

令和6年度

「次世代計算基盤に係る調査研究」

(運用技術調査研究)

委託業務成果報告書

令和7年5月

国立大学法人 東京大学

本報告書は、文部科学省の科学技術試験研究委託事業による委託業務として、東京大学が実施した令和6年度「次世代計算基盤に係る調査研究」（運用技術調査研究）の成果をまとめたものです。

目次

I.	委託業務の目的	4
II.	2024 年度の実施内容	4
1.	実施計画	4
2.	実施内容	7
2.1	プロジェクトの総合的推進	7
2.2	施設・設備	7
2.3	カーボンニュートラル	10
2.3.1	カーボンニュートラルに関する技術動向	10
2.3.2	導入コストの分析	16
2.3.3	次世代計算基盤にむけた HPC データセンターの構成	19
2.3.4	温水冷却による PUE および OPEX の試算	21
2.3.5	カーボンニュートラル化に関するまとめ	23
2.4	資源管理	24
2.4.1	資源管理の調査研究内容	24
2.4.2	既存計算基盤の調査	25
2.4.3	一体的運用の調査研究の進め方	27
2.4.4	機能と構成（案）	28

2.4.5	資源管理まとめ	37
2.5	データ利活用	38
2.6	Society5.0 運用	39
2.6.1	複数拠点 HPC システム間の統一的管理とワークフロー	39
2.6.2	JupyterLab からスパコンシステムへの透過的なジョブ実行	43
2.7	HPCI 運営	44
2.7.1	ユーザの利便性向上	45
2.7.2	次世代認証基盤の調査	48
2.7.3	ストレージの端境期対応	49
(1)	端境期問題の背景と課題	49
(2)	ストレージ端境期問題の性質：更新タイミング	50
(2-1)	ストレージ独立調達・運用モデル	50
(2-2)	大規模共用ストレージ基盤の活用モデル	51
(2-3)	データ転送・共有ツールの活用	52
(3)	ストレージ移行に伴う技術的課題の整理	53
(4)	ストレージの技術動向	54
(a)	オブジェクトストレージ	54
(b)	AI ワークロード向けストレージ	55
(5)	運用モデルのコスト比較	56
(6)	将来のストレージ運用の理想像	57

(7) ストレージの端境期対応まとめ	58
2.7.4 センター間連携に向けた調査、今後の導入計画の調査	58
2.7.5 HPCI の機能拡充に向けて	60

I. 委託業務の目的

ポスト「富岳」時代の次世代計算基盤実現を目指し、国内主要スパコン、データ基盤、ネットワークが一体となって、デジタルツイン技術の進化を支え、Society 5.0の推進、SDGsの達成に貢献するプラットフォームを実現するにあたり、次期システム群の設計・運用、それに伴う施設整備・管理、ソフトウェア管理、利用者サービス、ネットワーク整備、セキュリティ対応など、運用技術の観点から調査研究を実施することを目的とする。

このため、国立大学法人東京大学を代表機関として、参画機関である国立研究開発法人理化学研究所、大学共同利用機関法人情報・システム研究機構国立情報学研究所、国立大学法人東京工業大学（現在、東京科学大学）と密接に連携し、再委託により研究開発を実施する。

II. 2024年度の実施内容

1. 実施計画

2024年度には以下の項目について実施する。

1. Society5.0運用

スパコンとクラウドのように、異なるワークロードに適したプラットフォームをそれぞれ選んだ上でそれらを連携する機構について、具体的なユースケースに基づいてプロトタイプ的设计を継続する。複雑な処理内容をスマートに実現するワークフローや、大規模データの取り扱いについて、データ利活用のグループとも共同して検討を継続する。Fuzzballによる複数コンテナ制御機能を用いて、クラウドとスパコン間の連携を想定したワークフローの試行を実施する。

2. 資源管理

国内主要スパコン、データ基盤、ネットワークが一体的に運用される基盤の構築・運用とシームレスなクラウド連携を可能にする上での課題検討を継続する。一体的な運用に向け、統一ユーザインタフェースの検証、各スパコンから収集する運用情報の詳細化を更に進めるとともに、国内主要スパコンやクラウドとの連携に向けて各拠点へのヒアリングを進め、HPC 共通セキュリティガイドラインを作成する。

3. 施設・設備

国内外の高温温水冷却の運用例、さらに進んだ冷却方式について調査を継続する。昨年度までに東京大学・東京工業大学（現・東京科学大学）に設置したスパコン用電力測定環境を用いた、実際のデータ計測・蓄積を行い、プログラムからリアルタイムで電力計測値が参照・活用できることを確認する。

4. カーボンニュートラル

カーボンニュートラルを実現するための要素技術やエネルギー調達手法および排熱の有効活用について詳細な調査を継続する。また、地理的条件である気候やインフラの整備状況および周辺環境などを考慮した具体的なシナリオに基づくデータセンター建屋のカーボンニュートラル化について検討するとともに、次世代フラッグシップの導入に向けてデータセンター建屋に関する要求要件を整理する。

5. データ利活用

NII が運用する研究データ基盤をハブとして活用しながら、各種ストレージや計算基盤を連携させるための調査検討を継続する。大規模データを移動することなく、適切なセキュリティを維持しながら、データ保全やデータ連携、アーカイブも含めて効率的に行う方法について検討を継続する。さらに、従来から用いられてきたコマンドラインによるジョブ実行に加えて、Jupyter Notebook や Open OnDemand など新しいスタイルのジョブ実行も含めた計算資源の利用方法を考慮し、様々なデータ利活用シナリオを想定した調査分析を行う。

6. HPCI 運営

HPCI において計算資源を提供している各センターについて、横断的な運用ポリシーや、サービスレベルの考え方、利用負担金の考え方について、センター間連携、シームレスな利用の観点から調査・ヒアリングを継続する。富岳からポスト「富岳」への移行期間において、各センターの導入計画をもとに、HPCI 全体の計算資源の見積もり

と計算資源不足への対策について検討を継続する。また、欧州の事例に引き続き米国についても、計算資源の整備・運用の支援、アプリケーションの保守・高度化の促進機能について、より詳細な調査を実施する。加えて、本調査研究業務の終了後も、HPCI 運営の改善を継続的に推進する体制の構築を検討する。また、端境期におけるユーザデータのストレージ間移行について、既存の大規模ストレージ群、データ共有サービスとの連携も踏まえて、なるべく端境期をユーザに意識させない利用シナリオと技術検討を実施する。

7. プロジェクトの総合的推進

プロジェクト全体の連携を密としつつ円滑に運営していくため、本調査研究全体の運営委員会や技術検討会に参加し、参画各機関との連携・調整にあたる。

特に、プロジェクト全体の進捗状況を確認しつつ計画の合理化を検討し、必要に応じて調査或いは外部有識者を招聘して意見を聞くなど、プロジェクトの推進に資する。

プロジェクトで得られた成果については、積極的に公表し、今後の展開に資する。

2. 実施内容

2.1 プロジェクトの総合的推進

プロジェクト全体の連携を密としつつ円滑に運営していくため、本調査研究全体の運営委員会や技術検討会に参加し、参画各機関との連携・調整にあたった。

提案の際には、前述の通り 6 つのワーキンググループにおいて個々の調査研究を実施するとしていたが、昨年度プロジェクト全体の進捗状況を確認しつつ計画の合理化を検討した結果、本年度もワーキンググループの体制は残しつつ、以下の 3 つのグループで議論を進めた。

1. 施設・設備、カーボンニュートラルなどのインフラに関する議論
2. 資源管理、Society5.0 運用、データ利活用などのシステム設計・システムソフトウェアの開発に関する議論
3. HPCI 等の運用ポリシー、センター間連携、ユーザの利便性向上などの運用やユーザ視点での改善に向けた議論

プロジェクト全体での打ち合わせは月に 1 度、ワーキンググループのリーダーを中心とした幅広いメンバーで実施し、2024 年 4 月から 2025 年 3 月まで 10 回実施した。内容は、プロジェクト全体の進め方、ワーキンググループ間での情報共有、本調査研究全体の運営委員会や評価委員会に向けた報告のまとめやフィードバック、マイルストーンを踏まえた進捗管理を中心に実施した。

2.2 施設・設備

本節ではこれまでに実施してきた HPC データセンター設備に関する調査に関する報告をまとめる。

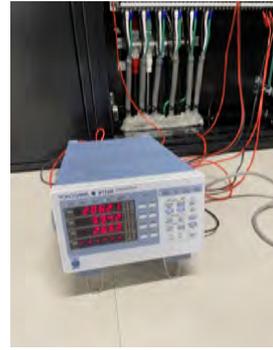
省電力運用実験基盤の整備においては、調査研究機関の前半に東京大学および東京科学大学(旧東京工業大学)の HPC 基盤に精度および時間分解能の高い電力計(EEHPC Power Measurement Methodology Level 3 相当)を導入しており、イーサネット経由でのリアルタイムデータ収集基盤を整備した。電力計がシステムのごく一部ではなく全体をカバーし、リアルタイムにデータを収集・蓄積・提供できる基盤を整備した。これらの産業用計測機器においては SNMP プロトコルもしくは MODBUS プロトコルでのデータ供給が一般的であり、各プロトコルに対するアダプタプログラムを作成することで、他サイトや他種の計測機器(例: 電力計→気象センサ)へのデータ収集部の移植は



①設置外観



②接続の様子



③測定の様子

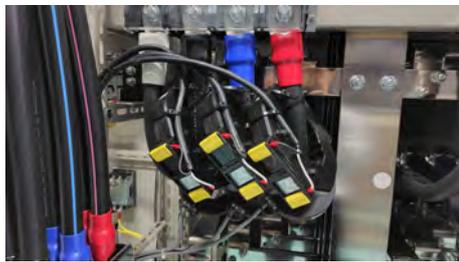


図 1 電力計測装置の設置状況（上段：東京大学、下段：東京科学大学）

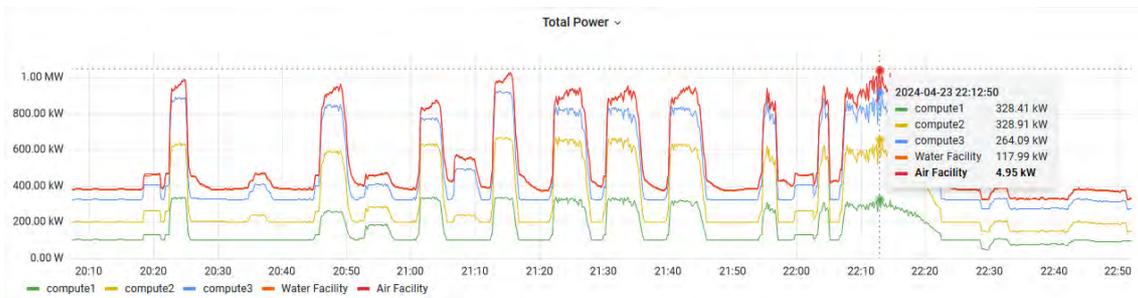


図 2 電力計測装置による計測データリアルタイム蓄積結果（東京科学大学）

相当に容易であることが確かめられている。システムの消費電力や気象センサーなどの環境情報を安定的に収集する基盤を活用し、施設運転の自動化やさらには人工知能技術を活用した最適化に活用できると期待される。



図 3 最先端共同 HPC 施設: Miyabi (東京大学)

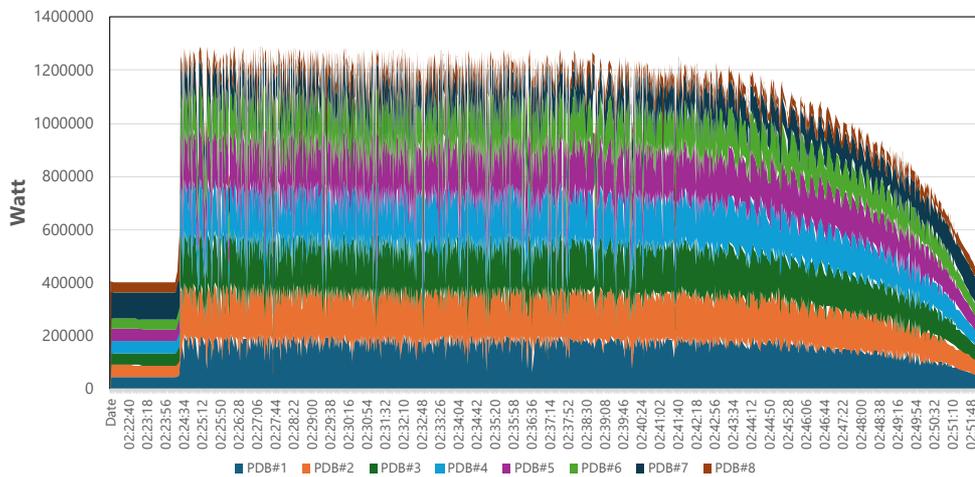


図 4 電力計測装置による Miyabi の Green500 計測 (東京大学)

また、国内外のデータセンター設備の状況調査を通じて、データセンター設備に求められる要件の精査を継続して実施した。冷却技術の詳細検討に関しては次節に詳細な検討結果を記載しているが、世界的に水冷、特に冷却塔を活用した比較的高温での冷却が主流となりつつあり、また計算機の高密度化・消費電力増大のトレンドが進行していることから、計算機本体よりも想定使用年数の長いデータセンター建屋においては、財政面での制約の範囲内で、冷却設備(冷却機器本体・配管)設置に十分なスペースや、計算機本体や冷却水を支えるために必要な耐荷重について、例えば液浸技術の導入など、2世代後の計算基盤における設備要求要件が厳しくなった場合にも受容可能となるよう、余裕を持った設計とすることが求められる。

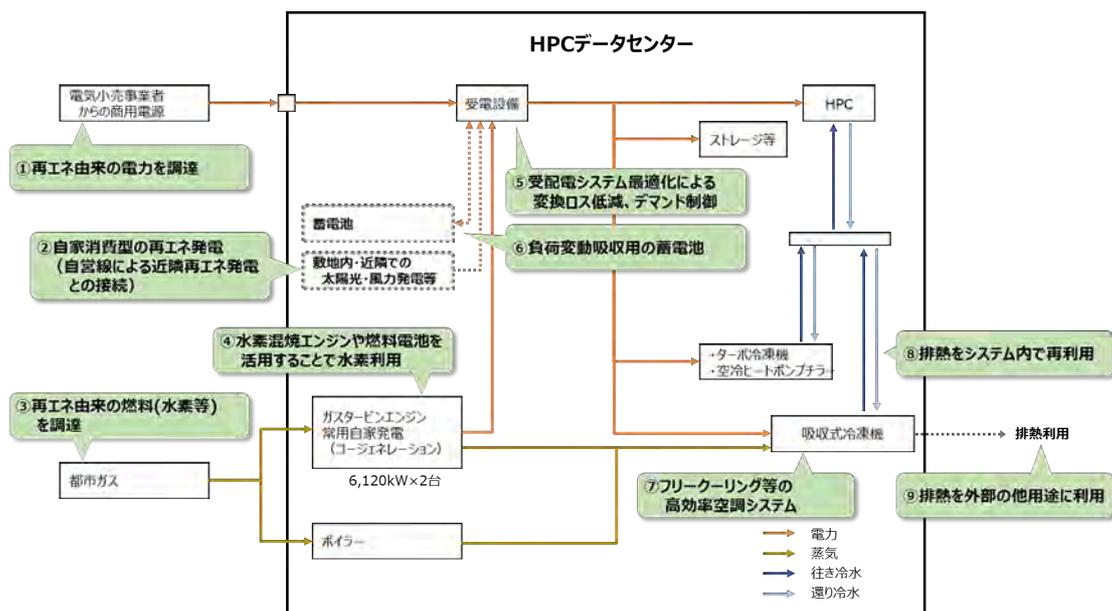
2.3 カーボンニュートラル

本節ではこれまでに実施してきた HPC データセンターにおけるカーボンニュートラル化の調査に関する報告をまとめる。政府は 2050 年のカーボンニュートラル（温室効果ガス実質排出ゼロ）達成を目標として掲げ、2030 年までに 2013 年度比で 46%の排出削減（50%も視野）を目指している。この目標達成に向け、社会基盤であるデータセンターやスーパーコンピュータ（HPC）の分野でも脱炭素化が求められている。特に HPC データセンターは大規模な電力消費を伴うため、再生可能エネルギーの活用や省エネルギー技術の導入等によるカーボンニュートラル（CN）化が重要な課題である。

今年度は昨年度までに整理した CN 化技術の調査項目 9 項目について動向の更新および検討の深化を行い、技術的な適用可能性や導入効果を詳細に評価した。また、各種 CN 化技術導入に伴う設備投資額（CAPEX）および運用コスト（OPEX）の特徴を分析し、従来型データセンター（ベースモデル）との費用比較を実施した。さらに、得られた知見を踏まえ、次世代計算基盤を設置するデータセンターの構成案を検討した。

2.3.1 カーボンニュートラルに関する技術動向

ここでは HPC データセンターの CN 化に関連する主要技術の動向について整理する。昨年度の調査で特定された下図の 9 項目の技術・方策について、国内外の最新動向や導入可能性を概観する。これら 9 項目のうち、1～7 はエネルギーの調達方法・利用方法に関するものであり、8 と 9 は HPC から排出される未利用熱の活用に関する項目である。項目ごとに、技術の現状と課題、HPC データセンターへの適用可能性について述べる。



1) 再生可能エネルギー由来の電力調達

データセンターの電力を再生可能エネルギーでまかなう代表的な方法として、電力会社等から再生可能エネルギー(再エネ)電力を購入する手法がある。具体的には、小売電気事業者が提供する再生可能エネルギーメニューの利用や、非化石証書(トラッキング付証書等)を購入して既存の電力をカーボンニュートラル電力とみなす方法、発電事業者との間で電力直接取引契約(PPA)を締結する方法(オフサイト PPA)等が挙げられる。HPC データセンターのように大規模な需要家でも、これらの手段を組み合わせることで実質的に再エネ 100%の電力調達が可能となる。

現状、日本国内では非化石証書の市場供給量が需要を上回っており、比較的安価に電力を購入時に再エネ属性を付与できる状況にある。例えば電力市場からの直接調達に近い形で、新電力による再エネ電力供給では 2024 年調査時点の価格で 1kWh あたり 0.7~1.4 円程度の追加コストで済む一方、一般電気事業者の再エネメニューでは約 2 円/kWh の上乘せとなる。非化石証書による付与コストは現在の電気料金の数%~10%程度に留まっている。したがって、現時点では最も経済的な CN 電力調達手段は、通常の電力に非化石証書を付与する方法といえる。しかしながら将来的な証書価格・再エネ電力価格の動向は不透明であり、2030 年頃には再エネ電力需要の増大から価格高騰の可能性も指摘されている。一方で、市場を通さず発電事業者と直接契約するオフサイト PPA は長期固定価格で安定調達が可能だが、送配電網の制約や契約交渉の障壁がある。いずれの方法でも、再エネ電力調達は比較的導入しやすい CN 化施策だが、電力調達コストが現在より上振れするリスクや地域の電力系統事情を考慮する必要がある。

2) 自家消費型再生可能エネルギー発電

本方式は、データセンターが自ら再生可能エネルギー源を設置・運用し、敷地内または近隣で発電した電力を直接利用する方式である。敷地内の建物屋上や遊休地に太陽光パネルを設置するオンサイト発電や、データセンター近傍に風力・太陽光発電所を建設して自営線で接続するケース(ローカル PPA)などが含まれる。これらは送配電網を介さずに電力を調達するため、利用料等の負担を抑えつつ再エネ電力の直接利用が可能となる。

自家消費型発電の利点は、長期的に見て安価で安定したエネルギー源を確保できる点である。太陽光や風力による自家発電は、一度設備を導入すれば燃料費ゼロで発電でき、化石燃料価格変動の影響を受けずに運用できる。特に将来的な電力市場価格の不確実性に対するリスクヘッジ手段となり得る。一方で、再エネ資源の出力は天候等に左右されるため、電力供給が不安定になる点が課題である。HPC の安定稼働には瞬時の電力確保

が欠かせないため、大規模な蓄電池（後述）やバックアップ電源との組み合わせが前提となる。また、都市部の既存施設では大面積の太陽光パネル設置は難しく、導入効果は限定的になりがちである。一方、地方に新設する場合には周辺に大規模発電所用地を確保するなど計画的な取り組みが可能となる。

