

## 1-2 ストレージ

### 1-2 Storages in Satellite Data Processing System

今村 哲生\*<sup>1</sup>      高田 左知男\*<sup>2</sup>  
Tetsuo Imamura      Takata Sachio

#### Abstract

The Japan Meteorological Agency (JMA) began operating the Numerical Analysis and Prediction System 9 (NAPS9)—its new supercomputer system—in June 2012. Accordingly, the Meteorological Satellite Center (MSC) improved the Satellite Data Processing System in response to an increase in data to be processed at the center.

This paper describes the structure of the storages in the new satellite data processing system.

#### 要 旨

気象庁(JMA)は2012年6月より新規スパコンシステムNAPS9を稼働開始した。これに合わせ気象衛星センター(MSC)では、センター内で行うデータ処理の増加に合わせ、衛星データ処理システムの改良を行った。

この章では新規衛星データ処理システムにおけるストレージの構成について記述する。

#### 1. はじめに

衛星データ処理システムには、スーパーコンピュータを含めた全体システム内でデータを共有するために、ストレージを外部記憶装置としてサーバに接続させている。業務用途、ファイルサイズ、アクセス頻度等に応じ、「衛星データ処理ストレージ」「データバンクストレージ」「バックアップストレージ」「共用ストレージ」の4種類で構成されている(以下「本ストレージ群」)。

本ストレージ群のうち、バックアップストレージ以外は、NAS(Network Attached Storage)ヘッドと磁気ディスクアレイ装置 Hitachi Adaptable Modular Storage2500 (AMS2500)で構成されており、NFSでサーバ等に接続してデータを共有している。これらの装置は、スーパーコンピュータシステム内のシステム管理サーバで一元的に管理され、「POSIX 準拠」「複数のパーティションを設定」「ホットスワップ」「Quota 設定」など共通の機能を有している。

衛星データ処理ストレージは、衛星システム専用の装置であるため、基本的に衛星処理サーバ類のみ接続されている。それ以外のストレージ群は、入出力データの送受信専用のストレージネットワーク

を構成し各サーバに接続している。

なお、各ストレージは相互に独立で、あるストレージの障害の影響が他のストレージには及ばない。いずれのストレージも、運用を停止することなく保守ができ、かつ、他のストレージやそのストレージを接続しているサーバの保守に影響を受けることなく運用できる。

#### 2. ストレージ構成

各ストレージの記憶部構成や性能・利用可能容量を表1に示す。

##### 2.1 衛星データ処理ストレージ

衛星データ処理ストレージは、衛星データ処理を行うサーバの2次記憶装置となる。SATA インタフェースの磁気ディスクによる磁気ディスクアレイ装置 AMS2500 を使用し、RAID5 構成でデータ保存を行う。

衛星データ処理ストレージは、NAS として NFS により衛星一次処理サーバ、衛星二次処理サーバ及び外国衛星処理サーバとファイル共有を実現する。

\*1 気象衛星センターデータ処理部システム管理課 (現 データ処理課)

\*2 気象衛星センターデータ処理部システム管理課

衛星データ処理ストレージの各領域を衛星一次処理サーバ、衛星二次処理サーバ及び外国衛星処理サーバで NFS マウントすることにより、同一のファイルにアクセス可能とする。

同様に、衛星一次処理サーバ、衛星二次処理サーバ及び外国衛星処理サーバ上で実行されるユーザプログラムからファイル単位で直接入出力を可能とする。

また、ストレージ本体は主系・副系で冗長性を持たせた構造を持つ。副系の AMS はユーザーからは見えないように隠蔽されているが、ユーザーにより書き込まれたデータは、主系・副系両方にほぼ同時に記録される。

## 2.2 共用ストレージ

共用ストレージは主に一般ユーザ向けの開発領域であり、一般ユーザのホームディレクトリや開発ディレクトリが置かれ、RAID6 構成でデータ保存を行っている。

比較的サイズの小さなファイルを多数格納し、アクセス頻度が高い傾向があるため、それに対応したスペックとなっている。

共用ストレージの各領域をスーパーコンピュータ主系・副系 I/O ノード、衛星二次処理サーバ、外国衛星処理サーバ、業務処理サーバ A・B 及びデコードサーバで NFS マウントすることにより、複数のユーザが同じファイルにアクセスすることが可能である。同様に前述のサーバ上で実行されるユーザプログラムからファイル単位での直接入出力を可能とする。

## 2.3 データバンクストレージ

データバンクストレージは、衛星データ処理システムとして重要な衛星観測データや衛星プロダクトを保存する目的のストレージであり、他と比較して若干アクセス頻度は低いものの、比較的サイズの大きなファイルを格納する。スパコンの高速ストレージなど、他のストレージで処理した結果を後日利用のために、保存・蓄積するのもこことなる。原則このストレージを最終保存領域とする。

いずれのデータバンクストレージも RAID6 構成でデータ保存を行う。

データバンクストレージは、その用途に合わせて(1)(2)(3)の3式存在するが、衛星データ処理システムで使用するのは(1)(3)の2式である。

(1)は、気象衛星センター・本庁予報部が、共同で様々な用途に使用されている。なお、このストレージのみ NAPS9 システム稼働前に先行納入されており、NAPS8 システムのテープライブラリのバックアップとして運用データの保存に使用されていた。気象衛星センター第一庁舎に設置され、NAS ユニット2台1組、ストレージ基本筐体2台1組で一式を構成する。

(3)は、主に気象衛星センターのデータバンクとして使用されている。気象衛星センター第三庁舎に設置され、NAS ユニット2台1組、ストレージ基本筐体2台1組で一式を構成する。

