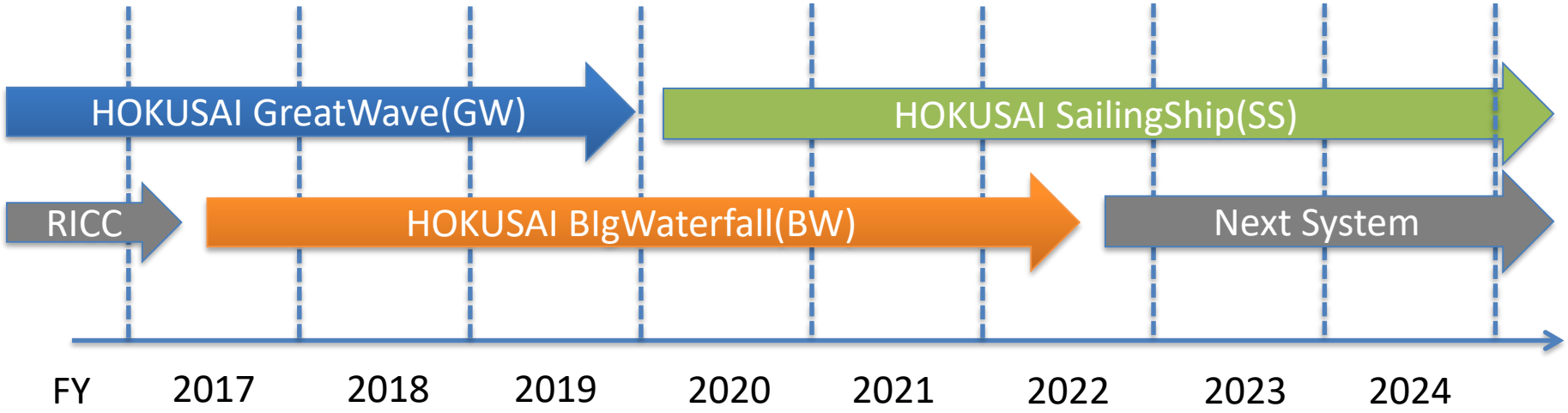
資料1

HOKUSAI SailingShipのトライアル運用

情報システム本部
情報システム部
情報化戦略・基盤課

共同利用計算機運用スケジュール

- 2015年4月にHOKUSAI GreatWave (GW) システムを運用開始し、2020年3月末に運用終了。
  - 1080 nodes, CPU: SPARC64-XIfx, 2PB
- 2017年10月にHOKUSAI BigWaterfall (BW) システムを運用開始し、2022年9月に運用終了予定。
  - 840 nodes, CPU: Xeon Gold 6148, 5PB
- 2020年6月頃に理研データ科学基盤HOKUSAI SailingShip(SS)システム運用開始予定。10月頃に本運用開始予定。
  - 440 nodes, CPU: Xeon Platinum 8260, 30PB



