

スーパーコンピュータ「不老」のシステム構成と性能

大島 聡史¹⁾, 永井 亨¹⁾, 片桐 孝洋¹⁾

1) 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

ohshima@cc.nagoya-u.ac.jp

System overview and performance of supercomputer “Flow”

Satoshi Ohshima¹⁾, Toru Nagai¹⁾, Takahiro Katagiri¹⁾

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

概要

名古屋大学 情報基盤センターでは 2020 年 7 月 1 日に新しいスーパーコンピュータ「不老」のサービスを開始した。スーパーコンピュータ「不老」は、多様化する計算需要に対応するため、特徴の異なる複数の計算サブシステムを備えた複合型のシステムとし、新しい制度も多くとり入れた。本稿ではスーパーコンピュータ「不老」のシステム構成とその性能について紹介する。

1 はじめに

名古屋大学 情報基盤センター（以下、当センター）は名古屋大学 情報連携統括本部に含まれる組織であり、学内外の様々な情報サービスを一元的に扱っている。当センターでは全国共同利用施設として学内外の研究者に計算資源を提供してきており、近年では、2013 年に導入した FUJITSU Supercomputer PRIMEHPC FX10 システム（以下、FX10 システム）[1] を 2015 年にアップグレードした FUJITSU Supercomputer PRIMEHPC FX100 システム（以下、FX100 システム）[2] と、同様に 2013 年に導入し 2015 年にアップグレードした FUJITSU Server PRIMERGY CX400 システム（以下、CX400 システム）を中心としたサービスを提供してきた。これらのシステムは安定して計算資源の提供を続けてはきたものの、運用期間が当初の予定を超え、またノードあたりの性能も現在主流の機種に劣るものとなってきたことから、2020 年 3 月末日にてサービス提供を終了した。

当センターでは、2018 年頃から本格的に次期システムの仕様について調査検討を開始し、2019 年 9 月に仕様書を公開した。その後、2020 年 1 月に富士通株式会社は落札し、2020 年 7 月 1 日に新システムのサービスを開始した。従来、当センターでは提供するスーパーコンピュータシステムに愛称をつけてこなかったが、利用者にさらに身近に感じてもらえるシステムとなることを目指して、今回初めて公募を行い愛称を付けた。愛称である「不老」は、名古屋大学の所在地名である

愛知県名古屋市千種区不老町に由来し、このシステムによって人類が得た恩恵がその後末永く人類の文明の中に生き続けることを願って命名されたものである。なお、英語名称は処理の流れなどを示す「Flow」であり、「不老」の読みも「フロウ」ではなく「フロー」である。スーパーコンピュータ「不老」（以下、「不老」）は、従来のスーパーコンピュータシステム利用者に引き続き計算資源を提供するという役割に加えて、多様化する利用者の需要・要望や社会的ニーズ等に応えるために、特徴の異なる複数の計算サブシステムと 2 種類のストレージシステムを備えた複合的なシステムとなっている。また、電力モニタリング機構とそれを活用した節電対応運転を行う仕組みを有し、消費電力削減のために湧水を活用した冷却機構を備えるなど、先進的で挑戦的なシステムとなっている。

2 スーパーコンピュータ「不老」のシステム構成

2.1 全体構成

「不老」の全体構成を図 1 に示す。また、当センター本館地下 1 階にある計算機室に設置された「不老」の写真を図 2 に示す。

多様化する計算需要に応えるため、「不老」は特徴の異なる複数の計算サブシステムと 2 種類のストレージシステムを中心とした複合的なスーパーコンピュータシステムとなっている。「不老」の主要な計算サブシステムとストレージの主な仕様を表 1 および表 2 に示す。システム全体の理論演算性能 (Type I, II, III, お

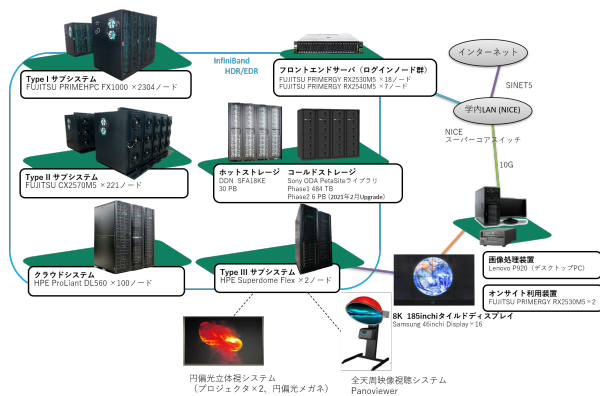


図 1 スーパーコンピュータ「不老」の全体構成



図 2 スーパーコンピュータ「不老」の外観写真 (計算機室の奥からと手前から)

