

大島 聡史（名古屋大学 情報基盤センター 准教授）
ohshima@cc.nagoya-u.ac.jp

最近の「不老」更新状況

- 2020年度から2021年度に変わる際に変更・更新したことや、その後のメンテナンス等で更新したことの中から、重要なこと・利用者に影響が大きいことを中心に幾つか紹介します
 - 基本的には「速報」やWebの「お知らせ」等でお伝えしていることですが、是非再確認して「不老」をフル活用してください
- 皆様への情報通知方法
 - 2021年の「お知らせ」一覧と各告知内容の概要
 - 課金ルールとリソースグループ構成の更新
 - 利用申請の電子化（試行）
 - 利用説明会や講習会の資料公開、紹介ビデオの公開
 - 利用ポイントを使い切った場合の対応
 - 富士通MPIと組み合わせて利用可能なGCCの公開
 - Type Iサブシステムのノード割り当て方法の変更
 - Type IIサブシステム向けMPIについて
 - その他

皆様への情報通知方法

- 主にWebとメールで情報を発信
- Web
 - 定期保守（メンテナンス）情報・アップデート情報・イベント情報など様々な情報を掲載
 - 定期保守の予定はページの上部
 - 最新情報は定期保守情報の次の「お知らせ」欄
 - ポイント残量（課金情報）やHPCポータル、講習会の案内などにもすぐにアクセス可能
 - コンパイラの更新や障害修正などについてはHPCポータルに詳細情報を掲載
- メール
 - 特に重要と思われる情報はメールでも配信
 - 月に1回程度、まとめてお知らせ（速報）を送信
 - 利用者以外に講習会やイベントに参加された方にも送信
- ○○がわかりにくい、などのご意見があればお知らせください

The screenshot shows a website navigation menu with several sections. Red dashed circles highlight the 'お知らせ' (Notice) and '利用方法' (Usage) sections. The 'お知らせ' section lists various events and updates, including a 2021 notice, AVS/Express seminars, and events. The '利用方法' section lists resources like account confirmation, manuals, and project archives. The '利用者支援' (User Support) section lists system reports, seminars, and consultation services. The '研究成果入力' (Research Results Input) section is also visible at the bottom.

2021年の「お知らせ」一覧（講習会などのイベント開催告知を除く）

- 2021.08.19 スーパーコンピュータへの接続不通について 8/19の時刻が変更になりました
- 2021.07.16 Type IIサブシステム利用者向けに、GPUと組み合わせて利用するMPIのmoduleとその性能に関する詳細資料を公開しました
- 2021.07.15 TypeI サブシステム FX1000ノード割り当て方法の変更について
- 2021.05.20 富士通MPIと組み合わせて利用可能なType Iサブシステム向けGCC公開のお知らせ
- 2021.04.27 ポイントを使い切ったあとのご利用について
- 2021.04.08 新型コロナウイルスを含む感染症対応HPCI臨時公募課題の募集について
- 2021.04.02 2020年度から継続して利用する際の注意点について
- 2021.04.02 スーパーコンピュータシステム「不老」の紹介ビデオを公開しました
- 2021.03.18 利用負担金一部改定について
- 2021.03.12 2020年3月末でファイルシステム（/uvdata）サービス終了しました
- 2021.02.02 「スーパーコンピュータ「不老」利用説明会（一般および企業利用）」の資料を公開しました
- 2021.01.29 【締切日延長】2021年度 名古屋大学HPC計算科学連携研究プロジェクト公募の締切日が2/7まで延長しました
- 2021.01.13 スーパーコンピュータ「不老」利用申請の電子化（試行）について
- 2021.01.07 クラウドシステムの混雑状況について(リソースグループの使い分けとUNCAIのご案内)

課金ルールとリソースグループ構成の更新

- 2020年度と2021年度では一部課金ルールが異なる
 - ログインノードのCPU利用によるポイント消費は廃止
 - ログインノードは共有の限られた資源であるため高負荷な作業はご遠慮ください
 - グループ利用の申し込み単位：20人まで20万ポイント20万円→10人まで10万ポイント10万円
 - コストパフォーマンスは一緒だが小規模の利用者にも使いやすくした
- リソースグループ構成は利用状況に合わせて少しずつ更新
 - 多くの利用者が短い待ち時間でジョブを実行できるように調整
 - 一部のリソースグループは休止・廃止
 - fx-extra, cx-extra, cl-extraは節電対応が不要なため休止
 - クラウドシステムはUNCAIより通常のバッチジョブ利用の方が圧倒的に多いため、大半のノードをバッチジョブ利用向けに設定。それでもUNCAIは空いているため、UNCAIで実行できる処理はUNCAIを使った方が早く実行開始できる可能性あり。