3) カーボンニュートラル燃料の調達

本方式は、再生可能エネルギー由来の燃料を導入し、データセンターのエネルギー源として活用するアプローチである。燃料の例としては、水素、バイオ燃料（バイオマス由来の液体燃料やバイオガス）、カーボンニュートラル LNG（CO₂ 排出相当量をオフセットした天然ガス）などが挙げられる。HPC データセンターにおいては非常用発電機やコージェネレーション設備の燃料をカーボンフリーなものに置き換える形で利用可能である。例えば非常用ディーゼル発電機の燃料としてバイオディーゼル燃料（FAME 等）を使用する試みや、ガスコージェネ設備にカーボンニュートラル都市ガスを適用するケースなどが国内外で現れ始めている。

水素については、再生可能エネルギー由来のグリーン水素が将来のエネルギーキャリアとして期待され、政府の水素基本戦略でも 2030 年以降の本格普及が目標とされている。しかし現状では供給インフラが限定的で、安価に大量の水素を調達することは難しい。仮に外部から水素を調達する場合、パイプライン敷設や圧縮・液化水素の輸送が必要となり、実現には地域の水素サプライチェーン整備が前提となる。一方、従来型エネルギーインフラを活用できるカーボンニュートラル都市ガス（メタン）は比較的導入しやすく、既にガス事業者がクレジットを付与した CN ガスを販売している例もある。しかしながら、いずれの燃料も化石燃料に比べ現時点で割高であり、燃料コスト増が課題となる（後述の導入コスト分析参照）。総じて、カーボンニュートラル燃料の普及はまだ初期段階にあり、HPC 用途への本格導入は燃料の安定調達性と価格競争力の確保が前提となる。

4) 水素利用（混焼エンジン・燃料電池発電）

前項で触れた水素等のカーボンフリー燃料を実際にエネルギー源として活用するため、データセンター内に発電装置を導入するアプローチである。具体的には、ガスエンジンコージェネレーションやガスタービン発電設備を水素・バイオ燃料対応型にする、あるいは水素燃料電池システムを導入するケースが想定される。これらの設備を用いることで、データセンター自らがクリーンな電力（および熱）を生み出し、外部から購入する電力量を削減できる。とりわけコージェネレーションは発電時の廃熱を冷暖房等に活用できるため、総合エネルギー効率の向上が期待できる。

水素対応エンジン・タービン技術は近年各社で開発が進み、一部実証段階に入っている。日本国内でも川崎市や福島県などで水素発電実証プラントが運転中であり、海外ではデータセンターの非常用発電機を水素燃料電池で置き換える試験が報告されている。一方で、HPC データセンターのように常時高負荷がかかる施設で水素発電を主電源として成立させるには、燃料供給コストおよび信頼性の課題が大きい。現状の水素価格（工業用途向け水素で約 100 円/Nm³ 程度）は発電用燃料としては極めて高価であり、運用コスト増につながる。また、水素燃焼エンジンは NO_x 排出や燃焼安定性の技術課題が残る。燃料電池についても初期費用が非常に高額で大型化に課題がある。しかし、これら技術の導入は CO₂ 排出削減効果が大きく、また発電時に発生する熱を有効利用できればエネルギー利用効率が飛躍的に高まる利点がある。今後、水素供給インフラの整備状況やエネルギー価格次第では、HPC データセンター施設での採用も検討対象となり得る。

5) 受配電システム最適化とデマンド制御

本項目は、データセンター内の電力受配電方式を工夫し、エネルギー変換ロスの低減や電力需要の平準化を図る技術である。具体的には、高圧直流給電や高効率な無停電電源装置（UPS）の採用による変換ロス削減、高度なエネルギー管理システムによる負荷平準化（ピークシフト）などが挙げられる。また、電力需要側での制御として、電力価格や再エネ供給量の変動に応じて HPC におけるジョブ実行タイミングやリソース割り当てを調整するデマンドレスポンス的な運用も考えられる。

受配電システム面では、データセンター用途向けに高圧直流給電(380V DC 等)の採用事例が海外で増えつつあり、変換段数を減らすことで数%の電力損失削減が可能と報告されている。また国内では一部の HPC センターで高効率電源や配電監視の高度化が検討されている。需要制御の面では、電力市場価格が高騰した時間帯に計算資源を抑制し、安価な時間帯に集中的にジョブを実行することも技術的には可能である。実際、2022 年の日本卸電力市場(JEPX)においては昼夜間で最大 20 円/kWh 程度の価格格差が生じており、この差を蓄電池や柔軟な負荷制御で吸収できれば大きなコスト削減余地となる。ただし HPC の運用においては計算需要を満たすことが求められるため、全てを価格優先で制御するのは困難である。将来的には、ジョブスケジューラとエネルギー管理システムを連携させる研究開発が進めば、一定のサービス品質を維持しつつ電力需要を調整する高度な運用も可能になると考えられる。

6) 蓄電池の導入

本項目は、大容量蓄電池をデータセンター設備に組み込み、エネルギー利用の効率化と非常時バックアップを両立する施策である。蓄電池は余剰再エネ電力や夜間の安価な

電力を蓄えてピーク時に放電することで、購入電力のピークカットや電力コスト平準化に寄与する。また停電時には非常用電源として機能し、従来の非常用発電機と組み合わせることで無瞬断での電力供給を可能にする利点もある。蓄電池の種類として、エネルギー密度やコストに優れたリチウムイオン電池、高い安全性と長寿命が特徴のレドックスフロー電池、大容量で放電時間の長いナトリウム硫黄電池（NaS 電池）などがある。

蓄電池技術は近年急速に進歩しており、リチウムイオン電池のコスト低下に伴いデータセンターへの導入も増えている。特に欧米のハイパースケールデータセンターでは、大規模 UPS システムに蓄電池を組み込み、電力網との相互作用（デマンドレスポンスや予備電力の提供）を行う動きもある。HPC データセンターでも、再エネ自家消費を高めるため昼間の太陽光発電余剰分を蓄電して夜間に利用する、電力料金の高い時間帯の需要を蓄電池で賄う、といった運用が考えられる。一方、課題として蓄電池設備そのものの初期投資が非常に高額であること、設置スペースの確保、および電池寿命（数千サイクル程度）に伴う定期的な更新コストが挙げられる。またリチウムイオン電池は高出力特性に優れるが発火リスクへの安全対策が必要であり、NaS 電池は大型・高温動作のため慎重な据付が求められるなど、電池種別ごとの技術的留意点もある。総じて、蓄電池導入は CO2 削減に加えエネルギーコスト低減やレジリエンス強化につながる有力手段だが、その恩恵を最大化するには電力市場動向と電池劣化を織り込んだ運用最適化が必要である。

7) 高効率な冷却システムの導入

本項目は、データセンターの冷却方式を改善し、消費電力の大幅な削減と排熱温度の上昇（後述の排熱利用性向上）を図る取り組みである。代表的な高効率化策としては、外気や冷却塔を利用したフリークーリング、熱交換器の高度化、液浸冷却の導入、自動制御化による最適運転などが挙げられる。HPC システムでは近年、水冷（冷却水）の採用が一般化しつつあり、従来の空冷に比べはるかに高い冷却効率と排熱回収効率を実現している。

多くの HPC システムで採用される直接液冷（コールドプレート方式）は、高発熱なプロセッサやメモリでも冷却水で効率よく熱を奪えるため、施設全体の PUE 値改善に大きく寄与する。さらに冷却水温度を高め（温水冷却）、チラー（冷凍機）の稼働を極力減らす動きがある。東工大の TSUBAME3.0 や産総研の ABCI では、32℃程度の冷却水温でも安定稼働する実証がなされており、外気温がそれ以下の期間は冷却塔のみで熱を捨てるフリークーリング運転が可能となっている。気候条件の良い地域では年間を通じてチラー非稼働とすることも視野に入り、この場合にはチラーの導入コストも削減できる。

一方、液浸冷却は近年注目を集める技術である。ファンが不要となりシステム全体の PUE を 1.05 前後まで低減し得ると期待されるが、現時点では実用上の課題も多い。例えば冷媒の種類によって冷却性能や取扱い難易度・環境影響が異なり、適切な冷媒選定が簡単ではない。また液体による機器腐食や膨潤、配管からの揮発損失や安全性（難燃性や有害性）の問題もあり、大規模な実運用 HPC への適用には慎重な検討が必要である。現状では一部の GPU クラスタ等で試験導入例があるものの、汎用 HPC システムでは、通常の温水冷却に一部空冷機器はリアドアクーラーの併用、という構成が現実的と評価されている。総じて、冷却システム高効率化は技術的選択肢が豊富で最も効果も大きい分野であり、HPC データセンターにおける CN 化の鍵を握る領域である。

8) データセンター内での排熱再利用

本項目は、データセンターで発生した排熱を施設内の他用途に再利用する試みである。計算機や冷却システムから排出される温水・温風を回収し、再び冷却プロセスに役立てたり建物内の暖房・給湯に転用したりするケースが想定される。具体的には、計算機からの温水をオフィススペースの暖房用に温水コイルへ送る方法、シャワーや給湯設備の湯沸かしに利用する方法などがある。

排熱の再利用はエネルギーの二次利用という点で理想的だが、実現にはいくつかのハードルがある。第一に、計算機からの排熱の温度が低い点である。空冷式のシステムでは排気温度が数十℃程度に留まり、そのままでは利用用途が限定的である。水冷の場合でも温水はせいぜい 30～45℃程度であり、暖房や給湯に用いるには温度不足となる。このためヒートポンプ等を追加して 50～70℃程度まで昇温する必要があるが、そのための追加電力や設備コストが新たに必要になる。第二に、適切な利用先が必要という点がある。排熱を有効利用するには需要側があって初めて成り立つため、データセンターと同じ建物内または近接地に熱需要（暖房負荷や温水需要）が存在することが望ましい。例えば HPC センター併設のオフィスや研究棟の暖房・給湯に充当すれば比較的容易に再利用が可能である。しかし専用の HPC センターのみでは十分な需要が見込めず、下手をすると回収した熱を結局廃棄する結果にもなりかねない。第三に、経済性と役割分担の問題がある。追加設備を導入してまで排熱を利用する場合、その費用負担を誰がするのか、投資に見合うだけのエネルギー費削減効果があるのかといった検討が必要である。これらの課題から、現在のところ排熱再利用は一部の好条件下でのみ成立している。例えば寒冷地にあるデータセンターでは排熱で融雪やビニールハウス加温を行う例があり、国内でも北海道でデータセンターの熱を農業ハウスに提供する実証が行われている。しかし一般的な温暖な気候条件では有効利用が難しく、無理に行えばヒートポンプ駆動

のためかえって電力を消費する状況にもなりうる。したがって、排熱再利用は HPC データセンター単体で完結する取り組みには限界があり、後述のように地域との連携や周辺需要の開拓と合わせて検討すべき施策である。

9) 外部への排熱供給

本項目は、データセンターの排熱を施設外の需要家へ供給し、地域全体のエネルギー利用効率向上に役立てる取り組みである。具体的には、病院・ホテル・商業施設等の他産業に温水を供給する、あるいは地域熱供給プラントへ熱源として提供する、といった形態が考えられる。欧州では Facebook のデータセンターが近隣住宅の地域暖房にサーバの排熱を活用する例や、データセンターが積極的に都市インフラと連携して地域暖房の主要熱源の一部となるなど、先行事例が増えつつある。日本国内でも、自治体によるスマートシティ計画の中でデータセンター排熱の地域利用が検討され始めている。

外部への熱供給の最大の課題は、適切な受け手とインフラの構築である。データセンター近隣に相応の熱需要を持つ施設が存在しなければ成立せず、仮にあっても両者を繋ぐ配管設備が必要となる。たとえば都市中心部で既に地域冷暖房ネットワークがあるエリアでは、そこに熱を供出する形で比較的容易に連携できる。しかし郊外型のデータセンターでは、一から隣接産業を誘致するか、長距離配管を敷設してまで供給する採算が合うか検討せねばならない。さらに制度面の問題として、データセンターが地域に熱供給事業を行う際の法規制や契約スキームが未整備である点が指摘されている。現行制度では、民間事業者が熱供給を行うには熱供給事業法に基づく手続き等が必要となる場合があり、データセンターが副業的に熱販売するにはハードルが高いとの指摘もある。こうした課題から、海外では行政が間に入ってマッチングやインフラ整備を支援する例がみられる。カリフォルニア州では新設データセンターに対し熱回収の検討を義務付ける動きもあり、政府が規制面から促進するケースも出てきている。日本においても、周辺自治体の積極的な協力や制度整備がなければ、大規模な外部排熱利用は実現が難しいと考えられる。もっとも、実現すれば地域全体の炭素排出削減に資するカーボンネガティブの取り組みともなり得るため、大規模なデータセンターでは早い段階から検討を開始し、関係者間の調整を進める価値は大きい。

2.3.2 導入コストの分析

1) 各技術導入時の CAPEX および OPEX

ここでは、前節で整理した CN 化技術を HPC データセンターに導入する場合の資本的支出（初期導入コスト：CAPEX）と運用コスト（OPEX）の特徴について比較分析する。

以下の表に各技術・方策の導入コストと運用コストへの影響をまとめる。

技術・方策	初期導入コスト (CAPEX)	運用コスト (OPEX) への影響
再エネ電力の調達 (非化石証書購入等)	ごく小 (既存電力契約の調整のみ)	電力料金に証書コストが上乗せ (数%程度の増加)
オフサイト PPA 契約	小 (契約手続き・交渉に伴う費用)	一般電力よりやや高い単価での長期電力購入 (コストは中程度増)
オンサイト再エネ 発電 (太陽光パネル 等)	大 (発電設備の設置、用地確保費)	電力購入量の削減による電気代低減効果 (余剰電力があれば売電収入も)
大型蓄電池システム	極めて大 (大容量電池設備の導入、設置スペース確保)	電力ピークカット・夜間充電等で電力購入 費を削減可能 (但し電池寿命に伴う更新費用発生)
CN 燃料の利用 (水 素・CN ガス調達)	中～大 (燃料供給設備の導入、契約コスト)	燃料コスト増: CN 燃料は現状高価で運用費 が増加
水素発電設備 (H2 エンジン・燃料電池)	極めて大 (専用発電設備導入、改修工事)	燃料コスト増: 水素は現状高価で運用費増 (排熱利用で一部エネルギー回収可)
受配電最適化・需 要制御	中 (制御機器導入、エネルギー管理システム開発)	最大需要電力の低減・損失削減により電力 費用が減少 (需要家メリット大)
高効率冷却 (フリ ークーリング等)	中 (冷却塔追加や配管改修など の設備投資)	冷却電力の大幅削減により電力費用が減少 (PUE 改善効果大)
排熱再利用 (内部/ 外部)	中～大 (熱交換器・配管敷設等 の追加工事)	他エネルギーの使用量削減効果 (施設内暖 房燃料等の節約) (外部供給する場合、収入 可能性もあるが不確実)

上表より、CN 化技術の導入にはそれぞれ特徴的なコスト構造があることが分かる。まず、再生可能エネルギー電力の調達や PPA 契約は、新規設備を要しないため初期費用がほぼ不要である反面、運用時に調達電力単価が割高となる (証書購入費やグリーン電力プレミアムの負担)。オンサイト再エネ発電や大型蓄電池、水素利用設備などエネルギー

ギー供給インフラ系の施策は、総じて巨額の初期投資を伴う代わりに、長期的に購入エネルギー量や電力料金の低減効果が期待できる。中でも水素発電設備は設備・燃料とも現状高コストであり、実質的に経済メリットを得るには相当のスケールメリットや他用途への熱利用など複合的な工夫が必要となる。一方、需要側の効率化策である受配電最適化や高効率冷却は、中程度の追加投資で比較的確実に運用コストを減らすことができ、特にフリークーリングの導入などは投資回収期間が数年程度と見込まれる。排熱再利用は、省エネ効果そのものは間接的だが、暖房用熱源の代替等による他燃料費の節約や環境価値の向上といったメリットをもたらす。ただし収支面では追加設備投資に見合うだけのリターンを得にくく、経済インセンティブが弱い点が課題である。

2) ベースモデルとの費用比較

CN 化技術導入時の費用を、何も対策を講じないベースモデルの場合と比較すると、運用コスト(OPEX)は一定程度低減する一方で、導入コスト(CAPEX)は従来比で格段に大きくなる傾向がある。もっとも、技術によっては導入後のコスト削減効果で一定期間内に初期投資を回収できる可能性も報告されている。実際、本調査で実施した試算では一部の施策で「数年で損益分岐点を迎える(投資回収が可能)」との結果が得られた。例えば、大容量蓄電池を用いたピークシフト運転は、電力価格の差額が十分大きければ数～十年程度で電池投資を回収できるシナリオも描ける。また、高効率冷却は気候条件が合致すれば、大幅な運用コスト低減のみならずチラーの導入費用の減少も可能であり、ほぼ水冷で今後も運用が想定される HPC データセンターにおいては最も優先して検討する項目である。他方、水素エネルギーの利用については現状コスト回収まで相当な長期を要する。総じて、各技術の費用対効果は前提条件(エネルギー価格、市場制度、設備利用率等)によって大きく変動するため、個別ケースごとの精緻な検討が必要である。

加えて、全体のスケジュール面でも留意が必要である。再エネ電力の大規模導入や排熱利用の実現には地理的条件の考慮や関係者との協議に時間を要し、具体的な設置場所の検討開始から運用開始まで3年以上の準備期間が必要と見積もられる。そのため、次世代計算基盤にCN化オプションを組み込む場合、できるだけ早期の段階から調整・計画立案を進める必要がある。以上のように、CN化技術の導入は費用面・時間面で大きな挑戦を伴うが、長期的に見ればエネルギー価格高騰リスクの低減や安定運用確保といったメリットも得られる。

これらを踏まえ、次節では具体的なスペックを想定した次世代計算基盤について、導入コスト・運用コストの両面である程度現実感のあるHPCデータセンターの構成シナリオを示す。

2.3.3 次世代計算基盤にむけた HPC データセンターの構成

本節では、40MW 級 HPC データセンターの CN 化に向け、現行のフラッグシップマシンである「富岳」の設置されている神戸エリアと、寒冷な気象条件を想定した北海道エリアを想定した 2 つのシナリオについて検討する。それぞれの構成技術と特徴を示した上で、CO₂排出削減効果、設備投資 (CAPEX)・運用コスト (OPEX)、実現性、所要期間、課題といった観点から定性的評価を行う。

1) 想定シナリオの概要

シナリオ 1 (神戸エリア)： データセンターを神戸地域に据え置きつつ、省エネ化と再生可能エネルギー電力の活用によって CN を目指すシナリオである。「富岳」の施設構成をベースモデルとし、設備投資面で現実的な更新を検討した。データセンターが消費する電力の 100%を太陽光・風力等の再生可能エネルギーで調達し、非化石証書等を活用して実質的な CO₂排出ゼロ電力とする。加えて、冷却効率の向上など省エネ施策を徹底する。具体的には、次世代計算基盤の水冷比率を「富岳」以上にすることを想定し、温水冷却を導入することで冷却に必要な電力を大幅に低減する（設計上の部分 PUE は 1.1 以下と国内トップクラスの省エネ性能）。これにより全体の消費電力量そのものを削減し、再エネ電力の必要量も抑制する。施設内には屋上や敷地を利用した蓄電池を設置し非常用電源として活用する。通常時のバックアップ電源には現行の「富岳」とは異なりディーゼル発電機を用いる。余剰な排熱については、熱源温度が比較的低いため大規模な外部利用は難しいものの、冬季における施設暖房等への一部利用を想定する。

シナリオ 2 (北海道エリア)： データセンターそのものを北海道など再エネ資源に恵まれた寒冷地に新設・移転し、地域の再生可能エネルギー電源を集中的に活用するシナリオである。具体的には、風力発電（陸上・洋上）や大規模太陽光発電、水力発電等のポテンシャルが高い北海道内にデータセンターを新設し、電力を当該地域内で発電された再エネで 100%賄うことを目指す。実際、苫小牧市に建設予定のデータセンターでは北海道内の再エネ電力のみで需要を満たす計画としており、地産地消型のグリーンデータセンターのモデルケースになるとされている。寒冷な気候はデータセンターに有利に働き、外気冷却の積極活用や海水・河川水を利用した冷却によって、チラー等の冷媒機器の稼働を大幅に低減できる。HPC の発熱密度は高いものの、低外気温を活かすことで PUE を 1.1 未満、場合によっては 1.05 程度まで改善し得る。計算資源の稼働によって生じる排熱は、周辺地域の暖房需要などに利用することで有効活用が可能である。特に北海道のような寒冷地では、データセンター廃熱は低温であっても農業用ハウスの加温や融