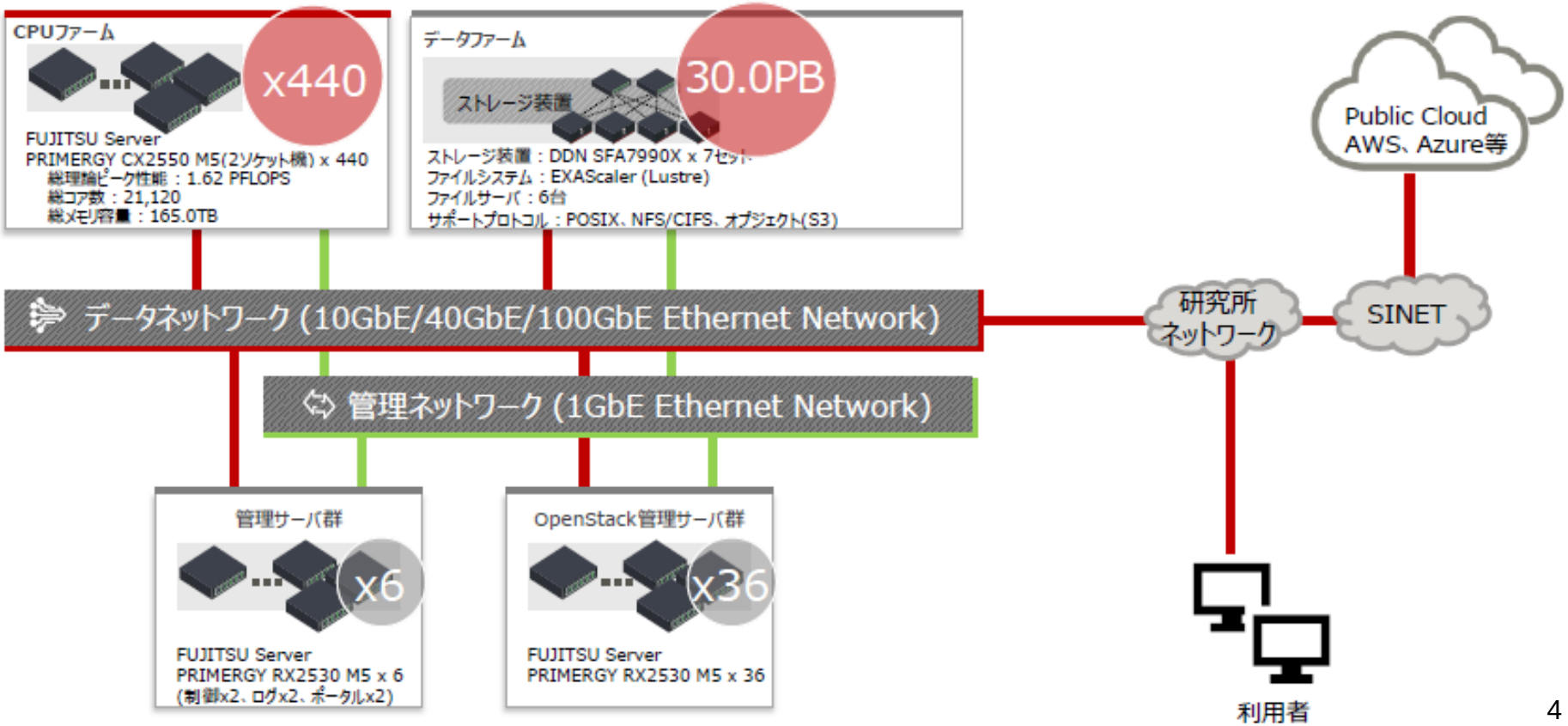
理研データ科学基盤HOKUSAI SailingShip(SS) 運用スケジュール

- 調達プロセス
  - 2018年11月 調達開始（RFI公告）
  - 2019年11月27日 開札→富士通の提案に決定
- 運用スケジュール
  - 2020年6月1日13:00 トライアル運用開始
  - 2020年10月 本運用開始
- 利用者説明会
  - 2019年10月3日 利用負担金導入
  - 2020年3月24日 データ科学基盤の紹介と利用負担金
  - 2020年6月1日 トライアル運用の案内
  - 2020年8-9月頃 本運用の案内

