

ALMA の大規模データ処理

小杉 城治^{*1}

Extensive data handling for ALMA (Atacama Large Millimeter/ Submillimeter Array)

George KOSUGI^{*1}

Abstract

The Early science operation with 16 antennas will soon be started in the ALMA (Atacama Large Millimeter / Submillimeter Array) Project. In the full science operation phase, 66 antennas produce enormous volume of data, which is correlated with each other in real time. Then the data is accumulated in time and binned in frequency to reduce the size, so that the data is permanently stored in the archive at a rate of 200TB per year. The data is first archived at the operation site in Chile and then mirrored to three ARCs (ALMA Regional Center) in Japan, North America, and Europe through the network. Data driven pipeline analysis runs in Chile to produce reduced data into the archive, while the user driven pipelines run in each ARC.

Keywords: ALMA, archive, data analysis, pipeline analysis

概要

巨大国際プロジェクト ALMA (Atacama Large Millimeter / Submillimeter Array) では、まもなく 16 台の望遠鏡を使った初期科学運用観測が開始される。本格運用期には 66 台の望遠鏡から出力される大量データをリアルタイムに相関処理、並びに、時間方向・周波数方向に積分・圧縮処理をおこない、恒久的にアーカイブ可能なデータ量である年間 200TB (テラバイト) までデータレートが落とされる。一旦チリの ALMA 観測所に保管されたアーカイブデータは、そこから日本、アメリカ、及び、ヨーロッパそれぞれの ALMA 地域センター (ARC: ALMA Regional Center) にネットワークを使って転送される。チリではデータ起動型のパイプライン解析処理が、また、各 ARC ではユーザー起動型のパイプラインが処理済みデータを生成する。

1. はじめに

日本、アメリカ (NRAO: National Radio Astronomy Observatory), ヨーロッパ (ESO: European Southern Observatory) 共同で 10 年以上かけて進められてきた巨大国際プロジェクト ALMA (Atacama Large Millimeter / Submillimeter Array) の建設は、最終段階にさしかかっている。ALMA は、南米チリ北部のアタカマ砂漠の高原 (標高 5000m) に 66 台の電波望遠鏡を組み合わせて巨大な電波干渉計 (口径 12m の望遠鏡 50 台を組み合わせるメインアレイと、口径 12m 望遠鏡 4 台と口径 7m 望遠鏡 12 台を組み合わせて短い干渉計基線を作るコンパクトアレイ:「いざよい」) を構成する。観測波長域は 0.3 - 3.6mm (80 - 950 GHz), 視野 7 arcsec (@900 GHz) であり、望遠鏡配置によって変化する角度分解能は 0.5 - 0.005 arcsec (@900 GHz) となっている。

ALMA の運用母体である合同 ALMA 観測所 (JAO: Joint ALMA Observatory) は、山頂施設 (AOS: Array Operations Site: 標高 5000m)、山麓施設 (OSF: Operations Support Facility: 標高 2900m)、および、ALMA サンチャゴ中央事務所 (SCO:

^{*1} 自然科学研究機構国立天文台 ALMA 推進室 (National Astronomical Observatory of Japan, National Institutes of Natural Sciences)

Santiago Central Office) から構成される。2011年3月には一部の望遠鏡を使った初期科学運用観測に対する一般研究者へのプロポーザル公募がおこなわれた。2011年9月からは最低16台の望遠鏡を使った初期科学運用観測が開始される。AOSには、2011年8月現在19台の口径12m望遠鏡が稼働している。初期科学運用観測で使われる望遠鏡数は本格運用期の約4分の1の規模であるが、それでも既に現在稼働中の電波望遠鏡を十分凌ぐ感度が達成される。本格運用期には66台の望遠鏡それぞれから大量の観測データがはき出される。

観測データはAOSにある相関器でリアルタイムに相関処理され、OSFに設置されているフロントエンドアーカイブに保管される。OSFで一時保管されたデータは、SCOのアーカイブシステムに順次転送され、そこから更に日・米・欧のALMA地域センター(ARC: ALMA Regional Center)のアーカイブシステムにミラーリングされる。

