**Heterogeneous Data Storage Management with Deduplication in Cloud Computing**

A PROJECT REPORT

in the partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

MAY 2020

**BONAFIDE CERTIFICATE**

**ABSTRACT**

Cloud storage as one of the most important services of cloud computing helps cloud users break the bottleneck of restricted resources and expand their storage without upgrading their devices. In order to guarantee the security and privacy of cloud users, data are always outsourced in an encrypted form. However, encrypted data could incur much waste of cloud storage and complicate data sharing among authorized users. We are still facing challenges on encrypted data storage and management with deduplication. Traditional deduplication schemes always focus on specific application scenarios, in which the deduplication is completely controlled by either data owners or cloud servers. They cannot flexibly satisfy various demands of data owners according to the level of data sensitivity. In this paper, we propose a heterogeneous data storage management scheme, which flexibly offers both deduplication management and access control at the same time across multiple Cloud Service Providers (CSPs). We evaluate its performance with security analysis, comparison and implementation. The results show its security, effectiveness and efficiency towards potential practical usage.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER No** | **TITLE** | **PAGE** |
|  | LIST OF FIGURES | Vi |
|  | LIST OF ABBREVATIONS | Vii |
| **1** | **INTRODUCTION** |  |
| 1.1 | About the Project |  |
| **2** | **SYSTEM ANALYSIS** |  |
| 2.1 | Existing system |  |
| 2.2 | Proposed system |  |
| **3** | **REQUIREMENTS SPECIFICATION** |  |
| 3.1 | Introduction |  |
| 3.2 | Hardware and Software specification |  |
| 3.3 | Technologies Used |  |
| 3.3.1 | Java |  |
| 3.3.3.1 | Introduction to Java |  |
| 3.3.1.2 | Working of Java |  |
| **4** | **Project Purpose and Scope** |  |
| 4.1 | Purpose |  |
| 4.2 | Project Scope |  |
| 4.3 | Product Perspective |  |
| 4.4 | System Features |  |
| 4.5 | Design and Implementation Constraints |  |
| 4.6 | Other Nonfunctional Requirements |  |
| **5** | **SYSTEM DESIGN** |  |
| 5.1 | Architecture Diagram |  |
| **6** | **SYSTEM DESIGN – DETAILED** |  |
| 6.1 | Modules |  |
| 6.2 | Module explanation |  |
| **7** | **CODING AND TESTING** |  |
| 7.1 | Coding |  |
| 7.2 | Coding standards |  |
| 7.3 | Test procedure |  |
| 7.4 | Test data and output |  |
|  | **SOURCE CODE** |  |
|  | **SNAP SHOTS** |  |
|  | **REFERENCES** |  |

**LIST OF FIGURES**

5.1 System Design

5.2 Sequence Diagram

5.3 Use Case Diagram

5.4 Activity Diagram

5.5 Collaboration Diagram:

5.6 Data Flow Diagram

5.7 Class Diagram

**LIST OF ABBREVATIONS**

**JDK**  Java Development Toolkit

**DEX** Dalvik Executables

**TCP** Transmission Control Protocol

**IP**  Internet Protocol

**HTTP** Hyper Text Transfer Protocol

**ADT** Android Development Tool

**CHAPTER 1**

**INTRODUCTION**

**Aim:**

The main aim of this project is to control file duplication in cloud computing.

**Synopsis:**

Cloud storage as one of the most important services of cloud computing. Data ownership proof is an essential process of data deduplication, especially for encrypted data. But this scheme does not provide flexible deduplication control across multiple Cloud Service Providers (CSPs). In this paper, They propose a multiple cloud service provider (CSPs) in which the data owner will upload the file and the hash MD5 algorithm is used to check data duplication during data storage at the cloud. CSPs. It can achieve data deduplication and access control with different security requirements. And also they have proposed a scheme called Provable Ownership of the File(POF). The result it is security, effectiveness and efficiency towards data storage management.

**CHAPTER 2**

**SYSTEM ANALYSIS**

**2.1 EXISTING SYSTEM**

In existing system a heterogeneous data storage management scheme, which flexibly offers both deduplication management and access control at the same time across multiple Cloud Service Providers (CSPs).They evaluate its performance with security analysis, comparison and implementation. They use Attribute Based Encryption(ABE) to realize deduplication data access control managed by data owner. This scheme was to solve the problem of access control .

Disadvantages:

* The security analysis and performance is not secure.
* They cannot overcome the issue of duplicated data storage in cloud computing.

They cannot solve the problem of access control.

**2.2PROPOSED SYSTEM**

In this work, we proposed to storage across multiple CSP’s and preserve data security by managing deduplication. we also introduced a scheme called Provable Ownership of the File(POF).They enhance user privacy and improve the performance of practical deployment. The random hash code challenge is applied to verify data ownership, which can guarantee that the data holder really have the original data rather than its hash code.

**Advantages:**

* They provide security, effectiveness and efficiency towards potential usage..
* They provide to save the cloud storage across multiple CSP’s and preserve data security in encrypted form.
* They specifies a set of attributes to identify user and encrypts based on it..

