ID number: 2010660

Disaster Recovery Plan in Exploration Datacenter – A path breaking initiative

Pradeep Kumar Kukreja* Anshuman Borah

Oil and Natural Gas Corporation Ltd. Western Offshore Basin, Priyadarshini Building, Sion, Mumbai -400 022.

Presenting author e-mail: kukreja_pradeepk@ongc.co.in

"Planning is bringing the future into the present so that you can do something about it now." Alan Lakein

Abstract

Natural and human-induced disasters can wreak havoc on data centres and our E&P companies that rely on them can be adversely affected. E&P companies begin to recognize the need of robust backup and recovery strategies in order to support business continuity following such events. With increased data growth, regulatory concerns and adoption of virtualization, single-point solutions such as data replication to a single server, virtual and physical tape libraries and image-based backup do not always work in these situations. Building a robust Disaster recovery planning (DRP) for information technology has become mission critical for IT personnel to mitigate risk and maintain critical business operations. This paper will analyse the advantages of deploying in house on-premise backup and recovery solutions and elaborate how such solutions can help organizations improve their information management recovery in an effective manner without much impact on the cost. This paper will also highlight the recent in-house DRP for any kind of possible disaster in Interpretation Centre, Western Offshore Basin, Mumbai, ONGC so that same day interpretation work can be started in Disaster recovery site without losing interpretation data or man-hours.

Introduction

With emerging technology and high availability of data specifically in G & G Centres all over ONGC, maintenance of this high volume of data has become a big challenge. Survival and re-establishment of normalcy of business in the event of a disaster would be one the primary objectives of any business. Business viability itself can be compromised if built in mechanisms of data loss are not in place and it means if assurance of recovery in minimum time span are absent. It has been observed that there is no fool-proof DR plan for most of the G & G Centres in ONGC in case of major breakdown of system and / or data storage caused by various factors such as fire, failure of Server, utility power outage, floods etc. To have an alternate Interpretation setup at a remote location that can be brought up and running within 2-3 days.

According to a recent study by the Aberdeen Group, organizations without robust recovery systems in place can expect to experience more than four business disruptions each year, with recovery times ranging from one hour to more than nine hours. At an average cost of USD138,000 per hour, that price is simply too high for most businesses. (Source IBM)

The current paper discussed a new method of DR developed in-house and successfully implemented and tested using real on-line data. The proposed DR solution has much less cost implication than commercial DR solution available in the market today

Experimental Details for Interpretation Centres:

Interpretation centres maintain precious G & G data for carrying out interpretation work central to ONGC's exploration activities involving enormous man hours and efforts.

For interpretation work to continue with bare minimum downtime in the event of any disaster, it is essential to save all the Data and interpretation in the form of daily incremental backup, apart from full

initial backup, and maintain at an alternate remote site, connected by leased line of reasonable speed. It is a fact that most of the interpretation centres in ONGC are not equipped to face any disaster like situation such as fire, data storage failure, server breakdown, floods and restoration plan for their data thereafter. The methodology being that one-time effort to take two master backups for complete file system will be taken, on data cartridges and preserved at original centre as well as DR Centre. Further daily incremental back-up of Oracle database, horizons, Seismic volume and other third party data to remote location via lease line connection or local connection to be implemented.

Methodology adopted at IIWS, Mumbai:

- One-time full backup of all the file system containing project data of all the application software such as Landmark R5000, Petrosys, OpendTect, Hampson-Russell etc., were taken on LTO-5 Cartridges. The backup of about 70 TB of data took around 20 days to complete spanning around 28 LTO-5 cartridges.

- All the full backup tapes, after suitably labelling, were sent to EPINET, Panvel, a remote location 40 KM away from main IIWS centre for safe custody by keeping in fire-proof cabinet. It is extremely unlikely that earthquake/flood would simultaneously affect, the current Centre & Panvel (the DR site).

- Incremental backup of Openworks projects, horizons, Seismic bulk volumes, ‘mfd’, ‘zgf’ files of IIWS is being carried out every night on a NAS storage located at GEODEC, Panvel, using the 34Mbps leased line in automated way for all the file system. The automated scripts was in-house developed and designed in such a way that all project data backed up and if any error reported giving alarm on Server system. The scripts were installed on IIWS, Mumbai Dell R910 Server . Incremented Backup runs every day from 07.00 PM onwards and last till 12.00-01.00 midnight depending on the size of the increments files.