SCO、及び、各ARCには、ALMAの巨大データを処理するためのパイプラインクラスター計算機が用意される。SCOでは、観測データの到着によってパイプライン処理が駆動され、定型処理されたデータもまたアーカイブされる。一方、各ARCでは、ユーザー、或いは、ARCスタッフが処理パラメータを任意に指定してパイプライン処理を駆動するが、処理データはアーカイブに保存されない。これは、SCOパイプラインが一般的な処理に基づいた均質なデータを生み出しアーカイブとして一般ユーザーに提供することを主眼としているのに対して、ARCパイプラインは各観測者のサイエンスに特化した処理を担っているからである。とはいえ、SCOパイプラインと全く同じ処理をARCパイプラインで実行することも可能である。SCOでパイプライン処理されたデータもまた各ARCにミラーリングされるため、ユーザーは生データ、パイプライン処理済みデータ共にARCのアーカイブシステムから利用することができる。

2. ALMA システム

2.1 オンライン処理系

望遠鏡パラボラで集められた天体からの電波は、受信機システムで電気信号変換・周波数変換・増幅されたのち、3bitのデジタル信号に変換される。各望遠鏡から出力されるこれらのデジタル信号は、約100Gbpsで相関器に送られる。このとき、望遠鏡の指向方向やケーブル長の違いによる遅延量や、光路中の水蒸気量による遅延を補正した上で、望遠鏡の組み合わせ(ペア)毎に信号がリアルタイムに相関処理される。そのため、相関器には高速な専用計算機が使われている。ALMAの相関器には、メインアレイ用の64アンテナ入力まで処理できるものと、コンパクトアレイ用¹⁾の16アンテナ入力を処理できるものの2種類ある。これらの相関器をあるときは独立に、また、あるときは同期させながら運用することによって、ALMAに要求されている様々な観測モードが実現される。

相関器から出力されたデータは、その後、ノイズを減らし感度を上げるため、また、出力データレートを下げるために、データを時間方向に積分したり、周波数方向に圧縮(ビニング)したりする。適当なサイズになったデータは、OSFにあるフロントエンドアーカイブシステムに転送・一時保管される。

2.2 アーカイブシステム

ALMAでは、論理的に4つのアーカイブシステムが稼働している。それらは、フロントエンドアーカイブ²⁾、バックエンドアーカイブ、Primaryサイエンスアーカイブ³⁾、及び、Secondaryサイエンスアーカイブである(図1)。フロントエンドアーカイブとバックエンドアーカイブはどちらもOSFで稼働しているが、フロントエンドアーカイブはオンライン処理系から出る間欠的なデータ(最大64MByte/s, 平均6.4MByte/s)を滞りなく受信して安全に一時保管する役割を担っている。バックエンドアーカイブは一時保管されたデータを、SCOのPrimaryサイエンスアーカイブに定常データレート(6.4MByte/s)以下で逐次転送する。OSF-SCO間のネットワークトラブルが運用に影響を与えない(ネットワークが切れても運用を継続できる)ように、OSFのアーカイブには観測データを数ヶ月間にわたって保管できるだけの容量が確保されている。SCO Primaryサイエンスアーカイブからは日・米・欧それぞれのARC Secondaryサイエンスアーカイブへデータがミラーリングされる。サイエンスアーカイブには年間200TBのデータが保管される予定である。

アーカイブデータの保全是観測所にとって最も重要な要件の1つであるため、定期的にバックアップを作成する等の措置がとられることが多い。ALMAの場合、同じデータが複数箇所(日・米・欧のARC)に保管されているため、バックアップは必要ない。将来どこかのARCでデータが破損した場合には、SCOからARCにデータをミラーリングしたのと同様の方法で、他のARCからミラーリングがおこなわれる。

