

スーパーコンピュータ「不老」の利用状況について

山田 一成¹⁾, 毛利 晃大¹⁾, 林 秀和¹⁾,

片桐 孝洋²⁾, 星野 哲也²⁾, 永井 亨²⁾

1) 東海国立大学機構 情報環境部 情報システム運用課

2) 名古屋大学 情報基盤センター 大規模計算支援環境研究部門

yamada@itc.nagoya-u.ac.jp

Regarding the usage status of the Supercomputer “Flow”

Kazunari Yamada¹⁾, Akihiro Mouri¹⁾, Hidekazu Hayashi¹⁾,

Takahiro Katagiri²⁾, Tetsuya Hoshino²⁾, Toru Nagai²⁾

1) Information System Operations Division, Information Technology Department, Tokai National Higher Education and Research System

2) High Performance Computing Division, Information Technology Center, Nagoya University

概要

名古屋大学情報基盤センターのスーパーコンピュータ「不老」は 2020 年 7 月から運用を開始して 2024 年 7 月で 4 年が経過した。本稿では運用開始から 2023 年度までにおける運用状況について報告する。

1 はじめに

名古屋大学情報基盤センター（以下、センター）が 2020 年 7 月より運用を開始した、スーパーコンピュータ「不老」は、後述する 4 つのシステム、

- ・ TypeI サブシステム
- ・ Type II サブシステム
- ・ Type III サブシステム
- ・ クラウドシステム

と大容量のストレージ群、可視化システムなどが高速ネットワークによって接続された複合型のシステムである。今年、2024 年 7 月で 4 年が経過した。そのため稼働状況や利用者の利用状況といったデータが蓄積されてきた。そこで本稿では、運用開始から 2023 年度末までにおける稼働状況や利用者の利用状況などの運用状況について TypeI サブシステム及び TypeII サブシステムを中心に報告するとともに、次期システムへのヒントを探る。

なお、稼働状況や利用状況については、毎日、ベンダーから提供される統計データを元に調査した。

2 導入されたシステムについて

2.1 スーパーコンピュータ「不老」の概要

名古屋大学情報基盤センターに導入されているスーパーコンピュータ「不老」は、先にも述べたように主に 4 つのシステムと共有ストレージ群などで構成されている。これらのシステム構成図を図 1 に示す。

まず TypeI サブシステムは FUJITSU PRIMEHPC FX1000 を導入し、総メモリ容量 72TiB、総演算性能は 2,304 ノードで 7.782PFLOPS である。TypeI サブシステムは、複数ノードを利用して大規模計算などに利用されている。

次に Type II サブシステムは、FUJITSU PRIMERGY CX2570M5 を導入している。TypeII サブシステムの総ノード数は 221 ノードで、メインの総メモリ容量 82.875TiB、総演算性能は、7.489 PFLOPS である。TypeII サブシステムは 1 ノードにつき NVIDIA Tesla V100 を 4 台と 6.4TB の SSD を搭載しており、機械学習や AI などの新たな研究分野の研究者に利用されている。

また Type III サブシステムは、HPE Superdome Flex が導入されている。総ノード数は 2 ノードとなっている。

総メモリ容量は 48TiB で 1 ノード当たり 24TiB の大容量メモリと 51.2TB の SSD の利用が可能である。総演算性能は 77.414 TPLOPS となっている。