- The methodology above is an active-passive solution wherein, recovery achieved by restoring from tapes and could take some hours depending on the project size and in case of complete restoration of Centre data would take minimum of 2 days.

- Disaster Drill entailed no additional expenditure except in expenditure lease line bandwidth enhancement from 8 Mbps to 34 Mbps. However, this enhancement is also benefits other data transfer from EPINET, GEODEC Panvel to IIWS, Mumbai. Also DR solution implementation did not require any additional hardware/software and solution provided with existing infrastructure and manpower.

Considering all aspects of recovery point objective and recovery time objective in the context of providing high-availability of data and system for G & G interpretation, and the DR solutions commercially available which promise to provide active-active solution at a very high cost, the recovery process carried out in the drill is being adopted for both IIWS well as GEODEC interpretation centre of WOB in case of emergency disaster situation.

Backup Checklist

The Centre has multiple copies of backups from which one copy is stored offsite EPINET Panvel around 40 KM away from the main centre in the event of fire, theft, or another type of disaster. Also, if one copy fails, centre still have another chance to avoid panic. EPINET, Panvel, the Offsite has fire proof cabinet for safe custody of tapes in the event of fire at disaster site.

This Centre planned to make sure for enough backup media to keep a safe rotation schedule and clearly label the media. To avoid the situation where accidentally overwrite current month's backup is overwritten over last month's backup without being sure one could restore last month's just in case something was wrong with the current month's backup. A good media rotation schedule is outlined above in the master backup plan section which is called Grandfather-Father-Son.

Estimated Media Quantity: (Table: 1)

Backup Type	Quantity at IIWS	Quantity at GEODEC
6-Monthly(Master Backup)	25 nos. LTO-5	25 nos. LTO-5
Weekly (Third Party)	05 nos. LTO-5 Monthly	05 nos. LTO-5
Monthly	4 nos. LTO-5 and 2 nos. LTO3	6 nos. LTO-5
Home Weekly	1 no. LTO-5 at Symantec	5 nos. LTO-35

Before making the DRP laying policy implementation the centre has developed a written plan as to how to organize files, how the Server files system and/or network is structured, and how backups will be implemented and the effective restore plan by using in-house java driven menu (Fig.-1) which is helpful in visualizing it and communicating the plan so all users understand it.

Under the backup and DR policy all the backups are scheduled properly. The responsibility for backup should be specifically delegated and suitable arrangements made if the person responsible cannot be available for a scheduled backup.

Roles & Responsibility (FPR): (Table: 2)

Description of work	IIWS (Onsite)	GEODEC(Offsite-DR Site)
DR Script maintenance	Mr. XYZ	Mr. XYZ
DR Logs check-up	Mr. ABC / Mr. NM	Mr. XYZ
Weekly TP S/W data backup	Mr. BCD/Mr. NM	Mr. XYZ
Monthly Backup	Mr. JRM/Mr. ABC	Mr. XYZ
6-monthly backup	Mr. ABC / Mr. NM	Mr. XYZ
Restoration	Mr. ABC	Mr. XYZ
Maintaining Records/ Safe Keeping of tapes	Mr. JRM	Mr. XYZ

Retention Policy of data at Local DR sites:

The dataset created as part of DR are backed up near line (from Disk) to tapes monthly along with System Backup. DR data is retained online at both the sites (Originating and DR) for Minimum one month and maximum up to 3 months. The Master backup dataset created every half yearly and all incremental data sets received daily are to be preserved in DR site for one year.

Media type **:** LTO 5 / IBM 3592

Backup Type	Tape	Retention policy
6-monthly (Master Backup)	LTO-5	Keep two set at DR site and every third will be copy of first
Incremental daily	With Monthly Backup	Retain for 3 months online. Retain media for one year till two full backups.
Monthly	LTO-5	Retain at DR site for three sets and overwrite after every third.
Home & Third party weekly	LTO-5 by Symantec	Retain for six months at local site.

Since tape media are used for backup and safe-keeping, in order to ensure reliability in storing and retrieval of data, tapes are discarded after reuse for 2 cycles and new tapes shall be used for backup.

Results and Discussion

Since, the Meta data goes to the My-SQL database also information about the latest files can be obtained from the My-SQL database.

For this, a simple front-end application to the My-SQL database named “DR_INFO **(Fig: 1)** “has been developed to search for the latest files across all the incremental backups, project wise.

This can be used to search for any files (latest or old) within the incremental backups.

