© 2019 IJSRSET | Volume 6 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET196274

Storage and Security Preservation Using Cloud Based Intelligent Compression Scheme

Dr. M. Chinnadurai, A. Jayashri

Head of the Department (CSE), E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu, India

CSE Department, E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu, India

ABSTRACT

Cloud computing is one of the important factoring that leads it into a productive phase. This means that most of the main problems with cloud computing have been addressed to a degree that clouds have become interesting for full commercial exploitation. However, permissions over data security still prevent many users from migrating data to remote storage. Client-side data compression in particular ensures that multiple uploads of the same content only consume network bandwidth and storage space of a single upload. Compression is actively used by a number of cloud backup providers as well as various cloud services. Unfortunately, encrypted data is pseudorandom and thus cannot be deduplicated: as a consequence, current schemes have to entirely sacrifice either security or storage efficiency. In this system, present a scheme that permits a more fine-grained trade-off. The intuition is that outsourced data may require different levels of protection, depending on how popular it is: content shared by many users. Then present a novel idea that differentiates data according to their popularity. In this proposed system, implement an encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data. Proposed data de-duplication can be effective for popular data, also semantically secure encryption protects unpopular content. Finally, can use the backup recover system at the time of blocking and also analyze frequent login access system.

Keywords : Secure File Storage, Duplicate Checking, Chunk based Similarity Checking, File Encryption, Backup Recovery

I. INTRODUCTION

Cloud computing is a computing paradigm, in which a large pool of systems are connected in non-public or public networks, to provide dynamically scalable infrastructure for utility, information and file storage. With the arrival of this system, the cost of computation, software web hosting, content storage and shipping is reduced substantially. It is a practical approach to experience direct cost advantages and it has the potential to transform a facts middle from a capital-intensive set up to a variable priced environment. The idea of cloud computing is primarily based on a very fundamental ideas of reusability of IT talents. The difference that cloud computing brings as compared to standard ideas of "grid computing", "dispensed computing", "application computing", or "autonomic computing" is to increase horizons across organizational obstacles. Forrester [1] defines cloud computing as: "A pool of abstracted, fantastically scalable, and managed compute infrastructure able to website hosting quit purchaser programs and billed with the aid of consumption". It is a generation that makes use of the net and vital faraway servers to maintain information and applications and allows consumers and companies to use packages with out installation and get entry to their personal files at any computer with internet get admission to. This era allows for a whole lot extra efficient computing via centralizing records storage, processing and bandwidth. Cloud computing examples are Yahoo email, Gmail, or Hotmail.

DEPLOYMENT MODELS OF CLOUD

Enterprises can pick to set up applications on Public, Private or Hybrid clouds. Cloud Integrators can play a vital part in figuring out the right cloud route for each agency.

1.5.1 Public Cloud

Public clouds are owned and operated by third events; they deliver advanced economies of scale to clients, because the infrastructure expenses are spread among a mix of customers, giving every person client an attractive low-cost, "Pay-as-you-go" version. All customers share the equal infrastructure pool with limited configuration, security protections, and availability variances. These are controlled and supported via the cloud company. One of the blessings of a Public cloud is that they'll be larger than an organizations cloud, therefore providing the capacity to scale seamlessly, on call for.

1.5.2 Private Cloud

Private clouds are constructed completely for a single company. They aim to cope with issues on information safety and offer more manage, that is usually lacking in a public cloud. There are two versions to a private cloud:

- On-premise Private Cloud
- Externally hosted Private Cloud Cloud

1.5.2.1 On-premise Private Cloud

On-premise personal clouds, also called internal clouds are hosted within one's own data center. This model affords a extra standardized manner and safety, but is restrained in elements of length and scalability. IT departments could additionally want to incur the capital and operational charges for the physical assets. This is exceptional applicable for programs which require complete manipulate and configurability of the infrastructure and safety.

1.5.2.2 Externally hosted Private Cloud

This form of personal cloud is hosted externally with a cloud provider, where the issuer facilitates an special cloud surroundings with full guarantee of privacy. This is fine suited for corporations that don't decide on a public cloud due to sharing of physical sources.

1.5.3 Hybrid Cloud

Hybrid Clouds is a combination of both public and private cloud models. With a Hybrid Cloud, service vendors can utilize 3rd party Cloud Providers in a complete or partial manner therefore growing the power of computing. The Hybrid cloud environment is capable of imparting on-demand, externally provisioned scale. The ability to reinforce a personal cloud with the assets of a public cloud can be used to control any unexpected surges in workload.

