

## V211b HSC カタログデータ検索高速化に向けたデータベースの応用

古澤順子, 高田唯史, 古澤久徳, 山田善彦, 大倉悠貴 (国立天文台), 鬼塚真, 伊藤竜一, 松本拓海 (大阪大学), 須賀秀和, 黒澤亮二, 神林飛志 (ノーチラス・テクノロジーズ)

天文観測技術は年々進歩し観測データおよび処理済データが大量に生成され続けている。これら観測データを取り扱うアーカイブシステムは汎用性の高いリレーショナルデータベース管理システム (RDBMS) を使用することが多いが、近年大量に生成されるデータに対しては、データトランザクションや複雑な分析の時間をいかに短縮するかが重大な課題となってきている。実際、すばる望遠鏡の HSC によって生成された科学的データ量は既に肥大化しており、現在もデータベースの分散化などによる性能改善を試みてはいるが、最終的な 3000 パラメータ以上の測定情報を展開した数百億行の表への応用を見越しては、さらなる技術開発を続ける必要がある。

現在開発中のクエリエンジンは、ビッグデータを処理するためのコンピューティングインフラストラクチャを効率的に使用するように設計されている。この新データベースを評価するには、科学アプリケーションのデモンストレーションが不可欠であり、その一分野として天文研究の観点から、現在 HSC による戦略枠観測データ (SSP PDR2) カタログでクエリのパフォーマンスを検証している。その一環として実際の操作から収集されたクエリワークロードに基いたデータベーススキーマを最適化する手法の確立を目指しており、また系統的な変光天体探査を実現するため、PDR2 データを基に時系列データの測定情報を構築し、一連の天体の明るさや形状などの測定値から外れ値を効率的に検索するエンジンも開発中である。

本講演では、クエリワークロードを用いた最適スキーマの推薦について実体化ビューを利用したクエリ最適化前後の性能差の検証試験、および時系列測定データベースの構築と外れ値検出に関する開発状況を報告する。