RESEARCH PAPER

Cloud-Based Storage for Education

Sushma Satpute
Deapartment of IT
Cummins college of Engg., R T M N University
Nagpur, India
sushma.satpute@gmail.com

Dr. Bharat Singh Deora
Department of Computer Science & IT
JRN Rajasthan Vidyapeeth University
Udaipur, India
bsdrvu@yahoo.com

Abstract: Technology is becoming an important element in today's educational system. Educational institutions and schools are now allowing students and faculty to use their own digital devices, computers, notebooks, and smart phones to study and to access the resources at the institution. The recent initiative by the government to bring in subsidized tablets to the school going children and college students are examples showing the importance of technology in education and the future direction. As the usage of digital devices will be increased in the educational institutions and among the students, the amount of data storage will increase exponentially and so as the need to effectively use this data and to have the storage capacity with the institution. This paper describes how cloud storage could be used for educational institutions, among the students and faculty to effectively store, share and distribute educational data, and also the benefits of using cloud storage and the challenges and issues with the same.

Keywords: Cloud computing; cloud storage for educational institutions; green initiatives; cache on device; S3

I. INTRODUCTION

Use of technology in education allows teachers to create new levels of interactivity that are ideal for individual and team learning. Mobile devices open up a universe of possibilities for science labs, distance learning, and student presentations. Despite the many possibilities of technology in education, many educational institutes face limited budgets and cannot afford maintaining the large computing systems and data storage devices. While many students now have laptops and students groups require collaboration of data [1]. To accommodate this huge storage need, the institution uses network-attached storage (NAS). Students use the network drives to store their data or share them amongst each other for collaboration. The institution needs to manage the infrastructure to handle multiple locations. Each location needs its own network-attached storage with proper protection scheme. This would put immense pressure on institutions to handle such large data storage and expenditures will increase to maintain the infrastructure for this huge data. Supporting faculty, staff, students and that too at different geographical locations with a small IT team and limited resources is very difficult for the educational institutions. In addition to the large demand coming from new students enrolling every year, schools and universities are required to retain students and other educational records for a certain number of years. This requires mounting more storage capacity and management of storage devices and that adds to the complexities of storage management. To take care of such issues where the work is moving around big data storage we suggest a model based on using cloud storage, an element of cloud computing Iass. We proposed the model that uses cloud storage for the users in educational institutes, but taking same concept forward any system could be designed that requires data in huge form by using storage in the cloud.

Cloud Storage concept is around for a while and is one of the important element of cloud computing. Cloud Storage could be very useful if used properly and can handle the needs of various institutions and organizations.

Unlike other cloud services, there are quite a few options to choose from for the storage services. The big name providers include technology majors like Microsoft, Google, Amazon and HP are now providing cloud based storage as a service and on pay what you use model [2].

II. CLOUD STORAGE

Cloud storage is growing in popularity due to the benefits it provides, such as simple, lower costs, anywhere access and the removal of maintenance and storage management. It is the delivery of data storage as a service, from a third party provider.

According to Gartner study, "the desire to share content and to access it on multiple devices will motivate consumers to start storing a third of their digital content in the cloud by 2016. Just 7 percent of consumer content was stored in the cloud in 2011, but this will grow to 36 percent in 2016" [3].

Cloud storage has access via the internet and the usage billing calculated based on capacity used in a certain period of time (e.g. per month) and the user need to pay for the storage they actually use. Organizations using cloud for storage do not need to install on campus physical storage devices in their own datacenter and the same could be accessed by any authorized user from any device According to National Institute of Standards and Technology (NIST): "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources.[4]

Cloud computing has five characters:

a) On-demand service means you can get what you need when you need it.
b) Broad network access, means heterogeneous thin or thick client platforms (e.g., mobiles, laptops, and PDAs, i.e. access is anywhere, anytime.
c) Resource pooling, means the subscriber is generally has no control or knowledge about the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.
d) Flexible resource allocation, means as demands changes, cloud services can scale rapidly. We don't have to worry about bringing new servers online or resources allocation.

e) Measured service, Mostly cloud usage is metered, as per user or per hour. You pay for what you use.

The other factor which affects computational performance in the cloud is contention. RAM and storage cannot be actually over-allocated but CPU can be over-allocated. The levels of contention vary considerably but essentially public cloud is able to use the CPU capacity of a physical host at more than 100%. We can use CPU contention ratios of over three times i.e. the total CPU capacity of all the virtual servers on the same physical machine might be three times its actual CPU capacity. As most virtual servers do not utilize 100% of their CPU allocation for most of the time but still contention ratios directly affect performance. If contention is high (i.e. at anything more than 200% CPU allocation) then cloud server performance may also deteriorate significantly. But this problem is one of the possibilities which can be minimized by using another drives. Another problem to storage is between performance and redundancy/reliability. The more redundant, the slower the performance .This also can be avoided by drawing a distinction between cloud storage that is persistent and temporary storage. For temporary storage we do not store critical data, so that the important data can be safe. We can improve the performance and reduce the impact of any drive failure. If any drive fails in an array we quickly replace it with another drive, and drives array is re-constituted in a much shorter period of time load of heavy disk access is spread across a greater number of drives. Hence storage performance can be improved.

