Cloud Pathway to Machine Learning for Subsurface Interpretation

Paul Schatz*, Interactive Network Technologies, Dan Kahn PhD, Amazon Web Services, Dmitriy Tishechkin, Amazon Web Services

Abstract

Over the past few years, the oil and gas industry has undergone a series of major changes. Companies must improve operational efficiency, optimize production, and minimize non-productive time. To achieve this, companies must make better use of their data. Data access, and the difficulty of finding the right data in a timely manner, remains a constant and ongoing challenge for Energy companies. The traditional approach to geoscience and petroleum engineering at energy companies creates multiple stand-alone data silos that sit in multiple locations. These problems prevent companies from introducing new workflows, such as machine learning, that can significantly speed up exploration and decision-making.

Data lakes are a way for companies to unlock their data, make it discoverable and searchable, and reduce the silos that currently plague the industry. There are multiple data lake offerings supporting OSDU, the E&P industry standard platform that promises to put all subsurface and well data at the center of a single Data Platform, utilizing standard data formats and providing a well-defined set of APIs to locate and access the data easily. Augmented by a modular visualization platform that supports multiple data sources including OSDU, custom data catalogs, and data lakes, the tools are available that allow companies to re-engineer their workflows and incorporate new technologies such as advanced visualization and machine learning.

Introduction

Data access is a constant and ongoing challenge for Energy companies. The bulk part of this data includes seismic volumes, well log data in various formats, and historical production logs. Access to the data can be a multi-dimensional problem. As prime examples, data sharing needs to occur for joint venture enterprises or for collaboration inside multi-discipline team. However, data are often siloed in different data centers or tape storage that might be located in different geographic regions. These factors further complicate access problems.

Another prominent problem is the difficulty of finding the right data in a timely manner. The traditional approach to geoscience and petroleum engineering at energy companies creates multiple stand-alone data silos that sit in multiple locations, often around the world. Data are usually stored by asset (field or geographic area), and within the asset, data are partitioned even further by discipline such as geophysics, well-planning, or production.

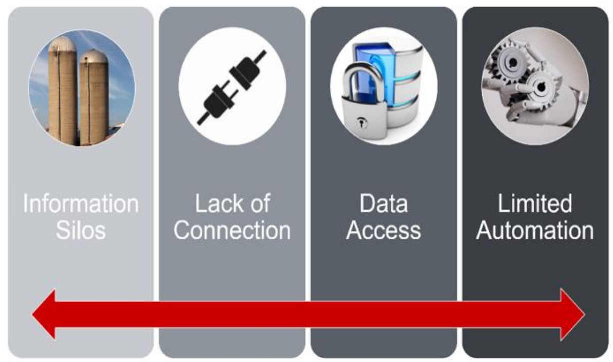

Figure 1: *The Digital Challenge*

Such problems prevent energy companies from introducing new workflows into their processes. New workflows like AI/ML can significantly speed up exploration and the planning cycle, which lead to better informed decisions and precise reserve estimations.

The solution to these problems is to bring data together in a centralized location where all interested parties can obtain access for their various purposes. This process consists of 2 parts: Migration to the cloud, and Indexing. There are multiple tools to move data from datacenters, tape storage and local offices to a central storage in the cloud. For large data volumes, offline data migration services are available such as AWS Snowball Edge. If online data migration is necessary, there are services that can move large amounts of data via public and private network connections. Even large transfers of up to 100 PB of data can be managed.

While data are being transferred to the cloud, they need to be carefully indexed and cataloged. This is a critical step in the migration process, because indexing enables the solution for our second problem: data discovery and search. For example, all seismic, wells and production data need to have geo tags to view and search on the map. SEGY data needs to have metadata on how to read and interpret binary blobs inside the file, legal tags will provide data entitlements on who, what, and for how long data can be accessed.

Method

Data lakes are the recommended solution to handle large storage, access, transformation and indexing of data.

