Big Data: The Sky Is the Limit for Oil and Gas Companies

A data explosion has taken place in the oil and gas industry over the last fifteen to twenty years with significant acceleration in the past few years. With worldwide data collected in zetabytes now, ‘big data’ is just getting bigger. Encompassing a wide array of data types, big data has the potential to spark big changes in the industry. Mark Allen and Chris Van Dyke offered their thoughts and insights in an interview about big data in oil and gas.

What is “big data” and how is it different from previous types of data?

Mark: In the 1980s and 1990s companies were deploying applications that were backed by relational databases. We had large amounts of transactional data from applications like enterprise resource planning and supply chain, and we had a lot of customer-related data in customer relationship management systems. This data was structured, wasn’t very complex, and it was at the gigabyte level in terms of volume.

Around the turn of the millennium new types of data started to come online, much of it unstructured with high complexity. In addition to the transactional data that was still being generated, we started gathering time series data. Equipment measurements, for example, were being captured over large spans of time and in sub-second increments. We also started to gather data around how customers were using web-based applications, and, on top of all that, audio and video files were being added to the mix. It should be no surprise that databases went up an order of magnitude into terabytes.

Chris: There is also a hardware component to the picture. Today we work on an “internet of things.” Data is no longer being gathered only on computer workstations. Tablets, cell phones, and other smart devices are also part of the process, and all these devices need to be connected. So, bottom line, we have more (and more complex) data being gathered with a range of connected devices. And we have jumped two orders of magnitude to petabytes. This is what is meant as “big data.”

Where is the industry now in relation to big data?

Mark: We are asking, “How can we look at all of this data together and in meaningful ways to gain new and better insights?” It’s a question of how to store the structured/unstructured, simple/complex data and then how to retrieve it to make the most difference. There are new types of data repositories that allow a user to assemble bunches of data and then perform
searches depending on what that user is looking for and what he or she is trying to pull out of it.

**Chris:** As an example in upstream oil and gas, you could pull every piece and type of data available about wells into one big data repository. The data could include documents, work orders, well header data, production data, schematics, and other information. Then you might say, “I want to know everything about onshore wells that have had a decline in production over the last three months.” The result of that search will give you all the data, whether structured, unstructured, or real time, available within those parameters. As we get smarter with our queries, we could then pull that same dataset and compare it to a second dataset of other wells which saw similar declines but then saw production improve. Comparing the two datasets we would then be able to see what corrective action caused the production to improve and apply that to our recent decliners. This allows the industry to start using their big data as an asset that can be mined to gain insights on how production can be improved. By the way, this could be done in the cloud.

**Q** Generally, when we think of really big data stores, we tend to think in terms of upstream operations. Is big data applicable only to upstream, or will it apply across the enterprise?

**A** Chris: Across the enterprise, no question. This is just as important in the downstream space as in the upstream space. For upstream the focus is on production, for downstream the focus is on supply chain and marketing, but the issues that big data can help address are essentially the same.

**Q** What would that look like? How would the cloud be used to access and analyze big data?

**A** Chris: High performance computing was historically limited to onsite capability, with a thousand nodes that turned on and off. High performance computing is now viewed as a cloud-based service or utility. I can load all my data up into the cloud, using a utility. I can mark my report, do my querying, and pull back the results. Then I can spin back down after that. This is part of the big data concept.

The major attraction to big data capability is around reliability. Reliable data translates to shortened cycle times and risk mitigation, which can have really significant positive impacts on the bottom line.

**Mark:** I think that upstream is going to gravitate to big data faster, mainly because dealing with large data stores is already part of that space. Big data will eventually roll out to the downstream organization, but I think it will take longer to get firmly established compared to upstream. The place we are beginning to see rumblings of big data in downstream is in the pricing area. We have a lot of customers who are engaged in putting better pricing systems in place, and they will use big data repositories to help get better pricing.
How do you see this playing out in the industry? How will oil and gas companies move toward use of big data?

Chris: I think that we will see big data repositories in companies throughout the industry by 2016 or 2017. The first step will be improved enterprise searches; that is, the ability to search across all the different data siloes in an organization and get one set of results back. We’ll probably see that sometime this year. The next step is building predictive analytics into the queries, which will probably be another two years after that. The final element will be complex event processing, where data is analyzed on the fly rather than by query.

Q One final question: What is the compelling benefit of big data to oil and gas companies?

A Mark: The major attraction to big data capability is around reliability. Being able to make operational and strategic decisions using robust, accurate and reliable data translates to shortened cycle times and risk mitigation, which can have really significant positive impacts on the bottom line.

Chris: That’s right. Here’s just one example to illustrate the benefit: I monitor a well that contains equipment from which real time data is being captured. In a big data scenario, when I review equipment data over time I see that a decline of production historically corresponds to a higher operating temperature. I also see that, according to work order data, the last time I performed an intervention was seven months ago. Based on the performance of other wells, this kind of data indicates high risk of that well failing. I can then go on to access information about how risks were mitigated (and how the well was kept operating) in similar circumstances elsewhere. This may seem like an isolated example, but it is significant. If big data was used just for this one example in a company, there would be big benefits attached. Expand big data capability across all operations, and the benefit compounds significantly.

A Mark: And this is just the beginning. Once big data has become “business as usual” in oil and gas companies, there are likely to be business benefits we haven’t even thought of yet.