Effective Resource Utilization for Caching as a Service in Cloud

Samir S. Rathod, G. Niranjana

Abstract— with the growing popularity of cloud based data centers as the enterprise IT platform of choice, there is a need for effective management strategies capable of maintaining performance. Caching technology improves the performance of the cloud. Cache as a service (CaaS) model is an additional service to Infrastructure as a service (IaaS). The cloud server process introduce, pricing model together with the elastic cache system. This will increase the disk I/O performance of the IaaS, and it will reduce the usage of the physical machines. The emerging cloud applications provide data management services allowing the user to query the cloud data, paying the price for the infrastructure they use. Cloud management necessitate an economy that manages the services of multiple users in an efficient, but also, resource economic way that allows for cloud profit. The cloud caching service can maximize its profit using an optimal pricing scheme. This scheme requires an appropriately simplified price-demand model that incorporates the correlations of structures in the cache services. The pricing scheme should be adaptable to time changes. This paper proposes a novel price demand model designed for a cloud cache and a dynamic pricing scheme for queries executed in the cloud cache. This will estimates the correlations of the cache services in a time-efficient manner and improve the efficiency of resources in cloud storage infrastructure to deliver scalable service.

Index Terms— Cloud Computing, CaaS model, Virtual Machine, Remote Memory, Optimal pricing scheme.

I. INTRODUCTION

Cloud Computing plays an important role in today's business organizations in order to make the business effective. Cloud is a picture that describes the entire infrastructure we need to utilize in order to move information from one location point to another. It is not only represent the data access from distance computer but actually it also consists of big servers that run applications and might have a huge data storage place [1]. If some high end calculations need to be done cpu utilization of computer is not much but in cloud there is whole clusters of computers that can be used known as utilization computing or cpu utilization. At current stage, the cloud computing is still evolving and there exist no widely accepted definition. Based on our experience, we propose an early definition of Cloud computing as follows: A cloud computing is a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing infrastructures on demand, which can be access typically through a web based interface.

Manuscript received on April, 2014.

Mr. Samir .S. Rathod, Department of Computer Science and Engineering, SRM University, Chennai, India.

Asst Prof. G .Niranjana, Department of Computer Science and Engineering, SRM University, Chennai, India.

Cloud services as virtualized entities are essentially an elastic providing an unlimited resource capacity. This elasticity with utility computing brings efficiency which is the primary driving force behind the cloud. In this paper we investigate how efficiency in the cloud can be further improved. Here, caching plays an important role in improving the performance.

Caching is used in storage systems to provide fast access too recently or frequently accessed data, with non-volatile devices used for data safety and long-term storage. Much research has focused on increasing the performance of caches as a means of improving system performance. In many storage system configurations, client and server caches form a two-layer (or more) hierarchy, introducing new challenges and opportunities over traditional single-level cache management. These include determining which level to cache data in and how to achieve exclusivity of data storage among the cache levels given the scant information available in all but the highest-level cache. Addressing these challenges can significantly improve overall system performance.

The following sections describe the cloud computing model and motivation for cache management. Section II details the related work of the proposed system. In section III Explain the concept of Cache as a Service model. Section IV shows the use of Virtualization concept. In Section V proposed work is discussed with pricing scheme of cache with respect to change in time followed by modeling structure correlation. Section VI states the architecture of proposed work with various measures for calculating the cost from user side and performance from server side.

A. Cloud computing model

The Fig.1 shows the different types of services provided by the cloud service providers.

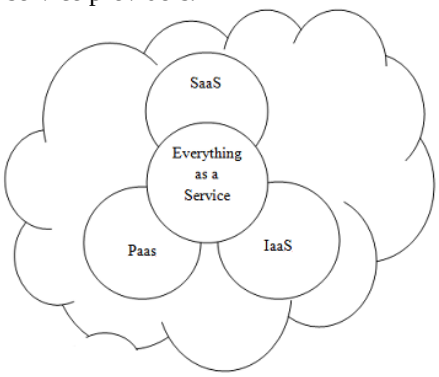

Fig.1 Cloud model

i. *Software as a Service (SaaS):* The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser, or a program interface.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Eg: Google docs etc.

ii. *Platform as a Service (PaaS):* The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider.

iii. *Infrastructure as a service (IaaS):* The capability provided to the consumer is to provide processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

Eg: Amazon S3 storage.

