SPE-201725-MS

Digital Transformation of Well Completion Selection and Design Through Data Insights

Varma Gottumukkala, Matthew Worthington, and Debasmita Basak, Schlumberger; Renuka Radhakrishnan, Jay Srinivasan, and Smitha Ramesh, Infiswift Technologies

This paper was prepared for presentation at the SPE Annual Technical Conference & Exhibition originally scheduled to be held in Denver, Colorado, USA, 5 – 7 October 2020. Due to COVID-19 the physical event was postponed until 26 – 29 October 2020 and was changed to a virtual event. The official proceedings were published online on 21 October 2020.

Abstract

Well completion selection and design involve the selection of optimum well completion architecture and associated downhole equipment to deliver hydrocarbon to surface in a safe and efficient manner. A number of well architectures can be conceived for a given application, and a plethora of equipment is available across the industry, with hardware to meet a wide range of operating conditions including hole size, pressure, temperature, flow rate and fluid type. This wealth of choice results in a highly complex and challenging selection process that today is done manually, relying on subject matter experts and local best practices through trial and error approach. As a result, the process can be quite inefficient, designs can be suboptimal and fail to consider unique reservoir and well conditions leading to premature equipment failure causing loss of production and well integrity. These failures can have impact ranging from unplanned well intervention, equipment pull outs, fishing operations, extended rig time, workovers, or even complete well loss—costing the oil and gas industry billions of dollars. The shortcomings in design are therefore ripe for innovative digital solutions.

This paper describes how manual completion selection process can be seamlessly transformed into an intuitive digital solution providing insights for the well completion selection process. The proposed digital solution describes software tools and architecture used to consolidate thousands of historic, unstructured, completion schematic data into a structured database. It automatically maps the completion architecture and equipment details to relevant operating environments, captures nonproductive time and highlights installation challenges. The solution also identifies correlations and data trends across various types of well designs and equipment categories, using advanced artificial intelligence and machine learning algorithms to provide insights into equipment reliability, operational efficiency, total cost of ownership and production performance. A minimum viable product consisting of 24,000 wells from across the world has been successfully developed to demonstrate key value propositions.

New data coming in from recently completed wells can be seamlessly integrated with the existing data bases and the algorithms constantly improvise its learning process to provide better accuracy. The digital solution proposed for well completion selection and design process ultimately enables oil and gas companies

to optimize well completion configurations and equipment that can deliver maximum value. It allows them to identify offset well issues, derisk operational concerns, check compatibility of equipment with respect to dimensional constraints, pressure and load rating requirements, thread configurations, metallurgy constraints, seal elements, complete well on digital file with tracing and accountability.

Introduction

Well completions play a key role in providing a fluid conduit for the reservoir fluids to be produced to surface while maintaining the integrity and safety of well operations over the life of the well. On average 50,000 wells get drilled every year and an estimated 2 million wells are currently active across the world. Every single one of these wells requires completion equipment, and each equipment is selected to address a key purpose. These selections are captured in millions of well configuration files, each of which contain information in unstructured formats across a wide range of formats. A total of 30% to 40% of the well construction budget goes into completion equipment for both surface and downhole hardware. It is estimated that $7 to 8 billion dollars are spent every year to procure completion hardware. This hardware includes the following equipment (Fig. 1):

- christmas trees
- casing, tubing, tubing hangers
- valves hardware including safety valves, isolation valves and flow control valves
- liner hangers, packers
- screens, inflow control devices (ICDs) and autonomous ICDs
- sliding sleeves
- grave pack
- artificial lift including ESP, PCP, SRP, gas lift and ESPCP
- surveillance hardware including downhole gauges and fiber optics cables
- other completion accessories such as nipples, plugs, entry guides, shoes, flow couplings

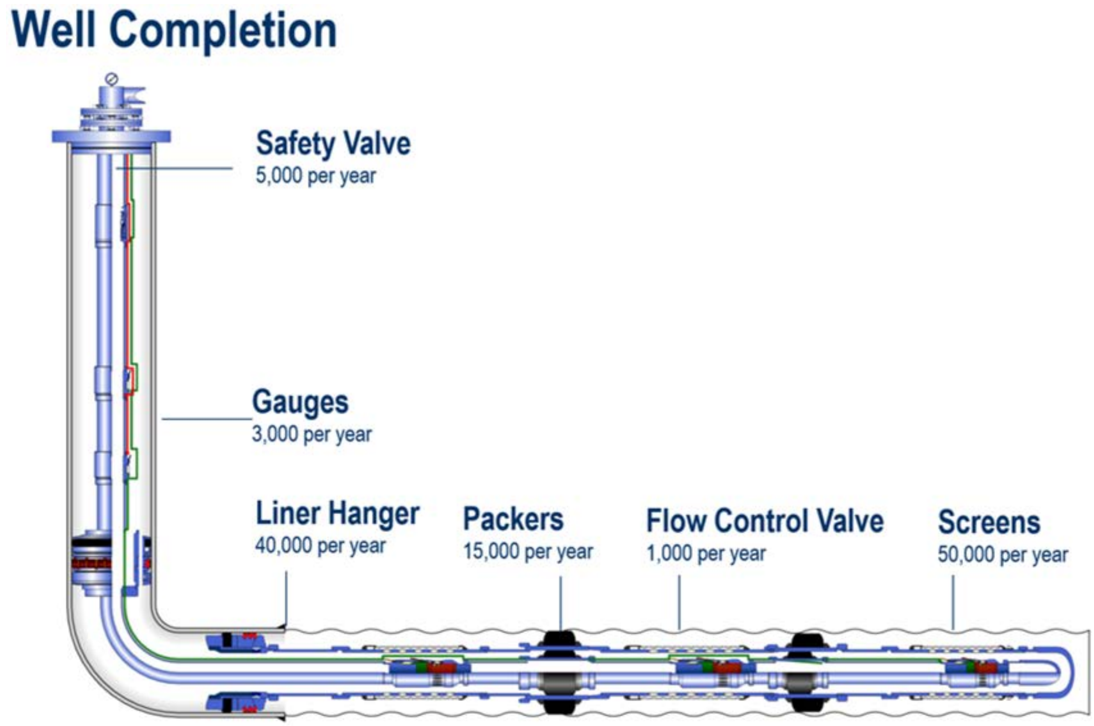

Figure 1—Well completion describing key equipment

Well completions concept selection is evaluated at inception of field development planning through reservoir simulations utilizing exploration and appraisal data. Various scenarios of well types (vertical, horizontal, multilateral) are evaluated for maximizing reservoir recovery, well life and NPV. Sinha, Yan, and Jalali have done some extensive work on manual completion architecture selection process. Field experience across similar environments is also considered via best practices. Unfortunately, most of the simulations do not model detailed completion equipment, their reliability undermining the design's impact on flow behavior, operations and performance. This leads to suboptimal completion equipment recommendations and poorer well designs. From the time equipment configurations and selections are made to the point of installation, the cycle could be multiyear depending on the complexity of the completion design. Any poor choices made cannot be reverted and the well will be delivered as planned with lower than expected performance.

The lack of proper selection tools, which can incorporate parameters of well, reservoir, production, drilling, procurement deadlines, reliability and operations throughout the entire well completion life cycle performance is key to maximizing the value of an oilfield asset. The benefits of lessons learned captured in historic data across various assets is frequently disregarded or is siloed to specific work environments or within individual companies. Collaboration between teams of different disciplines is inefficient due to the need for frequent exchange of noncontextual data in unstructured data formats, resulting in miscommunications, missed opportunities and poor user experiences.

The completion design decision tree Fig. 2[12] guides users through the traditional well completion selection journey where users are forced to gather data from various sources and domains to achieve their goal, using a stage gate process. This traditional approach can typically take months for a team to finalize a single standard design, and the nonsystematic approach across multiple teams often results in mistakes and design oversights. These traditional approaches are not scalable, inefficient, fail to always lead to optimum designs and do not result in a standard approach to design.