また、大容量メモリを使用する可視化処理のプリポストサーバとしての利用が考えられ、可視化システムと連携利用されている。

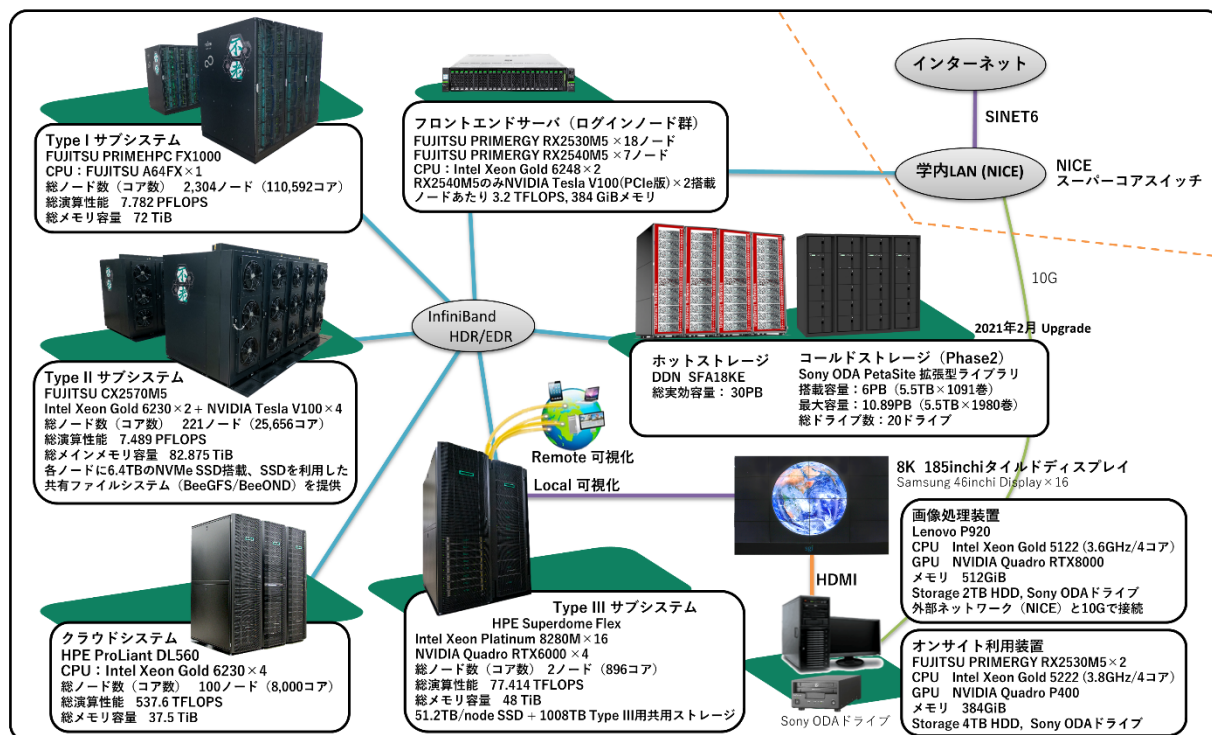


図1 スーパーコンピュータ「不老」システム構成図

次にクラウドシステムは、ノード数が 100 ノードである。クラウドシステムは、リソースの予約システムである UNCAI も導入されており、システムの一部のノードを予約により利用が可能となっている。

またこれら 4 つのシステムから利用可能な共有ストレージシステムがあるが、従来のハードディスクであるホットストレージの他に、光ディスクライブラリで構成されるコールドストレージも導入している。ホットストレージの総実行容量は、30.44 PB となっている。またコールドストレージはシステム 2 段階導入となっており、運用開始時

点のフェーズ I では総物理容量は 484 TB であった。2021 年 2 月よりフェーズ II として、総物理容量が 6 TB に増強された。なおコールドストレージ内の最大搭載容量は 10.89PB であり、利用者が光ディスクカートリッジを持ち込むことを可能としている。

その他に 8K のタイルドディスプレイ、オンサイト端末、画像処理装置を導入しており、利用者であれば誰でも利用することができる。なおスーパーコンピュータ「不老」の詳細な構成や性能については[1][2][3]で詳しく紹介している。

3 運用状況

3.1 稼働状況

スーパーコンピュータ「不老」は、24 時間計算サービスをおこなっているが、良好な運用を目的に年 3 回程度 (7 月, 11 月, 年度末)、合計 6 日間程度、計算サービスを停止して定期保守を実施している。また、すぐに対処をしなければならない

場合には臨時保守を実施している。これらの保守時間を除く時間、計算サービスをおこなっている。

(以下、サービス時間)

図 2 に各サブシステム及び年度ごとのノードサービス率(%)を示す。ノードサービス率(%)とは、式(1)で表される。

A = ノードごとのサービス時間 (合算)

B = ノードごとの保守時間 (合算)

$$\text{ノードサービス率(\%)} = \frac{A}{(A+B)} \times 100 \dots(1)$$

式(1)から、故障が少ないほど 100%に近い値となる。

図 2 に、ノードサービス率を示す。図 2 から、各サブシステムとも、安定稼働していることがわかる。

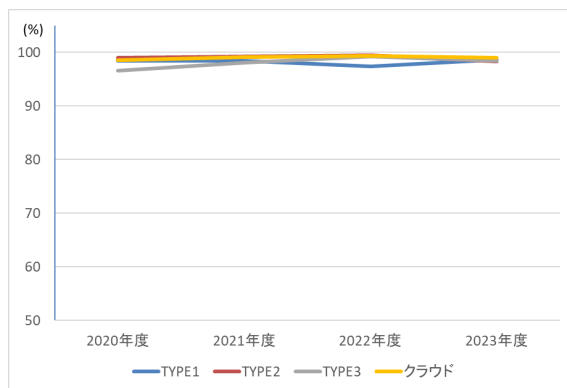


図 2 ノードサービス率(%)

図 3 に、TypeI サブシステムにおける月ごと年度ごとのノードサービス率(%)を示す。定期保守実施月と 2022 年 8 月に電源障害が原因のノード停止が発生したため、値が下がっているが、その他は大きな障害は無く稼働していることがわかる。

