Minimizing Payment Cost of Cloud Service Provider

Tarun Sharma[1], and Nagaraju Bogiri[2],
[1]Student, [2]Professor of K. J. College of Engineering and Management Research
[1]K. J. College of Engineering and Management Research, Savitribai Phule, Pune, India
[2]K. J. College of Engineering and Management Research, Savitribai Phule, Pune, India

Abstract

Todays, more and more businesses are shifting their workloads to cloud storage to save capital costs for building and maintaining the hardware infrastructures and avoiding managing data centers complexity. Now the cloud computing becomes commercial service to data storage and access. The CSSP (Cloud Storage Service Provider) provides data storage services in which includes Get and Put function, using this geographically distributed data centers around the world. Selection of different CSSPs datacenters and cloud customers facing two challenges, First one is how allocating data on the datacenters in worldwide to satisfy SLO (Service level objectives) requirement which includes both data availability and retrieval latency and second is how allocate reserve resources and data in the datacenters, which belongs to different CSSP to minimizing payment cost. Find out the solution of these challenges firstly, used integer programming techniques for handles cost minimization problems. Here, used three techniques for reducing service latency and payment cost first multicast-based data transferring; the second coefficient based data reallocation and third is request redirection based congestion. In this also using PPM-C (Prediction by Partial Matching-Cloud) data compression technique this helps to reduce storage cost and data transfer computing time.

Keywords: Service Level Objectives, Cloud Computing, Resources Reservation, CSSP, Payment Cost Minimization, PPM-C compression, and Data Availability.

1. Introduction

Cloud storage provides an infinite amount of storage spaces for clients to protect their data backups on a 'Pay-as-You-Go' basis, so remote back-up is an Internet of dynamically scalable resources as a filling service. Because of this cloud services, now you can remotely archive your data form the third-party cloud storage provider instead of managing the data center on its own. Deployment in cloud classified from the physical location model. More case studies show that the use of cloud storage to back up remote data. Apart from the companies and government agencies, individuals can also use tools like Dropbox to store personal data on the clouds. In the advent of smartphones, the storage resources of smartphones are limited, so you can use the Dropbox-like tool to transfer Audio/Video files to your smartphone.

Web services face many challenges due to poor abstraction and the lack of performance guarantees offered by cloud services. First time every cloud storage service offers an isolated storage pool in each of its own data centers, replication between data centers applications. Second, the limited low-level interface offered by cloud storage makes it difficult for applications to implement any replication protocols. Third, popular public cloud storage services today face high latency, which can significantly reduce the average performance of applications. Migrating data to a single data warehouse makes it easier for users to perform SLA/QoS to extent, but it faces a number of limitations. The data warehouse unavailability it causes data to be unavailable to users if data is store in the same data warehouse. It considered one of the ten main obstacles to cloud adoption [2]. Dependency on one data warehouse makes difficult to moving data from the one storage to another in the face of rising cloud service provider prices, the emergence of a new data warehouse with a lower price, user mobility, and changing workloads that require data migration.

However, a single CSP have not been data centers at all location necessary for global web applications. In addition, the use of a single CSP can introduce a blocking problem for the data storage provider; in this a customer cannot be free to the changes optimal supplier due to the prohibitive exchange costs. In this storage provider it charges clients by bandwidth of data transfer, data requests for Gets / Puts and storage. The CSP provides various storage services such as Get and Put services help to geographically distributed data centers around the world. As with cloud computing, there are two main contributors, there are two aspects of cost optimization: cost optimization and suppliers cost optimization is carried out by users.

Now continuously increase the data volumes and details capturing by the use of organizations, like Internet of Things (IoT), multimedia, and the rise of social media, released overwhelming data flow in the unstructured/structured format. Data creation is the record speed, named here as a big data, and emerges as widely recognizing trend bigdata set mainly use three aspects such as a) Data stored on the cloud cannot be categorizing in the regular relational databases. b) Data is in numerous formats, and c) Data is captured, processed them rapidly and generated. In addition, Big-Data is now used much application such as Healthcare centers, Engineering, Science, Finance, and many businesses. Advance data storage techniques and mining technologies enable saving increasing amounts of data, changes in the nature of an organization's data. Fig. 1 shows that the cloud storage service provider's overview. Most enterprises shift the load to cloud storage to save capital costs for maintenance and hardware infrastructure and avoid the complexities of managing data centers.