B. Incentive for cache management

Over the past decades, caching has become the key technology in bridging the performance gap across memory hierarchies via temporal or spatial localities; in particular, the effect is prominent in disk storage systems. Currently, the effective use of cache for I/O-intensive applications in the cloud is limited for both architectural and practical reasons. Due to essentially the shared nature of some resources like disks, the virtualization overhead with these resources is not negligible and it further worsens the disk I/O performance. Thus, low disk I/O performance is one of the major challenges encountered by most infrastructure services as in Amazon's relational database service, which provisions virtual servers with database servers. At present, the performance issue of I/O intensive applications is mainly dealt with by using high performance (HP) servers with large amount of memory, leaving it as the user's responsibility.

II. RELATED WORK

Existing clouds focus on the provision of web services targeted to developers, such as Amazon Elastic Compute Cloud (EC2), or the deployment of servers. There are two major contests when trying to define an optimal pricing scheme for the cloud caching service. The first is to outline a simplified enough model of the price demand dependency, to achieve a feasible pricing solution.

A static pricing scheme cannot be optimal if the demand for services has deterministic seasonal fluctuations. The second challenge is to define a pricing scheme that is adaptable to Modeling errors, time-dependent model changes, and random behavior of the application. The demand for services, for instance, may depend in a unpredictable way that are external to the cloud application, such as involving a combination of social and economic factors.

Static pricing cannot guarantee cloud profit maximization its result is unpredictable and, therefore, uncontrollable behavior of profit. Closely related to cloud computing is research on accounting in wide-area networks that offer distributed services, an economy for querying in distributed databases is researched. This economy is limited to offering budget options to the users, and does not propose any pricing scheme.

III. CACHE AS A SERVICE - A CONCEPT

The CaaS model consists of two main components: an elastic cache system as the architectural foundation and a service model with a pricing scheme as the economic foundation. The basic system architecture for the elastic cache aims to use RM, which is exported from dedicated memory servers. It is not a new caching algorithm. The elastic cache system can use any of the existing cache replacement algorithms. Near uniform access time to RM-based cache is guaranteed by a modern high speed network interface that supports RDMA as primitive operations. Each VM in the cloud accesses the RM servers via the access interface that is implemented and recognized as a normal block device driver. Based on this access layer, VMs utilize RM to provision a necessary amount of cache memory on demand [7].

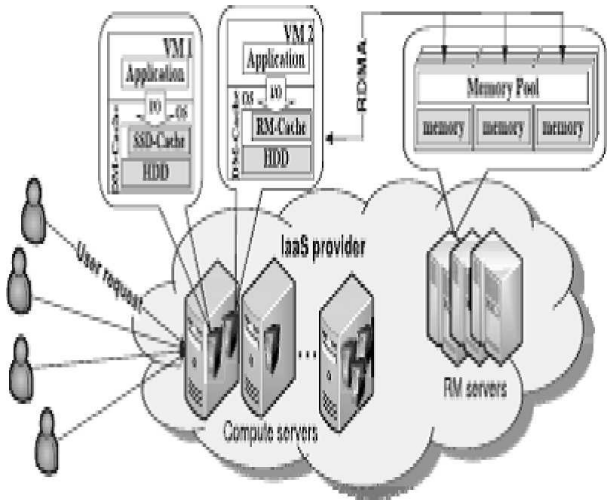

Fig.2 Overview of CaaS

As shown in Fig. 2, a group of dedicated memory servers exports their local memory to VMs, and exported memory space can be viewed as an available memory pool. This memory pool is used as an elastic cache for VMs in the cloud. For billing purposes, cloud service providers could employ a lease mechanism to manage the RM pool.

To employ the elastic cache system for the cloud, service components are essential. The CaaS model consists of two cache service types based on whether LM or RM is allocated with. Since these types are different in their performance and costs a pricing scheme that incorporates these characteristics is devised as part of CaaS.

