Data readiness and standardization for data analytics studies.

Cesar Velasquez, Ivan Olea, Charles Cheung, Russell Roundtree

Abstract

In this work, we present a consistent workflow and preliminary steps required to organize, calibrate, filter and standardize the different types of well records, tables and databases involved in the use of analytics tools, Artificial Intelligence (AI) or Machine Learning (ML) techniques. For this purpose, different data sources and databases were merged into a single and integrated environment, which was used as the master reference indispensable to investigate the relationship of well completion parameters over the long-term production profile of an unconventional asset.

Introduction

Access to accurate, and timely data is critical to reach the success in the early development of an unconventional play to minimize costly mistakes and lost opportunities. For this purpose, there is a need to consolidate different sources of information into a coherent and robust platform that can be trusted inside and outside the organization for benchmarking, evaluation of investment opportunities, and field development optimization purposes. Also, as our industry becomes ever more "data driven", the need to efficiently make this information available for geoscience, engineering and data analytics tools (to extract information and knowledge) requires a new approach.

In the energy sector, companies commonly have data stored in a variety of formats which could also come from different vendors, isolated or separated by area of expertise and, consequently, without a uniform format. This situation represents an enormous challenge to implement new techniques designed to extract the value of the data due to the massive amount of manpower required to update and maintain those independent databases and the laborious processes required to identify opportunities.

Method

Different databases from the early development of an unconventional play in North America (public and proprietary production data stored in Harmony Enterprise, petrophysical properties from shapefiles and completion parameters stored in AccuMap) were used in this work,

each one containing data from diverse sources but all related to wells within the same asset (figure 1).
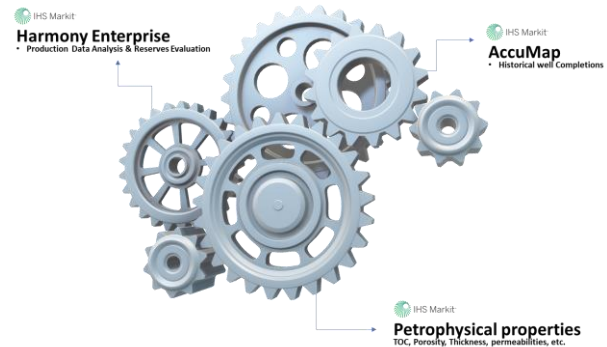

Figure 1 –Databases to be used as data sources to build an integrated and standardized data management environment.

A fully integrated data management solution was used with the aim of implementing a customizable, rules-based, engine to validate, manage and consolidate the upstream data. Many different arrays and formats were ingested into a standardized single trusted version capable of being seamlessly integrated with data analytics tools, as well as delivering reliable and valuable conclusions.
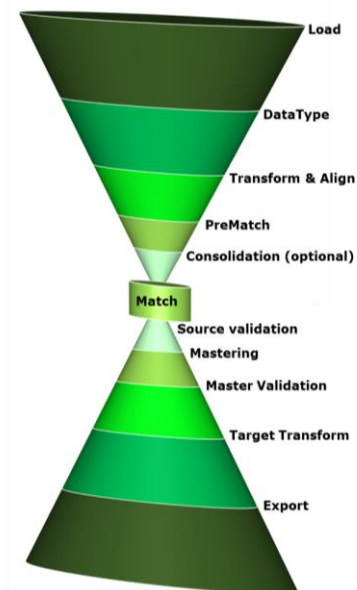

Figure 2 – Enterprise Data Management (EDM) standard workflow.

The standard workflow (figure 2) is divided in three main processes:

1. Pre-Match Processing.
2. Match.
3. Post-Match Processing.

The Pre-Match Processing was initiated with the data loading workflow, where predefined steps were implemented to pre-process the data:

- Store – Untyped data: Allows to import data as it stands to be stored in separate tables.

- Transform – Transform Data: Allows the transformation of the data into standard formats and the alignment of the source data to house standard or set of domain values contained in a reference table (reference lookup, formatting, etc). See figure 3.

- Data Type: Allows the testing of data to check if it conforms to the predefined data types. Exception rules are applied for this purpose.

- Validate: This is considered the pre-match inspection step, which allows tests to be applied to the data to determine validity and quality of the data.

- Filter: Allows rows in the data to be diverted into an additional table and excluded from the main flow.

- Function Rule: Allows the return of a table of data or values, and it is used to connect with analytics functions.

- Source Monitor: Allows the identification of changed rows at the point of loading the input data and filtering out unchanged rows.

During the matching process, it is extremely important to pay attention to the quality of the data, and any missing information. Once sources are matched, a unique identifier will be assigned.

The Post-Match Processing is related to:

- The source validation (inspection procedures).

