iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

Offline Archive Catalog for Robotic Tape Library

SVSRK Kishore[1], ANSV Murali Krishna[2], K Mounika[3], C Pradeep[4] , C Chandrasekhar Reddy[5]
[1, 2, 3, 4, 5] *Computing, Storage, Network and Visualization Facility, Data Processing, Products Archival and Web Applications Area, NRSC/ISRO Hyderabad.*

Abstract: ***Remote Sensing Satellite data has been a proven source in providing vital information pertaining to monitoring and management of natural resources, understanding weather and climate, support and management during disaster events ISRO carries out the operation of Earth Observation i.e., Remote Sensing through NRSC. The center receives data from remote sensing satellite missions, archive the data, process the data and generate products and disseminate the data to various users across the globe. Considering the utilization and need for satellite data, so far, ISRO has launched 23 IRS satellite missions out of which 13 are active and 10 have completed their service. The data products from the completed missions and data being received from the current missions are more valuable assets, managing them is a big challenge for the organisation. At, NRSC earth observation data and products from satellite missions are managed inside the III TIER SAN (Storage Area Network). The SAN handles all prospects of managing huge volumes of data-sets – it incorporates major data management capabilities like data archiving, data administration and data access inside the Data center. It is the central storage and archiving infrastructure facility and currently hosting around 6.5 Petabytes of remote sensing data. The III tier SAN uses high capacity robotic tape library for data storage, data archiving, cataloging and data retrieval. The robotic tape library is well-designed for handling data volumes of mission data from current and future IRS satellite missions. The tape library operates under the management of the tape library Data Server known as 'Media Storage Manager' (MSM). The media storage manager maintains a database that contains system managed volume information, volume status information required to dispatch the access data, mount volume, store data into a volume or to eject a volume. This paper describes the design and implementation of an offline archive database catalog developed using mysql database for supporting the TIER III – Tape Library minimizing its impacts on operational set back. It can be brought into activation during the absence of the actual active online Library catalog.***

Keywords: ***Data Archival, Storage Area Network, Media Storage Manager, Volume, database***

I. INTRODUCTION

National Remote Sensing Centre, an entity of Indian Space Research Organisation that is responsible for performing remote sensing operations on Indian Remote Sensing (IRS) missions that are launched with various objectives. These satellites are primarily programmed to serve the collection of Geospatial data and use them significantly towards the development of the society. The data that are collected from these satellite sensors of different characteristics are to be processed under various segments that converge into data product. The Satellite data being an important asset and the associated operations are the functionalities of the organisation, prime set of supporting technologies are integrated to form the *Data Center* that houses three major IT infrastructure systems: Computing, Networking and Storage. In the computing layer, the applications live and run. The network layer provides connectivity between computing nodes. Finally, the storage layer is where all the data resides.

In a large distributed enterprise like Integrated Multi-mission Ground Segment for Earth Observation Satellites (IMGEOS) the data storage was very complex and challenging task. Keeping track of all the challenges and requirements a centralised *STOARGE AREA NETWORK (SAN)* that can handle huge voluminous data with high availability was brought into existence.

Internally, SAN was constructed as III TIER architecture that comprises of Disk Arrays and Magnetic Tape storage devices. To manage such huge voluminous data, SAN adopts the *Hierarchical Storage Management (HSM)* – otherwise known as *Tiered Storage*. In HSM environment, the system automatically moves data between the high-cost and low-cost storage devices with minimal impact on data operations. Policies were created to migrate massive amounts data between different logical and physical locations. The migration of data is based on how often the data is most likely to be needed in the future.

Performance and availability are the two major challenges faced by storage infrastructure. On the performance front, is constituted by the collective conduct of disk and tape storage devices to keep pace with their peer devices to support operations. So storage performance isn't a simple matter and requires a good level of knowledge of the underlying components. On the topic of availability, because of the central and crucial nature of storage, it is designed to deal with both hardware and software failures. Even though the technology in the data world is evolving very fast, leveraging data needs and maintaining data integrity means functioning with well built applications to overcome shortcomings in this vast and wide domain.

