ONLINE BACKUP AND RECOVERY USING AMAZON S3

Ameya Kanitkar[1], Shreyas Joshi[2], Sumit Chitnis[3], Mayur Gite[4]

[1]*Student, Information Technology Department, MMCOE, Maharashtra, India*
[2]*Student, Information Technology Department, MMCOE, Maharashtra, India*
[3]*Student, Information Technology Department, MMCOE, Maharashtra, India*
[4]*Student, Information Technology Department, MMCOE, Maharashtra, India*

Abstract

The project on File recovery and backup involves using of the emerging cloud based technology Amazon Simple Storage Service (S3).This project caters the purpose of safe and secure storage of the data as well as backup of that data for an Educational Institute such as a school or a college. The process in this project involves registering the client using Facebook or Google Accounts or simple sign up form. Then client gets his own manageable storage space for storing the data on the Amazon servers. The project involves creating server side webpage which eases the Signup process. For Backup and recovery of the customer data, project uses the java platform to create client side software. The software involves the simple UI for uploading the data onto the Amazon servers. A person can share the particular data by making that data public, data is stored safely on the cloud as they use different encryption methods. And it involves a feature known as time URL which makes selected data on the cloud available to person or group of persons for a specified interval of the time. The project uses Amazon S3 Technology because Amazon servers provide Easy robust and durable storage of the data and they provide complete (99.9999 percent) availability. And also they are inexpensive and reliable than the physical storage options.

Keywords: Amazon S3, Cloud Computing, Backup, etc

---***---

1. INTRODUCTION

The storage and security of the data is one of the most important aspects of the modern day computing. Keeping the data safe and secure is one of the top most priorities for most of the important aspects as data should be readily available every time the user needs it.

Traditionally storage of the data was limited to the hardware components like floppies, CD, DVDs, flash drives etc. Although all those hardware storage media shared problems and they were prone to the hardware failure. Cloud computing based recovery and backup systems provide the memory space for the storage of the user's content on the servers of the Cloud service provider. Each user is assigned an account, it grants access and authorization for the user to use the data. Cloud computing is emerging as a trending technology in the technological world, as it provides high durability and reliability of the data. It also provides several innovative features such as direct server streaming media, timed URL, sharing of the data within the clouds etc.

1.1 Aim and Features

Purpose of this project is to provide an 'easy to use' client side application for managing users files, while ensuring user safety and security of the data.

1.1.1 One Account Multiple Users

This software is made for Educational Institute wherein one Administrator (For eg: Head of Department) owns the account and his/her staff and students use it for uploading , downloading and sharing files.
The users must register once using the Easy Login Mechanism provided by the software.

1.1.2 Easy Login Mechanism

The users can either register by filling a simple form on the webpage hosted by the institute's server or even register using their current Gmail account. For Gmail Account registration method, we have implemented the OAuth 2.0 protocol.

1.1.3 Groups

Administrator can easily create, edit and delete groups of users from his/her institute with a single click.
Also adding and deleting members in a group has been made fairly easy.

1.1.4 Sharing

Users can share files with a single user or with a whole group with a single click.

For example: A teacher can share a presentation file with the whole class which can be included in a single group.

1.1.5 Timed URLS

Users can share a file for a specified time or permanently according to the need.

For example: If a user shares a file for 3 hours, after 3 hours, the URL for that file will not yield the file, the URL will expire and Amazon Server will reject the request.

1.1.6 Portability

User can easily install the software (basically Java Swing Client side software) and access the account files.

1.1.7 Centralized Control

Administrator has full control over a user's account. He may remove the user and define access permissions for users using the Amazon S3 bucket access policies. He may also assign storage spaces for particular class of users. For example: Teachers get 500MB Storage while Students get 100MB Storage space.

1.1.8 Extra utilities

- **Chat**

Users can use the live chat feature to chat with other group members in real time. Also group chat feature is up and running.
This not only limits this software to handle files but enables users to communicate important messages.

- **Reminder**

Students can set a reminder for submission of an assignment or planning when to retrieve or upload files for future use. The reminder feature notifies the user at appropriate time to perform certain tasks.

2. THE AMAZON CLOUD

The core concepts of Amazon cloud include objects and buckets.

- Object: An object is a file uploaded on the Amazon cloud. It has several properties like name, size, date modified etc. also called its metadata .Metadata is created automatically by the cloud service when a user creates a file. Each object can be retrieved even by a URL with the condition being that it should be signed with account owner's keys and a signing method used by Amazon.
- Buckets: A bucket acts as a container for holding objects. There is a limit on number of buckets a user can create. Also, bucket names are unique so as to avoid naming clashes. This is because the buckets are backed up in different regions to provide a great deal of availability.

