digital energy journal

Cybersecurity - time for a new approach?

Exploration – time to be optimistic in oil and gas

How data managers can learn scripting

AI - keep your objectives simple

Moving subsurface data between cloud software

Why digital isn't going as expected

January 2019

Reports from Digital Energy Journal's forum in Kuala Lumpur

Finding Petroleum

Official publication of Finding Petroleum

Teradata – getting data out of the applications

One of the biggest hurdles for oil and gas companies in their analytics projects is finding ways to release data which is 'locked' in software applications. Teradata's Jane McConnell explained why the problem exists and how to tackle it

One of the biggest hurdles for oil and gas companies in their analytics projects is finding ways to release data which is 'locked' in software applications.

The problem arguably exists because of the industry's preference for "buy" over "build" over recent years, preferring to purchase the software applications available on the market, rather than build their own, said Jane McConnell, practise partner oil and gas with Teradata, speaking at Digital Energy Journal's October forum in Kuala Lumpur.

For example, many companies have subsurface data locked in Petrel, business data locked in SAP, and operations / facilities data locked in engineering applications, she said. They also have a number of proprietary systems for storing data over the long term, including well data archives, borehole data archives, seismic data archives, operational data archives.

Companies bring data from these archives into their subsurface modelling projects, drilling projects and data science projects as needed, involving the development of 1:1 connections.

This all makes it harder for companies to get the benefits of analytics. They could find ways to produce oil faster, cheaper or more efficiently, or improve the success rate of exploration, produce a higher percentage of their reserves, improve safety, use less energy in the process. They might use analytics to see that drilling can be done safely in a part of the world which most drillers would not touch due to concerns they might be drilling into very high pressure areas.

Companies should aim to gradually migrate their data management into one integrated system, which the various apps would draw data from as they need it, to support the work people want to do, she said. In other words the mantra could be "data first, apps second".

Integrated data strategy

This is something companies need to think strategically about. And when it comes to data strategy, a common problem is that companies bring in consultants who advise them to try to 'monetise' their data, doing extensive analytics on it and copying Uber and Netflix, she said.

But oil and gas companies are not like Uber and Netflix, who gain competitive advantage from changing the way that products are sold. Oil companies are not looking to do this, but improve the way that they produce it.

A better data architecture for the oil and gas industry might have continuous flows of data going in, being checked, being analysed, and then being made available for the various software applications people work on, she said.

Where you have software applications, they are just 'consumers' of data, helping to do specific tasks or analytics on it, such as for subsurface interpretation, well planning, production forecasting and simulation. They would interact with the data architecture through an API.

This is a big change from the software applications of today like SAP, which handle all the tasks of data acquisition, organisation, storage, analysis and visualisation.

Ms McConnell suggests changing your structure gradually, first having an integrated company wide data acquisition system, then adding data storage and analysis to this integrated system, and removing these tasks from the software applications.

Receiving data

A critical part of such an architecture is the way that new data is entered into the system.

The oil and gas industry has many sources of data, generated by business transactions, people, interactions and machines (sensors).

A good architecture would have a system for checking and integrating the data, and storing it in a "reference information architecture" in a standard format.

Data should be picked up automatically and automatically 'ingested' through predefined data pipelines built by data engineers, following a data flow defined by a data manager. The pipelines can determine where files need to be parsed or split, what to index, what to load into databases and what quality checks to run.

This replaces methods where data is manually 'imported' into the petrotechnical software with a fixed import procedure, and data must be in the right format, and loaded in the right way, otherwise it doesn't work or errors creep in.

Jane McConnell, practise partner oil and gas with Teradata

Data should also be freed from its historical file formats, such as tape, which are hard to interrogate.

Data should be standardised as much as possible. including standardising master data, reference data, units of measure, geospatial data. You can add business and metadata, and data quality checks, along the way. So it all delivers data which is ready to use in analytics.

It might be useful to compare an oil and gas data architecture to the way water is provided to our homes, with data quality management being equivalent to processes to check water quality, and IT being equivalent to processes to manage the integrity of the plumbing systems, she suggested.

Connecting different domains

This common digital architecture should also have data from all parts of the company, rather than having separate data stores for business, subsurface and operations, as is common today.

If the data stores are separate, it makes it much harder to do a business process which might involve data from two domains. For example if you want to analyse business performance but with data about actual production operations, not just data in the business systems.

The problem is made more difficult by the different working styles and language the different domains use.

In the 'business' departments, most companies use SAP heavily, and it stores a lot of transactional data. This is normally well known by the IT departments, but not necessarily data management people. There are also often data science people looking at SAP data drawing out business intelligence.

In the subsurface meanwhile, data management is often done "library style", receiving data on tape or disk with a requirement to store it safely and make it available when needed. And a lot of the data is stored in software applications from companies like Halliburton and Schlumberger. “I don’t think I’ve ever seen an oil company where subsurface data has been managed by the same people who manage SAP and E-mail,” she said.

The facilities side has two main “chunks” of data, sensor / control system data, and documents such as CAD drawings and project plans. There are no models for linking data across multiple facilities, as you might need for making predictive maintenance models. Automation systems were built for controlling plant – no-one expected them to be used to generate data for predictive maintenance models, she said.

