Data Processing of Consistency Service in Cloud Computing

R. Nandhagopal[1] R.Mohanabharathi[2]
[1]PG Scholar [2]Associate Professor
[1,2]Department of Computer Science & Engineering
[1,2]Selvam College of Technology, Namakkal, Tamilnadu, India

Abstract— The cloud service provider is a key-value data storage system, where each piece of data is identified by a unique key. A cloud service provider maintains multiple replicas for each piece of data distributed servers. To provide always-on services, the cloud service provider replicates all of the data on multiple geographically distributed cloud servers. A cloud consists of a group of users that cooperation a job. Consistency as a service model, which consists of a large data cloud and multiple small audit clouds. The load rebalancing task to storage nodes by having the storage nodes balance their loads spontaneously. The outsourcing to data cloud, the audit cloud and the data cloud will engage in a service level agreement, which promised level of consistency that should be provided by the data cloud. This eliminates the dependence on central nodes. The outsourced data will be splitted based on the storage servers parameters and the splitted data will be stored on different servers. High consistency implies high cost and reduced availability. The implementation of the data cloud is opaque to all users due to the virtualization technique.

Key words: Cloud Computing, Consistency Model, Data Replication

I. INTRODUCTION

The Key enabling technologies for clouds include the MapReduce programming paradigm, distributed file systems, virtualization, and so forth. These techniques emphasize scalability, so clouds can be large in scale, and comprising entities can arbitrarily fail and join while maintaining system reliability. Distributed file systems are key building blocks for cloud computing applications based on the MapReduce programming paradigm. In such file systems, nodes simultaneously serve computing and storage functions; a file is partitioned into a number of chunks allocated in distinct nodes so that MapReduce tasks can be performed in parallel over the nodes.

Ensuring consistency is the primary requirement for all the replication technologies. Load balance among storage nodes is a critical function in clouds. In a load balanced cloud, the resources can be well utilized and provisioned, maximizing the performance of Map Reduce based applications. The centralized approach simplifies the design and implementation of a distributed file system.

However, recent experience concludes that when the number of storage nodes, the number of files and the number of accesses to files increase linearly, the central nodes e.g., the master in Google GFS become a performance bottleneck, as they are unable to accommodate a large number of file accesses due to clients and Map Reduce applications. Specifically, in this study, we suggest offloading the load rebalancing task to storage nodes by having the storage nodes balance their loads spontaneously. This eliminates the dependence on central nodes.

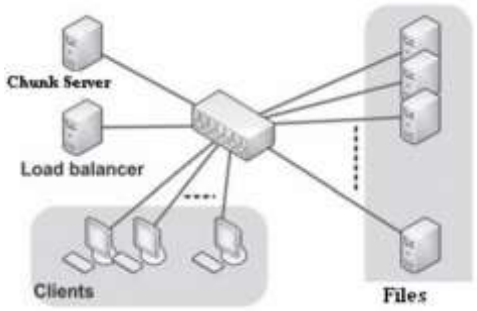


Fig. 1: Data Processing of Consistency Service Architecture

The storage nodes are structured as a network based on distributed hash tables DHTs discovering a file chunk can simply refer to rapid key lookup in DHTs, given that a unique handle or identifier is assigned to each file chunk. DHTs enable nodes to self-organize and repair while constantly offering lookup functionality in node dynamism, simplifying the system provision and management. we will conduct a thorough theoretical study of consistency models in cloud computing.

II. RELATED WORK

Cloud computing is computing in which large groups of centralized data storage and online access to computer services or resources. To provide ubiquitous always-on access, a cloud service provider CSP maintains multiple replicas for each piece of data on geographically distributed servers. Cloud Data Storages CDS basically focus on the maintenance of the large data which can't be stored on the end-user or clients system. A key problem of using the replication technique in clouds is that it is very expensive to achieve strong consistency on a worldwide scale.

Consistency as a service model, which consists of a large data cloud and multiple small audit clouds. A data cloud is maintained by a CSP, and a group of users that constitute an audit cloud can verify whether the data cloud provides the promised level of consistency or not. We propose a two-level auditing architecture, which only requires a loosely synchronized clock in the audit cloud. Then, we design algorithms to quantify the severity of violations with two metrics: the commonality of violations, and the staleness of the value of a read. Finally, we devise a heuristic auditing strategy HAS to reveal as many violations as possible. Extensive experiments were performed using a combination of simulations and real cloud deployments to validate HAS.