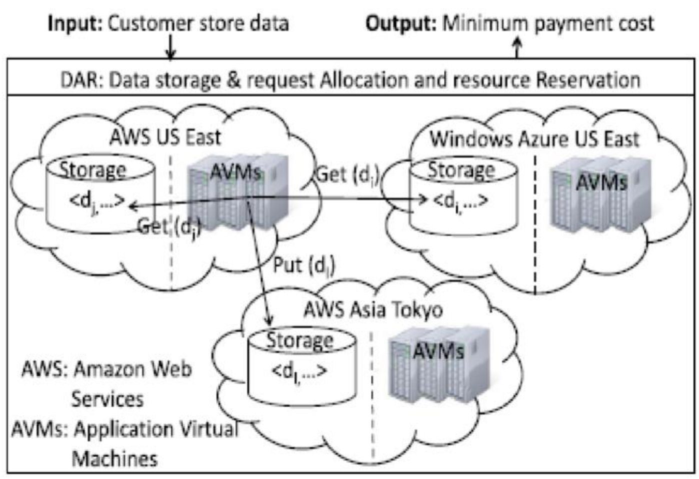

Figure 1. Shows the cloud storage system.

2. Literature Survey

In this section discuss the literature review in detail about the multiple cloud storage providers:

In this article [1], this is a key and value store that exports a consolidated view of the storage service to a geographically dispersed data center. Application first, SPANStore spans multiple cloud providers, increasing the geographic density of data centers, and taking advantage of price differences between them to minimize costs. Then, by evaluating the application workload at the appropriate level of detail, the large geodistributed replication required to achieve the latency target finally, SPANStore can be used to test the application workload at the appropriate level of detail. It is based on a two-phase lock required to provide a global view of the storage service and a computer for tasks such as data distribution.

To solve the configuration nightmare, they used MINERVA tools used to store data automatically. For data storage, MINERVA used declarative specifications of device capabilities and application requirements constraint-based formulations of the various sub-problems; and optimization techniques to explore the search space of possible solutions. This paper evaluated the design decisions that went into MINERVA, using specialized micro and macro-benchmarks [2].

Cloud services use Volley, sending the query logs of the data center. Volley analyzes logs use of iterative optimization technique based on customer locations, data access patterns, and provides recommendations. To scale to the volume of cloud service log data, Volley provides a scalable MapReduce scope-style platform that enables Volley to perform calculations of more than 400 machine hours in less than a day [3].

Cloud storage service provider provides the platform for storing large scale data on it with the minimum prize. Massive Parallel Processing (MPP) is using for bridges the gap between the modern cloud storage and traditional data warehouse. K. Liu, they implement Open Source Prototype GPCloud for Load/Upload data on the cloud storage this technique is based on cloud storage Amazon S3 and MPP data warehouse Greenplum [4]. An experimental result shows that the performance of this technique better than the existing system. This system is supported for INSERT and SELECT operation.

Resource provisioning to the computational task is the major challenge for cloud computing. Christoph Hoch Reiner and Stefan Schulte they propose prediction of cloud resources utilization on the pre-resource level and pre-task [5]. In this, they used machine learning techniques for prediction of resource utilization on the cloud storage. For this experiment, they used a dataset on the GitHub Travis CI and Travis CI. The performance of this system is compared with the simple learning regression approach and results show that the increase the accuracy of this system.

S. H. Gary Chan and Zhangyu Chan firstly study the problems of jointly optimizing resource allocation and video management for the large scale Video-on-Demand (VoD) cloud [6]. They propose RAVO (Resource Allocation and Video Management Optimization) model for jointly manage videos to achieve low cost and allocate system resources. For managing large video pool they used clustering algorithm. The performance of this system is compared with the other state-of-art techniques such as iGreedy, IPTV-RAM (internet protocol television-resources allocation and management) and super-optimum.

Cloud computing is a powerful technology for performing the operation such as complex and massive-scale computing. At this time the size of data increases and also verity of data generated and expanding every day. The use of a cloud service provider is to process, store and analyze data. Ibrar Yaqoob, Samee present classification techniques for big data, cloud service model and the conceptual view of data [7]. A review is conducting on the scalability, volume, data protection, availability, data transportation, data heterogeneity, regularity/ legal issues, governance and data access.

