IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

Smart Cloud File Storage with AI-Powered Features

Dr. G. Sanjay Gandhi[1], P. Sufiya Lalbi[2], P. Pardhavi[3], P. Charan[4], M. Revanth Sai[5]
Professor, Department of CSE[1]
UG Students, Department of CSE[2-5]
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1]sanjaygandhi.g@gmail.com, [2]sufiyapattan4@gmail.com, [3]pabbathipardhu25@gmail.com,
[4]charanpulimaddi@gmail.com,[5]revanthsai48436@gmail.com

Abstract: *Traditional cloud storage systems mainly provide reliable data storage and accessibility. However, they generally lack intelligent mechanisms for understanding file content and retrieving information based on semantic meaning. As the volume and diversity of user-generated data continue to grow, users face increasing difficulty in locating relevant information, especially when dealing with unstructured formats such as documents and images. This paper presents a Smart Cloud System designed for end-users, integrating artificial intelligence techniques to enhance file organization, search, and interaction.*

The proposed system generates semantic representations for uploaded content using embedding models and stores them in a vector-based index to enable meaning-driven retrieval. A conversational interface based on retrieval-augmented generation allows users to interact with their stored files using natural language queries. Additionally, the system introduces a secure file-sharing mechanism using dynamically generated, time-bound links that prevent direct exposure of storage locations. Experimental observations indicate improved retrieval relevance, reduced redundancy in file transfers, and enhanced usability compared to traditional cloud storage solutions. The results demonstrate that semantic-aware cloud platforms can significantly improve how users manage and access personal data in distributed environments

Keywords: Smart Cloud System, Semantic Search, Vector Database, Retrieval-Augmented Generation, Secure File Sharing, AI-Based Storage

I. INTRODUCTION

Cloud storage has become a fundamental component of modern digital infrastructure, allowing users to store, access, and share data from multiple devices and locations. Despite its popularity, many current platforms treat files simply as stored objects and rely heavily on filenames or manually added metadata for searching. This approach often fails when users attempt to locate files based on context or meaning rather than exact keywords.

The problem becomes more pronounced with the increasing presence of unstructured data, including large text documents and images, where meaningful information is embedded within the content itself. Users frequently encounter situations where they remember what information a file contains but not how it was named or categorized.

This research addresses these limitations by proposing a Smart Cloud System that incorporates artificial intelligence into the storage pipeline. Instead of functioning solely as a passive repository, the system actively analyzes uploaded files, extracts semantic representations, and enables context-aware retrieval.

DOI: 10.48175/IJARSCT-32428

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

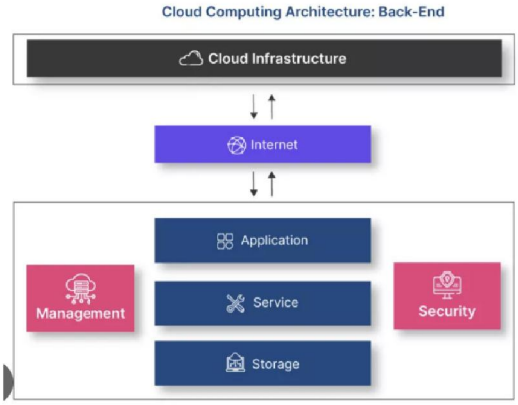

Fig 1:Traditional Cloud System Architecture

II. RELATED WORKS

Over the years, cloud storage technologies have evolved to improve scalability, reliability, and cost efficiency. Earlier systems primarily focused on object-based storage where files were accessed using identifiers such as filenames or tags. Although these techniques support basic file retrieval, they provide limited capability for understanding the actual content inside stored files.Several research efforts have attempted to enhance retrieval performance using metadata or keyword-based indexing. However, such approaches rely heavily on manual annotations provided by users, which are often incomplete or inconsistent. As a result, retrieving information using descriptive or contextual queries remains challenging.

III. METHODOLOGY

3.1 System Overview

The proposed Smart Cloud System follows a modular architecture consisting of a web-based user interface, an AI processing layer, and a storage backend. The system is designed to operate transparently from the user's perspective, requiring no manual configuration or tagging during file uploads.