III. RELATED WORK

There are some suggested models designed for the education purpose. Mohssen M. Alabbadi proposed C3F – complete cloud computing formation model for educational institute [8] which mainly is the further implementation of CCM proposed by JERIKO FORUM [7]. Both the models have distributed the cloud into three parts according to the activities in IT organizations such as mission criticality and sensitivity. Each class of IT activity is placed into appropriate position in their C3F model. [8]

Another models which proposed commercial and non commercial approaches for cloud computing are mainly framed around the general utility of cloud computing and not on improving the functionality of suggested models. [9, 10]. All models are focused on designing the architecture and using cloud into the model.

In this paper we have suggested a model for data caching on the user's device that would improve the functionality and performance of those models and would make them more usable and adaptable for institutions where user mobility is more.[11]

IV. PROPOSED MODEL

Cloud storage has several advantages over traditional data storage. For example, if data is stored in cloud storage system; the data could be accessed from any location where Internet is accessible. There is no need to carry around a physical storage device or use the same computer to save and retrieve the information. Since data is accessible from anywhere an educational project could be turned into a collaborative effort by sharing the data real time among a group of students.

Figure 1. Cloud Storage space accessed by multiple devices

Whether it is course curriculum or student assignments, lot of paper work need to be shared among parents, students, teachers, and administrative groups. Cloud Storage is the most secure and easiest way to address this [6].

The following aspects of the cloud storage are relevant and can be used for educational institutions and among the students and faculty.

a) Data Sharing

The first is sharing of information among various sub units. Cloud storage is very useful in sharing data and information. The cloud storage provider gives tools to share the data stored in the storage account. User of the storage account can control what data is to be shared and can publish URLs for others to access the files. A faculty can use such powerful feature of sharing, to easily share or publish assignments, homework and to distribute notes among the students without use of any paper and in real time. Students can have access to important curriculum and other educational content published by the faculty from any device like phones, tablets or laptops and from any physical location.

b)Data Synchronization

One of the useful features of cloud storage is Data Synchronization. In schools, any file saved to cloud storages can automatically be synchronized to all devices such as a desktop, laptop, and tablets. There are tools available to automatically sync all of files to number of devices in real time. Also the storage providers could also provide such tools. For example, if a document is added to a cloud-based storage from a desktop, the file can be viewed and edited on other devices like a phone or a tablet or any other device, which has connectivity to the internet, as it is already saved on the storage account in cloud-based storage. The changes made to the document will be reflected on other devices which are synchronized to the same storage account. The students can sync their devices with their faculty and can get any information updated without even doing any manual action to get the same.

c)Collaboration

Cloud storage acts as a good tool for collaboration. We use this in our model. Consider a scenario where a group of students are working on a project and collaborating. The project requires a set of documents to be shared and distributed among the project group. The files need to be stored at a common place accessible to all the project group members.

With cloud based storage, the collaborators can independently access the documents from any device, work on them, and save them to the one centralized place which is accessible to the group. The collaboration group can also maintain the versions and the change history for reference.

d)Data Backup

Data backup is maintained at low cost in cloud computing. Cloud backup work in much the same way as traditional backup, the difference is that the process is much simpler and the data is moved to a server over the internet, rather than copied onto a local server. The advantage is that the data is physically away from the site. It also doesn't rely on any individual and is less prone to human errors that generally happen while taking scheduled backups with conventional methods. Also whenever required the data could be restored without having the media physically present and is generally a one click process. The institutions need not to maintain in house infrastructure to preserve the data that could require huge investment otherwise.

Following figure 2 shows an Institute campus, it's remote office locations and remote users (faculty, students or staff) with internet connectivity using cloud as the storage medium.

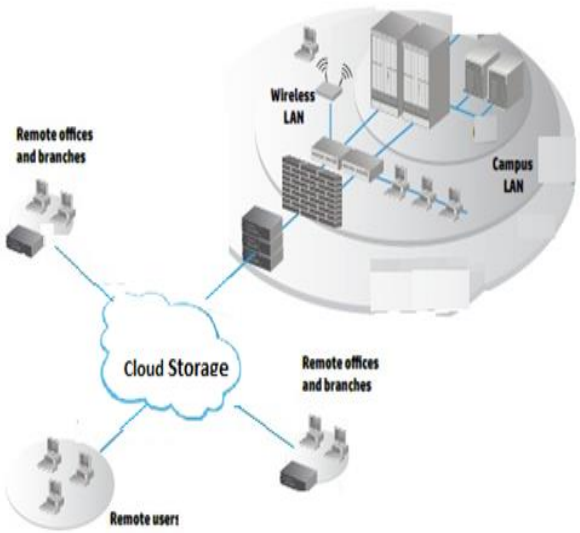

Figure 2. Institute campus, it's remote office locations and remote users (faculty, students or staff) with internet connectivity using cloud storage.