Zhiming Shen, Qin Jia, they present techniques such as “Supercloud”, it is deployed the use of resources from the several cloud service providers includes Rackspace, Amazon EC2, and HP Clouds [8]. Superclouds enable organizations, businesses, and individuals in cloud computing environments. In particular, cloud users manage the live location and migration of their storage, computing, and networking without owning the entire underlying infrastructure.

Boyang Wang, Jiqiang Liu propose privacy in frequent itemset mining on the encrypted cloud [9]. In this, they use three protocols such as Protocol 1 achieving higher mining performance and Protocol 2 providing the strong privacy guarantee and Protocol 3 for improved efficiency. Mining performance is achieved separate form protocol 1 and protocol 3. Performance of system compared with the association rule mining and they also used a chess database for this experiment, it has totally 3196 transactions and 74 possible attributes. Performance of comparison carried out into two different security levels: Database privacy and Item Privacy.

Miguel Correia, Alysson Bessani, they present DEEPSKY's methods for improving the privacy, integrity, and availability of information storing on the cloud and for this used encryption techniques for the data replica in a cloud. For achieving these objectives they used building Clouds of Clouds on the set of storage clouds and it also combines Cryptographic Secret Sharing with the Byzantine Quorum System Protocols [10].

Den Bossche, Jan Broeckhove develops techniques they also combine automated times series with the load prediction techniques which is based on the Double-Seasonal HoltWinter [11]. Load forecasting with automatic time-series forecasting based on Holt Winters provides cost-effective sourcing across a wide range of contract types, takes into the account the organization's current contract portfolio. This analysis explores the impact of various forecasting methods on the cost vs. perspective forecast, compares the performance of the algorithm against the fixed contract update approach, can significantly reduce the algorithm of the results of the system, the cost IaaS resources can be automatically procured contract.

Michael Borkowski, Christoph Hochreiner they are presenting techniques to the predicting use of the cloud resources at the task and resource level [12]. To do this, they use a machine learning prediction technique which is based on an extensive evaluation, they reduce the prediction error in 20% and improvements in this they achieved 89% accuracy.

S.H. Gary Chan, Zhangyu Chang they represent how to reduce the cost of deployment of optimizing video management like t hosting and searching video on the server and resource allocation in the clouds such as processing, linking, and storage it is subject to a specific user to defer the requirement to access the video [13]. First, the joint optimization problem was formulated, and it was shown as NP-hard. To solve this problem, they offer resource allocation and video management optimization (RAVO), which are based on linear programming with proven optimality gaps.

They proposed SCC (Storage Configuration Conclusion) to compile these inputs into cost-effective cluster configurations. This SCC technique is applied on different application workloads and storage options show that SCC captures enough detail to assign the right mix of storage and server hardware at the right scale; architecture change or decrease in magnitude leads to a significant decrease in performance. To meet application needs, SCC often predicts heterogeneous cluster architectures, resulting in significant cost savings compared to simply scaling homogeneous architectures [14].

This is primarily a departure from the current situation, which depends on human intervention. This technology works in key conditions to manage today's DCNs, given the increasing no of devices in DCNs and trend towards marketable equipment. NetPilot identifies the potential set of affected components that can cause the problem and repeats the mitigation steps for each until the problem is resolved. [15].

The presnt implementation of COPS is a key-value of store that provides models that is consistent across a wide area. It can provide cause and effect dependencies between keys stored across a cluster rather than a single server, as in previous systems. The core approach of COPS is to track and explicitly verify causal relationships between keys in the local cluster before providing a record, and the COPS-GT introduces a get transaction to get a consistent view of multiple keys without a lock or lock. As a result of the experiment, it was found that the cops can complete the operation in less than millisecond and provide the same throughput as the previous system, when a single server is used for each cluster, and can be extended as the case of increasing the number of servers in each cluster. This makes the COP-GT common workload for scaling, throughput, delay to other [16].

In this paper [17] they present the architecture of the storage system and that is by reducing the existing black box SSP Inter 9-network downtime for a given replication costs, and provide a durable storage system at low price. It is a traditional storage system with low availability and performance. To evaluate these ideas SafeStore based on the file system NFS interface.

This is a system that frees the cloud client from the burden of determining which service to use when deploying MapReduce compute to the cloud. The system also automatically selects the best cloud service to use, deploy the compute according to the selection, and adapt to changing conditions during deployment. The design of this guide includes a system for managing cloud computing through a variety of services and a system for providing a unified interface to these services.