なお、(2)は気象衛星センター第三庁舎に設置されており、主に本庁予報部の数値予報ルーチンに利用されている。

## 2.4 バックアップストレージ

バックアップストレージは、データバンクストレージの一部のデータを障害などでのデータ亡失に備えてバックアップする。

装置構成は、日立製作所製の磁気テープライブラリ装置 L56/3000 と管理サーバ(EP8000/520)2台1組となっている。磁気テープ媒体は、非圧縮で 800GB を記録可能な LTO Ultrium4 規格の磁気テープ 1900 巻を格納することができ、データ領域の実効容量は約 1500TB となっている。

管理サーバは EP8000/520 が使用され、バックアップデータ管理ソフトウェアとして JP1/VERITAS NetBackup を搭載している。

JP1/VERITAS NetBackup により、データバンクストレージ内の指定ディレクトリ以下のファイルが後述するストレージネットワークを経由して、L56/3000 に自動的に複製される。バックアップは、1日1回決まった時間に差分で行っており、複製元のファイルシステムを再構築した場合などに、バックアップされたファイルによる復元(リストア)が可能である。また、リストアは、バックアップされているデータからディレクトリ毎やファイル単位で行うことも可能である。

対象ストレージ	主要ハードウェア	主要記録媒体	記憶部の構成	利用可能容量 (うち衛星システム 利用可能容量)
衛星データ処理 ストレージ	NAS ユニット×4 台 AMS2500×2 台		RAID5 15グループ	69TB (mssc_rtn など : 衛星システム専用)
共用ストレージ	NAS ユニット×4 台 AMS2500×4 台	HDD:575.3GB (SAS, 15,000rpm)	RAID6 82グループ	754TB (mssc:56TB)
データバンク ストレージ(1)	NAS ユニット×2 台 AMS2500×2 台	HDD:1968.52GB (SATA, 7,200rpm)	RAID6 41グループ	807TB (mssc:75TB)
データバンク ストレージ(2)	NAS ユニット×2 台 AMS2500×1 台	HDD:1968.52GB (SATA, 7,200rpm)	RAID6 31グループ	610TB (衛星システムには不使用)
データバンク ストレージ(3)	NAS ユニット×2 台 AMS2500×2 台	HDD:1968.52GB (SATA, 7,200rpm)	RAID6 77グループ	1515TB (mssc1:747TB) (mssc2:491TB)
バックアップ ストレージ	EP8000/520×2 台 AMS2100×1 台 L56/3000 ライブラリ	テープメディア : 1本あたり 800GB	磁気テープ 1900本	1520TB

表1 各ストレージ側の性能比較

衛星システム利用可能容量にはNASファイルシステム名も追記。

全利用可能容量を衛星システムのみを利用できるとは限らない(各課ルーチンなどにも使われる)。

### 3 入出力性能

#### 3.1 サーバ側の入出力性能

試験段階でのサーバ側の入出力性能を表2に示す。

表2の計測は、各サーバ1台毎ないしはスーパーコンピュータの系ごとに独立して測定を行う(一式を構成する複数サーバは、同時に測定する)。

測定に使用するファイルは、1ファイルあたり10GBとする。

#### 3.2 ストレージ側の入出力性能

試験段階でのストレージ側の入出力性能を表3に示す。

表3において、共用ストレージおよびデータバンクストレージ3種は、4つ同時に測定した結果を元にして示している。

測定に使用するファイルは、1ファイルあたり10GBとする。また、入力と出力は、個別に測定を行う。

なお、ストレージネットワーク装置AX6608Sの最大スイッチング性能は384Gbpsであり、入出力性能を計測する上でその性能がボトルネックになることはなかった。

#### 3.3 実務的な性能チェック

衛星データ処理ストレージは、20MBの異なる160個のファイルを60秒以内に読み込み、60秒以内に書き込みが可能である。

データバンクストレージ(1)(2)(3)についても、任意の「100個」のファイル(各10MB)を、「100秒以内」に読み出せるかどうかの試験が行われており、可能であることが確認されている(ストレージ毎に独立に測定を行った結果)。

対象サーバ	入出力性能	接続デバイス
衛星二次処理サーバ 外国衛星処理サーバ 業務処理サーバA 業務処理サーバB	サーバ 1 台あたり 30MB/s 以上	1000BASE-T (実効帯域は 80MB/s 程度)
デコードサーバ	サーバ 1 台あたり 200MB/s 以上	10GBASE-SR (実効帯域は 400MB/s 程度)
スーパーコンピュータ 主系 I/O ノード 副系 I/O ノード	I/O ノード 1 つあたり 200MB/s 以上、 系あたり 400MB/s 以上	10GBASE-SR

表 2 サーバ側の入出力性能  
対象サーバ～ストレージネットワーク装置間の入出力性能を表示する。

対象ストレージ	入出力性能	接続デバイス	シーケンシャル性能値
共用ストレージ	合計 1000MB/s 以上	10GBASE-SR	1 クラスタあたり Write 性能：1080MB/s Read 性能：780MB/s
データバンク ストレージ(1)	合計 100MB/s 以上	1000BASE-T	NAS ユニット 1 台あたり Write 性能：215MB/s Read 性能：220MB/s
データバンク ストレージ(2)	合計 200MB/s 以上	1000BASE-T	NAS ユニット 1 台あたり Write 性能：215MB/s Read 性能：220MB/s
データバンク ストレージ(3)	合計 200MB/s 以上	1000BASE-T	NAS ユニット 1 台あたり Write 性能：215MB/s Read 性能：220MB/s
バックアップ ストレージ	テープドライブ 1 台あたり 120MB/s (合計で 600MB/s)	—————	—————

表 3 ストレージ側の入出力性能  
対象ストレージ～ストレージネットワーク装置間の入出力性能を表示する。  
バックアップストレージは、非圧縮データを取り扱った場合の転送速度である。