The scripts are developed in Java applets and menu are designed in such a way that every system person mentioned in (Table: 2) can restore either full or partial backup at site depending upon the requirement from the G & G users.

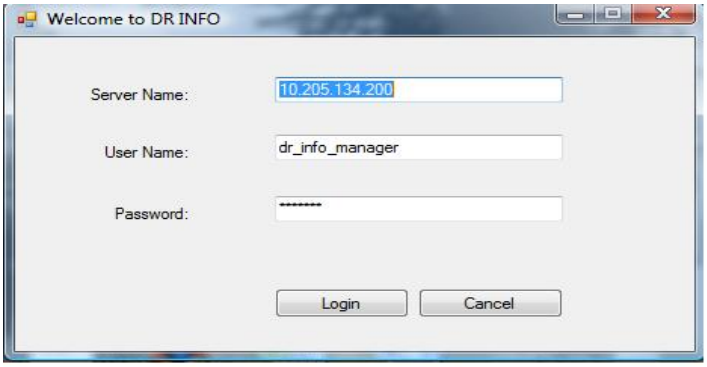
Type of Restoration

- **Project Restoration**: Everything within a project to be restored, namely OpenWorks Project and related seismic projects.
- **Particular File Restoration:** Particular file(s) with in a seismic project or a group of files related to a seismic or third party project.

These steps are for Openworks files restoration.

1. Project Restoration

- Identify the OpenWorks Project to be restored. Restore the latest OW back file from the DR files. Only latest oracle backup available must be restored.
- After OpenWorks restoration, identify all the seismic directories associated with the OpenWorks project, 2D and 3D both. Incase only a one or fewer seismic directories need to be restored then also identify the directories as needed.
- Now, restore these directories from the base tape backup. Since base backup are done as per file system, not as projects, directories may exists in different tapes. Restoration of all the required directories is essential.
- After restoration of the base backup from tapes, now it’s time to apply all the incremental backups available till date to the restored projects. Search for the latest files available in the set of incremental backups. These latest backups must be applied to the base project restored from the tapes.
- Searching for the latest files in the projects can be done from the text log files generated daily along with the incremental backup.



