THE DESIGN OF DATA DISASTER RECOVERY OF NATIONAL FUNDAMENTAL GEOGRAPHIC INFORMATION SYSTEM

ZHAI Yong[a], CHEN Jie[a], LIU Lei[a], LIU Jin[a]

[a] National Geomatics Center of China,Beijing,100830,China-(zhaiyong, chenjie, liulei, liujin)@nsdi.gov.cn

Commission

KEY WORDS: Geographic Information; Disaster Recovery; Storage; Remote Replication; RTO; RPO

ABSTRACT:

With the development of information technology, data security of information system is facing more and more challenges. The geographic information of surveying and mapping is fundamental and strategic resource, which is applied in all areas of national economic, defence and social development. It is especially vital to national and social interests when such classified geographic information is directly concerning Chinese sovereignty. Several urgent problems that needs to be resolved for surveying and mapping are how to do well in mass data storage and backup, establishing and improving the disaster backup system especially after sudden natural calamity accident, and ensuring all sectors rapidly restored on information system will operate correctly. For overcoming various disaster risks, protect the security of data and reduce the impact of the disaster, it's no doubt the effective way is to analysis and research on the features of storage and management and security requirements, as well as to ensure that the design of data disaster recovery system suitable for the surveying and mapping.

This article analyses the features of fundamental geographic information data and the requirements of storage management, three site disaster recovery system of DBMS plan based on the popular network, storage and backup, data replication and remote switch of application technologies. In LAN that synchronous replication between database management servers and the local storage of backup management systems, simultaneously, remote asynchronous data replication between local storage backup management systems and remote database management servers. The core of the system is resolving local disaster in the remote site, ensuring data security and business continuity of local site.

This article focuses on the following points: background, the necessity of disaster recovery system, the analysis of the data achievements and data disaster recovery plan. Features of this program is to use a hardware-based data hot backup, and remote online disaster recovery support for Oracle database system. The achievement of this paper is in summarizing and analysing the common characteristics of disaster of surveying and mapping business system requirements, while based on the actual situation of the industry, designed the basic GIS disaster recovery solutions, and we also give the conclusions about key technologies of RTO and RPO.

1. INTRODUCTION

With rapid developments of information technology, aerospace technology and network technology, China's surveying and mapping is seeing a unprecedented evolution. With satellite positioning system, RS, GIS and computer technology as the main all-digital mapping technology system， it completely replace the traditional simulation system of surveying and mapping technology. The rapid development from informatization surveying and mapping technology system (Liu, 2012), at the same time, the data security of the geographic information system is facing more and more challenges.

The geographic information of surveying and mapping is a fundamental and strategic resource, which is applied in all areas of national economic, defence and social development. Especially, the classified geographic information directly concerns Chinese sovereignty, and its national and social interests. It is especially vital to national and social interests when such classified geographic information is directly concerning Chinese Sovereignty. Once data cannot recovered, it will cause great harm (Zhai, Liu, 2008).

At present, the GIS storage of the classified geographic information included in the database is of 1:4,000,000 scale, 1:1,000,000 scale, 1:250,000 scale, 1:50,000 scale, mapping of western china, island surveying, geodetic datum and so on. During more than 30 years, it is accumulated gradually to form the geographic information that is the basis of the industry of surveying and mapping geographic information.

Several urgent problems needs to be resolved in this field which is mass data storage and backup, establishing and improving disaster backup system especially after sudden natural calamity accident, and ensuring all sectors rapidly restored on information system will operate correctly. For overcoming various disaster risks (Xu, 2013), protect the security of data, and reduce the impact of the disaster, it's no doubt the effective way is needed to design and establish a data disaster recovery system.

2. NECESSITY

With the advent of the era of big data, data is increasingly important. According to IDC predicts that by 2020, the global amount of data will reach 35ZB (Gong, 2012). Large amounts of structured data, unstructured data, semi-structured data will come to pervade every aspect of our lives. Subsequently, the issue of data protection is becoming increasingly prominent, a variety of catastrophic events given industry a wake-up call.