A common problem is understanding data from historians, where you don’t have a good way of identifying the tags (the piece of equipment which has the tag number in the historian data). Data analysts might want to connect performance with for example equipment installation date or its maintenance record. Sometimes the only record of a tag list is hanging on the wall in an office. “I’ve seen that way more than once,” she said.

The different departments run on different timescales, with subsurface data valuable forever, business departments making management reports usually for the past few months.

To illustrate the differences, consider the way the word ‘model’ is used by different departments. Business IT people might expect to see a financial model, subsurface people would expect to see a subsurface model, facilities people might expect to see a CAD drawing, data scientists expect to see a regression model. So there can be communications difficulties.

The word ‘asset’ also has different definitions. To a business department it means money, to subsurface people it means an oilfield, to facilities people it means machinery or an offshore platform.

Data management organisation

If companies are going to manage data themselves rather than just manage data within software applications, then they will need competent data management staff and governance systems, which can work in all company departments.

The oil and gas industry should see data management as a core skillset it needs, in the same way as it sees IT architecture as a core skill, even though some data managers specialise, such as in subsurface geotechnical data. There should be a proper career path in it at oil companies.

Roles can include data archiving, security, ownership, metadata management, managing reference and master data (so the same well name is used in all computer systems).

The enterprise data management department should be responsible for setting rules for data quality and making sure they are implemented, leading to gradual improvement in data quality as it is measured.

They should be managing core tasks like standardising data models and metadata management. You could have a chief data engineer for specific areas, such as subsurface, facilities and business data.

Many companies manage their digital transformation by establishing a “digital office” established separate to the rest of the company. The digital office might manage a ‘data lake’ but not link directly to anyone’s day to day work.

But for this “digital office” to be sustainable, it needs to gradually become an integral part of the company, she said. For the same reason, outsourcing it is probably not a good idea.

Data ownership needs to be carefully thought through. In most oil companies, the subsurface data is owned by the exploration departments, because they had it first. But they are not the people who will need to use it over the lifecycle of the oilfield.

Data stewardship is always going to be important, checking that data is about what it is supposed to be, in the right standards, and properly managed. “Someone who cares about the data.”

Why oil and gas is different

One question which often comes up in data management projects is how different the oil and gas is to other industries – or whether an approach which worked well in other industries should work here, too.

Ms McConnell has seen some of Teradata’s work doing similar tasks for other industries such as retail and e-commerce, and finds the oil industry isn’t as far behind as commonly thought.

“Sometimes they are getting stuck on stuff that’s pretty simple compared to what we deal with,” she said.

But many other industries moved away from storing data in software applications some years ago. “We’ve stayed trusting applications to do the work for us for quite a while. We’ve got quite a little bit of catch-up.”

The oil industry also has very complex technical terms to describe its data, something which is seen much less in banking data for example.

The types of data in oil and gas is always increasing, for example from new sensors being installed.

Another problem unique to oil and gas is the importance of measurement data. In banking, the only unit of measure is the currency. The oil and gas industry also has to deal with masses of data in old formats.

The oil and gas industry doesn’t need a large number of different technical solutions to improve, but it does need to do “a lot of work” simplifying and integrating the software structure it currently has, she said.

Delegates at our Kuala Lumpur forum in October 2018

How data managers can do scripting

Alvin Alexander, geo technician from JX Nippon, taught himself to do scripting with Python to automate common data management tasks. He explained how he does it at our KL conference

Alvin Alexander, geo technician from JX Nippon, taught himself to do scripting with Python to automate common data management tasks, and believes other data managers can do the same.

He shared his experiences and advice with a talk at the Digital Energy Journal forum in Kuala Lumpur in October.

It can be a lot more fun working by making scripts, rather than doing lots of manual work with a keyboard, such as making and copying folders, or copying data from one spreadsheet to another.

Alvin Alexander, geo technician from JX Nippon

Mr Alexander also has a bad wrist pain which emerges when he uses the mouse too much. So he writes scripts to automate as much as possible.

Mr Alexander has only been programming for 7 years, learning mainly from "a lot of online courses". But he says the fastest way to learn is to encounter real problems and have to solve them.

And when you automate tasks, there are usually a lot less mistakes than from in manual work, such as from pressing the wrong key.

But the most important motivator for scripting can be that it is more fun finding a better way to do something.

"If you're not having fun you're not doing something right for you," he said. But when you have fun, it increases your motivation, you are more productive, so it is better for the company too.

A starting point is to recognise that there is a very big difference between scripting and programming, Mr Alexander said. Programming is complicated work of building software. Scripts are something designed for a specific problem, such as "generate folders from a list", or doing machine learning on log data. Data managers should focus on scripting, not programming.

When considering building an automated solution, a first question is how often you spend doing a task. If it is a five second task, it probably doesn't need to be automated. "If it takes three days, it probably needs to be automated." If you have done something twice, maybe you will need to do it again, and it makes sense to automate it.

