Authenticated Transfer of Files with Storage and Backup within a Cloud Environment

B V Akash[1], Dr. Murugan R[2]

[1]PG student, Department of MCA, Jain Deemed-To-Be-University, Bangalore, India
[2]Associate Professor, Department of MCA, Jain Deemed-To-Be-University, Bangalore, India

---

Abstract - The cloud computing service is valuable for various segments of human activities, making it a vital technology that has been adapted by businesses and education sectors. By using cloud storage services, clients can effortlessly store data and take advantage of high-quality on-demand cloud applications without managing software, hardware, or data on a constant basis. The cloud service provider moves data to the cloud storage servers so as to prevent users from having to deal with a lot of burdensome data-related tasks such as controlling physical data possession. Due to the physical possession of information that is outsourced, cloud services are safer, but there are new threats related to data security. Cloud storage allows users to put away their sensitive data and since users no longer control the service they use, it is imperative that robust security strategies are implemented to prevent unauthorized access to the system, or user's information. For this reason, a strong authentication scheme and data encryption scheme have been implemented for online cloud storage to address security threats. A user's data contents are encrypted before leaving the user's computer and an authentication scheme is used to verify each user and protect unauthorized access to all functions of the system.

INTRODUCTION

The ability to store data is crucial in today's world. Many companies, industries, and other organizations are so concerned about the safety and security of their data that it is the best method for storing files in the cloud. Cloud Computing gives you the best security for your data, or in other words, it ensures that the data is protected between users. Data can be stored in the cloud in huge amounts, which makes it the best way to store large datasets. Furthermore, data storage does not require additional internal power and resources, thereby enabling users to achieve additional cost savings. Cloud computing allows for multiple users to access and edit the same document or file because it enables people to access and edit the same data from anywhere. Using the system, users can collaborate in real-time on any file or document from anywhere in the world. In order to ensure that data read from Cloud

Storage is exactly the same as data written to Cloud Storage, Storage Transfer Service uses metadata available from the source storage system, such as checksums and file sizes. Cloud computing is a complex infrastructure of software, hardware, and processing. All of these components are made available as a service. Customers benefit from tangible cost savings and increased speed when using the cloud. The cloud enables companies to deploy applications rapidly because they can easily expand and contract their core technology components at any given time in response to the highs and lows of their business cycles. Cloud enablers such as virtualization and grid computing, which enable applications to be dynamically deployed onto the best infrastructure based on runtime requirements, can make this possible. It is worth pointing out that while this might seem attractive, there are still privacy and security concerns, as well as issues regarding reliability.

PROBLEM STATEMENT

Backups are essential to your disaster recovery plan, so you should ensure your backup process works as it should. Here are eight reasons disaster recovery strategy fails due to backup

Problems:

***Problem*: corrupted backup.**

There are several reasons why backups can become corrupted. Several factors can cause old media to become damaged or corrupt, such as handling error or simple aging. There is a possibility that read-only backups may not contain application-consistent data; as a result, when you restore files, applications may fail to run correctly.

Solution:

You should make multiple copies of your backups, preferably on different types of media. To ensure your data is consistent, you should employ application-aware backup strategies.

***Problem*: inaccessible backup.**

If The backup you have is the only copy you have, and you cannot access it during a disaster and can't restore your data.

Solution:

Ensure you make multiple copies of the backup and store them at least two different places.

***Problem:* backup job failed to start.**

It's that simple, if the job doesn't run, the backup won't be created.

Solution:

Backup jobs could be scheduled to run automatically rather than requiring a staff member to manually initiate them. In addition to preventing errors in parameter settings, automation also reduces the possibility of making them when you manually start the job.

***Problem:* backup job failed to complete.**

The job can run into problems during its run, which leaves critical information unprotected.

Solution:

Implement monitoring so that staff is notified if backup jobs fail to complete. You should regularly check your storage capacity to ensure you have enough space for the backup files. Maintain a record of database growth to make sure you have enough storage space for backups.

***Problem:* incomplete backup.**

It is not always the case that a job did not run to completion when a backup is incomplete. Files which cannot be backed up may not be added to backup procedures, either because they were not thought necessary or because they were not inserted into the scripts.