雪設備に利用するなど地域エネルギーシステムと連携した活用も期待される。電力供給の信頼性確保のため、商用電力システムとも接続するが、北海道電力ネットワークの電源構成は本州より CO₂排出係数が低く、系統電力利用時の排出も抑えられる。以上により運用段階の CO₂排出は実質ゼロにできる可能性が高い。

2) シナリオ比較

2つのシナリオを主要評価項目で比較した結果を下表に示す。各シナリオの特徴が一目で分かるよう、CO₂削減効果、初期投資規模、運用コスト影響、実現性について相対的に整理した。

評価項目	シナリオ 1 神戸エリア	シナリオ 2: 北海道エリア
CO ₂ 削減効果	高い（再エネ電力で運用：削減率：90%以上）	非常に高い（地域再エネで 100%電力供給・排熱再利用でカーボンネガティブも）
初期投資規模	中程度（チラー費用要・温水冷却）	中程度（チラー費用減・排熱再利用インフラ整備増）
運用コスト	中程度（再エネ調達・証書による電気料金増）	小-中程度（電力コスト低減も遠隔運用コスト増を想定）
実現性	高い（既存技術で実施可能）	高い（既存技術で実現可能）
PUE	1.1	1.05-1.10

各シナリオとも相対的な評価はほぼ同等であり、現時点で最適解が一つに定まるものではない。PUE に関しては両者とも 1.1 程度であり「富岳」の 1.3-1.4 からの大幅な低下が可能とみている。これら PUE だけ見ても現行世代のスパコンから大幅な省エネが実現できることがわかる。シナリオ 1 は温水冷却で、シナリオ 2 はフリークーリングを想定しており、温水冷却を採用する場合は計算機の徹底した水冷化が必要であるが、これらは次世代計算基盤ではすでに実証され実現可能な技術と想定している。

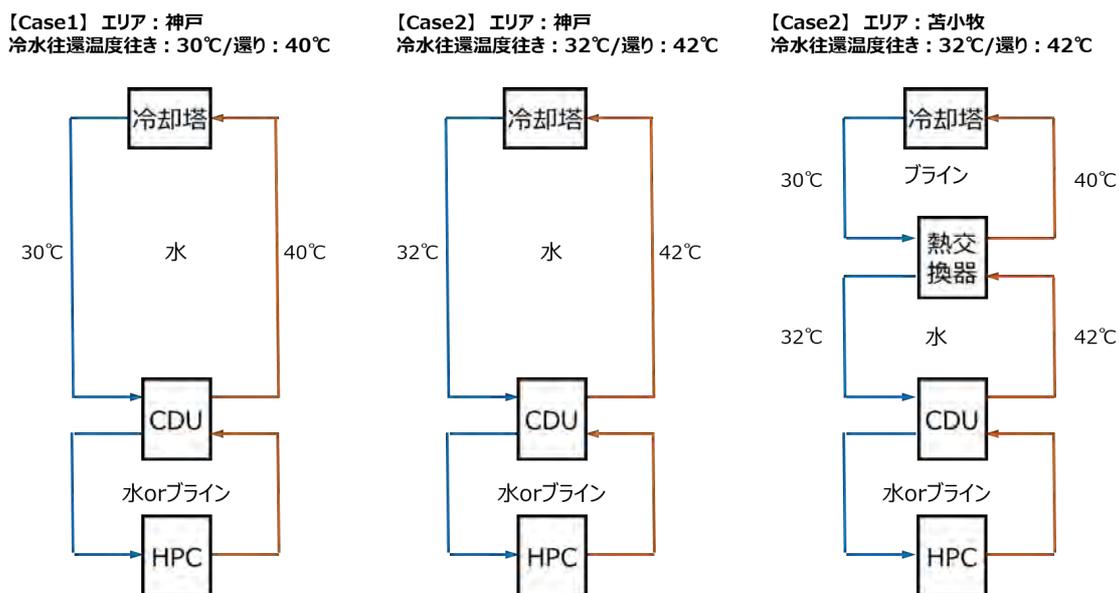
シナリオ 1 は技術的ハードルが低く早期に取り組める反面、長期的には非化石証書頼みの電力調達に課題が残る。シナリオ 2 は排熱再利用によりカーボンネガティブ化が見込まれ究極的な脱炭素データセンターを実現し得るが、立地に関して慎重な計画策定が必要である。今後の政策動向や技術革新、電力需給の情勢を注視しつつ、これらシナリオを組み合わせた段階的な CN 化戦略も検討すべきであろう。

2.3.4 温水冷却による PUE および OPEX の試算

前節において、冷却方式の変更が PUE の低減に著しい効果をもたらすことを述べた。ここでは、次世代計算基盤における冷却方式として温水冷却を採用したときの、PUE や運用コストを具体的に試算した結果を報告する。ケーススタディとして、以下の3つの冷却方式案(ケース A~C)を設定し、それぞれについて消費電力や運用コスト(電気代・水道代)を評価することを目的とする。比較検討の結果は、次世代データセンター設計に向けた最適な冷却システム構成の検討指針を得るための基礎資料となる。

1) ケース設定と前提条件

本比較検討では、データセンターの立地と冷却システムの温度について3つのケースを設定した。各ケースの概要を下図に示す。前提条件として冷却塔にて 30℃冷水を生成可能な湿球温度 27℃条件で、冷却塔のみで冷却システムを構築する場合を想定した。



Case1 エリア：神戸、送水往還温度行き：30℃/還り：40℃

Case2 エリア：神戸、送水往還温度行き：32℃/還り：42℃

Case3 エリア：苫小牧、送水往還温度行き：32℃/還り：42℃

神戸の気象条件においては、湿球温度 27℃を上回る発生日数(最大)は 16 日、最大継続時間は 17 時間であるが、計算機は常に最大の冷却能力を必要としているわけではないため、上回る時間帯が発生することは許容されるとして、追加の熱源設備(チラー)

は必要ないものとして試算する。Case 3においては、冬季の凍結対策として、ブラインを使用する条件とした。

2) それぞれのケースによる比較と考察

各ケースにおける光熱費（電気料金、水道料金、下水料金）を下表に示す。2023年度の気象庁データをもとに、計算機の負荷容量は8MW（80%）×5室の条件としている。なお、各ケースにおける変化を把握するため、同一の条件（ポンプや配管など）による消費電力量は含めていない。

【Case1】 エリア：神戸 冷水往還温度往き：30℃/還り：40℃				【Case2】 エリア：神戸 冷水往還温度往き：32℃/還り：42℃				【Case2】 エリア：苫小牧 冷水往還温度往き：32℃/還り：42℃			
年間消費電力量（1室）	kWh	125,509	116%	年間消費電力量（1室）	kWh	108,384	100%	年間消費電力量（1室）	kWh	87,930	81%
年間給水量（1室）	m ³	20,738	109%	年間給水量（1室）	m ³	18,967	100%	年間給水量（1室）	m ³	12,045	64%
年間排水量（1室）	m ³	6,913	109%	年間排水量（1室）	m ³	6,322	100%	年間排水量（1室）	m ³	4,015	64%
年間消費電力量（5室）	kWh	627,545	116%	年間消費電力量（5室）	kWh	541,920	100%	年間消費電力量（5室）	kWh	439,651	81%
年間給水量（5室）	m ³	103,689	109%	年間給水量（5室）	m ³	94,833	100%	年間給水量（5室）	m ³	60,225	64%
年間排水量（5室）	m ³	34,563	109%	年間排水量（5室）	m ³	31,611	100%	年間排水量（5室）	m ³	20,075	64%
電気料金	円	9,413,180	116%	電気料金	円	8,128,805	100%	電気料金	円	8,793,027	108%
水道料金	円	2,996,610	109%	水道料金	円	2,740,680	100%	水道料金	円	1,204,509	44%
下水料金	円	9,781,324	109%	下水料金	円	8,945,934	100%	下水料金	円	5,940,237	66%
合計	円	22,191,114	112%	合計	円	19,815,420	100%	合計	円	15,937,773	80%
pPUE		1.0072		pPUE		1.0062		pPUE		1.0050	

試算条件は以下とした。

- 神戸の電気料金は15円/kWh、水道（工業用水）料金は28.9円/m³、下水料金は283円/m³にて試算
- 苫小牧の電気料金は20円/kWh、水道料金は20円/m³、下水料金は295.9円/m³にて試算
- 電気料金・水道料金・下水料金に基本料金は見込まず
- ポンプの循環動力や凍結防止（ヒーター等）は見込まず

Case1を参照すると、Case2と比較して年間消費電力量が約16%増加、給排水量が約9%増加するという試算結果が得られた。これは、冷水往還温度を高く設定することで冷却効率が向上し、省エネルギーおよび節水効果が得られることを示している。冷却塔のみで冷却をまかなう構成であり、熱源機器を必要としないことからもともと非常に高効率であるが、pPUEについても約0.001向上するという結果となった。

また、Case3を参照すると、Case2と比較して年間消費電力量が約19%削減、給排水量が約36%削減されるという試算結果が得られた。冷涼な気候に立地することにより、冷却負荷が軽減され、省エネルギーおよび節水効果が得られることが確認された。pPUEにおいても、Case2と比較して約0.001の改善が認められた。

一方で、電気料金については、関西エリアと比較して北海道エリアでは電力単価が高いため、光熱費ベースでは約8%の増加が試算された。上下水道料金については、各地域における工業用水単価および下水処理単価に依存するため、地域条件により大きく異なることがわかった。

2.3.5 カーボンニュートラル化に関するまとめ

本節ではこれまでの技術的な調査結果を踏まえ、HPC データセンターのカーボンニュートラル（CN）化に関連する制度的・財政的・運用的な論点について整理する。これらの視点は、CN化に向けた諸施策を支える基盤として重要であり、今後の施策検討の参考となると考える。

1) 制度面での論点

- 電力調達に関する制度整備：HPCI等の研究インフラにおいて、再生可能エネルギー電力を柔軟かつ安価に調達できる環境整備が課題となっている。電力市場からの直接購入や発電事業者とのPPA契約における制約、系統利用の地域的制限などが、再エネ導入の障壁として確認された。
- 排熱利活用に向けたルール形成：大学や研究機関による熱供給は法的な整理が十分でなく、地域との費用負担・責任分担の明確化、計画段階からの熱需要とのマッチングが重要である。制度的には地域熱供給法や都市計画手続きとの整合性の検討が必要である。
- CN化達成度の評価指標：再エネ比率や排熱再利用率を含めた評価指標の導入により、施設ごとの脱炭素進捗を客観的に評価し、運用改善につなげる枠組みの整備が望まれる。

2) 財政的側面での論点

- 初期投資への支援枠組み：CN化に伴う高額な初期導入コストを補完する制度的枠組みの必要性が再認識された。初期導入費は増えても運用費で回収できる施策もあり、単年度でなく長期的な全体最適化に向かうことが望まれる。
- 運用費への対応：CN化を目指せば自ずと再エネ電力や水素導入に伴いOPEXが増加する。財政的に考えると、CN化を目標としない方が有利であり、CN化を推進するためにはOPEXの増加を補うメリットが必要となる。

3) 運用体制・連携に関する論点

- 産学官連携の重要性：HPCセンター単独ではCN化の全てを担うことは難しく、電力事業者・自治体・大学/研究所本部との連携体制の構築が鍵となる。

- 知見集約とガイドライン化：個別事例に依存した属人的な CN 化対応ではなく、再エネ導入率、冷却指針、排熱連携の標準化に向けたどの HPC センターでも利用可能なガイドライン整備の必要性が挙げられる。

2.4 資源管理

2.4.1 資源管理の調査研究内容

「次世代計算基盤に係る調査研究」公募要領（以下、「公募要領」と略）の調査研究項目、ならびに次世代計算基盤検討部会中間取りまとめによる「ポスト「富岳」時代に目指すべき姿」（以下、「目指すべき姿」と略）を基に、機能、構成等を検討した。

「目指すべき姿」を図エラー！参照元が見つかりません。に示す。様々な計算資源や実験施設、およびデータ基盤との連携が可能な環境を、研究分野や研究目的毎に利用プラットフォーム上の仮想専用マシンに構築可能な計算環境と考えられる。

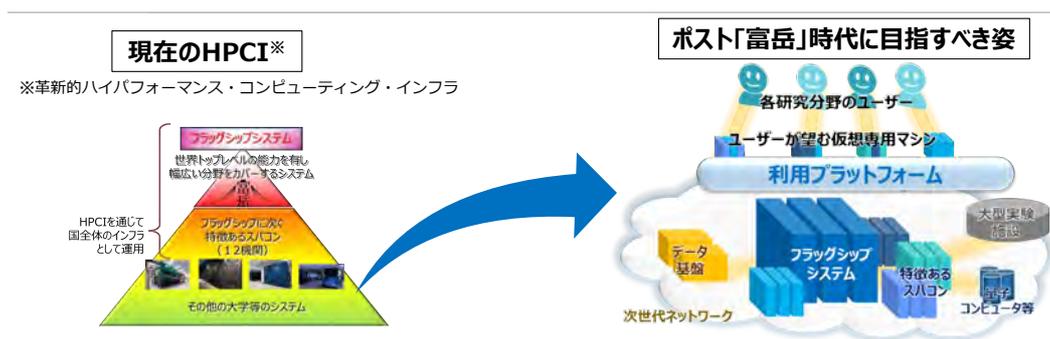


図 1 次世代計算基盤検討部会による「ポスト「富岳」時代に目指すべき姿」

公募要領が求める調査研究内容から資源管理に関連する課題を抽出した。後の説明の便宜上、各課題を採番する。

- 公募要領>2. 事業の概要>(1) 調査研究の内容より
 - 果たすべき役割
 - p1: 国内主要計算基盤、データ基盤、ネットワークを一体的に運用、持続的に機能
 - p2: 最先端のシミュレーション、AI 用学習データの生成、大量データ処理
 - p3: 実験系の研究者も利用しやすい計算環境
 - 求められる性能・機能
 - p4: ジョブと計算資源のアロケーション最適化
 - p5: 他のシステムとの連携・融合を可能とする機能拡張性

p6: ジョブを最適に実施するメタスケジューラ

- 公募要領>2.事業の概要>(2)実施体制>③公募により選定する機関>Ⅲ.運用技術調査研究チーム>研究内容より

p7: Society5.0 推進 (基盤間連携、デジタルツイン、緊急ワークロード等)

p8: SDGs の達成貢献 (カーボンニュートラル、持続的に機能等)

p9: 一体的運用に伴うセキュリティリスクへの対策

p10: 機密性の高いデータの安全な処理

p11: 仮想化と資源管理

p12: クラウド連携

p13: 高可用性

p14: リアルタイム処理

2.4.2 既存計算基盤の調査

欧米の代表的ないくつかの計算基盤を調査した。特に、計算資源間の連携、実験施設と計算資源との連携に注目した。

(1) 欧米の計算基盤の違い

欧州の EGI¹ と米国の Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS)² は、それぞれグリッドコンピューティング時代の代表的な計算基盤である DataGrid と TeraGrid をルーツに持ち、思想、構成、機能の多くを継承している。DataGrid は HTC 向け、TeraGrid は HPC 向けであったが、EGI や ACCESS では両者とも HPC、HTC、Cloud (VM)、コンテナをサポートしている。しかし、構成や機能は大きく異なっている。EGI は、計算基盤内に様々な機能を持つ一体型である。一方、ACCESS は計算基盤内には主に計算機能しか持たず、計算資源同士の連携機能も少ないが、有用な機能を持つサポートツールが提供されており、分散型と言える。このツールを利用して、Science Gateway³ と呼ばれる研究分野や研究目的毎に専用の利用者環境が開発されている。これは、「目指すべき姿」の利用プラットフォームに相当するとも考えられる。

¹ EGI: Advanced Computing for a Data-Driven Future. (オンライン) (引用日: 2025 年 1 月 28 日.) <https://www.egi.eu>.

² ACCESS. ACCESS. (オンライン) NSF. (引用日: 2025 年 1 月 28 日.) <https://access-ci.org>.

³ Science Gateways. (オンライン) NSF. (引用日: 2025 年 1 月 28 日.) <https://support.access-ci.org/tools/science-gateways>.

(2) 実験施設との連携

欧米とも実験施設と計算資源の連携機能を持つ計算基盤はない。ただし、実験施設を含む研究基盤と計算基盤とを連携させることが可能な総合的な研究基盤や、実験施設と計算資源とを連携する外部機構は存在する。

欧州の European Open Science Cloud (EOSC)⁴ は「ヨーロッパの科学のための、FAIR 原則に基づいてデータとサービスのウェブを構築する」ことを目的とした、様々な基盤やサービスを連携する総合的な研究基盤である。各種基盤間の連携機能を開発する EOSC Future プロジェクト⁵では、当初「目指すべき姿」と同様に実験施設や計算資源の資源レベルでの連携を目指していたようである。しかし、実現が困難であり基盤レベルでの連携に変更した。これを図 2 に示す。(a) 旧アーキテクチャ図の EOSC Exchange が「目指すべき姿」の様々な資源のクラウドのような部分に相当すると考えられる。新アーキテクチャは、様々な基盤の単なる寄せ集めのようにも見えるが、各基盤の同種の機能レイヤ間での相互運用性が検討されている⁶。新アーキテクチャは、2024 年 4 月よりサービスが開始されている。

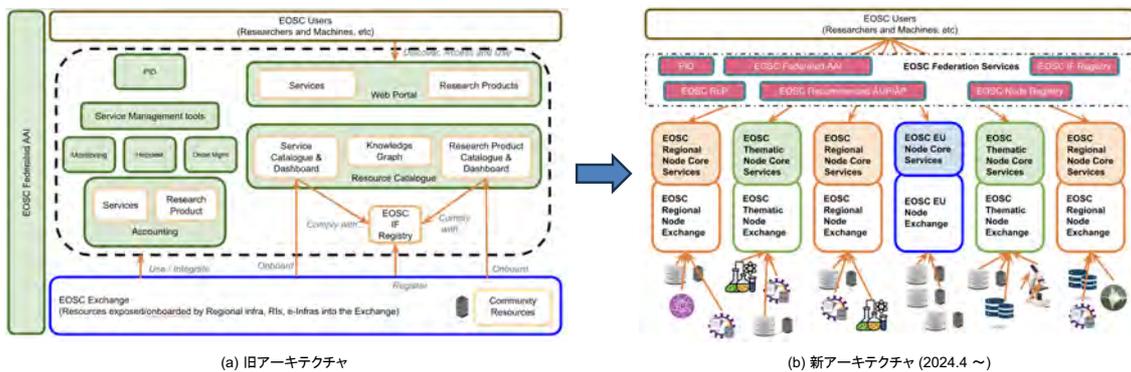


図 2 EOSC のアーキテクチャ

(DietrichMark. Core & Federation in EOSC Future. より)

米国では、実験施設と計算資源を連携するための外部機構が存在する。DOE はスパコンと実験施設を連携させる Integrated Research Infrastructure (IRI) フレームワーク

⁴ EOSC. (オンライン) (引用日: 2023 年 5 月 10 日.) <https://eosc.eu>, <https://eosc-portal.eu>.

⁵ EOSC Future. (オンライン) (引用日: 2025 年 5 月.) <https://eoscfuture.eu>.

⁶ Dietrich Mark. Core & Federation in EOSC Future. (オンライン) 2023 年 9 月 20 日. (引用日: 2025 年 3 月 23 日.) https://symposium23.eoscfuture.eu/wp-content/uploads/2023/09/Mark-Dietrich_EOSC-Future-the-EOSC-Platform-and-the-Federation.pdf.