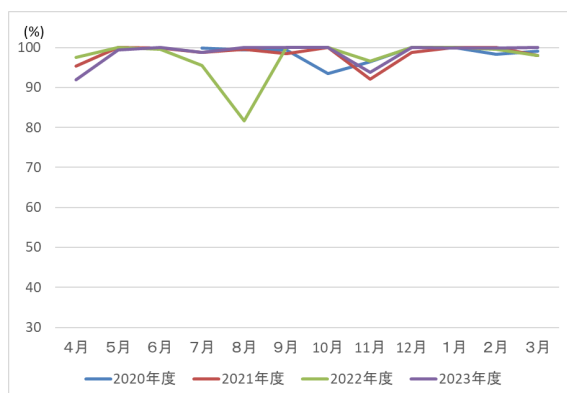


図 3 TypeI サブシステムのノードサービス率(%)

図 4 に、TypeII サブシステムにおける月ごと年度ごとのノードサービス率(%)を示す。定期保守実施月の値が下がっているが、その他は大きな障害は無く稼働していることがわかる。

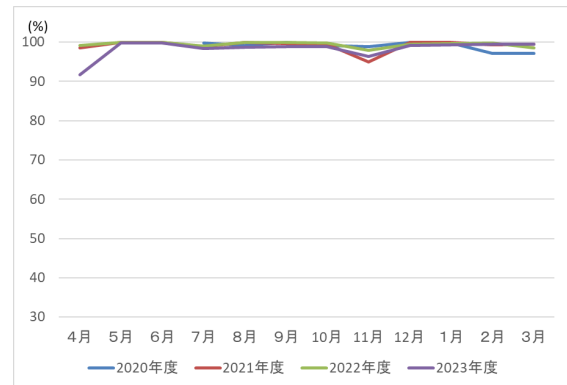


図 4 TypeII サブシステムのノードサービス率(%)

図 5 に、各サブシステムにおける 2022 年 4 月以降の月ごとの最大利用稼働率(%)を示す。最大利用稼働率とは、月ごとに「一瞬でもノードを利用されれば利用されたこととする」最大利用稼働率である。なお、TypeI サブシステム及び TypeII サブシステムにおいては、後述する「縮退運転ノード」分を除いた値となっている。クラウドサブシステムが年間を通して 100%に近い利用稼働率となっており、安定して利用されていることがわかる。TypeI サブシステム及び TypeII サブシステムは、60%~100%ぐらいの利用稼働率で良く利用されていることがわかる。TypeIII サブシステムはノードが少ないため変化が現れやすいが、高い利用稼働率となっている。

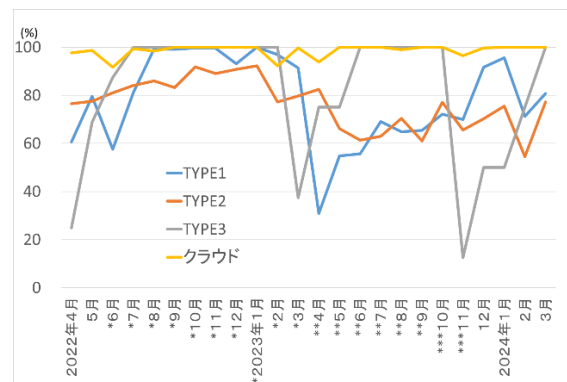


図 5 最大利用稼働率(%)の推移

図 5 における「縮退運転ノード」について

- * TypeI は縮退運転ノードを除いた値 (6 ラック中 2 ラック停止)
- ** TypeI は縮退運転ノードを除いた値 (6 ラック中 2 ラック停止) + TypeII 40 ノード停止

*** 2023 年 10 月 20 日から TypeI 全ノード稼動+TypeII 40 ノード停止 (TypeII は 2023 年 12 月 1 日より全ノード稼動)

次に、スーパーコンピュータ「不老」の 2020 年 7 月から 2024 年 3 月までの PUE (Power Usage Effectiveness : 電力使用効率) 値をシステム全体の消費電力と冷却設備を除いた消費電力から算出した。式(2)に PUE の計算式を示す。

なお、消費電力の値は、システム仕様内の「中央電子 (株) 製の環境監視システム」の計測データを利用した。

$$\text{PUE 値} = \frac{\text{システム全体の消費電力}}{\text{冷却設備を除いた消費電力}} \dots (2)$$

式(2)から計算すると

$$\frac{20,446,827.9\text{Kwh}}{15,187,926.5\text{Kwh}} = 1.35$$

となるため、PUE 値=約 1.4 である。

3.2 利用状況

図 6 に、年度ごとの総利用者数(人)及び学内比率(%)を示す。総利用者数は講習会用アカウントなどの臨時アカウントを除く人数とした。2020 年度から 2022 年度までは増加傾向であったが、後述する電力事情により、2023 年 1 月から値上げを実施した影響があり 2023 年度は減少した。また、利用者の内、50%弱が学内利用者となっている。