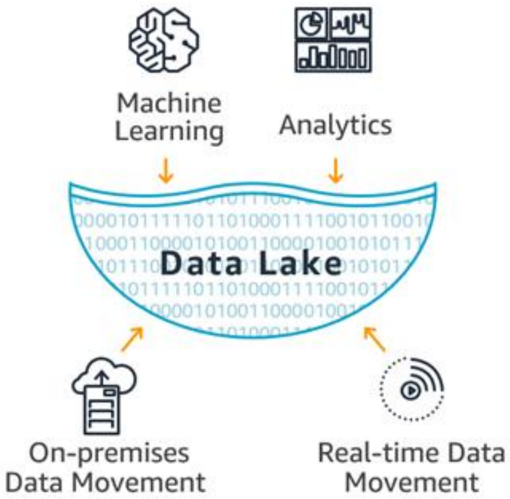

Figure 2: *Data Lake*

There are multiple data lake platforms available that support OSDU, the E&P industry standard data platform. There are also services that make it easy to set up a secure data lake in a matter of days, and DIY data lake solutions for those whose needs are not addressed by the available platforms. Data put on the cloud can have outstanding durability, because available platforms can automatically create and store copies of all objects across multiple systems. For applications that need Windows file systems, there are fully managed, scalable file storage solutions accessible via Windows standard SMB protocol. Likewise, for Linux based systems, customers can use managed file systems that provide access via NFS protocol.

In addition to being accessible, data needs to be easier to find. Explorationists spend up to 40% of their time not just searching for data but searching for the right data. We need improved data delivery so that explorationists have the right data at their fingertips when they need it. There are serverless data integration services that make it easy to discover, prepare, and combine data for analytics, machine learning, and application development. Dynamic search routines provide a fast search experience for applications, websites, and data lake catalogs, allowing users to quickly find relevant data. Adoption of cloud technologies and cloud data lake services make it possible to organize all the data in a project – thereby facilitating intelligent searches and quick access to the exact data needed for analysis and interpretation.

Most oil companies have geophysical teams, geological, drilling teams, etc. Each team uses their own system to perform analysis, and they often have different data formats. We need a central repository with standard data types and metadata. Centralized data, when organized by good metadata and indexing, also makes it easier to integrate disparate data for analytics or machine learning.

Once data is stored, indexed, and cataloged it needs to be made available to the user via a convenient user interface. Data visualization is a key component of many applications and workflows. For this, a visualization platform such as INT IVAAP provides value. IVAAP is a modular platform that supports multiple sources of data such as OSDU, custom data catalogs, and data lakes.

For example, the Open Subsurface Data Universe™ (OSDU) Forum was formed to establish an open subsurface data model and reference architecture with implementations for all major cloud service providers. The OSDU promotes application standards (APIs) to ensure that all applications, developed by various parties, can run on any OSDU data platform. The goal of OSDU is to deliver the same value and services while running on different cloud service providers and in different data centers.

The interpretation of seismic data to determine subsurface structure presents challenges that require months of processing and, in the end, some intuition. Using tools such as Amazon's SageMaker machine learning program with INT's IVAAP upstream data visualization platform, the complexities of analysis are dramatically reduced. Horizon picking can be automated through user-tuned, deep machine-learning methods, to give dependable results while saving significant time. From this rapid and accurate processing, the stratigraphic features such as faults, salt deposits, and productive reservoirs can be determined which will increase efficiency and productivity. The machine learning platform can also be applied to well-log data to determine subsurface attributes such as porosity, lithology, and stress regimes, among others. Lithology and porosity allow for the identification of layers that will allow oil and gas to accumulate. Stress regimes will indicate how the rock will crack under pressure and where fluids are most likely to flow. These combined with other attributes found from well-log data and seismic interpretation can be used in the process of inversion for classifying the best reservoirs for drilling. SageMaker machine learning coupled with INT's IVAAP improves the quality of the inversion process, enabling rapid decision making with dependable results. These results provide efficiencies from seismic and well log interpretation through complex production models which predict future productivity.

