

Hash Tabled Intelligent Keyword Storage Cloud Document System

Bellam Sai kishan¹, Iruvuri Bhargav Ram², Aravapalli Pavan Kumar³, Mr Chitti babu.K⁴
^{1,2,3,4}*M.E(Ph.d),Department of CSE,R.M.K College of Engineering and Technology*

Abstract- The major task of the cloud storage servers is the information sharing. The recent generations of computing is purely based on searching of documents or contents in huge databases. The cloud storage server allows storing of fragile and large volumes of data with less cost and fast access benefits.the storing of data and retrieving it in an efficient manner becomes a challenging task.in this case the security must be given the due importance and provide confidence to the data owner. Hence the utilization of data through plaintext search should be changed.Therefore an excellent strategy is required to equal the keywords with encrypted cloud data.The proposed concept similarity measure of coordinate matching combined with inner product similarity quantitatively assess and matches all applicable data with search keyword to retrieve the foremost results. In this approach each document is given with a binary vector to replace a keyword that is contained in the document.By the inner product of the query vector ,the similarity could be exactly measured with the data vector.The two multi-keyword ranked search and the inner product computation over the encrypted data ensures data privacy and provides information about the dynamic operation on the data set and index and hence this will improves the search event of the user.

I. INTRODUCTION

The term Cloud refers to a Network or net. In alternative words, we will say that Cloud is some things that are gift at remote location. Cloud will give services over network. Service Models are the reference models on that the Cloud Computing relies. These may be classified into three basic service models as listed below:

Infrastructure as a Service (IaaS)

Platform as a Service (PaaS)

Software as a Service (SaaS)

There are several alternative service models all of which may take the shape like SaaS Anything as a Service. This can be Network as a Service, Business

as a Service, Identity as a Service, information as a Service or Strategy as a Service. The Infrastructure as a Service (IaaS) is the simplest level of service. Every of the service models create use of the underlying service model. The planned approach similarity live of “coordinate matching” combined with “inner product similarity” quantitatively evaluates and matches all relevant information with search keyword to make best results. Then that user can able to upload the same document with changes in that document that document modified words are updated in the individual page.

2. EXISTING SYSTEM

Existing system will have a very simple document based search system. Users can search for the documents with the name of the document. For instance, typhoid.doc can be retrieved by searching for the keyword “typhoid”.In case the word penicillin is available in the document typhoid.doc it wont be available in the search. Ranked keyword search enhances the feature of retrieving the keyword in the documents too But the documents were stored in the file system.Moving the documents and maintaining the documents becomes tougher in this case.Encrypted information were stored securely in cloud and we don't have an option of hash table to access the data faster and effectively.No clear process flow between the document owner and document accessor. In the existing system. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy.Ranked Searchable encryption allows data owner to outsource his data in an encrypted manner while maintaining the selectively search capability over the encrypted data. In the existing system, the documents and data is stored into

a secure cloud storage whereas the documents were scanned with the keywords and an index of information were stored for future searching options. The existing system have used symmetric encryption algorithm to store the data securely. The large number of data users and documents in cloud is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. The searchable encryption focuses on single keyword search or Boolean keyword search, and rarely differentiates the search results. By stop word concept the unwanted keywords will be removed. The document search by name not by content. So we get relevant information and irrelevant information. We are using MD5 algorithm in existing system.

3. PROPOSED SYSTEM

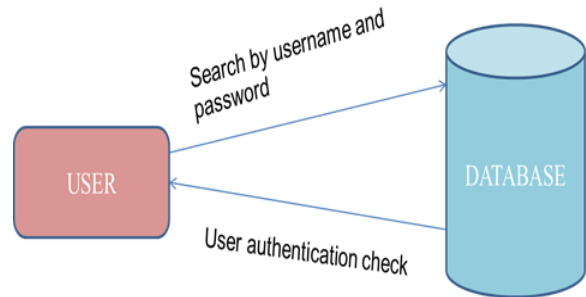
Proposed system emphasizes and address most of drawbacks of the existing system. Below are the items, Storage Efficiency - > Instead of the file system, the data will be stored in the format of BLOB (Binary Large Objects) in the database. This enhances storage efficiency and maintainability too. Access Efficiency -> Data will be stored in the format of keywords by automatic intelligent data retrieval process and the data will be indexed efficiently by hashing technique. So that, the data will be accessed at a very high speed (On a whole a high end intelligent data system is build in cloud) Process Efficiency – A clear process flow of document owner uploading the documents with the indexing keywords. Document viewer request for the keywords and document access. An automatic request will be initiated to the owner and in turn on confirming the document. Keys with RSA based encryption emphasized will be sent to the end user to access the documents. Technology Efficiency – key updation can be achieved via Whatsapp or SMS to the end users. Intelligence Efficiency – Semantic information were stored and the data will be retrieved accordingly in a semantic manner. We define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle

of “coordinate matching”. In proposed system define public or private page and will be stored. Individual page updation is in this system. We are ranking the document (abc.doc) by multi key word concept. Checksum value for each page.