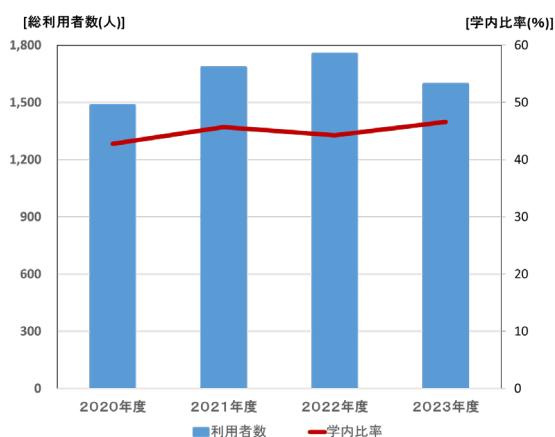


図 6 年度ごとの利用者数及び学内比率

次に、サブシステム (TypeI, TypeII) ごと、及び年度ごとに、総ジョブ件数を 100%とした場合におけるリソースグループごとに占める割合(%)を調査した。なお、スーパーコンピュータ「不老」のリソースグループの詳細については、文献[4]の一覧を参照されたい。

図 7 に、TypeI サブシステムのリソースグループごとの年度別利用変化を示す。

図 7 では、元データが件数のため、fx-debug や fx-small といった比較的資源量が少ないリソースグループの占める割合が高いことがわかる。また、fx-debug は、デバッグをすることを主な目的に作成されたリソースグループの為か、年度が進むにつれて、割合が小さくなっている。fx-small については、年度が進むにつれて、割合が高くなっている、なお、fx-debug と fx-small の資源 (概要) は次のようになっている。

fx-debug

- 最小ノード数 1
- 最大ノード数 36
- 最長実行時間 1 時間
- 最大メモリ容量 28 GiB x 36

fx-small

- 最小ノード数 1
- 最大ノード数 24
- 最長実行時間 168 時間
- 最大メモリ容量 28 GiB x 24

TypeII サブシステムについて調査した結果を、図 8 の TypeII サブシステムのリソースグループごとの年度別利用変化(その 1)及び、図 9 の TypeII サブシステムのリソースグループごとの年度別利用変化(その 2)に示す。

図 8 の TypeII サブシステムのリソースグループごとの年度別利用変化(その 1)の cx-share における 2021 年度の割合は 70.7%と高い割合となっている。そのため、他の年度でも cx-share が高い

