スーパーコンピュータ「不老」コールドストレージと民間利用制度について

○山田 一成 [1)] 田島 嘉則 [1)] 髙橋 一郎 [1)] 毛利 晃大 [1)] 林 秀和 [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

yamada@itc.nagoya-u.ac.jp

Supercomputer "Flow" industry support services and cold storage system

Kazunari Yamada [1)], Yoshinori Tajima [1)], Ichiro Takahashi [1)],

Akihiro Mouri [1)] , Hidekazu Hayashi[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

名古屋大学情報基盤センターでは，令和 2 年 7 月より，スーパーコンピュータ「不老」の運用を開始した．本稿では，令和 3 年 2 月に増強されたスーパーコンピュータ「不老」のコールドストレージの運用状況および，民間利用制度の利用状況について報告する．

1 はじめに

近年，スーパーコンピュータを取り巻く環境は大きく変わってきている．スーパーコンピューティングは従来の計算科学シミュレーション中心から，データ科学，機械学習との融合による新しいスタイルへと移行しつつある．さらに，従来の実験データや観測データなど日々膨大なデータも発生している．研究者がこの増え続ける膨大なデータを収集して活用するためには，大量のデータを長期保存して管理する低コストのストレージが必要になる．また，スーパーコンピュータは学術研究の発展だけではなく，ものづくりの現場など，企業競争力の強化にとって重要な基盤となっている．このような背景から，名古屋大学情報基盤センター（以下，本センター）では，社会貢献の一環として，平成 19 年から文部科学省先端研究施設共用イノベーション創出事業などを実施し，民間企業の課題に対してスーパーコンピュータの資源を提供してきた．さらに，これら事業の終了後においても，民間利用の取り組みを継続しており，自主事業として民間利用制度を推進している．

本センターのシステムもスーパーコンピュータを取り巻く環境に対応するため数年おきに更新しており，令和 2 年 7 月にスーパーコンピュータ「不老」の運用を開始した．スーパーコンピュータ「不老」は 4 つのサブシステムと大容量のホットストレージ群およびコールドストレージシステム，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的なシステムである．

さらに，令和 3 年 2 月にはコールドストレージシステムの増強（以下，コールドストレージシステム Phase2）を実施した．

本稿では，スーパーコンピュータ「不老」のコールドストレージシステム Phase2 の運用状況（システム・サービス概要，利用状況，整備状況，活用事例など）について報告し，民間利用制度の利用状況についても報告する．

2　スーパーコンピュータ「不老」の紹介

本センターでは，令和 2 年 7 月にスーパーコンピュータ「不老」の運用を開始した．図 1 にスーパーコンピュータ「不老」の全体構成を示す．スーパーコンピュータ「不老」はアカデミック利用などの他に民間利用制度の利用者など，すべての利用者が利用可能となっている．

4 つのサブシステムの性能諸元を表 1 に示す．各サブシステムの特徴は以下の通りである．

- Type I サブシステムは超並列大規模計算用であり，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た FUJITSU PRIMEHPC FX1000 により構成される．
- Type II サブシステムは機械学習や AI などの研究分野向けであり，1 ノードにつき NVIDIA Tesla V100 を 4 台搭載した FUJITSU PRIMERGY CX2570M5 により構成される．
- Type III サブシステムは，大容量メモリを使用する可視化処理用のプリポストサーバであり，2 ノードで総メモリ容量が 48TiB の HPE Superdome Flex により構成される．
- クラウドシステムは，前システムの FUJITSU PRIMERGY CX400 の後継として HPE ProLiant DL560 を導入している．バッチ処理以外に，Web システムから時刻を指定しての利用にも対応している．

利用者は目的に合わせ，これらのサブシステムから適切なサブシステムを選択して計算に利用する．

コールドストレージシステムは，国内のスパコンで初となる業務用次世代光ディスクを使用したソニー社製のオプティカルディスク・アーカイブ PetaSite 拡張型ライブラリーシステム（以下，ODA ライブラリ装置）を採用した大容量・低コストのコールドストレージであり，令和 2 年 7 月，総物理容量 484 TB にて運用を開始した．その後，令和 3 年 2 月，総物理容量 6 TB ，最大搭載容量 10.89PB に増強して運用している．