Together, we consider the following scenario. The service provider sets up a dedicated cache system with a large pool of memory and provides cache services as an additional service to IaaS. Now, users have an option to choose a cache service specifying their cache requirement and that cache service is charged per unit cache size per time. Specifically, the user first selects an IaaS type as a base service. The user then estimates the performance benefit of additional cache to her application taking into account the extra cost, and determines an appropriate cache size based on that estimation. We assume that the user is at least aware whether her application is I/O intensive and aware roughly how much data it deals with. The additional cache in our study can be provided either from the local memory of the physical machine on which the base service resides or from the remote memory of dedicated cache servers. The former LM case can be handled simply by configuring the memory of the base service to be the default memory size plus the additional cache size.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

The cost benefit of our CaaS model is twofold: profit maximization and performance improvement. Clearly, the former is the main objective of service provider. The latter also contributes to achieving such an objective by reducing the number of active physical machines. From the user's perspective, the performance improvement of application can be obtained with CaaS in a much more cost efficient manner since caching capacity is more important than processing power for those applications.

IV. CACHE MANAGEMENT WITH VIRTUALIZATION

Virtualization leads to Cloud Computing, but cloud computing is more than just virtualization Cloud computing (Public or Private) produce virtual servers or applications, virtualization does not create clouds. Virtualization is a mechanism used in a data center to consolidate and fully use physical devices. For example, a single server for each processing application performed then would be make the proposition very expensive. The CPU of each machine would usually not be fully utilized; resulting in wasted resources. In addition the size of data centers and the growth would be difficult, at best to manage. Virtualization can place several virtual servers on a single physical device, minimizing the need for physical devices; thus lowering cost in processing resources, space in the data center, environmental conditioners and overall power usage. Virtualization allows for consolidation and cost savings. The Fig.3 Shows this concepts

Fig.3 Virtual Machine

The Cache management methods are used to maintain the shared data in transmission process. Data values are maintained under RAM and disk cache environments. Dynamic pricing scheme is used for cache storages. Cache assignment is performed with location details.

V. PROPOSED WORK

The cloud caching service can maximize its profit using an optimal pricing scheme. Optimal pricing necessitates an appropriately simplified price-demand model i.e. Demand replacement Algorithm (DRA) that incorporates the correlations of structures in the cache services. The pricing scheme should be adaptable to time changes.

A. Adaptive price change

Profit maximization is pursued in a finite long- term horizon. The horizon includes sequential non- overlapping intervals that allow for scheduling structure availability. At the beginning of each interval, the cloud redefines availability by taking offline some of the currently available structures and taking online some of the unavailable ones. Pricing optimization proceeds in iterations on a sliding time-window that allows online corrections on the predicted demand, via re-injection of the real demand values at each sliding instant. Also, the iterative optimization allows for re-definition of the parameters in the price-demand model, if the demand deviates substantially from the predicted.

B. Modeling structure correlation

This approach models the correlation of cache structures as a dependency of the demand for each structure on the price of every available one. Pairs of structures are characterized as competitive, if they tend to exclude each other, or collaborating, if they coexist in query plans. Competitive pairs induce negative, whereas collaborating pairs induce positive correlation. Otherwise correlation is set to zero. The index- index, index column, and column-column correlations are estimated based on proposed measures that can estimate all three types of correlation. We propose a method for the efficient computation of structure correlation by extending a cache based query cost estimation module and a template-based workload compression technique.

C. Demand Replacement Algorithm

In our processing we have three level of memory they are remote memory (RM), cache memory (CM) and hard disk memory. If you providing data into server i.e. CM and RM. maintaining data in cache server as for DRA.
Steps involved in DRA are mentioned as follows:

Step 1: **Global**: cache structures *S*, prices *P*, availability Δ
Step 2: Demand Analysis ( )
Get the demand for request
Improve the demand value as per request
Step 3: Demand Replacement ( )
While require size>free size
Get lowest demand factor
Analyze equalization of demand factor
If equal then
Consider the size of the equal demanded files
If equal size then
Consider the date of the equal size files
End if
End if
End while
Step 4: Query Execution ( )
If *q* can be satisfied in the cache then
(Result, cost)←runQueryInCache (q)
Demand Analysis ( )
Else
(Result, cost)←runQueryInBackend (q)
Demand Replacement ( )
End if
S←addNewStructures ()
Return result, cost
Step 5: Optimal Pricing (horizon *T*, intervals *t*[*i*], *S*)
(Δ, *P*)←determineAvailability&Prices (T, t, S)
Return Δ, *P*
Step 6: Main ()
Execute in parallel tasks T1 and T2:
T1:
For every new *i* do
Slide the optimization window

