SPE-205761-MS

A Step Change in the Digital Oilfield Arena: Cloud Computing and Workflow Integration for Production Operations Solutions

Aditya Kotiyal, Guru Prasad Nagaraj, and Lester Tugung Michael, Schlumberger

This paper was prepared for presentation at the SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition held virtually on 12 - 14 October, 2021. The official proceedings were published online on 4 October, 2021.

Abstract

Digital oilfield applications have been implemented in numerous operating companies to streamline processes and automate workflows to optimize oil and gas production in real-time. These applications are mostly deployed using traditional on-premises systems; where maintenance, accessibility and scalability serves as a major bottleneck for an efficient outcome. In addition to this challenge, the sector still faces limitations in data integration from disparate data sources, liberation of consolidated data for consumption and cross domain workflow orchestration of that data.

The dimensional change brought by digital transformation strategies has paved a path for the Cloud-based solutions, which have recently gained momentum in the oil and gas industry pertaining to their wider accessibility, simpler customization, greater system stability and scalability to support larger amount of data in a performant way.

To address the challenges mentioned earlier, we have embarked on a journey with **Production Data Foundation** which brings together production and equipment data from across an organization.

In this paper, we will highlight how Production Data Foundation, hosted on the cloud, provides the underlying infrastructure, services, interfaces required to support and unify production data ingestion, workflow orchestration, and through the alignment of the common domain and digital concepts, improve collaboration between people in distinct roles, such as production engineers, reservoir engineers, drilling engineers, deployment engineers, software developers, data scientists, architects, and subject matter experts (SME) working with production operations products and solutions.

Introduction

Production operations is highly complex, spanning a range of business roles, geographies, and workflows, often addressed by a diverse, disconnected, set of software applications and tools. All these

factors meet at a single junction - data. Data gathered in production operations can become problematic and highly complicated if not managed properly. Some of these challenges are [Melo, 2019]:

- The collection of data, and its availability, from disparate data sources and systems

- A wide range of data frequencies, ranging from seconds to yearly
- An array of conventions, standards and nomenclatures used in software and workflows
- The maintenance, accessibility, performance, and scalability of on-premise software applications and tools

The above challenges may result in the significant impairment of data value extraction:

- Non-productive time, when data is not being utilized
- Data quality issues
- Inconsistencies in data and data-driven workflows, such as running calculations
- Missing or incomplete data

Key business drivers that prevail in production operations are the reduction of costs to remain competitive and to enhance production. Given cost pressures, achieving these goals should come from existing infrastructure, rather than through new investment.

The constant changes in technology and industry efficiency demand an easier and smarter solution that covers both infrastructure and digital capabilities. In a data-centric solution, such as the one that production operations demand, a consumer aims to overcome the complexity of managing different and disparate data sources to reach to faster and smarter decision making with additional capabilities, which includes integrated analytics, machine learning, achieving greater levels of access and consumption through a continuously enriched context.

In this paper, we have discussed these challenges and demands, and how Production Data Foundation is solving them efficiently. For ease of understanding, the capabilities of Production Data Foundation are divided into three areas:

1. Data ingestion
2. Production domain model
3. Consumption and workflows

Data Ingestion

A good advisory software (solution) needs to have data available in a reliable and timely fashion, and thus relies on a robust data-centric foundation. When dealing with production operational data, number of datatypes, coupled with data definitions, and data tags can get exponentially complex.

Some of the common data-sources found in production operations are:

- Production Data Management Solutions (PDMS)
- Corporate historians with high-frequency data flowing in from instrumented wells and sensors
- Edge devices streaming in high-frequency data
- Manual data entries via ticket or mobile apps
- Spreadsheets and CSV (Comma Separated Values) files
- Calculated data via simulations or physical models

The ability to capture and continuously integrate data from the data sources is a challenge. Based on the article published at Harvard Business Review [Dallemule et al, 2017], an employee typically spends around 80% of his or her time looking for data and unifying data sources.

