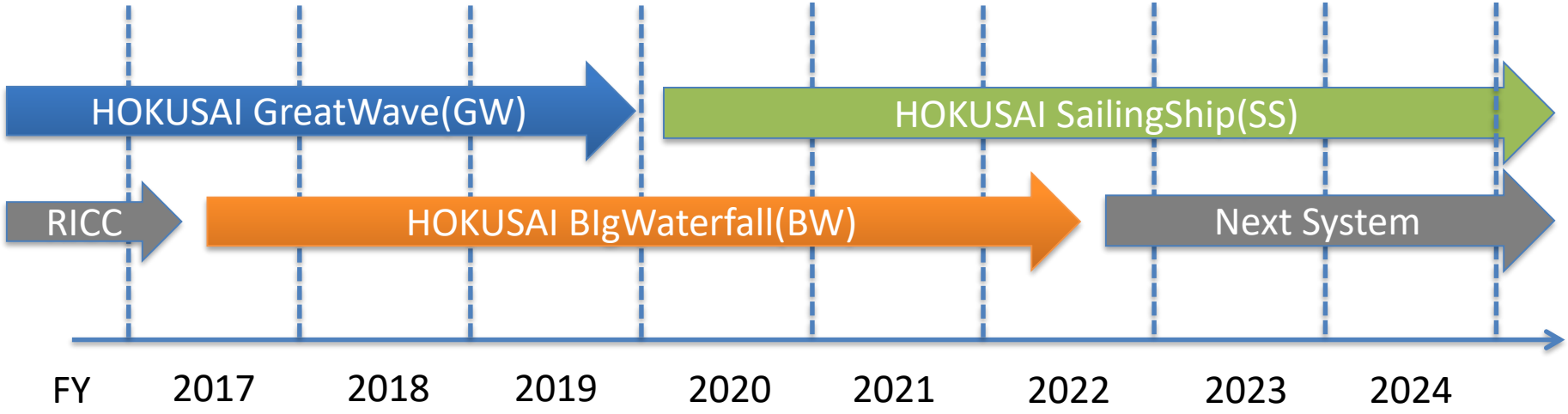
Handout 1

Trial Operation of HOKUSAI SailingShip

Information and Communication Infrastructure Section,
Information Systems Division,
Head Office for Information Systems and Cybersecurity

Operation schedule of shared use computers

- HOKUSAI GreatWave (GW) system started operation in April and will end in March 2020.
  - 1080 nodes, CPU: SPARC64-XIfx, 2PB
- HOKUSAI BigWaterfall (BW) system started operation in October 2017 and will end in September 2022年.
  - 840 nodes, CPU: Xeon Gold 6148, 5PB
- Data Science Infrastructure HOKUSAI SailingShip(SS) system will start operation in June 2020. Regular operation will start in October.
  - 440 nodes, CPU: Xeon Platinum 8260, 30PB



