PARS

White Paper

Published: January 2005

Application Project Archiving with EMC^2 Centera

Protecting assets through capture of point-in-time knowledge

CONTACT DETAILS

United Kingdom Office:

2-7 Horsted Square
Bellbrook Business Park
Uckfield
East Sussex TN22 1QG
United Kingdom

Phone: +44 (0) 1825 745100
Fax: +44 (0) 1825 767031

United States Office:

Enigma Data Systems Inc.
16000 Barkers Point Lane
Suite 202
Houston
Texas 77079
United States of America

Phone: 1 281 497 0440
Fax: 1 281 497 0402

Online Details:

Enigma Web Site:
www.enigmadata.com

Sales E-mail:
sales@enigmadata.com

1 Executive Summary

Oil and gas companies rely on hundreds of applications to find and develop their reservoir assets. The knowledge generated from these applications is used daily to make critical business decisions. In fact, knowledge is the lifeblood of upstream oil and gas companies, and re-using that knowledge; making it available enterprise wide is a key competitive edge. Traditionally, these companies have had to balance the cost of managing their archives in a detailed and efficient manner with that of cheaper homegrown solutions.

Now petroleum companies can harvest meaningful knowledge from the raw seismic data, well data, and various interpretation projects used in the pursuit of finding oil and gas. This is accomplished by using a new solution made possible through the partnership of Enigma Data Systems and EMC. This totally new solution helps to justify a robust archiving strategy by combining IT and Data Management concepts that can turn up the value of historic data for years to come.

EMC and Enigma's archiving solution enable customers to leverage their current and historic reservoir data assets in an efficient manner such that those assets are always easily searchable, immediately available, and reliably stored on the most cost effective media. This technology can be implemented quickly to help customers increase their return on investment more rapidly.

2 Business Need

Since they started using computers, oil and gas companies have been generating digital data related to their reservoir assets. Over the years, the computing infrastructure has become increasingly more powerful; following Moore's Law, the number of transistors in microprocessors has doubled every couple of years since its inception in 1965. Online storage capacity and network bandwidth have also increased in a similar exponential fashion. Combine that with the current plethora of petrophysical data processing and analysis application packages, and you end up with Terabytes of reservoir asset data being generated today. This whole cycle will continue for the forseable future. The question is: how do we manage all this data now and in the future?

It is important to note that each of these applications has their own methodology for storing, and accessing data related to the reservoir[s]. This data then represents the Oil Companies' knowledge of the reservoir asset at any point in time.

Throughout the reservoir's life cycle, new data must be incorporated into the existing reservoir model to enhance the knowledge of it, while archiving previously exploited information assets that are no longer needed for daily operations. Today most companies recognize the value of retaining this knowledge for several different reasons, but have struggled with a comprehensive, vendor neutral, cost effective solution.

In general, data is most used and therefore most valuable when it is new. This is also when knowledge about the data is most fresh in all the geoscientists' minds that have recently accessed the data. Over time, new data is created from the old data, such that the old data is referred to less often. This is when archiving becomes important.

Another important point is that the reservoir asset data is usually analysed in what is called an interpretation or application project. This geophysical project contains a collection of data related to a specific business purpose - in this case, finding an oil or gas reservoir. There are many application projects that can exist during the life cycle of a reservoir.

2.1 So, Why Archive a Project?

The general reason to archive projects is to preserve their state indefinitely, with the intent that they can be recovered at any later date. In support of this process, metadata should be added to the archive to ensure that there are many ways to search and qualify a project before restore. Another good reason to archive is to preserve little-used or unused data on less expensive media. Some might say, "We backup our data. Why should we then archive it as well?" The following table best explains the answer by contrast:

	Backup	**Archive**
Purpose	To ensure against system or human error, allowing restore on an exception basis. The intent is not to delete data to free disk space.	Catalog and index relevant data so that it can be stored indefinitely and later retrieved. Can optionally free disk space by removing project.
Time span	Finite, according to policy. Typically 1 year for full backups.	Unlimited
Granularity	File or physical volume (i.e. file system)	Project (i.e. multiple, related files, databases etc.)
Method	Usually scheduled automatic. IT Dept.	User initiated as business requires – Data Managers
Restore Criteria	Searches based on path and filename	Searches based on descriptive metadata and indices.