The solution to this problem has multiple parts, starting with getting data in from all the sources into a single data storage system based on common domain concepts that adopts industry standards. Production Data Foundation addresses the problem by supporting diverse data types and frequencies while enabling data contextualization with concepts provided by production domain model.

The foundation provides a variety of performant data adaptors and an API (Application Programming Interface) framework, allowing ingestion of data into a scalable storage from the sources mentioned above. The framework provides an efficient way to ingest data, irrespective of the quality and size, and based on its scaling capabilities, enables the deployment and setting up of automatic historical ingestion or ingestion of any incremental changes. This capability increases confidence amongst the users as there is no need to spend time and effort on collecting data and the process is streamlined by facilitating data consumption in the way the domain workflows require it [Isah et al, 2018].

Another significant concern is moving data securely from a source system to cloud-based storage. Production Data Foundation provides the data security during its flow, whether it is an on-premise deployment of the data adaptors or a cloud-based connector, by authorizing the connection through the concept of cloud service accounts and making sure that the data pipeline is encrypted from a source database to the cloud storage. The storage technologies used in Production Data Foundation for operational data related to different measurements and associated asset models are adopted after extensive research to satisfy multiple use cases and workflows in the production domain.

The Foundation supports either push or pull-based data ingestion mechanism, which increases data flow efficiency and reduces latency, which in turn allows an application or a workflow to generate results on the fly with the latest available data. Fig.1 illustrates the workflow.

