Mirror File System for Cloud Computing

Twin Peaks Software

Abstract

The idea of the Mirror File System (MFS) is simple. When a user creates or updates a file, MFS creates or updates it in real time on the local file system and on a file system on a remote server in the Cloud. MFS can also replicate the same updates in real time to an additional server, or to multiple servers, one-way or bi-directionally. The replication operation is transparent both to users and to applications. This paper describes the applicability of MFS to Cloud Computing for users and for service providers.

Introduction

Mirror File System (MFS) is a patented file system developed and implemented on Solaris and Linux by Twin Peaks Software [1][2]. A Windows version is currently under development. The Linux and Solaris versions allow real-time file replication between two systems, one-way or bi-directionally. The design and architecture of MFS allows replication to be propagated to multiple systems. MFS is used in the areas of Disaster Recovery, High Availability, Clustering, Load Balancing and Online File Backup.

This paper consists of two parts, Part 1 describes and explains how MFS can facilitate the adoption of Cloud Computing. For the Service provider, MFS can make Cloud Computing infrastructure more robust, efficient, and available. Part 2 examines similar technologies, such as Storage and Volume replication, Clustering systems, and Clustering file systems, and compares them to MFS. It also explains why these technologies cannot offer the availability, scalability, and efficiency that MFS can.

Technical details on how MFS works internally can be found in the MFS patent [1] and the various technical presentations about MFS in the references [3][4].

Part 1: MFS and Cloud Computing

MFS for Users

The Cloud Computing model shifts both the hardware and software resources used by conventional desktop computing from the local environment to the Cloud environment, including computing power, storage, applications and data. For the user or enterprise considering whether to adopt the Cloud Computing model, the main concerns are data security, reliability, and availability. One of the reasons people are reluctant to adopt Cloud Computing is that, once data is shifted to the Cloud, control also goes with it. MFS addresses this very legitimate concern. By making the transition from Desktop computing to Cloud Computing safe, seamless, easy, and secure, MFS eliminates the “leap of faith” that users must now make in adopting Cloud Computing.

From Local to Cloud Computing

MFS lets the user save the same file on both a local desktop system and a server in the Cloud transparently and in real time. The user does not have to choose between local and Cloud Computing, but rather uses them both at the same time. With the same live file stored on both the local and Cloud systems, the user can continue using the local computing as before and can also take advantage of Cloud Computing. There are many benefits that MFS can provide for this new computing and information model. They include:

- Performance and Flexibility
  Some applications perform better with the CPU power and storage capacity of Cloud Computing, but others may run faster on local systems. With MFS, users have the option of running their applications on the most suitable platform, whether local or Cloud.

- Availability
  If the Cloud is not available or accessible while a user is traveling or in an area of poor Internet access, the user can still use local computing and sync the updated data back to the Cloud later. On the other hand, if a user's computer is not available, data and applications residing on the Cloud can still be accessed with another device and synchronized back to the user's computer later.

- Cloud online backup
  For both user and enterprise, all important files are backed up on the Cloud easily and seamlessly. No tedious manual or scheduled backup is required.

Private Cloud

Although Cloud Computing offers many advantages to the user and enterprise, users and businesses may have doubts about the reliability and security of Cloud Computing. They may be reluctant to entrust their files and access to their applications to remote servers [5]. To ease such concerns, MFS offers a Private Cloud solution.

The Private Cloud resembles public Cloud Computing, but instead of using remote servers to host files and applications, it uses a server located in the customer's home (for example, a Router/Gateway with storage) or an enterprise's local facility.

The Private Cloud is connected to the Internet, but it is located behind a firewall. Users can replicate their data from desktop/laptop to the server in real time, so there are always duplicate copies of their important files within the Private Cloud. A Private Cloud server can still be accessed from the outside, but only by going through the firewall. Private Cloud Computing offers several advantages:

- Better security
  Important private information stays behind the firewall.

- Improved network speed and bandwidth
  Applications and file replication run much faster in a private domain than across the Internet.