**CHAPTER 3**

**REQUIREMENT SPECIFICATIONS**

**3.1 INTRODUCTION**

CLOUD computing allows centralized data storage and online access to computer services or resources. It offers a new way of Information Technology (IT) services by rearranging various resources and providing them to users based on their demands. Cloud computing has greatly pervasive services and become a promising service platform due to a number of desirable properties [40], [41], such as scalability, elasticity, fault-tolerance, and pay-per-use. Dataj8 storage service is one of the most widely consumed cloud services. Cloud users have greatly benefited from cloud storage since they can store huge volume of data without upgrading their devices and access them at any time and in any place. However, cloud data storage offered by Cloud Service Providers (CSPs) still incurs some problems. First of all, various data stored at the cloud may request different ways of protection due to different data sensitivity. The data stored at the cloud include sensitive personal information, publicly shared data, data shared within a group, and so on. Obviously, crucial data should be protected at the cloud to prevent from any access of unauthorized parties. Some unimportant data, however, have no such a requirement. As outsourced data could disclose personal or even sensitive information, data owners sometimes would like to control their data by themselves, while on some occasion, they prefer to delegate their control to a third party since they cannot be always online or have no idea how to perform such a control. How to make cloud data access control adapt to various scenarios and satisfy different user demands becomes a practically important issue. Access control on encrypted data has been widely studied in the literature [10], [11], [12], [13], [14], [15], [16], [17], [33]. However, fewof themcan flexibly support various requirements on cloud data protection in a uniform way, especiallywith economic deduplication management. Second, flexible cloud data deduplication with data access control is still an open issue. Duplicated data could be stored at the cloud [39] in an encrypted form by the same or different users, in the same or different CSPs. From the standpoint of compatibility, it is highly expected that data deduplication can cooperate well with data access control. That is the same data (either encrypted or not) are only stored once at the cloud, but can be accessed by different users based on the policies of data owners or data holders (i.e., the eligible data users who hold original data). Although cloud storage space is huge, duplicated data storage could greatly waste networking resources, consume plenty of power energy, increase operation costs, and make data management complicated. Economic storage will greatly benefit CSPs by decreasing their operation costs and reversely benefit cloud users with reduced service fees. Obviously, cloud data deduplication is particularly significant for big data storage and management. However, the literature still lacks studies on flexible cloud data deduplication across multiple CSPs. Existing work cannot offer a generic solution to support both deduplication and access control in a flexible and uniform way over the cloud

**3.2 HARDWARE AND SOFTWARE SPECIFICATION**

**3.2.1 HARDWARE REQUIREMENTS**

* Hard Disk : 80GB and Above
* RAM : 4GB and Above
* Processor : P IV and Above

**3.2.2 SOFTWARE REQUIREMENTS**

* Windows 7 and above
* JDK 1.8
* J2EE
* Tomcat 7.0
* MySQL

**3.3 TECHNOLOGIES USED**

* JAVA

**3.3.1 JAVA**

Java is an object-oriented programming language developed initially by James Gosling and colleagues at Sun Microsystems. The language, initially called Oak (named after the oak trees outside Gosling's office), was intended to replace C++, although the feature set better resembles that of Objective C.

**3.3.1.1 INTRODUCTION TO JAVA**

Java has been around since 1991, developed by a small team of Sun Microsystems developers in a project originally called the Green project. The intent of the project was to develop a platform-independent software technology that would be used in the consumer electronics industry. The language that the team created was originally called Oak.

The first implementation of Oak was in a PDA-type device called Star Seven (\*7) that consisted of the Oak language, an operating system called GreenOS, a user interface, and hardware. The name \*7 was derived from the telephone sequence that was used in the team's office and that was dialed in order to answer any ringing telephone from any other phone in the office.

Around the time the First Person project was floundering in consumer electronics, a new craze was gaining momentum in America; the craze was called "Web surfing." The World Wide Web, a name applied to the Internet's millions of linked HTML documents was suddenly becoming popular for use by the masses. The reason for this was the introduction of a graphical Web browser called Mosaic, developed by ncSA. The browser simplified Web browsing by combining text and graphics into a single interface to eliminate the need for users to learn many confusing UNIX and DOS commands. Navigating around the Web was much easier using Mosaic.

It has only been since 1994 that Oak technology has been applied to the Web. In 1994, two Sun developers created the first version of Hot Java, and then called Web Runner, which is a graphical browser for the Web that exists today. The browser was coded entirely in the Oak language, by this time called Java. Soon after, the Java compiler was rewritten in the Java language from its original C code, thus proving that Java could be used effectively as an application language. Sun introduced Java in May 1995 at the Sun World 95 convention.

Web surfing has become an enormously popular practice among millions of computer users. Until Java, however, the content of information on the Internet has been a bland series of HTML documents. Web users are hungry for applications that are interactive, that users can execute no matter what hardware or software platform they are using, and that travel across heterogeneous networks and do not spread viruses to their computers. Java can create such applications.

**3.3.1.2 WORKING OF JAVA**

For those who are new to object-oriented programming, the concept of a class will be new to you. Simplistically, a class is the definition for a segment of code that can contain both data (called attributes) and functions (called methods).

When the interpreter executes a class, it looks for a particular method by the name of **main,** which will sound familiar to C programmers. The main method is passed as a parameter an array of strings (similar to the argv [] of C), and is declared as a static method.

To output text from the program, we execute the **println** method of **System.out,** which is java’s output stream. UNIX users will appreciate the theory behind such a stream, as it is actually standard output. For those who are instead used to the Wintel platform, it will write the string passed to it to the user’s program.