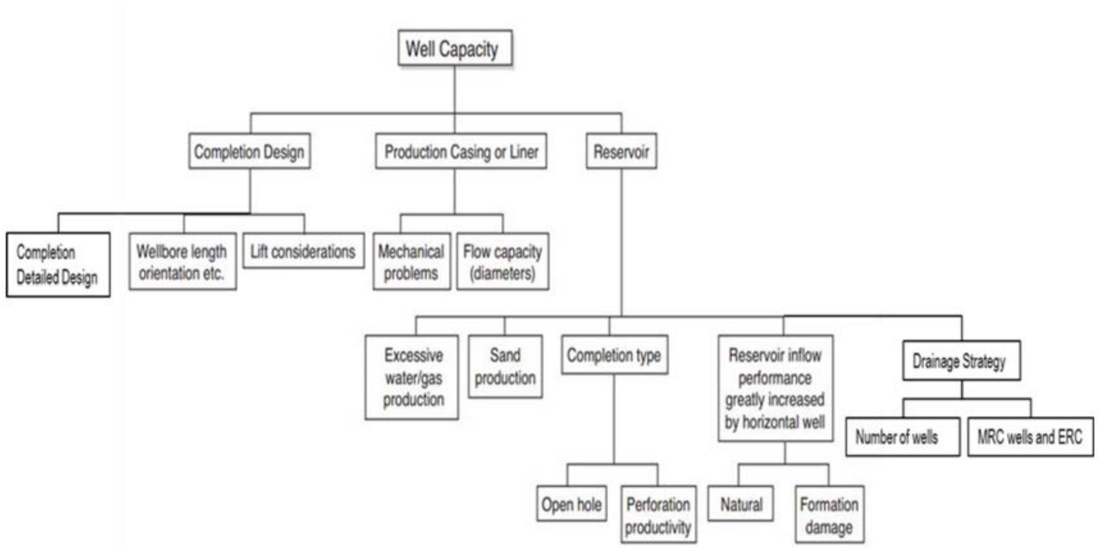

Figure 2—Completion design decision tree.

Traditionally, well completion planning is a multidisciplinary team exercise with involvement of various domains (Completion, Drilling, Production, Reservoir, Procurement) from operating companies and completion equipment providers.

The goals and objectives differ across participating parties, from recovery optimization to safety, rig time and reliability. The challenge of reconciling all factors and developing a comprehensive well completion strategy remains complex. Traditional approaches have led to a planning and design process that is highly subjective and error prone. Knowledge and best practices are siloed within geographies and between domains. Experts struggle to reconcile the diverse objectives that must be considered when attempting to optimize both the well plan and design.

The result is a suboptimal plan and design that does not take into consideration all aspects of the challenge at hand. Reservoir and well conditions are not fully considered in the process and cause unnecessary equipment failures. Lessons are not learned from previous installations with avoidable errors repeated in both neighboring wells and similar basins. Operators and service providers are left struggling through unplanned operations to address mistakes that could have been avoided. And, the process is unfortunately not evolving to prevent repeats. Fig. 3 walks through an example of the high-level complex process of gathering inputs from various domains for a successful well completion design.

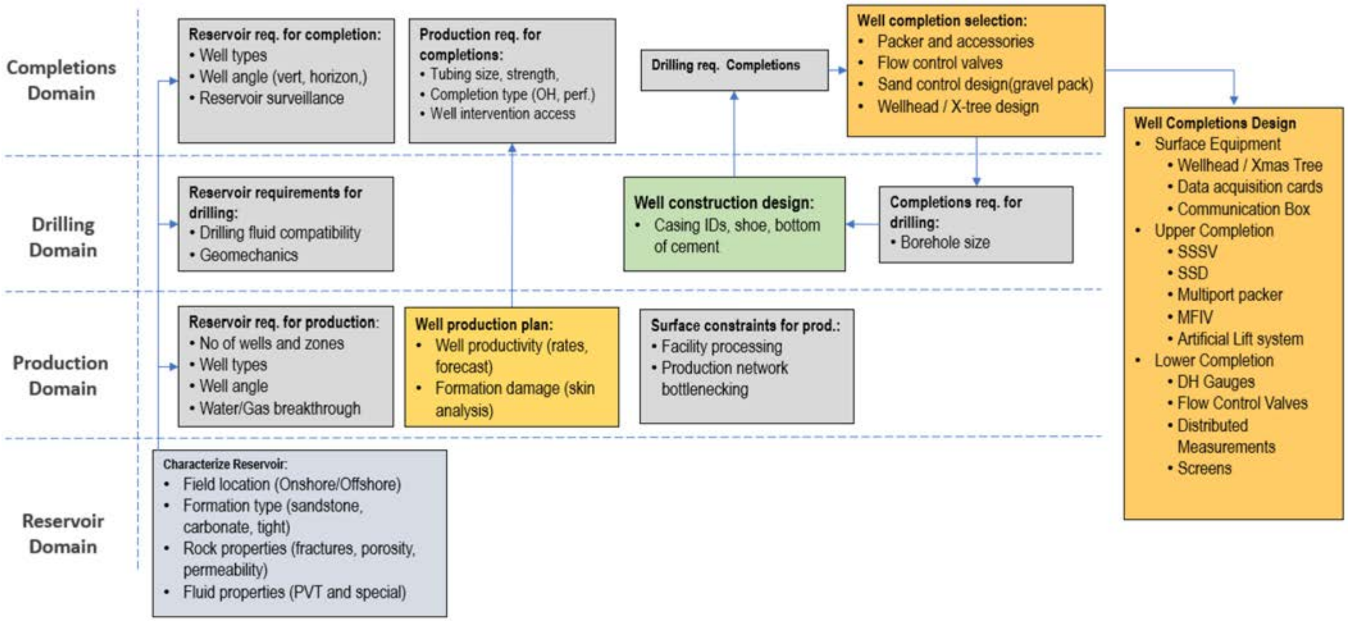

Figure 3—High level view of the cross-domain inputs required for design

A new approach is required. A forward-looking approach that leverages the tools and resources have become available in the digital age. The science required to validate designs from a purely technical standpoint exist today, even if they are not being used in a systematic fashion. What is lacking is an ability to understand what has been done historically, including what has worked and what has failed, and the analytics to assess how this impacts future decisions. The industry is built on a wealth of data that has been collected for decades, and which is rarely used in the completion planning and design phases. This must change to deliver new platforms that are able to analyze track records and determine optimal future decisions. Szemat-Vielma, W., Murray, T., Kiaie, A., Bolchover, P., and Yao, J.[3] have demonstrated a digital solution for well construction planning with a focus on drilling.

This paper proposes an innovative data-driven approach to well completion planning and design. Harnessing advances in machine learning and natural language processing (NLP), historic data can be transformed into a wealth of best practices, lessons learned and fit for basin designs that have a demonstrated track record. Taking into consideration the diverse criteria that drive both planning and design, this approach can provide unbiased recommendations into how well completions should be designed and completed to produce optimal solutions.

Solution Architecture

The digital revolution of the last three decades has transformed how traditional problems have been tackled across nearly all industries. The oil and gas industry is no exception. Each player within the space must either go through the digital transformation or risk becoming obsolete. This rapidly changing landscape presents numerous opportunities for innovation and novel ideas. Traditional problems can be tackled from different and previously impossible angles, allowing for step changes in how solutions are delivered. Truly optimal solutions can now be achieved through the use of novel approaches in machine learning and artificial intelligence. Leveraging these digital tools is key to delivering a solution to the problem of optimal well completion design.

