OvationData

THE ULTIMATE GUIDE
to
Data Cleaning & Ingestion

Best practices in the oil and gas industry

Clean data

Streamlined organization of data

Usable data

Improved ability to manipulate and use data

Quality data

Smarter decisions based on more reliable information

2.7Mpvs   40   17min

Introduction

The oil & gas industry is an extensive sector known for its high stakes, capital-intensive projects, and a notable pioneer in the domain of Big Data. Given its expansive operations, this industry generates colossal amounts of data from diverse sources, both in digital and non-digital formats.
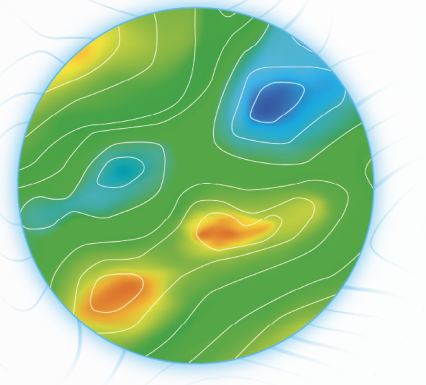
As companies venture into the evolving digital age, they are experiencing a transformative shift catalyzed by advancements in predictive analytics, data science, and machine learning. These technologies hold immense potential for **optimizing operations**, streamlining **cost management,** and augmenting **decision-making processes.** The power of these innovative technologies can be unleashed only when the underlying data is managed efficiently.

Operational efficiency

The oil & gas industry produces petabytes of data, ranging from operational data to sensor readings from equipment and maintenance records. The vast volumes and complexity of data types in oil and gas operations can pose significant challenges, and understanding these is crucial for effective data management and operational efficiency.

The cleaning and preparation of data - a process known as Extract, Transform, and Load (ETL), can account for up to 50 percent of a data scientist's time. An ETL workflow extracts data from various sources, transforms it into a format that can be easily analyzed, and loads it into a database for further use. This workflow is commonly used in data warehousing, business intelligence, and analytics applications. In this process, data is extracted from various sources, such as databases, spreadsheets, and APIs. The data is then transformed by cleaning, filtering, and structuring it into a format that can be easily analyzed. Finally, the transformed data is loaded into a database for further analysis and reporting.

Improved decision-making

The ETL workflow is critical for organizations to gain insights from their data and make informed decisions. A recent Ovation study examined challenges associated with ETL and data management, specifically related to upstream oil and gas operations.

Ovation's research used existing industry data models for reference, particularly the **Open Subsurface Data Universe (OSDU)** and the Professional Petroleum Data Management (PPDM) model. These models represent well-established frameworks for managing oil & gas data, and understanding their structure and functionalities can provide valuable insights for an effective ETL process.

OSDU is an ongoing innovation in industrial data management, addressing a major hurdle faced by several industries—the siloed nature of data. The platform embodies the dual aspirations of the industry - operational efficiency and innovation. By standardizing data management and supporting diverse data types, platforms such as OSDU allow the oil & gas industry to truly harness the power of its data and drive forward in the digital age. The platform supports most data types prevalent in the energy industry, and its open-source nature promotes the integration of workflows, thereby enabling faster deployment and better decision-making.

Cost management

The Geographic Information System (GIS) data cleanup technology plays an essential role in the ETL process by integrating, analyzing, and visualizing spatial data. This discipline has traditionally been dominated by ESRI, a long-time industry leader. However, open-source tools like QGIS are rapidly gaining traction. These tools provide excellent functionality for ETL processes, offering a high-quality solution at a substantially lower cost than ArcGIS.

Recently, Ovation addressed the time and cost challenges for Carbon Capture & Storage (CCS) monitoring projects and determined that it is essential to integrate data management with data analytics for the type of long-term projects that often occur in the oil & gas industry. Key considerations include:

- Affordable petabyte-scale data storage spanning decades.
- Indexing and management of data history and data dependencies using a 5-layer data model.
- Automations on metadata extraction and processing to reduce the human intervention needed for review.
- A Web GIS portal to facilitate data and information sharing among all project stakeholders.

Ovation's approach aims to seamlessly integrate data management with data analytics, ensuring that both knowledge and data are cohesively archived. This is particularly critical for monitoring projects that might span over half a century.

Data Clean-up Checklist

Techniques used in Data Cleaning may vary depending on how your company stores data. As a great start in formulating a framework that suits your specific needs, Ovation's Data Cleaning process includes these steps:

Step 1. Audit your data

- Perform system health check (QC) to identify departmental data siloes.
- Run data quality assessment to spot bad data.
- Automate the process to accommodate massive databases.
- Search for duplicates and perform deduplication on multiple records or repetitive data.
- Validate your data and accomplish data migration.