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Optimal Pricing (*T*, *t*[*i*], *S*)
End for
T2:
While new query *q* do
(Result, cost)←query Execution (q)
End while
If *q* executed in cache then
Charge *cost* to user
Else
Calculate total price and charge price to user
End if

VI. ARCHITECTURE

A. *Framework infrastructure for user module*

In this module, the web page for users will be designed, who access the cloud services on the cloud. In the web page, the pricing details, login such as services are included. The user is registered user will give user name and password and then login to the cloud services. If the user is new, he/she will registered on the cloud services, and the username and password provide to them, after that they can login to the cloud services.

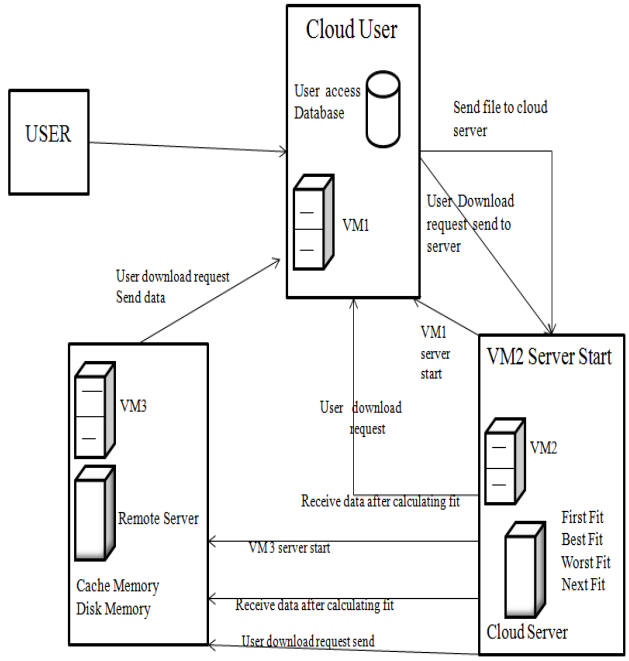

Fig.4 Design Model

B. *Deployment of cloud infrastructure*

After user login into the cloud services, cloud provider will provide access of the services to the user. The service provider give the file upload option to the user, the file can be browse by the user and send the file to the cloud server. User upload the file to the server, before that, the user will check the server is started or not. After the server is running, the user will upload their files to the server.

The cloud cache is a full-fledged DBMS along with a cache of data that reside permanently in back-end databases. The goal of the cloud cache is to offer cheap efficient multi-user querying on the back-end data, while keeping the cloud provider profitable. Service of queries is performed by executing them either in the cloud cache or in the back-end database. Query performance is measured in terms of execution time.

C. *Job Scheduling*

In the job scheduling, the file has been received in the cloud which the user uploaded to the server. The cloud server checks the performance of the cloud virtual servers; the performance value is the processing time of the server and also the system performance. If the performance of the server is idle the cloud server sends the file to the idle virtual server for the particular process.

The faster the execution, the more data structures it employs, and therefore, the more expensive the service. We assume that the cloud infrastructure provides sufficient amount of storage space for a large number of cache structures. Each cache structure has a building and a maintenance cost.

D. *Simulate Infrastructure*

We create the simulation environment as the connection of multiple Virtual Machine warehouse (VMWARE) to create the cloud environment for sharing of uploaded file through the connection of ip address.

E. *RM-Performance*

The file process accomplished, it will be stored on the remote memory through Remote Direct Memory Access interface. If the user wants to download the files from cloud he has to select the file and download it. It will retrieve from the remote memory and database. The users can query the cloud data, paying the price for the infrastructure they use. Cloud management necessitates an economy that manages the service of multiple users in an efficient, but also, resource economic way that allows for cloud profit.

