

ARCHIVING AND MANAGING REMOTE SENSING DATA USING STATE OF THE ART STORAGE TECHNOLOGIES

B Lakshmi^a, C Chandrasekhara Reddy^{a,*}, SVSRK Kishore^a

^aNRSC, SDAPSA, Hyderabad, India – (lakshmi_b,sekharareddy_cc,kishore_svsr}@nrs.gov.in

Commission VIII, WG VIII/9

KEY WORDS: Virtualized Storage, Multi-tiering, Data Archival, Data Management

ABSTRACT:

Integrated Multi-mission Ground Segment for Earth Observation Satellites (IMGEOS) was established with an objective to eliminate human interaction to the maximum extent. All emergency data products will be delivered within an hour of acquisition through FTP delivery. All other standard data products will be delivered through FTP within a day. The IMGEOS activity was envisaged to reengineer the entire chain of operations at the ground segment facilities of NRSC at Shadnagar and Balanagar campuses to adopt an integrated multi-mission approach. To achieve this, the Information Technology Infrastructure was consolidated by implementing virtualized tiered storage and network computing infrastructure in a newly built Data Centre at Shadnagar Campus. One important activity that influences all other activities in the integrated multi-mission approach is the design of appropriate storage and network architecture for realizing all the envisaged operations in a highly streamlined, reliable and secure environment. Storage was consolidated based on the major factors like accessibility, long term data protection, availability, manageability and scalability.

The broad operational activities are reception of satellite data, quick look, generation of browse, production of standard and value-added data products, production chain management, data quality evaluation, quality control and product dissemination. For each of these activities, there are numerous other detailed sub-activities and pre-requisite tasks that need to be implemented to support the above operations.

The IMGEOS architecture has taken care of choosing the right technology for the given data sizes, their movement and long-term lossless retention policies. Operational costs of the solution are kept to the minimum possible. Scalability of the solution is also ensured. The main function of the storage is to receive and store the acquired satellite data, facilitate high speed availability of the data for further processing at Data Processing servers and help to generate data products at a rate of about 1000 products per day. It also archives all the acquired data on tape storage for long-term retention and utilization. Data sizes per satellite pass range from hundreds of megabytes to tens of gigabytes

The images acquired from remote sensing satellites are valuable assets of NRSC and are used as input for further generation of different types of user data products through multiple Data Processing systems. Hence, it is required to collect and store the data within a shared, high speed repository concurrently accessible by multiple systems. After the raw imagery is stored on a high-speed repository, the images must be processed in order for them to be useful for value-added processing or for imagery analysts. The raw image file has to be copied on to data processing servers for further processing. Given the large file sizes, it is impractical to transfer these files to processing servers via a local area network. Even at gigabit Ethernet rates (up to 60MB/s), a 5GB file will take at least 83 seconds. For this reason, it is useful to employ a shared file system which allows every processing system to directly access the same pool where raw images were stored. Concurrent access by multiple systems is ensured for processing and generation of data products. With the above reasons, it was chosen to have high speed disk arrays for acquisition and processing purposes and tape based storage systems for long-term huge data (Peta Bytes) archival in a virtualized multitier storage architecture.

This paper explains the architecture involved in a virtualized tiered storage environment being used for acquisition, processing and archiving the remote sensing data. It also explains the data management aspects involved in ensuring data availability and archiving Peta bytes sized, remote sensing data acquired over the past 40 years.

1. INTRODUCTION

Prior to 2011, NRSC was storing Satellite data in various types of physical media for archival purpose. This media was stored physically in racks at a distance from the work centers. When ever the data was required, tape would be moved to the respective work centre and copied on to the system. This process takes quite some time and this delay gets added to the product turn around time. To avoid such delays, it was planned to archive the entire data on online storage systems. The entire data acquired is archived onto three tier storage and used as and when required.

