



El camino de valoración de los datos

Por **Leandro Chammaá** (YPF S.A.)

Este trabajo fue seleccionado en las 3^o Jornadas de Revolución Digital para Petróleo y Gas.

La evolución del Data Management está intrínsecamente asociada a las nuevas tecnologías, a la modernización de quienes trabajan en la disciplina y el crecimiento de los lineamientos de gobierno que pregonan por las mejores prácticas sobre los datos. En este marco, la gestión de datos intermedios, el data staging, permite que diferentes actores que participan del ciclo de vida exploten o aumenten el valor de los datos hasta su resguardo definitivo. En este trabajo se exponen casos de aplicación operativa con datos técnicos y procesos del Upstream, con una mirada práctica y conceptual.



Valorización de datos desde su adquisición hasta su resguardo

La valorización de los datos comienza desde el momento de su adquisición. En la etapa inicial, los datos crudos deben ser recolectados con la mayor precisión posible, utilizando tecnologías avanzadas que aseguren su calidad. Herramientas como Python son fundamentales en esta etapa para la automatización de procesos y análisis preliminares. Posteriormente, estos datos deben ser validados, procesados y almacenados adecuadamente. Aquí, plataformas de almacenamiento intermedio juegan un papel crucial en el resguardo temporal hasta que los datos sean adecuadamente curados, y se asegure que permanezcan íntegros y accesibles para futuros análisis en las bases corporativas. Durante este proceso, la interacción con proveedores de información es vital para asegurar el flujo de los datos obtenidos. Las herramientas livianas de automatización pueden facilitar esta interacción, optimizando parte de la recolección de datos críticos y la actualización de datos provenientes de diferentes fuentes. En cada etapa del proceso, desde la adquisición hasta el resguardo, se añaden capas de valor a los datos, aumentando su utilidad y relevancia para la toma de decisiones.

Importancia del manejo de los datos intermedios y su crecimiento en importancia a medida que se tratan

Los datos intermedios, aquellos que se generan y utilizan entre la adquisición inicial y el resguardo/análisis final, son de creciente importancia en la industria del petróleo. Estos datos intermedios pueden incluir resultados preliminares de análisis, datos de campo temporales y registros de mantenimiento. Su correcta gestión y tratamiento incrementan su valor, ya que proporcionan una base sólida para análisis a corto plazo. Herramientas de tratamiento liviano como R o Python, permiten el procesamiento y extracción de estos datos, mejorando su calidad, relevancia y agrupamiento. A medida que estos datos se tratan, su madurez aumenta, permitiendo una mejor interpretación y toma de decisiones. Además, es crucial manejar distintos formatos y familias de datos, tanto datos estructurados como no estructurados, para asegurar su disponibilidad y potencial análisis. Hay casos donde datos aun no gobernados tienen valor en sí mismos, y detectar esa oportunidad de explotación rápidamente es clave para el desarrollo de repositorios intermedios que permitan en primer lugar almacenarlos, comenzar a curarlos y explotarlos tempranamente. Se entra así en un ciclo de captura de valor incremental a medida que se trabajan los datos, con el data management adecuado.

Rol de los data managers en el incremento de valor para llevar a los datos a condición bronce, plata y oro

Los data managers juegan un papel crucial en la valorización de los datos. Se encargan de transformar los datos desde su estado bruto hasta alcanzar niveles más altos de calidad y utilidad. En la condición bronce, los datos son inicialmente validados y organizados, permitiendo rapidez en la consulta inmediata, aunque con

Problemática

Contexto del manejo de datos técnicos en la industria del petróleo

En la industria del petróleo, el manejo de datos técnicos es fundamental para la operación eficiente y segura de las actividades extractivas y de producción. La cantidad y calidad de los datos obtenidos, que van desde sísmica y registros de pozos hasta análisis de fluidos y datos de producción, determinan en gran medida la capacidad de los equipos técnicos para tomar decisiones informadas y bien contextualizadas. Estos datos son críticos no solo para la explotación y el desarrollo, sino también para la optimización de la producción, la gestión de riesgos y la sostenibilidad operativa. La gestión de la información en tránsito, es decir, los datos que se transfieren entre diferentes sistemas y ubicaciones, es esencial para garantizar que sean accesibles y útiles en todo momento.