In this, they propose CALMS (cloud-assisted live media streaming), a universal platform that facilitates migration to the cloud. Soothes adaptive leasing coaxing, according to the level of satisfaction of the temporal and spatial dynamics of demand from live streaming users, cloud server resources they also have different capacity and cost of rent, as well as the cost of rent. The actual cloud platform, the lease of the beginning and end of the potential delay of the cloud server in the work.

3. Proposed System

4.1. Problem Statement

Develop the techniques to minimize the payment cost under the aforementioned constraints using data compression technique with PPM (Prediction by Partial Matching) this helps to reduce storage cost and data transfer computing time.

4.2. Proposed System Overview

Here, present a novel cloud storage system for the request allocation, storage data on cloud; and reservation of resource across the multiples cloud storage service provides CSSPs in DAR. It is helps for customers to minimize their payment cost with the guaranteeing their SLOs. Here, can avoid vendor lock-in problems by building distributed cloud storage across multiple CSPs.

A data allocation technique based on the dominant cost, this dominant cost of each data element (storage, Get or Put) find the optimal resource reservation algorithm that is assigned to the data center at the lowest unit cost of this dominant cost, and pay-as-you-go payment method, such as payment reservation, while avoiding over booking.

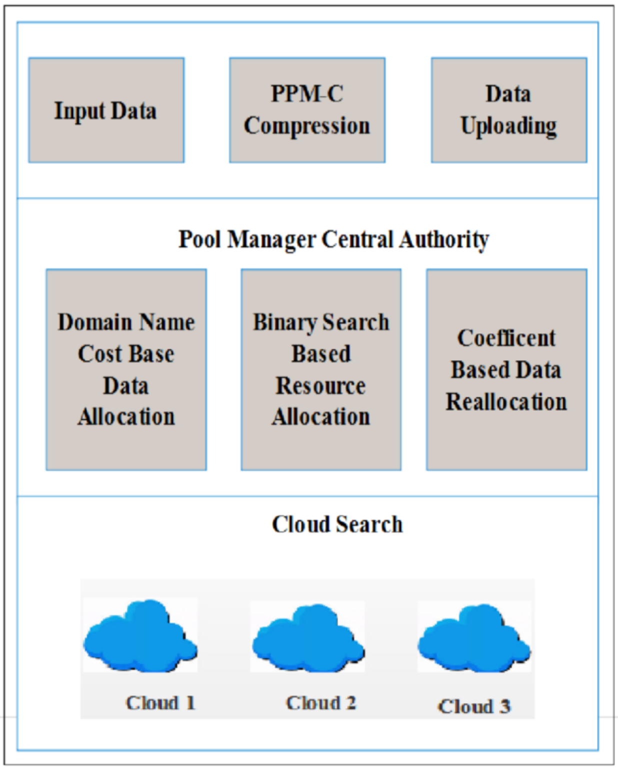

Figure 2. Proposed System Architecture

Reallocation of data based on coefficients, it aims to balance the workloads among all of the billing periods in order to minimize the cost of payment by maximizing the advantage of the reservation. Transfer of data in multicast, which builds a tree of minimum expansion to create new replicas of data in order to minimize the cost of Transfer for the creation of replicas on a new deployment of data mapping.

Request redirection congestion control it redirects Get requests from the overloaded datacenters to under loaded data centers that have received Gets more than (or less than) their expected number of Gets after data allocation to minimize the payment cost, respectively.

4.3 Algorithm

Algorithm: PPM-C (Prediction by Partial Matching - Cloud) Algorithm

```
  Begin;
    While (not last character) do;
      Begin;
        Readsymbol();
          Shorten Context;
      While (Context not found and context length -1) do;
     Begin;
      Output (escape sequence);
       Shorten context;
  End;
     Output (Character);
     While (Context length not -1) do;
  Begin;
    Increase count of character (Create a node of nonexistent)
     Shorten Context;
  End;
 End;
End;
```

Algorithm 2: Extract

- **PPM-C algorithm Compression working:**

In ppm needs only to store the context it has been occurs in the sequence it is encoded. In the beginning of the encoding needs to the encode letters that not occurs previously in context. The ppm use "ESCAPE SYMBOL" <ESC> is using to signal that letter to be encoded has not been seen in the context.

