High Availability Server Solution for Medium-Scale Television Program Production Networks: Postprint

Authors: Guan Leilei

Date: 2023-10-08T00:00:00+00:00

Abstract

This article first analyzes the concept of network server high availability, secondly constructs a medium-scale program production network model within a realistic environment by reviewing past practical experiences, and finally conducts a comprehensive analysis and evaluation of the characteristics and performance of two availability solutions—SANergyHA and MSCS cluster services—through two common server types in such network configurations, for peer reference.

Full Text

High Availability Solutions for Servers in Medium-Scale Program Production Networks at Television Stations

Guan Leilei (Technical Bureau, China Media Group, Beijing 100859)

Abstract: This article first analyzes the concept of network server high availability. By reviewing past implementation experiences, it constructs a real-world medium-scale program production network model. Finally, through examination of two common server types in such network configurations, it provides a comprehensive analysis and evaluation of the characteristics and performance of two availability solutions: SANergyHA and MSCS cluster services, offering a reference for industry peers.

Keywords: television station; program production; server; SANergy system; SQL Server

Classification: TN948.4
Document Code: A
Article ID: 1671-0134(2021)01-110-03
DOI: 10.19483/j.cnki.11-4653/n.2021.01.034

Citation Format: Guan Leilei. High Availability Solutions for Servers in Medium-Scale Program Production Networks at Television Stations [J]. China Media Technology, 2021(01): 110-112.

Against the backdrop of rapidly evolving science and technology, advanced technologies such as network interconnection and virtual storage have been continuously developed with increasing maturity. Network-based production, as a new model, has been gradually adopted by television stations, and the probability of implementing high network availability has consequently increased. A key question that industry practitioners continually investigate in practice is: what factors influence server availability? [1] From a logical functional perspective, numerous types of servers are available on the market today, such as those providing mainstream network services. In the television industry, networked program production represents a major future trend. Compared with previously established distributed independent production systems, network-based production systems improve work efficiency, creative capabilities, and production flexibility during operation, but impose more stringent requirements on system comprehensive availability. Server high availability largely determines overall network availability, and this article focuses its analysis on this critical factor.

1. Server High Availability

In the industry, network availability is more commonly defined as the ability to maintain normal operation and provide robust designed functions. Using time as the unit of measurement, the proportion of time that a network operates normally within a given cycle serves as an important metric for evaluating its availability. From this perspective, network availability reflects its capacity for continuous operation. Television program production demands high timeliness and work efficiency, making this continuous operational capability essential for supporting efficient business processes. Macroscopically, two types of events affect whether a network operates normally: planned events within network management control, such as scheduled downtime for inspection and maintenance, system structural adjustments, and functional testing; and unplanned events outside management control, such as unexpected failures and repairs during network operation. Comparatively, the latter poses greater potential damage to comprehensive network availability due to its unexpected and unpredictable nature.

Servers constitute critical network infrastructure. Unlike most workstations on the network that provide user operation interfaces, backend servers deliver functional services that better meet user requirements. If servers stop operating, network availability disappears entirely even when other network components remain intact. Based on the above discussion, it is evident that server availability directly impacts overall network availability, indicating that improving server availability enhances comprehensive network availability.

Within server clusters, two primary component types exist: cluster constituent services and usable resources within the cluster. Physical disks, IP addresses, and service programs are all elements that compose resource groups. Different resources within a resource group can form dependency relationships, which represent an important concept in cluster technology and determine the online/offline sequence of different resources within the group (see Figure 1 [Figure 1: see original paper]).

Regarding common availability issues across different server types, we can analyze them from the perspective of elements required to maintain availability properties: one element is the application, which refers to the service provider itself; the other is data, which corresponds to the content that the service needs to process.

2. Medium-Scale Program Production Network Model

Before exploring high availability solutions for servers, we construct a model of a medium-scale program production network for a television station for investigation purposes. This article uses a program production network solution implemented by a municipal-level television station in July 2004 as the model to analyze its basic network topology. [2]

The network model encompasses both fiber optic and Ethernet loops. Ethernet connects all stations and manages users and information transmission processes within the network. The fiber optic network is configured with a Storage Area Network (SAN) structure, where both the central storage system and workstations with video/audio codec cards connect directly to the fiber optic network switching infrastructure, enabling high-quality material access. Workstations with codec cards handle the primary audio and video editing tasks, while cardless workstations provide functions such as voice-over, subtitle presentation, and material browsing, effectively reducing the operational burden on card-equipped stations.

