Cloud Enables and Accelerates Next-Generation Collaboration
Cloud Enables and Accelerates
Next-Generation Collaboration

Author: Chandra Yeleshwarapu, Director, Information Management and Platform Technologies

Abstract

Information management as a service (IMaaS) is a ground-up framework involving technology (information management solutions and a foundational platform), people, and processes to manage and deliver petro-technical data and visualization securely through the cloud, and ensure safe, fast, and more accurate reservoir decisions. IMaaS will unlock data’s potential and value to the shareholders of upstream E&P companies by accelerating reserves replacement and maximizing asset recovery. This paper outlines an approach for E&P companies to manage and deliver collaborative information management from the cloud.

IMaaS must support:

- Hyper-collaboration
- Domain-specific data stores and multiple applications
- Master data management
- Integration using data federation and a common data model
- Ubiquitous data delivered through multiple deployment models, and support-connected and disconnected use of information.

IMaaS is made possible with adherence and adoption of an open, modular, and flexible enterprise platform.

Executive Summary

Upstream E&P generates and accumulates extreme amounts of data through the life cycle of a prospect and asset from data acquisition to decision and archival. This data comes in various sizes, formats, standards, sources, regions, speeds, frequency, and quality, and is used by different technology domains – geophysics, geology, earth modeling, reservoir engineering, drilling and completions, and production. This myriad of sources, sizes, and types of data creates a business challenge to generate, access, retrieve, use, and archive the data to maximize value to the shareholders. The varying data needs posed by deep water, unconventional reservoirs, and mature assets, and the continuing trend of data explosion in every domain and data store in the company adds more complexity. There is a need for a closed-loop system between the domains, but the data challenges result in a linear flow of the decision-making process at each stage. There is data everywhere in many data stores and files. There is, however, limited information.

These challenges are overcome by business users, data managers, and the information management organization by resorting to manual data loading and quality control, moving and duplicating data, working on small data sets, domain application proliferation, and enforcing governance in the form of inflexible business rules. There is a constant desire to improve data quality and business flexibility, and upstream E&P companies are moving legacy data and architecture to next-generation solutions.
This paper outlines the technology framework to combat upstream E&P data challenges by delivering information management as a service to end users and domain applications, and enable safe, fast, and more accurate reservoir decisions, accelerate reserves replacement, and maximize asset recovery. IMaaS embraces the diverse application portfolio of a company and their need to create, develop, and protect intellectual property, and an ability to migrate to the framework over time to mitigate risk, suit their growth, strategy, and operations.

IMaaS has five principals:

1. Hyper-collaboration ensures that data in one application and domain is available in near real time for visualization to other applications and domains to form a continuous closed-loop system.
2. Support domain-specific data stores to ensure business flexibility and continuity.
3. Include master data management to load, author, and manage entities (e.g., seismic, location, well, borehole, and production) and ensure system-wide data quality, security, and audit.
4. Strong adherence to leveraging data federation and a common data model to enable integration:
   - Diverse application choices made by the business for their domain workflows result in multiple domain data stores.
   - Technology considerations while dealing with different data types and sources (seismic, interpretation, engineering and completions, real-time data, borehole, time series data or reservoir, production historians, documents, images, and videos), resulting in the need to manage relational, semi-structured, columnar, data warehouse, and unstructured data.
   - A strict adherence to the principal of storing data in only one place, and the need for automating data integration between multiple data stores and applications to improve ease of data loading or eliminate data loading altogether.
   - Need to support industry-standard formats for data transmission and data access by applications that embrace and support standards.
5. Information usage and delivery in multiple deployment models—Web, mobile, laptop, collaboration centers, and usages—connected and disconnected.

IMaaS supports multiple scientific applications by adopting an open, modular, and flexible multitier enterprise platform that enables:

- Basin-scale data sets and visualization
- Connected and disconnected usage modes for applications and information management
- Multiple deployment and delivery models through a private, virtual private, or public cloud to any device running petrotechnical visualization applications, such as Web, mobile, workstation, or laptop
Delivering solutions using these principles provide business users with applications that deliver science on trusted data, cross-domain workflows, and visualization.