2.3 オフライン処理系

2013年以降のALMA定常運用時には、観測データは全てSCOの解析パイプラインクラスターで自動的に処理され、処

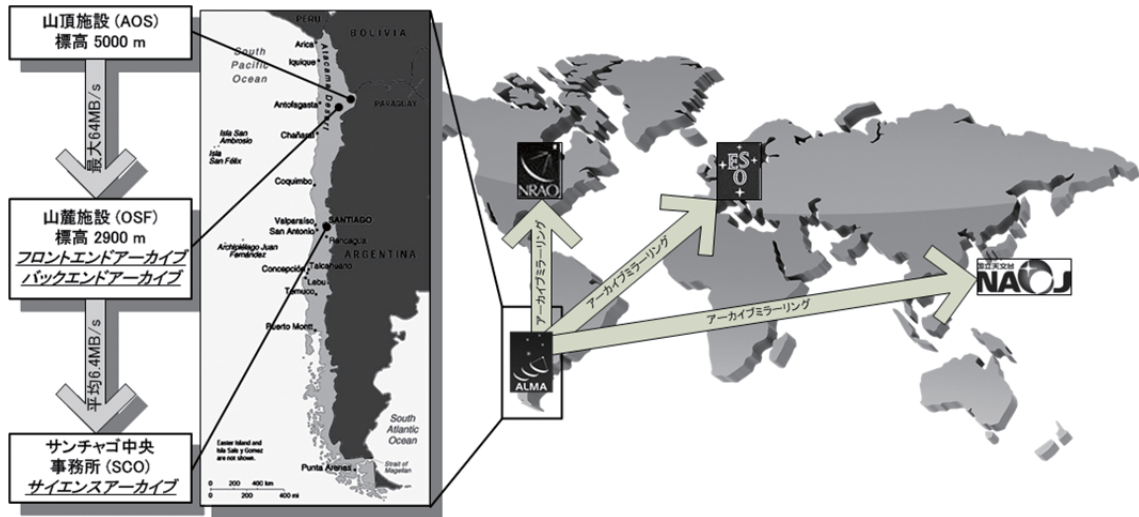


図1 アーカイブデータの流れ

処理済みデータもそこでアーカイブされる。初期科学運用期には、まだ最適なパイプライン解析処理アルゴリズムやパラメータが決められないため、JAO、或いは、各ARCのALMAサイエンススタッフがPrimary、或いは、Secondaryサイエンスアーカイブから取り出した生データを手動で解析処理し、その処理済みデータが観測者の元に送られる。また、処理済みデータはSCO Primaryアーカイブにも登録され、ARCにミラーリングされる。ALMAの定常運用時の典型的なデータサイズは、1データセットあたり25-250GB程度と考えられている。これは平均的なデータレートで観測した場合、観測時間で1時間から10時間に相当する。SCOのパイプラインに求められる性能は、観測データをかかった観測時間以下で処理することである。そのため、計算機ハードの高性能化と処理ソフトの効率化が合わせて求められる。

2.3.1 ALMAデータ解析パッケージ「CASA」の並列化対応

ALMAでは、解析ソフトウェアパッケージCASA(Common Astronomy Software Applications)⁴⁾を開発している。前述の解析パイプラインもCASAを解析エンジンとして利用している。ここでは、クラスター計算機での並列処理を視野に入れたCASAにおける取り組みを紹介する。

図2にCASAのソフトウェア構造を示す⁵⁾。CASAの現段階での並列化は、高レベルのpythonレイヤーに実装されている。すなわち、ひとまとまりの処理(Task)や個々の処理(Tool)に対して分割したデータを与え、それらの処理を並列に実行する。そのためにはデータファイルに複数のプロセスから同時にアクセスさせる必要がある。ただし、ファイル内の個々のデータ部分についてはそれぞれ別のプロセスに渡されるため同時アクセスは生じない。このようなデータアクセスの仕組みが並列化のためにCASAには追加されている。参考として、図2に処理を複数ノードに分散処理させたときの実行時間の推移を載せる。ノードの数にほぼ反比例して実行時間が短縮されているのがわかる。現在、更なる並列化の試行、及び、アルゴリズムやその実装の最適化が精力的に進められている。

2.3.2 ALMAのパイプライン処理

ALMAパイプライン⁶⁾⁷⁾の解析処理エンジンには上述のCASAが使われている。ALMAパイプラインはCASAの中に実装されたTask(図1)と似たような位置づけになっていて、パイプラインの個々のステージ(1つ1つのパイプライン処理をステージと呼び、多数のステージを経て全処理が完了する)はCASAの中からも利用可能になっている。

大規模データをパイプライン処理する場合、各ステージの処理結果が妥当かどうかをユーザーに「見せる」ことが、パイプライン処理済みデータを配布して2次利用させる上で非常に重要となる。ALMAパイプラインでは、

