

Data Membership Identification using Bloom Filter in Cloud Storage for Effective Resource Allocation

Niraja Jain¹, Dr B Raghu², Dr V Khanaa³

^{1,3} research Scholar, Computer Engineering Department, Bharath University, Chennai, India

Principal, SVS Group of Institutes, Warangal, India

² Dean, Bharath University, Chennai, India

Email: ¹nijajain@gmail.com, ²balrajraghu@gmail.com, ³khanaakrishna@gmail.com

Abstract: Cloud computing has recently been the buzz word that had changed the entire software industry with its PaaS, IaaS and SaaS architecture model. The legacy systems operational in the organizations across different industries have been using the database that had an overhead in terms of cost of data storage, runtime operation and frequent data maintenance. Cloud database concept had challenged the existing storage and operational norms of data. In the distributed environment, resources used in the cloud databases need to identify whether the requested data belongs to the data nodes of a cluster. With databases began to be ubiquitous, the data storage needed to satisfy heterogeneous data structures rather the unstructured data storage support is looked for. The Use of Bloom's filter for data membership identification is the novel approach and can effectively improve the resource organization strategy on cloud. Dynamic resource organization can further improve the query efficiency as well. The concern raised during this is the data privacy which can also be ascertain by maintaining the data access authority levels. Bloom filter uses less memory space against the large dataset to store it's information. A Bloom filter is proposed to be used to determine whether an element is part of a reference set. It is a very compact hash-based data structure with efficient look-up times and a manageable risk of giving false positives.

Keywords: Dynamic resource organization, Blooms Filter, Cloud Databases, Query Optimization

1. Introduction

The NIST definition of Cloud Computing [1]: "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models." The World is creating large amount of data. It's predicted that the cloud storage data usage in the year 2020 is estimated to be 14% resident and 34% passing through the

cloud by IDS. Source: IDC Digital Universe, May 2010. Studying in detail the NIST definition we can further state that "Database Infrastructure as a Service" is a model that is highly recommended in the cloud based applications developed and used worldwide with the advancements in the Virtualization, WEB 2.0 technology. Cloud storage systems have become the most successful cloud computing applications in use. This cloud era has given rise to the most generic problem of making the resources available to the Cloud Service Requesting Clients. With the well-known Cloud models in practice today namely [6]:

1) Infrastructure as a Service: provides virtual storage, virtual machines, virtual infrastructure & other hardware assets as resource to the client on demand. The service provider is responsible for providing & managing the entire required infrastructure whereas the client will look after the deployment.

2) Software as a Service: this is complete operating environment applications, management and the user interface. Client is responsible to enter & manage the data through the user interface. The service provider will provide the application through a thin client interface.

3) Architecture as a Service: provides virtual machines, operating systems, applications, services, development frameworks, transactions & control structures. Client is responsible for installing & managing the applications whereas service provider will manage the infrastructure, operating system & enable software.

The problem of resource organization becomes many folds in the cloud computing world. The resources defined / used in every model are different and its optimum utilization needs to be ensured to guarantee the Cloud computing characteristics features. These features are characterized with the following :

- 1) Pay-as-you-go
- 2) ability to operate in heterogeneous environment
- 3) ability to use encryption techniques on stored data
- 4) on demand self-service
- 5) multi-tenancy
- 6) elasticity & scalability

Cloud DBMS Wish-list [2]:

- 1) Efficiency
- 2) Fault tolerance

- 3) Ability to run in heterogeneous environmental
- 4) ability to support business intelligence

Cloud Database Architecture:

Cloud computing architecture compose of the following components [3]: Cloud Services, Cloud Infrastructure, Cloud Platform and Cloud storage