CLOUD COMPUTING CHALLENGES

Despite its growing have an impact on, concerns regarding cloud computing nonetheless remain. In our opinion, the benefits outweigh the drawbacks and the version is well worth exploring. Some common challenges are:

- Data Protection
- Data Recovery and Availability
- Management Capabilities
- Regulatory and Compliance Restrictions

1.7.1 Data Protection

Data Security is a critical element that warrants scrutiny. Enterprises are reluctant to shop for an guarantee of business data security from carriers. They fear losing data to opposition and the data confidentiality of consumers. In many times, the actual storage location isn't always disclosed, adding onto the security concerns of corporations. In the existing, firewalls throughout facts facilities (owned by corporations) shield this sensitive records. In the cloud version, Service providers are chargeable for retaining records security and corporations would must depend on them.

1.7.2 Data Recovery and Availability

All business programs have Service Level Agreements which can be stringently observed. Operational groups play a key role in management of service level agreements and runtime governance of programs. In manufacturing environments, operational groups assist

- Appropriate clustering and Fail over
- Data Replication
- System tracking (Transactions monitoring, logs monitoring and others)
- Maintenance (Runtime Governance)
- Disaster restoration
- Capacity and performance control

If, any of the above cited offerings is under-served via a cloud issuer, the damage & impact can be intense.

1.7.3 Management Capabilities

Despite there being more than one cloud providers, the management of platform and infrastructure remains in its infancy. Features like “Auto-scaling” as an instance, are a critical requirement for more businesses. There is big capability to improve at the scalability and cargo balancing features provided these days.

1.7.4 Regulatory and Compliance Restrictions

In a number of the European countries, Government policies do now not allow one's personal statistics and other sensitive statistics to be physically positioned outside the state. In order to satisfy such necessities, cloud carriers want to setup a data center or a storage web page exclusively inside the country to conform with regulations. Having such an infrastructure won't usually be possible and is a massive project for cloud companies. With cloud computing, the movement moves to the interface this is, to the interface among service providers and a couple of agencies of provider consumers. Cloud offerings will call for understanding in disbursed services, procurement, chance assessment and carrier negotiation areas that many businesses are best modestly geared up to deal with.

II. RELATED WORK

L. Wang, et.al,.[1] this studies contributions are threefold: first, we endorse an modern public cloud usage version for small-to medium scale scientific communities to utilize elastic resources on a public cloud website even as retaining their flexible gadget controls, i.e., create, prompt, suspend, resume, deactivate, and smash their excessive-stage control entities—carrier control layers without understanding the info of management. Second, we design and implement an revolutionary system—DawningCloud, at the middle of that are lightweight provider control layers going for walks on pinnacle of a common control provider framework. The common management provider framework of DawningCloud not simplest enables building light-weight carrier management layers for heterogeneous workloads, however also makes their management work easy. Thirdly, examine the systems comprehensively using each emulation and actual experiments. Here located that for 4 strains of two common workloads: High-Throughput Computing (HTC) and Many-Task Computing (MTC), DawningCloud saves the solution intake. First, with corresponding to Reservoir, the CSF of Dawning- Cloud facilitates building lightweight service control layers for heterogeneous workloads. This takes a backside-up technique to constructing DawningCloud. The commonplace units of capabilities for special runtime surroundings software program are delegated to the CSF. CSF helps building skinny carrier management layers—TRE for heterogeneous workloads, and a TRE most effective implements center functions for a particular workload.

B. Li, E. Mazur, et.al,.[2] implemented a step towards bringing the many benefits of the MapReduce model to incremental one-pass analytics. In the new model, the MapReduce system reads input data only once, performs incremental processing as more data is read, and utilizes system resources efficiently to achieve high performance and scalability. The goal is to design a platform to support such scalable, incremental one-pass analytics. This platform may be used to support interactive information analysis, which may additionally contain on line aggregation with early approximate solutions, and, inside the future, flow question processing, which affords near actual-time insights as new statistics arrives. Here argue that, so as to support incremental one-skip analytics, a MapReduce device should keep away from any blockading operations and also computational and I/O bottlenecks that save you information from "easily" flowing through map and decrease stages on the processing pipeline. We further argue that, from a overall performance perspective, the device desires to perform rapid in-reminiscence processing of a MapReduce question application for all, or maximum, of the facts. In the event that a few subset of facts has to be staged to disks, the I/O value of such disk operations have to be minimized. Our latest benchmarking take a look at evaluated current MapReduce structures together with Hadoop and MapReduce Online (which plays pipelining of intermediate records).

