

## 理化学研究所の計算資源利用について

理化学研究所

最先端研究プラットフォーム連携 (TRIP) 事業本部  
科学研究基盤モデル開発プログラム (AGIS)  
／計算科学研究センター (R-CCS)  
／TRIP 事業推進部／計算科学研究推進部

### 0. はじめに

科学技術振興機構 (JST) にて公募される AI for Science 革新的研究推進事業 (ARiSE) において、理化学研究所の計算資源を一部利用することができます。利用に際して必要となる経費は ARiSE 事業の研究費から支出可能です。計算資源の利用にあたっては、必ず本紙および各 Web ページ (本紙内にリンクを記載) を確認の上、利用計画を立案ください。なお、以下は現時点のものであり、今後変更される可能性があります。また、ユーザからの要望に応じて必要なソフトウェアの導入も検討しております。さらに、計算資源の利用に際しては、理化学研究所が定めるデータポリシー、アカウント申請手順、および利用規定を遵守してください。アカウントの取得・管理、データの取り扱い、ならびに計算資源の利用方法については、所定の手続きおよび規定に従う必要があります。なお、ARiSE 事業全体で利用される計算資源量の確保の観点から、理化学研究所の計算資源管理を担当する上記組織の関係者が提案書類の一部を確認することがあります。

本紙は JST ARiSE 事業向けの計算資源概要資料です。システムの詳細仕様・ソフトウェア一覧等は、別途公開の Web ページにて最新情報を提供しており、本紙からリンクしています。

最新情報は随時更新されていきますので、申請時必ず最新の情報をリンク先からご確認ください。

### 1. システム概要・提供形態

- **AI for Science 開発用スーパーコンピュータ (仮称: AI4S スパコン)**
  - 詳細: <https://github.com/RIKEN-RCCS/AI-for-Science-Supercomputer>
  - OpenOnDemand (Web フロントエンド) : ブラウザ経由でログインし、インタラクティブ利用・開発・可視化が可能なフロントエンド環境を整備

**【提供形態】**以下の通り、段階的に展開予定 (ただし、今後 ARiSE のスケジュールに沿うように一部変更を行う可能性があります)

- 先行提供フェーズ（試験運用）：ARiSE 事業公募期間中に、AI4S ワークロードの各種評価用に無料の試験アカウントを提供できるよう調整中です。申請方法等詳細は今後上記リンク先に掲載します。
- 本格運用フェーズ：段階的に一般提供へ移行予定
  - 先行提供フェーズから本格運用フェーズへ移行の際、ソフトウェア環境の改修が行われ、動作環境に変更が発生する可能性が高いことを予めご了承ください。また、AI4S クラスタは、富岳と異なり、アカウントの引き継ぎは保証されません。

※ARiSE 事業の研究遂行にあたっては、AI4S スパコンに加え、スーパーコンピュータ「富岳」の CPU 資源を併用することも可能です。既存の富岳のアカウントを持っていない場合は、富岳の「試行的課題」などへの申請をお勧めします(2-3 週間で受理、無料)

詳細: <https://github.com/RIKEN-RCCS/Fugaku/>

## 2. システム構成・性能 (富岳のスペックは上記の HP をご覧ください)

- 計算ノード：NVIDIA Grace Blackwell Superchip 搭載ノード 400 台 (GPU 総数：1,600 基)
- ノード間接続：NVIDIA InfiniBand XDR
- 理論性能：(FP64) 64.16 PFLOPS 以上、(FP8) 15.539 EFLOPS 以上

詳細（構成図・性能表・計算ノード詳細含む）は下記 Web ページを参照ください：

<https://github.com/RIKEN-RCCS/AI-for-Science-Supercomputer>

## 3. 利用形態

### • AI 推論・エージェントサービス向け利用（提供開始当初より段階的に提供）

大規模 AI 研究における推論、エージェント、ツール連携型ワークフローを支援するため、固定的に割り当てられた計算資源上で、利用者が独自のコンテナ、AI モデル、推論サービス、エージェント基盤等を展開・検証できる利用形態を提供します。また、商用 AI サービスからの移行やオープンソースモデルの活用を支援するため、共通的な推論サービス、AI コーディング支援、エージェント実行環境等についても順次整備します。これらのサービスは後述される RiVault 経由での利用を推奨します（現時点で、RiVault は理研 R-CCS ネットワーク内で試験運用中ではありますが、将来的にサービスを公開する予定です。）。