⁷を提唱している(図 3)。その実現方法の1つである Globus(注: 過去の Globus Toolkit とは異なる)を使用した、核融合実験施設(DIII-D)や放射光施設(APS)と NERSC や ALCF の大規模スパコンとの連携が進められている。

IRI はネットワーク基盤と計算基盤が連携したプロジェクトであり、ESnet が IRI を支えるオンデマンド大規模データ転送可能にするネットワーク基盤の提供を行っている。日本では、学術情報ネットワーク SINET においてネットワーク帯域をオンデマンドで確保する機能が提供されているが⁸、各種実験装置や計算基盤との連携は十分に行われておらず、機能の拡充が必要である。

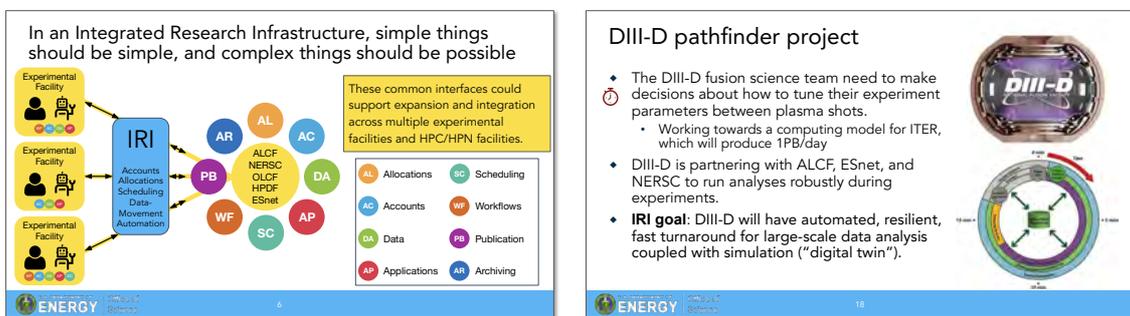


図 3 DOE/IRI

(GuokChin. Connecting Science with the DOE’s Integrated Research Infrastructure. より)

2.4.3 一体的運用の調査研究の進め方

(1) 前提

- 次世代計算基盤は、国内学術機関の計算資源の一部を時空間的に借用して構成する。計算資源は提供機関のローカルユーザとの共有であり、運用は所有機関の方針やポリシーに従う。ただし、連携にあたり HPCI と同様に最低限の統一は行う。また、商用クラウドの IaaS インスタンスも計算資源として利用可能とする。
- 公募要領の調査研究内容の1つである「ジョブと計算資源のアロケーション最適化」に対応するため、利用申請時にアーキテクチャ指定など幅を持った資源の指定を可能とする。

(2) 調査研究方針

⁷ Guok Chin. Connecting Science with the DOE’s Integrated Research Infrastructure. (オンライン)(引用日: 2024年9月18日.) <https://grpworkshop2024.theglobalresearchplatform.net/program/M1.3-5GRP-2024-GUOK-IRI.pdf>.

⁸ Atsuko Takefusa, et al., IaaS Migration Using the FELIX Federated Testbed, Proc. 5th IEEE CloudNet 33-38 2016年10月.

- 公募要領から抽出した調査研究内容の資源管理関連の項目を実現するための機能、構造等の検討を行う。
- 「目指すべき姿」から推測される計算資源と実験施設の自由度の高い資源レベルの連携は、実験施設毎の運用方法、利用方法、ポリシーなどの違いにより EOSC 同様の困難が予想される。このため、IRI と同様に実験施設など外部基盤側から計算基盤を計算アクセラレータ的に利用する連携の検討を行う。
- 利用エンドポイントを、利用形態（計算のみ／実験連携）と利用機能（従来機能／新規機能）の 2 軸で検討を行う。ここで、実験連携とは実験施設との連携だけでなく IoT システムなどデータ収集機能を持つ機構との連携を含む。また、新規機能とは従来機能を含む後述の必要機能を指す。

2.4.4 機能と構成（案）

(1) 利用エンドポイント

利用形態（計算のみ／実験連携）と利用機能（従来機能／新規機能）を組合せた 4 種類の利用エンドポイントを検討した。構成を図 4 に示す。

- EP1: 次世代計算基盤ポータルからの利用（計算のみ・新規機能）
次世代計算基盤ポータルから新規機能を利用してワークロードを投入・管理する。
- EP2: 計算基盤外部からの利用-1（実験連携・新規機能）
分野別研究基盤など外部から次世代計算基盤の計算資源を計算アクセラレータ的に利用する。次世代計算基盤 I/F 経由で新規機能を利用してワークロードを投入・管理する。
- EP3: 計算基盤外部からの利用-2（実験連携・従来機能）
EP2 同様に外部から計算アクセラレータ的に利用する。計算資源の抽象化ジョブスケジューラにワークロードを投入・管理する。このため、分野別研究基盤などの外部基盤が所有するジョブスケジューラベースのクラスタなどとの置換えが比較的容易となる。
- EP4: 次世代計算基盤計算資源の直接利用（計算のみ・従来機能）
従来と同様に、計算資源に直接ログインし抽象化ジョブスケジューラにワークロードを投入・管理する。

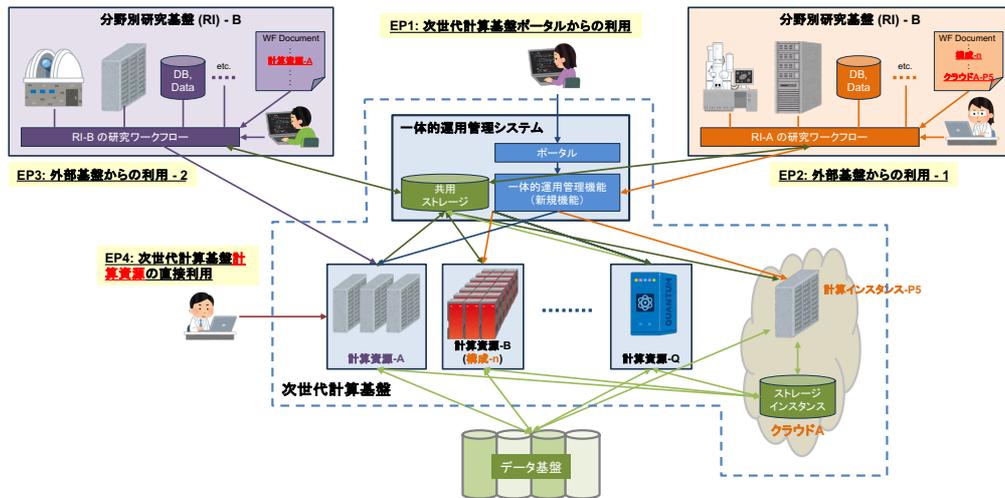


図 4 次世代計算基盤の利用エンドポイント

(2) 機能検討

単一ワークロード実行に必要な基本項目、下記検討項目 A-D、その他関連項目についてユースケースを設定、その実現に必要な機能を検討した。これを表 1 に示す。また、公募要領の課題との関係、他の計算・研究基盤での状況も示す。

- A) 「フラッグシップシステム」と国内主要スパコンの資源管理システム間でジョブやデータを高効率で融通する技術。
- B) 「フラッグシップシステム」他とリアルタイムデータ・蓄積データ基盤との連携
高度なデータ連携について Society 5.0 運用調査検討、データ利活用調査検討グループとともに検討。
- C) 「フラッグシップシステム」他とクラウドとの連携
連携する際の性能、可用性、ユーザビリティ、セキュリティの点で課題抽出（クラウドバースティングを例に）。
- D) 上記連携を可能にするためのネットワーク技術
学術情報ネットワーク SINET に求められる性能・機能要件を調査。

表 1 ユースケースと必要機能

検討分類	ユースケース		ユースケースから抽出した必要機能	説明	公募要領の課題との関係	他計算・研究基盤					
	目的	操作・動作				欧州・EGI	欧州・EOSC	米国・ACCESS	米国・DOE	日本・HPCI	参考・NAREGI
基本機能	資源選択～実行	利用者が選択した任意の資源でシンプルなワークロードを実行	1. 統一された ID、アカウントの管理	次世代計算基盤全体で統一された利用者 ID、アカウントの管理	p1	○	○	△	×	○	×
			2. 全ての資源、サービスに対応する認証・認可	ポータル、任意の資源・サービスへ SSO 可能な認証・認可	p1, p4, p7, p8, p11, p14	○	○	○	×	○	○
			3. 全資源、全プロバイダ等の静的、動的な情報の管理、表示等	WL 実行時の最適資源選択や管理のために、資源情報等の収集、管理、利用者や管理者への可視化、検索	p1, p4, p6, p7, p8, p11, p14	○	○	○	×	×	△
			4. 実行形式プログラム管理	異なる計算資源間での実行形式プログラムのポータビリティのため、コンテナ化や資源毎のビルドの管理	p1, p4, p7, p8, p11	○	*	△	×	×	○
			5. 抽象化された WL 実行管理	資源のジョブスケジューラに依存しない、抽象化された WL ドキュメントと操作方法による WL の実行管理	p1, p4, p7, p8	○	*	△	×	×	○
			6. 共用ストレージ	全計算機資源から共有可能なネットワークストレージ	p1, p4, p7, p8, p14	○	×	×	×	○	○
			7. 大規模データ、大量ファイルの高性能・セキュアなデータ転送	共有ストレージ、クラウドストレージ、リポジトリ、データ基盤などとの大規模、大量データの高速度・セキュアな転送。リモートからの指示による資源間の直接データ転送も可能	p1, p4, p7, p8, p14	○	△	○	○	○	○
			8. 統一的なアカウント管理	全計算資源で統一的なアカウント管理	p1, p4	○	×	○	△	×	×
			9. 一体的運用ポータル	利用者向け機能: 全資源のコンソール、様々な利用者機能の操作が可能な利用エンドポイント 管理者向け機能: 利用者向け機能＋一体的運用全体の管理機能	p1, p3	○	○	○	×	×	○
A	最適資源選択	WL 実行要件、利用者要件、資源状況などから、利用者が使用する資源を選択	基本機能 (#3)								
		利用者に異なる利用可能資源から、実行に最適な資源をシステムが選択	10. 資源ブローカー/メタスケジューラ	利用可能資源から、WL 実行条件、利用者要件、資源・環境状況、一体的運用資源割当ポリシーに従い最適資源の自動選択	p3, p4, p6, p7, p8, p13	○	*	△	×	×	○
	実行	複数の資源に渡る WF 実行	11. 複数プロバイダ対応 WF システム	WF ステップ毎に適した資源で実行。異なるプロバイダの資源も利用可能。単一 WL も 1 ステップの WF として実行	p2, p3	○	*	○	○	×	○
		複数の資源のコンテナオーケストレーション	12. コンテナオーケストレータ	複数資源のコンテナのオーケストレーション。異なるプロバイダの資源も可能	p3, p11		*			×	×
		実行する WL の管理	基本機能 (#4)								
		WL 投入・管理操作	基本機能 (#5)								
		複数の従来型計算資源による連成計算、システム分散計算	13. 従来型計算機資源間での連成計算のための実行開始時刻同期	連成計算を実行する計算機資源間、または分割実行する計算機資源間で実行開始時刻を同期させた実行	p2	○	*	×	○		○
HPC・QC ハイブリッド計算	14. HPC・QC ハイブリッド計算制御	HPC 資源と QC で連成計算するための同期実行制御。QC の利用率低下防止のため、#13 は使用できない	p2		*		△	△	×		
備考	<ul style="list-style-type: none"> ○, △, ×: 機能項目内での相対評価、空白: 不明 *: 連携する基盤依存 										

(WL: ワークロード、WF: ワークフロー、QC: 量子コンピュータ)

次ページに続く

検討分類	ユースケース		ユースケースから抽出した必要機能	説明	公募要領の課題との関係	他計算・研究基盤					
	目的	操作・動作				欧州:EGI	欧州:EOSC	米国:ACCESS	米国:DOE	日本:IPCI	参考:NAREGI
A (続)	状況変化対応	WL 実行待ち時の資源や環境の状況変化への対応	15. 実行待ち WL の時空間的移動	資源故障、電力事情による縮退運転発生、負荷集中などの時に、実行待ち WL の実行順や実行資源を変更することによる可用化や負荷の平滑化	p8	○	*	○	△	×	×
	防災・減災	災害時における緊急 WL 優先実行	16. チェックポイント/リスタート	強制中断/再開のための WL 実行状態管理。イベント駆動型 WL で許容レイテンシが長い場合などに利用	p7	△	*	△	○	×	×
			17. スポット WL	キャンセル可能な低利用コスト WL。再開、再投入はされない。イベント駆動型 WL で許容レイテンシが中程度の場合の資源確保などに利用	p7		*	△	○		×
B	データ利用・蓄積	一体的運用環境からデータ基盤の利用	18. データ基盤 I/F のサポート	データ基盤に蓄積されているデータの利用、データ基盤へのデータ蓄積	p2, p5, p7	○	*	○	○	×	×
	計算機資源提供	データ基盤など外部基盤から次世代計算基盤の計算機資源の利用	19. データ蓄積用メタデータのサポート	蓄積データへのメタデータ添付サポート、メタデータ作成サポート	p5, p7	△	*	△	△	×	×
			20. 計算基盤 I/F	分野別研究基盤やデータ基盤の計算・解析サービスなど、外部基盤から次世代計算基盤の計算資源利用をするための I/F	p5, p7, p14	○	*	△	△	△	△
	研究加速	実験施設、IoT システム連携のための低遅延 WL 起動・再開(リアルタイム処理、ストリームデータ処理等)	21. プリエンプティブ WL	計算ノード共有可能な WL タイプ。イベント駆動型 WL で許容レイテンシが中程度の場合などに利用	p2, p7, p14		*	○	○		×
22. ギャングスケジューラと同期型の WL 高速切替え			プリエンブション WL を切替えるギャングスケジューラと、イベント発生時のイベント駆動型 WL への高速切替え	p2, p7, p14	△	*	△				×
C	クラウド資源利用	クラウドインスタンス(IaaS)の管理	23. インスタンス管理の抽象化	クラウドプロバイダ毎に異なるインスタンス管理方法を抽象化し、操作を統一	p4, p5, p7, p12	○	*				×
			24. 秘匿情報管理	利用者固有の秘匿情報(クラウドプロバイダのクレデンシャル等)の安全管理	p4, p5, p7, p12	○	*				
	資源提供	クラウドから一体的運用資源の利用	基本機能+#20	一体的運用計算資源のクラウド的な利用(例. 富岳クラウド)	p5, p7, p12		*			○	×
D	データ転送	計算データと計算結果の転送 データ転送時間保証	基本機能(#7)								
			25. ネットワークのバンド幅保証と保証時間管理	計算データの転送遅延による WL 実行遅延防止のために、データステージングに要する時間(推定)のネットワークバンド幅を確保	p2, p4, p7, p8, p14	○	*		○	×	×
その他	セキュリティ	一体的運用としてのセキュリティ	26. プロバイダセキュリティと一体的運用セキュリティ	各プロバイダでのセキュリティ対策と、侵害が発生した時に一体的運用全体への波及を防ぐための構成や手順	p4, p7, p9, p10	○	○				×
	高可用性	一体的運用環境の 24/7 の利用	27. 一体的運用管理システムの高可用性	実行中 WL の管理や WF の引き継ぎが可能な高可用性	p4, p13	△	△	×	×		×
	資源プロバイダ構築	資源プロバイダ環境の構築作業	28. 資源プロバイダ機能のコンテナ化/アプライアンス化	資源プロバイダの容易かつ確実な構築のための、一体的運用関連モジュールのコンテナ化/アプライアンス化		○	○	△	○		×

備考
 ・ ○, △, ×: 機能項目内での相対評価、空白: 不明
 ・ *: 連携する基盤依存

(WL: ワークロード、WF: ワークフロー、QC: 量子コンピュータ)

(3) 特徴的な機能

いくつかの特徴的な機能について説明する。〈〉内の数字は、表 1 の機能番号または 2.4.1 節で採番した調査研究内容番号である。

● 資源ブローカ/メタスケジューラ 〈#10〉

ワークロード投入時に、ワークロード実行要件、コストやデッドラインなどの利用者要件、資源運用状況・予定、負荷や消費電力などの資源の動的状況、次世代計算基盤としての資源割当ポリシーなどを基に、実行に適した計算資源を選択決定しワークロードを投入する。

この機能は、公募要領の調査研究課題である、ジョブと計算資源のアロケーション最適化〈p4〉、SDGs の達成貢献〈p8〉、高可用性〈p13〉などの実現に必要もしくは影響する。これらの課題を実現するには、まず次世代計算基盤の全資源を対象とした資源割当ポリシーや資源状況による資源の優先順位付けが重要である。そして、ワークロード要件や利用者要件など個々の条件により最終的に資源を選択する。このため、ワークロードや利用者にとって最適な資源であるとは限らない。ただし、緊急性を要するワークロードについては、ワークロードの要件を優先する必要がある。

資源ブローカ/メタスケジューラが機能するには、少なくとも次の関連機能が必要である。全ての資源、サービスに対応する認証・認可〈#2〉、全資源、全プロバイダの静的、動的な情報の収集、管理、表示等〈#3〉、実行形式プログラム管理〈#4〉、抽象化されたワークロード実行管理〈#5〉。実行形式プログラムのポータビリティは、多くの場合コンテナ化が有効な手段となるが、詳細なアーキテクチャや物理構成の違い等により完全ではない。最終的には資源毎のビルドが必要になることがあり、その管理が必要となる。また、このような詳細な違いも静的情報に含め、資源ブローカリングの判断に加えることも考えられる。構成と動作の例を図 5 に示す。この図では、資源ブローカリングで各資源の詳細まで判断することを仮定している。

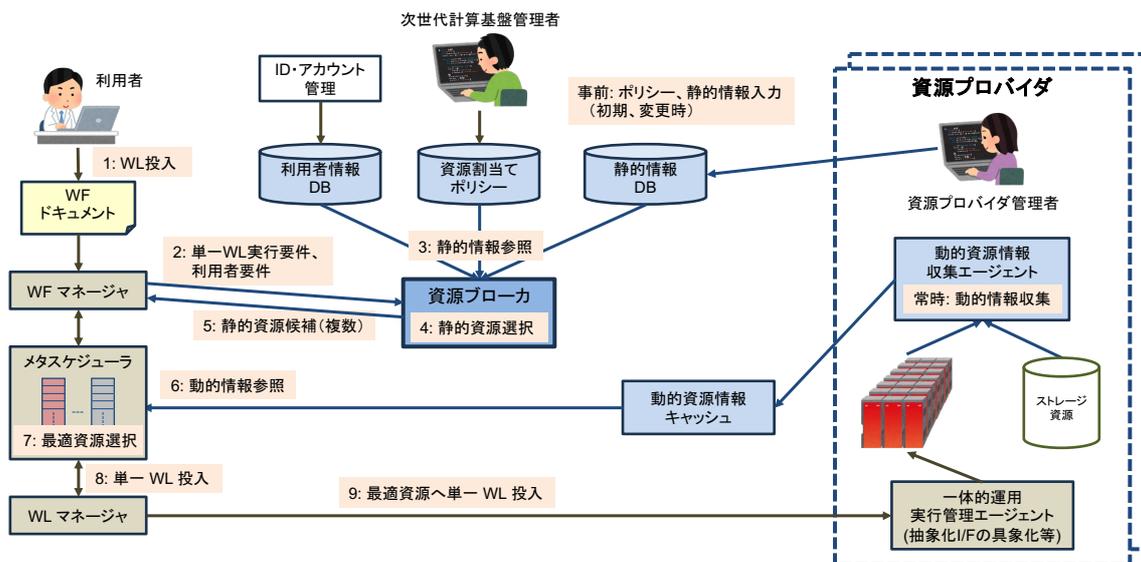


図 5 資源ブローカ/メタスケジューラの構成と動作の例

● イベント駆動型ワークロード

緊急ワークロード、実験施設連携ワークロード、リアルタイムデータワークロード、ストリーミングデータワークロード、LLM 推論ワークロードは、イベント発生から一定時間以内（以下、「許容レイテンシ」と表記）にデータ処理開始/再開が必要なワークロードである。このため、イベント駆動型ワークロードとしてまとめて検討した。表 2 に、ユースケースとそのデータ処理の起動・再開レイテンシを示す。なお、各ワークロードタイプともユースケースは一例に過ぎず、その許容レイテンシは代表的な値ではないことに注意されたい。