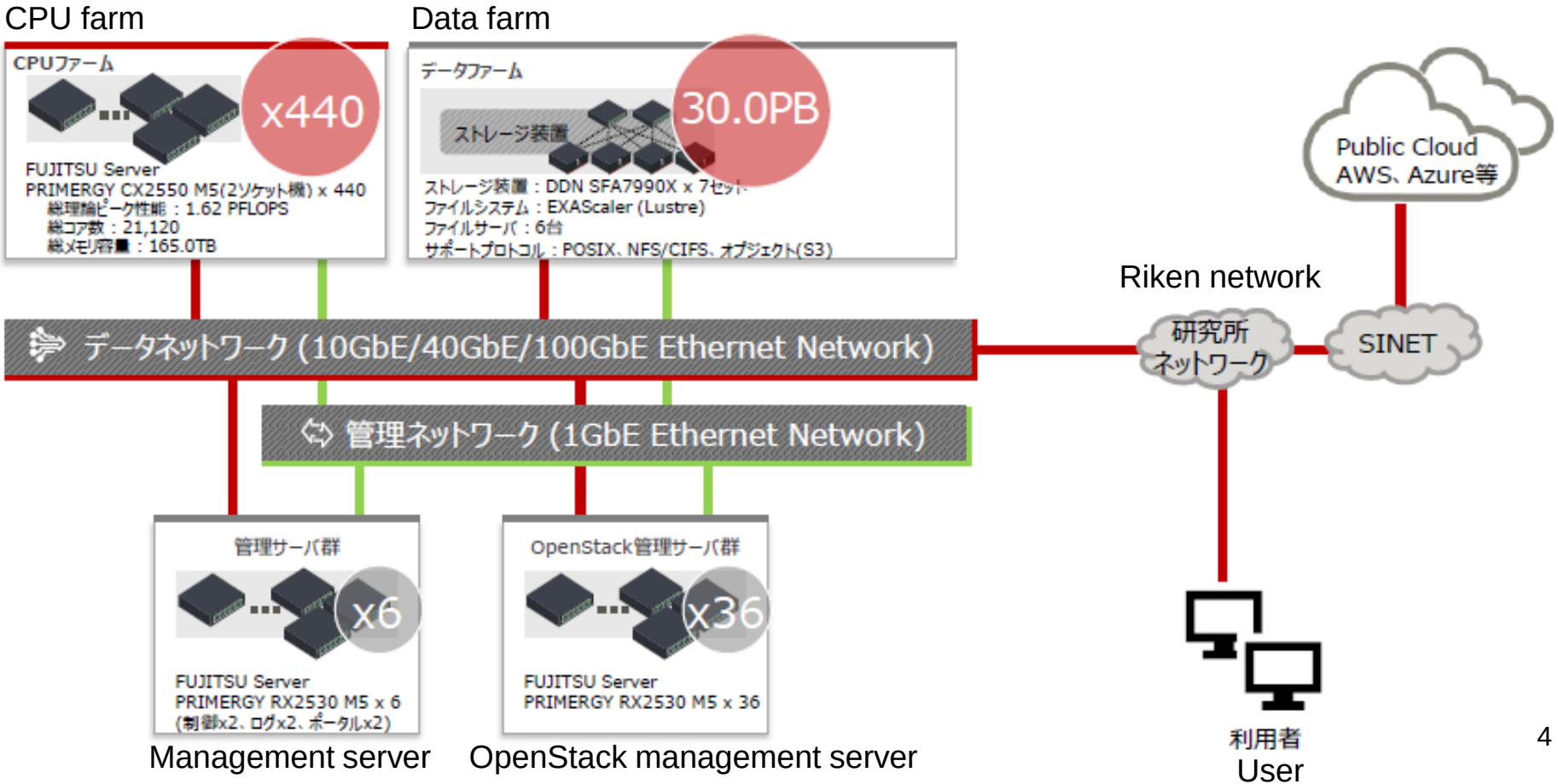
Operation schedule of Data Science Infrastructure HOKUSAI SailingShip(SS)

- Procurement process
  - November 2018 Procurement started (RFI public notice)
  - November 27, 2019 Bid opening→Fujitsu's proposal
- Operation schedule
  - June 1, 2020 1:00 p.m. Trial operation starts
  - November 2020 Start of regular operation
- User meetings
  - October 3, 2019 Introduction of beneficiary pays
  - March 24, 2020 Outline of SS and usage fees
  - June 1, 2020 Trial operation guide
  - August-September 2020 Regular operation guide

Even though the schedule has been delayed due to the new coronavirus, the trial operation started just in time. However, please understand that we are not fully prepared.