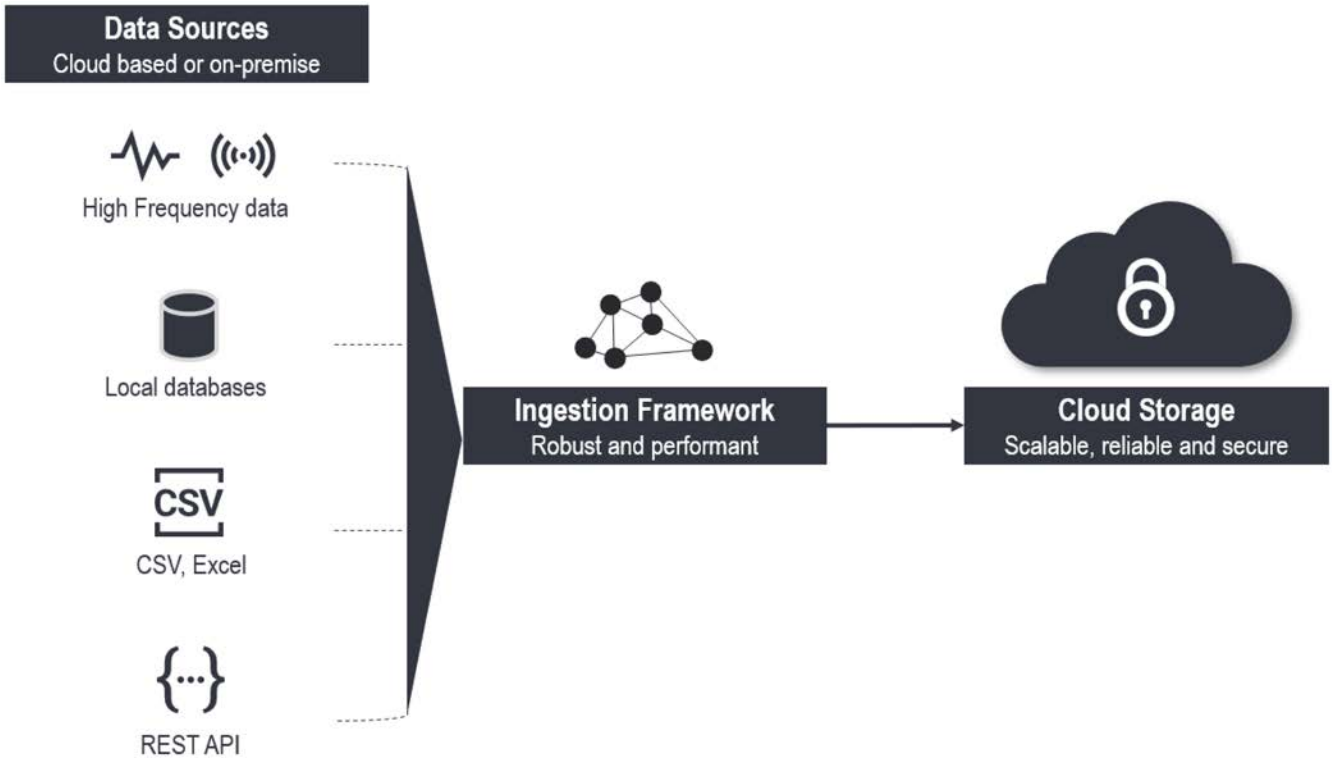

Fig. 1—Ingestion framework workflow.

A performant ingestion framework needs to be followed up with a standard schema which every application, workflow, solution, and more importantly a user can understand. This standard schema is represented by Production Data Foundation's domain model and is used as a universal language to map

source system details to the industry adopted domain concepts like entities, relationships of those entities and associated properties.

Production Domain Model

The production domain model is a canonical model that represents a collection of entities, related properties, and relationships in which the entities can be linked together following different rules and validations. The production operations domain supports fields, assets, surface and subsurface equipment, wells, boreholes, completions, among other types of entities widely used in the industry. It allows unification of data from dissimilar data sources and represents a model for both raw and calculated data.

The production domain model leverages decades of proven understanding of domain data models used in different applications related to production domain and combines all of them together to formulate multiple concepts supporting different verticals of the production domain, from upstream to midstream.

Three major sections of the data model include:

a. **Entities**, which are objects that represents anything that has a physical or a virtual boundary, such as a well or a compressor.
b. The **associated properties** representing the collection of data points for a particular entity. For example, *oil rate* which has different facets like *product, measurement* and *units* associated with it and can be used in multiple workflows based on different methods of acquisition like *allocated oil flowrate* coming from an allocation engine to *estimated oil flowrate* coming from a simulation engine.
c. The **relationships**, which represents the hierarchical linking between entities based on how a consumer wants to use them. For example, user A may use a relationship of entities described under their organization structure, whereas user B is more interested in technical workflows and is looking only for equipment relationships. Fig.2 shows an example of the two use cases.