Introduction

E&P upstream operators strive to accelerate reserves replacement and maximize asset recovery. The resulting need for lifecycle optimization—from prospect to production—for the assets and reservoir creates data management challenges that are unparalleled in other industries. These challenges stem from the scope of information management in the industry. The scope is defined and differentiated by the following:

- The E&P domains and their differing needs
- Types and sources of data
- Real-time data and the need for integrated, closed-loop domain systems
- The unique E&P challenges of deep water, unconventional plays, and mature fields, and their differing data needs
- Proliferation of applications in each company and the conflict between application versions and data store principles
- Data standards and their implementation and interpretation

E&P Domains and Their Differing Needs

The prospect-to-production life cycle and the need for optimization of a company’s assets has created a need for a cross-disciplinary skill set that blends earth science and engineering disciplines, and has created domains in the industry and departments in the companies. These domains and disciplines can be classified into the following: exploration, reservoir evaluation, drilling and completions, production, and economics. These domains are further classified and broken into geophysics, geology, earth modeling, reservoir engineering, drilling and completion, and production. The science and heuristics used in these disciplines, and the skills needed to work on these, create diverse data and visualization needs—from large seismic acquisition, processing and visualization of pre-stack and post-stack data, visualization of borehole data,
real-time drilling and completions information from downhole telemetry all the way to dealing with real-time visualization from historians and field supervisory control and data acquisition (SCADA) systems in production.

The asset life cycle, however, requires these domains and departments to work in concert to create a closed-loop process of making decisions. The differing data and visualization needs have resulted in numerous software applications and data stores being used, resulting in fragmented data and information. The cycle time from exploration to production remains long, stretching from years to decades. This requires raw and processed data to be managed and archived for long periods of activity and inactivity.

**Figure 2: Sub-domains in upstream E&P**

### Types and Sources of Data

The data sources for the industry remain varied, and the data is generated, enriched, and augmented at different points in the life cycle of the assets.

- **Exploration Data** – These sources include seismic acquisition firms, national data repositories, seismic processing firms, land and lease data from federal and state agencies, and location and image data from myriad sources. The challenge in this stage is to easily access raw data and corresponding metadata, and to process and visualize data in different ways to determine the reservoir characteristics, and to map the subsurface accurately.

- **Drilling and Evaluation** – These sources include many domain applications, but the prime source of the information is from public sources (national data repositories), content aggregators (IHS and drill net), and service suppliers working on the various stages of the well construction and testing. The challenge in this stage includes having access to the best information to efficiently and safely drill and complete the producing wells.

- **Production and Economics** – The sources of this data include the field and instruments, valves, gauges, flow, pressure, and temperature equipment, and the result of the many actions taken to optimally lift and transport the liquids and gas.
The E&P Challenges and Their Differing Data Needs

The challenges of dealing with deep water, unconventional and mature fields are different.

- Deep water challenges focus on large amounts of seismic data and processing, with telemetry information from downhole tools and field instrumentation. Data transmission to central command centers and the governance and risk compliance data needs set by local regulatory environments are highly prevalent.

- The unconventional reservoir and its exploitation is especially complicated due to the speed required to obtain the maximum returns from such an asset. Intelligent drilling mandates cost and efficiency requirements (e.g., the ability to integrate geosciences to drilling and completions and production process to refine the next set of wells and maximize recovery, as well as the ability to coordinate service crews, equipment availability, logistics and transportation, and legal and lease operations).

- Mature fields need the capability to monitor and manage the entire field. The operational response of the reservoir, through historians and field instrumentation, against the original plans and assumptions of the asset helps optimize production and improve recovery. The use of production data in geology and formation evaluation will result in the company's ability to affect needed changes to refine the operations on the field.

Real Time and the Need for Integrated, Closed-Loop Domain Systems

E&P data lives in silos. It is imperative to close the information gap between the various domains to improve safety, speed, and accuracy of the reservoir decisions. There is a need for cross-domain workflows, and the ability to exchange information and data easily between steps in the workflow independent of the domains or sources that generate or use the data. The domains and silos within a company must move towards collaboration, where the data in one domain is available in real time or almost real time in other domains.

With more information being pushed by real-time data coming from downhole tools at the time of drilling and completions, or field data coming from historians and on field instrumentation in production, an opportunity is created to refine the assumptions made in preceding domains and related models, and help the company optimize its workflows and continuously improve its processes.

Proliferation of Applications with Differing Principles for Data Storage

Upstream E&P responded to the challenge of reserve replacement and reservoir optimization by providing to their domain scientists and users a plethora of tools and science—both grown organically and acquired through independent software vendors. These tools have been meticulously put together and manually integrated to perform a variety of complex workflows. In a company’s petrotechnical portfolio, these tools can number from a dozen to over a thousand applications, creating a highly complex environment for technical computing and information management.