4. MODULE DESCRIPTION

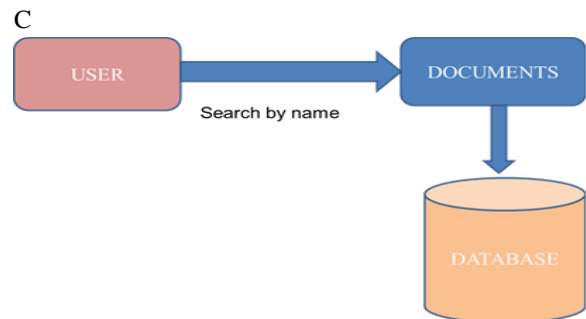
4.1 USER AUTHENTICATION MODULE

User’s information is stored in database to check whether the user is authenticated or unauthenticated user. When the user is authorized person means document is searched. User’s information is searched by using name and password in a database. Both the name and password is matching with the database then only documents are searched in a database. User is unauthorized person, documents not searched and it will produce errors.



4.2 NAME SEARCH MODULE

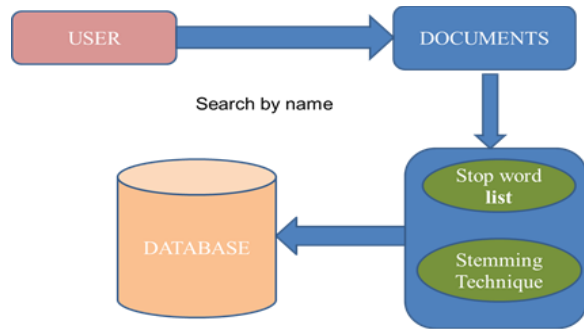
After the verification of user’s information in a Database. User is allowed to search a document Using document name only. Each and every word is sorted and produce output based on name search in a database. Content-wise search is allowed in this Proposed System.



4.3 STOP WORD REMOVAL MODULE

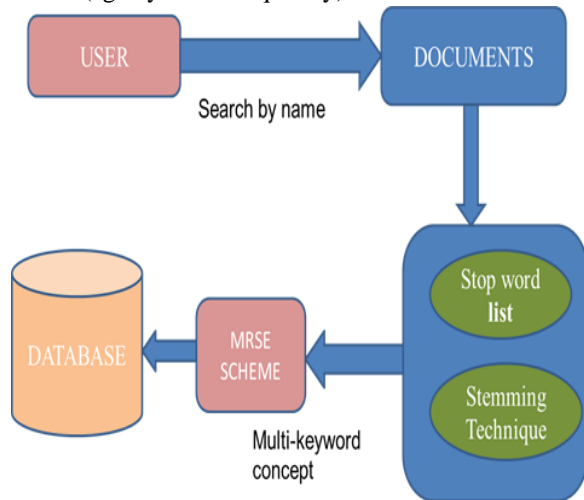
Documents searched in a database by the order of document’s name and its content. Document’s content

is sorted by using stop word removal technique in a database. Stemming technique used to list the words by the removal of stemming words in a document.



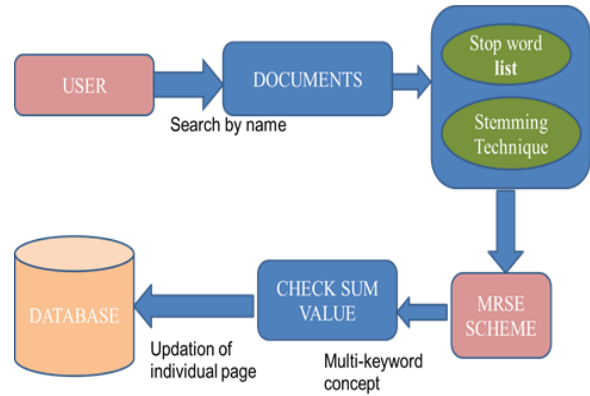
4.4 MULTI-KEYWORD SEARCH MODULE

Users search the document by using multiple keywords in a database. Removal of words in a database finally sorts out some multi-keywords for the document. These Multi-keywords are sorted by means of using priority basis. Multi-keyword ranked search over encrypted (MRSE) data is a technique involved and returning files in a ranked keyword order regarding to certain relevance criteria (eg: keyword frequency).