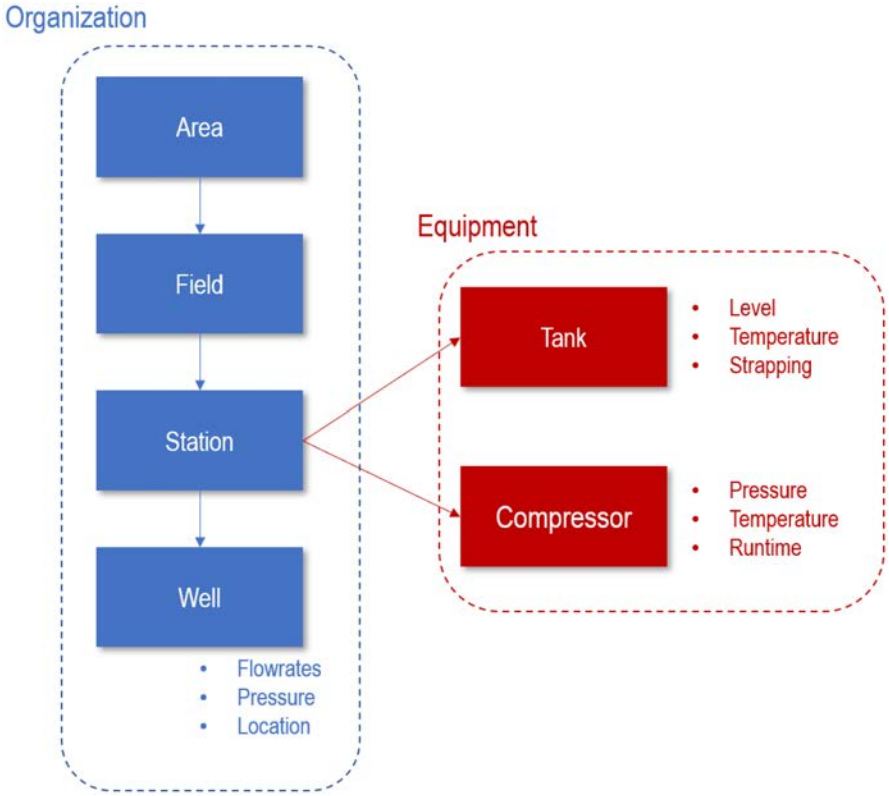

Fig. 2—Relationship between entities (solid boxes) and properties (bullet points)

The use of a common domain reference via a proven model helps deliver a sustainable and cost-effective solution, as it minimizes the effort to maintain and keeps consistency across teams and applications.

To demonstrate the value of having the production domain model, an example is chosen below. Imagine two different sources:

- Source A is a high-frequency data source streaming in electrical submersible pumps (ESP) frequency every minute through a sensor set on the pump assembly and stored behind a tag name called **esp.frequency.Hz.**
- Source B is a low-frequency data source storing ESP frequency manually added every day by a field operator and stored under a property called **esp_freq_Hz.**

The two different sources above, represent a single domain concept - **ESP frequency**. If a consumer wants to use the data, she or he needs to deal with two independent data streams (from source A & source B separately), which requires additional development to handle properly in an application i.e. in order to represent a single line in a chart or provide a single input to a calculation. This results in more cost of development, increased complexity in the system and the entire consistency of data.

For ingestion, consumption and an end user, the property of interest is **ESP frequency.**

This is where having a canonical model of domain concepts helps. A simple mapping of two different source properties to a common domain model property, named *ESP frequency* solves the problem of unifying data from two different sources of varied frequency and allows an application to include a consistent and reliable mechanism to use data from the storage via different consumption mechanisms (discussed in the workflows section). The illustration of the workflow is shown in Fig.3 where:

- Source A's **esp.frequency.Hz** is now mapped to domain model's *ESPfrequency*
- Source B's **esp_freq_Hz** is also mapped to domain model's *ESPfrequency*