The figure shows that various branches of the institute need not to have their in house data storage mediums and instead using the storage in the cloud provided by third part vendor like Amazon S3. Accessing data from cloud storage requires internet and this could be an issue where internet connectivity is not easily available especially in smaller cities and towns. Also the performance will be decreased if the internet speed is not good enough for large data transfers. User mobility in educational institutions is more as compared to an organization where most of the users are stationed at their defined physical locations. The model we propose is based on providing on-device cache to the user so that data availability and accessibility is much more even when the user is not hooked to the internet or to the in house network. User will still get the data that he frequently uses on his own device rather than having to go to the cloud.

Now a days as the use of digital devices is increasing in education, a lot of study materials is in form of videos, pictures or digital books. This digital data once created is hardly changed or edited over time. Our on-device cache model will enable the users to access the data from the local cache instead of getting it from the cloud every time is is required.

The application model has following components:

a. Cloud storage provided by a vendor like Amazon S3 (Simple Storage Service)
b. Cache storage on user's own device
c. Database on user's device to save cached data information

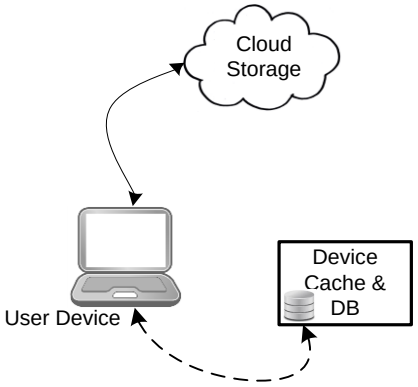

Figure 3. User's device connected to cloud storage and having own cache and database.

Following sections describes the working of our proposed application model:

- The application is installed on the user's device.
- User logs in to the Cloud Storage account and as soon as data is accessed, application logic checks for the accessed data information in the cache database on user's device.
- If the data is already present in the cache, the data is fetched and presented to the user.
- If data is not present in the cache database, data is fetched from cloud storage and saved in the cache. The cache information is replicated to the cache database on user device.

User can control the growth of cache size and can define the time till the cached data remain in the local device. The background service will periodically check for the elapsed time for the cached data and flush cache to free up the space. There are algorithms available for cached data replacement to control the cache size. The application logic could be implemented that best suits the requirement of users.

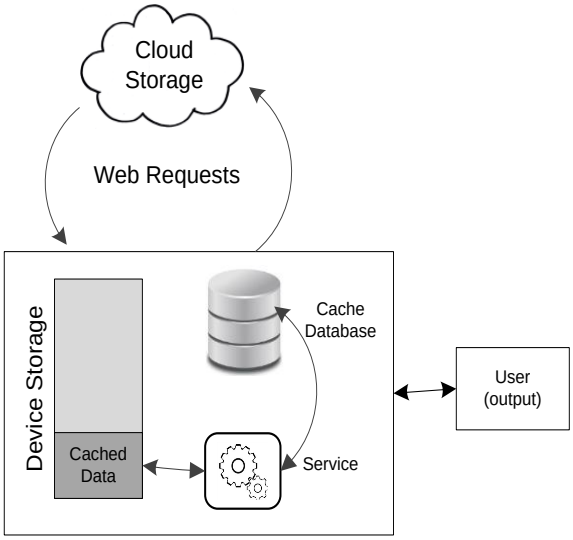

Figure 4. Architecture of the device cache model.

The proposed model for caching cloud data on user device not only reduces the latency in data access but also helps in controlling the cost of data access from the cloud storage. The network traffic will be reduced and will have direct impact on the bandwidth requirement for the institution. The users will have access to their frequently accessed data even if the internet is not available or the user is not connected to the in-campus network. Accessing data while user is offline is very important since there are connectivity issues for the educational institution in smaller towns and cities and the user mobility is more as compared to other organizations.

We used AWS's (Amazon Web Service) Simple Storage Service (S3) for our proposed application model.[12] AWS is being considered as the most utilized and reliable storage service. REST APIs given by AWS are used to create the web requests to read and write data to the cloud storage.