IMGEOS consolidates data acquisition, processing & dissemination systems and is built around scalable Storage Area Network based 3-tier storage system to meet current and future Earth Observation missions.

Work flow manager, user services, unified schedulers and processes of Data Ingest (DI) / Ancillary Data Processing (ADP) / Data Processing (DP) / Value Added Data Services (VADS) / Data Quality Evaluation (DQE) /Product Quality Check (PDC) are reengineered to utilize resources in a multi-mission mode.

Online availability of all the acquired data for all missions is ensured to minimize data-ingest times. All the movements of intermediate data and products are eliminated through usage of centralized SAN storage.

The storage is organized in three tiers. Tier-1 contains recent data and most frequently used data. Tier-2 will have data acquired in the last 15 months and off-the shelf products. Tier-3 will hold all the acquired data on tapes. All the three tiers are managed through HSM / ILM based on configurable set policies. These storages are presented in high-available and scalable configuration to all the storage clients through high speed FC network. Centralized monitoring and management of the SAN is done.

All the servers and workstations are connected on high-available and expandable Gigabit Ethernet comprising of core and edge switches. Centralized monitoring and management of the entire network is done.

A set of four dedicated servers are provisioned for Data Ingest (DI) to acquire data in real time from four antenna systems. One additional server is provisioned for redundancy. Three dedicated servers are provisioned for ADP to process ancillary data in near real-time. The redundant DI server is also a standby for ADP.

Data screening is done for all passes acquired and corrective feed back mechanism will be provided on data quality related issues.

Data Processing is classified into optical, microwave and non-imaging categories. A set of servers will be provisioned for each category and will be operated in multi mission mode. Servers are provisioned to cater peak load processing requirement of 1000 products per day. A set of work stations are provisioned for handling different functions like DQE, VADS, PQC and Media Generation.

2. STORAGE ARCHITECTURE

The architecture has taken care of choosing the right technology for the given data sizes, their movement and long-term lossless retention policies. Operational costs of the solution are kept to the minimum possible. Scalability of the solution is also ensured.

The main function of the storage is to receive and store the acquired satellite data, facilitate high speed availability of the data for further processing at DP servers and help generate data products at a rate of about 1000 products per day. It also archives all the acquired data on tape storage for long-term retention and utilization. Data sizes per satellite pass range from hundreds of megabytes to tens of gigabytes.

These raw images are most valuable assets of NRSC and are used as input for further generation of different types of user data products through multiple DP systems. Hence, it is required to collect and store the data within a shared, high speed repository concurrently accessible by multiple systems.

Earlier, the raw image file had to be copied on to data processing servers for further processing. After standard processing, the images were processed further to generate value-added products or for image analysis.

Given the large file sizes, it takes time to transfer these files between work centers via a local area network. Even at gigabit Ethernet rates (up to 60MB/s), a 5GB file will take at least 83 seconds. For this reason, it is useful to employ a shared file system which allows every processing system to directly access the same pool where raw images were stored. Concurrent access by multiple systems is ensured for processing and generation of data products.

Disk based systems provide high performance front-end for an archive. Disk based systems are not cost-effective to archive huge data for long term data archival and protection. Hence tape based solutions are planned for data protection for longer periods.

With the above reasons, it was chosen to have high speed disk arrays for acquisition and processing purposes and tape based storage systems for long-term huge data archival.

2.1 Tiered Storage Systems

Today's storage devices provide only point solutions to specific portions of large data repository issues. For example, high-end Fiber Channel based RAID solutions provide excellent performance profiles to meet demanding throughput needs of Satellite data acquisition and processing systems. Serial ATA (SATA) based RAID systems provide large capacity needs for longer-term storage. Tape storage provides enormous capacity without power consumption. Each of these storage technologies also has unique downsides to their utilization for certain data types. Utilization of Fiber Channel RAIDs for all data types is cost prohibitive. SATA-based RAID systems do not have the performance or reliability to stand up to high processing loads. Tape technology does not facilitate transparent, random access for data processing.