- Better control of hardware and software resources

MFS for Service Providers

When service providers build the data center as the basic Cloud infrastructure to provide service to the user and enterprise, they all face the following challenges:

- How to make the service available 24/7/365

- How to handle a disaster and recover from it, or better yet, how to prevent a disaster from happening

- How to scale the service up and down in data centers to meet the dynamic demands of Cloud Computing usage

- How to run the Cloud Computing service most efficiently and economically

Availability

MFS enables all data centers to contain the same data and to be in sync with each other. When a file in one data center gets updated, the same update is replicated to all other data centers in real time. Data centers can be separated by hundreds or thousands of miles. This is what conventional technologies such as clustering and storage replication cannot do.

When one data center becomes unavailable due to natural disaster or human error, other data centers in different geographical regions remain up and running and can continue to provide the same service to users and enterprises.

Performance

When multiple data centers are deployed in the Cloud, service can be balanced and distributed among them to provide better and faster performance. For instance, users can get better performance by accessing the service from the nearest regional data center instead of a single, centralized data center. The best conventional technologies can provide live, real-time data in a single location but not in multiple locations. This single location information model becomes increasingly susceptible to performance bottlenecks as demand on the service increases.

Scalability

MFS enables the service provider to add a new data centers as needed, even in a geographically distant location, while existing data centers continue functioning without interruption. As soon as the new data center finishes synchronizing data with other data centers, it can start to provide the same service to the new region. In case one data center is shut down, due to a disaster, maintenance, or the need to scale down, it does not affect the operation of the other data centers.

Part 2: Limitations of Other Approaches

Any Single Resource is Vulnerable

Our assumption is that to make Cloud Computing more attractive, efficient, and reliable and to protect against single resource failure, service providers must establish redundant resources across multiple data centers. Relying on the single data center model to provide Cloud Computing services makes them vulnerable. This has been an issue since the inception of the IT industry, and the introduction of Cloud Computing makes it even more critical today.

Many technologies have attempted to address the single source of failure issue on different levels. Some of the most widely used are listed here and described in the following sections.

- Disk Array
  This technology addresses the issue on the disk level by providing redundant disks for a single storage system.

- Storage and Volume Replication
  This technology addresses the issue on the storage level by providing redundant storage/volume for a single system or shared storage clustering systems [6] [7] [8].

- Shared Storage Clustering System
  This technology addresses the issue at the system level by providing redundant system nodes that share a storage system.

- Clustering File System
  This technology addresses the issue at the file system level, allowing redundant nodes to mount the file systems on the shared storage system and access the file at the same time.

Disk Array

The scope of disk array solutions is limited to single storage system and is thus not relevant to a discussion of redundant resources or multiple data centers.

Storage and Volume Replication

Storage and volume replication technology, implemented in software or hardware, replicates raw block data from one storage/volume to another that serves as a backup. There are several restrictions in this technology, but we will only discuss the restrictions that limit its effectiveness in providing redundancy for Cloud Computing.

This technology only allows the primary storage/volume to be used [6]. The backup storage cannot be used, even if it is connected to the same computer system. The only time that the backup storage/volume can be used is when the primary one is shut down and the replication stops. The computer system that was using the primary storage/volume then performs recovery procedures, such as `fsck`, mounts the storage/volume, and then starts to use the backup storage/volume. In some configurations, there is another backup server system that is physically connected to the backup storage/volume system. The backup computer system also needs to perform the same `fsck` and *mount* operations on the backup storage before using it. This file consistency checking process (`fsck`) that checks and fixes inconsistencies of backup storage/volume systems can be very time consuming, especially if the primary system or primary storage/volume was shut down abnormally.

Besides the time-consuming recovery procedure, the biggest limitation of storage/volume replication is that the user cannot view and use the replicated data on the backup system right away. If the user and enterprise can replicate their data from the desktop to the storage/volume system in the Cloud, but cannot use it right away, it discourages the use of Cloud Computing services.

In addition, service providers must purchase two storage/volumes but can only use the primary one. The backup storage waits passively in the same or a remote data center but cannot be used. Clearly, this is an inefficient use of resources.