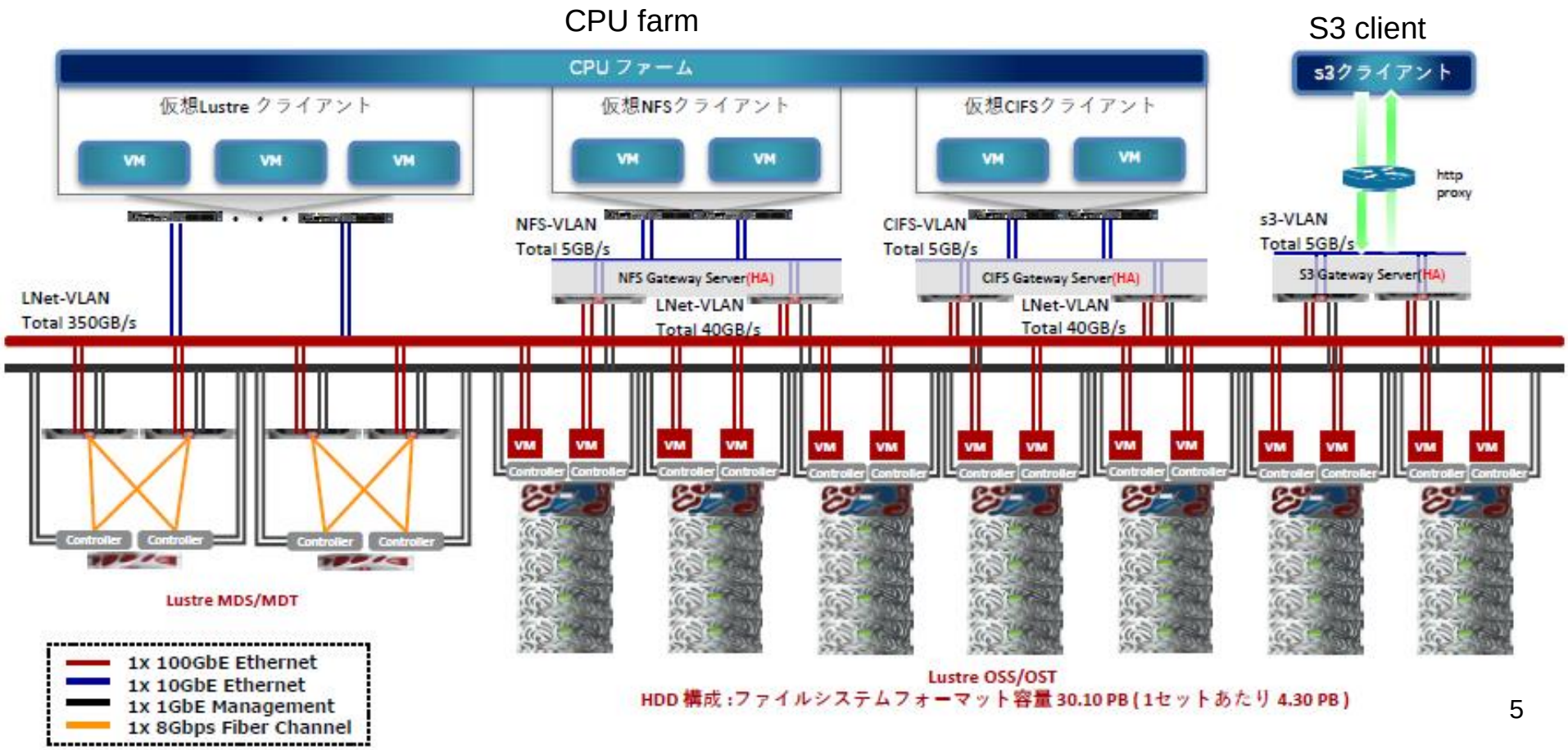
Outline of SS

- The lease period is from June 1, 2020 to May 31, 2026
  - Data farm
  - CPU farm
  - PIaaS (Private Infrastructure as a Service)



Data farm

- DDN ES7990X 7 unit
  - 30 PB、EXAScaler (Lustre) filesystem, 350 GB/s
  - NFS/CIFS gateway, URL access gateway 2 node each



CPU farm

- FUJITSU Server PRIMERGY CX2550 M5 440 node
  - Intel Xeon Platinum 8260 (2.40 GHz、24 core)
  - 2 CPU/node、21,120 core、1.62 PFlops
  - 384 GB (DDR4-2933)、SSD 1.92 TB、10GBASE-Tx2

システム全体構成	
総理論ピーク性能 (FP)	1.62 PFLOPS (3.68 TFLOPS x 440ノード)
総コア数	21,120コア (48コア x 440)
総メモリ容量	165.0 TB (384GB x 440 / 1024)