Java consists of two things :

* + Programming language
  + Platform

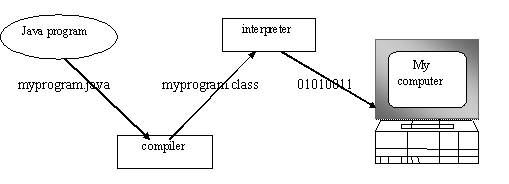
**3.3.1.3 THE JAVA PROGRAMMING LANGUAGE**

Java is a high-level programming language that is all of the following:

* + Simple
  + Object-oriented
  + Distributed
  + Interpreted
  + Robust
  + Secure
  + Architecture-neutral
  + Portable
  + High-performance
  + Multithreaded
  + Dynamic

The code and can bring about changes whenever felt necessary. Some of the standard needed to achieve the above-mentioned objectives are as follows:

Java is unusual in that each Java program is both co implied and interpreted. With a compiler, you translate a Java program into an intermediate language called **Java byte codes** – the platform independent codes interpreted by the Java interpreter. With an interpreter, each Java byte code instruction is parsed and run on the computer. Compilation happens just once; interpretation occurs each time the program is executed. This figure illustrates how it works:

****

**Fig.3.1**

You can think of Java byte codes as the machine code instructions for the **Java Virtual Machine (JVM).** Every Java interpreter, whether it’s a Java development tool or a Web browser that can run Java applets, is an implementation of JVM. That JVM can also be implemented in hardware. Java byte codes help make “write once, run anywhere” possible.

You can compile your Java program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the JVM. For example, that same Java program can e run on Windows NT, Solaris and Macintos

Complier

Interpreter

Interpreter

Interpreter

**  **

**PC-Compatible Sun Ultra Solaris Power macintosh**

**Windows NT System 8**

**3.3.1.4 THE JAVA PLATFORM**

A platform is the hardware or software environment in which a program runs. The Java platform differs from most other platforms in that it’s a software-only platform that runs on top of other, hardware-based platforms. Most other platforms are described as a combination of hardware and operating system.

The Java platform has two components :

* The Java Virtual Machine (JVM)
* The Java Application Programming Interface (Java API)

You’ve already been introduced to the JVM. It’s the base for the Java platform and is ported onto various hardware-based platforms.

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries **(packages)** of related components. The following figure depicts a Java program, such as an application or applet, that’s running on the Java platform. As the figure shows, the Java API and Virtual Machine insulates the Java program from hardware dependencies.

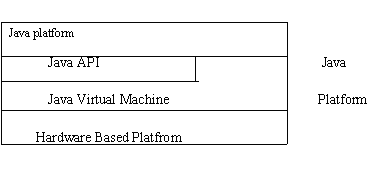
****

Fig.3.3

As a platform-independent environment, Java can be a bit slower than native code. However, smart compliers, weel-tuned interpreters, and just-in-time byte compilers can bring Java’s performance close to that of native code without threatening portability.

**3.3.2 APACHE TOMCAT SERVER**

Apache Tomcat (formerly under the Apache Jakarta Project; Tomcat is now a top level project) is a web container developed at the Apache Software Foundation. Tomcat implements the servlet and the JavaServer Pages (JSP) specifications from Sun Microsystems, providing an environment for Java code to run in cooperation with a web server. It adds tools for configuration and management but can also be configured by editing configuration files that are normally XML-formatted. Because Tomcat includes its own HTTP server internally, it is also considered a standalone web server.

**Environment**  
 Tomcat is a web server that supports servlets and JSPs. Tomcat comes with the Jasper compiler that compiles JSPs into servlets.

The Tomcat servlet engine is often used in combination with an Apache web server or other web servers. Tomcat can also function as an independent web server. Earlier in its development, the perception existed that standalone Tomcat was only suitable for development environments and other environments with minimal requirements for speed and transaction handling. However, that perception no longer exists; Tomcat is increasingly used as a standalone web server in high-traffic, high-availability environments.

Since its developers wrote Tomcat in Java, it runs on any operating system that has a JVM.

**Product features**

Tomcat 3.x (initial release)

* implements the Servlet 2.2 and JSP 1.1 specifications
* servlet reloading
* basic HTTP functionality Tomcat 4.x
* implements the Servlet 2.3 and JSP 1.2 specifications
* servlet container redesigned as Catalina
* JSP engine redesigned as Jasper
* Coyote connector
* Java Management Extensions (JMX), JSP and Struts-based administration
* Tomcat 5.x
* implements the Servlet 2.4 and JSP 2.0 specifications
* reduced garbage collection, improved performance and scalability
* native Windows and Unix wrappers for platform integration
* faster JSP paring

**History** Tomcat started off as a servlet specification implementation by James Duncan Davidson, a software architect at Sun. He later helped make the project open source and played a key role in its donation by Sun to the Apache Software Foundation.

Davidson had initially hoped that the project would become open-sourced and, since most open-source projects had O'Reilly books associated with them featuring an animal on the cover, he wanted to name the project after an animal. He came up with Tomcat since he reasoned the animal represented something that could take care of and fend for itself. His wish to see an animal cover eventually came true when O'Reilly published their Tomcat book with a tomcat on the cover.

**CHAPTER 4**

**4.1 Project Purpose and Scope**

**4.1 Purpose**

The main aim of this project is to control file duplication in cloud computing.

**4.2 Project Scope**

In this work, we proposed to storage across multiple CSP’s and preserve data security by managing deduplication. we also introduced a scheme called Provable Ownership of the File(POF).They enhance user privacy and improve the performance of practical deployment. The random hash code challenge is applied to verify data ownership, which can guarantee that the data holder really have the original data rather than its hash code.

* 1. **4.3 Product Perspective**

Cloud storage as one of the most important services of cloud computing. Data ownership proof is an essential process of data deduplication, especially for encrypted data. But this scheme does not provide flexible deduplication control across multiple Cloud Service Providers (CSPs). In this paper, They propose a multiple cloud service provider (CSPs) in which the data owner will upload the file and the hash MD5 algorithm is used to check data duplication during data storage at the cloud. CSPs. It can achieve data deduplication and access control with different security requirements. And also they have proposed a scheme called Provable Ownership of the File(POF). The result it is security, effectiveness and efficiency towards data storage management.

