

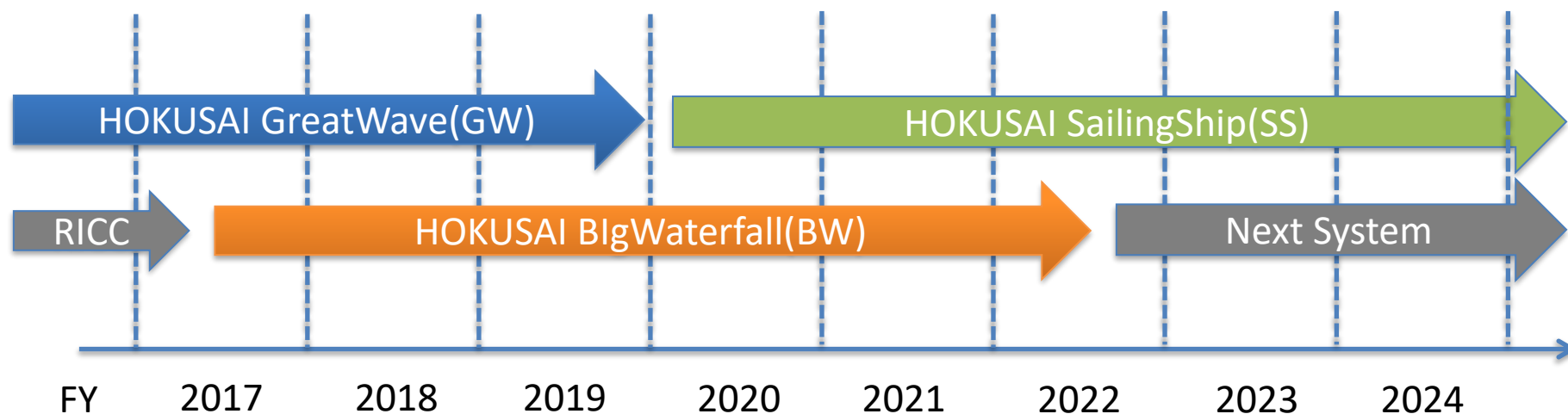
資料1

HOKUSAI SailingShipの本運用と HOKUSAI利用負担金

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19日に本運用開始予定。
 - 440 nodes, CPU: Xeon Platinum 8260, 30PB