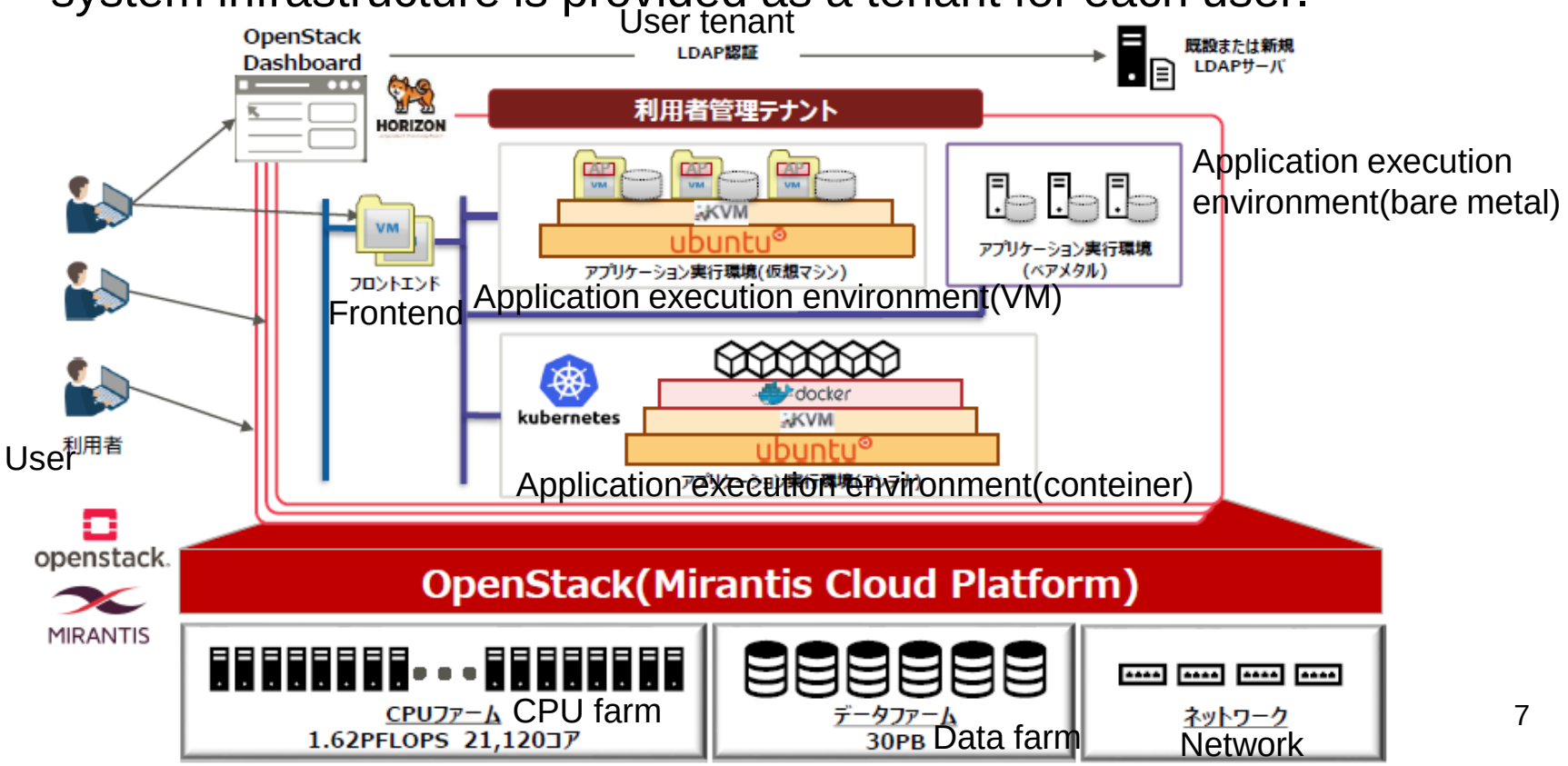
PRIMERGY CX2550 M5 [水冷]

PRIMERGY CX400 M4 [水冷]