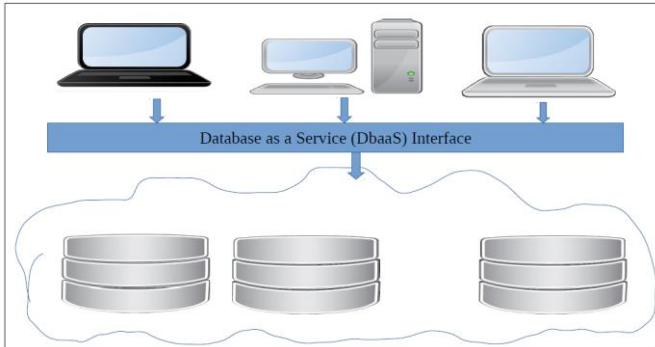


Fig 1. Cloud Database Architecture

2. Literature Survey

When the organization shift the responsibility of saving and storing its data to third-party, the Cloud service provider comes into picture. In a publication titled "Secure Data Transfer and Deletion from Counting Bloom Filter in Cloud Computing," authors C. Yang, X. Tao, F. Zhao and Y. Wang,[11] suggested that Cloud service users are compelled to frequently change the service provider due to its inability to stand-by the SLA terms. The secure data migration is the fundamental requirement and Bloom filter can be deployed to address this issue. The reason for doing so can range from finding an alternative to cost-effective storage, scalable alternatives to storing files on-premise or SAN (Storage Area Network), to because as stored data grows, companies have to invest in adding servers and infrastructure to accommodate increased demand. Zhibo Li, Kejing He, Funan Lin and Wenguo Wei,[10] in the article "Deduplication of files in cloud storage based on differential bloom filter," after analysing the existing cloud systems have identified that avoiding duplication of files in cloud storage is the problem so far unattended. The authors in the article proposed use of Bloom filter to accelerate the lookup speed and reducing computational cost and the results shows improved performance against computational time needed, memory utilization & deduplication efficiency. So all the data that the organization decides to store off-site becomes the responsibility of a third-party Cloud service provider. Shengjun Zhang, Hang Zhou, Yahui Yang and Zhonghai Wu, [12] have written a paper "A joint Bloom Filter and cross-encoding for data verification and recovery in cloud," that discusses the cloud storage security challenges. To ensure the integrity of the data storage in cloud the authors proposed a novel Cross-encoding recovery method using Dynamic Storage Table that uses Bloom filter to practically verify data

integrity. This data can further be accessed either through public internet or private dedicated network connection. The Cloud service provider will then host, secure, manage and maintain the storage servers and required infrastructure to ensure that the user/organization can access the data on demand. A. Singh, S. Garg, K. Kaur, S. Batra, N. Kumar and K. R. Choo, [13] "Fuzzy-Folded Bloom Filter-as-a-Service for Big Data Storage in the Cloud," suggests that existing storage solutions are deterministic in nature. Authors present an effective space-effective strategy for massive data storage using extended fuzzy folded bloom filter (FFBF), that improved efficacy without compromising false positive rate and query time. The service providers make use of the virtual servers commissioned on the physical servers to fulfill the user storage requests. J. Jeong, J. W. J. Joo, Y. Lee and Y. Son, [14] "Secure Cloud Storage Service Using Bloom Filters for the Internet of Things," in the article identify the need for data integrity verification which is usually done by third party. The authors proposed a provable data possession model using Bloom Filter. As the user demand grows, the service provider keeps adding the virtual servers to meet the access need. Thus cloud storage has elastic storage capacity that can grow with increase in demand and can shrink with decrease in demand without adding up any physical server cost. C. Wu, S. Horiuchi and K. Tayama, [17] "A Resource Design Framework to Realize Intent-Based Cloud Management," identified that the cloud service user need to pre-book the needed Virtual Machines (Vms) and the other computing resources necessary to allocate each VM and designed a Resource Design Framework (RDF) that uses cloud based machine learning scenario. This helps in perfect resource allocation and conforming the SLA policies. Y. Gao, K. Li and Y. Jin, [15] "Compact, Popularity-Aware and Adaptive Hybrid Data Placement Schemes for Heterogeneous Database as a Service (DbaaS) Interface Cloud Storage," states that the cloud services are overloaded due to input/output imbalance or space demand. Hashaware cloud storage can be the solution for this frequent problem ensuring reduced overload probability in cloud storage systems. The organization also saves on capital expenditure needed for storage commissioning and maintenance. The over-provisioning is taken care of by the service provider at the marginal cost increase saving user losses. T. Mengistu, A. Alahmadi, A. Albuali, Y. Alsenani and D. Che, [18] "A "No Data Center" Solution to Cloud Computing," presents opportunistic Cloud Computing system to demonstrate "no data center" solution to the problem of Infrastructure as a service. Also X. Chen et al.,[16] "Publicly Verifiable Databases with All Efficient Updating Operations," proposed VDB – Verifiable Database that enable resource-limited client to interact with the untrusted cloud server effectively that uses committed invertible Bloom Filter and supports all kinds of updating operations.

