

V201a 国立天文台共同利用 大規模観測データ解析システムⅠ概要と設計

古澤 久徳、磯貝 瑞希、巻内 慎一郎、田中 伸広、大倉 悠貴、高田 唯史 (国立天文台)

近年、すばる望遠鏡の Hyper Suprime-Cam (HSC) をはじめとする観測装置や観測プログラムの大規模化が進み、それらから得られるデータサイズも飛躍的に増大している。多くの観測計画が国際的な協力と競争の枠組みの中で行われるようになり、科学成果獲得のためには、適切な時間内に取得データを処理し研究に必要な情報を得ることがますます重要になっている。一方で、取り扱わなければならないデータサイズは生データの状態で数十テラバイトの規模、出力に至っては数十テラバイトからペタバイトの規模に及び、もはや個人研究環境で対応することは容易ではない。こうした状況において、大規模な観測データ解析のための共同利用計算機環境の提供が観測者から強く望まれてきた。

そこで、国立天文台天文データセンターでは2018年度から大規模観測データ解析に最適化した共同利用計算機システム（大規模システム）の構築を始めた。これまでも、我々は個々の観測者が小中規模の観測データ処理を行うための環境を提供してきたが、各ユーザーが多コアを長時間占有することは想定されておらず、HSCなどの1観測プログラムで大きな計算機資源を要するデータ解析の要求に十分に対応できなかった。大規模システムはこうした要求に応えるため、1000コア以上を持つ計算機クラスターと高速な大容量共有ファイルシステムを接続し、バッチシステムによるジョブ管理を行うことで現行の共同利用環境で対応できなかった規模の解析を受け入れる。

初期運用の課題はHSCの戦略枠観測と一般共同利用観測のデータ解析を本システムで実行できるようにすることであり、2019年前半の初期運用試験を目指してハワイ観測所と共同でシステム構築や運用の準備を進めている。