表 2 イベント駆動型ワークロードのユースケースと起動・再開レイテンシ例

ワークロードタイプ	ユースケース	起動 再開	ユースケースでの 許容レイテンシ
緊急WL	津波のダメージ予測（東北大） 200 個以上の WL から構成されておりワークフローで制御されている。地盤が落ち着くまで WL 実行を待つ必要がある。計算基盤やネットワークが被害を受けることも考慮する必要がある。	起動	地盤が落ち着くまでの数分
実験ループ内解析 WL	核融合実験炉の実験ループ内解析（DIII-D, 核融合研） 核融合実験の実験ショットは 3 分～20 分間隔で繰り返される。この間に、実験パラメタセット、プラズマ発生、データ転送、解析、次実験ショットのパラメタ決定を行う。	再開	最短実験間隔が 3 分のため数 100m 秒以下と推測
ストリーミングデータ WL	ビデオデータ 一般的なビデオデータは 30 フレーム/秒。RAW ビデオかつ毎フレームの処理が必要な場合で、パイプライン処理するなら 30msec 以下でデータ転送と処理を完了する必要がある。	再開	データ処理内容依存。数 m 秒単位と推測

LLM 推論ワークロード	人間によるオンデマンドな推論リクエスト リクエスト間隔は不定期。	再開	人間に遅れを感じさせないためには 100m 秒程度と推測
--------------	-------------------------------------	----	---------------------------------

様々な応用で実際に求められる許容レイテンシのレンジはかなり広く、統一した方法で管理することはできない。許容レイテンシ別の機能や実現方法を検討した。資源利用効率の低下を避けるために、可能な限りイベント検出処理とデータ処理を別ワークロードとして実行することを想定している。

➤ 許容レイテンシ：数分～

許容レイテンシ内に他ワークロードの実行を中断し、その資源でデータ処理ワークロードを実行する。終了後、中断した他ワークロードを再開する。中断・再開にはチェックポイント／リスタート<#16>機能を利用する。ノードを占有できる利点がある。

➤ 許容レイテンシ：100m 秒～数分

中断するワークロードのチェックポイントを取得する時間的余裕はない。そこで2通りの方法を検討した。

◇ スポットワークロード<#17>

キャンセルされる可能性があるが利用コストが安いスポットワークロードを導入。イベント検出時、スポットワークロードをキャンセルして、データ処理ワークロードを実行する。ノードを占有できる利点がある。

◇ プリエンプティブワークロード<#21>

プリエンプティブワークロードを導入し、イベント駆動型ワークロード（データ処理ワークロード部）と一般ワークロードが計算ノードを共有する。プリエンプション機能またはギャングスケジューラが必要となるが、前者は既実行ジョブの性能の課題があり、後者はスケラビリティ等の問題で同期型のワークロード高速切替え <#22> が必要となる。

➤ 許容レイテンシ：～100m 秒

このレイテンシでの、イベント検出とデータ処理ワークロードを分離しての起動・再開は難しい。資源利用効率は低下するが、両者を同一計算ノード上で常時実行する。

なお、人命に関わる緊急ワークロードは、イベント駆動型ワークロードの中でも高優先で実行される機能が必要である。

- プロバイダセキュリティと一体的運用セキュリティ <#26>

データセンターのセキュリティは、境界防御から侵入や内部不正（オペレーションミスを含む）を前提とするサーバーレジリエンスへと移行しつつある。そこで、サイバーレジリエンスのセキュリティフレームワークである米国国立標準技術研究所の Cybersecurity Framework 2.0 (CSF 2.0)⁹を基に、HPC データセンター向けセキュリティガイドラインのドラフトを作成し、HPCI に参加している 10 機関にコメントを求めた。このコメントを反映し、ガイドラインを作成した。なお、一体的運用としてのセキュリティポリシーや侵害された時の手順は、開発する構成や機能に大きく依存するため、今後設計を基に策定する必要がある。

- 一体的運用管理システムの高可用性 <#27>

次世代計算基盤の一体的運用管理システムは、地域的に分散した複数の同一システムによる高可用運転を行う。一体的運用管理システムの停止やネットワークの不通により、次世代計算基盤全体が利用不可能になるのを防ぐためである。次世代計算基盤では、災害時の緊急ワークロードも実行するため、全資源を管理するシステムが地域的に広く分散していることが重要である。また、緊急ワークロードのなかには 200 以上のワークロードのワークフローで構成されているものもあるため（東北大、津波被害予測）、実行中ワークロードの管理とワークフローの実行を引き継げる高可用性が必要である。

(4) 構成案

利用エンドポイント（図 4）、必要機能（表 1）を基にした高機能構成、利用者が資源を指定する必要があるがワークフロー等で資源間の連携が可能な資源連携構成、従来の HPCI と同様に計算資源を単独で利用する基本構成の 3 種類の構成を検討した。下記の構成案の灰色部分は、その構成案ではサポートしない機能であり、他の構成案との差分となる。

- 構成案 1：高機能構成

表 1 の必要機能の全てを実装し、公募要領に挙げられている課題の多くを実現可能な構成である。

⁹ CYBERSECURITY FRAMEWORK. (オンライン)(引用日: 2025 年 5 月.)
<https://www.nist.gov/cyberframework>

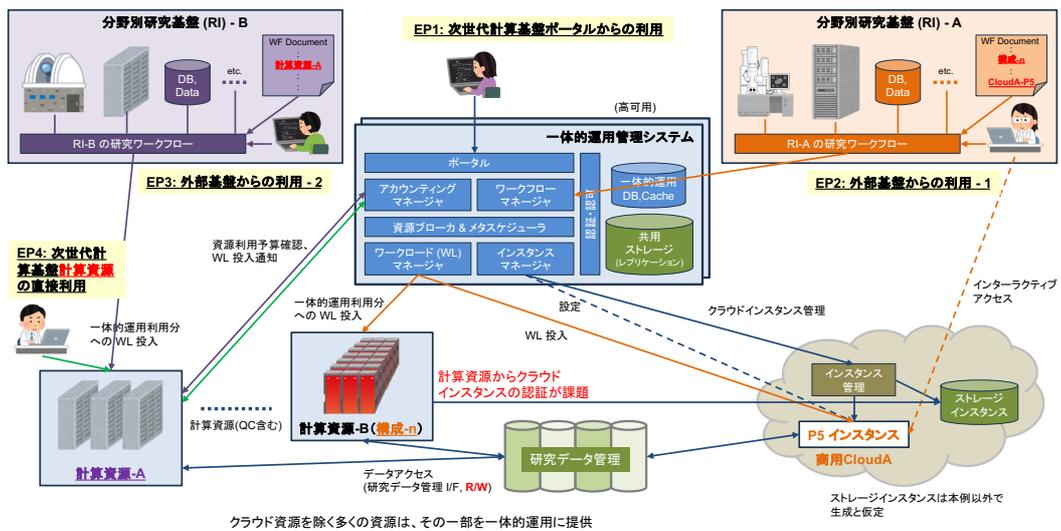


図 6 構成案 1：高機能構成

● 構成案 2：資源連携構成

高機能構成から資源ブローカ／メタスケジューラ関連の機能を省いた構成である。利用者が資源を指定する必要があるが、資源間を連携させたワークフローの実行や連成計算等が可能である。資源ブローカ／メタスケジューラがないため、資源の全体最適化が必要な負荷分散、カーボンニュートラル、計算資源の高可用などの課題への対応は困難となる。

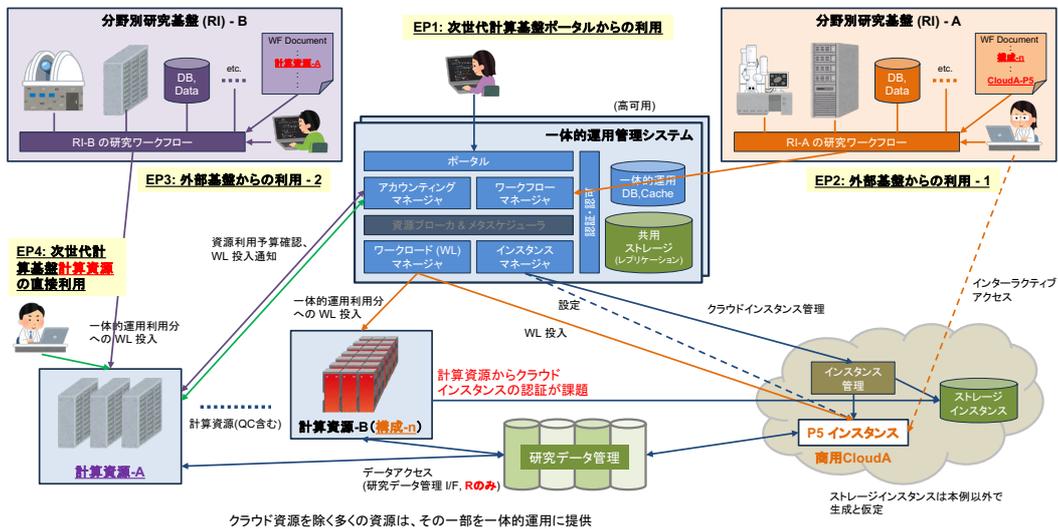


図 7 構成案 2：資源連携構成

● 構成案 3：基本構成

構成案 2 から計算資源間の連携機能を省いた構成である。従来の HPCI と同様に計算資源を単独で利用する。

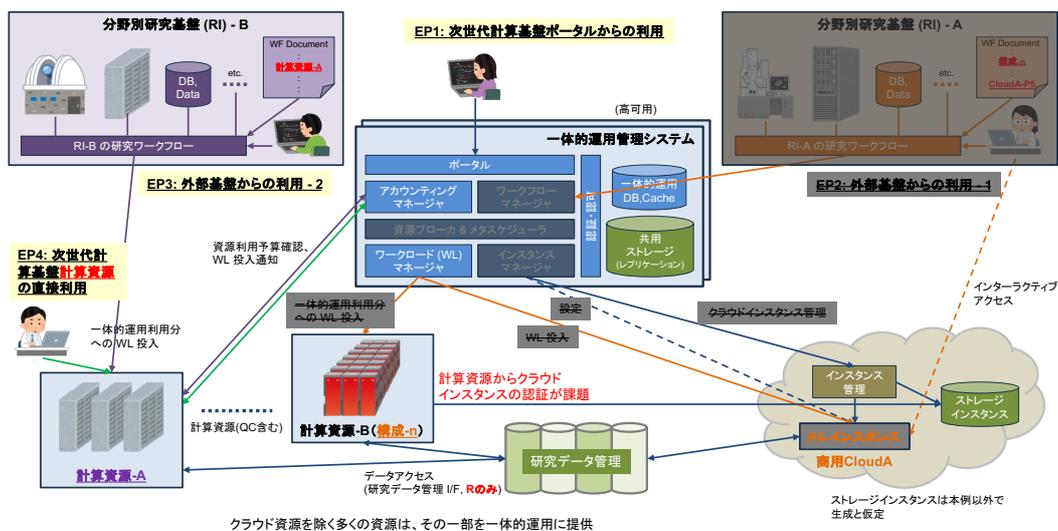


図 8 構成案 3：基本構成

2.4.5 資源管理まとめ

国内主要計算基盤、データ基盤、ネットワークの一体的運用検討をテーマに計算資源（計算機、ストレージ、ネットワーク等計算要素）間の連携、ならびに実験施設と計算資源の連携について検討を実施した。また、従来の HPCI ではサポートされていないタイプのワークロードについても検討を実施した。さらに、HPCI 参加機関からのコメントをいただき、サイバーレジリエンスを考慮した HPC 向けセキュリティガイドラインを作成した。

欧米の計算基盤では計算資源間の連携機能に違いがあり、欧州の EGI は内部に様々な連携機能を持つ一体型、米国の ACCESS は計算に特化し連携機能を外部に持つ分散型となっている。次世代計算基盤の構成として、一体型を採用すると複数の外部基盤との連携、研究資源間の連携が容易になるが、HPCI からの移行には多くの変更が必要になる。分散型は外部基盤との連携機構が複雑になる可能性があるが、HPCI からの移行は小変更で済む構成を取ることもできる。そこで、構成的には一体型に近いが、従来と同様な利用方法も可能な計算基盤の構成を検討した。さらに、機能を絞り込むことで開発工数を軽減する構成も検討した。

緊急ワークロードや実験ループ内リアルタイム計算ワークロードなど、HPCI とは異なるワークロードタイプの検討も行った。これらはイベント発生から処理開始までの時間に制約のあるワークロードであり、その許容レイテンシ毎の機能を検討した。

結果として、20 ユースケース+イベント駆動型ワークロードの 4 ユースケースから 28 の必要機能を抽出し、連続性がある 3 種の構造を検討した。一体的運用は、機能が豊富かつ複雑になると考えられるため、次世代計算基盤ユーザに対してだけでなく、資源プロバイダのローカルユーザにも影響を与えないよう十分注意して設計する必要がある。

2.5 データ利活用

2024 年度は、2022 年度に開始した各種の調査と分析の継続作業を行うとともに、他グループと連携して Open OnDemand やワークフローツールの調査・普及活動を実施した。前者の調査と分析については、新たに稼働を開始したスーパーコンピュータシステムに対するデータ入出力性能等の継続的調査であるため、詳細は割愛する。

Open OnDemand は Web ブラウザから計算資源を利用できるツールであり、海外のスパコンセンターを中心に利用が広がっている。またパブリッククラウドでは Web ブラウザを利用した計算資源の利用が一般的である。日本国内でも「富岳」を含む HPCI 資源に対して Open OnDemand の導入が進みつつある。Open OnDemand は従来のスパコンで使われてきた「SSH 対話処理+バッチジョブ実行」に慣れていないユーザに対するスパコン利用のハードルを下げるのみならず、Jupyter Lab やリモート可視化など GUI を伴うソフトウェアの利用にも適しており、今後もさらに導入が増えていくことが期待される。そこで本グループでは、グループをまたいで Open OnDemand に関する情報の共有を行うとともに、PC クラスタコンソーシアムが開催するワークショップに参加するなどして情報の共有と収集および普及活動を行った。

参考：PCCC HPC OSS ワークショップ in 九州大学

<https://www.pccluster.org/ja/event/2025/01/250127-hpc-oss-ws.html>

(本調査研究の参加者が講演者としても参加)

本グループでは、NII が運用する研究データ基盤 (NII Research Data Cloud: RDC) をハブとして活用しながら、各種ストレージや計算基盤を連携させて利用させるというデータ利活用のシナリオを想定している。既知の活用事例としては、例えば JupyterLab から研究データ管理基盤 (GakuNin RDM, RDC を構成する 3 基盤のうちの 1 つ) にアクセスしてデータ解析を行うコンテナ環境の利用情報などが提供されているが、HPCI システム上では Docker が使えなかったりジョブスケジューラが対応してい

なかつたりするため利用が難しいのが現状であった。そこで本グループではオープンなワークフローツールである Nextflow に修正を加えて「富岳」などで利用されている Technical Computing Suite に対応させ、GakuNin RDM のファイル操作とジョブ実行を含むワークフローを実行できるようにした。この手順は今後 Web サイト等にて詳細を公開する予定である。また Nextflow に対する修正（機能追加）は Nextflow の開発リポジトリへの提供を進めている。

2.6 Society5.0 運用

2.6.1 複数拠点 HPC システム間の統一的管理とワークフロー

複数の機関が提供する計算資源や、クラウドとスパコン間を横断的に利用することを想定し、米国 CIQ 社によって開発中の Fuzzball による複数コンテナ制御機能を用いて、ワークフローの試行を実施した。

具体的には、東京大学に設置されている mdx と、Wisteria/BDEC-01 Aquarius (以下 Aquarius) により、複数のスーパーコンピュータシステムを模擬し、この間で統一的な管理及び活用の概念実証を行った。

「統一的な管理」としては、mdx 上のゲスト OS にインストールされた Fuzzball Orchestrate (Control Plane) を通して、拠点 1 として mdx 上の VM 1 台、拠点 2 として Aquarius 上の計算ノード 2 台を管理することがそれに相当する (図 1)。また、ここでのワークフローは、複数のジョブから構成される計算パイプラインである。ワークフローの構成単位であるジョブとは、従来のバッチジョブスケジューラにおけるジョブに相当する。ここでは、1 つのワークフローの中で、以下に列記するジョブを実行した (図 2)。

(1) Fuzzball について

Fuzzball は、ワークフロー管理を簡素化し、移植性を向上させ、従来の HPC システムの複雑さを解消するために設計された、最新のコンテナベースの HPC プラットフォームである。従来の HPC システムが旧来のインタフェースに依存しているのに対し、Fuzzball は直感的な Web UI と CLI を提供し、初心者から上級ユーザまで容易に利用可能である。

Fuzzball は 2 つの主要コンポーネントで構成されている。Orchestrate は Kubernetes 上で動作し、ワークフローを管理する。一方の、Substrate は計算ノード上でジョブを

実行する軽量なサービスであり、低オーバーヘッドである。ワークフローは Fuzzfiles (実際には YAML による記述) で定義される。

Fuzzball の主な利点は、開発元の CIQ 社によると以下の通りである。

- 使いやすさ：SSH やバッチスケジューリング、Linux ファイル管理が不要。
- 再現性：完全コンテナ化されたワークフローにより、共有とデプロイが容易
- クラウド/オンプレミス互換性：異なる環境間でシームレスに実行可能。
- ソフトウェア管理の不要化：コンテナ URI を指定するだけでソフトウェアを利用できる。
- 幅広い MPI サポート：さまざまな MPI 実装に対応し、再コンパイルの手間を削減。
- 簡単なユーザ・データ管理：直感的な階層と権限管理システムを提供。

これらの特徴により、Fuzzball を用いることで、HPC とクラウドのギャップを解消することが容易になると考えられる。AI for HPC を始めとした計算ワークフローの複雑化、計算リソースの多様化・複雑化の中で、Fuzzball は、ユーザに計算環境を定義する自由、セキュリティを担保した実行環境、データ・ハンドリングの自動化など、これまでの HPC システムに不足していた機能を提供しようとするものである。

(2) Fuzzball の機能

Fuzzball はこれらの機能を統合することで、HPC ワークフローを効率化し、管理負担を軽減する。

- 使いやすいインタフェース
 - Web UI と CLI：複雑なコマンド操作なしで、ワークフローの送信・監視・管理が可能。
- (1) ワークフロー管理
 - Fuzzfiles：ワークフローは YAML ベースの Fuzzfile で定義され、ジョブの依存関係、コンテナイメージ、リソース要件などを記述可能。
 - ジョブ管理：ジョブの作成、実行、監視ができ、ログの取得やステータス確認が容易。
- (5) 高度な機能
 - 対話型シェルアクセス：実行中のジョブ内でコマンドを実行し、リアルタイムで操作可能。
 - コラボレーションツール：他のユーザのワークフローと連携し、共同作業を促進。
- (1) カスタムホスト名：ジョブに特定のホスト名を割り当て、識別・管理を容易

にする。

(6) データ管理

- (1) ボリューム：一時・永続ボリュームを利用し、データの入出力を効率的に管理。
- (2) シークレット管理： S3、HTTP(S)、コンテナレジストリなど、さまざまなタイプのシークレットを安全に管理。

(7) 並列・分散処理

- ▶ ジョブアレイ： 複数のジョブインスタンスを同時に実行し、並列ワークフローを実現。
- ▶ MPI および PGAS 対応： MPI（メッセージパッシングインタフェース）や PGAS（分散共有メモリ）を活用し、分散コンピューティングを支援。

(8) ハードウェア統合

- (1) GPU 対応： GPU リソースを活用し、計算負荷の高いタスクを加速。
- (2) 高速ネットワーク： 高速ファブリックとの統合により、データ転送速度を向上させ、遅延を低減。

(9) 組織管理

- (1) エンティティ構造： 組織・アカウント・ユーザの明確な階層を定義し、アクセス管理を容易にする。

(10) 開発者向けサポート

- (1) クライアント SDK： API と連携するためのソフトウェア開発キット（SDK）を提供し、ワークフローのリスト取得、送信、ステータス確認などが可能。

(11) ワークフロー例

- (1) サンプルワークフロー： OpenFOAM、LAMMPS、BLAST などの高度なシミュレーションまで、多様なサンプルを提供。

(12) クラスタ管理

- (1) デプロイと構成： Kubernetes クラスタ上で Fuzzball を導入・設定・テストするためのガイドを提供。
- (2) ストレージ管理： ストレージドライバーやクラスの設定・管理方法を詳述。

(3) 関連するソフトウェア・フレームワーク・サービスとの機能比較

本節では、スーパーコンピューティング・インフラで用いられる関連するソフトウェア・フレームワーク・サービス（以下、ソフトウェア）との比較を行う。Fuzzball は、既存のソフトウェアが其々に持つ様々な機能を複合・自動化したフレームワークであるため、既存のソフトウェアが持つ機能と Fuzzball が持つ機能とを列挙する。

Fuzzball System Overview

Fuzzball clusters allow users to submit workflows to the Fuzzball Orchestration services, which manages the data, container image management, and volumes and submits the job graph to the scheduler to reserve resources.