3. Problem Formulation

Typically the user will be given access to the storage data through Internet or private connection or using web portal, website or mobile app. As a service level policy, the provider will ensure the data availability through redundancy. Cloud storage can either be unmanaged or managed. Cloud storage service can be viewed in the light of its off-site data management, quick implementation, cost-effective, scalable solution for business continuity. While verifying the AWS cloud policies, it's identified that the Amazon have broadly classified Cloud storage into three categories [9]:

1) Object storage – scalability is the characteristic feature of this class. Amazon Simple Storage Service (S3) is

used in for its scale and flexibility, and ability to import existing data stores for analytics, backup, or archive.

2) File storage – which is often supported with a Network Attached Storage (NAS) server. Amazon Elastic File System (EFS) are ideal for use cases like large content repositories, development environments, media stores, or user home directories.

3) Block storage - is analogous to direct-attached storage (DAS) or a Storage Area Network (SAN) mostly in demand by enterprise applications like databases or ERP systems often require dedicated, low latency storage for each host. Amazon Elastic Block Store (EBS) are provisioned with each virtual server and offer the ultra-low latency required for high performance workloads[11].

Amazon Service	Service Features
Amazon Simple Storage Service (Amazon S3)	A service that provides scalable and highly durable object storage in the cloud.
Amazon Glacier	A service that provides low-cost highly durable archive storage in the cloud.
Amazon Elastic File System (Amazon EFS)	A service that provides scalable network file storage for Amazon EC2 instances.
Amazon Elastic Block Store (Amazon EBS)	A service that provides block storage volumes for Amazon EC2 instances.
Amazon EC2 Instance Storage	Temporary block storage volumes for Amazon EC2 instances.
AWS Storage Gateway	An on-premises storage appliance that integrates with cloud storage.
AWS Snowball	A service that transports large amounts of data to and from the cloud.
Amazon CloudFront	A service that provides a global content delivery network (CDN).

Table 1: Amazon Cloud Services Features

With the development of high capacity storage in 1990s, the Storage Service Providers opened up the datacenters worldwide fueling the dot com business. File transfer utility available then was the simplest form of unmanaged cloud storage. Unmanaged storage is presented to the client as if it is ready to use disk drive. The widespread use of this service further demanded sophistication in services such as synchronization and back-up. The second class of the

cloud storage services were the managed cloud storage which support the pay-as-you-go model. Managed storage involves provisioning of raw virtualized disk and use of it to support applications that use cloud based storage. This system supports virtualized storage component. Amazon S3, EMC Atmos, Nirvanix, Rackspace cloud, Iron Mountain are the well-known examples of managed cloud storage. Online Storage Service Providers: [19]

Cloud Service Provider	Storage Policy suitable for	Data Security Policy	Pricing	File Management
pCloud	Storing large files	TLS/SSL encryption	Free 10GB	Can be done from web, PC or mobiles
IDerive	Backup of files	256-bit AES encryption	Free 5GB	Real time sync with devices
DropBox	Light data users			
Google Drive	Teams & Collaboration		Free 15GB	Real time sync with devices
Microsoft OneDrive	For Windows users	SSL encryption	Free 5GB	Offline access allowed from any synced device
iCloud	Private users (Apple)		Free 5GB	Real time sync
OpenDrive	No file upload		Free 5GB (for	Business tools available

	limitation		personal use)	
Amazon S3	Suitable for large data	High security & performance	Priced	Scalability & data availability

Table 2: Online storage service providers

4. Data Storage Resource Organization In Cloud

In evaluating the cloud storage solutions, following factors are of important consideration:

- 1) Client self-service
- 2) Strong management capability
- 3) Performance characteristics
- 4) Appropriate block-based protocol
- 5) Seamless maintenance and upgrade

To successfully adjust the cloud service providers capacity, the system workload and characteristics of the data storage requested need to be understood thoroughly. The goal of capacity planner is to identify the critical resources and to keep adding more resources to avoid bottleneck dynamically. Scaling up vertically or to scale out horizontally is a critical architectural decision that affects the cloud performance largely. Data Storage need and the corresponding services available from Amazon are studied for the research work.:[20]

Storage Need	Service provider
File system	Amazon EFS
Structured data with query	Amazon DynamoDB Amazon RDS Amazon CloudSearch
Rapidly changing data	Amazon EC2 Amazon EFS Amazon DynamoDB
Archival data	Amazon Glacier
Dynamic website hosting	Amazon EC2 Amazon EFS Amazon DynamoDB Amazon RDS Amazon EFS