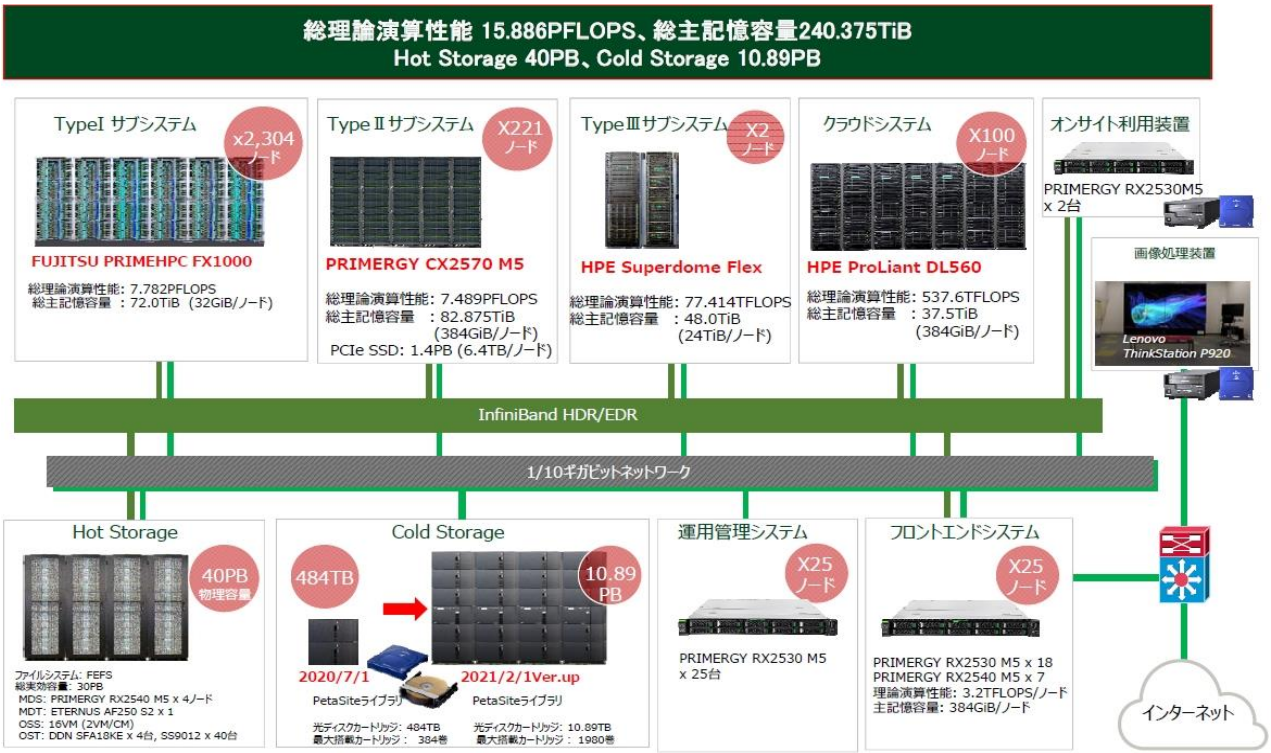

図.1　スーパーコンピュータ「不老」

表 1 計算サブシステム諸元表

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 CX2570MS	HPE Superdome Flex	HPE Proliant DL560
	プロセッサ	A64FX 48+2コア、2.2GHz	Intel Xeon Gold 6230 (20コア、2.10GHz)×2	Intel Xeon Platinum 8280M (28コア、2.70 GHz)×16	Intel Xeon Gold 6230 (20コア、2.10GHz)×4
	GPU	----------	Tesla V100×4 (Volta)	Quadro RTX6000×4	----------
総ノード数		2304 (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		72 TiB	82.875 TiB	48 TiB	37.5 TiB
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

