

V242c 天文データ・アーカイブをクラウド・コンピューティングにより安価に構築する方法の検討

江口智士（福岡大学）

Atacama Large Millimeter/submillimeter Array、すばる望遠鏡の Hyper Suprime-Cam、Large Synoptic Survey Telescope、Square Kilometre Array など、天文学分野においてもビッグ・データの波が押し寄せている。国際連携の大型プロジェクトでは、自前で超高性能なハードウェアを調達し、その性能を限界まで引き出すソフトウェアを一から開発することができる。けれど通常は観測装置の開発で手一杯であり、ハードウェア/ソフトウェアの両面でデータの公開系は後回しになりがちである。また、想定外の運用期間の延長により、計算機資源の後付が必要になることもある。いっぽう現在では、その時々に必要な計算機資源をクラウド・コンピューティングという形で動的かつ安価に手に入れることが可能になった。本研究ではクラウド・コンピューティングの雄である Amazon Web Services に注目し、この上で動く分散処理フレームワーク “Hadoop” および SQL ライクなデータベース・システム “Hive” を用いて非常に原始的な巨大天文データ検索システムを構築し、クラウド・コンピューティングを天文学分野に適用する際に必要となる技術の基礎的調査を行った。通常のリレーショナル・データベースと異なり、Hive には「パーティション」と呼ばれる、通常ファイル・システムにおけるディレクトリに相当する概念が存在し、この設計がシステムの性能を大幅に左右する。将来的に様々な波長のデータを格納することを前提に、天球を等立体角に分割するアルゴリズム “HEALPix” をパーティション分割に用いたところ、分割パラメータ $N_{\text{side}} = 2^6$ が最適と判明した。また、Hive のファイル形式に Optimized Row Columnar 形式を用いることで、データベースの大幅なサイズ削減と大幅な検索性能の向上を両立させた。