In the basic algorithm have to keep in mind the following three points:

1. If the symbol has not occurred into the context then an escape symbol is encoded.
2. Attempt to use of next smaller context i.e reduce the size of symbol one by one.
3. Each time the symbol is count corresponding to that entire symbol is updated in all tables.

Consider the phrase to be encoding examples ("this is the wonderfull day") of the given phrase have to make the encoding table until getting the higher order. For examples, higher order is 2 then making of the table from -1 order to 2nd order. To make the table need the fields like context, letter count and cum-count. Where;

The letter is the letter in the word phrase.

Context is the symbol (letter) which is present before that letter for -1 and 0 order table there is no such field context.

For 1st order: context is the symbol which comes before the corresponding letter i.e. in the above example of the phrase is the context't' for letter 'h'.

For 2nd order context= 'th' for the letter = i;

Count = it means the number of time that letter present in the word before <ESC> Cum-count = represents the addition of cum-count of previous letter and count of the current letter.

- Step1: in the initial state have to set the order of PPM model to 'n'.
- Step2: Here consider the order of ppm is n that means highest context order is 'n'. To achieve the highest order have to calculate the table of context order from -1 to n.

- step3: Once done with the calculation of context table calculate the Lower order value (l) and Higher-order value (u) by using the following equations:

$$L^n = L^{n-1} + [(u^{n-1} - l^{n-1} + 1) * cumcount]/total\ count$$

$$u^n = L^{n-1} + [(u^{n-1} - l^{n-1} + 1) * cumcount]/total\ count$$

- Step4: Then assume the word length of the arithmetic coder to the length of a word. In arithmetic, coding tag lies between [0, 1]. For example, if consider phrase is "this is" then word length is 6. That means the value of higher and lower order is 6 bit i.e (for example like 000010).

- Step5:
  I. For calculations of its letter consider 2nd order context table first and then determines whether the corresponding l context is encountered for that letter.
  II. If it encountered then take count and cum- count of that letter and find the value of lower
  III. Order and higher by using above equations and then check MSB of both values are same then transmit that MSB bit to the right-hand side and update the values of that letter in the 2nd context table.
  IV. Then move to 1st order context table. Note: if the context of the corresponding letter present then also move directly1st order context table.
  V. Similarly, calculate lower order value and higher order value update order c table if the context is encountered in the 0th order context table. It will repeat step 3, 4, and 5 until reach the end of the phrase.

3.4 Mathematical Model

- **Input:** Customer datacentres denoted by Dc

$$D_c = \{D^d_{c1}, D^d_{c2}, \ldots, D^d_{cn}\} \quad (1)$$

- **Process I:**

I = Cost Calculation

It is the sum of the total Transfer, Storage, Get, Cost and Put in the reservation time T:

$$C_T = \{C_c + C_s + C_g + C_p\} \quad (2)$$

The cost storage in the data center is product of the size of data and unit storage price of each data center. Then, total storage cost is calculated by:

$$C_s = \Sigma_{t_k \in T} \Sigma_{d_t \in D} \Sigma_{P_j \in D_s} X^{d_t,t_k}_{P_j} * P^s_{P_j} * S_{d_1} \quad (3)$$

Where S_{dq} denotes the size of data,

d_t *and* $P^s_{P_j}$ Denote the unit stor age price of datacentres P_j and denotes the binary variable: it is 1 if d_t is store in P_j during t_k; and 0 otherwise.

$$C_C = \Sigma_{t_k \in T} \Sigma_{d_t \in D} \Sigma_{P_j \in D_s} \Theta^{d_t} * P^t_{P_j} * S_{d_1} \quad (4)$$

Where, $P^t_{P_j}$ is the replicating cheapest unit transfer price d_t *to* P_j among all datacentres storing d_t.

$$C_g = \Sigma_{t_k} \Sigma_{p_j} \left(Max\left\{\Sigma_{c_i} r^{tk}_{c_i,p_j} * t_k - R^g_{p_j}, 0\right\} + \alpha R^g_{p_j}\right) * P^g_{p_j} \quad (5)$$

$X^{d_t,t_k}_{P_j}$

$$C_p = \Sigma\Sigma\left(Max\left\{\Sigma \omega^{t_k}_{c_i p_j} * t_k - R^p_{p_j}, 0\right\} + \alpha R^p_{p_j}\right) * P^p_{p_j} \quad (6)$$

The Get/Put billing is based on reservation and Pay-As-You-Go. The reserving numbers of Gets/Puts and is decided in beginning of the reservation $R^g_{p_j}$ time period $R^p_{p_j}$ T. The prices of reservation for the Gets and Puts are specific percentage of their unit prices in the Pay-As-You-Go manner [7].

