Storage and Retrieval of Records into Different Types of Devices for Better Performance: Access Frequency Approach

Kavyashree H N[1], S G Makanur[2]

[1]IV sem M-Tech (CSE), STJ IT, Ranebennur
[2]HOD, Department of ISE, STJ IT, Ranebennur

Abstract-Today, the world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly. With the growing volumes of data, there is a need to increase the storage capacity which in turn affects the storage cost too. But financially, it is impractical to increase the storage costs at the same rate as that of storage growth. This is because all the data are not of equal importance in their usage. Storing of data which is rarely used in high speed and high cost storage devices is impractical. Thus, there is a need of a methodology, which can reduce storage cost, by storing the data into different storage devices, based on their usability. The paper aims at emphasizing the reduction of storage cost. The idea here is Tier storage technology, it stores the high frequency usage record in fast device and less used records in comparatively lesser speed devices based on the access frequency and thus providing right information at right time at right cost.
Keywords-Access Frequency, Digital Information, Storage Cost, Storage Devices, Tier Storage.

I. INTRODUCTION

We live in an information-centric world. Data is a precious thing and will last longer than the systems themselves, says, Tim Berners Lee. Data today is growing unimaginably at a huge pace. As a society, we have a growing reliance on creating and consuming data, which must be available when and where it is needed. As the data is growing vaster and vaster, there is a huge explosion of data.

The global explosion of data has unleashed unprecedented data management responsibilities. As a result of this increasing reliance on information, both for home and personal use along with business and professional needs, more data is being generated, processed, moved, stored, and retained in multiple copies for longer periods of time. The net result is that IT organizations of all sizes have to do more with what they have including maximizing available IT resources while overcoming common footprint constraints.

More data must be preserved indefinitely and made continuously accessible throughout its lifespan. Data growth is vastly outpacing many organizations' ability to manage and keep the data. Today's purchase decisions in storing, managing, and archiving data will have long-term consequences for businesses. Data storage technology needs to be upgraded — but for many organizations data lives on indefinitely and must be continuously migrated across storage platforms and tape archives as new technologies emerge. As storage professional, you need trusted, familiar, and flexible archiving solutions to preserve data indefinitely while non-disruptively expanding and upgrading data archives. Because there is exponentially more data to manage, you also need systems that come fully configured, operate as specified, and are ready to deploy.

With the growing volumes of data, there is a need to increase the storage capacity which in turn affects the storage cost too. To store the huge amount of digital data there is a need of high speed and high cost storage devices like RAID. But financially, it is impractical to spend too much on storage along with increasing volume of data. This is because out of all the data available there are millions of data that are not used often. Such data can be data of students of a school/college, who are presently not the students of that school/college, earlier editions of a text book, data of people admitted in a hospital who have been discharged from the hospital or are not alive etc. Meanwhile

other data like data of current students of a school/college, new editions of text books, data of people who are currently admitted in hospital etc are more used and are required at faster pace. Thus, all the data available are not of equal importance in terms of their usage. When this is the case, storing of data which are rarely used in high speed, high cost devices is very much impractical. Data need to be classified based on their usage and then storage of those data in the devices must be decided.

The aim here is to build a data storage and retrieval system based on the access frequency of data. The idea here is to store the data of high frequency usage in high speed high cost devices like RAID, the data which are not that frequently used in comparatively less speed & less cost storage devices like hard disks and the data which are not accessed for a long time in still more less speed and less cost devices like tape drives. Retrieval of data here will be efficient as the data frequently accessed will be retrieved faster. Thus, doing this will allow efficient use of data in a very cost-effective manner.

II. RELATED WORK

Storing and retrieval of data is very important and basic requirement of today's life. The data we create today is growing at a faster and faster, whether we like it or not information is being stored about us. Applications are demanding for more and more storage space then memory requirement is increasing. . Hence to meet this requirement day by day new storage techniques are coming in to the picture and we are reached to holography and 3D optical storing techniques.

Holographic storage is one of the storage techniques for mass storage that uses three-dimensional holographic images to enable more information to be stored in a much smaller space. And holographic storage is capable of recording and reading millions of bits in parallel and it uses a Laser light.

In 3D optical storing techniques, information can saved or retrieved with 3D resolutions. This new and latest innovation can provide very high mass storage of petabyte level. The main challenge is the, development of commercial products are taking a significant length of time. And also this 3D Optical Data Storage has some of issues as Destructive Reading, Thermodynamic stability and Media sensitivity.

Today we are moving forward to hologram and 3D optical storage technique to overcome the drawbacks of memories and improving the existing memories to meet the storage requirement. But we cannot claim to understand everything about the data storage technologies available today, and we need to think about cost of the storage, what works and what doesn't. So if you are interested in talking about a data storage solution then why not you use Tier storage technology, in which different storage devices are used and data is automatically moves between high-cost and low-cost storage media. This technology is good in terms of performance, capacity and functions and also even saves your money.