R. Kienzler, et.al,.[3] carried out an incremental facts get admission to and processing technique for data-in depth cloud applications that may cover information transfer latencies even as preserving linear scalability. Similar in spirit to pipelined query assessment in conventional database systems, data is accessed and processed in small increments, thereby propagating data chunks from one stage of the data analysis assignment to some other as quickly as they are to be had instead of waiting until the complete dataset becomes available. This manner will system statistics commonly in reminiscence (for this reason, reduce time-eating I/O to local disk and cloud garage, and keep away from storage amount) as well as achieving pipelined parallelism (similarly to the present partitioned parallelism), main to a discount in average challenge of completion time. In our method, data is accessed in a "stream-as-you-go" style rather than in whole batches, making a stream-primarily based information management architecture a good base for implementation. We have designed our "stream-as-you-go" approach for a sure magnificence of cloud packages characterized via the following. First of all, the records evaluation algorithms concerned within the utility ought to be appropriate for incremental processing (like handiest non-blocking off operators can be evaluated in a pipelined style in traditional databases).

C. Olston, et.al,.[4] a search index from a circulate of crawled net pages. Some of the numerous steps are compression, link analysis for unsolicited mail and exceptional classification, joining with click-based reputation measurements, and file inversion. Processing semi-based records feeds, e.g. Information and (micro-)blogs. Steps consist of de-duplication, geographic place decision, and named entity recognition. Processing alongside those traces is an increasing number of performed on a brand new of bendy and scalable records control systems, consisting of Pig/Hadoop. Hadoop is a scalable, fault-tolerant device for jogging person map-reduce processing operations over unstructured records files. Pig provides better-stage, based abstractions for statistics and processing. In Pig, a language known as Pig Latin is used to describe arbitrary acyclic records float graphs produced from two sorts of operations: (1) integrated relational-algebra-style operations (e.g. Filter out, be part of); and (2) custom person-defined operations (e.g. Extract net page hyperlinks, compute quintiles of a set of numbers). Despite the achievement of Pig/Hadoop, it's far turning into apparent that a new, better, layer is wanted: a work manager that offers with a graph of interconnected Pig Latin applications, with statistics exceeded amongst them in a non-stop style. Given that Pig itself offers with graphs of interconnected facts processing steps, it's natural to ask why one might layer any other graph abstraction on top of Pig.

K.H. Lee, et.Al,.[5] MapReduce is a programming model created by way of Google. It became designed to simplify parallel statistics processing on massive clusters. First model of the MapReduce library changed into written in February 2003. The programming version is stimulated with the aid of the map and reduces primitives observed in Lisp and other useful languages. Before growing the MapReduce framework, Google used masses of separate implementations to method and compute large datasets. Most of the computations have been pretty simple, but the input statistics became regularly very massive. Hence the computations needed to be allotted throughout loads of computer systems in order to complete calculations in an affordable time. MapReduce is relatively green and scalable, and for this reason can be used to method massive datasets. When the MapReduce framework turned into added, Google completely rewrote its web search indexing device to use the brand new programming version. The indexing gadget produces the statistics structures utilized by Google web search. The parallelization doesn't always have to be completed over many machines in a community. There are distinct implementations of MapReduce for parallelizing computing in exceptional environments. Phoenix is an implementation of MapReduce, that is geared toward shared-reminiscence, multi-center and multiprocessor systems, i.e. Single computers with many processor cores.

III. IMPLEMENTATION

MODULES

- Cloud storage framework
- File encryption
- Similarity checking
- Alert system
- Backup recovery system

MODULES DESCRIPTION

CLOUD STORAGE FRAMEWORK:

Cloud computing and storage solution offer users and companies with numerous talents to shop and system their data in both privately owned, or third-party data centers that may be placed far from the user–ranging in distance from across a city to the world over. Cloud computing is predicated on sharing of assets to attain coherence. In this framework, we can have two types of users such as data owner and data provider. The person or enterprise that legally owns a cloud service is known as a cloud data owner. The cloud storage provider can be the cloud owner, or the cloud provider that owns the cloud inside which the cloud storage resides. Cloud service provider provides the storage space to the users. Storage space can be shared by multiple data owners. Data owners can be upload the files in storage system for future use.

FILE ENCRYPTION:

Encryption is the simplest manner to acquire data safety. To examine an encrypted report, have to have get admission to a secret key or password that permits you to decrypt it. Unencrypted facts are known as undeniable text; encrypted statistics is called cipher text. There are two principal varieties of encryption:

asymmetric encryption (also known as public-key encryption) and symmetric encryption. Here can implement symmetric encryption for encrypt the data files using single key approach. The keys can be same or there may be a easy transformation to go among the 2 keys. The keys, in practice, constitute a shared secret among two or more events that can be used to maintain a personal data hyperlink. Encrypted facts can be saved in cloud server.