It is expensive to manage, upgrade, and maintain these tools in such complex environments. The web of dependencies across the tools, applications, and the various data stores form an intricate and complex software portfolio, making upgrades across the various technology and applications challenging at best.

The underlying data storage philosophy of most petrotechnical applications is to create a closed system or store data in legacy storage formats. This creates workflow integration challenges.
Data Standards and Their Implementation

Standards organizations (such as by Energistics and PPDM) provide a standard data dictionary for use in both the domain workflows as well as for exchanging data between domains and business partners. However, the standards are also a challenge. In a highly competitive environment, to differentiate from one another and increase value, operators and service providers must innovate at a rapid pace. The innovation demand outweighs the company’s ability to stick to a standard or wait for a new standard before forging a path forward. The priorities of an individual company or partner, the implementations and interpretations of the standards, or the mass adoption of the standard for cross-domain integration is disruptive to business and time consuming. This typically results in the standards being rich and comprehensive in a few sub-domains, and the adoptions of such standards are high in those domains as compared to others.

Information Management as a Service

IMaaS provides the capability and framework of the internal business (IT and business) or a solution partner to deliver high-quality and relevant information from a private or virtual private cloud. This eliminates the need to search for various data files, blend, and create the right version of the data for consumption by the domain user. IMaaS reduces the constant need of loading data into various tools, and the need to create data sets for transfer to other domains. Information is always available at the user’s fingertips. IMaaS does this with a holistic approach to managing information in context of the asset workflow and the need of the domain users.

Using IMaaS, a company is able to acquire and load the data automatically or manually into data staging areas. This data is cleansed and augmented using various data quality tools and routines to load and update the master data store of the company. This data then seamlessly moves and updates the various domain data stores for the domain users’ consumption. Once a particular milestone in the workflow is reached, the data automatically propagates to the master data stores and is available to other domain data stores that need this data. This loading, publishing, subscription, and the hydration mechanism ensure that a golden record of the company’s asset data is always available to domain applications. The individual domain data stores may or may not be a schema extension of the master data store.

Keeping the master data and domain data in one single store ensures that data quality is not degraded with data movement and data copy. Modern architectures of distributed and cluster computing make this scenario possible.

Some deployments, however, may require the data to be replicated for ease of mobility or performance, such as when the data is being used by a mobile workforce or the large teams across multiple geographic locations. The segregation of the domain data stores from the master data store may also enable the domain users to make edits and updates to the data that may not be feasible in a controlled environment, such as a master data store.
In a cloud environment, the visualization needs of the differing applications, such as 3D and high-end computing, may require the need to keep the applications distributed from domain and master data stores. Remote rendering and visualization streaming to any device from a private or virtual private cloud alleviates this need in some use cases.

**Enterprise Architecture and Platform**

Enhancing the delivery of IMaaS is the responsibility of the Enterprise Platform, where the approach is to federate domain and master data stores (Information Foundation) by providing application and data management access through an integration layer. This data is made available by the Integration Foundation layer to the Application Foundation layer that enables end-user visualization and deployment in a hyper-collaborative environment.

Finally, the Ecosystem Foundation—a cloud and role-based dashboard layer—summarizes knowledge and understanding in an easily and efficiently searchable workflow structure, particularly when this requires integration across proprietary and commercial applications and extensions supporting both private and virtual private cloud environments.

![Figure 3: E&P Enterprise platform](image)

The multi-tier enterprise architecture facilitates the deployment options needed to enable the flexibility needed by domain users and their visualization needs while delivering information as a service.

![Figure 4: Enterprise platform and the deployment flexibility](image)
IMaaS Options

Figure 5 represents an IMaaS solution where the master data management solution is in the private or virtual private cloud, while the domain data and application is available to the domain users in their devices—central servers, workstations, desktops, or mobile devices.

![Figure 5: Master data management in the cloud]

Figure 6 represents an IMaaS solution where the master data and the domain data stores are both available in the cloud, while the application is available to the users in their devices, central servers, workstations, desktops, or mobile devices.

![Figure 6: Master and domain data stores in the cloud]

Figure 7 represents an IMaaS solution in a software-as-a-service scenario where the master data, domain data stores, and the applications are all being managed in the cloud, and the visualization is available to users using remote rendering and streaming technologies.
Figure 8 represents a scalable deployment paradigm in petrotechnical computing where a single user, cross-domain application is running in a user’s device supporting a multi-tier architecture with strong integration and information foundations. This system frequently and automatically syncs with the IMaaS solution, if needed, with both virtual private or private clouds.

