A Cloud Enviroment for Backup and

Data Storage

**ABSTRACT**

Currently derived from advances and technological developments can have Input-Output devices ever better able to store more information. The use of the disks of the nodes of a cluster as global storage system is an inexpensive solution for a cloud environment. The need for the available of information from anywhere is increasing; this represents a problem for many users who use applications such as databases, media, personal file, documents, etc. The I/O data demands of these applications get higher as they get larger. In order to improve performance of these applications can use parallel file systems. PVFS2 is a free parallel file system developed by a multi-institution team of parallel I/O, networking and storage experts. In this paper we present the design of an implementation for cloud environment for able to store and back up data through using remote servers that can be accessed through the Internet. The implementation aims to increase the availability of data and reduce in loss of information.

**EXISTING SYSTEM:**

Nowadays due to the growth of technologies, we have different services and applications that allow users to perform tasks that improve productivity in their daily activities. However, the need to access any time and / or from a remote site has grown from a theoretical proposal a genuine need. This has given rise to seek alternatives to solve this problem. One of the alternatives is known by the term cloud computing; cloud computing 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.).

**PROPOSED SYSTEM:**

We propose a cloud environment to backup and store data on a private cloud using PVFS2 like file system for storage data in order to increase the performance of these applications. This option allows input / output parallel, so that will reduce the access times to data. On the client-end, a multiplatform application is developed using free software that allows data transfer fast and simple way.

**Advantage:**

The advantages of this implementation is that it can reuse existing infrastructure (servers, cluster, and other devices) to reduce costs, becoming an advantage compared to having to deploy a private cloud from nowhere.

**PROBLEM STATEMENT:**

The different services and applications that allow users to perform tasks that improve productivity in their daily activities. However, the need to access any time and / or from a remote site has grown from a theoretical proposal a genuine need. This has given rise to seek alternatives to solve this problem. One of the alternatives is known by the term cloud computing; cloud computing 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.).

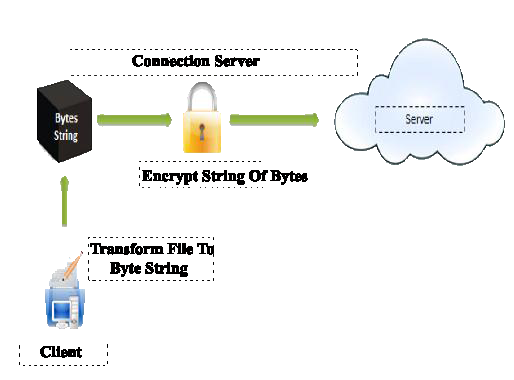
**SCOPE:**

Cloud Computing and cloud storage have become the preferred method for delivering information and online functionality. The implementation of a larger

number of servers time access to data is reduced with respect to the implementation of a single server. This is due to the distribution of data across multiple servers, allowing to access parts of the file parallel. Better still is the fact that it allows to manage in a file different customers because each of them can

manage a part of the file, so further increasing parallelism. This implementation is being developed in order to give users the experience of implementing and managing a private cloud environment that facilitates the backup and data storage, using infrastructure already available or low cost. Avoiding payment of fees or memberships required to contract this service. This allows full control of those who access the information, so maintaining the confidentiality of the data.

**PROCESS:**

****

**MODULE DESCRIPTION:**

# **Number of Modules**

After careful analysis the system has been identified to have the following modules:

1. **Cloud Storage Module.**
2. **Backup File System Module.**
3. **Server Application Module.**
4. **PVFS2 Module.**

**1. Cloud Storage Module:**

A network model where the data is backed up and stored by using an internet connection on remote servers is known by the name of cloud storage, generally are hosted by third parties. Hosting companies operate large data centers, and people who require their data to be hosted buy or lease storage capacity from them. The data center managers often use virtualization to reserve resources according to customer requirements and expose them as storage spaces to store files or data objects. Physically, the resource can extend along multiple servers and multiple locations. The safety of the files depends upon the hosting companies, and on the applications that leverage the cloud storage.

**2. BackUp File SystemModule:**

This application that serves as storage and backup environment in the cloud, we selected as a file system PVFS2 for our cloud environment; to be free and open source, we have the freedom to use and modify according to our needs. So far, some features of the application-level cloud and the benefits to be gained by paying for them have been mentioned. But it is important to know how to actually manage the backup and storage of files within the input / output.

**3. PVFS2 Module:**

PVFS2 distinguishes three different components as it is

1) *Data servers or I/O servers* constitute the PVFS store subsystem. In order to calculate the available store space the system first determines the available space in each data server and then it multiplies the smallest obtained value by the number of data servers.

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

3) *Clients or processing nodes* that access the file system through one of three application programmer interfaces (APIs): a kernel Linux module that allows access 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.

**4. Server Application Module:**

The server application will contain one thread per user, provided by the Thread class in Java, which through Sockets application will contact the client with whom share data and metadata, depending on the rules described in the previous section. The application server will receive data and metadata as strings of bytes encrypted the decrypt, write, replace or delete as applicable, and also send data and metadata in encrypted byte streams to the client application. This

application will also feature a web application which can view and download the files in the directory of each user, and simple steps can make the user account, and update personal information and password. Metadata that will keep are: file

name, size, path, last date modified, deleted mark, this information is stored in a MySQL table.

**SOFTWARE REQUIREMENTS**:

Operating System : Windows

Technology : Java and J2EE

Web Technologies : Html, JavaScript, CSS

IDE : My Eclipse

Web Server : Tomcat

Tool kit : Android Phone

Database : My SQL

Java Version : J2SDK1.5

**HARDWARE REQUIREMENTS**:

Hardware : Pentium

Speed : 1.1 GHz

RAM : 1GB

Hard Disk : 20 GB

Floppy Drive : 1.44 MB

Key Board : Standard Windows Keyboard

Mouse : Two or Three Button Mouse

Monitor : SVGA