

# 国立天文台・天文データセンター 大規模観測データ解析システム

○磯貝瑞希、古澤久徳、巻内慎一郎、田中伸広、高田唯史（国立天文台）

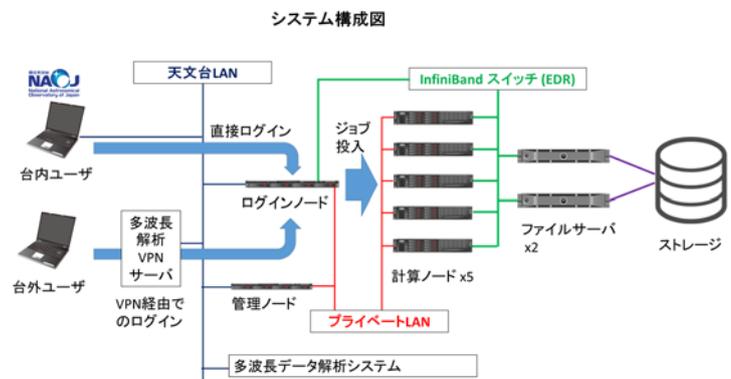
## 概要(Abstract)

国立天文台・天文データセンターでは、ハワイ観測所すばる望遠鏡の超広視野カメラ HSC など、解析処理に多くの計算資源を必要とする大規模観測データ用の解析システム（以下大規模解析）を構築中である。本システムは IBM 社の Spectrum Scale (旧 GPFS) を採用し、大容量かつ高速 I/O を実現している。計算ノードの総コア数は現在 280 で、少なくとも 2000 コアまで増強予定である。本講演ではこの詳細と主に初期試験結果を報告する。

## 1. システム情報・構成・運用方針

本システムは、ログインノード 1 台、計算ノード 5 台、ファイルサーバ 2 台、ストレージ、管理ノード 1 台で構成されている。OS は管理ノードのみ CentOS 7 で、それ以外は Red Hat Enterprise Linux 7 である。計算ノードは 1 台あたり CPU (Intel Xeon gold 6132, 2.6GHz, 14core) を 4 個(合計 56 core/台)、メモリを 1TB 搭載しており、全 5 台の総コア数は 280 である。ストレージは容量 5PB でファイルシステムは IBM 社の Spectrum Scale である。下にシステム構成図を示す。

本システムは多波長データ解析システム（以下多波長解析）と連携して運用する予定であり、本システムの利用には、多波長解析のアカウントを必要とする。本システムの計算資源はバッチシステムで管理され、ユーザはログインノードからジョブを投入することで計算ノードを使用する。本システムは 2019 年 2 月より試験運用を開始し、2019 年 8 月より本運用を開始予定である。



## 2. 初期試験

性能評価およびシステム構築後の動作確認を目的に、先行して納入された計算ノード 2 台(総コア数:112)とファイルサーバ、ストレージの構成で 3 種類の試験を実施し、多波長解析(開発系)の結果と比較した。多波長解析(開発系)は CPU (Intel Xeon E5-2667v4, 3.2GHz 8core)を 2 個、メモリを 256GB 搭載したサーバ 3 台から構成されており、以下の(1)と(2)の試験ではその 3 台を、(3)の試験では運用系のバッチサーバの台数と揃えるために 2 台を使用した。試験に用いたストレージは 1 台のサーバと 16Gbps のファイバーチャネル 2 本で接続されており、容量 12TB の領域である。本領域は他のサーバに対しては 10Gbps の LAN を通して NFS ver.3 プロトコルで共有されている。

### (1): ファイルの Write/Read 速度

ベンチマークソフト IOR を使用してファイルシステムの Write/Read 速度を測定した。右の表がその結果で、単位は MiB/s、平均値と標準偏差(括弧内に)を掲載している。I/O の並列数は 112(大規模)、96 (多波長)で、

操作	大規模解析	多波長解析
Write	23,774 (449)	274 (16)
Read	22,449 (845)	1,542 (55)

