

# Using Big Data Analysis Tools To Understand Bad Hole Sections on the UK Continental Shelf

Joe Johnston, CGG, and Aurelien Guichard, Teradata

Research and analysis show that 90% of all data in the world today was created in the past 2 years alone. Already the growth of data volume outpaces the conventional capacity to analyze and understand, and the trend is only accelerating.

This trend is also occurring in the oil and gas industry with, for example, the continued growth in seismic channel counts, integration of multiphysics information, logging while drilling, and the constant flow of information from “intelligent wells” in “digital oil fields.”

Current data analysis and interpretation approaches follow well-established and rigid workflows, which study the same small set of relationships between entities. It is common, for example, to use core data in well log analysis or well-bore acoustic measurements in surface seismic interpretation. However, core

data is never used alongside seismic data. As the data is growing in volume and variety, it is overwhelming traditional methods. What business opportunities are being missed?

## Big Data

The term “big data” refers to more than simply a large volume of different types of data, both structured and unstructured, with varying degrees of accuracy. It also includes a suite of applications providing solutions and analysis. But big data is really a movement. The big data approach is said to be a data-centric method adept at uncovering otherwise invisible patterns and connections by linking disparate data types. It can search and analyze all data and size with great agility and without regard to user group or project area. Examples from other

fields include a retailer analysis of buying patterns and automotive manufacturers predicting faults and failures.

Large volumes of data are nothing new to the oil and gas industry. The seismic business in particular has been successfully dealing with rapidly increasing volumes for a long time. For example, work by seismic equipment manufacturer Sercel suggests the channel count available for acquiring a seismic survey has been steadily increasing by one order of magnitude every 10 years.

Until now, much of our understanding of reservoirs has come from the study of physics-based models rather than directly from data itself. The challenge we face today is converting ever-increasing data volumes into models in decreasing time frames, ideally in “real time.” The big data approach will instead

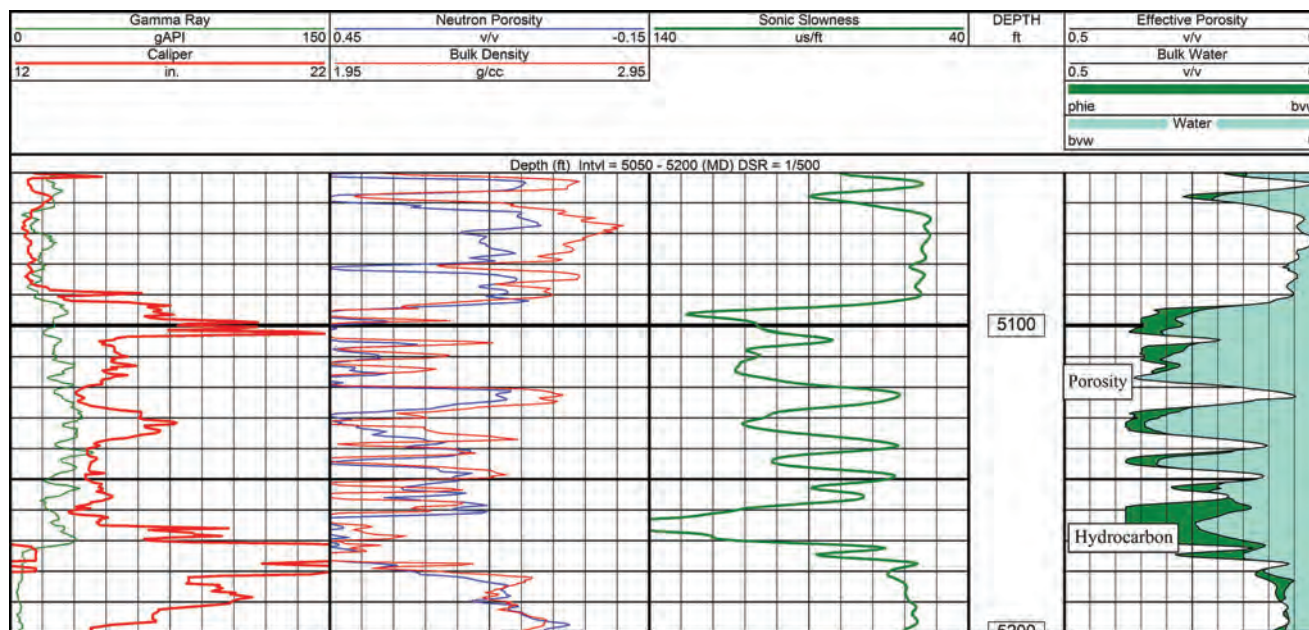


Fig. 1—The caliper curve (red in left track) shows an erratic borehole. The rest of the data is poor. The sonic velocity will be wrong for use in seismic analysis; the hydrocarbons and high porosity shown on the right are completely incorrect.

Copyright 2015, Society of Petroleum Engineers. Reprinted from the Journal of Petroleum Technology with permission.

allow us to construct new types of data-driven models to bypass these traditional bottlenecks. It is also expected to lead to different views of standard models, providing new and valuable insights in the process.