2.1 Why Amazon S3

2.1.1 Pricing

Amazon AWS has introduced a new free usage tier which provides cheapest services among the other cloud systems.

2.1.2 Mechanisms

Amazon S3 provides mechanisms to control the access of your account. Following mechanisms are provided:

2.1.2.1 Identity and Access Management (IAM USERS)

AWS IAM allows the account owner to create multiple users and manage their permissions.

2.1.2.2 Access Control Lists (ACL)

Amazon ACL is applicable to every object   (an object is a file on cloud) and bucket (a bucket is a container or folder for files). It determines which groups or users can access certain objects and buckets.

2.1.2.3 Bucket Policies

Bucket Policies are used to choose which permissions to allow and which permissions to deny. Only a bucket owner has the authority to write bucket policies.

2.1.2.4 Query String Authentication

Query String authentication is used to authenticate browsers or HTTP requests.
These are authenticated by signing the URL with Owner's Credentials, appending a signature to the URL which should match with the signature computed by Amazon S3.

2.1.3 Storage Types

Amazon S3 offers several storage types which are different in terms of availability, reliability, storage regions etc.

The types include Standard Storage, Reduced Redundancy and Glacier Storage.

3. SYSTEM DESIGN

3.1 Users

Users can use their storage areas from the client machines using this Software. This same software provides Login Mechanism for user by communicating with the institute's server.

The software for use is hosted by the Institute Server for free downloading. The software includes User Interface for

Sharing Files, both with groups or single users, and for uploading or downloading files with a single click of a button!

3.2 Institute Server

The server is essential for authenticating users, managing users. Server also does the job of routing requests from users. Each request always goes to the server, gets authenticated by the use of signature generated from administrator keys and is sent to Amazon Cloud for validation and a reply.

The Server is totally controlled by the Administrator who can assign Storage limits to users. The Server runs PHP code which uses the Amazon S3 API for PHP for creation of Command objects.
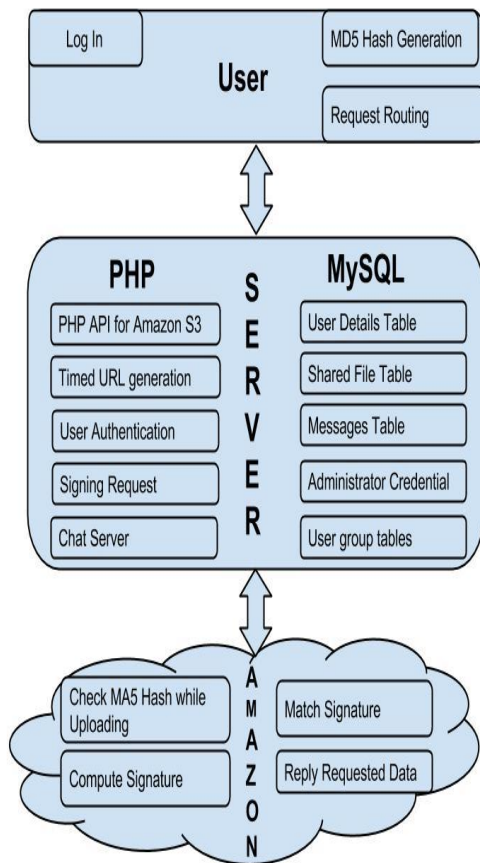

Fig -1: System Design

3.3 Amazon Cloud

The Amazon Cloud provides request servicing, request authentication, executing backup operations, object and bucket management, applying appropriate policies etc.

The Amazon Cloud also provides options to select a region to store your files which is geographically near to your location.

4. IMPLEMENTATION

The implementation of the system is described in this section in complete detail:

4.1 Login

User Logs in using username and password created during registration.

4.2 Send Request to Institute Server

User sends appropriate request like uploading, downloading, deleting, viewing or sharing the files, which is received by the institute's server.

4.3 Create Command Object

Institute Server creates a command object using PHP API for Amazon S3.

4.4 Send Request to Cloud

Institute Server forwards the complete request to Amazon Cloud.