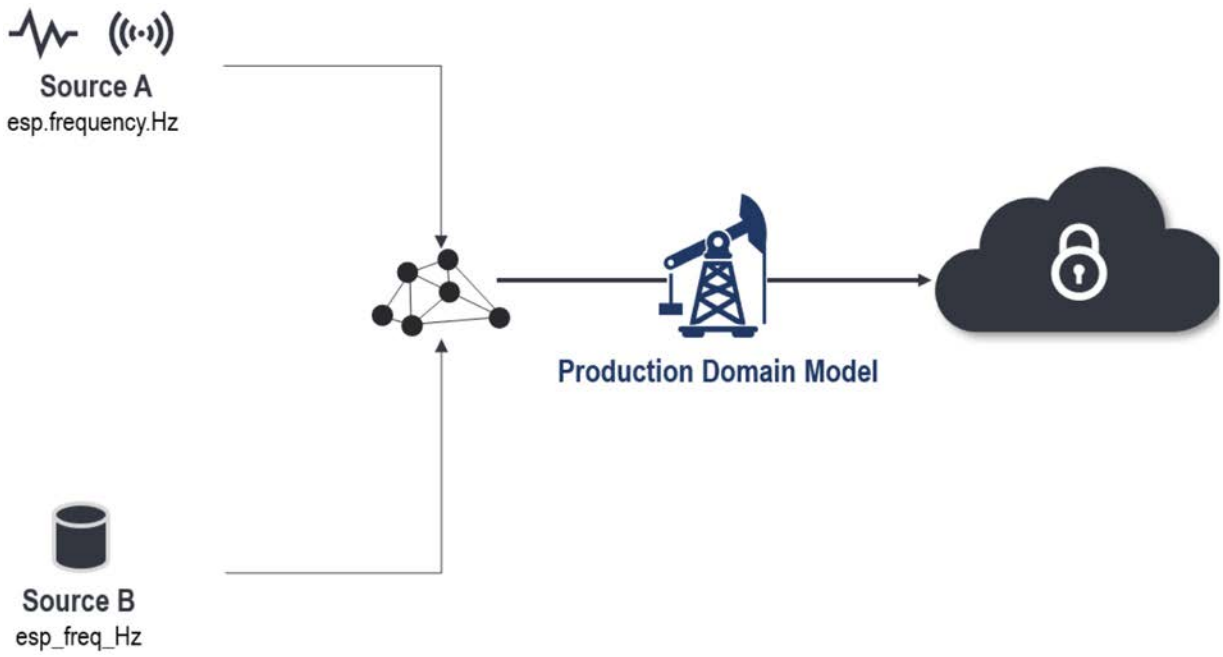

Fig. 3—Linking Ingestion framework and Production library

For any software solution, user experience is an important dimension which cannot be ignored as interaction and ease of using an application defines its global adoption and footprint. Production data foundation has also made the experience of these mappings and configuration activities user-friendly and

seamless through its next generation interfaces following a governance for styling, user workflows and administrative tasks. Fig. 4 shows some of the interfaces used today.

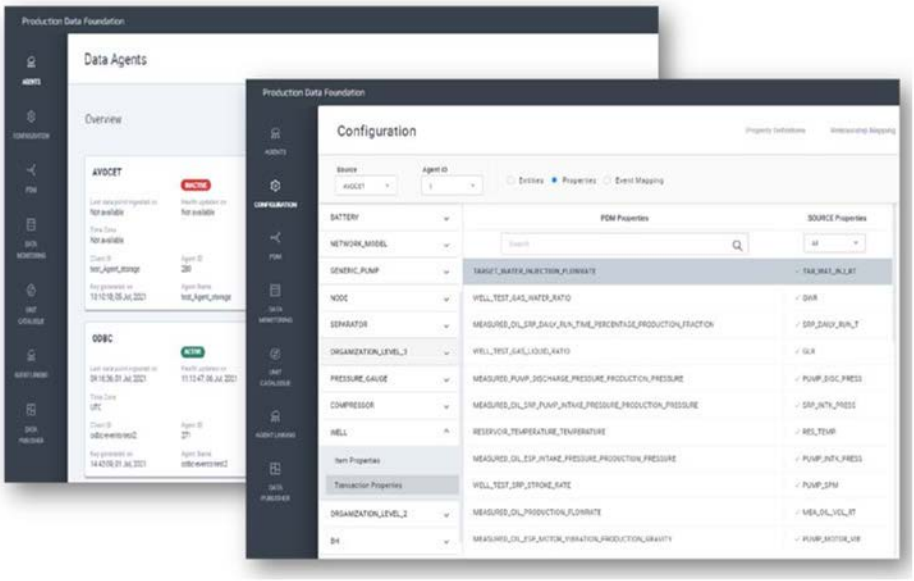

Fig. 4—Production Data Foundation: Administrative interface

Consumption and workflows

Data ingestion and storage on top of a robust model is only half of this data flow journey. This data needs to be conditioned, transformed and then prepared for the workflows to run on top of it using reliable and performant consumption services. The way of expressing a use case and the way of delivering a solution around that have changed in this full circle of digital transformation. The most important factors to successfully deliver the requirements on this data flow journey, besides time, are efficiency and quality.

The consumption and workflow services allow users to identify what data has been ingested and stored and retrieve it the way they need it, for example:

- Looks for entity and associated properties using the canonical model
- Looks for a relationship, and traverse it as required
- Calculate, aggregate, consume, writeback time-series data
- Apply data quality attributes
- Consume data and apply it to a calculation engine
- Throughout any of these activities, the process must be fast, reliable, and scalable.