The American "9.11" incident caused the collapse of the trade building, after the incident, the Deutsche Bank and Bank of New York had quite distinct outcome. Deutsche Bank business recovered rapidly because of establishing the center of remote disaster recovery. On the contrary, after a few months Bank of New York forced liquidation due to the loss of data. From "9.11" incident, Kobe earthquake in Japan, Southeast Asia Tsunami and 5.12 Wenchuan earthquake, previous disasters that constantly waken people the importance of disaster backup system (Wei ,2013).

As early as 2003, state authorities on issued a "National Informatization Leading Group on strengthening information security work advice", explicitly requested not only to fully consider disaster recovery capabilities of important information system, but also develop and constantly improve the information security emergency response plans (Wei, 2013).

At present, national fundamental GIS has not yet built an effective disaster recovery mechanism, once the system suffer the unpredictable incidents it will result in software and hardware crash of database management system, and lead to serious consequences, causing irreparable damage to the country. In order to protect the data security of fundamental geographic information, establishing disaster recovery system is necessary, and ensure the smoothness of dissemination service and emergency works.

3. ACHIEVEMENTS ANALYSIS

Currently, a variety of achievements total 562TB, including national geodetic surveying achievements, national fundamental geographic information data achievements, other thematic data and high-current potential and high access frequency RS achievements. Among them, the RS data reach 549TB. Frequent use of such data than other areas, covering an area of 3.5 million square kilometres, current potential from 2007 to 2013, covers the majority of China's economically developed areas, populated areas, disaster-prone areas and other key areas. In addition, according to the geographical national conditions surveillance scheduled, the next five years will increase by an annual average of the various types of achievements about 300TB, such as the geographical national conditions achievements and the basic scale database. About 120TB achievements that is focal monitoring area of high-resolution image, grid and vector are using online storage. The rest 180TB of data with near-line storage. Therefore, the new online data achievements in the next five years a total of 600TB. Based on mapping production, support for emergency mapping and requirements for distribution services, total over 1162TB (562TB +600 TB) data outcomes using online storage.

According to the actual demand, need online disaster recovery system to protect the core foundation of GIS data achievements, include achievements of 1:250,000 scale, 1:50,000 scale, island surveying, geographical national conditions and geodetic datum and so on. As of early 2013 data about 20TB, expected to increase to 40TB after five years. The above core data achievements mainly using Oracle relational database for storage management.

Other data are mainly renewable strong RS achievements, using offline remote storage and recovery.

4. DATA DISASTER RECOVERY SYSTEM

4.1 Design Principle

Design of Disaster Recovery System stick to the economic and advanced principles:
(1)Oriented mature network technology, correctly handle the contradictions between WAN private network bandwidth narrow and spatial data storage capacity.
(2)Oriented advanced storage technology, designed economic data remote transfer mode, to meet a wide range of spatial data and high requests of I / O.
(3)Oriented system integration technology, designed scientific technology integrate, configuration reasonable related hardware and software products.

4.2 System Architecture

Database construction mode of national fundamental geographic information database using foliation dataset, the spatial data warehouse need to establish a spatial index and the calculated property relations. For example in database storage of NGCC, according to the core data into foliation dataset way daily average data about 4GB, herein as a basis for the analysis.

In order to ensure fundamental geographic information data integrity and service continuity, established a hot standby site in remote which can backup and recovery data via IP network. That is, through the IP network to a combination of synchronous and asynchronous way to back up local site data to a remote site, when a disaster occurs in the local site, you can choose to restore data from a remote site or service from the remote site to replace the local site. System architecture shown in Fig.1.

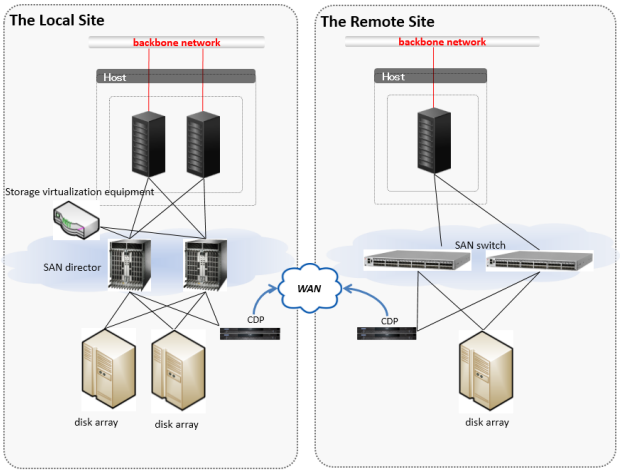

Figure 1. Data disaster recovery system structure

4.3 Design Index

Data disaster recovery system is mainly preventive irresistible or unpredictable catastrophic events brought about non-availability of data and applications, and loss caused minimized. Two technical indicators are RTO and RPO used to measure the system.