Table 3: Amazon Service solutions to different data storage needs

5. Research Methodology

For any cloud storage system to be user-friendly, it is very much essential to understand the client side storage requirement with proper analysis of the data with respect to the following features: storage space requirement, type of data, frequency of data upload, update and retrieval, data availability strictness, data security measures requested, predicted hikes in data operations, risk associated with data storage failure and the list goes on. Based on the data collected from the end user and the thorough analysis of it gives the service provider an insight about the data. In case of mismanaged data collection or storage, the entire service provider's job will be a nightmare.

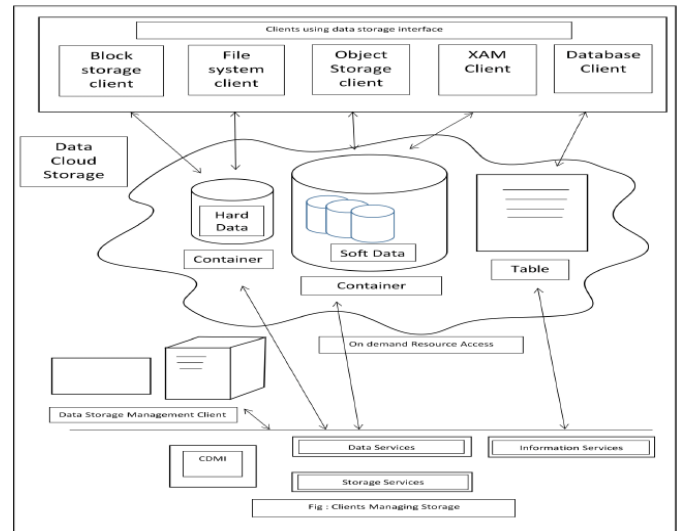


Fig. 2: Cloud Storage Data Model (Client managing the Storage)

Cloud storage services are basically untrusted storage. The chances of tampering with the data are high that leads to unavailability of necessary data. Such a data storage organization is primarily based on the data flow architecture. Hadoop is the widely used architecture in Big data industry that is identified as the best for processing huge amount of data simultaneously on large clusters in reliable, fault tolerant way. Capacity planning is an iterative process that deals with the inherent inefficiencies of the systems hampering the cloud service provider's market share due to fluctuations in SLA terms and actual service leveraged.

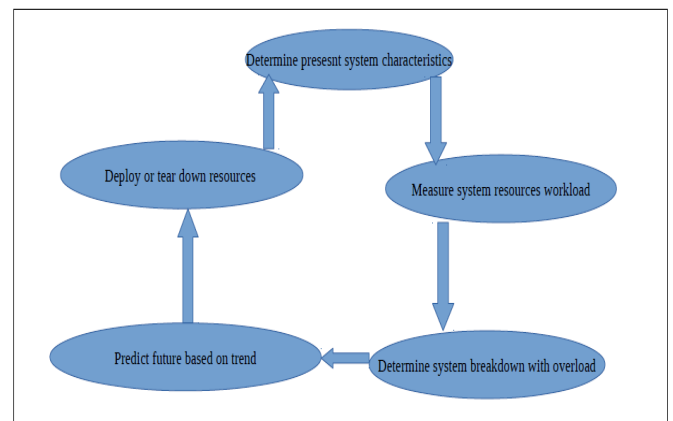


Fig 3: Cloud Resource Allocation Process

The authors of this paper in earlier publication titled "The future of database services: Cloud database", have Determine present system characteristics Determine system breakdown with overload Measure system resources workload Deploy or tear down resources Predict future based on trend

defined the Cloud storage system using the mathematical foundation as follows:

Data Storage Resource organization in the cloud environment can be mathematically defined as follows:

Let S the Cloud data storage system defined as $S = \{i,c,e,n,d | F(d), F(n), F(e)\}$

Herein ,

i= data storage index

c= data storage capacity

e= authority level for data accessed

n= data node

d= data available on cloud storage

for every $i \in S$, the functions need to be defined such

as:

$F(d): i \rightarrow d$

$F(n): i \rightarrow n$

$F(e): i \rightarrow e$

This suggests that the index defined on the data storage be such that it can identify whether the data is available on the cloud storage or not, if available then on which data node the data can be located and what is the authority level defined for data access.

