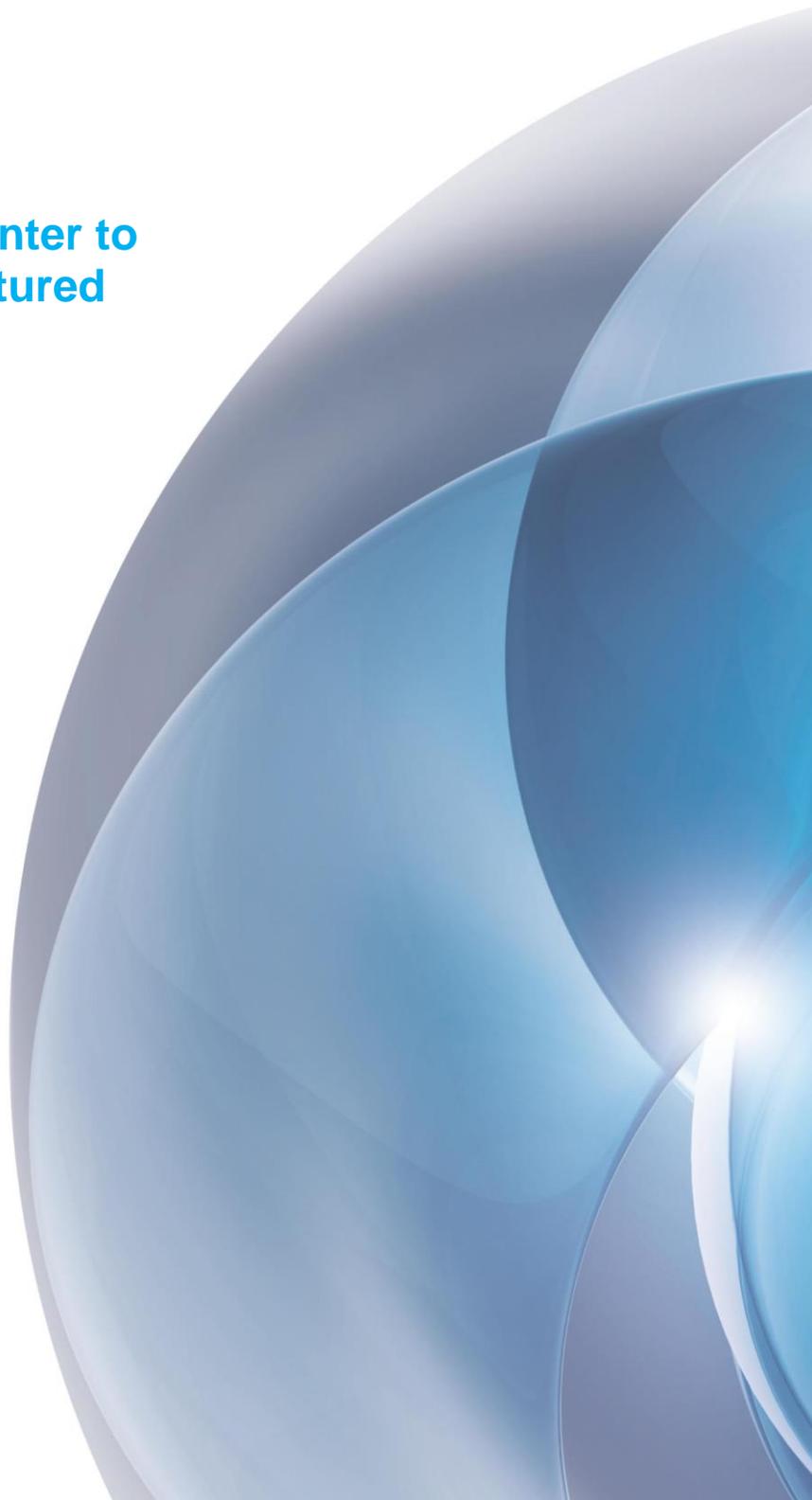


CASE STUDY

Using Informatica PowerCenter to
Extract Value from Unstructured
Data

altius[®]



SUCCESS STORY

Automatically Extracting Data from PDF Files into a Database

Our client, one of the world's largest oil and gas companies, captures drilling performance data from its external vendor operated wells. Since data is received as PDF files in many different formats, reporting was previously cumbersome and complex. Their process for basic reporting was manual and required users to read and enter key data into another system. Further, the reporting process was not scalable, making it time-consuming and in-effective. Our client needed a better solution.

Extracting value from unstructured data files has always been a tricky task with no obvious solution. Attempts to design a custom solution would undoubtedly take more time than was practical or worthwhile to do and would involve a large amount of hard coding. To fit the client's needs, any implemented solution would have to be flexible and cope with frequently changing source files as well as be quick to implement.

Enter Informatica PowerCenter and the Unstructured Data Option. solution.

Client Profile

Country: United States / Global Team
Industry: Mineral Extraction

Overview: The Exploration and Production (E&P) segment of one of the world's largest petroleum and petrochemicals companies, operating in offices in 26 countries.

