Survey on High-Performance Distributed File System for Microsoft Windows Server 2012

Prajakta Gosavi[1], Ankita Doiphode[2] , Prof. Vinayak Shinde[3]

Research scholars, Dept. of Comp, L. R. TIwari College of Engineering, Mira Road (E), Thane, Mumbai University,

India[1,2]

Assistant Professor, Dept. of Comp, L. R. TIwari College of Engineering, Mira Road (E), Thane, Mumbai University,

India[3]

ABSTRACT: Distributed file System is a collection of client and server services that allow an organization using Microsoft Windows servers to organize many distributed Server Message Block file shares into a distributed file system. The goal of a distributed file system is to allow users of physically distributed computers to share data and storage resources by using a common file system.General configuration for a DFS is a group of workstations and mainframes which are connected by alocal area network (LAN).
This research is about Microsoft's implementation of DFS. For general discussion of the concept and other implementations. It mainly focuses on transparency, fault tolerance, and scalability and discusses them in the context of DFSs. This paper is study of the present state of design and implementation of distributed file systems.
It is divided into parts: an Introduction of DFS, implementation,architecture and component of DFS ,their design goals for better distributed file system and case study of Microsoft windows server 2012.

KEYWORDS: Distributed file System; Microsoft Windows servers; DFS Management; DFS types.

I. INTRODUCTION

The goals of information technology groups in small as well as large organizations is to manage file servers and their resources efficiently while keeping them available as well as secure for users. As organizations developed they are able to include number of users and servers—whether they are located in one site or either in geographically distributed sites—administrators find it difficult to keep users connected to the files which they need. Instead of that, storing files on distributed servers makes files available to more users as well as also decreases latency and bandwidth use when the servers are located near users[1]. As the number of distributed servers increases, users have facing problem in locating files they need, and operational costs increase. The DFS solution in the Microsoft® Windows Server™ 2012 operating system helps administrators address these challenges by providing two technologies named as DFS Namespaces and DFS Replication, when used together, offer simplified, fault-tolerant access to files, load sharing, as well as WAN-friendly replication.
This research also illustrated the Distributed File System DFS Replication functionality that is new or changed in Windows Server 2012 and the DFS Replication and DFS Namespaces functionality that was new or changed in Windows Server 2012.

II. RELATED WORK

In [2] authors describes how do computational scientists access files in a widely distributed computing environment.And what needs to be provided to enable them to access files easily and be productive in these environments? They provide one data point from the NSF TeraGrid towards answering the first question and propose a system that believe provides a simple answer to the second question.In [3] TidyFS exposes data to clients using a stream abstraction.A stream is a sequence of parts, and a part is the atomic unit of data understood by the system. Each part is in general replicated on multiple cluster computers to provide fault-tolerance. A part may be a single file accessed using a traditional file system interface or it may be a collection of files with a more complex type—for

example, TidyFS supports SQL database parts which are pairs of files corresponding to a database and its log. The operations required by TidyFS are common across multiple native file systems and databases so this design is not limited to Windows-based systems.

III. IMPLEMENTATION RESULTS

DFS architecture is client-server architecture,Design of a computer network in which many users requests and receive service from a central server i.e host computer. Client computers establishes connection to allow a computer operator to request services of the server and show the results the server returns. Servers wait for requests to arrive from users and then respond . Ideally, a server provides a standardized transparent interface to clients so that clients need not be aware of the particulars of the system (i.e., the hardware and software) that is providing the service. Nowadays clients are often located at workstations or on computers, while servers are situated elsewhere on the network, generally on more machines.
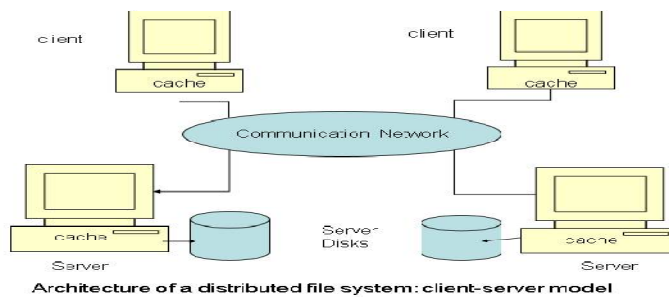

Fig.1. DFS Architecture[4]

Distributed File System is implemented as a role service of the File Services role. DFS technologies in Windows Server 2012 provide a easy way to access geographically dispersed files throughout an organization. DFS able to provides WAN-friendly file replication between servers. Technologies provided with DFS include as:
•DFS Namespaces
•DFS Replication
•Remote Differential Compression
As we know that DFS Namespaces and DFS Replication are separate technologies, so both can be used together to provide high availability as well as redundancy of data. The following process illustrates how DFS Namespaces and DFS Replication work simultenously:
1. User accesses folder in the configured namespace.
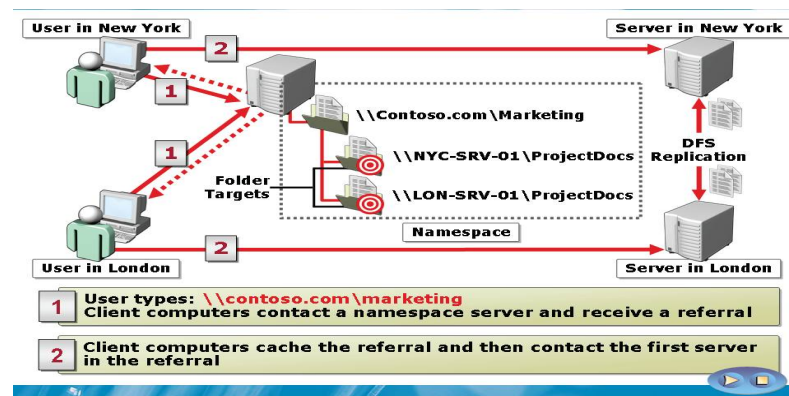2. Client computer accesses the first server in the referral