Solution:

Be Consider choosing what to back up in a comprehensive manner rather than selectively. When a new application or database is deployed, make adding it to the backup process a part of your deployment and change management processes.

***Problem:* slow backup.**

As part of your end-of-day process, backup jobs need to be completed. It is possible for slow backup procedures to cause delays the next day that affect productivity.

Solution:

In order to perform backups over the network, ensure that your network connection is reliable and you have enough bandwidth. Keep an eye on database development and how long reinforcements take to total; you will have to reschedule the work or do incremental reinforcements in order for the work to total, with a full reinforcement conceded to ends of the week or other longer periods of downtime.

EXISTING SYSTEM

A backup server is a dedicated server that is used to back up data from several client PCs on the same network. The server has a large amount of disc space as well as backup scheduling and management tools. Hard drives on backup servers are frequently configured to safeguard backup data and offer redundancy so that backups can continue even if a hard drive fails. Small businesses can save money by using onsite backup servers, but they can't safeguard your backup data against local failures or natural disasters.Your data and applications are backed up to a physical or (more likely) virtual backup server at a faraway data center managed by your organization, a hosting provider, or a cloud services provider over a corporate network or internet connection.

The most versatile method of backup is usually cloud backup. It may backup files, application data, and even complete physical or virtual servers. Backups can be scheduled as frequently or seldom as you like. Because cloud backup servers are often virtualized, you may quickly and affordably scale up and down as needed. For protection against local power outages or calamities, cloud backup eliminates the need to physically relocate backup media to another site (and the much greater RTOs and RPOs that might ensue).

PROPOSED SYSTEM

A cloud computing platform is a new type of computing that delivers hardware and software to users in a similar way to how electricity is provided to households today. AWS Solutions has a solution where Customers can utilize the implementation of the solution to build their own media analytical solutions within the AWS Cloud by modifying the implementation to fit their own workflows.

Basically, I came up with this Project during this pandemic crisis, Due to Pandemic mostly Businesses, Industries and Big Companies had to do everything online storing data, backup and security of that Particular data. So I came up with this project. My project title is "**Authenticated Transfer of files with storage and backup within a cloud environment** " where people can upload their data which can be accessed around the globe like High Availability. I deployed my Project in AWS(Amazon Web Service) which is a High leading Platform which provides services to the customers where cloud users can perform labs and projects. Also can achieve the desired output they need.

It took me Lots of days and months to do such project , Might look simple but there lots of programming languages like JavaScript, PHP and Backend for designing HTML and CSS. This can be done by any of you, if you have Knowledge of what exactly you are doing in it.

Implement Process:-

Users get a platform for uploading their data along with storage and backup options .

Furthermore, with encryption protocol, users can rely on that Data in a safe and secure manner. The main advantage of this project is that it can be available around the globe and they will get the data with low latency.

I made this project publicly available for everyone for example:- if suppose one user stored his data from Mumbai and the other user who is in the U.S he can use that data by downloading it.

Users can store data with large GB's and it can be easily accessible.

Creation Process:-

Under this I will be explaining, How I begin with creating the backend part to create such a wonderful project by using languages such as HTML(Hyper Text Markup Language) and CSS(Cascading Style Sheets).

HTML language I used to give a style of a font Basically to create a structure for a website. Which contains a heading, Paragraph, Links, etc. There are six types of header types like H1,H2,H3…….H6. I used H2 for a title.

<p> tag for Paragraph to write something which I will use further in a project for writing a Blog . I will create a section where I can see what they can do with such a website.

CSS has been used here for making a design or styling the webpage. Basically, we can call css as a supporter of HTML. Which describes how HTML elements are to be displayed on screen, paper, or in other media. You can style your sheets according to your preference . I used a normal sheet where I just added a Button and gave a design for it. You can color your website easily with styling.

Javascript is a high level programming language which is used to give an action to a particular website which we have created using HTML and CSS.

PHP(PHP is an acronym for "PHP: Hypertext Preprocessor" PHP is a widely-used, open source scripting language.) It is used to develop dynamic and interactive websites.