IMGEOS storage systems architecture leveraged the aspects of each device's capabilities into a seamless, scalable, managed data repository as required. The architecture is shown in figure 1.

2.2 Storage Virtualization using SAN File System

SAN File System (SFS) enables systems to share a high speed pool of images, media, content, analytical data, and other key digital assets so files can be processed and distributed quicker. Even in heterogeneous environments, all files are easily accessible to all hosts using SAN or LAN. When performance is key, SFS uses SAN connectivity to achieve throughput. For cost efficiency and fan-out, SFS provides LAN-based access using clustered gateway systems with resiliency and throughput that outperforms traditional network sharing methods.

For long term data storage, SFS expands pools into multi-tier archives, automatically moving data between different disk and tape resources to reduce costs and protect content. Data location is virtualized so that any file can easily be accessed for re-use, even if it resides on tape.

SFS streamlines processes and facilitates faster job completion by enabling multiple business applications to work from a single, consolidated data set. Using the SFS, applications running on different operating systems (Windows, Linux, Solaris) will simultaneously access and

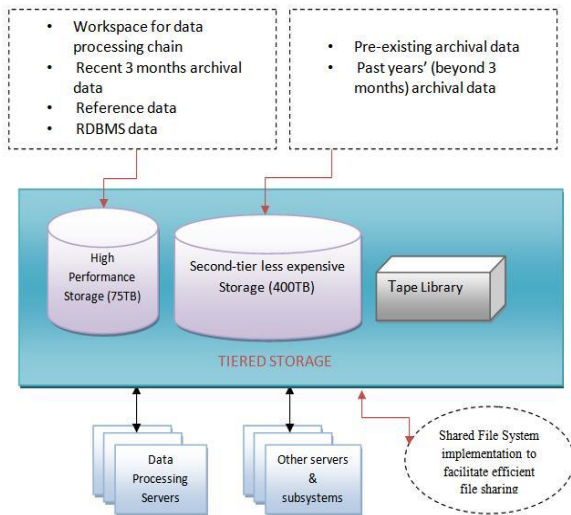


Figure 1. Tiered Storage

modify files on a common, high-speed SAN storage pool. This centralized storage solution eliminates slow LAN-based file transfers between workstations and dramatically reduces delays caused by single-server failures. The IMGEOs SAN has a high availability (HA) configuration, in which a redundant server is available to access files and pick up processing requirements of a failed system, and carry on processing.

2.3 Hierarchical Storage Management (HSM)

HSM is one of the features of SAN file System. HSM (Hierarchical Storage Management) is policy-based management of file archiving in a way that uses storage devices economically and without the user needing to be aware of when files are being retrieved from storage media. The hierarchy represents different types of storage media, such as redundant array of independent disks systems (FC, SATA), tape etc. Each of this type represents a different level of cost and speed of retrieval when access is needed. For example, as a file ages in an archive, it can be automatically moved to a slower but less expensive form of storage. Using an HSM it is possible to establish and state guidelines for how often different kinds of files are to be copied to a backup storage device. Once the guideline has been set up, the HSM software manages everything automatically.

The challenge for HSM is to manage heterogeneous storage systems, often from multiple hardware and software vendors. Making all the pieces work seamlessly requires an intimate understanding of how the hardware, software drivers, network interfaces, and operating system interact. In many ways, HSM is the analogue to heterogeneous computing, where different processing elements are integrated together to maximize compute throughput.

In the same way, HSM virtualizes non-uniform storage components so as to present a global storage pool to users and applications. In other words, the tiers must be abstracted; users and applications must be able to access their files transparently, without regard to their physical locations. This allows the customers and their software to remain independent of the underlying hardware. HSM has been a critical link in tiered storage systems, since these tools must

encompass the complexity of the hardware and all the software layers.