UNCAIの便利な使い方

- 10コア実行
 - バッチジョブの最小実行単位はcl-share (1ソケット実行、20コア)
 - UNCAIの最小実行単位はVSテンプレート (10コア)
 - コア数に応じて課金されるため、実行したい内容によってはUNCAIの方が経済的
 - コア数を増やしてもあまり性能が向上しない問題など
- 時刻指定バッチジョブ実行
 - 仮想マシンが立ち上がると同時に指定のスクリプトを実行してくれる機能
 - JupyterLabのサーバ自動起動などに有用
 - 使い方次第でバッチジョブ実行の代わりにも利用可能
 - 自動終了機能 (予約時間の終了やWebからの明示的な操作以外で仮想マシンを終了させる機能)がない点には注意が必要

JupyterLabの利用例（詳細はWebのFAQを参照）

- UNCAIからの起動の例（時刻指定バッチジョブ実行）
 - 以下のファイルを~/.batch/以下にstart_scriptから始まるファイル名で保存し、UNCAIの資源予約時に「時刻指定バッチスクリプト」から選択。仮想マシンが起動した後、ログファイルをみてtokenを確認し、SSHポート転送してブラウザから接続。

```
#!/bin/bash
module load gcc/8.4.0 python/3.9.2 jupyterlab/3.0.9
jupyter notebook --notebook-dir=${HOME} --ip='*' --port=8888 --no-browser --allow-root
```

- Type IIサブシステム、クラウドシステムのバッチジョブ、クラウドシステムの時刻指定インタラクティブ実行からの起動の例
 - 以下のバッチジョブを投入、開始後にログファイルを見てtokenを確認し、SSHポート転送してブラウザから接続。cx-shareやcl-shareのような「共有」ノードも利用可能。GPUも利用可能。

```
#!/bin/bash
#PJM -L rscgrp=cx-share
module load gcc/8.4.0 python/3.9.2 jupyterlab/3.0.9
jupyter notebook --notebook-dir=${HOME} --ip='*' --port=8888 --no-browser --allow-root
```

利用申請の電子化（試行）

- 名古屋大学に所属している方のみですが、ペーパーレスでの利用申請が可能です
 - 名大IDを使って本人確認などを行っているため現時点では名大関係者のみ
 - 学外の利用者への対応については現時点では未定

[ホーム](#) » [スーパーコンピュータシステム](#) » [利用手続き](#) » [スーパーコンピュータ「不老」新規利用電子申請（試行）](#)



スーパーコンピュータ「不老」新規利用電子申請（試行）

注意事項

- スーパーコンピュータ「不老」の電子利用申請（**新規利用**）は、名古屋大学所属の教職員及び学生限定となる
- 新規の一般利用は基本負担金として年額10,000円を、グループ利用は1口100,000円を徴収いたしますので、経理責任者の承認を事前に得て下さい。（[経理担当一覧表](#)）
- 各利用申請書を事前に記入して準備しておいて下さい（各押印は不要です）
 - ・ [一般利用申請書](#)
 - ・ [グループ利用申請書](#)
 - ・ [研究分野コード表](#)
- 提出するファイルはpdfファイルとしファイル名は「提出者.pdf」として下さい。
- 申請方法は、提出画面にて各項目入力後に利用申請書を入力画面に書かれているNUSSにアップロードして下さい。
- 提出画面の入力項目は以下の通りです。

利用説明会や講習会の資料公開、紹介ビデオの公開

- 当センターでは従来は講習会資料などをあまり積極的に配布してこなかった（らしい）
- 利便性向上のため資料の公開を進めています
 - 「講習会・説明会・その他イベント」のページを参照
 - 一部の資料は参加者と「不老」ユーザ限定（ホットストレージ上に資料が置いてある）
- 「不老」の紹介ビデオをYouTubeに掲載
 - 情報基盤センターのYouTubeチャンネルを開設しシステム構築ビデオやPV、バーチャルツアー動画などを掲載
 - 今後もイベント動画などを掲載していく予定

利用者支援

- システム速報
- 講習会・説明会・その他イベント
- IT相談室
- 問い合わせ先



利用ポイントを使い切った場合の対応

- 従来
 - ポイントを使い切った時点で新たにログインノードへログインすることができなくなる
 - 翌日以降にアカウントポータルにアクセスして課金情報を確認すればポイントを使い切ったことが確認可能
 - ファイルを引き上げたいなどの場合は個別に問い合わせ
- 現在
 - ポイントを使い切ってもログインノードへのsshログインおよびログインノード上での一般的な操作が可能
 - もちろんジョブの投入はできず待ち状態のジョブが実行開始されることはないが、すでに実行中のジョブはそのまま継続される
 - 残りポイントはアカウントポータルまたはchargeコマンドで確認可能
 - 従来と同様に前日までの状況のみ確認可能なため、使い切った当日の時点では使い切ったことを確認できない