* 1. **4.4 SystemFeatures**

We motivate to save cloud storage across multiple CSPs and preserve data security and privacy by managing encrypted data storage with deduplication in various situations. We propose a heterogeneousdata management scheme to support both deduplication and access control according to the demands of data owners, which can adapt to different application scenarios. Our scheme can support data sharing among eligible users in a flexible way, which can be controlled by either the data owners or other trusted parties or both of them. We justify the performance of the proposed scheme through security analysis, comparison with existing work and implementation based performance evaluation. The results show its security, advantages, efficiency and potential applicability.

**4.5 Design and Implementation Constraints**

**4.5.1 Constraints in Analysis**

* Constraints as Informal Text
* Constraints as Operational Restrictions
* Constraints Integrated in Existing Model Concepts
* Constraints as a Separate Concept
* Constraints Implied by the Model Structure

**4.5.2 Constraints in Design**

* Determination of the Involved Classes
* Determination of the Involved Objects
* Determination of the Involved Actions
* Determination of the Require Clauses
* Global actions and Constraint Realization

**4.5.3 Constraints in Implementation**

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modeling and many object oriented methods.

* 1. **Other Nonfunctional Requirements**

**4.6.1 Performance Requirements**

The application at this side controls and communicates with the following three main general components.

* embedded browser in charge of the navigation and accessing to the web service;
* Server Tier: The server side contains the main parts of the functionality of the proposed architecture. The components at this tier are the following.

Web Server, Security Module, Server-Side Capturing Engine, Preprocessing Engine, Database System, Verification Engine, Output Module.

**4.6.2 Safety Requirements**

* 1. The software may be safety-critical. If so, there are issues associated with its integrity level
  2. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions.
  3. If a system must be of a high integrity level and if the software is shown to be of that integrity level, then the hardware must be at least of the same integrity level.
  4. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable.
  5. If a computer system is to run software of a high integrity level then that system should not at the same time accommodate software of a lower integrity level.
  6. Systems with different requirements for safety levels must be separated.
  7. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

**CHAPTER 5**

**5.1 Architecture Diagram:**

Data Storage

Cloud Storage

Service

Deduplication Check (CSV)

Login

Registration

File Upload

Reference

Access polices

Download Request

Same File

Block with signature

Cipher Text

POF-CSV

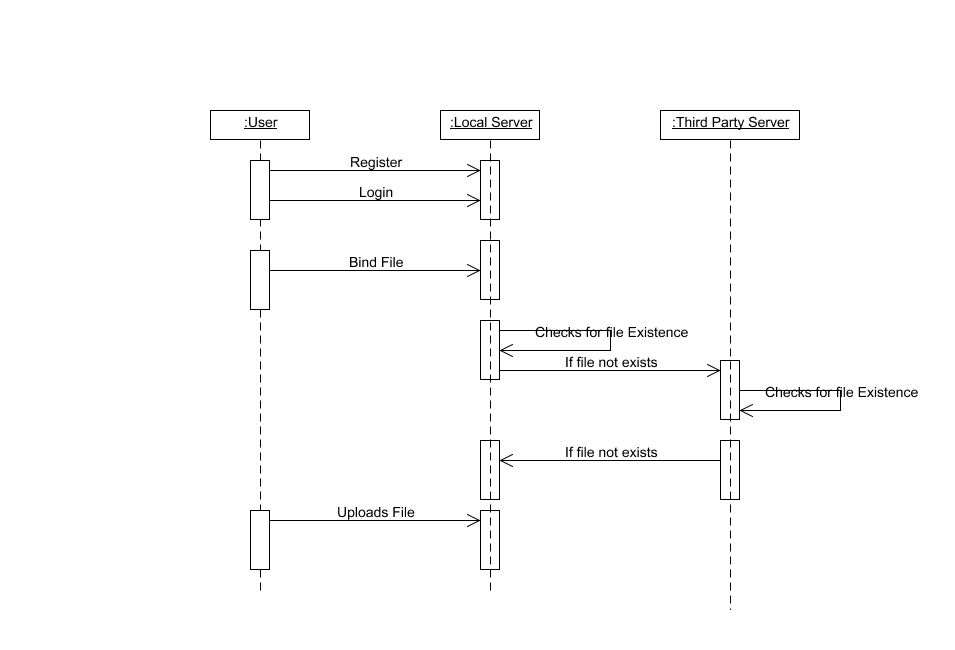
Security Question

Verification

**Fig: 5.1**

**5.2 Sequence Diagram:**

A Sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of Message Sequence diagrams are sometimes called event diagrams, event sceneries and timing diagram.

****

**5.3 Use Case Diagram:**

Unified Modeling Language (UML) is a standardized general-purpose modeling language in the field of software engineering. The standard is managed and was created by the Object Management Group. UML includes a set of graphic notation techniques to create visual models of software intensive systems. This language is used to specify, visualize, modify, construct and document the artifacts of an object oriented software intensive system under development.

