International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

DE Duplication Using MD5 in Cloud Based File Handling System

Dr. A. Kanimozhi[1]*, Suhail Farish M*[2]

[1]Research Scholar, Assistant Professor, Department of computer science
[2]Department of computer science, Sri Krishna Adithya college of arts and science, Kovaipudur, Coimbatore, Tamil Nadu, India
suhailfarish2004@gmail.com

ABSTRACT

In the era of big data and cloud computing, efficient file management and deduplication are critical for optimizing storage resources and enhancing data integrity. Traditional file systems often face challenges in managing large volumes of data and identifying duplicate files accurately. This project proposes a cloud-based solution leveraging MD5 hashing for improved file handling and duplication removal.

The primary objectives of this project are twofold: first, to develop a scalable cloud-based file management system capable of handling large datasets efficiently, and second, to implement a robust deduplication mechanism using MD5 hashing to identify and eliminate redundant files.

The project aims to address the challenges of file management in cloud environments by combining the scalability of cloud storage with the accuracy of MD5 hashing for duplicate detection. By implementing this system, organizations and users can effectively manage their data resources, reduce storage costs, and maintain data integrity. This abstract outline the goals, methodology, and expected outcomes of the project, emphasizing its contribution to enhancing file handling and deduplication in cloud-based environments using MD5 hashing.

Existing system

The existing system used in a single website or app. This user can post their reviews and rating below their product. The reviews can be posted only by the users who buys the product. In the existing system the fake reviews are identified only on the basis of repeated review posting from the same IP address on the same product Many works have been done in past in order to save the storage problem that is caused by data duplication. Data duplication has been the major problem and the technology developed in past was not able to solve the problem due to improper management of technology. For example, the confidential data in an enterprise may be illegally accessed through a remote interface provided by a multi-cloud, or relevant data and archives may be lost or tampered with when they are stored into an· uncertain storage pool outside the enterprise. Therefore, it is indispensable for cloud service providers to provide security techniques for managing their storage services.

Disadvantages

- More processing time.
- Chance of false result.
- Not user friendly.
- System maintenance is difficult.

Proposed system

The proposed system will utilize cloud storage services for scalable and reliable data storage. File uploads and downloads will be managed through a web-based interface, providing users with seamless access to their files from anywhere. MD5 hashing will be employed to generate unique fingerprints for each file, enabling quick comparison and identification of duplicate files based on their hash values. Data de duplication increases the amount of unwanted data in the storage unit by storing the multiple copy of same file. Data duplication removal technique uses file checksum technique to find duplicate or redundant data quickly. The technique calculates the checksum of the file when the file is uploaded and checks the newly calculated checksum with the file that are already store in database. They proposed a lightweight PDP scheme based on cryptographic hash function and symmetric key encryption, but the servers can deceive the owners by using previous metadata or responses due to the lack of randomness in the challenges.

Advantages

- Faster file searching.
- Reduce storage space by eliminating data redundancy.
- Ease to download and upload file.

System specification

SOFTWARE SPECIFICATION

Operating System: Windows95/98/2000/XP

Application Server: Tomcat5.0/6.X

Front End: HTML, Java, Jsp

Scripts: JavaScript

Server-side Script: Java Server Pages.

Database Connectivity: MySQL

HARDWARE SPECIFICATION

Processor - Pentium –III

Speed - 1.1 GHz

RAM - 256 MB (min)

Hard Disk - 20 GB

Floppy Drive - 1.44 MB

Key Board - Standard Windows Keyboard

Mouse - Two or Three Button Mouse

Conclusion

This project focuses on implementing this technique in NoSQL key-value stores, aiming to reduce storage consumption, improve performance, and ensure horizontal scalability on a Cloud Platform. While the project offers several advantages such as efficient storage utilization and simplified data management, it also requires an active internet connection as a potential limitation.

Scope of future enhancement

As part of future work, we would extend our work to explore more effective CPDP constructions. Finally, it is still a challenging problem for the generation of tags with the length irrelevant to the size of data blocks. We would explore such an issue to provide the support of variable-length block verification. Furthermore, we optimized the probabilistic query and periodic verification to improve the audit performance. Our experiments clearly demonstrated that our approaches only introduce a small amount of computation and communication overheads. Therefore, our solution can be treated as a new candidate for data integrity verification in outsourcing data storage systems.