よびクラウドに搭載された CPU と Type II に搭載された GPU の倍精度浮動小数点演算性能の合計値) は 15.88 PFLOPS であり、これは旧システムの全体性能 (約 3.7 PFLOPS) の約 4.3 倍である。各サブシステム間には InfiniBand HDR/EDR にて相互に接続されているものの、複数のサブシステムにまたがって 1 つの計算ジョブを実行することは想定されていない。

「不老」の最大消費電力は約 1900kVA であり、Type I サブシステムと Type II サブシステムのみが水冷による冷却方式を用いている。「不老」における省電力に向けた取り組みと実消費電力については文献 [3] を参照されたい。

2.2 Type I サブシステム

Type I サブシステムは、理化学研究所と富士通社が共同開発している「富岳」[4] の商用版として知られる FUJITSU Supercomputer PRIMEHPC FX1000(以下、FX1000)[5] により構成されている。FX1000 は A64FX CPU を 1 ノードあたり 1 基搭載しており、その理論演算性能 (倍精度浮動小数点演算) は 1 ノードあたり約 3.38 TFLOPS である。A64FX の主記憶は HBM2 であり、1,024 GB/s という高いバンド幅を有する一方、容量は 32GiB に限られる。ノード間は高速で低遅延な Tofu インターコネクト D によって相互接続されている。Type I サブシステムは 6 ラック

に 2,304 ノードが搭載され、総理論演算性能は 7.782 PFLOPS と全サブシステムの中で最も高い。

Type I サブシステムは従来の FX100 システム利用者向けの高速な後継機としての期待が大きい。さらに、現在は試行運用中かつ本格運用においても審査があり利用者が限られる「富岳」に対して、「不老」は利用資格を持つ者であれば利用できることから、「富岳」における超大規模実行を目指す研究者・開発者にとっての開発・テスト環境や「富岳」向けに開発された高性能ソフトウェアの利用環境として、また新しいアーキテクチャ向けの研究開発環境としてなど、様々な役割が期待される。

2.3 Type II サブシステム

Type II サブシステムは、FUJITSU Server PRIMERGY CX2570 M5 により構成されている。本サブシステムは、各計算ノードに 2 基の Intel Xeon CPU と 4 基の NVIDIA Tesla V100 GPU を搭載した、いわゆる GPU スパコンである。現行の最新世代といえる CPU と GPU を搭載しており、約 33.89 TFLOPS という高いノードあたりの理論演算性能 (倍精度浮動小数点演算) を備えている。本サブシステムは 221 ノードが 10 ラックに搭載されており、ノード間の接続は InfiniBand EDR、総理論演算性能は 7.489 PFLOPS と Type I サブシステムに匹敵する。

Type II サブシステムは、当センターでは初の大規模な GPU 搭載システムである。高性能な GPU を多数搭載するとともに、各ノードに高速な NVMe SSD を搭載、コンテナ機能も提供しており、当センターではこれまで十分に対応できていなかった機械学習やビッグデータなどデータ科学分野を中心とした様々な分野の研究への活用が期待される。

2.4 Type III サブシステム

Type III サブシステムは、2 ノードの HPE Superdome Flex によって構成されている。本サブシステムは各計算ノードに 16 基の Intel Xeon CPU と 4 基の NVIDIA Quadro GPU を搭載しており、ノードあたりの理論演算性能 (倍精度浮動小数点演算) は約 38.71 TFLOPS と全サブシステムで最も高い。しかし本サブシステム最大の特徴はノードあたり 24 TiB という大規模な主記憶を搭載している点であり、大規模なプリ・ポスト処理や可視化処理への活用が期待される。本サブシステム内のノード間は InfiniBand EDR によって相互接続されているが、1 ノードは会話型処理、もう 1 ノードはバッチ型処理と異なる利用方式で運用している。

