Optimization of Cloud Data Storage

Priti Kandewar
Assit. Professor in Computer Science & Engineering
Christu Jyothi Institute Of Technology & Science
Jangaon, Telangana-506167
pritikandewar09@gmail.com

Kurella Mamatha
Assit. Professor in Computer Science & Engineering
Christu Jyothi Institute Of Technology & Science
Jangaon, Telangana-506167
mamatha7k@gmail.com

Abstract-Cloud computing is a paradigm of distributed computing to provide the customers on-demand, utility based computing services. Cloud users can provide more reliable, available and updated services to their clients in turn.Cloud storage needs to look seamless to users, consistent in access through redundancy and distribution of resources, and consistent in version control.Data optimization is probably the most significant recent software innovation in storage because you can save more information in a smaller physical space.In this paper we present the optimization of cloud data storage.

Keywords: cloud storage, optimization, interoperability.

I. INTRODUCTION

Cloud computing is a paradigm of distributed computing to provide the customerson-demand, utility based computing services. Cloud users can provide more reliable,available and updated services to their clients in turn. Cloud itself consistsof physical machines in the data centers of cloud providers. Virtualization is providedon top of these physical machines. These virtual machines are provided to thecloud users. Different cloud provider provides cloud services of different abstractionlevel. E.g. Amazon EC2 enables the users to handle very low level details whereGoogle App-Engine provides a development platform for the developers to developtheir applications. So the cloud services are divided into many types like Softwareas a Service, Platform as a Service or Infrastructure as a Service. These servicesare available over the Internet in the whole world where the cloud acts as the singlepoint of access for serving all customers. Cloud computing architecture addressesdifficulties of large scale data processing. Types of clouds are as follows.

- *Private Cloud:* This type of cloud is maintained within an organization andused solely for their internal purpose. So the utility model is not a big termin this scenario. Many companies are moving towards this setting and expertsconsider this is the 1st step for an organization to move into cloud. Security,network bandwidth are not critical issues for private cloud.
- *Public Cloud:* In this type an organization rents cloud services from cloudproviders on-demand basis. Services provided to the users using utility computingmodel.
- *Hybrid Cloud:* This type of cloud is composed of multiple internal or externalcloud. This is the scenario when an organization moves to public cloudcomputing domain from its internal private cloud.

A) View Points

1. Cloud Providers' Point Of View

- Most of the data centers today are underutilized. They are mostly 15%utilized. These data centers need spare capacity just to cope with thehuge spikes that sometimes get in the server usage. Large companieshaving those data centers can easily rent those computing power to other
- Organizations and get profit out of it and also make the resources neededfor running data center (like power) utilized properly.
- Companies having large data centers have already deployed the resourcesand to provide cloud services they would need very little investment andthe cost would be incremental.

2. Cloud Users' Point Of View

- Cloud users need not to take care about the hardware and software theyuse and also they don't have to be worried about maintenance. The usersare no longer tied to someone traditional system.
- Virtualization technology gives the illusion to the users that they arehaving all the resources available.
- Cloud users can use the resources on demand basis and pay as much asthey use. So the users can plan well for reducing their usage to minimizetheir expenditure.
- Scalability is one of the major advantages to cloud users. Scalability isprovided dynamically to the users. Users get as much resources as theyneed. Thus this model perfectly _ts in the management of rare spikes inthe demand.

B) Interoperability and Quality of Service

As cloud computing becomes popular and competition between CSPsincrease, the customers should have the freedom to switch betweenservice providers for ensuring sectoral growth. The customers shouldnot be locked on a single cloud provider[1]. This would give customersthe freedom to switch providers as their computing needs grow orshrink, and the ability to build more complex business applications tooptimize their business requirements.

The ability to have multiple machines sharing resources on a singlephysical server and the ability to treat a machine as a file opens upnew possibilities. Virtualization is a first step toward movingapplications to the cloud. While, the lack of standards isn't stoppingcustomers from moving to the cloud, it is likely to slow them down.Interoperability and standardization is a major concern after securitytowards faster adaptation of cloud.