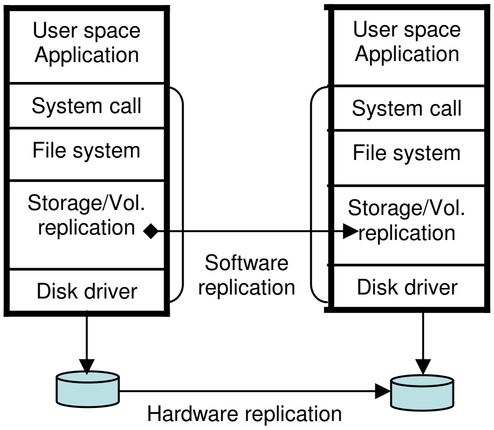

Figure 1: Storage/Volume Replication

Both MFS and storage/volume replication create replication between the storage systems. Why, then, can data replicated to the remote backup storage system by MFS can be viewed and used right away, while data managed by storage/volume replication cannot?

The key is in the file system software module, which manages the data in the storage. The file system software module is the brain of storage. The storage is the plain I/O device, the body. Without the file system software module, the data in the storage is raw and unstructured, and not usable as far as the user is concerned. Storage/Volume replication replicates raw block data directly into the storage/volume without going through the file system software module that manages it. So:

- The file system software module on the backup system has no updated knowledge of the content in the storage/volume.
- Even if the file system software module wants to know what the current content of the storage/volume is by mounting the storage/volume, it cannot find out.

As the primary storage/volume keeps replicating raw block data to the backup storage/volume without going through the file system software module, the file system software module will not have up-to-date bookkeeping information about the backup storage/volume. So, while the mount may succeed at the beginning, as more raw block data is replicated to the backup storage/volume directly, the bookkeeping information of file system software module about the backup storage/volume will become more and more out of sync with the actual data on the backup storage/volume soon. Thus, the result of using the out-of-sync data is unpredictable.

In the OS kernel layer structure, the Disk Replication, Storage Replication, Volume Replication and IP-based Replication are all below the file system module layer. Any replication performed in those layers will have the same restrictions as those of storage/volume replication [Figure 2].

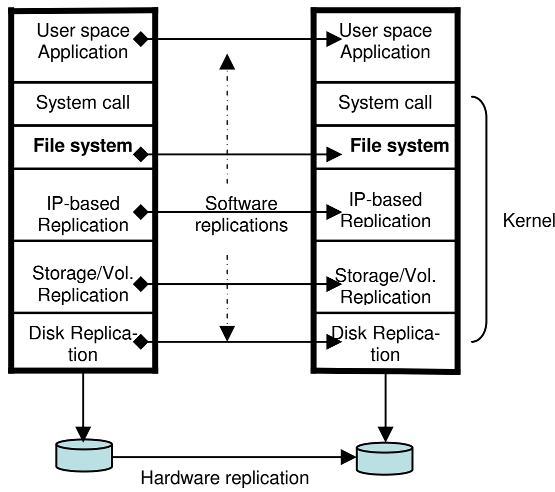

Figure 2: Replication Layers

MFS, in contrast, does replication on the file system software module level. Both primary and backup storage/volumes are already mounted with local file system software modules. All replicated data goes through and is managed by the file system software module, which enables the replicated data/file to be viewed and used by the backup system immediately.

For the user, the file replicated from the Desktop/Laptop to the server in the Cloud is live and ready to use. For the service provider, MFS enables the same copy of a live file to be updated in both, or in multiple, data centers. This MFS capability greatly enriches the value and reliability of any Cloud Computing service.

Shared Storage Clustering Systems

Conventional shared storage clustering systems provide redundancy at the system level to protect the clustering system from single system failure.

A two-node clustering system consists of one primary node and one backup node, both of which are physically or virtually connected to a storage system; however, only one node, the primary node, can access the storage system. The backup (passive) node cannot mount its file system software module on the shared storage while the primary node is accessing the storage through its file system software module because the storage is being updated by the primary node, and the primary system node does not inform the backup system node about what data is being updated to the shared storage. The only time the backup system can mount its file system software module on the shared storage is when the primary node is shut down and stops updating to the shared storage. When that happens, the backup node mounts the file system software module on the shared storage and takes over the operation. This is called a failover operation in the clustering system. The clustering system and its shared storage system are normally connected physically and located in one physical data center.

Shared storage clustering systems do not provide any benefits to Cloud Computing users, who still have to choose either local computing or Cloud Computing. For the service provider, using two systems to perform one service is neither efficient nor economical. The fact that the whole Clustering system normally resides in one data center and cannot provide redundancy across multiple data centers in the Cloud is a serious disadvantage in the event of a disaster, such as an earthquake, which would affect all the systems in a data center.