Type III サブシステムは当センター 1 階の可視化室に設置されている 185 インチ 8K 高精細ディスプレイなど可視化用の設備と組み合わせて使えるよう設計されているため、コンテンツの作成や閲覧などにも活用することができる。

2.5 クラウドシステム

クラウドシステムは、100 ノードの HPE ProLiant DL560 により構成されている。クラウドシステムは、Type II サブシステムと同様の Intel Xeon CPU を 1 ノードあたり 4 基搭載している。そのため、ノードあたりの理論演算性能 (倍精度浮動小数点演算) は、CPU だけで比較すれば約 5.38 TFLOPS と Type I, II の各サブシステムよりも高く、特にスレッド並列化のみ適用されている計算インテンシブなプログラムの高速実行に適している。クラウドシステム内のノード間は InfiniBand EDR によって相互接続されている。

クラウドシステムは、スーパーコンピュータシステムでは一般的であるバッチジョブシステムを介した利用に加えて、Web ベースの予約システムを介した時刻指定実行にも対応している。そのため本システムは、Type I, II 各サブシステムの混雑緩和とノードあたり CPU 性能の高い x86 CPU 環境の提供に加えて、インタラクティブ処理環境の提供により研究室の PC サーバ/ワークステーションなどの日常的な作業環境を代替するなど、様々な需要に対応できる。

2.6 共有ストレージ

「不老」は 2 種類の大規模な共有ストレージを有している。

ホットストレージは、スーパーコンピュータシステムでは一般的である、多数の HDD によって構成される共有 RAID ストレージである。保存されたデータの情報を保持するメタデータサーバ (MDS)・メタデータストレージ (MDT) は FUJITSU PRIMERGY RX2540 M5 および FUJITSU ETERNUS AF250 S2 により構成されており、データ自体は DDN SFA18KE および DDN SS9012 に保持される。ホットストレージのデータは RAID6(8D+2P) により保護されており、総容量は物理容量 40.32 PB、実効容量約 30.44PB である。従来のストレージは実効容量約 6PB であったため、約 5 倍に増量されている。

ストレージ容量は約 5 倍に増量されたものの、近年はデータ科学研究の発展にともない非常に大容量のデータを生成する研究も増加している。特に、観測機器 (センサ) から得られるデータや自動運転に用いる環境データなど、生成されたら変更は行わない (変更

してはならない) 大容量データや、アクセス頻度は低い重要性の高い大容量データを扱う需要が急速に増加している。HDD や SSD から構成されるストレージは最大容量にあわせて消費電力も増加し、また記録媒体を取り出して持ち出すことも現実的ではない。そこで「不老」ではこれらの需要に効率よく対応するため、追記書き込みのみ可能で耐久性の高い光ディスク・アーカイブ (Optical Disc Archive) を用いたコールドストレージを、スーパーコンピュータシステムとしては世界に先駆けて導入した。コールドストレージはソニー社の PetaSite 拡張型ライブラリにより構成され、2020 年 7 月の運用開始時点 (Phase1) では総容量 484 TB、2021 年 2 月 (Phase2) には約 6 PB に拡張される予定である。コールドストレージの詳細については文献 [6] を参照されたい。

2.7 その他の構成ハードウェア、ソフトウェア

以上の計算サブシステムやストレージに加えて、FUJITSU PRIMERGY RX2530 M5 および RX2540 M5 から構成される合計 25 台のフロントエンドシステム (主にログインノード群として利用)、USB 記憶装置などを持ち込んで「不老」のデータを読み書きする際などに便利なオンサイト利用装置、「不老」と可視化設備に接続されている Windows 端末である画像処理装置、その他の管理ノード群などにより「不老」は構成されている。

「不老」は商用ソフトウェアを含め様々なソフトウェアを提供している。特に、Type I サブシステムは CPU アーキテクチャが他の計算サブシステムとは異なり、また Type II サブシステムは GPU を搭載しているため、ソフトウェアごとに利用が推奨される計算サブシステムが異なる。計算サブシステムとソフトウェアの対応情報など最新の情報については Web ページ [7] にて公開している。