I used this script for the database process which is phpmyadmin.

Cloud Implementing Process:-

There are service platforms like Google Cloud Platform(GCP), Amazon Web Services(AWS), Microsoft Azure. But I chose AWS to deploy my project in it.

There are services like :-

- Amazon EC2 (Elastic Compute Cloud)
- Elastic Load Balancer
- Amazon RDS (Relational Database Services)
- Amazon IAM (Identity Access Management)
- Amazon S3 (Cloud Object Storage)

I will explain these services which will help you to make better understanding

Amazon EC2 (Elastic Compute Cloud):-

We got EC2 which stands for Elastic Compute cloud. This is a pure infrastructure as a Service or IAAS. This is what people thought cloud was all about initially. In a nutshell, EC2 can give machines or servers for your application needs. You can pick between different hardware sizes ranging a single core and 500 megs for Ram to 128 cores and 5TB of RAM. Of RAM not storage. Imagine assembling such monster size machines on premises. You can spin such machines in minutes.

You can choose between different vanilla operating systems images Linux, windows images and configure them any way you want to use them as web server or databases, deploy your applications to them. You can also choose from 1000 of preconfigured application stack images from amazon's marketplace to set everything up with a few clicks.

All of this is pay per use. If you are not using the machine, you can just stop the machine and your billing stops there. You bring it up again and your billing resumes.

So, let's say, I bring it up for 15 minutes. My boss calls me for a meeting. So, I stop it, finish up that meeting and then come back, start it up again and run again for another 15 minutes. Next, I head out for a tea break. So, stop it and then when I come back and start it back up again. Let's say I do these stop/start activities 3 times in an hour. But the actual run time of the machines is 45 minutes. What would my bill be.?

The answer to this question in 2017 would be guess what 3 hours. Because this was Amazon's minimum billing cycle. Whenever I bring up my machine and my machine goes into a running state, I would pay for the entire hour irrespective of whether I run it or not. When I stop it and start it up again that's a new hour.

Think of your mobile operator like Airtel. Their minimum billing cycle for calls may be a minute. Let's say I call up my mom to talk for 30 secs. I disconnect and call her back. I pay for the entire minute.

This was how Amazon's billing cycle functioned. And a lot of people didn't like it. They would have scaling environments where machines would come up and down in the same hour.

Newer clouds like Azure and Google learnt from this and introduced the per minute billing. With this model you could run a machine for 20 minutes to pay for exactly that amount of runtime.

A lot of people moved away from Amazon because of this. Then guess what Amazon does, it comes back in and smashes the competition by announcing the per second billing.

As of today, any machine you launch up you pay per second of usage time. Later Google and azure were forced to announce the per second billion too.

But look how we are talking. Earlier on premises I pay an upfront payment for my data center, contact my hardware vendors for any additional capacity that I need. Wait weeks or months for it to arrive and then rack and stack and start using it. This upfront cost is called CapEx or Capital expenditure. Here in the cloud world, I can spin up machines in minutes and start using it. There is no CapEx, what is called Opex or Operational Expenditure, Where I pay when using the service only. Which is every second I keep that machine running.

EC2 more information

AMI(Amazon Machine Image) :- This is a type of image which contains our operating machines like Linux, windows.

You can choose an instance type like t2.micro which will be your default instance which is used to give proper CPU performance.

You can generate a key pair which is used to access your ssh command which is basically like a linux command where you can set up your website which is what I did . Basically it's a server to manage your websites.

Elastic Load Balancer:-

Let's say I have a website, uploadmanager.com. I have a domain called www.uploadmanager.com and I have a DNS record pointing to my app server's IP. So, I usually get around 2000 hits a day and a good 65% average CPU utilization from my server. Suddenly many people started visiting my site and now my site goes from 2000 hits to 2,00,000. Can my server take it? No! What should I do? Scale. Yes. There are two kinds of scaling.

One is I can add a bigger machine. More cores, more memory, bandwidth support. This is called vertical Scaling. Vertical scaling in the cloud as we will see later on involves downtime. You will have to stop the machine, change your hardware profile and bring it up again.