(**Fig. : 1)**

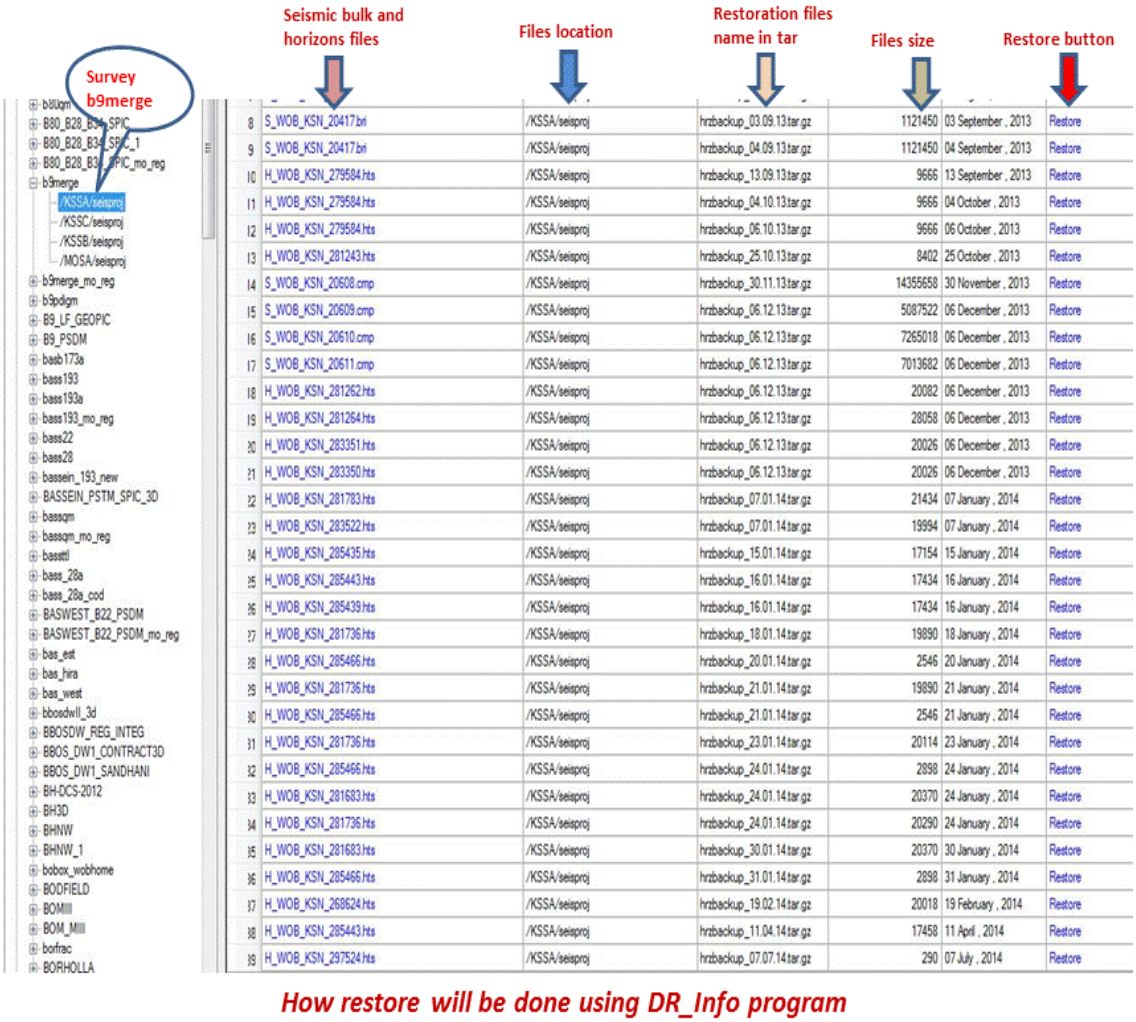
2. Specific File Restoration

- Identify the file to be restored and also the associated project directory (same file can exist in different directories).

- Searching for the latest occurrence of the file in the projects can be done from the text log files generated daily along with the incremental backup.
- Since, the Meta data goes to the My-SQL database also; information about the latest files can be obtained from the My-SQL database. Application “DR_INFO” can be used to search for any files (latest or old) within the incremental backups.

In case of disaster (for example at IIWS), required projects on which the work is to be continued, can be restored from the project backup and subsequently required external files from the Master backup are to be restored. Most recent incremental backup pertaining to the required projects is to be restored to come back up to the disaster point. Once this restoration process is completed, the interpretation work can be further continued from the point where the disaster occurred. Restoration time will vary depending upon the size of interpretation project.

The incremental backup script provides information from ‘My-SQL’ database to quickly locate the correct incremental backup (Fig: 2), relevant project data to be restored and correct tape or disk. The database contained all information of daily incremental backed up horizons, external backup etc.



(Fig: 2)

Example: (Please refer Fig: 2 above)

Seismic Survey “b9merge” (column 1) has Seismic data available in directories /KSSA/seisproj, /KSSC/seisproj, /KSSB/seisproj and /MOSA/seisproj including seismic bulk and horizons files. Next column 2-6 under backup files to be restored at desired location with size. Restore button available can be pressed for restoration of backup (Ref column 6 fig: 2). The program is designed in such a way

that even without the present of system person, G & G user can restore files automatically with their files location. But in case of major disaster i.e. Earthquake, Fire etc. the said DR system will act 'near online' or 'cold DR site' and the data recovery time may vary from 8 hours to 1 day and last modified interpretation work will be restored.

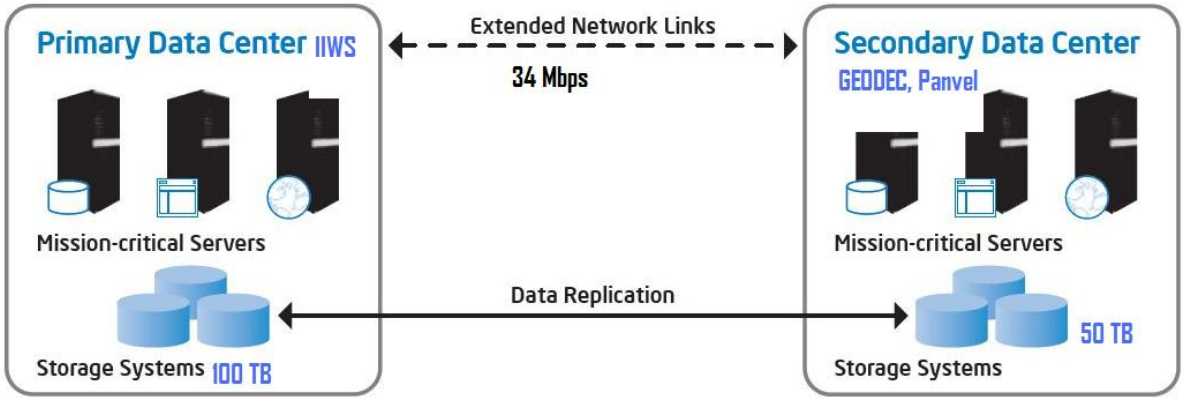
To achieve the goal, a disaster drill was successfully carried out by taking one running project NELP7 of Kutch-Saurashtra block from IIWS, Priyadarshini and complete recovery process was done at GEODEC, Panvel, an Exploration Centre having similar infrastructure and around 40km. away from existing place. The drill included the process of taking complete set of backup of all the file system about 40TB and daily incremental backup of Seismic volume, Horizons, database and other third party software to remote location at Panvel connected through 34 Mbps lease line.