The jobs are run via Fuzzball Substrate service which manages those workflows on unlimited scale.



図1 Fuzzball cluster 概念図

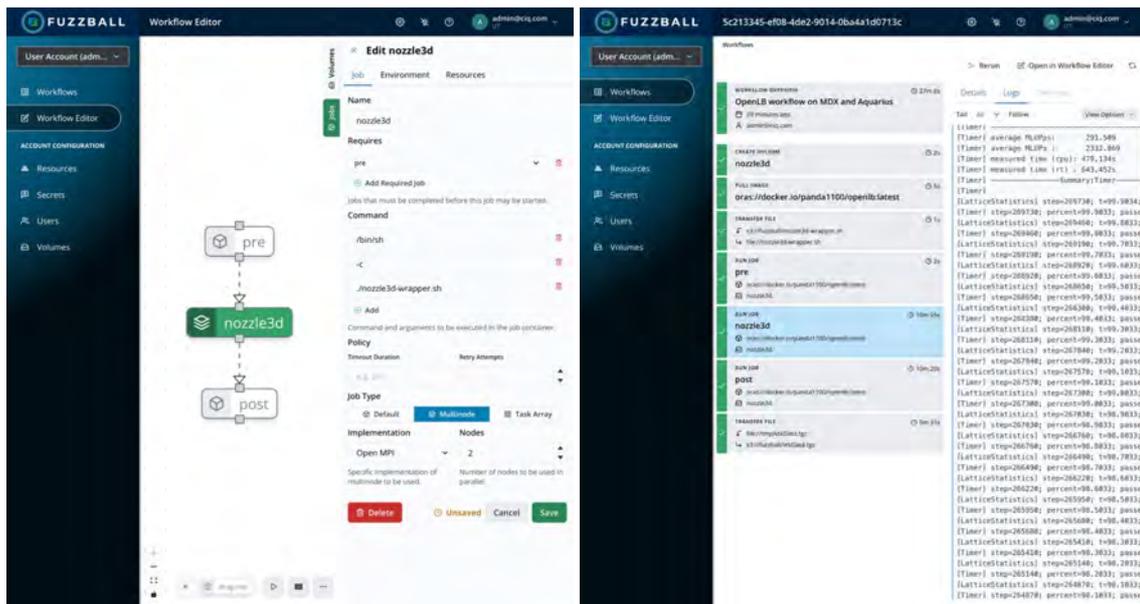


図2 Fuzzball を用いたワークフローの作成と実行

表1 Fuzzball と他のソフトウェアとの比較 (CIQ 調べ)

	Fuzzball	Slurm	Nextflow	Open OnDemand
Web UI	☑	☑		☑

CLI	✓	✓	✓	✓
ワークフロー	✓		✓	
ジョブ管理	✓	✓		✓
対話型シェルアクセス	✓	✓		✓
コラボレーションツール	✓			
カスタムホスト名	✓			
ボリューム管理	✓			
シークレット管理	✓		✓	
ジョブアレイ	✓	✓	✓	
MPI、PGAS	✓	✓	✓	
GPU	✓	✓	✓	
高速ネットワーク	✓	✓		
エンティティ管理	✓			
クライアントSDK	✓	✓		
サンプルワークフロー	✓	✓	✓	✓
デプロイと構成	✓	✓	✓	✓

2.6.2 JupyterLab からスパコンシステムへの透過的なジョブ実行

複数の機関が提供する計算資源を横断的に利用するような使い方を考えたとき、統一されたワークフローツールが無ければ、独自に各システムに合わせてジョブスクリプトを記述し、それぞれの間を繋ぐようなスクリプトをシステム間の状況に応じて用意する必要がある。ワークフローツールがあれば、共通の記述によって簡便にはなるが、各システムへのインストールやツールの習熟などが必要である。

JupyterLab は de-facto standard なインタラクティブ Web UI であり、特にデータ解析や機械学習の研究者は日常的に用いている。JupyterLab によって記述したフローを、ある処理単位でスパコンシステムに対してジョブ実行を透過的に行えるような仕組みがあれば、より自然にワークフローを実現することが可能になる。

昨年度、JupyterLab からスパコンシステムへの透過的なジョブ実行環境を試作した。実際には、mdx のゲスト OS 上に JupyterLab 環境を用意し、東京大学情報基盤センターの Wisteria/BDEC-01 のバッチジョブシステムへ透過的な利用を可能にした。JupyterLab を拡張し、ジョブスケジューラへの接続、keycloak によるスパコンアカウントに対する鍵の管理、ジョブステータスの管理機能などが実現できた一方で、(1) マルチユーザへの対応、(2) JupyterLab 上でセルを跨いだ変数を保持する機能、(3) ジョブ実行結果ファイルの自動転送、(4) 外部ストレージとの連携機能が必要、との課題があった。

そこで今年度はこれらの課題の解決を図った。

- (1) マルチユーザ対応: JupyterHub への対応と Keycloak により、マルチユーザ対応を実現した。
- (2) セルを跨いだ変数の保持: Python の dill ライブラリを利用して、セル内の値を保存しセルを跨ぐ際の値の再利用を可能にした。
- (3) ジョブ結果ファイルの自動転送: rsync コマンドによりジョブ結果ファイルをコピーし JupyterHub での表示を可能にした。
- (4) 外部ストレージとの連携: S3 互換オブジェクトストレージを想定し、s3cmd の同期コマンドにより連携を実現した。

HPCI における次期認証基盤との共通化に向けて、今回開発したユーザ認証機構において、認証局を hpci-oauth へ変更することで、HPCI が提供するトークンベースの認証・認可システム (HPCI 用 OAuth ssh) を活用したアクセストークンベースの仕組みが活用できると見込まれる。実際には、HPCI アカウントを利用した検証が必要となるため現状では未実施である。

2.7 HPCI 運営

HPCI は、国内の大学や研究機関の計算機システムやストレージを高速ネットワークで結ぶことにより、全国の HPC リソースを全国の幅広い HPC ユーザ層が効率よく利用できる科学技術計算環境を実現することに加え、HPCI の運用を通じて多様なユーザーニーズに応えるとともに画期的な研究成果を創出し、科学技術の発展や産業競争力強化に資するとともに、人材育成やスーパーコンピューティングの裾野の拡大にも貢献することを目的として、2012 年 9 月 28 日の共用開始から、多くの顕著な成果を上げてきた。

しかし、共用開始から 10 年以上が経過し、制度的な改善を求める声も少なくない。次世代計算基盤の検討を行う上で、HPCI システムのあり方についての課題について踏み込んだ検討を実施すべきである。

HPCI 運営における課題は、大きく分けると 2 つの側面がある。

- ユーザの利便性向上：各センターの異なるシステムにおいて共通化された利用方法の提供
- システム運用：上記を提供するための実現方法、計算資源の融通やポリシーの共通化

2024 年度は、ユーザの利便性向上について検討を継続した。また、システム運用については、認証についても調査を行った。加えて米国の事例を中心に、計算資源の整備・運用の支援、アプリケーションの保守・高度化の促進機能について、より詳細な調査を実施した。加えて、ユーザのストレージ利用の視点でシステム移行に伴う端境期をなくすための検討を実施した。

2.7.1 ユーザの利便性向上

従来のスパコンの利用方法は、いずれのシステムにおいても

- CUI ベースの利用：ssh を用いてターミナルソフトからシステムのログインノードにログイン
- バッチジョブ方式：ジョブスクリプトを記述しバッチジョブスケジューラにジョブを登録、ジョブ実行後に出力ファイルを確認

という点で共通していた。

しかし、既存のユーザに加え、幅広い分野のユーザへの対応を考えると、よりインタラクティブな環境が望まれている。特にビッグデータ解析、機械学習等のユーザには、JupyterLab¹⁰などの Web UI が好まれる。これまでもスパコン上で JupyterLab を利用可能にしている例は、東工大 Tsubame3.0 や東大 Wisteria/BDEC-01 などがあったが、前者では、ジョブを割り当てた後に計算ノード上で JupyterLab を起動する必要があり、後者では専用サーバを設けているがバッチジョブ実行との連携が困難であるなど、利便性の面では大きな問題があることがわかっている。

¹⁰ Jupyter Lab: <https://jupyter.org>

2022年度には、ユーザの利用形態の理想像を共有するために、図9のような将来のHPCIユーザの利用イメージを定めた。国内における研究データ管理基盤としては、国立情報学研究所が運用するGakuNin RDM¹¹が整備されており、GakuNin RDMをユーザのポータルとして活用することを念頭に検討を進めた。GakuNin RDMでは、mdxや商用クラウドをはじめとした様々なストレージやレポジトリを登録してユーザが使い分けることができる他、計算資源と連携させることも可能である。

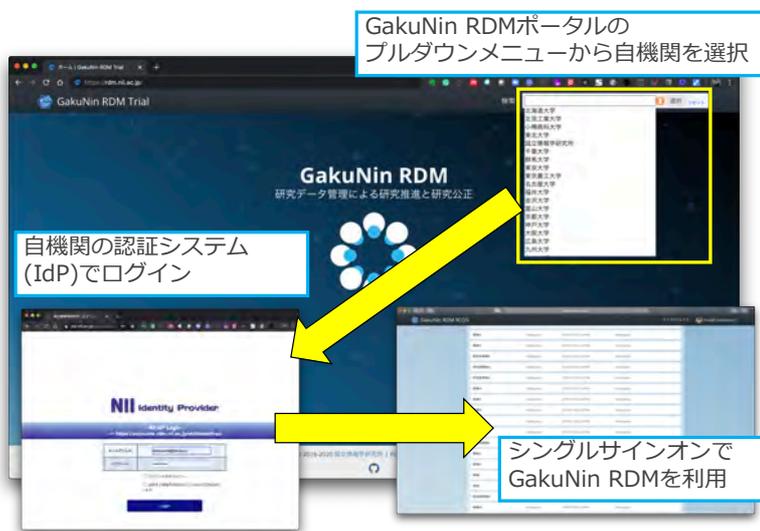


図9 今後のHPCIシステム利用イメージ

オハイオ州立大学を中心に開発されているOpen OnDemand¹²は、スパコンシステム向けにWeb UIを提供するものであり、各国で広く使われてきている。2023年度からは、富岳においてOpen OnDemandによるWeb UI環境の提供を開始し、Webブラウザから直接、富岳のジョブスケジューラにジョブを投入でき、GUIを持つアプリケーションを利用できるようになった¹³。さらに、GakuNin RDM、HPCI共用ストレージそれぞれとの間でデータの転送が可能になった¹⁴。図10は、富岳におけるOpen OnDemandのトッ

¹¹ GakuNin RDM: <https://rcos.nii.ac.jp/service/rdm/> , <https://rdm.nii.ac.jp>

¹² Open OnDemand <https://openondemand.org>

¹³ 「富岳」Open OnDemandの提供を開始 ～Webブラウザで「富岳」の操作が可能に～
<https://www.r-ccs.riken.jp/outreach/topics/20230530-1/>

¹⁴ 「富岳」Open OnDemandにおけるGakuNin RDMとのデータ転送アプリケーションの開発

ページを示したものである。図の下部にそれぞれ GakuNin RDM と HPCI 共用ストレージを示すアイコンがあり、それをクリックすることでデータ転送アプリケーションが起動し、認証ののちにデータ転送が実行される。



図 10 「富岳」 Open OnDemand トップページ

2024 年度には、新規に導入された東京科学大（旧・東京工業大）TSUBAME 4.0、九州大 玄界、最先端共同 HPC 基盤施設 Miyabi、加えて、産総研 ABCI においても Open OnDemand が利用可能になっている。

Open OnDemand 中のアプリケーションの一つとして JupyterLab も使用することができる。しかし、通常はジョブとして計算ノードに割り当てた中から利用することになるため、ユーザにとって利便性が向上するとはいえ、本質的な JupyterLab の利用形態としては以前と変わらない。ジョブ管理の外側でワークフロー的に扱うためには、2.6.2 節で試

<https://www.r-ccs.riken.jp/outreach/topics/20230726-1/>

「富岳」 Open OnDemand における HPCI 共用ストレージとのデータ転送アプリケーションの開発 <https://www.r-ccs.riken.jp/outreach/topics/20230907-1/>

行しているように Jupyter 環境からジョブを実行し結果を回収して表示、再利用できる仕組みが必要になる。

2.7.2 次世代認証基盤の調査

HPCI を構成する複数システムをまたがったワークフローなどを実現するには、共通の認証基盤が必要である。現在、HPCI では独自の認証系を整備しシングルサインオンを実現しているが、国立情報学研究所を中心に次世代認証基盤の設計が行われているのに伴って、HPCI においても次世代認証基盤の利用について検討を継続している。

HPCI における認証方式は、当初 X.509 証明書に基づく認証方式(GSI 認証)で開始したが、2024 年度から、より安全な、アクセストークンに基づく認証方式(以下、OAuth 認証)に移行した。これにより、従来の代理証明書によるアクセスに代わって、アクセストークンをエージェントから取得した上で処理を行うことになる。SSH についても GSI-SSH から OAuth-SSH になり、より標準的なパッケージでのインストールに対応でき、各システムの管理コストの低減にもつながる。

また、全国の大学等と NII が連携して構築されている学術認証フェデレーション「学認」では、次世代認証基盤として OAuth 2.0 と OpenID connect 1.0 の使用が検討されており、HPCI においてはこれらを取った形となっている。

現状の「学認」では必ずしもアカウント発行時に本人確認を実施していないが、次世代「学認」では本人確認を含めた発行プロセス(IAL2/AAL2)を推奨する予定であり、これは HPCI のアカウント発行基準に合致する。従って、次世代認証基盤では、各ユーザの所属組織が IAL2/AAL2 に準拠して発行したアカウントであることを確認した上で HPCI アカウントとして用いることができると考えられる。

一方、次世代「学認」についても基本的に学術機関のみに限られるため、それ以外の企業等のユーザをどのように扱うかが課題である。それに対しては、認証プロキシサービスを準備しようとしており、gBizID のように法人アカウントを管理できる認証サービスとの連携も可能になる。現状の HPCI プロジェクトと同様に代表者が対面認証の責任を持つような形態も実現できる。

次世代認証基盤の導入に向けては、今後、具体的に HPCI の運営にどう取り込んでいくか、議論を深めていくことにしている。複数システム、ストレージ間での柔軟な連携を進めるにはユーザ管理情報の共通化が必須であり、次世代「学認」のスパコンシステムへの適用は必然であろう。

2.7.3 ストレージの端境期対応

本節では、現行システムから次期システムへの移行において生じるストレージの端境期問題に着目し、その背景と影響、課題、および移行を円滑化するための方策について検討する。スーパーコンピュータの世代交代時には計算機本体の更新に伴い、一時的に計算資源が利用できなくなる端境期が生じることが指摘されている。特にストレージに関しては、計算機とストレージの更新サイクルの不一致や大容量データの移行作業などがボトルネックとなり、利用環境の連続性が損なわれるリスクがある。

以下では、まず端境期問題の概要と HPC センターにおける課題を整理し、次世代計算基盤への移行期に想定される影響とリスクを述べる。次に、ストレージ更新のタイミングによる問題を分析し、ストレージを計算機から独立して調達・運用するアプローチの検討や、HPCI 全体で共有可能な大規模ストレージ基盤（HPCI 共有ストレージ）の活用モデルについて議論する。また、ユーザデータの円滑な受け渡しを実現するための Globus 等の高速データ転送・共有ツールの活用とその効果、ストレージ移行に伴う技術的課題（データ転送性能、異機種間の互換性、アクセス権管理など）について整理する。さらに、オブジェクトストレージや AI ワークロード向け高速ストレージといった技術動向を概観する。最後に、ユーザが計算機の更新を意識せず継続利用できる理想的な運用像を示し、次世代計算基盤の設計者にとって有用な知見をまとめる。

(1) 端境期問題の背景と課題

本報告での端境期とは、システムの入替えに伴って旧システムの提供停止から新システムの本格稼働までの間に生じる空白期間を指す。HPC 分野では、スーパーコンピュータの世代交代に際し計算資源が一時的に大幅減少する問題として認識されている。例えば、「京」から「富岳」への移行時には、「京」の停止から「富岳」の稼働までの端境期に国内の計算資源が一時的に不足し、研究の継続性に支障をきたす懸念が生じた。この経験を踏まえ、次世代計算基盤ではこのような端境期を極力生じさせないことが重要な論点となっている。このことからポスト富岳の計画においても、新旧システムの稼働時期をオーバーラップさせ計算資源を継続提供することが目標として示されている。

端境期問題としては、計算資源の継続性が主に注目されるが、同様に重要なのがストレージの継続性である。ストレージは大規模なシステムであるほど旧システムから新システムへのデータ移行に長時間を要する可能性が高い。実際、近年の大型科学計算では単一プロジェクトで数ペタバイト級のデータが生成・保存されることもあり、システム更新時にこのような巨大なデータを移動・復元する作業は利用者と運用者双方にとって大きな負担となる。ストレージの移行が円滑に進まなければ、新システム稼働後も旧シ

システム上のデータアクセスや移行作業に追われ、利用者が計算資源を十分活用できない期間が生じる懸念もある。したがって、端境期問題の影響を最小化するにはストレージを含む計算環境や利用者環境の継続性を如何に確保するかが重要な課題となる。

(2) ストレージ端境期問題の性質：更新タイミング

ストレージの端境期問題が生じる根本要因の一つに、計算機本体とストレージ装置の更新タイミングが同一ということが挙げられる。一般にスーパーコンピュータは5年程度のサイクルで更新されることが多いが、計算機本体とストレージをひとまとまりのシステムとして更新すると、新旧システム間でのデータ移行が必須となる。このため、「計算機の更新＝ストレージも同時更新」とする従来の一体型調達を見直し、ストレージのみを別更新とする戦略が考えられる。

しかし、計算機とストレージの更新時期を意図的にずらす場合、逆に世代ミスマッチによる問題も発生し得る。新世代の計算機が旧世代のストレージにアクセスするためには、インターフェースやプロトコルの互換性が求められる。このため、新世代の計算機に最適なストレージが選択できないといった可能性が生じる。またファイルシステムも、旧システムで採用していた分散ファイルシステム（例：Lustre や GPFS など）と新システム側で最適なファイルシステムが異なる場合に、両者間の互換性確保が課題となる。

以上を踏まえ、本調査ではストレージ端境期問題に対する具体的な解決策・緩和策として、(1) ストレージ独立調達モデル、(2) 大規模共有ストレージ基盤の活用、(3) データ転送・共有ツールの導入、の各アプローチを順に検討する。また、それらを支える最新の技術動向と、コスト・運用面での評価についても議論する。

(2-1) ストレージ独立調達・運用モデル

一つ目のアプローチは、ストレージを計算機本体から独立して調達し、計算機の複数世代にわたり運用するモデルである。具体的には、現行システムのストレージ装置を次期システムでも引き続き使用するか、あるいはストレージのみ別途先行して更新して常に最新の大容量ストレージ基盤を維持し、計算機の更新時にはその共通ストレージに新旧計算機を接続する形を取るものである。このモデルでは、ユーザデータを格納するストレージが計算機の世代交代で途切れないため、端境期におけるデータ移行作業を大幅に簡素化できる利点がある。旧計算機で生成されたデータも新計算機から直接参照・利用でき、ユーザはデータの移行を意識せずに済む。

