Advances in Computer Science Research (ACRS), volume 54
International Conference on Computer Networks and Communication Technology (CNCT2016)

Get Rid of Data Disaster: An Active-active Model with VRN Network for Smart Grid DFS

Jun YU[1], Xi-gao LI[2,*] , Heng-mao PANG[3], Min BU[4] and Lin QIAN[5]

[1]NARI Group Corporation, Nanjing, China,

[2] State Grid Electric Power Research Institute, Nanjing, China,

[3, 4, 5] NARI Group Corporation, Nanjing, China

[1]yujun@sgepri.sgcc.com.cn, [2] xigao_li@foxmail.com

*Corresponding author

Keywords: Active-active disaster recovery, Smart grid, Big data, Distributed file system.

Abstract. Data storage security issues are expanding to global area, while "Internet+" Smart Energy is prospering in high speed. Therefore, the security backup for the distributed file system (DFS) has become the key factor in the construction of smart grid in electric power. However, due to remote data synchronization across multiple data centers, the reliability and robustness of the storage system is relatively weak. Our work proposed a novel active-active disaster recovery model, with a Virtual Redirection Network (VRN). Under the automatic control of VRN, the model provides a solution of data synchronization across large scope of power grid center. The results show that the system not only improves the IO performance in massive small files, but also increases the total system robustness over a long distance data transfer.

Introduction

With the evolution and expansion of distributed computing theory and the applications of distributed systems, it is now becoming more and more widely recognized for distributed architecture as the infrastructure of most enterprises. Therefore, a distributed file system (DFS) with a stronger robustness is widely utilized for storage [1, 2, 7]. Currently DFS almost took over all the big data storage by networking various data center, and provide high-speed access to those data.

Disaster tolerance of distributed systems has long been valued by high-end enterprises, especially in security critical areas such as power grid, therefore, carefully designed disaster recovery architecture in the DFS will aid engineers by quickly reboot or switch from a remote backup when a data center is devastated, and enhances overall system reliability and service quality. However, our present technology is still limited in the reliability requirement of the distributed file system for the State Grid Corporation of China[3]. Currently, most of the mature Disaster Tolerance systems are from U.S. or Germany, which are usually not open-source software. The lack of disaster tolerance program for DFS on the high-end storage, with none dedicated disaster tolerance design in open-source DFS, leads to great obstacles for its application in national smart grid[5,6].

The model proposed in this work will enhance the applications of new technologies on smart grid information systems, such as distributed storage, long-range recovery, etc., and ensure a high availability of business data. Based on the presented open-source DFS, our work will devote to resolve the prevalent technical defects of open-source

DFS, including the single-point failure on metadata node, poor security of data storage, lack of special disaster tolerance design, etc. More future applications will firmly assure the availability and robustness of smart grid system data as a disaster occurs.

Related Work

A. Oracle DataGuard Solutions

The Oracle DataGuard solution from Oracle is one of our model inspiration sources. The DataGuard places the backup data in both local and remote data center, where named as primary and standby database[10]. If user executes a SQL command/transaction, the transaction will be firstly sent to primary DB, and then it will be sent to standby DB, and do the same action. Therefore, the standby DB will be a synchronous mirror of primary.. However, this method relies on the redo log of distributed database, therefore it cannot be directly applied to the distributed file system, which has none structural data and logs.

B. Direct Data Fetch Technology

Our former work of a disaster-tolerance model for a specially distributed file system, MooseFS, using the Direct Data Fetch (DDF) technology, which solved the transfer suspension problem during the disaster happens[4, 8]. The DDF technology will seek partially file index from data node, and directly send the data to remote data center. However, this model did not fix the problem of quick switch. Also, the old model cannot handle massive amount of data flow, and though the files are intact, the pattern of file directory slightly changed during the switch, if applied with DDF technology.

C. Disaster-tolerance in Lightweight DFS

Other disaster tolerant technology applied in various lightweight distributed file systems, such as Lustre, iRods, GlusterFS, etc., are also having their own advantages like scalable or non-deploy. However, most of the lightweight DFSs strongly rely on the RAID or external software supports[4,9]. Due to the limit in either performance or disaster recover range, these methods are not applicable in the larger scale of smart grid power data, and we are unable to make the comparison with our proposed model.