In the cloud the preferred way is to use something called horizontal scaling. So instead of having one big machine and growing it up and down, I have multiple smaller machines. And? There's something missing here. Add a load balancer on top.

So, the machine is no longer the point of contact to the external world. The load balancer is. Which means my DNS records now point to my load balancer. So, user requests hit the load balancer and the load balancer spreads out the traffic to all of the machines under it according to different algorithms like round robin, weighted round robin, content based etc. Amazon has 3 fully managed highly available load balancers, the Classic, Network and Application Load balancer. We will be discussing more of this detail later.

Now back to indianhandicrafts.com, let's say I added the machines, setup the load balancer and was able to handle those 2,00,000 requests. But as with most things in life, my popularity died down and now I'm sitting at 50,000 hits. Good news about horizontal scaling is that I don't have to keep those 4 machines running all the time. I can just get rid of some of the machines and keep just enough capacity running under the load balancer. This wasn't the case with vertical scaling where I was stuck with that huge machine whether I'm using it or not.

But this is still a manual process of adding and removing capacity. What if I have a sine wave pattern which a lot of these ecommerce sites get, where during the day I have 100,000 requests and during the night I have 10,000 requests.

Amazon RDS (Relational Database Services):-

So, there are many products made by different vendors in the database world. We have Oracle, MongoDB, MySQL HBase, Cassandra and so on.

Chances are your company may be using one of these flavors of database right now.

If you want to set up something like this on cloud, you have two options. Spin up an EC2 machine, install it and configure, manage it yourself or use one of these managed database services that Amazon has to offer.

So, there's RDS, ElastiCache, DynamoDB, Neptune and Key Spaces.

Now we can broadly split the database world into two main categories

The older SQL database or Relational databases like Oracle Postgres and MS SQL server and the newer NoSQL or non-relational databases like MongoDB, Redis and Neo4j. Earlier we also used to refer to NOSQL databases as non-relational databases because they didn't support relationships and joins. Nowadays you don't call them non-relational anymore because there are some NoSQL databases like Neo4j which are just about relationships.

RDS is an AWS relational database service. I can install and configure Oracle or I can set up PostgreSQL. You can stop doing that because now most clouds have these fully managed database offerings like the Amazon relational database service or RDS.

A couple of clicks is all that it takes to set up a fully functional production database. Cool thing about this is all of the patching, backups and maintenance are all taken care of by Amazon.

You can set up patching windows and the latest patches will be applied to your database. They offer you 2 kinds of backups. A snapshot-based backup where I can just capture the state of my database and restore it with one click but also a point in time transaction log-based backup where I can go back to the last 35 days to every second if I want to.

What's also nice is they give you some nice high availability options too. With a click of a button, I can spin up slave in the other AZ and this will take over if anything happens to the master.

Good news about this is the endpoint to your database remains the same throughout this operation.

They also give you something called a read replica which you can serve your reads from.

You can choose between MySQL, Oracle, PostgreSQL, Microsoft SQL Server and Amazon's own Aurora. No SQLite or other flavors as of now. We will talk about this in more depth later on in the course.

Btw these are all relational databases and RDS only supports these flavors, if you want other relational flavors like DB2 or Times ten then you will have to setup and manage them yourself on an EC2 machine.

Amazon IAM (Identity Access Management):-

You can say Identity security and management**.** Anybody with a credit/debit card can go to AWS and sign for an account. This is the root account and it is the master account. He has full access to all of aws services. Imagine if I am in a team and we all share the root account. The architect as well as a new intern joining the team will get privileges to not spin up services but also delete resources. This is dangerous and this is the reason we have IAM.

IAM roles

Then we have a role. So, what is a role? A role gives you temporary token-based access to resources. Roles contact a service called STS which gives a token with an expiration time. This a token can be used to access services. What services can be accessed is determined by the policy we attach to the role.

Roles have a variety of use cases. First of all, it is used by AWS services to contact another AWS service on your behalf. For example, I have an application running in

EC2 and this app needs to upload data in S3. All I do is associate a role to the ec2 instance with a policy granting access to S3.