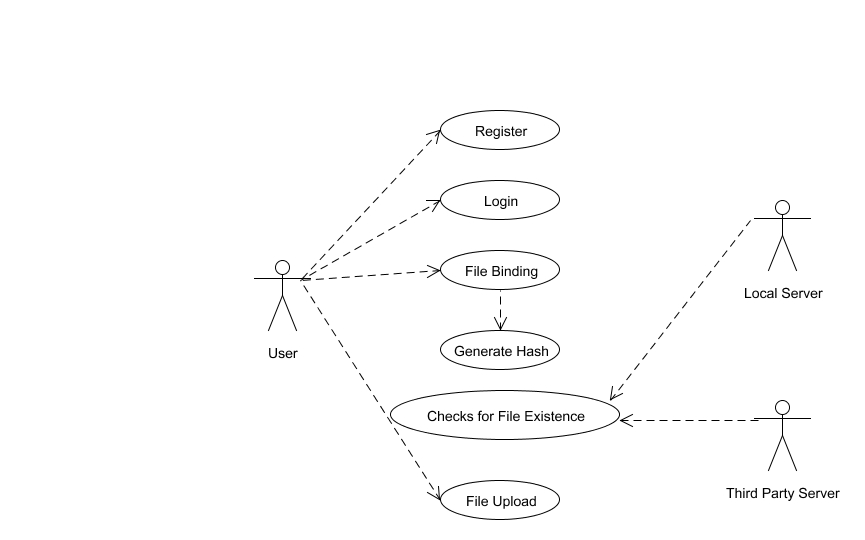
**5.3.1. USECASE DIAGRAM**

A Use case Diagram is used to present a graphical overview of the functionality provided by a system in terms of actors, their goals and any dependencies between those use cases.

Use case diagram consists of two parts:

**Use case:** A use case describes a sequence of actions that provided something of measurable value to an actor and is drawn as a horizontal ellipse.

**Actor:** An actor is a person, organization or external system that plays a role in one or more interaction with the system.

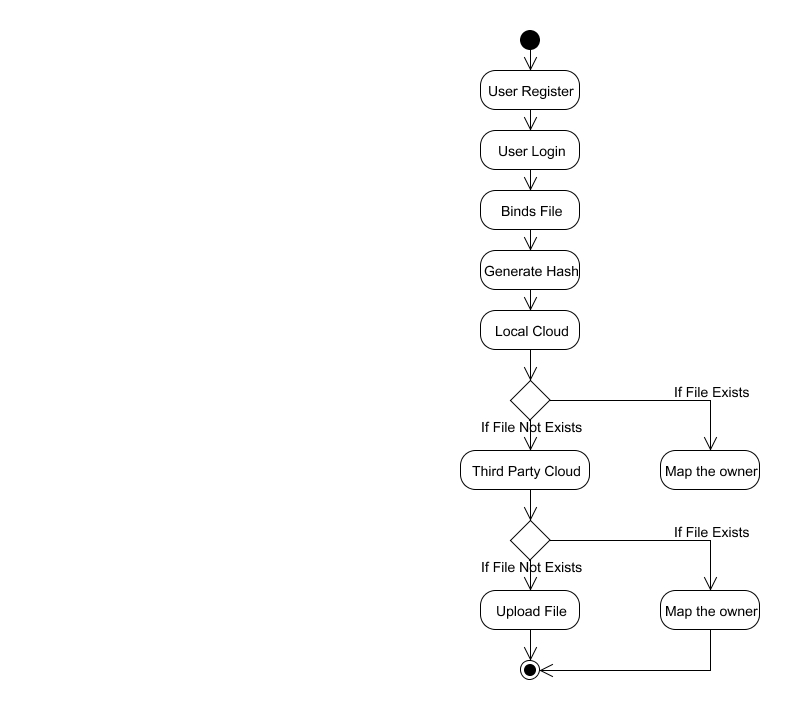
****

**5.4 Activity Diagram:**

Activity diagram is a graphical representation of workflows of stepwise activities and actions with support for choice, iteration and concurrency. An activity diagram shows the overall flow of control.

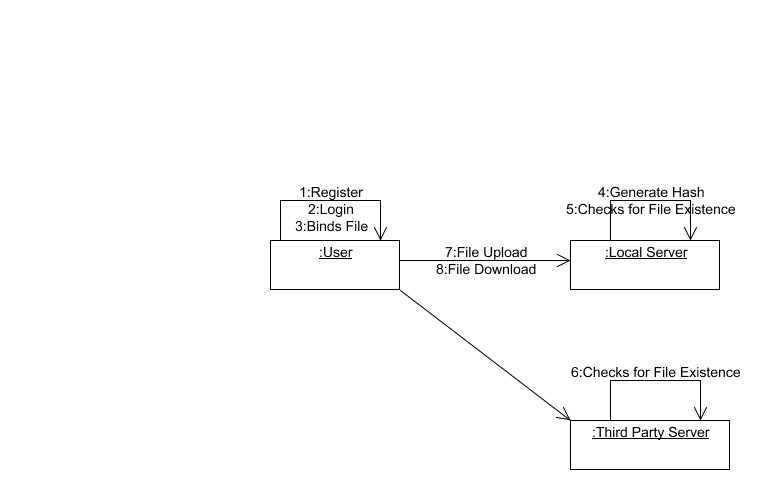
The most important shape types:

* Rounded rectangles represent activities.
* Diamonds represent decisions.
* Bars represent the start or end of concurrent activities.
* A black circle represents the start of the workflow.
* An encircled circle represents the end of the workflow.

****

**5.5 Collaboration Diagram:**

UML Collaboration Diagrams illustrate the relationship and interaction between software objects. They require use cases, system operation contracts and domain model to already exist. The collaboration diagram illustrates messages being sent between classes and objects.

****

**5.6 DATA FLOW DIAGRAM:**

A Data Flow Diagram (DFD) is a graphical representation of the “flow” of data through an information system, modeling its aspects. It is a preliminary step used to create an overview of the system which can later be elaborated DFDs can also be used for visualization of data processing.

