

基于NAS架构的数据容灾备份系统的设计与实现

叶虹余

(中海石油(中国)有限公司深圳分公司, 广东 深圳 518054)

摘要: 石油行业地质资料数据具有数据量大、增长快和数据种类复杂等特点, 如何备份这些海量的数据成了一大难题。以中国南海东部海域石油勘探数据为例, 结合石油行业数据特点, 设计了“两地三中心”的数据备份系统架构, 建成了三级数据备份系统。该系统以NAS存储数据快照技术为基础, 各级备份系统采用不同的备份策略, 对应不同的风险级别, 通过对各级备份数据进行恢复验证, 各级备份数据都是完整有效的, 取得了预期的效果, 达到了系统建设的目的。

关键词: NAS; 数据备份系统; 异地容灾; 数据恢复

中图分类号: TP309.3 **文献标识码:** A

Design and Implementation of the Data Disaster Recovery System Based on NAS Architecture

YE Hongyu

(Shenzhen Branch of CNOOC Ltd., Shenzhen 518054, China)

Abstract: The geological data of petroleum industry has the characteristics of large data volume, fast growth and complex data types. How to backup these huge data is a difficult problem. Taking the data of oil exploration in the eastern South China Sea as an example, combining with the characteristics of the oil industry data, the paper designs the *Three Centers for Two Places* data backup system architecture, and constructs the three-level backup system. The system is based on NAS storage data snapshot technology, and different backup strategies at all levels are adopted, corresponding to different levels of risk. Through the study of recovery verification, all levels of the backup data are complete and effective, and the expected results and the purpose of system construction are achieved.

Keywords: NAS; data backup system; remote disaster recovery; data recovery

1 引言 (Introduction)

随着中国南海东部海域石油勘探逐渐进入深水领域, 勘探所涉及的海域面积越来越大, 由此产生的石油勘探开发数据量也越来越大, 这些数据是进行石油勘探开发的基础, 也是企业最重要的财富。因此, 为了保障勘探开发科研数据的安全, 降低因人为误操作导致数据丢失的风险, 防范因各种自然灾害而造成硬件设备损毁导致的数据丢失, 建设一套有效的数据容灾备份系统显得尤为重要。

传统的数据备份多以磁带库为介质, 这种备份模式具有介质不易管理、存储容量有限和数据查找恢复不便等缺点, 已不能满足新形势下的企业级数据备份需求。文中设计的数据容灾备份系统以NAS存储为基础, 建立了多级备份系统, 可满足从过去几小时到几年内的数据恢复需要, 并在异地建立了数据灾备中心, 为企业的重要数据建立了多重安全保障^[1]。

2 现状分析(Analysis of the situation)

南海东部海域经历30余年的勘探开发, 已实现连续20年

油气产量超千万方, 经过多年的积累, 形成了海量的勘探开发数据资料库, 目前共有数据大约170TB, 数据类型及数据量如表1所示。

表1 南海东部海域勘探开发数据类型及大小

Tab.1 Types and sizes of exploration and development data in the South China Sea

应用系统	数据类型	存储形式	数据大小
GeoFrame	地震资料解释数据	结构化数据	120TB
Landmark	地震资料解释数据	结构化数据	10TB
Eclipse	油藏开发数据	非结构化数据	38TB
Oracle	数据库文件	非结构化数据	200GB
信息系统	导航数据	非结构化数据	200GB
信息系统	应用软件	非结构化数据	2TB

从上表来看, 勘探开发数据主要是GeoFrame应用数据, 该类数据是GeoFrame软件应用平台下的综合地质资料数据, 采取结构化的存储形式, 结合Oracle数据库, 利用GeoFrame

软件平台进行地质资料综合解释，其次是Eclipse油藏数据，该类数据用来进行油田数值模拟，采取非结构化的存储形式。

由于数据量大，数据种类多，数据增长快，数据读取频率高，目前所有数据的存储使用模式已由过去的单一机器存储，变成了分布式的网络存储，其存储形式为NAS存储(网络附属存储)，利用磁盘阵列和专业的软件来管理数据^[2,3]。以此为基础再结合应用服务器、数据库服务器、应用客户端和千兆光纤网络，形成了分布式的应用系统架构^[4]，其系统架构图如图1所示。由于所有数据都是存储于NAS设备上，应用人员只需要一台瘦客户机(PC)即可使用各种专业软件，调用所有数据。

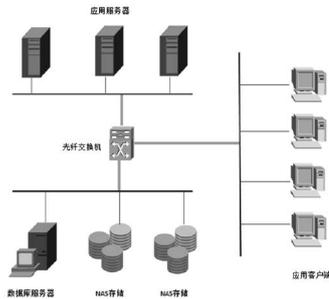


图1 数据存储应用系统架构图

Fig.1 Architecture diagram of data storage application system

3 系统建设(System construction)

3.1 设计要求

根据目前的系统架构和数据容量，以及每天的数据增量，综合考虑建设成本和网络带宽限制，结合系统故障风险承受能力和需求分析，提出了“两地三中心”的系统架构设计方案，即在深圳建设两套数据备份系统：近线备份系统和同城备份系统，在北京建设异地数据容灾备份中心，其数据流图如图2所示。