The above shows that for un-accessed or end of life (project) data that might be needed in the future, and is taking up valuable disk space needed by other projects, it might be prudent to archive it and

then remove it from disk. In the future, a user initiated archive search would include meaningful metadata, and the potential restore process would bring back all the files needed for the project. These related files could potentially span multiple file systems. Restoring an archived project from a project archive is much easier than restoring a project from a file system backup. This is because a user may not know the locations of all the files for a particular project in order to select them from the file system backup.

2.2 The Problem

The geoscientists would like all of their data to be online, all of the time. However, this can (and does) lead to competition for storage resources. And if those resources are not available in a timely fashion, then the rate at which oil companies can find reservoirs is detrimentally impacted.

Production disk is fast but expensive, especially if one is trying to keep all their growing data online all the time.

Tape based and secondary disk systems are cheap, and generally do not have the high availability of modern production disk sub-system.

The process of retaining reservoir knowledge is difficult to manage due to disparate applications that generate data over multiple file systems in various data types and formats. This basically means that the corporate knowledge of a reservoir is spread all over the place in various stages of organization. The geoscientists who work on this data are often berated by the IT or application support groups for not cleaning up files related to dead prospects and or not archiving. But the truth is that they are not measured by how well they manage this data. Often there is no place or easy way that a geoscientist can archive their exploited project data with any reliable assurance that it can be restored to production disk.

Very few of the countless Petrophysical applications come with their own archive solution. These solutions are more or less developed, and are definitely not vendor neutral, nor do they address the back end issues of reliability, security, and availability. So, many companies are forced into developing home grown, scripted solutions that do not scale with the number of applications, or the burgeoning data behind them. Additionally, the solutions usually do not involve a centralized, database to capture relevant metadata about the archives for meaningful searches in the future. The result is a system that courts data loss through human error.

3 Framing a Project Archive Solution

Constructing a robust archiving solution requires a good hardware infrastructure, good archiving software as well as the policy and procedures to back it up. Archiving is a human process involving the capture and addition of knowledge metadata to the archive. No human process is as good as it is when there are written and enforced, policies and procedures to govern it.

To start with, the hardware should support the following concepts:

- Read only or fixed content data – The data cannot be changed once archived.
- Data availablility – The data will always be "online" with no access wait times.
- Data reliability – The data will always exist in its original uncorrupted state.

The software should support these concepts:

- A project – the fact that subsets of reservoir asset data are related by the application that accesses it and in that respect should be considered a project.
- Metadata – for easy archive searches in the future when those who have knowledge of the project have "moved on".
- The ability to verify an archive.
- The ability to restore an archive to the original or different location.
- The ability to restore a subset of files from an archive.

Finally a policy and procedures document should be drafted to cover and answer the following concepts:

- Policies:
  - Roles and Responsibilities
  - What data should be archived
  - When should data be archived
  - Naming standards and conventions for metadata
  - Archive security and robustness
- Procedures (tuned to the policies):
  - How to archive a project
  - How to restore a project
  - Metadata maintenance

4 The Enigma, EMC Value

Everyone knows that retaining information assets is important, and becoming more important every day. The combination of PARS, Linear Disk, and Centera provides the software and hardware needed to implement a solution to protect reservoir asset data through the process of archiving.