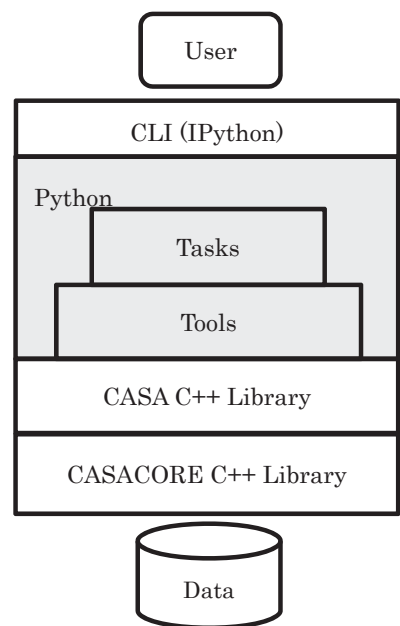


図2 CASAソフトウェア構造

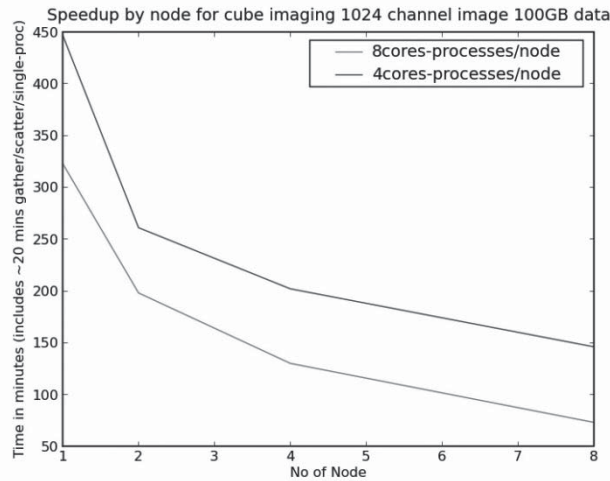


図 3 並列処理ノード数と実行

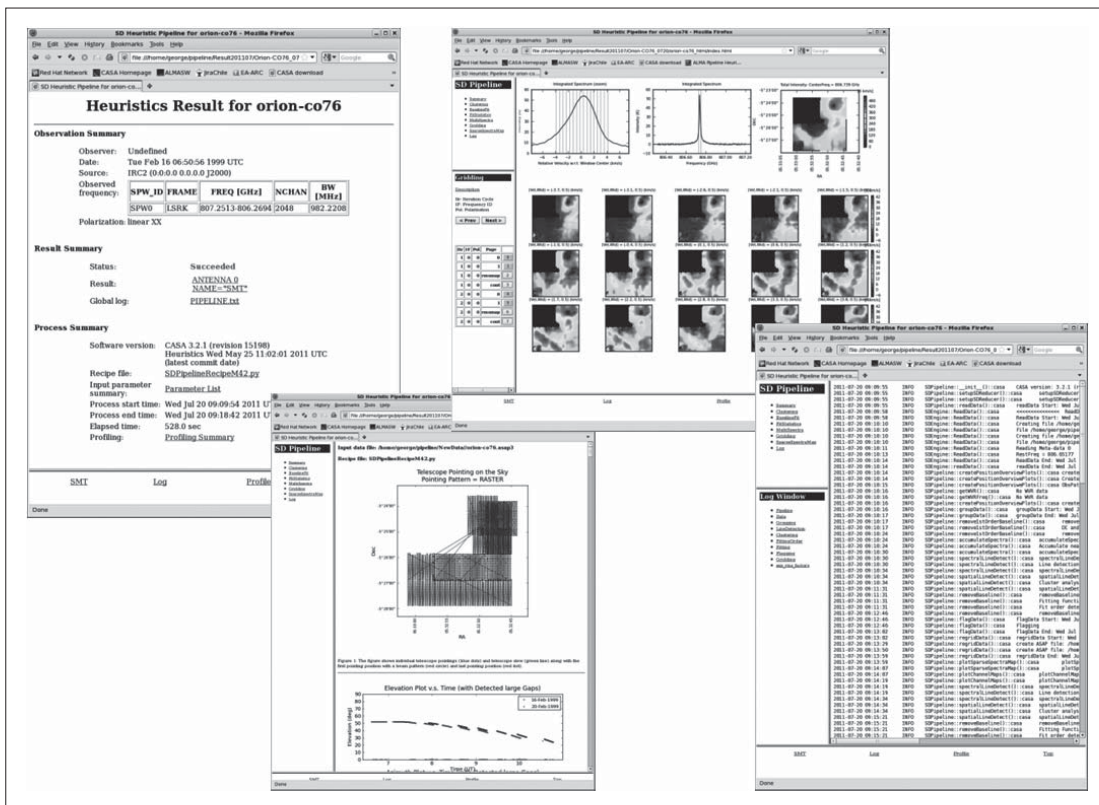


図 4 パイプライン結果出力例

中間結果を含む全処理の情報がネットワーク経由で Web ブラウザーから確認できる (図 4)。この情報は、パイプライン処理済みデータと同梱されてユーザーに配布されるため、ユーザーは各処理の内容やパラメータの妥当性を自分で確認することができる。

2.3.3 ALMA のパイプラインクラスター計算機

パイプラインクラスター計算機の要求仕様は、定常運用期のデータレート (年間 200TB) を遅滞なく処理することである。ただし、現段階においては、まだパイプライン処理が完全には確立されていないため、計算量を正確に見積もることはできない。図 5 が現状想定されている規模のパイプラインクラスター計算機である。パイプラインは各ステージをできるだけ独立に動かせるようにするため、ステージ毎にファイル入出力が生じる。ディスク I/O が処理のボトルネックにならないよう、