- **Process II:** Formulating the SLO objective, its need calculating actual percentage of Gets/Puts which satisfy the latency requirement in the billing period tk

$$Q^g * Q^p = 1 \quad (7)$$

$Q^g * Q^p = 1$ The customer's deadline SLO is satisfied.

$$\forall_{c_i} \forall_{t_k} \forall_{d_l} \in G_{c_i}^{t_k} \sum_{p_j \in s_{c_i}^g} X_{p_j}^{d_l,t_k} \geq \beta \quad (8)$$

Each customer's datacentre maintains a table and it maps each data item to be replica datacentres it assigns to ratios of request serving.

$$\forall_{p_j} Min\left\{\phi_{p_j}^g, \phi_{p_j}^p\right\} \geq 0 \quad (9)$$

$$\forall_{c_i} \forall_{p_j} \forall_{d_l} H_{c_l p_j}^{d_l,t_k} \leq X_{p_j}^{d_l,t_k} \geq 1 \quad (10)$$

$$\forall_{c_i} \forall_{t_k} \forall_{d_l} \sum_{p_j} H_{c_l p_j}^{d_l,t_k} = 1 \quad (11)$$

Constraints (7), (8) and (9) satisfy the deadline requirement and data availability requirement in the SLO and the data centre capacity constraint, as explained previously. Constraints (10) and (11) together indicate that any request should be served by a replica of the targeted data.

- **Output:**

Data allocations and data storage Ds

$D_s = \{D_{s1}, D_{s2}, \dots \dots \dots, D_{sn}\}$

4. Result and Discussion

4.1 Experimental Setup

For this work, the required technologies:

- **Software Technology:**
  1. Operating System: Windows 7
  2. Coding Language: Java/J2EE
  3. Tools: Net Beans 7.2.1
- **Hardware Technology:**
  1. System: Pentium dual core
  2. Hard Disc: 120 GB.
  3. RAM: 1 GB.
  4. Input Device: Keyboard, Mouse

4.2. Result and Analysis A. Result Analysis:

Table I show that data effectiveness and efficiency of SPAN-Store, DAR-C and PPM- C system. The data store cost of PPM-C is less than the SPAN-Store, DAR-C.

Table 1. Effectiveness and Efficiency of Cost Reduction

	Dat a compression in % Vs Data Size			
DAR	100	100	100	100
DAR-PPMC	55	60	60	55

Figure 4 shows the graph of effectiveness and efficiency DAR and DAR-PPMC system. The result of the graph shows the data compression in % of original data as found in stimulation observed that original data size is reduced up to approximately 44.3% of original data.

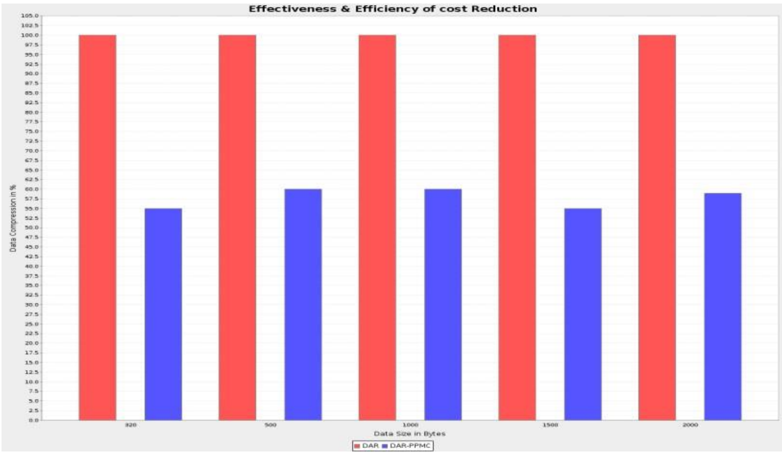

Figure 3. Effectiveness and Efficiency of Data Size Reduction

Table II shows that computing time required for data allocation of SPAN-Store, DAR-C, and PPM-C system. The PPM-C required computational times are less than the SPANStore, DAR-C.

