To Provide Privacy and Storage Backup for Users Using Personal Data Vault

B. Kavya Chandra, J. Jaya Lakshmi

[1] PG Student, Department of CSE, Narayana Engineering College, Andhra Pradesh, India
[2] Senior Assistant Professor, Department of CSE, Narayana Engineering College, Andhra Pradesh, India

---

Abstract: *The advances in technological developments have made Input-Output devices even better to store the information. Global storage system is used an inexpensive solution to the cloud environment. There is an increase in the need for information the user requests .The demand for these I/O data applications get higher as they get larger. Parallel file systems are used to improve the performance of these applications. In this paper we present an implementation for personal data vault that is able to store and back up data by using remote servers that can be accessed through the Internet. The main aim is to increase the availability of data and reduce in loss of information.*

Key Words: Cloud Environment, data vault, Backup Storage, Parallel File System, Virtual File System

1. INTRODUCTION

Nowadays there is a huge growth in technologies and as a result we have many services and applications for users to carry on with their daily tasks with more productivity. It has become an authentic need to get the required information from a remote site at any circumstance. This had resulted in searching for alternatives to the service providers. The best known alternative is cloud computing. Data vault can be defined as that service (software, platform or infrastructure) located on the Internet and is accessed from a mobile device or desktop computer, giving users a wide variety of applications (databases, middle office software, storage, etc.).

For delivering information and for providing online functionality [1] cloud computing and cloud storage are used. The reason behind using cloud storage is, the time required to access the data is reduced. It is easy to implement a single server which stores all the required data than using multiple servers. Still the fact is that, it allows parallel processing where multiple customers or users can access different parts of a file and thus further increasing parallelism. By using the infrastructure that already exists, users are given the experience of cloud computing at a very low cost. At this cost they can implement their private cloud environment and back up storage. Since this is an open source software users have full control to access the information, and thus maintaining the confidentiality of the data.

Private Mists [2] is a choice for some organizations that require high confidentiality and security. This kind of cloud is a decent option. The customers are allowed under permissible limits based on their identity. Thus by limiting users we can achieve high confidentiality. There is one more sort of cloud is known of half breed models join open and private mists. In this model it claims and imparted segments other in a controlled way.

2. RELATED WORK

Many organizations make use of personal data vault. Previously websites were to be under web 1.0 which restricts RIA (Rich Internet Applications), and hence some client server methodologies cannot serve their purpose

[3]. Some cloud services focus on providing consumers a wide range of services and functionalities. Others provide cloud storage to consumers for free or charge some type of subscription-based fee for example Windows Azure [4] is an open cloud platform in a global network of data centers run by Microsoft. Let's compile applications in any language, tool or framework for integrate your public cloud applications. In a hybrid cloud [5] environment both use of private and public cloud services are in use. Many Companies are realizing that they need many different types of cloud services in order to satisfy the customer needs. However, as many of the solutions of this type have a considerable cost. Dropbox [6] is a file hosting service which provides cloud storage, file synchronization, and client software. In Dropbox users are requested to create a special folder on their devices and Dropbox then synchronizes the folder (with the same contents) so that it can be used regardless devices used to view it. In cloud environment there are some file systems for backup and storage of files within input/output [7, 8]. Panzura CloudFS file System [9] is a file system developed to provide integration with cloud and NAS environments. It offers functionality transparent to users, as everyone can see the same file from any location .Without having to delete the original file, it allows data sharing .Cloud File System Oracle [10] is a file system for private cloud environments, designed to manage general purpose file store outside of an oracle database across multiple operative system platforms with one management interface.

The Parallel Virtual File System project is a multi-institution collusion effort to design and implement a production parallel file system for HPC applications [11]. It is an open source system which is available to all users. A parallel file system distributes file data across multiple servers and thus concurrent access is provided. PVFS has been designed for use in large scale cluster computing. The client library provides for high performance access via the message passing interface (MPI). The extension provided by the second PVFS version, PVFS2, is that it improves modularity and flexibility among modules, and provides a strong integration with MPI-IO [12].

3. PROPOSED SYSTEM

A cloud environment had been proposed for backup on a private cloud using PVFS2. This file system increases the performance of these applications. This allows parallel input / output operations, to reduce the access period. A multiplatform application is built using free software which allows data transfer fast and simple way on the client-end.

3.1 Data vault and Backup File System

Data vault is a network model in which the data is backed up and stored using an internet connection on remote servers. The service providers (hosting companies) provide large data centers and users who want to access their data or to store it buy or lease storage space from the servers. Virtualization is used by the data center managers often to reserve resources based on the requirements of the customer and storage spaces are allotted to back up their files or data objects. Physically, These resource can protract along pooled servers and locations. The confidentiality provided to the user data purely depends on the service provider. PVFS2 file system had been selected for our data vault; as it is open source, we have the flexibility to use and modify according to our compulsion. Let us know how these file systems actually manage the backup and storage of files within the input / output.

3.2 Working of PVFS2

PVFS2 distinguishes three different components as

- *Data servers or I/O servers* aggregates the PVFS store subsystem. The storage space available in individual data servers is obtained and the

smallest obtained value is multiplied by the number of data servers to know the available data storage.

- *Metadata servers* store all the metadata information of the PVFS namespace files (metafiles).