SIMILARITY CHECKING:

In computing, data compression is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and synonymous terms are compressed for single-instance (data) storage. This technique is used to improve storage utilization of server and can also be applied to network data transfers technique to reduce the number of bytes that must be sent. In the compression process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. As the analysis keeps, other chunks are as compared to the saved copy and every time a healthy occurs, the redundant chunk is replaced with a small reference that points to the saved chunk. In this module, we can check the files using file name with file contents. Encrypted files are spilted into chunks. Service provider checks the chunks at the time of uploading files. Data owner only upload original file so save storage space in cloud system. We can compression in text file, document file and image files.

ALERT SYSTEM:

In this system, we can design application for alert system for every week. After four weeks completed, if there is no access means the files are automatically sent to alternate mail and mobile which are stored at the time of registration. Server can save huge amount of storage and provide to other users. And also provide acknowledgement about to view the mail by recipients.

BACKUP RECOVERY APPROACH:

Admin can check access time for each user login. If user login to the system means, activity is registered in storage. And also monitor each user access. If the user access is paused more than 3 days means, admin automatically send alert to user based on registered mobile numbers. Finally if there is no access in storage system means, backup is generated. And flush the storage space and save storage for server for future use.

IV. METHODOLOGY

RSA for file encryption

Key Generation Process

Step 1: Generate two large random primes, p and q, of approximately equal size such that their product n=pq is of the required bit length, e.g. 1024 bits.

Step 2: Compute n=pqand $\phi=(p-1)(q-1)$.

Step 3: Choose an integer e, $1<e<\phi$, such that $\gcd(e,\phi)=$.

Step 4: Compute the secret exponent d, $1<d<\phi$, such that $ed\equiv 1 \bmod \phi$.

- The public key is (n,e)and the private key (d,p,q). Keep all the values d, p, q and ϕsecret.
- n is known as the *modulus*.
- e is known as the *public exponent* or *encryption exponent* or just the *exponent*.
- d is known as the *secret exponent* or *decryption exponent*.

Encryption

Sender A does following:

Step 1: Obtains the recipient B's public key (n,e).

Step 2: Represents the plaintext message as a positive integer mm with 1<m<n1<m<n.

Step 3: Computes the ciphertext $c=m^e \bmod n$.

Step 4: Sends the ciphertext cc to B.

Decryption

Recipient B does the following:

Step1 : Uses his private key (n,d)to compute m=c^{d}mod n.

Step 2 : Extracts the plaintext from the message representative mm.

V. EXPERIMENTAL RESULTS

The Mean Squared Error (MSE):

The difference among unique photograph information and compressed photo records is referred to as suggest rectangular blunders (MSE). MSE is inversely proportional to PSNR, as MSE decreases the PSNR will increase. PSNR suggest fine of photograph. Image compression is lossless while MSE is 0. Its higher to have less MSE. F(x,y) is the pixel value of the unique photograph, and f'(x,y)is the pixel price of the decoded image. The MSE calculated through,

$$MSE = \sqrt{\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}[f(x,y) - f'(x,y)]^2}$$

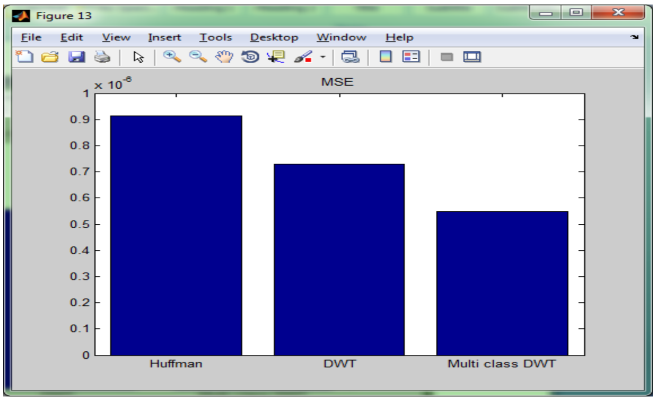

Fig 1 : Comparison chart of different compression algorithm using MSE measurements.