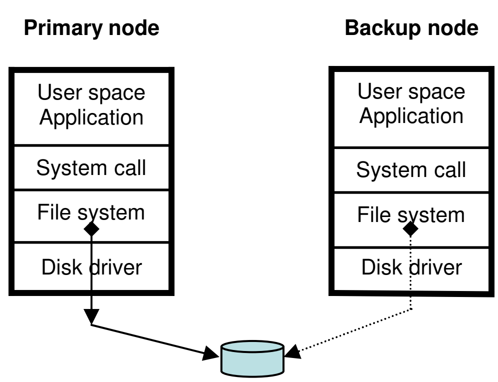

Figure 3: Shared Storage Clustering

Shared Storage Clustering File System

A Shared Storage Clustering File System enables all participating nodes to mount the shared storage and access the file on it at the same time [9] [10]. This is a step forward compared to the conventional clustering system, in which only one node can mount and access the file on the shared storage. Unfortunately, these benefits are not extended to users of Cloud Computing. Shared storage clustering does allow providers to run Cloud services more efficiently within one data center, but cannot go beyond that.

Mirror File System

MFS has two unique features that enable it to provide benefits for Cloud Computing users and service providers.

- Replication at the File System Software Module Level
  This feature allows file replication between two separate systems in real time. A user can have two copies of the same file on two systems, one on the local desktop, the other one on the Cloud server. This helps to make transition or migration from local computing to Cloud Computing easy and seamless. For the service provider, copies of the same file can be replicated in multiple data centers in real time, which helps to make the operation of the Cloud service very efficient and always available.

- No Shared Storage
  MFS does not restrict the storage systems to be shared and confined physically to one data center, the service providers can scale the operation up from one data center to multiple data centers easily as needed and vice versa. Each data center has its own system and its own storage, but contains information identical to that of the other data centers. The distributed capability of MFS makes the Cloud Computing service much more reliable and efficient.

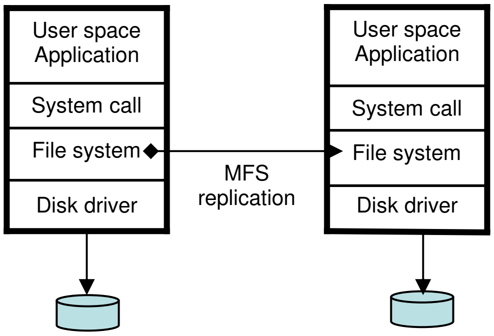

Figure 4: Mirror File System Replication

Conclusions

MFS's unique ability to use the file system software nodule to replicate live files between geographically distant systems in real time distinguishes it from conventional storage/volume replication and redundant clustering technologies. It allows the user to have an identical copy of the same file on both the desktop system and the remote system in the Cloud simultaneously and in real time, so that users can freely use both Desktop computing and Cloud Computing at the same time. More importantly, it enables Cloud Computing service providers to run multiple, geographically distant data centers that contain identical data and synchronize with one another in real time across the Cloud. This new information deployment model enables them to run their services more reliably, robustly, and efficiently.

References

[1] U.S. Patent number 7418439.

[2] Twin Peaks Software Inc. http://www.TwinPeakSoft.com

[3] MFS presentation at usenix.org fast 08, http://www.usenix.org/events/fast08/wips_posters/slides/wong.pdf

[4] MFS presentation at Connectathon.org, February 2007: http://www.connectathon.org/talks07/twinpeaks.ppt

[5] "Don't buy the hype: Firms not ready for cloud storage", Jan. 26, 2010, Sam Narisi, Docucrunch: http://www.docucrunch.com/dont-buy-the-hype-firms-not-ready-for-cloud-storage

[6] EMC Symmetrix Family: http://www.emc/products/family/symmetrix-family.htm

[7] Veritas Volume Replicator: http://www.symantec.com/business/volume-replicator

[8] DRBD, Disk Replication Block Device: http://www.drbd.org/home/mirroring

[9] Veritas Storage Foundation Cluster File System: http://www.symantec.com/business/storage-foundation-cluster-file-system

[10] Red Hat Global File System: http://www.redhat.com/gfs/