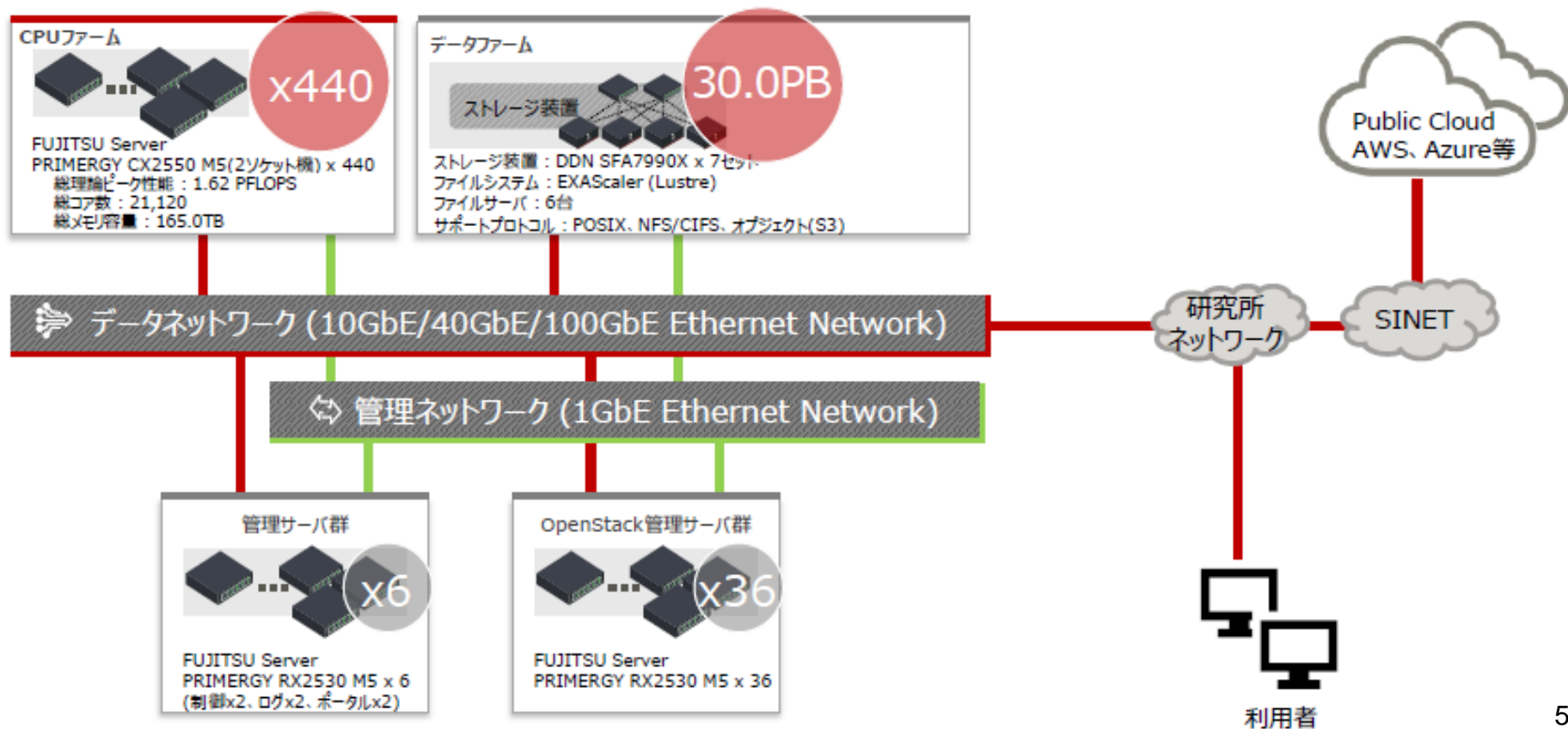
HOKUSAI SalingShipの本運用

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 – 10月9日 17:00 トライアル運用
 - 2020年10月19日 15:00 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9月25日 本運用と利用負担金の案内
 - 2020年9月29日 SS利用方法の講習会