Interoperability is essentially the ability to communicate acrossdifferent systems. It requires that the communicated information isunderstood by the receiving system. In cloud computing, it means theability to write code that works with more than one cloud serviceprovider simultaneously, regardless of differences between theproviders.

Portability is the ability to run components or systemwritten for one environment in another environment; this includesboth software and hardware requirements.

Need for interoperability: Every cloud Service provider creates its ownprocesses for a user or application interaction with the cloud leadingto cloud Application Programming Interface (API) propagation. It leadsto the issues such as vendor lock-in, portability and inflexibility to usemultiple vendors in the cloud including the inability to use anorganization's own data center resources seamlessly[6]. There is a needfor consistent data handling and predictable performance acrossdisparate cloud providers within a cloud ecosystem enabling hybridcloud. Such scenarios give rise to interoperability issues at business,security and functional interfaces [2].

II. CLOUD STORAGE

Enterprise-level cloud data storage is built on a virtualized infrastructure of easy-to-access interfaces, fast and reliable scalability (in both directions), and the ability to keep multiple clients' data in the same place but totally separated, and, in some instances, the ability to measure how much storage a client uses. Cloud storage used to be object storage (i.e. it manages data as objects), but now it extends to methods like block storage (i.e. it manages data as blocks within sectors and tracks) [3]. Cloud storage needs to look seamless to users, consistent in access through redundancy and distribution of resources, and consistent in version control. The data is stored in logical pools; physical storage resources are aggregated into storage pools and the logical storage is created from those storage pools [5].

Scalability must be provided not only for the storage itself (functionality scaling) but also the bandwidth to the storage (load scaling). Another key feature of cloud storage is geographic distribution of data (geographic scalability), allowing the data to be nearest the users over a set of cloud storage data centers (via migration). For read-only data, replication and distribution are also possible (as is done using content delivery networks). This is shown in Figure 1.

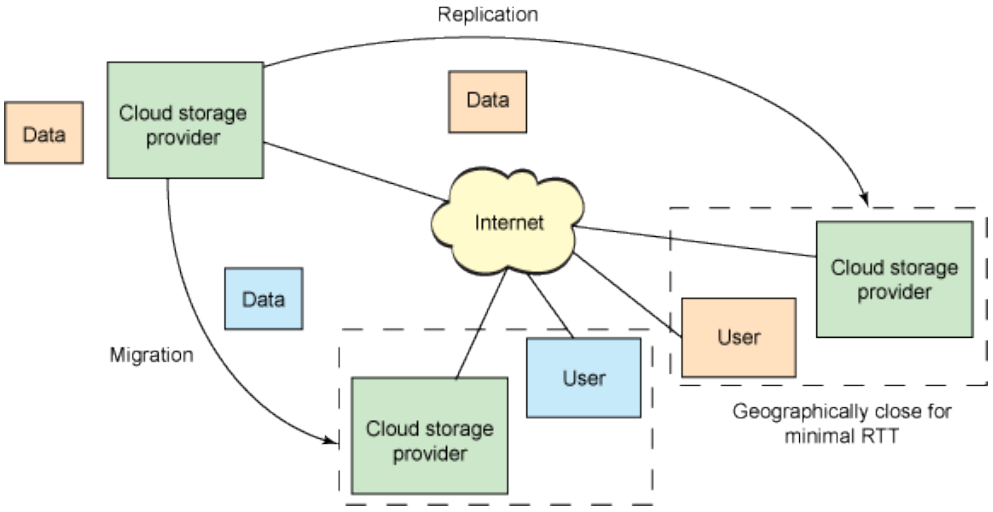

Figure1.scalability of cloud storage

A) Cloud-Optimized Storage

The major method for optimizing storage for a cloud system isoptimizing the data. Data optimization is probably the most significant recent software innovation in storage because you can save more information in a smaller physical space. The three tools of data optimization are deduplication, compression, and thin provisioning.