3 スーパーコンピュータ「不老」の性能

「不老」の各サブシステムを用いて測定した主要ベンチマークの性能を報告する。本稿に掲載している性能値はいずれも実機による性能値である。Type I サブシステムについてはいずれも Fujitsu Software Technical Computing Suite V4.0L20 に含まれるコンパイラや MPI を用いた。Type II サブシステムとクラウドシステムについては特に断りがない限り Intel Parallel Studio XE 2019 Update 5 に含まれるものを用いた。

3.1 STREAM ベンチマーク

STREAM ベンチマーク [8] は 2 つまたは 3 つの配列に対する単純な処理の性能を測定するベンチマークであり、メモリ転送性能を測定するものとして広く用いられている。STREAM ベンチマークに含まれる項目のうち、配列の積和計算を行う Triad の性能を示す。GPU の性能については、OpenMP により並列化されている CPU 版 (オリジナル版) を OpenACC を用いて GPU 化し、1GPU 内のデバイスメモリに対する処理の性能を測定した。

Type I サブシステムの性能は 826GB/s、理論性能に対して約 80.7% の性能を得た。特にコンパイル時オプション-Kzfill および実行時の環境変数 XOS.MMM.L.PAGING.POLICY を demand:demand:demand に設定することが性能に大きく影響した。Type II サブシステムの性能は、1 ノード 2CPU で 202GB/s (理論性能の約 71.7%)、1GPU で 777GB/s (理論演算性能の約 86.3%) の性能を得た。CPU のメモリ性能はメモリ種別が異なるため Type I サブシステムと比べて低い。GPU のメモリ性能は、同じ HBM2 でも A64FX より理論性能の低い構成であるため性能値では劣るが、理論性能に対する実測性能値の割合は高く、またノード内に 4GPU 搭載しているため、ノード全体のメモリ転送性能では Type I サブシステムを凌駕する。クラウドシステムは 1 ノード 4CPU で 338GB/s (理論性能の約 60.0%) と、Type I、II サブシステムと比べて理論性能に対する実測性能値の割合が低かった。

3.2 HPL ベンチマーク

High Performance Linkpack (HPL) ベンチマーク [9] は、LU 分解により連立一次方程式の求解を行う性能を求めるベンチマークである。このベンチマークは、倍精度浮動小数点データに対する行列積和演算 (Level-3 BLAS DGEMM) の性能に大きく影響を受けることが知られている。また、世界中のスーパーコンピュータシステムの性能を順位付けする TOP500 および Green500[10] の指標としても、本ベンチマークが用いられている。

TOP500 ランキング 2020 年 6 月版において、Type I サブシステムは 6,617.8 TFLOPS で 36 位 (2020 年 9 月 30 日現在) にランクインしている。日本国内の順位では、「富岳」、AI Bridging Cloud Infrastructure (ABCI)、Oakforest-PACS、TSUBAME3.0 につづく 5 位である。このときの問題サイズ (Nmax) は 2,621,440、理論演算性能 (Rpeak) に対する測定値

(Rmax) は約 85.0% である。なお、前回の 2019 年 11 月版における FX100 システムの性能と順位は 2,910 TFLOPS で 84 位であった。

Type II サブシステムは、導入時期の都合により TOP500 ランキングには掲載されていないが、221 ノード全てを用いた実測値として 4,880 TFLOPS が確認されている。この性能は TOP500 ランキング 2020 年 6 月版の 50 位に相当する。このときの問題サイズ (Nmax) は 1,880,064、理論演算性能 (Rpeak) に対する測定値 (Rmax) は約 65.2% である。コンパイラや MPI については、Intel Parallel Studio XE Cluster Edition 2018 update4、CUDA10.2、OpenMPI-3.1.3 を用いた。

クラウドシステムについては、Intel コンパイラ (mkl, 2019 Update 5) に付属のプログラム (mp_linalgpack) にて測定したところ、1 ノード 4CPU で 3.25 TFLOPS、実行効率は約 60.5% であった。

3.3 HPCG ベンチマーク

HPCG ベンチマーク [11] は、有限要素法から得られる疎行列を対象として共役勾配法 (Conjugate Gradient method, CG 法) を用いて連立一次方程式を解く性能を求めるベンチマークである。本ベンチマークは行列積和演算の性能に大きく左右される HPL ベンチマークと比べて現実のアプリケーションに近いものとされている。