A modern solution depicted in Fig. 4 would need to be developed on cloud technology using the latest tools available for understanding and analyzing large data sets. The solution would be broken into three distinct areas:

1. Data ingestion.
    a. Unstructured schematic files: the numerous sources of data from which the analysis will be performed
    b. Schematic ingestion service: pipeline to ingest all unstructured data from numerous sources
    c. Monitoring service: monitor data sets for changes and continually ingest and propagate changes throughout the system
2. Data analytics
    a. Schematic analytics service: analysis and cluster the structured data to identify and extract correlations within and across well architectures
    b. Schematic search service: search and explore the expansive data
3. Visualization and insights
    a. Recommendation interface: deliver recommendations and insights to the end user
    b. Schematics analytics dashboard: visualization and understand the well architectures
    c. Schematic search: search and explore the graphical well representations

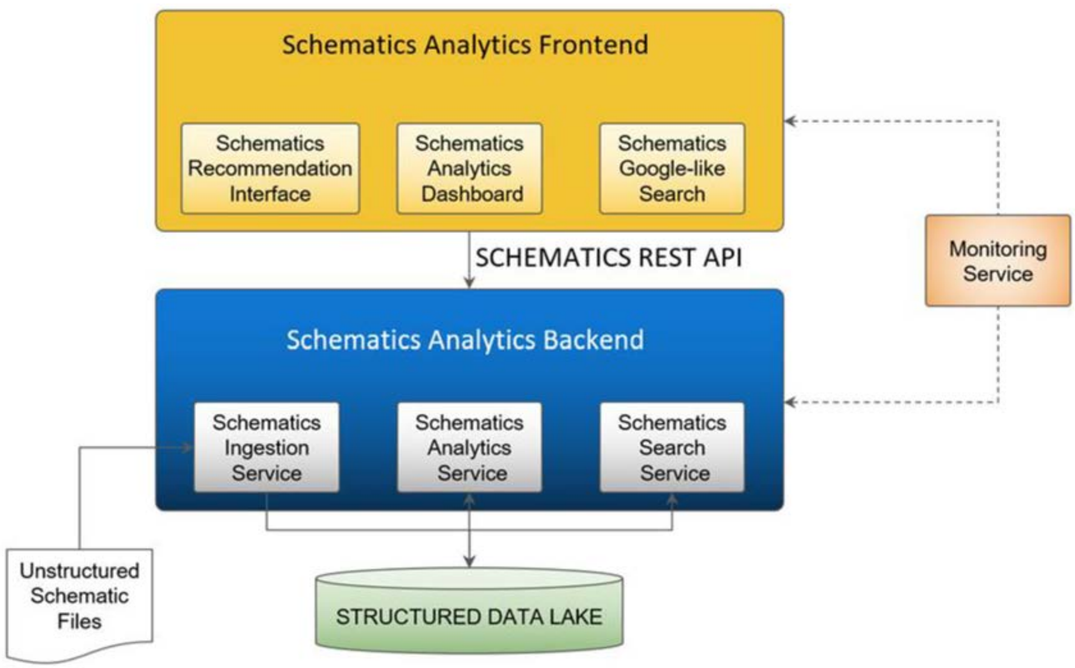

Figure 4—High-level solution architecture

Data Ingestion

Ingestion Strategies

The key to any digital solution is the ability to intelligently use the wealth of historic data that has been gathered over decades of operations within the oil and gas space. For well completions, this comes in the format of completion schematics and their corresponding well contexts. Unfortunately, despite years of use and the creation of millions of documents in this field, there has been no agreed industry standard for the representation of the data. The WITSML* standard is the closest to an agreed industry standard for the representation of all downhole equipment, but this has unfortunately not been widely adopted.

The result of this failure to standardize on data formats has led to thousands of disjointed databases containing huge volumes of unstructured data describing the global well completions. Thus far, the data has been gathered for the purpose of record keeping rather than usage in forward-looking designs. A particular well completion design could be looked up, analyzed and used as input for a future design in a highly manual and subjective manner. Knowledge of where and how to search these databases to discover best practices and past failures depended on the end users understanding of historic wells.

The first challenge is therefore clear—liberation of the source data into a structured and reuseable format. The key to achieving this goal is the usage of the advances in optical character recognition (OCR) packages, NLP tools and machine learning (ML). Combining all three techniques has allowed for the creation of ingestion pipelines that process a wealth of well completion formats and output structured data that clearly captures the design. This information includes the equipment run in the well, from casing and liners to individual jewelry throughout both the upper and lower completion. The volume of data required the solution to be developed on cloud technology with scalable resources and the ability to connect liberated data to a multitude of platforms and solutions.

In addition to the well schematic data, there was also a need to ingest historic service quality reports that highlighted any incidents that might have occurred during the installation and operation of the well completion. These incidents could highlight issues with hardware, process, design and incompatibilities between equipment and context. Again, this data had historically been used for auditing purposes, but never leveraged towards a solution that would consider all past successes and failures to develop truly optimal well completion designs. Their addition provides key equipment performance and reliability insights that help guide the end user toward inherent best practices and designs fit for the target basin.

Ingestion Pipeline

Data ingestion engine is key for deriving any value from unstructured data. Figure 5 below shows the organization of the service to ultimately convert data from various formats into structure data.

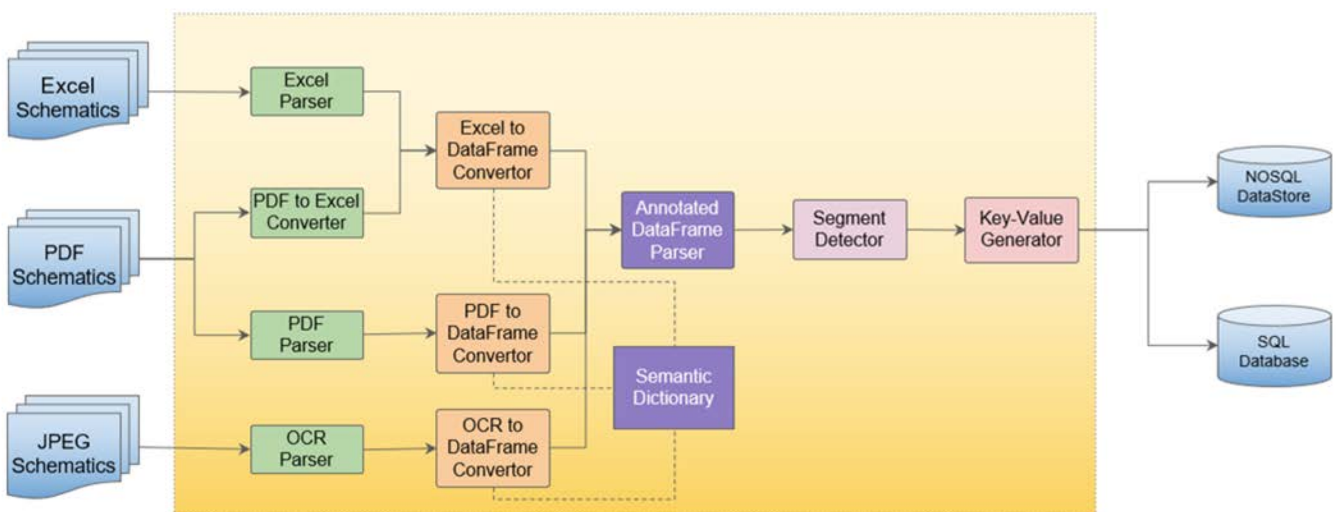

Figure 5—Analytics pipeline

The ingestion pipeline supports ingestion of document formats such as Microsoft Excel* sheets, PDF* documents and JPEG* images of schematics. The various types of documents themselves are exported to a shared folder on the cloud as and when they become available. A cloud process that has subscribed to folder activity gets notified of the presence of the new documents. This process then kicks off the appropriate parser for the particular document type and converts it into a key-value pair and stores it in a data lake. Parsing of data involves the analysis of sentences into their parts and describes their syntactic roles. After processing each document, it archives the processed document in another folder.

Microsoft Excel* Schematics. Extraction of schematic information from Excel documents requires converting the data in rows and columns into key value pairs. Since there is no standard template in the input documents, the algorithms have to infer the presence of different tables and the different ways in which information can be present in those tables. For example, the following, Fig. 6, shows a small variation in formats in the same document where information is presented in different ways and how the robust algorithm was able to handle the parsing of information irrespective of the format type.