©IJRASET (UGC Approved Journal): All Rights are Reserved

This paper describes the design, development and deployment of the database records which is called as offline archive catalog for robotic tape library.

A. Key Features of the application are

1) Preserve a independent copy of information against technological obsolescence and physical damage
2) Fast response on searches
3) Information on media, data movement, retention management and media lifecycle
4) Easy access and navigation of records
5) Enhanced user services
6) Early retaliation on risks
7) Improved operational competence
8) Detailed data integrity and consistency reports

The system is designed to handle two categories of functions. The first category of function is meant for substituting the operations in the absence of the online library catalog. The second category of functions is provided for internal use to carry out administrative operations and background processes that support to improve workflow.

B. The set of accessible functions are

1) Administrator Authentication
2) Catalog view
3) Import or export database copy
4) Status Monitoring

C. The internal functions include

1) Export the database
2) Add an entry to the catalog
3) Update Information periodically
4) File maintenance report
5) Collect Audit report
6) Save audit reports

D. Architecture

From the data sources, processing and applications, several characteristics of remotely sensed data that impose requirements on their storage and retrieval can be derived.

It is important to maintain information about the data and keep track of them and search for them when needed. Considerable techniques and expenses are involved in satellite systems and instruments for acquiring the data, which, obviously, cannot be reacquired for a time period that has passed. Therefore, all data deemed of sufficient value need to be permanently preserved.

The Robotic Tape Library is deployed as the III tier of out III levels in the HSM based SAN storage architecture. The Tape media continues to be a popular choice for long term storage requirements such as backup and archived data. Although it is ancient, it usually offers high capacity at a decent price point.

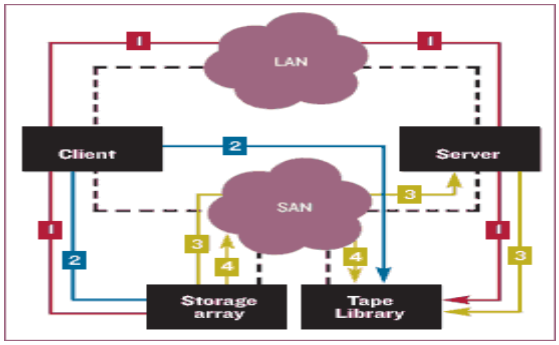


Figure 1: SAN storage architecture

©IJRASET (UGC Approved Journal): All Rights are Reserved

Offline Catalog for Robotic Tape Library is designed based on component based architecture. The component based architecture focuses on decomposition of design into logical components that bring forth defined interfaces containing methods, events and properties. The architecture focuses on grouping related functionality within the application that enables flexibility and maintainability.

E. Presentation Layer

It facilitates the administrator to interact with the application for browsing and viewing the records followed by visualization results. The presentation layer is designed to capture the uploaded inputs which are customized based on the copy obtained and the outcome of the latest database resulted after a specific time interval. The interactive and monitoring services are rendered through a rich graphical user interface. Visualization features for archived records, Media information, Media Summary and health status of Media designed using JSP an open source mapping framework based on Java.