Program materials are allocated to card-equipped workstations with video/audio codecs, all connected to Ethernet via fiber optics. Materials are uploaded and stored in the S2A8000 storage system and shared SCSI disks at different quality levels (high and low). After upload completion, any card-equipped station can retrieve and edit program materials, maintaining a unified editing interface regardless of whether editing high- or low-quality materials. Edited program files can be played directly or output after packaging.

In this network, the Metadata Controller (MDC) plays the role of shared storage owner. From the MDC' s perspective, shared storage appears as its local disk, manageable and configurable through disk management software. For other workstations on the network, shared storage is accessed as a mapped disk through Ethernet sharing. When workstations access shared storage based on logical volumes, they must use the MDC to capture metadata about the logical volumes, such as file names and Access Control Lists (ACLs), while actual data

transmission occurs via fiber optics. This explains why Fiber Channel (FC) provides faster, more stable bandwidth compared to Ethernet. [3] The MDC can be considered a special type of server with high status but low I/O traffic, necessitating separate physical layer configuration.

In the constructed network model, using SANergy HA high-availability software effectively facilitates achieving MDC high availability. The SANergy file system has built a close relationship with the network operating system. Although not purely application software, SANergy HA belongs to the same product family. [4] Its installation and configuration processes are straightforward: simply grant ownership of all logical volumes to the primary MDC, install SANergy HA software on the standby MDC, and configure the logical volume standby MDC settings.

During operation, SANergy HA software on the standby MDC dynamically monitors the primary MDC's status. Once the primary MDC fails unexpectedly, the software automatically transfers ownership of logical volumes to the standby MDC while adjusting disk mapping conditions on the workstation side to ensure accurate targeting. This process is transparent to users, with a time cost of approximately 1 minute. From the user' s perspective, if the primary MDC fails unexpectedly, workstations accessing shared storage will automatically report errors but can resubmit access requests after a brief wait, returning to normal operation. This MDC construction approach thus provides high availability.

Comprehensive analysis of multiple factors suggests that in current network models, implementing MDC high availability using SANergy HA software represents a practical solution. However, efficiently applying other high-availability software to enable MDC high availability may expose issues when combined with SANergy, such as lack of transparency to clients and certain time delays in fault recovery. Nevertheless, some network integration vendors have recently made efforts and attempts in this area, such as enabling MDC high availability with MSCS cluster assistance, which holds significant practical value for repeated testing before actual deployment.

3.2 Using MSCS Clustering to Improve SQL Server Availability

MS SQL Server is a database software developed by Microsoft that primarily provides a universal data interface to client applications, precisely separating core data from applications to maintain independent, stable, and timely upgradeable operation states. In our network model, SQL Server mainly manages program data and user management information within the editing system—both essential data types for maintaining normal network operation. Although SQL Server does not store program data but only manages it, database failures render data on storage systems unavailable to clients, interrupting the connection between client applications and data.

Database servers occupy an important position in networks, a fact that has been

undeniable for a long time. However, professionals in the broadcasting and television industry have not historically attached sufficient importance to availability requirements. Previous network implementations generally used SQL Server' s built-in backup function for timed intelligent database backups. While this method improved availability to some extent, it could not determine automatic backup frequency—overly short intervals increased network load and introduced instability, while overly long intervals risked untimely recovery of large data volumes. In the VDE network model constructed in this article, we continue using the data backup method to periodically transfer data to data stream tapes, compensating for data security shortcomings exposed by the clustering method.

MSCS is a cluster service function developed by Microsoft that can be considered a relatively standard high-availability software platform. For single cluster configurations between two servers sharing the same SCSI disk array, MSCS provides robust support. The dedicated private network configured between them monitors resource application status, while both connect to external public networks to provide access services. [5]

In our network model, SQL Server is essentially a service program built on an MSCS cluster. The SQL Server virtual server possesses a unique network name and IP address, which clients use to access the database via the SQL Server server and external public network. The MSCS cluster is configured in active-standby mode, with the primary node synchronously controlling the SQL Server resource group, which runs on this node. If this node fails and renders the resource group unavailable, MSCS cluster management software automatically transfers the resource group to the standby node. When the primary node recovers, MSCS cluster software moves the resource group back to the primary node according to user settings. Post-failure switchover time is approximately 1 minute, with SQL Server program startup accounting for over two-thirds of this time—highly consistent with cluster technology principles described earlier. [6]

In real-world environments, after numerous simulation tests and observations, SQL Server cluster performance proves highly effective, reducing switching failures and preventing normal recovery due to cluster errors. When network system failures trigger switchover, clients accessing databases will automatically report errors but resume normal operation upon repeating the operation process, requiring no manual intervention.