No matter the model, the IMaaS solution must adhere to the tenets of information management by providing the following data services:

- **Data Quality** – Continuously run data quality audits and checks to ensure the data in the master and domain data stores are not out of sync. Additionally, ensure data being loaded and synchronized is error-free, and conflicts are resolved using automated or manual techniques such as data rules, data mining, and statistics.
• Auto Synchronization—When data is copied into multiple data stores to enable mobility or performance, these data stores must constantly and automatically update and synchronize with other similar data stores and/or master data stores.

• Data Loading – IMaaS provides the ability to load bulk data from various public and private sources, as well as the ability to easily update or load data for the domain users. These data-loading techniques must include direct connectivity to the sources and the ability to load data using capability file-hosting of service where data is loaded in a particular Web or file store. Also, the data-loading routines must be able to automatically pull the data, manually load data for advanced users, and provide spreadsheet-like loading for end users.

• Cross-Reference and Metadata – Since the data is stored in various formats, and the applications need the data in varying formats, IMaaS provides full capability to first store the data in the format it is received, as well as to cross-reference data via units of measure, cartographic, and file format conversions in the form the application and the user expects.

• Data Search – IMaaS implements domain-specific search capabilities across the master data, domain data, and applications, as well as integrates with enterprise-wide data search of unstructured data and documents, but in the language and semantics of the domain user.

• Full Data and Workflow Data Audit – IMaaS implements a full data and workflow audit to be able to track the life cycle of data from acquisition to archiving.

IMaaS is designed for hyper-collaboration, the ability to deal with domain-specific data stores, and master data management. These features are made possible by implementing common data model and data federation. Without these capabilities, the deployment paradigm and characteristics of IMaaS discussed above are not possible.
Hyper-Collaboration

Hyper-collaboration is the ability to collaborate across domains and share the same data without the need to make multiple copies of data, resulting in a closed loop decision-making process. Implementing a messaging service that updates the integration layer when data changes in a domain application, in the Information Foundation, or in any domain data store or master data entity, results in a hyper-collaborative environment.

The use of real-time data and the ability of the IMaaS solution to triage the data using complex event processing and other predictive and classification techniques helps send relevant information in real time or near real time to required domain applications. It also helps to adjust and continuously optimize the process. For example, variations in assumptions in engineering and completions plans versus what is being read from downhole tools will help re-compute the engineering plans in real time and create a closed-loop system. Readings that indicate deviations from assumptions made in subsurface and fault positions must result in correction of the subsurface model or reservoir and formation model and re-compute the well placement and well designs.
Domain-Specific Data Stores

The current industry needs regarding agility and flexibility necessitate the targeted usage of a small number of widely used commercial database products in the various domains. These are often augmented by specific domain data stores and customer-developed IP. These databases serve multiple purposes, including:

- Domain data stores serving data to individual or multiple applications
- Central hub to consolidate information for collaborative use
- Preserved archive for securing longer-term definitive data for reporting and other purposes

IMaaS supports and enables these data stores and applications by passing along high-quality data and continuously monitoring and managing the data from the master data stores.

Master Data Management

Master data management is a requirement in most environments, particularly for some data classes, such as well, borehole, seismic, production, and interpretation. Figure 12 below illustrates the architectural components of the master data store for E&P.

The well master defines the unique identifier for the well in the company.

![Figure 12: Well master](image)

Key drivers for the presence of master data are:

- The need to synthesize multi-sourced information into definitive, quality-assured data sets. Several workflows in the borehole, production, interpretation, and seismic generate this requirement
- The creation of BI cubes for business analytics, in which underlying data sources are filtered to create static or real-time data volumes suitable for engineering and operational analysis.
- The requirement to retain core data associated with decisions and reporting that document auditable outcomes, necessitating preservation beyond the typical release cycle of a project or operational data solution.
- The ability to change overall workflows or applications and pull relevant data from master data and start the workflow again.
Seismic data management helps catalog raw and processed seismic files and formats for use by domain applications.

**Figure 13: Seismic data management**

Borehole data management helps catalog raw and processed log and core data formats for use by domain applications.

**Figure 14: Borehole data management**
Interpretation, engineering, and production master defines the golden record for the company’s asset—well, project, or field.