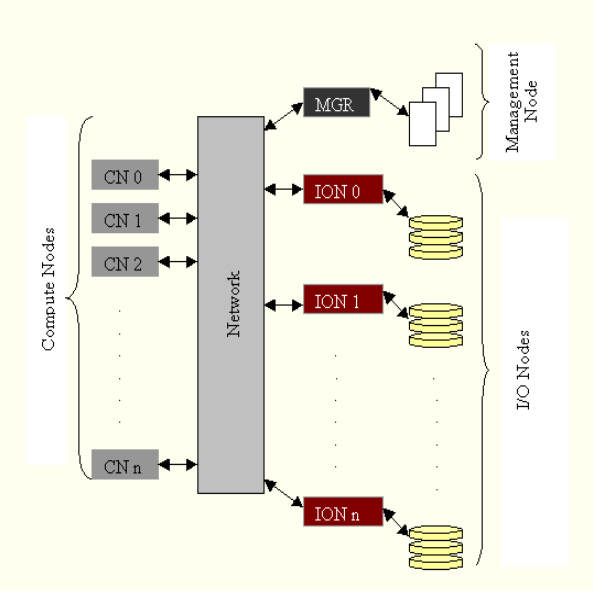

Fig-1: PVFS2 Architecture

- *Clients or processing nodes* that approach the file system through one of three application programmer interfaces (APIs): a kernel Linux module that allows approach through the Linux virtual file system (VFS) with the POSIX I/O API; a user-space library that provides access through what it is called the PVFS system interface (pvfs2lib); and a high level library, MPI-IO.

3.3 Operation of Server Application

Each user is provided with a thread through which Sockets application get in contact with the client for sharing the data, based on the described rules in the precedent section. Data and metadata are received by application server as string of bytes encrypted. A web application had been designed for users to view and download their data by creating the user account and thus maintaining personal information. File name, size, path, last date modified, deleted mark, this information is stored in a MySQL table.

4. RESULTS

The test program used for concurrent write performance is pvfs-test, which is included in the PVFS release package. In all tests, each compute node writes and reads a single contiguous region of size 2N MB, where N is the number of I/O nodes in use. The results obtained prove that the performance of PVFS2 is far better than the existing file system. The same is shown in the table given below and the results are depicted in the graph.

Table-1: Comparision between existing file system and proposed PVFS2

File System	Read (MB/S)	Write (MB/S)
PVFS2	45	35
Ext3fs	25	20

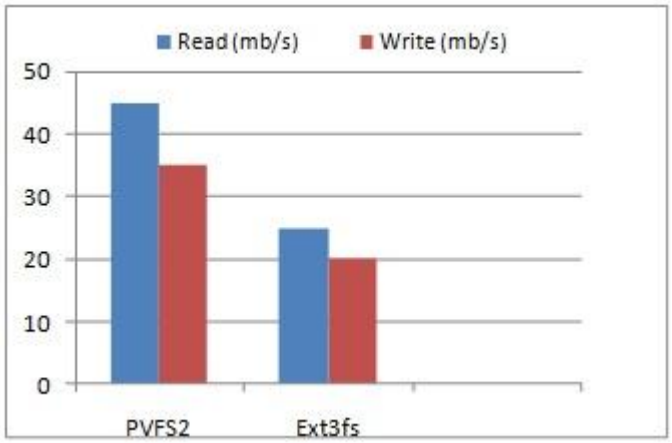

Chart-1: Comparison of results

5. CONCLUSION

We propose a design & implementation for the server integration to backup and store data on a private cloud using PVFS2 file system for storage data in order to increase the performance of applications that requires high I/O data demands. PVFS2 is file system which is open source, will reduce access time to data as it allows input / output parallel. On the client-side, an application is developed that allows data transfer fast and simple way. The advantages of this implementation are that it can reuse existing infrastructure (servers, cluster, and other devices) that reduces the cost and increases the throughput.

REFERENCES

1. Camacho, Hugo E., et al. "A cloud enviroment for backup and data storage."*Electronics, Communications and Computers (CONIELECOMP), 2014 International Conference on*. IEEE, 2014.
2. Gourshettiwar, Palash M., and Dhiraj Shirbhate. "A Survey paper on Cloud Environment for Backup and Data Storage."
3. P. J. Braam, "The Lustre Storage Architecture," November. 2002.
4. Windows Azure : http://www.windowsazure.com/es-es/
5. Hybrid Cloud: http://www.redhat.com/products/cloudcomputing/cloudforms/
6. Dropbox: https://www.dropbox.com
7. SkyDrive: http://windows.microsoft.com/es-es/skydrive/download
8. GoogleDrive: https://support.google.com/drive/answer/2424384
9. Panzura CloudFS file system (White paper) http://panzura.com/products/global-file-system/
10. Oracle Cloud File System (White paper): http://www.oracle.com/us/products/database/cloud-filesystem/ overview/index.html.
11. J. M. Kunkel and T. Ludwig, "Performance evaluation of the PVFS2 architecture," in 2007, pp. 509-516.
12. Camacho, H.E.; Nieto, E.; Anguita, M.; Díaz, A.F.; Ortega, J., "Client cache for PVFS2," Parallel Distributed and Grid Computing (PDGC), 2010 1st International Conference on , vol., no., pp.38,43, 28-30 Oct. 2010. doi: 10.1109/PDGC.2010.5679607.