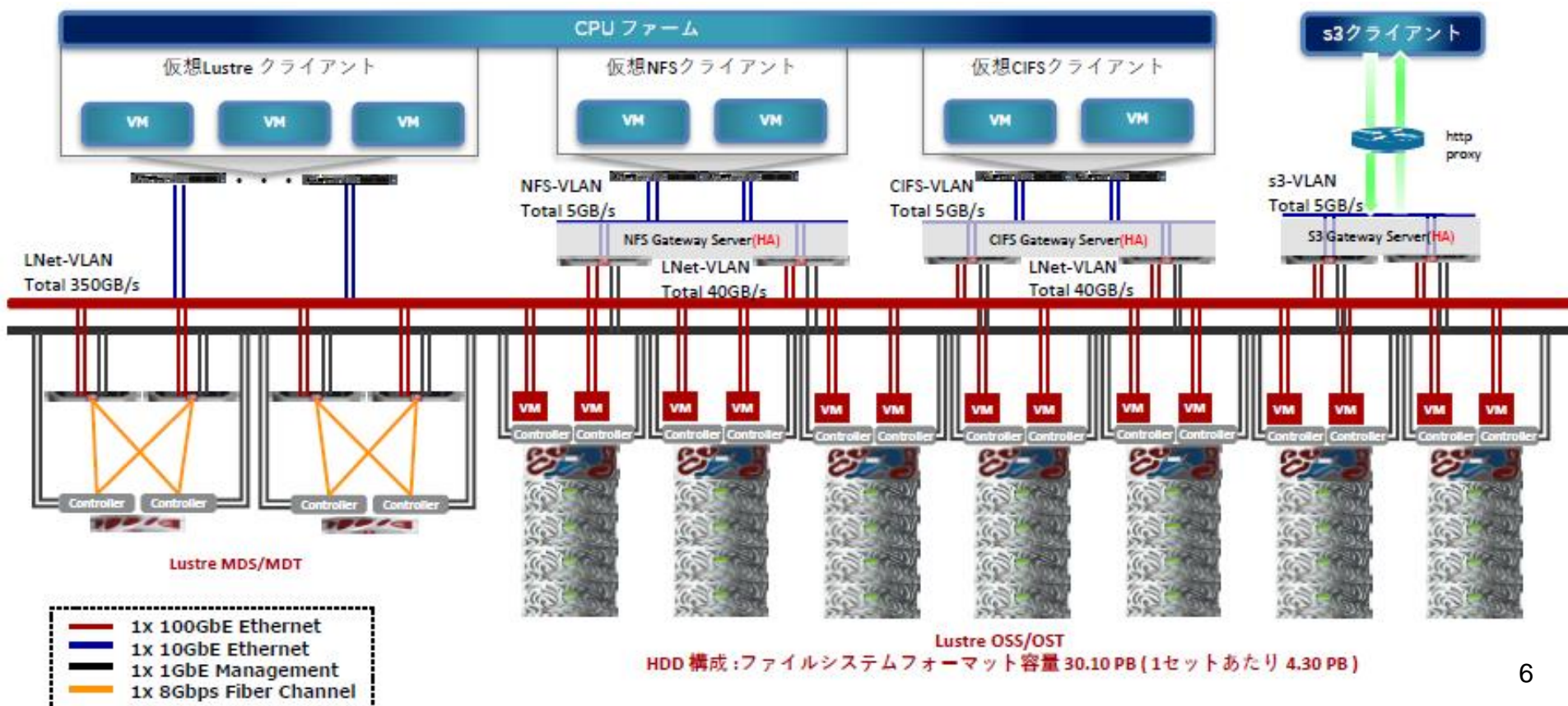
SSの概要

- 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
 - PlaaS (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]