**Level 0:**

User

Register

Admin

Verify User

**Level 1:**

Approved User

Login

Binds File

Generate Hash

**Level 2:**

Local Server

Third Party Server

User

File Binding

Generate Hash

Checks for File Existence

**Level 3:**

User

Binds File

If File already exists

Third Party Server

Local Server

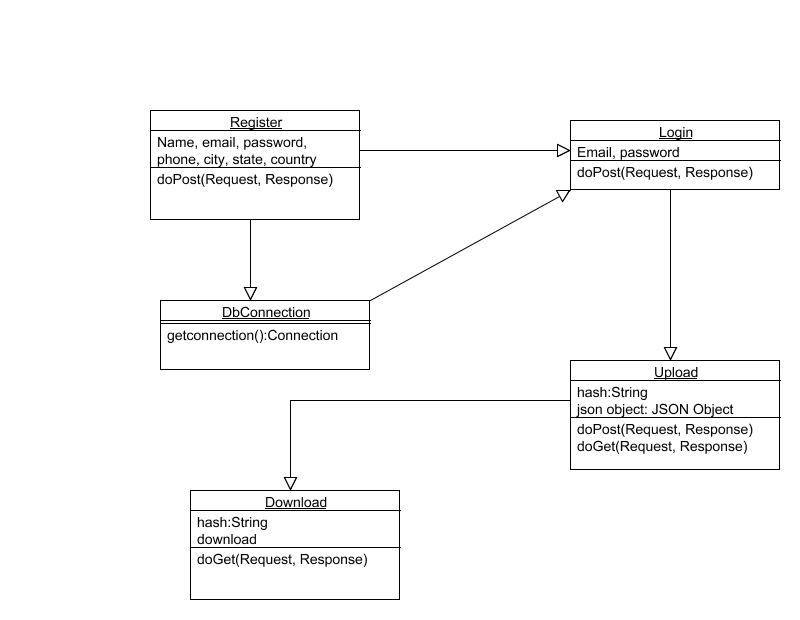
Map the owner

If File not exists

Upload the File

**5.7 Class Diagram**

A Class diagram in the Unified Modeling Language is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among objects.

****

**CHAPTER 6**

**SYSTEM DESIGN**

Owner has an initial level Registration Process in Cloud Service Provider(CSP). The users provide their own personal information for this process. The server in turn stores the information in its database. Then they have the Login process for the further access in cloud service Provider. The data Owner create their account under the Public cloud and upload the file in cloud storage. Here the Provable Ownership of the file (POF) scheme is proposed. While uploading the file by data owner, the hash key is generated based on MD5 algorithm. The hash key is unique for all the upload files. But if the same file is upload by the other data Owner it will not allow the file to upload rather then it will replace the reference id through Mapping of index. It also checks the file for physical present or not by both the data Owner. User will chooses the file and uploads to Storage where the HDFS storage system .In the system will generate a signature in particular file and then split into multiple block. Each block will be generating signature with key. In the signature by using MD5 message-digest algorithm is cryptographic hash function producing a 128-bit hash value typically expressed in text format as 32 digit hex value so that files of same are de-duplicated. After that generate convergent keys for each blocks splitting to store CSV file .like filename, file path, blocks, username, password and block keys. The data owner will download the file from cloud service provider. If they do not find the file then they will request to download the file from different Cloud service provider and also check whether the file is present or not then it gives the response to data owner.

**6.1 MODULES**

* + - **Cloud User Authentication.**
    - **File Upload and Comparison**
    - **Set Access Policy for File**
    - **File Download Request and Handling.**

**6.2 MODULE EXPLANATION:**

**1.Cloud User Authentication**

Owner has an initial level Registration Process in Cloud Service Provider(CSP). The users provide their own personal information for this process. The server in turn stores the information in its database. Then they have the Login process for the further access in cloud service Provider.

2. **File Upload and Comparison**

In this module, the data Owner create their account under the Public cloud and upload the file in cloud storage. Here the Provable Ownership of the file(POF) scheme is proposed. While uploading the file by data owner, the hash key is generated based on MD5 algorithm. The hash key is unique for all the upload files. But if the same file is upload by the other data Owner it will not allow the file to upload rather then it will replace the reference id through Mapping of index. It also check the file for physical present or not by both the data Owner.

**3. Set Access Policy for File**

In this module User will chooses the file and uploads to Storage where the HDFS storage system .In the system will generate a signature in particular file and then split into multiple block. Each block will be generate signature with key . In the signature by using MD5 message-digest algorithm is cryptographic hash function producing a 128-bit hash value typically expressed in text format as 32 digit hex value so that files of same are de-duplicated. After that generate convergent keys for each blocks splitting to store CSV file .like filename, file path, blocks, username, password and block keys.

**4. File Download Request and Handling**

In this module ,the data owner will download the file from cloud service provider. If they do not find the file then they will request to download the file from different Cloud service provider and also check whether the file is present or not then it gives the response to data owner.

**CHAPTER 7**

**CODING AND TESTING**

**7.1 CODING**

Once the design aspect of the system is finalizes the system enters into the coding and testing phase. The coding phase brings the actual system into action by converting the design of the system into the code in a given programming language. Therefore, a good coding style has to be taken whenever changes are required it easily screwed into the system.

**7.2 CODING STANDARDS**

Coding standards are guidelines to programming that focuses on the physical structure and appearance of the program. They make the code easier to read, understand and maintain. This phase of the system actually implements the blueprint developed during the design phase. The coding specification should be in such a way that any programmer must be able to understand the code and can bring about changes whenever felt necessary. Some of the standard needed to achieve the above-mentioned objectives are as follows:

Program should be simple, clear and easy to understand.

Naming conventions

Value conventions

Script and comment procedure

Message box format

Exception and error handling

**7.2.1 NAMING CONVENTIONS**

Naming conventions of classes, data member, member functions, procedures etc., should be **self-descriptive**. One should even get the meaning and scope of the variable by its name. The conventions are adopted for **easy understanding** of the intended message by the user. So it is customary to follow the conventions. These conventions are as follows:

**Class names**

Class names are problem domain equivalence and begin with capital letter and have mixed cases.