A hash function is a mathematical function that converts an input value into a compressed numerical value – a hash or hash value. Basically, it's a processing unit that takes in data of arbitrary length and gives you the output of a fixed length – the hash value. Hash functions are widely used to identify or compare files or databases. Fundamentally, Hashing is defined by two distinct characteristics – irreversibility and uniqueness. Ir-reversibility underlines the fact that once we apply the hash function to the given data, it's not possible to regenerate the data using any de-hash function like it happens usually with encryption/decryption techniques. The hash function never produces same hash values for the two different data sets. Just in case hash-collision happens ever, that hash function is immediately discarded as it's useless.

As we have defined the Cloud storage system, the resource organization problem can be identified under the optimization problem categories, wherein it is to be identified that value of data storage index (i) can help in identifying the correct value of data node (n) where the requested data is stored. It's also to be ascertain that whenever a new data storage request is received by the Cloud service provider, it must identify the data storage capacity needed to accommodate the new data. This data also need to be organized very very close to the most relevant available data so that the further query retrieval rates can be reduced. While deciding on to the data node for data storage as per the indexing mechanism defined on the available data, the resource organization strategy also need to assume the authority level associated with that data. This together can suggest a dynamic resource organization strategy in the cloud environment for the data storage.

Bloom filter to detect data membership

Blooms filter is a resultant space efficient probabilistic data structure for approximate set association problem.

- Bloom filter consists of an array of x bits, $Z[0] \dots Z[n-1]$.
- Initially the array elements are all set to 0.
- The Bloom filter will utilize m independent random hash functions defined in the range of $\{0, \dots, x-1\}$ such as x_1, \dots, x_m .
- For every element $s \in S$, the bits of $Z[hi(s)]$ are set to 1 for $1 \leq i \leq m$.
- A bit location can be set to 1 multiple times, but only the first change will have an effect.
- So to check whether the requested element m is in s, all the array locations $Z[hi(n)]$ need to be checked if set to 1.
- If its not set to 1, then it clearly signifies that the requested data element n is not the member of S.

The Bloom filter thus is very much similar to a hash table. [3] The hash functions used in a Bloom filter should be independent and uniformly distributed. They should also be as fast as possible.

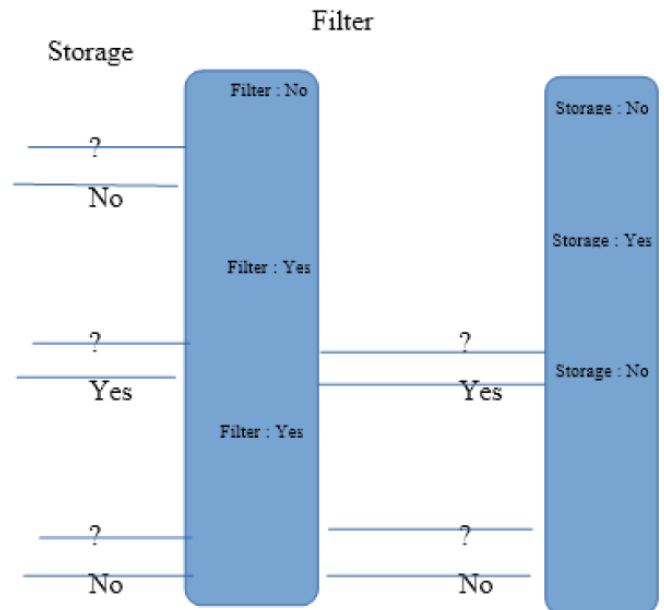


Fig 4. Bloom Filter usage to check data existence in storage system

Google BigTable, Apache Hbase, Cassandra etc make use of the Bloom Filter to reduce the cost of searching the non-existent data by multiple disk access. This in turn improves the database query operation.

6. Probability Of False Positives In Bloom Filter

Assume that we have written an hash function that will select each array element with equal probability. If there are m number of bits in an array, then the probability that a

typical element is not set to 1 by the defined function during the insertion of an element can be given by

$$(1 - 1/x) = e^{-mq/x}$$

$$\text{Let } Y = e^{-mq/x}$$

Assuming that a fraction p of the entries are still 0 after all the elements of S are hashed into Bloom filter.

The probability of False Positive is then Y .

$$(1 - (1 - 1/x)mq)^m \approx (1 - e^{-mq/x})^m \approx (1 - Y)^m$$

Bloom filter allows a constant probability of a false positive while keeping x/q the number of bits of storage required per item constant.