- *Deduplication*

Cloud computing is a paradigm shift in the Internet technology. Datadeduplication can save storage space and reduce the amount of bandwidth of data transfer.Data deduplication often called intelligent compression or single-instance storage is a process that eliminates redundant copies of data and reduces storage overhead. Data deduplicationtechniques ensure that only one unique instance of data is retained on storage media, such as disk, flash or tape.

- *Compression*

Storage space, such as that provided by computer hard drives, comes at a price. Compressing data files allows you to store more files in the storage space you have available.

- *Thin Provisioning*

Thin provisioning is a storage area network (SAN) management process where the storage capacity for a device is reserved and allocated on demand through a shared storage pool. Thin provisioning is also known as virtual provisioning. However, thin provisioning relates to physical computing environments.Cloud provisioning is the allocation of a cloud provider's resources and services to a customer. The growing catalog of cloud services that customers can. Provision includes infrastructure as a service, software as a service and platform as a service, in public or private cloud environments.

III. CLOUD STORAGE ADVANTAGES

Companies need only pay for the storage they actually use, typically an average of consumption during a month. This does not mean that cloud storage is less expensive, only that it incurs operating expenses rather than capital expenses.

Businesses using cloud storage can cut their energy consumption by up to 70% making them a more green business. Also at the vendor level they are dealing with higher levels of energy so they will be more equipped with managing it in order to keep their own costs down as well [4].

Organizations can choose between off-premises and on-premises cloud storage options, or a mixture of the two options, depending on relevant decision criteria that is complementary to initial direct cost savings potential; for instance, continuity of operations, disaster recovery, security, and records retention laws, regulations, and policies.

Storage availability and data protection is intrinsic to object storage architecture, so depending on the application, the additional technology, and effort and cost to add availability and protection can be eliminated.

Storage maintenance tasks, such as purchasing additional storage capacity, are offloaded to the responsibility of a service provider.

Cloud storage provides users with immediate access to a broad range of resources and applications hosted in the infrastructure of another organization via a web service interface.

Cloud storage can be used for copying virtual machine images from the cloud to on-premises locations or to import a virtual machine image from an on-premises location to the cloud image library. In addition, cloud storage can be used to move virtual machine images between user accounts or between data centers.

Cloud storage can be used as natural disaster proof backup, as normally there are 2 or 3 different backup servers located in different places around the globe.

IV. CONCLUSION

In cloud Data storage is no longer a warehousing issue; implicit in the new world of data everywhere is the implied ability to find, access, and use that data in an efficient manner. Much of that data exists on a cloud, so you need to know how to make sure data storage is optimized and doesn't become a weak link in your cloud platform. Storage techniques and software tools can help you achieve data and database optimization, and help to manage virtualized data storage through the software layer. This article explores what storage optimization tools can do for you.

REFERENCES

[1] N. Loutas, E. Kamateri, F. Bosi and K. Tarabanis, "Cloud Computing Interoperability:the state of play," in *IEEE Third international conference on cloud computing technology and science*, 2011.

[2] Narendra, N. C., Koorapati, K., and Ujja, V. (2015). "Towards cloud-based decentralized storage for internet of things data", IEEE International conference on cloud computing in emerging markets.

[3] Xu, L., Chen, W., Wang, Z., and Yang, S. (2012). "Smart-drs: A strategy of dynamic resource scheduling in cloud data center", IEEE International conference on cluster computing workshops .

[4] Beloglazov, A. and Buyya, R. (2010). "Energy efficient resource management in virtualized cloud data centers", 10th IEEE/ACM international conference on cluster, cloud and grid computing.

[5] J. McKendrick, "Does Platform as a Service have interoperability issues?," 24 May 2010. [Online]. Available: http://www.zdnet.com/article/does-platform-as-a-service-have-interoperability-issues/

[6] A. Sheth and A. Ranabahu, "Semantic Modelling for cloud computing. Part I and II," *IEEE Internet Computing magazinevol 14*, pp. pp 81-83, 2010.