Proposed Model

By applying the distributed file systems into business data storage, the capacity and access concurrency is greatly improved to some extent. However, the traditional method of disaster-tolerance would meet a great challenge when the system expands to a large scale. Ttraditional disaster recovery issue of the power system is facing more difficulties in its transplanting in a decentralized DFS. If the data center was destroyed accidentally due to accidents (such as fire, lightning, earthquakes, etc.), the contents stored in the data center will be lost.. For DFS, there is not yet a strong enough, open-source disaster recovery technology available.

Based on our former work, the model uses data-copy as its backup principles. For the disaster recovery part of the production center, firstly a "request operation record" message from the collaboration server is handled by log resolver, and then a bulk of data from the master file pool is sent to collaboration server;. During the data sync, the synchronization tool acquire information from the log resolver, and collect data by multi-threading, then perform the appropriate file operations to the remote center, and completing the synchronous backup function of files.

Active-active Hot Standby Model Using Virtual Redirection Network

The model presented in our works mainly includes two parts, the VRN network and a powerful sync tool. The architecture of the model is similar to most distribution systems, which includes a master node Charlie (MD), a spare node Delta (SN) and a bunch of data nodes (DN). The mount point of the client side take I/O operations from the user side, while the user's I/O requests are finally processed by MD, as shown in Fig. 1.

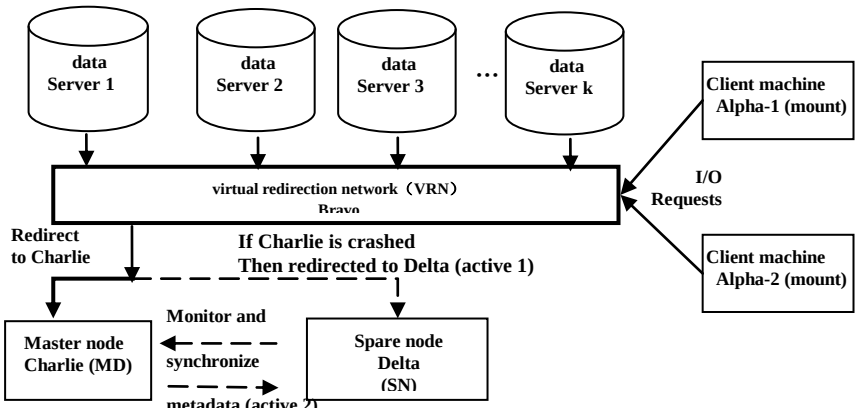

Figure 1. Active-active Hot standby switch model

Based on a native infrastructure, we proposed an active-active hot-standby model, which can quickly trigger to start the hot standby server when MD crashed. To achieve this goal, a virtual redirection network (VRN) will be set between MD and other parts, providing a floating access point for the data server. The client and the data server only knows the VRN address of MD, and also only establish communication with it, as if they are communicating with an actual master node.

At the same time, the VRN is partially managed by ucarp, an open-source software which redirects all the access to VRN to MD, as shown in the Fig.1. The active-active script on ucarp will run on VRN, to determine the heartbeat signal of both MD and SN. Once the master node Charlie goes down, the monitor thread will be notified immediately, and performs a two-step switch.

1. a quick switch is performed in a few milliseconds to redirect all messages to server Delta (Active-1).
2. Then, the script will perform the operations of meta data acquisition (Active-2), which happens a few seconds after the active-1.

The focus of this program is that the metadata can be immediately used for the following transaction recovery, as the metadata for restoring is stored right on server Delta. After that, ucarp will perform the handover of VRN to let it directing to the backup server Delta. Then, the backup server Delta will start MD program to take over the position of MD. In the meantime, the previous master node Charlie will automatically switch itself to the backup mode to wait for network to restore (either automatically or manually by engineers).

Apply the Model into Disaster Recovery Tool

Under the architecture above, the remote disaster recovery model based on data copy, which can provide additional redundancy for the data of the entire system, is established by constructing a collaboration server and establishing a long range connection to disaster tolerance center. Generally, the disaster tolerance center is deployed in another

city or region, in order to prevent the threat to the overall data security from local data disaster. The collaboration server provides data exchange between the local system and the remote center, and handles the data tranfer function. Thus, the model will finish the data disaster recovery of distributed file systems.