Peak Signal to noise Ratio (PSNR):

PSNR is the ratio between most sign powers to noise appear in sign. PSNR is related to excellent of image. For good exceptional of photo the PSNR of photo need to be high. PSNR is relies upon upon the suggest rectangular error (MSE) of picture. When the distinction between the unique photograph and compressed is less the PSNR is excessive so subsequently the nice of image is likewise high.

$$PSNR = 10log\frac{MAX^2}{MSE}$$

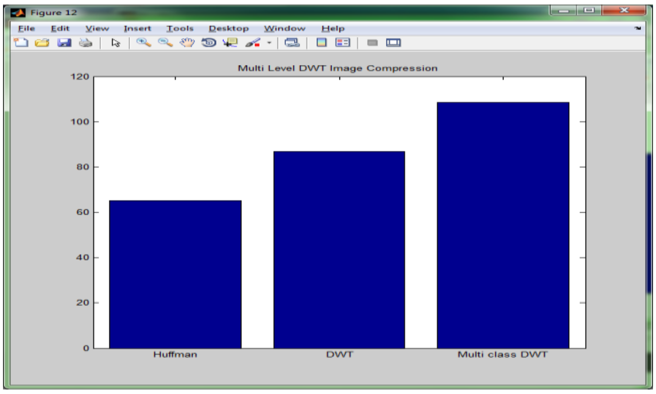

Fig 2 : Comparison chart of different compression algorithm using PSNR measurements.

VI. CONCLUSION

In this work proposed the distributed compression systems to improve the reliability of data while achieving the confidentiality of the users and also shared authority outsourced data with an encryption mechanism. Four constructions were proposed to support file-level and block-level data compression. The security of file consistency and integrity were achieved successfully. Here implemented compression systems using the secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations. In this work, we have identified a new privacy challenge during data accessing in the cloud computing to achieve privacy-preserving access authority sharing for similarity files. Authentication is established to assure data confidentiality and data integrity. Data anonymity is achieved because the wrapped values are exchanged at some stage in transmission. User privacy is more suitable by access requests to privately tell the cloud server approximately the customers get access of data. The backup recovery scheme is to improve the recovered scheme to avoid the blockages and also refund the amount to unused spaces in cloud system.

VII.REFERENCES

[1]. L. Wang, J. Zhan, W. Shi and Y. Liang, “In cloud, can scientific communities benefit from the economies of scale?” IEEE Transactions on Parallel and Distributed Systems 23(2): 296-303, 2012.

[2]. B. Li, E. Mazur, Y. Diao, A. McGregor and P. Shenoy, “A platform for scalable one-pass analytics using mapreduce,” in: Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'11), 2011, pp. 985-996.

[3]. R. Kienzler, R. Bruggmann, A. Ranganathan and N. Tatbul, “Stream as you go: The case for incremental data access and processing in the cloud,” IEEE ICDE International Workshop on Data Management in the Cloud (DMC'12), 20124C. Olston, G. Chiou, L. Chitnis, F. Liu, Y. Han, M. Larsson, A. Neumann, V.B.N. Rao, V. Sankarasubramanian, S. Seth, C. Tian, T. ZiCornell and X. Wang, “Nova: Continuous pig/hadoop workflows,” Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'11), pp. 1081-1090, 2011.

[4]. K.H. Lee, Y.J. Lee, H. Choi, Y.D. Chung and B. Moon, “Parallel data processing with mapreduce: A survey,” ACM SIGMOD Record 40(4): 11-20, 2012.

[5]. X. Zhang, C. Liu, S. Nepal and J. Chen, “An Efficient Quasiidentifier Index based Approach for Privacy Preservation over Incremental Data Sets on Cloud,” Journal of Computer and System Sciences (JCSS), 79(5): 542-555, 2013.

[6]. X. Zhang, T. Yang, C. Liu and J. Chen, “A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization using Systems, in MapReduce on Cloud,” IEEE Transactions on Parallel and Distributed, 25(2): 363-373, 2014.

[7]. N. Laptev, K. Zeng and C. Zaniolo, “Very fast estimation for result and accuracy of big data analytics: The EARL system,” Proceedings of the 29th IEEE International Conference on Data Engineering (ICDE), pp. 1296-1299, 2013.

[8]. T. Condie, P. Mineiro, N. Polyzotis and M. Weimer, “Machine learning on Big Data,” Proceedings of the 29th IEEE International Conference on Data Engineering (ICDE), pp. 1242-1244, 2013.

[9]. Aboulnaga and S. Babu, “Workload management for Big Data analytics,” Proceedings of the 29th IEEE International Conference on Data Engineering (ICDE), pp. 1249, 2013

Cite this article as :

Dr. M. Chinnadurai, A. Jayashri, "Storage and Security Preservation Using Cloud Based Intelligent Compression Scheme", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 417-424, March-April 2019. Available at doi : https://doi.org/10.32628/IJSRSET196274
Journal URL : http://ijsrset.com/IJSRSET196274