others

HITACHI SR11000 model J1 32-node model



SR11000 Cabinet Design Concept - Scalable Slat

The seemingly endless array of slats and the vibration of the colors and lights reflected in the slats symbolize the high system scalability and extreme node performance.

# Achieves Node performance of 121.6GFLOPS. State-of-the-art technologies enable ultra-high speed calculation.

## High-performance CPU

The SR11000 model J1 adopts the 1.9GHz POWER5 chip, which has been developed using state-of-the-art technologies including copper wiring and SOI.\*1 The major characteristics of the POWER5 are super scalar processing, large capacity on-chip L2 cache, high memory throughput and high-speed inter-processor connection interfaces. The POWER5 chip is a key factor in the high performance of the SR11000 model J1.

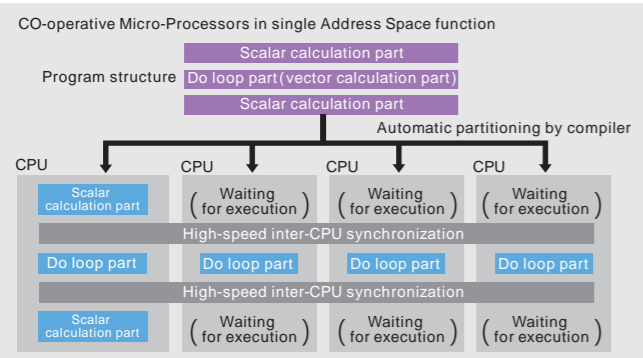
\*1 SOI: Silicon on Insulator

## Large capacity L3 cache

Each node has a total of 288MB of L3 cache. With this large capacity L3 cache, memory latency is shortened enhancing effective performance.

## CO-operative Micro-Processors in single Address Space (COMPAS) function

The SR11000 model J1 adopts the CO-operative Micro-Processors in single Address Space (COMPAS) function which simultaneously and rapidly synchronizes the multiple CPUs in a node. COMPAS enables high-speed start/end of DO loop calculations by hardware. COMPAS enables element-wise parallel processing like that found in vector processors and enables effective parallel processing.



## Memory pre-fetch function and software pipeline

“Hardware memory pre-fetch” is a function in which hardware checks automatically for memory data access patterns, and pre-fetches the memory data into the cache. This function is particularly effective to speed up programs that cannot be analyzed by the compiler. Automatic parallelizing compilers use hardware memory pre-fetch and software memory pre-fetch to allow arithmetic units to operate in a pipelined manner by means of software pipeline techniques and thus realize high-speed processing.

## Multi-stage Crossbar Network

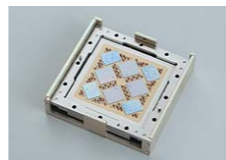
The network that interconnects the large number of nodes is another key technology in parallel processing. The SR11000 model J1 adopts a multi-stage crossbar network that has the advantage of fewer data transfer collisions. Since the maximum transfer rate of the inter-node network is 12GB/s (in each direction)x2, the ratio of the inter-node transfer rate to the node arithmetic performance is 1:10 which represents an excellent system balance.

## High-speed memory system

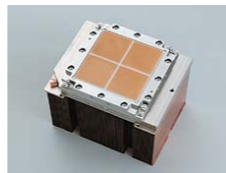
Low memory latency and high memory throughput are realized by connecting CPUs and high-bandwidth memory by switches with sufficient bandwidth not to act as a bottleneck on performance.

## World class advanced packaging technologies.

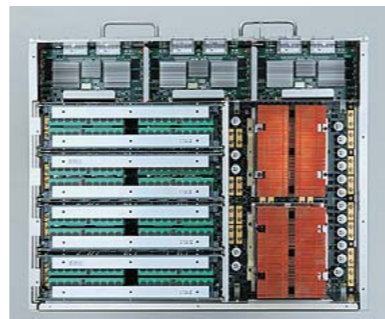
Thanks to advanced packaging technologies used, the theoretical peak performance per unit floor space of the SR11000 model J1 is seven times that of its predecessor. The SR11000 model J1 achieves a world class theoretical peak performance per unit area of 640GFLOPS/m<sup>2</sup>, enabling a large-scale high-speed computer with TFLOPS-level processing capability to be installed in a small area. Eight processors and 144MB of L3 cache are mounted on a 120mm wide by 120mm long high density module. Two high density modules, memory, memory controllers and inter-node network controllers are mounted on a high-density printed circuit board (package). A node with a theoretical peak performance of 121.6GFLOPS and a maximum of 128GB of memory is accommodated in this 460mm wide by 570mm long package.