- The mastering process, in which hierarchy rules and priorities are defined to populate certain fields in a record.

- Master Validation: This represents another inspection workflow where the data is compared constantly (daily) between different sources or vendors. The intention is to determine if the data is consistent within a certain range of tolerance, or if the data is changing constantly.

- Target Transformation: The data is manipulated and transformed as the end user has requested.

- Export: The final product (file format, data base, messages, etc.) is sent to clients or end users.
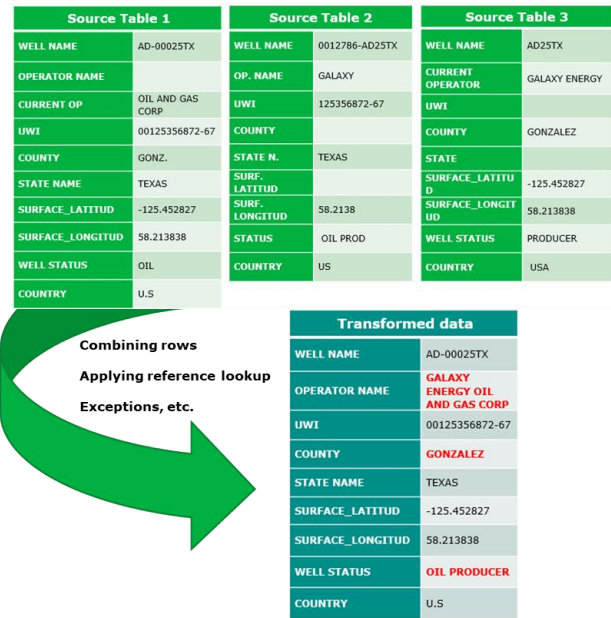

Figure 3 – *Pre-Match visualization of a transformed data*

Once a standardized product was generated in the format of a SQL database, it was connected to an advanced data analytics and visualization tool to investigate the effect of changes in rock properties by area, differences in completion designs, reduction of well spacing, clusters and stages, completion technology used, well interference, frac hits, etc. This enabled an assessment of its impact on the evolution of long-term production profile and the ultimate recovery of wells completed across the field.

Results

Before the application of the fully integrated data management solution recommended in this work, one of the workarounds previously implemented required a laborious manual process to generate a static master table that included all records together, which was almost impossible to change or update without breaking the connection between columns or records. That process made the implementation of automatic data loading-validation process and its consequent analytical solution very unlikely and unpractical.

This manual process could take a very important part of the execution of any project and, in most cases, could lead to adding human error in loading the data without any warning related to it. By the time the analytical study was in progress, the detection of these errors would be very difficult to fix and could easily cause a major delay on the study.

The integrated data management solution utilized here involves a dynamic workflow to load and inspect the quality and representativeness of the data each time new information is loaded, using business rules specifically created for each attribute. This is controlled into the sequence enclosed by the green dashed lines shown in figure 4. If the record does not pass through this cycle

successfully, then the loader process will store this record in a designated table for it. Additionally, a warning associated to the status of this record is sent to the user. However, the process of loading the following record does not stop. If the record is loaded successfully, then it is archived and it is sent to the stage table to start the match process.
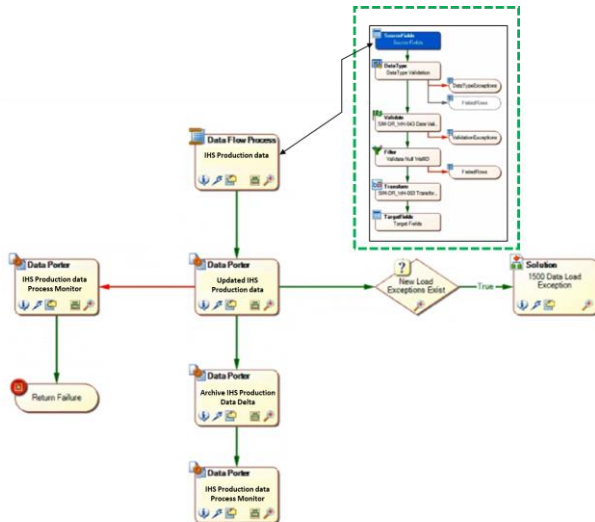

Figure 4 – *Pre-Match steps required to store and inspect the quality of the data.*

One of the advantages of this work is the fact that only IHS Markit databases were used which facilitated the interconnection between tables due to the standardization of most column names, primary key parameters, foreign keys, etc. However, the data is still collected from different sources and, for that reason, it was necessary to implement business rules to update data types, formats, etc., in different tables.