The synthesis workloads stress test the load-balancing algorithms by creating a few storage nodes that are heavily loaded. The computer simulation results are encouraging, indicating that our proposed algorithm performs very well. A simultaneously serve computing and storage functions; a file is partitioned into a number of chunks allocated in distinct.

Files can be dynamically created, deleted, and appended. This results in load imbalance in a distributed file system the file chunks are not distributed as uniformly as possible among the nodes. Emerging distributed file systems in production systems strongly depend on a central node for chunk reallocation.

III. EXISTING SYSTEMS

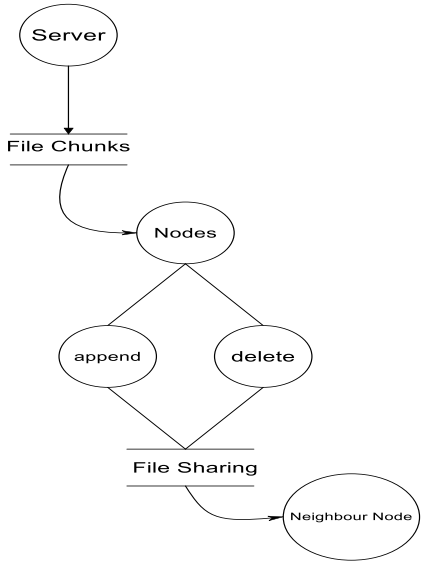
To meet the promise of ubiquitous 24/7 access, the cloud service provider CSP stores data replicas on multiple geographically distributed servers. A key problem of using the replication technique in clouds is that it is very expensive to achieve strong consistency on a worldwide scale, where a user is ensured to see the latest updates. Actually, mandated by the CAP principle3, many CSPs only ensure weak consistency, such as eventual consistency, for performance and high availability, where a user can read stale data for a period of time.

The domain name system DNS is one of the most popular applications that implement eventual consistency. Updates to a name will not be visible immediately, but all clients are ensured to see them eventually.

IV. PROPOSED WORK

The consistency as a service model. Then, we describe the structure of the user operation table UOT, with which each user records his operations. Finally, we provide an overview of the two-level auditing structure and related definitions. The cloud service provider is a key-value data storage system, where each piece of data is identified by a unique key. To provide always-on services, the CSP replicates all of the data on multiple geographically distributed cloud servers.

An audit cloud consists of a group of users that cooperation a job, e.g., a document or a program. We assume that each user in the audit cloud is identified by a unique ID. Before outsourcing the job to the data cloud, the audit cloud and the data cloud will engage in a service level agreement (SLA),which stipulates the promised level of consistency that should be provided by the data cloud. The audit cloud exists to verify whether the data cloud violates the SLA or not, and to quantify the severity of violations.



A. Consistency Model

Cloud applications typically use data that is dispersed across data stores. Managing and maintaining data consistency in this environment can become a critical aspect of the system, particularly in terms of the concurrency and availability issues that can arise. You frequently need to trade strong consistency for availability. This means that you may need to design some aspects of your solutions around the notion of eventual consistency and accept that the data that your applications use might not be completely consistent all of the time. This module is developed to provide the consistent data for different users from different regions.

Multi Agent Systems are basically used in artificial intelligence area as a technique for finding solution to the problems. In cloud computing they are used to develop an architecture for integrity of the data present at data centers. Data encoding is one of the basic mechanism of providing security, so we combine these two techniques to provide better integrity of data centers and data within the data centers. In this module we will first provide an introduction about the cloud computing and methodologies which will be used by us. Then we will be giving the work we have done related to our proposed model and further describe about the future work in this area. Cloud Computing can be defined as a computing paradigm that provides dynamic computing environment for end users that is reliable and customized and also guarantees quality of service.

Cloud computing with its acceptance also has some growing needs which affect the complete working of cloud, and one of those needs is the need for “security”. Cloud at present is lacking in its security needs in terms of data integrity, authorization and confidentiality. Data centers as the name suggests are the “house for data” for the purpose for data storage, management, analysis and dissemination. Data centers may exist in physical environment or virtually and can be organized as a public data center for large scale usage or a private data center specific to an organization.

$$n_{i+1} = \min(l, k \times n_i), \quad n_i \geq \alpha$$
$$n_{i+1} = \max(1, \tfrac{1}{k} \times n_i), \quad n_i < \alpha$$

Data centers today are one of the main needs for the increasing information technology services and have an important role in cloud computing. The end-users provide their data to cloud to access it whenever required on the rental basis, therefore, the data provided is stored at data centers of cloud known as cloud data storages.