4.1 EMC^2 Centera

EMC Centera is the world's first Content Addressed Storage (CAS) solution and the world's leading on-line archival storage media. Centera is purpose-built to address the unique information storage requirements of long-lasting, unchanging digital objects—or "fixed content"—such as check and medical images, enterprise content, e-mail, and more. Centera resolves two major challenges that have prevented organizations from bringing fixed content on-line – the increasing volume of fixed content and the cost and complexity of managing it. The Centera solution is self-configuring, self-healing and self-managing to deliver cost effective on-line access to fixed content assets with assured content authenticity. Centera enables one full-time employee to manage up to 350 terabytes of fixed content. Centera's unique features such as authentication, has prompted its industry-wide recognition as the first disk-based WORM (write once/read many) storage media and has established it as a compelling solution for addressing the needs of corporate governance and regulatory requirements.

4.2 ENIGMA PARS

Flexibility: By providing the flexibility to bundle data related to different applications and business needs, the oil and gas enterprise can ensure the organized long-term retention of corporate knowledge.

Metadata Database: PARS gives users the ability to, flexibly, add both manual and automatic metadata to the archive that describes things like:

- The archive type
- The archive date
- The archiver
- The archive initiator (project lead)
- The reason for the archive
- The application version under which the project data was created and interpreted
- And much, much more….

Now the enterprise can give its interpreters what they desire – all their data online, all the time – without keeping it all on the most expensive storage systems. The PARS archiving solution enables the archiving of lesser priority projects, and finalized projects to the Centera storage. After which, theses projects can be deleted, freeing up production disk for new, more active projects.

5 How it Works

The archive solution is comprised of the following components: Enigma's PARS and Linear Disk applications in conjunction with EMC's Centera storage solution.

PARS is Enigma's Project Archive And Retrieval software. It is a web based project oriented package designed to query GeoScientific applications for project lists and project directories to find the project data to archive. PARS gives users the flexibility to bundle and manage project specific data spread over several file systems on multiple servers. Users can define exactly which types of data comprise a given project, and archive their selections.

Linear Disk is a virtual disk file system. It interfaces with the Centera Application Program Interface (API) in order to present the Centera storage system as a normal Unix file system. See figure 1. Linear disk operates with a small slice of local disk connected to the Enigma server. Additionally, it has process daemons monitoring the Linear Disk (/LD) file system, and a database to link the archive files with the BLOB data objects on the Centera. The clip entry shown in figure 1 represents the pointer to the actual archived data object stored in the Centera.

EMC Centera is the first content addressed storage (CAS) system designed exclusively for fixed content. It dramatically simplifies management, ensures content authenticity, and delivers petabyte level scalability. It is used in this archiving solution as the destination for the project archives.

