

スーパーコンピュータ AOBA 新サブシステム AOBA-S の紹介

小野 敏¹⁾, 山下 毅¹⁾, 森谷 友映¹⁾, 大泉 健治¹⁾, 高橋 慧智²⁾, 滝沢 寛之²⁾

1) 東北大学 情報部デジタルサービス支援課

2) 東北大学 サイバーサイエンスセンター

ono@tohoku.ac.jp

Introduction of AOBA-S, a new subsystem of supercomputer AOBA

ONO Satoshi¹⁾, YAMASHITA Takeshi¹⁾, MORIYA Tomoaki¹⁾, OIZUMI Kenji¹⁾,
TAKAHASHI Keichi²⁾, TAKIZAWA Hiroyuki²⁾

1) Digital Services Support Division of Information Department, Tohoku Univ.

2) Cyberscience Center, Tohoku Univ.

概要

東北大学サイバーサイエンスセンターは、全国共同利用機関として大規模科学計算システムの整備と、HPCIの資源提供機関としての役割を担っている。本稿では2023年8月から運用を開始したスーパーコンピュータ AOBA の新サブシステム AOBA-S について紹介する。

1 はじめに

東北大学サイバーサイエンスセンター(以下、本センター)では、2020年10月からスーパーコンピュータ AOBA の運用を開始した。スーパーコンピュータ AOBA はサブシステム AOBA-A(SX-Aurora TSUBASA、日本電気株式会社製)、サブシステム AOBA-B(LX 406Rz-2、日本電気株式会社製)の2種類の計算機システムと、ストレージシステム等で構成される。

当初の導入計画では2022年にシステム増強を行う予定であったが、半導体製造に関する世界的な需給状況により、必要な設備の調達が困難であること分かり、約1年間導入計画を順延した。ただ、計算需要は処理能力を大きく超える状況が続いていたため、クラウドサービス[1]を活用し、2022年10月から2023年7月までクラウドサービス AOBA-C[2]として計算環境を提供し、計算機システムの混雑緩和を図ってきた。2023年8月からは当初の導入計画で増強予定であった新しいサブシステムとして、AOBA-S(SX-Aurora TSUBASA、日本電気株式会社製)の運用を開始した。

図1に本センターの計算機室に設置されたサブシステム AOBA-S の外観写真を示す。また、図2に現在のスーパーコンピュータ AOBA のシステ

ム構成図を示す。

本稿では、新たに運用を開始したサブシステム AOBA-S の構成、利用者環境、高速化支援活動について紹介する。



図1 AOBA-S の外観写真

2 AOBA-S の構成

AOBA-S は、演算サーバ、ストレージシステム、ネットワーク装置のハードウェア群、およびプログラミング環境、利用者管理と課金統計管理のソフトウェア群で構成される。