Examples

Example 1: Virtual Data Showroom

Standardized architecture and data models can be completed with an application layer that will enable all your data to be accessible from a single place, make data searchable and discoverable, provide tools to integrate domain expert workflows and deliver a collaborative work environment with advanced visualization to quickly QC data and drive better decisions.

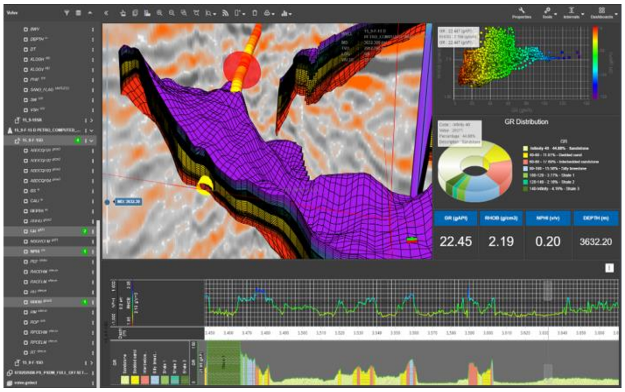

Figure 3: *Collaborative Work Environment with Advanced Visualization*

With the recent developments in algorithms, computation power, and availability of enormous amounts of data, the implementation of machine learning brings data science and analytics into the forefront of our future workflows. The idea of using automated algorithms to determine the rock facies is not new. However, recent advancements in machine learning methods empowers new algorithms in rock facies classification from well log measurements.

Example 2: Characterize facies from well logs

Readily accessible data makes it easy to search for the right data for input to further analysis. Here we can select the well from either a tree view or map view, and confirm that this is the data we want to use

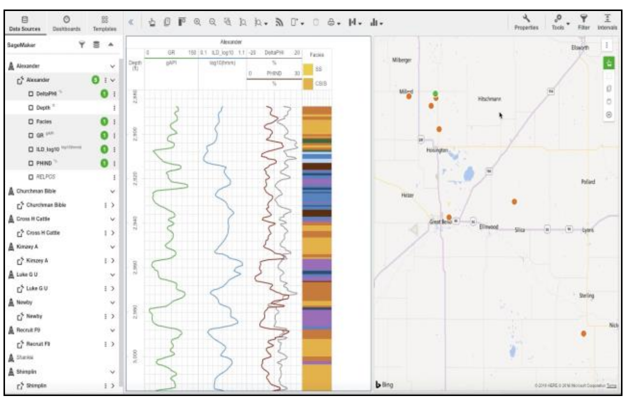

Figure 4: *Selected well, with GR, ILD, DeltaPhi, and PHIND curves displayed for QC*

We can then bring up a menu to specify the propagation model, confirm the well and curves to be used as input, specify output, then execute the job.

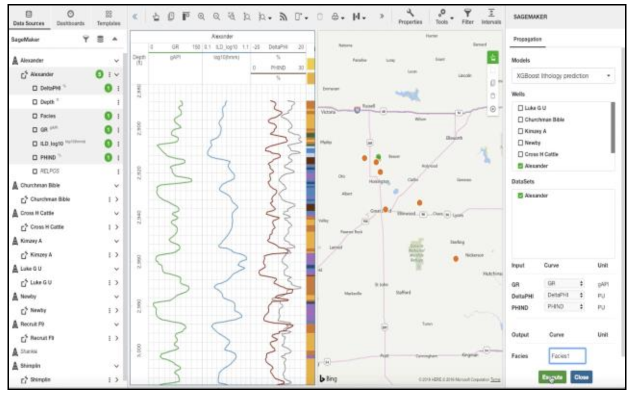

Figure 5: *Tools menu on right side of display to specify model inputs and parameters*

After the job has run to completion, the new facies model can be written back as a new curve and also displayed along with the original data for interpretation.