In some cases, when proprietary data from clients is used, there are no primary keys assigned to columns to be interrelated; this is a key issue to fix to avoid delays and to efficiently implement a fully integrated data management solution.

In this project, the UWI (Unique Well Identifier) was the matching attribute selected for the core matcher and mastering process, in which hierarchy rules and priorities were applied. The outcome of the core matcher process was to assign a unique identifier parameter to intercorrelate all different tables from all databases available (figure 5). The final standardized product generated (SQL database) was then exported to be interconnected with a data analytics tool (Spotfire®) to execute the detailed analysis of the information available with the aim to take full advantage of the value of it for the asset under evaluation.

The resulting data visualization environment was used to perform more granular analysis to investigate the relationship between wellbore geometry, completion practices per service company, and resulting well production performance indicators (e.g. cumulative production at six months) for a specific region.
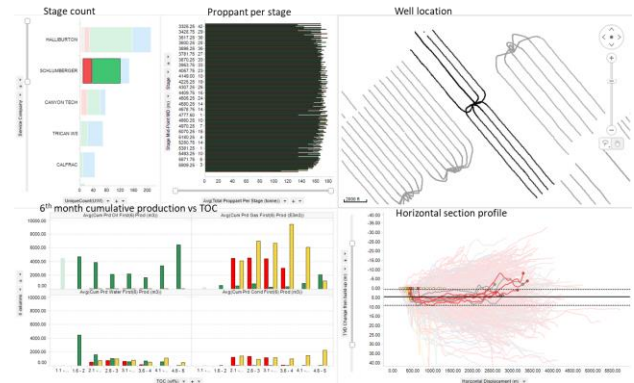

Figure 5 – *Interrelation between multiple tables through unique well identifiers.*

The outlined methodology is flexible enough to include petrophysical properties, well logs, and specialized analysis (e.g. rate transient analysis forecasts, EUR, decline parameters, etc.) on a per well basis. In this way, this visualization environment can be used as a benchmarking tool to assist in the decision-making process of the field development of a region with the understanding and big picture obtained from the whole dataset interpretation.

Similarly, this system can be extended for the purposes of hosting training datasets, required for artificial intelligence and machine learning purposes, which can benefit in having consistent and granular information.

For this work, the capability of automatically loading new public data (production and completion information) was activated in order to keep the model updated in the future.

Conclusions

- The time consumed in any data management process is usually disregarded. However, depending on the complexity of the problem, it typically represents 40%-80% of the execution of any data management / data analytics / AI or ML project.

- The success and quality of the interpretation in any project is based on the quality and representativeness of the data.

- The implementation of the fully integrated data management solution recommended in this project is extremely useful to optimize the time required and application of any data analytics solution, artificial intelligence, or machine learning algorithms.

- The integrated data solution has added a high level of confidence on the consistency and representativeness of the data analysis performed. The learnings and workflow used to analyze early development data can be used as

a guide to improve the field development plan of a region and/or play.

References

IHS Markit Enterprise Data Management (EDM) Reference manual. P. 20 – 75

CURTIS, T., The Speed of Data, SPE Workshop: Subsurface Data Analytics, 26-27 February, Calgary, Alberta, Canada, 2019

VAN DER SPRUY, M., AI Driven, Data Fueled: Data Preparation, Management and Storage, SPE Workshop: Subsurface Data Analytics, 26-27 February, Calgary, Alberta, Canada, 2019

GRANT, L., Under the Hood: Montney Integrated Analytics, SPE Workshop: Subsurface Data Analytics, 26-27 February, Calgary, Alberta, Canada, 2019

FARID, K., BHATTI, K., DASSINGER, J., FUDICKAR, B., FUGE, C., GARTRELL, K., KRISTENSEN, E.L., KUZMA, H., MCLAUGHLIN, S., and NICHOLS, D., Production Data Management and Surveillance in Shale Operations. SPE Intelligent Energy Conference and Exhibition, 1-3 April, Utrecht, The Netherlands. SPE-167863-MS, 2014

FUGE, C., DASSINGER, J., BHATTI, K., FARID, K., NICHOLS, D., and MCLAUGHLIN, S., Information Technology and Field Data Required to Successfully Manage Shale Production Operations. SPE Digital Energy Conference, 5-7 March, The Woodlands, Texas, USA. SPE-163705-MS, 2013

DENNEY, D, Forties Infill Drilling 8 Years Later: Development Driven by 4D Seismic, Journal of Petroleum Technology, Vol. 64, Issue 03 (March). SPE-0312-0114-JPT, 2012

BRULE, M., Fusion of Engineering and IT Transforms the E&P Business, Journal of Petroleum Technology, Vol. 62, Issue 04 (April). SPE-0410-0032-JPT, 2010