Fig.2. Working Model[5]

IV. TYPES OF DISTRIBUTED FILE SYSTEM

NFS/CIFS

Network File System (NFS) is mainly used by Linux/Unix, while CIFS (Common Internet File System) is used by MS Windows. NFS is a distributed file system which developed in 1984 by Sun Microsystems. NFS can be able to usedto share and use files from other hosts via a network. In 2003, NFSv4 was released, and
Presented better performance as well as safety. Instead of that CIFS is a distributed file system Microsoft applied to MS Windows, which was developed based on Server Message Block but with improved functionalities.CIFS is a distributed file system Microsoft applied to MS Windows, which was
developed based on Server Message Block (SMB) but with improved functionalities. NFS and CIFS comply with the POSIX standard. Thus, applications by using NFS or CIFS can be use a distributed file system like a local file system. This means, when you implement or run an application, not necessary to prepare a local file system and distributed file system independently.

HDFS (Hadoop Distributed File System)

Google developed Google File System (GFS), is an unique distribution system, which kept information about webpages monitored by Google. Google published a paper on the Google file system in 2003. Hadoop Distributed File System is an open source system which was developed using GFS as a reference. For this reason, HDFS has the same features as Google File System. HDFS divide a huge file into portions, and kept three of them into every data node. In other words, one file is store in several distributed data nodes. This also means that one file has three copies. The usual size of a chunk is 64 MB. The metadata about which data node stores the chunk is stored in the NAMENODE. This lets you to read data from distributed files and do processes by using MapReduce.

V. DFS FEATURE ROLE IN WINDOWS SERVER 2012

DFS Namespaces which is also called as DFSN or DFS-N and DFS Replication that is DFSR or DFS-R are important role services in the Windows Server 2012 File as well as Storage Services role.
The Distributed File System (DFS) in Windows Server 2012 is comparatively very unrelated from the previous versions of Windows. First of all, it is important to note that DFS Replication and Namespaces can be implemented which are independent of each other, SO basically as these serve two different purposes[6]. Below is a quick overview of DFSN and DFSR:
 DFS Namespaces — Allows you to group shared folders that are located on different servers into one or more logically structured namespaces. Every namespace appears to users as a single shared folder with a group of subfolders. However, the defined structure of the namespace can consist of various file shares that are located on independent servers and in multiple sites.
 DFS Replication — Allows you to efficiently replicate folders that means including those referred to by a DFS namespace path across multiple servers as well as sites. DFS Replication uses a compression algorithm known as remote differential compression (RDC), which able to detects changes to the data in a file, and it enables DFS Replication to replicate only the modified file blocks instead of the entire file.

REFERENCES

1.   Weil, Sage A, Scott A. Brandt, Ethan L. Miller, Darrell DE Long, and Carlos Maltzahn. "Ceph: A scalable, high-performance distributed file system." In*Proceedings of the 7th symposium on Operating systems design and implementation*, USENIX Association, pp. 307-320,2006.
2.   Walker, Edward. "A Distributed File System for a Wide-Area High Performance Computing Infrastructure." In *WORLDS*, vol. 6, pp. 9-9, 2006.
3.   Fetterly, Dennis, Maya Haridasan, Michael Isard, and Swaminathan Sundararaman. "TidyFS: A Simple and Small Distributed File System." In*USENIX annual technical conference*, pp. 34-34, 2011.
4.   Yadav, Jitender Singh, and Mohit Yadav Ankit Jain. "Distributed File System." *International Journal Of Scientific Research And Education* 1, no. 06,2014.
5.   Controllers, Using Read-Only Domain, "Part 1 Windows Server 2008 Administration Fundamentals." ,2008.

6.   Levy, Eliezer, and Abraham Silberschatz, "Distributed file systems: Concepts and examples." ,*ACM Computing Surveys (CSUR)* 22, no. 4, 321-374,1990.

BIOGRAPHY

Prajakta Gosavi is a Research scholar in the Computer engineering Department, L. R. Tiwari College of Engineering, Thane. Her research interest is in parallel, cluster computing, grid computing, operating system etc.

Ankita Doiphode is a Research scholar in the Computer engineering Department, L. R. Tiwari College of Engineering, Thane. Her research interest is networking, database etc

Prof. Vinayak Shinde is Assistant Professor in Computer engineering Department, L. R. Tiwari College of Engineering, Thane. His research interest is in Parallel Computing.