スーパーコンピュータ「不老」のサービスとエコシステム

田島 嘉則 [1)], 山田 一成 [1)],高橋 一郎 [1)], 毛利 晃大 [1)],
片桐 孝洋 [2)], 大島 聡史 [2)], 永井 亨 [2)]

1) 名古屋大学 情報推進部 情報基盤課
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

ytajima@itc.nagoya-u.ac.jp

Service and ecosystem of Supercomputer "Flow"

Yoshinori Tajima [1)], Kazunari Yamada [1)], Ichiro Takahashi [1)], Akihiro Mouri [1)],
Takahiro Katagiri [2)],Satoshi Ohshima [2)], Toru Nagai [2)]

1) Information Infrastructure Division, Information Promotion Department, Nagoya University
2) High Performance Computing Division, Information Technology Center, Nagoya University

概要

2020 年 7 月より、名古屋大学情報基盤センターでは、スーパーコンピュータ「不老」の運用を開始した。新システムでは、新たなサービスや前システムの運用を踏まえ消費電力を抑えるための仕組みを取り入れている。これら新システムの概要や新サービス及び、その仕組みについて述べる。

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、新スーパーコンピュータ「不老」は、４つの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システムである。本稿では、システムの概要や新たに開始したサービスそして消費電力を抑えるために導入した仕組みなどについて説明する。

2 新システムについて

2.1 システムの概要

新システムの性能諸元を表 1 に、またシステム構成を図 1 に示す。まず Type I サブシステム（以下、Type I）は、FUJITSU PRIMEHPC FX1000 を導入している。このシステムは、理化学研究所のスーパーコンピュータ「富岳」と同じ計算ノードを搭載しており、総ノード数は 2,304 ノードで総メモリ容量 72TiB、総演算性能は 7.782 PFLOPS となっている。このシステムは前システムの FX100 の後継で、超並列大規模計算用としての利用が期待される。

システム名		Type Ⅰ 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
ノードあたり	機器名称	FUJITSU PRIMEHPC FX1000	FUJITSU PRIMERGY CX2570M5	HPE Superdome Flex	HPE ProLiant DL560
	プロセッサ	A64FX (2.2GHz ,48+2コア)	Intel Xeon Gold 6230 (2.10GHz,20コア)×2	Intel Xeon Platinum (2.7GHz,28コア)×16	Intel Xeon Gold 6230 (2.10GHz,20コア)×4
	GPU	----------	NVIDIA Tesla V100×4	Quadro RTX6000×4	----------
総ノード数（総コア数）		2,304 (110,592コア)	221 (8,840コア)	2 (896コア)	100 (8,000コア)
総演算性能		7.782 PFLOPS	7.489 PFLOPS	77.414 TFLOPS	537.6 TFLOPS
総メモリ容量		72TiB	82.875TiB	48 TiB	37.5TiB
ノード間接続		Tofuインターコネクト D	InfiniBand EDR×2	InfiniBand EDR	InfiniBand EDR
冷却方式		水冷	水冷	空冷	空冷

表 1 新システム諸元表

次に Type II サブシステム（以下 Type II）は、FUJITSU PRIMERGY CX2570M5 を導入している。この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 となっている。このシステムは 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の利用が期待される。

また Type III サブシステム（以下、Type III）は、HPE Superdome Flex が導入され、総ノード数は 2 ノードとなっているが、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 の利用が可能となっており、総演算性能は

77.414 TPLOPS となっている。

Type III は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が期待されている。

次にクラウドシステムはノード数が 100 ノードとなっており、前システムの CX400 の後継として導入されている。またこのシステムには、リソースの予約システムである UNCAI が導入されており、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。

次にストレージシステムであるが、新システムでは、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ーであるコールドストレージを導入している。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構成図からも分かるように、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 であり、2021 年 2 月頃までにフェーズⅡとして、総物理容量が 6 TB になる予定である。なおコールドストレージ内の最大搭載容量は 10.89PB となっており、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。

最後に可視化システムであるが、前システムの可視化システムを継承し、新たにオンサイト端末や画像処理装置を導入している。

なお新システムの詳細な構成や性能については[1][2]で詳しく紹介している。

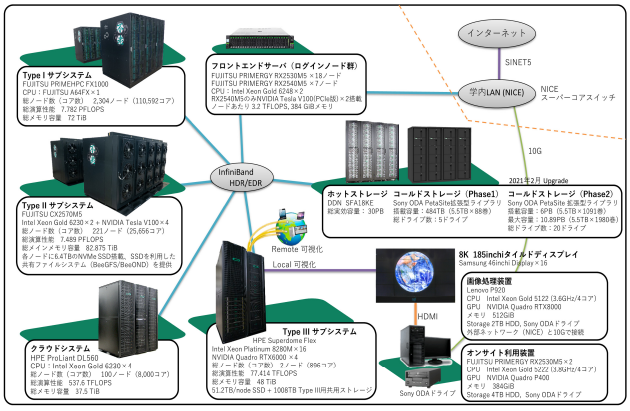

図 1　システム全体構成図

3 新たなサービス

運用を開始したスーパーコンピュータ「不老」では、前システムで行っていた主なサービスは継続しているが、変更したものや新たに開始したものもある、これら新たなサービスについて述べる。