新型コロナの影響でスケジュールが遅れるも、
ぎりぎり予定通りにトライアル運用開始。
ただし、準備不足な点はご理解下さい。

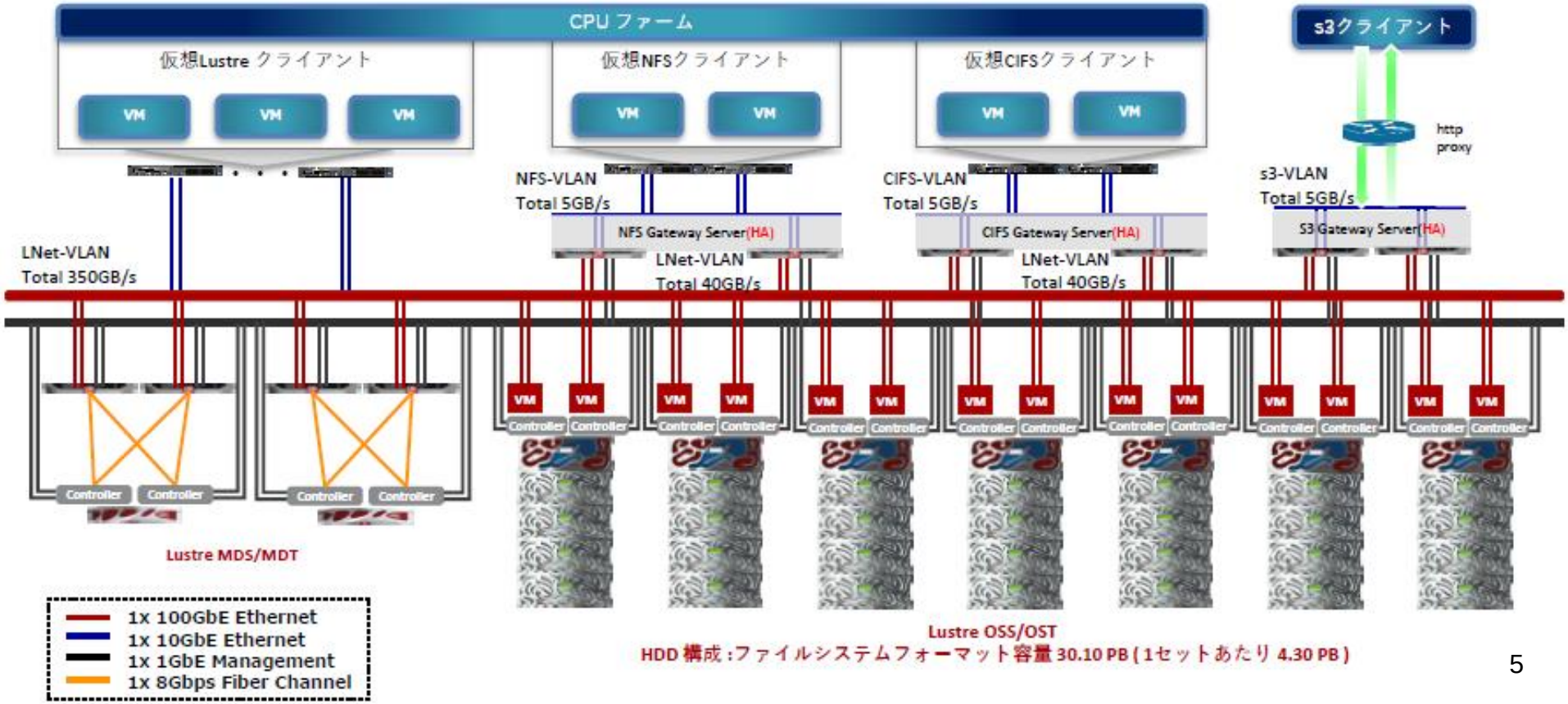
SSの概要

- 借入期間は2020年6月1日から2026年5月31日
  - データファーム
  - CPUファーム
  - PIaaS (Private Infrastructure as a Service)



データファーム

- DDN ES7990X 7台
  - 30 PB、EXAScaler(Lustre)ファイルシステム、350 GB/s
  - NFS/CIFSゲートウェイ、URLアクセスゲートウェイをそれぞれ2 node



CPUファーム

- FUJITSU Server PRIMERGY CX2550 M5 440 node
  - Intel Xeon Platinum 8260 (2.40 GHz、24コア)
  - 2 CPU/node、21,120コア、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)

