

統数研 スパコンの活用による 共同利用・共同研究の利用促進

Supercomputer System
for Data Assimilation

統数研は統計学を専門に研究する唯一の研究機関として、統計科学の研究に必要な不可欠な大型計算機を保有し、国内の研究者に広く利用されてきました。これは大学共同利用機関としての重要なミッションです。統数研では平成26年度、タイプの異なる3つのスーパーコンピュータを順次導入しました。4月から共有クラウド計算システム（C）や世界最大規模の共有メモリ領域を持つデータ同化スパコン（A）が導入され、7月からは最新の分散メモリ型スパコン（I）が稼働開始しました。これら3つのスパコンは愛称を一般公募し、統数研元所長故赤池弘次先生の業績にもちなみAICと名付けられました。URAステーションの重要なミッションの一つにこれらの計算機の利用促進があります。大学共同利用機関ならではの研究力強化の一つとして今後一層の共同利用、企業との共同研究、研究者コミュニティへの支援を行っていきます。



統計科学スーパーコンピュータ

Supercomputer System
for Statistical Science

「I」は、分散記憶型のスパコンです。平成26年7月に稼働開始していて、ノードあたりの演算速度が490GFlops以上、主記憶容量が128GB以上でシステム全体での演算速度は196TFlops以上（1年以内に336TFlops以上に増速予定）、主記憶容量は50TB以上（1年以内に100TB以上に増量予定）という仕様になっています。また、システムには統計モデルのシミュレーションには不可欠な物理乱数発生装置や視覚化システムが組み込まれます。統計数理研究所では、これまでも共同研究・共同利用のための計算資源として最新のスパコンを活用してきました。「I」もこれまでに以上に統計数理の基礎研究での貢献が期待されています。



統計科学スーパーコンピュータシステム

データ同化スーパーコンピュータ

「A」は、シングルシステムとして世界最大の64TBの共有メモリを搭載したSGI社製大規模共有メモリ型サーバを中核とするシステムです。本システムでは、この「SGI UV 2000」を2台構成で構築し、システム合計5,120プロセッサコア、総メモリ容量128TBという共有メモリ型としてはこれまでにない規模のスパコンになります。データ同化は、気象学、海洋学の分野で生まれたシミュレーション手法で近年急速に研究領域が広がっています。「A」においても広大なメモリ領域を活用した局地天気予報モデルや大規模な個人化データを用いた需要分析のような次世代のシミュレーション手法の開発での利用が期待されています。また、「A」は大学共同利用機関としては初の試みとして、全国の主要な大学・研究機関が保有するスーパーコンピュータを高速ネットワークで結んだ共同計算環境「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（<https://www.hpci-office.jp>）」に計算資源を提供します。



データ同化スーパーコンピュータシステム

HPCI（High Performance Computing Infrastructure）とは

「京」と全国の大学や研究機関に設置されたスパコンを高速ネットワークで結んだ「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」（<https://www.hpci-office.jp>）です。国内の主要なハイパフォーマンスコンピューティング（HPC）リソースを効率よく利用できるように仕組みと体制が構築されています。特にHPCIはSSO(Single Sign On)と呼ばれる機能を実現していて、HPCIユーザは一つの資源提供機関で一度認証処理を実行すればHPCI内の利用権のあるリソースをシームレスにアクセスできます。HPCIはこのような横断的な利用方法の提供と全国規模でニーズとリソースのマッチングを行い萌芽的研究から大規模研究まで幅広いHPC活用を実現しています。統数研では本年度7月より計算資源として「A」の計算ノード、128ノード（49.152TFLOPS）を提供予定です。

スパコンの愛称が決まりました。

統数研では今年度新たに導入するスパコンをよりみなさんに親しんでもらえるように愛称を募集しました。（期間：平成25年12月25日～平成26年1月20日）所長、及び、スーパーコンピュータシステム仕様策定委員会（以下委員会）による厳正なる審査の結果、下記のように愛称を決定いたしました。
データ同化スーパーコンピュータシステム 愛称：「A」
理由：Assimilation（同化）、Advanced（高等の、上級の）から
統計科学スーパーコンピュータシステム 愛称：「I」
理由：Intelligence（知能）、Investigate（研究する）から
共有クラウド計算システム 愛称：「C」
理由：Cloud（クラウド、雲海）、Community（共同社会）から
各システムの愛称が、アルファベット1文字と覚えやすく、外国人の方にも呼びやすい愛称となりました。今後、本研究所ではこの愛称をもとに様々な広報活動を実施していきます。



共有クラウド計算システム

Communal Cloud
Computing System

共有クラウドシステム

「C」は、Dell社製サーバ69ノード（合計138CPU、1380コア、16.4TBメモリ）、大規模共有ディスク装置（合計364TB）を中心として構成されます。クラウドミドルウェアを導入し、統計解析のための仮想環境を利用者に提供します。この仮想環境にはデータマイニングや機械学習による分析がすぐに実行できるように並列パッケージが組み込まれたRやHadoop、Mahoutがあらかじめ構築されています。また、広大な記憶領域を活用した分散データベースによる大規模テキストデータ処理等の利用も可能です。「C」は、デスクトップ環境でのプログラム資産をシームレスに移行できることから、これまでに以上に共同利用研究や教育分野での利用が期待されています。



統計数理研究所
The Institute of Statistical Mathematics

運営企画本部 URAステーション

ask-ura@ism.ac.jp