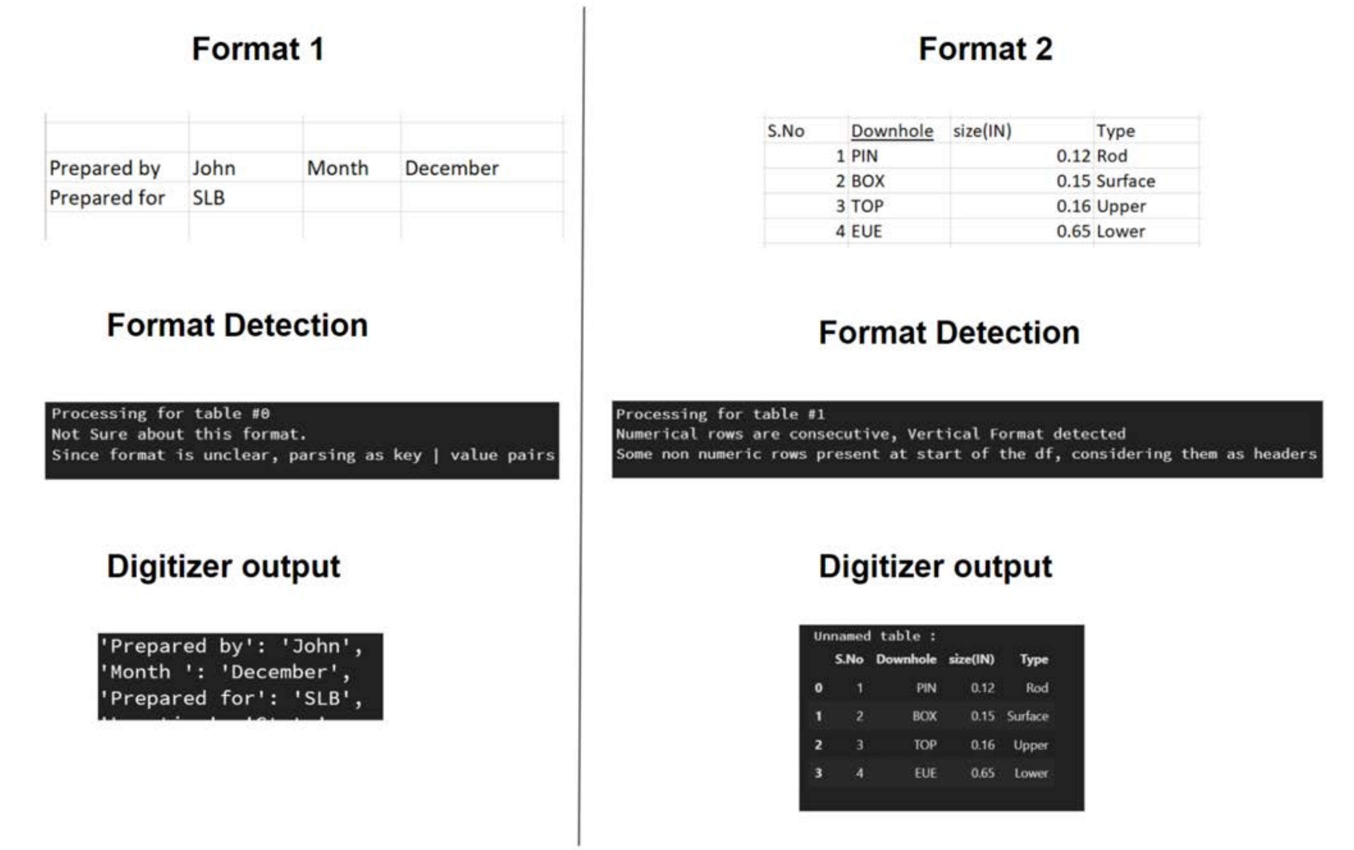

Figure 6—Managing ingestion formats

Contextual disambiguation is also very critical. This is performed by making use of the metadata wherever available and machine learning. Metadata is a set of data that describes and gives information about how other data is organized in context. Fig. 7 below shows how a typical row-major order parsing were parsed incorrectly by showing Location as State and Houston as Texas. By modifying the algorithm to learn the context and by allowing it to recognize the first row as a header column with the distinct background color of the cells, the Location is now correctly stored as Houston and State as Texas. There were several other challenges such as merged cells and section headers in the middle of a series of rows, which had similar issues. The algorithm continuously evolved and improvised to overcome all these common obstacles enabling it to ingest multivariant data with a 99% accuracy.

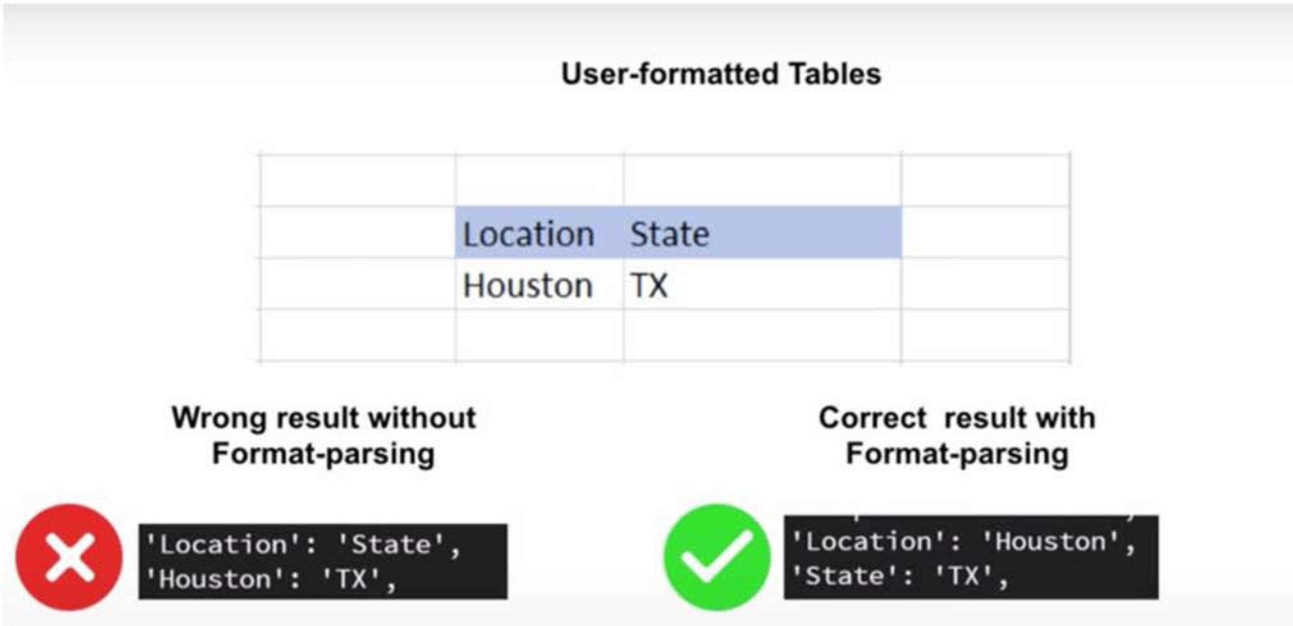

Figure 7—Context disambiguation

PDF* Schematics. Extraction of schematic information from PDF documents requires the ability to infer the cell relationships, which is normally available in Excel. Fig. 8 below shows a PDF schematic where casing and tubing data columns at the header level are merged with detailed information (e.g. outer diameter, inner diameter weight, grade) in multiple columns. A standard PDF parser would consolidate all the columns data below casing and tubing as one set loosing the granularity of the information around each specific attribute. Thanks to the general-purpose format recognition algorithm that was developed for Excel schematics, it was fairly straightforward to adapt the algorithm for PDF without heavily relying on the cell-level information. Fig. 9 below shows how a PDF schematic was successfully parsed with details of individual columns of casing, tubing and equipment data.