A test setup of 10 user devices having installed device cache application and saved around 100 files on Amazon S3 which includes documents, pictures and videos and all devices randomly accessing the files during a defined period. The usage report from Amazon S3 shows substantial reduction in DataTransfer-Out-Bytes and GET requests (Requests-Tier2) made to the Amazon S3 cloud. Following Figure 4 shows the graphical representation of approximate reduction in Get requests and DatTransfer-Out-Bytes from Amazon S3 on using data cache on user's device.

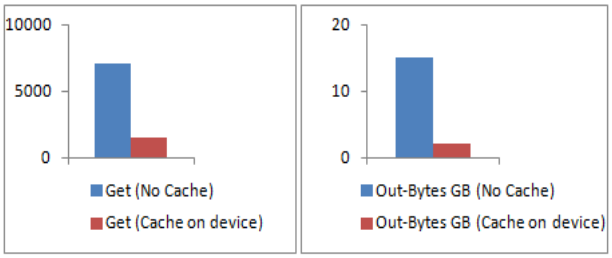

Figure 5. Reduction in Data Transfer-Out-Bytes and Get requests from Amazon S3 using data cache on device.

V. CONCLUSION

The importance of data storage cannot be overstated and with the increasing use of technology by the school going population growing up with the latest technology and devices, it is essential for educational institutions to adopt cloud based storage. With our proposed application model of Caching Cloud Storage Data on Device, the institutions can get better services than traditional computing and with reduced cost using cloud based storage. Now schools can accelerate learning, increase flexibility, and cut costs using such new technology models. Not only the cost, but there are many other reasons why educational institutions of all sizes and types should adopt the model of cloud storage with on device caching. Adopting to new technology like cloud storage can help institutions in taking Green Initiatives by reducing the use of paper and help environment. Along with caching, further study needs to be done to minimize the effect of issues related to internet speed, availability and cost. Also to enhance our proposed model of cache on device, further study needs to be done on how the APIs provided by the storage providers could be used to create custom applications for educational institutions and student's specific needs. The proposed application model in this paper addresses issues related to cost, availability and accessibility of cloud data and would certainly help institutions to adapt cloud storage model. The benefits of cloud based storage and some of its challenges like usage cost, internet availability and access not only opens up an area for the scholars to research but also provides huge opportunity for IT entrepreneurs to create applications on cloud based storage model that could be used by educational institutions and students.

VI. REFERENCES

[1] (March 2012) HP Business White Paper: BYOD in Education [Online] Available: http://www.nikeconsultants.co.uk/wp-content/uploads/2012/07/Business-White-Paper_HP-BYOD-in-Education.pdf

[2] Cloudtweaks website [Online] Available: http://www.cloudtweaks.com

[3] Jonathan Strickland (2008), "How Cloud Storage Works" website [Online]. Available: http://computer.howstuffworks.com cloud computing/ cloud-storage.htm

[4] NIST-Draft cloud computing synopsis and recommendations: Special Publication 800-146

[5] Gartner Press Release [Online] Available: http://www.gartner.com/it/page.jsp?id=2060215

[6] CloudDip, "Cloud storage for Educational Institutions and schools." [Online] Available: http://www.clouddip.com/education

[7] Jericho Forum, "Cloud cube model: selecting cloudformations for secure collaboration," Version 1.0, April, 2009, San Francisco CA, USA TheOpenGroup. [Online]. Available: https://collaboration.opengroup.org/jericho/cloud_cube_model_v1.0.pdf\

[8] Mohssen M Alabbadi, "Cloud Computing for Education and Learning: Education and Learning as a Service (ELaaS)".IEEE 14th International Conference on Interactive .Collaborative Learning (ICL2011) 11th InternationalConference Virtual University (vu'11) 21–23 September 2011, Pieš_any, Slovakia

[9] Y. Khmelevsky and V. Voytenko, "Cloud computing infrastructure prototype for university education and research," In the Proceedings of the 15th Western Canadian Conference on Computing Education. (WCCCE 2010), May 7 - 8, 2010,

Kelowna, British Columbia, Canada, ACM, Inc., New York, NY, USA, 2010. [Online]. Retrieved (Aug. 30, 2011): http://dl.acm.org/citation.cfm?id=1806524&bnc=1.

[10] Deepanshu Madan et al., "E-learning based on Cloud Computing", Volume 2, Issue 2, February 2012 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering. Research Paper [Online]. Available: http://www.ijarcsse.com/docs/papers/february2012/volume_2_issue_2/V2I2048.pdf

[11] David Burgett, MSDN Magazine, Web Farms:"Use Data Caching Techniques to Boost Performance and Ensure Synchronization" [Online]. Available: http://msdn.microsoft.com/en-us/magazine/cc188799.aspx

[12] Amazon Simple Storage Service: Developer Guide. (API Version 2006-03-01) [Online]. Available: http://awsdocs.s3.amazonaws.com/S3/latest /s3-dg.pdf