High density module (with the cap removed)



High density module (view from pin side)

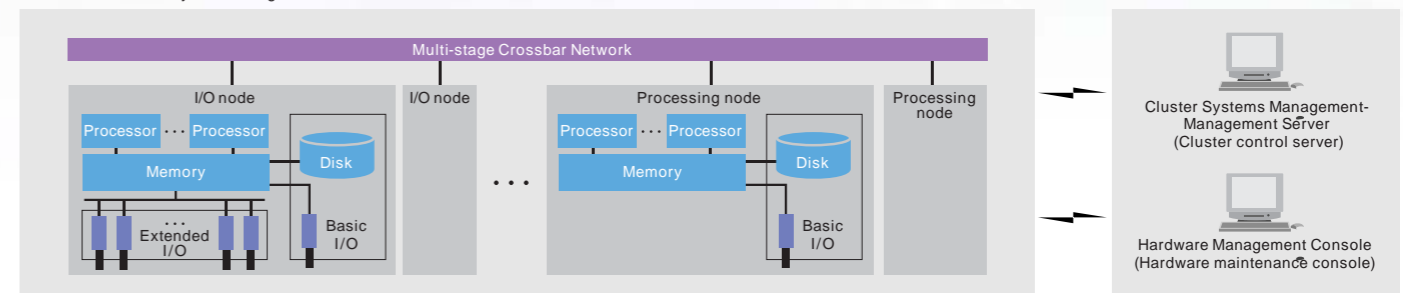


A node package.

## Flexible system configuration

The system configuration of the SR11000 model J1 is shown below. Each node is a 16-way POWER5 SMP and has a peak performance of 121.6 GFLOPS.\*2 Each node is equipped with one PCI slot and two or four integrated disks. Two types of nodes, an I/O node that connects to extended I/O interfaces and a processing node that does not connect to any extended

SR11000 model J1 system configuration



I/O interfaces exist. The high-performance network connects the nodes together. The peak performance of the SR11000 model J1 scales from 486.4GFLOPS(4nodes) to 62.2TFLOPS (512 nodes) and thus allows the user to choose the appropriate size of machine that meets the application's requirements.

\*2 1GFLOPS is the ability to execute one billion floating-point arithmetic operations per second.

## High reliability functions

As shown below, the SR11000 model J1 is equipped with high reliability functions in each component.

### Memory

- Memory patrolling  
In the background, hardware automatically searches for one-bit errors and automatically writes back corrected data.
- Memory bit steering  
When the number of one-bit errors on a memory line reaches a set threshold, the memory line is automatically replaced by a spare chip.
- Memory multi-bit error correction  
A multi-bit error occurring in a single chip can be corrected.
- On cache ECC  
L2/L3 caches are protected by ECC allowing correction of one-bit errors.

### Inter-node network

Network data is protected by CRC (Cyclic Redundancy Code). In addition, a resend function resends any failed network data.

## High availability functions

As shown below, the SR11000 model J1 is equipped with high availability functions in each component.

### Node

Node maintenance function to keep the system operational enables node replacement while the machine is operating.

### I/O

PCI adapter hotswap function allows replacement of PCI adapter in Extended I/O while the machine is operating.

### Automatic failure notification to remote maintenance center

When a hardware failure occurs or when the number of recoverable errors reaches a set threshold, occurrence of the event is automatically reported to the remote maintenance center.