2.1 ハードウェア

SX-Aurora TSUBASA は、AOBA-A と同じベクトルアーキテクチャを採用している。アプリケーション演算処理を行うベクトルエンジン(以下、

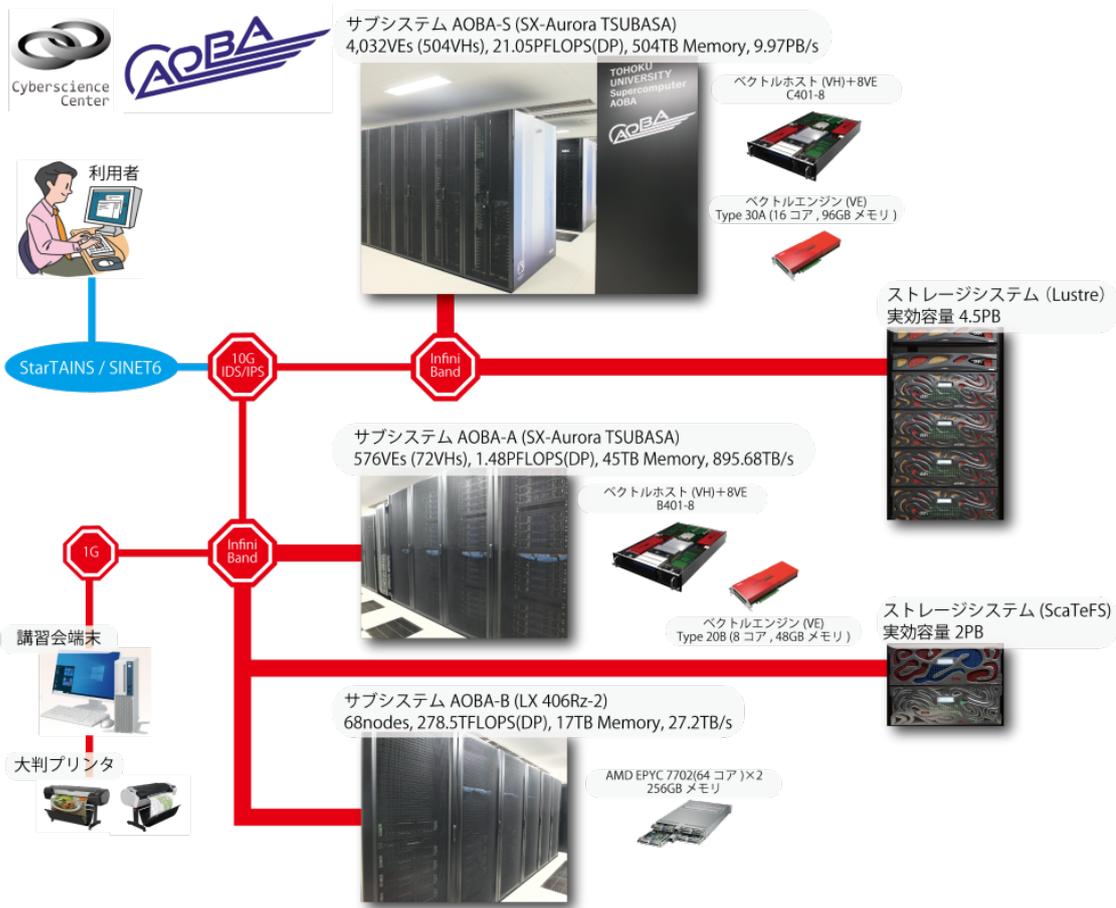


図2 スーパーコンピュータ AOBA の構成図

VE)部と、主に OS 処理を行うベクトルホスト(以下、VH)部により構成される。PCIe カードに搭載される VE 部はベクトルプロセッサ、および高速メモリから構成され、x86/Linux である VH と PCIe 経由で接続される。

今回導入した VE(Type 30A) は、新規に開発された第三世代の VE で、SX-Aurora TSUBASA の心臓部である。理論演算性能4.91TFLOPS(倍精度)となるマルチコア(16 コア) ベクトルプロセッサを1台、主記憶は 96GB の HBM2e メモリを搭載し、2.45TB/s という高メモリバンド幅でプロセッサと接続されることで、高い演算性能とメモリ性能の両立を実現している。VE の性能評価については文献[3]を参照されたい。本センターの AOBA-S は、1 台の VH と 8 台の VE が構成単位となる C401-8 モデルを採用し、サブシステム全体では 504 台の VH と 4,032 台の VE で構成される。AOBA-S を構成する C401-8 の外観写真を図3 に示す。

VE と VH を合わせたシステム全体の理論演算性能は、21.05PFLOPS(倍精度)、総主記憶容量は

504TB、総メモリバンド幅は 9.97PB/s となる。504 台の VH は InfiniBand NDR ネットワークで接続され広帯域・低遅延なデータ通信を可能としている。