### • Kubernetes (K8s) 環境（試験的提供を準備中）

推論サービス、エージェント基盤、学習ジョブ等を柔軟かつ動的に管理するため、Kubernetes を用いた資源制御・サービス運用環境の試験的提供を準備しています。将来的には、Slurm によるバッチ実行環境も含め、Kubernetes と連携した動的な資源配分や、学習ジョブのサービス化、実行履歴・来歴管理等の機能を段階的に整備していく予定です。

### • Slurm によるバッチ／インタラクティブ利用（提供開始当初より）

提供開始当初は、Slurm を用いたバッチジョブおよびインタラクティブジョブの実行環境を提供します。主に、大規模 AI モデルの学習、シミュレーション連携、性能評価・ベンチマーク、FugakuNEXT 関連の研究開発等で

の利用を想定しています。なお、初期運用段階ではネットワークや運用上の制約を踏まえ、資源割当は原則としてノード単位で行います。また、Slurm は AI エージェントからの利用も推奨します。

- **上記の実行環境全て、本格稼働時には、Virtual Fugaku (<https://www.r-ccs.riken.jp/en/fugaku/virtual-fugaku/>) によって富岳と互換性の高い実行環境を提供し、フロントエンドも Open OnDemand (<https://www.r-ccs.riken.jp/outreach/topics/20230530-1/>) や RiVault/OpenWebUI が統合された環境での利用を実現します。**

#### 4. OS・ソフトウェア環境 (概要)

本システムでは、大規模 AI 研究に必要なソフトウェアスタックを体系的に整備し、以下の分野を幅広くカバーします。

- 基本ソフトウェア (OS: Rocky Linux 10, コンテナ: Apptainer / Podman, CUDA Toolkit) ・GPU 計算基盤ライブラリ (CUDA, cuDNN, NCCL, TensorRT 等) on Virtual Fugaku
- 深層学習フレームワーク (PyTorch, DeepSpeed, Megatron-LM, NeMo, JAX 等)
- LLM 推論サービス基盤 (vLLM, TensorRT-LLM, Triton, NVIDIA NIM 等)
- オープン AI モデル群 (Llama 4, Qwen, Nemotron, DeepSeek-V4 等)
- RAG・エージェント基盤 (RiVault, LangGraph, RagFlow, NeMo Guardrails 等)
- 科学計算向け AI ツール (サロゲートモデル, PINN, コード変換支援 等)

実際の提供は必要に応じて段階的に整備・拡充される予定です。

各ソフトウェアの詳細仕様・一覧は下記 Web ページを参照ください：

<https://github.com/RIKEN-RCCS/AI-for-Science-Supercomputer>

AI 推論ゲートウェイ RiVault の詳細は <https://github.com/RIKEN-RCCS/RiVault> を参照ください。

その他、Virtual Fugaku 環境を提供し Open OnDemand 経由での利用も予定しており、その利用を推奨します。

#### 5. 利用料金 (先行提供フェーズでは無料)

- 計算資源：300 円/GPU 時間 (税抜き ※別途消費税 10%)
- ストレージ：
  - グループ共有ストレージのみ課金対象 (※ ストレージ課金詳細は <https://github.com/RIKEN-RCCS/AI-for-Science-Supercomputer> (にて後日案内))
  - 個人ディレクトリは課題有効期間中無料

#### 6. ストレージ構成

- All NVMe 構成 Lustre (ES400X3-NDR200: 下記参照)

詳細は <https://github.com/RIKEN-RCCS/AI-for-Science-Supercomputer> 内「Storage」節を参照ください。

## 7. 本紙に関する問い合わせ先

- 理化学研究所 AI for Science 開発用スーパーコンピュータ サポート窓口
- E-mail: [rccs-ai4s-support@ml.riken.jp](mailto:rccs-ai4s-support@ml.riken.jp)
- 将来的に、富岳で実運用が開始されているサポートサイト([Zendesk](#)) や AI を活用したサポートシステム ([AskDona](#)) を AI4S スパコンでも導入することを計画しております。

以上