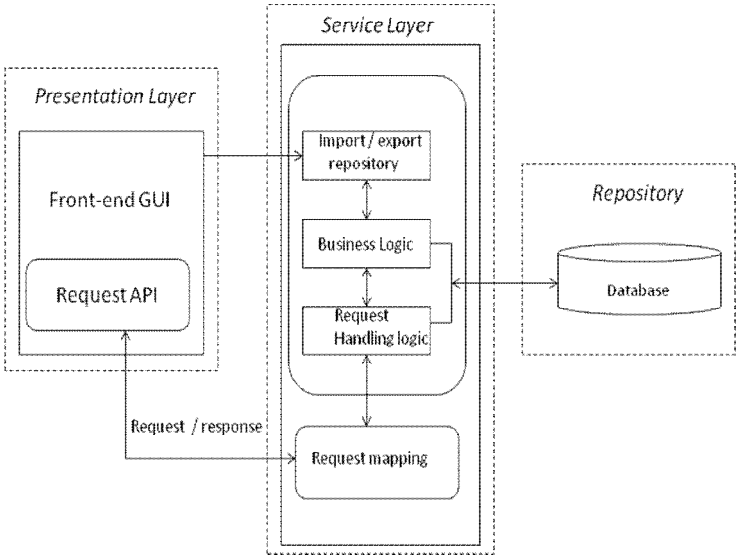


Figure 2. Offline Catalog Architecture

F. Service Layer

The service layer is embedded with business logic / controller; it consists of functional building blocks for processing the inputs captured by the presentation layer. The inputs received are channelized to various database tables with varying granularity, comprising of processing modules and Data Access Objects that manipulate the repository. The results are subsequently formatted as frames and made available to the presentation layer. The Front-end GUI is responsible for rendering visualization to the presentation layer. It provides display of respective table area.

G. Repository

The repository is the database that contains heterogeneous information comprising of tables under various granules to be managed by indexing. It provides the administrator with the ability to search for a data or media of interest and to narrow down the searches by specifying the parametric criteria, to minimize the extraneous load on to the system.

H. Design

1) *User Interface :* The presentation layer is designed to provide user interface for browsing and viewing from collection and archived catalog, status monitoring, related ancillary services like uploading and downloading files when required, administrative and control functions. The aim of designing presentation was user friendliness, appealing GUI, easy navigation and flexibility for re-configuring user defined changes. Validation of user inputs is done thoroughly and file uploads and format checks are very specifically done through the JavaScript validations. The GUI layout is brought to display and managed to be presented in blocks of information progressively within a single page minimizing navigation of the page to corners.
2) *Operational Logic:*The operational logic incorporates various components like the business logic, request mapping services and request handling logic. The business logic deals with set of rules and defines interactions between business objects. It executes the workflows defined to process inputs captured and queries identified at presentation layer and sends results back to

©IJRASET (UGC Approved Journal): All Rights are Reserved

presentation layer. The three layer architecture is implemented using MVC design pattern using JSP and java concepts. Model-view-controller isolates business logic representing workflows and rules for data manipulation from user interface and controller to manage communication between the business logic and user interface by listening to events and routing them to the appropriate adapter and subsequently sends corresponding results back to the GUI for visualization.

Record Creation–information created, received and maintained as evidence and information by an organisation, in fulfillment of policy obligations or in the transaction of data around in the chain.

Records management–file information management is the practice of formally managing records within a file system (electronic and or paper) including classifying, capturing, storing and disposal.

Digital continuity is the ability to use digital information in the way that you need, for as long as you need.

II. DATABASE / REPOSITORY

Large volumes of archived data catalog are archived over years since the first inception of the Tape Library in the year 2011. The repository is designed considering the data to be stored, transactions to be performed and their frequency. The database and their tables are organized into relations and its integrity and dependency is defined using primary key and reference constraints.