These are present at different locations and store the complete data present on cloud. there are also one of the rising trends in IT field and suffer from the issue of security within it. Even though there are many security issues related to Data centers or data storages but one of the most important issue is integrity of the data.

B. Data Replication

The data outsourced by the client should be accessed from anywhere at anytime for their use. The failure of the server will not interrupt the outsourced data accessibility over cloud computing. So we design this module for higher data availability in distributed cloud environment. The outsourced data will be splitted based on the storage server's parameters and the splitted data will be stored on different servers. Every data block will be stored in more than one server to improve the availability of the data. If one server is fails or not accessible, the corresponding data will be collected from the another server which having the replicated data.

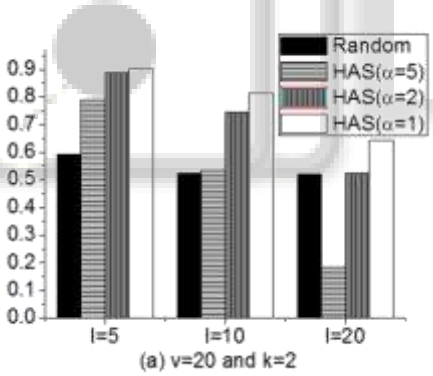


Fig. 2: Comparison of percentage of revealed violations have the same threshold values

Typical service-level agreement states what the provider has agreed to deliver in terms of availability and response to demand. The service level might, for example, specify that the resources will be available 99.999% of the time and that more resources will be provided dynamically if greater than 80% of any given resource is being used.

Within a private IaaS model, renting takes on a different focus. Although you might not charge each user to access a resource, in the charge-back model, you can allocate usage fees to an individual department based on usage over a week, month, or year. Because of the flexibility of the IaaS model, you can charge more of the budget to heavy users.

A computational task is typically replicated in space, i.e. executed on separate devices, or it could be replicated in time, if it is executed repeatedly on a single device. Replication in space or in time is often linked to scheduling algorithms. The access to a replicated entity is typically uniform with access to a single, non-replicated entity. The replication itself should be transparent to an external user. Also, in a failure scenario, a failover of replicas is hidden as much as possible. The latter refers to data replication with respect to Quality of Service aspects.

C. Data and Client Centric Design

A cloud is essentially a large-scale distributed system where each piece of data is replicated on multiple geographically distributed servers to achieve high availability and high performance.

Data-centric consistency model considers the internal state of a storage system, i.e., how updates flow through the system and what guarantees the system can provide with respect to updates. However, to a customer, it really does not matter whether or not a storage system internally contains any stale copies. As long as no stale data is observed from the client's point of view, the customer is satisfied. Therefore, client-centric consistency model concentrates on what specific customers want, i.e., how the customers observe data updates.

In, data replication across datacenters with the objective of reducing access delay is proposed. The Optimal replication site is selected based on the access history of the data. A weighted k-means clustering of user locations is used to determine replica site location. The replica is deployed closer to the central part of each cluster. A cost-based data replication in cloud datacenter is proposed in. This approach analyzes data storage failures and data loss probability that are in the direct relationship and builds a reliability model. Then, replica creation time is determined by solving reliability function.

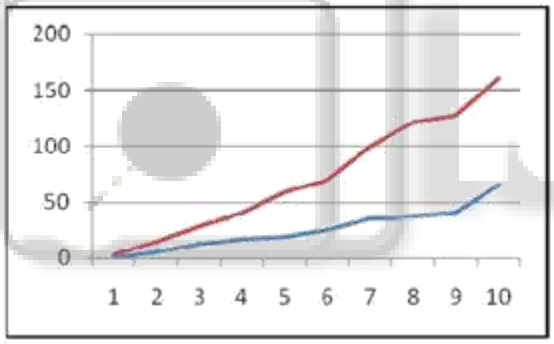
The approach presented in this paper is different from all replication approaches discussed above (a) by the scope, which implements data replication both within a data center as well as between geographically distributed data centers, (b) by the optimization target, which takes into account system energy consumption, network bandwidth, and communication delays. In this module we assume multiple cloud computing Data centers geographically distributed across the globe. Each datacenter has a three tier topology. Its interconnection network comprises of the core, aggregation, and access layers. The core layer provides packet switching backplane for all the flows going in and out of the datacenter. The aggregation layer integrates connections and traffic flows from multiple racks. The access layer is where computing servers are arranged into racks.