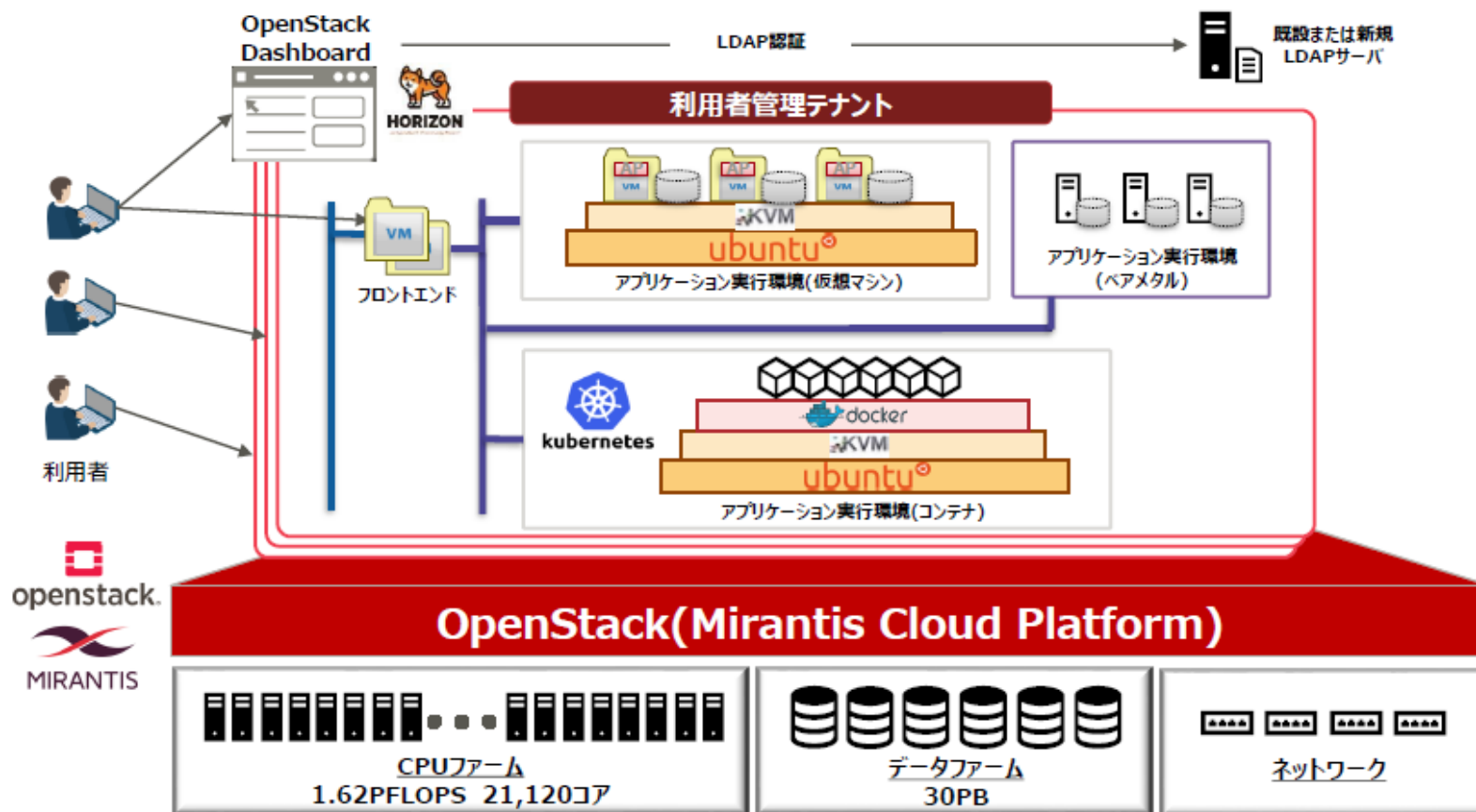


1台あたりの構成

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ーフェース	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
 - 計算リソースを常時確保しておき、利用に関わらずその数で課金される方法
 - 計算リソース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
 - 計算リソースを必要な時だけ確保し、利用した分だけ課金される方法
 - 計算リソースが空いているときだけ計算資源を利用可能

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
- オンライン申請システム(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
 - Pool type (hssa cluster): <https://hssa.riken.jp/>
 - On-demand type (hssb cluster): <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
 - VMの起動やネットワーク設定

SSの利用者区分と管理

- プロジェクトメンバー
 - 支払責任者:プロジェクトの管理、承認、テナントの管理
 - プロジェクトと予算支払の責任者で所属長であること。
 - 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
- 利用者の管理
 - 支払責任者責任でプロジェクトメンバーとVM利用者を管理
 - 非居住者の利用についても含まれる
 - HOKUSAI SSの使用に係るキャッチオール・チェックシートを作成・保存
- 計算リソースやストレージの容量の追加
 - まずポイントを購入(1ポイント1円)
 - そのポイントを使って計算リソースとストレージの容量の追加

SSの可用性とデータ保全

- 可用性

- 年間の97%以上のサービス提供時間を目指す。
- システム停止のタイミング(なるべく1か月前に予告)
 - 和光地区の停電時:計画転電10月の3連休
 - 例外的に2020年度は3回
 - 緊急メンテナンス:緊急性の高い脆弱性への対応などに対応するためのアップデートを行うためのメンテナンス
 - 停電や瞬電などのトラブル:可能な範囲で速やかに復旧

- データ保全

- ストレージ装置を含む装置全体は和光地区情報基盤棟にある。
- データは冗長性を持たせてあり、プール内のHDD2台の故障まで復旧可能。

SSのサポート範囲とセキュリティ対策

- サポート範囲
 - プロジェクトメンバーの利用者管理や提供しているイメージによるVM起動までの技術支援を行う。
 - テナント内のことは各プロジェクトで管理していただく。
- セキュリティ対策
 - フローティングIPが割り当てられたVMは理研中からアクセス可能なのでセキュリティ対策を行うこと。
 - グローバルFIPの場合はより気を付けること。
 - セキュリティグループは必要最小限だけ許可する。
 - ポートは必要最小限だけ開くようにする。
 - SSHやHTTPSなどの利用しているサービスは、定期的なアップデートと安全性の確保を行うこと。
 - 共有ディスク領域(Lustreファイルシステム)をマウントしているVMは、カーネルを最新に保つことが難しいのでより注意して管理すること。
 - カーネルをアップデートするとLustreファイルシステムをマウントできなくなるため。
 - 安全性の確保ができない場合は、共有ディスク領域をマウントしていないVMをゲートウェイサーバとして利用すること。

SS本運用の新規プロジェクト作成

- 事前の準備
 - 支払責任者に利用と予算の確認を取っておく。
 - プロジェクトメンバーは全員hss-deskでssアカウントを取得
 - ssアカウント名が登録に必要
- hss-deskでプロジェクトを申請
 - 支払責任者の情報入力
 - プロジェクトの情報入力
 - 予算番号とその名称を入力
 - 利用内容の入力
 - テナント利用形態をプール型とオンデマンド型から選択
 - プロジェクトメンバーの登録
- 支払責任者の確認前に一定量の計算資源を利用可能
 - 10,800ポイント(10,800円分)まで、支払責任者の承認なしでポイント購入し、ポイントをリソースに変換可能
 - 支払責任者に確認完了すれば上限以上のポイント購入可