III. PROPOSED SYSTEM

Here, we have implemented one of the most important enabling technologies for big data applications are that of tiered storage. The tiered model enables users to build storage systems from heterogeneous technologies, which allows applications to be optimized for price, performance, capacity, and functionality. Here, we are partitioning records based on the access frequency which is not a key nor a attribute of table and placing them into appropriate type of storage device, periodically moving the data across the device based on new access frequency automatically.

Divide database physically into separate layers but logically together. The database can be organized into layer based on its access speed. Three Tier model is shown below.
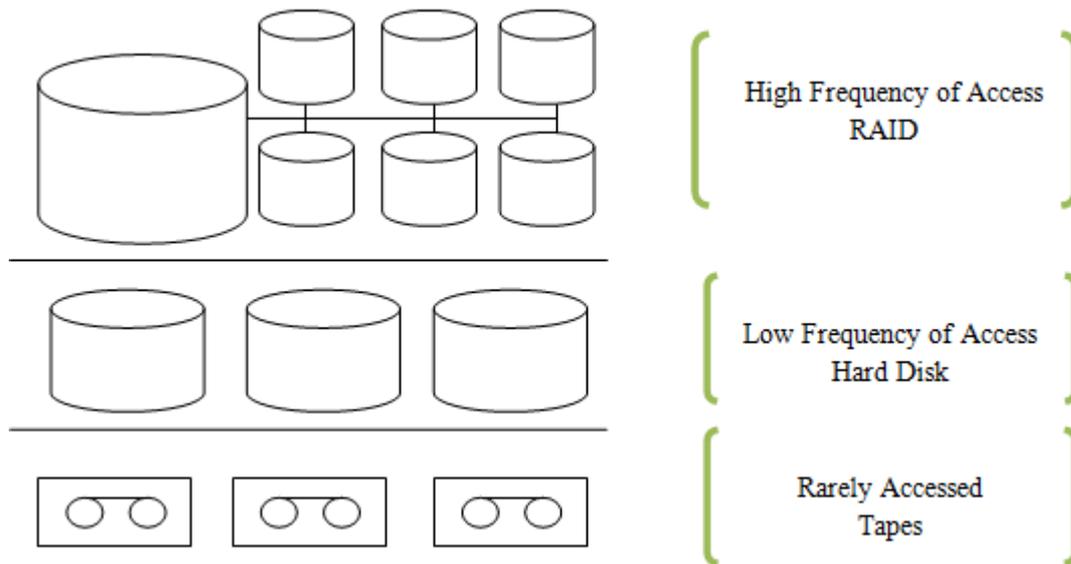
Figure: three layer tier storage architecture

Data is moving from one layer to other layer, measuring its usage by how many times particular data is accessed. Based on access frequency name the data as hot, warm and cold data.

- High Reference Data stored in Faster (Expensive) storage device like RAID – Hot Data.
- Low Reference Data stored in low speed (Less Expensive) storage device like Hard Disk – Warm Data.
- Rarely Reference Data stored in low speed (Low cost) storage device like Tapes – Cold data.

IV. CONCLUSION

Data in today's world is very important and thus needs to be stored safely. As the data is growing in volumes, there is a very much need of storage devices with high storage capacity and high speed. Now, when the data needs high capacity storage devices for storage, it is obvious that we need to invest so much on storage devices. But it is impractical to invest too much on storage devices just because there is a need for storage devices to store the data, as all the data that need to be stored and retrieved are not of equal importance based on their usage. Thus there is a requirement of a methodology that does this. With this in mind, the paper here proposes a tired storage methodology which stores the data based on its access frequency. The data with high frequency usage is stored in high speed high cost devices and the data that is not used often will be stored in lesser speed and less cost devices and thus provides the right information at right time and at right cost.

REFERENCES

[1]   S. Kawata and Y. Kawata, "Three-Dimensional Optical Data Storage Using Photochromic Materials" Chem. Rev. 2000, 100, 1777.
[2]   Han Hu, Yonggang Wen, Tat-Seng Chua, and Xuelong Li,"Towards Scalable Systems for Big Data Analytics: A Technology Tutorial". Journal of latex class files, vol. 6, no. 1, june 2013.
[3]   East Brunswick Public Library, "Computer Storage Devices" http://www.ebpl.org.
[4]   Jinal H. Tailor ,"3D Optical Data Storage Techniques", Volume : 3 Issue : 4, April 2014 ISSN No 2277 – 8160

BIOGRAPHY

Kavyashree H N is a student pursuing her Master degree in Computer Science and Engineering department at STJ Institute of Technology, Ranebennur, Karnataka, India. Her research interests are Computer Science related aspects such as storage area network, data base , dot net, networking.

Prof. S G Makanur M.Tech , HOD in the department of information Science and Engineering at STJ Institute of Technology, Ranebennur, Karnataka, India. His research interests are information Science related aspects such as storage area network, data base ,data mining.