For any production operations workflow, all the above features are a necessity. Any missing part represents a bottleneck for the entire solution. The microservices architecture allows the required scalability and helps distribute consumption & workflow services efficiently across the system based on which aspect of the data is being called for. The illustration of the workflow is shown in Fig.5.

Another key factor is to make sure that multiple end users have their data secure from each other, and different users can access only certain data to which they are entitled. For example, company XYZ has two assets; asset A and asset B. A user should only access data for asset A because he or she is not authorized to view or consume data from asset B. The reasons can be specific to location, contracts, employee's profile, department and data category. This data security and privacy protection is handled by data entitlements

implemented in the foundation, thus allowing users to view, share and consume data they are entitled to and making sure that the confidentiality and integrity of data storage and transmission is maintained [Chen et al, 2012].

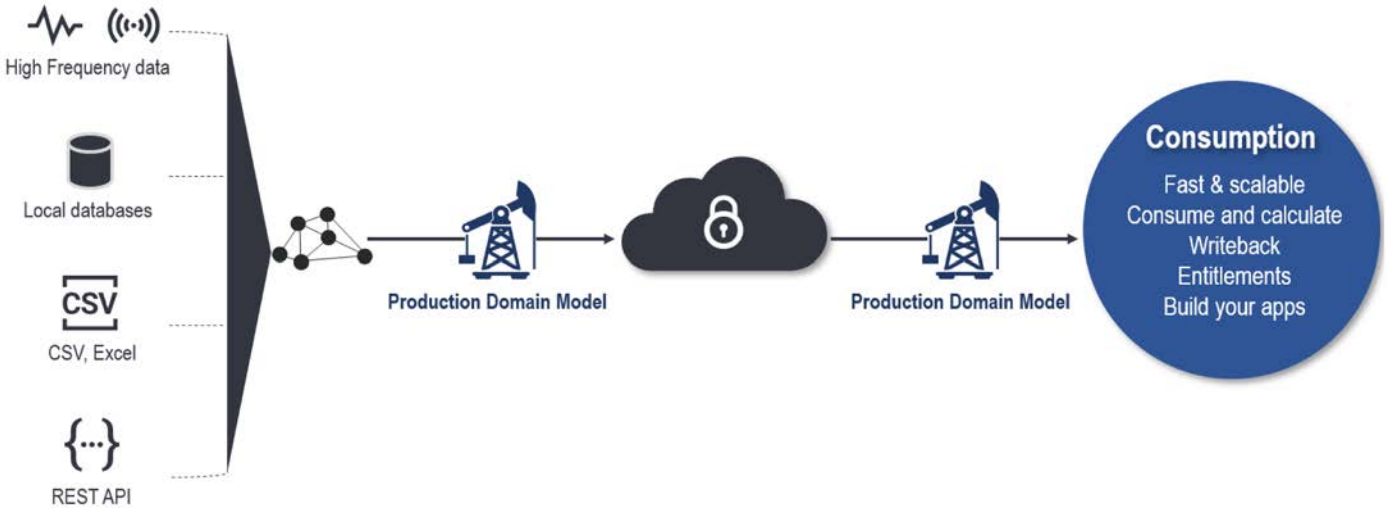

Fig. 5—Linking Ingestion framework, Production Library, and Workflow services

Applications and Integration

Extracting data and using it for visualization or running calculations is the final piece of the puzzle. One of the key issues with existing solutions is the consistency of data along with holdups related to performance. In addition, legal and audit aspects related to data storage are as important as any other factors governing data in a system. That is handled by having two distinct time dimensions in the storage, *the application time* which refers to the physical time of measurement in the actual world and then *the storage time* representing the time when a datapoint is stored in the storage. This concept is referred to as Bitemporality. Bitemporality of structure and time-series data helps to provide consistency, reproducibility and allows for executing long-running calculations due to the immutability aspect [Kaufmann et al, 2015].

