

Data Intensive Astronomy に向けた取り組み

国立天文台・天文データセンター

大石 雅寿

[アブストラクト]

ALMA やすばる望遠鏡の新世代カメラ(HSC)は膨大な量の観測データを生み出す。世界でも数 100 ペタバイトを超える大量データを生み出すサーベイ計画や、テラビットを越えるデータフローを必要とする装置建設が進んでいる。講演では、天文学研究の新たな時代である「データセントリック」な天文学の構築に向けた取り組みをかいつまんでご紹介する。

[キーワード]

データ爆発、巨大ストレージ、高速データ処理、統計処理、パラダイム変革

はじめに

天文学は、人類にとっての最初の学問と言われる。自らの目によって星々を観察し、その動きを調べるようになり、やがて、天体の動きから暦を作成できる専門家である天文学者が王室などのお抱え学者として活躍した。ガリレオが望遠鏡を用いて観測を行って以降、ティコ・ブラーへによる惑星の動きの精密な観測データを解析したケプラーは有名なケプラーの法則を見だし、地球は宇宙の中心から追いやられた。一方、ケプラーの法則を物理学的に説明できるニュートン力学は大成功を収め、理論的に予測された新惑星である天王星や海王星が発見され、理論研究が観測研究と両輪の役割を持つようになった。第 3 のパラダイムであるシミュレーションについては、本会合への参加者の方々はよくご存じであろう。もはや科学研究をシミュレーションなしで語ることはできない時代となった。

天文学者は、より高感度な観測をより高速に実施し、また、多波長データ同士を比較することを通じて宇宙に関する知見を拡大できることを理解していた。シミュレーションは IT の発展がもたらしたパラダイムシフトであったが、その IT がデータ爆発を生み出しつつある。(良い意味で) どん欲な我々は、最近の IT 技術が、多数の天体を観測し、統計的な研究ができる日を待ち焦がれていたのである。

世界でのデータ共有 - ヴァーチャル天文台

天文台は世界の様々な場所にあると共に、地上から観測できない波長帯（赤外線の一部、X線など）は衛星によってデータを取得する。このため、世界各地の天文台や宇宙機関が保有する観測データを共有するためには技術的な壁があった。特に、データを記述するデータであるメタデータについては記述方式が定まっていなかった。観測データそのものは天文用の FITS 形式で記述されるため、データ発見方法とデータ送受信プロトコルを標準化すれば、ネットワーク網をデータ検索対象に早変わりすることになる。

これを実現するのがヴァーチャル天文台システムであり、標準プロトコルを国際ヴァーチャル天文台連合 (IVOA, <http://www.ivoa.net/>) が定めている。IVOA には、日本も含め、18 の国や機関が参加し、IVOA プロトコルに準拠したシステムが世界各地で動いている。日本のヴァーチャル天文台システムは、国立天文台が中心となって構築し、<http://jvo.nao.ac.jp/> からアクセスできる。

巨大ストレージとソフトウェア望遠鏡

日米欧が参加してチリに共同で建設を進めている ALMA 望遠鏡が、その初期科学観測を 9 月 30 日より開始した。ALMA 望遠鏡は、年間数 100TB の観測データを取得し、チリにあるオンサイトパイプラインシステムにより一次処理され、マスターアーカイブを経て、日米欧 3 カ所に設置される地域センターに配布される。ALMA 東アジア地域センターは国立天文台三鷹に設置され、オフサイトパイプライン処理と（永続的）アーカイブシステムについての検討が進んでいる。ペタバイトスケールのストレージシステムが必要になるが、ALMA アーカイブは、ALMA に関するあらゆるデータを置く場所であるため、地域センター機能の中核になるものである。

世界では、さらに巨大な計算機資源を必要とする望遠鏡計画が進行している。欧州にある LOFAR や、いずれ南半球に設置されると期待される SKA である。SKA には豪州と南アフリカがサイト候補として残り、いずれの国も、SKA の 1/100 スケールのパイロット望遠鏡を建設している。豪州のパイロットシステムである ASKAP はしかし、望遠鏡からのデータ処理のために、約 2PFlops の演算能力と Tbit クラスのデータ伝送能力を必要としている。これはもはや、ソフトウェア望遠鏡と呼んでもよいシステムである。そのストレージについても、高速アクセスのための SSD、データ格納のための HDD、バックアップのための磁気テープを組み合わせた分散ストレージが検討されている。実際の SKA では、この 100 倍の能力が必要であり、技術的なブレークスルーが必須である。

第 4 のパラダイムでは、いわばスパコンの演算能力に加え十分なメモリバンド幅や巨大ストレージを組み合わせたモンスターマシンが必要となるのである。

以上