Business Situation: The client's Drilling Performance team needed a solution to help them capture and report on drilling and performance data. They received many differently formatted PDF files and had to manually read the data – this was inaccurate and slow.

Solution: Altius recommended the client implement Informatica PowerCenter to automate the reading of data from the PDF files into a Microsoft database where an advanced reporting solution could be developed.

Benefits :

- Automated the end-to-end process, thereby saving countless person-hours.
- Enabled a comprehensive and flexible reporting platform to be built where drilling performance could be analysed more effectively.
- Scalable solution provided to cope with any number of drilling reports and any number of formats of source PDF files.

SUCCESS STORY

What we say

“Having Informatica PowerCenter as part of an enterprise company’s data management toolkit can help develop solutions for a whole host of challenges that were previously literally unsolvable.

Companies can now gain access to powerful information from their data assets at a cost effective price point compared with the limited alternatives.”

Noel Phillips, VP – Consulting Business, Altius

Situation

Our oil & gas client received dozens of drilling reports each month in multiple formats and were manually extracting the key data they required by hand.

The data stored within the files was used for three main purposes:

- Monitoring operating performance
- Accurately determining when new wells came online and started producing
- Tracking and controlling spend

Despite the plethora of information available in the reports, only a small subset was being recorded because of the high overhead associated with capturing the data. In mid-2009, due to an acquisition, the volume of drilling reports received monthly by our client increased dramatically to many thousands per month. Clearly it was not practical to continue capturing the data by hand and an alternative approach was required.

Solution

After examining the few available options, we advised our client to invest in Informatica PowerCenter with the Unstructured Data Option. This product gave us the ability to seamlessly access, discover, and integrate data from virtually any business system, in any format, including data locked in documents and industry-specific data formats. The ability to access and process any data format increased the team’s productivity and saved costs by eliminating the need for manual intervention and improved business responsiveness and agility in their competitive market.

In our solution we implemented the following process to import the data for reporting and further analysis:

1. Various multi-format PDF documents were received into a Windows file system folder (As an alternative to the Windows folder it is possible to extract file straight from an Exchange Mailbox)
2. The folder regularly parsed for new files, processed each file, and extracted the data into a SQL Server database
3. SQL Server Reporting Services (SSRS) was used to provide various on-demand reports

SUCCESS STORY

What they say

“We have been an enthusiastic proponent of combining support for both unstructured and structured data in a single data integration product. With this functionality, PowerCenter helps us achieve greater productivity, drive down costs, minimize operational risks, and accelerate the time to market for new data integration intensive enterprise applications.”

Senior Vice President and CIO

Benefits

This solution solved the immediate needs of the business and from here many further improvements could be made. The key point is that once the data was in the database, it could be used in many different ways to provide value to the business. Some of the primary benefits realized follow:

Expand Data Visibility

The solution expanded visibility into all enterprise data, thereby increasing its business value, reducing risk whatever the source file format, and uncovering formerly hidden value from previously unusable data sources.

Increase Productivity

By automating the end-to-end process, countless person-hours were saved. The solution removed the need to manually process every PDF file and allowed the business to get on with running the business. Each unique template needed only to be defined once and then data extraction was automated.

Comprehensive Reporting Platform

Storing the data in the Microsoft SQL Server platform enabled a comprehensive and flexible reporting platform to be built and allowed drilling performance to be analysed more effectively. By targeting limited resources where they were needed, most investments could be made more effective for less.

Flexible & Scalable Solution

Informatica PowerCenter is a fully scalable solution that can cope with any number of drilling reports and any number of formats of source PDF or other unstructured data files. Knowing the solution can grow with your changing needs gives you a future proof investment.

Integrate into Existing Platform

Informatica PowerCenter can easily integrate into your existing ETL processes and form part of a larger workflow. By no means is this a standalone solution and the options are as impressive as they are extensive.

Software Products

- Informatica PowerCentre with Unstructured Data Option
- Microsoft SQL Server
- Microsoft Analysis Services
- Microsoft SharePoint Server

SUCCESS STORY

For further information

For more information about Altius services and products, contact your local office:

Corporate Headquarters

2245 Texas Drive
Suite 150
Sugar Land, TX 77479

Alaska Office

3300 C Street
Suite 102
Anchorage, AK 99503

+1 800 816 4850
info@altius-usa.com
www.altius-usa.com

If understanding your business better is a high priority, or you would just like to discuss what Business Intelligence and Performance Management options may work for your business, please contact us today.

About Altius

Altius helps companies better understand the data that drives their organizations. We are practitioners who have personally managed successful Business Intelligence initiatives. From this, we have developed a suite of products that helps you save time, reduce errors and quickly execute projects. At the end of the day, we help you make intelligent business decisions from your data.

Altius has the experience of over 500 project implementations, spanning companies in oil, gas, energy and utilities. Altius is head-quartered in Houston, Texas with a second office location in Anchorage, Alaska.