1台あたりの構成		
CPU	プロセッサ	Intel Xeon Platinum 8260(2.4GHz/24コア) ※インテル64アーキテクチャ(x86_64) ※インテルバーチャライゼーション・テクノロジー含む(intel VT)
	プロセッサ数	2プロセッサ
	コア数	48コア(24コア x 2CPU)
	理論ピーク演算性能(FP)	3.6TFLOPS (2.4GHz x 32浮動小数点演算 x 24コア x 2CPU / 1000)
	理論ピーク演算性能(INT)	1.843TINOPS (2.4GHz x 16整数演算 x 24コア x 2CPU / 1000)
主記憶	種別	32GB DDR4 2933MHz RDIMM x 12 (ECC)
	容量	384GB (32GB x 12)
	コア当たりの容量	8GB (384GB / 48コア)
	メモリバンド幅	281GB/s (2933MHz x 8バイト x 12チャネル / 1000)
内蔵ディスク	OSブート用兼仮想環境の起動用	SSD-1.92TB x 1
データアクセスネットワークインターフェース	種別	10GBASE-T x 2ポート
	理論性能	20Gbps(10Gbps x 2ポート)
	接続先	データアクセスネットワーク用スイッチ
管理ネットワークインタフェース	種別	1000BASE-T x 1ポート
	接続先	管理・制御ネットワーク用スイッチ
電源		1600W x 2 (80PLUS PLATINUM)
筐体内監視		iRMC (CPU,メモリ,HDD,カード,ファン,電源,温度,電圧等)

PlaaS(Private Infrastructure as a Service)

- Mirantis Cloud Platform (MCP)
  - Commercial distribution of OpenStack that can provide VM, container and bare metal environment.
    - No bare metal environment in the operation
  - CPU farm, data farm, and network are virtualized by MCP, and system infrastructure is provided as a tenant for each user.



Cluster configuration of CPU farm and two types of usage

- Independent two clusters operation of CPU farm（hssa/hssb)
  - Dividing 440 nodes into 220 nodes each
- hssa cluster: pool type
  - User applies for the total number of vCPUs (the maximum number of vCPUs that can be used at the same time) and be charged by that number regardless of use.
  - Computational resources are pre-assigned and available at any time.
- hssb cluster: on-demand type
  - User applies the vCPU time product (the product of total vCPU and usage time) and be charged for the amount actually used within that range.
  - Computational resources can be used only when they are free.

Trial operation of SS and user classification

- Schedule
  - From 1:00 p.m. on June 1 to the start of the regular operation in October
- Differences from the regular operation
  - Basically the same
  - We will actually not collect the usage fee, but will simulate the procedures of the usage fee (usage points).
- User classification

Project member

  - Payment representative(PI)： project management, approval, tenant management
  - Assistant: the same authority as payment representative
  - Tenant manager: project application, tenant management
  - Sub manager: limited tenant management
    - Usage monitoring, start/stop instance, and using console
  - VM user: User of VM within tenants, managed by each project

User management sites of SS

- Web site of Information System Division
  - http://i.riken.jp/en/
  - User guide, etc.
- Online Submission System(hss-desk)
  - https://accc-desk.riken.jp/
  - Users access using the RIKEN authentication infrastructure
    - Login with Shibboleth authentication on the RIKEN authentication platform
    - The same User ID and Password as AIR100
  - Application and management of project
  - Management of computational resources and storage
  - Various documents
- OpenStack Control Console
  - https://hssa.riken.jp/ or https://hssb.riken.jp/
  - Accessible from inside RIKEN (or RIKEN VPN).
  - Tenant management

Application of project for trial operation of SS

- User can apply new project through the Online Submission System (hss-desk)
  - Registration of projects members
    - The registration requires member's account names
      - User registration by all member are required in advance at hss-desk.
    - Payment representive should be one person and PI（Manager)
      - No need to enter a budget number during the trial operation period.
    - If non-residents are present, permission review about security export control must be conducted.
  - Choose tenant type: pool or on-demand
    - For the pool type: apply the total number of vCPUs
    - On-demand type: apply the vCPU time product
  - A certain amount of computing resources are available immediately after application
    - 16 vCPU (8 physical core), memory 64 GB, storage 1TB
    - Available for 1 month

Project management for trial operation of SS

- Project management
  - Changing project information
  - Changing project member
  - Purchasing points(not actually charged during trial operation)
    - Buy 1 point for 1 JPY
    - Points are valid until the start of regular operation
    - Up to 1 million JPY during trial operation
  - Changing resources（without approval by payment representative）
    - Converting points to computational resources and storage capacity
    - Points can be returned from resources.
    - Applications for global IP addresses are accepted at hpc@riken.jp
- Approval of the application by payment representative
  - The application are valid after approval, except for resource changes.