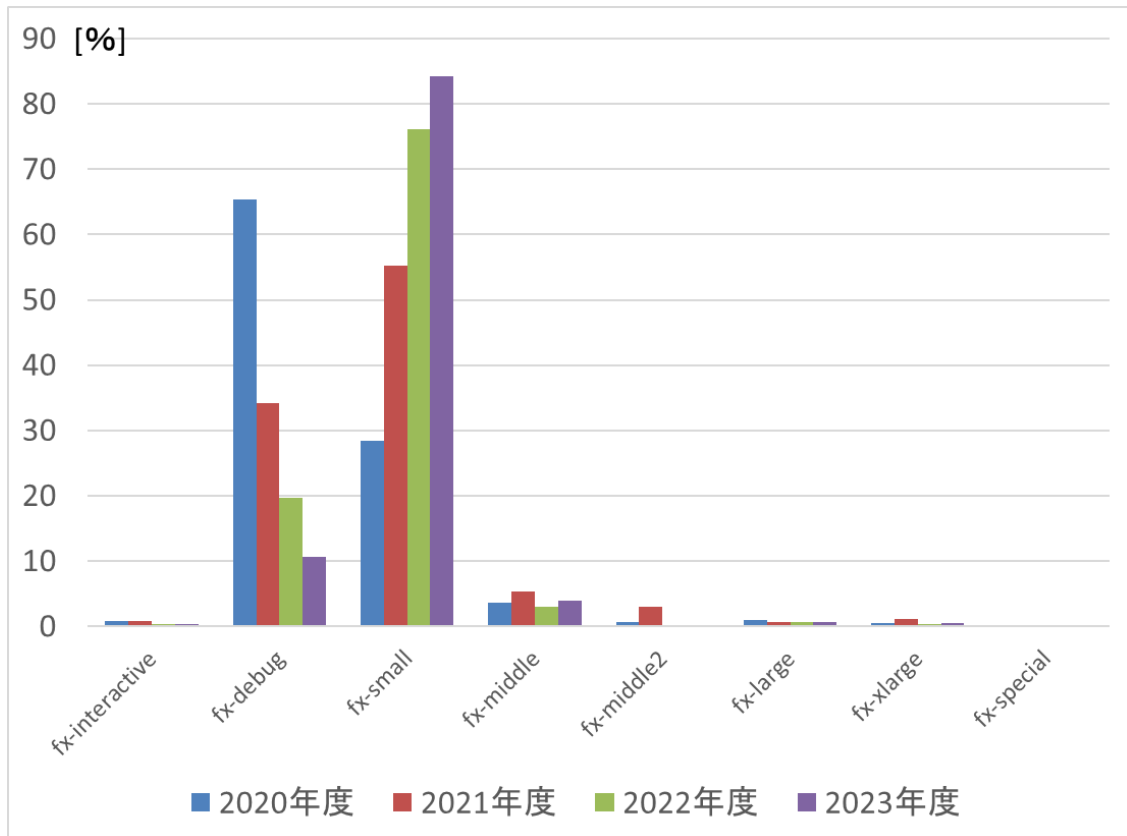


図 7 TypeI サブシステムのリソースグループごとの年度別利用変化

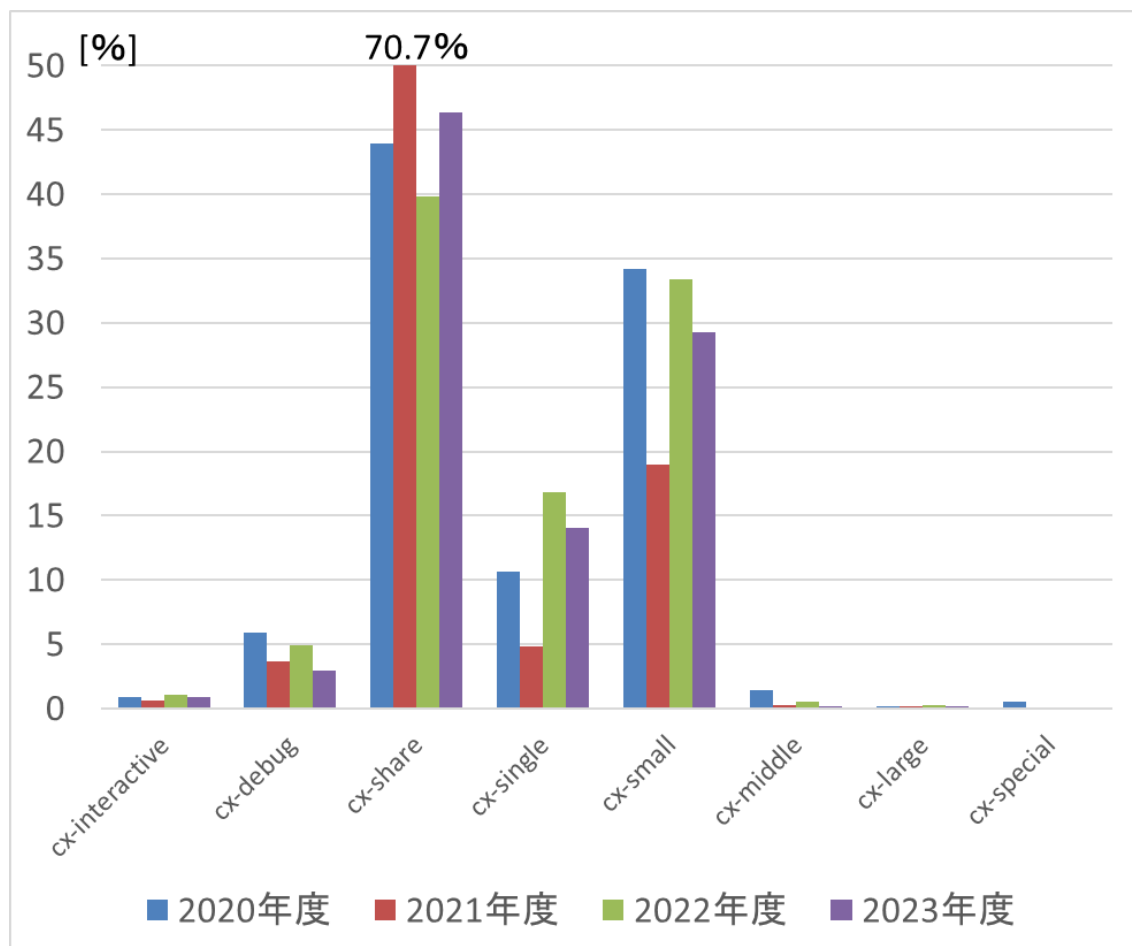


図 8 TypeII サブシステムのリソースグループごとの年度別利用変化(その 1)

