International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 14
October2016

Secure Distributed Deduplication Systems with Improved Reliability

Pagidimarri Sushma
Department Of Cse
Madhira Institute Of Technology And Sciences

G.L. Chandrashekar Rao
Assistant Professor
DEPARTMENT OF CSE
MADHIRA INSTITUTE OF TECHNOLOGY AND SCIENCES

ABSTRACT: *Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce the storage space and upload the bandwidth. Whenever There is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. As a result, de duplication system improves and increase storage with a utilization while reducing reliability. Furthermore, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud storage and Aiming to address are given the security challenges, this paper makes the first attempt to formalize the conception of distributed reliable de duplication system. We propose new distributed de duplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers.*
The security requirements of the data confidentiality and tag consistency are also achieved by introducing a deterministic with secret sharing scheme in distributed storage those systems, instead of using convergent encryption as in previous de duplication systems. The Security analysis is a demonstrates that our de duplication systems are secure and private in terms of the definitions specified in the proposed security model. As a proofof concept, we implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments

INTRODUCTION.

With the explosive growth of digital data, deduplication techniques are widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication has received much attention from both academia and industry because it can greatly improves storage utilization and save storage space, especially for the applications with high deduplication ratio such as archival storage systems. A number of deduplication systems have been proposedbased on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications. A brief review is given in Section 6. Especially, with the advent of cloud storage, data deduplication techniques become more attractive and critical for the management of ever-increasing volumes of data in cloud storage services which motivates enterprises and organizations to outsource data storage to third-party cloud providers, as evidenced by many real-life case studies [1]. According to the analysis report of IDC, the volume of data in the world is expected to reach 40 trillion gigabytes in 2020 [2]. Today's commercial cloud storage services, such as Drop box, Google Drive and Mozy, have been applying deduplication to save the network bandwidth and the storage cost with client-side deduplication. There are two types of deduplication in terms of the size: (i) *file-level deduplication*, which discovers redundancies between different files and removes these redundancies to reduce capacity demands, and (ii) *blocklevel deduplication*, which discovers and removes redundancies between data blocks. The file can be divided into smaller fixed-size or variable-size blocks. Using fixedsize blocks simplifies the computations of block boundaries, while using variable-size blocks (e.g., based on Rabin

fingerprinting [3]) provides better deduplication efficiency.

EXISTING SYSTEM:

1. A number of de duplication systems have been proposed based on various de duplication strategies with ideas such as server-side or client-side de duplications, file-level or block-level de duplications.
2. Bellare et al. formalized this primitive as message-locked encryption, and explored its application in space efficient secure outsourced storage. There are also several implementations of convergent implementations of different convergent encryption variants for secure de duplication.
3. Li addressed the key-management issue in block-level de duplication by distributing these keys are across the multiple servers using after encrypting the files.
4. Showed how to protect the data confidentiality by the transforming and predictable message into a unpredictable messages.

DISADVANTAGES OF EXISTING SYSTEM:

1. Data reliability is actually a very critical issue in a de duplication storage system because there is only one copy for each and file stored in to the server and shared by all the owners.
2. Most of the previous de duplication systems have only been considered in a single-server setting.
3. The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems.

EXTENSION

As the extension of this project, we go for cryptography for our application. So that the data that is there in normal human understanding form can be converted to cipher text.

For converting plane text to cipher, we do use a secrete key. Then will generate each file having separate secrete key can be used for converting the cipher text back to plain text. In this way, we can make our data more secure.

IMPLEMENTATION

MODULES:

- System Model
- Data Deduplication
- File level Deduplication Systems
- Block level Deduplication systems

SOFTWARE ENVIRONMENT:

Software Environment Java Technology: Java is a programming and platform independent language. The Java is Programming Language: This is a programming language is a high-level language as well as it can be characterized by all of the following Keywords:

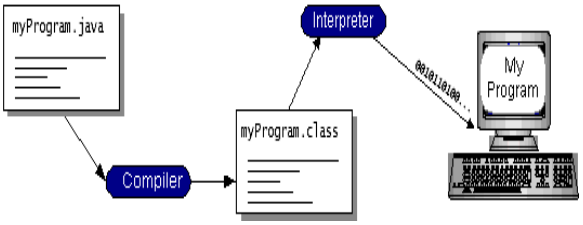
- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

The most programming language is java, we can

compile or run a programs so that we can compile it once run anywhere.

