

## Research of Distributed Systems on Local Area Network

Wenliang Qu<sup>1</sup>, Fumei Liu<sup>2</sup>

<sup>1</sup> Beihua University, Jilin, Jilin, China, 132013

28941577@163.com

**Keywords:** LAN; Distributed System; Data Storage

**Abstract.** Large-scale distributed storage system has the advantage of high scalability, has become a research and development direction of the storage areas and has a wide range of applications in the major internet companies and corporate data centers. Due to complex function of large-scale distributed storage systems, a sound management system is the realization of storage systems management, an important prerequisite for highly reliable operation.

### Introduction

Today, humanity has entered the era of big data, the amount of data is growing at an unprecedented rate surge in the amount of data generated by individuals and companies have to "terabytes" (trillion) as a unit to calculate, but also to more than the rate of Moore's Law continues to grow. The amount of data analysis research agency IDC released the latest report "Extract Value from Chaos" [1] show that the amount of data around the world will double every two years, in 2011 to create and copy of 1.8ZB (equivalent to 1.8 trillion GB), this figure is equivalent to have to do the world's 215 million per day times the amount of data of high resolution magnetic resonance imaging (MRI) scans produced; the equivalent of China everyone send three micro-blogging per minute and to continuously It continued 6217 years. And this growth is accelerating in 2015 to reach nearly 8ZB.

With the explosive growth of data volumes, large data applications are becoming more common. Compared with traditional applications, large data storage system applications put forward higher requirements:

**High capacity.** Many current applications are required to have a storage capacity of PB level, including digital libraries, seismic data processing, oil exploration, the storage system must be able to provide high-capacity storage space to meet application demand for storage space.

**High performance.** For like Baidu, Google this processing large amounts of data, and there are strict requirements of real-time applications, access the delay must be controlled within a certain range, and also requires storage system has a high throughput.

**Scalable.** Many applications in the course of all need to continue to expand its storage capacity, storage systems need to be able to without interrupting the running of the business premise of achieving online to join and manage more storage devices.

**Sharable.** In the background of big data, massive data replication and migration need to spend a great price, and the cost of the redundancy costs of large data storage system is also high, in order to avoid the unnecessary expense, storage systems need to ensure that data is shareable.

**Easy to manage.** With increasing storage system capacity, CPU, memory, and disk and other hardware devices will be more and more, so in order to facilitate maintenance and management, design of the storage system must follow the principle of easy management.

### Development Status of Distributed File System

To meet these new requirements big data applications, mass storage for distributed file system has been more and more attention, and now has set off a worldwide boom in research and development of distributed file systems. Distributed File System are the Client / Server architecture, data stored in the server, the client can be like access the local file system, files located on remote servers accessed through a computer network. Typically, a distributed file system uses a separate

server or server cluster on the entire file system metadata to manage, because it can not only improve system scalability, we can also ensure that the system has a high in the provision of services availability. Additionally, the metadata server should also save the mapping between the actual storage address files and documents, and provides access to the file management. Typical distributed file system such processes are generally used to access data: First, the client data to the metadata server or file management server sends a request to obtain documents or objects are stored in a data distribution server; then, according to data from the client yuan position information on the file server to obtain the requested data directly to the data server; Finally, the data server does not go through the metadata server transfers data directly to the client. In a distributed file system, the client needs with the metadata server and data servers to interact, so that the client needs to install the appropriate software modules.

### **Typical Distributed File System**

To meet these new requirements big data applications, mass storage for distributed file system has been more and more attention, and now has set off a worldwide boom in research and development of distributed file systems. Distributed File System are the Client / Server architecture, data stored in the server, the client can be like access the local file system, files located on remote servers accessed through a computer network. Typically, a distributed file system uses a separate server or server cluster on the entire file system metadata to manage, because it can not only improve system scalability, we can also ensure that the system has a high in the provision of services availability. Additionally, the metadata server should also save the mapping between the actual storage address files and documents, and provides access to the file management.

PVFS (Parallel Virtual File System) is Clemson University, developed high-performance, scalable, distributed parallel file system. It requires no special hardware device or kernel support, it can run directly on an ordinary PC. Currently, many popular distributed file systems have draws PVFS design. Wherein the management node is responsible for maintaining file metadata information, and provides access to the file validation feature. The so-called meta-data files, file refers to the file name, file location directory hierarchy, the file's owner and the file in each node of distributed file systems and other information distribution. Management node realizes PVFS file metadata information atomic operations, it avoids the use of those complex locking mechanism to ensure the consistency of concurrent access to process information in the metadata. I / O node is responsible for storing the real PVFS files. In fact, I / O nodes are the PVFS files stored in the local file system, so I / O nodes can, through the use of traditional read, write commands to access those files. This means that the I / O nodes can use any local file system to store PVFS files, such as EXT3, EXT4, ZFS etc. In addition, I / O nodes can also RAID [15] or by using hardware RAID software to achieve PVFS disk level fault tolerance.