In constructing the network model, the selected digital compression encoding format determines storage capacity requirements. Under MPEG-2 I-frame 50Mb/s bit rate, one hour of program material storage reaches 22.5GB. Although this storage capacity is substantial, database size remains highly controllable—typically not exceeding hundreds of MB after proper optimization, with dozens of accessing workstations. Both database size and access requirements are relatively small compared with conventional LAN databases. Configuring a two-node SQL Server cluster with only a single system providing services at any given moment offers sufficient support. Under such network model conditions,

MSCS clustering based on SQL Server demonstrates excellent effectiveness in helping servers achieve high availability.

3.3 Configuring Array Cards and Disk Arrays

The configuration method can be summarized as follows: power on the server and, following on-screen prompts during self-testing, enter the server's NetRAID Express Tools to configure the array card. Key procedures and considerations in practice include: [7]

1. During array card and disk array cabinet configuration, both servers cannot simultaneously connect to the disk array. Only after installing cluster software on the servers can both simultaneously connect to the array cabinet.
2. The two SCSI IDs must not be identical.
3. After modifying ID-related parameters, servers must be restarted before entering HP NetRAID Express Tools to configure RAID.
4. When enabling the ID on the first array of Server A, select "New configuration" ; when enabling the ID on the second array of Server B, be sure to select "View disk configuration" to ensure consistency between array card configuration parameters.
5. If the cluster needs to manage multiple application services, logical disks must be created during ID configuration. These can be identified by the operating system as physical hard drives and thus serve as independent resources within the cluster.

Conclusion

Based on the discussion of server high availability concepts, this article constructed a real-world network application environment model and conducted a detailed analysis of the characteristics and practical application considerations of high-availability solutions for MDC servers and database servers. As users, in practice we should select the most suitable solution by comprehensively considering actual requirements, solution characteristics, and constraints imposed by objective conditions. By acknowledging problems exposed during implementation, adjusting solutions accordingly, employing reasonable methods to avoid issues, or using alternative approaches, we can make greater contributions to achieving server high availability.

References:

[1] Xiang Jiemin. Design and Research of High Availability Systems [J]. Computer Programming Skills & Maintenance, 2020(07): 33-34+71.
[2] Zheng Jinxiu, Wang Qiang, Wang Wanyu. Design of Virtualization Platform for Satellite Receiving Station Monitoring and Management System [J].

Electromechanical Technology, 2020(03): 36-38+55.
[3] Liang Xiujing. Siemens Sinamics Connect 300 Expands New Intelligent Functions [J]. Light Industry Machinery, 2020, 38(01): 84-85.
[4] Wu Xiang. Research and Solution on Oracle Cluster Technology for Measurement Information System [J]. China Inspection and Testing, 2019, 27(06): 19-22.
[5] Su Jian, Zhang Jinmeng, Li Jidong. Design and Implementation of Server Operation Monitoring System Based on Mobile Terminal [J]. Computer Knowledge and Technology, 2019, 15(33): 14-16+26.
[6] Shen Fengxia, Yang Weijie, Mu Linsen, et al. Application of Server Virtualization in Hospital Informatization Construction [J]. Medical Information, 2019, 32(21): 20-21.
[7] Luo Yunjun, Yao Zhaobao. Application of High Availability Dual-Machine Hot Backup Technology in News Production and Broadcasting Network System [J]. Digital Communication World, 2019, 14(07): 20+68.
[8] Rao Dong. Application of Server Cluster Technology in University Office Environment [J]. Electronic Technology & Software Engineering, 2019, 26(09): 169-170.

Author Biography: Guan Leilei (1981-), female, Beijing, intermediate engineer, research direction: radio and television program directing.
(Responsible Editor: Zhang Xiaojing)

Note: Figure translations are in progress. See original paper for figures.