In this paper we assume multiple cloud computing datacenters geographically distributed across the globe. Each datacenter has a three tier topology. Its interconnection network comprises of the core, aggregation, and access layers. The core layer provides packet switching backplane for all the flows going in and out of the datacenter. The aggregation layer

integrates connections and traffic flows from multiple racks. The access layer is where computing servers are arranged into racks. In this module we assume multiple cloud computing datacenters geographically distributed across the globe. Each datacenter has a three tier topology. Its interconnection network comprises of the core, aggregation, and access layers. A cloud is essentially a large-scale distributed system where each piece of data is replicated on multiple geographically distributed servers to achieve high availability and high performance. Data-centric consistency model considers the internal state of a storage system, i.e., how updates flow through the system and what guarantees the system can provide with respect to updates. However, to a customer, it really does not matter whether or not a storage system internally contains any stale copies. As long as no stale data is observed from the client's point of view, the customer is satisfied. Therefore, client-centric consistency model concentrates on what specific customers want, i.e., how the customers observe data updates.

D. Threat-Aware Data Maintenance

Cloud Computing is a flexible, cost-effective, and proven delivery platform for providing business or consumer IT services over the Internet. However, cloud Computing presents an added level of risk because essential services are often outsourced to a third party.



This shift in control is the number one reason new approaches and techniques are required to ensure organizations can maintain data security. When an outside party owns, controls, and manages infrastructure and computational resources, how can you be assured that business or regulatory data remains private and secure, and that your organization is protected from damaging data breaches and feel you can still completely satisfy the full range of reporting, compliance, and regulatory requirements. Data protection tops the list of cloud concerns today.

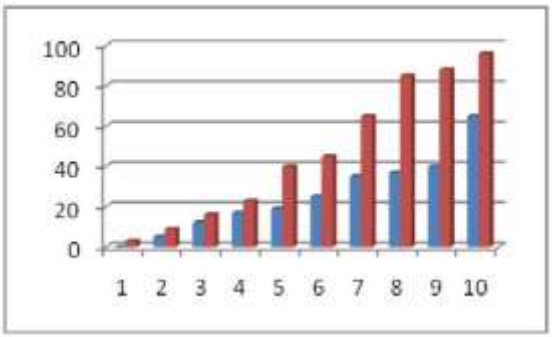
Data Privacy Directive Lack of standards about how cloud service providers securely recycle disk space and erase existing data. Auditing, reporting, and compliance concerns Loss of visibility to key security and operational intelligence that no longer is available to feed enterprise IT security intelligence and risk management . A new type of insider who does not even work for your company, but may have control and visibility into your data Such issues give rise to tremendous anxiety about security risks in the cloud.

Enterprises worry whether they can trust their employees or need to implement additional internal controls in the private cloud, and whether third-party providers can provide adequate protection in multitenant environments that may also store competitor data. There's also ongoing concern about the safety of moving data between the enterprise and the cloud, as well as how to ensure that no residual data remnants remain upon moving to another cloud service provider.

Unquestionably, virtualized environments and the private cloud involve new challenges in securing data, mixed trust levels, and the potential weakening of separation of duties and data governance. The public cloud compounds these challenges with data that is readily portable, accessible to anyone connecting with the cloud server, and replicated for availability. And with the hybrid cloud, the challenge is to protect data as it moves back and forth from the enterprise to a public cloud.

IaaS lets the consumer provision processing, storage, networks, and other fundamental computing resources and controls operating systems, storage, and deployed applications. As with Amazon Elastic Compute Cloud, the consumer does not manage or control the underlying cloud infrastructure. Data security is typically a shared responsibility between the cloud service provider and the cloud consumer. Data encryption without the need to modify applications is a key requirement in this environment to remove the custodial risk of IaaS infrastructure personnel accessing sensitive data.

Meanwhile, conventional security considerations must be addressed in the cloud environment. These include implementing best practices and real-time security intelligence, protecting data security, and preventing advanced persistent threats (APTs) or attacks that exploit social engineering. It's also critical to plan for the added risks posed by big data mined across different cloud environments and mobile devices that store information in the cloud infrastructure.