Lustre file system developed at Carnegie Mellon University, object-based storage of large-scale distributed file system. Lustre Cluster name by Linux and evolved, is a new distributed file system to solve the problem of mass storage designed to support tens of thousands of client systems, PB-class storage capacity, hundreds of GB of polymerization I/O throughput. It is the world's most widely distributed file system, mainly deployed in high-performance computing cluster. Lustre file system also consists of three parts, two MDS (Meta-Data Server), multiple OST (Object Storage Target) and several client. MDS is the Lustre file system is the most complex part of which offer global namespace, and is responsible for managing access to directories and files. MDS metadata operations are based transaction, and extensive use of existing capabilities log file system. In addition, MDS is also responsible for maintaining data consistency throughout the file system and respond to requests Client. To avoid single points of failure, Lustre provides two MDS, MDS occurs when one fails, the other can take over the function of the MDS. Lustre file system is the most important concept is that object storage, OST is responsible for the storage and management of object metadata and object data. Metadata objects, including the length of the object's data blocks and objects, which is the traditional inode metadata is very similar. Data objects are divided into

strips, stored on multiple OST, when an OST fails, can still ensure that the data stored on the OST is not lost.

GoogleFS is Google company in order to meet their application needs and the development of large-scale data-intensive applications-oriented, scalable, proprietary distributed file system based on Linux, and can be applied to large files, sequential read, not modified, high concurrency storage scenarios. It runs on inexpensive common hardware devices, and provides effective disaster redundancy capability and can provide a high-performance storage services to a large number of clients. In addition, GoogleFS also uses a loose consistency model, not only good support highly distributed applications, but also relatively simple to implement and efficient. GoogleFS now fully meet Google's demand for storage and as a storage platform has been widely deployed within Google.

## **The Management System Design of BlueOcean**

**The Overall Architecture Design.** BlueOcean management system is designed to easily and efficiently configure, monitor and manage the various components of the distributed file system, but in fact there are a variety of interaction between the various components of the internal nodes distributed file system, including metadata interactive nodes and data between nodes, interactive data nodes and data between nodes. If the system administrator operating management system resulted in a considerable impact on the internal interaction of distributed file system itself would be more harm than good, so in order to minimize the impact due to the operation and management system produced interactive distributed file system of internal nodes we can use a centralized control and management solutions.

**Configuration Management Design.** Configuration management is the core BlueOcean management system, its function is, the backup node metadata node or data node is added to the specified BlueOcean distributed file system, and then install the appropriate metadata node program or data node program, When the node program is installed, an administrator can make the appropriate administrative actions, such as stopping the node processes or delete nodes.

Data node before making configuration management does not belong to any of a distributed file system, distributed file systems to any one can put it into his own part, we call this state data node for backup, to be in this data node status is called backup node. Data node in most cases as part of a distributed file system uptime and achieve its function, we call this state run. Become operational status from the reserve node node must go through to build the node configuration file, then the data node executable copy to the node and start the process of data node process. In fact this process do ones, and the time it takes too long, in order to balance the junction time switch from one state to another state, we will process this big divide into two smaller processes. One is the configuration process, is the control node for generating a corresponding data node configuration file, we call node status at this time is already configured; the other is to install and start the process of data node executable process. In fact, doing so can also reduce the waste of time due to mis-configuration caused on. In such a distributed file system, data node has a high crash rate, it must provide a down state, be said that as a result of abnormal data node process can not be used. In addition to the data node process can not be used due to an exception, as well as a system administrator for some special reasons the initiative to close the case data node process, such as the need to maintain the system software update data on the node, providing services data excessive number of nodes, etc., and this time the system administrator is able to take the initiative to stop the data node process, in order to distinguish between state and downtime, we offer a Stopped state. In summary, we provide data node following five states: reserve, has been configured, run down and stopped.

**Status Monitor Design.** Condition monitoring is another important part of BlueOcean management system, whose main task is to collect, process useful information monitor objects and user-friendly way to show to the system administrator, perform configuration management operations to provide the basis for the system administrator, such as , if found insufficient storage space, the administrator needs to add nodes to perform the operation or to add disk operation; If the disk is damaged, the administrator needs to perform delete disk operations; data node If there are

problems, administrators need to delete node operation. Status Monitor to monitor levels of different objects can be divided into three categories, system-level status monitoring, status monitoring node level and disk-level status monitoring.

## **Conclusion**

In configuration management, analysis and design of our discovery node, configure the node, the node installation, updating nodes, stop nodes, delete nodes, add a disk and delete disk operations specific process, then in terms of condition monitoring, from system-level status monitoring, status monitoring node level and disk-level status monitoring and analysis of the three aspects need to be monitored and how to obtain information on the status of these contents.

## **References**

- [1] John Gantz, David Reinsel. Extracting Value from Chaos. Idc. June 2011.
- [2] Frank B. Schmuck, Roger L. Haskin. Gpfs: A Shared-Disk File System for Large Computing Clusters. Proceedings of the Conference on File and Storage Technologies, January, 2002:231-244.
- [3] Feiyi Wang, Sarp Oral, Galen Shipman, Et Al. Understanding Lustre Filesystem Internals. April 2009.
- [4] Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung. The Google File System. Proceedings of the 19th Acm Symposium on Operating Systems Principles, October 19-22, 2003.
- [5] Brent Welch, Marc Unangst, Zainul Abbasi, Et Al. Scalable Performance of The Panasas Parallelfiler System. Proceedings of The 6th Usenix Conference on File and Storage Technologies, P.1-17, February 26-29, 2008, San Jose, California.