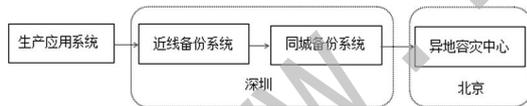


图2 备份系统数据流图

Fig.2 Data flow diagram of backup system

按照“两地三中心”的系统架构设计方案，需要在深圳和北京建设三套备份系统，其中近线备份系统是对生产应用系统的在线备份系统，部署在与生产应用系统的同一地点，同城备份系统为近线备份系统的半同步备份系统，部署在与近线备份系统同一城市的不同地点，异地容灾中心为同城备份系统的异步备份。三级备份系统分别对应不同的风险等级^[5]，其中近线备份系统对应一级风险，主要包括人为误操作和软硬件及系统故障导致的短期内数据丢失，需要利用近线备份系统恢复数据；同城备份系统对应二级风险，主要是水灾、火灾等灾难导致主机房硬件设备损毁导致的数据丢失，需要利用同城备份系统恢复数据，以及需要找回过去更长周期内的数据；异地容灾中心对应三级风险，主要是地震、恐怖袭击等不可抗力因素导致所在城市毁灭，继而导致当地所有硬件设备损毁和数据丢失，需要利用异地容灾中心恢复数据。

三级备份系统分别采取不同的备份策略进行数据备份：近线备份系统的备份策略是数据保留周期短密度高，同城备份系统的备份策略是数据保留周期长密度适中，异地容灾中心的备份策略是数据保留周期短密度低^[6]。

3.2 技术方案

根据设计要求，备份数据流传输路径依次为生产应用系统至近线备份系统，近线备份系统至同城备份系统，同城备份系统至异地容灾中心。目前生产应用系统数据采用NetApp存储设备，通过进行调研和需求分析，并结合前期的各项测试结果，最终决定近线备份系统也采用NetApp存储设备，同城备份系统和异地容灾中心采用DELL Compellent存储设备，生产应用系统至近线备份系统之间数据传输利用NetApp Data ONTAP的SnapMirror模块实现，近线备份系统至同城备份系统的数据传输利用Nexenta NexentaStor的Auto-Tier模块实现，同城备份系统至异地容灾中心的数据传输利用DELL ENTERPRISE MANAGER的Replication实现，其架构图如图3所示。



图3 备份系统架构图

Fig.3 Architecture diagram of backup system

3.3 方案实施

近线备份系统采用NetApp存储设备，主要硬件配置为NetApp FAS8020 控制器，以及配套的NetApp DS4243磁盘柜，磁盘柜配置4T/7200RPM的sata硬盘，软件许可配置为SnapMirror、SMO、OSSVhe和FlexClone等模块。

由于近线系统采用NetApp存储设备，与生产系统存储设备属于同构设备，因此数据传输可利用SnapMirror实现。SnapMirror是NetApp ONTAP的一项集成功能，可在各存储端之间提供一种高效的数据复制解决方案，通过配置SnapMirror参数，可满足从几分钟到几小时不等的的数据恢复点目标要求，从而满足了近线备份系统周期短密度高的数据备份需求。

同城备份系统采用DELL存储设备，主要硬件配置为Dell Compellent SC8000控制器，Dell Compellent SC200磁盘柜和3T/7200RPM的sata硬盘，由于近线备份设备与同城备份属于异构设备，因此还配置一台Dell Poweredge R720控制器，用于安装Nexenta公司的NexentaStor软件产品，该软件能够解决不同厂商的设备兼容和数据传输问题，其中的Auto-Tier模块实现了数据从近线备份系统至同城备份系统的传输，并能对数据进行重复删除和压缩，Auto-Snap模块能对文件系统做快照备份，从而满足了同城备份系统数据保存周期长的需求。

异地容灾中心也采用DELL存储设备，硬件配置与同城备份系统一样，作为同城备份系统的异地备份，通过iSCSI协议

利用DELL ENTERPRISE MANAGER的Replication模块从广域网进行数据传输^[7]。

4 关键技术(The key technology)

4.1 磁盘阵列

RAID是英文Redundant Array of Independent Disks的缩写,中文简称为独立冗余磁盘阵列,简单来说,RAID就是一种把多块独立的硬盘按不同的方式组合起来形成一个大的存储资源池,从而提供比单个硬盘容量更大、可靠性更高的数据存储技术。硬盘的不同组合方式,对应RAID不同的级别,提供了不同的数据读写速度和可靠性。

在此次数据备份系统建设中,各级备份系统都采用了RAID技术,通过RAID技术,有效保障了各级备份系统的数据安全。近线备份系统采用RAID 5,即将所有硬盘形成一个大的存储池,将所有数据和奇偶校验信息条块化的存储在所有硬盘上,并通过设置Hot-Spare(热备份)盘来提高数据的安全性,即使RAID有硬盘出现故障,也能保证存储数据的安全。同城备份系统和异地容灾中心由于采用了ZFS文件系统,因此使用了RAIDZ解决方案,RAIDZ可以利用ZFS文件数据的校验和其他机制进行错误检测和更正,可以对所有数据进行校验,并可在有正确副本时对错误数据进行恢复。