5.1 The Archive

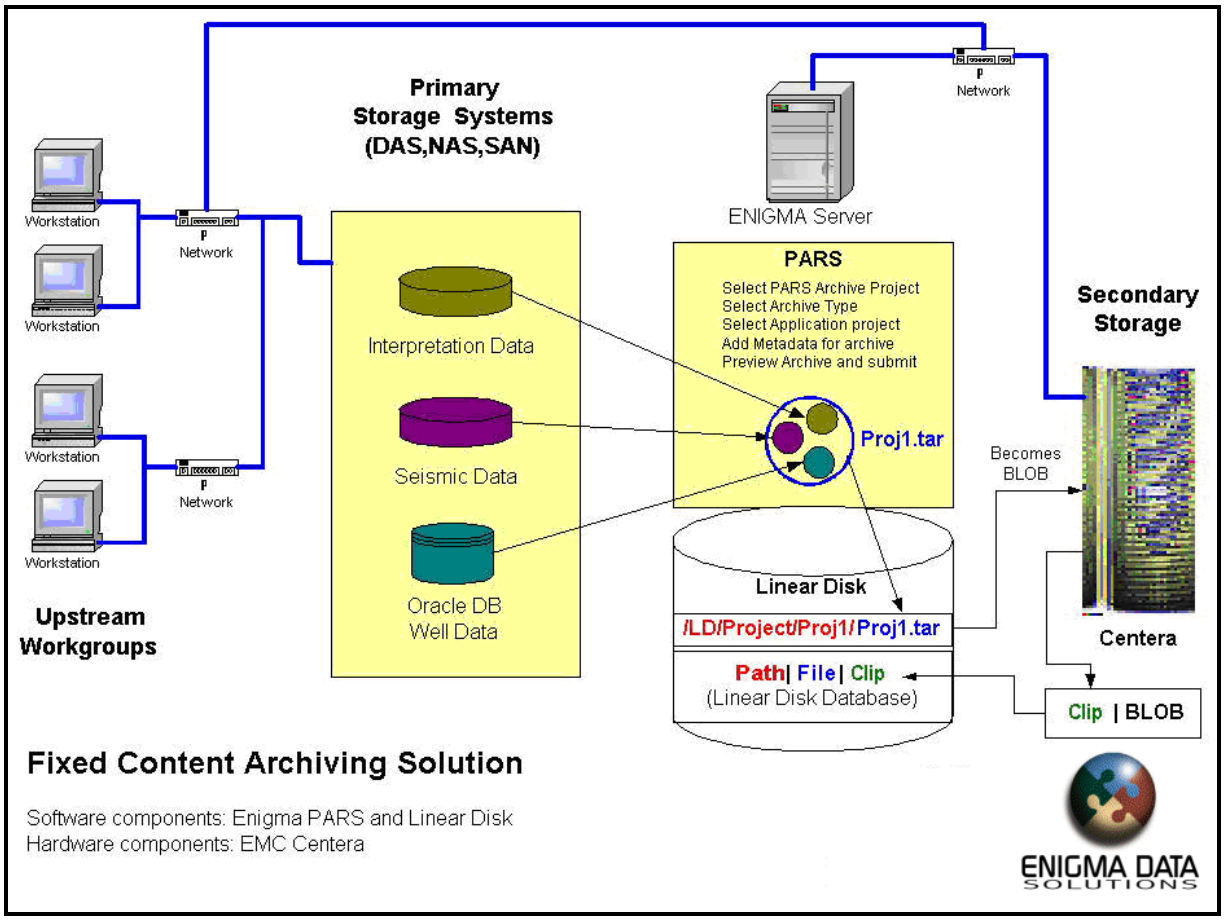
A user logs into the PARS application from a web browser. They then select the archive option. Next, they select a PARS project (logical archive container) from a list. These are usually created in advance based on an area of interest encompassing at least one seismic survey. Next they select the application type that they would like to archive (e.g. SeisWorks, GeoFrame, etc.) and then the create stores option. They are presented with a list of projects for the application type selected. After selecting one application project, they prepare the archive job by adding additional metadata about the archive such as the reason for the archive, the application release, the destination LD file system, etc. They may preview the archive at this point to see how many files it will contain and how big it will be. Finally they submit the archive job.

The job begins gathering the project data and opens an archive saveset on the linear disk file system. The project data begins to stream into the archive saveset. Linear disk 'passes' the data to the Centera, after asking for a content address pointer (clip) under which to store the new archive. The Centera returns the clip and then the Linear Disk processes begin to stream the data to the Centera, where it will be stored as a BLOB or Binary Large Object.

5.2 The Restore

The restore of an archive is done largely in reverse to the archive. The user still logs in to PARS, but then selects the restore option. They can then use filter and sorting tools to find the desired archive stored by PARS. Once the archive is found, they select the retrieve stores option. At this point they can preview the restore job to see how big it will be and if any files will be overwritten. Space should be made available if necessary at this point. Once satisfied that space is available for the archive, they submit the restore job.

The job begins to open the archive saveset on the Linear Disk file system, which in turn requests the appropriate BLOB from the Centera, based on the clip file associated with the tar file. The data is then extracted by the restore job and placed back on the production storage from which it came.



Getting Started

The seamless integration of Enigma Data Solutions' PARS and EMC^2 Centera is available now.

To learn more about this solution, please call an Enigma Data Solutions' sales representative.

US: +1 281 497 0440

UK +44 1825 745 100

Email: sales@enigmadata.com
Web Site: www.enigmadata.com