HPCG ランキング 2020 年 6 月版において、Type I サブシステムは 230.594 TFLOPS で 16 位にランクインしている。日本国内の順位では、「富岳」、AI Bridging Cloud Infrastructure (ABCI)、Oakforest-PACS に次いで 4 位である。HPCG は HPL と比べてピーク性能に対する実測性能の比が非常に低いことが知られており、「不老」 Type I サブシステムの性能比は 3.0% である。この性能比は現在のスーパーコンピュータにおいては非常に高い値であり、「不老」よりも上位で 2.0% に到達しているのは HPCG スコア世界一の「富岳」(2.6%) のみである。なお、2019 年 11 月版における FX100 システムの性能と順位は 86.534 TFLOP で 31 位であった。

Type II サブシステムについては導入時期の都合により全ノードを利用した測定が行えていない。現時点では 100 ノードを用いたテスト実行 (ランキング登録には 3600 秒以上の実行時間が必要だが、今回はテスト実行として 60 秒だけ動かした) を行った結果として 48254.4 GFLOPS の性能が確認できている。今後システム全体での測定を行う予定であるが、100 ノー

ドの結果から単純に推測すると 100 TFLOPS 程度の性能が期待される。実行には HPCG ベンチマークの Web サイトで配付されている 2019 年 12 月 5 日版の HPCG 3.1 バイナリを用いた。

クラウドシステムについては、Intel コンパイラ (mkl, 2019 Update 5) に付属のプログラムを用いて測定したところ、1 ノード 4CPU、60 秒試行で 61.7766 TFLOPS、実行効率 1.15% であった。

4 OSU Micro-Benchmarks

OSU Micro-Benchmarks[12] は MPI 通信性能を測定するベンチマークである。本ベンチマークは様々な通信パターン (通信関数) に対応しているが、ここでは 2MPI プロセス間通信のレイテンシと MPI Allreduce 集団通信の性能を報告する。

Type II サブシステムの OpenMPI については HPC-X v2.6 を用いた。

図 3 に Type I サブシステムと Type II サブシステムの 1 ノード内 2 プロセスによる通信のレイテンシを示す。Type I サブシステムについては同一ノード内の異なる CMG 間、Type II サブシステムについては同一ノード内の異なる CPU ソケット間の通信を対象としている。測定の結果、通信サイズ 8,192Byte までは Type II の Intel MPI が最も低レイテンシ、通信サイズ 16,384Byte から 262,144Byte までは Type I が最も低レイテンシという結果となった。GPU 間通信は他よりもレイテンシが大きい、262,144Byte 以上では高速という結果であった。

図 4 には 2 ノード間の通信のレイテンシを示す。16,384Byte までは Type I と Type II の OpenMPI のレイテンシが同程度に小さく、それ以上のサイズでは Type I がやや劣るという結果が得られた。

図 5 には集団通信 MPI Allreduce の通信レイテンシを示す。同一ノード数で比較すると、通信サイズが非常に小さい場合は Type I サブシステムが高速であるが、それ以外は Type II サブシステムが高速という結果となった。また GPU 間の通信時間は CPU 間より大幅に遅かった。