3 コールドストレージシステム概要

令和 3 年 2 月に増強したコールドストレージシステムは 4 つの同じシステムから構成され，1 つのシステムは ODA ドライブ 5 台と光ディスクライブラリ制御サーバ（以下，ODA ライブラリ制御サーバ）を兼ねるログインノード（以下，フロントエンドサーバ）から成る．

コールドストレージシステム Phase2 の構成を図 2 に示す．

ODA ライブラリ装置は，ODA カートリッジを格納するキャビネット，ドライブ，キャビネット内の ODA カートリッジを識別してドライブにセットするロボット機構，ODA カートリッジを ODA ライブラリ装置に収容および搬出する機構で構成される．図 3 に，ODA カートリッジ，表 2 に，ODA ライブラリ装置の諸元を示す．

ODA カートリッジは，データメディアコストが低価格なだけでなく光ディスクならではの後方互換性があるため長期間利用可能であり，温湿度の環境変化に強く水濡れ，紫外線，電磁パルスなど外的影響を受けにくいため，長期データ保全の

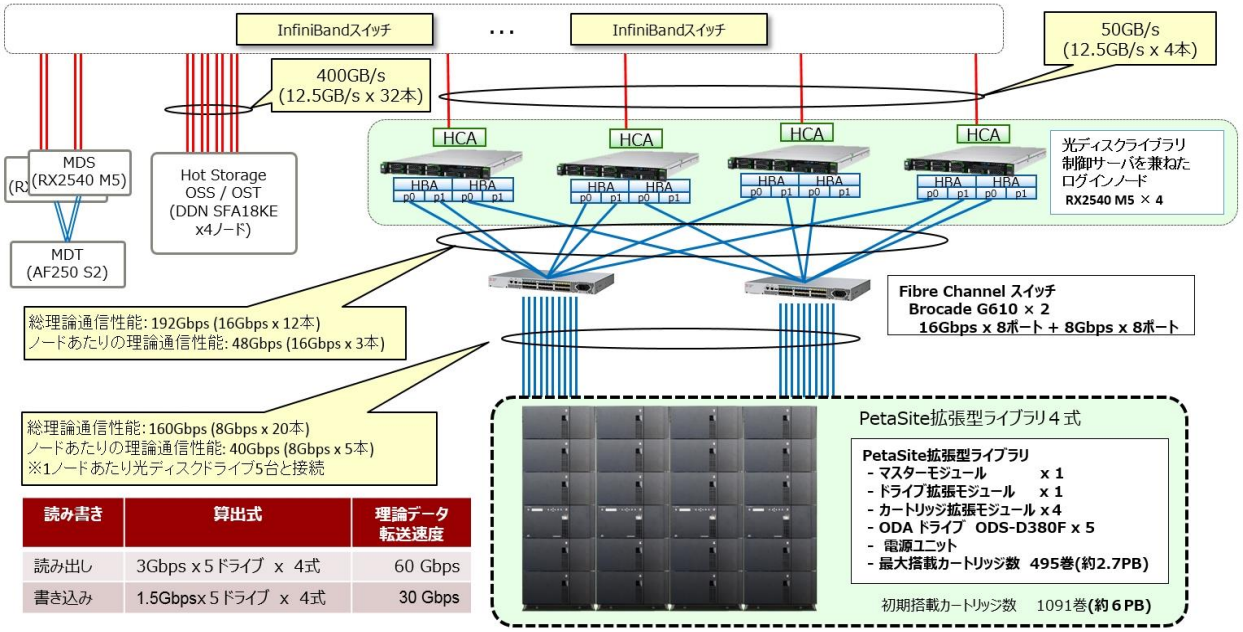

図 2 コールドストレージシステム構成

信頼性が高いのも特徴である．また，利用者はODA カートリッジを持ち込んだり取り出したりして利用することも可能である．

図 3　ODA カートリッジ

表.2　ODA ライブラリ装置の諸元

項目	仕様
機種名	PetaSite 拡張型ライブラリシステム
総物理容量	2.72PB
総スロット数	495
総ドライブ数	5
ロボット機構	1

この ODA ライブラリ装置は，Fiber Channel でフロントエンドサーバを兼ねている ODA ライブラリ制御サーバに接続されている．表 3 に，ODA ライブラリ制御サーバの諸元を示す．