富士通MPIと組み合わせ可能なType Iサブシステム向けGCCの公開

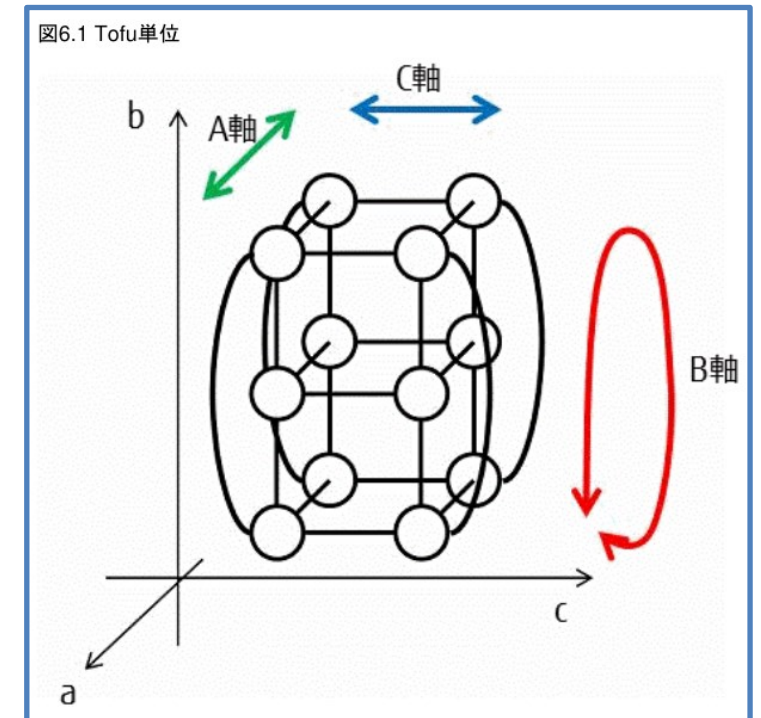
- 富士通コンパイラ向けに最適化されていないプログラムなどでは富士通コンパイラよりGCCを用いた方が良い性能が得られることもあるようだ
 - オープンソースソフトウェアなど
- 従来もType Iサブシステムで富士通製以外のコンパイラを使うことは可能であったが、富士通MPIと組み合わせる利用することができなかった
- 富士通MPIと組み合わせる利用可能なGCCの提供を始めたため、興味のある方は試してみてください
 - moduleをloadするだけで良い
 - ただしネイティブコンパイラのための提供のためログインノードでは利用できない（バッチジョブorインタラクティブジョブで利用する必要がある）

```
$ module load fjmpi-gcc/8.3.1
```

Type Iサブシステムのノード割り当て方法の変更

- ノード割り当て方法のデフォルト値をtorus（トーラス）からnoncont（離散）に変更
- Type Iサブシステムでは複数ノード利用時にノードの割り当て方針を指定できる
 - torusはTofu-Dの性能をフル活用できる期待度が一番高いがtorus形状に合わせたノードが確保できないとジョブが実行開始されず、12ノード単位に利用ノード数が切り上げられる（利用ポイントも切り上げられる）
 - 従来はtorusがデフォルト設定だったが、ジョブの実行待ちが起こりにくく利用ノード数の切り上げもないnoncontをデフォルト設定に変更した

MPI使用手引書から引用
→



Type IIサブシステム向けMPIについて

- GPUやInfiniBandをフル活用するには適切なソフトウェア環境が必要
- CUDAなどのバージョンにも影響を受ける（適切な組み合わせでないと動作しなかったり遅かったりする）
- 資料PDFをWebで、性能評価グラフをホットストレージ上で公開中
- 基本的なポイント
 - 主な通信方法：GPUに対応したOpenMPI、NVIDIA社が提供するHPC-Xというライブラリ、NCCL
 - 「不老」の構成との相性なのか、どの通信方法がベストかは状況による
 - 利用するノード数や通信パターンによって性能の優劣が異なる
 - CUDA-Aware MPIよりもcudaMemcpyとMPIを明示的に分けて書いた方が速いこともある
 - 残念ながら決定的な選択方法などは確立されていないため、公開している資料や実測結果を見て判断してください

ソフトウェア環境の変更：「不老」全体的なこと

- ソフトウェアの更新は主にメンテナンス（定期保守）のタイミングで実施
 - 影響の大きいような更新の情報はメールでも通知していますが、それ以外は通知していないため、メンテナンス時には念のため注意してください
 - moduleの使い方などによっては参照されるライブラリのパスの問題などが生じる可能性あり
 - デフォルトより新しいバージョンのmoduleが先行して導入されているものもいくつかあるため、新機能を試したい場合などに試してみてください