Next role can be used to grant temporary access to users to services in your account or to resources in another account. This is called a cross account role.

Let's say I have a corporate data center with 50,000 users.

Usually, all this information would be stored in an identity store like Microsoft AD or LDAP.

Now let's say I want to give all these users access to AWS. Who's going to create 50,000 IAM accounts in AWS. That's going to be somebody's full time job. Maintain 2 sets of accounts.

Rather you can implement something called federation and have the user sign through his corporate credentials and use a role in AWS that gives him access to AWS resources.

This is called corporate federation, use your existing identity store to log you in AWS and give you access to AWS resources.

What about mobile applications who need access to an AWS resource? Let's say an app like Dropbox that needs to store data in S3.

You can use Web federation which allows you to use OAuth and OpenID connect, have a user sign in using google, Facebook and twitter and then assume a role that gives you access to AWS services.

Amazon S3 (Cloud Object Storage):-

In the storage space we have Amazon S3. Which stands for simple Storage service. This is Amazon's poster child, one of the first services released by them and by far one of the most popular. Its unlimited storage and at a very high durability which

99.999999999. one of highest offered on the planet. It used mainly static content, image, video media, backup logs etc. It's your internet scale storage. One major caveat is and we'll discuss this later on is that it shouldn't be used for content that changes a lot like a database.

LITERATURE SURVEY

Amazon EC2 Service Security and Safety Cloud affiliates like Amazon Elastic Compute Cloud and IBM SmartCloud are changing the direction of leading companies by regulating their IT structures and providing an online cloud per the AMI EC2 Service study. It's not difficult to know what's happening today. you'll be able to usually buy or rent online and upload, retry, and package virtual images using the APIs provided by the cloud relationship. Consumers must be prepared to make, update and share virtual photos with other customers, a really important task in online cloud affiliates.

Cloud storage service provides a fast and convenient solution through the storage and sharing of objects of various types. However, the high-level management interface, while ensuring ease of use, obscures system implementation details and performance metrics. In this work, we performed a test study on the performance of the cloud user network for the Amazon S3 cloud storage service, as it was perceived by a group of home users distributed around the world. Since the dataset was obtained using the Bismarck platform, we report the overall performance rating of this service.

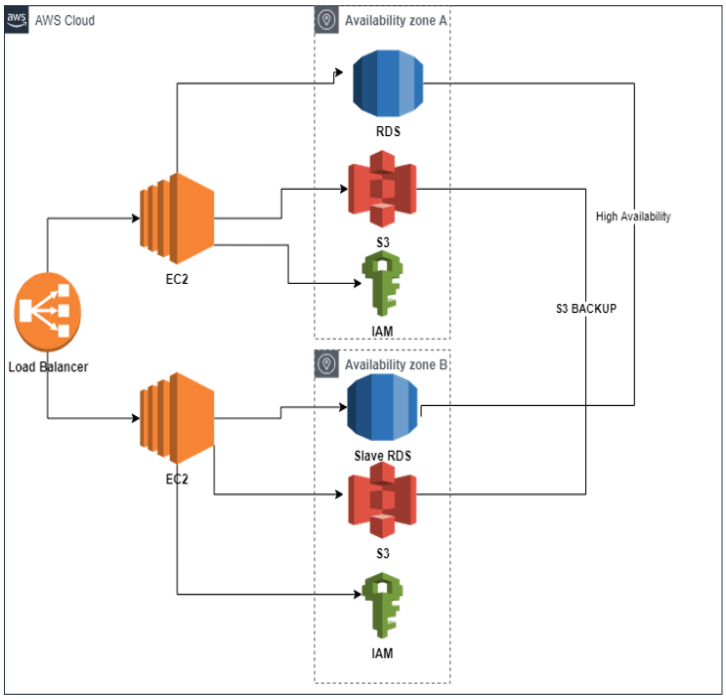
Load balancing in cloud computing data centers has been a main challenge and a full of life area of research in recent years. During this paper we've presented a survey on current load balancing techniques and solutions which are proposed just for cloud computing environments. Consistent with our study, cloud load balancing mechanisms are categorized into three main groups supporting their designing perspectives: General Algorithm-based, Architectural-based and Artificial Intelligence-based load balancing approaches.