2.4 Storage Manager

Storage Manager (SM) enhances the solution by reducing the cost of long term data retention, without sacrificing accessibility. SM sits on top of SFS and utilizes intelligent data movers to transparently locate data on multiple tiers of storage. This enables us to store more files at a lower cost, without having to reconfigure applications to retrieve data from disparate locations. Instead, applications continue to access files normally and SM automatically handles data access – regardless of where the file resides. As data movement occurs, SM also performs a variety of data protection services to guarantee that data is safeguarded both on site and off site.

2.5 Meta Data Servers

SAN File System (SFS) is a heterogeneous shared file system. SFS enables multiple servers to access a common disk repository. It is ensured that the file system is presented coherently and I/O requests are properly processed (e.g. Windows kernel behaves differently from a Linux kernel). Because multiple servers are accessing a single file system some form of “traffic cop” is required to prevent two systems from writing to the same disk location and guarantee that a server reading a file is not accessing stale content because another server is updating the file. Like this there are many features in SFS as listed above. The SFS is installed on Meta data servers and SFS functions are performed using a Meta Data Servers.

The Meta Data servers configuration is 4-CPU, 32 GB memory etc. These servers are made high available by connecting them in cluster. This avoids single point of failure of SAN File System.

2.6 High Availability

The High Availability (HA) feature is a special configuration with improved availability and reliability. The configuration consists of two similar servers, shared disks and tape libraries. SFS is installed on both servers. One of the servers is dedicated as the primary server and the other the standby server. File System and Storage Manager run on the primary server. The standby server runs File System and special HA supporting software.

The failover mechanism allows the services to be automatically transferred from the current active primary server to the standby server in the event of the primary server failure. The roles of the servers are reversed after a failover event. Only one of the two servers is allowed to control and update SFS metadata and databases at any given time. The HA feature enforces this rule by monitoring for conditions that might allow conflicts of control that could lead to data corruption.

The HA is managed by performing various HA-related functions such as starting or stopping nodes on the HA cluster.

2.7 Storage Implementation

All the data acquired from the satellites in Data Ingest systems will be transferred to the Storage Area Network

(SAN). This is a three tiered storage. The location of given dataset is determined by its latest usage. This is based on frequency of usage of the data. The Frequently used data will be stored in FC storage; less frequently used data will be stored in SATA storage. Third level of storage is tape based storage. The tape storage is used for backing up the data of the disk based tiers. The storage is divided into nine logical storage partitions for functional use and management. All the servers access this storage for various computing requirements.

The acquired data is processed for auxiliary data extraction, processing, generating Ancillary Data Interface File (ADIF) and inputs to browse image and catalogue generation. Browse images, catalogue information, ADIF and Framed Raw Extended Data Format (FRED) data are the outputs of pre-processing system.

There are four antenna terminals and each terminal is associated with one data ingest server and one common additional server as standby for all the four. The five servers are connected to the SAN. The Ancillary Data Processing, Data Processing and other servers access the storage. The setup is shown in figure 2.

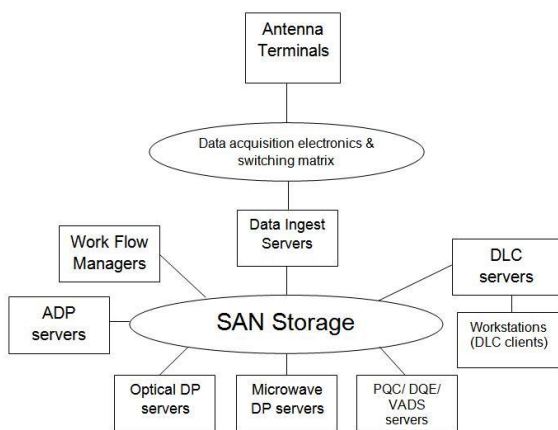


Figure 2. Storage Setup

It is found from the satellite data usage pattern, that recently acquired (<18 months) data is more frequently ordered than the earlier data. High performance FC storage of 75 TB can accommodate about three-months of data acquisition from all the satellites in addition to other work space required for processing servers. The tier-2 storage (SATA drive based) can accommodate all the acquired data for about 15 months. All the data is available online in the tape library.