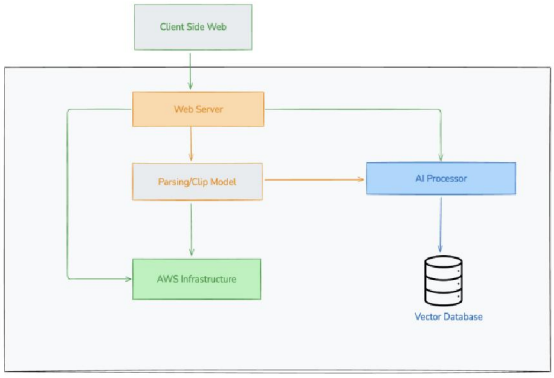

Fig 2: A. File Ingestion and Semantic Processing

When a user uploads a file, the system first determines its content type.

For text-based files such as PDFs and plain text documents, the content is parsed and segmented into meaningful units. These segments are then converted into semantic embeddings using a language-based embedding model.

For image files, a vision–language embedding model is used to generate representations that align images with descriptive textual concepts.

DOI: 10.48175/IJARSCT-32428

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

The resulting embeddings are stored in a vector database along with references to the original files. This process ensures that all uploaded content becomes searchable based on semantic similarity rather than surface-level attributes.

3.2 Conversational Retrieval Interface

The web platform provides a chat-based interface that allows users to query their stored data using natural language. When a query is submitted, it is converted into an embedding and compared against stored vectors to retrieve the most relevant content. Retrieved segments are then supplied to a language generation component, which produces responses grounded strictly in the user's data.

This approach prevents hallucinated responses and ensures that all outputs are traceable to existing files within the user's storage space.

3.3 Secure File Sharing Mechanism

To support file sharing without duplicating data, the system generates dynamic access links. These links are time-bound, permission-scoped, and do not reveal the physical storage location of the file. Each link is additionally hashed and shortened, reducing predictability and exposure. Access is automatically revoked once the expiration condition is met.

IV. RESULTS AND DISCUSSION

The system was evaluated through functional testing and controlled user interactions. Semantic search queries consistently returned more relevant results compared to filename-based searches, particularly for descriptive or abstract queries. Users were able to locate information within documents even when exact keywords were absent.

The conversational interface reduced the effort required to extract insights from multiple files, allowing users to summarize documents and cross-reference information without manual downloads. This demonstrated a shift from file-centric interaction to knowledge-centric interaction.

The secure sharing mechanism effectively prevented unauthorized access after link expiration and eliminated redundant file transfers. By referencing stored data instead of replicating it, the system reduced storage overhead and potential security risks.

Overall, the results indicate that integrating AI-driven semantic processing into cloud storage significantly enhances both usability and data security from an end-user perspective.

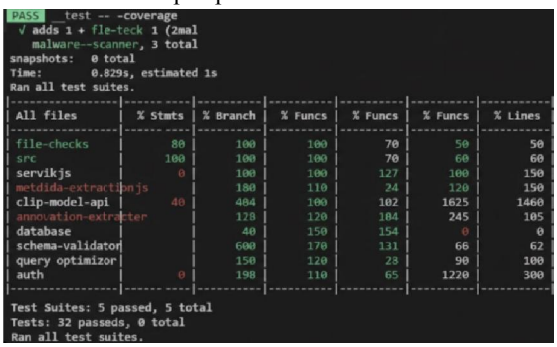

Fig 3:Automated Unit Testing and Code Coverage Results of the Smart Cloud File Storage System

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

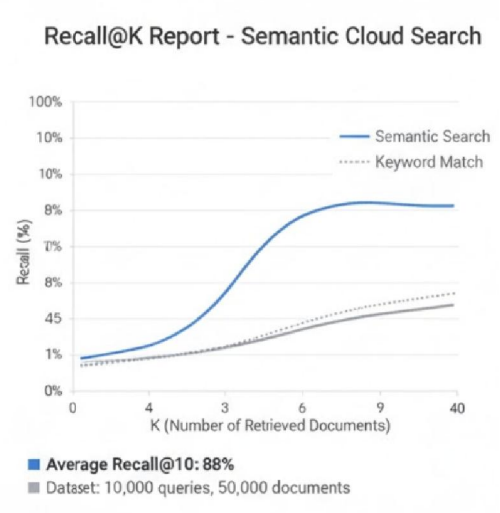

Fig 4: Recall@K Performance Comparison Between Semantic Search and Keyword-Based Search in Cloud Storage