## Case Study

To gain perspective on how a big data approach could be used in the oil and gas industry, we selected a real business case as a test bed. We thought it was important to look for specific answers using data from very different disciplines and sources. Drilling a well is a complex operation with significant risk, both economic and health, safety, and environment-related. Typical modern drilling operations generate vast quantities of data and metadata.

In this study, we chose to seek the trends and correlations that could improve drilling results using predictive relationships to indicate drilling efficiency and high-risk situations. We aimed to find ways to save on drilling time and, hence, cost, along with enhancing safety aspects. We also noted that poor borehole conditions make the subsequent logging operations difficult and often render the acquired data useless (Fig. 1).

Our study made use of data relating to approximately 350 wells from across the UK North Sea, provided by CGG as official UK Continental Shelf data release agents on behalf of the UK Department of Energy & Climate Change. The Teradata Aster platform was used for data loading, quality control, and analysis.

The combined input were drilling parameters, well logs, geological formation information, and well locations/deviations. It took the form of approximately 20,000 files in a range of common formats including LAS, TXT, CSV, XLS, and PDF. The location enabled a geospatial analysis while the stratigraphy allowed an exploration of the vertical dimension.

We needed our data management expertise to address various challenges encountered with the input data. We were careful to apply business rules to data preparation and quality control. We resolved inconsistent well log mnemonics (for example, CAL, CALI, CAL1, and C1

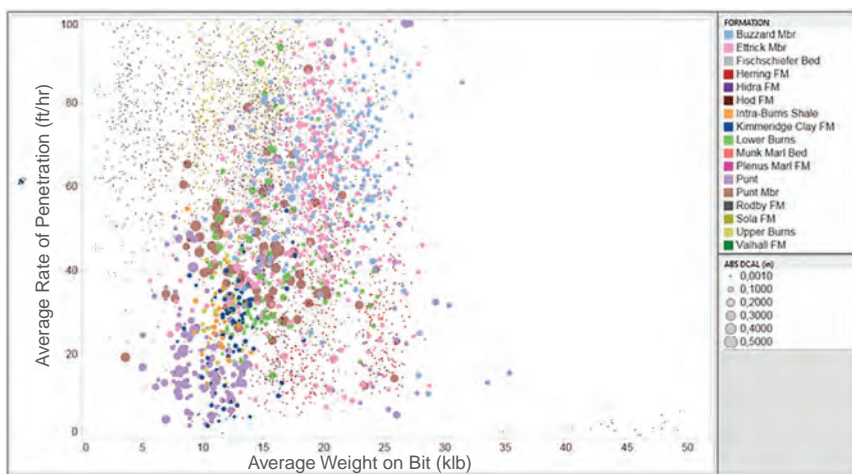


Fig. 2—The visualization of multidimensional data presents a challenge. In the trial, the size of the circle shows the variation from the ideal case and the colors represent the formations encountered. The Punt formation (brown) shows an enlarged borehole.

all referred to the same kind of measurement from different sources). We also encountered and solved issues with mislabeled formation tops, missing or misplaced data columns, and widely differing scales and granularity.

## Results

Once we had thoroughly prepared and loaded all the many types of input data, we began to form connections between drilling observations, measurements, and the subsurface environment. Our objective was to try and link drilling parameters to wellbore condition. In this study, we specifically considered weight on bit (WOB), rate of penetration (ROP), torque, and caliper, with the latter parameter normalized using the bit size to define the differential caliper, which can be used as an indicator of hole condition and flagging of “bad holes.”

An essential concept of this type of analytics is that there is no a priori data model or correlations and algorithms between the data types. All data, whether sampled every 6 in. as in a well log or at random intervals such as formation tops, is treated in the same way. This allows for rapid associations between disparate data types not normally connected (Fig. 2).

The results were surprising and instructive. It was possible to visualize where changes in drilling parameters affected

the borehole quality by a single well or by formation or by formation and geographical location or any combination.

An example of the type of discovery is illustrated in Figs. 3 and 4. Plots of WOB vs. torque (Fig. 3) and of WOB vs. ROP show an anomalous set of bad data points associated with an increase in WOB. This is found to be a single well. The two plots suggest a problem with the well. The various reports and plots associated with this well were checked.

The answer was found in the log plot (Fig. 4), which shows a dramatic decrease in the caliper. The device was apparently closed because the logging tool was becoming stuck owing to the borehole conditions. A significant increase in the tension on the logging cable was reported, a potentially dangerous situation. An additional wiper trip, representing an extra cost and time, was made to recondition the hole before running casing to eliminate potential problems with that operation. All this was noted in the reports.

In addition, the logs themselves were affected by the sticking tool, with spikes and straight lines. This effect was not noted in any report but would have been discovered in the subsequent interpretations.