**Member Function and Data Member name**

Member function and data member name begins with a lowercase letter with each subsequent letters of the new words in uppercase and the rest of letters in lowercase.

7**.2.2 VALUE CONVENTIONS**

Value conventions ensure values for variable at any point of time. This involves the following:

* Proper default values for the variables.
* Proper validation of values in the field.
* Proper documentation of flag values.

**7.2.3 SCRIPT WRITING AND COMMENTING STANDARD**

Script writing is an art in which indentation is utmost important. Conditional and looping statements are to be properly aligned to facilitate easy understanding. Comments are included to minimize the number of surprises that could occur when going through the code.

**7.2.4 MESSAGE BOX FORMAT**

When something has to be prompted to the user, he must be able to understand it properly. To achieve this, a specific format has been adopted in displaying messages to the user. They are as follows:

* X – User has performed illegal operation.
* ! – Information to the user.

**7.3 TEST PROCEDURE**

SYSTEM TESTING

Testing is performed to identify errors. It is used for quality assurance. Testing is an integral part of the entire development and maintenance process. The goal of the testing during phase is to verify that the specification has been accurately and completely incorporated into the design, as well as to ensure the correctness of the design itself. For example the design must not have any logic faults in the design is detected before coding commences, otherwise the cost of fixing the faults will be considerably higher as reflected. Detection of design faults can be achieved by means of inspection as well as walkthrough.

Testing is one of the important steps in the software development phase. Testing checks for the errors, as a whole of the project testing involves the following test cases:

* Static analysis is used to investigate the structural properties of the Source code.
* Dynamic testing is used to investigate the behavior of the source code by executing the program on the test data.

**7.4 TEST DATA AND OUTPUT**

**7.4.1 UNIT TESTING**

Unit testing is conducted to verify the functional performance of each modular component of the software. Unit testing focuses on the smallest unit of the software design (i.e.), the module. The white-box testing techniques were heavily employed for unit testing.

**7.4.2 FUNCTIONAL TESTS**

Functional test cases involved exercising the code with nominal input values for which the expected results are known, as well as boundary values and special values, such as logically related inputs, files of identical elements, and empty files.

Three types of tests in Functional test:

* Performance Test
* Stress Test
* Structure Test

**7.4.3 PERFORMANCE TEST**

It determines the amount of execution time spent in various parts of the unit, program throughput, and response time and device utilization by the program unit.

**7.4.4 STRESS TEST**

Stress Test is those test designed to intentionally break the unit. A Great deal can be learned about the strength and limitations of a program by examining the manner in which a programmer in which a program unit breaks.

**7.4.5 STRUCTURED TEST**

Structure Tests are concerned with exercising the internal logic of a program and traversing particular execution paths. The way in which White-Box test strategy was employed to ensure that the test cases could Guarantee that all independent paths within a module have been have been exercised at least once.

* Exercise all logical decisions on their true or false sides.
* Execute all loops at their boundaries and within their operational bounds.
* Exercise internal data structures to assure their validity.
* Checking attributes for their correctness.
* Handling end of file condition, I/O errors, buffer problems and textual errors in output information

**7.4.6 INTEGRATION TESTING**

Integration testing is a systematic technique for construction the program structure while at the same time conducting tests to uncover errors associated with interfacing. i.e., integration testing is the complete testing of the set of modules which makes up the product. The objective is to take untested modules and build a program structure tester should identify critical modules. Critical modules should be tested as early as possible. One approach is to wait until all the units have passed testing, and then combine them and then tested. This approach is evolved from unstructured testing of small programs. Another strategy is to construct the product in increments of tested units. A small set of modules are integrated together and tested, to which another module is added and tested in combination. And so on. The advantages of this approach are that, interface dispenses can be easily found and corrected.

The major error that was faced during the project is linking error. When all the modules are combined the link is not set properly with all support files. Then we checked out for interconnection and the links. Errors are localized to the new module and its intercommunications. The product development can be staged, and modules integrated in as they complete unit testing. Testing is completed when the last module is integrated and tested.

**7.5 TESTING TECHNIQUES / TESTING STRATERGIES**

**7.5.1 TESTING**

Testing is a process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an as-yet –undiscovered error. A successful test is one that uncovers an as-yet- undiscovered error. System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently as expected before live operation commences. It verifies that the whole set of programs hang together. System testing requires a test consists of several key activities and steps for run program, string, system and is important in adopting a successful new system. This is the last chance to detect and correct errors before the system is installed for user acceptance testing.

The software testing process commences once the program is created and the documentation and related data structures are designed. Software testing is essential for correcting errors. Otherwise the program or the project is not said to be complete. Software testing is the critical element of software quality assurance and represents the ultimate the review of specification design and coding. Testing is the process of executing the program with the intent of finding the error. A good test case design is one that as a probability of finding an yet undiscovered error. A successful test is one that uncovers an yet undiscovered error. Any engineering product can be tested in one of the two ways:

**7.5.1.1 WHITE BOX TESTING**

This testing is also called as Glass box testing. In this testing, by knowing the specific functions that a product has been design to perform test can be conducted that demonstrate each function is fully operational at the same time searching for errors in each function. It is a test case design method that uses the control structure of the procedural design to derive test cases. Basis path testing is a white box testing.

Basis path testing:

* Flow graph notation
* Cyclometric complexity
* Deriving test cases
* Graph matrices Control

**7.5.1.2 BLACK BOX TESTING**

In this testing by knowing the internal operation of a product, test can be conducted to ensure that “all gears mesh”, that is the internal operation performs according to specification and all internal components have been adequately exercised. It fundamentally focuses on the functional requirements of the software.