ciertas limitaciones reconocidas por los técnicos. En la condición plata, se integran y procesan para asegurar consistencia y fiabilidad. Herramientas de automatización low code pueden ser utilizadas para automatizar la validación y el procesamiento de datos, incrementando la eficiencia y reduciendo errores, y liberando a técnicos de una tarea repetitiva. Finalmente, en la condición oro, los datos son completamente validados, analizados y preparados para su uso en aplicaciones críticas. Los data managers aseguran que cada etapa del ciclo de vida del dato entre en un proceso de valorización, que se ejecute de manera eficiente y efectiva, utilizando diferentes plataformas para organizar y acceder a los datos de manera segura, íntegra y sostenible.

Data staging, validaciones y buenas prácticas

El data staging, o la preparación de los datos para su análisis, es una etapa clave en el proceso de valorización. Esta etapa incluye la limpieza de datos, la eliminación de duplicados, la corrección de errores y la normalización de los datos, entre otros procesos de data management. Las validaciones son esenciales para asegurar la integridad y precisión de los datos, y las buenas prácticas, como el uso de estándares, que garantizan que los datos sean confiables y útiles. Hay varias herramientas poderosas para el data staging (PowerShell, , permitiendo la optimización de la limpieza y validación de datos. Implementar buenas prácticas en el data staging y las validaciones ayuda a maximizar el valor de los datos y su utilidad en el análisis y la toma de decisiones. Trabajar con el negocio para entender los criterios técnicos que permitan validar automáticamente datos es fundamental para que este proceso se ejecute solo mediante reglas o bien mediante aprendizaje automático.

A su vez, entre las buenas prácticas, se encuentra la remoción de los repositorios intermedios de los datos que ya alcanzaron su calidad oro y fueron resguardados definitivamente en los repositorios corporativos. Esta práctica es muy importante para evitar duplicaciones y degradaciones de edición en el camino del dato, o bien confusiones en su utilización.

Desarrollo técnico

A continuación, se expondrán una serie de casos testigo cuyos resultados han agregado valor a los datos en

su camino de validaciones hasta el resguardo definitivo, que han permitido a los usuarios del Upstream tomar decisiones o contextualizar información, ya sea desde el punto de vista técnico como del operativo.

1. Captura de metadatos de perfiles en formato PDF:

Para una rápida catalogación de información disponible, que aceleró el proceso de interpretación: En muchas operaciones, los perfiles de pozos y otros documentos técnicos se encuentran en formato PDF. Utilizando Python, se desarrollaron scripts para extraer metadatos de estos documentos automáticamente. Esto permitió una rápida catalogación y organización de la información, acelerando el proceso de interpretación por parte de los técnicos. La mejora en la eficiencia resultó en una toma de decisiones más rápida y precisa, optimizando la operación y reduciendo tiempos de espera.

2. Rápida incorporación de datos de bombas mecánicas para comparación de equipamiento disponible mediante visualización en tableros BI:

La integración de datos de bombas mecánicas provenientes de diferentes proveedores se realizó utilizando herramientas low code e IA para automatizar la recolección y actualización de datos. Estos datos fueron luego visualizados en tableros de Power BI, permitiendo a los ingenieros comparar rápidamente el rendimiento y disponibilidad de diferentes equipos. Los datos estaban incrustados en archivos adjuntos enviados por email en distintos formatos y llevarlos a un repositorio común intermedio fue fundamental para el proceso. La capacidad de visualizar y analizar estos datos en tiempo real mejoró la selección de equipos y la planificación de mantenimiento, resultando en una mayor eficiencia operativa y reducción de costos.