ストレージ独立運用モデルを実現するにあたっては、新旧計算機とストレージを接続するネットワークとプロトコルの標準化が重要となる。例えば計算ノードとストレ

ージの接続に InfiniBand や高速 Ethernet を用いる場合、次期システム側でも互換のネットワークアダプタを搭載し、既存ストレージをマウントできるようにしておく必要がある。またファイルシステムについても、旧システムで使用していた並列ファイルシステムを新システム上でサポートするか、もしくはオープンなネットワークプロトコル（NFS や FUSE など）経由でアクセス可能にする等の対応が求められる。

運用面では、計算機更新時にもストレージ装置群は電源投入状態を維持し、新旧計算環境双方からアクセス可能な共通データ領域を提供することになる。この際、旧計算機から新計算機へのアクセス権限やユーザ ID の統一を図ることが不可欠である。理想的には、HPC センター全体で統一された認証基盤を構築し、計算機を更新してもユーザから見たファイルの所有者や権限が変わらないようにすることが望ましい。

本モデルには課題も存在する。一つは世代間性能ギャップであり、新計算機の性能向上分を旧ストレージでどこまで支えられるかという問題である。しかしながら、多くのユーザは性能面で旧ストレージよりも新ストレージを利用するモチベーションを持っており、旧ストレージを使い続けなければならない制約もないため、大きな課題とはならない。一方で、新旧計算機とストレージを長期運用する場合には、それに伴う故障増加リスクに備え、十分な冗長構成やバックアップ計画を用意しておく必要もある。加えて、ストレージ機器の保守サポート期間を越える長期運用となる場合には、ベンダーとの協議や予備部品の確保など、運用コスト・体制面での備えも求められる。

以上のように留意点はあるものの、ストレージ独立モデルは端境期ゼロを目指す有力なアプローチと考えられる。特に利用者にとっては、計算環境の移行期にもデータへのアクセス手段が途切れない点で大きな利便性向上となる。

(2-2) 大規模共用ストレージ基盤の活用モデル

次に検討するアプローチは、HPCI 共用ストレージを活用し、計算機個別のストレージへの依存度を下げるモデルである。日本では HPCI の一環として各スパコンから利用可能な HPCI 共用ストレージが提供されており、広域分散ファイルシステム Gfarm を用いて総容量 45PB のファイルシステムを HPCI に参加する各スパコンに提供している。HPCI 共用ストレージは理研 R-CCS（神戸・西拠点）および東大情報基盤センター（柏・東拠点）にそれぞれデータストレージとメタデータサーバを備え、100Gbps クラスの広域ネットワークで各拠点と接続された高性能・高可用性な基盤となっている。ユーザは単一の認証でこのストレージを利用可能であり、センター間で共通のデータ領域として機能している。

このような共用ストレージを活用することで、各計算機固有のストレージに依存しないデータ管理が可能となり、結果として計算機の更新時にもデータの受け渡しが容易になると期待される。具体的には、ユーザは日常的に HPCI 共用ストレージ等の共有基盤上にデータを保存・共有することで、仮にローカル（計算機固有）のストレージが更新されてもデータは共有基盤上に残存する。このモデルでは、計算機の世代交代時にユーザは新システムから共有ストレージ上の自分のデータにそのままアクセスでき、ファイル転送の手間を大幅に削減できる。また異なる計算機間でのデータ共有が進むことで、複数の計算資源を組み合わせた解析や他機関との協働も円滑になる利点がある。

もともと、共用ストレージの活用にも解決すべき課題は存在する。第一に性能面の課題である。共用ストレージは汎用性・広域接続性を重視するあまり、各計算機に直結のローカルストレージと比べると I/O レイテンシやスループットで劣る。例えば HPCI 共用ストレージのように広域ネットワーク越しにアクセスする場合、各ジョブからの高頻度な入出力やメタデータ操作を行うには適さず、計算実行用のスクラッチ領域は引き続き各計算機ローカルのストレージで担う、といった使い分けが必要になる。また共用ストレージへの集中アクセスによるスケーラビリティの課題もある。さらに、「富岳」のストレージは 150PB であるのに対し、共用ストレージの容量は 45PB のため、そもそも容量的にデータ移行には利用できない。

これら課題はあるものの、データを計算機と切り離して管理するという思想は HPC の利便性・持続性を高める上で重要である。「富岳」からポスト富岳への移行においても、HPCI 共用ストレージは一つの受け皿として機能し得る。実際、「富岳」利用者は HPCI 共用ストレージをマウントしてデータを保存でき、移行期間中に計算資源が不足する場合でもデータ共有基盤として活用できる。

(2-3) データ転送・共有ツールの活用

ストレージの端境期問題に対処する第三のアプローチは、高速・高信頼なデータ転送ツールを導入し、ユーザによるデータ移行作業を支援・効率化することである。計算機の更新に伴いユーザが自身のデータを退避・コピーしなければならない場合でも、適切なツールを用いることで作業負担を大きく削減できる。

代表的なツールの一つが Globus である。Globus はもともとグリッドコンピューティングのために開発されたツール群であるが、現在は遠隔での高速・高信頼のデータ転送サービスとして世界的に用いられている。ユーザは Web ブラウザ上の単一インタフェースから複数拠点間で大容量データを高速かつ安全に転送・共有することができ

る。また、ユーザは専用クライアントをインストールする必要がなく、ブラウザ上で転送元・転送先を選択するだけで TB～PB 級のファイル群を自動再試行・チェックサム検証付きで移動できる。このように Globus は HPC コミュニティで事実上の標準的データ転送ツールとなりつつある。

Globus 以外にも、各種データ転送ソフトや、最近では HPC センターに S3 ストレージを導入して汎用のプロトコルでデータ共有や転送を実現する動きもある。たとえば国内の学術認証基盤 (GakuNin) と連携した研究データ管理サービスを利用し、ユーザがブラウザ経由で HPC システムから直接データをアップロード・共有できる環境整備も進んでいる。これらツールを導入する効果はユーザ自身が容易にデータを扱える利便性向上にある。専門的なコマンド操作なしに GUI でデータを整理・移送できれば、計算機リテラシーに差のある利用者間での作業格差も縮まる。

もっとも、これらツールを最大限活かすには HPC センター側のインフラ整備も必要である。高速データ転送には計算機から切り離された専用のデータ転送ノード (DTN) を設置し、十分な外部接続帯域を確保することが望ましい。実際、「富岳」では外部システムとのデータ転送用に 100Gbps のネットワークインタフェースを持つ専用のデータ転送ノードを整備しており、対外回線も 400Gbps の帯域を備えている。一方で、HPCI アクセスポイント東京 (AP 東京)、神戸 (AP 神戸) では利用者がストレージを持ち込むことで、アクセスポイントに備えた大容量の回線を利用してスパコンのデータをストレージにダンプし物理的に持ち運ぶことも可能である。このようなインフラとサービスの整備は移行期におけるデータ退避を円滑にするため非常に有用であり、新システムへの移行時にも継続・拡充して提供されることが望まれる。

総じて、データ転送・共有ツールの活用は端境期のユーザデータ移行リスクを低減する有効な手段である。よって、次世代計算基盤では、これらソフトウェア基盤への投資や運用体制の構築も、ハードウェア整備と並んで検討すべき重要事項といえる。

(3) ストレージ移行に伴う技術的課題の整理

上述のアプローチを講じる際、具体的に直面する技術的課題を改めて整理する。ストレージ移行期における主な技術課題は次のとおりである。

1) データ転送速度・所要時間

端境期内に膨大なデータを移行させるには、ネットワーク帯域とストレージ I/O 性能がボトルネックとなる。仮に 100Gbps (12.5GB/s) の回線を終始フル活用できた場合でも、1PB の転送に約 1 日を要する計算となる。実運用ではプロトコルオーバーヘッドや再送なども発生するため、数十 PB 規模のデータ移行には数週間～数ヶ月

単位の計画が必要になる。端境期を短く抑えるには、並列転送の工夫（複数ストリームや複数ノードでの同時コピー）や事前の段階的データ退避（古いデータから順にあらかじめ共有ストレージやテープに移しておく）などの対策が求められる。

2) 異機種ストレージ間の互換性

新旧でファイルシステムの種類が異なる場合の互換性も課題である。例えば旧システムで Lustre を用いており新システムで別のファイルシステムを採用する場合、計算機側で両方のストレージをマウントできるのか、マウントできない場合は NFS マウントを介して利用するなど互換性の課題がある。NFS を用いる場合はストレージの性能が十分に出せない課題もある。

3) アクセス権・認証の移行

旧システム上のユーザ ID やグループ ID と新システムの ID 体系が一致しない場合、ファイルのアクセス権設定が引き継がれず移行後に権限エラーが発生する可能性がある。このため先述したようにユーザ管理基盤の統一が望ましいが、難しい場合は移行時に一旦すべて管理者権限でコピーし直し、新システム側で改めて正しいファイル所有者に戻す、といった運用上の暫定措置も検討される。

4) データの整合性

大量データ移行では、一部のファイル破損や抜け漏れが生じても気付きにくい。これを防ぐため、チェックサム検証やログ比較による厳密な整合性チェックが欠かせない。Globus のようなツールには自動チェックサム機能があるが、それ以外にも移行前後でファイル数・ディレクトリ構造をリスト化して照合するなどの多層的な検証を行うことが望ましい。

以上の課題に対しては、HPC 運用担当者とユーザの綿密な協力のもと計画的に対応することが求められる。特に「富岳」のような大規模システムでは、移行対象データ量も飛躍的に多いため、早めの準備と段階的移行が重要になる。

(4) ストレージの技術動向

ストレージ端境期問題を考える上では、関連する技術動向も把握しておく必要がある。近年、HPC 分野のストレージ技術は従来の並列ファイルシステム一辺倒から多様化しつつあり、オブジェクトストレージやクラウド連携、AI ワークロード対応など新たな潮流が生まれている。本節では、それらの代表的なトレンドを整理する。

(a) オブジェクトストレージ

オブジェクトストレージは、ファイルを階層構造で管理するのではなくオブジェクトとしてフラットに扱い、付随メタデータとともに格納するストレージ方式である。もともとクラウドサービス（Amazon S3）で一般化した技術であり、近年は HPC においても大規模な非構造化データの保存や共有に適している点から導入が進み始めている。オブジェクトストレージはスケーラビリティに優れ、PB～EB 級まで容量をシームレスに拡張しやすいこと、また HTTP ベースの標準 API である S3 インタフェースを通じ外部アプリケーションからもアクセスしやすいことが利点である。一部の HPC センターではオブジェクトストレージを導入しており、従来の並列ファイルシステムによる高速ストレージ環境と実現すると同時に、S3 API で外部サービスとデータ連携できる柔軟性を獲得している。

オブジェクトストレージ活用が端境期問題に示唆するのは、データフォーマットの標準化と可搬性向上である。S3 API は事実上データ公開・共有のデファクト標準となりつつあり、多くの分析ミドルウェアや科学アプリケーションが直接 S3 アクセスに対応し始めている。仮に HPC システムの更新時にファイルシステムが変更になっても、オブジェクトストレージ上に保存したデータは HTTP 経由で継続利用できるため、ユーザ側の移行作業を大幅に軽減できる。もともとスループット・レイテンシや細かな更新処理は従来の POSIX ファイルシステムに劣るため、HPC では高速ファイルストレージとオブジェクトストレージのハイブリッド運用が現実的である。このようにオブジェクトストレージは計算資源と独立したデータプールを構築する手段として有望であり、今後の HPC ストレージ計画に組み込む価値が高い。

(b) AI ワークロード向けストレージ

近年の HPC システムは科学技術計算に加え AI/機械学習ワークロードを統合的に処理する方向に進んでおり、それに伴いストレージにも新たな要求が生まれている。AI の学習フェーズでは数百万～数億点に及ぶ小さなファイル（画像やテキスト）を高速に読み込む必要があり、従来の大規模データとは異なる I/O パターンが発生する。これに応えるため、ストレージベンダ各社はオールフラッシュストレージや分散メタデータ処理等によりメタデータ性能と小ランダム読み出し性能を飛躍的に高めた製品を投入し始めている。また、NVMe over Fabrics や NVMe キャッシュを活用した階層ストレージによって、学習に必要なデータを高速な NVMe から供給し、コールドデータは大容量ディスク層に退避するといったマルチティア構成も登場している。さらに、NVIDIA 社の GPUDirect Storage 技術など、ストレージから GPU メモリへデータを直接転送することも可能になりつつある。これらの技術は HPC と AI の融合を背景に発

展しているが、結果として次世代の HPC ストレージ全体の性能底上げにつながっている。一方で、このような新しいストレージは、旧計算機からアクセスする際の互換性に注意する必要がある。

(5) 運用モデルのコスト比較

ストレージの端境期対策を検討する上では、コスト最適化の視点も欠かせない。巨大ストレージ基盤の整備には多額の設備投資と運用経費が伴うため、限られた予算で最大限の効果を実現することが重要である。本項では、いくつかの調達・運用モデルを比較し、コストと効果のバランスを考察する。

(1) 計算機・ストレージ一体更新（従来型）モデル

計算機の更新時にストレージも同時に更新するモデルは、技術的には最新の性能を一度に揃えられるメリットがある。調達コスト面でも一括調達による経費削減や、計算機とストレージを一体としてシステム全体の最適設計が期待できる。しかし、端境期問題への対策としては最も弱く、データ移行コスト（人的リソース含む）が別途発生する。さらに一般的に計算機は運用開始からフル稼働が期待できるが、ストレージは徐々にデータが増えて利用率も年々向上するといった傾向があるため、投資効率の観点から改善の余地も検討できる。このような、移行コストや投資効率のような隠れたコスト（ユーザの移行作業時間や研究停滞による機会損失）も含めて評価すれば、必ずしも一体更新モデルが安価とは言い切れない。むしろストレージ共用化や複数世代間運用によって端境期を無くすことが、長期的に見て研究生産性の向上とトータルコスト削減につながる可能性がある。

(2) 計算機・ストレージ別更新モデル

このモデルは、計算機とストレージを一体調達・更新せず、それぞれ独立に更新サイクルを設計する方式である。このモデルの最大の利点は、計算機の更新に際して既存ストレージ基盤を継続利用できるため、端境期におけるデータ移行・退避作業が不要となる点である。新システムは既存ストレージにアクセスしながら段階的にデータを取り込みつつ、旧システムの計算資源停止後もデータをそのまま利用できる。ユーザにとってはデータ移行の負担が大きく軽減され、継続的な研究活動が確保される。

一方、設計・運用面での課題としては、計算機更新時に旧ストレージとのインタフェース互換性を確保する必要がある。例えば、次期計算機が新しいストレージプロトコルやインタフェースにしか対応していない場合、旧ストレージがそれに追従できず本モデルは適用できない。また、ストレージへのユーザやアクセス権限を新旧

システム間で統一する必要もある。そのため、ストレージの更新時には将来の計算機更新も見据えた冗長性、拡張性、汎用性の高い設計が求められる。

また、コスト面では計算機とストレージとの一括調達によるスケールメリットは得られにくくなる可能性があるものの、更新タイミングを分けることで CAPEX を分散できる利点もある。

このように、計算機とストレージを別々のライフサイクルで更新する方式は、柔軟性と移行リスク低減に優れた運用モデルであると考えられる。HPC システムの更新計画においては、従来の一体更新モデルだけでなく、本モデルのような分離更新方策も併せて検討することで、費用対効果とサービス継続性のバランスを取ることが可能となる。

(3) 共用ストレージ・クラウド活用モデル

HPCI 共用ストレージやクラウドストレージを活用する本モデルでは、各 HPC センターが個別に大容量ストレージを抱え込む必要性が軽減されるため、全体最適の観点からストレージコストの削減が期待される。例えば、HPCI 全体としてのストレージ総容量を一定範囲で集約・共有することで、個別センターごとの余剰リソース確保を抑制し、投資の重複を回避することが可能となる。

しかしながら、HPC ジョブの実行時には、計算ノードとストレージ間の高スループット・低レイテンシ通信が求められることから、実運用においては各センターのローカルストレージに計算データが存在していることが望ましい。このため、共用ストレージを利用する場合でも、ジョブ実行前後に共用ストレージとローカルストレージ間のデータ転送が発生する運用（ステージング）となる。現行の学術情報ネットワーク SINET6 では 400Gbps 級の帯域が提供されているが、各 HPC センターが同時にジョブを走らせる状況を想定すると、データ転送の集中による帯域ボトルネックや性能劣化の懸念が残る。

したがって、本モデルを実運用に適用する場合には、各センターが一定規模のローカルストレージを保持する前提のもと、複数センターで共有すべきデータや長期保管データなど、一部の用途に特化して共用ストレージを活用する形が現実的である。これはまさに現在の HPCI 共用ストレージの利用方法である。

またクラウドストレージの活用についても同様に、オンプレミス的高速ストレージと併用する形で設計すべきであり、主に一時的なバックアップやプロジェクト終了後の成果データ保存先としての利用が現実的と考えられる。

(6) 将来のストレージ運用の理想像

これまでの調査結果を踏まえて、ユーザが計算機の更新を意識せず継続利用できる理想的な運用像を描く。本運用像は、本調査が目的とする次世代計算基盤のみならず将来の各センターのシステム更新にも適用可能であると考ええる。

理想像では、HPC システムの更新に際して計算サービスとデータサービスがシームレスに継続される。具体的には、旧計算機と新計算機が一定期間並行稼働し、ユーザのジョブは両者にまたがってスケジューリングされる。ストレージに関しては旧計算機・新計算機で共通の論理空間が提供される。つまり、新ストレージ・旧ストレージの両方も旧計算機・新計算機からアクセスできる。新旧計算機間でデータコピーをする必要は利用者には生じず、仮にハードウェア的なデータ移設が必要な場合でも運用側でバックグラウンド実施される。統一的な認証基盤によりファイルのパーミッションや属性も維持され、移行後にスクリプトやワークフローなどを修正する必要はない。

このような運用を実現するには、本節で議論した数々の方策の組み合わせが必要となる。まずは新旧の両システムから使えるような仕様のストレージの整備である。要は、計算機更新＝大規模な環境変化、という従来の慣習を打破し、徐々に、途切れなく HPC システムを進化させ続けることが求められる。こうしたユーザ本位の運用を積み重ねていくことで、将来的には「計算もデータも常に使える」真に持続的な HPC 基盤が構築されていくと考えられる。

(7) ストレージの端境期対応まとめ

本節では、次世代計算基盤への移行に関連して懸念されるストレージの端境期問題と、その解決策となり得る方策について包括的に検討した。端境期問題の背景には、計算機とストレージ更新サイクルの不一致や大容量データ移行の困難さがあり、それに起因してユーザの研究活動中断や計算資源活用の低下といったリスクが存在する。特に「富岳」のような大規模なシステムでは、前世代からの円滑なデータ継承が、利用者の利便性と計算科学力の維持に直結するため、対策の優先度は高い。

解決策として、ストレージの別調達方策や汎用化を柱に議論した。また、ソフトウェア面では Globus などのツール導入は比較的低コストで今からでも実行可能な対策であり、実例からもその有用性が確認できた。また、端境期問題に付随して近年のオブジェクトストレージやクラウド、AI 対応といったストレージ技術に関する概観した。

2.7.4 センター間連携に向けた調査、今後の導入計画の調査

HPCI を構成する National Infrastructure System (NIS、いわゆる第 2 階層システム)のうち、9 大学の基盤センター群について、各センターの運用状況、消費電力、利

利用負担金等、相互利用の可能性を検討するために調査を実施した（Society5.0グループと連携）。加えて、今後の各センターの導入計画についても調査した。

調査は以下の項目について行った。

- カーボンフットプリント
- 冷却方式
- 電気代推移（2022年）
- 利用負担金によるコスト回収モデル
- 将来の導入計画（期間、性能目標、電力）

カーボンフットプリントについては、2022年度におけるCO₂排出係数(kg-CO₂/kWh)が、北海道電力：0.535、東北電力：0.460、東京電力：0.376、中部電力：0.440、関西電力：0.420、九州電力：0.462となっており、地域による差が大きくなっている。年度によっても大きく変動するため一概には言えないが、管内の再エネ発電設備の構成割合や、再エネ証書の購入量などが関係しているものと思われる。

冷却方式については、チラー+冷却塔（一体型含む）による水冷主体が4拠点、チラーのみによる水冷主体が4拠点、空冷のみが1拠点であった。ただし、2024年度には冷却塔を使用する拠点は減り、3拠点になった。また、一部のシステムに外気冷却を使用している拠点も2拠点あった。