Casing Data									
Type	OD (in)	Weight (lbm/ft)	Grade	ID (in)	Drift ID (in)	Connection	Depth (m)	OD (in)	Weight (lbm/ft)
Conductor	28	288.36		26.000	25.812		210.00	7-5/8	29.70
Surface	20	133.00		18.730	18.542		823.00	8-5/8	36.00
Production Csg	13-3/8	72.00	P110	12.347	12.250	VAM® TOP CASING	1220.00	7	29.00
Production Liner	10-3/4	55.50	P110	9.760	9.625	VAM® TOP CASING	1940.00		
Open Hole	9-1/2			9.500			2353.50		

Item	Description	Model Number (P/N)
		Upper 7-5/8 Completion *
K1	7-5/8 29.70 L80 VAM® TOP TUBING	
K2	7-5/8" 29.7# L80 VAM® TOP TUBING PUP JOINT	
K3	7 5/8" 29.7ppf BOX TO 7" 29ppf PIN CROSS-OVER	
K4	7" 29# VAM TOP TUBING PUP JOINT	
K5	7 IN SLIMTECH-5, 6.000 DB, 8.900 OD, 5K	RR100536119

Figure 8—Well schematic PDF data format example

Table_name : Casing Data

	Type	OD(in)	Weight(lbm/ft)	Grade	ID(in)	Drift ID(in)	Connection	Depth(m)
0	Conductor	28	288.36	NaN	26	25.812	NaN	210
1	Surface	20	133	NaN	18.73	18.542	NaN	823
2	Production Csg	13-3/8	72	P110	12.347	12.25	VAM® TOP CASING	1220
3	Production Liner	10-3/4	55.5	P110	9.76	9.625	VAM® TOP CASING	1940
4	Open Hole	9-1/2	NaN	NaN	9.5	NaN	NaN	2353.5

Table_name : tubing data

	OD(in)	Weight(lbm/ft)	Material	Yield Stress(ksi)	ID(in)	Drift ID(in)	Connection
0	7-5/8	29.7	L80	80	6.875	6.75	VAM® TOP CASING
1	8-5/8	36	L80	80	7.825	7.7	VAM® TOP CASING
2	7	29	L80	80	6.184	6.059	VAM® TOP CASING

Table_name : Upper 7-5/8 Completion *

	Item	Description	Model Number (P/N)	Top Connection
0	K1	7-5/8 29.70 L80 VAM® TOP TUBING	NaN	7-5/8 29.70 VAM® TOP CASING BOX
1	K2	7-5/8" 29.7# L80 VAM® TOP TUBING PUP JOINT	NaN	7-5/8 29.70 VAM® TOP CASING BOX
2	K3	7 5/8" 29.7ppf BOX TO 7" 29ppf PIN CROSS-OVER	NaN	7-5/8 29.70 VAM® TOP CASING BOX
3	K4	7" 29# VAM TOP TUBING PUP JOINT	NaN	7 29.00 VAM® TOP CASING BOX
4	K5	7 IN SLIMTECH-5, 6.000 DB, 8.900 OD, 5K	RR100536119	7 29.00 VAM® TOP CASING BOX
5	K6	7" 29# VAM TOP TUBING PUP JOINT	NaN	7 29.00 VAM® TOP CASING BOX

Figure 9—Extracting key value pairs from unstructured data

JPEG* Schematics. Extraction of schematic information from JPEG documents requires employing Optical Character Recognition (OCR) techniques, in addition to the techniques discussed above. OCR is the electronic or mechanical conversion of images of typed, printed data into machine encoded text. To get high accuracy, a specific OCR metadata library relevant to well completions with a machine learning model was used to recognize equipment and patterns.

The biggest challenge using OCR is associating the text with a column header. For example, column headers could be centered while fields could be left justified. Even when they use the same justification, the text could still be of different lengths in each column, and it required use of several heuristics to disambiguate, which column a particular piece of text belongs to. The semantic dictionary, which is built specifically for the schematics knowledge base, also comes in handy to correct the spelling of several words that could be broken due to inaccuracies in the OCR-level parsing.

Free form text. A similar process is followed for the ingestion of completion operational data into the data lake. Using NLP (Natural Language Processing), ambiguous data in text format is converted into structured format and associated with the context. NLP uses identifiable common text, words, language and converts them into computer enabled text for storing data. Fig. 11 below shows operational data and their failure mode's with well, model and equipment part numbers embedded into the text in thousands of rows. To identify the part number or model number has been failing more frequently on a specific well, the users have to go through the entire list of wells, identify part numbers, record them and output into chart, which is extremely cumbersome.

Casing Data									
Type	OD (in)	Weight (lbm/ft)	Grade	ID (in)	Drift ID (in)	Connection	Depth (m)	OD (in)	Weight (lbm/ft)
Conductor	28	288.36		26.000	25.812		210.00	7-5/8	29.70
Surface	20	133.00		18.730	18.542		823.00	8-5/8	36.00
Production Csg	13-3/8	72.00	P110	12.347	12.250	VAM® TOP CASING	1220.00	7	29.00
Production Liner	10-3/4	55.50	P110	9.760	9.625	VAM® TOP CASING	1940.00		
Open Hole	9-1/2			9.500			2353.50		

Item	Description	Model Number (P/N)
	Upper 7-5/8 Completion *	
K1	7-5/8 29.70 L80 VAM® TOP TUBING	
K2	7-5/8" 29.7# L80 VAM® TOP TUBING PUP JOINT	
K3	**7 5/8" 29.7ppf BOX TO 7" 29ppf PIN CROSS-OVER**	
K4	7" 29# VAM TOP TUBING PUP JOINT	
K5	**7 IN SLIMTECH-5, 6.000 DB, 8.900 OD, 5K**	**RR100536119**

Digitizer Output

		CasData					Tubin	Data		
OD	Wdght	Drift m		Depth	OD	Wdght		Yidd Stress	Drift m	
Conductor	28	288.3626.000	25.812	210	7-5/8	29.70 L80		80	6.875	6.750 VAM@ TOP CASING
Surfice	20	133.0018.730	18.542	823	8-5/8	36.00 L80		80	7.825	7.700 VAMO TOP CASING
Production Csg	13-3/872.00 m	12.347	12150 VAM@ 1	1220	7	29.00 L80		80	6.184	6.059 VAM@ TOP CASING
Production Line	10-3/455.50 Pl	9.76	9.625 VAMO T(	1940						
Opal Hole	9-1/2	9.5		2353.5						
Item		Descri	tion			To Connection	Bottom Connection			
7-5/8 29.70 L80 VAMO TOP TUBING						7-5/8 29.70 VAM@ TOP	7-5/8 29.70 VAM@ TOP CASING PIN		7.625	6.875256.01
7-5/8-in 29.7# LSO VAM9 TOP TUBR•-JG PUP JOINT						7-5/8 29.70 VAM@ TOP	7-5/8 29.70 VAM@ TOP CASING PIN		7.625	6.8572.00
S/8-in 29.1ppfBOX TO 7-in 29ppr PIN CROSS-OVER						29.70 VAM@ TOP CASr	129 VAM@ TOP CASING PIN		1.625	
7-in VAM TOP TUBING PUP JOINT						7 29.00 VAM@ TOP CA	7 29.00 VAM@ TOP CASR•-JG		8.9	6.8572.00
IN SLIMTECH-S, 6M•o DB, WOO OD, 5K					RRIOOS36	129.00 VAM@ TOP CAS	129.00 VAM@ TOP CASING		1	6000
7-in VAM TOP TUBING PUP JOINT						7 29.00 VAM@ TOP CA	7 29.00 VAM@ TOP CASR•-JG		7.699	6.8571 .00
7-in VAM TOP TUBING PUP JOINT						7 29.00 VAM@ TOP CA	7 29.00 VAM@ TOP CASR•-JG		7	6.8572.00
T-in Vam Top Box 5/8-in29.-ft Vam Top Pin CROS...						129.00 VAM@ TOP CASING BOX			1.699	
7-5/8-in 29.7# LSO VAM9 TOP TUBR•-JG PUP JOINT									7.625	6.8571 .00
KIO	7-5/8 29.70 L80 VAM@ TOP TUBING					7-5/8 29.70 BOX	7-5/8 29.70		7.625	6.8751607.00
7-5/8 29.70 L80 VAM@ TOP TUBING PUP JOINT						7-5/8 N/A BOX	7-5/8 N/A			
KII									7.625	6.8752.00
K12	138-in 29.1ppfBOX to 7-in 29ppfPIN CROSS-OVER					NIA			1.669	
K13	1, 29.00, SGM-RI-FS ASSY, XPQG/NPQG/NHQG/NMQG,...					129.00 VAM TOP PIN	129.00 VAM TOP PIN			1.75