ストレージシステムは、DDN 社製の Lustre ファイルシステムを採用し、高速アクセスかつ高密度ストレージである DDN ES400NVX2(DDN 社製)を導入した。AOBA-S 利用者専用のホーム領域として利用し、RAID6 で構成され、実効容量は 4.5PB である。



図3 AOBA-S を構成する C401-8 の外観写真

2.2 ソフトウェア

プログラミング言語は AOBA-A と同じく、アプリケーションの実効性能を向上させる高度な自動ベクトル化・自動並列化機能を備えた Fortran/C/C++コンパイラが利用できる。自動並列化機能および OpenMP による共有メモリ並列実行と、システム構成に最適化された MPI ライブラリによる、分散メモリ並列実行が可能である。また科学技術計算ライブラリとして、VE に最適化された数学ライブラリコレクション NEC Numeric Library Collection(NLC)が利用可能である。

AOBA-A で動作していたプログラムを AOBA-S で実行する場合には、Type 30A に最適化された実行モジュールが必要なため、そのプログラムを再度コンパイルする必要がある。

AOBA-S では、オープンソースソフトウェア (OSS) の Quantum Espresso やオープンソースのコンテナ型プラットフォームである Singularity、第一原理計算コードの一種である FPSEID²¹ が利用できる。また、今後も VE 向けに移植されたアプリケーションを拡充する予定である。

3 利用者環境

表 1 に AOBA-S 用の利用者向けサーバ名とホスト名、および用途を示す。

表 1 AOBA-S 用の利用者向けサーバ

サーバ名	ホスト名	用途
フロントエンドサーバ	sfront.cc.tohoku.ac.jp	コンパイル、ジョブ投入、小規模データ転送
データ転送サーバ	sfile.cc.tohoku.ac.jp	ローカル PC、既存システムとの大規模データ転送
フロントエンドサーバ(HPCI 用)	shpcif.cc.tohoku.ac.jp	HPCI 課題用のフロントエンドサーバ

フロントエンドサーバ上でプログラムをコンパイルし、演算サーバへバッチリクエストとして投入しプログラムを実行する形態は、既存の AOBA-A および AOBA-B システム(以下、既存システム)と同じである。ただし、フロントエンドサーバは AOBA-S 用のサーバを利用する。また、利

用者のファイルを置くストレージも AOBA-S 用のストレージシステムを利用する。

AOBA-S でプログラミングの際は、あらかじめプログラムファイルやデータファイルを AOBA-S 用のストレージシステムへ転送しておく必要がある。

3.1 ログイン認証

AOBA-S 用のフロントエンドサーバ、および AOBA-S 用のデータ転送サーバへのログイン方法は、既存システムと同様に公開鍵認証方式による SSH 接続を採用している。利用者の公開鍵は既存システムで利用していたものと同一としているため、ローカル PC に保存済みの秘密鍵とパスフレーズによるログインが可能であり、既存システムの利用者は改めて鍵の作成をする必要はない。新規利用者は、本センターウェブサイト上に提供される利用者ポータル「SSH 公開鍵登録」機能を用いてログインに必要な秘密鍵を作成することで、既存システム、AOBA-S の両システムへログインすることが可能である。また利用者の利便性を重視し、利用者のローカル PC から直接 AOBA-S 用のフロントエンドサーバへログイン可能としている。利用者は AOBA-S 用のフロントエンドサーバ上でソースコードのコンパイルやバッチリクエストの投入を行う。

3.2 ファイル転送

AOBA-S のストレージシステムは既存システムとは別システムとなる。そのため AOBA-S で実行するプログラムファイルやデータファイルは、あらかじめ転送しておく必要がある。

小規模なデータ転送は、フロントエンドサーバ、大規模なデータ転送はデータ転送サーバの複数構成としており、転送方法としては Windows、macOS のターミナルアプリケーションおよび Linux 上でのコマンドによる操作と、SCP(Secure Copy Protocol)や SFTP(SSH File Transfer Protocol)に対応した FTP クライアントアプリケーションを用いる 2 つの方法がある。既存システムとの間でデータを転送する方法は、AOBA-S 用のデータ転送サーバからコピーコマンド(cp) 等によるデータ転送が可能である。