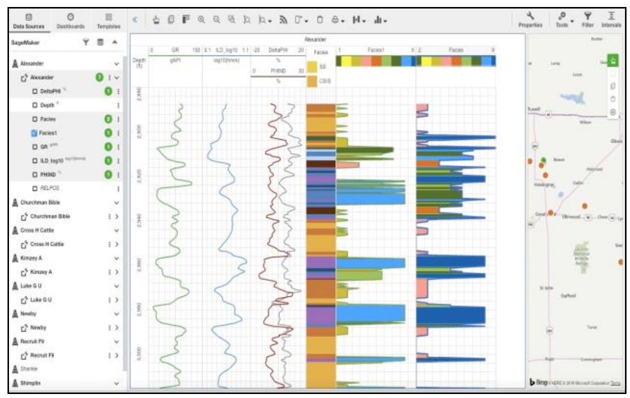

Figure 6: *Well data displayed along with new Facies classification*

Incorporating Jupyter notebooks will allow a non-expert data scientist to execute more complex machine learning and modeling workflows. A Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data transformation, statistical modeling, data visualization, machine learning, and much more.

Example 3: Empowering subsurface interpretation with machine learning integrated into the workflow

Searching for the appropriate data. We can start by bringing up a map view of available wells and selecting those to be used for analysis

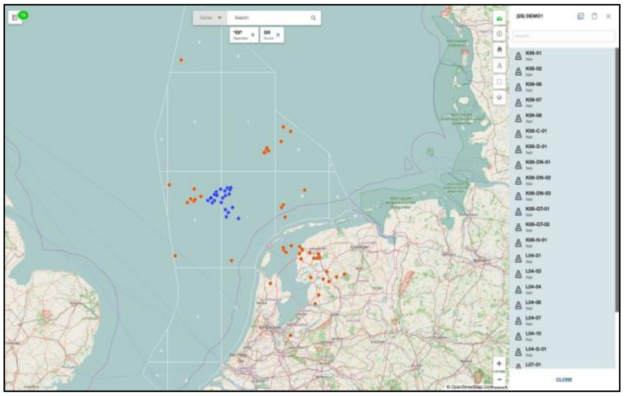

Figure 7: *Select wells for analysis*

Once the wells have been selected, we can then perform some quick views to QC the data with a variety of relevant displays.

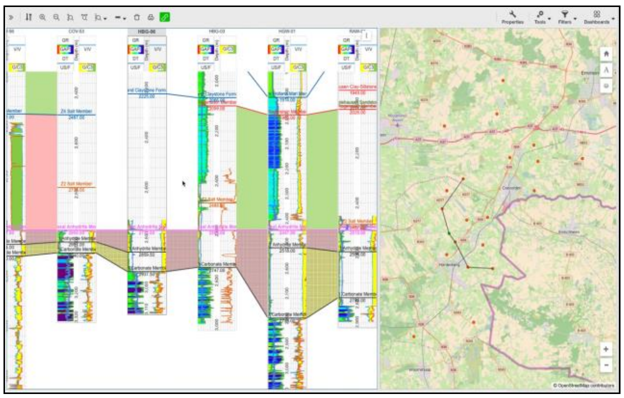

Figure 8: *Visualize wells and correlations*

Augment the data analysis with machine learning integrated into the workflow. The required parameters can be selected graphically or entered in a Jupyter notebook. Since a jupyter notebook can include all the steps of a modeling process, -essentially like a processing wizard - with an appropriate interface it will allow a geoscientist to run data science workflows.