Figure 10—Subset of the output of the digitizer for a sample JPEG schematic

ShortDesc	FullDesc
LH setting tool mobilization charges- OIL	meeting held on 3rd Jan with client- Liner hanger Service tools along with 5 sets
STC-6 can't pass through MFIV	Liner string with MFIV (P/N 100917506) was set in slips to run washpipe string v
E02y Lesson Learnt Bullnose lost in the hole.	Minimum 8 facts are: 1) Client: BP Azerbaijan 2) Brief description: BHA that w
Not Interchangeable Vam Top Connection	Initially it was designed to connect the 7" LBFV, 17-4PH(105), 4-1/2" 11.6 Vamp
Unsuccessful ESP packer backside pressure test	1. Name of the customer, location: PERENCO GABON 2. Brief description: Up
Wrong Stinger on Location	When running in the inner string for the MH patch the rig was unable to get the
Damaged Threads During Assembly Make-Up	While making up the washdown assembly (TOTAL-07C) for the TOTAL L/T L1 wel
ST Collect unable to Interact w/7in Multi Zone ext	Completions plan for MHN-2 on kalimantan shelf require the installation of MZ-
LH system leak from RCB during circulation at surf	Client PPL Well ARO-KHAN, upon connecting the liner hanger system to the line
Partial LH Running Tool LIH (Updated)	Update as of 13Jan2014 @ 06:20 GMT: Preparation for fishing BHA currently in
4-1/2" JFE Bear Coupling damaged	At mentioned time, found the 4-1/2" JFE-UHP-15CR-125 Pup Joint Coupling was
Upper Shearable Lock Collar too difficult to break	Sometimes when disassebling the Upper Shearable Collet Lock Collar from the L
11-3/4" LWP stuck inside liner joint-PDP stuck	Update - 6-Mar-14 As per intouch ticket (6426331) the only possible cause for F
Revenue and time lost due to setting tool failure	Setting tool failed to stroke. Wireline systems and power charge worked. There
Not reading the all of the operating procedure	While setting a 244.5mm MH patch on a MOST setting tool i did not read all of t

Figure 11—Operational data with key value pair extraction

Using NLP, this data is now moved into structured format where data can be queried and analyzed at the end user's fingertips. The process involves tokenization, developing a semantic dictionary of well names, part numbers and then application of a fuzzy parser for disambiguating and recognizing multiple variations of representations referring to the same piece of information.

Data Analytics

The creation of structured data across both well architecture and context provides for the foundation of the analytics work. With structure, it is possible to explore complex correlations across the vast sea of historic designs. The best practices, lessons learned, and domain knowledge intrinsically captured in each design can be liberated to develop generalized insights into optimal design. To extract these complex correlations, advanced data science techniques and machine learning need to be coupled with experienced subject matter expert input. Combining the two allows for both the development of intelligent engines that can consider all possibilities within the vast data sets and draw conclusions that are beyond the means of human designers. The approach eliminates the highly subjective and siloed problems that face traditional methods, and allow for us to truly develop optimal designs that consider all past successes and failures.

Clustering. Other novel techniques were employed to extract value and relationships from the data. Clustering allowed for the discovery of larger and hidden relationships which may not be known or expected in advance. The clustering is driven both through domain rules and via hierarchical tree algorithms. Extensive domain rules provided a foundation to help guide the clustering in the initial stages. These rules captured simple priorities that inherently exist in the selection of equipment, for example, that a gas lift mandrel is of far more importance than a pup joint. These rules provided a lauching point for the data mining algorithms to then create increasingly complex clusters that allowed for the identification of relationships that were not apparent to the completion engine. These clusters eventually described discrete architectures that had been employed to address a myriad of specific well and reservoir conditions while following best practices and design procedures.

Heuristics Engine. Once key value pairs are available from the analytics engine, there are a number of heuristics that are applied. These are described below.

Standardization of Units. The quantities referred to in the schematics do not use a standard unit. For example, fields such as ID or OD, may be represented in centimeters or inches. Fields such as length and depth may be represented in feet or meters depending on the well location or convention used by a particular operator. The units may appear as (m), (ft), (in), meters, mts, in, feet, mts, in an almost endless variation. They may appear in one place or multiple places in the document or in some cases do not appear at all. We therefore had to employ a heuristic algorithm that uses a fuzzy parser that can consider all these variations and establish the right unit for each metric with a very high confidence value. Once a unit was established, all quantities are converted to a standard unit system for the purposes of storing in the data lake and subsequent visualization.

Categorization of equipment. As there are too many individual equipment variations (300,000), it is important to categorize them by the family they belong to. When we perform data analysis, we can gain insights on the equipment categories first, before drilling down into individual equipment. All these are done by a combination of unsupervised learning and domain-informed rules.

Data Synthesizer. The data synthesizer (Fig. 12) component takes data from multiple sources such as the well database, service quality reports and the key value pairs in the digitized version of the schematics and combines all the relevant pieces of information found in each of these into a single synthesized record for each well. The main challenge here is that often there's no single primary key to do a direct join. This is due to the nature of the manual data input associated with these records and the lack of enforcement of any sort of industry wide consistency. The data synthesizer uses unsupervised machine learning to accomplish this by understanding the various patterns in these data sources and combining them, only when it can establish a very high confidence level.

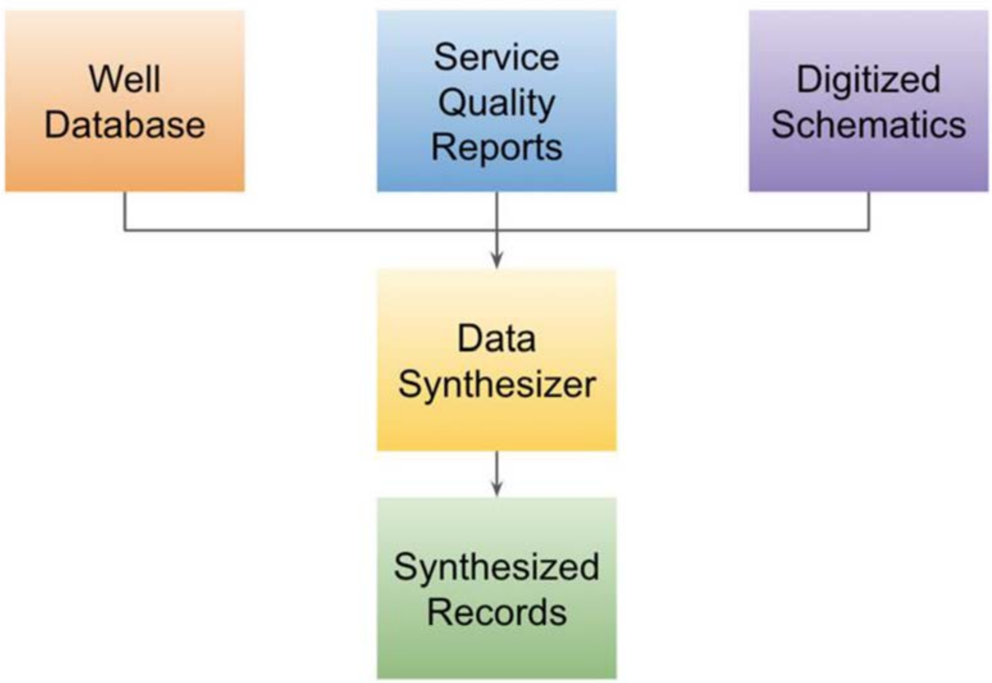

Figure 12—Data synthesizer

Thanks to the data synthesizer, we now have a structured database where every well has consolidated information about its operating environment including, for example, pressure, temperature, field and produced fluid, along with the equipment used at various depths and any failures reported with those equipment during the servicing period. We also calculated a reliability score for each equipment.