4.5 Validate Request

Amazon receives and interprets the request. It stores the signature provided with the request and computes request signature by the same method as that of the Institute Server with administrator's keys. This signature is then matched with received signature. If signatures match, it means the requester also has necessary keys and has the authorization to access a certain resource. While uploading files, MD5 hash code of the file is sent to Amazon for verification of data integrity. It ensures that the file attempting to be uploaded is not a victim of attacks like "Man in the middle". If signature does not match, Amazon Server sends a 403 response code: Forbidden Request.
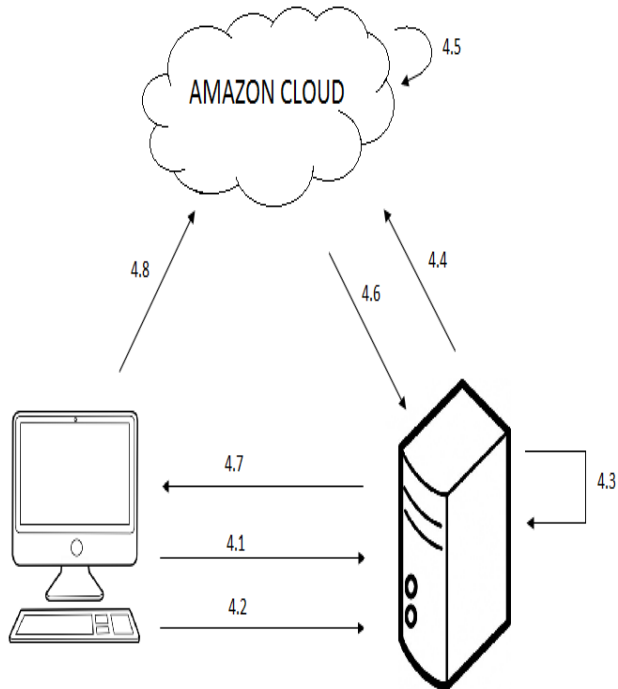
Fig -2: **Request work flow**

4.6 Response from Cloud

The reply is then sent back to institute server. It contains either requested data or an error code

4.7 Reply from Institute Server

The institute server returns it back to the client who made the request. Efficient communication mechanism (JSON/XML) is used for communicating data between client and the Institute Server.

4.8 User Execute the URL

The user executes the received URL which completes the task required by the user. So user gets his work executed.

5. CONCLUSIONS

Amazon cloud server's high Reliability and Availability can be harnessed for building better performing , secure and easy to use applications in daily life. Our work can be applied in different other domains as well. This paper demonstrates the use of client side applications for increased user friendliness which previously was not offered by the Old -school browser based applications.

ACKNOWLEDGEMENTS

We take this opportunity to thank our project guide Mrs. Preeti Joshi and Head of the Department Rupali Chopde for their valuable guidance and for providing all the necessary

facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information technology of MMCOE, PUNE for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

REFERENCES

[1]. Amazon S3. http://aws.amazon.com/s3/.
[2]. Arthur Rahumed, Henry C.H. Chen, Yang Tang, Patrick P. C. Lee, and John C. S. Lui: A Secure Cloud Backup System with Assured Deletion and Version Control.
[3]. Hobin Yoon, Ada Gavrilovska, Karsten Schwan, Jim Donahue: Interactive Use of Cloud Services: Amazon SQS and S3.
[4]. Jing Nie: Design the Desktop Backup System Based on Cloud Computing.
[5]. M. Noureddine, R. Bashroush: A Provisioning Model towards OAuth 2.0 Performance Optimization.
[6]. SNIA. Cloud Data Management Interface(CDMI). http://cdmi.sniacloud.com.131-133. (references)
[7]. Ghemawat,S.,Gobioff,H.,AND Lueng,S.-T. Google File System. In 19th ACM Symposium on Operating Systems Principles.518-519.
[8]. ZuquanTang,MinshenQu,ZhijingXu. Design the File Backup System Based on FileSystem Watcher. Electronic design project. Volume 18 phase 3: page 12.
[9]. Wuhan Zhao, Xiaoqin Hu, HongminDeng, Xiaoxu Ma. The Efficient file backup methods Based on Internet . Electronic design project.Volume 36 phase 2: page 247.
[10]. "AWS Forum." Available: https://forums.aws.amazon.com/forum.jspa?forumID = 12 & start=0
[11]. Bitcurrent, "Cloud performance from the end user perspective," 2011. Available: http://www.bitcurrent.com/download/cloud-performance-from-the-end-user-perspective
[12]. M. Kagan, "Global cloud performance data," 2011. [Online]. Available: http://www.cloudconnectevent.com/2011/presentations/free/76-marty-kagan.pdf
[13]. "AWS Import/Export." [Online]. Available: http://aws.amazon.com/importexport
[14]. "AWS Direct Connect," 2011. [Online]. Available: http://aws.amazon.com/directconnect
[15]. A. Gavrilovska, "High Performance IP-Based Transports,"September 2009.
[16] S. L. Garfinkel, "An Evaluation of Amazons Grid Computing Services: EC2, S3, and SQS," Center for, Tech. Rep., 2007.

BIOGRAPHIES

Ameya Vidyadhar Kanitkar, Student, Department of IT MMCOE, Pune

Shreyas Ramesh Joshi, Student, Department of IT, MMCOE, Pune

Sumit Suhas Chitnis, Student, Department of IT, MMCOE, Pune

Mayur Bhagwat Gite, Student, Department of IT, MMCOE, Pune