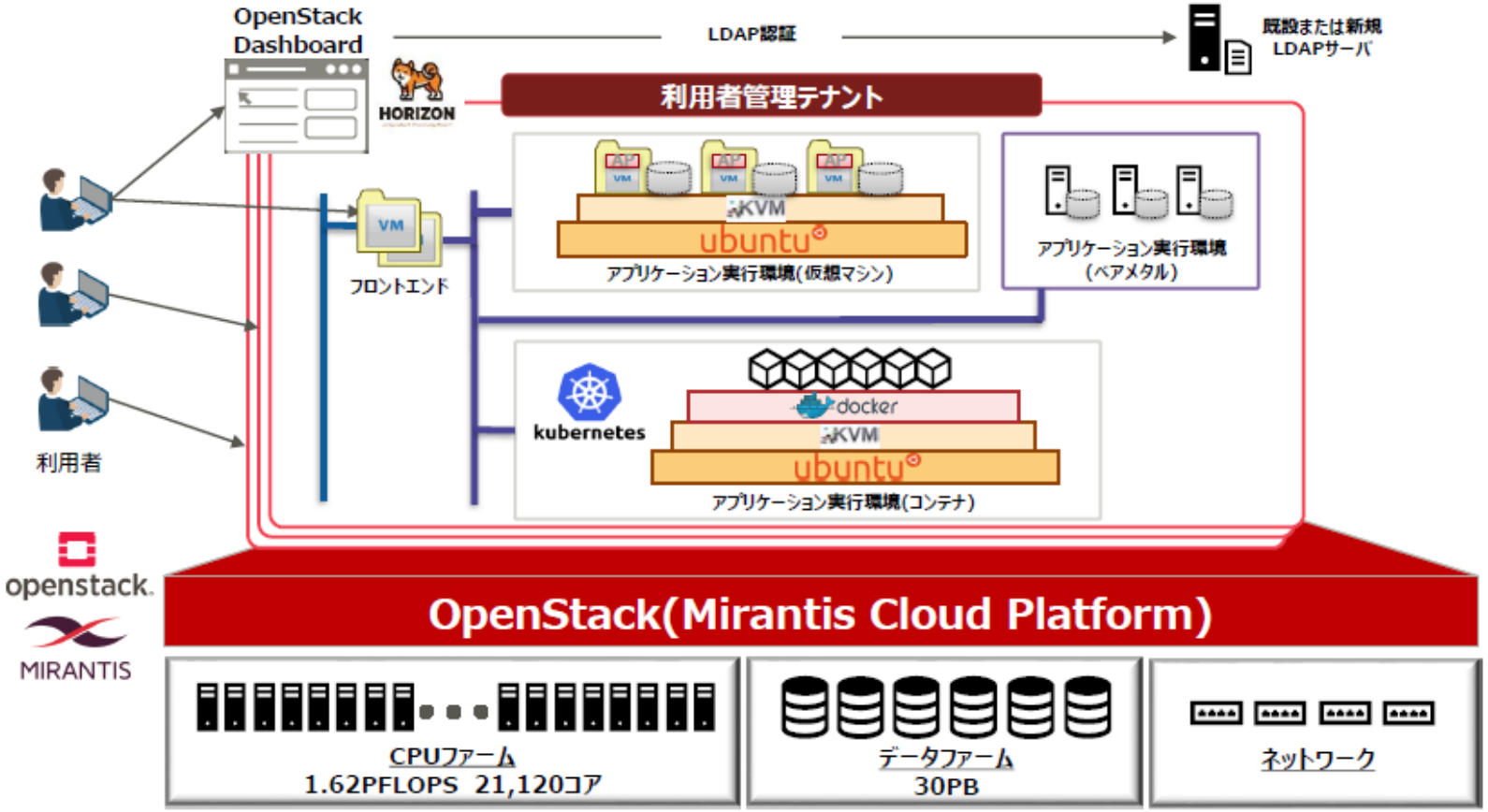
PRIMERGY CX2550 M5
[水冷]

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)
  - VM、コンテナ、ベアメタル環境を提供可能なOpenStackの商用ディストリビューション
    - 本運用ではベアメタル環境の提供はなし
  - CPUファーム、データファーム、ネットワークをMCPで仮想化し、利用者ごとにシステムインフラをテナントとして提供