HITACHI

# A rich variety of software to extract the best performance from the SR11000 model J1.

## Basic control

With 64-bit addressing, large scale calculations with 100TB files and more than 100GB of memory can be processed at high speed. In order to execute jobs efficiently, various job management functions are provided. The dynamic load distribution function dynamically assigns jobs to under-loaded systems and executes them. A function that designates job execution time and method can be used as a flexible system administration function, for example ,batch job execution at night. These functions ensure effective utilization of hardware resources. Job execution control functions such as job start/cancel/condition display, checkpoint/restart and accounting logs are also provided. Operating status monitoring functions, graphical display functions, tools for diagnosing processing performance problems, tools for analyzing system resources and operating status and so forth are provided. These functions optimize processing performance of the entire system. Centralized operation and collective administration for control of the nodes forming the SR11000 model J1, installing software and so forth are provided. In addition, the system can be partitioned into multiple groups of nodes and can be used in units of a group. Files can be distributed over multiple disks connected to each node in a cluster system, and parallel file I/O operations can be realized by file striping. In addition, file sharing by multiple nodes has been enabled. Budget and result management per user, group or account number has been realized. And thus system use per user can be managed. In addition, automatic system operation, in which the system starting time or ending time can be specified, has been realized. An extended memory function which enables users to use memory in each node as an extended memory region is provided. With this function, high speed data transfer between jobs is possible by making a virtual disk image file system in memory.

## Development support

Message passing libraries such as MPI which are used for inter-task communication of parallel programs, parallel debugger, and analysis tools for parallel applications are provided as effective development and execution environments for C, C++ and Fortran parallel programs. Parallel programs can be effectively developed.

## Compiler (Language)

Provided languages include a C compiler conforming to industry standards such as ANSI standards and OpenMP specifications, a C++ compiler conforming to ANSI standards, a Fortran compiler conforming to Fortran 77/90/95 standards and program development environments such as debuggers.

### Optimizing FORTRAN90

Provided as development/execution environments for making applications are an automatic parallelizing function that uses high level analysis of DO loops, and a program tuning assistance function that assists users with optimized coding at compile time, for example a diagnostics message output function. In addition, this compiler is optimized so that the highest degree of hardware performance can be extracted from the SR11000 model J1.

## Numerical calculation library

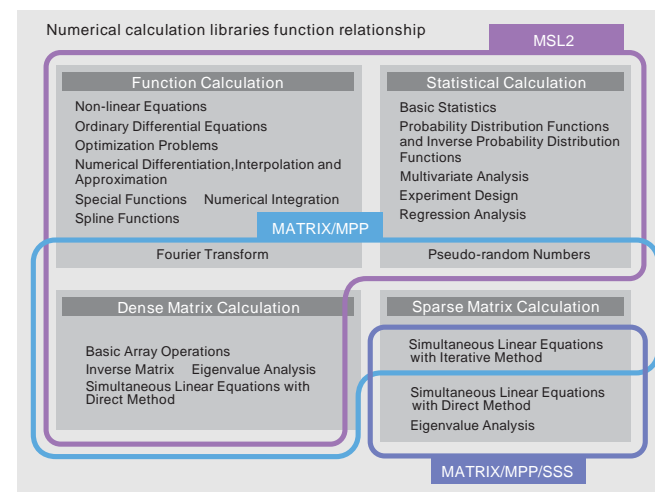
General purpose numerical calculation libraries are provided that correspond to scientific and engineering applications that have achieved high performance in various fields. See the numerical calculation libraries function relationship shown below.