表.3　ODA ライブラリ制御サーバの諸元

項目	仕様
CPU	Intel Xeon Gold6248 (2.5GHz/20 コア)×2
理論演算性能	3.2TFLOPS
主記憶	384GiB(32GiB DDR4×12)
内臓 SSD	システム：2TB ， 作業用：19.2TB
インターコネクト	InfiniBand EDR×1
SAN インターフェース	FC16Gbps×4
Network インターフェース	10G Twinax×2, 1000BASE-T×2

この ODA ライブラリ制御サーバには，他のフロントエンドサーバと同様に，全システム共有のホットストレージが InfiniBand EDRx8 で接続され，富士通 FEFS ファイルシステムでマウントされ高速にアクセスすることができる．

コールドストレージの利用負担金は 1 口 50TB 利用で，初年度はファイル負担金（19 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で利用できる．

3.1　ODA 単体ドライブ

情報基盤センター内の利用者支援室と画像処理室のサーバには，ホットストレージとのデータ転送用に ODA 単体ドライブが USB 接続され，Linux 環境と Windows 環境で利用できる．利用者は ODA カートリッジを持ち込みで利用する．ODA 単体ドライブを図 4 に，利用者支援室の Linux 環境を図 5 に示す．

図 4　ODA 単体ドライブ

図 5.　利用者支援室の Linux 環境

Linux 環境では，GNOME のデスクトップ環境が利用でき，ODA 単体ドライブに挿入した ODA カートリッジは Linux のオートマウント機能を使って自動的にマウントされ，フロントエンドサーバと同様に接続されたホットストレージ間で高速

にファイル転送を行うことができる．

また，Windows 環境では，ベンダー提供の Windows Explorer 相当の GUI を使ったファイル操作が行えるソフトウエアが利用できる．

4　利用形態

コールドストレージシステムの主な利用形態は，次のとおりである．

・1つのODAカートリッジを1ボリュームとして利用者に割当てて名前(ボリューム通番)を付けて，DB を使って管理する．本センター提供のマウントコマンドを使ってマウント操作を行って利用する．
・ODA ライブラリ装置内のカートリッジの記録フォーマットは，外付けの ODA 単体ドライブと同じフォーマットである．このため，ODA ライブラリ装置内で記録したカートリッジを取出して，外付けの ODA 単体ドライブで利用できる．また，外付けの ODA 単体ドライブで記録したカートリッジを，ODA ライブラリ装置に収納して利用する．
・ODA カートリッジの所有者情報やカートリッジの記録情報は DB で管理される．
・利用申請情報を基に利用者が使用できる ODA カートリッジやフロントエンドサーバを限定して利用する．
・サービス終了時，利用者に ODA カートリッジを返却できる．

5　整備状況

コールドストレージシステムの利用拡大と利用者の便宜を図るために，令和3年度は新たに単体ドライブの貸出しサービスや利用者向け コマンドを作成して公開した．

5.1　単体ドライブ貸出しサービス

令和3年７月からスーパーコンピュータ「不老」の利用者向けに単体ドライブの貸出しサービスを開始した．図6に単体ドライブ貸出し機器構成を示す．

単体ドライブを収納するジュラルミン・ケースの外形寸法は1,050mm (270×530×250)，Drive搭載時の重量は約11kgで，搬送に宅配便が利用できる．

図6　単体ドライブ貸出し機器構成

現在，この単体ドライブ貸出しサービスは，手元に大量のデータがある利用者やネットワーク環境が悪い利用者等に貸出している．データを記録したカートリッジは，5.2　カートリッジ持ち込み利用サービスを使って，利用者支援室や画像処理室のODA単体ドライブやODAライブラリ装置に収容して利用できる．