CPUファームのクラスタ構成と2種類の利用方法

- CPUファームは独立した 2 クラスタ運用（ hssa/hssb ）となっています。
  - 440ノードを220ノードずつに分割
- hssa クラスタ:プール型
  - 総vCPU数（同時に使える最大のvCPU数）を申請し、利用に関わらずその数で課金される方法
  - 計算資源はあらかじめ確保されており、いつでも利用可能
- hssb クラスタ:オンデマンド型
  - vCPU時間積（トータルのvCPUと利用時間の積）を申請し、その範囲で利用した分だけ課金される方法
  - 計算資源が空いているときだけ計算資源を利用可能

SSのトライアル運用方針と利用者区分

- 日程
  - 6月1日13時から10月の本運用開始まで
- 本運用との違い
  - 基本的には同じ。
  - 実際の利用負担金の徴収は行わないが、利用負担金（利用ポイント）に関する手順や使い方などのシミュレーションを行う。
- 利用者区分

プロジェクトのメンバー

  - 支払責任者(PI):プロジェクトの管理、承認、テナントの管理
  - アシスタント:支払責任者と同じ権限
  - テナント管理者:プロジェクトの管理、テナントの管理
  - サブ管理者:テナントの一部の操作
    - 使用状況の参照、インスタンスの起動/停止、コンソールの利用
  - VM利用者:テナント内のVMの利用者、各プロジェクトで管理

SSの利用者管理サイト

- 情報システム部のwebサイト
  - http://i.riken.jp/
  - 利用案内など
- オンライン受付システム(hss-desk)
  - https://accc-desk.riken.jp/
  - 利用者が理研認証基盤を利用してアクセス
    - ログインに理研認証基盤によるShibboleth認証を利用
    - AIR100と同じUser IDとPassword
  - プロジェクトの申請と管理
  - テナントで利用する計算資源の管理
  - 各種ドキュメント
- OpenStack管理コンソール
  - https://hssa.riken.jp/ or https://hssb.riken.jp/
  - 理研内（もしくは理研VPN）からのみアクセス可能
  - テナントの管理

SSトライアル運用の新規プロジェクト作成

- オンライン受付システム(hss-desk)でプロジェクトを申請
  - プロジェクトのメンバーの登録
    - 登録にはメンバーのアカウント名が必須
      - 全てのメンバーは事前にhss-deskでユーザ登録
    - 支払責任者は必ず一人でPI(所属長)に限る。
      - トライアル運用期間中は予算番号の入力は必要なし。
    - 非居住者がいる場合は安全保障輸出関連の所内審査を行う。
  - テナント利用形態をプール型とオンデマンド型から選ぶ。
    - プール型は総vCPU数を申請
    - オンデマンド型はvCPU時間積を申請
  - 申請後すぐに一定量の計算資源を利用可能
    - 16 vCPU (8 物理コア), メモリ 64 GB, ストレージ 1TB
    - 1ヵ月利用可

SSトライアル運用のプロジェクト管理

- プロジェクト管理
  - プロジェクト情報変更
  - メンバー変更
  - ポイント購入申請（トライアル運用中は実際の料金請求は行わない）
    - 1ポイントを1円で購入
    - 本運用開始まで有効
    - トライアル運用中は100万円分を上限
  - リソース変更（支払責任者の承認なしで即時反映）
    - ポイントを計算資源とストレージ容量に変換
    - リソースからポイントに戻すことも可能
    - グローバルIPアドレスの申請は当面はhpc@riken.jpで受付
- 支払責任者による申請の承認
  - リソース変更以外は、承認後に申請内容が反映