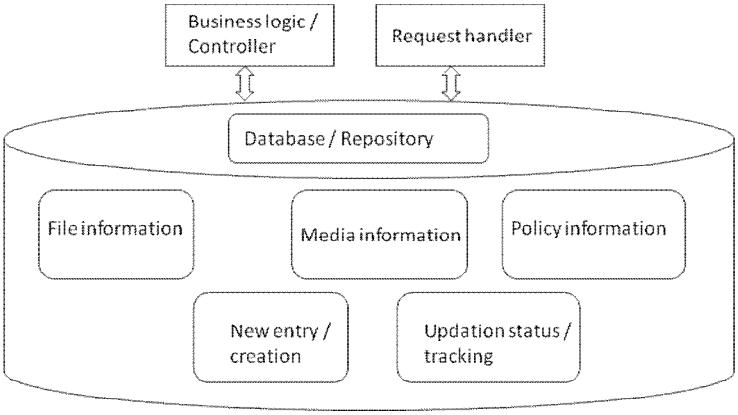


Figure 3. Offline Catalog Repository

III. WORK FLOW

There are three types of work flows. The first one is the browsing and view of the archive catalog. The second one is the copying and updation of the repository in par with the original catalogue and the third one is the replacement or substitution to work combination with the other storage components in place the actual archive catalog database.

The work flow in each case is optimised with the turnaround time for delivering a service with minimal delay. Re-entry points of services have been designed for easy resumption of system that was waiting for any archived data from the library.

A. It starts with the simple dashboard to keep track of, make records and those in entire archives, discoverable to all queries.

B. Record copying begins with discovery of records and copying them into our catalogue. Once it has found a document within discovery, you can request digital or paper copies of it.

C. All requests start with a page check. The page check involves a member of record finding the original document and reviewing it to assess if the information can be located and its suitability for copying.

D. Record Addition–to add new record we collect the following information –details from policies, data set, associated detailed parameters and media information.

IV. SECURITY

As the application is deployed in internal network, the vulnerability of being attacked or security being breached is negligible. Mostly this application is designed to be used by the administration team in the storage domain. The other aspects of security is handled through the following points

authentication and authorization are achieved through password protected login mechanism and verification so that only valid users are allowed to access the application.

©IJRASET (UGC Approved Journal): All Rights are Reserved

Another copy of the application and database are kept in a separate location other the hosted server to provide another layer of restore for the application.

A. Implementation and Deployment

Front end components are designed using JSP, CSS and JavaScript. Netbeans IDE, used for front-end design for this application, and this application is light weight and it is hosted with a local glassfish server and this is available to the user on internal network. On successful implementation, the application can be accessed over the web user interface that provides the login prompt and then the followed functionalities of the application.

Click on the + sign either to create an entry or upload a file...

Click on the update button to take latest information / update file to self update the repository.

To use the API, we need to have a hostname, port number and credentials to perform any action with respect to the application.

The Application is deployment on a windows based server machine with all recommended hardware settings.

V. CONCLUSIONS AND RECOMMENDATIONS

Based on the idea that archive data information can be made as a catalogue application designed to be operational in the environment with groups of data sets such as satellite raw data, FRED and products at various levels, the advantages of this collection is to propose an support alternate system to support data search and identification of data from a huge collection. Catalogue is useful where in the case of problems related to missing, lack or connection loss; we always find a way to convey data. The results from simulation show that the proposed idea improves the all data retention and preserves accordingly to the data delivery ratio. This solution involves three levels of communication: Metadata servers (HSM) to tape library, Tape Library to catalogue software and node to node.

VI. DIRECTIONS OF FUTURE WORK

For future work, we can plan to evaluate it using other metrics and then develop a cross layer so as to make it compatible with existing catalogue and make it more adaptive to new changes.