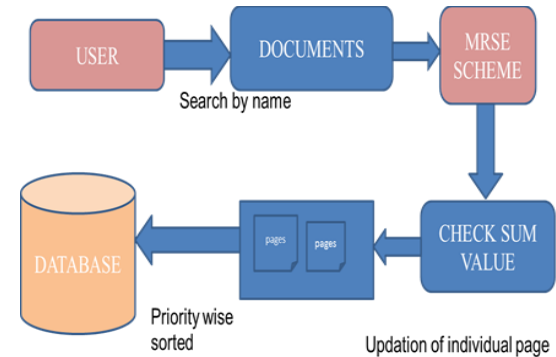
4.5 INDIVIDUAL PAGE UPDATION MODULE

Once the user checks about documents they are getting contents of various pages as an output. User wants to do the updation in particular page in a document, each page searched in a database and particular page is sorted out first and it is modified by the user. Those modification of particular page is done by using check sum technique in a database. It checks the pages and giving priority to the content.



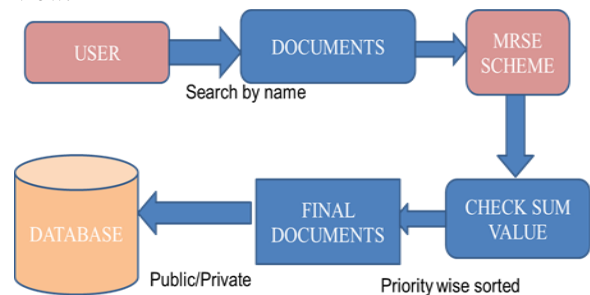
4.6 PRIORITY BASED SEARCH MODULE

User searched the document by using multi-keywords and finally outputs produced based on large number contents. Each and every content in a page is analyzed in a document and priority given to huge pages in a database. Priority wise pages are sorted in this module and cost is reduced because of using this technique in a cloud server.

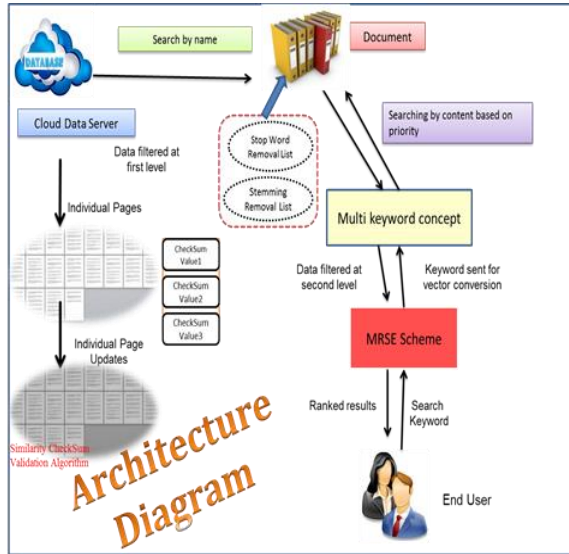


4.7 PERFORMANCE EVALUATION MODULE

Document is searched by name of document. And outputs produced based on the multi-keyword search in a document. Finally, documents saved in a database using public or private methodology. When the document is saved as public, its related to everyone's view.



5 ARCHITECTURE DIAGRAM



6 SYSTEM TESTING

TESTING OBJECTIVES

The purpose of testing is to discover 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-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

6.2 TYPES OF TESTS

6.2 Unit testing

Unit testing involves 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.

6.2.2 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 satisfactory, 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.

6.2.3 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 centred on the following items:

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

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

Functions: identified functions must be exercised.

Output: identified 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, systematic 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.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed. Features to be tested
- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

6.3 SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

6.4 Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format

- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

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

8 CONCLUSIONS

Main goal of this project is evaluation of storing data in cloud more secure. This section include various test conducted on data stored in cloud, these test are conducted on the basic of various parameters. Due to loss and Damage of Data Transmission, We Proposed one Concept. In order to overcome that, We Proposed a Technique to Transferring the Image or Video form Source to Destination Without any Loss of Data and Leakage of Data.