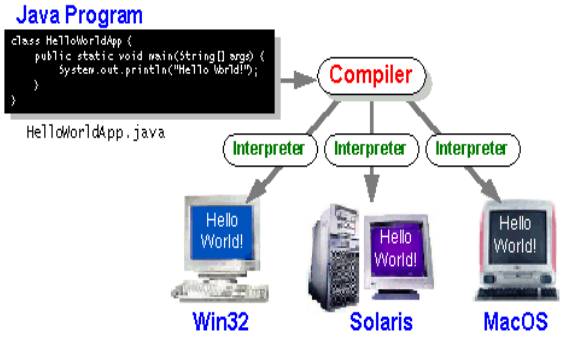
The Java is a programming language is done in that a program is both compiled and interpreted. Then compiler, first we can translates a program into an intermediate languages called Java is converted byte codes to independent codes interpreted by the interpreter on the Java. The interpreter parses and runs each Java is converted to code instruction on the computer. Compilation or run happens just once.

Interpretation every and each time the program is executed by java compiler. The following bellow figure we can show how it works.



We can think of Java is converted to codes or programs as the machine language code instructions by the Java Virtual Machine (Java VM). Every Java interpreter, whether it's a development tool or a Web browser that can run applets, it is an implementation of the Java VM.

Java is converted to codes are help makes "write once, we run anywhere" possible. We can compile we are programs in to byte codes on any platform that has to be a Java compiler. The byte codes or programming it can then be run on any implementation of the Java VM. That means that is as long as a computer has to be a Java VM, the same codes written in the Java is a programming language can run on Windows 2000 or above versions, the Solaris workstation or on an iMac.



Exceptions:

Exception is an object which is created by JVM when it encounters a **logical error** in the programming.

We have two type of errors in the programming

1) Syntax Errors

2) Logical Errors

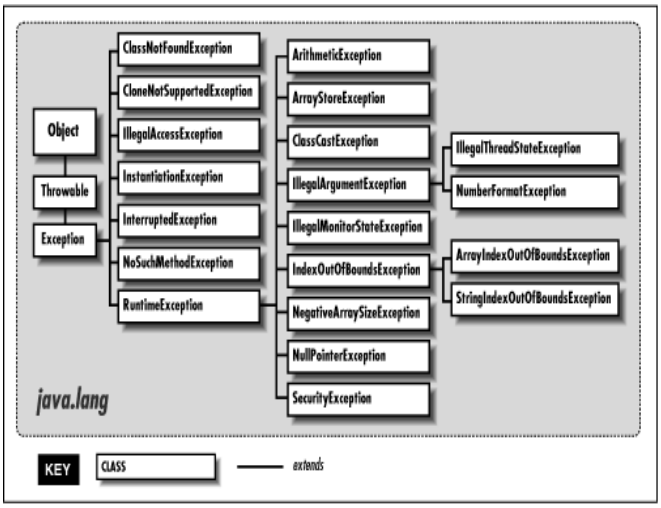
Advantages Of Exception Handling

→We can know in which part of the program the logical error has actually occurred.

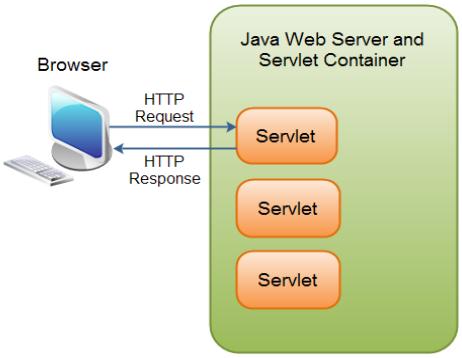
→We are not allowing the JVM to terminate the entire program. Only try block in which the error has occurred would be terminated.

→Thus we can debug the program very easily.

Hierarchy of Exceptions:



Servlets inside a Java Servlet Container:



Local host, the unique address of the server, the most of the times it's the hostname of the server that maps to unique IP address. Sometimes we can multiple hostnames point to same IP addresses and web server side virtual host of the takes care of sending a request to the particular server instance.
The Port numbers is 0 to 1023 are reserved ports for well known services, for example 80 for HTTP, 443 for HTTPS, 21 for FTP etc. First Servlet Project/jsps/hello.jsp – Resource requested from server. It can be static html, pdf, JSP, servlets, PHP etc.

Why we need Servlet and JSPs?

Web servers are good for static contents HTML pages but they don't know how to generate by the dynamic content or how to save data into databases, so we need another tool that we can use to generate dynamic content.

There are several programming languages for dynamic content like PHP and Python, Ruby on Rails, the Java Servlets and JSPs. Java Servlet and JSPs are server side programming technologies to extend to the capability of web servers by providing to the support for dynamic send the response and data persistence. Web Container Tomcat is a web container, when a request is made of from Client to web server side, it passes to the request from the web container and it's web container job to find the correct to the resource to handle the request (servlet or JSP) and then use the response from the resource to generate the response and provide it to web server.

Then web server sends the like response to back to the client side. When web container to gets the request send and if it's for servlet then container creates two Objects HTTP Servlet Request and HTTP Servlet Response. Then it finds the correct servlet is based on the URL and creates a thread for the request.