Type IIサブシステムのソフトウェア更新

- OpenACC対応コンパイラはPGIコンパイラからNVIDIA HPC SDKに変わっている
 - PGIコンパイラがNVIDIA HPC SDKに引き継がれたと思ってください（コンパイラオプションもある程度引き継がれている）
 - `module load hpc_sdk`
- CUDAやHPC SDKは更新頻度が高いが、「不老」は完全には追従できていない
 - （メンテナンスのタイミングとあわない、V100ではあまり関係ない、などの理由により）
 - CUDAやHPC SDK自体はホームディレクトリに導入して利用できるため、最新のコンパイラの機能等を利用したい場合は試してみてください
 - バージョン差が大きくなるとCUDAドライバ等の更新も必要となり使えなくなる
- MPIやSingularityなども少しずつ更新している
 - 高速にノード間GPU通信したい場合やSingularityとMPIを組み合わせて使う際には注意が必要

Singularity利用時の注意点など

- Singularityを3.5系から3.7系に更新したことによる影響
 - /data/group/\${USER}以下でコンテナを実行する際はバインドオプションの追加が必要
 - /data/group/\${USER}が見えない状態になってしまいホームディレクトリに移動させられる
 - コンテナ実行 (run, exec, shell) 時に以下のオプションを付けておけば/data/group/\${USER}が見える状態になるため問題は起きない
 - `--bind /data/group1/${USER}:/data/group1/${USER}`
 - 環境変数引き継ぎ用の新しいオプションが利用可能
 - 従来からの方法：SINGULARITYENV_で始まる環境変数を設定する
 - 3.6からの新機能：--envオプションで指定する

Singularity + MPI

- コンテナ環境とMPIを組み合わせる場合、うまく動作するかはコンテナの作りにも依存
- コンテナ上でmpirun：現状ではマルチノードでは動かない可能性が高い

```
$ module load singularity/3.7.4  
$ singularity exec --bind /data/group1/${USER}:/data/group1/${USER} --nv ./container.sif  
  mpirun -n 4 python ...
```

- mpirunでコンテナを実行：正しく設定すればマルチノードでも動く可能性が高い

```
$ module load cuda/11.2.1 openmpi_cuda/4.0.5 singularity/3.7.4  
$ export OMPI_MCA_btl_tcp_if_include=ib0  
$ mpirun -n 8 -machinefile $PJM_0_NODEINF -map-by ppr:4:node  
  singularity exec --bind /data/group1/${USER}:/data/group1/${USER} --nv ./container.sif python ...
```

コールドストレージの更新と新サービスについて

- 2月にコールドストレージPhase2が導入され大幅に容量が増加
 - ドライブ数が増えたため同時に読み書きできるメディア数も増加



フェーズ1: 2020年7月1日より稼働開始

機種名	PetaSite 拡張型 Library
カートリッジ数 (総カートリッジ数 / 最大搭載可能カートリッジ数)	88巻 / 88巻
総物理容量 / 最大搭載可能容量	484 TB / 484 TB
総ドライブ数	6
ODAサーバ数	1



フェーズ2: 2021年2月1日より稼働開始

機種名	PetaSite 拡張型 Library
カートリッジ数 (総カートリッジ数 / 最大搭載可能カートリッジ数)	1,092巻 / 1,980巻
総物理容量 / 最大搭載可能容量	6 PB / 10.89 PB
総ドライブ数	20
ODAサーバ数	4

新サービス：コールドストレージ単体ドライブの貸し出し

- 従来からのコールドストレージ利用に加えて、単体ドライブの貸し出しを開始
 - コールドストレージを利用したいが手元にドライブがないのが悩み、という場合などに是非ご利用ください
- コールドストレージ課金（通常の利用ポイントとは別です）
 - 1口：50TB、1回だけ書き込める（追記可能）の光ディスク×10枚（1枚約5TB）
 - ファイル負担経費（初回利用時のみ必要）：1口 190,000円
 - ファイル管理経費（毎年必要、基本負担金とは別）：1口 10,000円
- コールドストレージ単体ドライブ
 - 利用料：無料
 - 送料：利用者による負担（通常の宅急便等で送れます）



ジュラルミン・ケース

ドライブユニット(ODS-D380U)

Read: 3Gbps, Write: 1.5Gbps(ペリファイ)

外形寸法: 1,050mm(270×530×250) 重量: 6.2kg((Drive搭載時 約11kg)

まとめ

- 2021年になってからの「不老」の主な更新情報を紹介しました
- 「不老」は稼働から1年以上が経過しとても安定してサービスを提供できていますが、今後も利便性を高めるための設定更新やソフトウェアのアップデートなどは行っていくため、メールやWebを用いて提供される情報にはご注意ください
- 不便に思った点などのご連絡ください
 - 全てに対応できるかはわかりませんが、可能な範囲で改善していく予定です