5.2　カートリッジ持ち込み利用サービス

利用者側でODA単体ドライブやODAライブラリ装置を購入してデータを蓄積している場合，このサービスを利用して利用者支援室のODA単体ドライブやODAライブラリ装置に収容して利用することができる．この方式では，データ移行時の作業ミスが防げ，データ転送時間も不要となり，データ移行の負荷とリスクを軽減できる．

現在，コールドストレージシステムでは，ODAの最新の第3世代の「ドライブユニット」と「ライブラリ装置」が利用できるが，後方互換があり旧世代のODA単体ドライブで記録したカートリッジもODAライブラリ装置内で共存して利用することができる．持ち込み利用サービスの場合は，管理費のみで利用することができる．

5.3　利用者向けコマンドの追加

コールドストレージを利用するためのソフトウエアとして，カートリッジのマウント・アンマウント，ドライブ状態表示，バッチジョブ計算依頼コマンド，ファイル情報の管理等のコマンドを提供してきたが，新たに次の機能を追加した．

・　ファイル又はディレクトリを対象としたバックアップ（変更されたファイル又はディレク

トリのバックアップも可能).
- ファイル又はディレクトリを tar コマンドでまとめて非圧縮形式でアーカイブ.
- ディレクトリ同士の比較（ファイル情報とスペースサイズを比較).

6 利用状況と活用事例

コールドストレージシステムPhase2のODAライブラリ装置の初期搭載カートリッジ数は 1092 巻，記憶容量は約 6PB である．ODA カートリッジを追加購入すれは，最大搭載数は 1980 巻，約 10.8PB まで拡張可能である.

令和 3 年 9 月の本稿執筆時では，4 機関と 4 グループ，15 ユーザが，合計 530 巻，約 2.9PB を利用している．また，この他に，第 2 世代の ODA カートリッジ 100 巻, 330TB の持ち込み利用がある.

コールドストレージシステムの主な用途は，アーカイブ又はバックアップである．主な活用事例を，以下に示す.

1. 世界最大級の乱流場の直接数値シミュレーションの解析結果の格納.
2. デジタルユニバーシティ化に向けたデータマネジメントの実証実験として，宇宙地球環境研究所が収集している宇宙地球環境データを格納.
3. 自動運転の映像データの格納．主に観測データが格納されており，その内容は，映像・点群・GPS/IMU センサ・その他車両信号である．点群データの解析結果なども一部含まれる.
4. JHPCNの採択課題 jh170034-ISH「HPCと高速通信技術の融合による大規模データの拠点間転送技術開発と実データを用いたシステム実証試験」で発生する気象衛星ひまわりのデータの格納．日々発生するひまわり衛星のデータは，年間150TB，画像化すると10分あたり741ファイル370MBである.

7 民間利用制度の概要

令和 3 年度，本センターの民間利用制度は，有料の 2 種類と無料の 1 種類があり，スーパーコンピュータ「不老」のすべてのシステムが利用できる．ただし，一部のアプリケーションはアカデミックライセンスのため，利用できない.

・成果公開型（有料）

申し込み後，専門委員会による審議を経て，承認される．採択後，企業名と課題が公開される．基本負担金（1 口，200,000 円，10 ユーザまで利用可）を負担することにより，1 口当たり 100,000 ポイントの計算資源が与えられる.

利用終了後，利用報告書を提出しなくてはならない．利用報告は HP で公開される．最大で 2 年間の公開延期が可能である.

・成果非公開型（有料）

申し込み後，非公開審査ワーキンググループによる審議を経て，承認される．企業名と課題は公開されない.

基本負担金（1 口，400,000 円，10 ユーザまで利用可）を負担することにより，1 口当たり 100,000 ポイントの計算資源が与えられる．利用終了後，利用報告書を提出しなくてはならないが，一般には公開されない.