Table 2. Computing Time Vs Data

	Computing Time Vs Data Siz e			
DAR	65	75	85	95
DAR-PPMC	60	70	80	90

Below figure 4 shows that computing time comparison graph between DAR and DARPPM-C. As observed data size of original data is reducing approximately 40-44.3% of original data size by using PPM compression method with DAR. Hence while performing Get/Put operation if the size of data is less than why it requires less computing time for Get/Put operation.

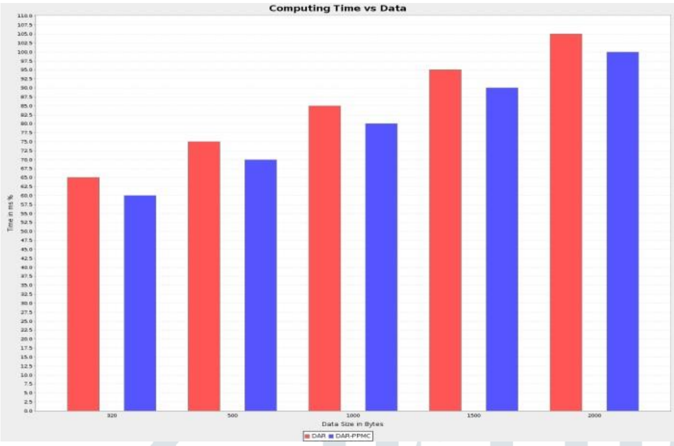

Figure 4. Graph of Computing Time Vs Data in Bytes

Below figure 5 shows that the cost ratio of DAR and DAR-PPMC. DAR compared with and without PPM-c Compression method v/s of data size. This is because the PPM-C compression method reduces the data size to approximately 40- 44.3 % of the original data size. So that computing cost of data will be reduced.

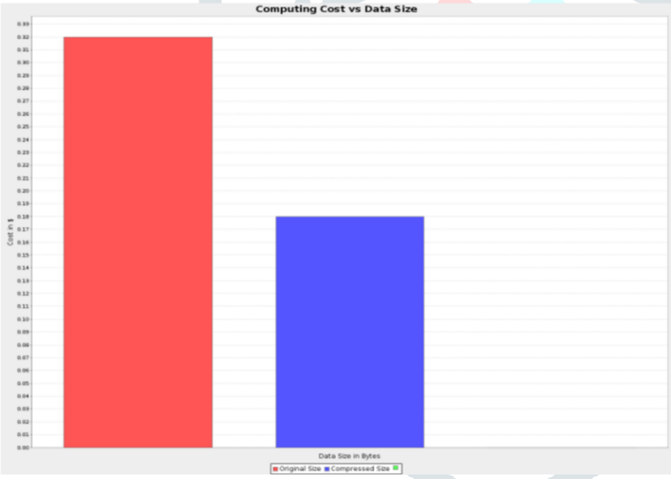

Figure 5. Computing Time vs Data

5. Conclusion

The aim of this paper is to minimize the cost of the CSSPs providers with the guarantee of services used worldwide distributed data centers, which belong to the different CSPs with the different resource unit prices. Here, firstly used an integer programming for cost minimizing problems. It is NP-hard problem; that why here introduce request allocation, data storage and resource reservation across the multiple CSSPs (DAR) techniques for the Dominant Cost Based Data Allocation Algorithm is used to reduce the cost of each datacenter. Reducing service latency and payment cost datacenters used three methods 1. Multicast-Based Data Transfer, 2. Coefficient based Data Allocation and 3. Request Redirection based Congestion Control. The PPM data compression technique it helps to reduce storage cost and data transfer computing time.