As shown in Fig.2, local production center deploys a modified DFS with the active-active model, which will establish a connection with VRN, and transfer all its operation logs to the collaboration server. At the same time, the DFS can provide other system and users their original features. Since then, the collaboration server receives logs and parses them to generate a plan queue, and ready to get the file information from local production center. At last, the sync tool in the collaboration server retrieves the actual data, and send it to remote center through a trusted long range network.

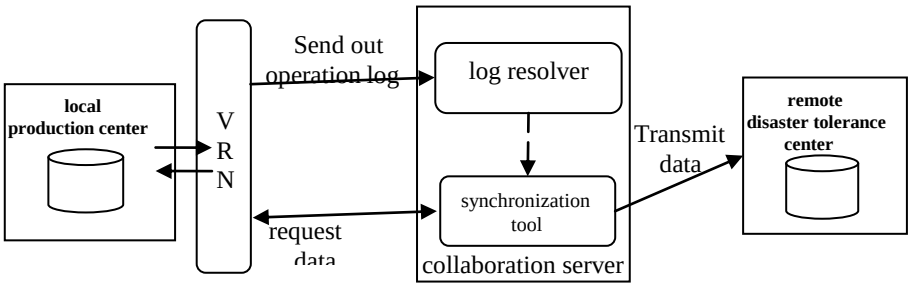

Figure 2. DFS disaster recovery model

Experiments and Results

Test Overview

A series of tests are designed to assess the model performance and availability. Two main scenarios in the tests are respectively based on the file size and the integrity of files. The test environment is based on server cluster with Red Hat Linux Enterprise 7.1 (x64) operating system. In this experiements, we mainly use the hash verification and comparison with other models to evaluate and verify the robustness and performance of the model.

Table 1. Environment overview of test cluster

ID	Server function	Memory	Hard Drive	IP address	Misc
1	Master Node	64G	900GB	192.168.2.36	Up VRN
2	Spare Node	64G	900GB	192.168.2.37	Up VRN
3	Data node 1	4G	800GB	192.168.2.63	
4	Data node 2	4G	800GB	192.168.2.64	
5	Data node 3	4G	800GB	192.168.2.65	
6	Client Machine	8G	500GB	192.168.2.66	

VRN File Integrity Test

The quantitative test for virtual redirection network is necessary before the performance test for the ovearll model. This test will be an integrity test for the model to determine whether data loss will occur in its normal data replication process. During the test, about 25,000 randomly sized files will be produced and put in a folder with random depth (levels) ,and the maximum depth is 5, according to the usual folder levels. The

depth of the folder is evenly distributed by following U (0,5), while the file size distribution is following the normal distribution, which μ = 500 (KB).

As shown in Fig.3, A specific shaped directory/file tree will be generated by the script on client machine. Then according to the model, those files and directories will be generated on both local and distant data center. Then, we will collect the tree information on both sides, and make a comparison of those trees to check the file and directory numbers.

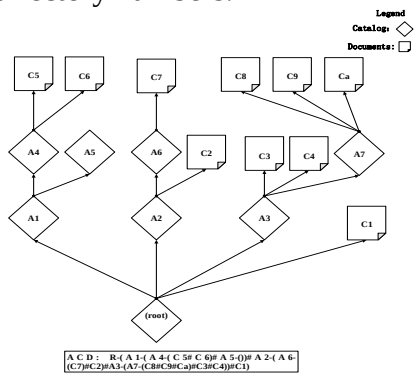

Figure 3.The generation of the Unique Tree String

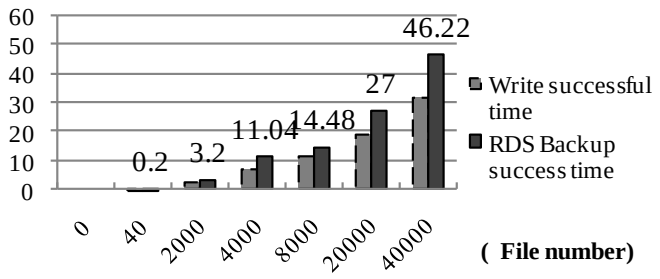

Figure 4.Performance Test of Active-active Model

During the comparison, we use a faster algorithm to make the comparison, the unique tree string (UTS). The UTS generation rules are defined as follows:

1. Root directory as the start of UTS uses R- as the leading character.
2. Directories, except the root directory, use "(directory name)-(filename)" as the leading character. All files use their filename as identification.
3. Marker "#" is used as the separator between peer files, and no delimiter between files in different directories.