Example of creating VM

1. Log in to the OpenStack Management Console
    1. hssa cluster for pool type and hssb cluster for on-demand type
2. Create and launch instance
    1. Source (image of instance): select CentOS-7.6-Application
    2. Flavour (setting of VM): select 1Core-8GiB-36GiB
    3. Network: select <project name>-network
    4. Network port: select <project name>-storage-port
    5. Security group: select <project name>-security-group
    6. Key pair: register public key to access from outside of SS
    7. Launch instance
3. Allocate floating IP address(to access from inside of RIKEN)
    1. Click compute-instance menu
    2. Click “allocate floating IP” at the right pull down menu of the instance
    3. Select “floating IP address” at address pull down from private-network pool

Example of using VM

- Login VM with SSH
  - SSH connection to the configured floating IP
  - $ ssh –l centos –i <private-key> <floating-ip>
- File transfer
  - Connect with SCP to the configured floating IP
- File system environment
  - /home: home area
    - Mount Lustre are as /home area in the case of -Lustre or *-Application images
  - /APL: application area
    - Mount application are as /APL in the case of *-Application image
- ISV/OSS environment
  - When the application area is mounted, the following ISV/OSS can be used by the module command
  - ISV: Intel compiler, Gaussian, GaussView
  - OSS: GROMACS, Python

Q&A

- Thank you very much for your attention.

- Q&A is only accepted by chat after the explanation is completed.
- We will answer verbally or via chat.

- Japanese and English versions of the Q&A will be available on the web at a later date.
  - If we need to check something, take it back and answer the Q&A.
  - The wrong answer is also corrected in the Q&A

APPENDIX

Provisions for services provided Head Office for Information Systems and Cybersecurity of RIKEN

(New regulation)

Provisions for services provided Head Office for Information Systems and Cybersecurity of RIKEN

Basic idea

Handling of usage fees for software licenses managed by the Information System Division (Notice at July 3, 2018)
・Usage support of software license

Target expansion

Handling of usage fees related to services provided by the Information System Division(Notice)
(will be reviced)
・Usage support of software license
・Usage support of computers
・Others (Large format printer)

The price of usage fees

Fees for shared use computer will be determined by the Notice.

Resources of shared use computers in FY2020

- HOKUSAI BigWaterfall(BW)(2017/10-2022/9)
  - BW-MPC: 840 nodes
    - CPU: Xeon Gold 6148(40 cores/node)
      - 2.58 PFlops、33,600 core
    - Memory: 96GB
  - Shared disk: 5PB
  - テープ: 8PB
- HOKUSAI SailingShip(SS)(2020/6-2026/5)
  - CPU farm: 440 nodes
    - CPU: Xeon Platinum 8260(48 cores/node)
      - 1.62 PFlops、21,120 core、42,240 v(virtual)CPU
      - Available 46 core/node (2 core for MCP hypervisor)
    - Memory: 384GB
  - Data farm（shared disk）: 30PB
  - Private Infrastructure as a Service(PIaaS)
    - Mirantis Cloud Platform(MCP)
      - Commercial distribution of OpenStack that can provide VM, container and bare metal environment

From “handling policy on beneficiary pays of ICT services”