Cloud Computing is a flexible, cost-effective, and proven delivery platform for providing business or consumer IT services over the Internet. However, cloud Computing presents an added level of risk because essential services are often outsourced to a third party, which makes it harder to maintain data security and privacy, support data and service availability, and demonstrate compliance. This module will remove the security bottleneck over the outsourced data. We split the outsourced data and store each data block in different

servers, the hacker can't get the outsourced private data by compromising the single storage server.
This module will remove the security bottleneck over the outsourced data. We split the outsourced data and store each data block in different servers, the hacker can't get the outsourced private data by compromising the single storage server.

E. Load Rebalancing

Nodes simultaneously serve computing and storage functions; a file is partitioned into a number of chunks allocated in distinct. Files can be dynamically created, deleted, and appended. This results in load imbalance in a distributed file system; that is, the file chunks are not distributed as uniformly as possible among the nodes. Emerging distributed file systems in production systems strongly depend on a central node for chunk reallocation. The chunk reallocation is done in this module.

V. CONCLUSION

A consistency as a service CaaS model and a two-level auditing structure to help users verify whether the cloud service provider CSP is providing the promised consistency, and to quantify the severity of the violations, if any. With the CaaS model, the users can assess the quality of cloud services and choose a right CSP among various candidates, e.g., the least expensive one that still provides adequate consistency for the users' applications. For our future work, we will conduct a thorough theoretical study of consistency models in cloud computing. High consistency implies high cost and reduced availability. The implementation of the data cloud is opaque to all users due to the virtualization technique. Thus, it is hard for the users to verify whether each replica in the data cloud is the latest one or not. Inspired by the solution in the users in the audit cloud to verify cloud consistency by analyzing a trace of interactive operations. Unlike their work, we do not require a global clock among all users for total ordering of operations.

REFRENCES

[1] A Light-weight Data Replication for Cloud Data Centers Environment, Author: Mohamed-K Hussein, Mohamed-H Mousa, 2012
[2] Consistency in Distributed Storage Systems, Author: David Bermbach and Jorn, Kuhlenkamp Karlsruhe Institute of Technology, Karlsruhe, Germany, 2011
[3] Consistency-Based Service Level Agreements for Cloud Storage, Author:Douglas B. Terry, Vijayan Prabhakaran, Ramakrishna Kotla, Mahesh Balakrishnan, Marcos K. Aguilera, Hussam Abu-Libdeh, Microsoft Research Silicon Valley Cornell University, 2013
[4] Consistency Models for Replicated Data, Author:Alan D.Fekete and Krithi RamamrithamB. Charron-Bost, F. Pedone, and A. Schiper (Eds.): Replication, LNCS 5959, pp. 1–17, 2010.C Springer-Verlag Berlin Heidelberg 2010
[5] Countermeasures and Security Threats in Cloud Computing, Author: Vahid Ashktorab, Seyed Reza Taghizadeh, Department of Computer Engineering, Islamic Azad University of NajafAbaad, Department of Information Technology, Kahje-Nassir-Toosi University of Technology – Iran, 2012
[6] Energy-Efficient Data Replication in Cloud Computing Datacenters, Author: Dejene Boru, Dzmitry Kliazovich, Fabrizio Granelli, Pascal Bouvry, Albert Y. Zomaya, CREATE-NET, DISI - University of Trento, University of Luxembourg, School of Information Technologies, University of Sydney, Australia, 2013
[7] Enforcing Policy and Data Consistency of Cloud Transactions, Author: Marian K. Iskander Dave W. Wilkinson Adam J. Lee Panos K. Chrysanthis, Department of Computer Science, University of Pittsburgh, 2010
[8] On Limitations of Using Cloud Storage for Data Replication, Author: Christian Cachin, Birgit Junker, Alessandro Sorniotti, IBM Research, Open Systems AG-Switzerland, 2009
[9] Quality-of-Service for Consistency of Data Geo-replication in Cloud Computing, Author: Sergio Esteves, Joao Silva, and Luıs Veiga, UTL-Portugal, 2012
[10] Security and Privacy Challenges in Cloud Computing Environments, Author: Hassan Takabi and James b.d.Joshi, 2010
[11] Security Issues with Possible Solutions in Cloud Computing, Author: Abhinay B.Angadi, Akshata B.Angadi, Karuna,C.Gull, International Journal of Advanced Research in Computer Engineering & Technology 2013