RPO (Recovery Point Objective): the data recovery-point objective, mainly refers to the business system that can tolerate data loss. RTO (Recovery Time Objective): the recovery time objective, mainly refers to the service that can be tolerated the

maximum time out of service, or recover from disaster to service the minimum time period required by the service.

RPO and RTO is computed as follows:
(1) RPO is approximately 100MB-4GB
The system using two steps to recover data are synchronous within local sites and asynchronous between the local and remote sites. The achievements need to replicate in accordance with offsite replication strategy after every 100MB achievements storage, combined with the secrecy WAN bandwidth to 20Mbps (actual transfer rate is about 2MBps), then:
1) single remote data replication time = 50 sec
2) single-day remote data replication time 4GB/2MBps=2000 sec

The remote data transmission based on 20Mbps secrecy WAN bandwidth, single day can be transmitted data 168.75 GB (2MBps×3600×24), more than 4GB data generated by the construction of single days. Therefore, the Wan is able to guarantee completely transfer a single-day storage data at the same day, storage data transmitted does not form the delays and accumulation, loss data due to disasters not more than one day of storage data.

Based on the above analysis and calculation, it can calculate the RPO between 100MB-4GB.

(2)RTO is approximately 15-30min
The system adopt CDP (continuous data protection) technology, increased data in remote maintenance. Meanwhile, you can separate the data copy relationship between the local site and the remote site when the disaster happened in local site. Spend about 15-30 min in database recovery from the remote site which achieve take over the service quickly. After the local site hardware and software system to return to normal, recovering remote local site data from remote site.

4.4 System Composition

National fundamental geographic information data disaster recovery system is mainly composed of the WAN, local disaster subsystem, remote disaster subsystem and data remote replication strategy.

4.4.1 WAN: Design of renting a 20Mbps bandwidth Special communication cable between the local disaster recovery subsystem and the remote disaster recovery system, using SDH (Synchronous Digital Hierarchy) optical fibre line, supports the TCP/IP protocol, the transmission efficiency is about 90%, the rate can reach 1.8Mbps (6.33GB/ hours).

4.4.2 Subsystem Equipment Allocation: Local disaster recovery subsystem is responsible for the daily management of the data achievements of Disaster Recovery Strategy. Remote disaster recovery subsystem is responsible for receives achievements, the outcome of the data achievements recovery in case of a disaster, as well as alternative local service in case of emergency.

Local disaster recovery subsystem is mainly composed of hardware and software. The hardware part mainly consists of 1 disk array, 1 storage virtualization equipment, 1 CDP equipment, 2 SAN (storage area network) switches and 1 special remote backup server. The software part mainly consists of a database management software and GIS software, spatial data engine management software each of the 1 sleeve.

Remote disaster recovery subsystem is mainly composed of hardware and software. The hardware part mainly consists of 1 data management server, 1 disk array, 1 CDP equipment, 2 SAN switches and 1 tape library. The software part mainly consists of a database management software and GIS software, spatial data engine management software each of the 1 sleeve.

Storage virtualization equipment through SAN for virtual and shield the host storage layer. Realizing the data real-time replication between production data of local site and data of disaster recovery database system through traffic mirroring. When disaster database systems to accept the new write I / O in the local site, using CDP equipment through its data split point features, the data is updated through CRR（continuous remote replication ） features asynchronous replicate to disaster recovery database system in the remote site.

4.4.3 Disaster Recovery Strategy: With combination of hardware and software, the data storage in SAN that will be replicated to remote site through IP network, once the event of a disaster, it able to recovery data from local or remote site based on the nature of catastrophic, or directly to enable remote site for data services.