References

6.1. Journal Article

[1] Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha : SPANSTORE: COSTEFFECTIVE GEO-REPLICATED STORAGE SPANNING MULTIPLE CLOUD SERVICES. in Proc. SOSP, pp. 292-308, **[2013].**

[2] G. A. Alvarez et al. : MINERVA: AN AUTOMATED RESOURCE PROVISIONING TOOL FOR LARGE SCALE STORAGE SYSTEMS, ACM Transaction on Computer System, vol. 19, no. 4, pp. 483-518**, [2001].**

[3] S. Agarwal et al.,: VOLLEY: AUTOMATED DATA PLACEMENT FOR GEO-DİSTRİBUTE CLOUD SERVİCES, in Proc. NSDI, p. 28, **[2010].**

[4] Y. Yao, Haozhou Wang, and K. Liu : ON STORING AND RETRIVEING GEOSPATIOAL BIGDATA IN CLOUD, SIGSPTIAL International workshop of use of GIS in the emergency management, ACM, **[2016].**

[5] Christoph Hoch Reiner, Stefan Schulte and Michael Borkowski, : PREDICTING CLOUD RESOURCE UTILIZATION, 9th ICUCC-International Conference on the Utility and Cloud Computing**; [2016].**

[6] S. H. Gary Chan and Zhangyu Chang : VIDEO MANAGEMENT AN RESOURCES ALLOCATION FOR A LARGE SCALE VoD CLOUD, Nature Communication, 24 January 2016 ACM Transaction on Multimedia Computing Communication Application, Vol. 12, Article 72, [**2016].**

[7] Ibrar Yaqoob, Samee Ullah Khan, S. Mokhtar and A. Gani : THE RISE OF BIG DATA ON CLOUD COMPUTING REVIEW AND OPEN RESEARCH ISSUES, Elsevier-**[2015].**

[8] Zhiming Shen, Qin Jia, Weijia Song : SUPERCLOUD OPPORTUNITIES AND CHALLENGES, SIGOPS Oper. Syst. Rev**, [2015].**

[9] Boyang Wang, Jiqiang Liu, Ming Li, and Shuo Qiu : TOWRODS PRACTICAL PRIVACY PRESERVING FREQUENT ITEMSET MINING ON ENCRYPTED CLOUD DATA, IEEE Transactions on Cloud Computing, [**2017].**

[10] Miguel Correia, Alysson Bessani, and Bruno Quaresma: DepSky DEPENDABLE AND SECURE STORAGE IN A CLOUD OF CLOUDS, ACM Transaction Storage, **[2017].**

[11] Den Bossche, Jan Broeckhove, and Kurt Vanmechelen: OPTIMIZING IAAS RESERVED CONTRACT PROCURRENT USING LOAD PREDICTION, 7th International Conference on Cloud Computing (ICCC), IEEE, **[2014].**

[12] Michael Borkowski, Christoph Hochreiner and Stefan Schulte: PREDICTING CLOUD RESOURCES UTILIZATION, 9th International Conference on Utility and Cloud, IEEE/ACM**, [2016].**

[13] S.H. Gary Chan, Zhangyu Chang, : VIDEO MANGEMENT AND RESOURCES ALLOCATION FOR A LARGE SCALE VOD CLOUD;. Transactions on Multimedia Computing, Communications, and Applicant, **[2013].**

[14] Harsha V. Madhyastha, and John C. McCullough: SCC: CLUSTER STORAGE PROVİSİONİNG INFORMED BY APPLİCATİON CHARACTERİSTİCS AND SLAS; FAST, p. 23, **[2012].**

[15] Xin Wu, Daniel Turner, and Lihua Yuan: NETPİLOT: AUTOMATİNG DATACENTER NETWORK FAİLURE MİTİGATİON, in Proc. SIGCOMM, pp. 419–430, **[2012].**

[16] Wyatt Lloyd, Michael J. Freedman, and Michael Kaminsky: DON'T SETTLE FOR EVENTUAL: SCALABLE CAUSAL CONSİSTENCY FOR WİDE-AREA STORAGE WİTH COPS, 23rd Symposium on Operating Systems Principles (SOSP'11)-ACM, **[2011].**

[17] Ramakrishna Kotla, and Mike Dahlin: SAFESTORE: A DURABLE AND PRACTİCAL STORAGE SYSTEM, ATC, pp. 129–142, **[2007].**

[18] A. Wieder, P. Bhatotia, A. Post, and R. Rodrigues, "ORCHESTRATİNG THE DEPLOYMENT OF COMPUTATİONS İN THE CLOUD WİTH CONDUCTOR," in Proc. NSDI, pp. 367–381**, [2012].**

[19] F. Wang, and M. Chen, "CALMS: CLOUD-ASSİSTED LİVE MEDİA STREAMİNG FOR GLOBALİZED DEMANDS WİTH TİME/REGİON DİVERSİTİES," in Proc. INFOCOM, pp. 199–207, **[2012].**