SS本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
 - 購入の最小単位は360ポイントを360円
 - ポイント購入の上限は100万円
 - 上限の変更はメールで受付
 - リソースの割り当て(支払責任者の承認なしで即時反映)
 - ポイントで計算リソースとストレージ容量に変換
 - IPアドレスの追加やグローバルIPアドレスの申請はメールで受付
- 支払責任者による申請の承認
 - リソースの割り当て以外は、承認後に申請内容が反映

SSの本運用への移行と利用負担金の導入

- 10/19から本運用に移行予定
 - ポイントとCPUリソース(オンデマンド型)は本運用開始時に0にする。
 - CPUリソース(プール型)とストレージは10月中は無料とする。
 - BWのデータ領域(/data)も含む
- 本運用開始前の準備
 - SSTライアル利用のプロジェクトについて
 - 支払責任者にメールを送付
 - 継続利用の場合は予算番号を登録していただく。
 - BWのデータ領域(/data)を利用しているプロジェクトについて
 - 課題代表者にメールを送付
 - 継続利用の場合は支払責任者の設定と予算番号の設定と年度末までの利用料金の購入をしていただく。
- 本運用開始時の処理
 - トライアル運用だけのプロジェクトを停止する。
 - CPUファーム(プール型)とストレージの期限を10月末にする。
 - BWのデータ領域(/data)を継続利用されない場合利用不可にする。
- 本運用開始後から10月末までの猶予期間
 - 猶予期間中にプール型の計算リソースとSSのストレージを年度末まで購入していただく。

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

- アクセス

- フローティングIPを割り当てられたVMに対して、理研内からSSHやSCPでアクセス可能

- \$ ssh -l centos -i <private-key> <floating-ip>

- ファイルシステム環境

- /home: ホーム領域 (共有ディスク領域)

- *-Lustreイメージの場合は、Lustre領域を/homeとして自動的にマウント

- *-Lustreイメージ以外の場合は、/homeはローカルな領域

- /APL: アプリケーション領域

- CentOS-*-Lustre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して利用可能

- 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

研究情報管理サービス連携用 CIFSアカウントの発行

- 申し込み時にプロジェクトLustreストレージとクオタを共有する CIFS/SMB (共有ファイルシステム)のアカウントが自動発行されます
 - 研究情報管理サービスで使える予定です(想定)
 - ・ 今年度末をめどに発足する研究データ管理及び利活用データの公開基盤

10月中旬ごろから、サービス内容、利用法について告知していく予定ですので、CIFS アカウントを保管ください。

HokusaiSSのVMからは見えません
(見せたいときはNextCloudをWebDAVでマウントすることになります)
HokusaiSS Lustre領域のクオタに利用可能容量は含有されます
詳しくは 情報システム本部研究開発部門データ管理システム開発ユニット 實本
(hideyuki.jitsumoto@riken.jp)までお問い合わせください

HOKUSAI利用負担金

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(PlaaS)
 - 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

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月頃から、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
 - 1コアを720時間の利用で90円
 - 1コアを1年間の利用で約1,080円
 - 1ノード(40コア)1年間の利用で約43,200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。

SSの利用方法とCPUファームの利用負担金

- テナントを貸し出し、利用者がVM環境を構築して利用
 - 利用方法はプール型とオンデマンド型の2種類
 - SSでは1コアあたり2v(virtual)CPUを割り当て
 - 2vCPUあたり、メモリ8GB、ローカルストレージ(SSD)36GBを割り当て
- プール型は資源を常時確保して利用
 - 年度末まで月単位の利用
 - 2vCPUを1ヵ月の利用で360円
 - 2vCPUを1年間の利用で4,320円
 - 1ノード(92vCPU)1年分で198,720円
- オンデマンド型は必要な時に資源を確保して利用
 - 実際に利用する分について課金
 - 2vCPUを720時間の利用で360円
 - 2vCPUを1年間の利用で約4,320円
 - 1ノード(92vCPU)1年間の利用で約198,720円

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

- BWの/dataとSSの共有ディスク領域について課金
 - BWの/homeは無料
- 確保するディスク領域について1TB当たり1月180円
 - 利用開始月から年度末までの利用料金
- BWの2019年度末までに申請された領域の特例
 - 2020年度10月以降は、1TB当たり1月90円
 - 2021年度以降は通常料金

- テープ領域は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

- テープは長期保存用として、利用負担金の対象外とする
- BWの入れ替えの際にコールドメディア調達の可能性

利用負担金表と想定見積額(2年目以降)