Step 2. Manage your data

- Standardize and assimilate your data.
- Use a single platform for easy fast access.
- Manage all data types including any legacy format.
- Archive data not currently in use.
- Manage risk with safe cloud storage and disaster recovery capabilities.

Step 3. Enjoy the results

- Experience the benefits of Data Cleaning.
- Gain a clearer understanding of the data you have and how to find it.
- Save time and money by eliminating tedious unproductive searches.
- Improve quality of business decisions based on enriched information.
- Breakdown information silos to promote knowledge sharing.
- Reduce production bottlenecks.
- Enable adoption of cutting-edge digital solutions.
- Unlock greater value that is currently hidden in your data.
- Examine and apply the latest technology innovations.
- Explore real-life strategies gained through digital transformation.
- Streamline integration and simplify data management.
- Lower downtime and reduce operation costs.
- Promote a culture of innovation to support greater profitability.

Even greater benefits can result further down the line when your organization is set up for success. Data Cleaning has many potential domino effects that can boost ROI.

Top 5 Questions to Ask Your Data Management Partner

When evaluating a potential partner for any of your data management needs, be sure to ask about the details. Below are five questions to consider when selecting a transcription provider:

#1: Does the provider digitize the data in such a way that it is actually useable?

Some companies can digitize data and identify the metadata, but they do it in an inferior way. Then they enclose it within a wrapper that clients cannot penetrate unless they buy additional services. Ovation has a different approach. When clients want to retrieve their data from Ovation, we recognize that they have already paid for tapes and maintenance. We do not use a wrapper to store the data, except of course if the client requests encryption by choice. We never charge egress fees. Other good news is that when other providers ask their clients to pay extra costs to retrieve their own data, we can break the data loose so clients are able to use it.

To ensure your assets are preserved, protected and accessible on demand, we have architected a unique blend of traditional and cloud-based solutions backed by security, performance and immediate access.

#2: How useful are "bells and whistles?"

Make sure you actually need what some providers brag about. For example, some companies are attracted by the idea of managing their data with viewers via a web portal, but this has little or no value when you see magnified views of only inferior data.

#3: Will a "one size fits all" approach meet your needs?

Different organizations have different goals, and you want customized solutions to fit your specific needs. Even if you are looking for only transcription services right now, keep in mind the bigger picture. Our suite of customized solutions provides the full end-to-end data management environment at every stage of the data lifecycle.

#4: Does the provider have broad experience across multiple industries and geographic locations?

Take advantage of Ovation's international reach. With data service projects spanning 50 countries, our team is effective with cross-border projects, often mobilizing to the client's country to perform data services.

#5: Is your provider fully equipped to handle full transcription?

Some transcription providers may generate a "bit copy," but they use the same old format that is no longer readable, forcing clients to wait longer to read the data. With the most comprehensive selection of tape drives and devices, our equipment supports more than 60 years of technology. We transcribe data from historical tape media to the current industry standard media and incorporate real-time quality control. Our processing facilities include a comprehensive inventory of historical equipment and proprietary equipment to read and recover data from every type of media. A special arsenal of rare and customized equipment and an inventory of over 800 devices set us apart, along with the ability to handle 250+ different media types including large volume seismic formats. Scroll down below for a more complete list.

About Ovation Data

Ovation Data is a full-service provider of digital transformation and data repository solutions. It provides secure and scalable data management services and support from basic and complex infrastructure to cloud-based solutions for data transmission, storage, stewardship and loss prevention. For more than 45 years, Ovation Data has securely delivered high-quality, connected, and accurate data that informs clients' business decisions, optimized with speed and confidence.

The Ovation Method

Our suite of customized solutions provides the full end-to-end data management environment at every stage of the data lifecycle. Liberate and integrate your data effectively with our expert guidance and proprietary technologies. Here is how it works:

1. Discover
Audit, validate and verify data type and quality. Check entitlements so you can prioritize and evaluate your inventory.

2. Transcribe
Automated methods that allow information of all kinds, in all formats to be intermingled through accessible technology.

3. Enhance
Sort, filter and aggregate by using advanced quality control and processing techniques. Interrogate and optimize data for cloud and analytical tools.

6. Preserve
Protect the safety and integrity of your data through secure accessibility. Maintain and preserve the longevity of your digital inventory.

5. Connect
A highly secure, scalable, on-demand network that delivers any data anywhere, at anytime.

4. Analyze
Custom-built systems that apply machine learning & artificial intelligence to accelerate insights and drive innovation.