The cloud paradigm is beneficial for these applications, but favors data access beyond individual key-value pairs. Therefore, they rely on traditional database systems.

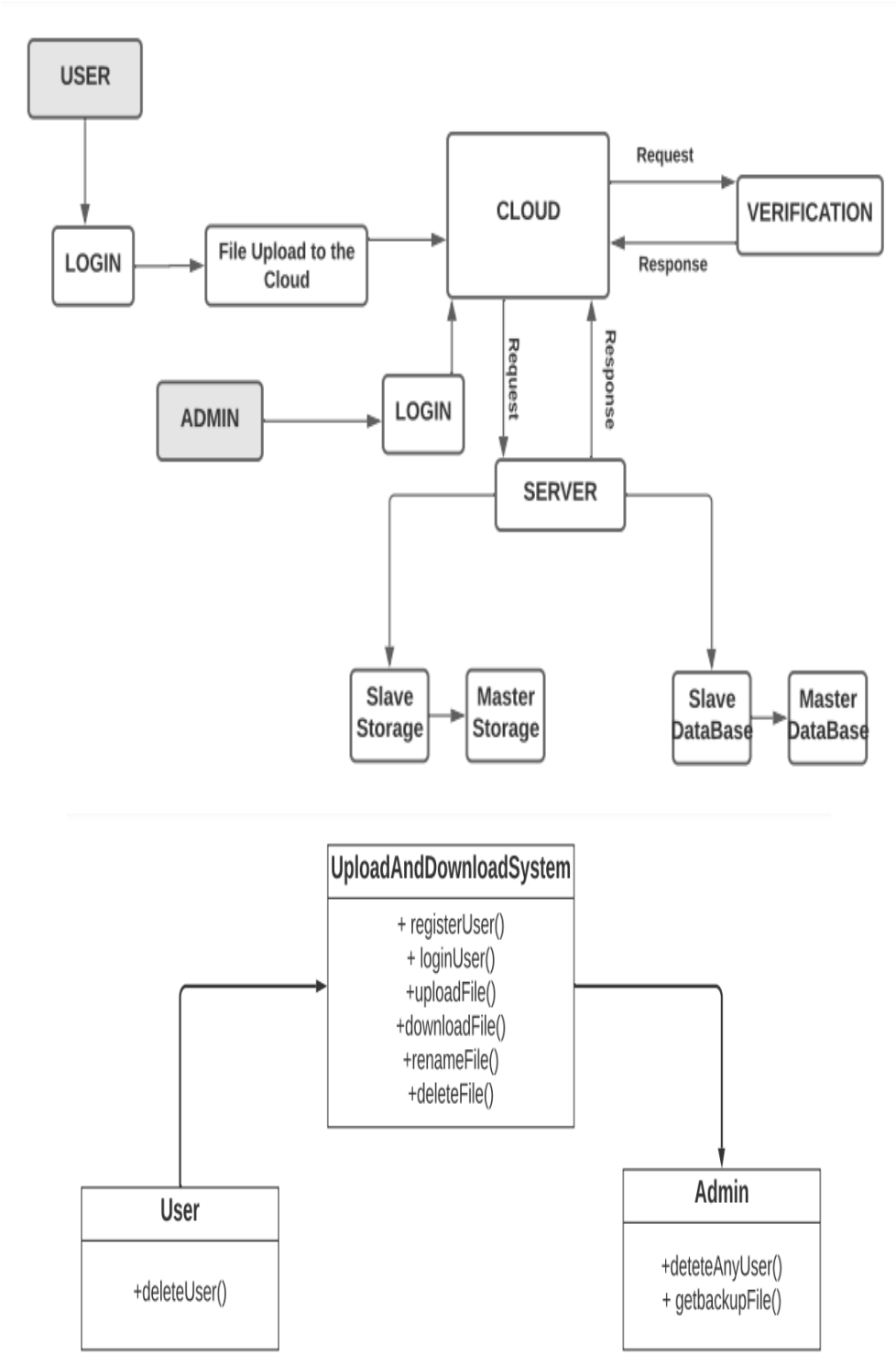
However, traditional database systems run on legacy hardware rather than cloud-based database services. Therefore, there is a gap between traditional cloud databases and modern cloud databases. This gap must be filled if the cloud is to support all kinds of existing and future applications.