SSテナント利用（プール型）	2vCPUを1ヵ月	360	確保する資源に対して年度末まで月単位
SSテナント利用（オンデマンド型）	2vCPUを720時間	360	利用した分に対して
BWバッチ利用（優先実行）	1coreを720時間	90	優先実行に対して
ストレージ（SSとBW[/data]）	1TBを1か月	180	確保する資源に対して年度末まで月単位
ストレージ（BW[/data]の 2019年度末までに申請された領域）	1TBを1か月	90	2020年年度のみ
テープ	0	0	コールドストレージ化の予定

	有料利用(node)	円/node	計（円）
SS（440 node）	220	198,720	43,718,400
BW（840 node）	420	43,200	18,144,000
	有料利用(PB)	円/TB	
SS Disk（30PB）	15	2,160	32,400,000
BW Disk（5PB）	2.5	2,160	5,400,000
合計			94,262,400

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。

SSの利用負担金の計算方法

- CPUファーム(プール型)とストレージは、利用期間分を事前に購入
 - 利用期間は月単位で年度末までの利用
 - 日割りは行わず、月の途中でもその月の分の課金
 - CPUファーム(プール型)の例
 - 7月20日から年度末(9ヵ月)まで92vCPUの利用の場合
 - $\frac{92}{2} \times 9 \times 360 = 149,040$ 円
 - 共有ディスク領域の例
 - 7月20日から年度末(9ヵ月)まで10TBの利用の場合
 - $\frac{10}{2} \times 9 \times 180 = 16,200$ 円
- CPUファーム(オンデマンド型)は、利用予定分のコア時間を事前に購入
 - 購入単位は1440vCPU時間(2vCPUを720時間に相当)
 - 1440vCPUを500単位の場合
 - $\frac{500}{2} \times 360 = 180,000$ 円
- ポイントシミュレーターをhss-deskのメニューに設置予定
 - vCPU数、TB、vCPU時間の単位数を入力し、必要なポイントを計算

SSのリソースとポイントの購入方法

- SSの計算リソースとストレージの購入はポイントでのみ可能
- ポイントの購入
 - プロジェクト作成時に予算番号を登録。
 - 組織、プロジェクト、費目のコードと名称を入力。
 - コード:6桁の数字-12桁の数字-6桁の数字
 - 名称:空白でなければOK
 - 予算番号を指定してポイントを購入。
 - 登録されている予算番号から選ぶ。
 - 予算毎の年度末の購入期限に注意。
 - ポイントの購入単位は360ポイント(360円)
 - ポイントは年度末まで有効。
- 取消と返却
 - ポイント購入は支払責任者の承認後は取消しない。
 - 割り当てられたリソースはポイントに返却しない。
 - ただし、理研からの離職や外部資金の期限がある場合は例外とする。

BWのリソースの購入方法

- ・ BWについては計算リソースとストレージをポイントを間に入れて直接購入
- ・ リソースの購入方法
 - － Webフォームで受付予定
 - ・ 支払責任者と予算番号の登録
 - ・ 優先実行用のコア時間かデータ領域(/data)の容量を申請
- ・ 優先実行の計算リソース追加
 - － 優先実行については専用の課題を作成
 - ・ $Q \rightarrow P$ 、 $G \rightarrow F$
 - ・ 専用の課題のジョブは、通常の課題のジョブより優先的にジョブを開始する。
 - ・ 専用の課題にコア時間を設定した分だけ、元々の課題から差し引く。
- ・ データ領域(/data)のデータ保持
 - － データ領域(/data)のデータは利用終了後のデータ保持は保証しない
 - ・ 年度末に継続利用の確認
 - ・ ホーム領域(/home)は課題終了後6ヵ月はデータを保持

予算番号の登録と予算の振替

- 予算番号の登録
 - 外部資金については支払責任者が事前に外部資金室に確認
 - 支払可能か、支払い方法、年度末の締切日、その他条件など
- 予算振替のタイミングは6月末、9月末、12月末、年度末
 - 各締切前に購入された分をまとめて振替。
 - ポイント購入は基本的には12月末までに。
 - 年度末の振替は予算の支払期限に間に合うように予算毎に調整。
- 想定される外部資金の取り扱い
 - 比較的問題が少ない: JST、AMED、文科省系
 - 個別対応が必要: NEDO、経産省系
 - 先払いが難しいので利用した分だけ振替。
 - ポイントやオンデマンド型の計算リソースを使い切る必要性。

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

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