

天文学の大規模データへの応用

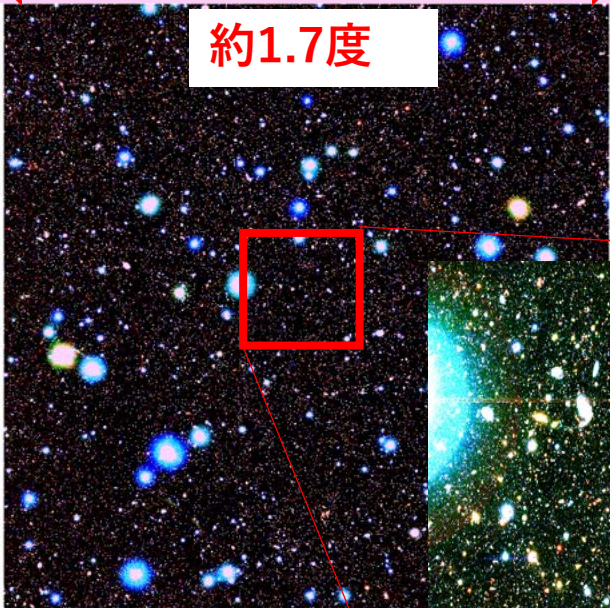
高深度宇宙観測データ処理DB

高田唯史、古澤久徳、古澤順子、大坪貴文
(国立天文台天文データセンター)

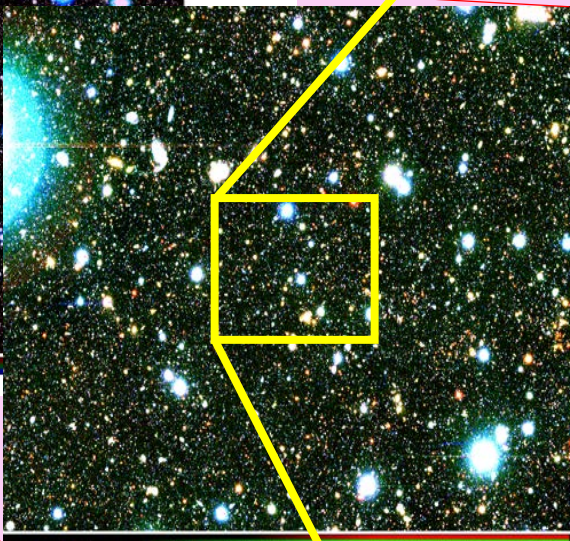
高速データベースと天文学の大規模データ

- 口径8メートルの光学望遠鏡（すばる）に装着された超広視野カメラ（Hyper SuprimeCam:HSC）による深宇宙探査観測による宇宙の進化の研究
 - ◆ 1000平方度を超える広い視野の観測（視野数は1000以上）
 - ◆ 各露出により50～100万程度の天体（+ α ）に関する明るさ、位置、形の測定値
 - ◆ 測定値に関連するパラメータは数百～数千
 - ◆ 露出数は10～数百程度（同じ天体に対する時系列情報）
 - ◆ 波長方向には5つ（または10以上）のフィルター
- 高速検索する必要のあるデータベーステーブルの規模は
レコード数：（観測する視野の数）×（露出あたりの天体数）×（露出数）×（フィルター数）
カラム数：測定パラメータ数
- **予定では、**
 - ✓ 8億以上の天体に対して0（ 10^3 ）個のカラム×5種類以上の波長情報
 - ✓ 300～400億天体（延べ数）で1000個程度のカラム×5種類の波長情報のようなテーブルに対する高速検索が必要になる。

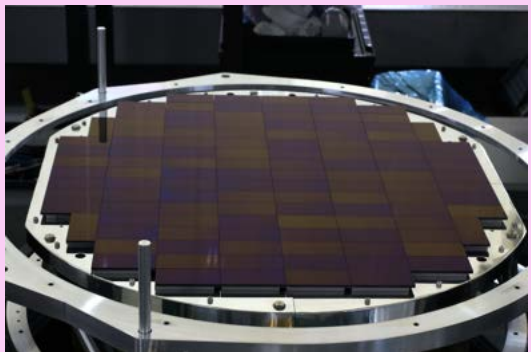
約1.7度



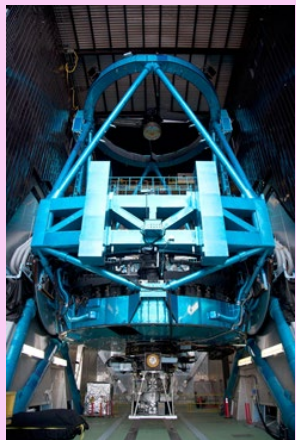
1領域で約13億
ピクセルの画像



約15分角



超広視野カメラ HSC



すばる
望遠鏡

約2分角 (0.033度)



-3.07e-13 -9.46e-14 1.20e-13 3.32e-13 5.46e-13 9.70e-13 1.18e-12 1.40e-12

Tsurugiプロジェクトにおける天文データの応用

- **現在のPostgreSQLクラスターでは遅い（遅すぎる）クエリーの高速化**
 - ✓12時間以上かかるようなクエリーを高速化して、研究の高効率化
 - ✓クエリーログ解析によるクエリー高速化情報の構築（スキーマ推薦）
 - ✓今までできなかった検索結果を用いたトライ&エラー・チャレンジ
- **巨大テーブルを用いての時間軸方向の探索的データ分析**
 - ✓時系列データ作成による変光天体や移動天体の検出、それによる科学研究
 - ✓数十から数百回の観測・測定->レコード数が多すぎて既存DBエンジンでは苦しい
 - ✓変光天体（明るさの変化する天体）の（自動）分類
 - ✓多数（数百万超）の巨大ブラックホール天体による観測的宇宙進化の研究
 - ✓重力波天体・超新星等の突発天体検出と即時アラート発信システムにつなげる
 - ✓移動天体の検出 -> 既存天体の軌道情報改善、新天体の発見（危険小惑星） + α

技術的な観点での方針と現在地

- Asymmetrical HTAPのプロトタイプとして開発中
Asymmetrical HTAP/Tsurugi(OLTP node)-Spark (OLAP node)の構成
- プロジェクト進捗
 - 以前/PostgreSQLクラスターでのシステム
 - プロジェクト現時点/Sparkクラスターへの移行完了・検証中
 - プロジェクト目標/Tsurugi-Sparkへの移行・データの逐次性の確保

現時点アプリケーション

- そもそもちゃんと高速化が可能なのか？
PostgreSQL→Sparkによる高速化（現時点で完了）