Visualization and Insights

Now that we have the consolidated data, the next step is to make this data available for mining for insights by experts where they can answer any question they may have on these wells and confirm any hypothesis about well performance. This is done by connecting the synthesized database to Microsoft Power BI*. We created different numerous views (Fig. 13) of the data including well level and equipment level views, from which all details about data for a particular equipment can be explored.

Figure 13—Example of the Power BI Insights

We also created a rich set of filters (Fig. 14) to be able to query any subset of data and find patterns within those subsets. These filters can also be customized by the individual users.

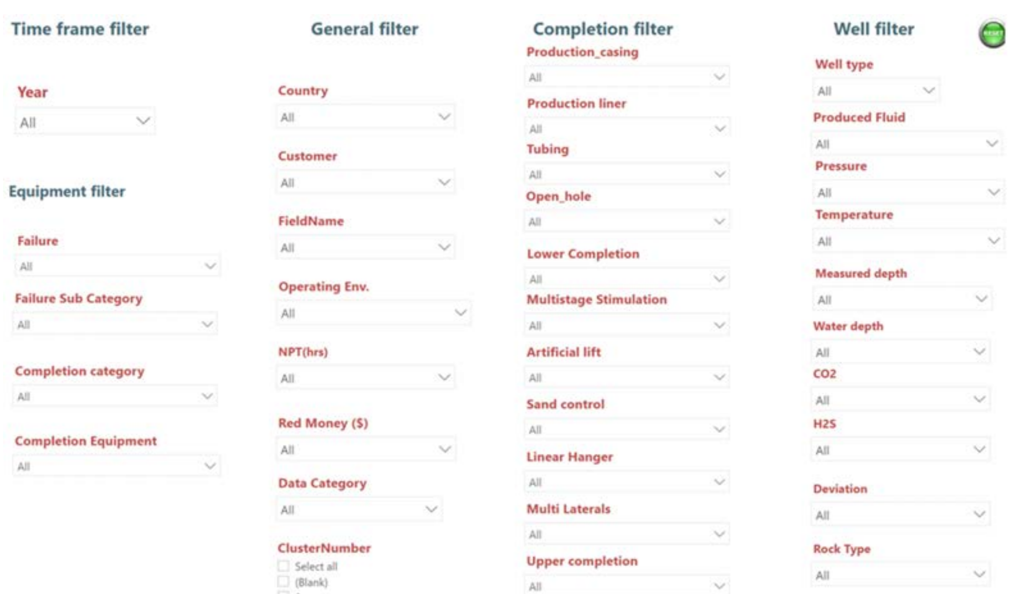

Figure 14—Data filters for end user interaction

Visualization of clusters. As part of the data analytics process, clusters of data created the regrouped similar architectures and allowed for the delivery of generalized insights based on both hardware and completion context.

The 3D plot (Fig. 15) depicts a higher-level similarity between the schematic architectures. Similarity is calculated based on features like operating conditions (e.g. pressure, temperature, and reservoir type), equipment and order of installation. Different colors represent different clusters. The optimum number of clusters is determined based on the distortion metric using the elbow method. A K-means algorithm is used for clustering and the clusters are visualized in three dimensions with the help of the principal component analysis (PCA) technique. PCA helps in reducing dimensionality by extracting important features from the input. Clustering can be done again on the clusters where intra-cluster distances are high. This gives more granularity. For clusters that are very close (inter-cluster distance is low), they can be clubbed together, which provides a good representation of the cluster.

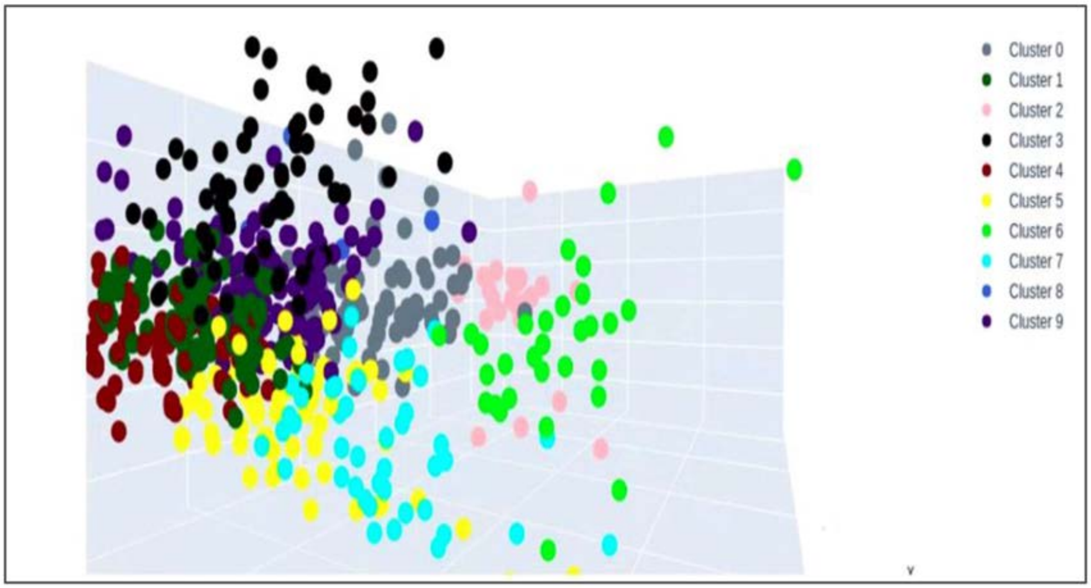

Figure 15—Visualization of clusters using PCA

Results

Harnessing pioneering natural language processes and analytics to build a scalable ingestion pipeline for completion schematics (Excel*, PDF* and WITSML*), track record and performance data, the project has successfully structured diverse data sets into a single data lake. Machine learning and analytics mapped correlations between equipment, performance, and reliability. Models harnessed all available data and provide recommendations on optimum well completion designs that are not only fit for environment but also fit for the specific well and its unique characteristics. A total of 24,032 wells have been successfully migrated, mapped, and used to develop a machine learning algorithm that can now provide optimum completion configurations and designs.

Fig. 16 shows the minimal viable product (MVP) of insights into completion configurations and designs incorporating nonproductive time, reliability of each well with respect to various operating environments and root cause failures. It showcases the completion equipment configuration and equipment detail with depths, ODs, IDs, and other key parameters. The user has the ability to define high-level operating environment constraints—allowing them to explore optimum solutions. Each solution is coupled with data insights that speak to both the reliability of the equipment and the root cause for any issues that might have previously been encountered when running these designs.