And if you have good skills in a programming language, particularly Python, maybe you will want to automate everything. Or there might be scripts available free online you can just copy.

Examples of tasks

Mr Alexander gave some examples of common oil and gas data management tasks which could be done faster with scripts.

A "plain boring" task data managers might be asked to do is generate 200 folders from a list of folder names you want created. "It takes forever", he said. "Copy one item, create a folder, rename, paste. In Python you can do it in 5 lines of code."

A second example is to consolidate 5 Excel spreadsheets into one. Instead of doing this, you can convert the spreadsheet into code. "It doesn't matter whether you have 5 or 5000 excel sheets, it is the same amount of code," he said.

A third example is extracting data from a well log (LAS) file. For example, you want vertical permeability data at a certain depth. Usually this is done by opening the LAS file with a text editor, and copying the columns into Excel. You can buy software to do this, but it is probably not free.

Mr Alexander wrote his own "parser", a software tool to break up well log data into useful elements.

You can make an automated solution with free code which will look for key words in the text. So it can copy pressure data only for a certain range of depths. A more involved question is to show how the permeability is varying by depth.

A fourth example is drawing curves. You can get Python code to draw perfect Sine waves, squares, perfect angle spiral, much better than drawing it by hand

In one project, Mr Alexander needed to trace a typical gamma ray curve. He tried doing it by hand using a mouse, or by looking for software which could do it digitally, but neither worked well. So he wrote a script to trace the curve image.

A fifth example was a tool which can create a list of all of the data files on a CD, no matter what the folder structure is. This is very useful to a data manager who has been handed a CD with unknown contents and wants to understand what is on there.

Using Python

"Python, among all programming language that I learned, is the most friendly, almost like talking to you, almost human language," he said. "Not like Java or JavaScript."

It can be amazing how little code you need to write with Python, because most of the code you might need is freely available. "The community is so nice, big, friendly, they provide almost everything for us. We almost never need to write any code from scratch."

Mr Alexander recommends the Anaconda Distribution software for installing Python, with all its modules, and a store of tools to run on it, "like Google Play store for Android".

You don't need to use an IDE (Integrated Development Environment), as you do in computer programming.

There is an open source web application called Jupyter Notebook which can be used to run specific lines of code. "It's very easy to use," he said. "Since I learned this, I learned Python a million times faster. IDE is so confusing. There's so many unnecessary things. In this is it is only code. A lot of the scientific community really favour this."

Original work

Mr Alexander concluded with the Gustave Flaubert quote, "Be regular and orderly in your life, so that you may be violent and original in your work." - this sums up the approach to data management, if you are orderly in how you go about managing data, it is possible to do much more with it.

In oil and gas, the people doing the "violent and original work" are the people who work with the data, such as the geologists, geophysicists, petrophysicist, reservoir engineers, other engineers – and many data managers are themselves also geologists so users of the data.

"Simple objectives best" with ML

Machine learning projects can do better if you keep your objectives simple, said Manoj Goel at our KL forum – with a case study of automatically extracting valve data from a P&ID

It makes sense to keep your objectives simple when embarking on machine learning projects, taking one step at a time, said Manoj Goel, director of Reliable Business Technologies, a software systems Integrator in Kuala Lumpur, speaking at the Digital Energy Journal KL forum in October.

"We have to keep simple to succeed," he said. "You have to take one baby step at a time and solve results. Nothing can be solved in one day. For a marketing perspective we like to sound complicated because we do not want others to do what we are doing."

Computers are only capable of a fraction of what human intelligence can do, he said, but computers can do some of the drudge work which people do, perhaps the work which does not draw deeply on human intelligence but does take up a lot of time.

Mr Goel presented a case study of how his company developed a machine learning application to read valve data from piping and instrumentation (P&ID) diagrams, drawn on paper and scanned, on behalf of a major oil and gas engineering contractor.

These drawings can be large and complicated at times. The client was paying consultants to extract the data from the diagrams, a task which was proving expensive, slow, repetitive and error prone.

The main computer challenge was understanding the different symbols used on the charts, describing different types of valves and lines, and which can be drawn at different scales and orientations. Sometimes the labels are handwritten. Some of the scans of the paper diagrams were not very high resolution, made at a time when scanning technology was less advanced.

The project was split into two phases, with the first phase objective just to create a list of valves, including all the metadata, size and rotation in a spreadsheet. It is a computer vision problem – you get an engineering drawing as an input," he said.

The project team decided to use the TensorFlow open source software developed by Google, which can build neural networks for image processing. It took 5000 lines of Python code on TensorFlow altogether. Most of the design work is working out what kind of model you want to build.

Manoj Goel, director of Reliable Business Technologies

An alternative is Keras, which is also open source, written in Python, and can run on TensorFlow and other systems. It has a better user interface, it is easier to make models," he said.

There are proprietary tools available, such as IBM Watson, but these are not necessarily the best tools for starting off with, he said.

The project team found that a 3 layer neural network model could get results 100 per cent correct, with P&IDs scanned at 200 dots per inch (dpi).

Where the computer was not sure what the symbol was, it could guess and assess the probability that it was correct.