3.3 ジョブの実行

フロントエンドサーバでコンパイル作業を行って作成したプログラムの実行は、バッチ処理と呼ばれる方法で計算機に実行を依頼する。本センターではバッチ処理に NEC Network Queuing

System V（以下、NQS_V）を採用している。

AOBA-S のバッチリクエストの投入は、AOBA-S 用のフロントエンドサーバで行い、既存システムとは別システムとなっている。またバッチリクエストの基本的な操作手順については、既存システムと同じである。表 2 に AOBA-S のキュー構成を示す。利用者の利便性を考慮し、既存システムと同様に最大経過時間を規定値 72 時間、最大値 720 時間として長時間のリクエストを実行可能とした。利用 VE 数が 1 台という制限はあるが、1 時間以内の実行を無料とするデバッグ用のキューも準備している。また、qlogin コマンドにより会話型キューを利用することで、VE 向けプログラムの会話型実行も一部のノードで可能である。

3.4 利用負担金

スーパーコンピュータ AOBA の基本利用負担金表を表 3 に示す。この表は大学・学術利用に適用され、民間企業利用は成果公開型の場合で本表記載の金額の 2 倍、成果非公開型の場合で本表記載の金額の 4 倍となる。共有利用の利用負担額単価は、課金対象時間あたり 100 円とし、後払いの従量制と負担金を前払いすることで一定の課金対象時間まで利用できる定額制がある。繁忙期など他の研究グループのジョブによる影響を受けない方法として、3 ヶ月単位での占有利用も準備している。

4 高速化支援活動

本センターでは 1997 年より、ユーザアプリケーションの高精度化、大規模化の支援を目的とした高速化支援活動を、また 1999 年より共同研究制度を実施している。利用者、計算機科学を専門とするセンター教員、技術職員、およびベンダー技術者が連携してアプリケーションの高速化に取り組んでいる。図 4 に 1999 年から本センターで取り組んでいるセンター独自の共同研究、学際大規模情報基盤共同利用・共同研究拠点(JHPCN) 課題および革新的ハイパフォーマンス・コンピューティング・インフラ(HPCI)課題採択数の推移を示す。本センター独自の共同研究は恒常的に年 10 課題ほど実施されていることがわかる。近年では HPCI を介した研究課題数が少ない状況であるが、システム更新が遅れ、必要十分な計算機環境の提供が遅れたことも一因ではないかと考えられる。今後はセンターの共同研究を通してユーザアプリケーションを高度化・大規模化し、JHPCN、

HPCI 採択課題へとステップアップしていくように、我々の高速化支援活動を継続的に続けていきたいと考える。

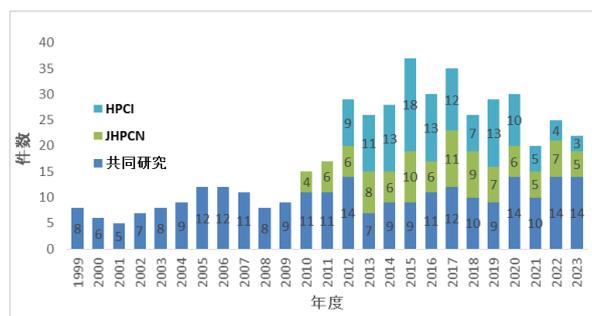


図 4 課題採択件数

5 おわりに

本稿では 2023 年 8 月に運用を開始した、サイバーサイエンスセンターのスーパーコンピュータ AOBA の新サブシステム AOBA-S の構成、利用者環境、高速化支援活動について紹介した。研究室のサーバでは実行できなかったアプリケーションやアイデアを実現する研究の強力なツールとして、最新鋭の AOBA-S を含むスーパーコンピュータ AOBA をご活用いただければ幸いである。今後も本システムを用いた研究をサポートし、様々な研究分野の発展に寄与していきたいと考える。