Cost comparison with traditional DR solution:

In order to minimize Interpretation downtime and associated loss of revenue, IIWS team has successfully implemented a cost-effective business continuity (BC) and disaster recovery (DR) solution at GEODEC, Panvel by "stretching" mission-critical systems between two different data centres within the same facility. Even a small glitch in Interpretation operation can cost millions of rupees and impact annual work plans. A cost-effective and efficient DR solution is vital to our business—so that if downtime does occur for any reason, an interpretation centre can resume operation as quickly as possible.

Cost Effectiveness and Investment on DR are directly revolve on two main questions -

a) What is the Recovery Time Objective (RTO) that is, how quickly one needs to recover from a data disaster?
b) What is the Recovery Point Objective (RPO) How much data can one afford to lose between the disaster and the last backup. With the current workflow described in this paper in the forgoing we have been able to demonstrate that maximum advantage leveraging existing infrastructure and investment resources is possible to satisfactorily adore current business.



(Fig: 3) Schematic Representation of the DR Solution

Please refer to Fig: 3 above where 34 Mbps lease line is already connected between IIWS & GEODEC for backup and other transfer of data work cost 4.54 lacks annually which are using for transfer the DR data to and fro primary and secondary data Centre. Existing infrastructure including Linux Server Dell R910, NetApps 50 TB Data storage are already in-placed on DR site which are serving the purpose without incurring the expenditure on hardware. Existing Cartridges are used for master backup of data and finally No other cost is involved for this active–passive DR solution whereas traditional DR cost would in between INR 30-40 million with average AMC cost of 10 million annually.

Conclusions

Many disaster recovery solutions focus on replicating data to a remote site, leaving IT departments burdened with the complex task of reconstituting servers, applications, network configurations, and replicated data into a functional set of data centre services for business continuity. Ensuring that Data Centres deliver with high Availability - Securely and Efficiently - is one of the foremost concerns for a business to sustain and ideal growth with NIL downtime. IT strategy and competitive advantage is built-up on optimizing applications and supporting the immediate needs of the business. Unfortunately, DR services come either at very high cost or with weak guarantees about minimal amount of data lost and long time required restarting operation after a failure unless at a very high cost for short recovery times post disaster.

The current work reported is one way achieving effective Disaster Recovery entailing minimum additional financial outlay, with nil additional deployment of hardware and HR personnel. The method demonstrated is simple and is practically maintenance free to run. Our dry runs have demonstrated that all the design goal have been met with successfully.

References

1. Feasibility Report on Disaster Recovery for IIWS for all G & G Centre in ONGC 2012.
2. IT@Intel Perspective, IT Best Practices Data Centres and IT Continuity, November 2010.
3. Best Practices in Business Continuity and Disaster Recovery by Laura Dubois February 2011, Riverbed Technology, MA, USA P508.8200
4. Automated Service-oriented Disaster Recovery, Technical White Paper, FalconStor Software, Inc.

Acknowledgments

Authors express their gratitude to Shri N K Verma, Director (Exploration) & Shri A.K. Dwivedi, ED-Basin Manager, Western Off-Shore Basin, Mumbai, ONGC, for giving kind permission to present this paper. Authors are fully indebted to Shri Jai Singh, GM (GP)-Head Database, Shri A V Satyanarayana, DGM (Prog.) and other fellow colleagues in Database Group.

We are also grateful to all our colleagues who have helped us with valuable support throughout the writing of this paper. Last, but not the least, we would like to thank God and our families without whose blessings and constant encouragement, the fulfilment of this paper could not have been envisioned."