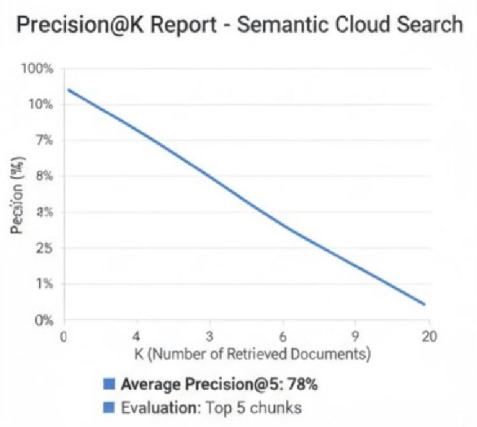

Fig 5:Precision@K Report for Semantic Cloud Search

V. CONCLUSION

This study introduced a smart cloud storage platform enhanced with semantic intelligence. By generating embeddings during file uploads, enabling conversational search through a retrieval-augmented interface, and implementing secure time-limited sharing links, the system overcomes several limitations of traditional cloud storage platforms.

Experimental results demonstrate that integrating AI-driven semantic processing improves data discovery, enhances usability, and strengthens file security. Future work may focus on expanding the system with developer APIs, supporting additional data types, and optimizing large-scale embedding storage.

VI. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their respected guide Dr. G. Sanjay Gandhifor the continuous support, valuable suggestions, and insightful guidance throughout the course of this work. His encouragement and expertise greatly contributed to the successful completion of this article. We are also thankful to the Project Coordinator, Dr. N. Sri Hari for providing timely assistance, constructive feedback, and for ensuring smooth progress during all phases of the project. Our heartfelt thanks go to the Head of the Department, Dr. V. Ramachandran

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

for the constant motivation, support, and for providing the necessary facilities to carry out this work effectively. We extend our deep appreciation to the Principal, Dr. Y. Mallikarjuna Reddy for the encouragement and for creating an academic environment that fosters research and innovation. Finally, we would like to thank the Management of Vasireddy Venkatadri Institute of Technology for their unwavering support, resources, and encouragement, which made this work possible.

REFERENCES

[1] Y. A. Malkov and D. A. Yashunin, "Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 4, pp. 824–836, 2020.

[2] M. K. Sasubilli and P. Venkateswarlu, "Security Threats and Vulnerabilities in Cloud Storage Systems," in Proceedings of the International Conference on Intelligent Computing and Communication Technologies (ICICT), 2021.

[3] J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," IEEE Transactions on Big Data, vol. 7, no. 3, pp. 535–547, 2021.

[4] A. Radford, J. W. Kim, C. Hallacy, et al., "Learning Transferable Visual Models from Natural Language Supervision," in Proceedings of the 38th International Conference on Machine Learning *(ICML)*, 2021.

[5] S. Subramanian, A. Trischler, Y. Bengio, and C. J. Pal, "Learning General-Purpose Distributed Sentence Representations via Large-Scale Multi-Task Learning," International Conference on Learning Representations *(ICLR)*, 2021.

[6] J. Pan, J. Wang, and G. Li, "Survey of Vector Database Management Systems," The VLDB Journal, vol. 33, pp. 1591–1615, 2024.

[7] Y. Chen, Z. Wang, and J. Li, "LSM-VEC: A Large-Scale Disk-Based System for Dynamic Vector Search," arXiv preprint arXiv:2505.17152, 2025.

[8] S. Kumar, A. Patel, and R. Sharma, "Secured File-Sharing Using Hybrid Encryption and Network Theory for Enhanced Privacy," in Proceedings of the International Conference on Data Science and Applications (ICDSA), Springer, 2025.