At a minimum, *native out-of-the-box* applications are plug-and-play with production data foundation, but the possibilities do not just stop there. The consumption and workflow APIs are open to allow bring your own and build your own capabilities with independent engines, applications, and models supported by reliable integration between different systems. The consumption services allow analytics applications to query data in a performant and coherent manner allowing ease of visualization and empowering the users to make smarter decisions.

Some of the production-related application solutions that the production data foundation supports today include a consolidated production overview, well surveillance for producers, injectors, and disposals, well intervention management, production forecasting, model management and analytics for production optimization and process facility operations. Operators can monitor and keep track of daily production in comparison with production target. Automated and actionable well insights allow user to proactively schedule, execute and track daily well optimization initiatives. Engineers no longer need to focus on data gathering and data quality check but are now able to focus on higher-value-adding tasks and production challenges [Hassan et al, 2019] such as production enhancement initiatives. Fig. 6 illustrates the application benefited with foundation's offerings.

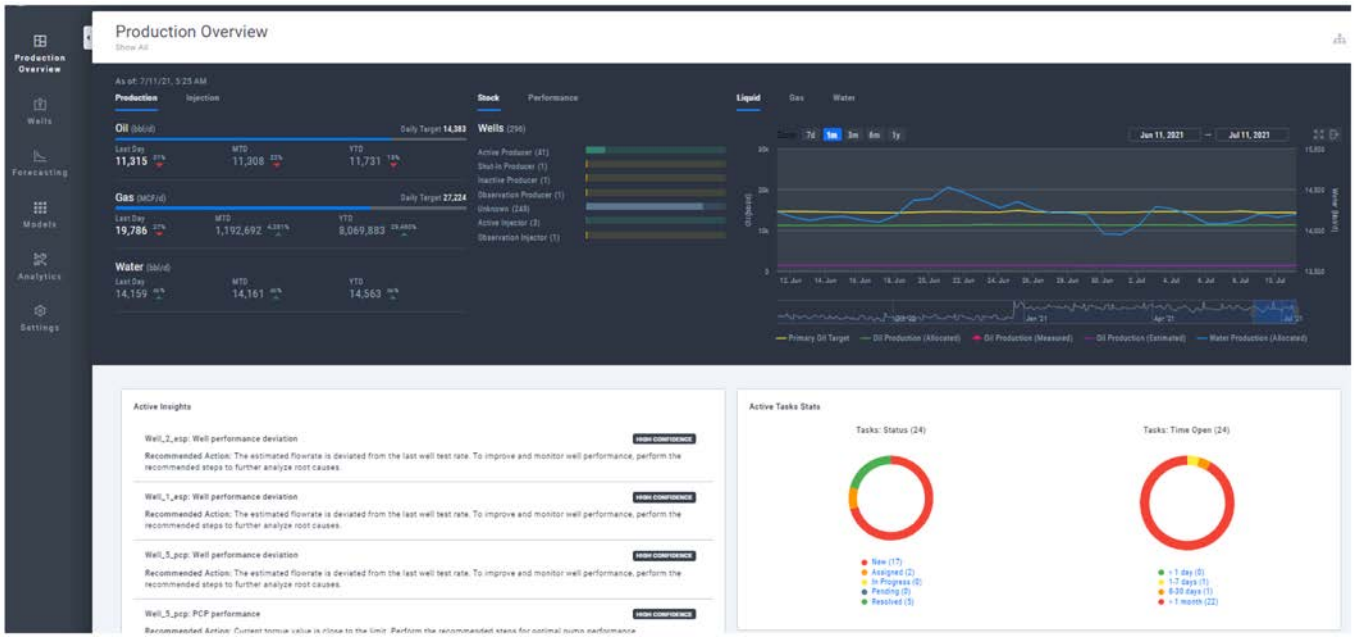

Fig. 6—Tuned production operations solution

A well performance application is focused on the key value activities to operate an asset through management by exception. Well portfolio optimization is a comprehensive opportunity management system to mature production enhancement opportunities from candidate recognition to job execution. The inclusion of past intervention best practices and lessons learned into this system predicts the chance of success for every intervention job [Hassan et al, 2019]. Engineers are now able to spend more time maturing production enhancement opportunities instead of looking for them in the first place.