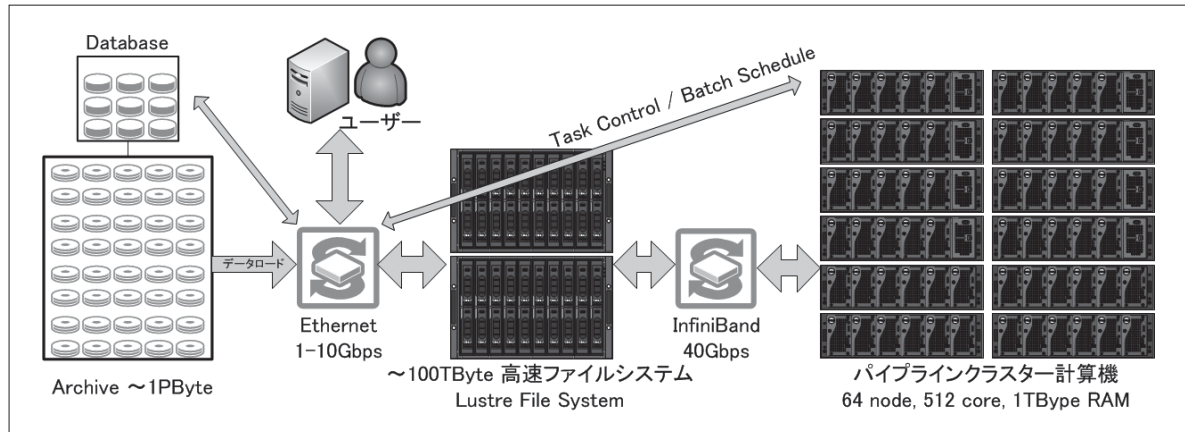


図5 パイプラインクラスター計算機

InfiniBandTM*2 で高速ディスクを接続した LustreTM File System*3 を使用する。また、必要十分な計算機リソースを見積もるために、初期科学運用期には定常運用期のパイプラインクラスターを 10 分の 1 程度にスケールダウンしたものを導入する。

3. まとめ

初期科学運用を間近に控え、全システムをつないだ運用試験が急ピッチで進められている。望遠鏡 16 台以上を使った観測データも取得されつつあり、実際のデータを使用して解析パッケージやパイプラインを本格的に試験する環境がようやく整ってきた。理想の解析を求めると、アルゴリズムも複雑になり、その分、計算量も増加する。計算量が増加すると、計算機ハードウェアへの要求が高まる。決められた予算の中で、解析の質を最大にするよう、ハードウェア、ソフトウェア、アルゴリズムに投入すべきリソースのバランスを最適化していくことが、今まさに求められている。

参考文献

- 1) Okumura, K. S., et al., Atacama Compact Array Correlator for Atacama Large Millimeter / submillimeter Array, Proc of Asia-Pacific Radio Science Conference, 2010
- 2) Wicenc, A., et al., ALMA Front-end Archive Setup and Performance, Proc of Astronomical Data Analysis Software and Systems XIX, ASP Conference Series, Vol. 434, 2010, P.457
- 3) Etoka, S., et al., A Look at the ALMA Science Archive, Proc of The Eighth Pacific Rim Conference on Stellar Astrophysics, ASP Conference Series, Vol. 404, 2009, P.324
- 4) <http://casa.nrao.edu/>
- 5) 中里 等, ALMA データ解析ソフトウェア CASA の開発 4, 日本天文学会秋期年会, 2010/9
- 6) Lightfoot, J., et al., ALMA Pipeline Heuristics, Proc of Astronomical Data Analysis Software and Systems, ASP Conference Series, Vol. 394, 2008, P.573
- 7) Davis, L. E., An Overview of the ALMA Pipeline System, Proc of Astronomical Data Analysis Software and Systems, ASP Conference Series, Vol. 411, 2009, P.306

脚注

- *2 InfiniBandTM: InfiniBand とは、IBTA (InfiniBand Trade Association) によって定義された仕様で、サーバーやストレージを相互接続するためのオープンスタンダードなポイント・ツー・ポイント双方向シリアル接続である。<http://www.infinibandta.org/>
- *3 LustreTM File System: 大規模ストレージのための分散並列ファイルシステムであり、GNU GPL のもとでオープンソース・ソフトウェアとして開発、維持されている。<http://wiki.lustre.org/>