Then it invokes the servlet service() method and based on the HTTP method service() method invokes doGet() or doPost() methods. Servlet methods generate the dynamic page and write it to response. Once servlet thread is complete, container converts the response to HTTP response and sends it back to client.
Every JSP in the an application is a compiled by the container and converted to Servlet and then it's a container manages them like other servlets. Miscellaneous Task data Web container like a manages the resource pool,it's does memory optimizations and run garbage collector, it's provides a security configurations, this is support for multiple applications, it's hot deployment and several other tasks behind the scene that makes our life easier.

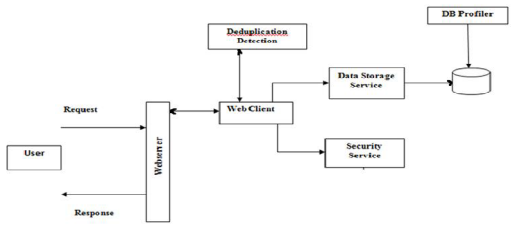
Web Application Directory Structure Java Web Applications are packaged as Web Archive (WAR) and it has a defined structure. We can export above the dynamic web project as a WAR file and unzip it to check the hierarchy. It will be something like below image. / Deployment Descriptor web.xml file is the deployment descriptor of the web application and contains mapping for servlets (prior to 3.0), welcome pages, security configurations, session timeout settings etc.
The SQL is a standard has been evolving the since 1986 and several versions exist. In this manual, the SQL-92 refers to the standard released in 1992, the SQL:1999 it's refers to be the standard released in 1999, and "SQL:2003" refers to the current version of the standard. We can use the phrase like a "the SQL standard" to mean the current version of the SQL is Standard at any time. MySQL software is Open Source.

Open Source means that it is possible for anyone to use and modify the software. Anybody can be download to the MySQL software from the Internet

and use it without paying anything. If we wish, we may study the source code and change it to suit wer needs. The MySQL is software uses the GPL (GNU General Public License), http://www.fsf.org/licenses/, to define what we may and may not do with the software in different situations.

SYSTEM DESIGN

SYSTEM ARCHITECTURE:



SYSTEM TESTING:

The purpose of testing is to during a search errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub gathered , gathered and/or a finished product It is the process of exercising software with the intent of make sure that the

Software system meets its requirements and user expectations and does not fail in an unacceptable a way. There are different types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing:

Unit testing have the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : establish classes of valid input must be accepted.

Invalid Input : establish classes of invalid input must be rejected.

Functions: establish functions must be exercised.

Output : establish classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, plan coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

Acceptance Testing:

User Acceptance Testing is a critical phase of any project and requires significant taking by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

CONCLUSION

We proposed the distributed is the deduplication systems to improve and secure the reliability of data while achieving the confidentiality of the users and outsourced data without an encryption data mechanism. The Four constructions are proposed to support fine-grained and file-level block-level data deduplication. The security of the tag consistency and integrity were achieved. We are implemented our deduplication systems using the Ramp secret sharing scheme and demonstrated that it incurs small encoding and decoding overhead compared to the network transmission overhead in regular upload and download operations.

REFERENCES

[1] Amazon, "Case Studies," https://aws.amazon.com/solutions/casestudies/# backup.

[2] J. Gantz and D. Reinsel, "The digital universe in 2020: Big
data, bigger digi tal shadows, and biggest growth in the
far east," http://www.emc.com/collateral/analyst-reports/idcthe-
digital-universe-in-2020.pdf, Dec 2012.

[3] M. O. Rabin, "Fingerprinting by random polynomials," Center
for Research in Computing Technology, Harvard University, Tech.
Rep. Tech. Report TR-CSE-03-01, 1981.

[4] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer,
"Reclaiming space from duplicate files in a serverless distributed
file system." in *ICDCS*, 2002, pp. 617–624.

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided
encryption for deduplicated storage," in *USENIX Security
Symposium*, 2013.

[6] ——, "Message-locked encryption and secure deduplication," in
EUROCRYPT, 2013, pp. 296–312.

[7] G. R. Blakley and C. Meadows, "Security of ramp schemes," in
Advances in Cryptology: Proceedings of CRYPTO '84, ser. Lecture
Notes in Computer Science, G. R. Blakley and D. Chaum, Eds.
Springer-Verlag Berlin/Heidelberg, 1985, vol. 196, pp. 242–268.

[8] A. D. Santis and B. Masucci, "Multiple ramp schemes," *IEEE
Transactions on Information Theory*, vol. 45, no. 5, pp. 1720–1728,
Jul. 1999.

[9] M. O. Rabin, "Efficient dispersal of information for security, load
balancing, and fault tolerance," *Journal of the ACM*, vol. 36, no. 2,
pp. 335–348, Apr. 1989.

[10] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11,
pp. 612–613, 1979.