Naturally, the maximization of cloud profit given some guarantees for user satisfaction presumes an appropriate price-demand model that enables optimal pricing of query services. The model should be plausible in that it reflects the correlation of cache structures involved in the queries.

F. *Pricing analysis*

The analysis of pricing is the rate of the services compared to each service for example, RM-cache, SSD-cache, no-cache. The values of each services and the graph will be plot against, it shows the performance of each service rates. Since sorting is the most important step in building an index, the cost of building an index is approximated to the cost of sorting the indexed columns. In case of multiple cloud databases, the cost of data movement is incorporated in the building cost. The maintenance cost of a column or an index is just the cost of using disk space in the cloud. Hence, building a column or an index in the cache has a one-time static cost, whereas their maintenance yields a storage cost that is linear with time.

VII. CONCLUSION

Caching technology improves the performance of the cloud. Here, caching plays a crucial role in improving their performance. This work proposes a novel pricing scheme designed for a cloud cache that offers querying services and aims at the maximization of the cloud profit. We define an appropriate Price-demand model (DRA) and formulate the optimal pricing problem. It ensures that with the proposal of a method that estimates the correlations of the cache services in a time-efficient manner will improve the

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

efficiency of the resources in cloud storage infrastructure.

ACKNOWLEDGMENT

I am thankful and owe my sincere gratitude to my project guide Mrs.G.Niranjana along with the entire department of Computer Science & Engineering of SRM University for their overwhelm support to carry out this project successfully.

REFERENCES

[1] X. Zhang and Y. Dong, "*Optimizing Xen VMM Based on Intel Virtualization Technology,"* Proc. IEEE Int'l Conf. Internet Computing in Science and Eng. (ICICSE '08), 2008

[2] H. Kim, H. Jo, and J. Lee, "*XHive: Efficient Cooperative Caching for Virtual Machines, "* IEEE Tran s. Computers, vol. 60, no. 1, Jan. 2011

[3] G. Jung, M. A. Hiltunen, K. R. Joshi, "*Amazon Elastic Compute Cloud"*, Amazon Web Services, 2010.

[4] Verena Kantere , Debabrata Dash , Gregory Francois , Sofia Kyriakopoulou, Anastasia Ailamaki Ecole Polytechnique F'ed'erale de Lausanne, *"Optimal service pricing for a cloud cache"*, vol. 23, no. 9, pp. 1345-1358, 2011.

[5] M.D. Dahlin, R.Y. Wang, T.E. Anderson, and D.A. Patterson, "Cooperative Caching: Using Remote Client Memory to Improve File System Performance," Proc. First USENIX Conf. Operating Systems Design and Implementation (OSDI '94), 1994.

[6] A. Menon, A.L. Cox, and W. Zwaenepoel, "Optimizing Network Virtualization in Xen," Proc. Ann. Conf. USENIX Ann. Technical Conf. (ATC `06), 2006.

[7] Hyuck Han, Young Choon Lee, Woong Shin, Hyungsoo Jung, Heon Y. Yeom and Albert Y. Zomaya, "*Cashing in on the Cache in the Cloud"*, IEEE Transactions on Parallel and Distributed Systems, Vol. 23, no. 8, August 2012.

[8] C. Park, P. Talawar, D. Won, M. Jung, J. Im, S. Kim, and Y. Choi, "A High Performance Controller for NAND Flash-Based Solid State Disk (NSSD)," Proc. IEEE Non-Volatile Semiconductor Memory Workshop (NVSMW `06), 2006.

[9] J. Ousterhout, P. Agrawal, D. Erickson, C. Kozyrakis, J. Leverich, D. Mazie'res, S. Mitra, A. Narayanan, G. Parulkar, M. Rosenblum, S.M. Rumble, E. Stratmann, and R. Stutsman, "The *Case for RAMClouds: Scalable High-Performance Storage Entirely in DRAM/"* ACM SIGOPS Operating Systems Rev., vol. 43, pp. 92-105, Jan. 2010

[10] C.A. Waldspurger, "Memory Resource Management in VMware ESX Server," Proc. Fifth USENIX Conf. Operating System Design and Implementation (OSDI `02), 2002.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.