### MATRIX/MPP

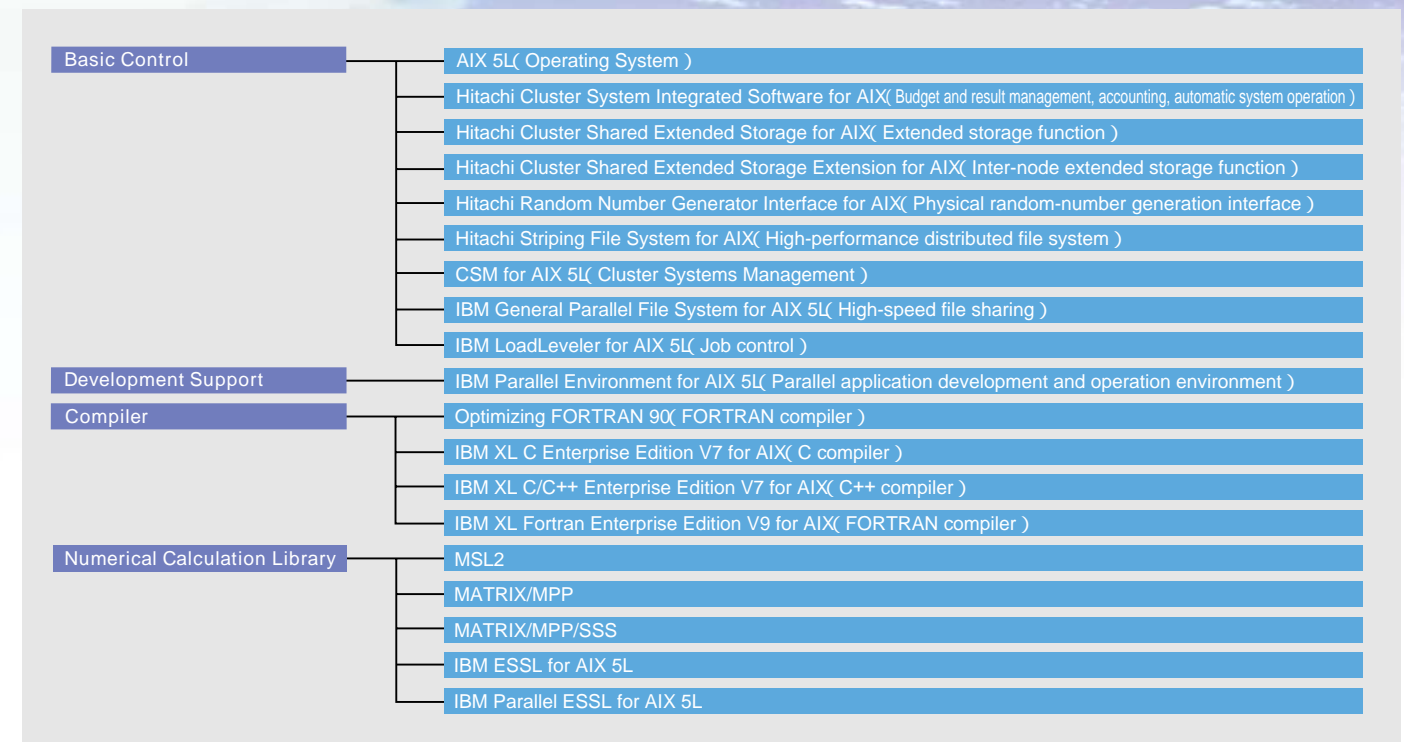
MATRIX/MPP is a numerical calculation subprogram library which includes matrix calculations such as iterative methods for simultaneous linear equations with large scale sparse coefficient matrix and fast Fourier transforms for either real or complex number data. This library is optimized to achieve high performance of the SR11000 model J1.

### MATRIX/MPP/SSS

To solve simultaneous linear equations with large scale sparse matrix used in numerical simulation of structural analysis, MATRIX/MPP/SSS provides the skyline method and sparse solver. This library is optimized to achieve high performance of the SR11000 model J1.



## Software configuration



## Broad Service Menu

Combining Hitachi HPC technology and Hitachi's long experience, Hitachi offers a broad service menu including system introduction, system configuration and program tuning. The service menu is shown in the following table.

Service name	Contents
Scientific and engineering system prototype evaluation service	Evaluating the performance of a customer's applications in advance of system introduction.
Scientific and engineering system introduction service (introduction plan and design)	Investigating the schedule and configuration for system introduction and introduction designing of program products.
Scientific and engineering system introduction service (hardware)	Confirming loading, installation, and connection for hardware introduction.
Scientific and engineering system introduction service (environment setting)	Executing various environment settings such as PP installation and system-defined information setting and confirming system operation.
Scientific and engineering system specialized consulting service	Consulting about configuration and operation of applications for customers who use high performance computers, scientific technical servers and/or grid computing system.
Scientific and engineering system education service	Explaining usage techniques, basic programming methods and numerical calculation techniques for high performance computers and/or scientific technical servers.
Scientific and engineering programs parallelization/vectorization tuning service	Porting programs running on a current system to a high performance computers, scientific technical servers and/or grid computing system according to the customers' needs and speeding up the execution of programs by parallelizing and vectorizing.
Scientific and engineering system stable operation service (Basic service,Optional service)	Providing problem solution assistance in the case of failure in a customer's system, system modification/maintenance, operation management assistance, system administration assistance and so forth.