The ease and speed with which this was uncovered using big data analysis contrasts with the traditional approach of a

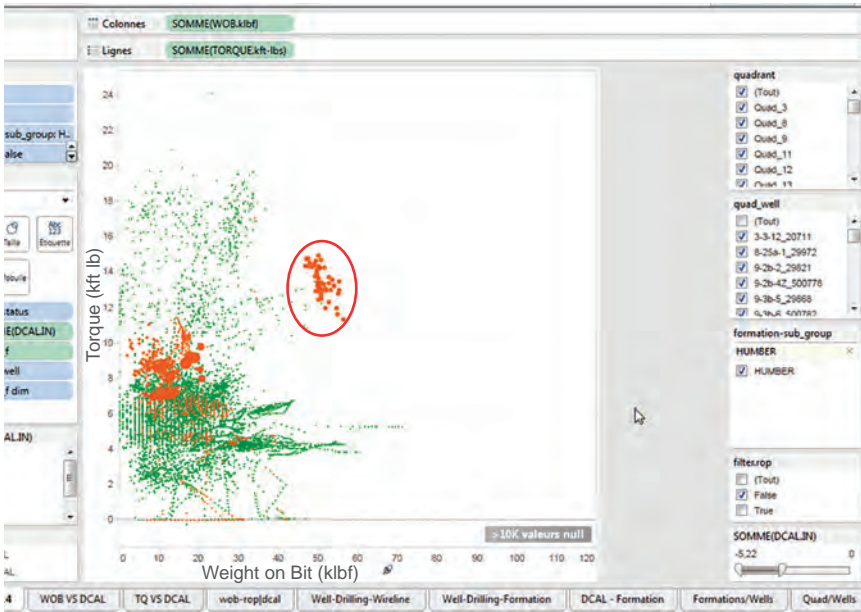


Fig. 3—In a single formation called Humber, the colors show whether the DCAL value is good (green) or bad (orange), based on some modifiable criteria. The circled cluster of points represents a single well. This “exception” uncovered with the big data approach revealed an anomaly.

detailed examination of well reports and other documents. In the case of multiwell phenomena, this task is extremely time consuming and prone to various types of errors.

The example is one of many to come out of this study. Other examples show variations on a regional basis in the same formation. This suggests that drilling parameters giving a perfectly good bore-

hole in one place are not necessarily correct in other places.

### Conclusion

We successfully deployed big data techniques on the diverse data used in this study and used advanced visualization capabilities to display multiple types of data. Multivariate analysis was performed on the data without preconcep-

tions. Unexpected correlations were exhibited and differences were discovered both spatially across areas and vertically through formations. The correlations allowed predictive statistics to be computed.

This can provide the drillers with indications of difficult areas. A set of operational parameters can be calculated enabling them to drill future wells with fewer problems. In addition, innovative quality control techniques applied on disparate data types were developed, which saved specialist time.

Our experience suggests that oil and gas data are highly suited to big data analysis. However, expertise is required to properly prepare and control the quality of the input. The preparation of data using big data techniques allows for a number of quality control steps to be rapidly performed.

It is better to focus any analyses on specific questions because it allows the user to obtain quick answers rather than fuzzy generalizations. It is no longer a question of whether big data has arrived, but how the oil and gas industry will use them and what insights and breakthroughs they will provide. **JPT**

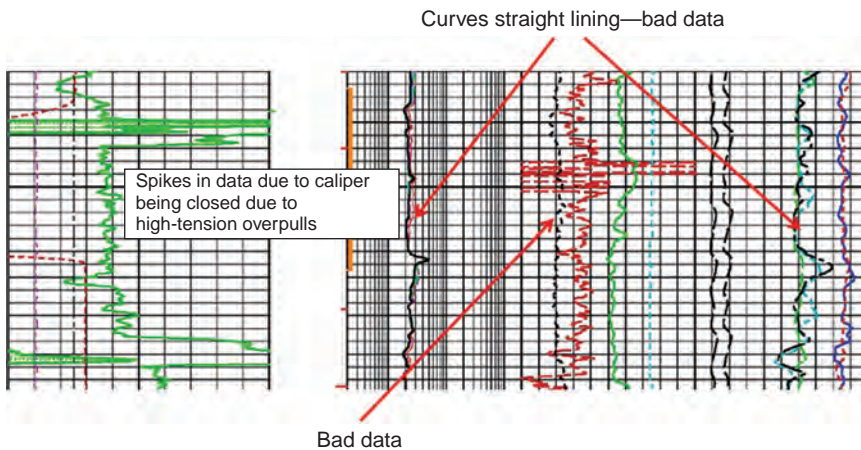


Fig. 4—The log shown is from the anomalous well discovered in Fig. 3. The caliper curve (red in left track) goes sharply to the left, indicating a dramatic reduction in apparent hole size. The logging problems are shown on the other tracks.

**Joe Johnston** is chief petrophysicist with CGG. He has worked in the industry for more than 40 years, 5 of them with CGG, in a number of positions around the world. He is the author of a number of papers on petrophysical topics and holds a BSc degree in physics from Imperial College London.

**Aurelien Guichard** is a senior consultant at Teradata where he focuses on uncovering business value from data and analytics in oil and gas. Previously, he worked for Schlumberger in business consulting, reservoir engineering, well testing, and seismic. He holds a master’s degree in ocean engineering from Texas A&M University, a master’s degree in civil engineering from Ecole Speciale des Travaux Publics in Paris, and an MBA from Tulane University.