Cloud computing is a new computing paradigm that provides organizations with innovative business models. Cloud computing moves application software and databases into large data centers where data management and maintenance can be inadequate. Security is an important aspect of quality of service. Cloud storage is much more cost-effective and profitable than previous traditional storage systems, especially in terms of scalability, cost savings, portability and functional requirements.

SYSTEM ARCHITECTURE



FLOW DIAGRAM

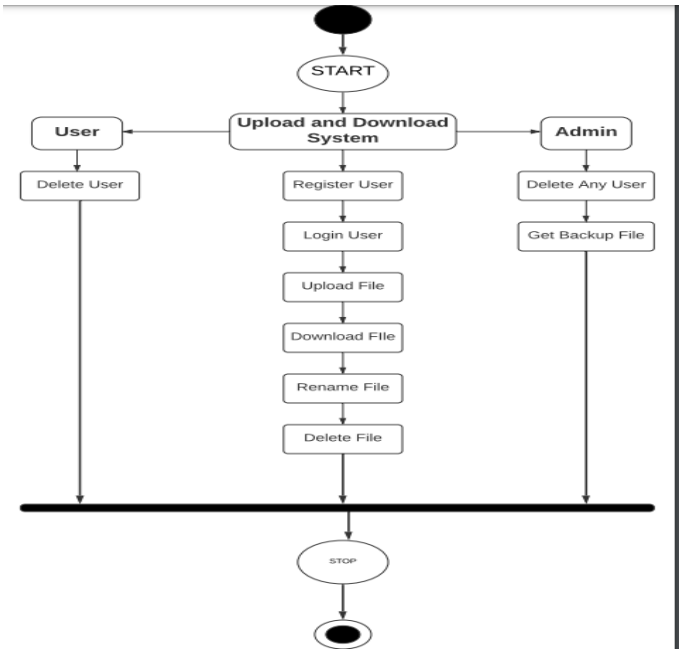


CONCLUSION

It is possible to securely store the large amounts of data generated by the cloud each day using encryption, as well as to retrieve them in case the cloud files are lost or deleted for any reason or in case the cloud crashes. Performing encryption ensures the safety of data stored in the cloud.

REFERENCES

[1] K. Ravi Chythanya, G. Sunil, K. Sudheer Kumar, Seena Naik Korra, A. Harshavardhan "Security and Safety in Amazon EC2 Service- A Research on EC2 Service



AMIs", 2019.

[2] Valerio Persico,Antonio Montieri,Antonio Pescape "On the Network Performance of Amazon S3 Cloud-storage Service",2016.

[3] Mohammad Reza Mesbahi, Amir Masoud Rahmani "Load Balancing in Cloud Computing: A State of the Art Survey", 2016.

[4] Pankaj Deep Kaur, G itanjali Sharma "Scalable database management in cloud computing Pankaj Deep Kaur , G itanjali Sharma",2015.

[5] Chengzhang Peng, Zejun Jiang "Building a Cloud Storage Service System", 2011.

[6] Clemens Zeidler, Muhammad Rizwan Asghar "AuthStore Password-based Authentication and Encrypted Data Storage in Untrusted Environments", 2018.

[7] Sai Shobh R, Dinesh Arpitha R "Data Storage, Security and Techniques in cloud computing",2018.

[8] chun-tinghuang, leihuang, zhongyuanqin ,hangyuan ,lanzhou ,vijayvaradharajan and c.-c.jaykuo "Survey on securing data storage in the cloud",2014.

[9] G.Pavani Durga, N.VijayaGopal "Efficient And Control Data Procedure Authentication In Cloud Storage",2021.

[10] Debajyoti Mukhopadhyay, Gitesh Sonawane, Parth Sarthi Gupta, Sagar Bhavsar, Vibha Mittal "Enhanced Security for Cloud Storage using File Encryption",2013.