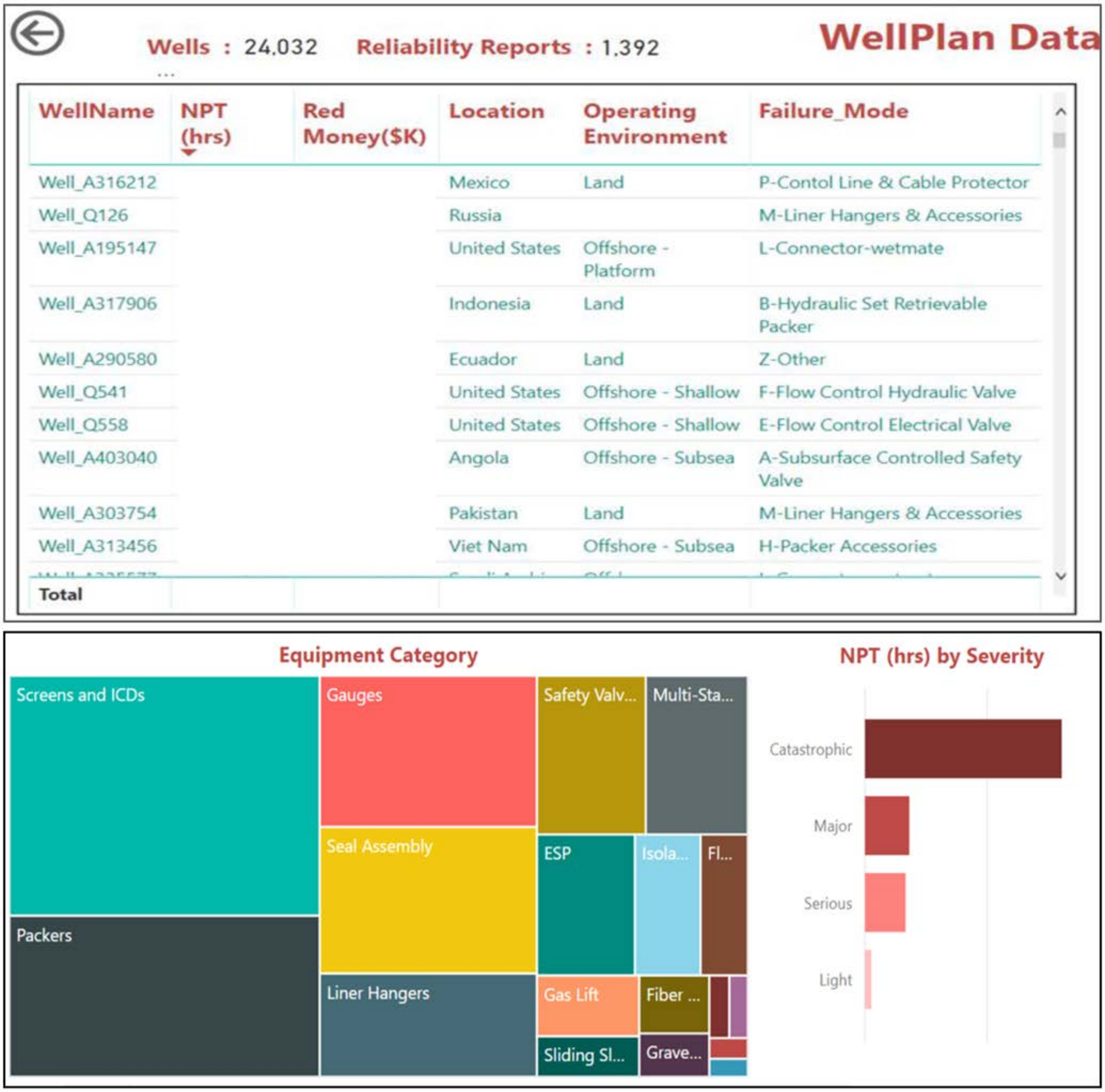

Figure 16—Visualization of clusters using PCA

The current solution was developed in a highly agile fashion targeting a true MVP. It therefore doesn't yet have the means of displaying the well schematic in a standard visual format although this is currently in development. Future versions of the solution will migrate the frontend to Angular and deliver an enhanced user experience addressing key user workflows. The core value though is delivered through the data, and future iterations will target expanding the ingestion pipelines to include more diverse data sources while also expanding on the insights available. Data remains the foundation of the product, and as the solution is agnostic of the location of the source data, it should be straightforward to couple this to diverse data sets from players across the industry so that they can liberate their own data and extract value. With the development of more insights, the value grows exponentially and delivers optimum solutions taking into consideration all knowledge and data relating to the target basin.

Conclusion

Traditional well completion selection suffers from the numerous fundamental shortcomings highlighted throughout this paper. A forward-looking solution is required to address this problem. Leveraging the latest offerings from the digital transformation have allowed for the creation of a novel approach to solving well completion design:

1. Advances in NLP, OCR, ML, AI and cloud technologies provides an ideal opportunity for transforming the well completion selection process
2. Data ingestion using NLP and OCR have successfully migrated unstructured completion schematic data into structure data
3. Structure data is mapped and correlated to diverse data sources and operating parameters for development of complex rules engines
4. Clustering of data allows for the identification of similarities in equipment failure patterns, equipment type, configurations, operational efficiencies
5. Advances in data mining and analytics modeling deliver insights into the optimum completion design
6. Cloud infrastructures provides the scalable infrastructure to deliver a solution that can be extended and organically grown as more data is made available to the solution

Acknowledgments

The authors wish to acknowledge Infiswift Technologies, a key digital partner in the elaboration and development of ideas discussed herein. Nevertheless, this paper only reflects the opinion of its authors and does not imply endorsement by the company to which acknowledgments are made.

Nomenclature

ACRONYM	MEANING
OCR	Optical character recognition
NLP	Natural language processing
ML	Machine learning
AI	Artificial intelligence
MVP	Minimal viable production
OD	Outer diameter
ID	Inner diameter
WITSML*	Trademark data exchange format from Energistics
PDF*	Trademark of Adobe
Excel*	Trademark of Microsoft
Power BI*	Trademark of Microsoft
JPEG*	Joint Photographic Experts Group

References

1. *Well Cube, Rystad Energy*
2. *Extensive formats covering Word, Excel, PDFs and custom data files from proprietary software*.
3. Szemat-Vielma, W., Murray, T., Kiaie, A., Bolchover, P., & Yao, J. (2019, November 11). *Innovative Digital Well Construction Planning Solution Enhances Coherency and Efficiency of the Drilling Design Process*. Society of Petroleum Engineers. doi:10.2118/197942-MS

4. Wetzel, R. J., Mathis, S., Ratterman, G., & Cade, R. (1999, January 1). *Completion Selection Methodology for Optimum Reservoir Performance and Project Economics in Deepwater Applications*. Society of Petroleum Engineers. doi:10.2118/56716-MS
5. Idiodemise, E. J., & Dosunmu, A. (2007, January 1). *A Model for Completion Selection for Multilateral and Multibranched Wells*. Society of Petroleum Engineers. doi:10.2118/111884-MS
6. Sinha, S., Yan, M., & Jalali, Y. (2003, January 1). *Completion Architecture: A Methodology for Integrated Well Planning*. Society of Petroleum Engineers. doi:10.2118/85315-MS
7. Jacob, S., & Naimi, K. M. (2014, April 21). *Advanced Well Completion Designs to Meet Unique Reservoir and Production Requirements*. Society of Petroleum Engineers. doi:10.2118/172215-MS
8. Pimenta, G., Bux, A., Ahmad, M., Baslaib, M., Al Jeelani, O., Al Rumaithi, F., & Turniyazov, N. (2019, November 11). *Selection of Tubing and Completion for High Sour HPHT Gas Wells – A Success Story*. Society of Petroleum Engineers. doi:10.2118/197797-MS
9. Akpenyi, C., Yuan, Z., Carson, D., & Hebert, Z. (2019, April 8). *Digital Technology Optimizes Unconventional Well Planning by Integrating Cross-Domain Expertise*. Society of Petroleum Engineers. doi:10.2118/195202-MS
10. Burrafato, S., Maliardi, A., Spagnolo, S., Cappuccio, P., & Poloni, R. (2019, May 6). Digital Disruption in Drilling & Completion Operations. Offshore Mediterranean Conference.
11. Goh, K. F. G., Gottumukkala, V., Dyer, S., Basak, D., Taor, C., & Musayev, R. (2020, January 13). Digital Well Completion Solution to Complement Extreme Reservoir Contact Drilling. International Petroleum Technology Conference. doi:10.2523/IPTC-20280-MS
12. Renpu, Wan., *Advanced Well Completion Engineering*, 3rd Edition
13. Ray Smith (2007). *An Overview of the Tesseract OCR Engine*
14. Joshua Harris (2018, February). *Improve OCR Accuracy with Advanced Image Preprocessing*
15. Megha Mathews (2018, July 17). *9 Steps to Improve OCR Accuracy*
16. Learntek (2019, February 12). *Categorizing and POS Tagging with NLTK Python*
17. Bird, S., Loper E. & Klein E. *Natural Language Toolkit*