3.1 グループ利用サービス

グループ利用サービスは、前システムで行っていた機関利用サービスを機関単位ではなく、研究グループ単位で申請できるように改訂したサービスである。1 口 20 万円で 20 名まで登録可能で利用可能ポイントは 20 万ポイント付与される。この制度は研究グループ内であれば他機関の研究者と利用可能ポイントを共有して利用することができる[3]。

3.2 ノード準占有サービス

ノード準占有サービスは、前システムの FX100 でサービスをしていたノード占有サービスとは異なる制度であり、対象システムは Type II 及びクラウドシステムである。このノード準占有サービスとノード占有サービスの異なる点は、ノード占有サービスの場合では申請者が占有用ノードを利用しない限りノードは稼動しないため稼動率が低下することになる。しかしノード準占有サービスの場合では申請者が準占有用ノードを利用していない時、一般利用者の短時間（1 時間以内）ジョブを実行させることでノードを稼動させ、ノード準占有の申請者がジョブを投入した時に、一般利用者の短時間ジョブが実行されていた場合には、一般利用者のジョブの終了後に優先して実行させる。つまり繁忙期であっても 1 時間以内にジョブが実行できることを保障する制度である。これにより計算ノードが有効活用され稼動率の低下を防ぐこともできる[3]。

3.3 リソース予約サービス

クラウドシステムでは、リソース予約システム（UNCAI）のサービスを開始している。このサービスはクラウドシステム内で、UNCAI 用に割り当てられたリソースを予約して利用するものであり予約は Web 画面から行え、VS(10cores,45GBmem),VM(20cores,90GBmem),VL(40cores,180GBmem),VX(80cores,360GBmem)の 4 種類の仮想ノードが予約可能となっている。

3.4 コールドストレージサービス

コールドストレージサービスは、今回新たに導入した、光ディスクライブラリーのサービスであり、専用のログインノードで利用することができる。このログインノードにはホットストレージもマウントされているため、計算結果やデータをコールドストレージに保存する事が可能である。利用負担金は1口 50TB 利用で、初年度はファイル負担金（20 万円）とファイル管理費（1 万円）を徴収するが、継続利用した場合 2 年目からは管理費のみ徴収する。この光ディスクは追記型であるため媒体を購入する利用形態となり、利用終了時には希望により光ディスクを持ち帰る事を可能としている。これにより研究室などで専用ドライブを購入すれは計算結果などのデータを活用することができる。

4 消費電力を抑える取り組み

前システムでは 7 月～9 月に掛けて、名古屋大学のピーク電力を抑えるため昼間時間帯で縮退運転を実施し協力してきた。今回導入したスーパーコンピュータ「不老」には、消費電力を抑えるための仕組みが導入されている、その仕組みについて紹介する。

4.1 縮退運転機能

前システムでは 7 月から 9 月にかけて FX100 システムの 14 ラック中 4 ラックを停止して大学のピーク電力を抑えるために協力をしていたが、新システムでは縮退用のノードを Type I に 384 ノード、Type II に 48 ノード、クラウドシステムに 18 ノード設定し、そのノードを使用するためのリソースグループ extra を新設した。これはシステムの消費電力を測定し、ある値を超えた場合、extra 専用ノードを縮退させ、大学全体の消費電力が低くなる夕方から翌朝まで夜間の 15 時間程度のみ縮退を解除して計算サービスを実施する仕組みで、これにより extra 専用ノードが縮退状態となった時、システムの消費電力を抑えることができる[4]。

4.2 湧水の利用

スーパーコンピュータ「不老」の冷却設備の構成図を図 2 に示す。

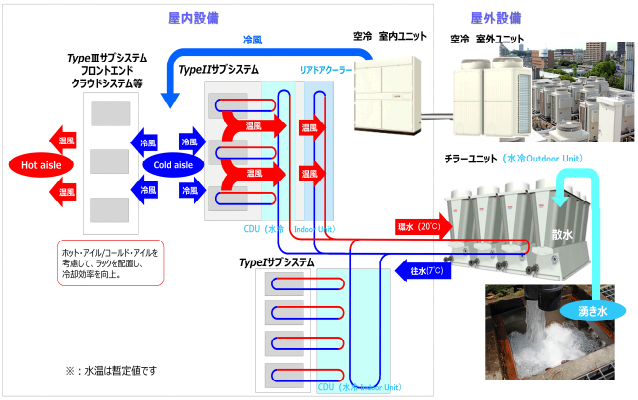

図 2　冷却設備構成図

図 2 から分かる様に、各システムは Type I、Type II は水冷方式、その他のシステムは空冷方式となっている。また新たな試みとして、センター地下に湧き出している地下水を利用している。センターの湧き水は年間通して 18℃前後で毎秒 0.55L 程度の水量が湧き出しており、センター地下の釜場に溜まっており一定の水位になるとポンプでくみ上げ排水溝に流している。この湧水を冷却装置に利用し消費電力を抑えることができないかと考え、今回導入したシステム仕様に盛り込んだ。現在情報基盤センター南側に置かれている水冷設備のチラー装置に、気温が 20℃以上になった時、湧水を噴霧しチラー装置及び、その周辺温度を下げている。これにより 7 月～9 月のチラー装置の消費電力の低下が期待できる。