Integration and the Common Model
The common model is used to translate data models from underlying data stores and, in turn, convert them to an application model or industry-standard models. In Figure 16 below, two separate data stores, D1 and D2, store a well in different formats. These formats are both converted to a common model to represent a well for two separate applications, App 1 and App 2. At the same time, the common model is used to convert to an industry-standard model. The applications themselves might convert the common model to an application format, or retain it as a common model if the application was built using the common model.

The approach in exploiting a common or extensible common model provides flexibility and openness to address industry-standard data models, as well as customer/proprietary extensions associated with customer-developed intellectual property, workflows, and business processes.

The integration foundation is not an afterthought to integrate disparate data sources and applications, but is built into the foundation of the architecture, enforcing that all data moves through this integration layer.
Data Federation

The presence of the multiple data stores results in the need for data federation. The data services and integration layer, therefore, serves data from project, operational, and master repositories to domain applications and users’ role-based dashboards.

Data services are driven by common models with requisite services to simplify integration to underlying data stores and cross-domain workflows.
Considerations for Implementing IMaaS

IMaaS framework design and implementation requires strong stakeholder and executive support to manage the change needed to move from essentially a two-tiered architecture to a multi-tier architecture, and enforces strong corporate data governance. A portfolio management process is needed to focus the company’s efforts around core applications and data stores.

Data governance implementations vary by the maturity of the organization. It is essential to start IMaaS with a multidisciplinary team composed of architects, data managers, business analysts, and domain users. Focus the efforts of the team in defining and implementing the domain workflows using the integration layer. When implementing the workflows, ensure that the workflows are coded with flexible business process management tools, and provide the ability to make changes to the workflows easily. This helps identify data stores and applications and helps determine milestones and methods for data propagation into the master data stores. Select key entities to create strong master data-management practices for the company, and implement the common model around these entities. Identify and integrate the core applications required, and integrate with these entities using the common model. For companies that have not implemented master data management, it is desirable to first implement well and borehole master data, as these entities are the most common cross-domain master entities.

Start with master data management and analytics in the cloud before venturing to try and manage all core domain data stores and applications in the cloud.

Focus on the right balance between strong governance, user flexibility, and desired data intimacy. Letting the users have flexibility in dealing with the application domain stores would help data managers and IT focus on master data and other entities essential for cross-domain workflows.

Figure 18: Suggested responsibilities for defining scope and managing IMaaS
In Conclusion

IMaaS provides a framework for a comprehensive, flexible, standards-based, and lowest total-cost-of-ownership (TCO) solution to manage information in the upstream industry. It suggests implementing a closed-loop system to enable hyper-collaboration using a multi-tier architecture.

IMaaS enables the upstream E&P company to safely and securely manage information in a central cloud environment and allows domain users to focus on asset life cycle and reserves replacement.

For a company trying to make safe, fast, and accurate reservoir decisions, IMaaS provides balance between the flexibility needed by domain users while ensuring compliance needed for corporate data governance, performance, and scale at IT data centers.

References


Acknowledgements

Key Contributions – Joe King, Nagaraj Srinivasan, Duane Moonsamy, Janet Hicks, Gay Miller, Nick Purday, Dave Butcher, Dale Sperrazza, Dale Blue, Ted Kicey, Jeff Pferd, and Olivier Germain.

About the Author

Chandra Yeleshwarapu, Director of Information Management and Platform Technologies, is responsible for evolving and delivering the DecisionSpace® platform, the energy industry’s first complete, open, and modular platform for geosciences, engineering and completions, and production software. The key tenets of the platform are interoperability, integrated workflows, performance, and scalability, and it provides deployment choices such as mobility, private, and virtual private cloud.

Prior to joining Landmark in 2011, Yeleshwarapu was Senior Director of Solutions Management for Ariba (now SAP), a Cloud commerce company, and spent 10 years with Oracle in a number of senior positions, including Director of Product Strategy and Product Management. Prior to Oracle, he was with Tata Group as a consultant at Tata Consultancy Services in their Manufacturing and Process Industries Practice and head of the Industrial Engineering Department at Tata Press/Tata Donnelley.

Yeleshwarapu is co-author of four U.S. patent applications. He earned his bachelor’s degree in industrial engineering from Pune University in India, his MS in software engineering from Carnegie-Mellon University, and his MBA from Northwestern University's Kellogg School of Management.