Many applications like password application accept this 1-2% false positive rates in practice.

7. Preprocessing Methodology:

Assume that the user is willing to store M blocks $M = \{m_0, m_1, \dots, m_{n-1}\}$

When the data is stored, the user needs to calculate the number of blocks of data M to be stored.

This number will influence the Bloom filter size and will also affect the positive error rate.

Before storing the data on the server, a key-value pair is generated which is a unique identification of the data known only to the data owner.

The bloom filter will be allowed to generate data verification index.

$$\text{Key-Value pairs} = \{(x_1, y_1), (x_2, y_2), \dots, (x_r, y_r)\}$$

Based onto the indexed value generated by the Bloom filter and comparing with the corresponding Key-value pairs representing a user block to be stored on the server the third party can also ascertain the validity of the data.

The malicious user also can not tamper with the data stored on the server due to the encryption of the key-value pair generated randomly corresponding the individual data block.

The verification is very much similar to an index creation where each block of data is the key building block. The randomly generated number of data blocks are verified for the data integrity and sent for verification. Bloom filter auditor will verify the data possession and accordingly the index will be verified. The constraint with this verification approach is that the indexing needs to be flexible enough to accommodate the addition or alteration to the key-value pairs.

8. Analysis And Remarks

The use of Bloom filter can be beneficial in the cloud storage systems management. We propose a scheme wherein the Global Bloom Filter will be defined in the storage cloud that will store the location details of the metadata on the servers. Whenever the user requests for the data, the Global bloom filter will take that data to the local bloom filter that is associated with each server. The hashed functions defined on the basis of the index will help to locate the exact server location for the data relevant to the requested query result. The

data flow technique is commonly used in the Map-Reduce infrastructure to improve on the efficiency. The similar techniques used for query execution will improve the cloud storage performance. When data storage organization strategy needs to be finalized by the cloud service provider, the ownership of the data by the server needs to be ascertained. The Cloud storage space is divided in to logical partitions of say n blocks. The need is to check on the set membership of each logical block using the Bloom filter. As this methodology primarily a probabilistic model, the threshold value needs to be decided on so that the data ownership verification process will not have insignificant impact of it. We are proposing to pre-process the data and verify its integrity.

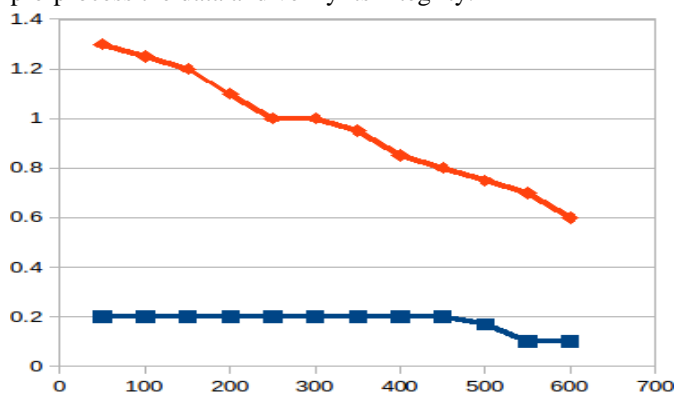


Fig Time taken to discover resources with and without Bloom Filter

9. Conclusion:

To determine the cloud computing performance it's essential to understand the demand pattern and predict the requirements. The cloud flexibility and scalability can be further improved using the probabilistic data structure. This paper presents an idea of using the Bloom filter to check the existing data stored on the cloud storage systems. The managed systems can be effectively checked for accommodating the newer data as per the indexing done on the available data. A Bloom filter can be used to determine whether an element is part of a reference set. It is a very compact hash-based data structure with efficient look-up times and a manageable risk of giving false positives.

