

V227b HSC-SSP データベースの高速化検討

瀧田怜, 古澤久徳, 田中賢幸, 高田唯史, 山田善彦, 小池美知太郎, 峯尾聡吾, 大倉悠貴, 池田浩之, 宮崎聡, 岩田生 (国立天文台), 他 HSC データ解析ソフトウェアチーム

我々は、すばる望遠鏡に搭載されている超広視野主焦点カメラ (HSC: Hyper Suprime-Cam) を用いた戦略枠観測 (SSP: Subaru Strategic Program) を 2014 年より行っており、得られたデータを HSC データ解析用パイプライン (hscPipe) で処理し、解析済みデータや天体カタログを SSP 共同研究者および一般利用者向けに公開している。このうち天体カタログを取めたデータベースはバックエンドとして PostgreSQL を用いているが、最新のデータでは天体数が 7 億超となっており単一のデータベースサーバでは検索にかかる時間コストが長いものでは 2 時間程度と非常に大きくなっていった。そこで検索効率を改善するために、データベースを複数の計算機に分散するシャーディングという手法に着目し、Citus (<https://www.citusdata.com>) を利用することにした。これは PostgreSQL の機能拡張として提供されており、ユーザー側は従来の SQL と同様の利用環境のままスケールアウトが可能である。我々は 8 台の計算機にデータベースを分散してパフォーマンスの試験を行った結果、メモリキャッシュのない状態では倍程度、キャッシュの乗った状態では最大で十倍程度の速度向上を確認している。より多くのサンプルを使ってシャーディングによる性能評価を行うために、共同研究者向けに試験を継続中である。また、カタログデータは 5000 個程度のカラムを含んでいるが PostgreSQL の 1 テーブルに作成できるカラム数の制限のために複数のテーブルに分割する必要があり、ユーザー側で JOIN をする等の工程が必要となっている。そこでこのような工程の軽減のために、よく利用されているカラムだけに絞った厳選テーブルの作成も行った。