各システムの利用法の詳細、本センターからのお知らせ、問い合わせ、利用相談および高速化の依頼方法などについては本センターのウェブサイト [4]を参照いただきたい。

参考文献

- [1] https://jpn.nec.com/press/202204/20220425_01.html.
- [2] 大泉 健治, 森谷 友映, 佐々木 大輔, 齋藤 敦子, 山下 毅, 小野 敏, 滝沢 寛之, クラウドサービス AOBA-C の紹介, 大学 ICT 推進協議会 2022 年度年次大会, 2022 年 12 月.
- [3] Keichi Takahashi, Soya Fujimoto, Satoru Nagase, Yoko Isobe, Yoichi Shimomura, Ryusuke Egawa, and Hiroyuki Takizawa, Performance Evaluation of a Next-Generation SX-Aurora TSUBASA Vector Supercomputer, ISC High Performance 2023: High Performance Computing.
- [4] <https://www.ss.cc.tohoku.ac.jp/>

表2 AOBA-Sのキュー構成

利用形態	キュー名	VE数※	実行形態	最大経過時間 既定値/最大値	メモリサイズ
無料	sxsf	1	1VE	1時間/1時間	96GB
共有	inter	1~8	会話型	1時間/1時間	96GB×VE数
共有	sxs	1	1VE	72時間/720時間	
		1~2048	8VE単位で確保		
占有	個別設定				

表3 基本利用負担金【大学・学術利用】

区分	項目	利用形態	負担額及び課金対象時間
演算 負担経費	AOBA-S	共有 (無料)	利用VE数 1(実行数、実行時間の制限有) 無料
		共有 (従量)	課金対象時間 = (利用VE数÷8を切上げた数) × 経過時間(秒) 課金対象時間 1時間につき 100円
		共有 (定額)	負担額 10万円につき課金対象時間 1,000時間分使用可能
		占有	利用VE数 8 利用期間 3ヶ月につき 216,000円
	AOBA-A	共有 (無料)	利用VE数 1(実行数、実行時間の制限有) 無料
		共有 (従量)	課金対象時間 = (利用VE数÷8を切上げた数) × 経過時間(秒) 課金対象時間 1時間につき 75円
		共有 (定額)	負担額 10万円につき課金対象時間 1,400時間分使用可能
		占有	利用VE数 8 利用期間 3ヶ月につき 162,000円
	AOBA-B	共有 (従量)	課金対象時間 = 利用ノード数 × 経過時間(秒) 課金対象時間 1時間につき 22円
		共有 (定額)	負担額 10万円につき課金対象時間 4,600時間分使用可能
		占有	利用ノード数 1 利用期間 3ヶ月につき 47,000円
	ファイル 負担経費	AOBA-S 固有領域	5TBまで無料、追加容量 1TBにつき年額
AOBA-A, B 共有領域		5TBまで無料、追加容量 1TBにつき年額	3,000円
出力 負担経費	大判プリンタによるカラープリント		フォト光沢用紙 1枚につき 600円
			クロス 1枚につき 1,200円

備考

- 負担額が無料となるのは専用のキューで実行されたものとし、制限時間を超えた場合は強制終了する。
- 演算負担経費の課金対象時間については半期毎(4月から9月及び10月から3月)に合計し、1時間未満を切上げて負担金を請求する。
- 演算負担経費について定額制を選択した場合は AOBA-A 及び AOBA-B を課金対象時間の範囲内で共用できる。
- 占有利用期間は年度を超えないものとし、期間中に障害、メンテナンス作業が発生した場合においても、原則利用期間の延長はしない。
- ファイル負担経費については申請日から当該年度末までの料金とする。運用期間が1年に満たない場合は、月割りをもって計算した額とする。占有利用に申込した場合は 10TB まで無料とする。