References

- [1] Peter Mell "The NIST Definition of Cloud Computing, Recommendations of the National Institute of Standards and Technology NIST Special Publication 800-145 , Sept 2011
- [2] Niraja Jain, A V Deshpande, B Raghu, V Khanaa. "The effectiveness of cloud databases", 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017
- [3] <https://ilimlib.github.io/bloomfilter-tutorial/>
- [4] Barrie Sosinski, " Cloud Computing Bible" Wiley Publications

- [5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [6] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [7] Arun Swaminathan et al, A Design for a Bloom Filter Variant Based Cloud Database Validation System, (IJCSIT) *International Journal of Computer Science and Information Technologies*, Vol. 7 (5), 2016, 2329-2331
- [8] K. C. Davis, "Teaching Database Querying in the Cloud," 2019 IEEE Frontiers in Education Conference (FIE), Covington, KY, USA, 2019, pp. 1-7, doi: 10.1109/FIE43999.2019.9028440.
- [9] <https://aws.amazon.com/what-is-cloud-storage/>
- [10] Zhibo Li, Kejing He, Funan Lin and Wenguo Wei, in the article "Deduplication of files in cloud storage based on differential bloom filter," 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2016, pp. 111-114, doi: 10.1109/ICSESS.2016.7883027.
- [11] C. Yang, X. Tao, F. Zhao and Y. Wang, "Secure Data Transfer and Deletion from Counting Bloom Filter in Cloud Computing," in *Chinese Journal of Electronics*, vol. 29, no. 2, pp. 273-280, 3 2020, doi: 10.1049/cje.2020.02.015.
- [12] Shengjun Zhang, Hang Zhou, Yahui Yang and Zhonghai Wu, "A joint Bloom Filter and cross-encoding for data verification and recovery in cloud," 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, 2017, pp. 614-619, doi: 10.1109/ISCC.2017.8024596.
- [13] A. Singh, S. Garg, K. Kaur, S. Batra, N. Kumar and K. R. Choo, "Fuzzy-Folded Bloom Filter-as-a-Service for Big Data Storage in the Cloud," in *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2338-2348, April 2019, doi: 10.1109/TII.2018.2850053.
- [14] J. Jeong, J. W. J. Joo, Y. Lee and Y. Son, "Secure Cloud Storage Service Using Bloom Filters for the Internet of Things," in *IEEE Access*, vol. 7, pp. 60897-60907, 2019, doi: 10.1109/ACCESS.2019.2915576.
- [15] Y. Gao, K. Li and Y. Jin, "Compact, Popularity-Aware and Adaptive Hybrid Data Placement Schemes for Heterogeneous Cloud Storage," in *IEEE Access*, vol. 5, pp. 1306-1318, 2017, doi: 10.1109/ACCESS.2017.2668392.
- [16] X. Chen et al., "Publicly Verifiable Databases with All Efficient Updating Operations," in *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2020.2975777.
- [17] C. Wu, S. Horiuchi and K. Tayama, "A Resource Design Framework to Realize Intent-Based Cloud Management," 2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Sydney, Australia, 2019, pp. 37-44, doi: 10.1109/CloudCom.2019.00018.
- [18] T. Mengistu, A. Alahmadi, A. Albuali, Y. Alsenani and D. Che, "A "No Data Center" Solution to Cloud Computing," 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), Honolulu, CA, 2017, pp. 714-717, doi: 10.1109/CLOUD.2017.99.
- [19] <https://www.softwaretestinghelp.com/cloud-storage-providers>
- [20] AWS Storage Services Overview A Look at Storage Services Offered by AWS December 2016
- [21] Niraja Jain, Dr B Raghu, Dr V Khanna, Dr A V Deshpande, "The future of database services: Cloud database", 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017
- [22] Niraja Jain, Dr B Raghu "Dynamic resource organization and query optimization strategies for cloud databases", *International Journal of Innovative Trends and Emerging Technologies* ISSN 23499842(online), 2015
- [23] Niraja Jain, Dr B Raghu "DaaS (Database as a Service) in Cloud Computing", *International Journal of Innovations & Advancement in Computer Science IJIACS* ISSN 2347 – 8616, 2018
- [24] Niraja Jain, Dr B Raghu, Dr V Khanna, Dr A V Deshpande, "Need of Privacy Preservation In Cloud Databases", *Multidisciplinary Journal Of Research In Engineering And Technology*, Issn:2348 - 6953
- [25] I. Odun-Ayo, O. Ajayi, B. Akanle and R. Ahuja, "An Overview of Data Storage in Cloud Computing," 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS), Jammu, 2017, pp. 29-34, doi: 10.1109/ICNGCIS.2017.9.
- [26] W. Xiao, W. Bao, X. Zhu and L. Liu, "Cost-Aware Big Data Processing Across Geo-Distributed Datacenters," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 11, pp. 3114-3127, 1 Nov. 2017, doi: 10.1109/TPDS.2017.2708120.