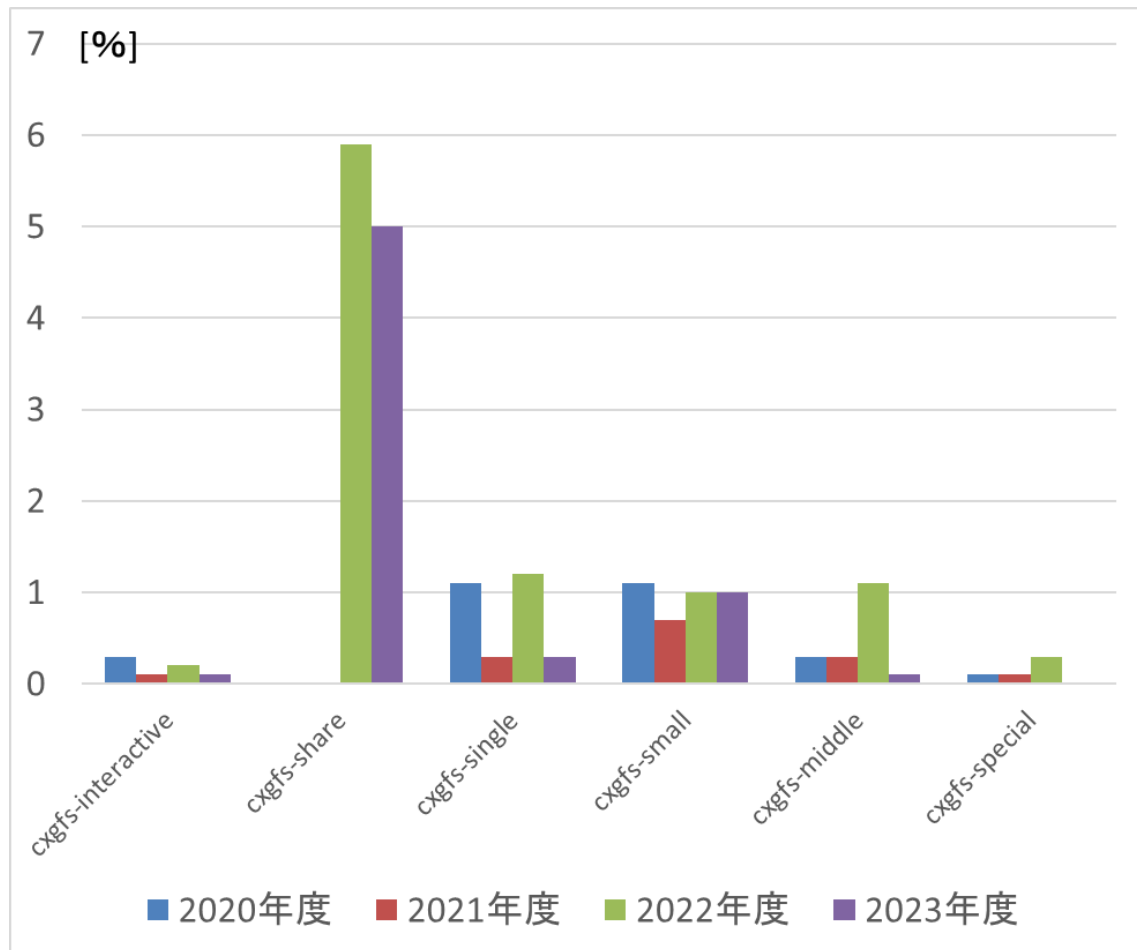


図 9 TypeII サブシステムのリソースグループごとの年度別利用変化(その 2)

割合であることが分かる。cx-share は 1GPU 利用となっており、1 ノード(4GPU)を 4 本のジョブで共有するため、料金(ポイント)が他のノード占有型のリソースグループに比べ 1/4 となっている。そのため、利用割合が高くなっていることがわかる。その他、比較的資源の小さい cx-small, cx-single も利用割合が高くなっていることがわかる。

なお、cx-share, cx-single, cx-small の資源(概要)は次のようになっている。

cx-share

- 最小ノード数 1/4(共有)
- 最大ノード数 1/4(共有)
- 最長実行時間 168 時間
- 最大メモリ容量 84 GiB

cx-single

- 最小ノード数 1
- 最大ノード数 1

- 最長実行時間 336 時間
- 最大メモリ容量 338 GiB

cx-small

- 最小ノード数 1
- 最大ノード数 8
- 最長実行時間 168 時間
- 最大メモリ容量 338 GiB×8

TypeII サブシステムの NVMesh が利用できるリソースグループについて調査した結果を、図 9 TypeII サブシステムのリソースグループごとの年度別利用変化(その 2)に示す。図 9 から、全体的には割合は少ないものの、2022 年に利用を開始した 1GPU の cxgfs-share が高くなっていることがわかる。

次に、サブシステムシステム (TypeI, TypeII) 及び年度ごとに、リソースグループごとのジョブ件数を調査した。その結果、次の特徴があること

が判明した。

図 10 に、TypeI サブシステムの中で一番大きなリソースグループ fx-special の年度ごとのジョブ件数を示す。なお、fx-special の資源（概要）は次のようになっている。