The steps involved in black box test case design are:

* Graph based testing methods
* Equivalence partitioning
* Boundary value analysis
* Comparison testing

**7.5.2 SOFTWARE TESTING STRATEGIES:**

A software testing strategy provides a road map for the software developer. Testing is a set activity that can be planned in advance and conducted systematically. For this reason a template for software testing a set of steps into which we can place specific test case design methods should be strategy should have the following characteristics:

* Testing begins at the module level and works “outward” toward the integration of the entire computer based system.
* Different testing techniques are appropriate at different points in time.
* The developer of the software and an independent test group conducts testing.
* Testing and Debugging are different activities but debugging must be accommodated in any testing strategy.

**7.5.2.1 INTEGRATION TESTING:**

Integration testing is a systematic technique for constructing the program structure while at the same time conducting tests to uncover errors associated with. Individual modules, which are highly prone to interface errors, should not be assumed to work instantly when we put them together. The problem of course, is “putting them together”- interfacing. There may be the chances of data lost across on another’s sub functions, when combined may not produce the desired major function; individually acceptable impression may be magnified to unacceptable levels; global data structures can present problems.

**7.5.2.2 PROGRAM TESTING:**

The logical and syntax errors have been pointed out by program testing. A syntax error is an error in a program statement that in violates one or more rules of the language in which it is written. An improperly defined field dimension or omitted keywords are common syntax error. These errors are shown through error messages generated by the computer. A logic error on the other hand deals with the incorrect data fields, out-off-range items and invalid combinations. Since the compiler s will not deduct logical error, the programmer must examine the output. Condition testing exercises the logical conditions contained in a module. The possible types of elements in a condition include a Boolean operator, Boolean variable, a pair of Boolean parentheses A relational operator or on arithmetic expression. Condition testing method focuses on testing each condition in the program the purpose of condition test is to deduct not only errors in the condition of a program but also other a errors in the program.

**7.5.2.3 SECURITY TESTING:**

Security testing attempts to verify the protection mechanisms built in to a system well, in fact, protect it from improper penetration. The system security must be tested for invulnerability from frontal attack must also be tested for invulnerability from rear attack. During security, the tester places the role of individual who desires to penetrate system.

**7.5.2.4 VALIDATION TESTING**

At the culmination of integration testing, software is completely assembled as a package. Interfacing errors have been uncovered and corrected and a final series of software test-validation testing begins. Validation testing can be defined in many ways, but a simple definition is that validation succeeds when the software functions in manner that is reasonably expected by the customer. Software validation is achieved through a series of black box tests that demonstrate conformity with requirement. After validation test has been conducted, one of two conditions exists.

\* The function or performance characteristics confirm to specifications and are accepted.

\* A validation from specification is uncovered and a deficiency created.

Deviation or errors discovered at this step in this project is corrected prior to completion of the project with the help of the user by negotiating to establish a method for resolving deficiencies. Thus the proposed system under consideration has been tested by using validation testing and found to be working satisfactorily. Though there were deficiencies in the system they were not catastrophic

**7.5.2.5 USER ACCEPTANCE TESTING**

User acceptance of the system is key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with prospective system and user at the time of developing and making changes whenever required. This is done in regarding to the following points.

* Input screen design.
* Output screen design.

**Conclusion**

Thus we achieve data de-duplication and access control with different security requirements. Security analysis with secure ,efficient and advanced has performed.

**Future enhance**

In addition,we will conduct game theoretical analysis to further prove the rationality and security of the proposed scheme

**Source Code**

**Screenshots:**

**REFERENCES**

[1] R. Chow, et al., “Controlling data in the cloud: outsourcing computation without outsourcing control,” in Proc. ACM Workshop Cloud Comput. Secur., 2009, pp. 85–90.

[2] S. Kamara, and K. Lauter, “Cryptographic cloud storage,” Financ. Crypto. Data Secur., 2010, pp. 136–149.

[3] Q. Liu, C. C. Tan, J. Wu, and G. Wang, “Efficient information retrieval for ranked queries in cost-effective cloud environments,” in Proc. IEEE INFOCOM, 2012, pp. 2581–2585.

[4] M. Kallahalla, E. Riedel, R. Swaminathan, Q. Wang, and K. Fu, “Plutus: scalable secure file sharing on untrusted storage,” in Proc. USENIX Conf. File Storage Technol., 2003, pp. 29–42.

[5] E.-J. Goh, H. Shacham, N. Modadugu, and D. Boneh, “SiRiUS: securing remote untrusted storage,” in Proc. Netw. Distrib. Syst. Secur. Symp., 2003, pp. 131–145.

[6] J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute- based encryption,” in Proc. IEEE Symp. Secur. Privacy, 2007, pp. 321–334.

[7] V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-based encryption for fine-grained access control of encrypted data,” in Proc. 13th ACM Comput. Commun. Secur., 2006, pp. 89–98.

[8] S. Muller, S. Katzenbeisser, and C. Eckert, “Distributed attributebased encryption,” in Proc. 11th Annu. Int. Conf. Inf. Secur. Crypto., 2008, pp. 20–36.

[9] A. Sahai, and B.Waters, “Fuzzy identity-based encryption,” in Proc. 24th Int. Conf. Theory App. Cryptographic Tech., 2005, pp. 457–473.

[10] S. C. Yu, C. Wang, K. Ren, and W. J. Lou, “Achieving secure, scalable, and fine-grained data access control in cloud computing,” in Proc. IEEE INFOCOM, 2010, pp. 534–542.