各センターが設定している利用負担金は、多くは、電気代+水道代などのランニングコストをベースに算出していることが多い。一部のセンターでは、一部の利用プログラム、特に民間企業向けの利用プログラムに対して、運用に携わる人件費や、設備更新のコスト等も加算している場合がある。センター間での連携や、リソース融通などを検討する上では、これらの利用負担金の考え方を踏まえて議論する必要があると考えられる。また利用ユーザに対しても、利用負担金の使途を明らかにし透明性を確保する必要がある。

将来の導入計画については、昨今の電力コスト増大、スパコン構成部品自体の高騰、円安などの外的要因が続いており、加えて各センターでは、センター全体の予算が以前から変わらないか減少する中で、ソフトウェアライセンス等スパコン以外の部分への支出も増加しており、各センターともシステム導入計画について苦慮している。現行機種の利用期間延長を表明しているセンターも多く、また、次世代システムの運用期間もシステム予算確保の観点から従来システムより長期間にする（6年程度）というセンターも増えている。とは言うものの製品供給・保守の観点から6年以上の運用は現実的ではない。従来はシステム更新のたびに性能が数倍に向上することが期待できたが、近年は、システム更新をすると逆に性能が下がるケースも出始めてい

る。GPU 主体に舵を切ること一度は大幅な性能向上を見込めるものの、その後は予算が一定であれば性能向上を見込むことが困難である。また、計算ノード当たりの単価が高くなるため、ユーザの需要に対して十分な計算ノード数を確保することが困難になっている。

2025 年度の 9 センターの理論ピーク性能合計値は 280 PFLOPS に留まっており、これらを踏まえると、大学基盤センター群の合計値で富岳と同等の 500 PFLOPS を達成するのは、現状のままでは 2030 年でも困難、と言わざるを得ない。

2.7.5 HPCI の機能拡充に向けて

今年度は、次世代の HPCI のあり方を検討する上での参考として、昨年度の欧州に続いて、米国における最新の状況について調査した。以下に簡単にまとめる。

米国政府は、科学技術研究開発 (R&D) 活動の多様な要求に対応するため、強力な国内の HPC エコシステムを強化・支援する長期的な戦略を確立している。過去 50 年間にわたり、市販されている HPC 関連技術を単に購入するのではなく、HPC 関連技術の直接 R&D 資金提供、技術パートナーシップ、労働力開発プログラム、科学コミュニティとの協力、先端技術の探求などの政策を実施してきた。また、輸出管理規制、貿易・関税制度、国内 HPC サプライチェーンへの財政支援、国際的な政策調整などの技術的ではないが HPC に関連する政策を通じた活動の支援も行っている。

米国政府の HPC ミッションの重要な目標は、国内 HPC サプライヤーを保護し、国家安全保障や経済的な課題に対応するための性能要件を満たすことである。新技術、国家安全保障の課題、商業的な動向に応じて、新たな HPC 関連政策が実現されてきた。将来の戦略としては、量子システム、AI、重要なサプライチェーン問題、ハイパースケーラーの役割、ポストムーアの法則の半導体生産能力などの新たな技術的課題に対応していこうと考えられている。

主な HPC に関連した組織は、以下の通りである。

(a) エネルギー省 (DOE) :

Advanced Scientific Computing Research (ASCR) プログラム: 応用数学とコンピュータサイエンスの進展を目指し、最先端の計算科学アプリケーションを提供する。主要な HPC サイトには、Frontier システムをホストする Oak Ridge Leadership

Computing Facility (OLCF) や、Aurora をホストする Argonne Leadership Computing Facility (ALCF) が含まれる。2024 年度の予算要求総額は約 11 億ドルである。

National Nuclear Security Agency (NNSA) : 核兵器の開発、管理、非拡散活動を担当する。主要な HPC サイトには、Los Alamos National Lab. (LANL) のペタスケール HPC “Venado”、Lawrence Livermore National Lab. (LLNL) のエクサスケールマシン “El Capitan”、Sandia National Lab. (SNL) をはじめ、合わせて 14 PFLOPS の性能を持つ 13 の HPC クラスタが含まれる。2024 年度の予算要求総額は 240 億ドル以上である。

(b) **国防総省 (DOD)** :

HPC Modernization Program (HPCMP) : 科学者やエンジニアに計算資源を提供し、ミッション関連の技術的課題を解決する。主要な DOD スーパーコンピューティング研究センター (SRC) は、陸軍、海軍、空軍のミッション要件に対応する。FY2024 の予算要求は 2 億 5500 万ドル。

(c) **National Science Foundation (NSF)** :

Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) : 研究者や教育者が先進的な計算システムとサービスを利用できるよう支援する。予算は 5 年間で 5200 万ドル。

NSF は Leadership-Class Computing Facilities (LCCF) として、Texas Advanced Computing Center (TACC)、それ以外にも Pittsburgh Supercomputer Center (PSC)、National Center for Supercomputing Applications (NCSA)、San Diego Supercomputer Center (SDSC) などを支援している。

2023 年度、2024 年度の調査結果も踏まえつつ、今後の機能拡充に向け、これまでの HPCI システムの運用を評価することに加え、具体的には以下の 5 点について検討が必要と考える。

(a) HPCI 計算資源の整備

HPCI 計算資源の整備については、フラッグシップシステムについては、開発プロジェクトが立ち上げられ、その中において詳細な検討が行われる一方で、国立大学の基盤センター群を中心とする National Infrastructure System (NIS, いわゆる第 2 階

層システム)については、大まかな整備計画の共有はあるにせよ、各資源提供機関が独自の方針に基づいて整備するケースがほとんどであった。結果的に、HPCI システム全体として見ると、機能や特徴が重複するシステムが存在することになり、全体的に最適化がなされているとは言い難い。これは、システム整備の財源が、フラッグシップシステムについては整備補助金、NIS については各資源提供機関の運営費交付金であることに起因するもので、すぐに改善することは困難であろう。

一方で、筑波大学と東京大学とが連携して整備・運用した Oakforest-PACS システムや、北海道大学、東北大学、筑波大学、東京大学、東京工業大学（現・東京科学大学）、名古屋大学、京都大学、大阪大学、九州大学、NII、産総研が連携して整備した mdx システムは、機関間連携によって財政的な困難を克服しつつ、先進的な技術を積極的に取り込んだ野心的な試みであり、今後もこのような取り組みは国として積極的に支援していくべきであろう。

そのためには、フラッグシップと NIS の中間に位置するようなシステム整備のためのプロジェクト公募を戦略的に実施することが効果的と考える。すなわち、今後 10 年程度において、各システム間での機能や役割の分担を考慮しつつ、いつどのようなシステムを整備するかといった大まかなロードマップを国として策定・共有し、それに基づいて整備する個別のシステムについては、都度開発・整備プロジェクトを公募し、解決すべきテーマや必要となる技術的な要素を明確にするとともに、機関間の連携を推奨する。これにより、機関間の連携協力を促進するとともに、我が国全体の計算資源配備の最適化・効率化が実現できる。

また、このような施策に必要な財源については、整備費用の一部を整備補助金で措置することも検討されるべきである。残りの部分については各資源提供機関の負担で実施することになるが、民間や地方自治体などからの出資も受け入れ可能であることが望ましい。それにより、システム開発のフェーズから産業利用、地域振興の視点がより反映されやすくなり、システム整備後の産業利用や社会基盤整備の促進にも極めて有効であろう。すでに触れたように、EuroHPC JU の場合、競争を経て採択された導入計画に対しては、整備費用および運用にかかる費用の半額を補助（残りの費用は設置機関が調達）している。これは、より挑戦的で、新規性の高い提案をエンカレッジすると共に、設置国政府や地方自治体、さらには民間からの資金拠出を促進する効果がある。

加えて、スパコン調達の制度をより柔軟にすべきである。具体的には、企画競争を許容することを可能にする必要がある。現在のスパコン調達の制度は、市販製品を指定の納期に合わせて設置する、ということしか許されておらず、納入期限・検収条件

を満たせないと最初からやり直しを求められる。サプライチェーン問題が顕在化している現在、規模、納期、予想金額、全てにおいてマージンを取る必要があり、保守的な提案に終わる上に、極めて非効率であり、調達作業担当者にとっての負担も多大である。米国、欧州いずれのスパコン調達も、企画競争が基本となっており、日本の調達制度は極めて不利である。一方、mdx-1においては、サプライチェーン問題、セキュリティリスクが顕在化し、企画競争によって実施した経緯がある。当然ながらWTOの求める調達ルールにも合致しており、現状のスパコン調達ルールが時代錯誤であり不当に厳しいと言わざるを得ない。

(b) HPCI 計算資源の運用

HPCI の各提供資源の運用は、原則として資源提供機関の責任で実施されているが、運用に関する機関間の連携は、AXIES(大学 ICT 推進協議会)、情報処理学会 HPC 研究会、個別の機関間連携（東京大学、名古屋大学、JAXA、富士通、R-CCS が参加する、富岳に類似するシステムの運用に関する検討会など）など、極めて限定的である一方で、日々の運用を支え、高度化を担う人材は、民間を含め、常に払底しており、安定的かつ最新の計算サービスの提供を維持するためには、HPCI 全体としての運用体制の改善が喫緊の課題である。

多くの資源提供機関は、少数のプロパーのスタッフに加えて、システムを開発・製造したベンダが提供する保守サービスを受けているケースがほとんどであるが、民間側も人材の確保に苦慮している状況もあり、現場の感覚としては保守サービスのレベル低下が急速に進行していることは否めず、今後ベンダに依存した保守体制は維持できなくなる可能性が高いと考えている。一方で、米国の HPC センターのように、保守作業を内製で行うための体制を今から構築することも現実的ではない。

一案として、民間も含めた運用・保守の人材をひとつの組織に集約し、HPCI の各計算資源の保守を一元的に行うような体制づくりを検討すべきと考える。HPC システムの運用は、システムそれ自体だけではなく、冷却設備、電源設備など、多岐に渡り、省エネルギーや低炭素排出等の観点での効率的な運用に関する高度な知見・ノウハウの蓄積をひとつの機関で行うことは不可能になってきている。多数の HPC システムや関連設備を並行して運用することで、運用技術そのものの高度化が促進されるとともに、運用を担う人材獲得と育成も格段にやりやすくなると期待できる。そのような運用を実施しやすくするために、仮想化の機能等についてシステムの調達仕様の中に盛り込むように促すことも効果的であろう。

(c) HPCI 計算資源の高度化

HPCI システムによる成果創出を促進するには、より使いやすく、より高度な機能を継続して開発し、提供していくことが不可欠である。

そのためには、HPCI システムを透過的に利用するための統一的な認証・認可機能、ソフトウェア利用環境の共通化、クラウド機能の整備やアプリケーションのプラットフォーム化も必要である。バーチャル富岳プロジェクトの成果等も活用しながら、着実に進めることが肝要である。加えて、HPCI システム全体を統括するメタスケジューラの導入による HPCI 全体の利用の最適化・効率化、利便性向上を目指すべきであろう。また、統一的な利用サポートの提供も極めて重要である。利用者から問い合わせが来てから対応するのではなく、利用者が知りたい情報に到達しやすくすることで自己解決を促進する対応を志向すべきである。そのために生成 AI の活用は、利便性の向上と運用側の負担軽減に極めて有効であると思われる。富岳関連ではいち早くヘルプデスクに生成 AI を利用しはじめており、これを HPCI 全体に拡大していくことから始めていくべきである。

(d) 国プロで開発したアプリケーションの保守および高度化

これまでも国プロで開発されたアプリケーションが、プロジェクトが終了した後に保守体制が維持できず、ソフトウェアとしての価値が失われるケースが多々あった。現在は RIST が中心となって、HPCI システムに国プロアプリを整備する取り組みがあるが、アプリをインストールするのみで、利用サポートや計算サービスとしての付加価値を高めるという観点では不十分であると言わざるを得ない。アプリケーションの高度化などの開発タスクについては研究プロジェクトで措置されるべきだが、その後のバグフィックスなどの保守、新アーキテクチャへのポーティング、チューニング、利用者対応、計算プラットフォーム化、マネタイズなどについては、開発者とは別の者が長期的に担うべきである。結果的にそれがソフトウェアの価値を高めることになり、間接的に HPCI システムの価値も高め、成果創出の促進につながる。

(e) 日本の研究力を支える計算基盤として

本調査研究の当初目標の一つに掲げられていた「日本の研究力向上に資する研究 DX を支える計算基盤」に向けては、HPCI の機能の拡充および認知度の向上を図っていく

必要がある。例えば、昨今、研究データマネジメントのための費用については、プロジェクト予算への計上が義務化されるような議論もされているが、計算基盤についても同様の議論があって然るべきではないかと考える。大型のプロジェクトごとに個別のクラスタが導入されていくことになるのは、国全体で考えると無駄が多い。計算資源を HPCI に集中させることで、今後の国全体の脱炭素化に向けた取り組みにも有効であろう。一方で、当然、システムそのものの研究や、セキュリティの観点では一定のオンプレミス環境は必要だろう。

将来的には、例えば、研究データマネジメントと連動して、最初からある程度の計算資源を利用可能にするような制度も考えられる。一方、計算資源を集約する代わりに、さまざまなタイプの計算基盤を整備する必要も生じると考えられる。さらに学術ユーザの爆発的な利用増加も見込まれるため、次世代「学認」の利用が前提になると考える。

(f) 上記を実現するための体制

ここまですべてをまとめると、以下の機能の強化が必要であると考えられる。

- ① HPCI システムの開発・設置に関するプロジェクトの公募をマネジメントし、HPCI システムの全体最適化を促進する機能
- ② HPCI システムの運用を集約し、安定的なサービス提供と高度化に資する機能
- ③ 国家プロジェクトで開発されたアプリケーションの保守および高度化を推進する機能
- ④ 日本の研究力を支える計算基盤としての機能

これまでも HPCI の高度化に向けた取り組みは進められて来てはいるが、スピード感の不足や責任の所在が曖昧なことから、十分な進捗には程遠いと感じている。組織を一元化し、責任を付与することで、HPCI の高度化に向けたさまざまな改善の取り組みが促進されると期待する。

一方で、上記の機能は多岐に渡っており、ひとつの者がすべてを担うことは稀であろう。組織間の連携を密にすることは重要だが、それぞれに強みを持つ複数の者でこれらの機能を分担するのが現実的ではないか。

特に欧州における EuroHPC JU の取り組みには学ぶべきところが多い。このような先行事例をさらに詳細に分析し、我が国における制度の中に、これらの機能をどのように実装していくのか、抜本的な見直しも含め、踏み込んで検討すべきである。

2024 年度には、HPCI コンソーシアムの下に「HPCI システムの今後の在り方に関する調査検討ワーキンググループ」、文部科学省 HPCI 計画推進委員会の下には「次世

代計算基盤を見据えた今後の HPCI の運営に係る検討ワーキンググループ」が設置され、今後の HPCI のあり方に向けた包括的な議論が進んでいる。本調査研究の成果も取り入れて改善を進めていきたい。

様式第 2 1

学 会 等 発 表 実 績

委託業務題目「次世代計算基盤に係る調査研究」（運用技術調査研究）

機関名 国立大学法人 東京大学

1. 学会等における口頭・ポスター発表

発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
Feasibility Study for Next-Generation Computing Infrastructure: Study of Operation Technologies（ポスター）	Toshihiro Hanawa, Keiji Yamamoto, Shin'ichi Miura, Akihiro Nomura, Atsuko Takefusa, Satoshi Ohshima	ISC-HPC 2024 (project poster)	2024年5月	国外
Update for Feasibility Study on Operational Technology Towards Next-Generation Computing Platform in Japan（口頭）	Toshihiro Hanawa	14 th European Workshop on HPC Infrastructure	2024年5月	国外
Operation Survey Study for Fugaku-next - Carbon Neutral & Energy Efficient HPC Data Center（口頭）	Shin'ichi Miura	14 th European Workshop on HPC Infrastructure	2024年5月	国外

Long Road to Miyabi: New Supercomputer on JCAHPC introducing GH200 (口頭)	Toshihiro Hanawa	14 th European Workshop on HPC Infrastructure	2024年5月	国外
TSUBAME4.0 Supercomputer: Introduction of System and Failures during the installation (口頭)	Akihiro Nomura	14 th European Workshop on HPC Infrastructure	2024年5月	国外
Long Road to Miyabi: New Supercomputer on JCAHPC introducing GH200 (口頭)	Toshihiro Hanawa	Japan HPC Infrastructure Workshop	2024年9月	国内
TSUBAME 4.0 Supercomputer: Introduction of System and Node Partitioning Strategies (口頭)	Akihiro Nomura	Japan HPC Infrastructure Workshop	2024年9月	国内
Green500 Level3 Measurement on Miyabi (口頭)	Toshihiro Hanawa	Green500 BoF, Supercomputing2024	2024年11 月	国外
次世代計算基盤におけ る利用環境の高度化に 向けて (口頭)	埴 敏博	第8回HPCものづくり 統合ワークショップ	2024年12 月	国内
最先端共同HPC基盤施設の新 スパコンシステムMiyabiの 運用開始に向けて / 東大情 報基盤センターの目指す 『計算・データ・学習』と データ利活用を融合した革 新的なスーパーコンピュー ティング (口頭)	埴 敏博	PCクラスタコンソー シウム第24回シンポ ジウム	2024年12 月	国内

富岳NEXTに向けたカーボンニュートラル化への取り組み（口頭）	山本 啓二	PCクラスタコンソーシアム第24回シンポジウム	2024年12月	国内
運用技術調査研究の概要と検討結果	埜 敏博	次世代計算基盤に係る調査研究に関する合同ワークショップ～フィージビリティスタディ結果報告～	2024年12月	国内
TSUBAMEシリーズにおけるブラウザベース利用機能の導入	野村 哲弘	PCクラスタコンソーシアム 第2回 Open OnDemandワークショップ	2025年1月	国内
ポスト富岳時代のHPCIに向けて（口頭）	埜 敏博	「次世代計算基盤のユーザビリティに関する提言」についての意見交換会	2025年2月	国内
Towards Efficient and Advanced Operation of Next-Generation Computing Infrastructure in Japan（ポスター）	Toshihiro Hanawa, Keiji Yamamoto, Shin'ichi Miura, Akihiro Nomura, Atsuko Takefusa, Satoshi Ohshima	SuperComputing Asia 2025	2025年3月	国外
Design and Operation of the TSUBAME4.0 Supercomputer for a Broader User Base（口頭）	Akihiro Nomura	ADAC16 Symposium	2025年3月	国外

2. 学会誌・雑誌等における論文掲載

掲載した論文 (発表題目)	発表者氏名	発表した場所 (学会誌・雑誌 等名)	発表した 時期	国内・ 外の別
次世代計算基盤 の資源管理に関 する調査研究の 中間報告	佐賀 一繁, 竹房 あつ 子, 合田 憲人, 高倉 弘 喜, 栗本 崇, 坂根 栄 作, 藤原 一毅, 田中 秀 樹, 大島 聡史, 山本 啓 二, 埴 敏博	情報処理学会研 究報告ハイパフ ォーマンスコン ピューティング (HPC), Vol. 2024-HPC-194, No. 9, p. 1-8	2024年5 月	国内
GH200の予備性能 評価	埴 敏博, 建部 修見, 中 島 研吾, 朴 泰祐, 三木 洋平, 下川辺 隆史, 山 崎 一哉, 住元 真司, 高 橋 大介, 額田 彰, 藤田 典久, 小林 諒平, 多田 野 寛人, 田浦 健次朗, 細川 颯介, 高橋 淳一 郎, 成瀬 彰	2024年並列/分 散/協調処理に 関するサマー・ ワークショップ (SWoPP 2024), Vol. 2024-HPC-195, No. 4, pp. 1- 11	2024年8月	国内
スパコンTSUBAME シリーズにおけ るリソース分割 戦略	野村 哲弘, 遠藤 敏夫	2024年並列/分 散/協調処理に 関するサマー・ ワークショップ (SWoPP 2024), Vol. 2024-HPC-195, No. 7, pp. 1-7	2024年8月	国内

(注) 発表者氏名は、連名による発表の場合には、筆頭者を先頭にして全員を記載すること。