VM作成の例

1. OpenStack管理コンソールにログイン
   1. プール型はhssa cluster、オンデマンド型はhssb cluster
2. インスタンスの作成と起動
   1. ソース（インスタンスのイメージ）：CentOS-7.6-Applicationを選択
   2. フレーバー（VMの設定）：1Core-8GiB-36GiBを選択
   3. ネットワーク：<プロジェクト名>-networkを選択
   4. ネットワークのポート：<プロジェクト名>-storage-portの中から1つ選択
   5. セキュリティグループ：<プロジェクト名>-security-groupを利用
   6. キーペア：SS外部からSSHで接続するための公開鍵を登録
   7. インスタンスを起動
3. フローティングIPアドレス割当（理研内からのアクセス用）
   1. コンピュート-インスタンスメニューをクリック
   2. 変更したいインスタンス右側のプルダウンメニューのFloating IPの割り当てをクリック
   3. IPアドレスプルダウンからprivate-networkプールからフローティングIPアドレスを選択

VM利用の例

- SSHログイン
  - 設定したフローティングIPに対してSSHで接続
  - $ ssh –l centos –i <private-key> <floating-ip>
- ファイル転送
  - 設定したフローティングIPに対してSCPで接続
- ファイルシステム環境
  - /home：ホーム領域
    - *-Lustre か *-Applicationイメージの場合はLustre領域を/homeとして自動的にマウント
  - /APL：アプリケーション領域
    - *-Applicationイメージの場合はアプリケーション領域を/APLとして自動的にマウント
- ISV/OSS環境
  - アプリケーション領域をマウントしている場合、moduleコマンドで以下のISV/OSSが利用可能
  - ISV：インテルコンパイラ、Gaussian, GaussView
  - OSS: GROMACS, Python

Q&A

- ご清聴ありがとうございました。

- Q&Aは一通り終わった後で、チャットでのみ受付
- 口頭もしくはチャットで回答

- 後日Q&Aの日英版をwebで公開予定
  - 確認が必要がものは持ち帰ってQ&Aで回答
  - 間違った回答などもQ&Aで訂正

付録

情報システム本部が提供するサービスに関する規程の制定について

（新規制定）

情報システム本部が提供する
サービスに関する規程

基本的考え方

情報システム部が管理するソフトウェアライセンスに係る利用料の取扱いについて（平成30年7月3日通達）

・ソフトウェアライセンス利用支援

対象拡大

情報システム本部が提供するサービスに係る利用負担金の取扱いについて
（通達）

（改正予定）

・ソフトウェアライセンス利用支援
・計算機利用支援

・他　（大判プリンタ利用支援等）

具体的な利用負担金額等

共同利用計算機利用負担金は通達で定める

2020年度の共同利用計算機のリソース

- HOKUSAI BigWaterfall(BW)(2017/10-2022/9)
  - BW-MPC：840 nodes
    - CPU: Xeon Gold 6148(40 cores/node)
      - 2.58 PFlops、33,600コア
    - Memory: 96GB
  - 共有ディスク：5PB
  - テープ：8PB
- HOKUSAI SailingShip(SS)(2020/6頃-2026/5頃)
  - CPUファーム：440 nodes
    - CPU: Xeon Platinum 8260(48 cores/node)
      - 1.62 PFlops、21,120コア、42,240 v(virtual)CPU
      - ただし2コアはMCPハイパバイザーに使用するので46コアが利用可
    - Memory: 384GB
  - データファーム（共有ディスク）：30PB
  - Private Infrastructure as a Service(PIaaS)
    - Mirantis Cloud Platform(MCP)
      - VM、コンテナ、ベアメタル環境を提供可能なOpenStackの商用ディストリビューション

受益者負担の取扱方針から

- 利用負担金の設定
  - 契約総額の15%を上限とする（1億円程度）
    - 契約総額はハードウェアのリース代、保守・サポート代などを含む
    - 建物代や電気代などは含まない
    - 計算機やストレージなどの利用について、個別ではなく全体としての上限
    - 補助や未利用分などがあるので、実際はこれより低い金額に
  - データ科学基盤の導入後に利用負担金の設定を始める。
    - 2020年度上期はテスト運用で、10月からの本運用から利用負担金も設定する予定
    - BWについても同じタイミングとする。
  - 利用負担金を払う場合は審査をなくし、利用内容の確認程度に
- 補助
  - 若手や萌芽的な課題や理研の戦略推進としての課題
  - 共同利用計算機の運用に協力した人への補助
- 従来の利用者に対する緩和措置を取る
  - 大規模利用者に対する救済措置
  - HOKUSAI BWのストレージにすでに保存されているデータ