・トライアルユース（無料）

1 ヶ月間有効のアカウントが発行され，試用する制度である．10,000 ポイントの計算資源が与えられる．通常の有料の利用方法と同様の審査手順を経て，利用承認がなされる.

利用終了後，利用報告書を提出しなくてはならない．利用報告は公開される.

8 民間利用制度の利用状況

平成 26 年度からの民間利用制度（有料）採択数を図 7 に示す．平成 27 年度以降，有料では成果公開型と成果非公開型の 2 種類を実施しているが，ここでは総数を示している．令和 3 年度は 9 月 20 日現在の件数である．また, 平成 27 年度から件数に変動があるが増加傾向であったが，令和 2 年度から採択数が減少している．これは，システム更新のため，アーキテクチャが変更となり，利用を検討している企業が多いためと思われる.

民間利用制度のシステムごとの利用割合（ノード時間積）を表 4 および表 5 に示す．表 4 は，前システムの利用割合．表 5 は，スーパーコンピュータ「不老」の利用割合．令和 3 年度は 9 月 20 日時点での割合である．ノード時間積は，利用ノード数×時間である．前システムのシステムごとの利用割合は，年度によって多少，ばらつきはある

が，両システムとも均等に利用されていたと思われる．スーパーコンピュータ「不老」となり，クラウドシステムの利用割合が多くなっている．

図.7　民間利用制度採択数

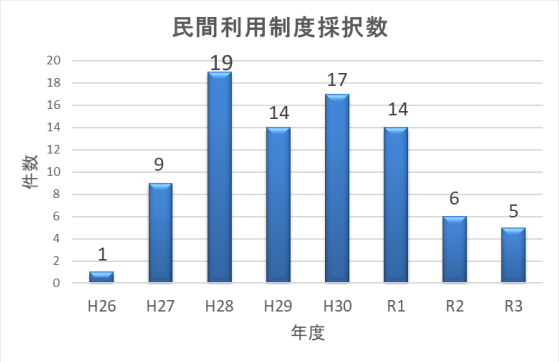

表.4　前システムの民間利用制度の利用割合（ノード時間積）

年度	FX100	CX400
H27	2.97%	3.01%
H28	5.61%	3.89%
H29	4.84%	1.95%
H30	4.40%	4.54%
R1	2.72%	1.79%

表.5　スーパーコンピュータ「不老」民間利用制度の利用割合（ノード時間積）

年度	TypeI	TypeII	TypeIII	クラウド
R2	0.69%	1.43%	0.00%	5.64%
R3	0.08%	0.20%	0.18%	12.23%

9 課題と今後の予定

本稿では，令和 2 年 7 月のスーパーコンピュータ「不老」のリプレースで導入し，令和 3 年 2 月に増強した，コールドストレージシステム Phase2 の概要，利用状況，整備状況，活用事例および，民間利用制度の概要，利用状況について報告した．

コールドストレージシステム Phase2 は，現在のところ大きな問題もなく順調に運用している．今後，利用者数の増加に伴って運用に支障をきたす障害の発生も考えられ，導入ベンダーと連携して安定運用に努めてゆきたい．また，利用者のご意見やご要望を取り入れて，次期スーパーコンピュータへのデータ移行を見据えてより使いやすいシステムへ改良したいと考えている．

民間利用制度については，令和 2 年度から採択数が減少している．これは，先に述べたとおり，システム更新のため，アーキテクチャが変更となり，利用を検討している企業が多いためと思われる．ユーザサポート担当者としては，新型コロナ対応で利用者と直接会っての利用相談がしづらい状況のため，遠隔会議システムなどを活用した利用者の利用準備の支援をさらに充実させ，採択数につなげたいと考えている．

参考文献

[1] 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.

[2] 山田 一成, 田島 嘉則, 毛利 晃大, 髙橋 一郎, スーパーコンピュータ民間利用制度について, 総合技術研究会 2019 九州大学, 2019.

[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」民間利用制度, https://icts.nagoya-u.ac.jp/ja/sc/riyou/industry/