5 IOR ベンチマーク

IOR ベンチマーク [13] はストレージの性能を測定するベンチマークである。ここではホットストレージに対するアクセスの性能を報告する。

「不老」のホットストレージは全サブシステムから共通にアクセス可能である。ただし、1 台の MDT、4 台

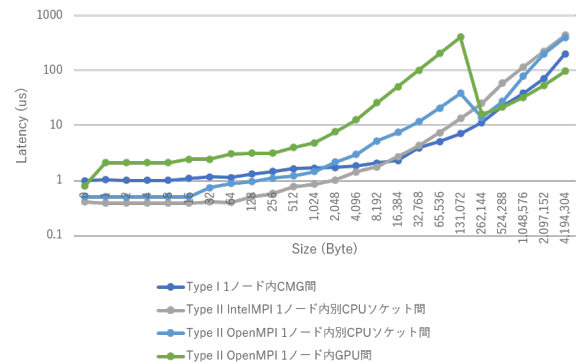


図 3 1 ノード内 1 対 1 通信のレイテンシ

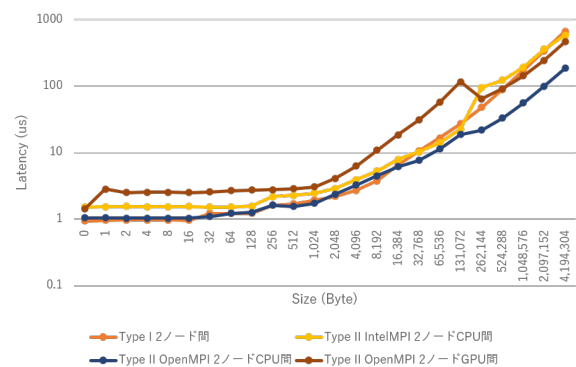


図 4 2 ノード間 1 対 1 通信のレイテンシ

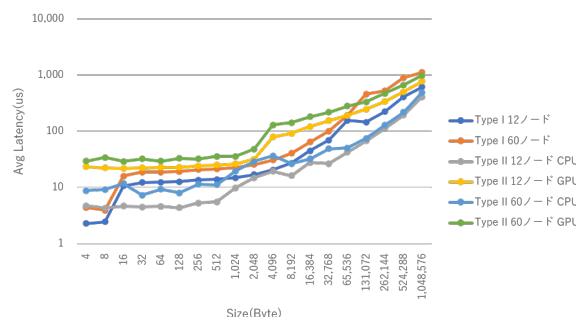


図 5 MPI Allreduce の通信レイテンシ (ノードあたり 1 プロセス)

の MDS、4 組の OSS/OST によって構成されており、その構成には幾つか選択肢がある。「不老」では、実際の利用シナリオを想定したストレージの負荷分散とスループット向上や耐障害性などを考慮し、OST_pool とマルチ MDS 機能を用いてアクセスの分散を図っている。構成の概要を図 6 に示す。/home には一般ユーザのホームディレクトリを配置し、/data/group1 は従来のシステムからの引き継ぎデータや追加で大容量のストレージを要求するユーザ用に利用、/data/group2 は特に I/O 負荷の大きなジョブ (ヘビーユーザ) を隔

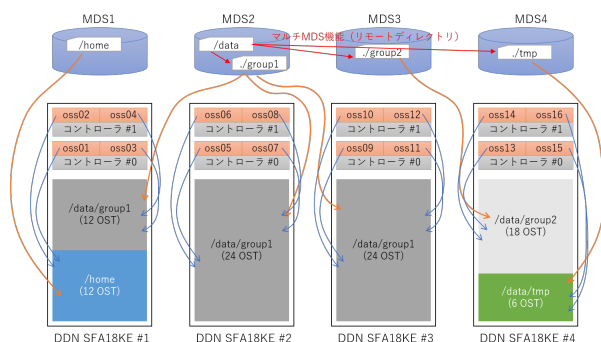


図6 ストレージの構成

離するために用いている。

ストレージの各領域に対して実施した IOR ベンチマークの性能を表3に示す。いずれも POSIX 版の結果である。性能評価結果から、多数のユーザが同時に利用する /data/group1 領域については読み込みと書き込みそれぞれ 100GiB/sec 程度、少数のヘビーユーザ向けの /data/group2 領域については 35GiB/sec 程度の性能を提供できることが確認できた。

6 おわりに

本稿では、2020年7月1日より名古屋大学情報基盤センターにてサービスを開始したスーパーコンピュータ「不老」の設計を紹介し、また主要なベンチマークプログラムにより得られた性能評価結果を示した。スーパーコンピュータ「不老」は、「富岳」型の計算ノードを世界で初めて本格サービス運用するとともに、全国共同利用のスーパーコンピュータではまだ提供数が少ない Tesla V100 GPU を多数提供し、光ディスクによるコールドストレージシステムをスーパーコンピュータでは初めて導入、さらに湧水を用いた冷却システムを導入するなど、非常に先進的で挑戦的なシステムである。今後も安定したサービス運用を行うと共に、適切な利用方法や性能情報など利用者に役立つ情報を提供し、ユーザによる本システムを用いた研究をサポートするとともに、様々な分野の研究の発展に寄与していきたい。