- 最小ノード数 1
- 最大ノード数 2304
- 最長実行時間 unlimited
- 最大メモリ容量 28 GiB x 2304

2020 年度が突出して多い理由は、2020 年 7 月の 1 ヶ月間は無償期間であったことが考えられる。そのため、利用者の多くが fx-special を利用した

ためと思われる。図 10 から、大規模ジョブの需要が高まってきていることがいえる。

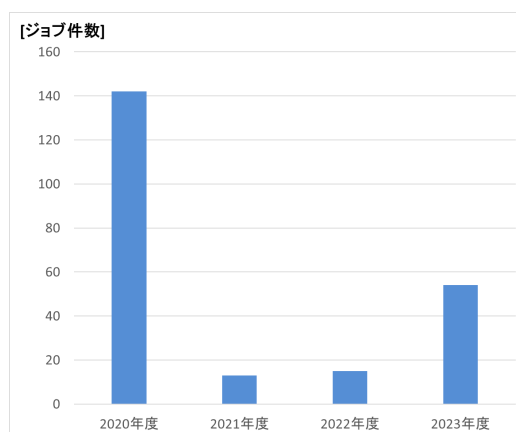


図 10 fx-special ジョブ件数

4 スーパーコンピュータ「不老」の電力事情

4.1 電力事情

スーパーコンピュータ「不老」は 2020 年 7 月より運用を開始した。2022 年以降、全国的に電気料金が上昇傾向になった。そのため、本学も電気料金が上昇し、スーパーコンピュータ「不老」の運用にも影響が出かねない状況となった。スーパー

コンピュータ「不老」の借用料金は、国から費用が賄われているが、電気料金は受益者負担の観点から利用者からの利用料にて賄われている。そのため、電気料金が値上がりすれば運用に影響が出る。

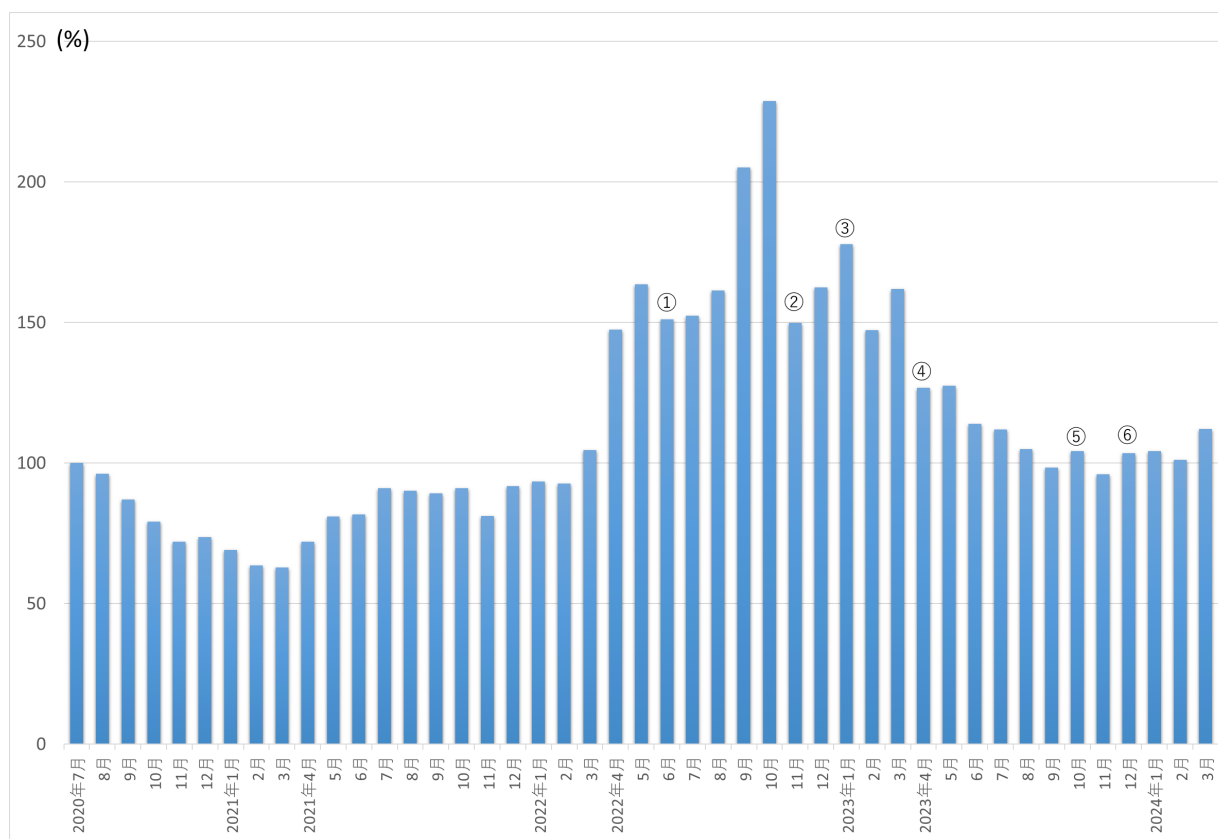


図 11 スーパーコンピュータ「不老」稼働時の電気料金を 100 とした場合の各月の割合 (%)