4.2 存储快照技术

存储网络行业协会SNIA(StorageNetworking Industry Association)对快照(Snapshot)的定义是:关于指定数据集的一个完全可用拷贝,该拷贝包括相应数据在某个时间点(拷贝开始的时间点)的映像。快照可以是其所表示的数据的一个副本,也可以是数据的一个复制品。快照的作用主要是能够进行在线数据备份与恢复,当存储设备发生应用故障或者文件损坏时可以进行快速的数据恢复,将数据恢复至某个可用的时间点的状态。

同样,在此次数据备份系统中都采用了存储快照技术,通过对备份系统中的数据卷按时间点做快照,即可形成该时间点的数据保护点(CDP),若该时间点以后出现数据丢失,则可利用快照恢复至该时间点,从而找回丢失的数据。快照的频率和保存时间长短决定了对数据的保护程度,在此次备份系统建设中,根据需要制定了不同的快照策略,即对近线备份系统采取密度高保存周期短的策略,满足近线备份系统短期的数据保护需求,而对同城备份系统和异地容灾中心,则采取密度低保存周期长的策略,满足数据保存时间长的要求。

4.3 数据卷克隆

数据卷克隆主要是针对数据恢复的,一般来说,数据快照是只读不可写的,因此要真正使用快照恢复的数据,需要进行数据卷克隆,即利用快照克隆出一份与相应数据集一样的可读可写的数据卷,从而达到数据恢复并可用的目的。

5 数据恢复与验证(Data recovery and validation)

数据容灾备份系统建成后,还需要配备相应的备用服务器,搭建应用软件环境,以便对备份数据定期进行数据恢复验证,确保备份数据是有效和可靠的。此次数据容灾备份系统包括三级备份系统,系统架构复杂,实施难度大,在系统建成后,制订了详细的备份系统日常管理办法和应急恢复预

案及数据恢复操作手册,形成了完整的制度规范。

按照制度要求,除了日常对各级备份系统进行日常维护、故障预警、日志记录和运行评估外,还需定期进行数据恢复验证,其中对近线备份系统每月进行一次数据恢复验证,同城备份系统每季度进行一次数据恢复验证,异地容灾中心每年进行一次容灾恢复演练,通过模拟生产应用系统出现数据丢失或系统崩溃,从备份系统恢复丢失的数据或进行系统切换,并对恢复过程和结果进行记录。

6 结论(Conclusion)

此次容灾备份系统的建设,从前期的需求分析、调研测试、方案设计,到方案的实施、软硬件的安装部署,前后历经半年时间,目前系统运行状态良好,所有数据都能按照设计的链路进行传输,数据压缩和重复删除功能也都得到了实现,各级备份系统的数据经过验证都是有效的,均符合前期设计要求。该系统建成后,为企业的重要数据建立了多重保障,已多次利用备份系统找回因误操作和系统故障而丢失的数据,受到了广大应用人员的好评,极大的提高了南海东部海域勘探开发数据的安全性。

在数据容灾备份系统的建设及日常维护过程中,笔者有几点经验体会:一是前期要做好调研,结合自身的应用需求和现状,制订一个符合自身需求的建设方案;二是系统建设要通盘考虑,要结合软件、硬件、存储、网络、机房等因素,确保方案从技术层面是可行的;三是备份系统数据一定要进行恢复验证,要定期进行数据恢复演练,检验备份数据的可用性;四是系统建设和日常维护要进行记录,要制订相应的管理规定和操作手册,形成系统的技术文档和总结报告,并归档保存。

参考文献(References)

- [1] YU J,REN K,WANG C,et al.Enabling Cloud Storage Auditing with Key-Exposure Resistance[J].IEEE Transactions on Information Forensics and Security,2015,10(6):1167-1180.
- [2] YU Y,LI Y N.Public Integrity Auditing for Dynamic Data Sharing with Multiuser Modification[J].IEEE Transactions on Information Forensics & Security,2015(10):1717-1726.
- [3] YU Rongwei,WANG Lina,WANG Dejun,et al.Study on Security Enhancement Technology for Disaster Tolerant[J].Wuhan University Journal of Natural Sciences,2009,14(1):019-023.
- [4] 郑如秋,梁庭玮.一种分布式并行文件系统的介绍及在海洋地震数据处理中的应用[J].电脑与信息技术,2015,23(4):44-46.
- [5] 汪生珠,何庆兵,欧阳欣.集群NAS存储技术在石油勘探高性能计算中的应用[J].中国科技信息,2016(6):29-31.
- [6] 陈胜华,伍德雁.异地数据容灾备份与恢复的应用研究[J].科技讯,2012(13):45.
- [7] 王春才,闫磊,李英韬.基于 i S C S I 的网络数据存储技术研究[J].吉林师范大学学报(自然科学版),2014(4):98-100.

作者简介:

叶虹余(1984-),男,本科,工程师/信息系统项目经理师.研究领域:信息系统开发、运维及管理。