An example of UTS is shown in Fig.3. The UTS contains the essential information of a folder tree. Since a UTS is equivalent to a folder tree, the comparison between UTSs is equivalent to the comparison between folder trees. The experiment results are shown in Table 2.

Table 2. Model integrity test results

Value	Expected Result	Actual Result	Conclusion
File number	25679	25679	No file losses
Tree inspection	Same UTS	Observed equal UTS in both sites	Folder structure remains intact
File size matching	RDC matching with the main server	Result matches in both server	File size is the same
File Hash MD5 match	File Hash Match	File Hash is Match	File content intact

The test results show that the VRN implements a precise match on file number, tree pattern, and file size/hash. Therefore, the VRN is stable in redirecting large amount flow of data, and kept their logical structure.

File-scale Test Based on the Number of Files

Same as the previous section, the script is used to generate a fixed number of files, and the file size is following the unified distribution of U (1, 51200) (in Bytes). The experiment collects three types of timestamp data, including script start timestamp, the master node writing successfully timestamp and the successful backup timestamp.

Fig. 4 shows tests based on the number of files in the standardized tests of data reading and writing speed (based on time-consuming). As the light gray bar indicates

the write successful timestamp in main server, the dark gray bar indicates the remote disaster site backup successful timestamp. Therefore, the model performance should be accessed by the distance between two bars. According to Fig. 4, the model performs a rapid data replica creation during high volume of data flow. Over 40,000 files, the time consumption is sharply increased. However, this increment may due to the limit on the physical 1000Mbps network, where 40,000 files transferred in 46 seconds is reaching its top.

Conclusion

Based on the native DFS for smart grid of State Grid Corporation of China, we proposed a novel active-active model with VRN network, which produces a quicker switch between master and spare (backup) nodes, as well as a promoted performance on data transfer. The experiment indicates that the VRN correctly redirected all access data to the right server, whether or not there is a switch, by which ensures the server data security in smart grid. Through the active-active model we proposed, large enterprises will get rid of the data disaster problem over long range, and more applications of this model in other fields are prospering in the near future.

Acknowledgements

This research is supported and funded by 2015 State Grid Corporation of China Technology Project.

References

[1] JN Foster, MB Greenwald, C Kirkegaard, BC Pierce, A Schmitt .Exploiting Schemas in Data Synchronization,pp. 669-689.

[2] Daniel P.Bovet and Marco Cesati. Application of Data Exchange Platform in Information Construction of State Grid Corporation of China. Electric Power Information Technology,2011.

[3] Lawler C M, Harper M A, Thornton M A. Components of Disaster Tolerant Computing[J]. IT Application Downtime & Executive Visibility," International Journal of Business Information Systems, 2008, 88(12):581 - 582.

[4] Xigao Li, Lin Qian. A hybrid disaster-tolerant model with DDF technology for MooseFS open-source distributed file system[J]. Journal of Supercomputing, 2016:1-17.

[5] Lihao Xu,Jehoshua Bruck.X-code: MDS array codes with optimal encoding. IEEE Transactions on Information Theory. 1999.

[6] B Saltzberg. Performance of an Efficient Parallel Data Transmission System, Communication Technology IEEE Transactions on, pp.805 – 811.1967.

[7] RT Beeston, EM Dawson, GS Johnson,K Katori, JM Swingler .Synchronous mode replication to multiple clusters.2015.

[8] Xigao Li, Lin Qian. A Direct Data Fetch technology applied in disaster-tolerant model of Distributed File System[C]// International Conference on Computer Science and Network Technology. IEEE, 2015..

[9] FW Chang, M Ji, STA Leung,J Maccormick, SE Perl .Cost-effective Disaster Tolerance. Proceedings of Fast, pp.103—116.

[10] Liu, Xiu-Ju, A brief analysis of the disaster recovery backup technology in oracle database DataGuard, 2010 2nd International Conference on Industrial and Information Systems, IIS 2010, v 2, p 234-236.