Data disaster recovery systems-oriented database management system (DBMS) in the three disaster recovery framework, namely the local site production database (first point) and local disaster database (second point) for data synchronous replication, then, remote asynchronous data replication between local disaster recovery system and remote backup database.

(1) Within Local Site Data Replication
The existing data achievements in the local production database which are using storage virtual equipment to complete the achievements migrate storage equipment to the existing disaster recovery storage equipment.

The new generate data achievements in the local production database, the first to establish the data replication relationship within the local site data, that is, through storage virtualization equipment within a SAN environment establish traffic mirroring relationship between the production disk array volumes and local disaster recovery subsystem disk array volumes, to synchronize data update replication. This data replication relationship built on SAN (16Gbps) network, enabling real-time replication of data increments.

(2)Data Remote Replication
After completed replicating the local site data achievements, the local and remote disaster recovery subsystem through the data split and CRR provided by continuous data protection equipment, establish volume group replication relation, asynchronous remote data replication.

Based on CDP equipment of CRR can be achieved for the RPO provides point in time recovery, while achieving complete site disaster protection, reached in RTO and zero data loss, and significantly reduce the bandwidth requirements via bandwidth reduction techniques.

When the local site breakdown, you can choose to select a continuous time point from the local disaster recovery

subsystem or select the important point in time recovery of data and applications from the remote disaster recovery subsystem, you can also enable remote site database system instead of the local site, in extreme cases associated systems for data services.

(3)Disaster Recovery System Topology

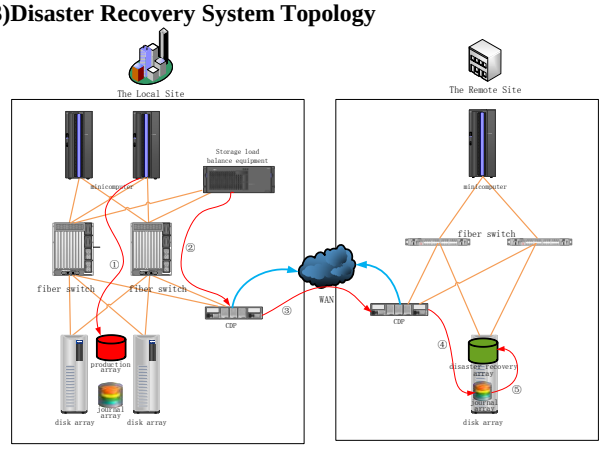

Figure 2. Data disaster recovery system topology

4.5 Key Technology

National fundamental geographic information data disaster recovery system is built on the basis of data on remote storage backup application-level disaster recovery, using a hot standby disaster recovery solutions which combine with the synchronous hardware and asynchronous replication. The key technologies include storage virtualization, continuous data protection and CRR.

Storage virtualization technology to achieve the internal local site across different types, different manufacturers uninterrupted storage, data movement transparent to the user, to simplify data movement processes, improve the efficiency of data movement.

Based on the IP network restore to any point in time, combine with the CDP and CRR technology achieved application-level disaster recovery between local and remote sites. All data changes are recorded in the log and take time tag, so data recovery can be rolled back to a consistent point in time. Rollback in the data while the application also consistent recovery, to ensure that the application-level disaster recovery implementations. Also implements two-way replication between the local and remote site, and allows each site as the recovery site.

5. CONCLUSION

This article analyses the features of fundamental geographic information data and the requirements of storage management, designs a disaster recovery system of DBMS plan. It employed the disaster recovery industry mainstream technologies, with good compatibility, scalability, robustness and economy. The system ensure the data security and the smoothness of business.

References

LIU Feng Jun, 2012. Thinking digital archives building. China Archives, Vol-12, PP43

ZHAI Yong & LIU Lei, 2008. The design of remote storage & backup system in national geomatic center of China. Science of Surveying and Mapping, Vol.33 Suppl, PP.98-100

XU Cheng, 2013. An analysis and improvement for disaster recovery situation of Shanghai airline's information system, Shanghai Jiao Tong University

GONG Lve, 2012. Our domestic enterprises reaction mass storage areas. Digital Communication World, Vol-7, PP28-29

WEI Jing, 2013. Billion market pull "Disaster recovery storage concepts". China Securities Journal