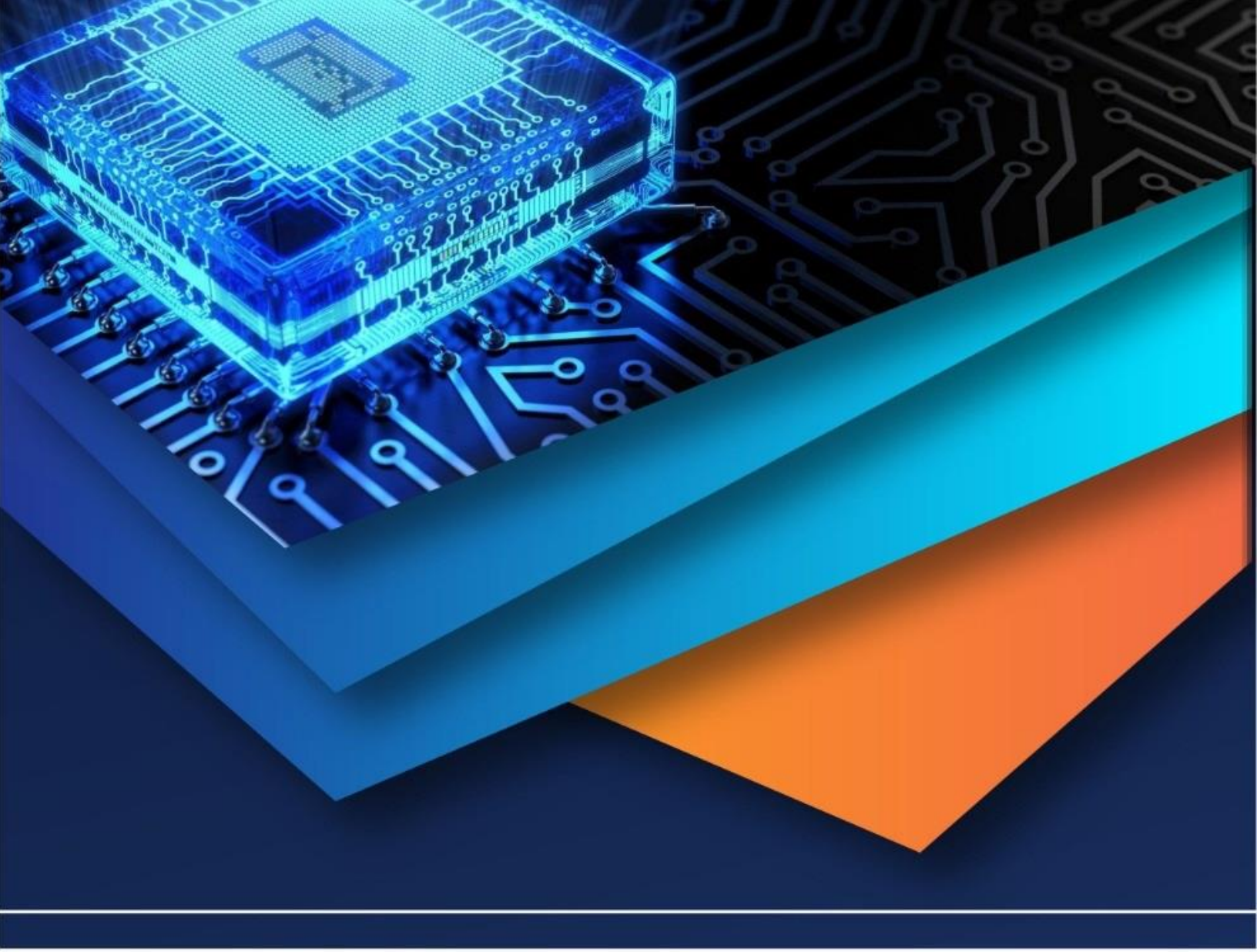
VII. ACKNOWLEDGMENTS

The authors would like to thank Deputy Director, NRSC for giving us the opportunity to realize the product. The authors would like to acknowledge the internal users of the storage services for their critical review to improve the infrastructure to meet the purpose for which it was envisaged.

REFERENCES

[1] http://irsa.ipac.caltech.edu/data/SPITZER/docs/spitzerdataarchives/
[2] http://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/documenting-collections/cataloguing-and-archives-networks/
[3] http://www.oracle.com/us/
[4] https://www.ibm.com
[5] https://www.cm-inc.com/
[6] Open source magazine on dependable databases and Devops way with the future
[7] Image Databases Search and Retrieval of Digital Imagery- Vitoria Castelli and Lawrence D. Bergman Wiley publications.

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)