4.3 消費電力の推移

新システムでは、前システムと同様に消費電力及びシステム室の温度/湿度を計測し監視している。消費電力は、Type I、Type II、Type III の各サブシステム、クラウドシステム、ストレージ、管理サーバ群及び冷却装置とシステム毎に測定しており、センター入口ホールに設置されたモニターに表示し、消費電力の見える化を実施している。次に 7 月～8 月の積算電力と平均消費電力の推移を図 3 と図 4 に示す。

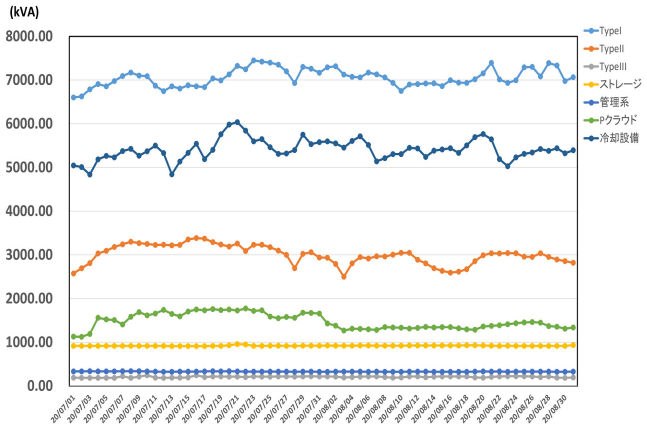

図 3 各システムの消費電力の推移（積算）

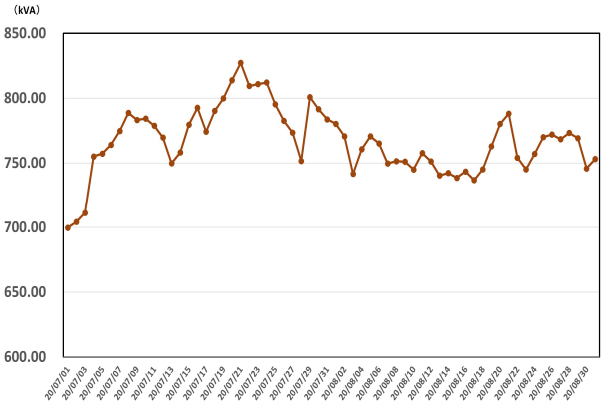

図 4 システムの 1 時間当りの消費電力

今回稼働を開始した、スーパーコンピュータ「不老」の定格消費電力の合計は約 1.9MVA となっている。図 4 の 1 日の平均省電力のグラフから、7 月から 8 月の間日毎での 1 時間当りの平均消費電力は 750~800kVA で推移していることが分かる。この値は定格電力の 30～40%となっており、通常運用の消費電力を前システムと同等の 1.4~1.5MVA と想定していたが大幅に低くなっていることがわかる。この要因として考えられるのは、新システムが立ち上がった直後であるため稼動率はまだ低いが、システム全体を地下室に持ってきたことにより外気の影響が低いこと、また湧水を利用する仕組みなどを取り入れた事などが考えられるが検証はできていない。

4 まとめ

2020 年 7 月より、スーパーコンピュータ「不老」の運用が開始され、2 ヶ月程経ったが新システムの立ち上がり直後のため、稼動率や消費電力が低く、この 7 月、8 月と縮退運転を実施しなかった。また湧水の利用であるが、7 月から 8 月のチラーなどの冷却設備の消費電力は、定格電力の約 20%前後で推移している。これが先にも述べたが、システム設置場所や湧水の効果であるかは、稼動率も低いことから、今のところ正確には検証できてはいない。また今後の消費電力の推移から縮退運転用計算ノードの設定割合も検討する必要が出てくるため、今後の各システム消費電力の推移を注視し効果を検証し、より良いシステム運用を行っていきたいと考えている。

謝辞

新システムの導入に当り新サービス及び利用負担金規程などを検討して頂いた、情報連携推進本部の教職員の方々に感謝いたします。

参考文献

[1] 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

[2] スーパーコンピュータ「不老」システム構成図
http://www.icts.nagoya-u.ac.jp/ja/sc/overview.html

[3] スーパーコンピュータ「不老」利用負担金規定
http://www.icts.nagoya-u.ac.jp/ja/sc/riyou/apps.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_limits.html