参考文献

[1] FUJITSU Supercomputer PRIMEHPC FX10 - 富士通 <https://www.fujitsu.com/jp/products/computing/servers/supercomputer/primehpc-fx10/> (accessed 2020-09-28)

[2] FUJITSU Supercomputer PRIMEHPC FX100 - 富士通 <https://www.fujitsu.com/jp/products/computing/servers/supercomputer/primehpc-fx100/> (accessed 2020-09-28)

[3] 田島嘉則, 山田一成, 高橋一郎, 毛利晃大, 片桐孝広, 大島聡史, 永井亨, スーパーコンピュータ「不老」のサービスとエコシステム, 大学 ICT 推進協議会 2020 年度 年次大会 予稿集, 2020

[4] スーパーコンピュータ「富岳」について — 理化学研究所 計算科学研究センター (R-CCS) <https://www.r-ccs.riken.jp/jp/fugaku> (accessed 2020-09-28)

[5] FUJITSU Supercomputer PRIMEHPC シリーズ : 富士通 <https://www.fujitsu.com/jp/products/computing/servers/supercomputer/> (accessed 2020-09-28)

[6] 高橋一郎, 大島聡史, 片桐孝洋, スーパーコンピュータ「不老」における光ディスクライブラリを用いたコールドストレージシステムの構築, 大学 ICT 推進協議会 2020 年度 年次大会 予稿集, 2020

[7] スーパーコンピュータシステム — 名古屋大学 情報連携推進本部 <http://www.icts.nagoya-u.ac.jp/ja/sc/> (accessed 2020-09-28)

[8] MEMORY BANDWIDTH: STREAM BENCHMARK PERFORMANCE RESULTS <https://www.cs.virginia.edu/stream/> (accessed 2020-09-28)

[9] HPL - A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers <https://www.netlib.org/benchmark/hpl/> (accessed 2020-09-28)

[10] Home — TOP500 Supercomputer Sites <https://www.top500.org/> (accessed 2020-09-28)

[11] <https://www.hpcg-benchmark.org/> <https://www.hpcg-benchmark.org/> (accessed 2020-09-28)

[12] MVAPICH :: Benchmarks <https://mvapich.cse.ohio-state.edu/benchmarks/> (accessed 2020-09-28)

[13] hpc/ior: IOR and mdtest <https://github.com/hpc/ior> (accessed 2020-09-28)

表 1 スーパーコンピュータ「不老」の性能諸元 1

Type I サブシステム		機種名: FUJITSU Supercomputer PRIMEHPC FX1000
CPU	型番と数量	FUJITSU A64FX (Armv8.2-A + SVE) × 1 ソケット
	コア数と動作周波数	48 コア +2 アシスタントコア, 2.2 GHz (I/O 兼計算ノードは 48 コア +4 アシスタントコア)
	1CPU あたり理論演算性能	3.3792 TFLOPS (DP), 6.7584 TFLOPS (SP), 13.5168 TFLOPS (HP)
	1 ノードあたりメインメモリ	HBM2 32 GiB
	1CPU あたり理論メモリバンド幅	1,024 GB/s (1CMG=12 コアあたり 256 GB/s, 1CPU=4CMG)
ノード数		2,304
ノード間接続		Tofu インターコネクト D 隣接ノード間 40.8 GB/s × 双方向 (6.8 GB/s/link, 6 link 同時通信可能)
冷却方式		水冷
総理論演算性能		7.782 PFLOPS
総メインメモリ容量		72 TiB

Type II サブシステム		機種名: FUJITSU Server PRIMERGY CX2570 M5
CPU	型番と数量	Intel Xeon Gold 6230 (Cascade Lake) × 2 ソケット
	コア数と動作周波数	20 コア, 2.10 - 3.90 GHz
	1CPU あたり理論演算性能	1.344 TFLOPS (DP), 2.688 TFLOPS (SP)
	1 ノードあたりメインメモリ	DDR4 2933 MHz, 384 GiB (32 GiB × 6 枚 × 2 ソケット)
	1CPU あたり理論メモリバンド幅	140.784 GB/s
GPU	型番と数量	NVIDIA Tesla V100 (Volta), upto 1,530 MHz × 4 枚
	1GPU あたりメモリ	HBM2 32 GB, 900 GB/s
	1GPU あたり理論演算性能	7.8 TFLOPS (DP), 15.7 TFLOPS (SP), 125 TFLOPS (Deep Learning)
	ホストとの接続	PCI-Express Gen.3 ×16 (16GB/s)
	GPU 間の接続	NVLink 2, 1GPU あたり合計 300 GB/s (1GPU から他の 3GPU に対してそれぞれ 50 GB/sec × 双方向)
ノード数		221
ノード間接続		InfiniBand EDR 100Gbps × 2 port
冷却方式		水冷
総理論演算性能		7.489 PFLOPS
総メモリ容量		メインメモリ 82.875 TiB, デバイスメモリ 28.288 TiB
ローカルストレージ		各ノードに NVMe SSD 6.4 TB, 総転送性能 707 GB/s 一部のノードにて BeeGFS, BeeOND, NVMesh を提供