3. Carga masiva de datos de Legajos a Dataverse en condición bronce, sin los cuales no sería posible su explotación ni tratamiento:

La carga inicial de grandes volúmenes de datos crudos en Dataverse permitió establecer una base sólida de información en condición bronce. A pesar de sus limitaciones iniciales, estos datos brutos son esenciales para

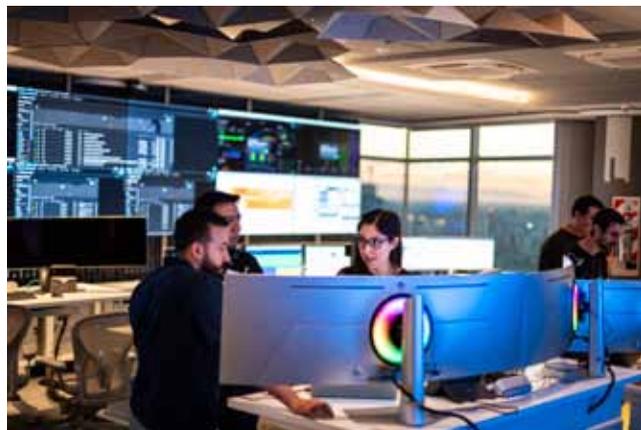


el análisis y la validación posteriores. Los data managers utilizan Dataverse para organizar y almacenar estos datos, facilitando su acceso y tratamiento. Sin esta carga masiva inicial, no sería posible explotar ni tratar los datos de manera efectiva, subrayando la importancia de gestionar adecuadamente los datos desde su etapa más temprana.

Resultados obtenidos

Los resultados obtenidos son muchos más que los representados por los casos expuestos. Entre los que figuran en el siguiente listado se incluyen resultados y beneficios de estos y otros casos de éxito en la valorización del dato hasta su resguardo definitivo:

- Aceleración de los tiempos de interpretación de datos técnicos raw
- Toma de decisiones con mayor contexto gracias a datos bronce
- Reducción de los tiempos de espera para contar con nueva información
- Optimización de las operaciones, destrabando la falta de datos cuyo ciclo de vida no concluyó aun
- Visualización y análisis de datos en tiempo real con tag “No validado” que permitió a los técnicos incorporarlos al cálculo o no, según necesidad
- Mejora en la precisión de la selección de equipos
- Optimización de la planificación de mantenimiento
- Identificación temprana de disponibilidad o no de equipos
- Mayor eficiencia operativa y reducción de costos por tiempos muertos sin información
- Consulta de información digital rápidamente vs chequeo presencial
- Organización eficiente de datos brutos para captura de metadatos
- Facilitación de acceso a los datos para los equipos de data management para su tratamiento posterior
- Revalorización del dato bronce: “al menos el dato está”
- Disponibilidad para analíticas y explotación en tableros
- Aceptación del riesgo de utilización de datos no



validados plenamente por sobre el tiempo de espera para completar el proceso

- Foco en la operación para los datos intermedios (Bronce/Plata), foco en analítica y mejora de procesos para datos definitivos (Oro)
- Entornos para pruebas de optimización de procesos de curado
- Mejora continua en la captura de valor de datos intermedios, a través de mejores prácticas de manejo de datos aplicadas a diferentes familias y proyectadas a otras
- Monetización del valor capturado por la puesta a disposición de datos rápidamente, basado en horas/hombre de los diferentes perfiles técnicos beneficiados.

Conclusión

El dato gestionado, en cualquier etapa del proceso, ya aporta valor y puede ser explotado.

En conclusión, la gestión técnica adecuada aumenta significativamente el valor de los datos desde su adquisición hasta su resguardo a medida que se tratan. Los datos gestionados correctamente pueden ser explotados en cualquier etapa del proceso, proporcionando insights valiosos, dando mayor contexto y mejorando la toma de decisiones en una industria petrolera cada vez más volcada a un modo data driven. La valorización de los datos es un proceso continuo y acumulativo, donde cada etapa de tratamiento añade valor, convirtiendo los datos en activos estratégicos una vez que son gestionados con sentido de negocio y puestos a disposición de los objetivos definidos, e incluso para objetivos aun no establecidos o visualizados al momento de su adquisición. La capacidad de manejar información en tránsito, interactuar eficazmente con proveedores de información y gestionar diversos formatos de datos garantiza que los datos sean un recurso estratégico y explotable en todas sus etapas.

Herramientas tecnológicas low code, de código abierto y de fácil utilización, tanto para gestión como para almacenamiento intermedio, son fundamentales en este proceso, facilitando la adquisición, validación, procesamiento y almacenamiento eficiente de datos, pero sobre todo su temprana captura de valor mediante su uso y explotación.