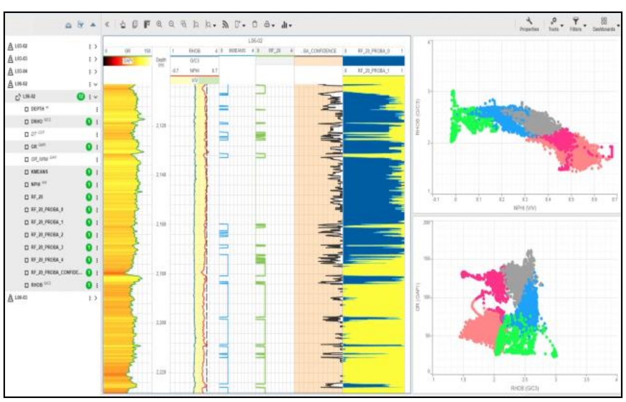

Figure 9: *View ML data and classification models*

Example 4: Pattern Recognition

We can see the benefit in this example of pattern recognition. Large operators can load a huge number of seismic sections from throughout the world, then apply pattern recognition to identify seismic features and attributes that are similar to a section from a known good field.

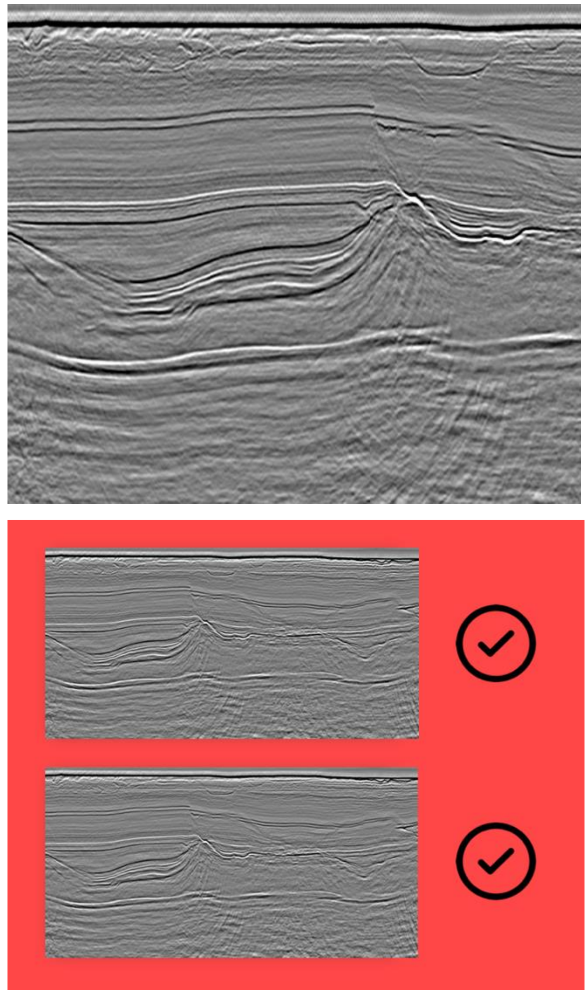

Figure 10: *Pattern recognition to identify similar seismic features and attributes on several volumes.*

The objective of this experiment is to speed up decision-making and lower the cost of exploration by “hi-grading prospects.” Rather than spending hundreds of hours staring at seismic volumes to find prospects that may have potential, interpreters can focus their time and effort on the prospects likely to be the most productive.

Conclusion

Over the past few years, the oil and gas industry has undergone a series of major changes, and exploration and production (E&P) companies are now operating in a radically different environment. Companies must improve operational efficiency, optimize production, and minimize non-productive time to compete successfully in the market. To achieve this, companies must make better use of their data. It needs to be more accessible and available to new workflows and technologies such as AI and machine learning. In many cases, this means migrating the data - properly - to a data lake, and unlocking the data to make it discoverable and searchable.

Our industry still tends to operate in silos, with geology, geophysics, drilling, and engineering data often separated by applications and vendors. One cannot empower collaboration and improve operational efficiency while still operating in silos. It is necessary to break down those silos in order to generate a more collaborative and cross-domain environment.

Acknowledgments

The authors would like to thank Interactive Network Technologies, Inc. and Amazon Web Services for the data and permission to publish this work.