Type III サブシステム		機種名: HPE Superdome Flex
CPU	型番と数量	Intel Xeon Platinum 8280M (Cascade Lake) × 16 ソケット
	コア数と動作周波数	28 コア, 2.70 - 4.00 GHz
	1CPU あたり理論演算性能	2.4192 TFLOPS (DP), 4.8384 TFLOPS (SP)
	1 ノードあたりメインメモリ	DDR4 2933 MHz, 24 TiB (128 GiB × 16 ソケット)
	1CPU あたり理論メモリバンド幅	140.784 GB/s
GPU	型番と数量	NVIDIA Quadro RTX6000 (Turing) × 4 枚
	1GPU あたりメモリ	GDDR6, 24 GiB
	ホストとの接続	PCI-Express Gen.3 x16 (16GB/s)
ノード数		2
ノード間接続		InfiniBand EDR 100Gbps
冷却方式		空冷
総理論演算性能		77.414 TFLOPS
総メモリ容量		48 TiB
ローカルストレージ		各ノードあたり 500TB の外部接続ローカルストレージ 可視化用ノードのみ 104TB の NVMe SSD を搭載

表 2 スーパーコンピュータ「不老」の性能諸元 2

クラウドシステム		機種名: HPE ProLiant DL560
CPU	型番と数量	Intel Xeon Gold 6230 (Cascade Lake) × 4 ソケット (Type II サブシステムと同じ CPU のため詳細は省略)
	1 ノードあたりメインメモリ	DDR4 2933 MHz, 384 GiB (32 GiB × 3 枚 × 4 ソケット)
	1CPU あたり理論メモリバンド幅	140.784 GB/s
ノード数		100
ノード間接続		InfiniBand EDR 100Gbps
冷却方式		空冷
総理論演算性能		537.6 TFLOPS
総メインメモリ容量		37.5 TiB

ホットストレージ

MDS	FUJITSU PRIMERGY RX2540 M5 × 4 台
MDT	FUJITSU ETERNUS AF250 S2 × 1 台
OSS/OST	(DDN SFA18KE × 1 台, DDN SS9012 × 10 台) × 4 セット
ストレージ容量	物理容量 40.32 PB, 実効容量 約 30.44 PB
データ保護	RAID6 (8D+2P), データバックアップなし
転送性能	384 GB/sec (最も転送速度が遅い部分の理論性能)

コールドストレージ

Phase 1 (2020 年 7 月 1 日から)	
機種名	Sony PetaSite 拡張型 Library
搭載カートリッジ数 / 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
搭載ドライブ数	6 台
Phase 2 (2021 年 2 月 1 日から)	
機種名	Sony PetaSite 拡張型 Library
搭載カートリッジ数 / 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	6.0 PB / 10.89 PB
搭載ドライブ数	20 台

表 3 IOR ベンチマークの結果

操作対象	OST	File Per Process 性能		Single Shared File 性能	
		write 平均値	read 平均値	write 平均値	read 平均値
ディレクトリ	クライアント数	[MiB/sec]	[MiB/sec]	[MiB/sec]	[MiB/sec]
/data/group1	576	93,130.36	106,612.15	84,112.94	83,286.16
/data/group2	288	33,770.22	38,685.99	32,984.10	36,369.85
/home	144	17,928.59	22,037.02		
/data/tmp	144	12,697.28	13,242.41		
		write+read 合計値	169,052.000	write+read 合計値	118,381.525