図 11 に、スーパーコンピュータ「不老」稼働時の電気料金を 100 とした場合の各月の割合(%)を示す

図 11 から、2020 年 7 月稼働以降、順調に運用し

ていたが、2022 年 4 月以降電力料金が上昇傾向となった。そのため、以下に示す様々な対策をおこなった。

- ① 2022 年 6 月 TypeI の縮退運転(約 1/3 停止)開始
- ② 2022 年 11 月 TypeI の TCS を使って、ノードが利用されていない（実行していない）時にクロックを下げる機能を有効化。
- ③ 2023 年 1 月 負担金規定を改定
10,000 円で 10,000 ポイント利用可能から
10,000 円で 6,500 ポイント利用可能に改定。
- ④ 2023 年 4 月 TypeI の縮退運転(約 1/3 停止)に加え TypeII 縮退運転(約 20%停止)開始
- ⑤ 2023 年 10 月 TypeI 全ノード稼働+TypeII 縮退運転(約 20%停止)継続
- ⑥ 2023 年 12 月 TypeII 全ノード稼働

2022 年 6 月の TypeI サブシステム縮退運転開始以降、スーパーコンピュータ「不老」全体の電力消費状況及び利用状況を考慮し縮退して運用してきた。縮退中は大規模ジョブが流れにくくなる

ため、利用者には大規模ジョブ利用時は、センターに連絡をしてもらうようにお知らせして運用している。

5 今後の運用について

スーパーコンピュータ「不老」は運用から 4 年が経過した。今後、2 年程度で運用を終了する予定である。そのため、TypeI サブシステムと TypeII サブシステムの利用状況から次期システムについてのヒントを考察する。

今回の分析結果から、以下が指摘できる。

- TypeI サブシステムについて、fx-small は年度ごとに利用割合が高くなっている。また、fx-special のジョブ件数が年度ごとに増加傾向にある。このことから、この大小 2 つのリソースグループの需要が高まってきている。
- TypeII サブシステムについて、1GPU で料金（ポイント）が安価な cx-share や比較的資源が小さい cx-small, cx-single が高い割合となっている。このことから、このようにノード当たりの資源を分割運用することで、安価な利用となるサービスの需要が、今後も見込まれると予想される。

今後、多くの計算機センターは、データ駆動型の研究推進を支援する、機械学習処理を強化したシステム導入が進んでいくと予想される。そのため、

機械学習処理が高速処理できる Graphics Processing Unit (GPU)を多数搭載したノードをもつサブシステムが形成されると予想される。一般に、GPU は高性能であるが電力単価が高く、ノード当たりの課金を引き上げる要因となる。そのため、上記の 2 項で指摘した、ノード資源を分割して運用し、料金を引きさげるような運用がなされる可能性が高い。その場合、高性能とスループットを維持できる運用技術の開発が鍵となると予想される。

6 おわりに

本報告では、スーパーコンピュータ「不老」の運用開始から 2023 年度末までにおける稼働状況、および利用状況などの運用状況について、TypeI サブシステム及び TypeII サブシステムを中心に、毎日、ベンダーから提供される統計データを元に調査した。その結果、以下が明らかになった。

- 各サブシステムともノードサービス率が高く、安定稼働している

- 2022 年 4 月以降における TypeI サブシステム及び TypeII サブシステムの最大利用稼働率は、60%～100%であり、良く利用されている
- スーパーコンピュータ「不老」の PUE 値は 1.35 であり、電力使用効率は良い
- 総利用者数は、各年度、1500 名～1800 名程度で、学内利用者は約 50%弱である

電力事情については、2022 年以降、電気料金が上昇傾向になり、負担金改定や縮退運転などを実施した。昨今、電気料金は一時期よりは落ち着いてきたものの、運用をする上で今後も注意である。この電気料金高騰と効果的な運用の施策については、今後の課題としたい。

参考文献

- [1] 山田一成，田島嘉則，高橋一郎，林秀和，片桐孝広，大島聡史，永井亨，スーパーコンピュータ「不老」の湧水噴霧による節電効果の評価，大学 ITC 推進協議会 2022 年度年次大会 予稿集，2022
- [2] 大島聡史，永井亨，片桐孝洋，スーパーコンピュータ「不老」のシステム構成と性能，大学 ITC 推進協議会 2020 年度年次大会 予稿集，2020
- [3] スーパーコンピュータ「不老」システム構成図，<https://icts.nagoya-u.ac.jp/ja/sc/overview.html>
- [4] スーパーコンピュータ「不老」リソースグループ一覧 http://www.icts.nagoya-u.ac.jp/ja/sc/resource_limits2.html