HOKUSAI BWの利用方法と利用負担金

- 2020年度は従来と同様課題審査を行う。
  - バッチジョブによる利用
  - 簡易利用と一般利用を募集
    - 一般利用は半年毎に区切って割り当て資源をリセットする。
  - 2020年10月から、一部に利用負担金導入
    - 基本的に無料で実行可だが、優先実行に負担金を導入
    - コア時間に対して課金
      - 1コアを1年分で1,000円（1ノード1年分で40,000円）
    - 各課題毎に優先実行用の課題を作り制御する。
- 2021年度以降のBWの利用方法
  - 原則利用負担金（料金は今後検討）対象とし、課題審査は行わない。

HOKUSAI SSの利用方法と利用負担金

- テナントを貸し出し、使用する計算資源量に対して課金
  - プール型は資源を常時確保して利用
  - オンデマンド型は必要な時に資源を確保して利用
  - 2vCPUを1年間の利用で、4,000円
    - データ科学基盤では1物理コアあたり2v(virtual)CPUを割り当て
  - 将来的にグローバルIPアドレスについて課金

物理コア	vCPU	メモリ(GB)	円/month	円/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

ストレージの利用方法と利用負担金設定

- ストレージの利用負担金
  - 割当てられたディスク領域について1TB当たり年間2,000円
    - BWの2019年度末までに申請された領域の場合
      - 2020年度10月以降は、1TB当たり半年500円
      - 2021年度以降は、1TB当たり年間2,000円
    - /homeは対象外
    - 不払いの場合の扱いの
      - センターに払っていただく、データをテープに移動、アクセス不可など
  - テープは長期保存用として利用負担金の対象外とする
    - 将来的に利用者サービスから外し、コールドメディアとして運用
      - データ預かりサービスはディスクに移動
    - BWの入れ替えの際にコールドメディア調達の可能性
  - 別途オンラインストレージ(Box)を準備中
    - 論文作成や一般的な業務で利用することを想定
    - 大規模でないデータを想定

利用負担金表と想定見積額(2年目以降)

	円/(core・year)	円/(node・year)	備考
SSテナント利用（プール型）	4,000	184,000	確保した資源に対して
SSテナント利用（オンデマンド型）	4,000	184,000	利用した分に対して
BWバッチ利用（優先実行）	1,000	40,000	優先実行に対して
	円/(TB・month)	円/(TB・year)	
ストレージ	167	2,000	
ストレージ（BWの既存データ）	167	2,000	2020年度は半年500円
テープ	0	0	コールドストレージ化の予定

ただし、HOKUSAI BWの2年目以降の利用料金は今後議論される。

	有料利用(node)	円/node	計（円）
SS（440 node）	200	184,000	36,800,000
BW（840 node）	400	40,000	16,000,000
	有料利用(PB)	円/TB	
SS Disk（30PB）	15	2,000	30,000,000
BW Disk（5PB）	2	2,000	4,000,000
合計			82,800,000

ただし、利活用データについては利用料金を取らないと想定される。

利用負担金の支払い方法(想定)

- 申請から利用開始までのハードルを下げることを検討中
  - 最初の少量の利用は無料もしくは、確認無しで利用可に。
- 利用の単位
  - プール型のテナント利用とストレージは、利用期間分を事前に購入。
    - 利用期間は基本的に月毎
  - オンデマンド型のテナント利用とバッチ利用は、利用予定分のコア時間を事前に購入。
    - コア時間の最低単位(数万円程度)を設定
- 支払方法
  - 資源の購入は随時で、3ヵ月に1回くらいの頻度で予算の振替を行う
    - 4半期末が考えられるが、年度末については考慮が必要
  - 共同利用計算機ポイント(トークン)
    - 一旦ポイントを購入後、ポイントで利用資源を購入。
    - ポイントは年度末まで有効