- Settings of user fees
  - Up to 15% of total contract amount (about 100 million JPY)
    - Total contract amount includes hardware lease, maintenance and support, etc.
    - Not includes bills for building and electricity.
    - The upper limit is for use of computers and storage and not individually but as a whole.
    - Expecting to be lower than this because of subsidization and unused resources.
  - A usage fee will be introduced after starting regular operation of SS.
    - SS will be operated as test operation in the first half of FY2020, then regular operation will start after October FY2020.
    - The usage fee will also introduced for BW at the same timing.
  - Eliminate the review process and just check the usage content.
- Subsidization
  - Young scientist and exploratory research and promoting the strategy of RIKEN
  - Cooperation in operation and user support
- Easing measures
  - for existing users using large-scale resources
  - for existing data on the storage of HOKUSAI BW

How to use BW and fees

- In FY2020, the project will be reviewed as before.
  - Usage by batch job
  - Quick Use and General Use
    - In General Use, the allocated resources will be adjusted as if half of the allocated core hour was consumed at the beginning of the second half.
  - Usage fee will be introduced from October 2020 for the part of the use.
    - Usage fee will be set for priority job execution.
    - Charge for core hour
      - 1000 JPY per 1 core per 1 year (40,000 JPY per 1 node per 1 year)
    - A project for priority execution will be created and controlled for each project.
- Usage of BW After FY2021
  - In principle, it is subject to the usage fee (the price will be determined in the future), and no project review will be performed.

How to use SS and fees

- Lend tenants and charge for the amount of used computing resources
  - For pool type, resources are always reserved.
  - For ondemand type, resources are reserved when needed.
  - 4,000 JPY per 2vCPU per year.
    - In SS, a physical core is allocated by 2v(virtual)CPU.
  - Fee for global IP address in the future.

Physical core	vCPU	Memory(GB)	JPY/month	JPY/year
1	2	8	333	4,000
2	4	16	667	8,000
4	8	32	1,333	16,000
8	16	64	2,667	32,000
16	32	128	5,333	64,000
32	64	256	10,667	128,000
46	92	368	15,333	184,000

How to use storage and fees

- Fees for storage
  - Fees for allocated disk area is 2,000 JPY per 1TB per year
    - For existing data in BW allocated by the end of FY2019
      - From October 2020 to March 2021, 500 JPY per 1TB per half year
      - After FY2021, 2,000 JPY per 1TB per year
    - Fees for /home area is free.
    - Possible treatment of non-payment data
      - pay by center, move to tape, or no access
  - Tapes are not subject to fees for long-term storage
    - Removed from user services in the future and operated as cold media
      - Data storage service will move to disk
    - Possibility of cold media procurement when replacing BW
  - Preparing online storage (Box) separately
    - Assumed to be used for writing papers and general work
    - Assuming non-large data

Usage fee table and estimated amount

	JPY/(core・year)	JPY/(node・year)	Note
SS Tenant(Pool type)	4,000	184,000	for the secured resouces
SS Tenant(Ondemand type)	4,000	184,000	for the amount actually used
BW Batch(Priority use)	1,000	40,000	for priority execution
	JPY/(TB・month)	JPY/(TB・year)	
Storage	167	2,000	
Storage(Existing BW data)	167	2,000	half a year 500 yen in FY2020
Tape	0	0	will be cold stroage

Fees for BW after FY2021 will be discussed in the future.

	Paid use(node)	JPY/node	Subtotal(JPY)
SS （440 node）	200	184,000	36,800,000
BW （840 node）	400	40,000	16,000,000
	Paid use(PB)	JPY/TB	
SS Disk （30PB）	15	2,000	30,000,000
BW Disk （5PB）	2	2,000	4,000,000
Total			82,800,000

It is assumed that “data for use and application” will not be charged.

Payment method for usage fees (tentative)

- Plan to lower obstacles to start usage
  - The first small amount is free or available without confirmation.
- Unit of use
  - For pool type tenant use and storage, purchase in advance for the usage period.
    - Usage period is basically monthly
  - For ondemand type tenant use and batch use, core time is purchased in advance.
    - Set the minimum unit of core time (about tens of thousands of yen)
- Payment method
  - Purchase resources at any time and transfer the budget about once every three months
    - The end of the quarter is possible, but adjustments are needed at the end of the fiscal year.
  - Shared use computer point (token)
    - First you purchase points, then purchase resources with points.
    - Purchased points are valid until the end of the fiscal year.