The size of the tape storage is 6.0 PB, which will be available online for data archival and processing. The tape storage is logically divided into two equal parts. Each part holds one copy of data to be archived. Additionally, one more vaulted copy will be maintained. With this, data archived will have three copies (two online copies from tape library and one vaulted copy) having identical data for redundancy / backup. A fourth copy is generated which is kept at the Disaster Recovery (DR) site.

Overtime, existing archived data (~ 950 TB) on DLTs and SDLTs, will be ported onto the tape library, so that there will be only one archive for all the data. As detailed before, all

the three tiers will be accessed through SAN file system installed on meta-data servers.

2.8 SAN Components

The SAN components are shown in the figure 3. The components of the SAN storage are three tiers of storage (FC, SATA and Tape), SAN switches and Metadata servers. SAN file system is installed on Metadata servers. Metadata servers are connected to the SAN switches using host bus adapters. Similarly all other servers, which are direct SAN clients, are also connected to the SAN switches. Storage tiers, SAN switches and servers are connected at 4Gbps bandwidth. SAN switches are configured in high available mode.

2.9 Storage Systems Management

One of the useful benefits of consolidating storage on to a SAN is the flexibility and power provided by the Storage Resource Management (SRM) software. Using the SRM software, management and administration functions like storage partitioning, RAID arrays creation, LUNs creation, logical volumes creation, file systems creation, etc are simplified. A variety of audit trails and event logging functions are supported, which aid in many management tasks including those connected with security. Storage provisioning is another important task of SRM, which allows the administrator to allocate the initial sizes to various server systems and flexibly enhance the allocations based on actual need. The SRM software also allows performance tuning by distributing the logical volumes across different arrays or disks.

2.10 Scalability

Scalability features built into any system ensure that the system is protected from short-term obsolescence while also optimizing the return on investment. The IMGEOS architecture is designed in accordance with this principle and scalability is inherent in all the envisaged subsystems.

By consolidating the disk system onto a high performance networked mass storage system based on FC SAN, a very high level of disk capacity scalability is achieved. Though initially configured for 75 TB disk capacity, the primary FC disk system is envisaged for scalability up to 400TB and more. The second-tier disk system is also configured on similar lines. These systems are mid-range high available storage systems. These systems were procured with 40% more expandability. If required, additional mid-range storage systems (SATA) can be added and consolidated.

All FC storage systems support disks capacities of 146GB, 300 GB, 450 GB and above. Mid-range storage systems support 750 GB, 1000 GB and above. The SAN switches are configured with 8 Gbps speed. But the above said storage systems are specified with 4 Gbps speed and can be upgraded to 8 Gbps as and when required.

The tape library is scalable from the initial 16-drive, 4000 slots to 64-drives, 10000 slots configuration, which translates from 6.0 PB to 15.0 PB (using LTO-5). Apart from the built-in scalability, modular augmentation of the disk and tape library systems with additional subsystems will further extend the capacities as well as the life of the infrastructure.

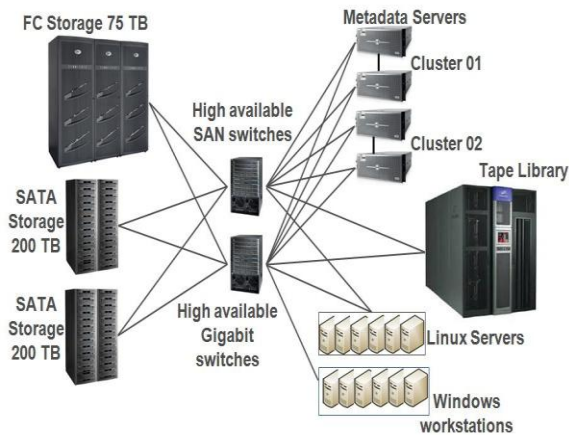


Figure 3. SAN Components

3. ARCHIVES MANAGEMENT

As of now, in IMGEOs SAN, the remote sensing data acquired at Shadnagar (India), Antarctica and Svalbard (Norway) is archived. The data from current missions viz. RISAT-1, Resourcesat-2, Cartosat-1, Cartosat-2, Oceansat-2, Resourcesat-1, NOAA-19, NPP and METOP-A/B is stored onto SAN.