Other applications such as production forecasting provides engineers with a continuous understanding of targets and planning thorough an automated, evergreen pipeline of forecasts. Model management ensures a seamless, automated way of calibrating and updating models for use in production-related workflows. Process facility operations application focuses on surface equipment surveillance related to oil, water and gas fluid handling. It also helps engineers to perform equipment prognostics health monitoring (PHM) analytics.

Through production data foundation, all these applications help operators to turn their production data into valuable insights in the cloud space, therefore allowing them to make timely decisions in production optimization and maximizing the true value chain of production operations.

Open frameworks on both sides of the system allow a collaborative canvas to integrate with existing solutions, which, in turn, allows extension to artificial intelligence workflows, data science and data quality attributes, thus allowing maximum value extraction.

Inbuilt and custom plugins can bring in significant amounts data sets over a given time, enabling creating and training machine learning models and writing back those calculations into production data foundation for future reference. In addition, the microservices architecture helps scale the solution flexibly and improves the speed of delivery of new features or functionality using agile methodology. The aim of liberating and orchestrating data solves not only existing industry issues but helps in exploiting areas that were not even accessible before. Fig 7. illustrates a custom analytics dashboard created on top of the foundation's consumption API.

Fig. 7—Custom analytics dashboard using foundation's consumption APIs and AI integration capabilities

Conclusions

Production Data Foundation allows unification and liberation of production operational data using a robust and scalable framework which provides data ingestion from live and operational data stores based on a rich and extensible domain model. This simplified approach supplies an automated pipeline of continuously enriched data (from different data sources). The scalable architecture is designed to be performant and allows data consistency, thus helping to manage long running calculations and out-oforder data handling. The history of changes to the data is maintained in the system which allows to request and consume data at a specific point and state and also helps in auditing and debugging the system. Quick data discovery via easy-to-use consumption APIs enables the fidelity required by various end-users.

The foundation constitutes an integration layer with a third-party artificial intelligence and machine learning application that reads the data, detects the events and then writes result back to the foundation's storage where the results can be consumed in business intelligence visualization. This end-to-end integration allows faster and smarter decision making. It empowers the consumers and end-users with the tools to define, build, share, operationalize and maintain their own workspaces and applications and create their own Intellectual Property (IP) which in turn helps to maximize return on investment and enables innovation.

References

1. Chen, Deyan & Zhao, Hong. (2012). Data Security and Privacy Protection Issues in Cloud Computing. Proceedings - 2012 International Conference on Computer Science and Electronics Engineering, ICCSEE 2012. 1.10.1109/ICCSEE.2012.193.
2. Dallemule, Leandro & Davenport, Thomas H."What's Your Data Strategy?" published at Harvard Business Review, May-June2017.
3. Isah, Haruna & Zulkernine, Farhana. (2018). A Scalable and Robust Framework for Data Stream Ingestion.
4. Kaufmann, Martin & Fischer, Peter & May, Norman & Ge, Chang & Goel, Anil & Kossmann, Donald. (2015). Bitemporal Timeline Index: A data structure for Processing Queries on bi-

temporal data. Proceedings - International Conference on Data Engineering. 2015. 471–482. 10.1109/ICDE.2015.7113307.

5. Marwa Hassan, David Rossi, Galina Ivanova. 2019: From digital oileld to operational excellence. *Journal of the Japanese Association for Petroleum Technology* Vol. **84**, No. 6 (Nov. 2019) pp. 394–402.
6. Melo, Dean.(2019). Data management evolves, but challenges persist. Published at JPT.