1 ファイルのサイズは 100MiB, 1 I/O Call 当たりのデータ転送サイズは 1MiB で、試行回数は 10 回である。ファイルシステムの W/R 速度は 22-23GiB/s で、多波長解析(開発系)の約 15-83 倍であった。

### (2): ファイルの操作速度

ベンチマークソフト mdtest を使用してファイルシステムのファイル操作(作成、状態確認、読み込み、削除)速度を測定した。右の表がその結果で、単位は操作数/s、平均値と標準偏差(括弧内に)を掲載している。操作の並列プロセス数は試験(1)と同じで、1 プロセス当たりのファイル作成数は 1000、試行回数は 3 回である。ファイルの操作速度は操作内容によって大きく異なるが、全ての操作で多波長解析(開発系)の 20 倍を超えている。試験(1)と(2)の結果より、本システムのストレージは非常に高速な I/O 性能を持つことが示された。

操作	大規模解析	多波長解析
File creation	44,343 (450)	2,106 (60)
File status	1,033,556 (178,465)	8,723 (2,499)
File read	208,597 (6,421)	4,160 (63)
File removal	86,015 (2,131)	3,874 (32)

### (3): HSC データの解析時間

すばる望遠鏡の超広視野カメラ HSC の観測生データからマルチバンド解析による検出天体カタログ作成までの一連の解析処理を行い、その処理時間を測定した。右の表がその結果である。解析は HSC データの解析処理ソフト HscPipe の最新版(ver 6.7)を使用し、データは HSC 解析講習会のウェブページで公開されているもの(6CCD 版と全(112)CCD 版の 2 種類、どちらも 3 バンド)を用いた。本データの Bias, Dark は 1CCD 当たり 5 枚、Flat, Object は 1 バンド・1CCD 当たり 5 枚(= 3 バンドで 15 枚/CCD)である。

処理内容	投入ジョブ数	6CCDの処理					全CCDの処理			
		入出力ファイル数 (容量 [GiB])		大規模		多波長	入出力ファイル数 (容量 [GiB])		大規模	多波長
		入力	出力	q112 x1	q32 x1	q16 x2	入力	出力	q112 x1	q32 x1
Bias作成	1	30 (0.5)	6 (0.3)	42	44	44	560 (9.5)	112 (6.3)	82	236
Dark作成	1	36 (0.8)	6 (0.3)	42	40	36	672 (16)	112 (6.3)	88	250
Flat作成	3	102 (2.2)	18 (0.4)	125	113	73	1904 (41)	336 (7.0)	333	719
Fringe作成	1	48 (1.3)	6 (0.3)	43	45	44	896 (24)	112 (6.2)	108	288
Sky作成	3	120 (2.6)	18 (<0.1)	109	126	82	2240 (48)	336 (<0.1)	417	1117
一次処理	1	144 (2.9)	960 (17)	216	398	725	2688 (54)	16960 (306)	2143	5593
Sky補正	1	108 (4.7)	90 (<0.1)	157	161	118	1896 (82)	1560 (0.3)	235	417
Stack	3	360 (4.7)	181 (6.8)	1717	1384	898	6240 (83)	1792 (103)	3572	4453
マルチバンド解析	1	142 (2.4)	128 (2.9)	4596	3927	4248	2026 (33)	1088 (38)	14404	22856
合計				7047	6238	6268			21382	35929

表中の q16, q32, q112 とはバッチシステムのキュー(queue)のことで、それぞれ最大で 16, 32, 112 CPU コアを使用可能である。なお、多波長解析で使用した q32 は開発系専用のキューで運用系には存在しない。6CCD の処理では、処理に必要な並列数が少ないため、総コア数・I/O 性能よりも CPU クロック数が大きい(2.6GHz に対して 3.2GHz)多波長解析(開発系)の方が 1 割程度早く処理を終えているが、112CCD の処理では、総コア数・I/O 性能の大きい大規模解析の方が 4 割ほど早く終了している。