NRSC has around 950 TB of data from earlier missions archived on optical/ tape media. The 95% of data porting onto IMGEOs SAN is completed. After this all the remote sensing data available with NRSC will be in online archives for data processing work centers. This requires managing the huge data efficiently to ensure automatic generation of data products irrespective of the date of pass.

Data Management involves storage, access and preservation of the data. In IMGEOs data management activities cover the entire lifecycle of the data, from acquisition on DI system to product delivery through FTP server and from backing up data (as it is created) to archive data (for future reuse).

Specific activities and issues that fall within the category of Data Management include:

- Data access
- Metadata creation
- Data storage
- Data archiving
- Data sharing and re-use
- Data integrity
- Data security

The data management is done using two broad components: SAN File System (SFS), a high performance data sharing software, and Storage Manager (SM), the intelligent, policy-based data mover.

3.1 Operational Scenario in Tiered Storage

Data received from Data Ingest systems is stored in high performance disk storage (Tier -1), which is being used for processing by Data Processing Servers. Subsequently, the data is copied automatically to the tape library based on the set policies. The data remains in the tape library as an online

archive for later usage. Three more copies of the same data is made on tapes, one for online backup of the archive (which shall remain on the tape library) and other two for vaulting. In the vaulted tapes one copy is being stored in the vaults and the other moved to other location for Disaster Recovery purpose.

Based on the set policies, the data in the tier-1, is automatically moved to another storage tier (tier-2), which is created using lower performance disks. The file path as seen by the processing server will remain same. In case, the file is removed from tier 1 or tier 2, the same file is restored from tape storage. After restoration, file path and name will appear as before without needing any changes from user-end.

3.2 Data Protection

As with any form of asset management system, cost control is only one of the key concerns. Data protection is paramount, especially in geospatial activities, because images cannot be recreated. The ability of an archive system to keep a digital asset safe in the long term starts with the ability to protect it in the short term. With file systems, even in relatively modest environments, growing into the 100s of terabytes, the traditional backup and recovery paradigms break down. Traditional backups require a nightly file system scan to identify new and changed files. If a file system has millions or 10s of millions of files stored within it, the time to scan that entire directory structure spans hours. Only after the scan has completed data will be copied to another location for safekeeping.

Traditional backup paradigms also require periodic (weekly or monthly) full file system backups. With 100s of terabytes, this too can take hours or days. And since geospatial data is relatively static – once created it is unlikely to change – repeated backups of this class of data add little value and instead increase costs as more and more media are used to stored the same, unchanging data. Using a traditional backup strategy for this class of data often results in throwing technology at the problem: deploying very high speed RAID and a sufficient number of tape drives to meet shrinking backup windows. While this strategy can work, it is quite expensive and contradicts the organization strategy of deploying technology to meet its needs. After all, backup copies are merely insurance policies against a problem with the primary copy. Recovery procedures are equally difficult as the process of restoring millions of files and terabytes of data can take days or weeks.

SFS takes a different approach to data protection. As each file arrives into a SFS volume, it is placed on a queue for the archiving component of SFS, Storage Manager, to determine how to protect that file. The archiving component uses data movement and protection policies, defined by the system administrator, which define how new or modified files should be replicated to another tier of storage, either disk or tape. When replication occurs, multiple copies of a file can be generated. Typically, this will be one copy for onsite storage (the copy for repurposing) and one copy for offsite storage (a disaster recovery copy). Through this process archiving occurs and there is never a need to perform a full backup. Instead, the organization automatically fulfills cost reduction strategies along with data protection. Additionally, SFS offers data integrity checks, performing a checksum on files as they are written to tape and then verified when restored back to disk. Through this process, SFS offers an

additional measure of data protection to prevent accessing damaged data. Restoring a large storage system after a disaster is equally efficient. SFS is able to restore the file system name space, the mechanism that allows users to access their data, at a rate of 20 million files per hour. Once the name space is recovered, the data is fully accessible. As users access their data, it is automatically pulled back from the replicate copy on the other storage tier. If the file is not accessed, it is not pulled back, eliminating any unnecessary data transfers.

3.3 Storage Policies

Storage is not only an object, but a service. Storage provides both short and long term business object storage and retrieval services for the organization. Defining clear storage management policies is more than just storage hardware and management software. The ultimate goal of storage service and any storage management policy tied to it is to deliver a cost-effective, guaranteed, uniform level of service over time. Achieving this goal is one of the harder aspects of storage management.

Storage policies act as the primary channels through which data is included in data protection and data recovery operations. A storage policy forms the primary logical entity through which data is backed up. Its chief function is to map data from its original location to a physical media.

The growth in data is making it increasingly difficult for administrators to manage routine maintenance tasks, such as back-ups and disaster recovery provision. This is made more acute by shrinking windows for these tasks. As data volumes keep increasing each year, it becomes increasingly important that data storage policies are put in place. A data storage policy is a set of procedures that are implemented to control and manage data.

A data storage policy is how the data should be stored, ie on-line, near-line, or off-line, as effective archiving can dramatically reduce the size of daily back-ups. A Storage Manager storage policy also defines how files will be managed in a directory and subdirectories.

Storage policies can be related to one or more directories. All files in that directory and sub-directories are governed by the storage policy. The connection between a storage policy and a directory is called the relation point. The storage policies are created based on the following parameters.

- Number of copies to create
- Media type to use when storing data
- Amount of time to store data after data is modified
- The amount of time (in days) before relocating a file
- Amount of time before truncating a file after a file is Modified

Separate Storage policies are created for each satellite and these policies are monitored for compliance of the policy.

3.4 Monitoring

The logs/ admin alerts of metadata servers, SFS and SM are continuously monitored for probable alerts. Health check is periodically done by running various diagnostic checks on the SFS. Whenever required, the current state of the system

is gathered as a snapshot on all component configured. This assists to analyze and debug problems in the storage system.

3.5 Verification

The data received from different remote sensing satellites is stored onto tape library as per defined storage policies. For data product generation, ADIF and FRED files are required. Hence, every month, availability of FRED data against ADIF is verified.

Over a period of time the FRED data of any satellite will reside only on tapes as the files will get truncated on disk. Therefore in future, if old data is required, invariably tape has to be accessed. As tape is a magnetic medium with electro-mechanical components, its reliability has to be checked regularly. To meet this requirement, daily few tapes are verified by performing a random read on each of the tape. This verification is a continuous process.

CONCLUSION

Using hierarchical storage that employs a mix of storage technologies allowed to build a cost -efficient multi-tiered storage system that meets the performance goals and recovery/archival requirements. Incorporating automatic migration and archiving of data files removes user intervention from the storage policy procedure and enhances data integrity for large - scale environments.

A good archive strategy is a vital part of organization Information Technology policy that can realize massive storage cost savings. The combination of intelligent archiving and data preservation software coupled with the latest high speed tape libraries gives best value, protection, and operational cost savings.

By balancing hardware performance and capacity requirements with the way data is actually managed, is the most cost-effective solution.

REFERENCES

- Jon William Toigo, 1999. The Holy Grail of Data Storage Management,
- EMC² series, 2009. The Information Storage and Management Book